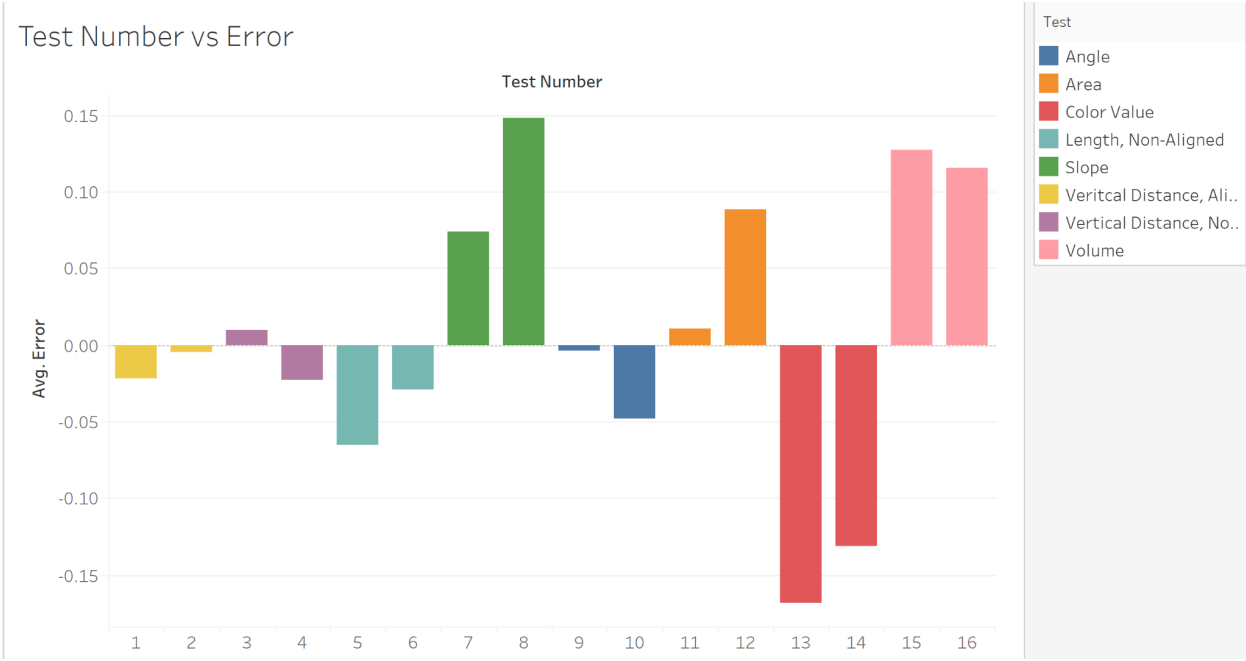
Elijah Caluya
DSC 465
5/15/2021

Assignment 3

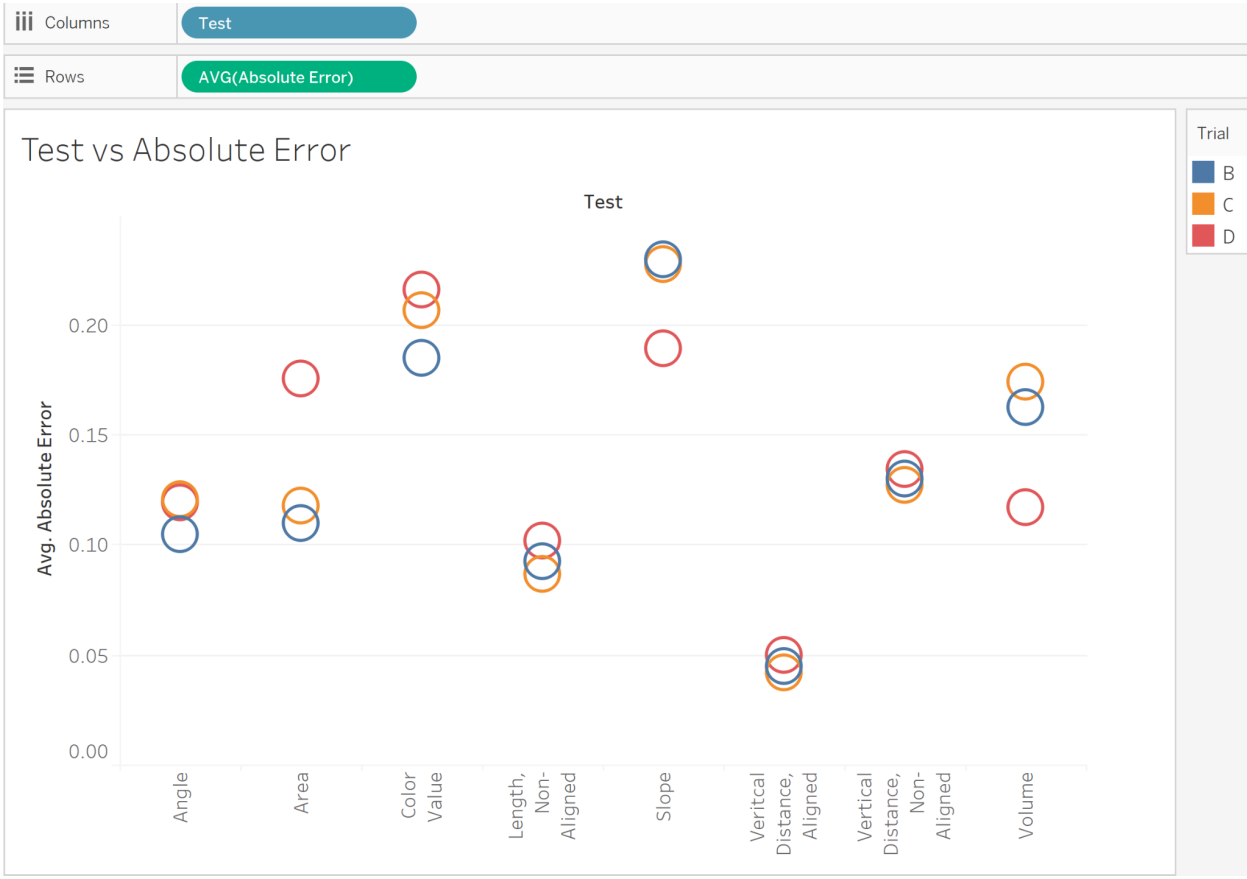
1) Problems that use PerceptionExperiment.csv

