RWorksheet_Aguas#6

Diana Marie Aguas

2022-11-30

```
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
```

```
library(ggplot2)

data(mpg)
dataA <- as.data.frame(mpg)
```

1. How many columns are in mpg dataset? How about the number of rows? Show the codes and its result. #Ans: The mpg dataset has 234 rows and 11 colums.

```
rowncol <- c(nrow(mpg), ncol(mpg))
rowncol
```

```
## [1] 234  11
```

2. Which manufacturer has the most models in this data set? Which model has the most variations? Ans: The manufacturer that has the most model in this data set is dodge with the model caravan 2wd

```
carvar <- dataA %>% group_by(manufacturer, model) %>%
  count()
carvar
```

```
## # A tibble: 38 x 3
## # Groups:   manufacturer, model [38]
##    manufacturer model                  n
##    <chr>        <chr>              <int>
##  1 audi         a4                     7
##  2 audi         a4 quattro             8
##  3 audi         a6 quattro             3
##  4 chevrolet    c1500 suburban 2wd     5
##  5 chevrolet    corvette               5
##  6 chevrolet    k1500 tahoe 4wd        4
##  7 chevrolet    malibu                 5
##  8 dodge        caravan 2wd           11
##  9 dodge        dakota pickup 4wd      9
## 10 dodge        durango 4wd            7
## # ... with 28 more rows
```

```
colnames(carvar) <- c("Manufacturer", "Model", "Counts")
```

a. Group the manufacturers and find the unique models. Copy the codes and result.

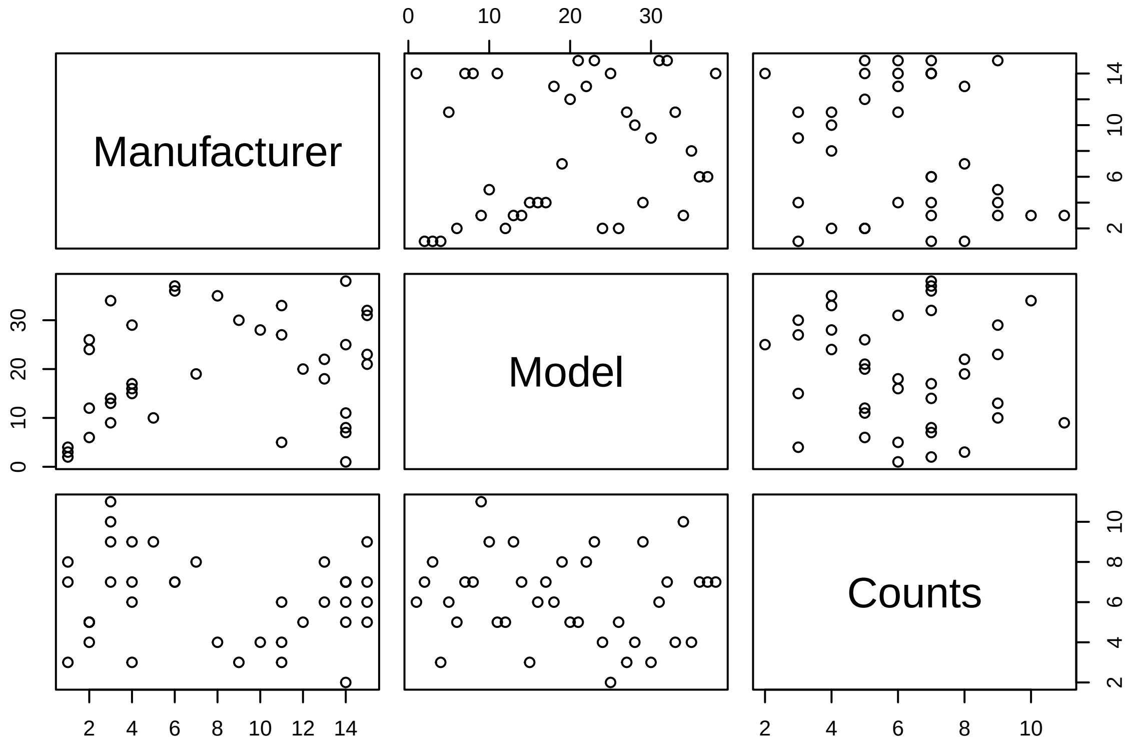
```
uniquemodels <- dataA %>% group_by(manufacturer, model) %>%
  distinct() %>% count()
uniquemodels
```

```
## # A tibble: 38 x 3
## # Groups:   manufacturer, model [38]
##    manufacturer model                  n
##    <chr>        <chr>              <int>
##  1 audi         a4                     7
##  2 audi         a4 quattro             8
##  3 audi         a6 quattro             3
##  4 chevrolet    c1500 suburban 2wd     4
##  5 chevrolet    corvette               5
##  6 chevrolet    k1500 tahoe 4wd        4
##  7 chevrolet    malibu                 5
##  8 dodge        caravan 2wd            9
##  9 dodge        dakota pickup 4wd      8
## 10 dodge        durango 4wd            6
## # ... with 28 more rows
```

