DSC 465
Homework 3
Daniel O’Brien

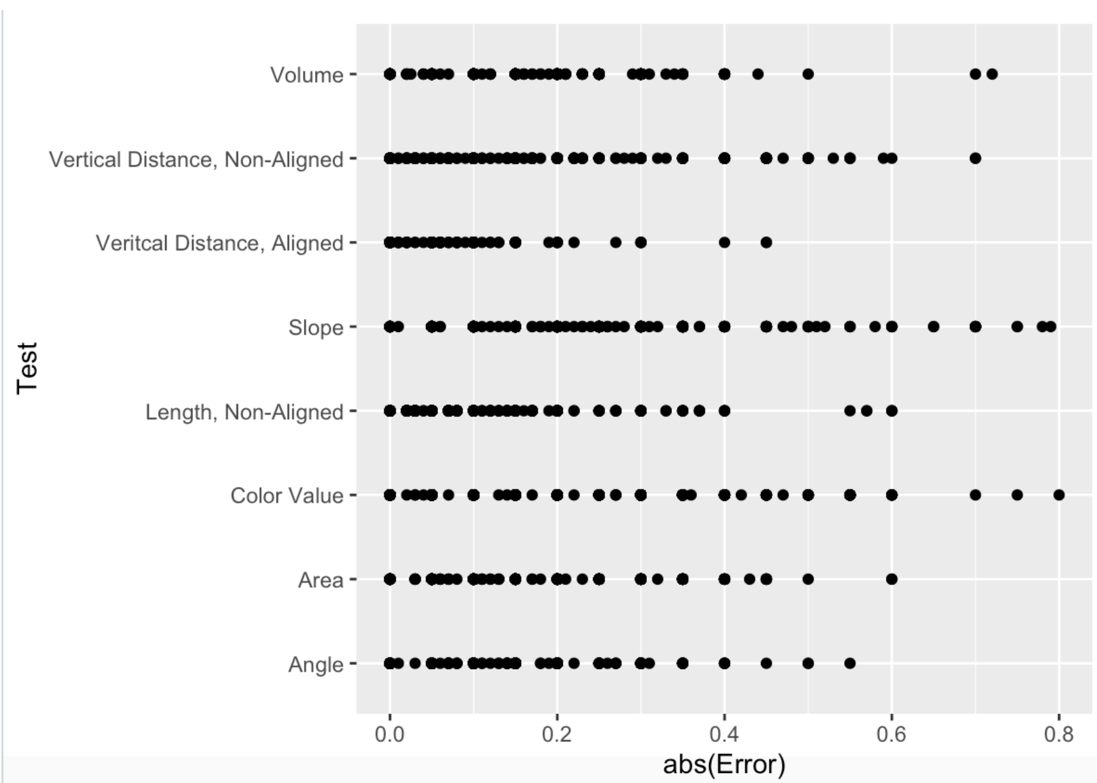
1.
a. In the color value test, people tended to underestimate the values more than the other tests. I found this by created an Error column consisting of the difference between the True Value and Response, and then summarized the columns grouped by Test, Display and Test Number. The largest average underestimation takes place in the Color Value Test, display 1 with an average error value of -0.168. This is followed closely by the other color value test, which has an error value of -0.131. This indicated that overall people tend to underestimate the data in the Color Value tests, more than the other tests. The largest overestimation on average takes place with display 2 of the slop tests, with an average error of 0.148. Following the slope test, both volume tests also show large error values of 0.127 and 0.116. This shows that people tend to overestimate the slope and volume tests. All of the above can be seen in the table below.

| | Test | Display | TestNumber | average |
|---|---|---|---|---|
| | <chr> | <int> | <int> | <dbl> |
| 1 | Angle | 1 | 9 | -0.00409 |
| 2 | Angle | 2 | 10 | -0.0479 |
| 3 | Area | 1 | 11 | 0.0108 |
| 4 | Area | 2 | 12 | 0.0884 |
| 5 | Color Value | 1 | 13 | -0.168 |
| 6 | Color Value | 2 | 14 | -0.131 |
| 7 | Length, Non-Aligned | 1 | 5 | -0.0656 |
| 8 | Length, Non-Aligned | 2 | 6 | -0.0295 |
| 9 | Slope | 1 | 7 | 0.0737 |
| 10 | Slope | 2 | 8 | 0.148 |
| 11 | Veritcal Distance, Aligned | 1 | 1 | -0.0219 |
| 12 | Veritcal Distance, Aligned | 2 | 2 | -0.00446 |
| 13 | Vertical Distance, Non-Aligned | 1 | 3 | 0.00946 |
| 14 | Vertical Distance, Non-Aligned | 2 | 4 | -0.0227 |
| 15 | Volume | 1 | 15 | 0.127 |
| 16 | Volume | 2 | 16 | 0.116 |

```
PerceptionExperiment <- read.csv("~/Desktop/PerceptionExperiment.csv")
head(PerceptionExperiment)

library(tidyverse)
```