   a) Based on the graph below, it seems that the tests for “Color Value” were notably underestimated and the tests for “Slope” and “Volume” were mainly overestimated. It is also important to note that the tests “Length, Non-Aligned” were all underestimated but not to the degree as “Color Value” and the same can be said for “Area” but for overestimation. What I used to measure this is the “Error” calculated field where a negative score would mean an underestimate and a positive score would mean an overestimate. When the error is close to 0, then the response was very accurate to the true value. I decided to use a bar graph where someone can see which tests, or more specifically test numbers, show a greater underestimate or overestimate than the other tests. This way, we can clearly see which tests overestimate and underestimate by how much their bar is below the 0 line.



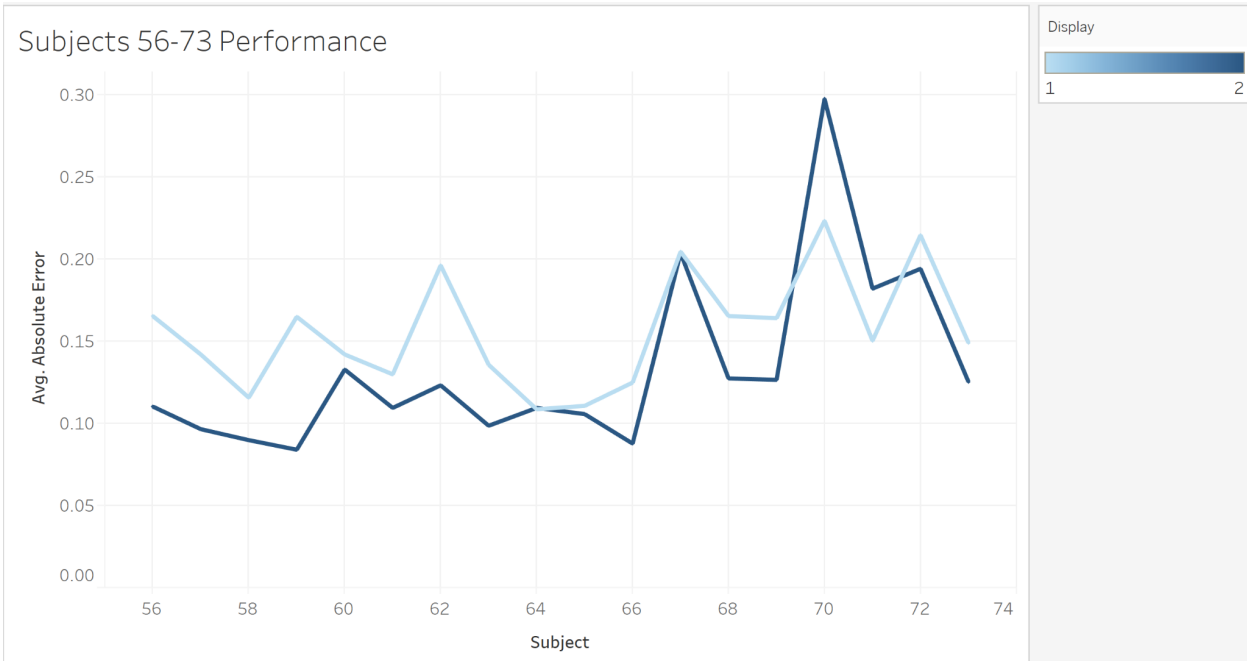
   b) Based on the graph below, we can see that “Color Value” and “Slope” had a higher absolute error value than the other tests while “Vertical Distance, Aligned” had the lowest absolute error value amongst all of the tests. It is also important to note that the majority of the tests had an absolute error that was in between 0.10 and 0.15. In terms of the perception test from which this data is based off of, we can see that “Color Value” and “Slope” are the two main factors in incorrectly perceiving the image. There seems to be some clumping of responses for both

