

COMPSCIX 415.2 Homework 3

Santosh Kanutala

6/22/2018

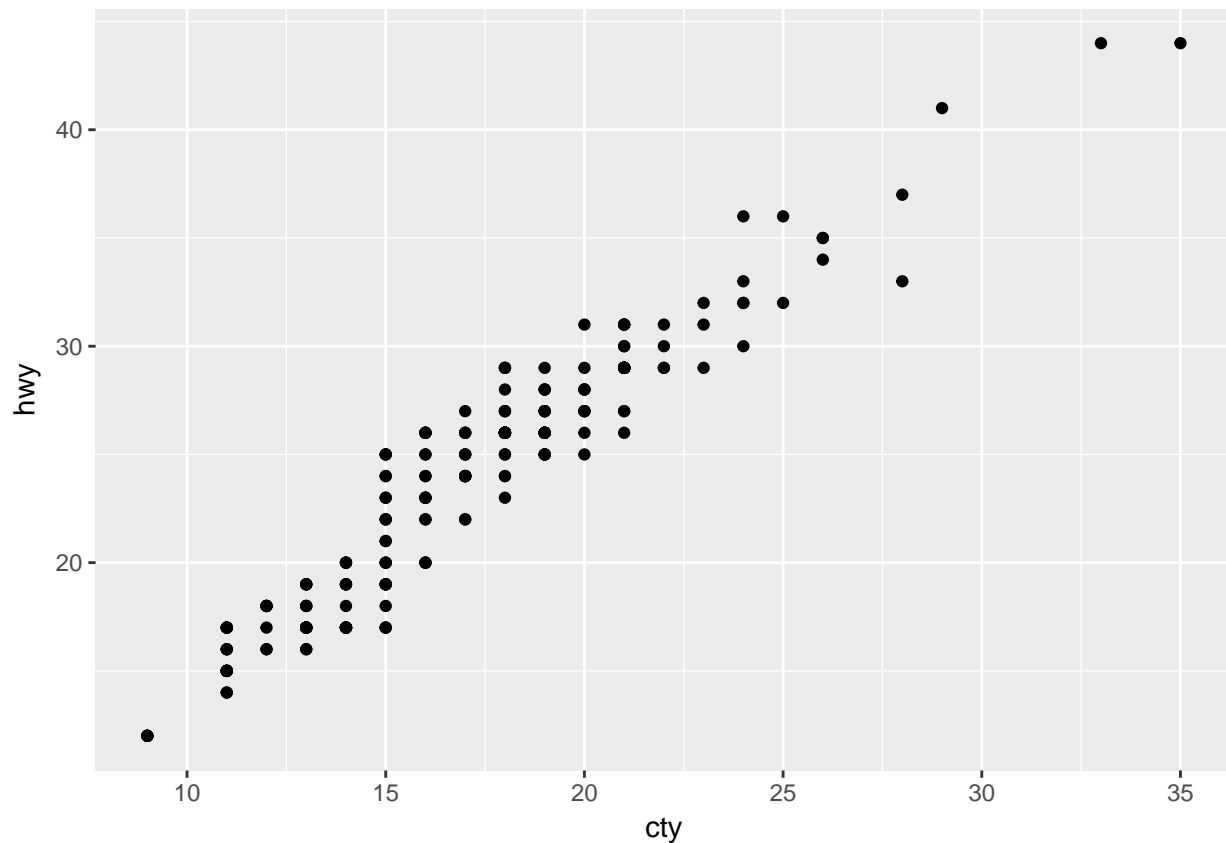
****My Github repository for my assignments can be found at below URL: (<https://github.com/santumagic/compscix-415-2assignments.git>)****

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

Section 3.8.1: all exercises

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.

