

# ggplot2 : Introductory course

Aurélien Béliard & Gabriel Pires

11/12/2020

## Loading packages

```
library(tidyverse)
```

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

```
## v ggplot2 3.3.2    v purrr   0.3.4
## v tibble  3.0.4    v dplyr   1.0.2
## v tidyr   1.1.2    v stringr 1.4.0
## v readr   1.4.0    v forcats 0.5.0
```

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

```
library(ggplot2)
```

## Visualizing dataset

```
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
```

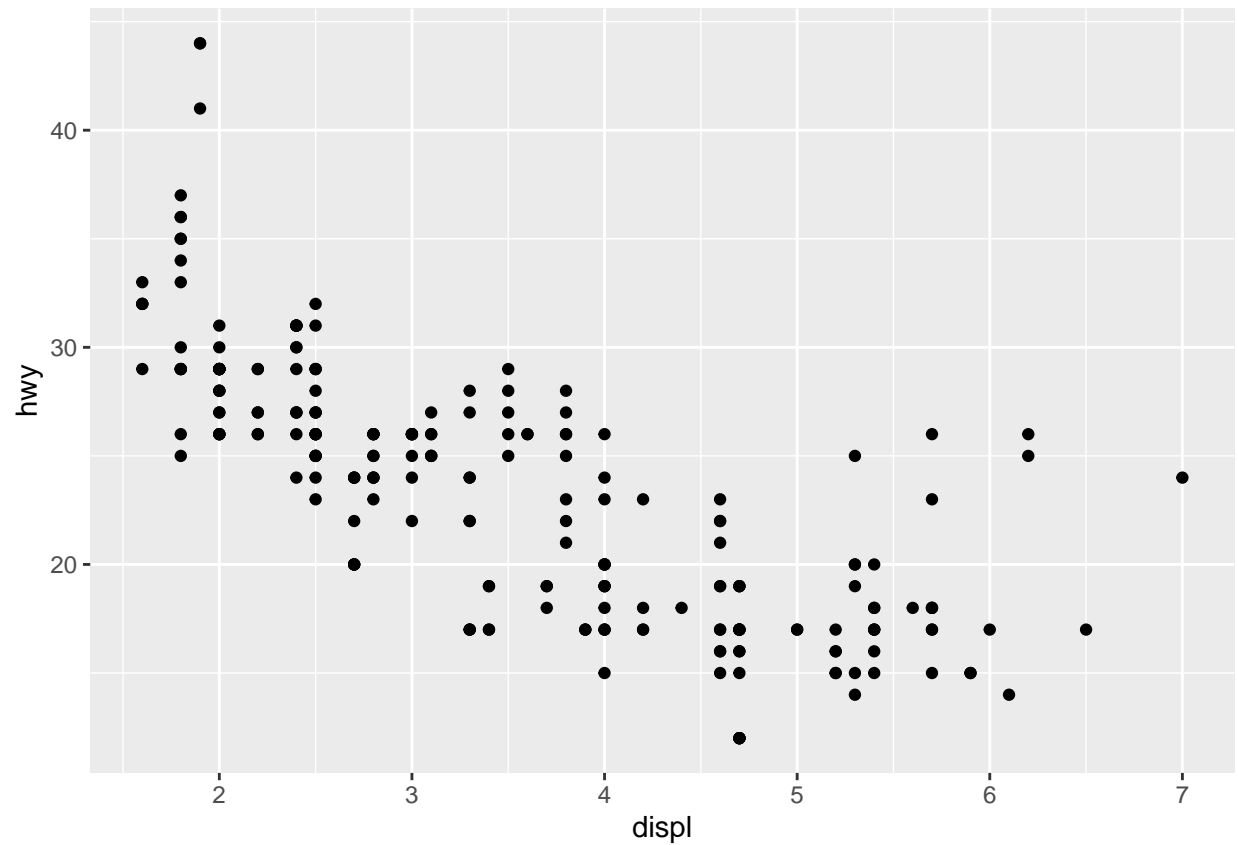
```
summary(mpg)
```

```
## manufacturer      model      displ      year
## Length:234      Length:234      Min.   :1.600      Min.   :1999
## Class :character  Class :character  1st Qu.:2.400      1st Qu.:1999
## Mode  :character  Mode  :character  Median :3.300      Median :2004
##                                     Mean   :3.472      Mean   :2004
##                                     3rd Qu.:4.600      3rd Qu.:2008
##                                     Max.   :7.000      Max.   :2008
##      cyl      trans      drv      cty
## Min.   :4.000      Length:234      Length:234      Min.   : 9.00
## 1st Qu.:4.000      Class :character  Class :character  1st Qu.:14.00
## Median :6.000      Mode  :character  Mode  :character  Median :17.00
## Mean   :5.889                                     Mean   :16.86
## 3rd Qu.:8.000                                     3rd Qu.:19.00
## Max.   :8.000                                     Max.   :35.00
##      hwy      fl      class
## Min.   :12.00      Length:234      Length:234
## 1st Qu.:18.00      Class :character  Class :character
## Median :24.00      Mode  :character  Mode  :character
## Mean   :23.44
## 3rd Qu.:27.00
## Max.   :44.00
```

