COMPSCIX 415.2 Homework 3

Santosh Kanutala

6/22/2018

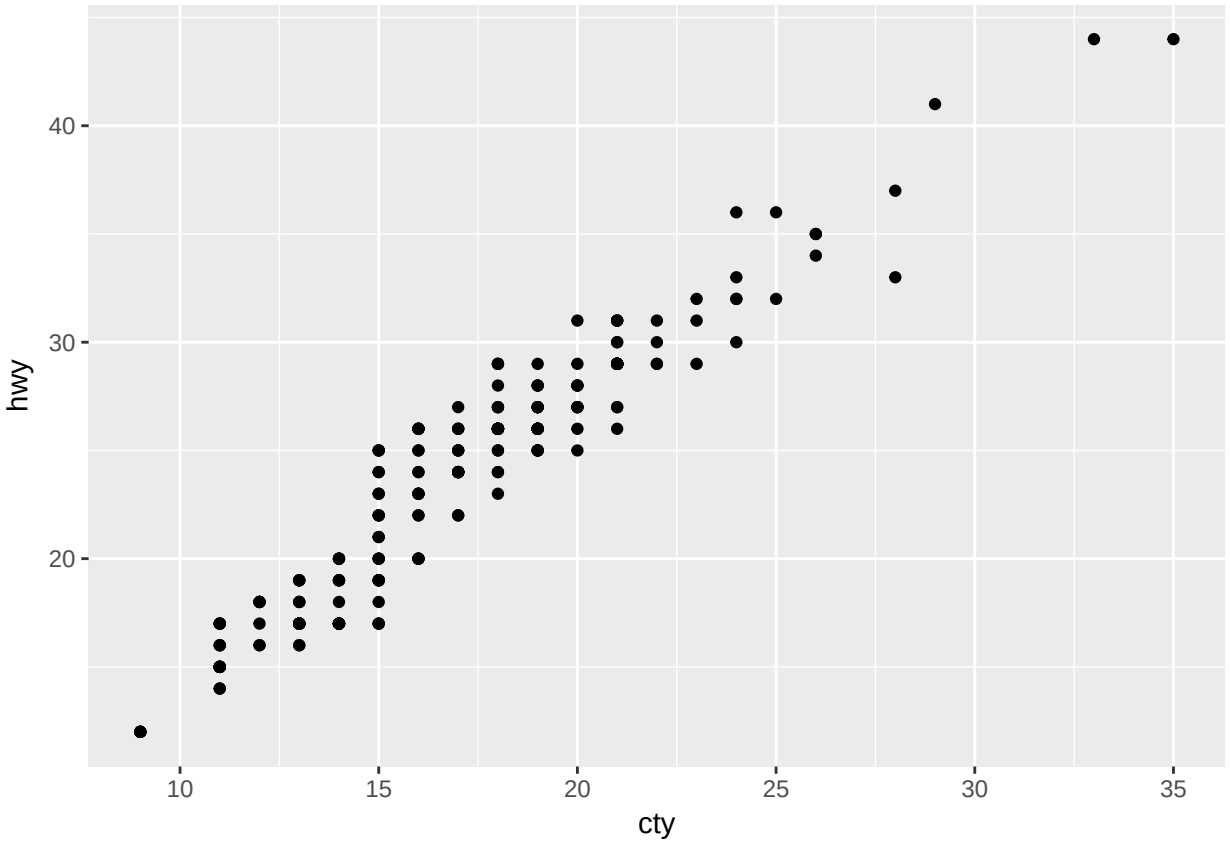
****My Github repository for my assignments can be found at below URL: (https://github.com/santumagic/compscix-415-2assignments.git)****

```
library(tidyverse)
library(mdsr)
```

Section 3.8.1: all excercises

QUESTION 1:

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



ANSWER:

From the mpg dataset we know that cty and hwy both are continuous variables and when we plot them in a single plot, many data points will be overlapped especially for larger datasets. We can resolve this issue

(overplotting) by using adjustment to jitter with position = “jitter” or by using geom_jitter () as shown below.

ggplot(data = mpg, mapping = aes(x = cty, y = hwy)) + geom_point() + geom_jitter()

QUESTION 2:

ANSWER:

Lets find from the help function.