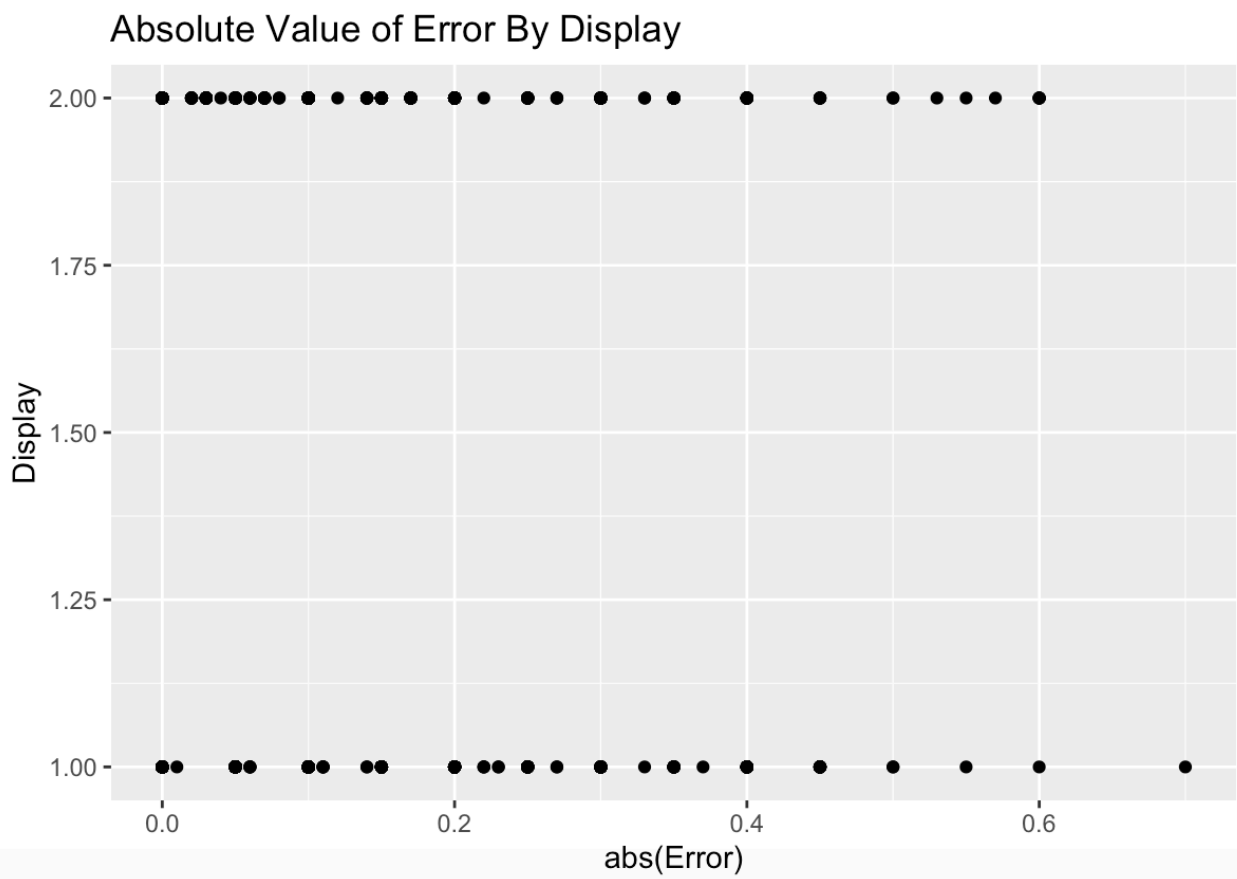
```
test = PerceptionExperiment
head(test)

test1 <-test %>%
  mutate(Error = Response - TrueValue)

head(test1)

summaryTest <-test1 %>%
  group_by(Test, Display, TestNumber) %>%
  summarise(average = mean(Error))

summaryTest
```

b. Below I created a univariate plot by test. The x-axis shows the absolute values of Error, while the y axis shows the different tests. Although this graph lacks pizazz, I think it clearly shows some tests had a higher error rate than others. The areas of clumping can be seen in almost all of the plots, indicating where the majority of people submitted their responses. Some tests have clumping areas at relatively low error rates like vertical distance aligned, and length nonaligned. And other tests show clumping areas at higher error rates, like slope and volume. We can also see that color value had the highest number of large error rates, although responses are more spread out for color value, and there is less clumping.



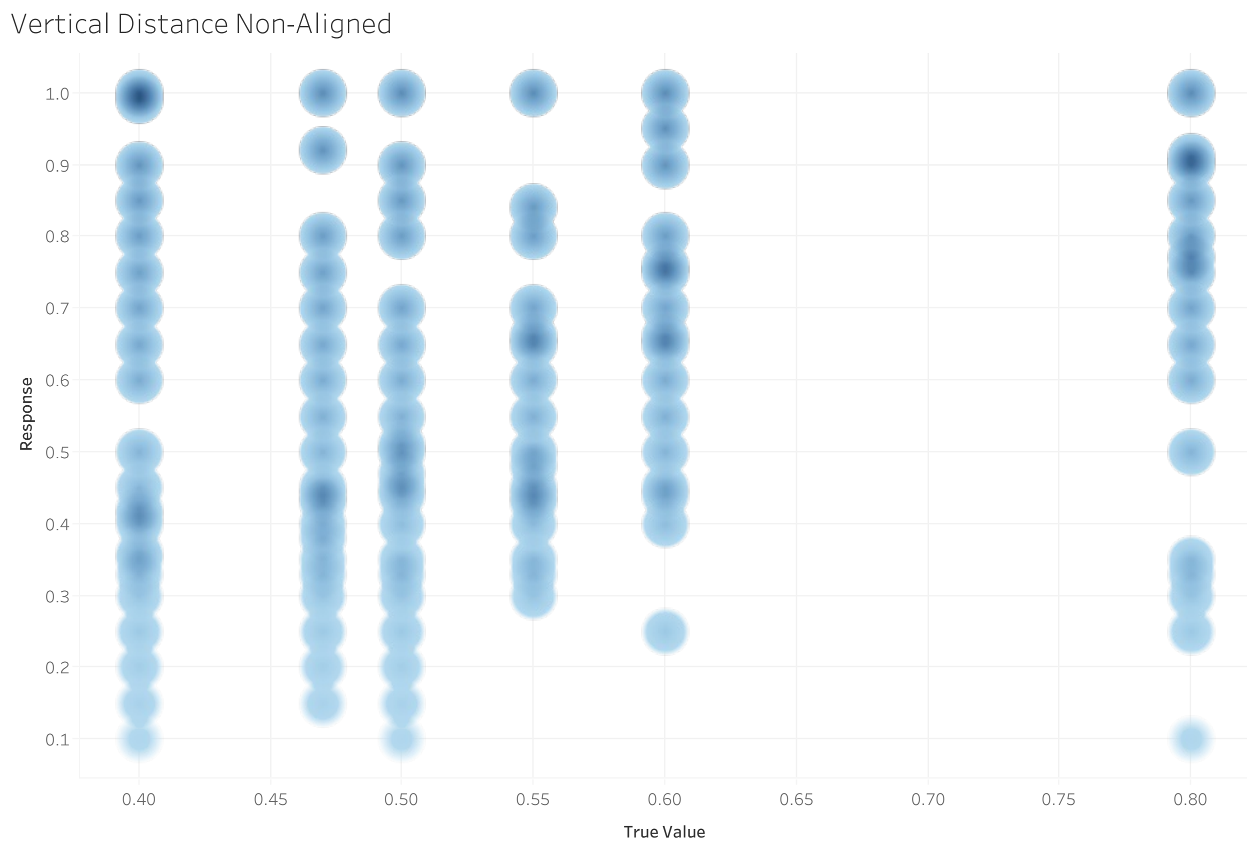
```
ggplot(test1, aes(x = abs(Error), y = Test)) +
  geom_point()
```

