Changing Minds to Changing the World

Mapping the Spectrum of Intent in Data Visualization and Data Arts

8 December 2013
Scott Murray
DRAFT

Introduction

The recent explosion in available data sources and data-processing tools has both scientists and artists diving into the world of data visualization. The result is a diverse, interdisciplinary field of practice, in which practitioners cultivate knowledge in other areas: Statisticians are learning about design, while designers are learning about statistics. All of these people are producing visualizations of data — objects of visual communication — but with widely varying intentions and goals for their creations.

Several primary goals for visualization — exploratory, informational, and rhetorical — are well-established. But in a field where artists (seeking to produce aesthetic, yet "accurate" work) are learning about science, and scientists (seeking to produce informational, yet "aesthetically appealing" work) are learning about art, how can we delineate between the range of types of intended communications, and can those delineations be made in any meaningful way?

One of the most exciting aspects of visualization today is the ease with which practitioners from different backgrounds collaborate and engage with each other. By

examining the discourse adopted by these practitioners, we can identify what processes they all have in common, and then map where practices overlap and where they diverge.

Getting Started

"How should I get started with data visualization?" This increasingly common question turns out to be quite difficult to answer. Visualization is inherently interdisciplinary; a true mastery of all its forms would require expertise in:

- visual design
- interaction design
- data analytics
- statistics
- mathematics
- psychology
- computer science

That list doesn't even include the technical skills required for implementing a project with specific tools, such as Excel, Tableau, Illustrator, Processing, R, D3, or — more commonly — some combination of tools, each optimized for a different step in the process.

Yet none of the practitioners I know are experts in all of the subjects and tools mentioned above. Many have formal backgrounds in one subject, then dabble in others. A computer scientist by training may "have a knack" for visual design, or a designer may discover she also excels at statistics. Thus, we pick and choose, and draw from whatever skill sets we are inclined to cultivate within the limits of our available time, interest, and abilities. I find that most people in data visualization are, by nature, very curious; we would prefer to learn everything and be skilled in all areas, but of course life gets in the way.

When beginners ask how they can get started, this interdisciplinary quality of the practice also gets in the way. There is no one best path into visualization; every practitioner has a different point of entry, such as:

- web design
- graphic design
- industrial design
- architecture
- mathematics
- cognitive science
- computer science
- journalism

With so many possible points of entry, the question is easier to answer on a personalized, individual level. To someone with a highly technical background, I might recommend some design books. To a journalist, I could suggest resources on data analysis and graphical storytelling. But of course even these are generic responses, and don't account for the individual's full range of prior experience. An interdisciplinary field can be exciting and stimulating for practitioners already who are already fully engaged. But for those just dipping in their toes, it can be frustrating to ask lots of questions and frequently hear the same answer: "Well, it depends."

Common Ground

In an effort to provide a more comprehensive answer to such questioners, I want to document the full range of experience and expertise found in the visualization community. As incomplete as this attempt may be, it should be valuable to see our differences mapped out, as well as the practices and language that we have in common.

While searching for this common ground, I also intend to propose an informal *taxonomy of practice*. Much prior work has been done to classify visual properties and common visualization elements (e.g., Bertin, *Semiology of Graphics*, and Segel and Heer, *Narrative Visualization: Telling Stories with Data*, 2010), but here I want to explore the community of practice itself. As the field grows, it becomes increasingly important to understand the range of its participants.

It is my sense that visualization practitioners, despite our diverse backgrounds and the interdisciplinary nature of the field, have quite a bit in common — it's just that we have a hard time describing exactly what that is. As evidence, I observe that many of the same people speak at or otherwise attend the following conferences:

• Eyeo Festival
• Resonate
• See Conference
• Strata
• Visualized
• IEEE VIS (formerly VisWeek, includes VAST, InfoVis, and SciVis)

While these conferences appeal to a range of audiences — from primarily academic researchers to decidedly non-academic professionals — there is significant overlap in attendance. How is it possible that a "creative coder" attending Resonate, a small annual gathering for exploring technology role in the arts, could be just as comfortable in that environment as at Strata, a more corporate environment focused on big data and data insights? Why does VIS, the essential event of the year for academics and researchers in visualization, also have an art program for exhibiting "artistic" uses of data alongside "practical" examples? The ACM SIGGRAPH conferences, too, while

focused more broadly on computer graphics and not just visualization, have extensive arts programs.

The fact that I've placed "creative coder," "artistic," and "practical" in quotation marks indicates that we have a language problem on our hands. This begins with how we identify ourselves. I have seen practitioners refer to themselves by the following titles:

• data visualizer
• data designer
• designer
• artist
• data artist
• code artist
• generative artist
• creative coder
• creative technologist
• graphics editor
• cartographer
• researcher
• storyteller

Each implies a slightly different emphasis — more fine arts, more code, more data — but, at gatherings, these people converse freely, communicate well with each other, and typically avoid using titles altogether. It is a common woe, especially toward the fine arts end of the spectrum, that these titles are essentially meaningless, except as cues to other practitioners already "in the know." The interdisciplinary data artist's elevator pitch is often brief and inaccurate, because the nuances of the process are not easily reducible to summary for outsiders. The result is a more tightly knit (and unintentionally insular, if still friendly) community.

I witnessed this label-aversion play out at a large scale at the first Eyeo Festival in 2011. The conference is held in Minneapolis each June, and invites presenters from a range of fields — data visualization, generative art, installation art, design, computer science. Its tagline, "Converge to inspire," is conveniently vague, and as such reflects the event's reluctance to pigeonhole its attendees. By the end of the week, I heard many people describing others not as artists, creative coders, or data visualizers, but as "you know, the kind of people who would go to Eyeo." For lack of a better umbrella term, we resorted to self-reference. I think we can explore this phenomenon, look at the principles and practices shared in common, and identify a clearer way of describing ourselves to others.

Mapping the Field

To frame the discussion, I will propose a series of ranges or spectra upon which practitioners and projects may be situated. For example, in the field of visual communication, there is an ongoing tension between the terms "art" and "design."

Art	⟵————————⟶	Design
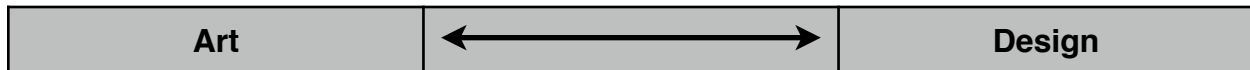
Work deemed to be on the "art" end of the spectrum may, for example, be considered purely aesthetic, have little or no "functional" purpose, and have little commercial or "practical" value (except, of course, as fine art, which, I would argue, is as practical a purpose as any). Work may be on the "design" end of the spectrum if it has obvious commercial value, communicates a specific message, and functions with an explicit purpose. Yet "design" is not without aesthetic value, and so shares that element with "art." And "art," such as illustration, may be employed within a "design" context, to communicate a message larger than the art itself.

At what point does an image cross over between art and design, or vice versa? While this distinction is in some sense arbitrary, it nevertheless carries value, at least by forcing us to struggle with the language we use to describe our work and the values we ascribe to it.

