Technology, Transparency, and the Digital Presidency

Sunlight is said to be the best disinfectant.

—LOUIS J. BRANDEIS, *OTHER PEOPLE'S MONEY*

Science and art have in common intense seeing, the wide-eyed observing that generates empirical information. [It] is about how seeing turns into showing, how empirical observations turn into explanations and evidence.

—EDWARD TUFTE, *BEAUTIFUL EVIDENCE*

On March 28, 2011, a group of representatives from five different government watchdog groups met at the White House to present President Barack Obama with an award in recognition of his efforts toward creating greater government transparency.[1] The presentation was intended to coincide with an annual event known as Sunshine Week, which has, since 2002, sought to raise awareness about greater access to and oversight of the government by the press and individual citizens.[2] In an odd public-relations gaffe, however, the White House chose not to make the presentation an official press event and consequently held the meeting in private. As perhaps should have been expected, the press immediately jumped on the irony of the situation, and for the next twenty-four hours, headlines like "Obama Accepts Transparency Award . . . in Private!" appeared across the media. While the award's presenters were critical of the discrepancy between the event and its public profile, they nonetheless reiterated their praise of the president's efforts to make the federal government more open, and pushed him to continue in his pledge to make his presidency the most transparent in history.

In spite of the absurdity of recognizing openness in private, however, a deeper irony underscoring the award seems to have gone unnoticed by all involved. In the spring of 2011, the US government was perhaps more open than at any point in its history, but it wasn't just Obama who deserved the credit. Although his administration strove to overhaul policies and procedures in order to push

mountains of data into the public domain, its efforts were overshadowed by a series of high-profile unofficial leaks by the organization WikiLeaks. Combined, the 2011 WikiLeaks releases placed more than a million previously secret documents online. Alongside the surveillance programs uncovered by Edward Snowden in 2013, the leaks opened up a heated debate about the need for secrecy as well as transparency in public life. While Obama's policies were meant to signal a shift from the secrecy that had characterized his predecessor's administration, the WikiLeaks scandal forced the administration to admit that there were limits to the degree of openness it was comfortable with.

This chapter explores the threads of openness and transparency as they are woven into debates around the high-profile "data dumps" (both official and unofficial) that emerged from US government archives during the first years of the Obama administration. Although government transparency has long been considered an ideal within democratic politics, different forms of media—from newspapers and photographs to film and television—have been celebrated as the best means of achieving this goal at various points. It was clear from the start that for the Obama administration, transparency and digital technology went hand in hand. Out of this marriage, data visualization emerged as the favored medium to carry out the administration's ambitious proposals, creating a new form of digitally driven documentary media in the process. WikiLeaks shared Obama's belief in transparency and his faith in technology as the means to achieve it, but from an anarchic, oppositional position. Given the size and scope of its releases, journalists and amateurs alike also turned to new graphical tools like data visualization in order to mine and display the WikiLeaks data.

Long used to represent scientific and financial data, data visualization, with its ability to represent vast quantities of information at a glance, offered an ideal medium to capture the government's complex inner workings in a legible way for the general public. While it may seem that data visualization subverts the representational and indexical media that previous chapters considered, image-based media continue to reassert their importance and influence. Not only do still and moving images continue to appear alongside data visualization's interactive charts and graphs on government websites, but, as its evolution will demonstrate, data visualization's claim to representing reality rests on the same "seeing is believing" foundation that photographic media first established. Data visualization is an outgrowth and inheritor of the same documentary impulse that has for many decades fueled the production of nonfiction film and television in general.

While the US government and WikiLeaks were at odds over the nature and purpose of releasing information to the public, both were nonetheless part of a larger historical debate over the definitions of "open" and "transparent"—one geared toward expanding notions of publicity and public space. Echoing Louis Brandeis and other Progressive Era reformers from a century earlier, these

new champions of transparency and government accountability believed that "sunlight is the best disinfectant" for clearing away the miasma of government corruption and public distrust. Unlike their forerunners, this new generation of transparency advocates believed that the best technology of visibility was not the camera or the muckraking exposé, but instead the networked computer—a tool capable of visualizing and distributing the vast amounts of data being made public through official and unofficial channels. Combined with a Depression-era belief in the power of documentary evidence to inform the public, groups on both sides of the law worked to usher in an era of transparency and accountability akin to early- and mid-twentieth-century efforts, but this time with a distinctly twenty-first-century twist.[3]

DATA VISUALIZATION: THE ROOTS OF A REVOLUTION

The use of data visualization by both the Obama administration and WikiLeaks to make their transparency efforts legible to the public should perhaps come as no surprise. Data visualization as a medium is currently in the midst of an immense level of popularity and prominence across contemporary visual culture. Examples can be found everywhere, from advertising (IBM's "Smarter Planet" campaign)[4] to journalism (CNN's "Magic Wall" and NYTimes.com's interactive infographics)[5] to academia (the journal *Nature* and others, as well as the burgeoning field of analytics). Visualization is celebrated on blogs like informationisbeautiful.net and visualisingdata.com and made available to the masses via tools like Wordle, Tableau, and Many Eyes. Even as they struggle for added revenue amid falling circulation rates, newspapers like the *New York Times* and *The Guardian* have invested heavily in the medium. They regularly offer online features and interactive visualizations not possible in print editions, fostering the new field of "data journalism" in the process.[6] Visualizations are so ubiquitous that as early as 1982, in a pre-Internet era of publishing, Edward Tufte was able to claim, "Each year, the world over, somewhere between 900 billion (9×10^{11}) and 2 trillion (2×10^{12}) images of statistical graphics are printed."[7] This staggering figure might still be rivaled by the number of photographic images captured every year by the increasingly ubiquitous camera-enabled smartphone, but it's the contention of this chapter that the two should be thought of as being in the same category of media.

As a contemporary phenomenon, the recent explosion of data visualization relies on a range of newly emergent technologies, from the widespread deployment of inexpensive sensors to collect data, to advancements in cloud computing and commodity-level clusters for processing and analyzing this data, to high-resolution displays and increased graphics processing for rendering it visible.[8] But in spite of its recent popularity in contemporary media and its reliance on cutting-edge

technology, data visualization broadly conceived (the impulse to collect information about the world and display it visually) is among the oldest of pursuits.

EARLY HISTORY AND "GOLDEN AGE"

The current renaissance of computationally produced data visualization is only the latest iteration of an aesthetic form that stretches as far back as the first scientific tools and earliest forms of human writing.[9] Many contemporary data scientists and visualization researchers chart the roots of their work as far back as 6200 BC, to early maps found in Konya, Turkey, that demonstrate the desire to graphically depict the physical world.[10] This same impulse continues through Ptolemy's *Geographia* in AD 150 and reappears in various scientific and technical representations throughout the Renaissance, extending into William Playfair's experimentation with line charts and bar graphs in the early 1800s. Playfair's work stands at the beginning of what historian Michael Friendly refers to as the "golden age" of data visualization—a period during the second half of the nineteenth century when simultaneous developments in a range of fields fostered an explosive period of innovation and a volume of output rivaled only in the last fifteen years.[11] For Friendly, this period "deserves to be recognized—even revered—for the contributions that it made to statistical thought and practice in that time and for the legacy that it provides today."[12] As with our contemporary moment, these innovations touched on each phase of the transformation from data to graphics, including data collection (improvements in scientific measurement, instrumentation, and cartography), analysis (developments in the fields of statistics and demographics), and display (the invention of processes like lithography to print and distribute full-color graphics on a mass scale).

This golden age of infographics stems from a larger nineteenth-century revolution in science and mathematics, one in which new ways of observing the world were developed alongside new forms of processing and understanding those observations. The nineteenth century, as Ian Hacking argues, was bookended by the determinist models of Newtonian mechanics on one end and the open-ended indeterminacy of quantum mechanics on the other. Through this metamorphosis, the natural world came to seem "regular and yet not subject to the universal laws of nature."[13] Statistics, through the development of probability theory and what Hacking memorably describes as the "taming of chance," gained the ability to describe the world in regular, repeatable patterns. As in our current moment, data began to penetrate and influence the inner workings of social and political life in a way that it hadn't previously, touching on everything from medicine to policing to agriculture. Through the work of figures like Adolphe Quetelet and Charles Babbage, Joseph Fourier and Frédéric Villot, Charles Dupin and others, detailed records for virtually every aspect of the modern nation-state were recorded, tabulated, and of

course graphically represented for the first time.[14] The rise of probability theory in statistics as a scientific model went hand in hand with the sorts of vast data collection and visualization that Friendly celebrates, resulting in what Hacking refers to as an "avalanche" of printed numbers. This avalanche provided the foundation for and entrenchment of the modern bureaucratic state, interpolating in the process a new view of the public that "has affected not only the ways in which we conceive of a society, but the ways in which we describe our neighbor."[15]

For Michel Foucault—pointing out as he did that "statistics is knowledge of the state, of the forces and resources that characterize a state"—the nineteenth century bound modern statecraft to statistics even more directly. Part of governing a population, then, relied upon the collection and administration of "its quantity, mortality, natality; reckoning of the different categories of individuals in a state and of their wealth; assessment of the potential wealth available to the state, mines and forests, etc.; assessment of the wealth in circulation, of the balance of trade, and measure of the effects of taxes and duties, all of this data, and more besides, now constitute the essential content of the knowledge of the sovereign."[16] Indeed, the rise of what Foucault has called "the sciences of man" (and the statistical models and visualization tools they utilize) plays an essential role in the state's administration of its power. Numbers, and their visual forms, enable states to control and regulate the flow of people and resources within and through their borders. As we will see when we turn to more-contemporary examples of data visualization and their role in the transparency debates surrounding the Obama administration, the flow of the "knowledge of the state"—who is allowed access, who is subject to it—remains at stake.

Statistical data collection and graphical representation thus moved from describing natural phenomena in the physical world (e.g., Galileo's records of the movements of heavenly bodies) to describing, and eventually influencing, the nature of the social and political worlds (e.g., Minard's celebrated map of Napoleon's ill-fated march into Russia). This shift signaled not only an aesthetic change in how data was processed and displayed, but also a conceptual shift about what types of phenomena could produce data and to what uses the information could be put. This move from representing the physical to the social world is perhaps the most important legacy our current moment inherits from its nineteenth-century roots. It is also one that places techniques of data visualization in league with two other modes of representation—namely, photography and film.