“Vertical Distance” tests as the absolute errors for those responses are very close together.



c) The graph below shows the performance of subjects 56-73. We can see that they mostly performed better on the second display because there is a lower Absolute error value for that display. The only case where the performance did not get better were for subjects 70 and 71. Besides these two subjects, we can assume that the subjects got better at judging after seeing the first display because the absolute error for the second display is lower than the absolute error for the first

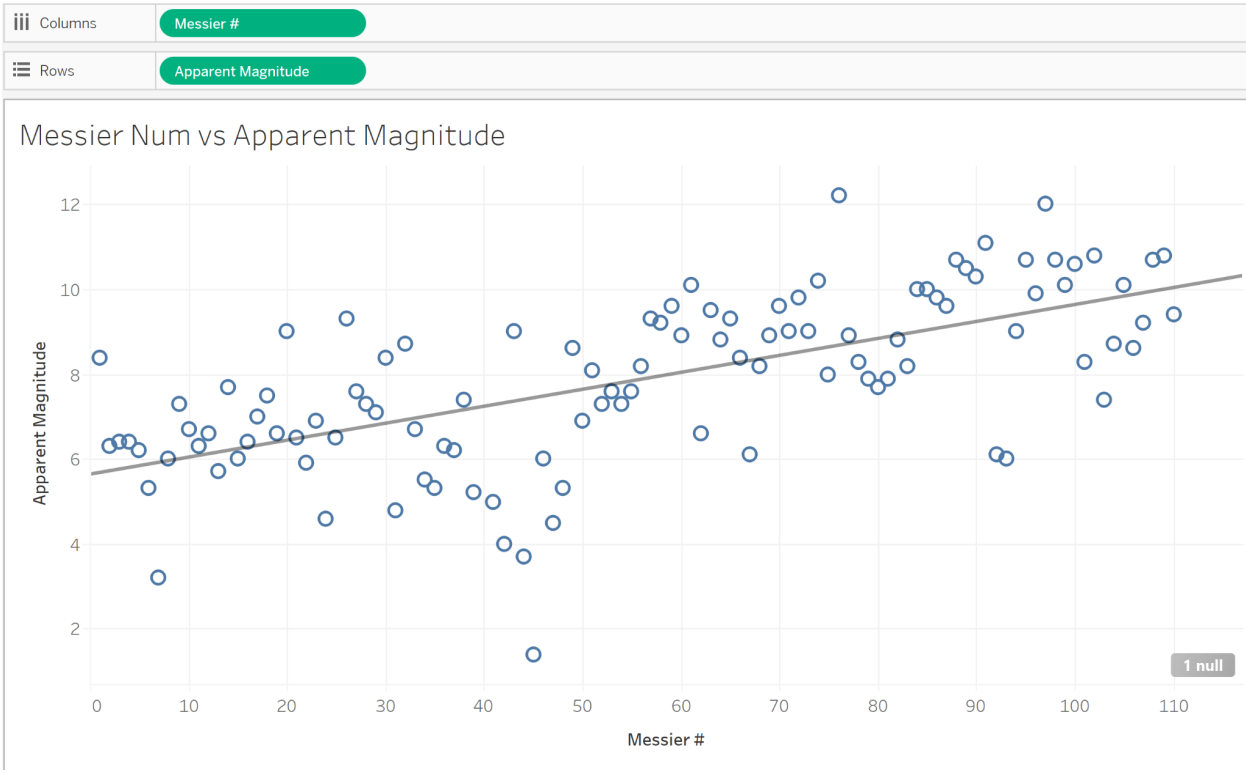
display.



d) In the graph below, we can see the continuous range of “1” responses for Trial B, C, and D. Most notably, the numbers of these subjects are 54-73. We can see that the points highlighted are the only ones that are on the 1 line. I wanted to make some sort of box and whisker plot that could potentially highlight these outliers better but I couldn’t find a good way to do so.

