Navigating Interdependence

Helping people see connections in a complex world

Erica Gunn

Thesis Presented by

Erica Gunn

to The College of Arts, Media & Design
in Partial Fulfillment of the Requirements for the Degree
of Master of Fine Arts in Information Design & Visualization

Northeastern University
Boston, Massachusetts
April 2017

Thesis advisor: **Ann McDonald**

Readers: **James Grady, Daniel Muller**

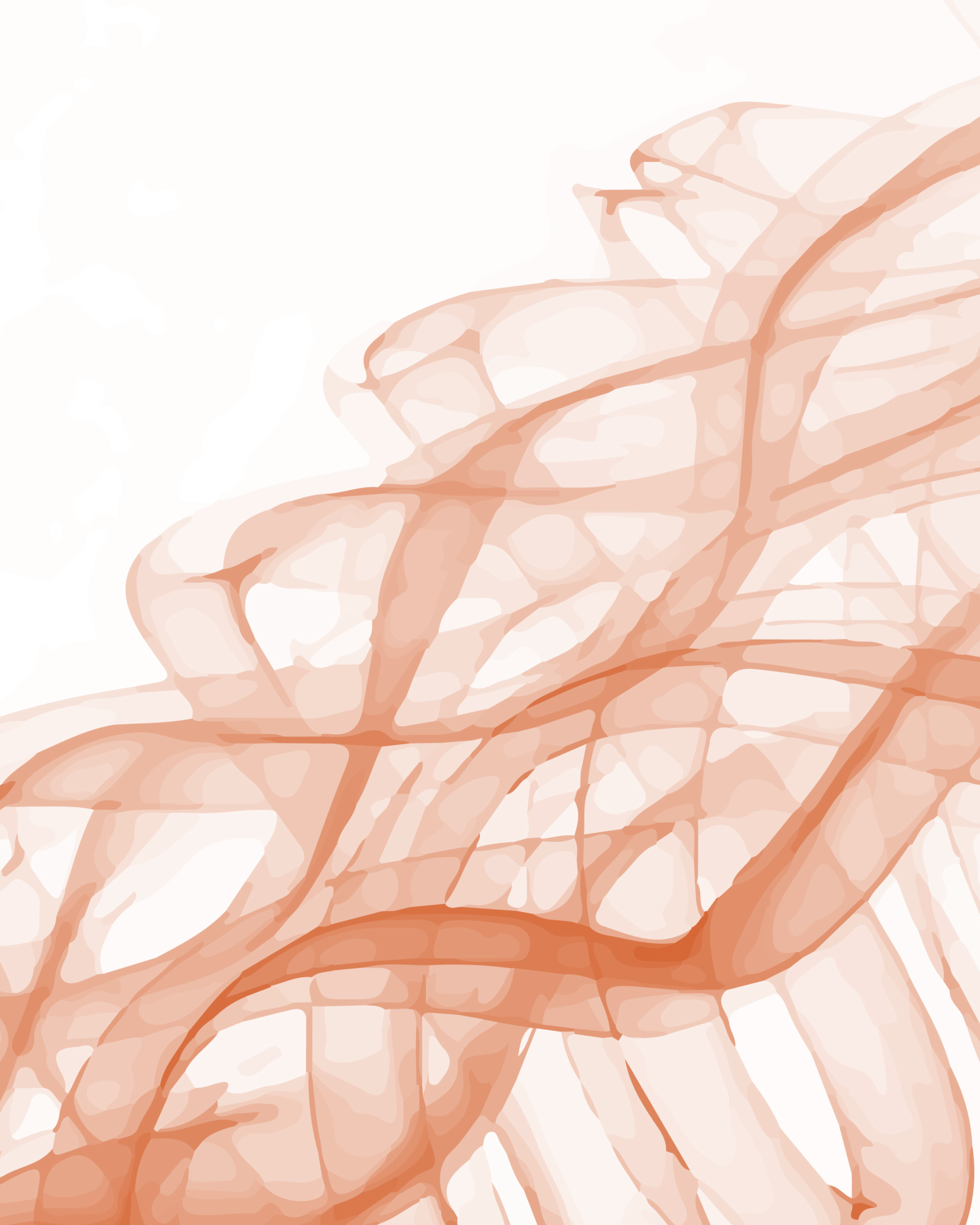
Respondent: **Dietmar Offenhuber**

Thesis Committee: **Pedro Cruz**

Advising faculty: **Thomas Starr**

Abstract

Complex systems are inherently interdependent and must be understood across many length scales, making them difficult to define and communicate. Scientific questions are often complex in nature, and can be difficult to convey to a public audience for this reason. Communication is often further confounded by institutional barriers, cultural attitudes, and occasionally distrust between scientists and the public. Designers are ideal partners to work with scientists in bridging these barriers, exposing conceptual frameworks and facilitating dialogue. This thesis reports the development of an interactive website that explores issues related to soil health, global land use,and population growth. The design follows a series of recommendations identified through a broad survey of recent work in information visualization, including specific guidelines for the tone, structure and content of the piece, as well as its support of user agency. A participatory exhibition project further investigates the potential for design to foster conversation and create community.



Acknowledgments

It takes a lifetime to create a piece of art, and this one bears the fingerprints of many makers. I'm sure I've left someone out. My apologies.

My family's wildly diverging viewpoints taught me just how many ways there are to look at the same thing. Experience with their different perspectives resides at the core of this thesis. My mother, my grandfather, Nita, and the ladies of the church plant sale initiated me into both the mysteries and the hard work of the garden at an early age. Soil has seemed like a miracle ever since.

Len found the scientist switch in me, turned it on, and gave it a name. Most of what I know about teaching comes from following his good example. Rich first turned my awareness to sustainability and environmental issues, an interest that continues to this day. Bart and Lian guided me in the ways of scientific research, instilling a deep respect for data, ethics, and the scientific process along the way.

For 15 years, my students gave me constant lessons in clarity, the importance of organization, sequence, and narrative, and taught me how (and how not) to appeal to an audience that is constantly wondering if science is really worth the effort of doing homework, anyway.

James kindly encouraged me when I was looking for a new career, and has paid the price of feeding a stray puppy ever since. I owe him a double gratitude for also agreeing to read this thesis, and for shaping its contents with his good advice.

Doug and Ernesto guided my first faltering steps into design. I didn't always believe that their suggestions would work, but they (almost) always did. My work is stronger, clearer and more beautiful because of them. Dietmar started this whole crazy thing by asking me to take an object apart and really look at it. I picked through a shovel full of dirt, and have been following it ever since. Siqi, Pedro, Noah, and Mauro showed me how to wield the black magic of code, and helped me build cool things with it. Pedro also created the space and the structure in which the bulk of this writing took place. Hugh's design theory class gave me a language to discuss the similarities between science and design, and pointed out how much designers and scientists have to offer one another. Miso and Alessandra helped me to think beyond an information design project and see how I might use this thesis as an opportunity to build community and foster dialogue. My thesis advisor, Ann, has always celebrated big ideas and encouraged me to go wherever the project took me. Her enthusiasm helped me take this further than I might have done alone. Paul started me thinking about the different ways that graphics are used, and was also instrumental in developing the exhibition portion of this project.

Daniel helped me plan out the last and the most complicated information design piece of the website. Having a very clear plan from the outset made it possible for me to complete the code part of the project on time.

J and T would want to be anonymous, but they have been constantly in my mind as representatives of my target audience. Their curiosity, enthusiasm, and everyday common sense questions have been invaluable in helping me see this project from the outside.

My fellow students challenged, supported, and inspired me every day. Watching their ideas develop in real time taught me more than classes ever could.

My husband Branden doesn’t always quite know what to do with my many projects, but is always excited to help. His quiet work in the background has helped me to make this project all that it could be.

My ideas about complexity were shaped in myriad ways by long walks through the Harold Parker State Forest, a beautiful place that represents so much of what I hoped to convey. It served as much as an advisor to me in this project as any human. I hope the result is worthy of its inspiration.

Many people participated in the exhibition project by sending in soil samples from all over the country. What started as a simple need for bags of dirt has now begun to sprout into the beginnings of a community. I am excited to see where it leads.

Preface

This thesis tackles a broad topic, and is by necessity written for a variety of audiences. I expect that the thesis itself will be most helpful to scientists seeking to improve their communication with a lay audience, and designers hoping to understand how they can contribute to this effort. As such, the first half of the book is both a practical guide and a call to action, an argument for the value of design thinking in scientific communication, and a discussion of the special considerations designers must make when they seek to engage with science.

Any thesis is written for an academic audience, but I hope that this one will be accessible to practitioners in both disciplines as well. The thesis project itself is both a personal exploration of the design space that has informed my thinking on these issues, and also an imperfect demonstration of the concepts that I hope to convey in the rest of this writing. Primarily, though, it is a product for public consumption. In adopting the methods and approaches outlined in the thesis document, I hope to convey the wonder, complexity, and implications of science to a general audience.

The thesis project addresses people who are perhaps curious about science but don't know much about my chosen topic: soil. I hope that this visualization will encourage people at all levels of scientific literacy to engage with something that we all see but seldom notice – the ground beneath our feet. Soil is a fascinating, complex ecosystem with broad implications for the future health and wellbeing of our planet. It intersects with many issues of scientific relevance, but it is also a human issue, directly related to questions of social justice and the survival of our species. As a result, it is also intimately tied to and affected by issues of public policy, and spans a number of different domains. I believe that re-integrating the discussion of science, human issues, and public policy poses a unique challenge and opportunity for designers, and hope to convey some small part of that possibility here.

On a more personal note, I was raised in a rather extreme religious household that adopted both strict creationist and anti-vaxxer views in the 1980s—before those viewpoints became as widespread as they are today. As an adult, I rejected many of the ideas that I was raised with, and went on to get a PhD in chemistry, which I taught at the university level for several years. I left the practice of science in 2015, in part because I sought a wider reach for the science that I taught in my classroom. I remain firmly allied to the ideals of scientific thought, and believe that it is both our responsibility and our privilege to convey this information outside of the halls of academia. I also recognize the unique challenges of appealing to a scientifically-resistant audience.

I came to design because it shares many characteristics with the teaching of science. Though the tools we use are often different, organizing information into coherent narratives that support understanding is a skillset that prevails in both worlds. In changing disciplines, I have come across many aspects of design that can inform a scientific audience, and that I believe will help in the urgent work of bridging the widening gap between scientific understanding and public response. As someone who has lived in these disparate worlds, I hope to offer a unified perspective that can begin to dissolve some of the seemingly impenetrable barriers that needlessly divide us.

Photo credit: Pexels.com

Contents

A complex system [is one] in which large networks of components with no central control and simple rules of operation give rise to complex collective behavior, sophisticated information processing, and adaptation via learning or evolution.

Melanie Mitchell

CHAPTER ONE

Introduction

Complex issues emerge from interdependent systems and require understanding of interactions at multiple levels, often requiring integration and synthesis of information across multiple domains. Designers are uniquely qualified to help scientists communicate complex topics to a lay audience.

A complex system is more than the sum of its parts. Relationships between interdependent components create synergies and feedback loops that cannot be understood unless we consider the system as a whole. Weather patterns, telecommunications networks, human intelligence, social systems, ecosystem function, and even life itself can be considered examples of complexity that we encounter every day. Understanding how these systems function and relate to human society is becoming increasingly important for making informed decisions in an interconnected world.

Unfortunately, complex systems are notoriously difficult to understand. Phenomena often occur across many different length and time scales. In order to see how the pieces fit together, subsystems need to be understood at multiple levels of detail and abstraction, often switching back and forth between mental models to gain a more comprehensive view. The information necessary to understand the entire system seldom exists within a single discipline or specialty, and so progress requires collaboration and formulation of shared language and conceptual frameworks before people can even begin to discuss an issue. The initial starting conditions or context for the problem usually play a dominant role: one size does not fit all, and what works once will not necessarily work again, especially if the starting conditions are different. Subtle changes often have large effects, but it can be difficult to predict the knock-on results of those effects in advance. Additionally, complex systems are best described using probabilities, which can be difficult for people to grasp.

Designers are uniquely positioned to provide assistance to scientists in communicating these complicated ideas. Despite their surface differences, similarities in approach and ethos between the two disciplines make them natural allies. Scientific data provides an interesting challenge for designers to develop their skills and stretch in new ways.

Visualization of complex, interdependent systems requires that we move beyond self-contained visualizations of an isolated dataset to an ecosystem model that clarifies the connections between disparate topics and multiple datasets. With a clear user interface and information architecture that support exploration and provide guidance, such a system can help people to navigate through a complicated issue at varying levels of detail to gain a better understanding of the system as a whole. Designing such systems requires different ways of thinking about narrative, structure, and form, as well as creating more integrated and adaptive experiences for the user.

Designers' experience in identifying and reconciling divergent or conflicting conceptual frames could also prove invaluable to scientists seeking to convince the public of the need for concern and directed action, especially for controversial topics. Understanding the reasons for public distrust of science and the cognitive processes that support everyday decision-making are key steps for those seeking to change modes of thought and behavior. Designers are also expert in communicating information clearly to a lay audience, and can help scientists to present their case in an appealing and understandable way.

To overcome public mistrust of science, scientists must move beyond a knowledge-dispersal model to one based on two-way discussions. In such a model, scientists and the public would collaborate as partners in the process of making sense of scientific information, especially as it relates to policy decisions. Well-executed design can facilitate meaningful, bidirectional conversations that go beyond didactic explanation or fun facts to create deeper engagement with both the process and the results of science. These conversations should educate the public about science, improve public appreciation for the meaning and limits of data, and also help to inform scientific research and policy agendas.

In this thesis, I will examine a broad base of visualizations spanning design, technical infovis, art, and graphic design to identify the features that are most important to science communication, and particularly to complex systems. From this collection of visualizations, I will build a set of guidelines, or best practices, for creating graphics that support scientific communication. I will compare these guidelines with current practices, synthesizing approaches from several different fields, such as science, education, journalism and design. By bringing these diverse perspectives together, I hope to create an integrated view of the role of design and information visualization in communicating science, especially to a lay audience.

To explore these principles more deeply, I will create a series of visualizations that allow users to explore multiple factors related to soil ecosystems, human population, and food production. By uniting these different perspectives in an exploratory website, I hope to promote awareness of this valuable ecosystem, and of the threats posed by human demands. Additionally, I will use the insights gained from this experience to further refine a list of considerations for designers seeking to support people in understanding complex systems.

Organisms are not billiard balls, propelled by simple and measurable external forces to predictable new positions on life's pool table. Sufficiently complex systems have greater richness. Organisms have a history that constrains their future in myriad, subtle ways.

Stephen Jay Gould

Everything should be made as simple as possible, but not simpler.

Albert Einstein

CHAPTER TWO

The Challenge of Complex Systems

Complexity poses several unique challenges. Issues of scale, debates over how to define a problem, differing attributions of causality, and difficulties with communication are a few of the issues that accompany interdependent systems.

Many of the challenges that we face as a society are multi-dimensional problems with implications for political, business, cultural, and environmental worlds. These issues often span many different contexts, providing myriad opportunities for disagreement over both cause and effect. The inherent interdependence of complex systems means that apparently local topics often turn out to have global implications, and vice versa, introducing questions of scale that lead to arguments over responsibility. Many of these debates are caused by differing definitions of the problem and its appropriate solutions, and of its magnitude and natural scale. Because the statement of a problem affects the level at which solutions should be proposed, lack of a single clear driving factor or responsible party can lead to confusion, disagreement, and entrenchment as opposing sides argue for their particular view. Competing definitions and conflicting perspectives also muddy the conversation and can lead to frustration or disengagement.

These barriers to communication are neither new nor surprising; they have been with us in some form since the dawn of civilization. Horst Rittel voiced a city planner's frustration in 1972 by calling these "wicked problems"; issues without clearcut solutions, where individual participants are not able to see eye to eye, even on the definition of the problem itself (and sometimes disagree on whether the problem exists at all) (Rittel, 1972). Recognition of "wicked problems" in Rittel's day laid the groundwork for studies of human motivation and decision making that erode the idea of humans as rational actors, and reveal that most people use culturally-encoded heuristics and

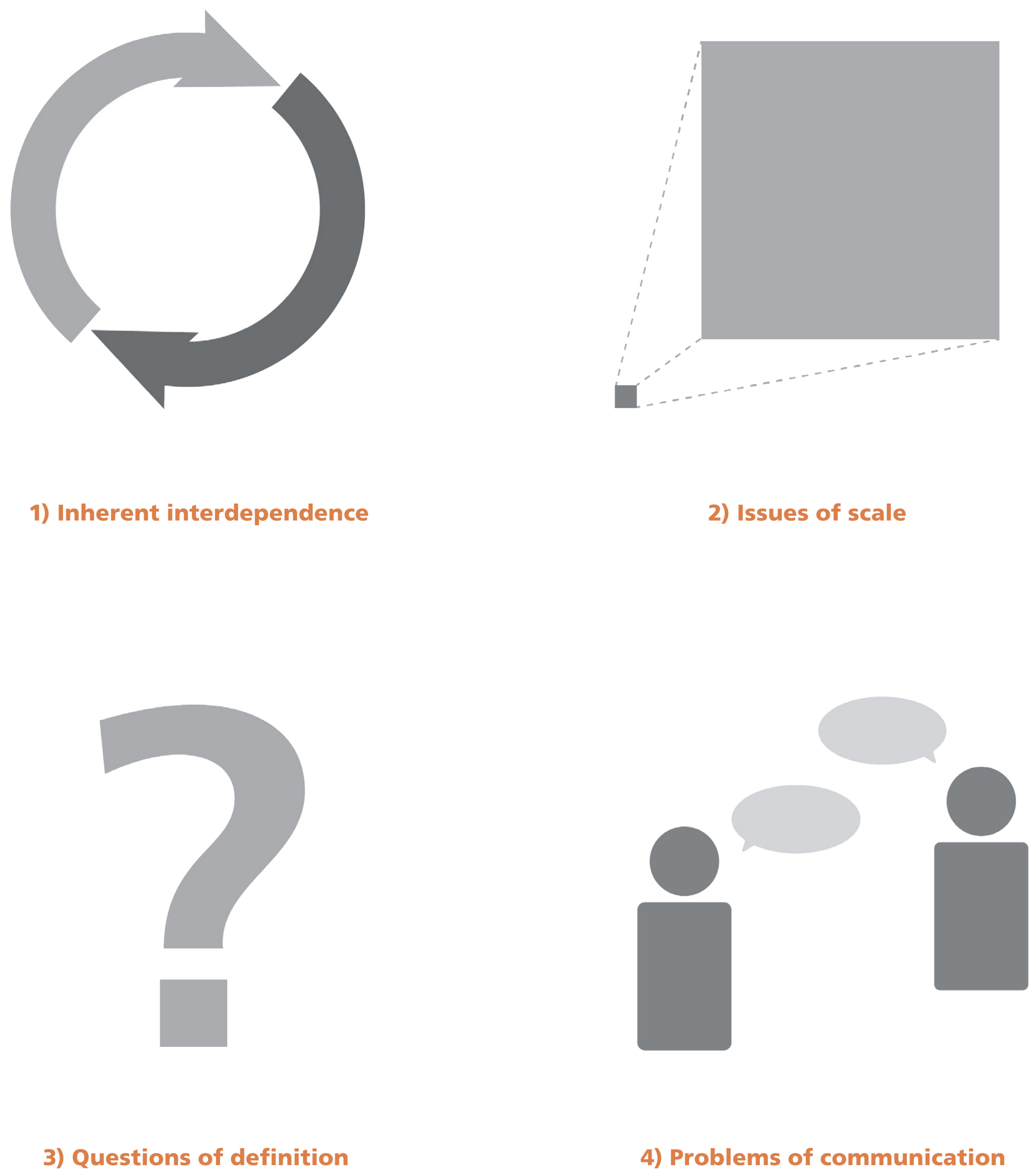


Figure 1. Complex systems are characterized by several challenges that pose specific requirements for design.

value judgments when making decisions, especially in cases of high uncertainty. These differing ideological or cultural frames are often unconsciously held, and may even be in direct conflict with a person's stated beliefs, making them extremely difficult to navigate. Designers use structured conversation and co-creation to uncover these unconscious biases and reveal the underlying conceptual models. Making these connections explicit can establish a common ground for conversation and comparison, and can help to break the deadlock that results from faulty or incomplete communication.

In fortunate cases, the data for one particular piece of a problem is fairly well understood and consensus can be reached on that one aspect of an issue. Still, it is often difficult to connect the results or conclusions from an isolated dataset seamlessly to knowledge from another domain. This difficulty stems from differences in disciplinary focus and methods of encoding, as well as the perspectives and disciplinary norms that each set of participants brings to the table. These intellectual habits are often deeply entrenched, and inform the methods, inquiries, data encoding, and boundary objects that represent a discipline. These ideas are often connected to a sense of personal or disciplinary identity and are thus resistant to change. Rather than reflecting a single issue or area of expertise, solutions to these topics require integration of knowledge across many domains. Conflicting or incompatible models, cultural differences, and disciplinary prejudices serve to complicate communication and can hinder attempts at synthesis.

Bridging these divides and negotiating a common language is a critical first step in establishing productive conversations around a topic. Designers are accustomed to working in a "mesoscopic" range of information: neither entirely general, nor overly detailed. Scientists and other disciplinary experts usually work at an atomic level of detail and can struggle to communicate at a higher level of generalization. The public, on the other hand, tend to focus on the general and have limited interest in specifics, except those that are perceived to be directly related to the problem at hand. Design can help to bridge these different levels of detail by creating structured environments that facilitate the exploration and understanding of complex issues.

A popular cliché in philosophy says that all science is pure analysis or reductionism, like taking a rainbow to pieces; and art is pure synthesis, putting the rainbow together. This is not so. All imagination begins by analyzing nature.

Jacob Bronowski

CHAPTER THREE

Communicating Science

Communicating science is both a technical and a human problem, and must be understood from both sides. This chapter enumerates some of the difficulties facing those who wish to communicate science to a general audience. Now is a perfect time to reconsider the nature of discourse between scientists and the public, and to improve methods of communication to support ongoing conversation.

This chapter explores some of the difficulties facing those who wish to communicate science, in preparation for the next chapter, which focuses on how designers and scientists can work together to overcome these obstacles. Science is often central to addressing the complex societal problems that we encounter today. Whether we consider questions of public health, environmental problems, or technological advance, scientific understanding permeates our view of the world.

However, there are many challenges facing those who seek to communicate science to the public. First is the inherent difference between the level of detail required to fully understand a problem at its most fundamental level and the zoomed-out view required to allow people to see broader connections. Communicating across these different scales while retaining the most important details and establishing a coherent argument presents a complicated technical problem. The second major challenge to science communication is the need to operate within the mental models and cultural framing of different worlds. Differing value systems and conceptual framings must be acknowledged and reconciled before real communication can begin.

Historically, science communicators have focused on the first problem, working under the assumption that clearer communication of technical information would be sufficient to convince the public on scientific issues. Failure to recognize that many of these disagreements arise from challenges of the second kind leads to ineffective solutions and increasing disengagement.

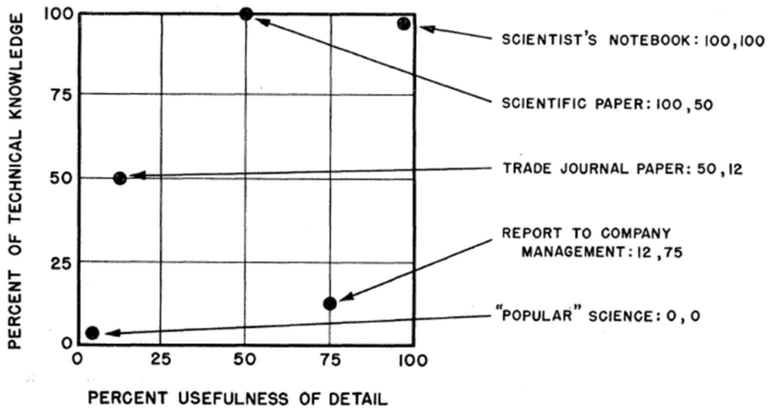


Figure 2. Graphic from a letter to the editor of the journal *Science*, showing the content recommended for different publication venues.

Bridging Distrust

Establishing robust science communication requires that we find ways to bridge the deep and rapidly widening distrust between scientists and the public. This distrust stems from a variety of factors, including historical events, cultural attitudes, and differing frames for assessing scientific outcomes. Both scientific and public attitudes contribute to the problem, and solutions must be found by changing perceptions on both sides.

Scientific attitudes

Scientists have not always considered public communication to be either possible or desirable. This reluctance is apparent in an irascible letter to the editor of the journal *Science*, written (perhaps somewhat tongue-in-cheek) in 1958 by M.W. Thistle—who was at the time the chief public relations officer for the National Research Council of Canada. Figure 2 reproduces an image from Thistle's letter, intended to show the amount of detail and technical knowledge that should be included in a paper on popular science: "science for laymen is in the area of 0,0: zero technical knowledge, and zero detail. Scientists please note."

Thistle presents several barriers that prevent the communication of science in a more general context: first, he cites the difficulty of symbolic encoding; the difference between physical objects or phenomena and the words that we use to describe them. It is difficult to find scientists who can communicate well, and the language of science is more precise than that of English, presenting a language barrier that loses details in the translation, usually in favor of generalities.

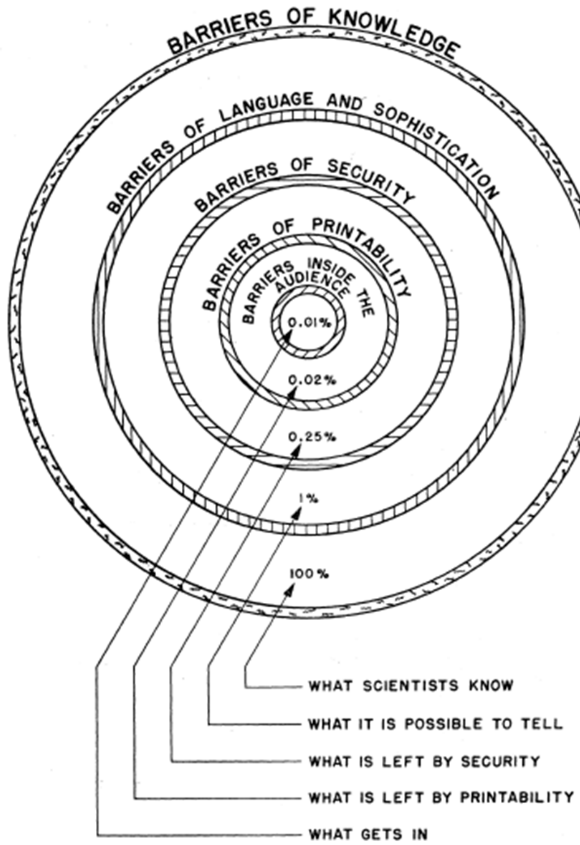


Figure 3. Visualization showing the barriers to scientific communication, according to M. W. Thistle.

Next, Thistle divides scientific literacy into 5 levels of understanding, and suggests that it is extremely difficult to communicate across the barriers created by the difference between levels. As an antidote to this apparently hopeless situation, Thistle states that we should choose to link scientific work to something that is well known to the non-specialist. Within the frame of this common interest, he suggests that one might manage to transmit a few broad trends and general conclusions.

The next barrier that Thistle identifies is that of security. He was writing in a time where atomic physics had recently changed the shape of military power, and at the height of the cold war between the US and Russia; it is perhaps not surprising that he felt it necessary to say that science might need to be silenced by concerns about national security.

The next barrier was printability, or the difficulty of pitching stories of scientific progress as a news item. On the whole, Thistle estimates the accuracy of transmission to a layperson to be one ten-thousandth of what scientists actually know. To achieve any level of communication, he recommended minimizing or removing the true scientific content.

Thistle took this position in response to a perceived public disinterest and inability to understand science, and these expectations became in many ways a self-fulfilling prophecy. Widespread oversimplification disengages the public from the process of science and makes it impossible for lay people to critically evaluate or engage with scientific results. With such attitudes prevalent and openly expressed at the highest levels of scientific institutions, long-term disaffection of the public audience is perhaps not surprising.

The attitudes in Thistle's letter are jarring to the modern ear, but their echoes are still recognizable in the attitudes of some members of the scientific community today. In particular, his framing of public communication as a necessary "report to management" in exchange for public funding is still widely held. The National Science Foundation added a required (and controversial) "statement of broader impacts" including teaching and public outreach as part of all funding proposals in 1997. This initiative was intended to recognize the efforts of scientists who do work to popularize their research, and to create incentives for others to do the same. In a relatively short period of time, the scientific community

People will forget what you said. People will forget what you did. But people will never forget how you made them feel.

Maya Angelou

has moved from a culture where it was considered impossible and even undesirable to communicate science to the public, to a culture where scientists are now required to plan outreach into their research, and often to publish their data online. This is a sea change in perspective, and the scientific community is still working to develop a different understanding of what a conversation with the public might mean.

The deficit model

In discussions that rely on scientific knowledge to inform policy and improve decision making, it is not uncommon to encounter a "deficit model" of science communication. This framing assumes that the public must be educated in order to remediate problems and prevent people from making bad choices (Nisbet and Scheufele 2009). This framing implicitly assumes that the audience is uninformed and in need of correction. In this model, the scientist shows up to fix misconceptions and disburse unambiguous facts, rather than to engage in a dialogue. This approach relies on universal acceptance of science as the ultimate authority and highest standard of truth, capable of overriding all personal experience and previous beliefs. It also assumes that humans are rational decision-makers who can be led to desired actions by logic alone. There is mounting evidence that this approach is not only ineffective, but may actively damage the public relationship with science (Niepold, Herring, and McConville 2007).

Rittel discussed exactly these issues in his 1972 paper (Rittel 1972). He argues that some problems are "simple" issues of fact, and have clear design solutions. Other problems, especially those involving issues of public policy, are more complicated to solve because people approach the problem from radically different frames. Not only does this lead to different perspectives on what the solutions should be, it also reflects fundamental differences in how the problem is defined and how causality is attributed. Rittel takes a particularly pessimistic view of these kinds of problems, calling them "wicked problems" that cannot be solved. In this context, he highlights the importance of conversation and continued engagement of both sides in a negotiation that seeks to establish common ground. Nisbet and Scheufeule call for a similar approach to science communication, identifying two-way conversation between scientists and the public as a critical component of improving the public perception of science, and in linking popular decision-making to scientific results (Nisbet and Scheufele 2009).

Navigating power structures

Open data has forced the relationship between scientists and the public to the forefront of conversation in recent years. Technological capacity and the increasing number of large-scale publically funded initiatives have encouraged the publication of scientific data in different formats and different fora than those that they traditionally inhabit.

Star and Griesemer discuss how controlled access to a particular collection of specimens or encoding methods functions as a critical component of establishing power and influence in a scientific context (Star and Griesemer, 1989). Using the example of a curated museum collection, they discuss how an "entrepreneur" can compile information about a particular group of items or methods, eventually becoming a recognized expert, or "obligatory channel," through whom others must go if they wish to work in the same domain. Core to this process is the creation of an encoded boundary object or "immutable mobile" that can be transferred between participants and encoded or decoded with high fidelity, creating a basis for collaboration and further development.

It is important to recognize that scientific data is a boundary object, imbued with all of the power dynamics that this involves. It takes years of work and cooperation to create an "immutable mobile," and much of a scientist's reputation and community standing is determined by the quality of his or her data. Scientists pride themselves on openly sharing information, but this usually takes place within the academic community, where the exchange is based on respect and reciprocation, which fosters trust and leads to more open exchange.

Some scientists are concerned that releasing data beyond its ecosystem of creation might lead to distortion and, occasionally, ethical issues from misinterpretation or misuse (Bezuidenhout 2013). By viewing scientific data as a boundary object, it is easier to see why scientists might be hesitant to extend the sharing of data beyond their own

For also knowledge itself is power.

Francis Bacon

communities. There are concerns about intellectual property and the responsible use of data, as well as resistance to the sometimes-hostile tone of "fact checkers," who can be perceived as outsiders who want to poke holes in research findings based on meta-analyses that may not adhere to standards of practice and evidence accepted within the community. Mistakes in the re-analysis or misunderstandings regarding the original data source can lead to public denouncements and media attention that erode the quality of scholarly discourse and force scientists to engage in a level of public debate that does not support thoughtful or systematic response.

Scientists also have cause to fear that the precedent of releasing all data on demand to anyone at any time will turn a valuable boundary object used to negotiate social value into a mere commodity that no one respects. The creation of a context-independent dataset is not a small amount of work; making this data interpretable to outsiders imposes an additional burden on the scientist, particularly if the data is to be released to a non-specialist audience that may not understand its limitations.

The open data movement has certainly benefited information designers, making interesting datasets easier to find than ever before. But anyone who has attempted to use these archives knows that there are also issues of data cleanliness, clarity of annotations and encoding, and issues of file format to overcome. The difficulty that a non-specialist encounters in interpreting "the data" reflects its nature as a product of a particular environment and a particular encoding system. Johanna Drucker emphasizes this point in her book Graphesis, where she refers to data as capta—products of a specific ecosystem that need to be respected as such, and interpreted in context (Drucker, 2014).

Despite these challenges, many scientists have benefited from the open sharing of data, and fervently support the widespread dissemination of processed data along with their research findings. The increasing availability of this kind of information has enabled rapid progress in many fields. It has also allowed designers and statisticians to engage with scientific data in unprecedented numbers, creating a broader reach than would have been realized in the scientific community alone. I firmly believe that both scientists and the public can benefit greatly from widespread dissemination of data, but non-specialists who work with data must understand that data sharing within the scientific community is an exchange built on trust. As designers, it is critically important that we understand the scientific ecosystem in order to respect the data that we receive.

Public perceptions

There are significant barriers to communication on the public side as well. Science is a highly abstracted domain, with many interconnected principles that build upon one another. Like the tip of an iceberg, a single scientific result is often only the smallest part of a much larger set of frames and assumptions that must be understood to truly grasp its meaning and implications. A robust science education would form the basis for this understanding, but not everyone receives such an education, and for many people the experience of studying science in school was a negative one. This may result in the belief that science is impossible for them to understand.

> The difference to the public between a scientist and a magician is the difference between understanding and not understanding and that is also the difference between respect and admiration on the one side, and hate and fear on the other.
>
> *Isaac Asimov*

Understandably, it can be difficult for non-experts to navigate the intricacies of a scientific argument, particularly when there are many interdependent variables in play. Interpreting such nuanced arguments requires practice and a certain level of context that can be difficult to obtain. When conclusions are limited by uncertainty or external contingencies, it can be even harder to see their implications for daily life or issues of public policy. In addition, a string of perceived failures in the past century combined with sometimes unfavorable impressions of science or scientists have made some members of the public wary of accepting scientific opinion as fact. This divide becomes deeper when people use wildly different frames to interpret the outcome of a situation.

A deeply ingrained mistrust often pervades both sides of the discussion; scientists believe that the public is unable or unwilling to understand, and the public believes that scientists are trying to avoid giving a straight answer; or worse, manipulating results in service to some private agenda. We see the reverberations of this distrust on a daily basis. Climate change skepticism, anti-vaccination campaigns, fears about GMO crops and public perception of many other issues related to scientific acceptance hinge on the cultural frames that people bring into the debate.

People are often hesitant to blindly accept scientific recommendations in these cases, especially when the topic involves uncertainty or ideologically contentious views. Politically and ideologically motivated actors make use of this distrust to further erode public confidence in science. Taken to an extreme, questions of fact become questions of loyalty to a particular authority or cause, rather than issues of reason or intellect, especially when people feel unqualified to make a decision on their own. When that loyalty is divided by competing ideals, it becomes more difficult for people to credit scientific facts when they contradict their personal conceptual frames.

The nature of scientific inquiry

Part of the difficulty in communicating science arises from unrealistic expectations of certainty founded on misunderstandings about the process of science itself. In his book Science in Action, Bruno Latour calls science a Janus with two faces: "one that knows, the other that does not know yet." The first represents the ready-made science that is usually presented to the public, and the other represents the process of science in the making, as it is practiced and experienced by the scientists who participate in the messy process of its creation (Latour, 1987). Non-experts often fail to appreciate the difficulties inherent in the struggle of knowledge creation, and can easily become confused or uncertain when exposed to this process of science in the making, rather than the science that is often presented as ready-made results.

Harry Collins and Trevor Pinch discuss the different frames used to represent science in their book *The Golem*, and propose a more humanized view of scientific progress. They argue that science is neither "a crusading knight" nor a "technological bureaucracy," but rather a golem—a "bumbling giant" that "knows neither his own strength nor the extent of his clumsiness and ignorance."

> *"The problem with experiments is that they tell you nothing unless they are competently done, but in controversial science no-one can agree on a criterion of competence. Thus, in controversies, it is invariably the case that scientists disagree not only about results, but about the quality of each other's work. This is what stops results from being decisive..."*

This difficulty is characteristic of Rittel's wicked problems; they are inevitably tied to consensus-building, and take significant time and energy to resolve. Even when a consensus has been reached, those who have not fully participated in the process may be hesitant to accept its results.

Reasons for hope

Despite the many challenges listed above, there are also several reasons to hope that communication between scientists and the public is not as hopeless as it may seem. First, there is a growing public awareness of scientific issues and their relation to policy decisions. Though engagement is not uniform, there are groups of non-specialists who are highly interested in science and scientific issues.

In 2007, Niepold, Herring, and McConville discussed the importance of conceptual framing in public opinions on climate change. They suggest an approach that addresses these different frames by tailoring communication to several target audiences: the uninterested or unaware, the climate science interested, the climate science attentive, and the climate science engaged. Their early experiences suggested that communication focused on changing conceptual framings—often through personal stories—were effective at moving people through these different levels, and toward more highly engaged attitudes.

Corbett and Durfee recently measured how media representations of climate change affect people's evaluations of the certainty in controversial topics (Corbett and Durfee, 2004). They found that presenting results in context created the greatest degree of confidence in science when compared to a control sample that simply stated results, or one that focused only on controversy, regardless of a person's incoming biases. Citing the results of a 2004 National Science Foundation survey, Nisbet also argues that these problems may not be as widespread as we often think; it may be that our current pessimism is driven in part by our own framing of the situation, grounded in the deeply-seated deficit model, which can harm both sides. Nisbet states:

> *"Relative to authority, deference, and respect, scientists have earned a rich bounty of perceptual capital. When controversies occur, the challenge is to understand how to use this capital to sponsor dialogue, invite differing perspectives, facilitate public participation, reach consensus when appropriate, learn from disagreement, and avoid common communications mistakes that undermine these goals."*

These and other studies suggest that many people do trust scientific information, and that improving public acceptance of science and its application to policy might simply be a matter of reframing the issues into a different context. It may be that people find the current methods of communication distasteful, but actually support the science itself, when communicated in a way that they find culturally acceptable.

Moving beyond the deficit model

Finding ways to presenting scientific data in a way that is accurate, approachable and culturally acceptable to the audience is an important part of beginning to communicate science. Consensus-building is a critical part of behavior change; in the absence of the consensus-forming process, people tend to revert to familiar heuristics and ignore scientific evidence. In order to turn understanding into belief and belief into action, science needs to do a better job of framing problems in a way that resonates with the values that people hold.

There is also room to consider the possibility that the public might contribute to the process of science in constructive and helpful ways. It is no longer the domain of scientists simply to understand and to tell; instead, we need to justify scientific rationale to the public and allow them to participate in the process of reaching conclusions.

A growing movement within the scientific community already seeks to engage the public by allowing "citizen scientists" to contribute data to scientific projects and to serve as activists to support science in favor of a specific cause. These initiatives allow the public to take an active role in science, rather than passively funding studies through government agencies and consuming the results. Another area ripe for development is to allow the public to participate in the broader framing of scientific research, and its place in public policy through robust and ongoing dialogue (Miller, 2001). Science communication is a critical part of preparing the public to take on these new roles. Redefining public contributions to science may also change the role and acceptance of scientific data in a public context, and designers are well positioned to facilitate this negotiation.

Now is the moment

2017 is a particularly good time to be discussing issues of science communication. Open data initiatives make an unprecedented quantity of scientific data available to the public for the first time. The prevailing popularity of data visualizations points to a widespread public curiosity and fascination with data when it is presented in new and interesting ways, juxtaposed against a deep suspicion of knowledge handed down from experts. The current political climate confronts established scientific norms and creates new incentives for scientists to focus on dissemination of their work. The political zeitgeist threatens the value of scientific expertise and reputation as a boundary object, and may help to convince scientists of the urgent need to take a more active role in representing themselves. All of these conditions present opportunities to change default assumptions and attitudes. It is time to find new ways to encode and represent scientific data, and to invite the public to engage with these issues.

Like a work of art, we exceed our materials. Science needs art to frame the mystery, but art needs science so that not everything is a mystery. Neither truth alone is our solution, for our reality exists in plural.

Jonah Lehrer

CHAPTER FOUR

Design for Science Communication

The design community has extensive expertise in high-fidelity transmission of conceptual information to a general audience, and represents a valuable resource for scientists looking to improve the perceived value and relevance of their work. Designers can help to identify points of contention, expose underlying conceptual frames, present differing perspectives clearly, and facilitate two-way conversations about scientific research and public policy.

Several features make designers and scientists natural partners in communicating with the public. Complementary knowledge domains and structural similarities between the disciplines create multiple opportunities for understanding and collaboration between designers and scientists seeking to communicate with a broader audience.

Design is not journalism

Science journalism already does an excellent job of translating scientific results to the public, and there is much that both designers and scientists can learn from studying this practice. Ultimately, effective science communication depends on collaboration among experts in all three areas. However, that is not the primary focus of this thesis. There are cases where the contributions of designers and journalists overlap, and there are many situations in which a single person might perform both roles, but the primary functions of these roles are different.

While journalists focus on researching and communicating information through narrative, designers focus on structuring information and conversations, in order to facilitate clear communication. Design can play an especially important role in cases where traditional narrative struggles: when information is highly nonlinear, or when large amounts of information need to be shown and compared side by side. In these situations, designers can help to create the infrastructures needed for journalists to convey a complicated story clearly. In this way, design can support rather than replace traditional journalism.

Structural similarities

Although science and design often differ markedly in their area of application, they share a similar focus on process and the structured application of internalized methods to the problem at hand. Both disciplines rely on an extended apprenticeship model, helping novices to develop experience and intuition through an extended period of training with a mentor or guide. In both disciplines, the product is usually the only thing that people see—the process is largely hidden from view. In both disciplines, clarity and truthfulness are paramount, and fidelity of information transfer is a centrally-held value. Both designers and scientists think actively about the appropriateness and use of models, the limits of metaphor and representation in communicating the essence of an idea, and the importance of finding an appropriate balance between abstraction and detail. Recognizing these common values and shared experiences can help to breed empathy and a sense of respect between scientists and designers, and leads the way toward productive collaborations between the two disciplines.

Contributions of designers

Designers have extensive experience in communicating with the public, and are practiced in facilitating conversations that clarify conceptual models and cultural frames. They also study different methods and approaches for converting these conceptual models to external forms, an important step in understanding and resolving differences in approach. Because they have often worked at the forefront of controversial issues, designers have developed generalized theories to encode the important principles supporting communication. Questions related to public policy and management in the past can easily be extended to inform discussion of science today, creating a rich body of literature and historical precedent that can be adapted to meet current demands. As outsiders to scientific practice, designers can also serve as a useful sounding board to help scientists refine the tone of their approach for an external audience.

Jean Trumbo discusses the extensive visual literacy requirements of science in a 1999 paper in the journal Science Communication (Trumbo, 1999). Science depends heavily on the use of models to represent things that we cannot see directly, and these models in turn influence how people think about a system. Mary Hegarty further explores how external representations affect internalized models, arguing that visual form plays an active role in shaping thought (Hegarty, 2002). Focusing on visual thinking, visual learning, and visual communication, Trumbo details how visual literacy relates to scientific understanding, but points out that there is currently no common theory to organize efforts at scientific visualization. As experts in visual communication, designers can help to build this unifying theory, based on principles of practice established over many decades of collective experience.

For designers, scientific data provides a new and interesting challenge. Creating new forms for abstract ideas and developing methods that scale to accommodate large quantities of information are interesting problems that push the boundaries of traditional design and present new opportunities for innovation (Ichikawa 2014). The overwhelming popularity of data visualizations in recent years speaks to the public interest in complicated and beautifully represented data. In addition, the applied aspects of the work bring a deep purpose and relevance to the product of such design. Extending this approach beyond presentation of a single dataset to include visualizations of the connections and interrelationships between disparate data presents further opportunities for design contributions. Approaching such a challenge requires attention to several aspects of a design, as will be discussed in the next chapter.

Much of the work on scientific data visualization has focused on accuracy and efficiency, sometimes to the detriment of aesthetics. To make scientific data interesting to the public, it must be clear, presented at an appropriate level, and attractive. In this situation, aesthetic appeal plays a significant functional role. In addition to encouraging the viewer to engage with the data, artistic visualizations can inspire a sense of awe and allow non-scientists to experience some of the wonder that drives most scientific work (Scott, 2003). Designers can help to reintroduce aesthetic values into scientific results, to improve their impact and enhance their appeal. Such approaches must also carefully balance aesthetic and functional considerations, to ensure that the value of the data itself is not lost.

Simplicity is complexity resolved.
Constantin Brancusi

Much like science, design is a practice based on process. Designers begin with a question, create a hypothesis about how best to answer it, run an experiment (in the form of sketches) to see whether form matches theory, test those experiments by showing them to people and gathering responses, analyze the results, and propose new solutions informed by their experience. Design is a difficult practice, because the stopping point can be poorly defined; an emotional response or a recognizable similarity is harder to identify or quantitatively assess than most results in science.

And, unlike science, in design there are no universal answers. There are many principles that guide good design, but no neat formulas that can be applied to all instances of a problem. The test for generality fails in a discipline where every problem is situated in a particular human context that changes from place to place. Unlike scientists, whose ideal results can be clean and clear, designers are accustomed to navigating the murkier waters of nuance and ambiguity.

I do not propose that designers possess ready-made answers; like science, genius in design comes from the ability to ask the right questions. These questions can only emerge through a robust and self-aware process that characterizes and responds to a particular variable space. It is only through a deep and engaged process that a designer can arrive at an integrated solution. Recognizing this, scientists can begin to see designers not as a mouthpiece or public relations representative, but as co-discoverers of the principles of communication needed for their particular issue.

Contributions of scientists

Before they can improve the public communication of science, the communicators themselves must first understand. Scientists can help designers (and others seeking to share scientific data with a public audience) to understand the context of a particular problem by explaining the implications and limits of the data in clear and accessible terms. Rather than relying on academic publications to convey their work, scientists can interact more directly with the public through their own writing, and by making consultation with designers and science journalists a priority. Teaching designers how to respect and work with data will help to ensure accuracy and avoid misinterpretation of results, benefiting both sides.

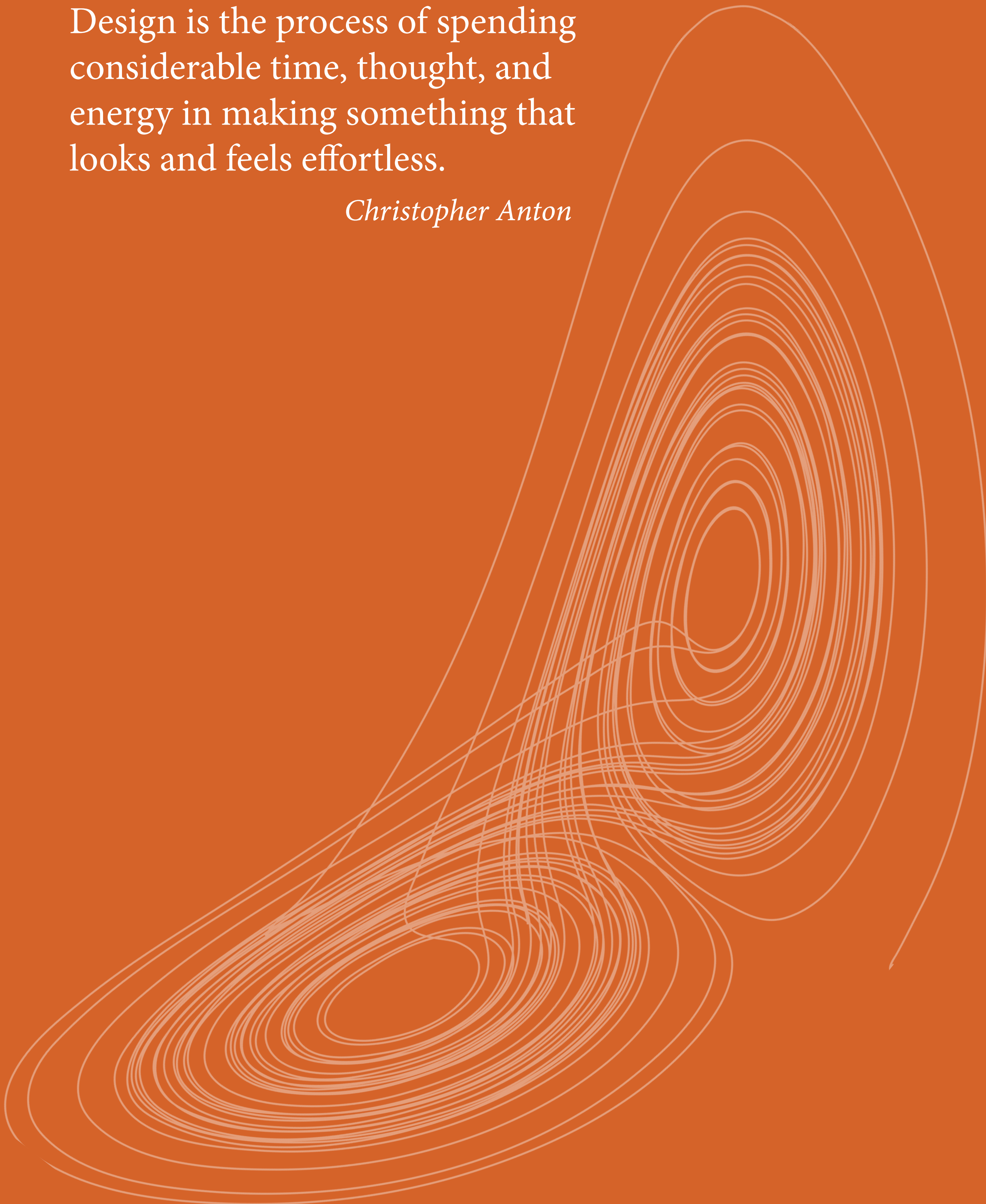
Scientists can also bring a sense of wonder or curiosity, and an understanding of the importance of scientific inquiry to the conversation. Well-trained in formal logic and reasoning, they can help to identify fallacies that result from incomplete understanding, evaluate and appropriately balance competing perspectives, and provide a critical assessment of the information fidelity achieved. Their experience in navigating differences between correlation and causation provides an essential framework for discussing the truthfulness of a particular graphic, helping to clarify the design message and prevent misattribution of cause. As experts in working with data, scientists can also help to catch inconsistencies between a scientific model or dataset and its representation in visual form.

Substantially engaging with designers and the public can help scientists as well. Conversations with outsiders can help them to understand people's concerns and opinions about scientific issues, and may create new opportunities and future avenues for exploration. Continued trust and goodwill between scientists and the public will ensure that high priority is placed on funding science, necessary to support large-scale initiatives, as well as the day-to-day business of science. Improving public understanding of the scientific process and its results can also increase the social standing of science and capture the public imagination, which ensures that this often difficult pursuit will continue to attract the best minds.

Finally, a deep and continuing engagement with the public allows scientists to shape their own reputations in the broader community. Direct, personal contact with non-scientists is a powerful opportunity for scientists to shape their own image, rather than relying on others to choose how they are portrayed in the world.

Design is the process of spending considerable time, thought, and energy in making something that looks and feels effortless.

Christopher Anton



CHAPTER FIVE

Visualizations of Interdependence

Graphics perform many different functions, depending on their intended audience. They can provide a sense of wonder, explain a complicated idea, or provide users with an invitation to explore. Discussions of context, sense of scale, data provenance, and uncertainty are especially important for visualizations of interdependent systems. User agency and the choice of narrative are also important considerations. In this section, I review current examples of information visualizations and create categories to help distinguish different approaches.

An information visualization can perform many different functions, depending on its intended purpose. A graphic that integrates many different approaches often has a broader reach than one that focuses on a single channel of communication. In this section, I will examine several recent visualizations and identify different approaches to creating visualizations that convey a sense of wonder, offer the user an opportunity to understand, or invite the user to explore a particular topic.

Visualizations of complex systems should emphasize connectedness and highlight relationships between interdependent parts.
Ideally, graphics for science communication should be simple and straightforward, but they should also convey the process and results of science for a sophisticated audience. In addition to presenting facts and data in an interesting way, these visualizations should also include information about knowledge generation in science, justifications for how scientists arrive at conclusions, and a discussion of how one should respond to uncertainty in scientific results. The weight given to each of these attributes varies greatly depending on the purpose of the graphic, and its intended audience.

To explicitly represent the purpose of different designs, I will group science visualizations for a general audience into three broad categories, based on whether their primary function is to explain a concept, to allow users to explore the data, or to excite users about the prospects and the process of science. Of course, these categories are not mutually exclusive, and any particular design will likely embody each of these aspects, to a different degree. Design is always a series of compromises between competing objectives; the uniqueness and value of a particular product emerges from the choices that the designer makes to accommodate conflicting needs. This compromise is not a flaw, but rather a desirable outcome of a robust process where the designer seeks to find the best fit solution for a particular set of needs.

