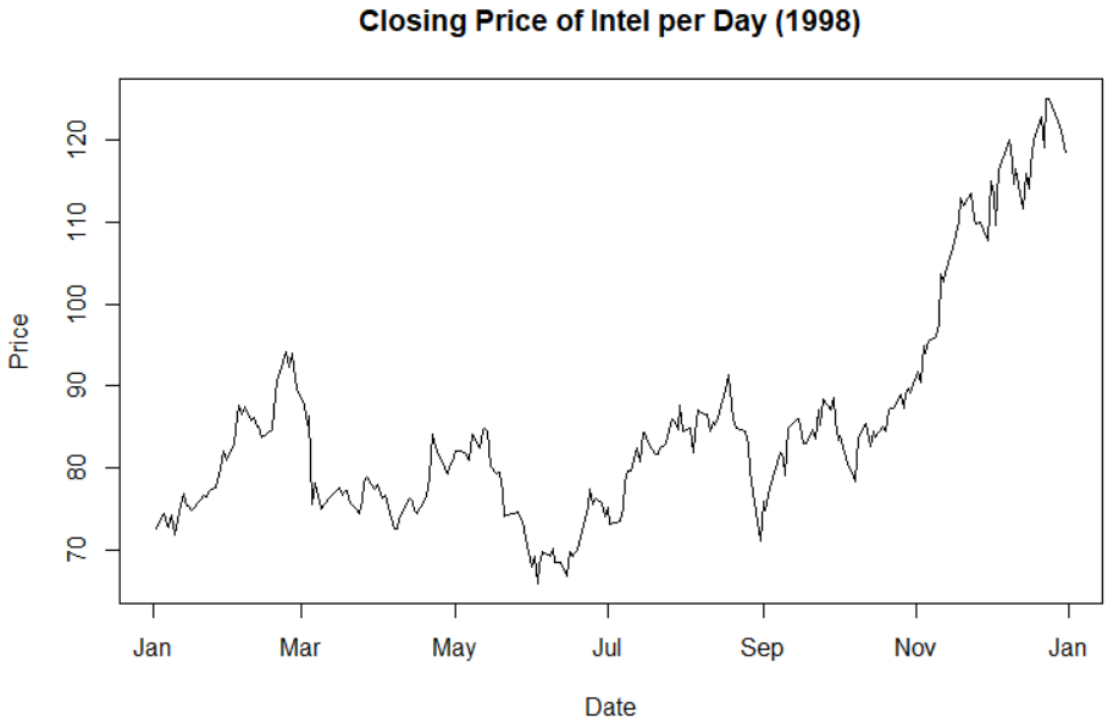
1a.

1a R – first MDY was used to set the DATE column. The plot created using the basic plot command and setting axis and date.

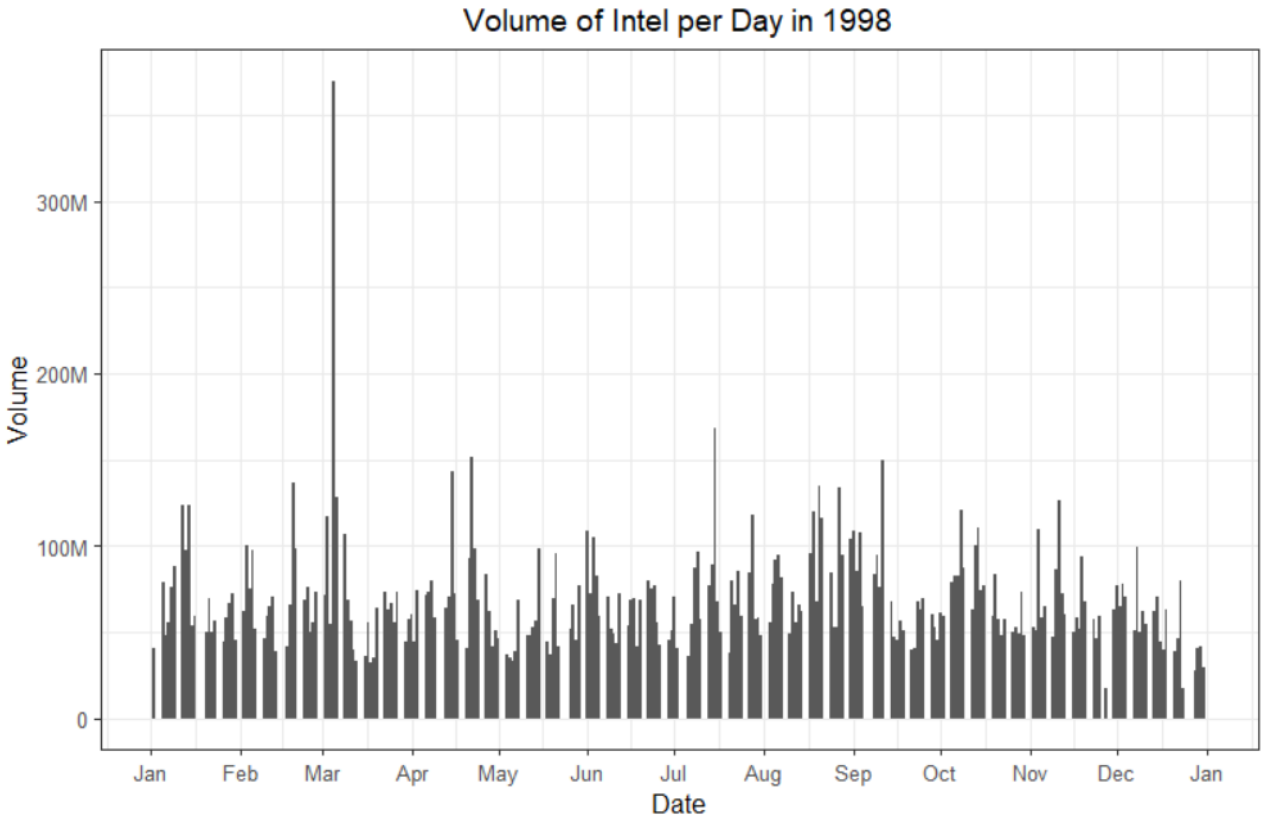


1a Tableau – simple to use. GUI allowed placement of Date in Columns and SUM(Close) in Rows. The labels/axis titles were edited as was the color and size of line.

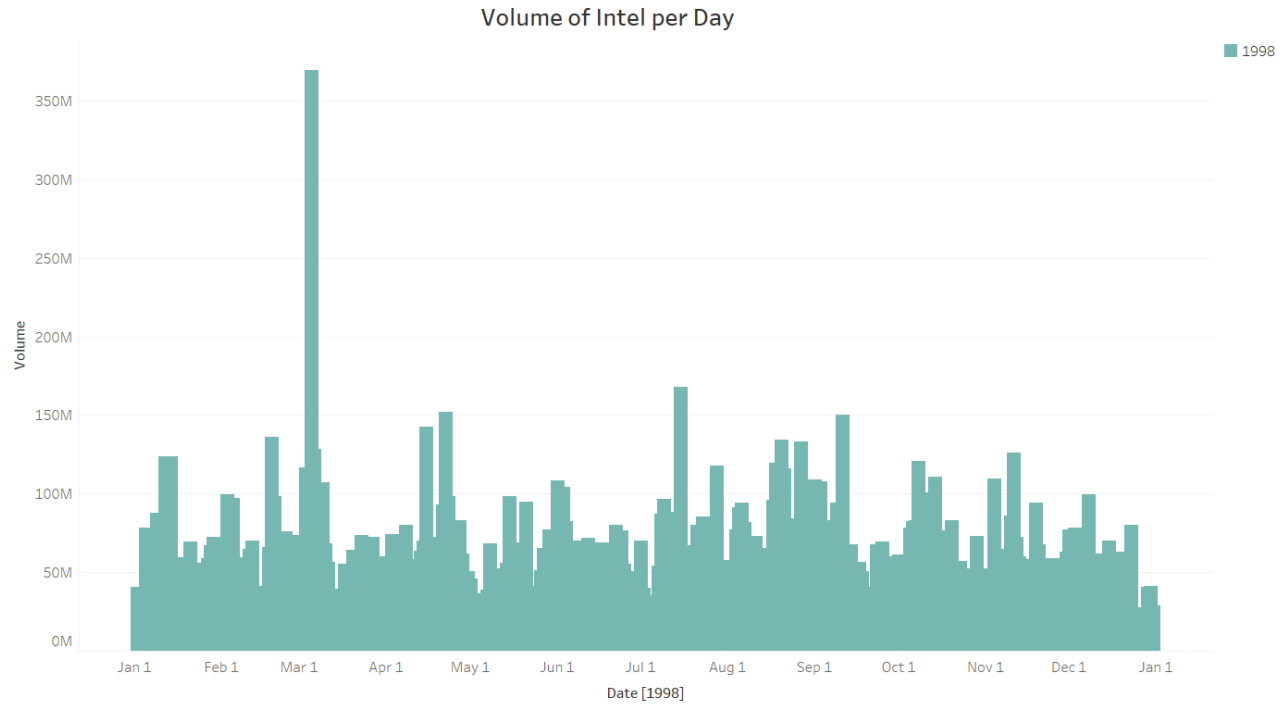
1a Power BI - the most frustrating of the three…Added the data, chose "line chart" and dragged "Date" to axis, unchecked "Date Hierarchy" and moved "Close" to values.

1b.

1b. R -using ggplot and setting titles, colors, and scales for volume to make the x axis less ugly and scale_x_date for the y axis to make the months appear clean.

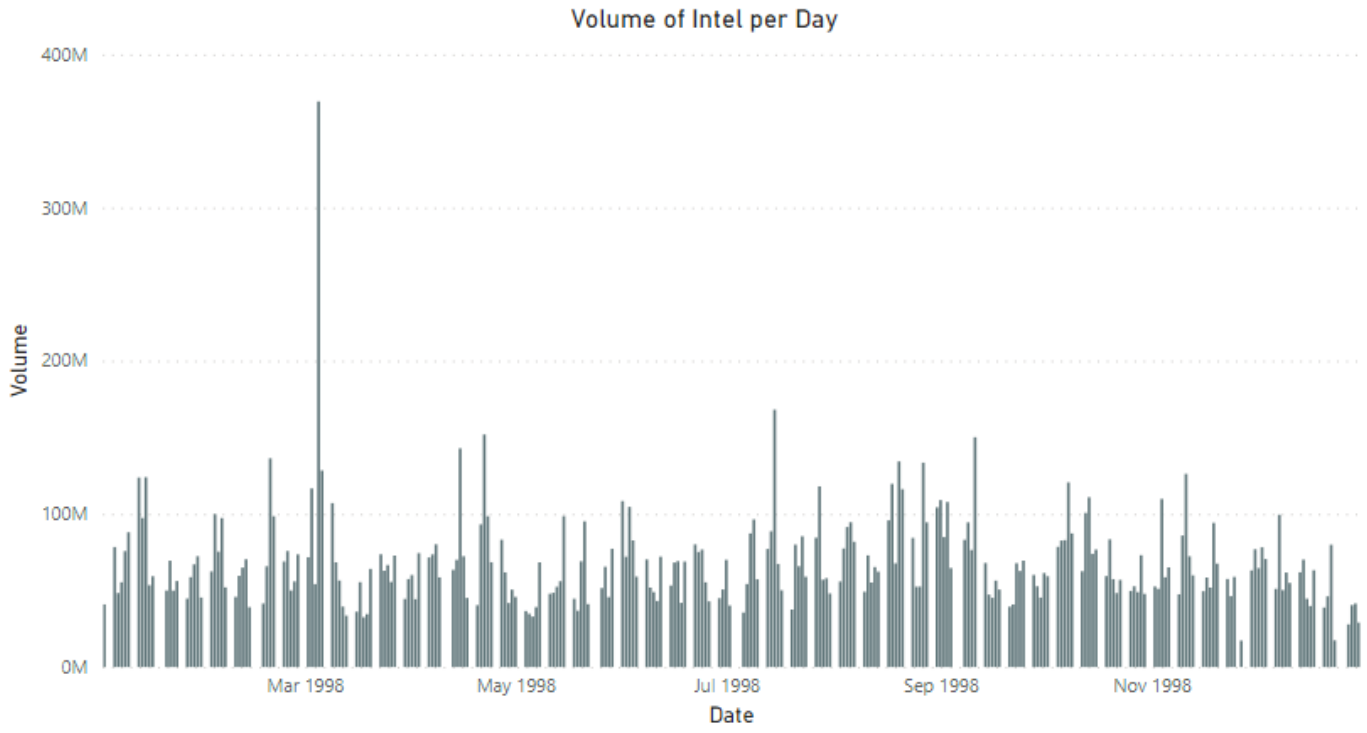


1b. Tableau – not much different than the last graph, just replaced Close with SUM(Volume) and changed to bar graph.



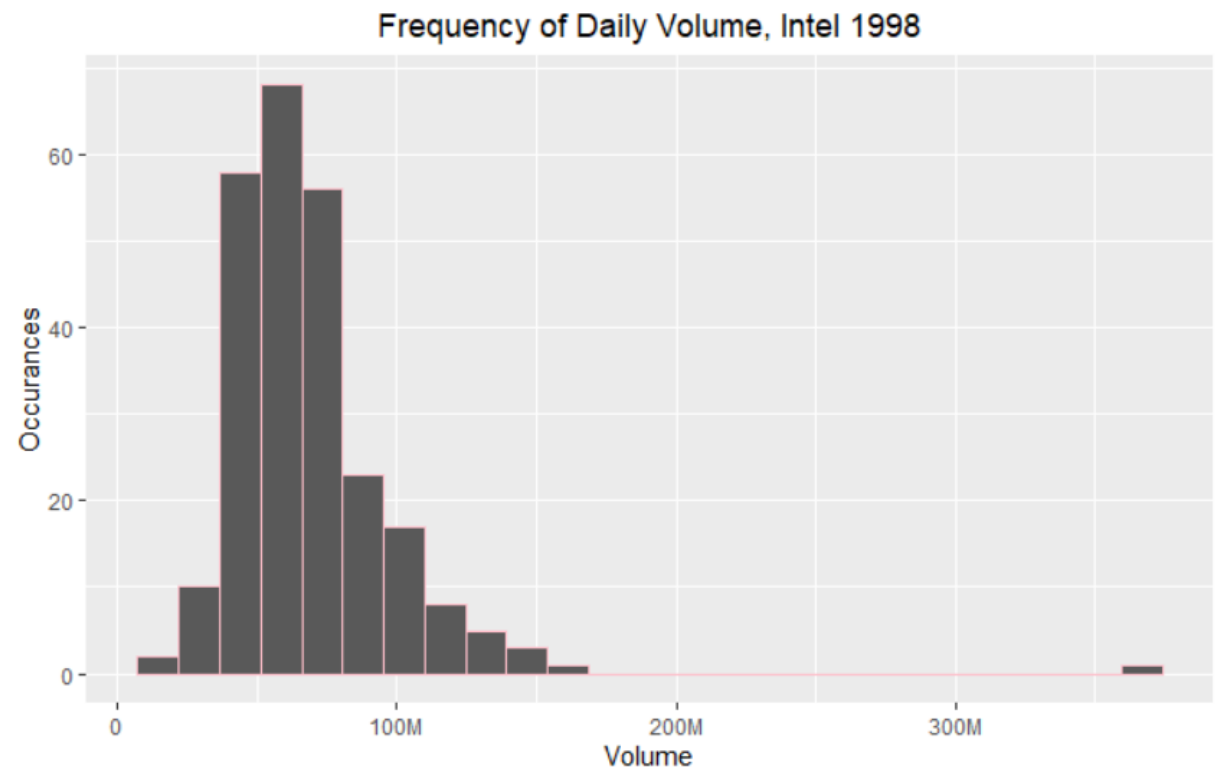
The plot of sum of Volume for Date. Color shows details about Date Year.

1b. Power BI- interestingly, Power BI wanted to automatically set the volume scale to Billions, leaving us with a left axis that went from 0.1B to 0.4B. Manually set it to millions like Tableau's auto did. The difference in impact of 100,000,000 is remarkable when shown as 100M vs 0.1B.

