PART IV
DATA VISUALIZATION

ON DATA VISUALIZATION

"The earliest seeds of visualization arose in geometric diagrams, in tables of the positions of stars and other celestial bodies, and in the making of maps to aid in navigation and exploration."[1]

"There are two primary sources of potential error in numerical algorithms programmed on computers: that numbers cannot be perfectly represented within the limited binary world of computers, and that some algorithms are not guaranteed to produce the desired solution."[2]

Accordingly, I fixed upon a tree of average bulk and flower, and drew imaginary lines—first halving the tree, then quartering, and so on, until I arrived at a subdivision that was too large to allow of my counting the spikes of flowers it included. I did this with three different trees, and arrived at pretty much the same result: as well as I recollect, the three estimates were as nine, ten, and eleven. Then I counted the trees in the avenue, and, multiplying all together, I found the spikes to be just about 100,000 in number. Ever since then, whenever a million is mentioned, I recall the long perspective of the avenue of Bushey Park, with its stately chestnuts clothed from top to bottom with spikes of flowers, bright in the sunshine, and I imagine a similarly continuous floral band, of ten miles in length.[3]

"Many different words can be used to describe graphic representations of data, but the overall aim is always to visualize the information in the data and so the term *Data Visualization* is the best universal term."[4]

"For no study is less alluring or more dry and tedious than statistics, unless the mind and imagination are set to work or that the person studying is particularly interested in the subject; which is seldom the case with young men in any rank in life."[5]

Graphical displays should

- show the data
- induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphic production, or something else
- avoid distorting what the data have to say
- present many numbers in a small space
- make large data sets coherent
- encourage the eye to compare different pieces of data

- reveal the data at several levels of detail, from a broad overview to the fine structure
- serve a reasonably clear purpose: description, exploration, tabulation, or decoration
- be closely integrated with the statistical and verbal descriptions of a data set.[6]

"So geographers, in Afric maps,
With savage pictures fill their gaps,
And o'er unhabitable downs
Place elephants for want of towns."[7]

"The interior decoration of graphics generates a lot of ink that does not tell the viewer anything new. The purpose of decoration varies—to make the graphic appear more scientific and precise, to enliven the display, to give the designer an opportunity to exercise artistic skills. Regardless of its cause, it is all non-data-ink or redundant data-ink, and it is often chartjunk."[8]

The process of seeking relationships within a data set—of seeking accurate, convenient, and useful summary representations of some aspect of the data—involves a number of steps:

- determining the nature and structure of the representation to be used;
- deciding how to quantify and compare how well different representations fit the data (that is, choosing a "score" function);
- choosing an algorithmic process to optimize the score function; and
- deciding what principles of data management are required to implement the algorithms efficiently.[9]

"One of the major themes for 2010 was using data not just for analysis or business intelligence, but for telling stories. People are starting to make use of the data (especially government-related) that was released in 2009, and there was a lot more data made available this year (with plenty more to come). There were also more visualization challenges and contests than I could count."[10]

Notes

1. Michael Friendly and D.J. Denis, "Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization." Online: www.datavis.ca/milestones (accessed December 20, 2010).
2. Micah Altman *et al.*, *Numerical Issues in Statistical Computing for the Social Scientist*, Hoboken, NJ: John Wiley & Sons, 2004: 2.
3. Francis Galton, *Hereditary Genius: An Inquiry into Its Laws and Consequences*, 2nd edn., London: Macmillan, 1892: 11.
4. Antony Unwin *et al.*, *Graphics of Large Datasets Visualizing a Million*, New York: Springer, 2006: 4.
5. William Playfair, *The Statistical Breviary*, London, 1801.
6. Edward Tufte, *The Visual Display of Quantitative Information*, Cheshire, CT: Graphics Press, 1983: 13.
7. Jonathan Swift's indictment of seventeenth-century cartographers, "On Poetry: A Rhapsody," *The Literature Network*, 1773. Online: www.online-literature.com/swift/3515 (accessed on December 21, 2010).

8. Tufte, op. cit.: 107.
9. David Hand *et al.*, *Principles of Data Mining*, Cambridge, MA: MIT Press, 2001: 3.
10. Nathan Yau, "10 Best Data Visualization Projects of the Year: 2010." Online: http://flowingdata.com/2010/12/14/10-best-data-visualization-projects-of-the-year-%E2%80%93-2010 (last updated December 14, 2010).

Bibliography

Altman, Micah, Jeff Gill, and Michael P. McDonald. *Numerical Issues in Statistical Computing for the Social Scientist*. Hoboken, NJ: John Wiley & Sons, 2004.

Friendly, Michael and D.J. Denis. "Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization." Online: www.datavis.ca/milestones (accessed December 20, 2010).

Galton, Francis. *Hereditary Genious: An Inquiry into Its Laws and Consequences*, 2nd edn., London: Macmillan, 1892.

Hand, David, Heikki Mannila, and Padhraic Smyth. *Principles of Data Mining*. Cambridge, MA: MIT Press, 2001.

Playfair, William. *The Statistical Breviary*. London, 1801.

Swift, Jonathan. "On Poetry: a Rhapsody." *The Literature Network*. 1773. Online: www.online-literature.com/swift/3515 (accessed on December 21, 2010).

Tufte, Edward. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphic Press, 1983.

Unwin, Antony, Martin Theus, and Heike Hofmann. *Graphics of Large Datasets Visualizing a Million*. New York: Springer, 2006.

Yau, Nathan. "10 Best Data Visualization Projects of the Year: 2010." Online: http://flowing-data.com/2010/12/14/10-best-data-visualization-projects-of-the-year-%E2%80%93-2010 (last updated December 14, 2010).

7 SUPERFUND365

Brooke Singer

Key Words: CERCLIS, Citizen Science, Co-production, Database, Data Visualization, EPA, Flash, Open Government, Superfund, XML

Project Summary

Superfund365 is an online, data-visualization application that consists of 365 Superfund sites—i.e., the nation's worst hazardous waste sites as designated by the Environmental Protection Agency (EPA). For one year, *Superfund365* featured a site a day, graphically depicting public data such as present contaminants, responsible parties, and regional demographics along with text and image contributions from users.

Project Developer Background

This is a project that fell into my lap. I did not necessarily have time for it. I was not looking to take on something new. But sometimes a statement of fact, a passing comment, or even an image can trigger such intense curiosity that you find you are in pursuit of answers and, before long, you are producing a new body of work.

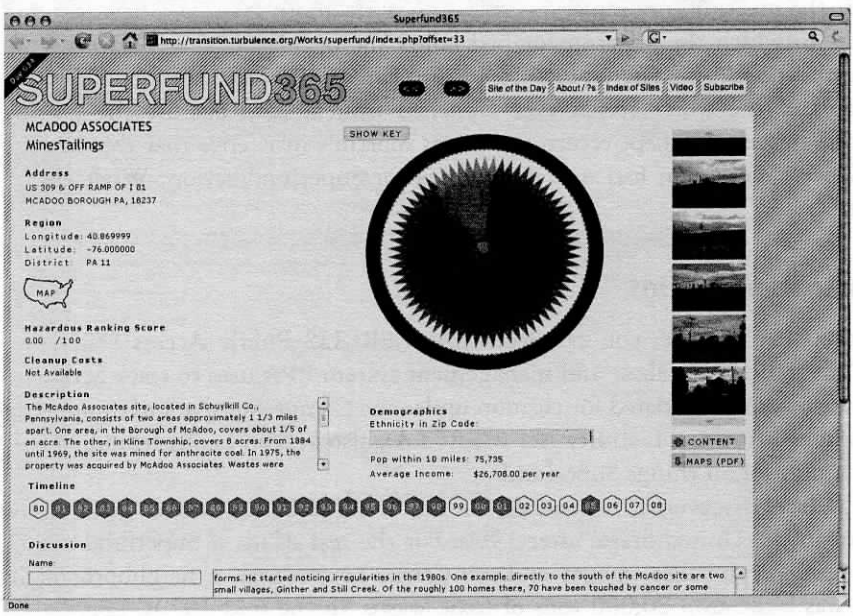

Figure 7.1 Superfund365.org screenshot (McAdoo Associates, day 34 of *Superfund365*).

