

# Homework 3

Kevin Jin

3/4/2022

## Problem 1

### Part (a)

```
# Load data
library(ggplot2, warn.conflicts = FALSE)
library(dplyr, warn.conflicts = FALSE)
data("mpg")
# Convert cyl variable from int into ordered factor
mpg$cyl <- factor(mpg$cyl,
                  ordered = TRUE,
                  levels = c("4", "5", "6", "8"))
```

### Part (b)

```
# Convert trans variable from chr into factor
mpg$trans <- substr(mpg$trans, start = 1, stop = 4)
mpg$trans <- factor(mpg$trans,
                   levels = c("auto", "manu"))
```

### Part (c)

```
# Convert drv variable from chr into ordered factor
mpg$drv <- factor(mpg$drv,
                 ordered = TRUE,
                 levels = c("f", "r", "4"))
```

### Part (d)

```
# Convert fl variable from chr into factor
mpg$fl[which(mpg$fl == "p" | mpg$fl == "r")] <- "gasoline"
mpg$fl[which(mpg$fl == "d")] <- "diesel"
mpg$fl[which(mpg$fl == "e" | mpg$fl == "c")] <- "other"
mpg$fl <- factor(mpg$fl,
                 levels = c("diesel", "gasoline", "other"))
```

## Part (e)

```
# Convert class variable from chr into ordered factor
mpg$class <- factor(mpg$class,
                    ordered = TRUE,
                    levels = c("2seater", "subcompact", "compact",
                               "midsize", "suv", "minivan", "pickup"))
```

