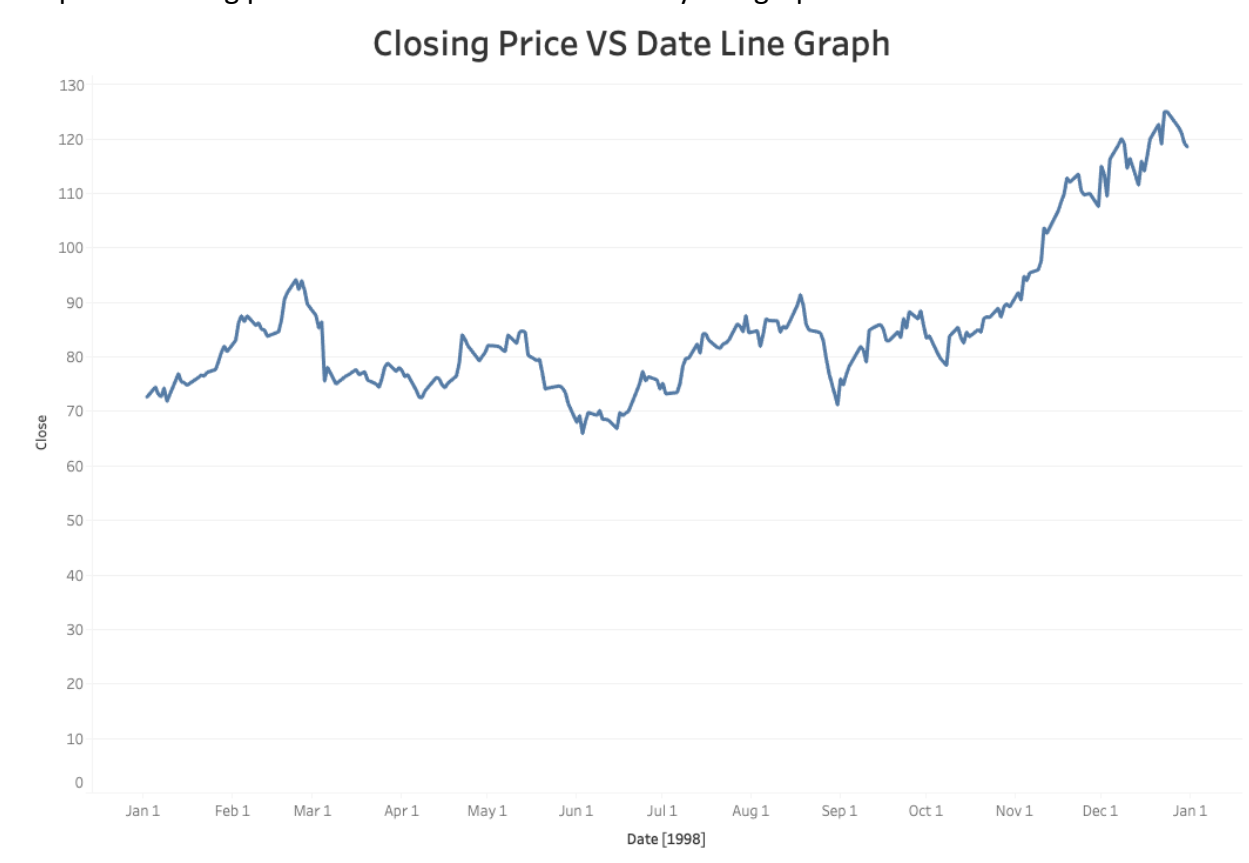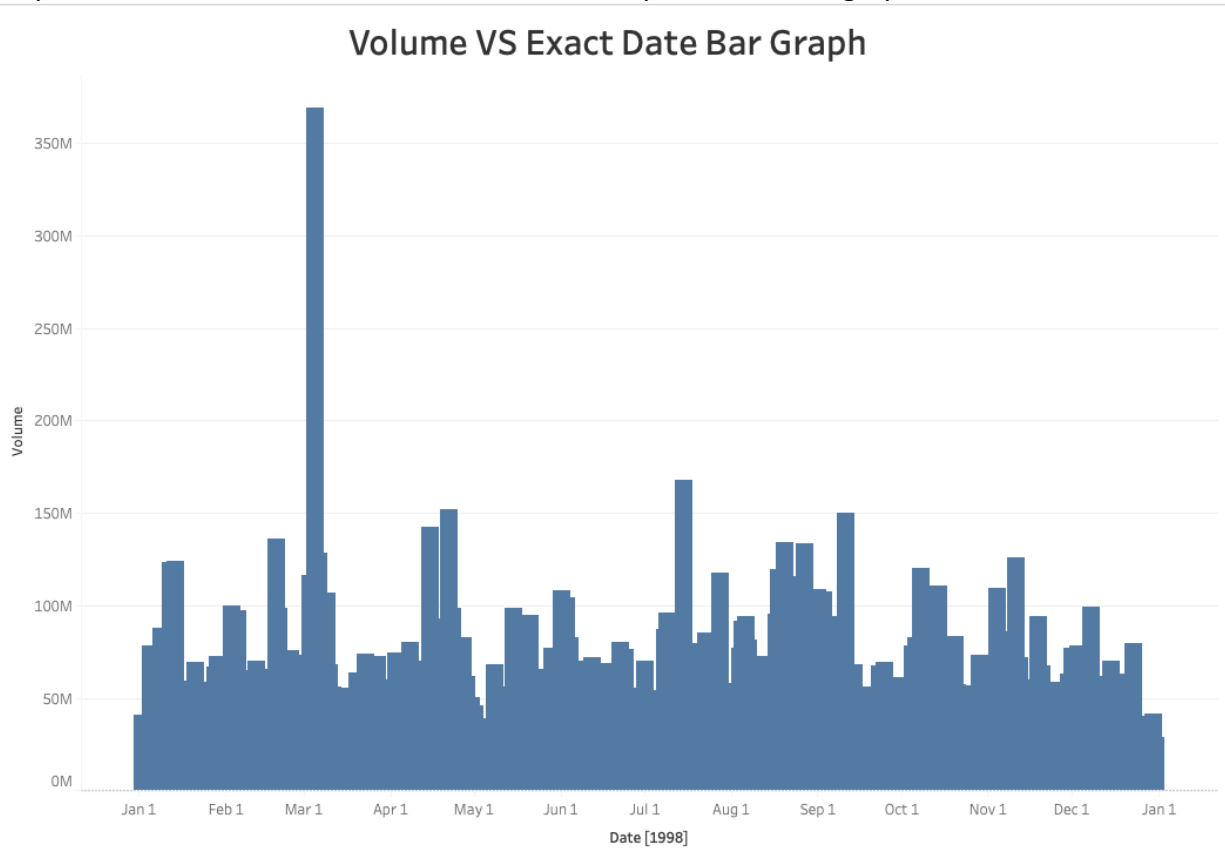**DSC 465**
HW1
Ximan Liu

