III

THE GRAPHICS

Which chart works best to show my data and insight?

Selecting the right visualization to present your data is complicated: the number of chart choices can distract you from the goal of communicating the key insight. This chapter reveals how the insight and data drive your selection of the right chart.

Each type of chart is designed to show a type of data in a particular way. For example:

- Horizontal bar charts show rank well by ordering bars from largest to smallest.
- Line charts convey a change over a specified period of time, such as the unemployment rate per month over a 12-month period.
- Point maps effectively demarcate precise locations, such as the address of each public school in a district.
- Filled or choropleth maps allow for the comparison of regions, such as the GDP of each African country. Each region is filled with a shade. The darker the shade the higher the value.

Two maps are presented in Figure 3.1. Chart A is a filled map showing the locations of recycling bins in NYC by an aggregated measure (zip code). Chart B shows the actual location of each recycling bin.

Which map communicates the data best?

The short answer is: it depends. Do you want to show the aggregated total number of recycling bins per neighborhood as shown in Chart A? Or would you rather show the concentration of the location of individual recycling bins within neighborhoods as shown in Chart B? As a data graphic designer, you decide.

The data graphics choices presented in this chapter will point you in the right direction based on your data. The type of chart you ultimately select is limited by the type of data. Data comes in many forms. Forms include categorical, univariate (a single variable), multivariate (more than one variable), geospatial, time series, network, or text. Certain charts display comparisons, distributions, proportions, relationships, locations, trends, connections, or sentiment better than others. Refer to Table 3.1 for guidance on the types of insights you can visualize based on your data.

This table should serve as a handy reference throughout the book.

Selecting the right chart
There are many available resources to guide you in determining the right chart for your data. Learn more at: http://becomingvisual.com/portfolio/chartpickers

Chart A: Recycling bins grouped by zip code

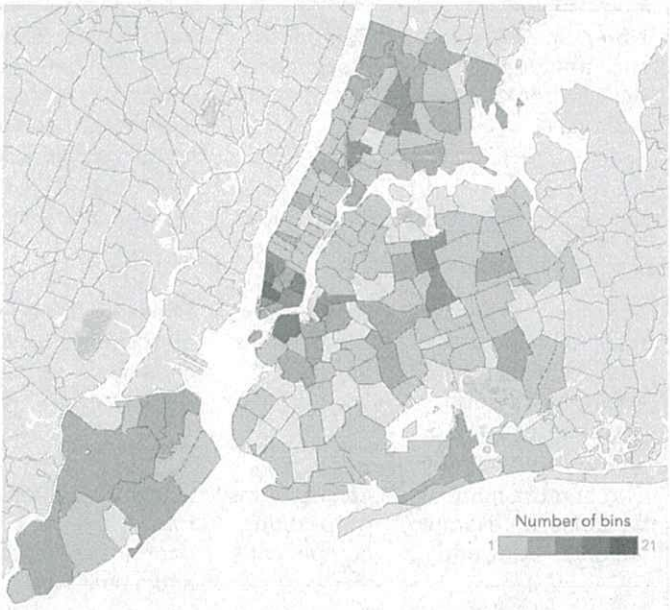
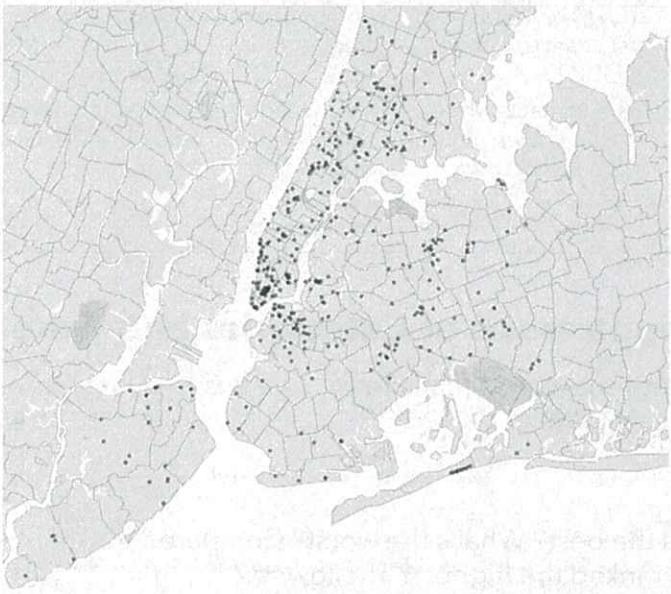


Chart B: Recycling bins grouped by individual locations



Kristen Sosulski | Source: NYC Open Data (2014)

Figure 3.1 Chart A and Chart B use the same data two different ways.

