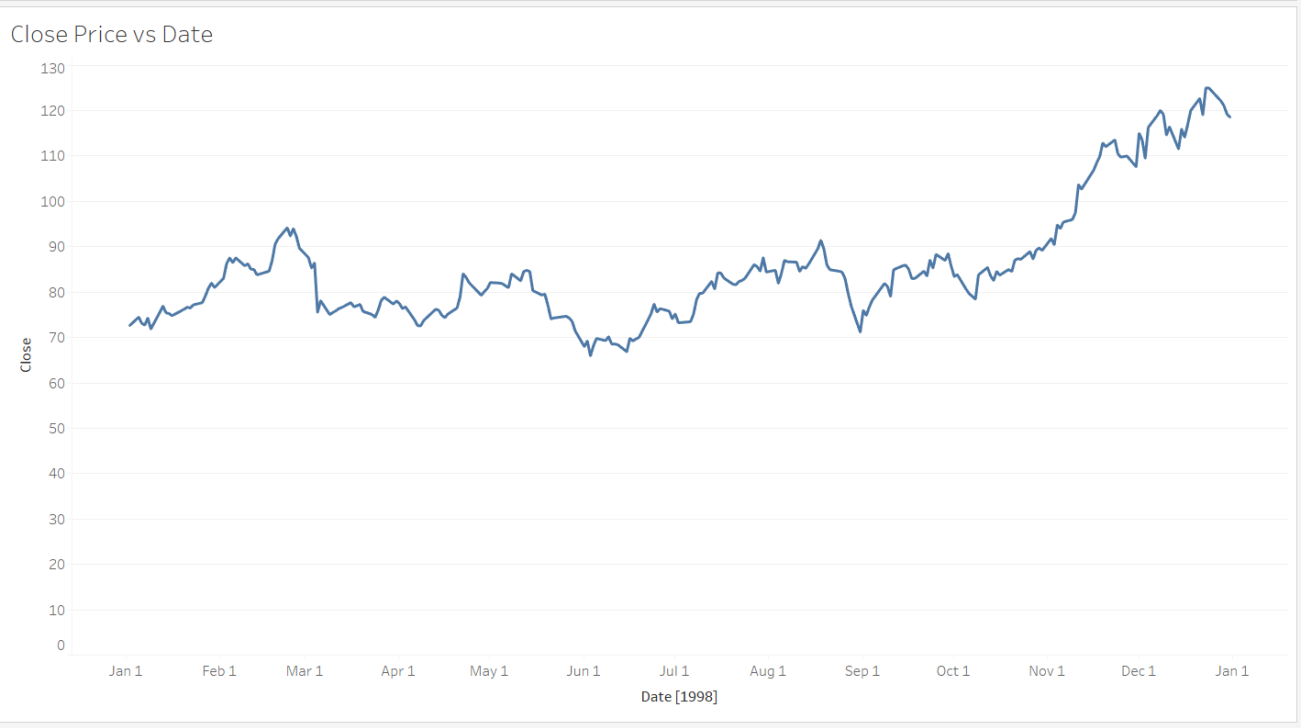
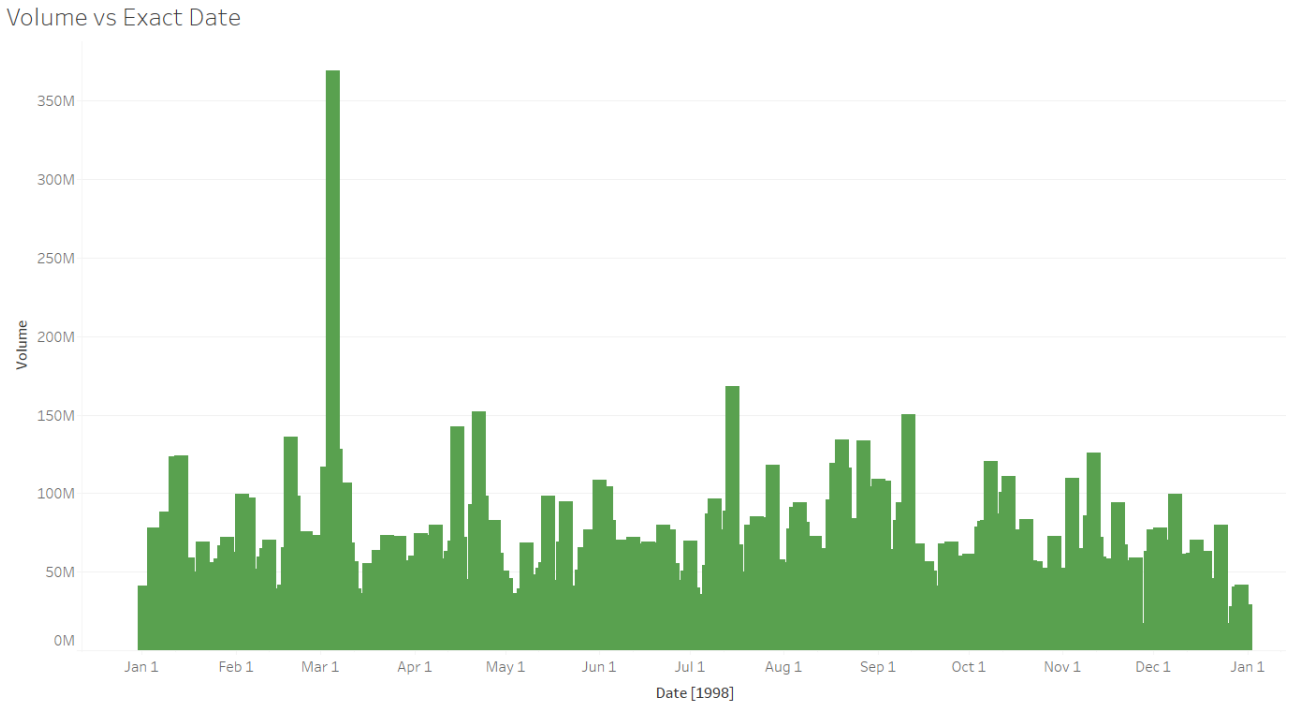
DATA VISUALIZATION ASSIGNMENT 1 Adithya Harsha

