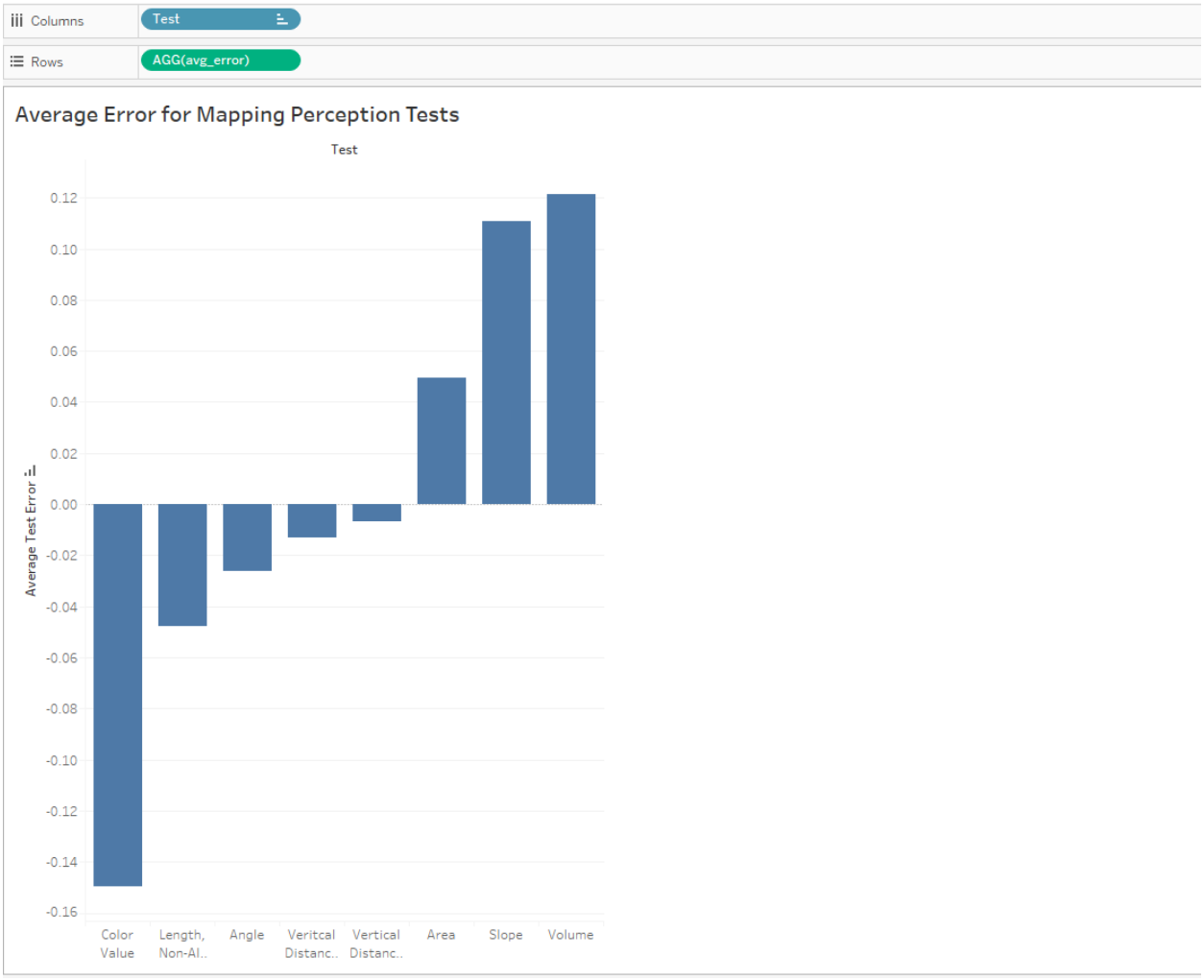
Yolanda Lewis
DSC 465
Homework 3

Problem 1

a. Were there any tests where people generally underestimated or overestimated the data? Explain what field you can graph to test this, what graphical method reveals this clearly. Analyze the results and explain in a short paragraph.

