II

THE TOOLS

Which software should you use to build data graphics?

To incorporate visualization into your practice, you must know which tools are best suited for the visualization task. The tools available for building visualizations fall into four categories: 1) basic productivity applications, 2) visualization software, 3) business intelligence tools, and 4) developer-based packages. Getting started with each is very straightforward. The difficulty comes in identifying what you want to visualize and ensuring your data is in the correct format. This chapter presents the options for creating data graphics and criteria for evaluating your software choices.

2.1 Basic Productivity Applications

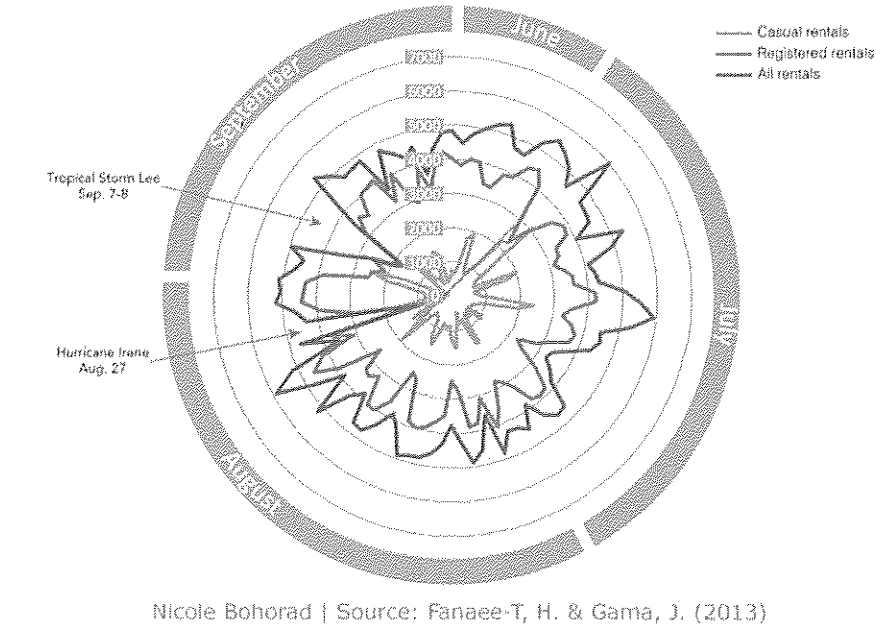
Common productivity tools are good enough for most visualization tasks. With Excel or the iWork suite, you can create basic chart types: bar, pie, line, and scatter plots in addition to more sophisticated displays such as stacked area and radar charts. Google Charts are also interactive and web-based.

MICROSOFT EXCEL

Microsoft Excel provides a sophisticated set of static charting options. These include column and horizontal bars, line, pie, area, radar, scatterplot, and spark lines. Excel is designed for working with data. Excel supports the pre-processing data and visualization in the same application. Charts created in Excel are easily ported to PowerPoint and Word. Excel charts require customization to adhere to many of the design standards presented in this book. For instance, the default charts contain unnecessary non-data elements such as gridlines, tick marks, and borders.

If you use Excel exclusively in your practice, consider creating chart templates to which you can apply your own chart style. http://becomingvisual.com/portfolio/excel

See Figure 2.1 for an example of a radar chart created in Microsoft Excel.



Nicole Bohorad | Source: Fanaee-T, H. & Gama, J. (2013)

Figure 2.1 A radar chart created in Microsoft Excel

Managers may do their analyses in Excel but present their charts in PowerPoint. There are additional plug-ins, for PowerPoint that extend the chart features and options. These include charting, layout, and additional data formatting features. Learn more at: http://becomingvisual.com/portfolio/powerpoint.

iWORK

Apple's own productivity suite, iWork, which includes Pages, Numbers, and Keynote, offers basic 2D and 3D charts in addition to animated vertical and horizontal bars, scatter plots, and bubble charts.

As with Excel, the default charts in iWork require that you reformat the default features to conform to your own aesthetic. The color templates provided simplify the process of removing non-data elements that may interfere with interpretation of the data.

See Figure 2.2 for an example of a chart created in iWork's Pages.