## Part (f)

```
# Create new country variable with manufacturer base location
mpg <- mutate(mpg, country = NA)
for (row in 1:nrow(mpg)) {
  if (mpg$manufacturer[row] == "audi" |
      mpg$manufacturer[row] == "volkswagen") {
    mpg$country[row] <- "germany"
  } else if (mpg$manufacturer[row] == "hyundai") {
    mpg$country[row] <- "south korea"
  } else if (mpg$manufacturer[row] == "land rover") {
    mpg$country[row] <- "great britain"
  } else if (mpg$manufacturer[row] == "honda" |
              mpg$manufacturer[row] == "nissan" |
              mpg$manufacturer[row] == "subaru" |
              mpg$manufacturer[row] == "toyota") {
    mpg$country[row] <- "japan"
  } else {
    mpg$country[row] <- "united states"
  }
}

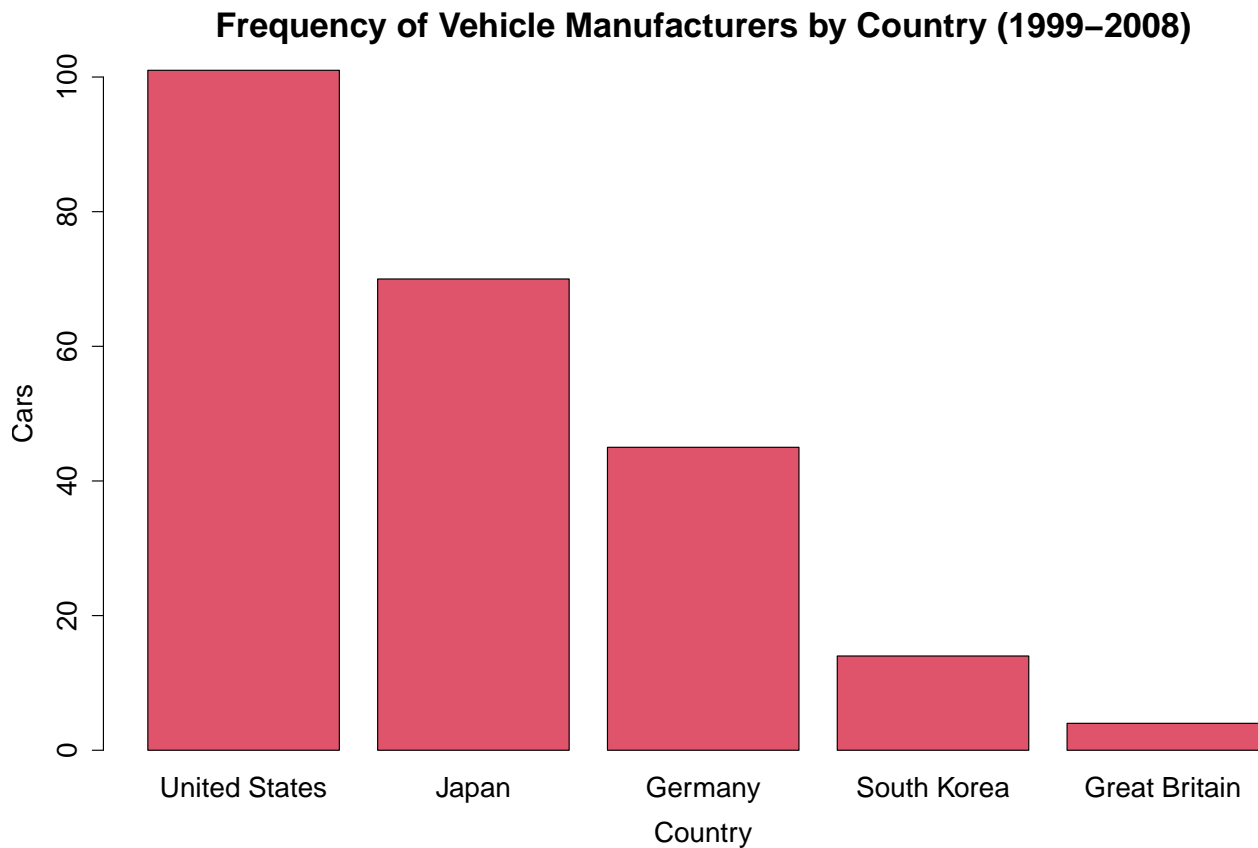
# Check object structure
str(mpg)

## tibble [234 x 12] (S3: tbl_df/tbl/data.frame)
##  $ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
##  $ model       : chr [1:234] "a4" "a4" "a4" "a4" ...
##  $ displ      : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
##  $ year       : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
##  $ cyl        : Ord.factor w/ 4 levels "4"<"5"<"6"<"8": 1 1 1 1 3 3 3 1 1 1 ...
##  $ trans      : Factor w/ 2 levels "auto","manu": 1 2 2 1 1 2 1 2 1 2 ...
##  $ drv        : Ord.factor w/ 3 levels "f"<"r"<"4": 1 1 1 1 1 1 1 3 3 3 ...
##  $ cty        : int [1:234] 18 21 20 21 16 18 18 18 16 20 ...
##  $ hwy        : int [1:234] 29 29 31 30 26 26 27 26 25 28 ...
##  $ fl         : Factor w/ 3 levels "diesel","gasoline",...: 2 2 2 2 2 2 2 2 2 2 ...
##  $ class      : Ord.factor w/ 7 levels "2seater"<"subcompact"<...: 3 3 3 3 3 3 3 3 3 3 ...
##  $ country    : chr [1:234] "germany" "germany" "germany" "germany" ...
```

## Problem 2

### Part (a)

```
# Draw barplot of country variable and arrange in decreasing order
barplot(sort(table(mpg$country), decreasing = TRUE),
        main = "Frequency of Vehicle Manufacturers by Country (1999–2008)",
        xlab = "Country",
        ylab = "Cars",
        names.arg = c("United States", "Japan", "Germany",
                      "South Korea", "Great Britain"),
        col = 2,
        cex.main = 1.8,
        cex.lab = 1.5,
        cex.names = 1.5,
        cex.axis = 1.5)
```



The most common country in this dataset is the United States, and the least common is Great Britain.

