Web-based Data Visualization:

Creating Effective Displays of Information

for

Master of Science

Information and Communications Technology

Daniel Feldman

University of Denver University College

May 28, 2017

Faculty: Janice Minder, Ed.D.

Director: Thomas Tierney, PhD

Dean: Michael J. McGuire, MLS

Abstract

Designers and developers struggle with finding effective workflow solutions when producing web-based data visualizations. Conceptual design problems surface in transforming raw data into effective visual messages. Technical development issues arise in choosing tools and frameworks to construct and distribute work. By examining the role of design guidelines and best practices to shape effective work, and assessing the benefits of leading data visualization production platforms, workflow optimization is possible. Solving these problems can save time and effort in production, and create improved work. This outcome benefits designers and developers, project stakeholders, and consumers. This research explains how the application of design best practices and guidelines, and choosing a web-standards framework for development and distribution can create effective and portable data visualizations.

Contents

Background

In the Information Age, there is an unprecedented abundance of data to inform us about our lives and business. As technology advances, our digital footprint grows exponentially larger, and with it the appetite for understanding how to capitalize on the wealth of data that surrounds us. In facing this profusion of data, the human brain is inefficient at processing text, yet hard-wired to quickly perceive visual information. Interpreting the meaning of data then largely relies on data visualization (Chen 2017d, 8-9).

Data visualizations are graphic artifacts that intuitively summarize key information from raw data. These charts, graphs, and maps let users quickly see the underlying messages and stories hidden in raw data. Effective data visualizations offer expansive insights into the vast data stores around us. This in turn enables us to make better inferences and decisions about our lives and business (Kirk 2012, 5-6).

Making use of these powerful and plentiful resources of data, however, is challenging. The long-established models of static visualization have changed drastically with dynamic web-based technologies. Designers and developers who produce data visualizations struggle with many facets of the design process. Knowing how to analyze and manipulate raw data to communicate key insights, and finding the right tools to disseminate work are difficult problems to solve. Understanding this production process is essential to create effective work.

Data visualization occupies an active space at the crossroads of art and science, and achieves two main purposes: data presentation and data analysis (Kelleher and Wagerner 2011, 822). Data visualization projects may involve a diverse network of stakeholders such as economists, geographers, statisticians, scientists, engineers, and biologists whose interests intersect with illustrators, designers, and developers during the production process. Yet in contrast to this complex workflow, a visualization simply involves a messenger, a message, and a receiver (Kirk 2012, 16-17). Crafting these messages means understanding how to analyze and manipulate data during the design process. Knowing how to apply best practices and guidelines helps solve these conceptual design problems. Next, designers and developers must choose supportive tools and technologies that align with their capabilities to construct and distribute web-based data visualizations.

Designers and developers have different workflows. Designers use flexible graphic user interface (GUIs) tools to instill creativity into projects. Developers fulfill functional requirements while working in code-based programming environments. Both interests need to be united to produce engaging and usable data visualizations. These disciplines often overlap while working toward the same end goal, and knowing which approach to take can be hard to discern. Paradoxically, the widest range of design solutions comes from a programming environment, rather than GUI tools (Bigelow 2014, 21-22). A web-standards framework accordingly provides a utilitarian production solution that offers optimal design flexibility and portability of work.

Finding effective workflow strategies for web-based data visualizations are complicated problems for designers and developers. As such, simple answers are not a realistic expectation. However, addressing these issues can save substantial time and effort needed to produce the work. Solutions benefit not only designers and developers, but all project stakeholders and their consumers. Applying design best practices and guidelines, and choosing a web-standards framework for development and distribution will help create effective and portable data visualizations.

Approach

The research focus here examines an overview of best practices and guidelines in data visualization design, and the tools and frameworks used in production. Understanding design concerns and the technical production options may help designers and developers choose a production path that aligns with their capabilities and resources, and can streamline the data visualization production process.

The literature review explores what best practices and guidelines are used in data visualization design and development, to provide fundamental instruction on how data is used as a source material. Following these conceptual underpinnings, the research explores the practical use of software tools and development frameworks to uncover technical production and distribution concerns. Open-source web standards and commercial software are compared to reveal how platforms align with specific design and development methods. Industry-standard visualization tools D3 and Tableau are analyzed as case studies. Literature primarily includes secondary research from scholarly resources such as journals, books or credible articles. The authors are data visualization practitioners, subject matter experts, or scholars. Supplementary research sources are credible websites, and nonpublished anonymous research from professionals in the statistics and design and computer-science fields.

The scope of research chosen here reflects practical design and development concerns needed to understand web-based data visualization production. It aims to inform developers and designers how best practices and guidelines can be used to build effective work, and examines why a web-standards development framework for distribution and development is recommended to produce portable web-based data visualizations.

Literature review

To decide if applying design best practices and guidelines and choosing a web-standards development framework for distribution can create effective and portable data visualizations, research was conducted into the conceptual design process and the application of software tools and frameworks. Guiding research questions in each topic area were:

1. Design—How can designing with data be made an effective process?
2. Development—What tools are used to construct web-based data visualizations?
3. Distribution—Do web standards offer an optimal distribution framework?

Exploring these areas supported the main research question, "*What can designers and developers of data visualizations do to ensure their work is effective and made widely accessible across the web*?"

History of Data Visualization

Highlighting developmental milestones in data visualization illustrates the important role it has played in the development of society, and offers insight into where future innovations may occur. According to Michael Friendly (2006, 3-4), Quantitative Methods Chair at York University and data analysis expert, there are no accounts that span the history of data visualization. The earliest records show use of geometric diagrams and tables to track celestial bodies, and by 200 BC systems akin to longitude and latitude were in place. In the seventeenth century, visual innovations truly became popular as explorers relied on maps to navigate a largely uncharted world. With these new discoveries came advancements in measurement and theories used to estimate time, space, and distance. In tandem, the fields of statistics, analytic geometry, and demographic studies were born, leading to a new way of visual thinking that ushered in The Age of Enlightenment (Chen 2017b, 5). New graphic forms were the next visualization milestones in the eighteenth century, when statistical theories, data collection, and novel ideas for graphic representation were developed. The work of William Playfair came to light, who created the first line graph and bar chart in 1786 (Friendly 2006, 7-8).

Building on this evolutionary foundation, the beginnings of the modern graphic period and the golden age of statistical graphics began in the nineteenth century. An explosive growth of visual innovations surfaced, unparalleled until recently. Nearly all modern forms of data display were invented, such as pie charts, scatterplots, histograms, 3D, etc. Advancements in thematic cartography were equally impressive with the creation of atlases as well as economic, medical, social, and physical maps (Friendly 2006, 9-14). With this boom in scientific graphics and visual thinking came action on the part of governments and societies to use the visualizations as strategic decision-making tools (Chen 2017b, 5). A famous example was the dot map created by Dr. John Snow in 1855, which revealed the source of an epidemic cholera outbreak to a cluster of deaths surrounding a water pump in London, confirming his hypothesis that cholera was not airborne but was a waterborne disease. This visual discovery paved the way for modern epidemiological mapping (Friendly 2006, 11-15) that now saves countless lives.

The 1900s began data visualization's modern dark age marked by a period of dormancy of visual innovation. This was in part due to developments in quantitative statistics, which produced impressively long sets of numbers, and visualizations were perceived as less accurate and powerful. The short period of latency changed quickly with the rebirth of data visualization from 1950-1975. (Friendly 2006, 20-23). As society moved from the industrial revolution to the digital age, computers were playing a large role in the transformation of data visualization.

Using computer technology, Princeton statistics professor John Tukey founded the revolutionary Exploratory Data Analysis (EDA) process in 1977. Tukey realized computer software could empower users to explore and interact with data using dynamic graphic visualizations to gain insight (Chen 2017b, 5-6). Throughout history data visualizations had been *explanatory* in nature; their function was to depict and explain a single pre-established narrative. With the advent of EDA, however, users could interact with, explore, and analyze their data through GUI software. This *exploratory* data analysis allowed users to find their own narratives and draw their own conclusions from their data (Kirk 2016, 29-37). This groundbreaking discovery still plays a crucial role in data visualization methods today. With Tukey's discovery of EDA, the two main functions of data visualization are now data presentation and data analysis (Kelleher 2011, 822).

From 1975 on, we enter the High-D, interactive and dynamic period still evolving today. This data visualization epoch has seen so many developments it is a challenge to categorize the breadth, depth, and rapid pace of innovations. Some of the key developments are highly interactive statistical computing systems, vastly increased computer processing speed and capacity, increased attention on the cognitive and perceptual aspects of data display, and an emphasis on creating multidimensional visualizations that depict data correlations in progressively higher dimensions. Perhaps the most important innovations, however, are software advancements that allow dynamic graphic methods of instantaneously manipulating objects and data together (Friendly 2006, 24-25). In creating the next wave of innovation, the powerful role that software plays strongly suggests this is a prime area for further research.

A key takeaway from these historical milestones is that data visualization has long accompanied and accelerated innovation in society. This is easiest to visualize in Figure 1, where the overall trend shows slow and steady growth until new graphic forms arose near the modern period. This was followed by an enormous proliferation of innovation, which peaks at the new millennium. Given the tight bond between data visualization innovation and societal advancement—and the rapid expansion of technology and widespread data availability—it seems an ideal time to explore how tools and techniques can be leveraged during the next epoch of data visualization innovations.

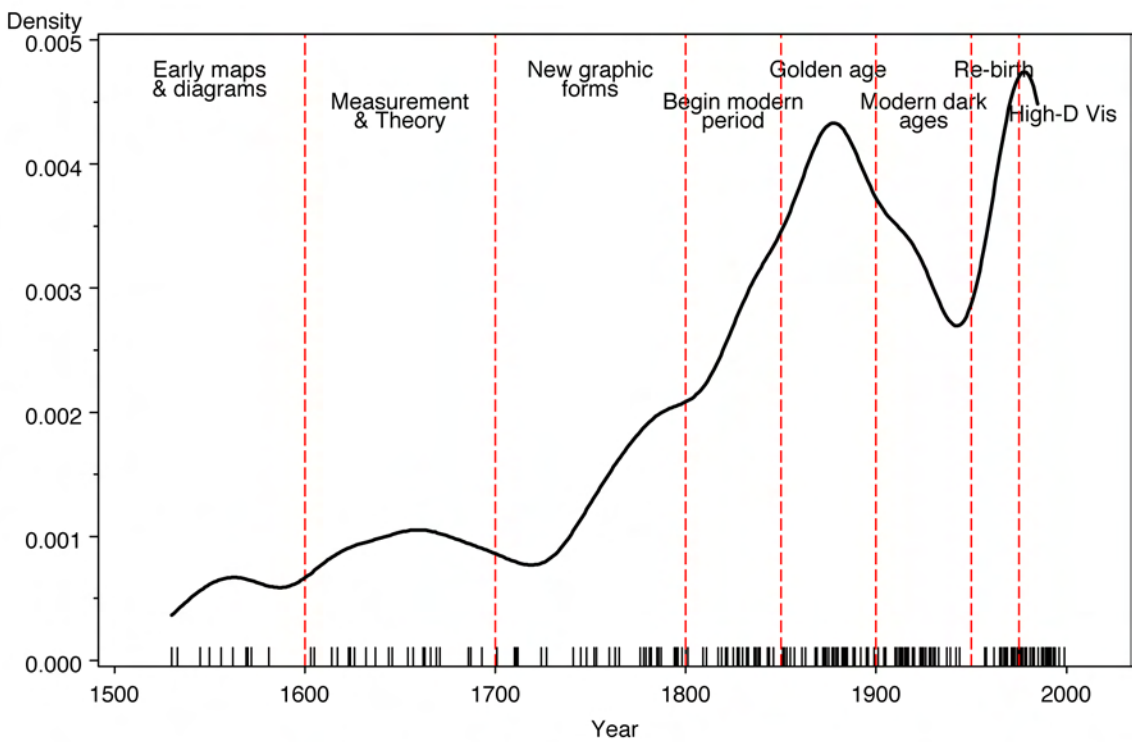


Figure 1. Visualization Milestones. The occurrence of innovation milestones during the history of data visualization, shown with a rug plot and density estimate. Source: Friendly 2006, 3.

