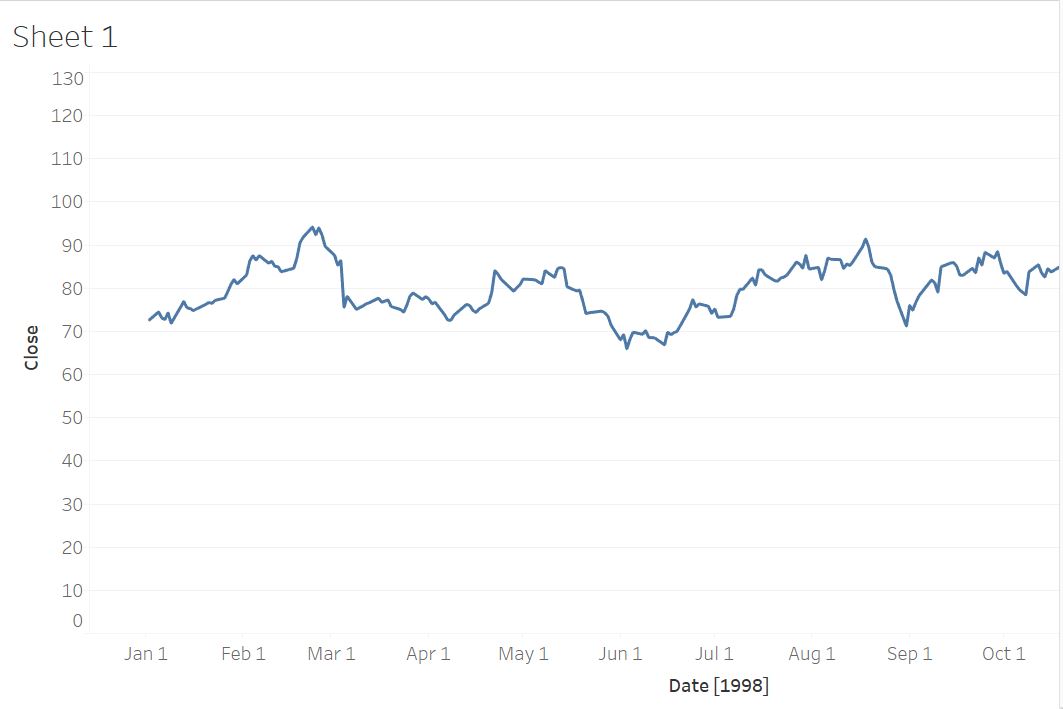
**CSC 465**

1) (20 pts) For this problem, we’ll look at data about Intel stock (Intel-1998 dataset from the zip file posted with this homework). The data covers stock market trading for the Intel corporation in 1998. Each row is a day, with the following columns: Date, Trading Day (integer day number, including skips), Open (price at market open), High (highest price of day), Low (lowest price of day), Close (price at market close), Volume (shares traded), and Adj. Close (adjusted closing price, meaning accounting for stock splits, which are not a problem in this data).

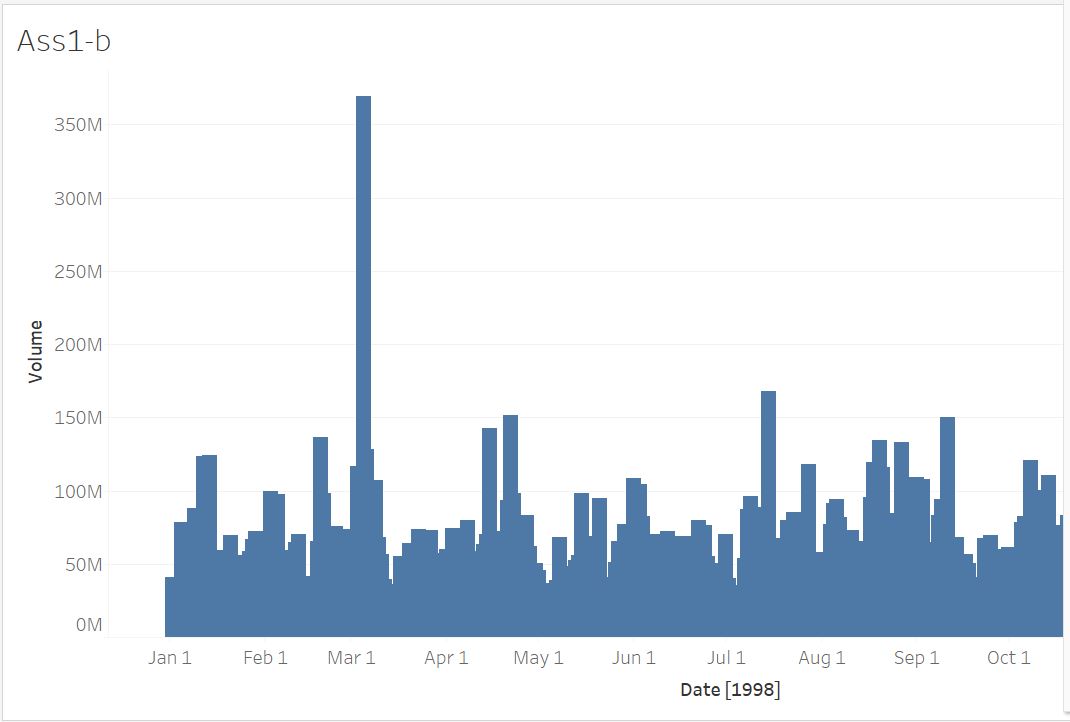
Make the specified graphs in either R or Tableau:

1. Graph the closing price vs. the date with an ordinary line graph. If you use Tableau, you need to right-click on the Date and choose Exact Date from the dropdown menu so that it uses the full date with "day".



