

Riley Payung

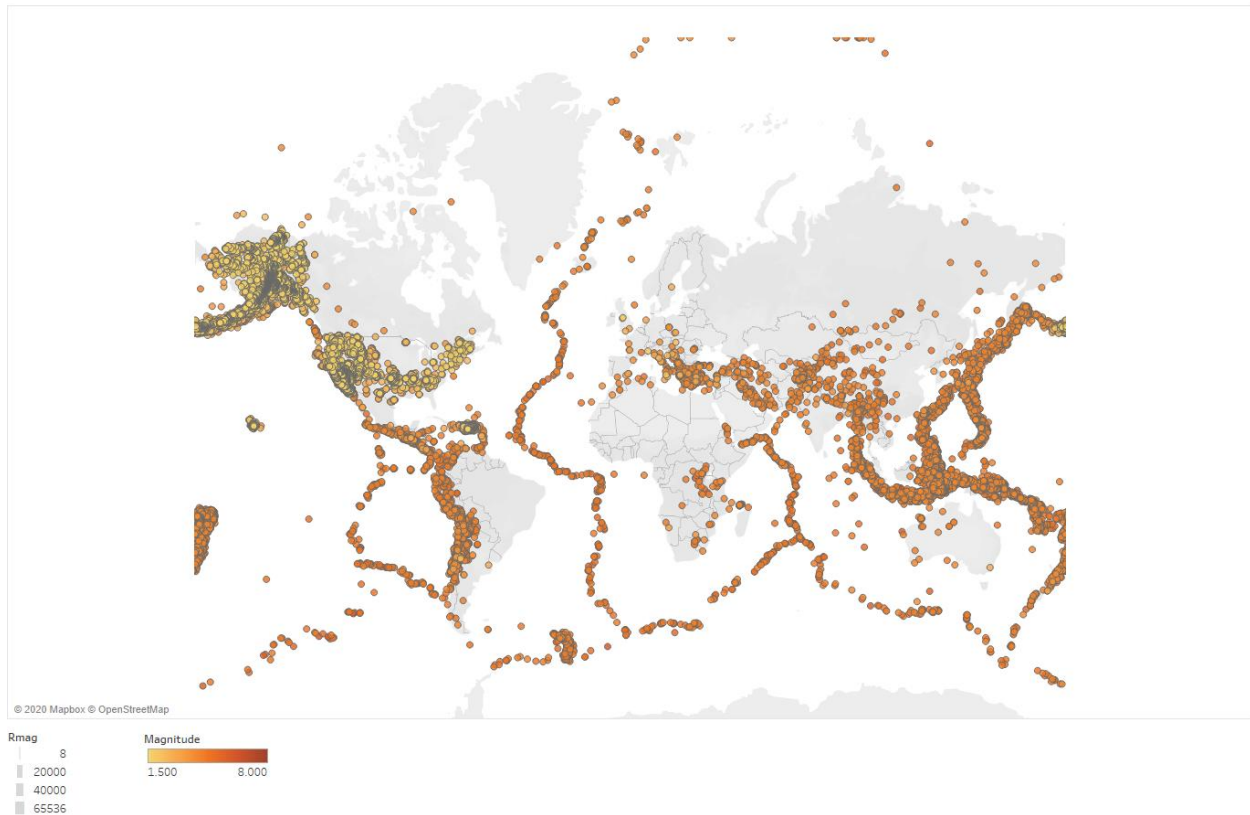
CDS 301

05/14/2020

Final Exam

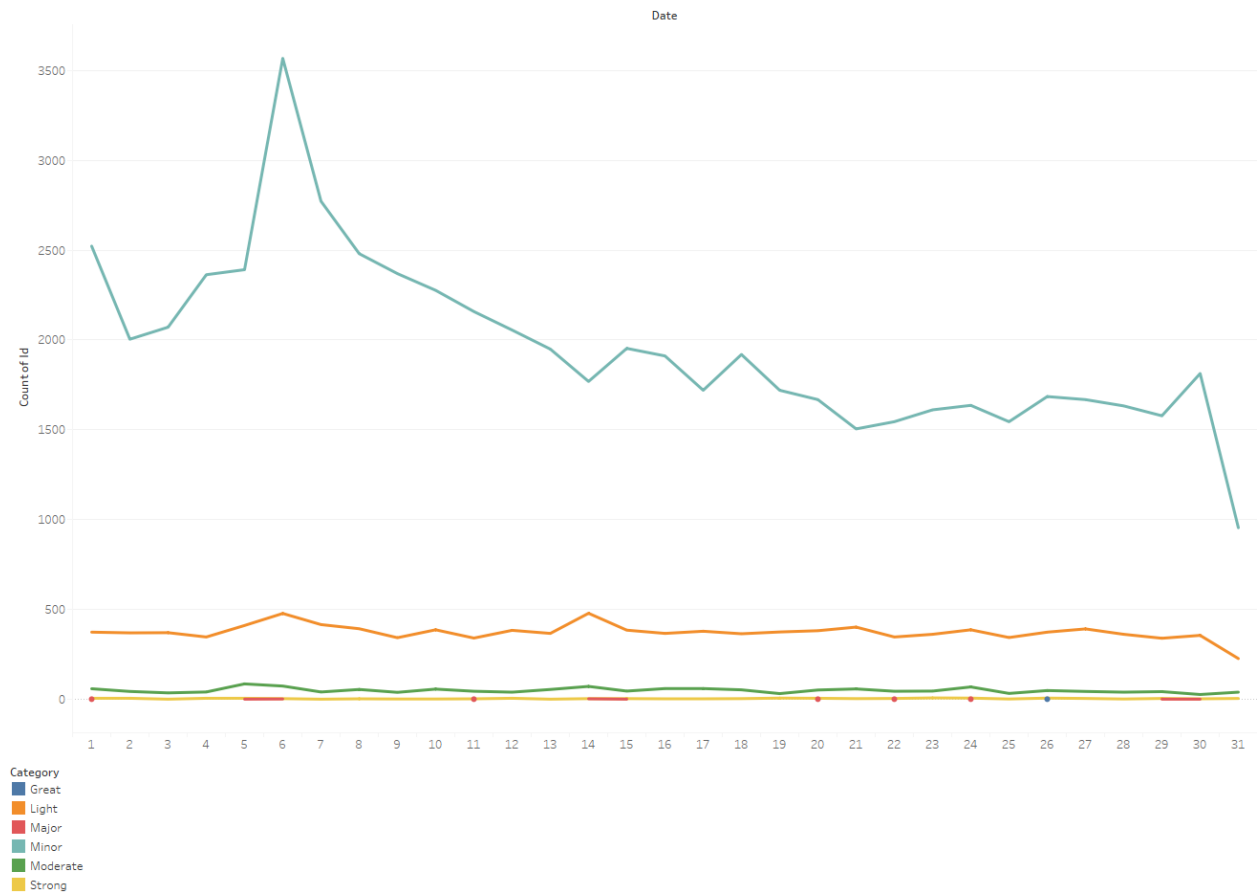
## Problem 1.1

Map of Earthquakes



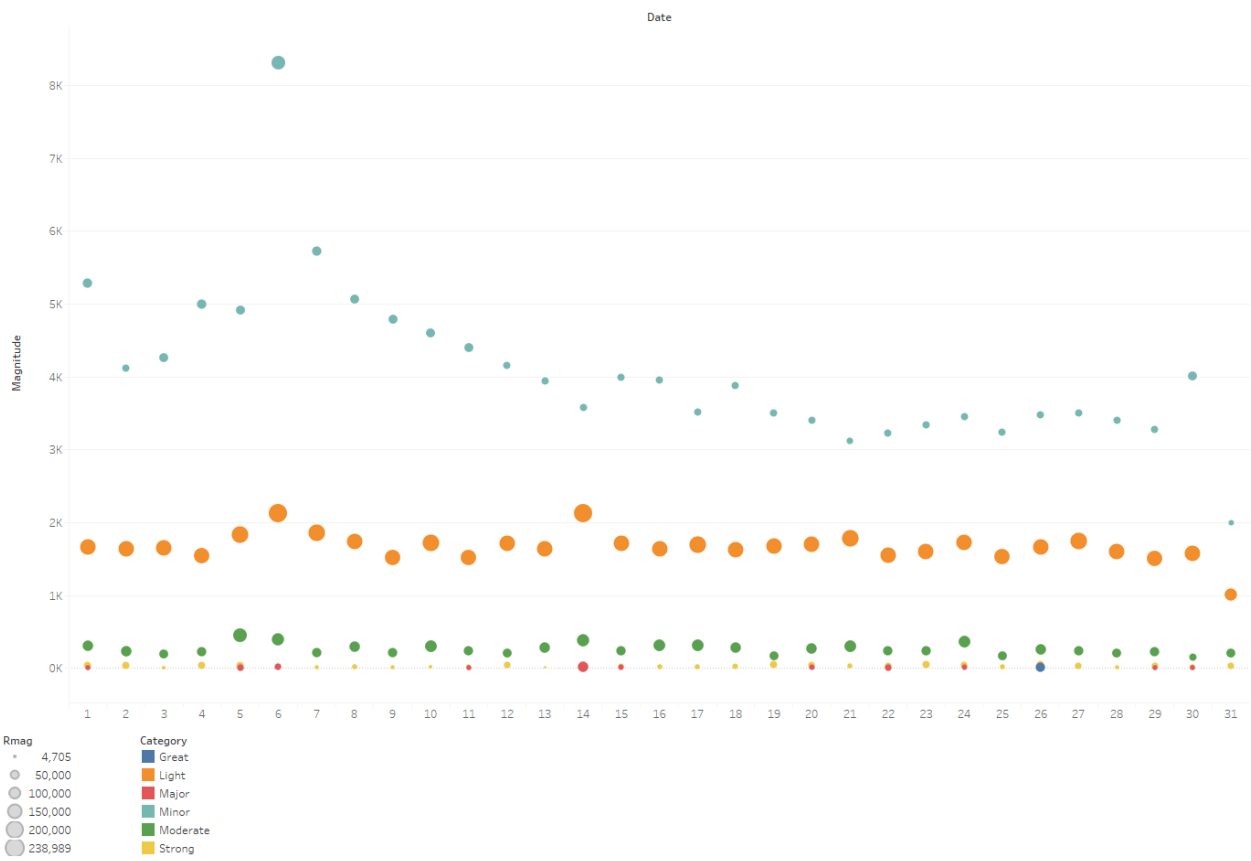
## Problem 1.2

Number of Events per Day



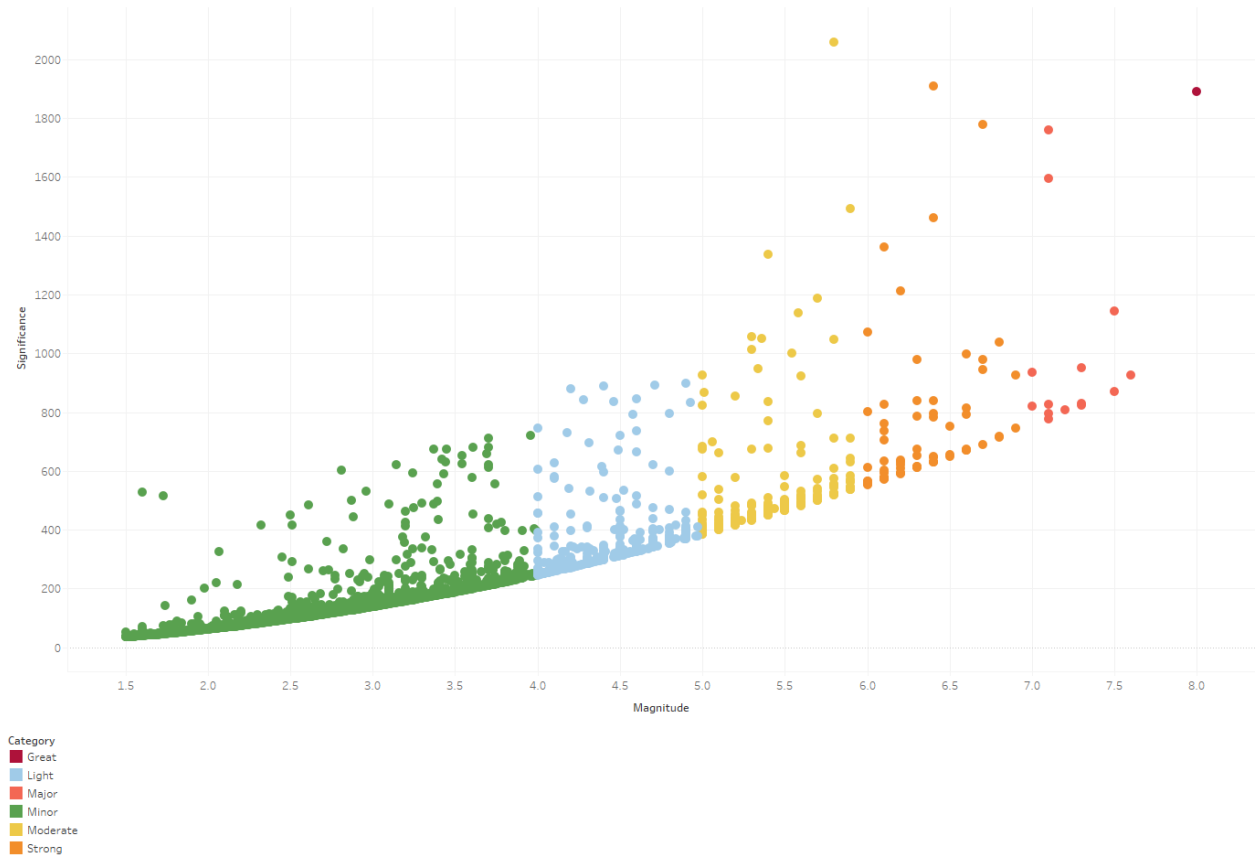
# Problem 1.3

Bubble Chart: Magnitude on Days of Month