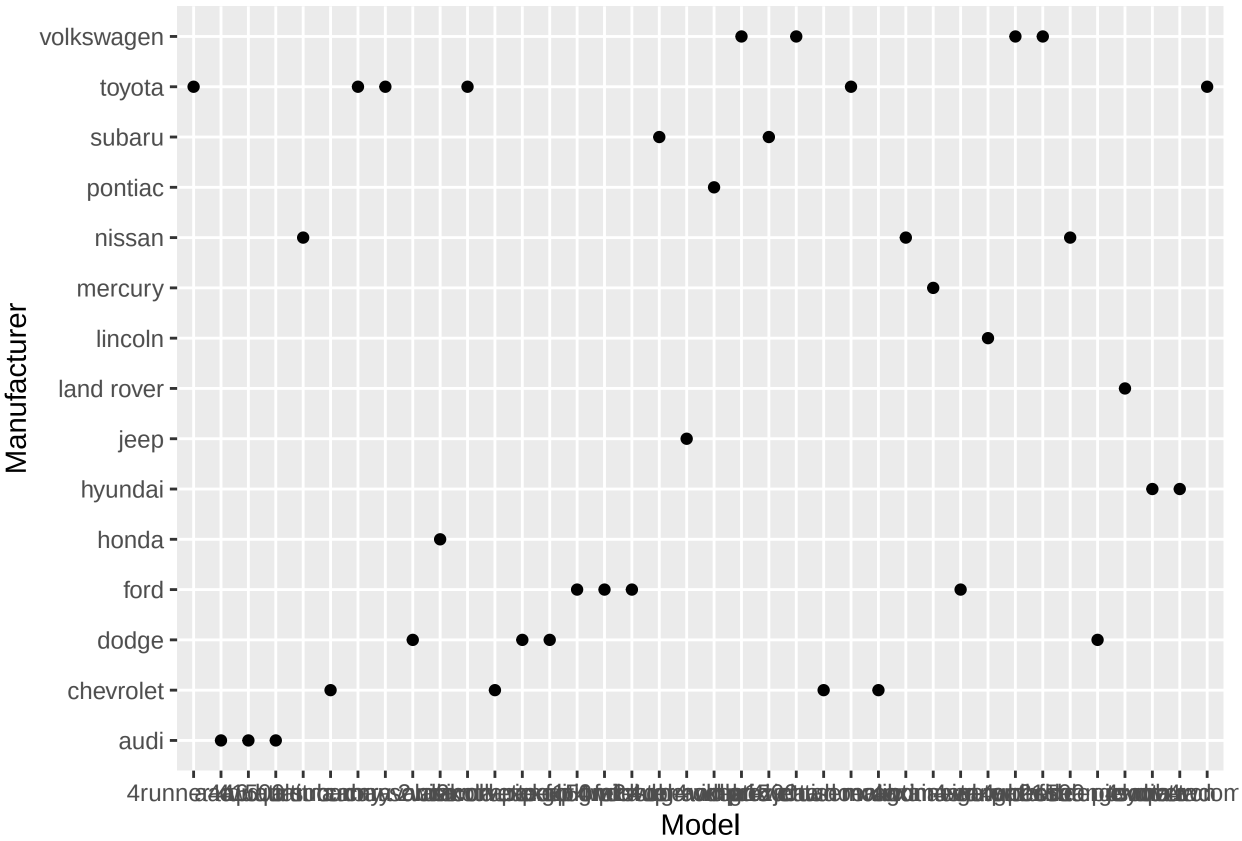
```
colnames(uniquemodels) <- c("Manufacturers", "Model", "Counts")
```

b. Graph the result by using plot() and ggplot(). Write the codes and its result.

```
plot(carvar)
```



```
ggplot(carvar, aes(Model, Manufacturer)) + geom_point()
```



3. Same data set will be used. You are going to show the relationship of the model and the manufacturer.

```
modnmanuf <- mpg

modelfact <- modnmanuf %>% group_by(manufacturer, model) %>%
  distinct() %>% count()
modelfact
```

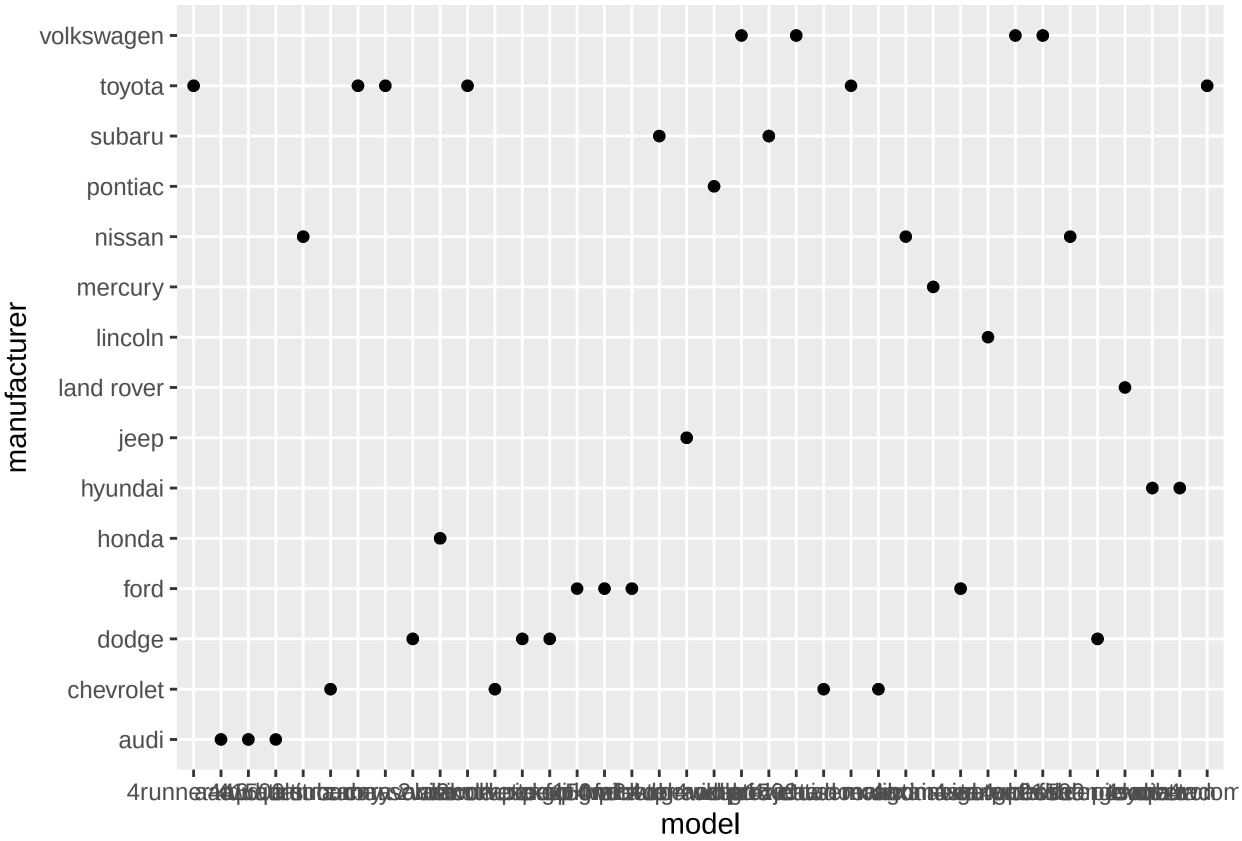
```
## # A tibble: 38 x 3
## # Groups:   manufacturer, model [38]
##    manufacturer model                  n
##    <chr>        <chr>              <int>
##  1 audi         a4                     7
##  2 audi         a4 quattro             8
##  3 audi         a6 quattro             3
##  4 chevrolet    c1500 suburban 2wd     4
##  5 chevrolet    corvette               5
##  6 chevrolet    k1500 tahoe 4wd        4
##  7 chevrolet    malibu                 5
##  8 dodge        caravan 2wd            9
##  9 dodge        dakota pickup 4wd      8
## 10 dodge        durango 4wd            6
## # ... with 28 more rows
```

```
colnames(modelfact) <- c("Manufacturer", "Model")
modelfact
```

```
## # A tibble: 38 x 3
## # Groups:   Manufacturer, Model [38]
##    Manufacturer Model                 ``
##    <chr>        <chr>              <int>
##  1 audi         a4                     7
##  2 audi         a4 quattro             8
##  3 audi         a6 quattro             3
##  4 chevrolet    c1500 suburban 2wd     4
##  5 chevrolet    corvette               5
##  6 chevrolet    k1500 tahoe 4wd        4
##  7 chevrolet    malibu                 5
##  8 dodge        caravan 2wd            9
##  9 dodge        dakota pickup 4wd      8
## 10 dodge        durango 4wd            6
## # ... with 28 more rows
```

