

DSC 465 Spring 2021

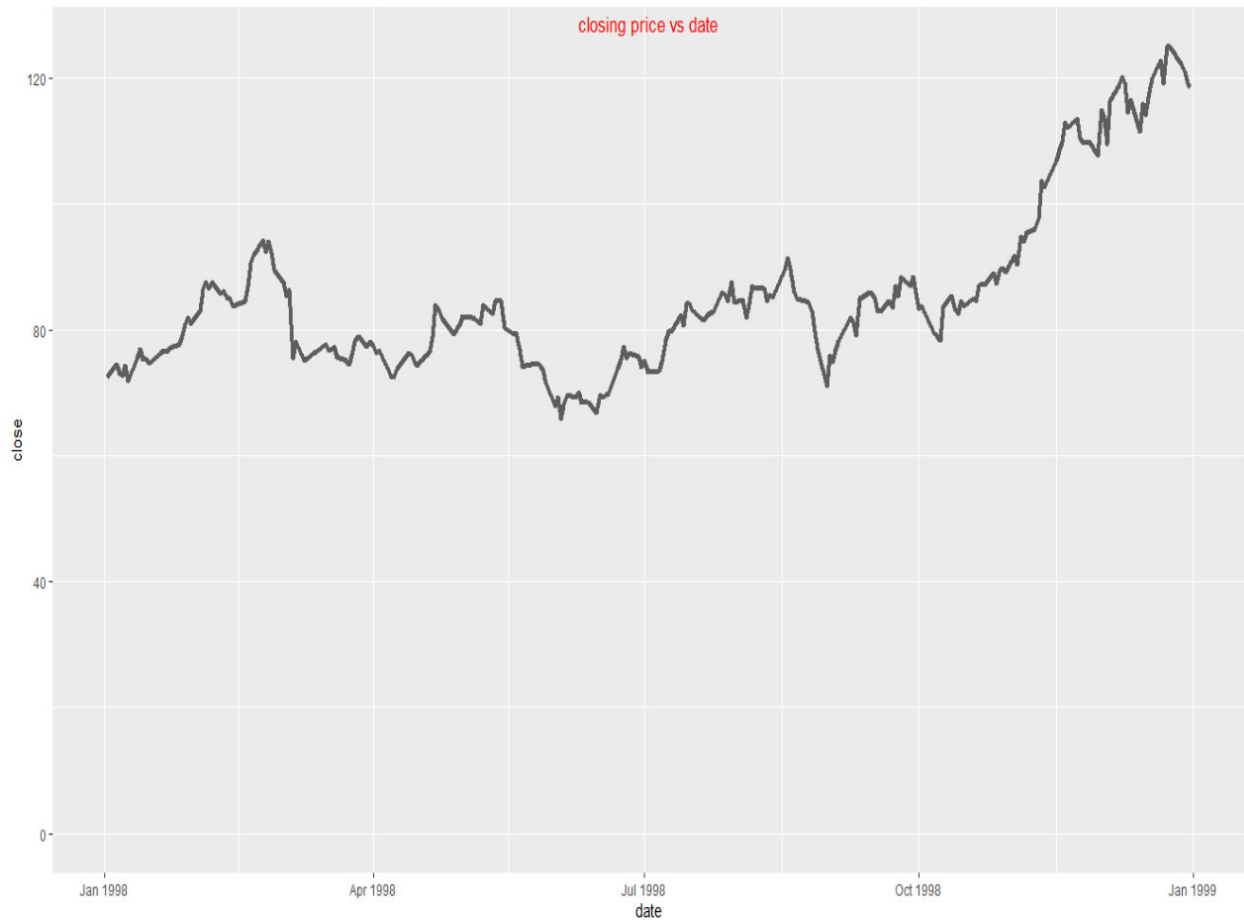
Homework 1

Student Number: 2020637

1. Intel stock

a. closing price vs. the date:

- In R:

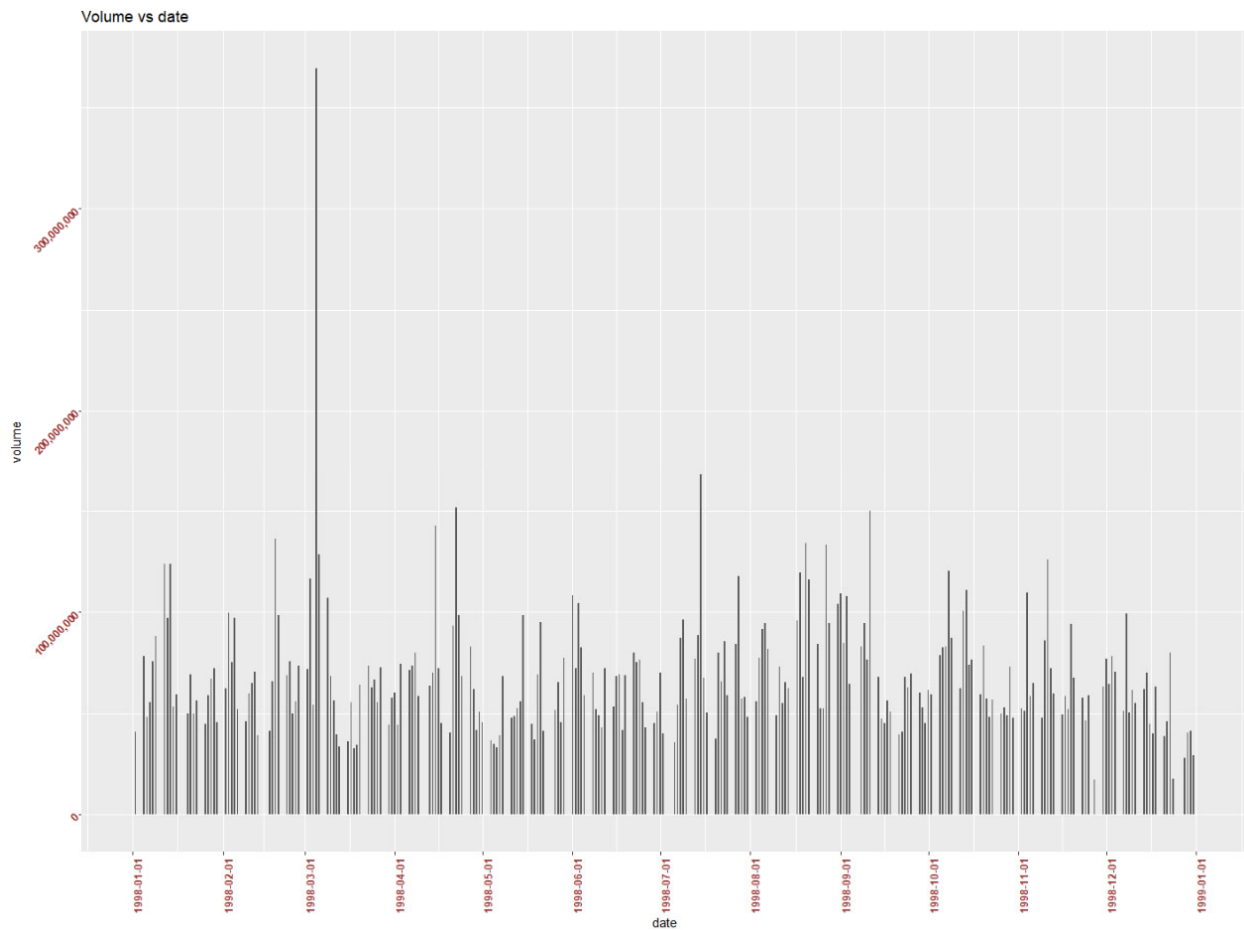


- In Tableau:

