CSC 465
Sarah Cummings

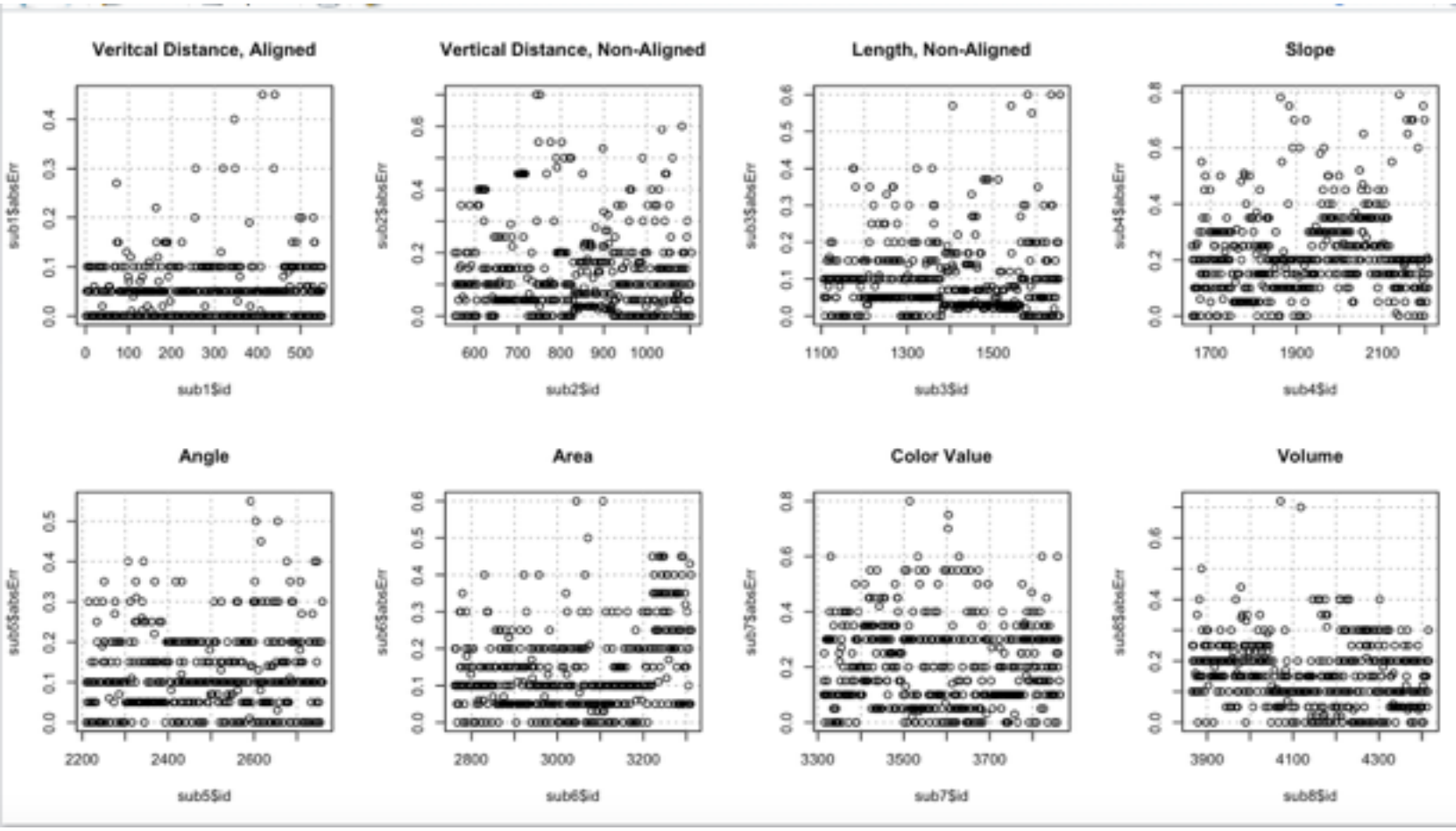
1) Perception test visualizations

a. See code.

```
data1$Error<-data1$Response-data1$TrueValue
data1$absErr<-abs(data1$Error)
```

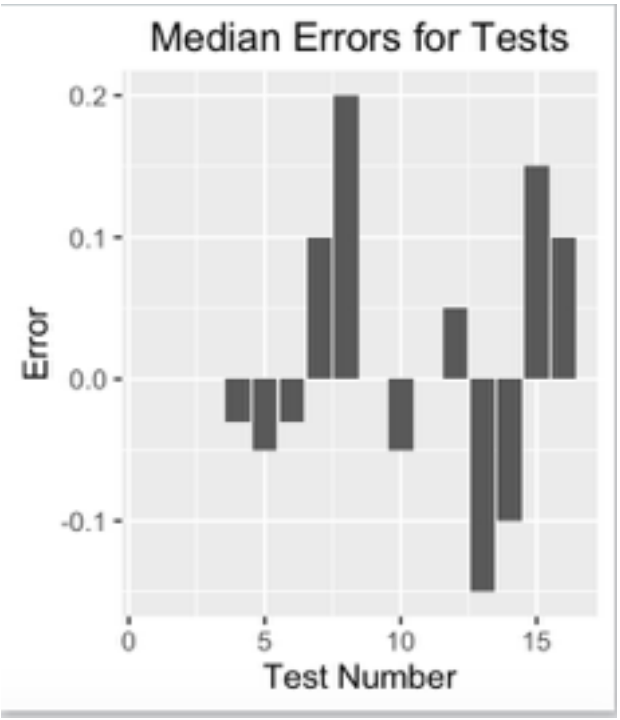
b. Distributions of absolute error for each test:

