**Chapter 10**

**The Milestones Project: A Database for the History of Data Visualization**

**Michael Friendly, Matthew Sigal, and Derek Harnanansingh**

If you would understand anything, observe its beginning and its development.—Aristotle **[QY: please provide source]**

**Introduction**

Questions regarding the history of data visualization are (or at least should be) of great importance to historians of science, to current developers of graphical methods for statistical analysis and the related info-vis community, as well those just interested in the history of ideas. In the history of science, diagrams, graphs, maps, and other visualizations have often played important roles in discoveries that arguably might not have been achieved otherwise.[[1]](#footnote-1) At the same time, in the fields of statistical graphics and information visualization, developers often create “new” methods without any appreciation that they have deep roots in the past.[[2]](#footnote-2)

These two perspectives provided the motivation for the development of the Milestones Project. This stemmed from the fact that historical accounts of events, ideas, and techniques that relate inter alia to modern data visualization were fragmented and scattered across a wide number of fields.[[3]](#footnote-3) When this work began in the mid-1990s, there were no accounts or resources that spanned the entire development of visual thinking and the visual representation of data across different disciplines and perspectives. 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.

In Section 2, we describe the evolution and structure of the Milestones Project. Section 3 presents some historical and modern approaches to one self-referential question: how can data visualization be applied to its own history? Section 4 introduces another self-referential topic we call statistical historiography, which entails the use of statistical and graphical methods for the analysis and understanding of historical innovations, developments, and trends. But first we give some brief vignettes of historical topics and questions for which the Milestones Project has proved invaluable in our own research.

*The First Statistical Graph*

In the history of statistical graphics, as in other artful sciences, a number of inventions and developments can be considered “firsts” in these fields (Friendly, “A Brief History of Data Visualization”). The catalog of the Milestones Project lists 70 events that can be considered to be the initial use or statement of an idea, method, or technique that is now commonplace, but there is probably no question more fundamental than that of the first visual representation of statistical data (Friendly and Denis, *Milestones*).

In Friendly et al., we argue that the one-dimensional line graph shown in Figure 10.1 by Michael Florent van Langen should be accorded this honour. The graph shows 12 estimates of the distance in longitude between Toledo and Rome, overlaid on a modern map. Toledo is located at lat/long (+39.86◦N, −4.03◦W), Rome is located at (+41.89◦N, +12.5◦W), both shown by markers (stars) on the map. This image makes clear what van Langren wished to communicate: the wide variability of the estimates, but also how far the estimates were biased. Van Langren used this to demonstrate that these estimates were all subject to large errors and to propose to King Phillip of Spain that only he had a sufficiently precise method for the determination of longitude for navigation at sea. The telling of van Langren’s story not only turned out to involve astronomy, archival research, the history of patronage in the seventeenth century, and even an unsolved problem of cryptography, but also serves as an example of statistical historiography. The Milestones Project provided the infrastructure for this research: through the use of a time-based, cross-referenced catalog of images, references, and links to related work, van Langren’s tale was able to be studied and reported upon.

**[Insert Figure 10.1—portrait]**

**Figure 10.1. Van Langren’s 1644 graph, re-scaled and overlaid on a modern map of Europe.**

*Who Invented the Scatterplot?*

Although there are earlier precursors, the main graphical methods used today— pie charts, line graphs and bar charts—are generally attributed to William Playfair in works around the beginning of the nineteenth century. All of these are essentially univariate displays of some aspect of a single variable. A logical next step would be to invent a method to reveal the relationship between two variables— what we now know as the scatterplot. By 1886, Francis Galton had utilized this truly bivariate display, which led to the discovery of correlation and regression, and ultimately to much of present multivariate statistics. However, he was not the first to use this graphical technique, and it is surprising that no one is widely credited with its invention.

