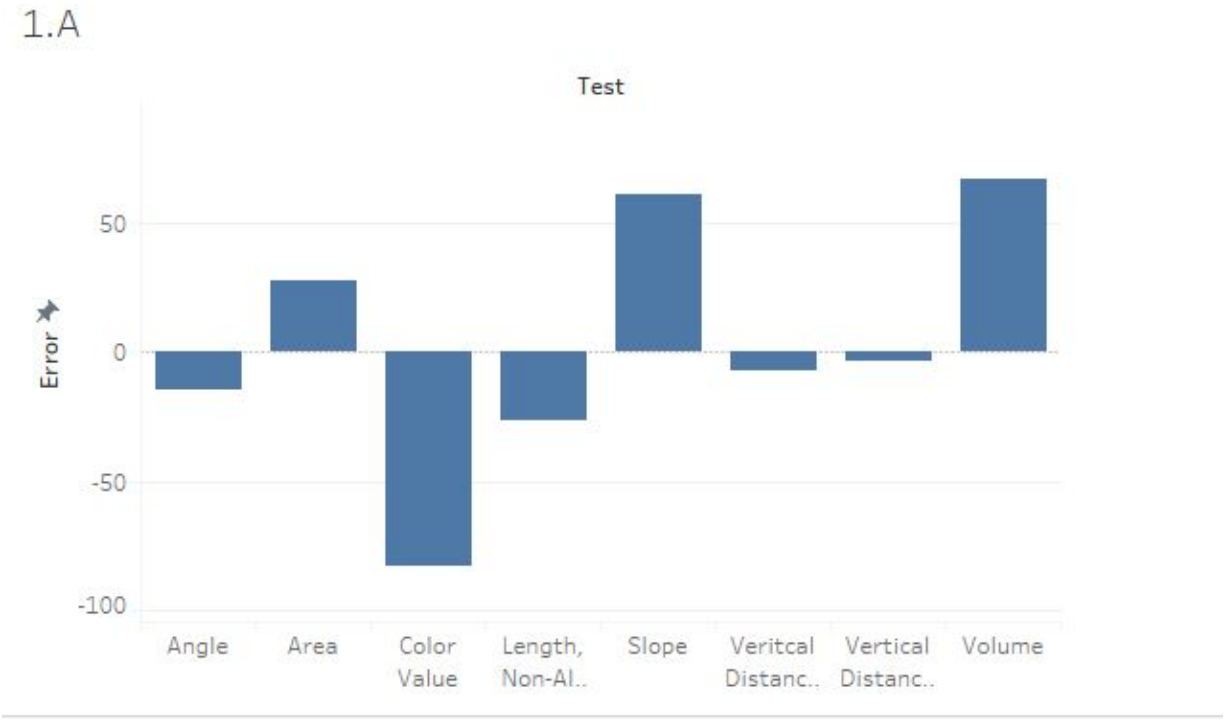
A) Were there any tests where people generally underestimated or overestimated the data? Explain what field you can graph to test this, what graphical method reveals this clearly. Analyze the results and explain in a short paragraph.

Ans:

To understand what is overestimation and underestimation of data, when response is more than or less than the true value we can say there is overestimated data.

- To test all fields error in estimation we know the equation Error = Response - True Value, so we can know which field is overestimating or underestimating.
- So I have used a bar plot to clearly reveal the hidden pattern in the data.
- You can clearly see that from the graph, Color value and length non aligned are highly underestimated, where slope and volume are overestimated.
- There are some minor errors Vertical Distance and Angle.
- Where the area is slightly over estimated.
- And, Length non aligned is slightly underestimated.



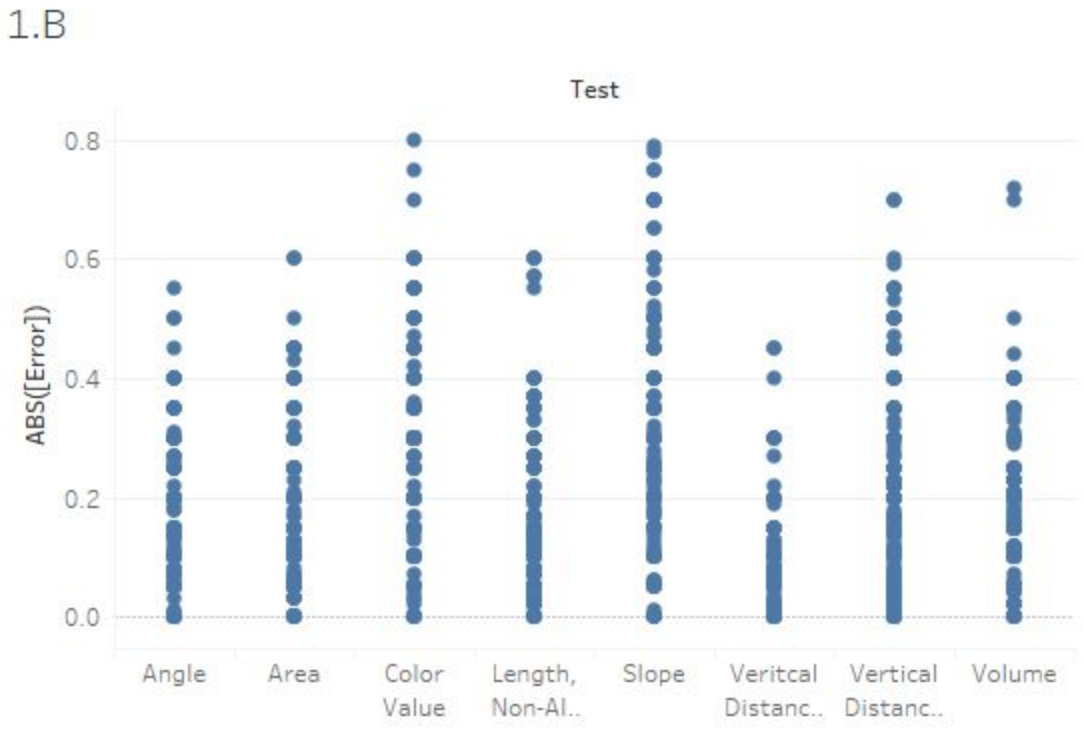
To be Continued

B) Use a univariate scatter plot or another technique that shows fine detail for a collection of distributions. For each Test (don't 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 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?

Ans:

From the below, you can see that I have used a univariate scatter plot which is showing fine distribution of the data. And haven't included 1& 2 or trial B,C and D.