Designing with Data

Understanding data design methodologies and their applications are foremost concerns in the data visualization production. Throughout the workflow, a designer navigates key stages, considerations, and tactics that inform the design process. These methods are technology-neutral and applicable to any data design problem. By conceptualizing and rationalizing decision-making choices, strategic direction for efficient, effective, and elegant design solutions can be gained (Kirk 2016, 15-17). Although this approach may seem foolproof, these methods are guidelines and not meant as explicit instructions. In practice, designers typically solve problems through heuristic and iterative methods (Bigelow 2014). Nonetheless, there are common approaches to conceptualizing effective data design solutions. The main themes include finding stories in raw data, choosing proper representational forms, and applying best practices to the presentation of data visualizations (Kirk 2016, 95-116; Kelleher 2011).

Finding Stories in Data

A principle conceptual design guideline is interpreting raw data to create stories that have a narrative focus (Kirk 2016, 53-78; Murray 2013; Segel and Heer 2010; Yau 2011). Later design guidelines emphasize *how* designs are implemented, but an established narrative focus shows *what* message might be crafted. Finding an editorial focus allows the most valuable and relevant dimensions of a story to surface from raw data. This focus is crucial in communicating a visualization effectively to an audience. Strategies for developing effective storylines include forming data questions, finding relevant data sources, and becoming familiar with data through preparatory analysis and exploration.

Forming data questions

Forming data questions is a key data design guideline. From the designer's perspective data questions are essential analyses and lines of inquiry into the meaning of the data. Exploring what data questions need to be answered is a foundational step to understand how to build an effective data visualization. It helps define the narrative story of the visualization and shapes the editorial focus of the message. From a user's perspective data questions are the lines of interrogation and dimensions of interpretation employed when reading a data visualization. Understanding the questions an audience might have is an important facet of the design process as it helps designers refine the key insights to present. Design questions are asked iteratively in this phase of the process and often change as the discovery unfolds (Kirk 2016, 61-62; Yau 2011, 20).

Andy Kirk (2016, 61-62), a data visualization specialist and design consultant, claims strategies to form data questions include deductive and inductive reasoning. Deductive reasoning is a targeted approach used to confirm or negate predetermined ideas about storylines. A designer examines data to find evidence to substantiate the key assumptions and dimensions of a story. Alternately, inductive reasoning uses an EDA approach to find unknown stories. A designer explores data without presuming to know what stories exist. These approaches are optimally used together to maximize the possible dimensions of a storyline. Forming meaningful data questions is an essential guideline during the data discovery process. It is the first step a designer takes in transforming raw data into a compelling story for an audience to engage with.

Sourcing data

Sourcing relevant data for stories is a data visualization best practice. Often clients or employers furnish data, but data may also be discovered independently to find or substantiate stories. In this case, common data sources and techniques to obtain data are:

- Open-data portals have data released publicly online under permissive licenses allowing for use, reuse, modification, and distribution.
- Application programming interfaces (APIs) allow developers to access data programmatically through web services.
- Websites may offer topical data, such as geographic, social, and world data.
- Governmental and political organizations offer many comprehensive data sets.
- Universities may allow public access to expansive data archives and statistics departments that furnish data.
- Direct contact with subject matter experts, authors, and academicians may yield first-hand data (Amr and Stamboliyska 2016, 61-65; Yau 2011, 23-27).
- Data scraping involves programmatically accessing unstructured data by crawling websites or extracting data from PDFs.

When sourcing data, a designer should always question the validity of the data. Nathan Yau (2011, 12), a popular data visualization author with a PhD in statistics, claims verifying data may be the most important part of the visualization design process. It is the designer's responsibility to confirm that errors are corrected before using the data, and not the responsibility of the data provider. Sourcing data is an important part of the data visualization production process. It emphasizes finding and creating effective stories to engage users.

Data preparation and exploration

After forming data questions and finding necessary data sources, the data preparation and exploration phase begins. This stage often involves the most time and effort in the data visualization design process. The designer now performs the roles of journalist and data scientist to examine the data for story leads while checking the data for omissions, accuracy, and formatting errors. The aim is to design compelling and technically accurate stories. Common guidelines and best practices at this stage are understanding data types, formatting data for quality and analysis, and exploring data visually (Kirk 2016, 53-58).

Understanding data types is a key tactic in data preparation and exploration. Examining the dataset structure reveals the types of data present. Knowing these data types informs which graphic representations may work best at later stages in the design process. A recommended best practice is to document the discrete data types in a dataset. An example using Olympic records lists the common data types, shown in Table 1. A similar technique is to create a sample of the data held against each field, shown in Table 2 (Kirk 2016, 55). These practices improve the designer’s understanding of data relationships and potential storylines. It also limits the range of possible solutions, which streamlines the design process.

Table 1. Comparison of common discrete data types

Types	**Examples**
Categorical nominal	Countries, gender, text
Categorical ordinal	Olympic medals, Likert scale
Quantitative (interval-scale)	Dates, temperature
Quantitative (ratio-scale)	Prices, age, distance

Source: Kirk 2016, 55

Table 2. Sample table of data, data types and ranges

Data	**Types**	**Range**
Event	Quantitative (interval-scale)	27 different years
Medal	Categorical ordinal	Gold, silver, bronze
Athlete	Categorical nominal	1,500 athlete names
Result	Quantitative (ratio-scale)	Race results (9.59s > 4:02:59)
Country	Categorical nominal	96 different country names

Source: Kirk 2016, 55

The next stage of exploratory preparation involves transforming data for quality and analysis. Transforming for quality involves data cleansing: finding areas where data is missing, removing duplicates, cleaning up erroneous values, and handling uncommon characters. Next, data is transformed for analysis. The goal is to find which categories and fields of data might be merged, split up, converted, calculated, removed, or added in to achieve greater refinement in anticipation of the projected use by users who will analyze the presentation (Kirk 2016, 56).

Another important consideration in the preparation process is to select the level of data display resolution, or story detail. As some data sets may be quite large, finding the right resolution corresponds to a fitting level of detail necessary to tell a story, such as:

- *Full resolution* plots all the available data.
- *Filtered resolution* excludes some records.
- *Aggregate resolution* rolls-up data into categories such as month, year, or category.
- *Sample resolution* has calculations applied to extract segments of the data. This is particularly helpful for testing, reducing and mocking up data visualizations.
- *Headline resolution* only shows overall statistical totals (Kirk 2016, 56-59).

EDA is another principle way to understand data. While the previous methods of analysis emphasize inspecting raw data, EDA focuses on visually exploring data to quickly gain insight into dataset structure. EDA reveals trends, patterns, and significant relationships in data stories through visually graphic methods, which is hard to achieve by solely reading raw data (Kirk 2016, 6-8). EDA uses interactive computer software, discussed in later sections.

A famous example, “Anscombe’s quartet,” shows what makes the EDA approach effective. Developed by statistician Francis Anscombe in 1973, he created a set of numbers that had nearly identical statistical properties such as mean, variance and correlation. When looking at raw data, it is hard to visualize what trends might emerge, but the data visualizations reveal insights instantly, shown in Figure 2 (Kirk 2016, 6-8).

Anscombe’s quartet proves how quickly the brain draws inferences from visual stimuli and patterns versus reading textual data. This phenomenon, known as pre-attentive processing, takes place instantaneously in sensory memory without the intervention of consciousness. Gestalt laws also highlight how visual perceptual patterns and the effective arrangement of visual symbols allow viewers to quickly draw holistic inferences from pieces of information (Chen 2017, 8-9). Effective and efficient visual processing is a key principle of data visualization, as it reveals insight into data stories with minimal cognitive effort (Kirk 2016, 63-67).

Forming data questions, finding relevant data sources, and preparing data for analysis and exploration are among the most important guidelines and best practices in data visualization design. Becoming familiar with and transforming the data for use allows designers and developers to begin designing stories that can effectively engage audiences.

x1	y1	x2	y2	x3	y3	x4	y4
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

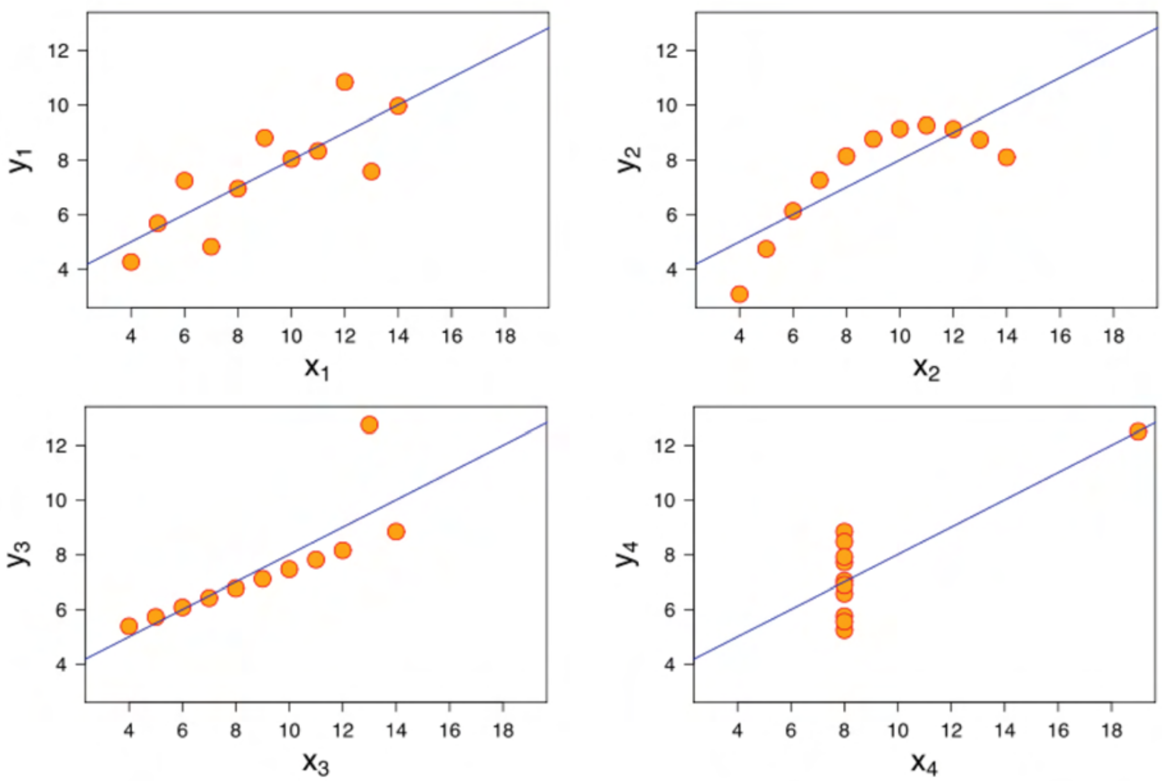


Figure 2. Anscombe's quartet. Two representations of the the same data reveal how quickly visual processing leads to insight. Source: Kirk 2016, 7.

Choosing Representational Forms

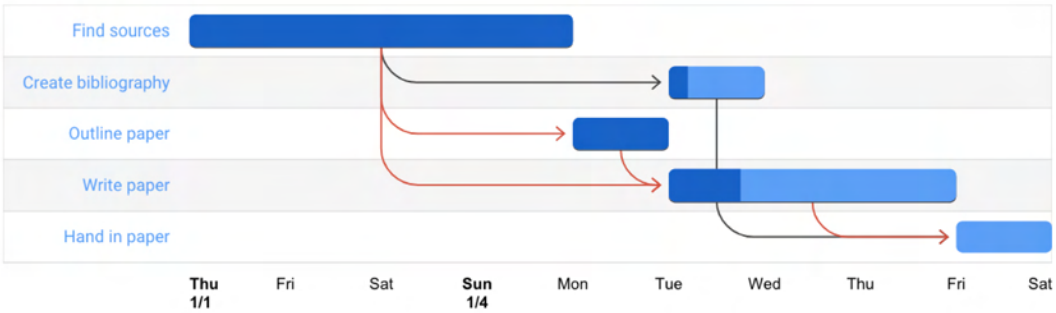
The next major phase of the data visualization production process is to find proper representational forms for the display of data. Choosing a proper chart type creates a connection between the story, the physicality of the data, and the visual representation. Selection relies on understanding a taxonomy of visualization methods. Sources agree that representational forms are not exhaustive, and that categories often overlap. However, there are common guidelines to find effective forms to represent data. The standard forms depict patterns over time, proportions, and relationships (Yau 2011, 91-269; Kirk 2016, 119-158).