To reflect this approach, I will examine aspects of different designs, each constructed as a continuum between opposite approaches. Figure 4 shows a summary graphic that will be used to rate each of the visualizations in the sections that follow, along several different dimensions. The primary attributes pertinent to science communication can be grouped into four broad categories: the tone of the visualization includes both its level of formality, and the balance between its appeal to reason and emotion. The structure of the visualization can be either linear or nonlinear, and can support either an expository argument or a narrative. In some cases, the visualization may not present any kind of argument at all, but instead serve as a collection of related facts that users can explore to draw their own conclusions.

The next category attempts to classify the agency given to the user, and explores the balance of control between user and designer. In many graphics, the user experience is tightly structured and carefully controlled. Others leave more room for the user to make choices and interact with the visualization in some way. Similarly, it is possible for the designer to leave room for the user to explore and reach his or her own conclusions, or the design could be focused on explaining a particular way of approaching the data. Finally, the graphic can present conclusions as facts, or it can discuss the reasoning behind those conclusions and invite the user to engage in their own reasoned argument in response.

In terms of content, a graphic can present data as entirely certain, or it can discuss alternate interpretations and possible weaknesses in the data. This discussion of uncertainty is still relatively uncommon in data visualizations, but is an important part of scientific discourse, even in cases where the greater conclusions are generally accepted and not in dispute. A critical part of this discussion is to acknowledge the sources

Figure 4. Template for analyzing important characteristics of the visualizations shown in the Related Works section.

for all data used, and to explain where each comes from and how it informs the final conclusions. Finally, it is also important to acknowledge alternate interpretations of the data shown, and to expose controversies or disagreements about the implications of the data. When viewed in aggregate, it is my hope that these distinctions will help to produce a more nuanced view of how a particular design is positioned within the design space.

In this chapter, I will examine several examples of recent visualizations in an attempt to frame the extreme ends of the spectra. I don't view these rankings as a "score" that shows how a visualization succeeds or fails; rather, they are an attempt to locate a particular example within the broader visualization community. By reflecting on this collection of works, I will identify general guidelines within each of these categories that can be used to inform and assess the thesis project.

SECTION ONE

Graphics that Explain

Visualizations can be used to help people understand a complicated idea or situation. Diagrams and illustrations help to highlight specific features of interest in a particular system, and can be used to guide the viewer's attention, while maintaining an approachable feel. Interactive graphics can be used to allow a user to play with the system, in order to see how it responds. Such visualizations can help people to understand the limits and meaning of models, to explore alternative scenarios, and to experience cause-and-effect relationships.

Graphics that explain focus primarily on presenting complicated ideas clearly, and creating opportunities for the user to better understand the topic. They are often used to support communication of complicated logic chains, or to build arguments that suggest causation. To improve clarity, these graphics often—but not always—limit alternatives, constraining the user's experience to a particular narrative or point of view.

The video "Wealth Inequality in America" is the only upload to an anonymous YouTube channel; it has no clear authorship, but has received 19 million views. The video begins with a highly personal tone: "There's a chart that I saw recently that I can't get out of my head," and then uses stacked bar charts to show the difference between how Americans perceive wealth distribution in the US, and how it compares to reality. The animation walks the viewer through each bar in the chart, explaining the significance of each and pointing out salient features as new data is added. Then, it shifts focus and shows the collective wealth of the country, and plots the distribution for individual percentage points, again comparing the same three distributions. Finally, the author rearranges the top ten percent of the data so that all of the wealth fits on the chart. To do this, he needs to isolate the person representing the top one percent in his own bar, which is wider than the top 10 percent combined.

This visualization is very effective because it presents a very clear explanation using simple graphics. Transitions are used effectively to complement the narrative and explain exactly what the viewer should be noticing in each stage of the animation. Using data to illustrate the narrative and citing the source of that data as a Harvard professor grants the visualization an authority that it might not otherwise have. The personal tone of the introduction and clear exposition make the data more approachable, and the piecewise revelation makes the data feel more dramatic. Overall, it is a highly persuasive piece.

The visualization appears to be appealing to the viewer's reason; the underlying argument is structured as an exposition that presents the data as it is. However, the haunted tone of the first sentence, barely-audible background music, and the way in which the author delivers the script contribute to an emotional message of loss, disappointment, or devastation. The argument is presented in a linear format, with no interaction, as expected for a video medium. There are no options other than the choice to agree or disagree. The primary argument is expository; one fact builds on another to reach an apparently unavoidable conclusion. The lack of alternate interpretations or qualifications that limit the statement suggest that alternate views may have been suppressed for the purpose of establishing a powerful narrative.

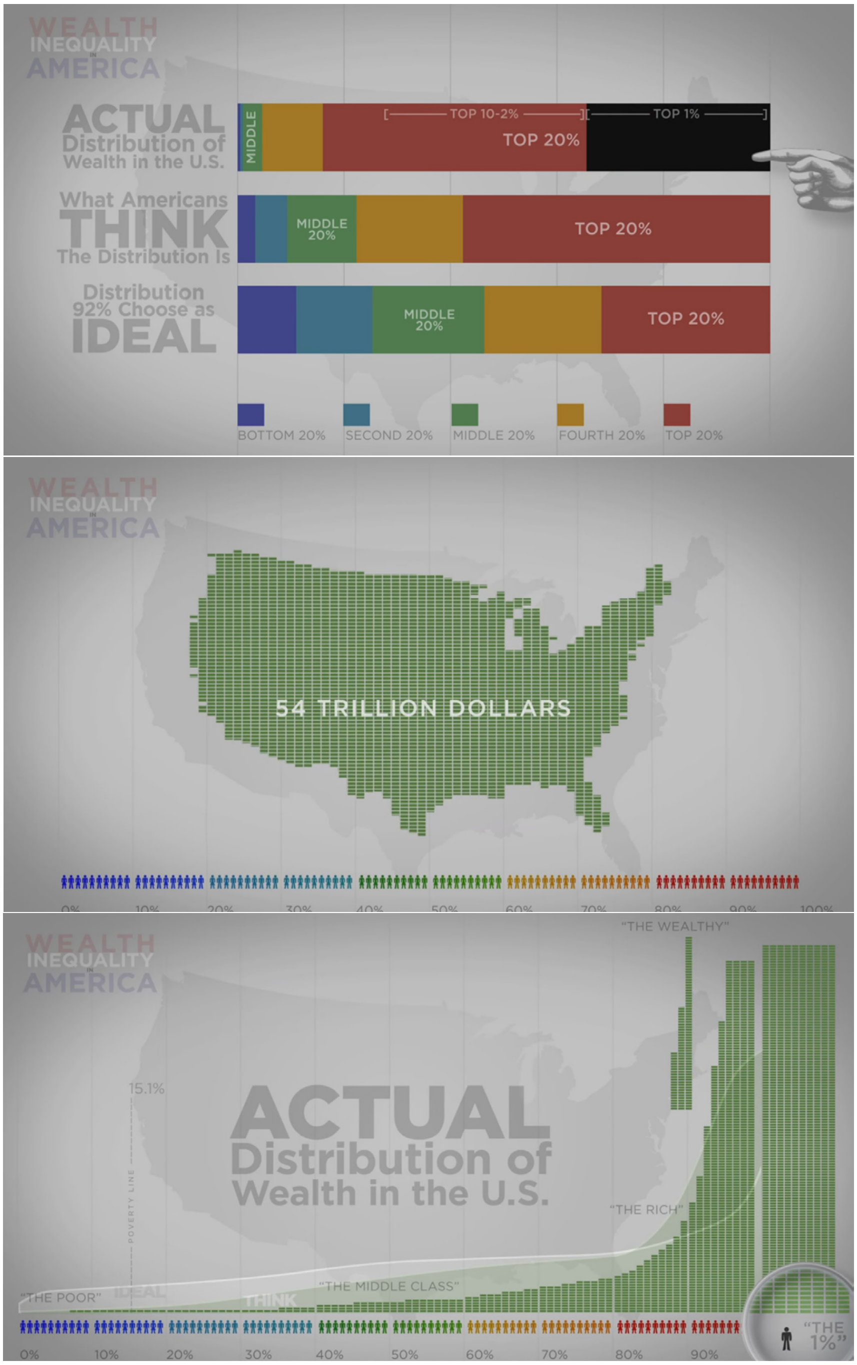
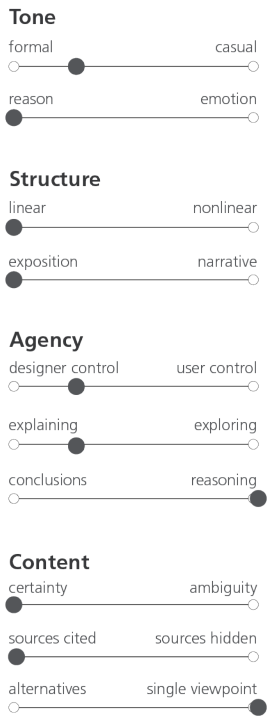


Figure 5. Different views from "Wealth Inequality in America," a visualization-based narrative explaining the statistical distribution of wealth in the United States.

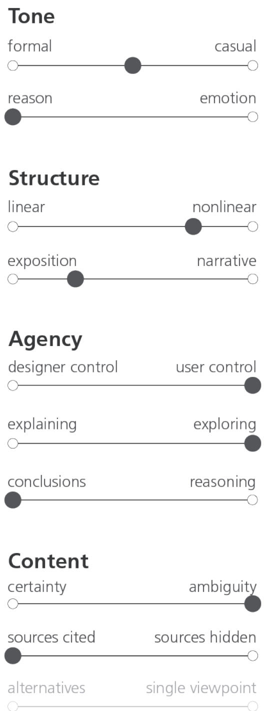
Figure 6. "Our World in Data" presents global statistics in a journalistic style, supported with information graphics to illustrate trends.



The website "Our World in Data" by Max Roser curates a broad collection of statistical data. For each section in the collection, the site lists a series of formal journalism articles that present and contextualize the data. The tone of the writing is expository, and each narrative presented is linear, although the user can choose to read in a nonlinear order. The primary purpose of the visualizations is to present the statistical data in a neutral way, while the text explains what the graphics mean.

Most of the graphics on the website include at least some interactivity, allowing the user to highlight specific datasets or read out values from a particular point on the graph. In some cases, the user is also able to change their view of the dataset to see an alternate representation. The influence of these interactions on the data shown is fairly minimal, so I would classify this set of visualizations as primarily designer-controlled. The visualizations are meant to explain the state of the world, while allowing some room for exploration within a particular topic selection. The sheer scope of the data presented on this website is impressive; it truly seeks to be a catalog of data on every topic, and explores the available data to some depth. In some cases, it also explains the relationships between datasets and links between narratives, though the flow is largely linear and sequential. Likely because of the comprehensive nature of the project, the narrative coverage of the data is uneven, varying from extremely detailed explanations to single-sentence image captions for different topics on the site.

When present, the narrative usually discusses the reasoning behind conclusions and summarizes critical features of the data. Generally, the discourse around a particular graphic assumes a high degree of certainty, and does not explore alternate interpretations. Source information is given for each graphic, and many include links to the full descriptions of each dataset from the original organization that published them, allowing users to analyze the substance and limitations of the data themselves.



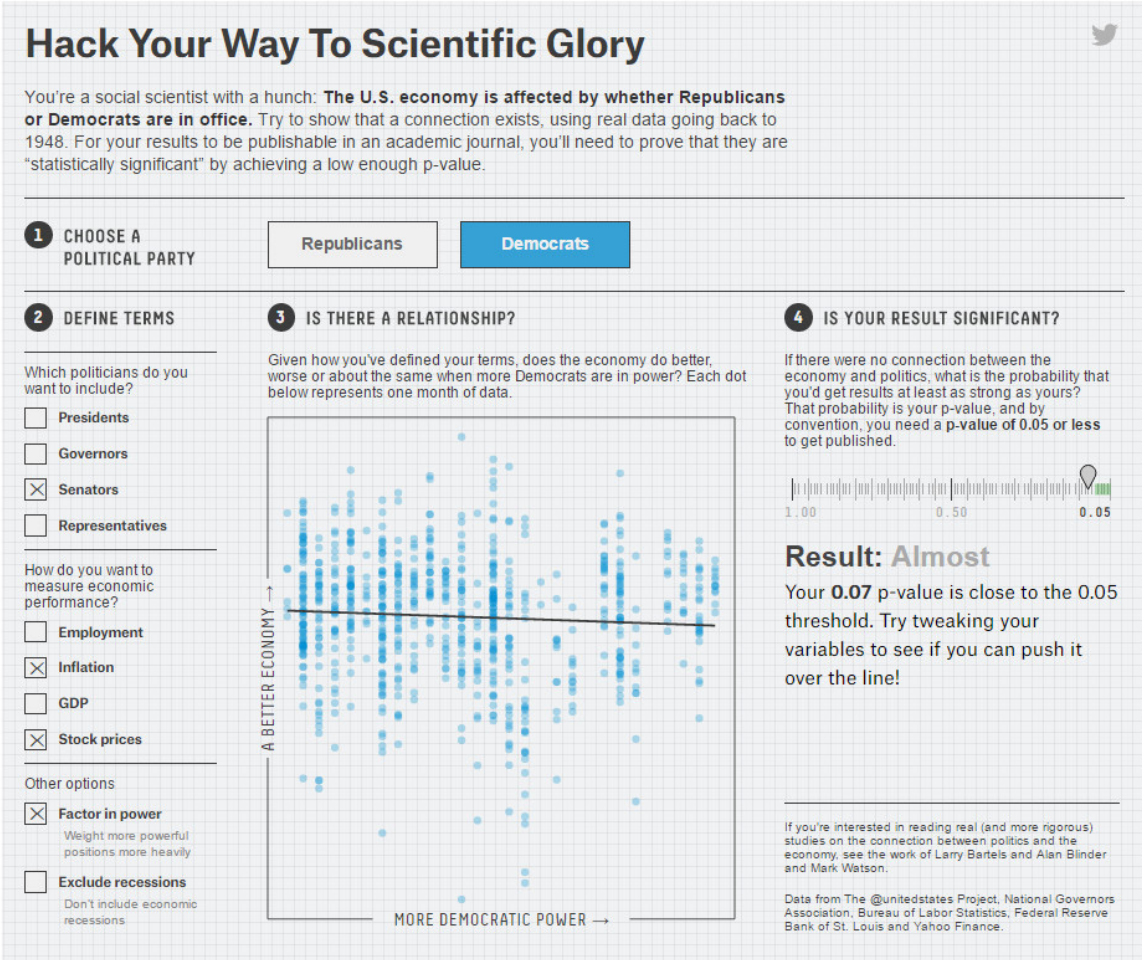
This visualization by Five Thirty Eight represents an attempt to educate the public about the meaning of statistical methods used in science, and the reasons that p-values are not entirely reliable as measures of certainty. The visualization is part of a larger article, entitled "Science Isn't Broken," which explores the reasons behind the so-called "reproducibility crisis" in science.

The visualization gives users the option to toggle different values in a political dataset, and shows the result of those changes on the analysis results. For each combination, the visualization provides a p-value score, as well as an interpretation of what that score might mean. By interacting with the graphic, users can get a better understanding of how seemingly unimportant or justifiable decisions can make a result seem more important than it really is.

This kind of visualization is important, both because it educates the public about some of the process behind science, and because it encourages healthy skepticism and critical thinking about the nature of statistical evidence. A secondary visualization on the site emphasizes the variability of results across different research methodologies; showing the different conclusions obtained from the same data by 29 research teams using different methods.

The casual tone and simple interface make the visualization easy to approach and interact with, and the experimental feel makes it engaging as well, with a high degree of user agency (given a limited number of controls). The visualization is designed to appeal to reason rather than emotion, and uses expository language throughout. The visualization itself imposes little narrative beyond explaining what the variables mean, but the broader article provides context for understanding the meaning of the graphic as a whole, and is positioned as an argument for the validity of science.

This graphic plays an important role by helping users to engage with ambiguity in scientific results. It is also somewhat risky—a simplistic take-away would be to assume that all numbers are faked and that scientific results are meaningless. The clear discussion surrounding the graphic works to counteract this view, and the piece as a whole manages to acknowledge uncertainty and the difficulty of interpreting data without undermining its authority.



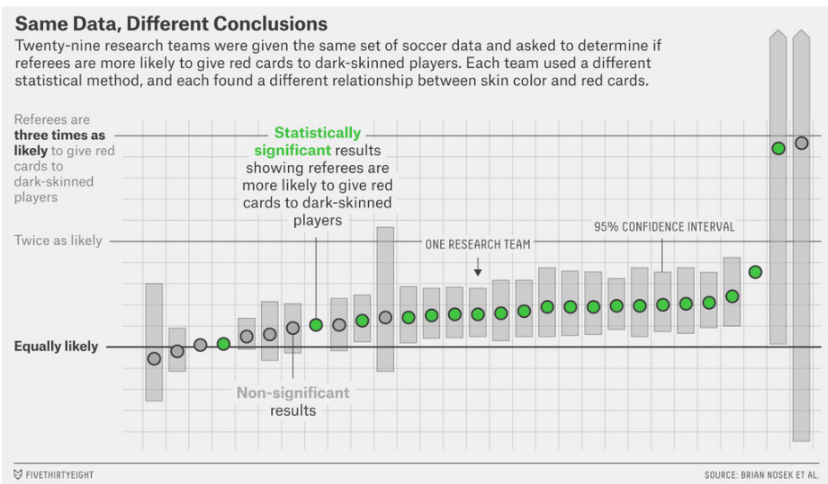


Figure 7. "Hack your way to scientific glory" explains the difficulty of interpteting statistical values by allowing users to adjust different components of the analysis and model the results.

SECTION TWO

Graphics for exploration

Visualizations can offer users the opportunity to explore a topic, either with a guided or a user-driven approach. These graphics tend to emphasize user controlled filtering, alternate or linked views of the data, and the opportunity to control the level of aggregation of the data. Such data-first approaches rely on the user’s interest, insights, and engagement to build understanding from the data, rather than constructing a single narrative chain. Here, user autonomy is very high, but the flexible information structure can make it difficult for people to know where to start. This challenge is often overcome by use of popups, sidebars, and other cues that are generated in response to actions, and help to orient the user and suggest possible directions to investigate.

In his "Better Life Index," Moritz Stefaner used small multiple glyphs to present an overview of multiple OECD life quality indices in a compact and attractive way. Each floral glyph is a radial bar chart, where petal length represents the size of that indicator for that country. When a user chooses a new category from the selection menu, the y-axis values update and the flowers rearrange themselves to reflect the ranking for each country. This engaging front page is backed up by a more traditional map view that shows how people from different countries rank each of the topics in the OECD list, by gender and age. Users are also able to compare values for two countries side by side, and to contribute their own rankings to the dataset. Each country also has a written description that summarizes the most important takeaways from the study, and shows its results in relation to other countries as a series of bar charts for each indicator.

Although the OECD website is a relatively formal venue, this graphic was clearly designed to be somewhat more casual and approachable. It still retains a data-heavy focus, and is driven more by reason than by emotion. Unlike most of the previous examples, this visualization is designed to enable exploration and full user control, and is almost entirely non-linear. The user can only interact with the visualization in prescribed ways, but the filters can be applied in any combination and any order that the user wishes to see. Although interdependence is not explicitly displayed, the use of shape, position, and small multiple encodings allows the user to pick out patterns in related variables, thus enabling them to search for similarities and strong relationships within the data.

The heavy emphasis on exploration means that narrative and exposition take a lesser role, at least in the primary visualization. The supplementary articles are expository in tone, and explain what should be understood from looking at the data. The discussion in these articles does show the reasoning behind at least some of the conclusions that the user can take away. An auxiliary executive summary minimizes the explanation of reasoning, and focuses on summarizing the conclusions in compact form.

Figure 8. The "Better Life Index" project by Moritz Stefaner represents economic indicators as floral glyphs to present statistical data from the OECD in an engaging and informative way.

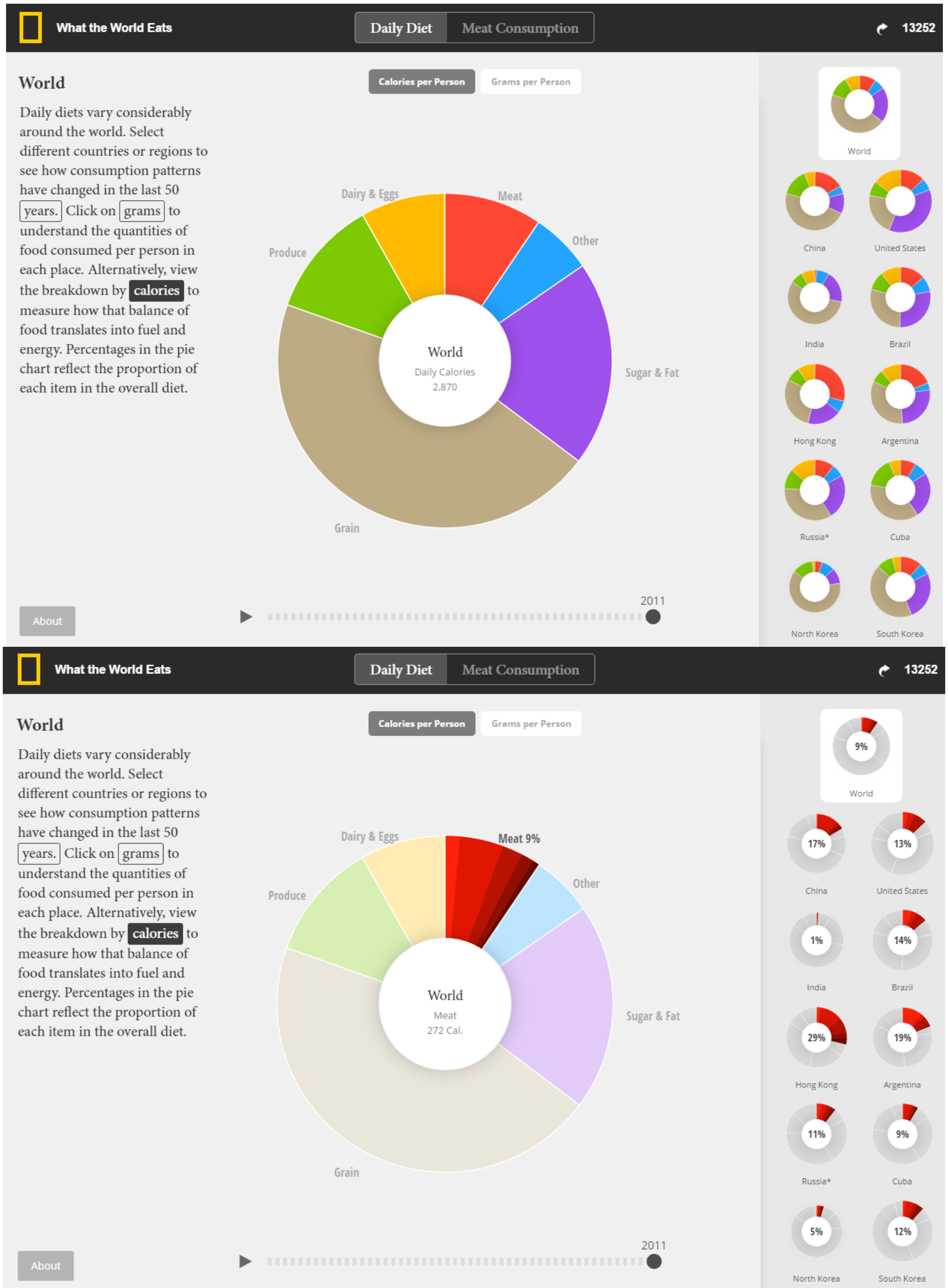
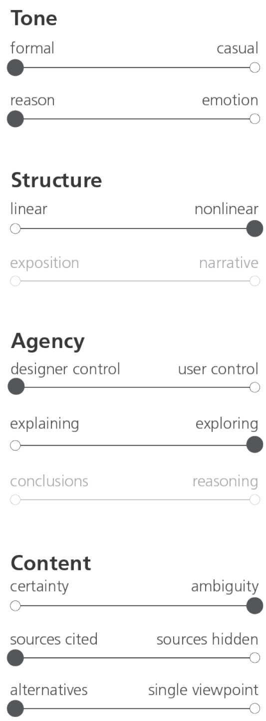


Figure 9. "What the World Eats" by Fathom Information Design is an interactive dashboard for exploring diets around the world.

The “What the World Eats” visualization designed by Fathom Information Design for National Geographic is an example of a dashboard interface that allows users to explore statistical information from the Food and Agricultural Organization of the UN. Using an interactive small multiple display, the visualization allows users to compare the relative proportion of food types consumed by people around the world. When a user clicks on a section of the bar chart, that category is subdivided into individual components and highlighted in each of the small multiples. Numerical labels are also added to facilitate comparison. The data can be viewed for a particular year, or historical data can be played back in sequence to observe changes in diet over time.

This visualization is somewhat more formal than the “OECD Better Life Index” in terms of its visual encoding of the data, but otherwise it is similar in tone and intent. Unlike the “Better Life Index,” this visualization is largely designed to be a standalone piece, unsupported by journalism pieces. The brief text description in the upper left provides some context for understanding the visualization, and updates for individual countries to highlight interesting features of the dataset.

As a predominantly exploratory piece, the emphasis is on allowing the user to control and interact with the data to discover trends for themselves rather than creating a narrative to guide the user on a particular path through the data. The “About” tab gives details of the data sources used, but does not explore alternate data or interpretations directly. This is due at least in part to the fact that there are so few interpretations presented to begin with.



Project Ukko is a data visualization dashboard developed for Euporias, a collaborative effort funded by the European Commission to create tools to visualize climate services. Of the examples that I am considering here, this example probably comes closest to a formal visualization tool, as it was designed in close collaboration with experts and intended primarily for use by an expert audience. As such, the graphic itself focuses on showing the data directly, and has less emphasis on providing interpretation or explanations.

The graphic itself is fairly simple in scope, though the techniques used to create it demonstrate sophisticated design. A mixture of color, orientation, and shape encoding is used to layer multiple pieces of information on a single graph. This central map view is supported by auxiliary graphics, including a clear legend indicating the function of each symbol and an expanding view that shows the results of different model predictions for a given location when the user clicks on the map.

The project has a formal tone in keeping with its intended purpose, and arguments in the written supplements appeal primarily to reason, using data as supporting evidence. Its structure is nonlinear; the user can choose to explore the available data in any order. However, there are no filtering or viewing options that the user can adjust. Instead, the visualization functions as a tightly controlled browser for showing the scientific results in an efficient and aesthetically pleasing form.

The supporting information includes a discussion of the significance of the different models shown, and outlines the procedures used to create the data in the visualization, including post-processing and estimates of the prediction skill of each model shown. This thorough representation and discussion of uncertainty is uncommon in the other visualizations we have seen, and likely reflects the more formal purpose of this design.

In addition to providing an overview of the data, the website also includes links to specific data sources, credits the designer, and includes information about sponsoring organizations and project goals. The graphic is presented as a tool from which the user can draw his or her own conclusions, and does not advance an argument. The few statements made in the discussion are used to justify design decisions or the scientific reasoning behind calculating the data in a particular way.

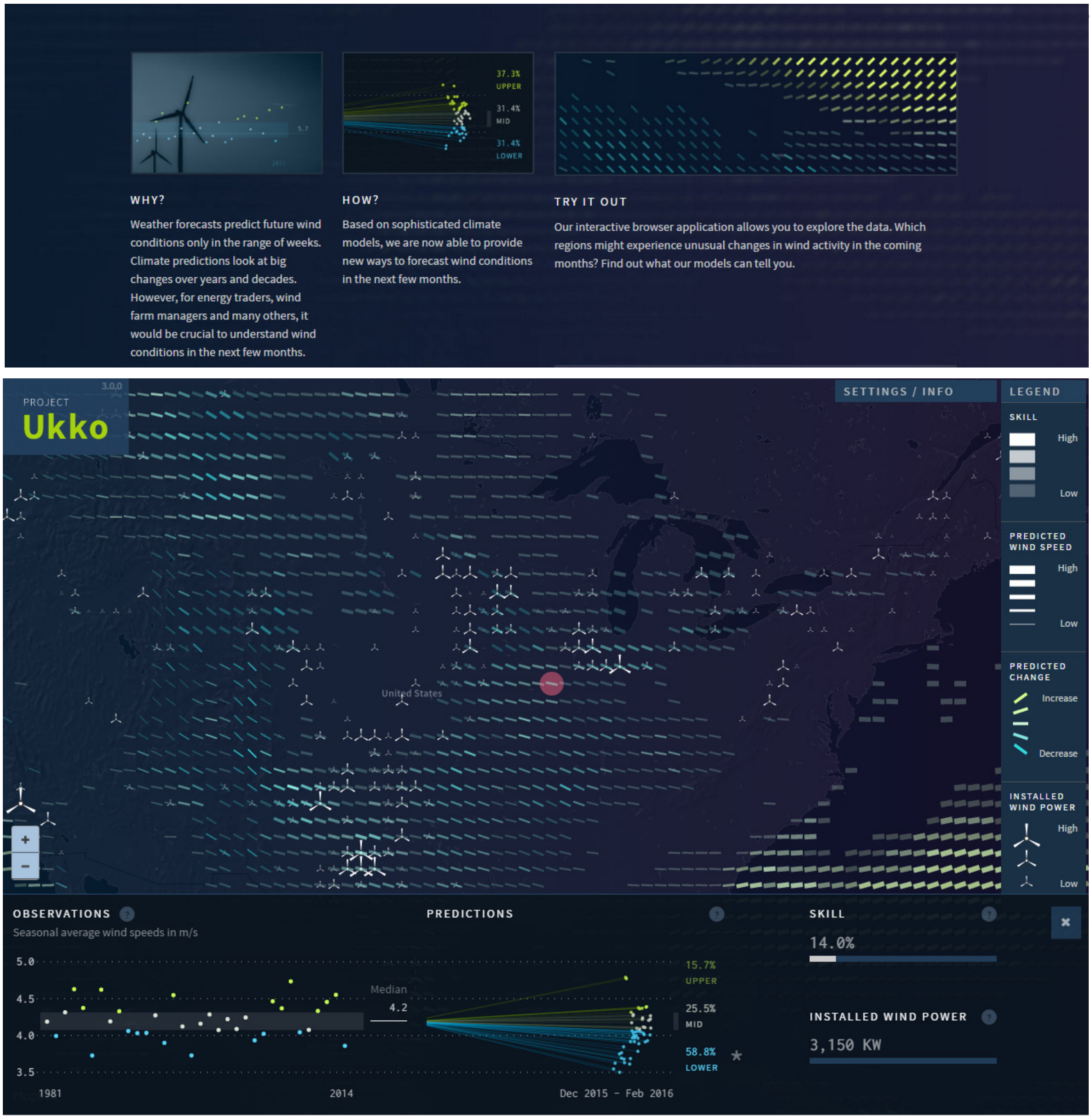


Figure 10. "Project Ukko" by Moritz Stefaner is a data visualization dashboard for browsing a multidimensional scientific dataset.

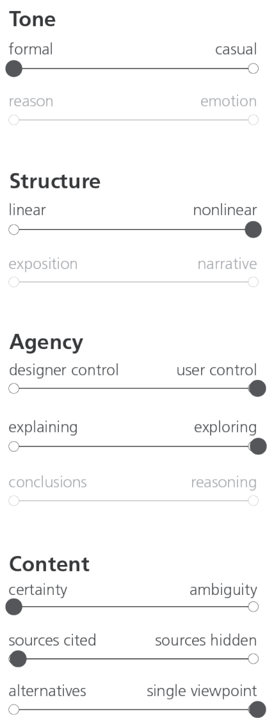
SECTION THREE

Graphics that evoke emotion

Information visualizations can inspire a sense of wonder or awe that excites people and encourages them to engage with a topic. These graphics can be used to excite a user about a topic, to encourage engagement with data that might otherwise be considered boring or unpleasant, or simply as a way to create a particular experience for the user. These graphics rely heavily on aesthetics and narrative to create an immersive environment that supports the desired response.

By communicating on both a logical and an emotional level, visualizations can facilitate perspective shifts that might not be possible with reason alone. Even if the viewer does not change his or her mind in response to a visualization, an aesthetically and emotionally powerful piece is often more memorable than simple charts and graphs.

Although these visualizations are sometimes considered “artistic” rather than data-driven, there are many cases where they serve an entirely quantitative purpose as well; there is no absolute divide between task-driven visualizations and aesthetic appeal. Many visualizations in this category are hybrids that exist in both worlds.



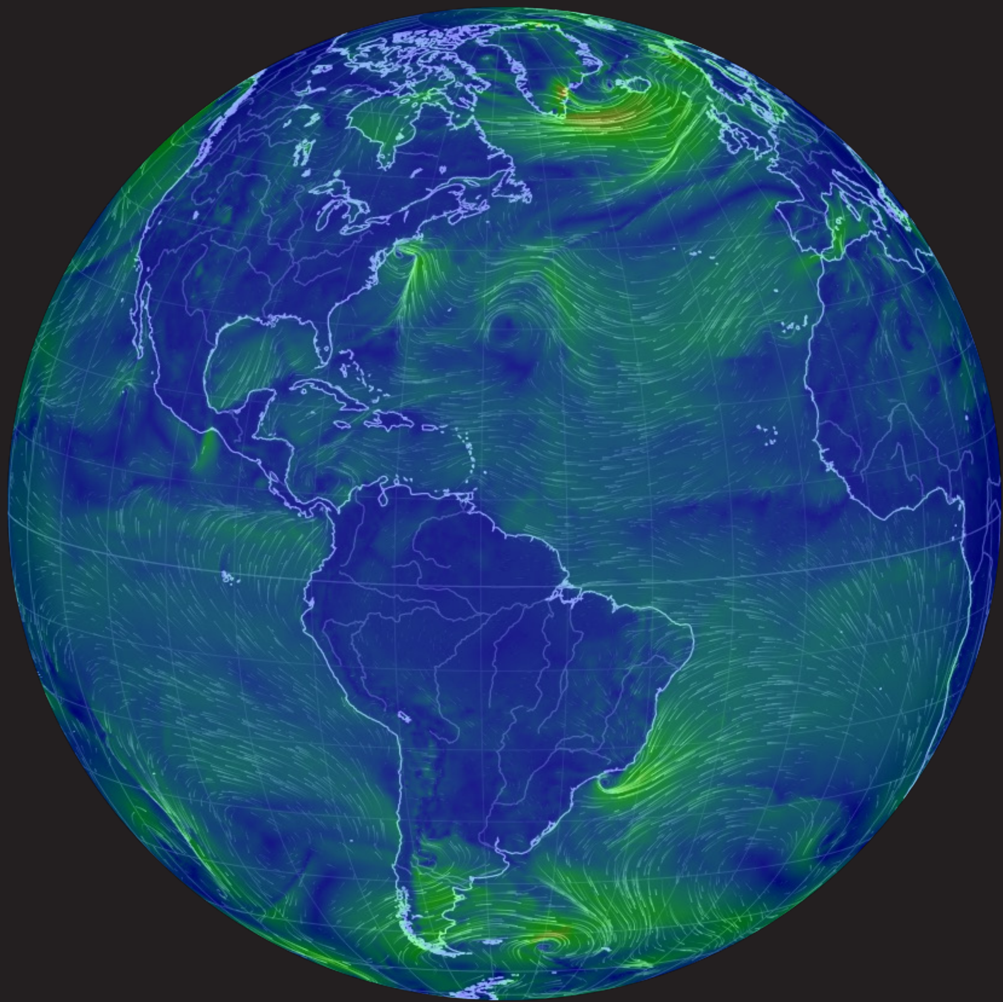
The "Earth" visualization by Cameron Beccario uses a 3D globe projection to visualize several different datasets related to global weather patterns, similar to the Wind Map by Fernanda Viegas and Martin Wattenberg. Users can choose to display wind and current patterns, chemicals and particulates in the atmosphere, air temperature, humidity, and a variety of other variables on the same map. Some variables are shown as a color gradient overlay on top of the moving particles that emphasize flow within the system. Users can also choose whether to view past or current data, and can show predictions for future patterns based on supercomputer models that are updated every few hours.

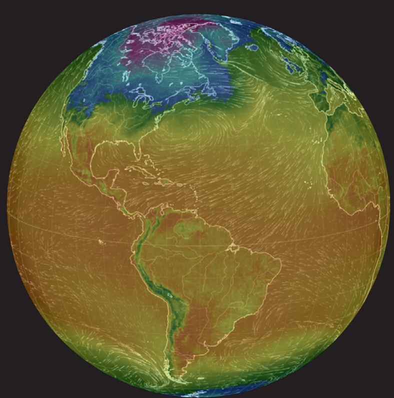
Like "Project Ukko," this visualization falls on the formal end of the scale, since its minimalist representation presents data with no interpretation and assumes that the user will know what to do with it. The visualization is designed to be explored in a nonlinear manner, mixing variables in any combination, to give the user maximum control. The ability to overlay different variables for comparison facilitates the detection of patterns and relationships between the different kinds of data, though a color map of values is not the most efficient way to search for correlations. Interpreting any variations detected would also require expertise.

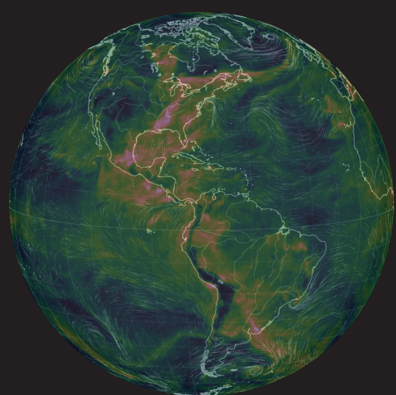
This visualization also provides a detailed "about" section that defines terms, cites original data sources, and provides basic information about how the data was processed for visualization. This section reads more like a footnotes section than a continuous narrative, but this is in keeping with its intended use as a tool for an expert audience.

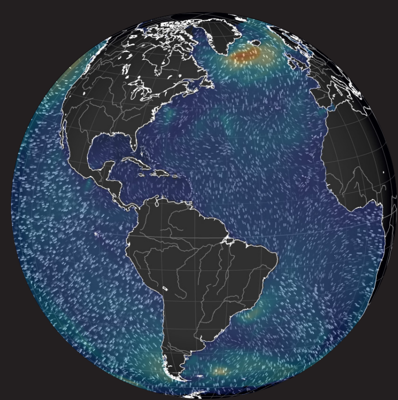
Despite the lack of contextual information to orient a lay user, this visualization is highly appealing to a general audience, simply because of its aesthetics. The way that the particles are used to illustrate the data makes the piece mesmerizing to watch, and draws the user in to engage more deeply with the data. The aesthetic excellence of the piece evokes a sense of wonder and emotion that would not emerge from the data alone. This example serves as a good reminder that factual and aesthetic purposes need not be at odds with one another in the context of a visualization, particularly when designing a graphic to have popular appeal.

Figure 11. "Earth" by Cameron Beccario is a 3D visualization of global wind and weather patterns that allows users to set parameters and view weather predictions from several different data systems.









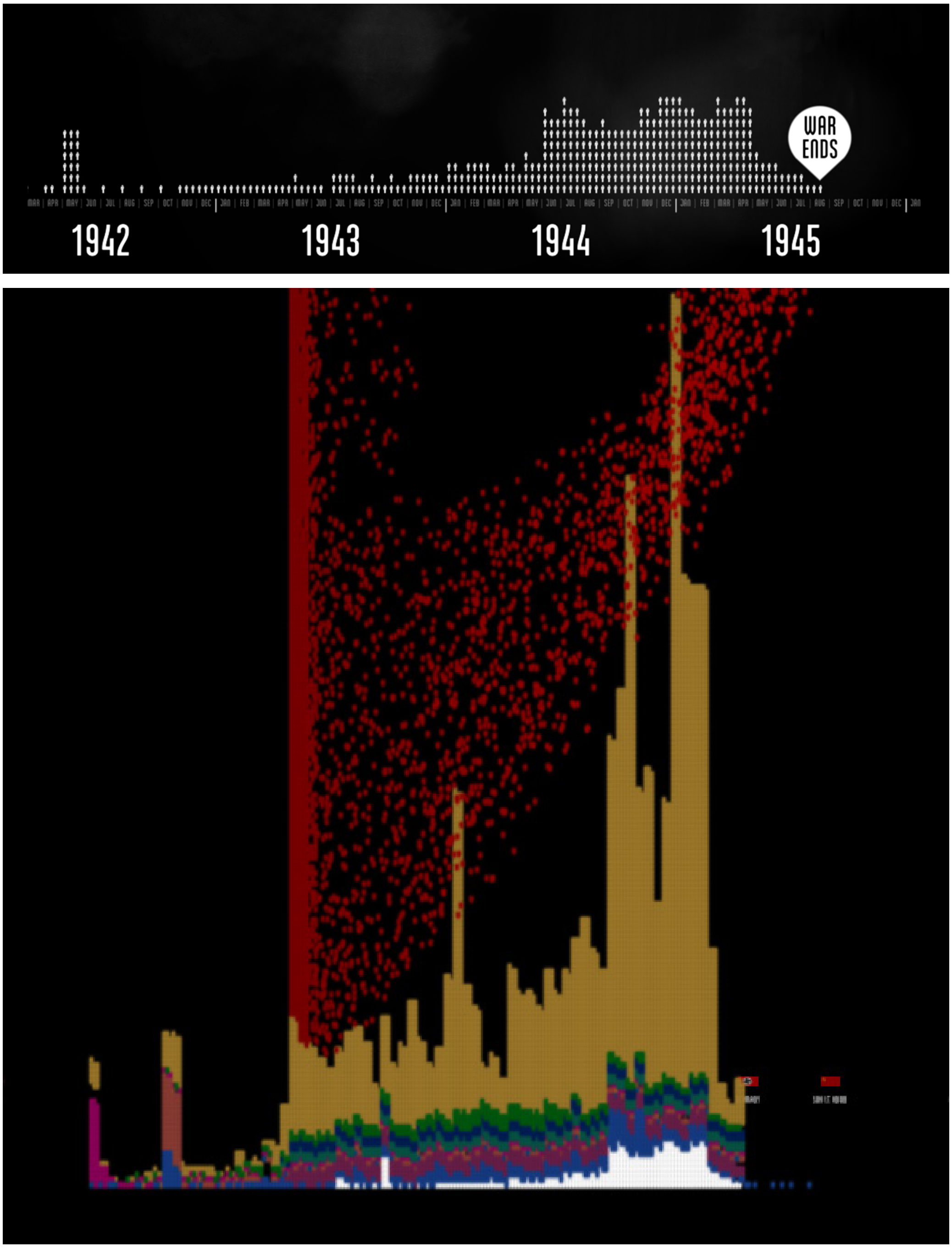


Figure 12. "The Fallen of World War II" is a data documentary by Neil Halloran that uses cinematic techniques to show the human costs of war.

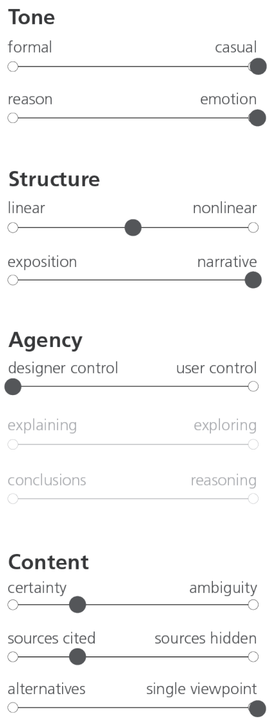
“The Fallen of World War II” is an animated data documentary by Neil Halloran, showing the casualties sustained during the second world war. Most of the visualization consists of a video using isotype people to build bar charts representing casualties for each country and in different battles during the war. At a couple of locations during the animation, the video pauses and allows the user to mouse over the graphics to reveal more details about the data presented.

Although it does focus on depiction of quantitative data, this visualization is clearly intended for a broad audience, and focuses more heavily on telling a story than on comparing quantitative values. Use of narrative and dramatic background music adds to its emotional impact. Halloran refers to his technique as “cinematic data visualization,” and borrows heavily from cinematic storytelling techniques in this film.

The visualization is a highly structured, linear narrative that guides the viewer through a story, using data to emphasize the human costs of war. The piece is designer-controlled, and focused on explaining the topic and creating a sense of scale. The narrator does discuss some ambiguities in the data, such as the difference between civilian and military deaths, but most of the visualization presents numbers with a high degree of certainty.

The project website does include links to the original datasets, but does not discuss any particulars of how they were processed or visualized. It is not clear whether there are alternate statistics or interpretations that should be considered, as the dataset appears to be fairly straightforward. There is no discussion of alternate views, in either case.

This visualization demonstrates that even difficult statistics and “simple” visualizations such as bar charts can be appealing to a wide audience, if designed in the right way. It also shows how effective narrative can be in framing a particular dataset.



The “US Gun Deaths” visualization by Periscopic shows the number of people killed by guns in the US in one of two different years. Each person is represented by a single line, which are drawn in one at a time. The first few lines are drawn in slowly to orient the viewer, and then the rest draw in faster, as the total counter ticks up. The curves are orange until the age of death, and gray thereafter. The visualization lists the person’s name, age, and their date and place of death, as well as statistically-derived predictions of what their future lifespan and causes of death might have been if their lives had not been cut short by gun violence.

This visualization is clearly intended to maximize emotional impact. Its focus on people rather than statistics gives it a narrative bent, and use of phrases such as “stolen years” makes the author’s position clear. Besides those indicators, the data is largely allowed to speak for itself. The framing is evocative, but the visualization itself is strongly data driven.

The introduction to this piece is entirely linear, and designer-controlled. Once the initial visualization is complete, the user is able to explore the data in a nonlinear fashion, using a variety of filters to explore different subsets of the data. There are also a series of high-level take aways displayed at the bottom of the screen after the introduction is complete; these give more statistics about rates of incidence and the distribution of victims.

The sources tab lists specific details of where the data come from and how the statistical projections were made. It also discusses differences in the data presented for the different years, based on changes in the source of the data used.

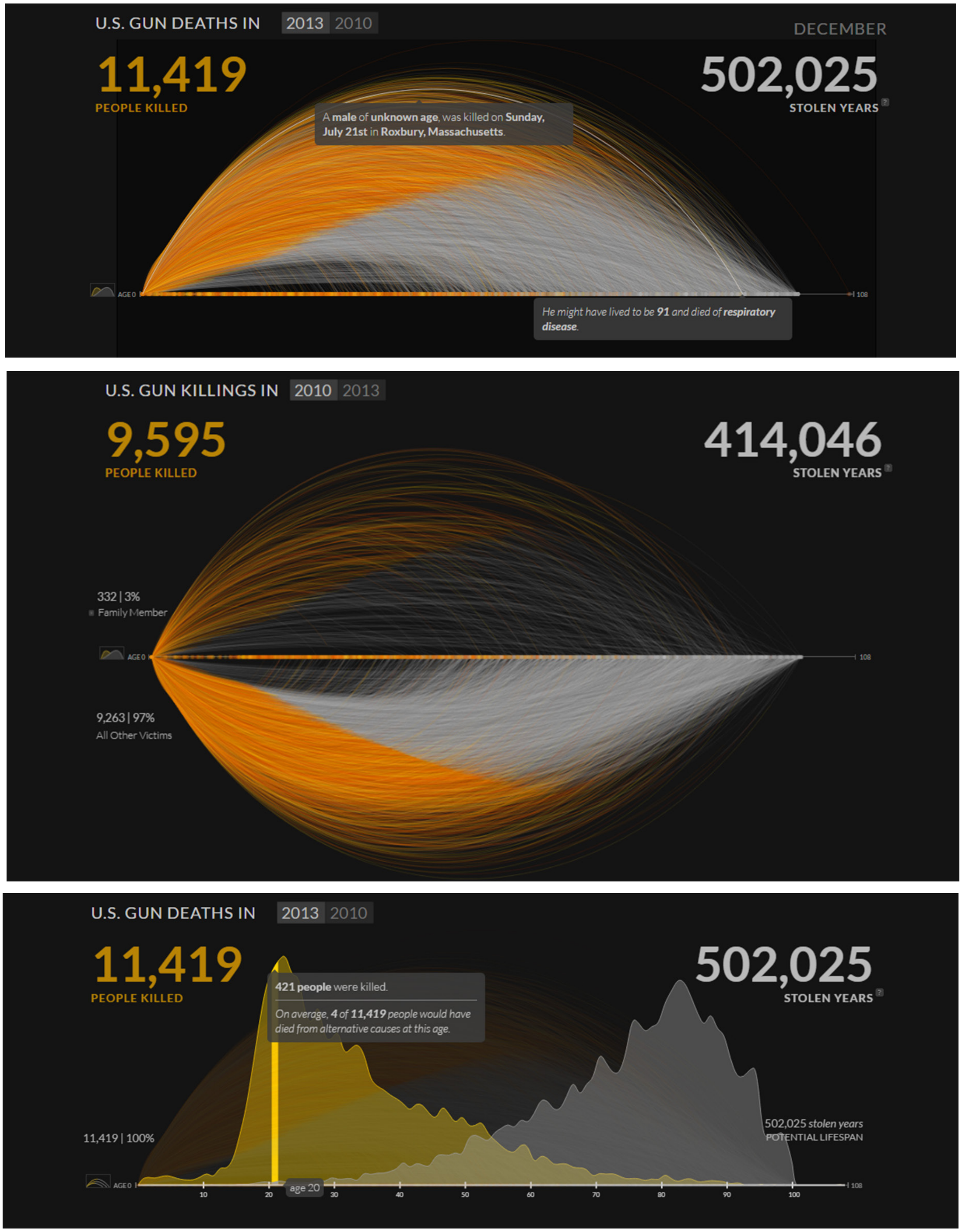


Figure 13. "US Gun Deaths" by Periscopic humanizes the statistics for gun-related deaths in the United States.

To manage a system effectively, you might
focus on interactions of the parts
rather than their behavior taken separately.

Russell L. Ackoff

SECTION FOUR

Design for complex systems

Here, I reflect on the examples from the previous section to create a list of considerations and best practices to guide my own experimentation in communicating interdependent systems. Following the categories established at the beginning of this chapter, the primary considerations can be grouped into questions of tone, structure, agency, and content.

Choosing an appropriate tone for a visualization is an important part of allowing users to engage with complexity, and can also create a strong emotional appeal. Curiosity and a sense of wonder help to increase interest and motivation. A clear sense of context and direction combined with high user agency allow people to engage in active exploration of the topic. Discussions of scale, data provenance, and uncertainty help to inform the comparison of different perspectives, and create opportunities for deeper discourse.

Tone: managing overwhelm

Complexity is overwhelming. People naturally want to minimize the level of detail and the number of connections within a dataset in order to reduce the system to something that feels manageable. Richard Saul Wurman discussed this phenomenon in his book *Information Anxiety* in 1989. Especially in topics related to science, people are unsure of their ability to understand. They are also anxious about the sheer quantity of information that there is to know, and afraid of the consequences of not understanding. This tendency has only increased in the intervening years. Anxiety narrows perspectives and encourages people to seek for quick answers and simplistic interpretations. When overwhelmed, people tend to revert to heuristic models and preconceived notions or ideologies. In order to counter this tendency, designers need to limit viewer's experience of this negative emotional state. In order to preserve an appreciation for complexity and interdependence, this simplification must be balanced against a need to fully inform.

Communicating complexity can be viewed as a cycle where the viewer needs to move through a series of stages of increasing detail in order to understand the system as a whole. A friendly, approachable tone is key to creating an environment in which a user can learn. Avoiding jargon, reducing the use of judgmental or dismissive framing, and explicitly encouraging reflection and curiosity can reduce anxiety and enable people to engage more fully with a confronting topic. Framing a topic in terms of wonder, intrinsic value or beauty rather than threats of disaster or loss can fuel curiosity and encourage interest. Once the viewer is open to receiving information, we can explain the data and discuss its implications, relate it to other knowledge and information sources, resolve disagreements, confront mistruths, and clarify difficult concepts. Finally, we complete the cycle by showing how this understanding creates opportunities for future improvements.

Often, visualizations for the public short-circuit this approach using "fun facts" that jump straight from wonder to exciting possibilities. Scott discusses the increasing role of spectacle in nature documentaries produced for public consumption in her 2003 paper (Scott, 2003). She argues that televised nature programs have moved away from

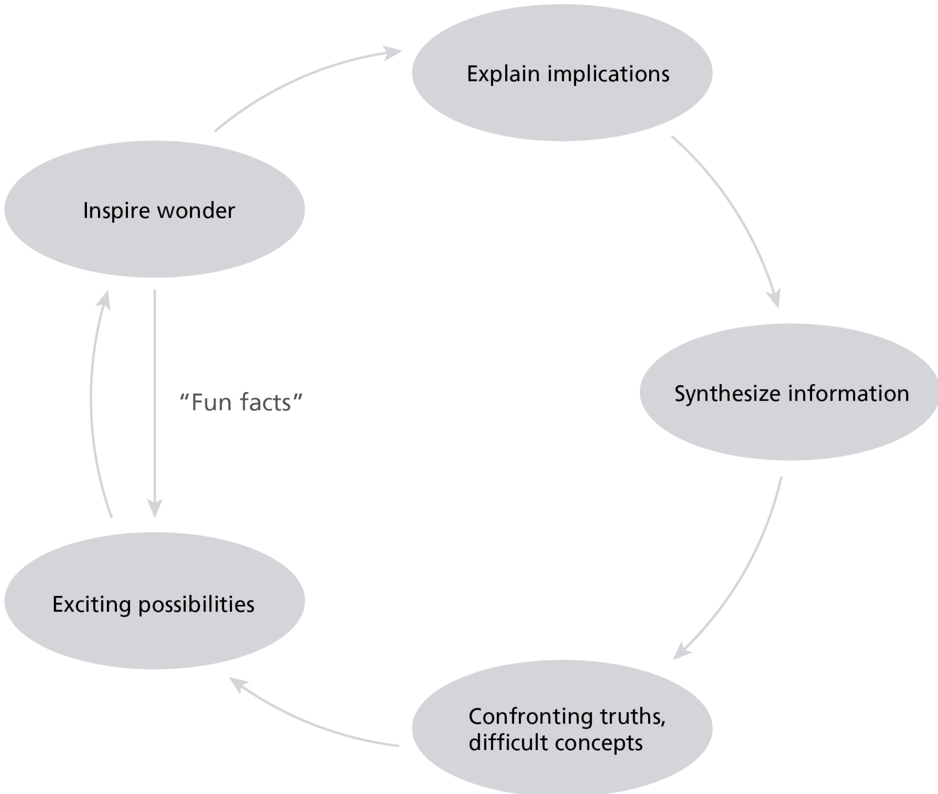


Figure 14. Schematic for leading an audience through a complex system in stages, allowing them time to absorb and understand the information.

a curiosity-driven approach to a showier, attention-grabbing mien. This approach seems quick, easy, and appealing on the surface, but it lacks the intellectual substance and value of deeper exploration and understanding. Over time, this fast food mentality erodes an otherwise healthy public understanding of science, which requires access to education about true scientific process, as well as basic data literacy.