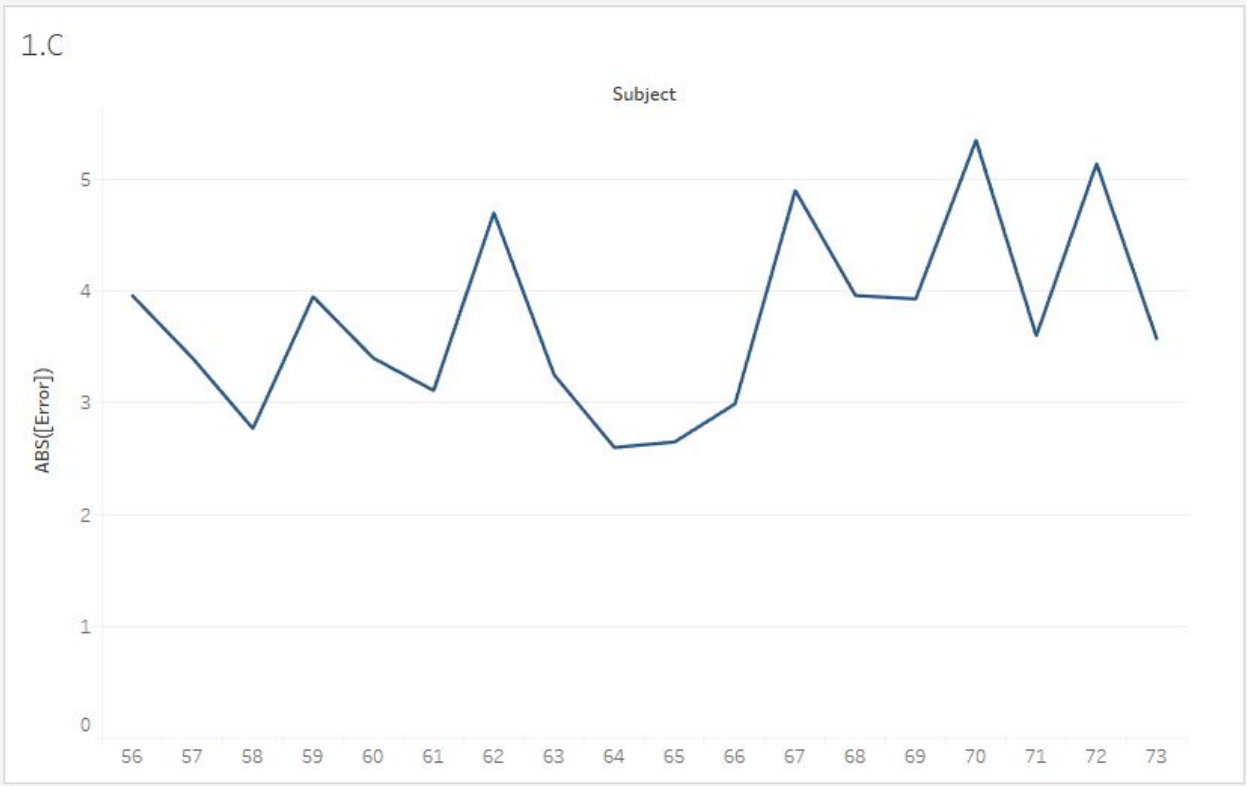
- There are overestimates and underestimates of the value so Error will be negative and positive value so we are using absolute value here. Hence, All the differences will be positive.
- Color value and Slope have high absolute error in the test and individual field. So it is hard to perceive human eyes to see the difference between color and slop.
- Where vertical distance has less absolute error if we ignore some outliers. SO we can perceive easily
- From the graph you can see that Angle, Area, Length and Volume have normal absolute error. For making this assumption I have neglected the outliers.

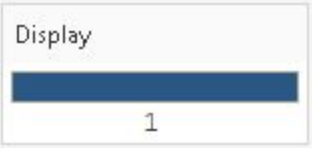


To be Continued

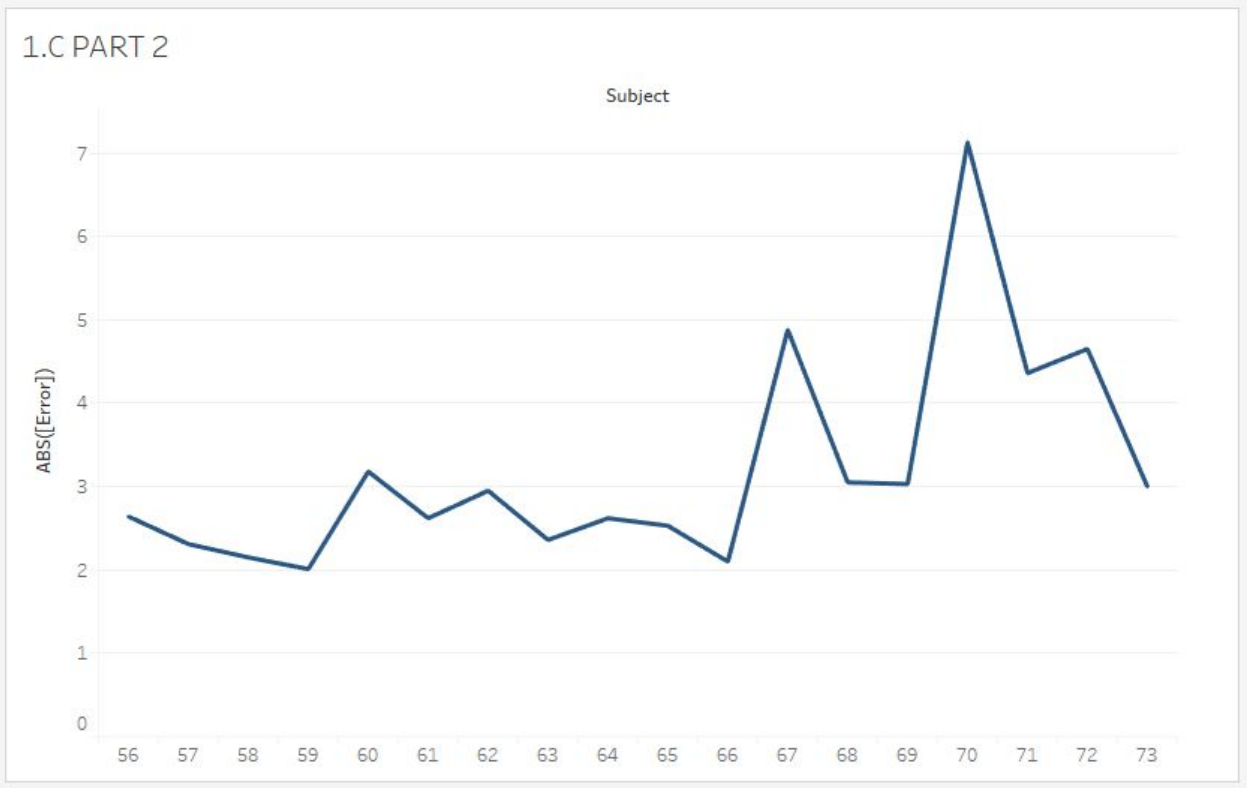
C) Compare the data for Displays 1 and 2 for subjects 56-73 (you will need to filter the data in Tableau or R). Create a visualization that shows any differences in the response patterns between the two. These subjects all saw the first set of Displays before the second set. Is there any difference in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

Ans: From the graph, you can see from display 1 and display 2 I have plotted two different graphs for subject 56 to 73 which is shown below. So I have shown the two different line graphs**.** You can see there is a difference in the pattern 62-67. Which is a huge difference in the graph. From the two graphs you can see a clear difference between display 1 and display 2 and it shows the first set of displays before the second set. There are differences between 62-67 from the displays 1 and 2. After closely observing the both displays we can say that participants are getting better at judging after having done it once.





