Lab: Data Wrangling Review 1

This lab will simply use the `mpg` dataset included with ggplot2.

```
library(tidyverse)
```

Part 1

1. Check to see if you have the `mpg` dataset.

```
data = mpg
head(mpg)
```

```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p     compa~
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     compa~
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     compa~
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     compa~
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p     compa~
## 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p     compa~
```

2. What data class is `mpg`?

```
class(mpg)
```

```
## [1] "tbl_df"     "tbl"        "data.frame"
```

3. How many observations (rows) and variables (columns) are in the `mpg` dataset?

```
# Install and load necessary package
install.packages("ggplot2")
```

```
## Warning: package 'ggplot2' is in use and will not be installed
```

```
library(ggplot2)

# Access the mpg dataset
mpg <- ggplot2::mpg
# It has 234 rows and 11 columns
```

4. Select the `manufacturer`, `model` and `year` columns from the `mpg` dataset.

```r
selected_data <- subset(mpg, select = c(manufacturer, model, year))
# OR
selected_data <- mpg[, c("manufacturer", "model", "year")]
View(selected_data)
```

5. Identify the subset of cars/rows where city fuel economy (`cty`) is greater than 20 and highway fuel economy (`hwy`) is greater than 30. Assign this output to an object called `eff`. How many cars/rows are present?

```r
eff <- subset(mpg, cty > 20 & hwy > 30)
   # Or using logical indexing
eff <- mpg[mpg$cty > 20 & mpg$hwy > 30, ]
nrow(eff)
```

```
## [1] 21
```

6. How many fuel efficient cars (in the `eff` dataset) were manufactured in the year 1999?

```r
eff_1999 <- subset(eff, year == 1999)
   # Or using logical indexing
eff_1999 <- eff[eff$year == 1999, ]
nrow(eff_1999)
```

```
## [1] 9
```

7. Filter cars from the overall `mpg` dataset that do not have 4 cylinder engines. How many cars/rows are there?

```r
not_4cyl <- subset(mpg, cyl != 4)
   # Or using logical indexing
not_4cyl <- mpg[mpg$cyl != 4, ]
nrow(not_4cyl)
```

```
## [1] 153
```

8. Filter cars to only those in the "suv" or "minivan" class. How many cars/rows are there?

```r
suv_minivan <- subset(mpg, class %in% c("suv", "minivan"))
   # Or using logical indexing
suv_minivan <- mpg[mpg$class %in% c("suv", "minivan"), ]
nrow(suv_minivan)
```

```
## [1] 73
```

9. Filter cars with displacements (`displ`) greater than 4 and that are all 4 wheel drive (`drv`) (a value of 4 for `drv`). How many cars/rows are there?

```r
filtered_cars <- subset(mpg, displ > 4 & drv == "4")
   # Or using logical indexing
filtered_cars <- mpg[mpg$displ > 4 & mpg$drv == "4", ]
nrow(filtered_cars)
```

```
## [1] 49
```