Homework 1

Note: Visualizations are at the end

1. These visualizations were created in Tableau, joined together in a Tableau with a border created around them for aid in grouping. Additionally, the question part (A, B, or C) was identified in Blue to make it more distinguishable from the title text. Finally, I added a note to the bottom defining "POI" and "Point Of Interest", in case it was unclear.
    a. Price vs Date was made in Tableau. I added date and close price (as continuous attributes). I added annotations of the start and end points in grey without border and high and low using a green (high) and red (low) curved border to distinguish. The goal was to make the high/low pop while allowing the start/end dates to be less noticeable as they seemed to be more ancillary and therefore warranted less attention.
    b. Volume by Date was created as a bar graph. I had to reduce the size of the bars to create spacing between them. In addition to the start/end annotations, I also pointed out two POIs: the high-volume day, which was significantly above the others, and 12/1/1998 because it was significant in graph C and it tells a better story in the context of the whole.
    c. Volume by Price Range was created as a scatter plot. I identified two outliers which I annotated in this graph (as well as graph B). I ensured range was in dollar form and set volume to display in millions.
2. These visualizations were created in Tableau. I experimented with different types of borders: I thought part A and B went well together while part D stood by itself. Part C was a discussion of part A & B so I added a border around it and used it to group the top two graphs, then I placed graph D and its discussion below. The discussion also had a border. In concert, the border around part C and discussion for D effectively put them into a group- in this case, a group of text, so within this chart there are 2 groupings: A group that consists of text, a group that consists of all graphs, and a group that consists of the graph subsets broken up by the text groups (A & B together, D by itself).
    a. Tree map was used using count of price which will identify each record with a price divided by make and model. This graph was relatively easy to make.
    b. The bubble chart was also relatively easy to make and was also made using the count of price. This one allowed for sorting by count, which did not seem to make sense, so I left it alone.
    c. This was text (it is in-line)
    d. The contingency plot by make/model was made using a count of records. I opted to leave the color schema as light to dark blue instead of blue to red as is often seen, to maintain the overall color scheme.
3. These visualizations were created within Tableau. The instructions were difficult to follow, so I did the best I could with what was presented.
    a. I could not determine whether the instructions asked for the population graphed over time or the y/y % difference in population, so I created a chart that included both (one on each axis). Using Blocks for the % change and lines for the total population allowed for contrast but may have confused the matters. Had the instructions asked for a single measure or measures that more readily worked into a single graph this would have been

cleaner (I would have likely created two separate graphs to show % change- the main issue is that the scale differences are large so it would require a logarithmic graph which does not work well with a linear graph which is needed for regular population growth).

b. This graph is a stacked area graph which readily shows population change over time of the entire reservation system broken up by individual reservations. It was created as a stacked area with analysis -> aggregation -> stacking turned on.

4. Answer with text:

a. Pre-attentive attributes are attributes that are immediately (<200ms) recognizable. These are not as effective in combination because it takes additional time for the brain to compute/understand the additional dimensions. Example from notes:

from Healey 2005
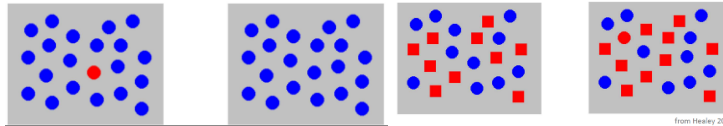
b. Weber's law states that "The just-noticeable difference between two stimuli is proportional to the magnitude of the stimuli" which means that people work by % difference. When working from an aligned scale it is more apparent where there are differences than when not, and without this baseline it's more difficult to mentally align objects for comparison.
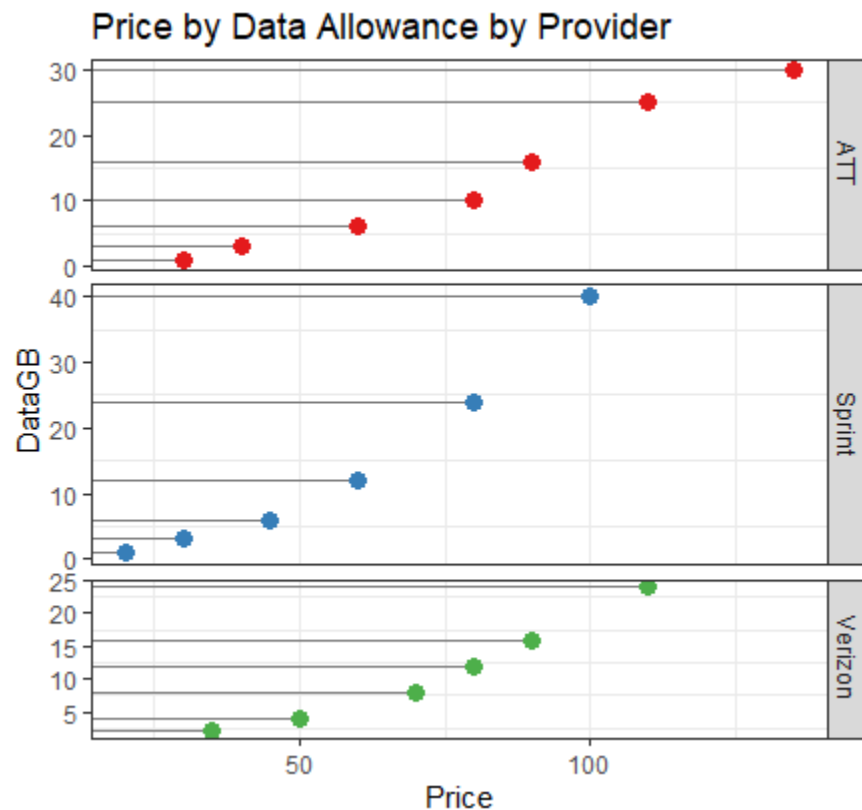
5. Phone Visualizations:

a. The question is, for a given GB need, which firm is the best? The visualization below allows the use to identify common price and data usage points by company using a common axis. It uses color to distinguish between firms (originally it used color and

