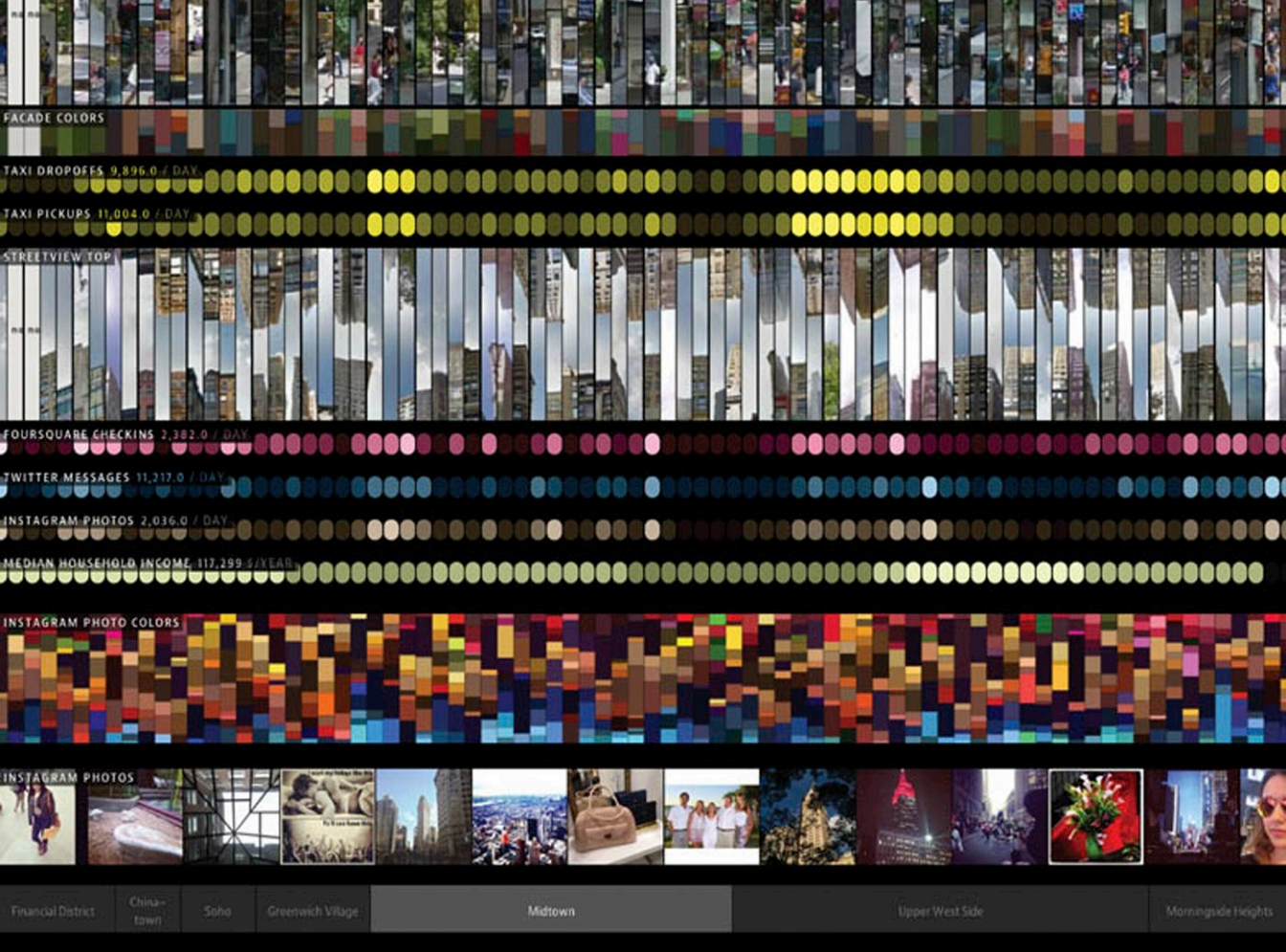
LEV MANOVICH

CULTURAL ANALYTICS



Cultural Analytics

Cultural Analytics

Lev Manovich

The MIT Press
Cambridge, Massachusetts
London, England

This book was set in ITC Stone Serif Std and ITC Stone Sans Std by New Best-set Typesetters Ltd.

Library of Congress Cataloging-in-Publication Data

Names: Manovich, Lev, author.
Title: Cultural analytics / Lev Manovich.
Description: Cambridge, Massachusetts : The MIT Press, [2020] | Includes bibliographical references and index.
Identifiers: LCCN 2020003045 | ISBN 9780262037105 (hardcover)
Subjects: LCSH: Culture—Research—Statistical methods. | Culture—Research—Data processing. | Mass media—Research—Statistical methods. | Mass media—Research—Data processing. | Information visualization.
Classification: LCC HM623 .M365 2020 | DDC 306.0285—dc23
LC record available at https://lccn.loc.gov/2020003045

10 9 8 7 6 5 4 3 2 1

Contents

Acknowledgments

I am very grateful to all the people and institutions that made this book possible:

The MIT Press: Doug Sery, senior acquisitions editor, Noah Springer, assistant acquisitions editor, Kathleen Caruso, manuscript editor, and Melinda Rankin, copyeditor.

Larry Star, director of the California Institute for Telecommunications and Information Technology (Calit2), Ramesh Rao, director of the UCSD Division of Calit2, and all the staff at Calit2, which has been supporting the work of our lab since its start in 2007.

Noah Wardrip-Fruin, who cofounded Software Studies Initiative (later renamed Cultural Analytics Lab) with me in 2007. Sheldon Brown, who invited us to the Center for Research in Computing and the Arts, which became our lab's home from 2008 to 2012. Mathew Gold who supported my work at The Graduate Center, CUNY, after I started teaching there in 2013.

Lab members, 2007–2018: Jeremy Douglass, William Huber, Tara Zepel, Cicero Inacio da Silva, Jay Chow, Everardo Reyes, Mehrdad Yazdani, Damon Crockett, Nadav Hochman, Alise Tifentale, and Agustin Indaco.

Lab collaborators and visiting fellows: Moritz Stefaner, Dominikus Baur, Daniel Goddemeyer, Miriam Redi, Nadav Hochman, Almila Akdag, Jean-François Lucas, Tristan Thielmann, Hijoo Son, Kay O'Halloran, Isabel Galhano Rodrigues, Falko Kuester, Jim Hollan, Matthew Fuller, Brynn Shepherd, and Leah Meisterlin.

Graduate and undergraduate students who worked in the lab: So Yamaoka, Sunsern Cheamanunku, Matias Giachino, Xiangfei Zeng, Cherie Huang, Chanda L. Carey, Daniel Rehn, Laura Hoeger, Rachel Cody, Devon Merill, Jia Gu, Agatha Man, Nichol Bernardo, Bob Li, Kedar Reddy, Christa Lee, Victoria Azurin, Xiaoda Wang, and Nadia Xiangfei Zeng.

The organizers of the UCLA IPAM Culture Analytics Institute (2016): Timothy Tangherlini, Tina Eliassi-Rad, Mauro Maggioni, and Vwani Roychowdhury.

The universities and educational programs where I have been permanent or visiting faculty between 2005 and 2020 (from the moment I first thought of cultural analytics to finishing this book): University of California, San Diego (UCSD); The Graduate Center, City University of New York (CUNY), National University of Singapore (NUS); Strelka Institute for Media, Architecture and Design; the European Graduate School (EGS); and the Institute of Social Sciences and Humanities, Tyumen State University (UTMN).

Illustrations

Examples of projects from our lab are used throughout the book to illustrate the concepts and techniques being presented. In my classes and workshops, I use the same approach because it allows me to show students the concrete steps involved in creating such projects, to discuss the multiple choices each step entails, and to point out what remained outside the analysis. Each of our projects has its own website or a web page where you will find descriptions, high-resolution color visualizations, and in some cases interactive interfaces with the datasets. You can access them from the Projects page of the Cultural Analytics Lab website:

http://lab.culturalanalytics.info/p/projects.html

Note that although some of the visualizations appear in the book as color plates and others as grayscale figures, the originals are all in full color. Most of the visualizations are the result of joint work between lab members, with a few people working on each project—creating the data, analyzing it and interpreting results, and making visualizations.

Introduction: How to See One Billion Images

> The impact of the computer in the human sciences, however, is likely proportionally to be more revolutionary in the long run [than in physical and life sciences]. . . . Some of it has to do simply with willingness to take advantage of the opportunity, or predisposition through already extensive use of processes, especially statistical, facilitated by the computer. More, perhaps, has to do with what a computer, in a sense like a telescope or a microscope, can enable us to *see*. In simplest terms, computer processing, properly prepared, can enable us to see relations and patterns in masses of data previously too large to comprehend; and to see the literal consequences of an idea applied to data, if not uniquely, then certainly far more inexorably and quickly.
>
> —Dell H. Hymes, "Introduction," in *The Use of Computers in Anthropology*, 1965[1]

This book is situated at the intersection of data science, media studies, and digital culture studies. It presents selected concepts and methods for computational analysis of cultural data. These methods can be used to explore digitized historical artifacts and contemporary digital media. While we can apply them to a single or a few artifacts, they become especially important if we want to explore millions of artifacts.

In fact, the astonishing scale of digital culture is what motivated me to start exploring these methods in 2005 and eventually write this book. How can we understand contemporary popular photography that grows by billions of images every day? Or contemporary music as represented by hundreds of millions of songs shared by twenty million creators on SoundCloud? Or the content of four billion boards on Pinterest?[2] Or patterns in the intellectual interests of people in 190 countries as represented by 330,000 Meetup groups and 84,000 events per month (as of 2019)?[3] This is also "digital culture" because these physical events are enabled by the Meetup web platform. In my view, the only possible way to study the patterns, trends, and dynamics of contemporary culture at that scale is to use data science methods.

You do not need to have a background in data science, programming, statistics, or math to use this book. My intended audiences are academic researchers and students in art, design, the humanities, social sciences, media studies, data science, and computer science; professionals working in design, photography, film, urban design, architecture, journalism, museum and library fields, curating, and culture management; and everybody who works with social media and the web in any role (creator, blogger, strategist, manager, developer, marketer, etc.).

Even if you have no interest in analyzing cultural datasets yourself, you are encountering such analysis on a daily basis. Maybe you are looking at your Facebook, Instagram, or Weibo analytics, or Google Analytics for your blog or website, or using a social media monitoring dashboard at work. And if you don't pay attention to such data, you are constantly interacting with the results of computational analysis when you do anything digital. For example, every time you capture a photo, the phone camera algorithms automatically choose the exposure and adjust the contrast of the photo and also identify the type of scene and objects in the photo.[4] Computational analysis of media artifacts and user interactions is what enables web search, recommendations, filtering, customization, interactions with digital devices, behavioral advertising, and other operations that form the "vocabulary" of digital culture. For example, web search engines such as Baidu, Bing, Yandex, or Google rely on continuous computational analysis of contents of billions of web pages, online images, and other web content to bring you relevant results.

I think that to be literate in such a society, you need to know the core ideas and principles that make such operations possible. This book is a gentle, nontechnical introduction to some of these ideas. Thus, it teaches you how you can explore cultural datasets yourself and also explains how our society thinks using data and algorithms.

Looking at Culture with Computers

In this book, you will find many examples of computational culture analysis from many researchers and also from my own lab. Right now, I want to describe two examples to illustrate the possibilities and challenges of this analysis.

The first example is the project *Elsewhere* that my collaborators and I have been working on since 2018. The project investigates the growth and diffusion of contemporary culture, taking into account many smaller cities, as opposed to only a handful of global capitals. Today, a small selection of these capitals gets a disproportional amount of attention in media, research studies, and various ratings. Therefore, it is easy to assume that a handful of "top" cities continue to act as the "centers" and the rest

of the world is still on the periphery, receiving new ideas after a delay. But what is the real picture? How did globalization and the rise of new communication technologies change the geography of culture? Can we find every contemporary cultural trend today in thousands of smaller cities? Is it possible that some of these cities are more culturally innovative precisely because of their distance from the capital and their smaller size? Are there big parts of the world left today that are not aware of these trends and do not innovate? How has contemporary culture developed and diffused around the world since the beginning of globalization? Was the growth even or uneven, accelerating or slowing down in some periods? Are the growth patterns the same for different cultural fields, or does each of them develop in its own way?

Certainly, no one project can answer all these questions. The goal of *Elsewhere* is to develop and a test a new methodology that uses public data about cultural events and the places that organize these events. The numbers of these places and events in the world today have become so large that we can now to treat them as "big data"—and use data science methods for their analysis. This perspective should allow us to create much more detailed maps and timelines of contemporary culture than what is provided in existing studies of culture industries or lists of cultural institutions. We also run text analytics on all texts that organizers publish about their events: descriptions of millions of exhibitions, lectures, workshops, festivals, meetings of interest groups, and other event types. This will allow us to look for patterns in themes, interests, and "keywords" across geographies and over long time periods.

Figure I.1 shows growth in the numbers of cultural events over time using selected platforms and networks via which such events are promoted or organized. Together, our dataset, assembled from events announcements on six such platforms and networks, contains 4,380,946 events in 21,072 cities in 200 countries on six continents. These platforms and networks are Behance, E-Flux, Arts and Education Network, Meetup, TED Local Events, and TimePad. (In the case of Behance, an *event* is the registration of a new user account.)

As we see in the graphs, the numbers of cultural events on each platform and network have been growing over time. For 2006, our dataset has 11,642 events (adding all sources together); for 2009, it has already 102,211 events; and for 2018, it has 781,697 events. Of course, there is one single global platform that lists all cultural events, and we have to be careful in generalizing the results obtained from our particular sources. This project exemplifies the challenges of using digital phenomena to learn about other phenomena. Does the growth we see in all six data sources over time represent the real growth of cultural events in many countries? Or is it only a sign of diffusion of digital culture itself, showing that more organizations in more countries were gradually

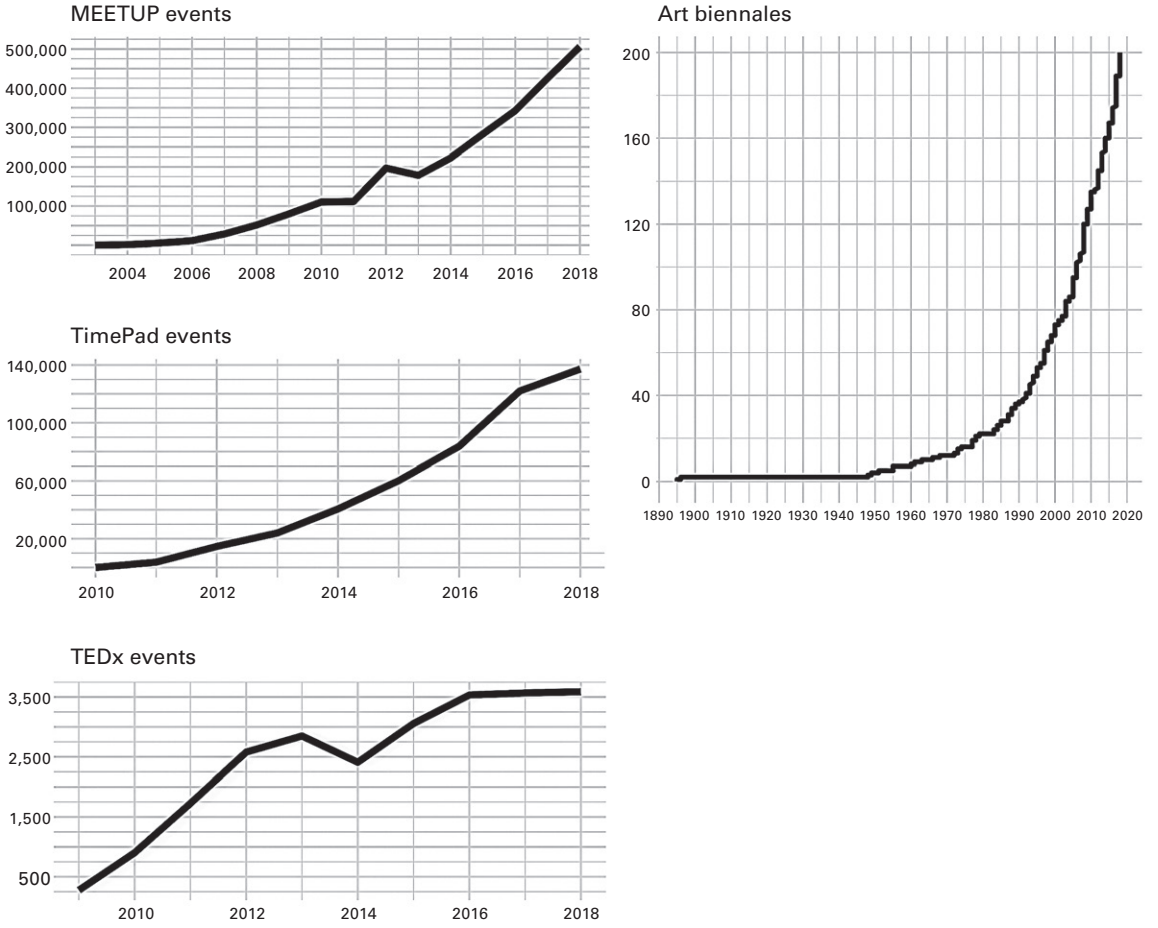


Figure I.1
Growth of numbers of cultural events over time as represented by particular data sources from the *Elsewhere* project.

using the platforms and networks we chose to look at to promote their events? Or do we simply see the winner-takes-all effect, in which certain information platforms become dominant and everybody starts using them? It is likely that all these effects are presented in our data. Some of the growth over time we see is due to the increasing popularity of the platforms themselves. Other growth reflects a real increase in the numbers of cultural places, actors, and events.