a. What does ggplot(mpg, aes(model, manufacturer)) + geom_point() show?

```
ggplot(mpg, aes(model, manufacturer)) + geom_point()
```



b. For you, is it useful? If not, how could you modify the data to make it more informative?

- it is useful, yes, but using legend would be very much useful to summarize the data.

4. Using the pipe (%>%), group the model and get the number of cars per model. Show codes and its result.

```
group_mod <- dataA %>% group_by(model) %>% count()
group_mod
```

```
## # A tibble: 38 x 2
## # Groups:   model [38]
##    model                  n
##    <chr>              <int>
##  1 4runner 4wd            6
##  2 a4                     7
##  3 a4 quattro             8
##  4 a6 quattro             3
##  5 altima                 6
##  6 c1500 suburban 2wd     5
##  7 camry                  7
##  8 camry solara           7
##  9 caravan 2wd           11
## 10 civic                  9
## # ... with 28 more rows
```

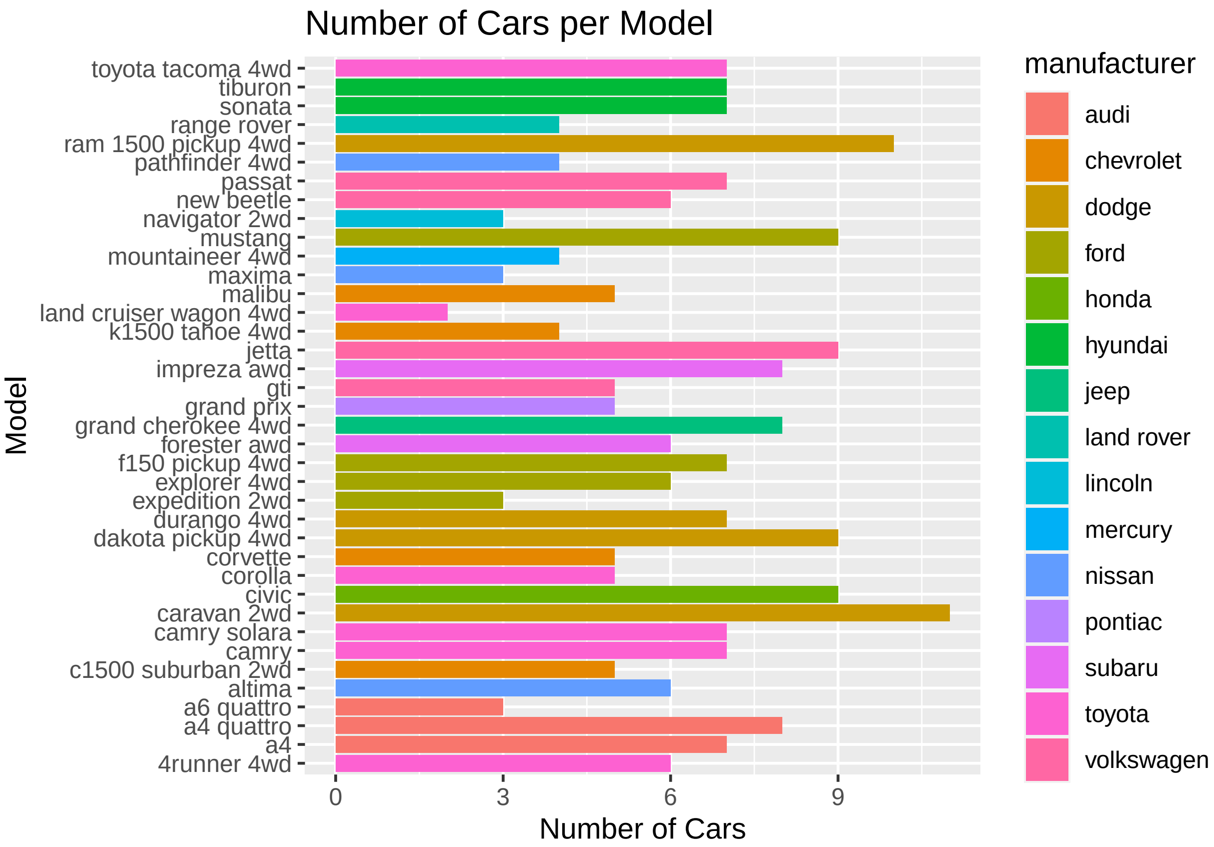
```
colnames(group_mod) <- c("Model", "Counts")
group_mod
```

```
## # A tibble: 38 x 2
## # Groups:   Model [38]
##    Model              Counts
##    <chr>               <int>
##  1 4runner 4wd             6
##  2 a4                      7
##  3 a4 quattro              8
##  4 a6 quattro              3
##  5 altima                  6
##  6 c1500 suburban 2wd      5
##  7 camry                   7
##  8 camry solara            7
##  9 caravan 2wd            11
## 10 civic                   9
## # ... with 28 more rows
```

a. Plot using the geom_bar() + coord_flip() just like what is shown below. Show codes and its result.

```
qplot(model,data = mpg,main = "Number of Cars per Model", ylab = "Number of Cars",
      xlab = "Model", geom = "bar", fill = manufacturer) +
  coord_flip()
```

```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
```



b. Use only the top 20 observations. Show code and results.

