day 6 code

Anil Kumar Yadav

10/5/2019

With reference from Hadley wickham book Loading required libraries

```
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

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

```
library(ggplot2)
library(data.table)
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
## The following object is masked from 'package:purrr':
##
##     transpose
```

Understading the data (data validation)

```
data("mpg")
names(mpg)
```

```
##  [1] "manufacturer" "model"        "displ"        "year"
##  [5] "cyl"          "trans"        "drv"          "cty"
##  [9] "hwy"          "fl"           "class"
```

```
dim(mpg)
```

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

```r
str(mpg)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    234 obs. of  11 variables:
##  $ manufacturer: chr  "audi" "audi" "audi" "audi" ...
##  $ model       : chr  "a4" "a4" "a4" "a4" ...
##  $ displ       : num  1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
##  $ year        : int  1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
##  $ cyl         : int  4 4 4 4 6 6 6 4 4 4 ...
##  $ trans       : chr  "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...
##  $ drv         : chr  "f" "f" "f" "f" ...
##  $ cty         : int  18 21 20 21 16 18 18 18 16 20 ...
##  $ hwy         : int  29 29 31 30 26 26 27 26 25 28 ...
##  $ fl          : chr  "p" "p" "p" "p" ...
##  $ class       : chr  "compact" "compact" "compact" "compact" ...
```

```r
head(mpg)
```

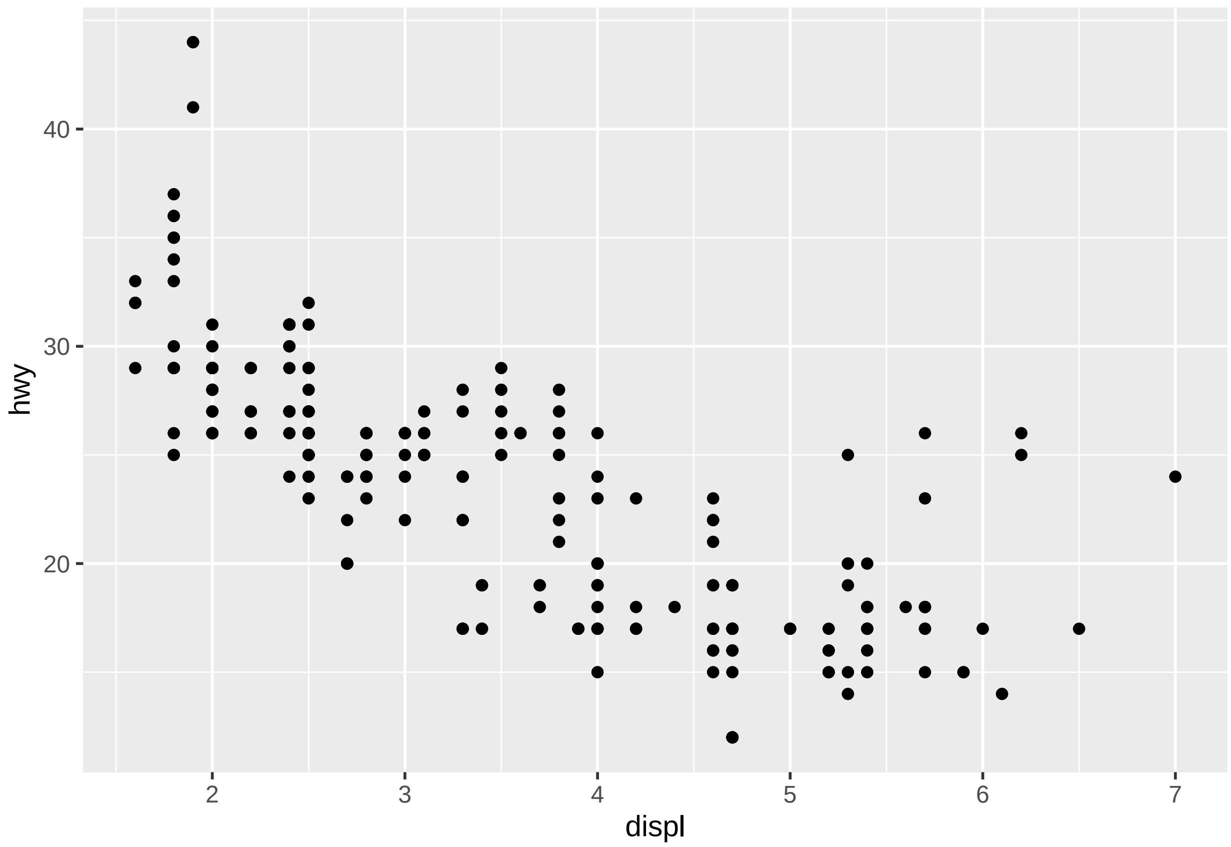
```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans  drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <int> <int> <chr>  <chr> <int> <int> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(~ f        18    29 p     comp~
## 2 audi         a4      1.8  1999  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

```r
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
```

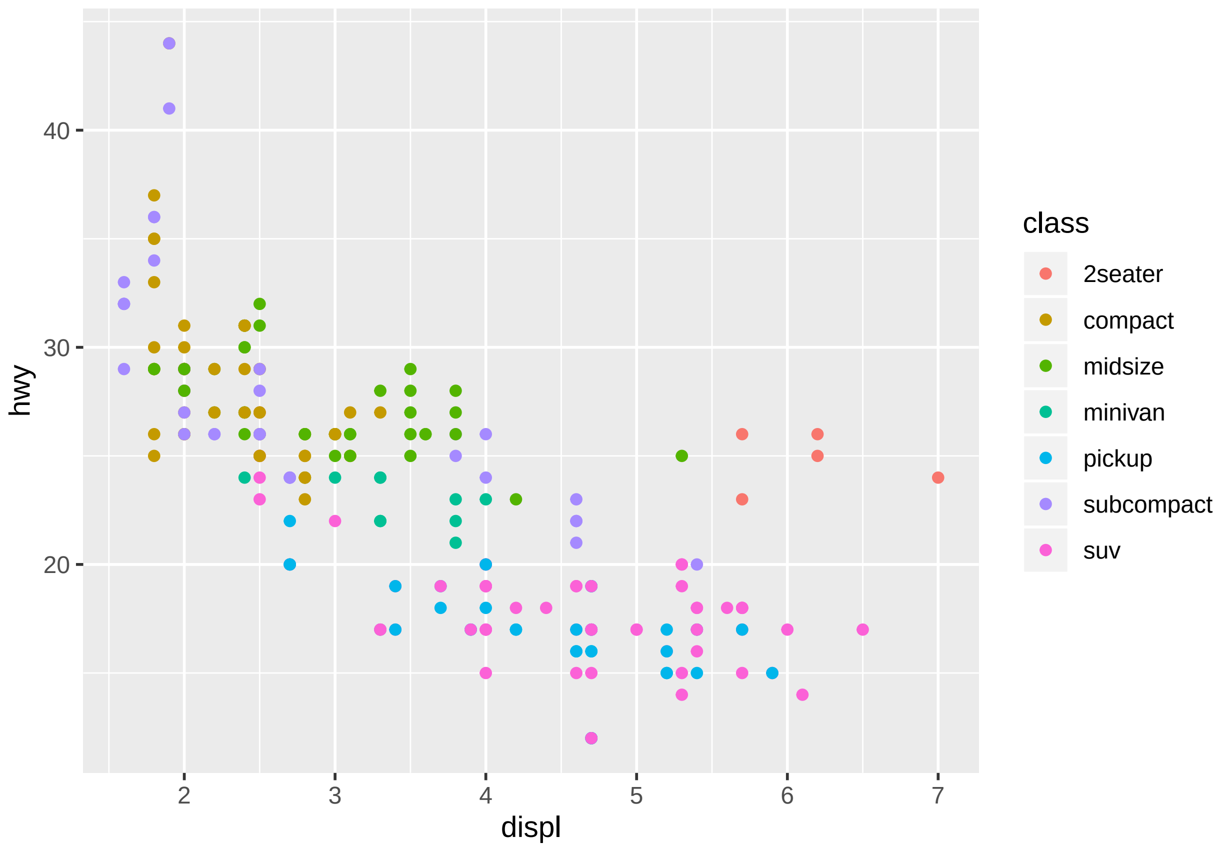
Plotting

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



gglpot with colours

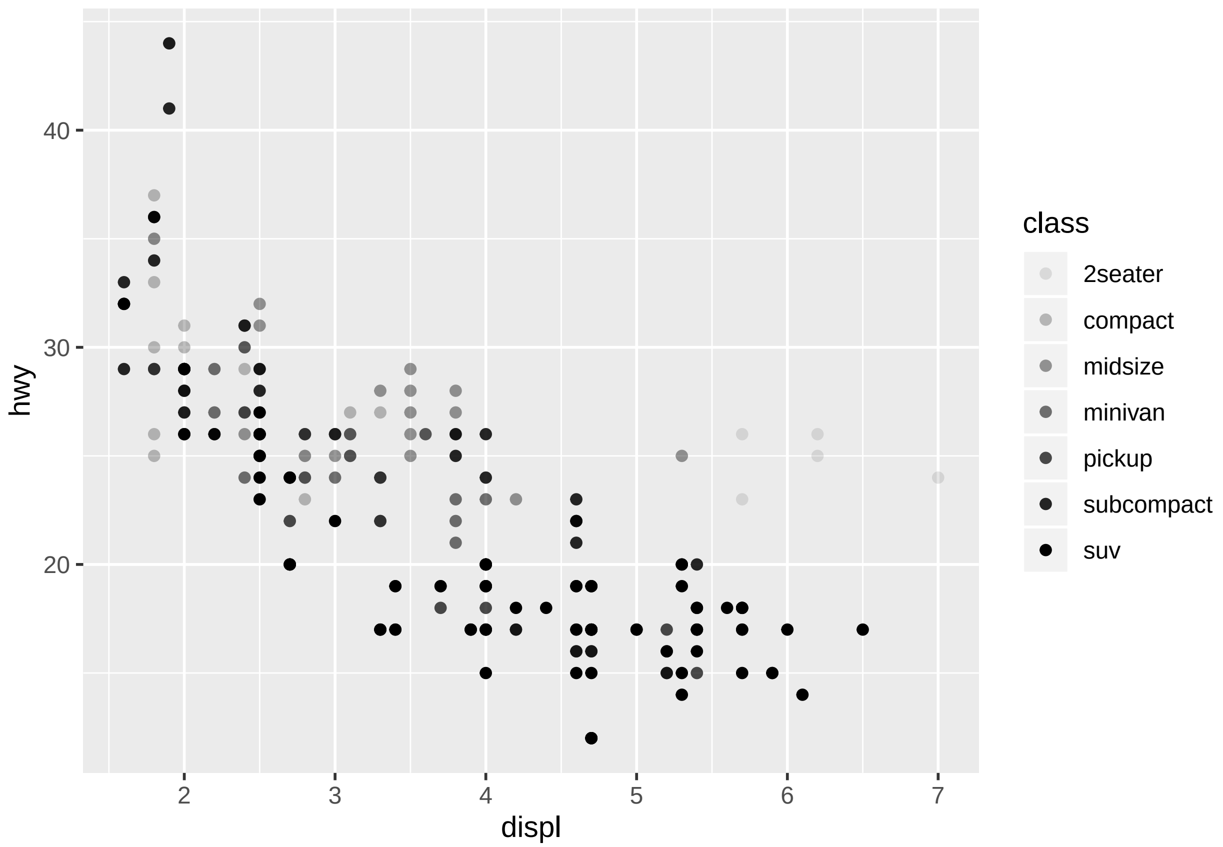
```r
ggplot(data=mpg)+ geom_point(mapping = aes(x=displ, y=hwy, color=class))
```



Alpha aesthetic gives transperecy of points

```
ggplot(data=mpg) + geom_point(mapping = aes(x = displ,y = hwy, alpha = class))
```

