

# graphing with ggplot2 part1

Code ▾

Hide

```
library(tidyverse)
library(ggplot2)
```

Hide

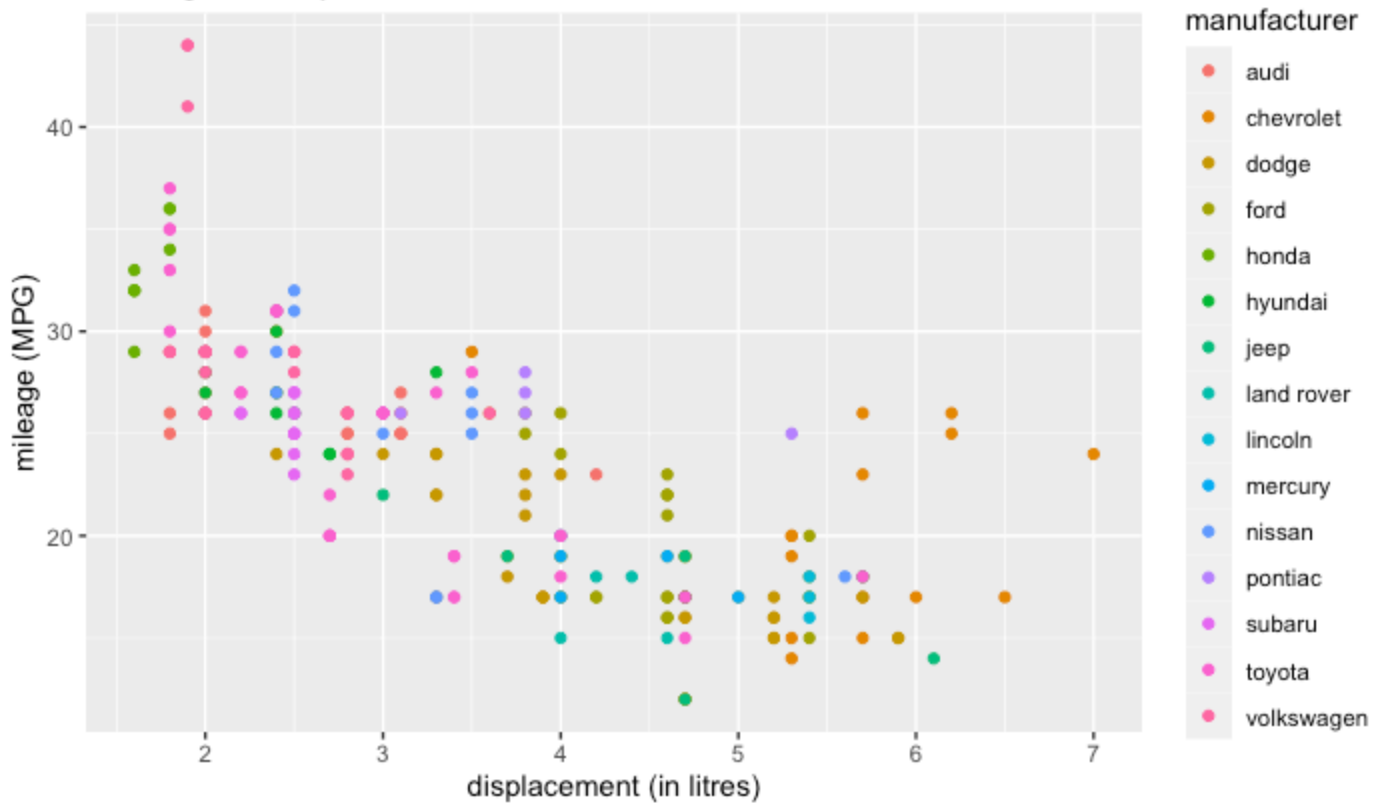
```
data('mpg')
mpg_data=mpg
mpg_data
```