In early 2006, I met Robert J. Martin, the former National Ombudsman for the EPA, at a pub near Times Square to discuss air sampling in New York City. I had just been awarded a commission with my collective, Preemptive Media, from Eyebeam Art + Technology Center and the Lower Manhattan Cultural Council to build portable air-monitoring devices. The project, which was launched in 2007 as *Area's Immediate Reading (AIR)*, started with the basic premise that the air-quality map of New York City would look radically different if the monitoring devices that supply the data were (a) mobile and (b) in the hands of citizens rather than located at fixed stations selected by the government. The purpose was to build micro air-quality reports that are based on individual paths through the city and that could be aggregated dynamically to determine general readings, an alternative to the Air Quality Index.[1]

A professor of politics and environmental studies at Ithaca College told me that anyone designing a project to measure air quality in New York should contact Robert Martin. From 1992 to 2002, Martin served as the EPA National Ombsudsman or mediator between the American public and the US government agency that is charged with protecting human health and the country's natural environment. On 9/11, Martin was driving to his office at the EPA headquarters in Washington, DC, when the World Trade Center collapsed and the streets of New York were blanketed in a record amount of the most toxic substances known to man.

Shortly after we met, Martin said that he could not help much with my technical questions about monitoring hardware and testing for contaminants; he referred me to his former colleague, a scientific advisor, for that information. Instead, he told me what it was like at the EPA in the days following 9/11. While Americans mourned and worried about further attacks, EPA employees immediately recognized the gravity of the environmental impact, swiftly mobilized a team of first responders, and internally debated whether all of Lower Manhattan should be declared a Superfund site, a site so hazardous that it warranted the relocation of inhabitants and immediate cleanup. Martin paused and rephrased his last comment: all of Lower Manhattan qualified for Superfund in 2001.

When I left the meeting, my mind was spinning with information and new insights into the extent to which the government had failed the people of New York and the extreme measures the administration took to ensure that Wall Street would be functioning within days of the worst attack ever launched from abroad against the United States.[2] The one statement I kept returning to was Martin's insistence that the damage inflicted on Lower Manhattan had met the criteria for Superfund action. What *exactly* did that mean?

Introduction to *Superfund365*

If you visit the EPA website, you can access the CERCLIS Public Access Database.[3] CERCLIS is "the national database and management system EPA uses to track activities at hazardous waste sites considered for cleanup under the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA), also known as Superfund."[4] This is the official source for all things Superfund.

Superfund, I have discovered, is generally an unknown term to younger generations or people born in the United States after 1980. For the rest of us, if Superfund means anything, it probably invokes images of Love Canal. Love Canal, a neighborhood in Niagra, NY, was built atop 20,000 tons of toxic waste. In the mid-1970s, Lois Gibbs started a movement in her community by going door to door polling people about their

health problems after her son fell sick and she learned his elementary school was directly above the contamination hot zone. After several years of demanding protection and fed up with government inaction, the community held hostage several EPA officials. This garnered national media coverage and public outrage.[5]

In the wake of these events at Love Canal, the Federal Superfund program was enacted "to address abandoned hazardous waste sites ... [allowing] the EPA to clean up such sites and to compel responsible parties to perform cleanups or reimburse the government for EPA-led cleanups."[6] At that time in the early 1980s, regulators thought there were only a small number of sites eligible for Superfund designation and that the original $1.6 billion trust would be ample to fund the program. But more sites qualified for Superfund status than anticipated, and each year many more sites were added to the list than cleaned up. As of May 2010, there were a total of 1,279 sites on the National Priorities List (NPL) of the worst hazardous waste sites as identified by Superfund.[7] Thousands more sites, specifically 13,253, were active or are under consideration for NPL designation, while another 35,593 were archived sites labeled as "no longer of interest" in the database.[8]

The interface to CERCLIS is a search form offering numerous points of entry into this massive database (see Figure 7.2). Like many other online databases, CERCLIS is not particularly welcoming: it is text heavy, it does not encourage browsing, and, although a lot of information can be extracted, it does not easily produce understanding. It reminds me of my dislike for shopping malls; I will enter only if I know exactly what I want and where it is so I can get in and out as quickly as possible. My initial curiosity about Superfund was dampened by the data deluge. I began to think about user interfaces and the experience of this information on screen versus my one-on-one conversation with Martin. How could I bring the two closer together? Just how public, in fact, was this public database?

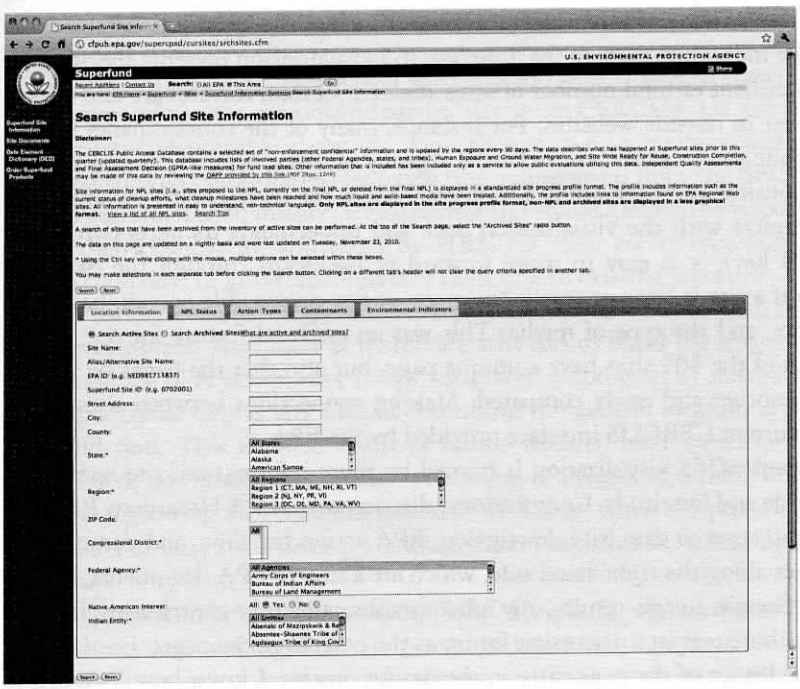

Figure 7.2 A view of the CERCLIS interface.

Technical Description

From the start, I was convinced that I should follow a one-a-day format for several reasons. There was so much data that highlighting a new site each day created a manageable pace and avoided overwhelming my visitors. A one-a-day format also allowed a narrative to unfold and I pushed this even further by structuring the project as a travelogue. I knew I would not be able to spend a whole year visiting a different Superfund site each day and the project did not actually require this, but the personal descriptions of place and the "just-snapped" photographs brought the sites to life. In other words, the data became more meaningful and had more of an impact when presented from a human perspective. My hope was that as the journey moved further from my home in New York City, I could use social networking and media attention to attract others to pick up the route with me and, eventually, for me.

When looking at the data, I decided that the main attributes of a Superfund site as described by CERCLIS consisted of the contaminants of concern, the contaminated media (or where the contamination is found), the total site acreage, and the responsible parties. Therefore, this set of data makes up the primary visual of *Superfund365*, which is colorful and flower-like. It is an animation that unfolds and beckons continuing exploration.

As I was designing, I was consciously using beauty as a lure. Specifically I was reacting to the BP logo and rebranding efforts. In 2000, BP adopted a bright green and yellow sunburst or helios for its logo because, in the company's own words, it symbolizes "dynamic energy in all its forms." The same year the company shed the name British Petroleum for BP to allow for new associations like "better people, better products, big picture, beyond petroleum."[9] BP's logo and rebranding are acts of visual veiling; we are instructed to look beyond (beyond the dirty and destructive practices of oil extraction) toward better (meaning select and cheery) outcomes.[10] My visualization is an ironic riff on the BP sunburst. My visual is not there to distract and deceive with manufactured marketing ploys, however, but rather it is intended to entice and lure users into content such as the CERCLIS database.

Scrolling over the individual parts of the *Superfund365* visualization presents specific data like a present contaminant or total number of acres affected. Certain features are also clickable and take a viewer to outside websites. For instance, many of the contaminants link to the appropriate public health statement from the Agency of Toxic Substances and Disease Registry and a responsible party (typically a business) link to the representative website.

After a user interacts with the visualization once and understands its parts (with help from tooltips and a key), it is easy to move forward or backward through the days and immediately grasp if a site is large or small, has many or few responsible parties, has many or few contaminants, and the type of media. This was an essential feature for me. It was important that each of the 365 sites have a unique page, but also that the pages be viewed in relation to one another and easily compared. Making connections between sites is not possible using the current CERCLIS interface provided by the EPA.

The primary *Superfund365* visualization is framed by more information: site name, site type, address, latitude and longitude, Congressional district, map, EPA Hazardous Ranking Score (HRS), cleanup costs to date, site description, EPA action timeline, and demographics. There are images along the right-hand side, which are a mix of EPA documents, historical photographs, Google image results, my photographs, and user-contributed images. Users can also contribute text in a discussion forum at the bottom of the page.