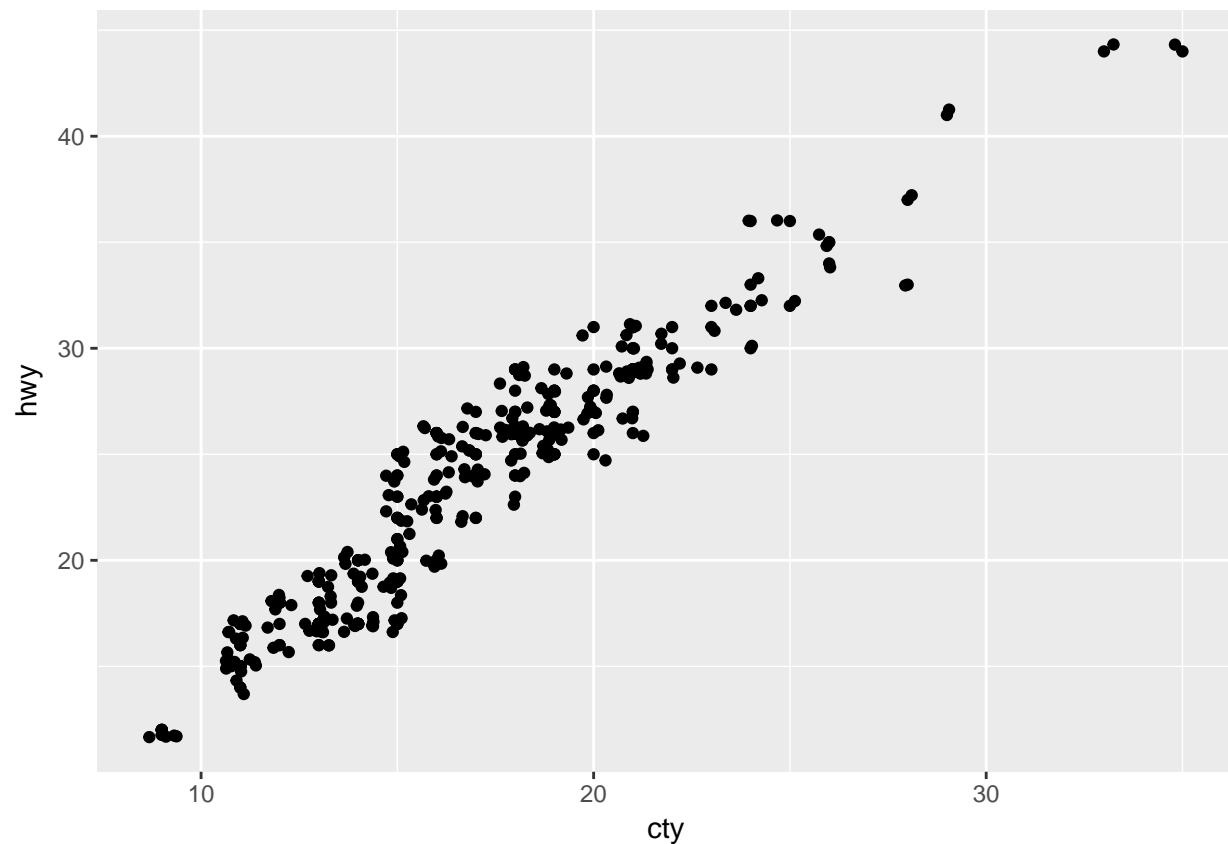
```
?geom_jitter
```

width and height are the parameters that control the jittering.