## First plot

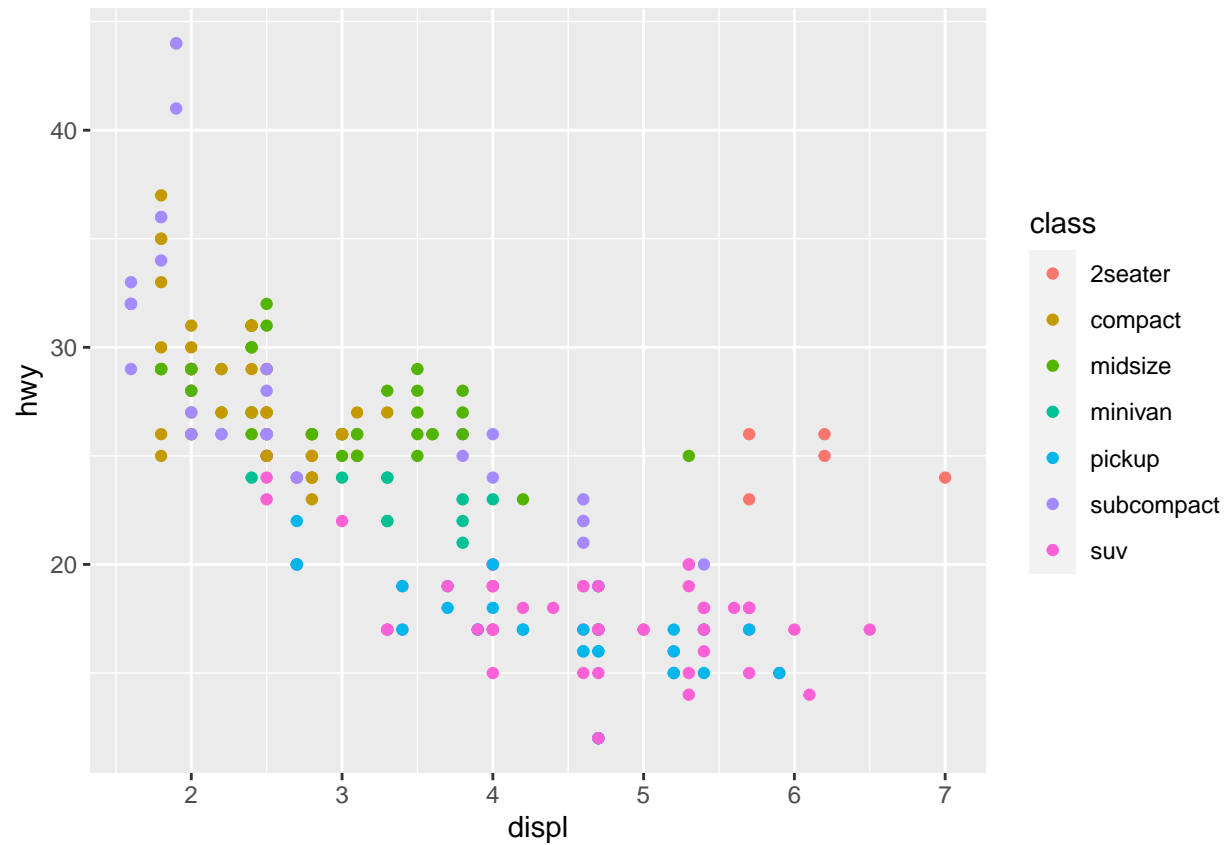
You can create a dot plot using the following function :