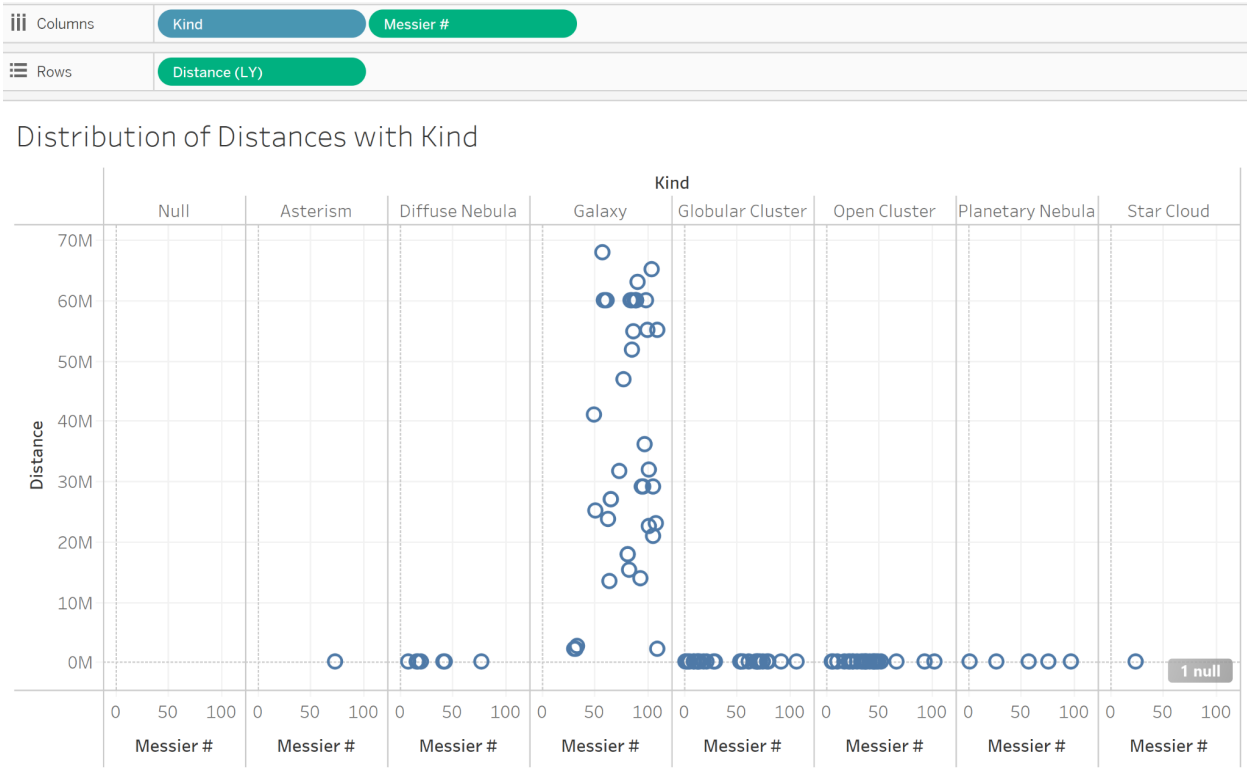
2) Problems using MessierData.csv dataset

a) With the graph below we can see that there is a slight upward trend between the Messier number and the apparent Magnitude. As you can see from the graph, as the Messier number increases, the Apparent Magnitude will also increase, thus showing a positive relationship.

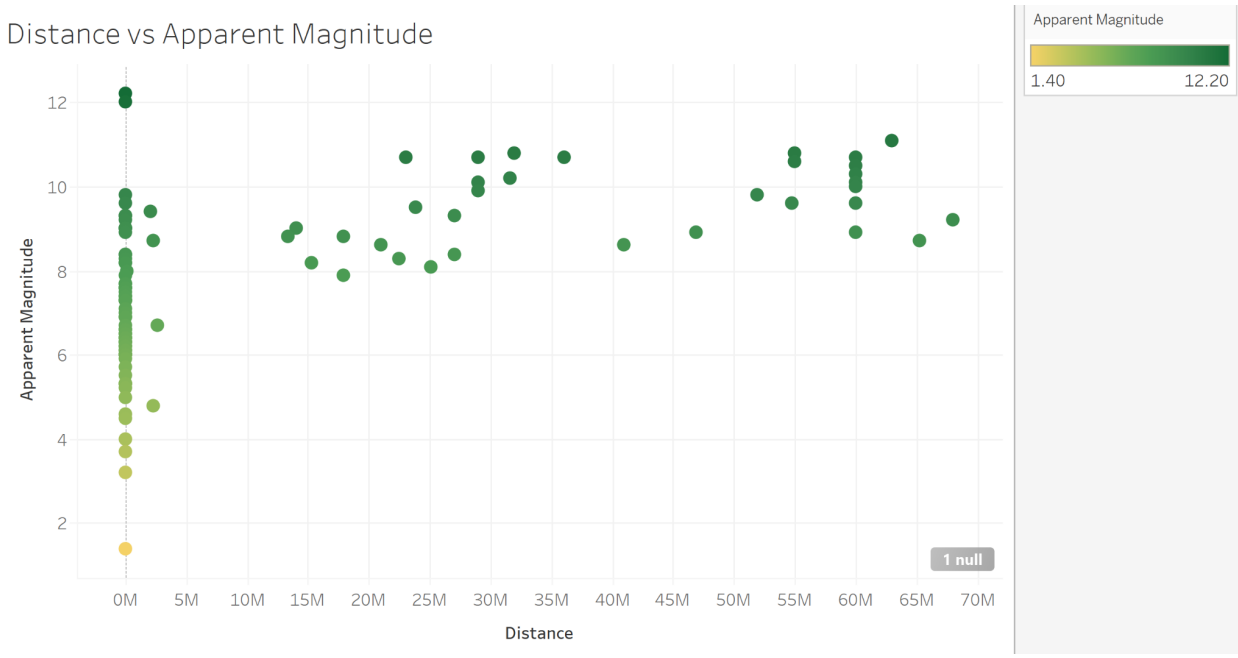


b) I struggled to find a good way to represent the distribution of Distances to each kind so I ended on this chart. From this chart below, you can still see that “Galaxy” has far greater distance than the rest of the Kind but it is so large of a