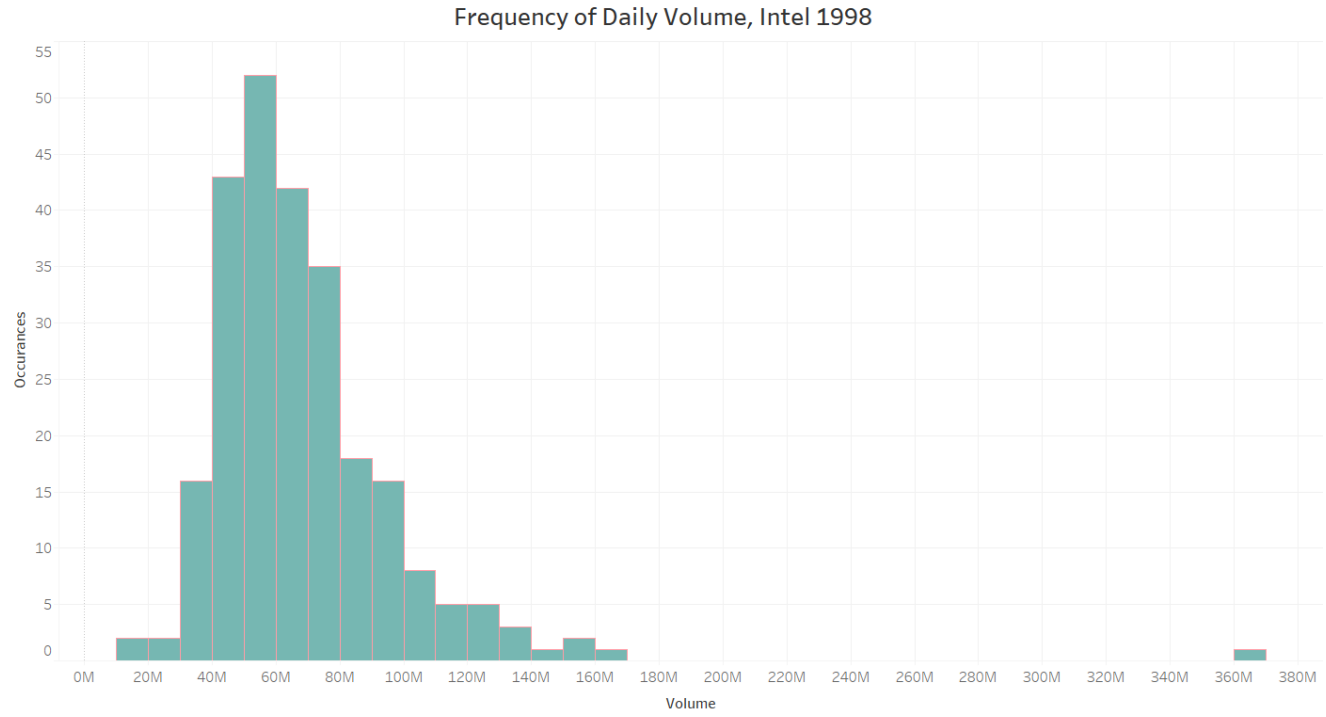


1c.

1c. R -using ggplot and setting titles, colors, and scales for volume to increase legibility and make the graph pretty.

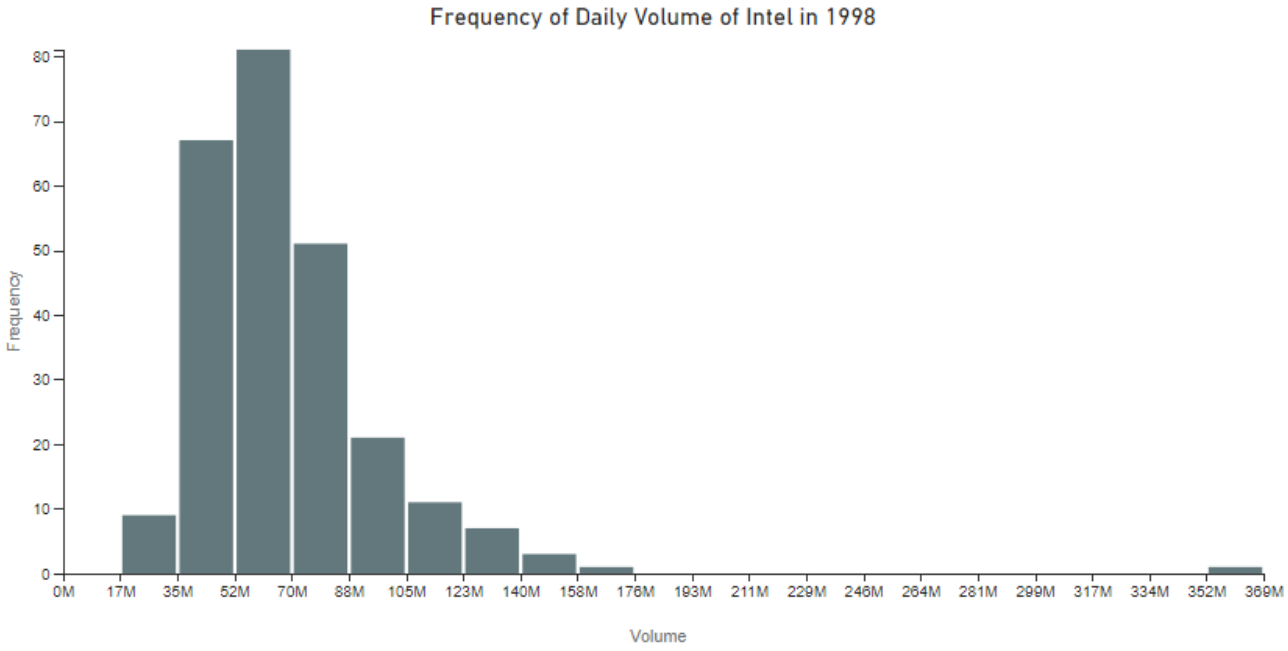


1c. Tableau – messed with border color options, took too long to figure out to change SUM(volume) to COUNT(Volume), used "Show Me" to change to histogram, changed titles.



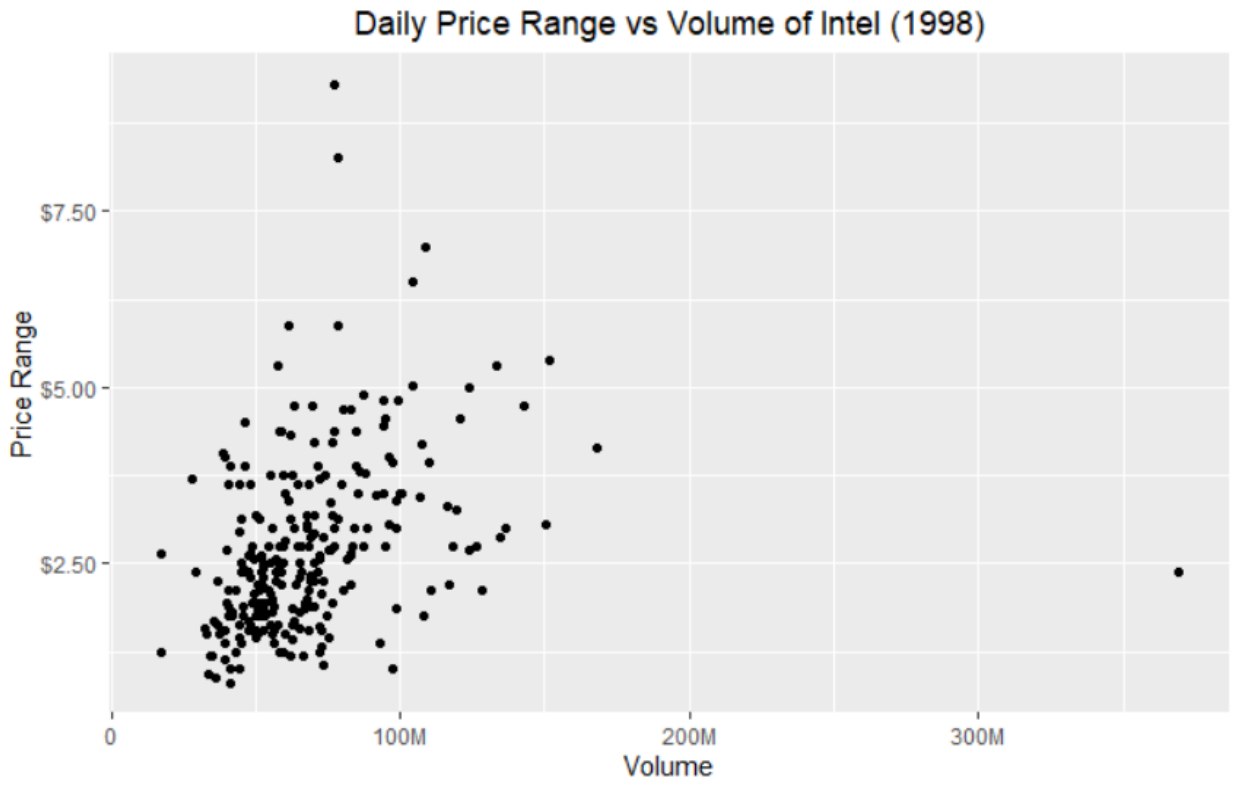
The trend of count of Volume for Volume (bin).

1c. Power BI- Apparently Power BI has no built-in functionality to create histograms because why not? Downloaded the histogram visual, and it presented a terrible autogenerated graph. The version below has less bins than the R and Tableau graphs (20), only because when I tried to add more bins the graph would show negative values to the left and I gave up trying to fix that.

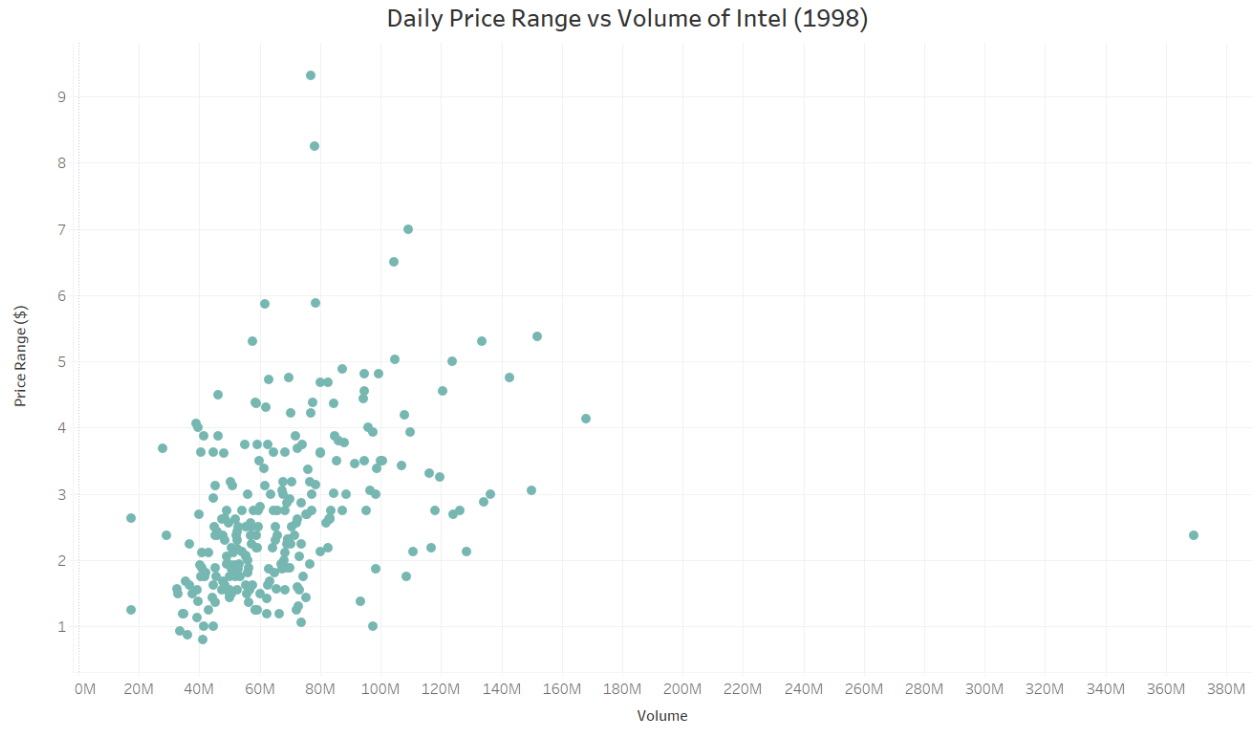


1d.

1d. R – A new column with high-low price range was added to Intel.1998 then ggplot geom_point was used to create the scatterplot

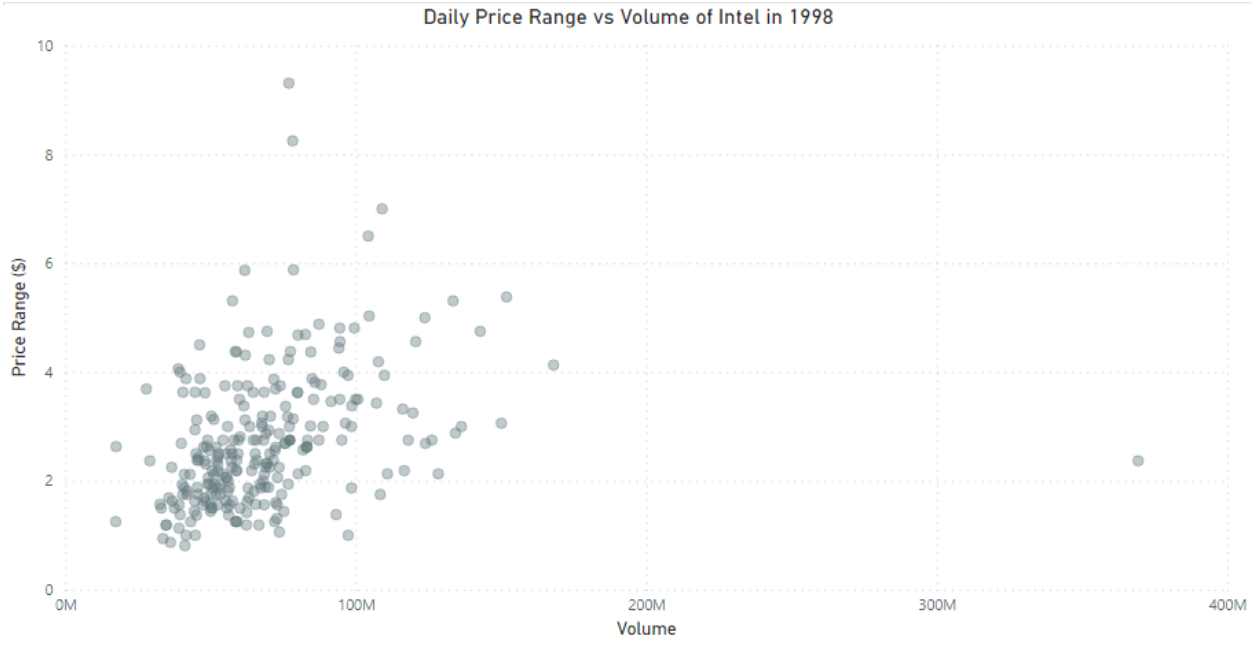


1d. Tableau – Did analysis calculation, very simple [HIGH] – [LOW], then used this as the y-axis



Volume vs. HighLowRange.

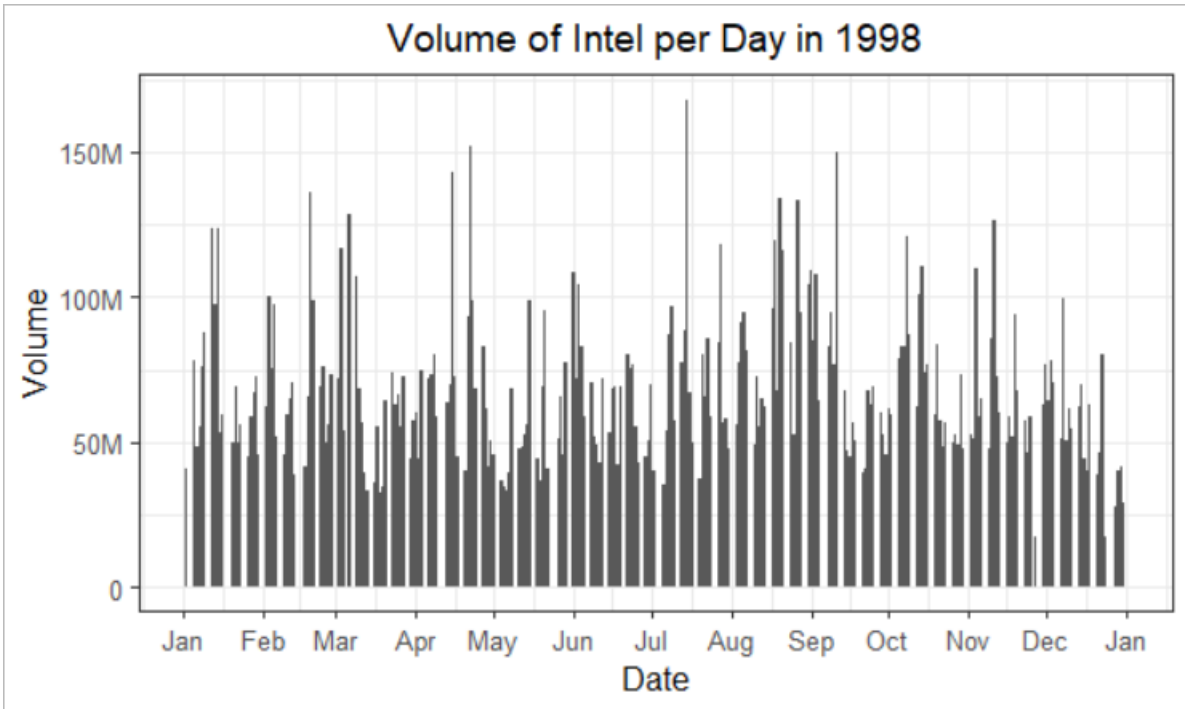
1d. Power BI- this was very intuitive, I hit "Transform Data", then "Add custom column" and was given the option to create the Range column based on [HIGH] – [LOW] calculation. The Range data had to be set to data type "Decimal Number", then the scatterplot was ready to be made.