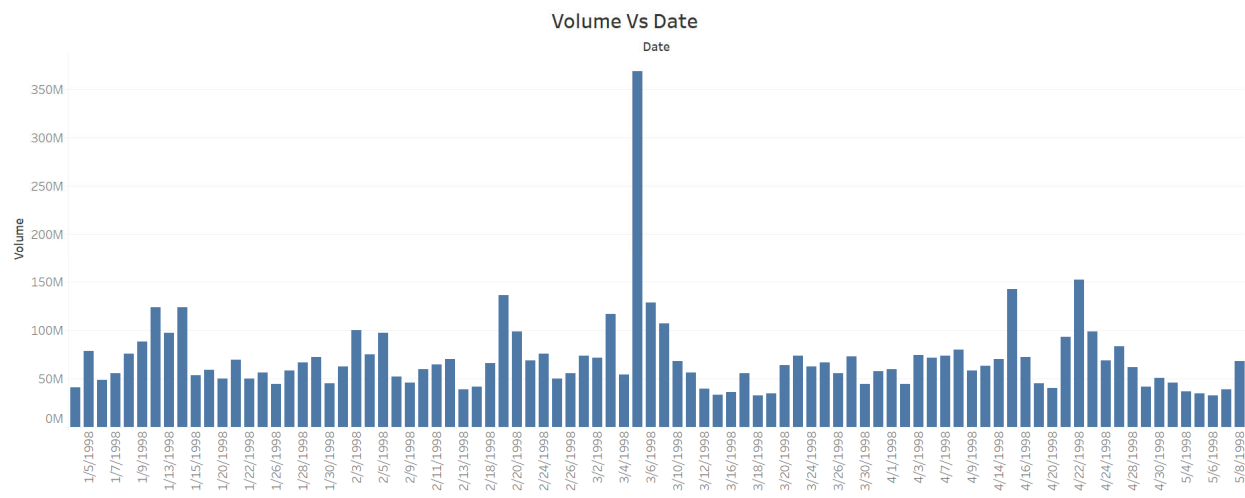


b. Volume vs. the exact date

- In R:



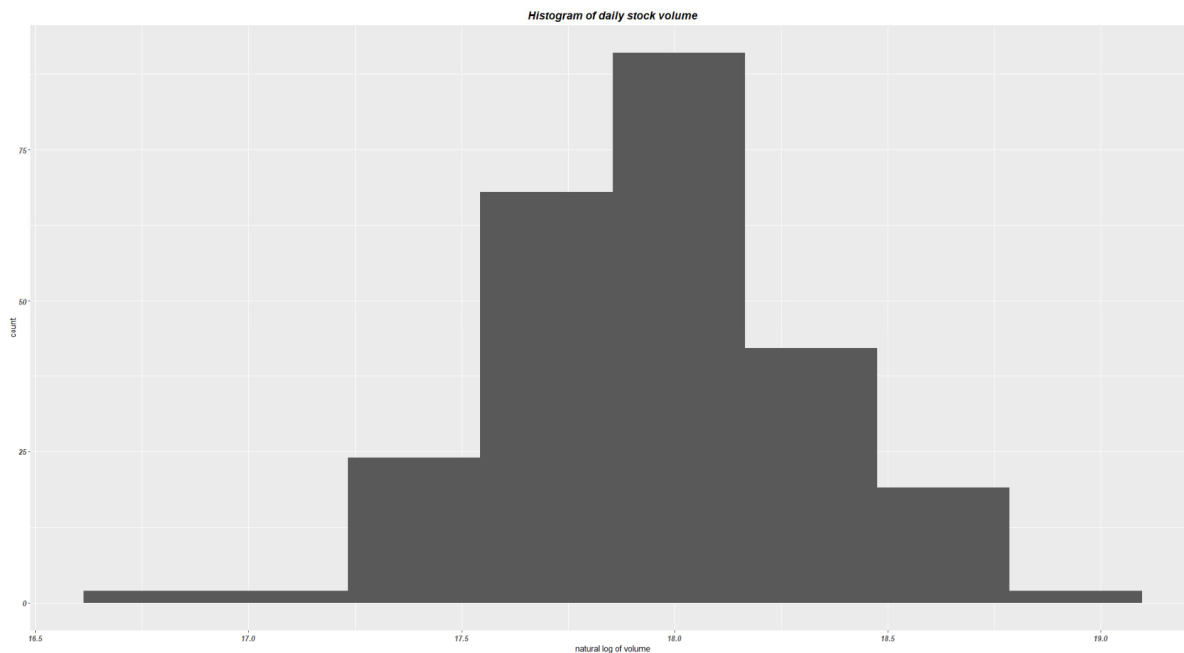
- In Tableau:



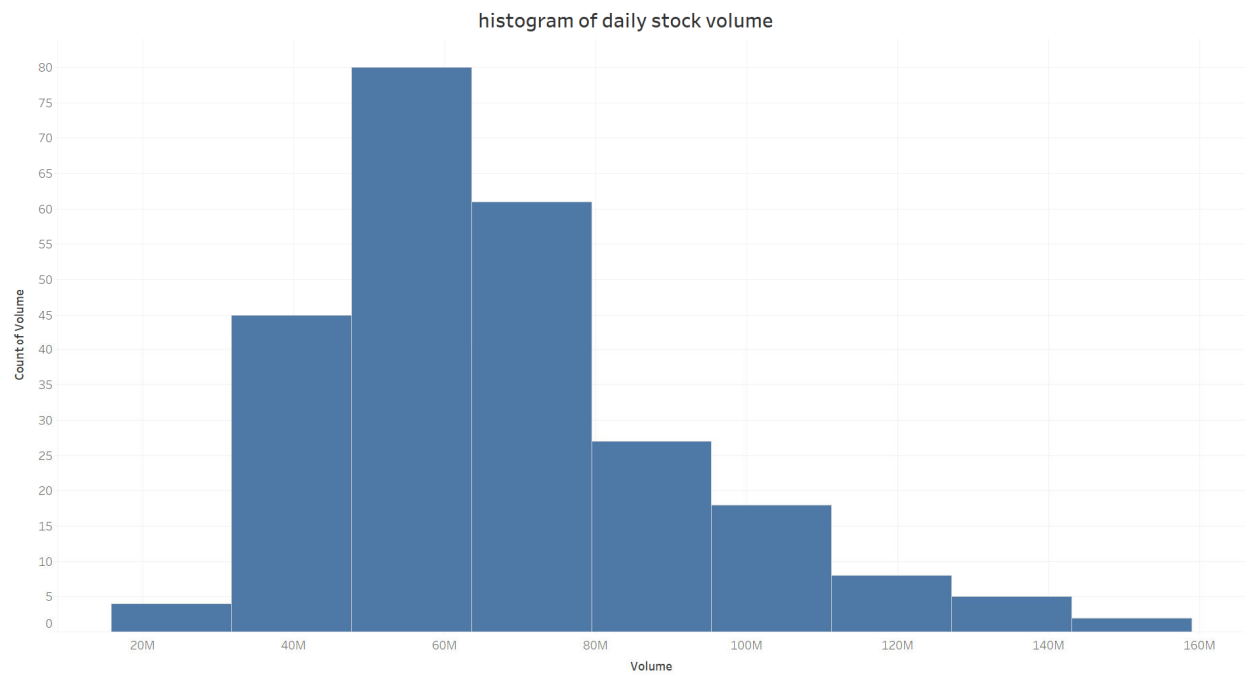
c. Daily stock volume histogram

- In R:

