

## NOTES

### PREFACE

1. The renovation was covered in the *New York Times*. Cotter, “The Met Reimagines the American Story.”
2. For an introduction to the concepts of “back stage” and “front stage,” see Goffman, *The Presentation of Self in Everyday Life*.
3. Ibid.
4. For an introduction to sociotechnical perspectives, see Bijker, *Of Bicycles, Bakelites, and Bulbs*.
5. Galison, “Limits of Localism.”
6. Turkle, *Simulation and Its Discontents*; Loukissas, *Co-Designers*.
7. Geertz, *The Interpretation of Cultures*.
8. David Small first proposed this idea.
9. In recent years, the annual meeting of the Society for the Social Studies of Science and the iConference have both incorporated sessions on “making” and “doing,” which are other terms for “design.” For more on the possible relations between design and science, technology, and society, see Vertesi et al., “Engaging, Designing, and Making Digital Systems.”
10. I see the design of such systems in the broad terms laid out by Herbert Simon many years ago: “Everyone designs who devises courses of action aimed at changing existing situations into preferred ones.” Since Simon wrote this in 1969, the term *design* has continued to gain ground, to a degree that might have surprised even him. Simon, *The Sciences of the Artificial*, 111.

### INTRODUCTION

1. *Mega meta collection* is a term first used by Dan Cohen to describe “collections of collections” like the US Library of Congress’s Global Gateway project. See LeFurgy, “Digging into a Slice of Digital History.”
2. In recent years, researchers in data studies have sought to illuminate the local conditions in which data are created. For an early study, see Bowker and Star, *Sorting Things Out*. On reused data, see Zimmerman, “New Knowledge from Old Data.” On aggregated data, see Edwards, *A Vast Machine*. On exchanged data, see Vertesi and Dourish, “The Value of Data.”
3. University of California at Berkeley, College of Environmental Design, “The Data Made Me Do It.”
4. Kleinman, “Artificial Intelligence.” For a more in-depth academic exploration of this subject, see Noble, *Algorithms of Oppression*. Strauss, “34 Problems with Standardized Tests.” For a more extensive treatment of this topic, see Muller, *The Tyranny of Metrics*. Confessore and Wakabayashi, “How Russia Harvested American Rage to Reshape U.S. Politics.” For an academic exploration of this subject, see Boczkowski and Papacharissi, *Trump and the Media*.
5. In statistical analysis, *p*-value (probability value) is a number between 0 and 1. Results are typically considered statistically significant when *p*-value < 0.05. This means the chances that the same results could occur at random is less than 1 in 20. Simmons, Nelson, and Simonsohn, “False-Positive Psychology.”

6. In the 1980s, Sherry Turkle documented the tension between “doing” and “doubting” around data that accompanied early efforts by scientists at MIT to incorporate computer simulations in their labs and classrooms. One physics professor explained that the machines changed “what it means to do a physical measurement.” See Turkle, *Simulation and Its Discontents*, 36. For an even broader exploration of the history of these debates, see Porter, *Trust in Numbers*.
7. Haraway, “Situated Knowledges,” 590; Nagel, *The View from Nowhere*.
8. For influential works on privacy within social media, see boyd, *It’s Complicated*, and Marwick and boyd, “Networked Privacy.”
9. Finn, *What Algorithms Want*.
10. Consider JavaScript Object Notation, which is an increasingly used standard that brings together collections thinking and object orientedness in order to serve a broad array of online services.
11. As Lewis writes, “The price of the end product was driven by the ratings assigned to it by the models used by Moody’s and S&P. The inner workings of these models were, officially, a secret: Moody’s and S&P claimed they were impossible to game. But everyone on Wall Street knew that the people who ran the models were ripe for exploitation.” Lewis, *The Big Short*, 99.
12. *Ibid.*, 6.
13. *Ibid.*, 99.
14. Loukissas and Mindell, “Visual Apollo.” For a full analysis of human-machine relationships during the first lunar landing, see Mindell, *Digital Apollo*.
15. Some of the findings of this study were synthesized into a conference paper. Loukissas et al., “Redesigning Postoperative Review.”
16. For more on the dominance of US models of computing, see Chan, *Networking Peripheries*.
17. Ortner, *Anthropology and Social Theory*, 42.
18. To learn more about this history of data, see Gitelman, “*Raw Data*” *Is an Oxymoron*. Lisa Gitelman’s book, though, does not deliver the necessary tools for scholarship and practice in the present. Meanwhile, Christine Borgman’s work seeks to thoroughly prepare us for the challenges of data today, but it does so only for academic fields. Borgman, *Big Data, Little Data, No Data*.
19. This book is informed by extended empirical fieldwork, including interviews, observations, and workshops conducted over the course of the past six years. In preparing to write this book, I undertook the following research. I conducted fieldwork between 2012 and 2014 with faculty and staff at the Arnold Arboretum to inform chapter 2. For chapter 3, I engaged staff at the DPLA as well as contributing institutions (libraries, museums, and archives) between 2012 and 2015. Between 2014 and 2017, I worked with practitioners and researchers within broadcast journalism to inform chapter 4. From 2015 to 2017, I studied the practices of real estate agents, planners, tax assessors, developers, and affordable housing advocates in Atlanta, Georgia, to inform chapter 5. Many of the subjects interviewed or observed in my fieldwork are not mentioned directly. Subjects are only named with permission.
20. I make use of a combination of local reading techniques. This hybrid model of analysis owes much to developments in close and distant reading as methods of interrogating texts in liter-

ary and cultural studies. When used as a method of analysis for literary texts, Jonathan Culler explains that close readings attend to “how meaning is produced or conveyed. Culler, “The Closeness of Close Reading,” 22. Paradoxically, distant reading aims not to read. Instead, the latter technique, pioneered in literature by Franco Moretti, aspires to “generate an abstract view by shifting from observing textual content to visualizing global features of a single or of multiple text(s).” Jaenicke and Franzini, “On Close and Distant Reading in Digital Humanities,” 2.

