DESIGNING PUBLIC VISUALIZATIONS OF LIBRARY DATA[1]

ANGELA M. ZOSS
DUKE UNIVERSITY LIBRARIES

As in many other organizations and fields of inquiry, the data generated by libraries becomes ever more complex, and the need to communicate trends both internally and externally has also been increasing. As visualizations become increasingly embedded in library assessment and outreach, it is crucial to take into consideration the audience of the visualizations and to design visualizations that are easy to interpret. This chapter will walk readers through the process of selecting a visualization based on a particular data representation need, designing that visualization to be optimized to its specific purpose, and combining visualizations into larger narratives to engage a public audience.

The process of choosing, designing, and combining visualizations in a way that is engaging to a broader public requires an understanding of how visualizations convert data into shapes or spatial arrangements that can be analyzed and understood. Various types of visualizations have been developed over the centuries to take advantage of human skills and acuities. To understand why some work better than others in certain situations, it helps to first examine the components of these graphs and how well they match up with what humans are good at.

Several studies have been done to explore how well individuals read different components of visualizations. Cleveland and McGill (1985) conducted a seminal study where users were asked to evaluate two data elements in terms of the proportion one represented of the other. This task was performed using multiple visual representations: position, length, angle, slope, and area. The results suggested that individuals are much better at assessing numerical data using position and length encodings than they are using slope or area of a shape. This study was replicated and extended by Heer and Bostock (2010), who confirmed that position and length afford more accuracy than slope or area, including both circular and rectangular areas.

A quote from Moritz Stefaner, a professional visualization designer and researcher, summarizes this and related academic and professional work around designing data visualizations: "position is everything, color is difficult" (Stefaner 2012). The following section will explore how these studies and inherent properties of different types of visualizations can be used to select an appropriate visualization for a data need.

VISUALIZATION SELECTION

This section will review common types of charts and maps, discussing their strengths and weaknesses in relation to different types of data and audiences. For example, some visualizations are matched well with the visual acuities of humans, some are good at filling space on a page, and some can be especially engaging for a public audience. Additional visualization-specific design suggestions will accompany the discussions of each visualization type.

---

[1] This is an author-produced version of a book chapter. The final version is available at:
Zoss, Angela M. (2016). "Designing Public Visualizations of Library Data." In *Data Visualization: A Guide to Visual Storytelling for Librarians*, edited by Lauren Magnuson. Lanham, MD: Rowman & Littlefield Publishers, Inc.

BASIC CHARTS AND GRAPHS[2]

Basic charts and graphs (as opposed to custom/novel visualizations or highly scientific plots) are the easiest visualizations to produce, as well as the easiest visualizations for a broad audience to understand. Their ubiquity does not, however, mean that all chart types can be equally well understood by all people. Firstly, different charts use different visual encodings, and those encodings match up differently with human perceptual abilities. Secondly, people do not have equal exposure to all chart types. Depending on someone's educational background, s/he may have more or less practice interpreting a particular chart type. Finally, the data being visualized may have properties that do not match well with a particular chart type, making the resulting chart difficult or impossible to read effectively.

Choosing an appropriate chart type requires familiarity with the chart options available, as well as with the properties of the data and the purpose of the chart. What trends in the data are most important? What chart will best show those trends? Who is going to be reading the chart? The following section will review six basic chart types: bar charts, pie charts, line charts, scatterplots, bubble charts, and heat maps. Each chart will be presented with a discussion of the best uses of the chart, the most common concerns with the chart, and any additional suggestions about the use of the chart.

BAR CHART

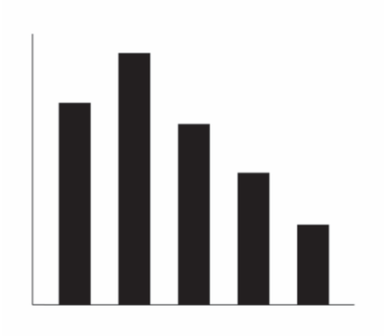

Figure 1. A basic bar chart.

Bar charts are a staple of data visualization – especially those that are intended for a broad audience. They are some of the most general types of charts, with few limits on the kind of data that can be used. There are just a few basic design tips to watch out for to make sure you're taking full advantage of how powerful it can be.

DATA: The most basic form of a bar chart[3] involves one categorical variable and one numerical variable. The categories from the categorical variable are each given a bar, and the quantities from the numerical variable are used to determine the length of the bar. An optional, additional categorical variable can be added to make either a grouped bar chart (where several bars appear next to each other for each category in the original categorical variable) or a stacked bar chart (where the original bar is split up into several colored segments, stacked on top of each other).

[2] Some authors make distinctions between charts and graphs. For example, Börner (2015) uses "chart" for visualizations that lack a formal reference system (e.g., word clouds). Börner and Few (2012) both reserve "graph" for visualizations that do use a reference system – typically a Cartesian or Polar reference system. In many software applications, however, these terms are used interchangeably, and this chapter will follow that convention.

[3] Sometimes bar charts are incorrectly referred to as "histograms." A true histogram visualizes only a single, numerical variable. The "bars" in a histogram are actually numerical ranges (e.g., 0 to 4.99, 5 to 9.99), dividing the values of the variable up into a series of chunks or "bins." The number of data points that fall within the range becomes the length of the bar for that bin. You can usually tell the difference between a histogram and a bar chart by looking for gaps between the bars. In a histogram, the lower value of one bins starts right where the upper value of the previous bin leaves off, so the bars in a histogram usually do not have any gap between them.

STRENGTHS: Bar charts are especially good for times when it's important to be able to read numbers accurately off the chart. The reason for this lies in the human visual system. Humans have excellent perceptual acuity for differences in alignment and line length (Cleveland and McGill 1985, Ware 2013). Bar charts are also extremely common. They are some of the first charts we learn to read and produce. Comparing the individual bars to each other or exploring the overall trend across categories are both good uses for a bar chart.

WEAKNESSES: Bar charts have a few common concerns, however. Firstly, labeling the bars with long phrases can cause design problems. If the bars are vertical, it can be difficult to make the bar labels horizontal; the words will be much wider than the bars, leaving large gaps or awkward line breaks in the labels. If, on the other hand, the bars are horizontal, it is very easy to have long labels for each bar. The text height and the bar height can be very similar, and the wide format for the chart fits much better with the aspect ratio of screens nowadays.

Sometimes one or two bars will be much longer than the others; consider creating a second chart where you exclude the long bar and zoom in on the shorter bars (Few 2012), or investigate whether several categories can be combined into an "other" category. Also consider what comparisons or evaluations are most important for your bar chart; if alphabetical order is not relevant for your data, for example, it might make sense to sort the bars instead by length (Cleveland 1994).

Most important, however, is to pay attention to the numerical axis of the bar chart. Because humans are so good at perceiving differences in length between bars, having the axis start at anything other than zero will distort how the length corresponds to the data. This renders the bars, which are a powerful visual cue, inaccurate and hard to ignore. The numerical axes for bar charts should always start at zero (Few 2012, Yau 2013).

PIE CHART

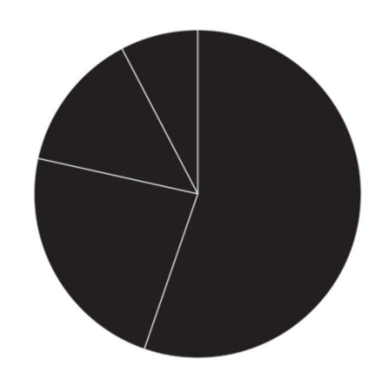

Figure 2. A basic pie chart.

Pie charts are also very common in infographics, reports, etc. They are the subject of much discussion and are often criticized because of the many problems associated with them, but they are still a valuable chart in certain contexts.

DATA: Like a bar chart, a pie chart is built from one categorical variable and one numerical variable. In a bar chart, each category becomes a separate bar. In a pie chart, each category becomes a separate wedge of the pie. The proportion of the total data associated with each category is represented by the area of the wedge. Typically, programs that create pie charts will start the first wedge at the "12:00" position of the circle and proceed clockwise. Programs also typically assign a different color to each wedge; this color is then used to match a wedge up with the legend entry for the wedge. If wedges are instead labeled directly, they can each have the same color, provided a contrasting border color is used to show the wedge boundaries (Wong 2010).

STRENGTHS: Pie charts are a favorite when people want to show a part-to-whole relationship. That is, when you have data on the total amount of some variable, like budget, it is natural to want to represent the subdivisions of that total in a way that shows that all of the parts add up to the total. Pie charts are also quite common and easy to produce, so audience familiarity should be relatively high.

It has been said that the best data set for a pie chart is a single number. Focusing on one category and how much of the total data it represents can be very compelling. This technique is used more frequently in an infographic context – one where data are reduced to bite-sized pieces – rather than a reporting context, where completeness may be important.