c. Below we can see that there is a significant cluster close to zero for Display 2. This would indicate that after participants observed the first display their perception increased when examining the second display.




```
test2 <- test1 %>%
  filter(Subject > 55 & Subject < 74)

head(test2)

ggplot(test2, aes(x = abs(Error), y = Display)) +
  geom_point() + labs(title = paste("Absolute Value of Error By Display"))
```

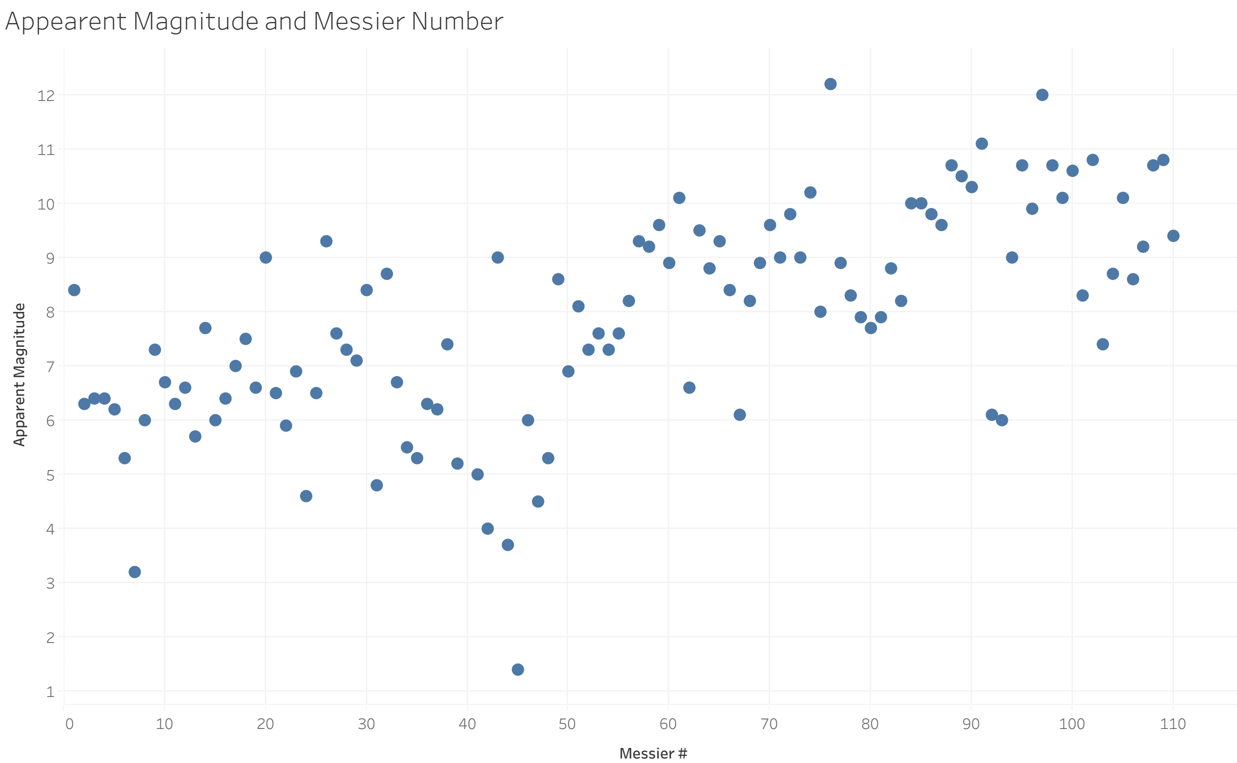
d. For this representation, I compared the Responses by the True Value for the Vertical Distance Non-Aligned test. The responses are also shown by density, so we can see that there are more responses near the true values, but there are also a higher concentration of responses at 1 for these tests as well.




True Value vs. Response. Color shows average of Response. The data is filtered on Test, which keeps Vertical Distance, Non-Aligned.

