Graphing Using ggplot Part-1

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

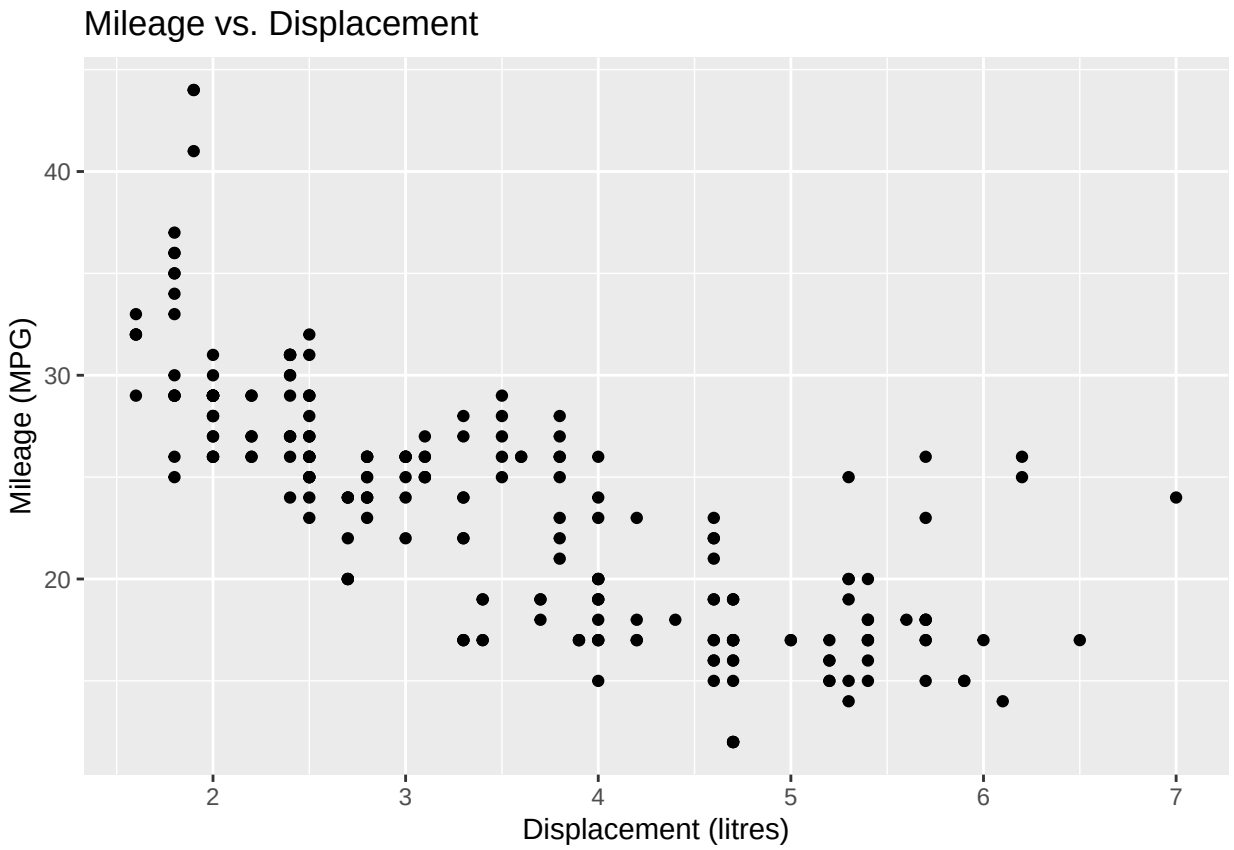
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# Load the mpg dataset
data('mpg')
mpgData = mpg