I determined the layout of the page early in the design process. I knew how I wanted it to look and function only after a few days of studying CERCLIS. It then took three months

of solid work with a research assistant, programmer, and business analyst to build the online application and get *Superfund365* up and running.

Most of the work was fairly tedious. The data contained in the CERCLIS database is public and freely available for download but only in a format called CSV or comma-separated values. This format is not the most current and there is a greater potential for making errors when handling it compared with XML. Also, not all of the data I wanted to incorporate in my application was available as one neat download. My assistant, Emily Gallagher, and I had to call the Superfund hotline many times for help.

After downloading numerous datasets from CERCLIS and calling the hotline, I realized that we would have to manually insert a substantial amount of data into our spreadsheets (which were to become the *Superfund365* database). This meant copying and pasting information from the EPA website or searching through PDF documents. When information was entirely missing, we had to request it by email or a phone call to the site manager.

I was most surprised to find that present contaminates for the sites are not in CERCLIS. All the Superfund sites included in *Superfund365* are on the NPL, which means the contaminant lists do exist. Sometimes the information was posted on the regional EPA website, but when we had to call a manager for the information, we would ask why it was not online. The most frequent response from site managers was that their contaminant list was not final and more in-depth testing was needed.

The site managers were generally very willing to help and pleased that someone was expressing interest in their site. When we told them what we were doing, they were interested and supportive mostly because we would be drawing attention to their work and the problem of Superfund.

Piecing the information together took a lot of time, as did cleaning up the data. We found a lot of inconsistencies and inaccuracies—so many that we did not attempt to fix them all. We tried to standardize the contaminant list since it was central to the project. One contaminant can be spelled in several different ways, which, of course, undermines the integrity of the database and makes it impossible to search for how often a single contaminant appears in all 365 sites.[11]

Kurt Olmstead, a business school graduate and skilled programmer, tackled the irregularities apparent in the responsible parties' dataset in CERCLIS. He was able to generate the most current responsible party (as these change due to buyouts and spinoffs or can be obscured through subsidiaries) by extracting information from the SEC EDGAR database, Wikipedia, and Yahoo finance. By cleansing and standardizing the responsible party names, it was then possible to rank them by HRS, list their entity type (e.g., public company, private company, or government agency) and link to existing websites.

This narrative should suggest the extreme tedium that comes with data entry and database development. And there is more! We still had to build the image database. I wanted to include all EPA documents (maps, diagrams, photographs) related to the sites in the image gallery as well as any images describing the general location or the actual site that we could find. This entailed hours of online research (and sometimes more emails and phone calls), downloading and batch processing of images, and then, finally, uploading the files to the growing *Superfund365* database. This way, when the project launched, each site would have at least a few descriptive images and I hoped more would be submitted by users over time.

The programmer on the project, John Kuiphoff, was responsible for taking the spreadsheets and writing PHP scripts to generate the *Superfund365* database. That was probably his easiest task. More difficult was translating my initial Illustrator sketch into Flash and making an application that functioned with the *Superfund365* database. The

final outcome looks implausibly similar to my sketch, which says a lot about John's programming capabilities. We decided to use the Flash programming environment because it would guarantee that the application would appear consistently across browsers and most browsers are equipped to view Flash content. Other emerging technologies, like AJAX, were considered but we did not feel they were up to the job at the time. In retrospect, this was a mistake. Flash is a proprietary and inherently closed system; developing the project in such a way that our information could be easily exported, used, and extended by others should have been a high priority.

Sources other than the EPA and CERCLIS were invaluable along the way and became part of the final application. I have already mentioned the Agency for Toxic Substances and Disease Registry (a divison of the Center for Disease Control that was created along with Superfund in 1980). The Center for Public Integrity's investigation, "Wasting Away: Superfund's Toxic Legacy," provided excellent background research, including a list of the most dangerous Superfund sites. A Superfund "not having human exposure under control and/or not having contaminated groundwater migration under control" is classified as most dangerous.[12] The "Wasting Away" research aided me in selecting my 365 Superfund sites from approximately 1,300 on the NPL in 2007. The US Census Bureau provided the demographic information. As Lois Gibbs emphasizes in her video interview with me, you cannot talk about Superfund without talking about class and race.[13]

I have shared a good bit of detail about the building of the visualization application because it is important to understand the level of accessibility and usefulness of the public data that is at the core of the project. EPA officials often apologized to us when we inquired for more and better source data. I was later told by agency officials that *Superfund365* succeeds because it underscored the errors and irregularities of CERCLIS; several officials told me they preferred using *Superfund365* over CERCLIS.

This was not my purpose, but interesting nonetheless. My goal was to visualize and make tangible what was obscure and disparate. Pulling together various sources of public material and thus enabling new pictures and questions to emerge was paramount for me. That is why a user comment on the final day of the project, day 365 or Pearl Harbor Naval Complex, was very rewarding. The user posted:

> Great work!! I have been tuned in regularly and have even worked on a few of the sites mentioned. I thought the Pearl Harbor description was a little weak. How does a sugar company become PRP [primary responsible party] for a naval base location?? Sounds like there is more to the story somewhere.[14]

Historical Perspectives

During the summer of 2007, when I was in the midst of producing *Superfund365*, English media artist, Graham Harwood, came for a studio visit. His reaction was: "Isn't this the business of government, not artists?"—which is a slightly different version of the more usual question, *why is this art?* I replied that I was filling a void, taking on the project in the absence of government, and, thus, aligning myself with environmental activists and non-government organizations (NGOs). In fact, artists—such as Mierle Ukeles, the Harrisons, and Mel Chin—have been doing the same thing for decades now.

With this work, I am not advocating that the EPA hire more sophisticated interface and interaction designers. That misses the point. I do, however, think the EPA (and likely every other US government agency) needs to do a much better job of maintaining (correcting, standardizing, updating) its data. In addition, the EPA needs to provide

data for the public in more useful formats, employing the newest trends in data sharing, manipulation, and visualization.

The fact that the EPA's data is inadequate and its reporting procedures are outdated probably does not come as a big surprise. Government agencies are notoriously slow and under-funded while the emerging tech world is neither. The government is aware that it needs to improve in this area. President Obama, on his very first day in office, released a memo in which he stated:

> All agencies should adopt a presumption in favor of disclosure, in order to renew their commitment to the principles embodied in FOIA [Freedom of Information Act], and to usher in a new era of open Government. The presumption of disclosure should be applied to all decisions involving FOIA.The presumption of disclosure also means that agencies should take affirmative steps to make information public. They should not wait for specific requests from the public. All agencies should use modern technology to inform citizens about what is known and done by their Government. Disclosure should be timely.[15]

The government should make it as easy as possible for *anyone* (including artists, activists, NGOs, students, businesses, watchdogs, homebuyers, the general public) to access, use, and benefit from data. Obama emphasized the importance of openness, modern technology, and timeliness in his January 21, 2009, memo. In May 2009, the administration delivered on its promise and launched Data.gov, a user-friendly repository for all information the government collects. This is a positive first step, but there is also the issue of quality. For the producers of *Superfund365* and EPA officials working on Superfund, my project reveals the failings of government disclosure even when there *is* openness and public reporting. This is not apparent to most users of the site.

Discussing data quality can be straightforward. The quality of the EPA Superfund data would be improved greatly if the agency adhered to consistent names for present contaminants. If some data is less reliable than other information, that should be noted and the data should be disseminated anyway for whatever value it might provide.

Thinking about quality begs more difficult questions. What are the assumptions and ideologies that determine inclusion? What are the categories for classification and who is doing the classifying? With the Superfund program, I was told repeatedly that obtaining Superfund status is always a highly politicized process. The players typically are the government (federal, state, and local), industry (potential responsible parties), developers, community organizations, individual residents, and others. The conflicting interests, negotiations, and struggles are reflected in the CERCLIS database. For example, Lois Gibbs of Love Canal shared a fascinating story of why in the 1980s it was determined that a site must have a HRS of 28.5 or higher in order to qualify for Superfund designation. The decision had everything to do with politics and nothing to do with science.[16]

Gilles Deleuze reminds us, "machines are social before they are technical."[17] Too often we believe our data is purely technical and forget its social dimension. Once someone or something (since automatic, computerized data creation is becoming the norm) translates our world into data, that data holds power over us. It makes decisions for us. It triggers action—or compels inaction. Lisa Gitelman writes about this problematic from a slightly different perspective in her book, *Always Already New: Media, History and the Data of Culture*. She states: "Media are frequently identified as or with technologies, and one of the burdens of modernity seems to be the tendency to

essentialize or grant agency to technology."[18] "Data" could replace "media" in this quotation and the words are still true.