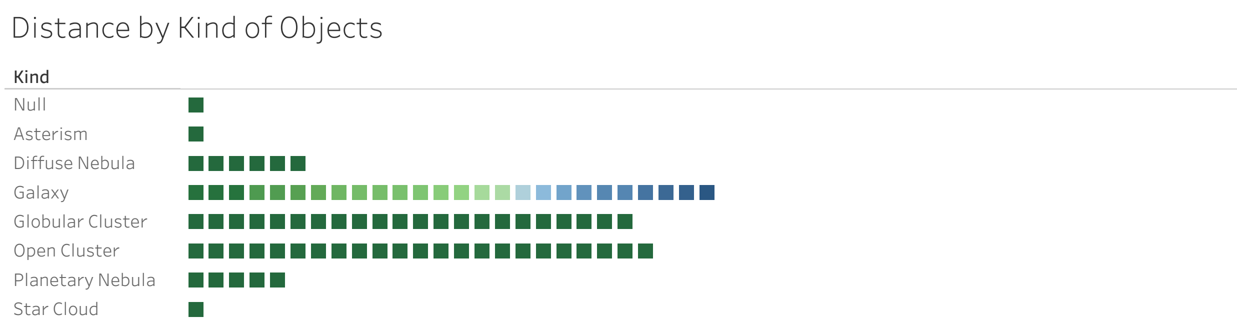
2.

a. I tested all of the attributes and the closest I came to finding a pattern was when I compared apparent magnitude and the Messier Number. It appears to show a slight positive correlation. Most of the other attributes did not show much of a pattern when compared to the Messier number.

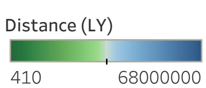



Messier # vs. Apparent Magnitude.

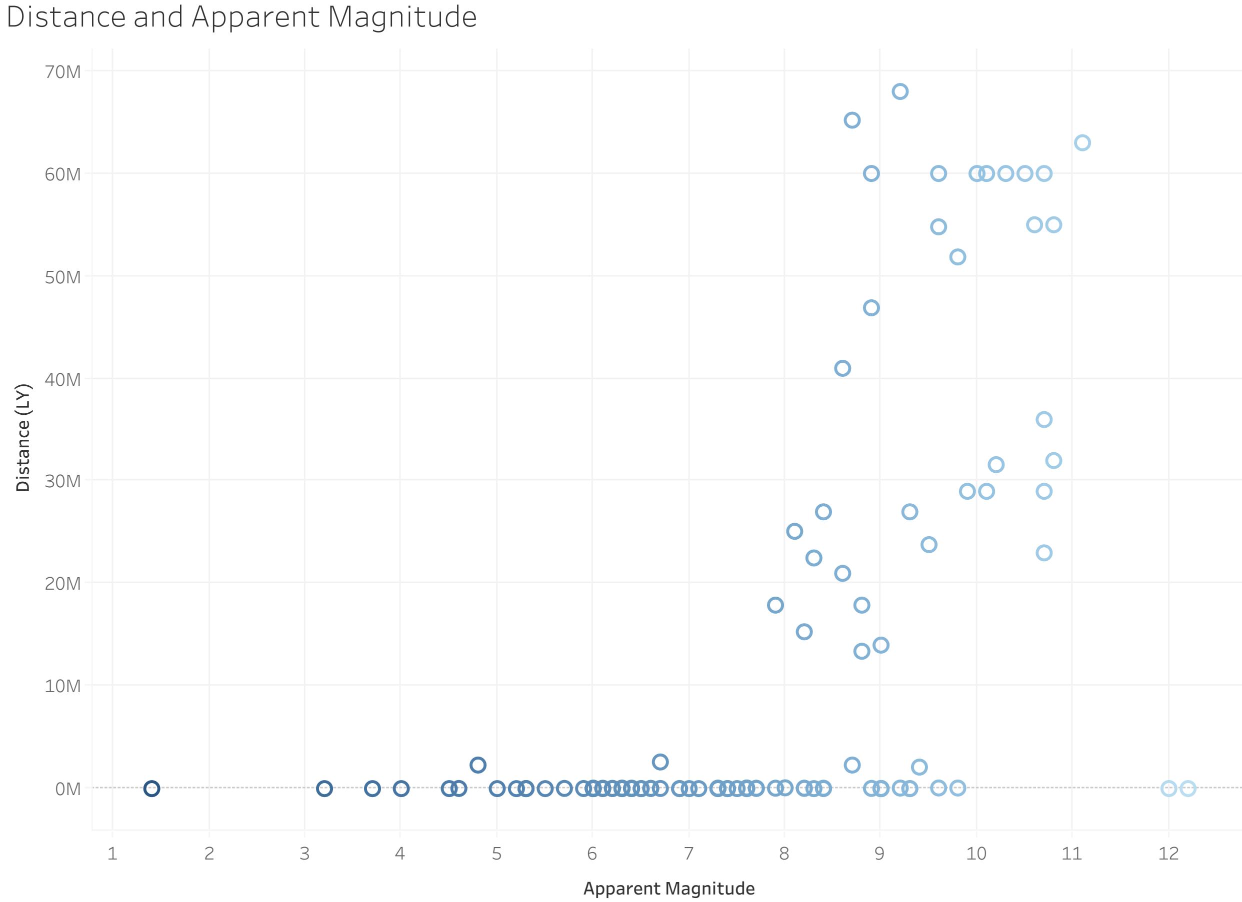
b.




Distance (LY) (color) broken down by Kind.




c. For this visualization, I reversed the sequential coloring to demonstrate that the higher magnitudes would be less visible by giving them a light color, when lower magnitudes are given a darker more visible color to demonstrate that they would be easier to see.