Patterns over time

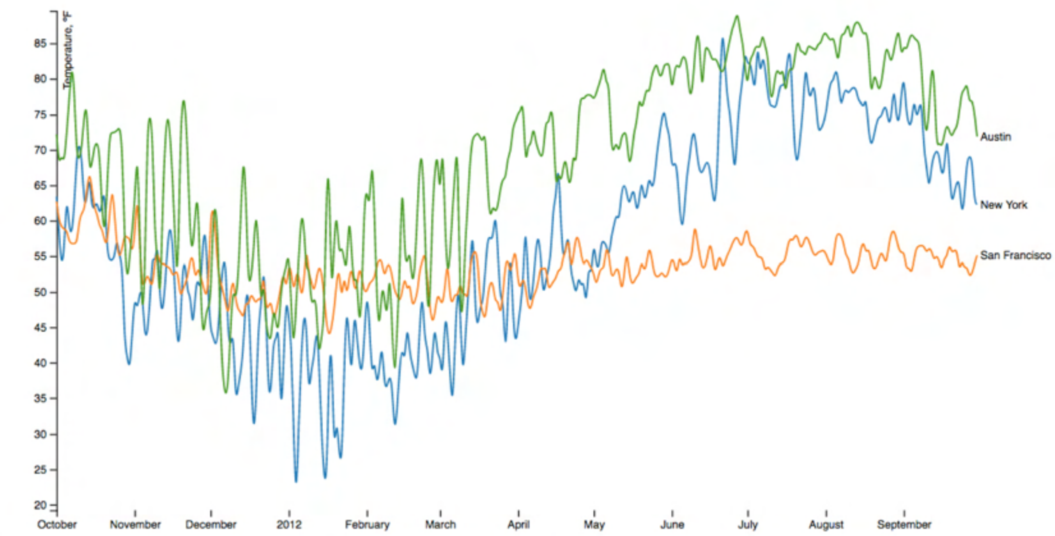
One of the most popular forms of representation are those that depict trends over time. The popularity of temporal forms is due to how often time is used in everyday situations. Examples of these visualizations are line charts, bar graphs, scatterplots, sparklines, area charts, horizon charts, Gantt charts, stream graphs, box-and-whisker plots, rug plots, and flow maps, shown in Figure 3 (Yau 2011, 91-269; Kirk 2016, 119-158).

To know if a patterns-over-time visualization is an appropriate form, observe the data set to see if it contains trends like increasing or decreasing values, seasonal cycles, or other time-bound variables. Temporal data are categorized as discrete or continuous. Discrete data are values that occur at certain points in time with a finite number of possible values, such as test scores percentages, which occur only once. Continuous data are on a continuum and can constantly change, such as temperature, which has no predetermined value range. Knowing these distinctions helps limit the possible choices for representation of data visualizations (Yau 2011, 92-93).

A. Gantt chart



B. Line graph



C. Column bar chart

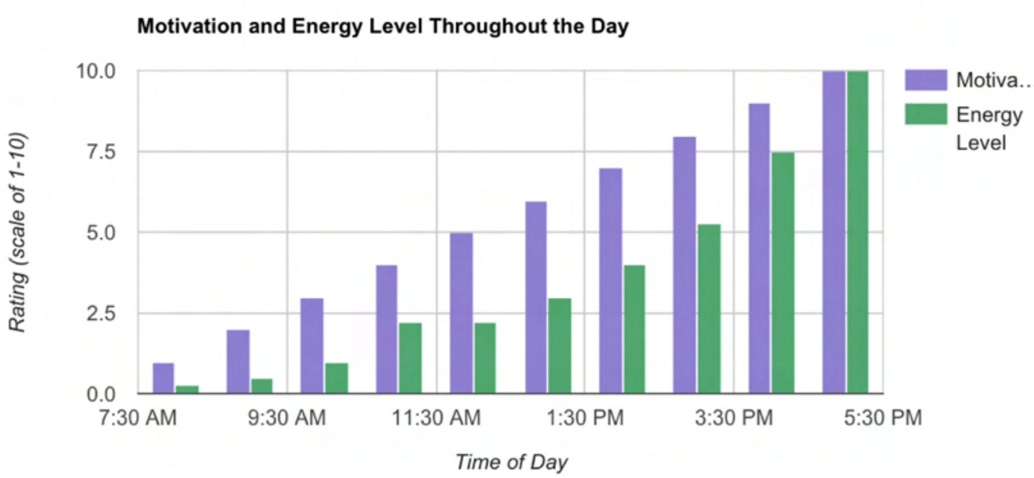


Figure 3. Patterns-over-time data visualizations. Changes and trends over time are among the most popular representational forms used due to the ubiquity of time. Sources: Google Charts 2017a (A), GitHub Gist 2017 (B), Google Charts 2017c (C).

Proportions

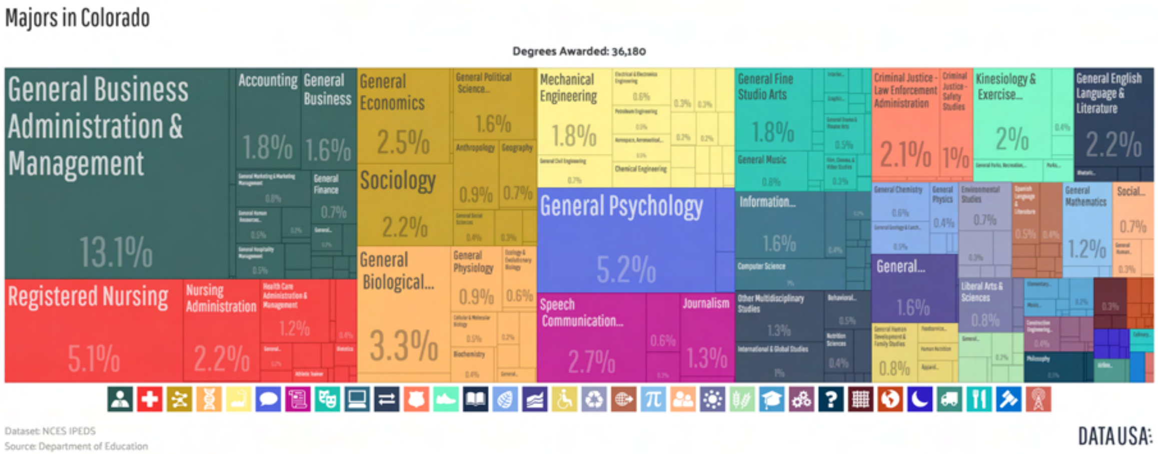
The next category of graphic forms are proportional, or parts-of-a-whole data representations. Proportional data are grouped by categories, subcategories, and by population of total choices or outcomes, where total proportions add up to one and percentages to one hundred percent. The data can usually be sorted by maximum, minimum, and overall distribution. The distribution of values tells the story, as each part shows how it relates to the whole. Examples of these representations are pie charts, donut charts, stacked bar charts, unit charts, tree maps, and bubble hierarchy charts, shown in Figure 4 (Yao 2011, 135-178).

Relationships

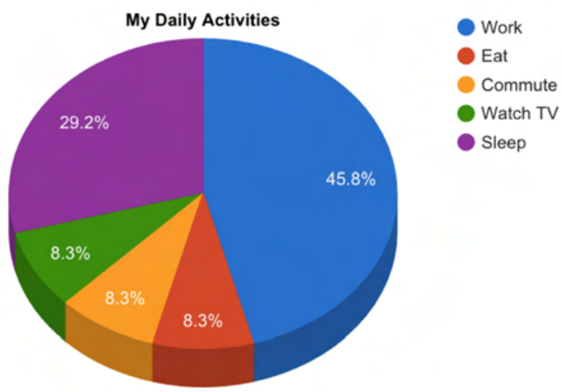
Finding relationships among data variables is a key facet of selecting appropriate representations. Relationship representations show correlations between groups and sub-groups of data. These representations are the most challenging to design, because critical thinking is necessary to understand data relationships, however, this often elicits the best stories. A common approach to visualizing relationships is to display all the data at once. Radar graphs, for example, are effective in showing relationships among multivariate data (Yau 2011, 228-244). Common representations for relationship visualizations are scatter plots and matrices, bubble charts, heat maps, parallel coordinates, network diagrams, stem-and-leaf plots, and histograms, shown in Figure 5 (Kirk 140-146; Yau 179-226).

Patterns over time, proportions, and relationships are the most common representational forms used in data visualization. Designers and developers that understand these representational guidelines are likely to create data visualization stories that are easiest for audiences to process and understand.

A. Tree map



B. Pie chart



C. Stacked bar chart

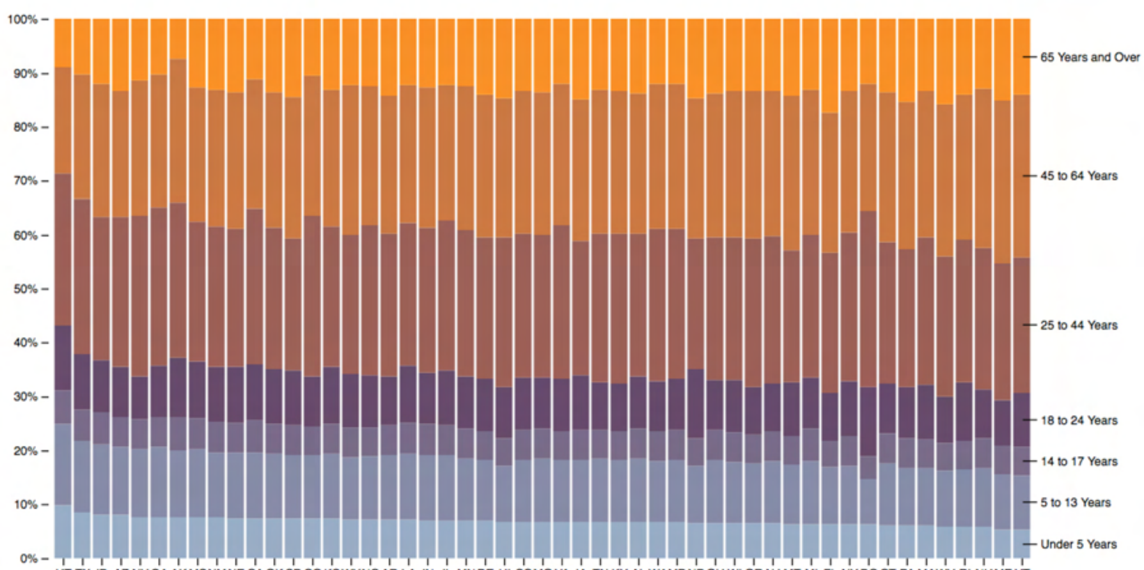
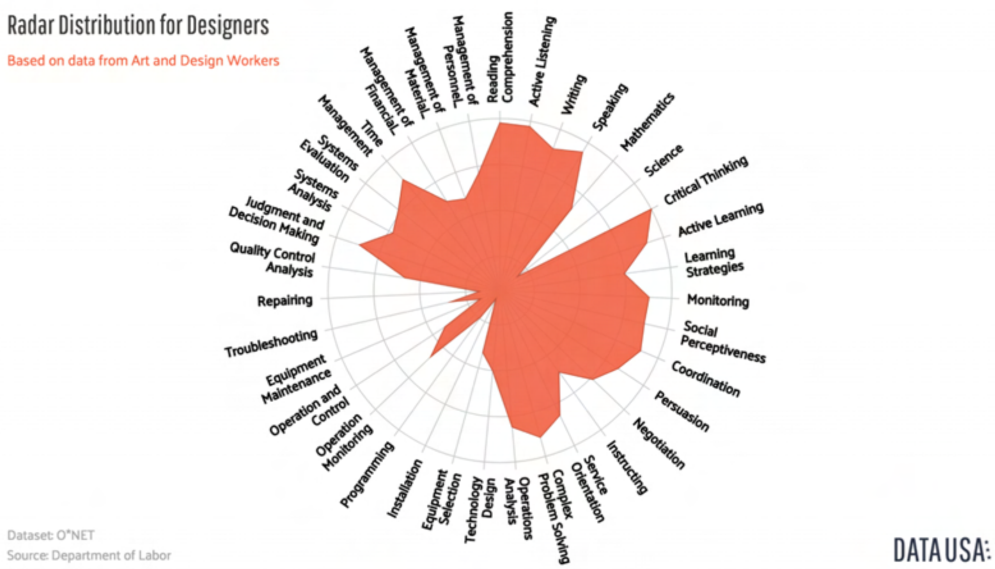


Figure 4. Proportion data visualizations. Each variable is depicted as a part of the whole. Sources: DataUSA 2017b (A), Google Charts 2017d (B), GitHub Gist 2016 (C).