PHOTOGRAPHY AND DATA VISUALIZATION: TWIN HISTORIES

The formal and procedural similarities between contemporary data visualization and photographic media being outlined here trace their roots back to the historical coincidence of the emergence of both forms. About the same time that

William Playfair and others were pioneering the techniques and models that would initiate Friendly's "golden age" of data visualization, Nicéphore Niépce and Louis Daguerre were beginning the collaboration that would produce the first reliable method for chemically recording and reproducing images captured in a camera obscura–style device.[17] By the time François Arago arose in the French Chamber of Deputies to formally announce the procedure perfected by Niépce and Daguerre in 1839, its benefit to scientific practice as a "valuable aid" of calculation and observation in fields as far-ranging as astronomy, microscopy, and anthropology were already clear.[18]

In claiming photography as a scientific tool, Arago was arguing one side in the debate that seems to have dogged photography from its earliest uses: whether it constituted a genuine art form or merely a mode of technological reproduction. Notably, it seems to have gone unquestioned that photography was always, at least, a tool for science and scientific observation. The automatic reproduction of reality without the mediating hand of the artist or scientist rendered photography both suspect as an art form and ultimately useful to scientists. Photography became both a symbol of the standard of scientific objectivity as a whole and one of the tools by which individual scientific results were documented. Through its use in the laboratory, a certain "facticity of the photograph" was secured through "a distinct form of scientific comportment that harnessed photography to rigid protocols," as Robin Kelsey and Blake Stimson put it.[19] In tones reminiscent of contemporary accounts of the importance of data visualization for generating insight, scientists utilizing early forms of photography to stop motion and freeze time hailed it as capable of offering humans a power of observation their eyes did not have. Its incorporation into the laboratory setting spurred a shift in scientific observation and documentation, bringing with it a standard of what Lorraine Daston and Peter Galison refer to as "mechanical objectivity".[20] As data visualization would a century later, photography seemed to cast light on an aspect of the world otherwise hidden from human observation.[21]

In the late nineteenth century, the world was becoming representable and represented, not simply calculable but increasingly quantified and measured, in means beyond the written word and the painted image. Photographic media (both moving and still) and data visualization were the means through which this transformation occurred. As a medium emerging in and among the same social and scientific changes that gave rise to the "taming of chance" and the establishment of the world as statistically calculable, photography was energetically identified as one of the forces of modernity shaping social life. Oliver Wendell Holmes's much cited celebration of stereoscopic photography placed the automatic reproduction of images alongside the railway and the telegraph as an invention whose "significance forces itself upon us daily."[22] For Holmes, the exchangeability of the image for the referent and its freedom to circulate made it capable of "annihilating time and space—a potential many would later grant to film as well. Holmes's prediction

of a time "when a man who wishes to see any object, natural or artificial, will go to the . . . stereographic library and call for its form" evinces the same desire for ubiquitous documentation of the world that drove the "avalanche of numbers" that Hacking and Friendly credit with the birth of data visualization during the same period.[23]

Given the close connection between scientific observation of the world and photographic documentation, it is unsurprising that the first steps on the path from still photography to motion pictures were undertaken by scientists seeking to perfect the still image as a means of observation. Étienne-Jules Marey and Edward Muybridge, the standard figures cited in histories of cinema, both stumbled on motion pictures while doing "other" scientific work. For Marey at least, it seems that motion reproduction, in the cinematic sense, was the least relevant by-product of his photographic work.[24] While he would later lament that the moving image's ability to capture "simply what the eye can see" rendered it useless as a tool of scientific observation, film's ability to accurately and automatically document the world cemented its place in the popular imagination as a substitute for physical presence in a given place.[25]

Almost from the moment they appeared, nonfiction moving images were deemed capable of documenting, preserving, and revealing the world. While documentary film in the standard definition of the genre wouldn't appear until the mid-1920s, from the Lumière brothers on, the impulse to document the world and consume the resulting footage has persisted in a variety of forms over the last century. Many of the same tropes from these early actualities reappear in the latest cell-phone footage on YouTube, including individual records like home movies as well as social/historical records like network news coverage of important events. As a tool, nonfiction moving images have always been utilized in the way scientists and data journalists utilize visualization today: to document and reveal the world to others.

Of course, like early forms of data visualization and the political and social frameworks they supported, photography and film have both been criticized for mediating and shaping the historical world as much as they record and reproduce it. Much of the critical work undertaken in the last fifty years has gone toward pointing out the lie behind photographic indexicality's seemingly untroubled connection with the physical world. A photograph is now thought to reveal as much about the photographer as it does about the subject—a by-product Kelsey and Stimson call its "double indexicality."[26] Moreover, as the work of John Tagg makes clear, photography itself was as much a part of the rise of the bureaucratic state and the normalizing of everyday life as was the data collection and statistical representation of the late nineteenth century. From the use of photography as a means of surveillance by police to the documentation of slums by governments seeking to rid society of their social ills, photography has played part in the power of the state equal to its more numerically driven counterparts: statistics and data

visualization.[27] If we add to this tools such as closed-circuit television, satellite imagery, and other optical forms of state surveillance—tools that have created a panopticon of power within modern social life that affects everything from urban planning to individual behavior—then moving images require as much scrutiny as tools of state power as their still forerunners.[28]

PHOTOGRAPHY AND DATA VISUALIZATION: DISTINCT MEDIUMS

In spite of their shared histories and the similar roles they have played in scientific, social, and political contexts, photographic media and data visualization are nonetheless distinct media that offer starkly different modes of representation. The differences between the two become apparent by considering the role that each has played in relation to the transparency movement. Although photographic media have been used as tools of the state in its administration of power, it would obviously be misleading to conclude that photographic images have benefited or been utilized by just the government. As an optical medium, photography was a natural fit for those who wished to keep an eye on the state and reveal its misdeeds to others. Data and data visualization, on the other hand, have until quite recently remained the province of the state. One characteristic of the modern transparency movement is its clear desire to make data visualization as accessible as photography has been for the past century.

A long tradition of social-issue photography exists in the United States, stretching as far back as Jacob Riis's documentation of New York slums in the 1890s and continuing through the work of Lewis Hine, Dorothea Lange, Margaret Bourke-White, and Walker Evans in the early twentieth century and extending on through the work of photojournalists like Kevin Carter, James Mollison, and others today.[29] While charges of exploitation and patronizing paternalism have always been levied against such work, the connections between the photographic representation of social issues and the healthy functioning of representative democracy continue to inform the impulse to visually document and distribute certain images as a tool for social justice.[30]

The work of Lewis Hine, for example, is often considered to have played a major role in the implementation of child-labor laws in the early twentieth century and in the rise of labor rights more generally throughout the 1920s and 1930s. In words reminiscent of Tufte's ambitious claims for data visualization, Hine often claimed a utopian strength for the power of documentary photography. He wrote:

> The picture is the language of all nationalities and all ages. The increase, during recent years, of illustrations in newspapers, books, exhibits and the like gives ample evidence of this.

The photograph has an added realism of its own; it has an inherent attraction not found in other forms of illustration. For this reason the average person believes implicitly that the photograph cannot falsify. Of course, you and I know that this unbounded faith in the integrity of the photograph is often rudely shaken, for, while photographs may not lie, liars may photograph. It becomes necessary, then, in our revelation of the truth, to see to it that the camera we depend upon contracts no bad habits. . . . The dictum, then, of the social worker is "Let there be light"; and in this campaign for light we have our advance agent the light writer—the photograph.[31]

Hine calls here on the full weight of long-standing Enlightenment connections between truth and social justice, drawing metaphorical and physical connections to the process of photographic exposure. In doing so, he tacitly channels the deep-seated liberal faith in the connection between the presence of information about a problem and resulting action toward social justice. This faith—that exposing a problem will result in its elimination—is one that runs throughout work on government transparency and political accountability, regardless of which medium is acting as the channel of transmission.

As Jonathan Kahana points out, much of the history of political documentary film in the United States rests on the notion that moving images can be used to achieve social change, whether this change is political accountability for those in power or greater justice for those without it.[32] Kahana covers a range of formal practices from the Depression era through the social crises of the 1960s and up to contemporary political work. He writes, "In an emancipatory gesture that [we] find repeated over and over . . . each style of documentary claims in its way to liberate its viewers from ignorance, prejudice, false consciousness, or illusion."[33] Indeed, much of the moving-image work considered in each of the previous chapters echoes this same gesture. From the overt skepticism of Errol Morris through the straightforward polemics of Robert Greenwald and on through the virtual walls of *Gone Gitmo's* Guantánamo Bay, each of the works considered uses a variety of formal approaches, but all share the basic aim of educating viewers on a set of issues. As a medium built on the same physical and photochemical principles utilized in still photography, political and social-issue documentary have inherited this faith in the connection between the light that exposes the film and the light that can expose social injustices—a spirit Hine so clearly articulated.

Turning to data visualization and the history of government transparency, we find a somewhat different history—one, in fact, that is far shorter. While notions of openness and transparency have been an ideal of democratic states from their earliest instantiations, the usual organ for achieving these goals has been the existence of the principle of freedom of the press. Access to government meetings and records at all levels by the press has been a consistent source of conflict, one in which degrees of access and publicity have varied throughout different countries and time periods. During the Progressive Era, for example, reformers like Louis

Brandeis and Woodrow Wilson called for greater information sharing on the part of the government in order to counteract the perceived corruption of politics by business interests. Brandeis's notion of "publicity" drew on his belief in the ability of the public gaze to root out corruption if given proper access to the information. Thus, one of the clear goals of Brandeis and others was making information known to the government available to the public as well.[34] Indeed, his well-worn observation that "sunlight is said to be the best of disinfectants" evokes metaphors similar to Hine's and directly connects access to information with optical visibility as a means to expose truth.

These early parallels notwithstanding, something like a comprehensive or even limited collection of information or records from and about the government that could be subject to analysis by independent groups and visualized for the public has emerged only within the last few years. This is in part the result of two factors that we have already seen in the history and development of data visualization. The first is that for much of its history as it relates to social and political life, data visualization has been a tool for the state to more effectively administer power. To be carried out, large-scale data collection and record keeping often required the resources of the state, and such records, moreover, were often kept secret by governments.[35] For example, one of the first uses of the early data-sorting machine that Herman Hollerith designed was in tabulating the 1890 and 1900 censuses for the US government, one of the only potential customers with a need and a budget to support the new technology.[36] Even as Hollerith's machine became the centerpiece of IBM and was eventually surpassed by digital mainframes and cloud computing, the company continued to rely heavily on sovereign governments around the world as the main customers for its cutting-edge computing technology. States were the only entities that had both the resources to purchase these technologies and sufficient quantities of data to justify them. But by the next centennial census, in 2000, technology like the Google MapReduce framework had made large-scale computing a commodity open to anyone capable of networking two or more computers together, and the ubiquitous presence of computers for the previous fifty years had created mountains of data that were increasingly available to anyone with an Internet connection.

The second change that enabled the current explosion of transparency-related activity was a further redefinition of what counts as data, and thus what can be visualized. In the same way that the golden age of data visualization was fueled in part by expanding the notion of data to include certain measures of social and political life, the current resurgence of the field is similarly expanding the idea of what falls within its purview. Part of the recent revolution in the field of data visualization, for example, has been the development of a related field known as natural-language data processing (NLDP for short). NLDP analyzes written texts and categorizes them according to various criteria, thereby generating data about them and hence the ability to visualize this material. Unlike conducting a census,

where data must be collected and classified, NLDP allows one to turn any written material into a data source (not just books, but e-mail, websites, tweets, etc.). In political circles, the same army of bureaucrats it takes to conduct a national census also produces an overwhelming archive of written material in the process of everyday governance. Even though reforms in the United States like the 1966 Freedom of Information Act (FOIA) and the Ethics in Government Act of 1978 (passed in response to the Watergate scandal) have long given citizens access to this information, until recently such records had been made available only in printed form, thus requiring an equally sized army of individuals to sort and make sense of such information. The advent of digital record keeping and NLDP makes this archive a viable source of data that can be analyzed and visualized by anyone with relatively little computing power.

Though photographic media and access to government data have both been invoked utilizing Enlightenment metaphors of light, optics, and vision as a direct access to truth and justice, it is only in the last decade that information visualization became a widely available form of visual media. Photography has enjoyed this status since the early twentieth century thanks to consumer-grade equipment like the Kodak Brownie. Until recently, free access to information as an avenue toward government transparency typically meant access to printed records in a specific government agency during prescribed hours. Any public insight into what such records contained was usually provided via written reports from the press or a specific watchdog group. Once these records become digital, however, they become open to investigation by a much wider segment of the population using a broader array of tools for summarizing and accessing their contents.