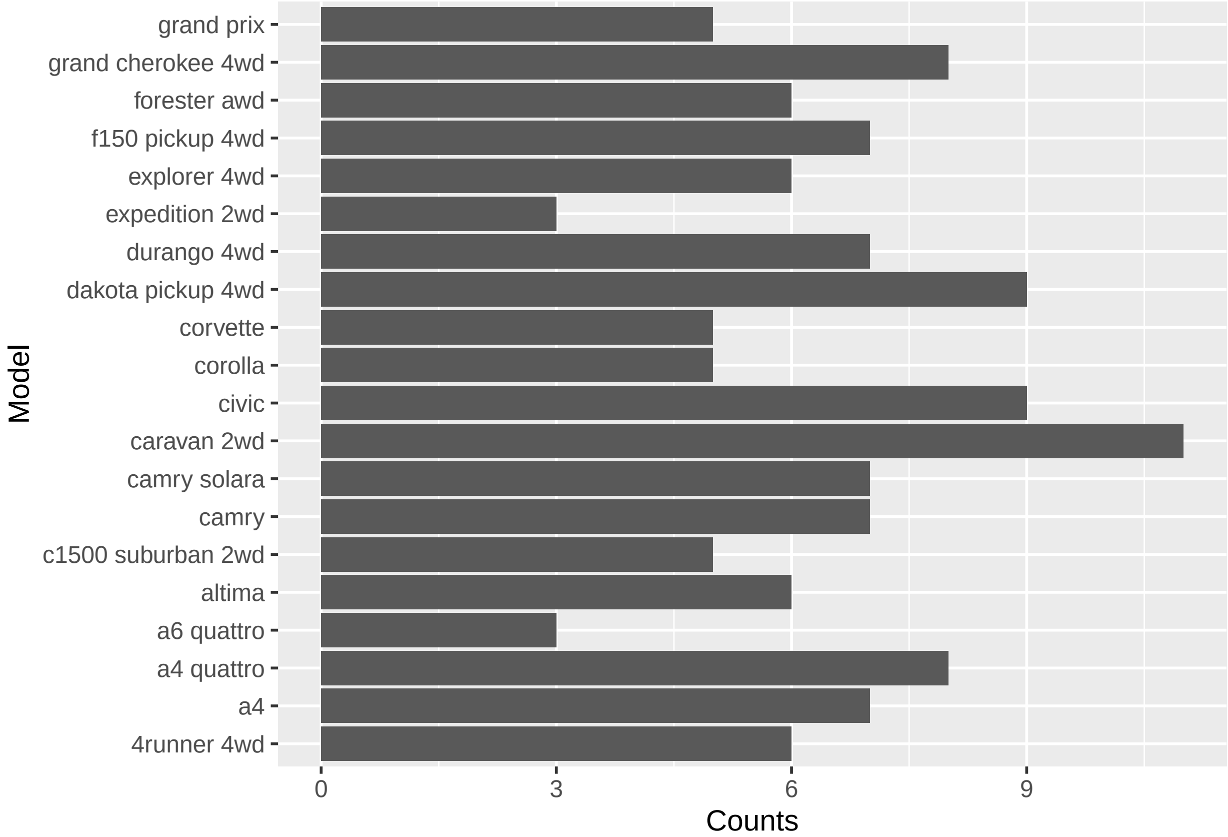
```
twenty <- group_mod[1:20,] %>% top_n(2)
```

```
## Selecting by Counts
```

```
twenty
```

```
## # A tibble: 20 x 2
## # Groups:   Model [20]
##    Model              Counts
##    <chr>               <int>
##  1 4runner 4wd             6
##  2 a4                      7
##  3 a4 quattro              8
##  4 a6 quattro              3
##  5 altima                  6
##  6 c1500 suburban 2wd      5
##  7 camry                   7
##  8 camry solara            7
##  9 caravan 2wd            11
## 10 civic                   9
## 11 corolla                 5
## 12 corvette                5
## 13 dakota pickup 4wd       9
## 14 durango 4wd             7
## 15 expedition 2wd          3
## 16 explorer 4wd            6
## 17 f150 pickup 4wd         7
## 18 forester awd            6
## 19 grand cherokee 4wd      8
## 20 grand prix              5
```

