Chapter2 : Overview

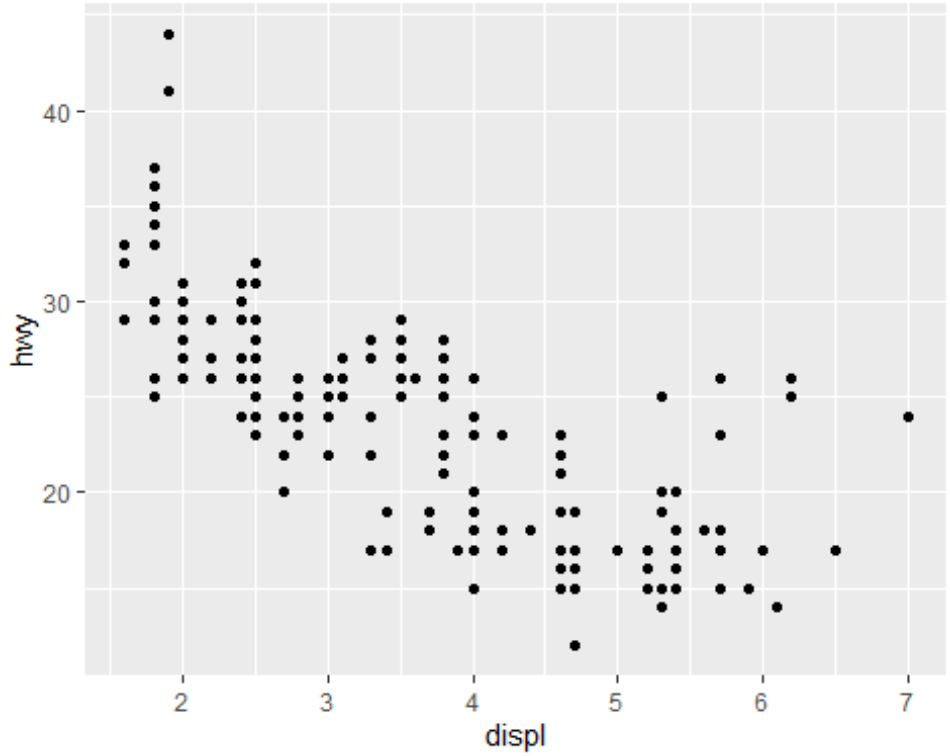
songminsoo

2.3 ggplot2's 3 key component

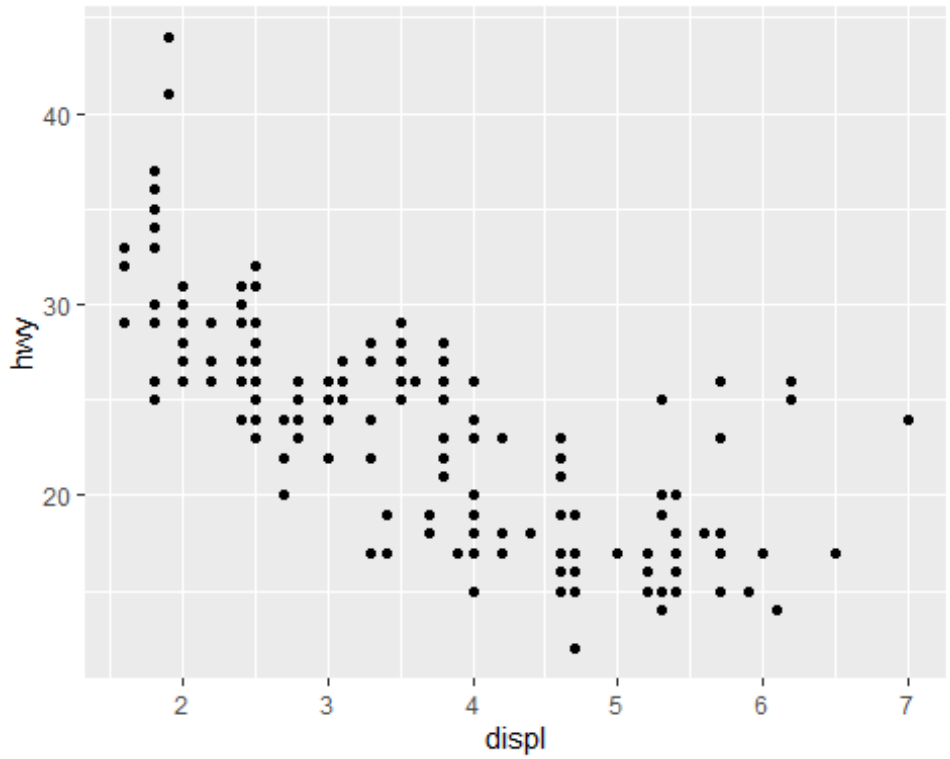
1. data
2. aesthetic mappings : between variable in the data and visual properties
3. geom : how to render each obs. layers are usaually created with a geom function

```
library(ggplot2)

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



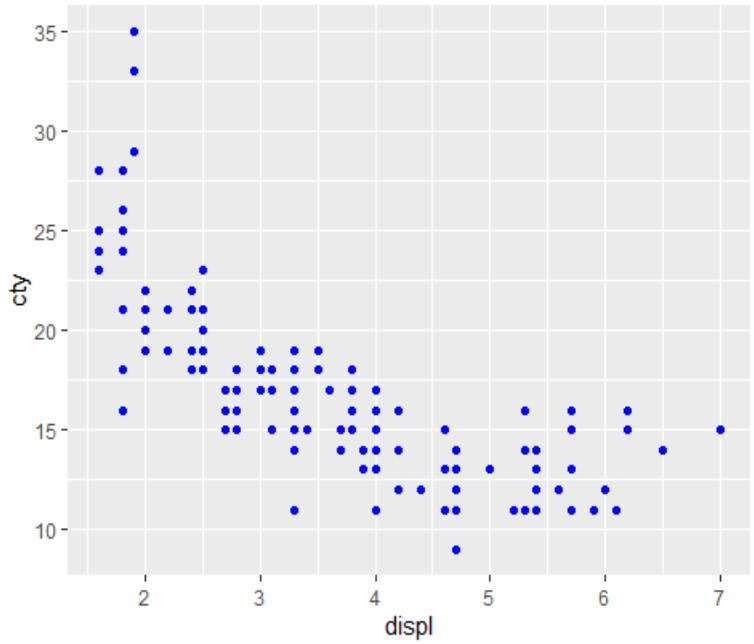