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

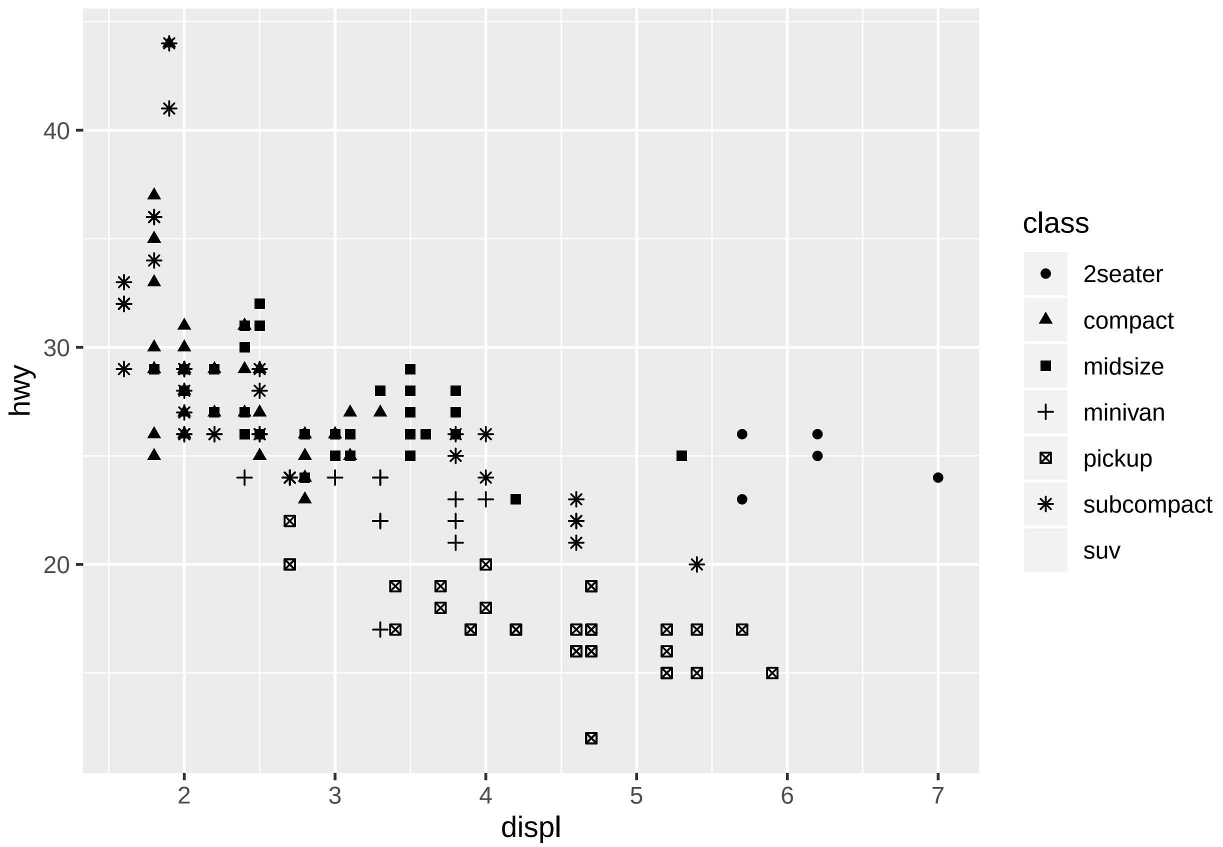


Shape aesthetic gives control of the shapes

```
ggplot(data=mpg) + geom_point(mapping = aes(x=displ,y=hwy, shape = class))
```

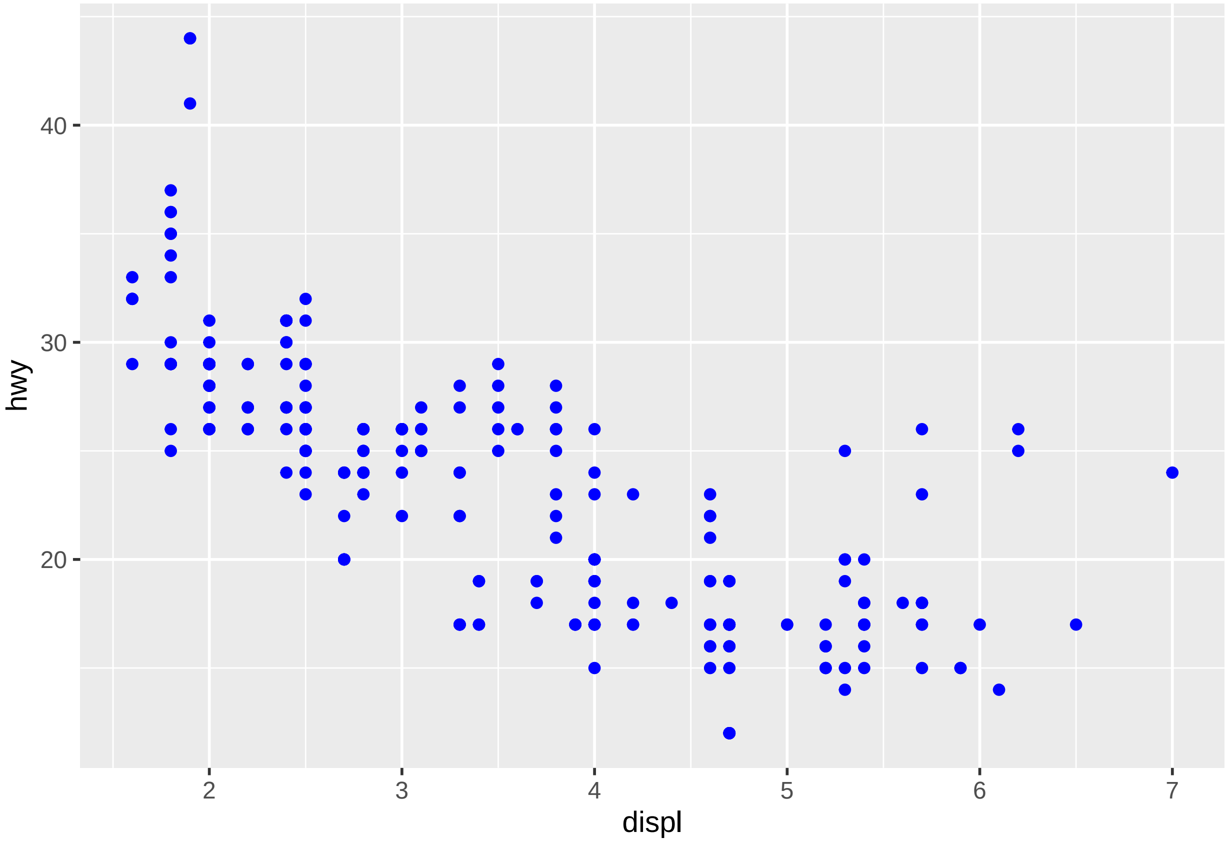
```
## Warning: The shape palette can deal with a maximum of 6 discrete values
## because more than 6 becomes difficult to discriminate; you have 7.
## Consider specifying shapes manually if you must have them.
```

```
## Warning: Removed 62 rows containing missing values (geom_point).
```



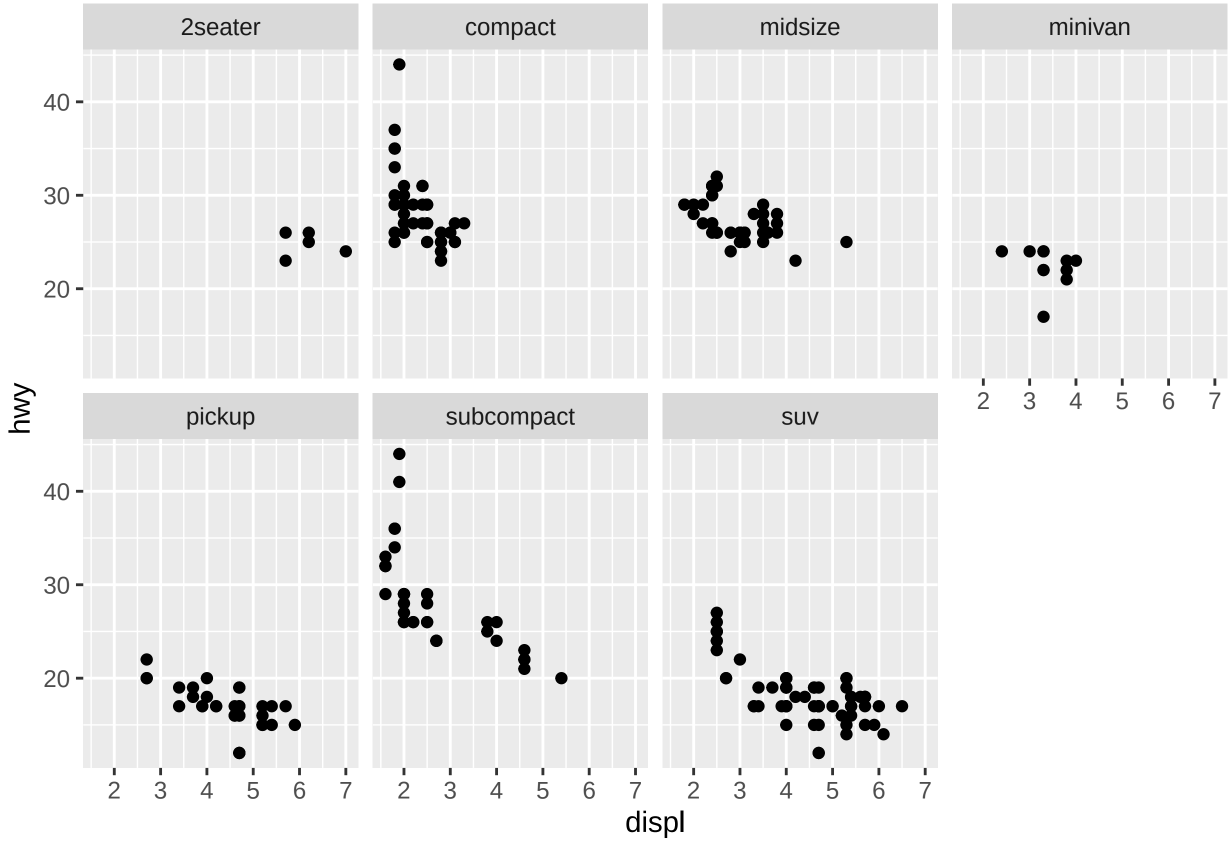
Single coloured

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



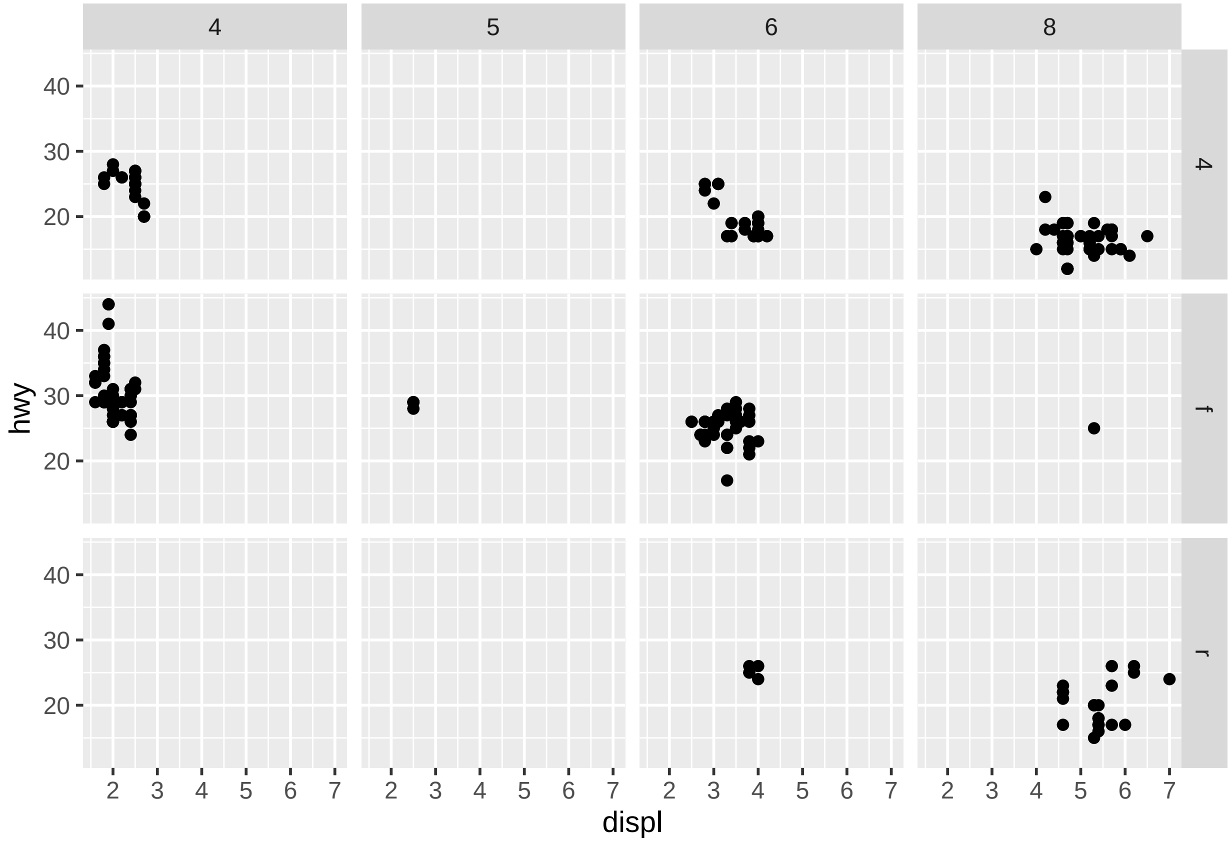
To display the subset of data

```
ggplot(data=mpg)+ geom_point(mapping = aes(x=displ, y=hwy))+facet_wrap(~class, nrow=2)
```



