

Umair Chaanda
DSC 465 Data Visualization
Assignment-1

Problem 1 (15 points): For this problem, we'll look at data about Intel stock (Intel-1998 dataset from the website). 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).

a) Closing price vs. the date with an ordinary line graph

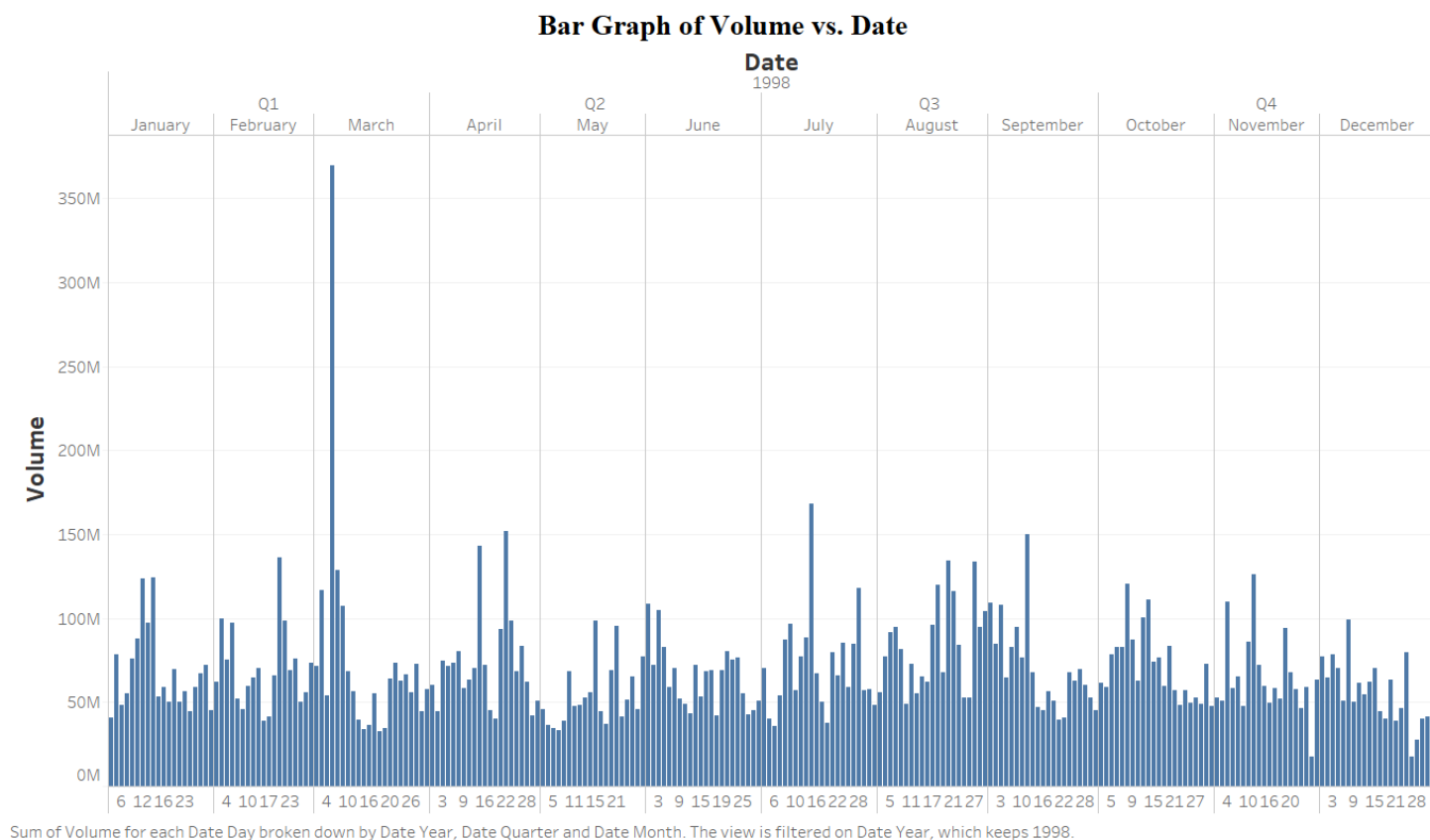
Intel Stock Prices for the Year 1998

Line Graph of Closing Price vs. Date

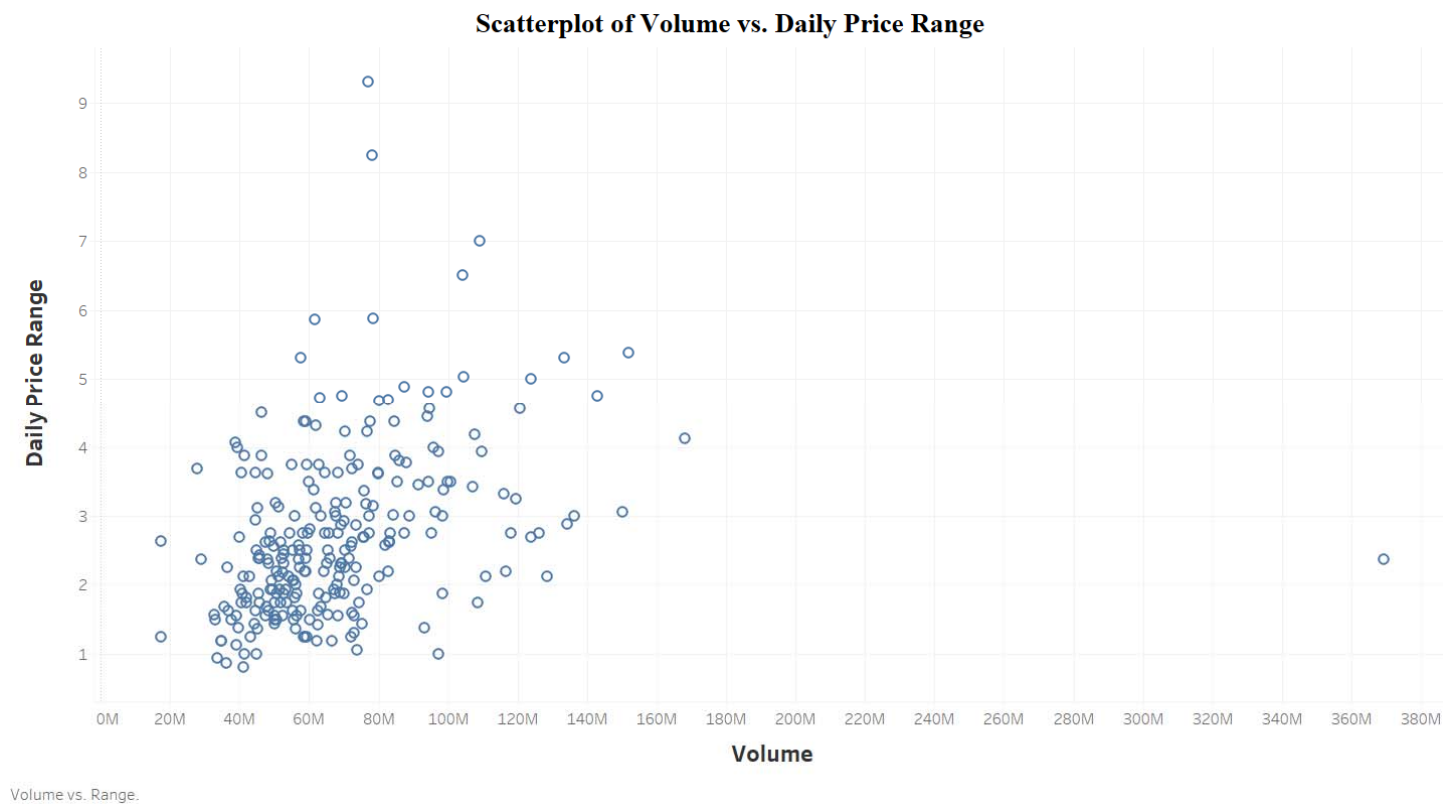


The trend of sum of Close for Date Day.