WEAKNESSES: Humans are notoriously bad at reading data values from pie charts. Our perceptual system is not very precise at evaluating angles, orientations, or two-dimensional areas (Cleveland and McGill 1985, Heer and Bostock 2010). With each wedge rotated in space, it is incredibly difficult either to read precise values or to compare wedges, especially those that do not appear next to each other.

This problem is compounded if the differences between data values are very small. For example, if you have five categories and each is only slightly above or slightly below 20% of the total, the fine distinctions will be lost to chart readers. If the goal of your chart is to help readers understand small differences, the pie chart may not be the best choice; consider switching to a bar chart. If, on the other hand, there are noticeable differences in the sizes of most of your wedges, you can make sure to help your readers as much as possible by putting the wedges in decreasing order by size (i.e., left-hand edge of largest wedge at "12:00", followed on the right by next largest wedge, etc.)[4]. Sorting by size can aid readers in making relative estimates of data values.

Also beware of data that do not have a true part-to-whole relationship. One example might be data that change over time – for example, where each "category" is a single year. In this case, there is a mismatch between the visual encoding or the visual metaphor (one that joins parts of the chart into a solid shape) and the nature of the data (one where a single variable is changing over time). Combining multiple years does not often result in a meaningful total, so encoding that total as a visual element (a completed circle) is seldom necessary or natural. Time-based variables are often a good match for line charts, which will be discussed next.

Pie charts have another notorious concern – the use of three-dimensional special effects. Some programs allow the chart creator to add a special effect that renders charts as though they are 3D objects. This may simply involve drawing a 2D pie chart to look like it is popping forward, but it may also involve actually tilting the chart "away" from the reader. This tilting effect causes an extreme distortion of the visual area of the chart (Skau 2011). The wedge at the bottom of the chart will end up taking up much more of the chart space than it would in a 2D version, and the wedges at the top will take up much less chart space than they should. Thus, in a tilted pie chart, neither the angle nor the area of the wedges accurately correspond to the data values in question. An already difficult chart to read has become highly distorted. Consider settling for the more boring, but much more faithful, 2D version (Few 2012).

The final problem with pie charts occurs if there are a large number of categories. (See Figure 3.) This can result in very small slices, including ones that may even be invisible depending on the size of the chart. This can also lead to difficulties with color, if the program you are using assigns a new color to each wedge. If the program has to repeat the same color multiple times, it become hard to match a wedge up to the correct entry in the legend. If wedges are labeled directly to solve the color problem, the many labels can add a lot of

---

[4]Wong (2010), however, recommends placing the largest wedge to the right of the 12:00 position and then starting the second largest wedge to the left of 12:00, continuing counter-clockwise after that. This places the two most important wedges at the top of the chart, instead of have the most and the least important wedges at the top. If wedge sizes are similar, however, it may not be obvious to readers that this nontraditional ordering is being used.

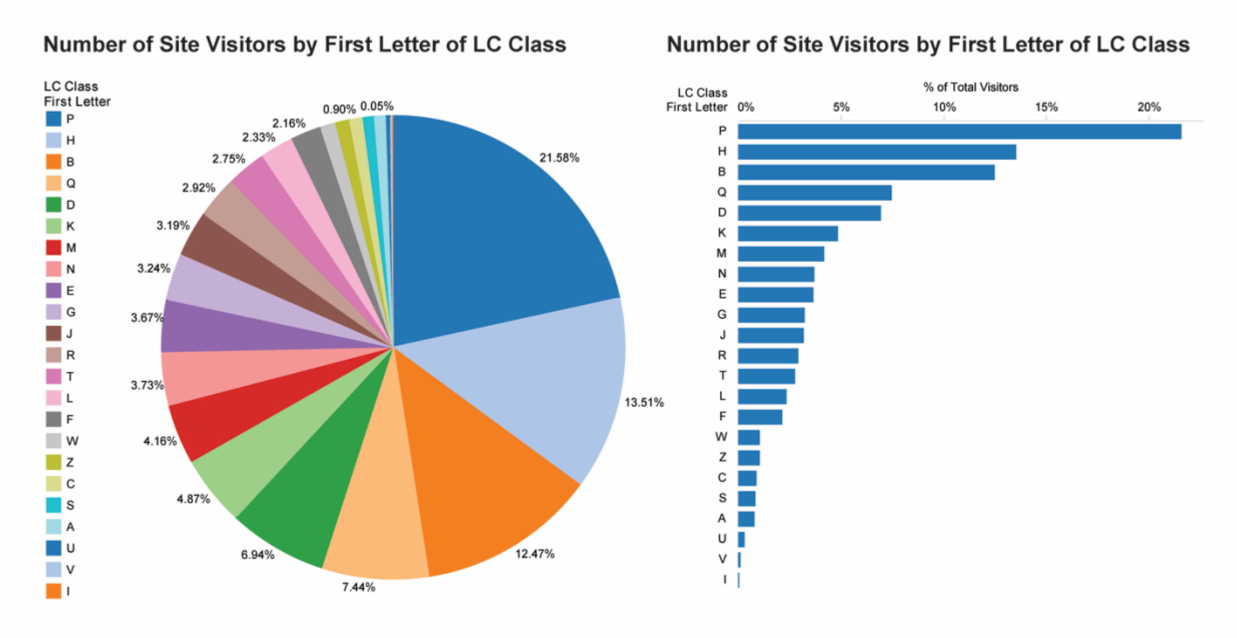

Figure 3. These two visualizations of the same data show how bar charts can be used to solve problems with pie charts. The data used are from a sample data set of visitors to catalog pages for an academic library. The categorical variable, the first letter of the Library of Congress classification for the catalog item in question, has too many categories to be well represented in a pie chart. A bar chart shows the data much more effectively.

visual complexity to the chart. If the data have a natural grouping, you could combine wedges into larger categories or simply create an "other" category for wedges under a certain size.

LINE CHART

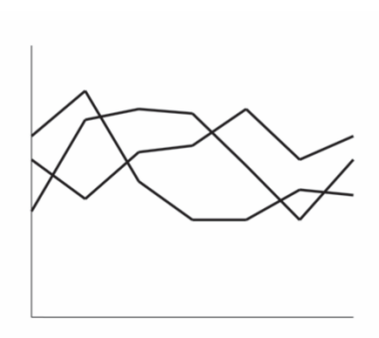

Figure 4. A basic line chart.

Line charts are more limited than bar charts and don't necessarily match well with proportional data, but they are very well suited to visualizing continuity and temporal data. The main concerns involve data complexity or consistency.

DATA: Line charts typically show change over time. As such, they often use one date variable (for the horizontal axis) and one numerical variable (for the vertical axis). If your software program does not require line charts to use dates, the program might treat that axis as a series of ordered categories (like days of the week) or just as another numerical variable (like treating years as integers). An extra categorical variable can be used to split the line into multiple lines, one for each category.

STRENGTHS: The continuity of the line in a line chart is a great match for data that change continuously over time. Pie charts, as previously discussed, should only be used for separate categories that add up to a meaningful whole. A bar chart is a good general chart and can be used either for parts of a whole or for quantities that change over time, but the separations between bars do not reinforce the idea that data are changing in a more fluid manner.

The line chart also offers practical features beyond a bar chart. Firstly, the line chart can be more elegant when you add a category variable than a bar chart. Imagine looking at a data set over time with a grouped bar chart. For each time point, you will have several bars, but they have to be placed next to each other. This will add width to the chart and make it difficult for the eye to connect the bars that represent a single category. The other option, a stacked bar chart, makes it very difficult to read the exact value of each category because all but the bottom-most category start at a value other than zero.

A line chart can fix both of these problems. Each data point is positioned from the bottom of the y axis[5]; data points within a single category are clearly connected by lines; and the data points from all categories line up vertically at each time point, keeping the width of the chart manageable.

A final benefit of using a line chart is that the date measurements do not have to be evenly spaced. Bars default to even spacing, which would distort how quickly data are changing if, say, some dates are a single year apart while others are two or three years apart. With a line chart, the horizontal position of the data points is (or should be) precise, corresponding exactly to the date and accurately representing the distance between measurements. (See Figure 5.)