# Print the first five rows (or samples) in the data frame
head(mpgData, 5)
```

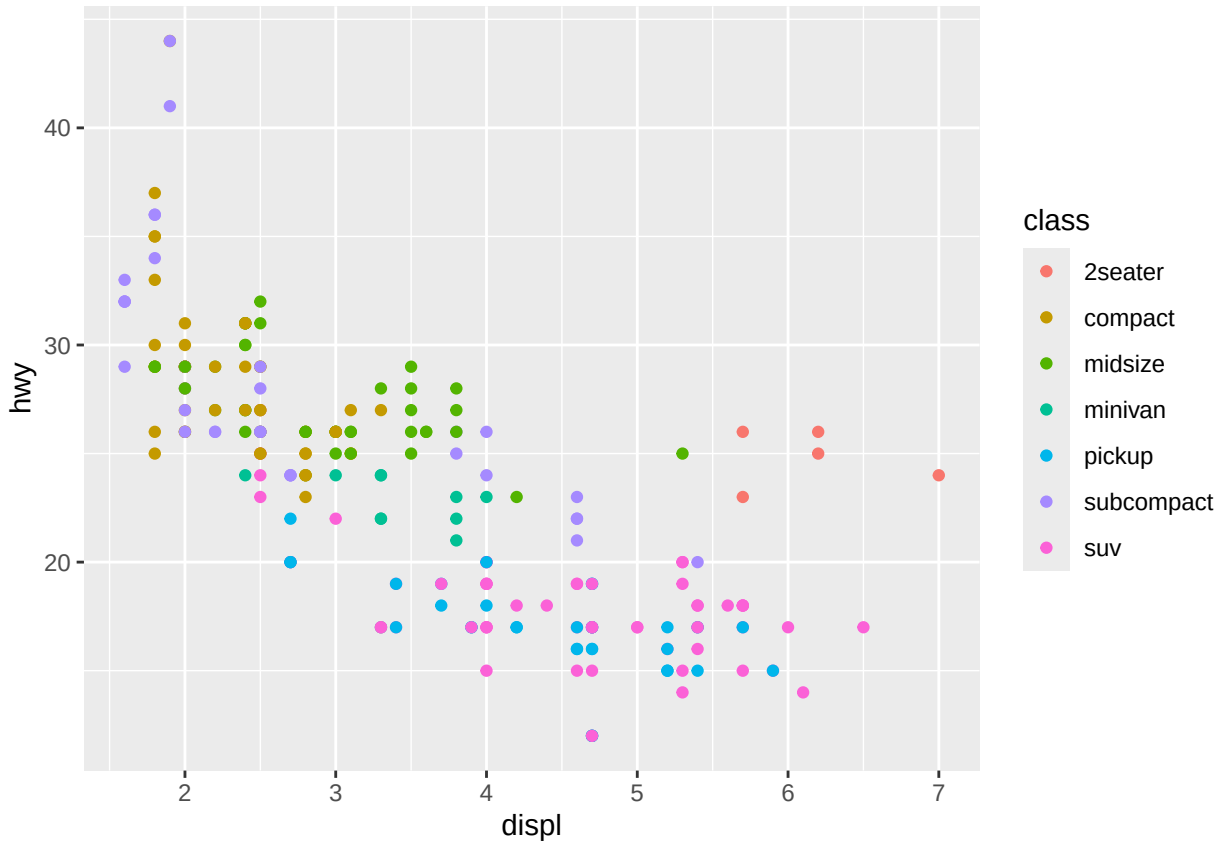
```
## # A tibble: 5 x 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p     compa~
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     compa~
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     compa~
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     compa~
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p     compa~
```

```
# Initiate the ggplot() function binding to the car data frame
ggplot(data = mpgData)
```

```
# Create a plot object
p1 = ggplot(data = mpgData)

# Use the aes() function to specify the aesthetic mapping, that is, which variables
# should be plotted
p1 = ggplot(data = mpgData, aes(x = disp, y = hwy))

# Use the geom_ type functions to add geometric elements
p1 = ggplot(data = mpgData, aes(x = displ, y = hwy)) +
  geom_point()