b) Volume vs. the exact Date as in the last part with a bar graph



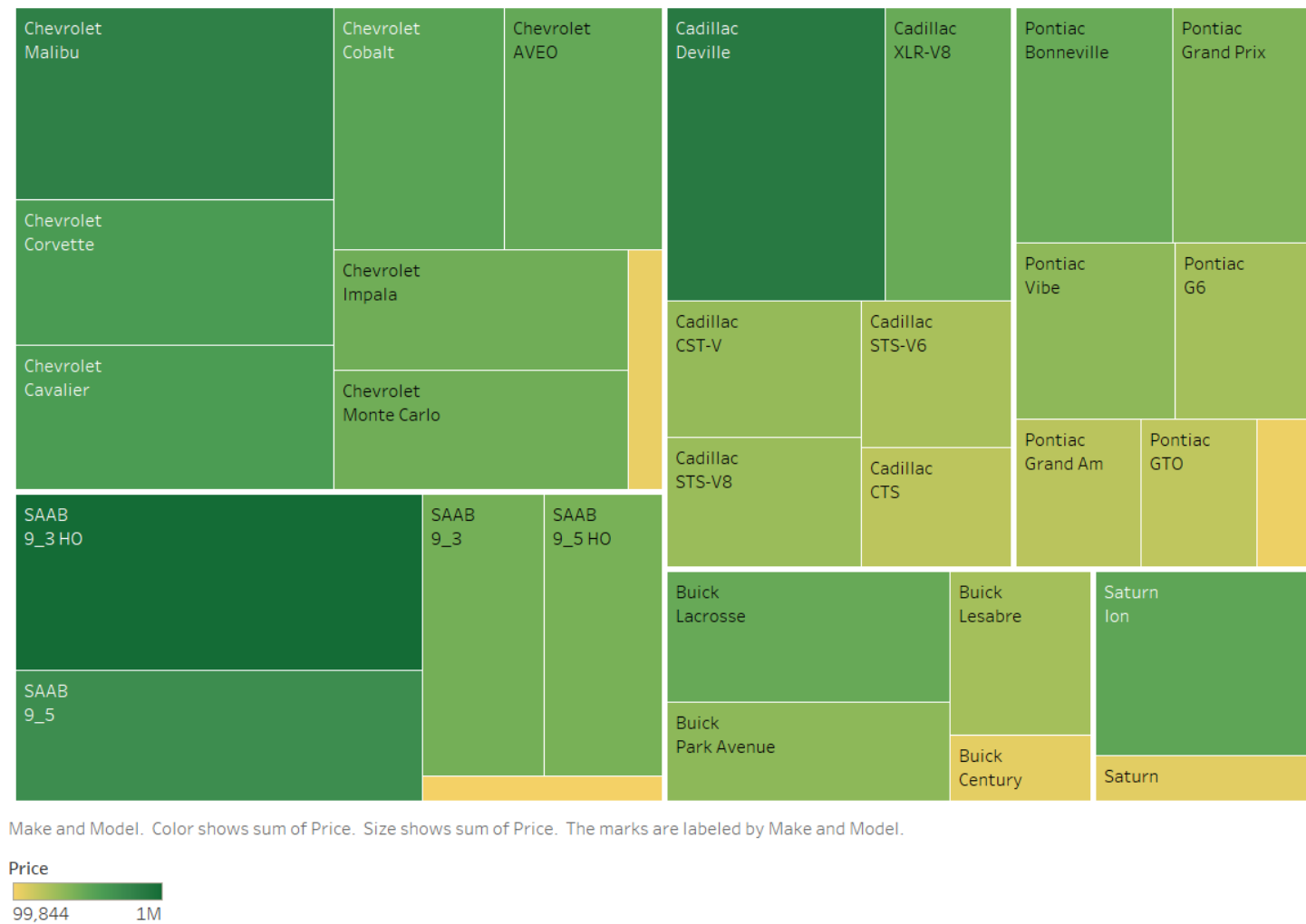
c) Scatterplot that graphs the Volume on the x-axis and the daily price range on the y-axis



Problem 2 (20 points): Tableau was used for this problem. We'll look at the data about GM cars (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.

- a) A treemap based on Price with a main subdivision for the Make of the car and a minor subdivision based on the Model.

Tree Map on Price with Make and Model



b) A packed bubble chart of the same type.

Bubble Chart on Price with Make and Model



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

c) Discussing the **differences** between two plots. Describing for each something that displayed more clearly than with the other.