In “The early origins and development of the scatterplot,” we delved into this mystery. This involved tracing the early origins of ideas related to the scatterplot, which led to two compelling narratives: how, in Playfair’s time, it was nearly impossible to think about and visualize bivariate relationships; and, later, how the scatterplot was essential for Galton’s visual insights that would lead to the rise of modern statistics and graphics. The resources available in the Milestones Project allowed us to focus upon the events in this period and attribute the essential ideas of the scatterplot to J. F. W. Herschel and two 1832 papers.

*The Golden Age of Statistical Graphics*

In our initial web presentation of the Milestones Project, it proved convenient to sub-divide the history of data visualization into epochs, each of which turned out to be describable by coherent themes. As we illustrate later, one period turned out to be particularly noteworthy, both for the sheer number of contributions, and for the beauty and elegance of their execution. We call this period, from roughly 1850 to 1900 (±10), the Golden Age of statistical graphics (Friendly, “Golden Age”).

Figure 10.2 shows the time distribution of the 260 significant events that had been included in the Milestones Project database by 2007, demarcated by the labels we used for epochs. The density estimate is based on n = 260 significant events in the history of data visualization from 1500–present. The developments in the highlighted period, from roughly 1840 to 1910, comprise the Golden Age of statistical graphics. We traced the origin of this period in terms of the infrastructure required to produce such an explosive growth of contributions to data visualization, and found three primary sources: the systematic data collection by state agencies, the rise in popularity of statistical and visual thinking, and the enabling developments of technological innovations (see Friendly, “Golden Age”).

**[Insert Figure 10.2—portrait]**

**Figure 10.2. The time distribution of events considered milestones in the history of data visualization, shown by a rug plot and density estimate.**

**The Milestones Project**

Direction is more important than speed. We are so busy looking at our speedometers that we forget the milestone.—Anonymous

An early overview of the content and aims of the Milestones Project appeared in Friendly’s “Milestones in the history of data visualization.” Here we update that description and provide a few technical details on some problems that were encountered in attempting to make the history of data visualization convenient for collecting, browsing, searching, and analysis.

*Origin, Structure, and Evolution*

The initial step in portraying the history of data visualization was to create a simple chronological listing of milestone items with capsule descriptions; bibliographic references; markers for date, person, and place; and links to portraits, images, related sources, and more detailed commentaries. The initial database contained the 105 developments listed by Beniger and Robyn in 1978, then incorporated additional records from Hankins, Tufte (see *Visual Display, Envisioning,* and *Visual Explanations*), and Heiser, among others.

[QY: Can you provide page numbers for Heiser bibliography entry? Heiser, W. J. “Early Roots of Statistical Modelling.” Social Science Methodology in the New Millennium: Proceedings of the Fifth International Conference on Logic and Methodology. Eds. J. Blasius, J. Hox, E. de Leeuw, and P. Schmidt. Amsterdam: TT–Publikaties, 2000. 000–000.]

This database began as a single LATEX file (with markup tags for all relevant bits of information), used to produce a hyper-linked PDF document. A variety of software tools (perl scripts, Unix utilities) allowed us to turn this single source directly into the web version.[[4]](#footnote-4) Other custom software tools allowed us to add new milestones items from text files using a template of tags (DATE:, AUTHOR:, WHAT:, REF:, IMG:, etc.) and to extract the information about milestones items, authors, images, and the like in a variety of forms (CSV, XML, JSON) that could be used as input for analyses and graphic displays. For example, Figure 10.2 was produced in SAS software by piping the output of a latex to csv translator:

itemdb -o milestones.csv < milestones.tex | sas -i milestones.csv mileyears.sas

It soon became apparent that such a text-based representation was inadequate. Updating the milestones data required that the LATEX file be shared among several collaborators; milestone assets, such as images, web links, and references, were not easily accessible by others, which made collaboration cumbersome. Further, each update to the web site required an inefficient number of steps of verification, re-building, and synchronization with the server, meaning the website was often out of date. Around 2005 we began to make the process more dynamic, convert the flat file into a relational database, create a Milestones administrative system, and completely redesign the Milestones web site. Specifically, we wanted to facilitate contributions by any number of trusted collaborators via an easy-to-use web administration area, and allow for the dissemination of milestones data via an easy-to-browse public user interface.