The importance of wonder

Wonder is a powerful tool in the fight against overwhelm, because it reduces the fear response and encourages curiosity. In this positive emotional state, the viewer is motivated to learn more about the topic, despite the possibility of discomfort caused by complexity. Designers can provide the aesthetic components necessary to provoke a sense of wonder, and they can also provide a structure that rewards the viewer's curiosity with new knowledge, encouraging them to continue their exploration. By starting with the obvious things that anyone can see, we can progressively reveal deeper layers of knowledge to expose the structures underneath. Similar to the appreciation of fine art, true understanding of complexity can be seen as a process of gradually reintroducing nuance, relationship, and technique into a person's understanding of a topic.

Structure

A viewer that is both curious and oriented to the possibilities of an information space will usually want to explore. The structure of a visualization should reflect the purpose it is intended to serve. A simple, didactic narrative will likely have a linear structure, while an immersive exploratory experience will usually be less tightly controlled.

Some visualizations rely on pre-programmed narratives to guide the user through the data. Although these visualizations do provide options to the user, they are usually limited in number. Once a selection is made, a user is simply along for the ride. If they become disinterested or bored, their only option is to switch narratives or stop using the tool entirely.

An alternate model is to allow the user to pre-select a narrative of interest and to provide curated recommendations for further exploration at each stage, based on that initial choice. In this way, the user can choose to follow the recommended path, or take detours along the way to explore topics that suit their interest. To further facilitate this selection, it is helpful for the interface to provide a preview of topics available, so that the user can make an informed decision about which to select. In order to create a fully modular solution, each story must function on its own, but should also be connected to others in a way that reinforces its relationship to the whole. The use of transitions and content that adapt to the information that the user has already seen helps to maintain a sense of continuity in the user experience. The interface design should reinforce the impression that these modular representations are part of a greater whole; it should reflect the uniqueness of each subtopic, but also indicate its connection to the broader context of the visualization.

Agency: allowing the user control

Returning decision-making power and agency to the viewer allows them to regain a sense of control and self-determination in their exploration of the data. Allowing the user to make choices at each stage of the journey is equivalent to asking their consent; rather than signing up for an experience that they cannot change, the viewer is able to negotiate his or her own experience in potentially powerful ways. Rather than feeling like passive recipients of knowledge, consensual interfaces allow the user to negotiate their interaction with the information, and may create an experience that is more meaningful for them. If a particular area or topic begins to feel overwhelming or unwelcome, the user is able to back out to a different level until they regain their sense of balance. Providing opportunities for people to periodically advance and withdraw allows them to manage their own engagement level in ways that feel acceptable to them, and reduces the need to dissociate entirely from the tool if the information becomes uninteresting or overwhelming in a particular section.

To facilitate this self-guided approach, an interface should have a clear structure, and a contextual overview that shows how different components of the system relate. The form of this contextual mapping depends on the narrative structure of the piece, and should reflect the possibilities offered by the interface. This allows the user to see what options are available, and may also record where they have been. It also helps the user to understand which features are important, and why.

The presence of such an interface can also help the viewer to engage with the narrative at multiple levels, which is essential for understanding a complex system. Figure 15 shows an image of a rabbit warren that was filled with concrete and then excavated as part of producing a nature documentary. From aboveground, it is impossible to tell how these complicated tunnels connect, and the motion of rabbits through the warren is a mystery. Collapsing all of the junctions into a single graphic makes a crowded visualization that is hard to read, and it may not be clear why or how a rabbit can move between one pair of nodes but not another. Separating the tunnels into stacked views—analogous to varying levels of detail in an information visualization—creates a better

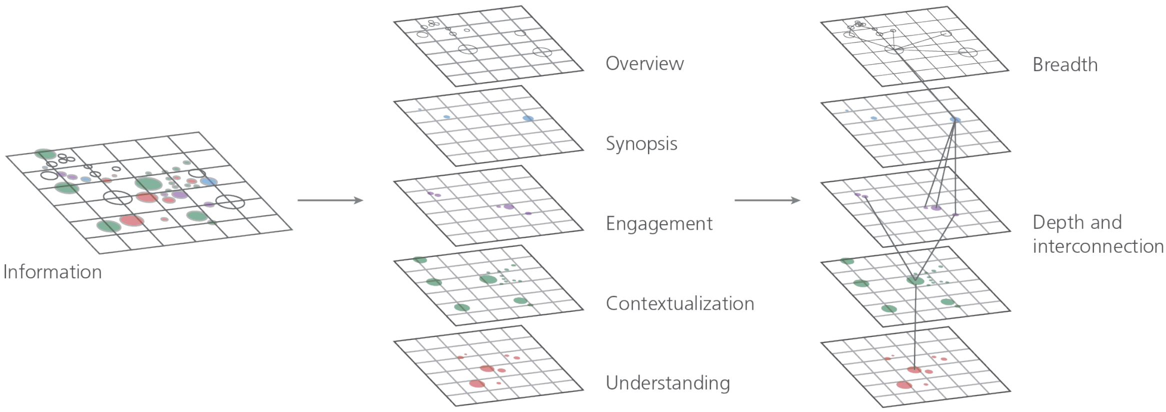
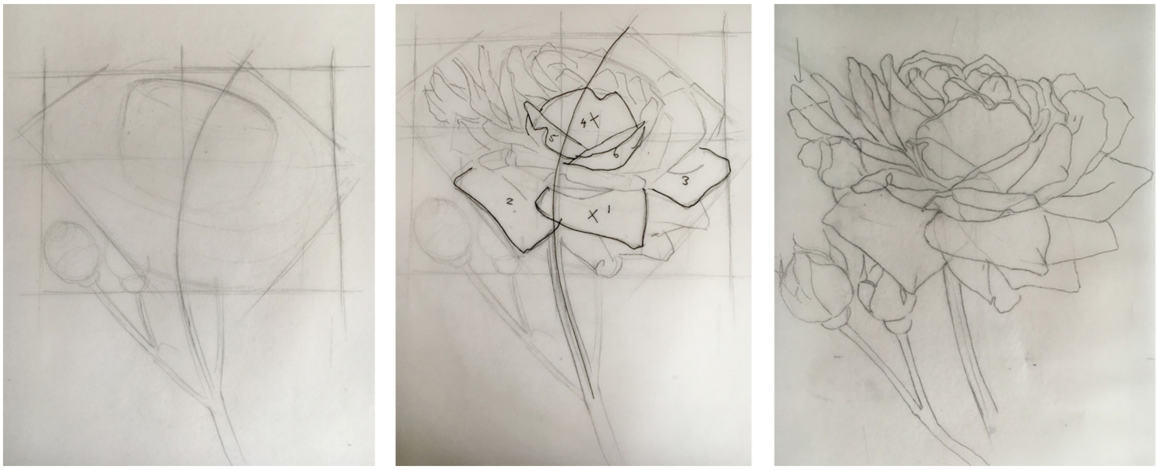


Figure 15. Excavated rabbit warren and a diagrammatic representation showing how the different levels of tunnels connect.

understanding of the space. From this stacked view, a person can then begin to trace out different paths that connect the tunnels, and can gain a better sense of how to move through that space.

With such an overview to guide a user in moving through different levels of detail or specialization, an information system can provide better opportunities for the user to explore connections between the data and compare topics at similar levels of detail. Subsections of the visualization would ideally include an overview or contextual information that reflects the current user position within the information space, and indicates the navigation options available at each point. Ideally, the interface will also allow the user to preview the destination for each option, to allow them to make fully informed selections.



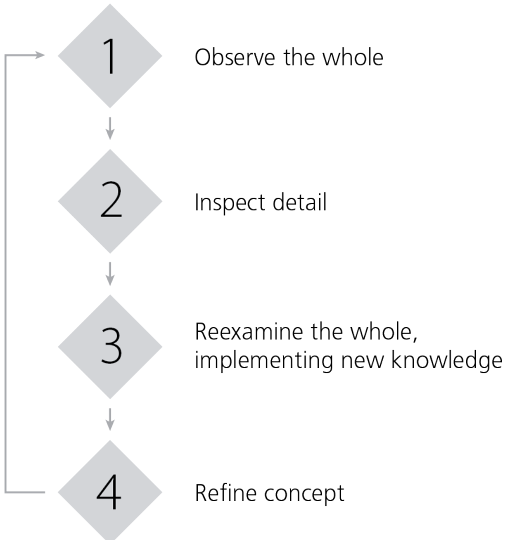


Figure 16. Drawing is an iterative process that reveals increasing levels of detail.

Content: explaining context, modulating depth

In many interdependent systems, it is difficult to achieve a single synthetic overview because the topic is so broad and each area of expertise is so deep. An information visualization needs to show enough depth to reveal connections that might not be apparent at the surface, but not so much that it obscures the overall structure. Designers can create this synthetic overview by drawing relations between these various levels of detail, and simplifying the overview to reveal the underlying structure.

In the ideal case, each level is self-contained and consistent, and could be viewed on its own. The different layers make reference to one another, but understanding the system does not rely on one particular approach to reading the visualization. In some cases, it might be beneficial for a person to move horizontally within a layer, or to adopt a purely vertical approach; the overview supports both options, within a consistent context.

Often, it is necessary to iteratively adjust the level of detail to gain a better understanding of a topic. An artist begins by sketching a rough outline of a shape, and then refines the drawing by working on one particular detail, and then steps back to assess the art as a hole. Creating a balanced piece of art (or science) depends on the ability to relate to a topic on both a microscopic and a macroscopic level. Switching between a broad overview and careful detail builds understanding of the bigger picture and constantly checks for consistency. The area of focus for the individual details changes throughout the process, gradually layering on new information to reveal the final image.

Where possible, information visualizations should support the user in this process of progressive knowledge accumulation. Proving a clear overview coupled with the ability to explore in detail as well as the agency to regulate their involvement makes the user an active participant in the learning process.

At each stage of the visualization, it is important to provide context to orient the user to the current visualization and the system as a whole. Transitions play a large role in shaping the narrative and supporting the user in moving through an interdependent system. An ideal visualization would be designed from a fully user-centered perspective, and could adapt itself by changing text or the starting conditions for the visualization based on the user's entry point, and based on the content that they had previously viewed. This kind of flexible structure would be specifically designed to allow multiple paths through the data, and creates a custom experience for each user, for a more seamless experience.

Regardless of how contextualization is achieved, at each stage of the visualization, the user needs to know how the current topic is related to the previous one, what it is showing, and why it is important in the broader story. When working with information in a scientific context, it is also important to include information on the origins and limits of the data, and to help people interpret what it means.

Finally, it is vital to communicate the extent of scientific consensus, and to explicitly discuss where and why people disagree about the interpretation of results. This may require us to differentiate between correlation and causation, to discuss the meaning of probabilities, and to use scenario-setting to help people understand the implications of different choices. Visualizations that incorporate simple models—such as the Ladder of Abstraction and p-hacking graphics in the previous section—can be extremely effective at helping people understand interrelationships, and can encourage them to engage in hypothesis-forming to try on different views. Throughout the discussion, it is important to acknowledge the existence of uncertainty in both the data and its interpretation, and to explain how those considerations relate to the conclusions that we present.

Great design is great complexity presented via simplicity.

M. Cobanli

Truth has many dimensions, and the way
you arrive at truth in complex
situations is through many perspectives.

Eric Kandel

SECTION FIVE

Visualization as discourse

The previous section focused on extracting recommendations about how to represent complexity based on the successes of recent data visualizations. In a sense, that section creates a set of guidelines based on design as it currently exists. This section is more speculative, and focuses on the design of interactive data visualizations as they could be: as a framework for facilitating dialogue and encouraging participatory engagement with complex information systems. To reflect this emphasis, I will introduce a new category to the list of guidelines: conversation. This section explores how designers might use interactive data visualizations to support conversation.

A defining feature of complex systems is that they can be viewed in many ways; "wicked problems" result from the wildly different frames that participants bring to the debate. To avoid accusations of prejudice, it is important to show how scientific models relate to different ideas or cultural frames in an approachable and constructive way. Acknowledging the existence of (viable) alternate interpretations is an important step in helping individual people connect their point of view with the one we hope they will choose. This does not need to be presented as a choice between equal alternatives, but it does need to respect and give credit to the beliefs that people bring to the table.

Persuasion requires that a new model be fit into an existing one: an argument that does not reflect a person's initial framing cannot change their minds. When trying to convince a skeptical public of scientific facts, we need to show how the things that they "know" can be viewed from a different direction, and that there is a logically consistent point of view that matches their lived experience. We cannot shy away from confronting errors or fallacious thinking, but we must also provide alternatives that are both convincing and reasonable to consider—from the point of view of the audience we seek to persuade.

The act of demanding evidence and questioning power is a central component of scientific inquiry; true science communication must embrace this tendency. Consensus cannot be created from either-or choices, arguments of authority, or threats of exclusion. Such responses only serve to reduce questions of fact to questions of popularity. Rather than shunning those who question scientific results, we need to find ways to help them bridge the gulf between their perceived realities and those that science holds to be true. Beginning from their own initial framings, we can then create opportunities for people to explore the evidence and examine alternative interpretations.

It is also critical that this public discussion be a true conversation, where both sides seek to learn and understand. Poorly-disguised proselytizing, legalistic debate, or intellectual salesmanship that seeks to coerce or convince rather than to understand will be perceived as manipulative and rejected as such. In the end, this approach leads to disengagement, resentment, and a deepening distrust.

Bridging these divides requires input and cooperation from both sides. Scientists should retain authority in questions pertaining to their particular disciplines, but they should also be willing to embrace skepticism and discourse with the public as a way of generating new ideas and approaches to scientific inquiry, and as a critical component of translating scientific fact into appropriate public policy—an area outside the realm of scientific expertise.

Robust public discourse can also help to inform scientific inquiry. Ongoing conversations with people affected by existing or proposed technologies and public policies can reveal new aspects of a problem,

> The best ideas emerge when very different perspectives meet.
>
> *Franz Johansson*

expose potential pitfalls, and help to identify likely causes for rejection. In turn, these concerns can create new avenues of research for scientists to pursue, or provide fresh perspectives on existing problems. Recognizing the potential value of public contribution to science and engaging in constructive, long-term discussions with a general audience will help to resolve thorny policy issues and can reignite public interest and engagement in scientific progress.

Designers can facilitate this conversation by moderating interaction between scientists and the public, by creating environments that help people to understand interdependence, and by creating visible representations that can be used to compare different approaches to a problem. Although a dialogue-based approach is not often incorporated as a primary component of interactive data visualizations, it is a well-established method in design. Gaver, Dunne and Pacenti published a case study illustrating such an approach in 1999 (Gaver, Dunne & Pacenti, 1999). They discuss the importance of using a series of "cultural probes" to gather information and perspectives from a focus group to guide a UN development project. These contributions were then analyzed and used by the design team to refine their understanding of the situation and guide their approach to create a more inclusive design.

Tools that create a tangible record of a person's interactions with and thoughts about a complex issue can be used as a starting point for mapping the mental models that people construct during the experience. Tracing different thought paths through the same information can be useful in studying and archiving the process of learning. Examining differences between these different traces would help to emphasize that there are multiple approaches to understanding the same problem, and can help the user to discover connections between topics that they might not expect. Inviting people to compare their results will identify points of agreement and contention, allowing these external representations to become starting points for further conversation and perspective-sharing.

Make it personal

The slow food movement could form a useful model and a modern cultural framing to help support critical engagement in science. "Slow food" grew out of the sustainability movement, and appeals to many people across the political spectrum who are interested in getting back to basics and regaining a sense of control in their lives. It repudiates mindless overconsumption of commodities and speaks to people who are suspicious of easy options and traditional authority figures, encouraging people to take individual responsibility and control over the things that they consume. People who are attracted to this movement tend to be highly engaged, and are willing to do a great deal of research and often make large lifestyle changes in response to their findings. These people tend to be tinkerers and experimentalists who want to find ways to optimize their health by becoming informed.

Although it can often be manipulated by people peddling suspicion and fear, this movement also speaks to people's desire to become personally informed and to improve the quality of their lives through research. With appropriate information sources and improved critical thinking, this section of the public could be a strong force for creating an engaged audience in support of science. A similar argument could be made for maker spaces, citizen science initiatives, and a host of other venues where thoughtful, concerned citizens seek to improve their lives. The ethos of personal responsibility and individual decision-making appeals to people on both sides of the political aisle, and could form the basis for improved public education in science. It may be that grassroots science already exists, and is just waiting for an appropriate venue to act.

Much of the power of the slow food movement is its focus on change at a personal and local level, where people feel that they can be in control. People need and want connections to local actions with direct results that support their convictions. Hypothetical statements of possibility ("if we all did this, X would change") fail in cases where people do not have faith in a communal endeavor, or where they have no sense of a shared social contract. Refocusing on personal choice and local actions can help to overcome feelings of irrelevance and empower individuals to act.

Connection to social media facilitates interpretation of personal choices as status symbols or social currency, and can motivate personal change through added accountability. The current popularity of personal data tracking initiatives like FitBit reflects this trend, and could provide an interesting model for future behavioral design. In developing data infrastructures to support change, designers, scientists and data-collecting institutions need to consider the scalability of data from local to global and individual to societal scales, to support action and representation both of problems and of progress.

One intriguing possibility is to create systems that not only represent, but that evolve in response to user interactions, perhaps allowing the user to adjust the visualization in lasting ways. Such visualizations of impact would allow a user to feel the weight of their own contributions in the context of a larger issue. Such responsive design might change form throughout the user interaction, evolving toward a more complex visualization over time. In this case, the interface must provide contextual information to reorient the users as they move through each stage.

In its simplest implementation, such a visualization would ultimately reveal a similar representation for each user, with subtle changes that reflect their own interactions with the piece. A more intricate version might create a drawing that is entirely dependent on the user's choices, and that could be used to represent their experience in some way. Using principles of parametric design and generative art, one could create a visualization that incorporates feedback from the user throughout the learning experience, and visually represents the construction of a user's mental model throughout the process of interacting with the piece. If these final visualizations maintained sufficient connection to the original visualization space, they could be used as external boundary objects to facilitate discussions of the user experience. Such representations could be used to identify similarities and differences in perspective, creating ground for deeper discussion of cultural frames and mental models.

Summary of guidelines

Tone

1. Use wonder, aesthetic interest and curiosity to encourage engagement

Structure

2. Provide narratives to guide the user

3. Support undirected exploration — use modular components, make multiple cross-references

4. Show relationships and use transitions to preserve context

Agency

5. Provide opportunities to switch between narratives. Allow users to preview selections.

6. Indicate user progress/history

Content

7 Clearly separate data from interpretation

8 Explain limits of certainty and compare alternate interpretations; may be supported by a scenario-setting or experimental approach.

Conversation

9 Create tangible records of interaction

10 Support action; incorporate the personal/local, allow user to contribute to information shown

The health of soil, plant, animal, and man is one and is indivisible.

Albert Howard

CHAPTER SIX

Soil as Case Study

Soil is a complex ecosystem that has social, environmental, and scientific implications at both a local and global scale. Soil health forms an ideal case study for developing methods of visualizing interdependence.

Soil is a precious resource. Without soil, there are no plants. Without plants, there is no food. Laid against the backdrop of a burgeoning world population faced with climate change, preservation of soil resources is a pressing concern. As with many ecological issues, soil health is a complex topic with many interrelated parts. Geographic distribution, chemical properties, ecosystems and natural cycles, and human behavior all have direct influences on soil health. By uniting these different perspectives in an exploratory graphic, I hope to promote awareness of this valuable ecosystem, and of the threats posed by human demands.

It takes a thousand years to make an inch of fertile topsoil, and erosion can remove 25 to 40 billion tonnes of topsoil annually. In a 2015 report the United Nations (UN) reports that roughly a third of the world's soil is moderately to severely depleted, primarily due to continued demands caused by population and economic growth (UN 2015). Defined broadly, soil degradation is any process that leads to a reduction in the soil's ability to support life, including erosion, chemical contamination, paving or other permanent structures, and a variety of other factors that contribute to loss of soil function. Estimating soil health is a tricky and sometimes contentious business, but most experts agree that soil is a vital resource under imminent threat. Communicating this threat to the public in constructive ways is a critical component of protecting this vital resource.

Though the situation it reports is dire, the UN report also leaves room for hope; many of the problems with soil degradation are caused by human activity, and can be ameliorated with moderate changes to behavior. The suggested changes include: reducing human impact by practicing no-till agriculture, minimizing chemical fertilizer use and focusing on restoring the carbon content of soils, increasing awareness about the importance and fragility of soil, and improving soil management practices. Enhancing public appreciation for the importance of soil and awareness of the steps needed to address this global issue should be at the forefront of the effort to protect the global food system and replenish soil resources.

When one tugs at a single thing
in nature, he finds it attached to the
rest of the world.

John Muir

SECTION ONE

Preliminary Experiments

This thesis began as a semester-long research project for my Research Methods class in the spring of 2016. The assignment was to get "under the surface" of an object to explore both its origins and its implications. I was inspired by an article written by Michael Pollan to take a closer look at something that I have always found fascinating: dirt (Pollan 2015). After a semester of preliminary investigations, I decided to convert the resulting poster into an interactive website, as part of my thesis work.

Project Overview:

The first phase of this project focused on discovering interesting facts about soil, identifying relevant datasets, and creating a narrative to connect soil health to topics that people can relate to. This process culminated in a final poster connecting global population and land use to soil degradation. In the fall, I developed the theoretical framework discussed in previous chapters, pursued some new experiments, and developed an interactive visualization of global food imports and exports to highlight the interconnectivity of a global society. The final semester of the project focused on unifying the previous experiments and creating a navigation system and site infrastructure that supports the different stories that I wanted to convey. This required clarification of the overall message of the site, consideration of the specific audience I hope to reach, and identification of individual narratives that match these goals. Previous experiments were refined and updated to match this sharper focus, and additional visualizations were developed to create bridges between related ideas. The overall site design evolved throughout this process in response to the needs of these different visualizations. The

Figure 17. Soil sample and sampling locations in my back yard. The sample was weighed, separated into individual components, and baked to drive off water. After baking, I weighed each of the components and found the percent composition of the sample.

final component of the project was the exhibition design, completed during the spring semester of 2017 as a way to explore how design can be used as a tool to begin conversations and support community engagement around important scientific issues.

Soil Autopsy

I began my soil investigation by performing an "object autopsy" to identify the physical components of soil that I could see by eye. I collected a half a cup of soil from my back yard garden for the analysis. The sample was collected in January, when the ground was frozen solid, so it was taken from the very surface of the topsoil and included more mulch and fewer soil organisms than I would expect from a sample taken in the summertime.

I weighed the soil sample on a kitchen scale, and then picked through it with a pair of tweezers to separate the various components. Small rocks and pebbles were collected using a kitchen sieve, but it wasn't possible to separate the very fine clay and sand from the smallest rocks. After grouping the different components into foil muffin cups, I heated them in an oven at 170°F for about an hour and a half to get rid of all the water.

Once the samples were dry, I weighed each of the different components to determine their percent mass, and compared that to the total to find out how much water had been lost. The categories that I measured are as follows: clay/fine rock particles (37%), water (35%), rocks (13%), fine mulch/rock particles (7%), and mulch (7%). I also found a few seeds, some bits of dead leaves and pine needles, a little bit of dried moss, a few plant roots, and a couple of hibernating insects.

In completing the soil autopsy, I was reminded how many important parts of soil are invisible; the chemical composition and microorganisms that make soil what it is can't be seen with the naked eye, or separated with a pair of tweezers. This perspective also pushed me to explore the invisible implications of soil, including its use in food production and climate change regulation. For something that seems so direct and concrete, it is really only a short step to much more abstract topics that are harder to measure, and to see.

Soil

What is it?

Scientific definitions of soil vs dirt, soil vs. sand, etc.

What's it made of?

Compare different types of soil and their compositions.

- Humus
- Rock/mineral
- Water
- Air

(become more specific about humus and mineral content, especially. Discuss location and composition in relation to other properties of soil (such as pH)

Where is it found?

Geographic comparison of soil types, land use, and land area. Where is the fertile land, and why?

Figure 18. List of questions to guide early conceptual framing and approach to soil research.

What lives in it?

Discuss the soil ecosystem and interdependence. How does the population of "healthy soil" differ from "depleted soil"? Major ecological players:

- Arthropods
- Nematodes
- Fungi
- Bacteria
- Worms
- Plants

Discuss how ecosystem balance affects nutrient availability and soil stability.

How much is there?

Sense of scale - how does soil volume compare to the volume of the oceans? The land area of the earth? The volume of the planet?

Why is it important?

Discuss political, social, and environmental implications of soil.

These topics include:

- Water management and aquifer depletion
- Crop production and use of fertilizers, as well as chemical runoff from phosphates/nitrates, which cause dead zones
- Climate change; soil as a carbon sink!
- Flood prevention
- International conflict as a product of food shortages
- Political questions such as farm subsidies, organic/conventional agriculture, GMOs, and how to optimize extraction of human nutrition from the soil, in a changing climate and for a growing population.

Preliminary questions

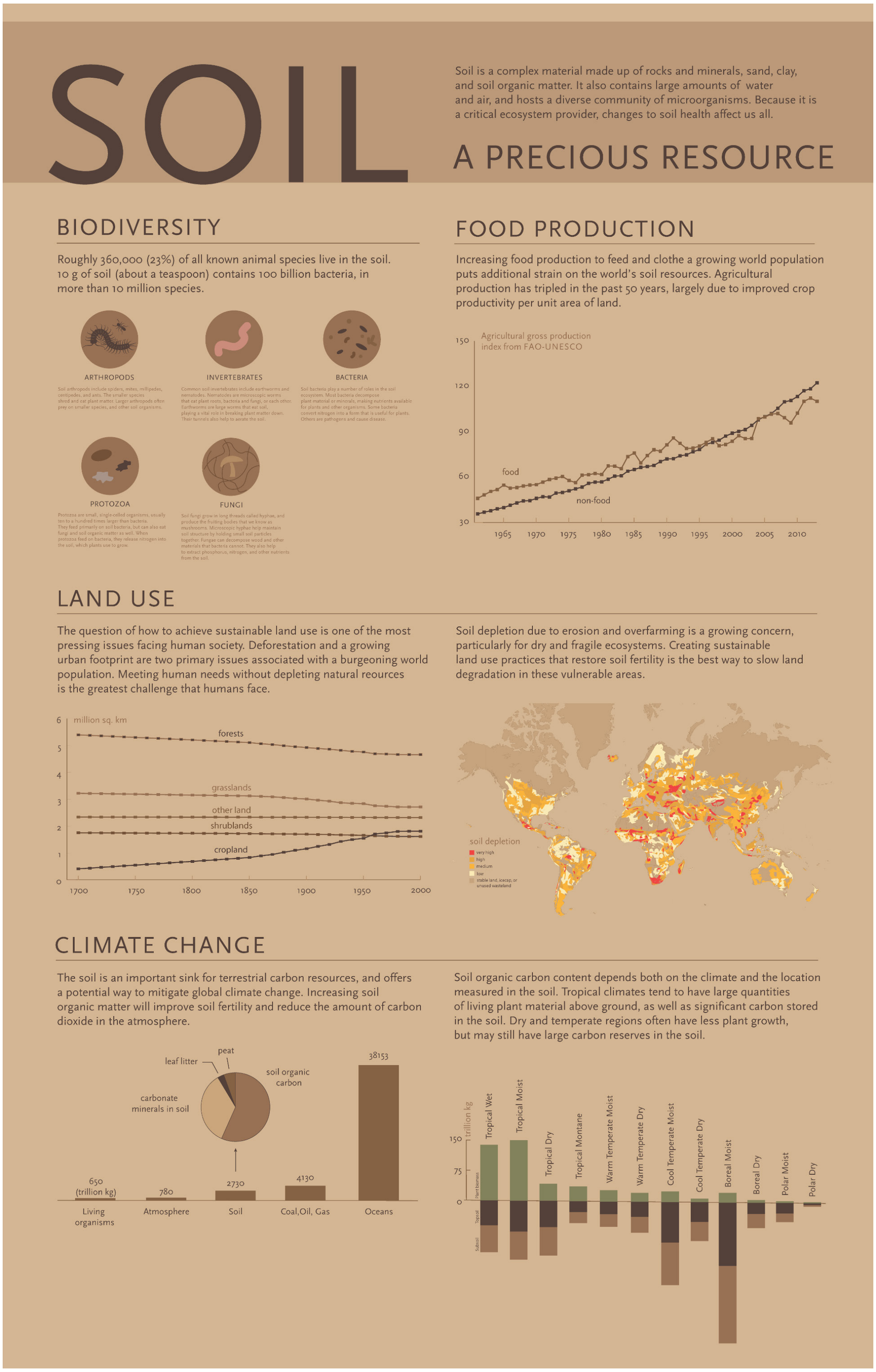
My early reading on soil issues uncovered a wide range of questions and topics related to soil health. I compiled a list of these first impressions to capture where I started the research process, as summarized in Figure 18. This early phase of the project was driven largely by the need to create definitions and identify quantitative measures that related to my topic. From this initial list of questions, I identified the impacts of agriculture on the climate and environment as a core set of issues related to soil. With this somewhat narrowed focus, I began performing a broad survey of the scientific literature on related topics.

I compiled the relevant data into an early poster draft (figure 19), focusing on global trends in land use patterns. My general sense from our class midterm review was that people were interested in the facts and understanding the connection of soil to the larger ecosystem, but were having difficulty grasping the scale of the issue or finding personal relevance from the "big picture" view presented by the global data. For the next draft, I chose to focus on the human issues related to soil health, and to explore the statistical data in more detail.

The World Bank publishes a collection of World Development Indicators that include land usage information for each country in the world, recorded between 1961 and the present. Plotting the 10 countries with the largest difference in forested land between 1961 and 2013 (the first and last years with complete data) indicated that it would also be important to look at the relative land areas reflected in these percent changes—a 10% change in the Virgin Islands doesn't mean the same thing as a 10% change in the US or India.

Environmentalists often talk about the rates of deforestation in Brazil and other developing nations, especially in the tropical regions, but I had also found reports that stated that forest area was actually increasing in some countries because of the switch from agricultural to industrial economies. I decided to look at the data for Brazil and the US, since they are of roughly equal size and are at very different stages of economic development. The US has had a stable total land area, a decrease of about 5% in agricultural land area, and an 0.8% increase in forested area over the time period measured. The population is steadily increasing, but

Figure 19. Midterm poster draft showing the primary narrative that I wanted to convey, along with some aggregated global statistics.


SOIL
Soil is a complex material made up of rocks and minerals, sand, clay, and soil organic matter. It also contains large amounts of water and air, and hosts a diverse community of microorganisms. Because it is a critical ecosystem provider, changes to soil health affect us all.
A PRECIOUS RESOURCE
BIODIVERSITY
Roughly 360,000 (23%) of all known animal species live in the soil. 10 g of soil (about a teaspoon) contains 100 billion bacteria, in more than 10 million species.
ARTHROPODS
Soil arthropods include spiders, mites, millipedes, centipedes, and ants. The smaller species shred and eat plant matter. Larger arthropods often prey on smaller species, and other soil organisms.
INVERTEBRATES
Common soil invertebrates include earthworms and nematodes. Nematodes are microscopic worms that eat plant roots, bacteria and fungi, or each other. Earthworms are large worms that eat soil, playing a vital role in breaking plant matter down. Their tunnels also help to aerate the soil.
BACTERIA
Soil bacteria play a number of roles in the soil ecosystem. Most bacteria decompose plant material or minerals, making nutrients available for plants and other organisms. Some bacteria convert nitrogen into a form that is useful for plants. Others are pathogens and cause disease.
PROTOZOA
Protozoa are small, single-celled organisms, usually ten to a hundred times larger than bacteria. They feed primarily on soil bacteria, but can also eat fungi and soil organic matter as well. When protozoa feed on bacteria, they release nitrogen into the soil, which plants use to grow.
FUNGI
Soil fungi grow in long threads called hyphae, and produce the fruiting bodies that we know as mushrooms. Microscopic hyphae help maintain soil structure by holding small soil particles together. Fungae can decompose wood and other materials that bacteria cannot. They also help to extract phosphorus, nitrogen, and other nutrients from the soil.
FOOD PRODUCTION
Increasing food production to feed and clothe a growing world population puts additional strain on the world's soil resources. Agricultural production has tripled in the past 50 years, largely due to improved crop productivity per unit area of land.
Agricultural gross production index from FAO-UNESCO
150
120
90
60
30
food
non-food
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
LAND USE
The question of how to achieve sustainable land use is one of the most pressing issues facing human society. Deforestation and a growing urban footprint are two primary issues associated with a burgeoning world population. Meeting human needs without depleting natural reources is the greatest challenge that humans face.
Soil depletion due to erosion and overfarming is a growing concern, particularly for dry and fragile ecosystems. Creating sustainable land use practices that restore soil fertility is the best way to slow land degradation in these vulnerable areas.
6
5
4
3
2
1
0
million sq. km
forests
grasslands
other land
shrublands
cropland
1700
1750
1800
1850
1900
1950
2000
soil depletion
very high
high
medium
low
stable land, icecap, or unused wasteland
CLIMATE CHANGE
The soil is an important sink for terrestrial carbon resources, and offers a potential way to mitigate global climate change. Increasing soil organic matter will improve soil fertility and reduce the amount of carbon dioxide in the atmosphere.
Soil organic carbon content depends both on the climate and the location measured in the soil. Tropical climates tend to have large quantities of living plant material above ground, as well as significant carbon stored in the soil. Dry and temperate regions often have less plant growth, but may still have large carbon reserves in the soil.
leaf litter
peat
soil organic carbon
carbonate minerals in soil
650 (trillion kg)
780
2730
4130
38153
Living organisms
Atmosphere
Soil
Coal,Oil, Gas
Oceans
trillion kg
150
75
0
Plant biomass
Topsoil
Subsoil
Tropical Wet
Tropical Moist
Tropical Dry
Tropical Montane
Warm Temperate Moist
Warm Temperate Dry
Cool Temperate Moist
Cool Temperate Dry
Boreal Moist
Boreal Dry
Polar Moist
Polar Dry

the gap between total and urban population remains relatively constant. Brazil, on the other hand, has experienced much larger shifts: a 6% decrease in forested land area and a 15.3% increase in agricultural land. The population curves show that there is a much larger move toward urbanization in Brazil: from 49% urban in 1960 to 85% urban in 2014 (an increase of 36 percentage points).

This exploration convinced me that it would be interesting to look at changes to the land use patterns for different countries. It wasn't immediately clear whether it would be better to focus on a few examples or to show information for all of the countries in the world. I didn't want to blindly choose the biggest and the smallest values, and I was hesitant to select case studies too early in the data exploration process. So, I decided graph data for all of the countries to start with. I focused on land use, degradation, and population metrics, which I compiled manually into a spreadsheet and processed using d3.

Land Use Data

My first draft used a bounding box to represent the total area of a country, and showed the different land use variables as colored areas inside it (figure 20), separated into 5 categories—red: degraded land, green: forest, dark gray: urban, dotted gray line: urban expansion predicted for 2050, light brown: agricultural land, dark brown: arable land. The set of squares on the right show population information; light gray is the total population, and dark gray again shows the number of people living in cities.

A partition layout would be a more standard approach, but I don't find them very readable, and it is hard to make comparisons based on area when the aspect ratio of the rectangles is constantly changing. I wanted to use the consistent square position to identify the different variables, so that you could look at squares on the top left to compare degradation, and all of the ones on the bottom right to see urban expansion.

Requiring that all of the variables be represented as squares meant that they had to overlap in some instances, and the overlap was not actually meaningful. The fact that the land areas plotted don't add up

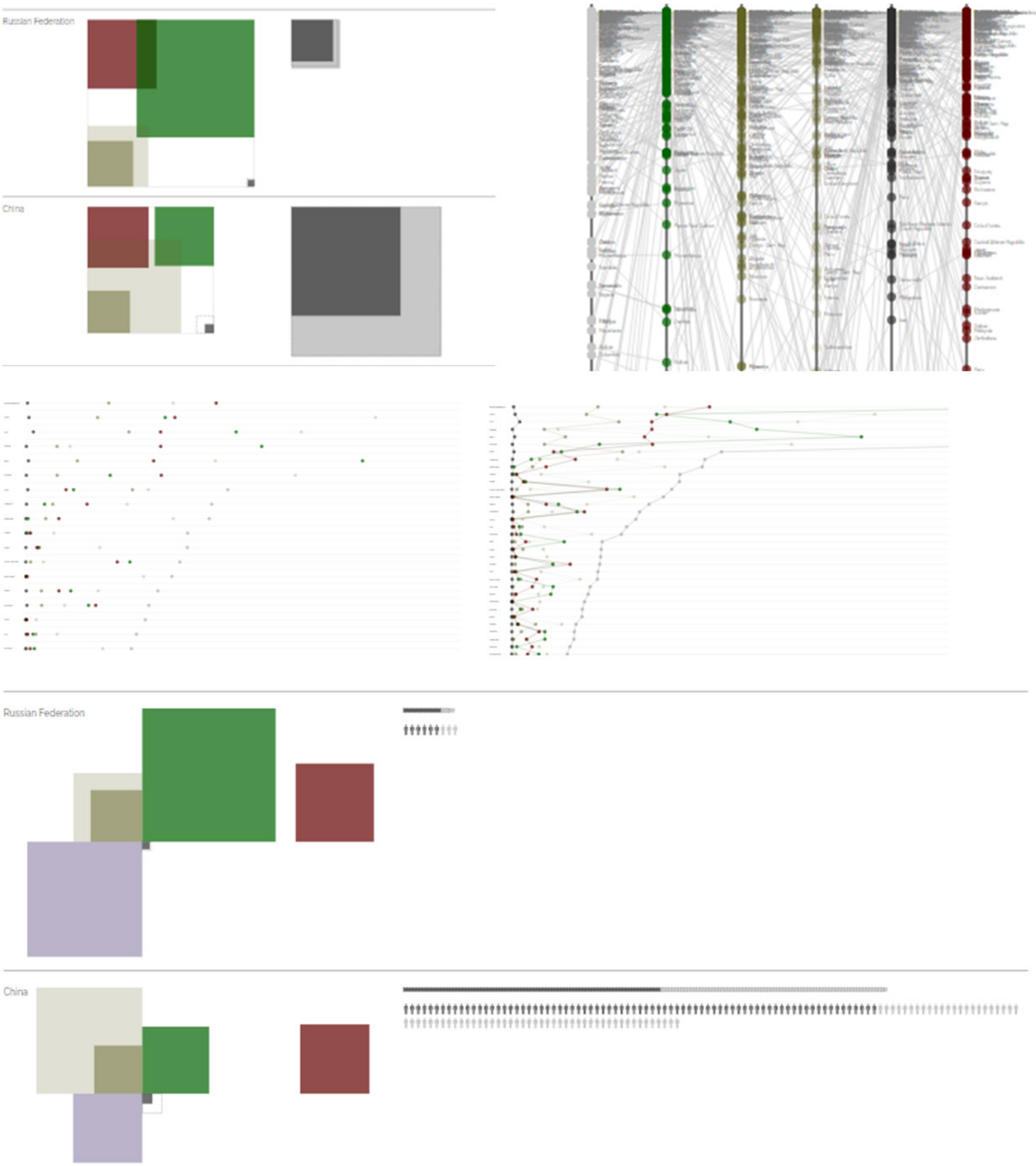


Figure 20. Several early sketches representing land area distribution in different ways.

to 100% created an additional complication. The land degradation and land use data are from different datasets. The degraded land area does overlap with the different land use categories, but is not quantified in the dataset. As I had feared, most people found the overlap distracting at best, and at worst confusing.

Plotting the different land use variables on multiple y axes created an unreadable graph. This might work in an interactive setting, but didn't belong here. I considered using side by side bar charts for comparison of data between countries; while it would be quantitatively superior to area maps, I felt that multiple bar charts would be hard to read and compare quickly across different countries, and wouldn't give the holistic shape signature that the square glyphs provide.

I also tried collapsing the bar chart data into a single line for each country. Since the variables share a single axis, it's possible to plot them all as colored dots on a single line, and then the reader can follow dots of the same color to compare the different countries. This was a more compact way to represent the data, but I found that this was also hard to read, especially with many spots of different colors.

I tried adding lines to help guide the reader through the forest of dots, but the connecting lines suggested trends or connections that didn't really exist—there is no reason to suspect that the land use values for different countries depend on one another, but connecting lines between the symbols tended to imply a dependent relationship. The other disadvantage of the line/bar plots was that they use a linear scale. That's much better for making quantitative comparisons than area encoding, but it also means that the length of the bars for the biggest countries need to be very large in order for the data for the small countries to be visible on the same scale.

In the end, I broke the bounding box and let the collection of squares stand on its own to represent the size of the country.
By creating an "other" category in the land use group, I was able to create a glyph from the collections of squares. Taking the red square out of the country area group eliminated the overlap problem, and helped to draw attention to land degradation as an important, but separate, variable.

I also re-configured the population data in later versions of the visualization. I didn't like using the same representation for population and land area in the first draft of the square glyphs. Instead, I represented population as a bar chart, and indicated the number of people per km of land for each country as a series of isotypes. I hoped that this would give a more tangible sense of crowding. When tested, this also proved problematic. People were confused by the fact that the people icons represented population density rather than total population, and often continued to misread the information in the graphic even after being corrected. The icons also took up a lot of space, and were the only dominant feature in the graphic for the small countries, where the land use squares are often too small to see.

To reduce confusion and create more space, I removed the icons and replaced them with constant-area population density boxes instead. This allowed the squares to be made bigger, so that they became more prominent on the page. Adjusting type style and size and incorporating quantitative labels also improved the readability of the graphic. The top 5 countries in both percentage and absolute land area were highlighted for each category, to provide a point of comparison for readers interested in ranking the different countries.

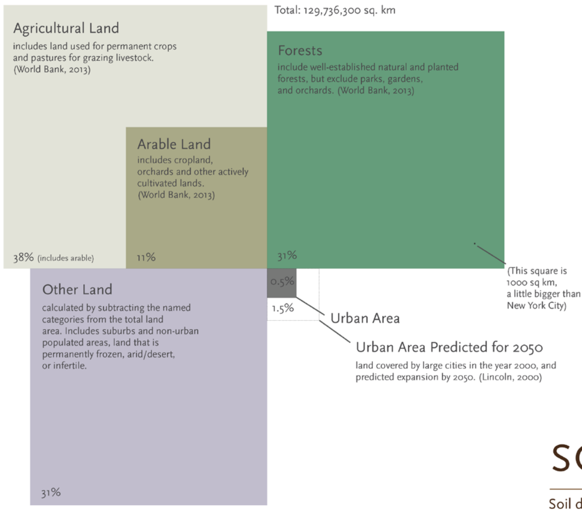
The final version of the poster for the Research Methods class focused on the population and land use data, and included other information about soil importance and biodiversity in smaller graphics that framed the main panel. The aggregated data for the entire world was shown on the left hand side of the poster, and served both as a general context and a key for orienting the user to the meaning of the country glyphs. The soil depletion data was also given a prominent place, to emphasize the relationship between the human and environmental factors.

Feedback on the poster was generally positive; people were interested to explore the glyph panel, and enjoyed searching for different countries and comparing outliers. Some people would have preferred a simpler graphic showing just the top ten countries. Others suggested that it would be better to eliminate the soil narrative and focus only on population and land use, but I felt that it was important to keep the connection to my original "object," and all of the complexity it represents.

GLOBAL POPULATION

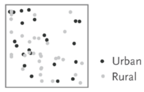
(World Bank, 2013)

GLOBAL LAND USE



People per square km

people living in one square km of land, if you spread the entire population out evenly over the globe. Really, most people live in cities or suburban areas, which have a higher density. (World Bank, 2013)



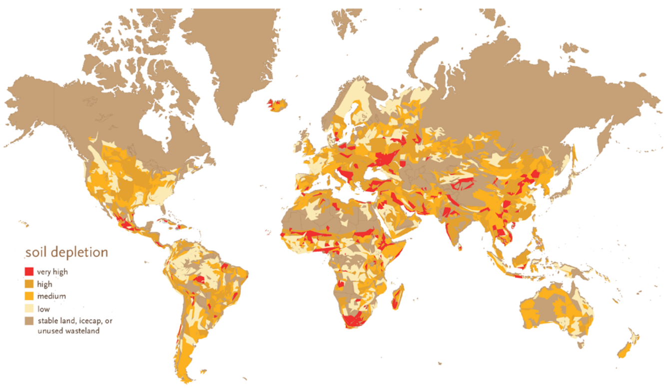
Only care about the biggest?

Countries with the biggest **percent land use** for each category are framed by a **box** matching the category color.

Countries with the biggest **absolute land area** are marked with a **star**.

SOIL DEPLETION

Soil depletion caused by erosion and overfarming is a growing concern, particularly for dry and fragile ecosystems. Creating sustainable land use practices that restore soil fertility is the best way to slow land degradation in these vulnerable areas. (UNEP/ISRIC, 1992)

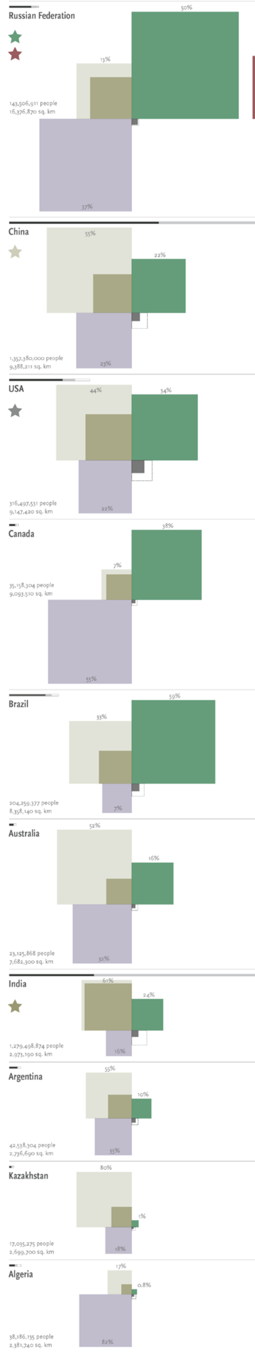


SOIL

A PORTRAIT OF LAND USE

Soil is a complex material made up of rocks and minerals, sand, clay, and soil organic matter, and hosts a diverse community of microorganisms. Soil reduces flooding, may help to mitigate climate change, and provides food for a growing world population. Changes to soil health affect us all.

LAND USE AND POPULATI



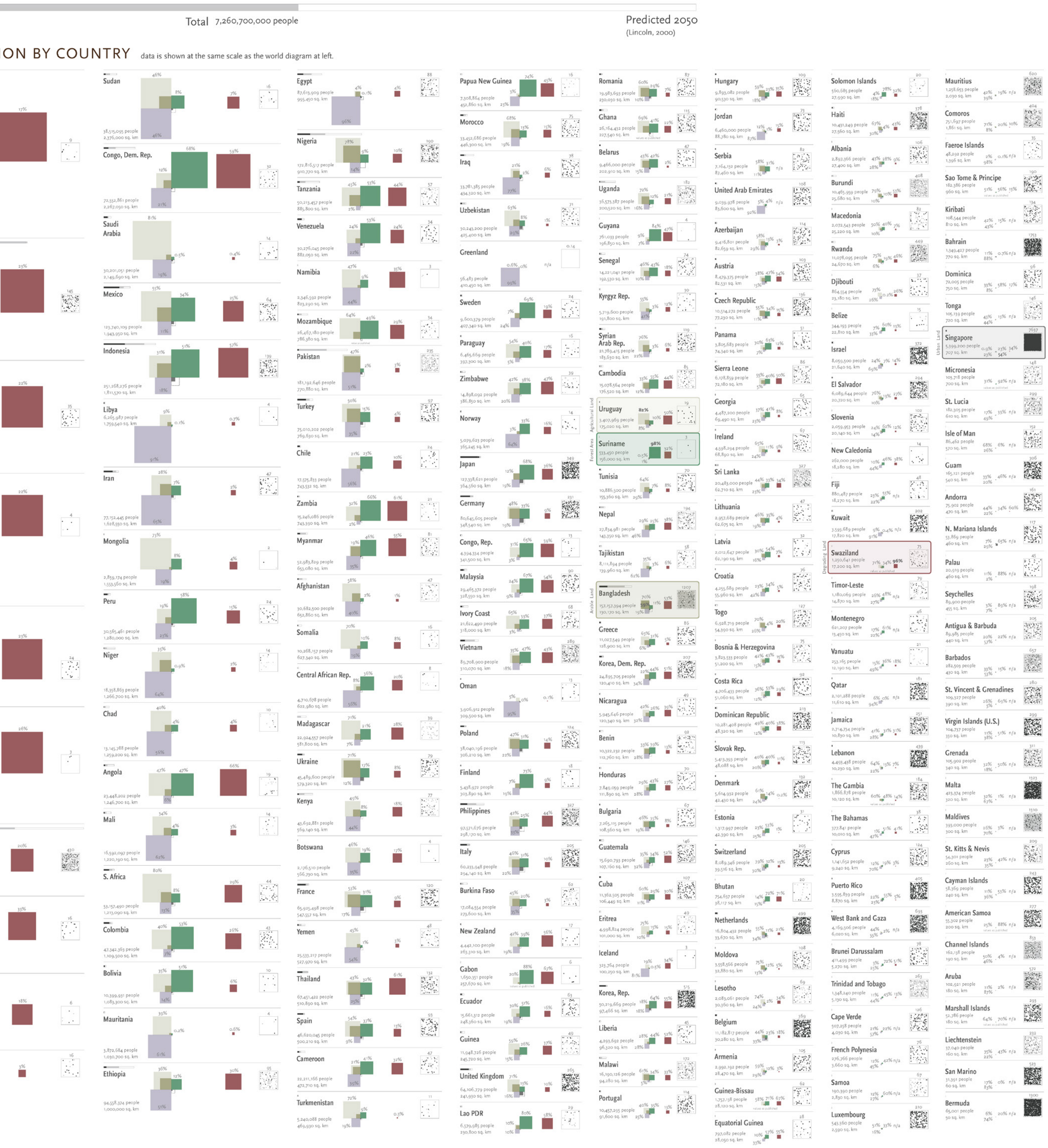


Figure 21. Final version of the land area graphics for the poster. This made up the top section of the poster.