1) (20 pts) 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). Make the specified graphs in either R or Tableau:

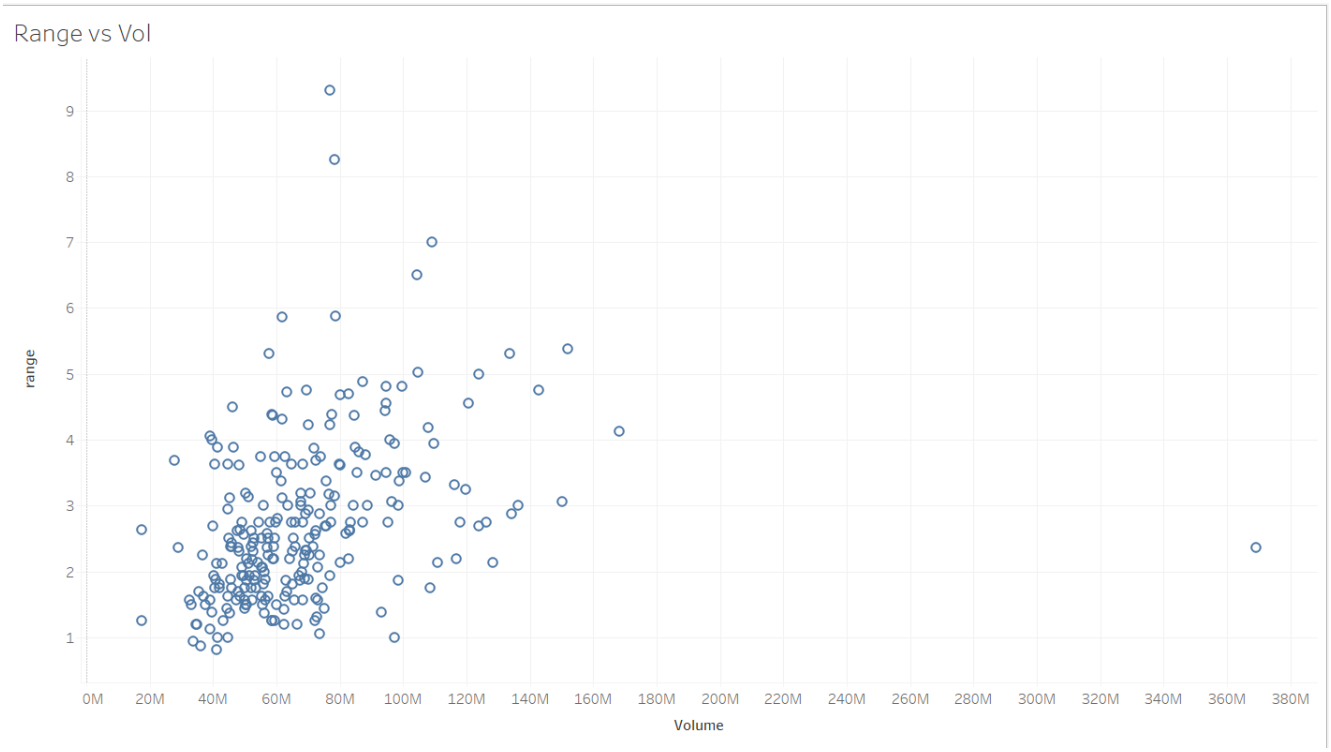
a. Graph the closing price vs. the date with an ordinary line graph. If you use Tableau, you need to right-click on the Date and choose Exact Date from the dropdown menu so that it uses the full date with "day"1