In the following graphs, titled by test type, the absolute error is plotting against the id number for the given observation.



Confirming what we knew from class, Vertical Distance aligned seems to have the lowest absolute error. It also has the the tightest shape of distribution, centered at the lower range in error with majority errors in the 0.0-0.1 range. Vertical Distance non-aligned and length have similar error distributions, both heavily clumped in the lower range. Slope, area, and angle are the tests with the most points in the upper range. Color Value appears to have the widest spread.

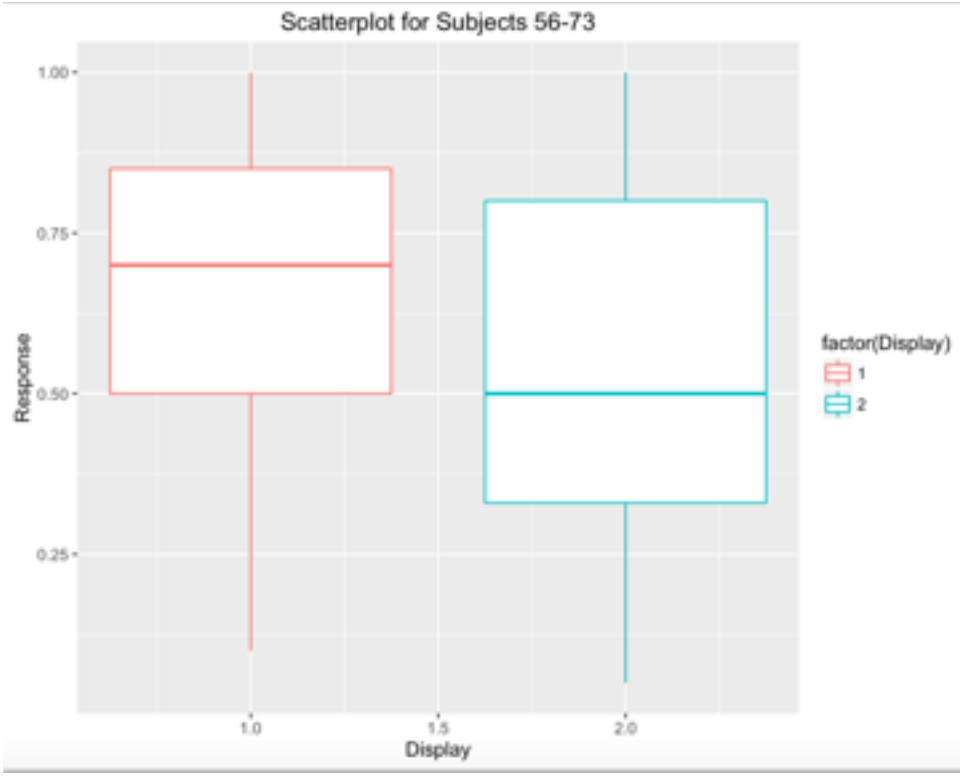
c. To see in which tests students tended to under or over estimate, we could check the median error for each test. Since we calculated then error as follows: data1$Error<-data1$Response-data1$TrueValue, we examine the median errors for the tests and whether or not those values are positive or negative will indicate the direction of over or under estimate. Positive implies an over estimate and negative is an underestimate. I've included a graph I used in the previous homework assignment, median errors for each test number.



Tests 15 and 16, which used volume as a measure, tended towards overestimation. Tests 14 and 13, which used color value, lead to underestimation. Test 7 and 8, which used slope, lead to overestimation.

d. Comparison of responses for Display 1 and Display 2 for subjects 56-73.

Subjects had lower response values on average for the second display. In the second display, more values are within the inner quartile range. Given the mean true value for this data set is 0.5910417, and display two values have a closer center to 0.5910417, we can also conclude the subjects were better with display 2.