Migrating the data to this format posed some challenges. First, the existing milestones data needed to be restructured logically and have redundancy minimized. To do this, we partitioned the data into its relevant entities: namely the milestone itself, and its descriptors, such as its aspect, author, subject, keywords, reference, and linked media items (such as images). The aspect, author, subject, keyword, and reference descriptors exist as a many-to-many relationship between it and the milestone. The main table (milestone) contains information regarding each of the items considered a milestone in the history of data visualization, linked to other tables (for example, reference, mediaitem) by unique (primary) keys. Other supporting tables (for example, milestone2aspect) provide for convenient lookups of descriptors of these milestones items (subject, aspect, keyword).]For example, an aspect can belong to one or more milestones, and the milestone can belong to one or more aspects. Media items, on the other hand, can only belong to one milestone at a time, with multiple media items possible for a single milestone. Figure 10.3 illustrates these relationships.

**[Insert Figure 10.3—portrait]**

**Figure 10.3. Simplified schema for the MySQL database for the Milestones Project.**

Normalizing the data in this way enabled us to free the database of modification anomalies and ensured that the database structure was scalable and could be extended with minimum modifications. Most importantly, it allows for future growth, and provides a query-neutral database model that could be used to power web presentation and customized indexed search (Codd). The last major benefit, which will be demonstrated in Section 3, is that this schema allows for any type of analysis of the Milestones data itself.

At present, the Milestones Project documents 288 contributions, with nearly 350 references, information on 336 authors, and 774 media items, made up of 371 images appearing online on the milestone site, and 403 hyperlinks to images and documents that are externally hosted. In addition, we maintain an offline image database comprising over 1,100 images collected from various sources. Over time, these too will be incorporated into the database.

*User Interface*

The second challenge related to how to display such a large amount of information in an easy-to-use interface that would provide overview, search, and details about these events in the history of data visualization. We decided to retain the time-based grouping of the milestones content by epochs (Pre- 1600, 1600s, 1700s, etc.), each with a theme (for example, 1600–1699: Measurement and Theory) and descriptive text. The visual design of the interface adopts Ben Shneiderman’s mantra: “Overview first, zoom and filter, then details on demand” (Could you provide page number?). To do this, we added a timeline view of the milestones items displayed on the overview landing page (see Figure 10.4).

**[Insert figure 4—portrait]**

**Figure 4. Timeline view of the Milestones Project.**

This timeline, based on the SIMILE Timeline Widget, allows multiple connected time bands, showing events at different resolutions. Each band can be separately panned by dragging left or right with the mouse pointer, scroll wheel, or keyboard arrow keys. In this view, the top panel shows a detailed view of the segment of history highlighted in the bottom panel, both of which can be separately scrolled. Items in the top panel show a brief text tag, colour-coded by category. Clicking on an item in this panel brings up a small description, which is further linked to the details of the milestone item.The timeline view, although most obvious, is just one of several possibilities for a visual overview or interaction with the display of the milestones database. The software design of the site, using open-source tool kits, makes it relatively simple to add new ones. For example, the database can also be navigated via a list view (with drop down quick links), and in Section 4.3 we will illustrate how it can be explored using a map-based display.

**Visualizing Time and History**

What does history look like? How do you draw time?—Rosenberg and Grafton (10)

The questions in this quotation introduce an important topic in the history of data visualization: how can such a history be visualized? What methods might be called upon to detail the richness of its past?[[5]](#footnote-5) Time provides an obvious dimension, but what else could be included in a static display that might reveal a story previously hidden? What kinds of dynamic or interactive displays might fascinate and intrigue viewers?