1)
[1a]
Graph the closing price vs. the date with an ordinary line graph.

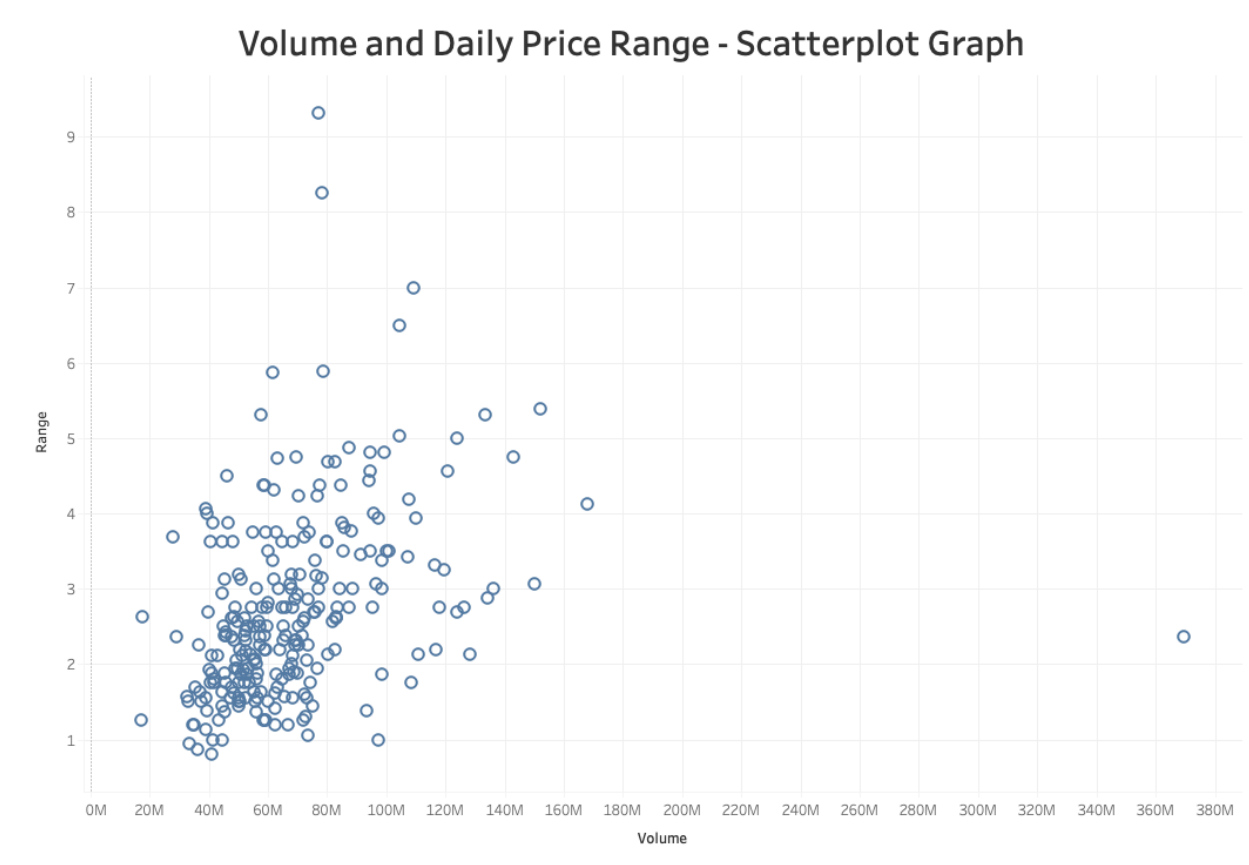**[1b]**
Graph the Volume vs. the exact Date as in the last part with a bar graph.