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

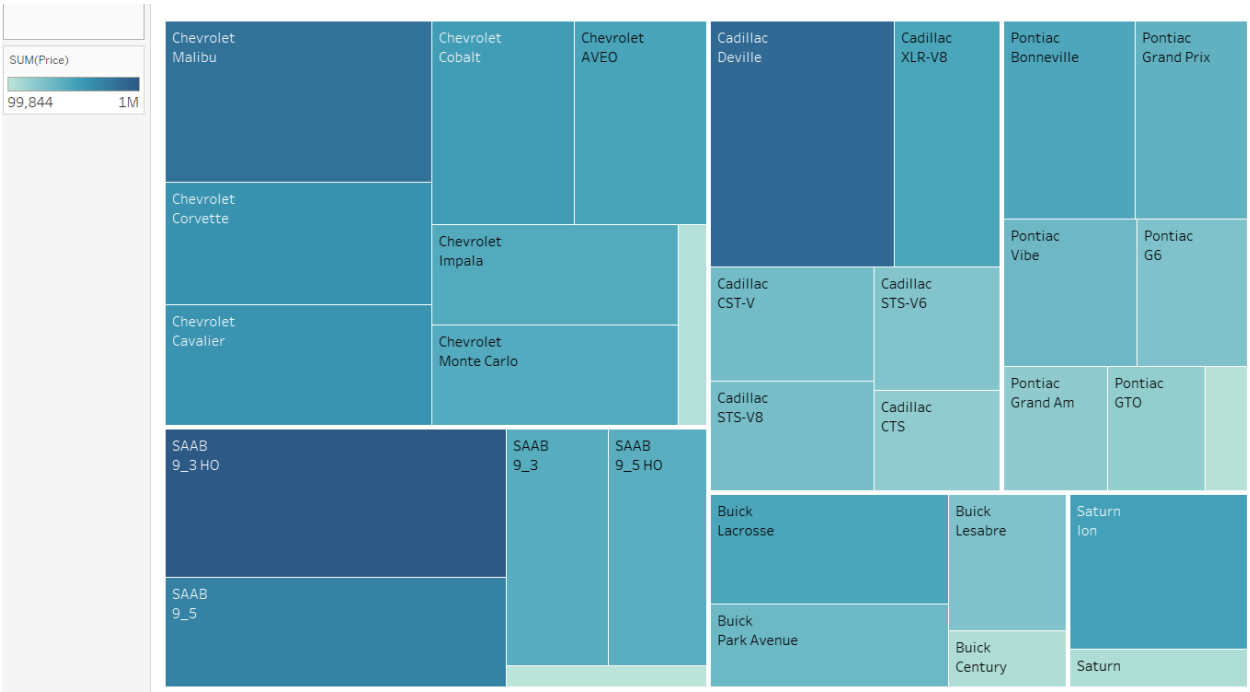


1c) Create a scatterplot that graphs the Volume on the x-axis and the daily price range on the y-axis. You will need to create an additional column that contains the "range" of the prices for the day as the difference between the fields High and Low. Range = High – Low



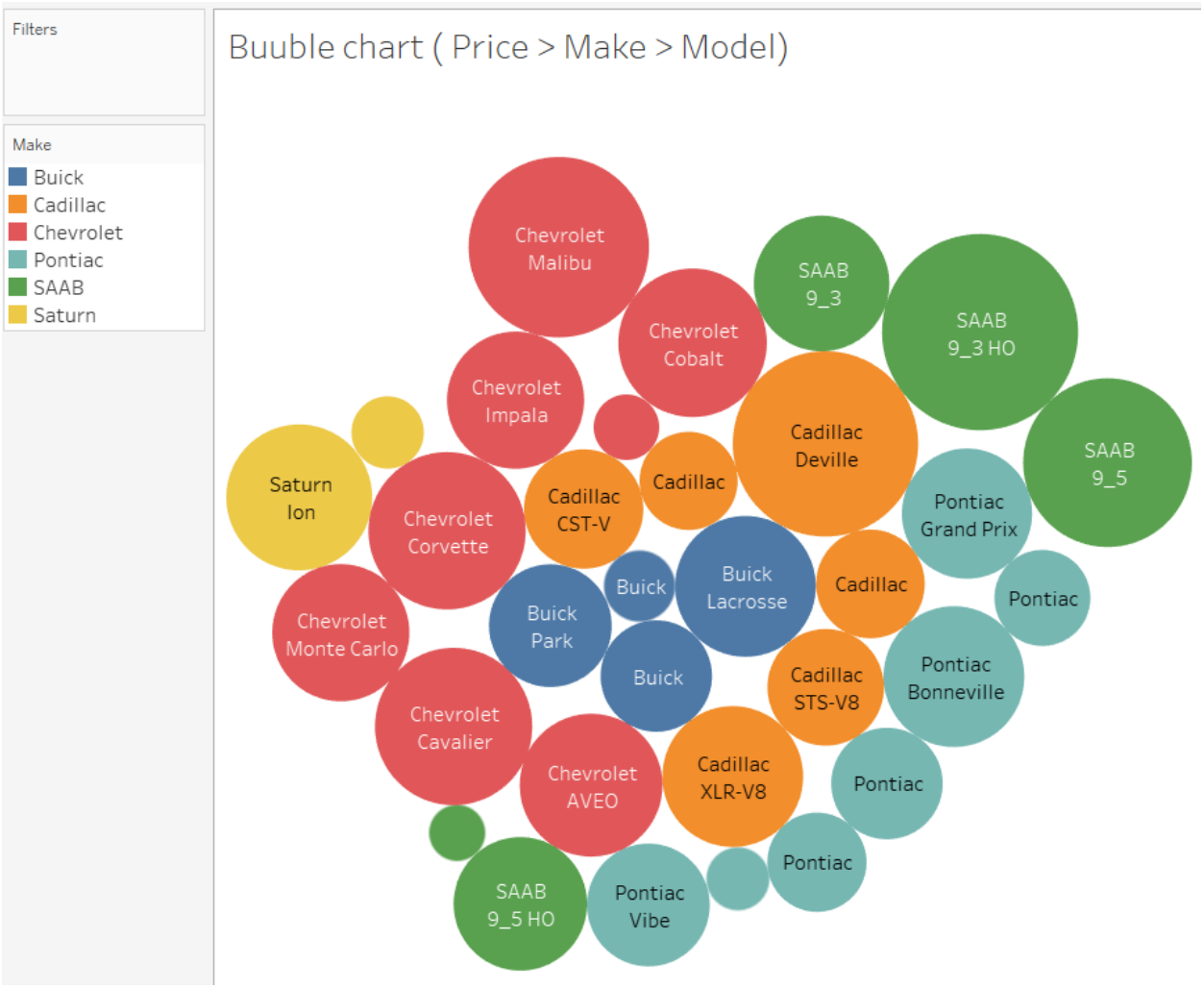
2) (20 pts) Use Tableau for this question. Open the GM cars dataset included with this assignment (gmcar_price.txt). Each row represents a different car that was sold and includes information about features like the mileage and the price of sale. Hint: use the “Show Me” menu.