To be Continued



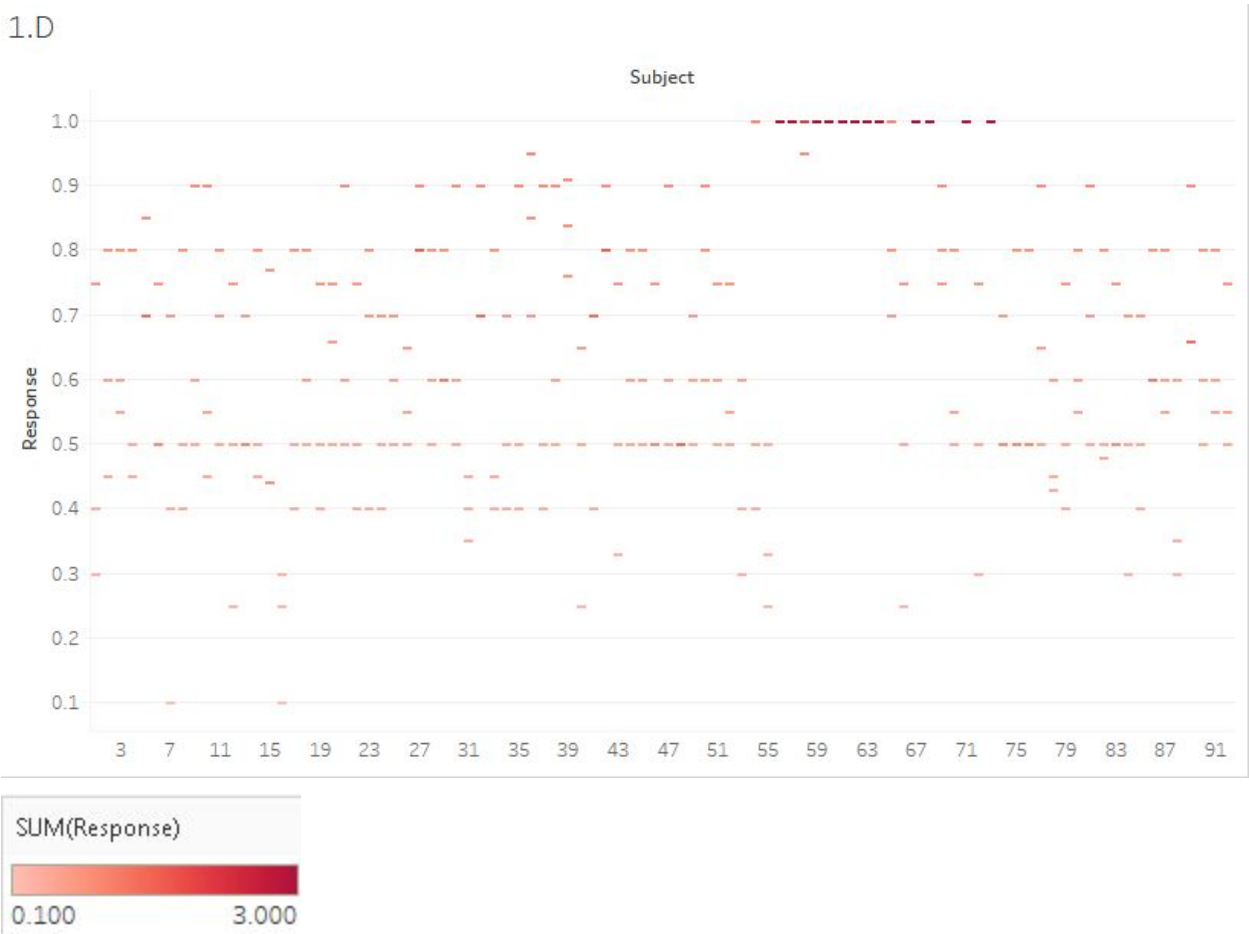
D) An erroneous stimulus was used for the first Display of “vertical distance, non-aligned” for a small subset of the subjects. They manifest themselves as an anomalous sequence of “1” Responses across Trial B, C and D. Look closely at the original raw scores and identify the sequence of subjects (hint: 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.

Ans: From the graph you can see that, we have shown a graph of subject and response. An erroneous stimulus was used for the first display of “Vertical distance, non-aligned”
For a small subset of the subjects. Graph can see from the graph original raw scores in a way that highlights these values and makes their anomalous nature clear. From that we can see outliers clearly not only that it also features that distinguish them from ordinary outliers.

You can see series of 1 from 53 to 73 so you can say that, that is the erroneous stimulus in the graph.

To be Continued

1.D



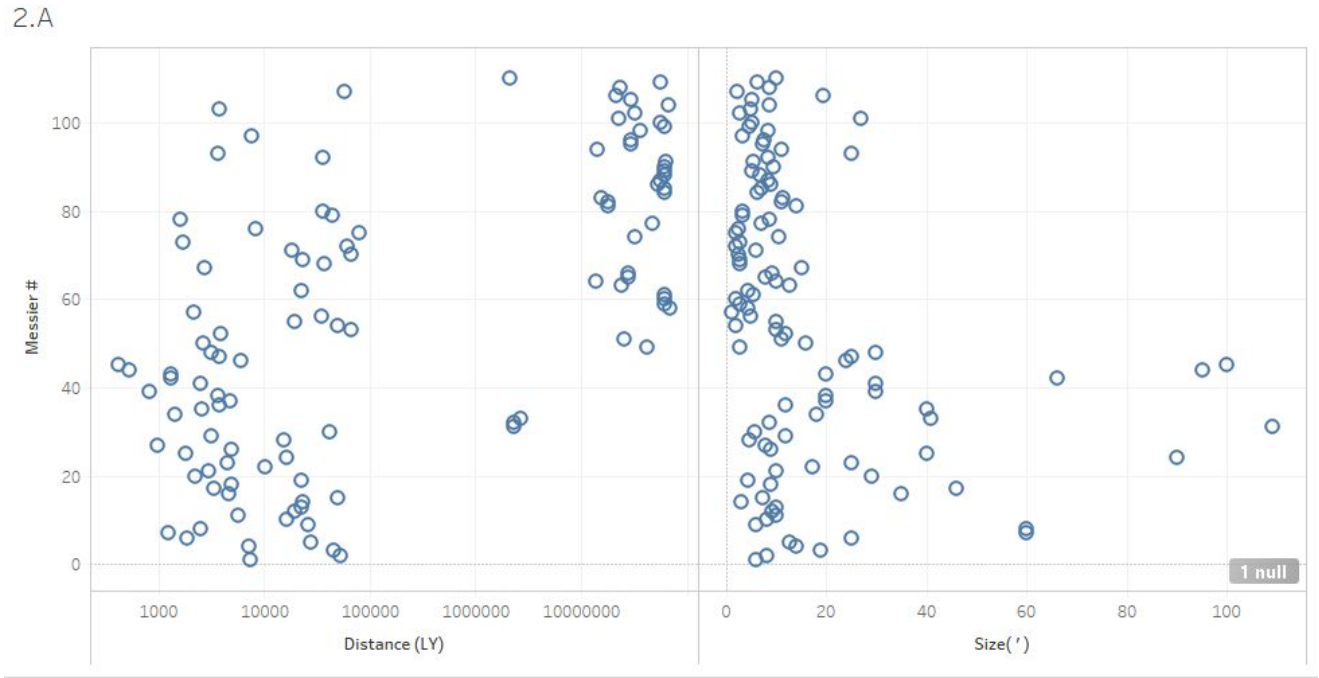
To be Continued

Question 2

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?