An annotated visual gallery of some timeline designs and visual histories can be found in our Data Visualization Gallery. The topics covered include early visual histories, encyclopedic charts, special purpose charts, correlated histories showing events in one domain in the context of events in other areas, non-linear scales for time and space, as well as dynamic, interactive timelines. Here we present a few inventive selections from this scholarship.

*The First Timelines, Reconsidered*

Although there are earlier precursors, the first timelines of modern design— featuring a horizontal, linear axis for time, and vertical positions for place, theme, or category of events—were produced in the mid 1700s. Most notable of these prototypes were Jacques Barbeau-Doubourg’s 1753 Carte Chronologique and Joseph Priestley’s 1765 Chart of Biography. Priestley first published a small “Specimen” of this chart as a proof-of-concept, showing the lifespan of famous men in the years 600 BC to 0 AD, classified as “statesmen” (from Solon to Augustus) and “men of learning” (from Thales to Ovid). Later that year, in 1765, Priestly published a detailed version that quickly became the most popular and influential timeline of the nineteenth century. The full graphic details the lifespans of more than 2,000 people from 1200 BC to 1750 AD, classified by their areas of achievement (statesmen & warriors, mathematicians & physicians, artists & poets, and so on). Priestly’s timeline charts can be seen on our Data Visualization Gallery, and we don’t reproduce them here. Instead we show (see Figure 10.5) a re-design, in his style, of the lifespans of 79 authors from the Milestones database who were born in France or the United Kingdom between 1500 and 2000. Rosenberg and Grafton called Priestley’s charts “masterpieces of visual economy” (117). Indeed, they were at the time. However, in his charts, the famous people were arranged haphazardly within category groups, so it is difficult to find specific individuals, and nearly impossible to uncover any trends, either over time or across categories.

In our version, authors are sorted by birth year within each country and the names are printed alternately at the year of birth and death. Authors are sorted within country by year of birth and labeled alternately at birth and death years, allowing better lookup and visual comparison.The result, which resembles a cumulative distribution plot, (a) allows easier visual lookup of names, (b) provides an overall “lifespan envelope,” and, (c) highlights a few individuals who lived conspicuously longer or shorter lives (for example, Willam Jevons, James Maxwell, John Snow, and Phillipe Buache) than their contemporaries. Of course, to display lifespan directly requires a different kind of plot, but one that would not have been even thinkable by Priestley in 1765. We return to this question later on.

**[Insert figure 10.5—portrait]**

**Figure 10.5. A modern re-design of Priestley’s 1765 Chart of Biography, using information on authors in the Milestones database born in France or the United Kingdom.**

*Universal Histories*

In addition to unrivalled thematic maps and statistical diagrams, the Golden Age of statistical graphics also gave rise to a variety of novel attempts to visualize history in a comprehensive manner, combining parallel, intertwined time-flows, text, illustrations, maps, and other visual forms. Among the most impressive is the series of Synchronological Charts of Universal History produced by Sebastian Adams between 1871 and 1885. The 1881 version is 23 feet long and captures 5,885 years of history, from 4004 B.C. to 1881 A.D. Rosenberg and Grafton call it “nineteenth-century America’s surpassing achievement in complexity and synthetic power” (172). Figure 10.6 shows the entire chart at the top (note the increasing visual density towards the right) and a small portion below.[[6]](#footnote-6) Adams used a linear scale for time, and so it is understandable why it took 23 linear feet to include all of recorded history.

[Insert Figure 10.6—portrait]

Figure 10.6. Top: the entirety of Sebastian Adams’ Synchronological Chart of Universal History, 1881. Bottom: an excerpt detailing the 600–1100 AD period. Horizontal bands trace developments in different countries, with detailed text describing significant events, and break up or merge according to political factors.