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



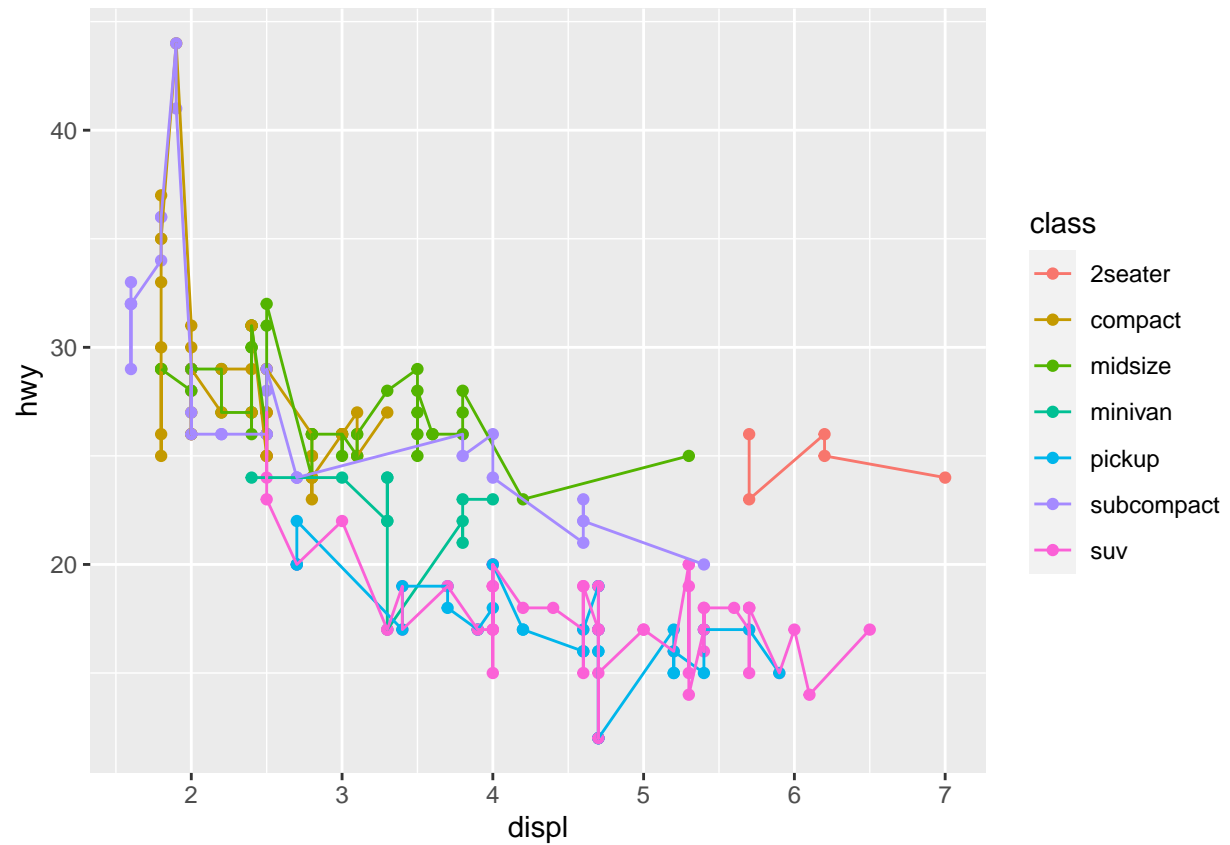
Pretty basic, now let's add some colors :

```
ggplot(mpg)+  
  geom_point(aes(x=displ, y=hwy, color=class))
```