I created a filter of volume which I applied in my sub selection to create a new data set with the volume of less than or equal to 151875200 then derived a histogram with bins of $\log_2(252) + 1$



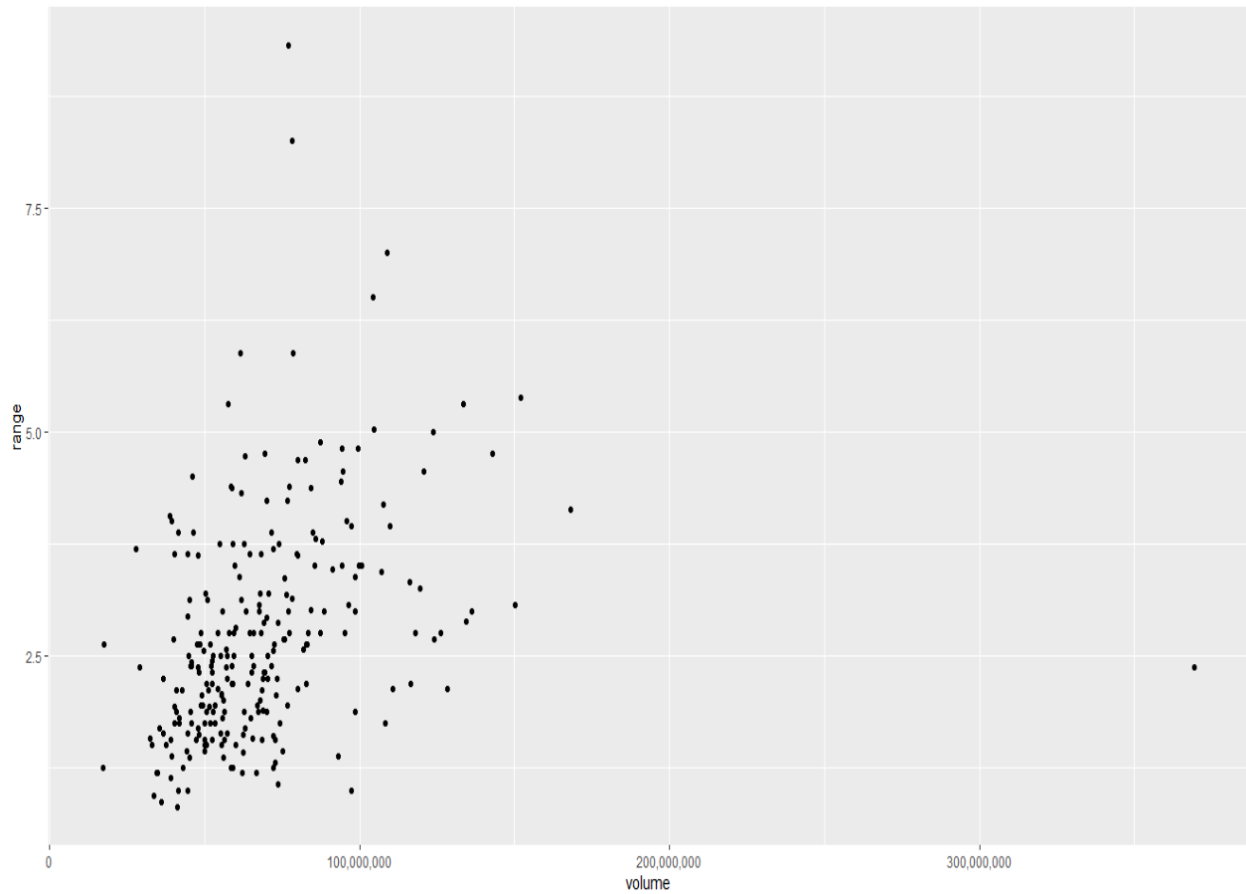
- In Tableau:



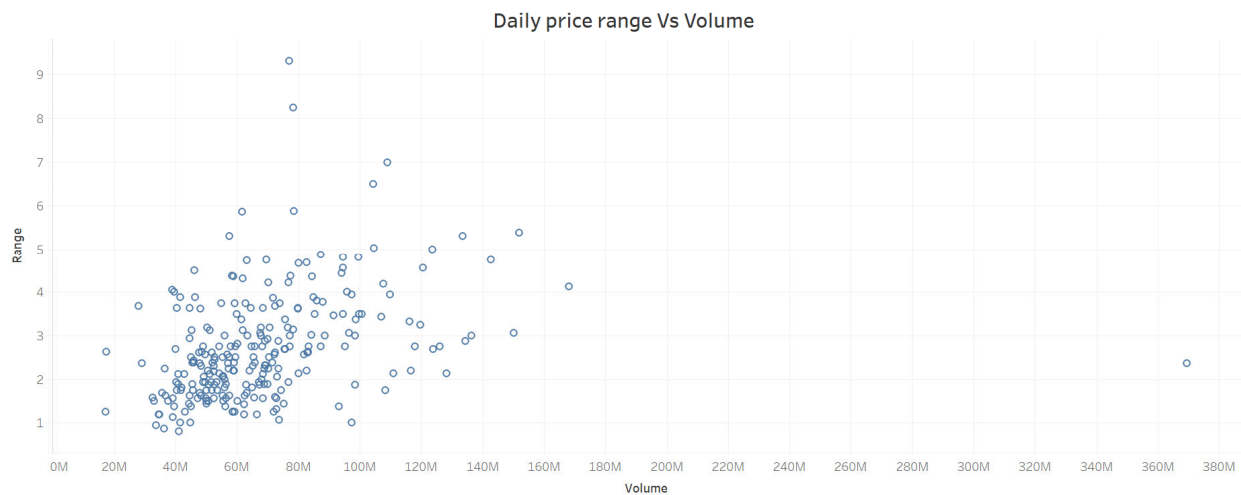
d. Scatter plot of Volume vs daily price

- In R:

Scatter plot of daily price range vs Volume



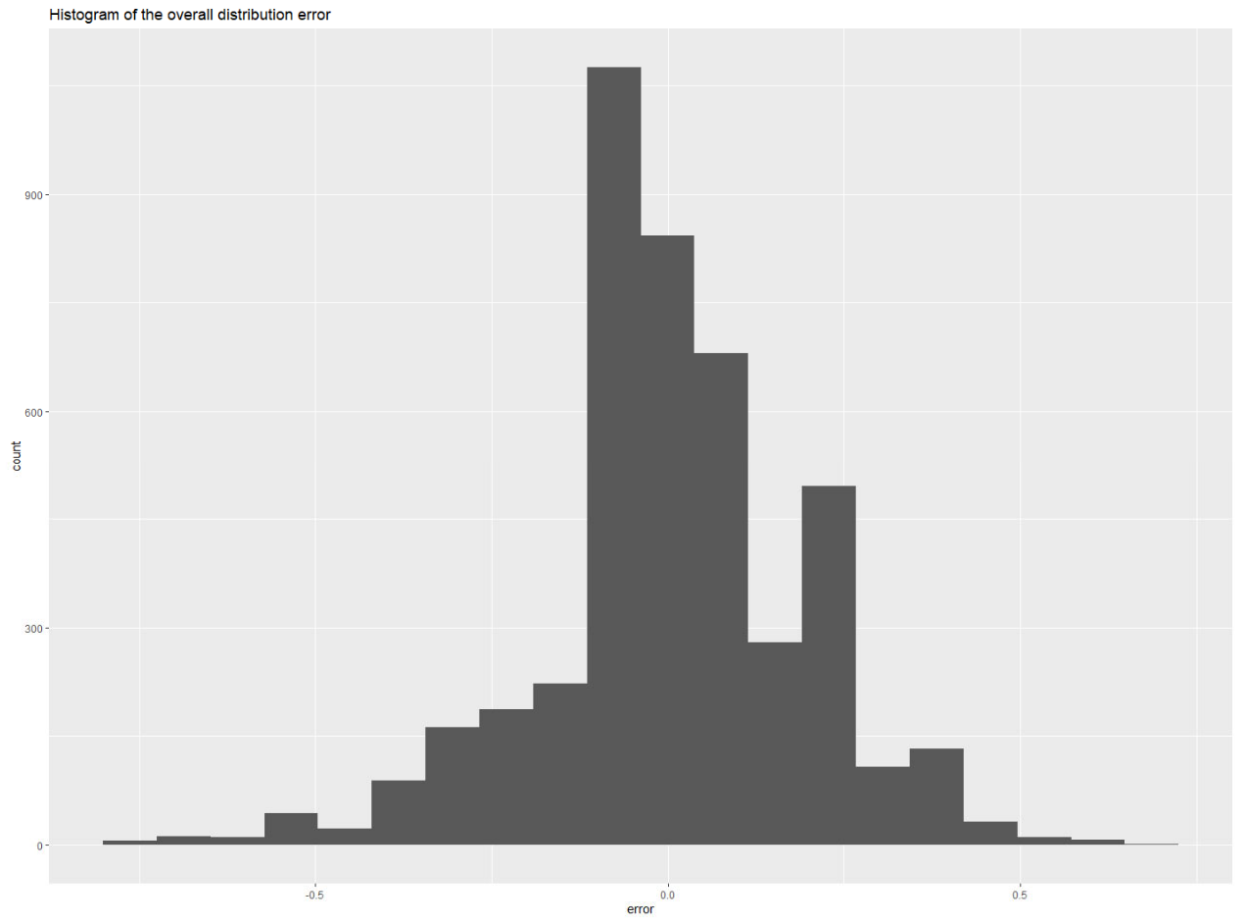
- In Tableau:



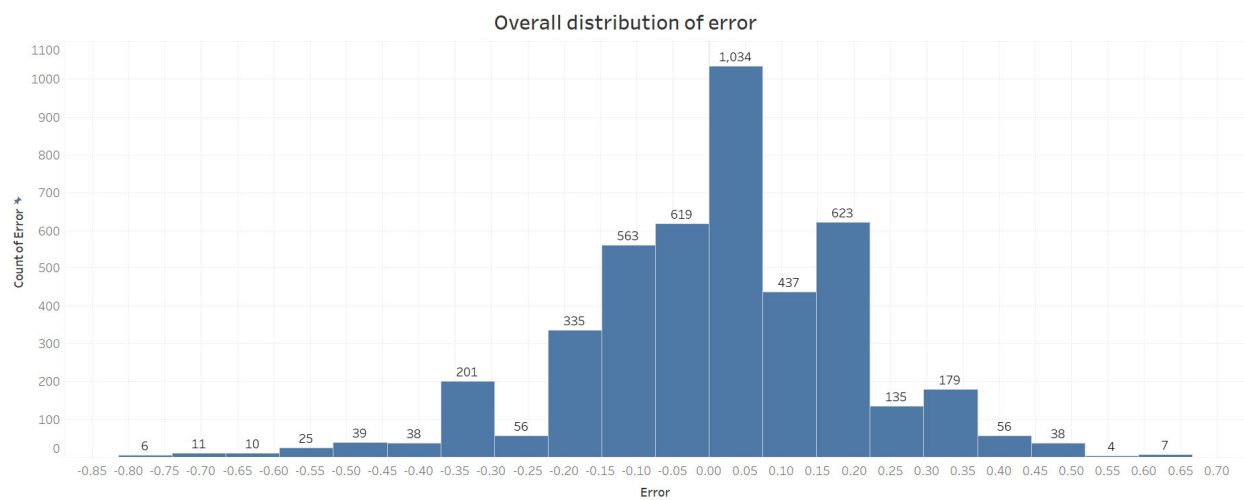
2. Perception Experiment data set

a. Histogram of the overall distribution of Error

- In R:

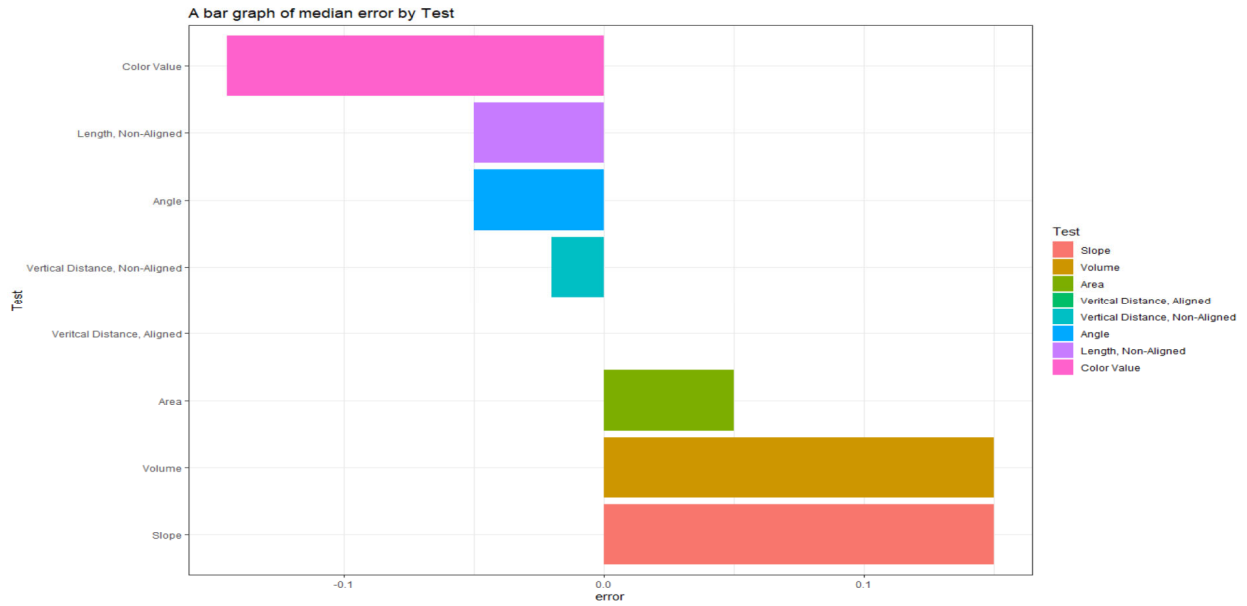


- In Tableau:

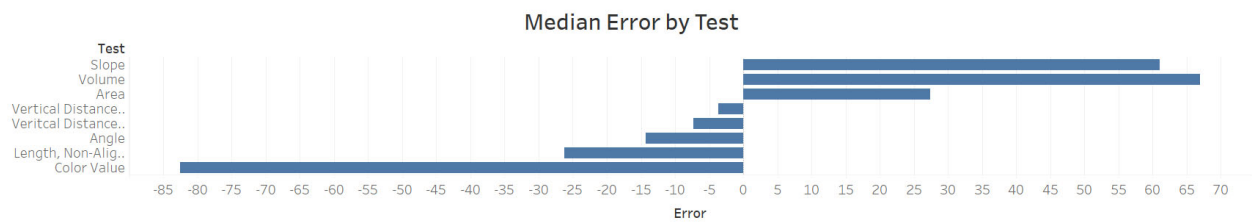


b. Median Error by Test

- In R:

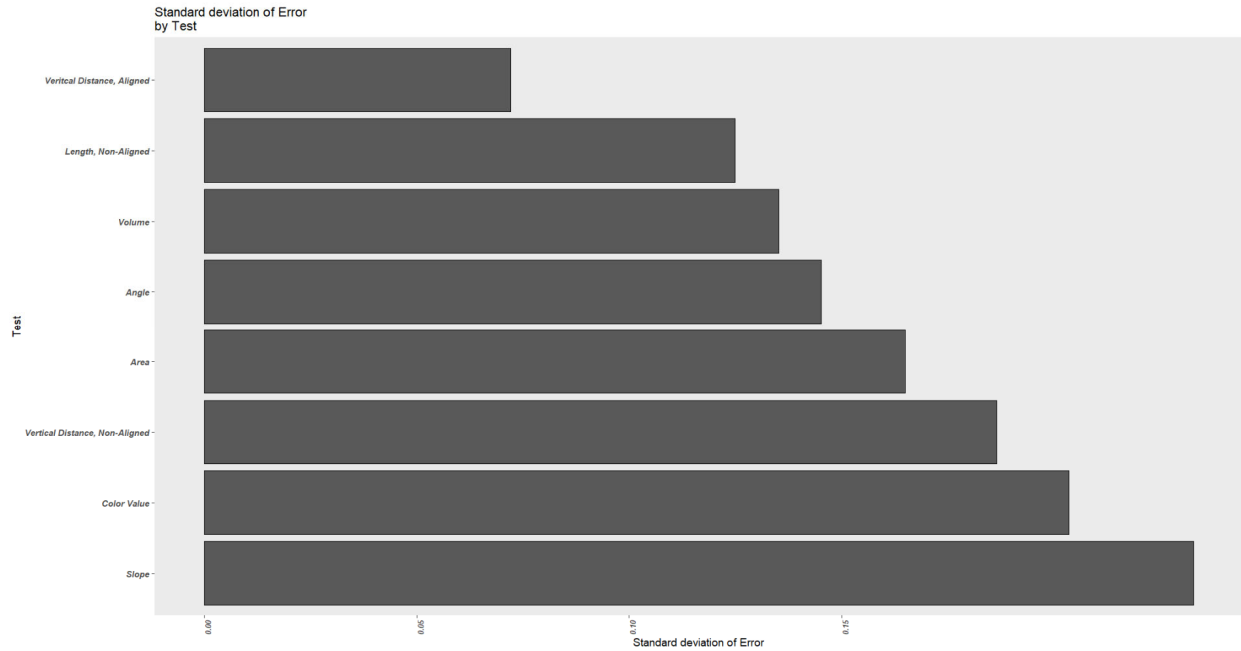


- In Tableau:

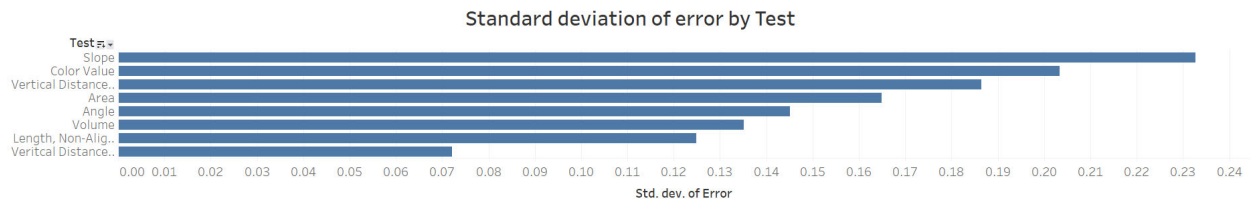


c. Standard deviation of error by Test

- In R:

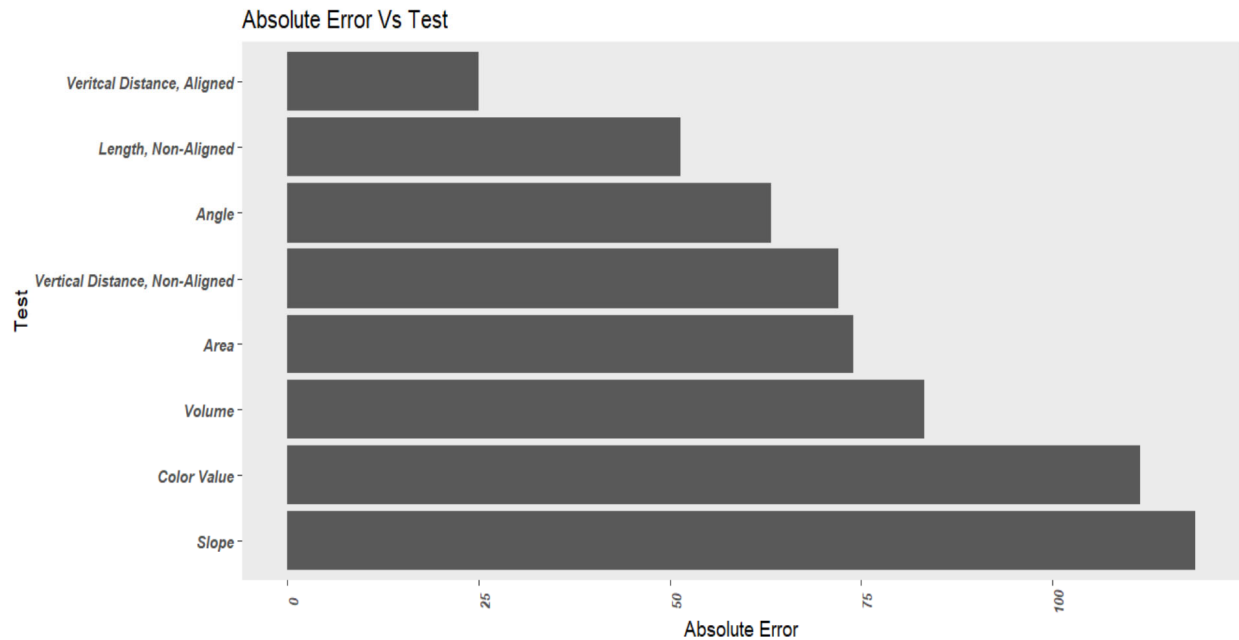


- In Tableau:

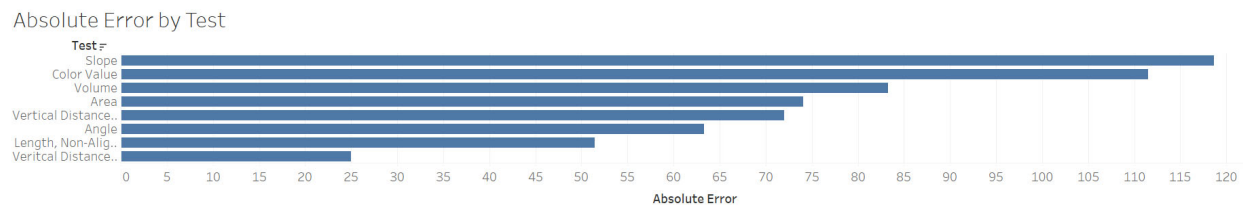


d. Absolute Error Vs Test

- In R:



- In Tableau:



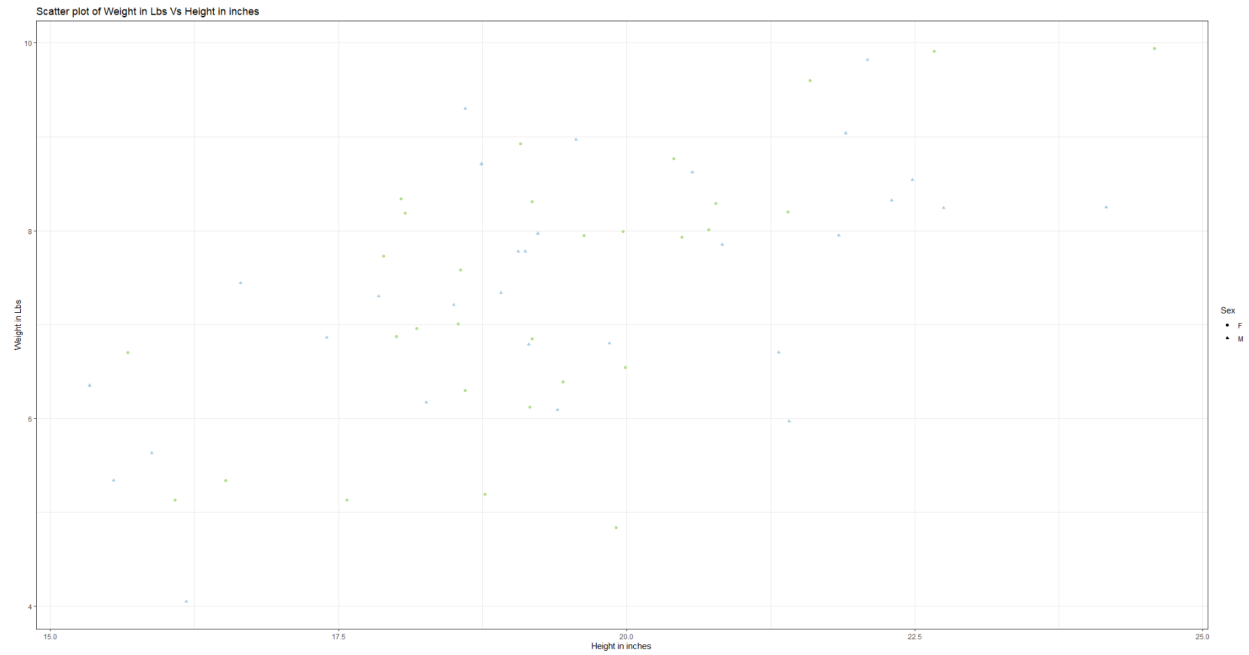
e. Explanations of the above graphs

- In part a, the histogram chart shows the distribution of error of all the test. The data is not normally distributed as they are potential outliers on the right and left, which need to be investigated.
- In part b, the chart shows the median error of each test. What we learn from the graph is that if the median error is positive, the response was greater than the true value; if the median error is negative, the response was less than the true value.
- In part c, the graph shows the standard deviation of error by test, it shows the spread of how widely the students varied in their responses for each test. From this graph, we learn that slope has a high standard deviation, which means that it has a wider spread of responses that need to be looked at.
- In part d, this chart is similar to the one in (b), the difference is that this chart doesn't define if the response was low or high than the actual value.

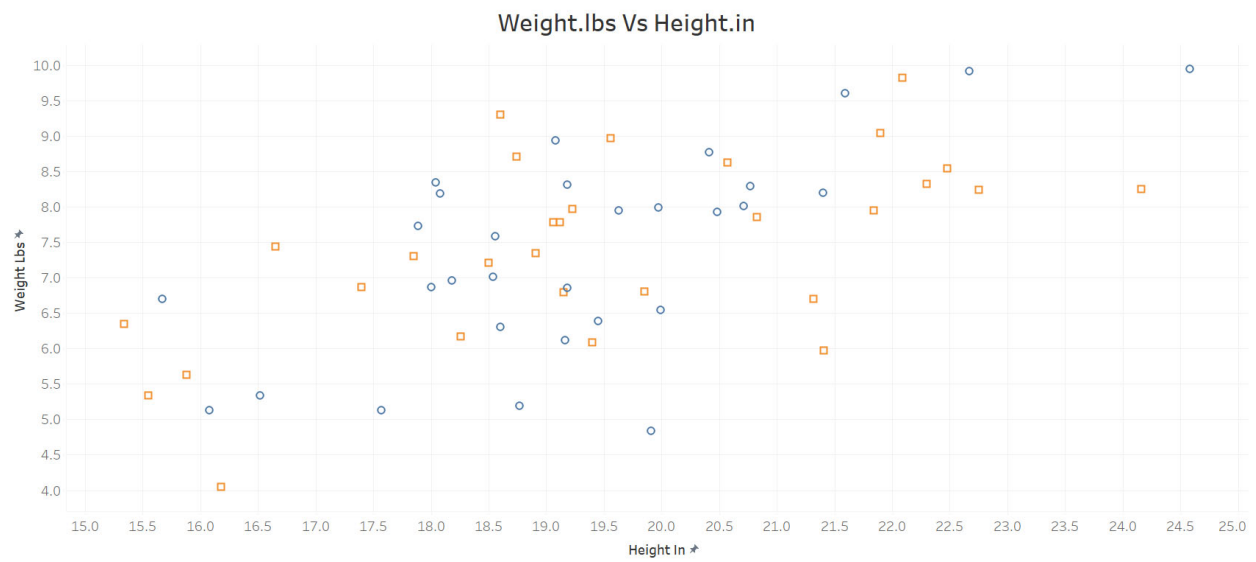
3. Infant data

a. Scatter plot of Weight.Lbs Vs Height.in

- In R:

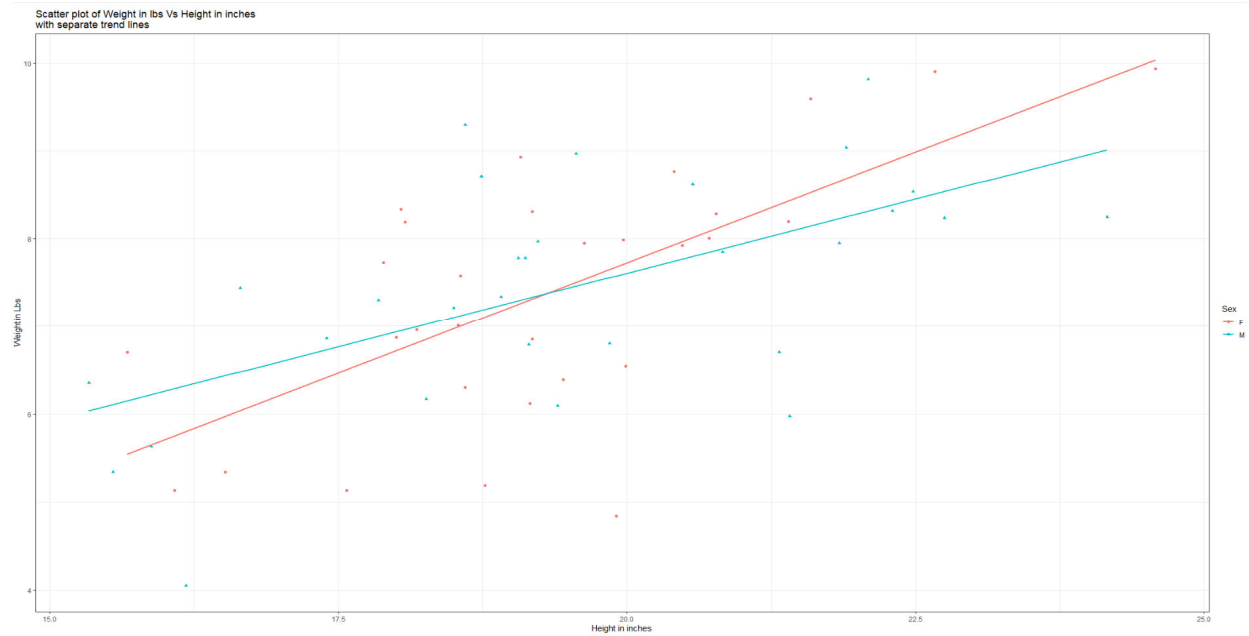


- In Tableau:

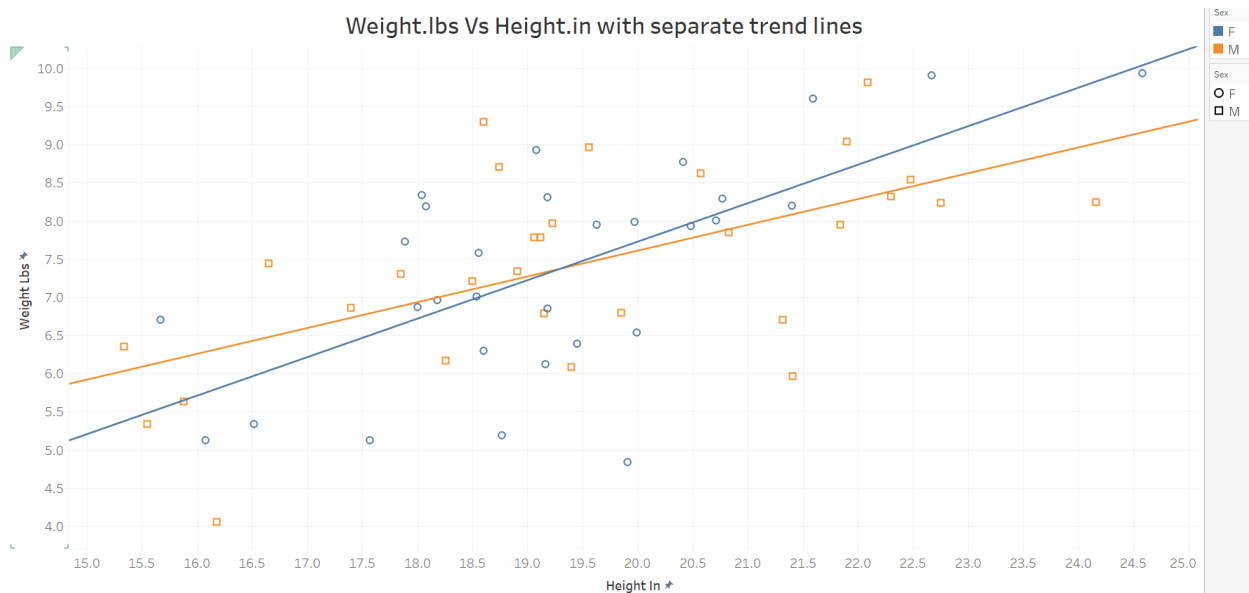


b. Separate trend lines for the two populations

- In R:



- In Tableau:



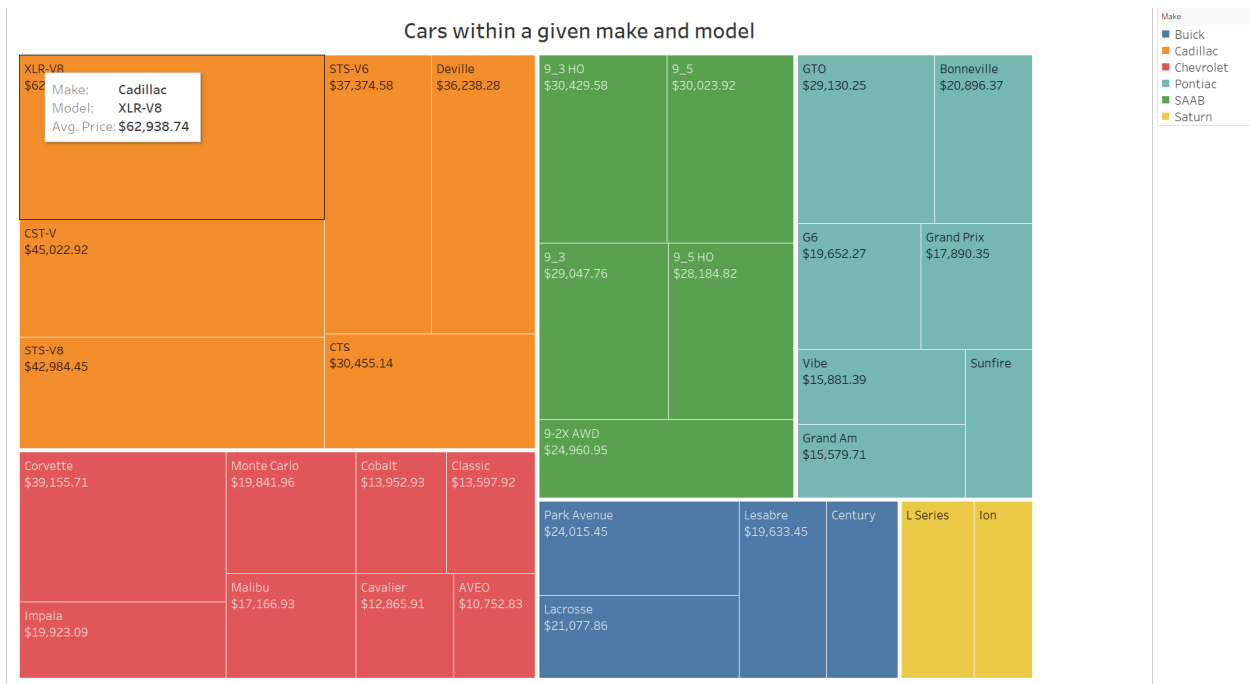
c. Explanations

- In part a, the scatter plot shows the strength, direction and form of the relationship between weight in lbs and height in inches. We observe that there is a strong positive correlation between weight in lbs and height in inches for both males and females.