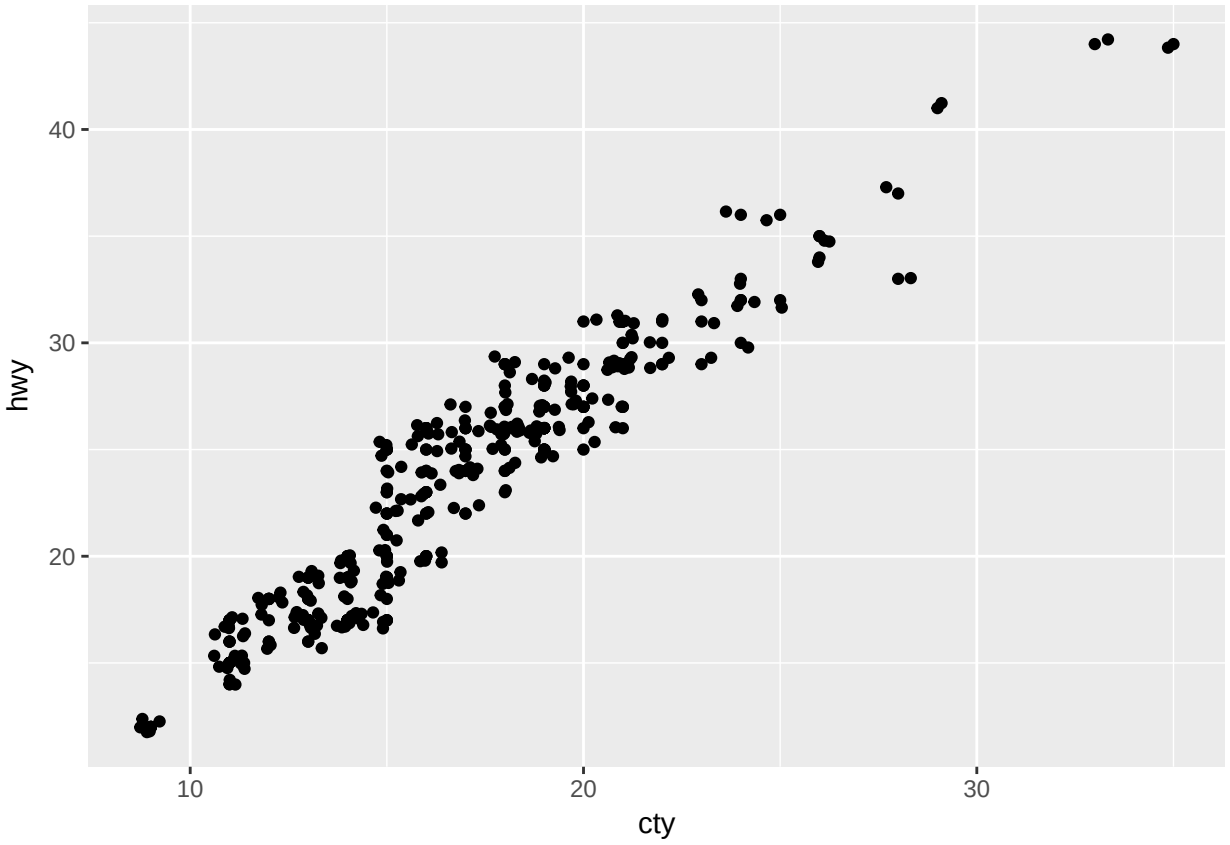
```
?geom_jitter
```

width and hight are the parameters that control the jittering.