Ans:

From the graph, you can see that we have graphed the size and distance with respect to messier data. It is clear that we can not find any relation between two properties size and distance so we can say that there are not any such patterns with respect to the ordering in the list.



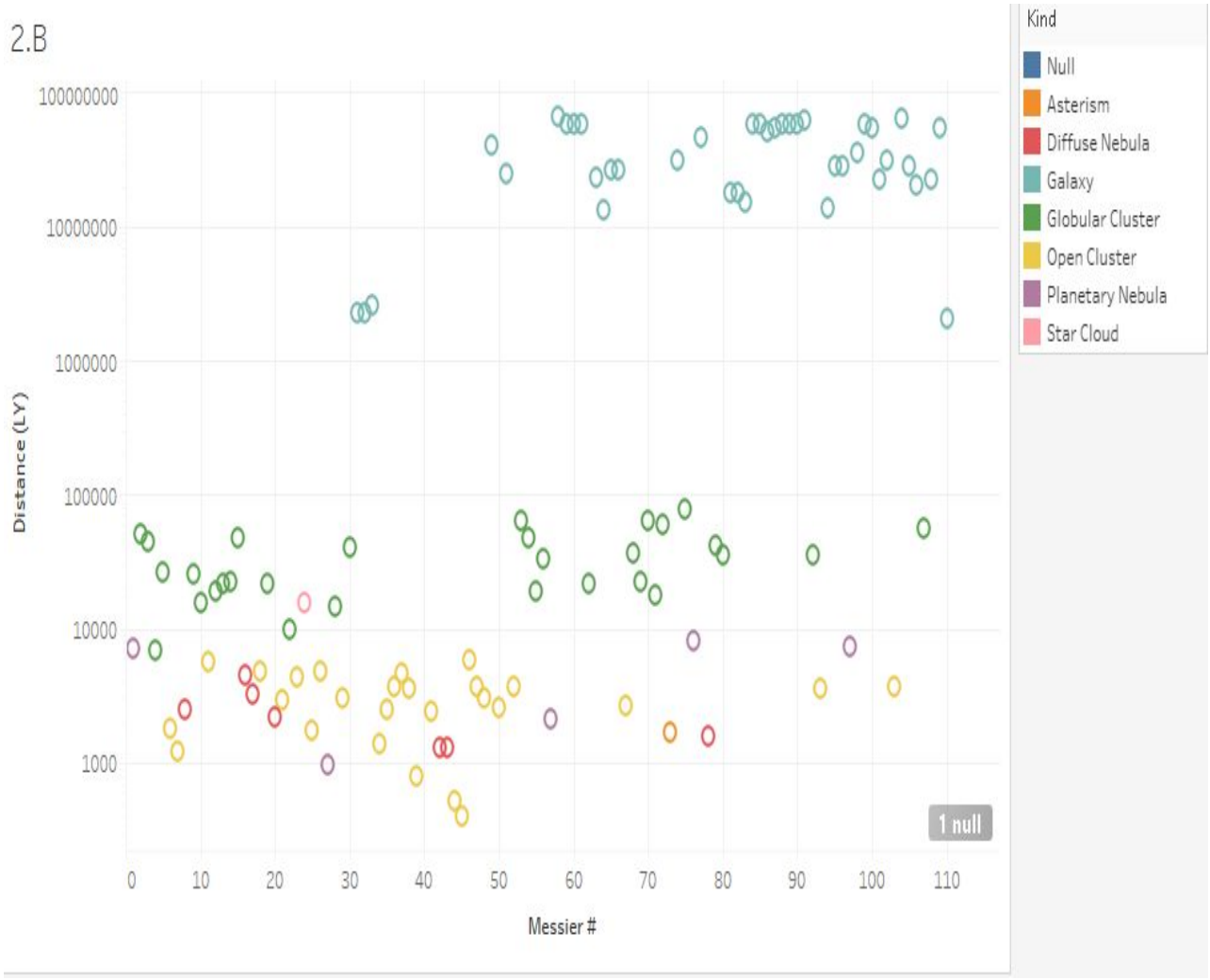
B) Create a visualization that compares the distributions of the distances to the objects in each Kind. Note that the Type variable is a very different category and is really a subcategory of Kind. Do not use that here. Sort the distribution displays in a way that makes the relationship clear.

Ans:

To be Continued

From the gra[h, you can see that it compares the distributions of the distances to the objects in each kind. From the the below graph you can see that,

- Omega Nebula and open clusters are only 1000 light years away from the earth
- Where globular clusters are 10000 to 1,00,000 light years away.
- Only galaxies are the farthest from the earth.



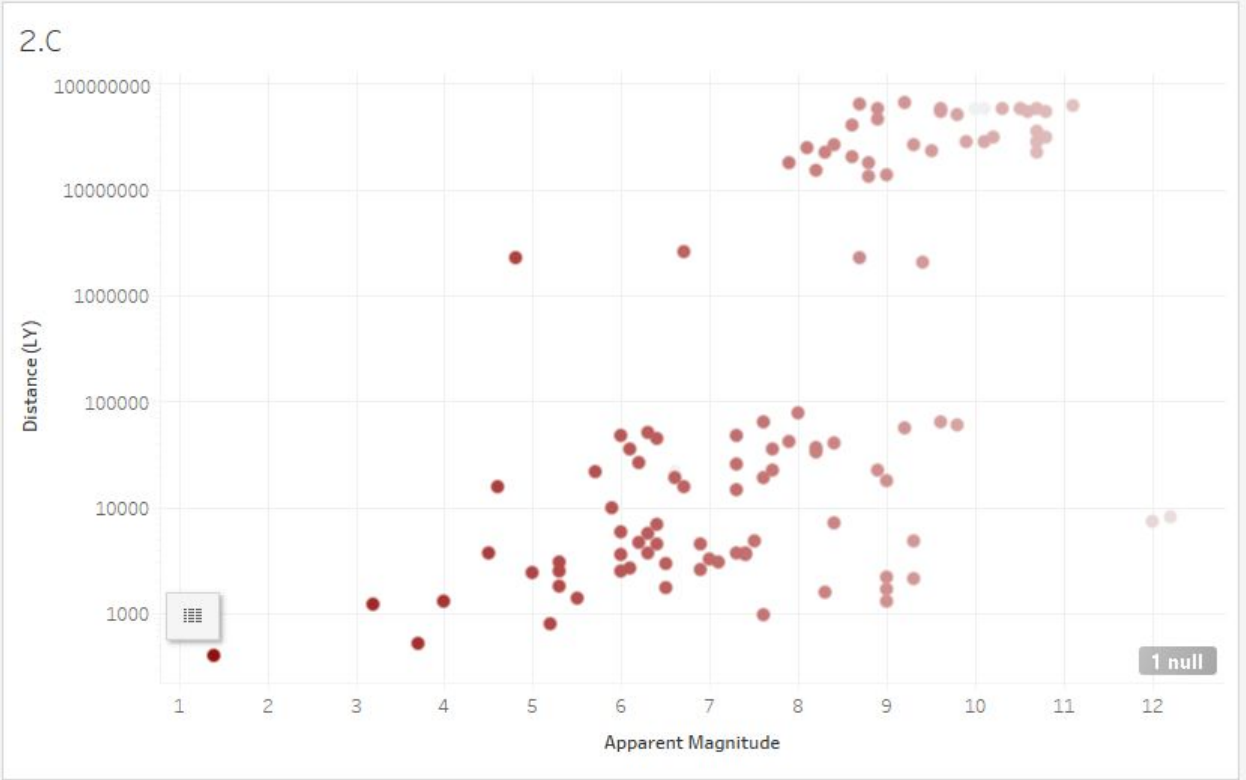
C) Create a scatter plot with the distance to the Messier objects plotted against their Apparent Magnitude (it's their visual magnitude, a measure of how bright they are in the sky). Note that these values may be… backwards from what you would think. The higher the number the fainter the object is in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