Table 3.1 The general data insight that corresponds to each data classification

Data	Example	Insight	Chart type
Categorical	Non-numeric data such as types of movies, books, or authors.	Comparisons, proportions	Vertical bar, column bar, horizontal bar, and bullet charts Pie, stacked bar, stacked 100% bar, stacked area, stacked 100% area, and a tree map
Univariate	One numeric variable, such as book price	Distributions, proportions, frequencies	Histogram, density plot, and a boxplot
Geospatial	Specific locations marked by the latitude and longitude, regions coded by zip code, city, state, country, or county boundaries	Locations, comparisons, trends	Choropleth filled-map, bubble map, point map, connection map, and isopleth map
Multivariate	Two or more numeric variables, for example, weight, height, and IQ	Relationships, proportions, comparisons	Scatterplot, scatterplot matrix, bubble, parallel coordinates, radar, bullet, and a heat map.
Time series	Years, months, days, hours, minutes, seconds, or date	Trends, comparisons, cycles	Line chart, sparkline, area, stream graph, as well as bubble, stacked-area, and vertical bar charts.
Text	Single words or phrases, such as keywords from restaurant reviews on Yelp	Sentiment, comparisons, frequency	Word cloud, proportional area chart using size bubbles or squares, histogram, and bar chart
Edge lists or adjacency matrices	Who contacts whom or who knows whom in a network	Connections, relationships, tie strength, centrality, interactions	Undirected network diagram and directed network diagram

3.1 Comparisons of Categories and Time

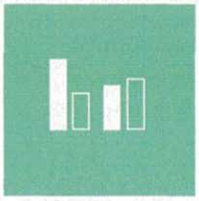
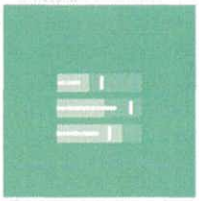
Table 3.2 presents charts that compare values or quantities either over time or to each other.

Questions:

1. What's the best? What's the worst? Compared to what?
2. Who's ranked the highest? The lowest?
3. How does performance compare to the target or goal? For example, did total sales exceed the forecast?

Insight: use comparisons to illustrate the similarities and differences among categories. This includes the minimum value, maximum value, rank, performance, sum, totals, counts, and quantities.
Data: aggregated categorical data, such as the number of books sold by author. Time series data can be shown as a categorical variable. For example, each year can be a category.
Chart options: vertical bar, column bar, horizontal bar, and bullet charts.

Table 3.2 Chart types to present categorical data

Chart type	Description and design considerations
Vertical bar 	Bars are arranged vertically on the x-axis. Each bar represents a category or sub-category. The bar height measures the quantity (count) or sum. • Keep bars the same color and shade when they measure the same variable (Wong, 2010). • Use a zero baseline for the y-axis. • Show negative values below the baseline. • Keep the width of the bar about twice the width of the space between the bars (Wong, 2010).
Column bar 	Column bar charts present two series for each category. • Use different color shading for each series. • Shade bars from lightest to darkest (Wong, 2010).
Horizontal bar 	Bars are arranged horizontally, rather than vertically. • Best used for ranking, such as first place, second place, third place. • Arrange bars in descending order, from largest to smallest.
Bullet 	Bullet charts display performance of a variable as a horizontal bar compared to a target or goal, represented by a vertical line. For example, a bullet chart could show whether the actual sales for a given period(s) are above/ below target sales. The performance measure (horizontal bar) overlays several shaded rectangles that represent qualitative ranges (e.g., 40% to the target goal, to indicate the performance progress).

3.2 Distributions

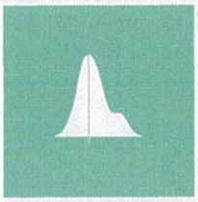
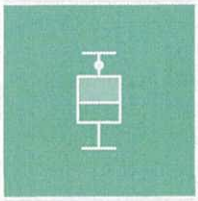
Table 3.3 presents options for showing possible values (or intervals) of the data and how often they occur. These types of charts can reveal the minimum and maximum values, median, outliers, median, frequency, and probability densities.

Questions:

1. What are the highest, middle, and lowest values?
2. Does one thing stand out from the rest?
3. What does the shape of the data look like?

Insight: use to distributions charts reveal outliers, the shape of the distribution, frequencies, range of values, minimum value, maximum value, and the median.
Data: univariate or a single numeric variable.
Chart options: histogram, density plot, and a boxplot.

Table 3.3 Chart types for showing distributions

Chart type	Description and design considerations
Histogram 	Histograms show frequencies of a single variable grouped into bins or frequency ranges on the x-axis. The y-axis of the histogram shows the frequency count or percentage. • A large bin size can obscure the data. • Adjust the size of the bins to best reveal the shape of the frequency distribution.
Density plot 	Density plots show probability densities and the distribution of a single variable. The area under the curve emphasizes the shape of the distribution of data. Annotate the mean to draw attention to the center of the distribution.
Boxplot 	Boxplots show the range of a single variable including the minimum, 25th percentile, 50th percentile, median (not the average), 75th percentile, and the maximum value. Boxplots are helpful to spot outliers.

3.3 Proportions

Table 3.4 presents options for displaying individual parts of a whole. This enables comparisons among subcategories by evaluating relative proportions, for example, demographics by neighborhood.

Questions:

1. What are the parts that make up the whole?
2. What part is the largest or smallest?
3. What parts are similar or dissimilar?

Insight: use to show summaries, similarities, anomalies, percentage related to the whole (by category, subcategory, and over time).
Data: single categorical variable with subcategories, two or more variables. A time dimension can also be included.
Chart options: pie, stacked bar, stacked 100% bar, stacked area, stacked 100% area, tree map, and doughnut chart.

3.4 Relationships

Table 3.5 presents options for displaying multivariate data. These charts show how one or more variables relates to other variables. For example, how do sales affect profitability by region?

Questions:

1. Is the relationship positive, negative, or neither?
2. How are x and y related to each other?
3. What makes one group or cluster different from another?