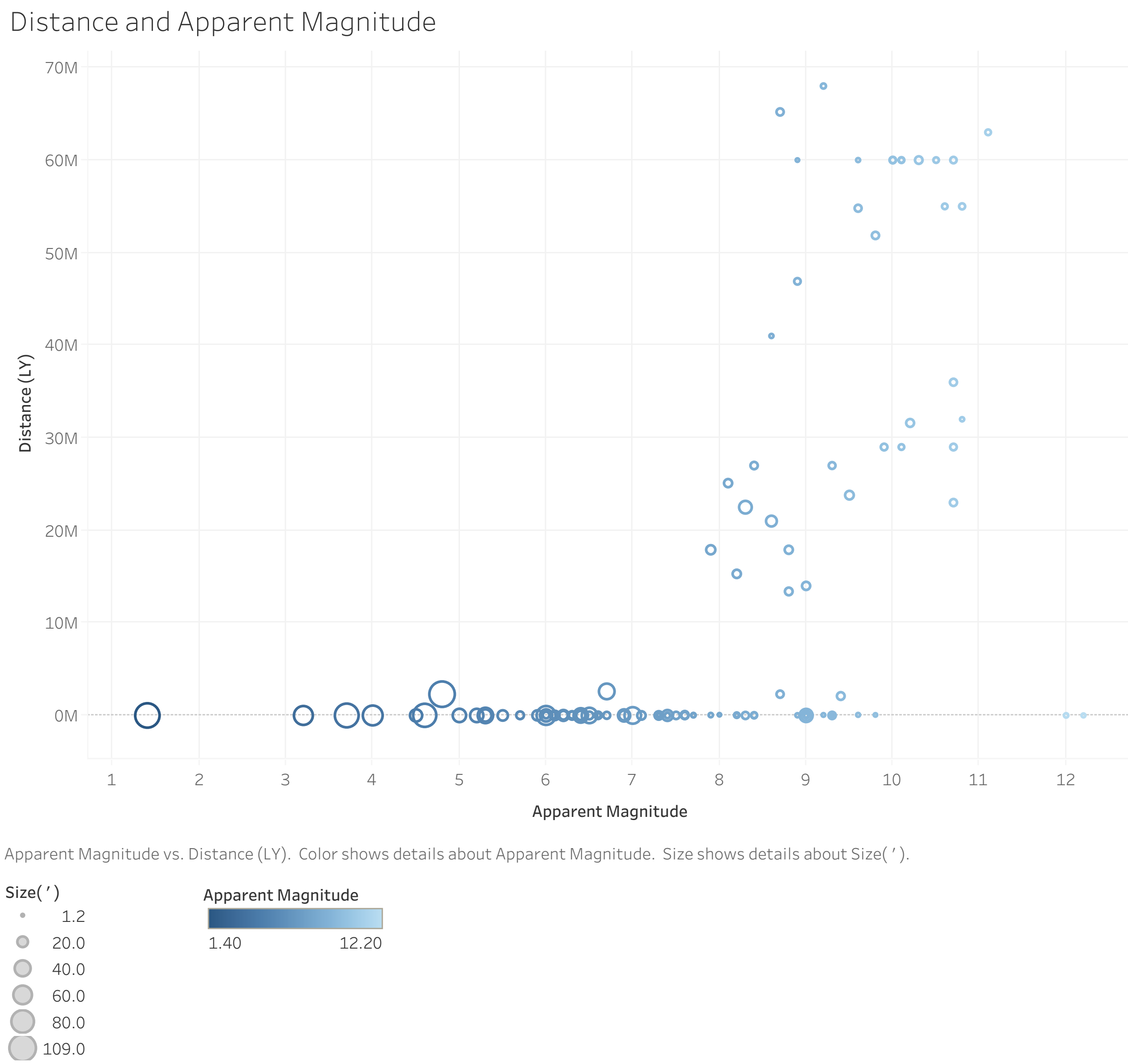



Apparent Magnitude vs. Distance (LY). Color shows details about Apparent Magnitude.