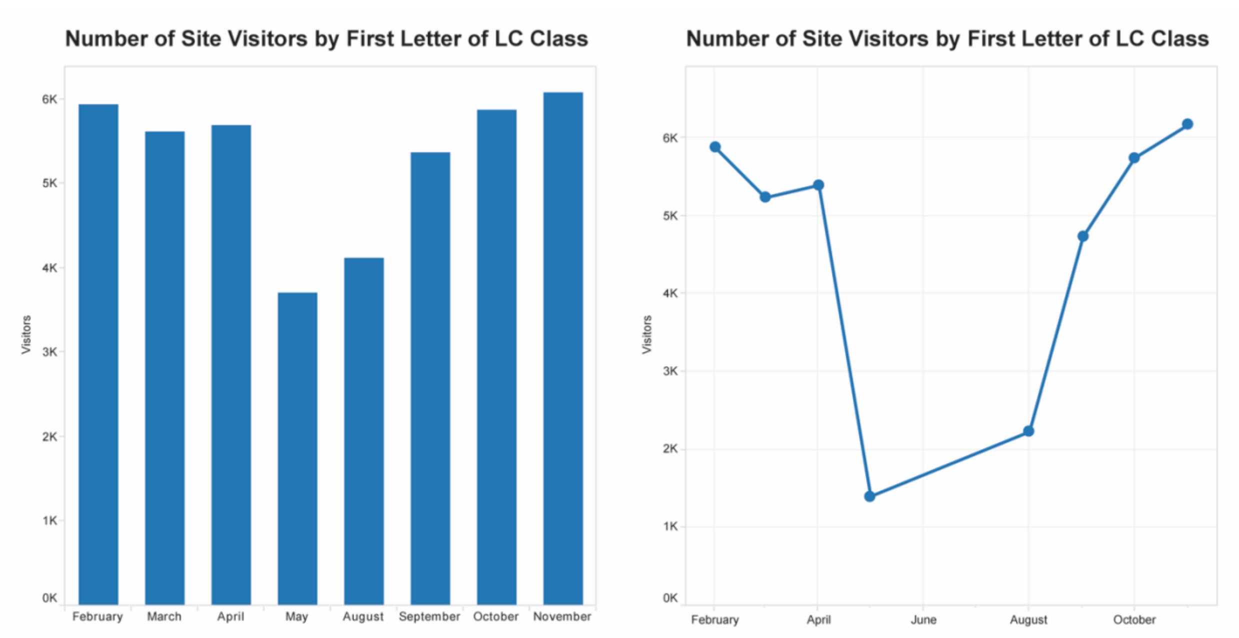

Figure 5. These visualizations illustrate the problem with using bar charts for temporal data that have gaps between temporal categories. Charting tools may draw all of the bars with equal spacing, ignoring the months that have no data. The line chart, instead, correctly shows how the months should be spaced.

[5] There is some debate about whether the y axis of a line chart should start at zero. Because there is no bar representing the number as a continuous visual area, it may be appropriate to limit the y axis to the range of interest, effectively "zooming in" on the line chart (Cleveland 1994). This can be especially useful for data series where small variations are important or significant.

WEAKNESSES: The strengths of the line chart become its weaknesses when it comes to data overlap. Unlike bar and pie charts, line charts have nothing that prevents the lines from covering each other up. Typically line charts use color to differentiate the lines, which can help when lines cross a lot, but this also makes it more difficult for users to read the chart. Users end up needing to look back and forth between the chart and the legend to identify the lines. Where possible, placing labels on the chart, next to the lines, can improve readability and potentially even eliminate the need to use color for identifying categories.

Depending on the software being used, you may still need to worry about the problem of uneven spacing between date measurements. If your dates look more like categories – for example, if you are using the names of months, you may have to create the complete list of all possible months and then leave "null" values or zeroes for the months that have no data. (For a good discussion about how to deal with missing data in a line chart, see Kandel et al. (2011))

Finally, if you use a simple style (i.e., a plain line) for your line chart, you may find that it is hard to tell where the data points are. The measured data will only be visible if the line bends one way or the other at that particular point. If the slope of the line is relatively consistent, it is hard to tell how evenly spaced the data are. Consider adding small dots on top of the line to clarify where the measurements are, especially if they are unevenly distributed (Few 2012). It is typically not necessary to use multiple shapes for these dots; a single shape (e.g., a circle) for all lines will help maintain a consistent style and ensure that each data point is equally visible.

SCATTERPLOT

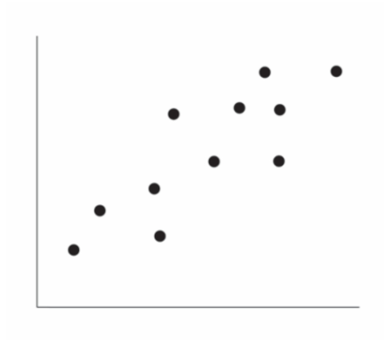

Figure 6. A basic scatterplot.

Scatterplots are some of the most precise and data dense charts available for visualization. Special patterns in scatterplots can reveal strong trends in a dataset, or even suggest errors or outliers in the data (Yau 2013). Scatterplots are a great match for our perceptual system, but they can fall short in terms of audience familiarity.

DATA: Scatterplots are typically used to show the relationship between two numerical variables. Each variable is assigned to an axis, and a point is placed in space so that it lines up with the correct values on each axis. An additional categorical or numerical variable can be added to change the color or shade of the dots in the scatterplot. A slight variation on the scatterplot, often called a bubble chart, will change the size (specifically, the area) of the dots according to yet another numerical variable.

STRENGTHS: Scatterplots offer a visual representation of the correlation of two numerical variables. If there is a strong (linear[6]) relationship between these variables, the dots will form something like a line, angled either from top left to bottom right (negative correlation) or from bottom left to top right (positive correlation). Scatterplots are also useful for showing clusters of data points (i.e., data points that are much closer to each other than they are to other data points)

[6] Two numerical variables could be related to each other nonlinearly, as well. Scatterplots can also show this, but broad audiences are less likely to understand nonlinear relationships.

or outliers (i.e., individual data points that are very different from the rest of the data points) (Yau 2013).

A scatterplot can also be used as a general form to make other types of charts. Take, for example, a dataset where you want to see the distribution – or range of values – for a particular variable. Maybe you have the number of pages each user read of a particular eBook. You don't necessarily need to know the name of each user; you just want to see how the different users varied from each other, what the lowest and highest values were, etc. Instead of making a bar chart, which could take up a lot of space, you could use a scatterplot where one axis is the number of pages read and the other axis is a single value – basically, a fake integer representing the name of the book. This is called a one-dimensional scatterplot (Cleveland 1994) or strip plot (Few 2012), and all of the dots representing the users will be places in a single vertical or horizontal line. This will help you see clearly whether there is a pattern across these eBook readers or if they vary wildly, and the chart itself will be very compact. The distribution of a variable can also be shown with a histogram (see previous note), which may be a better option if you have a lot of data points or if many of the data points lie on top of each other.

Note, however, that this simplified scatterplot technique would also work for multiple books. You could assign a different fake number to the each books, and the data points for each book would then show up in parallel lines on the chart. More generally, this is how you can use a categorical variable, like book title or genre, with a scatterplot. This can even be used as a replacement for a bar chart. Each category is assigned to a different integer, and each category has a single numerical value, or a single dot. Because the dot is not visually powerful in the same way as a bar, this type of chart – often called a dot plot – does not require a full numerical axis that goes all the way to zero (Cleveland 1994, Few 2012). See Figure 7 for examples of using scatterplots with different combinations of variables.

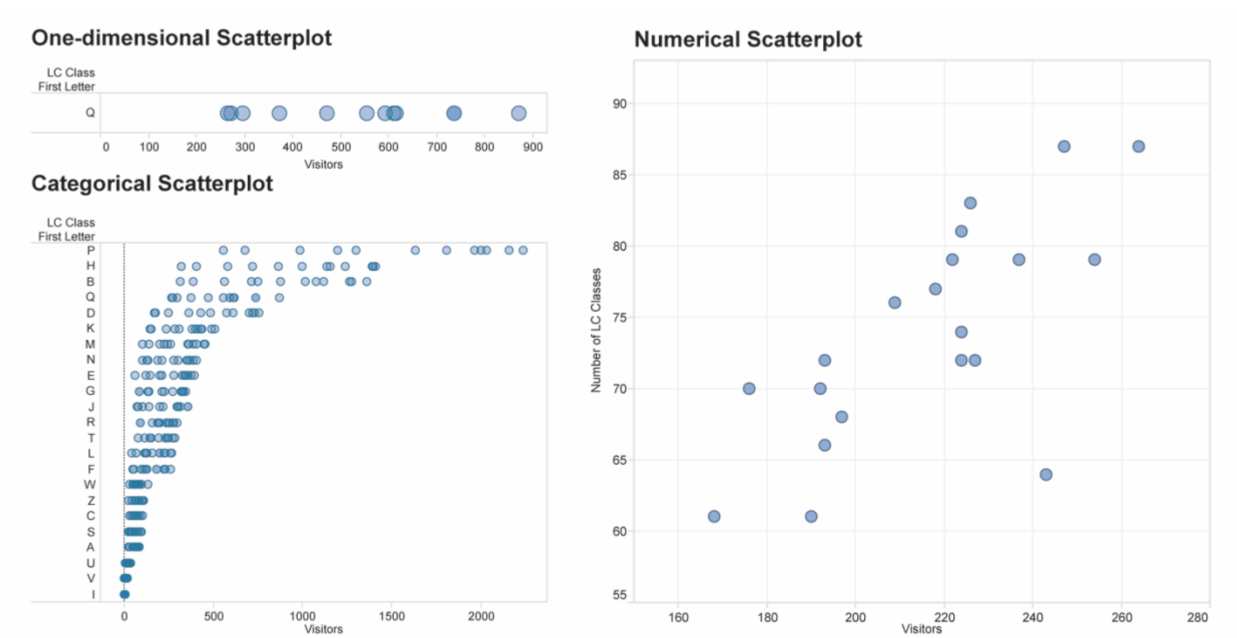

Figure 7. Three examples of scatterplots. The one-dimensional scatterplot shows the distribution of monthly catalog visits over a single Library of Congress class. The categorical scatterplot shows the distribution across multiple categories. The numerical scatterplot shows two numerical variables, number of visitors and number of LC classes, that have a strong positive correlation.