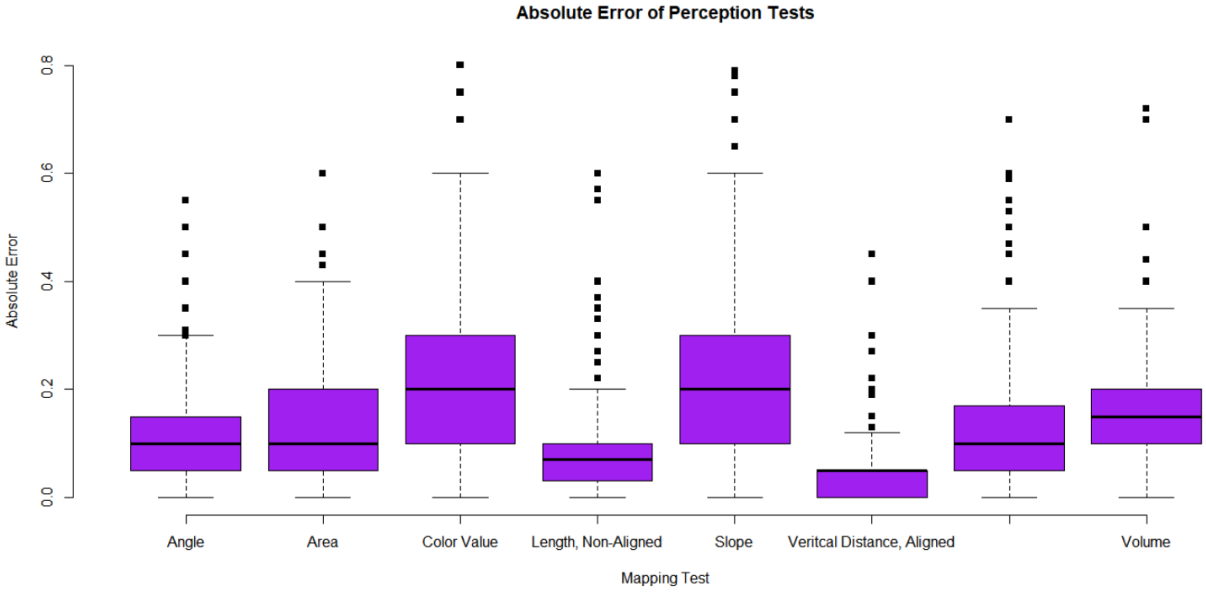


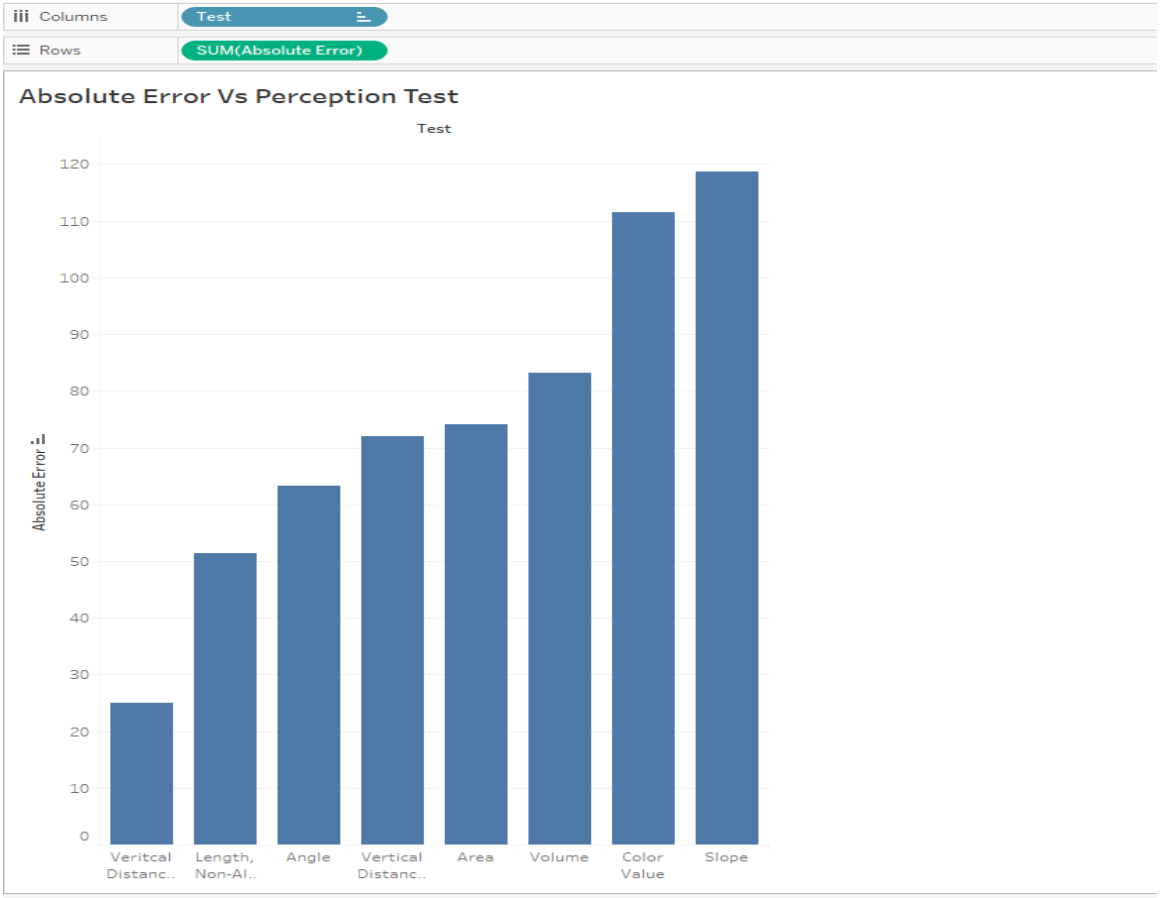
A bar graph plotting the average test error was used to graph the over and under estimation of mapping perceptions. In the bar graph above, we can see that color value the most underestimated perception and volume was the most over estimated perception. This could be due to the fact the color is not as easily distinguishable and volume can be easily mistaken on a graph. The bar graph clearly shows the amount of error less than zero or above zero.

b. Use a univariate scatterplot or another technique that shows fine detail for a collection of distributions. For each Test (do not divide between Display 1 & 2 or Trial B, C and D) plot the AbsoluteError (absolute value of Error). Then write a short paragraph of analysis. How do the

Yolanda Lewis
DSC 465
Homework 3

distributions of the data compare across the different methods our perception test studied for encoding numerical data visually? Is there any noticeable clumping of responses for any of the methods?

