

Lab 2: ggplot2

Spring 2018 - Multivariate Data Analysis

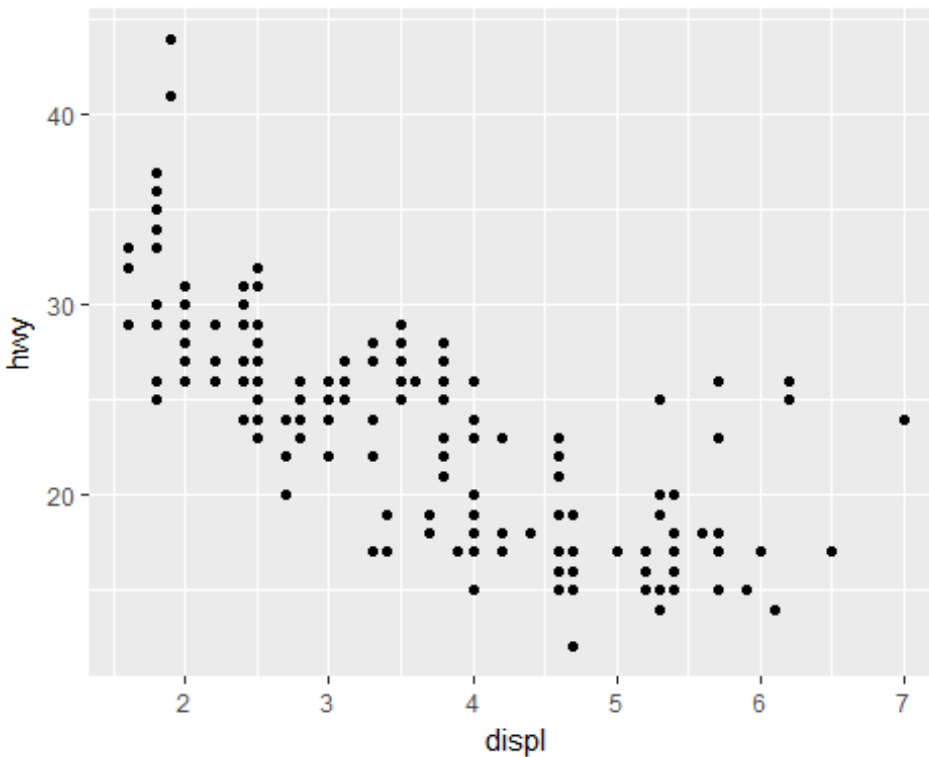
ggplot2 - intro

```
library(ggplot2)
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     1.80  1999     4 auto~ f     18    29 p    comp~
## 2 audi         a4     1.80  1999     4 manu~ f     21    29 p    comp~
## 3 audi         a4     2.00  2008     4 manu~ f     20    31 p    comp~
## 4 audi         a4     2.00  2008     4 auto~ f     21    30 p    comp~
## 5 audi         a4     2.80  1999     6 auto~ f     16    26 p    comp~
## 6 audi         a4     2.80  1999     6 manu~ f     18    26 p    comp~
```

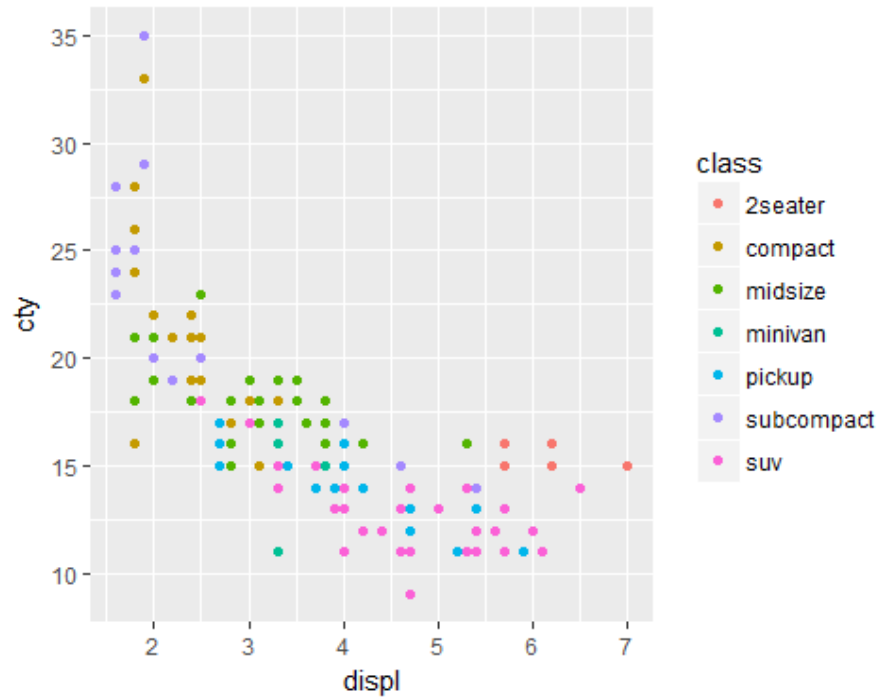
Key Components

```
ggplot(mpg, aes(displ, hwy)) + geom_point()
```

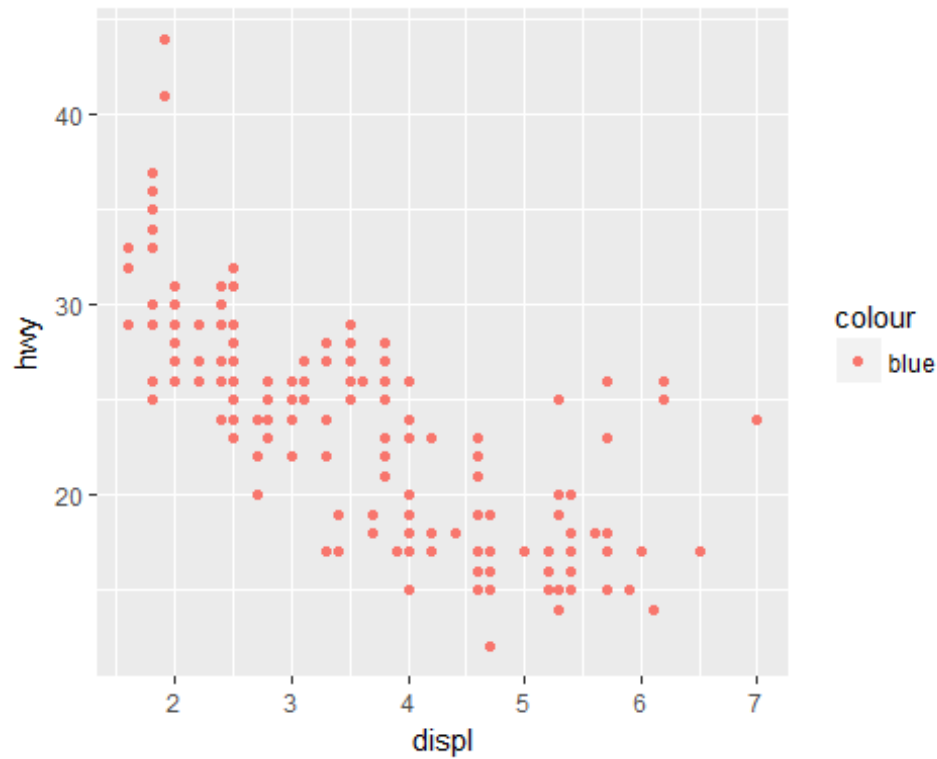


Colour, Size, Shape and Other Aesthetic Attributes

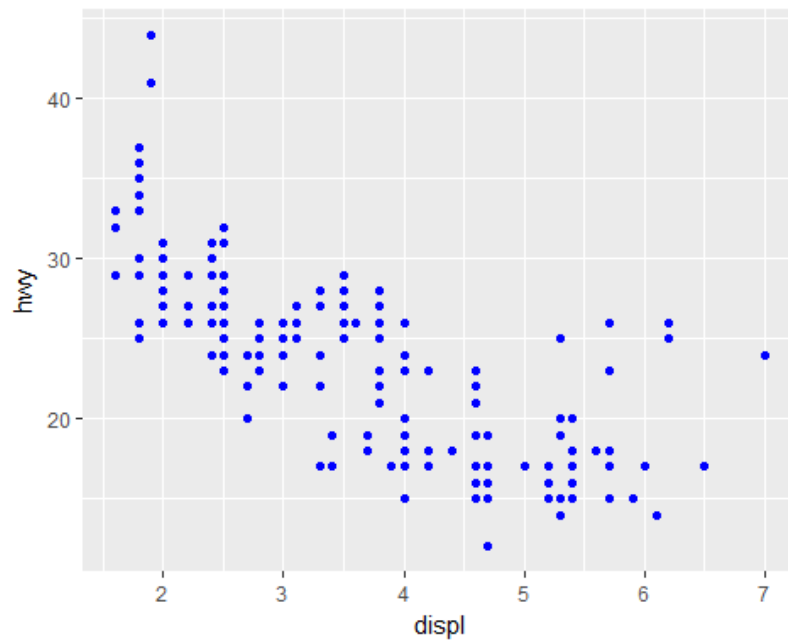
```
ggplot(mpg, aes(displ, cty, colour=class)) + geom_point()
```



```
ggplot(mpg, aes(displ, hwy)) + geom_point(aes(colour="blue"))
```