Elsewhere uses dates, locations, categories, and text content of the events announcements. In the second example I want to describe, we are looking at actual media artifacts

using data science and visualization. The goal is to be able to study various contemporary cultural fields by looking at many more works than media studies examine normally. The project in question is One Million Manga Pages that Jeremy Douglass, Wiliam Huber, and I started to work on in 2009. Figure I.2 shows one of the visualizations we created while exploring a dataset of 1,074,790 unique pages from 883 manga series. The pages come from the most popular fan manga site at that time—OneManga (onemanga.com). The site contained most pages for these series that fans scanned and translated into multiple languages.

The longest running manga series available on OneManga has been published continuously since 1976. The most popular series on this site were *Naruto* (8,835 pages for the 1999–2009 period) and *One Piece* (10,562 pages for the 1997–2009 period). Along with such long manga series, our dataset also contains shorter series that appeared in the 2000s and only ran for one to three years.

We used our own image analysis software to measure each page, converting its selected visual characteristics into numerical features. The visualization maps the pages onto x- and y-axes according to two of these features. The x-axis represents the standard deviation of the pixels' grayscale values, measured per page. The y-axis represents the entropy of all the pixels' grayscale values, also measured per page. What do these measurements mean in practice? The pages in the bottom part of the visualization are the most graphic and have the least amount of detail. The pages in the upper right have lots of detail and texture. The pages with the highest contrast are on the right, while pages with the least contrast are on the left.

In between these four extremes, we find every possible stylistic variation. This suggests to me that our basic concept of "style" may not be appropriate when we consider large cultural datasets. The concept assumes that we can partition a set of cultural artifacts into a small number of discrete categories. In the case of our One Million Manga Pages dataset, we find practically infinite graphical variations. If we try to divide this space into discrete stylistic categories, any such attempt will be arbitrary.

Visualization also shows which graphical choices are more commonly used by manga artists (the central part of the "cloud" of pages) and which appear much more rarely (bottom and left parts). We can ask why manga evolved in particular ways visually, with some choices used very frequently, others less frequently, and others almost never. And if we want to understand the visual originality of a new manga series (as represented by selected visual features we can measure), we can add its pages to such a visualization and—if we like—even quantify this originality. Later in the book, I present another analysis of this dataset, looking at connections between visual styles, gender of manga audiences, and manga genres (see figures 7.1 and 7.2 and plate 9).

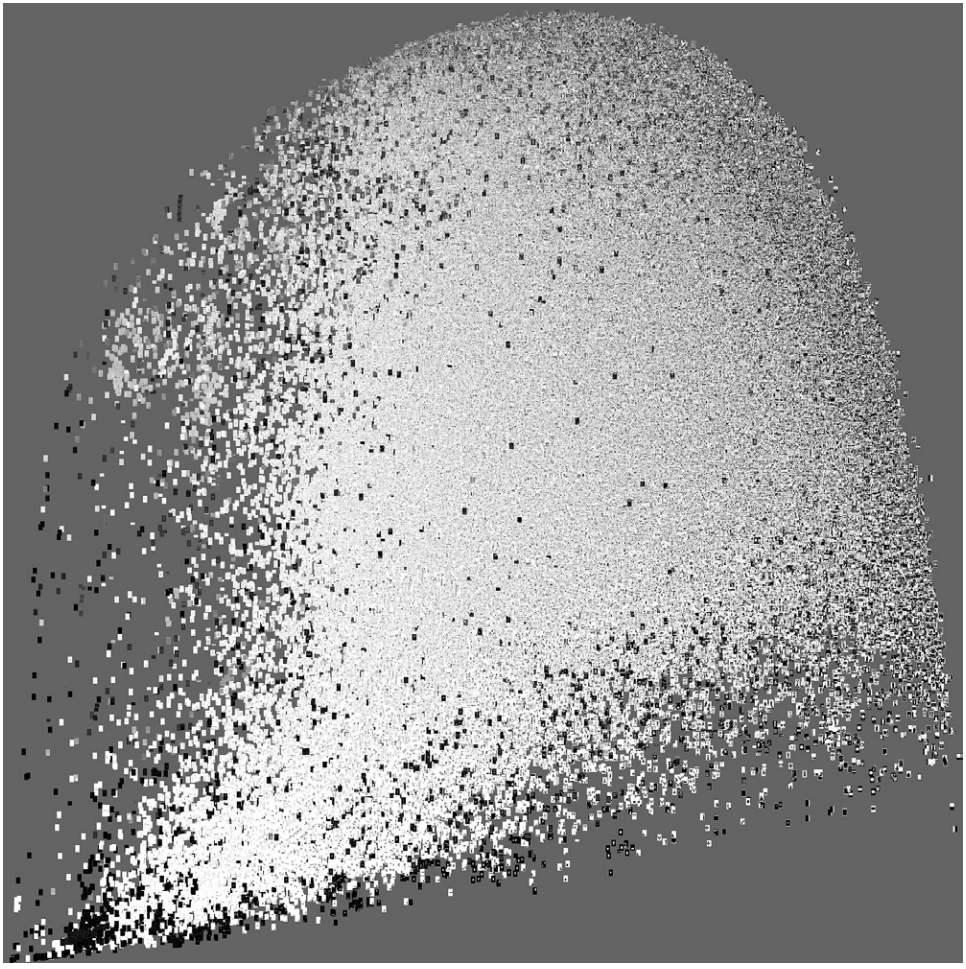


Figure I.2
Visualization of 1,074,790 manga pages sorted by two visual characteristics algorithmically computed on every page: standard deviation of grayscale values (x-axis) and entropy of grayscale values (y-axis).

Cultural Analytics: Five Ideas

I first started to think about analyzing and visualizing patterns in contemporary digital culture on a large scale in the fall of 2005. At that time, were there already examples of computational analysis of collections of websites and blogs carried out by researchers in computer science. However, computer science is a huge field with dozens of subfields. This research was appearing in journals and conferences of various subfields and did not have its own unique name. In the humanities, the term *digital humanities*, first introduced in 2003, was becoming known. But here as well, computational analysis of cultural datasets was only one part of this emerging larger field, which also included digitization and publication of historical collections, using digital tools in teaching, and other activities. The term "digital humanities" also had another limitation for me: digital humanities scholars were working almost exclusively with datasets of historical literary texts as opposed to other types of media or contemporary digital culture—so this term was both too broad and too narrow. Finally, I already knew of a number of very impressive artistic and design projects that visualized large cultural datasets, but this work was done outside the academic departments and also did not have a single term describing it.

In my mind, the existing work on analyzing and visualizing large-scale cultural data I saw happening across a number of fields was creating a new research paradigm—but it did not have its own name. I felt the need for a term that can refer to computational analysis of patterns and trends in *contemporary digital culture* (as opposed to only historical culture) and can cover analysis of all kinds of media (as opposed to only texts). When we were establishing our own lab to do this research in spring 2007,[5] I came up with the term *cultural analytics*.

Our lab pursued two goals. The first was practical: using methods from computer science, data visualization, and media art to explore and analyze different kinds of contemporary media and user interactions with them. The second was theoretical: we asked how the use of such methods and large datasets of cultural media challenges our existing modern ideas about culture and methods to study it. What exactly are the new possibilities they offer, and can they work with any kind of media? What are the limitations of computational methods and large-scale analysis?

In practice, our lab focused on analyzing and visualizing patterns in visual datasets, such as millions of photos shared on Instagram and Twitter, a million pages from manga publications, dozens of feature films, and thousands of magazine covers. This focus led many people who were following our work to associate cultural analytics with the use of computation and visualization for exploration of large visual collections. In

the last few years, the term *cultural analytics* (or *culture analytics*) started to be used more broadly by many other academics, as exemplified by two symposiums;[6] a four-month-long research program at UCLA in 2016 that brought together 120 leading researchers from universities and industry labs;[7] an academic peer-review journal, *Journal of Cultural Analytics*, established in 2016;[8] the use of the term in calls for conferences and academic job listings; and a number of academic programs and particular undergraduate and graduate courses.

In September 2019, I looked at these programs and courses: the Cultural Analytics BA course at King's College, London (UK); the cultural analytics concentration in the master of science in informatics and analytics at the University of North Carolina (US); the cultural analytics graduate certificate at Temple University libraries (US); the Center for Cultural Analytics at University College Dublin (Ireland); the Cultural Analytics course at Dartmouth College (US); the MA in data, culture and visualization at ITMO University (Russia); Cultural Analytics: The Computational Study of Culture, a course at McGill University (Canada); the CulturePlex Lab at Western University (Canada), which is doing research on "culture analytics and digital innovation"; the Distant Reading and Cultural Analytics course at the University of California, Los Angeles (US) and Cultural Data Analytics lab at Tallinn University (Estonia). The descriptions of all these programs and courses are using the term "cultural analytics" in a variety of ways related to the institutional context in which they are offered (e.g., a literature department, an information science department, etc.).

Looking at the papers published during its first three years by *Journal of Cultural Analytics*, we similarly see a range of subjects—although analysis of literary texts does dominate over other types of media. These subjects include nineteenth-century and contemporary English-language novels, nineteenth-century illustrated newspapers, late imperial Chinese literature, folklore classifications, menus of restaurants in the United States, and US television series.[9]

The term "cultural analytics" may mean different things to different people, and it can be used in different contexts to do different work. This is fine, and I am not interested in controlling its usage. This book does not try to summarize all cultural analytics research or to cover everything that is relevant to such research (e.g., inferential statistics, experiment design, supervised and unsupervised machine learning, text analysis, geospatial analysis, music data analysis, network analysis, agent-based simulation, and other topics that I do not discuss). Rather than trying to make an encyclopedia covering every relevant topic and technique, I decided instead to focus on a smaller number of topics and explore them in more depth.

The choice of these topics reflects the original motivations that led me to research how computers can help us see contemporary culture; my experience since 2008 of working on over forty practical projects in our lab; teaching practical cultural analytics every year in all my university classes since 2006 to undergraduate and graduate students; conducting many workshops in different countries; and learning from collaborations with other academic researchers and designers. You can see all the topics I included by consulting the table of contents. The topic organization follows a logical progression. Part I gives examples of computational cultural analysis and discusses the shift from "new media" of the 1990s to "more media" of 2000s, which motivated me to start thinking about cultural analytics. In part II, I discuss types of cultural data and techniques for representing cultural processes as computational data. Part III introduces concepts for exploring cultural datasets using data visualization and then focuses on recently developed methods for exploration of image and video collections.

Among all the topics and ideas presented in this book, there are five that I am most passionate about. Taken together, they describe a version of cultural analytics explored in this book—but other versions, visions, uses, and definitions of the term are certainly welcomed. Here are these five ideas:

(1) My original motivation for turning to computational methods and big data came from realization that the scale of culture in the twenty-first century makes it impossible to see it with existing methods. Therefore, in this book *cultural analytics refers to the use of computational and design methods—including data visualization, media and interaction design, statistics, and machine learning—for exploration and analysis of contemporary culture at scale*. One goal of these explorations is to *enable us to see what hundreds of millions of people around the world today create, imagine, and value*. This includes cultural activities of both normal people and professionals and students in creative industry: think of hundreds of design weeks and fashion weeks happening every year, thousands of film festivals, tens of thousands of educational programs, hundreds of thousands of cultural projects and exhibitions that announce themselves on the web and in social media, and so on. Thus, the key practical goal of cultural analytics for me is to work toward a more inclusive and democratic understanding of the cultural present and also of cultural histories. This means making fully visible the "long tail" of cultural life—and placing on our culture maps cities, countries, groups, individual creators, and artifacts that have been left out from both contemporary and historical cultural narratives.

The second goal is to come up with *new theoretical concepts* appropriate for the scale, speed, diversity, and connectedness of contemporary global digital culture. How is this

different from twentieth-century culture theory? Our new concepts should be not only theoretical but also *qualitative*; that is, they should allow us to measure dimensions of digital culture and compare styles, taste, imagination, and cultural behaviors in many geographic places, networks, and creative fields. (This means that we may also want to formalize and quantify existing concepts—such as, for example, *style*.) But we also need to be thinking about the limits of such quantification and be sensitive to dimensions and aspects of culture that existing measurements do not capture.

(2) The use of numerical representation and data analysis and visualization methods offers a new language for describing cultural artifacts, experiences, and dynamics. As I argue in chapter 7, the human languages that developed rather recently in human evolution are not good at capturing analog properties of human sensorial and cultural experiences. These limitations become particularly worrying if we want to compare thousands, millions, or billions of artifacts—that is, to study contemporary culture at its new scale. When we instead use numbers and visualization, we can better capture small differences between lots of artifacts and also between groups of artifacts. (Here data science, which has many methods to characterize relations between any number of objects—cluster analysis, dimension reduction, network analysis, and so on—becomes particularly relevant.)

Numbers and visualization also give us a language to represent *gradual and continuous temporal changes*. We can now describe the characteristics of cultural processes that are hard to capture linguistically—for example, gradual historical changes in visual culture over long periods or temporal changes in visual form in the career of an artist.

Given my commitment to analysis of visual culture, I contend that having a better language to describe its *analog dimensions* is invaluable. Digital computers that operate on numerical representations can better capture these dimensions that natural languages cannot describe adequately, such as motion or rhythm.

(3) While many ideas and discussions in the book are relevant for working with all kinds of data, I pay particular attention to *visual media*. I want to demonstrate through many examples how we can use computational and visualization methods to explore visual collections, asking interesting cultural questions. The larger portion of the computational work in the humanities so far has focused on literary texts, historical text records, and spatial data. In contrast, other types of media, such as still and moving images and interactive media, have received relatively little attention. This situation is gradually improving, but as I am writing this, analysis of visual media is still a small part of digital humanities.[10] You can see this yourself by browsing programs of annual conferences organized by the Alliance of Digital Humanities Organizations or by looking at the field journals. The field limitations are well summarized by the title of an

article published in 2017 in the *Digital Scholarship in the Humanities* journal: "Digital Humanities Is Text Heavy, Visualization Light, and Simulation Poor."[11]

This is surprising because computer scientists had already started to develop methods for analysis of images at the end of 1950s. Today they are implemented in numerous digital services and devices, including web image search engines, standalone photo cameras and cameras in mobile phones, image editing software such as Photoshop, image sharing networks, and so on. In computer science fields called computer vision and multimedia computing, researchers for many years have been publishing new algorithms for automatic detection of image content, artistic styles, photographic techniques, genres of TV and video, and applying them to progressively larger datasets.[12] In our lab, we have been using some of these methods to analyze many types of both historical and contemporary visual media—for example, twenty thousand photographs from the collection in the Museum of Modern Art (MoMA) in New York, films by Dziga Vertov from the Austrian Film Museum, sixteen million images shared on Instagram in seventeen global cities, 270 million images shared on Twitter globally, one million manga pages, and one million artworks from popular art network DeviantArt. In this book, I will refer to the details of some of these projects and to papers and work of other researchers (typically in computer science) who analyze visual content using algorithms.

(4) Can we explore and study collections of cultural media and records of cultural behaviors *without systems of categories* that languages impose on reality? Can we *avoid the quantification, measurements, and summarization* that comes with the use of statistics? Can we study big cultural data *without using numbers?*

The seemingly obvious answer to these questions is no. But the answer given by cultural analytics is yes (at least for some types of media, such as images and video). Any numerical measurements, linguistic categories, sets of tags, networks, or other *forms of representation*—regardless of their descriptive power and capacity to make visible similarities and other forms of relations between objects—are also *forms of omission*. For example, computer vision techniques today are able to detect thousands of object types in photographs[13]—but any art student or professional photographer knows that a photo is not simply a collection of objects or human figures in it. Rather than replacing human cultural observers with algorithms, the intention of cultural analytics is to augment our human abilities by *providing new interfaces and techniques for observing massive cultural datasets and flows*. (This is consistent with the vision of computing developed by Douglas C. Engelbart in his famous report from 1962, "Augmenting Human Intellect."[14]) Humans can notice more meaningful dimensions, recognize small details that really matter, and place information in more contexts than any algorithm at this point,

and unless the work on "artificial general intelligence" progresses sufficiently, this will remain true for the indefinite future. But can we extend these human capacities to deal with the scale of global cultural production and participation? In short: How can we see (for example) one billion images?

(5) Cultural analytics includes not only the application of currently available computational methods for data analysis to cultural datasets and flows, but also *critical examination of these data science methods and their assumptions*. Today our interactions with digital media, our access to information and each other in social networks, is mediated by software systems. They continuously analyze "big cultural data"—that is, the content of billions of media artifacts we share, our online interactions with these artifacts, and our other online and physical behaviors. In cultural analytics research, we often use similar methods for different purposes—for example, seeing patterns in cultural history, exploring the work of contemporary designers, and examining the content and styles of photos shared by billions of people online. What are the similarities and differences between the use of these methods in industry and in cultural research? Are there some assumptions and goals built into methods that are widely used in industry that we need to question if we adopt them? What historical developments led to the popularity of certain methods today? Addressing such questions is also the key part of the cultural analytics agenda.