To me, the most meaningful way to make this distinction is to identify the goal or *intent* of the creator. For art, the intent may be to elicit a purely aesthetic or emotional experience from the viewer/participant. For design, the intent is typically to communicate a specific message to the viewer/participant. So, regardless of the medium and context, an image made with intent to communicate a particular message or meaning falls near the "design" end of the spectrum. (This assessment is made independent of whether or not the design is successful in achieving its creator's goals.) An image with intent to elicit an emotional experience (without a specific message), can be called art (though, to further muddy the waters, art often has a message). Perhaps this could be simplified even further to say that a work's position on the spectrum indicates only the specificity of its intended message. The more open the message, the more artistic; the more specific, the closer it is to design.

The art/design spectrum, as well as the others I propose below, is presented as two-dimensional, but of course, reality is more complex and not suited to such clean definitions. (This is particularly true given the current rate of change in visualization practice, and the rapid development of new forms.) Please take these proposals as tools for framing discourse about the current state of the field, not attempts to define it in fixed terms.

I will address each of the following in turn:

- Avenues of practice
- Media
- Contexts
- Conceptual structures
- Goals

The first spectrum can be used to evaluate either practitioners or projects, while the others are specific to individual projects.

Avenues of Practice

As mentioned above, practitioners self-identify with a range of titles. Broadly speaking, each of these titles could be placed on a spectrum of *data arts* to *data visualization*. Both are visualization, of course, but data arts is more akin to fine arts, and data visualization is more akin to design — that is, it creates visualizations with specific messages, or with the intent to reveal messages intrinsic to the data (e.g., patterns and trends).

Data Arts	⟷	Data Visualization
• artists • creative coders • creative technologists	• designers • illustrators • storytellers	• data analysts • scientists • journalists • cartographers • researchers

To offer an example, I would file Memo Akten into the data arts end of this spectrum. Akten's project "Forms," done with the artist Quayola, is intensely data-driven or data-derived, yet it is more evocative than explicitly communicative.

Still image from "Forms," Memo Akten and Quayola, 2012

That said, Akten has done projects for corporate clients, such as the "McLaren P1 Light Painting," which I would classify as a data illustration: it functions primarily as an advertisement for a new automobile, and thus, the communications *intent* is different from that of "Forms."

Still image from "McLaren P1 Light Painting," Memo Akten and James Medcraft, 2012

Continuing further still to the right edge of this spectrum, we can look at geographic maps, a visualization of practice that has undergone massive changes in the past ten years. Stamen Design in San Francisco, which refers to itself as "a design and technology studio," is known for their wide array of explorations in maps. Their "Toner" tiles are intended for use when a map will be printed and photocopied. As such, they don't use any color, and gray areas are rendered with a halftone screen, to improve reproducibility by analog means. Yet, even with this constraint, the design functions effectively as a guide for orientation and directions — that is, as a traditional map.

Toner-style map, from MapStack, by Stamen Design, http://maps.stamen.com/m2i/image/20131117/toner_uwnIcEfPDtw

This particular intent and specificity of communication places the "Toner" map squarely on the data visualization end of the spectrum. Contrast that with Stamen's "Watercolor" tiles, which represent the same underlying data in a completely different form.

Watercolor-style map, from MapStack, by Stamen Design, http://maps.stamen.com/m2i/image/20131117/watercolor__Q2AD5HJgMk

The "Watercolor" maps are less precise by design, evoking an abstract sense of place for those already familiar with the place, as opposed to helping orient new visitors to specific locations within a place (e.g., cities, streets, addresses). So I would file the "Watercolor" maps on the data arts end of the spectrum.

But what about Stamen as an entity? Where do its designers and technologists fall, given their influential contributions all along the spectrum?

This highlights how difficult it is to classify individual practitioners. I may be acting as an artist today, but I put my designer hat on when it's time to update my website, and perhaps I have to think like an analyst when making sense of the data set underlying my next project. I will propose a solution to this classification problem, but first, it will be useful to discuss how to classify individual projects.

Contexts

Visualizations commonly have one primary context of presentation, although essentially every project is now documented and published online in some form.

Gallery Primary, Online Secondary	Online-Only	Print Primary, Online Secondary
• prints • video • interactive • installation	• personal website • project website • arts blog/site • visualization blog/site • technical blog/site	• newspaper • magazine • poster • flyer • academic paper/journal

Jonathan Harris and Sep Kamvar's "I Want You To Want Me" is a computationally intensive installation created for the Museum of Modern Art's *Design and the Elastic Mind* exhibit in 2008, curated by Paola Antonelli. As a commissioned piece, it was designed from the beginning for the gallery context. While the artists posted documentation online, it isn't feasible to adapt the project for the web, so it remains a gallery-only, in-person experience.
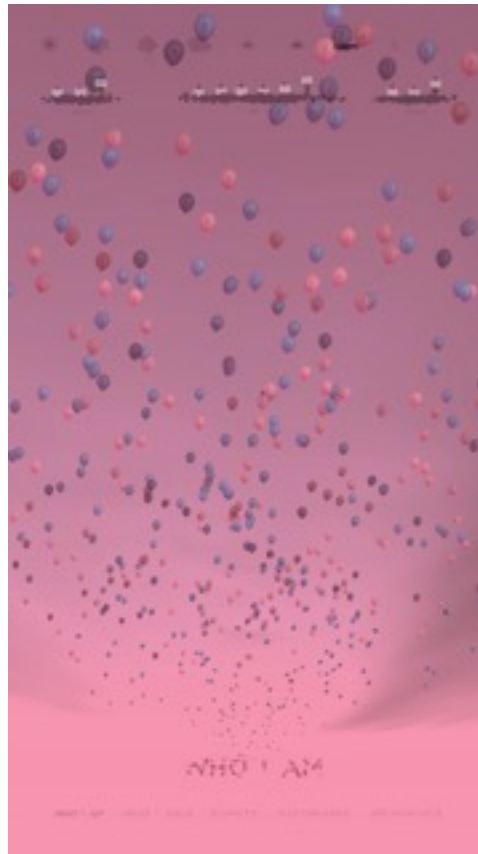
Still image from "I Want You To Want Me," Jonathan Harris and Sep Kamvar, interactive touch-screen installation commissioned for The Museum of Modern Art's Design and the Elastic Mind, February 14, 2008.