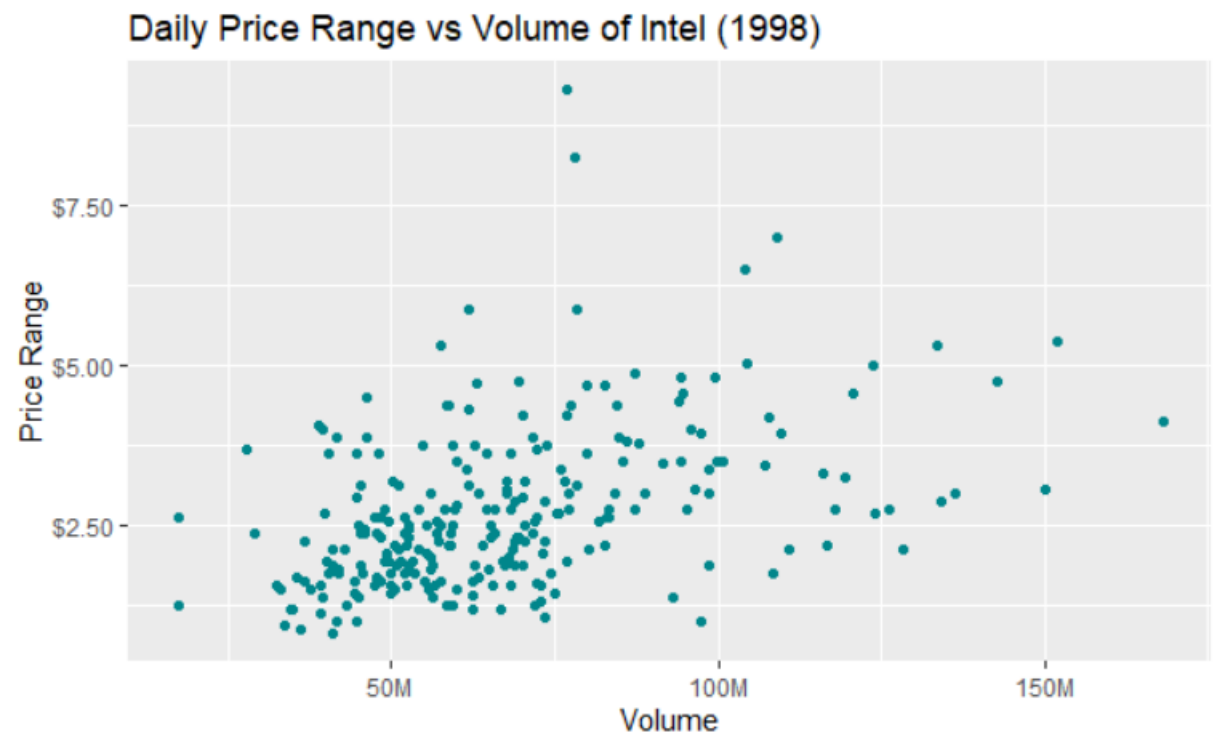
1e.

R: The first step was looking at the summary statistics of the Intel dataset as a whole using the summary function. From previous graphs I could already determine that volume had an upper outlier so dplyr was used to filter all values above the 99th percentile, which included 3 points. The maximum value was far above the other two data points and shared no noticeable connection with any other value, so Intel.1998 was filtered to include the subset of all values less than the maximum for Volume.

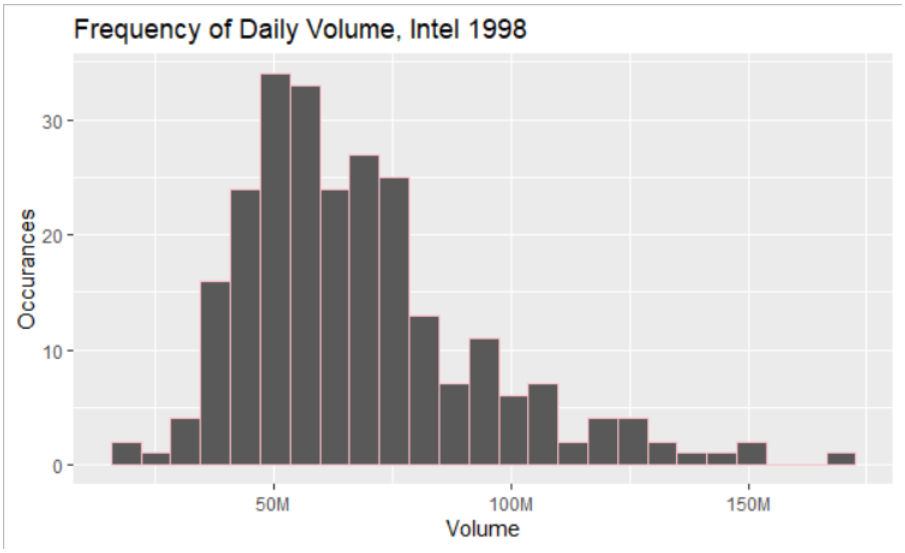
This exponentially helped the appearance of the Volume vs Date graph:



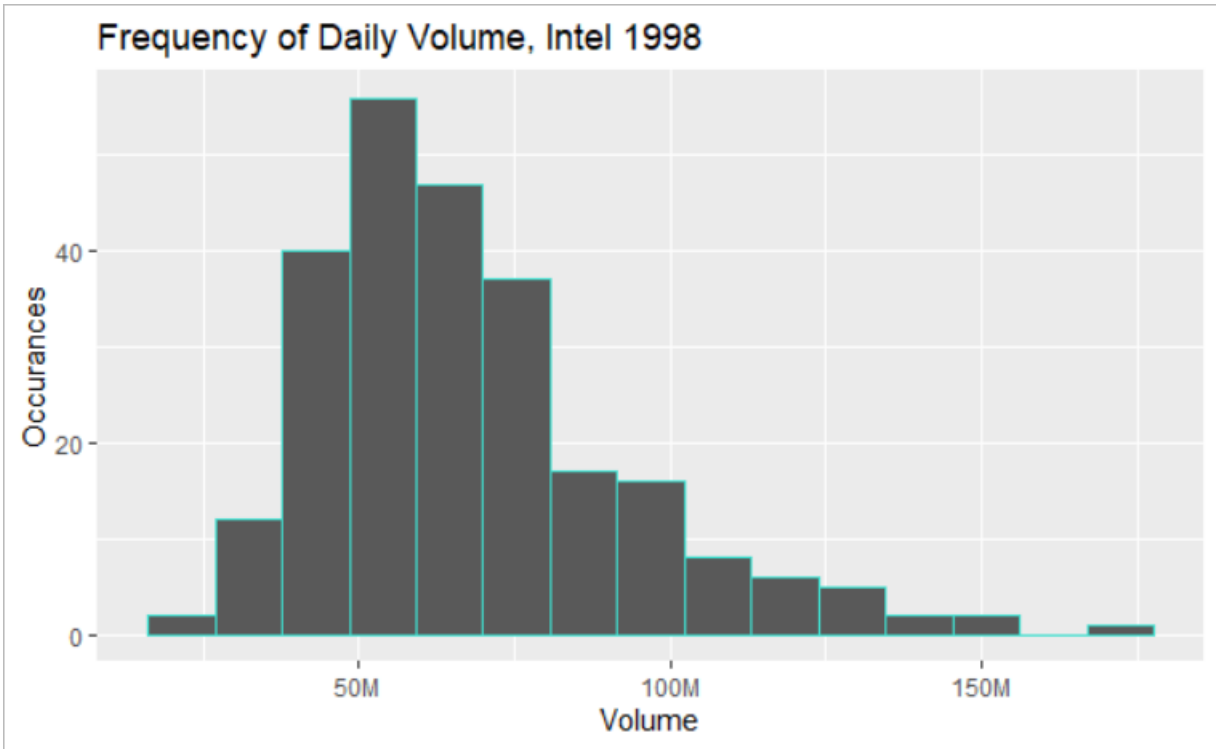
And the Price Range vs Volume Scatterplot:



But when using the same R code, the loss of the outlier made the histogram look like there was another outlier!



This was because the graph originally had 25 bins, which I chose because the huge gap between the clustered data and the outlier required it, lest the graph was three thick bars on one side and the outlier on the right. With the outlier gone, the data was now able to be shown with less bars.

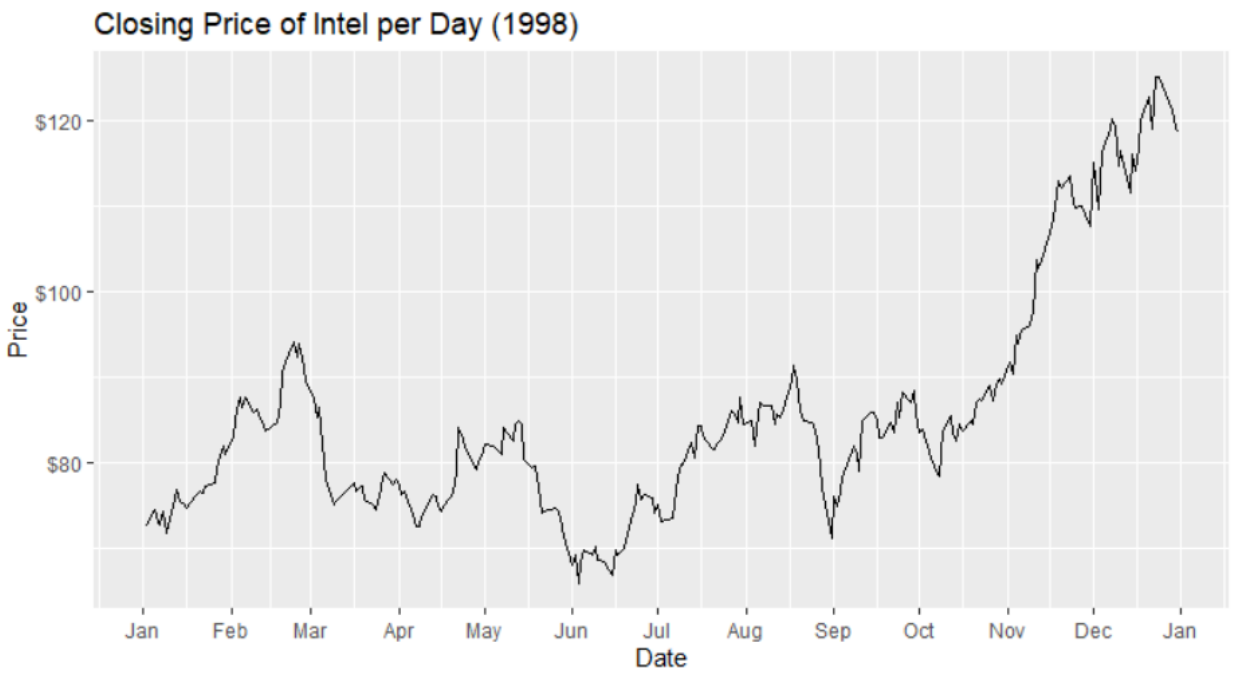


I think this shows the general distribution of Volume more relevantly, with a side note that there was the 300M outlier not pictured.

The same quantile method was used to look for outliers in the HighLowRange category. I decided the values above the cluster aren't outliers, since the two highest values occurred on back-to-back days, suggesting an important spike took place.

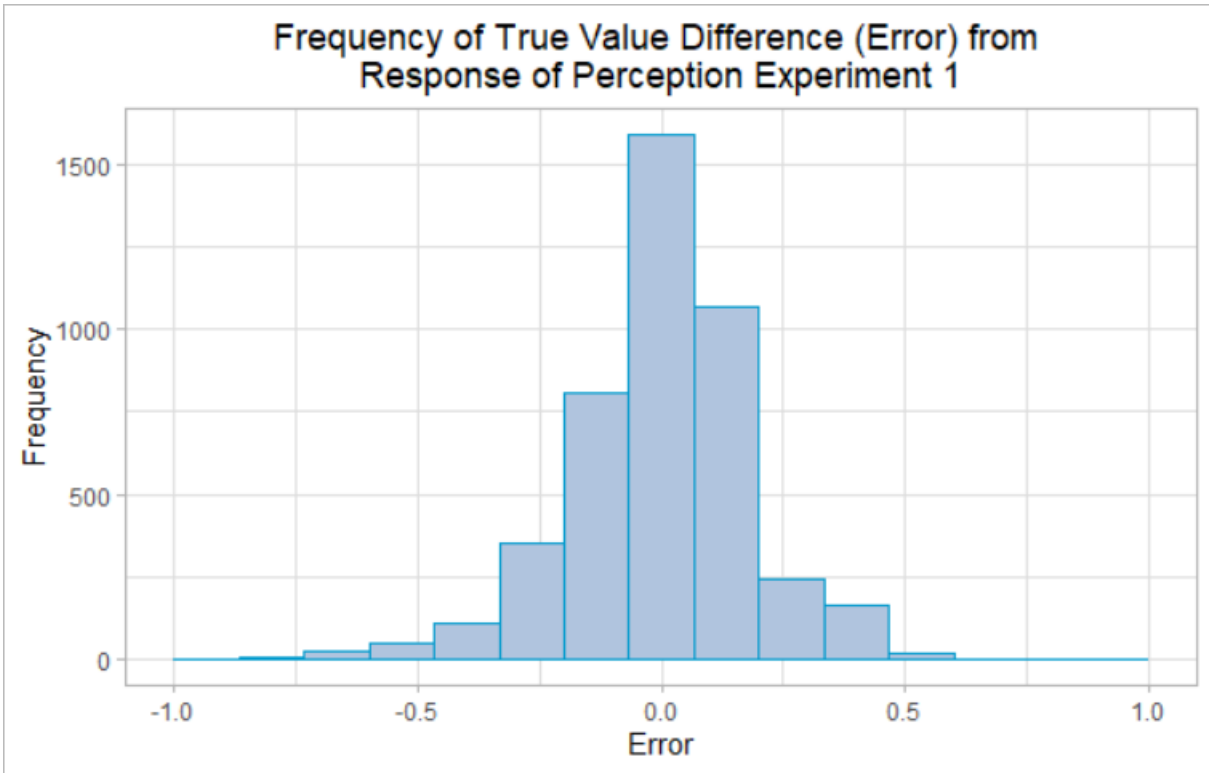
For Closing Price, the values at the top and bottom were checked, since there were values in the upper and lower bounds that could have been outliers, however in both

cases, the values in the smallest percentile occurred during the same time period, and those in the highest during another. Like with highLowRange, this convinced me that these were not outliers, but significant data points.

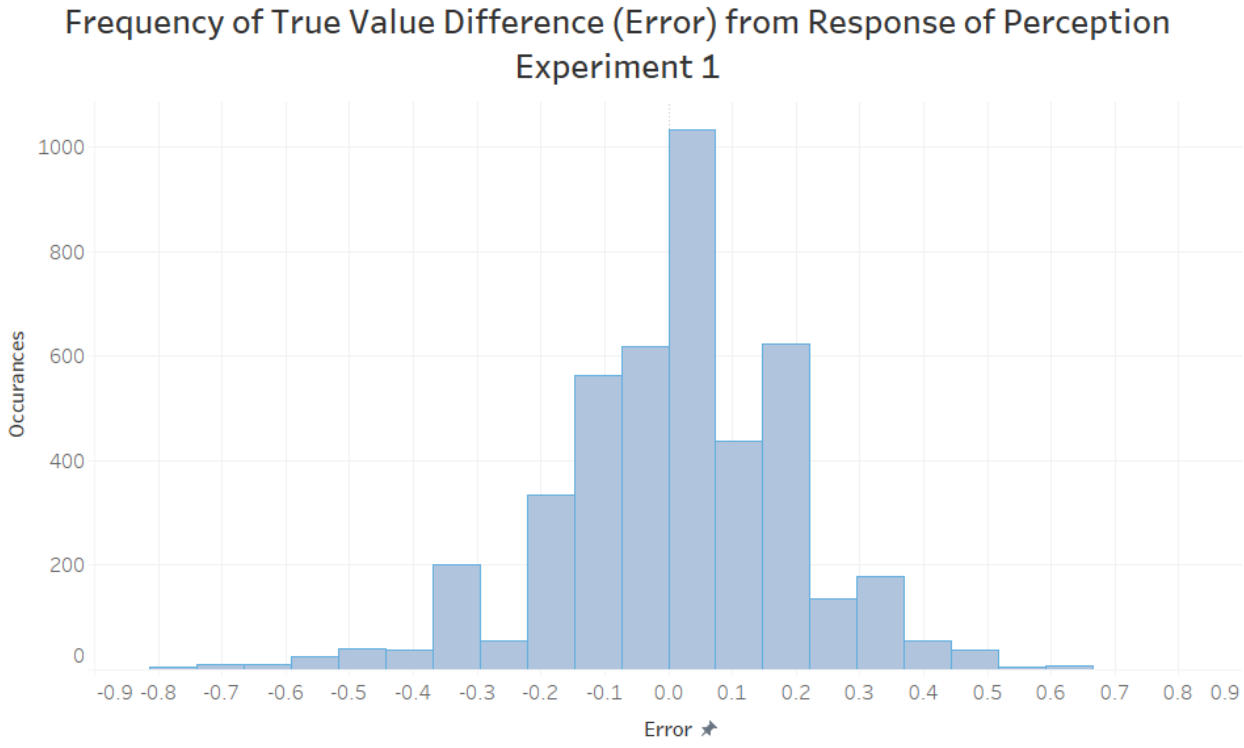


2a.