Volume VS Exact Date Bar Graph

[1c]
Create a scatterplot that graphs the Volume on the x-axis and the daily price range on the y-axis.

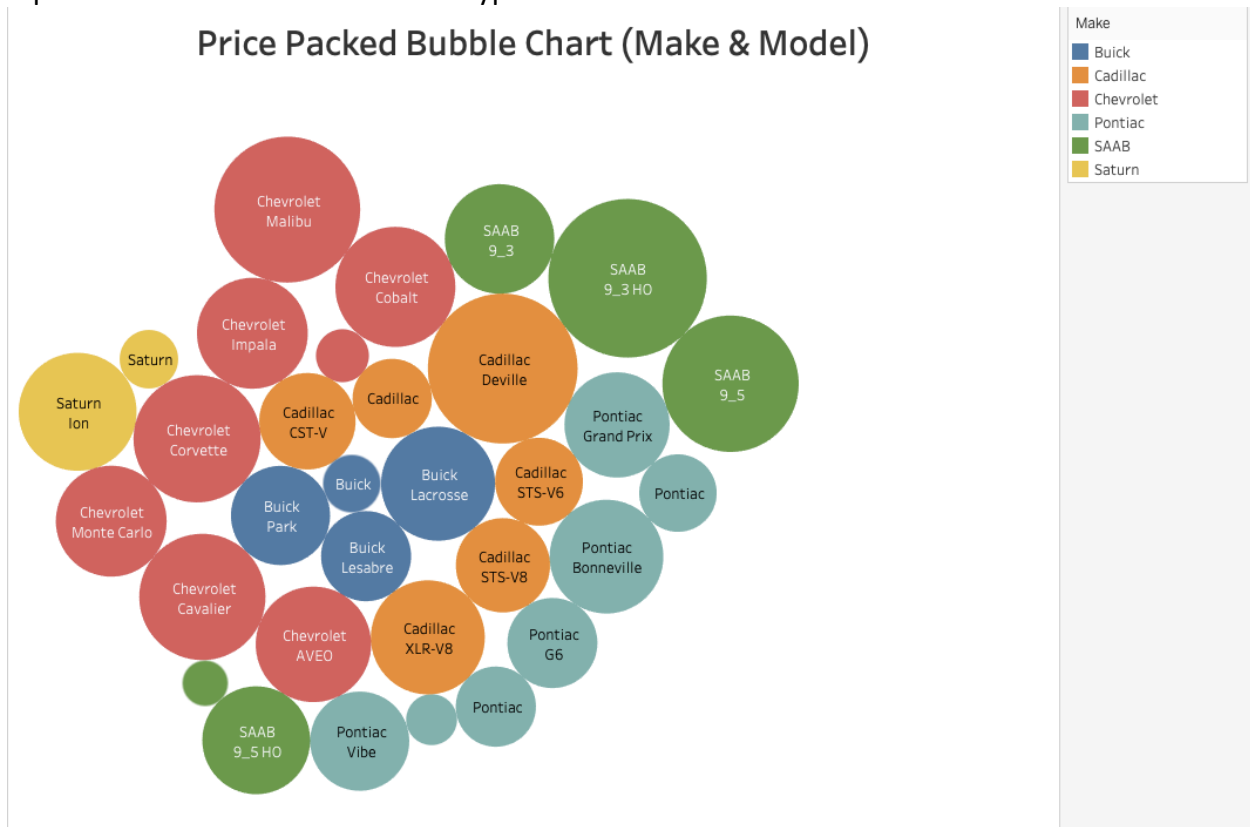Volume and Daily Price Range - Scatterplot Graph