The list of academics and artists critiquing our data culture is long, but the conversation has also entered popular media. John Allen Paulos discusses the supremacy of data and questions the aptness of pervasive reliance on data-driven decision-making in a recent article in the *New York Times* called "Metric Mania: Do We Expect Too Much from Our Data?" He ends his discussion with these words: "This doesn't mean we shouldn't be counting—but it does mean we should do so with as much care and wisdom as we can muster."[19]

What is needed is a change in cultural perception of data and its technical underpinnings. Such a transformation occurred in the last century with respect to photography. In 1973, Susan Sontag eloquently debunked the myth of photography in her book *On Photography*. She described our relationship to photographs in this way:

> Photographs furnish evidence ... A photograph passes for incontrovertible proof that a given thing happened. The picture may distort; but there is always a presumption that something exists, or did exist, which is like what's in the picture. Whatever the limitations ... a photograph—any photograph—seems to have a more innocent, and therefore more accurate, relation to visible reality than do other mimetic objects.[20]

I like to think *Superfund365* is playing a role in debunking the database myth. In other words, it challenges our perceived relationship to the data, it reveals the structure of the database and allows for new configurations and correlations. It begets new articulations and, as a result, new dialogues. It confuses as much as explicates. It challenges the very concept of incontrovertible proof. It simultaneously pays homage to the existence of public data and encourages government to improve. This is not about being for or against counting, but is rather a call for greatly expanding who is doing the counting and how results are circulated to support multivalent uses.

As I was writing the conclusion to this chapter, I ran into Tim Dye at a locative media symposium in San Francisco. Tim Dye is a Senior Vice President at Sonoma Technology, Inc., where he helped to develop, and now manages, AIRNow. AIRNow is an EPA program that makes accessible air-monitoring data from across the country via the Internet. Dye knew about Preemptive Media's work developing portable air-monitoring devices, and he told me that within five years or so he believes such affordable and off-the-shelf devices will be publically available. In anticipation of this, Dye envisions the AIRNow program will support the use of these devices by citizen scientists across the country and will encourage the upload and integration of citizen data into the government's air-quality database. AIRNow would be co-produced by government and citizens alike. Independent data collection and visualization projects would not be forced into an adversarial role with the government but rather the relationship could be collaborative and mutually informative.

Jason Coburn describes the importance of co-production like this in his book *Street Science: Community Knowledge and Environmental Health Justice*:

> [The] co-production model problematizes knowledge and notions of expertise, challenging hard distinctions between expert and lay ways of knowing ... [The] co-production model emphasizes that when science is highly uncertain, as in many environmental-health controversies, decisions are inherently trans-science—involving questions raised by science but unanswerable by science alone.[21]

This model and Dye's wishful approach to AIRNow is the future for data collection and visualization. It is democratic and responsive. It allows for numerous perspectives from professionals and citizens. Science is not accepted uncritically as fact but is co-produced by numerous stakeholders because the risks are high and solutions are urgently needed.[22]

Conclusions and Outcomes

The focus of this chapter has been the original purpose of *Superfund365* and the production process for developing it. I have described the project as a data visualization application and what it means for me, as an artist, to use government-generated data as my medium.

But *Superfund365* is also a platform, meaning it is not only a place to visit and learn but also to share and connect with others for possible action. On the day it was launched, September 1, 2007, the project was a starting point rather than a completed work. From that day forward, the data visualization gave way to collective storytelling.

It is beyond the scope of this chapter to discuss in any detail the life of *Superfund365* after it was launched, but I will say that after staring at the EPA data on screen for several months, it was almost startling to visit the locations in person. The first several weeks, I traveled to every Superfund site highlighted in *Superfund365*, beginning with Quanta Resources in Edgewater, NJ, which is directly across the Hudson River from the Upper West Side. I expected to find the sites cordoned off and dotted with scientists in Hazmat suits, but this was not the case (see Figure 7.3). Very rarely were the places even marked as Superfund sites and only once did I see any cleanup activity. I found myself in such places as abandoned lots with DO NOT ENTER signs, the largest shopping mall in New York State, an abandoned softball field, an empty square block in a city center, and a Home Depot store. These were, I concluded, common, everyday places.

For this reason, Alex Prud'homme's 2010 Op-Ed in the *New York Times* resonated with me. He wrote:

> We tend to think of oil spills as dramatic events—a sinking ship, a burning rig. So it's easy to forget that across the country, hundreds of spills, many left over

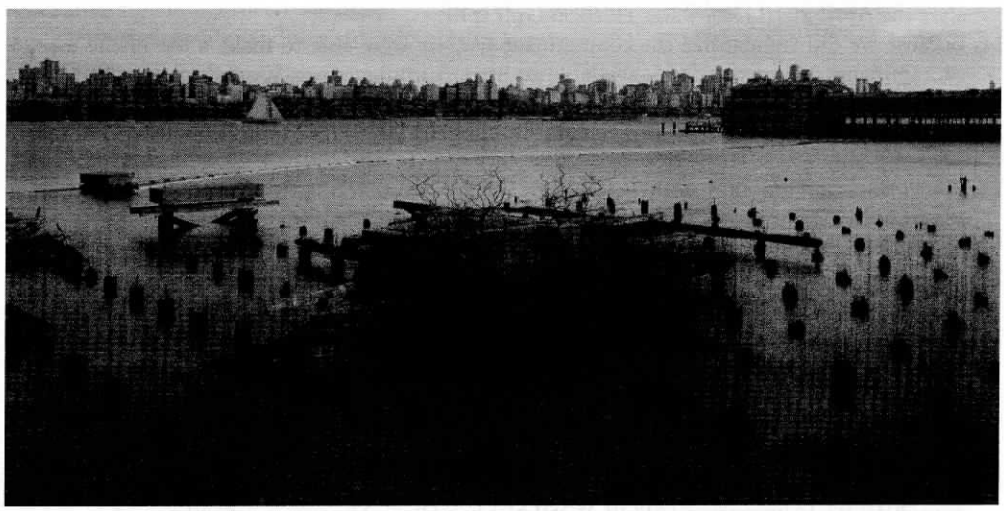

Figure 7.3 Quanta Resources Superfund Site, Bergen County, NJ (day 1 of *Superfund365*).

> from a less regulated time, continue to poison groundwater and leak toxic fumes. Instead of letting the Gulf spill divert our attention yet again from slow-moving disasters like Newtown Creek, we should take it as an impetus to address problems much closer to home.[23]

And who better to force attention to these sites, tell their histories, explain the consequences of contamination, and suggest solutions for building a better future than the people who call these places home. With one in four Americans living within four miles of a Superfund site, there is a lot of local expertise to be tapped.[24] The convergence of social media and off-the-shelf monitoring hardware offer tremendous potential for co-production. The challenge remains to transform this kind of knowledge production into actionable solutions.

Notes

1. See www.pm-air.net.
2. On Earth Day 2002, Robert Martin resigned as National Ombudsman. He left the EPA after Administrator Christine Todd Whitman eliminated his office's independence—and therefore its function as internal watchdog—through restructuring. Weeks before, Martin had issued a very public and frank condemnation of the EPA's handling of Lower Manhattan post-9/11 after holding two extensive public hearings. No replacement was ever hired and to this day there is no National Ombudsman at the EPA. To hear Robert tell his story, watch my video interviews with him (link).
3. Cfpub.epa.gov: http://cfpub.epa.gov/supercpad/cursites/srchsites.cfm (accessed May 18, 2010).
4. U.S. EPA: http://epa.custhelp.com/cgi-bin/epa.cfg/php/enduser/std_adp.php?p_faqid=173&p_created=1065036508 (accessed May 18, 2010).
5. To hear the story of Love Canal told by Lois Gibbs, watch the video interview on Superfund365.org: http://turbulence.org/Works/superfund/video.html.
6. Epa.gov: www.epa.gov/superfund/about.htm (accessed May 24, 2010).
7. Epa.gov: www.epa.gov/superfund/sites/query/queryhtm/nplfin1.htm (accessed May 24, 2010).
8. These numbers are in reports run one a month by the EPA and available for download from here: http://epa.gov/superfund/sites/phonefax/products.htm.
9. Bp.com: www.bp.com/sectiongenericarticle.do?categoryId=9014508&contentId=7027677 (accessed May 24, 2010).
10. The extent of the dirty and destructive oil-extraction practices of BP came into sharp focus after the April 2010 Deepwater Horizon Gulf spill.
11. Because we did standardize the contaminant list, we were able to make a list of the top 25 contaminants of Superfund365, available at: http://turbulence.org/Works/superfund/top_25_cont.html.
12. Publicintegrity.org: http://projects.publicintegrity.org/Superfund/HumanExposure.aspx (accessed May 25, 2010).
13. See video interview 003 with Lois Gibbs on Superfund365: http://turbulence.org/Works/superfund/video.html.
14. Turbulence.org: http://turbulence.org/Works/superfund/index.php (accessed May 25, 2010).
15. Eff.org: www.eff.org/deeplinks/2009/01/on-day-one-obama-demands-open-government (accessed May 25, 2010).
16. See video interview 001 with Lois Gibbs on Superfund365: http://turbulence.org/Works/superfund/video.html.
17. Gilles Delueze, *Foucault*, trans. Sean Hand, Minneapolis, 1988: 13.
18. Lisa Gitelman, *Always Already New: Media, History and the Data of Culture*, Cambridge, MA: MIT Press, 2006: 2.
19. John Allen Paulos, "Metric Mania," *New York Times*, May 10, 2010. Online: www.nytimes.com/2010/05/16/magazine/16FOB-WWLN-t.html?scp=1&sq=metric%20mania&st=cse (accessed May 25, 2010).

