

# Homework 14 Bar Graphs

Stephanie Neul

10/6/2021

\*Note that for all graphs we make this year you will be expected to include titles, labels and legends (when appropriate) unless told otherwise. I will not always prompt you to do this - it should become second nature! Also, colors are always good to see.

1. Load up the following data set which contains some information on recently sold used Toyota Camrys. The Toyota Camry is a sedan; to learn more about the car, go to [https://en.wikipedia.org/wiki/Toyota\\_Camry](https://en.wikipedia.org/wiki/Toyota_Camry).

The .csv file should already be in your Day 14 folder that we downloaded in class, and make sure that this .Rmd file is also saved there. Then look at the first 6 rows of the data.

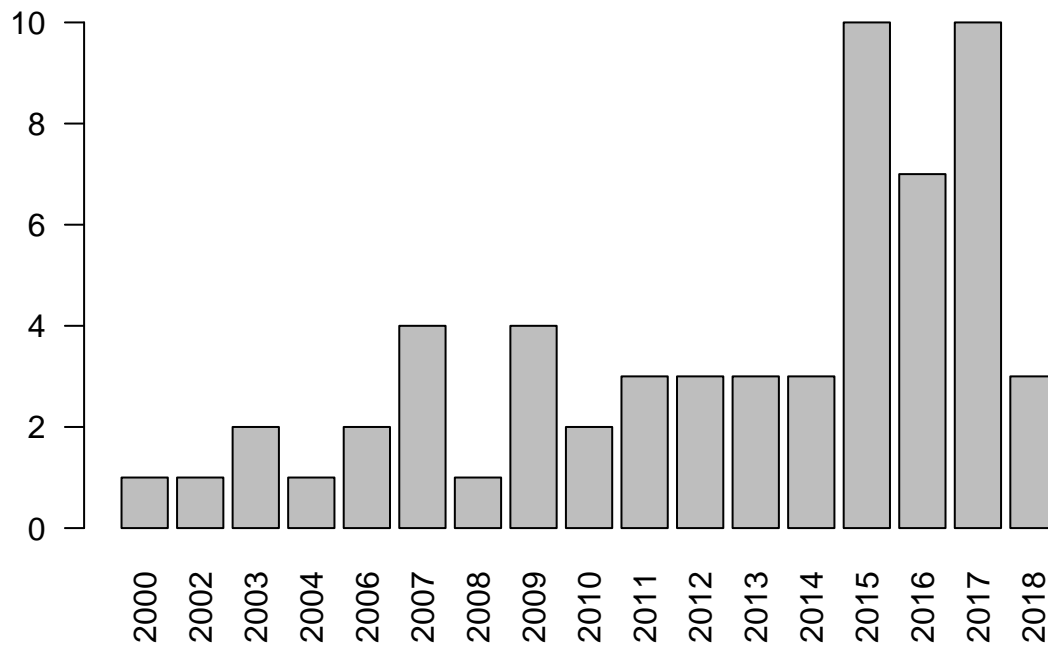
```
camry <- read.csv("Camry.csv")
head(camry)
```

```
##   ModelYear Trim Hybrid Mileage Price
## 1      2018   SE  FALSE   5275 17998
## 2      2015   LE  FALSE  43696 17898
## 3      2017   SE  FALSE  41522 17700
## 4      2018   LE  FALSE   5242 16998
## 5      2014  XLE  FALSE  38410 16965
## 6      2017   SE  FALSE   9930 16299
```

**About the Variables** ModelYear - year car was manufactured. Trim - trim or style of the model Hybrid - is the car a hybrid or not Mileage - miles on car when sold Price - price sold at

