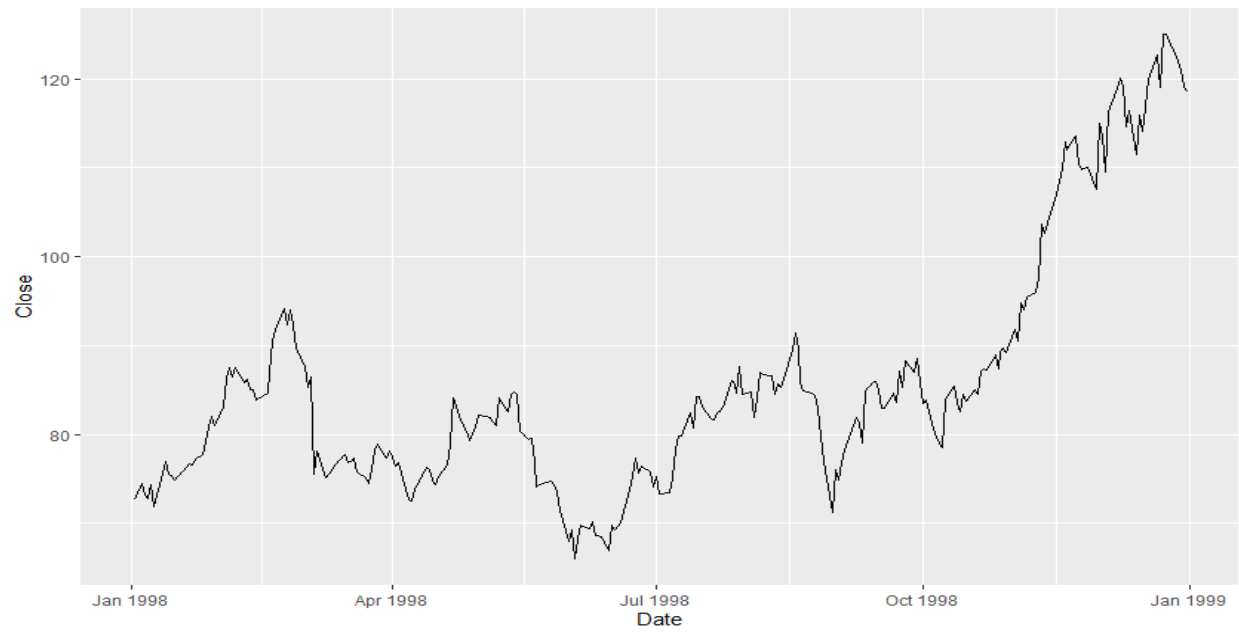


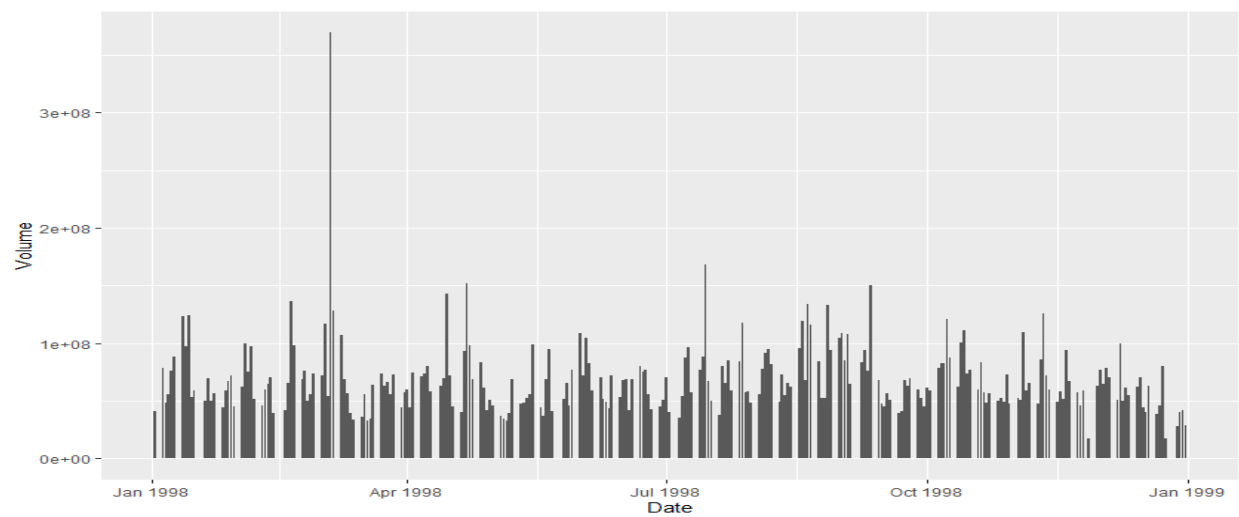
Name: Ashay Kargaonkar