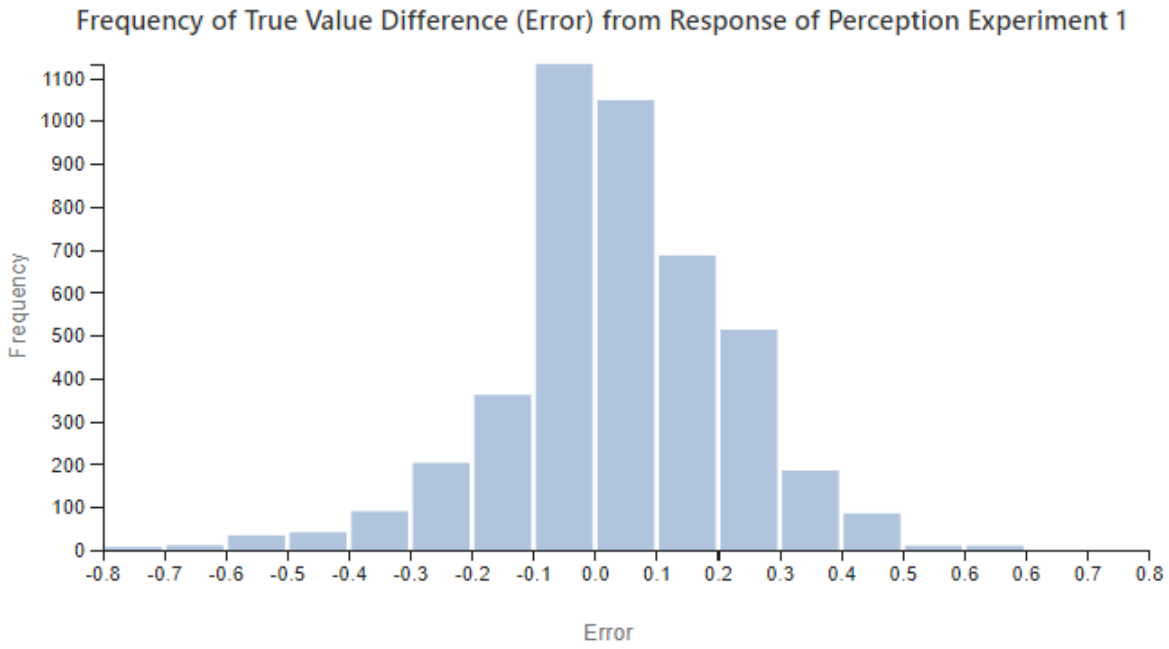
2a. R – updated theme to light, settled on 16 bins, set graph to run from -1 to 1



2a. Tableau- was messing with colors and realized Tableau has the amazing ability to "screen select" colors, so I took the exact colors that the r graph has. Set graph to run from -.9 to .9



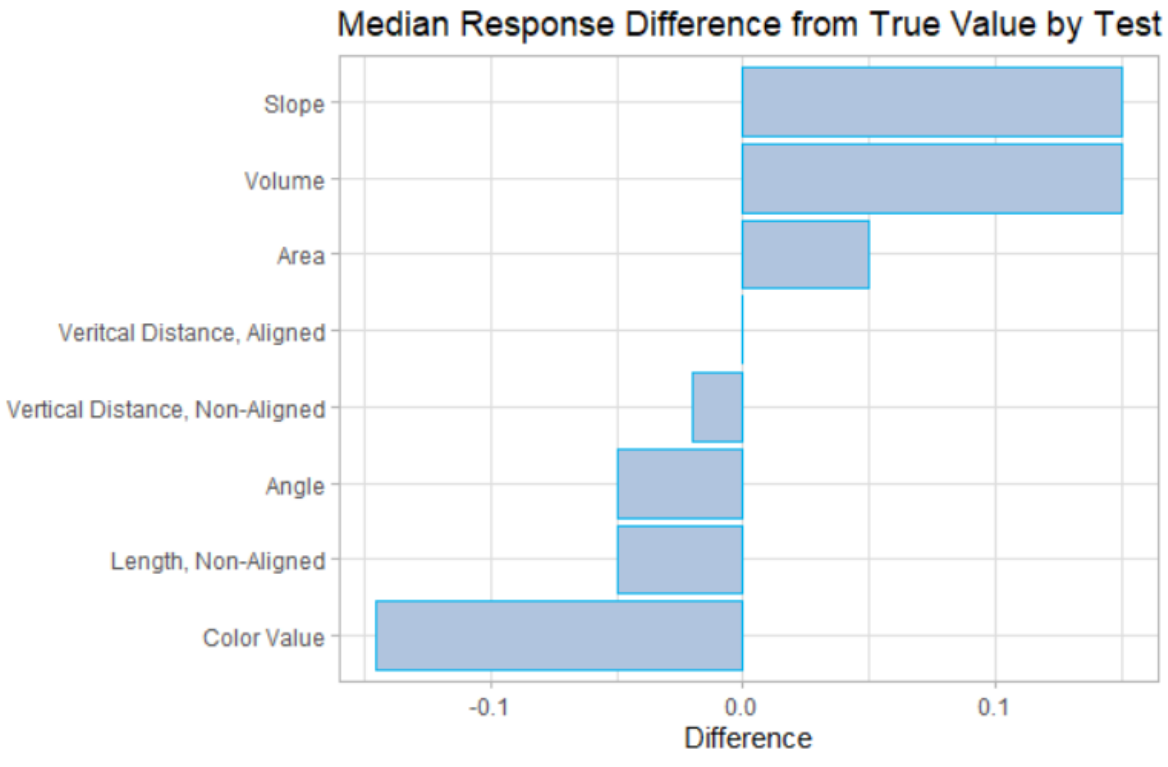
2a. PowerBI- PowerBI allows you to choose color based on HEX number, and Tableau shows HEX number for all colors. Thus: all graphs have the same fill color.



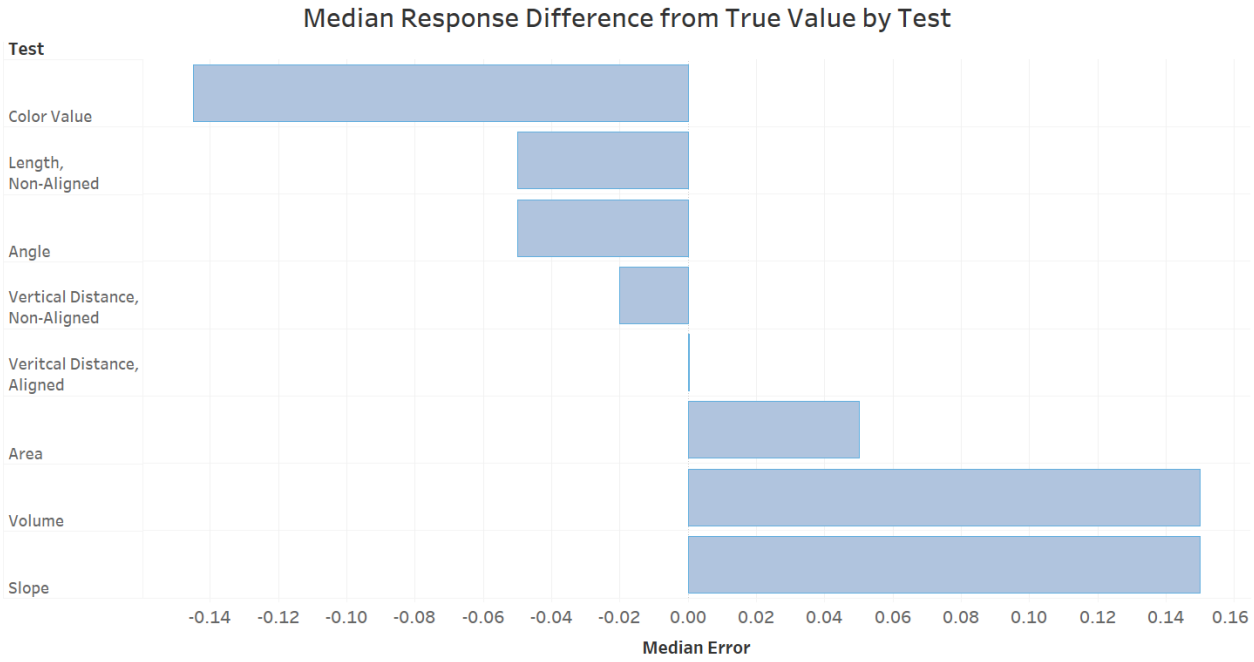
2e. (a) All three graphs are showing that the amount of error from Responses has a bell curve around the 0.0 mark, suggesting responses were largely accurate.

2b.

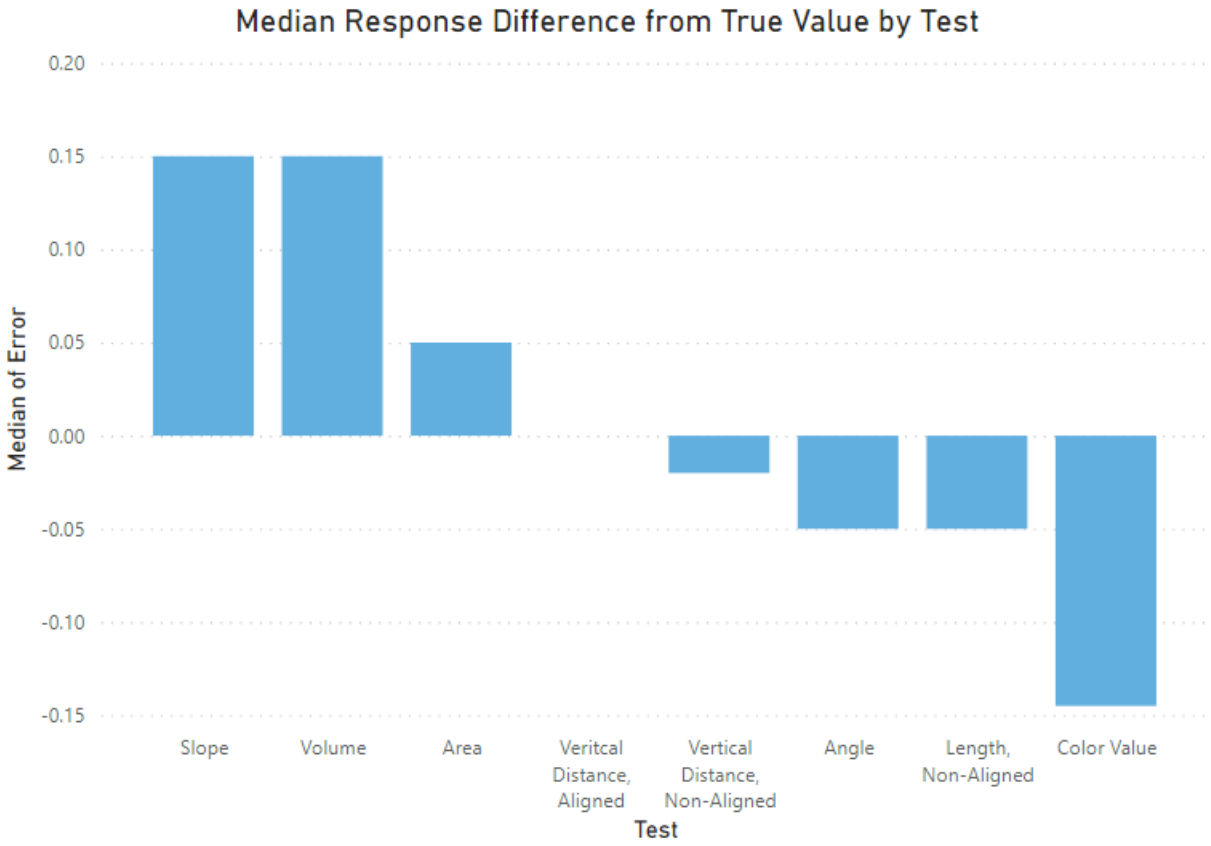
2b. R- Couldn't get the plot to look nice with median on the y axis, even though that's how I planned it in my head… Aggregated the dataset to find median and created a new dataset that this graph is created from. Used scale_y_discrete to sort tests in descending order of Median Value.



2b. Tableau – placed MEDIAN[Error] as Column, Test on Rows. Tweaked borders and label fonts.



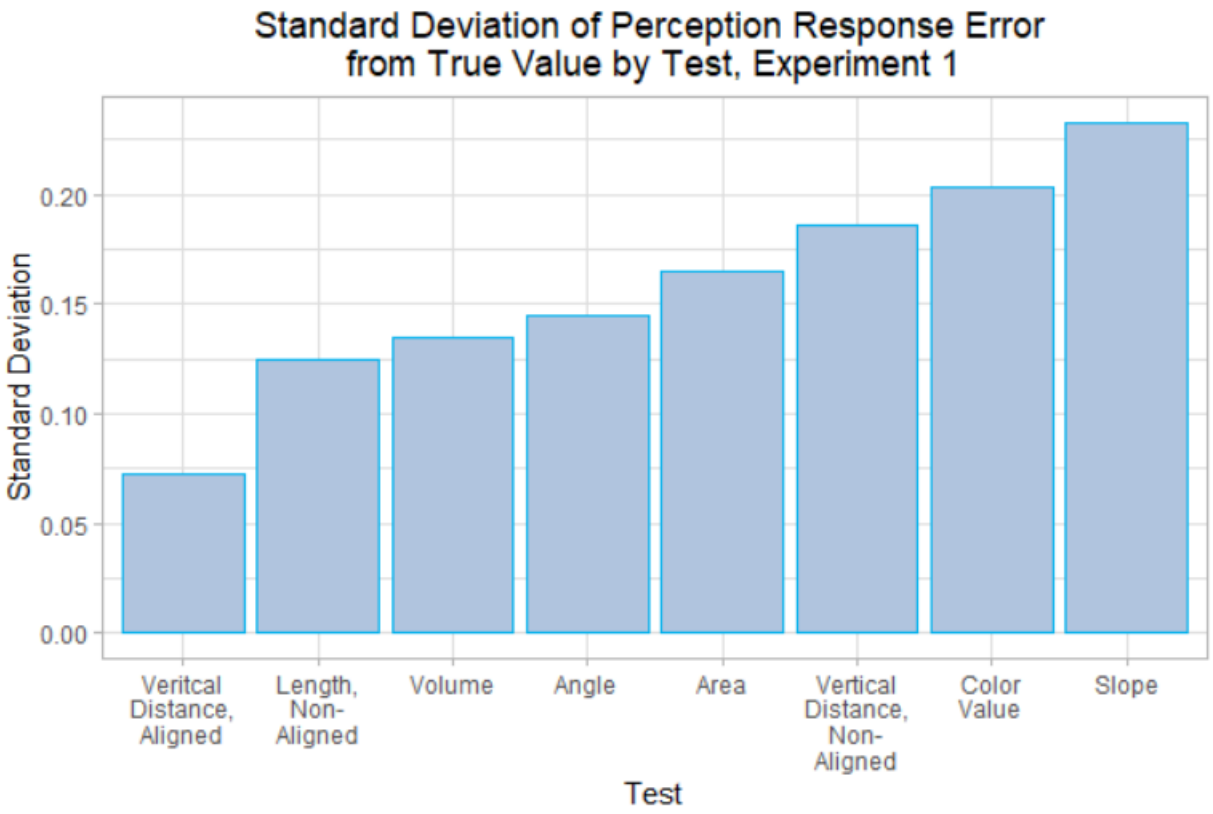
2b. Power BI – PowerBI auto aligned the Test names and allowed for the graph to a clean appearance with the flipped axis labels.