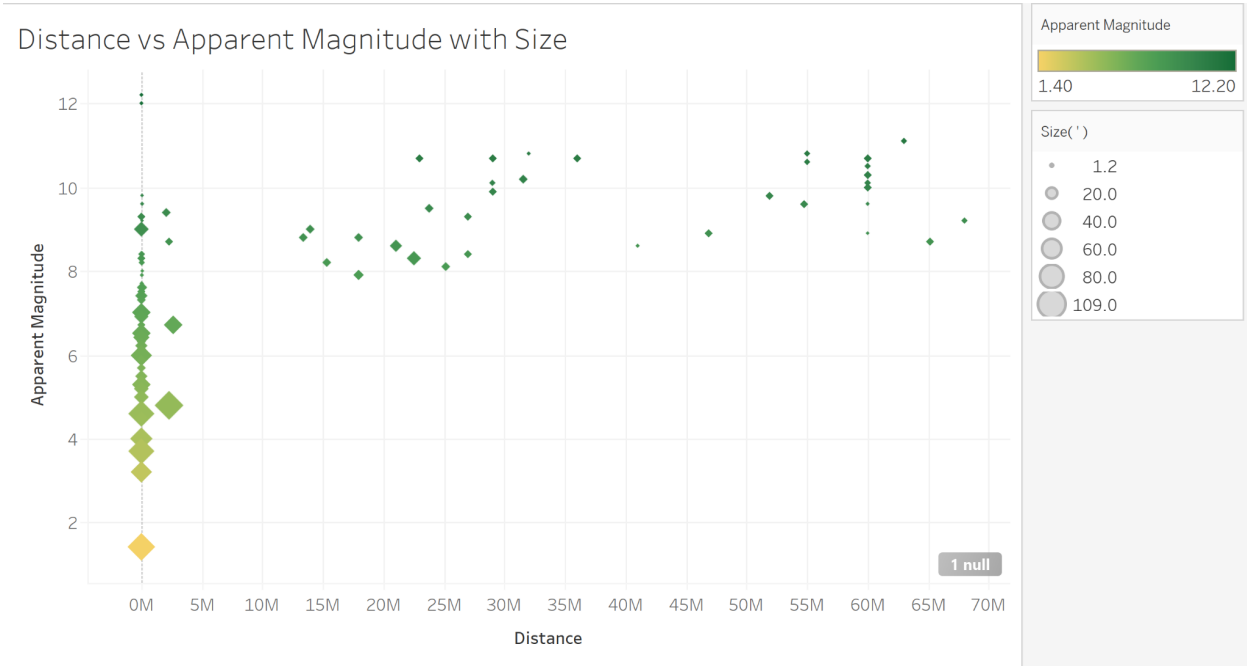
difference that it was hard to see the small changes amongst the other Kind.



c) For this chart, I wanted to use color so that it would be easier for the viewer to see which distances had a high “Apparent Magnitude”. This way aside from looking at the graph and comparing the point to where it lies on the Y axis, we can also look at how deep of a shade of green the point is and thus determine that it has a high “Apparent Magnitude” level.