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.



b)

b. A packed bubble chart of the same type.

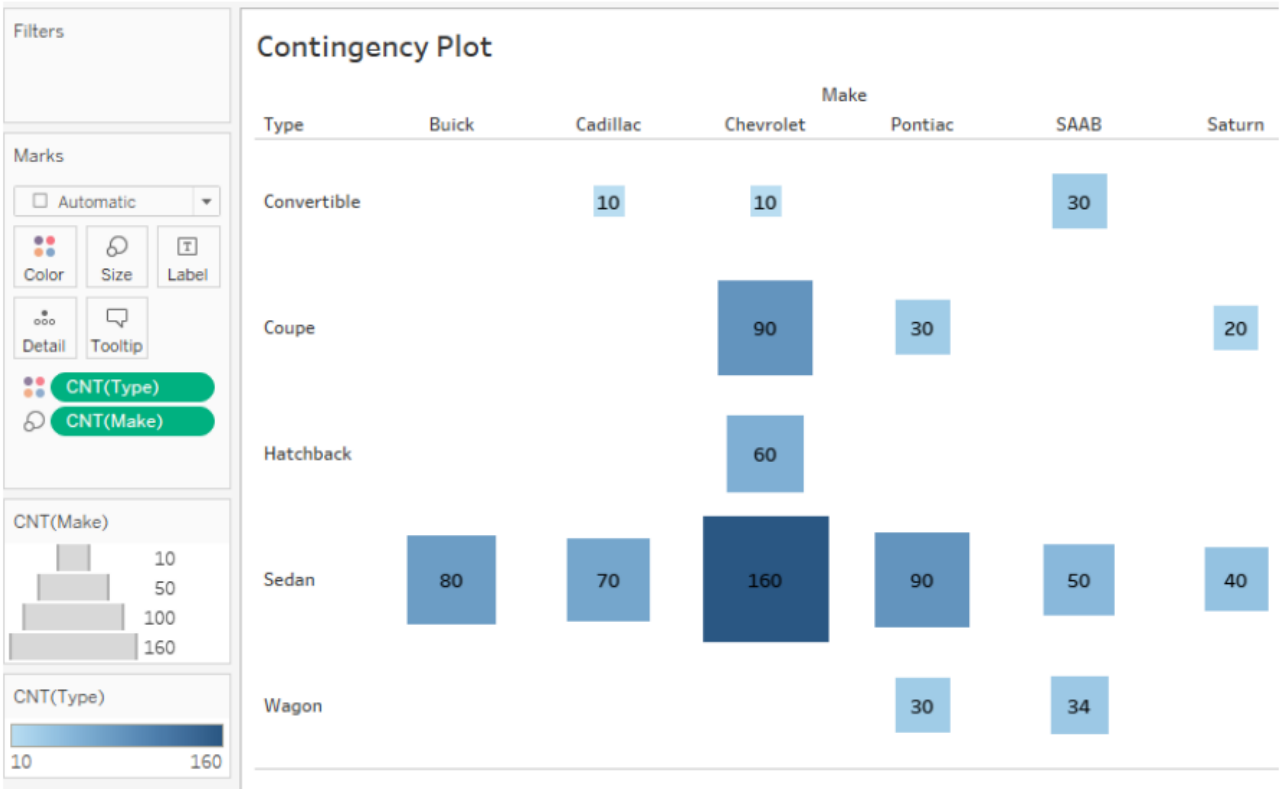


2 **c. Write a short paragraph discussing the differences between the two plots. Describe for each something that displayed more clearly than with the other.**

The treemap visualization has organized the data into four distinct groups based on the Make of the vehicles, and the saturation of each group's color corresponds to the price range. This allows us to easily identify the standout models within each group, compare the differences between models within a group, and compare models across different groups.