Both of these graphs give us the good indication of pricing of the cars with respect to their make and models. In Tree map, the coloring is based on the price values, however the bubble chart is using color based on each make of the car which is helpful in differentiation of the makers of the cars. Tree map grouped the makers at each corner. For example, the Cadillac make is at the top right corner of the graph and SAAB make is at the lower left corner of the graph which makes the grouping better and allows to also comparing price based on make of the cars as a whole. I think, the Tree map does a better job in differentiating the price because comparing the area of the rectangle boxes is easier then comparing the area of circles.

- d) Contingency plot (Tableau calls it a heat map under Show Me) showing with color the number of cars (Number of Records) of each Type sold by each Make. Explaining at least one observation about that data that this chart makes it easy to see.

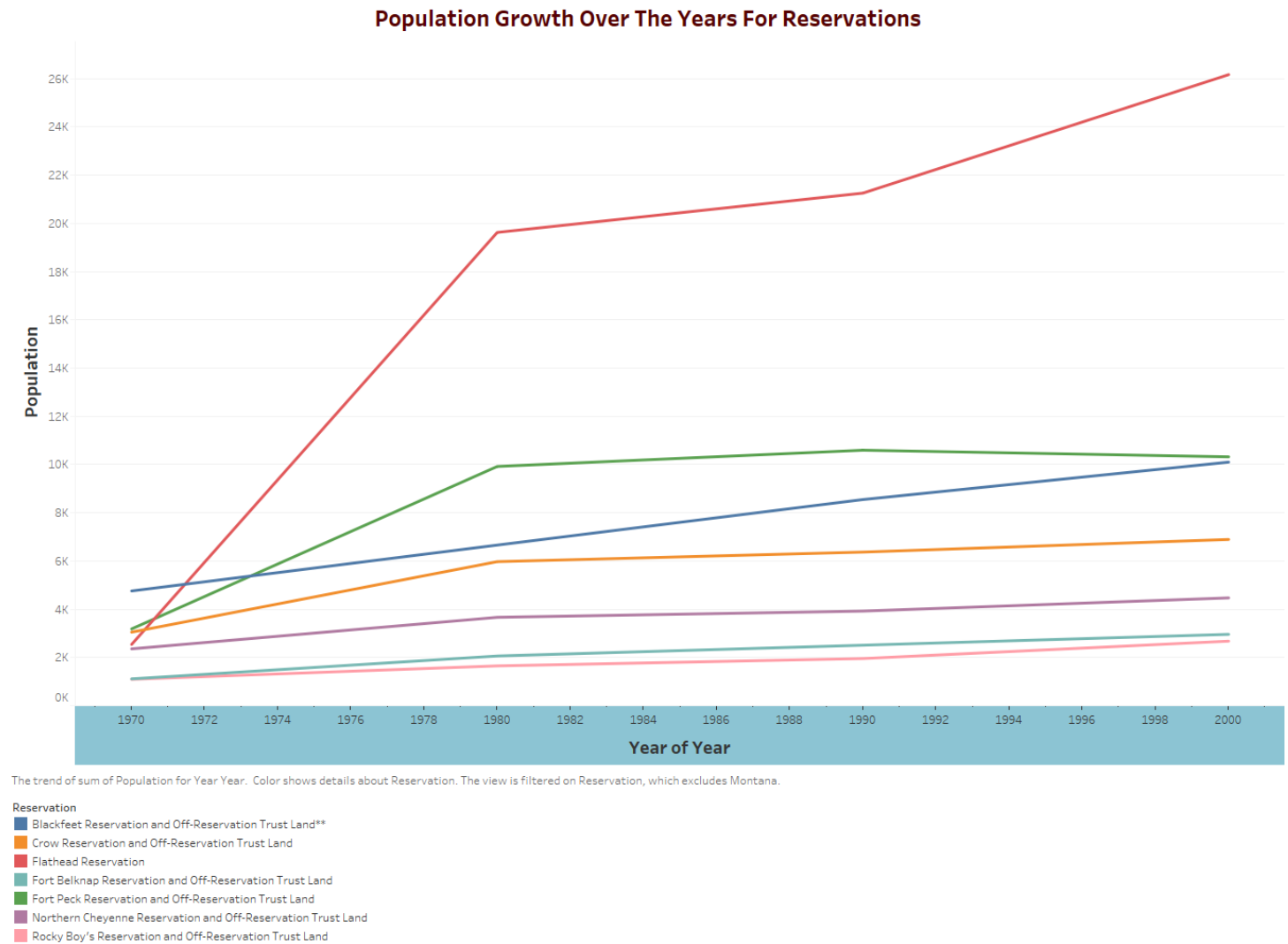
Heat Map / Contingency plot



Heat Map in Tableau helps us to compare the data by their color. This heat map helps us identify immediately the number of cars sold with respect to different make and type of the cars. The color and size of the boxes are effectively recognized by human perception helping to identify, which products have not been sold much and which products are doing best. We can clearly see that the Chevrolet Sedan was the most sold car.