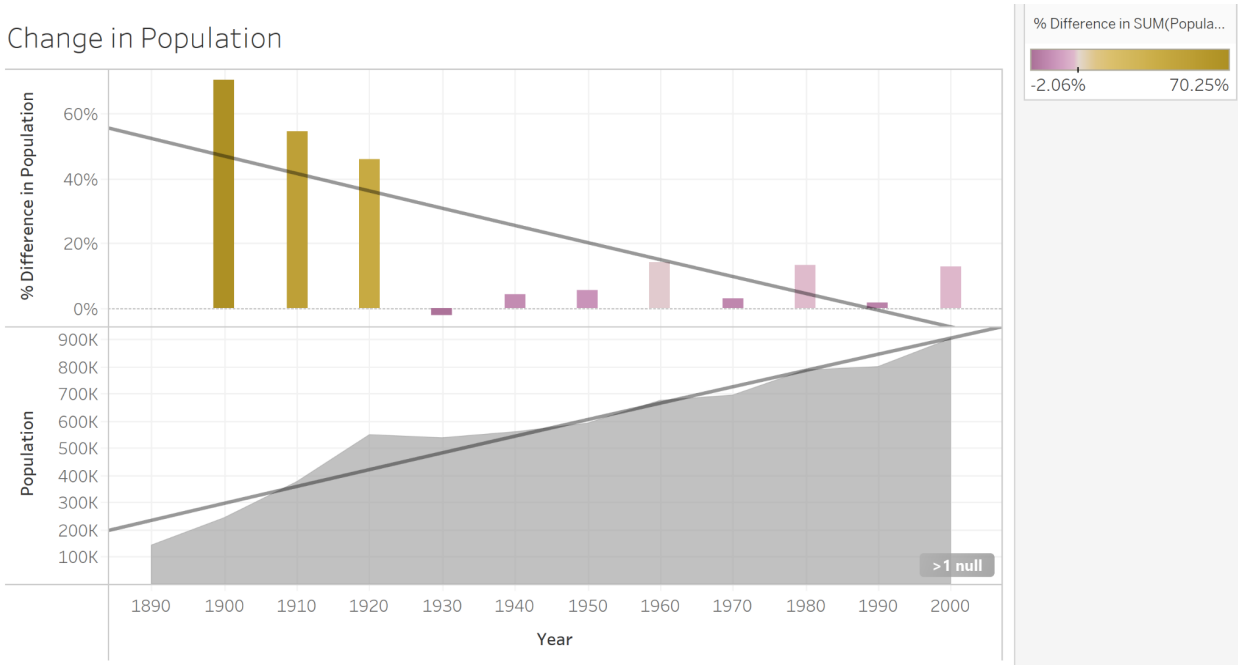


d) From the chart below, we can see that the size of the object in the sky is reflected in the size of the scatter point plot. I tried to change the shape of the point so that it may be slightly easier to tell the difference in size between a diamond than it is a circle. In order to display this information a bit more clearly, I may want to show a trend line that shows the relationship between the “Distance” and “Apparent Magnitude” where the farther the distance is, the easier it is to see.



3) Problems using MontanaPopulationData.csv
   a) The population has never doubled since 1890. The Population has increased since 1890 but the population has not doubled since the previous decade. I wasn’t sure if the question was asking if the population doubled when compared to the previous year (i.e. 1910 to 1920 then 1920 to 1930) or if the population doubled across multiple decades (i.e. 1890 to 1920). I went on the assumption that we are supposed to compare the previous years, thus the population never doubled since 1890. The graph shown below shows the growth of the population across the years and we can tell that the population never doubled because the percentage change between years is never above 100% which would mean the

population would have doubled.

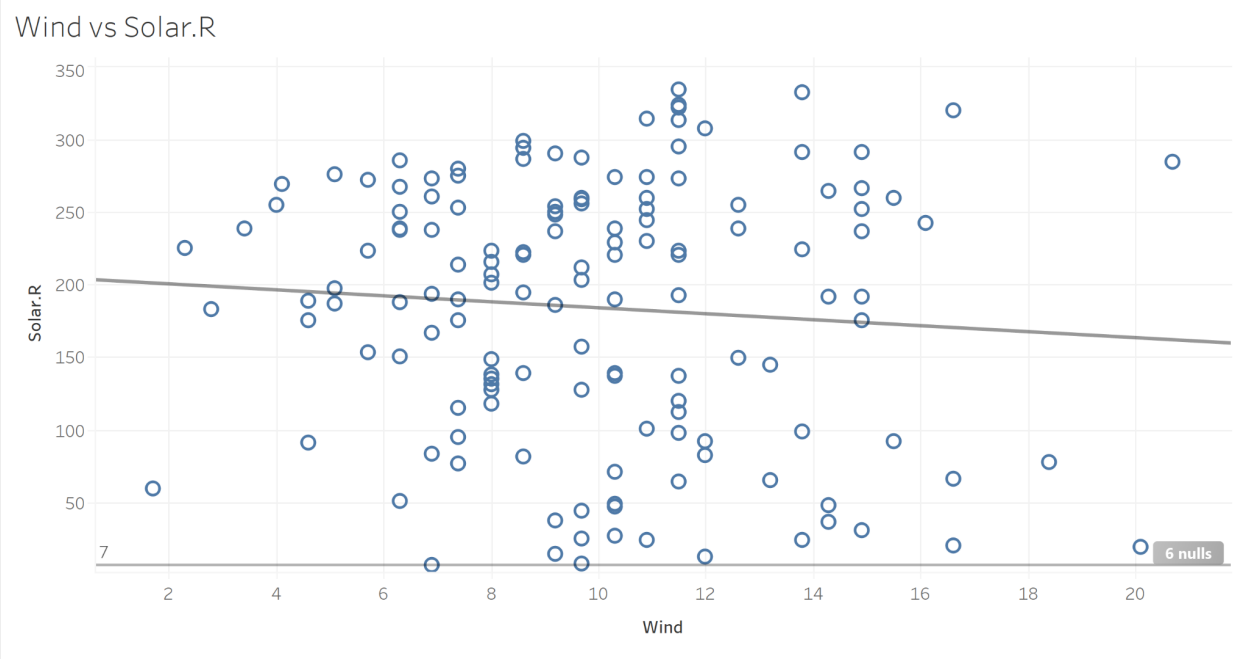


b) For this question, I used the same chart that was shown in question A. The rate of change has decreased over the years as shown by the decreasing trend line on the “Difference in Population” section. The year 1900 had the biggest increase in the rate of change as the population increased by 70.25%. The second highest increase in population was 1910 where the population increased by 54.55%.