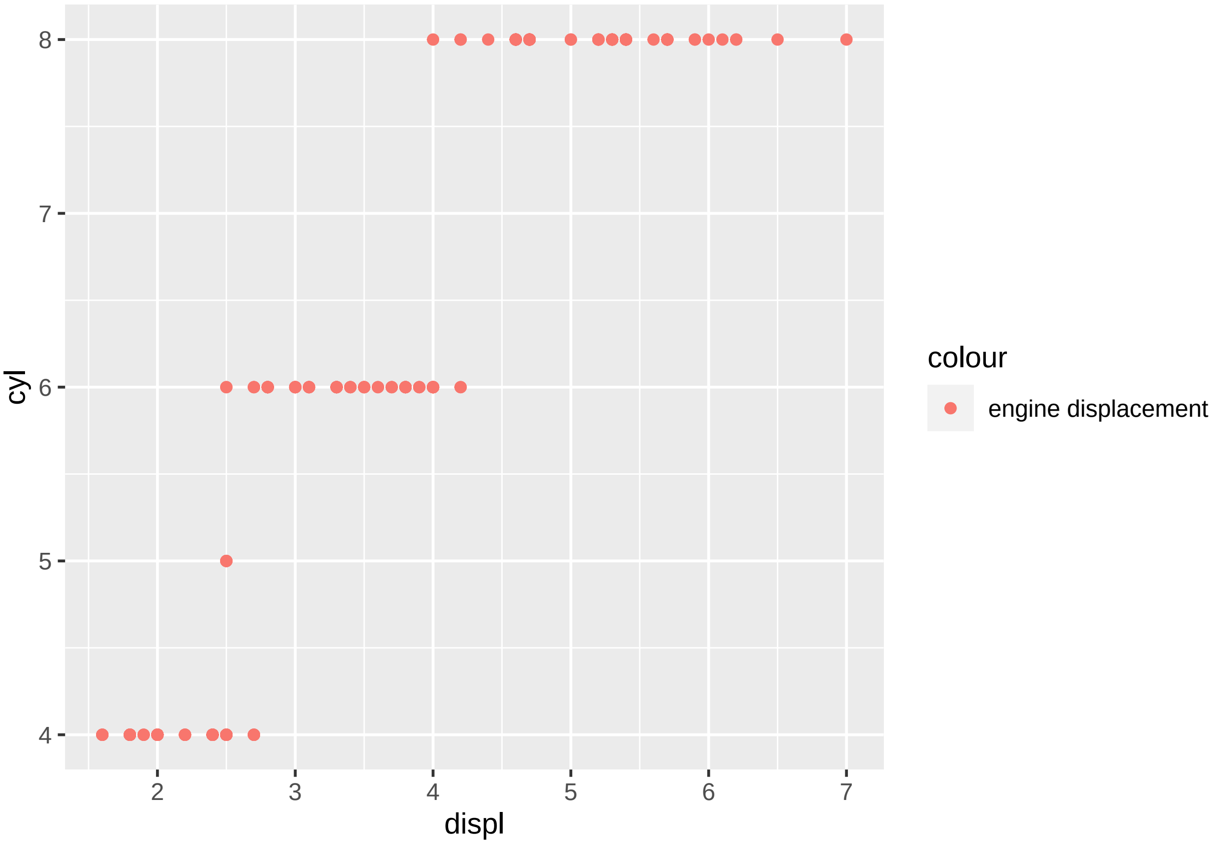
```
ggplot(twenty, aes(x = Model, y = Counts)) +
  geom_bar(stat = "identity") + coord_flip()
```



5. Plot the relationship between cyl - number of cylinders and displ - engine displacement using geom_point with aesthetic colour = engine displacement. Title should be "Relationship between No. of Cylinders and Engine Displacement".

a. Show the codes and its result.

```
ggplot(data = mpg , mapping = aes(x = displ, y = cyl, main = "Relationship
                                    between No of Cylinders and
                                    Engine Displacement")) +
  geom_point(mapping=aes(colour = "engine displacement"))
```



b.

How would you describe its relationship?

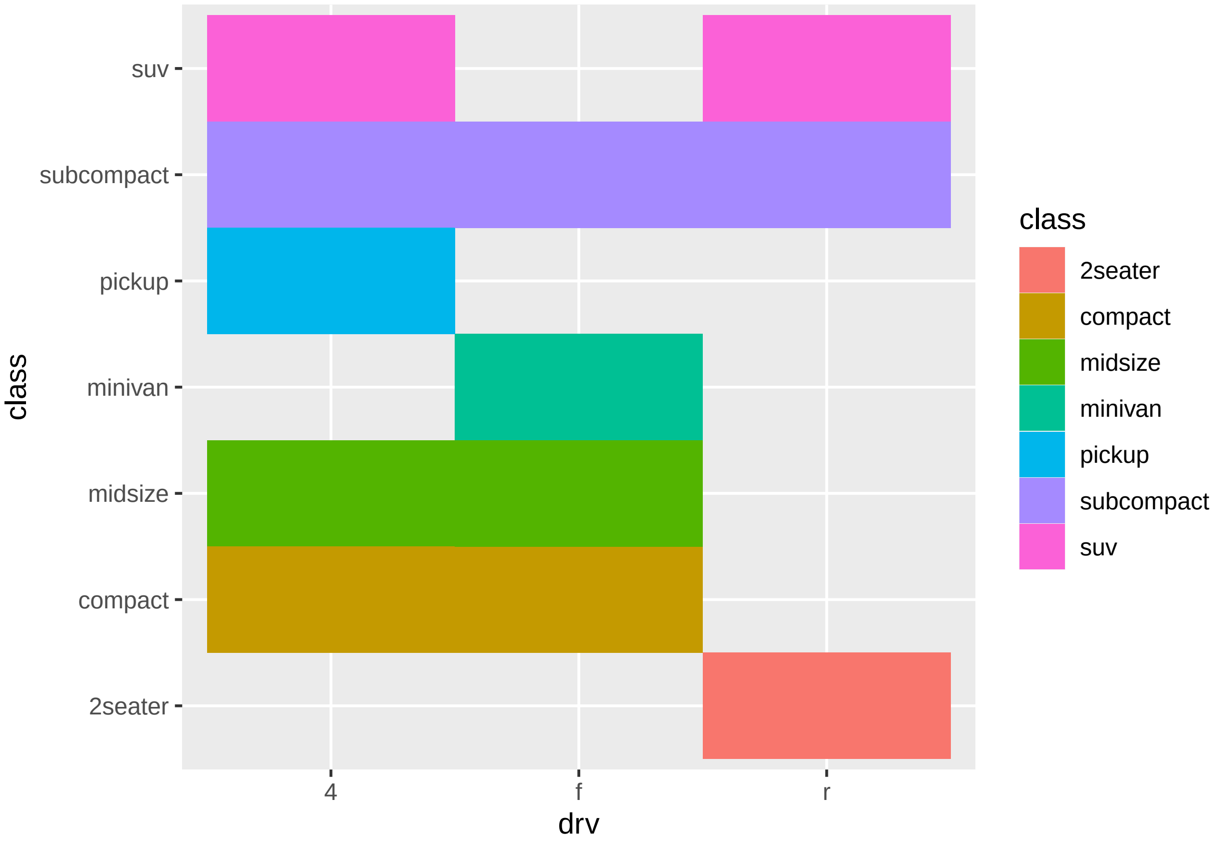
- cyl is the y in the chart which stands for the number of cylinders

and displ is the x/engine displacement.