DSC 465 Homework 1

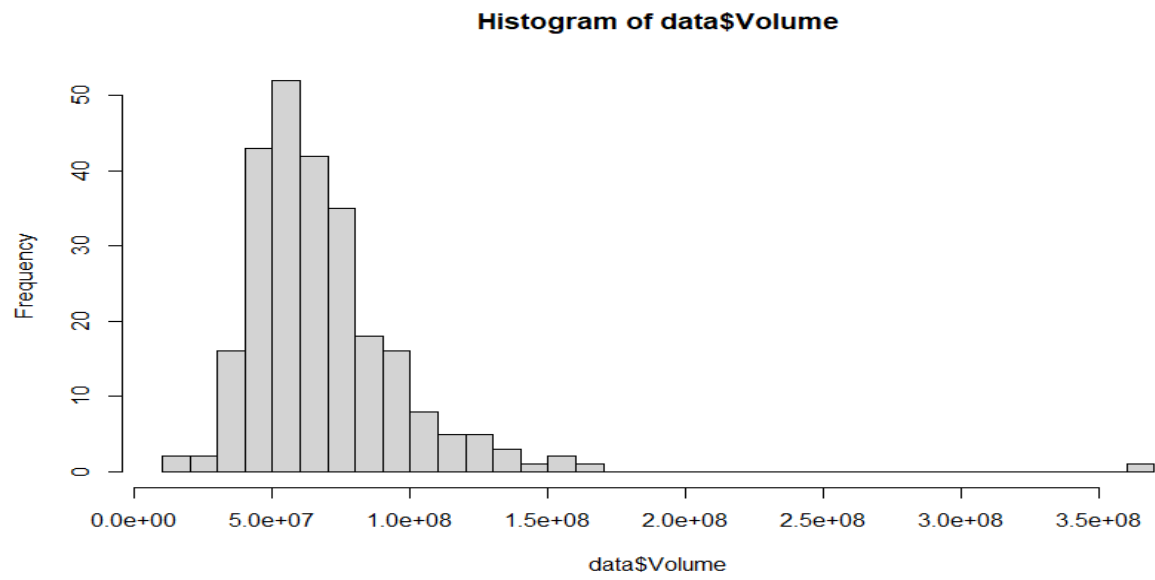
Q1.a.



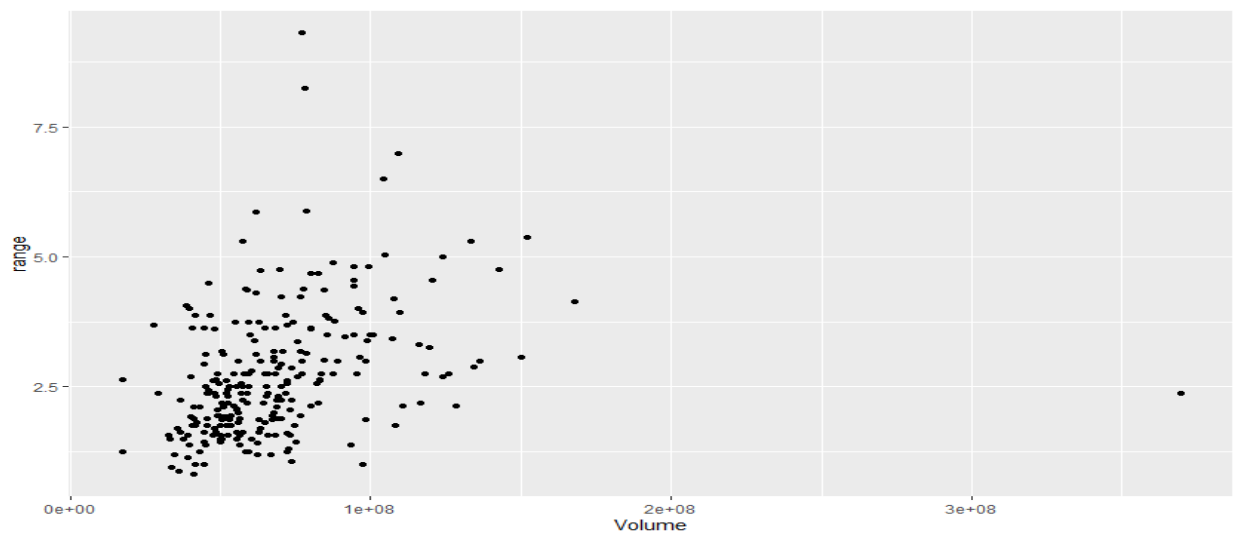
Q1.b.