In contrast, Santiago Ortiz's innovative portfolio interface was designed specifically for the online context, and wouldn't make sense in any other medium. It begins as a grid of

project image thumbnails, but visitors can drag a round cursor to adjust the visual weight given to projects across three axes: recent, favorites, or all projects. At a glance, we can watch projects resize to reflect, for example, which ones were completed most recently versus which ones Ortiz himself enjoys. This form of representation, unlike Stamen's "Toner" maps, isn't intended for print, and would cause confusion in paper form, due to clipped images and abbreviated text.
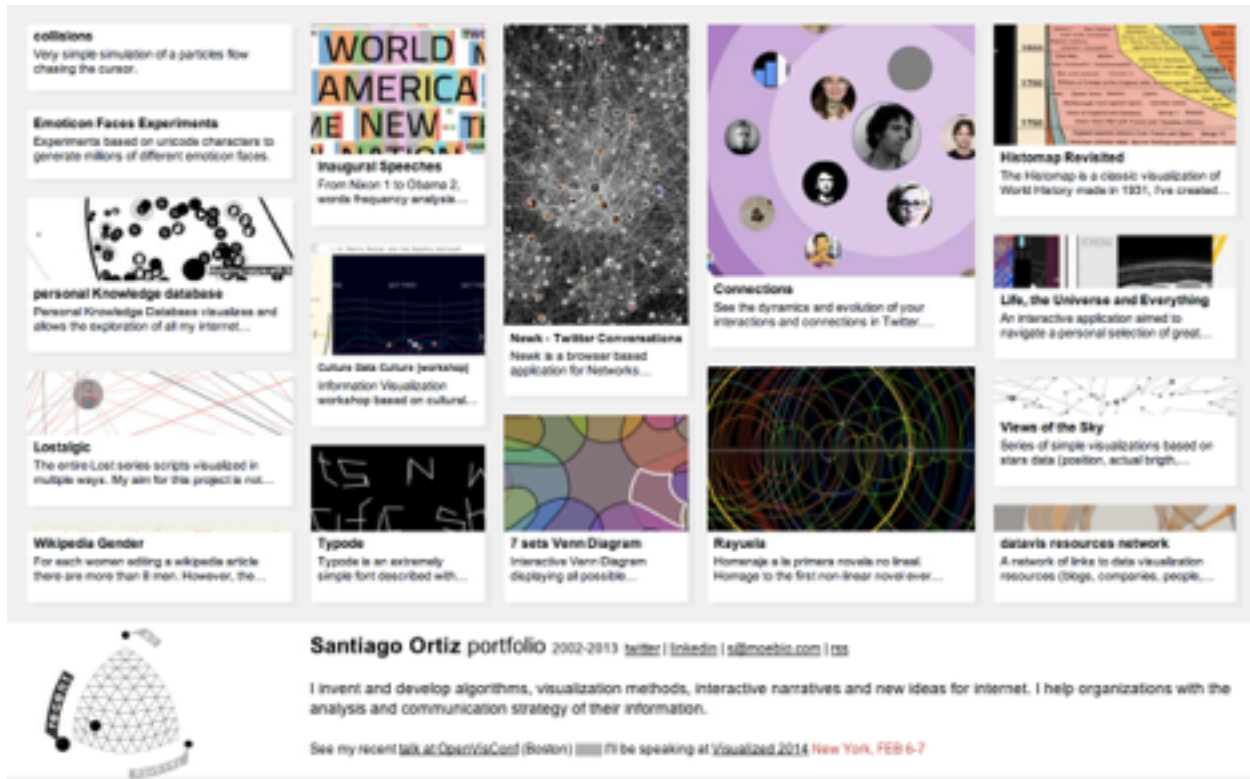
Portfolio, Santiago Ortiz, 2013, moebio.com

Continuing toward the print end of the spectrum, while many visualizations are designed primarily for print output, the dual approach taken by the *New York Times'* Graphics Desk is slowly becoming more common. At the *Times*, every graphic must work in both print and online. Typically, this means designing a default view that communicates the story. The default view works in the print edition of the paper, and also serves as the initial view of the online version. Interactivity can be used to make the piece explorable, enabling readers to dig deeper into specific data values. For example, this graphic on drought in the US includes annotations that highlight key trends in the data, but the online version also allows readers to mouse over any section of the graphic to reveal specific drought levels.

Drought and Deluge in the Lower 48

Last summer's drought, one of the worst in a century, has continued through the winter. This chart shows the proportion of what is now the contiguous U.S. in various stages of drought over 118 years of record-keeping. Roll mouse over individual months to see what percentage of the lower 48 was in drought.   Related Article »

During January, 56 percent of the contiguous U.S. was in moderate to extreme drought, the highest January level since 1955.

Dryness based on the Palmer Drought Index

← Wetter   Avg.   Drier →

Roll mouse over to isolate categories

Large, intense regional droughts that migrated across the country. The West saw persistent drought in the first half of the decade.

The drought of 1988 destroyed at least half the crops on the Great Plains and was one of the costliest U.S. natural disasters.

Years of persistent drought. At its worst, most of the country was involved. Hardest hit: the South, the Plains, Texas and New Mexico.

Four waves of very intense drought spreading over huge expanses of the country, then retreating about a year later. Known as the Dust Bowl days.

"Drought and Deluge in the Lower 48," Mike Bostock and Shan Carter, February 22, 2013.
*http://www.nytimes.com/interactive/2012/08/11/sunday-review/drought-history.html*

Media

Visualizations are often designed with at least one primary target medium in mind. Those media may be considered along a spectrum of *static* to *interactive*.

Static	⟵⟶	Interactive
• print • screen (static)	• screen (dynamic)	• screen (dynamic) • spatial (projected/ installed)

Note that the static/interactive spectrum is independent of a work's *context*. For example, not all *New York Times* graphics are interactive, even when published online. "Drought's Footprint" was only ever intended to be a static image, both for print and on the website.

Drought's Footprint

More than half of the country was under moderate to extreme drought in June, the largest area of the contiguous United States affected by such dryness in nearly 60 years. Nearly 1,300 counties across 29 states have been declared federal disaster areas. Areas under moderate to extreme drought in June of each year are shown in orange below.   Related Article »
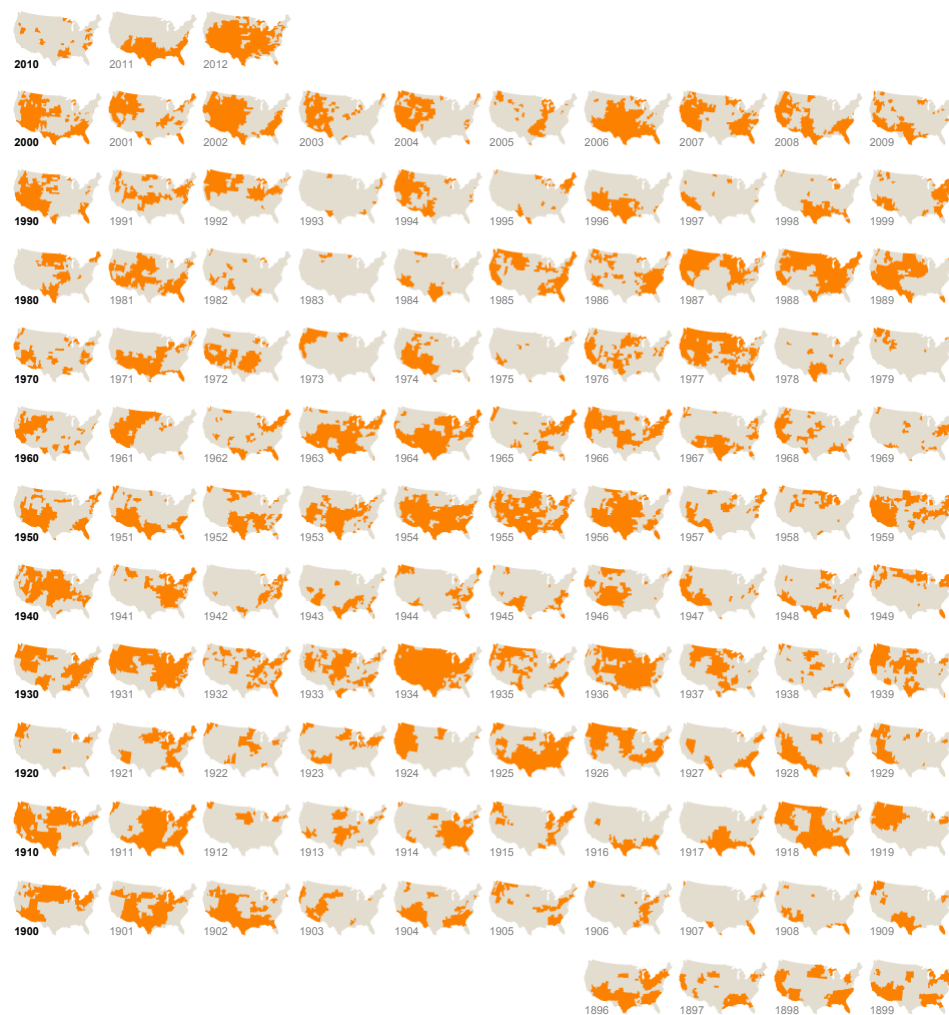