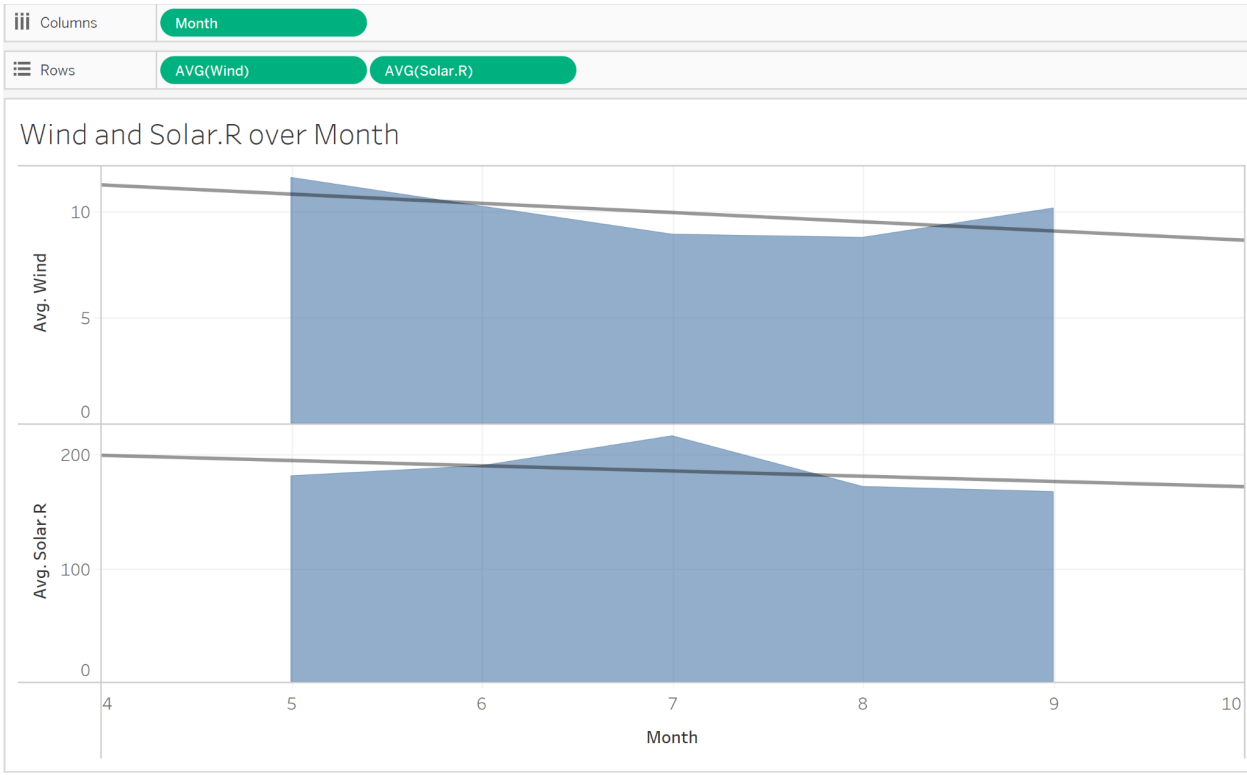
c) For this question, I used the same chart that was shown in question A. The years where the population increase percentage was greater than 15% was 1900, 1920, and 1920. The way you can tell in the graph is that they are the only bars that have a gold color. I categorized the color by which ones were above 15% and below 15%. The purple bars are below and the gold ones are above, thus easily showing which years had a percent increase greater than 15%.

4) Problems that use AirQuality.csv
   a) The chart shown below shows the relationship between “Wind” and “Solar.R”. There seems to be a slight negative trend when the wind increases, the Solar.R decreases. This is a very slight trend because the trend line shown on the graph