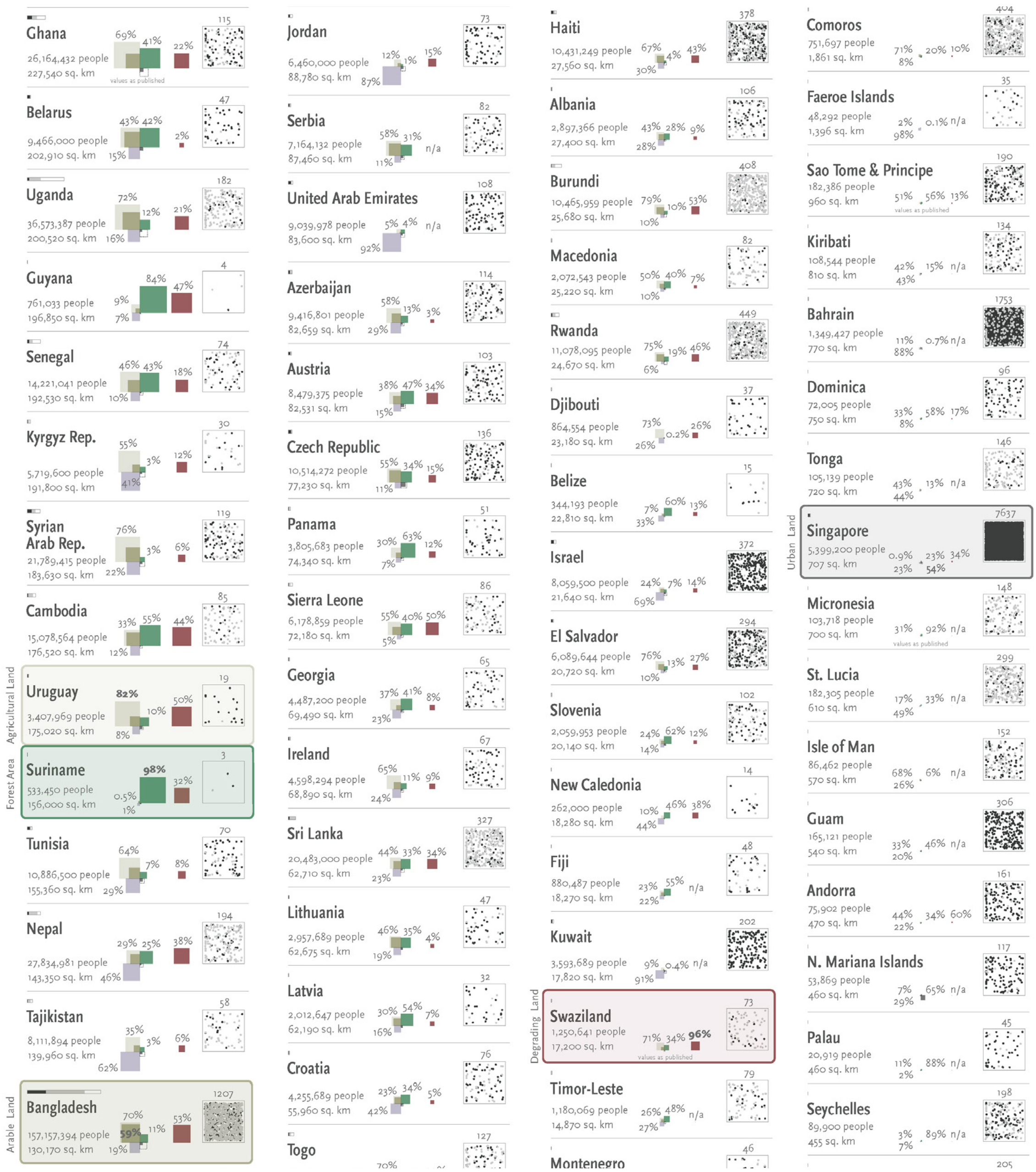
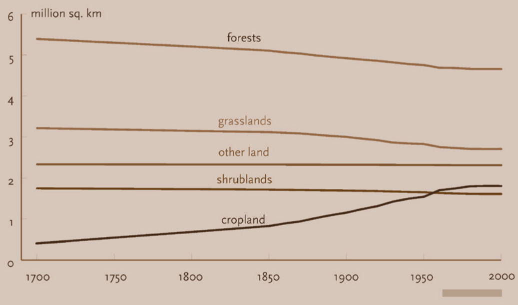


Figure 22. (above) Detail of the land use section of the poster, showing some of the smaller countries. Boxes highlight the countries with the largest percentage of a given land use area. (right) This section contained information on land use and food production changes over time, soil biodiversity, and climate change issues. It was originally shown as a single strip across the bottom of the poster.

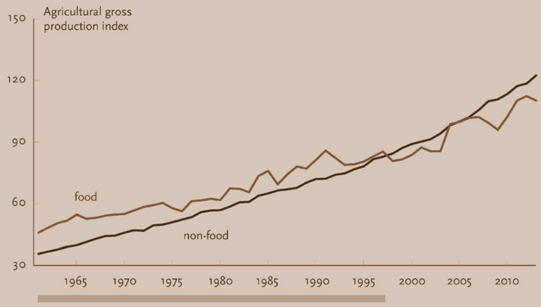
LAND USE

The question of how to achieve sustainable land use is one of the most pressing issues facing human society. Deforestation and a growing urban footprint are two primary issues associated with a burgeoning world population. Meeting human needs without depleting natural reources is an important challenge for our future. (Ramankutty, 1999)



FOOD PRODUCTION

Increasing food production to feed and clothe a growing world population puts additional strain on the world's soil resources. Agricultural production has tripled in the past 50 years, largely due to improved crop productivity per unit area of land. (FAO-UNESCO, 2016)



BIODIVERSITY

Roughly 360,000 (23%) of all known animal species live in the soil. 10 g of soil (about a teaspoon) contains 100 billion bacteria, in more than 10 million species.

ARTHROPODS

Soil arthropods include spiders, mites, millipedes, centipedes, and ants. The smaller species shred and eat plant matter. Larger arthropods often prey on smaller species, and other soil organisms.

INVERTEBRATES

Earthworms and nematodes are common soil invertebrates. Nematodes are microscopic worms that eat plant roots, bacteria and fungi, or each other. Earthworms are large worms that eat soil, playing a vital role in breaking plant matter down. Their tunnels also help to aerate the soil.

BACTERIA

Soil bacteria play a number of roles in the soil ecosystem. Most bacteria decompose plant material or minerals, making nutrients available for plants and other organisms. Some bacteria convert nitrogen into a form th is useful for plants. Others are pathogens and cause disease.

PROTOZOA

Protozoa are small, single-celled organisms, usually ten to a hundred times larger than bacteria. They feed primarily on soil bacteria, but can also eat fungi and soil organic matter as well. When protozoa feed on bacteria, they release nitrogen into the soil, which plants use to grow.

FUNGI

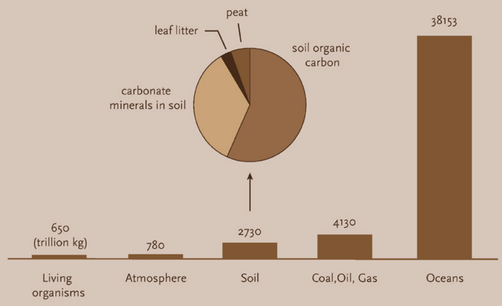
Soil fungi grow in long threads called hyphae, and produce the fruiting bodies that we know as mushrooms. Microscopic hyphae help maintain soil structure by holding small soil particles together. Fungae can decompose wood and other materials that bacteria cannot. They also help to extract phosphorus, nitrogen, and other nutrients from the soil.

PLANTS

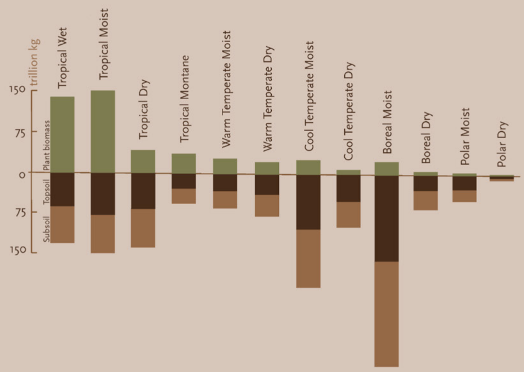
Plants constantly communicate with their environment by giving off chemical signals and hormones that encourage soil organisms to colonize their roots. Understanding the details of this interaction will help to improve plant and soil health, and increase food production.

CLIMATE CHANGE

The soil is an important storage area for terrestrial carbon resources, and offers a potential way to mitigate global climate change. Increasing soil organic matter will improve soil fertility and reduce the amount of carbon dioxide in the atmosphere. (Lal, 2008)



Soil organic carbon content depends both on the climate and the location measured in the soil. Tropical climates tend to have large quantities of living plant material above ground, as well as significant carbon stored below ground. Dry and temperate regions often have less plant growth, but may still have large carbon reserves in the soil. (Scharlemann, 2014)



Fertility of the soil is the
future of civilization.

Albert Howard

SECTION TWO

Continuing Experiments

After completing the soil poster in static form, I decided that I wanted to explore the possibilities of digital interactive media for representing multi-dimensional data of this kind. The poster format required that I present only a single snapshot of the data, and made it hard to provide additional detail and context on request. On-demand filtering would also help to address the suggestion to include only the top 5 or 10 countries from the list, and would give the user more control while exploring the data.

In addition to presenting the data that I had already collected, I also wanted to expand the information included to create a deeper view of the topic at hand. The first step was to include data from multiple years, so that users could observe changes over time. I also wanted to do a better job of connecting data between graphs, and to include more detailed storylines to discuss the evidence for the ecological importance of the soil.

Figure 23. Initial storyboard sketches organizing the final web implementation of the soil data

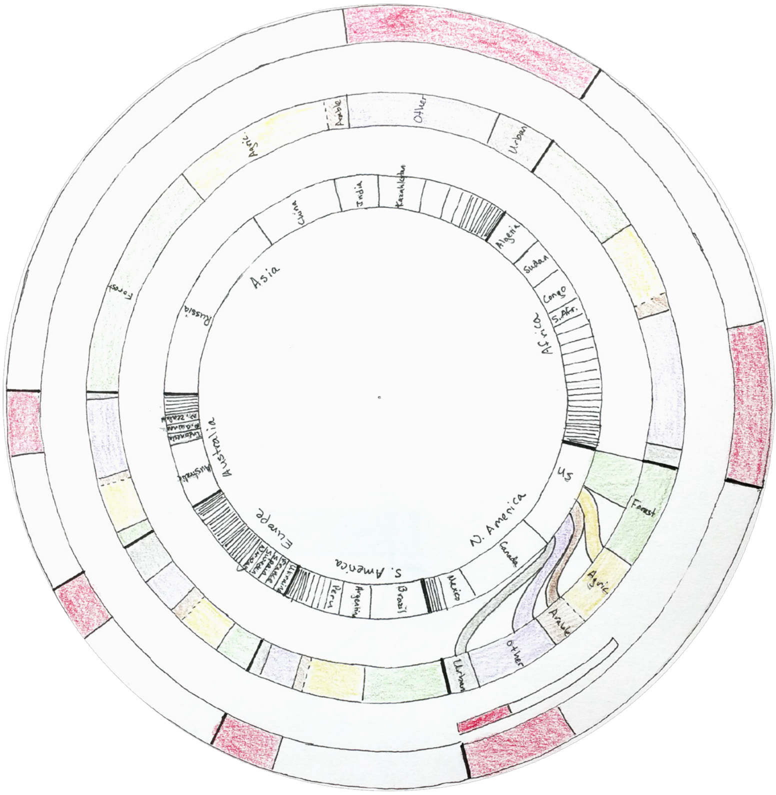
Early storyboarding

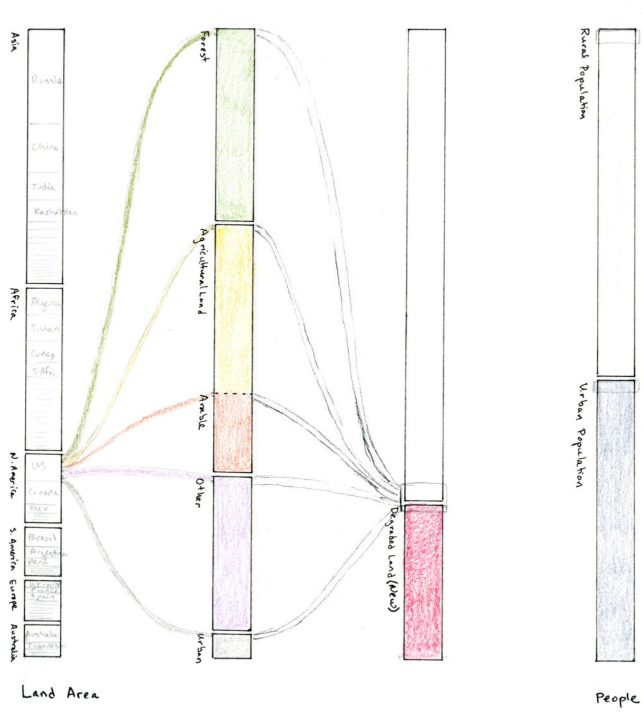
I began by sketching out a storyboard for a multi-page website, to guide my exploration of the design space. The narrative flow was very similar to the poster content, with a few high-level pages that discussed general information about the soil, and a more in-depth data exploration dashboard for exploring the World Bank dataset.

Land Use data: Reassessing design

Before diving fully into website construction, I wanted to step back and reconsider the design problem from the beginning, as a way to broaden my perspectives on the project and explore alternate directions. For this exploration, I wanted to push the edges of the design space a bit. I drew several sketches, and then finalized a couple for closerconsideration. Both of the final versions were derived from Sankey diagrams, and were intended to show how land was distributed within the different countries. The circular version required replication of data categories, so I rejected it and created an “unwrapped” version of the same information in linear form. I didn’t feel that the Sankey metaphor fit this particular

Figure 24. (opposite) Sketches of alternate representations of the land use data, showing land area, land use, and degraded land for each continent and country. Sankey flow ribbons connect the different variables for each country.


Asia
China
India
Kazakhstan
Russia
Africa
Algeria
Sudan
Congo
S.Afr.
N. America
US
Canada
Mexico
S. America
Brazil
Argentina
Peru
Europe
Australia
Forest
Agric
Arable
Other
Urban


Asia
Russia
China
India
Africa
Algeria
Sudan
Congo
S.Afr.
N. America
US
Canada
S. America
Brazil
Europe
Australia
Forest
Agricultural Land
Arable
Other
Urban
Degraded Land (New)
Rural Population
Urban Population
Land Area
People

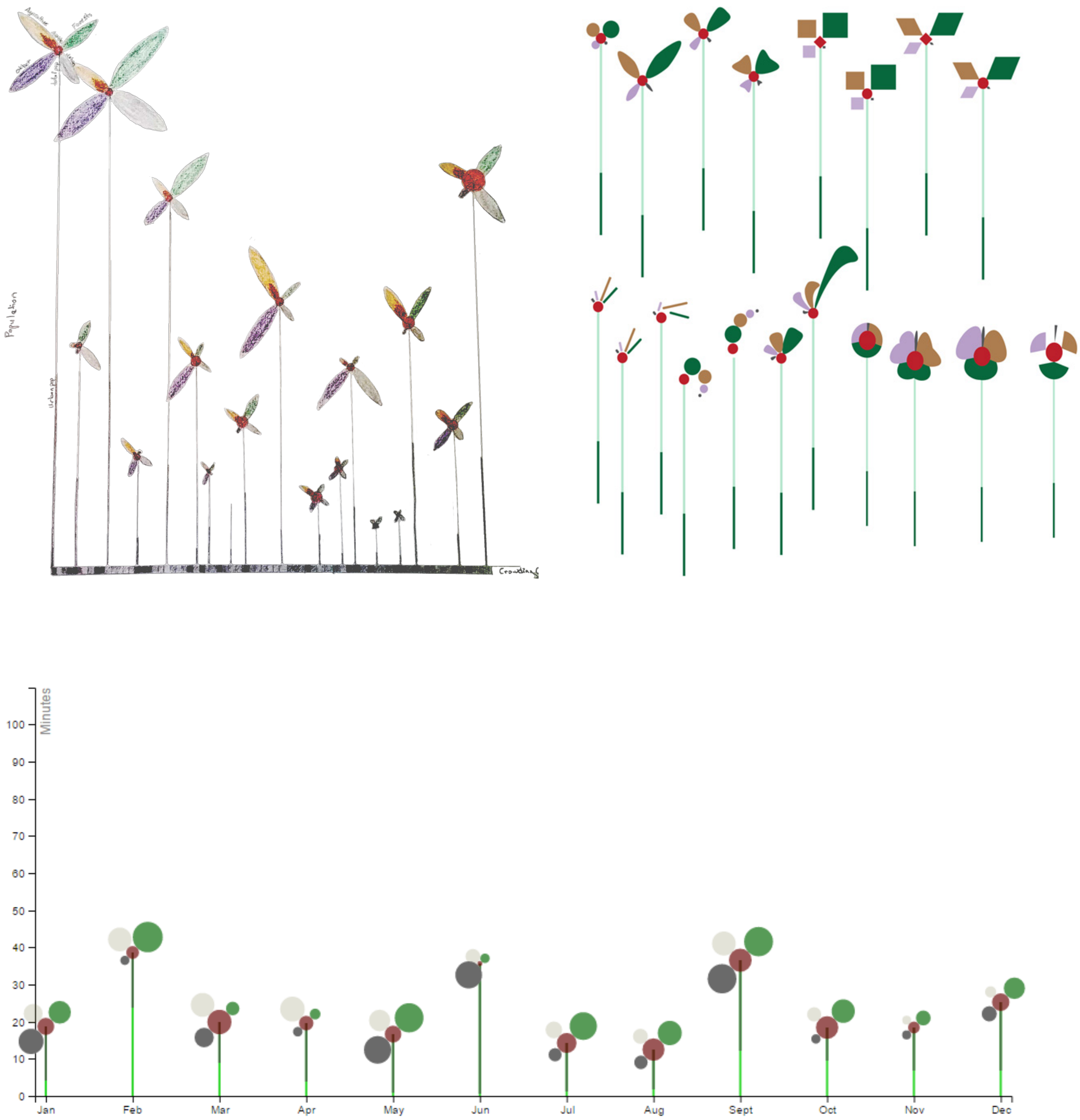


Figure 25. Sketches exploring the use of floral glyphs to represent land use data. Each petal represents a land use category, and the stem length reflects the population for a particular country.

dataset very well, since I'm more interested in static relationships than in flows. Its other disadvantage is that the smaller countries would be nearly impossible to select, since the linear scaling would make their widths so small.

In addition to considering alternate representations, I also experimented with making the original glyphs less literal. Inspired partly by Moritz Stefaner's OECD data, I considered using floral glyphs to emphasize the connection between land and soil data and the plants that we eat. I thought that these sketches were somewhat promising, so I tried representing the petals as other simple shapes as well. These sketches worked less well; I felt that they favored the metaphor over the quantitative values too much to be the main visualization for a project of this kind.

I did write a very simple version of the flowers in code, to see how they would work in an animated version. Since country populations generally grow over time, animating the flowers with data caused them to "grow" on the bar chart stems. I still felt that this was too metaphorical an approach, and the large differences in size between the petals made the flowers look rather lopsided.

Assembling longitudinal data

The flower animation also brought another issue to the forefront; while working on these sketches and considering the structure of the visualization that I wanted to create, I was also working on collecting data for all of the different metrics across multiple years. I assembled land use and population data from the World Bank Development Indicators database for all years available in the time period measured, between 1960 and 2015.

When I began examining this data more closely, it became clear that the datasets had very different levels of completeness. Forest data was not collected at all before 1990, and for some countries was not collected until much later. Urban data is only collected every 10 years. There is fairly good continuity for agricultural and arable land and for population statistics, but it was clear that I would not have access to all variables for all years. I mocked up a simple dashboard in d3 to begin

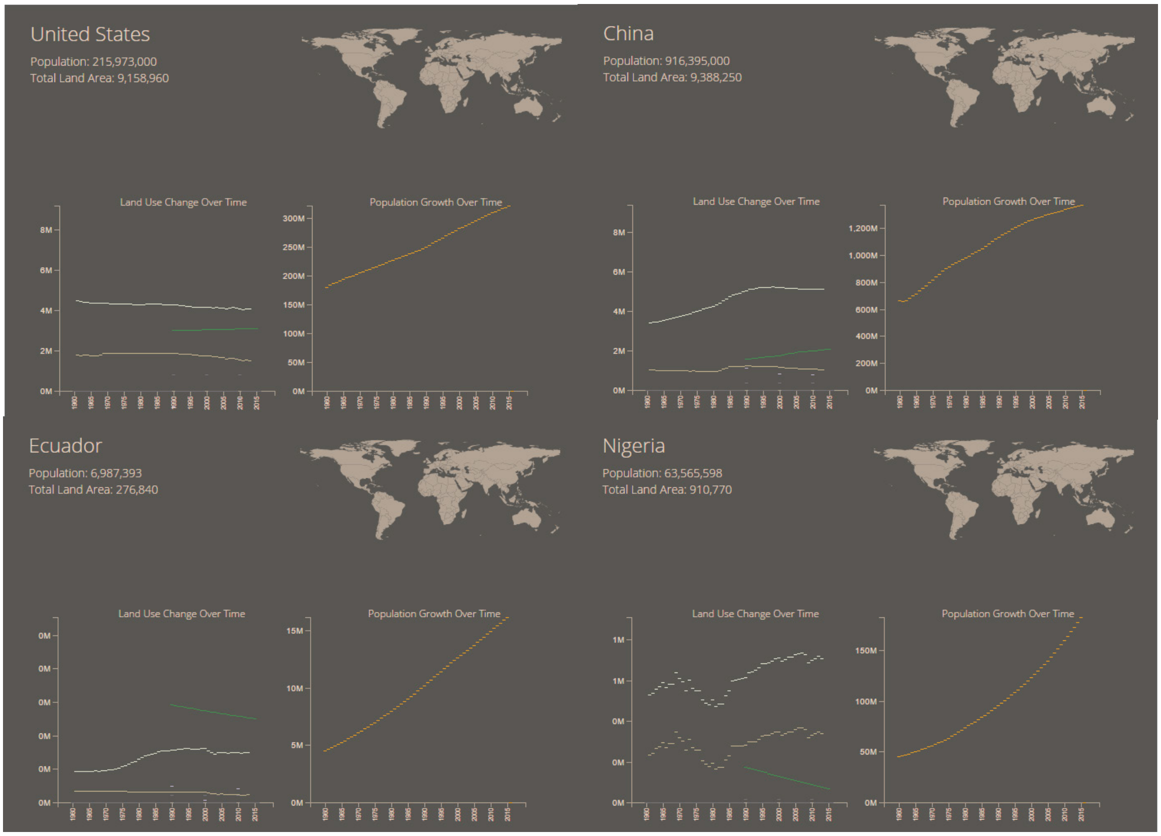


Figure 26. First draft of a visualization dashboard for visualizing data completeness and trends over time.

exploring the data and looking for trends. The initial version contained my original square glyphs, as shown in the storyboard sketches in figure 23, but the incompleteness of the data made them hard to read and compare, so I removed them in later versions.

Mapping options

I also began exploring different ways to represent the data geographically. The data dashboard used a map for country and continent browsing and selection, but didn't yet incorporate any data in the map itself. To start, I made a basic chloropleth (figure 27), just to see how the data would distribute. Though it was relatively simple to see which countries had the largest and smallest amount of agricultural land, this did not achieve the side-by-side comparison of the different values that I was hoping for. As I was most interested in looking at the relative quantities of the different land use categories, it didn't really work well.

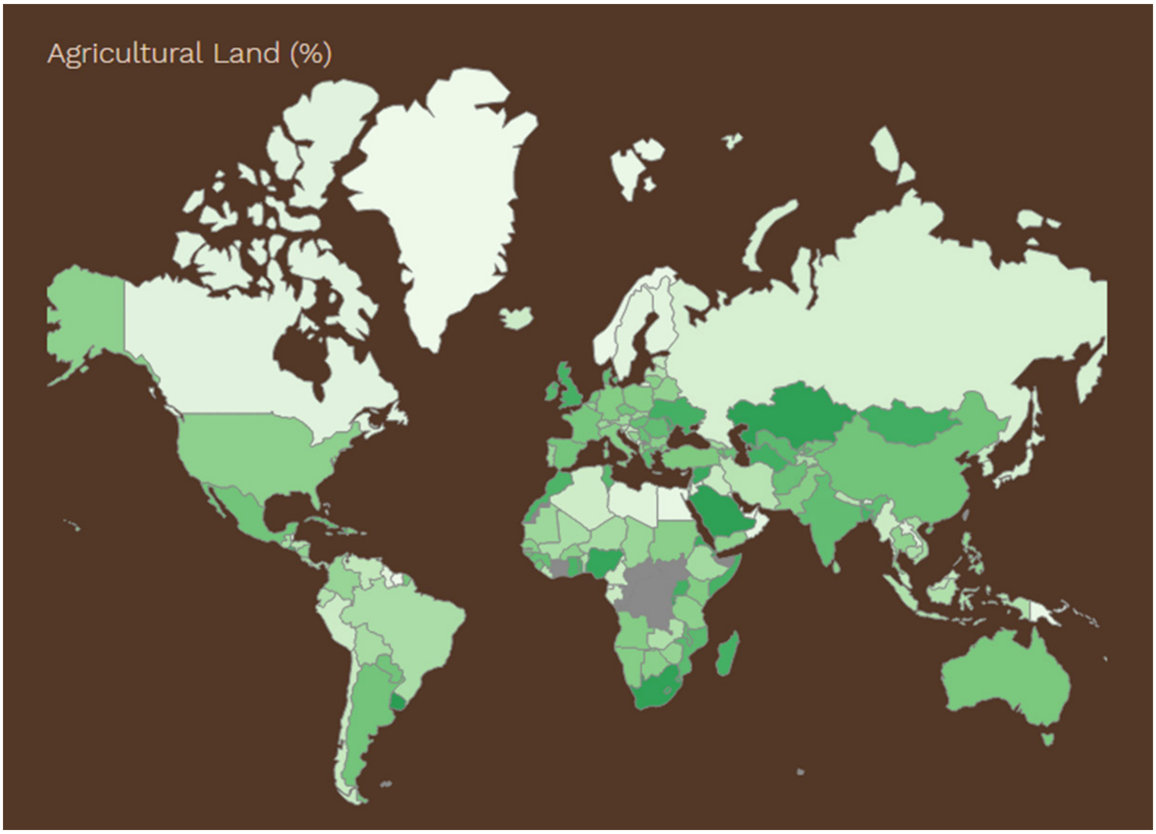


Figure 27. Early sketch exploring chloropleth mapping as a way to visualize land use variables.

I also implemented an animated area cartogram, where the area of a country is distorted according to the value of its associated data (countries with larger populations get larger than their actual size, etc.). The implementation was based on a heavily modified version of some sample code that I found online. I was unable to change the map projection due to the architecture of the code sample, so all of the cartograms were based on a Mercator projection.

Though it was possible to detect distortion in the animated version of the cartograms, the effect was subtle, particularly for the land area data. In the static versions shown in figure 33, it is difficult to discern any contrast at all for most countries. The population and population density data (bottom two panes) gave the most dramatic results, but I found it difficult to extract information about any particular country from this visualization, and it was similarly difficult to observe countries changing over time. Area cartograms rely on the user having a good sense of the

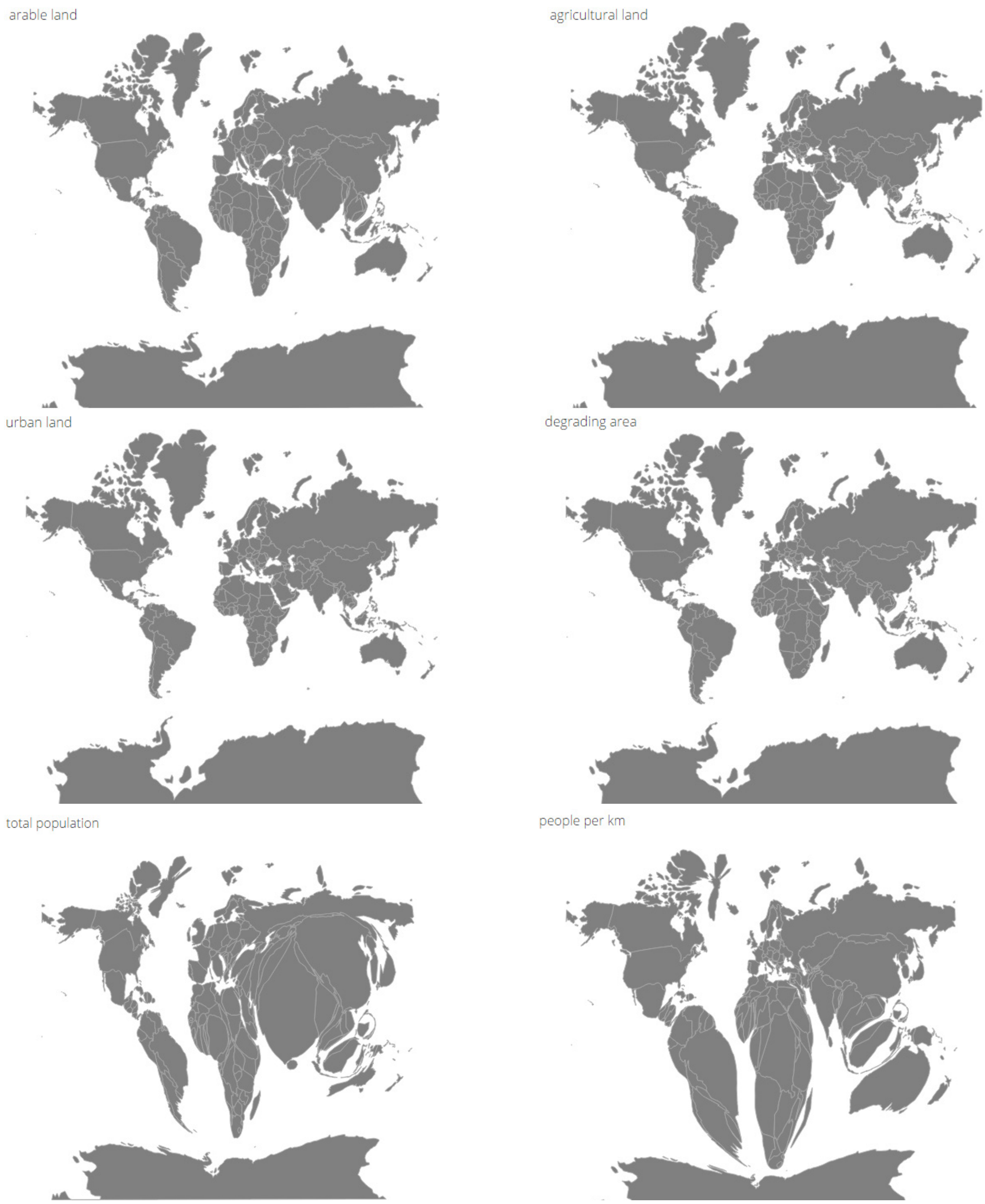


Figure 28. Area cartograms showing global land use and population data.

undistorted data, and I don't expect that my primary audience will be particularly well versed in geography, so I felt that the data would be better represented in another way.

At this point, I had explored several different options but hadn't yet found one that worked to my satisfaction. To gain some distance and allow the ideas time to develop, I chose to put the land use area aside for a while and focus on other things.

Visualizing food flow

In responding to the original land use poster, several people indicated that they would like to see only the top or bottom 10 for a particular metric, rather than seeing data for all of the countries in the world. To some extent, this disappears when the data is converted to digital form; it is not possible to show data for all countries at once on a small screen, so there will have to be a selection mechanism that allows the user to view just one or a few countries at a time. Ranking the data by value is also simple enough, and could easily be implemented.

I suspect that there's also a deeper issue beneath this request, one that I find problematic. Showing extremes is helpful for providing a fast overview and providing a sense of context; the biggest and smallest measures can be useful for creating anchors that help people to judge the severity or scope of an issue at a glance. But if I were to include only the biggest or smallest countries in the visualization, I would be implicitly adopting an exceptionalist perspective that suggests that the rest of the world is not important or does not exist. It also supports a tendency to look for the worst offender, making it easy to place blame and claim moral superiority rather than focusing on solutions that can improve a situation regardless of where a particular country falls in the rankings. In a project where I am attempting to emphasize interdependence, I feel that it is particularly important to expose the interplay between these different factors and acknowledge the contingencies of each case, rather than simply covering up the middle in favor of exposing the extremes.

I think that the impulse driving this feedback rests on a tendency to see the world in terms of national borders. This tendency is reflected in the structure of the data itself, and must therefore be a component of any visualization based on that information. However, emphasizing the importance of arbitrary national borders misses an important component of global systems: as with most ecological issues, the impacts of soil loss are not local, and cannot be contained within political boundaries drawn on a map.

It is also important in any discussion of this kind to realize that not all ecosystems are created equal. Soil ecology is critically dependent on geologic and climate factors that have nothing to do with who lives in a particular area or how they care for their land. Barring shifts to the global climate that affect its local weather, the Sahara will always be the Sahara; trying to increase soil carbon content in a desert is largely futile, because there is not enough moisture to support plant growth and prevent wind erosion. There are also human factors that need to be considered when looking at the soil as a global system. A facile interpretation might conclude that people in a particular country are simply bad stewards of their land, but this misses important contributions from the global economy. Understanding these factors provides an important piece of context needed to assess both the meaning and the implications of the global soil degradation map.

If we are interested in understanding where soil is most at risk due to human factors, we must also consider where the pressures on soil come from. Agriculture is a leading cause of soil degradation. In a global economy, food production is far from being a local event. For these reasons, I wanted to examine where our food comes from, and how products created from soil travel around the globe. When placed in relationship to the soil degradation map, I hope that this graphic will help to emphasize that soil degradation isn't just something that happens locally to people "over there," but that soil fertility loss anywhere affects all of us. Also, I hope to reflect the fact that the purchasing power of developed nations shapes the policy of emerging economies, for better or for worse. Relating this map to fair trade and the local food movement is probably outside the scope of this project, but I hope that this graphic helps viewers to realize the extent of their reach in a global economy and to consider the impacts of their choices.

I chose to focus on food import and export data, because food is the most direct soil product sold internationally, and because it is of natural interest to anyone who eats. The data is tabulated by the FAO-UNESCO as a detailed Trade Matrix that shows the origins and destinations of imports and exports for each country, by year. The Trade Matrix data is presented as individual categories for each kind of product traded, so that it is possible to look up how many tonnes of coffee beans were traded between Colombia and the United States in 2013, for example. The database also contains a Food Balance Sheet that gives the difference in total imports and exports for a given country, but does not identify individual trading partners.

I wanted this visualization to emphasize connections between countries and the motion of food products around the globe. For this reason, I chose to experiment with particle systems, with tiny dots moving around the map to represent food trade between countries. The data does not include information about individual shipments or the timing of trade within a particular year, so instead I related the particle emission rate to the overall trade value; a country with more exports emits more particles than a country with less active trade.

I began by creating a series of sketches to explore different options for the particle motion between countries (figure 29). In the first iteration, I gave the particles a fixed starting point, and then allowed them to spray out like water from a garden hose. This made it difficult to identify the destination country, and would make a very cluttered visualization if data were shown for multiple countries at one time. Emitting a single stream of particles helped to simplify the visualization. I considered animating the particles with differences in size or fill color or size, but found that this complicated the graphic without adding new information.

I also began experimenting with Bezier curves to trace the particle path and guide the eye. These curves could be used alone, in combination with particles moving along them, or in multiples to show the intensity of imports/exports. With these options in hand, I then began applying these different techniques to real data (in the last two panes of figure 29).

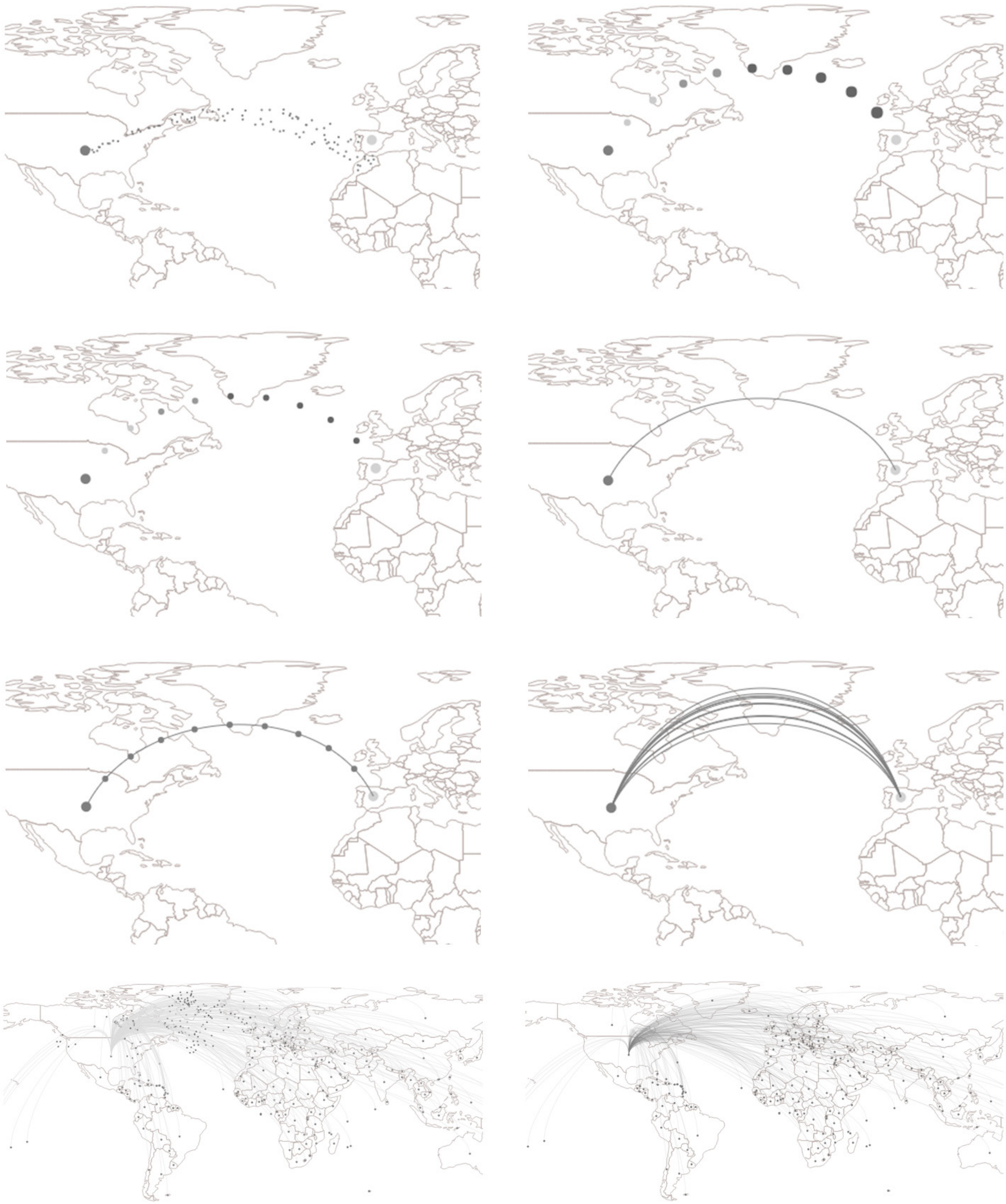


Figure 29. Development of a particle motion for a visualization of the global food network.

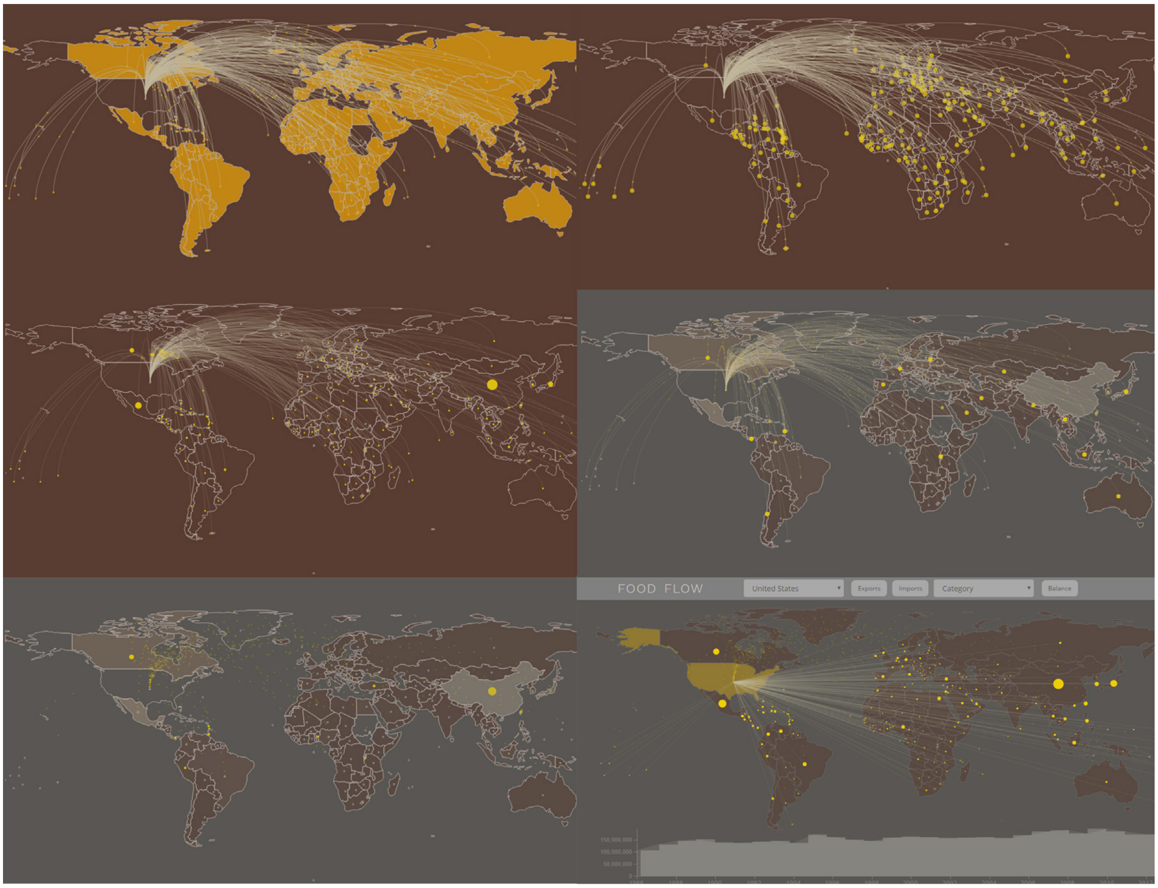


Figure 30: Experiments in designing the integrated visualization of global food flow

Next, I experimented with showing the import/export quantities and highlighting destination countries in different ways. First, I tried changing the fill color of all import/export partners for the selected country. For countries that are heavily connected, like the US, this highlights almost the entire map, leaving the occasional odd hole in the map, which draws more attention than the colored countries because it looks like a mistake.

As an alternative, I tried drawing dots at the centroid for each destination country. This was much more effective, as it draws attention to the partner countries rather than the map. I also tried taking advantage of the presence of these spots to reinforce the trade volume between countries. Initially, I programmed the spots to blink on and off as each particle reached the end of the Bezier curve, which resulted in a map that pulsed in time with the import/export values. Most people found the pulsing distracting, and it made it harder to see the sizes of the country spots, so I reverted to a static view. I also tried highlighting major trade partners using chloropleth shading on the background map, but I felt that that was redundant and that the spots alone did a better job.

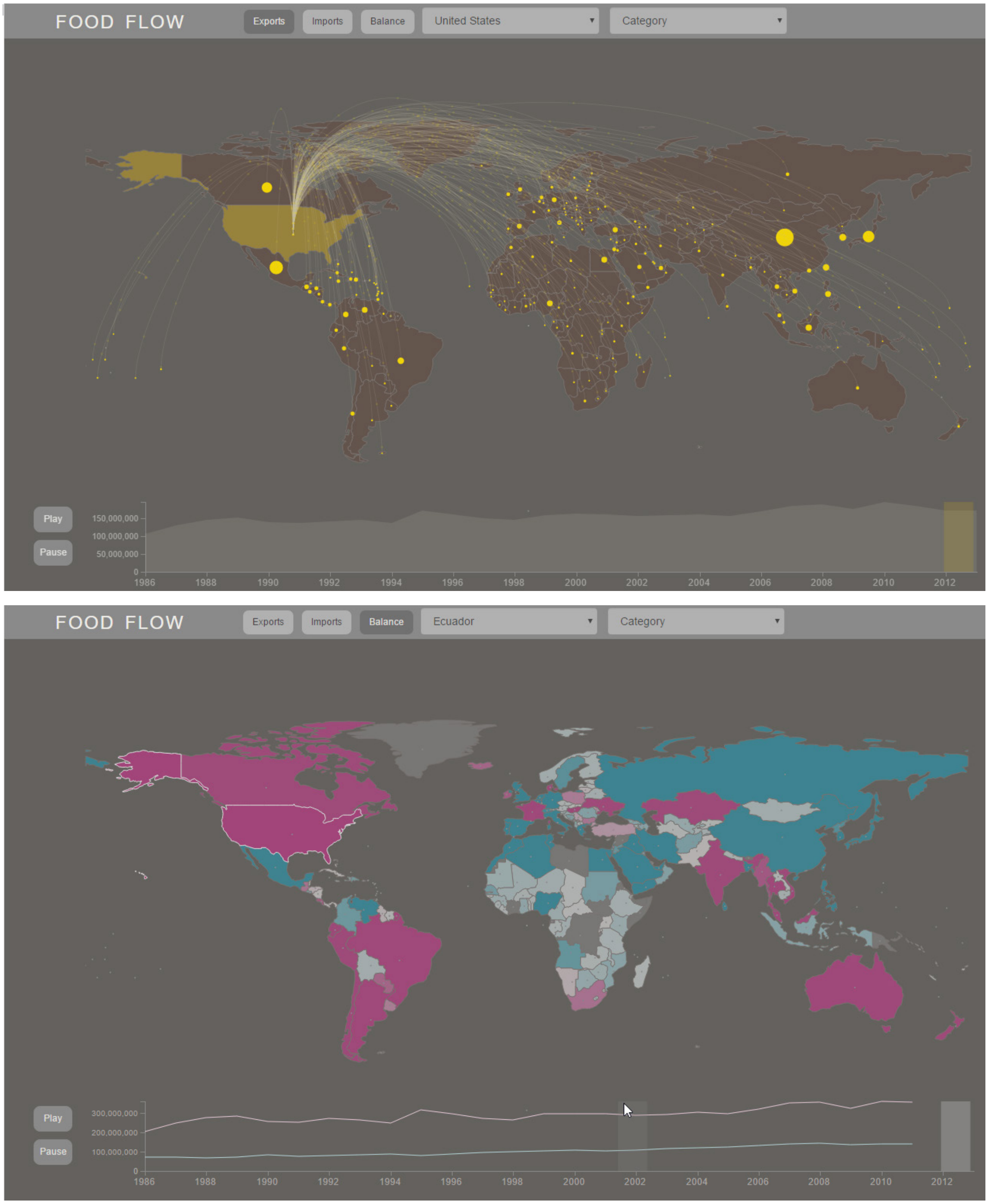


Figure 31. Late-stage version of the food network visualization, showing food exports from the United States, and import-export balance data

At this point, I decided to step back a bit and check whether the Beziers were actually necessary. Although I reduced the opacity to make them less prominent, they do still partially occlude the background map, especially when a country exports to many of its closest neighbors. Without the curves to highlight particle paths, it was difficult to follow the particle motion at all, and the map became much less visually interesting. The Beziers also help to highlight the connectivity within the graphic, which is one of the primary features I hoped to convey. Finally, I tried using straight lines to connect the different cities, to see if that might help with issues of occlusion while retaining connectivity. I did not feel that this approach was nearly as appealing or as successful as the Bezier version.

The final version of this visualization for the Studio 3 class used Bezier curves and particles to animate the motion of food from one country to its partners. Users can select the country that they want to view using the dropdown menu, or by clicking on the map. The user can also toggle back and forth between import and export data, and the particle motion changes direction as a result. Each origin/destination partner is marked with a yellow spot whose size reflects the quantity of food traded between the selected and partner nation. This makes it easy to see at a glance that China, Mexico, and Canada are primary food trade partners for the US, for example. The timeline across the bottom of the screen shows the historical values for all years available in the dataset. Hovering over a section in the timeline allows the user to load data for a specific year, to allow them to explore changes in trade relations over time. Consistency of scaling for the circles across all countries allows the user to get a rough qualitative sense of the relative size of trade relationships, but does not provide a quantitative view, especially for countries represented by the smaller dots. Food category data was also collected and cleaned, but category filtering was not implemented in this version of the visualization.

Finally, I included a second map in the visualization to show the net balance of imports and exports for different countries over time. Each country is color coded on a diverging scale to indicate net positive (magenta) or negative (teal) balance for a given year. The timeline shows the curves for both imports and exports for the selected country, to show trends over time. As the user highlights different sections of the timeline, the map colors update to show the world food balance for that year.

This visualization was implemented in Javascript, using the d3.js library to draw the map and destination markers, and to manage the data. A transparent canvas was overlaid on the d3 SVG, and was used to draw the Bezier curves and moving particles, which would otherwise make the browser lag. Mouse events from the canvas are linked to d3 using a force layout, to allow the user to select SVG items through the Canvas using mouse events. The background data filtering was done using Crossfilter.js, and the page formatting was done using Bootstrap.js.

Linking ecoregions datasets

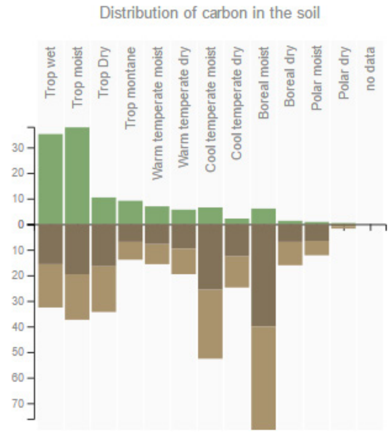
In addition to these experiments using large statistical datasets, I also wanted to do a better job of creating linked graphics for the smaller scientific data as well. Although the datasets are simpler, they provide an important reference point for discussing the land use and trade data. In the original poster, I presented a bar chart comparing the amount of carbon in living plants with the amount in the soil for different kinds of ecoregion around the globe. This data is particularly interesting to me, because the distribution is not what one might expect.

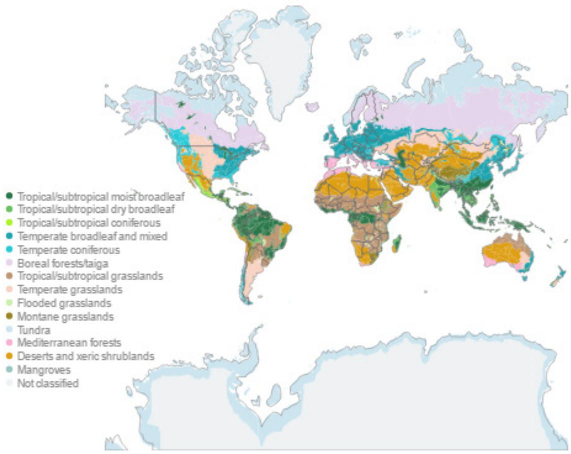
Tropical rainforests do have the most aboveground carbon, consistent with lush plant growth in a warm climate. However, it is the northern boreal forests that have the largest amount of stored soil carbon. It is interesting to realize that it is actually the cooler regions of the globe that lead in this regard.

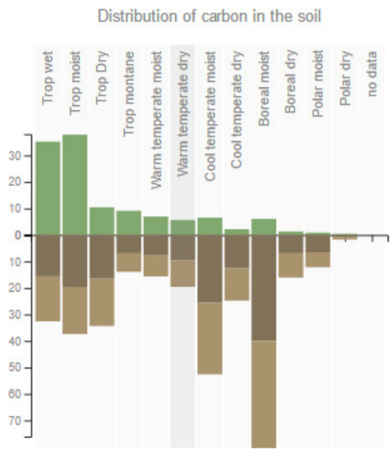
Figure 32. Multi-pane visualization linking soil carbon storage with world eco-regions and forest area change. Carbon stored in living plant material (green) is compared with that stored in the soil (brown) for different ecoregions of the world. The bar chart on the far right compares forest areas for countries within a selected region to highlight active deforestation.

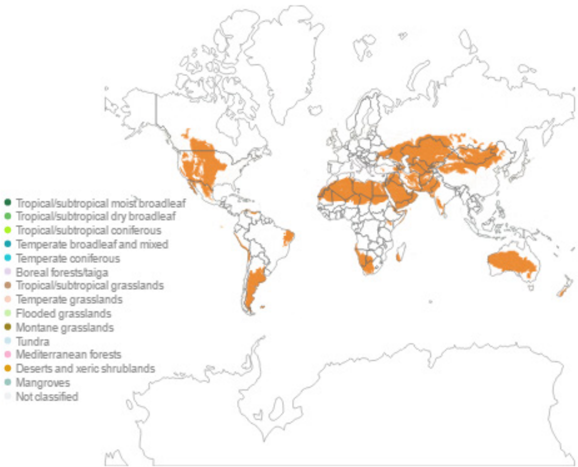
Besides a vague notion of where broad climate categories are found, I do not know very much about where specific climates are located, and so I assume that my viewers will not know this, either. I wanted to create a way for them to understand which regions of the globe each bar in the bar chart relates to, to create a better understanding of the geographic distribution of soil carbon.

Unfortunately, this was not trivial. Like soil, ecoregions are defined in different ways by different groups, and the definitions do not match one to one. The scientific paper that presented the soil carbon data states

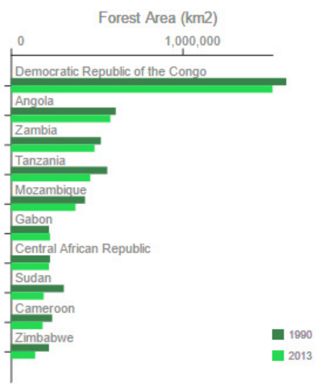

Distribution of carbon in the soil
Trop wet
Trop moist
Trop Dry
Trop montane
Warm temperate moist
Warm temperate dry
Cool temperate moist
Cool temperate dry
Boreal moist
Boreal dry
Polar moist
Polar dry
no data
30
20
10
0
10
20
30
40
50
60
70


Tropical/subtropical moist broadleaf
Tropical/subtropical dry broadleaf
Tropical/subtropical coniferous
Temperate broadleaf and mixed
Temperate coniferous
Boreal forests/taiga
Tropical/subtropical grasslands
Temperate grasslands
Flooded grasslands
Montane grasslands
Tundra
Mediterranean forests
Deserts and xeric shrublands
Mangroves
Not classified


Distribution of carbon in the soil
Trop wet
Trop moist
Trop Dry
Trop montane
Warm temperate moist
Warm temperate dry
Cool temperate moist
Cool temperate dry
Boreal moist
Boreal dry
Polar moist
Polar dry
no data
30
20
10
0
10
20
30
40
50
60
70


Tropical/subtropical moist broadleaf
Tropical/subtropical dry broadleaf
Tropical/subtropical coniferous
Temperate broadleaf and mixed
Temperate coniferous
Boreal forests/taiga
Tropical/subtropical grasslands
Temperate grasslands
Flooded grasslands
Montane grasslands
Tundra
Mediterranean forests
Deserts and xeric shrublands
Mangroves
Not classified


Tropical/subtropical moist broadleaf
Tropical/subtropical dry broadleaf
Tropical/subtropical coniferous
Temperate broadleaf and mixed
Temperate coniferous
Boreal forests/taiga
Tropical/subtropical grasslands
Temperate grasslands
Flooded grasslands
Montane grasslands
Tundra
Mediterranean forests
Deserts and xeric shrublands
Mangroves
Not classified


Forest Area (km2)
0
1,000,000
Democratic Republic of the Congo
Angola
Zambia
Tanzania
Mozambique
Gabon
Central African Republic
Sudan
Cameroon
Zimbabwe
1990
2013

that they use the IPCC climate region definitions, but gave no specific conversion between those definitions and the labels in the chart. The names for the labels seemed to match several other nomenclatures better than the IPCC standards, so it's possible that there is an error in the original paper. The paper references data in the Harmonized World Soil Database, which was named similarly to the FAO ecofloristic zones, so I used these values to map onto the ecoregions listed by the World Wildlife Federation (WWF). I downloaded map data from the WWF website, and manually merged the two categorization systems.