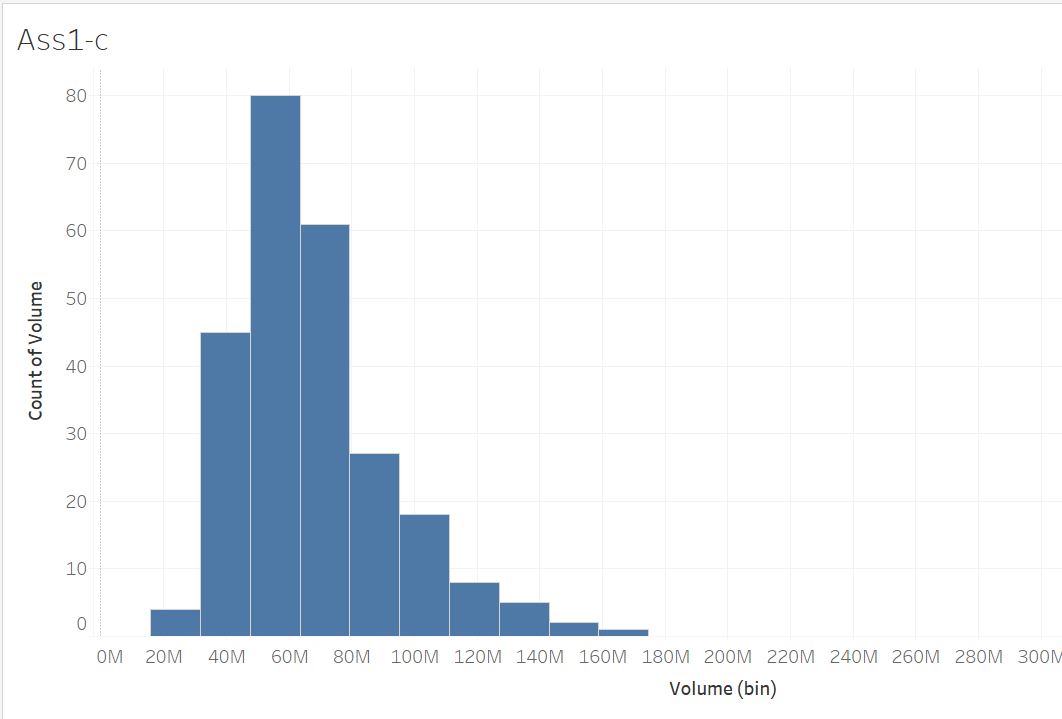
Select the data file in the data source. Placing Date to columns and Close to rows and selecting line chart from the Show Me menu.

1. Graph the Volume vs. the exact Date as in the last part with a bar graph.

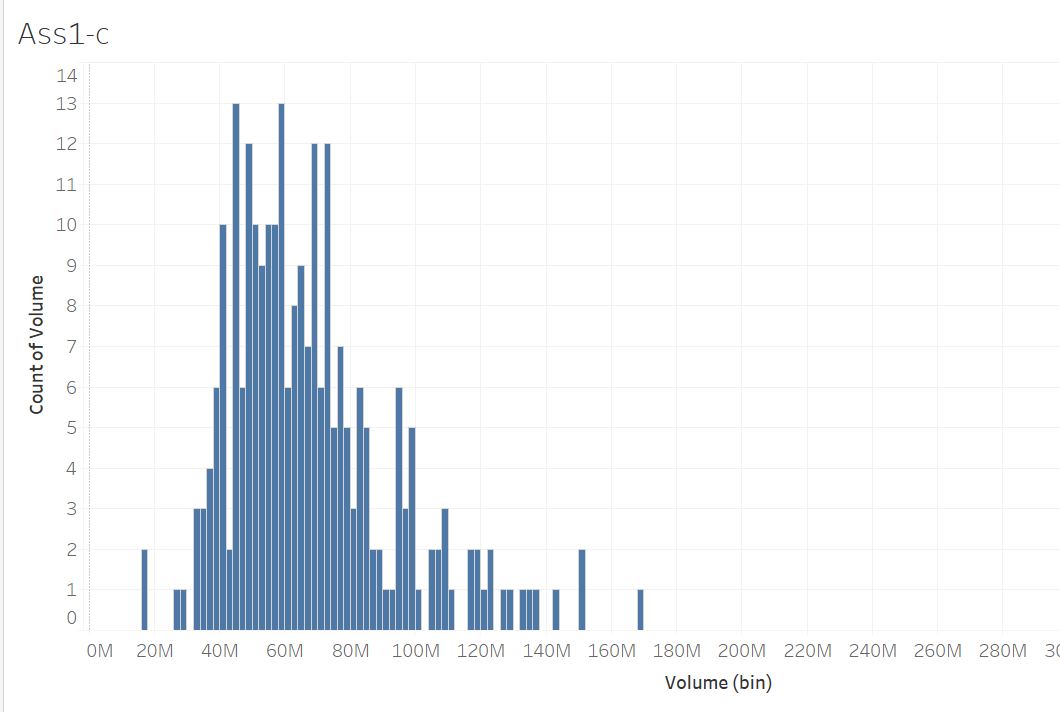


Placing Date in columns and Volume in rows and selecting bar chart from Show Me menu.

1. Create a histogram of the daily stock Volume. R has the hist command and a ggplot geom. In Tableau, the Histogram graph type in the Show Me box will be useful. Experiment with the bin size. It’s an optional parameter in the R functions (e.g. n=20 for hist or bins=20 for geom\_histogram). In Tableau, after you have the histogram, right click the “Volume (bin)” in the data bar on the far left and use Edit. In Tableau, it’s not the number of bins, but their width (in terms of data). You can set them that way in R as well with different parameters.



Placing Volume in rows and selecting histogram from Show Me menu. We can also select size of the bins.