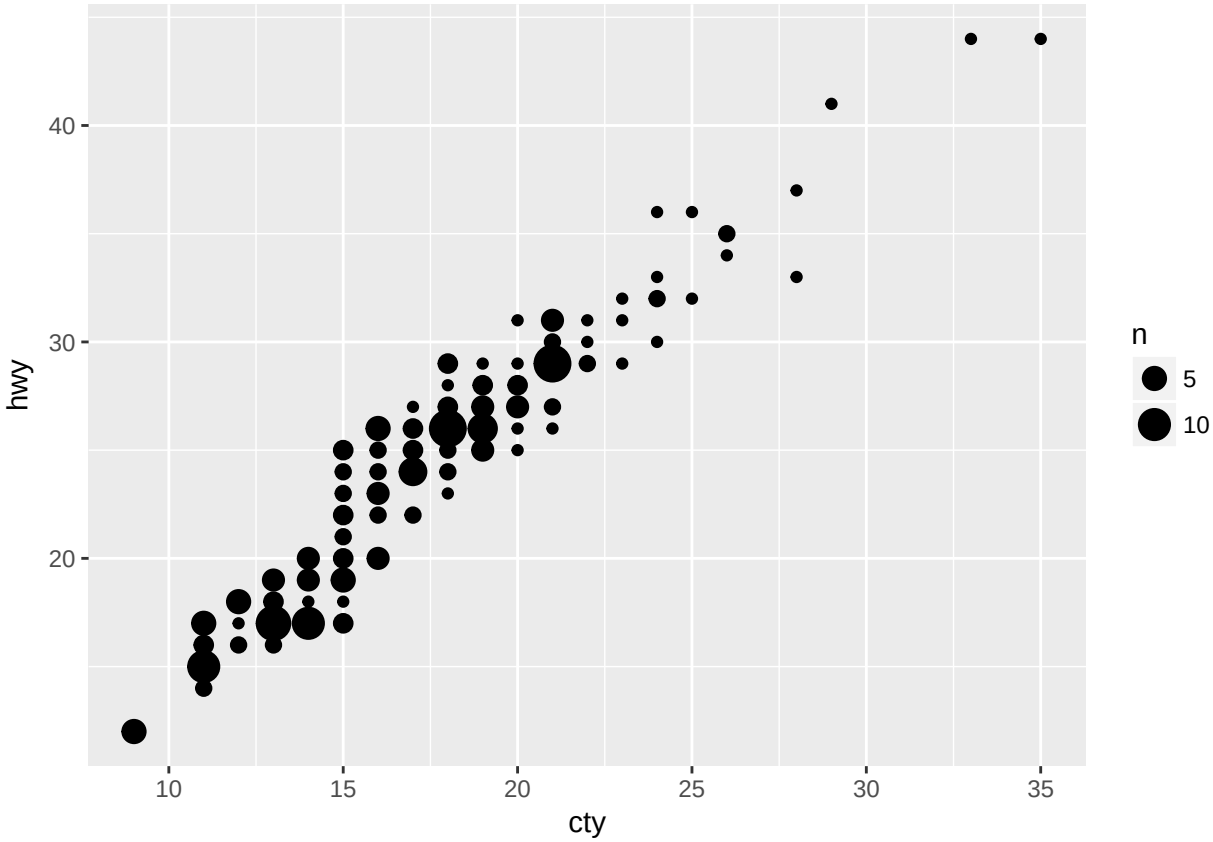
QUESTION 3:

ANSWER:

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



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



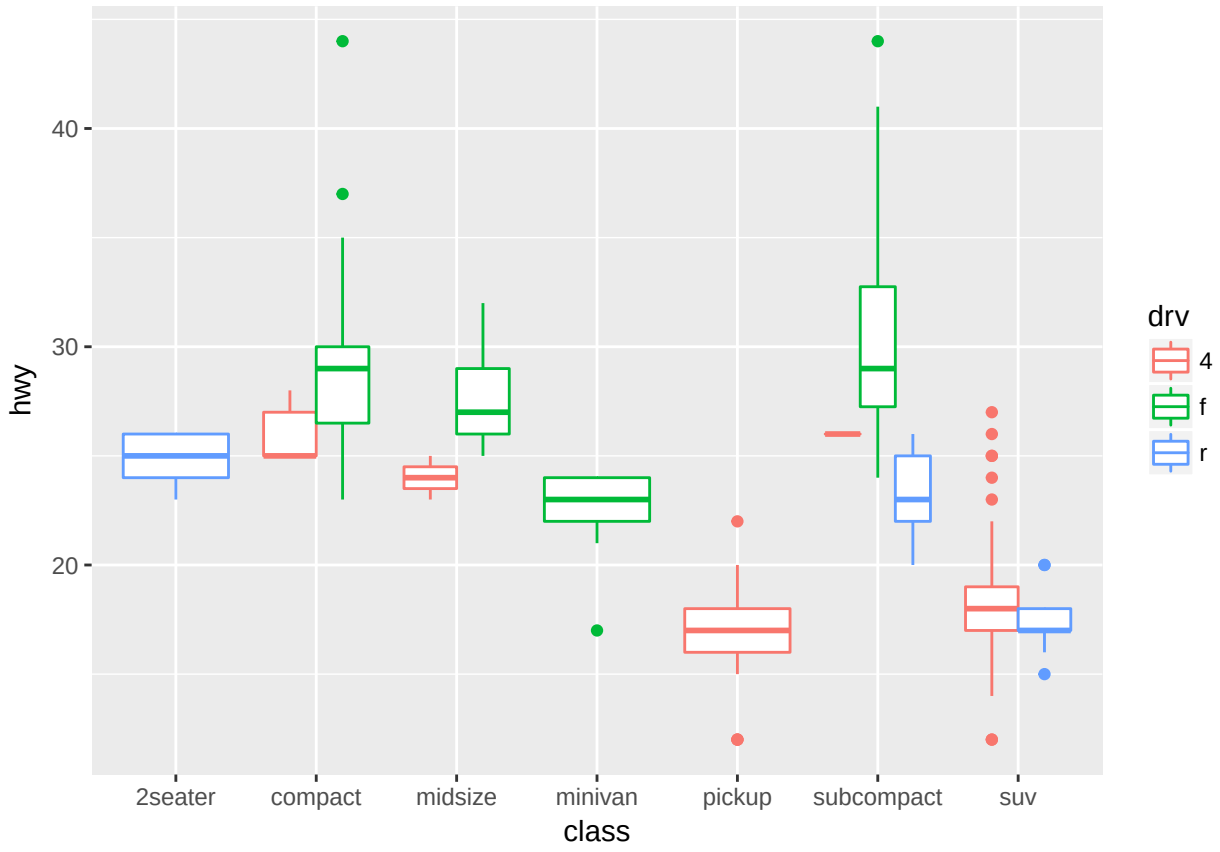
geom_count() is varient of geom_point and it basically it counts the number of data elements or observations at a point in the plot and then maps that count to the pointing area.

QUESTION 4:

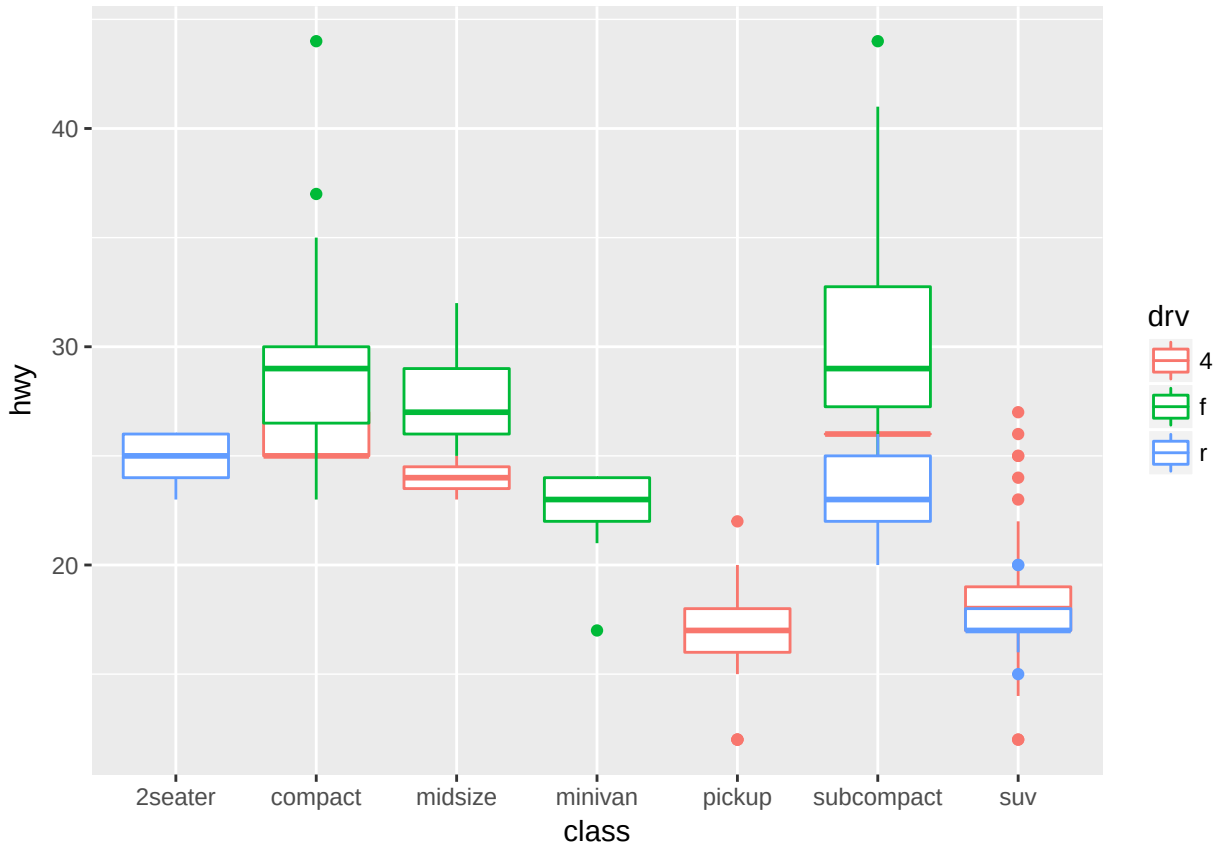
ANSWER:

By observing all the below graphs, we can conclude that **position = "dodge"** is the default position adjustment for a boxplot.

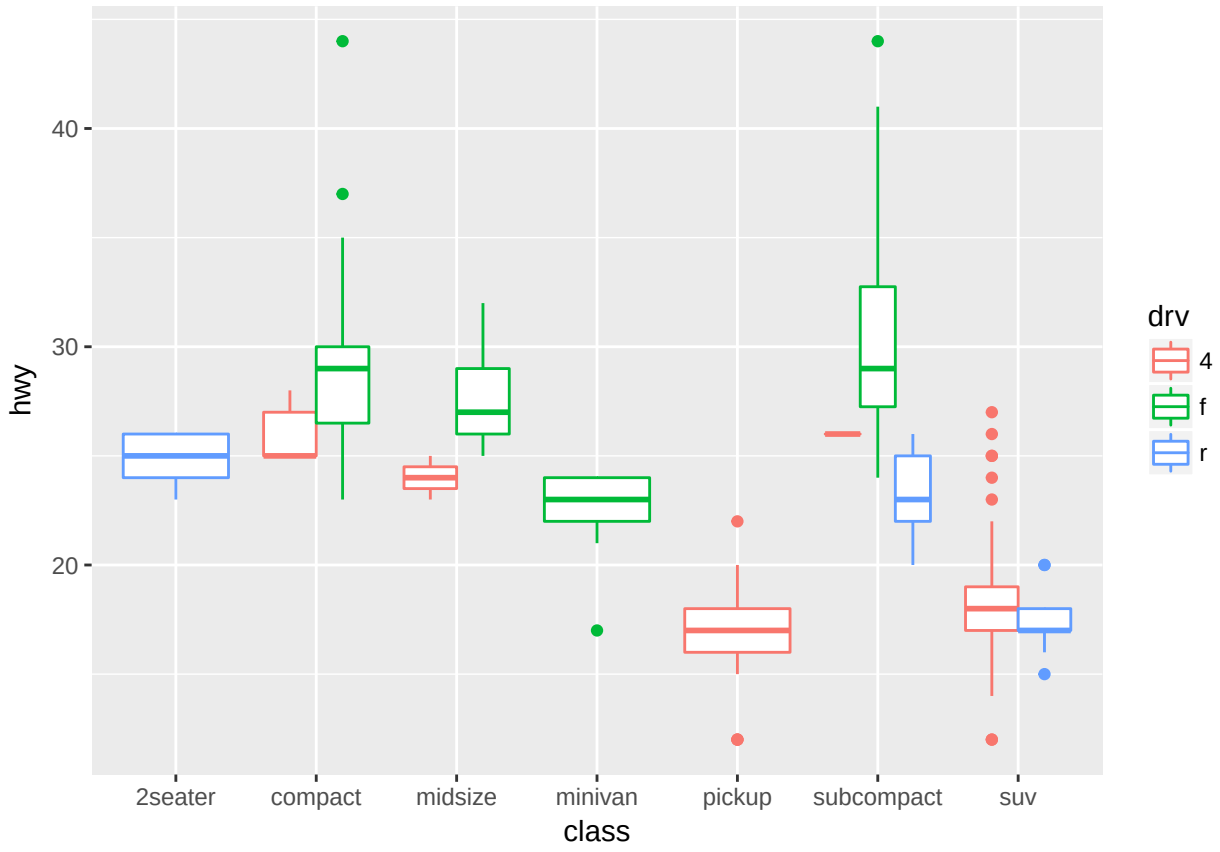
```
ggplot(data = mpg, mapping = aes(x = class, y = hwy, color = drv)) +
  geom_boxplot()
```