manufacturer <chr>	model <chr>	displ <dbl>	year <int>	cyl <int>	trans <chr>	drv <chr>	cty <int>	h... <int>	fl <chr>					
audi	a4	1.8	1999	4	auto(l5)	f	18	29	p					
audi	a4	1.8	1999	4	manual(m5)	f	21	29	p					
audi	a4	2.0	2008	4	manual(m6)	f	20	31	p					
audi	a4	2.0	2008	4	auto(av)	f	21	30	p					
audi	a4	2.8	1999	6	auto(l5)	f	16	26	p					
audi	a4	2.8	1999	6	manual(m5)	f	18	26	p					
audi	a4	3.1	2008	6	auto(av)	f	18	27	p					
audi	a4 quattro	1.8	1999	4	manual(m5)	4	18	26	p					
audi	a4 quattro	1.8	1999	4	auto(l5)	4	16	25	p					
audi	a4 quattro	2.0	2008	4	manual(m6)	4	20	28	p					
1-10 of 234 rows   1-10 of 11 columns					Previous	1	2	3	4	5	6	...	24	Next

Hide

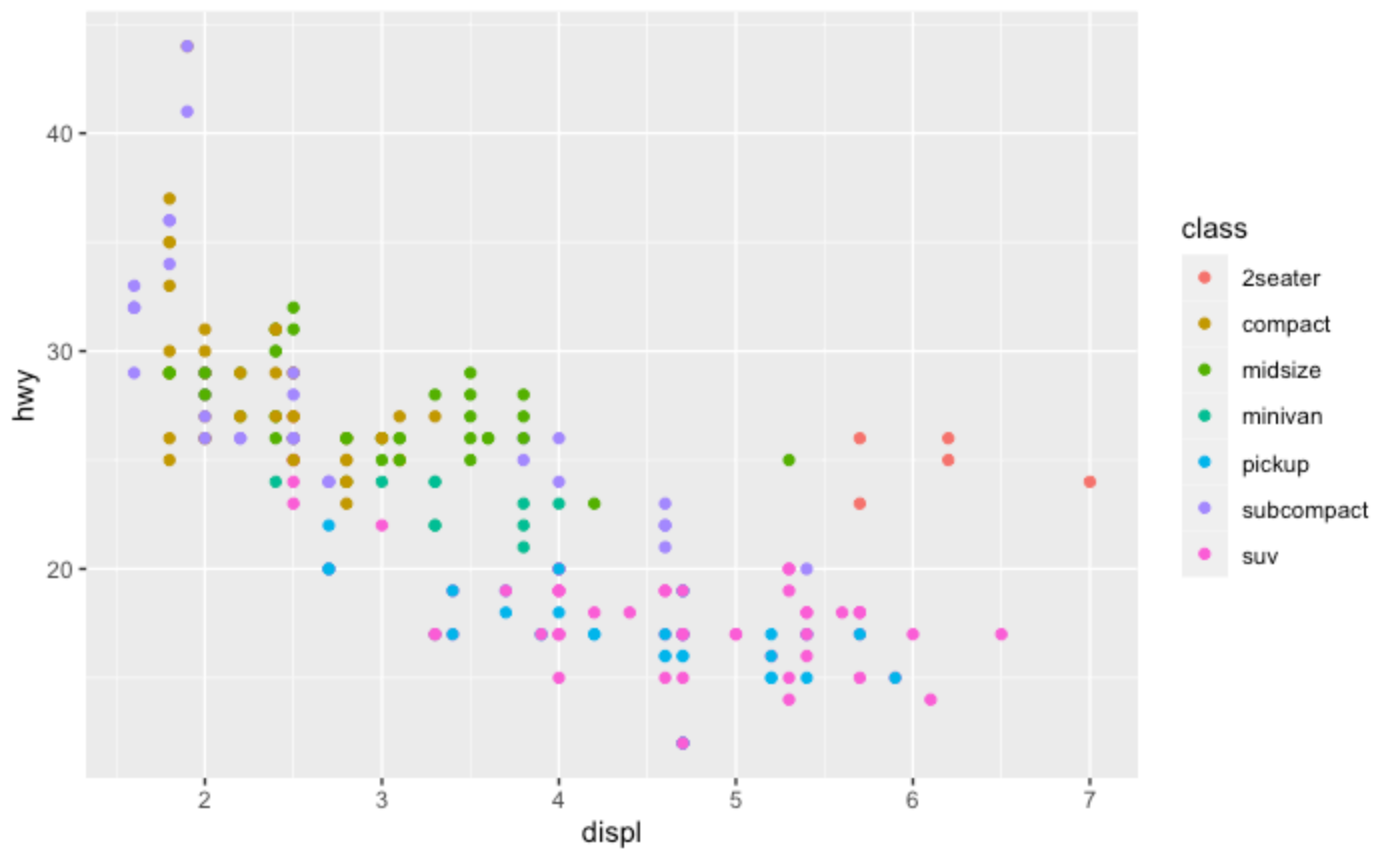
```
p1=ggplot(data=mpg_data,aes(x=displ,y=hwy,color=manufacturer))+geom_point()
p1=p1+labs(x="displacement (in litres)",y="mileage (MPG)",title = " mileage vs displacement")
p1
```