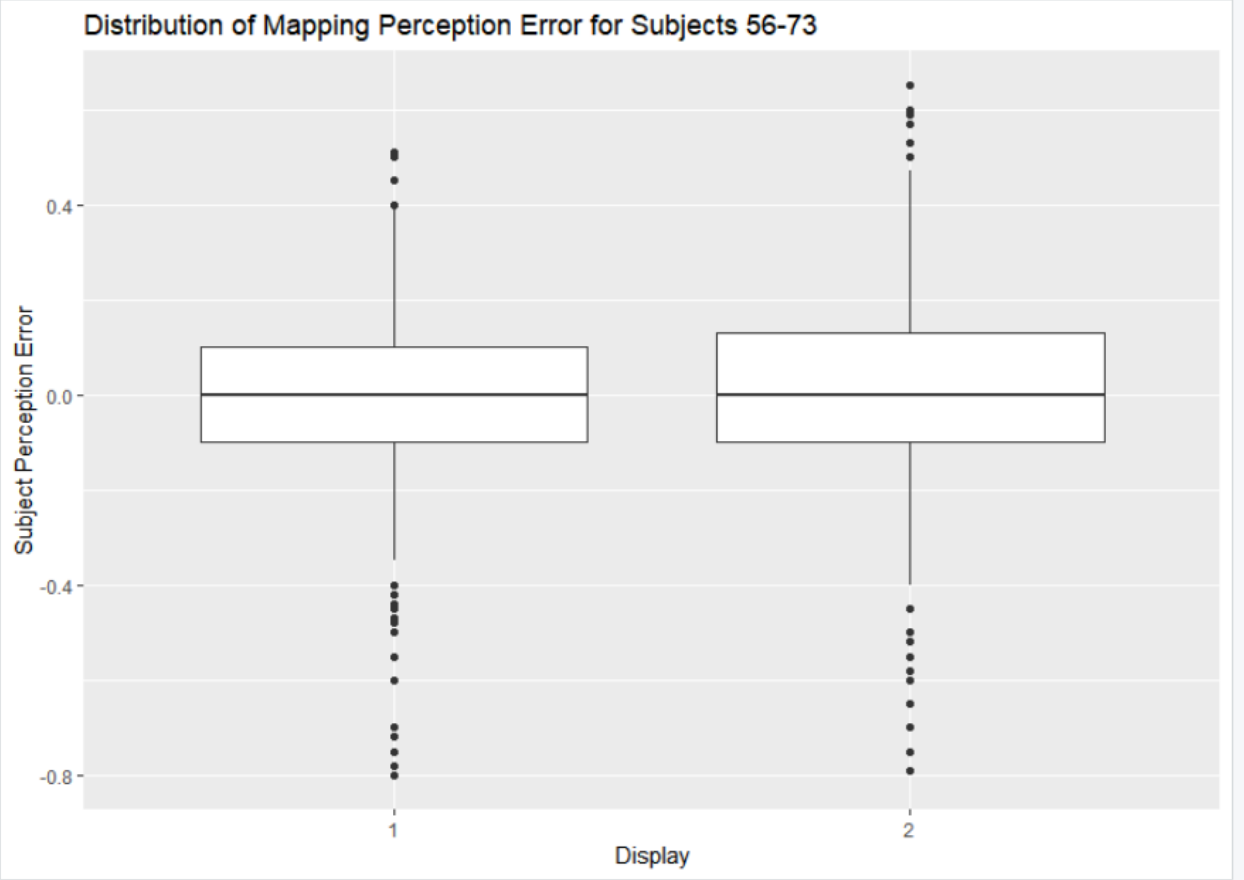




Yolanda Lewis
DSC 465
Homework 3

Boxplots and a bar graph were used to graph the distribution of the absolute error among each type of mapping distribution test. From each of the graphs we can see that the aligned vertical distance test yielded the least amount of error. The use of color and slope yielded the most amount of error. The distributions show that color along with slope are not as easily visualized in comparison to the aligned vertical distance or non-aligned length.

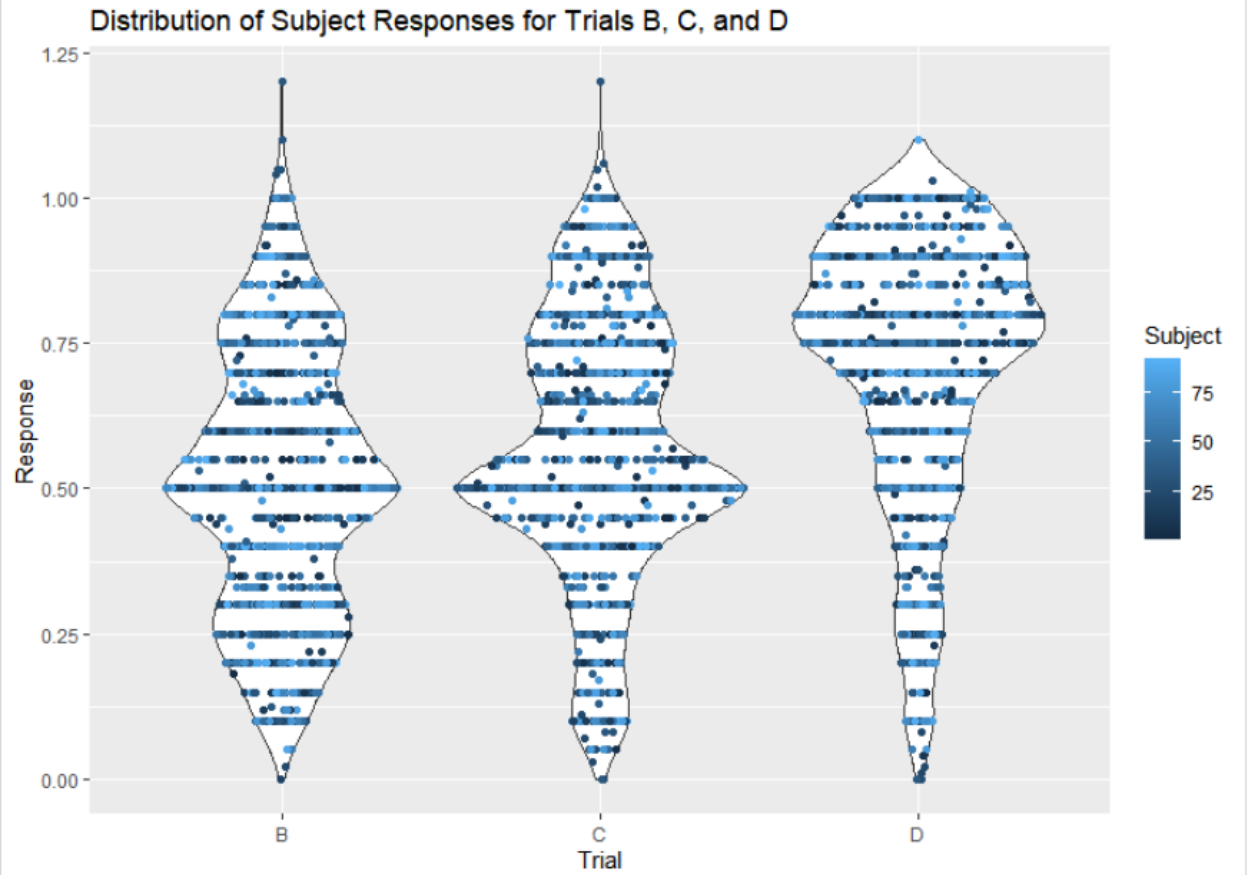
c. Compare the data for Displays 1 and 2 for subjects 56-73 (you will need to filter the data in Tableau or R). Create a visualization that shows any differences in the response patterns between the two. These subjects all saw the first set of Displays before the second set. Is there any difference in the values for Displays 1 and 2? Did the participants get better at judging after having done it once?



In the above plot, a boxplot was used to visualize the differences in response to display 1 and display 2 using in the perception tests. From the plot above, we can see that there is not a large difference in the amount of error during the first versus the second test. There is a slightly larger amount of error seen in display 2 with more values above and below the 1st and 3rd quartiles.

Yolanda Lewis
DSC 465
Homework 3

d. An erroneous stimulus was used for the first Display of "vertical distance, non-aligned" for a small subset of the subjects. They manifest themselves as an anomalous sequence of "1" Responses across Trial B, C and D. Look closely at the original raw scores and identify the sequence of subjects (hint: they are contiguous). Visualize the raw scores in a way that highlights these values and makes their anomalous nature clear. It should make it clear not only that they are outliers but should show any features that distinguish them from ordinary outliers. Some features that you might think about exploiting they are identical values across all three Trials, regardless of what the true values for the Trial is; they are only for a small subset of subjects.

