Lecture Assignment 3

Viraj Vijaywargiya

2022-04-08

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

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

Part 3.2.4

Question 1

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


Based on the output, it looks like R is not rendering the graph. It looks to me like I need the additional function code `+ geom_function()`

Question 2

```
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~
## # ... with 224 more rows
```

There are 234 Rows, and 11 columns

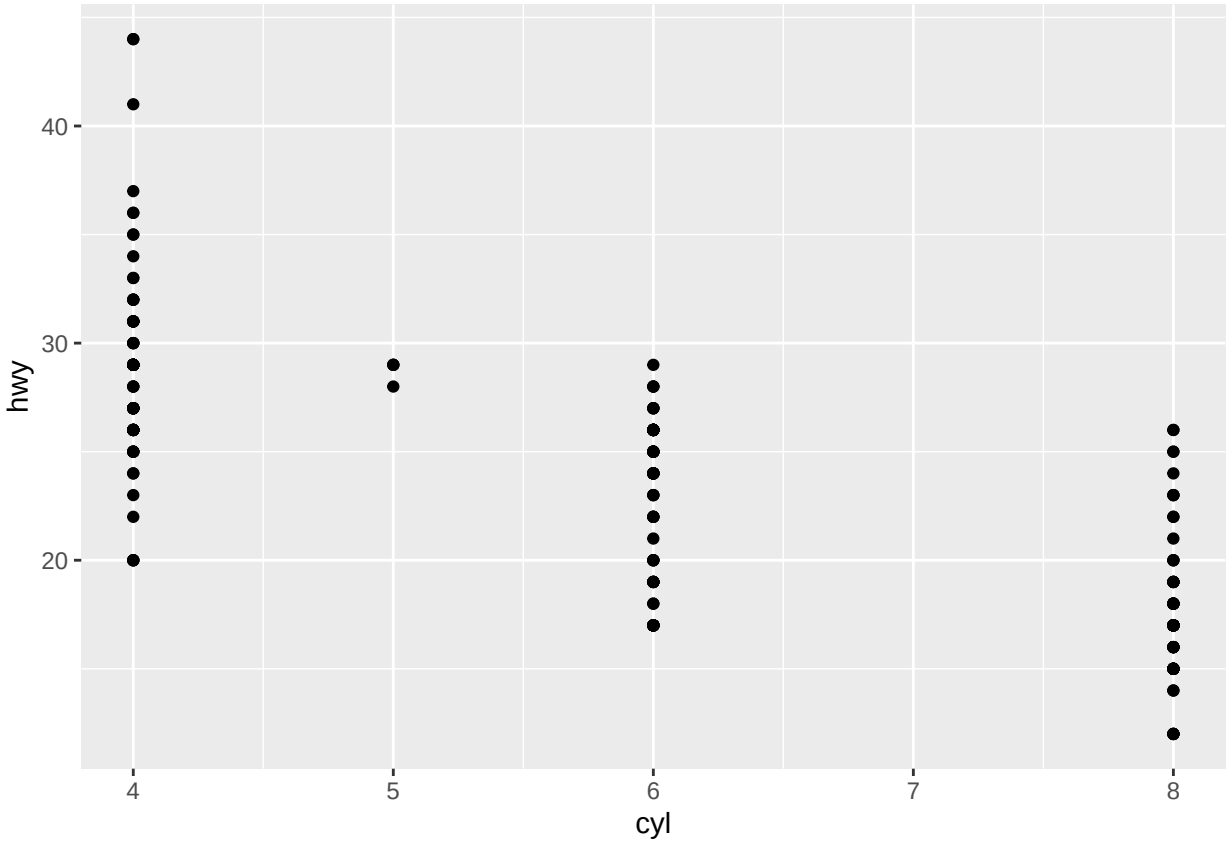
Question 3

```
?mpg
```

The drv variable is the type of drive train, where f = front-wheel drive, r = rear wheel drive, 4 = 4wd