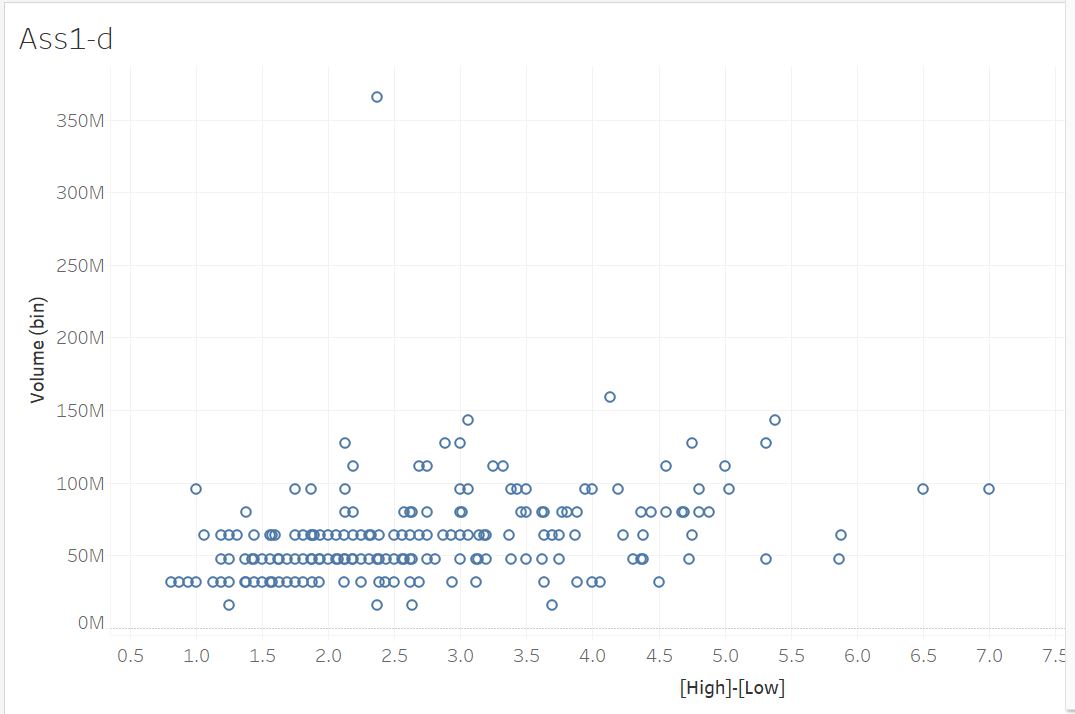


Placing Volume in rows and selecting histogram from Show Me menu, going all the way left and right click on Volume. Right click -> Edit -> Size of the bins. Change the size of the bins there.

d. Create a scatterplot that graphs the Volume on the x-axis and the daily price range on the y-axis. You will need to create an additional column that contains the "range" of the prices for the day as the difference between the fields High and Low.

Range = High – Low

Tableau can do it with a Calculated Field. In R you can do it by making a new column equal to the result from subtracting the two columns. In Tableau, to get a scatter plot, you will need to right click on both the Range and Volume entries in graph and change them to "Dimensions".



In the data source, right click on High -> Create calculated field. Setting field name as Range and value as [High]-[Low].

Go to the sheets.

Placing Volume to rows and Range to column. Right click on both the values and selecting Dimension.

Then selecting scatter plot from Show Me menu.

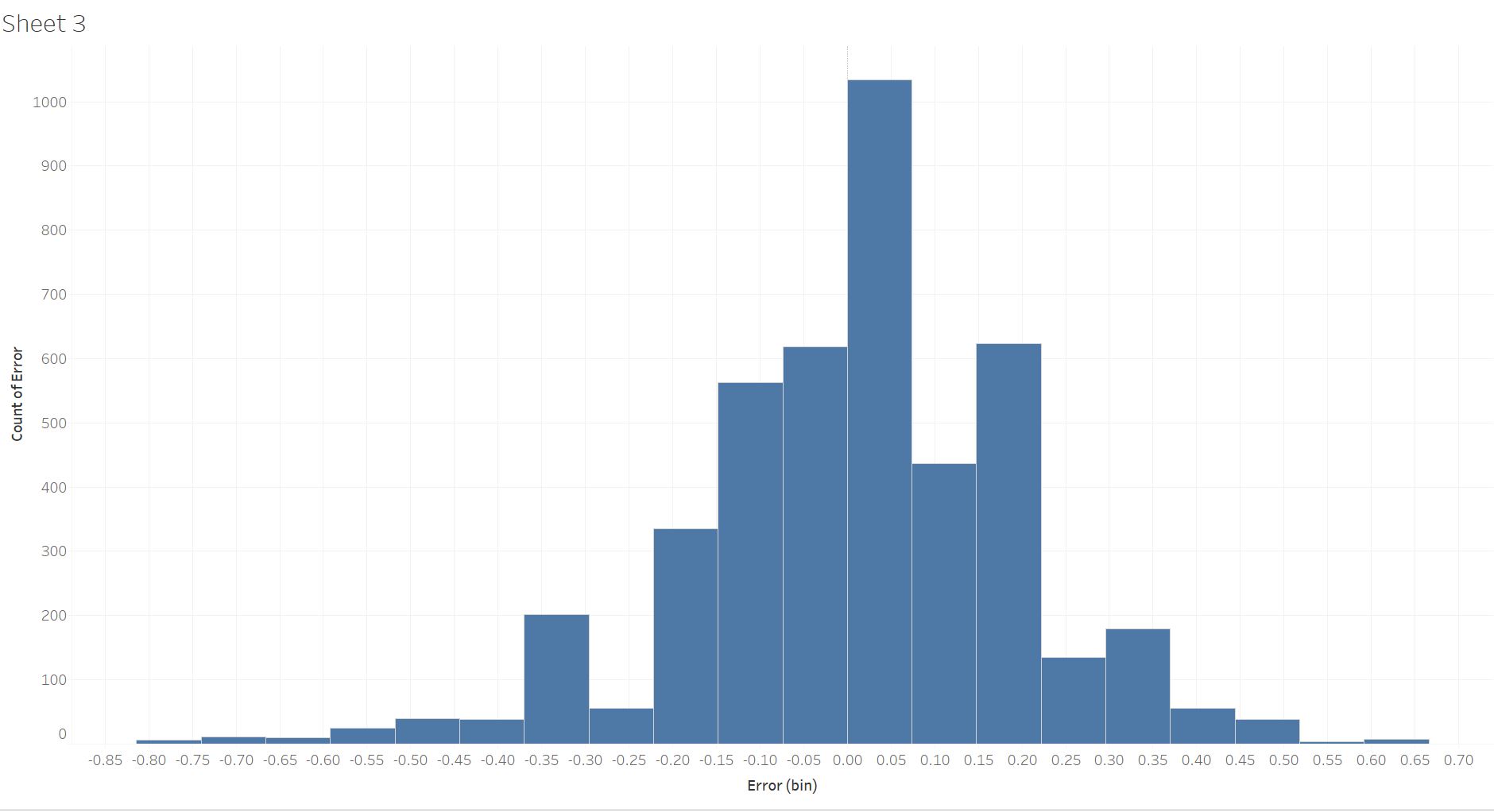
2) (20 pts) We will analyze the perception data collected in class to see how accurate students were at perceiving values with different encodings (aligned bar vs. unaligned bars vs. volume, etc.). Use the PerceptionExperiment.csv data file, which has data from 92 students in previous years’ classes. Remember that you saw a sequence of slides each with four encoded values, marked A, B, C and D. You were supposed to write down the values for B, C and D as a proportion of A.