Critical examination of data science and algorithms and data use in society in general also takes place in a number of research fields, including science and technology studies, digital humanities, digital culture studies, critical algorithm and data studies,[15] and software studies. Among many academic journals in social science and the humanities that publish articles about these topics, I can highlight *Big Data & Society*. I advise you to look at publications and conferences in these fields and read the papers that look particularly relevant to your interests.

Throughout this book, I will discuss what I personally see as the *most interesting and promising research directions in cultural analytics*. By *interesting*, I mean the analysis that allows us to think of contemporary culture in new ways and helps us to question concepts and methods for studying culture that we take for granted. Some of these directions can be illustrated by existing work, while others have not yet been pursued. So, if you are going to get into cultural analytics, I hope that you will find interesting ideas in this book for new things to try.

Cultural analytics is only one among the paradigms that emerged in the second part of the 2000s to take advantage of the availability of large cultural and social data. These include digital humanities, computational social science, social computing, digital anthropology, digital history, the science of cities, urban informatics, and

culturomics.[16] At the same time, big cultural datasets started to be analyzed and used in many areas of computer science, such as machine learning, AI, computer vision, natural language processing, and computer multimedia, as well as in network science and communication studies. In the early 2010s, the "quantitative turn" began in art history, with *International Journal for Digital Art History* starting its publication in 2015. In film studies, the first monograph that uses quantitative methods and data visualization to analyze works of a single film director appeared in 2018.[17]

In the same decade, a number of new research agendas emerged to address the growing use of algorithms, data, and artificial intelligence (AI) systems from the perspectives of social sciences and humanities. These research areas include machine behavior, and the already mentioned algorithm studies and critical data studies. (The 2017 open access anthology *The Datafied Society: Studying Culture through Data* collects a number of articles addressing methodological and ethical questions related to the use of data and algorithms in academic research.[18])

The research we undertook in our lab developed as a dialog with the work by researchers in such fields and all relevant projects I saw, and that research would not have been possible without everything I was learning from them. The goal of this book is not to draw a boundary around the field or to claim that we are the only ones who did research on particular topics. Cultural analytics for me is the instrument to question all categorical boundaries, so certainly it also would not be wise to draw a border around all cultural analytics research taking place across many academic and professional fields.

Cultural Analytics: Twelve Research Challenges

My thinking about cultural analytics and work in our lab has been guided by a list of theoretical and practical challenges that I defined around the time the lab was created. Only some of these questions will be discussed in detail in this book: they are the ones that I ended up spending more energy and time on. But I want to give you the full list because these questions can be also useful to you in your work. I made a list of first eight challenges in 2005–2007 (questions 1–8 below); after years of research, the new challenges become apparent (questions 9–12):

1. How can work with "big cultural data" help us questions our stereotypes, assumptions, concepts, and existing knowledge about cultures?
2. What are the fundamental new ways of understanding and studying visual and media cultures enabled by computational methods and large datasets?

3 How can we explore massive visual collections that may contain billions of images and video?

4. How can we combine computational media analysis with qualitative media studies methods and theories?

5. How do we use computational approaches to analyze interactive media and experiences (e.g., playing a video game, interacting with the Instagram app, experiencing an interactive installation), as opposed to only static media artifacts?

6. What theoretical concepts and models do we need to deal with the mega scale and velocity of user-generated content and user interactions online?

7. How can we analyze and visualize the diversity of contemporary global digital cultures, taking into account the activities of billions of creators and trillions of objects they are creating?

8. What will the "science of culture" driven by massive cultural datasets and computation look like, and what will its limitations be?

9. Can we define general quantitative measures of cultural variability, diversity, temporal change, difference, influence, and uniqueness that will be meaningful for many types of media and different periods and cultures—and in particular our own period?

10. Given that statistical and data science methods are based on data reduction and summarization, how can we also analyze computationally small differences and unique details of individual artifacts and experiences?

11. Can we describe cultures—both objectively and as we perceive them—as statistical distributions and combinations of elements, themes, and strategies? Or is culture about gestalts not reducible to their parts? (If the answer to the second question is yes, this may make cultural analytics impossible.)

12. Let's assume that we can detect a small number of themes, topics, styles, and cultural techniques in billions of cultural artifacts, experiences, and interactions. Let's further assume that we track these culture DNAs across the world, adding new ones when we detect these. What level of reduction is appropriate when we extract small number of topics from billions of cultural "objects," and what is lost at each level? For example, how much information is lost if we extract 10,000 topics, and then aggregate them into 1,000 topics, and then aggregate this data again to have only 100 topics? Is reduction a wrong approach? If contemporary authors want to create unique artifacts and experiences that cannot be duplicated, will we inevitably miss true uniqueness in our quest to track larger trends?

I see this last challenge as the most important. Should we aggregate big cultural data and reduce it to a smaller number of structures—only the most frequently occurring ideas, themes, styles, patterns, and behaviors? This is the paradigm we inherited from the history of statistics, and many quantitative studies of culture in computer science follow it. In this paradigm, we focus on what is common between a number of objects, and we do not include what occurs infrequently. Or should we develop the opposite paradigm—refuse aggregation and reduction, and instead focus on diversity, variability, and differences among numerous artifacts, behaviors, and individuals? In this paradigm, we include all the data, and we pay special attention to the infrequent and rare phenomena.

Together these twelve questions should further clarify in what ways my own cultural analytics motivations and interests that guide this book differ from those of other researchers working in this area. My main motivations in using "big cultural data" is to question what we think we know about cultures (1), as opposed to only making technical progress in already well-established paradigms; to address the challenges of thinking about new forms of digital culture (5); and to understand how its scale, speed, and diversity can be dealt with, both on the level of theoretical concepts and as characteristics we can measure (6, 7, 9). Other challenges arise when we confront the "data cognition" of contemporary society (i.e., data science assumptions and methods) with the types of subjects that these methods were not designed for—cultural life, experiences, and artifacts (2, 3, 4). My original optimism about the seemingly endless possibilities of cultural analytics gradually gave way to a more realistic view when I realized the limitations of statistical and computational approaches—and more generally, the limitations of thinking of cultures as combinations of elements that we can track (10–12).

Thus, this book aims to both convince you that these approaches are very useful and sometimes the only way to start dealing with the scale of culture today—and at the same time clearly explain what at least at present computers cannot see. And this is not because they are born, so to speak, with blindness to the aesthetic domain and cannot be cured. They are not. The problem lies with their teachers—that is, with us. If we ourselves do not understand why, out of hundreds of photographs of the same subject, the magazine editor choses a particular one, even though when a computer measures their many characteristics, they turn out to be practically identical, then how can we hope to teach such aesthetic judgments to a machine? Certainly, we can feed millions of such examples to a neural network and it will learn to predict with a certain precision the "best" photographs—but such probabilistic vision is not the same as understanding.

What Cultural Analytics Is Not

Having defined key ideas and twelve research challenges for cultural analytics as I see it, I also need to clearly state what cultural analytics is not. The rapid growth of many social networks and the availability of their data via application programming interfaces (APIs) between approximately 2007 and 2015 stimulated lots of research activity across a number of fields. In our lab, we have also taken advantage of having access to data from this new cultural universe. Between 2012 and 2015, we and our collaborators worked with datasets of publicly shared images or information about them from Twitter, Instagram, and VK, created a number of visualizations and published our research in a few papers. (Networks' APIs provided usernames of visual posts, but we never used this information in any publications or exhibitions.) Big social media data was also used during the years of open APIs by tens of thousands of scientists. In fact, for a number of years now, downloading data from Twitter and analyzing it is often used as an exercise in many computer science and data science classes.

The global growth of social networks and media sharing sites after 2007 certainly confirmed my ideas about the need for cultural analytics formulated earlier. However, cultural analytics is not "married" to this type of media and data. Social media networks emerged recently in the long history of media, and they may not exist in the same form in the future. In fact, 2005 I was not even thinking about using social media data, because social networks were not yet very popular, and there was no mechanism to download their data. Instead, I was thinking about collecting information from numerous websites belonging to individual designers, cultural centers, publications, art schools, museums, and analyzing the content of culture-related blogs that were already very popular. I imagined accessing content from as many sites as possible around the world and visualizing the changing patterns in real time (see plates 1 and 2).

The free access to content from leading social networks that became available in a few years was important in making visible opportunities and challenges in observing and analyzing cultural data on the scale—but the cultural analytics program does not depend on access to social media, or any particular source. If at some point in the future, websites and social media cease to exist in their present forms, some other forms of collective media publishing and sharing are likely to replace them.

It is almost certain that particular technologies will be changing, but the fundamental new conditions that were already established during the 2000s will remain: the new scale of culture and the growing presence of culture in contemporary societies. This includes both "more culture" (more cultural producers, more objects and events, more areas of society where aesthetics becomes very important, etc.) and "more

information" about culture (websites, posts, publications, datasets). The goal of cultural analytics is to address the challenge of seeing and thinking about culture at this new scale.

Cultural Analytics, Media Theory, and Software Studies

Cultural Analytics is a book of *media theory*. I argue that to systematically study global media cultures today, media theory needs to turn to data science. In fact, computational methods are necessary in order not only to analyze and theorize these global cultures, but even simply to *see* them in the first place.

If we rely only on our intuitions based on content we happened to see online, algorithmic recommendations, or intuitions about what is important, it is too easy for us to stay in our cognitive and historical *filter bubbles*, projecting our biases onto the world.[19] The academic fields may have their own filter bubbles, in which established research paradigms can make invisible to academics emerging types of cultural activities they should be attending to. For example, while interaction design has become central to our daily cultural experiences in the forms of apps, websites, and connected devices, it is not yet sufficiently discussed in media studies or humanities.

Of course, computational methods and large datasets do not automatically guarantee more objectivity and inclusion. However, they can help us to *confront our assumptions, biases, and stereotypes*. They allow us to notice what we overwise may not be able to see—the content and its creators that do not make it to the top results of search or recommendation engines, do not appear in top ten or top 100 lists, and thus remain invisible.

How do you go about researching the long tail of digital culture?[20] Many computer researchers have been using random sampling to select millions of text posts, images, and videos shared on Twitter, YouTube, Instagram and other networks and then analyze them. Such large samples capture a snapshot of global activity on such networks, but they may not represent well the substantial differences in content shared in different geographic areas or by different demographic groups. In several projects in our lab, we have followed a different strategy: selecting a small geographic area and then collecting all content shared in this area. For example, for the *On Broadway* and *Inequaligram* projects, we analyzed data on *all* geolocated Instagram images that were shared across all of Manhattan for five months.[21] We did not filter anything out. We did not start with any hashtags that represent a particular topic. We did not only look at images that have more likes. We did not separate "art" from "nonart" or "original" images from copies or "influencers" from regular users. Instead, all posts with location information

indicating that they were shared in Manhattan during this period—all 7,442,454 geolocated images, hashtags, and descriptions from 1,890,585 Instagram users—were collected in our lab and considered to be equally valuable for further analysis.

Before we can "theorize" contemporary media, we need to see it, and this is not possible anymore without computers because of its new scale. Thus, rather than only being the *subject of analysis*, as in my book *The Language of New Media*, the computation becomes the *practical tool* for studying media in this new book.

Since 1984, I have been working practically with digital media in different capacities: as computer animator, motion graphics designer, software developer, media artist, and professor of digital art. I have taught hands-on classes in digital art since 1992, data visualization since 2006, and data science since 2013. I also wrote software tools that I and others in our lab used to visualize large visual datasets. This practical experience designing, programming, and teaching digital media, visualization, and data analysis is reflected in this book. Exploring, manipulating, and visualizing datasets for me is a direct continuation of creating art and design with code that I have been doing since 1984.

In a parallel way, I see my cultural analytics research as a direct extension of my new media theory work of the 1990s and 2000s. The difference is the scale of digital culture then and now. In the middle of the 1990s, the number of artists working with algorithms was so small that we could all meet in a single conference. Two such key annual conferences were the *Ars Electronica* festival that began in 1979 and the annual International Symposium on Electronic Art (ISEA) that first took place in 1988. In the 1994 ISEA in Helsinki, approximately 150 participants gathered, and this motivated the international takeoff of new media art.

Today, hundreds of thousands of people identify themselves as digital artists, creative technologists, or creative coders, and a few billion people with camera phones have become digital photographers. A search for the phrase "How do I edit my Instagram" on YouTube (January 15, 2016) returned 150,000 videos with how-to advice by Instagram users. And the same search on October 11, 2017, returned 228,000 videos. The top videos in this genre have millions of views each.[22] This new scale of media production and participation calls for new research methods, concepts, and tools, and this motivates this book. In the world in which digital media is created by a few billion rather than by a few thousand, like it was twenty-five years ago, we need to reinvent what it means to study culture.

Cultural Analytics also incorporates perspectives of the software studies field, which asks how software shapes the world today. In *Software Takes Command*, I wrote: "If we want to understand contemporary techniques of *control, communication, representation,*

simulation, analysis, decision-making, memory, vision, writing, and interaction, . . . we [must] consider this software layer."[23] *Software Takes Command* presented theoretical and historical analysis of popular tools for media creation such as Photoshop and After Effects; *Cultural Analytics* looks at some of the core concepts and assumptions of the data-centric world view. It asks *how our society thinks and acts using data and algorithms* and *how the algorithmic analysis of user content and interactions by the industry shapes culture today*.

I believe that all members of the creative class, media researchers, humanists, and social scientists need to have *data science literacy*: knowledge of core principles of data analysis, machine learning, and predictive analytics methods and applications. Why? Because the software and code based on these principles is employed throughout our society, including digital culture industry, businesses, nonprofits, and government. And if data science has not yet started to be used in a particular academic field, it is only a matter of time.

In summary, I see my move from *media* to *data analysis* as a logical progression. *The Language of New Media* (written in 1999) described forms of digital culture that emerged in the 1990s. *Software Takes Command* (2007) covered the history of software programs for media creation and editing and the new visual languages they enabled as they were adopted around the turn of the twenty-first century. *Cultural Analytics* now investigates the new post-2005 stage in which billions of people create digital media and share it online. Also, at this stage, cultural software is given a new role. We delegate more agency to it. It is no longer only a tool, a medium, or an assistant. Instead, it now engages in *cultural behaviors* (e.g., deciding what new social media posts to show, improving quality of our photos, writing news articles, etc.). And while for now we still write posts, take photographs, and perform other cultural actions ourselves, gradually such actions may become fully automated in the future. (For example, in 2018 Google added an autocomplete feature to Gmail that automatically completes your email response as you start writing—you only need to press the Tab key to accept the suggestion.) This is why all cultural and media scholars and students need to acquire a good understanding of the data science and AI fields. Chapter 3 of this book presents an analysis of these new cultural roles of computers; my 2018 book *AI Aesthetics* discusses how the growing use of algorithmic systems may affect cultural diversity.[24]

Using This Book in Classes

The content and the structure of this book reflect my experience teaching hands-on cultural analytics classes to diverse groups of students. They included undergraduates

in digital art, media art, computer science, and art history, and graduate students in computer science and humanities and social science fields including art history, literature, musicology, communication, economics, sociology, anthropology, psychology, and digital humanities.[25]

The chapters of this book present a sequence of topics that can be covered in a semester or a quarter-long class (10–14 weeks). The goals are to familiarize students with examples of work with cultural datasets in humanities, computer science, design, and other fields, and explain why we need computational methods for analyzing contemporary culture (part I); to learn the conceptual operations, choices, and constraints involved in creating "cultural data" (part II); and to understand how to explore media datasets using data visualization (part III).

Why did I choose to include these topics and not others? More generally, what is the difference between cultural analytics and data analytics? Why don't I have chapters in the book on data analysis?

Consider the workflow for doing a research, design, or artistic project with data: (1) think of how some subjects can be analyzed or represented quantitatively; (2) research what suitable data is available or how to generate it; (3) assemble the data; (4) use visual methods to explore this data; (5) analyze the data using methods from statistics and data science (i.e., descriptive and inferential statistics, unsupervised and supervised machine learning, time-series analysis, network science, etc.); and optionally (6)—create interactive visualization tools for others to explore this data, or make other design and media outputs.

In my view, for cultural analytics, step 5 is not really different from what you would do with any other data. Because there are lots of good textbooks, online courses, and tutorials available to learn these methods, in addition to courses your university may offer, you should use these resources to learn them. So instead of covering the material that already exists elsewhere, this book deals with other parts of the workflow—1, 2, 3, 4 and 6. In other words, what is unique in cultural analytics is not what you do with data—but how do you get from this elusive thing we call *culture* to its data representation in the first place. That is, *how do you turn cultural experiences, events, actions, and media into data*? What is gained and what is lost in this translation? And once you get it into data form, how can you explore it on multiple scales, seeing both unique and infrequent as well as common and regular patterns?

However, while I am not going to teach you statistics and data analysis in this book, I will talk about the assumptions behind some of their methods, what they can enable us to see, and what they can't, at least at present. Therefore, ideally, you should read

this book in parallel to learning data science techniques, or after you have learned some of them.