The map data was exported as .shp files, and converted to geoJSON using mapShaper, a free online map conversion tool. All paths were simplified to reduce the number of points that needed to be projected by the browser in order to draw the map. The map outlines come from the Natural Earth database, and are drawn as SVG shapes on top of an HTML Canvas element, to allow d3 mouse selection and brushing behaviors.

All 15 WWF ecoregions were shown on the map. When a user hovered over a bar in the bar chart, the map would update to show only the ecoregions associated with that bar, allowing the user to see where on the globe those ecoregions are located. If the user selected a region on the map, then the application displayed a list of the top 10 countries in that region, and showed their forest land area for both 1990 and 2013. This allows the user to search for trends in the amount of forest area change within a particular region during that time.

I used this first draft to fulfill a homework assignment for my CS class, and chose to link the different graphs in a single, linked view partly because of the specifics of that assignment. I had also hoped to create a direct bridge between the soil carbon storage and deforestation.

In practice, though, I think the dual action of the central map—showing ecoregion coverage in one mode, and selecting country deforestation data in the other—is confusing. The ecoregions do not fit neatly within countries, and each country usually contains more than one ecosystem, so it is difficult to make any quantitative comparison from one end of the graphic to another. Each half of the graphic served a useful purpose, but I concluded from this experiment that they would be better off split in two for future versions.

Narrative animation

When searching for examples to include in the Related Work section, I was struck by how many visualizations made effective use of documentary-style narrative visualization in some way. Whether used as an introduction to orient the user, or used as the primary technique in the piece, many successful visualizations are based on cinematic techniques that incorporate music or animation. As discussed in the Related Work section, narrative can engage an audience and improve both the understanding and memorability of a graphic.

When my Sketching with Code class was asked to program a line or bar plot in d3, it seemed like a good opportunity to recreate one of the simple visualizations from my original poster, and to use transitions, music, and animation to increase interest in this simple bar chart, and to direct attention to the information it contains.

For me, realizing that soil stores more than half as much carbon as all of the global coal, oil, and gas reserves was a big surprise. It was also a turning point in shaping how I understood the relevance to soil health as a method of combating climate change. In the original poster, this surprising revelation was pushed to the final visualization in the bottom right corner, and didn't have the prominence that I felt it deserved. It's a simple piece of data, and doesn't require a complex visualization. A simple bar chart does quite well for the purpose, but it's easy for viewers to miss the important comparison because bar charts are so ubiquitous. For this reason, this visualization seemed like a particularly good candidate for animation.

The bar and pie charts were constructed in d3 using sequences of transitions and easing to draw one bar at a time, accompanied by text that interprets the meaning of each piece of data incorporated. The animated text is presented one sentence at a time, just before the corresponding data is drawn. It reads: "Carbon. It's a huge player in climate change. Most of the world's carbon is stored in the oceans. By comparison, a lot less is in the atmosphere. Coal, oil, and gas reserves contain about 5x as much carbon. All life on earth uses just a fraction of that amount. After fossil fuels, soil stores the most carbon on land. Several different forms of carbon are present in healthy soil. Increasing

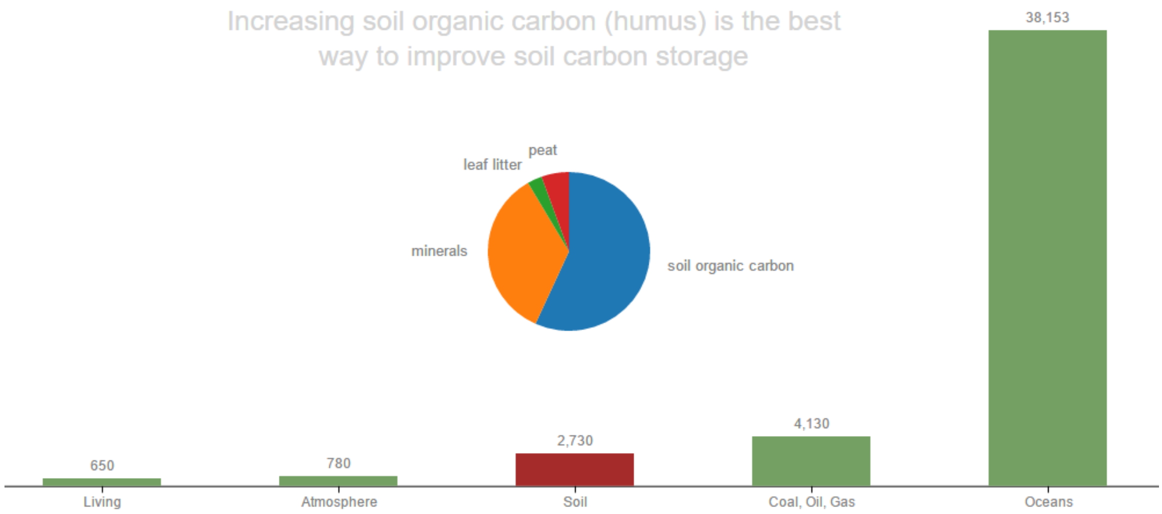


Figure 33. Animated visualization explaining the significance of soil for global carbon storage

soil organic carbon (humus) is the best way to improve soil carbon storage." The narrative steps the viewer through the data, interprets significance, and provides context for each piece of data shown. It also helps to connect the pie chart to the soil bar, and to link the organic carbon content of the soil to its storage capacity.

When developing the visualization without music, I frequently found myself annoyed with the delays caused by the animation, and impatient to rush the transitions between states. It was clear that it would be difficult to balance the speed of the animation against the time that people need to read the text and watch the animation, especially when viewing the animation for the first time.

Adding the background music made a dramatic difference in how I perceived the animation. The 500 ms transition time that had felt agonizingly slow now felt too fast, because it didn't match the tempo of the music. In the final version, the text dwell time on the screen was 2000 ms, with 3000 ms transitions for the fade in time. The entire visualization takes just over a minute, which felt impossibly long without the music, and almost too fast with it.

The music also added an emotional component to the visualization that was not there before. I used the Epilogue piece from Doug Hammer's album Travels: a slow, deliberate piano score with a lot of emotion. Pandora lists the following attributes for the track, courtesy of the Music Genome Project: new age aesthetics, an overall meditative sound, a prominent rhythm piano part, intricate melodic phrasing, composed melodic style, acoustic sonority, major key tonality, instrumental arrangement. This set the tone of the visualization and emphasized the subjective importance that I wanted to convey.

Overall, I think that this was a successful experiment. The animated visualization helped to improve the impact and emotional significance of an otherwise unremarkable graphic. The narration provides important context and places emphasis on the meaning of the data shown, rather than the presentation. Adding a voice-over component could also help to control pacing and make up for differences in reading speed.

This animation could make a successful introduction to the project or to a section of a website. I would not use this presentation style for every visualization in a series, because the enhanced emotional impact came at the expense of time. I felt that the additional time was justified in this case, because it was so easy to miss the important feature of the bar chart in a cursory examination.

If you change the way you
look at things, the things you
look at change.

Wayne Dwyer

SECTION THREE

Consolidating Design

With these new experiments completed and the theoretical thesis framework beginning to solidify, the end of the fall semester was a good time to step back and reassess progress, as well as identifying strategies for completing the rest of the thesis project. This final phase is largely one of consolidation and reinforcement, cleaning up narratives and visualizations, filling in the missing pieces, and checking the final outputs against the guidelines described in the previous chapter. I began by formally defining the thesis audience, which had been clearly identified in my own mind from the beginning but had never been stated in writing. Then I reconsidered the tone and structure of the piece, outlining the narrative approach I intended to use, as well as its advantages and drawbacks. With that decision in place, I began testing out narrative threads to connect the different pieces of the visualization into a coherent whole, and then returned to the code to implement these changes and update the individual designs.

Defining the audience

This thesis is aimed at presenting science to the general public. At its most basic level, the public could be defined as those who do not actively practice science. Within that category, there are many sub-divisions based on education, socio-economic status, and ideology, among others. I want this project to be accessible to as broad an audience as possible, across all of these divides. For the most part, I expect that viewers will be English-speaking North Americans, and the majority will likely come from the US.

I will outline a few of the important features that I imagine my audience may have below. Any such attempt necessarily relies on generalizations and stereotyping. I do not mean to imply that every member of my audience will have all or even some of these traits, or that they can be universally applied to the different categories of people that I discuss. The descriptions simply list some of the features that I have had in the back of my mind while developing the project website, and are usually based on specific individuals that I know who fit into these different categories. I have attempted to leave room within the descriptions for variations within groups, but have almost certainly left some people out, and characterized others too broadly.

Core audiences

Practically speaking, the audience for any interactive website is likely to be composed of young people who are technologically-savvy, college-educated, and relatively wealthy. They likely live in cities, and may never have thought about or engaged with soil or agriculture in a meaningful way. They likely come to the website through blog or forum posts, or through their social media page. The challenge with this group is to convince them that soil is worth paying attention to. It will be necessary to get past negative associations of soil as "dirt" and undesirable, squeamishness about bacteria and soil organisms that they likely find "gross" rather than fascinating, and to create an appreciation for the role of these organisms in the broader ecosystem.

On the positive side, many members of this audience love to be amazed, and are strongly drawn to narratives of wonder and awe; they want to be captivated by an important story. They are also drawn to technological and aesthetically pleasing forms, and like to think of themselves as data savvy (which they may or may not be). They also tend to have many distractions and short attention spans, and need to be rewarded with frequent insights and takeaways in order to maintain engagement. This audience is also very active on social media.

Visualizations that are eye-catching, interesting, and sharable enhance their social cachet, particularly if the representation incorporates their personal views in some way.

The next group is a bit older, and is defined by already-established interest more than age. This audience is composed of hobby gardeners and people attracted to topics related to sustainability. They may or may not be college-educated, but they are intellectually curious. They likely live in suburban areas and may have access to a yard or community garden, though they may also live in cities or rural areas, and may believe that they have a black thumb. They are likely very concerned about the environment and the effects of pollution on ecological and personal health. This interest may also extend to a curiosity about where food comes from, or even to a strong emotional need to control the source of their food. Some may view personal purchasing choices as a political or identity statement, purchasing only organic produce, only buying from certain stores, or through supporting local farmer's markets, and they may belong to social groups where these choices are considered evidence of belonging. This audience is probably less dependent on technology, and will tend to come to the website through established webpages or news feeds that they read, email forwards, personal recommendations, and possibly social media feeds.

The older members of this audience may not be comfortable with data or cutting-edge web technologies, and may be intimidated by a complicated interface or highly abstract visualization with unfamiliar forms. This sub-group is interested in the topic and curious to learn, but needs signifiers to show them how to interact and wants to see a reasoned analysis of what things mean. They are more motivated by learning something interesting than by quotable quotes or sharable personal art, and they want to connect this information to their own experience and beliefs. Once engaged, this audience will want to explore all of the details, and some will be excited to pose and answer their own questions.

Another sub-group within this domain contains people who may have scientific or technical training, but in a different field (this audience would include myself). They are fairly data-literate, and want to see a thorough, clear argument in order to be convinced. Fast facts and superficial summaries are not interesting to this group, who prefer to dig deep into the information and explore on their own. They likely come to the site through blogs or social media feeds, and are mainly interested in technical insights and statistical information, as well as links to resources and further information. They may also have pre-existing knowledge of the microbiome, soil quality measures, and many of the other topics covered in the introductory section of the site, which may feel too superficial to them. More detail and deeper discussion are necessary to engage their interest. Where possible, they also want to see how this topic relates to their own specialized domain, and may have additional professional expertise or insights to contribute.

Aspirational audiences

In addition to these relatively "easy" targets, I believe that this topic also has the ability to appeal to people with less formal education who live in rural areas. Many in this audience don't need to be convinced of the importance of soil, as farming is part of their daily lived experience; either personally or within their community. Some may not know much about microorganisms and the soil ecosystem, but others will be highly versed in understanding the soil community, and may even be leaders in creating new agricultural techniques to preserve soil health. They are innovators, pragmatic experimentalists, and experts in the economic realities of agriculture. This audience viscerally understands the threat of soil degradation, but may not see how to get around it. They also may not have thought about the truly global implications of soil, or may feel that their work as its caretakers is underappreciated.

The challenge for this audience lies in helping them identify opportunities to respond in positive ways to the problem of declining soil health, and helping them to connect with resources and support. They understand (and may be personally experiencing) the severe economic tradeoffs of producers in a commodity market, and may be very resistant to ecological preservation as an impractical ideal in direct contradiction to their bottom line — and possibly their personal economic survival. This audience may feel that they are being unduly blamed by any discussion that emphasizes the detrimental effects of modern agriculture, and may feel that it is impossible or detrimental to change their techniques. They may also be deeply suspicious of scientific advice, focusing on recommendations that have let them down in the past, or that seem to contradict their direct experience. They could also be highly attuned to new technical advances, and eager to experiment.

This audience is also deeply attached to their way of life, and has a strong emotional commitment to their role as stewards of the land. They have experienced the consequences of soil degradation more directly than anyone else, and they are the ones with the most power to make direct change, as well as the deep knowledge required to enact it. They may be hesitant to believe that their market is committed enough to support the change that needs to happen. This audience needs to see that there is a way forward, and they are actively looking for new techniques and technologies to solve this problem, without hurting their economic prospects. They also need to believe that others are interested in and committed to preserving the soil, and to feel respected as producers of the food that supports our society. For them, this project is less about educating, and more about creating a community where they can feel respected enough to share both their challenges and their expertise. This discussion may also involve questioning default assumptions and unfounded beliefs, but that would be better handled by soil and agricultural scientists with the expertise and authority to evaluate those claims, and is outside the purview of the project website itself.

Tone

Within the first few weeks of investigating soil topics, it became clear that I needed to make intentional choices about the overall tone and framing for my project. It is common practice in treatment of environmental issues to focus on the threats, because getting right to the heart of the issue demands attention and often creates a strong emotional response. As I discussed in the previous chapter, this approach can also lead to overwhelm and disengagement, and is often counterproductive. Rather than jumping directly into shocking statistics about soil degradation and global food supply, I wanted to approach the

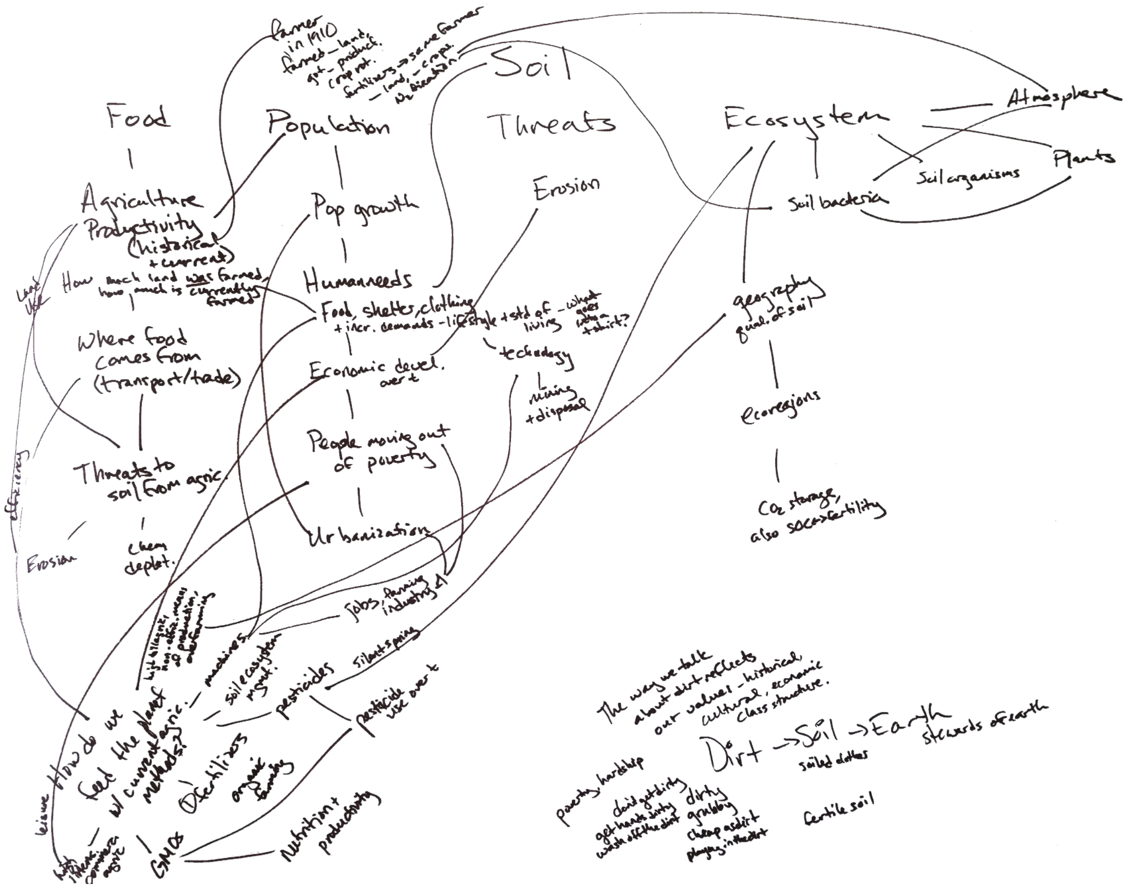


Figure 34. Concept map showing relationships between potential soil narratives.

topic with a sense of wonder. Helping people to see and understand the value of the soil and the scale of its influence can create a sense of desire and even ethical responsibility to protect this valuable resource. Recognizing the extent to which every one of us depends on soil for our very survival can also frame ecological concerns in terms of self-interest.

Framing the issue in terms of threats is more direct and focuses immediately on the core relevance of the topic. The wonder-first approach carries with it the risk that people will not be interested in learning about soil, and will not ever progress far enough into the project to understand its importance. Mitigating this possible disadvantage by providing frequent insights and rewards that encourage the reader to continue exploring is a critical component of success.

Structure

At this point in the project, I had identified many interesting storylines within the topic area, and needed to stop and choose the primary narratives to guide the development of the final thesis site. My first mind maps of the relevant topics reflect the complicated network of connections and relationships to choose from (figure 34).

At the end of the fall semester, this project was best described as a collection; a loose grouping of related topics based on a set of classification guidelines. My earliest narrative sketches showed the project as a network: connected, but with no particular directionality. An alternative structure would be a traditional narrative, which imposes hierarchy and a direction on the story (figure 35).

Developing my storylines was really a process of shaping this collection to line up with narrative lines. Ideally, I wanted the final result to reflect a narrated network; a hybrid between a traditional narrative and a looser node/network form. In this model, a user should be able to access each piece of the visualization as a modular whole, or to experience it as part of a continuous narrative. I wanted my visualization interface to support both modes; the user should be able to select a node directly to see what it contains, or choose a narrative to follow if they want more guidance. To get a better sense of the options, I wrote out all of the main topics on index cards, and arranged them in different ways.

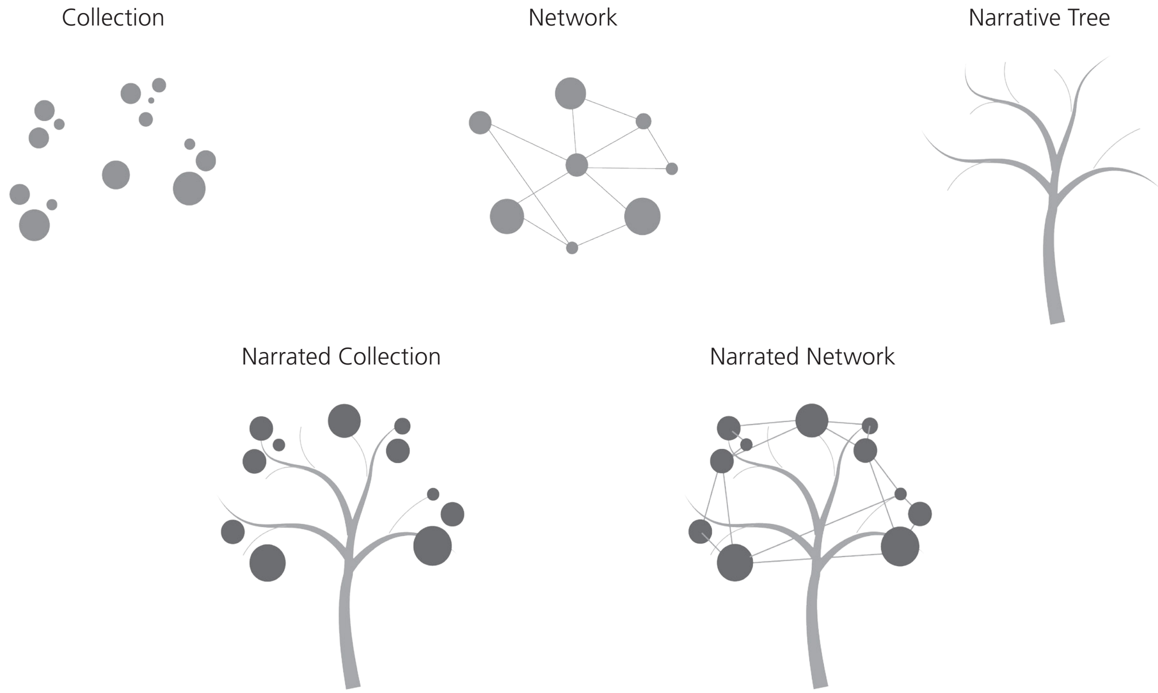
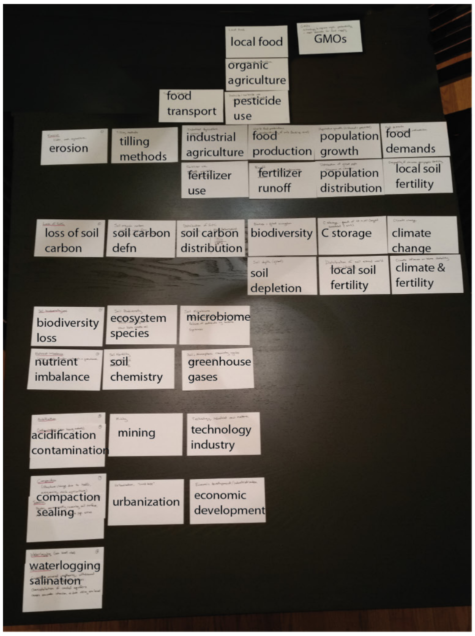


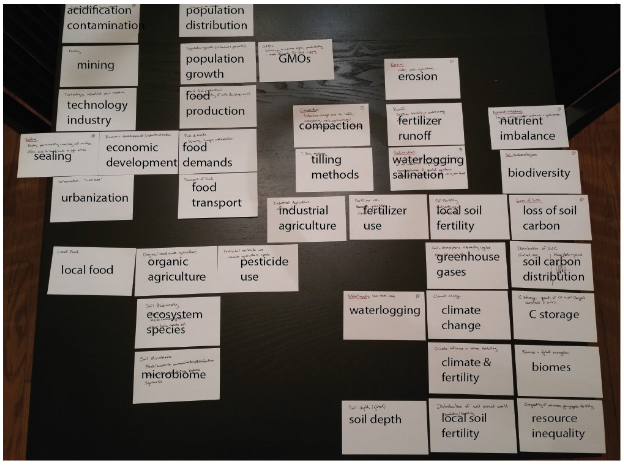
Figure 35. Different ways of organizing information, progressing from simple grouping to hierarchical ordering, and then showing hybrids of the two.

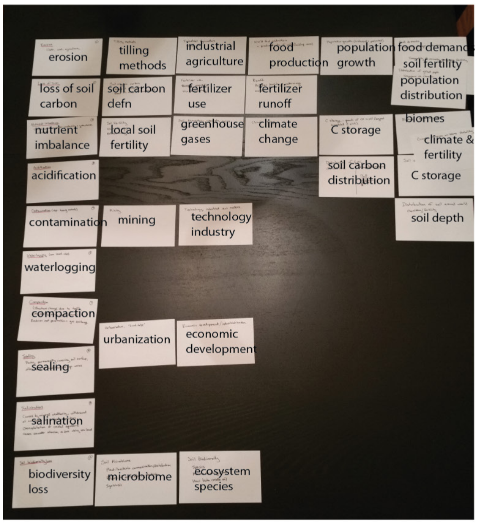
In analyzing the story lines created in this way, I identified several possible ways to frame the project. I could begin by presenting threats to the soil, and then focus on issues related to these. I could focus on agriculture and food production, and examine how the soil supports world food needs, accentuating its importance to human civilization. Focusing on economic development, technology, and industrialization would give more of a historical perspective, and emphasize the environmental costs of human actions. Or, I could lead with the ecological importance of soil, emphasizing its biodiversity.

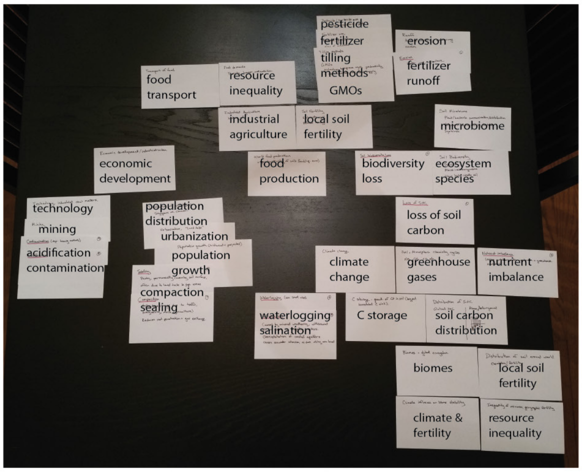
Within each of these topics, I began to distinguish between datasets and pieces of information that were central to the story and those that were more peripheral. I was also able to identify some sub-stories that worked as bridges to connect the different parts. Drawing out all of these different components helped to show where the different topics were clustered, and where the balance of my research lay.

I presented a simplified overview of these topics in my thesis proposal presentation in December (figure 37): it highlights the main topics that I thought were important, and some of the connections between them and the more peripheral subtopics. The items highlighted in red are current threats to the soil, as identified by the FAO-UNESCO.









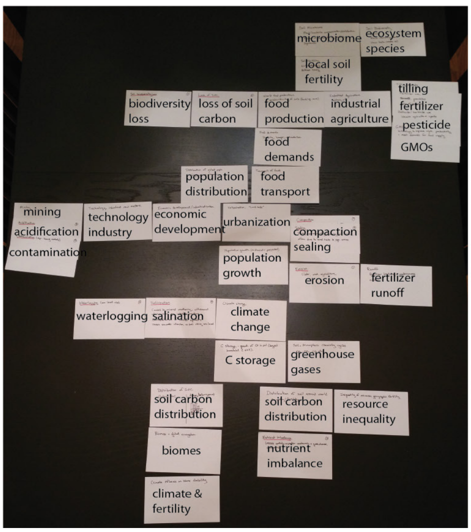


Figure 36. Arranging index cards to develop narrative flow and identify alternatives. Organizing by threats to the soil created parallel narratives; connecting by topic created a less linear flow. Agriculture, ecosystems and carbon storage were particularly information-dense areas, and were often connected by population and urbanization.

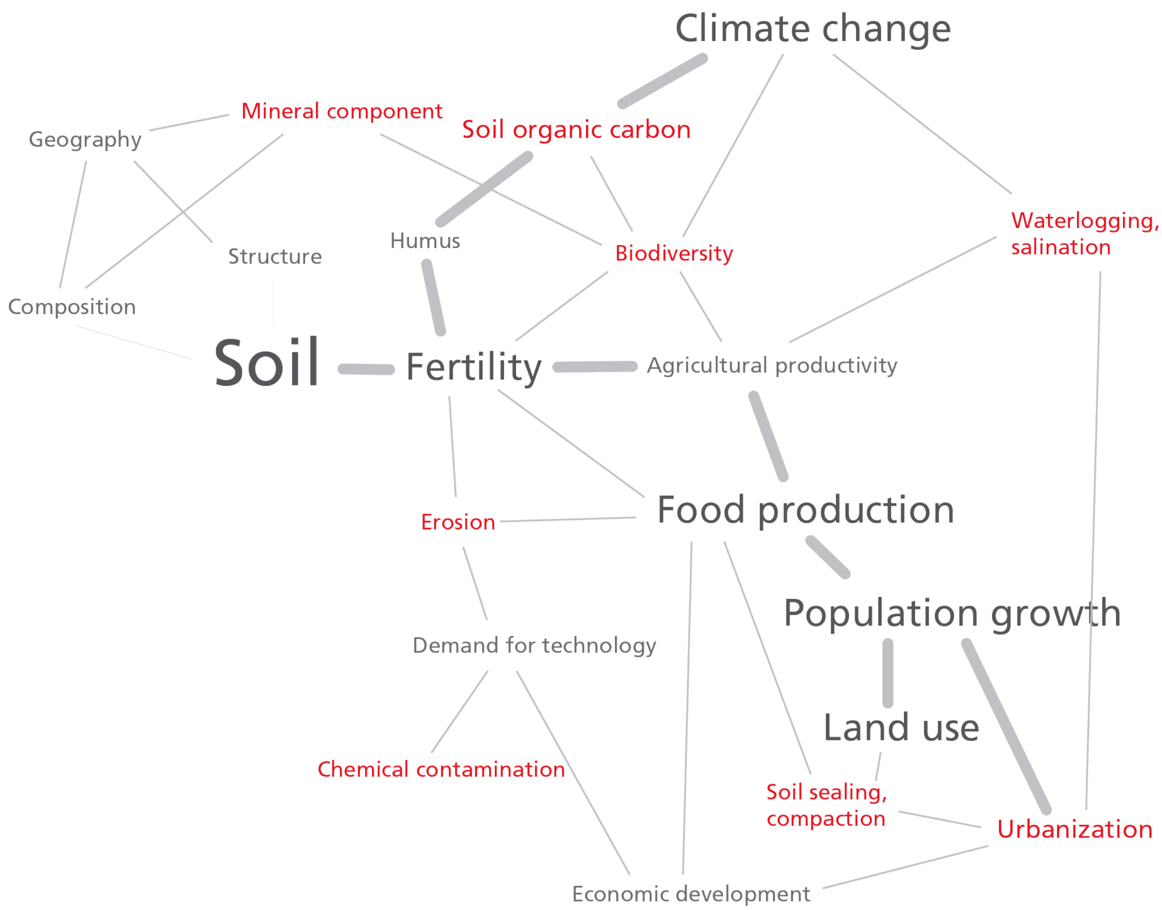


Figure 37. Early concept map extracted from the cards in figure 36, showing relationship between soil topics and UNESCO threats.

From this undirected network of ideas, I began to consider how a person might move between these different topics in a sequence, forcing the network structure into a more linear form (figure 38). I also began color-coding the different narratives, and identifying the areas where I had the most information readily available.

From these sketches, I extracted a set of common nodes and identified three different narratives that could connect them in different ways: soil, population, and food production. From there, I moved over into Illustrator to create a loosely-organized collection of nodes. Once I had a preliminary organization, I printed out several copies and highlighted the different paths that a user could take to move

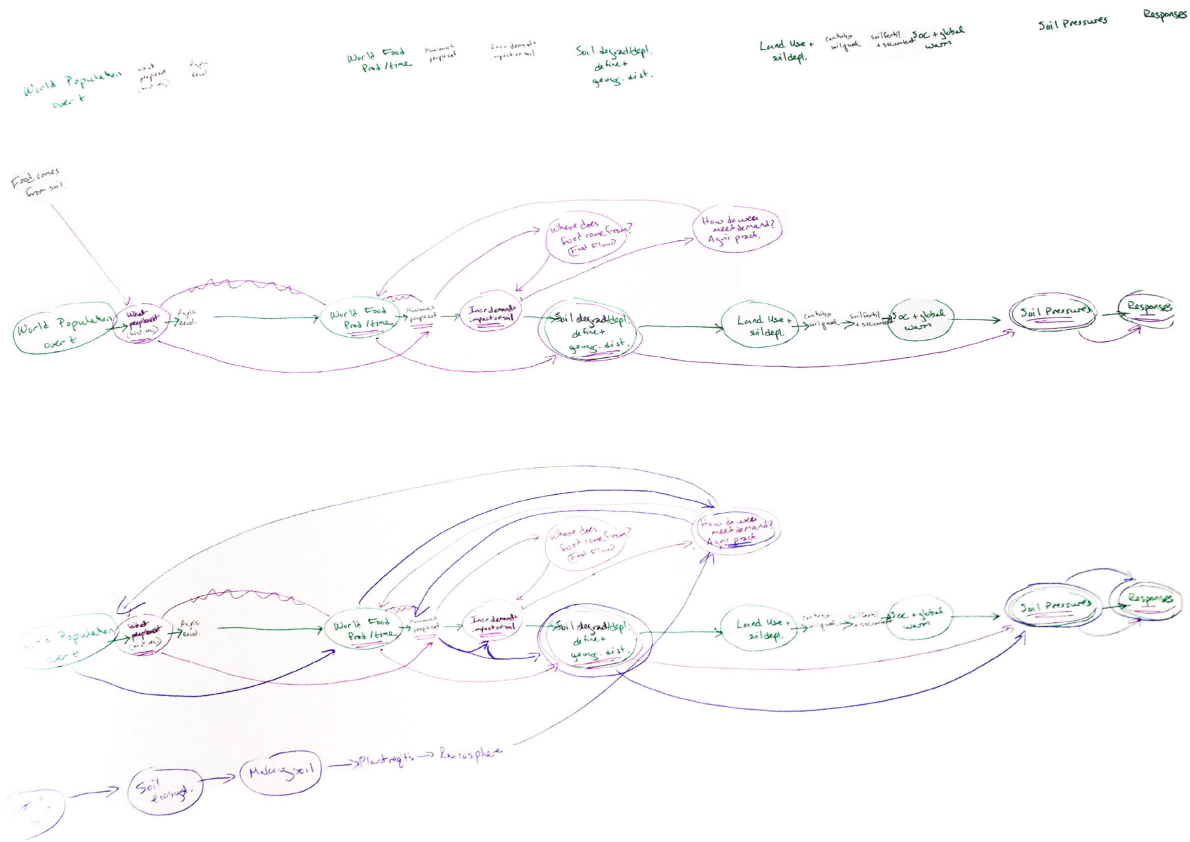


Figure 38. Reducing concept map to a set of common base nodes, connected by alternate narrative options.

between them (figure 39). The primary topics for each narrative were highlighted, to help create a hierarchy to support simplification. From these rudimentary drawings, it was possible to identify different emphases for the three narratives.

Agency

In a functioning narrated network, it should be possible for a user to move through the visualization in any order, or to choose a narrative to guide them through the data. The navigation interface for the website plays a critical role in supporting this function and granting the user agency within the visualization. The navigation interface must provide

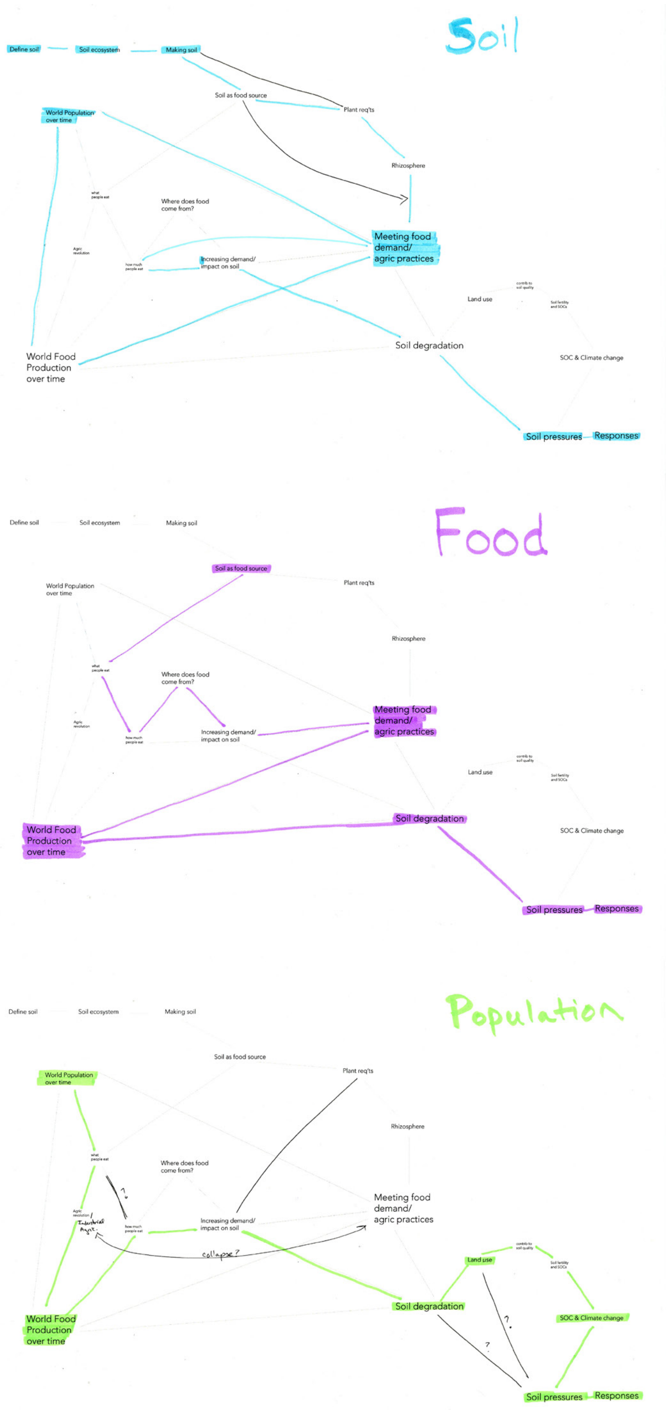


Figure 39. Three main narratives that connect the same nodes in different ways.

enough structure to orient the user and support exploration, but without forcing a particular series of choices. I made my visualizations modular in order to provide maximum flexibility in reading order; a self-contained visualization can be approached from any angle. I also decided to create a separate project title page that orients the user and show how the modular visualizations are connected, as a way of introducing the user to the system, and allowing them to browse the data available.

Designing the interface:

The earliest drafts of the user interface were developed mainly as a way for me to switch between different sites during development, and to set up the code that would switch between them. They also functioned as a way for me to test out different interaction paradigms as the site grew. From the rough layouts show in figure 39, I drew a combined node network in Illustrator. I consolidated some of the nodes, simplified others, and drew connections for each of the different narratives. Then, I began to untangle the different pieces step by step, trying to find the natural flow of the site. This initial sketch is shown in figure 40. The node size reflected the relative importance for a particular narrative, the stroke colors identified the main narratives that each node belonged to, and fill colors were used to highlight stories for which I had data or draft visualizations available.

At this point in the code development, I also needed a method for browsing the individual visualizations In order to keep some link between the different project representations. To minimize the amount of context-switching between the different sketches, I built a force layout in d3 with fixed node positions, calculated based on those in my original Illustrator document. This also allowed me to work out the node and link file structures needed to generate the visualization, and to test features like adding image and svg drawings inside d3 nodes from an arbitrary file.

Even ignoring the aesthetic limitations of this sketch, it is difficult to know where to start in an undirected network of this type. Adding narrative lines helped to connect the different nodes, but did not provide a strong sense of direction or sequence in this visual representation. To create a stronger sense of flow, I began rearranging the node positions in Illustrator to impose a loose sense of directionality, minimize line crossings and reduce unnecessary topic interconnections.

Riffing off of the tree-network relationship in figure 35, I started playing around with the possibility of using roots to connect the node network and provide a sense of directionality. I wanted to create an animated drawing that would begin on page load and add an extra layer of aesthetic interest to the piece, but I knew that this would be time-consuming to execute, and wanted to test whether the idea would work first. I began by tracing an image of tree roots in Illustrator, to see if I could create a quick mockup of the idea. The resulting drawing (figure 41) was far too busy to be of any use, and overwhelmed both the nodes and the narrative lines connecting them. It also wasn't possible to match the node positions to the photograph in a simple and responsive way, so the node coordinates and the image features were not aligned to one another, making it difficult to tell whether they were related.

Next, I played around a bit with recursive algorithms in p5.js, but found that calculating a fully ramified root structure was far too computationally intensive to be done in the browser, even when drawing to an HTML canvas rather than SVG. I tried converting the generative algorithms to Processing thinking that I might be able to export a points array to load into Javascript, but it was clear after some initial experimentation that it would require a lot of refinement to get aesthetically pleasing results with the desired effect. As a temporary measure, I settled on using an SVG file drawn in Illustrator for the roots, because it could be imported simply, scaled to match different screen sizes, and with care could be directly related to my node positions. After drawing up a concept sketch in Illustrator, I roughed up a quick sketch in code to see how it would work. I also began integrating screen shots of the different visualizations into the nodes on the front page, and added a title to give users a preview of what they might see. These replaced the image and description tooltips shown in the first root sketch. The orange icons are filled with a random png from my computer, which I used to make it very obvious which nodes were still

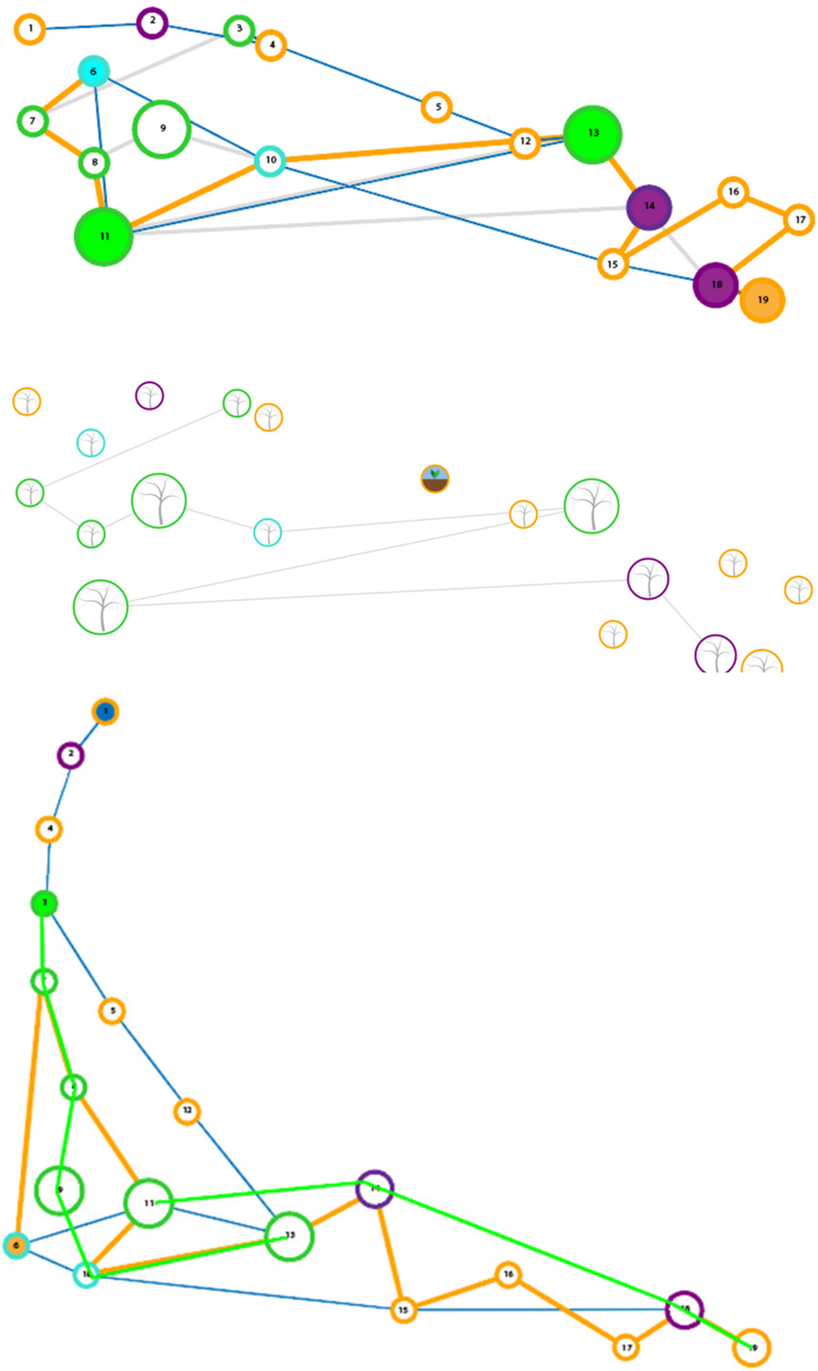


Figure 40. Developing complementary node networks in Illustrator and d3.

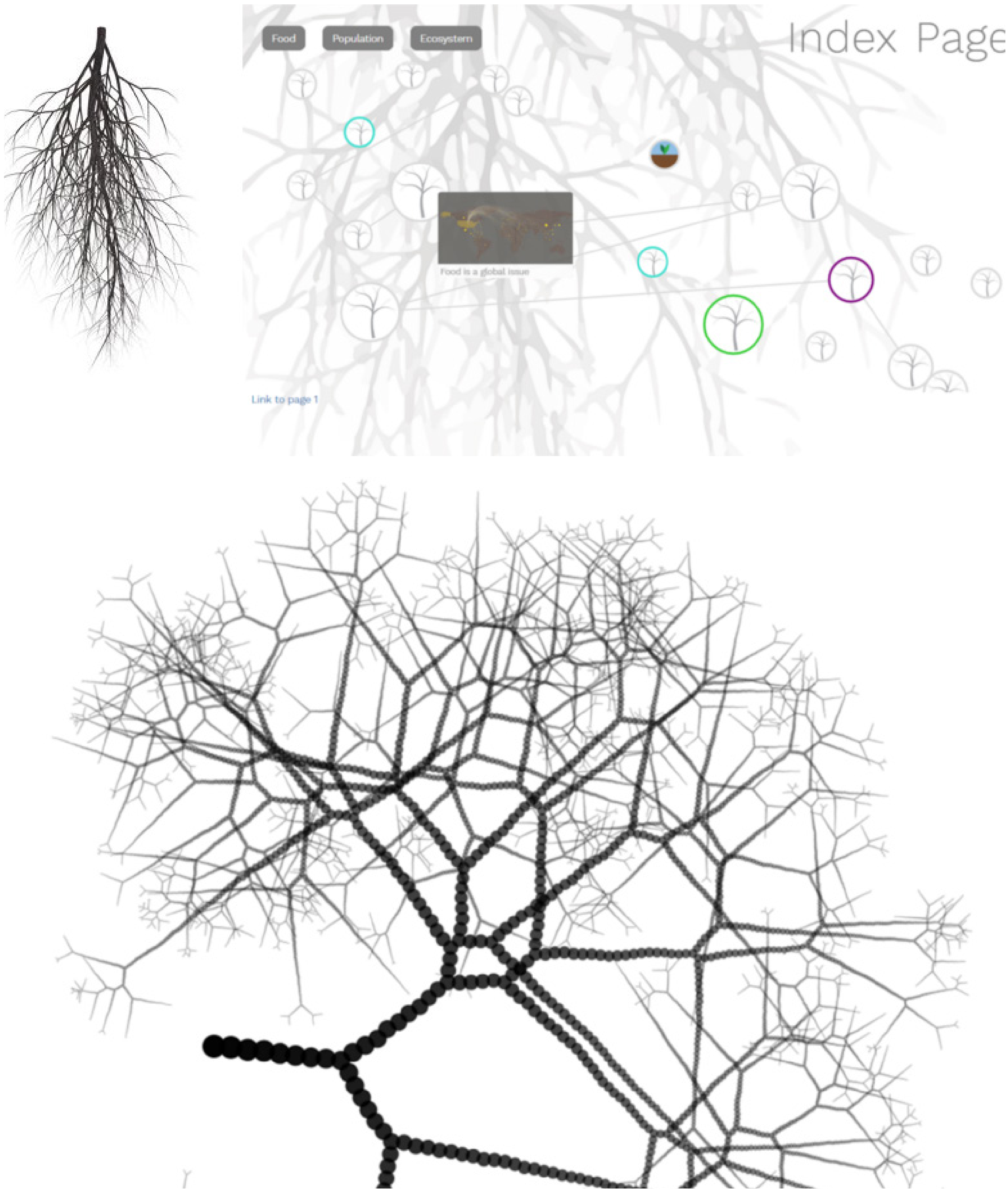
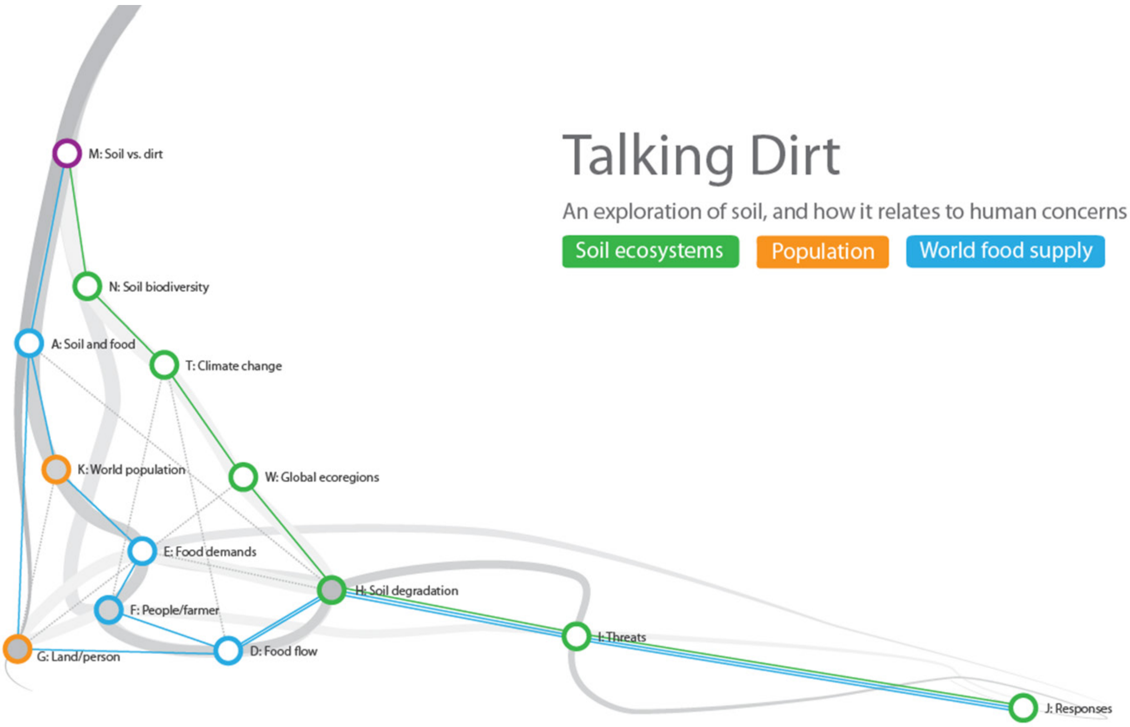
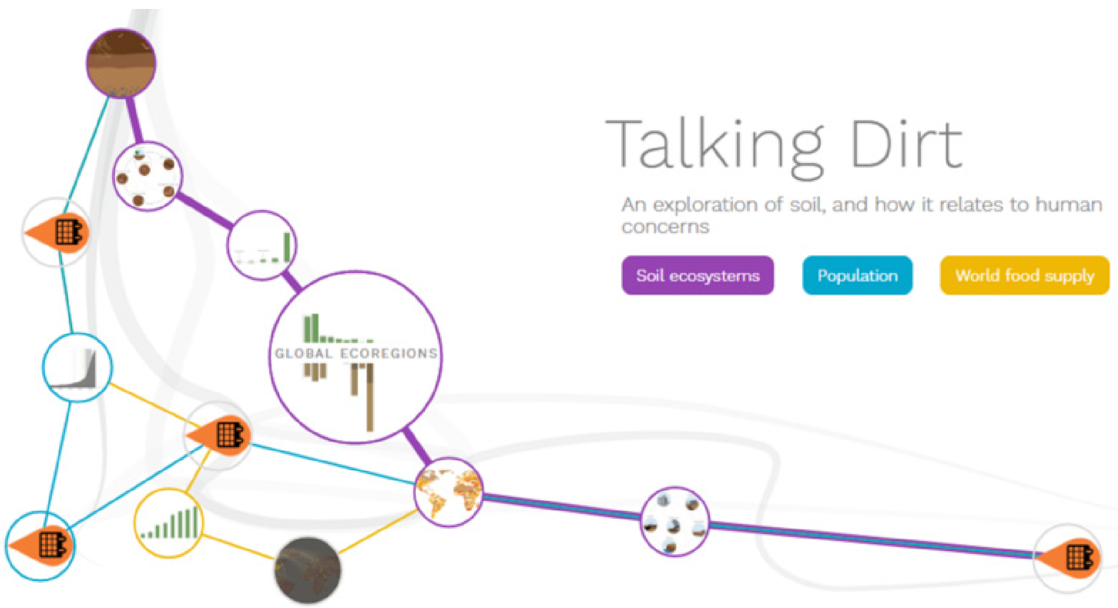


Figure 41. Experimenting with roots as an aesthetic component that gives directionality to the node network. (opposite) Mock-ups made in Illustrator and d3.