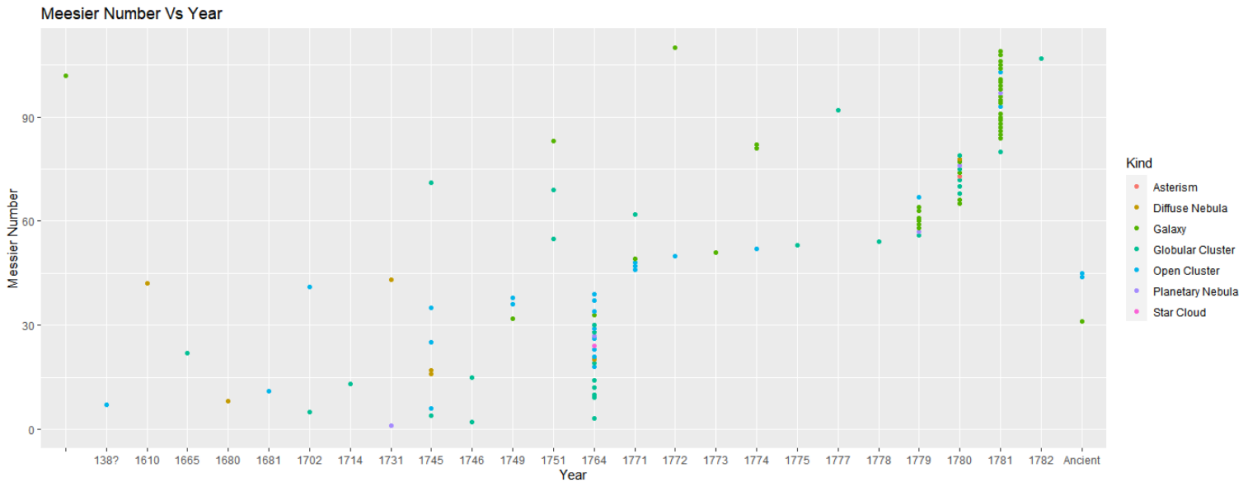


In the above plot, a violin plot was used to visualize the values that have a response value of "1" across the different trials. For trials B and C, the plot appears to narrow toward 1 for Subject numbers greater than 50. For trial d we can see that there is a greater distribution of responses between 0.75 and 1.00 for various subjects. From this plot we can see that many of the subject that guessed an answer of 1 were mostly subjects from Trials B and C.

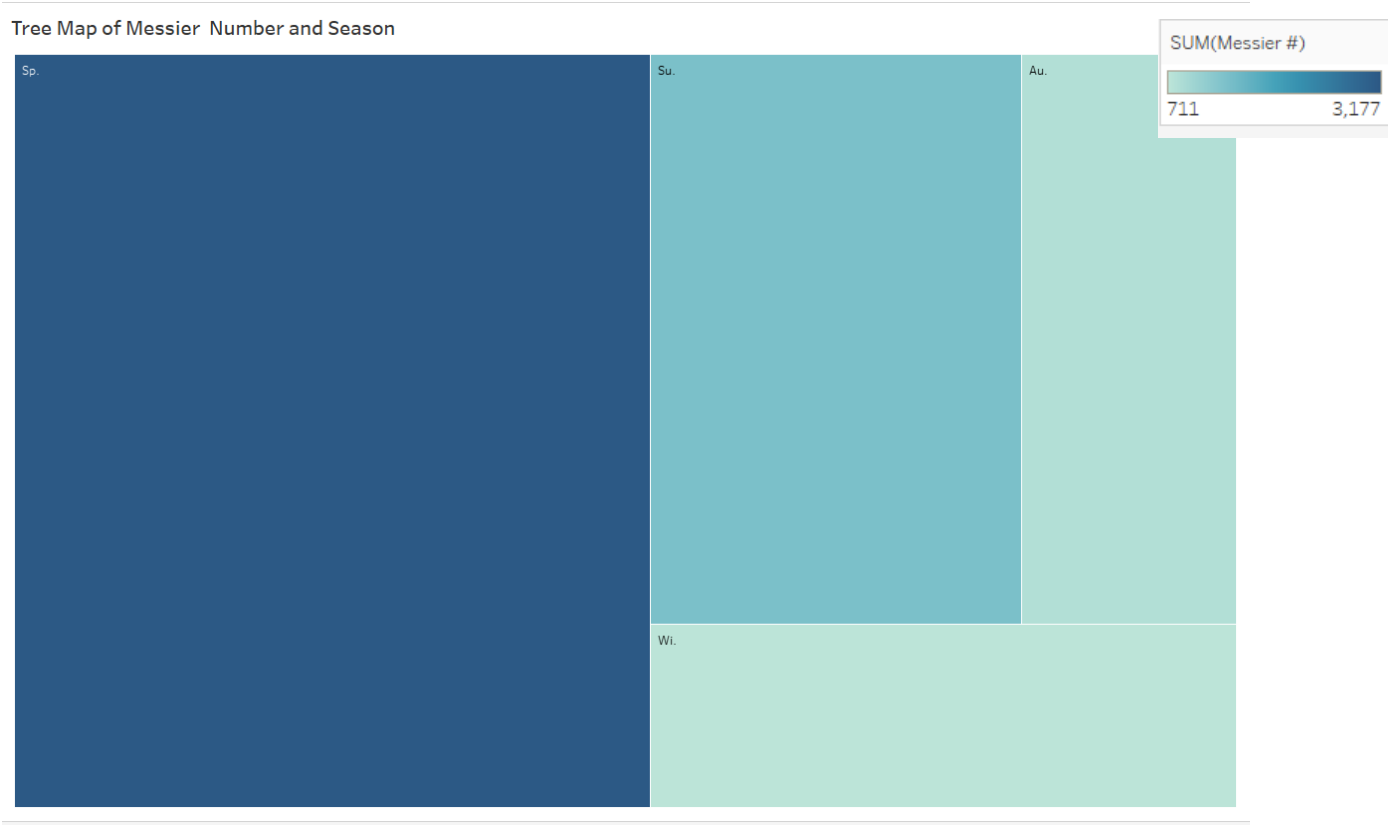
Problem 2

Yolanda Lewis
DSC 465
Homework 3

a. Start by trying to graph one or more properties of the objects against the Messier Number. Remember, there is nothing 'intrinsic' about this number, it is just the order of Messier's list. Is there any property that exhibits a pattern with respect to the ordering in his list?