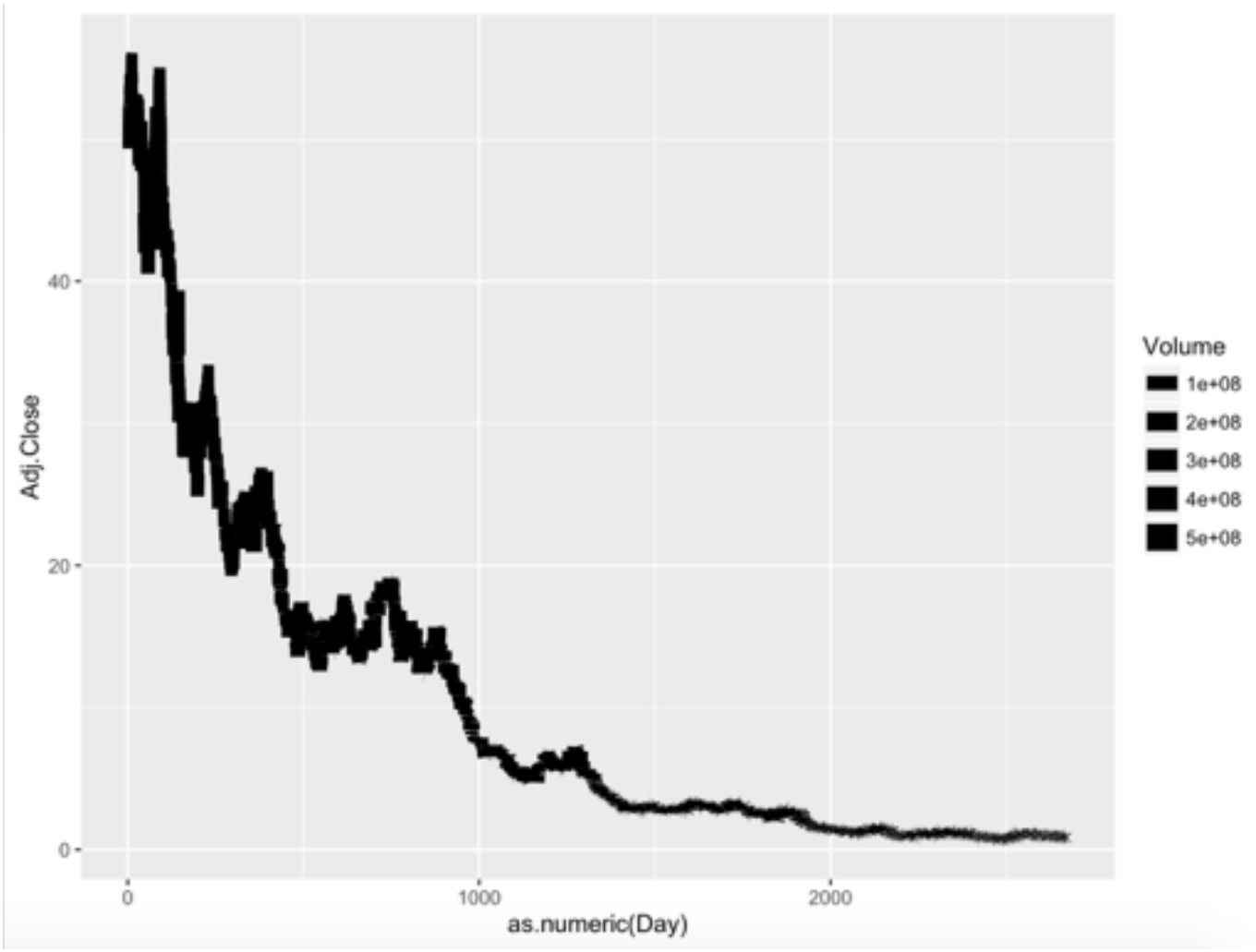


e. Subjects 56-64 responded 1 in Vertical Distance unaligned display one.