# Add labels and title
p1 = p1 + labs(x = 'Displacement (litres)', y = 'Mileage (MPG)', title = 'Mileage vs. Displacement')
p1
```

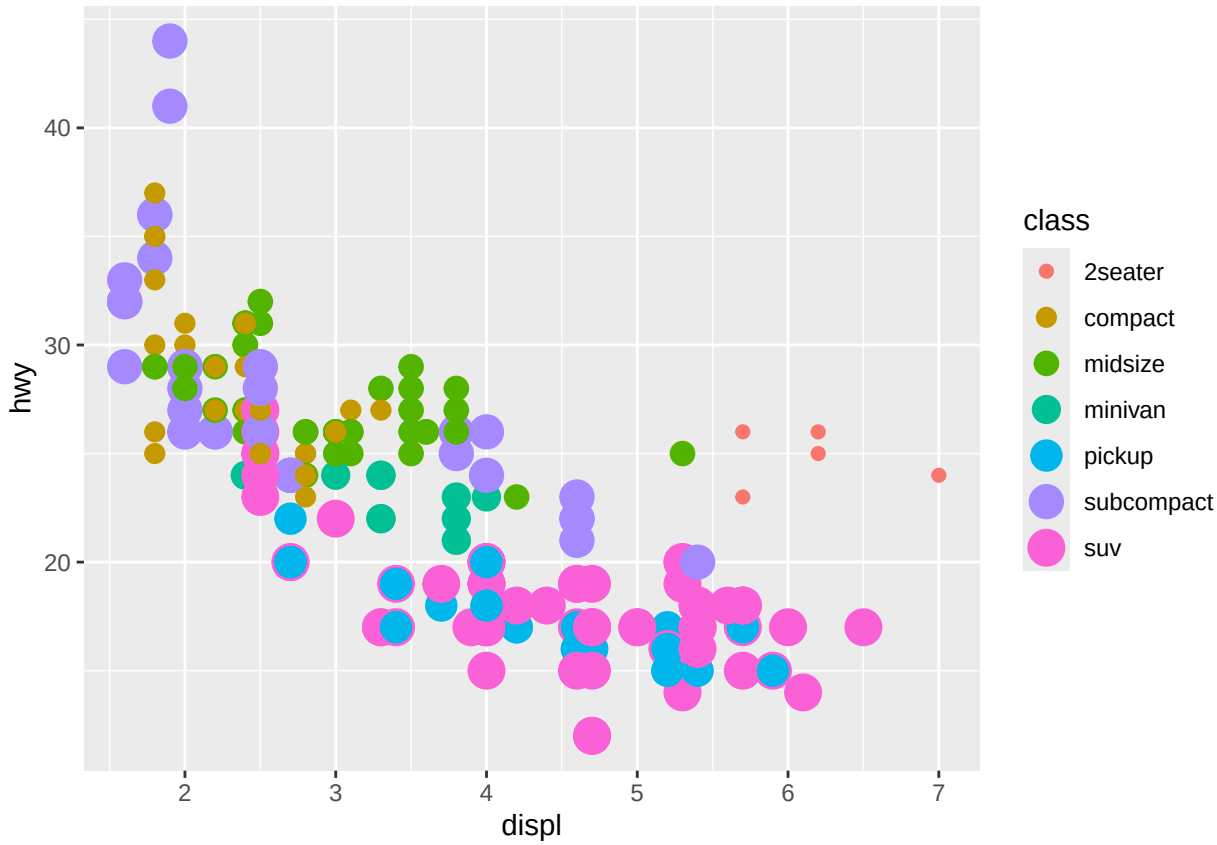