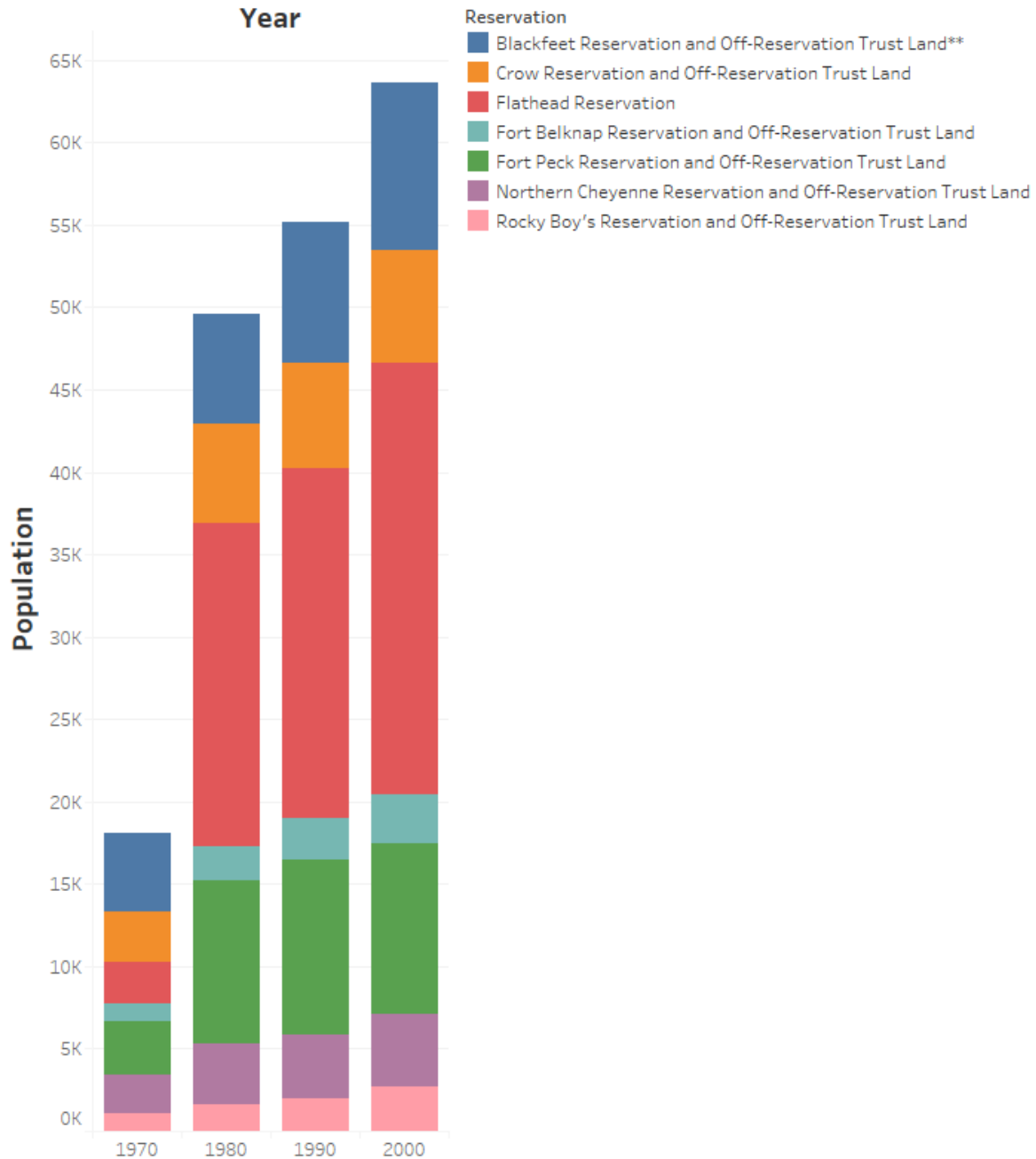
Problem 3 (20 points): This problem works with a dataset containing the population of Montana and of each of the 7 Native American reservations within it (reservation70-00.xlsx). There is a measurement for each decade between 1970 and 2000. Sheet1 has the original data. We will use Tableau for this question.