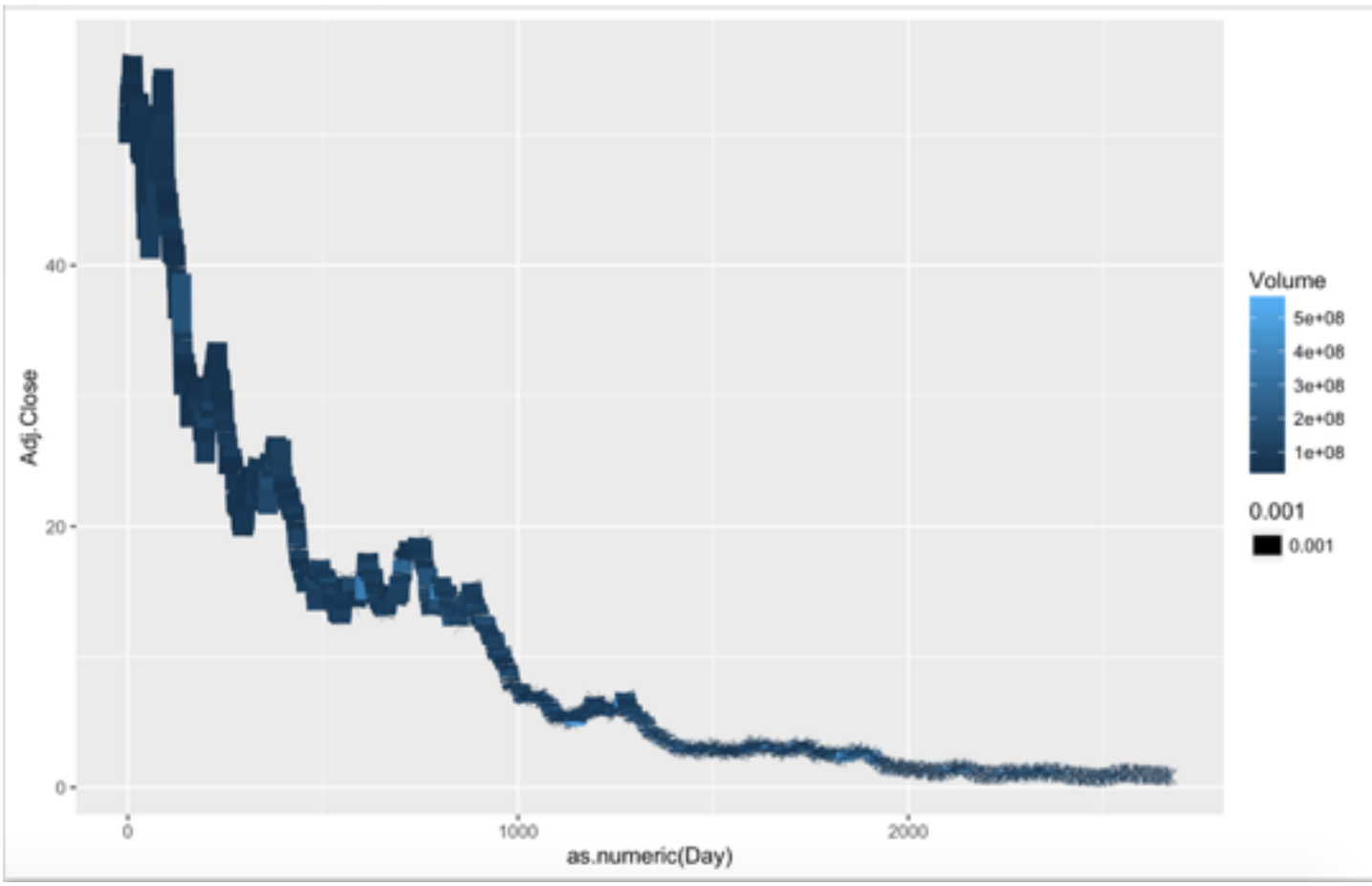
I created the following scatterplot, above, which highlights the anomaly. Note that for these subjects, the points are on top of each other for the different trials, so they are colored according to trial D.

2) Intel Stock Data

a.

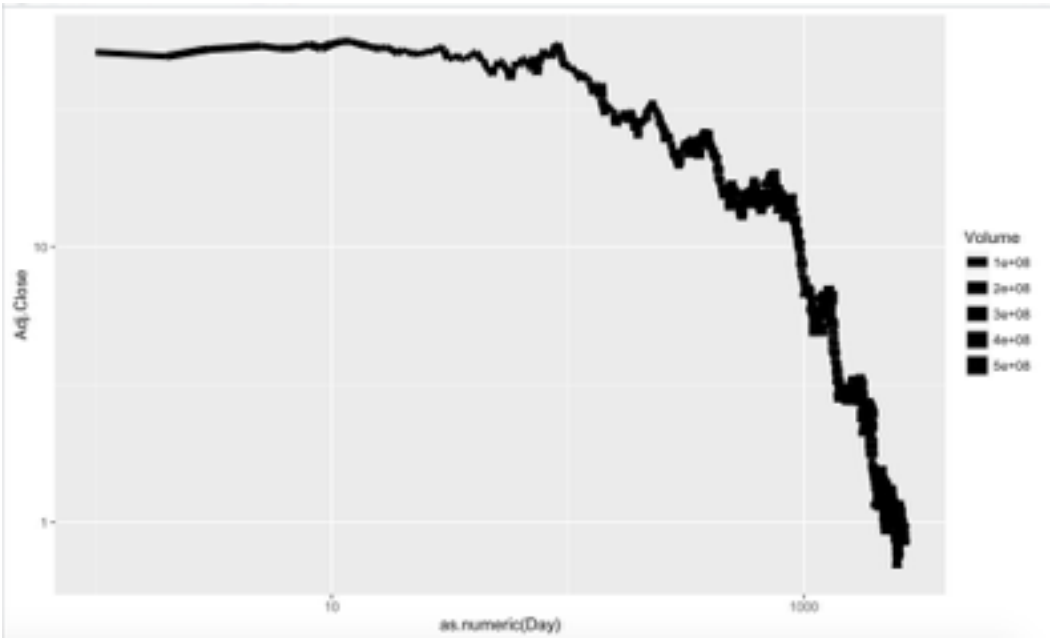


b.