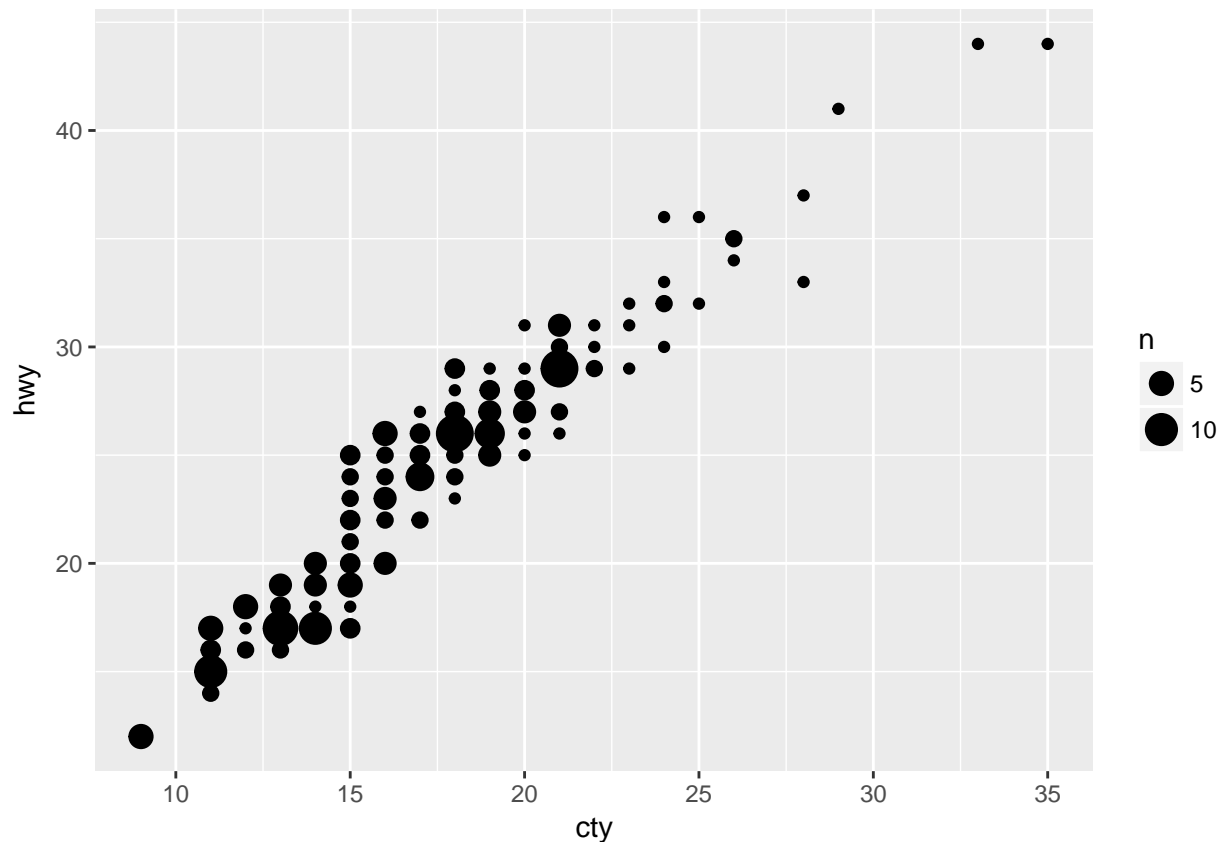
QUESTION 3:

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()
```



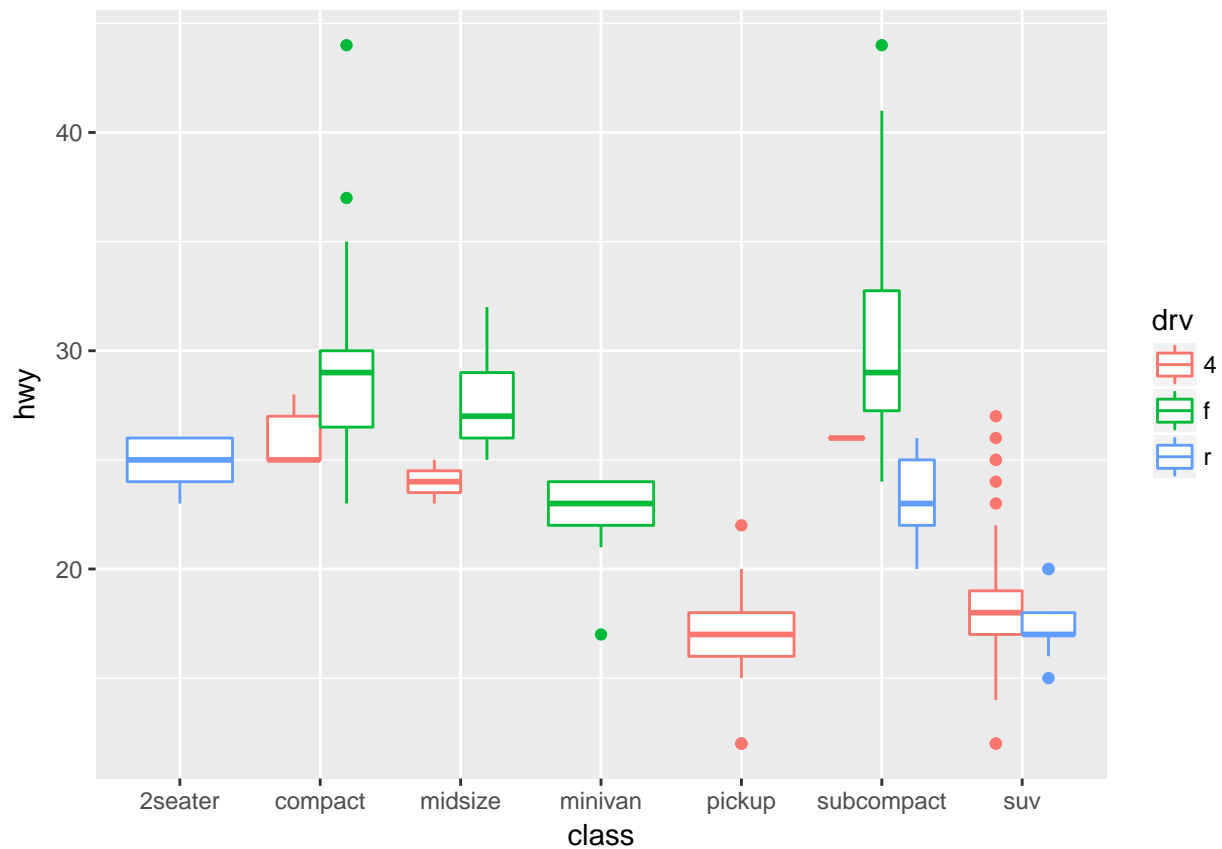
`geom_count()` is variant of `geom_point` and it basically it counts the number of data elements or observations at a point in the plot and then maps that count to the pointing area.

QUESTION 4:

ANSWER:

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

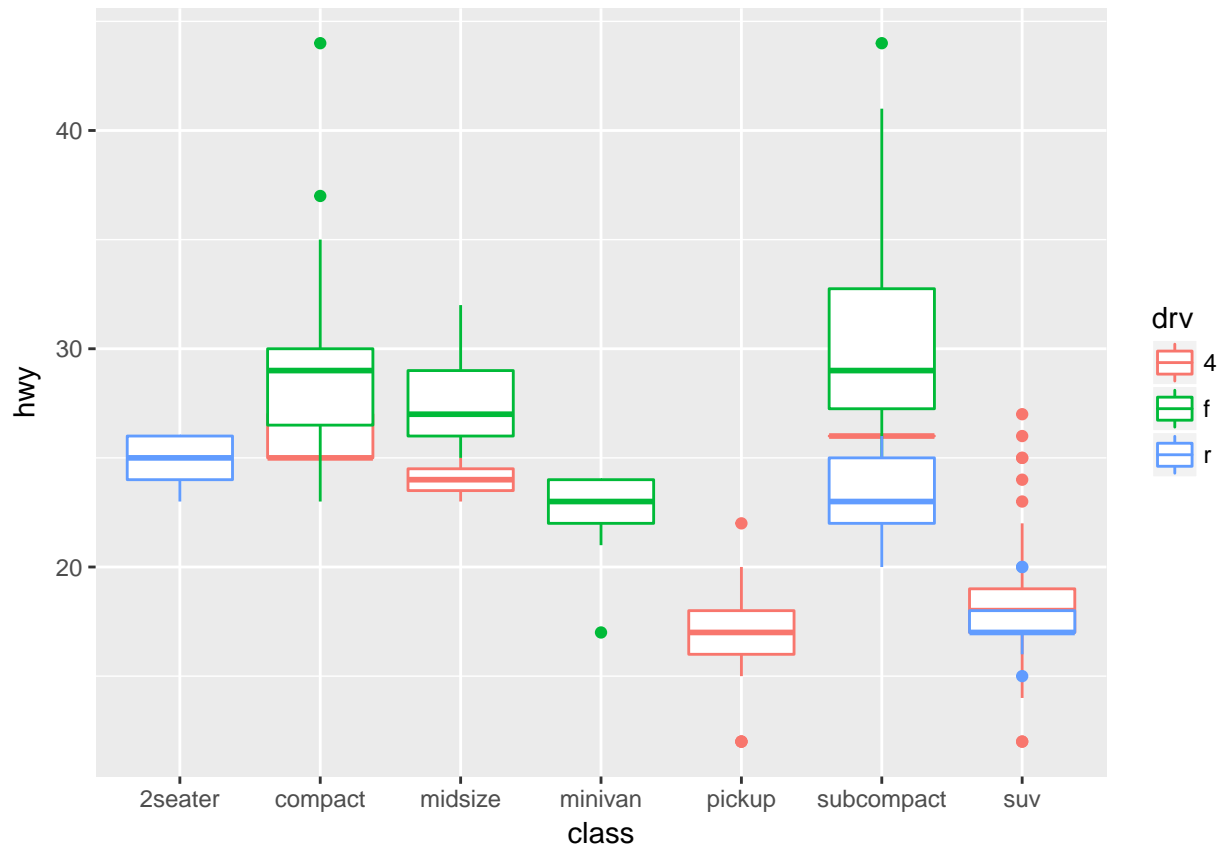
```
ggplot(data = mpg, mapping = aes(x = class, y = hwy, color = drv)) +
  geom_boxplot()
```



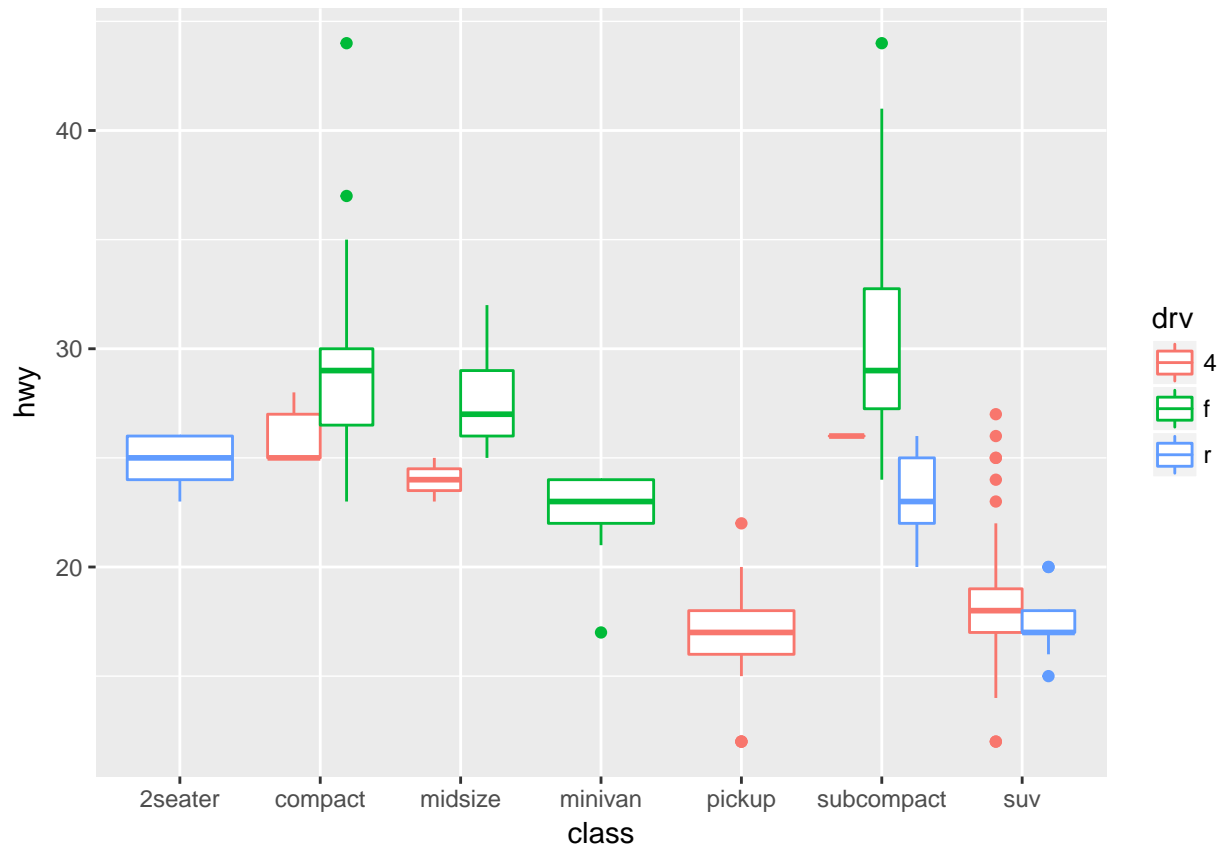
```
# lets try all types of position adjustments
```