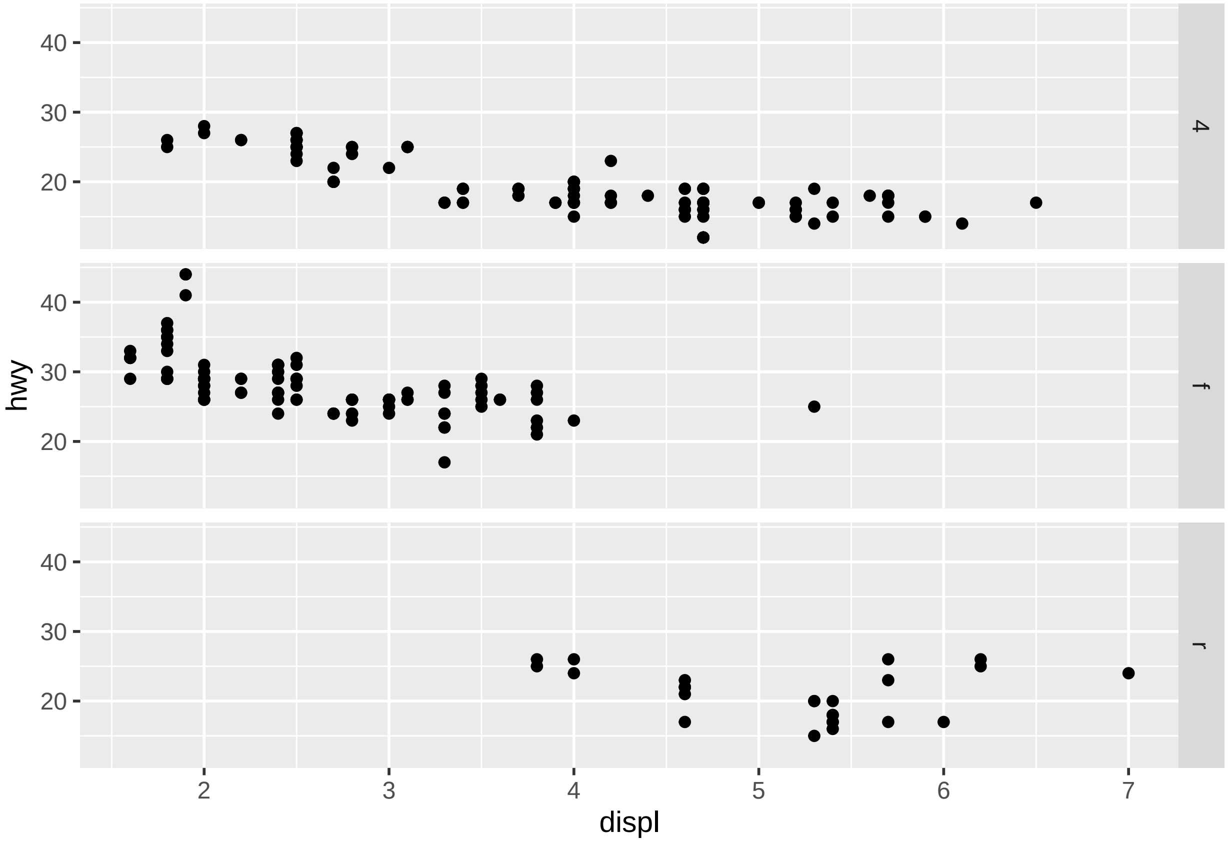
To add another set of variable on he same plotting

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

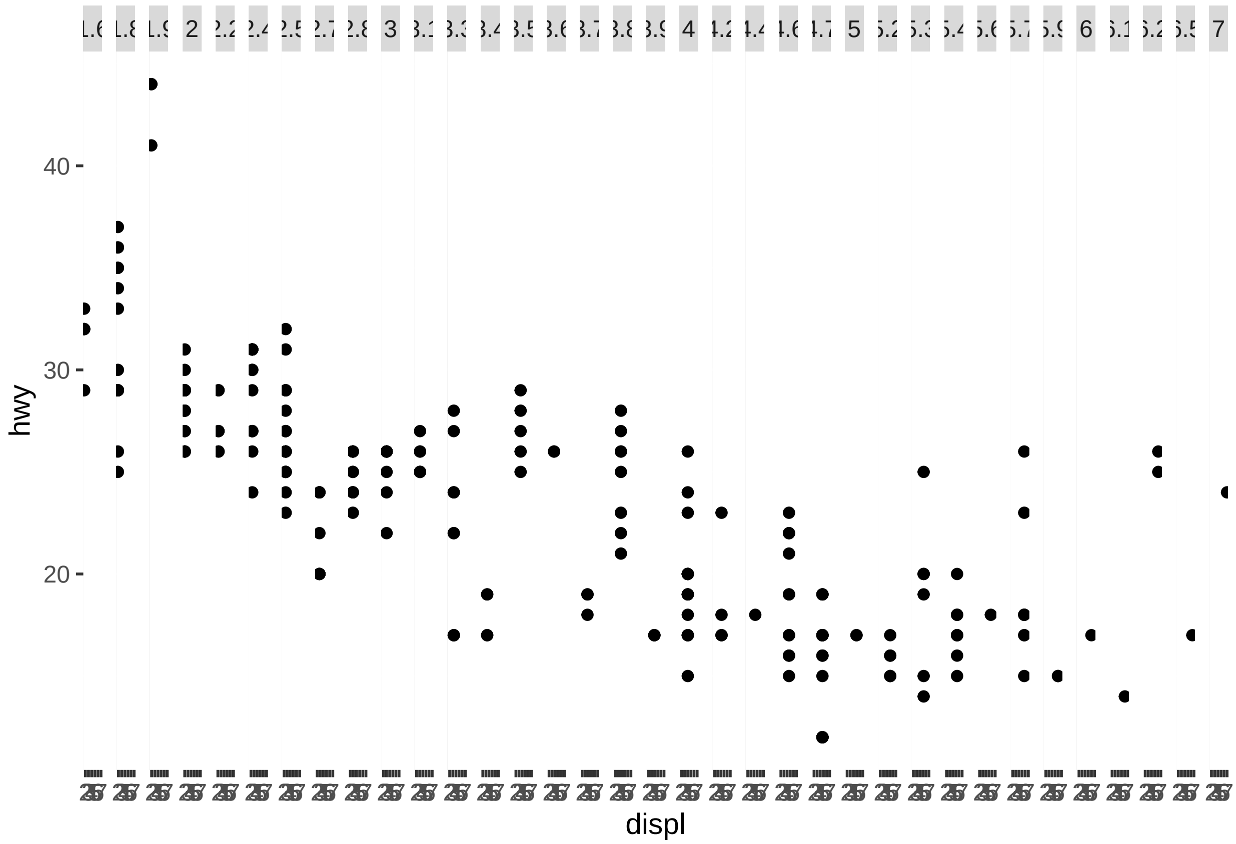


experminting exercises

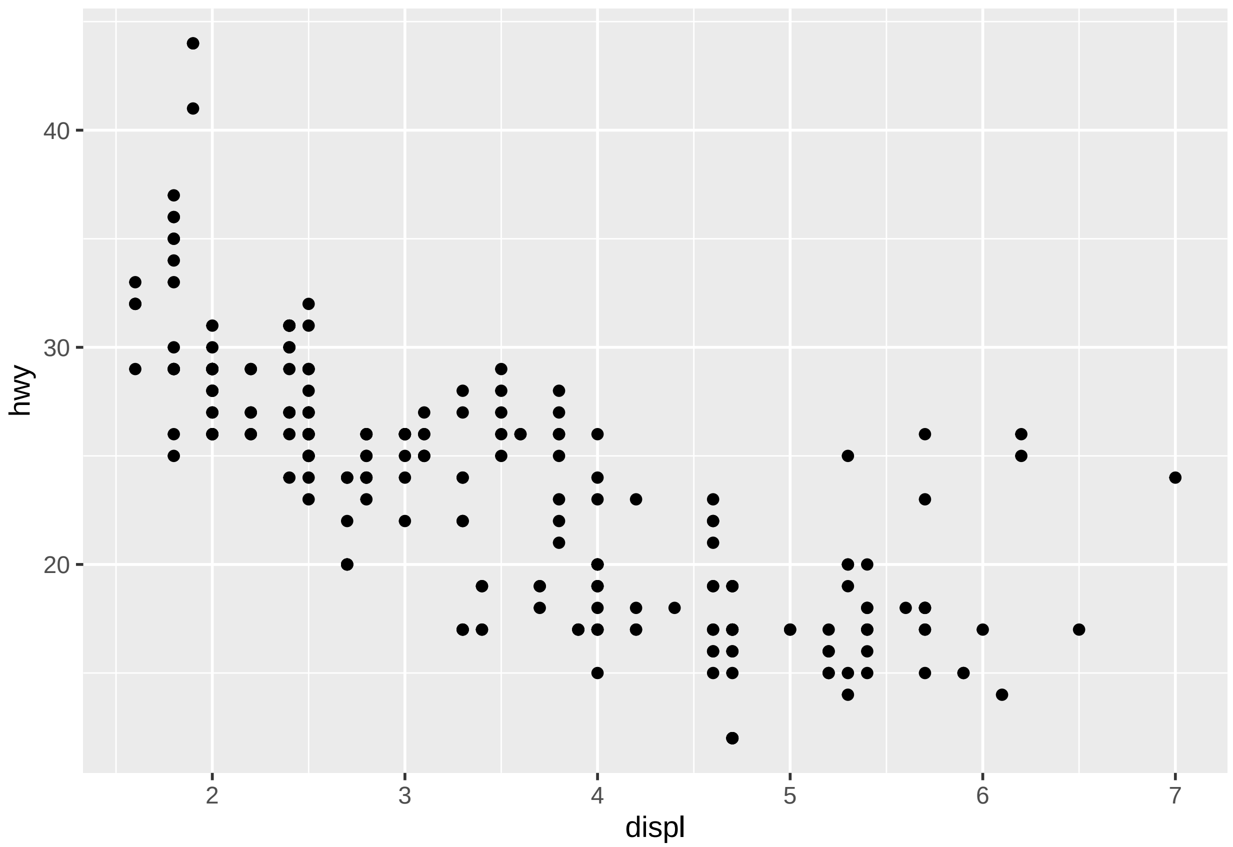
```
ggplot(data=mpg)+ geom_point(mapping = aes(x=displ, y=hwy))+ facet_grid(drv~.)
```



```
ggplot(data=mpg)+ geom_point(mapping = aes(x=displ, y=hwy))+ facet_grid(.~displ)
```

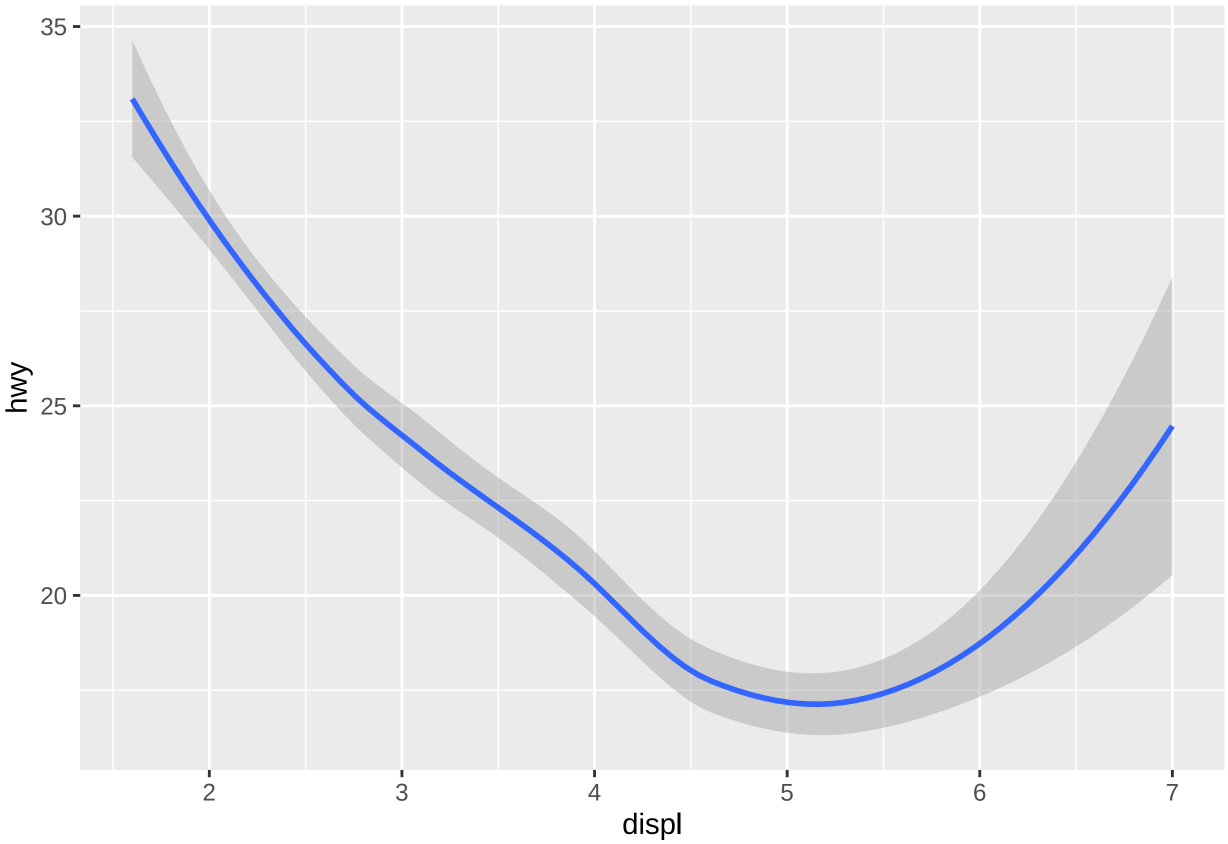


```r
ggplot(data=mpg)+geom_point(mapping = aes(x=displ, y=hwy))
```



```
ggplot(data=mpg)+geom_smooth(mapping = aes(x=displ, y=hwy))
```