mileage vs displacement



Hide

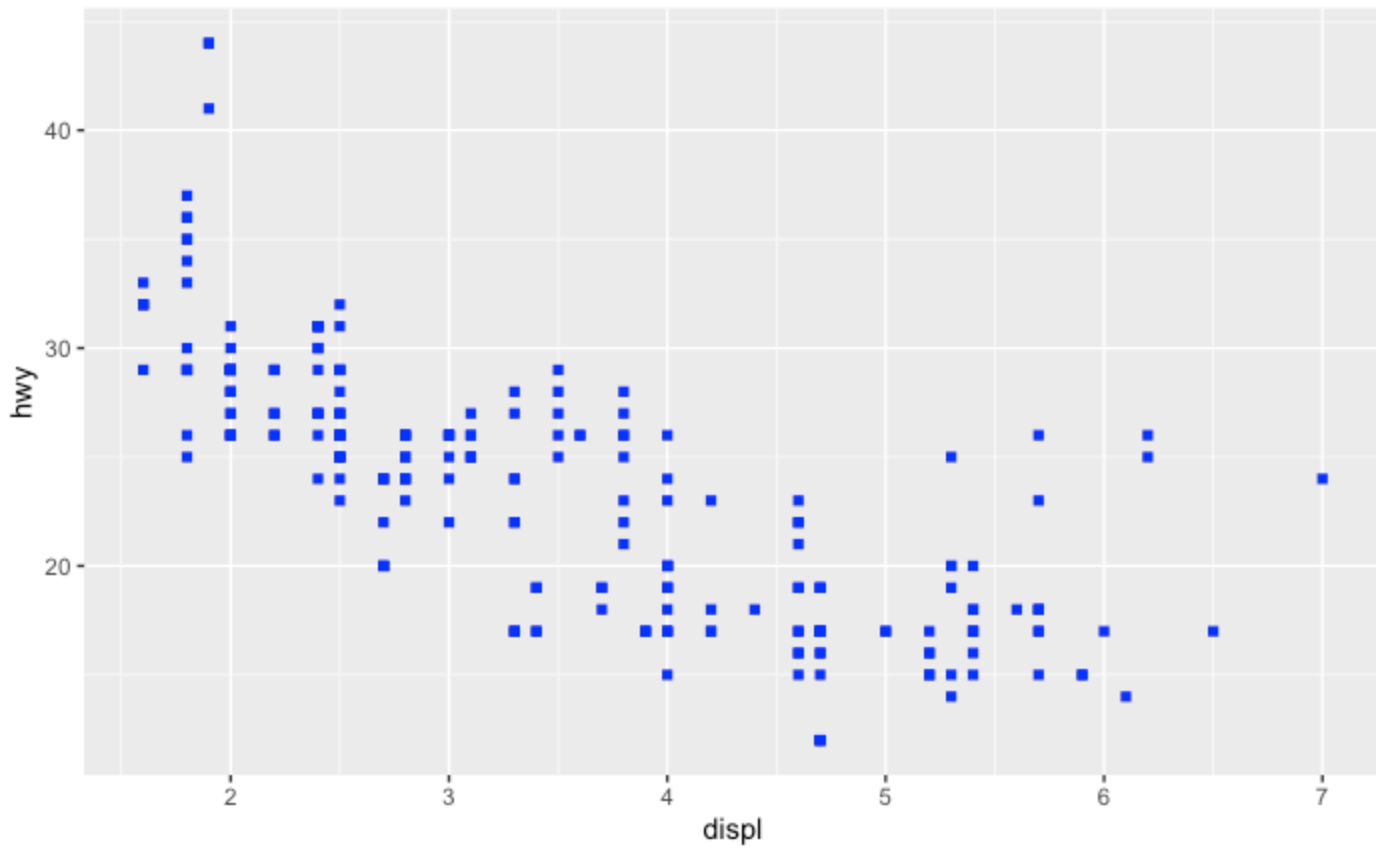
```
p2=ggplot(data=mpg_data)
p2=p2+geom_point(mapping=aes(x=displ,y=hwy,color=class))
p2
```