"Drought's Footprint," Haeyoun Park and Kevin Quealy, July 19, 2012. http://www.nytimes.com/interactive/2012/07/20/us/drought-footprint.html

Work intended for the screen can be dynamic without being interactive, such as IBM's "THINK Exhibit," which included a large-scale "data visualization wall." During its temporary installation in New York, the wall displayed real-time visualizations of data about the city, such as traffic flows, air quality, and water use. Since the display was generated live, it was dynamic, although not directly interactive. This stands in contrast to a pre-recorded video loop, which remains static in the sense that it merely repeats itself and its imagery does not change over time.

Photo documentation of the IBM THINK Exhibit at Lincoln Center in New York City, developed by Mirada, opened September 23, 2011. http://mirada.com/stories/ibm

Many screen-based visualizations are interactive, of course, but on the far end of the spectrum are works that are so dependent upon interaction with participants that, without it, they essentially cease to exist. "Shadow Monsters," an installation by Philip Worthington, for example, begins as nothing but a silent room with a plain white screen. Only when participants enter the space does the system spring to life, interpreting their shadows as horned, fanged creatures with creepy hair and nails.
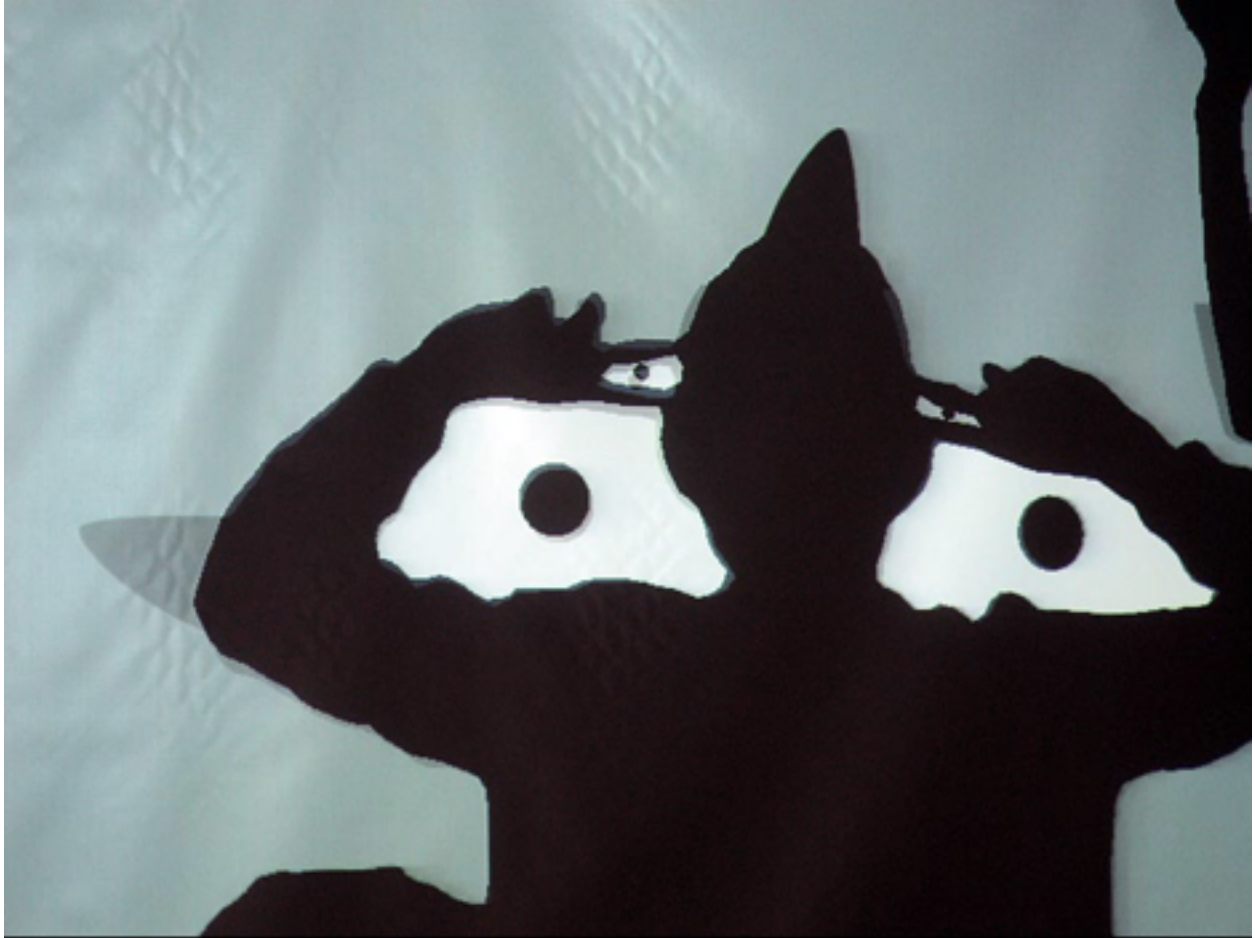
Shadow Monsters, by Philip Worthington, 2004–ongoing. http://moma.org/exhibitions/2008/elasticmind/#/229

Conceptual Structures

Exploratory tools are used to visualize data for the purposes of discovering what is interesting and valuable about that data. Explanatory visualizations take a point of view, and communicate to the viewer some pattern, trend, or discovery already observed.

