

# Case Study 1

Mohamed Megahed

4/19/2020

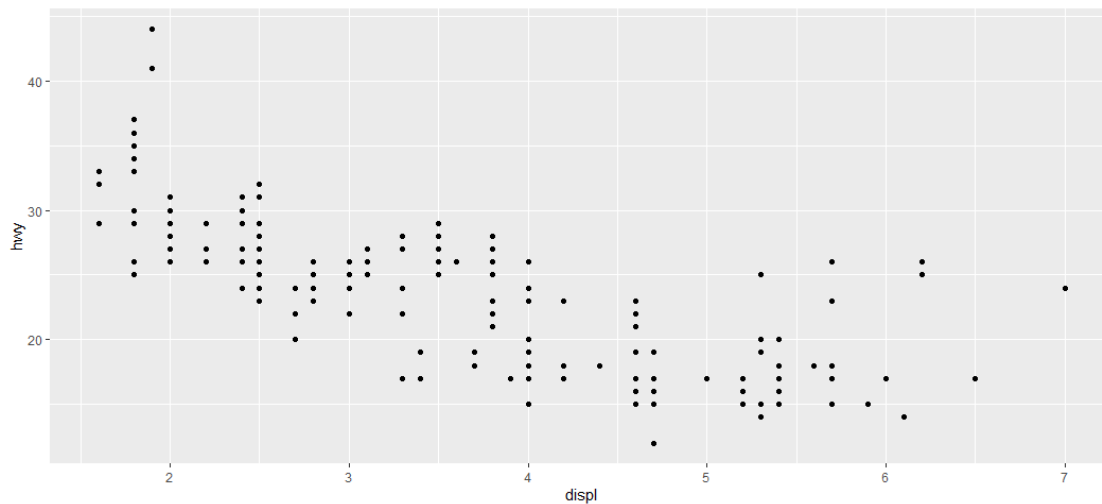
## Class Practice - 1

```
library(ggplot2)
```

### 1. Plot the following using ggplot():

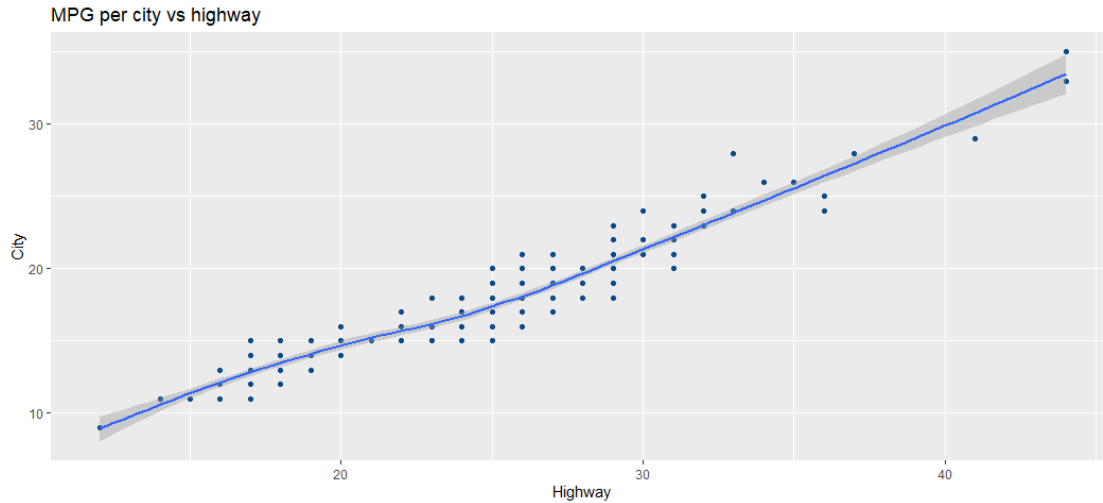
```
ggplot(cars, aes(x=displ, y = hwy)) + geom_point()
```

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



2. How would you describe the relationship between city and highway mileage (plot graph using ggplot)? Do you have any concerns about drawing conclusions from that plot?