```
# without x, y notation
ggplot(mpg, aes(displ, hwy))+
  geom_point()
```


hwy
40
30
20
2
3
4
5
6
7
displ

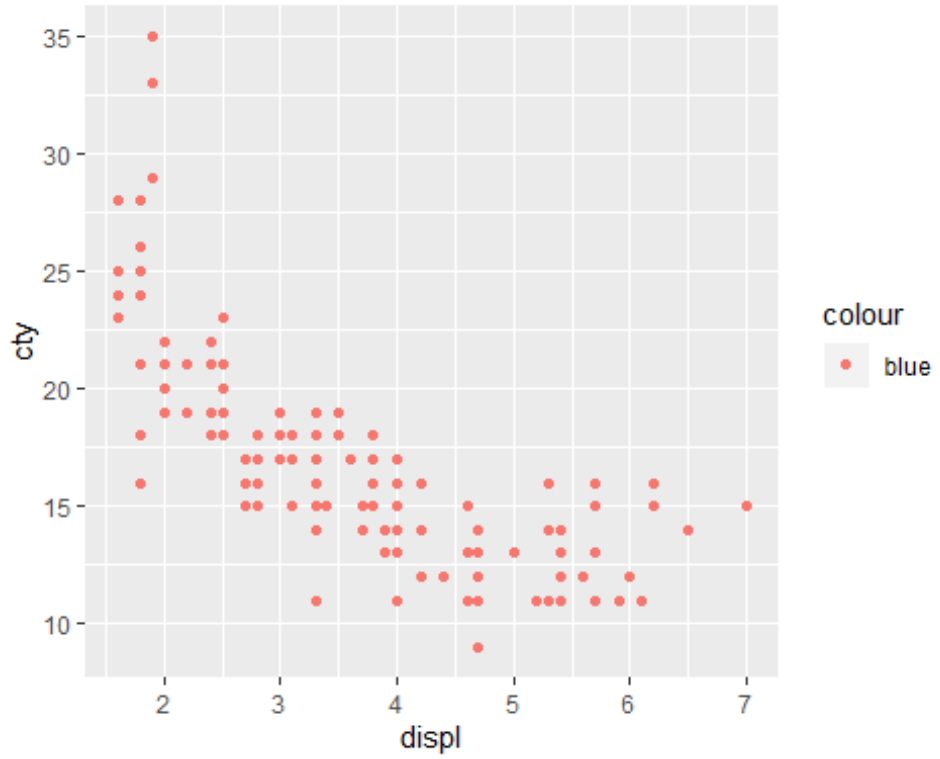
2.4 colour,size, shape and others

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


```
ggplot(mpg, aes(displ,cty))+
  geom_point(colour = 'blue')
```



```r
ggplot(mpg, aes(displ,cty)) +
  geom_point(aes(colour = 'blue'))
```

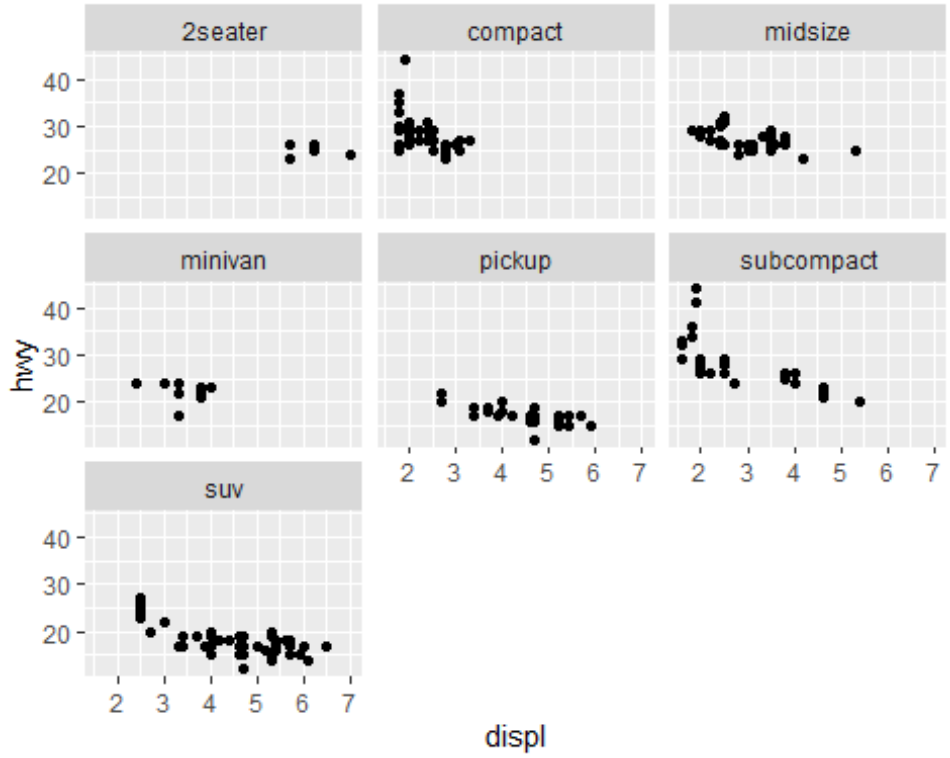