Talking Dirt
An exploration of soil, and how it relates to human concerns
Soil ecosystems
Population
World food supply
M: Soil vs. dirt
N: Soil biodiversity
A: Soil and food
T: Climate change
K: World population
W: Global ecoregions
E: Food demands
H: Soil degradation
F: People/farmer
G: Land/person
D: Food flow
I: Threats
J: Responses


Talking Dirt
An exploration of soil, and how it relates to human concerns
Soil ecosystems
Population
World food supply
GLOBAL ECOREGIONS

under construction (it is a habit of mine to make things that still need to be resolved very ugly; it makes them harder to miss, and increases the reward for dealing with the stickier problems).

This visualization was far from perfect, but it was enough to test whether the idea would work, and whether it was worth the additional investment of time that a full implementation would take. I thought that the roots did a fairly good job of structuring the node network and creating a sense of flow within the document, but several users tried to read additional meaning into the images and found them distracting.

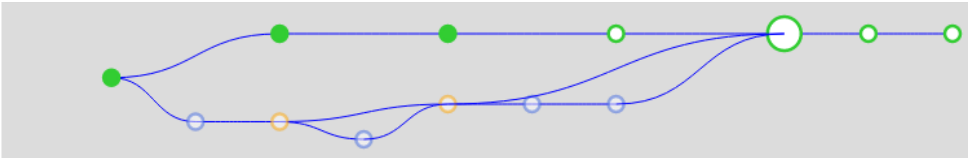
In addition to refining the system overview on the front page, I was also working on making a compact navigation pane that could be incorporated into each page. This navigation panel would provide a contextual overview of the system and help users to maintain their orientation as they browse through the site.

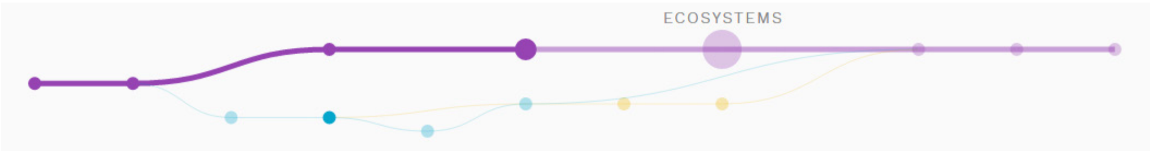
I began by experimenting with d3's tree layout to see if I could use it to organize the nodes in a compact way. The first sketch was focused mostly on making connections and distinguishing between nodes that had and had not been visited. D3 does not permit nodes with more than one parent in a tree layout, so I had to code connectors for the narrative crossings by hand. The resulting visualization maintained the same node structure as the front page, but in a simpler representation.

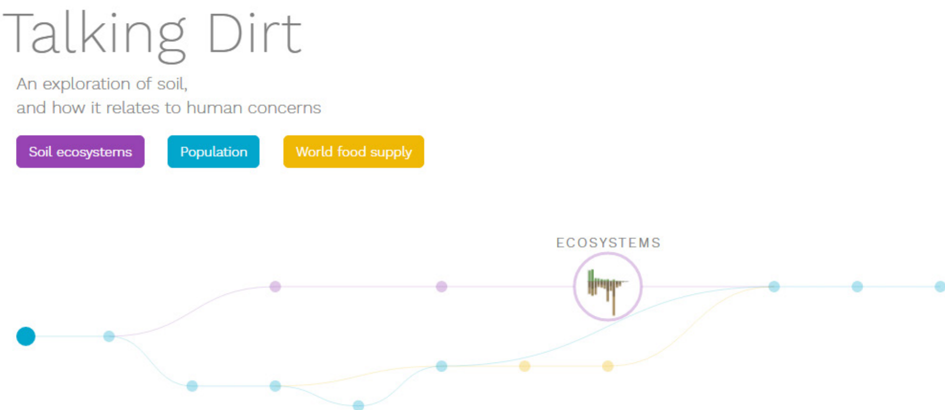
Additional styling and cleanup made the browsing interaction much clearer. If a narrative is selected, its entire path is highlighted with a heavy but semi-transparent line, in the color related to that narrative. Nodes that have been visited are a solid color (even if they are not part of the selected narrative), and the current node is larger than the rest. When the user hovers over a node, its radius increases and a label is drawn to indicate the content that it represents. This visualization feels a bit too much like a subway map for my taste, but it is clean and efficient and most people find it to be clear. I would like to replace this with a more aesthetically-interesting representation in the future, but it works well enough to suit my purposes here.

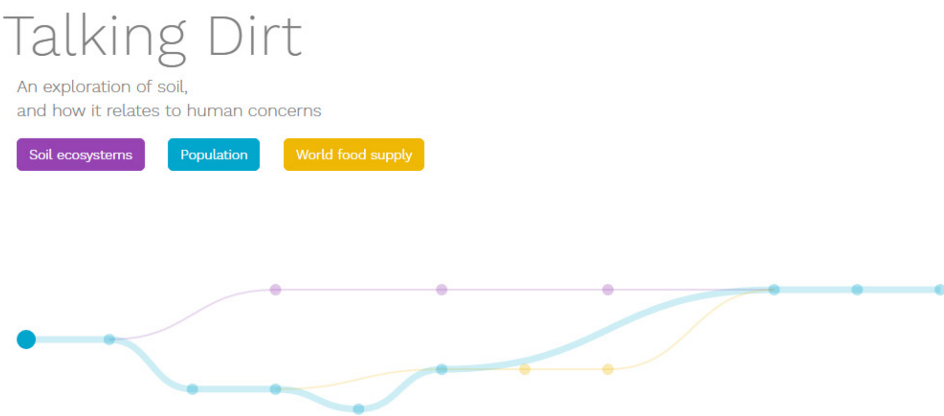
Figure 42. Simplified network diagram developed for page navigation menu and then applied to the index page.

Unfortunately, this representation clashed with the visualization on the original index page. Combined with lingering issues in the details of the earlier concept, it seemed that the roots had to go. In their place, I




ECOSYSTEMS


Talking Dirt
An exploration of soil,
and how it relates to human concerns
Soil ecosystems
Population
World food supply
ECOSYSTEMS


Talking Dirt
An exploration of soil,
and how it relates to human concerns
Soil ecosystems
Population
World food supply

used a modified version of the page navigation bar. Hovering over a node still shows the user a preview of its content, and hovering over one of the narrative buttons highlights the path that the user will take. Overall, this is cleaner, clearer, and works better for most people. I think it misses an opportunity to represent the interdependence and connectedness of the system, but that is a challenge for another day, and will be discussed in the future work section.

Content

In addition to revising old designs, it was also necessary to create new content to fill in the gaps and make smoother transitions between the points that I wanted to convey. Many of these pages didn't need to show a lot of data, but they did need to present background information in a simple and interesting way. I could have adopted an infographic approach and focused on fast facts or including more simple graphs, but I wanted to reserve my graph components for data that really had to be there. I also wanted these pages to feel welcoming and friendly to the user, since they are many of the earliest pages in the site, and jumping straight into abstract representations could be a bit much.

I chose to create a series of icons to represent the different topics for these pages, beginning with modified versions of the designs that I used for the soil poster in the early poster drafts. Incorporating the icons and text into the webpage directly made it feel too crowded, and I didn't think that users would be likely to read a big wall of text. Instead, I drew a diagram showing connections between the different icons in Illustrator, and imported the arrows into the background as an SVG. I converted the node coordinates to screen dimensions, and generated them in d3 as circles with an image pattern fill. Adding a hover effect means that the text only shows when the user chooses to look at a particular node, and looks much cleaner and less overwhelming. A paragraph of text in the sidebar gives an overview of the whole system, and summarizes the main takeaway for the page.

This arrangement keeps some interactivity on each page, and emphasizes the icons over the text as the primary visual elements for these contextual graphics. I've found the synergy between Illustrator

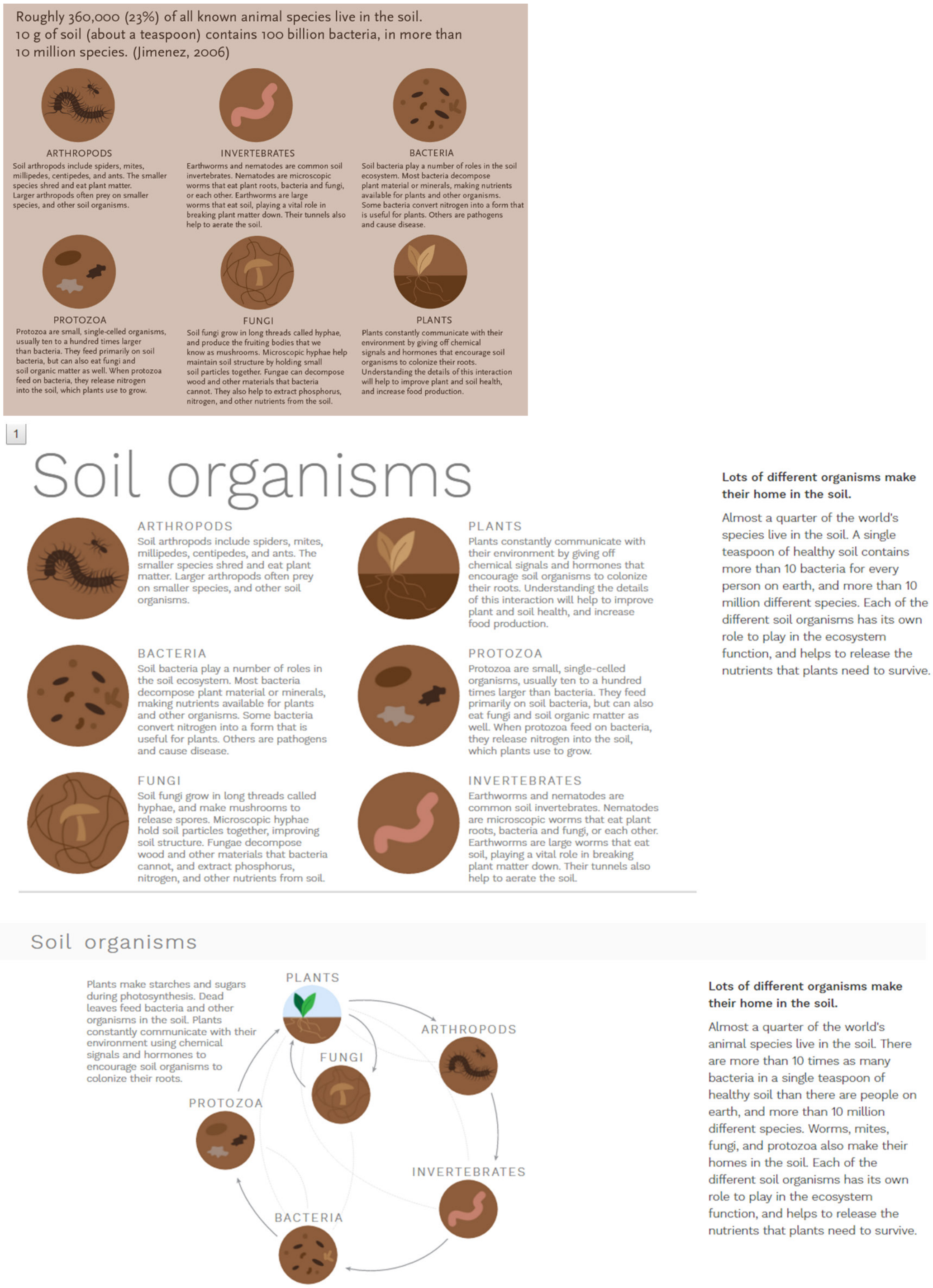


Figure 43. Developing icons to represent basic definitions.

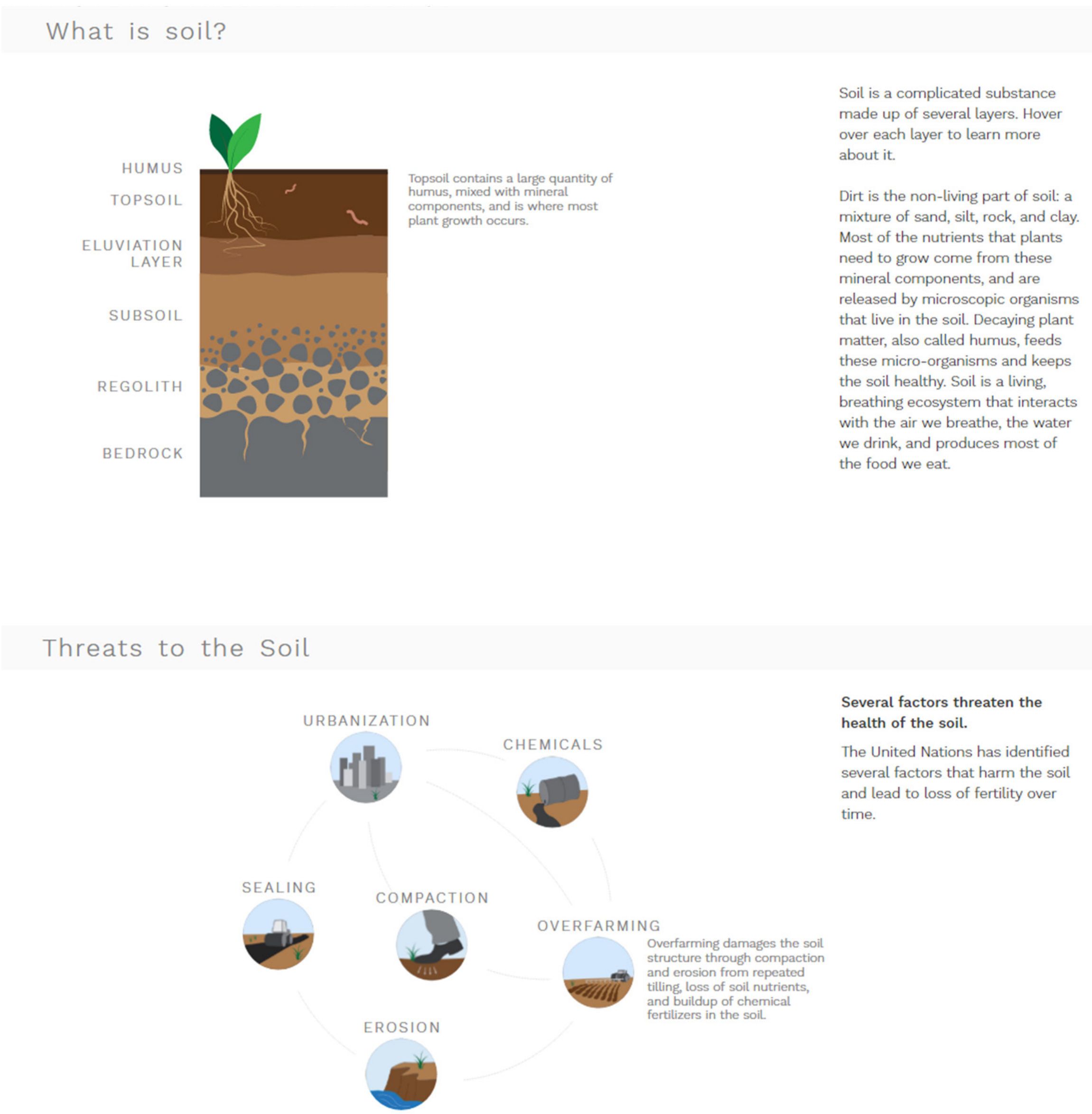


Figure 44. Definition pages using icons and simple hover effects.

and d3 to be very helpful for this part of the project; it's much simpler to draw complicated components in Illustrator and import them into the SVG, and then I use d3 to generate the specific pieces that need to be interactive or bound to data.

I used a similar approach to define different soil terms, and to represent threats posed to the soil by human demands. I also plan to use this method to make the connection between soil and food, and to highlight responses to threats in the final page of the site.

When I imported the icons into the actual website, some of them were somewhat difficult to see, because they were very small. Some of the colors were also difficult to distinguish on lower-contrast screens. To address these issues, I cropped many of the original icons more tightly, and used lighter colors to help the details pop.

Some of the pages needed to fill gaps in the narrative only required a simple visualization of data, with no interactivity. The pages for world food production, people per farmer, and the soil degradation map are examples of this. This data is presented for information rather than exploration, and is intended to give a broad overview of issues related to the more in-depth pieces. For users who are not highly data-literate, these simple, more conventional graphics may be easier to read and more approachable than a fully interactive view.

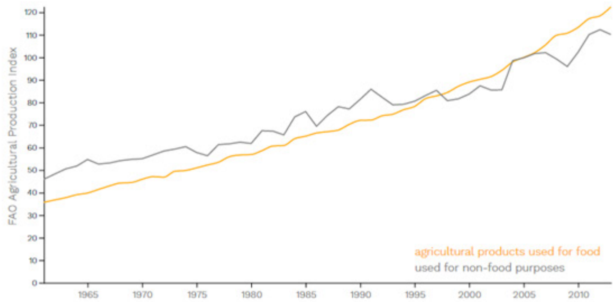
World population

The world population page had a similar purpose, but also presented an opportunity to introduce a bit of historical context using interactivity.

Viewed on its own, the shape of the simple population curve has a pretty big impact. The graph shows estimated values from 10000 BC to 2050. For most of human history, the world population was a few million (not even visible on the graph). It begins to grow larger around 1000 BC, and then gradually increases until about 1800 AD, where things really start to take off. That tall spike on the right shows the expected population increase in the next century, and the size of it is staggering.

On its own, this graph makes a powerful statement, but I found myself wanting more historical context, to better understand how this graph related to other events in human history. It was also an

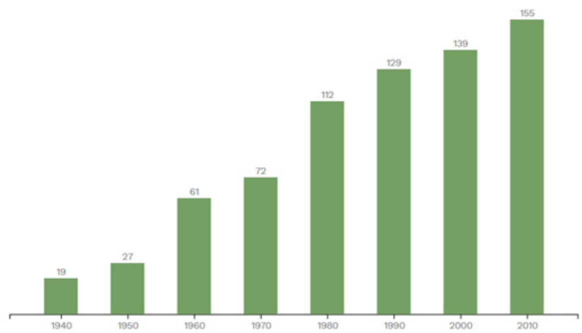
World Food Production



This chart shows world agricultural production. Source: FAOSTAT

Perspectives

People per farmer



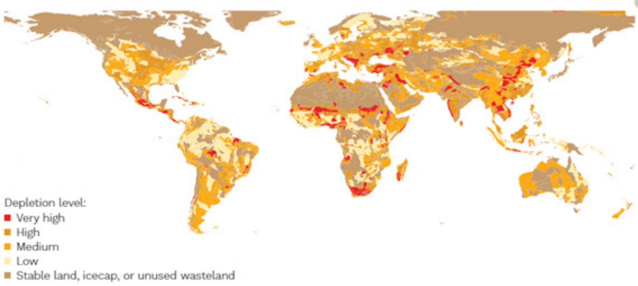
Through most of history, most people were farmers, and one farmer could feed 1-4 people. Increasing agricultural productivity and larger, commercial farms means that a single farmer can feed a larger number of people than they could in the past. This bar graph shows how the number of people in the US fed by one (US) farmer have changed in the past 80 years. Source: USDA

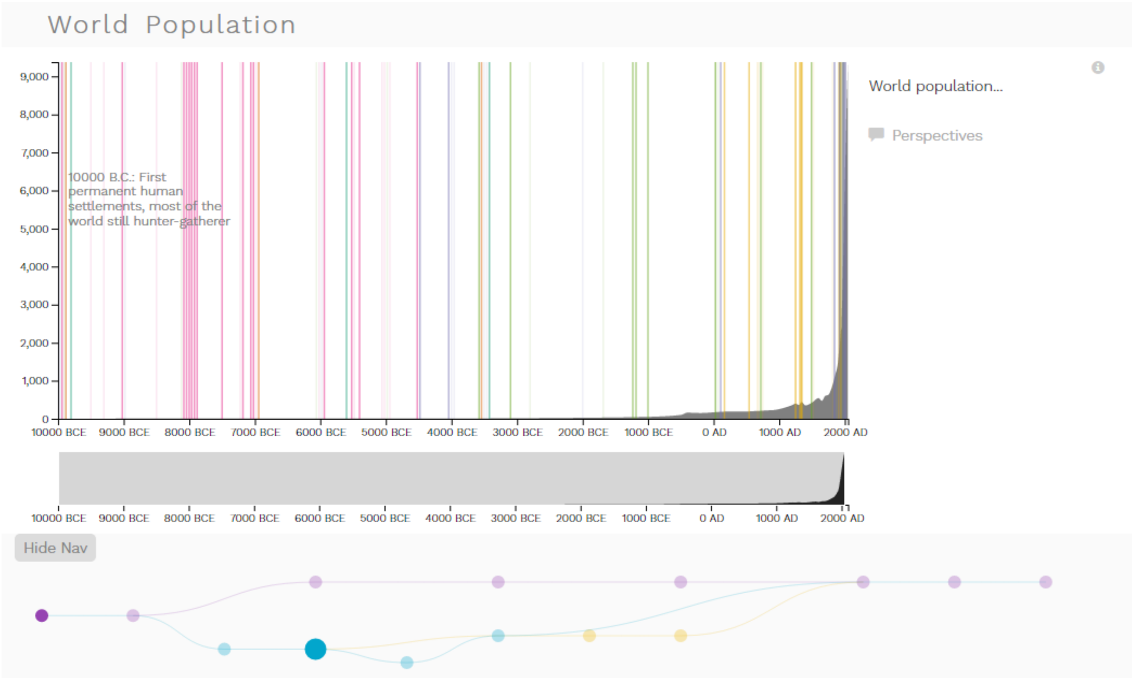
Perspectives

Soil Degradation

Soil degradation means...

Perspectives





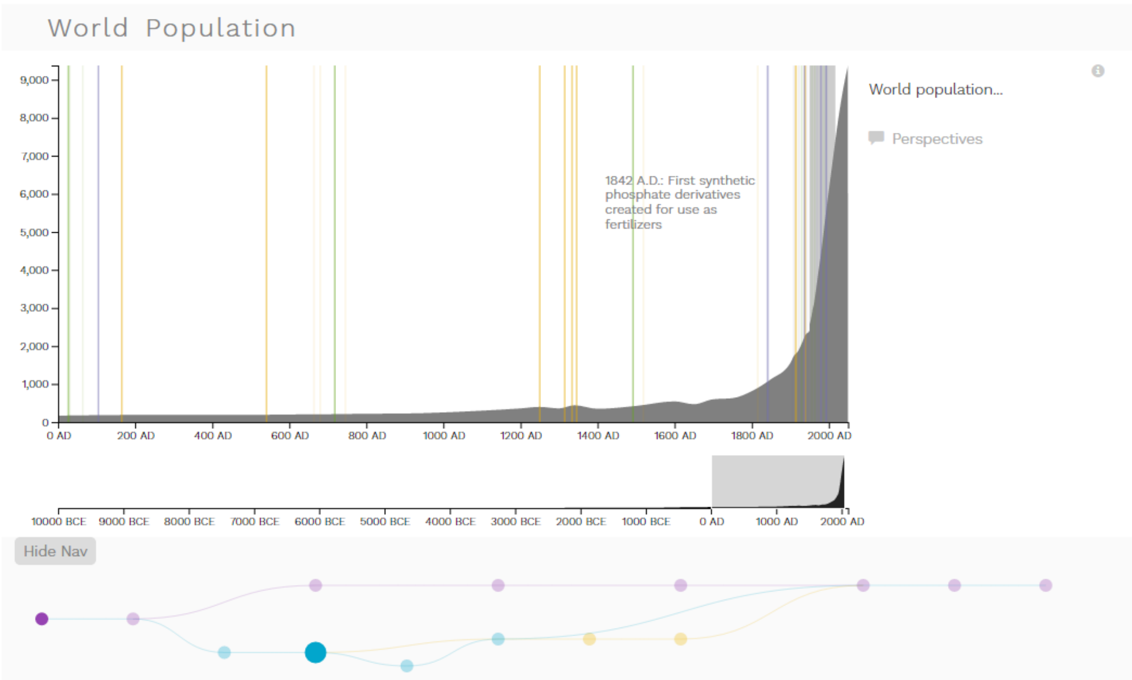


Figure 45. (left) Simple graphics with little interactivity provide contextual overviews. (above) World population graph shows historical context as well.

opportunity to provide some of the history of how agriculture has developed throughout human civilization, which is useful when discussing where things stand today.

Using Wikipedia, I handpicked events for the timeline that were related to landmarks in agriculture and human population. Early on, this means a lot of "firsts": first record of agriculture, first record of seed storage, first evidence of livestock, etc. Then, there are a lot of cultural/historical events that help to frame the history: settlement of Egypt, founding of Rome, etc. In the more "recent" history, there is a lot more data about plagues and famines, which are important because they're related to population loss. The modern period has several technological advances, with things like GMO crops and industrial fertilizers showing up on the scene. To help distinguish the different kinds of events, I gave each category a different color. Higher priority events are represented with slightly heavier lines, and lighter lines show lower priority events.

I considered using this as an example to discuss issues of uncertainty and data completeness; except for a very narrow window at the end of the last century, almost all of this data is based on predictions, models, and archeological data that is incomplete. The data gets better and better as we approach modern times, because tax records and other government reports help to supplement other ways of guessing populations for different places. Universal records have only been kept for about the past 50 years; the population numbers for the UN and Census Bureau data only go back to 1950.

In the end, I decided against including that level of detail in this version of the site. In the context of the broader story, the magnitude of population change is the most important part; focusing on the details of upper and lower bounds on the historical data would be an unnecessary distraction. I do still think that this would be interesting information to include, and it is a useful opportunity to engage withuncertainty and discuss data provenance. Had this data been more central to the main topic, I would have chosen to develop it further. Instead, I will discuss these variations in the data notes for this section (added as an information button at the top right of the page), and may revisit it in a more in-depth visualization at a later time.

For the integrated version of this page, I chose to zoom in on the past 2000 years for the default setting, leaving the other data available in case people want to explore. This maintains enough of the graph to show the shape of the curve, but focuses the default view on more recent events.

Ecoregion Carbon

As the webpage design and narratives progressed, I also revisited earlier experiments to update them to match the tighter focus of the integrated site. The map linking soil carbon storage and world ecoregions was one of these examples; the animated bar chart of global carbon distributions was another.

The first step was to import these visualizations into the new page template. The first image in figure 46 shows the Ecoregions visualization in its original form, on the homework page for my CS class. The second image is the result of importing it into my placeholder template form, which required converging three separate SVG elements and a Canvas into a single SVG/Canvas stack. I also decided to eliminate the bar chart on the far left, as discussed in the Preliminary Experiments section, and I removed Antarctica from the map to declutter the space. The third pane shows the visualization in a later version of the template, where the page styles are becoming more defined.

This revision focused on simplifying the visualization to show only its core storyline. I started by removing the WWF ecoregion designations from the map. Since the corresponding areas are highlighted when the user hovers over a bar in the bar chart, it's not necessary to name the individual regions separately, and the complicated color scheme and legend add a lot of unhelpful complexity to the visualization. This does mean that the user must interact with the page to see the map legend, but that seemed like a reasonable price to pay for the added clarity.

I also wanted the main story to be clearer when the user first arrives at the page. One of the interesting things to note in this graphic is that the distribution of above and belowground carbon is very different around the world; the tropics have the most plant material (aboveground carbon), but relatively poor soils. The northern ecoregions store much more carbon in the soil, and sequester far more total carbon than the

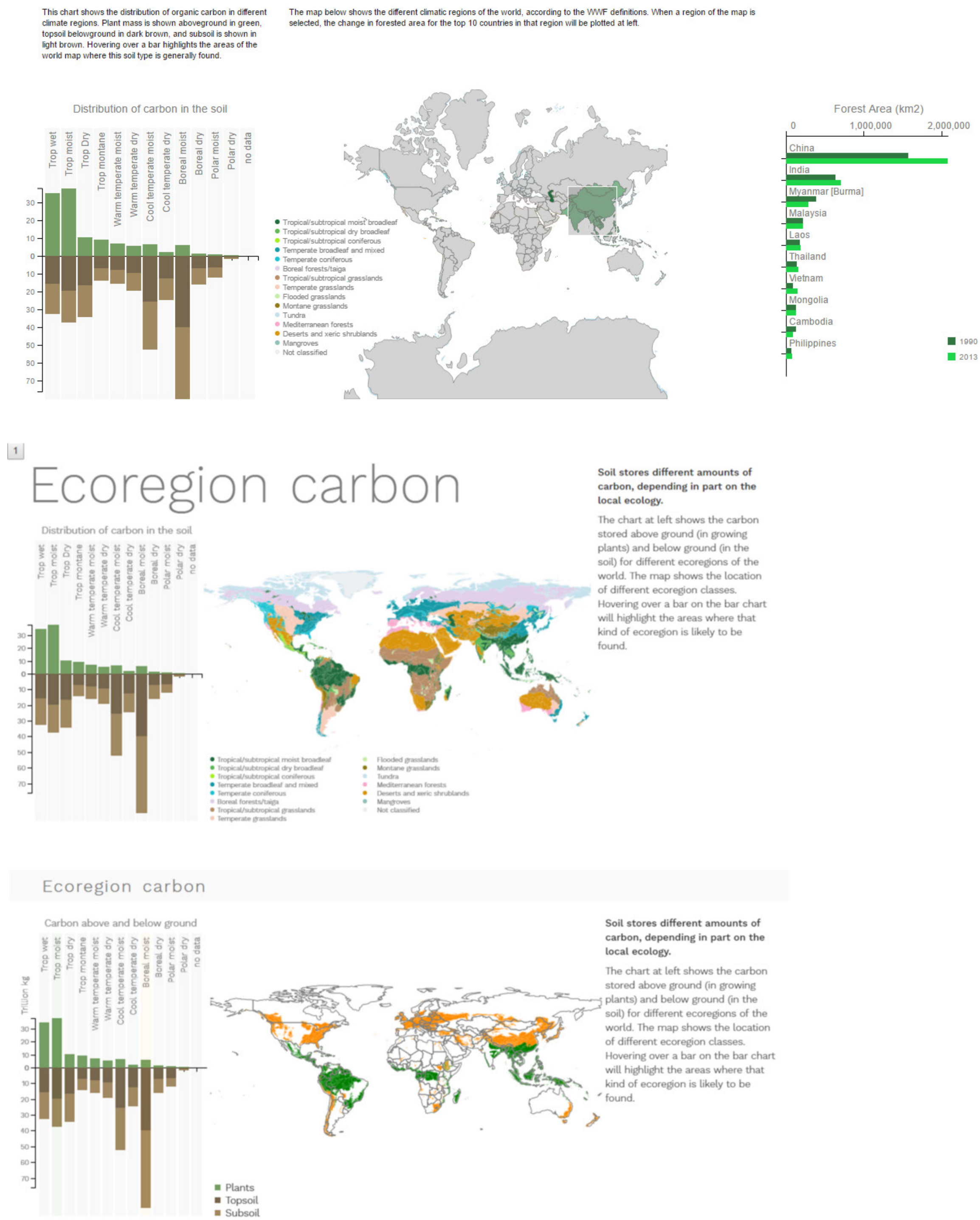


Figure 46. Simplifying linked views of a map and bar charts that highlight regions with different soil carbon content.

tropical regions do. In the revised version, these two bars are highlighted when the user first arrives at the page, and the difference will be described in the sidebar text in the final version. The user is then able to select individual bars and explore the dataset further. Additional improvements include making the bar chart highlights more visible, and possibly adding cues to clarify what the initial map colors mean.

Simplifying Global Carbon data

The Global Carbon visualization underwent similar simplification. The original version (figure 33, page 112) was an animated bar chart, which took about a minute and a half to run. As an isolated graphic, it worked well and had a strong emotional appeal, but I decided to eliminate the animation in the final version for two reasons. First, I find it annoying when a website starts playing background music without warning, and the animation really requires the music for optimal effect. Second, I thought that the pace of the animation was too slow for a situation where this visualization is just one stop in a longer narrative. I liked the gradual unfolding of information that the animation provides, but it wasn't the right paradigm for the rest of the visualizations and wouldn't feel like part of the same design.

Removing the animation meant that I needed to find another way to make the main points clear from the beginning. In the current version, I decided to get rid of the pie chart as interesting but non-essential information, and to add simple labels for the two most important bars. The comparison between the carbon stored in soil and in all living organisms on earth (including plants) was the main story that I wanted to tell here, and the other bars are really just added for context. The oceans and coal, oil, and gas bars will attract attention naturally because they are so large, so I really only needed to label the smaller bars to make them stand out. These updates were fairly small changes, but they help to clarify and simplify the design.

Most visualizations for public consumption would stop here; the graphic has presented real data, the narrative does not distort or exaggerate the information, and there is a clear takeaway point for the viewer. As a graphic for science communication, I would like it to go one step further. This simple narrative does not fully represent the nuance of this topic, and does not present alternate views. Its success relies on the authority of the source, supported by citing the original scientific paper. But the visualization does not yet do justice to the full content of the paper, which gives a comprehensive overview of global carbon cycles and how changes to soil carbon might affect climate change. As the author points out in the conclusion, the real question is: "which option is environmentally and economically viable under what ecological conditions?" That level of complexity is not yet reflected in my animated graphic. This graphic also ignores other perspectives within the scientific community as well. Other authors debate the relative gains in soil carbon storage compared to other methods, and argue (correctly) that the soil cannot absorb enough carbon to reverse climate change.

Though I was careful not to state it in this way, simplistic interpretation of my animated introduction could suggest that organic farming is sufficient to solve climate change. In reality, this is just one approach of many that must be taken to address an issue of this magnitude. There are many excellent reasons to focus on improving soil health in addition to its benefits for climate change, but it's important to realize that increasing soil humus is not a cure-all solution.

As with the world population data, this level of detail feels like too much for a page that is trying to get a simple point across, but I would like to include more detailed discussion in the site in some way. Some aspects will be discussed in the data information popup, and I am also experimenting with including "perspective" buttons that present different views of the data. The latter will be discussed more in the Reflections section.

Returning to land per person

The last major data component to incorporate into the thesis project site was the land use data. I described several early sketches for this visualization in previous sections, but had not yet resolved how all of these pieces would fit together. I wanted to create a visualization that would allow users to view data for a single country, compare two countries, and look at global rankings. I also wanted both longitudinal

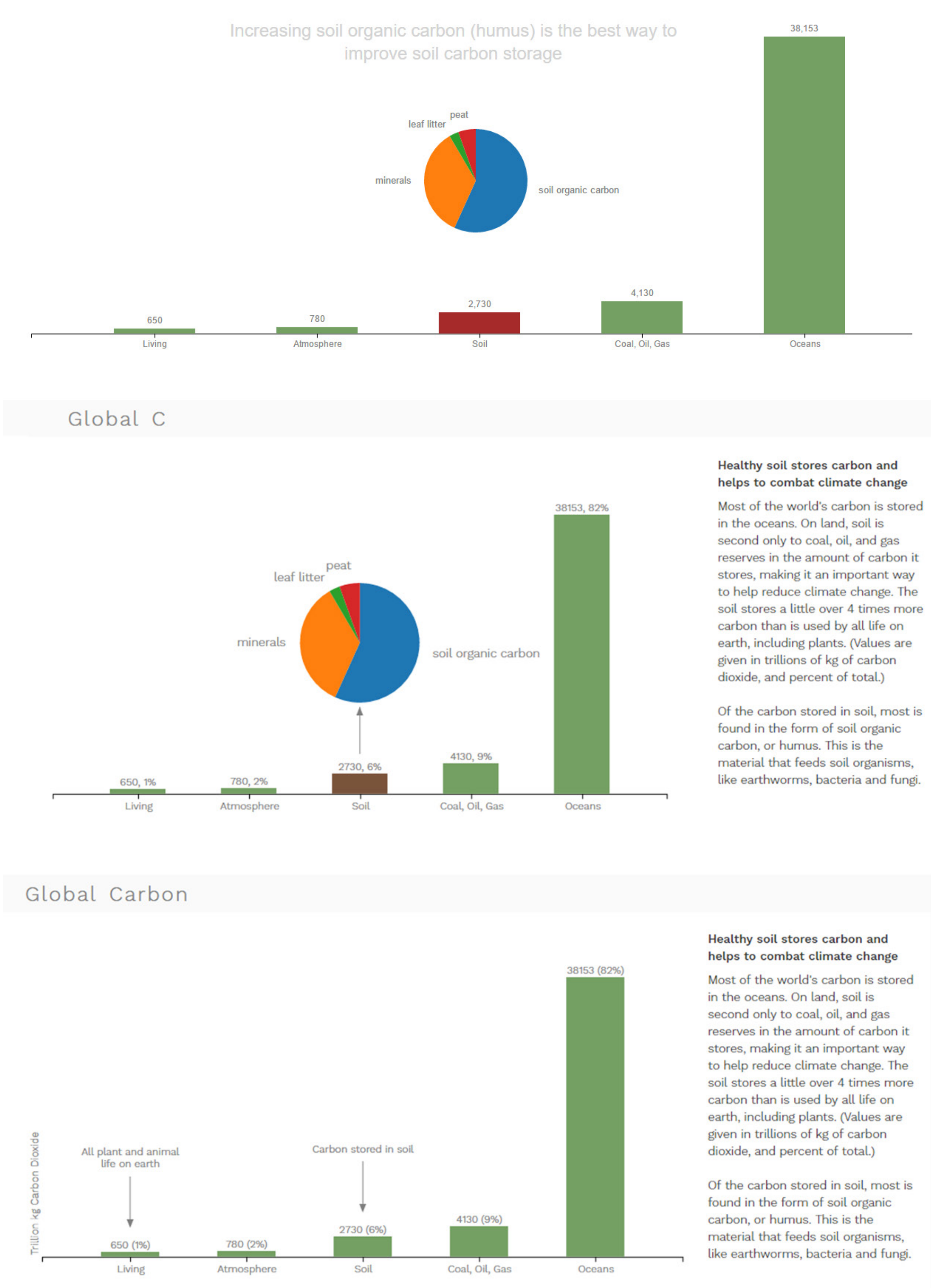


Figure 47. Simplifying animated bar charts that show where carbon is stored globally.

and comparative displays, if possible. Along with the Food Flow visualization, this page of the site would allow the user deeper access to the data to encourage hypothesis-finding and discovery of their own storylines.

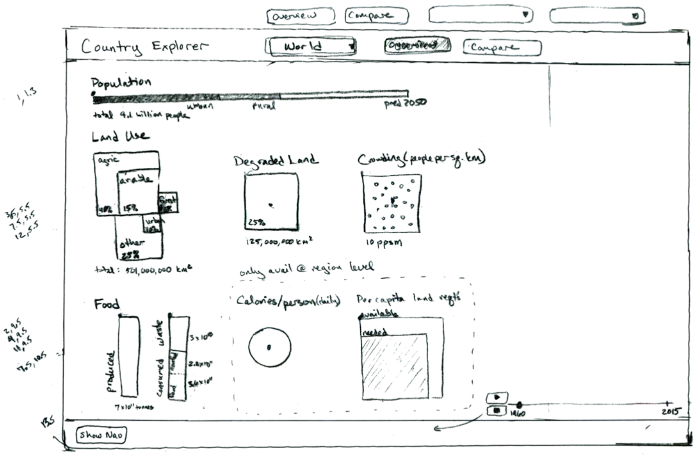
Since creating the dashboard to assess data completeness (figure 26, page 98), I had incorporated data about the balance of food imports and exports from different countries, and had also identified a new dataset that lists per capita food consumption and calculates the total land area required to feed the population of different regions of the world. By distributing this information across all of the countries in a region and using the country's own population values, it was possible to calculate compare the total land area of each country against the amount of land needed to feed its population. This combination of datasets weaves all three of the project narratives together, and provides a strong foundation for creating an interesting data story.

Because the data was already processed and well characterized and the site layout is fairly well identified, the initial sketches for this phase of the project could be quite detailed. The front page shows a view of all of the different variables for a particular country in a single year, to give a synthetic overview of all of the topics. The visualization begins with information for the entire world, and allows the user to select a country or region to explore in greater detail.

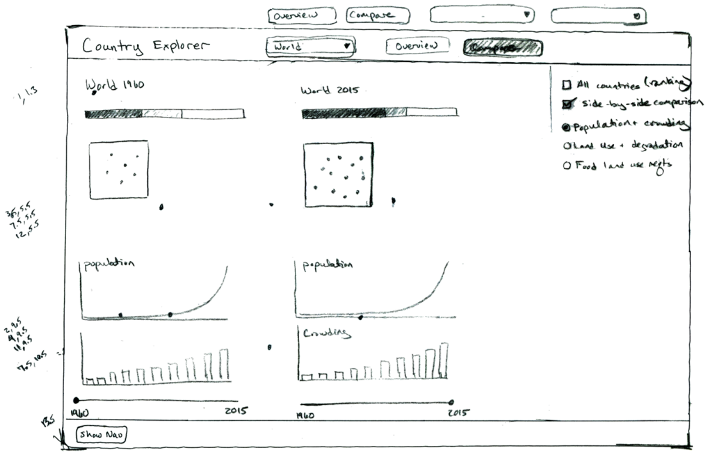
From top to bottom, the sketch for the front page gives the total population, percent land use for each of the major categories, amount of degraded land, crowding (people per sq. km), food trade information in two bars — one showing imports and production values, the other showing exports, consumption, waste, and non-food uses, the average caloric intake, and finally a comparison of the actual land available in that country/region, and the total amount needed to support the population's food needs. I also planned to include a time slider that would allow users to browse all of the years in the dataset, with a play button that would auto-advance so that they could watch changes happening over time.

The next page would allow side-by-side comparison of two countries, with the ability to select between different pairs of related variables — population and crowding, land use and soil degradation, and food and

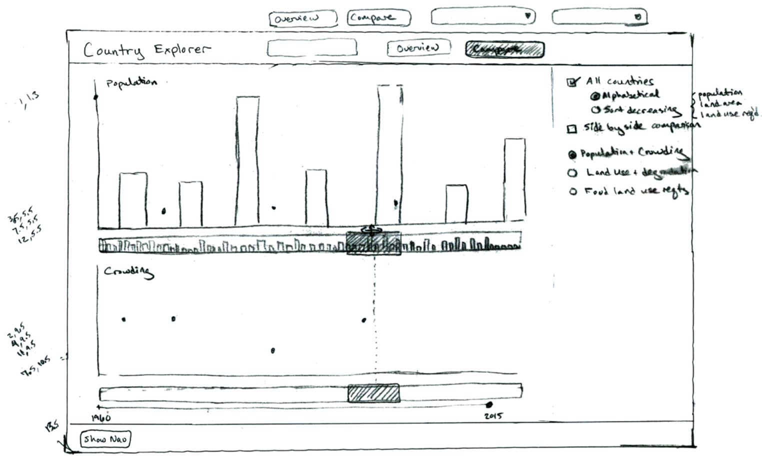
Figure 48. Sketches for updated land use visualization.



18
scale to the selected country
(max for all yrs)



18
Scale to the larger country
(max for all yrs)



total land requirements for the countries. The top section of the page focuses on visualizing a single year in the same format as they are shown on the front page, and the bottom section shows the same data as a line or bar graph over time. Time sliders at the bottom would allow you to compare values for the same country in different years, or for two countries in the same year.

Finally, I hoped to include a ranking page that compares all of the countries in the dataset, and lists them in order according to user-chosen filters. This is for the people who only want to see the top or bottom 5, and don't really care about the rest. For this, I decided to use a format similar to the one I used for the world population data, which use d3's brushing and linking functions to show an overview of the entire graph in a small menu visualization, while the main display zooms in on a selected region to show more detail.

Figure 49 shows an early version of the visual components for the front page. (The gray dots are positioning guides that I used to determine the relative placement of the different features while I worked on the code.) I learned several useful things from the mockup. First, the technique I chose for drawing the dots inside the crowding boxes meant that the images jumped around randomly from frame to frame. This made it extremely hard to tell whether the numbers were increasing, or whether the dots were just jumping around.

The final version will add new spots to a single series, so that there is constancy between frames. Also, it was obvious that the change in land area would be hard to see at this scale, especially because the data itself changes slowly over time. This was compounded by occasional holes in the data, which meant that the different visual elements sometimes blinked in and out. Finally, it was difficult to pay attention to changes in all of the variables as the time slider moved, especially since change occurs across multiple axes. I had expected most of these observations, but wanted to see the code in action before making a final decision about how to deal with the different difficulties. Because there was no clear change in some of the datasets and exaggerated or false change in others, and because it was difficult to track changes in this view anyway, I decided to eliminate the year selection slider,

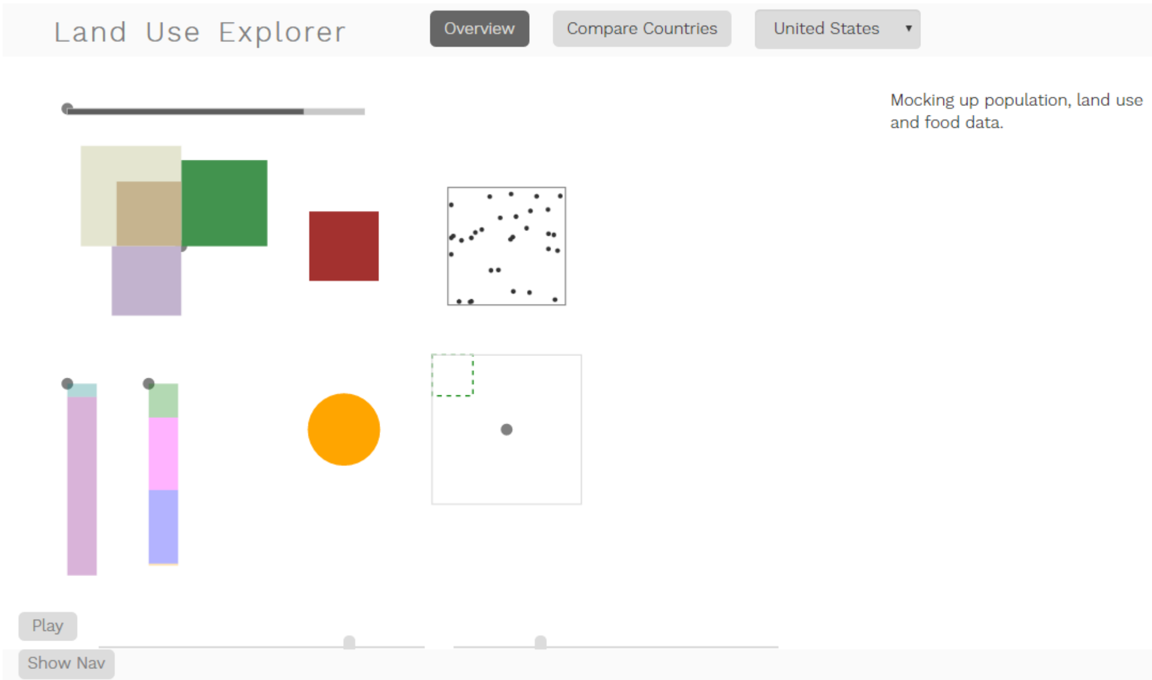


Figure 49. Early implementation of the land use visualization.

and choose a single year for display. The degraded land area is only available for 2003, so I chose that year as the basis for the rest of the visualization.

Removing the time slider also helped to resolve a second problem with this design. In the updating version, the land use squares needed to grow bigger from the center out, so that growth of one square wouldn't force the others to move over time. This meant that the land degradation and crowding squares also needed to be center-aligned. The food balance bars and the calories circle didn't fit well with this alignment, though; some components fit better to a top-aligned grid and others to a center-aligned grid, and nothing ever quite matched up.

Once the time slider was removed, it was no longer necessary to look at changes happening over time, so there was no quantitative reason that the land use squares had to have a consistent center origin. This meant that I could use labels to define the grid, and hang the visualizations off of those, making the whole view feel much more orderly.

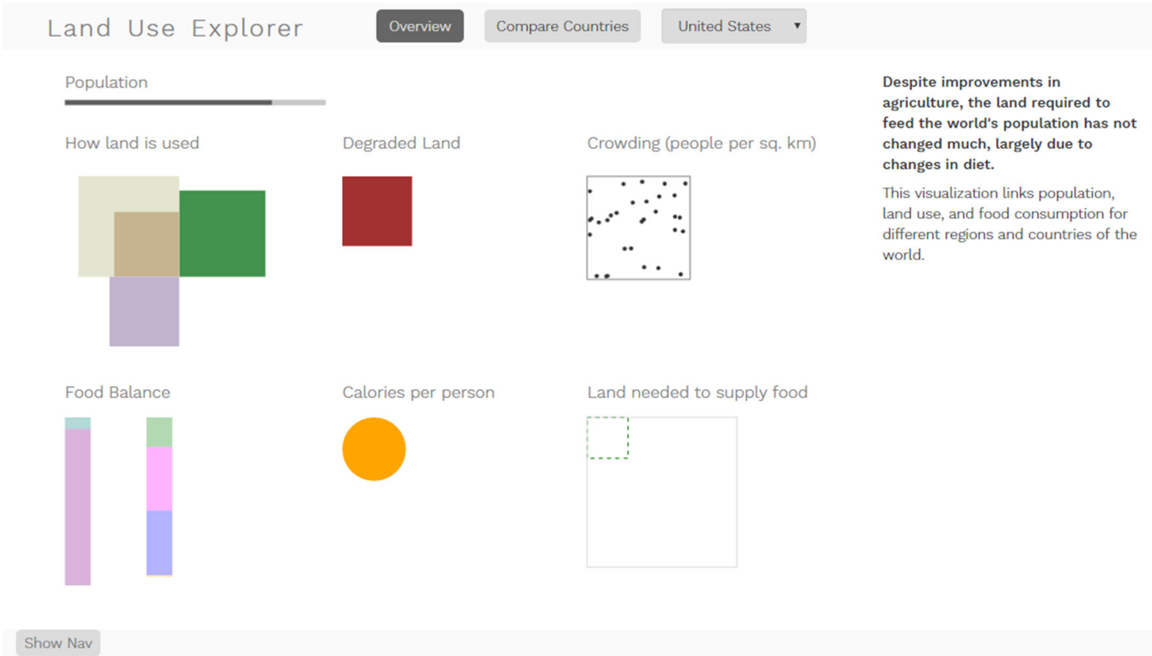
Adding the first series of labels also helped to make the overall story clearer, and exposed some issues with the flow. While I was roughing out the code, I kept roughly the same organization that I had used in the original land use poster (page 92), which was optimized for space efficiency, but only included a subset of these variables. In the first version of the code, I added the new variables in on the bottom row. Reading from top to bottom and left to right, that meant that I started with population, jumped to land use and degradation, back to population/crowding, then to food imports, food consumption, and total land required. It wasn't impossible to read the data in this way, but I felt that there were more intuitive ways to group the information.

Rearranging the same components, the newer narrative talks about people first; then total population and crowding, how much those people eat, and the land needed to make all that food. Then, it shows the food imports and exports, and the amounts produced, consumed and wasted. Finally, it shows how the total land area is currently used, and how much of it is under threat of degradation. Even before the final annotations are added, it is structurally much clearer.

I also changed the color of the square that represents the land required to grow food to match the color of the food narrative, and the calories circle. I switched it to a solid stroke to make it obvious that this was the more important of the pair, and made the total land area square dotted to help it recede. These are small changes, but they help to link those two components of the dataset. Comparing different countries in this visualization begins to uncover some very interesting stories.

Singapore has one of the highest population densities on earth, with more than 7500 people per square km, as of 2013. It's also a tiny nation, and one of the few cases in the dataset where the land required to feed a country's population far exceeds the amount of land available in the country itself (by about a factor of 10). Roughly 25% of its land has become degraded since 1982 (the red square shows how much land lost productivity between 1982 and 2003), and 50% of its land was listed as urban in 2010. Even if Singapore used every square inch of its soil to make food at the highest possible productivity, this country would be unable to feed itself without importing food from other places.

Using the visualization to look at the whole world, the land area use looks pretty good; we farm a little more land than we currently need to feed the world population, and it looks like there's plenty of other land available. Of course, when you factor in things like local climate, things are a little less rosy — that total land area includes Antarctica, Greenland,



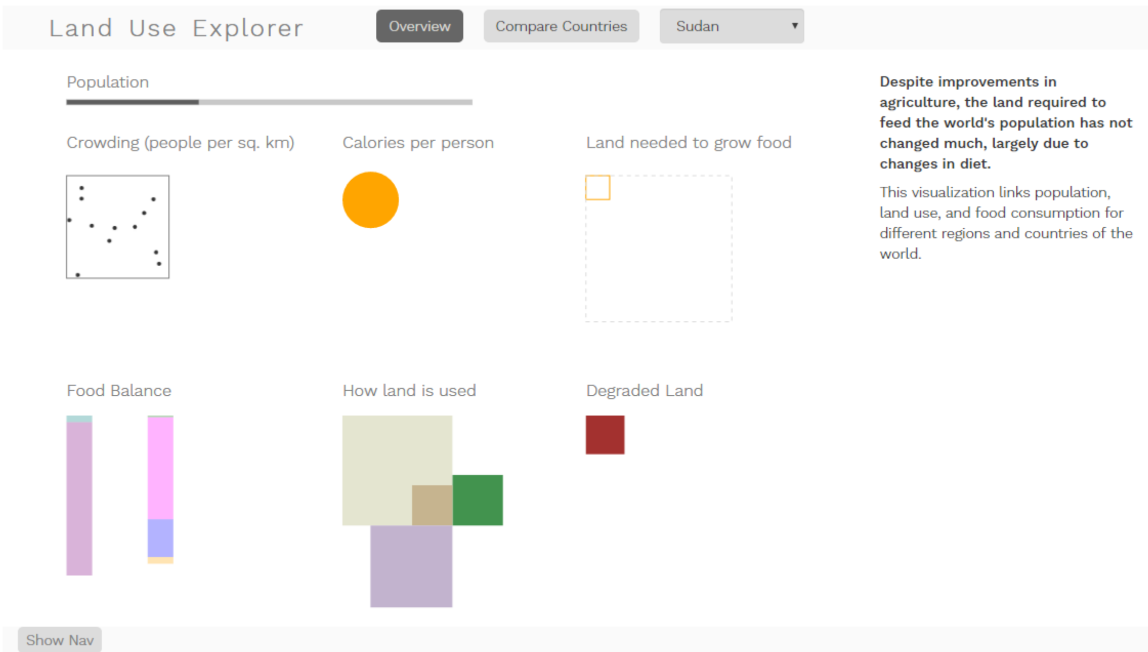


Figure 50. (opposite) Clarifying narrative and grid structure of land use visualization.