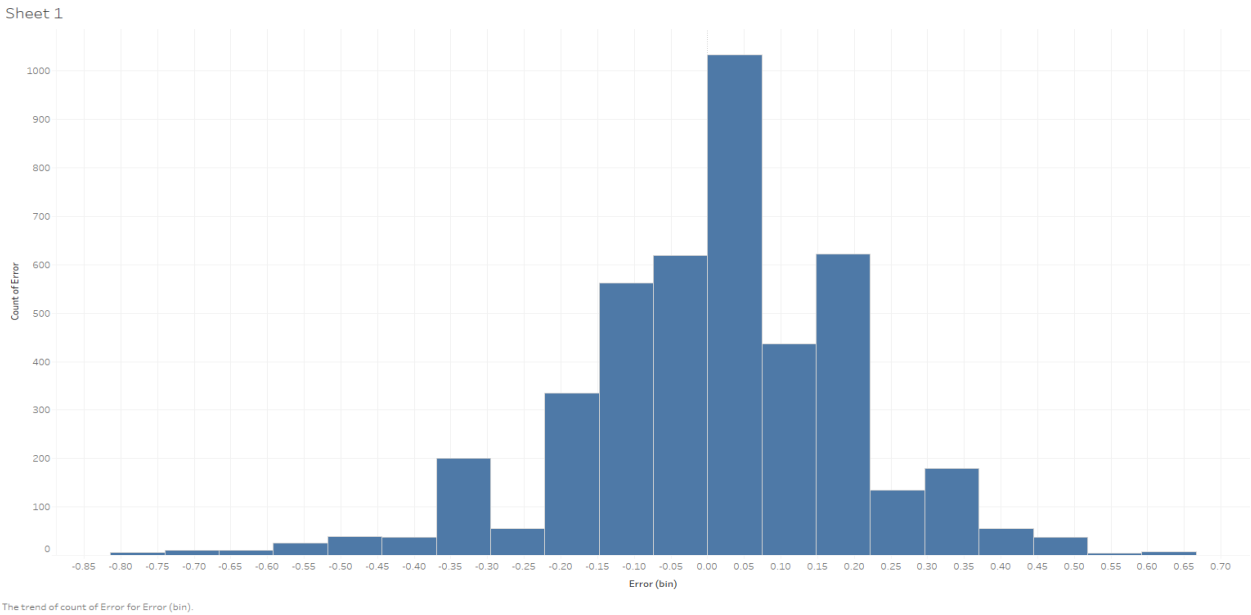
Q1.c.



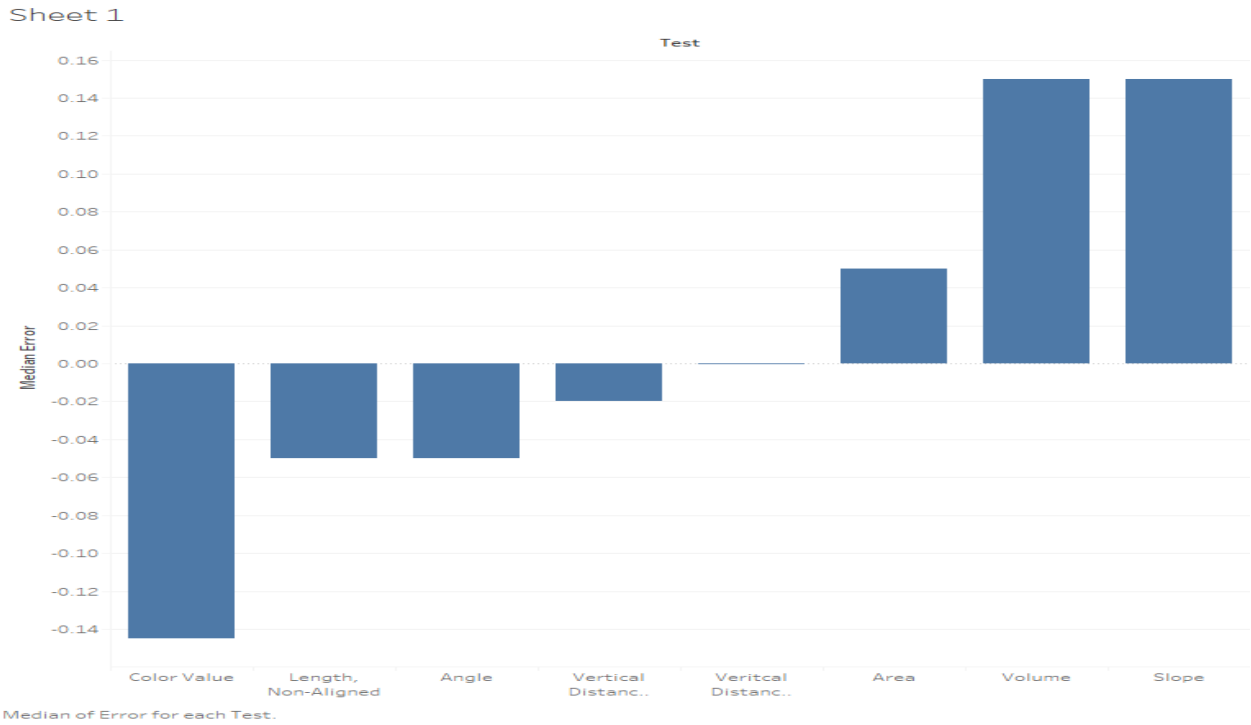
Q1.d.



Q2.a.

