Wasfi Momen
CSC 4730
Tuesday June 27, 2017

Homework 1

Choropleth of US unemployment rate

The first example comes from Mike Bostock's gallery of visuals from the javascript library D3.js. The visualization shows the unemployment rate of August 2016 over all counties in the United States. Differing values of blue express the data with white being the counties with low unemployment rates and dark blue displaying high unemployment rates—the rates are marked as percentages from zero percent to one hundred percent.

Within the data, the values of unemployment rate lie on the same row as the id of that county. The source of the data comes from the US Bureau of Labor Statistics and the Census Bureau. At the time, the political season of the next US election was underway, so the target readership consisted of voters paying attention to the unemployment rate in specific districts of voting consisting of several counties.

For the main readership of the visualization, the purpose of its design decisions convey straightforward meaning—showing differing unemployment rates at the smaller county to the larger state levels. In the visualization, areas of the counties make up the main visual unit, while value represents the main visual variable being manipulated to convey the data. In a table format, this would be:

Visual Variable Type	Data
Value	Unemployment rate (percentage)

The main design decisions consist of: choosing a projected map of the United States as a cognitive basis for the reader, value differences to depict the change of unemployment rate, and color value to depict boundaries between states and counties within the map.

Overall, I do not believe that this is a fully effective visualization of unemployment rates in the United States. Using a map of the United States provides a well-known cognitive basis for the reader, but introduce new visual variables. Since there is no standard of county size or shape, shape and size become implicit visual variables that might mislead the reader into thinking they have an effect on the data portrayed. However, the author clarifies the data being shown with a scale of color value, but looses the visual property of order when depicting counties with similar values close to one another.

Instead, the visualization should change the visual units and visual variable. The new unit should depict the area of voting districts instead of counties to match closer to the level of detail the main readership wishes. The visual variable of color value should be changed to a more extreme range for each section of the scale and perhaps range from color hue instead of saturation, thus making the visual variable more distinctive. For example, the districts would now have a separate color for each value in the scale. One to two percent would be blue, two to three percent yellow, and so on for each section to eliminate the issue of data loss due to similar values close to one another.

Average Days Spent during Adulthood

Flowing Data's infographic on the average days spent for adults uses data from the American Time Use Survey from 2011 to 2015. Colored squares stacked in a single vertical bar show the average time spent in days for each activity.

The American Time Use Survey data contains several categories of activities arranged in a parent-child relationship with the parent being the general description of the child activities and the children being specific target activities that are defined under a lexicon. An example would be a parent being the descriptor of "Sports" and a child being "Playing Baseball". The visualization aggregates each child under each parent to give the average time spent on each category of activity. The general main readership for the visualization is the same as the data source, adults from age 18 to 79 years of age that have an interest in seeing their time spent on certain activities.

For this visualization, the design decisions focus mainly on using the space in the graphic to describe the data. The main visual unit is the area of a square representing a day, so the visual variable of size is manipulated to show the aggregation of days in each category. The stackable nature of the visual unit provides the perspective of the size visual variable with larger sizes of many squares showing lots of days spent on the activity and smaller sizes showing less days spent. Color hue is a secondary visual variable that serves to differentiate between each activity in the data, allowing for selection for each category and association of colors for each activity in the visual search for the reader.

Visual Variable Type	Data
Size	Amount of Days Spent
Color Hue	Type of Activity

The design of the visualization accurately displays the data for each activity. The compare and contrast cognitive understanding of the data is reinforced through the size of multiple squares. The quantitative nature of the data is clearly expressed by the areas of multiple squares to contrast the sizes, but lacks clarity at the detail level between similarly sized categories. However, the author provides the actual number of days spent on the title of each activity as a secondary structural element, therefore alleviating loss of quantitative detail.

On the other hand, many quantitative elements of the visualization are strong, but could use some reinforcement. For example, the stacked nature of the bars of each activity impede the visual search for the reader since comparing the larger areas at the top and the smaller areas at the bottom is difficult. Instead, it'd be better to arrange the data either across both axes from top left largest area to bottom right smallest or even the largest area in the middle with all other areas surrounding it would help aid the user to compare and contrast the data.

New York Times Yield Curve

The yield curve of the various countries makes up the data for the New York Times 3-D view of the Yield Curve. The visualization depicts a three dimensional surface representing the yield curve of different countries and uses different views to show the same data but reach different conclusions.