On the other hand, the bubble chart provides an overview of the Makes, with each Make represented by a different color and the size of each bubble corresponding to the price. From the bubble chart, we can observe that Chevrolet has a larger number of models and a higher price range compared to the other Makes, while SAAB has a smaller number of models but a higher price range than Pontiac, which has more models. Therefore, the treemap provides a more detailed representation of the data while the bubble chart offers a broader overview.

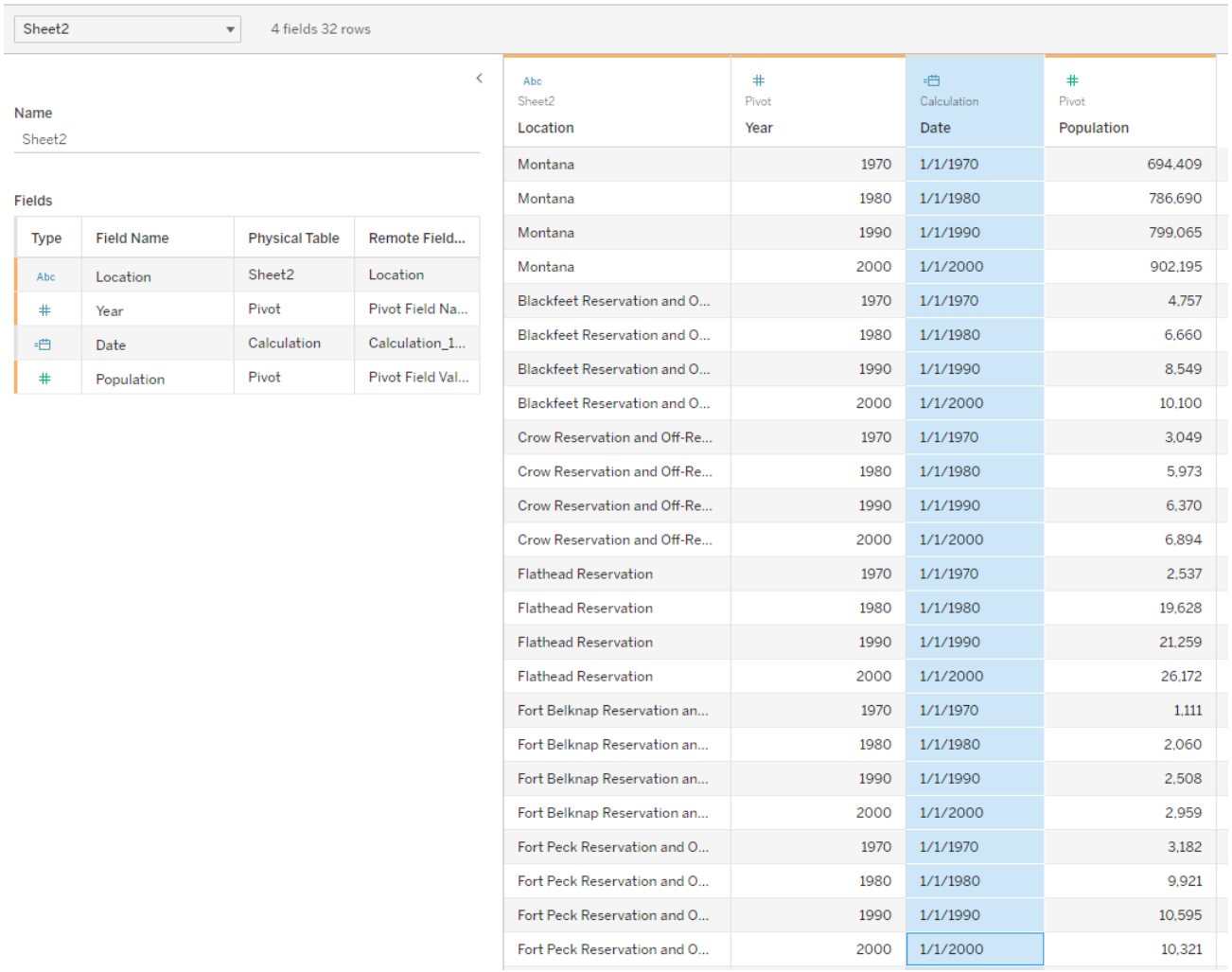
2 **d. Create a contingency plot (Tableau calls it a heat map under Show Me) showing with color the number of cars (Number of Records) of each Type sold by each Make. Explain at least one observation about that data that this chart makes it easy to see.**



The chart displays the number of models sold by each company, with a focus on sedans. For example, Chevrolet sold 160 sedans, while Pontiac sold 90 sedans, Buick sold 80, So by this visualization we can make out that Sedan car types are more likable to people so has better sold rate compared to any other type. By examining the chart, it is easy to identify which models are most popular and sold well and which are less popular by low sold rate for each company.