Question 4

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

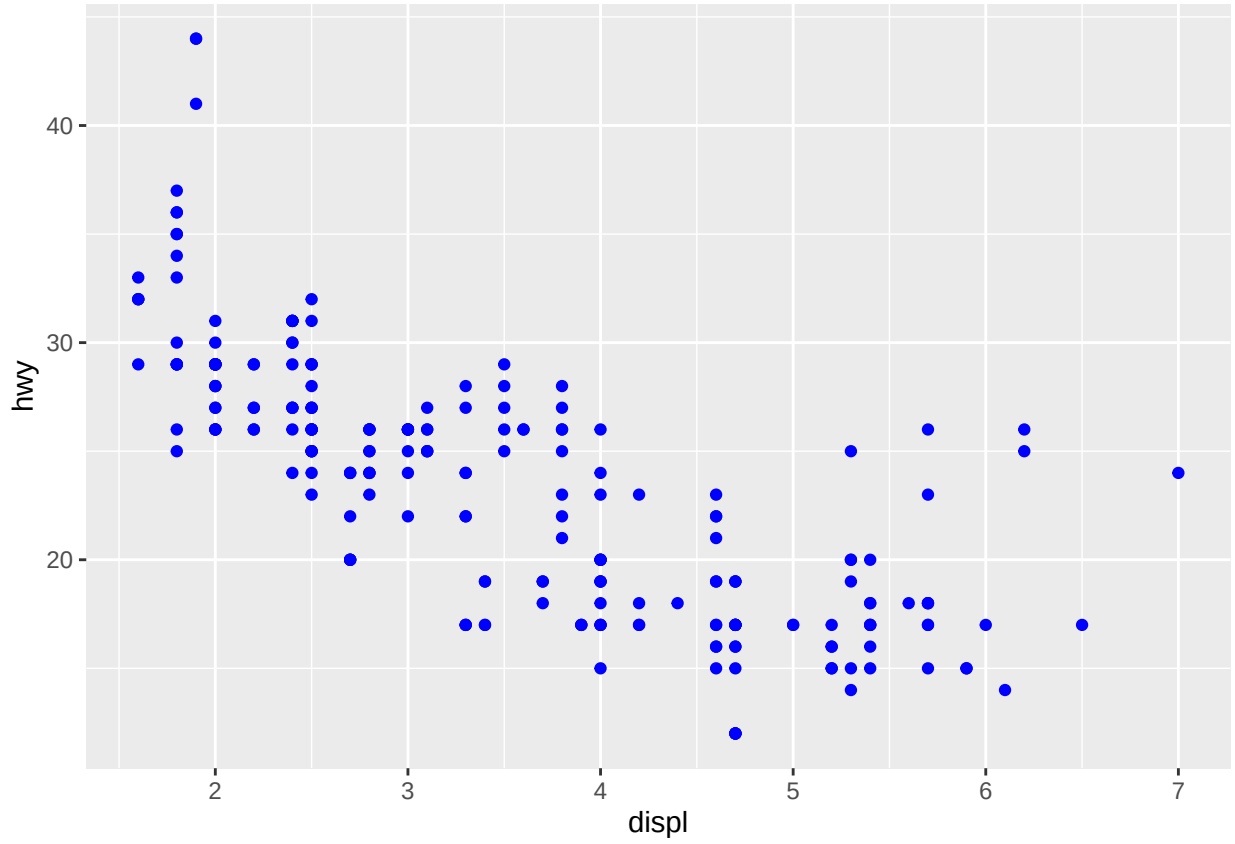


Part 3.3.1

Question 1

The points are not blue because color = "blue" is inside aes(). It should be set manually by setting it as an argument of the geom function i.e. it goes outside of the aes() like shown below.

