**Practice: Data Wrangling with `dplyr`**
STAT 450, Fall 2024

Here are some practice problems on data wrangling with the R package `dplyr`. These exercises will help prepare you for the midterm.

```
library(tidyverse)
```

All exercises use the `mpg` data set.

```
mpg
```

```
# A tibble: 234 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~ f        18    29 p     comp~
 2 audi         a4           1.8  1999     4 manu~ f        21    29 p     comp~
 3 audi         a4           2    2008     4 manu~ f        20    31 p     comp~
 4 audi         a4           2    2008     4 auto~ f        21    30 p     comp~
 5 audi         a4           2.8  1999     6 auto~ f        16    26 p     comp~
 6 audi         a4           2.8  1999     6 manu~ f        18    26 p     comp~
 7 audi         a4           3.1  2008     6 auto~ f        18    27 p     comp~
 8 audi         a4 quattro   1.8  1999     4 manu~ 4        18    26 p     comp~
 9 audi         a4 quattro   1.8  1999     4 auto~ 4        16    25 p     comp~
10 audi         a4 quattro   2    2008     4 manu~ 4        20    28 p     comp~
# i 224 more rows
```

Type `help(mpg)` to read about this data set in the help menu.

**Exercise 1**

Use `filter()` to subset all rows corresponding to cars that

(a) Have less than 17 city miles per gallon.
(b) Have less than 17 city miles per gallon and are midsize.
(c) Have less than 17 city miles per gallon and are midsize with front-wheel drive.
(d) Are manufactured by honda, hyundai, or subaru.

## Exercise 2

Use `mutate()` to add a new column called `mpg_diff` which is the difference between the highway and city miles per gallon. Make a histogram and density plot for this new variable.

## Exercise 3

Use `group_by()` and `summarize()` to create a data frame with the following columns:

(a) Count number of cars for each car type
(b) Mean city miles per gallon for each car type
(c) Standard deviation of city miles per gallon for each car type
(d) Mean highway miles per gallon for each car type
(e) Standard deviation of highway miles per gallon for each car type

Your R code should recreate the following table:

```
# A tibble: 7 x 6
  class        count cty_mean cty_sd hwy_mean hwy_sd
  <chr>        <int>    <dbl>  <dbl>    <dbl>  <dbl>
1 2seater          5     15.4  0.548     24.8   1.30
2 compact         47     20.1  3.39      28.3   3.78
3 midsize         41     18.8  1.95      27.3   2.14
4 minivan         11     15.8  1.83      22.4   2.06
5 pickup          33     13    2.05      16.9   2.27
6 subcompact      35     20.4  4.60      28.1   5.38
7 suv             62     13.5  2.42      18.1   2.98
```

## Exercise 4

For each car type (`class`), compute the proportion of cars that have over 25 highway miles per gallon. Arrange the rows in descending order according this proportion.