Width of data is generally easier to perceive than color scale, and I believe that to be the case with this data. The range of volume makes color scale way too hard to distinguish.

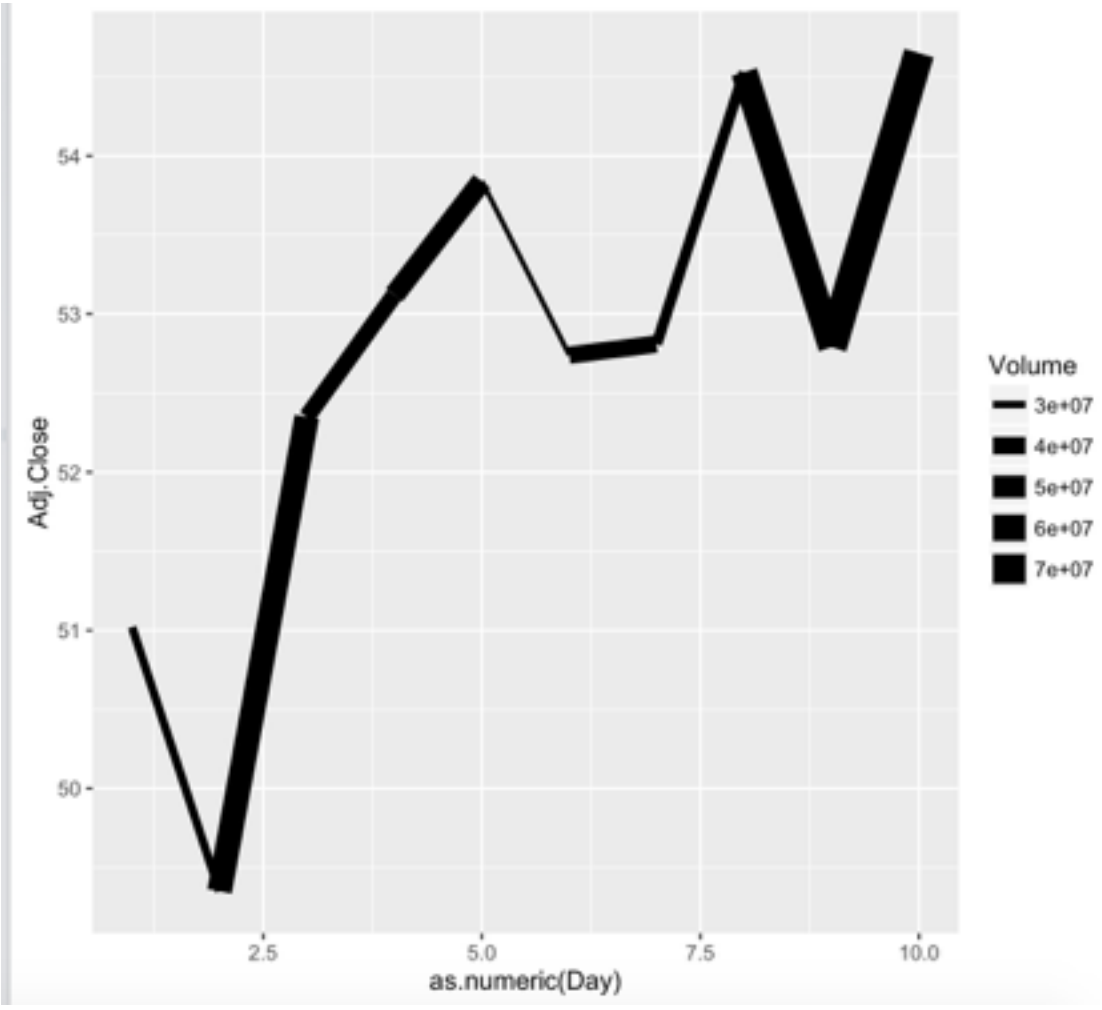
c. Graph with logarithmic scale



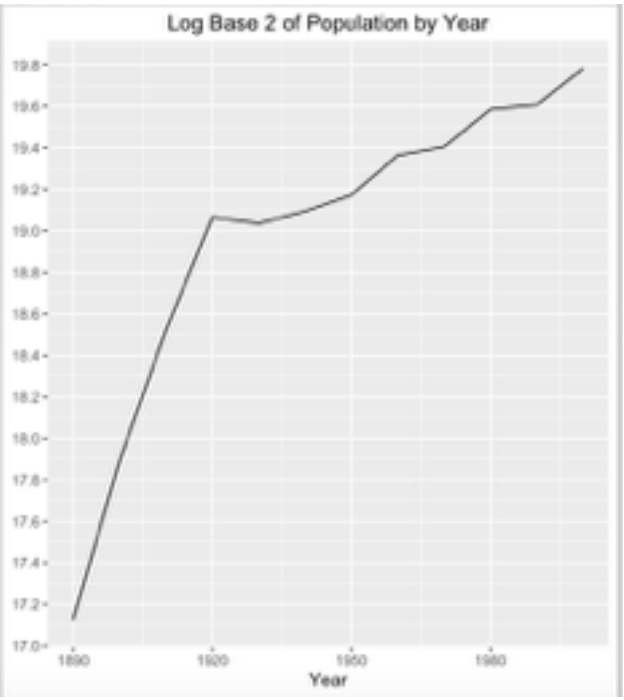
With the logarithmic scale, the shape of this graph is a downward sloping curve. This means the Adj close rate has dropped significantly and suddenly.