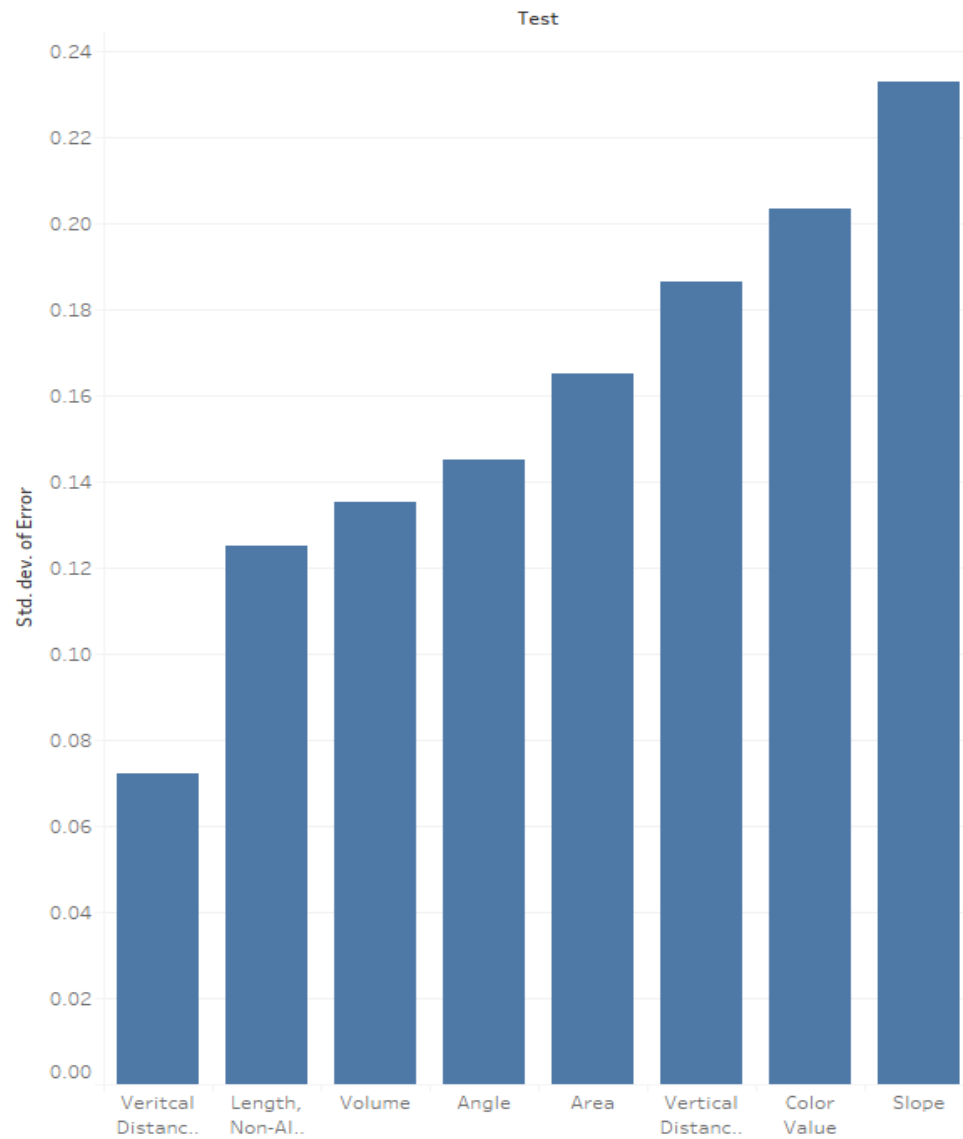


Q2.b.



Q2.c.