*Categorization and Non-linear Scales*

Linear time scales have the advantage that they provide uniform resolution and detail across the entire time span, but events in time, or our interest in them, are rarely uniformly distributed. As exemplified by the Milestones Project, most visual histories are rather sparse at their beginning and very crowded at their end. Utilizing non-linear scales can allow resolution to vary smoothly across the range, providing greater detail in regions of interest, which are most often the recent past.[[7]](#footnote-7)

Figure 10.7 is a proof-of-concept sketch for something that a graphic artist could use as a starting point for a chart of the history of data visualization. It uses the events from the Milestones Project, categorized by two correlated factors: Subject area, in which the content has been categorized as dealing with human populations, physical properties of the world, or mathematics and statistics; and, the milestone’s aspect or form, which has been categorized as dealing with cartography, graphs and diagrams, or technology. To provide greater resolution for more recent events, time (Year) is shown on a square-root scale, going backward from the year 2000. To provide greater resolution for more recent events, we have used a reverse square-root scale going backward from the year 2000. Specifically, “Year” on the horizontal time axis is actually plotted according to the formula , giving the more pleasing result that the modern period 1800–2000 occupies about 60% of the scale, despite only comprising 40% of the range. This is conveyed visually by the spacing between tick marks on the X-axis.

**[Insert Figure 10.7—portrait]**

**Figure 10.7. Sketch for a thematic timeline of milestones items, 1500–present, categorized by both the Subject (content) and Aspect (form) of the milestone item.**

**Using the Milestones Project for Statistical Historiography**

Vision is the art of seeing things invisible.—Jonathan Swift, 1711 **[QY: source?]**

*Statistical Historiography*

We use the term “statistical historiography” to refer to the use of statistical and graphical methods to explore, study and describe historical problems and questions.[[8]](#footnote-8) This topic has a delightful self-referential quality when applied to the history of data visualization itself, since we have often found ourselves using modern methods of statistical analysis and graphics to study the development of ideas in this area. As in the Swift quotation above, one goal is to make previously hidden aspects of this history visible.

At the same time, our examination of some of the most impressive graphic works of the past sometimes left us awe-struck by their exquisite beauty and visual design.[[9]](#footnote-9) On more than one occasion when looking at these elegant presentations, we wondered whether there wasn’t something lost with the advent of modern software. While we can now analyze massive data sets, and generate a multitude of graphics with a simple mouse click, we still feel that designing a truly effective visual display of information requires thought and manual intervention.

For this reason, it is often quite instructive to attempt to re-create or even re-vision a graphic work from the past (Friendly, “Visions”). We can learn from this undertaking an appreciation for the insight and hard labor of our graphic heroes, and can sometimes better understand or improve on their designs by a process we call “understanding through reproduction,” another facet of statistical historiography.

There is, of course, one principal requirement for statistical historiography: *data*. The Milestones Project database is the repository of all the information we have so far recorded, and modern database tools allow the possibility of simple or complex queries, limited only by the available information.[[10]](#footnote-10) In related work, we have collected and disseminated data sets of historical interest on a variety of topics in statistics and data visualization, for instance via the R packages HistData and Guerry (Friendly, *HistData*; Friendly and Dray, “Guerry”). These can be considered another source for data, pictures, and stories related to statistical historiography, and understanding through reproduction. This is the essence of the motto on the datavis.ca web site: Looking back, going forward.

In the subsections below, we describe a few applications of these ideas using the Milestones Project database and case studies that arose from this work. There is an interesting interplay between such historical analyses and these data collections. Some studies called for us to find and incorporate new data sources, such as our paper on Guerry’s moral statistics of France and the Guerry package, to which we added Angeville’s extensive 1836 data on social and economic characteristics of France (Friendly, “Visualizing”). In other cases, our analyses suggested new or different ways to visualize historical data.