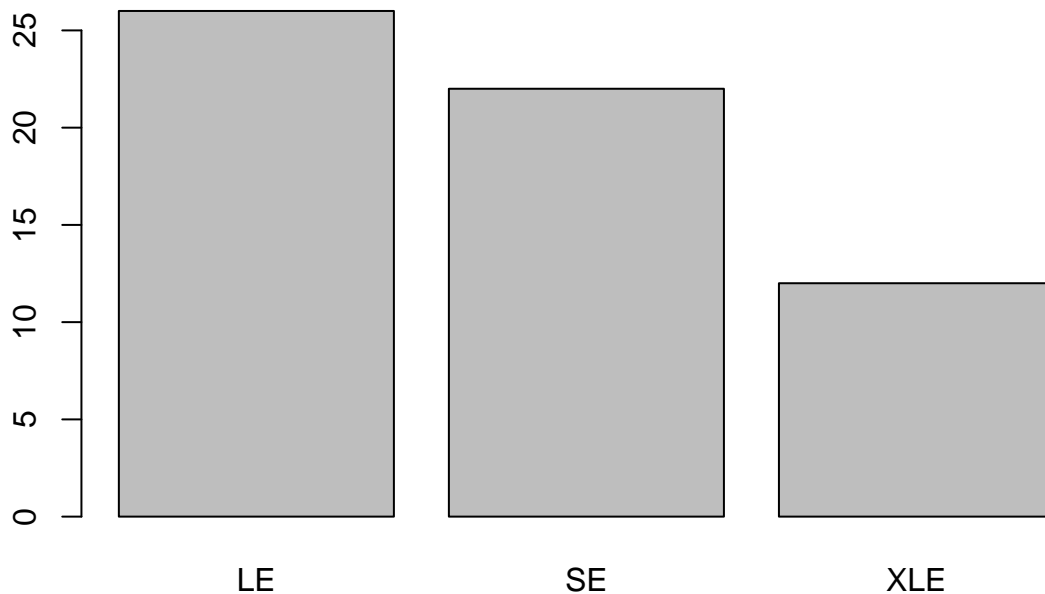
2. Create a bar graph showing how many Camrys were sold by year. You will need to do something to the x-axis so that all of the year labels are visible. You can look back at Day 8 to see how we did that when we made boxplots. There are two ways to fix this.

```
barplot(table(camry$ModelYear),
        las = 2
        )
```



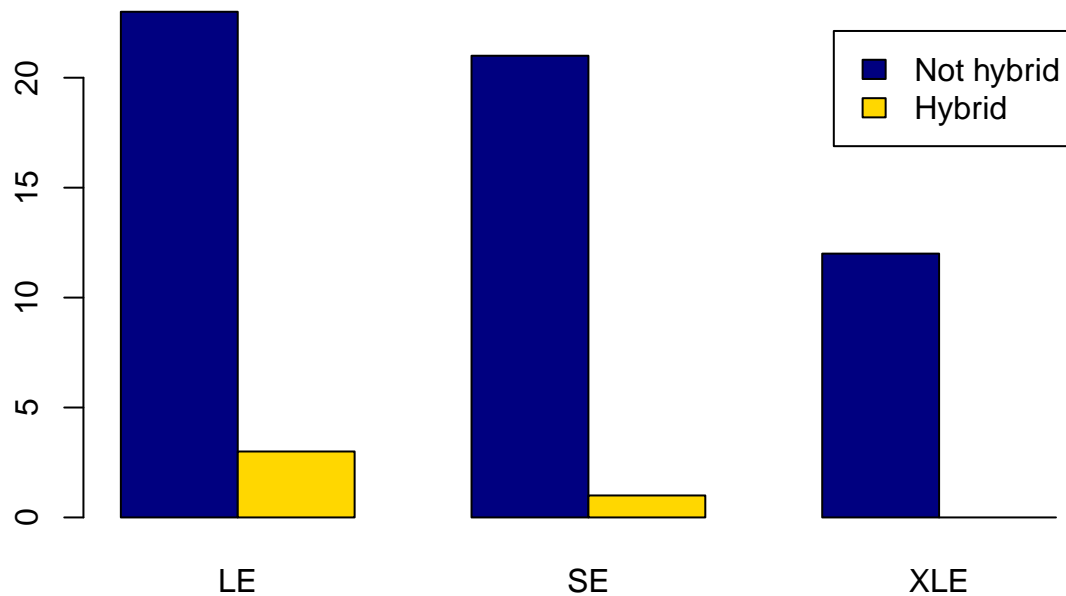
3. Create a bar graph showing the frequency of each Trim. If you don't know what Trim means for cars, read about it here: <https://www.caranddriver.com/research/a32766780/car-trim/>

```
barplot(table(camry$Trim))
```