```r
# they are all different. we will see this later.
```

2.5 Facetting

- displaying additional categorical variables on a plot
- creates tables of graphics by splitting the data into subsets and displaying the same graph for each subset
- two types of facetting : grid & wrapped
- see more details later

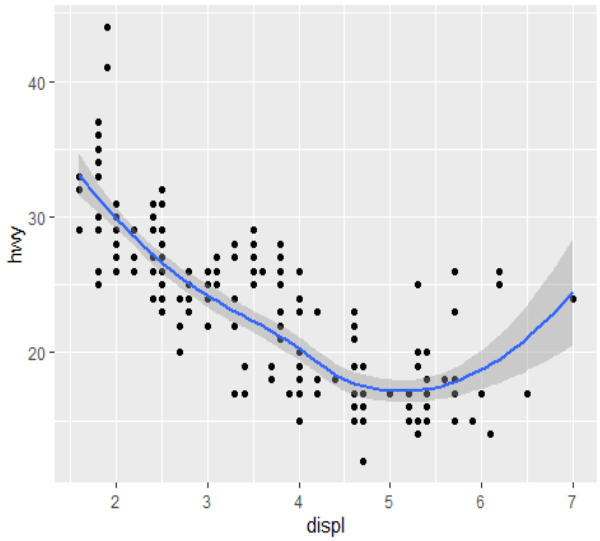
```
ggplot(mpg, aes(displ, hwy))+
  geom_point()+
  facet_wrap(~class)
```



2.6 plot Geoms

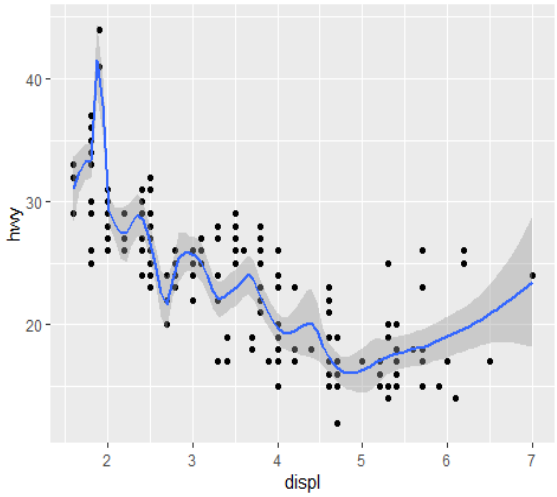
```
# add a smoothing
ggplot(mpg, aes(displ, hwy))+
  geom_point()+
  geom_smooth()
```

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



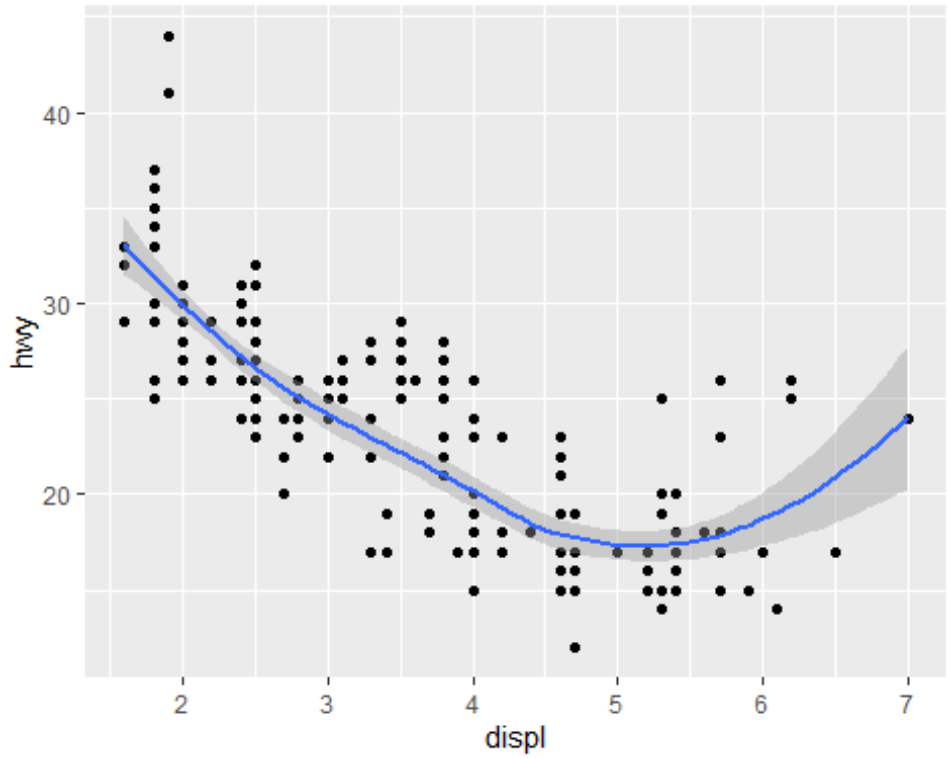

```
# method = 'loess'
ggplot(mpg, aes(displ, hwy))+
  geom_point()+
  geom_smooth(span=0.2)
```

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



```
ggplot(mpg, aes(displ, hwy))+
  geom_point()+
  geom_smooth(span=0.8)
```

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

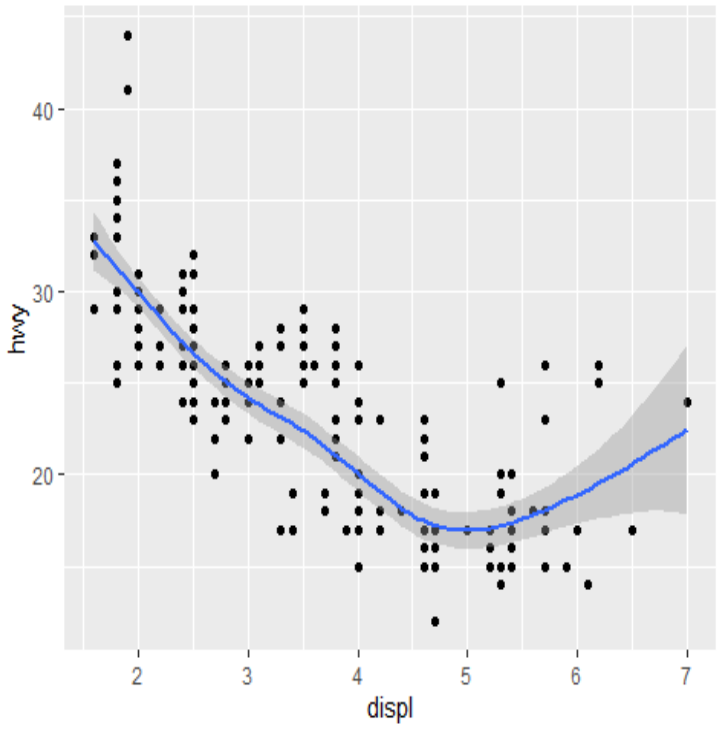



```
# method = 'gam'
# when n > 1000
library(mgcv)
```

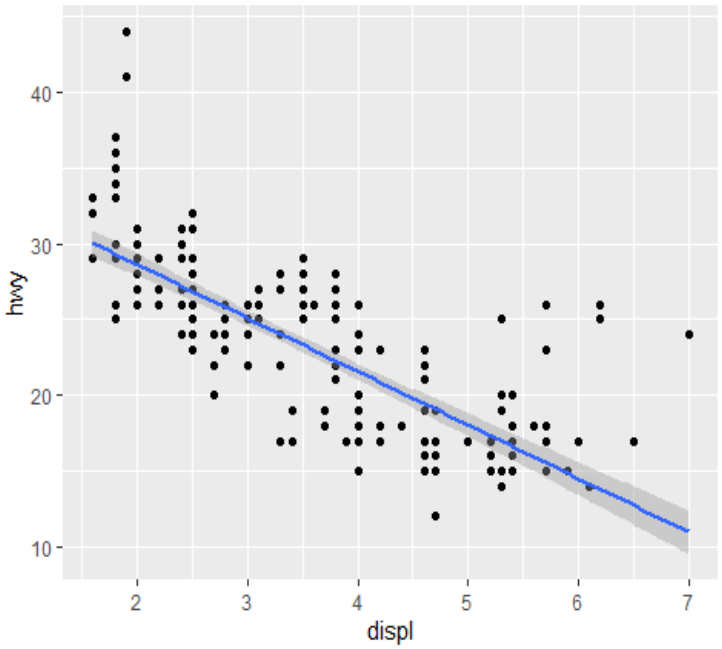
```
## Loading required package: nlme
```