Insight: use to show outliers, correlations, positive, and negative relationships among two or more variables.
Data: two or more numeric variables.
Chart options: scatterplot, scatterplot matrix, bubble, parallel coordinates, radar, bullet, and a heat map.

3.5 Locations

Geospatial visualizations require data that corresponds to geography (e.g., latitude and longitude). Table 3.6 presents the options for showing these location types. These geospatial displays can identify places, population differences, concentrations, and distances.

Table 3.4 Chart options for showing proportions

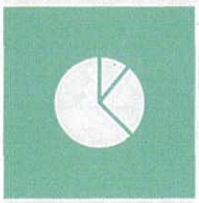
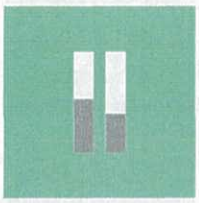
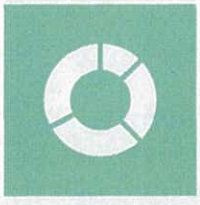
Chart type	Description and design considerations
Pie 	Pie charts show proportions within a whole. The slices are subcategories of a single category. Slices add up to 100% or 1. • Avoid using pie charts if all the slices are similar in size. • Limit pie charts to eight slices or less (Wong, 2010). • Label directly on the pie slices, rather than using a legend. • Keep pie slices the same color. Use the whitespace between slices to differentiate the slices.
Stacked bar 	Stacked bar charts show proportions and quantities within a whole category. They show absolute and relative differences. • Limit the number of subcategories to four or less. • Use stacked bars that add up to 100% to show the relative differences between quantities within each group.
Stacked area 	Stacked area charts highlight the absolute and relative differences between two or more series. They are line charts with the area below the line filled in with color. To show relative differences use a 100% stacked area chart. Label each series directly, if possible over using a legend.
Tree map 	Tree maps show parts of the whole by using nested rectangles. Each rectangle is designated a size and a shade of a color. This enables you to emphasize both the importance (usually shown by size) and urgency (usually represented by color) of a data point. • Used often for portfolio analysis to highlight similarities and anomalies. • Usually require interactivity such as mouse-over, to read the subcategory labels for the smallest rectangles. • This chart type is best used for analysis and exploration rather than presentation.
Doughnut 	Doughnut charts present proportions of a whole through slices of a doughnut shaped graphic It is just a pie chart with the center missing. This type of chart can contain multiple series, represented as doughnuts arranged inside one another.

Table 3.5 Chart options for showing relationships between two or more variables

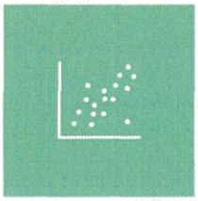
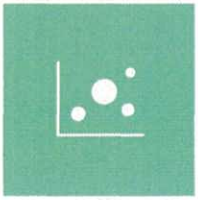
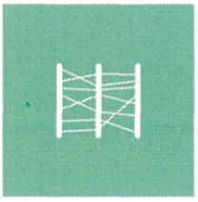
Chart type	Description and design considerations
Scatterplot 	Scatterplots show relationships between two variables. For example, they show the change in x given y. • Use to show positive or negative correlations, or linear and nonlinear relationships between variables. • Labeling of every data point reduces readability but increases interpretation.
Scatterplot matrix 	Scatterplot matrices help identify a correlation between multiple variables. It makes it easy to observe the relationship between pairs of variables in one set of plots. This chart type is best reserved for exploration versus presentation.
Bubble chart 	A bubble chart is a scatterplot that shows relationships between three or four variables. The position of the bubble shows the relationship between the x and y variables. • The bubble size is based upon a numerical variable, such as population, or sales. • The bubble color is best reserved for categorical data, such as region. • Bubble charts are best when the bubble sizes vary significantly.
Parallel coordinates 	Parallel coordinates map each column in a data table as a vertical parallel line with its own axis. Each observation (row) is represented by a point on the parallel line. That point is then connected to the next point on the next parallel line by a horizontal line. • Use the technique of highlighting the lines that touch any number of values in either of the categories, called brushing, to provide data context while focusing on select series. • This chart type is best reserved for exploration over presentation.

Table 3.5 (Continued)

Chart type	Description and design considerations
Radar 	Radar charts compare multiple numerical variables. They show which variables have similar values, and to spot outliers, high values, and low values. Each variable is provided its own individual axis, but the axes are arranged radially. Every observation connects to form a shaded polygon. • Limit the number of variables to reduce the number of axes to increase readability. • Scaling is affected when variables have dissimilar minimum and maximum ranges.
Heat map 	A heat map is a graphical representation of a table of data. The individual values are arranged in a table/ matrix and represented by colors. Use grayscale or gradient for coloring. Sorting of the variables changes the color pattern.
Small multiples 	A series of similar graphs that use the same scale. This allows for easy comparisons between variables. A single chart represents each categorical variable, such as sales personnel; the individual charts are grouped together on a single display. • Allows easy comparisons by using the same scale for each chart • Avoid showing too much detail in any individual chart.

Questions:

1. Where can the most or least be found?
2. How does one area compare to another?
3. What is the distance from one place to another?
4. How does a variable change by location?

Insight: use to demonstrate similarities and differences by location, density, distance, and counts (such as population).
Data: latitude and longitude, zip codes, census tracks, cities, states, countries, and regions.