```
# position = "Identity"
```

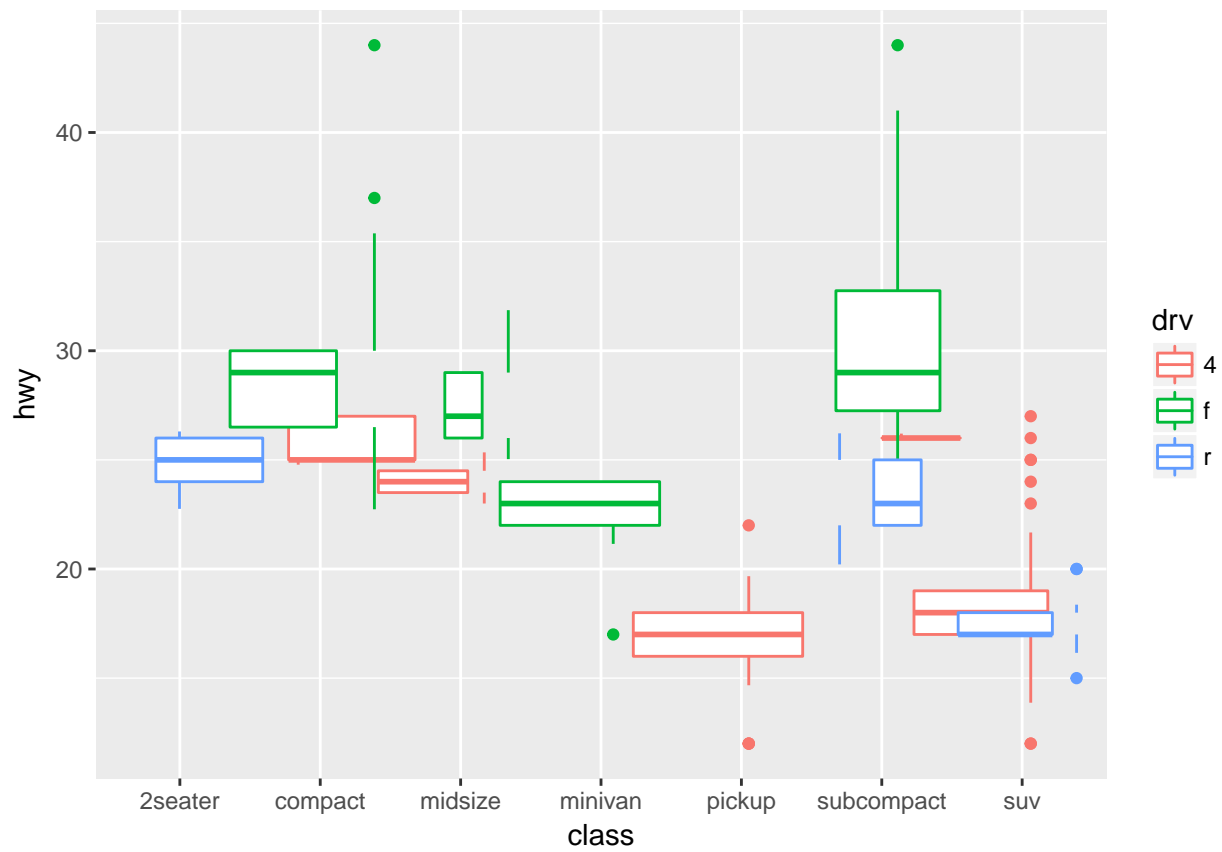
```
ggplot(data = mpg, mapping = aes(x = class, y = hwy, color = drv)) +  
  geom_boxplot(position = "Identity")
```



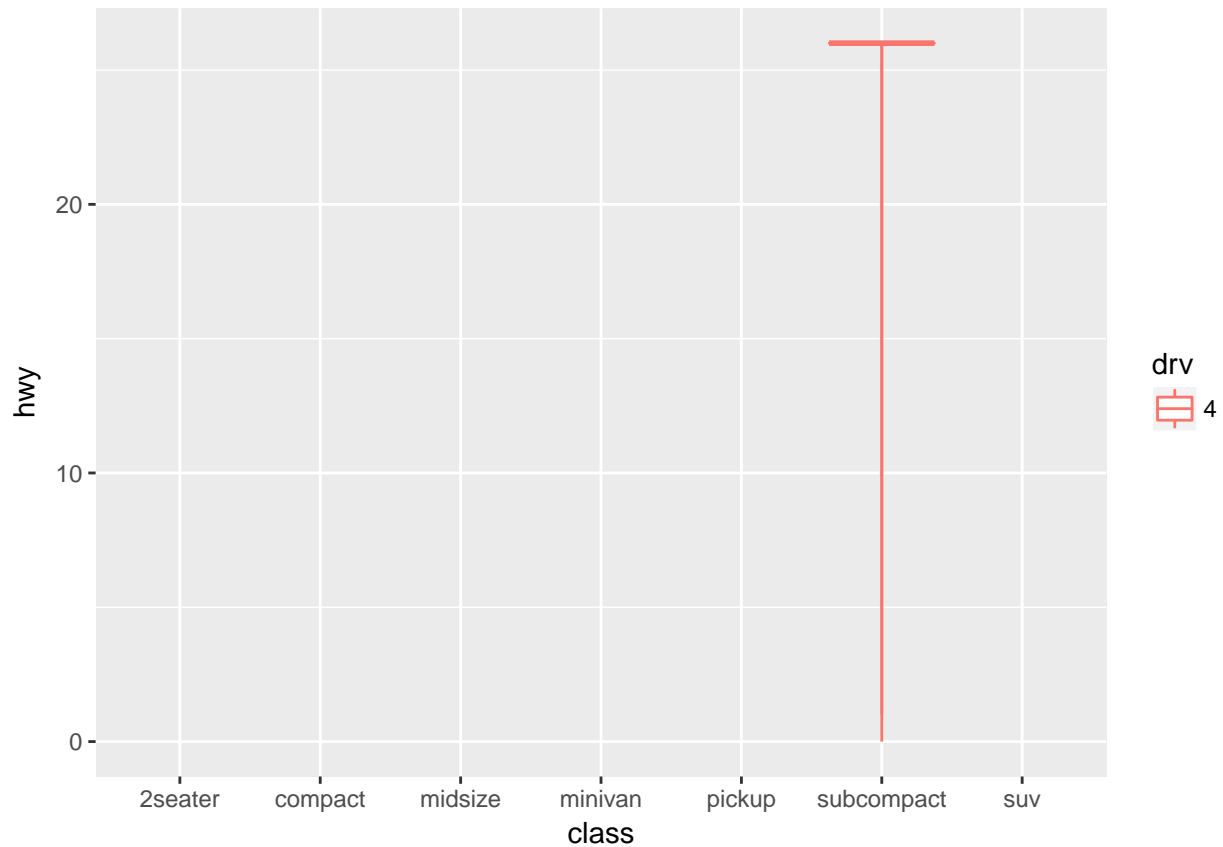
```
# position = "dodge"
ggplot(data = mpg, mapping = aes(x = class, y = hwy, color = drv)) +
  geom_boxplot(position = "dodge")
```