```
## This is mgcv 1.8-31. For overview type 'help("mgcv-package")'.
```

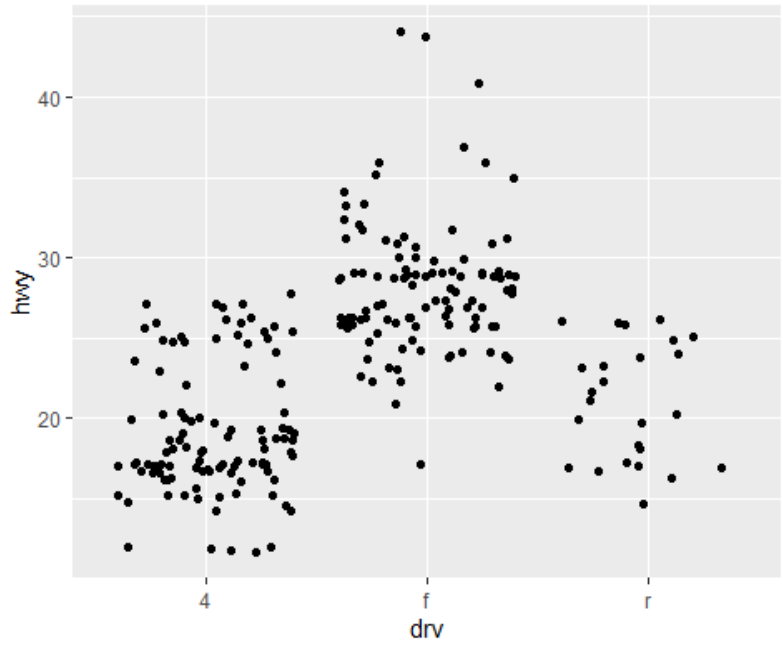
```
ggplot(mpg, aes(displ, hwy))+
  geom_point()+
  geom_smooth(method='gam',formula = y~s(x))
```



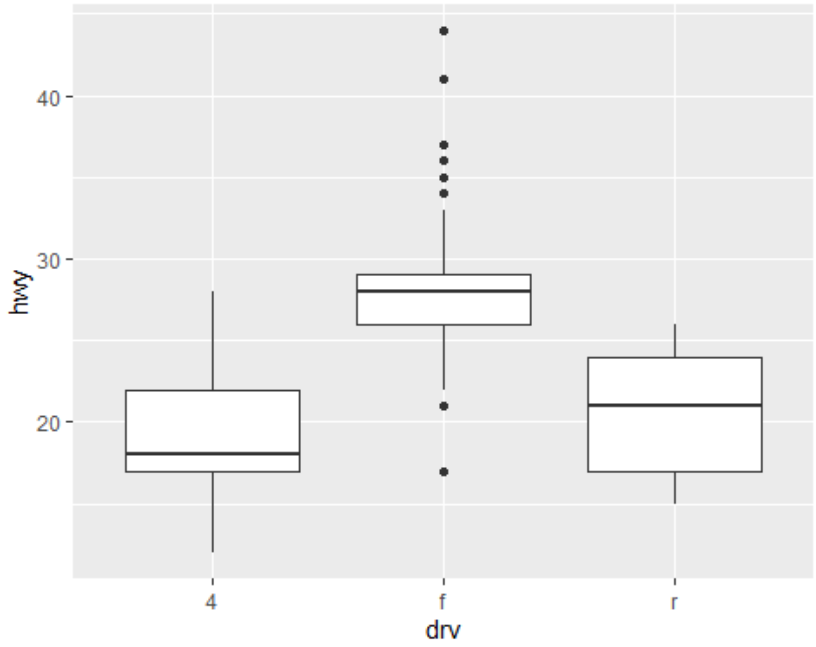

```r
# method = 'lm'
ggplot(mpg, aes(displ, hwy))+
  geom_point()+
  geom_smooth(method = 'lm')
```



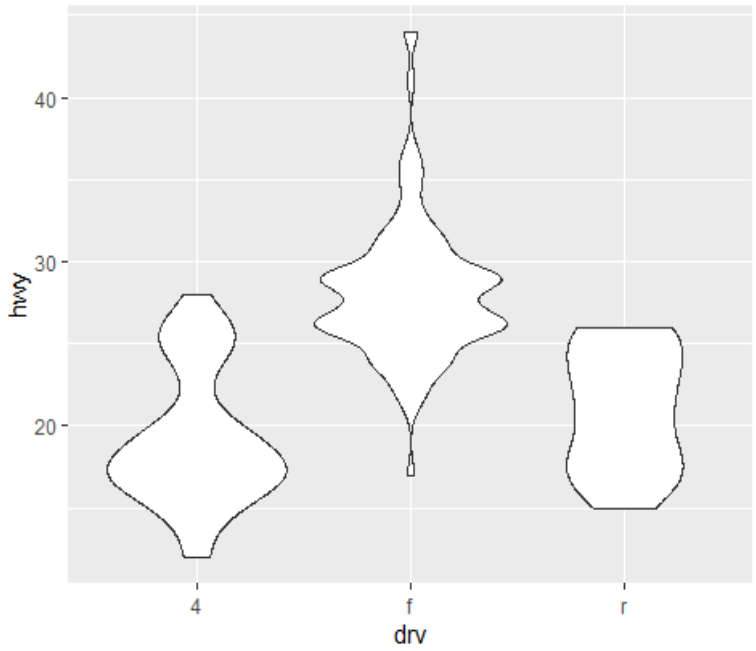
```r
# Boxplot & Jittered points
ggplot(mpg, aes(drv, hwy))+
  geom_jitter()
```




```r
ggplot(mpg, aes(drv, hwy))+
  geom_boxplot()
```

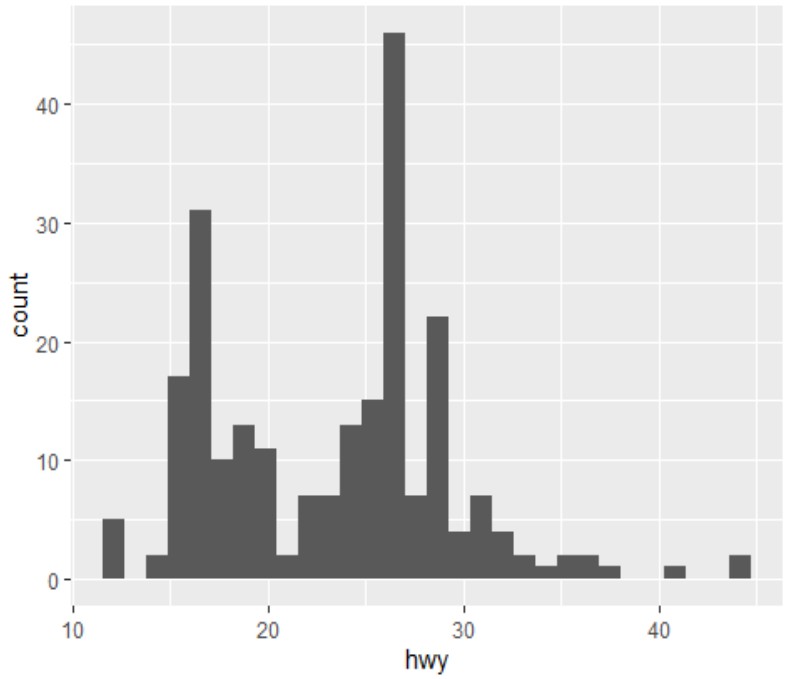