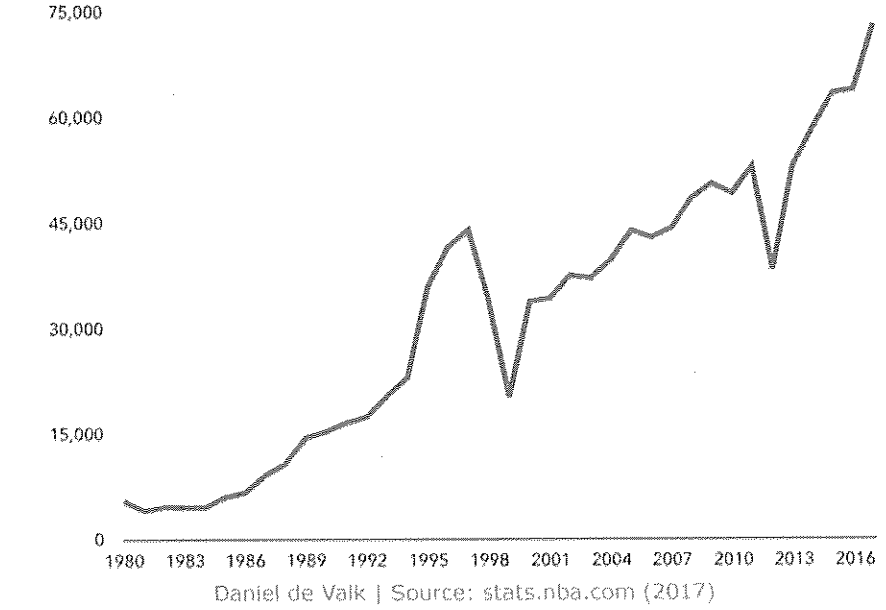


Figure 2.2 A time series chart created in iWork's Pages of the number of three-point shot attempts in the NBA

Users who work with data in Excel can easily import their data to iWork's Numbers, Pages, or Keynote. PC users can use Apple's iWork productivity suite using iWork for iCloud.

GOOGLE CHARTS

Google offers a free and open option for creating a variety of data graphics. The charts integrate seamlessly with the Google Apps suite (Docs, Sheets, and Slides). Google Charts offers more options than Excel or iWork including interactive, animated, and geospatial data graphics. For more robust reporting and visualization tools in one, see Google's Data Studio.

For a gallery of chart possibilities, go to: http://becomingvisual.com/portfolio/googlecharts

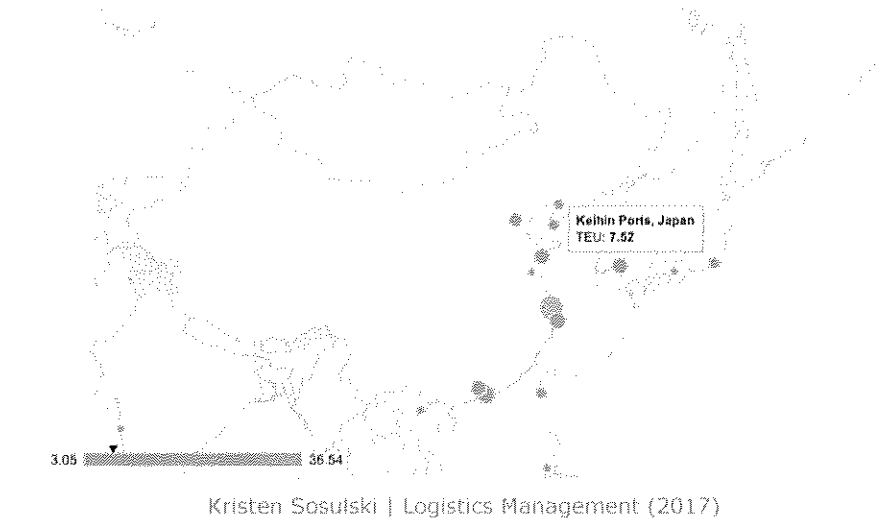


Figure 2.3 A GeoChart with bubble markers created in Google Charts

See Figure 2.3 for an example of a GeoChart (point map) created in Google Charts. The chart shows the locations of the world's top container shipping ports, measured in millions of TEUs (20-foot equivalent units).

Microsoft Excel, iWork, and Google Charts all enable you to create static charts. Interactive web-based charts can be created using Google Charts. These charts require coding in JavaScript and knowledge of HTML. Use basic productivity tools when working with single tables of data in. csv format or. xlsx. When files sizes approach a gigabyte, they become unmanageable using these tools. Large data sets are typically hosted externally, in the cloud, and queried using specialized business intelligence tools and programming platforms.