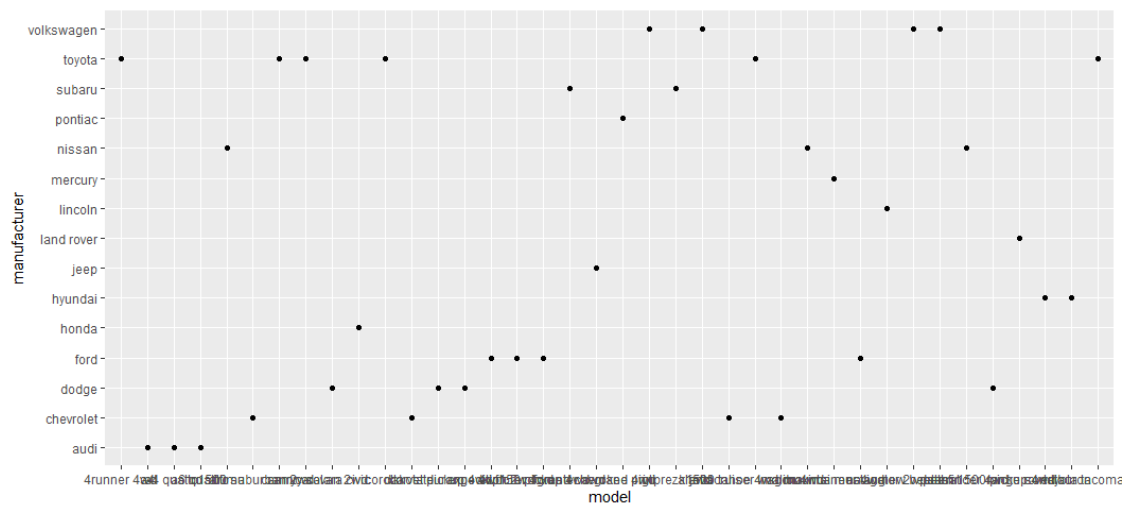
```
ggplot(mpg) + aes(x = hwy, y = city) + geom_point(size = 1.52, colour = "#0c4c8a") + geom_smooth() +  
  labs(x = "Highway", y = "City", title = "MPG per city vs highway")  
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



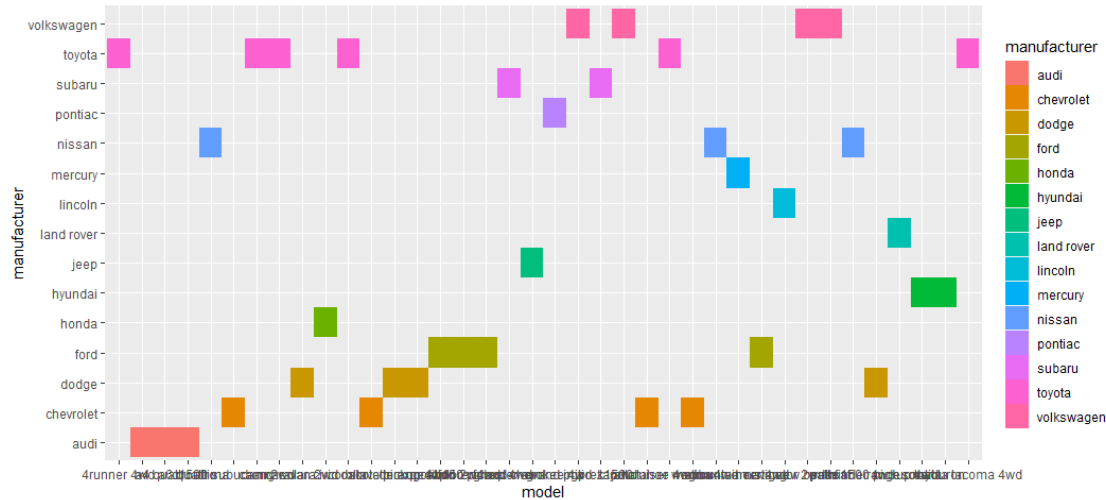
From the graph we can see a positive correlation between number of miles per gallon on the hwy and on the city.

3. Plot the following graph and explain the output:

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



```
ggplot(mpg, aes(x = model, y = manufacturer, fill = manufacturer)) + geom_bin  
2d()
```



*graph is not useful as the model and manufacturer are categorical variables, so we can use different plot to visualize the relation better*

4. Is the above graph useful? How could you modify the data to make it more informative?\_\_

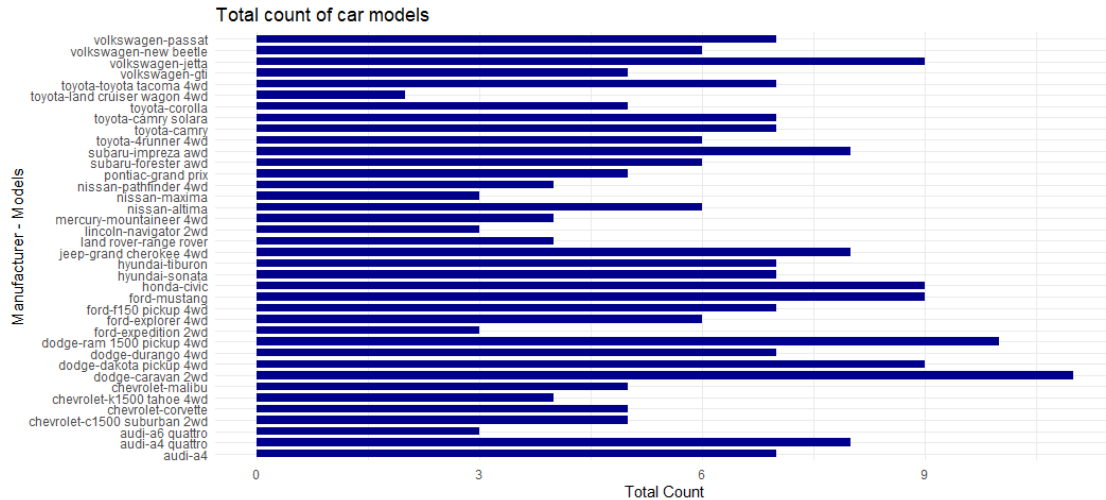
```
library(dplyr)
```