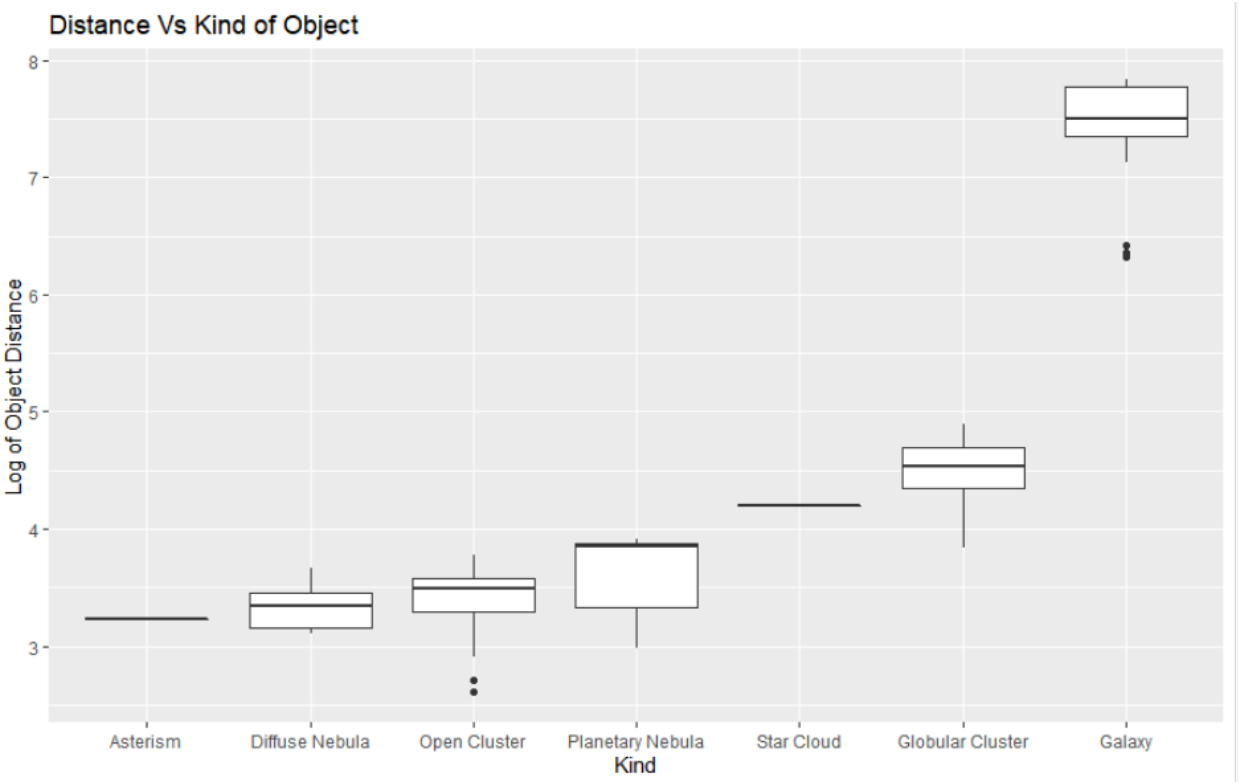


In respect to year versus the Messier number, from the scatter plot above, the Messier number increases as the year increases.



The above Treemap shows that majority of the Messier numbers were assigned in the Spring. The second largest amount of Messier number were assigned in the Summer. This could be because many of the messier objects could only be seen during these seasons which could have impacted the ordering on the list.

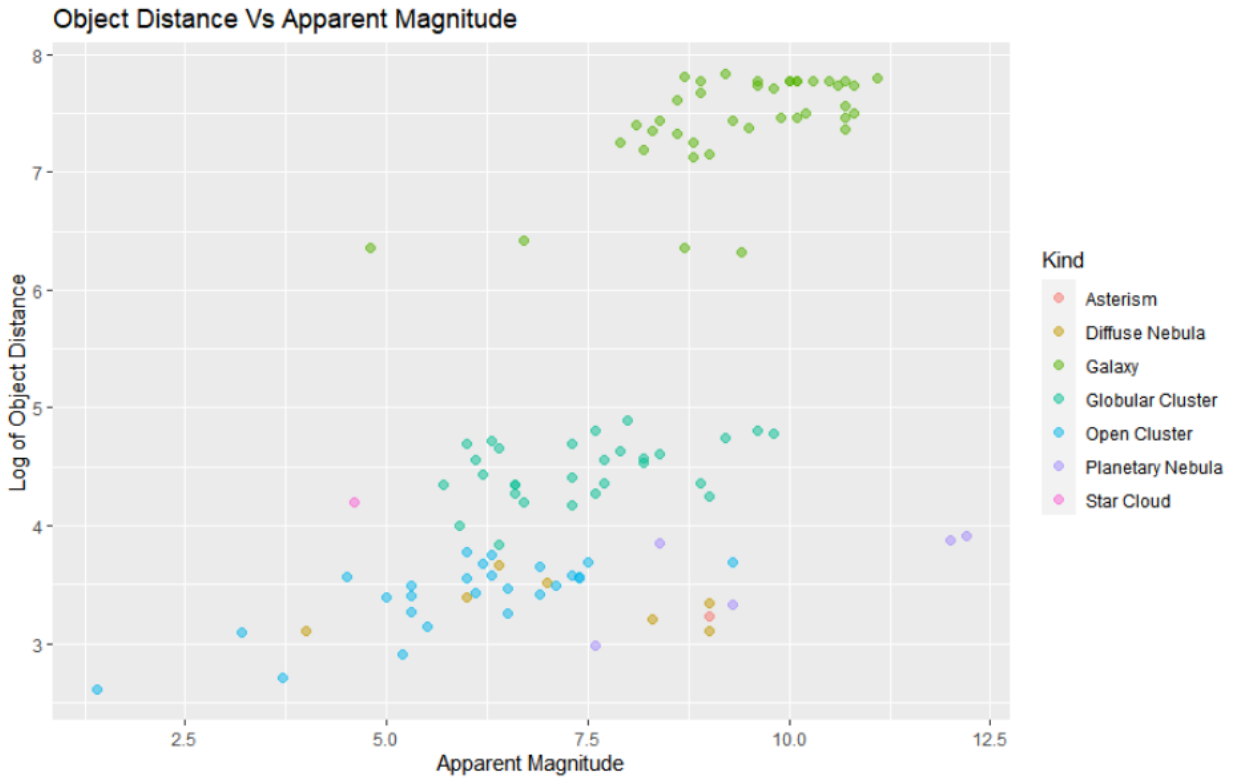
b. Create a visualization that compares the distributions of the distances to the objects in each Kind. Note that the Type variable is a very different category and is really a subcategory of Kind. Do not use that here. Sort the distribution displays in a way that makes the relationship clear.



Boxplots are used in the above plot to show the distribution of the distances of each kind of Messier object. The kind of object is sorted based on the distance. From the above plot it appears that in relation to the distance, the objects in the furthest galaxies have the furthest distance and the Asterism object appears to have the closest distance.

c. Create a scatter plot with the distance to the Messier objects plotted against their Apparent Magnitude (it is their visual magnitude, a measure of how bright they are in the sky). Note that these values may be… backwards from what you would think. The higher the number the fainter the object is in the sky. Try to incorporate that into your visualization to make the relationship clear.