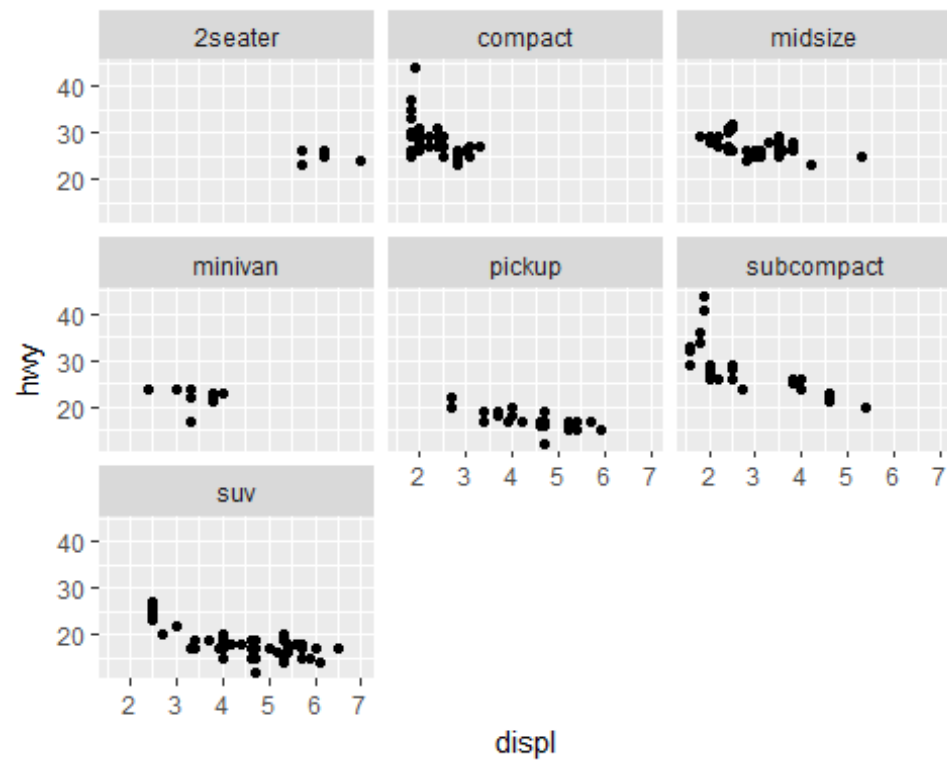


```
ggplot(mpg,aes(displ,hwy))+geom_point(colour="blue")
```



Facetting

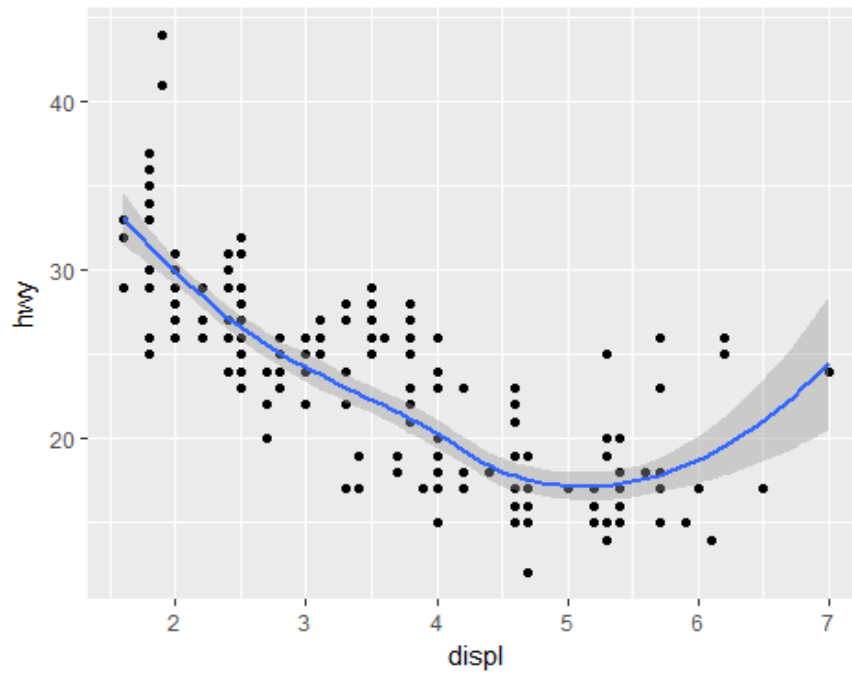
```
ggplot(mpg,aes(displ,hwy))+geom_point()+facet_wrap(~class)
```



Plot Geoms

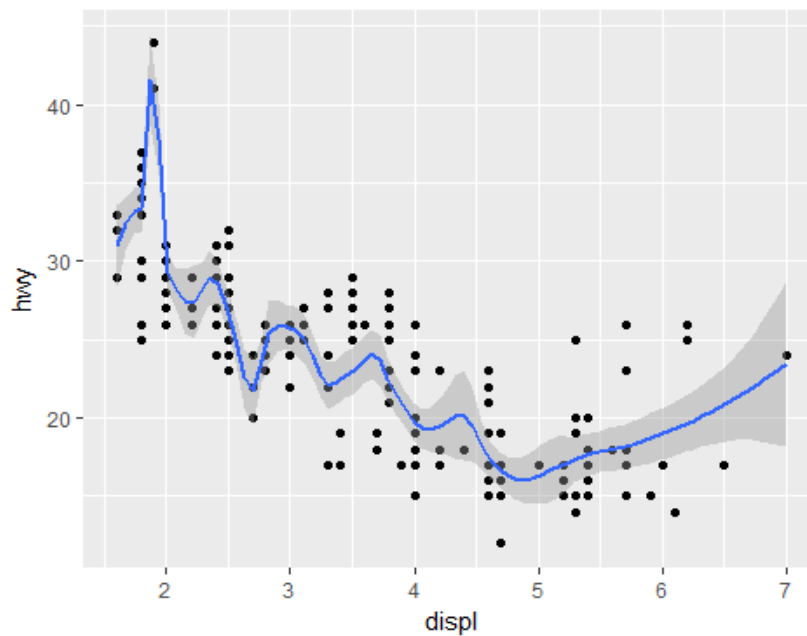
```
ggplot(mpg, aes(displ, hwy)) + geom_point() + geom_smooth()
```

```
## `geom_smooth()` using method = 'loess'
```

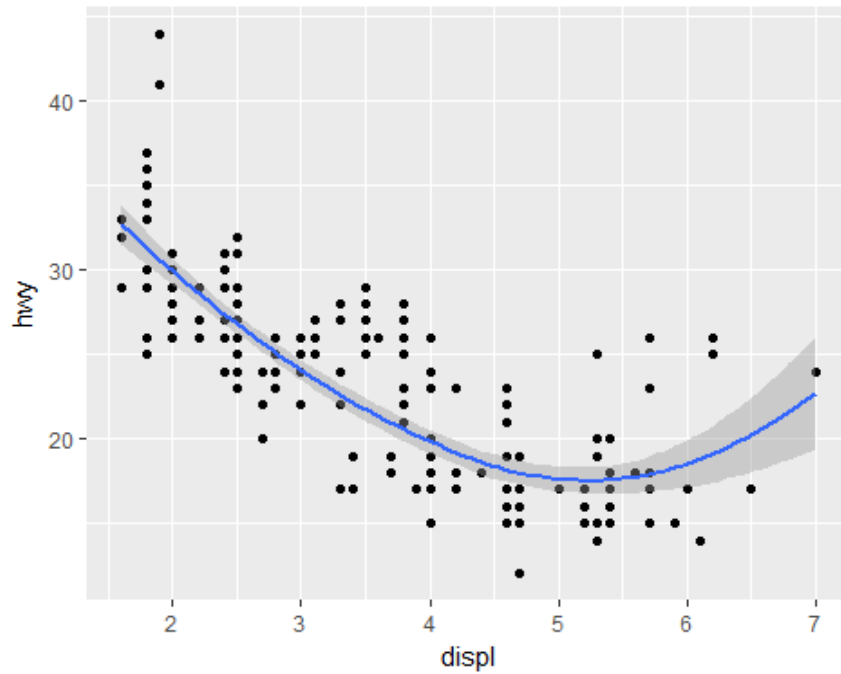


```
ggplot(mpg, aes(displ, hwy)) + geom_point() + geom_smooth(span=0.2)
```

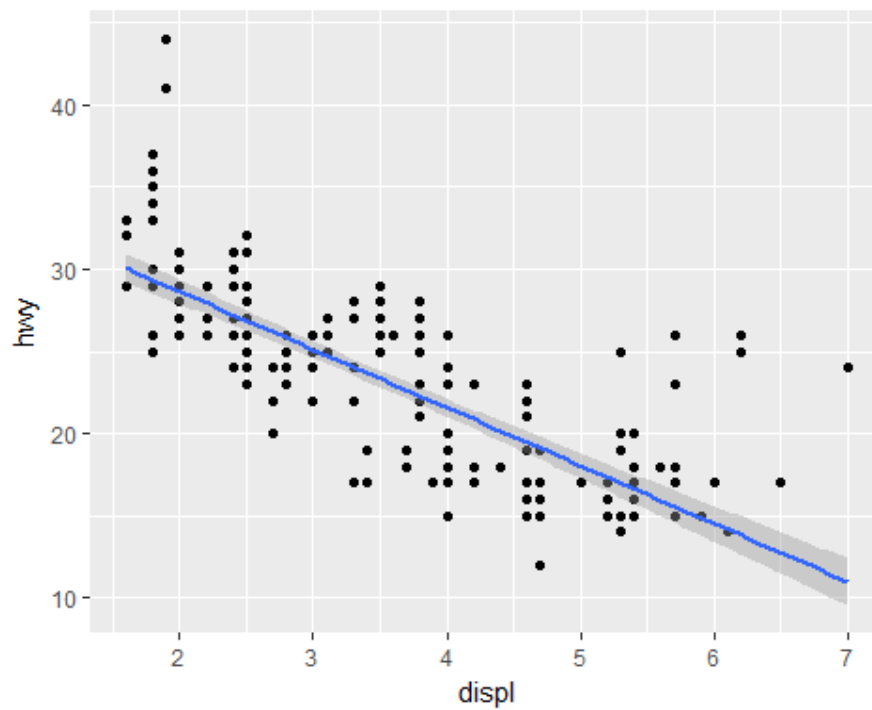
```
## `geom_smooth()` using method = 'loess'
```