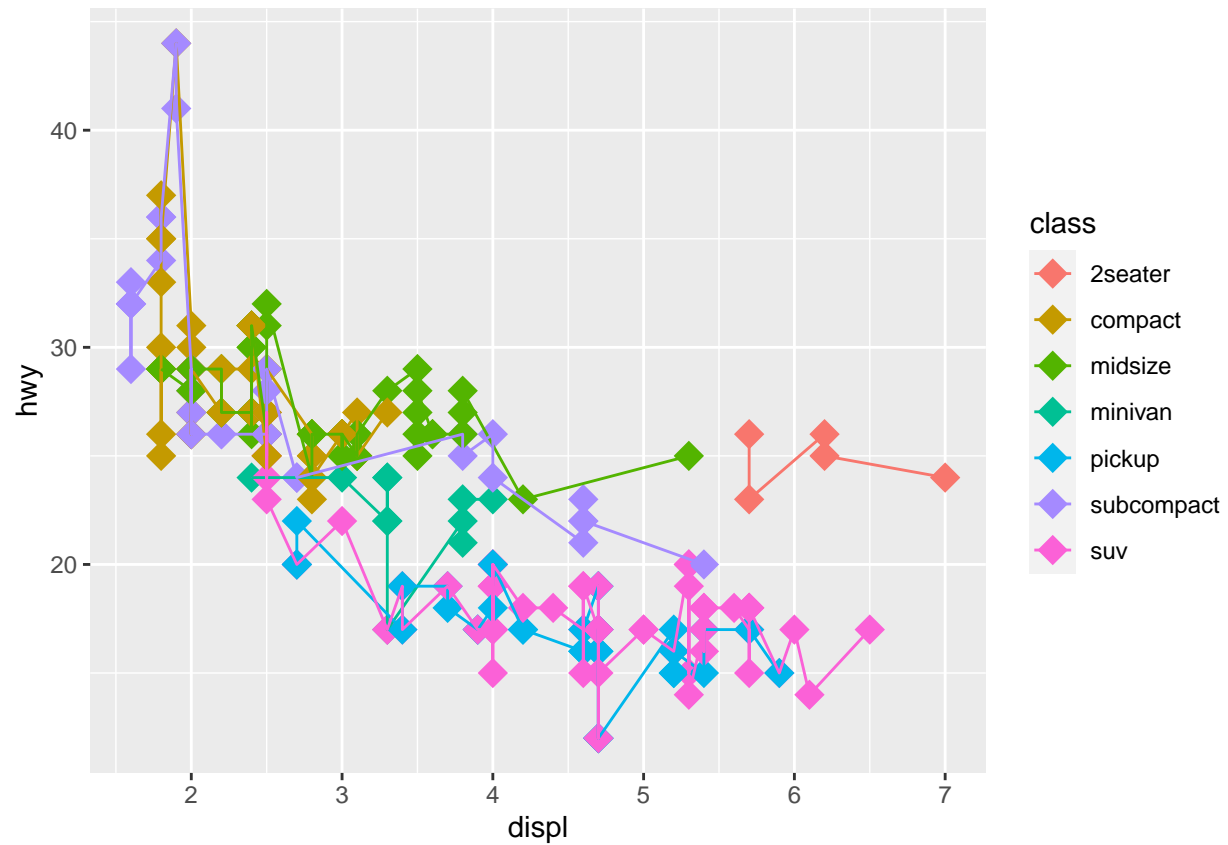
We can also add another layer as :

```
ggplot(mpg)+  
  geom_point(aes(x=displ, y=hwy, color=class))+  
  geom_line(aes(x=displ, y=hwy, color=class))
```