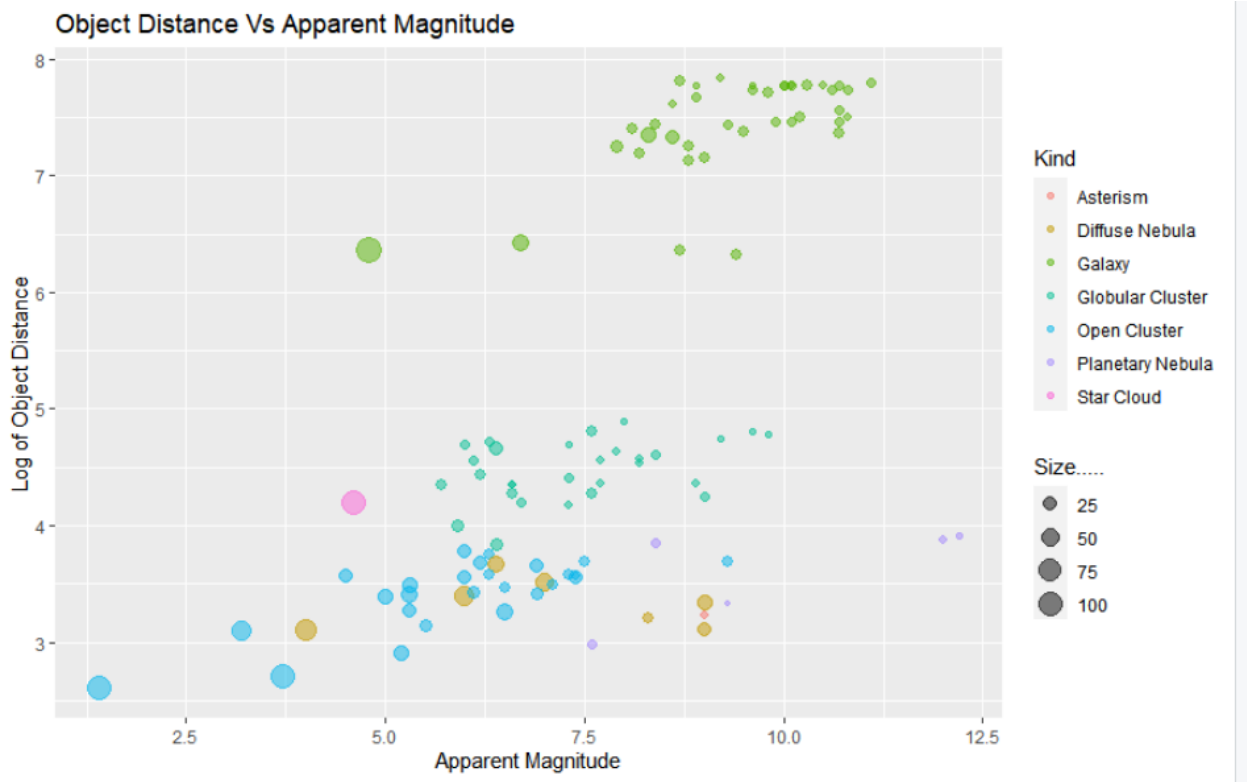
Yolanda Lewis
DSC 465
Homework 3



The above scatter plot shows the Messier object versus the apparent magnitude of the object according to the kind of Messier object. The galaxy objects shown in lighter green are the furthest away. The x axis is transformed to use a log 10 base. The open cluster object is the closest object. The objects with apparent magnitude between 5.0 and 10.0

d. Augment the visualization in (c) by adjusting the size of the points in the scatterplot based on the angular Size of the objects in the sky. Evaluate how easy it is to analyze all encoded aspects of the data from this graph and give a suggestion on how you might modify the graph to display all this information more readably.

Yolanda Lewis
DSC 465
Homework 3

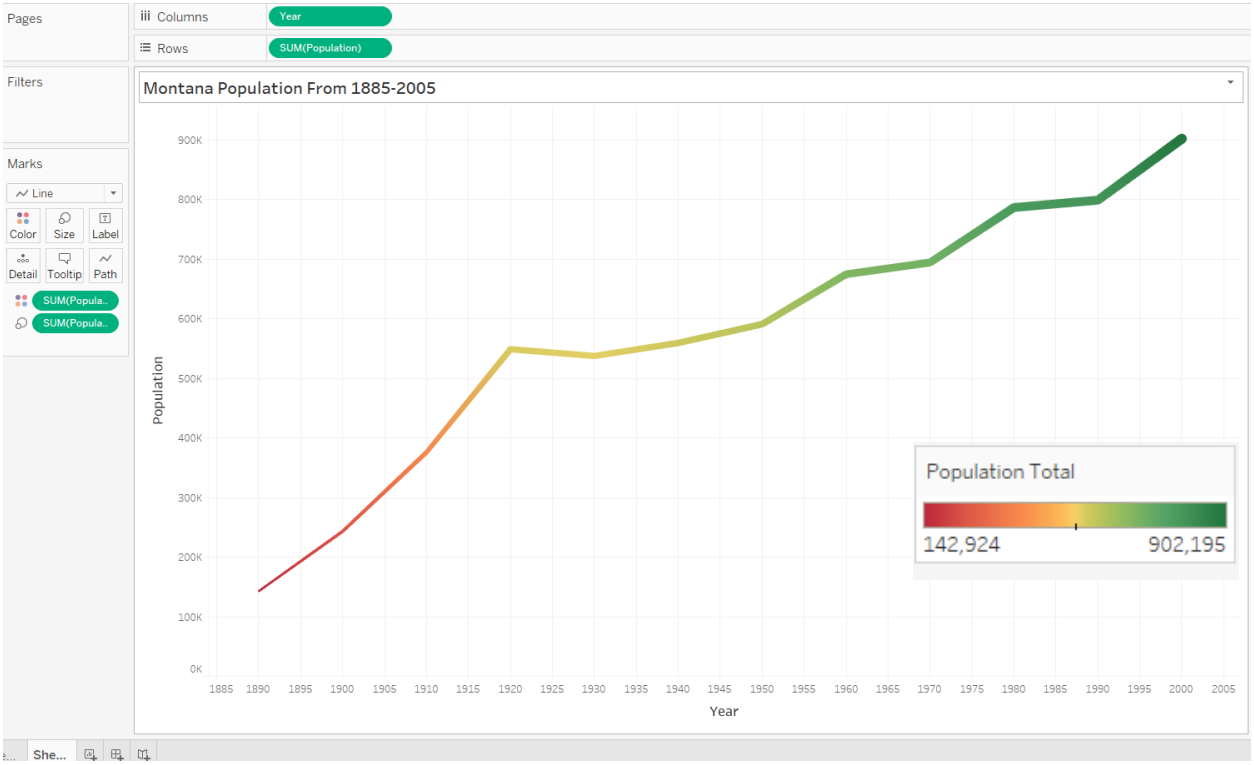


The above scatterplot shows the relationship of the Magnitude of an object versus its distance. The size of the object is adjusted based on the size of the object seen in the sky. So, the closer the distance of the object the larger its size. Adding the angular size makes it easier to see how visible the object might be the varying distances. I might modify the graph by changing the color scale along with the size, making the larger objects darker and the smaller objects lighter.

Problem 3

a. How many times has the population doubled since 1890?