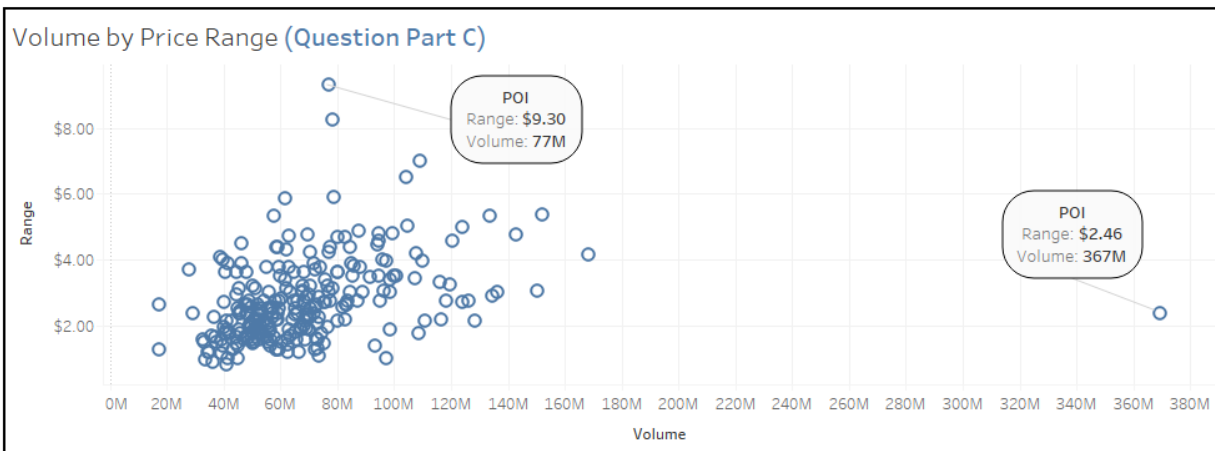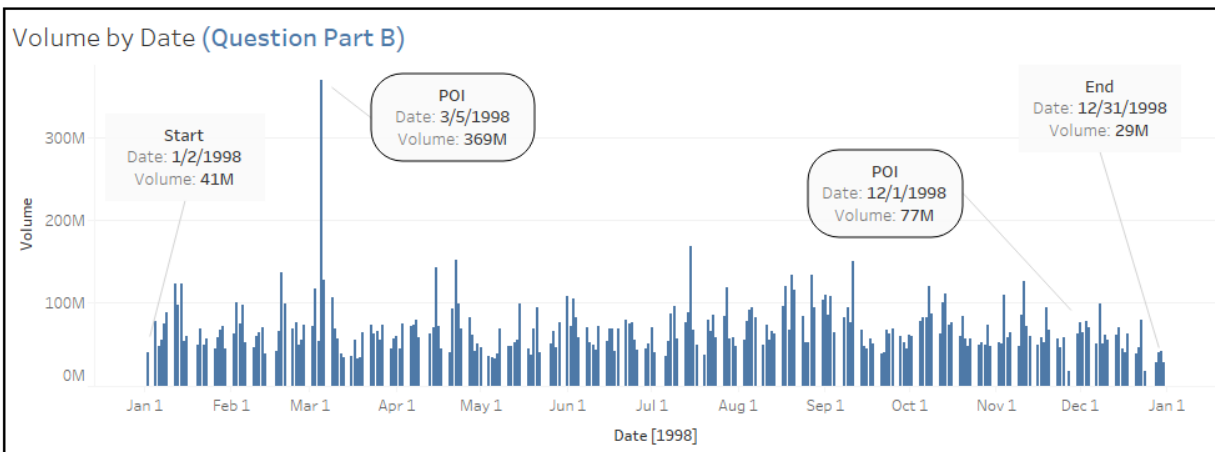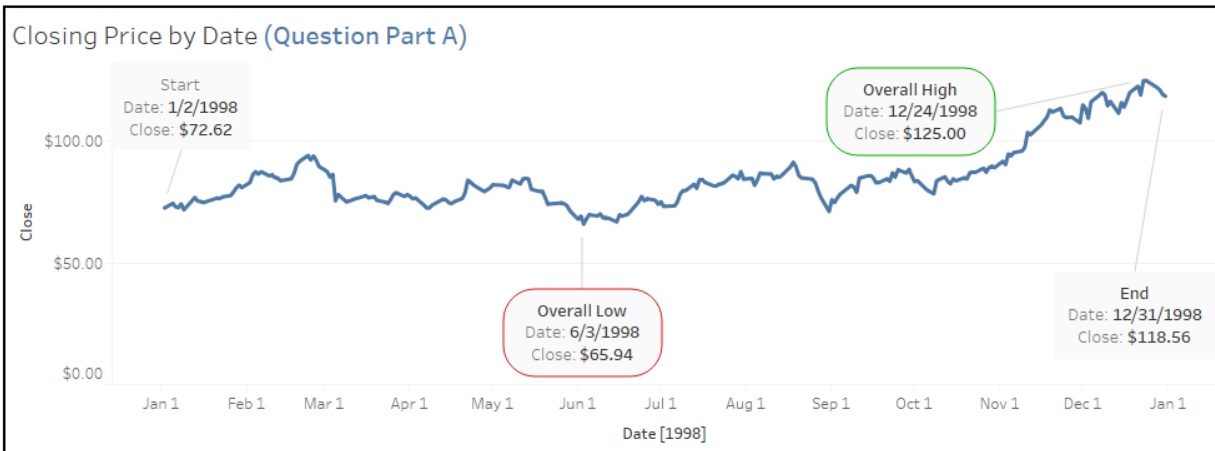

Price by Data Allowance by Provider
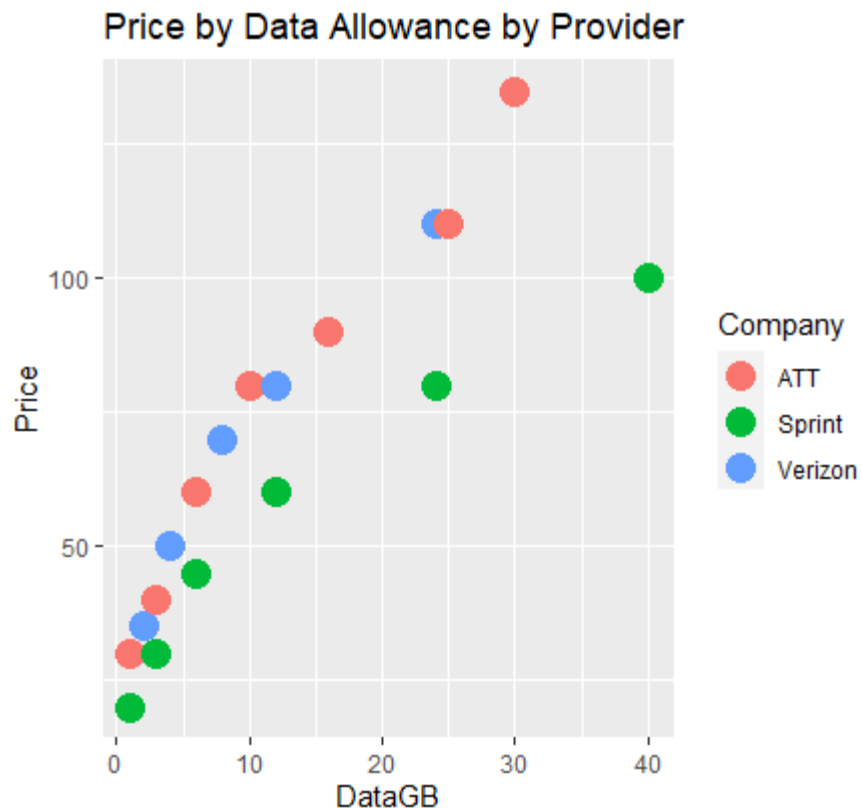
shape however review of pre-attention as well as channel effectiveness suggested using only color was more effective than the combination of the two. If I needed 40 GB it's easy to see that only one firm even has that plan, but it's also clear that if I need anything more than 20 GB I should *still* go with the 40 GB plan from sprint because it's cheaper.