Follow the tutorial in Exercise 1 on page xx to create the chart above.

2.2 Visualization Software

Data visualization software applications are ubiquitous. These applications focus on usability through a drag and drop interface. They are designed for everyone from novices to expert visualization designers

and analysts. Tableau Desktop and many other specialized data visualization software packages (e.g., QlikView, Domo) offer an interface for visualizing data. These applications offer a full-range of data graphics from basic charts to maps. These tools feature Interactive, static, animated, multiple-dimensional linked charts, and dashboard displays.

TABLEAU DESKTOP

Tableau is one of the leading data visualization software packages. It is designed to integrate with a range of data sources and file types. For example, you can import the basic. xls,. cvs, or. txt files and connect them to live data sources on Tableau Server, Oracle, Amazon, Cloudera, etc.

The drag and drop interface makes it easy to start visualizing your variables. The design of the charts and tables produced in Tableau are inspired by the *Grammar of Graphics* by Leland Wilkinson. Therefore, the graphics need little refinement in terms of the 10 design standards discussed in Chapter V–THE DESIGN. Interactive, spatial, animated, linked, and dashboard displays are all possible with Tableau Desktop.

Tableau has robust capabilities for filtering, grouping, clustering, aggregating, and disaggregating variables. Some programming knowledge is required for complex analytical tasks. As one would create a formula in Excel, users of Tableau can create new fields and perform mathematical computations.

Tableau workbooks easily publish to the web with Tableau public, a free service. Tableau workbooks can also be shared securely across an organization with Tableau server, an additional service. Figure 2.4 shows an interactive Tableau data graphic with an option to filter by year.

ARCGIS

There are specialized software packages that focus on specific data graphics such as geospatial displays. ArcGIS is a mapping platform available for desktop or online. It is a platform that visualizes and analyzes most types of spatial data. There are many types of ready to use base maps, demographic and lifestyle maps, historical maps, and layers for boundaries and places, landscapes, oceans, earth observations, transportation, and urban systems. ArcGIS offers 3D mapping as well. See Figure 2.5 for an example of a map of shipping container volume by country using ArcGIS online.

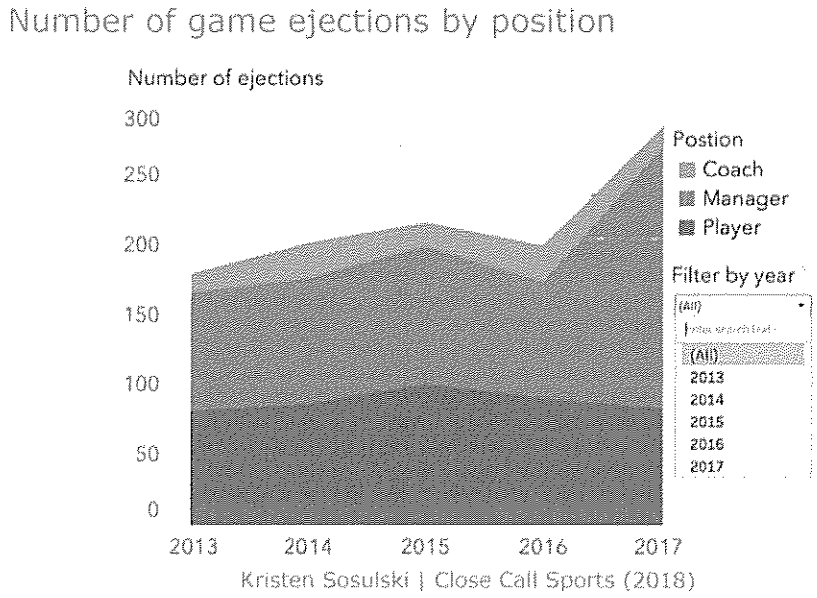


Figure 2.4 An interactive data graphic with a drop-down menu created in Tableau Desktop