```
# Map aesthetics to variables

# Map the color aesthetic to the class variable
p2 = ggplot(data = mpgData) +
  geom_point(mapping = aes(x = displ, y = hwy, colour = class))
p2
```

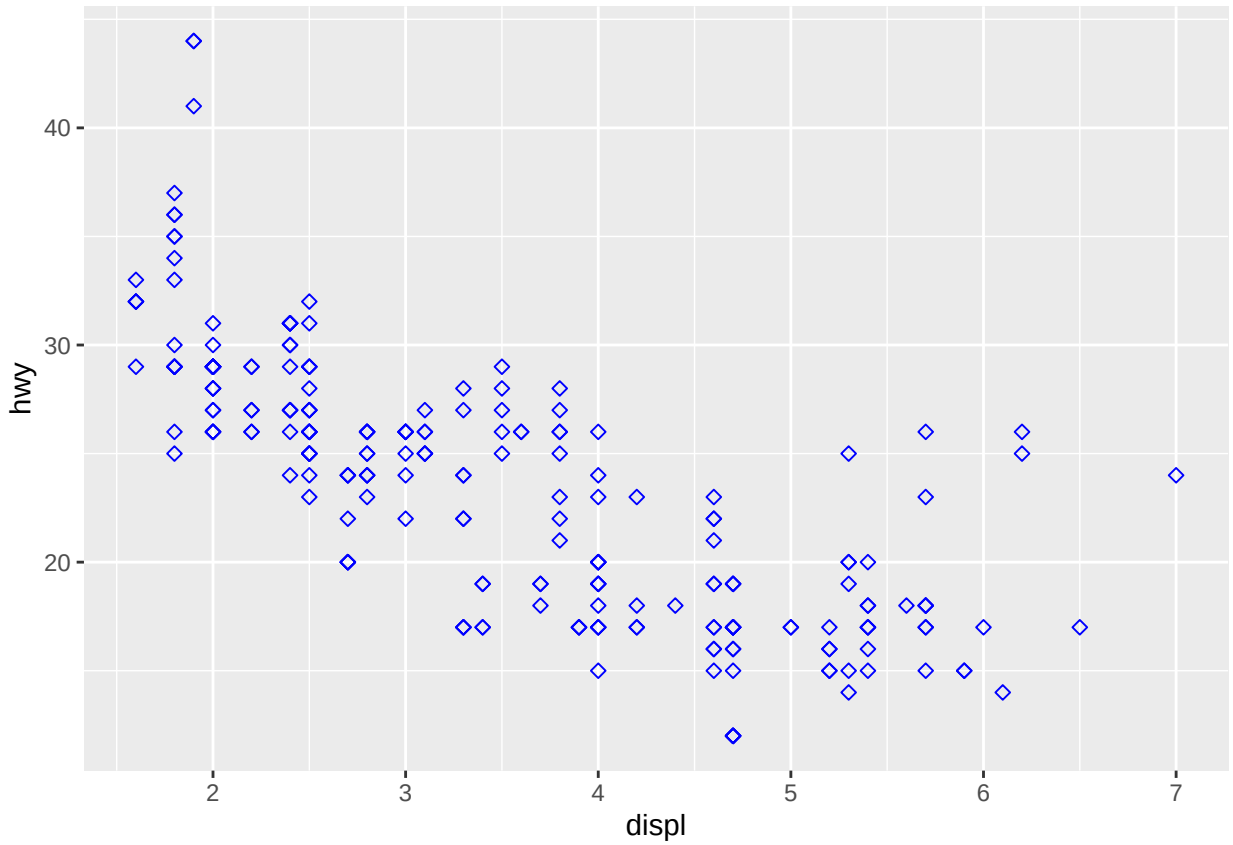