6. Get the total number of observations for drv - type of drive train (f = front-wheel drive, r = rear wheel drive, 4 = 4wd) and class - type of class (Example: suv, 2seater, etc.). Plot using the geom_tile() where the number of observations for class be used as a fill for aesthetics.

a. Show the codes and its result for the narrative in #6.

```
ggplot(data = mpg, mapping = aes(x = drv, y = class)) + geom_tile(aes(fill=class))
```



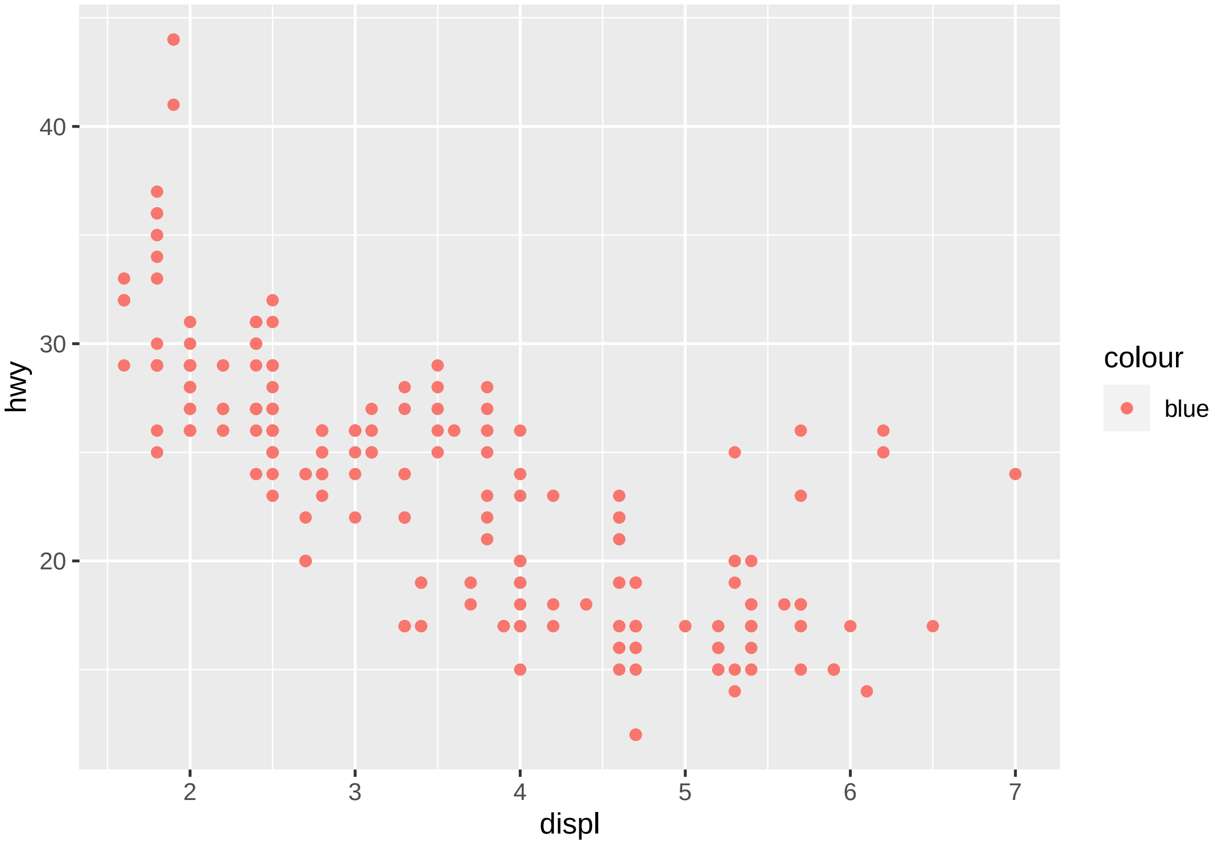
b. Interpret the result.

- geom_tile() with fill makes it easier to navigate through the chart and identify class(y axis) and drive trains(drv- x axis)