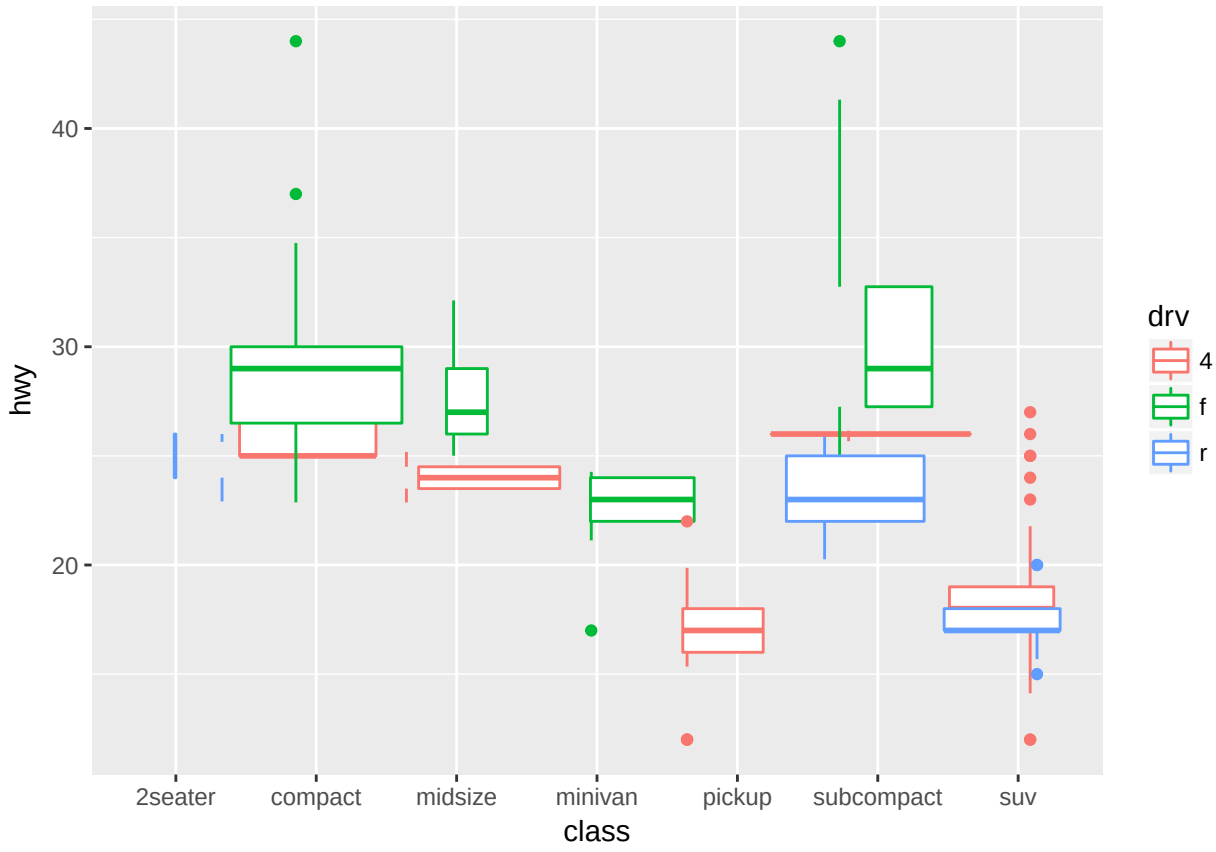

```
# lets try all types of position adjustments

# position = "Identity"
ggplot(data = mpg, mapping = aes(x = class, y = hwy, color = drv)) +
  geom_boxplot(position = "Identity")
```