THE REBIRTH OF DATA VISUALIZATION

The current body of practitioners, theorists, and researchers working on data visualization maintain a curious relationship with the medium's past. On one hand, the field demonstrates a clear awareness of its own history. Textbooks on visualization often begin with an introductory snapshot of successful examples and important milestones, each perhaps covering different periods but all invariably highlighting a fairly consistent canon of work. And yet, despite this historical awareness, the claim is often made that what unites all of these materials is a kind of timeless, universal aesthetic that appeals to an innate, almost biological aptitude for this particular mode of representation. As Edward Tufte puts it, "The principles of analytical design are universal—like mathematics, the laws of Nature, the deep structure of language—and are not tied to any particular language, culture, style, century, gender, or technology of information display."[37] Data visualization has a history, but it is simultaneously and consequently thought of as timeless.

At the heart of what connects early diagrams in Euclidean geometry with computationally rendered scientific charts and graphs in the latest issue of *Nature* is a

long-held faith, in Western culture, in the apparent connection between vision and human cognition. As Martin Jay, W. J. T. Mitchell, and others working on visual culture have argued, the connections between seeing as one of the human senses and cognitive concepts like understanding, knowing, and believing run deep. The correlation between vision and our formulation of abstract concepts of reality, truth, and rationality dates back to the ancient Greeks and persists in varying degrees to the present day.[38] Regardless of whether this connection is culturally/historically determined or more biologically based (and Jay, for example, has shown that there are certainly cases to be made for both), for current visualization researchers it is embraced as a foundational principle for their research.[39] In this sense, visualizations of phenomena become models capable of both defining and communicating knowledge about the world in a process that Ben Shneiderman calls both "external cognition" and "expanded intelligence."[40] Visualizations, in other words, help scientists think about the phenomena they study and reveal truths that would remain hidden in the data itself, at least according to the scientists who work in visualization research. For Shneiderman, diagrams, charts, maps, and any number of other visual displays are tools that represent our observations of the world and enable us to discover new forms of knowledge about it.

If the renaissance in data visualization has a modern birthplace, it is arguably the Human–Computer Interaction Lab at the University of Maryland, which Shneiderman founded in 1983. The lab focuses on developing new tools and forms for visualizing large-scale data sets.[41] In a series of textbooks for the field, Shneiderman and his colleagues stake an intellectual ground and a scientific importance for contemporary data visualization that even echo intellectual histories in the field of visual culture. *Readings in Information Visualization,* which bears the subtitle *Using Vision to Think,* begins:

> To understand something is called "seeing" it. We try to make our ideas "clear," to bring them into "focus," to "arrange" our thoughts. The ubiquity of visual metaphors in describing cognitive processes hints at a nexus of relationships between what we see and what we think. . . .
>
> The interweaving of interior mental action and external perception (and manipulation) is no accident. It is how we achieve expanded intelligence.[42]

For Shneiderman and others working in the field, there is a clear belief in the untroubled connection between seeing, thinking, investigating, and communicating. This degree of faith in visualization's capability to record and reveal the world not only echoes adherents of the documentary image but also pushed the medium to prominence at a critical historical junction in the contemporary transparency debates.

Outside of academia, this same faith in the connection between vision, visualization, and insight is further echoed by the person commonly thought of as the popular-culture guru of the field of data visualization: Edward Tufte. Trained as

a political scientist, Tufte began looking at the power of data graphics when he cotaught a seminar on statistics to a group of journalists alongside the famed Princeton statistician and graphics pioneer John Tukey. After completing a manuscript on the subject, Tufte became frustrated when an academic press refused to print his book to the exacting standards he maintained the subject matter demanded: high-resolution graphics, archival paper, reader-friendly formatting. He took out a second mortgage on his home to self-publish the book, which became *The Visual Display of Quantitative Information* in 1983. Instantly a runaway success, the book is now hailed as one of the touchstones in the field of data visualization, and his follow-up texts *Beautiful Evidence* and others continue in the same vein. Tufte retired from Yale in 1999 and now tours the world giving one-day seminars on data visualization to sold-out crowds of academics, graphic designers, software developers, and product managers.[43]

Considering his presence both inside and outside the academy (in addition to the seminars Tufte gives, he has consulted with dozens of private companies), Tufte's influence on the contemporary field of data visualization is difficult to overstate. And Tufte, in turn, stakes the greatest social and intellectual claims for it. Indeed, the opening lines of *Beautiful Evidence* might almost be the governing ethos of the field as a whole:

> Evidence is evidence, whether words, numbers, images, diagrams, still or moving. The intellectual tasks remain constant regardless of the mode of evidence: to understand and to reason about the materials at hand, and to appraise their quality, relevance, and integrity. Science and art have in common intense seeing, the wide-eyed observing that generates empirical information. *Beautiful Evidence* is about how seeing turns into showing, how empirical observations turn into explanations and evidence.[44]

The connections he makes between empirical observation, explanation, and evidence demonstrate that through data visualization what might otherwise be simple facts or observations about the world are transformed into narratives that account for a particular mode of existence. Given the potential for such evidence to mislead, he calls the creation and consumption of evidence presentations a "moral act" in which presenter and audience are tasked to hold one another ethically, intellectually, and factually accountable. While such claims might easily be dismissed as mere hyperbole, Tufte provides evidence of the mortal importance of the particular form in which we choose to present information. In his essay "Cognitive Style of PowerPoint," Tufte analyzes Boeing's use of the popular presentation software to present its analysis of damage to the space shuttle *Columbia* during its launch in 2003. He convincingly argues that the default hierarchical format of the software program was partially responsible for NASA's inability to isolate the problem that eventually led to the shuttle's destruction upon reentry.[45] While many have derided PowerPoint for its facile display of information, few have argued that it could also be a matter of life and death.

Given Tufte's focus on the practical implications and limitations of the data visualization, it is clearly no coincidence that he started his work in the field as a political scientist teaching the craft of visualization to aspiring journalists. Whereas Shneiderman and other computer scientists champion data visualization's ability to produce new cognitive insights, Tufte's background seems to temper his enthusiasm, forcing him to consider the uses and misuses of visualization in a wider social and political context. Beyond seeing data visualization as a tool for scientists to better understand their observations about the world, Tufte recognizes its impact on the realm of social and political discourse. Instead of straightforwardly representing "what" is in the world, he notes the extent to which data visualizations are further called upon to narrate "why" a given situation exists and dictate "how" an audience should respond.

What further separates Tufte from the more classically trained scientists currently working in the field is his emphasis on the decisive role that aesthetic form plays in this process. Significantly, he relies on photographic metaphors as measures of the relative merit of a given technique or form. Visualizations are evaluated on their "resolution," which for Tufte refers not to the pixel density of a photographic image, but rather to the amount of data contained within a given visual space. High-resolution graphics like the sparkline can present thousands of points of data in a space no bigger than the average printed word, while low-resolution formats like PowerPoint slides contain no more than a few dozen. Visualizations are further gauged on their "clarity," which refers not to photographic depth of field or focus but instead to the extent that essential information is brought to the fore and extraneous details are excluded. Entire chapters of his book are devoted to detailing the dangers of what he calls "chart-junk" and "PowerPoint Phluff"—extraneous formal features added to low-resolution graphics to hide their inadequacies. Throughout his work, data visualization is for Tufte an aesthetic form in which every choice should be made with an eye toward maximizing the amount of relevant data that can be represented by a given visual feature.

Tufte's prominence in the field of visualization was recognized in late 2008 when he was called to Washington to use his data visualization skills to help his fellow citizens understand one of the greatest challenges it had faced in nearly century: the rapid collapse of many of the country's largest financial institutions. The person issuing the call was a young, newly elected president seeking to make good on many of his ambitious campaign promises: Barack Obama.

OBAMA'S OPEN GOVERNMENT

With his election to the White House in November of 2008, then president-elect Barack Obama immediately set to work fulfilling what had been his campaign's most amorphous and perhaps most compelling promise: change. Capitalizing on voter dissatisfaction over the protracted wars in Iraq and Afghanistan and

increasing anxiety over the global economic downturn, Obama's promise of change stretched from the specific (closing Guantánamo within a year) to the general (bridging the polarized, winner-take-all politics of the previous eight years). Attempting to demonstrate his intention to make good on these promises, his transition team established the change.gov website to outline and track progress on many of his administration's nascent policies. Although some of these long-term ideas would prove to be, like most campaign promises, overly ambitious, his administration's use of digital technology as a channel of open communication was a change in and of itself. Signaling an immediate break from the perception of secrecy and dissimulation that had for many characterized George W. Bush's White House, the transition team's early utilization of web technology for public communication became a hallmark of Obama's approach to governing.

While this embrace of technology marked a clear policy departure from his predecessor, for both Obama and the left it was simply an extension of both his campaign and the eight years of progressive opposition that delivered him to the White House. His candidacy had been characterized by its utilization of the net-roots strategies of organizing and fund-raising that had become a pillar of progressive activism during the Bush era. From his move to post the president's traditional weekly radio address on YouTube to his very public desire to hold on to his Black-Berry while in office, Obama signaled early on that he intended to incorporate technology into every aspect of the new administration.[46] Just as George Bush had styled himself the first "CEO president," Obama clearly wanted to claim the title of the first online or digital president.[47]

Nowhere was the embrace of technology more apparent than in the new president's approach to government transparency. On January 21, 2009, on his first full day in office, Obama issued a presidential memorandum with the subject heading "Transparency and Open Government." Part of a move advocated by several government watchdog groups to embrace "openness on day one," the initiative outlined in the memo described his administration's approach to sharing information about the government and directed various individuals to coordinate an "Open Government Directive," which would revise standards for releasing information to the public.[48] Reminiscent of the idealism surrounding his campaign and echoing the heady optimism surrounding Web 2.0 initiatives that had launched nearly a decade earlier, the memo outlines three general principles to be followed by all federal agencies: (1) government should be transparent; (2) government should be participatory; and (3) government should be collaborative.

Along with moves to declassify documents and speed up the response to FOIA requests, these new policies were roundly applauded by the coalition of groups that made up what was by then becoming known as the transparency or "sunshine" movement.[49] Along with other moves by his administration over the next eight years, they placed the Obama presidency in line with a series of

efforts by democracies around the world to reinvent or refocus government and democratic participation using digital tools and open-data frameworks. Eventually, these would include the Public Sector Information Directive in Europe in 2008, the Open Government Partnership in 2011, and the G8 Open Data Charter in 2013, among many others.[50] Many of these public open-data initiatives were justly criticized for further institutionalizing a neoliberal framework of privatizing public infrastructure (in the form of data) and normalizing competition as equivalent to efficiency.[51] However, the general ethos of releasing public data for the public good would eventually prove to be quite fragile once Obama left office and a new administration sought to alter the relationship between openness and government data.