WEAKNESSES: One of the primary weaknesses of scatterplots is that they are less common than bar charts, pie charts, and line charts. It does require some training to be able to detect and make sense of trends in dot patterns, and the sheer number of data points may also be overwhelming.

Likewise, the more data you have, the greater the chance that dots will cover each other up. Unlike the line chart, where lines coming into and out of data points can help users infer an obscured data point, scatterplot points can simple disappear when drawn on top of each other. Some solutions involve using transparency to show when multiple dots are stacked, using logarithmic axes to space dots out more toward the lower ends of the data values, or even using aggregation to total up the number of points that are in exactly the same spot (Cleveland 1994). This gives you an additional variable that you can encode as either color or size (to make a bubble chart).

Bubble charts, though, have an additional weakness. Like a pie chart, a bubble chart requires users to compare the areas of different shapes to understand trends in the data. Comparing the areas of circles has been found to be extremely difficult (Cleveland 1994), and bubble size is thus not a good match for data that needs to be interpreted precisely. If you have three numerical variables and you would like to show them all on the same chart, it is best to use the axes for the two that are most important or that require the most precision.

HEATMAP

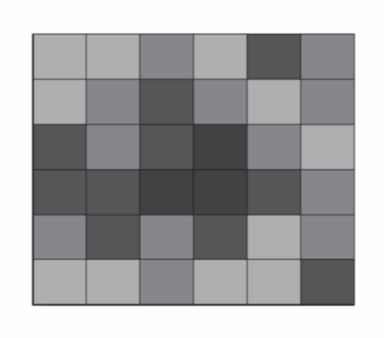

Figure 8. A basic heatmap.

Heatmaps are much less common than the other types of charts, but they can be extremely useful for representing a large amount of data. They are more appropriate for general trends than for precise data lookup.

Data: Heatmaps show data in rows and columns, much like a standard table. The difference between a table and a heatmap is that every number in a table cell is converted to a color. The grid of a heatmap can be created with categorical or numerical variables. For categorical variables, each category becomes a separate row (or column). For numerical variables that are easily separated, like integers, the values can just be treated like categories. For continuous numerical variables, though, it is typical to create number ranges (see previous note on histograms). Within the grid, a third numerical variable (often the number of data points associated with that particular row and column) is encoded as the color of the cell.

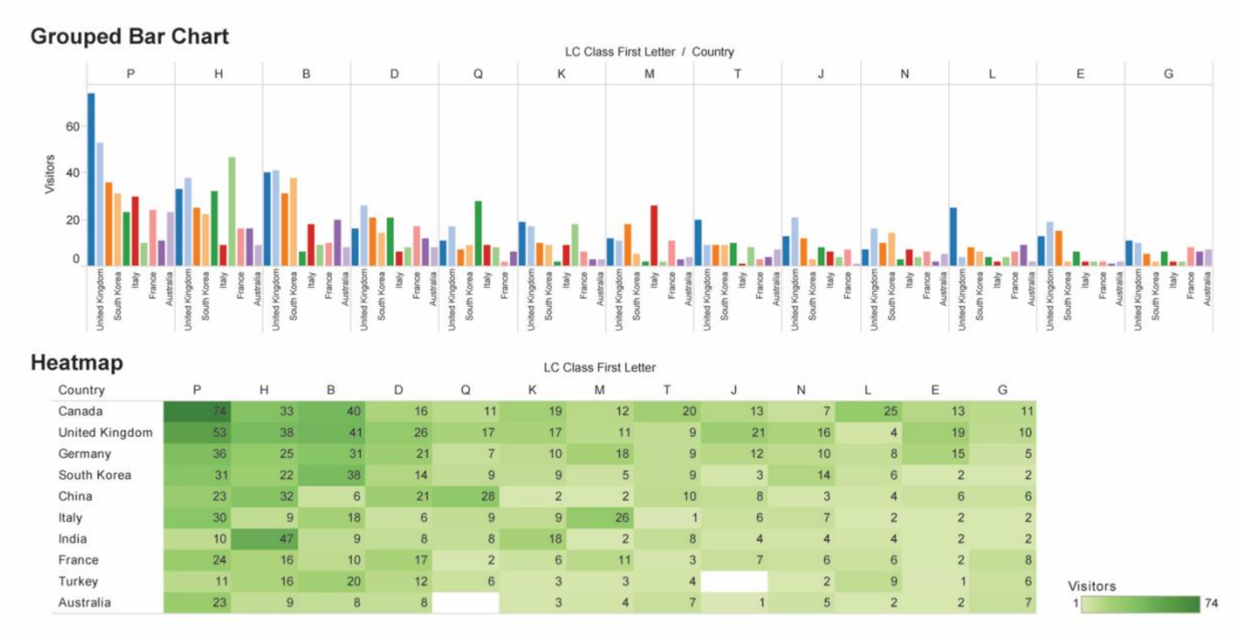

Figure 9. Using a grouped bar chart to show the interaction between two categorical variables yields a chart with many redundant labels and a lot of wasted space. While the bars allow for more precise data lookup, the width of the chart and the small bars make it difficult to compare all of the bars of a single color. The heatmap here offers a much more efficient way of comparing two categorical variables. Sorting the countries and LC classes by the grand totals for each category makes it easier to see highly active combinations.

Strengths: While not a common data visualization, heatmaps fill a gap in the ability of visualizations to highlight the relationships between two categorical variables. Grouped or stacked bar charts can also visualize two categorical variables along with a numerical variable, but both make it difficult to see interesting interactions between the categories, and both prioritize one variable over the other (see Figure 9). Laying two categorical variables out on a grid gives both of those variables equal focus. Finally, heatmaps are also very space efficient; they can compress an amazing amount of data into a small area (Munzner 2014).

Weaknesses: The primary weakness of a heatmap is that the numerical variable cannot be read precisely. The color encoding may give a sense of what categories are consistently high or low or where the interaction between two categories results in a high point or a low point. Like area encoding, however, (numerical) color encoding does not allow data to be matched precisely to a data value. Humans are simply not proficient at detecting small changes in the brightness or saturation of a color and matching the changes to a data range.

Another weakness of heatmaps is that they are relatively uncommon. While they can be produced even in common spreadsheet programs, they do not appear as an official chart type. This makes them harder to produce than other chart types, and the lack of popularity also means that users will be less familiar with them.

Finally, heatmaps are highly dependent on the order of the rows and columns to reveal similarities between different categories within the same variable. If the categories do not have a natural ordering – say, if they are institution names instead of months of the year – you may wish to inspect the data to see if some of the categories should be placed next to each other (Yau 2013).