21. Geertz, “Local Knowledge and Its Limits.”
22. For more on speculative approaches, see Dunne and Raby, *Speculative Everything*.
23. The visualizations in this book represent a small fraction of the total number developed during the course of my research. Each is custom created from computer code, primarily using Java or JavaScript, and sometimes D3 (Data-Driven Documents), a JavaScript library created by Mike Bostock and others. Many of the visualizations were made in collaboration with two graduate students of mine, Krystelle Denis (Harvard University) and Peter Polack (Georgia Tech). Often, I created an initial code sketch, and then my students elaborated on it. We did not rely on premade visualization templates or software. Yet I have left out the vast majority of the visualizations we created, for they were frequently rudimentary and of limited value to the reader. I have included only examples that help me make my points, and even then, only the last iterations.
24. Data visualization is significantly informed and enabled by, but more publicly oriented than, the subfield of computer science called information visualization. For an extensive history of these practices, see Drucker, *Graphesis*. My own particular approach to visual design and communication is shaped by my education in manual drafting at Cornell University’s Department of Architecture in the 1990s.
25. Discussions of what makes an effective data visualization are often kept separate from questions about the origins of data as well as their local stakes. On efficiency, see Tufte, *The Visual Display of Quantitative Information*. On memorability, see Borkin et al., “What Makes a Visualization Memorable?” On elegance, see Kirk, *Data Visualization*. “Data graphics,” writes Tufte, “are instruments for reasoning about quantitative information.” Tufte, *The Visual Display of Quantitative Information*, 9. “Whenever we analyze data,” explains Ben Fry, “our goal is to highlight its features in order of their importance, reveal patterns, and simultaneously show features that exist across multiple dimensions.” Fry, *Visualizing Data*, 1. “Data visualization,” instructs Andy Kirk, is “the representation and presentation of data to facilitate understanding.” Kirk, *Data Visualization*, 19. For one of the only books to embed a critique within data visualization, see Kurgan, *Close Up at a Distance*. Her monograph, however, is more a collection of creative work than a theoretical treatment of data.
26. This term was first used in Dalton and Thatcher, “What Does a Critical Data Studies Look Like, and Why Do We Care?”
27. Kitchin and Lauriault, “Towards Critical Data Studies.”
28. Bogost, *Persuasive Games*; Elmborg, “Critical Information Literacy”; Ratto, “Critical Making”; Dunne, *Hertzian Tales*.
29. Bellacasa, *Matters of Care*.
30. Maria Puig de la Bellacasa writes “understanding caring as something we do extends a vision of care as an ethically and politically charged practice, one that has been at the forefront of

- feminist concern with devalued labors.” Bellacasa, “Matters of Care in Technoscience,” 90. The need for care is widespread, but not simple. David Ribes explains that care can go wrong, specifically in practices of maintenance and repair: “Whether drawing attention to them, valuing them, or conceptually repopulating them with human work, care—practice or history—cannot be an end in and of itself if it means failing to ask the questions: repair and maintenance of what, serving whose interests, and at the expense of what people?” Ribes, “The Rub and Chafe of Maintenance and Repair.”
31. Negroponte, *Being Digital*, 165.
  32. Negroponte, “One Laptop per Child.”
  33. Robertson, “OLPC’s \$100 Laptop Was Going to Change the World.”
  34. Ananny and Winters, “Designing for Development,” 107.
  35. Berners-Lee et al., “The World-Wide Web.”
  36. Many eminent technologists, such as Ted Nelson, have criticized the founding assumptions of the web, which is now the dominant standard for the internet. See Nelson, “Complex Information Processing.”
  37. You do not go to a website. It is downloaded and displayed on your machine. Many people do not consciously realize that by the time they’ve seen a website, it is already stored on their computer and it will be indefinitely. Furthermore, depending on your network connection, hardware, software, and settings—not to mention any laws of your country that censor its content—the web page may appear quite differently that it does in other places.
  38. Chan, *Networking Peripheries*, 7.
  39. See Taub and Fisher, “Where Countries Are Tinderboxes and Facebook Is a Match.”
  40. For further reading on free market ideology, see Graeber, *Toward an Anthropological Theory of Value*.
  41. For a detailed explanation of computing at the periphery in Peru, see Chan, *Networking Peripheries*. For more on how data can become enlisted in colonialism, see Thatcher, O’Sullivan, and Mahmoudi, “Data Colonialism through Accumulation by Dispossession.”

## CHAPTER 1

1. Chalabi, “‘Data Is’ vs. ‘Data Are.’”
2. For more about how the term *data* can be used, see the *Associated Press Stylebook* or *Chicago Manual of Style*.
3. Rob Kitchin explains that the use of *capta* instead of *data* was also called for by H. E. Jensen in 1950. See Kitchin, *The Data Revolution*.
4. Dourish, *The Stuff of Bits*, 107.
5. *Ibid.*, 4.
6. Crampton, *Mapping*.
7. Bowker and Star, *Sorting Things Out*; Gitelman, “*Raw Data*” Is an Oxymoron; Dourish, *The Stuff of Bits*.
8. Latour, “Drawing Things Together,” 26.

9. Ibid.
10. See Latour and Woolgar, *Laboratory Life*; Lynch, *Art and Artifact in Laboratory Science*; Cetina, *Epistemic Cultures*; Keller, *Making Sense of Life*; Star and Griesemer, "Institutional Ecology, 'Translations' and Boundary Objects"; Bowker and Star, *Sorting Things Out*; Vertesi and Dourish, "The Value of Data"; Edwards, *A Vast Machine*.
11. Borgman, *Big Data, Little Data, No Data*.
12. Kitchen extends our understanding of data's diversity beyond the academic world and calls for more specific empirical studies of diverse data cultures. Kitchin, *The Data Revolution*, 4.
13. Recounted by the author from a public lecture by Culler at Cornell University in 1998.
14. If the examples used in this book seem accessible, outside their originally intended settings, it is only because of the standardization and coordination work that has been done by their keepers, who have prepared their data to "jump contexts" from one site of use to another. See Downey, "Making Media Work."
15. Kitchin and Lauriault, "Towards Critical Data Studies."
16. Kitchin, "Big Data, New Epistemologies, and Paradigm Shifts."
17. Kitchin and McArdle, "What Makes Big Data, Big Data?"
18. Crawford, "The Anxieties of Big Data."
19. Ibid.
20. Edwards, *A Vast Machine*.
21. Anderson, "The End of Theory."
22. See Borgman, *Big Data, Little Data, No Data*, 17. The early relation between data and faith has not entirely dissipated. Issues of purity still pervade the handling of data.
23. Buckland, "Information as Thing"; Borgman, *Big Data, Little Data, No Data*.
24. Borgman, *Big Data, Little Data, No Data*, 28.
25. Ibid.
26. Ribes and Finholt, "The Long Now of Technology Infrastructure."
27. See Posner and Klein, "Data as Media"; Lievrouw, "Materiality and Media in Communication and Technology Studies."
28. Marshall McLuhan writes, "This is merely to say that the personal and social consequences of any medium—that is, of any extension of ourselves—result from the new scale that is introduced into our affairs by each extension of ourselves, or by any new technology." McLuhan, *Understanding Media*, 9.
29. Manovich, *The Language of New Media*, 225.
30. McLuhan, *Understanding Media*, 3.
31. Gillespie, Boczkowski, and Foot, *Media Technologies*.
32. In everyday terms, *local* means "relating to a particular region or part, or to each of any number of these." *Oxford English Dictionary*, s.v. "Local, adj. 2," accessed December 4, 2017, <http://www.oed.com/viewdictionaryentry/Entry/11125>.
33. Geertz, *Local Knowledge*.