7. Discuss the difference between these codes. Its outputs for each are shown below.

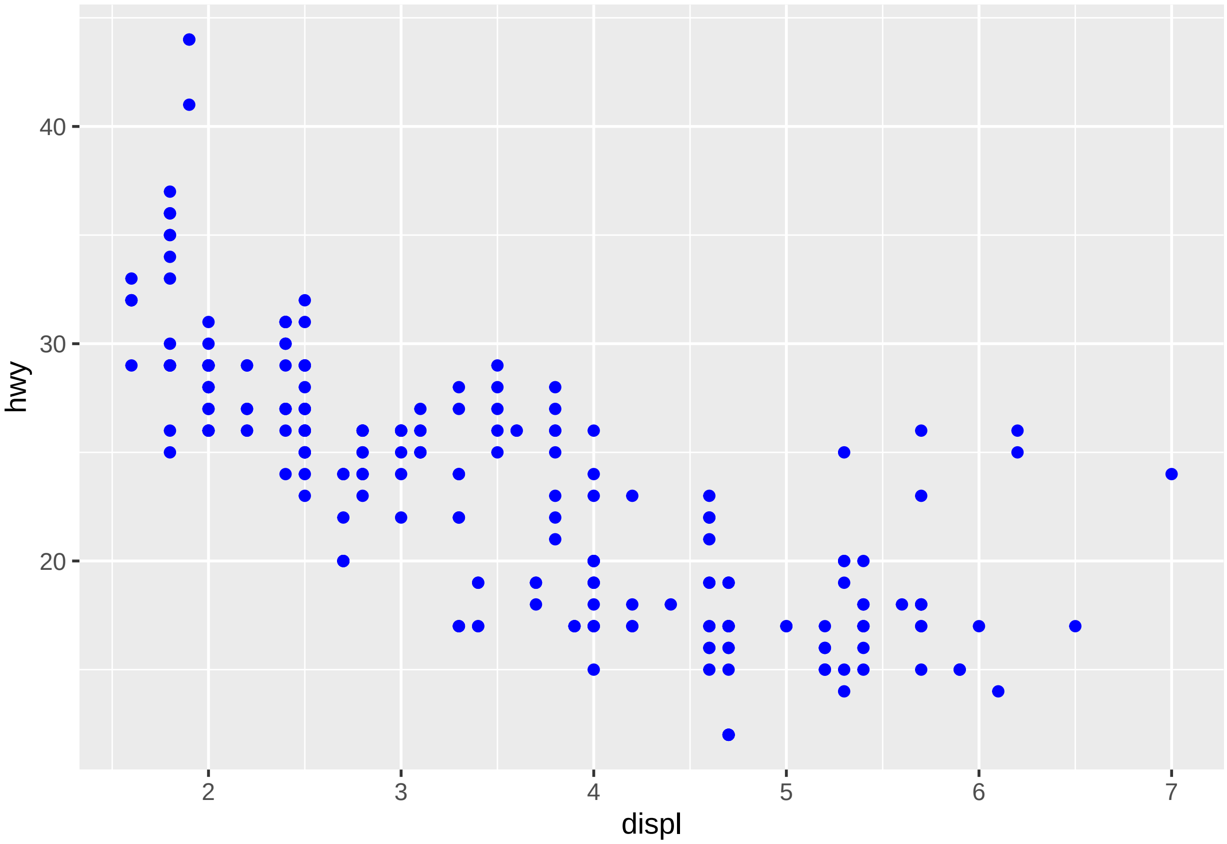
• Code #1

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, colour = "blue"))
```



- **Code #2**

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy), colour = "blue")
```



8.

Try to run the command ?mpg. What is the result of this command?

```
?mpg
```

- It showed the dataset that cointains a subset of the fuel economy data from 1999 to 2008 for 38 popular model cars.

a. Which variables from mpg dataset are categorical?

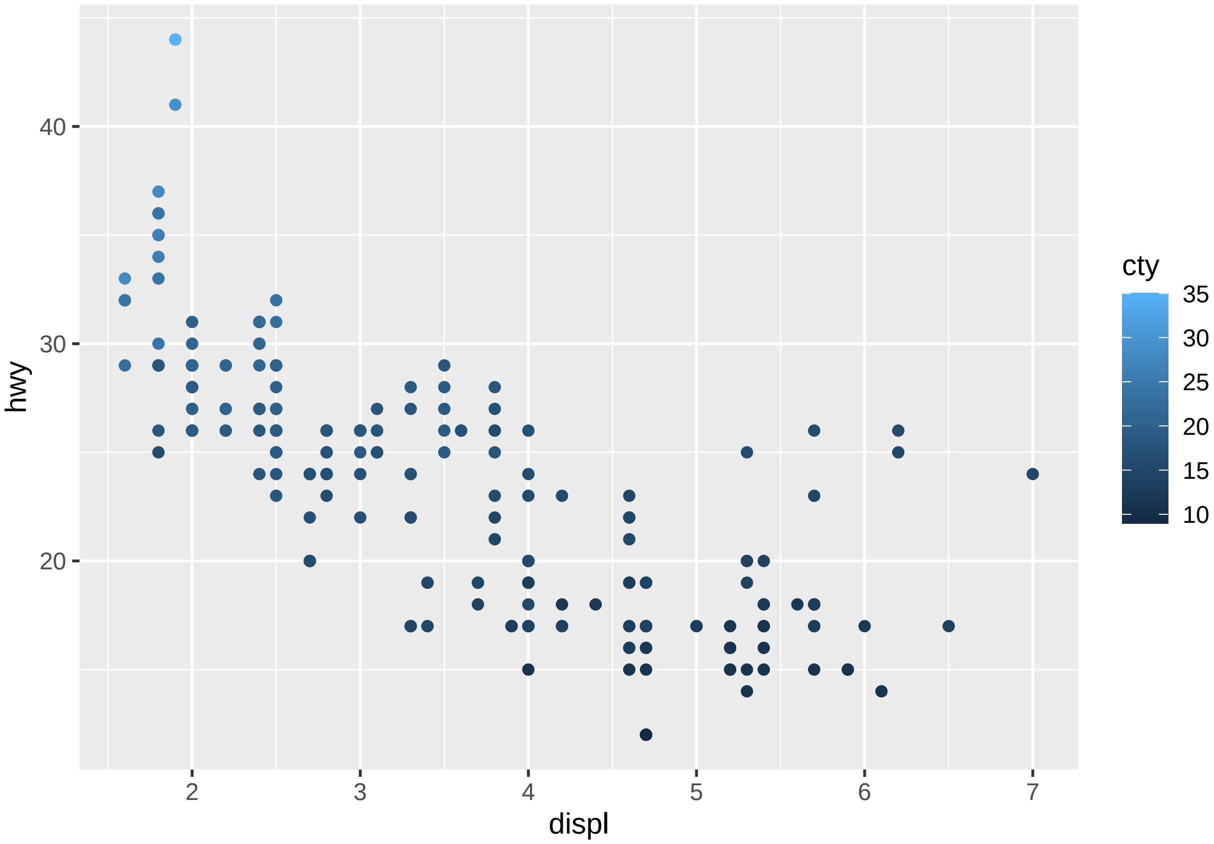
- Manufacturer trans ,model, drv, fl, and class

b. Which are continuous variables?

- displ, cyl, and hwy

c. Plot the relationship between displ (engine displacement) and hwy(highway miles per gallon). Mapped it with a continuous variable you have identified in #5-b. What is its result? Why it produced such output?

