syntax_comparison.R

Tracy.Shen

Sat Aug 18 11:35:31 2018

```
setwd("C:/Users/tracy.shen/Dropbox (Personal)/Data Tools learning/R exercises")
library(datasets)
data("mtcars")
# you can also just use attach(mtcars) to replace the above two lines

head(mtcars,10)
```

```
##                    mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360        14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D         24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230          22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280          19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
```

```
# ===Dollar sign Syntax: base R style====

# --summary statistics----
mean(mtcars$mpg)
```

```
## [1] 20.09062
```

```
# [1] 20.09062

table(mtcars$cyl)
```

```
##
##  4  6  8
## 11  7 14
```

```
#  4  6  8
# 11  7 14

# :::two categorical variables:
table(mtcars$cyl,mtcars$am)
```

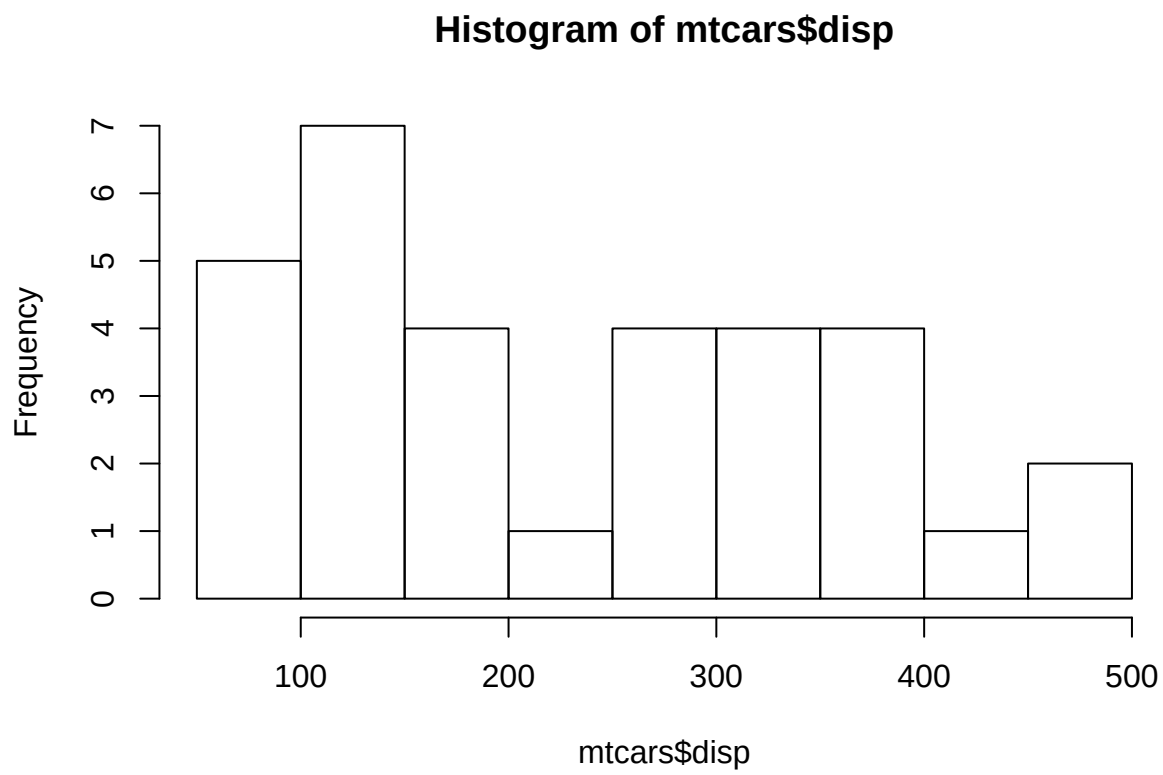
```
##
##      0  1
##   4  3  8
##   6  4  3
##   8 12  2
```

```
#    0  1
# 4  3  8
# 6  4  3
# 8 12  2
```

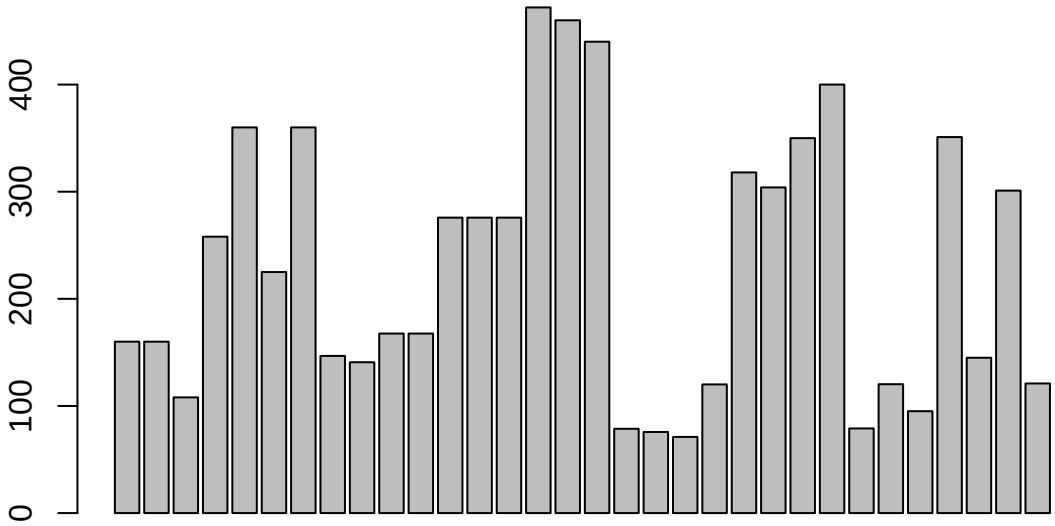
```r
# ::one continuous, one categorical::::
mean(mtcars$mpg[mtcars$cyl==4]) #conditional
```

```
## [1] 26.66364
```

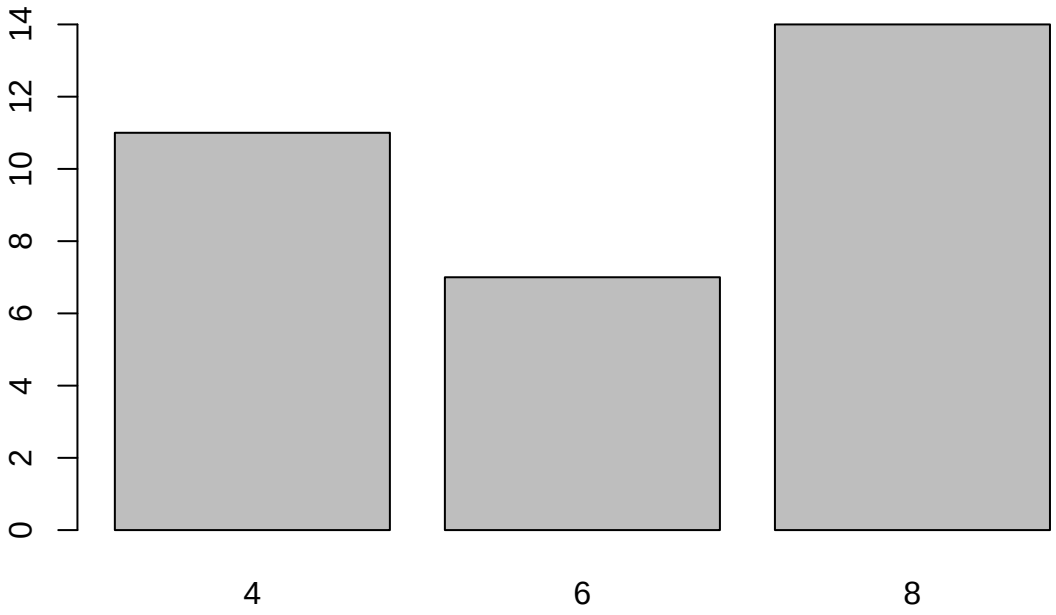
```r
#[1] 26.66364


# ---plotting:use base plot-----
hist(mtcars$disp)
```



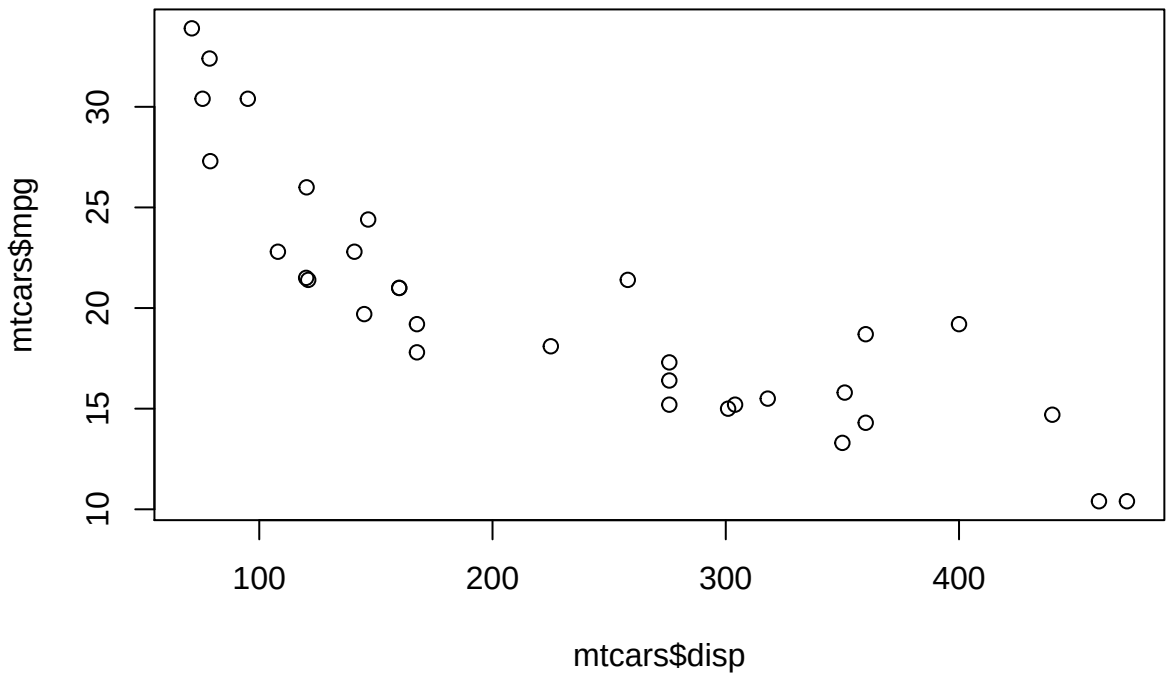

```r
barplot(mtcars$disp)
```



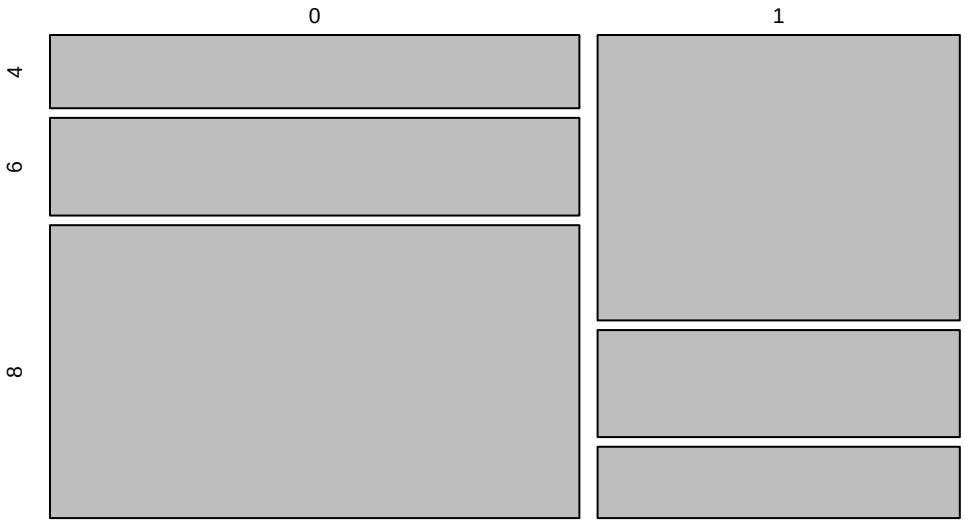

```
barplot(table(mtcars$cyl))
```



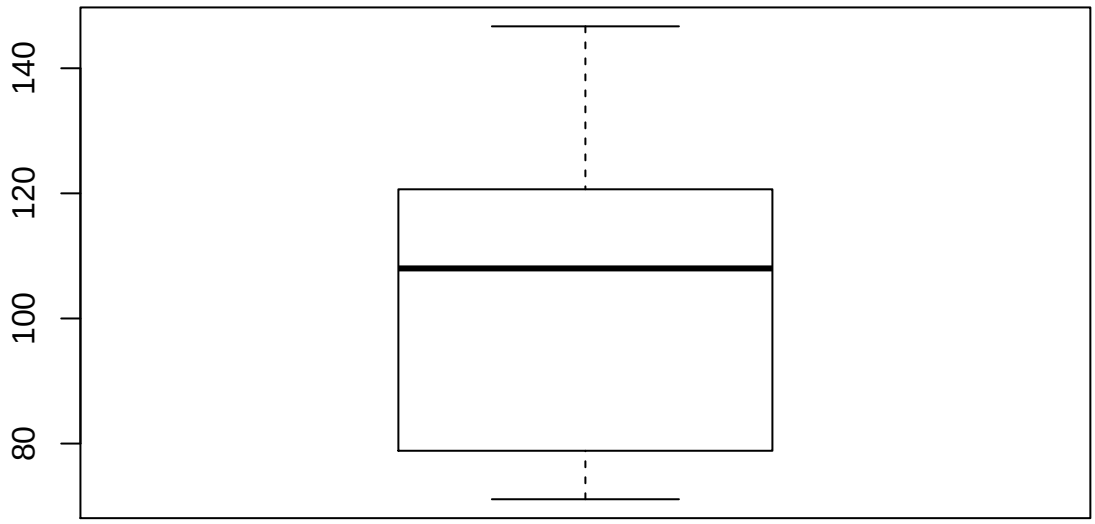

```
plot(mtcars$disp,mtcars$mpg)
```




```
# warnings()
mosaicplot(table(mtcars$am,mtcars$cyl)) #mosaic on two categorical variables
