Homework 2

John Chandara mrpotatofactory@gmail.com

2/3/2020

This homework is about syntax and basic plotting functions in R, and is **due by the end of Feb 7**.

You can use the RMarkDown file with these homework instructions as a template for your work. Make sure to replace "Your Name" and "your_github_username" in the YAML with your name and Github username.

Submission instructions: Create a folder named `ds202_hw2`, and name the RMarkDown file including your solutions `hw2.Rmd` under this folder. For submission, create a GitHub repository named `ds202_hw2` under your account, and push both `hw2.Rmd` and the knitted `hw2.html` to GitHub before the deadline. I will assume you use the same GitHub username as for your HW1 submission. The latest version of your homework appearing on GitHub before the deadline is graded. *It is crucial to follow the exact folder structure and file names*, so that your homework can be reproduced and graded by an automated script. Some questions will require the following answer format.

Example question and answer

0. What is the smallest positive odd integer? Storage it in a variable named `a0` and print it.

Example answer: In the RMarkDown file, type in

The knitted html file will show:

```
a0 <- 1
a0
```

```
## [1] 1
```

R commands

1. Today's temperature has a high of 24 and a low of 15 degree fahrenheit. Find the corresponding degrees in celsius and storage it in a (single) vector named `cel`. Print the variable.

```
high <- 24
low <- 15

celsius <- function(x) {
    (x - 32) * 5/9
}

cel <- c(celsius(low), celsius(high))

print(cel)
```

```
## [1] -9.444444 -4.444444
```

2. Create a vector of length 200. The odd entries should be 3 and even entries should be 1. Storage the created in a variable named `vec200` and print.

```
vec200 <- vector(,200)
for (i in seq(1:200)) {
  vec200[i] <- if(i %% 2 == 0) 1 else 3
}
print(vec200)
```

```
##   [1] 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3
##  [38] 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1
##  [75] 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3
## [112] 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1
## [149] 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1 3
## [186] 1 3 1 3 1 3 1 3 1 3 1 3 1 3 1
```

3. How many perfect squares are there between 1 and 2001? Storage the answer in a variable named `numPerfect` and print. You may want to look at `?round`.

```
numPerfect <- 0

for (i in seq(1:2001)) {
  if(i^(1/2) == floor(i^(1/2)))
    numPerfect <- numPerfect + 1
}

numPerfect
```

```
## [1] 44
```

Cars and mileage

For this part we will use the `mpg` data set, you can access the data by first installing (using `install.packages('ggplot2')`) and then loading the package `ggplot2` (using `library(ggplot2)`) into your R session. The `mpg` data includes information about fuel economy of popular car models

```
library(ggplot2)
```

for the years 1999 and 2008 as collected by the US Environmental Protection Agency. An answer should include a short paragraph of text for the explanation to a (human) reader, together with appropriate R code, outputs, and graphs.

Most exercises are from the ggplot2 book section 2.2.1. You might find it helpful for working on the homework to read through this section. *However, all questions can be solved by using just the basic R graphics functions.*

A data frame with 234 rows and 11 variables

- manufacturer
- model model name
- displ engine displacement, in litres
- year year of manufacture
- cyl number of cylinders
- trans type of transmission
- drv f = front-wheel drive, r = rear wheel drive, 4 = 4wd
- cty city miles per gallon
- hwy highway miles per gallon

- fl fuel type
- class "type" of car

1. What are the 3 most fuel-efficient models in terms of highway mileage? Store the data frame (with 3 rows) containing all information of these three cars in a variable named `top3` and print.

```
mostEfficient <- mpg$hwy
names(mostEfficient) <- mpg$model
top3 <- sort(mostEfficient, decreasing=TRUE)[1:3]

top3
```

```
##      jetta new beetle new beetle
##         44         44         41
```

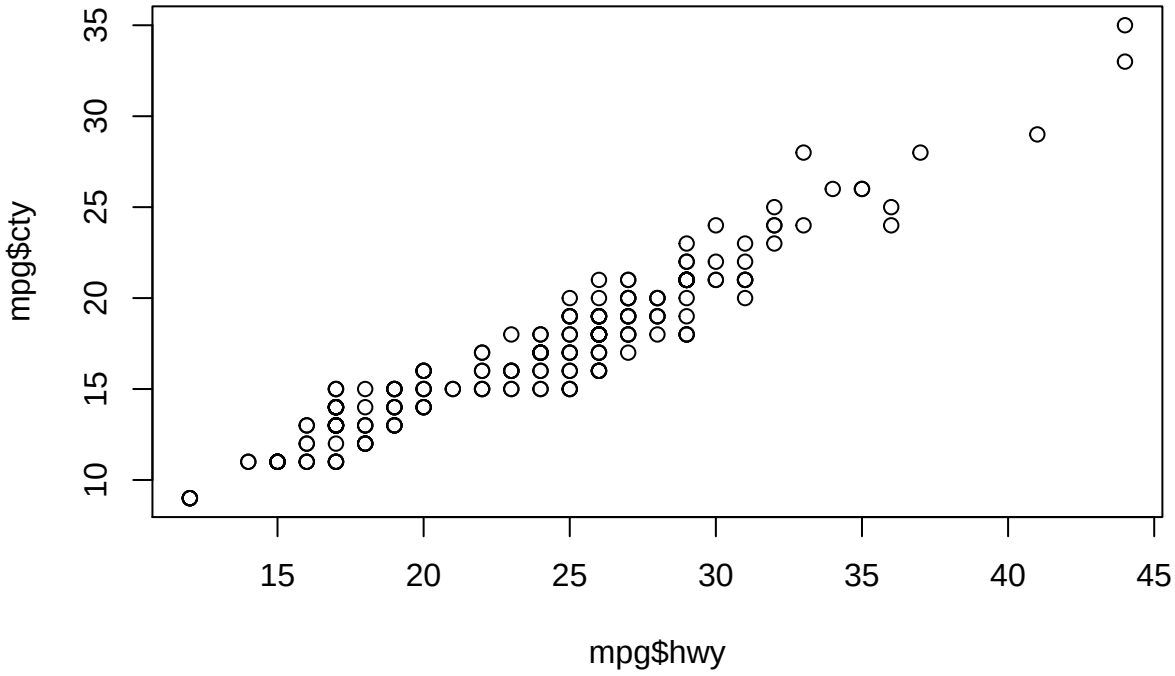
2. How many different compact models are included in this dataset? Storage the answer in a variable named `numCompact` and print.

```
length(mpg$class == 'compact')
```

```
## [1] 234
```