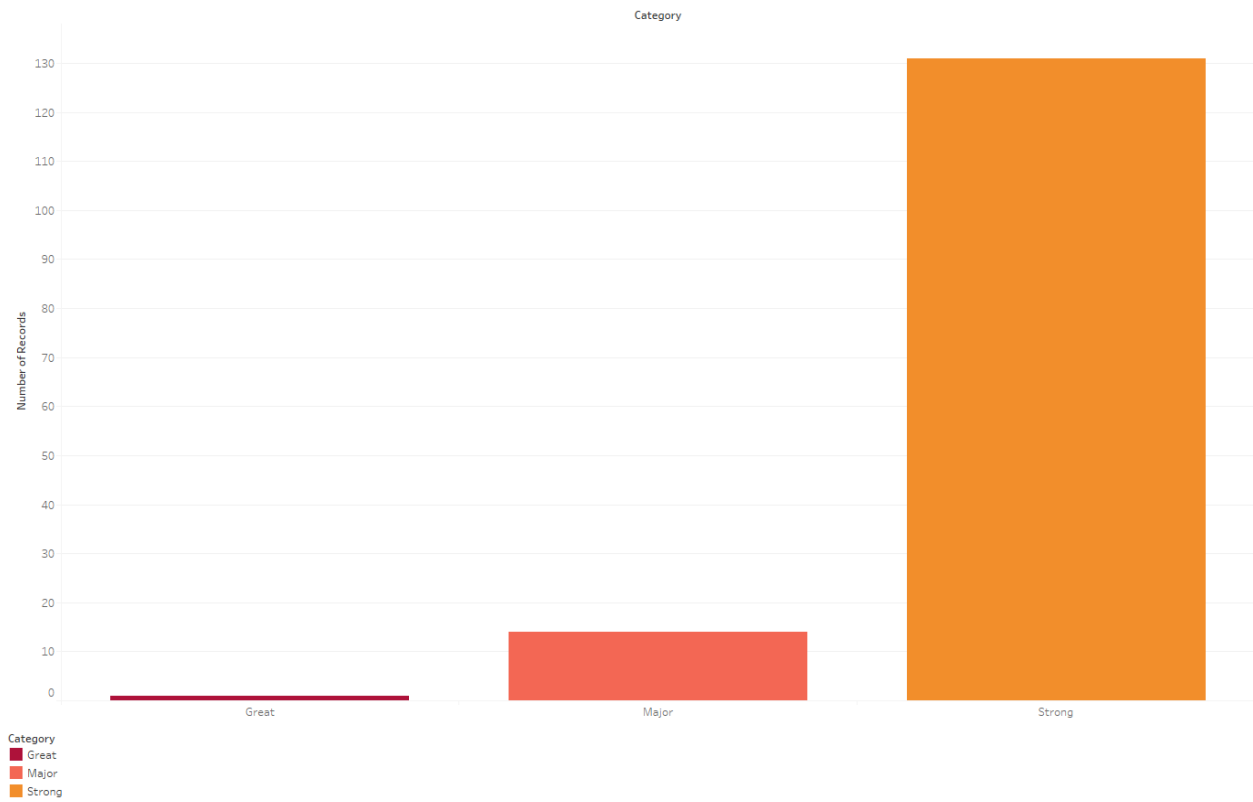
## Problem 1.4

Significance of Events by Magnitude



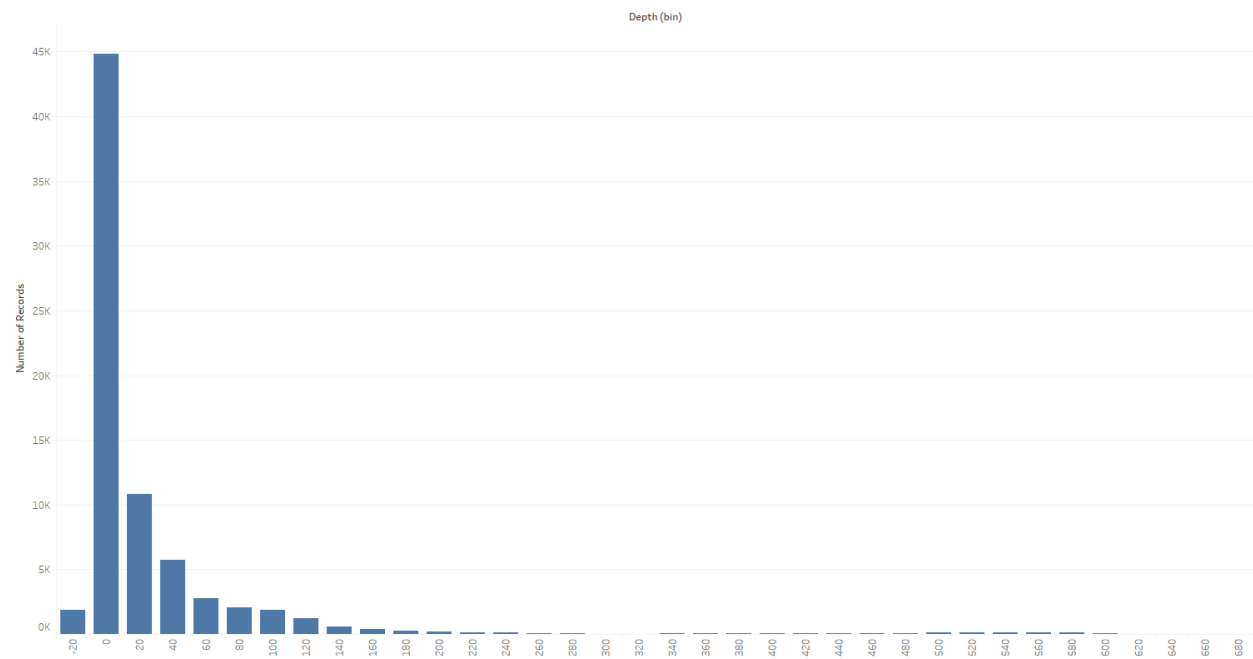
## Problem 1.5

Occurrences of Categorized Earthquakes



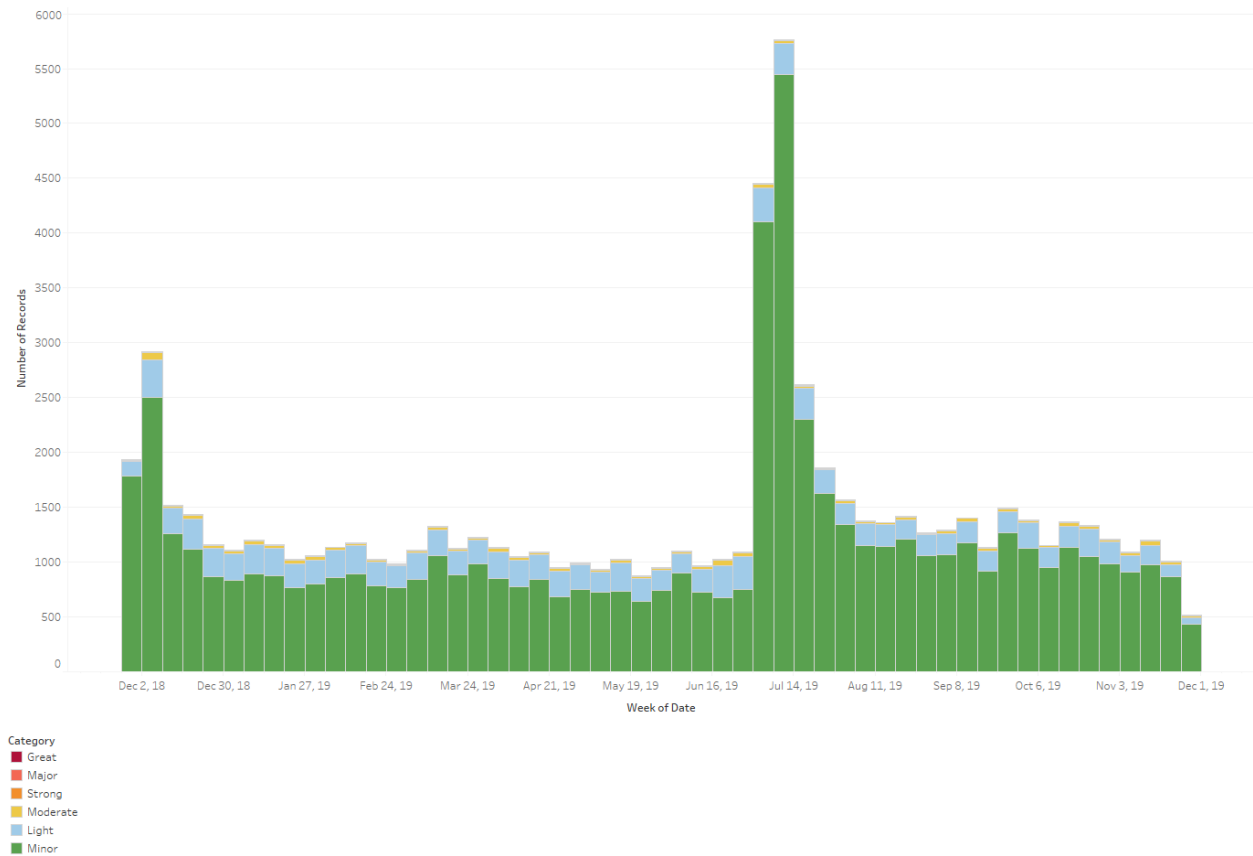
## Problem 1.6

Depth of Earthquakes (20KM Bin Size)



## Problem 1.7

Stacked Earthquake By Week



## Problem 2.1

There's color, which should be properly used according to your audience; this is meant for categorical data, and a scaled color is meant for numerical data (continuous).

There's size, which can have significance in changing the appearance of dramatic data, like in our bubble graph in the pages above, the significance of the event is based on its size.

There's labeling, which should only be used for a small amount of data.

There's the scaling of graphs, which should always start at either 0 or the lowest value if its negative; there's linear and logarithmic scaling.

There is also the type of graph and should be appropriate to the data. For example, the graph that you would use in a continuous data column is not the same graph that you would use for discrete data; A Pie Chart would be appropriate to mapping data that is discrete, whereas a Histogram would be useful for either one, and a line chart would be more suitable for continuous data.

## Problem 2.2

Qualitative (categorical) should normally only contain up to ten different colors, since it becomes a little too hard to understand otherwise.

Sequential color scales are based on a hue, and usually contain colors of only that hue. For example, I would be using approx. 10 different colors of blue, starting from dark and ending with light. There are some that contain multiple colors.