Table 3.6 Chart options for showing locations

Chart type	Description and design considerations
Choropleth or filled maps 	Choropleth maps fill regions with color. A color gradient and density distinguishes regions from one another. Use to compare different regions such as continents, countries, states, territories, zip codes, or census tracks. Provide a legend. Keep the gradient of colors within a limited range. This will allow the reader to easily compare the regions.
Point map 	Point maps show a specific location. These dots can vary in size, form, or color. Point maps illustrate density when the individual locations are easily distinguishable. Too many points can obscure the location. Consider the size of the points and the labeling of the points.
Symbol or bubble map 	Symbol maps are point maps that use different sized bubbles or shapes to mark a location. These symbols are sized by a certain variable. Too many or too large bubbles can obscure the locations referenced.
Connection or path maps 	Connection maps graph a line from one or more points to another. Use to show distances or pathways between one or more locations. Use high contrasting colors for the map projection and the lines that connect the points. Avoid too many overlapping lines.
Geographic heat map (Isopleth) 	Isopleth maps show gradual change over geography. This technique uses a color value (lightness/darkness) and hue to show density. The color value is not constrained by boundary lines (e.g., such as zip code). Use for events that are continuous and unbounded (e.g., such as temperature).

Chart options: use a choropleth-filled map for comparing regions; bubble or point maps to mark a location while showing quantities or categories; connection maps to show distances; and isopleth maps to show variables, such as temperature, that are unrestrained by boundaries (e.g., city, state, zip code).

About geocoding
Geocoding is the process of transforming a description of a location, such as an address, to a location on the Earth's surface. This will take the form of numerical coordinates such as latitude and longitude. Reverse geocoding is the process of converting numerical coordinates into a description of that location. Learn more at: http://becomingvisual.com/portfolio/geocode

About Map Projections
Geographical data is plotted on a map projection. The most common are Albers and Mercator. Select the most appropriate map projection for your data. Use the Mercator projection to preserve angles and shapes in small areas. The Mercator projection is also good for presenting connection maps that show routes and paths for directions. The standard Albers projection can be used to present U.S. census or government data. This equal-rectangular projection is a good option for thematic world mappings. Learn more at: http://becomingvisual.com/portfolio/mapprojections

3.6 Trends–Showing Comparisons Over Time or Composition Over Time

Table 3.7 presents several charts to show how one or more variables changes over time; for example, the growth of Internet usage by country over time or number of Twitter posts per minute over a 24-hour period. Relationships, locations, proportions, comparisons, and distributions can have a time dimension.

Questions:

- What changed today from yesterday?
- How does time of year affect sales, results, outcomes, etc.?
- What times are the most popular? Least popular?

Insight: change over time, cycles, or comparisons over time.
Data: time dimension such as year, month, day, hour, minute, second, date, quarter, season, century, decade, etc.
Chart options: line chart, sparkline, area, and stream graph. The bubble, stacked-area, and vertical bar charts are options as well.

Table 3.7 Chart options for showing trends

Chart type	Description and design considerations
Line chart 	Line charts show the change over time for one or more series (sales per hour). The line connects each data point in the series (shown or not). The y-axis baseline should be equal or less than the minimum value in the data. • Show four or fewer series of lines on a line chart (Wong, 2010). • Label each series directly or use an ordered legend.
Sparkline 	A sparkline is a line chart without axes or much detail. It is a small graphic designed to give a quick representation of change over time. • Not intended to provide the quantitative precision of a normal line graph. • Label the last data point to provide additional information.
Area graph 	Area graphs are line charts with the area below the line filled in with color. They can show a single series or multiple time series using stacked areas. Use the same color for the line and the area beneath it. See Table 3.4 for use of stacked areas to show proportional change over time.
Stream graph 	Stream graphs show changes over time for different data series. Color is used to distinguish the categories. Each stream represents a single category proportional change over time. Stream graphs are used to provide a general overview, not when accuracy is important. Use for large time series data sets with five or fewer categories.

With time series data, beware of interpolation, a method of constructing new data points within the range of a discrete set of known data points. If you do not have many known values, consider plotting each data point without a line; a line chart could lead to incorrect estimated values.

3.7 Word Frequency and Sentiment

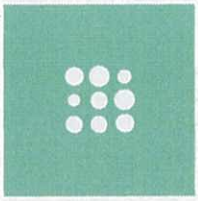
Table 3.8 presents options for visualizing textual data (for example, to show how frequently words appear in a given body of text by making the size of each word proportional to its frequency). Frequency of words may include added dimensions categorized by sentiment, such as positivity or negativity of each word.

Questions:

- How many times does a given word or phrase appear?
- What words or phrases appear most often? Least often?
- What words appear together?
- Are most words or phrases positive or negative?

Insight: frequency or counts of words and phrases. The count of the positive or negative direction of the sentiment of the words or phrases.
Data: text as single words, or n-grams (one or more words that appear together in text).
Chart options: word cloud, proportional area chart using size bubbles or squares, and histogram.

Table 3.8 Chart options for showing sentiment and frequency words

Chart type	Description and design considerations
Word cloud 	Words are arranged in a cluster or cloud of words. Words can be arranged in any format: horizontal lines, columns, or within a shape. Color is used to categorize words by sentiment, or another categorical variable.
Proportional bubble area chart 	Words are ranked by their frequency. The frequency is represented by sized bubbles or squares. The bubbles /squares are arranged in a grid with words on the x-axis and observation on y. Works well for the top 10 words (difficult to view beyond that).