The influence of entrepreneurial webspeak on the presidential memo turned out to be more than superficial. In the years that followed, the Obama administration instituted a range of new IT-driven policies and transparency initiatives—from websites and dashboards to blogs and data feeds—under the direction of the government's newly created chief information officer, Vivek Kundra.[52] Kundra had made a name for himself in the public sector by spearheading a number of open-government initiatives in Washington, D.C., under the reformist mayor Adrian Fenty, many of which would eventually come to characterize the federal government's approach to technological transparency. For example, one of Kundra's projects collected and published data, on a webpage called the "D.C. Data Catalog," that the city routinely collected.[53] It included access to everything from crime statistics and arrests to applications for building permits and city maintenance requests. Kundra realized that all of this data was both public information and potentially useful to citizens but that there was currently no way to connect the audience with its source. To address that problem, he borrowed what was quickly becoming a well-established model from the business world and held a contest that challenged people to come up with apps that would make the data more useful. The result of these efforts was a portal of tools and information that Kundra and the city hoped would be "a catalyst ensuring agencies operate as more responsive, better performing organizations."[54] Significantly, the page also included links and tools that would "allow users to create and share a variety of data visualizations" and informed each user that "you can create your own visualization using already uploaded datasets or slice and dice data the way no one has before."[55] The site was so successful that it eventually won the Harvard Kennedy School's Innovations in American Government Award in Urban Policy.[56]

After the 2008 election, Kundra was appointed to the transition team as a technology adviser, and part of his charge was to replicate the success he had had in Washington at the federal level. The result was the Open Government Initiative, which undertook a range of projects to push government transparency onto the web using a variety of tools. These included everything from sites like

USASpending.gov, a dashboard intended to track and streamline spending on government projects, to the OpenGov dashboard, a metachart tracking progress toward transparency across several dozen federal agencies. Among all of these initiatives, the two that best exemplify the government's current approach to transparency are the data.gov and recovery.gov projects that launched early in Obama's first term. Together, they illustrate the potentials and the pitfalls the government faces as it seeks to make itself more open to its citizens.

RECOVERY.GOV

Roughly one month after taking office, Obama signed into law the American Recovery and Reinvestment Act (known eventually as the "stimulus package" or the "recovery act"), which appropriated $787 billion of tax relief and government spending to counteract the effects of the ongoing global economic downturn. Two days later, on February 19, 2009, the website recovery.gov went live, welcoming visitors with the following message:

> Recovery.gov is a website that lets you, the taxpayer, figure out where the money from the American Recovery and Reinvestment Act is going. There are going to be a few different ways to search for information. The money is being distributed by Federal agencies, and soon you'll be able to see where it's going—to which states, to which congressional districts, even to which Federal contractors. As soon as we are able to, we'll display that information visually in maps, charts, and graphics.[57]

This initial page also featured a video of Obama explaining that the unprecedented size and scale of the recovery act demanded new methods of transparency and oversight to "root out waste, inefficiency and unnecessary spending."[58] Toward the bottom of the page was a simple bar chart breaking out the amounts dedicated to different spending projects. In spite of its rather limited initial offerings, the original home page for the site already contained all of the elements that would be essential to later versions. The site was a first step in what was intended to be a complete overhaul of the federal information infrastructure, a test case for his administration's transparency agenda.

After two years of updates, the site eventually became populated with a great deal of the information it initially promised. It became possible to search through hundreds of thousands of projects by size, geographic location, federal agency, or subcontractor. A number of different tools made this information available via charts, graphs, maps, and other visualization tools that allow one to analyze and interpret the data. The "video center" on the site listed over thirty clips (totaling just over three and a half hours of running time), each of which explained how to navigate the site and its overall mission. As an information source, recovery.gov represented an ambitious attempt to document and display

FIGURE 5.1. The initial recovery.gov home page.

the flow of federal revenue, dollar by dollar, from the Treasury into the multitude of projects it supported and the jobs it created—a virtual showcase of the recently reembraced Keynesian economic principles. But as an object of visual culture, the site was perhaps even more groundbreaking. In the economy of information it utilized to document the impact of the stimulus package, recovery.gov represented a clear faith in the ability of quantitative information to sufficiently represent reality, and thus signals the partial advent of a postphotographic form of the documentary impulse.

Placing a nonfiction, multimedia text like recovery.gov in the same conversation as documentary film is not without precedent. As Tom Gunning has convincingly argued, illustrated lectures and narrated slideshows by social reformers like Jacob Riis place early documentary still photography in the pre-evolution of later

FIGURE 5.2. Recovery.gov circa 2011, after a great deal of the data about the stimulus package had been generated and visualized.

forms of voyeuristic, observational ethnography.[59] In this formulation, a series of still images narrated by the speaker provides the same format (photographic evidence, timed delivery, narrative progression) that would later be united in the form of the sound documentary film. Similarly, the combination of evidence, argument, and political narrative that recovery.gov offered might be seen as a third iteration of the form—one in which all elements sit side by side rather than being delivered sequentially.

The purpose of the site, moreover, was one that places it squarely within the documentary tradition, or at least that part of the tradition populated by state-sponsored films that effortlessly (if overtly) conjoin civic edification with political persuasion. On an explicit level, the site's goal of openness and transparency might be read as providing information and issuing a call to action—both common items

on the documentary agenda. Indeed, the ominous presence of a large red button that adorned the top of every page that asks visitors to "report waste, fraud and abuse" simultaneously notifies them that they can take action while assuring them that action is being taken. Beyond its stated aims, the site also narrates for visitors, in basic beginning-middle-end structure, the story of how the stimulus package moved from being a piece of legislation in Washington to a series of concrete projects carried out in the real world.

The metanarrative at work here delivers the implicit political message that such spending works, and works for "real Americans." Part of a series of massive government spending measures intended to safeguard the economy from slipping into a depression, the stimulus package took its place alongside the Troubled Asset Relief Program and the auto-industry bailouts, which together set aside nearly $2 trillion to address the state of the economy. Of these, the stimulus package was the only clear example of a classic Keynesian stimulus investment by the public sector. While the other programs may have prevented the collapse of such iconic names in American business as General Motors and Bank of America, the stimulus package stood apart in that it was designed to inject new capital into the market. Within the Keynesian model, this public spending will then cascade across the economy as people paid by the government spend their paychecks on other goods and services, which income is then spent on other goods and services, and so on. The site's invitation to "track the money" is thus an invitation to witness what economists call the multiplier effect in action. Recovery.gov thus seeks not only to persuade skeptical conservatives that the historic spending levels were effective and necessary, but also to reassure taxpayers that they, too, would be the beneficiaries of the government's largesse.

Contrary to the fervor of the debate that surrounded it at the time, the size and scale of both the stimulus package and the economic threat it was meant to address were not unprecedented. The similarities between the Great Depression of the 1930s and what was quickly dubbed the "great recession" were widely discussed, and many parallels were drawn between the policy responses of Obama and of Franklin Delano Roosevelt. Indeed, these policy parallels might also be suggestively extended to the media that both administrations created in support of them.[60] After all, in attempting to carry out the controversial resettlement of destitute farmers and migrant laborers, Roosevelt's Resettlement Administration (RA) sponsored what eventually would become the era's most iconic and influential media representations. In addition to the well-known and widely distributed photographic work of Dorothea Lange, Walker Evans, and Gordon Parks, the RA also produced two thirty-minute documentary films meant to educate the public: *The Plow That Broke the Plains* (1936) and *The River* (1938), both directed by Pare Lorentz.

The River, in particular, offers both formal and rhetorical features that make its comparison to recovery.gov particularly productive. The film focuses on the
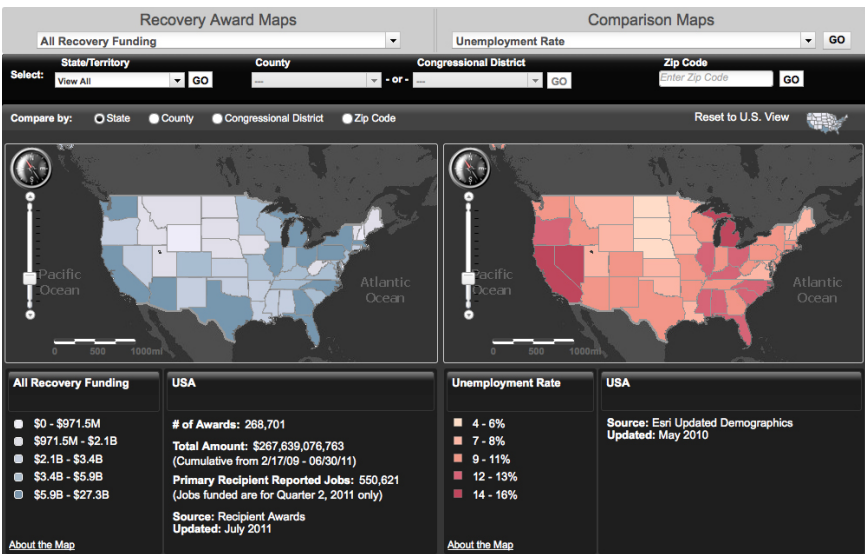
mismanagement of the Mississippi River watershed over the previous century as the nation pushed westward and documents the ecological and social destruction that resulted. Using a mix of statistics, maps, and images tied together with voiceover narration, the film argues for the need to control the river and restore the damage, both of which would involve the large public-works projects that have come to be associated with the New Deal. As Paul Arthur has argued, the film directly positions new forms of technology as the solution to bridging the existing conflict between man and nature but presents these solutions in a poetic, lyrical style that blunts the heavy-handed role government would necessarily play in carrying them out.[61] In rhetorical terms, while the film does contain sufficient logos-driven data to convince the audience of the size and scale of the problem (acres of farmland flooded, percentages of deforestation, tons of topsoil erosion, etc.), its primary appeals are the pathos-laden images of the destruction itself and of the people whose lives have been ruined.[62]

Returning to recovery.gov, we find a similar mix of elements with a decidedly different sense of proportion and emphasis. Whereas films like *The River* utilize data and statistics to support an overarching framework of photographic images, recovery.gov utilizes images to support what is otherwise intended as a data-delivery system. Rather than using a map to demonstrate the context and scale of the subject as *The River* does, the maps on recovery.gov instead become navigational tools through which specific data points may be accessed. While both texts seek to document specific flows and the impact they have on people's lives, for one this flow is the photogenic tempest of the nation's largest waterway, and for the other it is the flow of capital for the nation's largest fiscal outlay. In spite of historically similar origins and an overall shared political purpose, the two objects thus utilize radically different media to achieve their aims.

A great deal of this difference can be explained by the technological and social contexts in which they emerged. For the average audience member of the 1930s, moving images were the primary portal to the wider world. For the governments and institutions that sponsored them, documentary was seen as the best means of edifying and persuading the mass publics who gathered to see them. Like the hydroelectric dams that *The River* promoted, sound documentary films like it represented some of the most cutting-edge mass-media technologies of the time. Seventy years later, this slot has been filled by the Internet. While the information agnosticism of the Internet's design stipulates no difference between the types of data it carries—all packets are created equal—in the information economy of recovery.gov, data as data takes precedent over image as data. The site certainly does contain images both still and moving, but these clearly play a supporting role in relation to the data visualizations that are its main focus. The vast majority of the site's three and a half hours of video, for example, are dedicated to tutorials on using the site's data interface. Hence, most of the images the website presents are,

FIGURE 5.3. Data visualization in *The River* and on recovery.gov.

Recovery Award Maps

All Recovery Funding

Comparison Maps

Unemployment Rate   GO

Select: State/Territory   View All   GO   County   — — — or —   Congressional District   — — —   GO   Zip Code   Enter Zip Code   GO

Compare by: ○ State   ● County   ● Congressional District   ● Zip Code   Reset to U.S. View

Pacific Ocean   Atlantic Ocean   0   500   1000mi

Pacific Ocean   Atlantic Ocean   0   500   1000mi

All Recovery Funding

- $0 - $971.5M
- $971.5M - $2.1B
- $2.1B - $3.4B
- $3.4B - $5.9B
- $5.9B - $27.3B

USA

# of Awards: 268,701

Total Amount: $267,639,076,763
(Cumulative from 2/17/09 - 06/30/11)

Primary Recipient Reported Jobs: 550,621
(Jobs funded are for Quarter 2, 2011 only)

Source: Recipient Awards
Updated: July 2011

About the Map

Unemployment Rate

- 4 - 6%
- 7 - 8%
- 9 - 11%
- 12 - 13%
- 14 - 16%

USA

Source: Esri Updated Demographics
Updated: May 2010

About the Map

ironically, images of the website itself: screen captures, frame grabs, and so on. Both texts seek to "show" people what the government is doing to address their problems. But in earlier era, this meant photographic evidence. In the contemporary era, it means empirical evidence.

