

Lab 01: Data visualization

Due: Tu, Sep 28 at 11:59pm

Mehreen Junaid

Packages

```
library(tidyverse)
```

Exercise 1

1. How many rows and how many columns does mpg have? What are the variables included in the data frame?

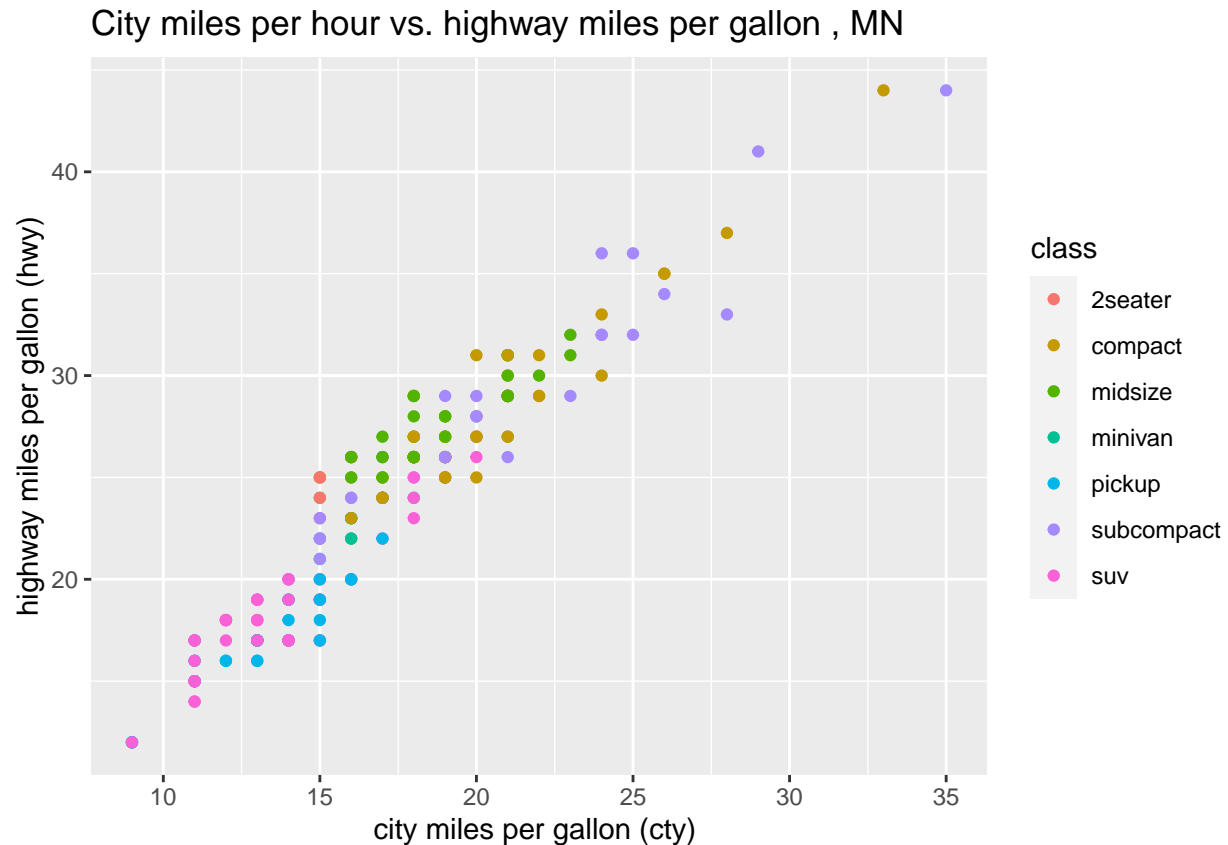
```
glimpse(mpg)
```

```
## Rows: 234
## Columns: 11
## $ manufacturer <chr> "audi", "audi", "audi", "audi", "audi", "audi", "audi", "~
## $ model <chr> "a4", "a4", "a4", "a4", "a4", "a4", "a4", "a4 quattro", "~
## $ displ <dbl> 1.8, 1.8, 2.0, 2.0, 2.8, 2.8, 3.1, 1.8, 1.8, 2.0, 2.0, 2.~
## $ year <int> 1999, 1999, 2008, 2008, 1999, 1999, 2008, 1999, 1999, 200~
## $ cyl <int> 4, 4, 4, 4, 6, 6, 6, 4, 4, 4, 4, 6, 6, 6, 6, 6, 8, 8, ~
## $ trans <chr> "auto(l5)", "manual(m5)", "manual(m6)", "auto(av)", "auto~
## $ drv <chr> "f", "f", "f", "f", "f", "f", "f", "4", "4", "4", "4", "4~
## $ cty <int> 18, 21, 20, 21, 16, 18, 18, 18, 16, 20, 19, 15, 17, 17, 1~
## $ hwy <int> 29, 29, 31, 30, 26, 26, 27, 26, 25, 28, 27, 25, 25, 25, 2~
## $ fl <chr> "p", "p", "p", "p", "p", "p", "p", "p", "p", "p", "p", "p~
## $ class <chr> "compact", "compact", "compact", "compact", "compact", "c~
```