20. Susan Sontag, *On Photography*, New York: Bantam Doubleday Dell, 1973: 5.
21. Jason Corburn, *Street Science: Community Knowledge and Environmental Health Justice*, Cambridge, MA: MIT Press, 2005: 41.
22. For more on this topic, see Silvio O. Funtowicz and Jerome R. Ravetz, "Science for the Post-normal Age," *Futures*, 25, 7, September 1993: 739–755.
23. Alex Prud'homme, "An Oil Spill Grows in Brooklyn," *New York Times*, May 14, 2010. Online: www.nytimes.com/2010/05/16/opinion/16Prudhomme.html?scp=1&sq=Alex%20Prud%27homme&st=cse (accessed June 10, 2010).
24. Americanprogress.com: www.americanprogress.org/issues/2006/06/b1777527.html (accessed June 18, 2010).

Bibliography

BP. "BP brand and logo." Bp.com: www.bp.com/sectiongenericarticle.do?categoryId=9014508&contentId=7027677 (accessed June 21, 2010).

Center for American Progress. "The Toll of Superfund Neglect." Americanprogress.org: www.americanprogress.org/issues/2006/06/b1777527.html (accessed June 21, 2010).

Center for Public Integrity. "Most Dangerous Superfund Sites." Publicintegrity.org: http://projects.publicintegrity.org/Superfund/HumanExposure.aspx (accessed June 21, 2010).

Corburn, Jason. *Street Science: Community Knowledge and Environment Health Justice*. Cambridge, MA: MIT Press, 2005.

Electronic Frontier Foundation, "On Day One, Obama Demands Open Government." Online: www.eff.org/deeplinks/2009/01/on-day-one-obama-demands-open-government (accessed June 21, 2010).

Deleuze, Gilles. *Foucault*. Trans. Sean Hand. Minneapolis: University of Minnesota Press, 1988.

Funtowicz, Silvio O. and Jerome R. Ravetz. "Science for the Post-Normal Age." *Futures*, 25, 7, September 1993.

Gitelman, Lisa. *Always Already New: Media, History and the Data of Culture*. Cambridge, MA: MIT Press, 2008.

Paulos, John Allen. "Metric Mania." *New York Times*, May 10, 2010.

Prud'homme, Alex. "An Oil Spill Grows in Brooklyn." *New York Times*, May 14, 2010.

Sontag, Susan. *On Photography*. New York: Bantam Doubleday Dell, 1973.

US Environmental Protection Agency. "Basic Questions." Online: www.epa.gov/superfund/about.htm (accessed June 21, 2010).

US Environmental Protection Agency. "Final National Priorities List (NPL) Sites—by Site Name." Online: www.epa.gov/superfund/sites/query/queryhtm/nplfin1.htm (accessed June 21, 2010).

US Environmental Protection Agency. "Frequent Questions." Online: http://epa.custhelp.com/cgi-bin/epa.cfg/php/enduser/std_adp.php?p_faqid=173&p_created=1065036508 (accessed June 21, 2010).

US Environmental Protection Agency. "Report and Product Descriptions." Online: http://epa.gov/superfund/sites/phonefax/products.htm (accessed June 21, 2010).

US Environmental Protection Agency. "Search Superfund Site Information." Online: http://cfpub.epa.gov/supercpad/cursites/srchsites.cfm (accessed June 21, 2010).

Links

http://cfpub.epa.gov/supercpad/cursites/srchsites.cfm
www.data.gov
http://turbulence.org/Works/superfund/index.php
http://turbulence.org/Works/superfund/video.html
http://projects.publicintegrity.org/Superfund/HumanExposure.aspx

8 PASTICHE

Christian Marc Schmidt

Key Words: Agency, Analytical, Collective Gestalt, Collective Memory, Data Visualization, Experiential Expression, Hypothesis-generating, Mythic city, Objective C, Psychogeography, Relevance, Rhetoric, Typography

Project Summary

Pastiche is a dynamic data visualization that maps keywords from blog articles to the New York neighborhoods they are written in reference to, geographically positioned in a navigable, spatial view. Keywords, assigned based on relevance, surround their corresponding neighborhoods. The result is a dynamically changing, three-dimensional description of the city, formed around individual experiences and perspectives.

Project Developer Background

For *DesignInquiry*, a design conference I attended in 2006, I created an experimental visualization of 12 so-called "branded communities" around the world that sought to represent the size, density, and perception of these developments. While size and density were known factors, perception was determined by comparing and ranking communities according to three factors: sustainability, status, and security, the relative values of which were derived from extensive research. Through this work I started forming an interest in the quantitative display of qualitative information, and in how to visualize a collective experience. Shortly thereafter I became aware of *We Feel Fine* (wefeelfine.org), by Jonathan Harris and Sep Kamvar. The piece parses instances of the phrase "I feel" from blog articles and visualizes every instance in several alternate views, called "movements." Some movements are more analytical, others more experiential, but the novelty of *We Feel Fine* is that it quantifies statements made about feelings and by doing so portrays the emotional state of an entire community. These were the two perhaps most formative influences for what would later become *Pastiche*.

Introduction to *Pastiche*

Pastiche is a software application I created in collaboration with Ivan Safrin, a Russian-born software developer. Ivan was the developer for the project, and we also collaborated closely on the concept, visuals, and interaction. Our goal was to model New York City according to what people were writing about it, creating an alternate representation by substituting the physical architecture of the city for people's associations, and in turn imbuing them with a kind of tangibility and permanence.

As a designer and media artist, I am interested in understanding the ways in which the aggregate of individual experiences may lead to insights about a larger cultural or societal condition. I think of my work as exploring what I call "collective gestalt"—an archetypal pattern or gesture emerging from a multitude of sources.

This idea is perhaps most clearly illustrated in several video works I completed between 2006 and 2010. These pieces aggregate hundreds of photographs uploaded to Flickr under a Creative Commons "Non-Commercial Share Alike" license, stitched together in a time-based narrative revealing a particular iconic gesture. In each case the subject matter was a familiar space or environmental condition, including prominent New York City landmarks such as the Brooklyn Bridge and the Grand Central Terminal, as well as images depicting horizons. By choosing iconic subjects that viewers would be intimately familiar with, I found that I could comment on both the individual and the collective. Analyzing each image in terms of location and orientation of the photographer, I was able to identify an inherent formal gesture for each subject, such as a particular radial gesture in the Main Concourse at Grand Central, the linear gesture of crossing the Brooklyn Bridge toward Manhattan, and a skyward-trending upward pan in the case of the horizon imagery.

In each of these pieces, the viewer is participatory in experiencing a space as perceived by a collective. *Pastiche* continues this trajectory by establishing a virtual environment formed by collective activity.

Technical Description

Pastiche depicts keywords parsed from blog articles to describe New York City neighborhoods via an immersive, three-dimensional view of the city (Figure 8.1). Visually, the project relies on a typographic display. Keywords are displayed as clusters around the neighborhoods they refer to, which are in turn mapped in relation to the center point of their geographic neighborhood boundaries. All of the typography in the piece is rotated 90 degrees and pointed upward to reference the vertical architecture of the city, while allowing the information to remain legible. Unlike a traditional, horizontal orientation it does not obscure the geographic locations of each neighborhood or keyword, and due to the narrower footprint, the text maintains higher legibility with fewer overlapping letters.

Pastiche is a desktop or "client" application written in Objective C that currently runs on Intel-based Apple Macintosh computers. Initially, we considered using other technologies such as Adobe Flash or Java that would allow viewing the piece in a Web browser; but decided that optimal performance would yield from an application, given the fluid, immersive experience we wanted to create and the complexity of the presented information.