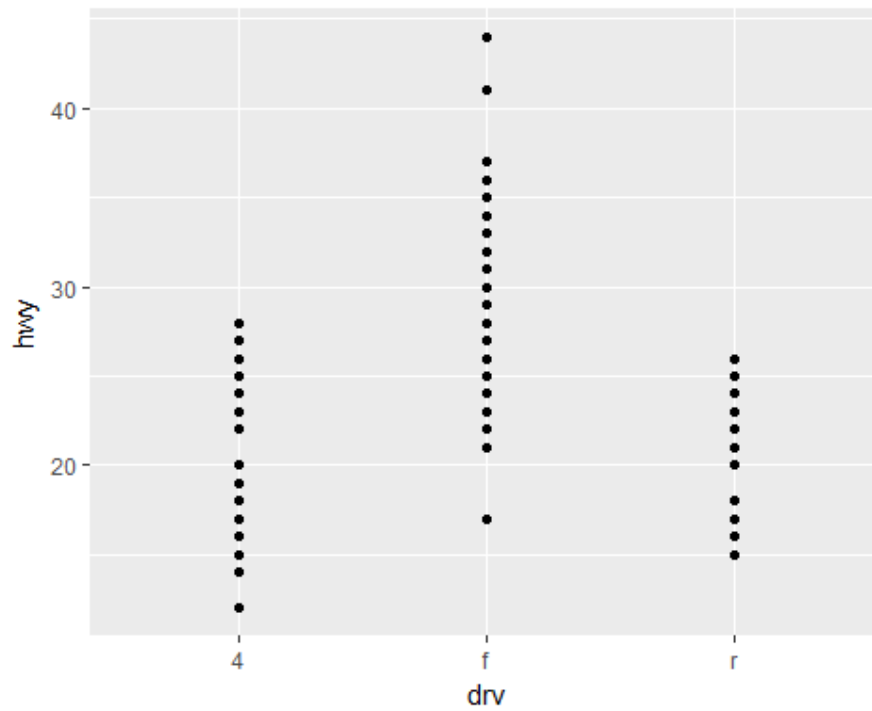
```
ggplot(mpg,aes(displ,hwy))+geom_point()+geom_smooth(span=1)
## `geom_smooth()` using method = 'loess'
```



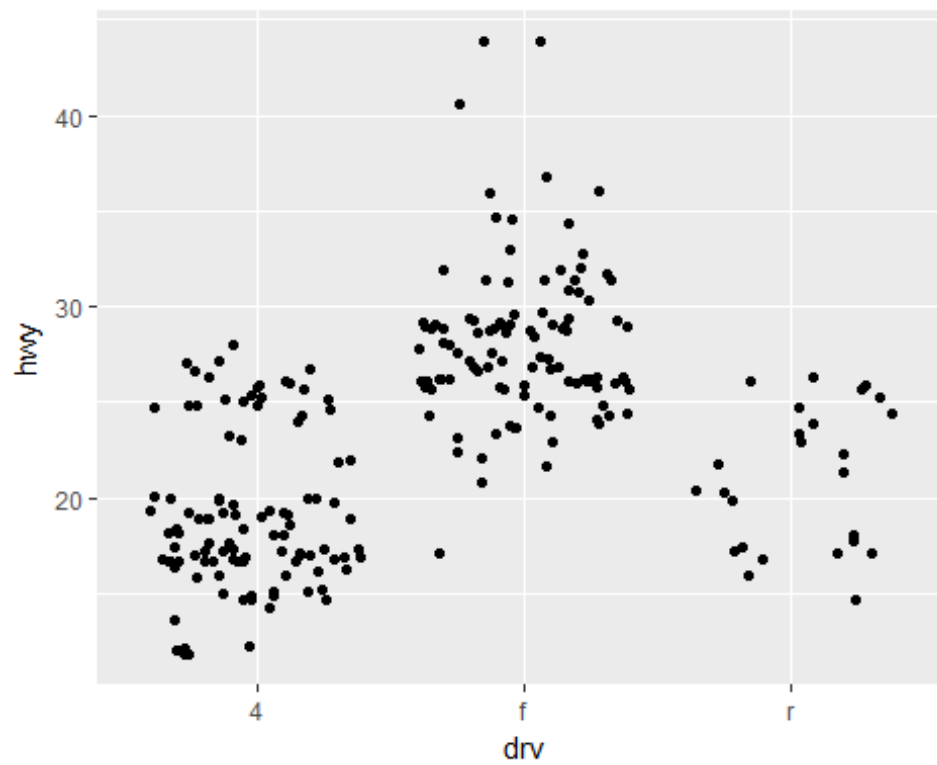
```
ggplot(mpg,aes(displ,hwy))+geom_point()+geom_smooth(method="lm")
```



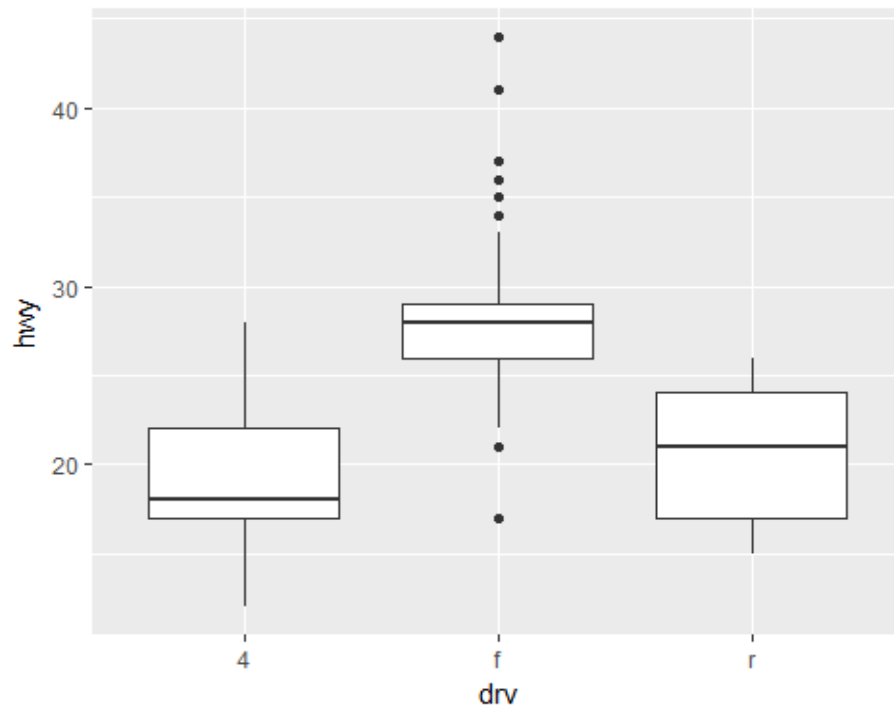
```
ggplot(mpg, aes(drv, hwy)) + geom_point()
```



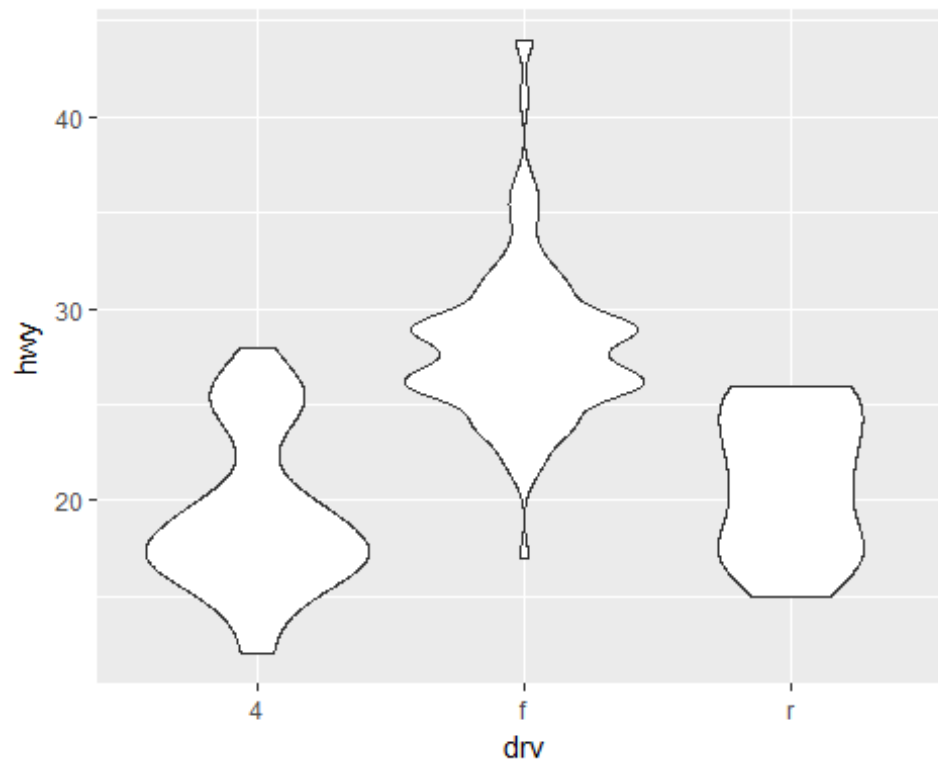
```
ggplot(mpg, aes(drv, hwy)) + geom_jitter()
```