*Milestone Authors: Lifespan*

As noted earlier, we record information relevant to the contributors of milestones events in an author table in the database. For most of these individuals, internet and biographical searches allowed us to determine the dates and places of their birth and death. One simple question that can be posed using this information is how long did these contributors live? As illustrated earlier in Figure 10.5, Joseph Priestley was the first to develop the idea of using a graphic representation to show the lifespan of famous men. His “charts of biography” did this in a particularly evocative form, representing each person by a line segment whose length was defined by the individual’s lifespan and then grouped by occupational category.

These “timespan” charts tell an interesting story, but they do not provide an answer to the question of how long, in general, these individuals lived. However, with the author table from the Milestones Project, it is a simple matter to calculate lifespan and to obtain a direct answer to this query.

Figure 10.8 shows one display of this information, using a combined density plot and rug plot, similar to the one used in Figure 10.2. Individual observations are shown by a (jittered) rug plot, and the three extremes on each end are identified by name. The red vertical lines show the quartiles of the distribution.Several features of this plot deserve comment, and also invite further inquiry: Most notable is that, by and large, milestones authors generally lived to a ripe old age— the median lifespan is 73.0, but the density plot peaks at around 79. This contrasts with a detailed study on famous people between 2400 BC to 1880 AD by David de la Croix and Omar Licandro. In their research, the typical lifespan fluctuated around a mean of 61 years for four millennia; it only gradually reached 69 toward the end of their sample. Such a discrepancy between the two studies might warrant further investigation—for instance, by classifying the individuals into occupational, locational, or otherwise more delineated groups, and looking for trends. Another interesting feature that becomes apparent in this graphic is the noticeable bump in the distribution around 45 years. This occurrence calls for some attempt at further explanation. We don’t pursue this here, but again note that such graphs often suggest further analyses (breakdowns by region or time period), or cry out for the collection of more data.

**[Insert Figure 10.8—portrait]**

**Figure 10.8. Density plot of the lifespan of the 172 authors in the Milestones Project database who were born after 1500 and for whom lifespan can be determined.**

Finally, although Figure 10.8 is just a summary graph, we have labeled a few extreme observations on each end, which may relate to telling parts of the story of the history of data visualization. Among these, Henry Moseley, who is known for the discovery of atomic number from a graphical display, died the youngest, as a consequence of serving in the British Army during World War I. But, we were surprised to see the noted and prolific French cartographer Phillippe Buache, and the German physicist and astronomer Johann Tobias Mayer, show up in positions two and three. On the other end, we were delighted to see that Adriano Balbi, a Venetian geographer and early collaborator of André-Michel Guerry had the longest lifespan, just exceeding the population geneticist, Sewall Wright, who invented path analysis and the path diagram around 1920 (Balbi and Guerry). By incorporating these details, the visualization is able to reveal narratives that otherwise would have been concealed.

*Milestone Authors: Geography*

The Milestones Project web site provides an initial page showing an interactive timeline of the events in this history as a visual overview (Figure 10.2). A long-term goal has been to provide other views of this history and other tools for searching and exploring the database. With recent technological developments, it became evident to us that one such method would be through the use of geographical data.

So far, the primary geographic information we have encoded in the database refers to the birth and death place of the milestone authors. This is an imperfect representation, as these locations may not accurately represent the author’s primary residence. For instance, Charles Joseph Minard was born in Dijon, and died in Bordeaux, but all of his work was done in Paris while he worked at the École Nationale des Ponts et Chaussées. Nevertheless, a geographic view of the available information is potentially useful. In this regard, we used the Google geocoding tools to provide latitude and longitude for the locations listed in the author table. Using this and the R package googleVis created by Gesmann and de Castillo, we developed the interactive map shown in Figure 10.9.

Like other instances of Google Maps, this graphic can be panned and zoomed using mouse controls. The place markers display tool tips when hovered over and, when clicked, link to a search page that details all of the Milestone items that are related to that author. This interesting visualization will soon be revealed on the Milestones Project website, with future work planned to incorporate other types of data in addition to the birth and death locations.