If data thus provides the core of the evidentiary claims that support the larger political argument the site levies, it also embodies the specific elements that form its basic narrative structure. Rather than the textual narration of the voiceover

heard on *The River*'s recorded soundtrack (groundbreaking technology itself at the time), recovery.gov's temporal beginning, middle, and end are laid out on the horizontal time axis of the charts and graphs detailing the allotment of funds and the completion of projects. While this varies depending on the particular statistical lens one chooses to use, the site's focus on procedures like funding allotment and project tracking means that one nearly always encounters graphs trending in an upward direction as they move toward completion, subtly implying notions of uplift and progress as time moves forward. This impression is both reinforced and potentially predetermined by the framing the site's title provides. "Recovery" in a general sense refers to a process of moving from a diminished state to an improved state one had previously inhabited; but used in its noun form in this context, it also implies that this process is an object that might be purchased and put on display. While a given chart might be labeled "Funds Allocated by the Department of Education," the larger channel of transmission continually reminds its audience that what they are seeing is the recovery of the US economy from one of the greatest economic threats it has ever faced.

The site's "Lights-On Map" in particular demonstrates the combination of ethos, pathos, and logos at work in the broader rhetorical framework of recovery.gov as a documentary corpus. Created by Edward Tufte as part of his role on the advisory board for Obama's data and transparency initiatives, the piece was a map-based visualization covering the disbursement of funds allocated through the spending program. The map itself depicts the outline of the continental United States, with Alaska and Hawaii in an unscaled inset below the southwestern border. Dark, neon-blue land masses are set against a black background, evoking well-known satellite images of the earth at night. This impression is reinforced as a series of over 150,000 tungsten-yellow points begin to slowly illuminate different areas on the map, each corresponding to a disbursement of funds for one of the recovery projects. As they appear, the visualization strongly conveys the impression of city lights coming on, illuminating the darkness cast across nation. Two insets further illuminate the map's scope and purpose. The first is a text box over the northern border that describes what the map depicts; the second is a small chart that indicates spending dates and dollar totals for those dates, the total trending in an upward direction. The entire animation lasts less than a minute, even though final version spans four years and tracks $285 billion in funding.

Although largely respected within the design community, Tufte's map was greeted with muted enthusiasm in professional design circles.[63] Many thought he had overly simplified the information by making every point of light equal one award (regardless of its size) and missed a huge opportunity to make the graphic interactive with the mountain of data underlying it. It had, in other words, emphasized style over substance, committing in the process a number of the errors that Tufte himself so often decried in his many books and lectures on good design.
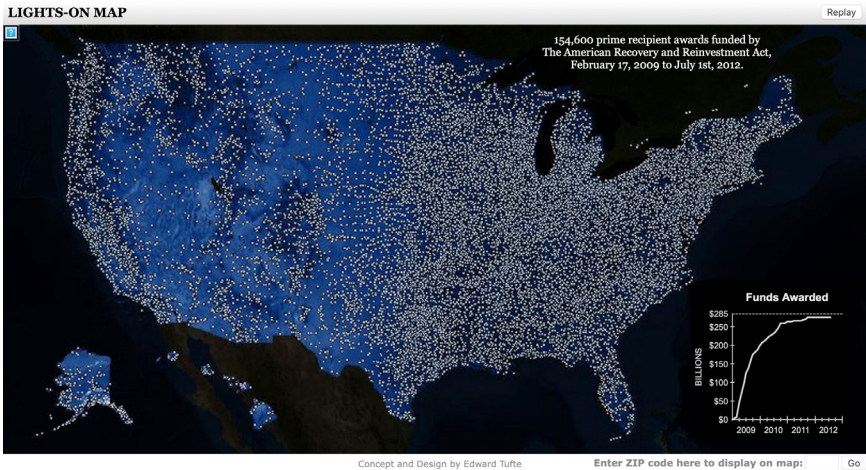
FIGURE 5.4. Edward Tufte's "Lights-On Map."

While not inaccurate, these critiques miss the larger rhetorical and historical work the map's aesthetic features reveal. The overt connection between spending money and bringing illumination to a space of darkness connects Keynesian economics with Enlightenment rationality and visual transparency. It further draws in metaphorical connections between the phrase "keeping the lights on," with its colloquial meaning of supporting struggling businesses, and the government's support of local economies and projects through the stimulus spending. As the animation progresses and the entire United States is illuminated, it draws on historical connections with the New Deal projects that *The River* had evangelized, including the Public Works Administration (which supported the construction of public infrastructure projects to generate electricity, among other things) and the Rural Electrification Administration (which supported extending electrical service to remote areas). Tufte's "Lights-On Map" might not have been the most data rich or interactive of the visualizations that appeared in the "Map Gallery" on recovery.gov, but it did aesthetically connect the policies behind the site's creation with the most ambitious efforts of Roosevelt's New Deal.

Beyond its evidentiary and narrative capacities, the overall site's transparency purposes bear out the extent of the government's faith in data's documentary capabilities. In addition to convincing people of the effectiveness of the program, the site also seeks to enlist the public's help in safeguarding these funds. By exposing the data to public scrutiny, any instances of "waste, fraud and abuse" will be highlighted and addressed. This implies a one-to-one correspondence between what's represented in the data and what's taking place in reality. This is a far more ambitious claim to transparency than the one advanced by the champions of film and

photography. Certainly a film might succeed in uncovering some level of malfeasance or graft via hidden-camera techniques of the sort pioneered by filmmakers like Nick Broomfield. But a project like recovery.gov claims to have already supplied all the relevant information. Wrongdoing in one instance must be captured; in the other, it need only be uncovered. It is only through the sort of totalizing archive of documentation that such claims might be made.

One final note worth mentioning on the documentary efforts of recovery.gov relates to the appropriateness of selecting data visualization as medium to explore this particular subject. The 2008 financial crisis revealed the nature of post-Fordist capitalism for ordinary observers. Starting with the collapse of Lehman Brothers and continuing on through the collapse of the US housing market, the "credit crunch" faced by major US banks, and various other cascading factors and effects, what slowly became apparent was the completely intangible nature of money and wealth in the modern global economy. While money itself has always been an abstraction of sorts based on various forms of value created through labor and the exchange of goods and services, this abstraction took on an extreme form in the various exotic financial instruments that were partially blamed with causing the crisis in the first place. In an odd way, this level of abstraction found its perfect corollary in the data visualizations on recovery.gov. Though intended to document the process of repairing the damage to people's jobs and lives, quantitative visual media of this sort also ideally reflect the intangible nature of the information-based economy that caused the damage itself. While this data was intended to be a gateway to the multitude of real-world projects on which the money was spent, viewed cynically, it might appear to offer data as the solution itself.

DATA.GOV

Launched several months after recovery.gov, on May 21, 2009, data.gov was intended to be a clearinghouse for all of the government data sets already being collected by the different agencies that make up the federal government. Unlike the event-driven nature of recovery.gov, data.gov was intended to provide a permanent access point for the release of US government data, and hence remains online nearly a decade after its launch. Prior to its creation, the default approach to information sharing by federal agencies was typically to err on the side of secrecy, owing either to legitimate concerns over national security and individual privacy or to a more general fear that such information could lead to criticism or embarrassment by the agency that collected it.[64] As the White House's own blog described the situation: "For years, agencies have collected data in support of their particular missions. But before the ubiquitous use of technology, data often sat in filing cabinets and agency basements."[65] Even when agencies did release information, it often took the form of reports published in PDF format or as charts and graphs without any of the underlying data exposed, neither of which could be read

by computers and other programs that might put the data to use. While this may have given agencies the sense that they were being open with their data, from a technological perspective it was no different than if they had been sitting in a basement or filing cabinet.

Responding to the president's memorandum on transparency, the release of the Open Government Directive in December of 2009 put in place official policies dictating how agencies should handle their existing data. Agencies were not only reminded of the three principles set out by the president ("transparency, participation and collaboration") but were also given a primer on the value of these principles, with details like: "Transparency promotes accountability by providing the public with information about what the Government is doing."[66] In addition to setting forth these general principles, the document sought to change the very culture of each agency, claiming:

> To increase accountability, promote informed participation by the public, and create economic opportunity, each agency shall take prompt steps to expand access to information by making it available online in open formats. With respect to information, the presumption shall be in favor of openness (to the extent permitted by law and subject to valid privacy, confidentiality, security, or other restrictions).[67]

The document also put in place specific deadlines each agency had to meet in order to be considered "in compliance" with the directive. These included milestones like placing three high-value data sets online within three months (which would then link back up to the data.gov site), as well as goals for targeting and publishing all of the data an agency collected. In order to hold the agencies accountable, the Open Government Initiative set up a dashboard measuring each agency on ten different benchmarks.[68]

The idea of creating a specific goal tied to a transparent progress report seems to have worked. When Data.gov launched prior to the directive, it featured just 47 different data sets. On its one-year anniversary, this number had grown to 250,028, and by the end of two years it was well over 379,000.[69] The site as it currently stands is a teeming mass of information, featuring lists of everything from the "Failed Bank List" published by the FDIC to the "Farmer's Market GEODATA" list put out by the Department of Agriculture. Although the updates in the site's various sections are admittedly uneven (the "Climate" section, for example, has had no updates since 2015, reflecting the priorities of the Trump administration EPA), many agencies are still publishing data on the site long after Obama has left office. Alongside all of the data sets are tools for viewing the data in different visualizations (charts, graphs, maps, timelines) as well as options for downloading the raw data in machine-readable formats (.csv and .xls). Beyond this, an "Applications" section featuring a number of ready-made tools and visualizations enables visitors to utilize the data and share their interpretations with one another, and a "Community" forum allows people to provide help to one another on using the site and feedback to the agencies for improvement.[70]

At first glance, it's difficult to see what use all of this data might be put to. In essence, this is part of data.gov's strategy. Rather than trying to anticipate what might be useful for people, the site is designed to give users complete access and allow them the freedom to create new views on the data by filtering and combining data sets to reveal insights. Responding to criticism that the site was too intimidating for average users (as one gender-biased review put it, it failed the "mom test"), visualization tools were added so that users could interact with the data right in their browsers and then save and share their visualizations with other users.[71] To further the aim of innovation and accessibility, agencies were required to create apps that visitors could use to navigate the data in meaningful ways, and several contests have been held by outside groups seeking to test the site's usefulness. Sunlight Labs, a project of the Sunlight Foundation, for example, held an "Apps for America" contest, which solicited projects from the community and awarded prizes for the best submissions.

Early on, the data.gov website had some high-profile successes—both in a general, public way and behind the scenes. Some apps, like Airport Status Service, enabled travelers to determine in real time whether a particular airport was experiencing delays, providing small conveniences to citizens. Others, like Hospital Report Card, enabled them to make important health-care-related decisions and push hospitals to improve their levels of care.[72] Behind the scenes, as agencies began standardizing and sharing data, some unexpected conveniences also emerged. The Department of Education (DOE) and the Internal Revenue Service (IRS), for example, found that by sharing data they could allow users to prepopulate the DOE's FAFSA (Free Application for Federal Student Aid) with information from their electronically filed income-tax returns.[73] Such things have the potential to be of enormous benefit to citizens simply by making data the government already has available to the public. All of these tools are designed to make the data that Foucault once deemed an essential tool of the nineteenth-century sovereign open to the twenty-first-century citizen.