Exercise 2

2. Generate a scatterplot of city miles per gallon (cty) versus highway miles per gallon (hwy) with points colored by class.

```
ggplot(data= mpg, mapping=aes(x=cty, y=hwy, color=class))+geom_point() +
labs(title="City miles per hour vs. highway miles per gallon , MN",
      x="city miles per gallon (cty) ", y="highway miles per gallon (hwy)")
```



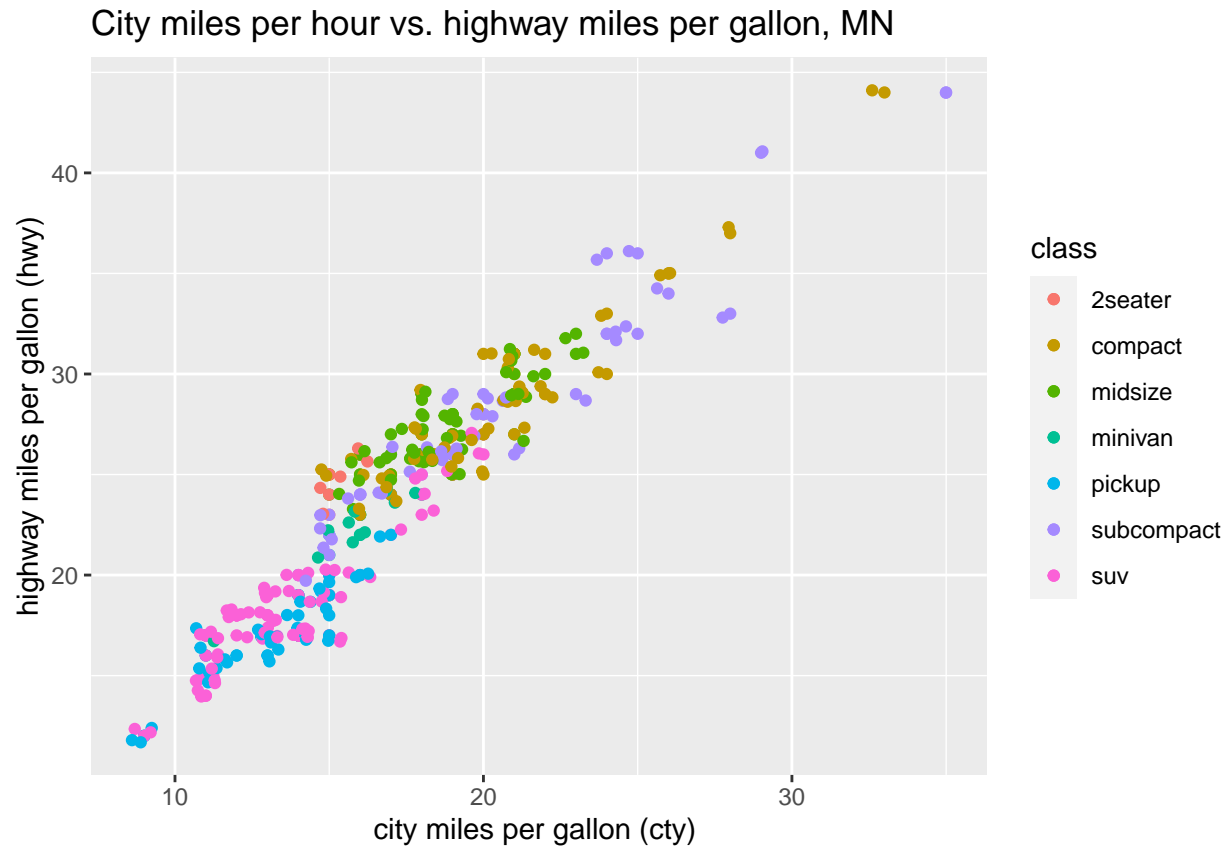
The plot produced by the code snippet above gives a visual representation of the relationship between city miles per gallon (cty) and highway miles per gallon (hwy). The advantage of this plot is that it allows for an easy comparison between different classes of cars [1]. For example, it can be seen at a glance which cars have the best mileage for city and highway driving. Additionally, the color-coding of the points allows for the differences between classes to be easily seen.