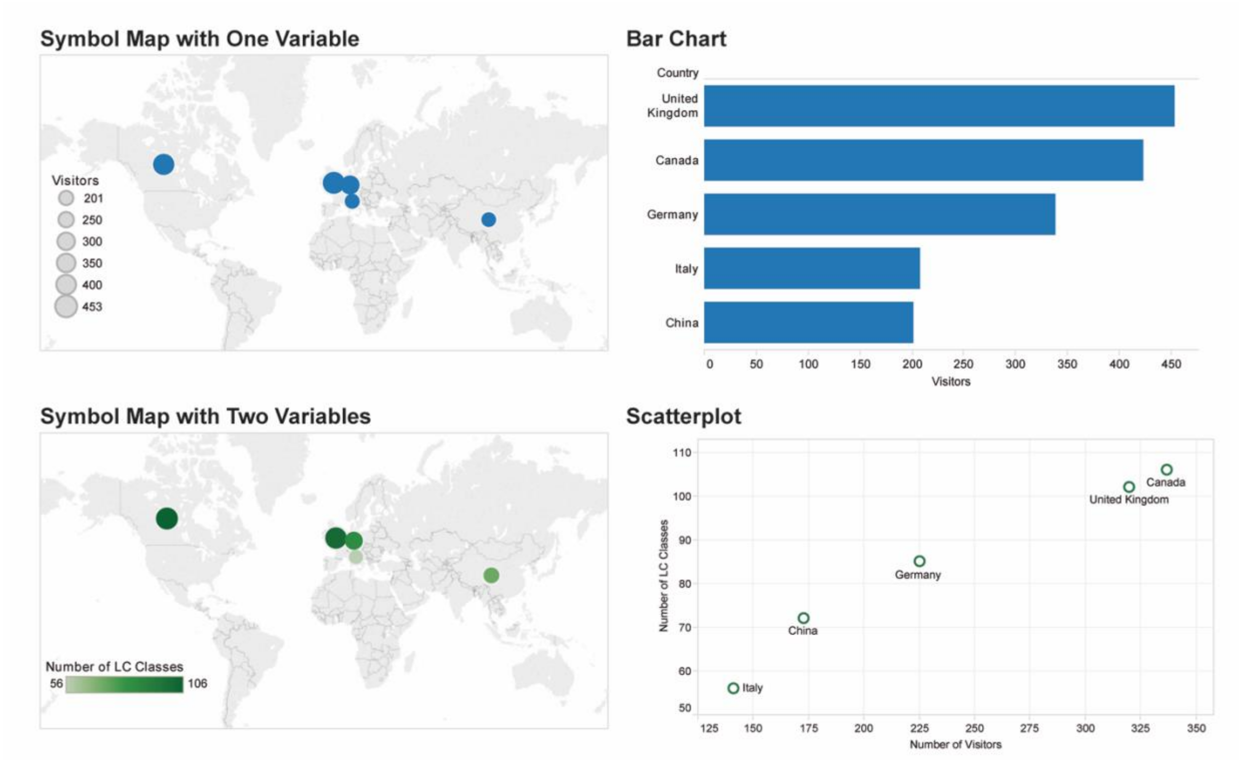

Figure 10. These two symbol maps both focus on locations that are geographically dispersed and variables that are unrelated to the location. For displays of one variable, consider a bar chart instead. If you are trying to show the relationship between two variables, a simple scatter plot may be more effective.

MAPS

Data often include a spatial component. Maps can be a very powerful way to engage an audience and to highlight patterns in data that relate strongly to location. Sometimes, however, location data is a red herring. Sometimes the patterns in your data are not inherently spatial (Wong 2010). Even if location is important, sometimes a map is too constraining to show the data well.

For example, you may have data on where website traffic is coming from (see Figure 10). In all likelihood, most of the traffic is coming from the same country as your university. Maybe a few other countries are represented, but there may be no spatial pattern to where those other requests are coming from. There may be a small number of those external requests, and they might be scattered across the rest of the globe, hard to see among the many countries without data and the large oceans. That would be a lot of space devoted to showing a very small amount of data, and unless location turns out to be important somehow (e.g., requests are only coming from places in a particular time zone), there may be a better visualization type than a map. Remember – position is everything in data visualization. When you use a map, you lose the ability to encode a variable (or two!) by position.

If spatial patterns are important, however, maps can often immediately pull a user in by helping them "find themselves" in the data. Different types of maps allow for different types of data analysis. This chapter will focus on the two most common types of maps for data presentation: choropleth maps and (proportional) symbol maps.

CHOROPLETH MAPS

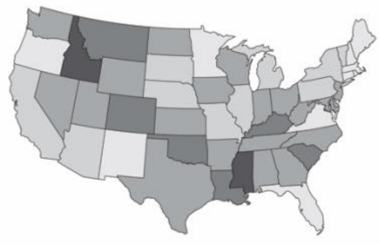

Figure 11. A basic choropleth map.

Choropleth maps are a basic and common map type. They fill the borders of geographic regions (e.g., counties, states) with colors that represent data. They are easy to create with a variety of tools and easy to read. While there are some problems to be aware of, knowing how to use these maps well can lead to maps that are very easy to read.

Data: Choropleth maps typically take two variables – the name of the region of interest and a numerical or categorical variable that gets converted into a color. When using a numerical variable with a choropleth map, cartographers often use number ranges (see the note on histograms) instead of a continuous color scale. This is a way to adjust for unevenness in the numerical variable. For example, if your data values are very skewed and you have a lot of very low values, you might want to limit how many of those low values show up as a very light color. You can set up number ranges that start out very narrow and gradually get larger, grouping together enough data points in the upper ranges that the map shows a good amount of differentiation.

Strengths: Choropleths are especially good for making sure that data points do not overlap each other. Map regions can get very complicated, but in most cases it is easy to draw boundaries and keep data separate from one region to another. Choropleths can also take advantage of audience familiarity with particular region shapes; the shape of a county, state, or country can sometimes stand on its own, without a need for a label which could add visual complexity.

Weaknesses: The major weakness of a choropleth map is that sometimes large regions take up a huge amount of visual space but don't necessarily warrant that much attention. On the iconic "red state/blue state" maps that emerge around election times in the United States, large but sparsely populated states like Montana end up making the country look like a lot of the population leans "red." In truth, the densely populated areas like large cities take up very little visual space but can have a much large effect on the results of an election.

The other major weakness of a choropleth map is that the only variables are a location variable and a single extra variable. Being wedded to both the accurate position and the accurate size of regions means that both of these encodings are unavailable. Special maps called "cartograms" play with this notion by changing the size of the regions while trying to keep them looking approximately the same as the true regions. For example, Montana might shrink down in proportion to the number of Electoral College votes it has, while New York and California would swell quite a bit. The distortion can make these difficult to read, depending on how different the region ends up looking, and it is also hard to find software that will create cartograms.

SYMBOL MAPS

Figure 12. A basic symbol map.

Symbol maps use a standard map in the background, but the data are encoded in circles that are placed on top of the correct locations. These maps are becoming more common, but they fall prey to some of the same problems experienced by scatterplots and bubble charts.

Data: Symbol maps, or proportional symbol maps, place a circle on top of a location of interest. The circles can then be sized by a (positive) numerical variable and colored by a numerical or categorical variable.

Strengths: Symbol maps avoid the limitations of choropleths and can encode two separate variables, in addition to location. They also make it easier to draw attention to small regions on the map that are more important than large regions, without distorting the underlying map. Finally, symbol maps are the standard way of creating maps for data related to cities. While cities do have borders and could be filled, typically the borders of a city are not recognizable, and most cities are so far apart that the filled regions would be much too small to show the data trends.

Weaknesses: When changing filled borders for area-encoded circles, however, there arises the danger of overlapping circles. As with a scatterplot, there are a few possible solutions. Transparency and aggregation are possible solutions. Also, this map requires making comparisons based on area, which is not an easy task. Precision judgments would be better served by a bar chart.

GENERAL DESIGN SUGGESTIONS

In addition to the visualization-specific design suggestions mentioned above, there are other design suggestions that apply across visualizations. This section will focus on the color and text used within visualizations. Color is one of the most complex considerations in the design of visualizations, and careful color selections are essential to the success of the visualization. Text may seem like a strange focus for a section on visualization design, but the text of a visualization plays an integral role in its interpretation.

COLOR

How color is used in a chart or map has a huge impact both on how the audience engages with the chart and also on how easily and accurately the data are interpreted. If chosen carefully, color differences can be perceived almost instantaneously and used easily for data interpretation. On the other hand, color that is chosen poorly can severely impede how well individuals can read the chart, either by creating optical illusions or by obscuring genuine trends in the data. Color contrast is an important consideration for all chart elements, but there are also specific concerns when color is used to encode either numerical or categorical variables.

COLOR CONTRAST

The first aspect of color to consider is how well the colors of the chart or map show up against the background of the chart. This is called contrast, and it is a way of ensuring that, at a very basic level, the elements of your chart will be visible and clear (Few 2012). In general, elements that have the most contrast with the background color will draw the most attention.

For light backgrounds, this means that the most important elements of your chart should be the darkest. Supporting elements, like grid lines, axes, etc., are typically not the most important parts of the chart. It often works well to make these elements lighter and thinner to allow users to focus on the data elements. For some charts and tasks, it is even possible to remove non-data elements entirely; if precise data lookup is not important, or if data points are labeled with the precise data values, it may even be possible to remove the axis.

After the contrast of the non-data elements has been reduced, you can begin to make choices about the chart itself. If you are not currently using color to visually encode a variable, you may have a completely monochromatic chart – all blue bars, for example. Consider using a medium contrast color for this default color. This creates an opportunity to use a high contrast color to highlight one area of the chart or one special data point (Few 2012). Say you want to compare your institution to other similar institutions. You could color the bars for the other institutions in a medium gray color and color the bar for your institution in a darker color – e.g., blue, red.