34. Geertz, *Available Light*, 140.
35. See Harding, "Is Science Multicultural?"; Wyer et al., *Women, Science, and Technology*; Barad, *Meeting the Universe Halfway*. For more on how feminist theory has been used to understand data and data visualization, see D'Ignazio and Klein, "Feminist Data Visualization."
36. See Harding, "Is Science Multicultural?," 55. As an example, Harding notes that "Western conceptions of laws of nature drew on both Judeo-Christian religious beliefs and the increasing familiarity in early modern Europe with centralized royal authority, with royal absolutism." Meanwhile, Chinese science regarded nature as a "web of relationships without a weaver," not an empire ruled by divine laws. See *Ibid.*, 316. For more on knowledge systems and their relation to local knowledge, see Turnbull, "Local Knowledge and Comparative Scientific Traditions"; Watson-Verran and Turnbull, "Knowledge Systems as Assemblages of Local Knowledge."
37. More recently, for instance, John Law and Annemarie Mol describe how scientific facts are "brought down to earth" through early laboratory studies, such as at an institute for biological studies in San Diego or plant protein research facility in Berkeley, both in California. Law and Mol, "Situating Technoscience," 610.
38. See Star and Griesemer, "Institutional Ecology, 'Translations,' and Boundary Objects"; Edwards, *A Vast Machine*; Vertesi and Dourish, "The Value of Data."
39. Haraway, "Situated Knowledges"; Suchman, *Human-Machine Reconfigurations*.
40. Cresswell, *In Place/Out of Place*, 3, 10.
41. Frampton, "Towards a Critical Regionalism."
42. Gibson, *Neuromancer*.
43. See Benedikt, *Cyberspace*; Mitchell, *City of Bits*; Castells, *The Information Age*. These authors have overturned a long history of social theories on the local. For studies of *gemeinschaft* and *gesellschaft*, see Tonnies and Loomis, *Community and Society*. On the distinction between particularism and universalism, see Parsons, *The Social System*.
44. Mitchell, *Me++*, 14.
45. Castells, *The Information Age*.
46. More recently, scholarship has shown the way that software and space mutually inform one another. See Kitchin and Dodge, *Code/Space*. We have yet to come to terms with this relationship at the level of data, however.
47. Ali, *Media Localism*, 5.
48. Kovach and Rosenstiel, *The Elements of Journalism*, 160.
49. Borgman, *Big Data, Little Data, No Data*.
50. Bowker and Star, *Sorting Things Out*.
51. Wagner, "The Americans Our Government Won't Count."
52. Bowker, "Biodiversity Datadiversity"; Gitelman, "*Raw Data*" Is an Oxymoron.
53. Geertz, *Local Knowledge*, 129.
54. Gieryn, "A Space for Place in Sociology"; Edwards, *A Vast Machine*.
55. For further reading on how nested locality can work, see Bratton, *The Stack*.

56. Star and Ruhleder, "Step toward an Ecology of Infrastructure," 114.
57. Edwards, *A Vast Machine*, 3.
58. See Galison, "Limits of Localism."
59. Pariser, *The Filter Bubble*.
60. Leurs, "Feminist and Postcolonial Data Analysis"; Houston and Jackson, "Caring for the Next Billion Mobile Handsets."
61. Tronto, *Moral Boundaries*, 103.
62. From a DPLA meeting attended by the author in 2013 in Cambridge, Massachusetts. See also Battles and Loukissas, "Data Artefacts."
63. Battles and Loukissas, "Data Artefacts."
64. Bellacasa, *Matters of Care*.
65. For more about gendered approaches to technical learning, see Turkle and Papert, "Epistemological Pluralism."

## CHAPTER 2

1. Data from the Arnold Arboretum are accessible online at <https://www.arboretum.harvard.edu/plants/data-resources>.
2. At the time, I was a Media Arts Fellow at Harvard University and member of the research group metaLAB@Harvard, part of the Berkman Klein Center for Internet and Society. Hailing from a range of humanistic subfields, this group was united by a determination to invent new counterdisciplinary approaches to working with collections of science and cultural heritage. Those present at Bussey Brook that day included Mathew Battles, Kyle Parry, and Jessica Yurkofsky as well as our host, arboretum senior researcher Peter Del Tredici.
3. For more on urban ecology in the Northeast, see Del Tredici, *Wild Urban Plants of the Northeast*.
4. For an extended contemplation of Bussey Brook Meadow, see Battles, *Tree*.
5. "Public Lab."
6. For more on aerial photography and its history, see Haffner, *The View from Above*.
7. From an interview by the author with Michael Dosmann, 2014.
8. Note that the provenance is not a place of origin but rather the name and address of a collector.
9. From an interview by the author with Peter Del Tredici, 2014.
10. Ibid.
11. This understanding reinforces prior studies of data that show institutionalized categories to be connected to specific social groups. Star and Griesemer, "Institutional Ecology, 'Translations,' and Boundary Objects."
12. From a record in BG-Base, the Arnold Arboretum's database of plant accessions.
13. Del Tredici and Kitajima, "Finding a Replacement for the Eastern Hemlock."
14. From an interview by the author with Peter Del Tredici, 2014.

15. For a detailed exploration of this concept in citizen science, see Gabrys, Pritchard, and Barratt, "Just Good Enough Data."
16. Edwards, *A Vast Machine*.
17. See Mayernik, Batcheller, and Borgman, "How Institutional Factors Influence the Creation of Scientific Metadata." See also Edwards et al., "Science Friction."
18. Gnoli, "Metadata about What?"
19. Kitchin and Lauriault, "Towards Critical Data Studies."
20. From an interview by the author with Peter Del Tredici, 2014.
21. For place-based perspectives on science and technology, see Galison and Thompson, *The Architecture of Science*; Livingstone, *Putting Science in Its Place*; Kirschenbaum, *Mechanisms*.
22. Haraway, "Situated Knowledges," 589.
23. Cosgrove, *Geography and Vision*; Crampton, *Mapping*; Kitchin, Gleeson, and Dodge, "Unfolding Mapping Practices"; Wilson, *New Lines*.
24. See Schivelbusch, *The Railway Journey*. STS scholar Laura Forlano first connected the use of this term to Schivelbusch's work in a phone conversation with the author.
25. For further reading on visualizations as instruments for reframing, see Hall, "Critical Visualization."
26. Accessible at <http://www.lifeanddeathofdata.org>.
27. From an interview by the author with Michael Dosmann, 2014.
28. Ibid.
29. Battles and Loukissas, "Data Artefacts."
30. Douglas, *Purity and Danger*; Mody, "A Little Dirt Never Hurt Anyone."
31. From an interview by the author with Peter Del Tredici, 2014.
32. Ibid.
33. From an interview by the author with Kyle Port, 2014.
34. The form of this visualization was influenced by a DensityDesign project, entitled "La Lettura," completed in 2012 in collaboration with Corriere della Sera. See <https://densitydesign.org/research/la-lettura/>, accessed July 27, 2018.
35. Morozov, "The Rise of Data and the Death of Politics"; Anderson, "The End of Theory"; Lohr, "The Age of Big Data."
36. Kalay and Marx, "The Role of Place in Cyberspace"; Mitchell, *City of Bits*; Graham, "The End of Geography or the Explosion of Place?"; Dourish, "Re-Space-Ing Place"; Irani et al., "Postcolonial Computing."
37. Dalton and Thatcher, "What Does a Critical Data Studies Look Like, and Why Do We Care?," 6.
38. Gieryn, "A Space for Place in Sociology." John Agnew writes of places as "discrete if 'elastic' areas in which settings for the constitution of social relations are located and with which people can identify." Agnew, "Representing Space," 263.
39. Buell, *The Future of Environmental Criticism*.