```
##
## 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

mpg %>% transmute(manufacturerAndmodel = paste(manufacturer, model, sep = "-"))
%>% ggplot(aes(manufacturerAndmodel)) + geom_bar(fill = "darkblue", width = 0.7) +
  labs(x = "Manufacturer - Models", y = "Total Count", title = "Total count of car models") + theme_minimal() + coord_flip()
```

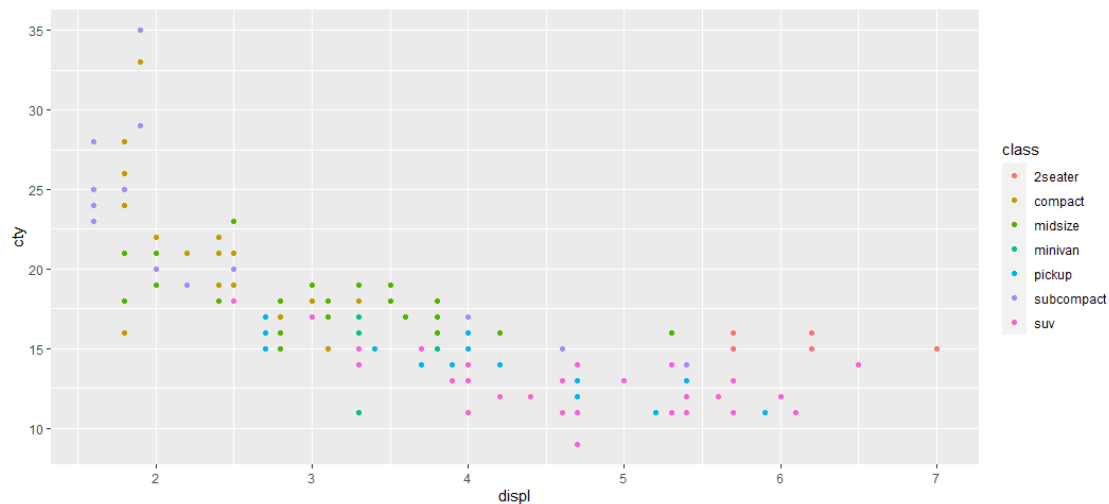


## Class Practice - 2

1. Plot the following using ggplot()\_

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

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

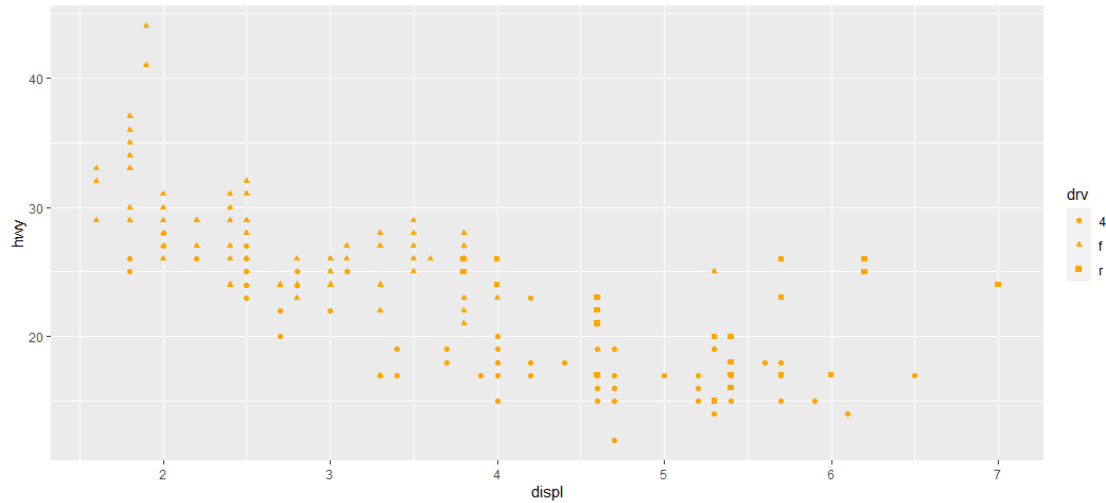


2. Plot

the following (adding one more dimension):

`ggplot(mpg, aes(x=displ, y = hwy, shape = drv)) + geom_point(colour = "orange")`

`ggplot(mpg, aes(x=displ, y = hwy, shape = drv)) +  
geom_point(colour = "orange")`



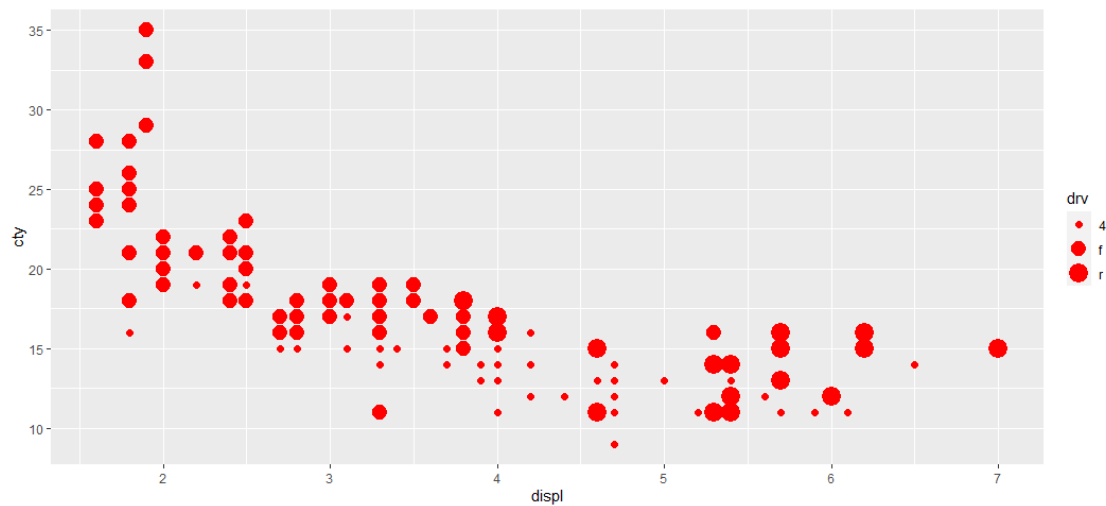
3. Plot

the following graph (adding more dimensions)