Even if you are using color to visually encode a variable, there are still choices to make. Contrast between the background color and the foreground (data) element colors must be maximized, but there must also be some contrast between data elements of different colors. The purpose of data visualization is to be able to quickly tell the difference between data elements that are, in fact, different. When data colors are too close to each other, especially for categorical data, the user will have to spend more time and attention to evaluate the patterns in the data. With proper contrast between the data colors, this process become what is called "pre-attentive," or something that happens so fast that we perceive it almost immediate and don't have to focus on it (Healey and Enns 2012).

Keeping data colors distinct from each other is much easier when the colors are mostly dark than when they are mostly light. That is, it is much easier to perceive differences between reds and purples than pinks and lavenders. In truth, some variation in lightness helps a lot both for readability and for the ability to print color graphics in grayscale, but data colors should typically be mostly dark or mostly light in order to select an appropriate background color. Because staying with mostly dark colors increases contrast options, using light backgrounds and dark colors is typically safe. Black on white charts may look boring and antiquated, but there are good perceptual reasons for dark on light, rather than light on dark. (Note, however, that yellow is always perceptually light, even though we normally consider it as a primary color like blue and orange. Yellow does not offer enough contrast against a white background to be visible.) The next two sections focus specifically on choosing colors for encoding either numerical or categorical variables.

COLOR FOR NUMERICAL VARIABLES

Using color to represent a numerical variable is not common for basic charts and graphs. Aside from heatmaps, most basic charts only use color for categorical variables. Choropleth and symbol maps, on the other hand, very commonly use color encoding for numerical variables. If your favorite software tool does allow you to encode numerical data as a color, these suggestions should apply equally well to maps and charts.

When converting a numerical variable to a color, there are several choices to make. First, does the variable have a natural middle, like a variable that starts negative, goes through a zero point in the middle, and finishes positive? If so, you may want to emphasize both extremes as if they are two categories and deemphasize values around the middle. In terms of contrast, that usually means the colors you use for the extremes should both have high contrast with the background of the chart, as well as high contrast against each other. The middle values, instead, should have relatively low contrast against the background[7]. This type of

[7] This actually may differ between maps and charts where the data are less densely positioned. In a choropleth map, for example, the fact that the background behind the map is white may not matter if you are filling in every state with a color. Each state will be surrounded by other states filled with some kind of color. If a state filled with a light color is surrounded by states filled with dark colors, that state may actually draw more attention instead of less attention. You may sometimes want to choose a darker center color – say, a medium gray – so it doesn't pop so much in context.

color scale is called a “diverging” color scale because two colors diverge from the center color.

When choosing the colors to use for the extreme high and low values, remember that a significant portion of the general population experiences red-green color blindness. Combinations like red-blue, orange-blue, orange-green, red-purple, etc., avoid this problem.

If there is no natural middle to your data, which side of the variable is the most important – the upper extreme or the lower extreme? Say you are making a map that is supposed to reveal areas of low income. You may want users to focus on high values if your variable is the percentage of people under the poverty line, but you may want users to focus on low values if your variable is average income. Use the most contrast for the end of the number scale that is most important.

If you have ever seen a weather report that shows the temperature across the country, you have likely seen a map that uses the wavelength spectrum, or a “ROYGBIV” rainbow, to encode a numerical variable. Figure 13 shows an example of this using the average daily high temperature per county in the contiguous United States (McClure et al. 2013). Reds indicate high temperatures, blues or purples indicate low temperatures, and the temperatures in the middle go through yellows and greens. Unfortunately, despite how common this is, using a rainbow for numerical variables is a terrible match for human perceptual abilities. Many excellent analyses of rainbow color scales have already been written (e.g., Borland and Taylor 2007, Munzner 2014), but the arguments that relate most to information visualizations like this are the arguments about visual artifacts, loss of detail, and salience.

Visual artifacts are like optical illusions; they look like real data, but using a different color scale would show a different pattern, or show that there is no real pattern there. When you use a rainbow color scale, the data tend to look striped – like there are big jumps in the values in certain parts of the number range (Borland and Taylor 2007). On weather maps, this usually shows up as a sharp line between red and yellow, between yellow and green, etc. If your data don’t have sharp breaks, it certainly isn’t helpful for the color scale to add them for you.

Loss of detail happens in some parts of the rainbow because of some of the properties of certain colors. The range that is especially problematic is the yellow to light green range (Borland and Taylor 2007). As we mentioned earlier, two light colors are not as easy to differentiate as two dark colors. Rainbow color

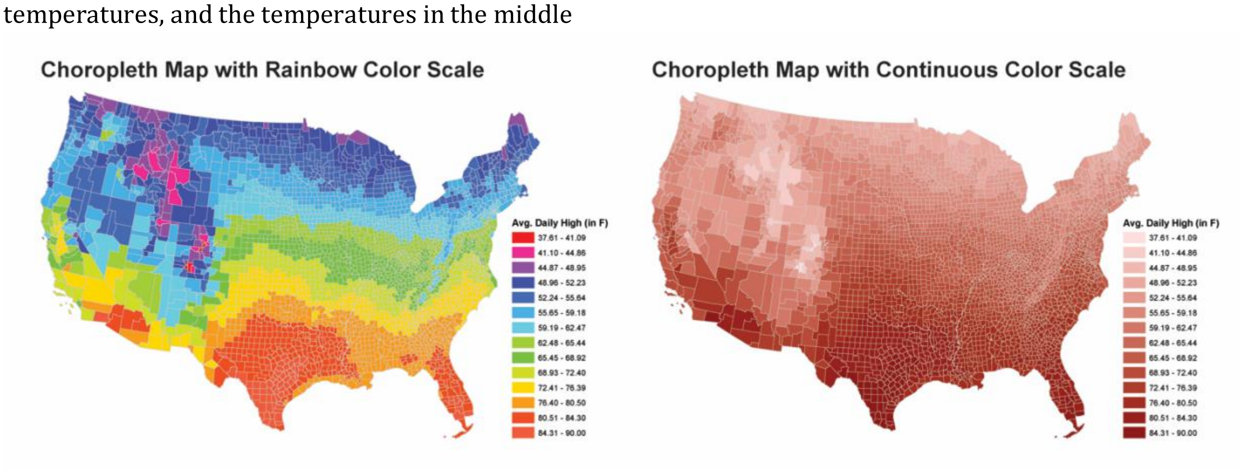

Figure 13. Above, two choropleth maps of the counties of the contiguous United States show how different colors scales can reveal different patterns. On the left, the numerical variable – average daily high temperature in Fahrenheit (McClure et al. 2013) – is encoded with a rainbow color scale. With this color scale, data points that are very close to each other can look very different because of the rapid changes in hue. Using a color scale that goes from a very light shade to a very dark shade is a more faithful representation and helps focus attention on the true regions of interest.

scales can devote as much as a third of the color scale to the colors yellow, light green, and sometimes cyan. The data points that fall into this range will be very difficult to tell apart, making it looks like all of these data points are much more similar to each other than they really are.

The final problem with rainbow color scales is salience (Wong 2011). Salience is a way of talking about how much a particular color "pops out" and draws attention. Salience is related to contrast; light colors against a dark background are an iconic example of visual salience. The rainbow color scale has a built-in salience problem – the yellow and light greens, again, will always draw attention away from the ends of the scale, which are typically dark reds and dark blues. You may notice that this make the rainbow sound more like a diverging color scale. Indeed, using a rainbow can create a similar effect to using a diverging color scale, but they are seldom used on data with a natural middle point, and even having a natural middle point does not resolve the striping and data loss issues.

For a numerical variable with no natural middle point, consider using a color scale that goes from white to a dark color – black, purple, red, etc. The continuously increasing (or decreasing) lightness is incredibly important for accurately representing data. If your data are not evenly distributed, though, you might consider using number ranges to give yourself more control over how data values are associated with shades of a color; see the previous section on choropleths for more information. Note that there is a limit to the number of different shades humans can differentiate. It is usually safe to use between five and seven discrete shades of a color for numerical variables (Gilmartin and Shelton 1990), assuming that the darkest version of that color is quite dark.

COLOR FOR CATEGORICAL VARIABLES

Using color to encode a categorical variable is quite common. The discussions above show that it occurs in almost every basic chart type. In charts like pie charts and line charts, color often simply stands in for the name of a data point, and each color is used only once. In bar charts and scatterplots, it is more likely that the category will appear multiple times and that color is the only thing connecting the various data elements within the category.

In either case, however, what matters is that the data colors can easily be distinguished from each other. The basic rules of contrast are the primary focus for categorical variables. Again, using a light background and dark data colors offers the most options. With numerical variables, we talked about shades of a single, dark color. With categorical variables, we are talking about different color classes, or "hues." Humans begin to have trouble distinguishing between hues when there are more than about seven separate hues (Healey 1996). Even at seven hues it is difficult to ensure high foreground/background contrast – seven hues often require using a fairly light color like yellow, which will begin to introduce salience concerns on top of the differentiation concerns.