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

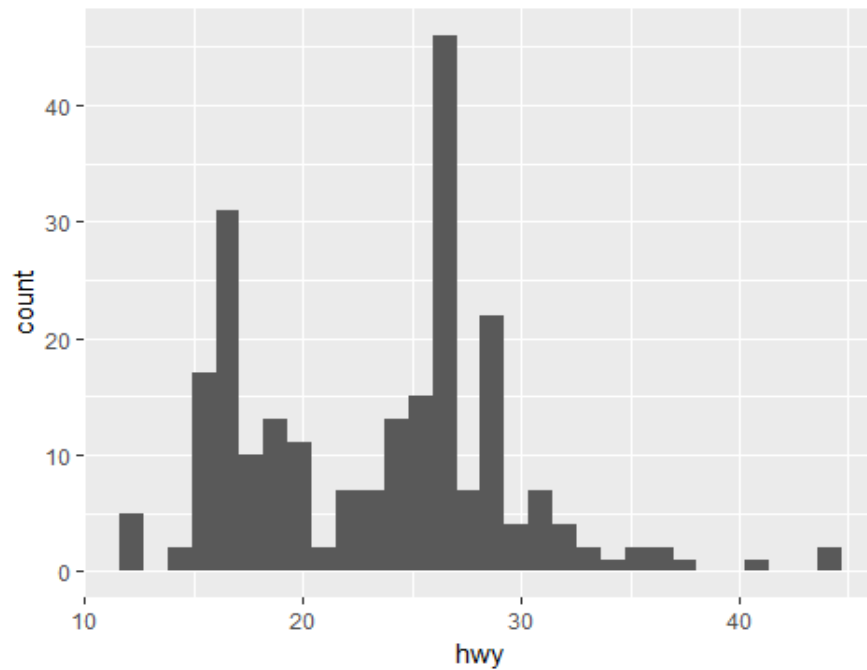


```
ggplot(mpg, aes(drv, hwy)) + geom_violin()
```



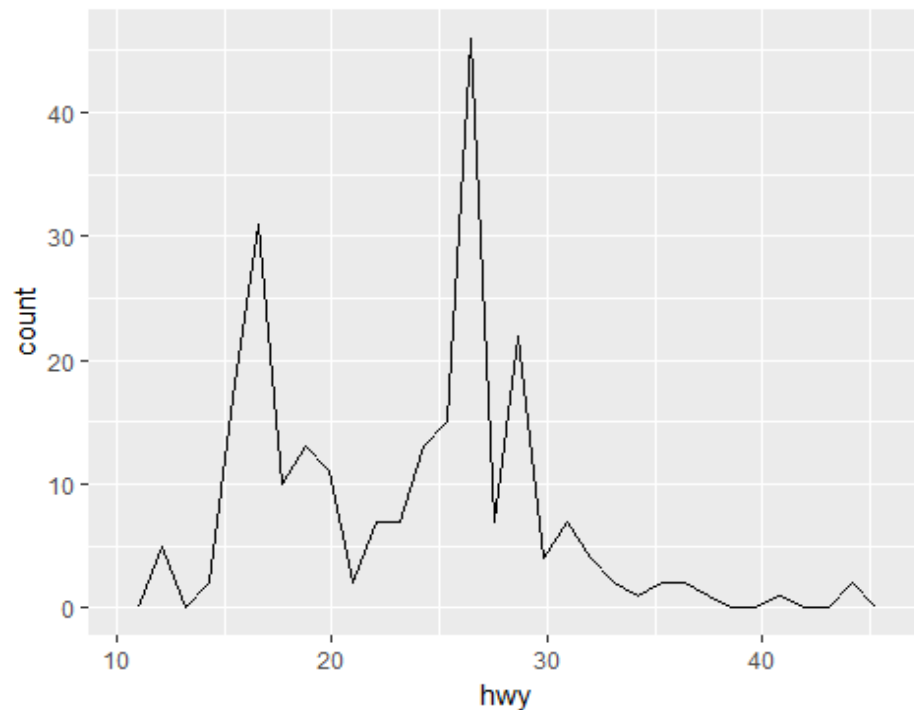
```
ggplot(mpg, aes(hwy)) + geom_histogram()
```

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

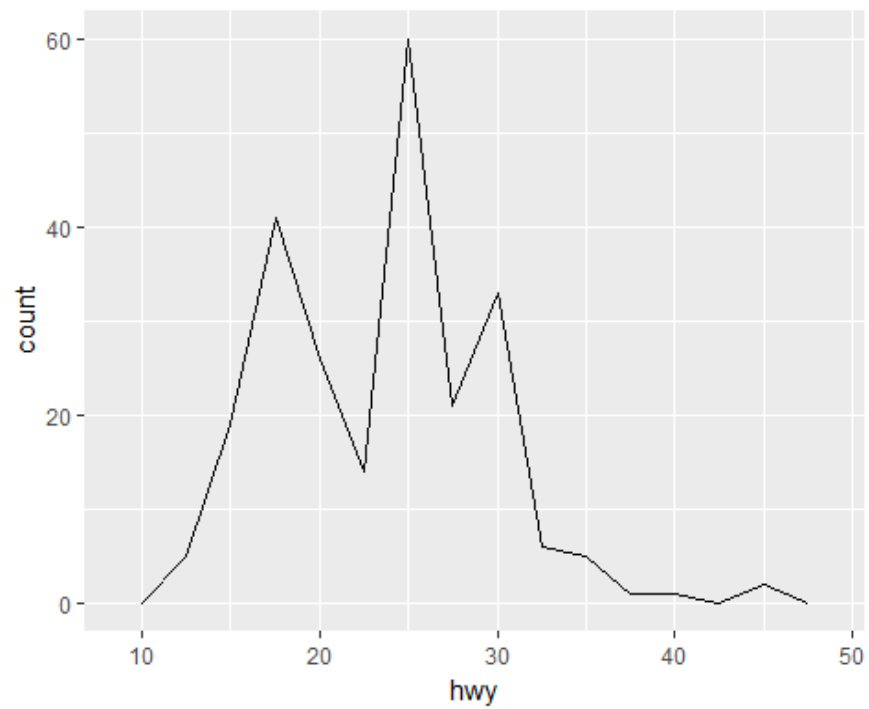


```
ggplot(mpg, aes(hwy)) + geom_freqpoly()
```

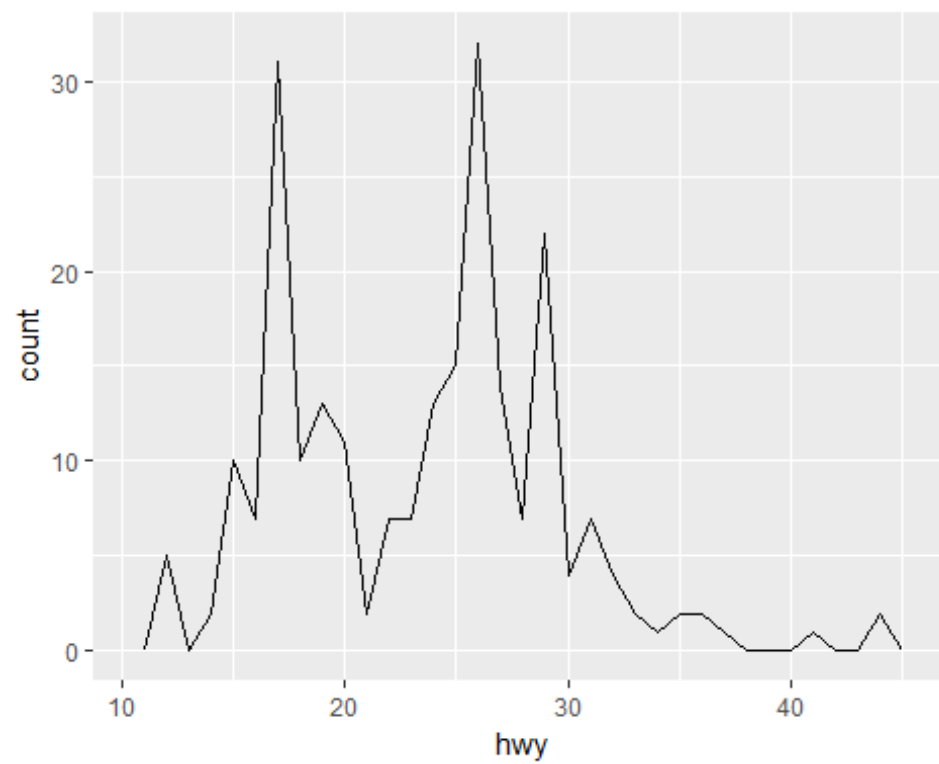
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



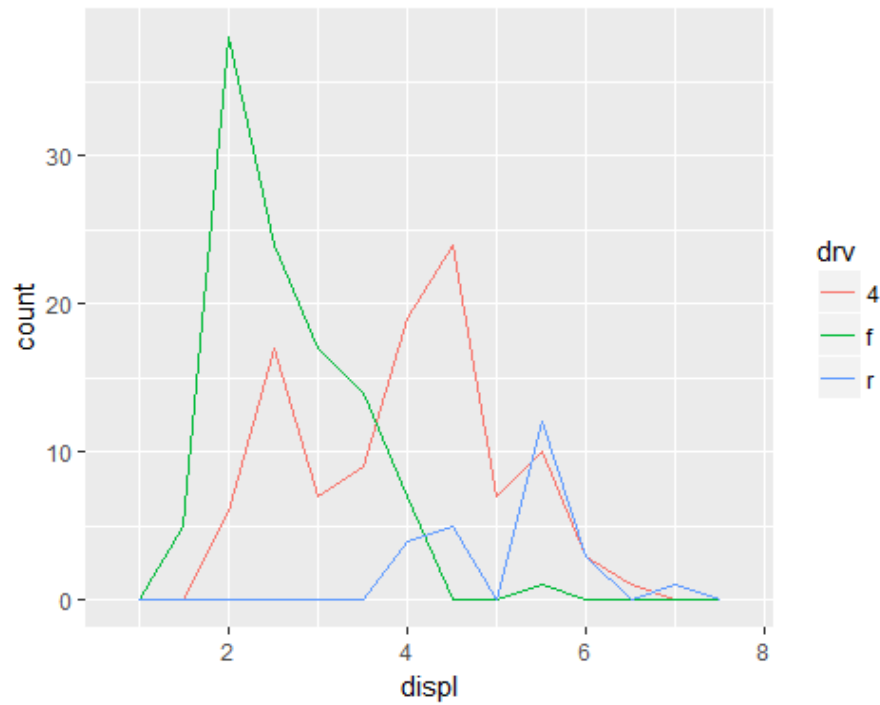

```
ggplot(mpg, aes(hwy)) + geom_freqpoly(binwidth=2.5)
```