```
ggplot(mpg, aes(x=displ, y = cty, size = drv)) + geom_point(colour="green")
```

```
ggplot(mpg, aes(x=displ, y = cty, size = drv)) +  
  geom_point(colour="red")
```

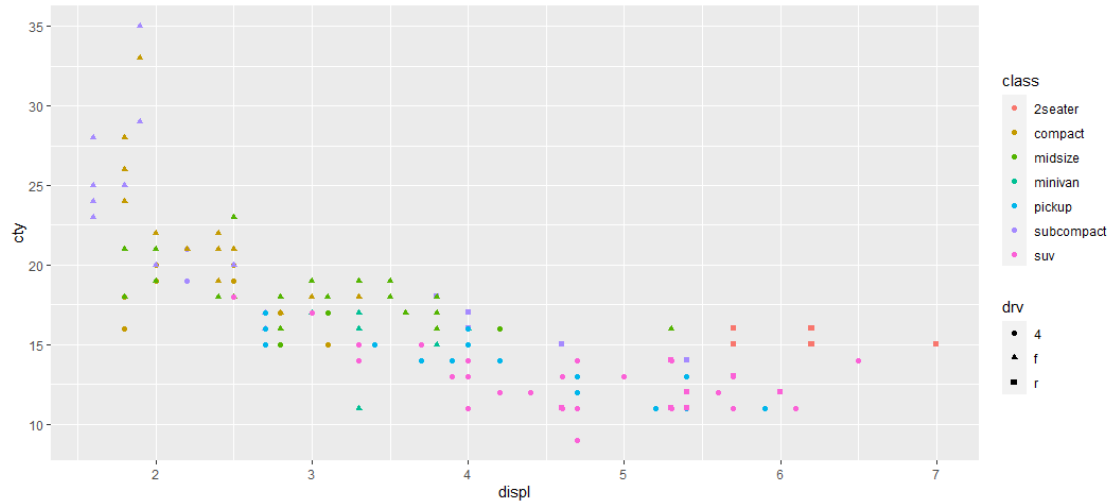
```
## Warning: Using size for a discrete variable is not advised.
```



4. Plot

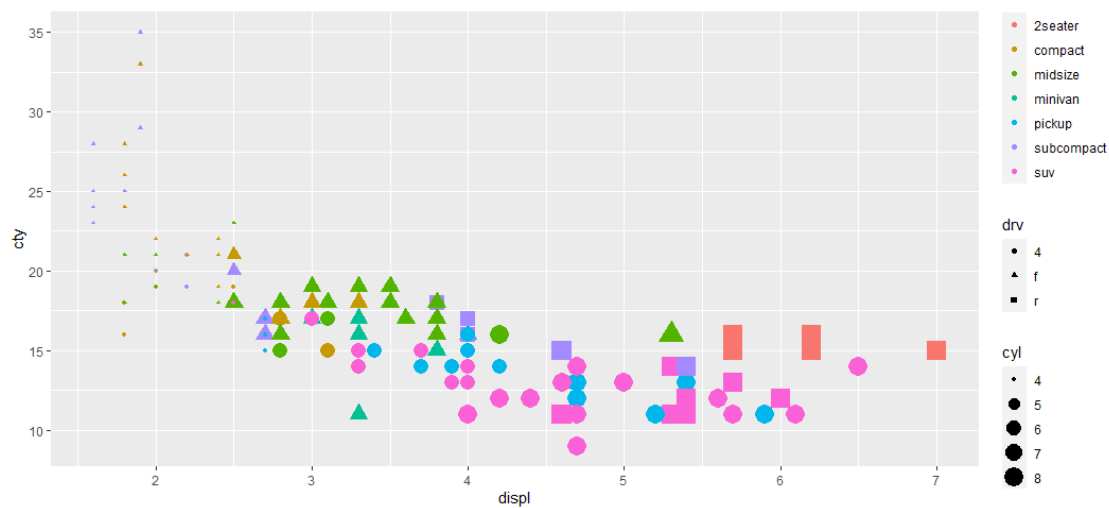
the following graph:

```
ggplot(mpg, aes(x=displ, y = cty, shape = drv, colour=class)) + geom_point()
```



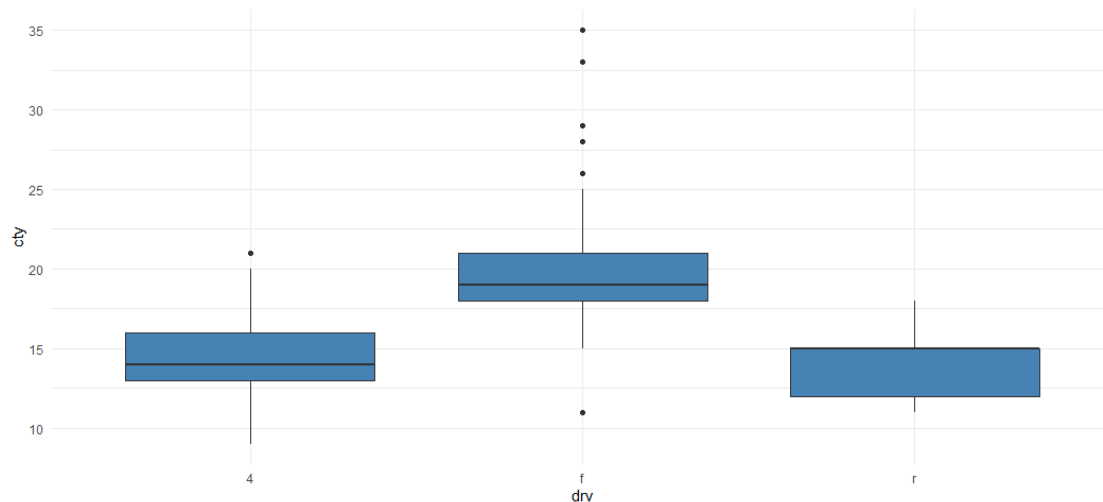
5. Plot the following graph:

```
ggplot(mpg, aes(x=displ, y = cty, shape = drv, colour=class, size=cyl)) + geom_point()
```



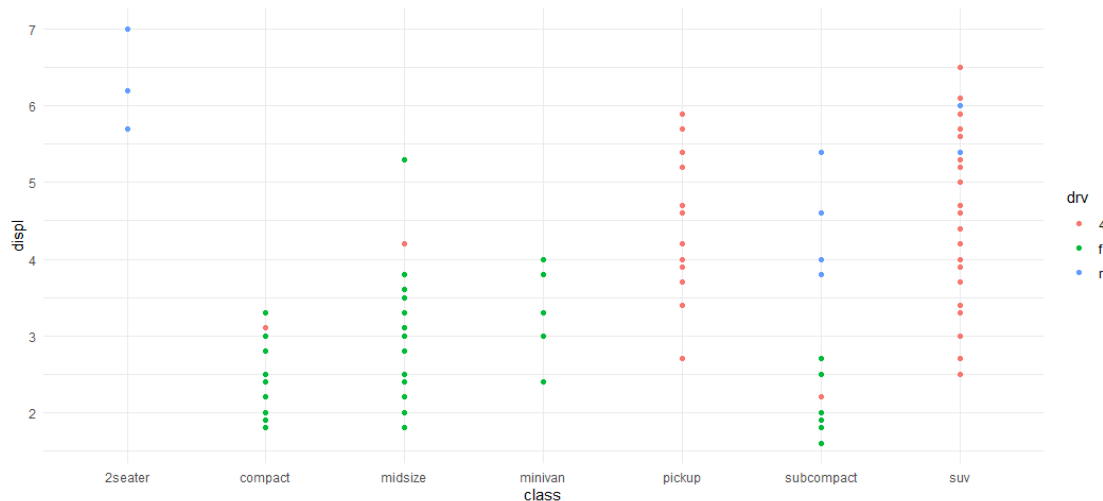
6. How is drive train related to fuel economy? Plot and show

```
ggplot(mpg, aes(drv, cty)) + geom_boxplot(fill = 'steelblue') + theme_minimal()
```



7. How is drive train related to engine size and class? Plot and show

```
ggplot(mpg, aes(class, displ, color=drv)) + geom_point(size = 1.5) + theme_minimal()
```



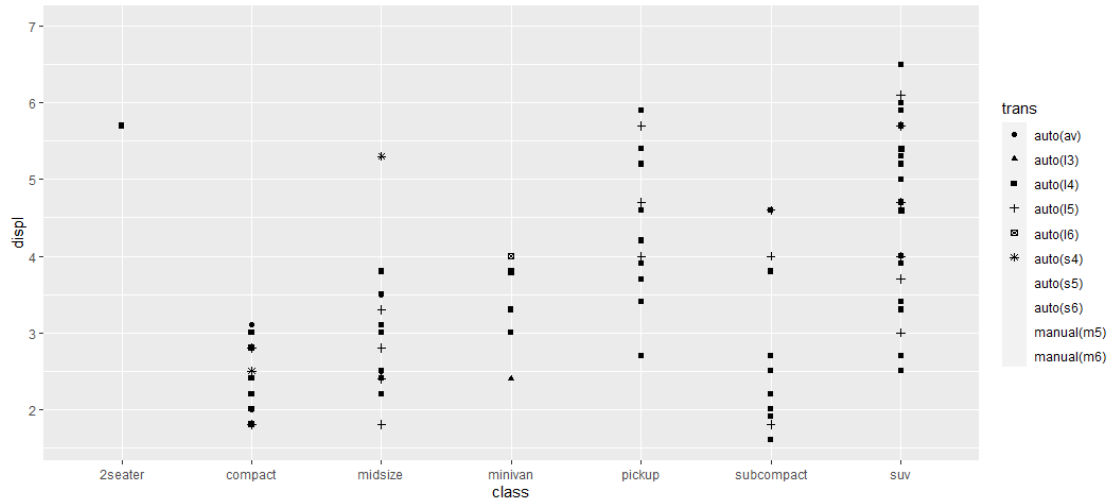
8. What happens if you map trans variable to shape? Why?