Choice of color will be highly dependent on the patterns and features in your data. Some categorical data have culturally determined colors; in the U.S., we often use pink and blue for Female and Male, or red and blue for Republic and Democrat. If your data have these associations, reinforcing them with color can improve the readability of your chart (Lin et al. 2013). It can even increase the number of categories you can represent; if you have a map of different types of natural resources, you may be able to use three different shades of green for different types of forests or wooded areas. This works best if the shades are positioned close to each other; if you have two different light green colors and they are on two different sides of your visualization, it will likely be difficult to tell them apart.

TEXT

Many people don't think about text as being a major component of a chart or a map, but good use of text can make or break a visualization. Especially when you are sharing data with a broad audience, informative text can help users understand how to read the chart and what trends in the data are most important.

DESIGN

Text design can enhance or detract from the look and feel of the chart. Paying attention to font choice, font size, and text rotation can create a professional look for your chart and even make the chart easier to understand.

Font choice is a way of influencing the style of your charts. If the chart appears in a report, article, or poster, there may already be a chosen font that you would like to continue to use or complement. If you are selecting a font from a blank slate, you just need to pay attention to the "mood" of the font. Try not to pick anything too stylized or too informal if you are going for a more professional look.

When choosing font size it is essential to consider the final size and context of the visualization. It is much more likely that you will discover your fonts are too small than the reverse. For example, when you use a visualization in a slideshow presentation, the chart needs to be easy to read from a long distance away, possibly in an improperly lit space. The same is true for posters. For reports and papers, visualizations may need to fit into a single column on a two-column page, or they may be resized to help the report flow better. Err on the side of very large text.

Having text that is easy to read means that text will act in service of the visualization, instead of distracting from it. Ways to make text easier to read include making sure text is horizontal wherever possible (Byrne 2002). As mentioned previously, rotated text problems occur often with bar charts. Try using a horizontal bar chart instead of a vertical bar chart so the labels will not have to be rotated or wrapped onto multiple lines. Rotated text is also common for y-axis labels. While leaving the text rotated for the axis label is less problematic than for category labels – users typically only need to read an axis label once – it is worth trying to keep the text horizontal if the words are short.

ACTIVE TITLING

One safe rule of thumb is to make sure that every axis (or visual encoding) is labeled. Sometimes chart designers will use the title of the chart to describe the variables in the chart and leave the axes unlabeled. If you label the axes directly, however, it actually becomes unnecessary to duplicate this information in the chart title. Instead, the chart title becomes an opportunity to help users understand the chart better.

Consider using a practice called "active titling." With active titling, the standard boring chart title transforms into an active statement of the importance of the chart. Instead of "Change in eBook transactions over time," you can make a concrete statement about the change *was* over time – for example, "eBook transactions have been doubling each year."

ANNOTATIONS

Annotations, or special informative text or reference points that have been added to visualizations, take the principle of active titling one step farther. Annotations can offer users additional contextual information, can focus attention on a single data point or specific trend, or can add an additional reference against which to evaluate the data. Examples include paragraphs of text placed on or near the chart, special labels for points, additional lines or boxes added to areas of graphs, even a "callout" box that zooms into or out of the data in the chart. While it's important not to overwhelm the reader or create too much visual complexity, it's also important to make sure the chart is understandable, especially if it is being designed for a broad audience (Yau 2013).

Different publication venues may limit the use of annotations, however. For example, a paragraph of explanatory text may not be appropriate for a PowerPoint presentation. Instead, consider creating several versions of the chart, each focusing on a small aspect or adding one small annotation at a time. On the other hand, web-based visualizations may offer wonderful opportunities for "on demand" annotations. Perhaps annotations can appear when the mouse hovers over a data point and go away again when they are no longer needed, thus reducing visual complexity.

Most programs that create visualizations will allow you to customize the text on the charts or maps. If not, consider removing all of the text from the chart and adding it back in later with a graphic design program or even a simple drawing or presentation program. The same is true for most annotations; even if you are adding a reference line, it may be easier to figure out how to add this in with a drawing program than to figure out how to get a visualization program to do it. Just make sure any manual edits look professional and are in the same style as the original chart.

STORYTELLING WITH DATA

These days, it isn't quite enough to produce a well-designed and accurate visualization. The focus is increasingly on storytelling with data, or on creating some kind of narrative arc that ties multiple visualizations together into a clear and compelling progression. This can happen sequentially over time, as in a digital slideshow, or by arranging the visualizations in space, as with an infographic, online dashboard, or poster.

NARRATIVE DESIGN

Designing a narrative with separate visualizations is a lot like studying how a user explores data with a single visualization. For many years, the field of data visualization has subscribed to Ben Shneiderman's "visual information seeking mantra" (1996), which states that the following information seeking tasks should be prioritized in the design of interactive visualization systems: "overview first, zoom and filter, then details-on-demand." Several additional tasks, including a "relate" task defined as "viewing relationships among items" (*ibid.*, p. 2), are listed in the paper but not included in the mantra.

Applying these tasks to a narrative, it makes sense to start with an overview of the data you want the user to understand. Where in an interactive visualization you could literally zoom into parts of the visualization, in a narrative you would create a separate visualization that is already zoomed or filtered – that is, limited in numerical ranges or categories or time periods. Perhaps to show this different level of detail, you will also want to change the chart type. For details on demand, you may be able to generate a series of small, highly focused charts that allow users to focus on the details that are most interesting to them. Finally, to summarize what the data suggest, using the additional "relate" task to build back to an interpretation of the data will help users understand why you are drawing certain conclusions from the data.

There are other ways to structure narratives across multiple visualizations. Few (2012) recommends designing visual communication by following this organization process: "1. Group (i.e., segment information into meaningful sections), 2. Prioritize (i.e., rank information by importance), 3. Sequence (i.e., provide direction for the order in which information should be read)" (p. 144). Going through this process is especially helpful when determining the spatial layout of your visualizations, which will be addressed in the next section.

Finally, a recent study by Hullman et al. (2013) presents a survey of transition types found within a sample of narrative visualizations from the field of journalism, as well as visualization blogs and repositories. The study found six transition types (presented in decreasing order of frequency): temporal, granularity, comparison, causal, spatial, and dialogue. Some kind of temporal transition, or moving from one time period to another, appears in almost 90% of the examined narratives. Transitions in granularity (i.e., from general to specific or vice versa) mirror closely the "overview first, zoom and filter" components of the visual information-seeking mantra (Shneiderman 1996). The final transitions – those of comparison, causality, spatial proximity, and dialogue, each extend the normal approach to designing a user experience with a visualization by exploring the different types of connection that can occur between different visualizations. When considering how to organize visualizations into a larger narrative, it may be necessary to move back and forth between general questions of organization ("What are my priorities?") and specific questions of arrangement ("What transition makes the most sense here?").

ARRANGEMENT

After the content of the visualizations has been selected, and the visualizations have been generated, it's important to lay the visualizations out in an appealing and understandable arrangement. Especially for posters, infographics, or dashboards, the size and arrangement of visualizations have a huge impact on how users read the display.

The arrangement of panels in a graphic novel or comic book offers a great example (McCloud 1993). The size and arrangement of the panels must guide the reader's eye so that he/she follows the narrative in the correct order. Cultural differences can also play a part here; is your audience likely to start in the upper left corner or the upper right corner? Is the primary reading directly horizontal or vertical? When you have a diverse audience, you will want to use visual elements to help guide the eye and encourage users to follow the narrative in the way or ways that will make the most sense. Typically, in Western cultures, the most important elements are placed at the top left, and additional elements follow left-to-right first and then top-to-bottom. Consider using size as well as position to emphasize which elements are most important.

Gestalt principles explain some of the native tendencies of humans when they are interpreting visual elements (Ware 2008). For example, the Gestalt principle of enclosure says that visual elements that are all contained within an enclosing shape will be seen as more similar to each other than to elements outside of that shape. To apply this principle, any elements contained within a box or circle will be interpreted together, separately from others. Another Gestalt principle, the principle of proximity, states that objects that are close together will seem more similar than those that are far apart. This offers another practical solution; you can actually simply use whitespace in your arrangement to guide the eye between elements. Make the spacing tight between elements that should be interpreted as a single group, then increase the space between that group and any other groups (Few 2012).

Finally, the arrangement of elements in your layout should take into consideration graphic design principles like the rule of thirds and the golden ratio (see Figure 14). When arranging elements in visual space, having perfectly centered and symmetrical arrangements is not always considered the most aesthetically pleasing. The rule of thirds suggests that the focal point of an arrangement should be determined by dividing a space into thirds both horizontally and vertically, and then placing visual elements along those divisions or at the intersections of the divisions. The elements of interest in the display thus get placed slightly off center but not all the way at the edge of the visual field.