b. The second graph answers the question "What can I get from each firm at a given price point?"? I chose a Cleveland Dot plot for this graph because it seemed to convey the image in the cleaned manner. The main drawback is that, while the price is using a common axis, the DataGB are not. It is important to note that while this is intended to be the focal point, it can be misleading to an individual not paying attention. and it therefore makes comparison slightly more difficult.

Price by Data Allowance by Provider

Question 1

Closing Price by Date (Question Part A)

Start
Date: 1/2/1998
Close: $72.62

Overall High
Date: 12/24/1998
Close: $125.00

Overall Low
Date: 6/3/1998
Close: $65.94

End
Date: 12/31/1998
Close: $118.56

Volume by Date (Question Part B)

Start
Date: 1/2/1998
Volume: 41M

POI
Date: 3/5/1998
Volume: 369M

POI
Date: 12/1/1998
Volume: 77M

End
Date: 12/31/1998
Volume: 29M

Volume by Price Range (Question Part C)

POI
Range: $9.30
Volume: 77M

POI
Range: $2.46
Volume: 367M

Point of Interest ("POI")

Question 2

Number of Feature Schemas by Make and Model (Question part A)

Count of Price
4 ▭▭▭▭▭▭ 60
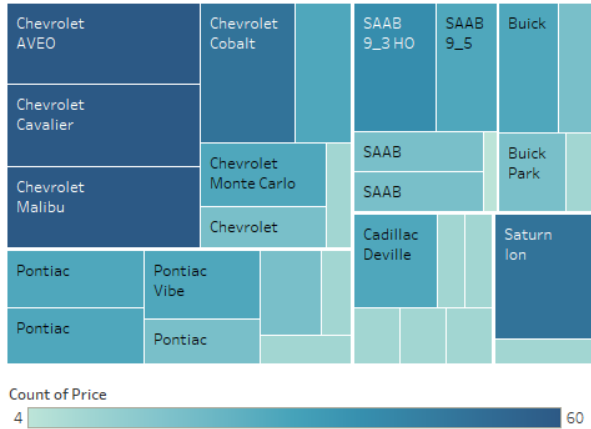
Distinct Feature Schema by Make and Model (Question Part B)

Make
- Chevrolet
- Pontiac
- SAAB
- Buick
- Cadillac
- Saturn
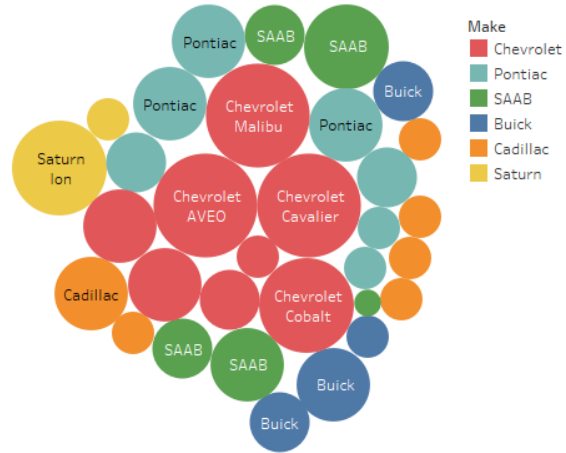
Part C - Discussion of Part A and B

Visualizations A and B are different in a number of ways. *A* separates car makes by grouping (white space between different car makes), while *B* separates them by color. Both show size using area, althought *A* also uses color (darker shades mean more cars). *A* does a great job of showing the relative sizes of the overall car *make* (aggregate of *models*) through the use of grouping (via white space), whereas that data is ambiguous in *B*. However, *B* makes it easier to compare sizes of individual models across makes.