In the yield curve data, the interest rates of the government is compared to time, in years. The three dimensional axes shows the x-axis as the short-term 1990 to 2014, the y-axis as the long-term growth rate, and the z-axis as the percent of yield per year for the government's interest rate. Every line in the graph is mapped to either a single year or multiple years for both the short term and long term views. Color value separates the high and low interest rates on the surface.

Looking at the time on the article, this visualization was made at the beginning of 2015. The beginning of each year signals an economic check, since most countries' fiscal year splits halfway at the beginning of the new year. It can be inferred that the main readership comprises of people interested in interest rates based on political and economic pressures in the United States.

Based on the main quantified data of interest rate yield, the design decisions of the visualization center on the ability for readers to see the interest rates in lengths of time. Since the data in the short-term view would be less flat for a yield curve, it made sense for the x-axis to contain smaller increments of time and have a longer axis than the long-term view. Extruding the long-term surface from the edge of the short term curve provides the top-down views in slides six and seven of the visualization, and relate the time in two different axes to the same quantified variable of interest rate. Other decisions include the gridlines for both the xy-plane and white lines on the surface for the top-down views to clearly show values and the changes in value across the surface define the high vs low interest rates.

For the significant variables in the visualization, the main visual unit is represented as a three dimensional surface. In the top-down views and axes views, the authors exposed the change of position and orientation of the surface visual unit in order to change the meaning of the data portrayed. Line and value are visual variables within the visualization that depict the data, however, line is the main descriptor of the data since it is relating the time to the interest rates while value only differentiates the surface values of interest rate along the y axis.

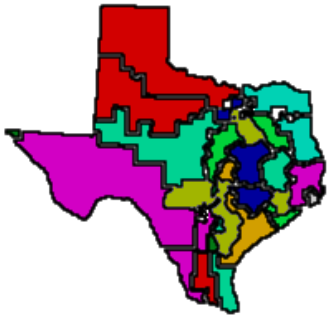
Visual Variable Type	Data
Line	Time vs Interest Rate Yield
Value	Interest Rates (percentage)

To depict data in multiple dimensions, three dimensional visualizations catch the eye of viewers, but most of the time contribute chartjunk and unnecessary extrusions that do not relate to the data. In this visualization, these problems are alleviated to an extent with the extrusion being an actual dimension related to the data in terms of long-term interest rate yield and the chartjunk is necessary since the color value of the interest rates needs to be separated by ticks in the scales to avoid confusion.

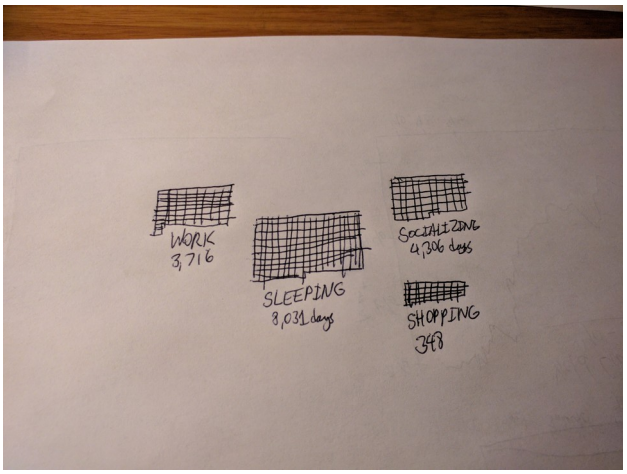
However, the visualization fails to work an important part of data visualization—visual navigation. It is very difficult to understand and interpret the data without assistance from the text boxes. Even if readers knew what a yield curve depicted and the axes were labeled, the visualization by itself would still be very difficult to handle. The similarity between the x and y axis of times, the fact that the color value only highlights the data on one axis, and the attempt of bottom-up guidance but never actually teaching the user how to read the data makes the three dimensional aspect of the visualization flawed. A better visualization would be a simple two line graph with two y-axes for time.

Revisualizations

Choropleth: More variation of values, districts instead of counties to decrease visual load.



Average Days Spent in Adulthood: Largest data in center, smaller data around it.



New York Times Yield Curve: better represented as two graphs instead of the 3d visualization.