**[Insert Figure 10.9—portrait]**

**Figure 10.9. Birthplaces of 188 milestone authors, shown on an interactive Google map, centered on France. Each geographic marker is linked to an author query on the datavis.ca web site that lists the contributions by that individual.**

*Milestones: Themes and Trends*

The records in the Milestones Project database also feature various text fields for each logged event. These include a brief item tag, a full description of the event, and relevant keywords, as well as categorical codes for the content (Subject), and form (Aspect) of the item. Treating this information as “data” allows us and others to study themes and trends in these developments. Modern methods of text mining and data visualization can provide insights into this history not available through other means.

As one simple illustration of this approach, Figure 10.10 shows two mosaic displays[[11]](#footnote-11) that explore the relationships among Epoch, Subject, and Aspect. The left panel shows changes in the distributions of milestone events by Subject over time. It can readily be seen that while most of the milestone innovations up to the end of the eighteenth century were about the physical world (astronomy, geodetic measurement, weather, etc.), this trend changed in the nineteenth century, where there was a large shift toward problems that related to human populations (for example, pertaining to mortality, births, disease, crime). Beginning in the early 1900s, the pattern changes again, with advances in mathematics and statistics becoming the dominating force.

The right panel shows the association between Subject and Aspect, pooled over Epoch. As is not surprising, maps and other cartographical representations were most often used to show data of the physical world, while graphs and diagrams were most often associated with mathematical and statistical subjects. Other statistical graphs and analyses could be used to explore these and other relationships in more detail. The key to this is of course the existence and availability of data—in this case reflected by the coding of graphical milestones in our database.

**[Insert Figure 10.10—portrait]**

**Figure 10.10. Mosaic displays for milestone items, classified by Epoch, Subject and Aspect. Left: mosaic for the marginal table showing differences in Subject across Epochs.**

**Right: mosaic for the marginal table showing differences in Subject across Aspect. Numbers in the tiles give the number of milestone items.**

**Conclusion and Future Directions**

The Milestones Project began as a simple attempt to collect a comprehensive history of innovations and developments in data visualization in a single location. Like Topsy, it grew over time, with images, historical papers and references, suggestions, and other contributions graciously provided by friends and collaborators, most notably from the members of Les Chevaliers des Albums de Statistique Graphique.

In this chapter, our primary goal was to introduce the second and latest iteration of this project. The redesign was undertaken to make this history more accessible for browsing and searching, and to attempt to make the database more amenable to additions, edits, and extensions among collaborators. However, we find that the most exciting aspect of the new structure is its flexibility in terms of data retrieval, and our newfound ability to use and manipulate the data for graphic-based statistical historiography.

One goal for the future, as we suggested earlier, is to extend the user interface to provide multiple views and advanced text search and filtering capabilities. One convenient path for this development is provided by the SIMILE Exhibit framework (Karger, Smith, et al). This provides web software libraries (Ajax, javascript, css) for timelines, interactive maps, tabular displays, image “tiles” and other visualizations. Various views can be composed for browsing as tabbed, alternatives or as faceted displays, showing for example an interactive timeline and a map.

Equally important, the Exhibit framework allows us to present some of the milestones tables to be used as filters for the items displayed in these views. Tables for subject, aspect, keywords, location, epoch, and so on would allow the user to select milestone events based on some or all of these criteria, providing a way to ask such questions as “what milestones events between 1700-1900 involving social science occurred in Europe?”

Finally, we would like to make the milestones database more publicly accessible for use by others on the history of data visualization. For the examples we have shown here, we connect to the milestones database directly via MySQL or ODBC interfaces to SAS and R, but this presents security risks. Happily, the Exhibit framework also provides methods for data export from various views, using JSON or CSV formats. In addition, we contemplate adding facilities for users in the data visualization community to add comments, notes, references and links to milestones items. These extensions will comprise the Milestones Project 3.0.