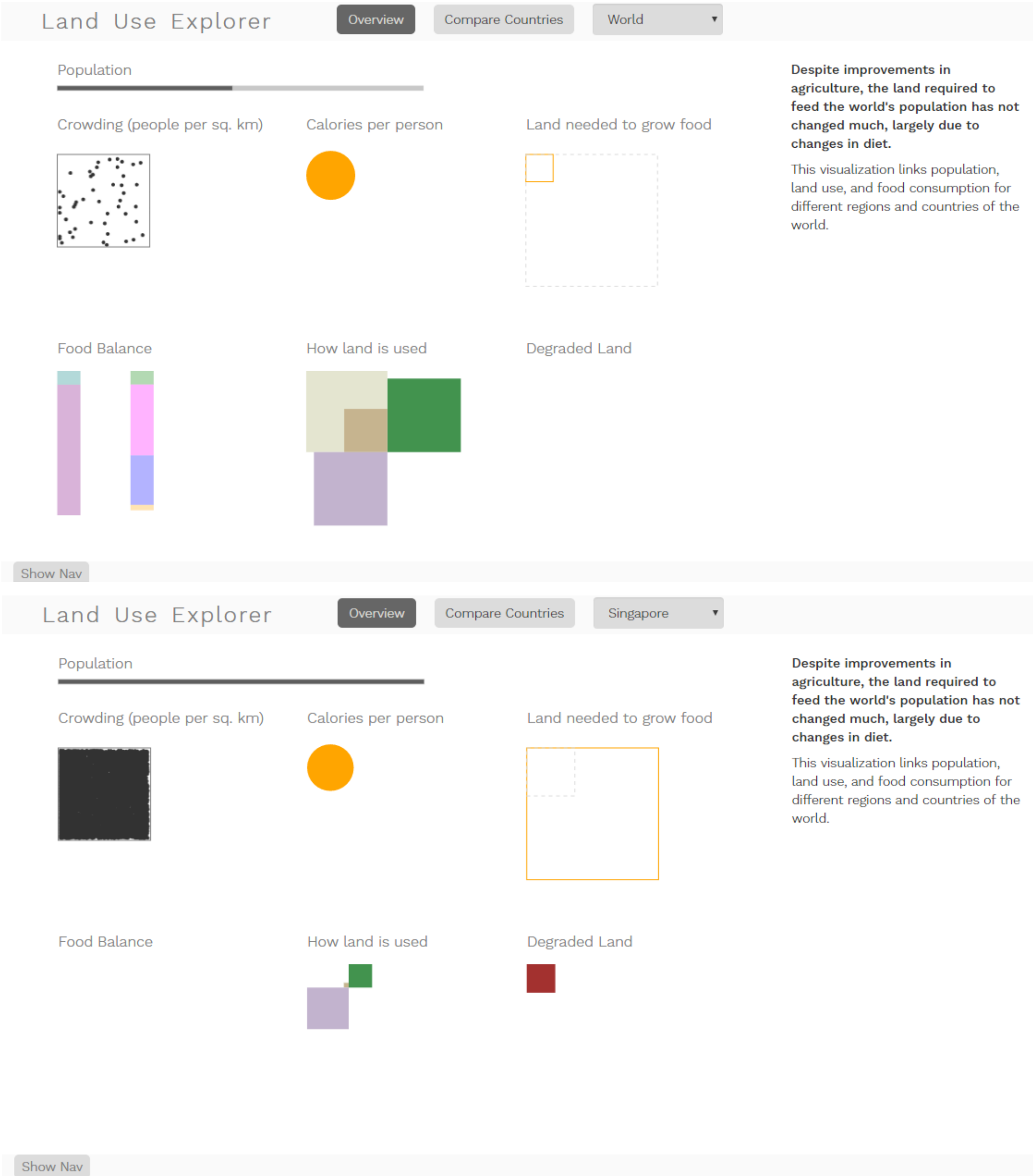


Figure 51. Comparing global land use statistics with those of Singapore. Values for crowding and the land required to grow food for the population show particularly large differences in this comparison.

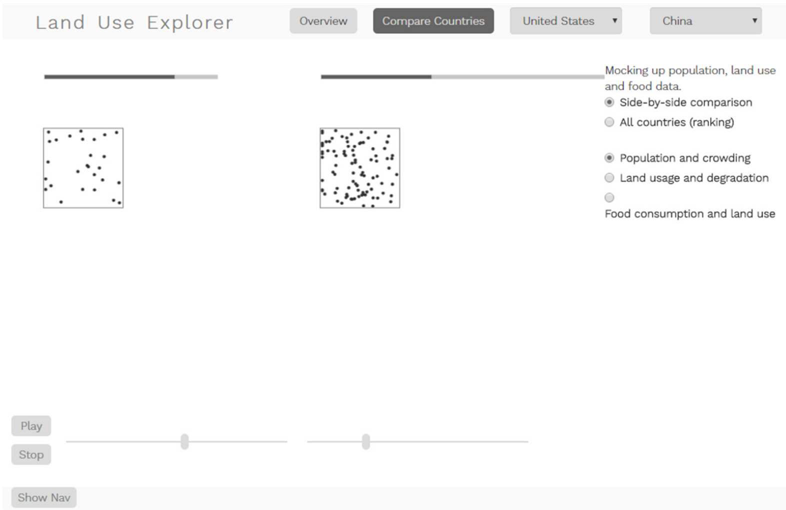
the Sahara, the Himalayas and Siberia, none of which are known for being great places to garden (the arctic, Antarctic, and desert regions make up about a third of the land area on earth). Most of those likely fall into the "other" category, though some are included in the agricultural lands estimate as well; if there's enough grass to feed a herd of goats, the land counts as agricultural, which is how Saudi Arabia ends up with an agricultural land area 82% of its total size.

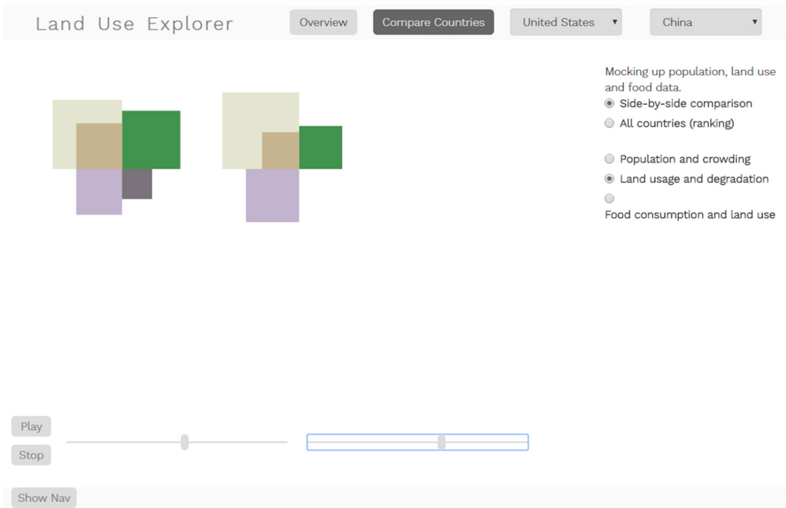
From that perspective, it looks like a lot of the fertile land available falls into the forest category, which makes it easy to see why deforestation is such a serious concern: by and large, the remaining fertile land is currently covered in forests. Thinking about the amount of carbon stored in those forests and all of the ecological reasons to protect them, it's easier to understand how all of these different topics start to interrelate.

The remaining tasks for this visualization include finding ways to guide users through these kinds of comparisons, so that they can gather useful insights from the datasets shown. Including quantitative labels and annotations that define each category of the visualization is also an important next step.

Once users have explored the relationship between different variables for a single country, the second page comparison view will allow them to get deeper into the individual datasets and look at two countries side by side. For this view, users choose a pair of variables to visualize for two different countries or years, and are able to view them side by side.

When you compare two countries side by side, the size scales to whichever of the two variables has a larger value, so that you can immediately see the relative distribution of values for the two countries selected. (This is also a good way of comparing a selected country to data for its region, or for the world.) Sudan and China have pretty similar land use distributions, and similar proportions of degraded land, but China's larger size means that its percent degraded land makes up a much larger proportion of the planet. Combining these different views helps to get a sense of how a country fares on its own, and how it stands in relation to the rest of the world.


Land Use Explorer
Overview
Compare Countries
United States
China
Mocking up population, land use and food data.
Side-by-side comparison
All countries (ranking)
Population and crowding
Land usage and degradation
Food consumption and land use
Play
Stop
Show Nav


Land Use Explorer
Overview
Compare Countries
United States
China
Mocking up population, land use and food data.
Side-by-side comparison
All countries (ranking)
Population and crowding
Land usage and degradation
Food consumption and land use
Play
Stop
Show Nav

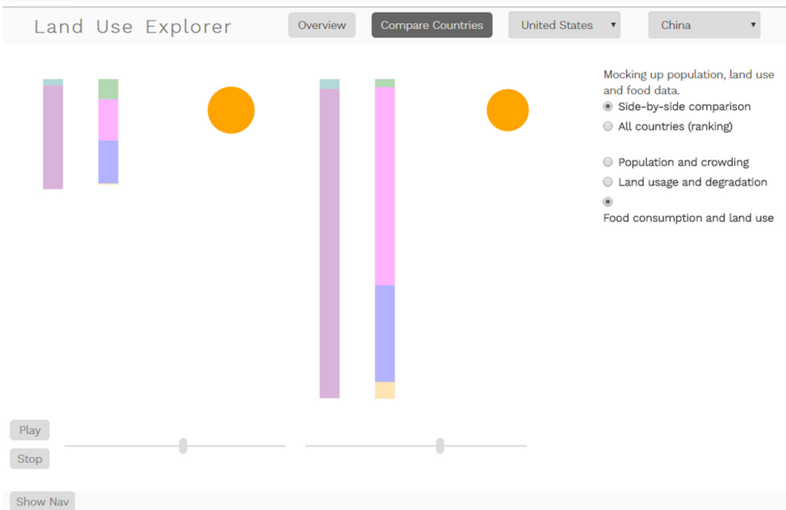

Land Use Explorer
Overview
Compare Countries
United States
China
Mocking up population, land use and food data.
Side-by-side comparison
All countries (ranking)
Population and crowding
Land usage and degradation
Food consumption and land use
Play
Stop
Show Nav

Figure 52. Side by side comparisons of land use statistics.

The extra space at the bottom of this view will show the longitudinal change in the plotted variables as a bar chart, so that users can also get a sense of how they change over time. The plots shown in the last panel of figure 52 are from an earlier version of the land use visualization, and will probably be the basic model for the charts in the new version as well.

The last view allows users to make comparisons in the value of a single variable across all countries in the dataset. The visualization starts out with a bar chart of the selected variable for all countries in the world, in alphabetical order. Mousing over a particular bar shows its name and clicking on a bar highlights it so that it's easier to follow that particular object through animated transitions.

Switching from alphabetical to ranking mode rearranges the bars by the selected variable, in descending order, making it easy to identify the largest and smallest countries for a particular dataset. The bottom pane is a brushable overview of the entire system, which allows the user to zoom in on a particular section of interest, while still retaining a contextual overview of the whole.

It was clear right away that I was going to have problems with labels on the x axis; even using two-letter country codes, the axis labels are illegible at anything but the closest zoom. I thought it might help to rotate the visualization so that I could use human-readable names for the axis labels, but when I tried it I found that the bar chart had to be over 2000 pixels tall before the labels were clearly legible, and I think it's important for the user to be able to see the whole range without scrolling. I decided to go back to the vertical bars instead, and am working to come up with another way to make the labels work.

These three modes constitute a series of visualizations that allows the user to slice the information in different ways. The overview page shows all the variables for a particular country, to explore relationships between them. These graphics are made up of simple shapes and allow qualitative comparisons, especially focusing on land area and population size. If people are just clicking quickly through the pages, this visualization unites several of the narrative threads from other parts of the site without requiring in-depth interaction and filtering. The

Figure 53. Comparing the ranking for all countries against different land use variables.

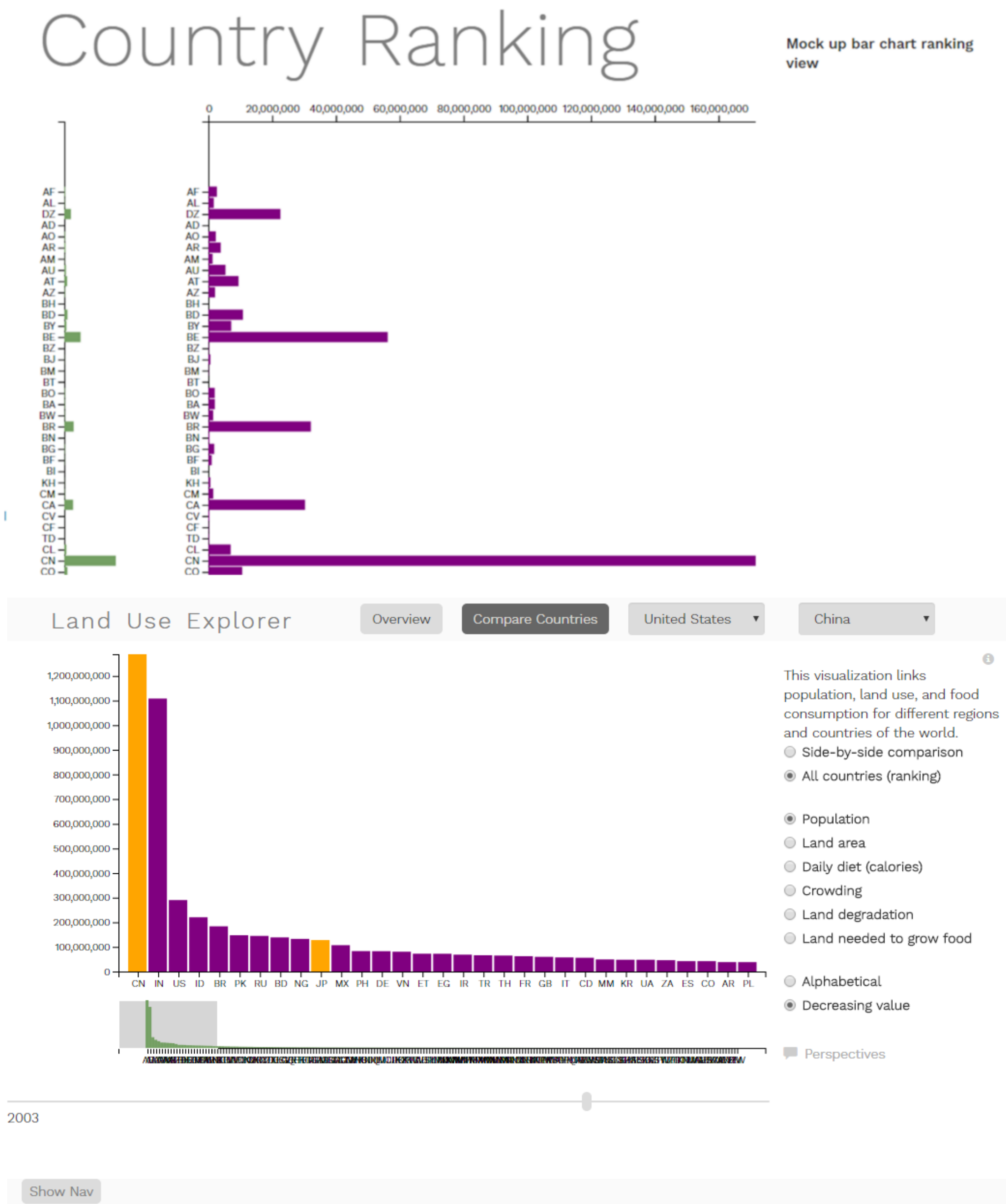


Figure 54. Testing bar chart orientations for the ranking page.

simple visual forms are hopefully more interesting and less intimidating to the average user than a series of bar charts would be, and the ability to interact with and filter the data will hopefully also help to inspire curiosity and further exploration.

The side-by-side view encourages users to compare the different countries and regions, and also allows comparison of two snapshots in time for the same country. The bar charts (coming soon) will help to give a more quantitative comparison of how the selected variables change over time, helping to identify trends that might be hard to read from the area encodings alone. I have not yet decided how I will handle the food import/export data in its longitudinal form; I could use a stacked bar chart, but I often find those hard to read and compare, especially when they're so small. The sketch shown in figure 52 works well for showing all of the land use data at once, but I'm not yet convinced that this will be a good solution for the food import data.

The ranking view helps people to compare a particular variable for all of the countries at once, and can also be used to get a "top 5, bottom 5" ranking. It's also interesting to highlight a country and watch its ranking change as the different variables are selected. In the final version, I plan to let users select two different countries and highlight them with different colors, so that they can compare their relationship across multiple rankings.

I doubt that most casual users will dig too deep into the comparison views, but I think it's important to allow people access to the data if their curiosity is piqued. The overview (in its final version) will give the quick takeaways that most users will want, and then the other views will be accessible to those who want to see more detail and make their own comparisons. This is the one section of the website where the user can interact with the data in a quantitative way, and I think it's important to be able to compare the different datasets side by side, especially since this collection of datasets reinforces the main narratives so well. I want this part of the website to facilitate active exploration of the data, and to encourage hypothesis-forming and testing, to help people investigate and develop their own understanding of the topic.

SECTION FOUR

Planning the Exhibition

Planning the exhibition piece is the final component of the thesis project. The primary challenge is to turn a detailed exploratory site into something that users can engage with in just a few minutes, while also taking advantage of the physical space to offer an in-person experience that couldn't be achieved online. The exhibit attendees will mostly be city people who never really see or think about soil, so I wanted to take this as an opportunity to help them experience soil on a physical level, as well as engaging in a conversation about the more abstract data side. Rather than attempting to convey the full content of the main website, I chose to focus instead on creating a memorable experience for the people who visit the gallery; one that will be intriguing enough to encourage them to visit the site on their own time.

I started by stepping back to the core messages that I hope people will take away from the exhibition. First, soil is interesting. It's not something that most people think much about, and just introducing the complexity of the soil ecosystem can begin to inspire the wonder that I want to lead this conversation.

Second, soil is important. I want people to understand that protecting the soil is in their best interest, and that problems with soil depletion affect us all. I also want people to realize that soil isn't some special thing that only shows up in gardens and nature preserves; that it's something we all interact with on a daily basis. It is, quite literally, the ground beneath our feet.

Figure 55. Plants growing in soil and soil substitute, in preparation for the thesis exhibition.

Third, I hope that people will see themselves in the exhibit. Awareness of global issues is important, but I also want viewers to consider where soil shows up in their own lives, and how they can protect the soil that they come in contact with in some small way, at a more local level. I believe that this self-identification is crucial to helping people stay engaged with the topic long-term.

If I could say only one of these things, I would choose to focus on the first. Creating a sense of wonder and curiosity is the most important thing that designers can do for science. Once a person has become curious, they can find additional information on their own, and they will be more motivated to do it. Without curiosity or interest, all of the facts and statistics in the world won't do much to change someone's mind. My goal for the exhibition was to find ways to convey the importance of soil in a visual way, while also taking advantage of physical space to provide viewers with a direct experience of the soil.

Just looking at a plant root system opens up an opportunity to start a conversation about what happens underground. In some ways, this relationship between soil and plants forms the real bridge to human concerns, but much of this relationship is underground and unseen. I decided to try using a clear, gelatin-based soil substitute as a way to make the root system visible and contrast it with the more normal view

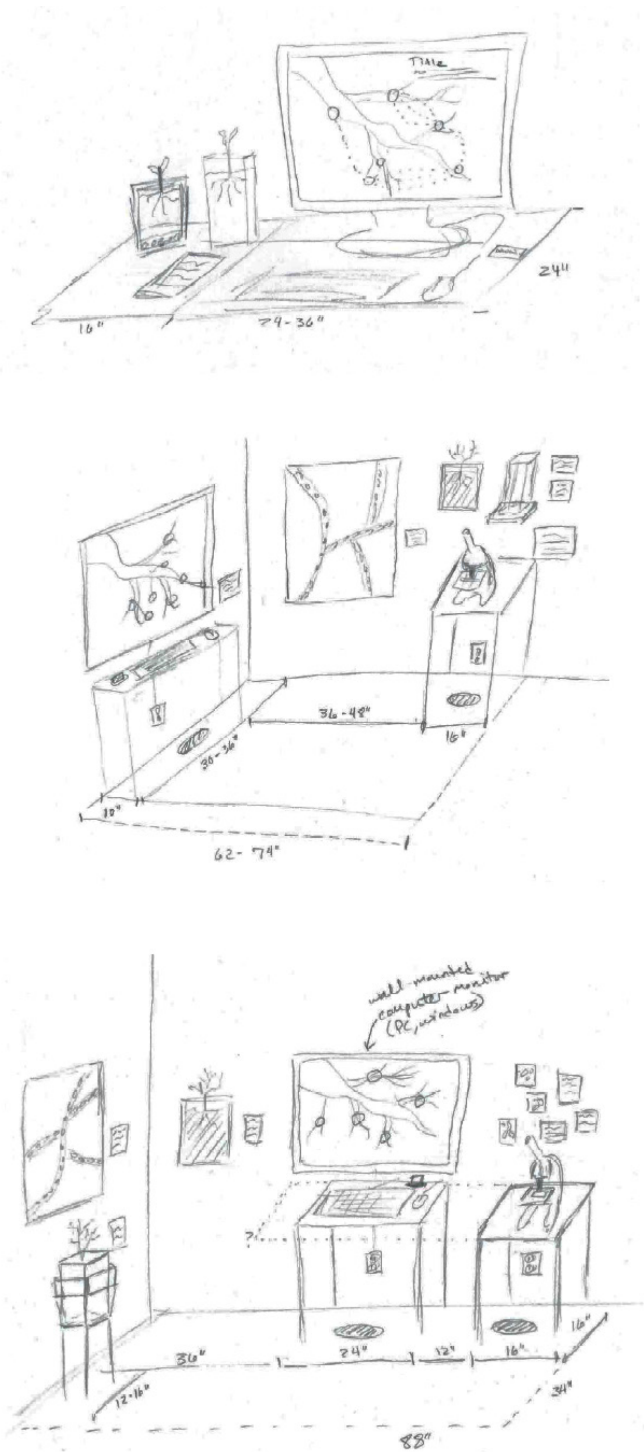


Figure 56. Early sketches of the exhibition layout.

of a plant in soil. I wanted both sets of roots to be visible, if possible, so I constructed a transparent container using a photo frame, and began rooting some plant cuttings in January, to give the roots time to develop before the exhibition in April (figure 55). Putting these two plants next to one another invites comparison, and poses the question of what soil does that we can't see, or, alternatively, what the world would look like without soil. The latter comes perilously close to using the threat narrative, depending on how the question is phrased. I wanted to avoid creating that emphasis, instead preferring to leave people with a sense of how amazing the soil is.

For our initial exhibition sketches (figure 56), we were asked to dream big, as if space and equipment were no issue. In that case, I thought I might consider having a microscope with soil samples available, to allow users to actually see some of the bustling life that isn't visible to us on a daily basis. This would emphasize the robust community that supports soil function, and could be a very memorable, "hands on" approach for people who may never have looked through a microscope before. A computer screen would show a video or webpage, as a teaser to encourage people to visit the website and explore more. I also considered adding photos or a poster with additional information to round out the experience, while also providing some breathing space between the long dwell-time displays (the microscope and the screen).

I had never tried looking at soil samples under the microscope, so I collected some samples and got access to a local microscope to see what I could find (figure 57). Sadly, it turns out that there isn't all that much going on in the soil in January. I was able to see soil particles, and lots of plant root hairs. At higher magnification, I was also able to make out what I'm pretty sure are protozoa swimming around (there's one in between the red lines on the length scale).

Bacteria would be even smaller than this, but I'd have expected to see other larger organisms as well, if the sample were taken at a more active time of year. Nematodes, tardigrades and rotifers are on the mm length scale, but I didn't find any of those in my samples. I'm not sure whether it's just because they migrate below the frost line in the winter, or whether I sampled in the wrong locations or prepared the microscope

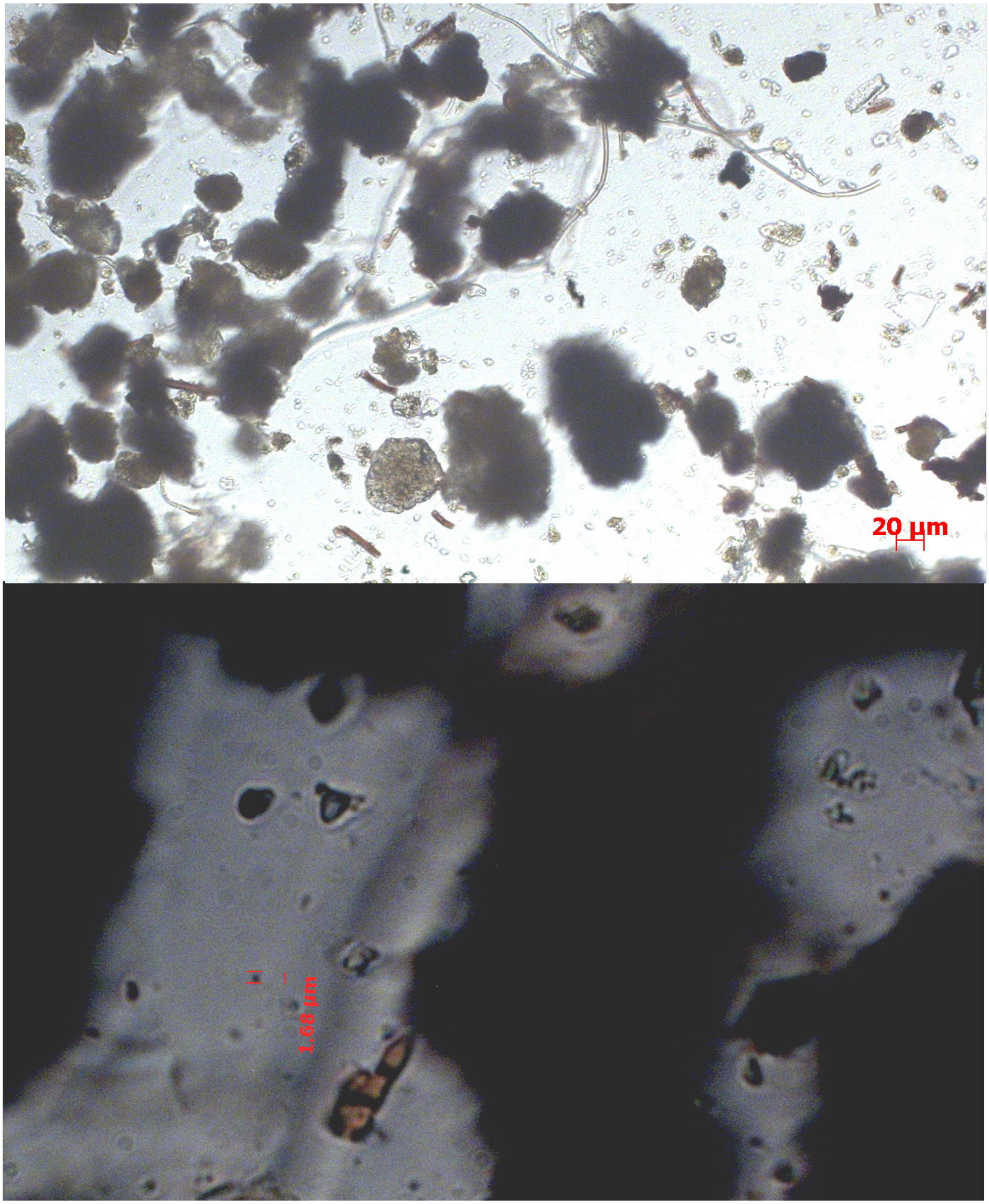


Figure 57. Microscope images of soil, showing soil particles and root hairs (top), and protozoa (bottom, between red lines)

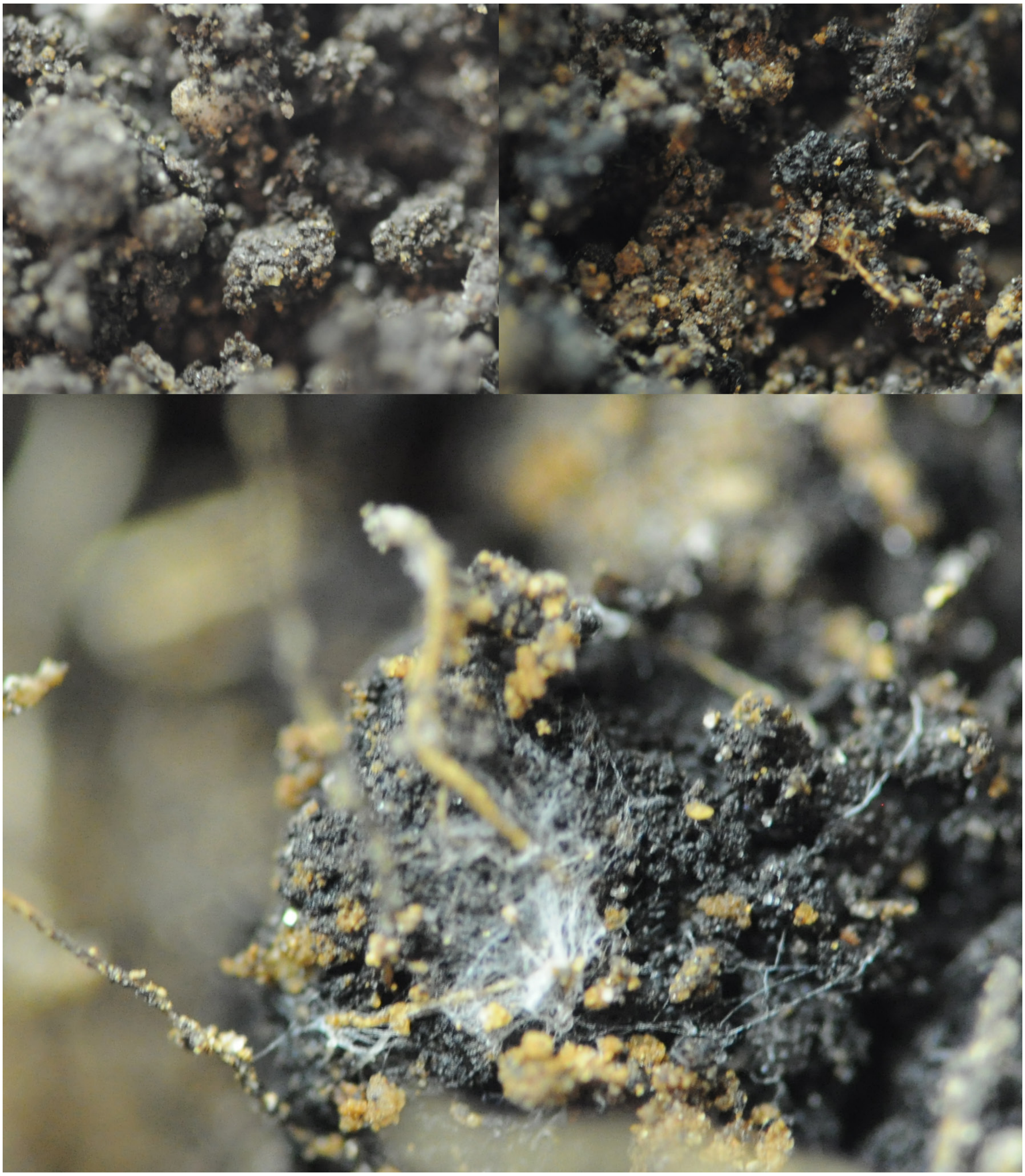


Figure 58. Comparison of a soil sample from my yard (top left) and one from Harold Parker State Forest. Mycorrhizal strands from soil fungi connecting different plant roots in the soil (bottom)

slides badly. Possible problems with sample prep, the unpredictability of working with living things, and the difficulty of obtaining a microscope for use at the exhibit made this a risky display. I considered replacing the live microscope with looping videos captured in advance, but this felt less connected to the physical experience that I wanted to create.

While preparing the microscope samples, I collected samples from my back yard garden, as well as a national forest where my husband and I go for walks. The contrast between the two soils was amazing, and readily visible even to people who don’t know a lot about soil. This suggested the possibility of collecting soil samples from different locations, and using that comparison to help visitors think about different qualities and types of soil.

Using a macro lens, I was able to capture some of the larger features of the soil microcosm; figure 58 shows a great example of mycorhizzal fungi growing around a plant root. It looks like a fine white spider web, made of tiny threads of fungi that allow plant roots to communicate and pass food to one another. These threads serve as “bacterial highways” in the soil, allowing bacteria to find and interact with one another.

I also experimented with other tangible ways that people could interact with soil in the exhibition. I took advantage the unseasonably warm weather in January to store a few soil samples from my garden in the freezer. I discovered during my soil autopsy last January that thawing soil is the smell of spring, and thought that experiencing that smell might help people connect with the soil on an emotional level as well. I suspect that most of the scent is released in the first gasp of life during the spring thaw, so I took 4 samples of hibernating soil, saved them in ziploc bags, and put them in the freezer to preserve a dose of spring air for the exhibition. I purchased coffee bags with one-way valves and a clear back panel that would allow many people to smell the different samples without losing the concentrated air or making a mess of the exhibition space.

I also prototyped a small moss terrarium that could be handed out at the exhibition, allowing viewers to take a small piece of the soil home with them. It consists of a bit of gravel, a teaspoon or so of soil, and a piece of terrarium moss in a plastic cosmetics container that can be screwed shut. Once the moisture is properly balanced, the terrarium is a closed ecosystem that can be sealed and should remain stable indefinitely, emphasizing the inherent balance of a soil ecosystem left to its own devices.

Participatory project

Comparing the soil samples from the forest and my back yard also inspired a new way to use the exhibition piece to support public dialogue around soil topics. First, I thought about personally collecting a hundred different soil samples from all over the state, to show the diversity of soils and the richness of the information that's all around us, if we have the interest and knowledge to understand it.

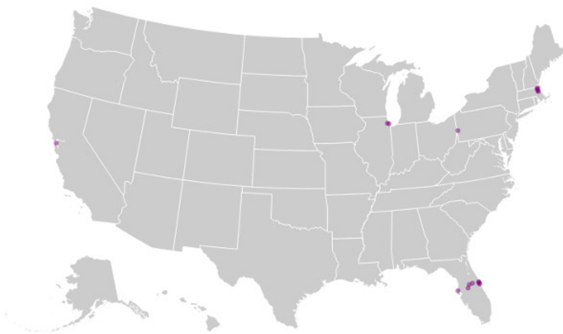
On reflection, I thought that it would be more interesting to compare soils from different geographic locations instead. I asked a few friends and family if they would be interested in mailing me a bag of dirt for the exhibition display. I was surprised at the number of people who wanted to participate, so I wrote up a simple submission form with instructions for collecting and mailing the samples and sent it out to anyone who was willing to dig up a sample from their back yard and mail it to me. In addition to a half a cup of soil from anywhere that seemed important to them, I asked participants to include a photograph of the location where it was collected, a title for the sample, and a brief story about the location. My husband and I posted a call for samples on Facebook, and within a few hours had more than 15 people who wanted to participate.

At this point, the question became how to manage all of this interest, and whether I could use it to generate excitement for both the topic and the project. I put together a simple web map to show where the samples are coming from, so that each person that contributes gets their own point on the map. This would both help people participating in the project to see how it is progressing, and might also be useful for filling in gaps in the map.

I also noticed an interesting thing happening in the comments of the original Facebook post. Two people from Ohio started talking about where they were from, so that they could make sure to get samples from different places. This is perhaps a simple thing, but it encouraged me

Figure 59. Moss terrarium refrigerator magnet and vented coffee bags for the exhibition.

Soil Sample Contributions



I asked a group of friends to send soil samples from all over the country for inclusion in my final thesis exhibition in April. These samples will be an important part of making the data in my project more personal, and will help show how soil touches all of our lives. This map shows where the samples have come from, and will be updated as they arrive. Click on a state to zoom in.

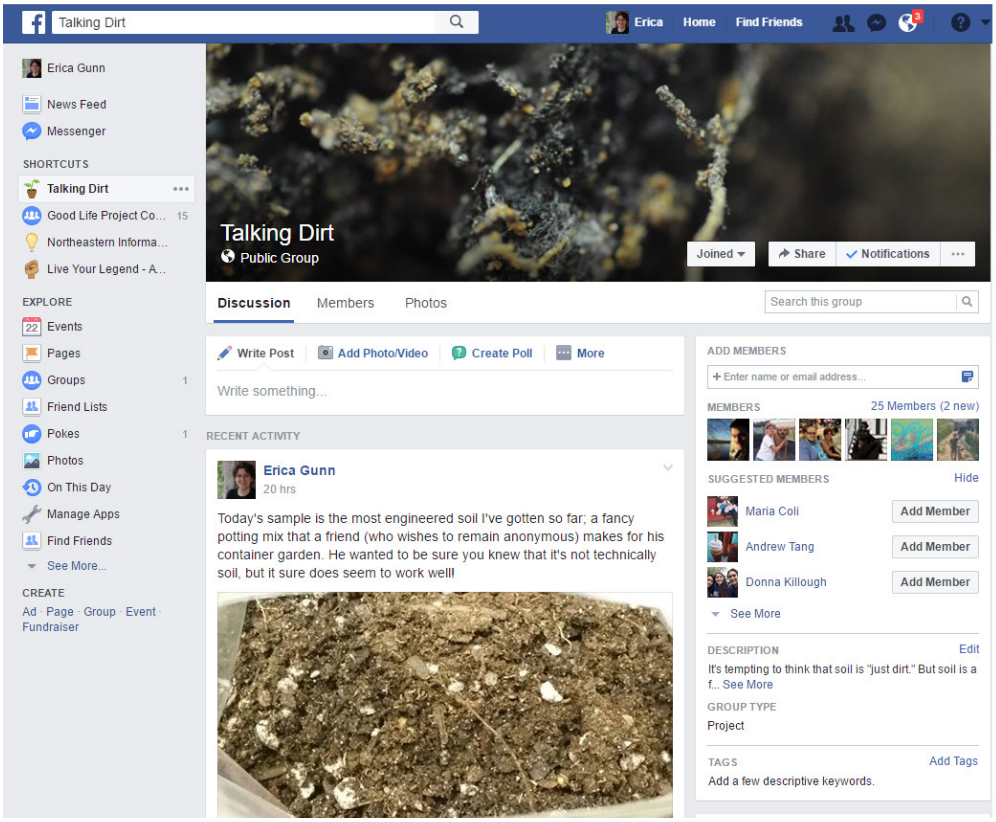


Figure 60. Map showing soil collection locations, and a Facebook group page to encourage participant conversation.

to find a way for all of these different people to talk to each other, and to post photos of their samples and stories as they send them. Rather than just mailing off a sample to me for an exhibition that they won't necessarily see, they would be able to share it with a growing community of people interested in the project—to see where other people are coming from, and how they experience the soil. Because it's web-based, it would also give my far-flung friends, relatives, and acquaintances a way to participate in the exhibition and the project, even if they can't be in Boston in April to stand in that physical space.

I created a Facebook group to host this conversation, post updates as samples arrive, and as a way to announce when the project website is released. Because social media posts are transient things, it is important to produce ongoing content to retain visibility and attract attention. Posting when samples arrived in my mailbox was an important acknowledgment of the people who contributed to the project, as well as a way to create opportunities for conversation. It also serves as an implicit reminder to people who join the group to actually send their samples. I also wanted to produce content that people could respond to, whether by liking or sharing posts, or by answering questions that could form the basis of a conversation.

To maintain interest while the group coalesced, it was also necessary to provide a variety of content. I had already collected many relevant quotes for use in my thesis book design, and created a simple page layout that I could use to share them in the group. I created a series of "did you know?" factoids, and began releasing a couple of these a week, along with articles related to the topic. The quotes have emotional resonance, and the factoids can be interesting, but too many of these might feel like telling rather than talking. Among these posts, I began to intersperse personal photos, stories from other participants, and questions about people's interactions with soil, in an attempt to get the conversation going.

The samples and stories that people contribute through this piece of the project will be incorporated into the exhibition piece, and eventually also curated on the thesis site as well. Early display prototypes are shown in figure 62. These samples will play an important role in the project by

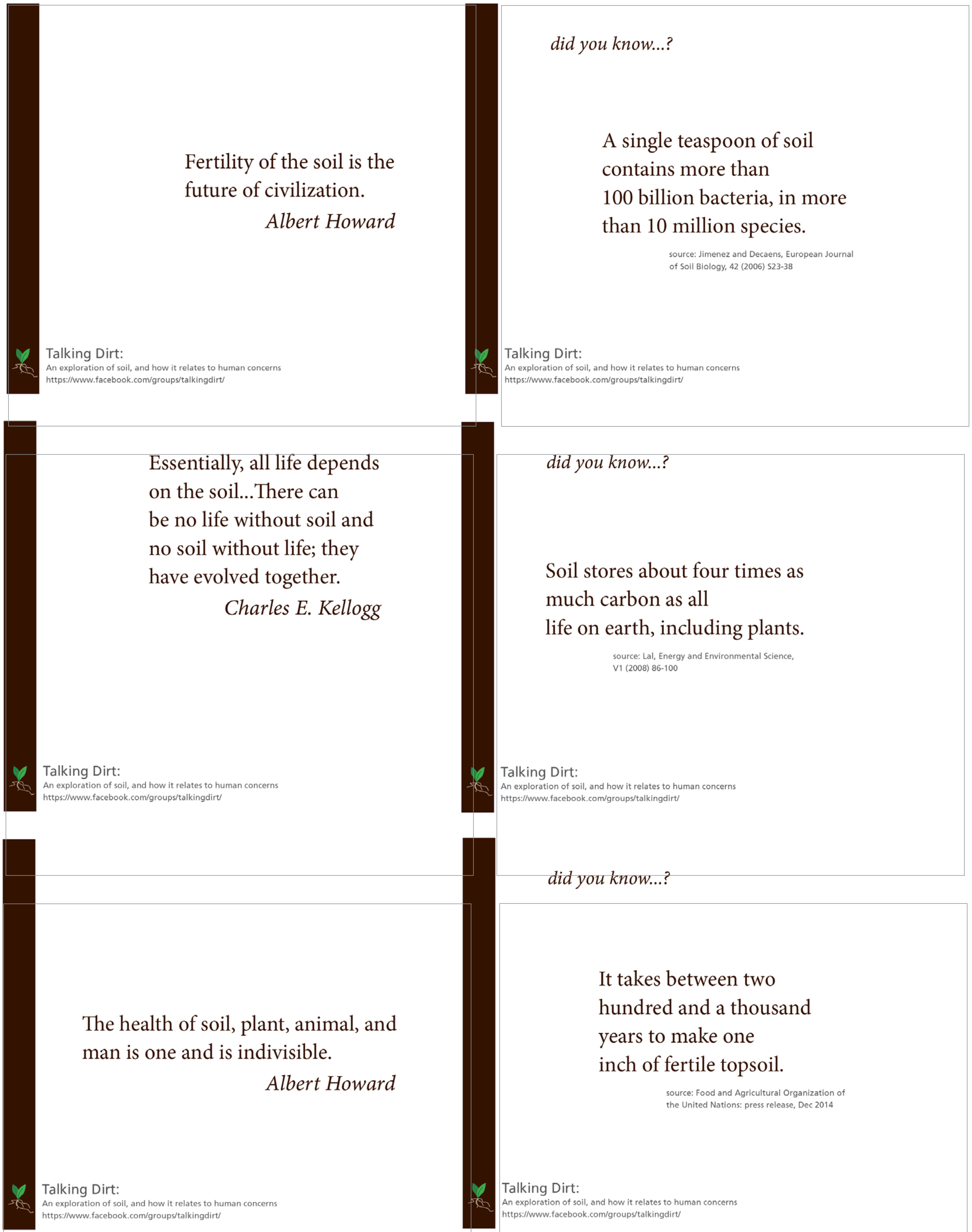


Figure 61. Selection of quotes and factoids for posting to the Facebook group.

bringing others' voices into the room. Most of the data on the project website focuses on the national or even global level; this individual sampling is a way to bring the local scale into the project, to help people connect those bigger ideas and statistical numbers with the soil they encounter in their own lives.

Showing soil samples from different places and with different information attached will help to show the heterogeneity of soil and highlight the variability of this particular ecosystem. The personal stories and sample titles will better reflect the full range of ways that humans interact with the soil. Taken together, these different exhibition pieces will help people to see everyday dirt as something worth thinking about.

Figure 62. Soil samples submitted for the exhibition project (above). Preliminary mockups of the exhibition display (right).


AT THE BOTTOM OF MISSION HILL
Zhengyan Yu
Boston, MA
I collected the soil on the sunny morning of 3/13 at the foot of Mission Hill. The place I collected the soil is also pretty near my apartment and the supermarket I usually go to. My apartment— which is my very first living place in United States —means the beginning of my dreams. That is why this place is special to me.
THE FARM
Anonymous
Reading, MA
This is a soil sample from our container garden. After several unsuccessful attempts to have a productive vegetable garden, we came up with the idea to move it to where the sun is in our yard—along the fence. To maximize efficiency of the space, we made it into a container garden and made use of the vertical space of the fence. The soil is a special mix used for container gardening. "The Farm" now supplies us with fresh vegetables and greens.

We abuse land because we regard it as
a commodity belonging to
us. When we see land as a community
to which we belong, we may
begin to use it with love and respect.

Aldo Leopold

SECTION FIVE

Reflections

As the project neared completion, I revisited the main components of the website to make sure that the narratives were balanced, the level of detail was where I wanted it, and the structural pieces fit together to make the experience that I want to create. This assessment focused mostly on the different levels of scale covered within the project. Several people have referred to Charles and Ray Eames' video Powers of Ten in reference to this project. It spans many different levels of scale, and it's important to help the user transition between them smoothly, and to understand what the different levels mean. To help myself assess the state of the project, I cut out screenshots from each of the website pages, and sorted them in different ways to show how the project as a whole is distributed.

Story

Soil itself spans a wide range of scales, and the project is built to reflect that aspect of its nature. Soil is important at a molecular level, because it contains the nutrients plants need. On a microscopic level, it is populated with a huge variety of microorganisms that make up 25% of all species on the planet, are central to helping soil function, and that may hold the key to new antibiotics, improvements in agriculture, information on disease, and a host of other things. At the macroscopic level, soil depends on local geology, is affected by its history of past use, and supports the plants that make the food we eat. Products from the soil are shipped globally, and it also has direct impacts on the quality of our air (through oxygen from plants, but also from methane and nitrous oxide release, among other things), and its role as a carbon storage mechanism makes it important for mitigating climate change as well. Each of the narratives that I chose to visualize walks through some subset of this story.

The data also has issues of scale associated with it. Most of the data I have is statistical information from the World Bank or FAO-UNESCO, which means that it's available at the global and sometimes regional and state levels, and only a very small amount is available at a finer degree of resolution. The scientific study data, on the other hand, tends to focus only on a few examples rather than broad analyses of the whole globe.

Detail

Sorting the different visualizations in terms of content depth, the project spans a pretty wide range. Starting out at the very top, I have general quotes and factoids for the Facebook group that are interesting but don't contain a lot of information about the system as a whole. Then there's the index page, which gives an overview and identifies the different narratives that the user can explore, and the topics within each one. The next level contains the bulk of the visualizations for the site. Because this project is aimed at a general audience who might not want to fully immerse themselves in the data, it's important that most of the visualizations are simple, and easy to read and understand.

A few of the visualizations go one step deeper, allowing the user to see a simple data story, but also to ask questions about where it applies (to which ecoregions or countries). The Land Use and Food Flow visualizations start out with a high level overview, and then allow users to break that information down even further to compare categories and look for patterns in the data. The Food Flow visualization takes the import and export data for a single country, and explores different partnership relations. These can be further filtered according to category, to see differences in grain exports, for example, vs. meat. I aggregated the 1000-plus categories available in the original data into 20 to simplify browsing, but there's still a lot of detail to be played with there. The

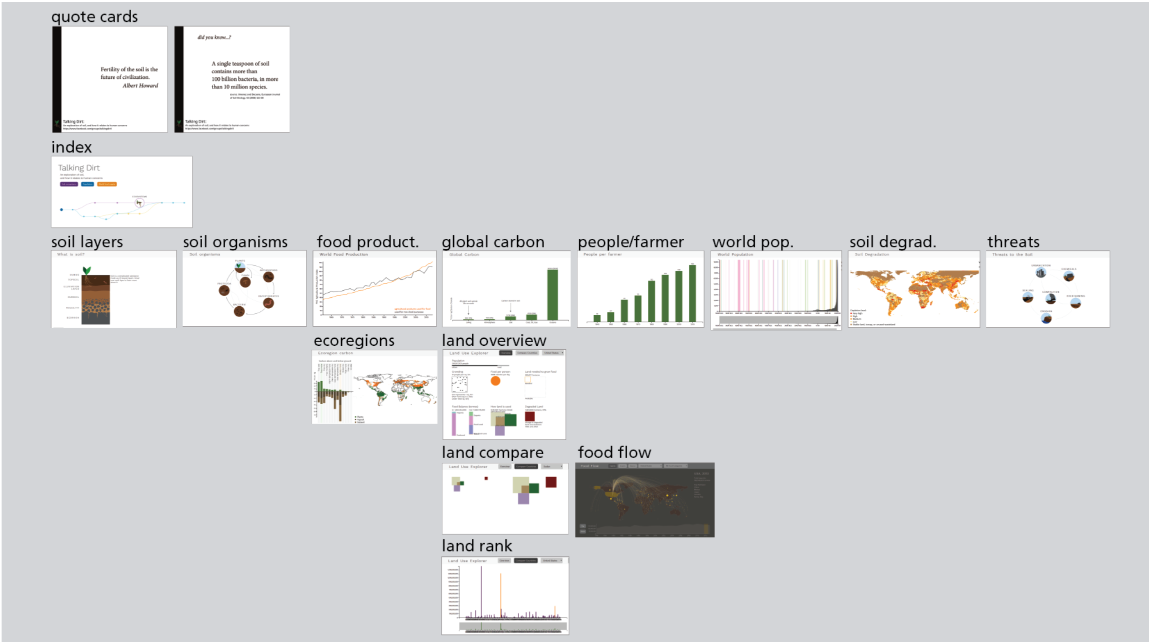


Figure 63. Assessing depth of content for different project components.

Land Use visualization goes the deepest in terms of making quantitative comparisons of multiple variables using different levels of aggregation. I expect that the majority of users will not spend the time to delve into these more advanced modes, but they are there for the people who are interested in more specific questions and more quantitative comparisons.

Comparing narratives

I was also curious about how the different narratives compared in terms of detail. The soil narrative stays relatively high-level, partly because there isn't a lot of detailed data available, and partly because introducing any detail at all requires introducing so much. It's easy to understand that the soil community is made up of different kinds of microbes; it's quite another thing to start talking about which ones are found in which layers of the soil, how they vary with moisture content, how microbial populations change along the length of a particular kind of plant root. In terms of information available, there isn't much middle ground between these two extremes. I chose to stay on the high-level view because of my audience, though my own personal fascination leans deep into the detail of this section.