### Part (b)

```
# Summarize typical US car
names(which.max(table(mpg$displ))) # Engine displacement (liters)
## [1] "2"
names(which.max(table(mpg$cyl))) # Number of cylinders
```

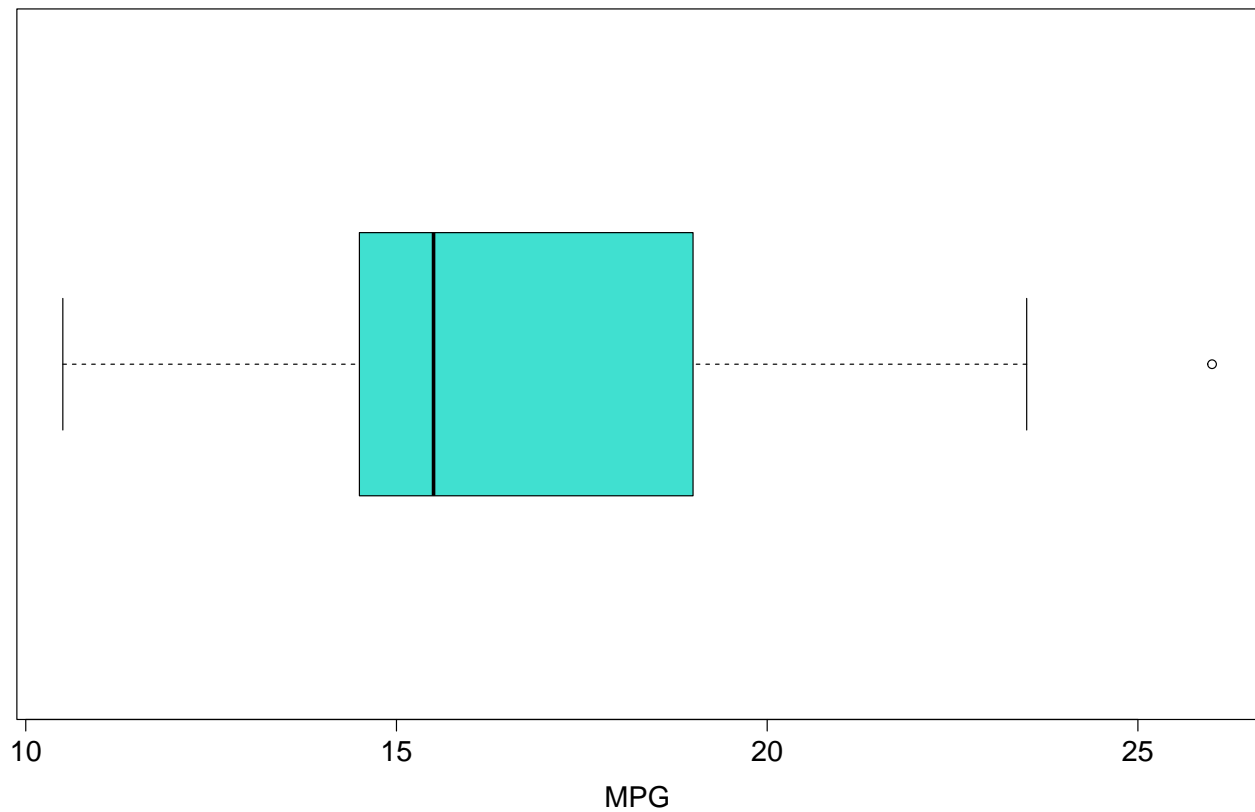
```
## [1] "4"
names(which.max(table(mpg$trans))) # Transmission type
## [1] "auto"
names(which.max(table(mpg$drv))) # Drive type
## [1] "f"
names(which.max(table(mpg$fl))) # Fuel type
## [1] "gasoline"
names(which.max(table(mpg$class))) # Car class
## [1] "suv"
```

Metric	Typical US Car
Displacement (liters)	2
Cylinders	4
Transmission	Automatic
Drive Type	Front-Wheel Drive
Fuel Type	Gasoline
Car Class	SUV