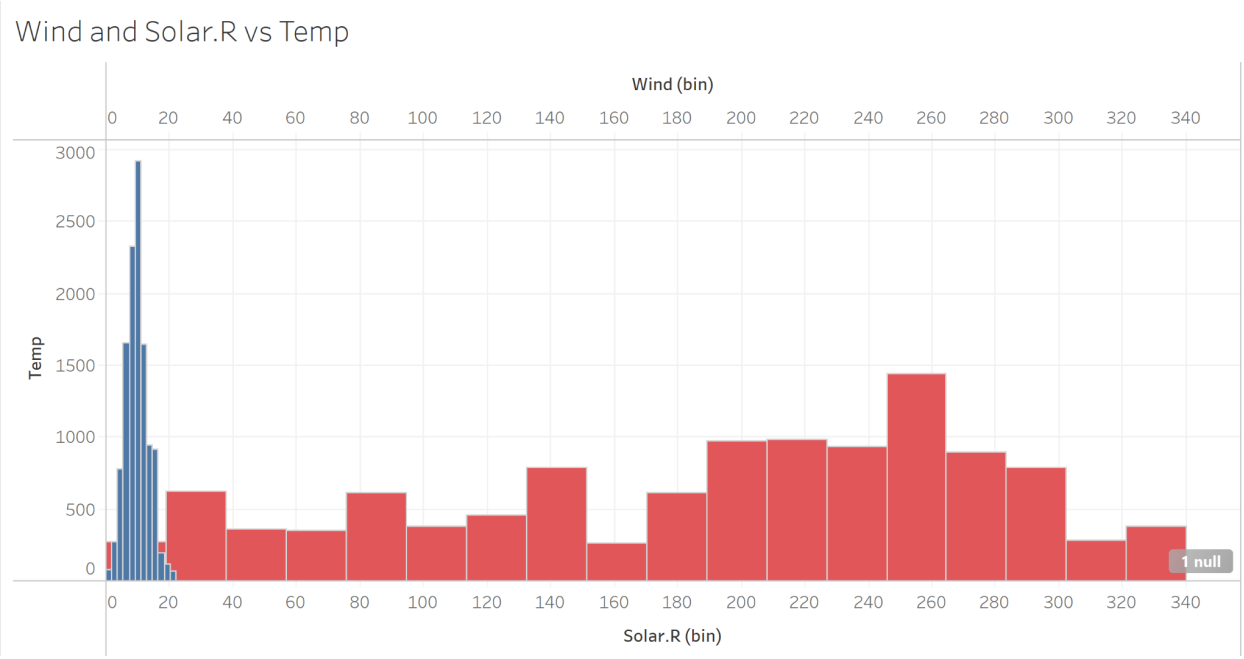
is not very steep.



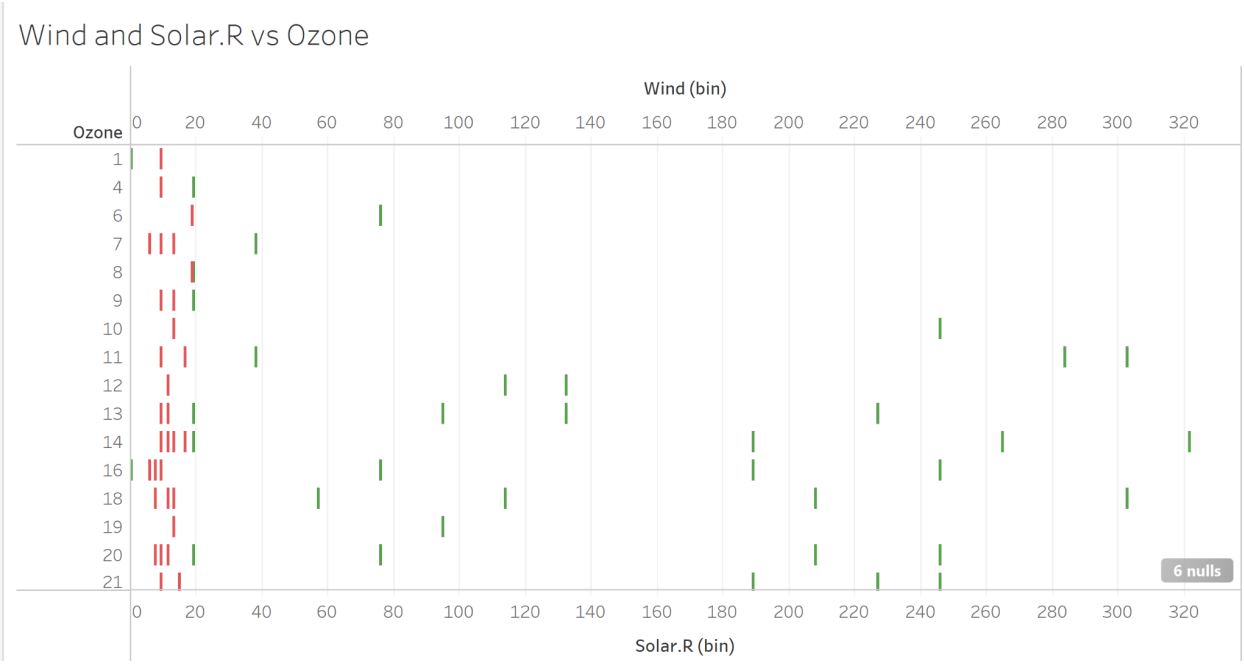
b) This plot shows the distribution of Wind and Solar.R across the months. I did not really know how I could make the plot more easily comparable between Wind and Solar.R so I tried to have their distributions next to each other with each of their trend lines showing. I found this question broad and felt that there would be many ways to go about it.



c) Here is Wind and Solar.R in respect to Temp.



Here is WInd and Solar.R in respect to Ozone.



I really had no clue how I would go about making these graphs or finding a way to make them look more presentable and easy to read, particularly, the plot comparing ozone. I tried to make it so that you can see the distribution of Solar.R and Wind on the same graph using overlaying plots and being separated by color but I don't think it turned out well. I couldn't find the tutorial that was described in the hint for this question and was totally stumped on how I would answer it. I

definitely remember learning about the ways we can compare multiple distributions in the lecture but I am just not aware of how to do them in Tableau.