40. Hacking, "A Tradition of Natural Kinds"; Kitchin and Lauriault, "Towards Critical Data Studies."
41. Rob Kitchin and Gavin McArdle define big data as high magnitude in volume (terabytes or petabytes), velocity, variety, scope, resolution, flexibility, and relations with other data sets. Kitchin and McArdle, "What Makes Big Data, Big Data?" This litany of attributes, however, accounts for only the most ambitious of contemporary practices with big data. My use of the term is more in line with the work of danah boyd and Kate Crawford, who characterize big data as a phenomenon with not only technological but also cultural dimensions. See boyd and Crawford, "Critical Questions for Big Data."
42. See Battles, *Library*.
43. Hayles, *How We Became Posthuman*.

### CHAPTER 3

1. Data from the DPLA are accessible online at <https://pro.dp.la/developers>.
2. At the time of the Appfest, the DPLA was a project of the Berkman Klein Center for Internet and Society, an institution known for its efforts to map and remap the legal problems of the internet. The reader might be familiar with "Creative Commons," its revolutionary copyright schema currently in use across the web.
3. See <http://dp.la/info/about/history>.
4. For more on data infrastructures, see Kitchin, *The Data Revolution*.
5. After the first year as a nonprofit, the DPLA was well aware of the challenges of ingestion. The ingestion process exposed problems in the consistency, enrichment, and validation of the final data. As a result, these processes required significant hands-on work from DPLA staff—an approach that was presented as unsustainable at the time. See Matienzo and Ruderstorf, "The Digital Public Library of America Ingestion Ecosystem."
6. Palfrey and Gasser, *Interop*.
7. "One of our major priorities is to improve the quality of data we receive from libraries, archives, and museums. After all, it is from this data that so much else flows, including our ability to search well across collections and develop popular interfaces such as our browseable map. We plan to work extensively with our partners to get better data into this core piece of DPLA, and to refine that data as part of our valuable ingestion processing." "Digital Public Library of America Strategic Plan, 13.
8. Much has been written on the challenges of interface design; see Negroponete, *Being Digital*; Manovich, *The Language of New Media*; Galloway, *The Interface Effect*; Hookway, *Interface*.
9. "DPLA Metadata Application Profile."
10. Latour, *Science in Action*.
11. Not unlike the kind of conversation with materials that Donald Schön described as design. See Schön, *The Reflective Practitioner*. In fact, Turkle and Schön developed the notion of the "technological bricoleur" together while working on a study of MIT's Athena computing system and its implications for pedagogy at the institute. See Turkle and Schön, "The Athena Project." Also see Turkle, *Life on the Screen*.
12. The other participants in my Appfest group were Matthew Battles, Joshua Cash, Jessica Donaldson, Summer Leimart, Jim Reece, and Jeremy Throne.

13. We nonetheless went on to receive a small grant from the DPLA to develop our interface into the Library Observatory.
14. "Our survey and analysis have strengthened our conviction that multiple copies of a pre-copyright book cannot be deemed redundant based on catalog information alone." Ruotolo et al., "Book Traces @ UVA."
15. Ibid
16. Howard, "Book Lovers Record Traces of 19th-Century Readers."
17. Cited in Lewis, "ASU, MIT to Develop New Approaches to Library Print Collections."
18. "Digital Public Library of America Strategic Plan," 3.
19. From an interview by the author with Jeffrey Licht, 2013.
20. Battles and Loukissas, "Data Artefacts."
21. From a public lecture in 2015 in Cambridge, Massachusetts, by Marya McQuirter at Beautiful Data II.
22. Klein, "The Image of Absence."
23. Gusterson, *Nuclear Rites*.
24. From an interview by the author with Thomas Ma, 2013.
25. Ibid.
26. On the museum, see, for example, Star and Griesemer, "Institutional Ecology, 'Translations,' and Boundary Objects"; Battles and Loukissas, "Data Artefacts." On the laboratory, see, for example, Latour and Woolgar, *Laboratory Life*; Cetina, *Epistemic Cultures*; Lynch, *Art and Artifact in Laboratory Science*.
27. Gitelman, "*Raw Data*" Is an Oxymoron, 3.
28. Until recently, the big data phenomenon has largely been debated independently of local sites of production and use. Scholars of big data have sought to define their subjects in more general terms. On technical definitions, in terms of their "volume, velocity, and variety," see Kitchin and McArdle, "What Makes Big Data, Big Data?" On historical definitions, as a step change from past practices, see Anderson, "The End of Theory." On practical definitions, focusing on the difficulty of managing them, see Shneiderman, "The Big Picture for Big Data." On ideological definitions, by illuminating their "mythology," see boyd and Crawford, "Critical Questions for Big Data." Moreover, popular articles about big data depict them as ubiquitous tools for research and decision making across domains. See Lohr, "The Age of Big Data"; Mayer-Schoenberger and Cukier, "The Rise of Big Data."
29. Mayer-Schoenberger and Cukier, "The Rise of Big Data."
30. Bowker and Star, *Sorting Things Out*, 34.
31. See Karadkar et al., "Introduction to the Digital Public Library of America API."
32. For more on the origins of this data format, see "Dublin Core Metadata Initiative."
33. Bush, "As We May Think."
34. Ibid. Orit Halpern writes that memex "is considered an important contribution to the dream of networked, hyperlinked, and personal computing machines." See Halpern, *Beautiful Data*, 73.
35. Bush, "As We May Think."

36. Halpern, *Beautiful Data*, 73.
37. Nelson and Brand, *Computer Lib/Dream Machines*.
38. It “would certainly beat the usual file clerk,” he exclaims. See Bush, “As We May Think.”
39. Kitchin, *The Data Revolution*.
40. Mayer-Schoenberger and Cukier, “The Rise of Big Data.”
41. Retrieved originally by Matthew Battles from Darton, “Jefferson’s Taper.”
42. Nye, *American Technological Sublime*.
43. boyd and Crawford, “Critical Questions for Big Data.”