A. Word cloud

B. Radar graph



C. Network diagram

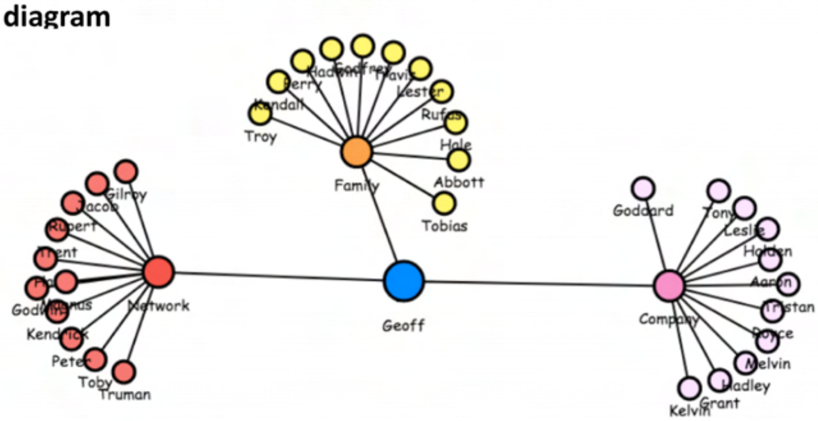


Figure 5. Relationship data visualizations. Relationships are revealed through correlation, distribution, and comparison. Sources: Davies 2017 (A), DataUSA 2017c (B), adapted from DecemberCafe Studio 2016 (C).

Applying Presentational Elements

The final guidelines and best practices in designing with data are applying the presentational elements and aspects of data visualizations. Effective presentation offers extra meaning, insight, and intuition to audiences. Graphic embellishments should be unobtrusive so the data is visually dominant and comprehensible. Key concepts are designing with enough space to allow discrimination between categories and values, and not adding design elements that could be misconstrued as data. The main best practices of presentation are the use of color, architectural arrangement, and adding interactive features (Kirk 2016, 91-92).

Color

Effective use of color creates attractive designs that leverage pre-attentive processing. Effective color schemes bring data to the forefront of the design, while striving for utility and elegance over novelty. Visual hierarchies are hard to see when the foreground and background colors are muddled. A best practice is to not use strong colors over large areas, but to highlight data in small areas of key significance, as shown in Figure 6 (Kirk 2016, 92-102).

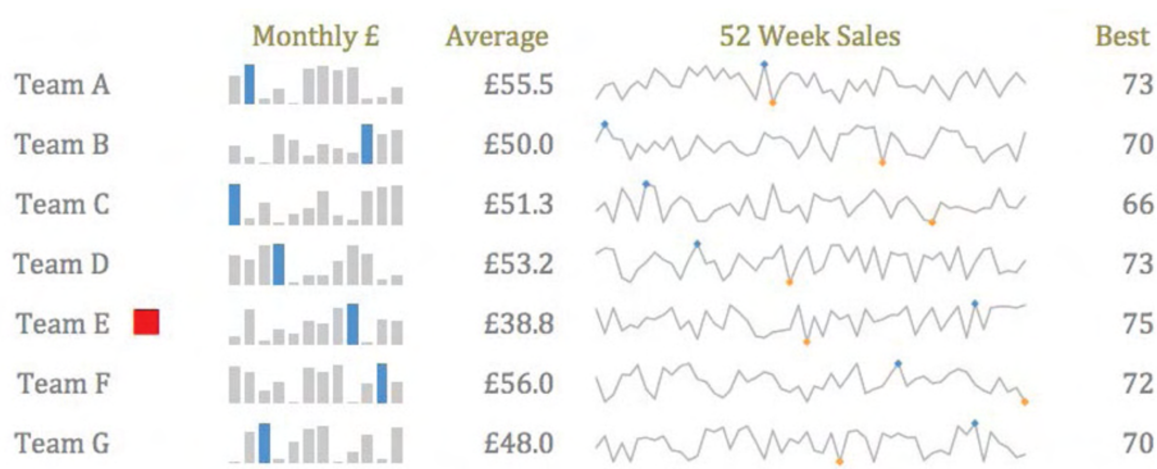


Figure 6. Visual hierarchy. Color palettes can be designed to emphasize key data, as seen in this sparkline chart. Source: Kirk 2016, 100.

One of the most misunderstood uses of color in visualizations is using hue to convey quantity. This technique is usually chosen to convey hierarchies or magnitudes, but it creates extra effort to process cognitively. A suggested a solution is using color to indicate quantity through a single gradated color using color lightness to represent value, as shown in Figure 7 (Kirk 2016, 94).

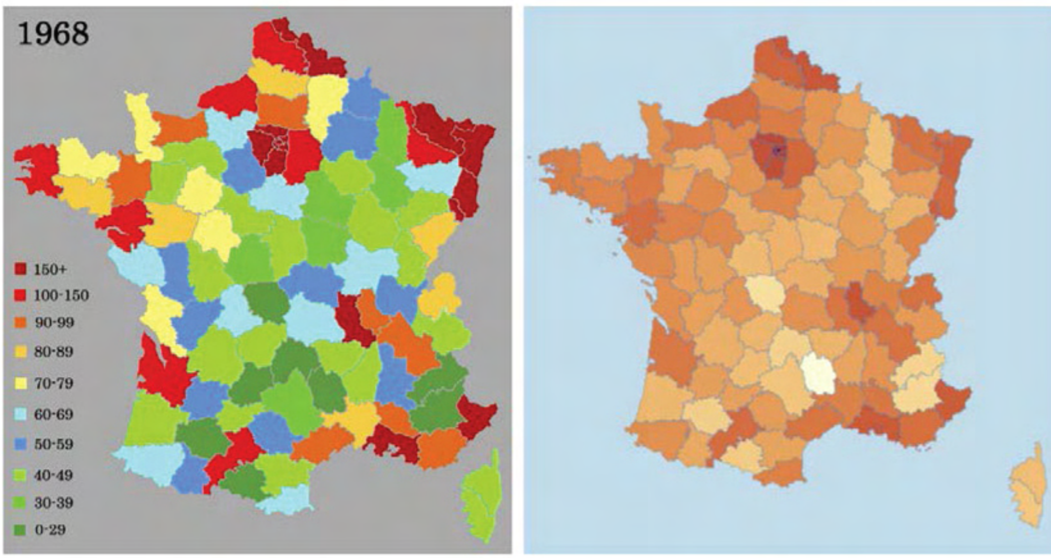


Figure 7. Gradation as value. Hue should not be used to convey value as it is a cognitive burden (left), but a gradation of one color is easily understood (right). Source: Kirk 2016, 100.

Other common problems with color involve perception. Around ten percent of the population has a red-green color deficiency, where that color combination is indistinguishable. A best practice is to replace green with blue in these situations. Additionally, the human eye can only distinguish twelve different color classifications at once, so it is beneficial to limit or combine categories of data so they do not exceed the limits of what can easily be perceived (Kirk 2016, 96-99). Color plays a dominant role in presenting data visualizations, and applying these best practices can help create more effective and useful designs.

Architectural arrangement

The architectural arrangement of a design can create an intuitive user experience. The logic and implied meaning behind the arrangement of design features can reduce the amount of effort needed to process and navigate hierarchies and sequences of display. Designers should be able to justify all the decisions they make regarding size, positioning, grouping, and sorting of design elements and features (Kirk 2016, 111).

Related to architectural arrangement are the use of annotations. Annotations are important to add another layer of usefulness and meaning to a design through the use of titles, introductions, labels, captions, and user guides. Visualizations should include legends, keys, units of measurement, data source credits and attribution. (Kirk 2016, 107-111). The careful architectural arrangement of design elements and use of annotations enriches the meaning and clarity of stories, making data visualizations easier to use.

Interactive features

Interactivity is an optional best practice that can add interest and functionality to data visualizations. Scott Murray (2013, 2-4), assistant professor of design at the University of San Francisco and teacher of interactive and data visualization design courses, claims interactivity adds substantial value to data visualizations. Static visualizations show only a pre-composed view of data. Consequently, multiple representations are used to present different perspectives of the same data. Interactive visualizations, however, enable EDA through the standard pattern *overview first, zoom and filter, then details on demand,* offering multiple perspectives in one application. Another value of interactive design is that GUIs can be very engaging and game-like, encouraging users to investigate topics they might not be interested in, shown in Figure 8.

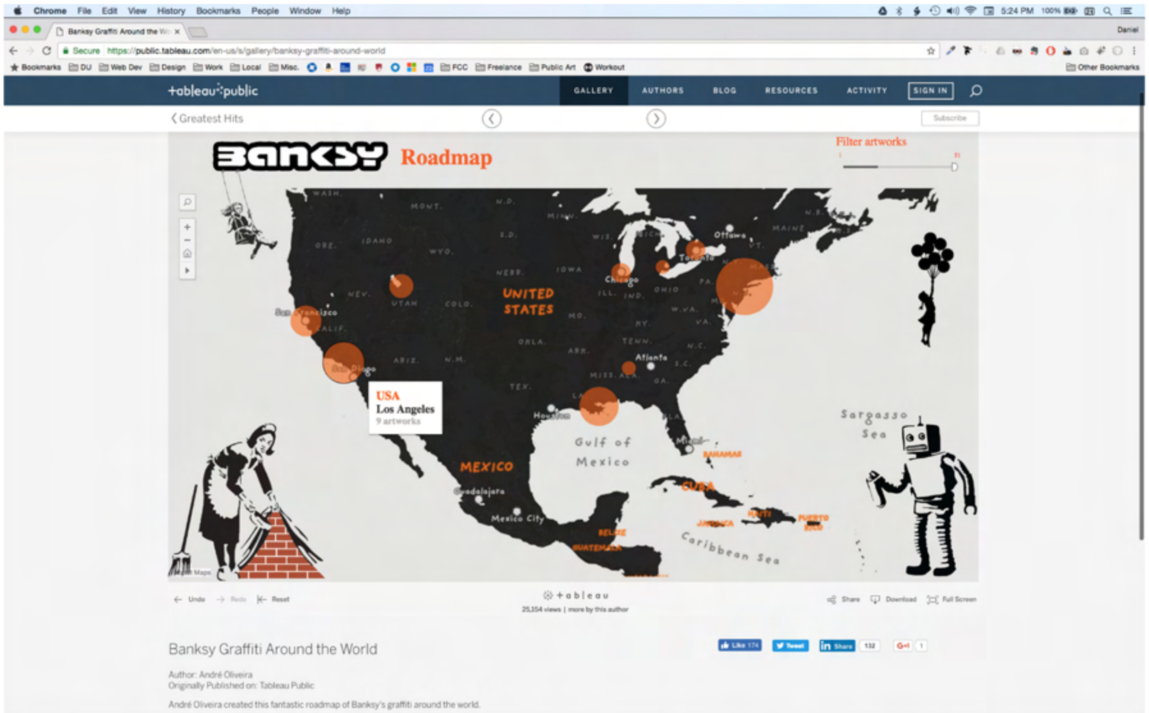


Figure 8. Interactive features. Data visualizations with game-like interactivity can persuade users to explore stories they might have passed by. Source: Oliveira 2017.

Designing interactive data visualizations marks a critical juncture in the workflow. Kirk (2016, 103-104) states that, "The development of interactive designs requires technical capabilities. There is no way of avoiding that… Other constraints such as platform compatibility, data loading speed, and server capacity need to be factored in as well." The inclusion of interactivity foreshadows many technical issues to consider when publishing interactive web-based data visualizations. These issues are discussed in following sections.

Applying presentational best practices are a key facet of the data design process, as they engage audiences, while adding utility and ease of use. Along with finding data stories, and selecting proper representational forms, these areas make up the main conceptual guidelines and best practices for effectively designing data visualizations. Beyond these considerations, the design process begins to merge with the technical development of data visualizations.

Development Tools

After the preliminary conceptual data design stages, the literature shows the next round of practical considerations is choosing software tools to construct data visualizations (Bigelow 2014, 17-23; Kirk 2016, 159-169; Simon 2014 51-74; Yau 53-89). Since data visualization spans many disciplines, a vast array of data visualization tools are available. Popular applications are shown on visualisingdata.com, which shows a total of 305 applications in eight use case categories (Kirk 2017). Evidently, choosing the right software can be a complex task.

Consensus is that developers and designers need to learn to use a wide range of tools—there is no one tool that can do everything. Ultimately, the development process is influenced more by capability and access to resources than the suitability of a given tool (Bigelow 2014, 17-23; Kirk 2016, 159-169; Simon 2014 51-74; Yau 53-89). Consequently, focusing on workflows and tools that best match the capabilities of designers and developers offers clearest insight into software selection strategies. Observational studies conducted with data designers discovered two methodologies that govern data visualization workflows: *visualization creation tools*, and *visualization programming environments* (Bigelow 2014, 18).