The disadvantage of this plot is that it does not provide any information about the distribution of the data. It can be difficult to interpret the exact relationship between the two variables without knowing the exact range and variation of the data points. Additionally, this plot does not provide any information about the trends or patterns in the data. It also does not provide any information about the underlying relationships between the two variables.

Exercise 3

- Note that there are only so many possibilities of highway and city miles per gallon, so some of the points are on top of each other. Using `geom_jitter()` or a `position =` argument in `geom_point()`, add a small amount of random variation to each point. Briefly comment on the differences between the plots you constructed in Exercises 2 and 3. What are the advantages and disadvantages of each?

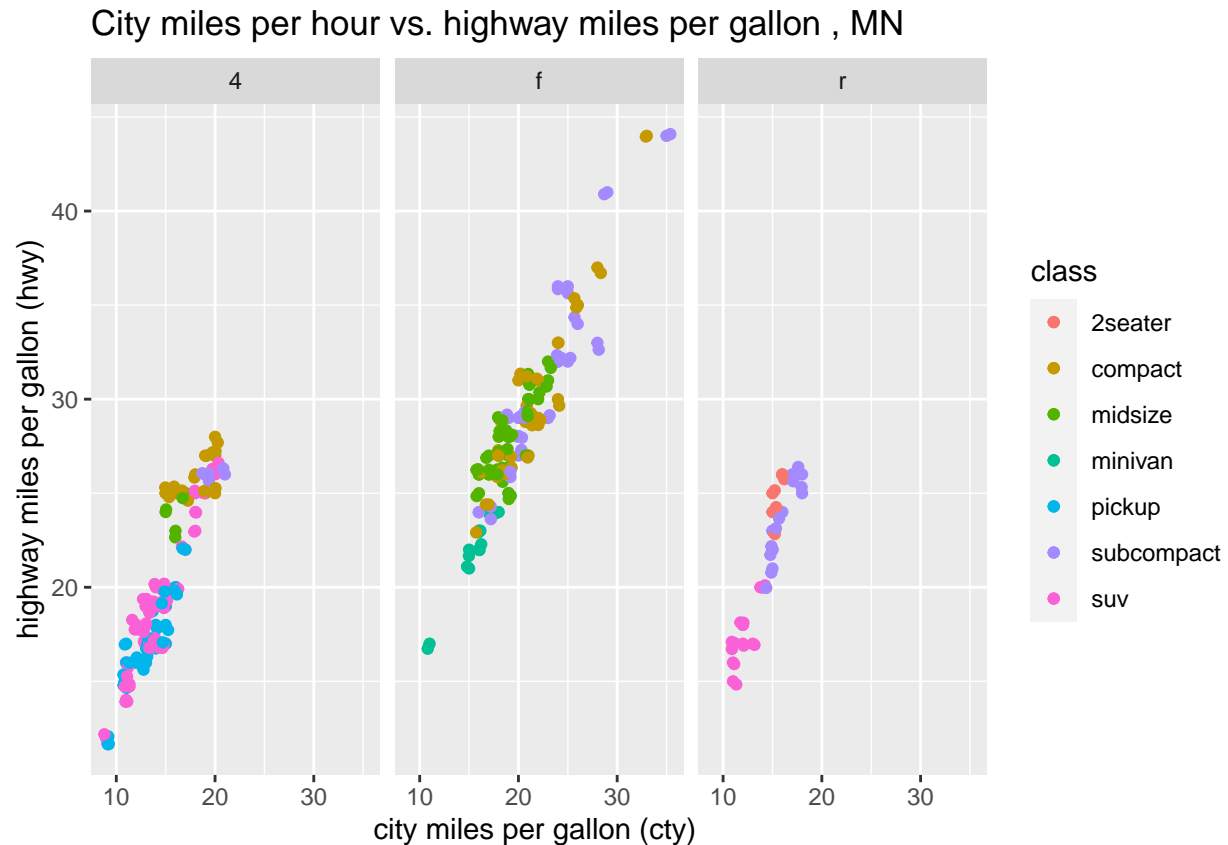
```
ggplot(data= mpg, mapping=aes(x=cty,y=hwy,color=class))+geom_point()+
geom_jitter()+labs(title="City miles per hour vs. highway miles per gallon, MN",
x="city miles per gallon (cty) ", y="highway miles per gallon (hwy)")
```