Distinct Feature Schemas Sold by Make (Question Part D)
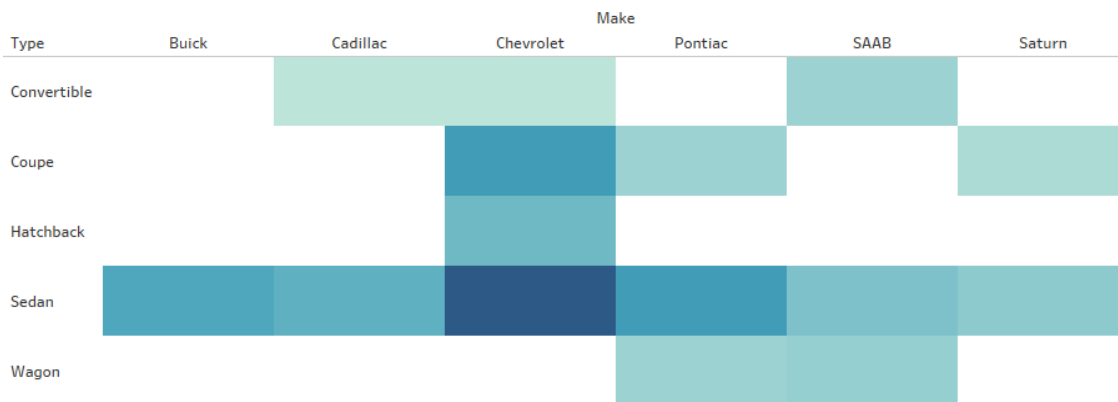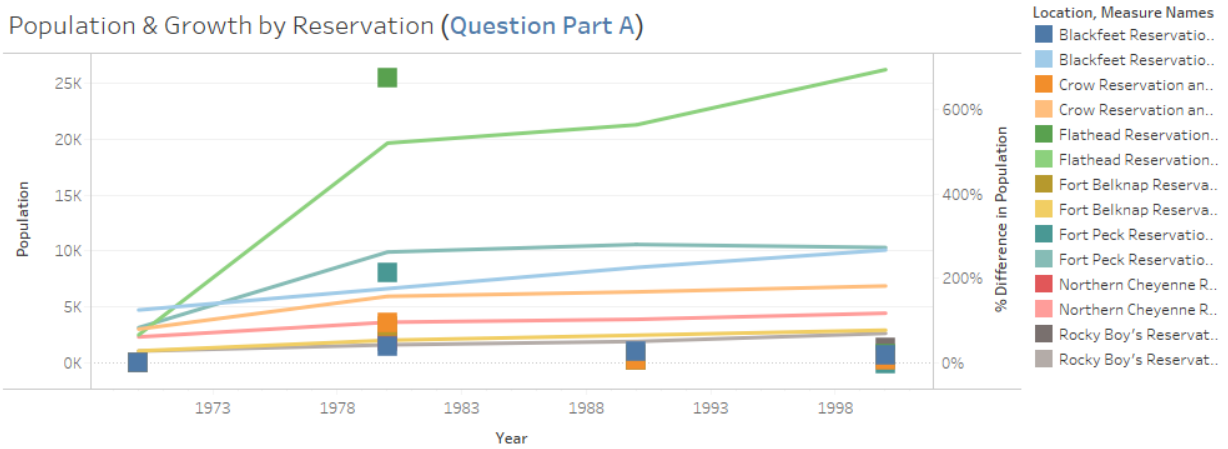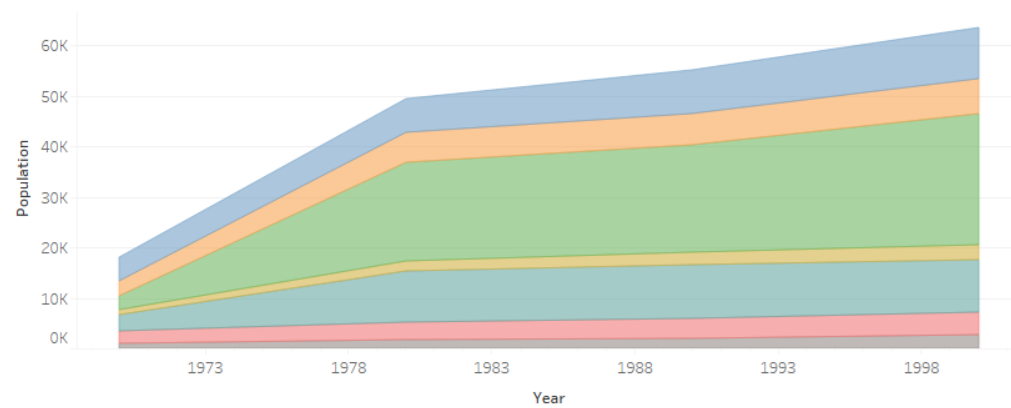