```
# Map aesthetics to variables

# Map the size (or alpha or shape) aesthetic to the class variable
p3 = ggplot(data = mpgData) +
  geom_point(aes(x = displ, y = hwy, size = class, colour = class))
p3
```

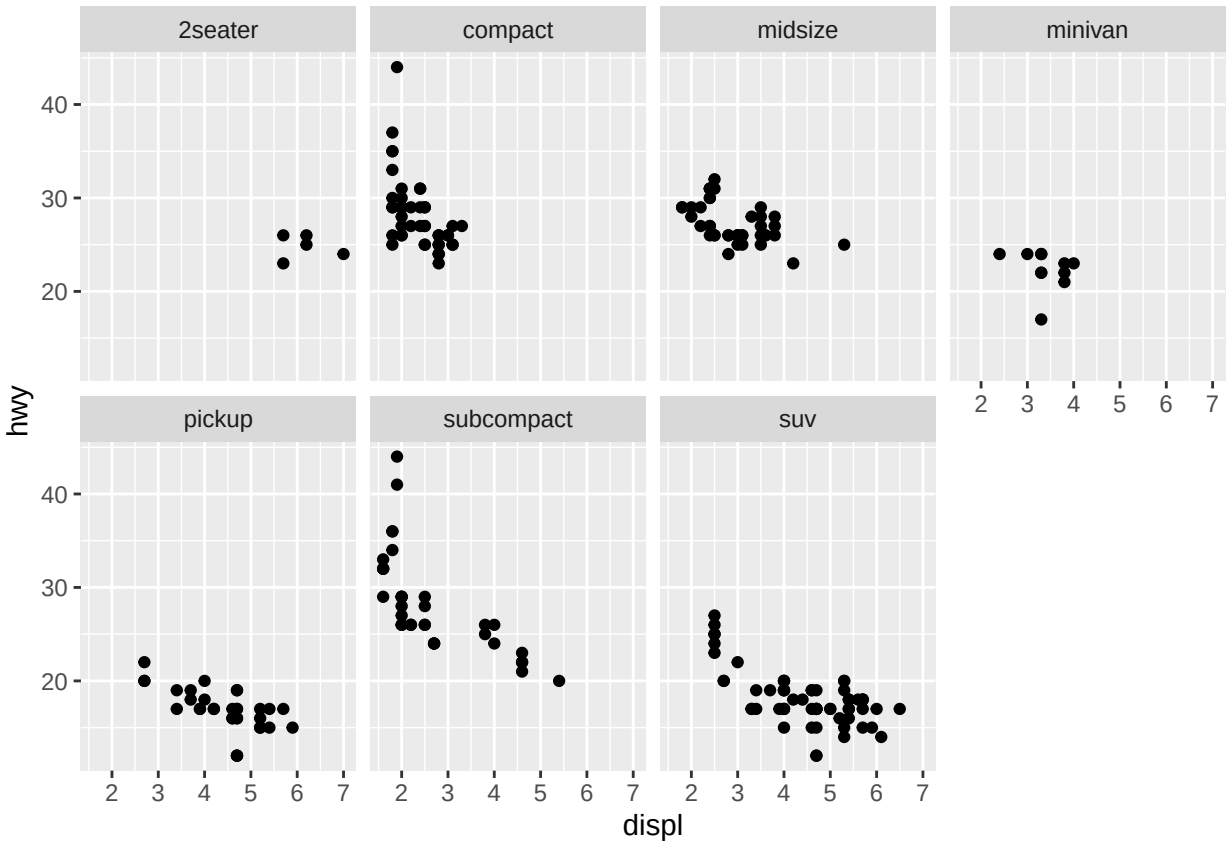
```
## Warning: Using size for a discrete variable is not advised.
```



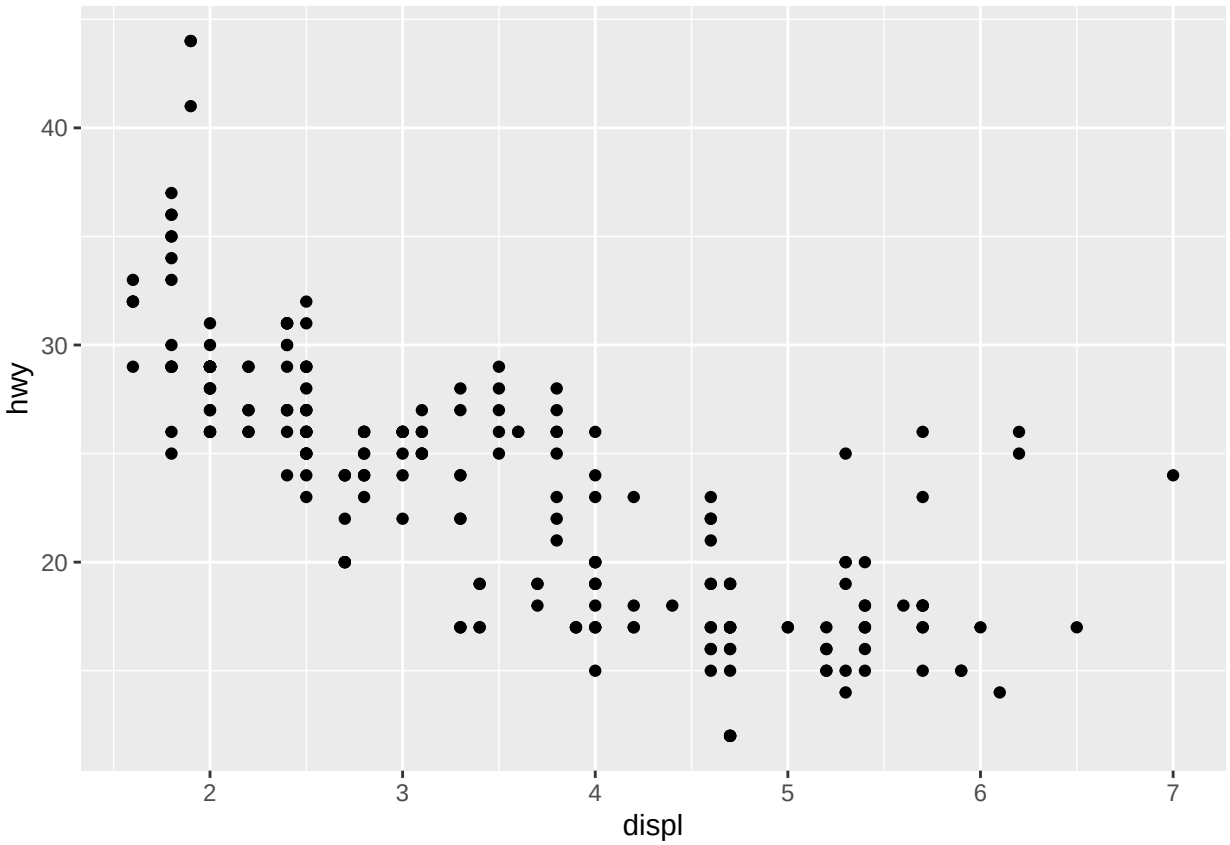

