A History of Data Visualization and Graphic Communication

Michael Friendly
Howard Wainer

HARVARD UNIVERSITY PRESS
Cambridge, Massachusetts, and London, England · 2021

Contents

Introduction

The only new thing in the world is the history you don't know.

—HARRY S. TRUMAN, quoted by David McCulloch

We live on islands surrounded by seas of data. Some call it "big data." In these seas live various species of observable phenomena. Ideas, hypotheses, explanations, and graphics also roam in the seas of data and can clarify the waters or allow unsupported species to die. These creatures thrive on visual explanation and scientific proof. Over time new varieties of graphical species arise, prompted by new problems and inner visions of the fishers in the seas of data.

Whether we're aware of this or not, data are a part of almost every area of our lives. As individuals, fitness trackers and blood sugar meters let us monitor our health. Online bank dashboards let us view our spending patterns and track financial goals. As members of society, we read stories of outbreaks of wildfires in California or extreme weather events and wonder if these are mere anomalies or conclusive evidence for climate change. A 2018 study claimed that even one alcoholic drink a day increased health risks,[1] and there is considerable debate about the health benefits or risks of green tea for lowering cholesterol, vitamin C for mitigating the common cold, marijuana for chronic pain, and (sadly) even childhood vaccination. But what do all these examples mean? As a popular t-shirt proclaims: "We are drowning in data, but thirsting for knowledge."[2]

These illustrations are really about understanding something systematic or the strength of evidence for a claim. How much does my blood sugar go up if I skip my morning run or eat a Krispy Kreme donut? Are there really more wildfires in California or more extreme weather events worldwide in recent years? Exactly how much does my health risk increase from drinking one or

two glasses of wine a day, as others had long recommended, compared with total abstinence?

For such questions, evidence can be presented in words, numbers, or pictures, and we can try to use these to evaluate the strength of a claim or argument. The purpose of scientific research is to gather information on a topic, turn that into some standard form that we can consider as evidence, and reason to a conclusion or explanation. A graph is often the most powerful means to accomplish this because it provides a visual framework for the facts being presented. It can answer the important, though often implicit, question, "compared to what?" It can convey a sense of uncertainty of evidence for the validity of a claim. Yet it also enabled viewers to think more deeply about the question raised and challenge the conclusion. A diagram can provide a visual answer to a problem and graphic displays can communicate and persuade.

As we illustrate in this book, graphs and diagrams have often played an important role in understanding complex phenomena and discovery of laws and explanations. To truly understand the impact of a visual framework, we must not only look at contemporary examples, we must also learn how it changed science and society. We must learn history.

A Long History

This book recounts a long history, a broad overview of how, where, and why the methods of data visualization, so common today, were conceived and developed. You can think of it as a guided tour of this history, focusing on social and scientific questions and a developing language of graphics that provided insights, for both discovery and communication.

This book has a long personal history as well. It began in October 1962, when we met as undergraduates at Rensselaer Polytechnic Institute. Sequentially we became math majors, house mates, and friends. We then did our graduate work at the same university (Princeton), both supported by Educational Testing Service's Psychometric Fellowship. There we came into contact with John Tukey, Princeton's widely celebrated polymath, who was in the process of revolutionizing the field of statistics with the idea that the purpose of data analysis was insight, not just numbers,[3] and that insight—seeing

the unexpected—more often came from drawing pictures than from proving theorems or deriving equations.

