# Homework 1: R for Data Science / 3.2.4

Nickhil-Sethi

8/31/2020

Before completing the questions, let's load the `tidyverse` library.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

### 1) Run `ggplot(data = mpg)`. What do you see?

Running the above command, we actually see a blank screen (see next page).

```
ggplot(data = mpg)
```

## 2) How many rows are in `mpg`? How many columns?

`mpg` has 234 rows and 11 columns, as can be seen from printing the truncated dataframe:

```
mpg
```

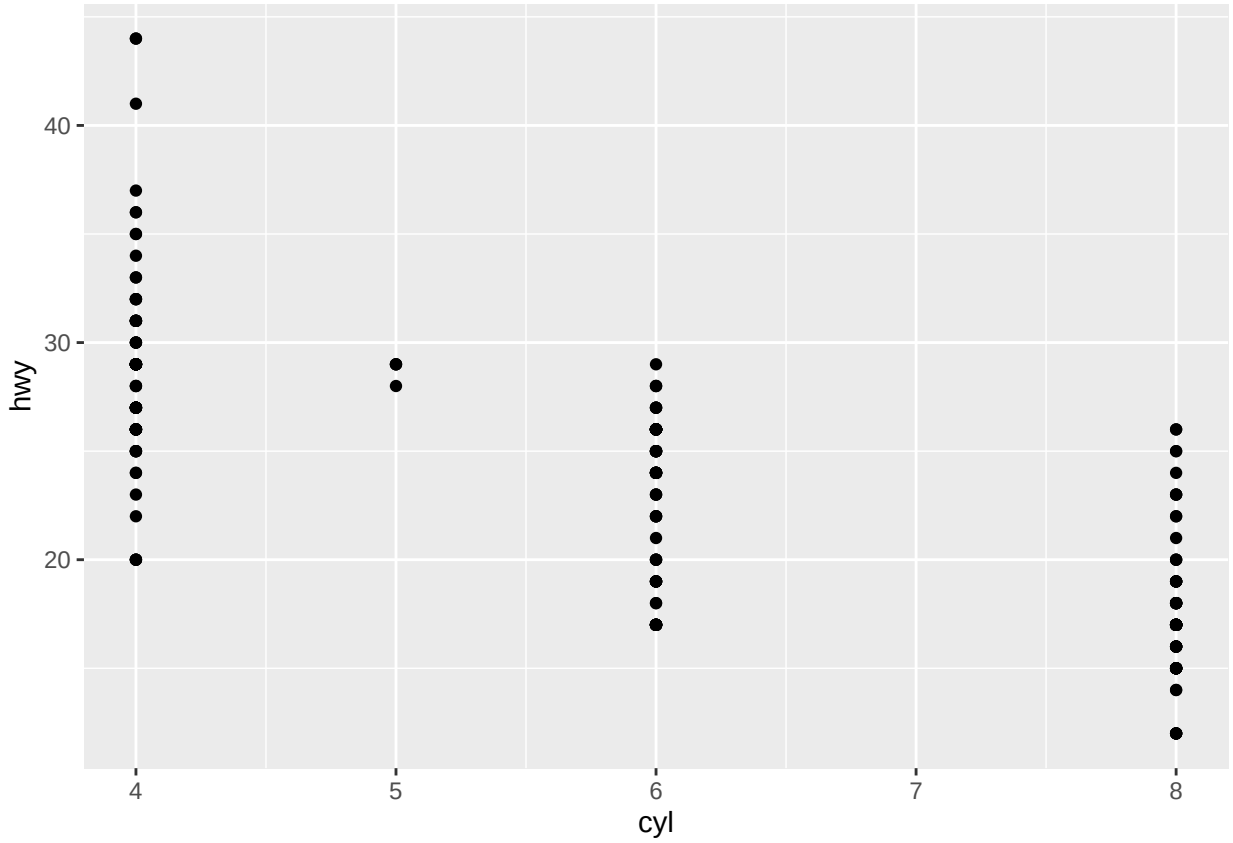
```
## # A tibble: 234 x 11
##    manufacturer model    displ  year   cyl trans   drv     cty   hwy fl    class
##    <chr>        <chr>    <dbl> <int> <int> <chr>   <chr> <int> <int> <chr> <chr>
##  1 audi         a4         1.8  1999     4 auto(l~ f        18    29 p     comp~
##  2 audi         a4         1.8  1999     4 manual~ f        21    29 p     comp~
##  3 audi         a4         2    2008     4 manual~ f        20    31 p     comp~
##  4 audi         a4         2    2008     4 auto(a~ f        21    30 p     comp~
##  5 audi         a4         2.8  1999     6 auto(l~ f        16    26 p     comp~
##  6 audi         a4         2.8  1999     6 manual~ f        18    26 p     comp~
##  7 audi         a4         3.1  2008     6 auto(a~ f        18    27 p     comp~
##  8 audi         a4 quat~   1.8  1999     4 manual~ 4        18    26 p     comp~
##  9 audi         a4 quat~   1.8  1999     4 auto(l~ 4        16    25 p     comp~
## 10 audi         a4 quat~   2    2008     4 manual~ 4        20    28 p     comp~
## # ... with 224 more rows
```

## 3) What does the `drv` variable describe? Read the help for `?mpg` to find out.

`drv` is a categorical variable describing type of drive train; values `f`, `r`, and `4` indicate front-wheel drive, rear-wheel drive, and four-wheel drive, respectively.