Hide

```
#this plot makes the point into blue color and of shape square, note that color and shape is outside aes() function
```

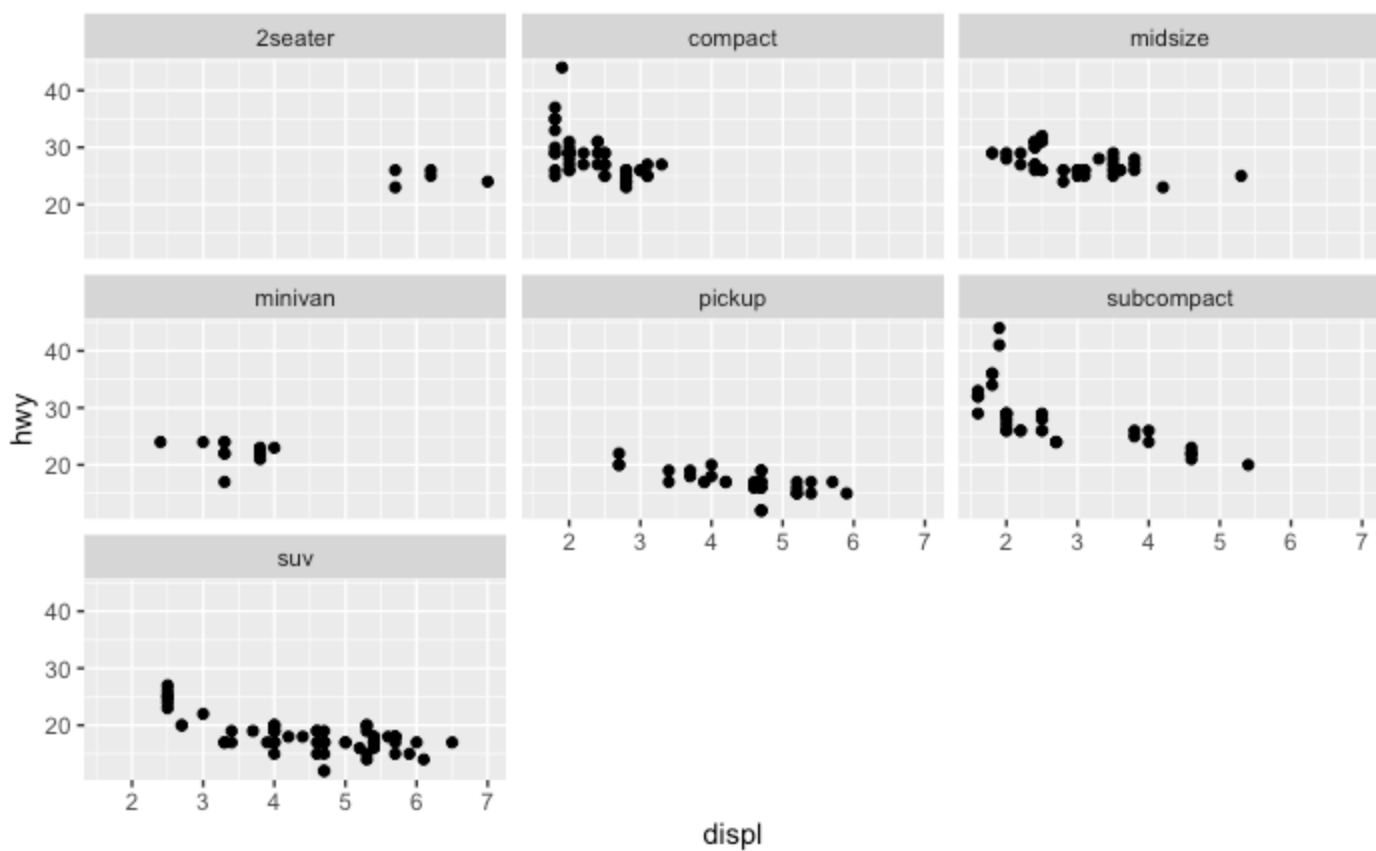
```
p3=ggplot(data=mpg_data) + geom_point(aes(x=displ,y=hwy),color="blue",shape="square")  
p3
```