- a) Chart that graphs the population growth over the years for the individual reservations.



- b) Chart that graphs the total reservation population for each year, subdivided among the different reservations. The difference between this and (a) is that in (b) we are not looking only at each population individually but at the growth of the total population of all of them together, then subdivided by the reservations.

Reservation Population Over Year For Different Reservations



Sum of Population for each Year Year. Color shows details about Reservation. The view is filtered on Reservation, which excludes Montana.

Problem 4 (10 points): For this question, answer only with text. You may include an illustration if you would like, but you do not need to visualize data for this question.

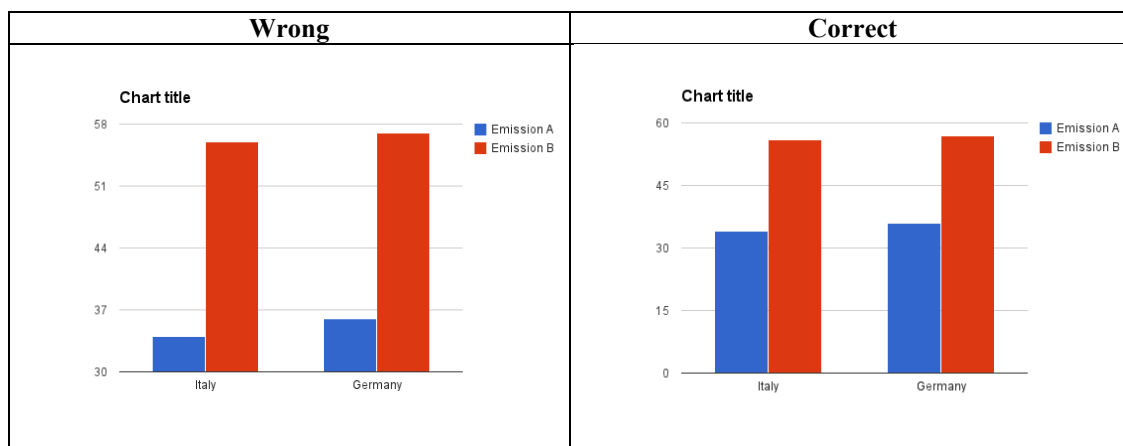
- a) Explain what we mean by ‘pre-attentive’ attributes. Are these as effectively recognized by human perception when they are used in combinations?

Pre-attentive or pre-intentional attributes are the key to better data visualizations. These are the attributes that are processed in our sensory memory without our conscious thought. These are visual cues that are visually apparent very quickly (~200ms) and do not require intentional processing to see. It gives us the opportunity to take advantage of the properties of human visual perception, which is the most developed among all our senses. There are usually four pre-attentive visual properties: Color, Form, Movement, and Spatial Positioning.