```
ggplot(mpg, aes(class, displ, shape=trans)) + geom_point()
```

```
## Warning: The shape palette can deal with a maximum of 6 discrete values be
cause
```

```
## more than 6 becomes difficult to discriminate; you have 10. Consider
## specifying shapes manually if you must have them.
```

```
## Warning: Removed 96 rows containing missing values (geom_point).
```



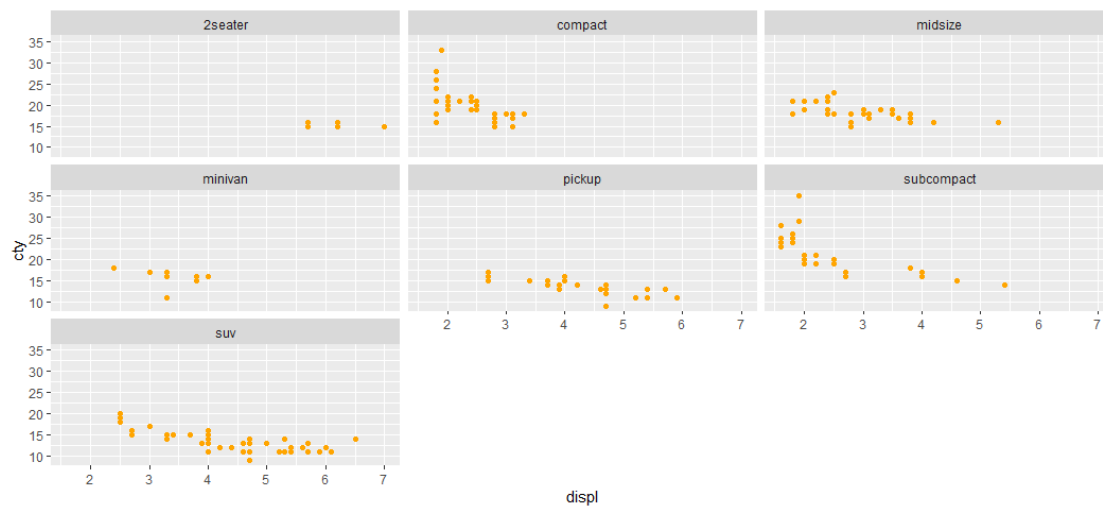
It is incomplete visualization as the trans variable has 10 different categorical records exceeds the limit 6 and the error message says that *“The shape palette can deal with a maximum of 6 discrete values because more than 6 becomes difficult to discriminate; you have 10. Consider specifying shapes manually if you must have them. Removed 96 rows containing missing values (geom\_point)”*.

### Class practice - 3

1. Plot this graph using ggplot()

```
ggplot(cars, aes(displ, cty)) + geom_point(colour="orange") + facet_wrap(~class)
```

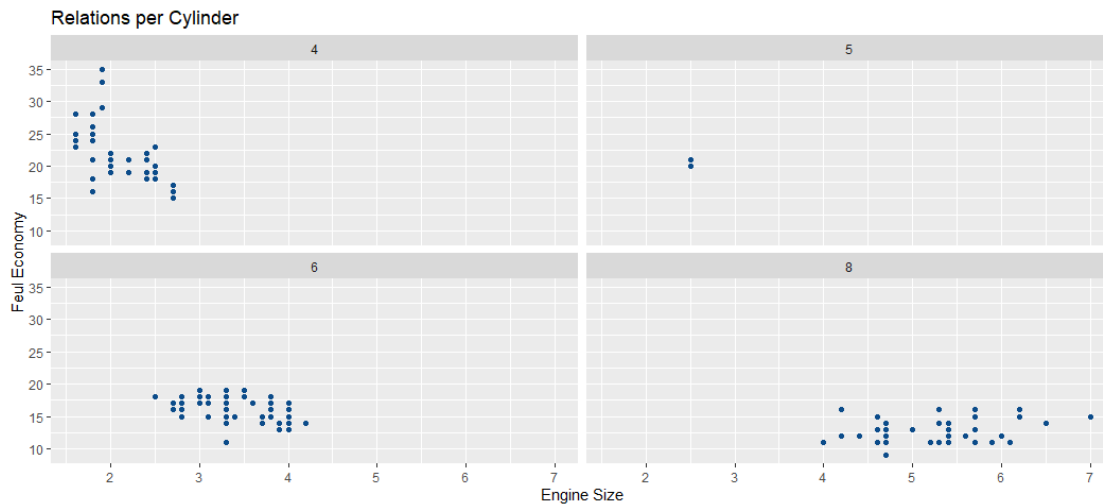
```
ggplot(mpg, aes(displ, cty)) + geom_point(colour="orange") + facet_wrap(~class)
```



2. Use facetting to explore the 3-way relationship between fuel economy, engine size, and number of cylinders.