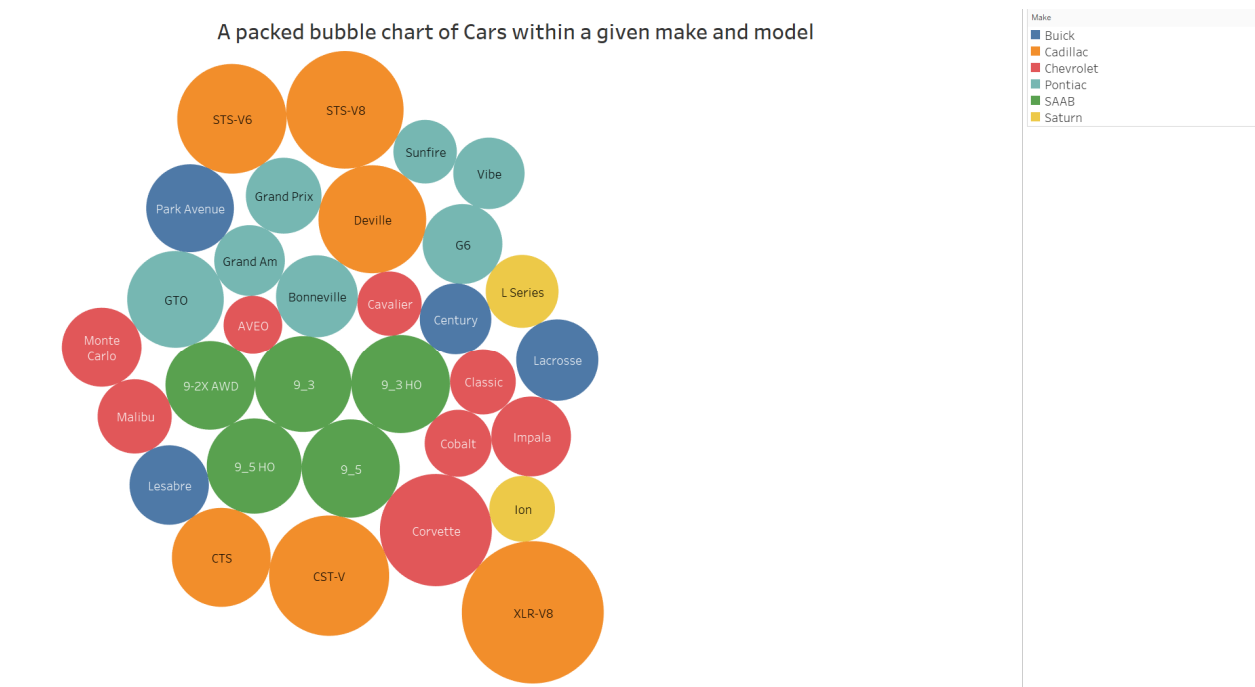
- In part b, a scatter plot has been drawn, fitted with two linear models between weight in lbs and height in inches; one for males and the other one for females. In R, the gray bands in the graphs represent confidence interval, which is set to 95% by default.

4. Gmcar_price data set

a. Tree map



b. Bubble chart

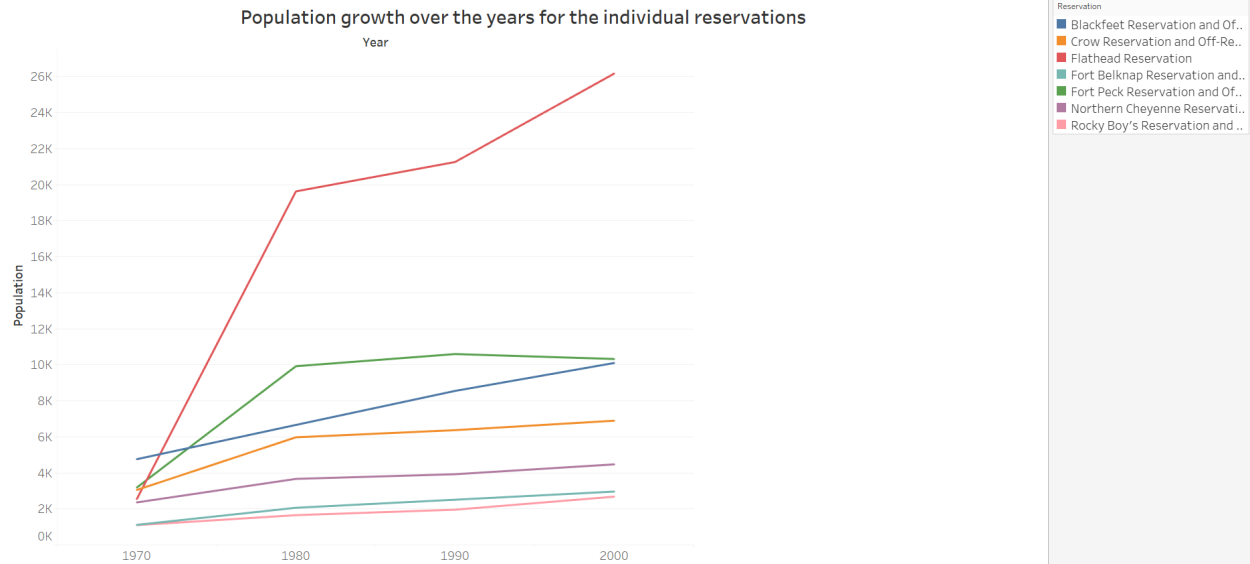


c. Explanations:

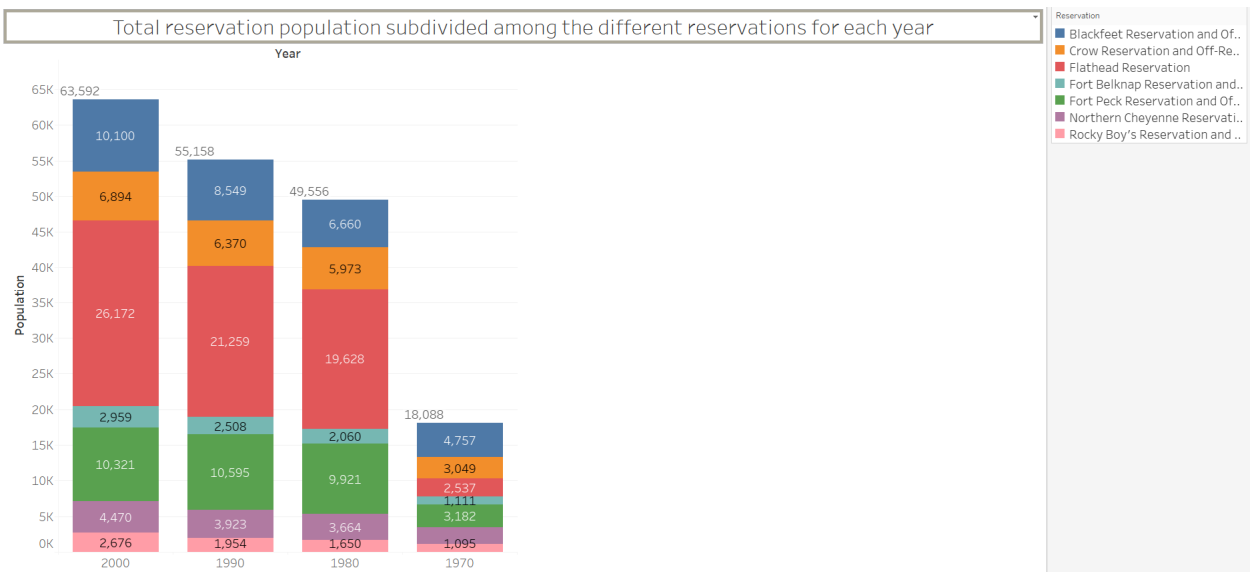
- In part a, we have a tree map subdivided into 6 price boxes by make and within a given make, there are various models. In the tree map, both size of the rectangles and their color are determined by the value of their prices. From the tree map, we learn that Cadillac has the biggest subdivision of price amongst the different GM brands whereas in part b, the packed bubble chart is divided into the prices of the different brands (make) and within a given a make, there are several models defined. We also learn the Cadillac has the biggest subdivision of price amongst the different GM brands (make). The size of the bubbles shows the total prices for the different models within a make.

5. Reservation70-00 data set

- Population growth over the years for the individual reservations.



b. Total population growth over the years.



c. Box-and-whiskers plot.

Distribution of population at reservation locations over the years