You can read the book chapters in sequence, or just move to any chapter that interests you. I have tried to make each chapter relatively self-sufficient. Throughout the book, you will find summaries of the already covered material and presentations of new material organized in numbered lists. I hope that this organization is helpful if either the whole book or separate chapters are used in classes.

Notes

Introduction

1. Dell H. Hymes, "Introduction," in *The Use of Computers in Anthropology*, ed. Dell H. Hymes (The Hague: Mouton, 1965), 29–30. Emphasis in original.

2. United States Securities and Exchange Commission, Form S-1 Registration Statement for Pinterest, Inc., March 22, 2019, https://www.sec.gov/Archives/edgar/data/1506293/000119312519083544/d674330ds1.htm.

3. "About," *Meetup*, accessed August 12, 2019, https://www.meetup.com/about/.

4. This feature was first introduced in Apple's iPhone 7 in 2016. However, even earlier some mobile apps and services, such as Google Photos, were already offering object detection in photos on mobile phones.

5. Cultural Analytics Lab, "About," accessed February 28, 2020, http://lab.culturalanalytics.info/p/about.html.

6. *Cultural Analytics: Computational Approaches to the Study of Culture*, symposium at the University of Chicago, Chicago, May 22–23, 2019, http://neubauercollegium.uchicago.edu/events/uc/cultural_analytics/; *Cultural Analytics 2017*, symposium at Notre Dame University Notre Dame, May 26–27, 2017, https://sites.google.com/nd.edu/ca2017.

7. "Culture Analytics program," Institute for Pure and Applied Mathematics (IPAM), UCLA, March 7–June 10, 2016, http://www.ipam.ucla.edu/programs/long-programs/culture-analytics/.

8. "About," *Journal of Cultural Analytics*, accessed July 23, 2019, http://culturalanalytics.org/about/.

9. "Articles," *Journal of Cultural Analytics*, accessed October 1, 2019, https://culturalanalytics.org/category/articles/.

10. Digital Humanities 2019 Conference, Utrecht, Netherlands, July 9–12, 2019, https://dh2019.adho.org.

11. Erik Malcolm Champion, "Digital Humanities Is Text Heavy, Visualization Light, and Simulation Poor," *Digital Scholarship in the Humanities* 32, s1 (2017): 25–32.

12. Miriam Redi, Frank Z. Liu, and Neil O'Hare, "Bridging the Aesthetic Gap: The Wild Beauty of Web Imagery," in *Proceedings of the 2017 ACM International Conference on Multimedia Retrieval*, (New York: ACM, 2017), 242–250.

13. Clarifai reported being able to detect eleven thousand types of objects reliably in 2017. See Clarifai, "Models," accessed October 13, 2017, https://www.clarifai.com/models.

14. Douglas Engelbart, "Augmenting Human Intellect: A Conceptual Framework," SRI Project no. 3578, October 1962, http://dougengelbart.org/content/view/138.

15. David Moats and Nick Seaver, "'You Social Scientists Love Mind Games': Experimenting in the 'Divide' between Data Science and Critical Algorithm Studies," *Big Data & Society* 6, no. 1, (2019), https://doi.org/10.1177/2053951719833404; Andrew Iliadis and Federica Russo, "Critical Data Studies: An Introduction," *Big Data & Society* 3, no. 2 (2016), https://doi.org/10.1177/2053951716674238.

16. Jean-Baptiste Michel et al., "Quantitative Analysis of Culture Using Millions of Digitized Books," *Science* 331, no. 6014 (2011): 176–182, https://doi.org/10.1126/science.1199644.

17. Adelheid Heftberger, *Digital Humanities and Film Studies: Visualising Dziga Vertov's Work* (Basel: Springer, 2018).

18. Karin van Es and Mirko Tobias Schäfer, eds., *The Datafied Society. Studying Culture through Data* (Amsterdam: Amsterdam University Press, 2017), https://oapen.org/search?identifier=624771.

19. Eli Pariser, *The Filter Bubble: What the Internet Is Hiding from You* (New York: Penguin Press, 2011).

20. Chris Anderson, "The Long Tail," *Wired*, October 1, 2004, https://www.wired.com/2004/10/tail/; Erik Brynjolfsson, Yu Jeffrey Hu, and Michael D. Smith, "The Longer Tail: The Changing Shape of Amazon's Sales Distribution Curve," September 22, 2010, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1679991.

21. "On Broadway," 2014; http://on-broadway.nyc/; "Inequaligram," 2016, http://inequaligram.net/.

22. "How I Edit My Instagram" (search results), *YouTube,* accessed January 15, 2016, https://www.youtube.com/results?search_query=%22how+i+edit+my+instagram%22.

23. Lev Manovich, *Software Takes Command*, rev. ed. (London: Bloomsbury Academic, 2013). An earlier version was released under a Creative Commons license in 2007, 15; italics in original.

24. Lev Manovich, *AI Aesthetics* (Moscow: Strelka Press, 2018).

25. Lev Manovich, "Teaching," accessed September 20, 2019, http://manovich.net/index.php/teaching.

1 From New Media to More Media

1. Gustav Theodor Fechner, *Elements of Psychophysics* (New York: Holt, Rinehart & Winston, 1966); L. L. Thurstone, "The Vectors of Mind," *Psychological Review* 41 (1934): 1–32; Jacob Levy Moreno, *Who Shall Survive? A New Approach to the Problem of Human Interrelations* (Washington, DC: Nervous and Mental Disease Publ. Co, 1934).

2. The later version of this text was published as Lev Manovich, "Cultural Analytics: Visualizing Cultural Patterns in the Era of 'More Media,'" *Domus* 923 (March 2009), http://manovich.net/index.php/projects/cultural-analytics-visualizing-cultural-patterns.

3. International Conference on Weblogs and Social Media (ICWSM) 2017, "Previous Conferences," accessed October 11, 2017, https://icwsm.org/2019/contact/previous-conferences/.

4. International Conference on Weblogs and Social Media (ICWSM) 2019, accessed October 11, 2017, https://icwsm.org/2007/.

5. "How Many Online Forums Are in Existence?," *Quora*, accessed October 11, 2017, https://www.quora.com/How-many-online-forums-are-in-existence.

6. "How Many People Use Quora?," *Quora*, accessed October 11, 2017, https://www.quora.com/How-many-people-use-Quora-7?redirected_qid=12824.

7. "Academia: About," accessed February 7, 2020, https://www.academia.edu/about.

8. WGSN, "Fashion," accessed October 11, 2017, https://www.wgsn.com/en/products/fashion/.

9. WGSN, "About WGSN," accessed October 11, 2017, https://www.wgsn.com/en/wgsn/#!/page/our-services.

10. Fern Seto, "How Does Trend Forecasting Really Work?," Highsnobiety, April 5, 2017, https://www.highsnobiety.com/2017/04/05/trend-forecasting-how-to/.

11. Uma Karmarkar and Hilke Plassmann, "Consumer Neuroscience: Past, Present, and Future," *Organizational Research Methods* 22, no. 1 (2019): 174–195.

12. Behance, "Year in Review—2015," accessed July 5, 2016, http://www.behance.net/yearinreview.

13. Jesse Alpert and Nissan Hajaj, "We Knew the Web Was Big . . . ," *Google Official Blog*, July 25, 2008, http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html.

14. Wikipedia, "YouTube," accessed August 1, 2008, http://en.wikipedia.org/wiki/YouTube.

15. Heather Champ, "3 Billion!," *Flickr Blog*, November 3, 2008, http://blog.flickr.net/en/2008/11/03/3-billion/.

16. John F. Gantz et al., The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth through 2011 (Framingham, MA: International Data Corporation, March 2008), https://www.atour.com/media/images/service/IDC-EMC-The-Diverse-and-Exploding-Digital-Universe-2008.pdf.

17. The number of design portfolios submitted by users to Coroflot.com grew from 90,657 on May 7, 2008, to 120,659 on December 24, 2008.

18. "A Brief History of Google Analytics, Part One," *Digital State*, May 1, 2014, http://digitalstatemarketing.com/articles/brief-history-google-analytics-part-one.

19. Lev Manovich, "Cultural Analytics Visualizations on Ultra High Resolution Displays," *Software Studies Initiative* (blog), December 24, 2008, http://lab.softwarestudies.com/2008/12/cultural-analytics-hiperspace-and.html.

20. "UC San Diego Unveils World's Highest-Resolution Scientific Display System," California Institute for Telecommunications and Information Technology, July 9, 2008, http://www.calit2.net/newsroom/release.php?id=1332.

21. Gapminder, "Gapminder Tools," accessed September 18, 2019, http://www.gapminder.org/world/.

22. Brandon Keim, "Map of Science Looks Like Milky Way," *Wired*, March 11, 2009, https://www.wired.com/2009/03/mapofscience/.

23. Sprout Social, accessed July 29, 2016, http://sproutsocial.com/.

24. Mention, accessed July 29, 2016, https://mention.com/en/.

25. Brandwatch, "Historical Data," accessed July 29, 2016, https://www.brandwatch.com/historical-data/.

26. DataSift, accessed October 14, 2017, http://datasift.com/.

27. *Operationalization* is the practice in natural and social sciences of defining concepts through measurement operations. For example, in psychology, emotion can be measured in a number of ways, such as by facial expression, body movements, choice of vocabulary, and tone of voice.

28. Lynn Gamwell, *Mathematics and Art: A Cultural History* (Princeton, NJ: Princeton University Press, 2015), 169–170.

29. Hymes, *The Use of Computers in Anthropology*, 29–30.

30. Martin Wattenberg, "The Shape of Song," 2001, http://www.bewitched.com/song.html; "History Flow," 2003, http://www.bewitched.com/historyflow.html.

31. Peter Eleey, "Mark Hansen and Ben Rubin," *Frieze*, May 6, 2003, https://frieze.com/article/mark-hansen-and-ben-rubin.

32. Jason Salavon, *The Top Grossing Film of All Time, 1 x 1*, 2000, digital C-print mounted to plexiglass, 47" × 72", http://www.salavon.com/work/TopGrossingFilmAllTime/.

33. George Legrady, Making Visible the Invisible, 2005–2014, six LCD screens on glass wall, 45" × 24", Seattle Central Library, Seattle, http://www.mat.ucsb.edu/g.legrady/glWeb/Projects/spl/spl.html.

34. Susan Hockey, "The History of Humanities Computing," in *A Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens, and John Unsworth (Oxford: Blackwell, 2004), 3–19.

2 The Science of Culture?

1. Friedrich von Hayek, "Prize Lecture" (Nobel Prize lecture, December 11, 1974), https://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/1974/hayek-lecture.html.

2. *The 22nd ACM Conference on Computer-Supported Cooperative Work and Social Computing*, Austin, Texas, November 9–13, 2019.

3. Joan Serrà et al., "Measuring the Evolution of Contemporary Western Popular Music," *Scientific Reports* 2, no. 521 (2012), https://doi.org/10.1038/srep00521; Maximilian Schich, Chaoming Song, Yong-Yeol Ahn, Alexander Mirsky, Mauro Martino, Albert-László Barabási, and Dirk Helbing, "A Network Framework of Cultural History," *Science* 345, no. 6196 (2014): 558–562, https://doi.org/10.1126/science.1240064.

4. *PLOS ONE*, accessed September 18, 2019, http://journals.plos.org/plosone/.

5. Miriam Redi, Neil O'Hare, Rossano Schifanella, Michele Trevisiol, and Alejandro Jaimes, "6 Seconds of Sound and Vision: Creativity in Micro-videos," in *CVPR '14 Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition* (Washington, DC: IEEE Computer Society, 2014), 4272–4279.

6. Yuheng Hu, Lydia Manikonda, and Subbarao Kambhampati, "What We Instagram: A First Analysis of Instagram Photo Content and User Types," in *Proceedings of Ninth International AAAI Conference on Web and Social Media* (Palo Alto, CA: AAAI Press, 2014), 595–598, https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8118/8087.

7. Saeideh Bakhshi, David A. Shamma, Lyndon Kennedy, and Eric Gilbert, "Why We Filter Our Photos and How It Impacts Engagement," in *Proceedings of the 8th International Conference on Weblogs and Social Media* (Palo Alto, CA: AAAI Press, 2015), http://comp.social.gatech.edu/papers/icwsm15.why.bakhshi.pdf.

8. Flávio Souza, Diego de Las Casas, Vinícius Flores, SunBum Youn, Meeyoung Cha, Daniele Quercia, and Virgílio Almeida, "Dawn of the Selfie Era: The Whos, Wheres, and Hows of Selfies on Instagram," paper presented at the ACM Conference on Online Social Networks 2015, Stanford University, Stanford, CA, October 19, 2015, https://arxiv.org/abs/1510.05700.

9. Kevin Matzen, Kavita Bala, and Noah Snavely, "StreetStyle: Exploring World-Wide Clothing Styles from Millions of Photos," June 6, 2017, https://arxiv.org/abs/1706.01869.

10. Babak Saleh et al., "Toward Automated Discovery of Artistic Influence," *Multimedia Tools and Applications* 75, no. 7 (2016): 3565–3591, https://doi.org/10.1007/s11042-014-2193-x.

11. Joan Serrà, Álvaro Corral, Marián Boguñá, Martín Haro, and Josep Ll. Arcos, "Measuring the Evolution of Contemporary Western Popular Music," *Scientific Reports* 2, no. 521 (2012), https://doi.org/10.1038/srep00521.

12. James E. Cutting, Kaitlin L. Brunick, Jordan E. DeLong, Catalina Iricinschi, and Ayse Candan, "Quicker, Faster, Darker: Changes in Hollywood Film over 75 Years," *i-Perception* 2, no. 6 (2011): 569–576, https://doi.org/10.1068/i0441aap.

13. Susan Hockey, "The History of Humanities Computing." in *Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens, and John Unsworth (Oxford: Blackwell, 2004), 3–19.

14. Ted Underwood, "A Genealogy of Distant Reading," *Digital Humanities Quarterly* 11, no. 2 (2017).

15. Rachel Sagner Buurma and Laura Heffernan, "Search and Replace: Josephine Miles and the Origins of Distant Reading," *Modernism/Modernity*, April 11, 1018, https://modernismmodernity.org/forums/posts/search-and-replace.

16. The Alliance of Digital Humanities Organizations (ADHO), "Conference," accessed September 18, 2019, http://adho.org/conference.

17. Ted Underwood, "Seven Ways Humanists Are Using Computers to Understand Text," *The Stone and the Shell* (blog), June 4, 2015, https://tedunderwood.com/2015/06/04/seven-ways-humanists-are-using-computers-to-understand-text/.

18. Ted Underwood, Michael L. Black, Loretta Auvil, and Boris Capitanu, "Mapping Mutable Genres in Structurally Complex Volumes," in *2013 IEEE Conference on Big Data* (Santa Clara, California), 95–103, http://arxiv.org/abs/1309.3323.

19. Underwood et al., "Mapping Mutable Genres."

20. Peter Klimek, Robert Kreuzbauer and Stefan Thurner, "Fashion and Art Cycles Are Driven by Counter-Dominance Signals of Elite Competition: Quantitative Evidence from Music Styles," *Journal of the Royal Society Interface* 16, no. 151, February 6, 2019, https://doi.org/10.1098/rsif.2018.0731.

21. Natasha Singer, "In a Scoreboard of Words, a Cultural Guide," *New York Times*, December 7, 2013, https://www.nytimes.com/2013/12/08/technology/in-a-scoreboard-of-words-a-cultural-guide.html.

22. New York Public Library, "Visualize the Public Domain," accessed October 21, 2019, http://publicdomain.nypl.org/pd-visualization.

23. New York Public Library, "Photographers' Identities Catalog," accessed October 21, 2019, http://pic.nypl.org.

24. Spotify, "Celebrating a Decade of Discovery on Spotify," October 18, 2018, https://newsroom.spotify.com/2018-10-10/celebrating-a-decade-of-discovery-on-spotify/.

25. Rossano Schifanella, Miriam Redi, and Luca Maria Aiello, "An Image Is Worth More than a Thousand Favorites: Surfacing the Hidden Beauty of Flickr Pictures," in *Proceedings of the 8th International Conference on Weblogs and Social Media* (Palo Alto, CA: AAAI Press, 2015), http://arxiv.org/pdf/1505.03358.pdf.

26. Katharina Reinecke and Krzysztof Z. Gajos, "Quantifying Visual Preferences around the World," in *Proceedings of the 2014 ACM CHI Conference on Human Factors in Computing Systems* (New York: ACM, 2014), 11–20, http://www.eecs.harvard.edu/~kgajos/papers/2014/reinecke14visual.pdf; Yuheng Hu, Lydia Manikonda, and Subbarao Kambhampati, "What We Instagram."

27. Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon, "What Is Twitter, a Social Network or a News Media?," in *Proceedings of the 19th International World Wide Web (WWW) Conference* (New York: ACM, 2014), 591–600, http://www.eecs.wsu.edu/~assefaw/CptS580-06/papers/2010-www-twitter.pdf.

28. Google Scholar, "Haewoon Kwak," accessed February 20, 2010, https://scholar.google.com/citations?user=M6i3Be0AAAAJ&hl=en.