## Part (c)

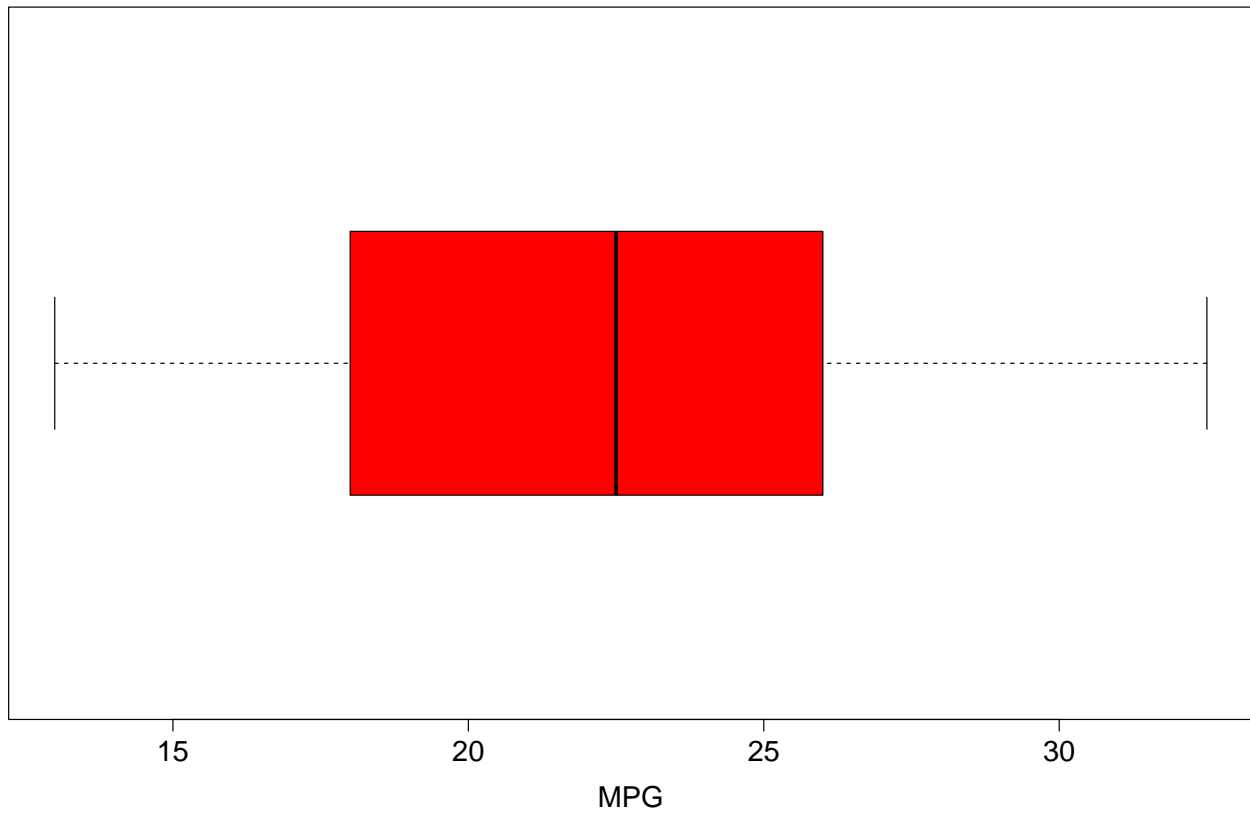
```
# Boxplots of combined mpg of US and Japanese cars
mpg <- mutate(mpg, value = (cty + hwy) / 2) # Create new mpg variable
us <- which(mpg$country == "united states") # Extract indices of US
jp <- which(mpg$country == "japan") # Extract indices of Japan
boxplot(mpg$value[us],
        horizontal = TRUE,
        main = "Combined MPG of US Cars",
        xlab = "MPG",
        col = "turquoise",
        cex.main = 1.8,
        cex.lab = 1.5,
        cex.axis = 1.5)
```

## Combined MPG of US Cars



```
boxplot(mpg$value[jp],  
        horizontal = TRUE,  
        main = "Combined MPG of Japanese Cars",  
        xlab = "MPG",  
        col = "red",  
        cex.main = 1.8,  
        cex.lab = 1.5,  
        cex.axis = 1.5)
```