3. Create a scatterplot of `hwy` and `cty` and describe the relationship. Why are there so few points visible? Explain your findings.

```
plot(mpg$hwy, mpg$cty)
```



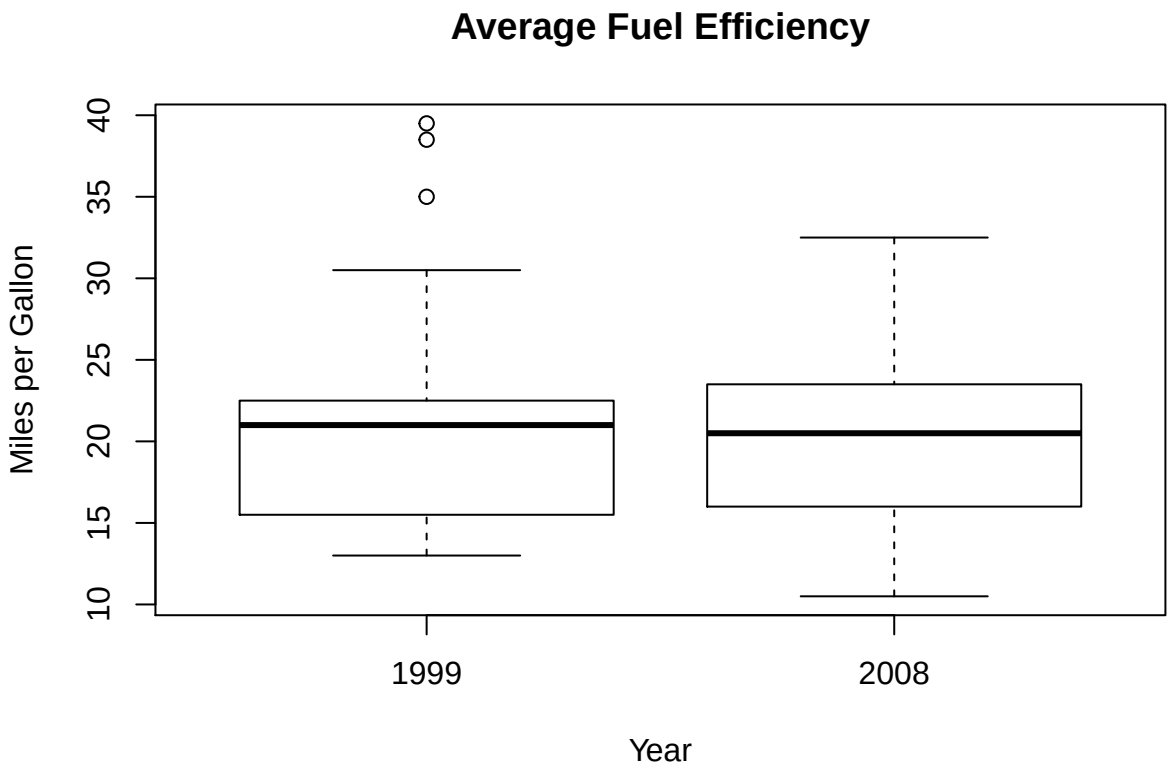
There appears to be a positive semi-strong correlation between highway fuel efficiency and city fuel efficiency. We likely do not find a good amount of points because they are overlapping. The overlap could be a result from little to no fuel efficency innovation over the years, the next question will provide evidence onto this conclusion.

4. Are cars in 2008 more fuel efficient than cars in 1999? Use graphs *and* summary statistics to answer this question. Explain your findings.

```r
mdl <- mpg[mpg$year == 2008 | 1999, ]

mdl$avgEff <- (mdl$hwy + mdl$cty) / 2

boxplot(avgEff ~ year, data=mdl, xlab='Year', ylab='Miles per Gallon', main='Average Fuel Efficiency')
```



We find very little movement in terms of the mean, first quartile and third quartile. This likely represents very little fuel innovation between the years of 1999 and 2008.

Note: your submission is supposed to be fully reproducible, i.e. the TA and I will 'knit' your submission. Make sure that you don't use any file structure that depends on your computing environment.