Discussion of Part D

This chart allows the user to see number of types of cars by make and model. Since it is a heat map, contrast can be used to identify patterns. In this case, the graph clearly shows that Chevy Sedans have the most cars.

Question 3

Population & Growth by Reservation (Question Part A)

Location, Measure Names
- Blackfeet Reservatio..
- Blackfeet Reservatio..
- Crow Reservation an..
- Crow Reservation an..
- Flathead Reservation..
- Flathead Reservation..
- Fort Belknap Reserva..
- Fort Belknap Reserva..
- Fort Peck Reservatio..
- Fort Peck Reservatio..
- Northern Cheyenne R..
- Northern Cheyenne R..
- Rocky Boy's Reservat..
- Rocky Boy's Reservat..

Total Population Growth for All Reservations (Stacked) (Question Part B)

R Code:

```
 1   # Written by Patrick Keener
 2   # 1/17/2021
 3   # HW1, Question 5
 4
 5
 6   library(ggplot2)
 7
 8   # DataFrame
 9   cellPlans = data.frame(
10     c("ATT", "Sprint", "Verizon", "ATT", "Sprint",
11       "Verizon", "ATT", "Sprint", "Verizon", "ATT",
12       "Verizon", "Sprint", "Verizon", "ATT",
13       "Verizon", "Sprint", "ATT", "ATT", "Sprint"),
14     c(1, 1, 2, 3, 3, 4, 6, 6, 8, 10, 12, 12, 16, 16,
15       24, 24, 25, 30, 40),
16     c(30, 20, 35, 40, 30, 50, 60, 45, 70, 80, 80, 60,
17       90, 90, 110, 80, 110, 135, 100))
18   names(cellPlans) = c("Company", "DataGB", "Price")
19
20   # Scatter Plot
21   ggplot(cellPlans, aes(x=Price, y=DataGB, color = Company)) +
22     geom_point(size = 5) + ggtitle("Price by Data Allowance by Provider")
23
24
25   |
26   # Graph 2; adapted from R Gfx Cookbook
27   # https://r-graphics.org/recipe-bar-graph-dot-plot
28   ggplot(cellPlans, aes(x = Price, y = DataGB)) +
29     geom_segment(aes(yend = DataGB), xend = 0, colour = "grey50") +
30     geom_point(size = 3, aes(colour = Company)) +
31     scale_colour_brewer(palette = "Set1", limits = c("ATT", "Sprint", "Verizon"), guide = FALSE) +
32     theme_bw() +
33     theme(panel.grid.major.y = element_blank()) +
34     facet_grid(Company ~ ., scales = "free_y", space = "free_y") +
35     ggtitle("Price by Data Allowance by Provider")
36
```

Written by Patrick Keener

1/17/2021

HW1, Question 5

library(ggplot2)

DataFrame

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")


# Scatter Plot

ggplot(cellPlans, aes(x=Price, y=DataGB, color = Company)) +

 geom_point(size = 5) + ggtitle("Price by Data Allowance by Provider")




# Graph 2; adapted from R Gfx Cookbook

# https://r-graphics.org/recipe-bar-graph-dot-plot

ggplot(cellPlans, aes(x = Price, y = DataGB)) +

 geom_segment(aes(yend = DataGB), xend = 0, colour = "grey50") +

 geom_point(size = 3, aes(colour = Company)) +

 scale_colour_brewer(palette = "Set1", limits = c("ATT", "Sprint", "Verizon"), guide = FALSE) +

 theme_bw() +

 theme(panel.grid.major.y = element_blank()) +

 facet_grid(Company ~ ., scales = "free_y", space = "free_y") +

 ggtitle("Price by Data Allowance by Provider")
```