Hide

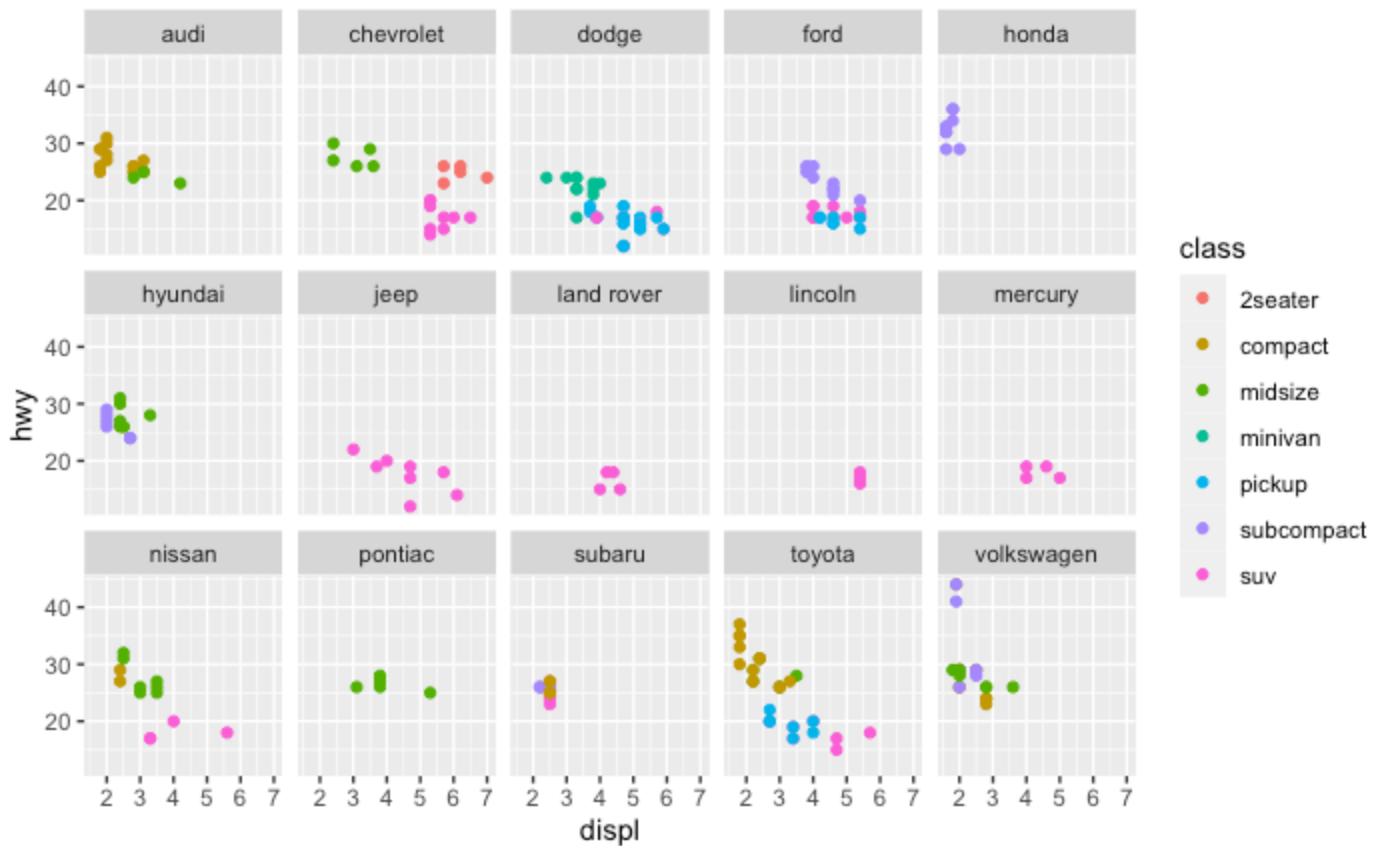
```
#facet is a subplot
```

```
p4=ggplot(data=mpg_data)+geom_point(aes(x=displ,y=hwy))+facet_wrap(~class,nrow=3)  
#~class means we're considering subplots based on class, and nrow means number of rows  
p4
```