29. Paul F. Lazarsfeld and Frank N. Stanton, eds., *Radio Research, 1941* (New York: Duel, Sloan and Pearce, 1942).

30. See Pierre Bourdieu, *Distinction: A Social Critique of the Judgement of Taste*, trans. Richard Nice (London: Routledge & Kegan Paul, 1979).

31. Maeve Duggan, Nicole B. Ellison, Cliff Lampe, Amanda Lenhart, and Mary Madden, "Demographics of Key Social Networking Platforms," Social Media Update 2014, Pew Research Center, January 9, 2015, http://www.pewinternet.org/2015/01/09/demographics-of-key-social-networking-platforms-2/.

32. Quoted in Philip Ball, *Critical Mass: How One Thing Leads to Another* (London: Arrow Books, 2004), 69–71.

33. Craig Smith, "By the Numbers: 400 Surprising Facebook Statistics (July 2016)," *Expandedramblings.com*, July 16, 2016, http://expandedramblings.com/index.php/by-the-numbers-17-amazing-facebook-stats/15/.

34. Christian Stefansen, "Google Flu Trends Gets a Brand New Engine," *Google AI Blog*, October 31, 2014, https://research.googleblog.com/2014/10/google-flu-trends-gets-brand-new-engine.html.

35. Michael Gavin, "Agent-Based Modeling and Historical Simulation," *Digital Humanities Quarterly* 8, no. 4 (2014); Graham Alexander Sack, "Character Networks for Narrative Generation: Structural Balance Theory and the Emergence of Proto-Narratives," in *Workshop on Computational Models of Narrative* (Dagstuhl, Germany: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2013), 183–197.

36. Douglas Fox, "IBM Reveals the Biggest Artificial Brain of All Time," *Popular Mechanics*, December 18, 2009, http://www.popularmechanics.com/technology/a4948/4337190/.

37. Nigel Gilbert and Klaus G. Troitzsch, *Simulation for the Social Scientist,* 2nd ed. (Maidenhead, England: Open University Press, 2005): 3–4; italics in original.

38. For the example of how agent-based simulation can be used to study the evolution of human societies, see Peter Turchin, Thomas E. Currie, Edward A. L. Turner, and Sergey Gavrilets, "War,

Space, and the Evolution of Old World Complex Societies," in *Proceedings of the National Academy of Sciences of the United States of America* 110, no. 41 (2013): 16384–16389.

3 Culture Industry and Media Analytics

1. Max Horkheimer and Theodor W. Adorno, *Dialectic of Enlightenment*, trans. E. Jephcott (Stanford, CA: Stanford University Press, 2002). Original book was published in Germany in 1947.

2. Janet Wiener and Nathan Bronson, "Facebook's Top Open Data Problems," Facebook Research, October 22, 2014, https://research.fb.com/facebook-s-top-open-data-problems/.

3. Chartbeat, "About," accessed July 1, 2015, https://chartbeat.com/about.

4. Nathan Bierma, "Amazon's SIPs Let Readers Search and Dip into Books," *Chicago Tribune*, May 24, 2005, http://articles.chicagotribune.com/2005-05-24/features/0505240239_1_improbable-phrases-books-word-pairs.

5. Greg Linden, Brent Smith, and Jeremy York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," *IEEE Internet Computing* 7, no. 1 (2003): 76–80.

6. Linden, Smith, and York, "Amazon.com Recommendations."

7. Gordon Donnelly, "75 Super-Useful Facebook Statistics for 2018," *WordStream* (blog), August 12, 2019, https://www.wordstream.com/blog/ws/2017/11/07/facebook-statistics.

8. J. Clement, "Most Famous Social Network Sites Worldwide as of July 2019, Ranked by Number of Active Users (in Millions)," Statista, accessed September 18, 2019, https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

9. Axel Bruns, "Facebook Shuts the Gate after the Horse Has Bolted, and Hurts Real Research in the Process," *Medium*, April 25, 2018, https://medium.com/@Snurb/facebook-research-data-18662cf2cacb.

10. "2019 Conference on Digital Experimentation (CODE). About," *MIT Digital*, accessed September 18, 2019, http://ide.mit.edu/events/2017-conference-digital-experimentation-code.

11. Matt Asay, "Beyond Hadoop: The Streaming Future of Big Data," *InfoWorld* (blog), March 23, 2015, http://www.infoworld.com/article/2900504/big-data/beyond-hadoop-streaming-future-of-big-data.html.

12. Spotify, "Get Audio Features for a Track," Spotify for Developers, accessed October 12, 2019, https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/.

13. Google, "How Google Search Works," accessed October 21, 2019, https://www.google.com/search/howsearchworks.

14. Alexis C. Madrigal, "How Netflix Reverse-Engineered Hollywood," *Atlantic*, January 2, 2014, http://www.theatlantic.com/technology/archive/2014/01/how-netflix-reverse-engineered-hollywood/282679/.

15. Stuart Dredge, "How Does Facebook Decide What to Show in My News Feed?," *Guardian*, June 30, 2014, https://www.theguardian.com/technology/2014/jun/30/facebook-news-feed-filters-emotion-study.

16. Françoise Beaufays, "The Neural Networks behind Google Voice Transcription," August 11, 2015, https://research.googleblog.com/2015/08/the-neural-networks-behind-google-voice.html.

17. Sundar Pichai, "TensorFlow: Smarter Machine Learning, for Everyone," November 9, 2015, https://googleblog.blogspot.com/2015/11/tensorflow-smarter-machine-learning-for.html.

18. Paul Sawers, "The Rise of OpenStreetMap: A Quest to Conquer Google's Mapping Empire," *TNW* (blog), February 28, 2014, http://thenextweb.com/insider/2014/02/28/openstreetmap/.

19. David Segal, "The Dirty Little Secrets of Search," *New York Times*, February 12, 2011, https://www.nytimes.com/2011/02/13/business/13search.html; Tom Vanderbilt, "The Science behind the Netflix Algorithms That Decide What You'll Watch Next," *Wired*, August 7, 2013, http://www.wired.com/2013/08/qq_netflix-algorithm/.

20. George Ritzer and Nathan Jurgenson, "Production, Consumption, Prosumption: The Nature of Capitalism in the Age of the Digital 'Prosumer,'" *Journal of Consumer Culture* 10 (1): 13–36. https://doi.org/10.1177/1469540509354673.

21. Mark Sanderson and W. Bruce Croft, "The History of Information Retrieval Research," *Proceedings of the IEEE* 100 (2012): 1444–1451, http://ciir-publications.cs.umass.edu/getpdf.php?id=1066.

22. Quoted in Eugene Garfield, "A Tribute to Calvin N. Mooers, a Pioneer of Information Retrieval," *Scientist* 11, no. 6 (March 17, 1997): 9, http://www.garfield.library.upenn.edu/commentaries/tsv11(06)p09y19970317.pdf.

23. This stage is analyzed in Lev Manovich, *Software Takes Command*, rev. ed. (London: Bloomsbury Academic, 2013).

24. Josh Constine, "How Instagram's Algorithm Works," *TechCrunch*, June 1, 2018, https://techcrunch.com/2018/06/01/how-instagram-feed-works/.

25. For other examples, see Celeste LeCompte, "Automation in the Newsroom," *Nieman Reports*, September 1, 2015, http://niemanreports.org/articles/automation-in-the-newsroom/; Shelley Podolny, "If an Algorithm Wrote This, How Would You Even Know?," *New York Times*, March 7, 2015, http://www.nytimes.com/2015/03/08/opinion/sunday/if-an-algorithm-wrote-this-how-would-you-even-know.html.

26. Mailchimp, "Use Send Time Optimization," October 5, 2017, https://kb.mailchimp.com/delivery/deliverability-research/use-send-time-optimization.

27. Twitter, "Follower Targeting on Twitter," accessed July 1, 2017, https://business.twitter.com/en/targeting/follower.html.

28. Felix Richter, "Digital Accounts for Nearly 70% of U.S. Music Revenues," Statista, September 30, 2014, https://www.statista.com/chart/2773/digital-music-in-the-united-states/.

29. According to the current default setting as of the time of writing, Facebook will show you only some of these posts, which it calls Top Stories, automatically selected by its algorithms. This setting can be changed by going to the News Feed tab and selecting Most Recent instead of Top Stories. See also Victor Luckerson, "Here's How Facebook's News Feed Actually Works," *Time*, July 9, 2015, http://time.com/3950525/facebook-news-feed-algorithm.

30. Corrado Mencar, "What Do You Mean by 'Interpretability' in Models?," ResearchGate, question posted July 7, 2013, https://www.researchgate.net/post/What_do_you_mean_by_interpretability_in_models.

31. Lev Manovich, "The Algorithms of Our Lives," *Chronicle of Higher Education*, December 16, 2013, http://chronicle.com/article/The-Algorithms-of-Our-Lives-/143557/.

32. Lev Manovich, *Software Takes Command*, rev. ed. (London: Bloomsbury Academic, 2013).

33. Wiener and Bronson, "Facebook's Top Open Data Problems."

34. Mikael Huss, "Data Size Estimates," *Follow the Data* (blog), June 24, 2014, https://followthedata.wordpress.com/2014/06/24/data-size-estimates/.

35. Alex Woodie, "The Rise of Predictive Modeling Factories," *Datanami* (blog), February 9, 2015, https://www.datanami.com/2015/02/09/rise-predictive-modeling-factories.

36. Gregory D. Abowd et al., "Towards a Better Understanding of Context and Context-Awareness," in *Handheld and Ubiquitous Computing 1999*, ed. H-W. Gellersen (Berlin and Heidelberg: Springer, 2001), ftp://ftp.cc.gatech.edu/pub/gvu/tr/1999/99-22.pdf.

37. David Carr, "Giving Viewers What They Want," *New York Times*, February 24, 2013, http://www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html.

38. Vanderbilt, "The Science behind the Netflix Algorithms."

39. Phil Simon, "Big Data Lessons from Netflix," *Wired*, March 2014, accessed February 28, 2020, https://www.wired.com/insights/2014/03/big-data-lessons-netflix/.

40. "Extracting Image Metadata at Scale," *Netflix Tech Blog*, March 21, 2016, https://netflixtechblog.com/extracting-image-metadata-at-scale-c89c60a2b9d2. .

41. Alex M., "Finding Beautiful Yelp Photos Using Deep Learning," *Yelp Engineering* (blog), November 29, 2016, https://engineeringblog.yelp.com/2016/11/finding-beautiful-yelp-photos-using-deep-learning.html.

42. Association for Psychological Science, "Political Polarization on Twitter Depends on the Issue," *ScienceDaily*, August 27, 2015, http://www.sciencedaily.com/releases/2015/08/150827083423.htm; Karen Kaplan, "Your Twitter Feed Says More about Your Political Views than You Think, Study Says," *Los Angeles Times*," September 18, 2015, http://www.latimes.com/science/la-sci-sn-twitter-political-conservative-republicans-20150917-story.html.

43. David A. Shamma, "One Hundred Million Creative Commons Flickr Images for Research," Yahoo Research, June 24, 2014, https://yahooresearch.tumblr.com/post/89783581601/one-hundred-million-creative-commons-flickr-images.

44. Miriam Redi, Damon Crockett, Lev Manovich, and Simon Osindero, "What Makes Photo Cultures Different?," in *Proceedings of the 24th ACM International Conference on Multimedia* (New York: ACM, 2016), 287–291, http://manovich.net/index.php/projects/what-makes-photo-cultures-different.

45. Cultural Analytics Lab, Phototrails, 2013, http://phototrails.info.

46. David Bordwell, Janet Staiger, Kristin Thompson, *The Classical Hollywood Cinema: Film Style and Mode of Production to 1960* (New York: Columbia University Press, 1985).

47. Max Horkheimer and Theodor W. Adorno, *Dialectic of Enlightenment*, trans. Edmund Jephcott (Stanford: Stanford University Press, 2002), 94.

48. Anant Gupta and Kuldeep Singh, "Location Based Personalized Restaurant Recommendation System for Mobile Environments," in *Proceedings of the International Conference on Advances in Computing, Communications and Informatics* (Mysore, India: Sri Jayachamarajendra College of Engineering, 2013), https://doi.org/10.1109/ICACCI.2013.6637223.

49. Renjie Zhou, Samamon Khemmarat, and Lixin Gao, "The Impact of YouTube Recommendation System on Video Views," in *Proceedings of the 2010 ACM Internet Measurement Conference* (New York: ACM, 2010), 404–410, http://conferences.sigcomm.org/imc/2010/papers/p404.pdf.

50. Nadav Hochman and Lev Manovich, "Zooming into an Instagram City: Reading the Local through Social Media," *First Monday* 18, no. 7 (July 1, 2013), http://firstmonday.org/ojs/index.php/fm/article/view/4711/3698.

51. Joan Serrà, Álvaro Corral, Marián Boguñá, Martín Haro, and Josep Ll. Arcos, "Measuring the Evolution of Contemporary Western Popular Music," *Scientific Reports* 2, no. 521 (2012), https://doi.org/10.1038/srep00521.

52. Matthias Mauch, Robert M. MacCallum, Mark Levy, and Armand M. Leroi, "The Evolution of Popular Music: USA 1960–2010," *Royal Society Open Science*, May 1, 2015, https://doi.org/10.1098/rsos.150081.

53. Academia.edu, accessed December 23, 2017, https://www.academia.edu/.

54. See also "List of Internet Phenomena," Wikipedia, accessed September 18, 2019, https://en.wikipedia.org/wiki/List_of_Internet_phenomena.

4 Types of Cultural Data

1. Samuel P. Fraiberger, Roberta Sinatra, Magnus Resch, Christoph Riedl, and Albert-László Barabási, "Quantifying Reputation and Success in Art," *Science* 362, no. 6416 (November 16, 2018): 825–829, https://science.sciencemag.org/content/362/6416/825.

2. "The 11th International AAAI Conference on Web and Social Media," accessed September 18, 2019, http://www.icwsm.org/2017/index.php.

3. Sandeep Junnarkar, "Bloggers Add Moving Images to Their Musings," *New York Times*, February 24, 2005, http://www.nytimes.com/2005/02/24/technology/circuits/bloggers-add-moving-images-to-their-musings.html; Nitecruzr, "Using Images in Your Posts," *The Real Blogger Status* (blog), October 6, 2006, http://blogging.nitecruzr.net/2006/10/using-images-in-your-posts.html.

4. Wikipedia, "DeviantArt," accessed August 2, 2016, http://en.wikipedia.org/wiki/DeviantArt.

5. DeviantArt, "About DeviantArt," accessed July 25, 2016, https://about.deviantart.com/.

6. Vimeo, "Motion Graphic Artists," accessed March 5, 2017, https://vimeo.com/groups/motion.

7. Daniela Ushizima et al., "Cultural Analytics of Large Datasets from Flickr," in *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media* (Palo Alto, CA: AAAI Press, 2015), http://manovich.net/index.php/projects/cultural-analytics-of-large-datasets-from-flickr.

8. Sirion Vittayakorn et al., "Runway to Realway: Visual Analysis of Fashion," in *Proceedings of 2015 IEEE Winter Conference on Applications of Computer Vision* (Waikoloa, HI: IEEE, 2015), 951–958.

9. Elena Garces et al., "A Similarity Measure for Illustration Style," *Journal ACM Transactions on Graphics* 33, no. 4 (July 2014). See also the Related Works section in this paper for more relevant research.

10. Yuji Yoshimura et al., "Deep Learning Architect: Classification for Architectural Design through the Eye of Artificial Intelligence," in *Computational Urban Planning and Management for Smart Cities*, ed. Stan Geertman et al. (Cham: Springer, 2019).

11. This topic is analyzed in detail in Lev Manovich, *Software Takes Command*, rev. ed. (London: Bloomsbury Academic, 2013).

12. In *Software Takes Command*, I argue that this constant evolution is the defining characteristic of computer media.

13. Schich et al., "A Network Framework of Cultural History."

14. Twitter Developer, "Tweet Objects," accessed March 12, 2017, https://dev.twitter.com/overview/api/users.

15. US Department of Health and Human Services, "Considerations and Recommendations Concerning Internet Research and Human Subjects Research Regulations," for SACHRP, March 13, 2013, http://www.hhs.gov/ohrp/sites/default/files/ohrp/sachrp/mtgings/2013%20March%20Mtg/internet_research.pdf.

16. Mike Schroepfer, "Research at Facebook," Facebook Newsroom, October 2, 2014, http://newsroom.fb.com/news/2014/10/research-at-facebook/.

17. Social Media Research Group, *Using Social Media for Social Research: An Introduction*, May 2016, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/524750/GSR_Social_Media_Research_Guidance_-_Using_social_media_for_social_research.pdf.

18. Venturini Tommaso and Richard Rogers, "'API-Based Research' or How Can Digital Sociology and Digital Journalism Studies Learn from the Cambridge Analytica Affair," *Digital Journalism* 7, no. 4 (2019): 532–540.

19. Ars Electronica, "Interactive Art +," accessed October 1, 2019, https://ars.electronica.art/prix/en/categories/interactive-art/.

20. Benjamin Tatler et al., "Yarbus, Eye Movements, and Vision," *Iperception* 1, no. 1 (2010): 7–27.

21. Alfred Yarbus, *Eye Movements and Vision* (New York: Plenum Press, 1967), 190.

22. Stanislav Sobolevsky et al., "Scaling of City Attractiveness for Foreign Visitors through Big Data of Human Economical and Social Media Activity," in *Proceedings of 2015 IEEE International Congress on Big Data* (Santa Clara, CA: IEEE, 2015), 600–607.