For the most part, however, data.gov remains more of a potential benefit to citizens than an actual one. While it is still relatively new, the site has generated two well-placed criticisms that raise significant transparency issues. The first of these is the basic trade-off in the flexibility of completely unformatted data (what's referred to as wholesale data) and the more user-friendly but less flexible prepackaged data (retail data). Given varying levels of technological fluency, raw data will be unusable for most people. Put another way, it remains opaque rather than transparent. But retail data in the form of user-friendly charts and graphs flies in the face of the entire ethos of open-government initiatives by creating a layer of mediation between the information and its audience.[74] This conundrum between direct access to unfiltered material and legibility for the average user is the same one confronted by advocates of the Direct Cinema school in the United States. Eschewing

manipulative intervention in favor of "fly on the wall" filmmaking without interviews, voiceover, or expository intertitles, figures like Frederick Wiseman nonetheless implicitly rely heavily on skillful editing and camerawork to shape their material into a comprehensible form. While the trade-off achieved a certain level of formal purity, it opened its adherents to charges of duplicity nonetheless.[75]

While proponents of transparency and open government tend to favor some combination of both methods (which data.gov offers), allowing a completely free and open interpretation of the data runs a significant risk of error and misinformation. Even supposedly user-friendly applications that make use of data.gov, such as datamasher.org, require a basic understanding of the way statistics work. The site, which won the Sunlight Labs "Apps for America" contest and is often held up as an example of the power that data.gov opens up, allows users to combine any two data sets, such as poverty levels and high school graduation rates, on a national map to reveal correlations between different facets of social life in the United States.[76] But the potential for misapplication of the tool's parameters and hence misinterpretation of the data is readily apparent. To use our previous examples, consider a map that correlates the location of local farmers' markets with failed banks, claiming to demonstrate some relationship between the two. While a numerical relationship between the data would be easy to map using the tools the app provides, it would be difficult to claim that any causal connection had been revealed. The simplicity with which these visualizations can be created, and their connections to supposed "facts" about the world, mask a complexity in the science underlying the meaningful information we always hope such things will provide.

Moreover, as these tools and visualizations migrate off of the data.gov website and into the blogs and forums where political discourse increasingly takes place, it becomes conceivable that numbers might be found to support any range of politically loaded interpretations. Kundra once claimed that he aspired to make data. gov so easy to use and share that it would play a role on blogs the way YouTube does, implicitly equating data evidence with moving-image evidence.[77] The open-source ethos Kundra brought to the project holds that any egregious misuse of information will be spotted and quickly corrected as it is in the Wikipedia model, where users act as editors and fact checkers for one another. But in the paranoid and relatively polarized atmosphere of online political discourse, it is equally plausible that such corrections will themselves be quickly dismissed. While the potential for misuse certainly isn't a reason to avoid providing the information in the first place, it is a cause to question the utopian aims to which the site aspires.

The second, equally serious criticism that data.gov faces deals with data that doesn't appear on the site. Many open-government proponents question the extent to which relatively useless data is giving the government the appearance of transparency even as the material most important for fostering greater accountability is left off of the site. This problem became an acute reality when the Trump

administration took office. Transparency watchdogs were concerned about his apparent indifference to open-government initiatives more broadly and his outright hostility to specific factual data (like climate data) that conflicted with his policies. Sounding the alarm, many began archiving data before it disappeared from federal websites. While the widespread fears of "Trump's War on Data" did not come to pass in the manner that many feared, the anxiety created revealed the underlying fragility of the framework itself.[78] Data may not have been removed, but many of the sections within Data.gov haven't been updated since 2015, and there is nothing to prevent their future removal. As noted in the discussion of Second Life in the previous chapter, the Internet and digital platforms provide an effective means of making things instantly available to a widespread portion of the public, but the trade-off is what Wendy Chun refers to as an "enduring ephemerality."[79] Digital objects online can be available to everyone in one instant but gone forever in the next.

The inherent problem with gauging progress toward transparency is that most citizens and independent groups don't know the extent of the government's information holdings to begin with. As Ellen Miller of Sunlight Labs put it, "We don't like high-value data that involves [things like] wild horse counts. . . . We suspect they have data that would be of more interest to citizens."[80] As Aliya Sternstein points out, this might include information like which sections of private industry had been cited by the government for failing to meet public safety standards.[81] The paradox that all open-government initiatives quickly hit upon is that without complete and total transparency, it is difficult to gauge just how transparent the government is actually being. And yet, issues like national security and individual privacy do dictate the need for some "defense of secrecy," even if such concerns are often overblown.[82]

The Obama administration proved itself open to the criticism surroundings its new initiatives and capable of responding, when it decided to act, with at least a version of the "change" it promised. For example, after an extensive audit of open-data programs timed to coincide with the celebration of "Sunshine Week" in March of 2010, the National Security Archive's executive director, Tom Blanton, stated, "The Obama Administration deserves an 'A' for effort but an 'Incomplete' for results."[83] Sensitive to the criticism, the administration quickly announced on its blog a redoubling of its efforts, providing a "Tour of the Horizon," which reiterated that "transparency is one of the core principles of democracy."[84] The result was the launch of 2.0 versions for many of its sites that addressed the concerns brought by the administration's critics. But in an odd twist of irony, the technology-driven principles of transparency would be quickly put to the test by another open-government initiative, one operating outside of the Washington, D.C., beltway and headed not by an Obama appointee but instead by self-appointed activist-anarchist Julian Assange.

WIKILEAKS

The months following Sunshine Week 2010 would indeed be, as the White House blog predicted, "chock full of examples of concrete efforts—not lip service—to making open government happen."[85] However, these efforts were in large part the result of a collaboration between Julian Assange and Chelsea Manning to release several hundred thousand classified documents and other media from the US government's Secret Internet Protocol Router Network, or SIPRNet.[86] As if heeding Obama's call for transparency, collaboration, and participation, the government's open data was at once hit upon by a number of groups utilizing a range of new techniques to analyze, visualize, and make sense of the avalanche of information. As the ensuing drama and debate over WikiLeaks played out, a number of the groups that supported the Obama administration's initial open-government efforts once again stepped up to support the new poster child of transparency, in the process creating an uneasy juxtaposition between idealistic government bureaucrats on one hand and renegade anarchist whistle-blowers on the other. In spite of clear legal and procedural differences, however, the two groups shared a belief in the need for governmental transparency and a clear faith in technology as the best means of achieving this.

Any discussion of WikiLeaks and its role in the transparency movement confronts several potential problems. The first is the overidentification of the organization with its complicated and controversial founder, Julian Assange. Even his most ardent defenders and allies admit that Assange is an overbearing, attention-seeking figure prone to taking credit for the entire transparency movement. As accounts like Daniel Domscheit-Berg's *Inside WikiLeaks* make clear, both the organization and Assange himself benefited from a broad array of actors working toward the same ends.[87] The second problem is the overidentification of WikiLeaks with the series of massive data leaks that have come to characterize the era of big data itself. These include incidents across a broad spectrum, from Edward Snowden's shocking revelation that the National Security Agency was spying on US citizens to the incompetent release of personal credit information by Equifax in 2016. While WikiLeaks played an important role early on, as I will demonstrate here, not all leaks are politically, ethically, and materially equal, nor are they all the result of WikiLeaks . And finally, WikiLeaks itself has shifted radically from its early politically ambivalent anarchist stance to one more directly focused on shaping existing government structures through election influencing.[88] While I believe WikiLeaks deserves a great deal of credit for pushing the transparency debate in 2010, I would also caution against a simple celebration of the impact that it has had on democratic governance over the past decade.

Although WikiLeaks had been on the fringes of the activist and hacker communities for several years prior to 2010, its largest leaks had focused on fraud in private entities like the European banking giant Julius Baer and on political cor-

ruption in places like Kenya and Peru.[89] The organization had attained a reputation among hackers and transparency advocates for creating technology that would allow anyone to securely and anonymously upload large caches of previously secret data for publication on its servers. Its early successes also earned it the attention and appreciation of both the media and the nonprofit sectors, which recognized its positive disruptive potential for information freedom and social justice.

Throughout this early phase, WikiLeaks seems to have largely positioned itself as a basic conduit for information, publishing any and all contributions it deemed authentic and leaving the interpretation and investigation of the material up to journalists and activists. The site attempted to bring principles from the open-source software movement like community collaboration (memorably expressed in Linus Torvalds's "with enough eyes, all bugs are shallow") and open information exchange (Stewart Brand's equally memorable "information wants to be free") to the practice of whistle-blowing and investigative journalism. In April 2010, however, WikiLeaks began a series of high-profile leaks that would eventually earn it intense international legal and media scrutiny. The leaks, taken from classified US government databases, related to the wars in Iraq and Afghanistan and US diplomacy around the globe. Sensational as these releases were, the subsequent treatment of the source material itself reveals a great deal about the peril and promise of total informational freedom. While WikiLeaks and the trials of its controversial founder, Julian Assange, became a model for lofty goals like freedom of speech and governmental transparency in the age of the Internet, the migration of the leaked material across the media forced these ideals to confront reality.

COLLATERAL MURDER

The first of its high-profile leaks—the *Collateral Murder* video—was released by WikiLeaks on April 5, 2010, on the site's main page and on a connected site set up by WikiLeaks (http://www.collateralmurder.com). Both sites featured a set of videos drawn from an encrypted video file the site had received several months earlier captured by a camera mounted on the gunsight of an Apache helicopter in Iraq. The footage shows the helicopter shooting and killing eighteen people, including two Reuters journalists, and wounding two young children.[90] After vetting the footage to ensure its authenticity and sending people to Iraq to conduct interviews and notify relatives of the victims, WikiLeaks broke the footage into several versions. Among these was the original thirty-eight minutes of almost unedited material and a second, shorter version that excerpted relevant portions of the film and included both a prologue and an epilogue to the footage to provide context. In a possibly prescient move, this shorter version also solicited donations for the organization itself. Both films are accompanied by radio transmissions between the crew of the helicopter and soldiers on the ground. Although not

exactly graphic, the footage is nonetheless horrifying to watch and troubling on a multitude of levels—not least because of the fact that, regardless of the context, it depicts violent death on a scale and in a manner that few Americans are ever forced to confront.

As an example of the WikiLeaks style of technological transparency, the *Collateral Murder* footage is something of an anomaly. In spite of the attention the film generated with its release in April of 2010, it seems to have been virtually forgotten in the wake of other subsequent high-profile leaks. These included the Iraq and Afghanistan war diaries and, of course, the now infamous "Cablegate" release, which exposed the dirty laundry of the State Department in a cache of diplomatic cables sent back and forth between various embassies and Washington. In all of the press surrounding the US diplomatic cables, the *Collateral Murder* film barely rates a mention except as part of the "other material" allegedly released by US Army Private Chelsea Manning through her access to the government's SIPRNet. Moreover, its status as video footage taken from a single incident sets it apart from the other WikiLeaks releases, which have almost exclusively consisted of document collections spanning broader time frames. And yet, there is a great deal this film and its treatment reveal, not only about the status of visual media and visible evidence in the information age, but also about the value of information more generally for citizens in the democracy where the political debates around WikiLeaks played out.