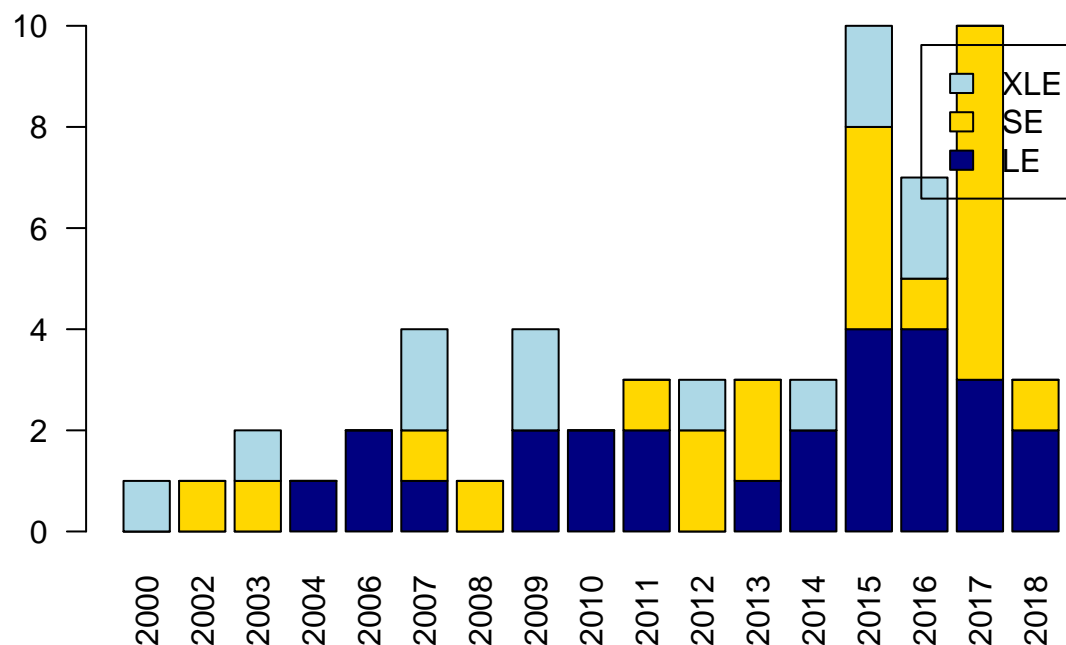
4. Take the graph you just made in problem 2, and now change it so that it shows the frequency of each trim that are hybrids and not hybrids, and make it *side-by-side*. Include a legend so we know which trim is represented by which color in the barplot. Hint: One way to figure out which color represents which trim would be to make a table using those two variables.

```
barplot(table(camry$Hybrid, camry$Trim),
        beside = TRUE,
        col = c("navy", "gold"),
        legend.text = c("Not hybrid", "Hybrid")
)
```