Exploratory	⟷	Explanatory
• flexible views • open structure • adaptable encodings		• constrained views • narrative structure • optimized encoding

Exploratory visualizations are often interactive, and many tools are designed primarily for this purpose, such as Tableau and R with ggplot2. Since exploratory designs are geared toward producing insights, they tend to be more literal and specific than purely

aesthetic. (A "data arts" visualization would not be likely to produce valuable insights.) "The Dynamic HomeFinder" was one of the very first such exploratory visualization tools.
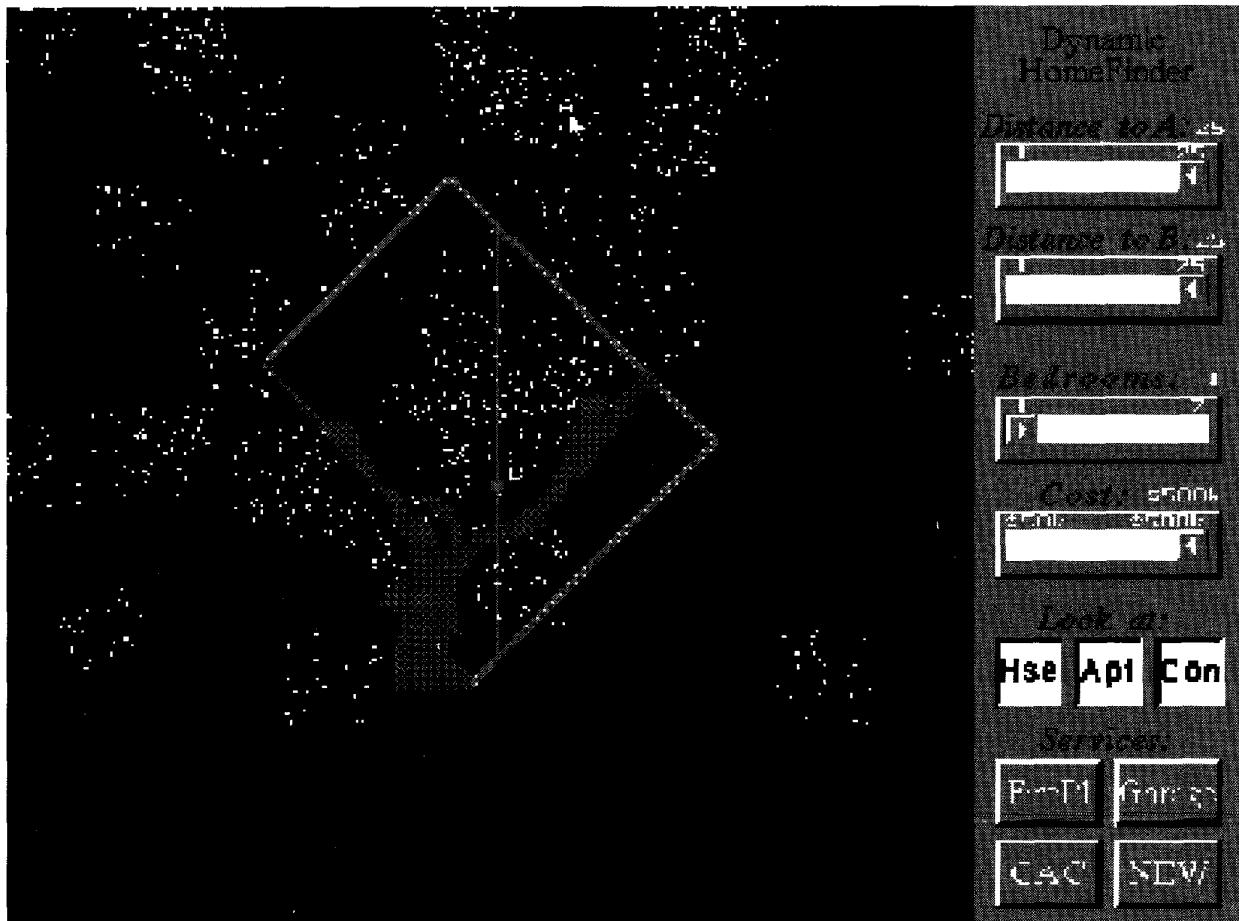
"The Dynamic HomeFinder: Evaluating dynamic queries in a real-estate information exploration system." Christopher Williamson and Ben Shneiderman. ACM, 1992.

Explanatory visualizations are more focused, with limitations imposed on the viewer and design elements to increase the specificity of the communications value. Discussions of visualization as storytelling are referring to explanatory images and interfaces. Journalistic graphics are typically very strong in this regard, such as in "The Cost of Water," a piece I worked on for the *Texas Tribune,* which explores why, despite record droughts, water is relatively inexpensive in Texas.

Rates are comparatively cheap …

The Texas cities shown here generally have lower water rates than others in the nation for households using a typical 7,500 gallons a month. In the rainy Northwest, Seattle and Portland (not shown) have among the highest rates of any cities. Texas' rates are also generally lower than other arid Southwestern cities like Phoenix and Los Angeles. These data are limited in that they apply only to usage of 7,500 gallons a month. Higher usage may be charged at rapidly increasing rates. One challenge in comparing water rates: "tiered" pricing systems in many cities levy higher charges for increasing usage.

Next: Per-Family Usage →

Price for 7,500 gallons/month

○ $56.18
○ $15.08

City	Typical Family Water Bill, 2011 Price for 7,500 gallons/month	Household Usage Gallons/year, single family	Level of Precipitation Inches per year avg.	Drought Time in drought, past 5 years
Seattle, WA	$56.18	46,720	37.4	24%
Boston, MA	$41.18			
Philadelphia, PA	$39.30			
Phoenix, AZ	$38.55			
Los Angeles, CA	$37.56			
Minneapolis, MN	$34.56			
New York City, NY	$31.80			
Houston, TX	$31.43			
Denver, CO	$24.08			
Detroit, MI	$22.95			
San Antonio, TX	$22.80			
Dallas, TX	$21.45			
El Paso, TX	$19.43			
Miami, FL	$17.03			
Chicago, IL	$15.08			

Still image from "The Cost of Water," a collaboration between Scott Murray, Geoff McGhee, and Kate Galbraith of The Texas Tribune, June 8, 2012. http://www.texastribune.org/library/data/cheap-water-in-texas/

Some visualizations or tools, of course, try to serve both exploratory and explanatory functions. Often, this employs a structure of a default explanatory view, followed by the use of interactivity to enable independent exploration.

Goals

Finally, each project is created with different goals, which may be placed somewhere on the spectrum of *inspire* to *inform*.

Inspire	←——————————————→	Inform
• aesthetic experience • emotional experience	• motivate • upset/disturb/enrage • delight/amuse/satisfy	• enlighten

An *inspiring* project may, like art, induce a kind of "a-ha moment" absent any concrete information. An *informative* project may communicate specifics of its data, but without any noticeable emotional impact.