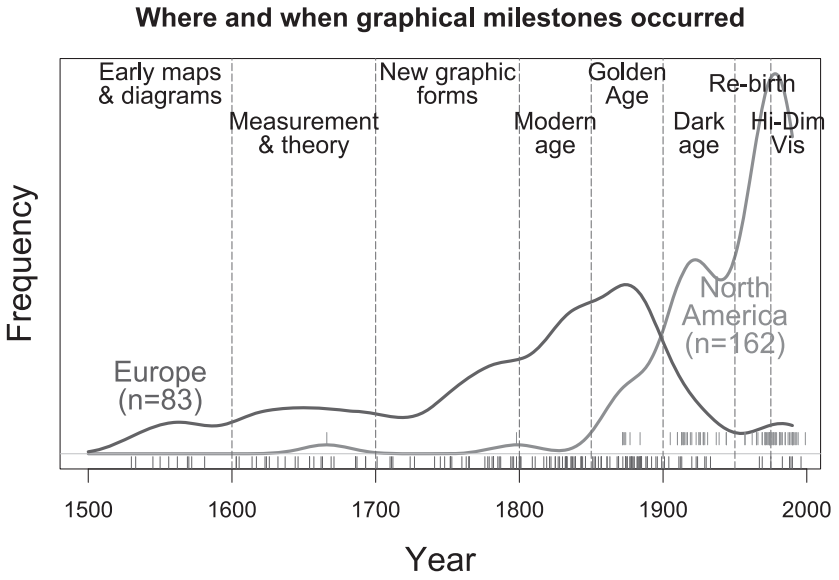
Tukey's guidance proved important and prophetic as we found that whatever substantive topic we worked on, our ability to understand and communicate the evidence we gathered almost always involved viewing the data in some graphic format. Our research led us both to gravitate toward aspects of the use and development of data visualization methods. This interest spanned their applications in scientific exploration, explanation, communication, and reasoning, as well as the creation of new methods for illuminating problems so that they can be understood better.

Remarkably, for both of us, our studies of graphical methods took us back ceaselessly into the past for a deeper and more thorough understanding. Much of what seemed commonplace today turned out to have deep historical roots.

There is also a long history of research, collaboration, and writing that informed this book and prompted this account. One initial foray was the 1976 National Science Foundation *Graphic Social Reporting Project* directed by Wainer.

One of the project's tasks was to assemble a coherent group of international scholars who worked on the use of graphics to communicate quantitative phenomena and create a social network to facilitate the sharing of information. This led to several conferences, a fair number of scholarly articles (e.g., Beniger & Robyn's 1978 history of graphics,[4] and the English translation of Bertin's iconic *Semiologie Graphique* [1973]).[5] Once republished in English, Bertin's ideas spread more broadly and became useful for the work of many other scholars, most importantly, Edward Tufte's transformative books.[6] Data visualization, as a field of study, was off to the races.

A second key event was Friendly's Milestones Project.[7] It has been substantially revised and now appears at http://www.datavis.ca/milestones/, which began in the mid-1990s. At that time, previous historical accounts of the events, ideas, and techniques that relate to modern data visualization were fragmented and scattered over many fields.[8] The Milestones Project began simply as an attempt to collate these diverse contributions into a single, comprehensive listing, organized chronologically, that contained representative images, references to original sources, and links to further discussion—a source for "one-stop shopping" on the history of data visualization. It now consists of an interactive, zoomable timeline of nearly 300 significant



I.1 **Time line of milestone events:** Classified by place of development. Tick marks at the bottom show individual events. The smoothed curves plot their relative frequency, in Europe and North America. *Source:* © The Authors.

milestone events, nearly 400 images, and 350 references to original sources, together with a Google map of authors and a milestones calendar of births, deaths, and important events in this history.

A happy, but unanticipated, consequence of organizing this history in a database was the idea that statistical and graphical methods could be used to explore, study, and describe historical issues and questions in the history of data visualization itself. This approach can be called *statistical historiography*.[9] Each item in the milestones database is tagged by date, location, and content attributes (subject area, form of the development), so it is possible to treat this history as data.[10]

For example, Figure I.1 shows the frequency distribution of 245 milestone events classified by continent. We can immediately see that most early innovations occurred in Europe, while most after 1900 occurred in North America. The bumps in the curves reflect some global historical trends that deserve explanation. The labeled time periods provide a framework of what we consider to be the major themes driving advances in data visualization.

Overview

The earliest event recorded in the Milestones Project is an 8,000-year-old map of the town of Catalhöyük, near the present Turkish city of Konya. The prehistory of visualization goes back even further. But, as you can see in Figure I.1, most of the key innovations occurred only in the last 400 years and showed exponential growth in the last 100 years.

Our central questions in this book are "How did the graphic depiction of numbers arise?" and more importantly, "Why?" What led to the key innovations in graphs and diagrams that are commonplace today? What were the circumstances or scientific problems that made visual depiction more useful than mere words and numbers? Finally, how did these graphic inventions make a difference in comprehending natural and social phenomena and communicating that understanding?