23. Senseable City Lab, accessed September 18, 2019, http://senseable.mit.edu/; Spin Unit, accessed September 18, 2019, http://www.spinunit.eu/; Habidatum, accessed September 18, 2019, https://habidatum.com/.

24. Todd Schneider, "A Tale of Twenty-Two Million Citi Bike Rides: Analyzing the NYC Bike Share System," *Todd W. Schneider* (blog), January 13, 2016, http://toddwschneider.com/posts/a-tale-of-twenty-two-million-citi-bikes-analyzing-the-nyc-bike-share-system/.

25. Patrick Nelson, "Just One Autonomous Car Will Use 4,000 GB of Data/Day," *NetworkWorld*, December 7, 2016, https://www.networkworld.com/article/3147892/internet/one-autonomous-car-will-use-4000-gb-of-dataday.html.

26. Jordan Gilbertson and Andrew Salzberg, "Introducing Uber Movement," *Uber Newsroom*, January 9, 2017, https://newsroom.uber.com/introducing-uber-movement/.

27. William H. Whyte, *The Social Life of Small Urban Spaces* (New York: Project for Public Spaces, 1980).

28. This section summarizes arguments developed in more detail in Lev Manovich, *Software Takes Command*, rev. ed. (London: Bloomsbury Academic, 2013).

29. Meetup, "Meetup: About," accessed March 11, 2017, https://www.meetup.com/about/.

30. Eventbrite, "Eventbrite: About," accessed March 11, 2017, https://www.eventbrite.com/about/.

31. J. Clement, "Number of Monthly Active Facebook Users Worldwide as of 4th Quarter 2019 (in Millions)," Statista, January 30, 2020, https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/; J. Clement, "Leading Countries Based on Number of Facebook Users as of July 2019 (in Millions)," Statista, February 14, 2020, https://www.statista.com/statistics/268136/top-15-countries-based-on-number-of-facebook-users/.

5 Cultural Sampling

1. Matthew Arnold, preface to *Culture and Anarchy* (1875), in *Arnold: Culture and Anarchy and Other Writings*, ed. Stefan Collini (Cambridge: Cambridge University Press, 1993).

2. Franco Moretti, "Conjectures on World Literature," *New Left Review* 1, no. 1 (January–February 2000): 54–68, http://newleftreview.org/II/1/franco-moretti-conjectures-on-world-literature.

3. Europeana, accessed August 3, 2018, http://www.europeana.eu/portal/en.

4. Internet Archive, accessed August 3, 2018, http://archive.org.

5. Google Arts & Culture, accessed July 26, 2016, http://www.google.com/culturalinstitute/beta/search/exhibit.

6. Lev Manovich, "How to Follow Global Digital Cultures, or Cultural Analytics for Beginners," in *Deep Search: The Politics of Search beyond Google*, ed. Konrad Becker and Felix Stalder (Innsbruck: Studien Verlag, 2009), 198–211, http://manovich.net/index.php/projects/how-to-follow-global-digital-cultures.

7. The very first large institutional collection that formed the core of Artstor was the slide library at the University of California, San Diego—the same university where I taught digital art and media theory since 1996. The library had over two hundred thousand slides, and they were all digitized and included in Artstor. In 2009, this was the largest single collection in Artstor. The slides were created either directly by art history faculty teaching in the Visual Arts Department or by art library staff via lists of images faculty provided. This collection is quite interesting because it reflects the biases of art history as it was taught over a few decades when color slides were the main media for teaching and studying art.

8. Manovich, "How to Follow Global Digital Cultures."

9. The New York Public Library Digital Collections, "Photographs of the Catskill Water Supply System in Process of Construction," accessed July 26, 2016, http://digitalcollections.nypl.org/collections/photographs-of-the-catskill-water-supply-system-in-process-of-construction.

10. The New York Public Library Digital Collections, "The Buttolph Collection of Menus," http://digitalcollections.nypl.org/collections/the-buttolph-collection-of-menus#/?tab=about.

11. The New York Public Library Digital Collections, "Catalogue of the Chiroptera by G.E. Dobson," accessed July 26, 2016, http://digitalcollections.nypl.org/collections/catalogue-of-the-chiroptera-by-ge-dobson.

12. Aleksandra Strzelichowska, "Maggy's Picks: New Content in Europeana," *Europeana Blog*, July 25, 2016, http://blog.europeana.eu/2016/07/maggys-picks-new-content-in-europeana/.

13. Lev Manovich, *Instagram and Contemporary Image* (self-published under Creative Commons License, 2017), http://manovich.net/index.php/projects/instagram-and-contemporary-image.

14. "Category: Geographic Region-Oriented Digital Libraries," Wikipedia, accessed September 18, 2019, https://en.wikipedia.org/wiki/Category:Geographic_region-oriented_digital_libraries; "List

of Digital Library Projects," Wikipedia, accessed September 18, 2019, https://en.wikipedia.org/wiki/List_of_digital_library_projects.

15. For an overview of different sampling methods, see Sam Cook, "Sampling Methods," Revise Sociology, May 4, 2011, https://revisesociology.wordpress.com/2011/05/04/5-sampling-methods. The longer list of methods is presented in Wikipedia, "Sampling (Statistics)," accessed August 2, 2018, https://en.wikipedia.org/wiki/Sampling_(statistics).

16. Nina Cesare, Christian Grant, Quynh Nguyen, Hedwig Lee, and Elaine O. Nsoesie, *How Well Can Machine Learning Predict Demographics of Social Media Users?*, ArXiv.org, February 6, 2017, https://arxiv.org/pdf/1702.01807.pdf.

17. Agustin Indaco and Lev Manovich, *Urban Social Media Inequality: Definition, Measurements, and Application*, ArXiv.org, July 7, 2016, https://arxiv.org/abs/1607.01845.

18. Manovich, *Instagram and Contemporary Image*.

19. Lydia Manikonda, Yuheng Hu, and Subbarao Kambhampati, *Analyzing User Activities, Demographics, Social Network Structure and User-Generated Content on Instagram*, Arxiv.org, October 29, 2014, http://arxiv.org/pdf/1410.8099v1.pdf.

20. National Gallery of Art, *The Art of the American Snapshot, 1888–1978: From the Collection of Robert E. Jackson*, accessed March 1, 2020, https://www.nga.gov/exhibitions/2007/snapshot.html.

21. Gallup, "Methodology Center," accessed August 2, 2016, http://www.gallup.com/178685/methodology-center.aspx.

22. Gallup, "How Does the Gallup U.S. Daily Work?," accessed August 2, 2016, http://www.gallup.com/185462/gallup-daily-work.aspx.

23. Rachel Donadio, "Revisiting the Canon Wars," *New York Times*, September 16, 2007, http://www.nytimes.com/2007/09/16/books/review/Donadio-t.html; Jan Gorak, *The Making of the Modern Canon: Genesis and Crisis of a Literary Idea* (London: Bloomsbury Academic, 2013).

24. Pew Research Center, "Internet User Demographics," accessed September 25, 2016, http://www.pewinternet.org/data-trend/teens/internet-user-demographics.

25. Brand Analytics, "Статистика по источникам," accessed September 20, 2016, https://br-analytics.ru/statistics/.

26. Statista, "Regional Distribution of Instagram Traffic in the Last Three Months as of April 2016, by Country," accessed September 20, 2016, https://www.statista.com/statistics/272933/distribution-of-instagram-traffic-by-country/.

27. Pierre Bourdieu, *Distinction: A Social Critique of the Judgement of Taste*, trans. Richard Nice (London: Routledge & Kegan Paul, 1979).

28. Frédéric Lebaron, "How Bourdieu 'Quantified' Bourdieu: The Geometric Modelling of Data," in *Quantifying Theory: Pierre Bourdieu*, ed. Karen Robson and Chris Sanders (Dordrecht: Springer, 2009), 11–29.

29. Christine A. Knoop, Valentin Wagner, Thomas Jacobsen, and Winfried Menninghaus, "Mapping the Aesthetic Space of Literature 'from Below,'" *Poetics* 56, no. 5 (June 2016): 35–49, https://doi.org/10.1016/j.poetic.2016.02.001.

30. Marc Verboord, Giselinde Kuipers, and Susanne Janssen, "Institutional Recognition in the Transnational Literary Field, 1955–2005," *Cultural Sociology* 9, no. 3 (September 2015): 447–465, https://doi.org/10.1177/1749975515576939.

31. Aurélie Van de Peer, "Re-artification in a World of De-artification: Materiality and Intellectualization in Fashion Media Discourse (1949–2010)," *Cultural Sociology* 8, no. 4 (December 2014): 443–461, https://doi.org/10.1177/1749975514539799.

32. James E. Cutting, *Impressionism and Its Canon* (Lanham, MD: University Press of America, 2006).

33. Jin Yea Jang et al., "Teens Engage More with Fewer Photos: Temporal and Comparative Analysis on Behaviors in Instagram," in *Proceedings of 2016 ACM Conference on Hypertext and Social Media* (New York: ACM, 2016), 71–81, https://doi.org/10.1145/2914586.2914602.

34. Manikonda, Hu, and Kambhampati, *Analyzing User Activities*.

35. Manikonda, Hu, and Kambhampati, *Analyzing User Activities*.

36. Émile Durkheim, *The Rules of Sociological Method* (New York: Free Press, [1895] 1982).

37. David Pierce, "Inside Spotify's Hunt for the Perfect Playlist," *Wired*, July 20, 2015, https://www.wired.com/2015/07/spotify-perfect-playlist/.

38. Franco Moretti, "Conjectures on World Literature."

39. "The Museum of Modern Art (MoMA) Exhibition and Staff Histories," GitHub, accessed September 18, 2019, https://github.com/MuseumofModernArt/exhibitions.

40. Cutting, *Impressionism and Its Canon*.

6 Metadata and Features

1. Ronald Fisher, *Statistical Methods for Research Workers* (Edinburgh: Oliver and Boyd, 1925), http://psychclassics.yorku.ca/Fisher/Methods/.

2. Theodore M. Porter, "Reforming Vision: The Engineer Le Play Learns to Observe Society Sagely," in *Histories of Scientific Observation*, ed. Lorraine Daston and Elizabeth Lunbeck (Chicago: University of Chicago Press, 2011), 281–302.

3. Michel Foucault, *Discipline and Punish: The Birth of the Prison*, trans. A. M. Sheridan Smith (New York: Pantheon Books, 1977), 129. Original book published in France in 1975.

4. Jean-Francois Lyotard, "The Field: Knowledge in Computerized Societies," in *The Postmodern Condition: Report on Knowledge* (Manchester, UK: Manchester University Press, 1984), 3–4. Original book published in France in 1979.

5. Museum of Modern Art, "Network Diagram of the Artists in Inventing Abstraction, 1910–1925," MoMA December 23, 2012–April 15, 2013, http://www.moma.org/interactives/exhibitions/2012/inventingabstraction/?page=connections.

6. Michel Foucault, *The Archaeology of Knowledge*, trans. A. M. Sheridan Smith (London: Routledge, 2002). Original book published in France in 1969.

7. "The Museum of Modern Art (MoMA) Collection," GitHub, accessed September 18, 2019, https://github.com/MuseumofModernArt/collection.

8. Alise Tifentale and Lev Manovich, "Selfiecity: Exploring Photography and Self-Fashioning in Social Media," in *Postdigital Aesthetics: Art, Computation and Design*, ed. David M. Berry and Michael Dieter (London: Palgrave Macmillan, 2015), 109–122, http://manovich.net/index.php/projects/selfiecity-exploring.

9. Saeideh Bakhshi, David Shamma, and Eric Gilbert, "Faces Engage Us: Photos with Faces Attract More Likes and Comments on Instagram," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York: ACM, 2014), 965–974.

10. Lisa Gitelman and Virginia Jackson, "Introduction," in *"Raw Data" Is an Oxymoron*, ed. Lisa Gitelman (Cambridge: MIT Press, 2013), 3.

11. OpenStreetMap, "OpenStreetMap: About," accessed September 19, 2019, https://www.openstreetmap.org/about.

12. Pew Research Center, accessed September 19, 2019, http://www.pewinternet.org.

13. Gallup, "How Does the Gallup U.S. Daily Work?"

14. Victor Ginsburgh and Sheila Weyers, "Persistence and Fashion in Art: Italian Renaissance from Vasari to Berenson and Beyond," *Poetics* 34, no. 1 (2006): 24–44.

15. Fionn Murtagh, *Origins of Modern Data Analysis Linked to the Beginnings and Early Development of Computer Science and Information Engineering*, ArXiv.org, October 30, 2018, https://arxiv.org/pdf/0811.2519.pdf.

16. MongoDB, "Industries," accessed September 28, 2016, https://www.mongodb.com/industries.

17. For an example of the alternative history that argues that conventional history misses some crucial figures, see Paul F. Lazarsfeld, "Notes on the History of Quantification in Sociology—Trends, Sources and Problems," *Isis* 52, no. 2 (1961): 277–333, https://www.jstor.org/stable/228683.

18. Alain Desrosieres, *The Politics of Large Numbers* (Cambridge, MA: Harvard University Press, 2002).

19. For historical examples, see Michael Friendly and Daniel J. Denis, "Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization," 2001, accessed March 1, 2020, http://www.datavis.ca/milestones/index.php.

20. Michael Friendly and Daniel Denis, "The Early Origins and Development of the Scatterplot," *Journal of the History of the Behavioral Sciences* 41, no. 2 (2005): 103–130.

21. Quoted in Garabed Eknoyan, "Adolphe Quetelet (1796–1874)—the Average Man and Indices of Obesity," *Nephrology Dialysis Transplantation* 23, no. 1 (2008): 47–51, https://doi.org/10.1093/ndt/gfm517.

22. Lazarsfeld, "Notes on the History of Quantification in Sociology," 297.

23. Émile Durkheim, *Le Suicide. Étude de Sociologie* (Paris: F. Alcan, 1897).

24. For a contemporary presentation and practical tutorial, see Oleksandr Pavlyk, "Centennial of Markov Chains," *Wolfram Blog*, February 4, 2013, http://blog.wolfram.com/2013/02/04/centennial-of-markov-chains/.

25. Lazarsfeld, "Notes on the History of Quantification in Sociology," 310.

26. Charles E. Spearman, "'General Intelligence' Objectively Determined and Measured," *American Journal of Psychology* 15 (1904): 201–293.

27. Raymond B. Cattell, ed., *Handbook of Multivariate Experimental Psychology* (Chicago: Rand McNally, 1966).

28. L. L. Thurstone, "The Vectors of Mind," *Psychological Review* 41 (1934): 1–32.

29. Neil W. Henry, "Latent Structure Analysis at Fifty," in *Proceedings of the Survey Research Methods Section* (American Statistical Association, 1999), 587–592, http://www.asasrms.org/Proceedings/papers/1999_102.pdf.

30. T. W. Anderson, *An Introduction to Multivariate Statistical Analysis* (New York: Wiley, 1958).

31. Warren S. Torgerson, *Theory and Methods of Scaling* (New York: Wiley, 1958).

32. Michael Baxandall, *Painting and Experience in Fifteenth Century Italy: A Primer in the Social History of Pictorial Style* (Oxford: Clarendon Press, 1972); Victor Burgin, *The End of Art Theory: Criticism and Postmodernity* (Basingstoke: Macmillan, 1986).

33. For example, see the program for the 2016 Workshop on Human Interpretability in Machine Learning, New York, June 23, 2016, https://sites.google.com/site/2016whi/.

34. For example, see Matthew D. Zeiler and Rob Fergus, *Visualizing and Understanding Convolutional Networks*, ArXiv.org, November 28, 2013, http://arxiv.org/abs/1311.2901.

35. Phil Schiller, quoted in Tonya Riley, "Apple's iPhone 7 Camera Uses Machine Learning to Look for People," *Inverse*, September 7, 2016, https://www.inverse.com/article/20677-iphone-7-camera-isp-phone.

36. Glenn Fleishman, "Two Cameras in iPhone 7 Plus Allow Synthetic Zoom, Soft-Focus Backgrounds," *Macworld*, September 7, 2016, http://www.macworld.com/article/3117258/iphone-ipad/two-cameras-in-iphone-7-plus-allow-synthetic-zoom-soft-focus-backgrounds.html.

37. EyeEm, "EyeEm Team," accessed September 23, 2016, https://www.eyeem.com/u/team.

38. Olga Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision* 115, no. 3 (2015): 211–252, https://doi.org/10.1007/s11263-015-0816-y.

39. See Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge." For the details of the results from all competing teams, see ImageNet, "Large Scale Visual Recognition Challenge 2015 (ILSVRC2015)," accessed March 1, 2020, http://www.image-net.org/challenges/LSVRC/2015/results.

40. Google Cloud Platform, "Cloud Vision API," accessed August 8, 2016, https://cloud.google.com/vision/.

41. Andrew Ng, Machine Learning Course, online course, week 6, accessed September 28, 2016, https://www.coursera.org/learn/machine-learning/home/week/6.

42. Kim Hye-Rin et al.,"Building Emotional Machines: Recognizing Image Emotions through Deep Neural Networks," *IEEE Transactions on Multimedia 20*, no. 11 (November 2018): 2980–2992.