2)
[2a]
A treemap based on Price with a main subdivision for the Make of the car and a minor
subdivision based on the Model.
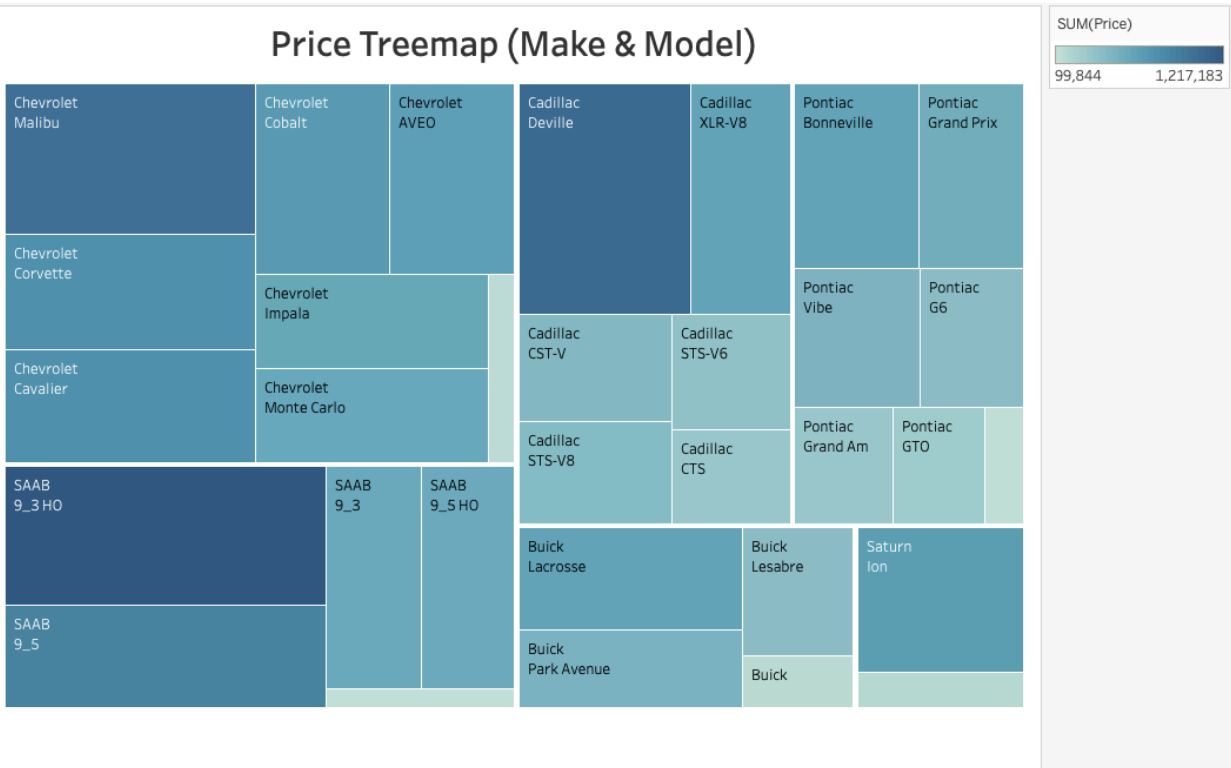
[2b]

A packed bubble chart of the same type.