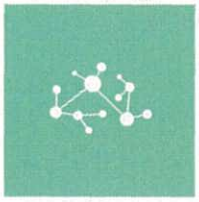
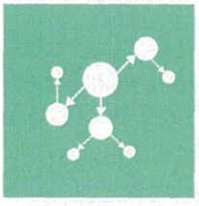
Learn more about software options for creating word clouds at:
http://becomingvisual.com/portfolio/wordclouds

3.8 Connections and Networks

Table 3.9 shows how connections between people or entities can be represented within a network (such as a school, organization, or structure). For example, a diagram may show who is following whom on Twitter. These are known as social graphs. They show the representation of the interconnection of relationships in an online social network.

In addition, a network diagram can help explain which employees are most critical to certain tasks, how students in a school are grouping, and so on.

Table 3.9 Chart options for showing networks

Chart type	Description and design considerations
Network diagram (undirected) 	Undirected network diagrams depict equal relationships between individual entities. Each entity is referred to as a node, represented as a bubble. The relationships between the entities are shown as lines known as edges. The thicker the line, the stronger the relationship. The position of the nodes shows centrality in the network and distance from other nodes. Minimize edge (line) crossing (Ognyanova, K, 2016).
Network diagram (directed) 	Use directed layout to show the orientation of the relationship between nodes. Communicate the strength of a relationship by the width and the direction of using arrow heads.

Questions:

- Who is closest to whom? Who is connected to whom?
- Who is the most popular? Who is the least?
- What communities exist and who are their members?
- What is the strength of the relationship between two entities?

Insight: see relationships, patterns, centrality, or interactions. This is shown by the width, color, or arrowheads on a line to communication relationships. The position of the nodes (that look like bubbles) show centrality in the network and distance from other nodes.
Data: edge lists or adjacency matrices show relationships between entities.
Chart options: undirected network diagram and directed network diagram.

Interview With a Practitioner

I interviewed Jack Hanlon from Jet.com who described how he uses data graphics in his practice.

Kristen Sosulski (KS)

Jack Hanlon (JH)

KS:

Who are you and what do you do?

JH:

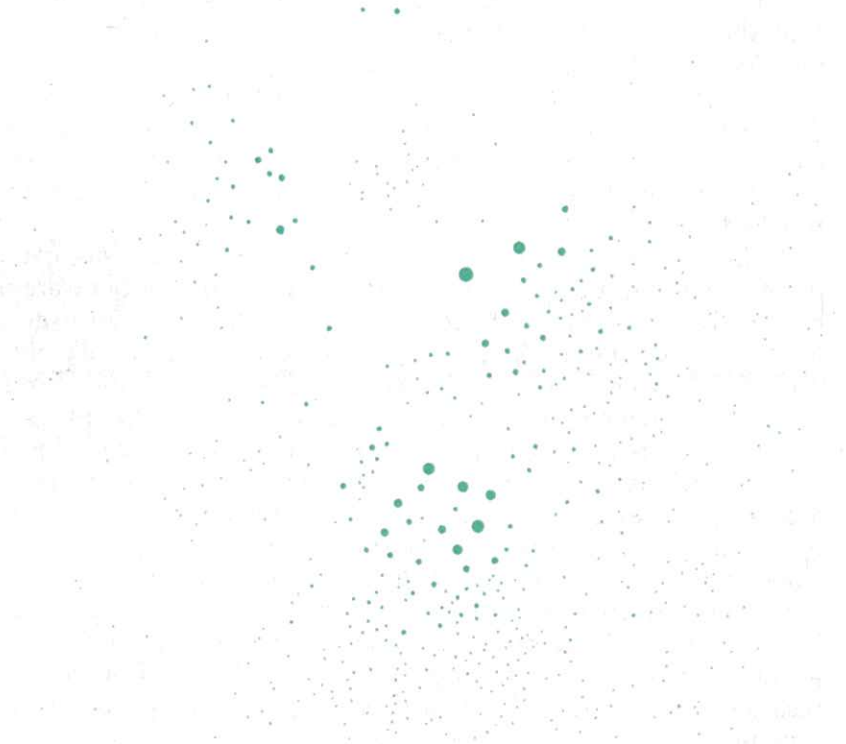
My name is Jack Hanlon. I started the Analytics & Insights (A&I) practice at Jet.com, where I serve on the management team. Our six teams in A&I are Customer Analytics, Marketing Analytics, Core Data Science, Research, Testing & Optimization, and Personalization. Our core belief is that if we can democratize access to insights, advance research and analysis techniques to all of Jet's employees, then Jet can create more amazing experiences for our customers and make stronger business decisions. That means our work involves everything from performing rigorous analysis, to doing in-person ethnographic studies every week, to teaching classes on things like Machine Learning, to creating large-scale user testing systems, to building pipes for data connections between systems. We want to be wherever information and automation create real impact.

KS:

How do you use data visualization in your practice?

JH:

Data visualization is critical for our practice to be successful, as it is one of the powerful mechanisms for storytelling. Most people's eyes glaze over if you discuss the different statistical methods in an analysis (even when this detail is critical), but if you create fantastic visuals with elite production value, you send a message about your credibility to support all the great work behind it. So, while we make sure our analytical methods are airtight, we also spend a great deal of time on data viz and other presentation skills because if you are a great storyteller, then you can make amazing things happen in an organization. In science, sometimes, we'd like to think that facts alone are sufficient, but time has shown how frequently that's not the case!