Visualization creation tools focus on importing data and making visual representations easy and largely automated through drag-and-drop methods. These out-of-the-box applications are easiest for beginners to pick up. Common examples of software applications are Tableau, IBM Many Eyes, and Microsoft Excel. The main downside in selecting these tools is the lack of design flexibility. Only basic design customization of colors, fonts, and titles are often supported. If a designer wants to depict a data visualization with a certain chart type it may not be available. Designers commonly opt for visualization creation tools, which is a paradox since there is limited design flexibility available (Bigelow 2014, 18; Yau 2011, 53-62).

Alternately, visualization programming environments support the creation of richly creative data visualizations. These code environments use programming languages to create work, are difficult to learn, and are the domain of developers. Popular environments are D3, Processing, and R (Bigelow 2014, 18; Yau 2011 62-75). This approach is also paradoxical since design flexibility is achieved through programming, which are not skills most designers have. Kirk (2016, 161) explains, “The ultimate capability in visualization design is to have complete control over the characteristics and behavior of every mark, property, and user-driven event on a chart or graph. The only way to fundamentally achieve this level of creative control is through the command of one or a range of programming languages.”

The differences between these approaches underscores an unresolved problem in the data visualization workflow, which is the trade-off between ease-of-use and design flexibility. This highlights an area for further research, as they are important issues to resolve for designers and developers. A comparative analysis of leading data visualization software Tableau and D3 is discussed later to show how these divergent approaches work in practical applications.

Software selection is an important decision in the data visualization workflow. Software tools allow designers and developers to produce data visualizations. Depending on the approach selected, varying degrees of design flexibility are expected. Gaining insight into the choices available may align designers and developers with optimal tools for their capabilities and skills. After selecting a method of development, designers and developers can produce web-based data visualizations that are ready to connect with their audiences.

Distributing Data Visualizations

The final stages of data visualization production are examining methods to distribute the work (Fox and Hendler 2011, 706-708; Kirk 2016, 165-172; Yau 2011, 67-71). From ancient times until the 1970s dissemination of data visualizations were through static means like printed materials (Friendly 2006). However, with the advent of personal computers and the explosive growth of the Internet as a communications medium this changed drastically. Recent data states that around forty percent of the world population has an internet connection today, up from one percent in 1995 (internetlivestats.com 2017). Regarding the power of the web as the primary medium to publish data visualizations, Murray (2013, 3) claims,

> Visualizations aren't truly visual unless they are *seen*. Getting your work out there for others to see is critical, and publishing on the Web is the quickest way to reach a global audience. Working with web-standard technologies means that your work can be seen and experienced by anyone using a recent web browser, regardless of the operating system (Windows, Mac, Linux) and device type (laptop, desktop, smartphone, tablet).

Jeffrey Zeldman (2010, 6-7), who initiated the web standards movement and directed the Web Standards Project (WaSP), states that web standards are software specifications that form the core technologies that allow users to access the Internet. The common standards are hypertext markup language (HTML), extensible markup language (XML), cascading style sheets (CSS), scalable vector graphics (SVG), the document object model (DOM), and JavaScript, which add structure, display, and functional behavior to websites. Created by expert work groups, web standards allow the greatest benefits to be realized by the greatest number of people with a unified roadmap for web technologies (Zeldman and Marcotte 2010, 6-7; Murray 2013, 4). These same core technologies can be used to create data visualizations (Amr and Stamboliyska 2016, 25).

Web standards are mainly governed by the Worldwide Web Consortium (W3C). The director of the W3C, Tim Berners-Lee, invented the Web in 1989, and in 1994 the W3C was created to ensure technologies worked together as the web evolved (Zeldman and Marcotte 2010, 4-7). The necessity for standardization was born from the earliest days of the web, when browser manufacturers fought to dominate the market with proprietary non-standard code. In response, web developers formed the standards advocacy group WaSP in 1994. WaSP successfully lobbied Netscape, Microsoft and other browser manufacturers to adopt standards by persuading them that interoperability was a necessity if the web was to move forward (Zeldman and Marcotte 2010, 50). The adoption of standards allowed the Web to become a worldwide open platform to share information. The W3C currently promotes the development of a wide range of web standards technologies, including those used in data visualization, such as HTML5, CSS3, SVG, and the DOM (W3C.org 2017a).

Phil Simon (2014, 53-63), technology consultant and organizational management expert, states that popular alternatives to open web-standard technologies are proprietary enterprise and best-of-breed applications, which are increasingly cloud-hosted. Distinguishing the lines between data visualization and other core functions in these applications is often blurry since they are marketed as data mining, data discovery, business intelligence, analytics, and enterprise reporting tools. One example of enterprise-level data visualization software is Microsoft Excel, which integrates over the web with Microsoft’s databases and data warehouses. An example of best-of-breed data visualization software is Tableau, which has formed partnerships with some of the world’s largest databases and data warehouses.

The distinction between distribution models is an important consideration. Today's data environment is changing rapidly. Most large companies no longer store data on-premises and are increasingly using proprietary cloud-based models that afford access to big data technologies. Smaller companies are using real-time data sources through web standards technologies and open APIs (Simon 2014, 63-73). These trends in data distribution are another integral issue to resolve in data visualization. The author discussed development and distribution issues with an anonymous source who is a professional data visualization developer. He claims developing data visualizations is equally costly with web standards or proprietary frameworks. It costs valuable time and effort to develop visualizations with web-standards technologies, but it is costly financially to choose proprietary software. Further research and future innovations in this area of data visualization may offer practical solutions.

The literature review helped answer the question, "*What can developers and designers of data visualizations do to ensure their work is effective and made widely accessible across the web*?" From a design perspective, applying techniques to analyze and find relevant stories in raw data, choosing proper representational graphic forms, and creating useful presentational displays can help create effective data visualizations. From a production perspective, designers and developers can select supportive work environments and tools that best match their capabilities. Finally, selecting a web-standards production model is a viable choice for distributing visualizations as it offers cross-platform portability and open-source tools to build rich data visualization design solutions. The following sections examine how applying these design best practices and guidelines and choosing a web-standards development framework for distribution can create effective and portable data visualizations.

Solution

Representing the interests of a cross-disciplinary range of stakeholders presents problems for designers and developers of data visualizations. From a design perspective, problems surface in understanding how to effectively extract intuitive messages and stories from technical data. From a development perspective, there are problems identifying the tools and platforms necessary to build and distribute web-based data visualizations. Both facets need to be resolved to create effective and usable work. Accordingly, a multipart solution involves applying best practices and guidelines during the design process, and selecting a web-standards framework to build and distribute the designs.

Applying design best practices and guidelines during the design process helps build effective work by offering parameters to constrain the range of design solutions to what is relevant and applicable. They model effective work practices and offer techniques to produce those results. Design best practices and guidelines come from varied sources. Many are de facto standards promoted through common use or market forces. An example is Google Charts, a well-documented free web application to create data visualizations (Google 2017b). De facto standards may be difficult to find or trust implicitly, but they are popular nonetheless. Alternately, professional standards bodies like the International Business Communication Standards Association or the W3C also promote design best practices and guidelines. These organizations develop their recommendations through governance models that build consensus among subject matter experts and industry consortia, often with a focus on documentation (IBCS Association 2015; www.w3.org). This assures a high level of confidence that the recommendations have been vetted and of practical use.

The W3C is the best solution for aggregating and hosting best practice and guidelines for web-based data visualizations. The W3C is the premier standards consortium for web technologies and is well known to the web development community. They develop and promote the standard specifications for the main languages and frameworks used on the web and in data visualization. They have a Data on the Web Best Practices standard, and host a Data Visualization Community that is actively lobbying the W3C to draft data visualization web standards (W3C.org 2015a; W3C.org 2017c). Not only does the W3C offer many best practices and guidelines for designing web applications, but they also help create the tools and frameworks for their development and distribution.

Web-standard technologies HTML, CSS, SVG, JavaScript, and the DOM offer methods to import data, construct visualizations, and distribute them over the web. This framework allows common languages to be used in development, which promotes further innovation. Using web standards for design and development allows data visualizations to be portable and freely accessible across platforms and devices, which complements the trends of open-data sources. By choosing a web-standards development framework, and web-standard languages such as D3, data visualization designers and developers can elevate their own capabilities to create effective design solutions, while ensuring that their work is accessible to all.

Applying design best practices and guidelines, and choosing a web-standards framework for development and distribution, can solve common problems that surface in data visualization production. This proposal increases the chances that designers and developers have the practical knowledge, tools, and techniques to make their data visualizations effective and portable across the web. The following section examines this solution in detail.

Discussion

The problems designers and developers producing data visualizations face are conceptual and technical, which requires a multipart solution. Accordingly, knowing how to apply design best practices and guidelines, and then choosing a web-standards development framework for distribution can help create effective and portable data visualizations. The following discussion examines the practicality of applying best practices and guidelines, and compares a commercial data visualization tool with an open-source web-standards framework to reveal the production value of each model.

Do Best Practices and Guidelines Work?

The application of best practices and guidelines initially seems a fitting solution to data visualization problems. Their main strength is that they are defined by consensus among subject matter experts and field practitioners, which seems a credible endorsement based on those merits alone. They offer practical guidance in navigating common problems and arriving at workable solutions. However, there are also reservations in their use that merit discussion.

A weakness in applying guidelines is they are general by nature, and often cannot adequately address specific situations. For example, despite guidance, it is still difficult to choose graphic forms to represent data. Kirk (2016, 115) prefaces his de facto standard chart options with, “A creative field like data visualization simply does not lend itself to finite classification,” then shows forty-two charts and graphs to consider when designing, any of which could be combined with other charting methods. So realistically, guidelines are to be applied at a foundational and conceptual level, and not with the expectation that they will solve specific data visualization problems. This admission may complicate the feasibility of proposing them as a solution, but if taken in proper context guidelines and best practices still play a lead role in understanding subject matter and using that knowledge to align with the intent of the design. Finding this level of balance may be challenging, however, especially for inexperienced designers and developers who are most likely to reference guidelines and best practices.

Guidelines and best practices also have opportunities for growth. Data visualization is a rapidly growing field, and as shown by the historic milestones, society is in a peak period of innovation and is racing to keep pace with technology. While there are long-established graphic forms for representation, innumerable new forms are being generated. To illustrate, Santiago Ortiz (2012), mathematician, developer, and educator, created 45 data visualizations for the two quantities 75 and 37. Ortiz did this to prove that creative visualization methods are an inexhaustible resource to communicate data. Since data visualizations are continually evolving, best practices and guidelines need to evolve with them to maintain relevance and promote the visual methods that communicate most effectively.

A threat to using best practices and guidelines is knowing where to turn to for guidance. Ironically, there can be multiple standards organizations vying to be the authoritative voice for industry. In the case of the web, the W3C is the premier consortium, but others have risen to prominence like the Web Hypertext Application Working Group (WHATWG). WHATWG (2017) was formed by in 2004 by members from Apple, the Mozilla Foundation, and Opera Software, who were concerned the W3C was not properly meeting industry needs. This divergence dilutes the mission of standards to unify industry. Additionally, the production of a data visualization crosses multiple professions and fields, which makes it hard to find a single authoritative voice for the whole design and development process.

Evolutionary trends in web standards show how best practices and guidelines have played a key role in the development of communications technology. Through the development of effective guidelines, standards advocacy groups like WaSP and the W3C have done incredible work to create a global communication platform with the Internet. The trends for web-standard applications to solve similar communication issues is still relevant today. The recently formed W3C Data Visualization Community Group conducted a meeting in September 2015 in Beijing, China, where Chair Xiaoru Yuan said, “We can see two clear trends in the datavis world. One is the datavis work [is] spreading from people with expertise to the public, while the other one is datavis [is] moving from native [applications] to the web. And that is why I think relating datavis standards with [the] Web should be a proper starting point.” (W3C.org 2015b). These trends underscore the importance of creating new standards and guidelines that are specific to the field of data visualization, and that can be readily used by the public. With the volumes of data generated now and certainly in the future, new standards might offer practical solutions to more easily use and visualize data over the web.

Applying best practices and guidelines is a crucial facet of the data visualization process. They promote solutions developed through consensus by credible workgroups and expert practitioners. They provide a benchmark to measure work by, and lend assurance that designs are produced and represented in a useful manner. Designers and developers can rely on best practices and guidelines to improve their subject matter knowledge, and can apply their techniques and methods to produce effective and engaging data visualizations. By resolving these conceptual issues more easily, designers and developers may concentrate on producing the work.