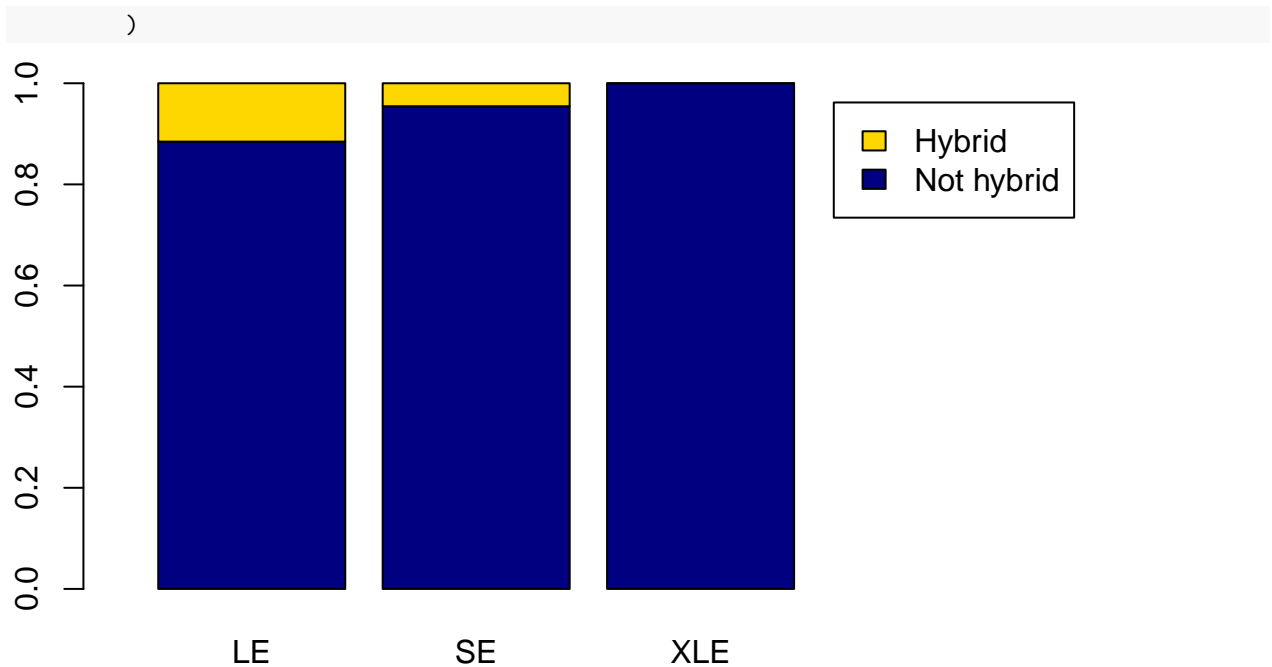
5. Create a *stacked* bar chart showing how many Camrys were sold by year, taking into account their trim. Make sure you have enough colors, and make sure to fix the x-axis labels so that all are visible!

```
barplot(table(camry$Trim, camry$ModelYear),
        col = c("navy", "gold", "lightblue"),
        las = 2,
        legend.text = c("LE", "SE", "XLE")
    )
```



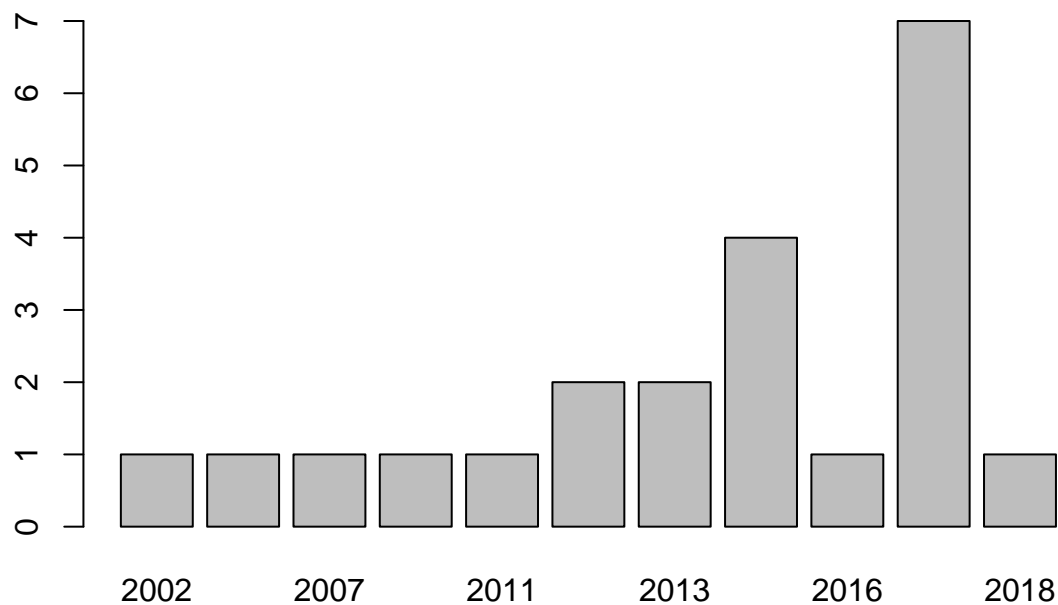
6. Create a segmented bar chart showing what proportion of Camrys of each trim are hybrids vs. not hybrids.

```
barplot(prop.table(table(camry$Hybrid, camry$Trim), 2),
        col = c("navy", "gold"),
        legend.text = c("Not hybrid", "Hybrid"),
        xlim = c(0, 5))
```