d. I have no idea what this question is asking me to do. Data presented as years? What?

e. Yes, reading for thickness is better with fewer points as seen below:



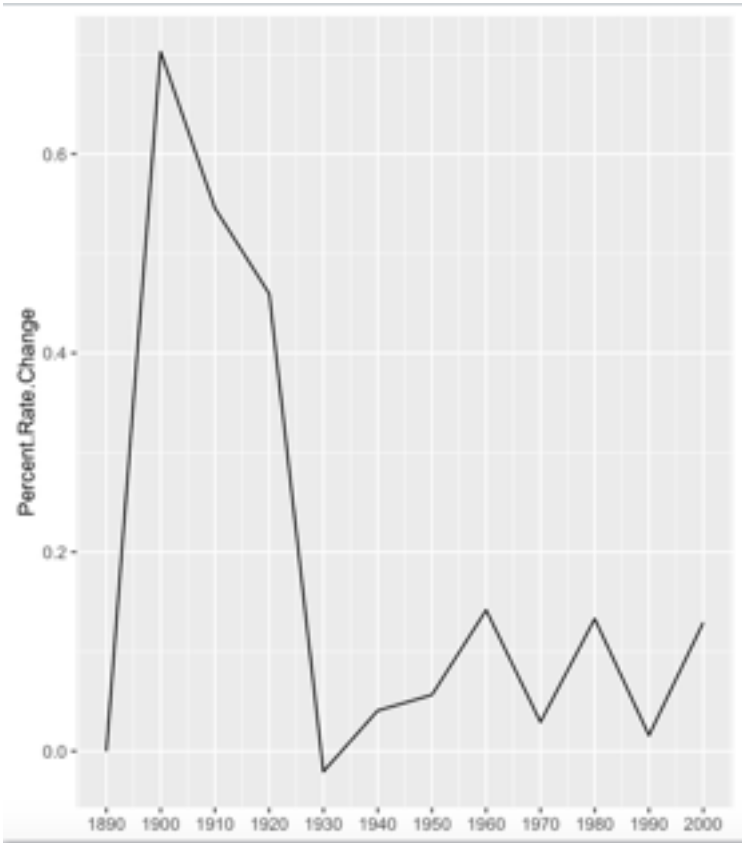
3) Montana Population Data



a) Left. Showing how many times the population has doubled, used log_2(population) on y axis

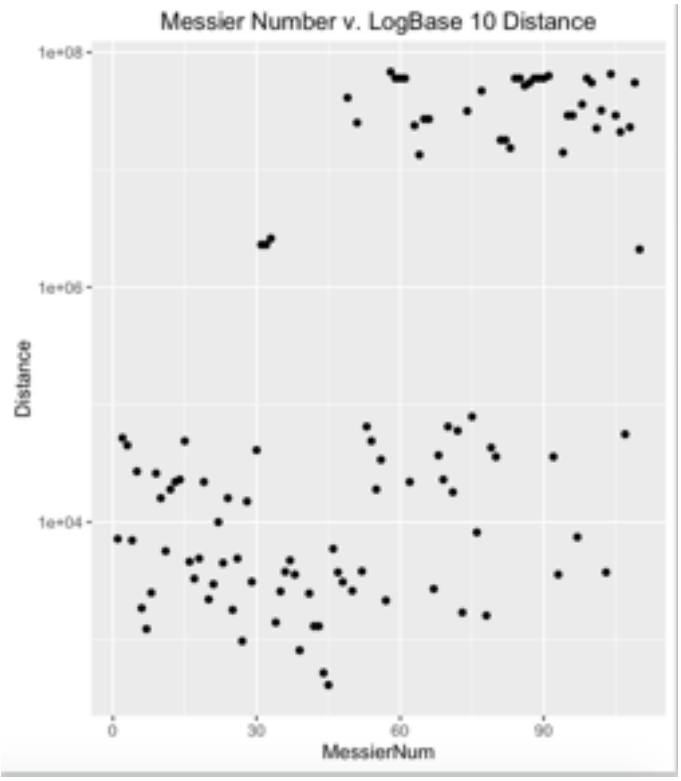
Answer: approximately 19.7 times.

b) Right.Showing percent Rate change on y axis, year on x axis. Population percent rate of change has increased and decreased over the years, seen in the slope of the graph. It had the biggest increase percentage wise in years 1900, 1910, and 1920.



c) Population percent increase was greater than 15% in year 1900, 1910, 1920, also seen in graph for b.