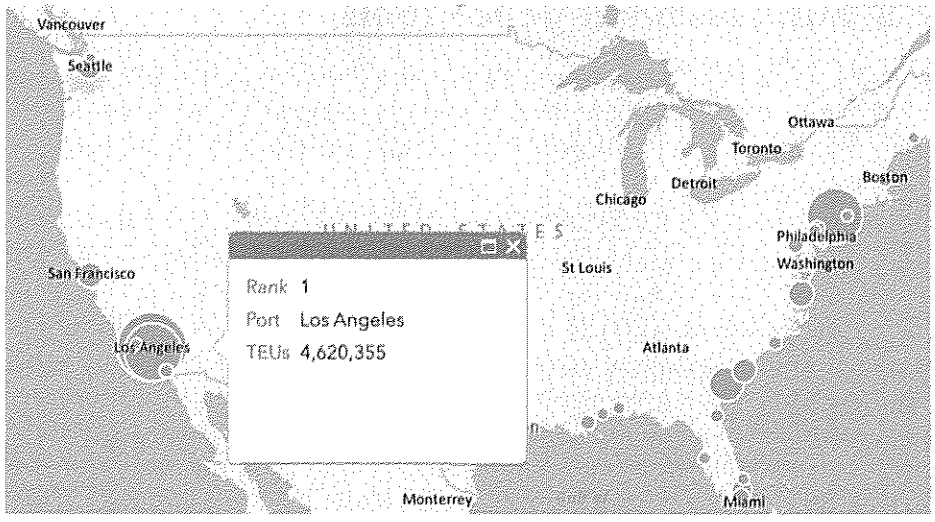


Figure 2.5 A simple bubble map created in ArcGIS of the top 30 shipping ports in the United States by container volume

CURIOUS HOW TO VISUALIZING BIG DATA IN 3D OR USING VIRTUAL REALITY?

QuantumViz is a big data visualization software company that allows data scientists and analysts to find insights in massive datasets and create amazing data stories in 3D, VR, or AR. For example, they helped develop a geo-visualization of container ships movement through the Panama Canal. In addition, their storyboard feature allows users to build a data-story that includes 360 images and videos, such as a 360 image showing a cargo shipping passing through a gate. All of this can be experienced in VR as well. QuantumViz's revolutionary tool transforms data visualizations into data experiences.

Learn more: http://becomingvisual.com/portfolio/quantumviz

The level of design and analytical sophistication with software like Tableau and ArcGIS is much higher than the basic productivity tools such as those offered by Excel.

2.3 Business Intelligence Tools

> At the next level, modern technologies have enabled the use of more dynamic and interactive business graphics, such as real-time dashboards and charts that update automatically as the data changes.
>
> (FORRESTER RESEARCH, 2012, p. 4)

Forrester Research (2012) describes these business intelligence tools as the next wave of advanced visualization software. They provide the ability to show dynamic content, visual querying, multiple dimensional-linked visualizations, animated visualizations, personalization, and alerts based on changing data. Examples of these business intelligence tools include IBM Watson Analytics, SAS, TIBCO's SpotFire, and Microsoft's Power BI. All require a paid subscription or license. Each provides an interface for data querying and exploration. Most of these tools offer visualization recommendations as well.

IBM WATSON ANALYTICS

Watson Analytics provides a platform for users to explore their data, ask questions of their data, and create data graphics. Watson is unique in that it guides the user in selecting the best method of inquiry to learn about the data. A set of exploratory visualizations called spirals are presented (see Figure 2.6). The spirals show the drivers of the target variable, for example, what factors contribute to the total dollar amount spent by a customer in a casino? The user can then build data graphics with recommendations from Watson.

2.4 Programming Packages

For developers, analysts, and designers who want to visualize data in their own programming environment, there are several contenders. Most programming languages have data graphic packages. Python and R have a sophisticated set of libraries or packages for data visualization. In addition, there are numerous JavaScript libraries for web-based data graphics.

R AND RSTUDIO

R is a free open-source statistical programming language. There are several packages that are used for visualization in R.

These include:

- graphics
- ggplot2
- car
- lattice
- ndtv
- ggvis
- plotly
- shiny

R is capable of both data analysis and data graphics. However, the default chart output requires refinements to aesthetic elements (such as colors). For example, look at Figure 2.7. This plot lacks a title. The gray background and gridlines do not add any information to the display. The black dots are harsh. They can be changed to a lighter color