As was noted at the time, the shortened version of the film represented something of a departure for WikiLeaks. Until this point, WikiLeaks had contented itself with summarizing and contextualizing the information it released but left the material itself entirely unaltered. With *Collateral Murder,* however, the released version of the film approached something closer to an analysis of the raw footage, thereby partially editorializing a particular interpretation. The move was criticized on two fronts. One set of critics felt WikiLeaks hadn't simply contextualized and interpreted the shooting, but rather had misrepresented it entirely by leaving out certain mitigating details, including the possession of weapons by several members of the group fired upon by the helicopter.[91] A second set of critics felt WikiLeaks had overstepped its role as a self-branded leaker of information and had become something more akin to a news organization rather than a simple conduit connecting sources of information with the public. Even Steven Colbert, in a rare moment of seriousness, confronted Assange on the issue, stating: "You have edited this tape, and you have given it a title called *Collateral Murder.* That's not leaking, that's a pure editorial."[92] Assange, both here and elsewhere, justified the move by stating that part of WikiLeaks's promise to its sources included generating what he called "maximum political impact" with the information it released—an aim echoed in statements by other members of the organization.[93] While this usually meant partnering with journalists at major news organizations like the *New York*

Times, Der Spiegel, The Guardian, and others, in this case WikiLeaks took on the job itself. In doing so, WikiLeaks wasn't playing the role of editor or journalist, but rather the role of filmmaker. In doing so, what it created was a documentary.

The case for claiming *Collateral Murder* as a documentary is a fairly straight-forward one. Whether we take John Grierson's oft-repeated if equally contested definition of documentary as "the creative treatment of actuality"[94] or Bill Nichols's more recent reworking that it "tell[s] stories with evidence and argument," documentary is generally accepted to consist, in varying degrees, of a creative or critical interpretation of events in the historical world, often with the intent of convincing viewers to accept this particular version of events as "true"—loaded as that last term may be. Regardless of the degree of nuance these definitions leave out, or the expansive domains into which documentary scholars and practitioners have recently pushed the canon, *Collateral Murder* clearly fits comfortably within these boundaries.

As evidence of this, consider the film's opening intertitles, which situate the footage historically and cast it within a particular critical frame. After a quote by George Orwell about the speciousness of political language,[95] we're given a brief synopsis of the event and informed that two of the men killed, Saeed Chmagh and Namir Noor-Eldeen, were Reuters news reporters, and we're shown images of both. We're then told that Reuters petitioned the US government to release the video under the Freedom of Information Act, and the ominous final title declares, "[T]his video has not been released . . . until now."

In documentary terms, this opening segment is clearly doing a great deal of work, or what we might call, after Jonathan Kahana, intelligence work.[96] It organizes the field of knowledge by orienting it both specifically—this particular event, these individuals—and generally, via the Orwell quote, as part of the larger struggle between truth and lies in political discourse. The pathos-laden background information on the reporters—that both were respected, talented, and, in Chmagh's case, survived by a wife and children—cements the event's status as a genuine tragedy. Indeed, this implicit emotional framework becomes explicit in the first still image we see of Chmagh, an image itself framed in the grief of the son who clutches it to his chest. Lest we miss the point, the film's very title has already rendered judgment on the event by declaring it not simply a tragedy but one involving murder—an overt allegation calling for a judicial response that would hold those responsible for the killing accountable for their actions.

Outside of the prologue and the title, the film goes on to annotate and edit its source material in a variety of ways, including cutting out nearly half of the video's original length, enlarging the image at specific points to highlight specific details, and supplementing its on-screen information with labels and arrows that identify certain figures and details in the frame even as the filmmakers, controversially, ignore others. These alterations and annotations further cement the piece's status

as a documentary. Most of the footage left out of the shortened version of the film contains a second offensive by the helicopter crew some twenty minutes later in which three missiles are fired into a building believed to contain enemy fighters and weapons. As Raffi Katchadourian reported in the *New Yorker,* this second attack was arguably the bigger story, and a more open-and-shut violation of the "rules of engagement" followed by the US military.[97] While WikiLeaks may have passed up the opportunity for a second "smoking gun," the omission yields a more coherent "beginning, middle, and end" structure that works in the service of—to return to the definition of documentary—telling a story with evidence and argument. Simply put, this is what separates the document of the raw footage from the documentary nature of *Collateral Murder.*

These alterations of the footage also evince a need to clarify for the viewer details that would not otherwise be evident—a need that points to a curious indexical duality inherent in the source footage itself. In terms of rendering the event visible for its viewers, the source footage offers both too much and too little. On one hand, its original purpose has endowed the footage with a wealth of informational artifacts visible on-screen: the camera's position in space, its angle relative to the horizon, the exact center of the frame, the time of day, and so forth. And yet, for us as viewers, all of this information is relatively meaningless. On the other hand, the resolution of the image is far too poor to yield the relevant details that we care about in our attempts to understand what is happening. As evidence of the footage's insufficiency, we might consider the failure of both the pilot and the gunner to distinguish between a weapon and a camera, both of which were actually present on the scene. The disturbing nature of the footage comes not from the images themselves, but rather from our secondary knowledge about what they depict. Ironically, the maximum political impact that WikiLeaks sought comes almost entirely from the context surrounding the leaked material, of which the material itself is a fairly faint signifier.

If *Collateral Murder* is operating here as a documentary film, one with a clear political position and a set interpretation of events, Colbert was right to distinguish between what he called "leaking" and the "straight editorial" of the film. The editorial section has of course traditionally been that portion of the broadcast where the unreachable ideal of objectivity is momentarily cast aside and the sources of the text are able to voice their particular opinions. As a contrast, for example, more traditionally objective treatments of the material were offered by every major news outlet, from *Al Jazeera* to *Democracy Now!*'s Amy Goodman.[98]

What seems less obvious, however, is why this additional layer of mediation was needed at all. Journalism's traditional role within a liberal democracy is putting eyes and ears on the ground where citizens can't be to provide them with the information they will need to make informed choices. As Ulrich Keller has demonstrated, it was at the exact moment in the nineteenth century when war

was no longer waged as public spectacle that the war correspondent was born.[99] If, as the rhetoric surrounding it claims, WikiLeaks can "transparently" connect the "source" of the information with the public, why has the middle man persisted?[100]

On one level, the answer is obvious and borne out readily enough in *Collateral Murder*. As the film demonstrates, the raw information itself is anything but readily intelligible. And if the average viewer can't understand what's going on in forty minutes of video footage without help, then she or he stands even less chance when the object of consideration is vastly more complex—say, a cache of ninety thousand documents. This is what makes the *Collateral Murder* video emblematic of the problems posed by the larger WikiLeaks project. Even with direct access to a visual recording of a single event, the need for interpretation, and hence mediation, is immediately apparent. In this sense, the public relies on journalists and filmmakers to process the information so that it can be made accessible to a general audience. By offering multiple versions, WikiLeaks is just fulfilling its desire to create what it calls "scientific journalism" by placing the original evidence alongside the analysis so that viewers might consult it to arrive at their own conclusions. But on another level, we might wonder if this leaves us any better off than we were before.

At the heart of what WikiLeaks offers is the belief that contained somewhere within organizational and institutional archives is information that can and should be made public, with the promise of justice, accountability, and, ultimately, truth on the other side. Steeped in the Enlightenment faith of reason's ability to deliver one to the truth and reminiscent of the eighteenth-century debate on the freedom of the press, it should come as no surprise that this promise is at once alluring and controversial.[101] Indeed, the term "scientific journalism" itself belies an uncritical faith that, simply provided with the evidence, all rational individuals might arrive at the same conclusions. Both sides of the transparency debate are in agreement on this connection between evidence and universal truth. Their disagreement stems from one side's belief that the public has a right to access the information and the other side's willingness to accept government secrecy as a valid trade-off for some other good—presumably, security. Whether the release of information is seen as liberating or threatening, however, the belief in the "truth" of the leaks persists.

This same logic even extends to the radical left, where Slavoj Žižek claimed, in what can only be read as a parody of Donald Rumsfeld: "The real disturbance [of WikiLeaks's release] was at the level of appearances: we can no longer pretend we don't know what everyone knows we know."[102] For Žižek, the truth contained in the disclosures isn't the mundane truth that this or that injustice occurred in the midst of war, which is already known even without direct evidence. Rather, Žižek's interest lies in what he refers to as the true face of the power wielded over us by the state: the power to commit murder or wage war and lie about it. This is a truth that is "made more shameful by being publicized."[103]

Indeed, the same sentiments were echoed with another high-profile release of secret materials from an unlocked archive—in this case, the Abu Ghraib images that emerged from the cameras of Charles Graner and Sabrina Harman in 2006. In relation to those images, Judith Butler argued that they, and the illicit manner in which they escaped, were a striking counterexample to the ongoing battle by the state to regulate the field of intelligibility for the public through its policies of embedded reporting.[104] For Butler, the problem of embedded journalism was the clear possibility that the journalist, the supposed objective party, would simply adopt the ideological viewpoint of the individual units or soldiers whose actions they were supposed to be covering.[105] And this possibility has been borne out by several recent analyses of the practice.[106] But alongside the high-profile role of reporters embedded within different units of the military there has also been a contingent of unilateral reporters covering the conflicts. Left on their own to cover the war as they see fit, these reporters are neither protected by the military nor subject to the subtle indoctrination this situation may create. Indeed, Chmagh and Noor-Eldeen were both operating unilaterally when they were killed, apparently attempting to cover events beyond the perspective, and hence the safety, of troop units moving through the area. Despite the long history of comparisons made between the camera and the gun, never has the connection between the two been so direct, so mistaken, and so tragic.

The irony, however, is that for all of the debate surrounding the role of both embedded and unilateral reporters covering wars, some of the most high-profile events have been the leaks themselves, from Abu Ghraib to WikiLeaks. If the recent war in Iraq is, indeed, the "most covered war in history,"[107] then the distinction owes as much to Julian Assange and Charles Graner as it does to journalists. And while *Collateral Murder* highlights the dangerous roles of both embedded and unilateral reporters on the ground, the next two WikiLeaks releases would bring a third type of journalist to prominence: the data journalist.

THE WAR LOGS: IRAQ AND AFGHANISTAN

In spite of the attention *Collateral Murder* received, WikiLeaks was apparently just getting started. Over the course of the next few months, the video would be joined by several subsequent leaks of classified material, including what became known as "the Afghan War Diary" and "the Iraq War Logs" on July 25 and October 22, 2010, respectively—collectively known as "the War Logs"—and a final release of US diplomatic cables on November 29, 2010. While each of these collections is unique and offers interesting points in its own right, I focus here on the War Logs collection for the contrasts it offers to both *Collateral Murder* and the US government's official transparency projects.[108]

Unlike its handling of *Collateral Murder,* WikiLeaks decided early on to partner with major news organizations for the War Logs, both to increase the impact of the

releases and to outsource the work involved in identifying the relevant information contained in the masses of data they offered. Shortly after the release of *Collateral Murder,* WikiLeaks contacted both *The Guardian* and *Der Spiegel,* offering them a scoop on the next set of data in exchange for a simultaneous publication date. (*The Guardian* eventually brought in the *New York Times.*) All three news organizations spent about a month decoding the cryptic terminology of the reports, verifying the data against other sources, and determining what material would be of public interest. On the day that WikiLeaks released the full dataset online, all three news sources went public with the reporting they had prepared over the prior month.