## Combined MPG of Japanese Cars



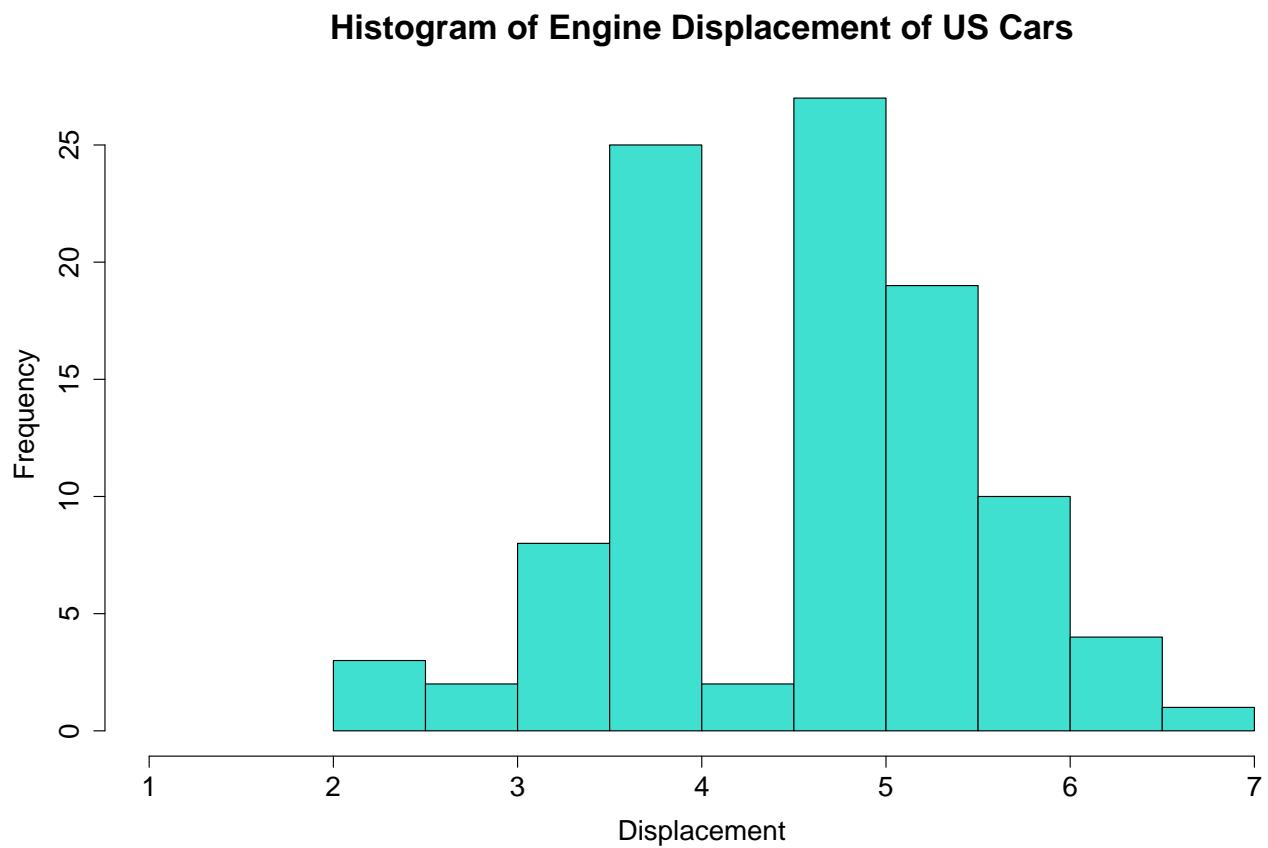
```
# Numerical summaries of combined mpg of US and Japanese cars
summary(mpg$value[us])
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    10.50  14.50   15.50   16.64   19.00   26.00
IQR(mpg$value[us], na.rm = TRUE)
## [1] 4.5
sd(mpg$value[us], na.rm = TRUE)
## [1] 3.302362

summary(mpg$value[jp])
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    13.00  18.38   22.50   22.66   26.00   32.50
IQR(mpg$value[jp], na.rm = TRUE)
## [1] 7.625
sd(mpg$value[jp], na.rm = TRUE)
## [1] 4.60208
```