Here is what the column names in the data file mean: for each Test, i.e. for each type of visual encoding from angle to volume, there were two slides called Displays. Each individual slide, i.e. each Display of each Test, has a unique TestNumber. Each sample that you estimated a value for was labelled B, C or D as its Trial. The Subjects are the students and the estimates they made are the Responses. Each row has a copy of the TrueValue, i.e. the correct value that the student should have entered (if the whole point weren’t how hard it is). The Responses themselves are not very useful for initial visualizations because they will naturally cluster around each True Value. The first thing you will need to do is to create a new column that contains the amount of error. Using the same procedure as in Question 1D, define:

Error = Response – TrueValue

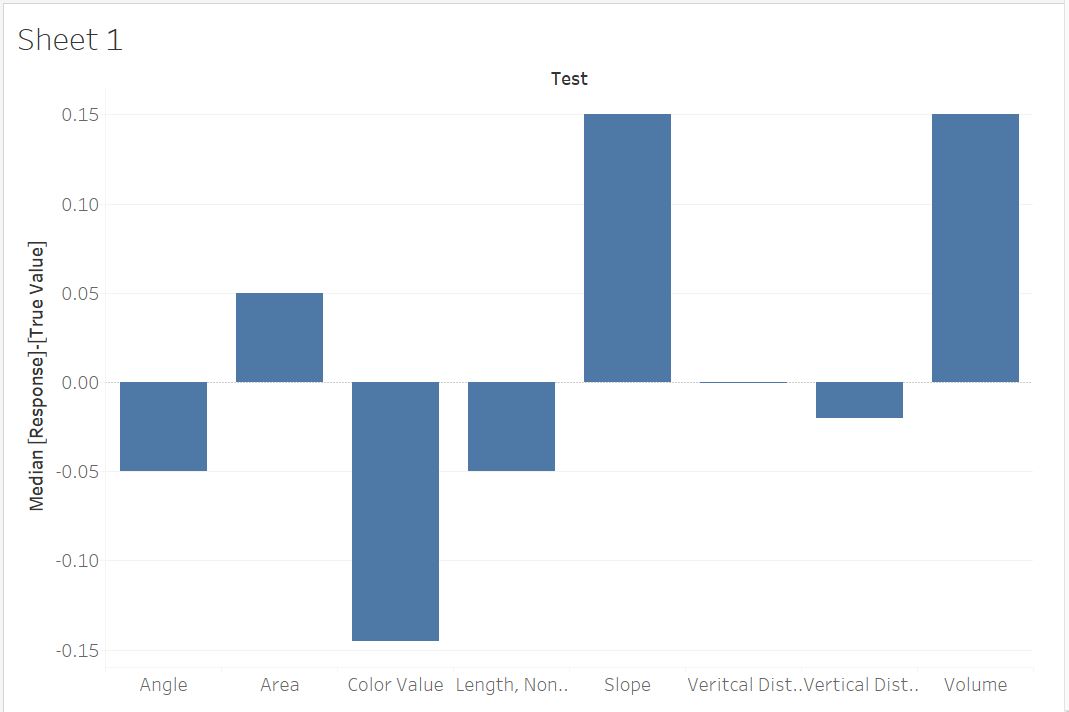
Using either Tableau or R, create the following graphs:

1. A histogram of the overall distribution of Error.



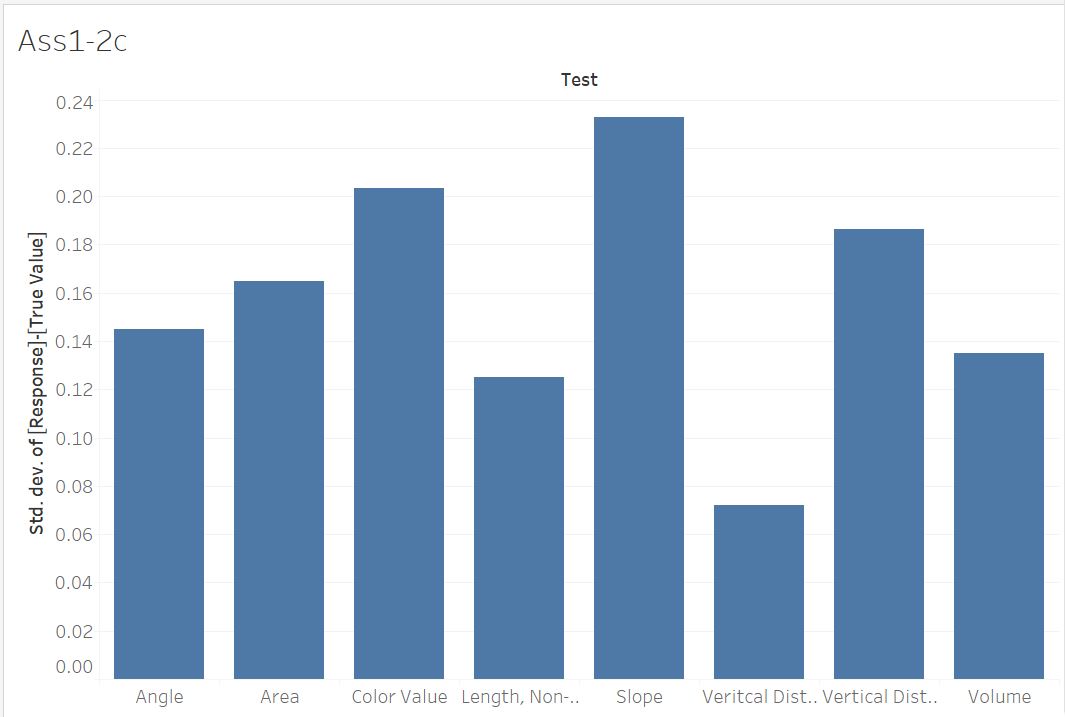
Creating new Calculated field same as explained in 1-d. In sheets placing Error to columns and selecting histogram from Show Me menu.

1. A bar graph of the median Error by Test. Do not subdivide by the display or the trial. Order the x-axis to make the display as clear as possible. Remember, for bar graphs in general, do not necessarily keep the default order (e.g. alphabetical) of the x-axis.



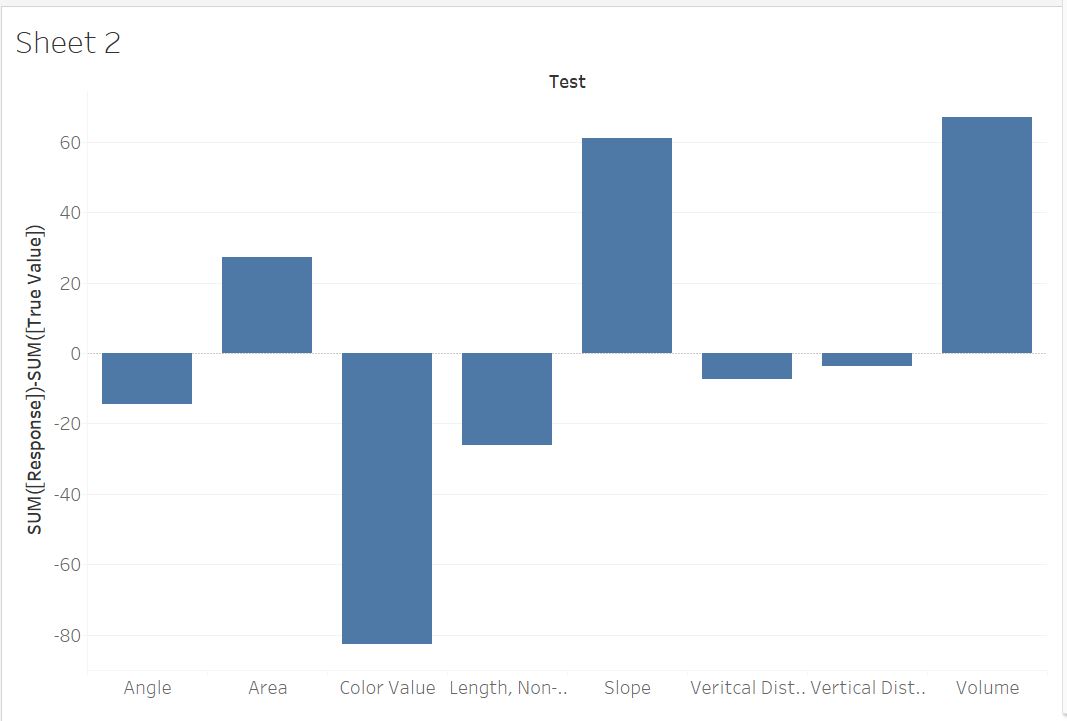
Placing Test to columns and Error to rows and right click on error and selecting Median from measure.

1. A bar graph of the standard deviation of the Error by Test. Remember that this measures the spread of how widely subjects varied in their responses. Again, order the x-axis to make the graph clear.



Selecting standard deviation from Measure.

1. Create a new field called AbsoluteError by computing the absolute value of the Error field you created. Then do the same as in (b) with the AbsoluteError.



For the Absolute Error, Calculating Sum[Response]-Sum[TrueValue] and doing the same procedure as in the last question.

1. For each of the above graphs, explain with a few sentences what the graph tells you.

- Histogram shows number of errors occurred with the use of bins. Each bin indicates number of times an error occurred.

- The second graph shows the Median values of error against Test.

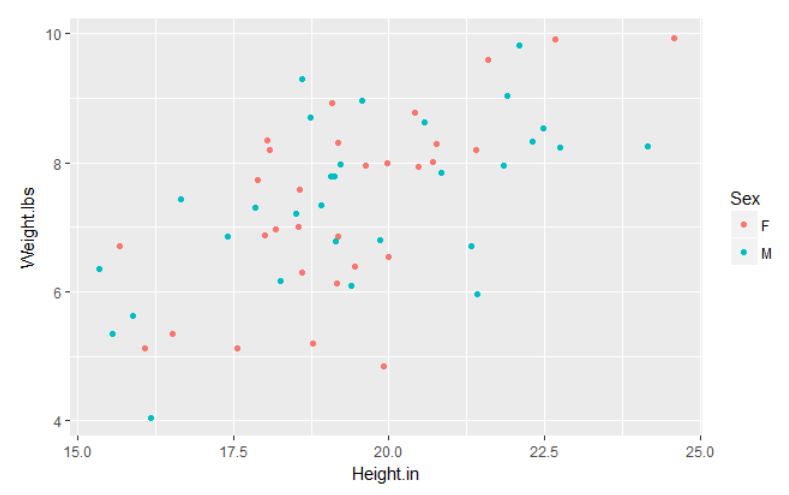
- The third graph shows standard deviation of the error against Test.