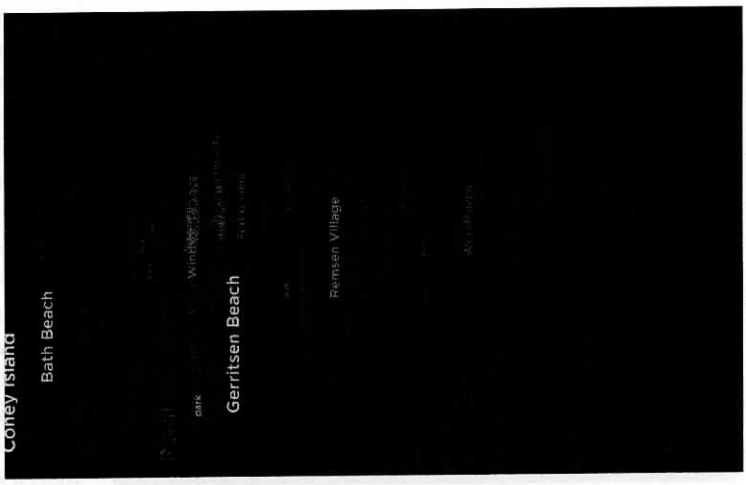

Figure 8.1 *Pastiche* displays keywords associated with New York City neighborhoods in an immersive spatial view.

Ivan wrote *Pastiche* in two parts, consisting of the client and a server *daemon*, a program that continuously runs in the background on a dedicated server. The daemon collects and processes information from Google Blog Search, storing the processed data in a file on the server, updated every 24 hours and read whenever the application is run. This two-tier approach allowed us to both reduce load times and achieve greater accuracy when parsing keywords, because the server daemon gathers data irrespective of the client.

The daemon, developed in the programming language *Python*, reads RSS feeds from Google Blog Search for a total of approximately two hours each day. It compiles a database of top keywords linked to New York City neighborhoods using a search algorithm with a combination of specific neighborhood names and a set of helper keywords that ensures the results are related to the neighborhood. The top keywords are then parsed by analyzing the blog data for the most frequently repeated words, while eliminating unimportant words (pronouns, prepositions, common unrelated words, etc.). These are omitted using a dataset of undesired words and a set of basic conditions, such as word length and capitalization. The final results are stored on the server as an XML (Extensible Markup Language) file, a common data format. The application was written in the programming language C++ using *Substance*, a framework that Ivan authored specifically for developing interactive applications. It downloads the XML file and displays the data using OpenGL, an open-source technology that enables the rendering of three-dimensional graphics.

As an expressive data visualization, *Pastiche* is both experiential and analytical. We asked ourselves what an alternate view of New York would look like—a view defined not by the architecture, but rather by a collective attitude toward city areas that might lead us to better understand a socio-cultural identity.

We arrived at the formal idea for the visualization during an initial meeting. As Rem Koolhaas outlined in *Delirious New York*, Manhattan is the product of hyper-density resulting in unabashed verticality, the city's singular urban ideology (termed "Manhattanism").[1] We were seeking to create a reference to this concept through the use of typography and to contrast *Pastiche* with the physical, architectural environment. This was a critical argument for the piece, and it led us to consider the use of upright typography, which became the central visual gesture (see Figure 8.2). In our minds, *Pastiche* would become less of an objective, analytical visualization tool. Rather, it would have a particular attitude toward the information it conveyed. Therefore, we allowed the formal idea to drive our process—from the interaction paradigm to the way the information was presented.

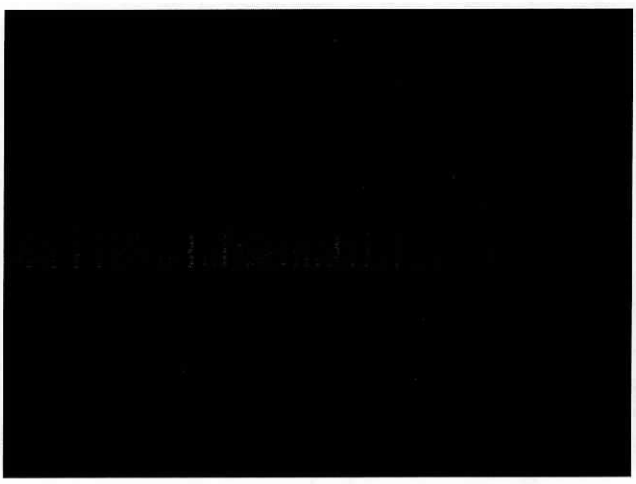

Figure 8.2 *Pastiche* references the vertical architecture of Manhattan.

While the formal idea took shape early on, the final expression resulted from working closely with the data. For instance, we decided to focus on individual neighborhoods, rather than specific addresses or streets and intersections, once we realized that they increased the likelihood of finding suitable content. We identi-

fied the geographic center points of each neighborhood with the help of a city base map and used them as anchor points for neighborhood labels. Positioning these labels relative to their actual geographic locations allowed the shape of the city to emerge.

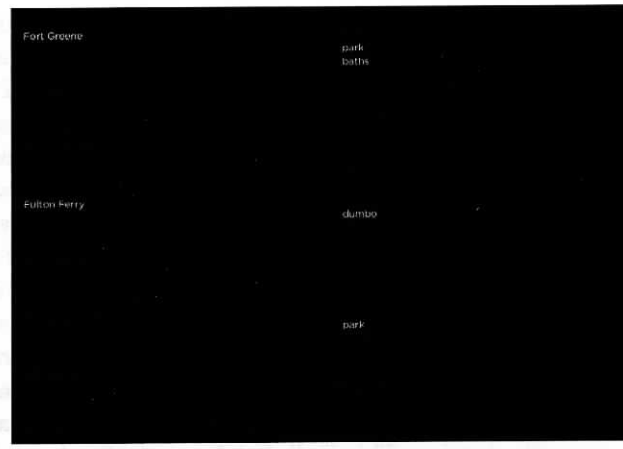

Figure 8.3 A list view allows alphabetical browsing of keywords by neighborhood.

For the blog content, we chose to use keywords rather than full text passages based on concerns about legibility. Over the course of several weeks we refined the algorithm to produce more meaningful keyword results, after which we began viewing the information in a spatial context and making adjustments to type size, opacity, color, position, and general visibility. We limited the number of keywords to a maximum of three per neighborhood, and arranged them radially around their respective neighborhood labels. Again to improve legibility, both keywords and neighborhood labels always face the user while displayed in three-dimensional space.

To compensate for compromised legibility within the spatial view, we later added a list view (Figure 8.3) to allow pivoting on a particular neighborhood, transporting the user back to the geographic view centered on the selected element. We also programmed the application to highlight neighborhoods that share keywords. Through this key interaction, the entire field of text becomes a stage on which relationships between neighborhoods can be selectively explored (Figure 8.4).

Our initial vision for *Pastiche* was that it update in real time, its form changing during runtime as new data became available. However, after running several tests we soon realized that the piece would be most effective as a cumulative visualization based on relevance. Parsing the keywords most related to their respective neighborhood, rather than simply looking for the most recent keywords, produced a more descriptive set of results and a more meaningful end result. Consequently, *Pastiche* evolved from a real-time model to one that is cumulative. The final outcome is now best described as a visualization of collective memory.

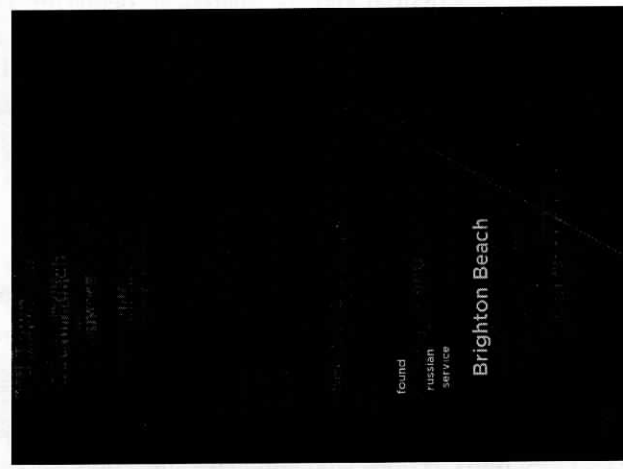

Figure 8.4 Neighborhood labels are surrounded by their related keywords extracted from blog articles.