```
ggplot(mpg, aes(drv, hwy))+
  geom_violin()
```



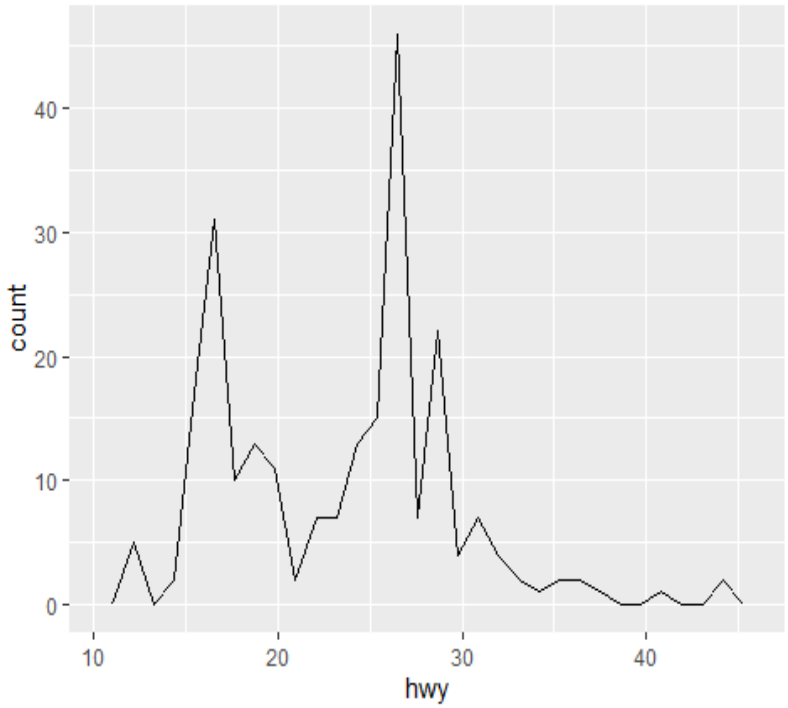

```
# Histogram, polygons
ggplot(mpg, aes(hwy)) + geom_histogram()

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

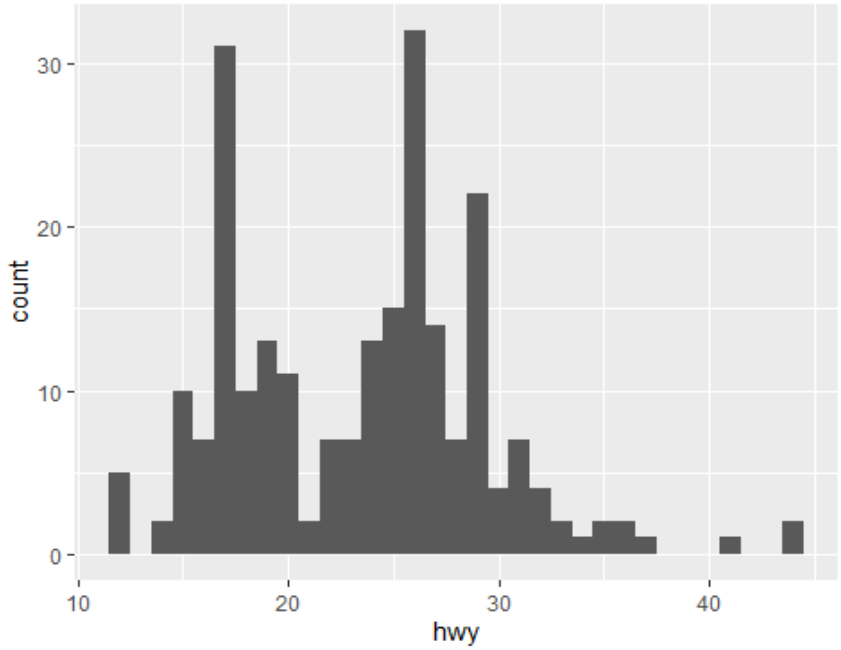


```
ggplot(mpg, aes(hwy)) + geom_freqpoly()
```

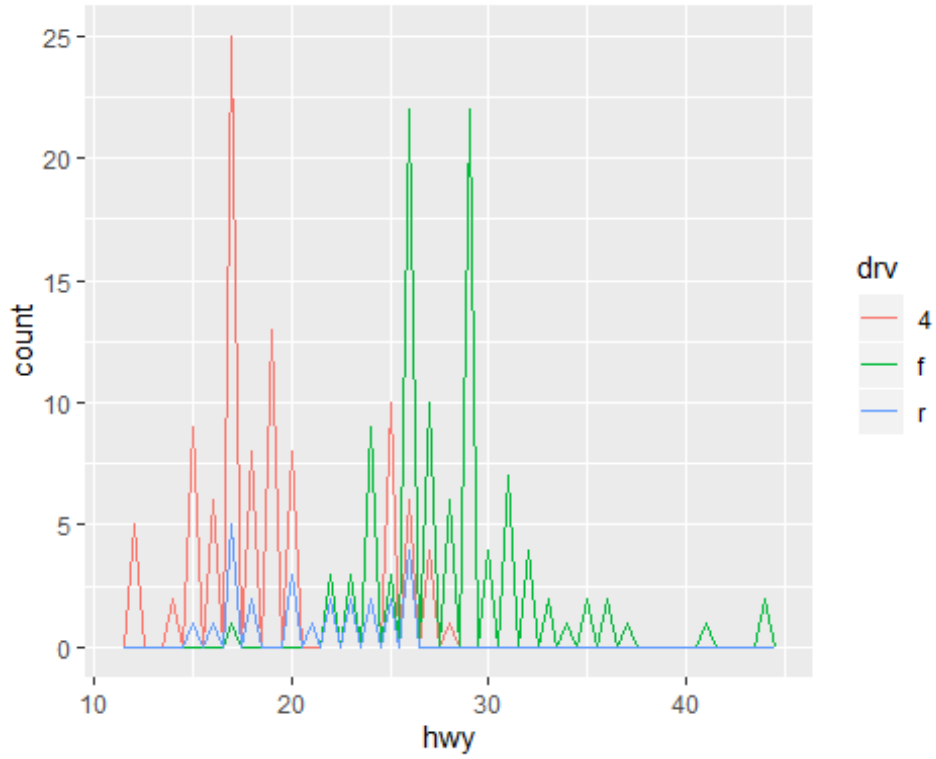
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