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy), color = "blue")
```



Question 2

```
?mpg
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~
## # ... with 224 more rows
```

Categorical variables: manufacturer, model, trans, drv, fl, class. Quantitative variables: displ, year, cyl, cty, hwy

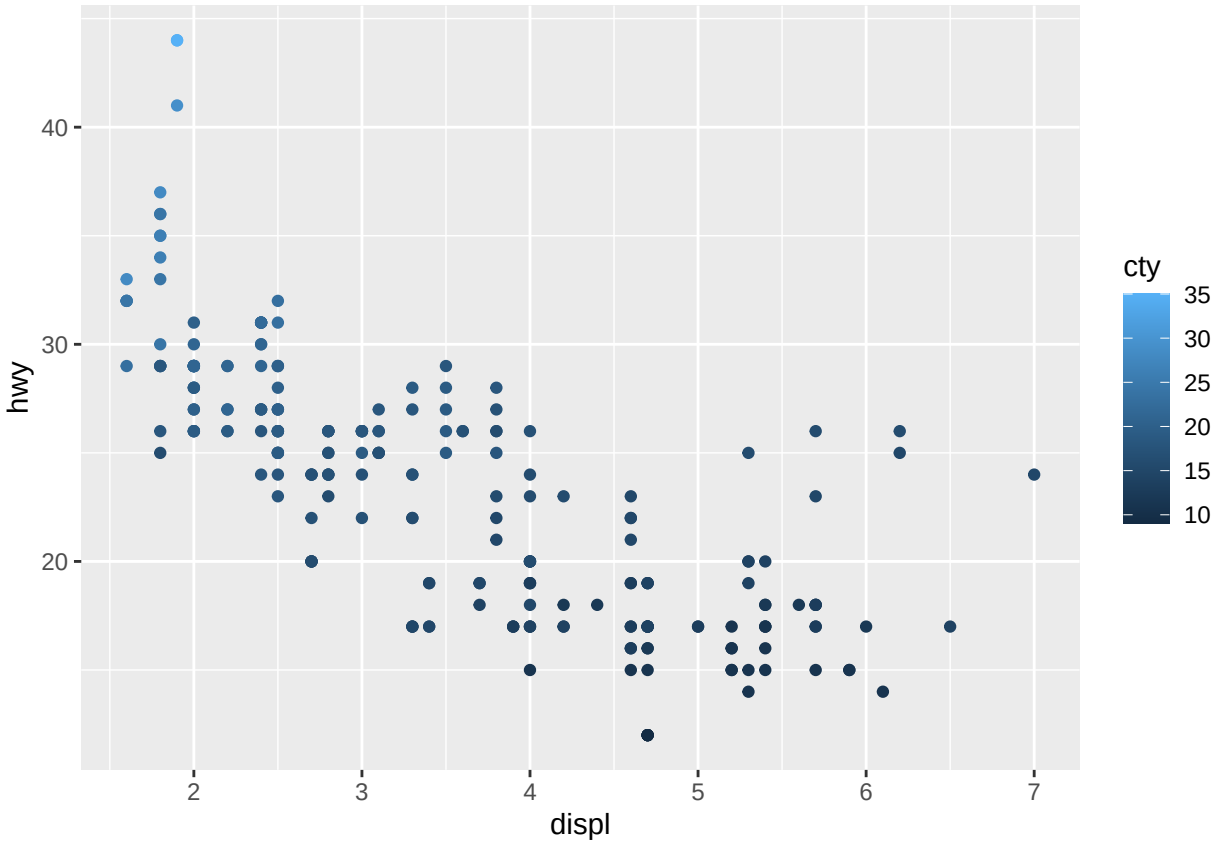
The columns in the mpg data frame represent the variables. There are 11 columns and therefore, 11 variables. Numeric variables are quantitative variables and non-numeric variables are categorical variables.

Question 3

Using the continuous variable, cty.

Mapping to color,

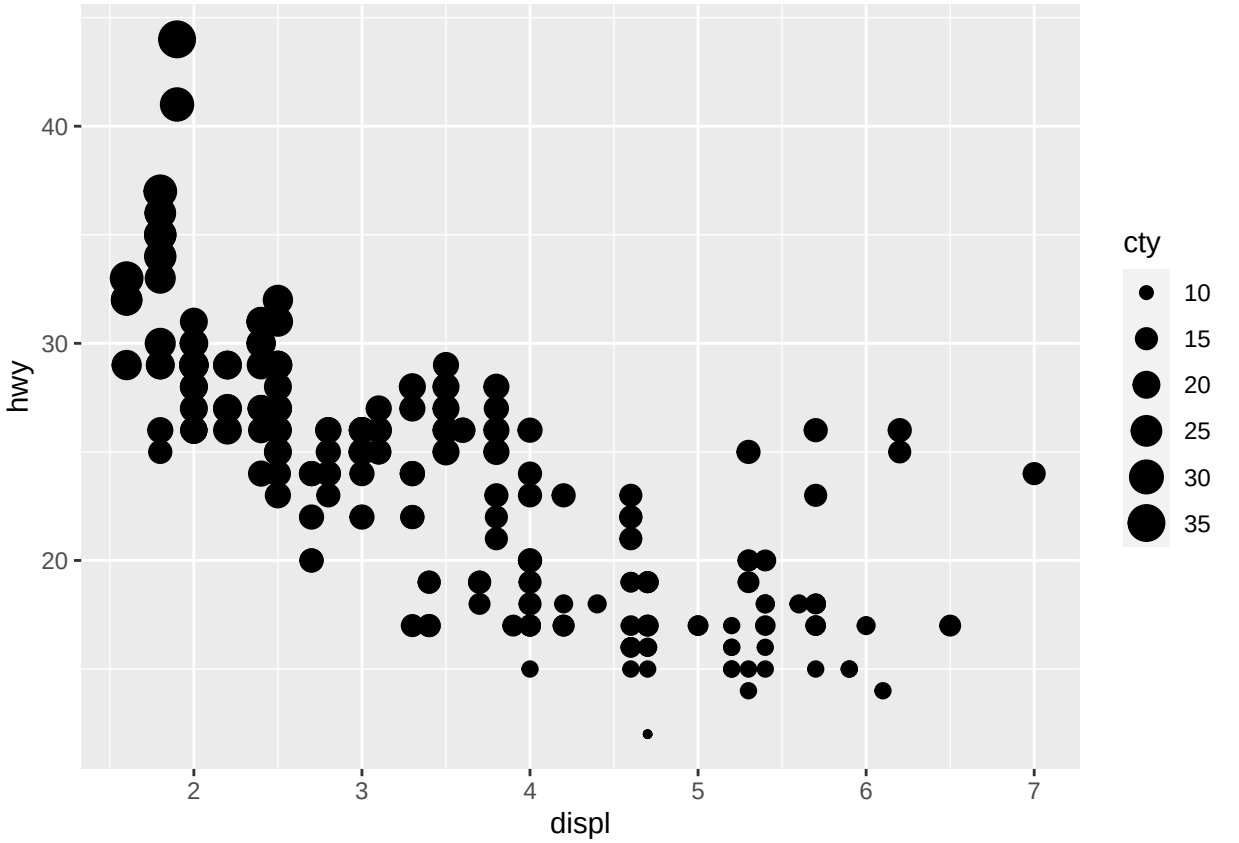
```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, color = cty))
```



For the continous variable, cty, the color scale used varies from light to dark blue. Whereas, for discrete variables, the color scale uses discrete colors.

Mapping to size,

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, size = cty))
```



For continuous variable, cty, the sizes of the points vary continuously corresponding to the function of their size.

For continuous variable, cty, mapping to shape gives an error saying continuous variables cannot be mapped to shape. This is possibly because it is unknown which shape is smaller/greater than the other.