The one exception that goes a tiny bit deeper into the soil data is the Ecoregions visualization, which shows data from a scientific paper that talks about the distribution of global carbon reserves, and links it to the World Wildlife Fund's designated ecoregions, to give users some

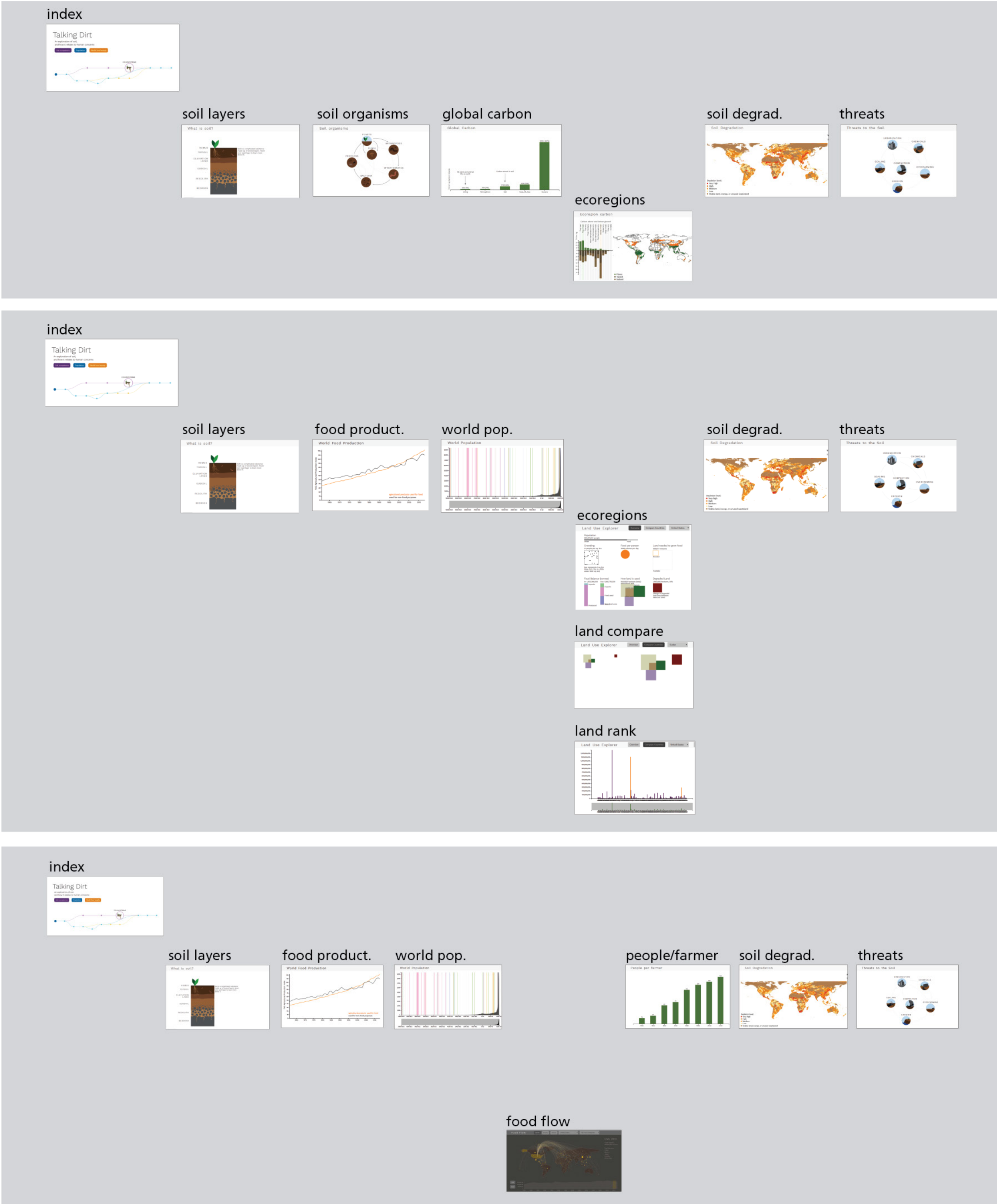


Figure 64. Comparing level of detail for different narrative choices within the project website.

sense of where soils with those different carbon contents might show up in the world. This additional level of detail and interactivity moves that visualization a little deeper into the information space.

The population narrative also stays mostly at a high level, and shares many of the visualizations from the soil narrative as well. It delves deep into population, land use, and food supply in the Land Use panel, but a user could easily choose to stay on the overview page of that module and ignore those deeper layers of detail.

The food narrative follows along with the population narrative quite closely, but jumps down to the Food Flow data instead of focusing on land use. Food Flow really could populate either the third or the fourth level, depending on whether the user chooses to get into individual data categories. Since the option to go deeper was available, I chose to rank it on the lower of the two levels.

Global vs local

Grouping the visualizations by their aggregation level makes the limited range of the data clear pretty quickly: the top row contains data that is aggregated to a global level or general information that is not site-specific. The lower row contains the visualizations whose data is grouped by country. The Ecoregions and Soil Degradation maps go to a slightly less aggregated level—at least for the bigger countries—because they ignore country boundaries and focus on natural features instead.

Personal/Impersonal

I also ranked the different components of the project in terms of their personal/impersonal nature. The quotes and factoids are at the top, and create some amount of interest but probably have a fairly small personal impact. Next come the global statistics, and then the country-level statistics, where people begin to see distinctions between their situation and that of others around the world. Next, I put in the map with all the soil sampling locations that I'm collecting for the exhibition; this piece of the project begins to look at the heterogeneity of soil within a single country, and even within a single state. It also begins to incorporate the personal and the local, which is the part that has the most relevance and emotional resonance for most people. At the very bottom, core to the dialogue part of my thesis, is the Facebook group, where people can interact with one another, tell stories, and learn more about this resource that we all should treasure.

Comparing against recommendations

I also felt that it was important to explicitly assess the final project against the list of recommendations generated in the related work section. It's not possible (or even desirable) for every component of the site to meet every objective, but I wanted to get a sense of whether those different goals were being met, where, and how.

Tone

1) Use wonder, aesthetic interest, and curiosity to encourage engagement and reduce overwhelm.

This recommendation is mostly reflected in the choice of framing for the project narratives. I chose to lead with an appreciation for the soil as an ecosystem and source of food rather than focusing first on the threat of soil degradation, to reduce the feeling of overwhelming problems that such ecological narratives can create. Allowing the user to interact with different pieces of the website and uncover information in small pieces can also help to make its absorption more manageable. Revealing an interesting piece of information in one area will also encourage users to explore more. Providing deep access to the data also helps to satiate curiosity once it is created, allowing people to pose questions and look to see whether the information matches what they would expect. The site narratives start out fairly shallow and gradually become deeper, and also allow the user to back out to a less detailed level if they are feeling overwhelmed.

Structure

2) Provide narratives to guide the user
3) Support undirected exploration, use modular components, make multiple cross-references
4) Show relationships and use transitions to preserve context

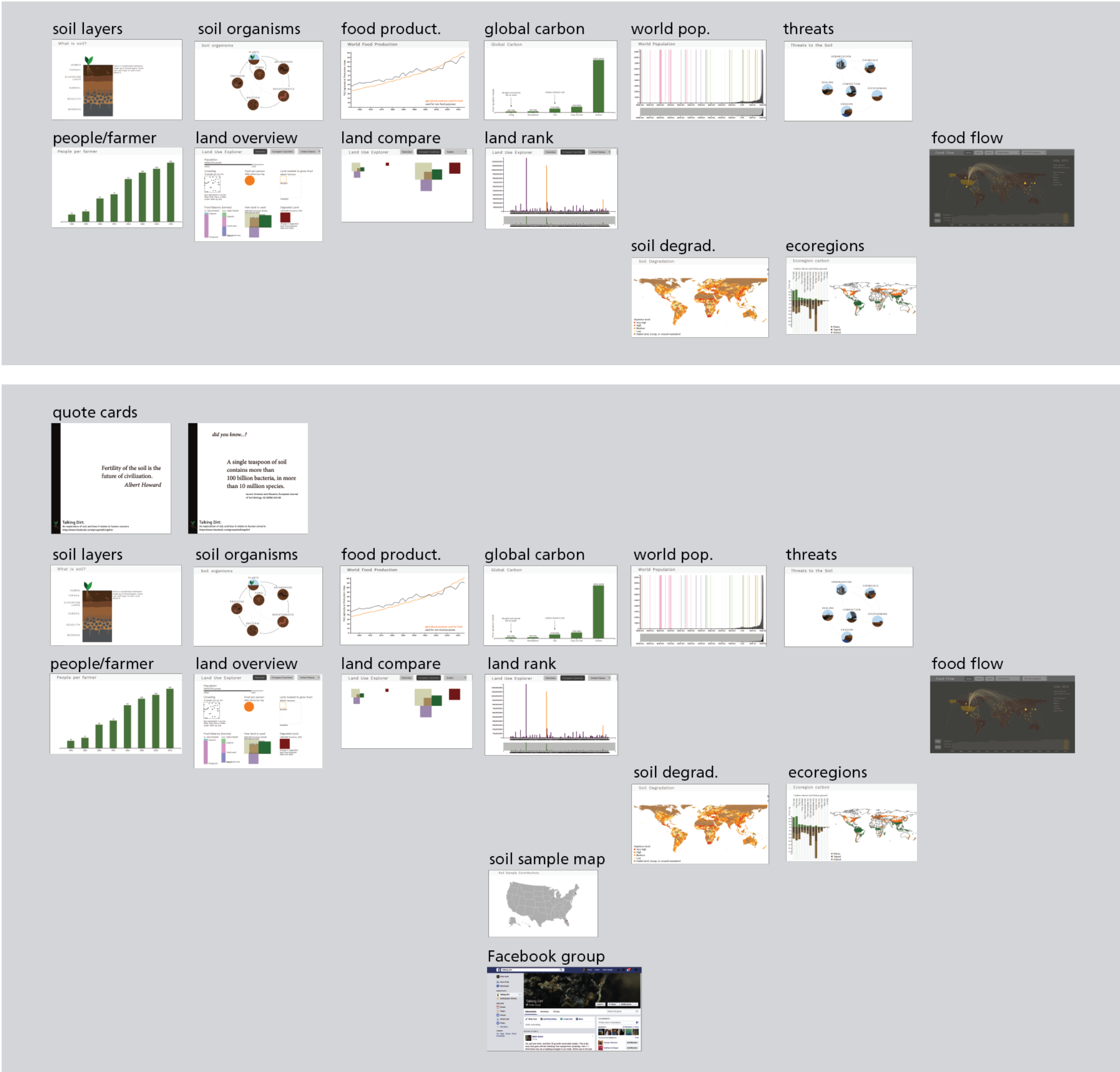


Figure 65. Comparing data aggregation level and level of personal relevance for different site components.

The site browsing interface is the primary feature that provides narratives to guide the user. Each page has a smaller sub-story in the sidebar that helps to reinforce this global narrative, but the navigation panel is primarily responsible for helping the user to experience the different pieces of the story in a particular sequence. The index page provides a broad overview of the system, and allows people to preview the title and visual form of the information that they will see in a given node. The similarity in design between the index page and the navigation panel at the bottom of the site pages helps to maintain a sense of context, as well as showing progress and the options available.

Each of the individual visualizations is designed to be modular, so that they can be read in any order without significant loss of information. This is particularly important in a site with so many different options; it is unlikely that a user will explore all of the different pages, so it is important that each individual page is also able to tell a story on its own. The code architecture currently tracks which system nodes the user has visited throughout their interaction, so it is also possible to adjust the page narratives presented in the sidebar to accommodate different approaches to the information shown. This is not yet implemented, but will be added to the final site if time allows.

In a future version, it would also be interesting to explore the possibility of showing connections between the current page and other related stories in the navigation panel, to emphasize the interconnections between the different nodes and suggest alternate paths for exploration.

Agency

5) Provide opportunities to switch between narratives. Allow users to preview selections.
6) Indicate user progress/history

User agency is also managed primarily by the navigation panel. It is possible for a user to select and follow a single narrative, or to jump around at random between the different nodes. Users can preview the title of the different options available to them, to get a sense of whether or not they are interested in a particular topic. Earlier versions of the index page also provided a brief overview of each node, to give the user more information about what it contained. This may be re-introduced in the final version, but is not currently present in the design. The interface also tracks where users have already been, and indicates their progress along a particular narrative, if one is selected.

Content

7) Clearly separate data from interpretation
8) Explain limits of certainty and compare alternate interpretations; may be supported by a scenario-setting or experimental approach

The site currently shows a good amount of data and summarizes it clearly, but does not provide as much context or detail in the discussion as I would like. I have added information buttons on each page that will discuss the sources of the data shown, along with any limitations or discrepancies that I have found. I would like the site to support more in-depth discussion of these aspects, but found it difficult to fit on a small page without compromising the site flow.

I considered allowing the page to scroll and providing additional details below the main graphic, but felt that a scrolling page paradigm fought the modular experience that I was hoping to accomplish. The single-view setup helps to keep viewer's attention close to the central narrative rather than pulling them off into discussions of the data details, but it also necessitates a certain degree of simplification. As a result, many of the pages in the site do not provide significant interpretation beyond describing what is represented in the data; adding this deeper narrative component would help to ensure that viewers know what to take away from the visualizations that they see on the page. This is an area that needs further development.

For an audience with a high degree of scientific or data literacy, including greater detail about uncertainty and varying interpretations of the data would have been a top priority. Because I do not expect that the majority of my audience will be technically inclined, I chose to focus on the pieces that I felt would be most relevant to the core group for now. In future, I would like to expand in this area, perhaps using linked blog posts, a behind the scenes view, or some other format where I could deconstruct what's really going on in the data for each individual piece.

Conversation

9) Create tangible records of interaction
10) Support action; incorporate the personal/local, allow user to contribute to information shown

As of this writing, the project site does not create any record of interaction visible to the user, beyond that shown in the navigation panel. I have included a perspectives button on each page, which shows the user two different perspectives on the data that they just saw, and allows them to agree or disagree with each. This is currently not stored on the server or visualized in any way, but that level of interaction could be implemented in the future. This could also be a way of showing interconnections between topics; if a user is not sure how to vote on a particular perspective, these summaries could include links to a longer discussion where they could learn more. Adding in commenting functionality or linking to a forum page would also encourage a deeper level of dialogue, rather than a binary up or down vote. The choice of perspectives will need to be made carefully, as it could heavily bias the visualization or create a sense of binary opposition.

The project site does not yet provide local information, and does not provide users with the ability to contribute information to the project. I plan to include a list of links and actionable items in the final page of the site, but this has not been implemented at this point. The participatory piece of the exhibition project is an attempt to bring personal stories into the project, and makes a first step toward incorporating a local, personal component, but there is still a long way to go in this regard.

Figure 65. Testing out user interaction and response gathering for future versions of the site.

If facts are the seeds that later
produce knowledge and
wisdom, then the emotions and
the impressions of the
senses are the fertile soil in which
the seeds must grow.

Rachel Carson

CHAPTER SEVEN

Conclusions and Future Work

Design can contribute to science communication by building systems for viewing multiple interrelated datasets. Future work on this project includes increasing the content depth, reinforcing connections between datasets, and creating a visual overview of the entire system to facilitate conversation and comparison of different views related to the topic of soil health.

This thesis demonstrates an approach to building data visualizations that communicate science to a general audience, and identifies several guidelines for the design of such systems. The project implementation combines information from a variety of sources and a range of different topics related to soil health and human land use to create a broad overview of the relationships between ecological systems and human activity. The information presented spans a range of scales, and provides users the opportunity to interact at different levels of detail with the data. Components are related but modular, and maintain maximum user agency while navigating through the site and exploring the information presented. The navigation interface and page templates are built to accommodate future expansion, including support for presenting alternate interpretations of the data and tracking user interactions and opinions on the different visualization pages. In the future, this information could be used to produce a personalized record of a user's journey through the site, and their response to the different components. These interaction records could be used as a way for users to identify similarities and differences of opinion, and could form a starting point for ongoing discussions about controversial issues. The exhibition piece allowed people to participate in the process of co-creation, and made their stories and experiences a part of the greater soil narrative. In addition to helping individuals to recognize where soil shows up in their own lives, this also helped to build excitement about the project, and seeded a community where some of these conversations could take place.

Design has a critical role to play in communicating scientific information to a lay audience. Good design can help users to navigate interdependence by creating clear structure and sense of context, as well as showing connections within and between related datasets. By inspiring a sense of wonder and aesthetic attraction, designers can encourage people to explore data that they might otherwise find intimidating or uninteresting. Integrated visualization networks can help users to create a synthetic overview of interdependent systems by showing connections between related ideas and supporting undirected exploration, and can inspire a deeper appreciation of complex topics.

Future work

There are many opportunities to improve on this project in the future. There are several modules where it would be helpful to provide additional context, whether through detailed notes in the information section, articles discussing the quality and limitations of the data, or auxiliary visualizations that take the analysis to a deeper level. Making the quantitative analysis more robust would make this a stronger educational tool, but would also require more prolonged engagement and would appeal primarily to a more technical user, and so was placed as a second-level priority in the initial site development. In future, it would be helpful to revisit these visualizations and provide closer investigations into the original data to enrich the scientific content of the site.

In many cases, a stronger sense of context for an individual visualization could be created through comparison with other datasets or sources of information. The visualization connecting soil carbon storage to global ecoregions is a good example of how different datasets can reinforce one another, as well as the difficulty of merging disparate data. Expanding this kind of comparison to other visualizations in the system would help to reinforce the connections between modules. For example, it would be interesting to compare the quantity of carbon stored in different forms globally with the amount that has been released by human activity, to better link soil carbon storage to climate change mitigation. Linking the ecoregions data to areas with the highest rate of deforestation or other human activity could also provide a stronger connection to the main topic. The soil degradation map could be connected to global rainfall patterns, population density, and other metrics.

It would be helpful to reinforce connections between visualizations more strongly as well. This could be accomplished by drawing additional links in the navigation panel, relating the currently-selected node with others in the system. It could also mean incorporating land use data into the food flow visualization more directly, and vice versa, or developing additional visualizations that compare the different representations side by side. For the first stage of development, it was important to create modules that were independent and visually distinct. For future development, it would be interesting to show the connections between these different views, to reinforce the system interdependence and allow the user to see how changes in one variable cause changes in another, like the p-hacking graphic shown in figure 7. This more predictive aspect might require modeling or projection from the base datasets, and so is a more complicated undertaking.

Incorporating contextual comparisons into the modules is a first step toward showing interdependence, but it would also be helpful to create an aesthetically-driven representation of the entire system with the flavor of a generative or abstract art piece. Ideally, this aesthetic representation would also form the core visual identity of the site, and would be incorporated into the page overview and navigation panels at a minimum, and possibly into the rest of the site as well. This visual overview would create interest and user appeal, and could also form the basis for personalized visualizations of the information space.

These representations would be helpful for facilitating conversation and encouraging people to compare their different perspectives in a synthetic, global way. Rather than getting bogged down immediately in the specifics of one viewpoint or another, people would have opportunities to compare which views they share and where they differ, making for a more nuanced conversation. I find the metaphor of roots to be a particularly compelling one for this project, but more abstract representations could also work, provided that the user is still able to identify specific ideas and opinions within the resulting graphic.

Finally, it is also important to consider how this project could become the basis of a community. A personalized visual overview of the system would be especially helpful for discussions in a focus group or other in-person setting, but it could also be used online to establish a point of discussion and build community. This community could take many forms, but would ideally connect people with different viewpoints, identify areas for further exploration, provide recommendations of resources and opportunities to participate in soil issues, and establish direct connections between experts and laypeople, either through ongoing publications, seminars, or individual discussion. It would also incorporate personal stories and experiences into the main site, beginning with the samples contributed to the exhibition project. It could also provide information and opportunities to take local action, whether through advertising citizen science projects related to the soil, creating local data initiatives, or by serving as a hub for community activism. This kind of organizing would have to be built from the bottom up, based on the needs and interests identified by users of the site. The visualizations would provide an attractive starting place and a shared experience from which to discuss these issues, and the community portion of the site would provide a space in which those conversations could take place.

Life doesn't make any sense without interdependence. We need each other, and the sooner we realize that, the better for us all.

Erik Erikson

References

Angel, S., J. Parent, D. L. Civco, A. Blei, and D. Potere. 2011. "The Dimensions of Global Urban Expansion: Estimates and Projections for All Countries, 2000–2050." *Progress in Planning, The dimensions of global urban expansion: Estimates and projections for all countries*, 2000–2050, 75 (2): 53–107. doi:10.1016/j.progress.2011.04.001.

Anonymous, "Wealth Inequality in America," accessed online Mar 2017. https://www.youtube.com/watch?v=QPKKQnijnsM.

Bai, Z. G., D. L. Dent, L. Olsson, and M. E. Schaepman. 2008. "Proxy Global Assessment of Land Degradation." *Soil Use and Management* 24 (3): 223–34. doi:10.1111/j.1475-2743.2008.00169.x.

Beccario, C. "Earth," accessed online Mar 2017. https://earth.nullschool.net.

Brath, R., M. Peters, and R. Senior. 2005. "Visualization for Communication: The Importance of Aesthetic Sizzle." In *Ninth International Conference on Information Visualisation* (IV'05), 724–29. doi:10.1109/IV.2005.145.

Brink, S. C. 2016. "Unlocking the Secrets of the Rhizosphere." *Trends in Plant Science* 21 (3): 169–70. doi:10.1016/j.tplants.2016.01.020.

Corbett, J. B., and J. L. Durfee. 2004. "Testing Public (Un)Certainty of Science: Media Representations of Global Warming." *Science Communication* 26 (2): 129–51. doi:10.1177/1075547004270234.

Decaëns, T., J. J. Jiménez, C. Gioia, G. J. Measey, and P. Lavelle. 2006. "The Values of Soil Animals for Conservation Biology." *European Journal of Soil Biology, ICSZSoil Animals and Ecosystems ServicesProceedings of the XIVth International Colloquium on Soil Biology,* 42, Supplement 1 (November): S23–38. doi:10.1016/j.ejsobi.2006.07.001.

Drucker, J.. 2016. *Graphesis: Visual Forms of Knowledge Production.* Cambridge: Harvard University Press. http://www.hup.harvard.edu/catalog.php?isbn=9780674724938.

Estrada, F. C. R., and L. S. Davis. 2015. "Improving Visual Communication of Science Through the Incorporation of Graphic Design Theories and Practices Into Science Communication." *Science Communication* 37 (1): 140–48. doi:10.1177/1075547014562914.

FAO-UNESCO, org. 1972. Food and Agricultural Organization of the United Nations Educational, Scientific, and Cultural Organization "Soil Map of the World," 1972, accessed online March 2016, http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/faounesco-soil-map-of-the-world/en/

FAO, org. 2012. Freddy Nachtergaele, H. T. van Velthuizen, L. Verelst, D. Wiberg, Niels Batjes, J. A. Dijkshoorn, et al. "Harmonized World Soil Database (Version 1.2)."

FAO-UNESCO, org. 2016. "Status of the World's Soil Resources. Technical Summary," accessed online February 2016, http://www.fao.org/documents/card/en/c/39bc9f2b-7493-4ab6-b024-feeaf49d4d01.

FAO, org. 2016. "FAOSTAT Emissions Database." Food and Agricultural Organization of the United Nations. http://faostat3.fao.org/home/E.

Fathom Information Design, "What the World Eats," accessed online Mar 2017. http://www.nationalgeographic.com/what-the-world-eats/.

Five Thirty Eight, "Science Isn't Broken," accessed online Mar 2017. https://fivethirtyeight.com/features/science-isnt-broken/#part1.

Friendly, M. 2008. "A Brief History of Data Visualization." In *Handbook of Data Visualization,* 15–56. Springer Handbooks Comp.Statistics. Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-540-33037-0_2.

Foley, J. A., R. DeFries, G. P. Asner, C. Barford, G. Bonan, S. R. Carpenter, F. S. Chapin, et al. 2005. "Global Consequences of Land Use." *Science* 309 (5734): 570–74. doi:10.1126/science.1111772.

Frankel, F. 2002. *Envisioning Science.* The MIT Press, Cambridge MA.

Gershon, N., S. G. Eick, and S. Card. 1998. "Information Visualization." *Interactions* 5 (2): 9–15. doi:10.1145/274430.274432.

Gibbs, H. K., and J. M. Salmon. 2015. "Mapping the World's Degraded Lands." *Applied Geography* 57 (February): 12–21. doi:10.1016/j.apgeog.2014.11.024.

Halloran, N. "The Fallen of World War II," accessed online Mar 2017. http://www.fallen.io/ww2/.

Johnson, J., and A. Henderson. 2011. *Conceptual Models: Core to Good Design.* Morgan & Claypool Publishers.

Jouquet, P., J. Dauber, J. Lagerlöf, P. Lavelle, and M. Lepage. 2006. "Soil Invertebrates as Ecosystem Engineers: Intended and Accidental Effects on Soil and Feedback Loops." *Applied Soil Ecology* 32 (2): 153–64. doi:10.1016/j.apsoil.2005.07.004.

Katz, Y. 2013. "Against Storytelling of Scientific Results." *Nature Methods* 10 (11): 1045–1045. doi:10.1038/nmeth.2699.

Kitchin, R. 2014. "Conceptualizing Data." In *The Data Revolution.* SAGE Publications, Inc. https://thedatarevolutionbook.wordpress.com/.

Kögel-Knabner, I. 2002. "The Macromolecular Organic Composition of Plant and Microbial Residues as Inputs to Soil Organic Matter." *Soil Biology and Biochemistry* 34 (2): 139–162.

Krzywinski, M, and A. Cairo. 2013. "Points of View: Storytelling." *Nature Methods* 10 (8): 687–687. doi:10.1038/nmeth.2571.

Lal, R. 2008. "Sequestration of Atmospheric CO2 in Global Carbon Pools." *Energy & Environmental Science* 1 (1): 86–100. doi:10.1039/B809492F.

Lavelle, P., T. Decaëns, M. Aubert, S. Barot, M. Blouin, F. Bureau, P. Margerie, P. Mora, and J. -P. Rossi. 2006. "Soil Invertebrates and Ecosystem Services." *European Journal of Soil Biology, ICSZSoil Animals and Ecosystems Services, Proceedings of the XIVth International Colloquium on Soil Biology,* 42, Supplement 1 (November): S3–15. doi:10.1016/j.ejsobi.2006.10.002.

Lepers, E., E. F. Lambin, A. C. Janetos, R. DeFries, F. Achard, N. Ramankutty, and R. J. Scholes. 2005. "A Synthesis of Information on Rapid Land-Cover Change for the Period 1981–2000." *BioScience* 55 (2): 115–24. doi:10.1641/0006-3568(2005)055[0115:ASOIOR]2.0.CO;2.

Lowe, R. K. 2003. "Animation and Learning: Selective Processing of Information in Dynamic Graphics." *Learning and Instruction, External and Internal Representations in Multimedia Learning*, 13 (2): 157–76. doi:10.1016/S0959-4752(02)00018-X.

Lowe, R. 2004. "Interrogation of a Dynamic Visualization during Learning." *Learning and Instruction, Dynamic Visualizations and Learning,* 14 (3): 257–74. doi:10.1016/j.learninstruc.2004.06.003.

Ma, K. L., I. Liao, J. Frazier, H. Hauser, and H. N. Kostis. 2012. "Scientific Storytelling Using Visualization." *IEEE Computer Graphics and Applications* 32 (1): 12–19. doi:10.1109/MCG.2012.24.

Mueller, L., U. Schindler, W. Mirschel, T. G. Shepherd, B. C. Ball, K. Helming, J. Rogasik, F. Eulenstein, and H. Wiggering. 2010. "Assessing the Productivity Function of Soils. A Review." *Agronomy for Sustainable Development* 30 (3): 601–14. doi:10.1051/agro/2009057.

Niepold, F., D. Herring, and D. McConville. 2007. "The Case for Climate Literacy in the 21st Century." *Fifth International Symposium on Digital Earth*. UC Berkeley, CA, USA. http://rose.geog.mcgill.ca/geoide/files/geoide/CaseForClimateLiteracy.pdf.

Nisbet, M. C., and D. A. Scheufele. 2009. "What's next for Science Communication? Promising Directions and Lingering Distractions." *American Journal of Botany* 96 (10): 1767–78. doi:10.3732/ajb.0900041.

Perry, M., "In response to the viral 'Wealth Inequality in America' video" *AEI Ideas*, published online Mar 2013, accessed December 03, 2016, https://www.aei.org/publication/in-response-to-the-viral-wealth-inequality-in-america-video/

Periscopic, "US Gun Deaths," accessed online Mar 2017. http://guns.periscopic.com/.

Pollan, M., 2015. "A Secret Weapon to Fight Climate Change: Dirt," *The Washington Post*, published Dec. 15, accessed online February 28, 2016, http://michaelpollan.com/articles-archive/a-secret-weapon-to-fight-climate-change-dirt/.

Pousman, Z., J. Stasko, and M. Mateas. 2007. "Casual Information Visualization: Depictions of Data in Everyday Life." *IEEE Transactions on Visualization and Computer Graphics* 13 (6): 1145–52.

Ramankutty, N. and J. A. Foley, *Global Biogeochemical Cycles*, vol. 13, no. 4 (1999): 997-1027. Data accessed online from the Nelson Institute Center for Sustainability and the Global Environment Mar 2016. www.nelson.wisc.edu.

Rees, K. 2015. Tapestry Conference, Georgia, USA. Published online Mar 13, 2015. Video accessed on YouTube Dec 6, 2016. https://www.youtube.com/watch?v=0VGARONcLVs.

Roser, M. "Our World in Data," accessed online Mar 2017. https://ourworldindata.org/.

Rittel, H. W. J. 1972. "On the Planning Crisis: Systems Analysis of the 'First and Second Generations.'" *Bedriftsøkonomen* 8 (0): 390–96.

Sanchez, P. A., S. Ahamed, F. Carré, A. E. Hartemink, J. Hempel, J. Huising, P. Lagacherie, et al. 2009. "Digital Soil Map of the World." *Science* 325 (5941): 680–681.

Scharlemann, J. P.W., E.V.J. Tanner, R. Hiederer, and V. Kapos. 2014. "Global Soil Carbon: Understanding and Managing the Largest Terrestrial Carbon Pool." *Carbon Management* 5 (1): 81–91.

Seto, K., B. Guneralp, and L. Hutyra. 2012. "Global Forecasts of Urban Expansion to 2030 and Direct Impacts on Biodiversity and Caron Pools." *Proceedings of the National Academy of Sciences* 109 (40): 16083–88. doi:10.1073/pnas.1211658109.

Stefaner, M. OECD Better Life Index, accessed online Mar 2017. http://www.oecdbetterlifeindex.org/.

Stefaner, M. Project Ukko, accessed online Mar 2017. http://project-ukko.net/.

Thistle, M.W. 1958. "Popularizing Science." *Science* Vol 127, No 3304, pp 951-955.

UNEP/ISRIC, 1992. World atlas of desertification, United Nations Environment Programme/International Soil Reference and Information Centre, Edited by Middleton et al., London. Data accessed online through the ArcGIS database Mar 2016.

World Bank, 2013. "World Development Indicators — Data," accessed online Mar, 2016, http://data.worldbank.org/data-catalog/world-development-indicators.

Worstall, T., 2013. "The Problem With The Wealth Inequality In America Video. It Commits Worstall's Fallacy." Forbes.com. Published online March 9. http://www.forbes.com/sites/timworstall/2013/03/09/the-problem-with-the-wealth-inequality-in-america-video-it-commits-worstalls-fallacy/#702d76a324fb Accessed December 3, 2016.

Essentially, all life depends
on the soil…There can
be no life without soil and
no soil without life;
they have evolved together.

Charles E. Kellogg

Appendix

SCIENTIFIC BACKGROUND

Despite its ecological significance, relatively little is known about soil on a global scale. Soil varies widely in composition and structure from region to region, and is extremely sensitive to local geology and the history of a particular piece of land. In addition to large-scale local variations due to rock formations and land use history, soil also displays hyperlocal features as well. The soil in my backyard garden is markedly different than the soil under my front lawn, or the compacted earth at the edge of the street. Although they all began as soil of the same type, structure, and chemical composition, the way that those tiny microenvironments are used affects their properties in subtle but important ways. Dirt packed down by the passage of many feet becomes hardened, and resists penetration by plant roots. It also tends not to absorb water as well, leading to local flooding in periods of heavy rain. In contrast, the looser, more fertile soil of the garden is easily washed away if there are no plant roots to hold it in place.

Of course, scientists usually sample at much greater depths that I am likely to turn over with a garden trowel, but this variation is a microcosm of the challenge that soil scientists face when trying to draw a map


Photo credit: Wikimedia Commons

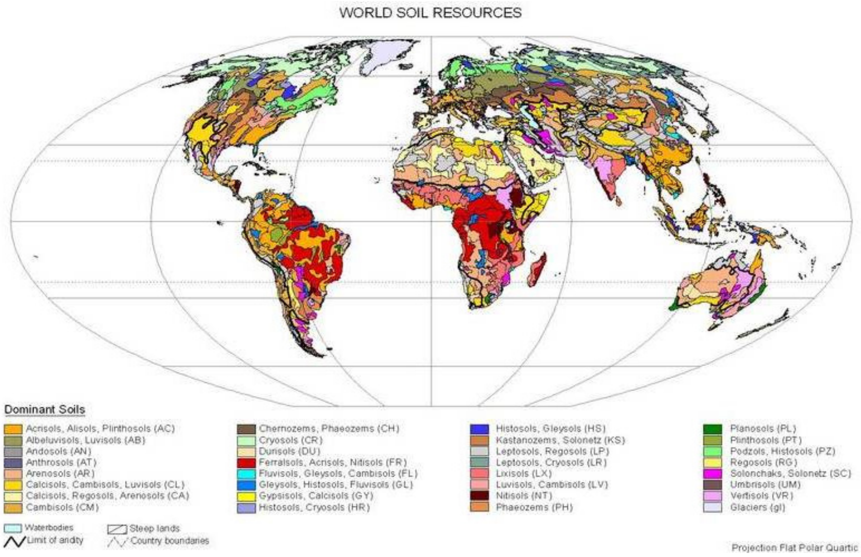
of the world's soil resources. Where do you sample, and when? How are the samples collected, and which features do you analyze? How much do you sample, and how deep do you go? Even if a scientist only collected one sample for every square mile, it would take 197 million samples to measure the surface of the earth, 3.8 million for the US alone. Considering that soil composition varies by depth as well as by location and that several samples need to be collected to obtain a representative measurement, one can begin to comprehend the magnitude of the task facing soil scientists in assessing the state of soils across the globe. Satellite and remote sensing can help with this daunting task, but they only measure surface properties, and much of the important information lies deep beneath the surface. Remote sensing is also of limited use in areas where the soil is covered with dense vegetation. In fact, tree coverage is often used as a proxy for measuring soil health, because it is so much easier to see.

Mapping the world's soils

Even if the barriers to soil sampling are overcome, the question still remains of which properties to measure, and how to map them. Different soil classification systems, incomplete data, and the constantly-evolving condition of the world's soil conspire against a unified map of soil resources. Despite these dififculties, many organizations are working to create a unified map of the world's soil. The Food and Agricultural Organization of the United Nations (FAO-UNESCO) created the most commonly used soil classification system, and published the first unified global soil map in 1981 (FAO 2012; FAO 1972; Sanchez et al. 2009).

Researchers from several international organizations have been working to compile, digitize, and update the global world soil map since 1995, adding onto the original map and incorporating data from several disparate databases across the world. The product of this international effort was the Global Harmonized Soil Database, first published in 2008 (FAO 2012).

A glance at the legend for the map shown in figure 66 is not terribly enlightening for the uninitiated; names like andosols and spodosols do not necessarily convey a strong sense of the objects that they represent.



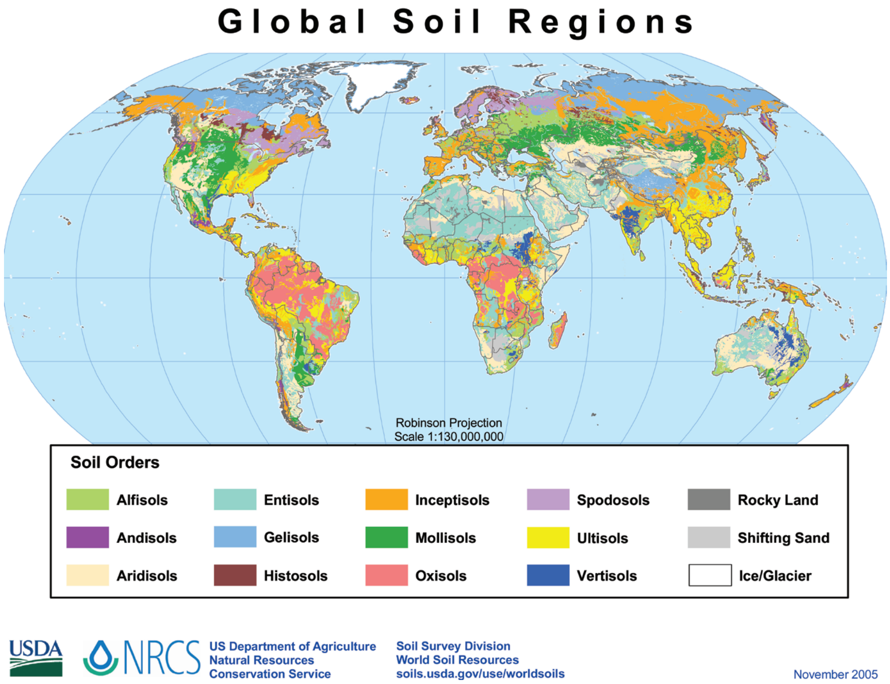


Figure 19. Soil classification maps from the Food and Agricultural Organization of the United Nations and the US Department of Agriculture. Soils are classified into 28 and 12 categories, respectively.

Instead, these names refer to a complicated classification system, meant to represent all of the important chemical and physical properties of a soil in a single name.

Soil is a variable mixture of rocks and minerals, sand, and clay, arranged in discrete layers of varying thickness. Soil classifications are based on detailed analysis of these individual layers (Mueller, 2010). Each layer is assessed for its organic content, drainage history, mineral content and the relative proportions of sand, clay and silt in the layer. Then, the physical properties are measured, including: the structure of the layer and its permeability by water and air, the consistency when lumps of soil break apart, its density and pH, and its ability to absorb water and other chemicals. The geologic and land use history are also included in the final designation.

The FAO-UNESCO published a full technical paper as the "legend" to accompany its Soil Maps of the World in 1997; the document is 140 pages long, and gives a brief history and definitions for each of the 28 major soil groupings and 153 different kinds of soil units. The USDA has a more accessible system with only 12 classes, but their definitions for what constitutes the different classes are still quite involved.

Soil microbiome

Agricultural interests have driven research focused on the chemical makeup of soil, and studies have historically focused on the specific macronutrients needed for plant growth. Introduction of chemical fertilizers in the latter half of the 20th century created an unprecedented increase in the amount of food that farmers are able to grow in a given area of land. Still, current approaches to plant nutrition and soil health are still relatively crude. It is well known that plants need a particular balance of phosphorous, nitrogen, potassium and calcium to survive, but the timing of release and the specific chemical forms of these nutrients must be tightly controlled to optimize absorption and ensure plant health.

In a natural soil ecosystem, this carefully tuned release of nutrients is usually carried out by soil microbes. Just as medical science is beginning to recognize the importance of bacteria in the human microbiome, biologists are discovering that plant health is intimately related to the kinds of bacteria and fungi present in the soil. Although we usually think of plants as relatively passive lifeforms, they constantly interact with their environment using chemicals to encourage and—to some extent—control the bacterial populations that support their health. These bacteria provide the nutrients that the plant needs to survive, and the right community of microorganisms helps to optimize plant growth. Some of the most exciting directions for current agricultural research involve using microbial communities to improve the effectiveness of chemical fertilizers and increase the productivity of plants.

It is estimated that 10 grams of soil (about a teaspoon) contains a teeming community of more than 100 billion bacteria, representing more than 10 million species (Decaëns et al. 2006). Many of these microorganisms specialize in breaking down dead plant material, releasing nitrogen and carbon dioxide in the process. Some bacteria also digest rocks and release the small quantities of mineral nutrients that plants need to survive.

Many soil microbes are impossible to culture in the lab, and have previously passed unobserved. Scientists have only recently discovered gene sequencing tools that allow them to characterize the astounding diversity of soil microorganisms, and efforts to analyze their diversity, habits, and contributions to the soil ecosystem have only recently

We must [it has been argued] go beyond reductionism to a holistic recognition that biology and culture interpenetrate in an inextricable manner.

Stephen Jay Gould

begun (Brink 2016). This is painstaking work, and it requires extensive sampling to understand interactions within the soil ecosystem, even for the simplest of cases.

In addition to microbial and fungal symbionts, the soil is also home to a wide range of arthropods (centipedes and millipedes), invertebrates (worms), and protozoa (single-celled organisms that release nitrogen into the soil), each with their own roles to play in the ecosystem function (Jouquet et al. 2006; Lavelle et al. 2006). Soil also serves as a habitat for almost a quarter of the world's species: more than 360,000 animal species make their home in the soil (Decaëns et al. 2006).

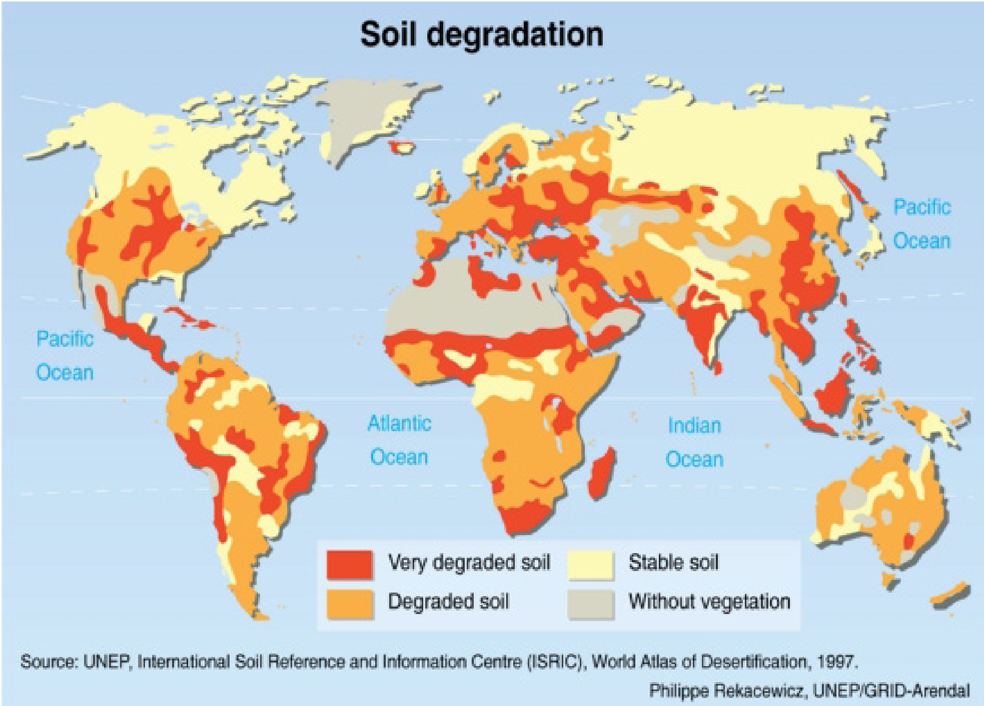
Soil as an ecosystem provider

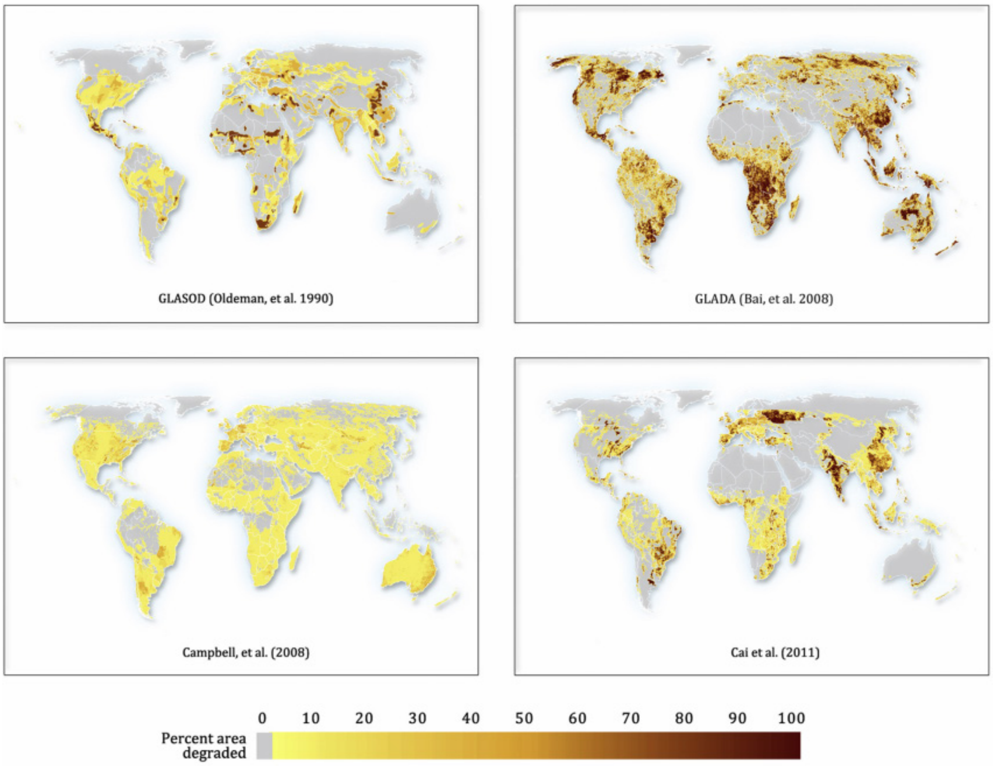
Soil provides a range of ecosystem services that are often taken for granted. Soil organic carbon—found mostly in the form of humus: a rich mix of discarded leaves, roots, decaying wood, and other dead plant materials—represents the world's third largest carbon reserve. The carbon stored in soil trails far behind the vast quantities of carbon stored in the oceans, but it still contains more than half as much carbon as the world's coal, oil and gas reserves, and 4 times more carbon than all of the living organisms on the planet combined (Lal 2008). In light of this information, restoring soil carbon stocks to their full capacity could be a partial strategy for combating climate change and mitigating the negative consequences of global warming.

Soil is also a critical participant in the global water cycle. A soil rich in organic material absorbs and holds onto rain water, preventing flooding. Soil that is compacted from frequent travel, disrupted by frequent tilling, depleted due to overfarming, or otherwise compromised cannot perform these critical ecosystem functions that we rely on for our basic survival.

Challenges to the soil

Human population growth has placed unprecedented demands on the world's soil resources in recent years. Extensive urbanization leads to the clearing of forests and widespread construction on previously open lands. Mechanized agriculture allowed farmers to produce food in large amounts, pushing the productivity of the soil to its limits. Increased


Soil degradation
Pacific Ocean
Pacific Ocean
Atlantic Ocean
Indian Ocean
Very degraded soil
Degraded soil
Stable soil
Without vegetation
Source: UNEP, International Soil Reference and Information Centre (ISRIC), World Atlas of Desertification, 1997.
Philippe Rekacewicz, UNEP/GRID-Arendal


GLASOD (Oldeman, et al. 1990)
GLADA (Bai, et al. 2008)
Campbell, et al. (2008)
Cai et al. (2011)
Percent area degraded
0 10 20 30 40 50 60 70 80 90 100

density and frequency of plantings lead to nutrient depletion in the soil, and the frequent tilling that often accompanies industrial agriculture can also lead to erosion and topsoil loss.

According to a 2015 report on the "Status of the World's Soil Resources" from the FAO-UNESCO, soil health is threatened by many different factors: erosion caused by wind, water, and agriculture; loss of soil organic carbon and mineral nutrients caused by excessive farming; acidification and heavy metal contamination due to mining and fertilizer use; waterlogging and salinization caused by encroaching seawater; compaction caused by heavy traffic, overgrazing, or mechanical agriculture; and paving or other building projects that permanently cover the soil surface, and which are usually the result of urban growth (FAO 2016). All of these factors contribute to declining soil health, and threaten both the environment and the global food supply.

Figure 20. Map showing the global distribution of soil degradation, published by the United Nations Environment Programme in 1997

Erosion

Generally speaking, soil can be divided up into two separate layers: the top few inches (and sometimes feet) are usually fertile topsoil, which has a fairly open structure and a lot of organic content. Under the topsoil, there is a denser layer of subsoil that often has a larger clay or mineral component. The quality of the topsoil is most important in determining the agricultural productivity of land, but it is vulnerable to erosion due to wind, water, or agricultural practices that disturb the native plant cover and leave bare soil exposed. Erosion is a serious problem in many areas of the world, and is responsible for soil productivity losses worldwide. Many human activities influence soil erosion, especially deforestation and high-till agricultural practices that rely on annual crops planted without cover crops (FAO 2016).

Sealing, Compaction, and Urbanization

Figure 21. A comparison of soil degradation maps based on different assessment methods. Discrepancies between the maps highlight how difficult it is to provide a definitive metric on this topic (Mueller et al. 2010).

The soil's ability to absorb water and reduce flooding can be compromised in two key ways. Compaction destroys the soil's pore structure in areas with heavy traffic or excessive mechanical agriculture, and in lands that are overgrazed. Soil can also be sealed by paving or permanently covering its surface. Compaction and sealing disturb the surrounding ecosystems and make it difficult or impossible for plants to recolonize the affected areas, because their roots are not able to penetrate the densely packed soils. Both of these problems tend to show up in urban or densely populated areas, and represent the most immediate threats to soil health from urban expansion (Mueller et al. 2010).

Chemical changes

Loss of soil organic carbon, excessive application of chemical fertilizers, soil acidification, and heavy metal contamination are just a few of the chemical changes that can affect the productivity of soils. Many areas are also experiencing salination, or an increase in salt concentrations in the soil, due to misuse of chemical additives or to infiltration with salt water as a result of sea level rise (Mueller et al. 2010).

The FAO report details these current threats to soil health, and suggests that it is critical to improve food production without damaging the soil by avoiding further degradation, improving agricultural techniques, and working to restore soil organic carbon and biodiversity through good management and land use practices (FAO 2016). The report identifies crop rotation, reduced tilling, and planting of cover crops as key strategies to protect and restore soil health in agricultural lands.

Soil Degradation

In order to properly contextualize threats to the soil, it is helpful to have an overview of where things currently stand. Soil degradation can be broadly defined as a loss of soil productivity over time. Degradation is a complex issue, influenced by soil quality, local climate, and land use history. If it is difficult to obtain a single, consistent, and detailed map of information about global soil resources, it is even more difficult to quantitatively measure the quality and productivity of those soils. The agricultural output of an area does depend on the inherent quality of the soil, but it is also heavily influenced by local climate and terrain.

Only about 25% of the global land area is suitable for crop farming (arable), while 40-50% of it is currently used for agriculture (including grazing herds). Soil moisture and temperature are the primary factors that

Photo credit: NASA

affect agricultural productivity, followed by low moisture retention, depth of soil, chemical properties, and challenging local topography. Roughly two thirds of the soils in developing nations suffer from low productivity (Mueller et al. 2010).

Identifying degraded lands is also controversial, because different methods yield different results. Expert opinions are important for providing context to laboratory results, but they are inherently low resolution (one opinion is applied to a broad area of land), and also highly subjective. Satellite mapping provides an alternate strategy, but can measure only a limited number of surface properties, and so neglects important subterranean effects. Satellite data is a secondary measurement; tracking the kind and density of plant material growing in a region at a given time, rather than directly measuring the soil itself.

Since plant cover can also be affected by droughts, land use change, and other human activities, this is not always a reliable measure of the inherent capacity of the soil. Biophysical models use computers to integrate information about the geology, physical properties, climate, and other metrics of a particular area to produce a rich analysis of soil quality, but they also rely on mathematical models, simplifications, and assumptions that can obscure the reality on the ground. No one model or method of measurement can capture all of the information about soil productivity. Instead, it is important to compare multiple techniques to arrive at a more holistic view (Gibbs and Salmon 2015).

Land Use

It is important to track the human contributions to changing soil health, as well as the geographic variations and climate dependencies. Much soil degradation stems from human activities and poor land use practices, and changes in land use are a natural part of societal change. Deforestation, accelerated erosion, chemical contamination, and urban expansion all contribute to soil degradation, and have also been responsible for roughly 35% of anthropogenic carbon emissions since 1850 (Foley 2016). Introduction of sustainable agriculture and intentional restoration activities can contribute to improved soil health and help to reverse these damages.

Analysis of land cover change is faced with many of the same issues encountered when mapping soil health. Differing definitions, inconsistent measurement techniques, and structural blindness to local context are also present here. Regional and boundary disputes also contribute to the difficulty of identifying changes in contested lands. Several organizations track land cover change, and many researchers have been attempting to synthesize those results into a coherent global picture (Lepers et al. 2005). Again, the FAO and other organizations are leading the way in making that data available to the public, through open data initiatives.

Although cities make up a relatively small percentage (roughly 1%) of global land use, urban expansion is an important consideration when examining soil issues. Densely populated cities can change the local climate through heat island effects. Large areas of sealed or compacted soil cannot absorb water runoff during storms, and so urban areas can contribute to accelerated erosion in surrounding areas. And, perhaps most important, large urban populations require large amounts of food to be imported from the surrounding area, putting strain on surrounding agricultural lands and requiring dense networks of roads to be built through otherwise open areas.

Urban land area is expected to increase dramatically in the coming years. If the current rates continue, the global urban area is expected to triple between 2000 and 2030. This increase will not be distributed uniformly around the globe, but is expected to be largest in developing nations, which currently host some of the world's most important biodiversity hotspots (Seto, Guneralp, and Hutyra 2012).

Popularizing Soil Science

In light of all these factors, it seems straightforward to assume that soil preservation would be a primary focus for humankind. And yet, most people know little to nothing about the soil, the services it provides, or the pressures that it faces due to our increasing need for food. This needs to change.

The United Nations declared 2015 to be "the international year of soils," in an attempt to increase scientific research in the area and to improve public understanding of this complex system. And yet,

Photo credit: Atmospheric Radiation Program preservation of the world's soil resources doesn't receive nearly the attention that it should, perhaps because worms and bacteria don't have the public appeal of the iconic polar bears, elephants, and great cats that have helped to popularize species conservation efforts in recent years. Part of the difficulty arises from the fact that soil is something that we just don't see. The rich diversity of soil ecosystems exist out of sight and underground, and many of the most fascinating soil processes happen at the microscopic or molecular level, out of the range of our daily lived experience.

I believe that data visualization is uniquely positioned to help bridge this gap between our knowledge of the soil ecosystem and the importance of what it provides. By creating visualizations that engage the user in exploring this complex ecosystem, I hope to inspire people to protect what I believe to be one of the most important global resources that we have.