### 4) Make a scatterplot `hwy` vs `cyl`.

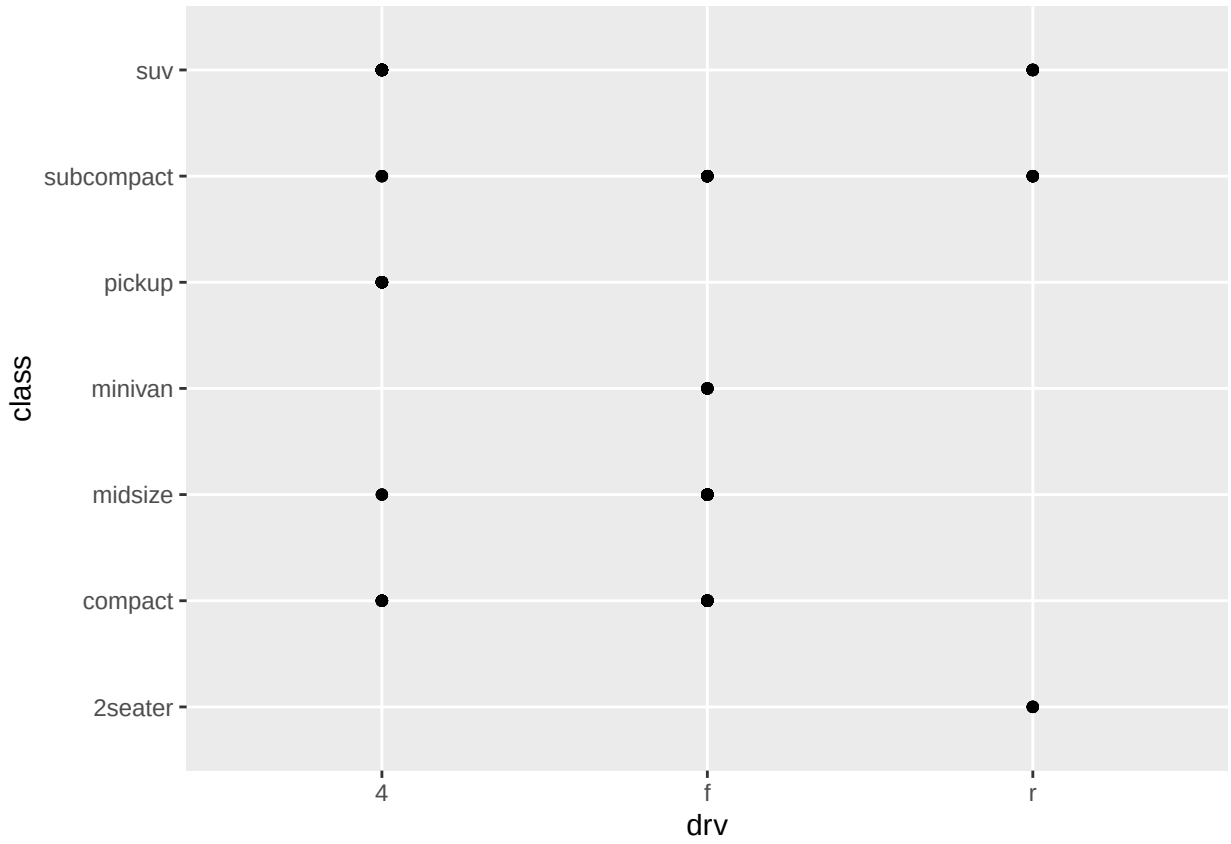
```
ggplot(data = mpg) + geom_point(mapping = aes(x = cyl, y = hwy))
```




### 5) What happens if you make a scatterplot of `class` vs `drv`? Why is the plot not useful?

See next page for plot. It's difficult to interpret this plot because both `drv` and `class` are categorical variables. There is no order structure on either of the variables i.e. the way we embed them into the x and y axes is arbitrary, and so any 'trend' or 'line of fit' here is meaningless.

```
ggplot(data = mpg) + geom_point(mapping = aes(x = drv, y = class))
```


suv
subcompact
pickup
minivan
midsize
compact
2seater
class
4
f
r
drv