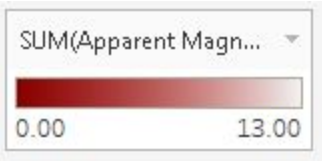
Ans:

Here, I have plotted the scatter plot with the distance to the messier objects plotted against their apparent magnitude. So basically it is their magnitude of the objects

To be Continued

and a measure of how bright objects are, so you can see in the graph that brighter the object nearer it is and lighter the object farther the object which you can clearly see from the below graph.



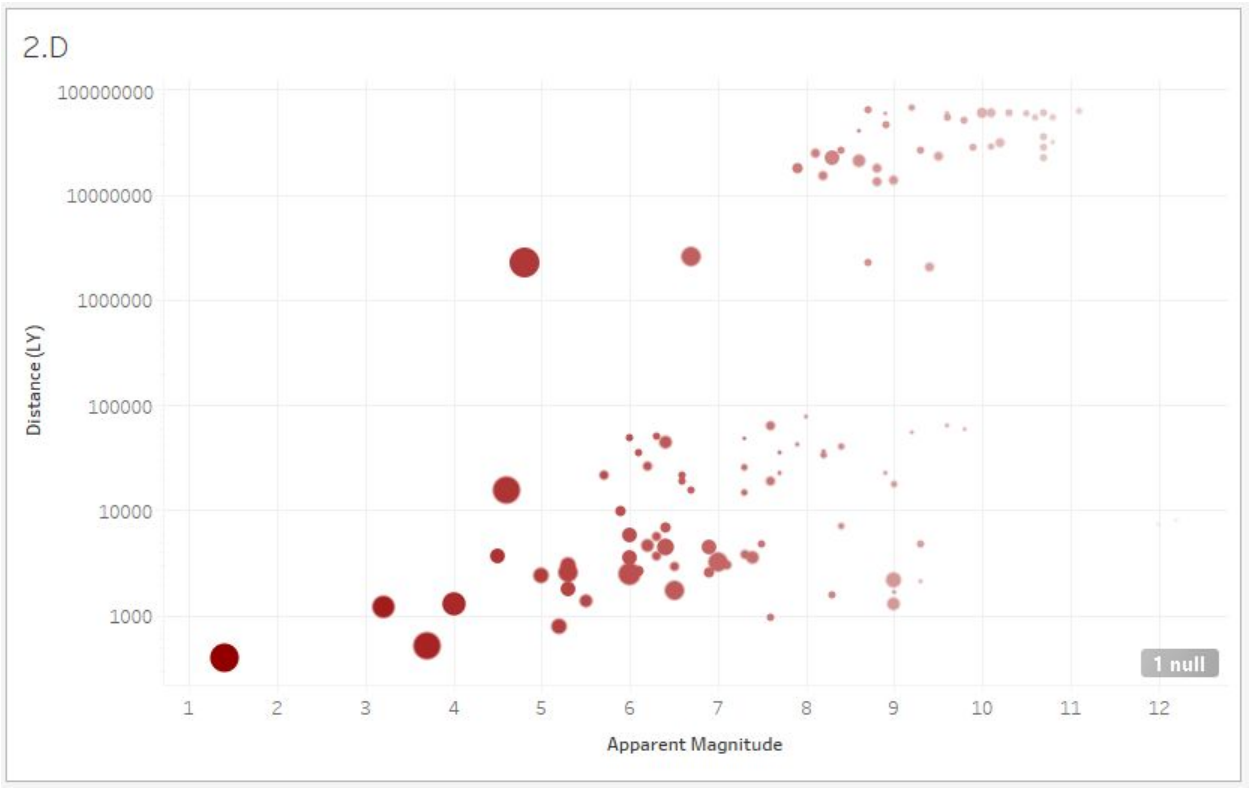


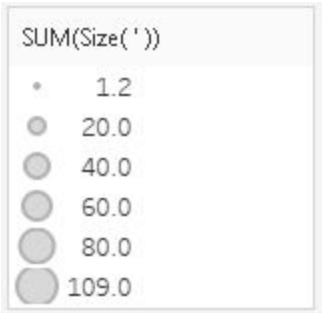
D) Augment the visualization in (c) by adjusting the size of the points in the scatter-plot based on the angular Size of the objects in the sky. Evaluate how easy it is to analyze all encoded aspects of the data from this graph and give a suggestion on how you might modify the graph to display all this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

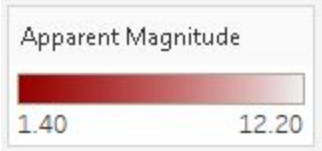
Ans:

From the below graph, you can see that I have selected the apparent magnitude vs distance of the objects. I have used the scatter plot to show that. From the graph it is really easy to evaluate the angular size of the objects in the sky. I have also provided an information graph along with the scatter plot. For the angular size of the objects I have used different shapes of the circle you can find the size of the circle from the second graph. And for apparent magnitude I have used color shading that also you can find in the graph.