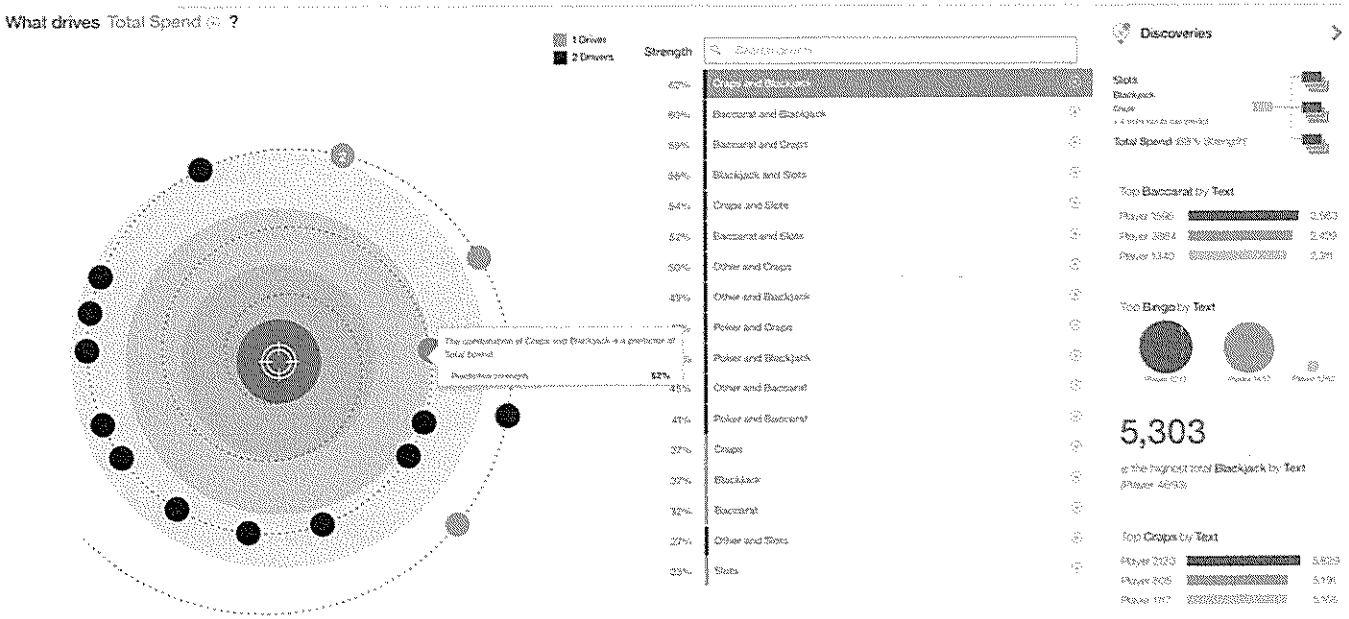


Figure 2.6 The spiral visualizations generated from Watson that shows the predictive strength of a given variable (denoted by gray and black bubbles) to the target variable (the center of the spiral)

Bicycle rentals are positively associated with warmer temperatures.

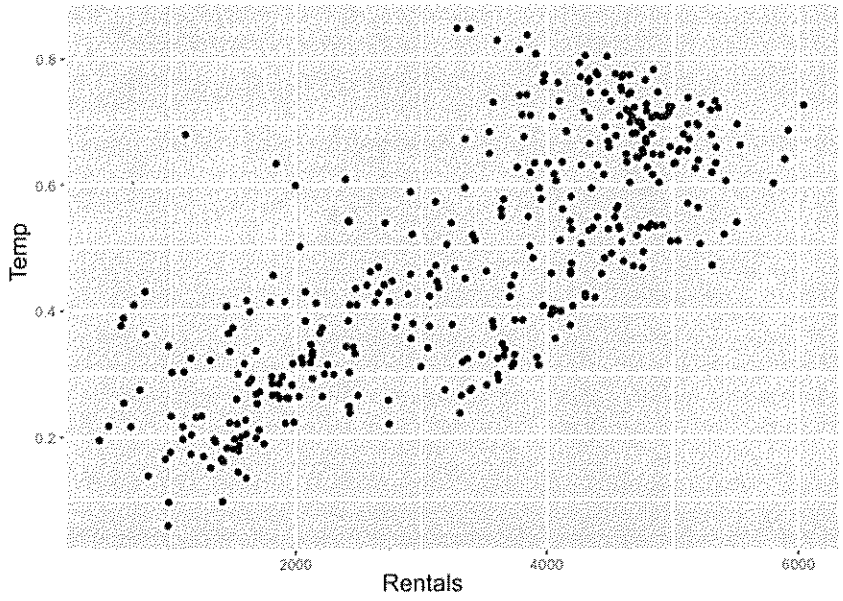
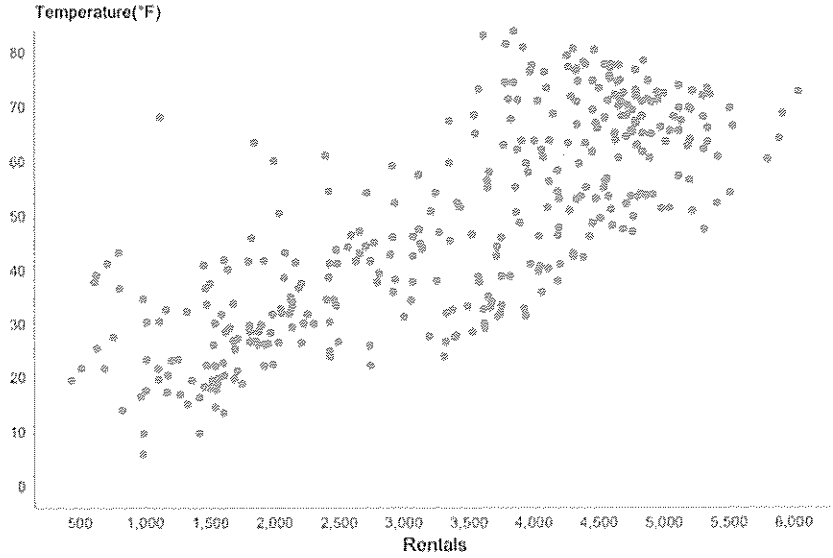