On the inspiring end of the spectrum, we may find "Tape Recorders, Subsculpture 12," an installation by Rafael Lozano-Hemmer. Sensors tracks the presence of visitors to the space, and the length of their visits is expressed through the lengths of tape measures. The work is data-driven, but the individual data values are meaningless; the aesthetic and emotional experiences are what matter.
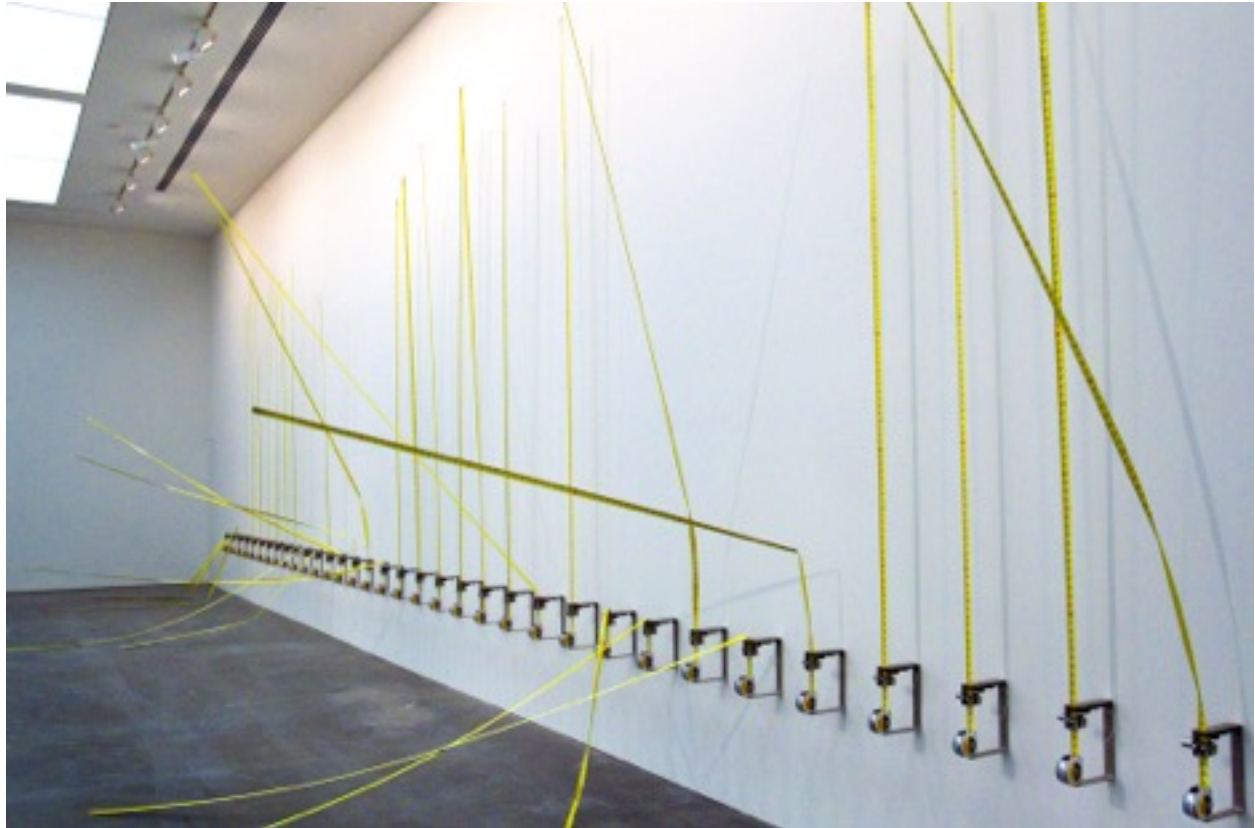
Photo documentation of "Tape Recorders, Subsculpture 12," Rafael Lozano-Hemmer, 2011.

On the informative end of the spectrum, we find wholly uninspiring charts and graphs, like this example from *The Economist*. This is in no way to pick on *The Economist*; when communicating specific data values, it is not necessary to inspire or delight. The chart below efficiently communicates the rise of text messaging, and includes several annotations, offering context of historically relevant moments. This chart is intended to inform, and it does so successfully.

Text messaging

Global short message service (SMS) traffic, trn   Global mobile subscriptions, bn

FIRST TEXT MESSAGE, "MERRY CHRISTMAS" SENT

FIRST COMMERCIAL SMS SERVICE LAUNCHED

PREDICTIVE TEXT, "T9" INVENTED

TWITTER FOUNDED, FACEBOOK OPENS TO GENERAL PUBLIC

US MOBILE SUBSCRIBERS SEND AND RECEIVE MORE TEXTS THAN PHONE CALLS

WHATSAPP, AN SMS ALTERNATIVE, FOUNDED

CHINA SENDS 825bn SMS MESSAGES

FORECAST

1992 93 94 95 96 97 98 99 2000 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16

Sources: Portio Research; Nielsen; Acision; national sources

Economist.com/graphicdetail

"OMG! Texting turns twenty," Economist.com, December 3, 2012. http://www.economist.com/blogs/graphicdetail/2012/12/daily-chart

Many projects, especially in data journalism, aim for a balance of inspiring and informing — such as when *informing* is essential, but achieving that end requires also engaging the reader on an emotional level.

One such landmark project is "We Feel Fine," another piece by Jonathan Harris and Sep Kamvar. Made in 2005, "We Feel Fine" is one of the early, online interactive visualizations. Still just as potent almost a decade later, it doesn't hurt that the data behind the project are themselves all about emotions and the human experience.

Still image from "We Feel Fine," Jonathan Harris and Sep Kamvar, 2005. http://www.wefeelfine.org

More recently, Periscopic's "US Gun Deaths" interactive visualization poetically and powerfully documents lives lost to gun violence — and projects an alternative future in which victims live out the rest of their lives (as algorithmically projected). The work performs a dual role, both informing us of the scale of tragedy as well as inspiring us to reflect upon and debate the significance of so many lives cut short.
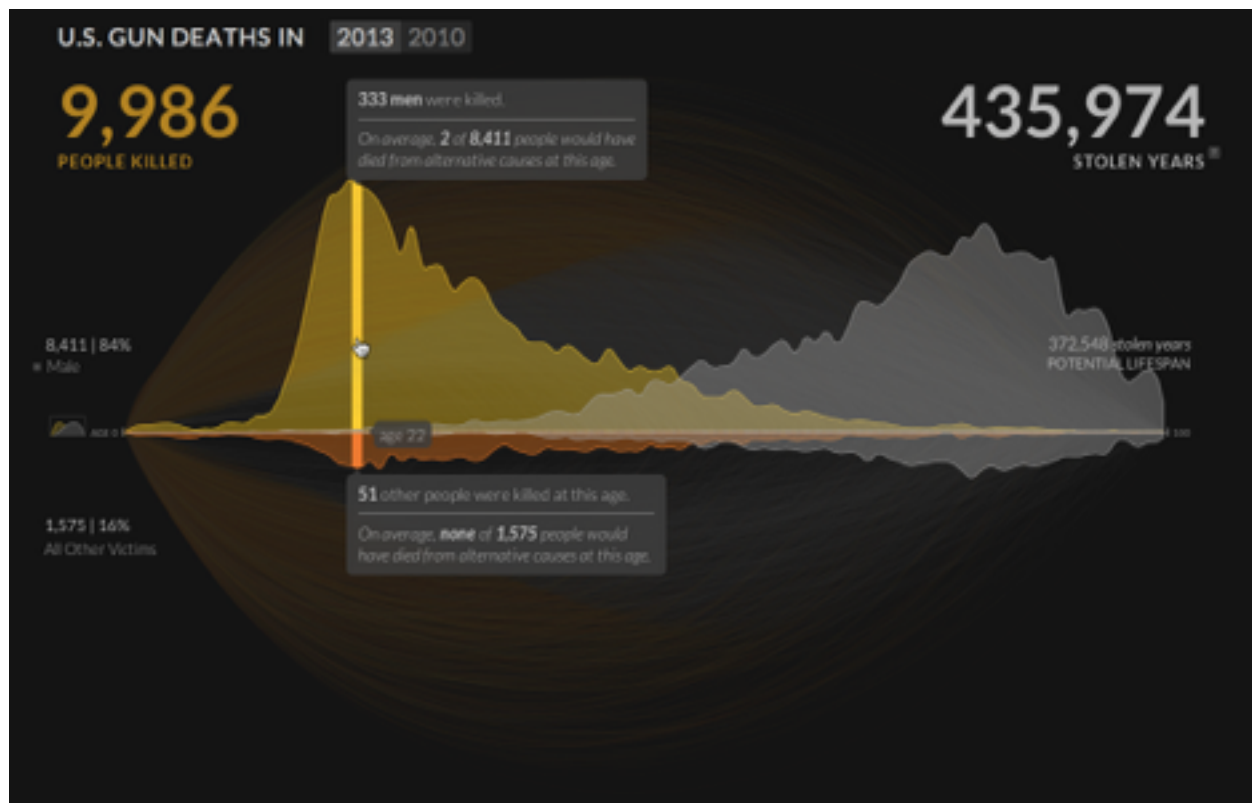
Still image from "US Gun Deaths," Periscopic, 2013. http://guns.periscopic.com/