Exercise 4

4. Examine the relationship between city and highway miles per gallon, with a separate plot for each type of drive train (drv).

```
ggplot(data= mpg, mapping=aes(x=cty,y=hwy,color=class))+geom_point()+  
geom_jitter()+labs(title="City miles per hour vs. highway miles per gallon , MN",  
x="city miles per gallon (cty) ", y="highway miles per gallon (hwy)") +facet_wrap(~drv)
```



Comment here on what you notice about the relationship between class and city miles per gallon:

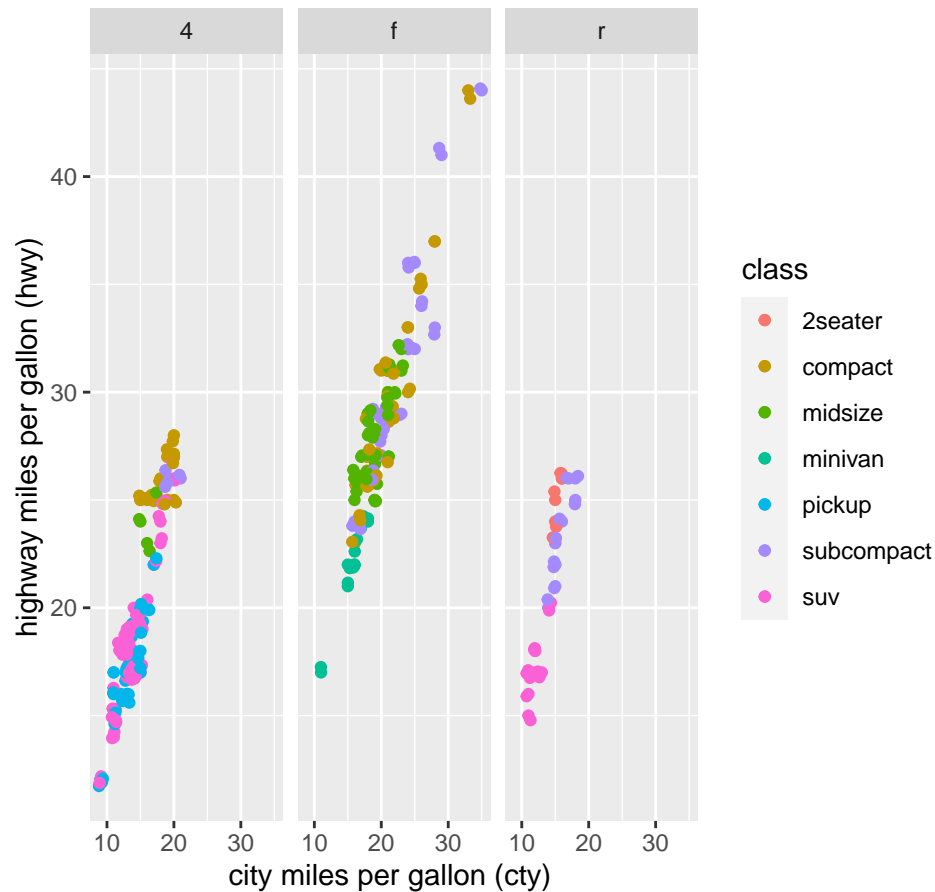
From the plot generated by the code snippet above, it is clear that the class of a car has an impact on the city miles per gallon (cty). In general, larger cars (such as SUVs, pickups, and vans) have lower city miles per gallon than smaller cars. Additionally, the relationship between class and city miles per gallon appears to be non-linear, with SUVs having the lowest city miles per gallon and compact/2seaters having the highest.