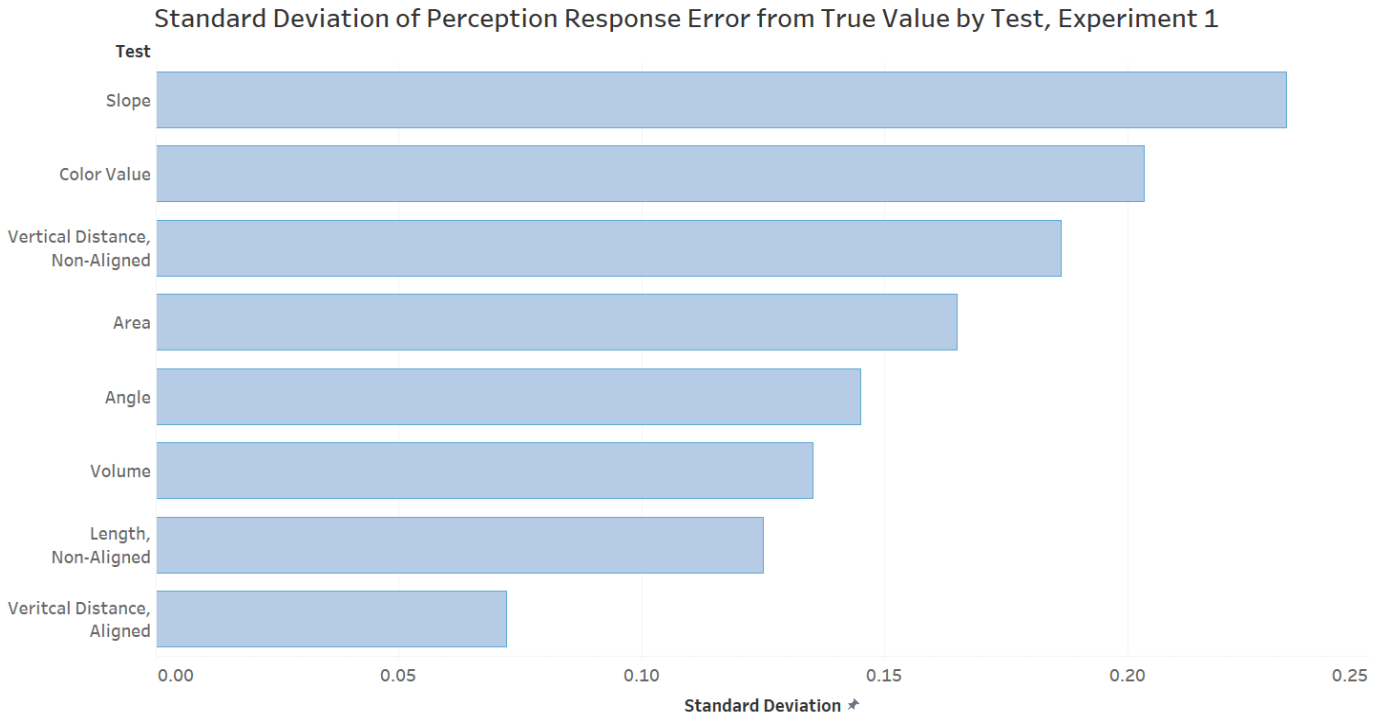
2e. (b): These graphs would work better with scientists than with a general audience. The graphs are showing that the median response differences were all relatively small, with slope, volume, and color value having the greatest medians. However, median is also arguably not the best method of comparison, so the graph might hide a greater skew in another direction.

2c.

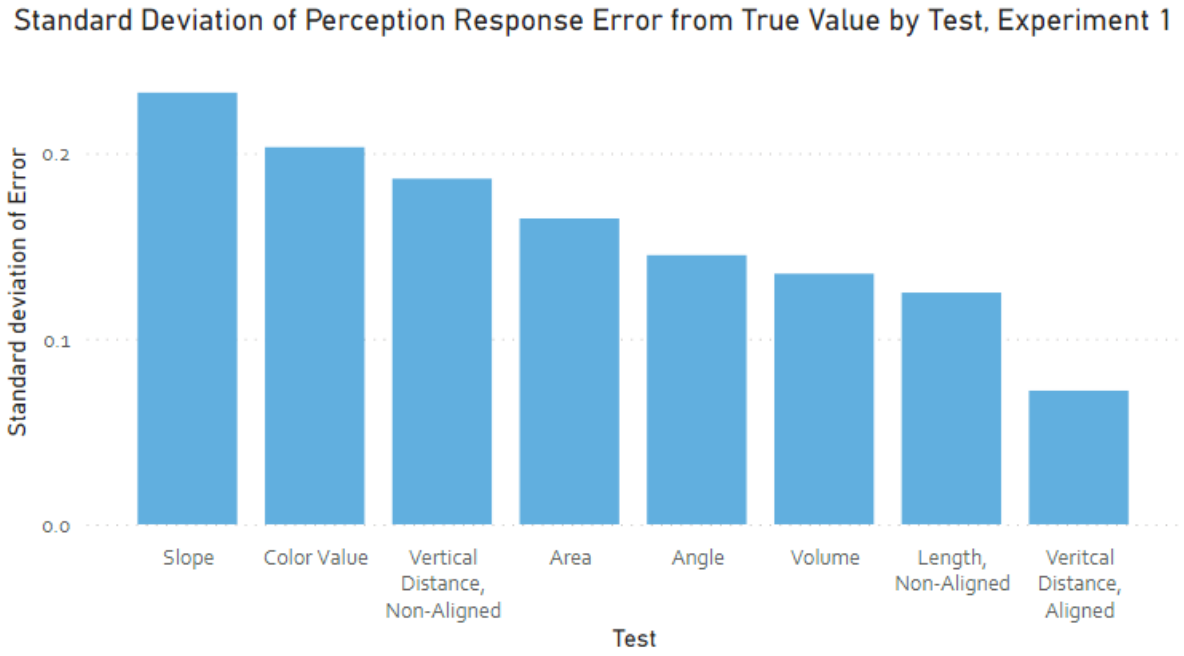
2c. R- Similar to part b, data was aggregated by Test to find standard deviation and saved to a new dataset (ds2), then sorted in ascending order. For this graph I also managed wrap the x-axis labels so the test type is on the x-axis!



2c. Tableau – columns were STDEV[Error], rows were [Test]. Adjusted gridline, start/end values, Test sort order, and ticks



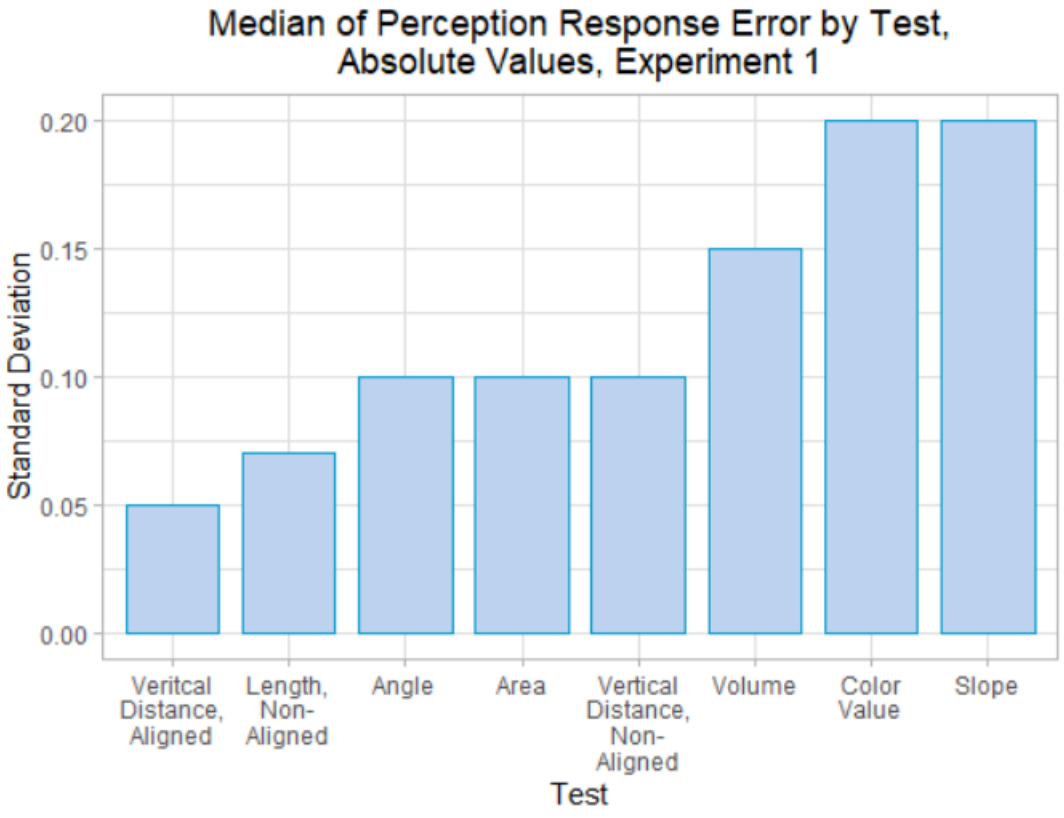
2c. Power BI – Automatically sorted the graph in descending order. Adjusted ticks and titles



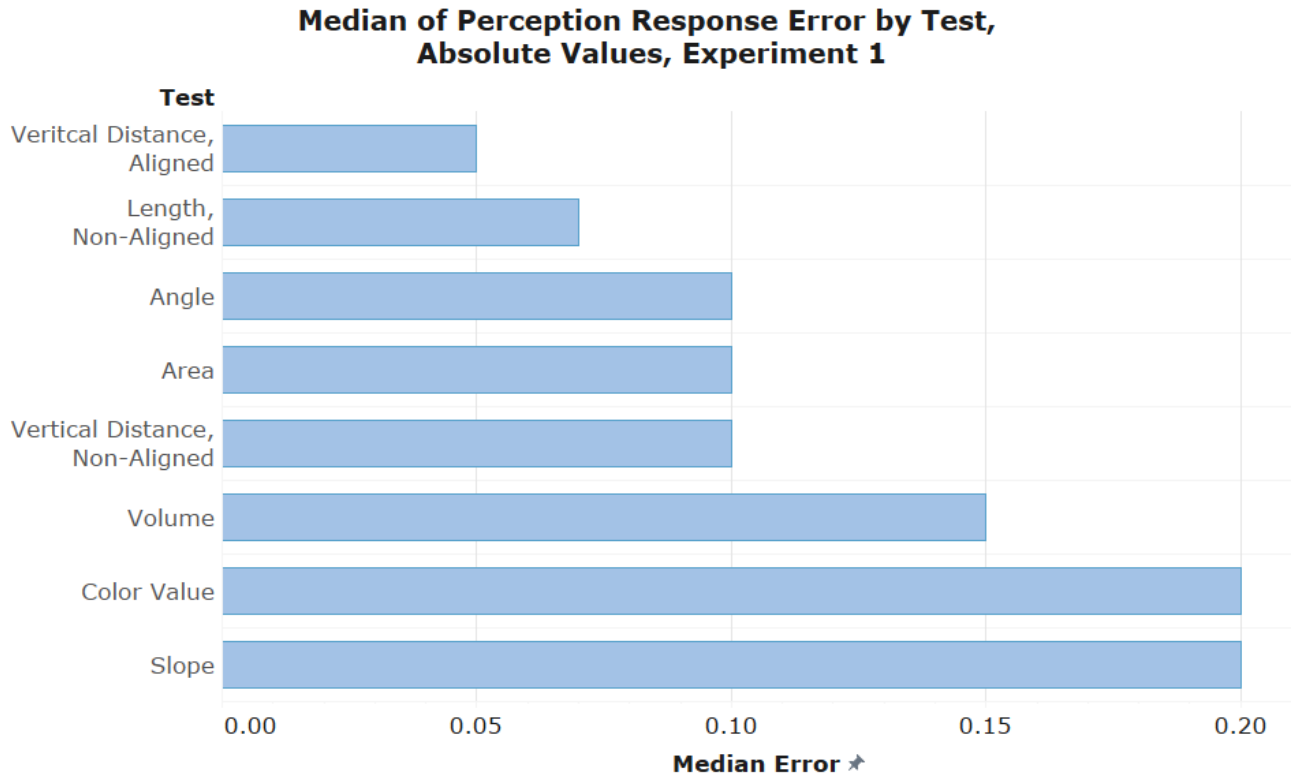
2e. (c): These graphs shows the standard deviation of error was greatest with slope, and least with vertical distance tests. This tells me that the highest range of error was occurring during the slope tests. Coupled with the first graph, I'm led to think that cases at the ends of the bell curve may have occurred during the tests with greatest STDEV of Error.

2d.

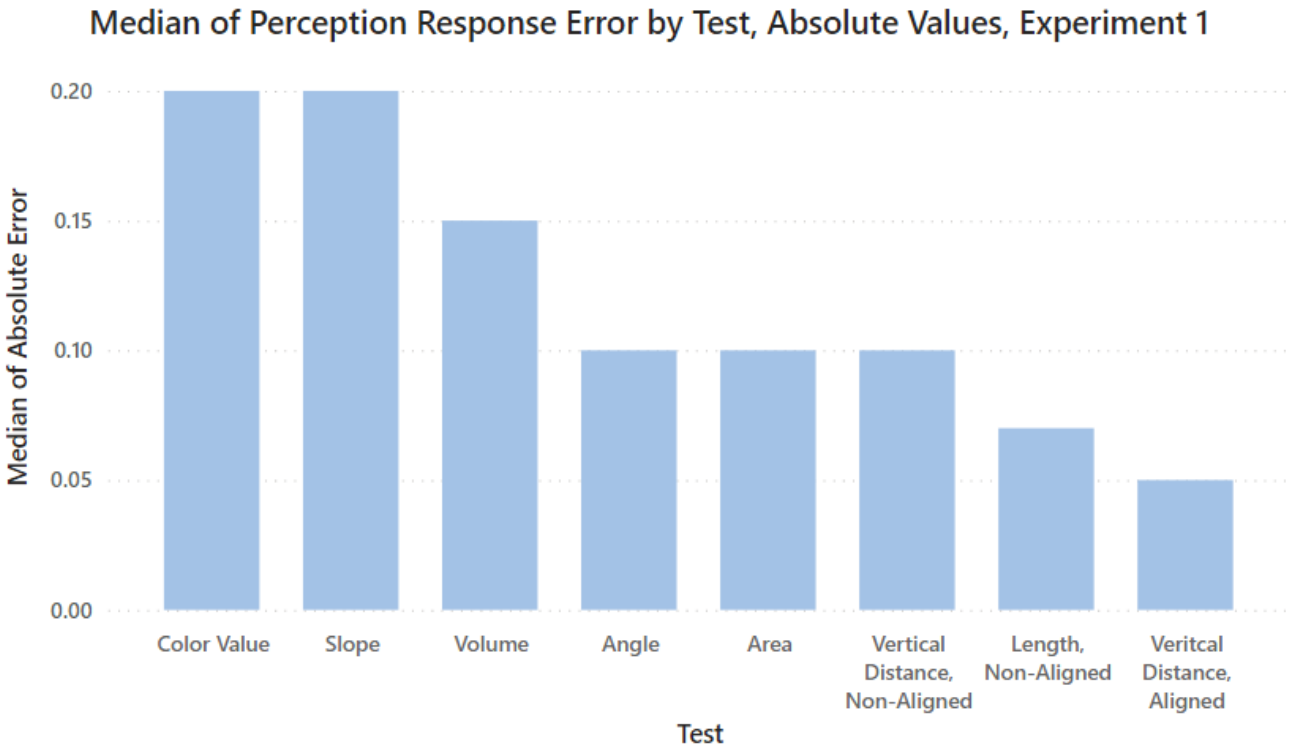
2d. R- made new column using abs(Error) then created a new df with aggregated data filtered by test and created plot using the new dataframe. Adjusted width of the bars looks less clunky.