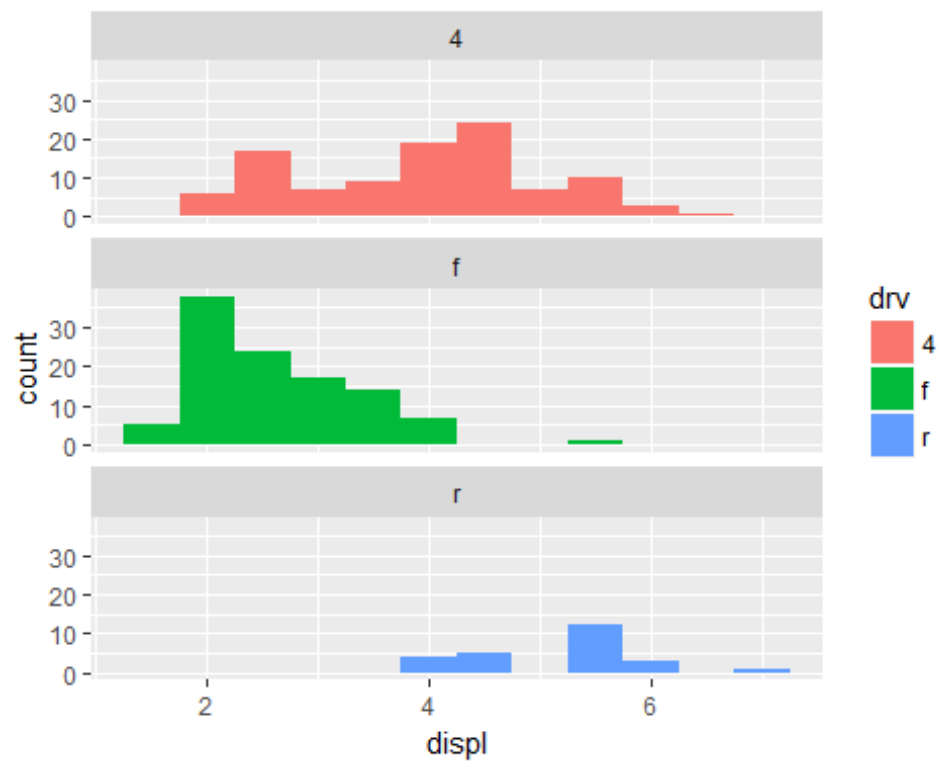
```
ggplot(mpg, aes(hwy)) + geom_freqpoly(binwidth=1)
```



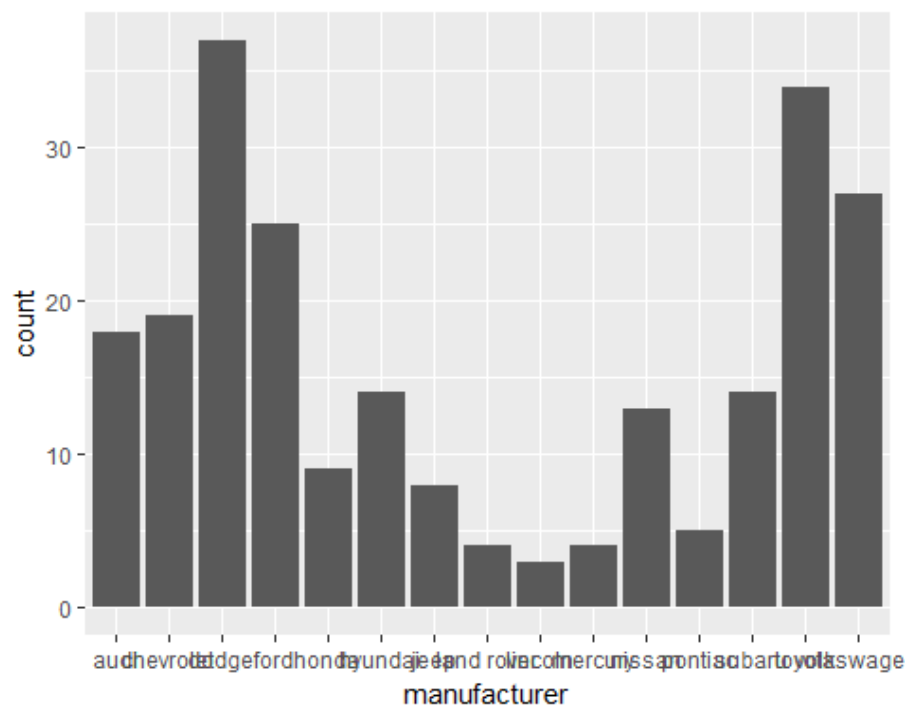
```
ggplot(mpg,aes(displ,colour=drv))+geom_freqpoly(binwidth=0.5)
```



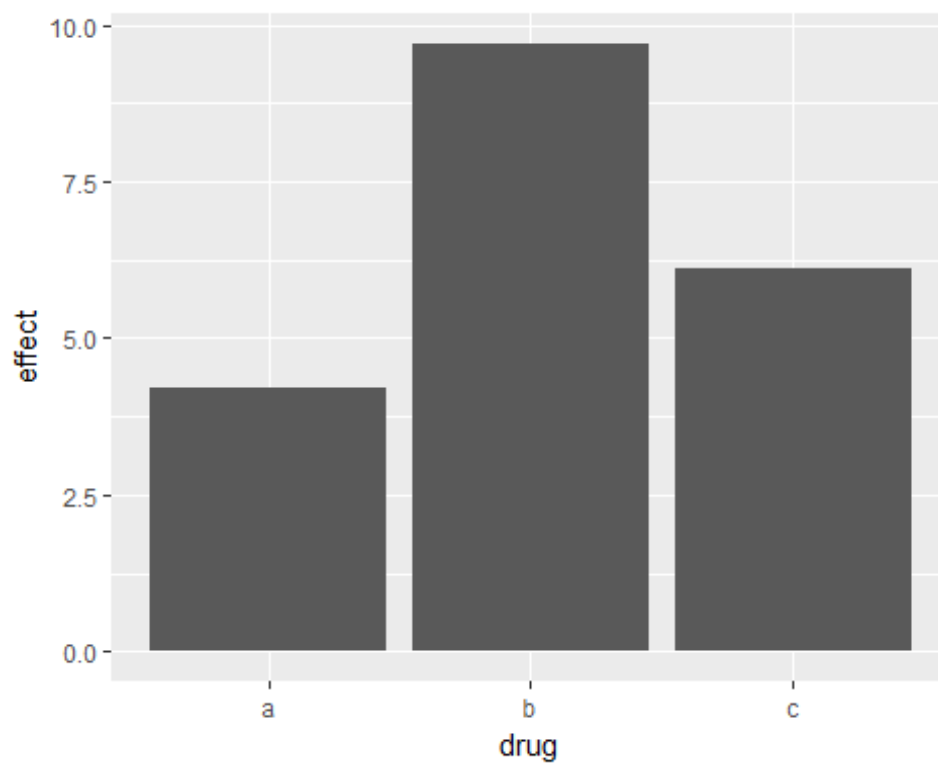
```
ggplot(mpg,aes(displ,fill=drv))+geom_histogram(binwidth=0.5)+facet_wrap(~drv,ncol=1)
```



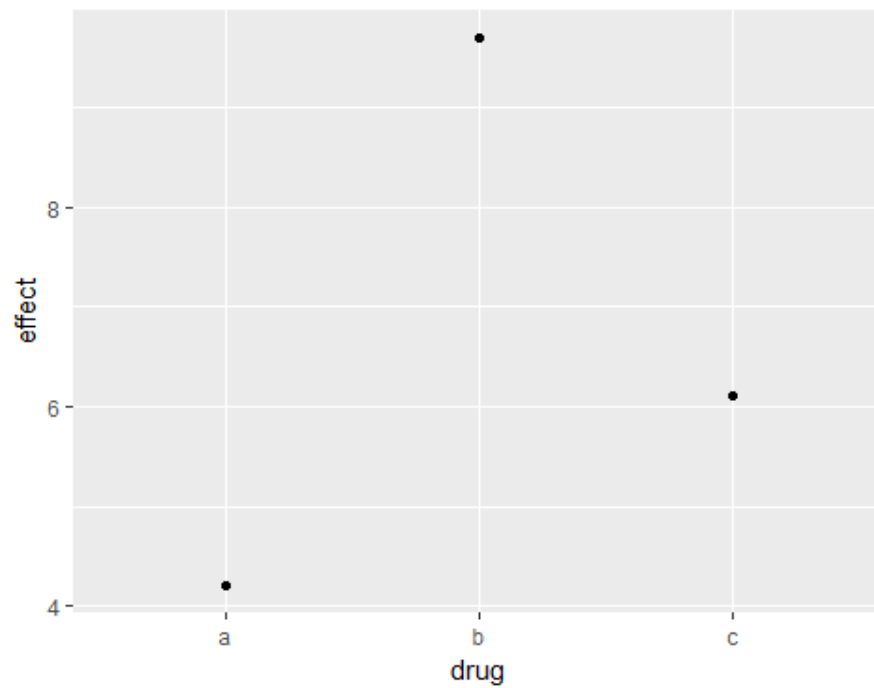
```
ggplot(mpg, aes(manufacturer)) + geom_bar()
```