Diverging color scales are usually two sequential scales put together and meeting in the middle with color. For example, Red on one side, and blue on the other, with a mix of the colors in the middle that is also lightened as we get more toward the middle.

Affective color in a visualization should be appropriate to the audience that you are presenting data to. Is the audience mostly colorblind? if so, what type? The colors that you choose should also be relative, meaning that you should not have random colors that do not really make sense. You should not have more than 10 colors in a discrete color scale, and shading using a sequential scale should not have so many values that its nearly impossible to differentiate the colors.

## Problem 2.3

Logarithmic scales are appropriate when you have data values that vary greatly. For example, multiple data points are of size 5, and many are of size 100000; using a logarithmic scale in this case would give you a better sense of the data without skewing things heavily. Log Scales should be used for greater variability and many records. For example, if I had 100,000 records that had varying counts on the y-axis, I could use a log-Y axis to make the data more presentable.

## Problem 2.4

B is in the wrong position. It should be on the positive Y-axis; it is on the negative side.

C is in the wrong position. It should be between the  $1e+00$  and  $1e+01$  on the x-axis.

F is in the wrong position. It should be higher on the Y-axis and between  $1e+00$  and  $1e-01$  on the X-axis.

## Problem 2.5

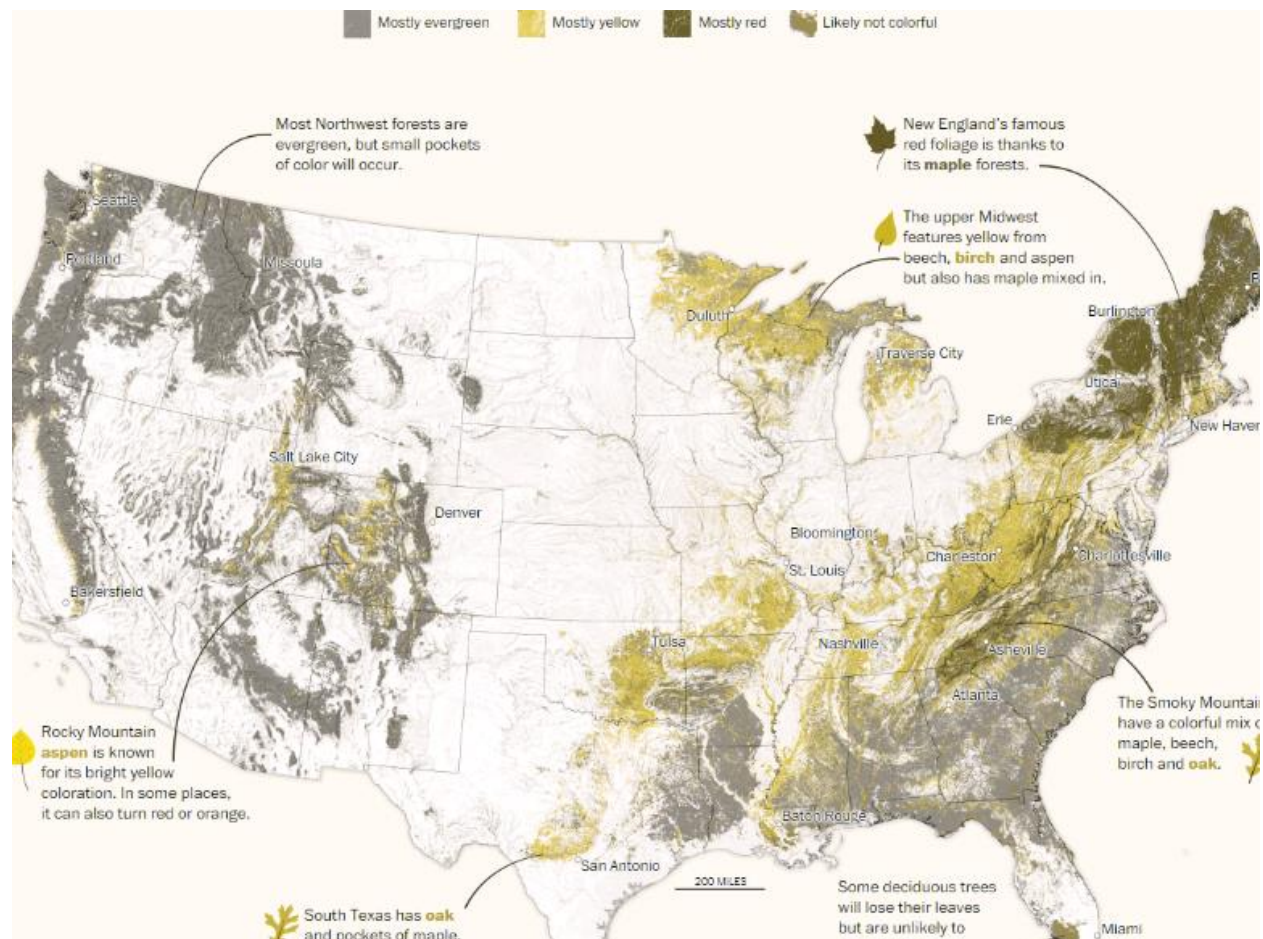
The problem with using a rainbow color scale is that it does not do well with patterns, it causes people to see false patterns. Changes are not uniform, skewing data value differences. This also causes grouping of similar colors, when in fact the color is not the same data value, so all shades of red are read as the same value. Rainbows are not good for colorblind people, because depending on the type, they cannot see certain colors, causing them to miss some of the information laid out for them.