KS:

Can you share an example of how you have used data visualization in your practice? (show data graphic itself). What insight is evidenced by the graph provided? What did you do with that insight? (make a decision, inform policy, predict the future, etc.)

JH:

You can imagine how well the owner of a General Store knew their customers–getting to see each transaction and talk to every customer. One of the amazing things about eCom is the scale and access but it adds a few layers of abstraction between the consumer and the seller. As a result, we need to be able to understand what purchase patterns look like for certain customers to be able to make their lives easier wherever we can. In this case from 2015, right after the Jet launch, we wanted to see what categories of products were purchased together in the same cart, focused on consumable purchases (which are grocery-type items, e.g. items you would consume and have to replenish at some point like detergent or paper towels). Did they only buy other consumables? Did they also buy durables in that session (products that you don't need to replenish, like a bike or a shoe)?

The circles (nodes) represent item categories, and the colors represent whether the category contains consumables (in dark grey) or durables (in cyan). The categories purchased together are linked by lines (edges). The thickness of the lines indicates the frequency of the connections between the two categories, and the size of each circle represents its number of connections. Items that are frequently purchased together are attracted, while items that are infrequently purchased together are repelled.

In this image, we saw a number of categories that were categorized as durables based on our organizational structure but were purchased like consumables based on customer thinking. For example, light bulbs or printer ink/toner were both part of Electronics but are purchased like consumables. Similarly socks and underwear were part of Fashion, but their purchase pattern does not operate like blouses or shoes, and instead would be purchased with other consumables. This sounds obvious now, but like so many things–until it is illuminated in a digestible way–it can be easily missed. This insight enabled us to understand this concept of "shopping missions" that people are on when they come to us, which had tangible tactical implications including rethinking what products we should carry in certain warehouses (carrying these above products with consumables would allow them to ship together) and rethinking how we should think about merchandising on the site to make it easier to find the right products at the right time.

KS:

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

JH:

This is called a force-directed network diagram and it was created by Jamie Fitzgerald, a very talented data scientist in our group. She is a career scientist, so she spent a great deal of time working through understanding the data. She started simple with tables with number values indicating strength between categories, but due to the volume of data and the multiple connections for each item, it became clear that we needed a stronger relationship-focused and visual method to get the story across to the widest audience. From there it became clear that a graph-oriented method was the right approach. She explored a variety of graph-related options, both in terms of network diagram types, and of tools used to generate the graphs, and found that the easiest and most flexible option was Gephi, an open source software solution for visualizing and analyzing graphs and networks. When Jamie finished the first draft it was clear we were going to have something special here, and the larger version of this diagram that includes all categories and all colors is framed and hung outside of our central boardroom at the Jet.com HQ in Hoboken.

3.9 Chart Interface

Up until this point, we have assumed that we were creating static, motionless, unresponsive data graphics. Visualizations that serve as an interface to navigate or traverse data elements feature animated and interactive components.

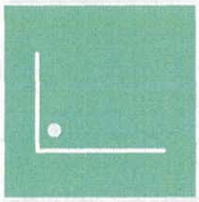
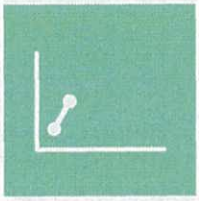
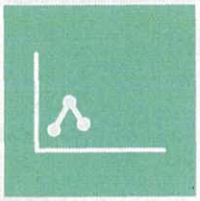
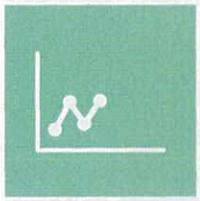
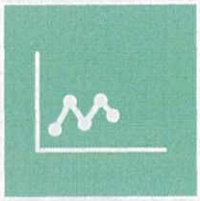
3.9.1 ANIMATION

Animation charts show the movement of data encodings, such as bubbles, lines, points, lines, fills, nodes, or areas, in a sequence. There are three types of animations used in data visualizations: trend, transition and trace. These are used to show a succession of changes.

Trend Animation

A trend animation shows all trends simultaneously building up from beginning to end. For example, Table 3.10 shows a time series revealed gradually, point by point, using trend animation.

Table 3.10 A trend animation shown frame by frame

January	February	March
		
April	**May**	**June**
		

Transition Animation

A transition animation shows how a data point within an entire series changes. Usually, a point or series is highlighted throughout the animation to provide context for how the points around it change. Table 3.11 shows a transition animation of city budget by city and population and budget subcategory. The city in question is highlighted in gray. The position of the data point changes by movement along the x-axis.

Trace Animation

Trace animation uses fade-in and fade-out points or bubbles to show the direction of the flow of data points while keeping the historical data points present but faded. Too many data points can create a cluttered interface. Trace animation can be used to present variables related to geography over time. Table 3.12 shows a map of store openings over a six-year period. For each year, the store locations are represented in green; the points fade to a gray color as the animation moves to the next year in the sequence.