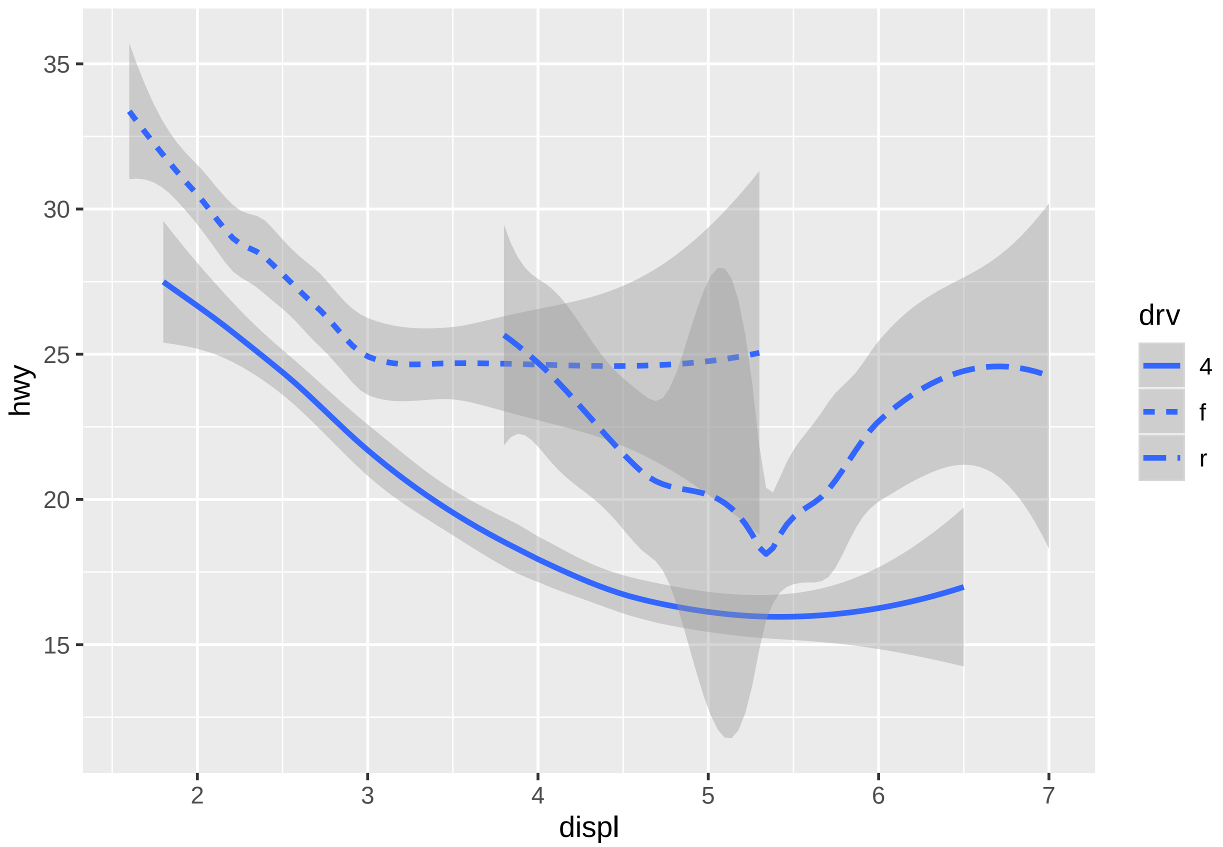
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Using smooth

```
ggplot(data=mpg) + geom_smooth(mapping = aes(x=displ, y=hwy,linetype =drv))
```

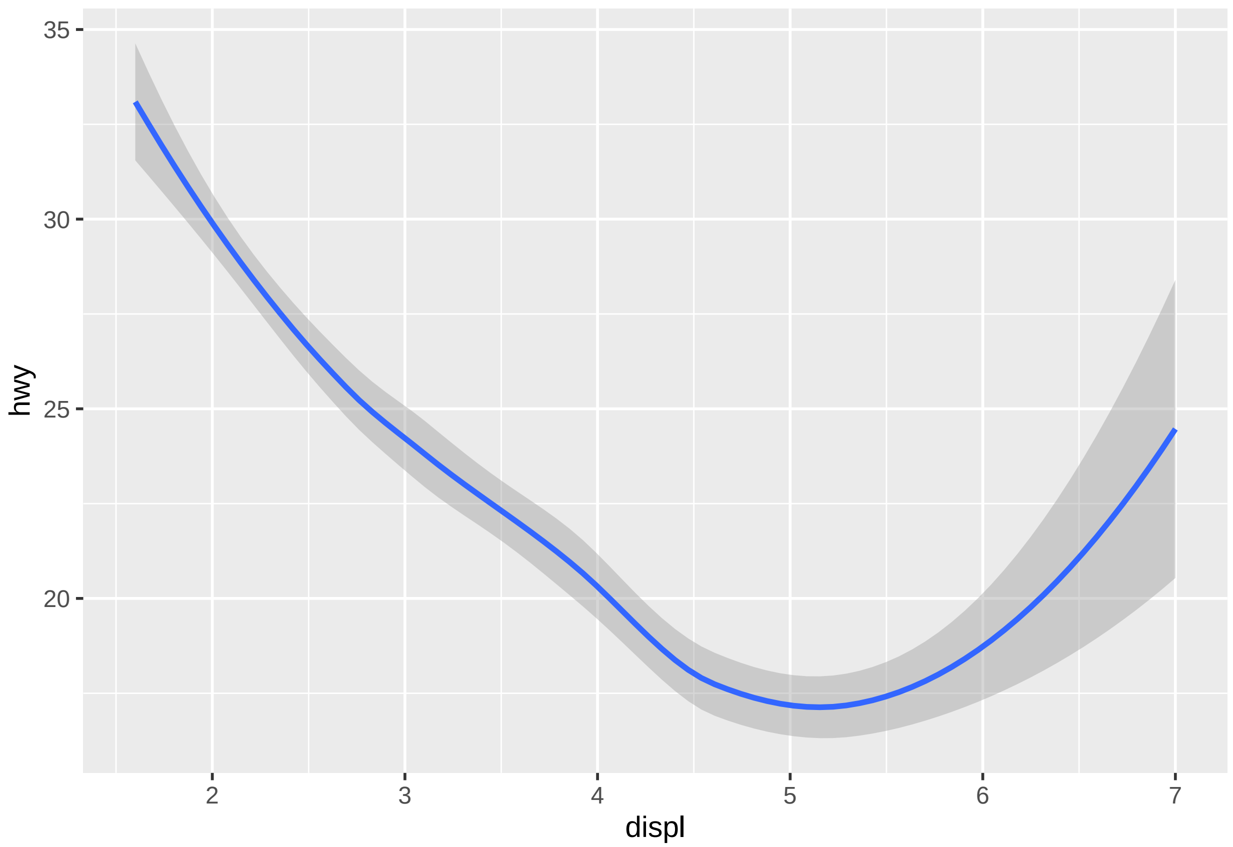
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



All about smooth

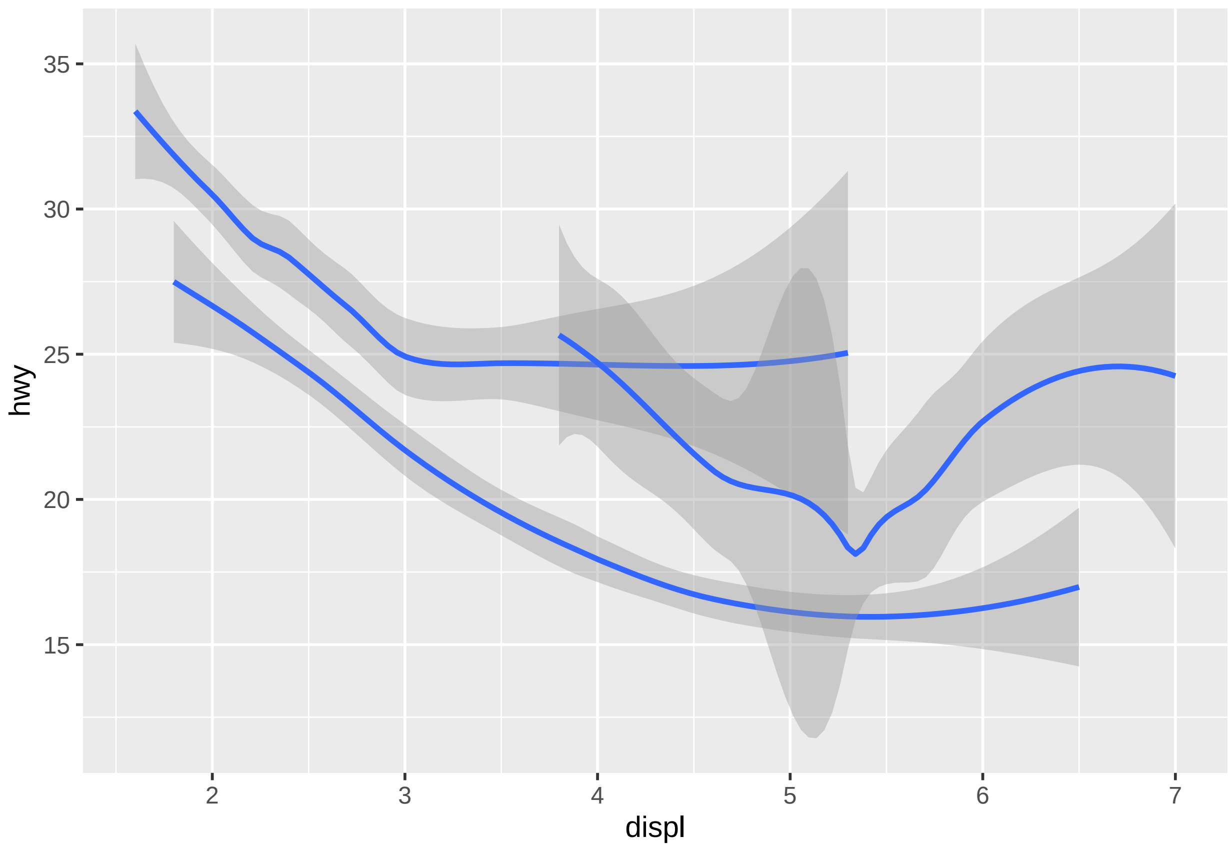
```
ggplot(data=mpg)+geom_smooth(mapping = aes(x=displ, y=hwy))
```

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



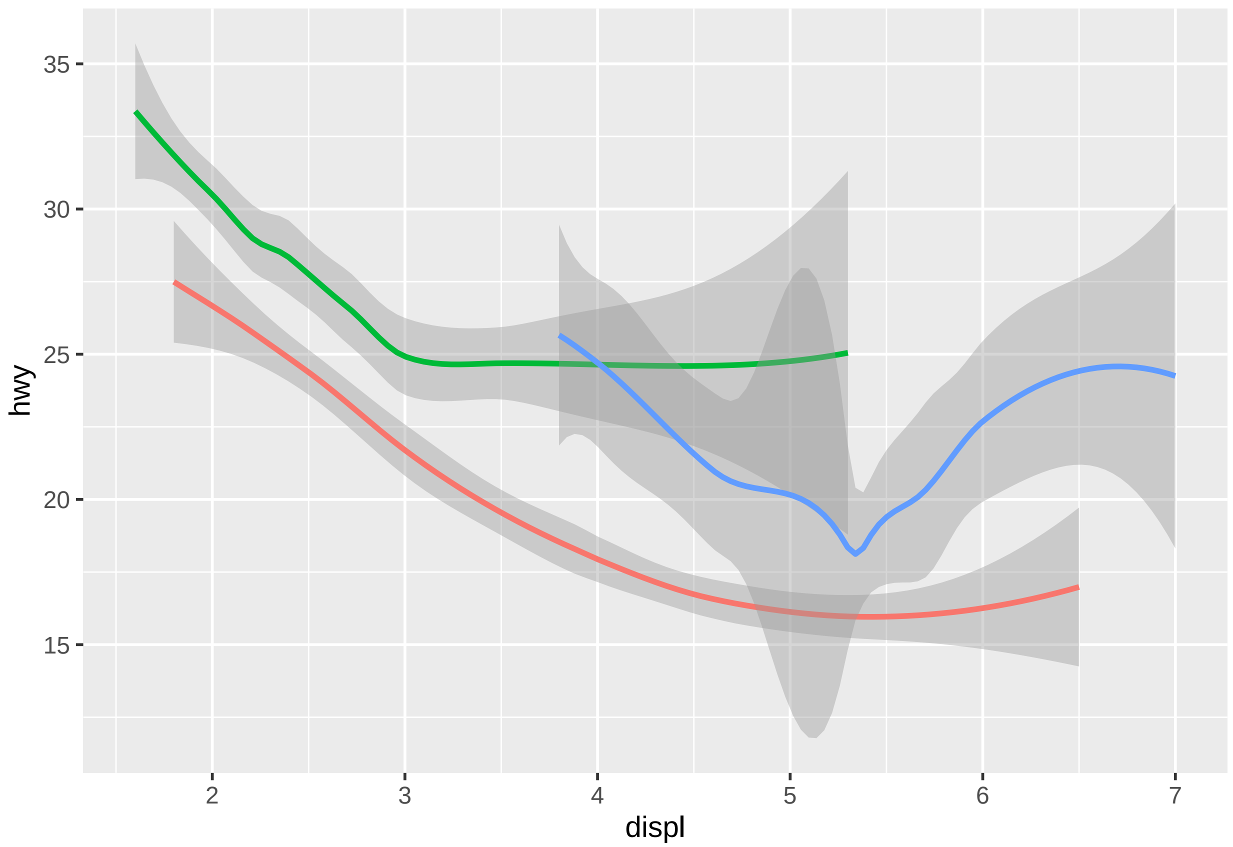
```r
ggplot(data=mpg)+geom_smooth(mapping = aes(x=displ, y=hwy, group=drv))
```