```




```
# ::one continous and one categorical:::
boxplot(mtcars$disp[mtcars$cyl==4])
```




```r
hist(mtcars$disp[mtcars$cyl==4])


# ---wrangling---
# subsetting:
mtcars[mtcars$mpg>30,]
```

```
##                 mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Fiat 128       32.4   4 78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic    30.4   4 75.7  52 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla 33.9   4 71.1  65 4.22 1.835 19.90  1  1    4    1
## Lotus Europa   30.4   4 95.1 113 3.77 1.513 16.90  1  1    5    2
```

```r
# making a new variable:
mtcars$efficient[mtcars$mpg>30]=TRUE
mtcars$efficient[mtcars$mpg<30]=FALSE
head(mtcars,10)
```

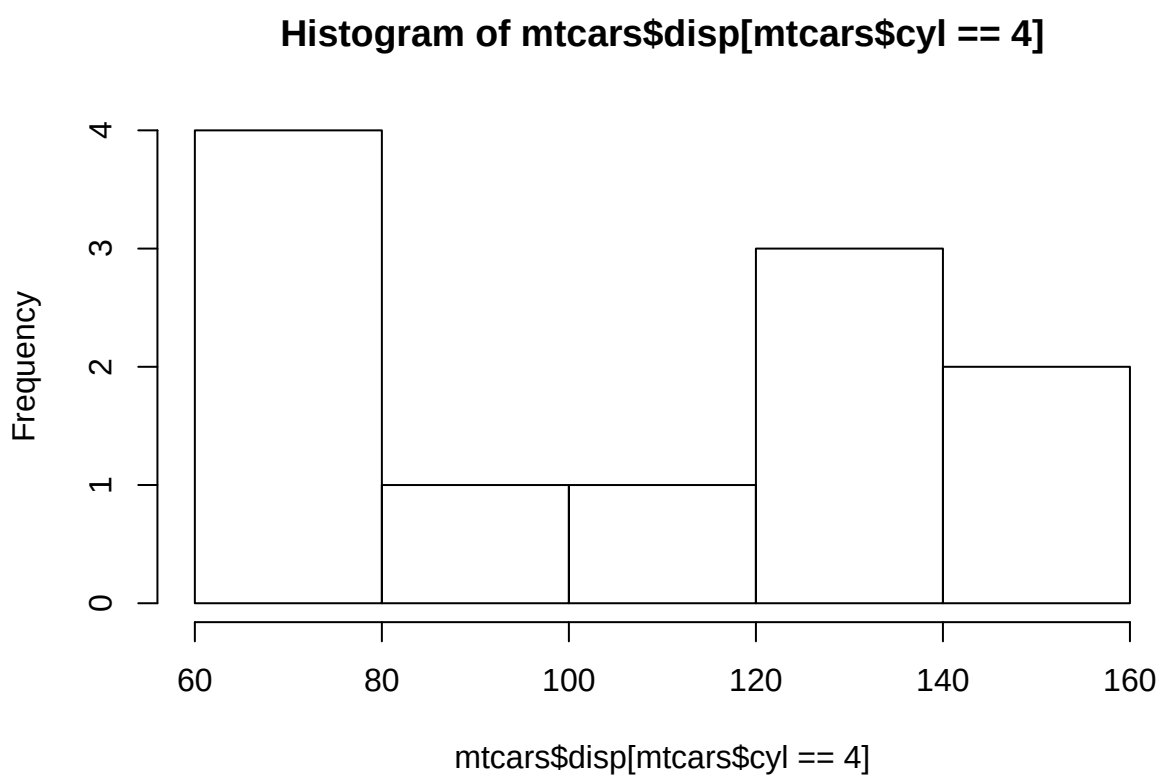
```