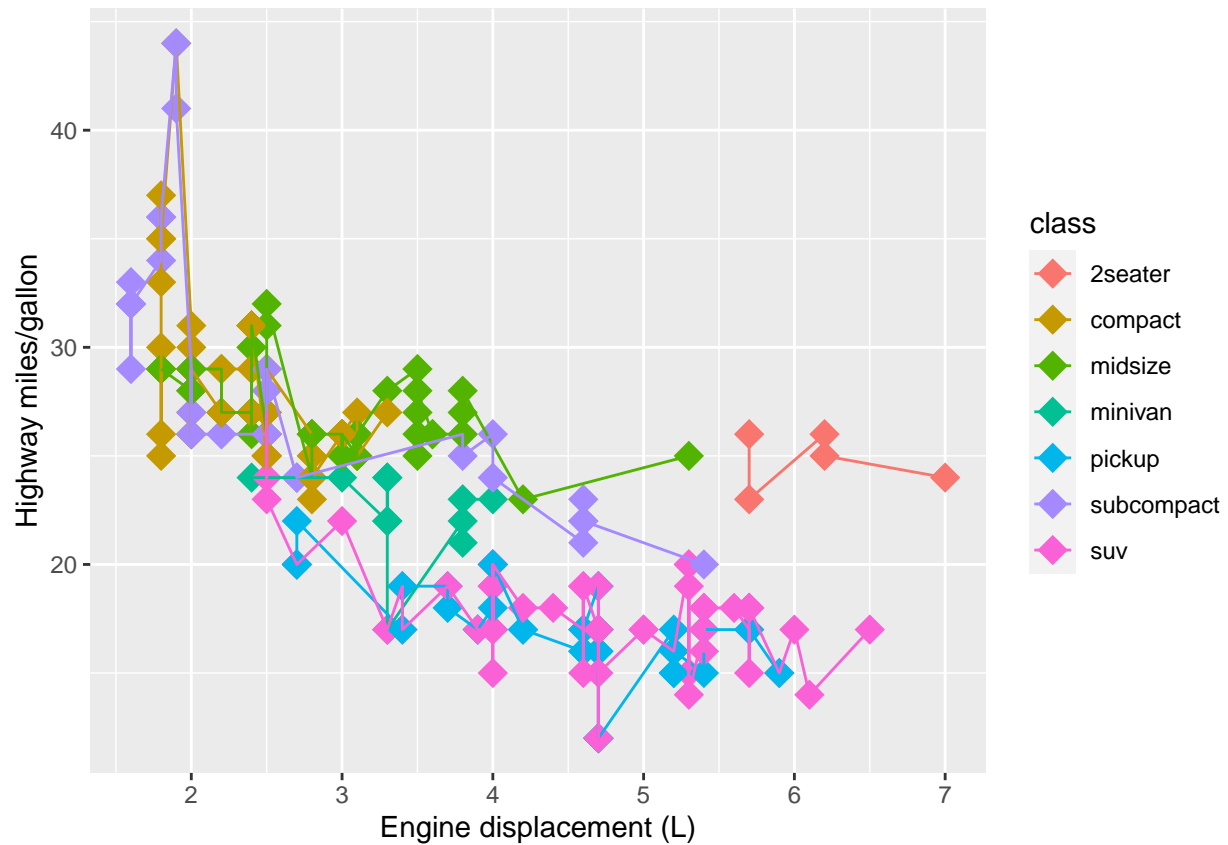
Let's now play with the shape and size !

```
ggplot(mpg)+
  geom_point(aes(x=displ, y=hwy, color=class), shape="diamond", size=5)+
  geom_line(aes(x=displ, y=hwy, color=class))
```