- The last graph finds absolute value for each Test case and display it as a bar graph.

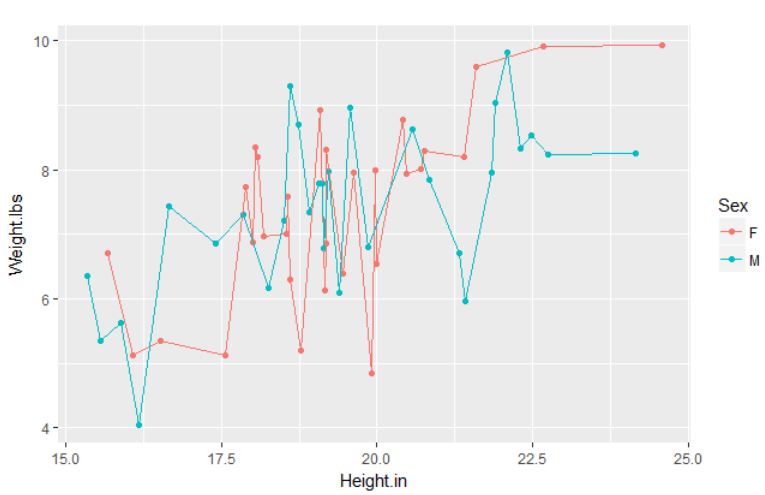
3) (20 pts) Use R for this problem. We will look at data on infant sizes at birth (InfantData.xlsx). There are libraries to help you import the Excel file directly, but in my experience, they are finnicky. The easiest thing to do is open the file with Excel or other compatible software and save it as a CSV file.

Create the following graphs:

1. Graph the data as a scatter plot of Height.in on the x-axis and Weight.lbs on the y-axis. Differentiate in the plot between M or F values for Sex, but graph both on the same plot.



1. Then create another single graph that has separate trend lines for the two populations on the graph. Adjust both the line and data-point weight and color to make the scatter plot and trend lines stand out.

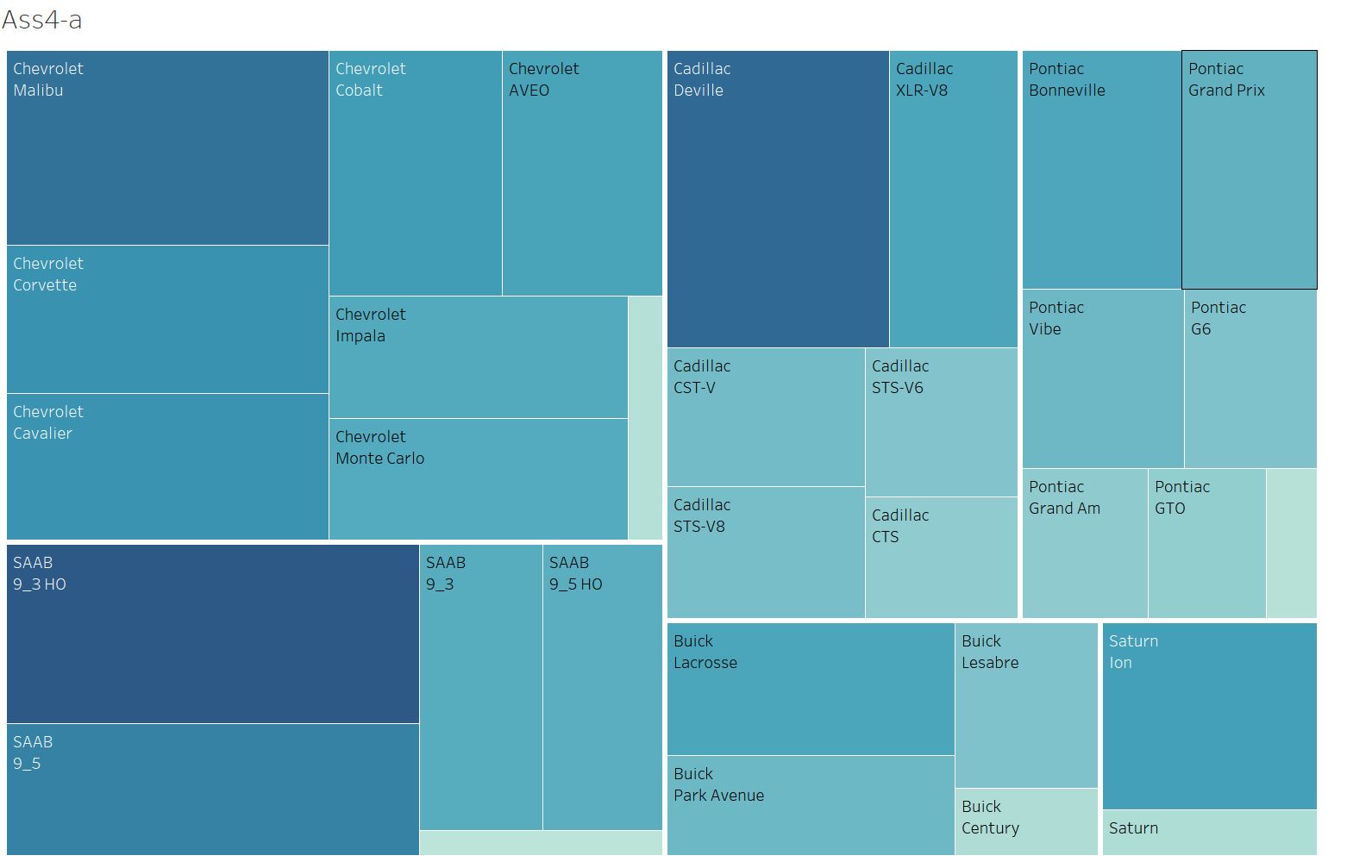


c. Explain in a short paragraph the decisions you made here and their impact on the visualization.