2d. Tableau- set columns to ABS([Error]) with measurement type Median. Messed with font styles and tick distance



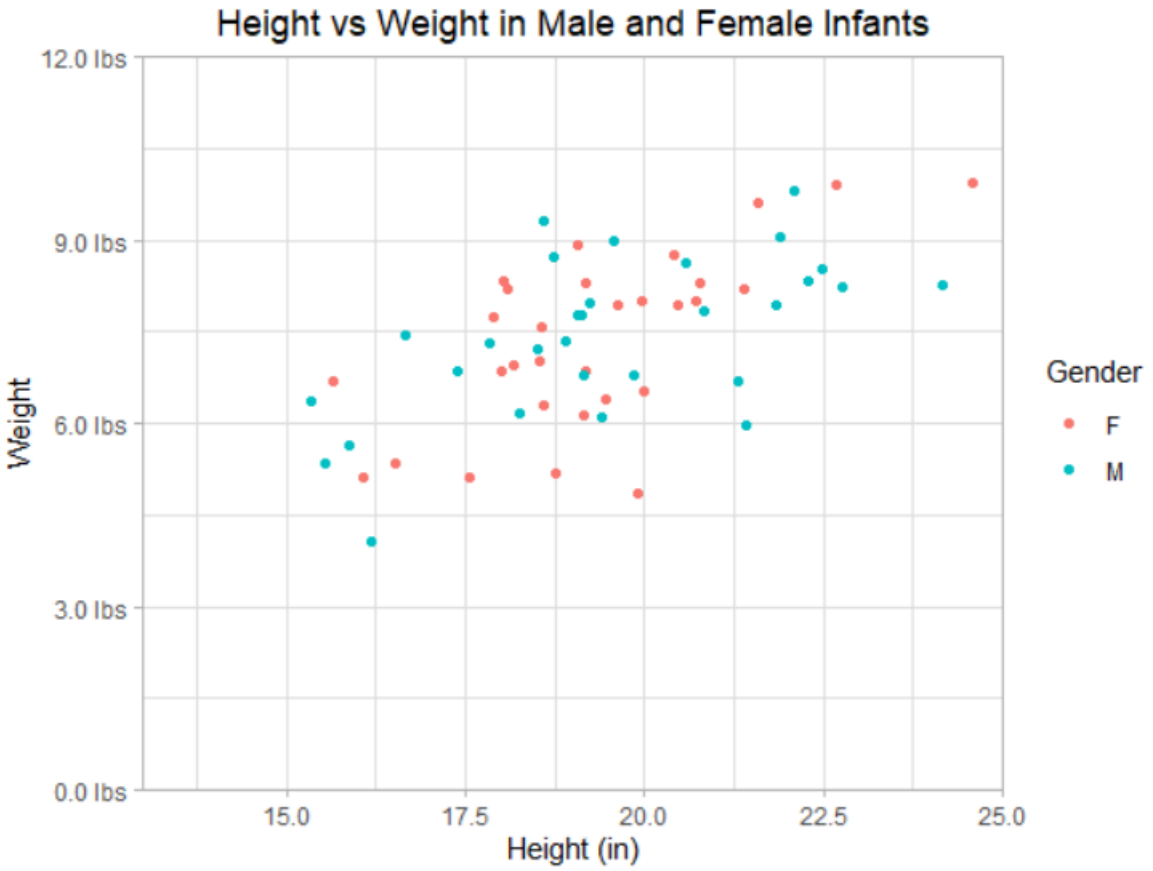
Power Bi – created a new column of type Number.abs([Error]), set the column to "decimal", then made the plot



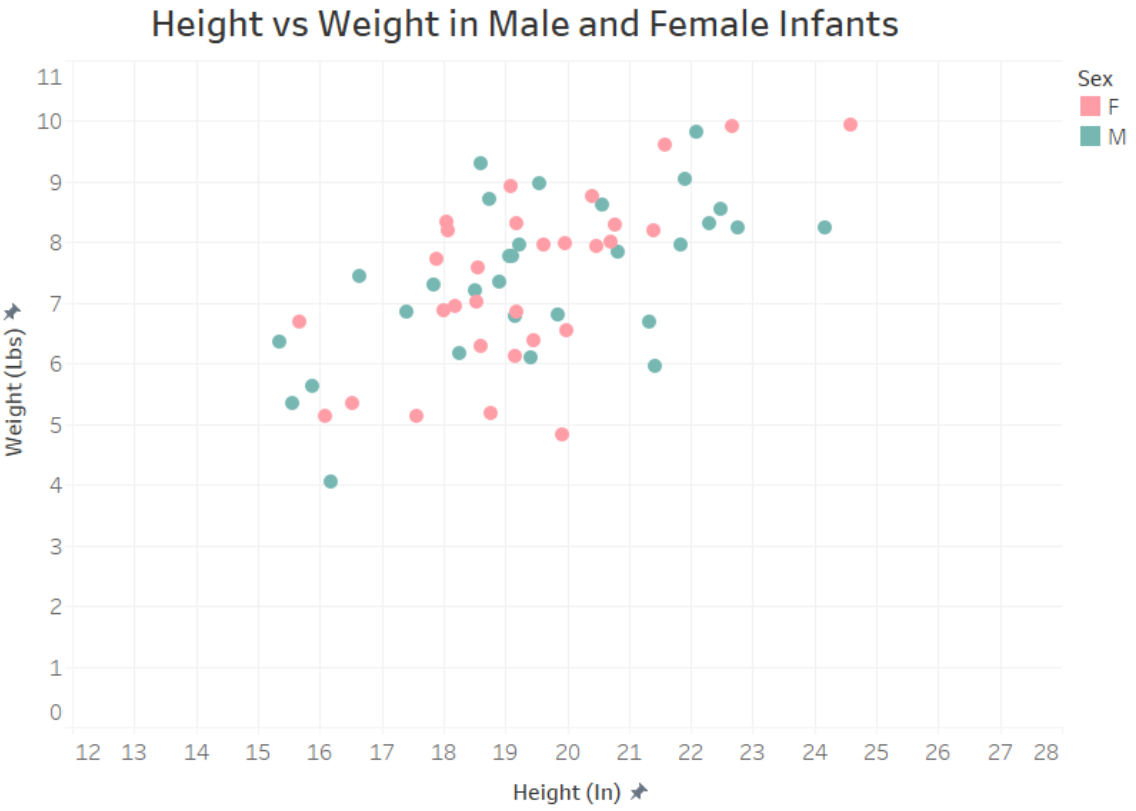
2e. (d): These graphs are less confusing than in 2b, as the axis flip was disorienting to the audience. It is also clearer that color value and slope have near identical error distance.

3a.

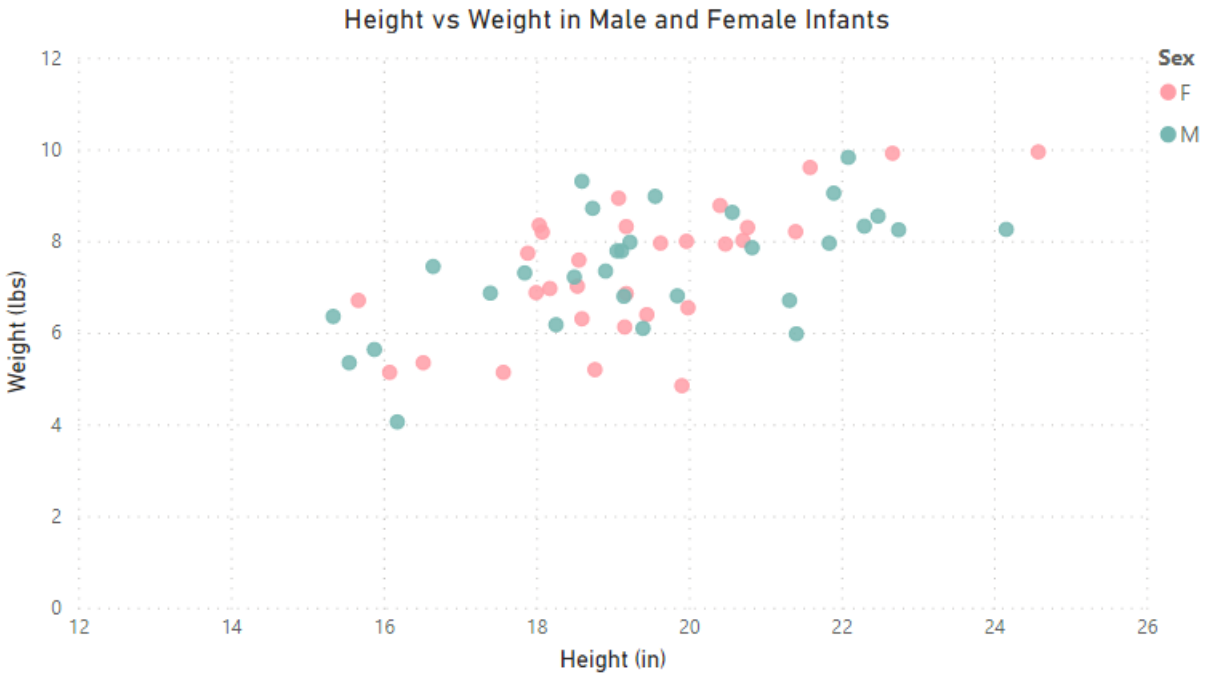
3a. R- Set the scales for x and y to begin further back, with y starting 0,0 in order to display the existence of a trend. Edited the y-axis to add "lbs" to the data ticks..



3a. Tableau – similar axis settings to R. The colors were implementing by dragging "Sex" into the Color Marks section.

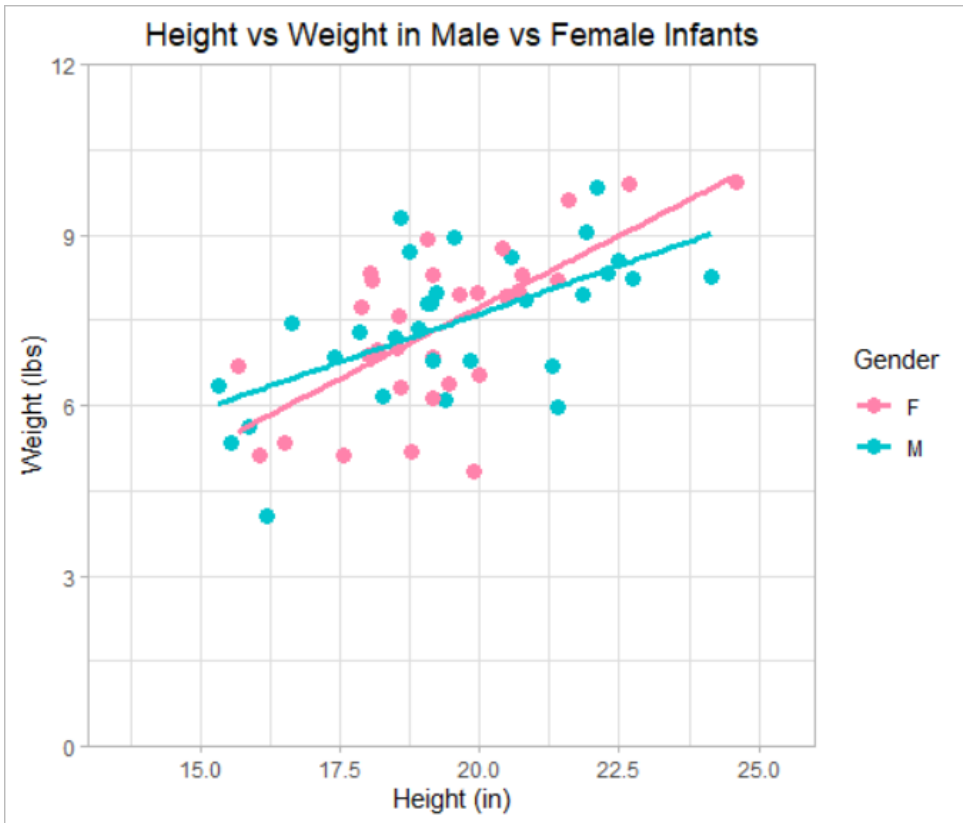


3a. Power BI – placed Weight in Y Axis, Height in X axis, and Sex in Label. Adjusted start and end values for both axis.

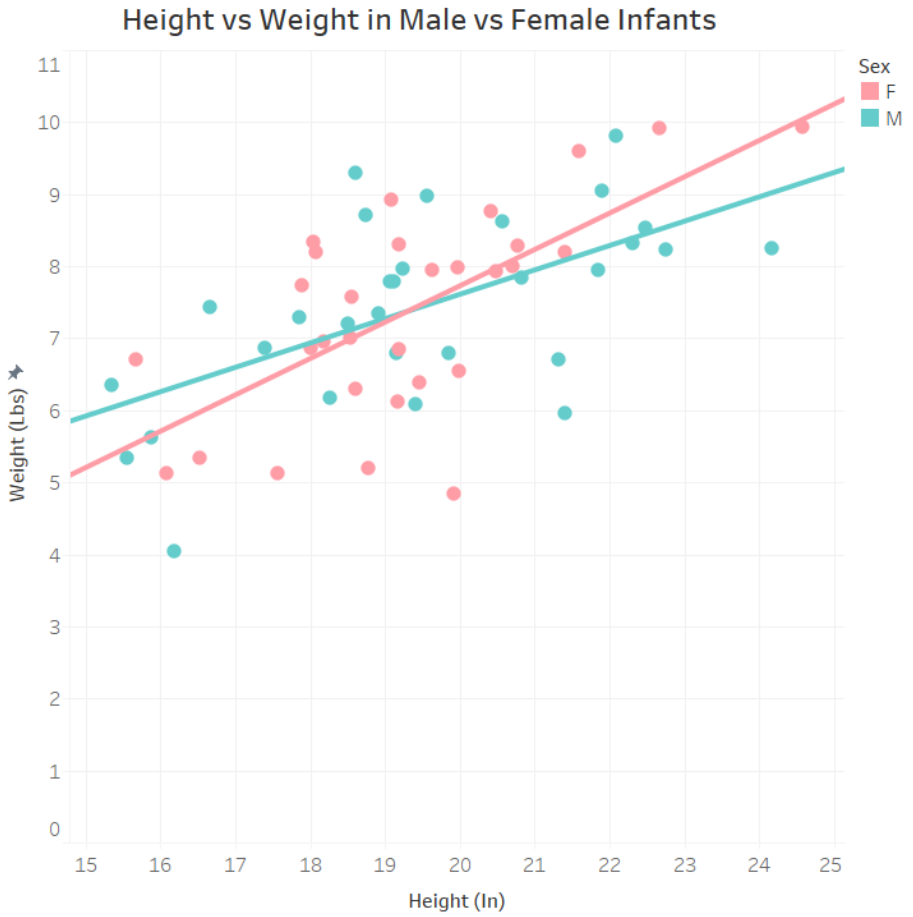


3b.

3b. R- Added treadline, adjusted colors and geom_point sizes.



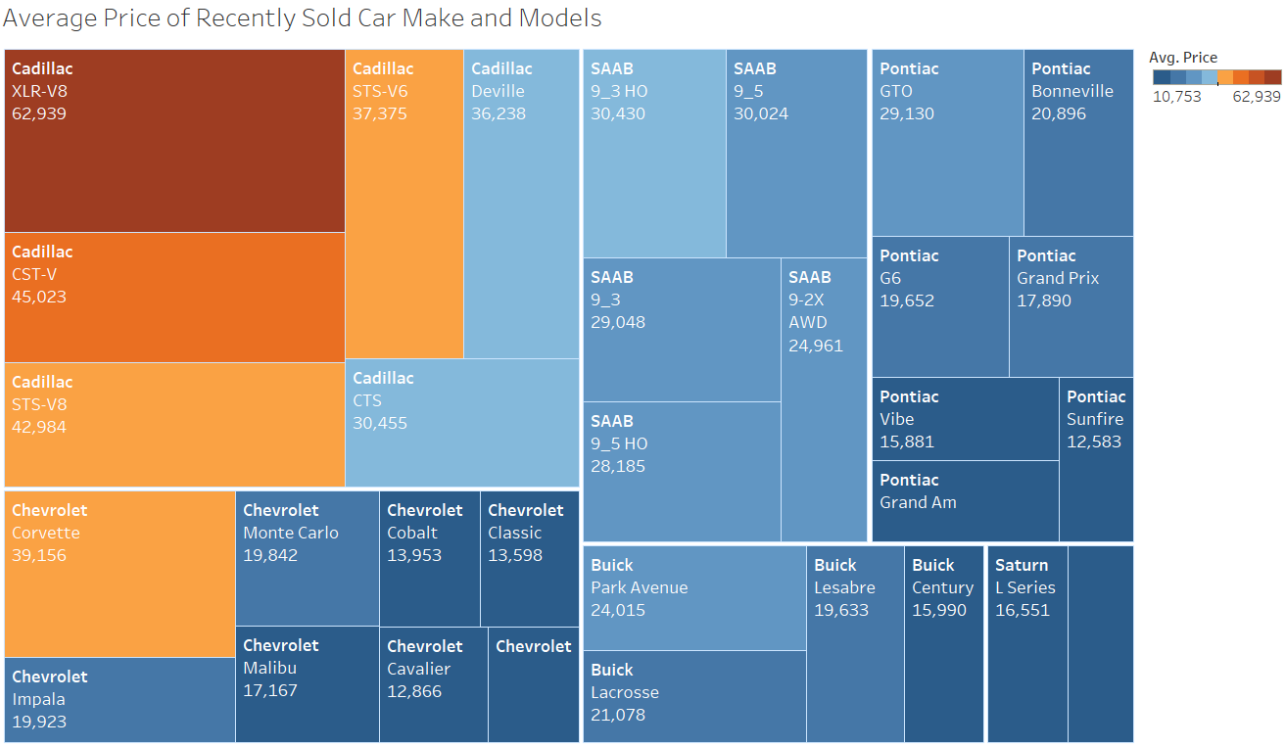
3b. Tableau – Similar to last graph, but with resized treadlines



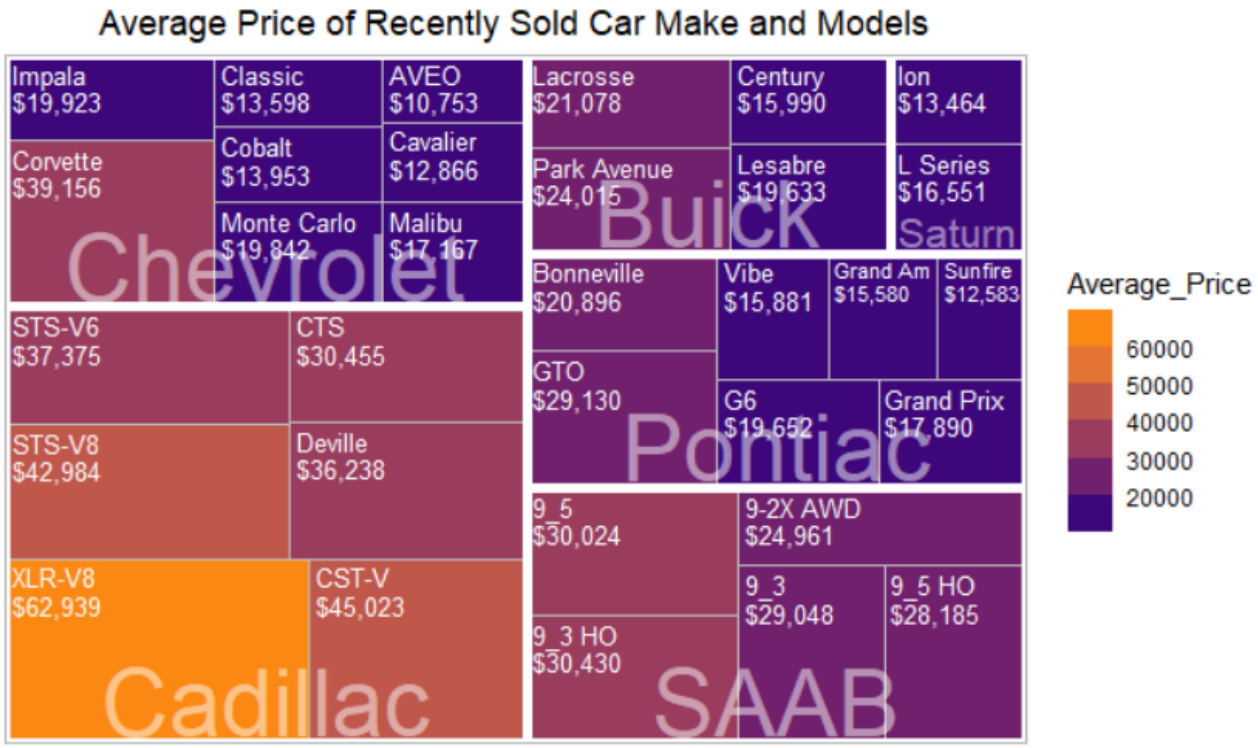
3b. Power BI- Adjusted axis again so that the treads were around 45 degree angles

3c. Firstly, the colors chosen work in greyscale and all 3 types of color blindness. There was a lot of tweaking in the axis, and I determined that starting at 0 on the y axis kept the direction of the scatterplot most legible, showing that there is probably a relationship between height and weight. The graphs are also relatively square, in order to maintain that 45 degree angle.