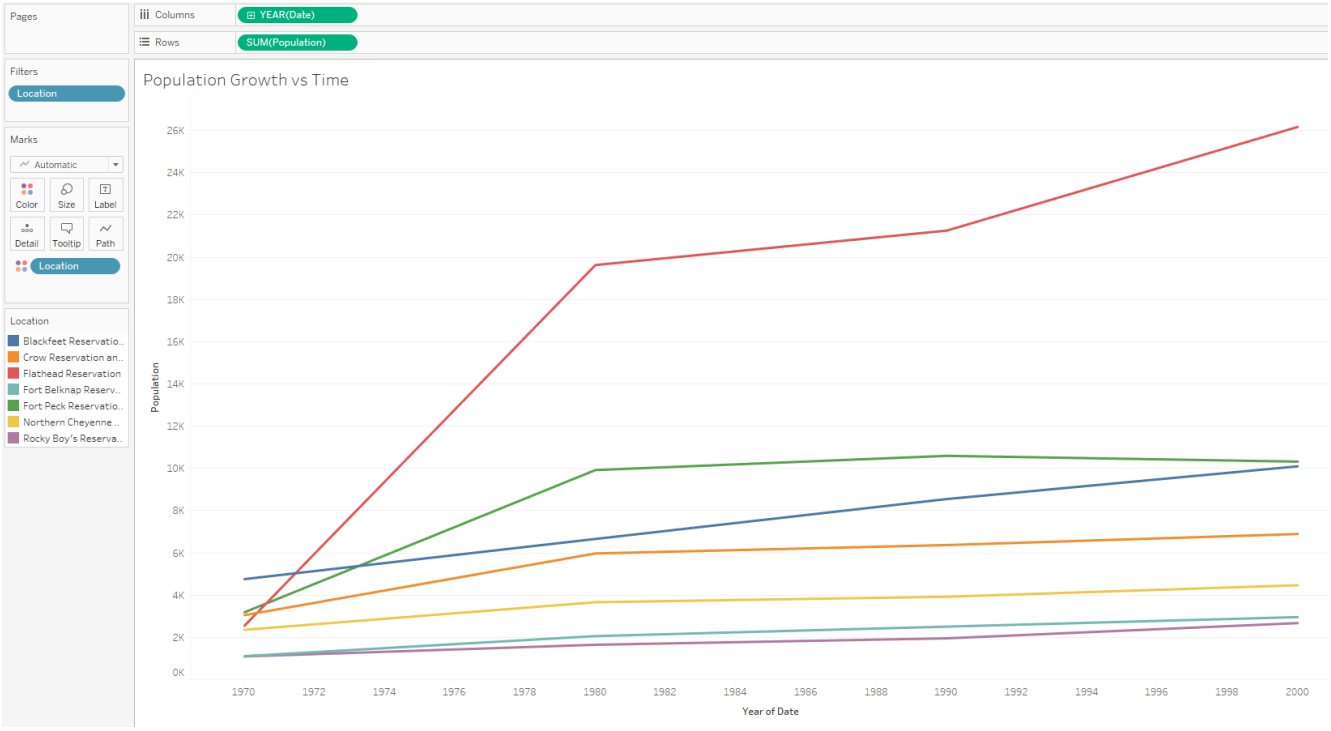
3) This problem works with a dataset containing the population of Montana and of each of the 7 Native American reservations within it (reservation70-00.xlsx). There is a measurement for each decade between 1970 and 2000. Sheet1 has the original data.