## Problem 2.6

The graph has multiple things going on. Its colored by category, is on a linear scale both x and y, and has sizing based the value of the sports team. The x axis is the number of championships that the sports team attended, and the y-axis is the number of years that team has been competing. Interestingly, soccer/football (non-american football) is the most grossing sport of all time as well as the biggest sport in terms of the number of players in championships, which is unsurprising. I don't really truly

understand the titling of this graph since it seems like it has more to do with the teams than the players, but maybe its just a play on words.

### Problem 3.1



### Problem 3.2

Exploratory help us figure out what's important, whereas explanatory only shows us the important things. For example, we may see more data in an exploratory graph with colored important points, or differently sized important points, whereas in an explanatory graph, we would only see the important data

### Problem 3.3

The chapter states that there should be a balance in the amount of data that you present, as well as a balance with the level of visualization that takes place. For example, a not-so-great balance is the graph that's in Problem 2.6, where theres just a lot going on in one area. I personally don't think that the colors are good in that graph, but its okay. There's a lot of data on that graph. An example of a good balance would be to just take the sizing out and change the monetary cost to the x-axis in order to focus



more on the actual most valuable teams, which gets rid of the number of years that they've been competing which I think is irrelevant information for the most part.

### Problem 3.4

3D is very difficult to read. Gratuitous 3D is just the needless use of 3D graphs. There is no point to them because all of the data that someone is attempting to create 3D visualizations of can easily be done in two dimensions. An easy alternative to a 3D positional scale is to add sizing on points based on attributes or split the graphic into multiple different graphics of the same thing, with categorical faceting. Another way to accomplish this is to just categorize based on color and stack data.

### Problem 3.5

Poorly positioned on the Y-axis, as the scale does not start at 0, so there is no real sense of 'scale'. No units, poorly chosen time scale, the x-axis is based on years, but only contains October to April? Which is another problem in it of itself, is this meant to be October 2011 to April 2013 or is it meant to literally be only the months of October to April, and in that case, the years are completely irrelevant. Overall, just a terrible graph. They needlessly colored the graph with a gradient.

### Problem 3.6

First off, far too many different colors for categories (States). The biggest problem here is that the x-axis is the population in 2000, and the y-axis is a percentage growth from 2000 to 2010? I don't really understand the point for that but whatever. A better option would be to do the following:

List the states on the Y-axis, and then show the population growth as the size of the bubble. Keep the population size in 2000 on the x-axis.

Overall I think that theres just too much going on in this graph to really 'fix' it.

### Problem 3.7

Problem: Its 3D. Its also completely missing the time between 40 minutes and 45 minutes. The easiest way to fix this is to just create a pie chart with the same data. Theres few enough categories that the color scales would be good, and the graphic is trying to show a percentage of drivers with commute times, so it just makes sense.