Yolanda Lewis
DSC 465
Homework 3

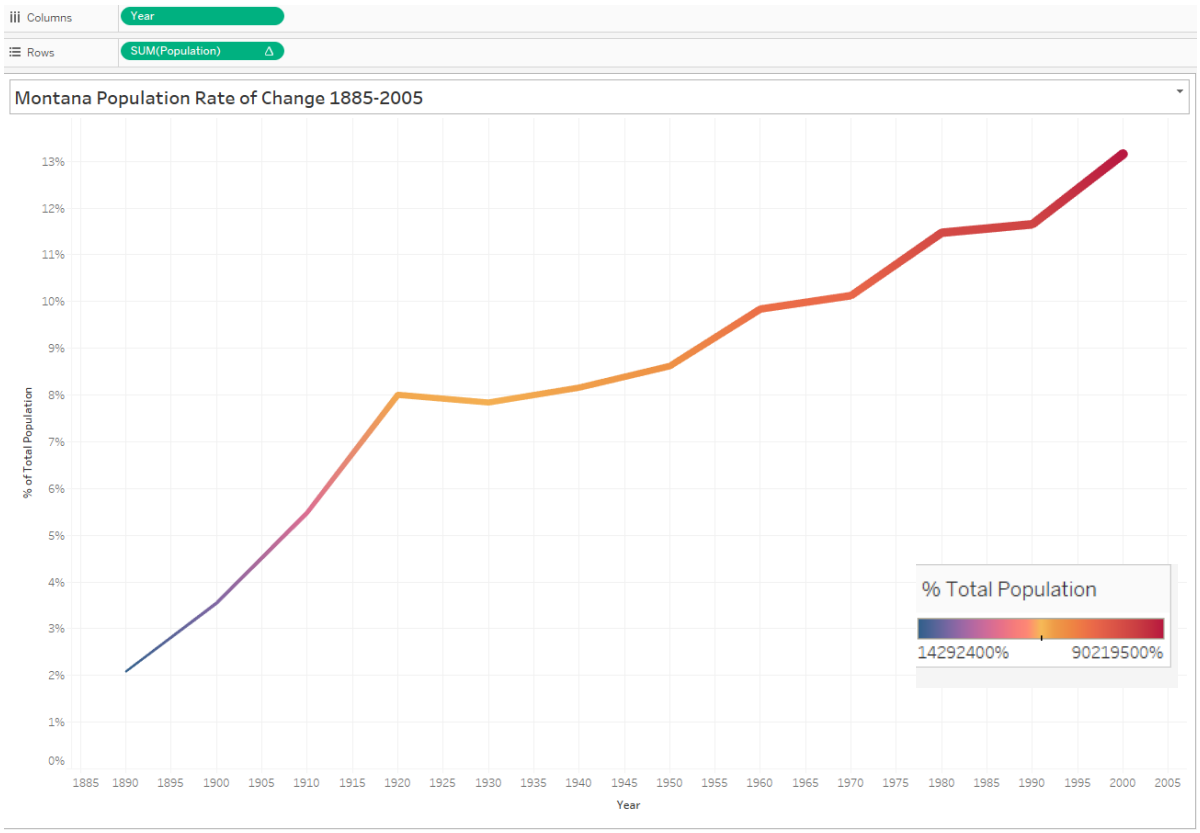


The population in Montana has doubled twice since 1890. The population appears to have doubled from 1890- 1920 and from 1920- 1970.

b. Has the percentage rate of change in the population increased or decreased over the years? What years had the greatest increase in population %-wise?

The percentage rate of change has increase over the years. The years 1890-920 saw the greatest increase in population percentage wise with a percentage increase from 2%-8%. See the plot below.

Yolanda Lewis
DSC 465
Homework 3

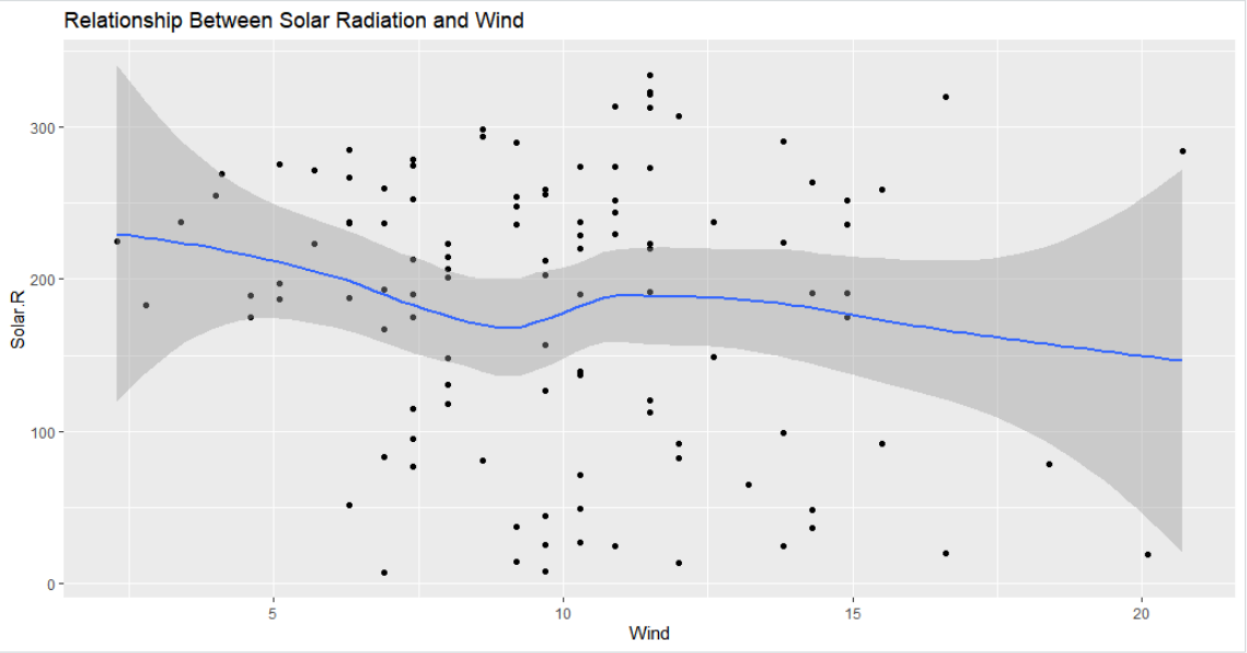


c. What years was the population percentage increase greater than 15%?

Using the plot above we can see that the population appears to be trending upward. There appears to be a percentage increase greater than 15% between the years of 1890 and 1915.