Exercise 5

5. Use R chunk options to customize the figure output dimensions as specified below. We can use `fig.height` and `fig.width` as chunk options to adjust the height and width of figures.

```
ggplot(data= mpg, mapping=aes(x=cty,y=hwy,color=class))+geom_point()+
geom_jitter()+labs(title="City miles per hour vs. highway miles per gallon , MN",
x="city miles per gallon (cty) ", y="highway miles per gallon (hwy)") + facet_wrap(~drv)
```

City miles per hour vs. highway miles per gallon , MN



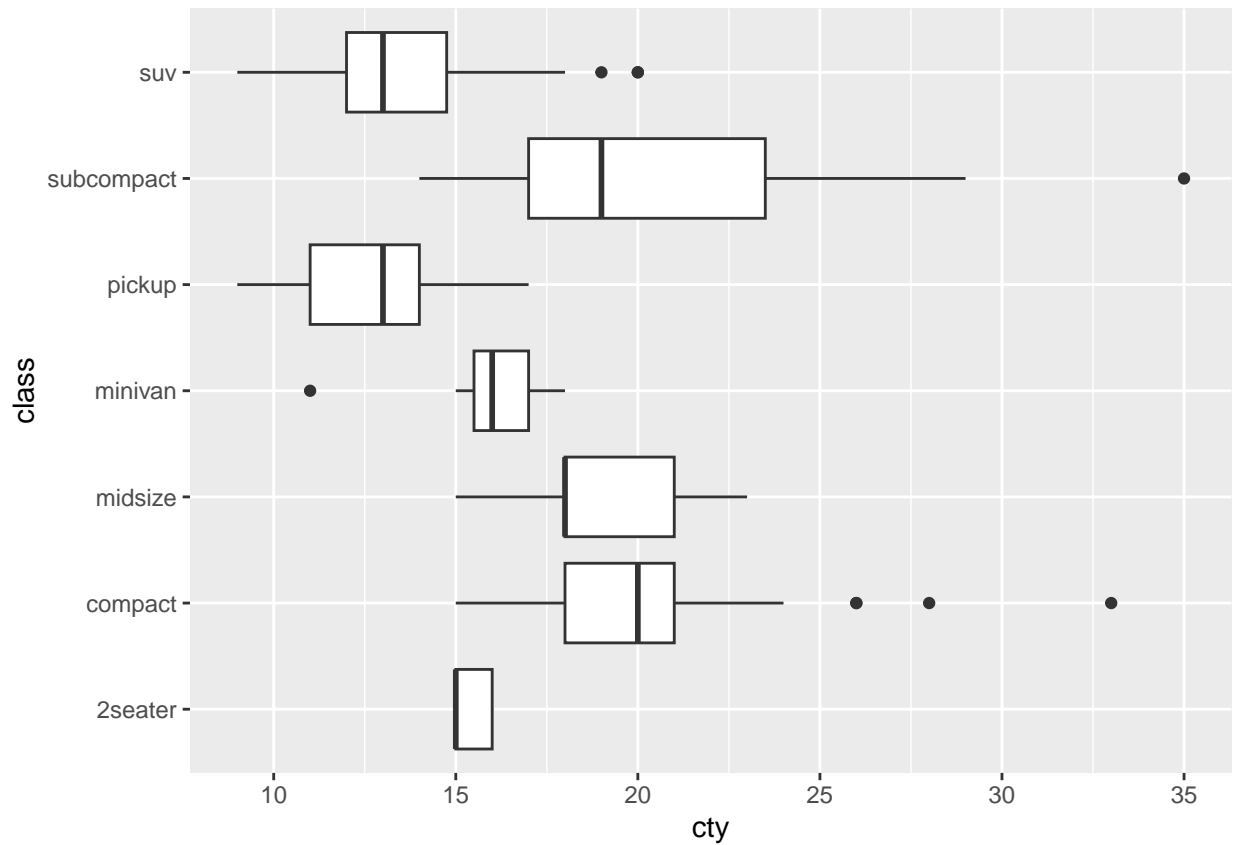
What do you notice from the bar chart above?

The bar chart above shows the relationship between city miles per gallon (cty) and highway miles per gallon (hwy) for different classes of cars. It is evident that larger cars generally have lower city miles per gallon than smaller cars. Additionally, the relationship between class and city miles per gallon appears to be non-linear. Additionally, the chart also shows that there is significant variation in the city miles per gallon of cars within the same class.