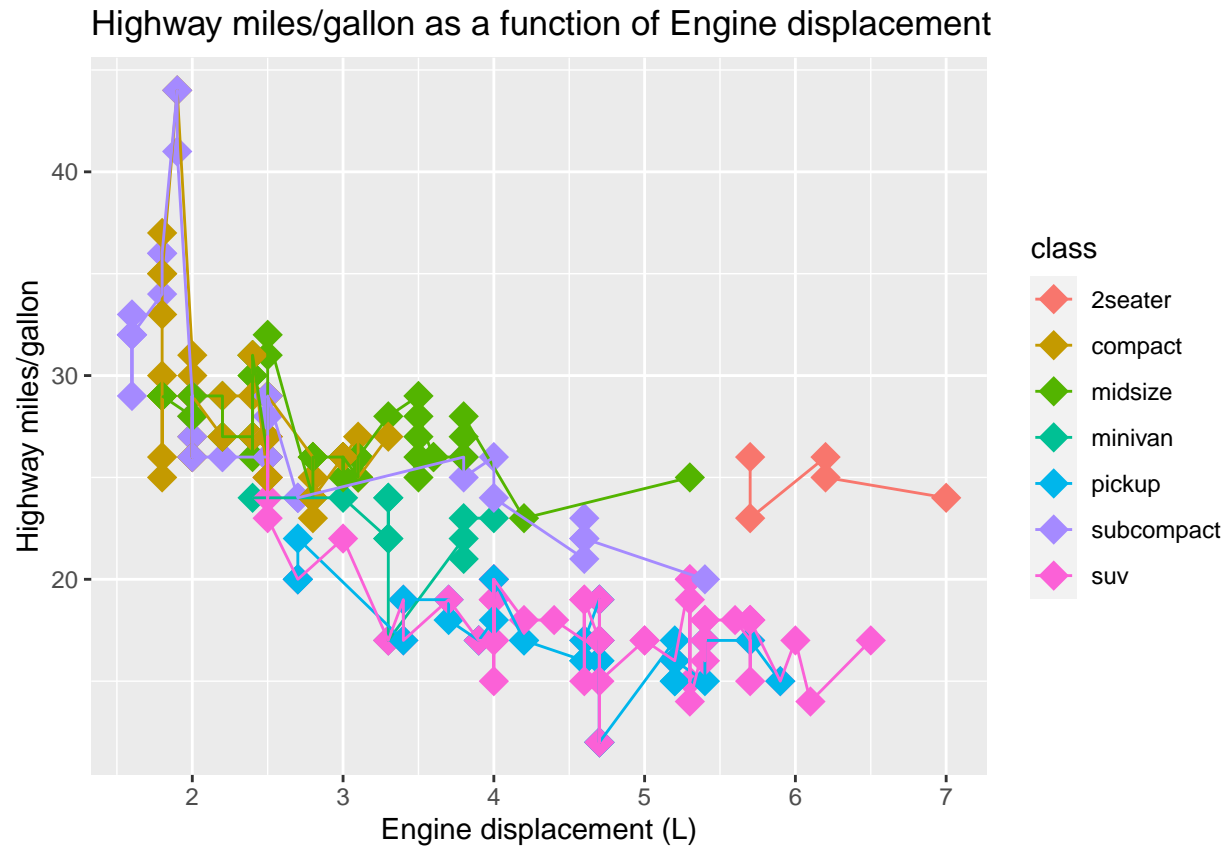
Now that we have the plot we wanted, let's arrange it to make it prettier and easier to read. First, let's change the names of the axes to make it clearer.

```
ggplot(mpg)+
  geom_point(aes(x=displ, y=hwy, color=class), shape="diamond", size=5)+
  geom_line(aes(x=displ, y=hwy, color=class))+
  xlab("Engine displacement (L)")+
  ylab("Highway miles/gallon")
```



We could also add a title.