PIVOTING

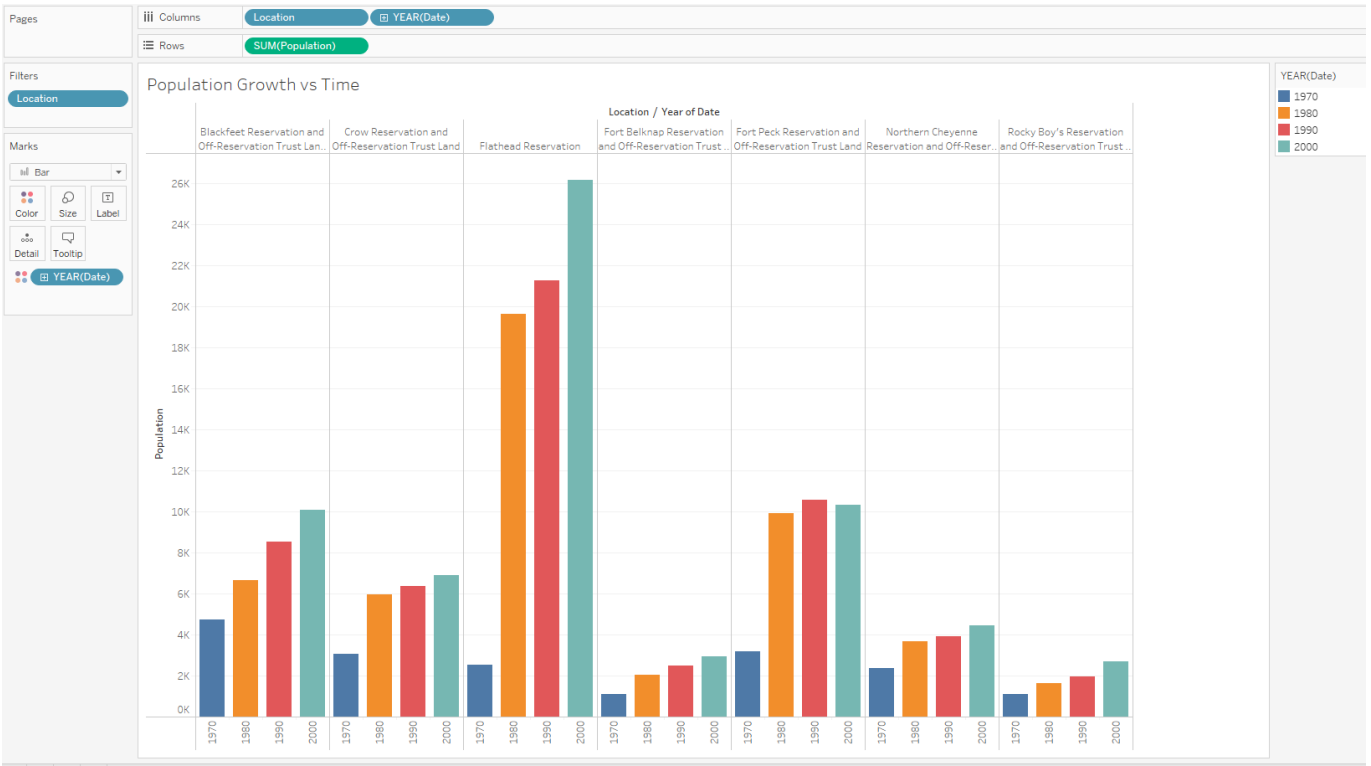


3) A)

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



3) B) **b. One 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.**



4**)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 attributes are visual properties of graphical elements that can be quickly and easily perceived by the human visual system without conscious attention or effort. Examples of pre-attentive attributes include color, size, shape, orientation, and motion.

When used individually, pre-attentive attributes can be effectively recognized by human perception, allowing viewers to quickly identify patterns and anomalies in data visualizations. However, when used in combinations, pre-attentive attributes can become more complex and may require additional attentional resources to accurately perceive and understand the information being presented.

For example, a bar chart that uses color to encode a categorical variable may be easy to read and interpret. However, if the chart uses color, size, and orientation to encode multiple variables, it may be more difficult for viewers to accurately interpret the information without additional attentional resources. Therefore, it is important to carefully consider the use of pre-attentive attributes in data visualizations to ensure that they are effectively communicating the intended information to viewers.

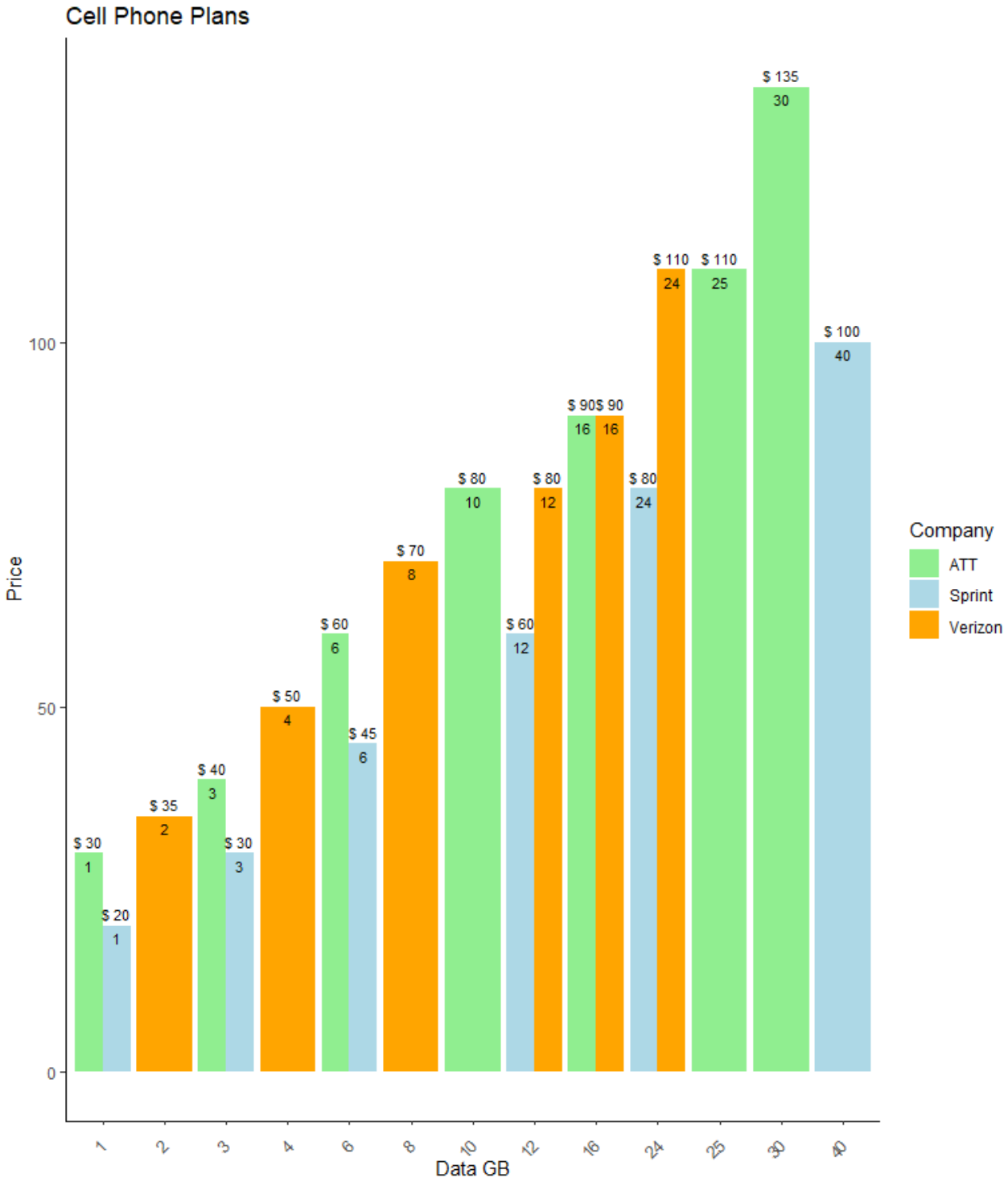
b) **b. Use Weber's Law to explain why it is important to include 0 in the numerical axis of a bar chart**