```
# 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:

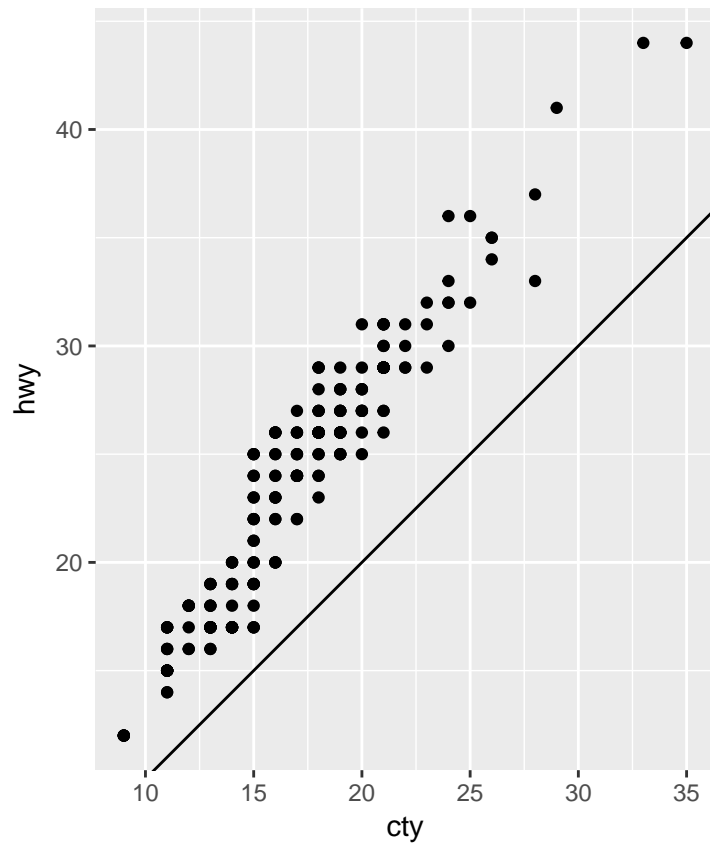
```
?labs ()
```

ANSWER:

labs will change the labels for the axes.

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.