Apparent Magnitude

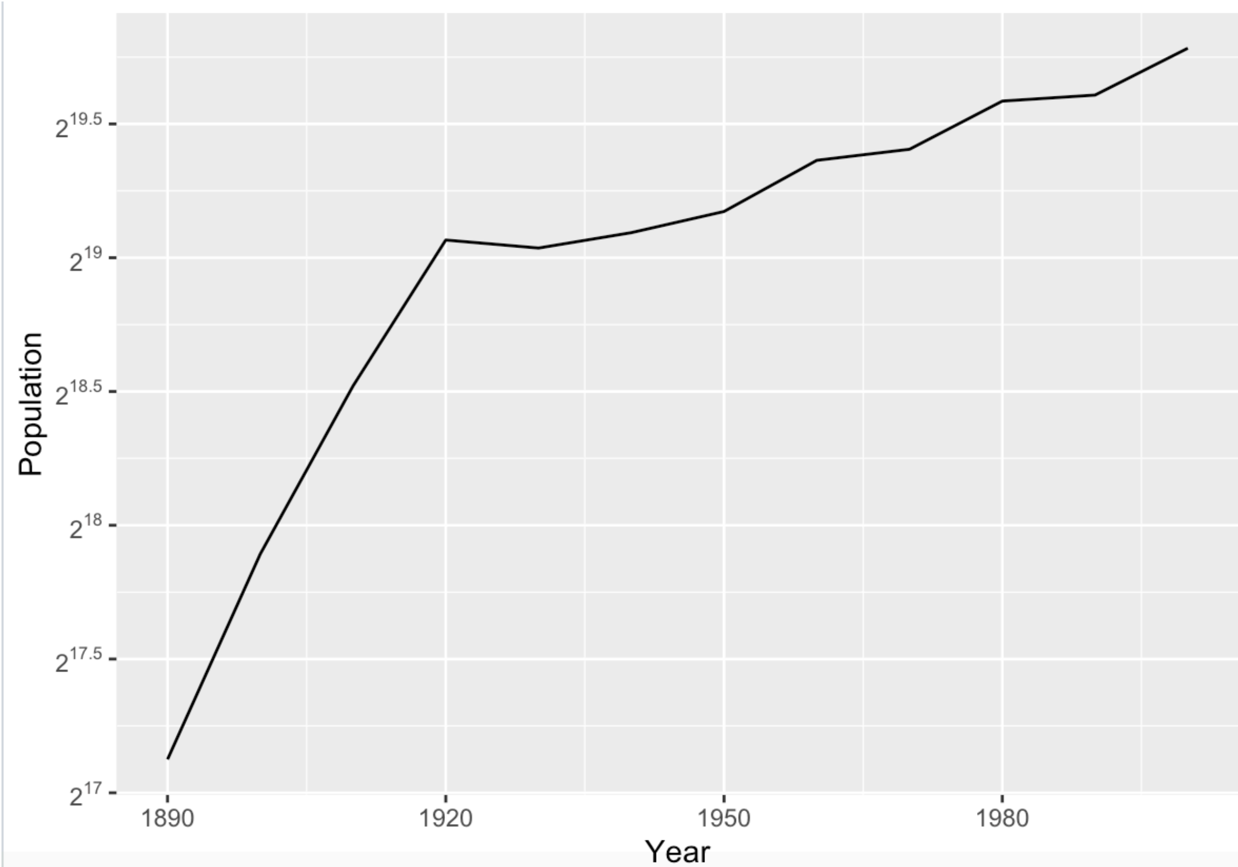
1.40 12.20

d. By augmenting the size of the points by the angular size of the objects, some of the information can be interpreted more easily. We can see that the objects that are further away, and of a smaller size, would have a higher apparent magnitude, meaning they would be harder to see, because they are a lighter color and a smaller size. The portion of the graph that is more difficult to encode is the bottom of the graph showing the more visible objects. When the size of the object is larger, and the apparent magnitude is smaller, the points are bigger and darker making them blend together and difficult to individually encode. One way to combat the concentration of points on the x-axis would be to change the scale of the x-axis and spread the points out a bit.



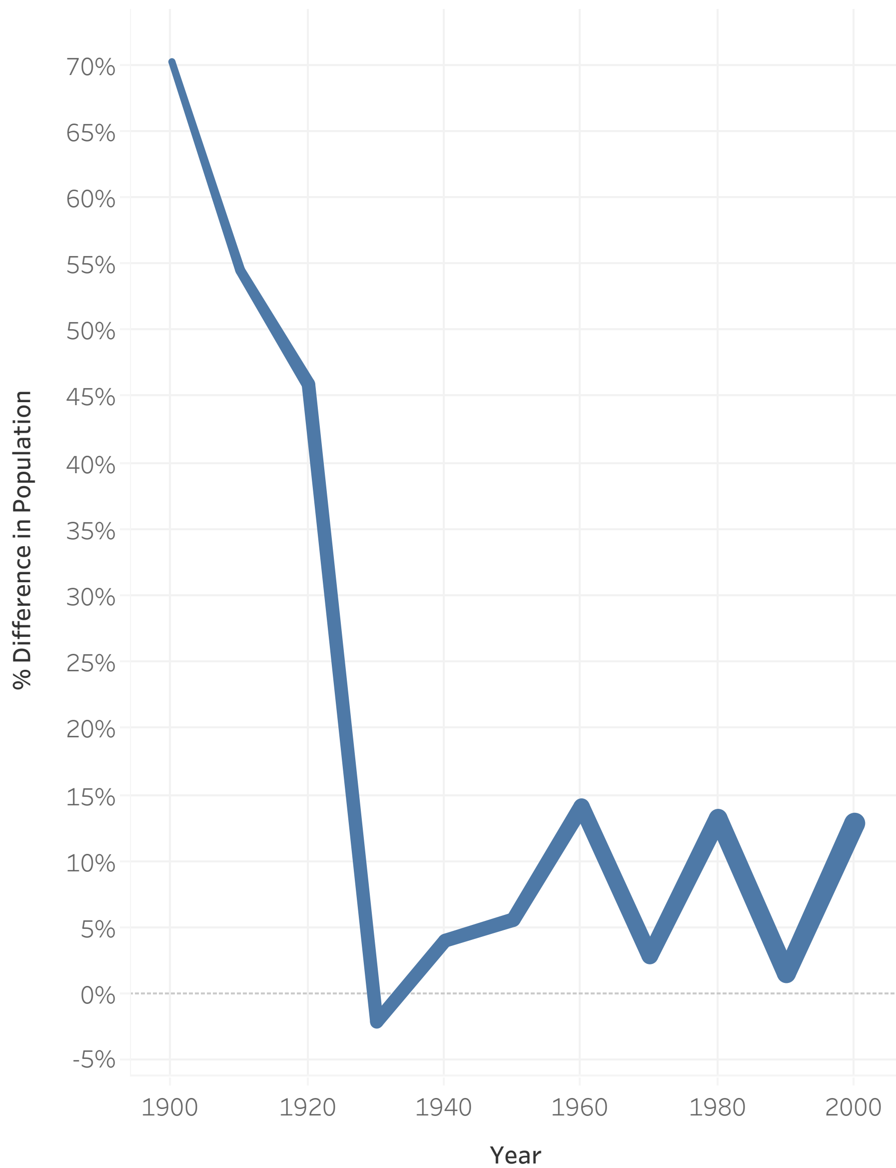

3.

a. Using a log base 2 scale we can track how many times the population doubled by seeing how many times the log increases. Since we start at log base 2 of 17, and reach log based 2 of 19, we can tell that the population doubled completely two times.




b and c. Below is a chart that shows the population change by year and the total population is indicated by the thickness of the line. We can see that the greatest percent increase occurred between 1890 and 1900, and there was only one year in the 1920s where the percentage change was negative. The y-axis, which represents the percentage change, is numbers by increments of 5% so it can clearly be seen which years the population growth was less than 15%.