The golden ratio is used to create rectangles where the long end is about 1.618 times the length of the short end, or almost 2/3 longer. The exact

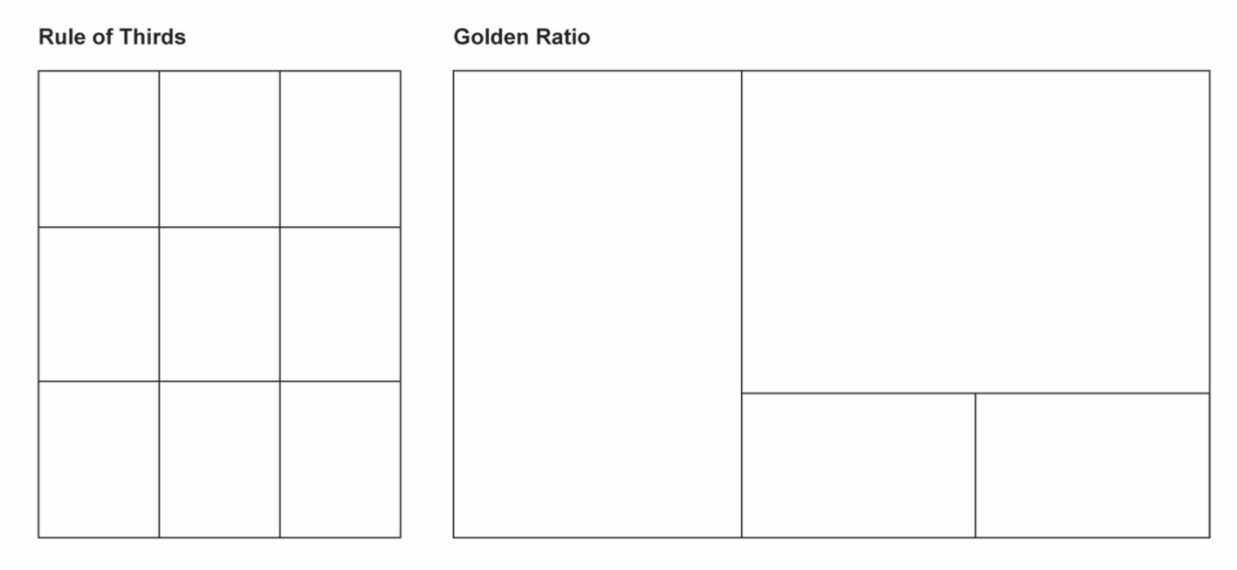

Figure 14. Consider using the rule of thirds or grids made up of the golden ratio to size and arrange visual elements.

ratio is perhaps less important than the idea that nestling golden ratio rectangles within other golden ratio rectangles creates another type of arrangement that keeps elements slightly off center and encourages changes in size, which in turn encourages users to focus attention differently on different elements. Combining the golden ratio and the rule of thirds with more basic grids can generate very visually appealing narrative visualizations.

DESIGN CONSISTENCY

In any situation where you are combining multiple visual elements, however, it is important to use graphic design to ensure consistency across all elements. This will present a more professional output, but it will also aid comprehension by allowing users to learn design conventions once and apply them successfully to multiple visualizations (Wainer 2008).

Design consistency extends from using the same or compatible fonts across all text and visual elements to the use of the same color scheme (or same specific colors) from one chart to the next. A special case of design consistency arises when you create multiple visualizations from the same variables. Small multiples, also known as trellis plots or lattice plots, are created when a complex chart like a line chart with too many lines is split up into a series of nearly identical charts, each showing one category. In this situation, it is not only essential that any data colors are kept consistent, but it is also necessary to make sure that the range on the axes, the gridlines, and other reference marks are also kept consistent (Cleveland 1994). To make accurate judgments about how the data differ from one category to the next, the reference systems for each chart have to be identical.

CONCLUSION

The considerations presented in this chapter will hopefully provide a foundation for developing visualizations that can appeal to a broad public. The process, however, is likely to be much less linear than the information here suggests. You may find while building a narrative that a new chart is necessary for a particular transition. You may realize that you have several different charts all using color in different ways and decide to try to re-encode at least one of those variables in some other type of graphical element. Specialized visualization software like Tableau (http://www.tableau.com) can be especially helpful during planning phases because of the ease with which additional charts can be created. The best advice for creating great visualizations is often to create as many draft visualizations as possible and to seek out feedback on your ideas. With practice, you will develop your own style and intuitions, and your visualizations will become increasingly clear and engaging.

REFERENCES

Borland, David, and Russell M. Taylor. 2007. "Rainbow color map (still) considered harmful." *IEEE Computer Graphics and Applications* 27 (2):14-17.

Börner, Katy. 2015. *Atlas of knowledge: Anyone can map*. Cambridge, MA: MIT Press.

Byrne, Michael D. 2002. "Reading vertical text: Rotated vs. marquee." Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting, Santa Monica, CA.

Cleveland, William S. 1994. *The elements of graphing data*. Murray Hill, NJ: AT&T Bell Laboratories.

Cleveland, William S., and Robert McGill. 1985. "Graphical perception and graphical methods for analyzing scientific data." *Science* 299 (4716):828-833.

Few, Stephen. 2012. *Show me the numbers: Designing tables and graphs to enlighten*. 2nd ed. Burlingame, CA: Analytics Press.

Gilmartin, Patricia, and Elisabeth Shelton. 1990. "Choropleth maps on high resolution CRTs: The effects of number of classes and hue on communication." *Cartographica* 26 (2):40-52.

Healey, Christopher G. 1996. "Choosing effective colors for data visualization." Proceedings of the 7th conference on Visualization '96 (VIS '96), San Francisco, CA.

Healey, Christopher G., and James T. Enns. 2012. "Attention and Visual Memory in Visualization and Computer Graphics." *IEEE Transactions on Visualization and Computer Graphics* 18 (7):1170-1188.

Heer, Jeffrey, and Michael Bostock. 2010. *Crowdsourcing graphical perception: using mechanical turk to assess visualization design*. New York, NY, USA: ACM.

Hullman, Jessica, Steven Drucker, Nathalie Henry Riche, Bongshin Lee, Danyel Fisher, and Eytan Adar. 2013. "A Deeper Understanding of Sequence in Narrative Visualization." *IEEE Transactions on Visualization and Computer Graphics* 19 (12):2406-2415.

Kandel, Sean, Jeffrey Heer, Catherine Plaisant, Jessie Kennedy, Frank van Ham, Nathalie Henry Riche, Chris Weaver, Bongshin Lee, Dominique Brodbeck, and Paolo Buono. 2011. "Research directions in data wrangling: Visualizations and transformations for usable and credible data." *Information Visualization* 10 (4):271-288.

Lin, Sharon, Julie Fortuna, Chinmay Kulkarni, Maureen Stone, and Jeffrey Heer. 2013. "Selecting semantically-resonant colors for data visualization." Proceeedings of Eurographics Conference on Visualization (EuroVis) 2013, Leipzig, Germany.

McCloud, Scott. 1993. *Understanding comics: The invisible art*. New York, NY: HarperPerennial.

McClure, Leslie , Mohammad Al-Hamdan, William Crosson, Sigrid Economou, Maurice Estes, Jr., Sue Estes, Mark Puckett, and Dale Quattrochi, Jr. 2013. North America Land Data Assimilation System (NLDAS) Daily Air Temperatures and Heat Index (1979-2011). CDC WONDER Online Database.

Munzner, Tamara. 2014. *Visualization Analysis & Design*. Edited by Tamara Munzner, *A K Peters Visualization Series*. Boca Raton, FL: CRC Press.

Shneiderman, Ben. 1996. "The eyes have it: A task by data type taxonomy for information visualizations." Proceedings of IEEE Symposium on Visual Languages, Boulder, CO.

Skau, Drew. 2011. "2D's company, 3D's a crowd." *Visual.ly blog*. http://blog.visual.ly/2ds-company-3ds-a-crowd/.

Stefaner, Moritz. 2012. Data Stories episode #5: How to learn data visualization (with Andy Kirk). In *Data Stories*, edited by Enrico Bertini and Moritz Stefaner.

Wainer, Howard. 2008. "Improving graphic displays by controlling creativity." *Chance* 21 (2):46-53.

Ware, Colin. 2008. *Visual thinking for design*. Burlington, MA: Morgan Kaufmann Publishers.

Ware, Colin. 2013. *Information visualization: perception for design*. 3rd ed. Waltham, MA: Morgan Kaufmann Publishers.

Wong, Bang. 2011. "Points of view: Salience to relevance." *Nat Meth* 8 (11):889-889.

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

Yau, Nathan. 2013. *Data points: Visualization that means something*. Indianapolis, IN: Wiley.