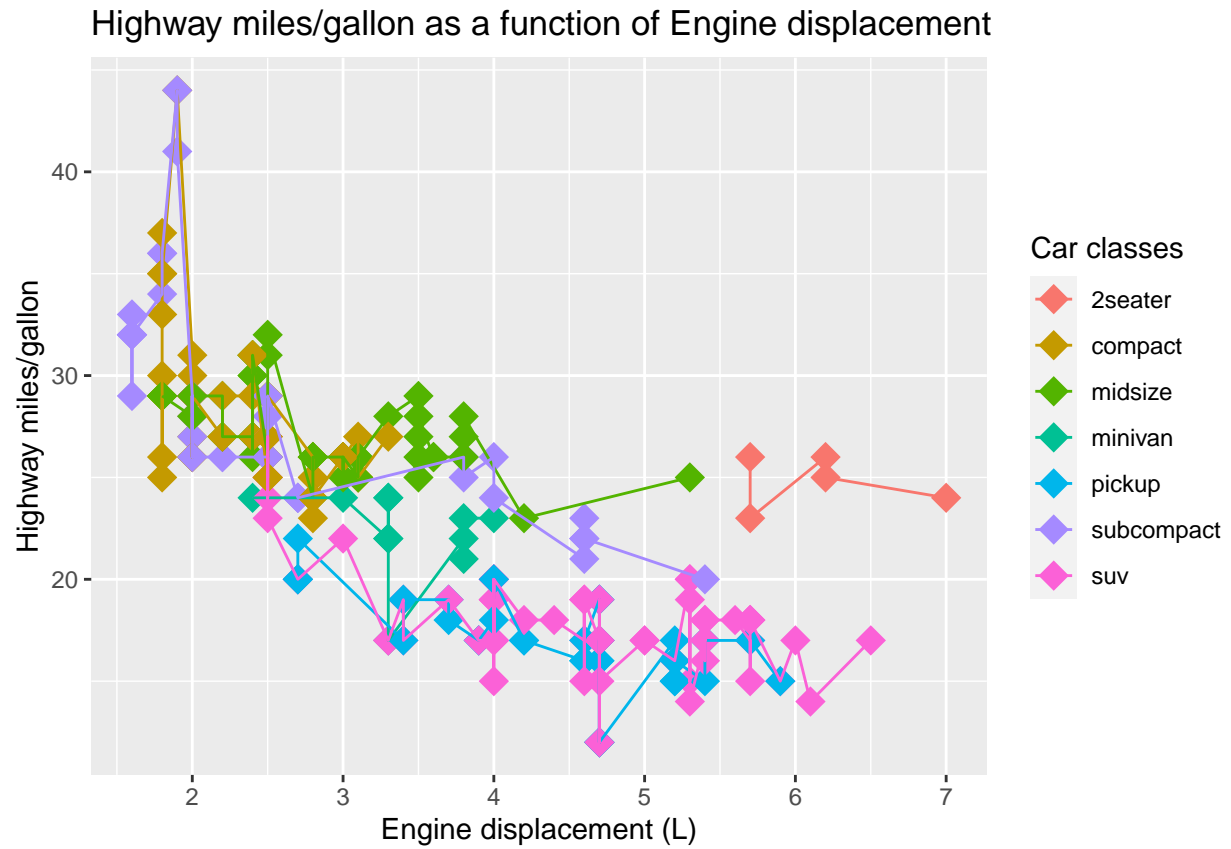
```
ggplot(mpg)+
  geom_point(aes(x=displ, y=hwy, color=class), shape="diamond", size=5)+
  geom_line(aes(x=displ, y=hwy, color=class))+
  xlab("Engine displacement (L)")+
  ylab("Highway miles/gallon")+
  ggtitle("Highway miles/gallon as a function of Engine displacement")
```



But that's a lot of functions to add to our ggplot object. We could also use the function `labs()`.

```
ggplot(mpg)+
  geom_point(aes(x=displ, y=hwy, color=class), shape="diamond", size=5)+
  geom_line(aes(x=displ, y=hwy, color=class))+
  labs(x="Engine displacement (L)",
       y="Highway miles/gallon",
       title="Highway miles/gallon as a function of Engine displacement",
       color="Car classes")
```





With this function we can also replace the name of the aesthetics, color, shape, size, fill, group, etc.