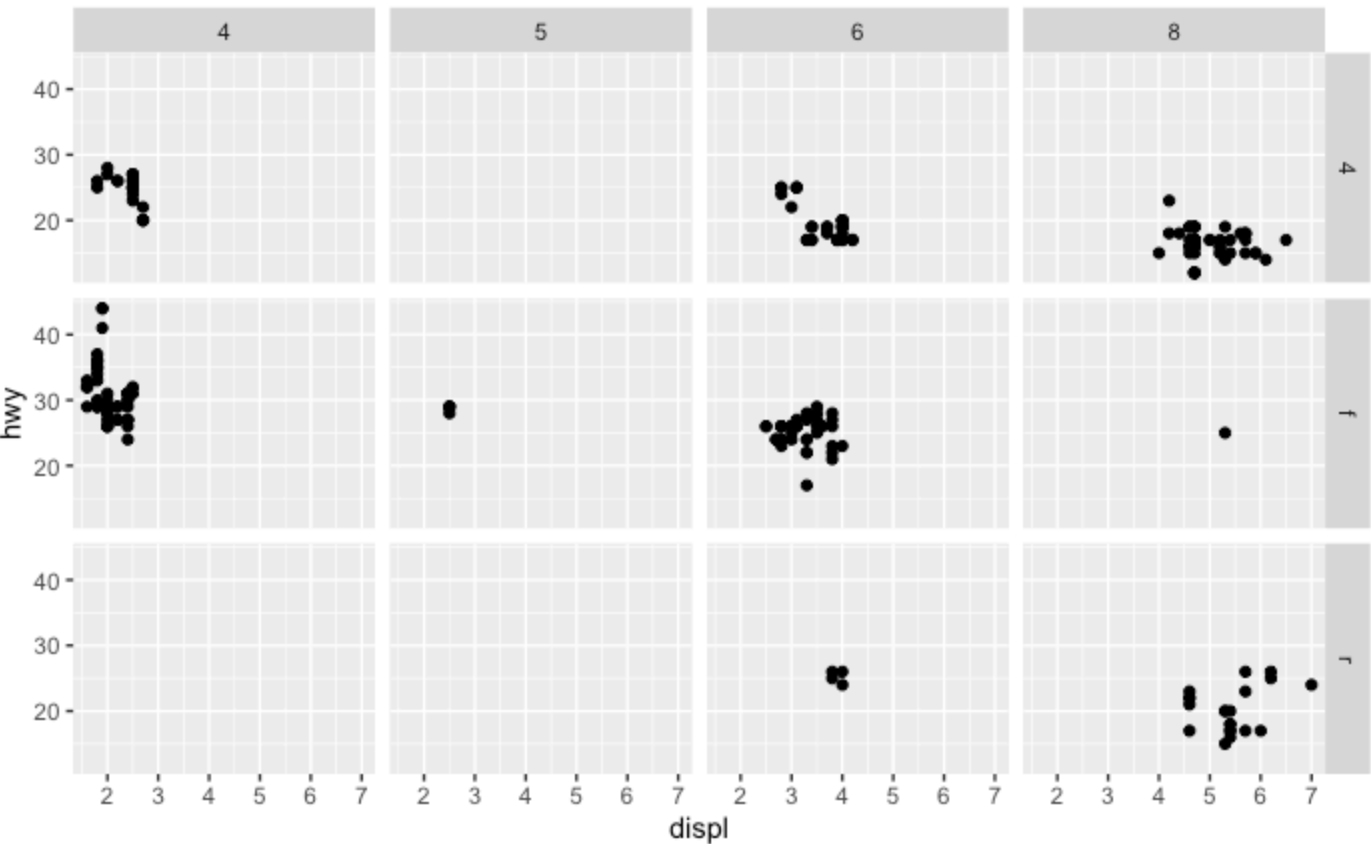
Hide

```
#this is a scatter plot that makes a subplot based on manufacturer and colors based on their class
#as you can see, 2 seaters are mostly from Chevrolet
p5=ggplot(data=mpg_data)+geom_point(aes(x=displ,y=hwy,color=class))+facet_wrap(~manufacturer,nrow=
3)
p5
```