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



```
ggplot(data=mpg)+geom_smooth(mapping = aes(x=displ, y=hwy, color=drv), show.legend=FALSE)
```

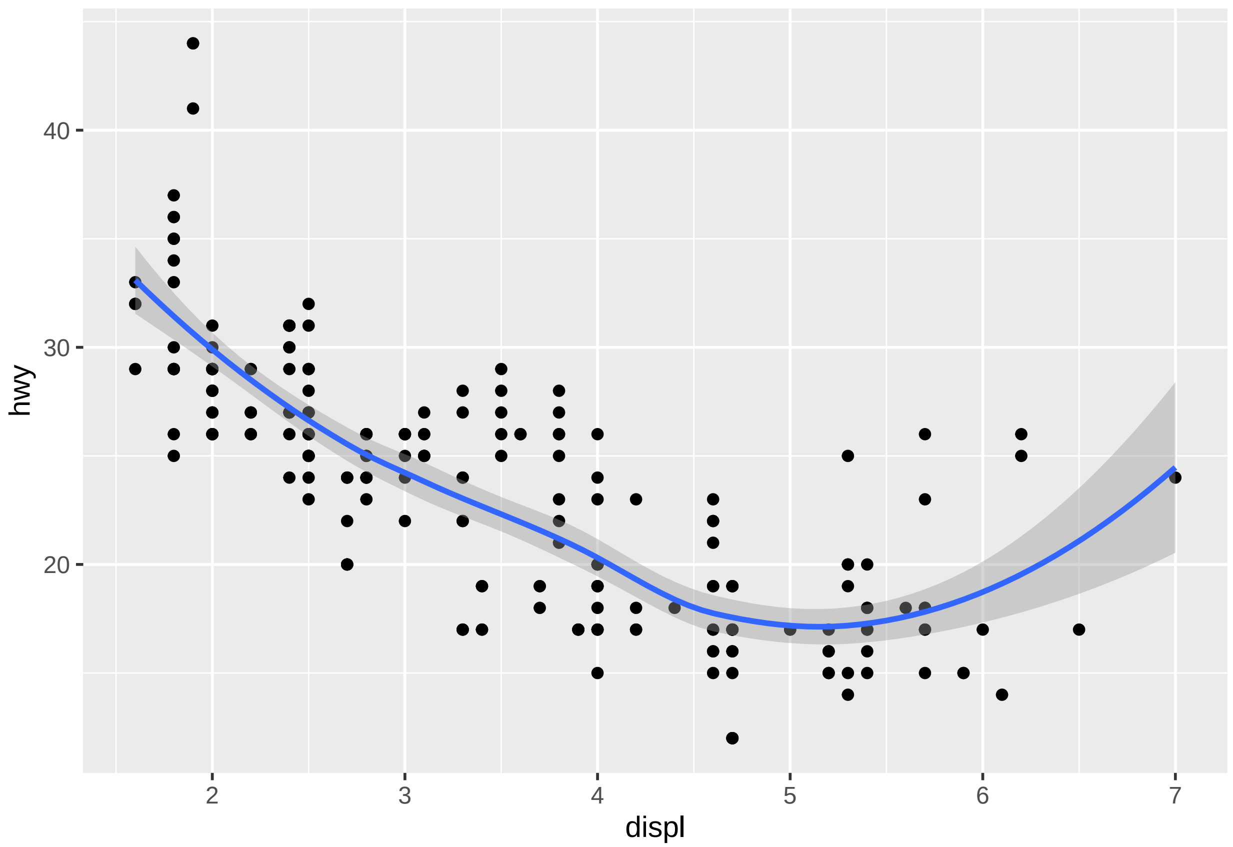
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Multiple geoms

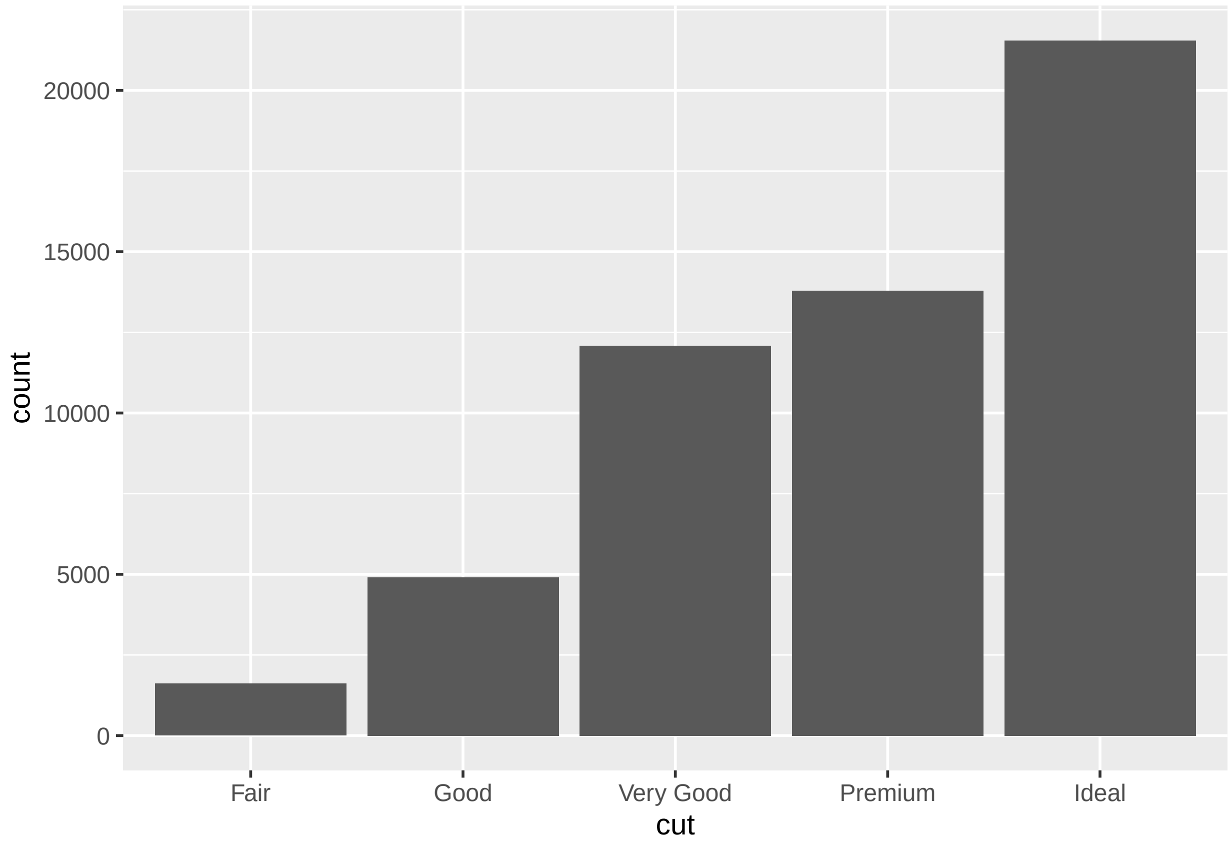
```
ggplot(data=mpg)+geom_point(mapping = aes(x=displ, y=hwy)) + geom_smooth(mapping=aes(x=displ, y=hwy))
```

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



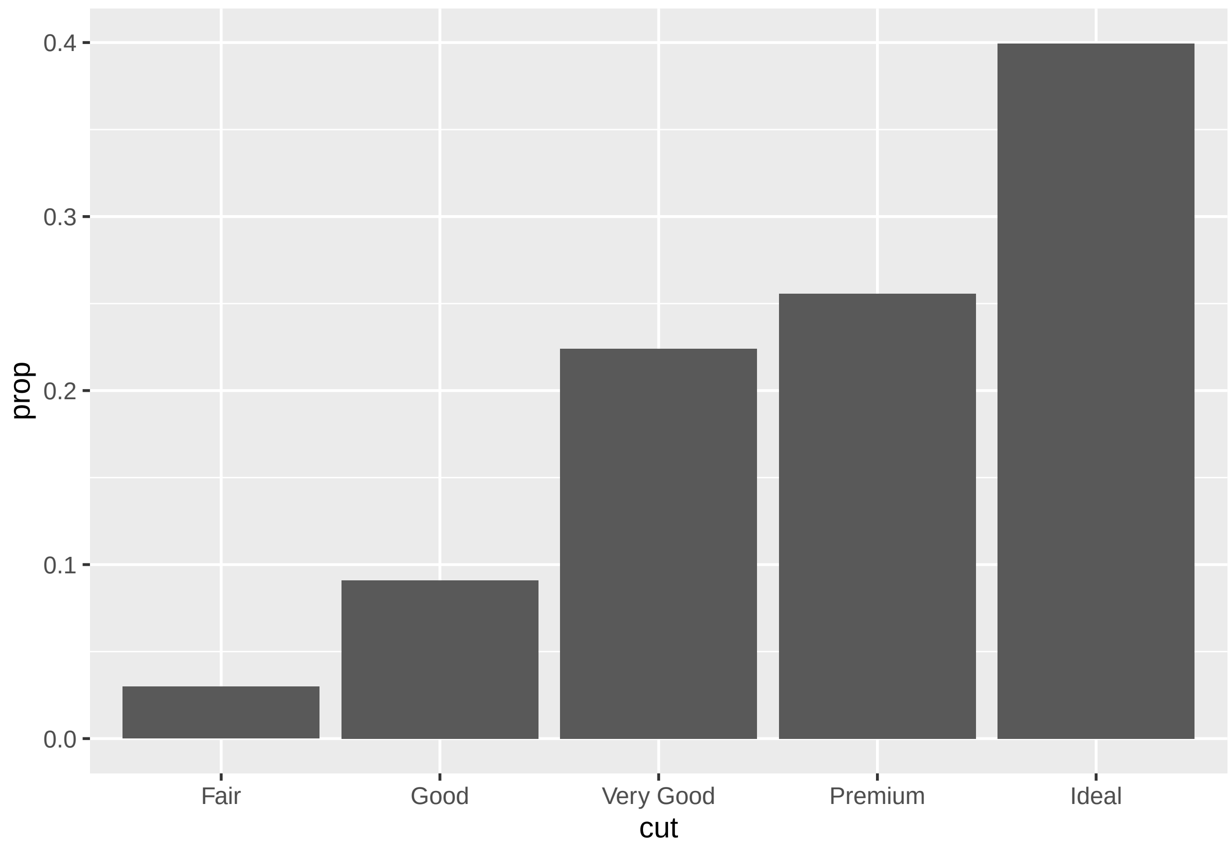
Using Diamond dataset

```
ggplot(data = diamonds)+geom_bar(mapping = aes(x =cut))
```



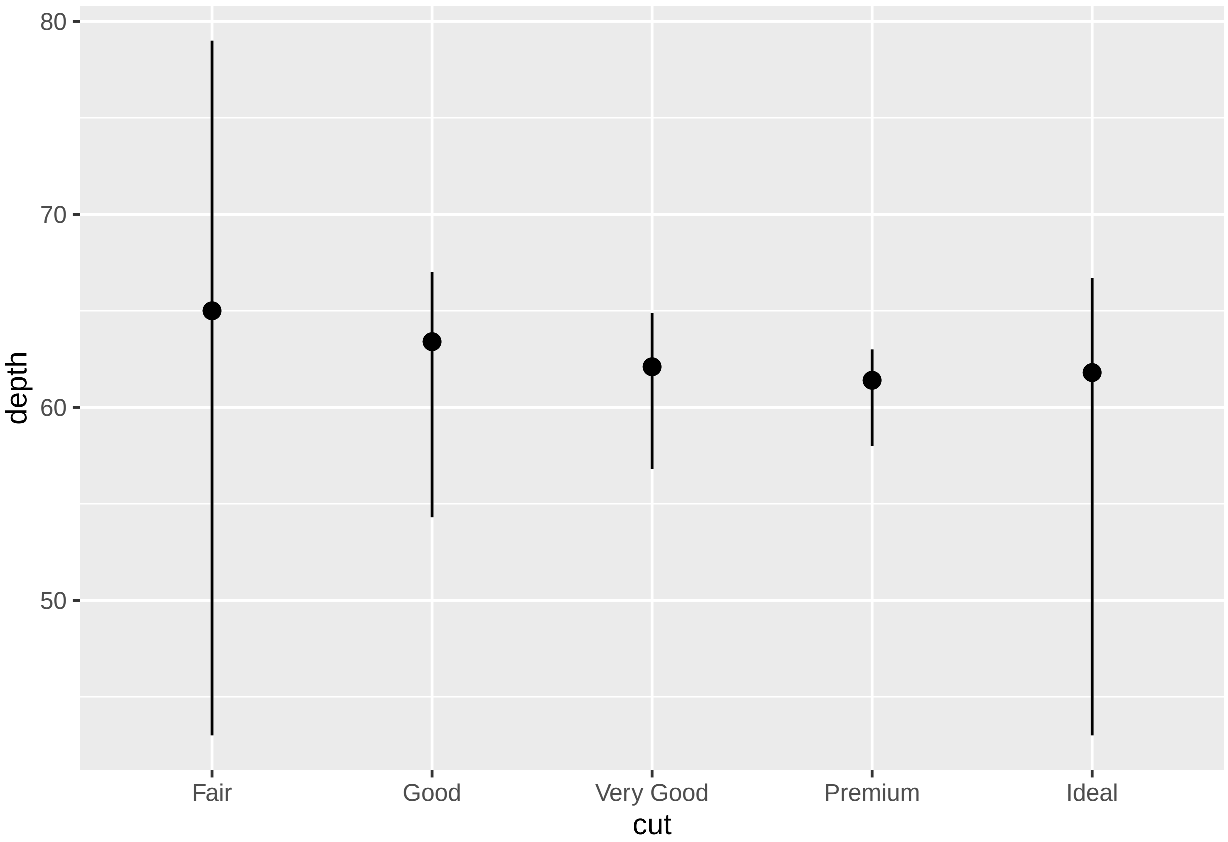
Barplot

```
ggplot(data=diamonds)+geom_bar(mapping = aes(x=cut, y = ..prop.., group=1))
```