Figure 2.7 A default scatterplot produced by the ggplot2 package in R

once the background is removed. Also, notice how the x-axis begins at 2000 and the scale for temperature is undefined. Fortunately, the plot can be revised as shown in Figure 2.8. Note the x and y scales, the labeling of axes, title, data source, and simple white background with green points.

The ggthemes package can be used to customize the non-data elements of graphs produced in ggplot2. The *bw()* theme was applied to Figure 2.8 for a simple black and white color scheme. The point color was changed from black to green.

The **ggvis** package in R produces graphs that apply many more of the accepted data visualization design principles. It leverages some of the interactive components of the **shiny** package for interactive web applications. Shiny applications can be published to the web and include animated or interactive visualizations. R's geomapping capabilities are somewhat limited in comparison with ArcGIS.



Kristen Sosulski | Source: Fanaee-T, H. & Gama, J. (2013)

Figure 2.8 A customized scatterplot produced by the ggplot2 package in R

PYTHON

Python is a powerful programming language with stellar data cleaning and data manipulation capabilities. Python's **matplotlib** package is used to plot basic charts. Packages such as Seaborn yield high-quality data graphics.

Some of Python's other data visualization libraries include:

- geoplotlib
- Bokeh
- Pandas
- Altair
- ggplot
- pygal
- plotly

JAVASCRIPT

JavaScript is a web-based scripting language that is used in combination with HTML.

Some JavaScript libraries include D3, rCharts, HighCharts, charts.js, dimple.js, and processing.js. These libraries allow users to create

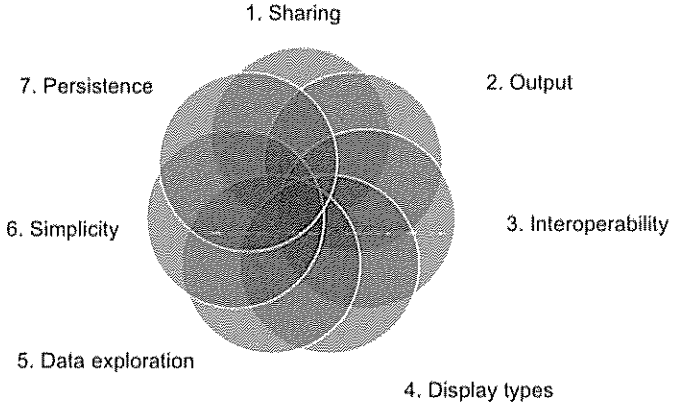


Figure 2.9 Criteria for evaluating software for visualizing data

highly sophisticated web-based visualizations. The libraries are freely available and interact with plotly as the platform for displaying the charts and graphs. The learning curve is very steep. Skills in working with HTML and JSON data are required.