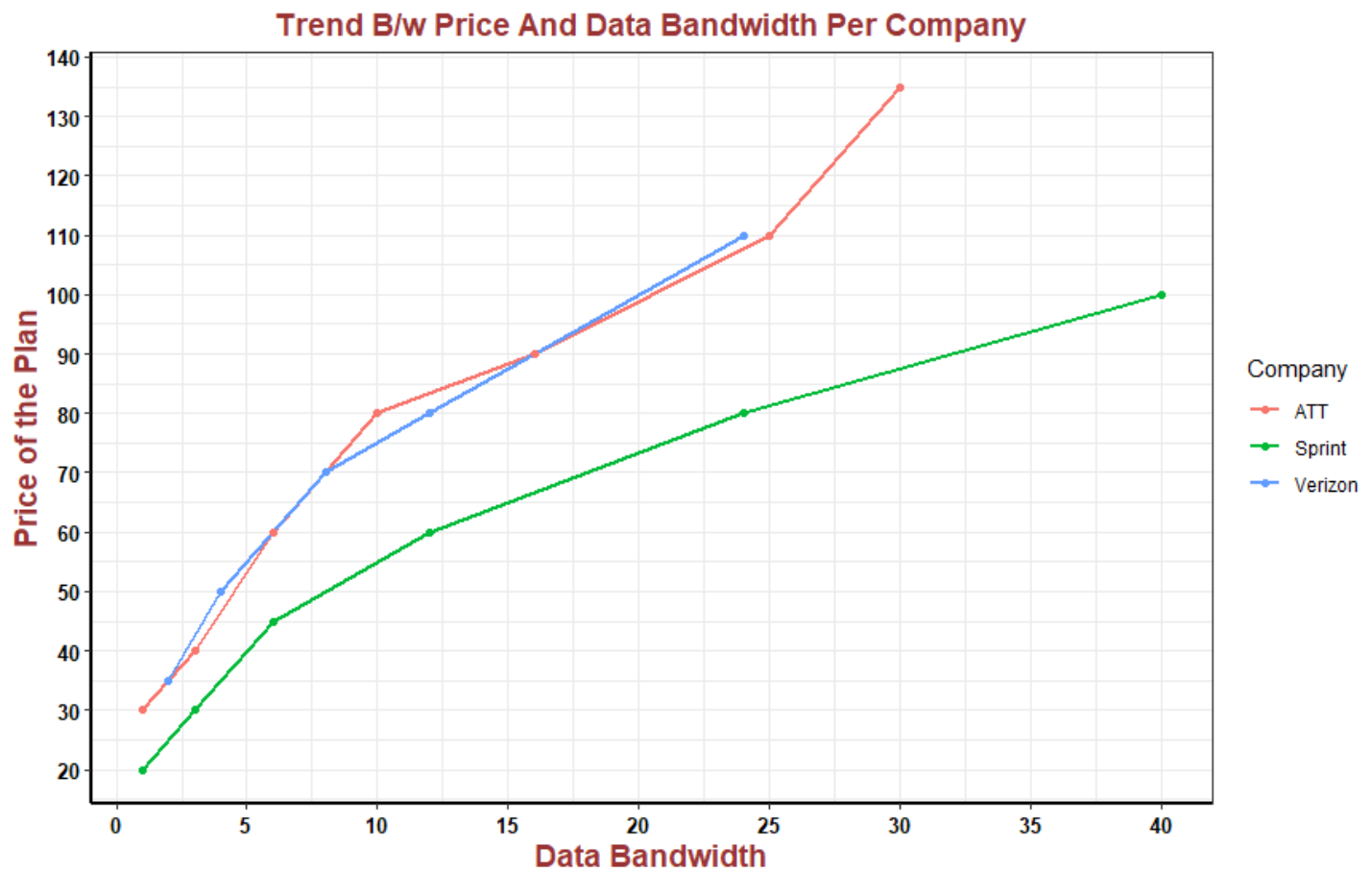
When we are developing the visualization, we need to emphasize on visual mapping to decide which properties to use for each attribute of the data set because this is the most significant thing. There are no universal rules or agreement, on which attributes are most effective in which combination for any particular data, however we can test them extensively to check if they deliver the appropriate results.

- b) Use Weber’s Law to explain why it is important to include 0 in the numerical axis of a bar chart.