3.9.2 INTERACTION

Interactive data visualizations allow the audience to explore a data set through a visual interface. Users can manipulate and transform

Table 3.11 A transition animation shown frame by frame

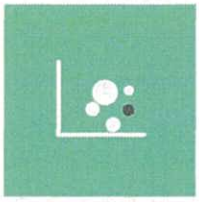
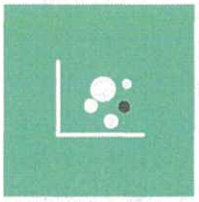
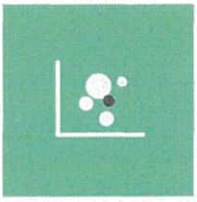
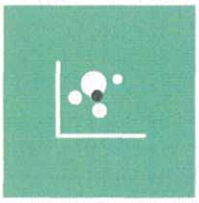
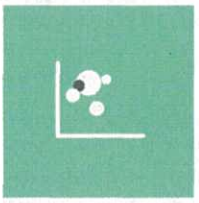
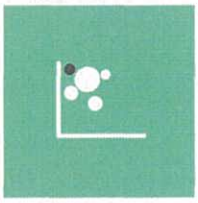
Public Safety	Admin.	Community
		
Government	**Public Works**	**Development**
		

Table 3.12 A trace animation shown frame by frame

2013	2014	2015
		
2016	**2017**	**2018**
		

the output of the display through a variety of mouse actions. User actions include: select, explore, reconfigure, abstract/elaborate, filter, and connect (Ji Soo Yi, Youn ah Kang, Stasko, & Jacko, 2007). These actions differ from animation because: 1) the data is manipulated with

the output being affected; or 2) the design of the visualization is transformed, which affects the presentation but not the data.

Select Interaction

The user marks one or more data points of interest (see Table 3.13). Click on a data point to mark or highlight it. This is used for tracking data points. For example, let's say we wanted to see the change in sales by store location over time. This can be illustrated in a bubble map (with bubbles sized by the sales). To see a particular store and sales, the data point is selected; through animation, the viewer sees how sales have changed for that store over time.

Explore Interaction

The user examines a subset of data through panning, zooming, or rotating the data graphic. Table 3.14 shows panning and zooming interactions on a point map of New York City.

Abstract and Elaborate Interaction

The user clicks on or hovers over a data point that reveals information. Table 3.15 presents a scatterplot with a point selected and a label with a description of the data point.

Reconfigure and Encode

The user can build interactive data graphics. For example, the user selects the variables to plot, determines the spatial arrangement of variables on the axes, and the visual appearance (size, shape, color, etc.). Figure 3.2 is a scatterplot with options to change the variables

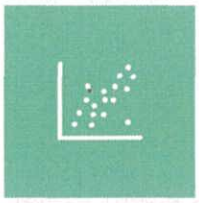
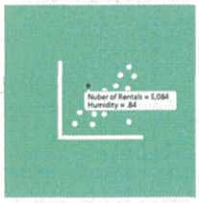
Table 3.13 Marking a point of interest to highlight and observe

Click on data point	Data point changes color	Point size changes based on time or another attribute
		

Table 3.14 Exploring a point map using panning and zooming

Original view	Panning	Zooming	Selecting and zooming
			

Table 3.15 Clicking on or hovering over a data point reveals information

Original view	Hover or click
	

shown on the axes, the encoding type (circles, squares, or diamonds) and the color of the encoding (black, gray or blue).

Filter Interaction

The user changes the set of data items being presented based on some specific conditions. Filter types include a single value list, single value dropdown, single value slider, multiple values slider, multiple values list, multiple values dropdown, or search. See Table 3.16 for the common filter types.

Connect Interaction

Users interact with one data display to control the output of another data display. This is commonly used in dashboards. The connect action is used to highlight associations and relationships between data items that are already represented, and it shows hidden data items that are relevant to a specified item.

For example, Table 3.17 shows the original view and the data changes as the user clicks on the bubble.

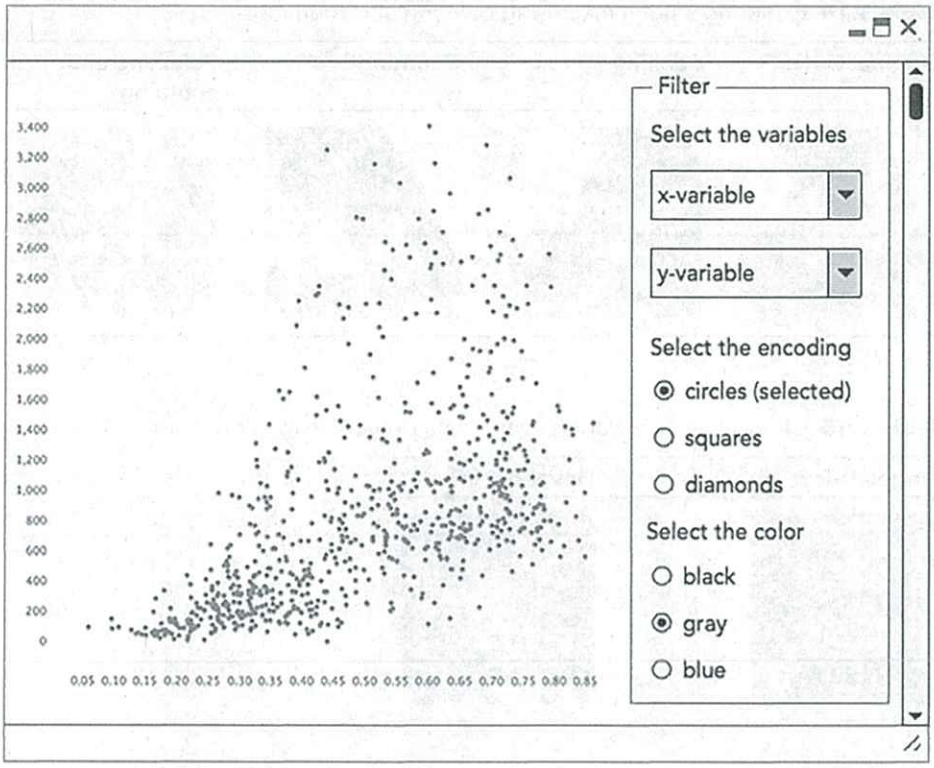
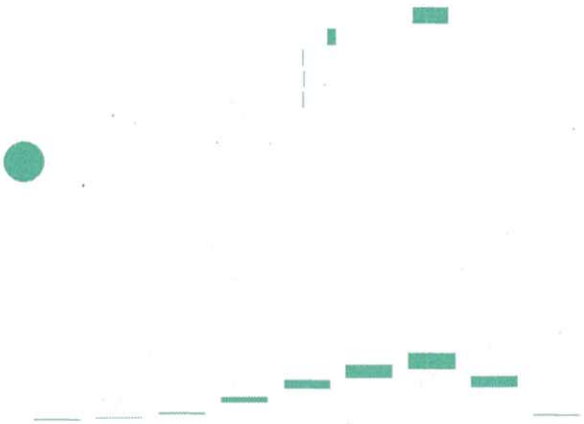