```
ggplot(mpg, aes(displ, cty)) + geom_point(colour="#0c4c8a") + facet_wrap(~cyl)
) + labs(x = "Engine Size", y = "Feul Economy", title = "Relations per Cylinder")
```



- How does facetting by number of cylinders change your assessment of the relationship between engine size and fuel economy?

***Relationship varies with respect to engine size among different cylinder.as the Cylinders 4 and 6 have Negative correlation nad Cylinder 5 has no relation. Cylinder 8 has a steep positive correlation between displ and cty***

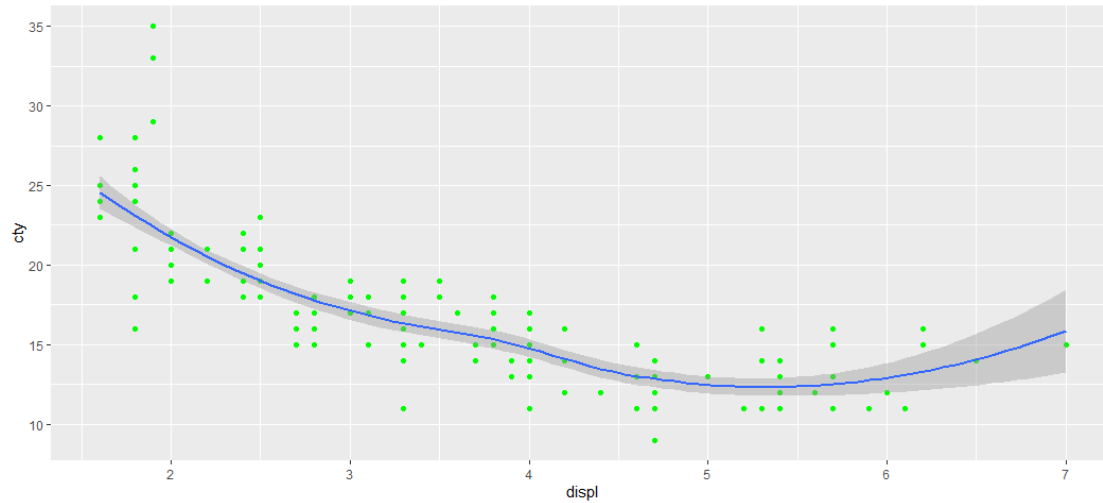
### Class Practice - 4

- Plot the following graph adding smooth()

```
ggplot(cars, aes(x=displ, y = cty)) + geom_point(colour="green") + geom_smooth()
```

```
ggplot(mpg, aes(x=displ, y = cty)) + geom_point(colour="green") + geom_smooth()
```

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

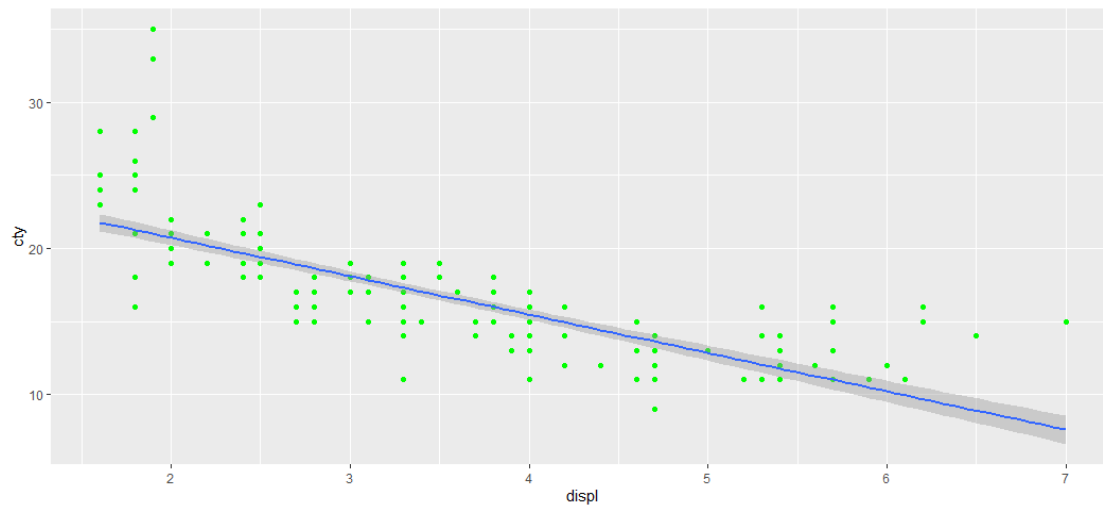


2. Change the method to linear (method='lm')

```
ggplot(cars, aes(x=displ, y = cty)) + geom_point(colour="green") + geom_smooth(method = "lm")
```

```
ggplot(mpg, aes(x=displ, y = cty)) + geom_point(colour="green") + geom_smooth(
  h(method = "lm")
```

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

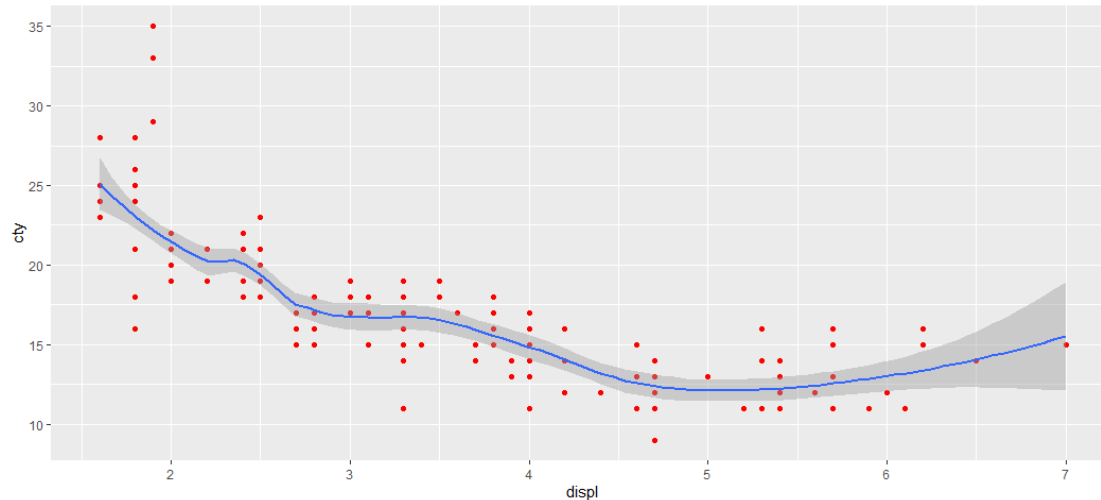


3. Changing the span parameter to control wiggliness of the line:

```
ggplot(cars, aes(x=displ, y = cty)) + geom_point(colour="red") + geom_smooth(span=0.4)
```

```
ggplot(mpg, aes(x=displ, y = cty)) + geom_point(colour="red") + geom_smooth(
  span=0.4)
```

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



4 How is this smooth() function useful in interpreting results?

*The smooth line gives us more clear visualization about the trend of the relation. The grey part in smoothing line is the confidence interval.*

5 Can you use this to explain this to a senior management executive?

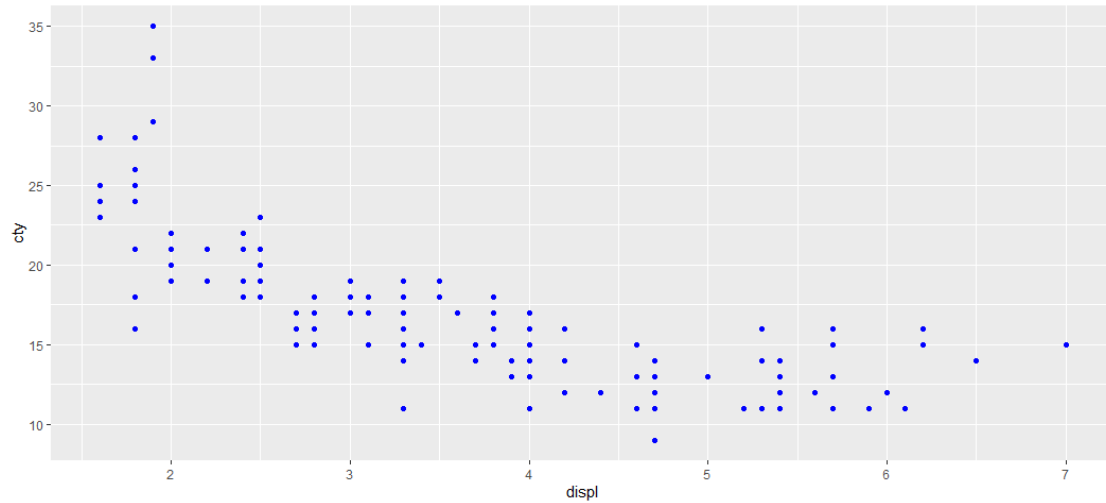
**From the above graph we can notice that the engine size affects the number of gallons per mile in the city roads, while the engine is small it consumes more gallons and decreases until a size of 5 then starts to consume more gallons when the size exceeds 6.**

6 Try changing span parameter, Values range from 0 to 1

```
ggplot(mpg, aes(x=displ, y = cty)) + geom_point(colour="blue") + geom_smooth(
  span=0)

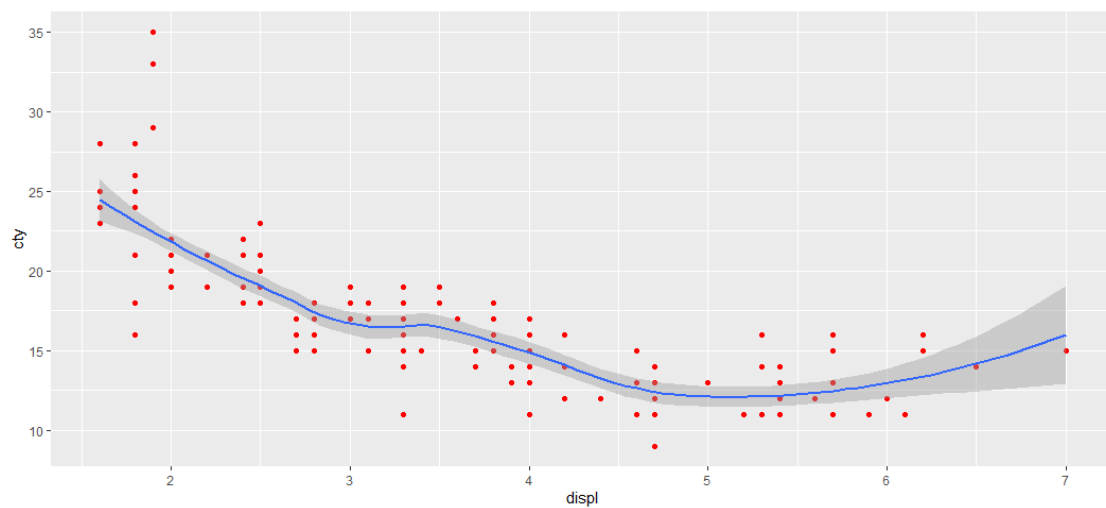
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Computation failed in `stat_smooth()`:
## span is too small
```



```
ggplot(mpg, aes(x=displ, y = cty)) + geom_point(colour="red") + geom_smooth(  
span=0.5)
```

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



```
ggplot(mpg, aes(x=displ, y = cty)) + geom_point(colour="red") + geom_smooth(  
span=1)
```

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