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

Tree Maps depict hierarchical data as a series of nested rectangles, with a rectangle assigned to each branch of the tree, which is then tiled with smaller rec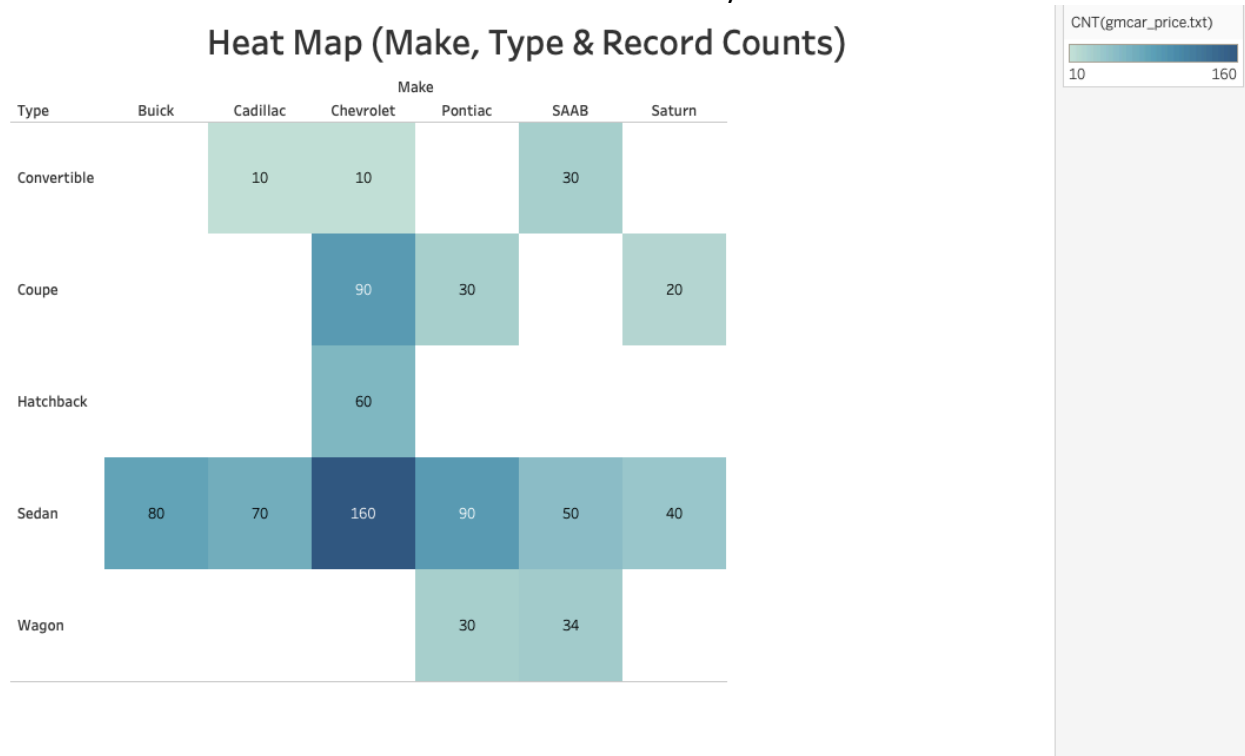tangles representing sub-branches. In a cluster of circles, a bubble map shows data in which dimensions identify the individual bubbles and measurements identify the size and color of the individual circles.
The Tree map for the dataset of GM cars helps narrow down the car's make and model at the highest price.
The packed bubble graph for the dataset of GM cars allows to restrict the car's make and model down to the highest price. To better distinguish between them, Make and Models are mixed and given an individual color.

[2d]

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.

Heat Map (Make, Type & Record Counts)

CNT(gmcar_price.txt)
10   160

| Type	Buick	Cadillac	Chevrolet	Pontiac	SAAB	Saturn
Convertible		10	10		30	
Coupe			90	30		20
Hatchback			60			
Sedan	80	70	160	90	50	40
Wagon				30	34	

From this heat map, we could see that Chevrolet's type Sedan is the most popular one by record counts. This is quite easily to be defined because it has the darkest blue shade in heat map, which represents the advantage of using heat map to show data.