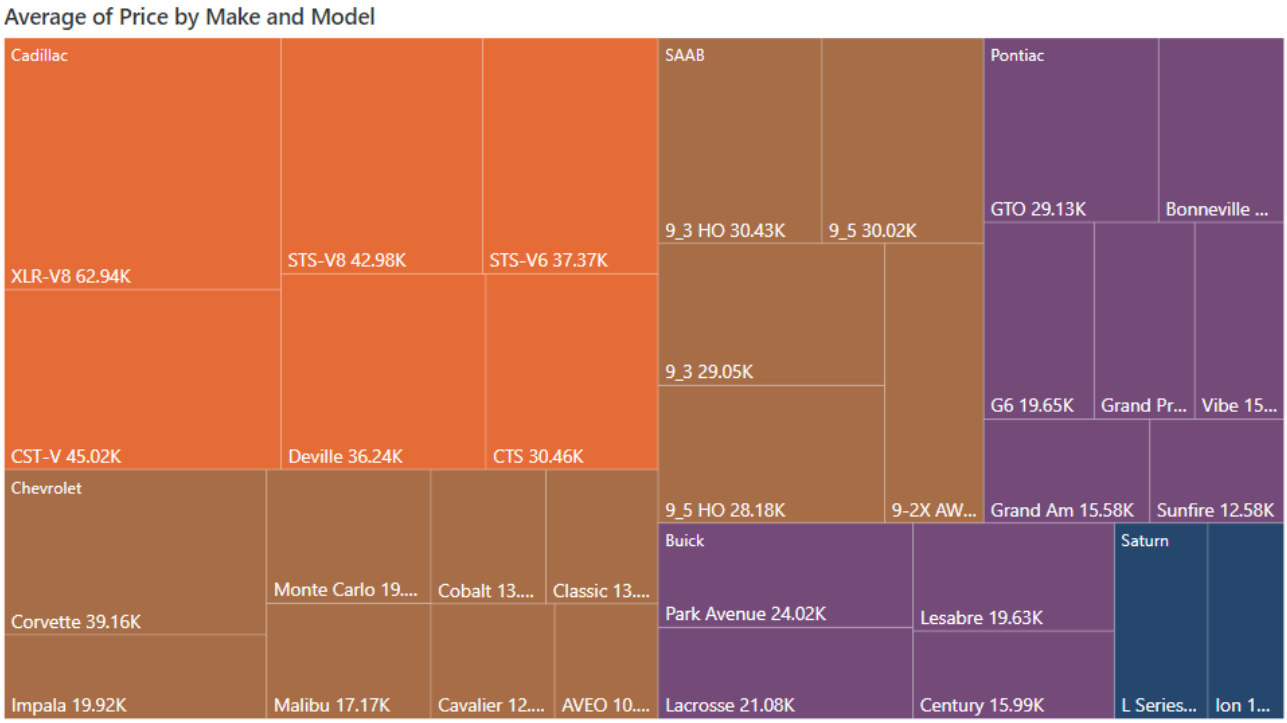
4a. Tableau- Super easy grab and drop, then tweaking to the labels, font sizes, and boldness.



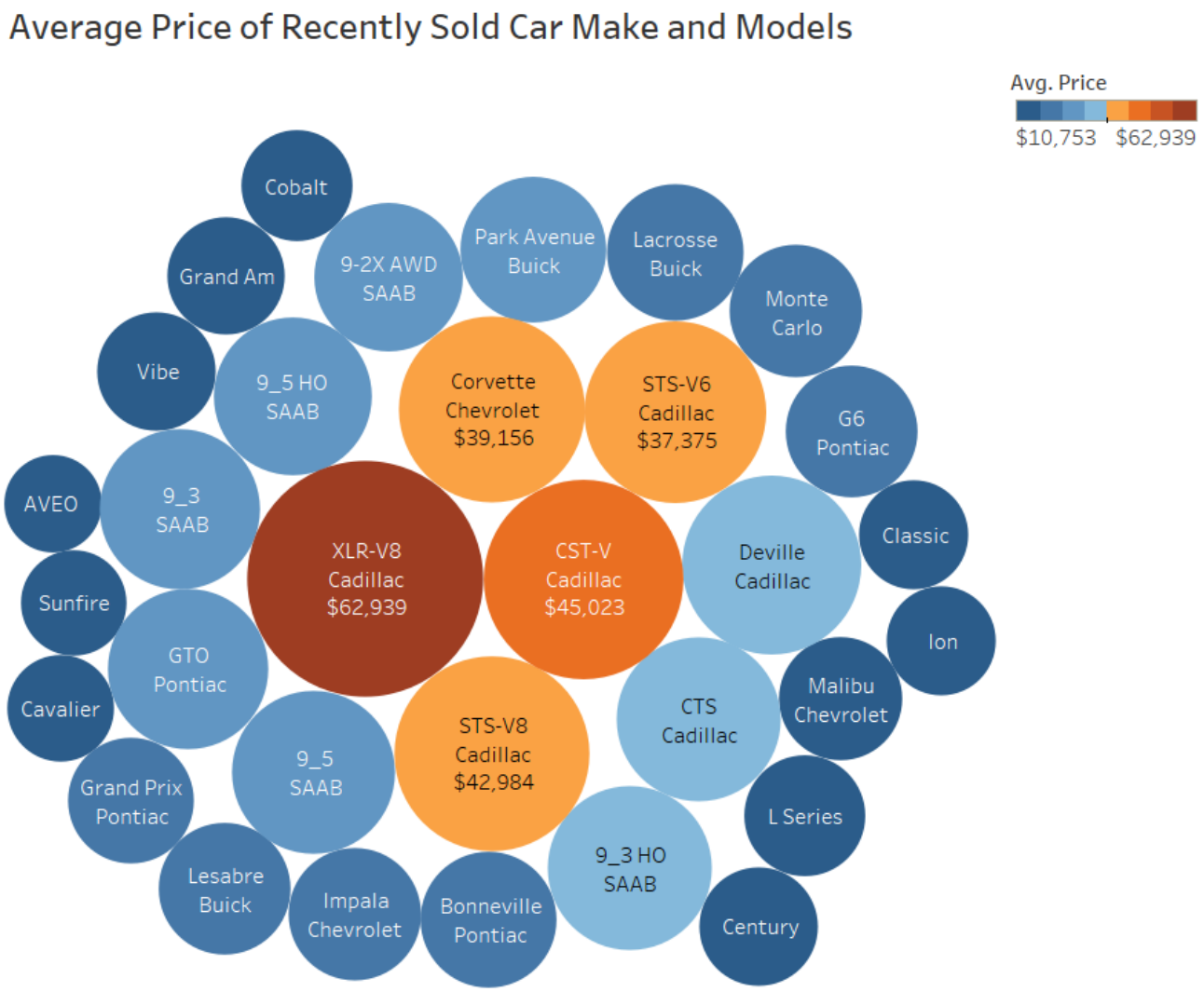
R- Used Tableau's treemap to confirm my piped grouped dataframe was accurate. After piping, a lot of tweaking to the script to make the Make appear large and transparent on the bottom, and set the color scale using ggplot's colors.



4a. Power BI- There just isn't enough functionality to make this treemap beautiful. This was incredibly frustrating to work with, and I never figured out how to set the individual average make price to set the color while still keeping the models grouped in the same make, so we settle for the sizes here

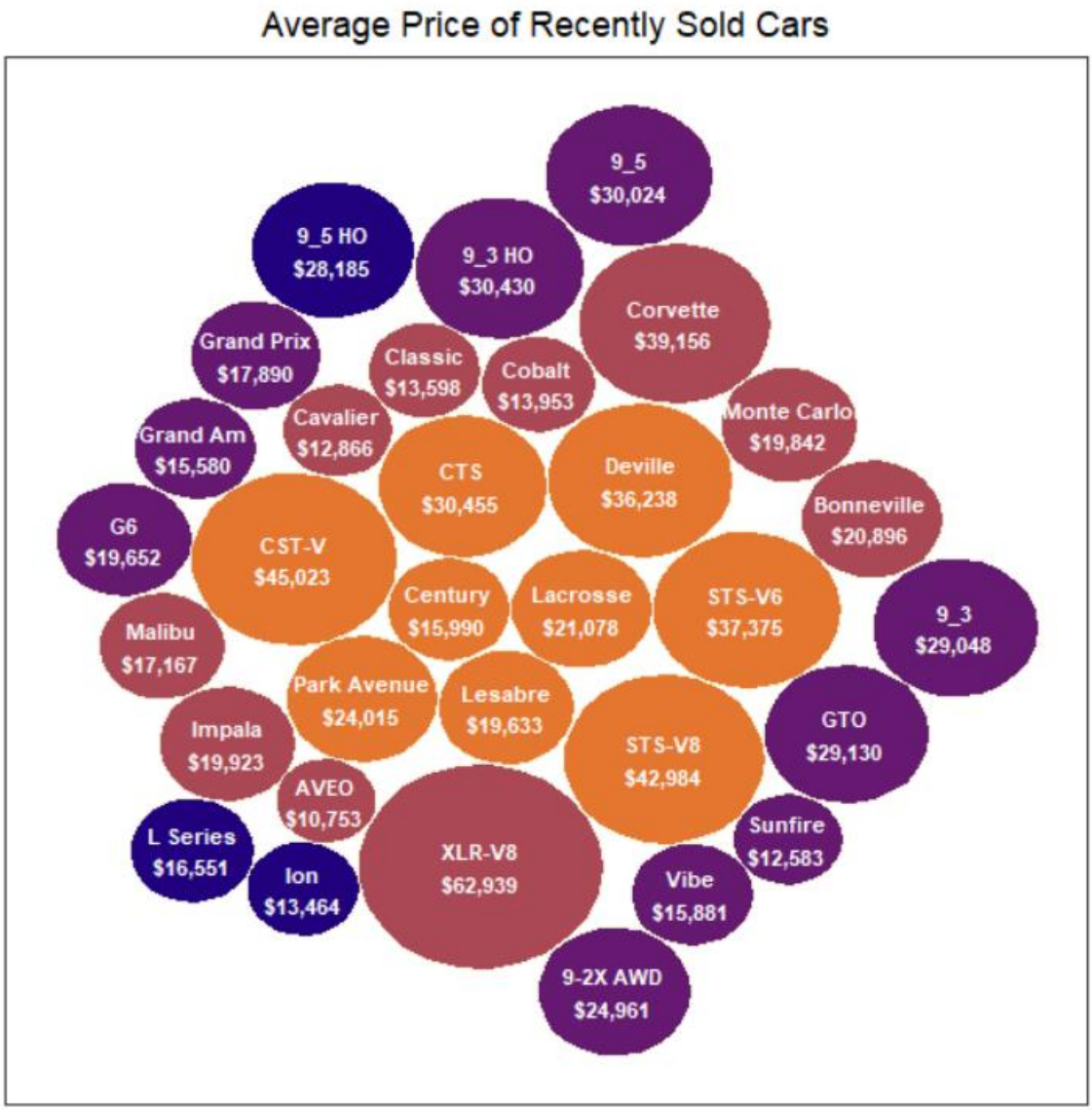


4b: Tableau -Because the assignment asks to observe the price of makes of models, I sorted the text in order of Make, Model, and Price. Figuring that the audience can recognize the make of a vehicle by the model name, this means that the smallest circles only have the make labeled.

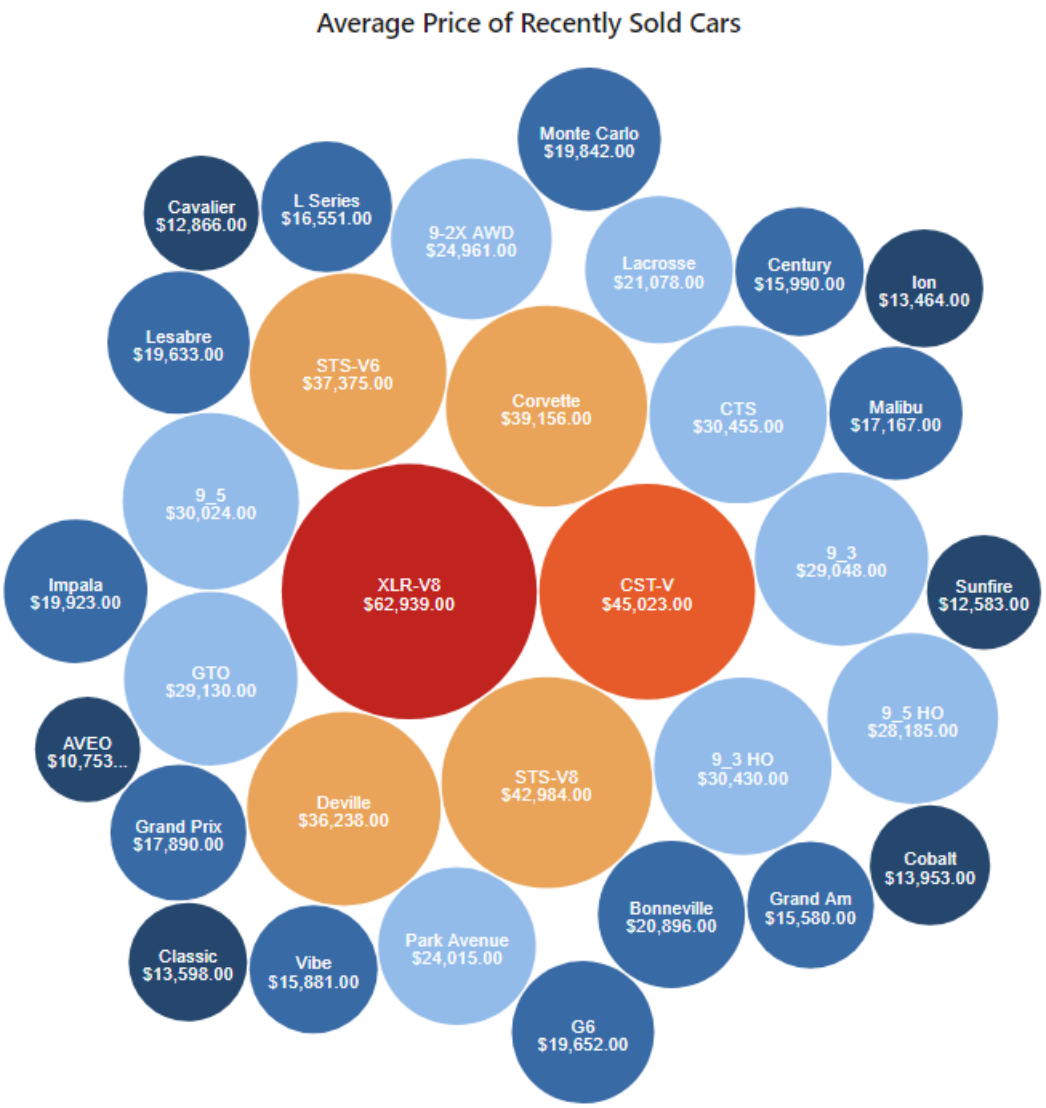


Model, Make and average of Price. Color shows average of Price. Size shows average of Price. The marks are labeled by Model, Make and average of Price.