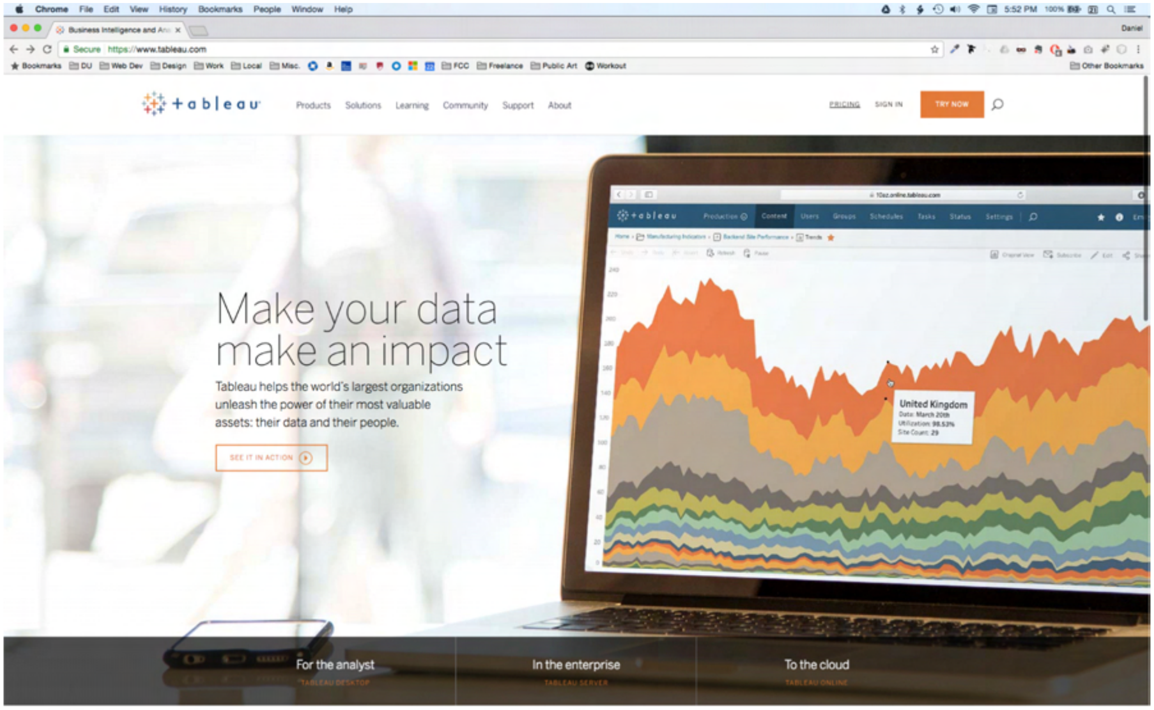
Which Software Tools and Frameworks Work Best?

The software tools and frameworks used to create and distribute data visualizations are key components that transform conceptual designs into practical applications. To understand why an open-source web-standards framework may offer optimal benefits, it is necessary to compare it to a leading proprietary model. Each has distinct advantages and disadvantages, which the following analysis covers in detail.

Tableau

Tableau is a proprietary visualization creation tool, which emphasizes ease-of-use in importing data and creating automated data visualizations via a drag-and-drop GUI (Bigelow 2014, 18), shown in Figure 9. Tableau is the undisputed leader in consumer off-the-shelf data visualization software (Simon 2014, 61). The company was formed in 2003 by Stanford computer science PhD candidate Chris Stolte; his advisor Pat Hanrahan, who had already co-founded Pixar and transformed animated film; and Christian Chabot, who was studying entrepreneurship at Stanford Business School. Together they realized computer graphics could empower people to better understand their data, and they set a precedent by merging two computer science disciplines: computer graphics and databases (Tableau 2003-2017c). Tableau went public on May 17, 2013, and when trading closed the company stock had risen 63 percent and their market capitalization exceeded two billion dollars (Simon 2014, xxi).

Tableau has several key strengths. It is the only best-of-breed software application available that is built exclusively for data visualization (Simon 2014, 61), and occupies a niche market in self-service business intelligence software (Rist 2016). It works well for EDA to quickly understand properties, shapes, and quality of data (Kirk 2016, 159). Basic customizations are



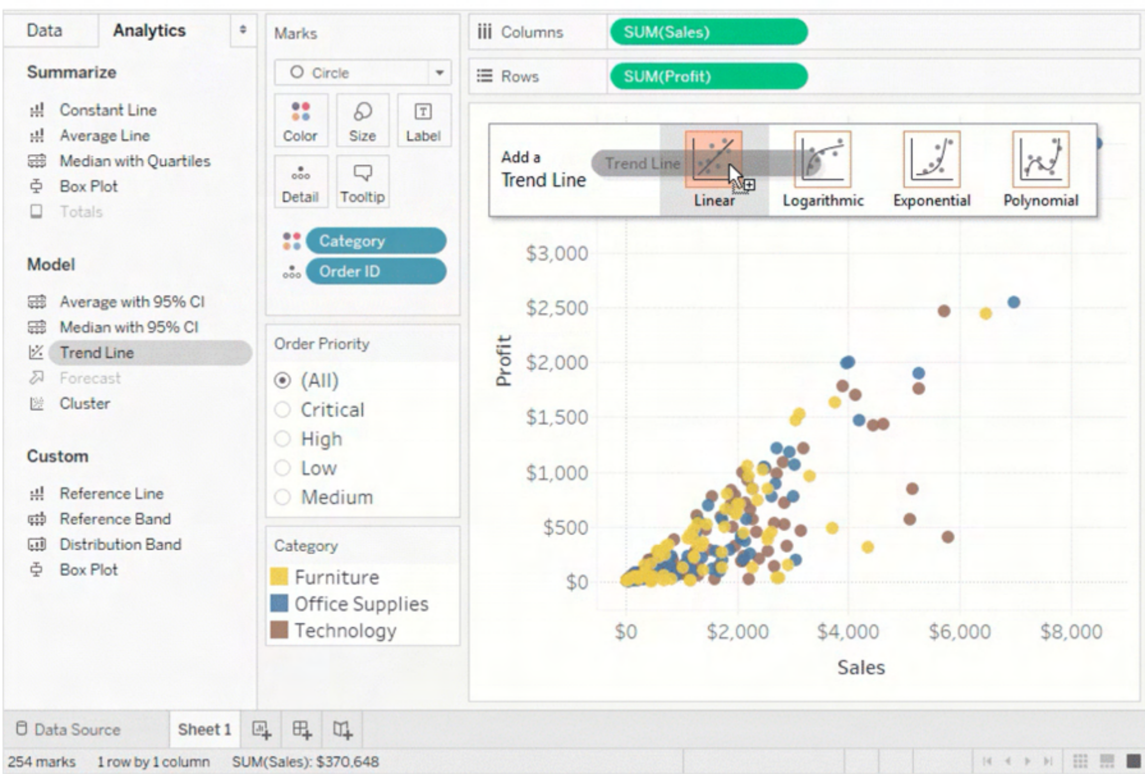


Figure 9. Tableau. The leading data visualization creation tool offers a user-friendly GUI that makes designing with data an efficient and easy process.
Sources: Tableau 2003-2017a (top), Tableau 2003-2017b (bottom).

possible such as changing colors and chart types. It includes an organizational feature called the Story that allows multiple visualizations to be presented with a narrative focus (Hamersky 2016, 150). Data connectivity is another key strength, as Tableau connects to many data sources such as Excel, text files, and database servers including big data platforms like Amazon's Elastic Map Reduce, Google BigQuery, and Cloudera Hadoop. Tableau also works well on mobile devices, supporting touch navigation, and changing visualizations for different screen sizes (5000fish 2014; Rist 2016; Yau 2011, 60-61).

Tableau's weaknesses are cost and limited design flexibility. Tableau has two desktop versions: the Professional edition costs $1,999 and supports forty data connections, and the Personal edition costs $999 and handles six data sources. Organizations needing collaborative visualization workflows use Tableau server. An entry-level server license costs $1,000 per user, per year with a ten-user minimum. Small- and medium-sized businesses that need licenses for one thousand people could expect to see charges exceeding $240,000 plus twenty-five percent maintenance fees may make it cost-prohibitive for sizable organizations (5000fish 2014; Rist 2016; Yau 2011, 60-61). The other main weakness is the lack of design flexibility. Depending on the project, this could prevent a project from moving forward altogether. An anonymous source, a professional data visualization developer, spoke with the author about how this problem surfaced in a recent Tableau data visualization he was producing. He could not use Tableau's built-in map shape files for an international political organization that insisted on not using Mercator map projections since they distorted the shape and size of their country in an unfavorable way. He had to abandon Tableau for the project since it wasn't flexible enough to make this design customization.

Tableau continues to find new opportunities for growth by connecting data visualization products to the public. In 2009, Tableau Public launched as a free model with similar features as the desktop version, but with fewer data sources available. Tableau has extensive online learning communities that have engaging and active forums, galleries, and tutorials available to help learn the software and extend its usefulness (5000fish 2014; Rist 2016). This emphasis on customer-centric marketing may be one of its key strategies to create further opportunity. Tableau is popular because the aesthetics and design are carefully thought out (Yau 2011, 60-61). A Gartner survey revealed seventy percent of Tableau's customers selected the product due to its ease of use (5000fish 2014).

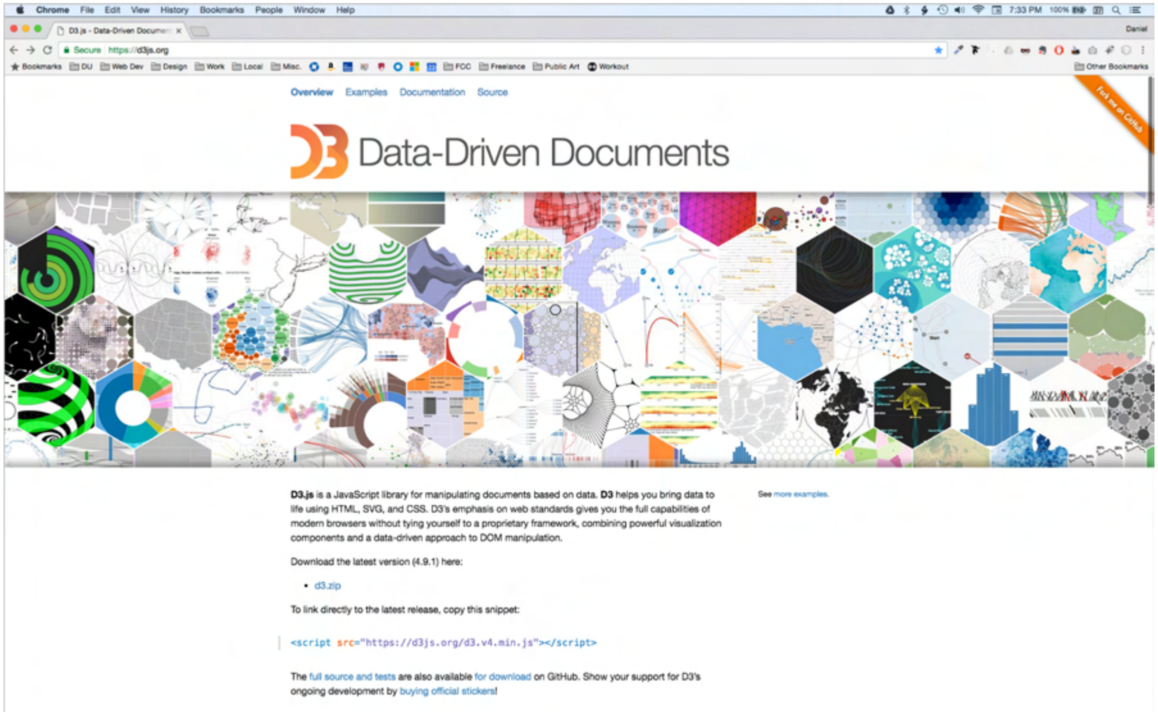
Threats to Tableau's business model come in the form of data governance challenges that compete with free open-source models. Tableau Public hosts free data visualizations, but users wishing to use this service must make their data public, however (Hamersky 2016, 148-149). Users with sensitive or business-related data may opt out of this distribution model to publish their data visualizations. Supporting data privacy relies on buying a commercial version of Tableau.