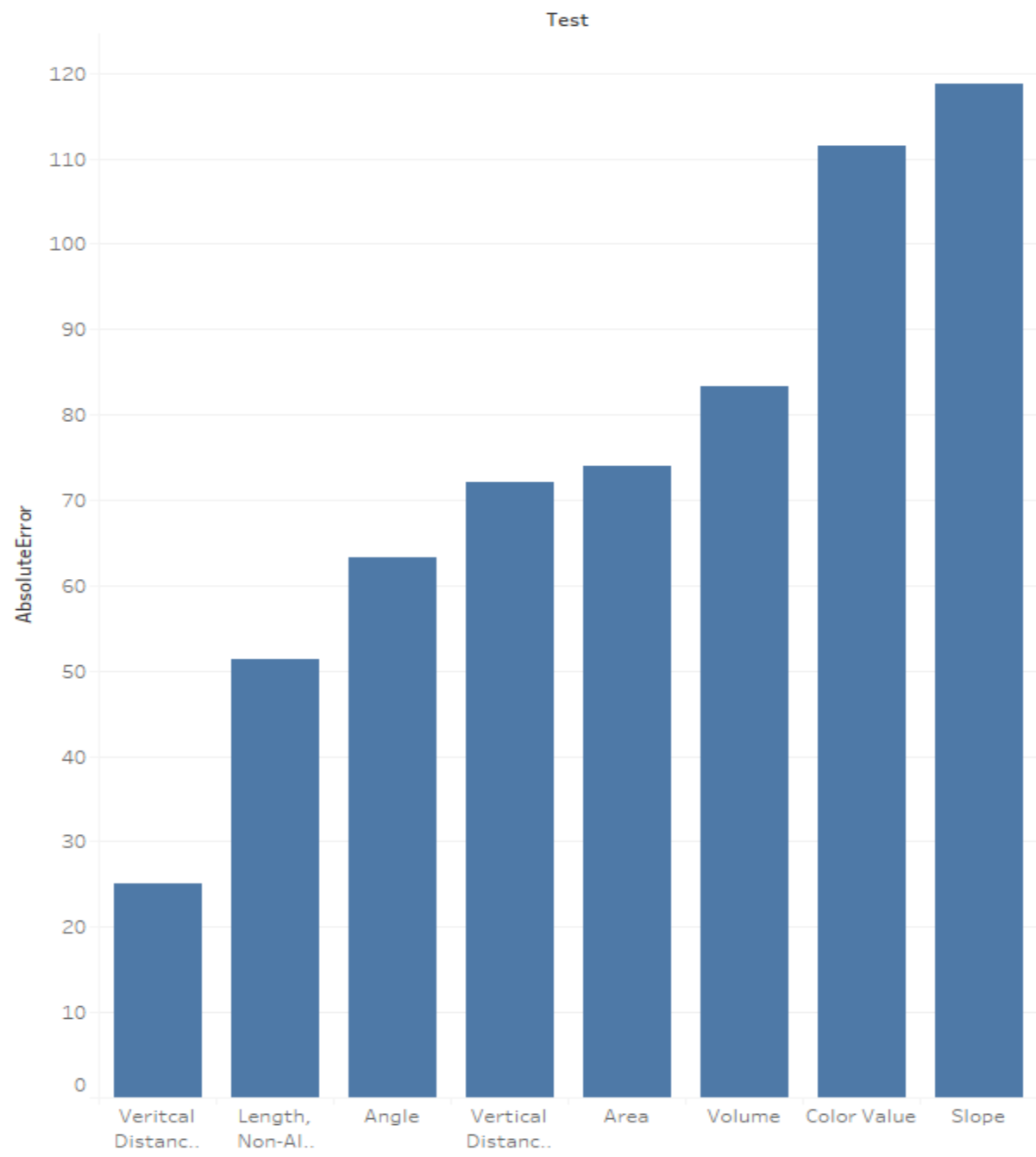
Sheet 2



Standard deviation of Error for each Test.

Q2.d.

Sheet 3



Sum of AbsoluteError for each Test.

Q2.e.

About Graph 2.a:

- Min error value is -0.8 and Max error value is +0.65
- This graph is normally distributed displaying the bell curve.
- Maximum count of error value is displayed for value between 0 to 0.5.

About Graph 2.b:

- Color value test has the most negative value of median of error.
- Slope value test has the most positive value of median of error. We can say that people face difficulty understand the different between the objects with these 2 features.
- As vertical distance has the least value of median of error which states that people can easily detect the difference shown in terms of vertical distance attribute.

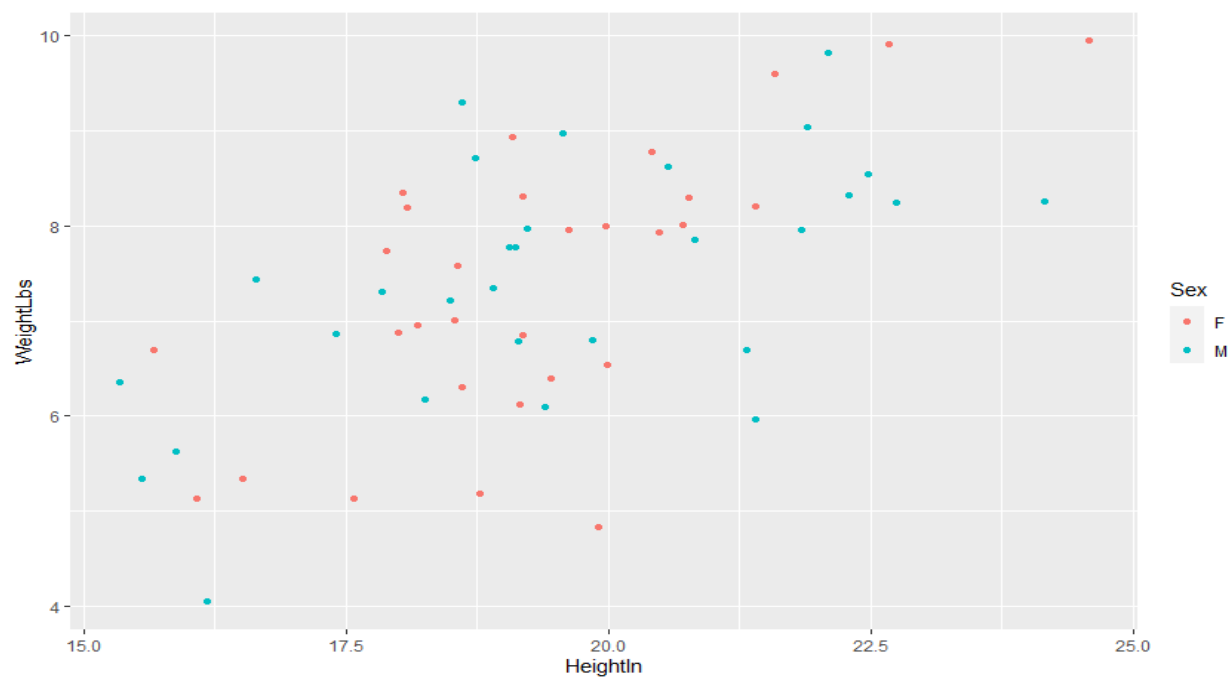
About Graph 2.c:

- From the graph we can see that vertical distance has the least value of Std. Dev. Of error.
- Slope test has the most.