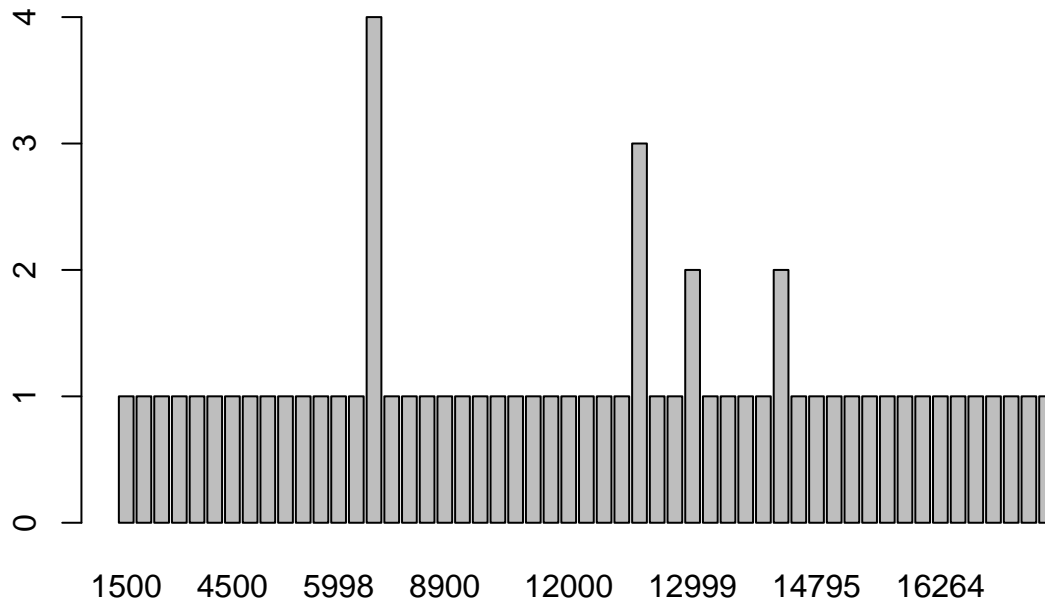
7. Create a bar chart showing only how many Camrys with the trim SE were sold each year. *Hint: you will need to make a subset to do this. The code for subsetting can be found in your Day 14 notes!*

```
camry.se = camry[camry$Trim == "SE",]
barplot(table(camry.se$ModelYear))
```



8. Why does it not make sense to make a bar graph of the `Price` variable? If you aren't sure, try it and it might help you figure out the answer.

```
barplot(table(camry$Price))
```