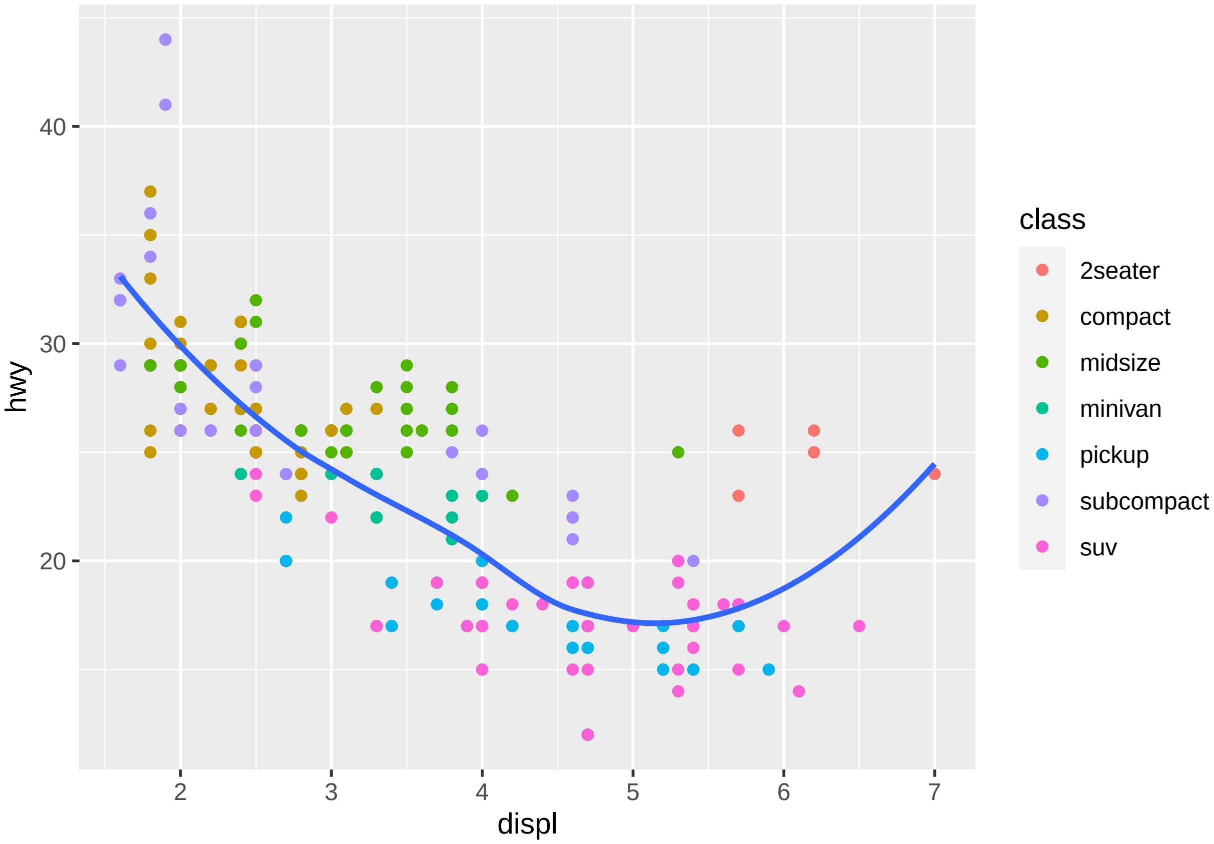
```
ggplot(mpg, aes(x = displ, y = hwy, colour = cty)) + geom_point()
```



9. Plot the relationship between displ (engine displacement) and hwy(highway miles per gallon) using geom_point(). Add a trend line over the existing plot using geom_smooth() with se = FALSE. Default method is “loess”.

```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +
  geom_point(mapping=aes(color=class)) +
  geom_smooth(se = FALSE)
```

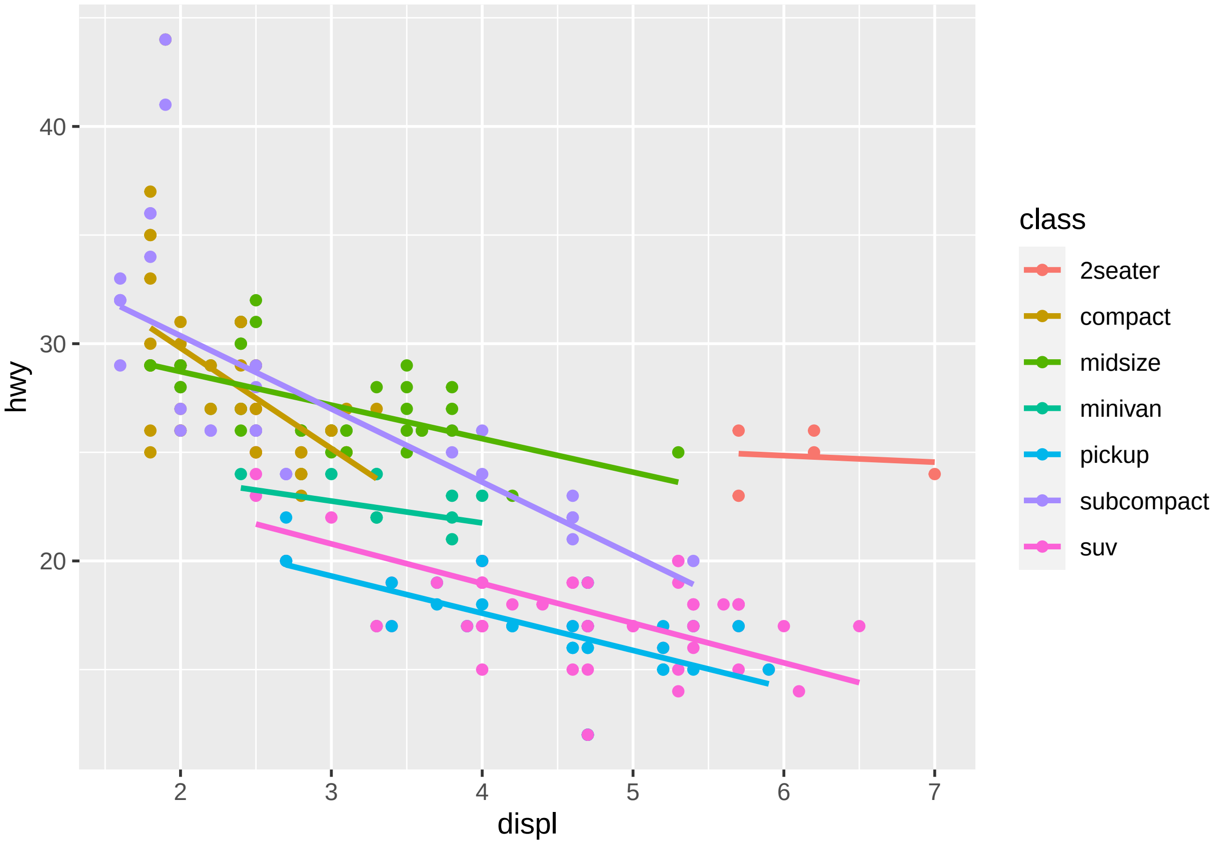
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



10. Using the relationship of displ and hwy, add a trend line over existing plot. Set these = FALSE to remove the confidence interval and method = lm to check for linear modeling.

```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy, color = class)) + geom_point() +
  geom_smooth(se = FALSE, method = lm)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```


40
30
20
hwy
2
3
4
5
6
7
displ
class
2seater
compact
midsize
minivan
pickup
subcompact
suv