Hide

```
p6=ggplot(data=mpg_data)+geom_point(aes(x=displ,y=hwy))+facet_grid(drv~cyl)
```



Hide

```
#exercise
#to find number of levels in cyl feature
factor(mpg_data$cyl)
```

[illegible]

Hide

```
#how many cars of each cyl type are there
mpg_data%>%count(cyl)
```

	<b>cyl</b> <int>	<b>n</b> <int>
	4	81
	6	79

	cyl <int>	n <int>
	8	70
3 rows		

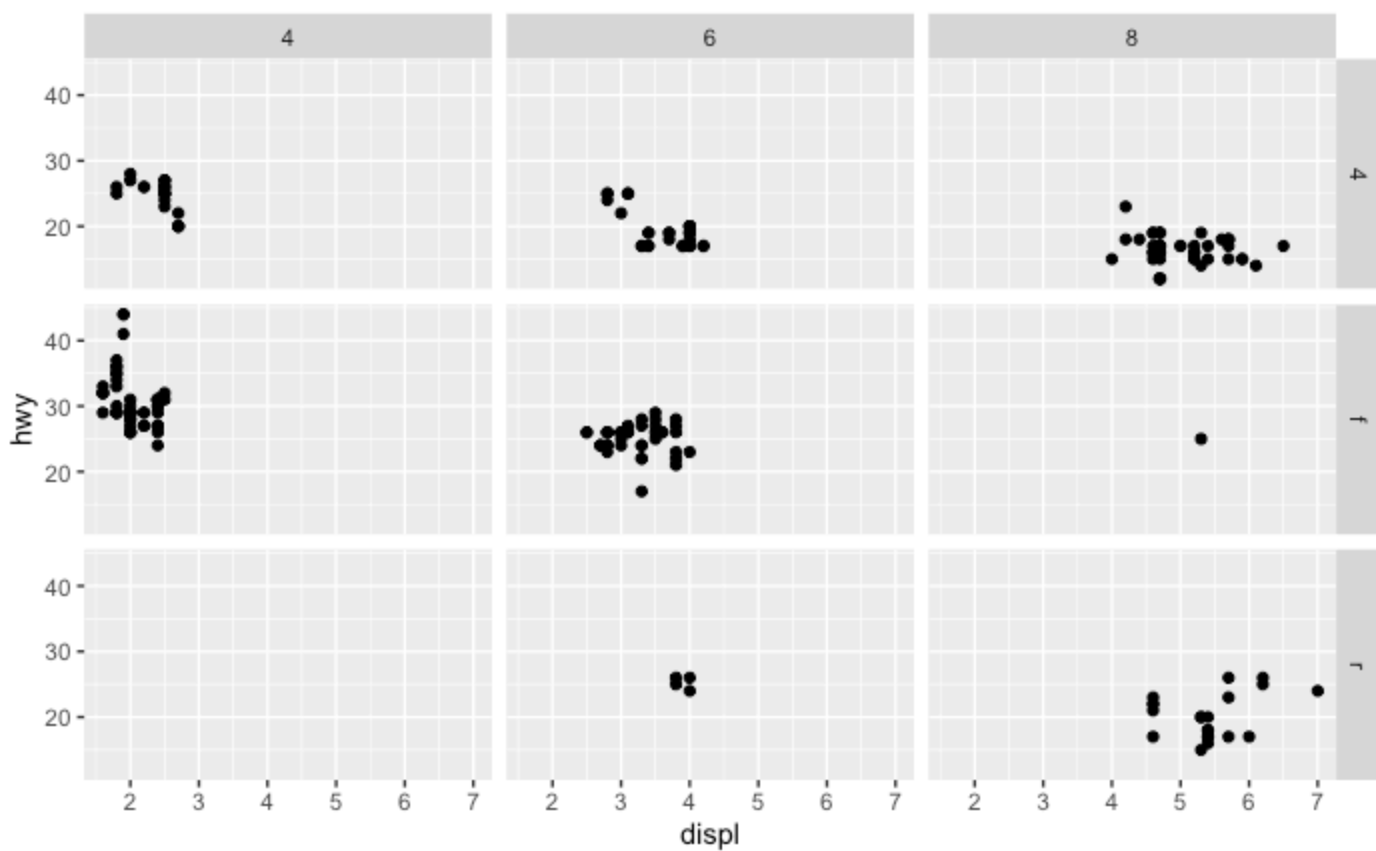
Hide

```
#filter those cars with 5 cyl
mpg_data=mpg_data%>%filter(!(cyl %in% c(5)))
mpg_data
```

manufacturer <chr>	model <chr>	displ <dbl>	year <int>	cyl <int>	trans <chr>	drv <chr>	cty <int>	h... <int>	fl <chr>	
audi	a4	1.8	1999	4	auto(l5)	f	18	29	p	
audi	a4	1.8	1999	4	manual(m5)	f	21	29	p	
audi	a4	2.0	2008	4	manual(m6)	f	20	31	p	
audi	a4	2.0	2008	4	auto(av)	f	21	30	p	
audi	a4	2.8	1999	6	auto(l5)	f	16	26	p	
audi	a4	2.8	1999	6	manual(m5)	f	18	26	p	
audi	a4	3.1	2008	6	auto(av)	f	18	27	p	
audi	a4 quattro	1.8	1999	4	manual(m5)	4	18	26	p	
audi	a4 quattro	1.8	1999	4	auto(l5)	4	16	25	p	
audi	a4 quattro	2.0	2008	4	manual(m6)	4	20	28	p	
1-10 of 230 rows   1-10 of 11 columns										
				Previous	1	2	3	4	5	6 ... 23 Next

Hide

```
p7=ggplot(data=mpg_data)+geom_point(aes(x=displ,y=hwy))+facet_grid(drv~cyl)
p7
```



Hide

#after filtering we can see that there's no car with 5 cylinders. that it has been filtered out

Hide

```
#map color aes to cylinders
p8=ggplot(data=mpg_data)+geom_point(aes(x=displ,y=hwy,color=factor(cyl),shape=class))
#if we use cyl, the color schema wouldn't show us the difference
#so we'll use factor(cyl)
p8
```