#### CHAPTER 4

1. Data from the NewsScape archive are accessible online at <http://tvnews.library.ucla.edu>.
2. Like several other data visualizations in this book, I created NewsSpeak in collaboration with a former Georgia Tech graduate student, Peter Polack.
3. “UCLA Library NewsScape.”
4. Others include Media Cloud, based at MIT, and the Internet Archive.
5. Dourish, *The Stuff of Bits*, 4.
6. Mayer-Schoenberger and Cukier, “The Rise of Big Data.”
7. For a comprehensive look at how newspapers went digital, see Boczkowski, *Digitizing the News*.
8. Gillespie, “The Relevance of Algorithms,” 167.
9. Manovich, *The Language of New Media*, 223.
10. See Eubanks, *Automating Inequality*; Noble, *Algorithms of Oppression*; Finn, *What Algorithms Want*. Social research is only beginning to address NLP algorithms specifically. See Hovy and Spruit, “The Social Impact of Natural Language Processing.”
11. Loukissas and Pollock, “After Big Data Failed.”
12. Office of the Director of National Intelligence, “Joint DHS, ODNI, FBI Statement on Russian Malicious Cyber Activity.”
13. CNBC reports that “at his first press conference since Election Day, President-elect Donald Trump answers questions about fake news and the media, while avoiding press questions from organizations (such as CNN) he considers ‘terrible.’” See “Trump to CNN Reporter.”
14. As CNN reports, “Executives from Facebook, Twitter and Google are set to testify before Congress on Tuesday about foreign nationals meddling in the 2016 elections.” See Fiegerman and Byers, “Facebook, Twitter, Google Testify before Congress.”
15. Bird, Klein, and Loper, “Preface.”
16. Recent interest in machine learning has been spurred by accessible online courses such as What Is Machine Learning? taught by Andrew Ng on Coursera.
17. This is an approach explicitly suggested by Rob Kitchin and Gavin McArdle: “Rather than studying an algorithm created by others, a researcher reflects on and critically interrogates

- their own experiences of translating and formulating an algorithm.” See Kitchin and McArdle, “What Makes Big Data, Big Data?,” 23.
18. See, for example “Stanford CoreNLP.”
  19. This pseudocode was written by Georgia Tech student Benjamin Sugar, based on a program collaboratively created by the author and Peter Polack.
  20. Steen and Turner, “Multimodal Construction Grammar.”
  21. Based on an interview by the author with a professional transcriptionist.
  22. One well-known example of this is the macro “Barry the bomber,” which became a shorthand for Barack Obama in the United Kingdom, where transcribers had trouble pronouncing his name. (Obama was known by this nickname at Occidental College, where he was on the basketball team and excelled at long-range shots.) There is at least one known occurrence when the closed captioning read the macro rather than the US president’s name, to the embarrassment of the broadcasters.
  23. From an interview by the author with Sergio Goldenberg, 2016.
  24. In the past, storing the news was only something done by large libraries. It was held in basements or on microfiche, rarely accessed, and typically only by serious researchers or journalists looking for a backstory.
  25. Rogers, “Wall Street Journal.”
  26. Based on *Wall Street Journal*, “2016 Presidential Election Calendar.”
  27. See, for example, Nasukawa and Yi, “Sentiment Analysis”; Pang, Lee, and Vaithyanathan, “Thumbs Up?”
  28. In earlier chapters, I explain the use of visualization for reframing as a kind of critical reflection. See Hall, “Critical Visualization.”
  29. The idea that conventional forms of reading are innately human is a bit naive. After all, the typography of modern printing, the bound form of the codex, the institutionalization of book collecting, and even the work of the artificial light bulb make the way we read its own form of human-machine interaction.
  30. From an interview by the author, 2016.
  31. Watzman, “Internet Archive TV News Lab.”
  32. Benkler, Faris, and Roberts, “Study.”
  33. Costanza-Chock and Rey-Mazón, “PageOneX.”
  34. See Snopes.
  35. Albright, “Itemized Posts and Historical Engagement”; Timberg, “Russian Propaganda May Have Been Shared Hundreds of Millions of Times.”
  36. Fake News Challenge.
  37. Jones, “Natural Language Processing.”
  38. As Peter Hancox writes, “Its power came from its very limited domain and any attempt to scale the system up would result in increasingly less effective systems.” See Hancox, “A Brief History of NLP.”

39. Boguraev et al., “The Derivation of a Grammatically Indexed Lexicon from the Longman Dictionary of Contemporary English.”
40. Ralph, “Information Extraction,” 2.
41. Fillmore, “Encounters with Language,” 711.
42. “Natural Language Toolkit.”
43. Even today, statistical NLP is still dependent on working in specific domains. Watson, the machine that won the TV gameshow *Jeopardy* over human contestants, is a modern marvel, but only when it has the right kind of data—data that are expensive to generate and not widely available.
44. Weizenbaum, “Computer Power and Human Reason,” 371.
45. For an example of how to read NLP as an otherworldly means of interacting with text, see Binder, “Alien Reading.”
46. For more on training algorithms, see Diakopoulos, “Algorithmic Accountability.”
47. Taylor, Marcus, and Santorini, “The Penn Treebank.”
48. For an explanation of the Penn Treebank, see “Wayback Machine.”
49. “The Penn Treebank (PTB) project selected 2,499 stories from a three-year *Wall Street Journal* (WSJ) collection of 98,732 stories for syntactic annotation.” See “Wayback Machine.”
50. For the complete NLTK book, see <http://www.nltk.org/book/ch05.html>.
51. Marcus, Marcinkiewicz, and Santorini, “Building a Large Annotated Corpus of English,” 316.
52. There are increasingly alternatives to the Penn Treebank. Today many technology companies, such as Amazon and Facebook, need their own domain-specific annotated corpora. They are hiring people specifically to help them build these.
53. Marcus, Marcinkiewicz, and Santorini, “Building a Large Annotated Corpus of English,” 313.
54. At the time of its use in the formation of the Penn Treebank, when it became the model for natural language in NLTK, the *WSJ* was only published five days per week. A weekend edition was added in 2005.
55. This is more true in English than in other languages.
56. Houston, “The Mysterious Origins of Punctuation.”
57. Eisenstein, “What to Do about Bad Language on the Internet”; Brock, “From the Blackhand Side.”
58. Kitchin, *The Data Revolution*.
59. Gillespie, “The Relevance of Algorithms”; Seaver, “Knowing Algorithms.” Samir Passi and Steven Jackson have explained this as a kind of data vision: “the ability to organize and manipulate the world with data and algorithms, while simultaneously mastering forms of discretion around why, how, and when to apply and improvise around established methods and tools in the wake of empirical diversity.” Passi and Jackson, “Data Vision.”
60. Manovich, *The Language of New Media*.
61. Borges, *Collected Fictions*.
62. Dourish, *The Stuff of Bits*.

63. Baudrillard, *Simulacra and Simulation*.
64. Based loosely on the notion that making media helps people make knowledge. See Piaget, *The Psychology of Intelligence*; Papert, *Mindstorms*.
65. Latour, *Reassembling the Social*; Law and Benschop, "Resisting Pictures"; Mol, *The Body Multiple*. For an example of actor-network theory applied to the news, see Turner, "Actor-Networking the News."
66. Latour wrote that "the set of statements considered too costly to modify constitute what is referred to as reality." The news, like science, is a contest to construct reality. "Scientific activity is not 'about nature,' it is a fierce fight to construct reality." See Latour and Woolgar, *Laboratory Life*, 243.
67. Data and algorithms are in a symbiotic relationship, argues Manovich. When algorithms are used to manipulate data, the more complex the data structure, the simpler the algorithm can be. But this way of thinking is reductive and overly generalized. How can we turn from claims about cultural algorithms to the exploration of algorithmic practices, and how they reshape both data and reality?
68. From an interview by the author with Sergio Goldenberg, 2016. Also see Goldenberg's work in Murray et al., "StoryLines."
69. Kovach and Rosenstiel, *The Elements of Journalism*.
70. Zuckerman, *Digital Cosmopolitans*.