Figure 3.2 An interface for selecting the variables and encoding attributes

Table 3.16 Six types of filters

Single value list	Radio buttons	Drop down list
List item 1 List item 2 List item 3 List item 4 List item 5	option 1 option 2 option 3 option 4	List item 2 List item 4
Checkbox	**Slider**	**Search**
Checkbox 1 Checkbox 2 Checkbox 3 Checkbox 4 Checkbox 5		

Table 3.17 An illustration of the connect interaction using one display (the bubbles) to control the output of the other displays (horizontal and vertical bar charts)

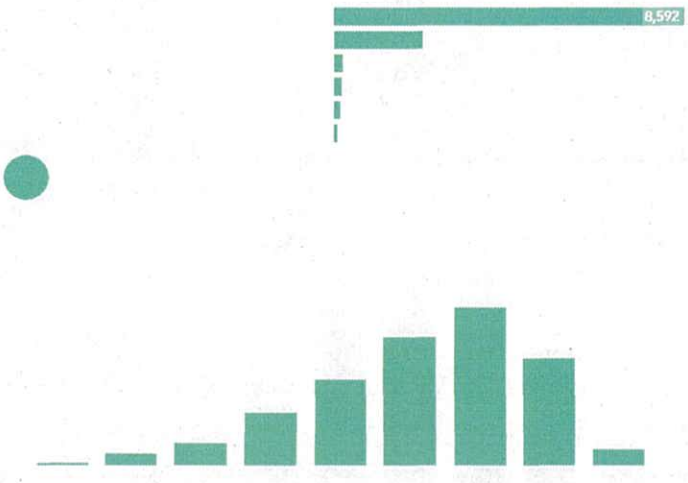
Original view

Dark gray bubble selected and series changes
Light gray bubble selected and series changes
Green bubble selected and series changes

(Continued)

Table 3.17 (Continued)

Original view

Light gray bubble selected and series changes



Throughout this chapter, we looked at how the data type drives the types of visualizations available to reveal insights, including comparisons, distributions, compositions, relationships, locations, trends, connections, and sentiment.

Almost every chart can be transformed into one of three formats:

1. *Static*. Unanimated or non-interactive charts. The data may change or update with refreshed data, but the display does not change based on user actions.
2. *Animated*. Animated displays work well for projected displays or screen displays where a presenter is describing the insight shown on the screen. Use animation to guide the pathway through a chart. Animated displays can take several forms: transition animation, trend animation, or trace animation.
3. *Interactive*. Displays that are designed to allow users to explore a data set on their own—often providing details on a mouse over, giving different coordinated views, or panning and zooming.

3.10 Exercises

1. Download the data on motor vehicle accidents from New York Police Department from https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95. This data set contains geographic, categorical, and time series data.
   a. Create one display for each type of data to present an interesting insight.
   b. Create an animated time series display that shows the accidents in Manhattan over time.
   c. Create an interactive display that uses the borough as a filter for the location of accidents shown on the map.

Bibliography

The City of New York. (2014). Public recycling bins. Retrieved from https://data.cityofnewyork.us/Environment/Public-Recycling-Bins/sxx4-xhzg

Corum, J. (2011). The class of 2011. Retrieved from www.nytimes.com/interactive/2011/06/10/education/commencement-speeches-graphic.html

Ognyanova, K. (2016). Basic and advanced network visualization with R. Retrieved from www.kateto.net/wp-content/uploads/2016/04/Sunbelt%202016%20R%20Network%20Visualization%20Handout.pdf

Ribecca, S. (2013). The data visualisation catalogue. Retrieved from http://datavizcatalogue.com/index.html

Wong, D. M. (2010). *The Wall Street Journal guide to information graphics: The dos and don'ts of presenting data, facts, and figures*. New York, NY: W. W. Norton & Company.

Yi, J. S., Kang, Y., Stasko, J. T., & Jacko, J. A. (2007). Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, *13*(6), 1224–1231. doi:10.1109/TVCG.2007.70515