##                   mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360        14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D         24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
```

```
## Merc 230          22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280          19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
##                   efficient
## Mazda RX4             FALSE
## Mazda RX4 Wag         FALSE
## Datsun 710            FALSE
## Hornet 4 Drive        FALSE
## Hornet Sportabout     FALSE
## Valiant               FALSE
## Duster 360            FALSE
## Merc 240D             FALSE
## Merc 230              FALSE
## Merc 280              FALSE
```

```
# ===========Formula Syntax: used in models, using lattice plotting=====

library(mosaic)
```

```
## Warning: package 'mosaic' was built under R version 3.5.1
## Loading required package: dplyr
## Warning: package 'dplyr' was built under R version 3.5.1
##
## 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
## Loading required package: lattice
## Loading required package: ggformula
## Warning: package 'ggformula' was built under R version 3.5.1
## Loading required package: ggplot2
## Warning: package 'ggplot2' was built under R version 3.5.1
## Loading required package: ggstance
## Warning: package 'ggstance' was built under R version 3.5.1
##
## Attaching package: 'ggstance'
## The following objects are masked from 'package:ggplot2':
##
##     geom_errorbarh, GeomErrorbarh
##
## New to ggformula?  Try the tutorials:
##  learnr::run_tutorial("introduction", package = "ggformula")
##  learnr::run_tutorial("refining", package = "ggformula")
## Loading required package: mosaicData
```