Metric	MPG of US Cars	MPG of Japanese Cars
Mean	16.64	22.66
Median	15.50	22.50
Standard Deviation (2 d.p.)	3.30	4.60
Inter-Quartile Range (2 d.p.)	4.50	7.63

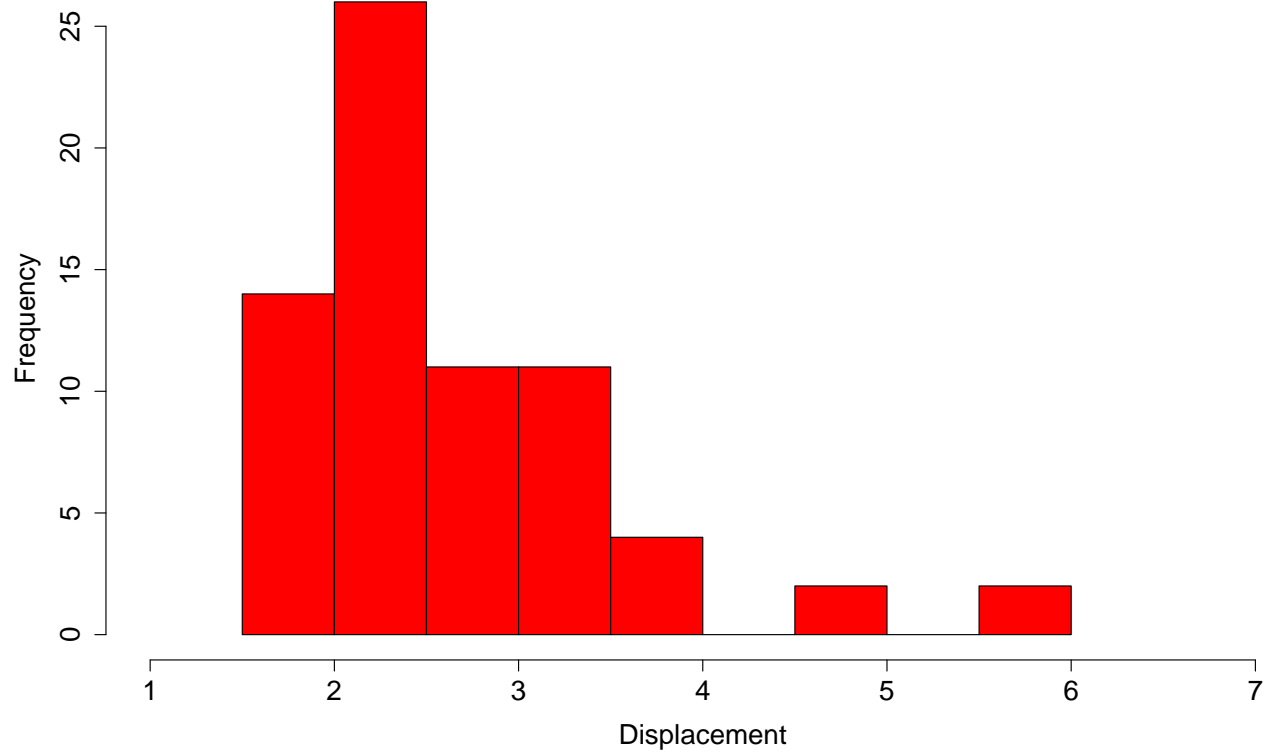
## Part (d)

```
# Histograms of engine displacement of US and Japanese cars
hist(mpg$displ[us],
     main = "Histogram of Engine Displacement of US Cars",
     xlab = "Displacement",
     col = "turquoise",
     xlim = c(1, 7),
     cex.main = 1.8,
     cex.lab = 1.5,
     cex.axis = 1.5
)
```



```
hist(mpg$displ[jp],
     main = "Histogram of Engine Displacement of Japanese Cars",
     xlab = "Displacement",
     col = "red",
     xlim = c(1, 7),
     cex.main = 1.8,
     cex.lab = 1.5,
     cex.axis = 1.5
)
```

**Histogram of Engine Displacement of Japanese Cars**



Histogram	Shape
US Cars	Bimodal
Japanese Cars	Left Skewed