To be Continued







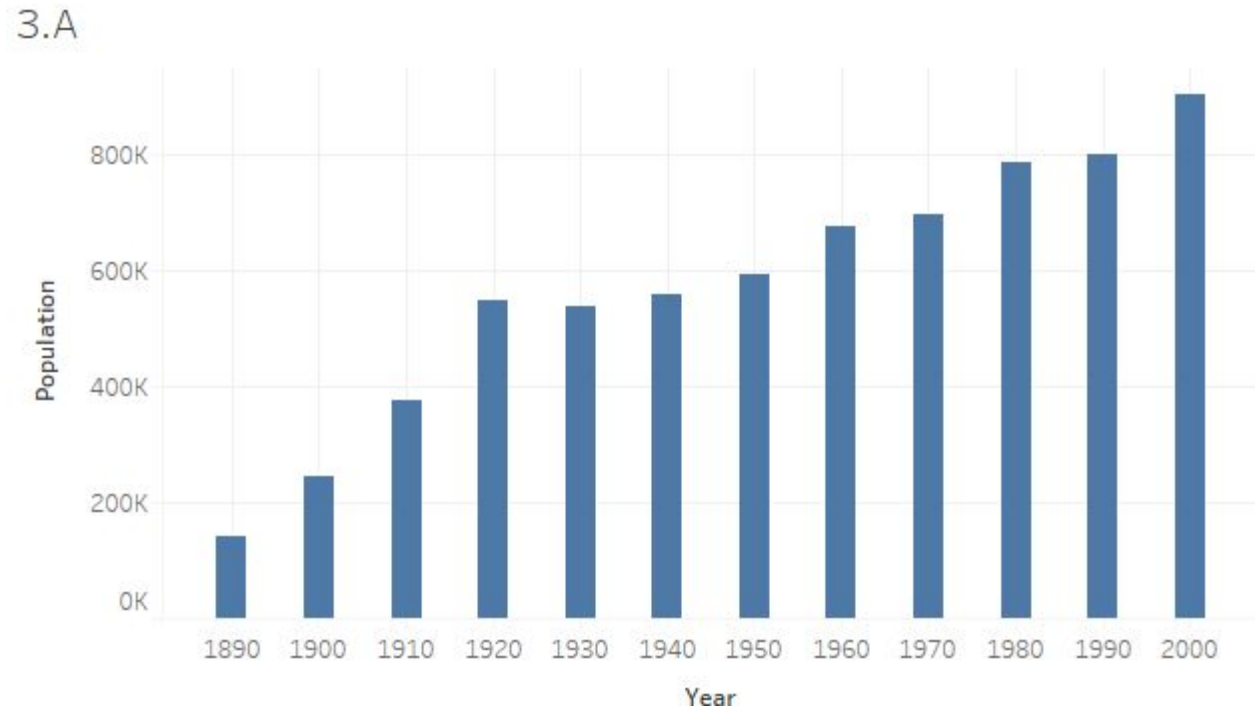
To be Continued

Question 3

A)How many times has the population doubled since 1890?

Ans: From the graph we can say that the population doubled question is a little ambiguous so if the question is that the population doubled in a decade next to the last decade then the answer is “No”. And the answer is for overall time that is “three times”.

So if we compare with previous decades then no doubled but for overall graph and different decades then it doubled two times.



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?

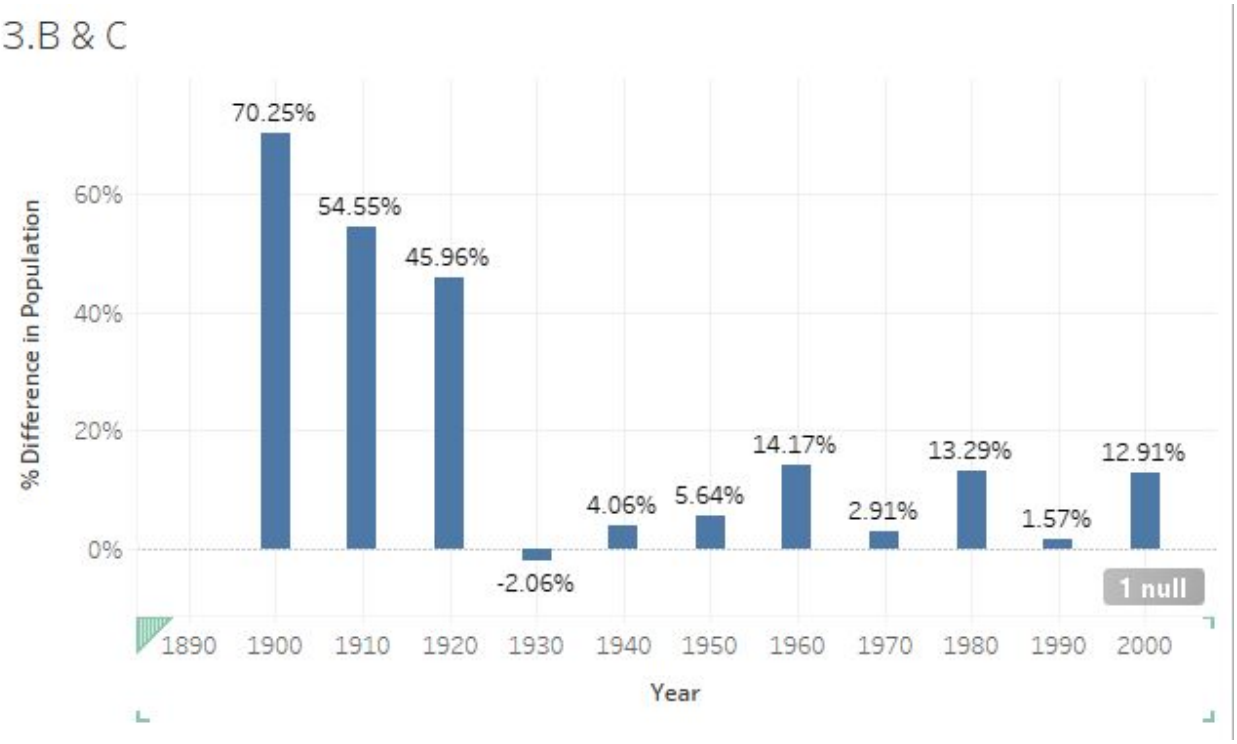
Ans: Yes, the percent rate of change in the population has increased over the years. It decreased only in the one decade which is 1930.

C) What years was the population percentage increase greater than 15%?

Ans: In 1900, 1910 and 1920 population percent was greater than 15% which we can say that, by graph from the next page.

To be Continued

3.B & C

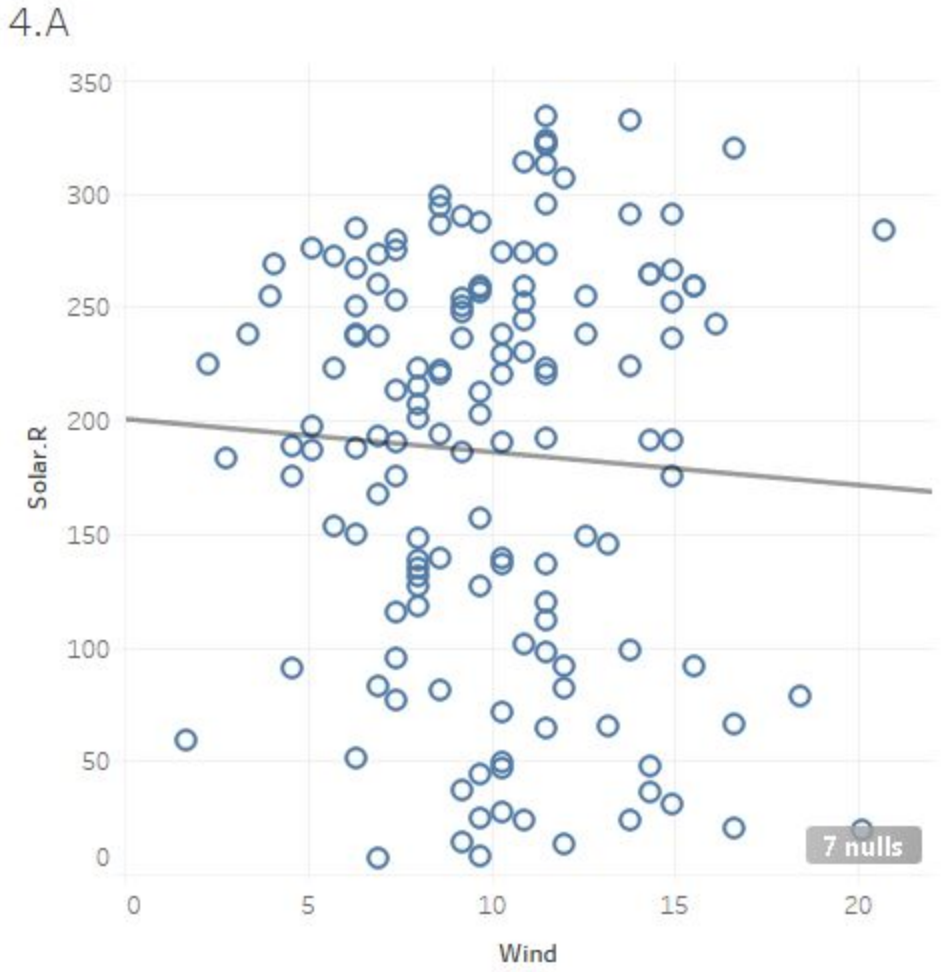


To be Continued

Question 4

A.)Use a scatter plot 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. 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).