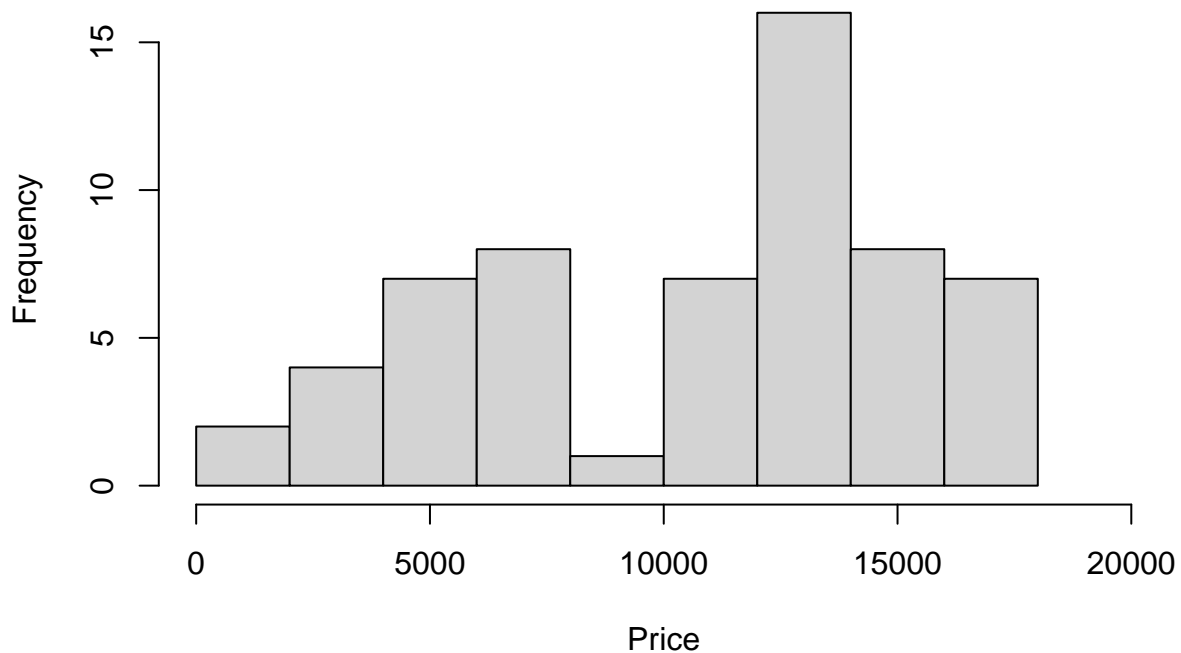


Because price isn't categorical; each car sold may go for a different price. Perhaps a histogram would be better.

- Now make a graph of the `Price` variable using a kind of graph we have learned about that IS suitable (there is more than one option for this).

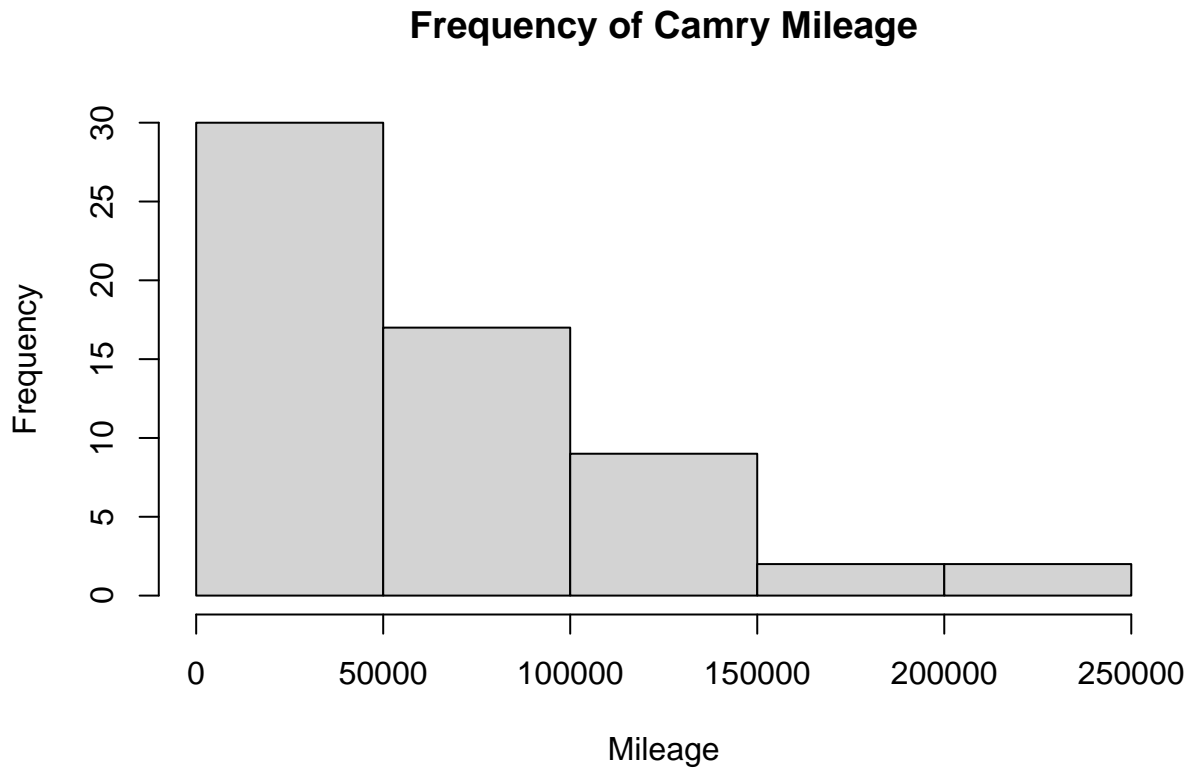
```
hist(camry$Price,
     main = "Frequency of Camry Prices",
     xlab = "Price",
     xlim = c(0, 20000)
)
```