```
# Set aesthetic manually

# Mark the points blue in color, square in shape, and fixed size
p4 = ggplot(data = mpgData) +
  geom_point(aes(x = displ, y = hwy), color = 'blue', shape = 5)
p4
```




```
# Add additional variables using facets

# Plot mileage w.r.t. each class individually
ggplot(data = mpgData) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_wrap(~class, nrow = 2)
```




```
# Plot mileage w.r.t drive train and number of cylinders
ggplot(data = mpgData) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_grid()
```




```
# A quick exercise on facet plotting with filtering of samples
# Investigate the number of levels in the cyl feature
  factor(mpgData$cyl)
```

```
##   [1] 4 4 4 4 6 6 6 4 4 4 4 6 6 6 6 6 6 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 4 4 6 6 6
##  [38] 4 6 6 6 6 6 6 6 6 6 6 6 6 6 6 8 8 8 8 8 6 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8
##  [75] 8 8 8 6 6 6 6 8 8 6 6 8 8 8 8 8 6 6 6 6 8 8 8 8 4 4 4 4 4 4 4 4 4 4 4 4 4
## [112] 4 6 6 6 4 4 4 4 6 6 6 6 6 6 8 8 8 8 8 8 8 8 8 8 8 8 6 6 8 8 4 4 4 4 6 6 6
## [149] 6 6 6 6 6 8 6 6 6 6 8 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 6 6 6 8 4 4 4 4 6 6
## [186] 6 4 4 4 4 6 6 6 4 4 4 4 4 8 8 4 4 4 6 6 6 6 4 4 4 4 6 4 4 4 4 4 5 5 6 6 4
## [223] 4 4 4 5 5 4 4 4 4 6 6 6
## Levels: 4 5 6 8
```

```
# #
# # # How many cars of each cyl type are there?
  mpgData %>% count(cyl)
```

```
## # A tibble: 4 x 2
##     cyl     n
##   <int> <int>
## 1     4    81
## 2     5     4
## 3     6    79
## 4     8    70
```

```
# Filter samples with 3 and 5 cylinder cars (which are very rare)
mpgData = mpgData %>% filter(!(cyl %in% c(5)))
```

```
#mpgData = mpgData %>% filter(!(? %in% c(?, ?)))
#
head(mpgData, 5)
```

```
## # A tibble: 5 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(15)   f        18    29 p     compa~
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     compa~
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     compa~
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     compa~
## 5 audi         a4      2.8  1999     6 auto(15)   f        16    26 p     compa~
```

```
#
# # Map the color aesthetic to the cyl variable
p5 = ggplot(data = mpgData) +
   geom_point(mapping = aes(x = displ, y = hwy, color = factor(cyl), shape = class))
 p5
```

```
## Warning: The shape palette can deal with a maximum of 6 discrete values because more
## than 6 becomes difficult to discriminate
## i you have requested 7 values. Consider specifying shapes manually if you need
##   that many have them.
```

```
## Warning: Removed 62 rows containing missing values or values outside the scale range
## (`geom_point()`).
```