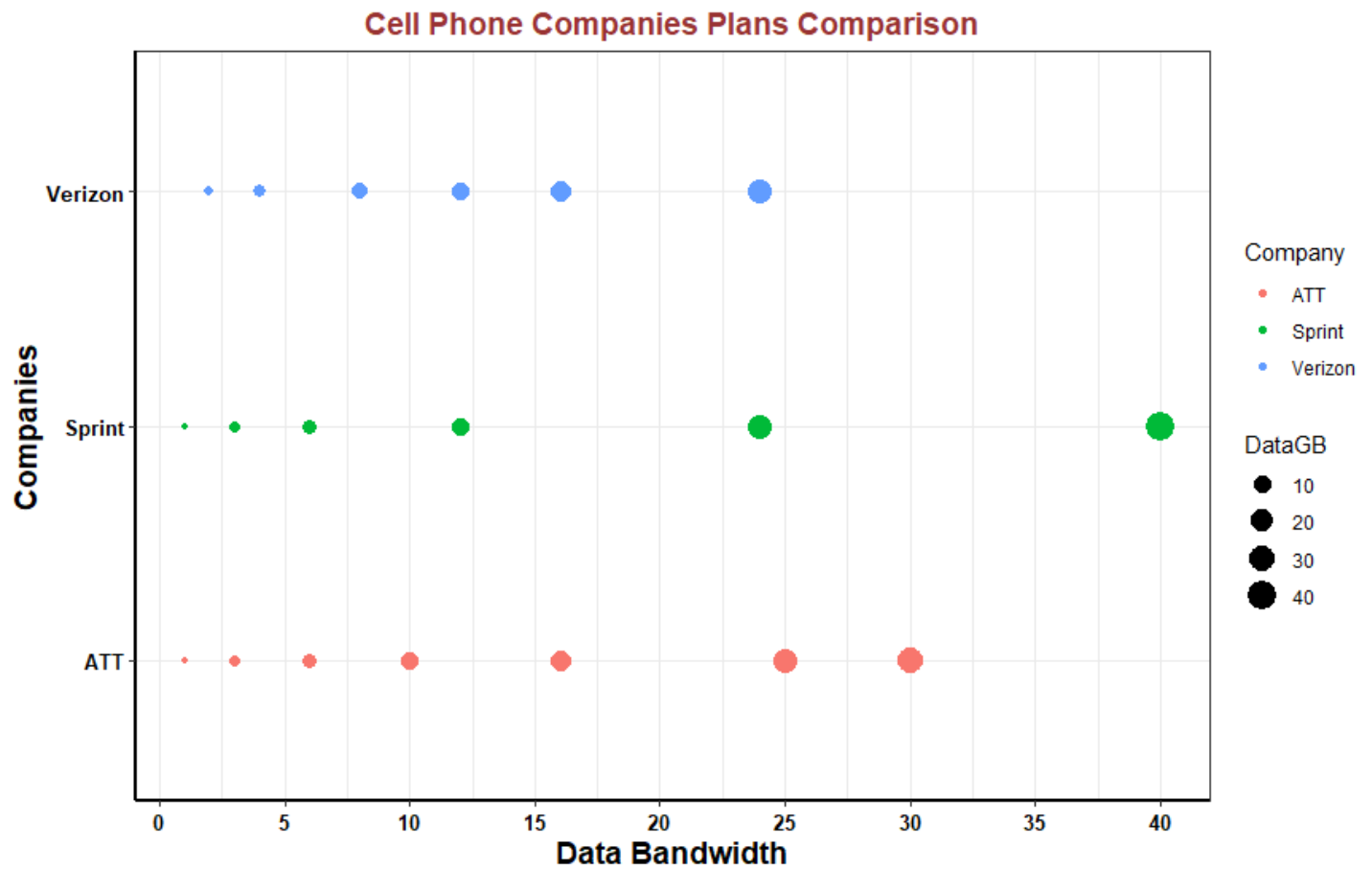
For bar charts, the numerical axis (often the y-axis) must start at zero. This is necessary because our eyes are very sensitive to the area of bars, and we draw inaccurate conclusions when those bars are truncated. Please refer to the good and bad example in below charts. Human eye is sensitive in comparing lengths. Comparing $x_1 - x_0$ to $x_2 - x_0$ is different from comparing $x_1 - 0$ to $x_2 - 0$.



Problem 5 (25 points): This problem works with a different cell phone companies pricing plans. We will use R for this question and recreate graph in two different ways. For each one, we will explain what we are trying to help the user see.



This first graph examines the trend of the relationship between price and data bandwidth. Different colors show details about different companies. The relationship holds overall but Sprint Company's pricing is more aggressive compare to ATT & Sprint. The overall pricing for Sprint for all plans are cheaper then ATT & Sprint. This graph makes the answer to our question immediately clear and it is also well implemented, i.e. following the guidelines for a clean graph.



The purpose of this second graph is to compare the cell phone companies to see what kind of different plans they offer. We have plotted available Data Bandwidth VS each Company. Different colors show details about different companies e.g. ATT, Sprint, and Verizon. The size of the circles represents Data GB bandwidth. The plans offered by companies are more or less in same range however, Sprint offers the biggest data plan which is 40GB.