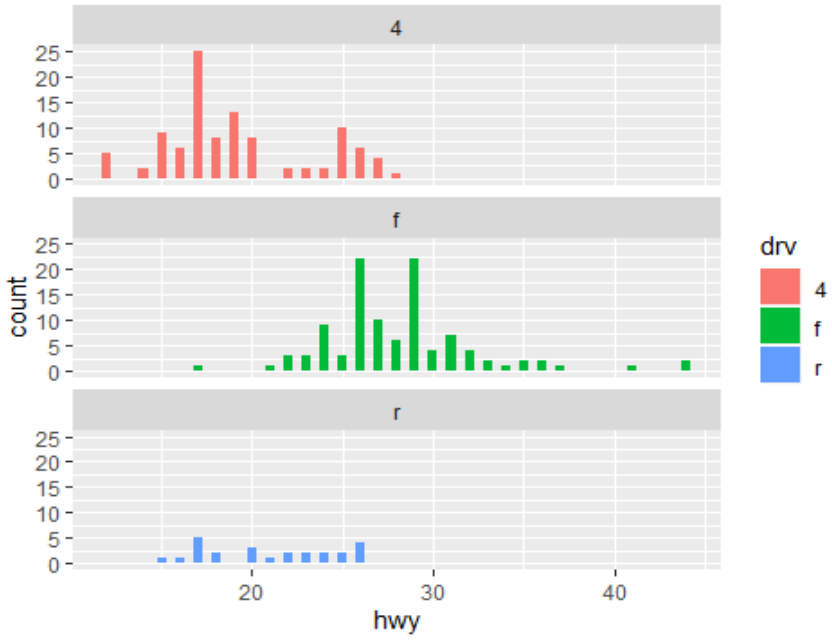

```
ggplot(mpg, aes(hwy)) +
  geom_histogram(binwidth = 1)
```



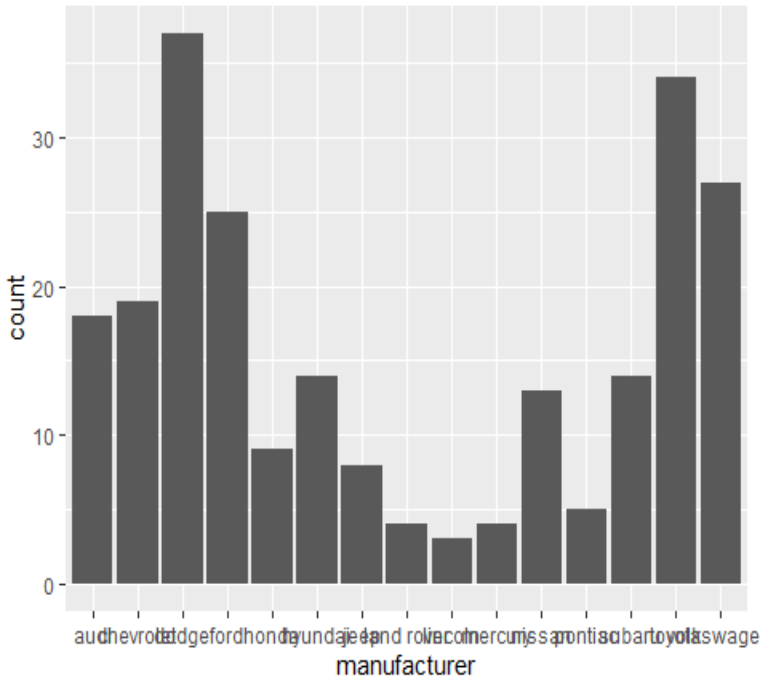
```
ggplot(mpg, aes(hwy, colour=drv)) +
  geom_freqpoly(binwidth = 0.5)
```



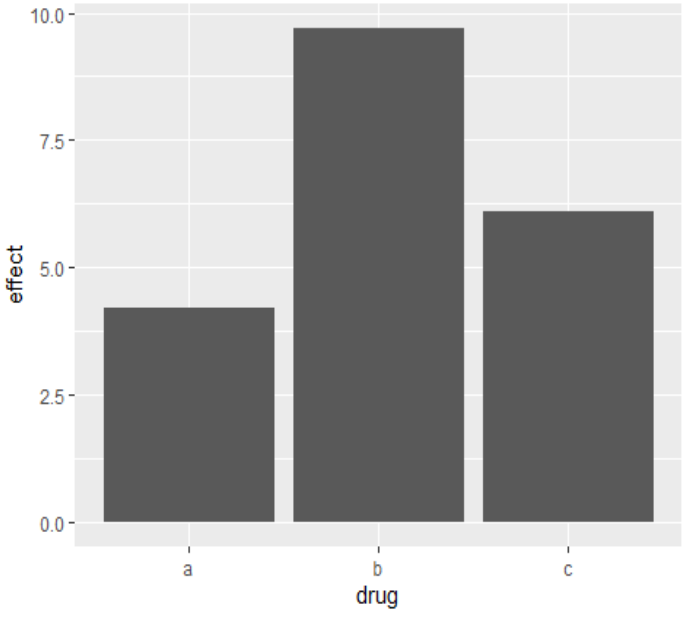

```
ggplot(mpg, aes(hwy, fill=drv)) +
  geom_histogram(binwidth = 0.5) +
  facet_wrap(~drv, ncol = 1)
```