## CHAPTER 5

1. Data from the Fulton County Board of Tax Assessors are accessible online at <http://fultonassessor.org/>.
2. According to "Rise of the Corporate Landlord," "Companies like the Blackstone Group, American Homes4Rent, Colony Financial, Silver Bay, Starwood Waypoint, and American Residential have spent approximately \$20 billion to purchase roughly 150,000 single family homes, nationwide, and convert them into rental properties." Also see Call, Powell, and Heck, "Blackstone."
3. Atlanta Intown neighborhoods are inside the I-285 loop.
4. Battles, *Tree*.
5. Hanan, "Home Is Where the Capital Is."
6. Immergluck, Carpenter, and Lueders, "Declines in Low-Cost Rented Housing Units in Eight Large Southeastern Cities."
7. Units costing \$750 per month or less are dwindling at the rate of 5 percent per year. See *ibid*.
8. See Finn, *What Algorithms Want*. The phrase "interface layer" is used by designer turned venture investor Scott Belsky. Belsky, "The Interface Layer."
9. Belsky, "The Interface Layer."
10. When Adele Goldberg and Alan Kay first wrote about the "personal" computer, they were imagining an expressive tool, not an instrument for consumerism. See Kay and Goldberg, "Personal Dynamic Media." Experimental humanities scholar Ed Finn explains that companies working to create the "interface economy" establish "wrappers around existing services,

- bundling, organizing, and demystifying them for a painless user experience.” Finn, *What Algorithms Want*, 129.
11. See Kitchin, *The Data Revolution*; Ramirez et al., “Data Brokers”; Small, *FTC Report Examines Data Brokers*.
  12. Yelp.
  13. Nextdoor.
  14. Uber.
  15. “Zillow.”
  16. Galloway, *The Interface Effect*.
  17. For a summary of the confusion over what counts as context in information technology design, see Seaver, “The Nice Thing about Context Is That Everyone Has It.”
  18. For example, this was a common refrain at the 2016 information design and visualization conference, Information+, in Vancouver, Canada.
  19. See O’Neil and Schutt, *Doing Data Science*. See also popular articles such as Richtel, “How Big Data Is Playing Recruiter for Specialized Workers.”
  20. For more on the problem of technological platitudes, see Paul Dourish’s critical studies of disembodied interaction. Dourish, *Where the Action Is*.
  21. Meanwhile, there are humans, particularly in communities that do not or cannot produce their own data, who actually lack a voice in deliberations over issues of utmost importance to them.
  22. As danah boyd and Kate Crawford write, “There is a value to analyzing data abstractions, yet retaining context remains critical, particularly for certain lines of inquiry. Context is hard to interpret at scale and even harder when data are reduced to fit into a model. Managing context in light of Big Data will be an ongoing challenge.” boyd and Crawford, “Critical Questions for Big Data.”
  23. Loukissas, “Taking Big Data Apart.”
  24. Today, context is a topic of widespread interest in communities that develop information technologies. Researchers and designers seek to shift the burden of accounting for the relevant circumstances of interaction from humans onto computers. In the course of their development of “context-aware” systems, they have, understandably, put forward definitions of *context* that fit well within a data processing framework. Some adopt a definition that presents *context* as data not taken into account, but that can be retrieved at a later point. Other definitions suggest a more dynamic approach, treating *context* as data that change based on the system and its use. Some see it as the users’ environment, while others view *context* as the environment of the computing system. Anind Dey offers an encompassing informational view: “Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.” Dey, “Understanding and Using Context,” 3. All these definitions frame *context* as a kind of data or information.
  25. Dourish, “What We Talk about When We Talk about Context,” 22.
  26. Ibid.

27. Although not the first to use the term, design theorist Donald Norman popularized the word *affordance* for explaining the properties of interactive objects. See Norman, *The Design of Everyday Things*.
28. Geertz, *The Interpretation of Cultures*, 14.
29. See foundational work on epistemic cultures and communities. Cetina, *Epistemic Cultures*; Keller, *Making Sense of Life*.
30. This thought experiment was first described by philosopher Gilbert Ryle and later adapted by Geertz. See Geertz, *The Interpretation of Cultures*.
31. *Ibid.*, 7.
32. Rascoff and Humphries, *Zillow Talk*, 17.
33. Or it might be a reference to a children's book by Dr. Seuss titled *There's a Wocket in My Pocket*, in which the last line is "The ZILLOW on my PILLOW always helps me fall asleep."
34. Many internet companies such as Uber and Airbnb similarly seek to intervene in a market without adhering to previously existing rules and regulations.
35. Realtor.com was the first.
36. For an overview of data brokers and their practices, see Small, *FTC Report Examines Data Brokers*.
37. From remarks by Stewart Brand at the Hackers Conference in Marin County, California, in 1984.
38. For a more in-depth history of mapping technologies, see Drucker, *Graphesis*; Crampton, *Mapping*; Kitchin, Gleeson, and Dodge, "Unfolding Mapping Practices."
39. The complex history of this phrase was the subject of a *New York Times* article. See Safire, "Location, Location, Location."
40. Robinson, *The Look of Maps*.
41. For a description of how representations produce reality, see Baudrillard, *Simulacra and Simulation*; Wood and Fels, *The Natures of Maps*.
42. See Kitchin, Gleeson, and Dodge, "Unfolding Mapping Practices," 15.
43. Sellers know that resetting the listing date can make their property more appealing. In this sense it is kind of like a digital paint job. After all, properties that have been sitting on the market for too long, like houses where the paint is peeling, look neglected and hence less desirable.
44. Much has been written about information technology and discourse that I will not reprise here. My use of the term *discourse* as a means of establishing and maintaining power relations is rooted in the work of Michel Foucault. See *The Archaeology of Knowledge: And the Discourse on Language*.
45. For the still-best reference for understanding the term *public* and its implications, see Dewey, *The Public and Its Problems*. For further reading on the concept of *publics* and its relationship to design, see LeDantec, "Infrastructuring and the Formation of Publics in Participatory Design." For more on the ethics of using big data that's publicly available, see boyd and Crawford, "Critical Questions for Big Data."