Making Meaning From All This

Given these groupings — avenues of practice, media, contexts, conceptual structures, and goals— it should be possible to (roughly) evaluate and score any given project along each spectrum. If we consider each spectrum as an axis, an arbitrary, normalized value could be assigned to each. For example, a project could be scored anywhere from 0.0 to 1.0 along the axis of *inspire* to *inform*. Scores could be requested from numerous reviewers, and then aggregated to produce mean scores for a project, one value for each axis.

Having converted our collective assessments to data, we could (of course) visualize the results. I would recommend a parallel coordinates plot, with each axis oriented vertically, and horizontal lines connecting the values for each project.
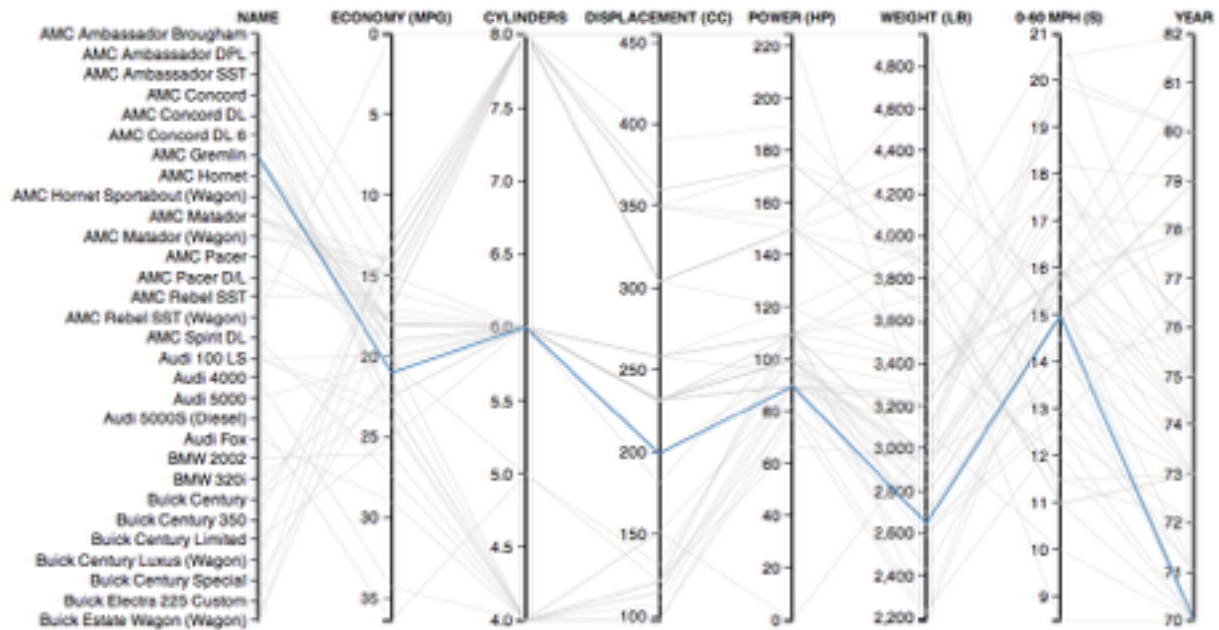
"Ordinal Parallel Axis" example, Kai Chang, 2012, http://bl.ocks.org/syntagmatic/3731388

Through interactivity, we could filter the view to show only projects by a particular creator, or by people from a specific subfield (say, only "researchers" or "artists"). This could enable us to discover places in the field where practices converge or diverge, either conforming to or challenging our expectations.

Independent of the visualization, after scoring all projects by a single creator, we could then calculate a "career average" with which to place them along the *data arts / data visualization* spectrum of practice. While acknowledging that we all move between many roles, it could be useful to see how heavily the field skews toward the arts or the other direction. (My sense is that there is so much interest in the field right now, from a diversity of perspectives, that there is a fair balance.)

This approach reminds me of two recent projects. First, a recent visualization by Pitch Interactive which visualized artists' careers with colorful star diagrams.

Star diagrams from "McKnight Artist Fellows: Visualizing Artists' Careers," Pitch Interactive with The McKnight Foundation, XXXX, http://diagrams.stateoftheartist.org/gallery

Second, a map by Jeff Clark of visualization practitioners on Twitter.

"Data Visualization Field on Twitter," Jeff Clark, 2012, http://www.neoformix.com/2012/ DataVisFieldSubGroups.html

Whether artists, scientists, journalists, or cartographers, practitioners of data visualization all seem to be in love with data. And what data could be more appealing than data about ourselves? Gazing into mirrors is fun, but beyond that, self-mapping is a great way to understand how we each fit into the field and, perhaps more importantly, it is a tool for explaining to others what the field is all about. With a map, we have a visual interpretation of the phrase heard so often at Eyeo. Now, instead of saying "the

kind of people who would go to Eyeo," we could just point to an image and say "data designers are the people on this map."

Common Elements

Other than a love of data and visuals, what do all visualization practitioners have in common?

Tools and Media — We all live and work in the same time, and have access to the same tools and media for sharing our work. Fine artists use Processing, but so do data journalists. Statisticians use R, but so do increasing numbers of people from other fields and specialties. Each tool is designed for a different audience or task, but there is a surprising amount of crossover between sub-fields. Nearly all of these tools involve code, so basic programming ability is a must. It is possible to create great visualizations without code, but it is difficult to articulate new visual forms without it.

Process — We all work with data, defined as structured information. It takes a certain mindset to appreciate a well-structured, honest data set. Ultimately, we encode that data into visual form, a process that requires another, similar mindset to appreciate. So we have data and data-visual mapping in common. But governing each of these steps are many rules, usually documented as algorithms in the software we write: scripts to parse data, programs to generate charts and graphs, and applications to share beautiful renderings with our audiences. The algorithm rules every step. Our core value is a love and appreciation for process itself.

Curiosity — I have never met an incurious practitioner. We love learning and we love being inspired by discovering things in the world around us, or perceiving old ideas in new ways. Data visualization is fundamentally about making the invisible visible, a shared goal for all practitioners. Where our work diverges is in the intent of our process, and in what means of visual rhetoric are employed to that end.