## Percent Change in Population By Year



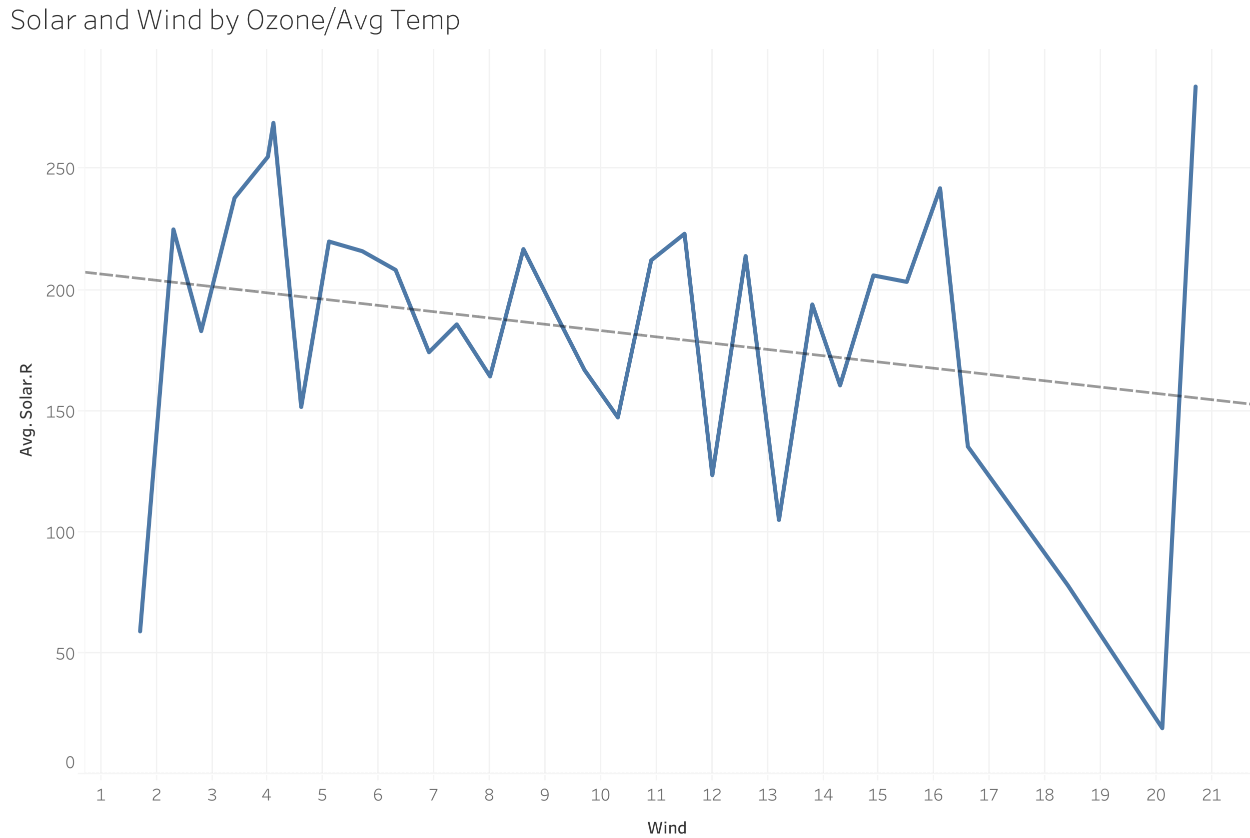

The trend of % Difference in Population for Year. Size shows sum of Population.

**Population**

- 142,924
- 400,000
- 600,000
- 800,000
- 902,195

4.