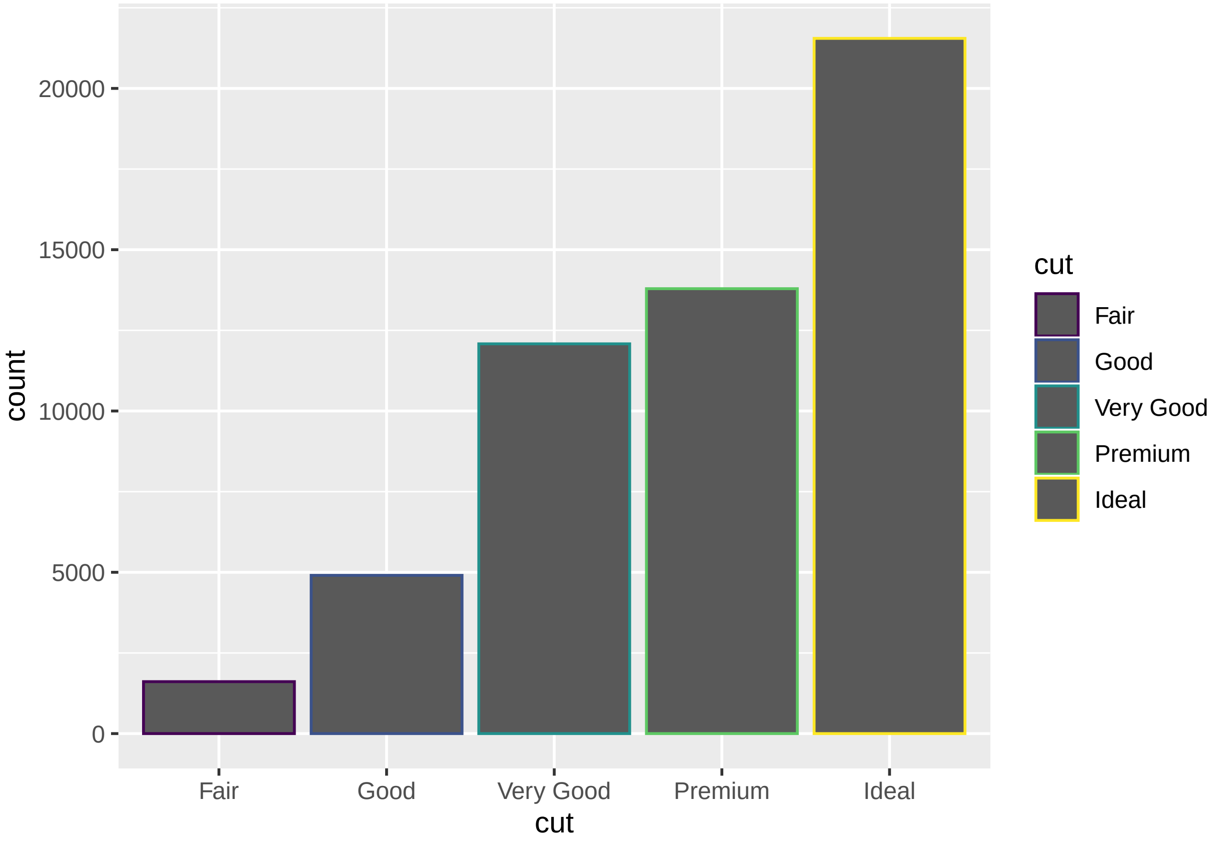
Summarising the values

```
ggplot(data = diamonds)+stat_summary(mapping = aes(x=cut, y=depth), fun.ymin=min, fun.ymax = max, fun.y
```

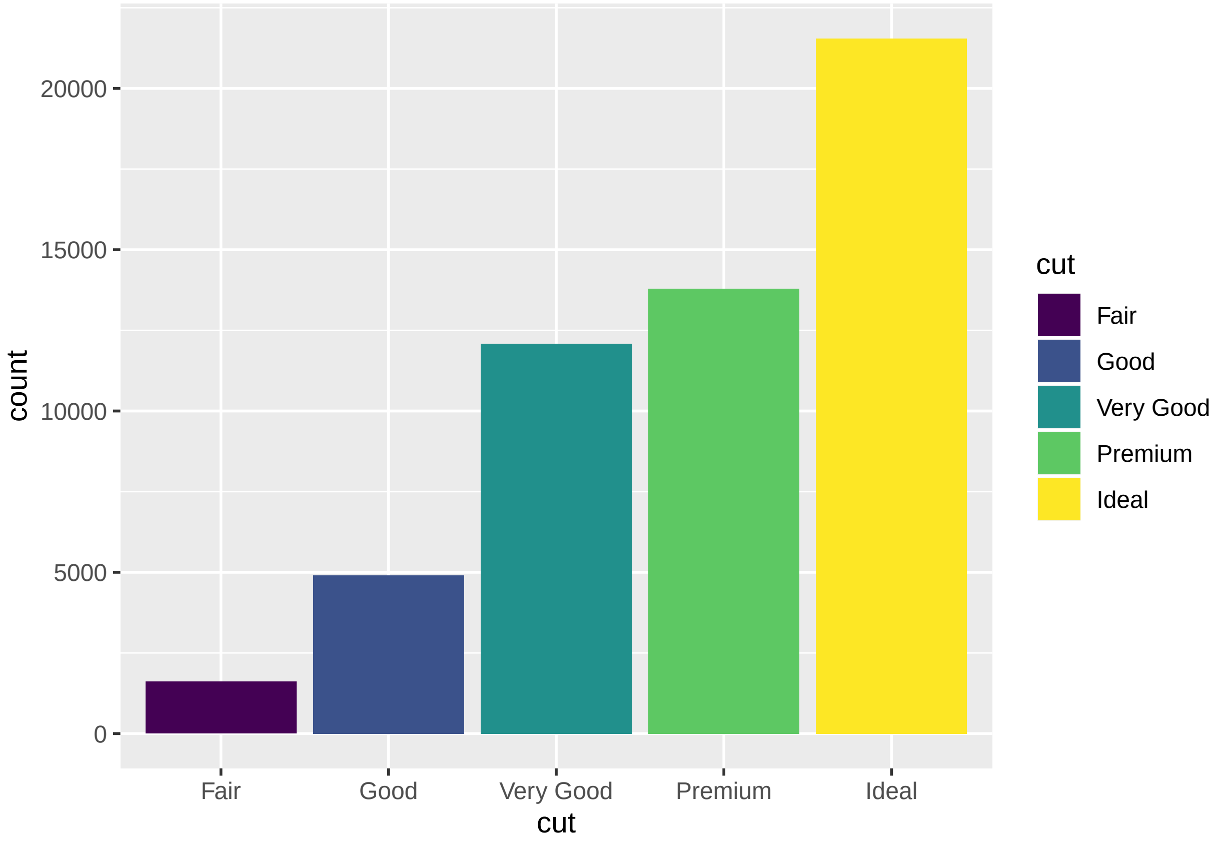


filling the colour

```
ggplot(data = diamonds)+geom_bar(mapping = aes(x=cut, colour =cut))
```

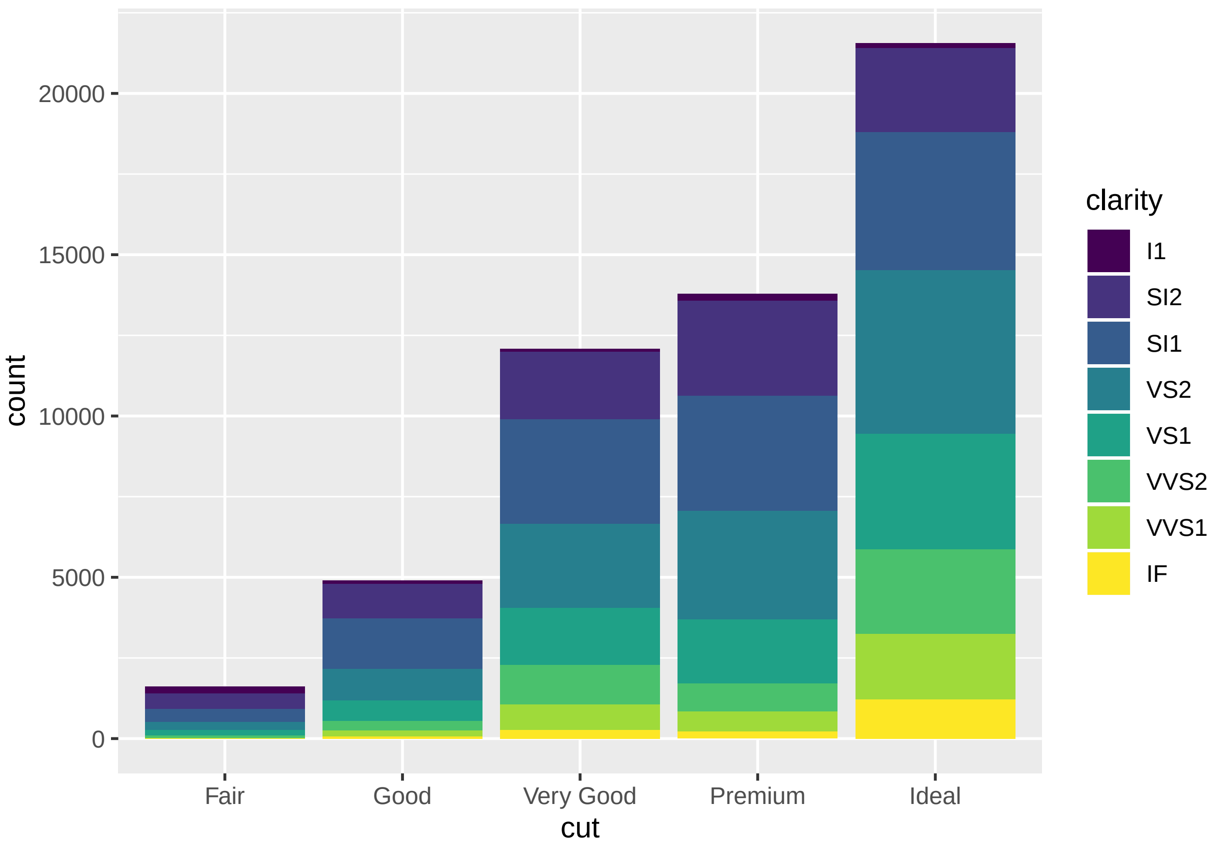


```
ggplot(data=diamonds)+geom_bar(mapping= aes(x=cut, fill = cut))
```