The Value of Interdisciplinary Practice

While the interdisciplinary nature of the practice makes it hard to summarize the field to outsiders, it is also one of our biggest strengths. By drawing on the discoveries and expertise of many fields, we can improve our processes and improve our designs. One concern, of course, is that we may be inclined to learn broadly, but not deeply. Yet, as I described earlier, many practitioners tend to have formal training in one or two different areas, but then more loosely explore others.

Contrast this interdisciplinary approach to a more specialized one. Certainly, there is value in being deeply focused on just a single research area, but such a focus will not by itself produce informative or inspiring visualizations. Domain-specific research — such as in human visual perception, computer graphics, and new visual forms —

however, is invaluable for visualization practitioners. The evolution of our own practice depends on the insights developed by such research.

The interdisciplinary mindset pervades practitioners' selection and use of tools, methods (processes), and domains of operation (uses of tools and methods). Data visualization practitioners are often hired by domain experts (the clients) to interpret and represent the client's data. When Pitch Interactive is hired by *Popular Science* to visualize historical government projections for energy independence, they are not expected to have prior knowledge on energy independence. When Stamen partners with the nonprofit organization Climate Central to map projected sea level rise, no one expects them to be climate change experts. When Fathom contracts with Thomson Reuters to map the power dynamics in China's political sphere, it is the client, not the design firm, who is expected to bring the domain-specific knowledge (and data, of course) to the table.

It's an odd role for a consultant, whose area of expertise is not the specific domain at hand, but an expertise in the *process of exploring* — that is, exploring both the new information provided by the client as well as a range of visual forms for representing and communicating that data.

This *exploration*, as fueled by practitioners' curiosity, rigorously structured processes, and media expertise, is what makes data designers so uniquely valuable today.

Does this mean that a data designer, if inserted into any industry or context, could bring value to the organization simply through her *interdisciplinary process*, even without a specific end goal in mind? Possibly. Designing is problem-solving, and the process itself may be just as important and valuable as the resulting product.

Finally, this also explains why practitioners struggle to articulate their daily work to outsiders: A *process* is much more abstract and difficult to explain than a *product*. It is easy to point to images and say "I made these." It is much harder to say "Through years of practice, I have developed a process that guides my decisions and actions, which results in a successful representation of data, more often than not." Good luck offering such an accurate, yet uninteresting explanation in a social context, such as at a cocktail party — you may not be invited back!

Future Challenges

Looking ahead, what are the future challenges for data design? I see several, each related to the issues addressed above.

Tools — The vast array of tools available will continue to grow and diversify. So the problem is not a dearth of tools; it's cataloging them and making efficient use of the existing tools appropriate to any given task. It seems every week a new framework or library is introduced that provides an improved solution to a very specific problem. Just

as we can't all be experts in every field, we can't all learn how to use every tool (much as we would like to). We need a better method to identify the best tools for a given task. Whatever that method is, it needs to fit into existing workflows.

Methods — Speaking of methods, we each have our own working process, and our challenge is to develop clearer language around those processes. With better language, we can compare processes and learn what, exactly, certain practitioners do that makes their work more successful (or not) than others. What are data design's best practices? How similar or different are they when making data art, as opposed to data visualizations? I hope that this essay is a small step toward framing that discussion.

Data Design Literacy — It is essential to clarify our best practices so that we can educate new practitioners. I began with the question, "How should I get started with data visualization?" Students and others new to the field deserve better answers to this important question. We need maps and taxonomies of practice (which this essay seeks to introduce), and we need more structure and consistency in our training programs. Although data design has a long history, in this rapidly changing environment, it often feels like we are just figuring things out for the first time.

Data Image Literacy — Practitioners are not the only ones who need to be educated; informed audiences are also essential. The consumers of data design must understand the possibilities and pitfalls of the images we create. Just as media literacy education seeks to ensure critical awareness of film, television, and radio, data image literacy is needed to ensure that the inherent biases of data images are well-understood.

Ethics — While there is a tendency to trust data images as fact, practitioners know that even minor changes to a design can strongly influence how the underlying story or information is perceived. Given the ease with which charts and maps can be made to lie, there may be a need for a professional code of visual ethics, a formalization of already well-known design principles advocating for representations that align with human perceptual abilities.

Given my colleagues' innate curiosity, enthusiasm, and love of process, I am optimistic that data explorers all along the spectrum will engage with each other to tackle these issues.

The Nature of Tools

Among hammers, there are minor variations in form, weight, and size. Yet all hammers share a similar fundamental form. Over time, a builder develops a feel for a particular hammer, sensing how much force is needed to move a nail into position.

Software-based tools are more diverse. They share only fundamental underpinnings, such as the use of computation and some common interface conventions. Despite expressing no obvious physical form, they encourage the development of limited muscle

memory, perhaps for common keyboard shortcuts or method patterns. Over months or years of use, a favorite tool or suite of tools will often emerge, and a data designer will gradually develop expertise with that tool, having cultivated a practiced sense for how to strike a particular type of nail.

Yet with so many software tools available, it can be overwhelming to know where to begin. New practitioners are not yet attached to any particular tool; they want to choose an approachable tool, the mastery of which will be transferable to other such tools in the future. Unfortunately, software is not as straightforward as hammers. Learning to code in one language may familiarize you with core concepts — variables, arrays, logic, functions — but switching to another language involves different syntax and methods, different best practices and frameworks, often a very different way of approaching the problem entirely. (Worst-case scenario: moving from Python to Java. So many semicolons!) Every time we switch tools, we have to re-learn how to strike the nail.

Even worse, our favorite software-based tools may change themselves right underneath our noses, auto-updating to add new features, remove old ones, modify syntax rules, or change operating requirements. For some people, this would be crazy-making, and certainly, in the physical world, it would be. Imagine a hammer that, after having been used successfully for years on multiple projects, is considered "trusty" — a reliable workhorse that has supported the builder in a variety of scenarios. But this hammer is an open-source hammer, with a core group of five or six dedicated contributors. They actively patch bugs and introduce new features, so every few months or so we get another point release — Hammer 1.1, Hammer 1.2, and so on. With each release, our hammer is still recognizable, but functions a bit differently; we must adjust the angle of our strike. Hammer 2.0 brings new operating requirements; our old, dingy workshop is no longer supported, so the hammer just sits there, inoperable, until we repaint the walls, install better lighting, or move to an entirely different neighborhood. Of course, Hammer 1.9 is still available for download, and we have a hundred copies sitting around on shelves, but it doesn't drive nails as quickly, precisely, or elegantly. Also, there is market pressure; the hot design firms are not interested in practitioners using old technology.

I present this software-hammer metaphor as further illustration of the *intense curiosity* and *enthusiasm for problem-solving* exhibited by data design practitioners. We enjoy exploring data and learning about the world around us, but we are also excited about new tools, as well as continuous evolution and change in our existing tools. If every project is just another puzzle to be solved, we also secretly enjoy the geekery of solving the *process puzzle*, the ongoing meta-challenge we all share, the operating context inherent to an interdisciplinary practice powered by computation.