Check out more options for data visualization JavaScript libraries: http://becomingvisual.com/portfolio/javascript

2.5 A Criteria for Selecting Tools to Build Data Graphics

A SOFTWARE EVALUATION CHECKLIST FOR DATA GRAPHICS

When evaluating a new data visualization tool, consider the following (aside from price):

- ☐ Sharing: can others view and edit your visualization and analysis? The ability to share your charts and graphs with others promotes collaboration on data visualization tasks.
- ☐ Output: can you publish visualizations to the web, create high-quality print graphics, and embed them in other applications? The ultimate destination of your visualization will dictate your tool

choice. For example, if your audience is viewing your graphs online, you may want to make them interactive to facilitate exploration.

- ☐ Interoperability: how easily can you connect to other data sources? For example, does the software allow you to import diverse file types, such as. xlsx,. csv, or. txt, and also link to databases?
- ☐ Display types: what types of visualizations do you need to build? Maps, networks, and text-based visualizations are not available in every tool.
- ☐ Data exploration: do you need a tool to explore your data and present it visually? Features such as visual querying are not standard in every tool.
- ☐ Simplicity: do you want to create charts and graphs quickly? Some tools require a steep learning curve, even to build a simple bar chart.
- ☐ Persistence: do you think you'll need to revise the visualizations you create? Choose a tool from a reputable company that you think will be around for a while.

Before using a free data visualization tool, know how your data are stored. A major drawback with most free apps is that they require you to make your data public in exchange for being a freemium member. Consider looking into a premium membership to protect your data.

There is no one-size-fits-all solution to visualizing geospatial, categorical, time series, statistical, and network data as static, animated, or interactive displays for the desktop, web, or a presentation.

Select the tools that work best for your workflow. If you do all of your analysis in Excel, consider learning the nuances of the chart options for the basic chart types, such as bar, pie, and line charts. Then, explore other tools, such as Tableau, to easily create maps, interactive data graphics, and animated charts.

Interview With a Practitioner

I interviewed Christian from Viant, who described how he uses data graphics to support his work.

Kristen Sosulski (KS)

Christian Theodore (CT)

KS:

Who are you and what do you do?

CT:

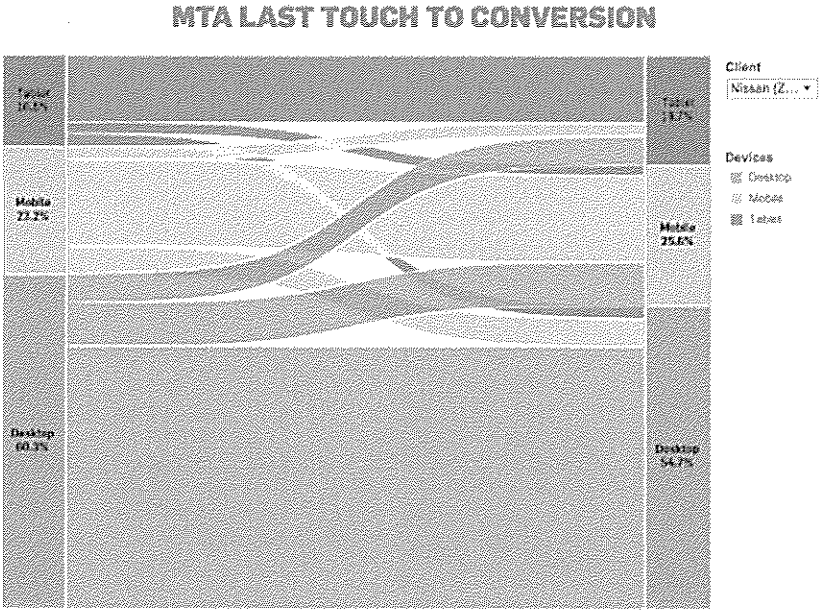
My name is Christian Theodore. I work as a data analyst on the Attribution Platform Team for Viant Inc.

KS:

How do you use data visualization in your practice?

CT:

I use data visualization to help end users move from the complex to the simple. My primary function is to build innovative client-facing products at Viant while applying the principles of data visualization. At Viant, we offer a suite of solutions to help our clients: 1) understand their audience segments, 2) gains insights on the performance of their digital marketing campaigns, and ultimately 3) use those insights towards better decision making.