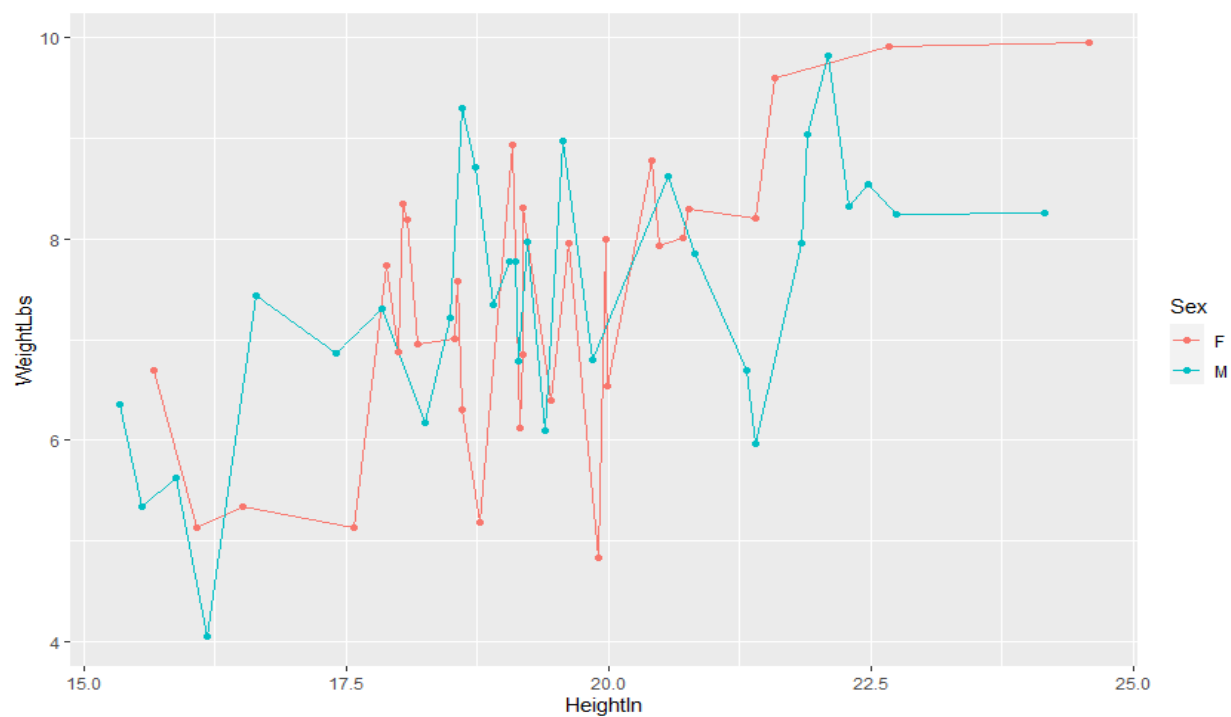
About Graph 2.d:

- Similar findings can be seen in this graph as the value of vertical distance is least and that of slope is most.
- If we want to generalize, we can say that people can detect the difference between the objects if there is difference in terms of vertical distance and people have difficulty to differentiate between the objects if they differ in terms of slope.

Q3.a.



Q3.b.



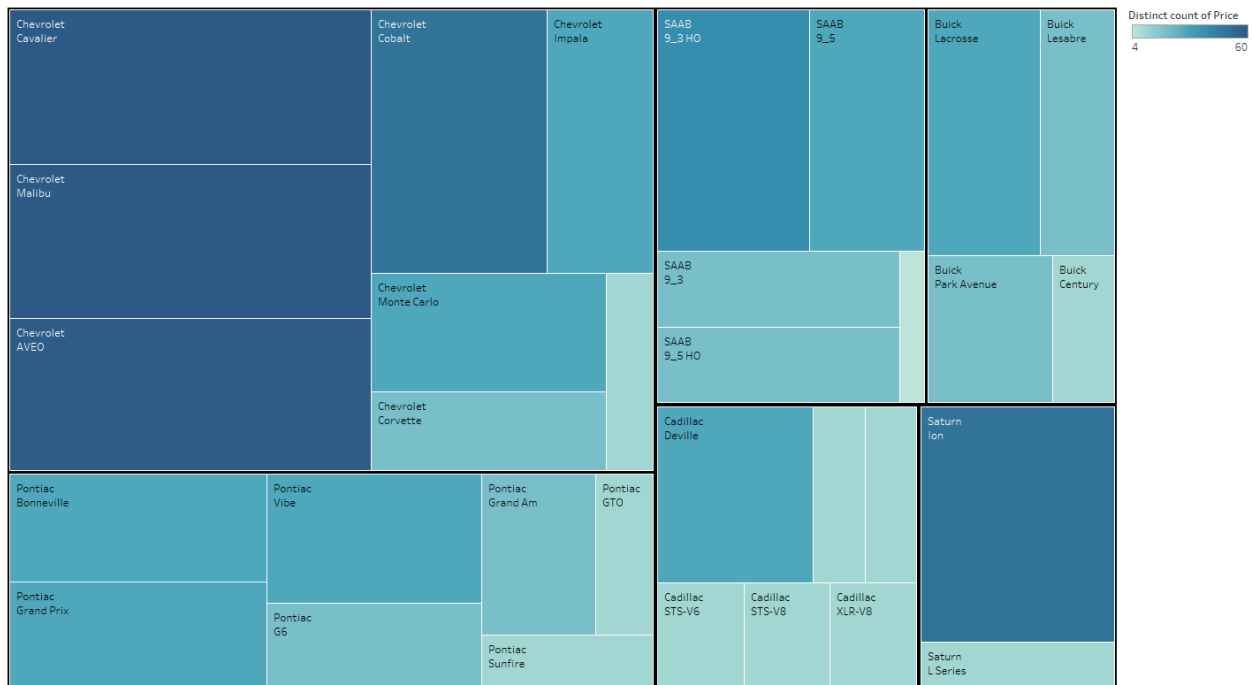
Q3.c.