46. See <https://zillow.zendesk.com/hc/en-us/articles/213218507-Where-does-Zillow-get-information-about-my-property>.
47. See <https://zillow.zendesk.com/hc/en-us/articles/212561938-What-is-a-property-page>.
48. From an interview by the author, 2016.
49. From an interview by the author, 2016.
50. See Rascoff and Humphries, *Zillow Talk*.
51. The only exceptions to this that I have found are in low-income neighborhoods. Some multi-unit properties do not get a Zestimate. Is Zillow not able to find any records on these units? We cannot be certain. But there are some homes that Zillow's algorithm implicitly does include in its otherwise-comprehensive perspective on the market. Thanks to Juan Carlos Rodriguez for first calling attention to such properties along the Buford Highway corridor in Atlanta.
52. In some states, the Multiple Listing Service, which works directly with realtors to collect data about properties on the market, also provides Zillow with bare-bones listings. This is the case in Georgia, where First Multiple Listing Service Inc. has partnered with Zillow to allow information listed in their databases to be used.
53. See Rascoff and Humphries, *Zillow Talk*.
54. Seaver, "Knowing Algorithms," 9.
55. See <http://www.zillow.com>.
56. "Pricing conversations that may be started with the review of the Zestimate should ultimately be augmented with the input of opinions from local real estate professionals (agents, brokers, and appraisers)." See <http://www.zillow.com/research/putting-accuracy-in-context-3255>.
57. Howell, "How Accurate Is Zillow's Zestimate?"
58. Humphries, "How Accurate Is the Zestimate?"
59. Rascoff and Humphries, *Zillow Talk*, 223.
60. Humphries, "How Accurate Is the Zestimate?"
61. "In September 2012, the Zestimate was just as likely to be too low as too high; now, it is roughly twice as likely to be too low." Howell, "How Accurate Is Zillow's Zestimate?"
62. Murray, *Hamlet on the Holodeck*, 71.
63. Bogost, *Persuasive Games*.
64. If Zillow is in fact indifferent to the truth value of its Zestimate, this would meet the formal definition of *bullshit*. Frankfurt, *On Bullshit*. After all, the company does not profit from being right. Instead, it profits from your repeat visits and the advertising that attention supports. Thanks to information scholar Brian Butler for pointing this out during a conversation at the 2018 iConference.
65. Zestimates are not calculated for a significant amount of low-cost housing. Why? What does this tell us about Zillow's agenda?
66. This definition of digital civics comes from Carl DiSalvo, during a presentation at Georgia Tech, September, 22, 2016. For more on digital approaches to civics, see Gordon and Mihailidis, *Civic Media*.

67. There are several ongoing efforts at Georgia State University to trace these trends. See <https://atmaps.com/>; <http://digitalcollections.library.gsu.edu/cdm/planningatl>; <http://dh2016.adho.org/abstracts/277>.
68. Housing Justice League and Research|Action Cooperative, “BeltLining.”
69. See Lands, *Culture of Property*; “From Herndon Homes to two Georgia Domes.”
70. Ponczek and Wei, “The 10 Most Unequal Cities in America.”
71. The images in this chapter were prototyped early on using the Processing development environment, a platform created by Ben Fry and Casey Raes in order to enable visually oriented people to program graphics using Java—a preexisting object-oriented programming language. They were later finalized using JavaScript in collaboration with Peter Polack, a graduate student at Georgia Tech.
72. To locate Fulton County tax data, see [http://qpublic9.qpublic.net/ga\\_display\\_dw.php?county=ga\\_fulton&KEY=14%20004700060091&show\\_history=1&](http://qpublic9.qpublic.net/ga_display_dw.php?county=ga_fulton&KEY=14%20004700060091&show_history=1&).
73. See <http://www.housingjusticeleague.org>.
74. Although the technique dates back to the nineteenth century, the term *small multiple* is best explained in Tufte, *The Visual Display of Quantitative Information*.
75. The Fulton County tax assessor explained in a 2016 interview with the author that such data is error prone.
76. These median values are reported by Trulia, so take them with a grain of salt. For Trulia market trends, see [https://www.trulia.com/real\\_estate/Atlanta-Georgia/market-trends/](https://www.trulia.com/real_estate/Atlanta-Georgia/market-trends/).
77. Atkinson, “Does Gentrification Help or Harm Urban Neighbourhoods?”
78. See Desmond, *Evicted*.
79. The Department of Housing and Urban Development defines affordable housing as housing that is accessible to those earning the median income. This definition is highly contested, for it means that half of the population can’t afford it. This is not the same as the median housing price, which is probably much higher. HUD, “United States Department of Housing and Urban Development.”
80. Even the founders of Zillow acknowledge that listings are full of “hidden meanings.” Rascoff and Humphries, *Zillow Talk*.
81. I am currently supervising work by Georgia Tech student Eric Corbett to study the coded language about gentrification present in Zillow listings. This work focuses on neighborhoods near the Old Fourth Ward. Corbett’s preliminary findings reveal a wide range of variation in how neighborhood change is represented. Corbett has looked at listings from 2012 to 2015. He collected 246 each from Cabbagetown and Reynoldstown. Of all the listings for Cabbagetown, 88 mentioned the neighborhood, and 68 of those characterized the neighborhood in social terms relating to discourses of gentrification. Meanwhile, Reynoldstown listings mentioned the neighborhood 82 times—65 of which portrayed the neighborhood in specific terms. The descriptors used in listings in these neighborhoods, however, were not equivalent. Cabbagetown was described as *historic* in 17 separate listings. The same term showed up only twice for Reynoldstown. Meanwhile, terms such as *new* and *hot* were much more prevalent in Reynoldstown. As of this writing, the full study remains unpublished.
82. See Zukin, Lindeman, and Hurson, “The Omnivore’s Neighborhood?”
83. See <https://dor.georgia.gov/property-tax-rates>.

84. For more on how mass appraisals are conducted, see <https://dor.georgia.gov/property-tax-rates>.
85. Anything over 30 percent of one's income is considered unaffordable by the Department of Housing and Urban Development. HUD, "United States Department of Housing and Urban Development."
86. See <https://dor.georgia.gov/property-tax-exemptions>.
87. The rules are different in other cities.
88. Landaw, "Atlanta Declared Renters State of Emergency."
89. Housing Justice League and Research|Action Cooperative, "BeltLining."
90. "The Atlanta BeltLine Project."
91. Housing Justice League and Research|Action Cooperative, "BeltLining," 32.
92. Ibid.
93. For further reading about the Atlanta BeltLine and its unequal impact on Intown neighborhoods, see Immergluck and Balan, "Sustainable for Whom?"