```
## Warning: package 'mosaicData' was built under R version 3.5.1
## Loading required package: Matrix
##
## The 'mosaic' package masks several functions from core packages in order to add
## additional features.  The original behavior of these functions should not be affected by this.
##
## Note: If you use the Matrix package, be sure to load it BEFORE loading mosaic.
##
## Attaching package: 'mosaic'
## The following object is masked from 'package:Matrix':
##
##     mean
## The following object is masked from 'package:ggplot2':
##
##     stat
## The following objects are masked from 'package:dplyr':
##
##     count, do, tally
## The following objects are masked from 'package:stats':
##
##     binom.test, cor, cor.test, cov, fivenum, IQR, median,
##     prop.test, quantile, sd, t.test, var
## The following objects are masked from 'package:base':
##
##     max, mean, min, prod, range, sample, sum
```




```
mean(~mpg,data=mtcars) #[1] 20.09062
```

```
## [1] 20.09062
```

```
tally(~cyl,data = mtcars)
```

```
## cyl
##  4  6  8
## 11  7 14
```

```
#  4  6  8
# 11  7 14

tally(cyl~am,data = mtcars)
```

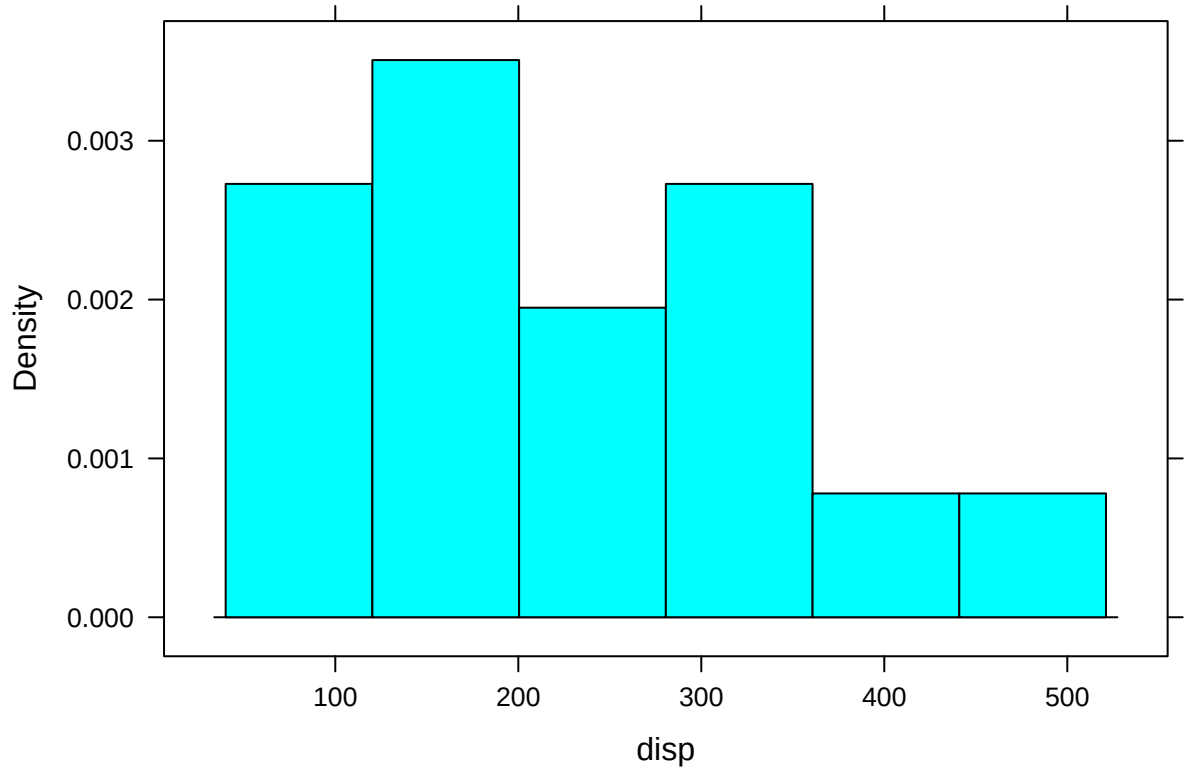