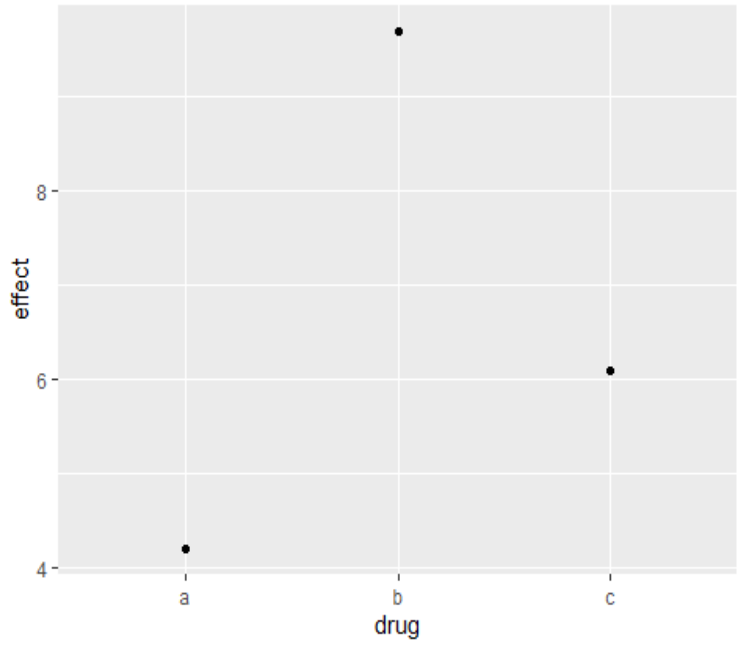
```r
# Bar
## not for summary data
## for count plot
ggplot(mpg, aes(manufacturer))+
  geom_bar()
```



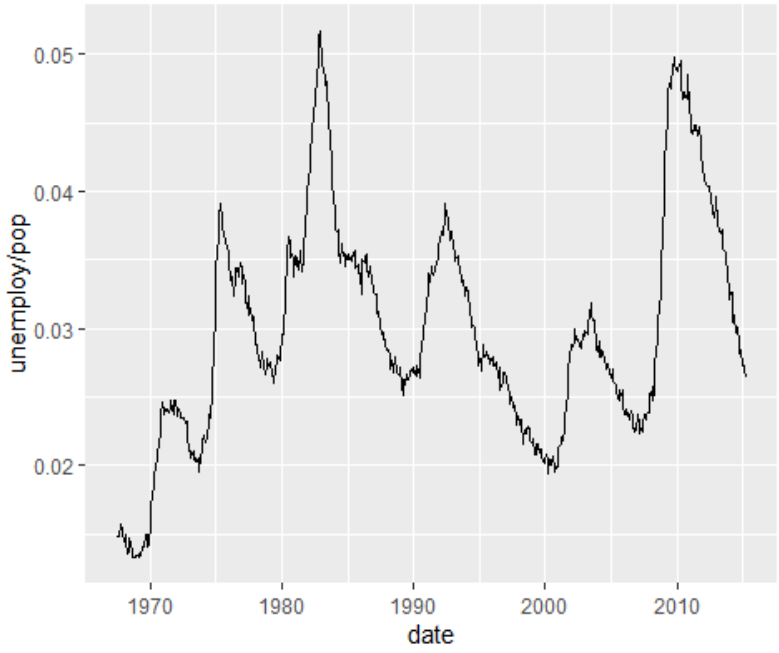

```r
# to use in summaried data
drugs <- data.frame(drug = c("a", "b", "c"),effect = c(4.2, 9.7, 6.1))
ggplot(drugs, aes(drug, effect)) + geom_bar(stat = "identity")
```



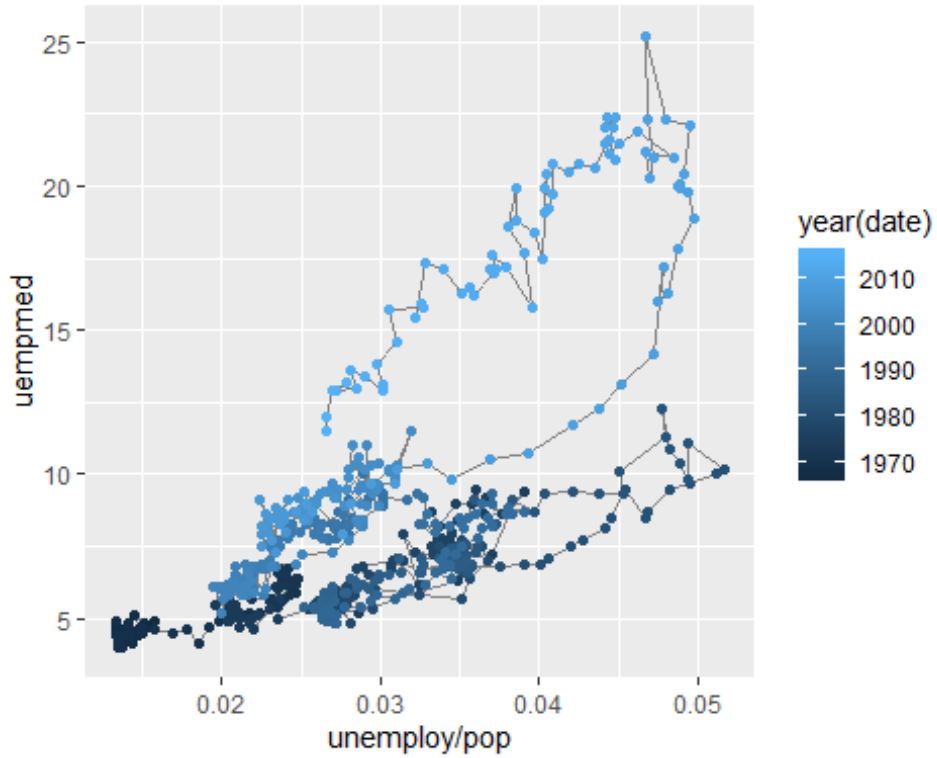
```r
## u can use point plot
ggplot(drugs, aes(drug, effect)) + geom_point()
```




```r
# line & path plot
## line : x-axis is usaully time
## path : two variables with time
ggplot(economics, aes(date, unemploy/pop))+
  geom_line()
```

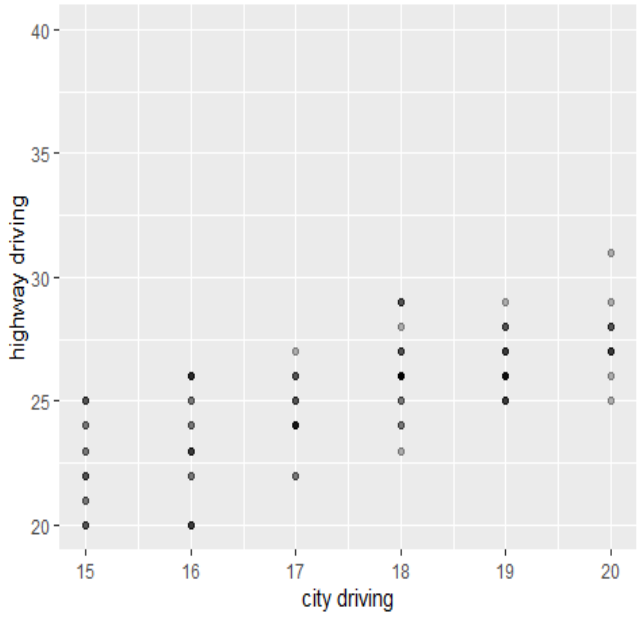


```r
year = function(x) as.POSIXlt(x)$year + 1900
ggplot(economics, aes(unemploy / pop, uempmed)) +
  geom_path(colour = 'grey50') +
  geom_point(aes(colour = year(date)))
```