Exercise 6

6. Create side-by-side boxplots of city miles per gallon for each class. Briefly comment on what you notice.

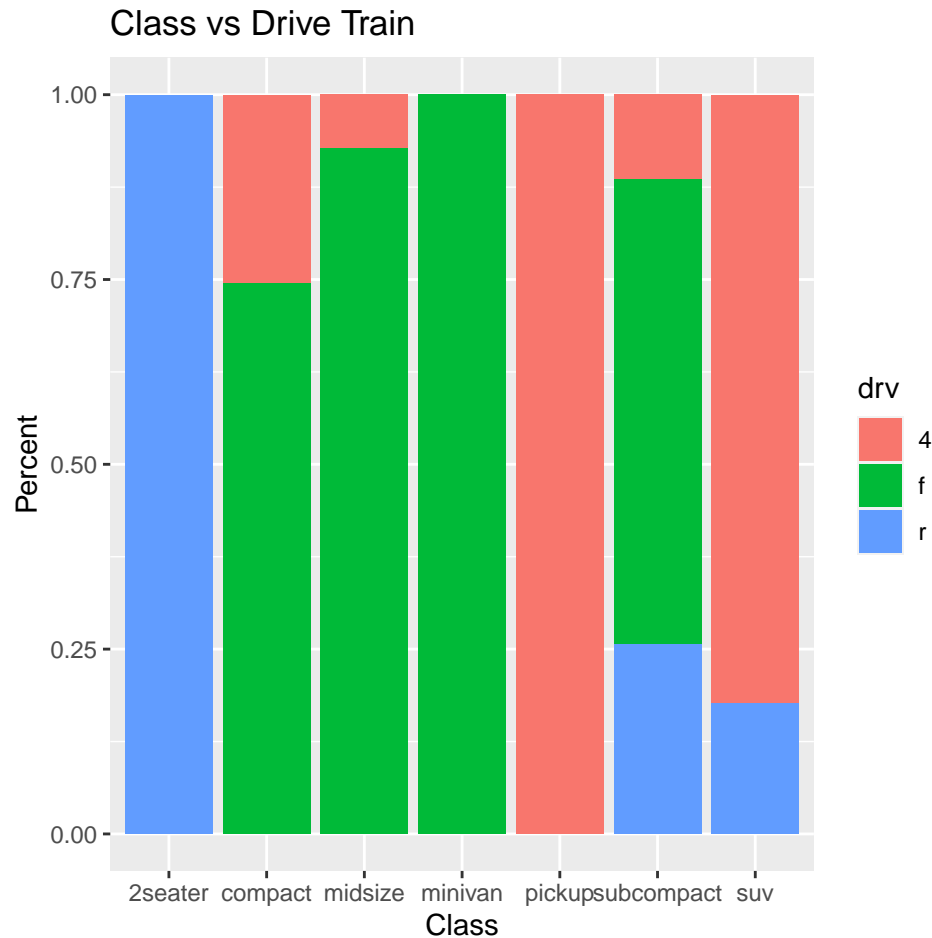
```
ggplot(data = mpg, aes(x = cty, y = class)) + geom_boxplot()
```



Exercise 7

7. Create a segmented bar chart with one bar per class, each bar going from 0 - 1, with the fill determined by the type of drive train (drv). What do you notice?

```
ggplot(data=mpg, aes(x=class, y=1, fill=drv)) +
  geom_col(position="fill") +
  labs(title="Class vs Drive Train",
       x="Class",
       y="Percent")
```



The code above will produce a segmented bar chart with one bar per class, each bar going from 0-1, with the fill determined by the type of drive train (drv). The y-axis will range from 0-1, and the title and x-axis labels will be set to “Class vs Drive Train”, “Class”, and “Percent”, respectively.

Exercise 8

- Recreate the plot below. The functions `theme_bw()` and `labs()` will be helpful. The size of the points is 0.50. Also, set the figure dimensions with appropriate code chunk options, where the width is 9 and the height is 6. Note: the figure dimensions below are distorted due to the formatting of this word file.

```
ggplot(data = mpg, mapping = aes (x=displ, y=hwy )) +
  geom_point(size=0.5) + labs(title = "Highway mpg vs. displacement",
  x = "displacement(litrs)", y = "Highway (mpg)") + theme_bw() +
  facet_wrap(~ class, nrow = 2)
```

Highway mpg vs. displacement