Tableau sets many trends in data visualization and continues to expand their functionality to accommodate big data, commercial databases, open APIs, and other web services and data streams. Their product lines now include Tableau Desktop, Tableau Server, Tableau Online, Tableau Reader and Tableau Public models (Hamersky 2016, 148). Tableau represents the epitome of the visualization creation tool, enabling novice users to explore and explain their data easily and quickly, but with significant design limitations and potentially high financial costs.

D3

D3.js, commonly known as D3, is a visualization programming environment, that uses code to build sophisticated, flexible, and creative data visualizations (Bigelow 2014, 18), shown in Figure 10. D3 was created by Stanford Vis Group members Jeffery Heer, Mike Bostock, and Vadim Ogievetsky. Heer had already contributed substantially to web data visualization efforts with the languages prefuse and Flare, and in 2009 was advising graduate student Bostock. In 2011, the trio succeeded in creating a data visualization language based solely on web standards that encouraged design experimentation (Murray 2013, 9-10). D3 now is the leading interactive web-standards data visualization framework, perhaps because it offers complete control over design production (Kirk 2016, 161).

D3 stands for Data-Driven Documents, and it main strengths are that it is free, open-source, and emphasizes portability and extraordinary design flexibility. D3 is a JavaScript library that uses HTML, SVG, CSS, and the DOM to create data visualizations, according to D3 principal creator, Mike Bostock (2017b). This framework allows for the efficient manipulation of documents based on data. D3 does this by loading data into the browser's memory, binding data to DOM elements, transforming them by interpreting each element's bound datum and setting its visual properties, then transitioning elements between states in response to user input. This process allows designers and developers full control over how data is expressed and presented. D3 targets the full capabilities of modern browsers to render the work instead of proprietary applications. When a data visualization is produced with D3, its file formats (HTML, CSS, JavaScript, and SVG) are web standards. Since there are no proprietary aspects to the framework, the visualizations are portable across platforms and devices. (Bostock 2017a).



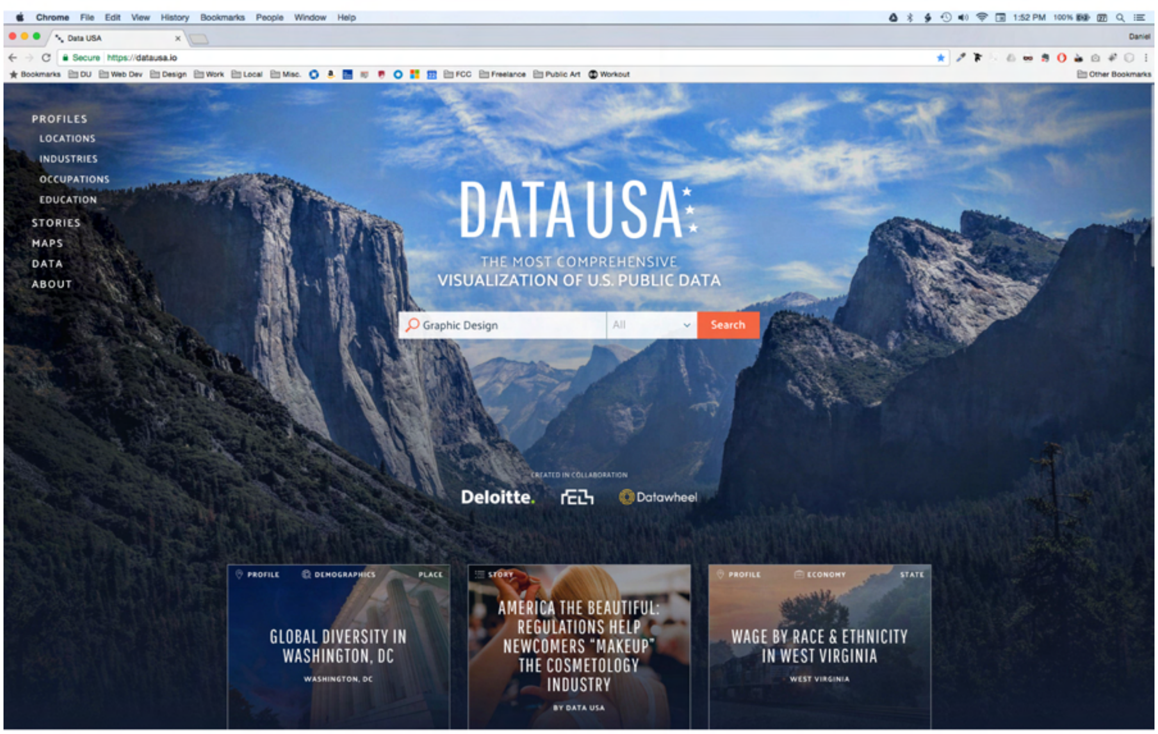


Figure 10. D3. The leading data visualization programming environment uses a dynamic open-source web-standards framework (top). The D3-built DataUSA.io website is the most comprehensive visualization of US public data to date (bottom). Sources: Bostock 2017b (top), DataUSA 2017a (bottom).

The main weakness of the D3 framework is that it is difficult to learn since it a code application. Using D3 comes with the considerable caveat that its use is predicated by understanding fundamental web development concepts and having programming skills (Murray 2013, 4). Bostock (2017b) claims,

> “Alas, coding is famously difficult! Even its name suggests impenetrability... we don’t manipulate the output of our program directly but instead work in abstractions. These abstractions may be powerful, but they can also be difficult to grasp… We continue to have great difficulty understanding code, and it can feel like a miracle that anything works at all.”

Bostock explains that code, while difficult to master, is still the best tool available because it has unlimited expressiveness due to its general and symbolic nature. Bostock claims that ideas we wish to express, explore, and explain through code may be irreducibly complex.

Opportunities for growth in D3 continue to be exploited by Bostock. Knowing the limitations of D3 as an explanatory visualization framework, and that code is inherently hard to understand, Bostock (2017a) has been developing d3.express, an *integrated discovery environment*. It incorporates EDA into the design process by using reactive code to immediately visualize data as it is entered, and interactive GUI widgets can update code directly. The effect is a hybrid application that combines visualization design tool features within a visualization programming environment. This aims to make the coding process easier through visual feedback, without sacrificing design expressiveness and flexibility.

Threats to this D3 framework revolve around the data governance. D3 data visualizations do not protect data, which are processed on a client-side browser. As there is no data security, it might not meet business or consumer requirements to keep data private. Additionally, there is debate that web browsers offer enough computing resources to efficiently process big data applications. These threats could be resolved, however, by incorporating server-side processing for big data and security issues, and delegating the browser to render only the final visualization (Bostock 2017a).

Trends seen in the D3 open-source community are the progressive improvements of base code with libraries, plugins, and utilities. A GitHub page that offers D3 code extensions contains roughly 65 charting libraries, 15 map applications, and 50 utilities (Klack and Möller 2017). With the popularity of D3 and the diverse community of designers and developers that supports the framework by contributing to its development, the foreseeable trend is it will continue to be the leading open-source framework for data visualization well into the future.

Tableau and D3 are the leading choices for data visualization production platforms. Tableau offers advanced functionality in a user-friendly format, but is expensive and offers few customizable design options. D3 is free and offers expansive design solutions that are portable across the web, but is hard to learn. Realistically, choosing between these platforms may be prompted by external factors such as time, budget, and scope of work. Designers and developers that choose a web-standards framework, however, will cultivate valuable programming skills applicable to future data visualization work, and can offer the widest range of design solutions. These are substantial advantages that proprietary solutions do not offer, making this choice a desirable choice, particularly for designers and developers.

Data visualization designers and developers face conceptual and technical challenges, which requires a multipart solution. Applying design best practices and guidelines and choosing a web-standards development framework for distribution can help create effective and portable data visualizations.

Recommendations

For designers and developers that produce data visualizations using best practices and guidelines is a recommendation due to the inherent complexity of producing web-based data visualizations. They help navigate technical and complex conceptual design issues by offering a level of assurance that data visualizations communicate effectively. Additionally, a web-standards development and distribution framework is recommended to allow the greatest degree of design flexibility, while allowing work to be portable across platforms and devices.

The following recommendations could spur innovation in data visualization production throughout the web development industry:

- The W3C should create comprehensive web-based data visualization best practices and guidelines. Little W3C documentation exists that is specific to data visualization. Consolidating and documenting these resources could be a substantial benefit to the web development and visualization community.
- The W3C should begin drafting a web-standard language specific to data visualization. A data visualization XML language, for example, could offer a human- and machine-readable semantic structure for defining data and chart types. This could simplify the programmatic aspects of visualization production.
- Data visualization experts should help d3.express be built. This framework ambitiously seeks to solve the ease-of-use vs. design flexibility problem with a hybrid design/code EDA approach. This could be the next major innovation in data visualization, but the creator has asked for help from developers. Those interested in open-source code have a perfect opportunity to contribute.

Conclusion

Data visualization has accompanied societal advancements throughout history and is one of the primary means by which we understand the world. With the widespread availability of data now and in the future, there is greater opportunity than ever to use data visualization as a tool to help us insightfully improve our lives and business. An often-repeated adage in the data science community is that *data is the new oil*, but a more fitting saying surfaced afterward: *Data is the new soil* (McCandless 2010). This implies data is not merely to be mined for exploitation, but it is the catalyst for growth, opportunity, and creative innovation.

Designers and developers understand the importance of nurturing this creative innovation. Their jobs rely on finding unique and useful ways to express and send messages that capture the imagination and create insight. Data visualizations embody this capacity to deliver beauty and utility, but to hit that mark takes knowledge, proper tools, and a clear path. Following established guidelines and best practices offers collective knowledge to help convey the message properly. Choosing open-source web-standard tools creates a framework for forging the message and firing it off. While there are certainly other ways to transmit a message, these methods best align with the core philosophies of design and development, which strive toward creative freedom in delivering innovative and open communications.

References

5000fish. 2014. "Straight Talk: Review of Tableau Software, the Pros and Cons." Accessed May 14, 2017. https://www.yurbi.com/blog/straight-talk-review-of-tableau-software-the-pros-and-cons/.

Anscombe, F. J. 1973. "Graphs in Statistical Analysis." *The American Statistician* 27, no. 1 (February): 17-21. Accessed May 5, 2017. http://www.jstor.org.du.idm.oclc.org/stable/pdf/2682899.pdf.

Amr, Tarek, and Rayna Stamboliyska. 2016. *Practical D3.js*. Berkeley, CA. Apress. Accessed April 14, 2017. https://link-springer-com.du.idm.oclc.org/book/10.1007%2F978-1-4842-1928-7.

Aparicio, Manuela, and Costa, Carlos. 2015. "Data visualization." *Communication Design Quarterly Review*, 3, no. 1. Accessed April 2, 2017. http://dl.acm.org.du.idm.oclc.org/citation.cfm?doid=2721882.2721883.

Barlas, Panagiotis, Ivor Lanning and Cathal Heavey. 2015. "A Survey of Open Source Data Science Tools." *International Journal of Intelligent Computing and Cybernetics* 8, no. 3: 232-261. Accessed April 23, 2017. http://du.idm.oclc.org/login?url=http://search.proquest.com.du.idm.oclc.org/docview/1697987721?accountid=14608. ProQuest.

Berkshire Encyclopedia of Human-Computer Interaction. 2004. Berkshire Publishing Group. Accessed April 9, 2017. http://du.idm.oclc.org/login?url=http://search.credoreference.com/content/entry/berkencyhci/data_visualization/0?institutionId=1676.

Bigelow, Alex. 2014. “Reflections on How Designers Design with Data.” Proceeding published in AVI ’14 Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces, in Como, Italy, May 27-29. ACM New York, NY. Accessed April 27, 2017. https://doi-org.du.idm.oclc.org/10.1145/2598153.2598175.

Bostock, Michael, Vadim Ogievetsky, and Jeffrey Heer. 2011. “D3: Data-Driven Documents.” *IEEE Transactions on Visualization and Computer Graphics* 17, no. 12: 2301-2309. Accessed April 2, 2017. https://doi.org/10.1109/TVCG.2011.185.

Bostock, Mike. 2017a. “A Better Way to Code.” Medium blog entry. Accessed May 19, 2017. https://medium.com/@mbostock/a-better-way-to-code-2b1d2876a3a0.

—. 2017b. “D3: Data Driven Documents.” D3js site homepage. Accessed May 18, 2017. https://d3js.org/.