43. David G. Lowe, "Object Recognition from Local Scale-Invariant Features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2 (Washington, DC: IEEE Computer Society, 1999), 1150–1157, https://doi.org/10.1109/ICCV.1999.790410. A good summary of the use of SIFT for object detection using the "bag of words" approach is provided in Gil Levi, "Bag of Words Models for Visual Categorization," *Gil's CV Blog*, August 23, 2013, https://gilscvblog.com/2013/08/23/bag-of-words-models-for-visual-categorization/.

44. Paul Viola and Michael Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1 (Los Alamitos, CA: IEEE Computer Society, 2001), 511–518, https://doi.org/10.1109/CVPR.2001.990517.

45. A. Huertas and R. Nevatia, "Detecting Buildings in Aerial Images," *Computer Vision, Graphics, and Image Processing* 41, no. 2 (1988): 131–152.

46. Yann LeCun, Yoshua Bengio, and Geoffrey E. Hinton, "Deep Learning," *Nature* 521, no. 7553 (2015): 436–444, https://doi.org/10.1038/nature14539.

47. The paper that started the trend of using deep networks for image classification is Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems 25* (New York: AMC, 2012), 1097–1105, http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.

48. Krizhevsky et al., "ImageNet Classification with Deep Convolutional Neural Networks."

49. The fact that certain data types are very popular has to do with the history of computer use in research and industry. Many algorithms were developed to analyze certain types, while other possible types did not receive the same attention. For each of the data types in my list, there are now

common methods of analysis, corresponding algorithms, and various data formats. For example, spatial data can be represented as coordinates, as shapefiles, or in other ways.

7 Language, Categories, and Senses

1. Peter M. Broadwell, David Mimno, and Timothy R. Tangherlini, "The Tell-Tale Hat: Surfacing the Uncertainty in Folklore Classification," *Journal of Cultural Analytics*, February 8, 2017, https://doi.org/10.31235/osf.io/a7dp8.

2. Ted Underwood, *Understanding Genre in a Collection of a Million Volumes*, Interim Performance Report Digital Humanities Start-Up Grant, Award HD5178713, December 29, 2014, https://figshare.com/articles/Understanding_Genre_in_a_Collection_of_a_Million_Volumes_Interim_Report/1281251.

3. "The Museum of Modern Art (MoMA) Exhibition and Staff Histories," GitHub, accessed September 18, 2019, https://github.com/MuseumofModernArt/exhibitions.

4. Wikipedia, "Statistical Data Type," accessed August 15, 2016, https://en.wikipedia.org/wiki/Statistical_data_type.

5. The literature on the introduction of metric time systems and their contribution to the rationalization of work and life is vast. See, for example, E. P. Thompson, "Time, Work-Discipline, and Industrial Capitalism," *Past & Present*, no. 38 (December 1967): 56–97; Jonathan Martineau, *Time, Capitalism, and Alienation: A Socio-Historical Inquiry into the Making of Modern Time* (Chicago: Haymarket Books, 2016).

6. Stanley S. Stevens, "On the Theory of Scales of Measurement," *Science* 103, no. 2684 (June 7, 1946): 677–680.

7. Affectiva, "Metrics," accessed September 18, 2019, https://developer.affectiva.com/metrics/.

8. Clement Greenberg, "Avant Garde and Kitsch," *Partisan Review* (1939): 34–49.

9. Roman Jakobson, "Verbal Communication," *Scientific American* 227 (1972): 72–80.

10. Marta J. Hardman, "Why We Should Say 'Women and Men' Until It Doesn't Matter Anymore," *Women and Language* 22, no. 1 (1999): 1–2.

11. Rensis Likert, "A Technique for the Measurement of Attitudes," *Archives of Psychology* 22 (1932–1933): 5–55, https://legacy.voteview.com/pdf/Likert_1932.pdf.

12. Maximilian Schich et al., "A Network Framework of Cultural History."

13. Lev Manovich, "There Is No Software," in *Nam June Paik Reader: Contributions to an Artistic Anthropology*, ed. Youngchul Lee and Henk Slager (Seoul: NJP Art Center, 2009), 26–29.

14. For a discussion of nineteenth-century debates about the meaning of the average and different types of averages, see Alain Desrosières, "Averages and the Realism of Aggregates," in *The Politics of Large Numbers: A History of Statistical Reasoning* (Cambridge, MA: Harvard University Press, 1998), 67–102.

15. On the change in phone designs from functional to highly aesthetic, see my 2007 article: Lev Manovich, "Information as an Aesthetic Event," *Receiver* 17, http://manovich.net/index.php/projects/information-as-an-aesthetic-event.

16. James Peckham, "Huawei P20 and P20 Pro Colors: What Shade Should You Buy?," *TechRadar*, March 27, 2018, https://www.techradar.com/news/huawei-p20-and-p20-pro-colors-what-shade-should-you-buy.

17. Huawei, "Huawei P20," accessed October 1, 2019, https://consumer.huawei.com/en/phones/p20/.

18. Paul Goldberger, "On Madison Avenue, Sometimes Less Is Less," *New York Times,* October 27, 1996, http://www.nytimes.com/1996/10/27/arts/on-madison-avenue-sometimes-less-is-less.html.

19. John Pawson, *Calvin Klein Collections Store*, Johnpawson.com, accessed June 1, 2017, http://www.johnpawson.com/works/calvin-klein-collections-store.

20. Walter Isaacson, "How Steve Jobs' Love of Simplicity Fueled a Design Revolution," *Smithsonian Magazine*, September 24, 2012, https://www.smithsonianmag.com/arts-culture/how-steve-jobs-love-of-simplicity-fueled-a-design-revolution-23868877/.

21. Ted Gibson and Bevil R. Conway, "The World Has Millions of Colors. Why Do We Only Name a Few?" *Smithsonian Magazine*, September 19, 2017, https://www.smithsonianmag.com/science-nature/why-different-languages-name-different-colors-180964945/.

22. Lev Nusberg, "Cybertheater," *Leonardo* 2 (1969): 61–62, https://monoskop.org/images/a/af/Nusberg_Lev_1969_Cybertheater.pdf.

23. Hadley Feingold, "Sculptural Fashion: Volume, Structure, and the Body," *Textile Arts Center* (blog), January 15, 2018, http://textileartscenter.com/blog/sculptural-fashion-volume-structure-and-the-body.

24. Tor D. Wager, Lauren Y. Atlas, Martin A. Lindquist, Mathieu Roy, Choong-Wan Woo, and Ethan Kross, "An fMRI-Based Neurologic Signature of Physical Pain," *New England Journal of Medicine* 368 (2013): 1388–1397.

25. Jeffrey Bardzell, "Interaction Criticism: An Introduction to the Practice," *Interacting with Computers* 23 (2011): 604–621.

26. Emotiv, "MyEmotiv," accessed September 18, 2019, https://www.emotiv.com/myemotiv/.

27. Affectiva, "Affectiva Automotive AI," accessed December 16, 2018, https://www.affectiva.com/product/affectiva-automotive-ai/.

28. Peter Weibel and Jeffrey Shaw, *Future Cinema: The Cinematic Imaginary after Film* (Cambridge, MA: MIT Press, 2003); Cristiane Paul, *Digital Art* (London: Thames and Hudson, 2003); Lucy Bullivant, *Responsive Environments: Architecture, Art and Design*, (London: Victoria and Albert Museum, 2006).

29. On progress in *neurocinema* (the field of measuring brain reactions to films), see Aalto University, "Nolan Film 'Memento' Reveals How the Brain Remembers and Interprets Events

from Clues," *Medical Press*, February 22, 2018, https://medicalxpress.com/news/2018-02-nolan-memento-reveals-brain-events.html.

30. Kevin Gray and Barry Gills, "South–South Cooperation and the Rise of the Global South," *Third World Quarterly* 37, no. 4 (2016): 557–574.

31. Martin Muller, "In Search of the Global East: Thinking between North and South," *Geopolitics*, October 2008, 1–22, https://doi.org/10.1080/14650045.2018.1477757.

32. Tuvikene, quoted in Wladimir Zbignev, "Theorizing Cities from/with a Global East," *Connections*, September 14, 2018, https://www.connections.clio-online.net/event/id/termine-38138.

33. Lev Manovich, *Software Takes Command*, rev. ed. (London: Bloomsbury Academic, 2013).

34. Wikipedia, "List of Subcultures," accessed August 12, 2016, https://en.wikipedia.org/wiki/List_of_subcultures.

35. Geoffrey C. Bowker and Susan L. Star, *Sorting Things Out: Classification and Its Consequences* (Cambridge, MA: MIT Press, 2000), https://www.ics.uci.edu/~gbowker/classification/. Emphasis in original.

36. Johan Bollen, Herbert Van de Sompel, Aric Hagberg, Luis Bettencourt, Ryan Chute, Marko A. Rodriguez, and Lyudmila Balakireva, "Clickstream Data Yields High-Resolution Maps of Science," *PLOS ONE* 4, no. 3 (2009), https://doi.org/10.1371/journal.pone.0004803; Katy Börner, Richard Klavans, Michael Patek, Angela M. Zoss, Joseph R. Biberstine, Robert P. Light, Vincent Larivière, and Kevin W. Boyack, "Design and Update of a Classification System: The UCSD Map of Science," *PLOS ONE* 7, no. 7 (2012), https://doi.org/10.1371/journal.pone.0039464.

37. Katy Börner at al, "Design and Update of a Classification System: The UCSD Map of Science."

38. Richard Klavans and Kevin Boyack, "Toward an Objective, Reliable and Accurate Method for Measuring Research Leadership," *Scientometrics* 82 (2010): 539–553.

39. Michel Foucault, *Les mots et les choses: Une archéologie des sciences humaines* (Paris: Éditions Gallimard, 1966).

40. Software Studies Initiative, "ImagePlot Visualisation Software," 2011, accessed March 1, 2020, http://lab.softwarestudies.com/p/imageplot.html.

41. Van Gogh Museum, "Meet Vincent," accessed July 31, 2016, vangoghmuseum.nl.

42. Van Gogh Museum, "Arles 1888–1889," accessed July 31, 2016, vangoghmuseum.nl.

43. Van Gogh Museum, "Arles 1888–1889," accessed July 31, 2016, vangoghmuseum.nl.

44. For additional examples, see Lev Manovich, *Style Space: How to Compare Image Sets and Follow Their Evolution*, 2011, accessed March 1, 2010, http://manovich.net/index.php/projects/style-space.

45. Franco Moretti, *Graphs, Maps, Trees: Abstract Models for a Literary History* (London: Verso, 2005). See also "Pamphlets," Stanford Literary Lab, accessed September 18, 2019, https://litlab.stanford.edu/pamphlets/.

46. Ted Underwood and Jordan Sellers, "The Emergence of Literary Diction," *Journal of Digital Humanities* 1, no. 2 (2012), http://journalofdigitalhumanities.org/1-2/the-emergence-of-literary-diction-by-ted-underwood-and-jordan-sellers/; Ted Underwood, Michael L. Black, Loretta Auvil, and Boris Capitanu, "Mapping Mutable Genres in Structurally Complex Volumes," arXiv.org, September 18, 2013, https://arxiv.org/abs/1309.3323; Ted Underwood, "The Life Cycles of Genres," *Cultural Analytics* 1 (May 23, 2016), http://culturalanalytics.org/2016/05/the-life-cycles-of-genres.

47. Lev Manovich, *One Million Manga Pages*, 2010 research report, March 1, 2020, http://lab.softwarestudies.com/2010/11/one-million-manga-pages.html.

48. Jeremy Douglass, William Huber, and Lev Manovich, "Understanding Scanlation: How to Read One Million Fan-Translated Manga Pages," *Image and Narrative* (Winter 2011), http://manovich.net/index.php/projects/understanding-scanlation. For history of the OneManga site, see http://fanlore.org/wiki/OneManga, accessed October 26, 2016.

49. Nanjing University of the Arts, "Disciplines," accessed December 28, 2018, http://en.nua.edu.cn/2639/list.htm.

50. The New School, "Undergraduate Academics," accessed December 28, 2018, https://www.newschool.edu/academics/undergraduate/.

51. Mehrdad Yazdani, Jay Chow, and Lev Manovich, "Quantifying the Development of User-Generated Art during 2001–2010," *PLOS One*, August 7, 2017, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0175350.

52. Eugene Garfield, "Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas," *Science* 122, no. 3159 (1955): 108–111, https://doi.org/10.1126/science.122.3159.108.

53. Tim Ingham, "World's Top 5 Music Publishers Now Control 11 Million Songs," *Music Business Worldwide*, May 25, 2015, https://www.musicbusinessworldwide.com/top-5-publishers-now-control-11m-songs/.

54. Kim Albrecht, "Cultural Development in Movie History," *Culturegraphy*, accessed September 18, 2019, https://www.culturegraphy.com/extras/findings/.

55. Kim Albrecht, "Cultural Development in Movie History."

56. Albrecht, "Cultural Development in Movie History."

57. Nadav Hochman and Lev Manovich, "A View from Above: Exploratory Visualizations of the Thomas Walther Collection," in *Object:Photo. Modern Photographs: The Thomas Walther Collection 1909–1949*, ed. Mitra Abbaspour, Lee Ann Daffner, and Maria Morris Hambourg (New York: Museum of Modern Art, 2014), 1–6, http://www.moma.org/interactives/objectphoto/assets/essays/Manovich_Hochman.pdf.

58. Gallup, *Gallup Global Wellbeing: The Behavioral Economics of GDP Growth* (Washington, DC: Gallup, 2010), http://www.gallup.com/poll/126965/gallup-global-wellbeing.aspx.

59. Jean-Paul Benzécri, *L'Analyse des Données*, vol. 2, *L'Analyse des Correspondances* (Paris: Dunod, 1973). Correspondence analysis is available in various statistical software environments, including R.

60. Weibo, accessed August 12, 2016, http://d.weibo.com/.

61. Mitch Joel, "We Need a Better Definition of 'Native Advertising,'" *Harvard Business Review*, February 13, 2013, https://hbr.org/2013/02/we-need-a-better-definition-of.

62. Presentations during Advertising Week NYC 2016, New York, September 26–30, 2016.

63. Andrew Bosworth, "What's the History of the Awesome Button (that Eventually Became the Like Button) on Facebook?," Quora, October 17, 2014, https://www.quora.com/Whats-the-history-of-the-Awesome-Button-that-eventually-became-the-Like-button-on-Facebook.

64. Bart de Langhe, Philip M. Fernbach, and Donald R. Lichtenstein, "Navigating by the Stars: Investigating the Actual and Perceived Validity of Online User Rating," *Journal of Consumer Research* 42, no. 6 (2016): 817–833.

8 Information Visualization

1. William Playfair, *An Inquiry into the Permanent Causes of the Decline and Fall of Powerful and Wealthy Nations: Illustrated by Four Engraved Charts* (London: Printed for Greenland and Norris, 1805).

2. Robert Venturi, Denise Scott Brown, and Steven Izenour, *Learning from Las Vegas: The Forgotten Symbolism of Architectural Form* (Cambridge, MA: MIT Press, 1977). Emphasis in original.

3. Bruno Latour, "Tarde's Idea of Quantification," in *The Social after Gabriel Tarde: Debates and Assessments*, ed. Mattei Candea (London: Routledge, 2010), 116.

4. Eric Rodenbeck, keynote lecture at O'Reilly Emerging Technology 2008 conference, March 4, 2008.

5. "Interview: Fernanda Viégas and Martin Wattenberg from Flowing Media," *Information Aesthetics*, May 7, 2010, https://flowingdata.com/2010/05/13/interview-fernanda-vigas-and-martin-wattenberg/.

6. Google, "Public Data," accessed September 18, 2019, http://www.google.com/publicdata/directory.

7. Daniel A. Keim, Florian Mansmann, Jörn Schneidewind, and Hartmut Ziegler, "Challenges in Visual Data Analysis," in *Proceedings of Information Visualization* (Piscataway, NJ: IEEE Computer Society, 2006), 9–16, 10, https://doi.org/10.1109/IV.2006.31.

8. Helen C. Purchase, Natalia Andrienko, T. J. Jankun-Kelly, and Matthew Ward, "Theoretical Foundations of Information Visualization," in *Information Visualization: Human-Centered Issues and Perspectives*, ed. Andreas Kerren, John T. Stasko, and Jean-Daniel Fekete (Berlin: Springer, 2008), 46–64.

9. Theusrus, "Mondrian: About," accessed September 18, 2019, http://www.theusrus.de/Mondrian/.

10. For example: "In contrast to scientific visualization, information visualization typically deals with nonnumeric, nonspatial, and high-dimensional data." Chaomei Chen, "Top 10 Unsolved Information Visualization Problems," *IEEE Computer Graphics and Applications* 25, no. 4 (2005): 12–16.

11. Fernanda B. Viégas, Martin Wattenberg, and Kushal Dave, "Studying Cooperation and Conflict between Authors with History Flow Visualizations," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, April 2004, 575–582, https://doi.org/10.1145/985692.985765.