See the R examples from the first lab for reference.

4) (20 pts) Use Tableau for this question. Open the GM cars dataset included with this assignment (gmcar\_price.txt). Each row represents a different car that was sold and includes information about features like the mileage and the price of sale. Create the following plots (we will look more closely at their meanings and design criteria later, but do the best you can to make them readable). Hint: use the “Show Me” menu.

a. A treemap based on Price with a main subdivision for the Make of the car and a minor subdivision based on the Model. Because each row of the data file represents a single car but each box in the treemap represents all the cars with a given make and model, pay very close attention to what kind of aggregation is being used.



Select data set from the data source and then in the sheets, placing Make, Model and Price and selecting TreeMaps from the Show Me box.

1. A packed bubble chart of the same type.



Selecting bubble chart from the Show Me box.

Write a short paragraph discussing the differences between the two plots. Describe for each something that displayed more clearly than with the other.

* In the above two charts, Data displayed more clearly in the Tree maps as it shows car model by make. So that the car models with the same company are sorted together. It can be identifying easily. And in the bubble chart according to my understanding, the car models are displayed according to Price from inside to outside.

5) (20 pts) This problem works with a dataset containing the population of Montana and of each of the 7 Native American reservations within it (reservation70-00.xlsx). There is a measurement for each decade between 1970 and 2000. Sheet1 has the original data. We will use Tableau for this question, but Sheet1 has a header that confuses Tableau. If you’re interested, check out the “Data Interpreter” feature in Tableau to learn how to deal with this. Otherwise, use Sheet2, where I’ve removed the header.

Even with the header and footer removed, you will have to transform the data by

1. Renaming the 1970\* field so it has no \* and can be converted to a number

2. “Pivoting” the year fields in a similar manner to the example in the first lab.

3. Changing the name of the pivot fields to Year and Population, and changing the type of the year field to “whole number”.

4. You can also hide the “Percent Change” field as it only contains information for change over the entire period, not per decade.

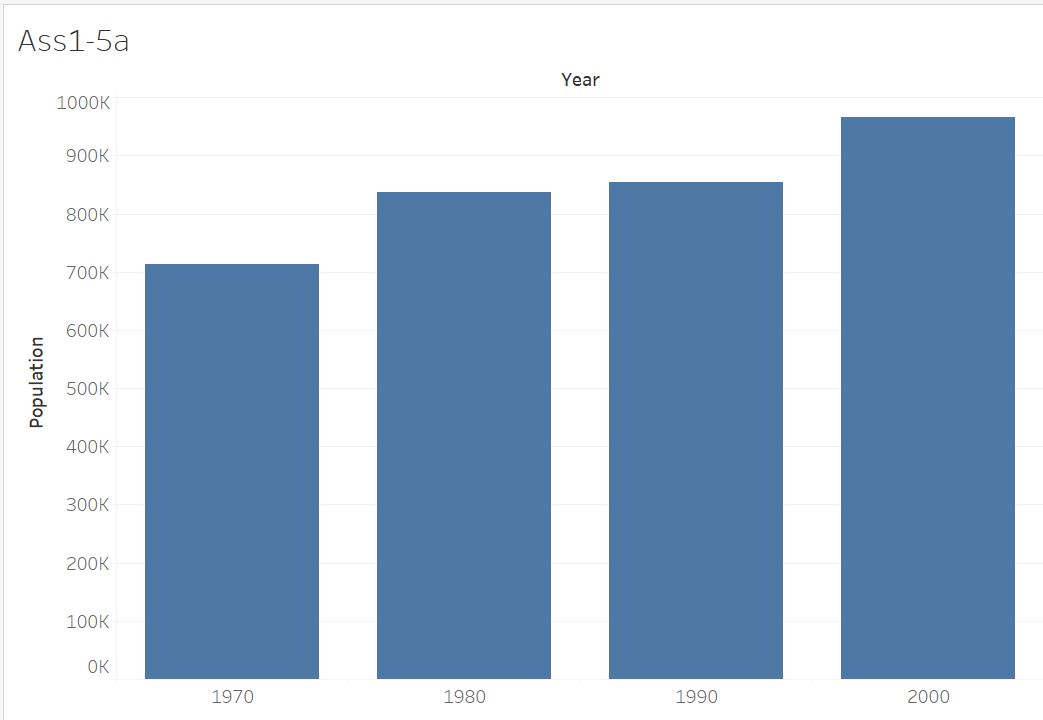
5. If you would like to have an actual Date field for the Year, you need to create a “Calculated Field”. It should construct a Date using the Year, i.e. make a Date field that is on January 1 of the specified year:

makedate([Year], 1, 1)

6. We are not interested in the Montana population, only the reservation populations. When you have used Location on your graph, you can right mouse click (or click the down arrow within it) to apply filters. You can also use “Exclude” from the right click menu on the legend just below the “Marks” configuration.