- From the above graph we can see that the trend lines and scatter points stand out separately.
- For getting this graph I added `geom_line` and `geom_point` to get both the graphs on a single graph.
- Code used to get the above graph:

```
ggplot( data=data3, aes(x=HeightIn, y=WeightLbs, color=Sex) ) + geom_line(alpha = 0.5) + geom_point()
```
- I have also used `alpha` to decrease the thickness of the lines for clear visibility of points and lines.

Q4.a.

Sheet 1



Make and Model. Color shows distinct count of Price. Size shows distinct count of Price. The marks are labeled by Make and Model.

Q4.b.

Sheet 2



Make and Model. Color shows details about Make. Size shows distinct count of Price. The marks are labeled by Make and Model.

Q4.c.

- Sheet 1:
 - We can clearly see which make of the car based on model has the greatest number of distinct value of price. From the graph we can say that 3 cars of Chevrolet have the most number of distinct values of prices.
 - But here differentiating the count of prices between different makes and model is difficult.
 - Also, we cannot easily figure out which makes has a greater number of models.

- Sheet 2:
 - Here it is easily visible to get to know which make has a greater number of models.
 - Also, the size of circle says that which 'make + model' has more number as the size and color helps use to differentiate.
 - Here differentiating between different make and model based on the distinct count of prices is easy.
 - Also, here the text here is more readable and user-friendly, and seems better than the sheet 1.

Q5.a.