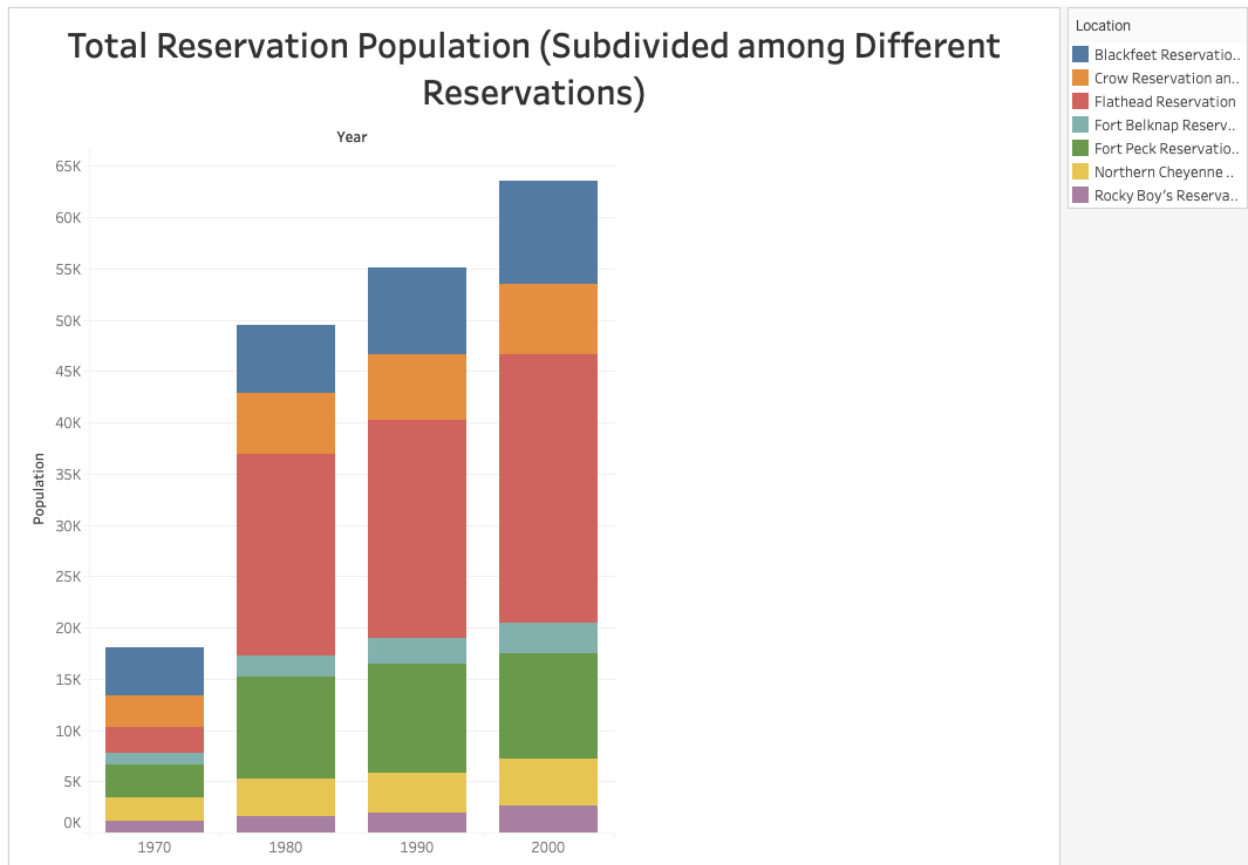**3)**
[3a]
One chart that graphs the population growth