Problem 4

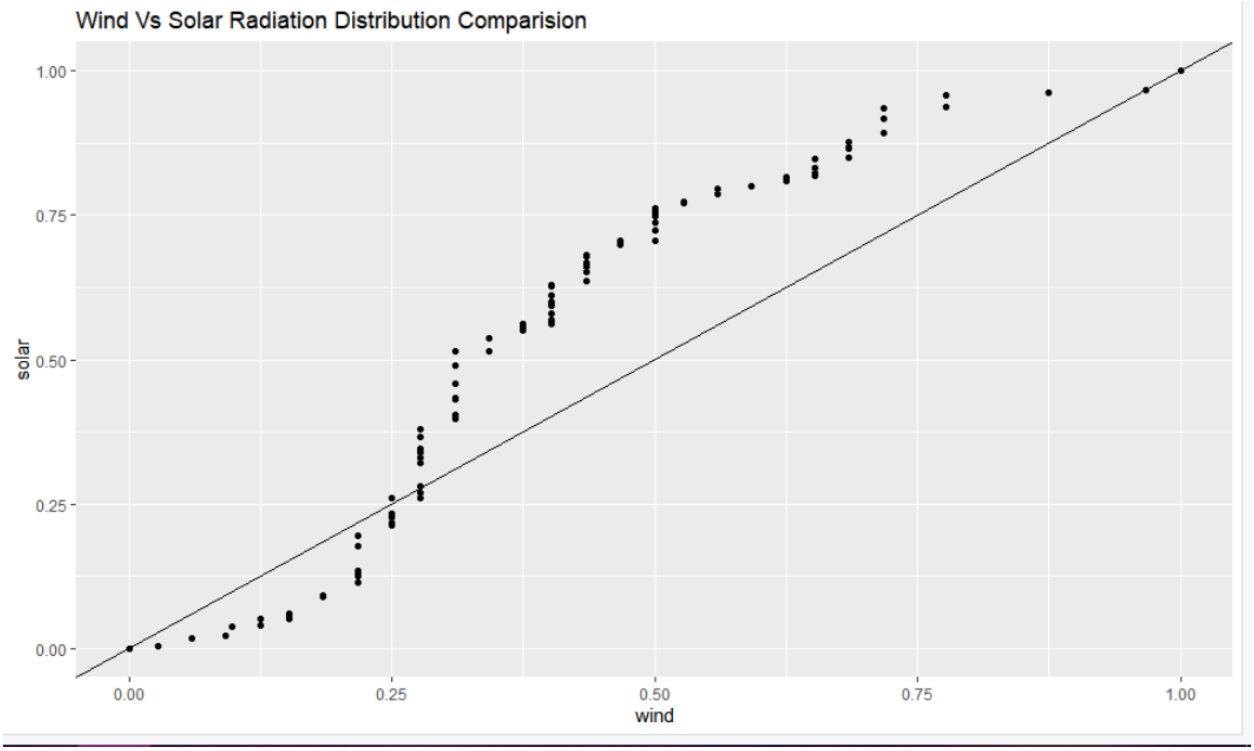
a. Use a scatterplot to look at the relationship between Wind and Solar.R (solar radiation).
Show a fit line. Make sure to produce a clean visualization with emphasis on the trend.
This provides one view of the relationship.



The plot above compares the distribution of Solar radiation and wind. The scatterplot shows that the wind pattern fluctuates as the solar radiation increase. It appears that the wind measurement narrows then expands as the wind measurement increases.

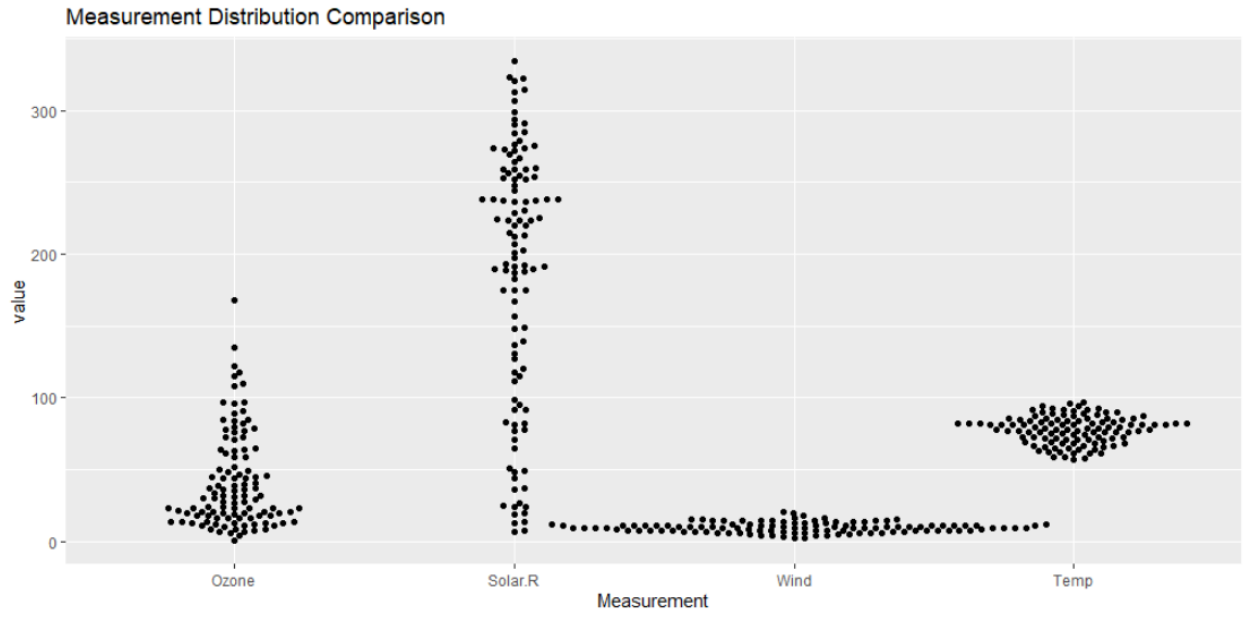
For help doing this in R, see Tutorial 5. In Tableau, this is available from the Analysis tab. It is one of the tabs along with Data for the panel on the far left (i.e. look at the top of the panel from which you drag variables).

b. Use a plot that will show the distributions of Wind and Solar.R and allow you to compare with fine detail.



The distribution of solar and wind appears to be similar. As the wind increases so does the solar radiation. The fit line shows that the data may not be normally distributed.

c. Finally, show these distributions in context of the rest of the variables by using a technique for comparing multiple distributions.
Note: you will need to transform the data in a particular way that we have studied. I it showed in the Tableau tutorial and in an R tutorial. Hint – you need to collapse the current variables into two: (1) stores the original variable name, and (2) stores the corresponding original value.



The measurement distribution in performed using a bee swarm plot. The measurement comparison shows the distribution of values among the different measurements versus the value given for each measurement. Ozone and wind appear to be more widely distributed with lower values. Solar radiation is fairly evenly distributed and the values increase. Temp appears to cluster together at a lower value.

d. For extra credit, compare Wind and Solar.R again with a QQ plot. What does this tell you?

Not completed