**Frequency of Camry Prices**



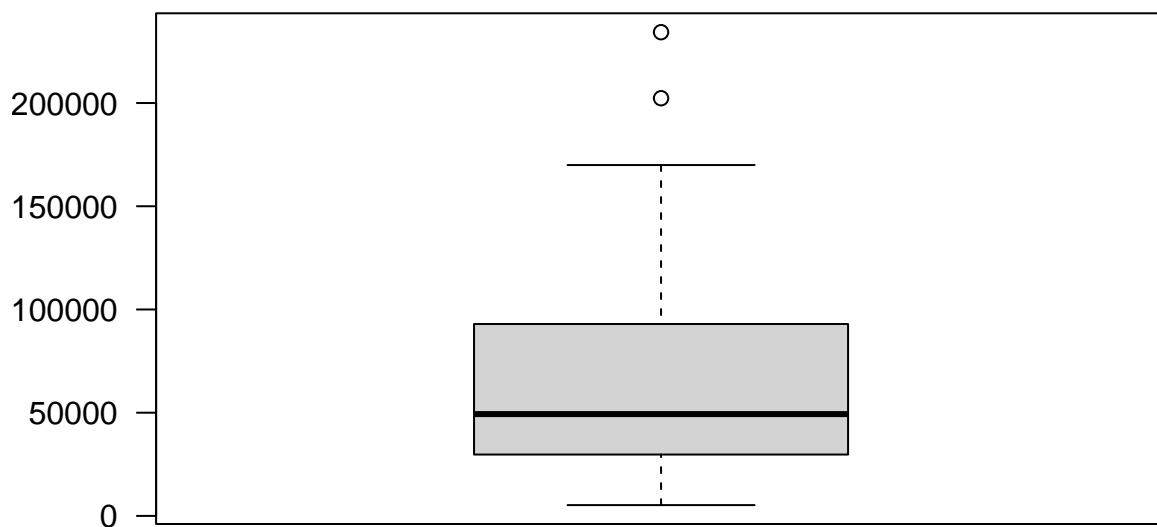
10. Make a histogram of the Mileage variable.

```
hist(camry$Mileage,  
     main = "Frequency of Camry Mileage",  
     xlab = "Mileage"  
)
```



11. Make a boxplot of the Mileage variable.

```
boxplot(camry$Mileage,  
        las = 2)
```



12. Using your graphs from 10 and 11, describe the distribution of the Mileage variable. Make sure to contextualize your description - what have you learned about the mileage of the used cars represented

in this data set from the histogram and boxplot that you made?

The mileage of Camrys has a median of 50,000. Based on the histogram, it seems like most Camrys have a mileage below 50,000 but in the boxplot we can see the IQR spans ~30,000 to ~10,000. Evidently there are two outliers on the boxplot, but that was my guess anyways looking at the histogram.

13. OPTIONAL Challenge: Make a new column in the dataset that takes the **Price** variable and changes somehow so that it DOES make sense to make a barplot of it. There are several ways to do this and some choices to make in the process. Then, make a barplot of your new variable.