over the years for the individual reservations.
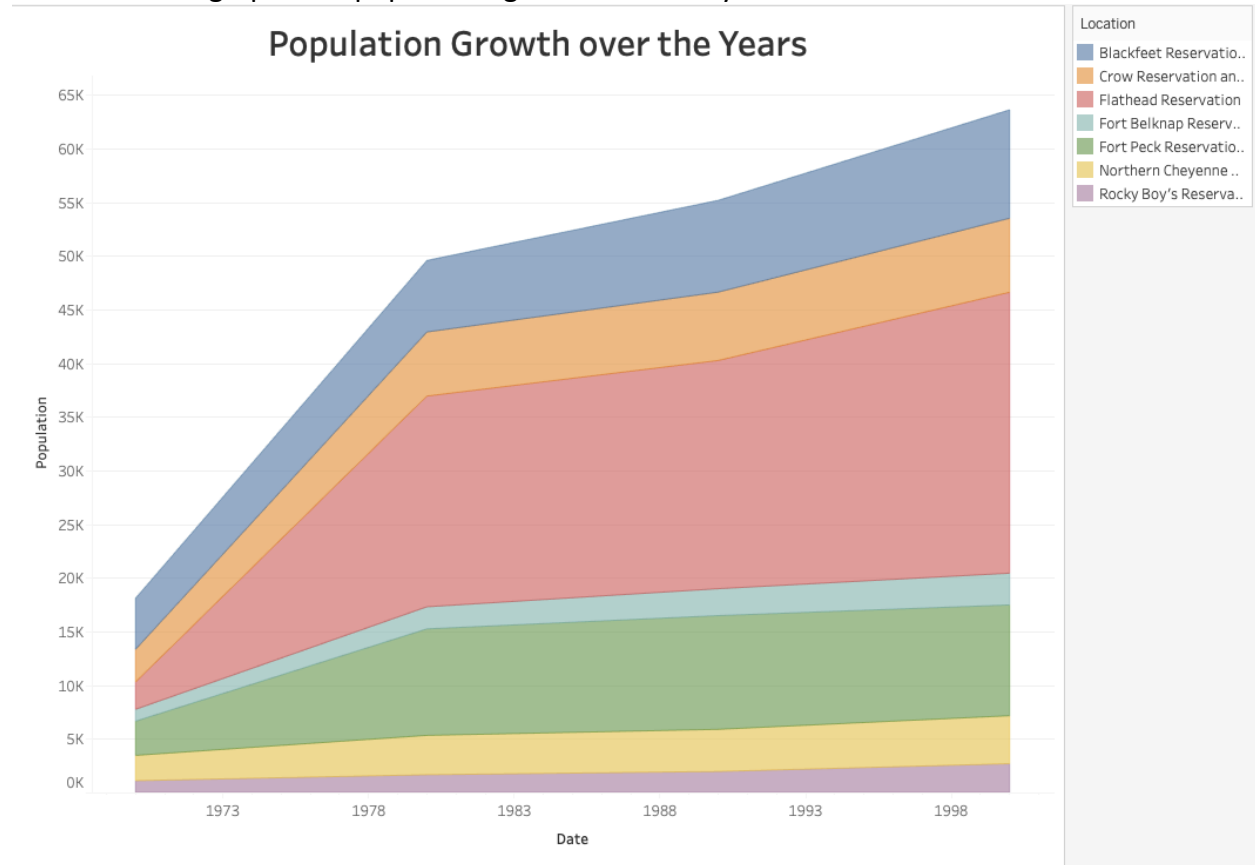
[3b]