4) Meisser project

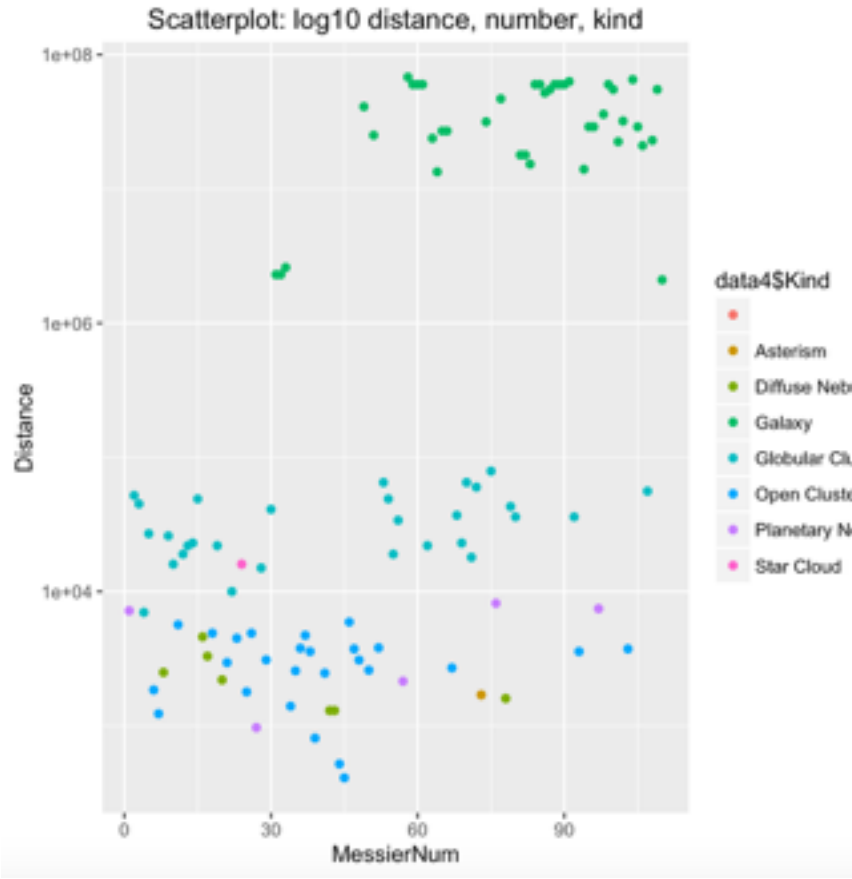


a. Graph one or more properties of the objects against the Messier number

Looking at distance on a log base ten scale, it appears that there are at lease two distinct clusters of points for this data. The pattern likely has nothing to do with the ordering of the Messier number, but it is interesting nonetheless.

It also appears as though all of the "far distance" cluster points are in the Messier number range from 45 on. Perhaps they are listed in the order in which they were discovered, at which point it would make sense for Messier to getting better at spotting these objects over time.

b. Visualization that compares the distributions of the distances to the objects in each "kind"