Theoretical Perspectives

Modeling the City

My fascination with cities stems from a desire to understand the multitude and magnitude of the forces shaping them. New York City, where I live, is the product of a complex array of interacting forces, from politics to the

real-estate market and the economy, from the dynamics of public and private transportation to cultural trends and traumatic events such as September 11, 2001. Cities are truly living organisms. They retain evidence of the forces at work within them in their very fabric—in the architecture, the streets and public spaces, as well as in the composition of city neighborhoods. Cities are a potent indicator of culture and society and give us a sense of the values and ideals of the people who shaped them.

My interest was perhaps first stimulated by two early texts by Paul Auster: *Moon Palace* and *City of Glass*. In the former, Auster paints a vivid picture of his adolescent experience in New York City, while the latter is a conceptual piece on serendipity and the power of chance, in one memorable instance making a direct analogy between writing and the act of walking in the city. To me, these titles outline two fundamental approaches to understanding cities: one experiential, the other analytical.

I would describe the analytical approach as a generalized set of predictions based on observation. This includes the work of urban planners, sociologists, architectural theorists, and philosophers, from Jane Jacobs to Manuel Castells, Susan Sontag to Rem Koolhaas, who have written about the modern city and the role and effects of politics, culture, development, and technology. I was particularly influenced by the Italian postmodern architect Aldo Rossi, who wrote about urban artifacts in *The Architecture of the City*.[2] As constants within the city fabric that result from and define a particular cultural context, urban artifacts are a pivotal element in the organic development of a city over time.

An example of the experiential approach is the French avante-garde movement, the *Situationist International* (SI). Members of SI subjected themselves, in psychogeographic experiments, to a generative algorithm while traversing the streets of Paris, allowing the program to dictate the course. Freed of intent, they were able to experience serendipity and rediscover the city.

The experiential and analytical sides are of course intrinsically related, cities being both the purveyors and the products of our experiences. Architect and writer James Sanders developed a similar hypothesis in his work *Celluloid Skyline*, in which he described how cinema has represented New York City over time (what he calls the "mythic city") and how the real city, in turn, has shaped itself according to presiding mythologies of New York.[3] The discipline of mapping effectively bridges the realms of the experiential and the analytical, depicting and facilitating cultural mythologies. As Denis Wood wrote in *The Power of Maps*, maps are neither neutral, nor are they authentically objective. Rather, maps have agency, and instead of merely describing a territory, they *become* the territory.[4] By modeling the city, we are not only describing it, but also participating in the act of shaping it.

Data visualization

We are inundated with information. The sheer amount of information we absorb daily demands solutions to help us determine what is relevant and meaningful in our lives. In this information age, we are increasingly searching beyond facts for guidance and interpretation to help us understand their significance. This may explain the recent attention given to data visualization, which presumes to help us understand large and complex datasets and derive meaningful conclusions from them.

Most people who followed the 2008 US presidential election will remember the interactive graphs and maps used by the media to analyze the results of both campaigns and help explain developments to the viewer. It was particularly remarkable that,

instead of simply describing a chart or graph, television anchors often interacted directly with visualizations on large touch-screens. This resulted in a theatrical quality reminiscent of Hans Rosling's famous speech at the 2006 TED conference, where he demonstrated the capability of his Gapminder World software to provide insight from time-series data. Theatrics were also evident in the 2006 documentary *An Inconvenient Truth*. Those who saw it remember Al Gore explaining a so-called "hockey stick" graph, named after its characteristic shape showing the sharp upturn in temperature over the last 90 years (Figure 8.5). As David Womack writes in "Seeing is Believing: Information Visualization and the Debate over Global Warming," the graph has since become an icon for the climate change debate.[5]

In the realm of art and design, Paula Antonelli curated *Design and the Elastic Mind* at the Museum of Modern Art in New York. Data visualization played a significant role in this 2008 exhibition that explored the relationship between science and design.

Several blogs have emerged over the last couple of years that document recent developments in data visualization. Two of the most widely read are Andrew Vande Moere's *Information Aesthetics* (infosthetics.com) and Nathan Yau's FlowingData (flowingdata.com). The *New York Times* also regularly publishes interactive, editorially driven visualizations on current topics (nytimes.com/multimedia), indicating that data visualization has become part of mainstream culture.

However, while there are countless well-known examples of data visualization, it still tends to elude singular definition. In *Readings in Information Visualization*, Stuart Card *et al.* describe visualizations as "computer-supported, interactive, visual representations of

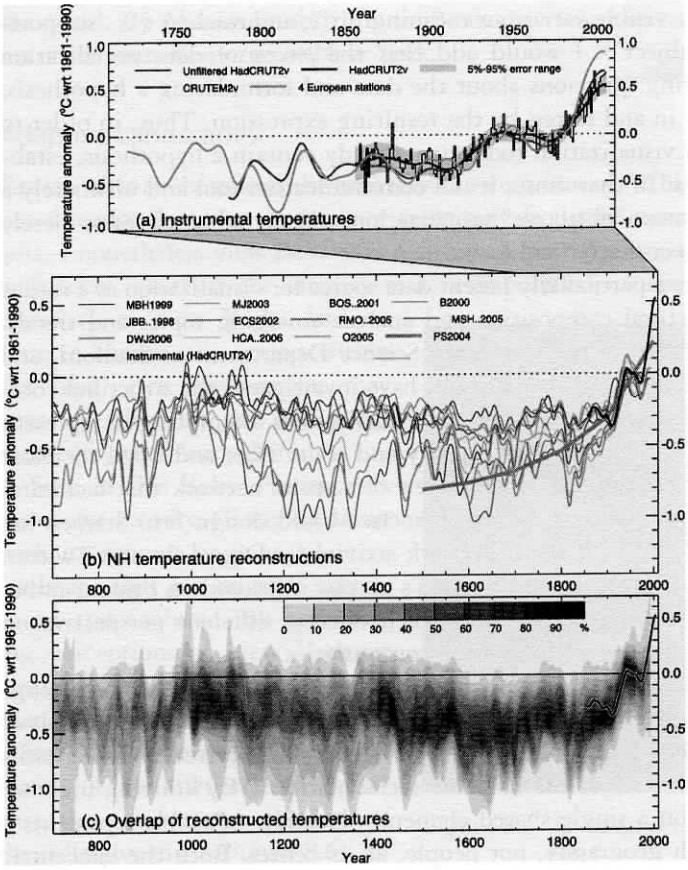

Figure 8.5
The "hockey stick" graph originally published by the IPCC and popularized by *An Inconvenient Truth*. Image credit: *Climate Change 2007: The Physical Science Basis*. Working Group I Contribution of the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Figure 6.10. Cambridge University Press. Used with permission from the IPCC.

data to amplify cognition."[6] Colin Ware, on the other hand, views visualization as an empirical science. In *Information Visualization*, he suggests that effective information presentation should be based on visual perception, including the perception of color, pattern, objects, and space.[7] Edward Tufte, perhaps the most prominent proponent of information design, regards visualization as a technology, describing techniques and processes that enable clear communication and data analysis.[8] While Ben Fry, MIT Media Lab graduate and former Broad Institute of MIT and Harvard postdoctoral candidate, considers storytelling to be one of the most valuable traits of data visualization.[9] Yet these views ultimately have their limitations. They don't satisfactorily encompass aesthetic data visualization or information art, which utilize data for more expressive ends. And when applied to a critical art context, it is apparent that a study of data visualization would include additional material to the rules of clear information presentation, due to its functioning as a medium in support of the artist's case. In "Critical Visualization," an essay written for the catalog of MoMA's *Design and the Elastic Mind*, Peter Hall recently described data visualization as a science, a technology, and an art.[10] This inclusive view finally points toward data visualization as a medium with a wide range of expressive potential, a view articulated by Eric Rodenbeck of Stamen, a San Francisco-based design firm specializing in data visualization.

On his website, interaction designer W. Bradford Paley addresses the question, "What makes something Information Visualization?" According to Paley, any visualization should be hypothesis-generating, meaning that it should support asking questions.[11] This test applies if the visualization is to function as a tool, after passing the basic tests of being *data-driven* ("Is there data in it?"), *information-driven* ("Is there information in it? ... are the visible variations meaningful?"), and *readable* ("... supporting discourse about the subject"). I would add that the *process* of data visualization should always begin by asking questions about the data and formulating a hypothesis, one that is both manifested in and tested by the resulting expression. Thus, in order to be hypothesis-generating, a visualization tool must already contain a hypothesis, establishing its breadth and focus. In that sense, it is a communication tool and ultimately a rhetorical device. Even the most "objective" visualizations display a bias of sorts, merely through their choice of subject matter and focus.