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

4)
[4a]
Explain what we mean by 'pre-attentive' attributes. Are these as effectively recognized by human perception when they are used in combinations?

My answer:
Pre-attentive attributes mean visual cues that are visually apparent very quickly (~200ms) and do not require intentional processing to see.
Pre-attentive attributes become less distinct as variety increases. Understanding pre-attentive attributes will help us create visuals that emphasize the most important information while keeping the rest of the information from competing for attention.

[4b]
Use Weber's Law to explain why it is important to include 0 in the numerical axis of a bar chart.

My answer:
Weber's Law is all relative. The just-noticeable difference between two stimuli is proportional to the magnitude of the stimuli. Essentially, human perception works by percentage increase. In that case, for a bar chart, it is important to include 0 in the numerical axis. We need a relatively standard starting line.

5)
[Method 1]
We can compare the cell phone companies to see that kind of plans they have. In that case, we could build up a histogram graph of company and its plans. Red shade pillars represent ATT company, green shade pillars represent Sprint Company, and blue shade pillars represent Verizon company.

From the chart, we can see different companies supply different range of services. Verizon covers below 25 DataGB, that means if consumer tends to choose plan with higher DataGB, they'd better not choose Verizon. On contrary, Sprint's DataGB coverage is more comprehensive, users can find all her/his choices, even their demand of DataGB is as high as 40DataGB, Sprint can also do good job. Moreover, ATT coverage area is also relatively balanced, it is complementary to Sprint in terms of DataGB settings.

```
ggplot(data=cellPlans, aes(x = DataGB, y = Price, fill = Company)) +
  geom_col(position = "dodge") + ggtitle("Histogram of Company and its Plans")
```
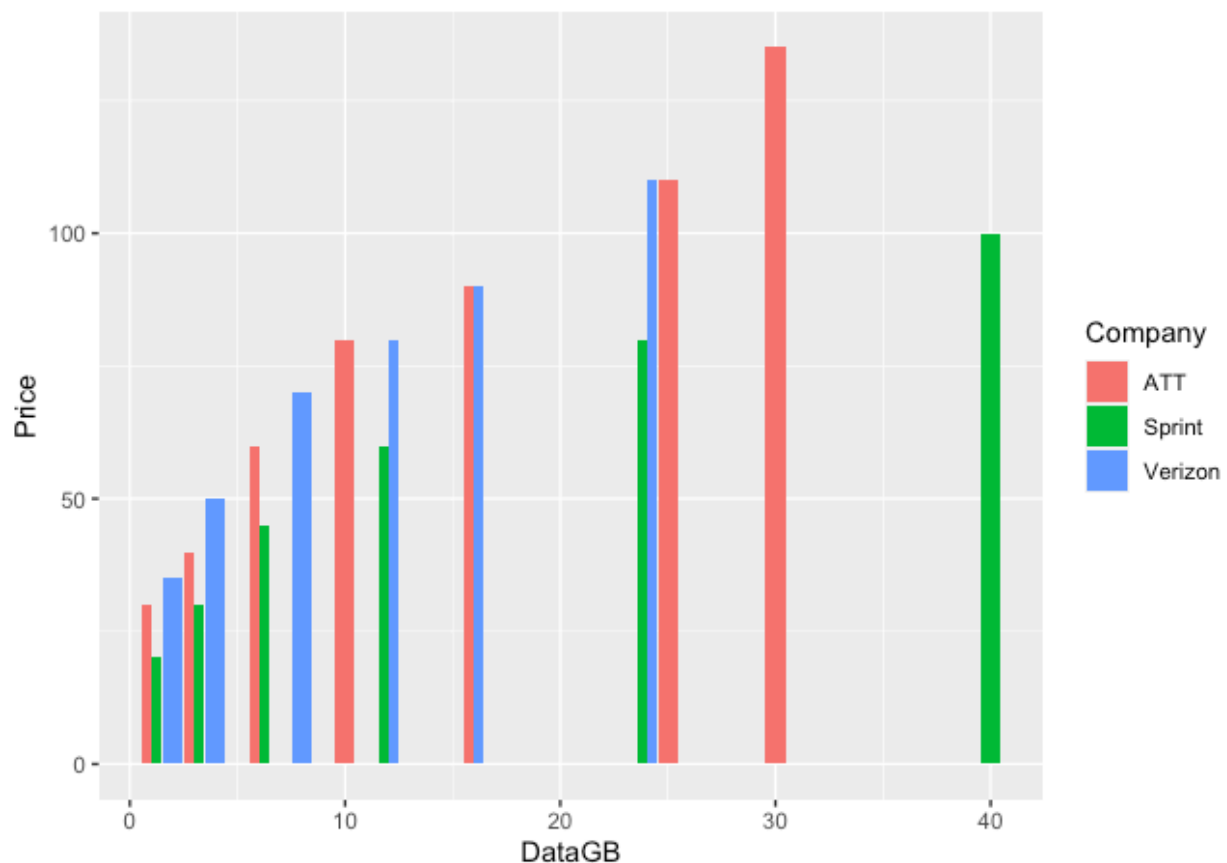
[Method 2]
We can also examine the trend of the relationship between price and data bandwidth. From line graph we can see that Sprint is extremely advantageous in price on even each different DataGB. We can definitely say that if consumers choose Sprint, she/he makes a wise choice. The pricing strategies of ATT and Verizon are more similar, below 25GB DataGB there is no obvious difference between these two companies. After 25GB DataGB, the price of ATT is extremely expensive compared with Sprint, that means it is not a good choice for consumers. More interestingly, Verizon does not choose to cover high-traffic frequency bands when DataGB is higher than 25GB, so its target to lower DataGB frequency bands.

```
ggplot(data=cellPlans, aes(x = DataGB, y = Price, colour = Company)) +
  geom_line() + ggtitle("Trend Relationship Line Graph of Company and its Plans")
```


Trend Relationship Line Graph of Company and its Plans