Looking at the data, what stands out immediately is the extent to which a professional, third-party source is needed to interpret its contents and make it legible to a wider pubic. While the term "transparency" implies an unobstructed or unmediated connection between observer and subject, the War Logs demonstrate the necessity of several layers of mediation between the two. If the optical transparency of *Collateral Murder* required basic identifying labels and arrows to clarify its contents, then the War Logs would need an outright translation. Of the approximately ninety thousand documents it obtained, for example, the *New York Times* identified and published on its website a selection of the most relevant material (about two dozen documents, in total). A typical line from one of the documents reads:

> JCC REPORTS THAT IP REPORTED THE IED STRIKE ON CF CIV VIC MB 4265 9065. IP CLAIM THAT 1X LN AMBULANCE DRIVER WAS KILLED BY UNCONTROLLED SMALL ARMS FIRING BY THE CF CIV CONVOY AFTER THE IED STRIKE (SEE ASSOCIATIONS FOR DETAILS OF IED STRIKE). JCC NOTIFIED 4/101AA AND REQUESTED THAT THE CF PATROL AT THE SITE INVESTIGATE.[109]

Needless to say, this is hardly material that's clear to anyone outside of the military communication channels in which this particular terminology is used. In addition to the necessity of decoding the terse acronyms and obscure identifiers (e.g., MB 4265 9065), there is also the need to identify what of the underlying information is actually new. Both the Iraq and Afghanistan wars were extensively covered by news media around the world with whatever degree of bias. This daily reporting had covered many of the daily events that the War Logs recapitulated. Much of the leaked information was, in other words, old news. Given the scale of the leaks and the quantity of outside information against which it could be cross-checked, it is no surprise that all three news organizations turned to "database experts" in order to "mine the data."[110] Thus, whatever level of truth these reports contained, it would apparently need to be unearthed by specialized machinery.

What was unearthed, it turns out, was at once surprising and unsurprising, at least in terms of the news media and its relation to the official and unofficial information coming from the government. The unsurprising aspect of both leaks

was the largely sensationless nature of the information they offered. Overall, the assessment of the three news organizations seems to be that the actual conflict was less promising than the official government assessment ("bleaker" in *The Guardian*'s terms, "more grim" in the *New York Times*'s). Aside from several specific revelations (e.g., that the Taliban likely used a surface-to-air missile rather than a rocket-propelled grenade in bringing down a US helicopter), the logs themselves reveal no smoking gun. In a post-Watergate, post–Iran-Contra era, this sort of official spin seemed to shock few. Ironically, the leak itself ultimately became the biggest story.

More surprising than the actual information contained in the War Logs was the coverage it received by the news organizations that had early access to it. Simply reading the headlines published by the three primary outlets, one would get the impression that the War Logs had unearthed the truth of combat itself. *The Guardian*'s website, for example, claims in headlines that the War Logs offer "the unvarnished picture" and "expose[s] the real war."[111] The *New York Times*'s coverage claims that the documents take us "Inside the Fog of War" and offer a "real-time history" of the conflict.[112] On closer inspection, however, all three of the outlets carefully qualified and circumscribed the information contained in the reports. *Der Spiegel,* for example, included an FAQ section on its website, detailing all of the qualifiers that should be taken into account when going through the documents (level of classification, source, etc.). The *New York Times,* the most circumspect of the three, notes:

> It is sometimes unclear whether a particular incident report is based on firsthand observation, on the account of an intelligence source regarded as reliable, on less trustworthy sources or on speculation by the writer. It is also not known what may be missing from the material, either because it is in a more restrictive category of classification or for some other reason.[113]

In short, the War Logs offer *another* account of the wars but not *the* account.

While the prospect of any definitive account is an impossibility, the cautious approach of the *New York Times* to the unofficially released official material is reflective of the uncertain nature of government transparency in general. So long as some information remains secret, a necessity argued for by all but the most radical advocates of transparency, there will always be the suspicion that true "truth" lies at an even deeper, more classified level. This sentiment echoes the concerns voiced by Ellen Miller of Sunlight Labs over data.gov cited earlier ("We suspect they have data that might be of more interest to citizens"). While any information may be potentially useful, the invisible information casts a shadow on the visible.

Outside of their narrative coverage of the War Logs' material, both *The Guardian* and the *New York Times* also put together interactive visualizations of the data they contained. Of these, *The Guardian*'s "IED" (short for "improvised explosive device") visualization stands out as an excellent example of visualizations gener-
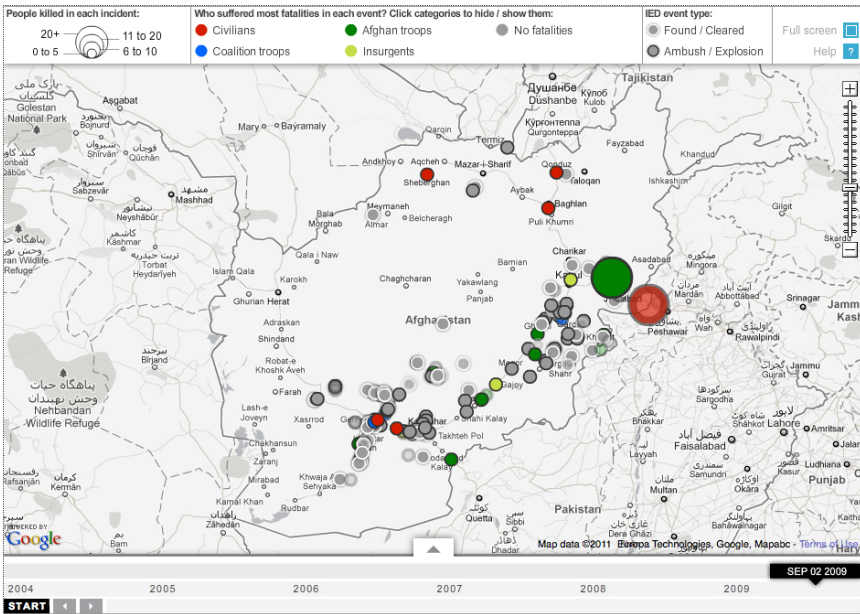
FIGURE 5.5. Interactive IED visualization tool on *The Guardian*'s website. In a multidimensional visualization like this, time and map location are overlaid with dots representing different attacks, with size representing the number of casualties and color representing the largest classification of those injured or killed (civilians, soldiers, etc.).

ated from the War Logs' data. Using data from the release of the Afghan War Diary, *The Guardian* sifted out reports of IED attacks in the country and used the embedded dates and GPS information to place them on a time-layered map of Afghanistan. Different targets and casualties were represented using both size and color codes (red for civilians, blue for coalition troops; the smallest dot for zero to five casualties, the largest for twenty or more, etc.). Viewed on a website, the resulting visualization can be played like a streaming video. Over the course of a few minutes, a map of Afghanistan is slowly dotted to reveal the number and location of IED attacks in a given time period.

In its own right, the visualization certainly opens up an interesting and otherwise unseen aspect of the conflict. The dots, starting slowly at first and building to a sustained crescendo, reveal in their concentrated location and apparent simultaneity both the contested regions of the war (primarily the beltway between Kabul and Kandahar) and the coordinated efforts of the Taliban insurgents. Even to a relatively attentive viewer of the news from the period, these aspects of the conflict become clear through the visualization in a way that they had not been previously.

Seen thus, a surprising quantity of information is condensed and made accessible in a relatively short time frame.

But beyond this initial impression, the insight such visualizations offer doesn't appear to penetrate very deeply. In other words, in spite of the quantity of information it provides (this would qualify in Tufte's terms as a "high resolution" graphic), such views of the event are hardly sufficient to understand the nature and meaning of the conflict itself outside of additional context. While one can certainly see that the number of IED attacks rises and falls at certain specific points, without additional sources of information it becomes difficult to ascertain the importance of any particular development. Seen through the lens of visualization, the War Logs reveal another account of these two wars, but it is very far indeed from revealing a definitive or total account. Even transparency, it seems, threatens to further obscure our view.

What remains most striking about the assertion of truth contained in the WikiLeaks material (both the video and the data) is the amazing capacity of the existing ideological frameworks to absorb this additional information as further evidence of their beliefs and move on. The left claimed that this material was evidence of the injustice of these wars, while the right claimed that it was a regrettable consequence of them. The extreme left saw evidence of a larger ideology of American empire, and the extreme right saw an instance of justice served. In spite of the millions of points of data added by the releases, the discursive framework around the events they record hardly changed at all.

Lest it seem that I'm sliding the debate over into the realm of an all too easy postmodern relativism, let me stress that I'm not claiming that all of these positions are factually or ethically equivalent, but rather that, despite their mutually irreconcilable positions, they remained unchanged in the face of new information. Nor am I claiming that transparency projects serve no purpose. As Micah Sifry points out, many such projects have achieved incremental improvements in government accountability.[114] But while the transparency movement may uncover cases of overt corruption and political manipulation, its potential for radical change against dominant ideology seems fairly limited. We certainly know more after WikiLeaks, but it seems we don't know any differently. In providing everyone direct access to the "truth," it seems, everyone's preexisting truths just become a little truer.

CONCLUSION

Writing on his blog "net critique," theorist, commentator, and sometimes agitator Geert Lovink published with Patrice Riemens what they called "Ten Theses on WikiLeaks."[115] Taken together, their observations amounted to an initial attempt to understand what WikiLeaks was by sidestepping the ongoing debate about whether or not WikiLeaks should be at all. The post contains a number of interesting

points, concluding that WikiLeaks (and organizations like it) amount to pilot projects in what will be an ongoing process of greater information overload. Insightful in their own right, their observations stand out even further in this context for the insight they offer on the US government's transparency as well as the work of WikiLeaks. In their third thesis, for example, Lovink and Riemens point out that while WikiLeaks deserves credit for opening up US government archives, its efforts cannot be seen as ushering in the "age of global transparency" that many have claimed, given the extent to which other equally large players (China and Russia, to name only two) remain beyond the grasp of whistle-blowing prowess. The same might easily be said of Obama's open-government projects, which do a great deal to open up what is arguably already the world's most scrutinized government. Important though these transparency initiatives are, they hardly lay out a map that other governments will follow. Lovink and Riemens's point that WikiLeaks is a classic single-person organization (SPO), and hence rises and falls with the fortunes of its founder, might be translated with a few caveats to Obama himself. While data.gov survived into the Trump administration, nothing guarantees that a future administration will not summarily pull the plug in the name of security or cost-cutting or both.

While Assange and Obama are obviously not interchangeable, a closer inspection of both reveals the extent to which their efforts utilize surprisingly similar rhetoric to justify parallel projects that face identical challenges. Both official and unofficial transparency projects saw themselves as part of a larger open-source hacker ethos, working to provide raw material to the public to make whatever use of it they see fit. Both further neglect the essential role that the traditional apparatus of investigative journalism must play in achieving any real insight from the information provided.[116]

The move to data transparency hence necessitates the embrace (or the reembrace) of visualization techniques to render and make sense of it. This move from image-based media (film and television) to data-based media (online databases, the Internet) represents a regime change of sorts that has been in the making for much of the last two centuries. Thus, much as François Arago and Oliver Wendell Holmes were establishing photographic images as the gold standard of objective observation and documentation in the mid–nineteenth century, quantitative data and their visual display were busy opening up this other window on the world more recently. Over the last century, the two fields have continued to develop alongside one another. The widespread diffusion of camera technology promises a panoramic if not panoptoconic view of the world as developments in digital technology continue to generate an exponentially expanding quantity of data about it.

Even as digital technology has continued to erode a faith in photography's ability to unproblematically represent reality, a faith that was itself never entirely unquestioned, it has increased the role that data plays within it. Consider, for

example, the extent to which debates about individual privacy have migrated from the fear that someone might optically witness and record one's physical actions to the fear that someone might access the information these actions inadvertently left behind: credit-card transactions, health records, and so on. If Lovink and Riemens are right in concluding that WikiLeaks is "a 'pilot' phase in an evolution towards a far more generalized culture of anarchic exposure, beyond the traditional politics of openness and transparency,"[117] then data visualization is the only means by which we might witness this evolution.