Looking over the history portrayed in the Milestones Project, it became clear that most of these key innovations occurred in connection with important scientific and social problems: How can a mariner accurately navigate at sea? How can we understand the prevalence of crime or poverty in relation to possible causal factors such as literacy? How well are passengers and goods transported on our railways and canals, and where do we need more capacity? These are among the questions that illustrate the descriptive labels we apply to the time periods in Figure I.1.

But the story of the rise of data visualization is richer than the stimulating problems. Questions like these provide the context and motivation for many graphic inventions in this history, but they don't fully answer the question "Why?" Principal innovations over the last 400 years arose in conjunction with a cognitive revolution we call "visual thinking," the idea that some problems and their solutions can be much more clearly addressed and communicated in visual displays, rather than just words or tables of numbers. Einstein, who was better known for theories of physics expressed in words and equations, captured this visual sense in his statement, "If I can't picture it, I can't understand it."

The history we relate here is exemplified in the stories of some key problems in the history of science and graphic communication, but told as an appreciation of some of the heroes in this history, for whom visual insight proved crucial. But this begs the larger question of how such visual thinking

itself developed. We provide some context for this in the initial chapters, but the essential idea is that this was bound to a concomitant rise in "empirical thinking"—the view that many scientific questions could better be addressed by gathering relevant data than by applying even the best abstract or theoretical thinking.

Re-Visions

The historical graphs we describe in this book were created using the data, methods, technology, and understanding that were current at the time. We can sometimes come to a better understanding of the intellectual, scientific, and graphical questions by attempting a reanalysis from a modern perspective.

Sometimes we come up sadly short because the software tools we have today don't allow us or make it very difficult for us to reproduce the essential ideas or the artistic beauty of important historical graphs and their stories. The hand-crafted graphs, thematic maps, and statistical diagrams of our heroes in this history often show that the pen is mightier than any software sword.

Our conscientious best efforts sometimes yield only a pale imitation of an original; in other words, we are unable to advance the understanding of the problem through reanalysis or the redrafting of graphs. One consequence is that we learn to admire the thoughtful and skillful work of our predecessors and the challenges of pen-and-ink drawings or copperplate engravings. Another consequence is that we can learn to appreciate the context of historical problems and the graphs created to present them, from both our modern successes and our failures.

We refer to these attempts as Re-Visions, meaning "to see again," possibly from a new perspective. We don't intend merely to try to see the past through present-colored glasses. Rather, we hope to shed some light on the strengths and weaknesses of the landmark developments in data visualization or understand them better in historical context. One small example illustrates this point: In Chapter 4 we show how John Snow could have made a more compelling graphic argument for cholera as a water-borne disease originating at the Broad Street pump.

Chronology versus Theme

The structure of this book requires a little explanation. In most nonfiction narratives there is considerable tension between chronology and theme, with chronology typically winning. The chronological narrative wants to move linearly from moment to moment, whereas topics scattered across eras sometimes cry out to be collected together by theme. Nevertheless chronology usually dominates, and has done so at least since narratives were recorded on papyrus scrolls.

In this book chronology dominates, but we tried to hold its force in check, fearing that if we didn't, the reader would be thematically left at sea, with the next instance far off on some foreign shore. The great themes of epistemology, scientific discovery, social reform, technology, and visual perception move with time, but not in lockstep. Consequently, much of our narrative is structured around key problems of a given time and the individuals—our graphic heroes—whose visual insight and innovations led to advances in data visualization and science.

What follows is a synopsis of the book.

Chapter 1, "In the Beginning . . . ," is an overview of the larger questions and themes that provide a context for the book. We consider the relations among numerical data and evidence for an argument and graphs, and then describe some of the prehistory of the visual representation of numbers and the early rise of visualization itself. The story continues to the rise of empirical thinking in philosophy and science around the sixteenth century and the concomitant remarkable development of the visual representation of numbers to communicate quantitative phenomena.

From there we explore a fundamental and difficult problem of the seventeenth century: the determination of longitude at sea. In Chapter 2, "The First Graph Got It Right," we show how Michael Florent van Langren had the idea to make a graph of historical determinations of the longitude distance from Toledo to Rome, in what is arguably the first graph of statistical data.