12. Aaron Koblin, *Flight Patterns*, accessed September 18, 2019, http://www.aaronkoblin.com/work/flightpatterns/.

13. Processing, accessed September 18, 2019, http://processing.org/.

14. Data-Driven Documents, accessed September 18, 2019, https://d3js.org/.

15. Hadley Wickham et al,, ggplot2 (software), accessed March 1, 2020, https://ggplot2.tidyverse.org/.

16. "Harry Beck's Tube Map," Transport for London, accessed September 18, 2019, https://tfl.gov.uk/corporate/about-tfl/culture-and-heritage/art-and-design/harry-becks-tube-map.

17. Edward Tufte, *The Visual Display of Quantitative Information* (Cheshire, CT: Graphics Press, 1983); *Envisioning Information* (Cheshire, CT: Graphics Press, 1990); *Visual Explanations: Images and Quantities, Evidence and Narrative* (Cheshire, CT: Graphics Press, 1997); *Beautiful Evidence* (Cheshire, CT: Graphics Press, 2006).

18. Several definitions of information visualization from the recent literature are available at "Information Visualization," InfoVis Wiki, accessed September 27, 2019, https://infovis-wiki.net/wiki/Information_Visualization.

19. Michael Friendly and Daniel J. Denis, "1800–1849: Beginnings of Modern Data Graphics," Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization, accessed March 1, 2020. http://www.datavis.ca/milestones/index.php?group=1800%2B.

20. Philip Ball, *Critical Mass* (London: Arrow Books, 2004), 64–65.

21. Michael Friendly and Daniel J. Denis, Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization, 2001, accessed March 1, 2020, http://www.datavis.ca/milestones/.

22. Historical data is from Friendly and Denis, Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization.

23. Ben Fry, *Distellamap*, August 2005, accessed March 1, 2020, http://benfry.com/distellamap/.

24. Marcos Weskamp, "The Movement," December 29, 2005, https://www.flickr.com/photos/pkeenan/79036462.

25. InfoVis Lab, "Research," accessed September 18, 2019, http://ivl.slis.indiana.edu/research/.

26. Edward Tufte, "Minard's Sources—from Virginia Tufte and Dawn Finley," Edwardtufte.com, August 7, 2002, http://www.edwardtufte.com/tufte/minard.

27. Visual Complexity, "The Evolution of The Origin of Species," accessed September 19, 2019, http://www.visualcomplexity.com/vc/project.cfm?id=696.

28. Google Trends, accessed September 19, 2019, http://www.google.com/trends.

29. One important case that does not fit my analysis is the use of different tones or colors to represent terrain elevation and relief in printed topographic maps already common in the eighteenth century. In these maps, tone or color codes quantitative data rather than categories.

30. Wikipedia, "Tag Cloud," accessed July 18, 2016, http://en.wikipedia.org/wiki/Tag_cloud.

31. As an example, open-source data visualization software Mondrian 1.0, running on my 2009 Apple PowerBook laptop with a 2.8 GHz processor and 4 GB of RAM, took approximately seven seconds to render a scatter plot containing one million points.

32. Many additional examples of direct visualization can be found in the field of motion graphics: film and TV titles and graphics, commercials, and music videos. In many motion graphics, text or images are animated to create dynamically changing, meaningful patterns made from these media objects.

33. Brendan Dawes, "Cinema Redux," 2004, accessed March 1, 2020, http://brendandawes.com/projects/cinemaredux.

34. Ben Fry, "Traces," 2009, accessed March 1, 2020, https://fathom.info/traces/.

35. I have created a few visualizations that show a whole book in a single image. See http://www.flickr.com/photos/culturevis/sets/72157615900916808/; http://www.flickr.com/photos/culturevis/sets/72157622994317650/. To display the whole text of Tolstoy's *Anna Karenina* in the smallest font that can be read, I had to make a 14,000 × 6,000 pixels image—well beyond current screen resolutions.

36. Roberta Smith, "Art in Review; Mark Hansen and Ben Rubin—'Listening Post,'" *New York Times*, February 21, 2003, https://www.nytimes.com/2003/02/21/arts/art-in-review-mark-hansen-and-ben-rubin-listening-post.html.

37. To see Manual Lima's taxonomy of network display methods, select "filter by method" from the "Filter by:" dropdown," accessed March 1, 2020, www.visualcomplexity.com/vc/.

38. Latour, "Tarde's Idea of Quantification."

39. Wikipedia, "Synechdoche," Wikipedia, accessed July 18, 2016, http://en.wikipedia.org/wiki/Synecdoche.

40. Stefanie Posavec, *Writing without Words*, 2008, accessed March 1, 2020, http://www.stefanieposavec.com/writing-without-words; Martin Wattenberg, The Shape of Song, 2001, accessed March 1, 2020, http://www.bewitched.com/song.html.

41. Lev Manovich, "Data Visualization as New Abstraction and Anti-Sublime," *SMAC! 3* (2002): n.p. (San Francisco, 2002), http://manovich.net/index.php/projects/data-visualisation-as-new-abstraction-and-anti-sublime.

42. David L. Small, *Rethinking the Book* (PhD thesis, MIT, January 1999), https://acg.media.mit.edu/projects/thesis/DSThesis.pdf.

43. Ben Fry, *Valence*, 2001, accessed March 1, 2020, http://benfry.com/valence/.

44. W. Bradford Paley, *TextArc*, 2002, accessed March 1, 2020, http://wbpaley.com/brad/projects.html.

45. Frank van Ham, Martin Wattenberg, and Fernanda B. Viégas, "Mapping Text with Phrase Nets," *IEEE Transactions on Visualization and Computer Graphics* 15, no. 6 (2009): 1169–1176, https://doi.org/10.1109/TVCG.2009.165.

46. Software Studies Initiative, "Image_Graphr Outputs," Flickr, accessed September 18, 2019, https://www.flickr.com/photos/culturevis/sets/72157617847338031/.

47. Wayne Rasband, ImageJ, accessed September 20, 2019, https://imagej.nih.gov/ij/.

48. Lev Manovich, "Cultural Analytics Visualizations on Ultra High Resolution Displays," Software Studies Initiative, December 24, 2008, http://lab.softwarestudies.com/2008/12/cultural-analytics-hiperspace-and.html.

49. Humanities+Digital Visual Interpretations Conference: Aesthetics, Methods, and Critiques of Information Visualization in the Humanities, Arts, and Social Sciences, conference at Massachusetts Institute of Technology, Cambridge, Massachusetts, May 22–22, 2010, https://www.iri.centrepompidou.fr/evenement/humanitiesdigital-visual-interpretations-conference-2010/.

50. It is possible, however, that our interactive interfaces with visualizations are effective precisely because they do provide certain reduction functions. I am thinking in particular about the zoom command. We zoom into direct visualizations such as *Time* covers to examine the details of particular covers. We zoom out to see the overall trends. When we do that, the images are gradually reduced in size, eventually becoming small color dots.

9 Exploratory Media Analysis

1. John W. Tukey, *Exploratory Data Analysis* (Reading, MA: Addison-Wesley, 1977).

2. "Chronicling America: Historic American Newspapers," Library of Congress, accessed July 7, 2019, http://chroniclingamerica.loc.gov/.

3. Internet Archive, accessed February 20, 2020, https://archive.org/.

4. "Art Now", Flickr group, accessed July 7, 2016, http://www.flickr.com/groups/37996597808@N01/.

5. Coroflot, "About Us," accessed July 7, 2016, http://www.coroflot.com/about.

6. "Prints & Photographs Online Catalog," Library of Congress, accessed July 7, 2016. http://www.loc.gov/pictures/.

7. Flickr, "The App Garden," accessed July 7, 2016, http://www.flickr.com/services/api/.

8. For more details, see Steve Stemler, "An Overview of Content Analysis," *Practical Assessment, Research & Evaluation* 7, no. 17 (2001), http://PAREonline.net/getvn.asp?v=7&n=17.

9. Calvin N. Mooers, *The Theory of Digital Handling of Non-Numerical Information and Its Implications to Machine Economics* (Boston: Zator Co., 1950).

10. Mooers, *The Theory of Digital Handling*, 1–2. Emphasis added.

11. Calvin N. Mooers, *Scientific Information Retrieval Systems for Machine Operation: Case Studies in Design* (Boston: Zator Co., 1951), 3.

12. Mooers, *The Theory of Digital Handling*, 2. Emphasis added.

13. Vannevar Bush, "As We May Think," *Atlantic Monthly*, July 1945, http://web.mit.edu/STS.035/www/PDFs/think.pdf.

14. Today, scientists do not keep up with research by moving from references in one article to other articles and sites; instead, they search giant databases and depositories of science publications and data, such as ACM, arXiv, IEEE, PubMed, ProQuest, Web of Science, ScienceDirect, and many others. And because science publishing is highly structured, with articles and conference papers including subject categories, keywords, overviews of related research with many citations, and ID numbers, all being generated automatically for each new publication, the database paradigm works quite well.

15. These features correspond to the Instagram UI as of March 2019. The interface can change in the future: new functions can be added and older ones altered.

16. This description applies to versions of these applications and apps as of March 2019.

17. Lev Manovich, Moritz Stefaner, Mehrdad Yazdani, Dominikus Baur, Daniel Goddemeyer, Alise Tifentale, Nadav Hochman, and Jay Chow, *Selfiecity,* a website and custom interactive app, 2014, and *Selfiecity London,* a website and custom interactive app, 2015, http://selfiecity.net/ and http://selfiecity.net/london/; Daniel Goddemeyer, Moritz Stefaner, Dominikus Baur, and Lev Manovich, *On Broadway,* interactive artwork for touch display, 2014, http://on-broadway.nyc.

18. Wikipedia, "Digital Image Processing," accessed June 6, 2016, http://en.wikipedia.org/wiki/Digital_image_processing.

19. "Text Analysis," Tooling up for Digital Humanities, accessed July 27, 2016, http://toolingup.stanford.edu/?page_id=981.

20. Voyant, accessed September 20, 2019, https://voyant-tools.org; Matthew Jockers, *Text Analysis with R for Students of Literature* (Berlin: Springer, 2014).

21. For an explanation of image features learned by convolutional networks, see Matthew D. Zeiler and Rob Fergus, *Visualizing and Understanding Convolutional Networks*, ArXiv.org, November 12, 2013, https://arxiv.org/abs/1311.2901.

22. For discussion of computer vision adoption in photography services and phone cameras, see Lev Manovich, *AI Aesthetics* (Moscow: Strelka Press, 2018).

23. Clarifai, "Models," accessed September 20, 2019, https://clarifai.com/models.

24. David Ramli and Shelly Banjo, "The Kids Use TikTok Now Because Data-Mined Videos Are So Much Fun," *Bloomsburg BusinessWeek*, April 17, 2019, http://www.bloomberg.com/news/features/2019-04-18/tiktok-brings-chinese-style-censorship-to-america-s-tweens.

25. Mehrdad Yazdani and Lev Manovich, "Predicting Social Trends from Non-photographic Images on Twitter," in *Proceedings of the 2015 IEEE International Conference on Big Data* (Washington, DC: IEEE Computer Society, 2015), 1653–1660.

26. Miriam Redi, Damon Crockett, and Lev Manovich, "What Makes Photo Cultures Different?," in *Proceedings of the 24th ACM International Conference on Multimedia* (New York: ACM, 2016), 287–291.

27. Yale Digital Humanities Lab, "Neural Neighbors: Capturing Image Similarity," accessed October 1, 2019, https://dhlab.yale.edu/projects/neural-neighbors/.

28. Konstantinos Rematas, Basura Fernando, Frank Dellaert, and Tinne Tuytelaars, "Dataset Fingerprints: Exploring Image Collections through Data Mining," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Washington, DC: IEEE Computer Society, 2015), 4867–4875.

29. Benoît Seguin, *Making Large Art Historical Photo Archives Searchable* (Lausanne: EPFL, 2018), https://infoscience.epfl.ch/record/261212?ln=en.

30. Seguin, *Making Large Art Historical Photo Archives Searchable*, 118.

10 Methods of Media Visualization

1. John Unsworth, "Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might our Tools Reflect This?," Humanities Computing Symposium, May 13, 2000, King's College, London, accessed October 1, 2019, http://www.people.virginia.edu/~jmu2m/Kings.5-00/primitives.html.

2. Tiago Ferreira and Wayne Rasband, *ImageJ User Guide*, last updated October 2, 2012, https://imagej.nih.gov/ij/docs/guide/index.html.

3. John W. Tukey, *Exploratory Data Analysis* (Reading, MA: Addison-Wesley, 1977).

4. Damon Crockett, "ivpy," GitHub, accessed October 1, 2019, http://github.com/damoncrockett/ivpy.

5. Apple, "Photos," accessed July 17, 2016, http://www.apple.com/ios/photos/.

6. M. Albanese et al., "Video Summarization," in *Encyclopedia of Multimedia*, ed. B. Furht (Boston: Springer, 2006), https://doi.org/10.1007/0-387-30038-4.

7. Stuart Hall, *Encoding and Decoding in the Television Discourse* (Birmingham: Centre for Contemporary Cultural Studies, 1973).

8. Stuart Hall, "Encoding/Decoding," in *Culture, Media, Language: Working Papers in Cultural Studies, 1972–79*, ed. Stuart Hall (London: Hutchinson, 1980), 128–138.

9. Hall, "Encoding/Decoding."

10. Drake Baer, "Why Data God Jeffrey Hammerbacher Left Facebook to Found Cloudera," *Fast Company*, April 18, 2013, http://www.fastcompany.com/3008436/takeaway/why-data-god-jeffrey-hammerbacher-left-facebook-found-cloudera.

11. Nick Yee, "The Demographics, Motivations and Derived Experiences of Users of Massively Multi-User Online Graphical Environments," *Presence: Teleoperators and Virtual Environments* 15, no. 3 (2006): 309–329.

12. Ferreira and Rasband, *ImageJ User Guide*.

13. Jeffrey M. Perkel, "Life Science Technologies: This Is Your Brain: Mapping the Connectome," *Science* 339, no. 6117 (January 18, 2013): 350–352, https://doi.org/10.1126/science.339.6117.350.

14. Johannes Schindelin, Curtis T. Rueden, Mark C. Hiner, and Kevin W. Eliceiri, "The ImageJ Ecosystem: An Open Platform for Biomedical Image Analysis," *Molecular Reproduction and Development* 82, no. 7–8 (2015): 518–529, https://doi.org/10.1002/mrd.22489; Ferreira and Rasband, *ImageJ User Guide*.

15. Ferreira and Rasband, ch. 28, sec. 6 ("Stacks") in *ImageJ User Guide*.

16. Ferreira and Rasband, ch. 28, sec. 6 ("Stacks") in *ImageJ User Guide*.

17. For a very good discussion of general sampling concepts as they apply to digital humanities, see Anthony Kenny, *The Computation of Style: An Introduction to Statistics for Students of Literature and Humanities* (Oxford: Pergamon Press, 1982).

18. Jesse Alpert and Nissan Hajaj, "We Knew the Web Was Big . . . ," *Google Official Blog*, July 25, 2008, http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html.

19. Marco Brambilla, *Civilization (Megaplex)*, 2008, high-definition 3D video, https://www.marcobrambilla.com/civilization-megaplex.

20. Wiktionary, "Browse," accessed July 28, 2016, http://en.wiktionary.org/wiki/browse.

21. Wiktionary, "Explore," accessed July 28, 2016, http://en.wiktionary.org/wiki/explore.

Conclusion

1. Lev Manovich, *AI Aesthetics* (Moscow: Strelka Press, 2018).

2. The concept of "cultural techniques" has been mostly used in recent German media theory. See Geoffrey Winthrop-Young, Ilinca Irascu, and Jussi Parikka, eds., "Cultural Techniques," special issue, *Theory, Culture & Society* 30, no. 6 (November 2013).

3. Nicolas Truong and Nicolas Weill, "A Decade after His Death, French Sociologist Pierre Bourdieu Stands Tall," *Guardian*, February 21, 2012, https://www.theguardian.com/world/2012/feb/21/pierre-bourdieu-philosophy-most-quoted.

4. The term *cultural omnivore* was developed by American sociologist Richard Peterson. See Richard Peterson, "Understanding Audience Segmentation: From Elite and Mass to Omnivore and Univore," *Poetics* 21, no. 4 (1992): 243–258.

5. Minsu Park et al., "Understanding Musical Diversity via Online Social Media," in *Proceedings of the Ninth International AAAI Conference on Web and Social Media* (Oxford: AAAI Press, 2016), 308–317, http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10570.

6. See, for example, Jordan DeLong, "Horseshoes, Handgrenades, and Model Fitting: The Log-normal Distribution Is a Pretty Good Model for Shot-Length Distribution of Hollywood Films," *Digital Scholarship in the Humanities* 30, no. 1 (2015): 129–136.

7. Franco Moretti, *Atlas of the European Novel: 1800–1900* (London: Verso, 1998), 150.

8. Agustin Indaco and Lev Manovich, *Urban Social Media Inequality: Definition, Measurements, and Application*, arXiv.org, July 7, 2016, https://arxiv.org/abs/1607.01845.

9. Inna Kizhner et al., "The History and Context of the Digital Humanities in Russia," paper presented at the Digital Humanities 2018 conference, Mexico City, June 26–29, 2018, https://dh2018.adho.org/the-history-and-context-of-the-digital-humanities-in-russia.

10. Internet Archive, accessed September 10, 2019, https://archive.org.