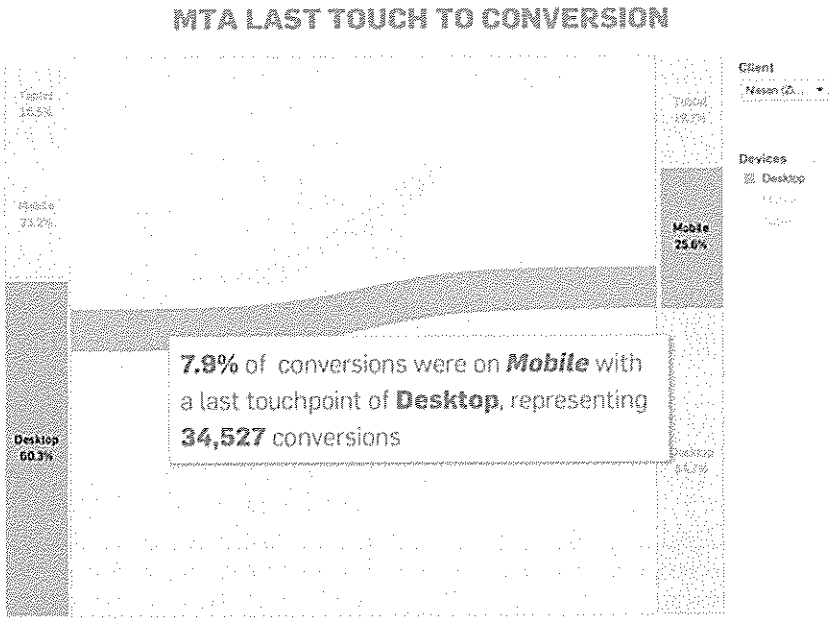
As a recent example of how I apply data visualization to my practice, I used a Sankey Diagram to illustrate how different devices drive conversions after users see a client's ad. The core idea was to illustrate how impressions across multiple devices can be attributed to a user's conversion (a "conversion" can be defined as the user taking whatever specified call to action the marketer intended, via the ad).

The chart is composed of three main elements (from left to right):

1. The Source Stacked Bar Chart
2. The Flow Ribbons
3. The Destination Stacked Bar Chart

In detail, here is what each element represents:

1. A "source" bar chart, which displays the proportion of last-touch (that is, immediately prior to conversion) ad impressions accounted for by device type.
2. The "flow ribbons," which show the percentage of conversions attributed by device (see screenshot below). The branches in each ribbon show how the source device differs from the destination device. In the below example, the highlighted strand indicates that **7.9% (34,527) of people who last saw an ad on a desktop computer converted on their mobile phone.** Overall, the volume of conversions (indicated by the thickness of each strand) shows that the majority of people convert on the same device on which they last saw an ad**, but a significant minority convert on a different device**.
3. The "destination" bar chart gives an aggregate picture of the overall distribution of conversions by device.



KS:

What insights are evidenced by the graph provided?

CT:

The chart delivers several powerful insights:

1. It illustrates the direct pathway users follow from the last touchpoint to conversion. This is valuable for advertisers, as it helps to answer the question:
   Do Customers Typically Convert on the Same Device on Which They Last Saw the ad?
2. It allows advertisers to make more informed decisions about their budgeting strategies. In cases where ad spend is skewed to a particular device, the chart helps to validate that strategy is best, or can suggest that the advertiser change its strategy, based on the devices where conversions occur For example, a desktop may drive the majority of conversions, but a mobile device may still provide value, and therefore not be discounted as merely a touchpoint (that is, **investment in mobile is still a good strategy**).

KS:

How did you create it? What was the data? What was the software? What would have been the alternative?

CT:

This chart was created using Tableau Desktop, leveraging its built-in Data Densification feature, creating Padded Bins, and finally, incorporating a mathematical function to trace the path of each conversion.
The data was obtained from a sample of data culled from a few of Viant's MTA (Muti-Touch Attribution) client reports.

This example shows how data graphics are used to show the pathways of conversions for specific Viant clients. These pathways are the visual evidence that inform decisions about which platforms to use to reach Viant's client's target market.

Software is not magic. As evidenced by the Viant case, there is quite a bit of technical detail involved in creating sophisticated data graphics. Use the tools that best suit you and your work environment. Use the checklist presented to guide you in the software selection process. There are many visualization software packages (desktop and cloud-based) available. However, with the cloud-based solutions, there may be data privacy issues. In addition, many software packages are not equipped to handle large data sets.