4b: R Studio The excruciating experience creating this hideously pixelated chart, which does area correctly but does not sort nor fill the circle colors properly just proves that R is not meant to make every type of graph…yet.



4b: Power BI- unfortunately most bubble chart settings are hidden behind a paywall, but by exporting the dataframe for the treemap and bubblechart created in R, I was able to make a decent bubble chart with Power BI.

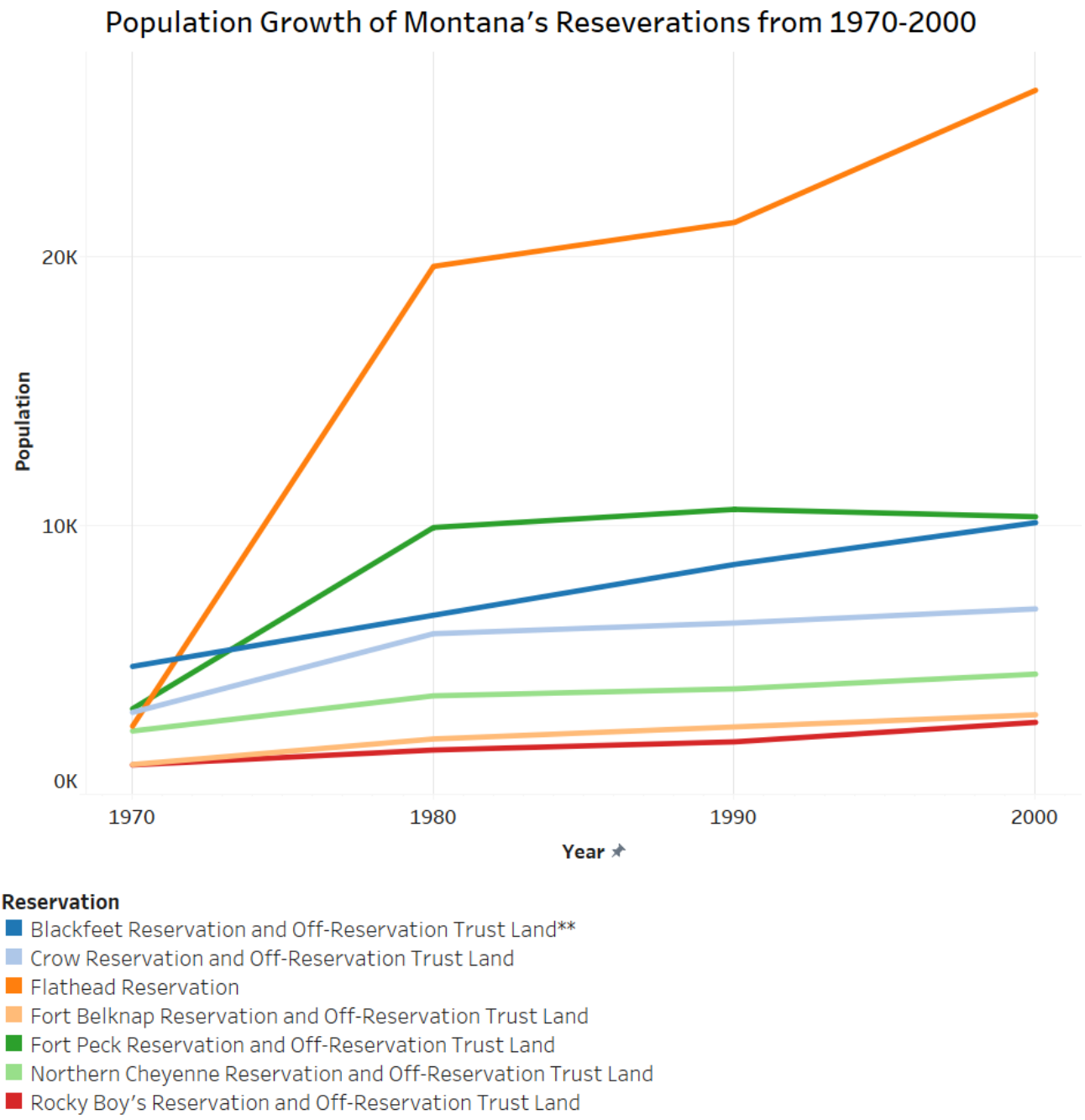


4c: The treemap does a great job of illustrating which brands of cars were cheaper overall. A single glance tells us that Caddys are a tier own tier of expense. To a certain degree, the number of models that appear per Make also suggests what Makes of cars are being sold in general, and would led me to think that Saturns are both cheap and unliked by buyers. As a dealership owner, I would become interested in having more Cadillacs and Chevrolets than Buicks and Saturns.

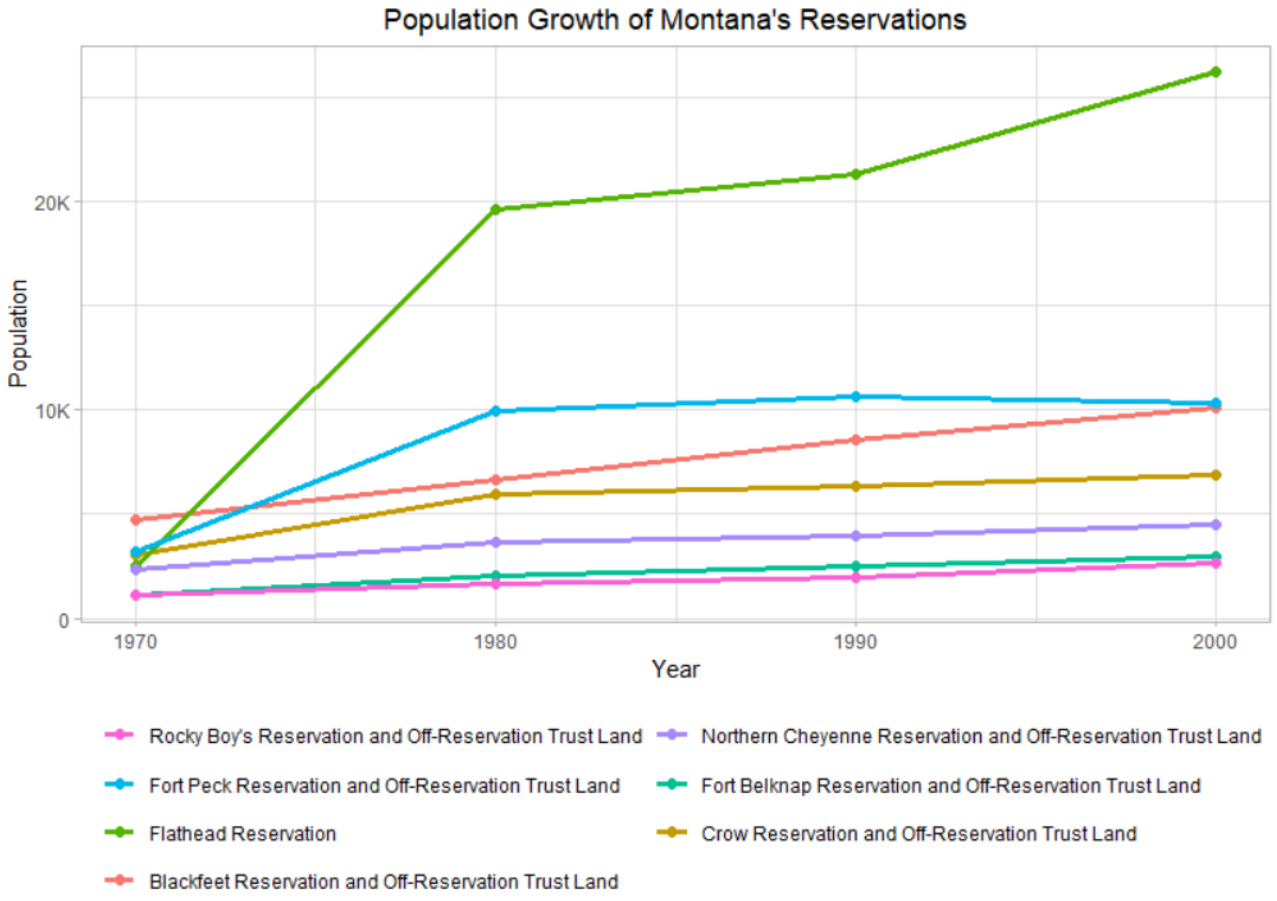
The bubble chart, meanwhile, would probably be more interesting to general audiences to let us know how expensive cars are in general. As a potential buyer, I would utilize the chart to give an idea of what model of cars are in my budget. The order of Model then price especially would be interesting to buyers who don't care for the Brand of car, which probably encompasses most car buyers.

5a.

5a: Tableau- Since this graph was made after watching 3 weeks of lectures, it is more beautiful. Sheet1 was used, and Data Interpreter fixed the header confusion instantly! Pivoting was done as shown in the tutorial, changed 1970* to a whole number in 2 clicks.

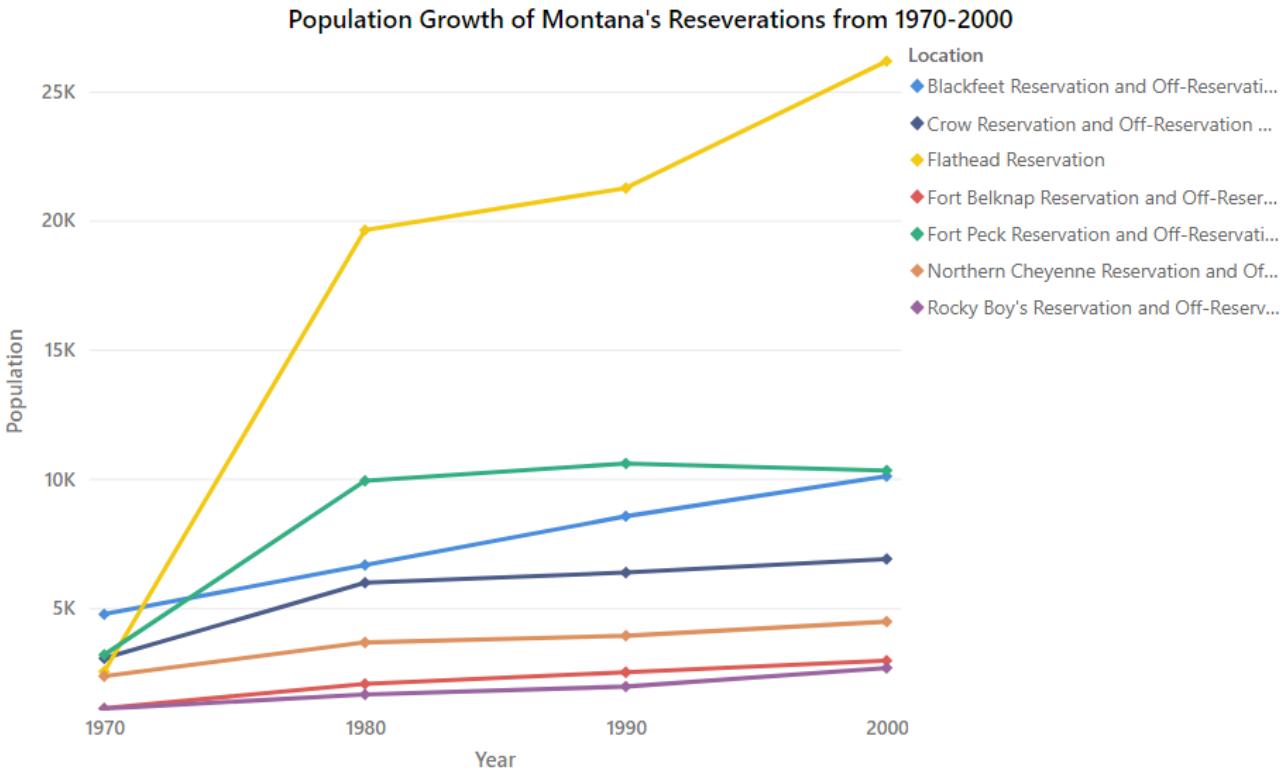


5a: R Studio: A lot went on with R. Sheet2 was used. The column name was changed, pivoting was done, lubridate changed the year column to date, Montana was deleted, the ** was removed, and finally the plot was made.

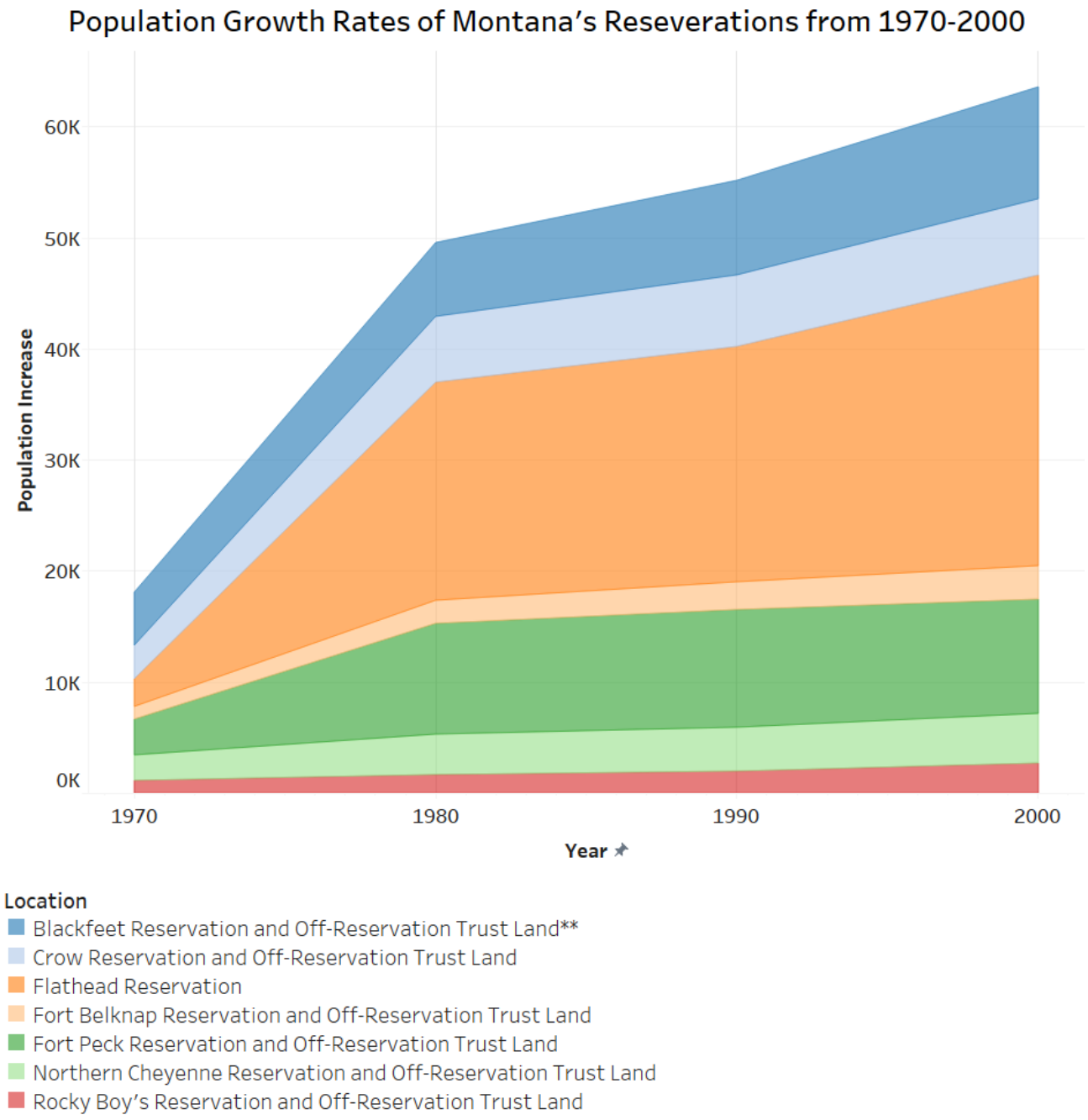


5a: Power BI:

Easiest tool by far. Simply clicked on the year columns and clicked "unpivot table", then transformed to date with another click. Then just removed the top 4 rows (Montana) and used the data from there.



5b: Tableau This is similar to the first graph, with the area chart to showcase which reservations had highest increases at each point. After 3 diligent weeks of work, my brain refuses to recreate charts in R and Power BI anymore.



5c: Tableau I still can't figure out why Tableau betrayed me here and wouldn't wrap the reservation names properly. Otherwise, this was made with Locations as column, population in rows, the Montana filter, year-date as a detail, and added color determined by reservation. The color was added just so that we could get that label box that shows that audience the full names of the reservations, since Tableau refused to show it to us.