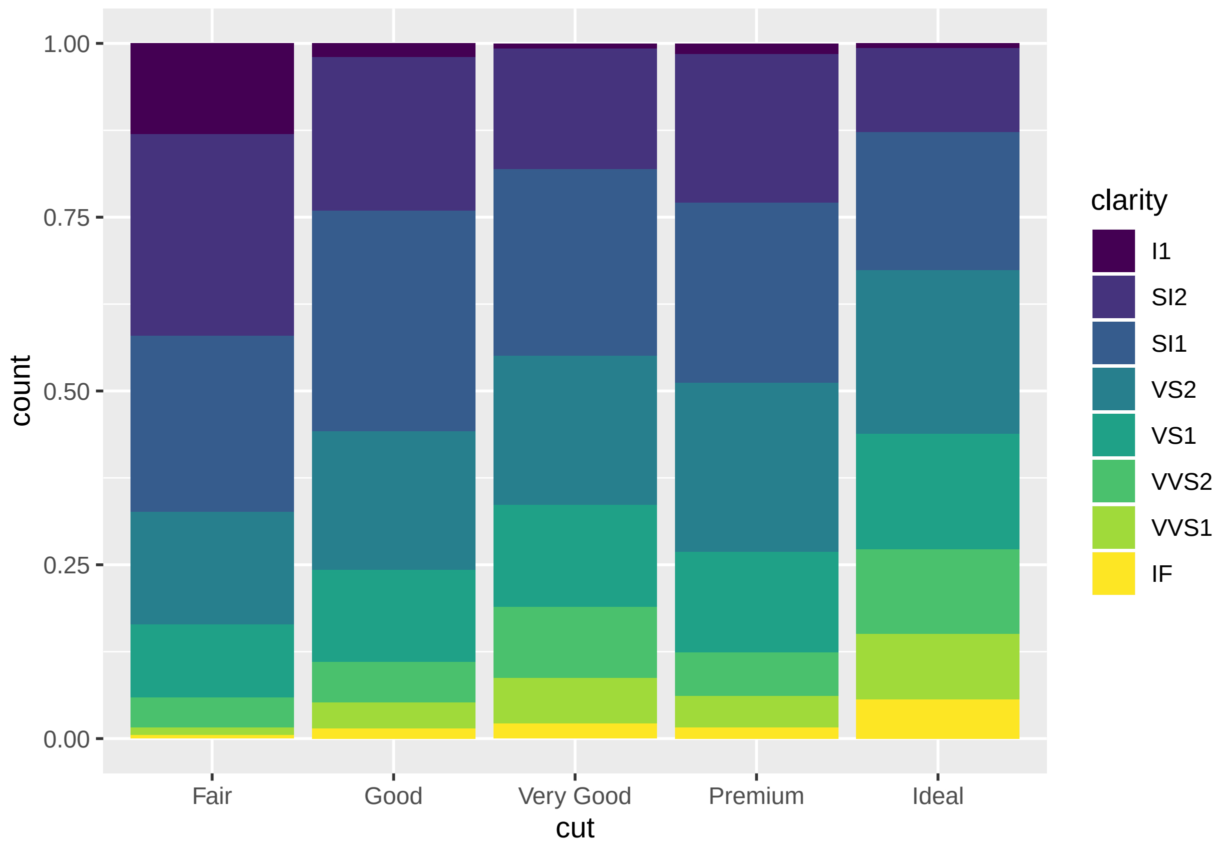
For clarity, the bars are automatically stacked

```
ggplot(data=diamonds)+ geom_bar(mapping=aes(x=cut, fill=clarity))
```



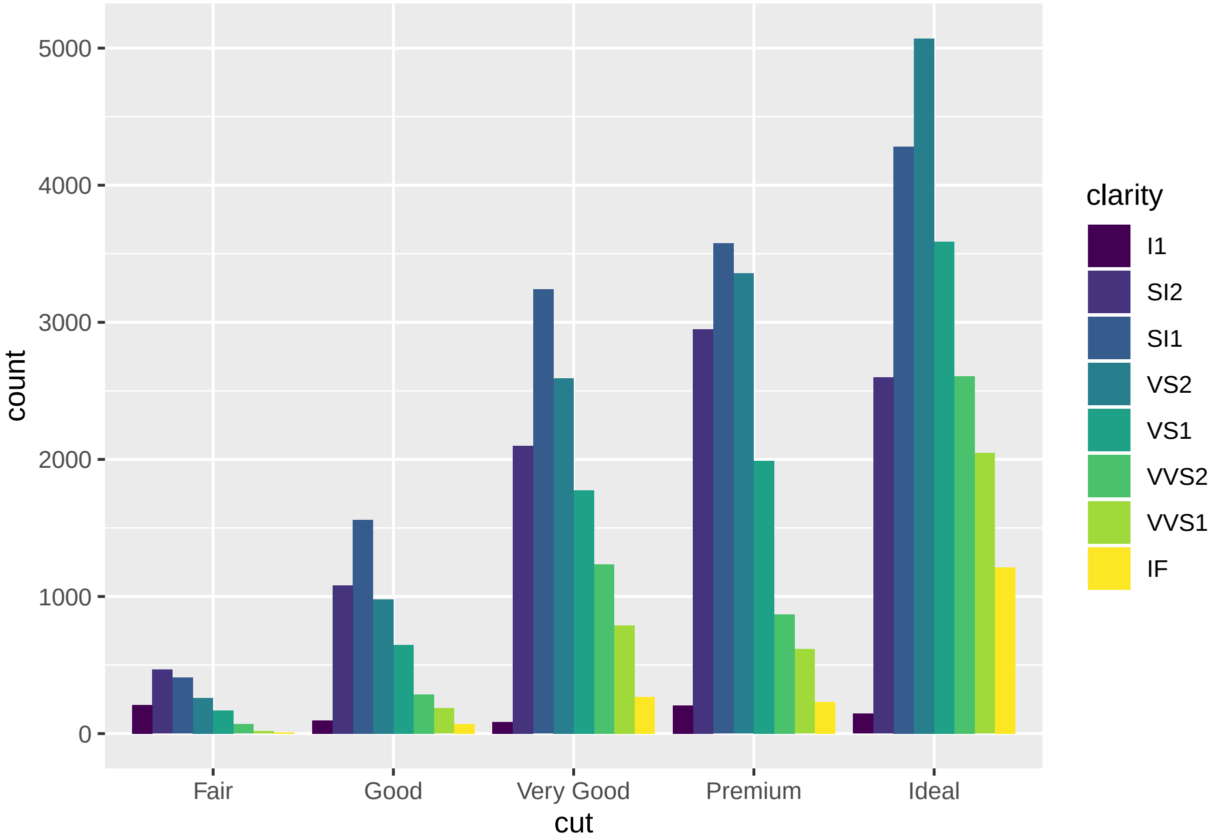
To compare proportions use fill

```
ggplot(data=diamonds)+
  geom_bar(mapping = aes(x=cut, fill=clarity), position = "fill")
```



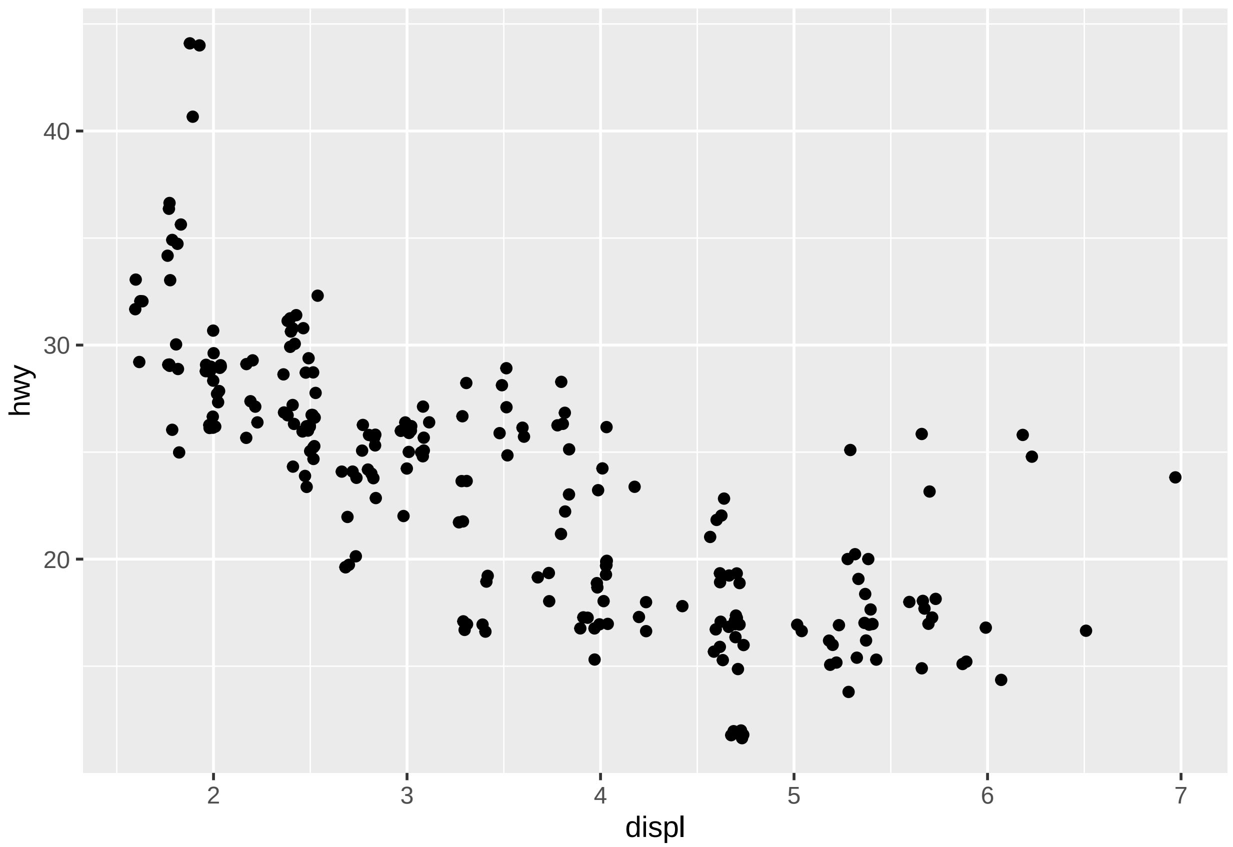
To compare individuals use dodge

```
ggplot(data=diamonds)+ geom_bar(mapping = aes(x=cut, fill= clarity), position = "dodge")
```