Acknowledgements

This work was supported by Grant OGP0138748 from the National Sciences and Engineering Research Council of Canada to Michael Friendly. We are grateful to Dan Denis, Stephen Stigler, and Ben Shneiderman for constructive comments on this chapter.

1. Some salient examples are: Francis Galton’s 1861 discovery of anti-cyclonic movement of wind around low-pressure areas from contour maps; Edward Maunder’s “butterfly diagram” of the variation of sunspots over time leading to the discovery of the “Maunder minimum,” from 1645–1715; and Henry Moselely’s 1913 discovery of the concept of atomic number, based largely on graphical analysis (a plot of serial numbers of the elements vs. square root of frequencies from their X-ray spectra). [↑](#footnote-ref-1)
2. For example, mosaic displays for frequency tables were thought to have been invented by Hartigan and Kleiner and extended to show the pattern of residuals in loglinear models by Friendly (“Mosaic Displays”). But it turns out that the essential idea behind this area-based display goes back to Georg von Mayr in 1877 (Friendly, “A Brief History”). [↑](#footnote-ref-2)
3. Among these are general histories in the fields of probability (Hald), statistics (Pearson; Porter; Stigler, *History of Statistics*), astronomy (Riddell), cartography (Wallis and Robinson). More specialized accounts focus on the early history of graphic recording (Hoff and Geddes, “Graphic Recording,” “Beginnings”), statistical graphs (Funkhouser, “Note,” “Historical Development”; Royston; Tilling), fitting equations to empirical data (Farebrother), cartography (Friis; Kruskal, “Visions of Maps”), thematic mapping (Friendly and Palsky; Palsky, *Des Chiffres*; Robinson, *Early Thematic Mapping*), and so forth. [↑](#footnote-ref-3)
4. The original web version was originally shown at http://www.math.yorku. ca/SCS/Gallery/milestone. [↑](#footnote-ref-4)
5. Another recent book, *Visualizing Time* (Wills), discusses a range of modern graphical methods for visualizing time-based data. [↑](#footnote-ref-5)
6. The entire chart can be viewed in high-resolution at http://www.davidrumsey. com/blog/2012/3/28/timeline-maps [↑](#footnote-ref-6)
7. Of course, interactive graphics offer the possibility to vary resolution dynamically, by moving a “lens" across the display, as in a hyperbolic viewer. [↑](#footnote-ref-7)
8. As far as we know, the initial expression of this idea appeared in a paper by Rubin discussing various ways in which statistical methods could be applied to historical topics. These included: the use of sampling methods to test historical theories; statistical distributions applied to historical data; and, the use of time series graphs with smoothed curves to study historical trends. More recently, many examples of the application of these ideas to statistical topics can be found in Stigler (*History of Statistics*, *Statistics on the Table*), as well as our own papers on the history of data visualization, cited inter alia. [↑](#footnote-ref-8)
9. Some examples are Charles Joseph Minard’s famous depiction of Napoleon’s March on Moscow (Friendly, “Visions and Re-Visions”), Francis Galton’s detailed study of weather patterns in Europe (see: Friendly, “Golden Age”), and Andr´e-Michel Guerry’s (Guerry, Plate 17) semi-graphic table depicting the relations of occurrence of crimes to a wide variety of social and demographic factors (see Friendly, “A.-M Guerry’s”). [↑](#footnote-ref-9)
10. It should be noted that, beyond the basics of recording milestones items, images and references, inputting the other meta-data (content and form categories, keywords, etc.) is highly labor-intensive. Thanks are due to many research assistants and graduate students who have and continue to work on the Milestones Project, including Dan Denis, Matt Dubins, Yvonne Lai, Avi Lipton, and Carolina Patryluk. [↑](#footnote-ref-10)
11. Mosaic displays show the frequencies in cells of a cross-classified table by the area of each tile. The tiles are shaded according to departure from a null model of no-association, using blue for cells with frequencies substantially greater than chance, and red for cells with frequencies that are lower than expected. [↑](#footnote-ref-11)