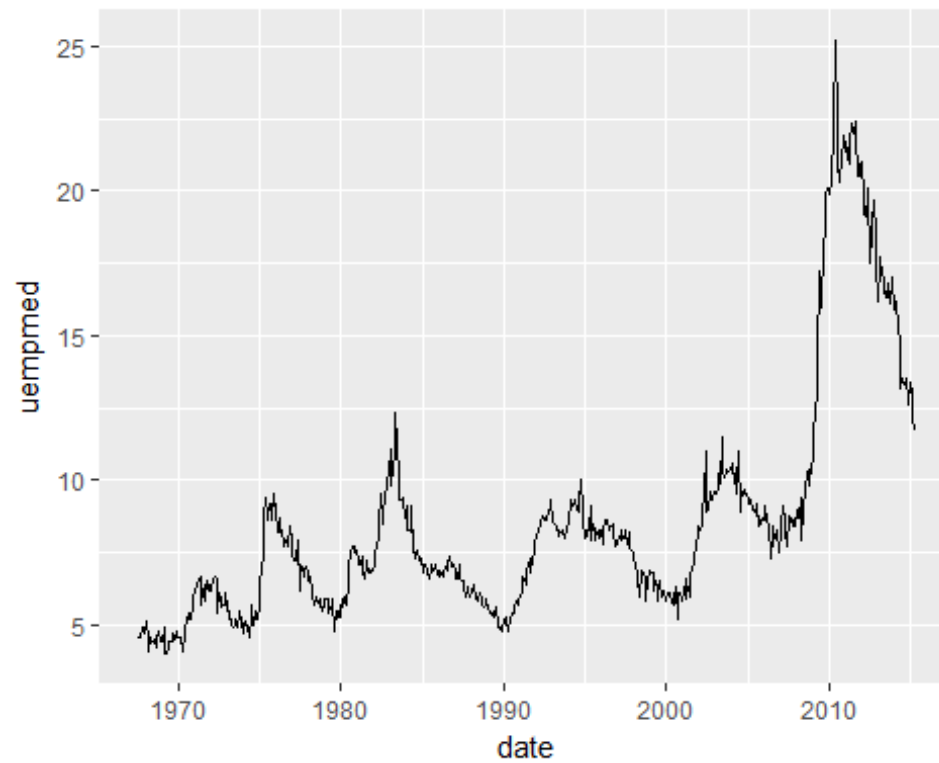
```
drugs<-data.frame(drug=c("a","b","c"),effect=c(4.2,9.7,6.1))
ggplot(drugs, aes(drug, effect)) + geom_bar(stat="identity")
```



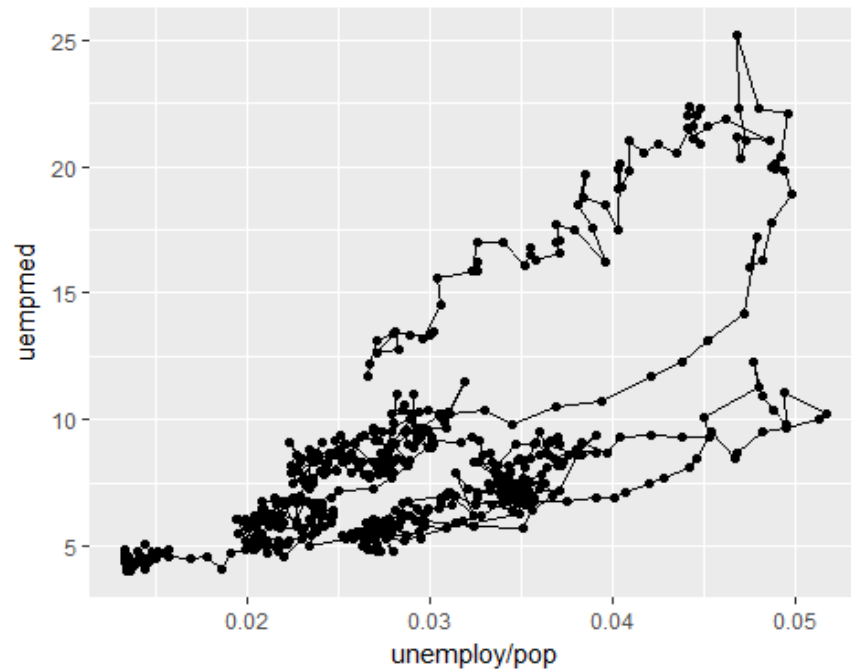
```
ggplot(drugs, aes(drug, effect)) + geom_point()
```



```
ggplot(economics, aes(date, uempmed)) + geom_line()
```

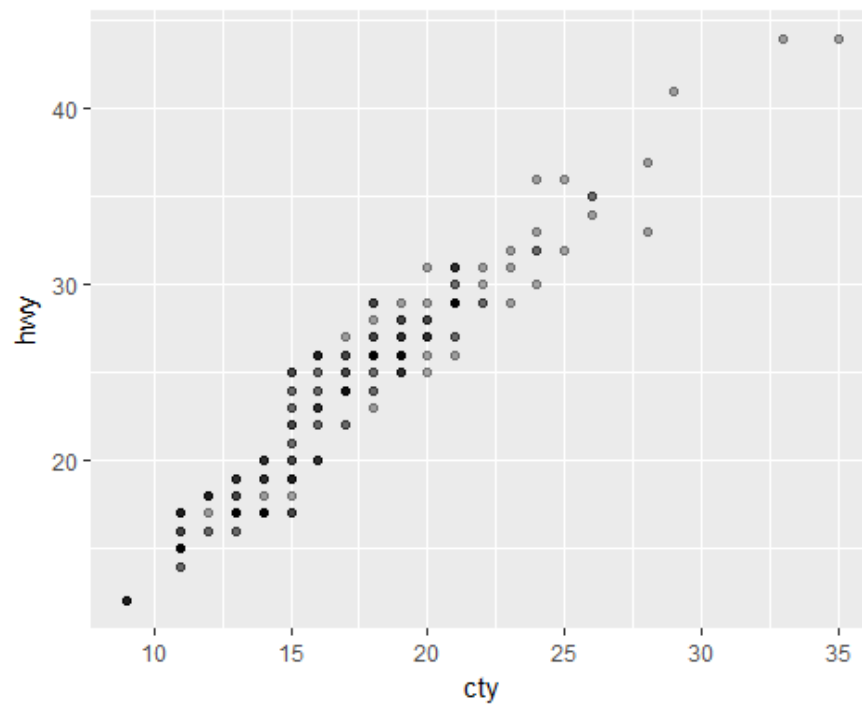


```
ggplot(economics, aes(unemploy/pop, uempmed)) + geom_path() + geom_point()
```

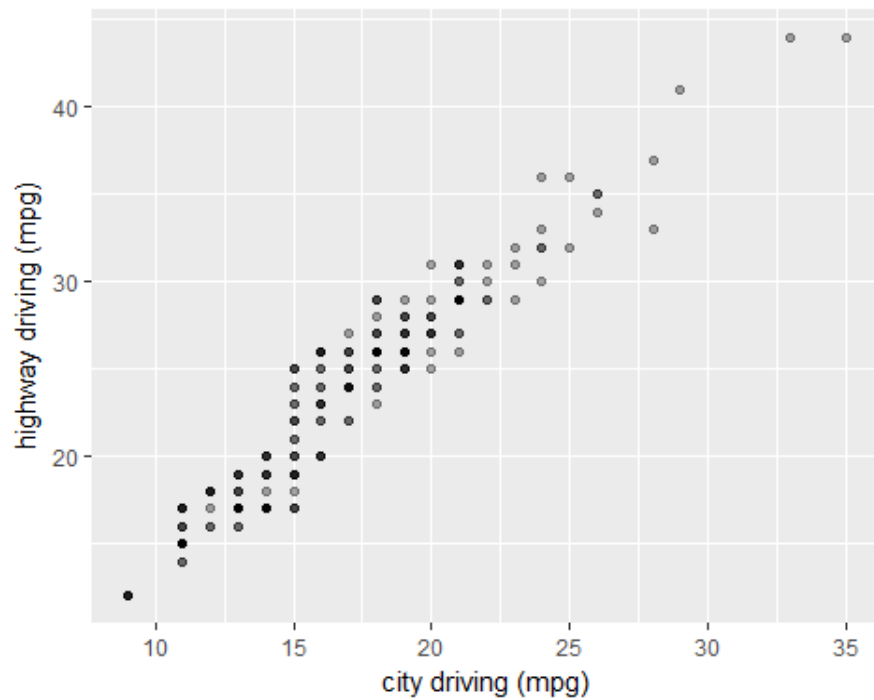


Modifying the Axes

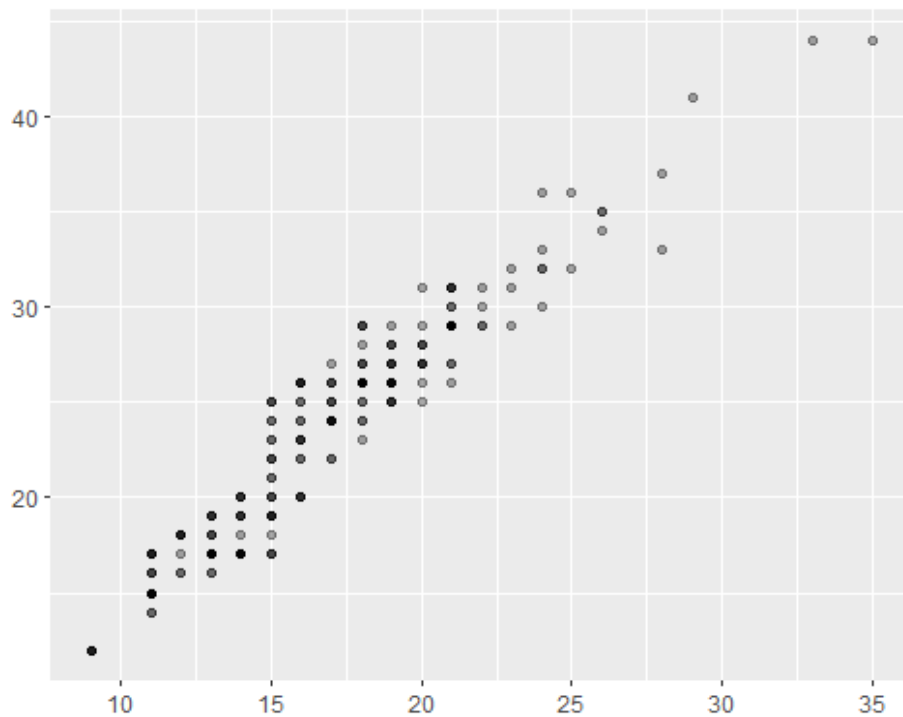
```
ggplot(mpg, aes(cty, hwy)) + geom_point(alpha=1/3)
```