## CHAPTER 6

1. See Porter, *Trust in Numbers*; Bowker, *Memory Practices in the Sciences*; Galison, "Limits of Localism."
2. Simon, *The Sciences of the Artificial*.
3. Schön, *The Reflective Practitioner*, 175.
4. See Sengers et al., "Reflective Design."
5. Agre, *Computation and Human Experience*.
6. Dunne and Raby, *Design Noir*.
7. DiSalvo, *Adversarial Design*. For more on how critical design has been taken up and debated among researchers in human-computer interaction, see Bardzell and Bardzell, "What Is 'Critical' about Critical Design?"; Pierce et al., "Expanding and Refining Design and Criticality in HCI."
8. For a review of the ways in which design can be critical without replicating humanistic approaches based in other media, see Pierce et al., "Expanding and Refining Design and Criticality in HCI."
9. Johanna Drucker provokes us to reconsider our use of the term *data* and offers suggestions for design. Drucker, "Humanities Approaches to Graphical Display." Geoffrey Bowker and Susan Leigh Star's book ends with insightful "design exigencies." Bowker and Star, *Sorting Things Out*, 324. Jeffrey Bardzell and Shaowen Bardzell also offer design implications for those seeking critical reflection. Bardzell and Bardzell, "What Is 'Critical' about Critical Design?" Catherine Ignazio and Lauren Klein supply prescriptions for a critical and feminist approach to data visualization. D'Ignazio and Klein, "Feminist Data Visualization."
10. Dourish, *Where the Action Is*.
11. One danger of offering principles is the "not invented here" problem. Practitioners want to make their own interpretations of theory in their own ways. Moreover, practitioners are

- more likely to remember models. As the *Oxford English Dictionary* defines it, “Model (n) A thing used as an example to follow or imitate.” For further reading on models in science, see Sismondo, “Models, Simulations, and Their Objects.” For further reading on models in education, see Papert, *Mindstorms*. And in design, see Loukissas, *Co-Designers*.
12. For more on how architectural precedents are used in order to convey design ideas, see Rowe, *The Mathematics of the Ideal Villa and Other Essays*; Le Corbusier and Etchells, *Towards a New Architecture*.
  13. In recent years, software such as Tableau and software development toolkits such as the JavaScript library D3 have significantly flattened out and generalized the way that data are presented. In many ways, these tools are extraordinary accomplishments and positive contributions, but they should not be accepted uncritically.
  14. Norman, *The Design of Everyday Things*, 3.
  15. Mendes and Allison, “Bear 71 VR.” For a full analysis of this project, see Ray, “Rub Trees, Crittercams, and GIS.”
  16. Mendes and Allison, “Bear 71 VR.”
  17. I taught several classes on the subject of data documentaries at Georgia Tech in 2015–2016.
  18. Gabrys, *Program Earth*.
  19. Thorp, “A Sort of Joy.”
  20. Ibid.
  21. Lupi and Posavec, *Dear Data*, ix.
  22. Ibid., x.
  23. Rhodes, “This Guy Obsessively Recorded His Private Data for 10 Years.”
  24. Lupi and Posavec, *Dear Data*, 2.
  25. Ibid., 5.
  26. Lupi and Posavec, *Dear Data*.
  27. See Laura Kurgan’s analysis of Google Earth. Kurgan, *Close Up at a Distance*.
  28. Dosemagen, Warren, and Wylie, “Grassroots Mapping.” “Civic science” is a reference to Fortun and Fortun, “Scientific Imaginaries and Ethical Plateaus in Contemporary U.S. Toxicology.”
  29. Dosemagen, Warren, and Wylie, “Grassroots Mapping.”
  30. Anderson et al., “A Grassroots Remote Sensing Toolkit Using Live Coding, Smartphones, Kites, and Lightweight Drones.”
  31. For an explanation of how algorithms used to predict recidivism by the criminal justice system can be racist, see Diakopoulos and Friedler, “We Need to Hold Algorithms Accountable.”
  32. For an example of how counterdata have been used to reveal the biases in facial recognition algorithms, see Buolamwini and Gebru, “Gender Shades.” For an example of a counteralgorithm designed to “obfuscate browsing data and protect users from tracking by advertisers,” see Nissenbaum, Howe, and Zer-Aviv, “AdNauseam.”

33. Brown, "Slave Revolt in Jamaica, 1760–1761," <http://revolt.axismaps.com/project.html>, accessed August 28, 2018.
34. As Brown explains, "Scholars working in subaltern history rarely have the kind of big databases that inspire projects in text mining, topic modeling, or network analysis." Brown, "Mapping a Slave Revolt," 138.
35. Ibid.
36. Manovich, "Database as Symbolic Form," 85.
37. Brown, "Mapping a Slave Revolt," 138.
38. Anti-Eviction Mapping Project, "About."
39. Ibid.
40. See Loukissas, "Keepers of the Geometry."
41. There is an extensive literature on participatory methods and their potential uses in civic design. For an example, see LeDantec, "Infrastructuring and the Formation of Publics in Participatory Design."
42. Simon, *The Sciences of the Artificial*.

## CHAPTER 7

1. Totty, "The Rise of the Smart City"; Jackson, "Columbus under Construction to Become America's First 'Smart City.'" For a more comprehensive look at smart cities see Townsend, *Smart Cities*.
2. Loukissas and Pollock, "After Big Data Failed."
3. University of Texas at Austin, Texas Advanced Computing Center, "The Future of Search Engines."
4. Cadwalladr, "Google, Democracy, and the Truth about Internet Search."
5. See Google's web page "Our Company | Google."
6. Noble, *Algorithms of Oppression*, 14.
7. "What these right wing news sites have done ... is what most commercial websites try to do. They try to find the tricks that will move them up Google's PageRank system. They try and 'game' the algorithm." Cadwalladr, "Google, Democracy, and the Truth about Internet Search."
8. For more on how to tame everyday algorithms for the public good, see Eubanks, *Automating Inequality*; O'Neil, *Weapons of Math Destruction*; Pasquale, *The Black Box Society*.
9. For a primer on adversarial approaches to design, see DiSalvo, *Adversarial Design*.
10. See Bearne, "Is the 'Gig Economy' Turning Us All into Freelancers?"; Armstrong, "The People Making the On-Demand Economy Work."
11. For an overview of open data initiatives, see Kitchin, *The Data Revolution*.
12. I have started to explore this question through the development of *civic data guides* in collaboration with Catherine D'Ignazio, an assistant professor of civic media and data visualization at Emerson College, and Bob Gradeck, who manages the Western Pennsylvania Regional Data Center project at the University of Pittsburgh.

13. See Gradeck, "Data User Guides." Gradeck and his colleagues at the Western Pennsylvania Regional Data Center have already produced more than forty guides at the time of this writing. Their work has been inspired by other guides to civic data, such as Pettit and Droesch, "A Guide to Home Mortgage Disclosure Act Data," and Smart Chicago Collaborative, "Crime and Punishment in Chicago." For parallel work in this area, also see Gebru et al., "Datasheets for Datasets."
14. See data guides offered by the US Department of Education and US Centers for Disease Control and Prevention.
15. See Gradeck, "Data User Guides."
16. For an overview of ethical issues in these areas, see boyd and Crawford, "Critical Questions for Big Data"; Kitchin, *The Data Revolution*.
17. Bowker, *Memory Practices in the Sciences*, 184.
18. Bellacasa, *Matters of Care*, 100.
19. On "instrumental rationality" and "datafication," see Mattern, "Methodolatry and the Art of Measure."
20. Ortner, *Anthropology and Social Theory*, 43.