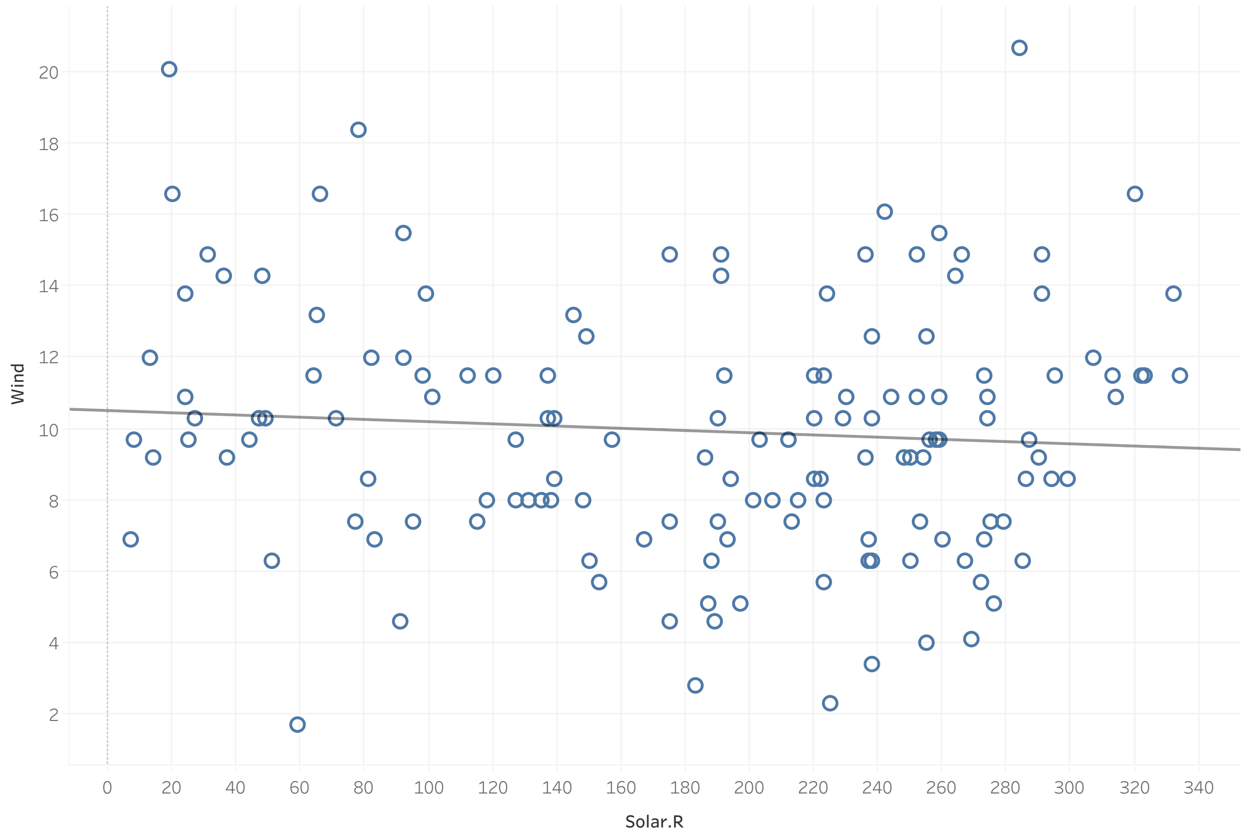
a.\




The trend of average of Solar.R for Wind.

b.

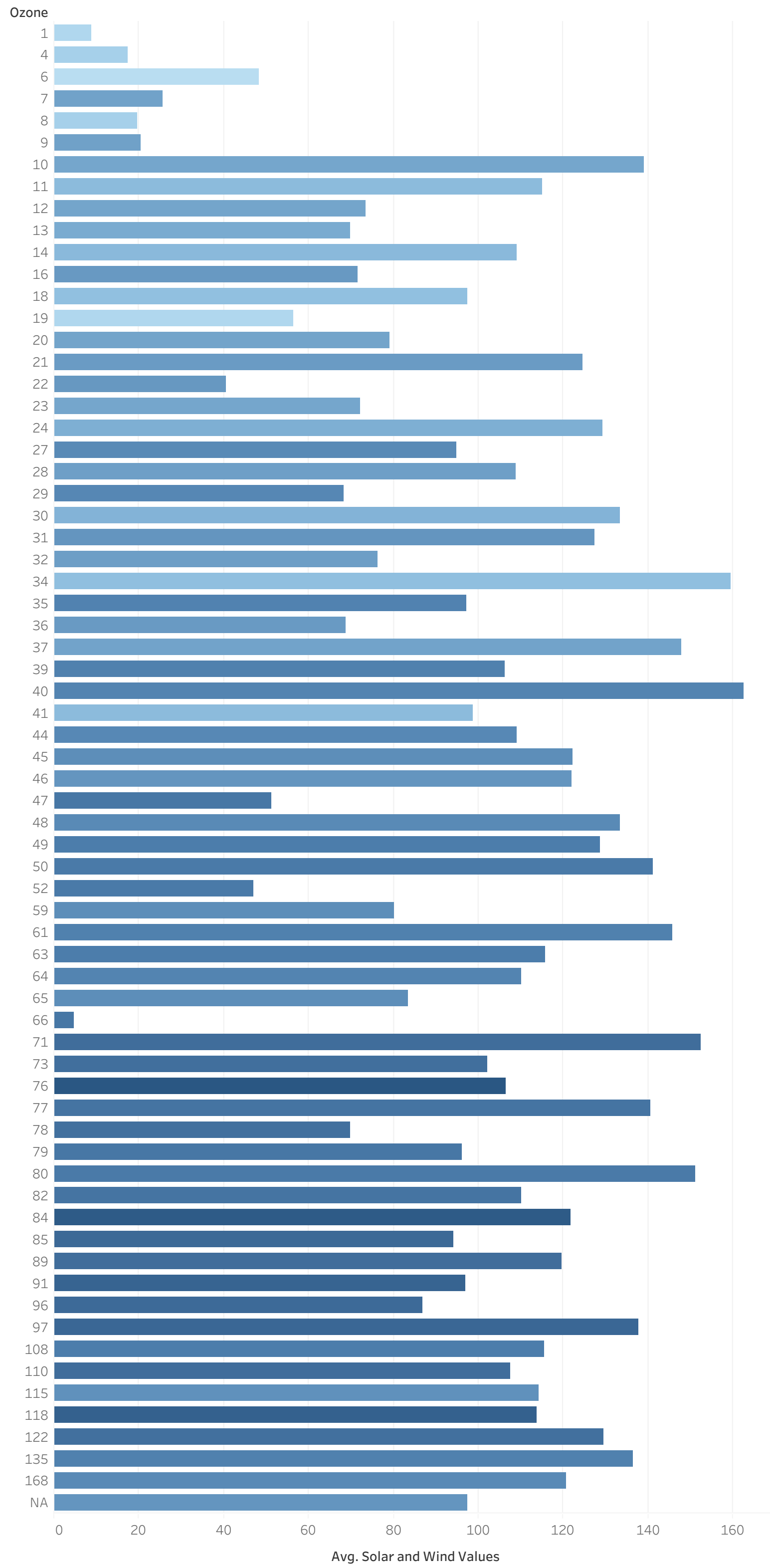
Wind by Solar Radiation




Solar.R vs. Wind. The view is filtered on Solar.R, which keeps non-Null values only.

c. For this visualization, I created a pivot table by combining the Solar Radiation and Wind variables, and then compared the pivot table to ozone and temperature. The average measurement for solar radiation and wind can be seen on the x-axis, and the y-axis is made up of the ozone values. Additionally, the average temperature is noted by the color of the bars. This shows us that with very low levels of ozone, the average wind and solar radiation is also lower, however it begins to vary once the ozone levels begin to increase. The temperature seems to correlate for with the ozone value than the solar radiation and wind values.

## Solar and Wind by Ozone/Avg Temp




Average of Solar and Wind Values for each Ozone. Color shows average of Temp.

Avg. Temp

57.00 97.00