Chen, Hsuanwei Michelle. 2017a. “Additional Resources and Final Remarks.” *Library Technology Reports* 53, no. 2 (April 2017): 28-30. Accessed April 29, 2017. http://eds.a.ebscohost.com.du.idm.oclc.org/ehost/pdfviewer/pdfviewer?sid=c86e35cc-cd12-4dec-99c3-a1771ad64842%40sessionmgr4007&vid=9&hid=4113.

—. 2017b. “An Overview of Information Visualization.” *Library Technology Reports* 53, no. 2 (April 2017): 5-7. Accessed April 29, 2017. http://eds.a.ebscohost.com.du.idm.oclc.org/ehost/pdfviewer/pdfviewer?sid=c86e35cc-cd12-4dec-99c3-a1771ad64842%40sessionmgr4007&vid=7&hid=4113.

—. 2017c. “Challenges and Concerns.” *Library Technology Reports* 53, no. 2 (April 2017): 26-27. Accessed April 29, 2017. http://eds.a.ebscohost.com.du.idm.oclc.org/ehost/pdfviewer/pdfviewer?sid=c86e35cc-cd12-4dec-99c3-a1771ad64842%40sessionmgr4007&vid=11&hid=4113.

—. 2017d. “Information Visualization Principles, Techniques, and Software.” *Library Technology Reports* 53, no. 2 (April 2017): 8-16. Accessed April 27, 2017. http://eds.a.ebscohost.com.du.idm.oclc.org/ehost/pdfviewer/pdfviewer?sid=c86e35cc-cd12-4dec-99c3-a1771ad64842%40sessionmgr4007&vid=7&hid=4113.

Coates, Kathryn and Andy Ellison. 2014. *An Introduction to Information Design*. London: Laurence King. Accessed April 9, 2017. http://du.idm.oclc.org/login?url=http://search.credoreference.com/content/entry/lkingid/introduction/0?institutionId=1676.

Clancy, Heather. 2014. “Why This Business Intelligence CEO Spends Oodles on Research.” Fortune.com, December 2. Accessed April 27, 2017. http://fortune.com/2014/12/02/tableau-software-ceo-research/.

DataUSA. 2017a. “DataUSA.” DataUSA landing page. Accessed May 27, 2017. https://datausa.io/.

—. 2017b. "DataUSA: Colorado.” DataUSA Colorado webpage. Accessed May 21, 2017. https://datausa.io/profile/geo/colorado/.

—. 2017c. "DataUSA: Designers.” DataUSA Designers webpage. Accessed May 21, 2017. https://datausa.io/profile/soc/271020/#skills.

Davies, Jason. 2017. “Wordcloud.” Wordcloud application webpage. Accessed May 21, 2017. https://www.jasondavies.com/wordcloud/.

DecemberCafe Studio. 2016. “Relationship Visualization.” Accessed May 21, 2017. Last modified December 24. http://www.decembercafe.org/demo/relation/.

Fox, Peter, and James Hendler. 2011. "Changing the Equation on Scientific Data Visualization." *Science* 331, no. 6018 (XXXXXX): 705-08. Accessed April 9, 2017. http://www.jstor.org.du.idm.oclc.org/stable/25790280.

Friendly, Michael, and D.J. Denis. 2001. “Milestones in the history of thematic cartography, statistical graphics, and data visualization.” Introduction. Accessed April 29, 2017. http://www.datavis.ca/milestones/.

—. 2006. “A Brief History of Data Visualization.” In *Handbook of Computational Statistics: Data Visualization*. Springer-Verlag. Accessed April 22, 2017. http://www.datavis.ca/papers/hbook.pdf.

—. 2009. *Milestones in the history of thematic cartography, statistical graphics, and data visualization*. Accessed April 2, 2017. http://www.math.yorku.ca/SCS/Gallery/milestone/milestone.pdf.

GitHub Gist. 2016. “Normalized Stacked Bar Chart.” GitHub Gist webpage, chart created by Mike Bostock. Accessed May 21, 2017. Last modified July 19. https://bl.ocks.org/mbostock/3886394.

GitHub Gist. 2017. “Multi-Series Line Chart.” GitHub Gist webpage, chart created by Mike Bostock. Accessed May 21, 2017. Last modified May 18. https://bl.ocks.org/mbostock/3884955.

Google Charts. 2017a. “Gantt Chart.” Google Charts webpage. Accessed May 21, 2017. Last modified April 25. https://developers.google.com/chart/interactive/docs/gallery/ganttchart.

—. 2017b. “Google Charts.” Google Charts landing page. Accessed May 26, 2017. https://developers.google.com/chart/.

—. 2017c. “Visualization: Column Chart.” Google Charts webpage. Accessed May 21, 2017. Last modified April 14. https://developers.google.com/chart/interactive/docs/gallery/columnchart.

—. 2017d. “Visualization: Pie Chart.” Google Charts webpage. Accessed May 21, 2017. Last modified February 23. https://developers.google.com/chart/interactive/docs/gallery/piechart.

Grammel, Lars. 2012. “User Interfaces Supporting Information Visualization Novices in Visualization Construction.” PhD diss., RWTH Aachen University, Germany. Accessed April 27, 2017. http://search.proquest.com.du.idm.oclc.org/docview/1441948038/.

Hamersky, Steve. 2016. “Tableau Desktop.” Mathematics and Computer Education 50, no. 2: 148-151. Accessed April 27, 2017. http://du.idm.oclc.org/login?url=http://search.proquest.com.du.idm.oclc.org/docview/1794881754?accountid=14608.

IBCS Association. 2015. “Association.” Association information webpage. Accessed May 20, 2017. http://www.ibcs-a.org/association.

Internetlivestats.com. 2017. “Internet Users.” Accessed May 6, 2017. http://www.internetlivestats.com/internet-users/.

Kelleher, Christa, and Thorsten Wagener. 2011. "Ten Guidelines for Effective Data Visualization in Scientific Publications." *Environmental Modelling & Software* 26, no. 6: 822-827. Accessed April 23, 2007. https://doi-org.du.idm.oclc.org/10.1016/j.envsoft.2010.12.006.

Kirk, Andy. 2012. *Data Visualization: A Successful Design Process*. Olton, GB: Packt Publishing. Accessed April 14, 2017. http://site.ebrary.com/lib/udenver/reader.action?docID=10642563. ProQuest ebrary.

Kirk, Andy. 2017. "Resources." Visualizingdata webpage. Accessed May 26, 2017. http://www.visualisingdata.com/resources/.

Klack, Moritz, and Christopher Möller. 2017. "Awesome D3." GitHub webpage. Accessed May 19, 2017. https://github.com/wbkd/awesome-d3.

Lane, Liz. 2016. "Data Visualization Best Practices." References OWL at Purdue, September 30. Accessed April 15, 2017. https://owl.english.purdue.edu/owl/resource/1014/1/.

McCandless, David. 2010. "The Beauty of Data Visualization." TED Talk website transcript. Accessed May 28, 2017. https://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization/transcript?language=en.

McCormick, Bruce H. 1987. "Visualization in Scientific Computing." *Computer Graphics*, 21, no. 6. Accessed April 22, 2017. http://www.sci.utah.edu/vrc2005/McCormick-1987-VSC.pdf.

Murry, Scott. 2013. *Interactive Data Visualization for the Web*. Sebastapol, CA: O'Reilly Media, Inc. Accessed April 1, 2017. http://search.proquest.com.du.idm.oclc.org/docview/1651823588?accountid=14608&rfr_id=info%3Axri%2Fsid%3Aprimo. ProQuest.

Ng, Cory, and Mason Pan. 2017. "Dynamic Decision Making Through Data Visualization." *Pennsylvania CPA Journal*, March 1, 2017, 5-9. Accessed April 27, 2017. http://eds.a.ebscohost.com.du.idm.oclc.org/ehost/pdfviewer/pdfviewer?sid=94b3a6f8-2589-4a46-8c43-4bc83f94184d%40sessionmgr4010&vid=1&hid=4113.

Ortiz, Santiago. 2012. "45 Ways to Communicate Two Quantities." Visual.ly design blog, July 27. Accessed May 13, 2017. https://visual.ly/blog/45-ways-to-communicate-two-quantities/.

Rist, Oliver. 2016. "The Best Data Visualization Tools of 2016." PC Mag, August 15, 2016. Accessed April 27, 2017. http://www.pcmag.com/roundup/346417/the-best-data-visualization-tools.

Rinisland, AEndrew, and Swizec Teller. 2016. *Learning d3.js Data Visualization*. Birmingham, UK: Packt Publishing. Accessed April 22, 2017. http://proquest.safaribooksonline.com.du.idm.oclc.org/book/databases/business-intelligence/9781785889042. Safari ProQuest.

Schneiderman, Ben. 1996. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations." Proceeding published in IEEE Symposium on Visual Languages, in Boulder, CO, USA, September 3-26. IEEE. Accessed May 7, 2017. https://doi.org/10.1109/VL.1996.545307.

Segel, Edward, and Jeffrey Heer. 2010. “Narrative Visualization: Telling Stories with Data.” *IEEE Transactions on Visualization and Computer Graphics* 16, no. 6 (Nov.-Dec.): 1139-1148. Accessed April 30, 2017. http://ieeexplore.ieee.org.du.idm.oclc.org/document/5613452/?reload=true.

Simon, Phil. 2014. *The Visual Organization*. Hoboken, New Jersey: John Wiley & Sons, Incorporated. Accessed May 1, 2017. https://ebookcentral.proquest.com/lib/ca/detail.action?docID=1638155. ProQuest Ebook Central.

Tableau. 2003-2017a. “Tableau.” Tableau landing page. Accessed May 27, 2017. https://www.tableau.com/.

—. 2003-2017b. “Tableau Desktop.” Tableau Desktop overview webpage. Accessed May 27, 2017. https://www.tableau.com/products/desktop.

—. 2003-2017c. “Tableau Mission.” About Tableau page. Accessed May 14, 2017. https://www.tableau.com/about/mission#founding-catalys.

Timms, Simon. 2013. *Social Data Visualization with HTML5 and JavaScript*. Birmingham, UK: Packt Publishing. Accessed April 22, 2017. https://ebookcentral-proquest-com.du.idm.oclc.org/lib/du/detail.action?docID=1389337. ProQuest Ebook Central.

Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphic Press.

W3C.org. “W3C Process Document.” Introduction webpage. Accessed May 20, 2017. https://www.w3.org/2005/10/Process-20051014/intro.

—. 2012. “Standards FAQ.” Accessed May 6, 2017. https://www.w3.org/standards/faq#std.

—. 2015a. “Data on the Web Best Practices Use Cases & Requirements.” W3C Working Group Note, February 24. Accessed April 27, 2017. https://www.w3.org/TR/dwbp-ucr/.

—. 2015b. “Data Visualization Community Group Meetup Event Minutes.” September 20. Accessed April 27, 2017. https://www.w3.org/2015/09/20-datavis-minutes.

—. 2016. “Main Page.” W3C Data Visualization Community Group wiki, July 31. Accessed April 16, 2017. https://www.w3.org/community/datavis/wiki/Main_Page.

—. 2017a. “All Standards and Drafts.” Standards list. Accessed May 14, 2017. https://www.w3.org/TR/.

—. 2017b. “Data on the Web Best Practices.” W3C Recommendation, January 31. Accessed April 17, 2017. https://www.w3.org/TR/dwbp/Overview.html#ProvideComplementaryPresentations.

—. 2017c. “Data Visualization Community Group.” Data Visualization Community Group home page. Accessed May 14, 2017. https://www.w3.org/community/datavis/.

Watson, Hugh J. 2017. “Data Visualizations, Data Interpreters, and Storytelling.” *Business Intelligence Journal,* no. 22, 1. Accessed April 27, 2017. http://eds.a.ebscohost.com.du.idm.oclc.org/ehost/pdfviewer/pdfviewer?sid=75910663-14ed-457d-bf6c-f13b80d17dd1%40sessionmgr4007&vid=2&hid=4113.

WHATWG. 2017. “FAQ.” WHATWG FAQ webpage. Accessed May 20, 2017. Last modified March 13. https://wiki.whatwg.org/wiki/FAQ.

Wilkinson, Leland. 2005. *The Grammar of Graphics*. New York, NY: Springer Science+Business Media. Accessed April 23, 2017. https://doi.org/10.1007/0-387-28695-0. Springer Link.

Yau, Nathan. 2011. *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics (1)*. John Wiley & Sons, Incorporated. Accessed April 15, 2017. http://site.ebrary.com/lib/udenver/detail.action?docID=10484696.

Zeldman, Jeffrey, and Ethan Marcotte. 2010. *Designing with Web Standards*. Berkeley, CA: New Riders. Adobe PDF eBook.