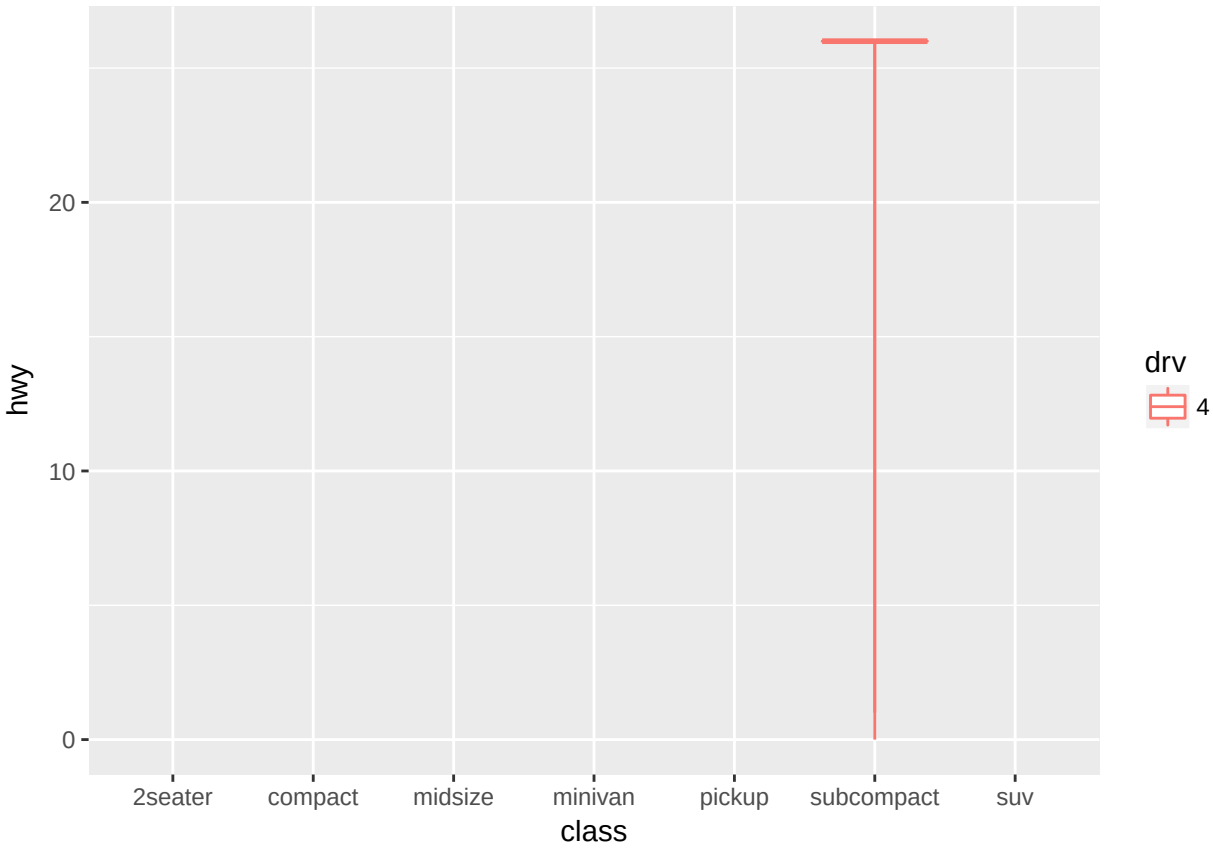

```
# position = "dodge"
ggplot(data = mpg, mapping = aes(x = class, y = hwy, color = drv)) +
  geom_boxplot (position = "dodge")
```




```
# position = "jitter"
ggplot(data = mpg, mapping = aes(x = class, y = hwy, color = drv)) +
  geom_boxplot (position = "jitter")
```




```
# position = "fill"
ggplot(data = mpg, mapping = aes(x = class, y = hwy, color = drv)) +
  geom_boxplot (position = "fill")
```