In Chapter 3, "The Birth of Data," we trace the role of data in the initial rise of graphical methods around the early 1800s. We focus attention on one important participant in this story: André-Michel Guerry [1802–1866], who

used an "avalanche of data" and graphical methods to help invent modern social science.

A short time later, analogous widespread data collection began in the United Kingdom, but this was in the context of social welfare, poverty, public health, and sanitation. In Chapter 4, "Vital Statistics," we see two new heroes of data visualization, William Farr and John Snow, who worked independently trying to understand the causes of several epidemics of cholera and how the disease could be mitigated.

Chapter 5, "The Big Bang," details how, at the beginning of the nineteenth century, nearly all the modern forms of data graphics—the pie chart, the line graph of a time series, and the bar chart—were invented. These key developments were all due to a wily Scot named William Playfair. He can rightly be called the father of modern graphical methods, and it is only a slight stretch to consider his contributions to be the Big Bang of data graphics.

Among all the modern forms of statistical graphics, the scatterplot may be considered the most versatile and generally useful invention in the entire history of statistical graphics. It is also notable because William Playfair didn't invent it. Chapter 6, "The Origin and Development of the Scatterplot," considers why Playfair was unable to think about such things, and it traces the invention of the scatterplot to the eminent astronomer John F. W. Herschel. Scatterplots achieved great importance in the work of Francis Galton [1822–1911] on the heritability of traits. Galton's work, visualized through statistical diagrams, became the source of the statistical ideas of correlation and regression and thus most of modern statistical methods.

In the latter half of the nineteenth century, enthusiasm for graphical methods matured and a variety of developments in statistics, data collection and technology combined to produce a "perfect storm" for data graphics. The result was a qualitatively distinct period that produced works of unparalleled beauty and scope, the likes of which would be hard to duplicate today. In Chapter 7 we argue, as the chapter title implies, that this period deserves to be recognized as the "Golden Age of Statistical Graphics."

Chapter 8, "Escaping Flatland," discusses the challenges of creating displays of data. Displays are necessarily produced on a two-dimensional surface—paper or screen. Yet these are often misleading at worst or incomplete at best. The representation of multidimensional phenomena on a two-dimensional surface was, and remains, the greatest challenge of graphics. In this chapter we

discuss and illustrate some of the approaches that were used to communicate multidimensional phenomena within the existing limitations.

Chapter 9, "Visualizing Time and Space," explores two general topics in the recent history of data visualization. First, graphical methods have become increasingly dynamic and interactive, capable of showing changes over time by animation and going beyond a static image to one that a viewer can directly manipulate, zoom, or query. Second, the escape from flatland has continued, with a variety of new approaches to understanding data in ever higher dimensions.

Graphs are justly celebrated for their ability to accurately present phenomena in a compact way while simultaneously providing their context. If this were all that they did, their place in scientific history would be secure. But with suitable data and the right design, they can also convey emotion. Indeed, in some instances graphs provide an emotional impact that can be likened to that of poetry. In Chapter 10, "Graphs as Poetry," we imagine a collaboration between the civil rights activist W. E. B. DuBois and the canonized graphic designer C. J. Minard to depict the Great Migration of 6 million African-Americans fleeing the racism and terror in the post-Confederacy South to the industrial North. The result of this *gedanken* collaboration provides a vivid example of how we can profit from studying the past to help solve the problems of the future. A final section, "Learning More," lists additional resources for those who wish to explore a topic in greater depth.

This print edition necessarily omits some materials that enrich our stories but fell to the cutting-room floor. Moreover, publishing constraints limited the number of color images. To partially compensate, we created an associated web site, http://HistDataVis.datavis.ca, containing all images in color, some of our more extended discussion, and biographical notes on some of our dramatis personae in this history. A happy consequence is that we can continue to keep this topic active with additional essays on related topics.

Thus, this book invites you to consider the history of data visualization from a larger perspective: a journey that began with the earliest visual inscriptions and progressed to social and scientific problems that could be understood in graphs and diagrams. Along this path, many innovations were forgotten or underappreciated, as Harry Truman noted in the opening quote. The following chapters highlight contributions that are imperative to the history of visual thinking and graphic communication.