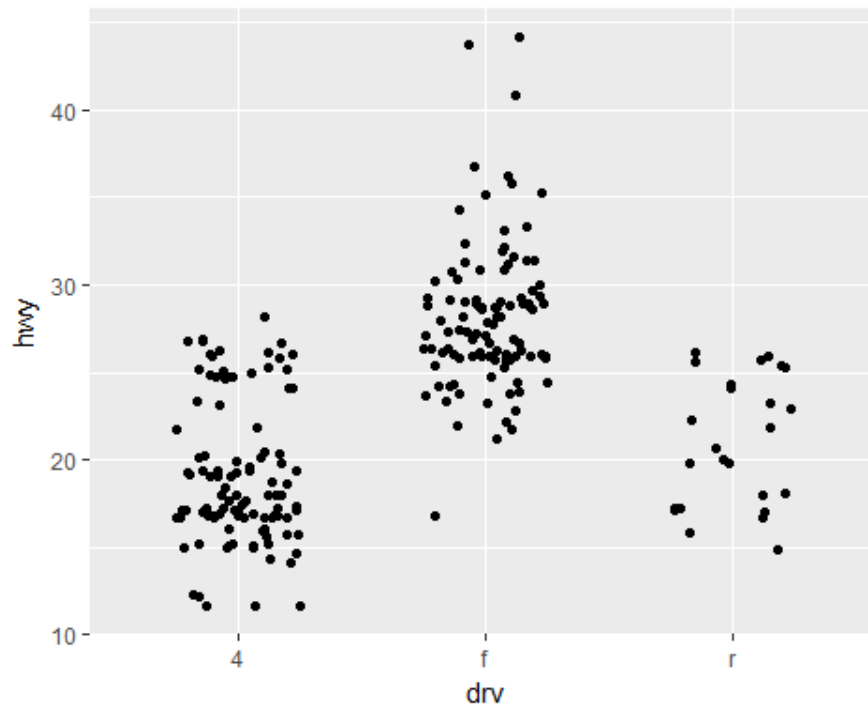
```
ggplot(mpg, aes(cty, hwy)) + geom_point(alpha=1/3) +  
  xlab("city driving (mpg)") + ylab("highway driving (mpg)")
```



```
ggplot(mpg, aes(cty, hwy)) + geom_point(alpha=1/3) +  
  xlab(NULL) + ylab(NULL)
```

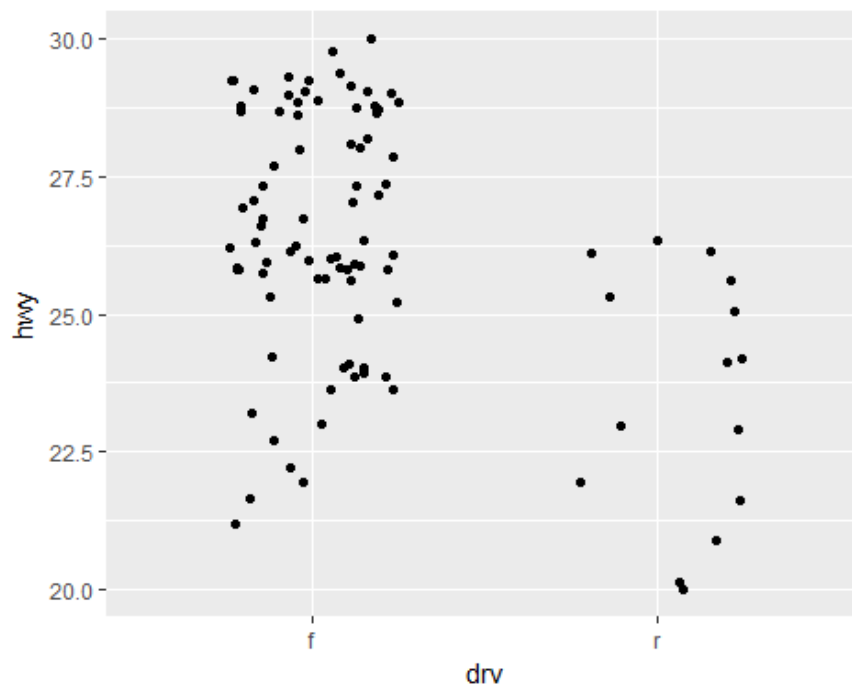


```
ggplot(mpg, aes(drv, hwy)) + geom_jitter(width=0.25)
```

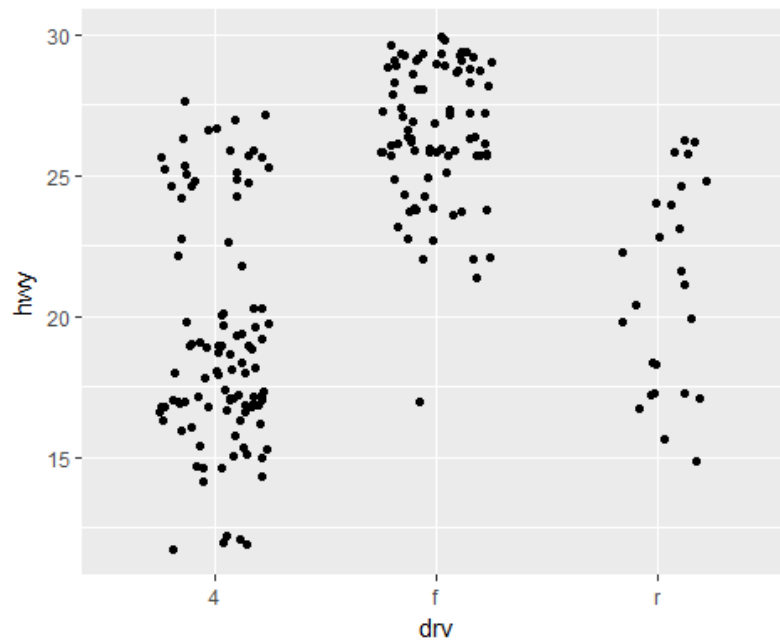


```
ggplot(mpg, aes(drv, hwy)) + geom_jitter(width=0.25) +  
  xlim("f", "r") + ylim(20, 30)
```

Warning: Removed 138 rows containing missing values (geom_point).

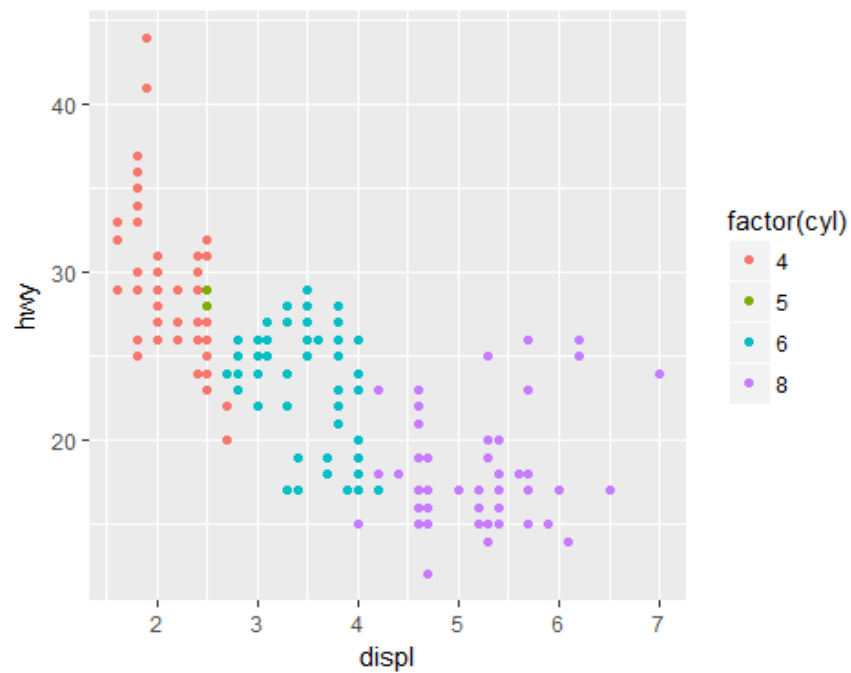


```
ggplot(mpg, aes(drv, hwy)) + geom_jitter(width=0.25, na.rm=TRUE) + ylim(NA, 30)
```



output

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



ggplot - definition of graph

```
library(ggplot2)
```

```
head(diamonds)
```

```
## # A tibble: 6 x 10
```