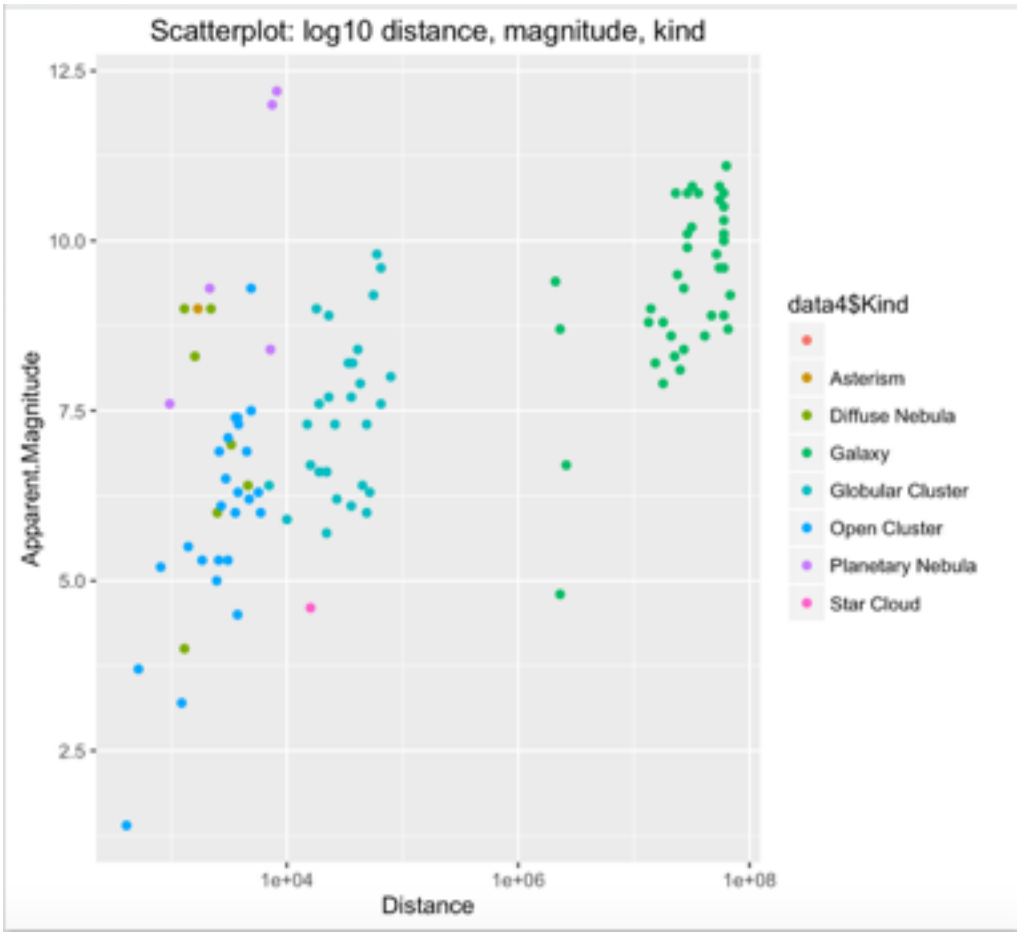


Three main clusters described only by kind: Galaxy, Globular Cluster, and Open cluster

The galaxy kind has the furthest distancefor its center and a fairly narrow spread.

The globular cluster kind has the second furthest distance as its center, and widest spread

The open cluster kind has the shortest distance at its center, and the narrowest spread of the three.

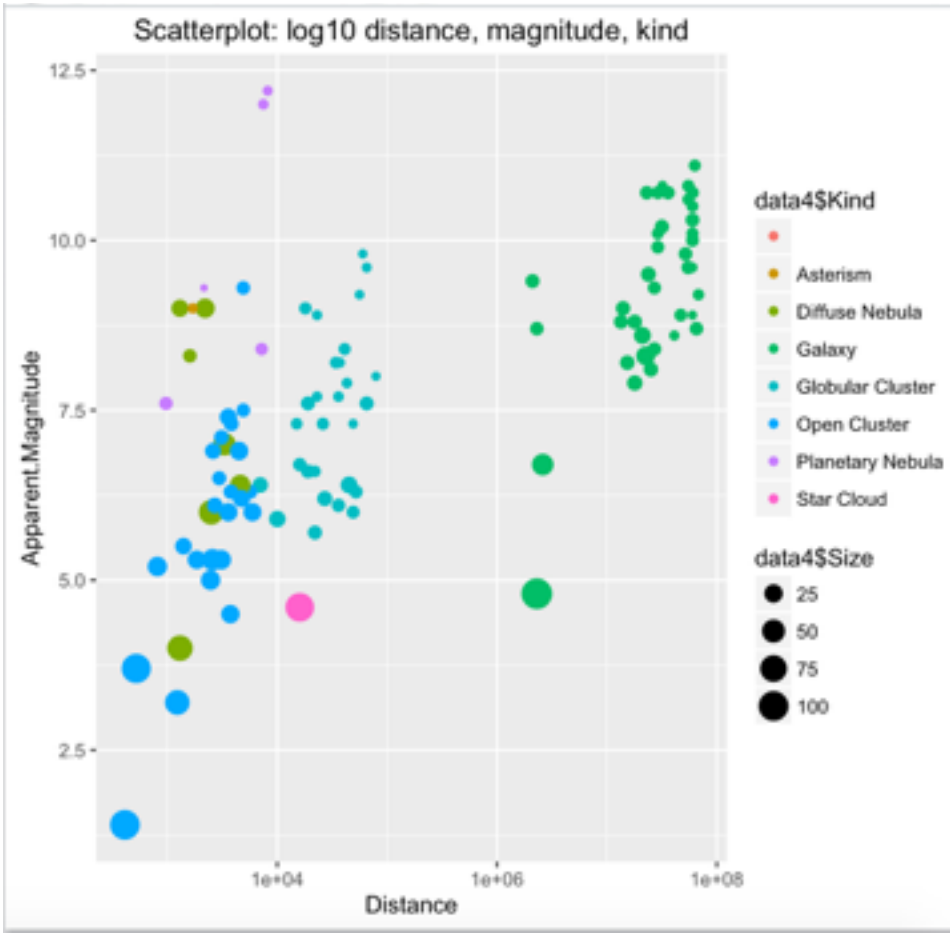


c. Scatter plot with the distance to the Messier objects plotted against their apparent magnitude

I'm seeing a positive correlation between apparent magnitude and distance, which is no what the homework implied we should find…

d. Augment the visualization in c) by adjusting the size of the points in the scatter-plot based on the angular size of the objects in the sky.

I think its pretty easy to see all four varaibles of this graph, but given the overlap in the clusters, I would definitely stretch the window of the graph so the smaller points can be seen better.



5) Portland Water Objects