Weber's law states that the smallest noticeable difference (or just-noticeable difference) between two stimuli is proportional to the magnitude of the stimuli. In the context of a bar chart, this means that the ability of a viewer to distinguish between the heights of two bars depends on the magnitude of the heights.

When a bar chart includes a numerical axis that does not include 0, the viewer's ability to perceive differences between the heights of the bars may be distorted. For example, if a chart displays bars with heights of 10 and 20, the viewer may perceive the difference as smaller than if the bars had heights of 100 and 110, even though the actual difference is the same. This is because the relative difference between 10 and 20 (a 100% increase) is much greater than the relative difference between 100 and 110 (a 10% increase).

By including 0 on the numerical axis, the viewer can more accurately perceive the relative differences between the heights of the bars, and make more informed comparisons between them. This is especially important when the data being displayed represents values that are close to each other, as even small differences can have a significant impact on the interpretation of the data.

5)


Cell Phone Plans
Price
100
50
0
$ 30
1
$ 20
1
$ 35
2
$ 40
3
$ 30
3
$ 50
4
$ 60
6
$ 45
6
$ 70
8
$ 80
10
$ 60
12
$ 80
12
$ 90
16
$ 90
16
$ 80
24
$ 110
24
$ 110
25
$ 135
30
$ 100
40
1
2
3
4
6
8
10
12
16
24
25
30
40
Data GB
Company
ATT
Sprint
Verizon

The plot shows the relationship between the amount of data GB used and the monthly price for three different cell phone companies (AT&T, Sprint, and Verizon).

Each bar represents a different amount of data GB used, and the height of the bar indicates the monthly price. The bars for each company are shown side-by-side in different colors (light green for AT&T, light blue for Sprint, and orange for Verizon).

To enhance the plot, the code includes labels on top of each bar indicating the price (in dollars) and on top of each bar with the data GB value. The x-axis label indicates the amount of data GB, and the y-axis label indicates the monthly price, from plots we can know sprint offers plan for lesser price compared to other two but doesn't provide most Data Gb plans where as other two provide more Data plans but ATT is higher priced than Verizon plans.

```
cellPlans = data.frame(
  c("ATT", "Sprint", "Verizon", "ATT", "Sprint",
    "Verizon", "ATT", "Sprint", "Verizon", "ATT",
    "Verizon", "Sprint", "Verizon", "ATT",
    "Verizon", "Sprint", "ATT", "ATT", "Sprint"),
  c(1, 1, 2, 3, 3, 4, 6, 6, 8, 10, 12, 12, 16, 16,
    24, 24, 25, 30, 40),
  c(30, 20, 35, 40, 30, 50, 60, 45, 70, 80, 80, 60,
    90, 90, 110, 80, 110, 135, 100))
names(cellPlans) = c("Company", "DataGB", "Price")
library(ggplot2)


ggplot(cellPlans, aes(x=as.factor(DataGB), y=Price, fill=Company)) +
  geom_bar(position="dodge", stat="identity") +
  scale_fill_manual(values=c("lightgreen", "lightblue", "orange")) +
  labs(title="Cell Phone Plans", x="Data GB", y="Price") +
  theme_classic() +
  geom_text(aes(label=paste("$", Price)), position=position_dodge(width=0.9), vjust=-0.5,
size=3) +
  geom_text(aes(label=DataGB), position=position_dodge(width=0.9), vjust=1.5, size=3) +
  theme(axis.text.x = element_text(angle = 00, vjust=0.5, hjust=1))
```