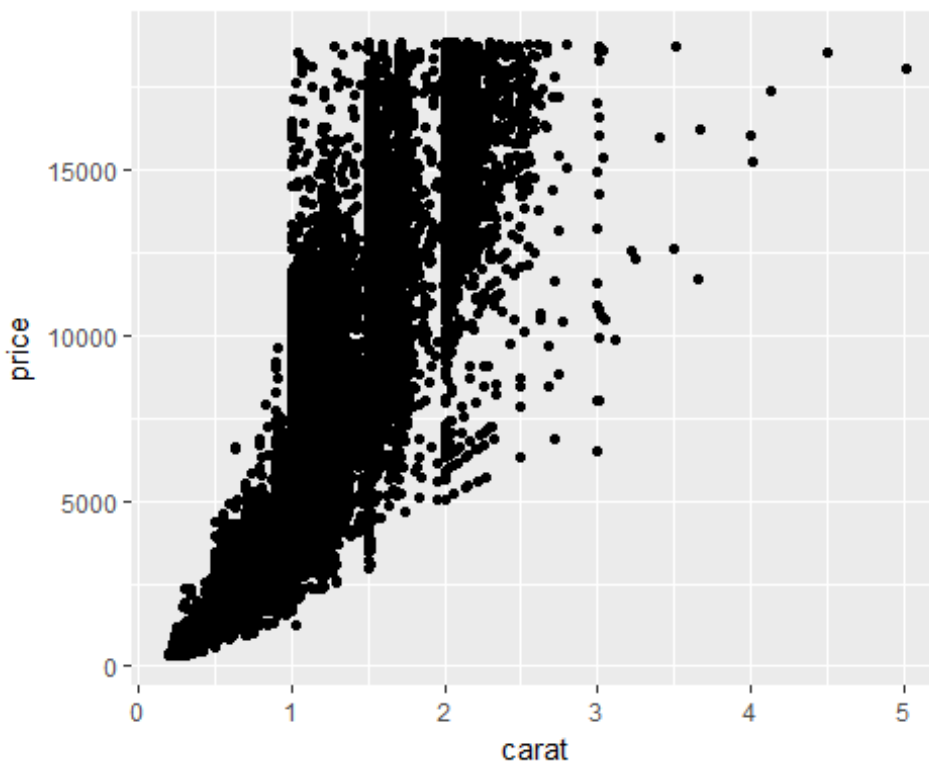
```
##   carat cut      color clarity depth table price      x      y      z
##   <dbl> <ord>    <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>
## 1 0.230 Ideal      E     SI2     61.5  55.0   326  3.95  3.98  2.43
## 2 0.210 Premium    E     SI1     59.8  61.0   326  3.89  3.84  2.31
## 3 0.230 Good       E     VS1     56.9  65.0   327  4.05  4.07  2.31
## 4 0.290 Premium    I     VS2     62.4  58.0   334  4.20  4.23  2.63
## 5 0.310 Good       J     SI2     63.3  58.0   335  4.34  4.35  2.75
## 6 0.240 Very Good J     VVS2     62.8  57.0   336  3.94  3.96  2.48
```

```
dim(diamonds)
```

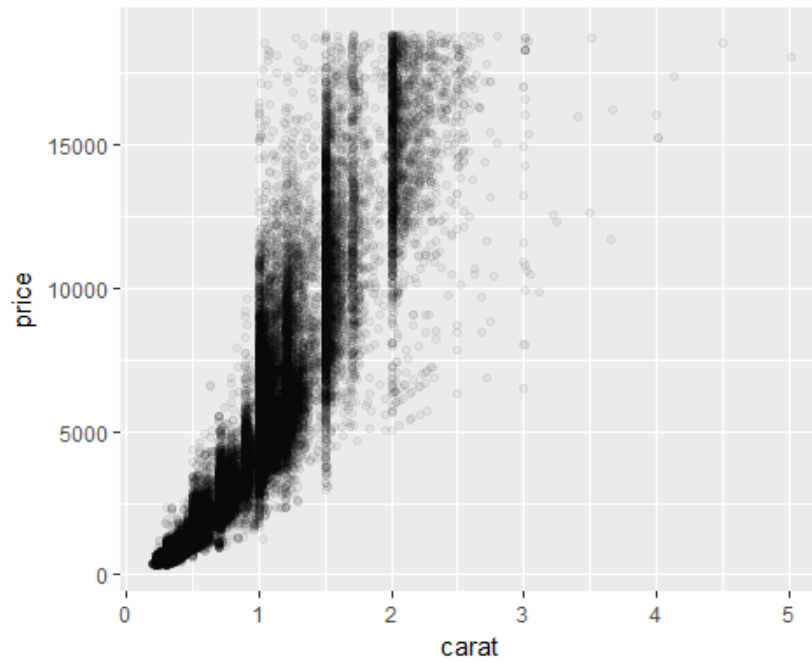
```
## [1] 53940    10
```

```
plot.basic<-ggplot(diamonds,
  aes(x=carat,
      y=price))
```

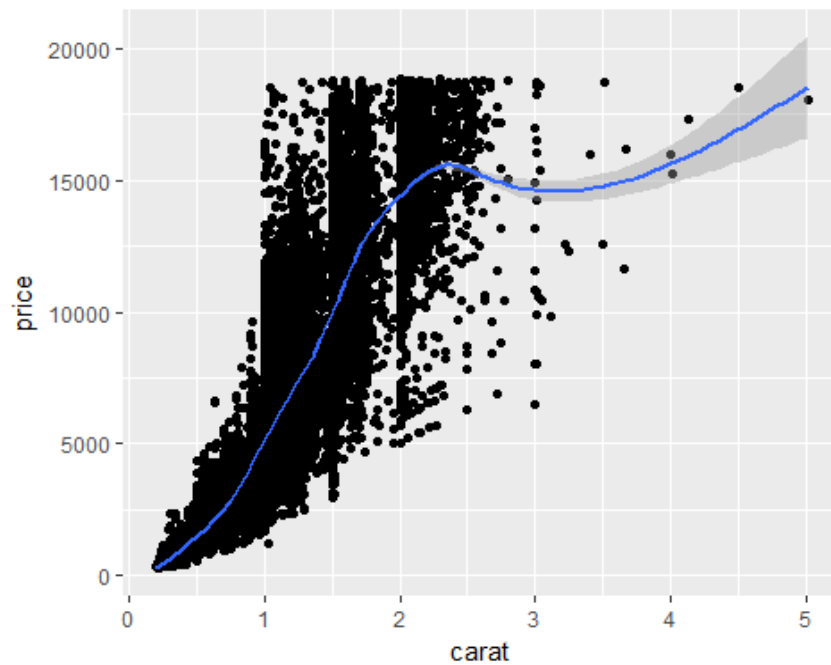
```
plot.basic+geom_point()
```



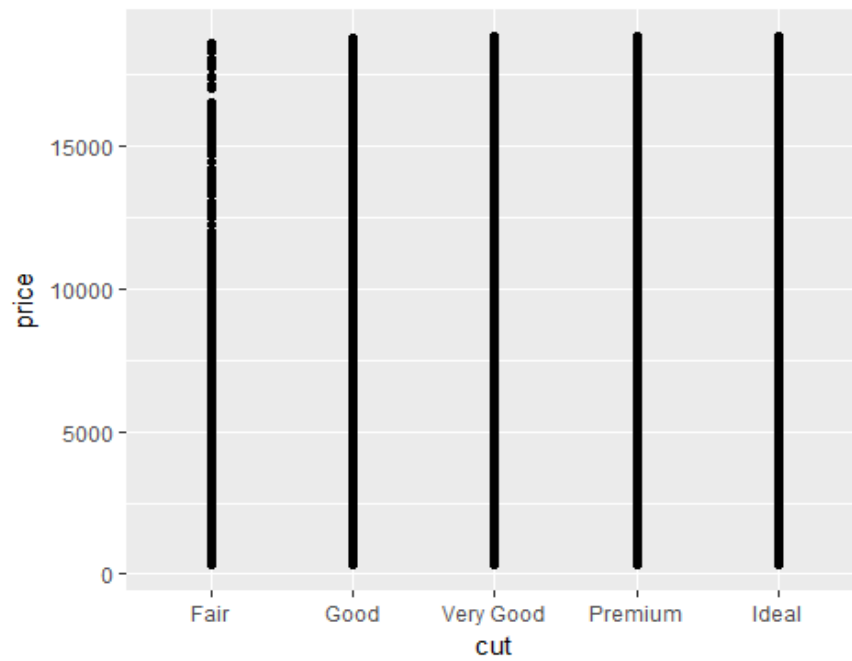
```
plot.basic+geom_point(alpha=0.05)
```



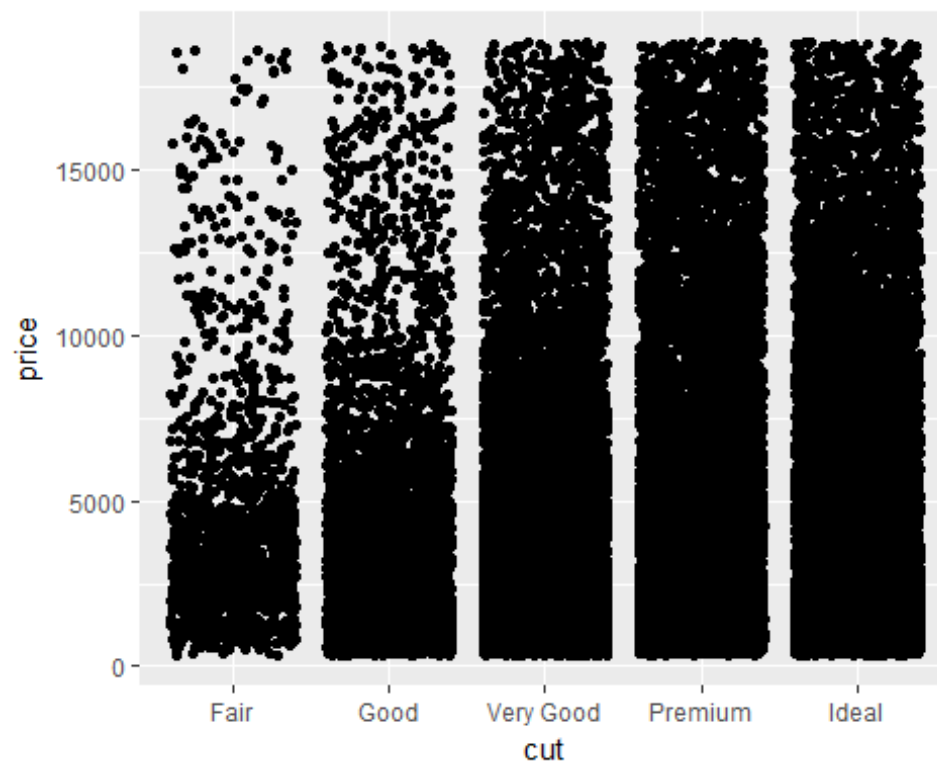
```
plot.basic+  
  geom_point()+  
  geom_smooth(span=0.2)  
## `geom_smooth()` using method = 'gam'
```



```
ggplot(diamonds,aes(cut,price))+geom_point()
```



```
ggplot(diamonds,aes(cut,price))+geom_jitter()
```



```
ggplot(diamonds,aes(cut,price))+  
  geom_jitter(alpha=0.05)
```