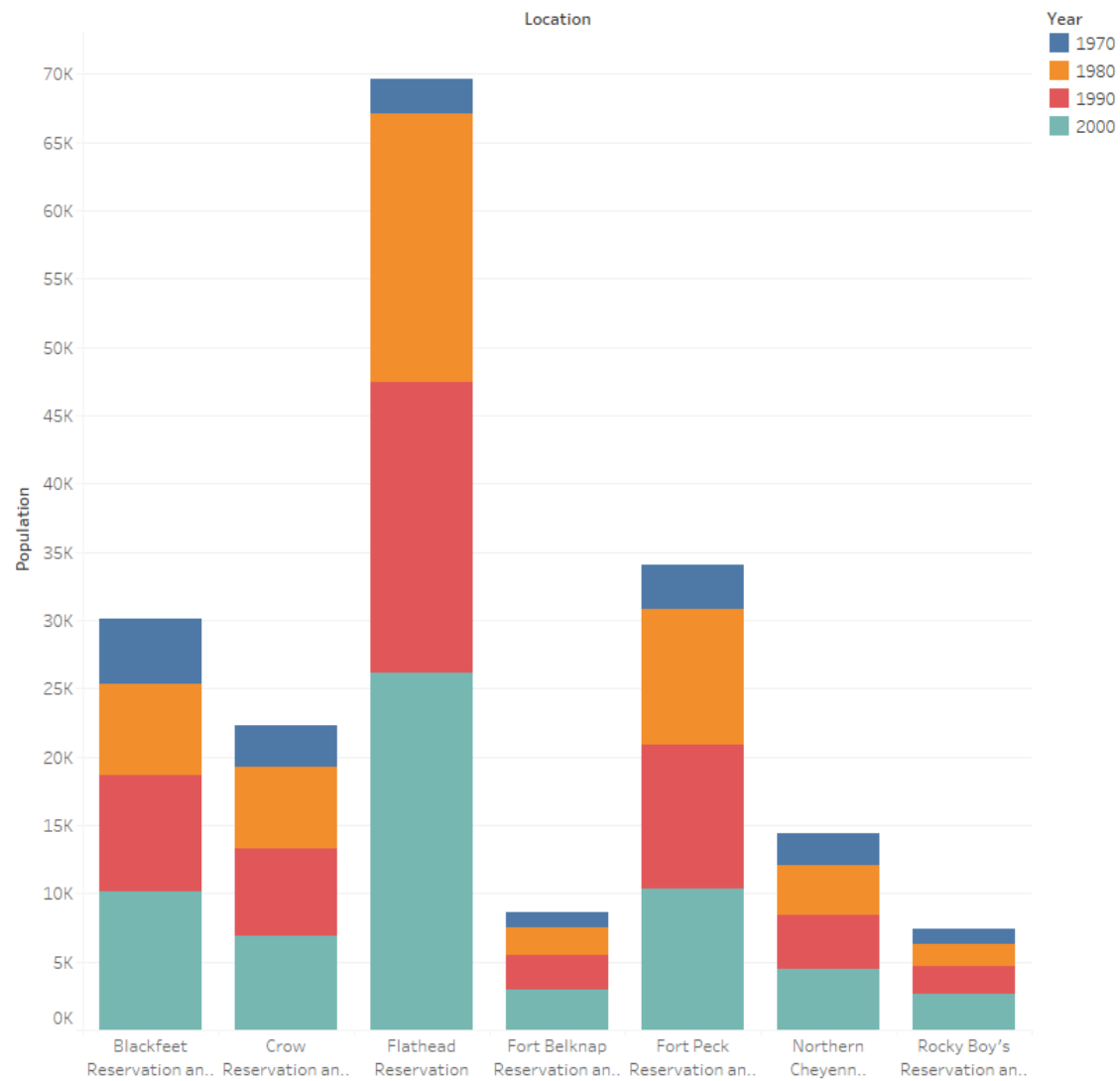
Sheet 1



Sum of Population for each Year broken down by Location.

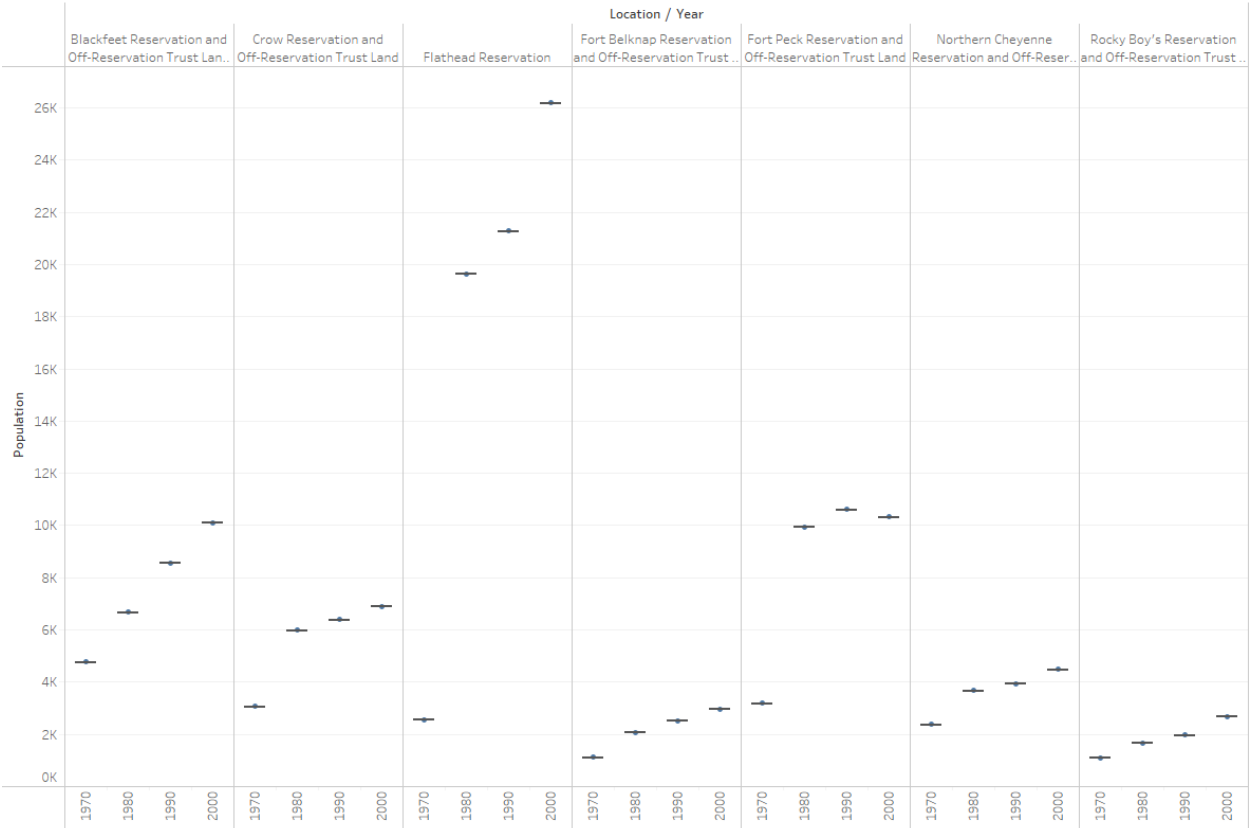
Q5.b.

Sheet 1



Q5.c.

Sheet 3



Sum of Population for each Year broken down by Location.