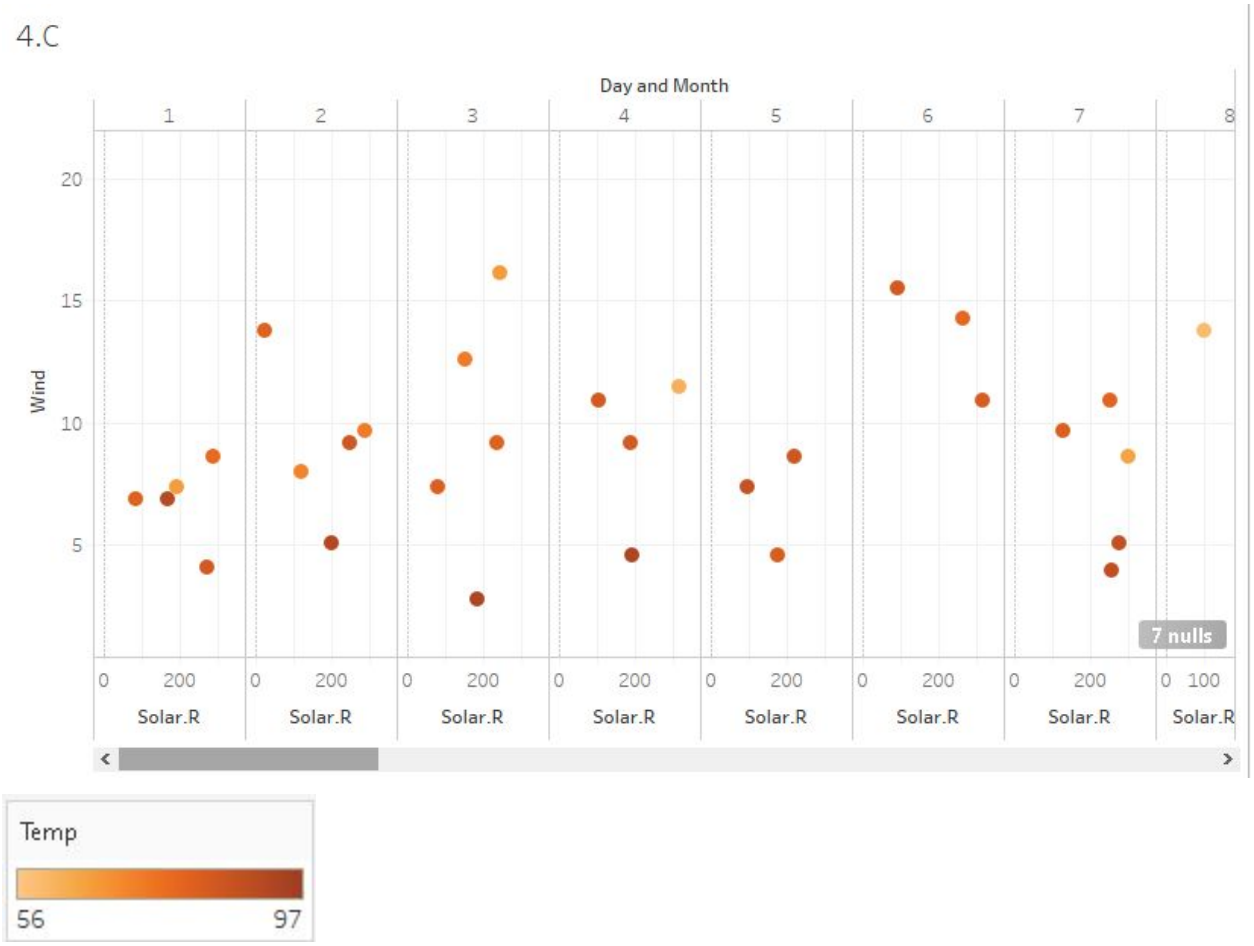
Ans: FROM THE SCATTER PLOT, we can say that the relationship between wind and solar is weak. You can see the graph is not strong and not strongly explaining any relationship. It is not linear so it's hard to say the effect of the wind on solar or vice versa. Because even when wind is high there is less and high solar radiation and vice versa. So we can come to the conclusion that there is not a strong relationship between wind and solar.R.



To be Continued

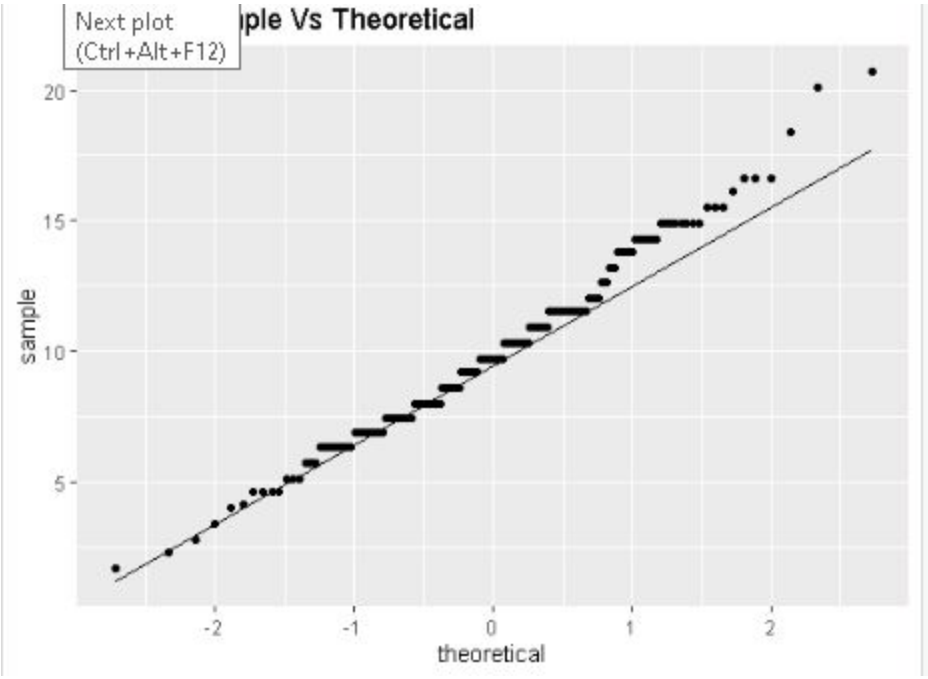
C) Finally, show these distributions in context of the rest of the variables by using a technique for comparing multiple distributions. Note: you will need to transform the data in a particular way that we have studied. It showed in the Tableau tutorial and in an R tutorial. Hint – you need to collapse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

Ans: From the graph, you can see that I have shown wind and solar on particular days of the month so you can see different values on different days although there is no strong relationship between both quantities . With that I have also put temperature so you can see different temperature with intensity on each day.