Section 3.9.1: #2 and #4 only

QUESTION 2:

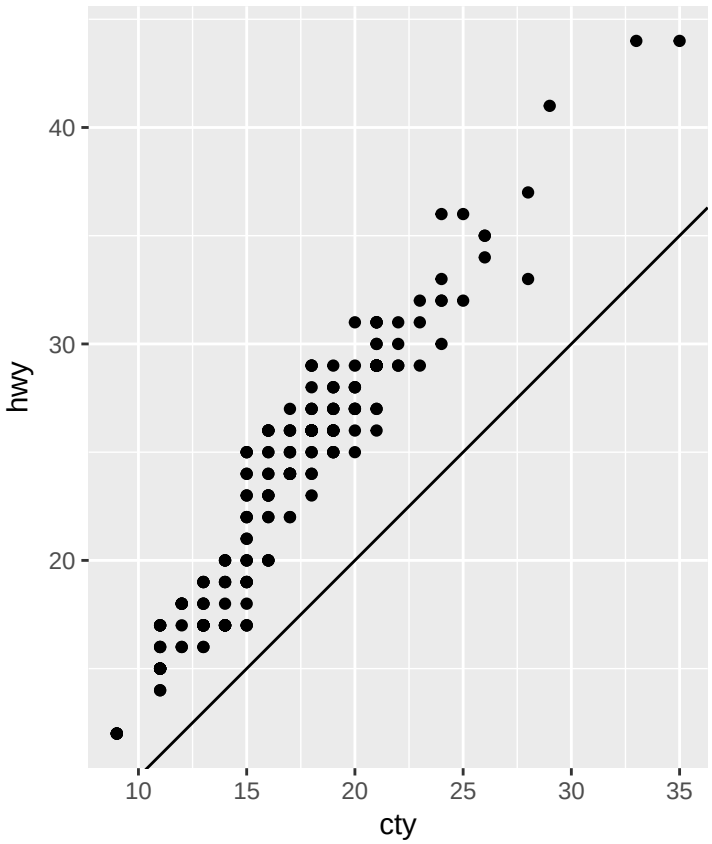
```
?labs ()
```

ANSWER:

labs will changes the lables for the axes. In addition we can use this for titles and substitles aswell.

QUESTION 4:

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



ANSWER:

– From the above graph it is observed that the both the variables are positively related to each other.
– coord_fixed is important because it is making sure that the coordinates for the both variables are fixed.
– geom_abline plots the slope between the variables cty and hwy.

Section 4.4: #1 and #2 only

QUESTION 1:

my_variable <- 10 my_var1able

ANSWER:

There is a type in the second line. It should be : my_variable

QUESTION 2:

ANSWER:

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

filter(mpg, cyl == 8)
filter(diamonds, carat > 3)

Data transformation

Section 5.2.4: #1, #3 and #4 only

```
library(nycflights13)
library(tidyverse)
library(mdsr)
filter(flights, month == 1, day == 1)
```

```
## # A tibble: 842 x 19
##     year month   day dep_time sched_dep_time dep_delay arr_time
##    <int> <int> <int>    <int>          <int>     <dbl>    <int>
##  1  2013     1     1      517            515         2      830
##  2  2013     1     1      533            529         4      850
##  3  2013     1     1      542            540         2      923
##  4  2013     1     1      544            545        -1     1004
##  5  2013     1     1      554            600        -6      812
##  6  2013     1     1      554            558        -4      740
##  7  2013     1     1      555            600        -5      913
##  8  2013     1     1      557            600        -3      709
##  9  2013     1     1      557            600        -3      838
## 10  2013     1     1      558            600        -2      753
## # ... with 832 more rows, and 12 more variables: sched_arr_time <int>,
## #   arr_delay <dbl>, carrier <chr>, flight <int>, tailnum <chr>,
## #   origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>, hour <dbl>,
## #   minute <dbl>, time_hour <dttm>
```