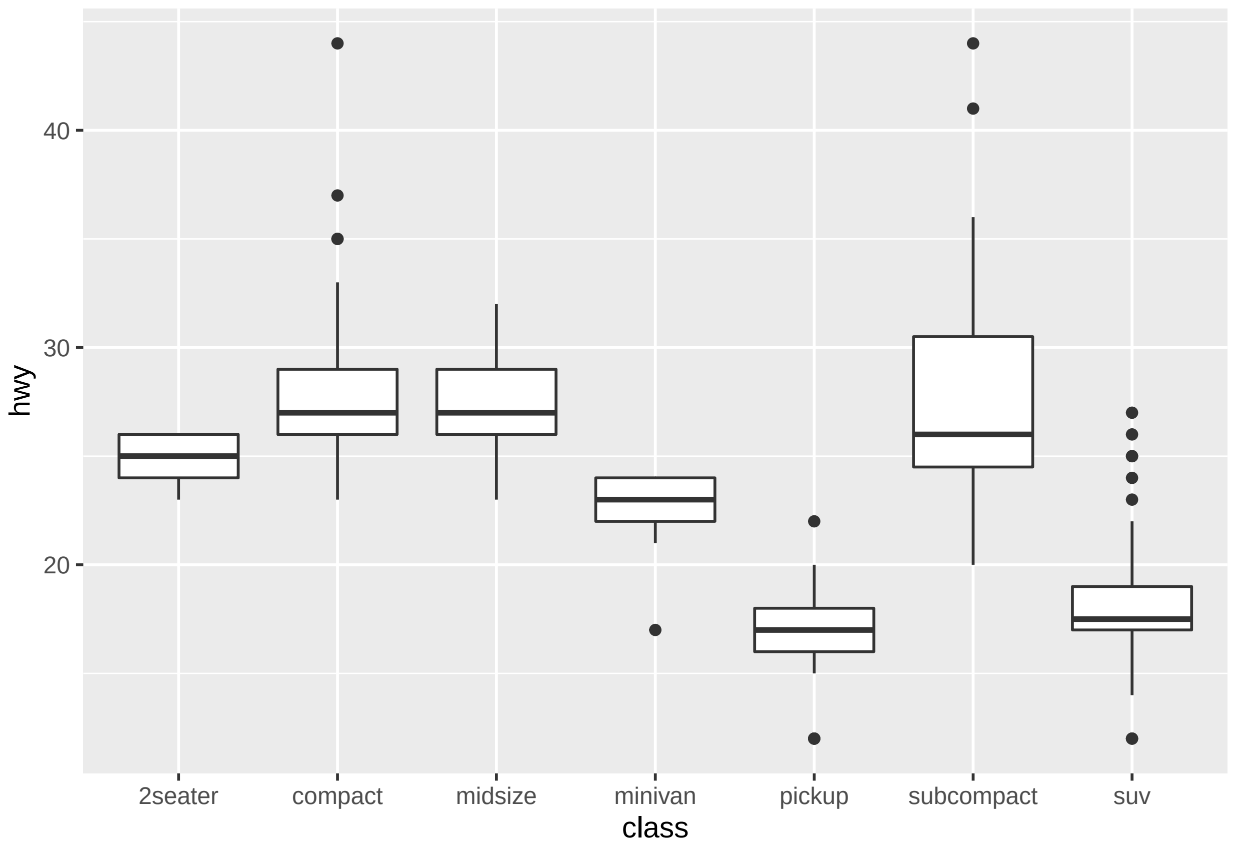
To add noise to point use jitter

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

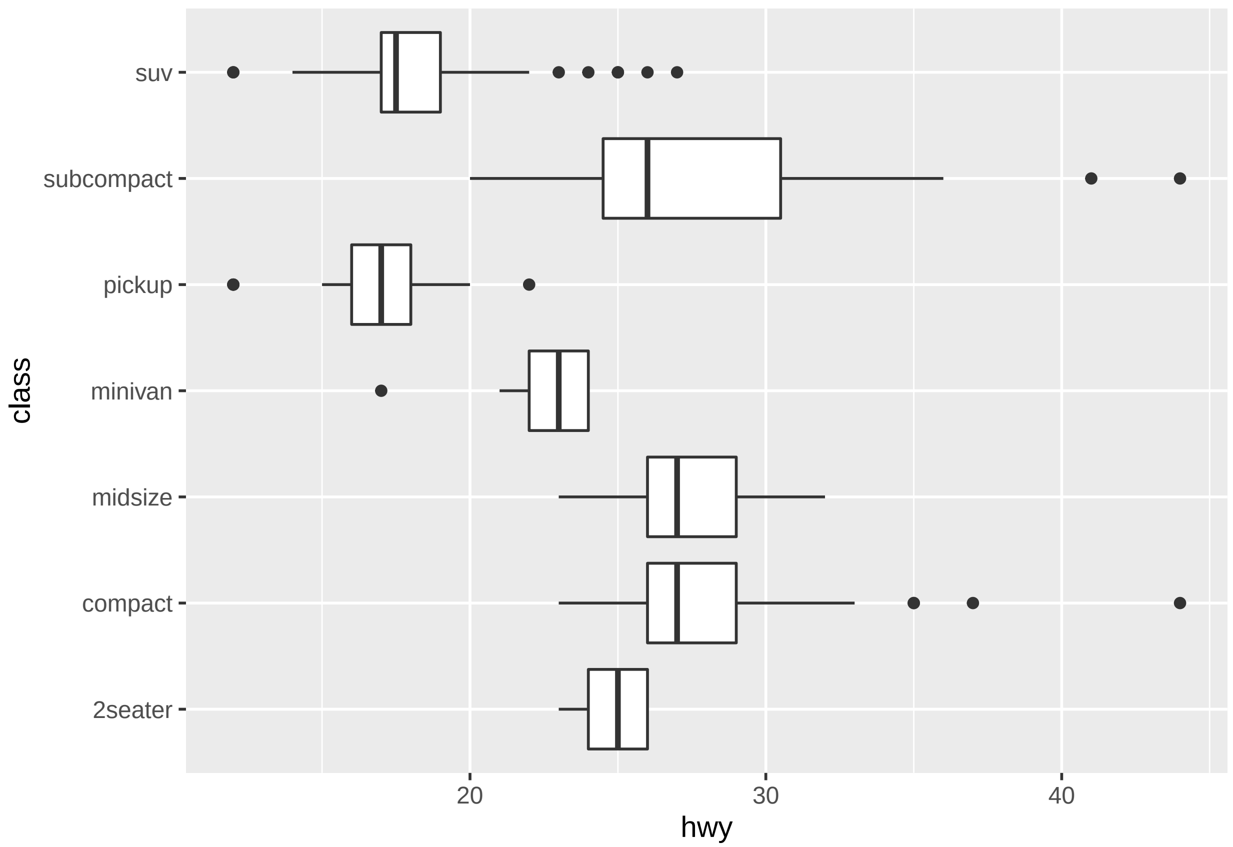


To switch x and y axes use the function, coord_flip()

```
ggplot(data=mpg, mapping = aes(x=class, y=hwy))+geom_boxplot()
```

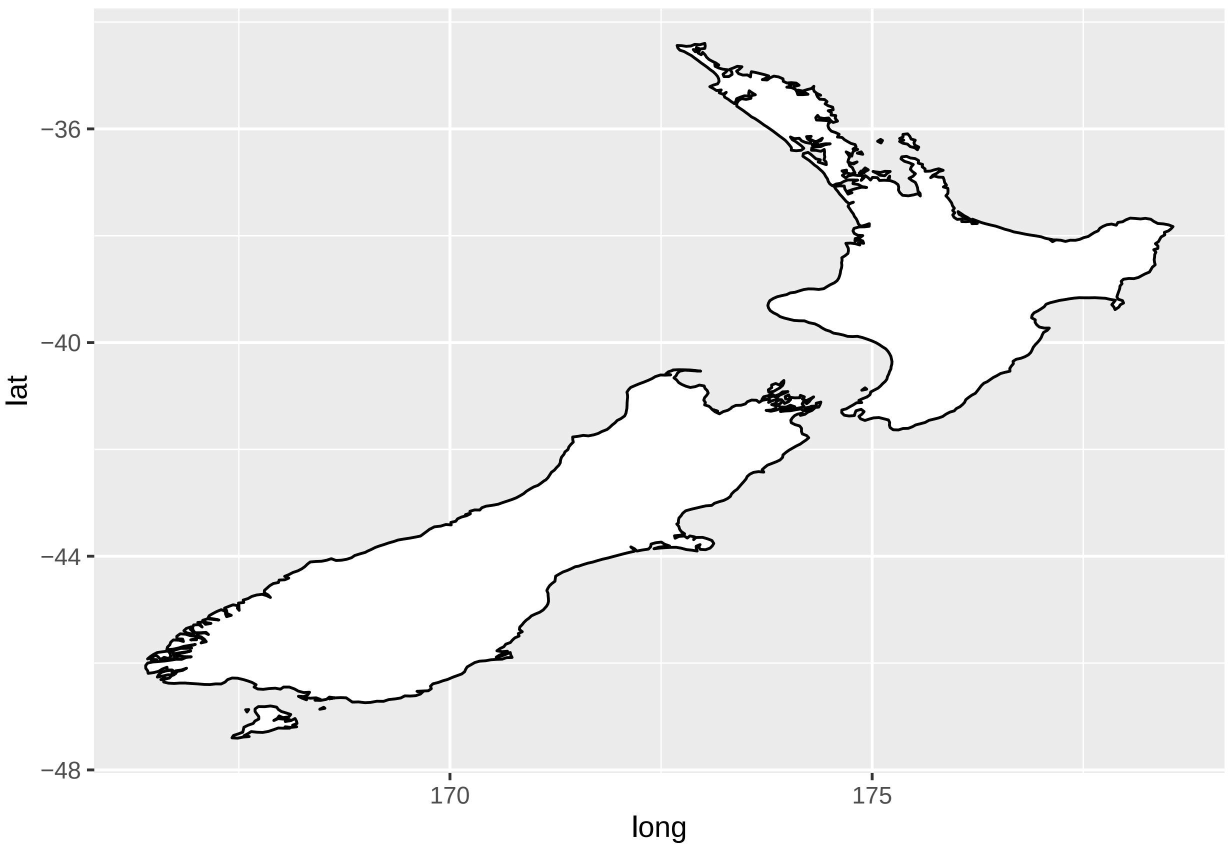


```
ggplot(data=mpg, mapping= aes(x=class, y=hwy))+ geom_boxplot() + coord_flip()
```

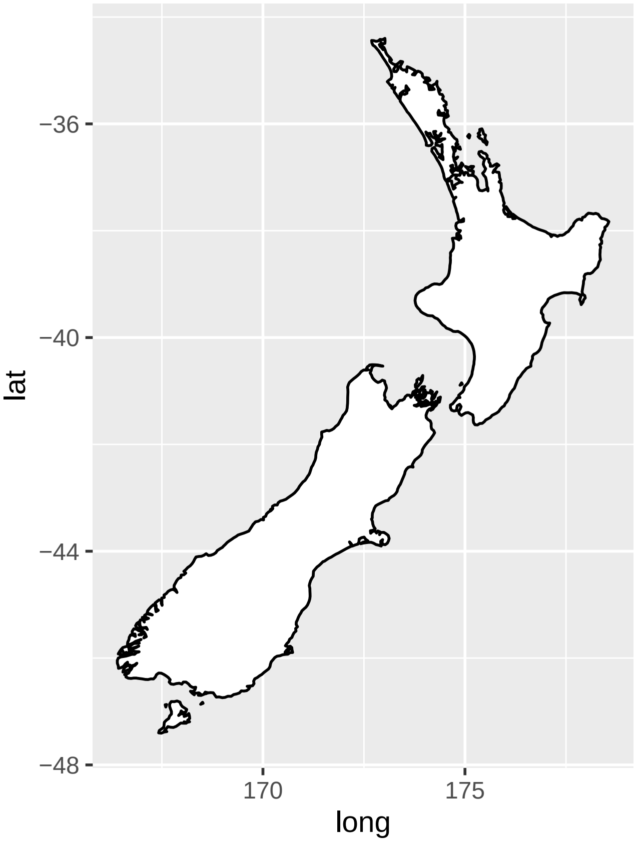


To set aspect ratio for maps use the function, coord_quickmap()

```
nz <- map_data("nz")
ggplot(nz, aes(long, lat, group=group))+ geom_polygon(fill = "white", colour = "black")
```



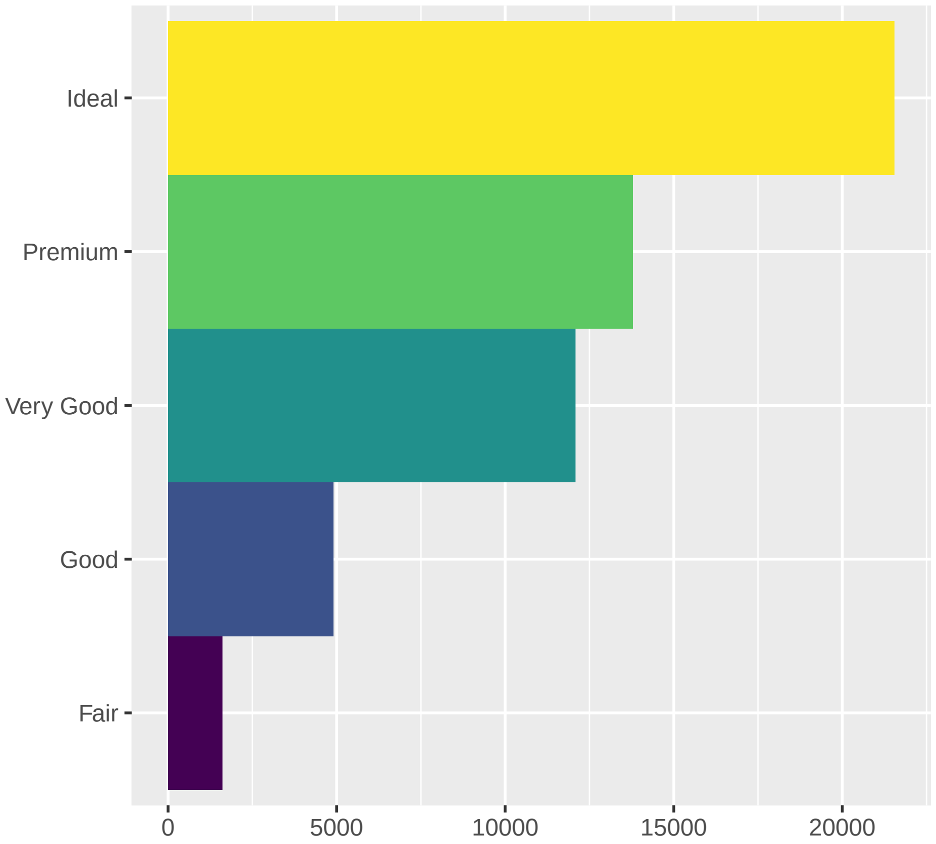
```
ggplot(nz, aes(long, lat, group=group))+geom_polygon(fill="white", colour = "black" )+ coord_quickmap()
```



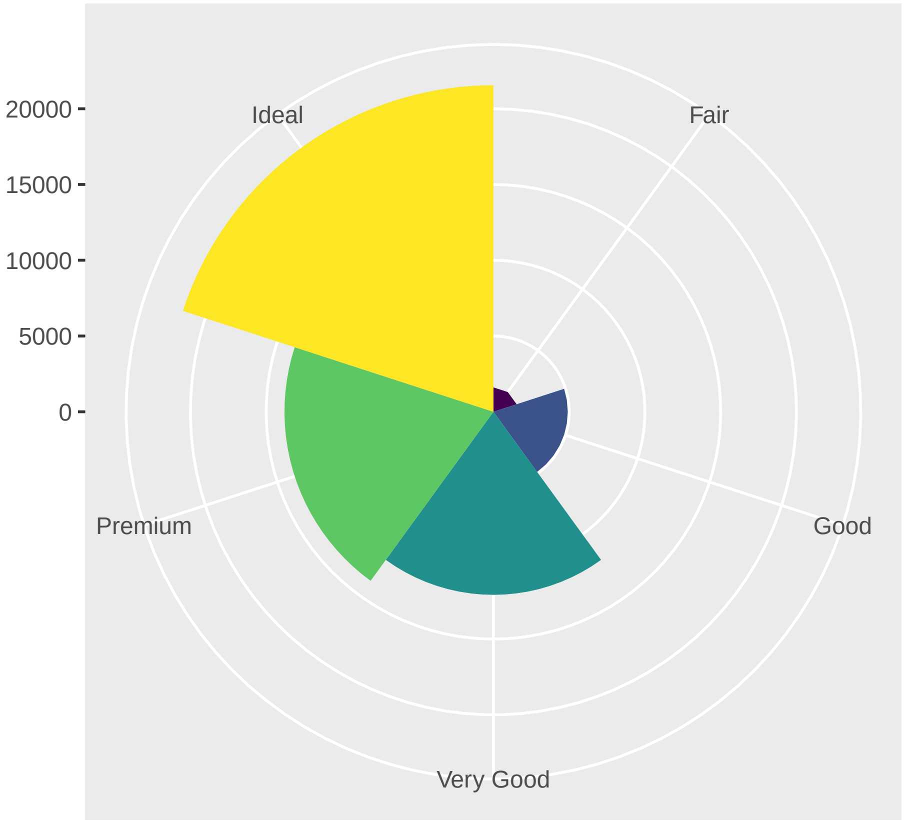
Combing box chart and coxcombo chart

```
bar <- ggplot(data = diamonds)+ geom_bar(mapping = aes(x=cut, fill=cut), show.legend= FALSE, width = 1)

bar+coord_flip()
```



```
bar+coord_polar()
```


20000
15000
10000
5000
0
Ideal
Fair
Good
Very Good
Premium