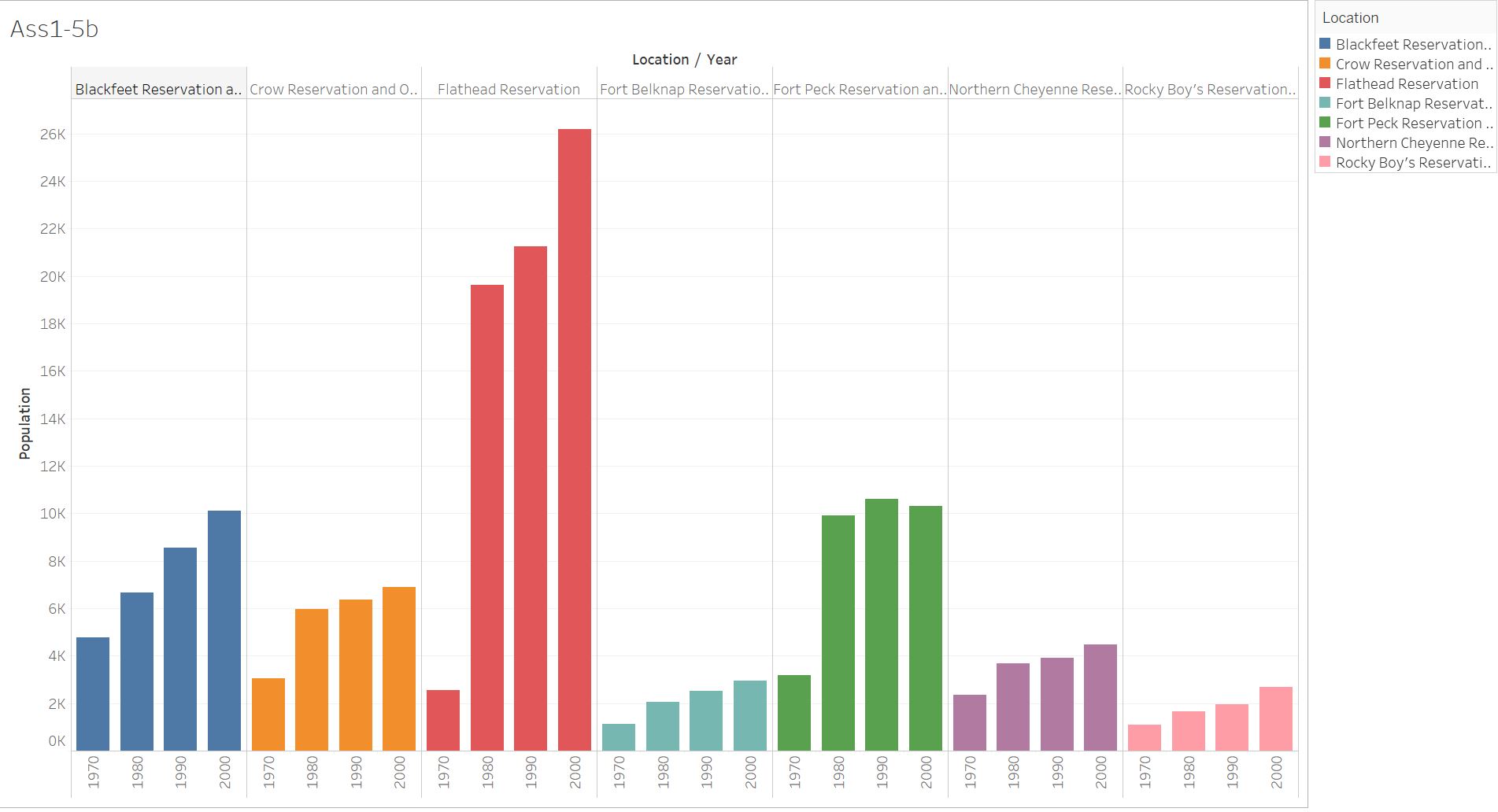
Create graphs to show the following information, using appropriate graph types:

1. One chart that graphs the population growth over the years for the individual reservations.



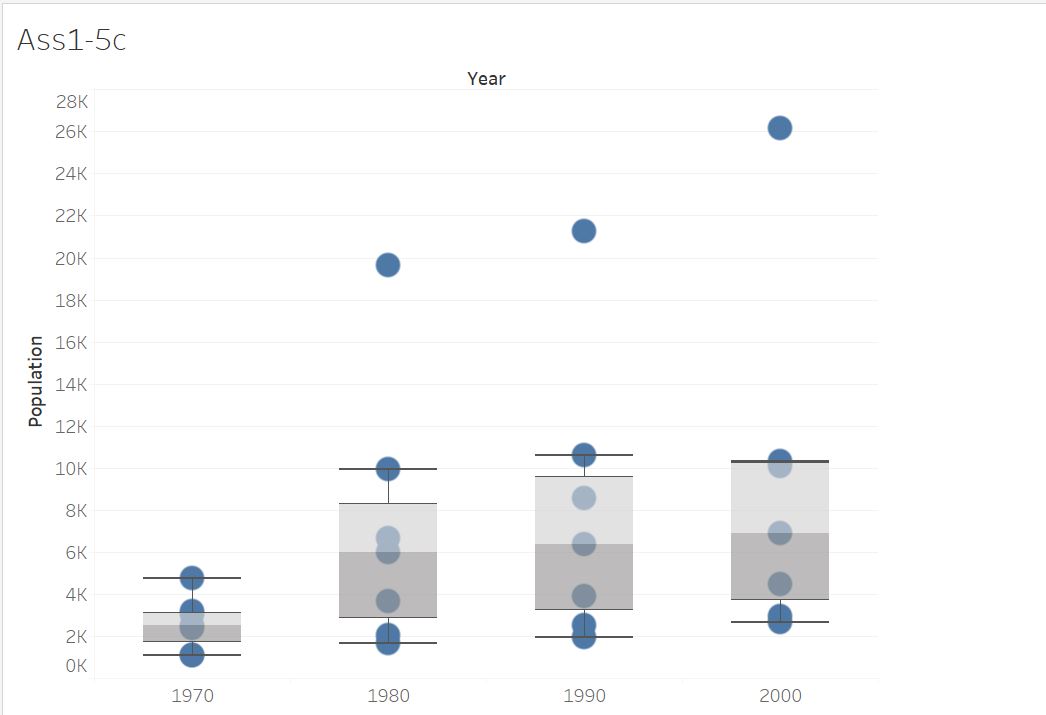
Selecting bar chart for the Show Me box.

1. One that graphs the total reservation population subdivided among the different reservations for each year.



Selecting bars from the Show Me box.

1. One that graphs the population distribution over years for each reservation with a box-and-whisker plot.



Selecting box-and-whisker plots form the show me box.

Make sure that the graphs are properly labeled and that the axis scales properly reflect the type of data represented.