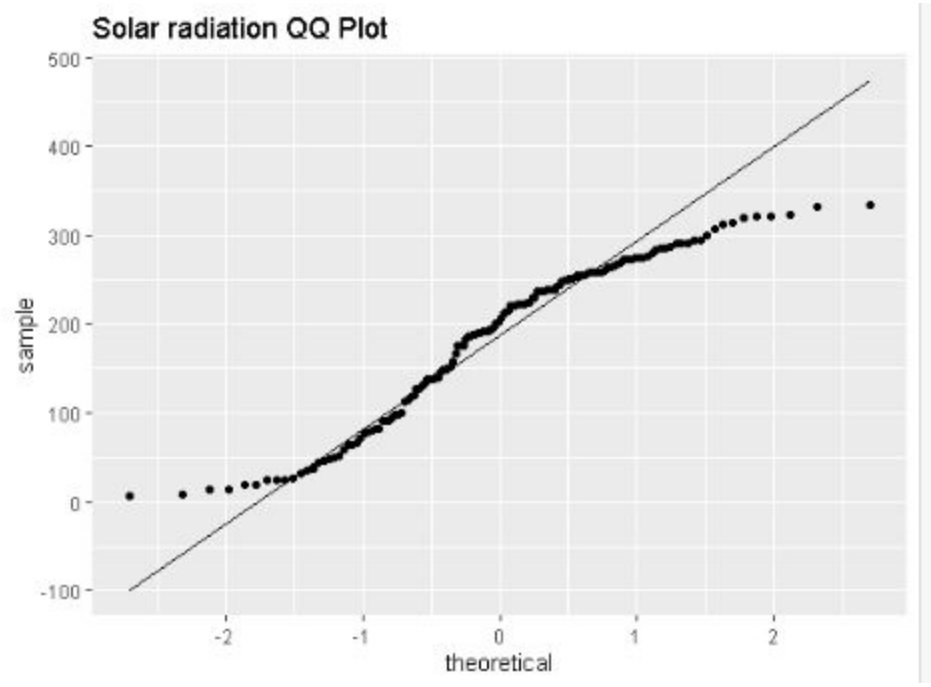


To be Continued

D) For extra credit, compare Wind and Solar.R again with a QQ plot. What does this tell you?
Ans: FROM THE Q-Q PLOT, we can say that the wind is not normal, there are steps in the graph. So we can infer that graph is not normal.



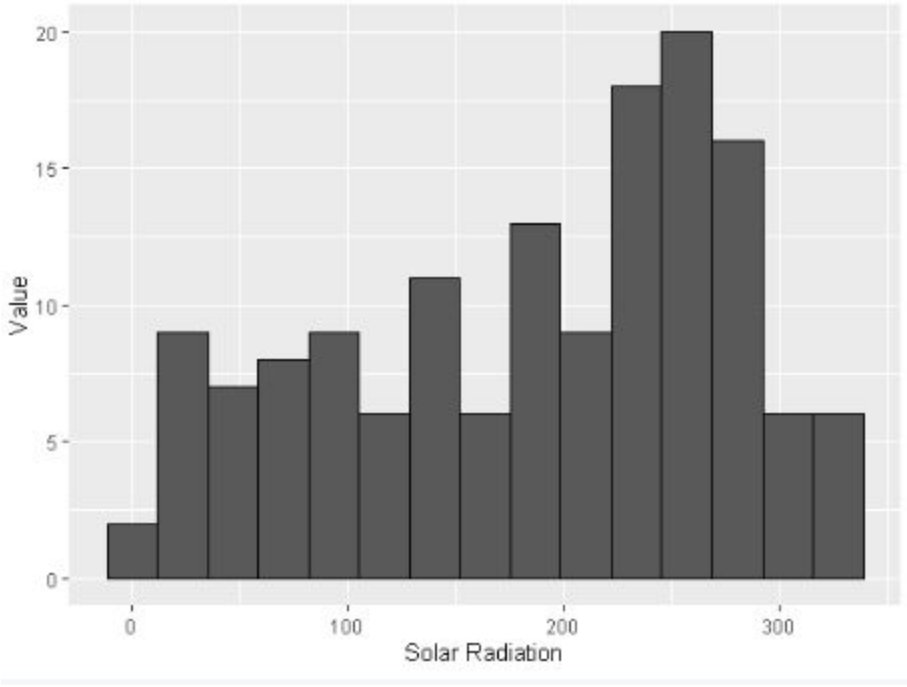
FROM THE Q-Q PLOT, we can say that the solar is not normal, there are s curves and steps in the graph. So we can infer that graph is not normal.



To be Continued

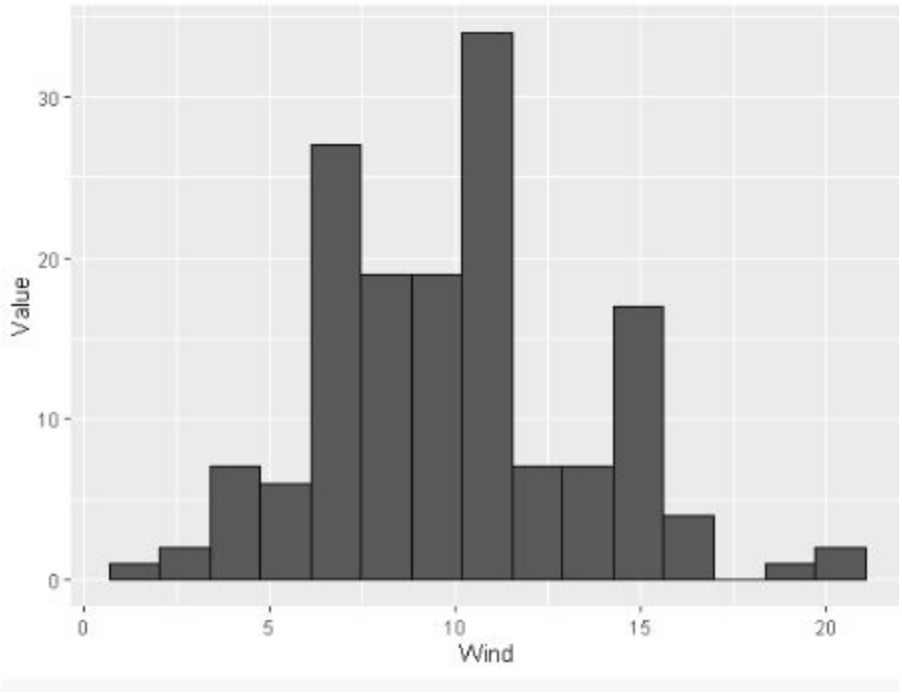
B.)Use a plot that will show the distributions of Wind and Solar. R and allow you to compare with fine detail.

Ans: From the graph, you can see that from the below bar graph that I have shown a plot of days vs solar radiation. So most of the day(Around 20 days) have solar radiation of 280. So we can say that, that is the common radiation.



From the below graph, I have shown the value of wind vs days so you can say that wind is 10 so which everyday. Where both extremes of the graph are shown rarely on any day of the month.

To be Continued



To be Continued