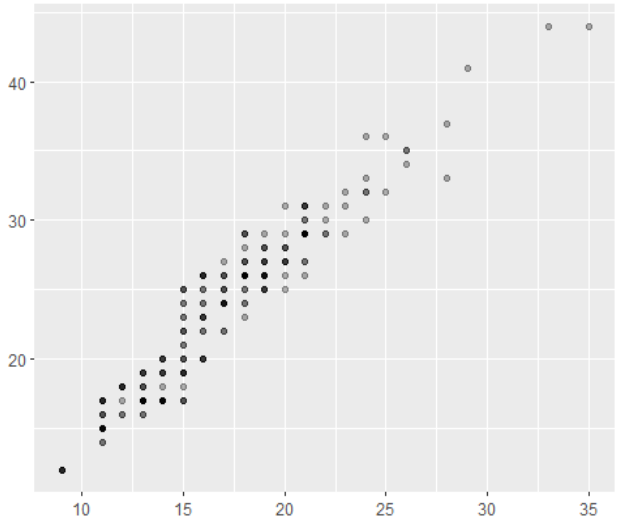


2.7 Modifying the axes

```
ggplot(mpg, aes(cty, hwy)) +
  geom_point(alpha = 0.3) +
  xlab('city driving') +
  ylab('highway driving') +
  xlim(15,20) +
  ylim(20,40)
```



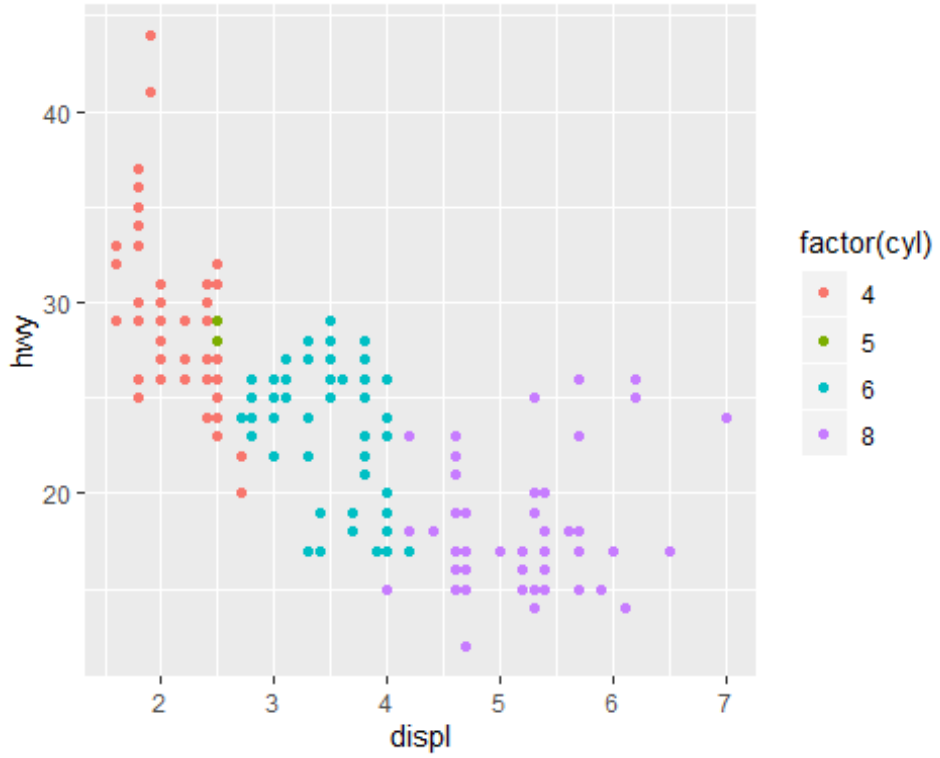

```
ggplot(mpg, aes(cty, hwy)) +
  geom_point(alpha = 0.3) +
  xlab(NULL) +  # u can remove the xlab, ylab by NULL
  ylab(NULL)
```



2.8 Output

```
p = ggplot(mpg, aes(displ, hwy, colour = factor(cyl)))+
  geom_point()

print(p)
```



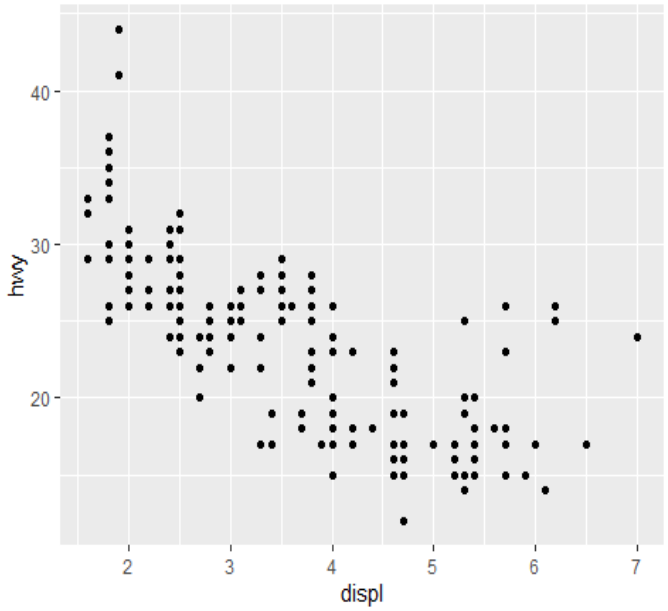
```r
# save
ggsave('plot.png', width = 5, height = 5)

# summmary
summary(p)
```

```
## data: manufacturer, model, displ, year, cyl, trans, drv, cty, hwy, fl,
##   class [234x11]
## mapping:  x = ~displ, y = ~hwy, colour = ~factor(cyl)
## faceting: <ggproto object: Class FacetNull, Facet, gg>
##     compute_layout: function
##     draw_back: function
##     draw_front: function
##     draw_labels: function
##     draw_panels: function
##     finish_data: function
##     init_scales: function
##     map_data: function
##     params: list
##     setup_data: function
##     setup_params: function
##     shrink: TRUE
##     train_scales: function
##     vars: function
##     super:  <ggproto object: Class FacetNull, Facet, gg>
## -----------------------------------
## geom_point: na.rm = FALSE
## stat_identity: na.rm = FALSE
## position_identity
```

2.9 Quick plot

```
# using qplot() -> picking a geom by default
# not that recommeded
qplot(displ, hwy, data = mpg)
```




```
qplot(displ, data = mpg)

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```