Online social networks are a particularly potent data source for visualization as a means to make sense of their structural composition and analyze emerging topics and trends. Jeffrey Heer, Assistant Professor in the Computer Science Department at Stanford, and Danah Boyd, Social Media Researcher at Microsoft, have investigated the structural composition of social networks. Many early social graph visualizations attempting to represent entire networks suffered from their sheer complexity and scale. Heer and Boyd therefore advocate an "egocentric," selectively expandable view of a social network that facilitates browsing and exploration.[12] Among others, San Francisco-based design firm Stamen has explored the visualization of real-time social network activity for Digg Labs and Twitter. Unlike the work of Heer and Boyd, most of Stamen's output is focused on content rather than the internal structure of social networks, which offers a different perspective on understanding the dynamics present within virtual communities.

Of course, the largest existing social network is the web itself, yet it can be challenging to navigate particularly from a social point of view, as it lacks the tight protocols of Twitter, Facebook, and others. Recently, Google and others have started building tools for websites that use the web itself as the ultimate social platform. By knitting together disparate blog posts based on a single shared element—location—*Pastiche* represents a form of social network with geography, not people, at its center. Both the egocentric

view of Heer and Boyd as well as the dynamic representation of social content set precedents for the resulting visualization.

Outcomes

Pastiche is based on two questions: What are people's associations with New York City and its neighborhoods? What are the relationships and connection points between neighborhoods? Both questions lead to the central hypothesis that New York City is, in fact, two cities—a physical city and a city "of the mind"—and that the two repeatedly inform each other.

Starting with this hypothesis, geography became a decisive factor for us. Associations had to be shown in spatial proximity to demonstrate a correlation between the familiar city landscape and the city of the collective mind.

In essence, *Pastiche* maps information from the realm of blog articles to a geographic plane. The piece also reflects an attitude toward data density, namely that a high density of information allows patterns to become legible. It is critically important that, at times, all information be shown at once, in order for the relationships and connections between neighborhoods to emerge. Furthermore, the data landscape encourages exploration; it functions not only at the macro-level, but also at the (quasi-) street level. This is an important aspect in conveying a parallel experience to the physical city, and becomes necessary for the legibility of individual keywords. *Pastiche* also adheres to a reductionist philosophy interested in representing only the most essential information. Purely typographic, *Pastiche* remains tightly focused on its core objective—representing the city of the mind. By remaining abstract, it allows the viewer to complete the experience with his or her own memories of New York.

Adaptive Landscapes

Pastiche ultimately reflects the activities of a community. While it might appear passive or one-directional in that it does not seemingly facilitate communication among participants, I nonetheless view *Pastiche* as a two-way, participatory experience. Not only does it help users construct the identity of a neighborhood, it also encourages the active (re-) discovery of New York. And, at least in theory, any action within the "real" city environment may have an effect on the one documented by the blogosphere.

Seen in this light, *Pastiche* is a canvas for activity. Changing along with its content, it is an adaptive landscape that can be both experienced and analyzed. By aggregating perceptions of New York, it creates an alternate reality, one that, in turn, has the potential to inform our perceived reality. Ultimately, it is an invitation to experience the city through the lens of the collective mind (Figure 8.6).

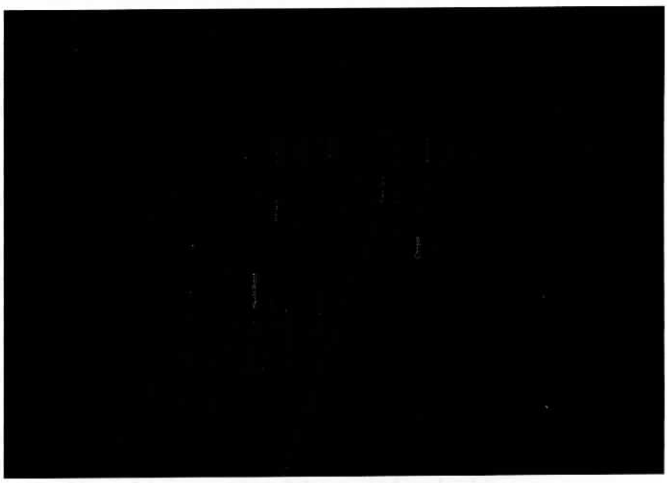

Figure 8.6 An aerial perspective view shows the contours of the city.

Notes

1. Rem Koolhaas, *Delirious New York: A Retroactive Manifesto for Manhattan*, New York: Oxford University Press, 1978: 10–11.
2. Aldo Rossi, *The Architecture of the City*, Cambridge, MA: MIT Press, 1982: 28–45, 127–130.
3. James Sanders, *Celluloid Skyline: New York and the Movies*, New York: A.A. Knopf, 2001: 13–22.
4. Denis Wood, *The Power of Maps*, New York: Guilford Press, 1992: 17–25.
5. David Womack, "Seeing is Believing: Information Visualization and the Debate over Global Warming," Adobe Systems Incorporated website, 2006: www.adobe.com/designcenter/thinktank/womack.html (accessed June 13, 2010): 1.
6. Stuart K. Card *et al.*, *Readings in Information Visualization: Using Vision to Think*, San Francisco: Morgan Kaufmann Publishers, 1999: 6.
7. Colin Ware, *Information Visualization: Perception for Design*, San Francisco: Morgan Kaufman, 2000: 1–27.
8. Edward R. Tufte, *The Visual Display of Quantitative Information*, Cheshire, CT: Graphics Press, 1983: 177–190.
9. Ben Fry, *Visualizing Data*, Sebastopol, CA: O'Reilly Media, Inc., 2008: 2–4.
10. Peter Hall, "Critical Visualization," in *Design and the Elastic Mind*, ed. Paola Antonelli, New York: Museum of Modern Art, 2008: 120–131.
11. W. Bradford Paley, "InfoVis tests," *Information Esthetics*, 2007. Online: http://informationesthetics.org/node/27 (accessed June 13, 2010): 1.
12. Danah Boyd and Jeffrey Heer, "Vizster: Visualizing Online Social Networks," Minneapolis: IEEE Symposium on Information Visualization (InfoVis 2005), 2005: 3.

Bibliography

An Inconvenient Truth/Paramount Classics and Participant Productions present a Lawrence Bender/Laurie David Production. Directed by Davis Guggenheim, 2006.

Auster, Paul. *City of Glass*. New York: Penguin Books, 1990.

Auster, Paul. *Moon Palace*. New York: Viking, 1989.

Card, Stuart K., Jock D. Mackinlay, and Ben Shneiderman. *Readings in Information Visualization: Using Vision to Think*. San Francisco: Morgan Kaufmann Publishers, 1999.

Fry, Ben. *Visualizing Data*. Sebastopol, CA: O'Reilly Media, Inc., 2008.

Hall, Peter. "Critical Visualization." In *Design and the Elastic Mind*. Ed. Paola Antonelli. New York: Museum of Modern Art, 2008: 120–131.

"Hans Rosling Shows the Best Stats You've EVER seen." TED Conferences, LLC website, June 2006. Online: www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html (accessed June 13, 2010).

Heer, Jeffrey and Danah Boyd. "Vizster: Visualizing Online Social Networks." Minneapolis: IEEE Symposium on Information Visualization (InfoVis 2005), October 23–25, 2005.

Koolhaas, Rem. *Delirious New York: A Retroactive Manifesto for Manhattan*. New York: Oxford University Press, 1978.

Lynch, Kevin. *The Image of the City*. Cambridge, MA: Technology Press, 1960.

Paley, W. Bradford. "InfoVis tests." *Information Esthetics*. May 19, 2007. Online: http://informationesthetics.org/node/27 (accessed June 13, 2010).

Rossi, Aldo. *The Architecture of the City*. Cambridge, MA: MIT Press, 1982.

Sanders, James. *Celluloid Skyline: New York and the Movies*. New York: A.A. Knopf, 2001.

Tufte, Edward R. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press, 1983.

Ware, Colin. *Information Visualization: Perception for Design*. San Francisco: Morgan Kaufman, 2000.

Wijk, Jarke van. "The Value of Visualization." *IEEE Visualization 2005 (VIS 2005)*. Minneapolis, 2005.

Womack, David. "Seeing is Believing: Information Visualization and the Debate over Global Warming." Adobe Systems Incorporated website, October 18, 2006: www.adobe.com/designcenter/thinktank/womack.html (accessed June 13, 2010).

Wood, Denis. *The Power of Maps*. New York: Guilford Press, 1992.

Links

www.christianmarcschmidt.com
www.climatecrisis.net
http://flowingdata.com
www.gapminder.org
http://infosthetics.com
www.moma.org/exhibitions/2008/elasticmind
www.nytimes.com/multimedia
http://stamen.com
http://substanceframework.org
www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html
wefeelfine.org