```
##    am
## cyl  0  1
##   4  3  8
##   6  4  3
##   8 12  2
```

```
# #    am
# cyl  0  1
# 4  3  8
# 6  4  3
# 8 12  2

mean(mpg~cyl,data=mtcars)
```

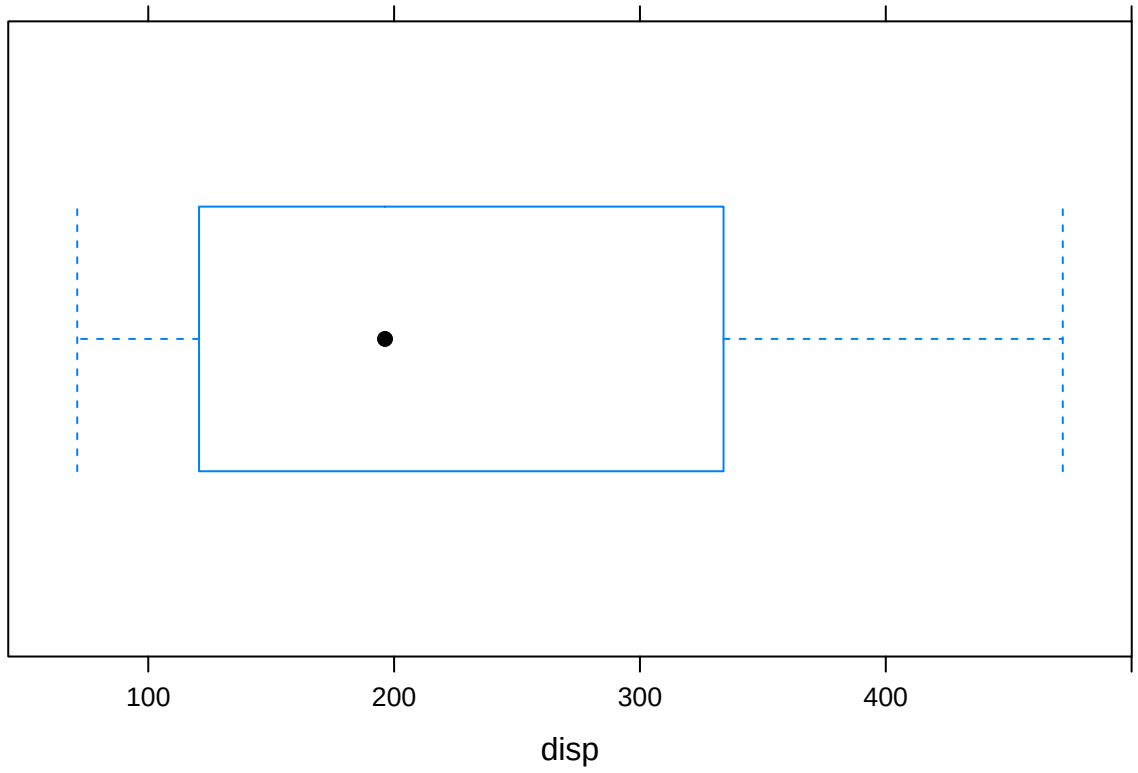
```
##        4        6        8
## 26.66364 19.74286 15.10000
```

```
#        4        6        8
# 26.66364 19.74286 15.10000
```

```
library(lattice)
histogram(~disp,data=mtcars)
```

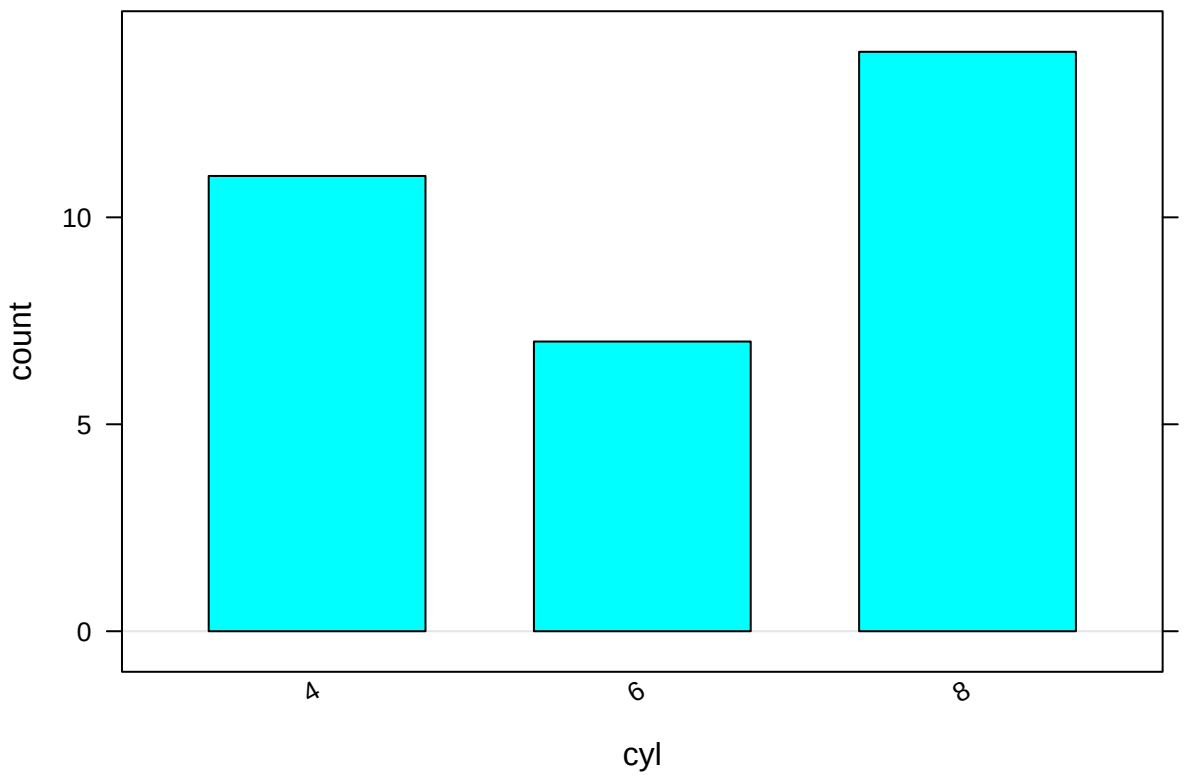



```
bwplot(~disp,data = mtcars)
```

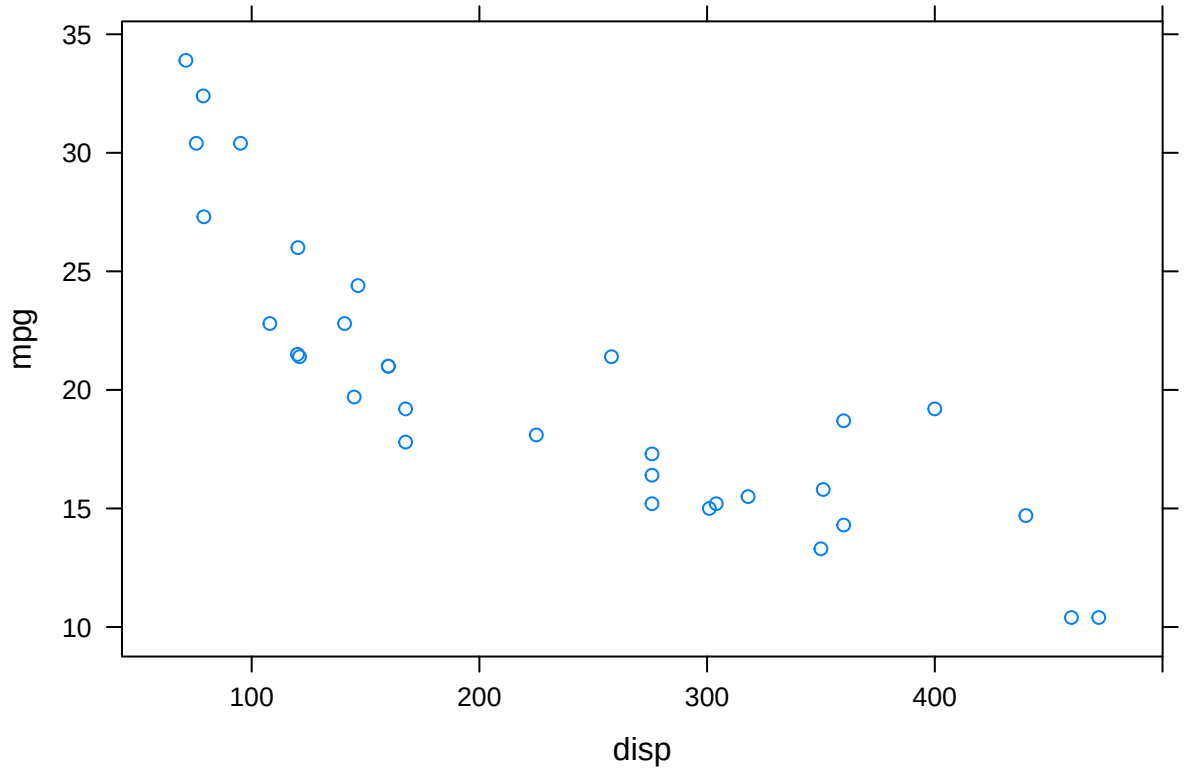



```
bargraph(~cyl,data = mtcars)
```

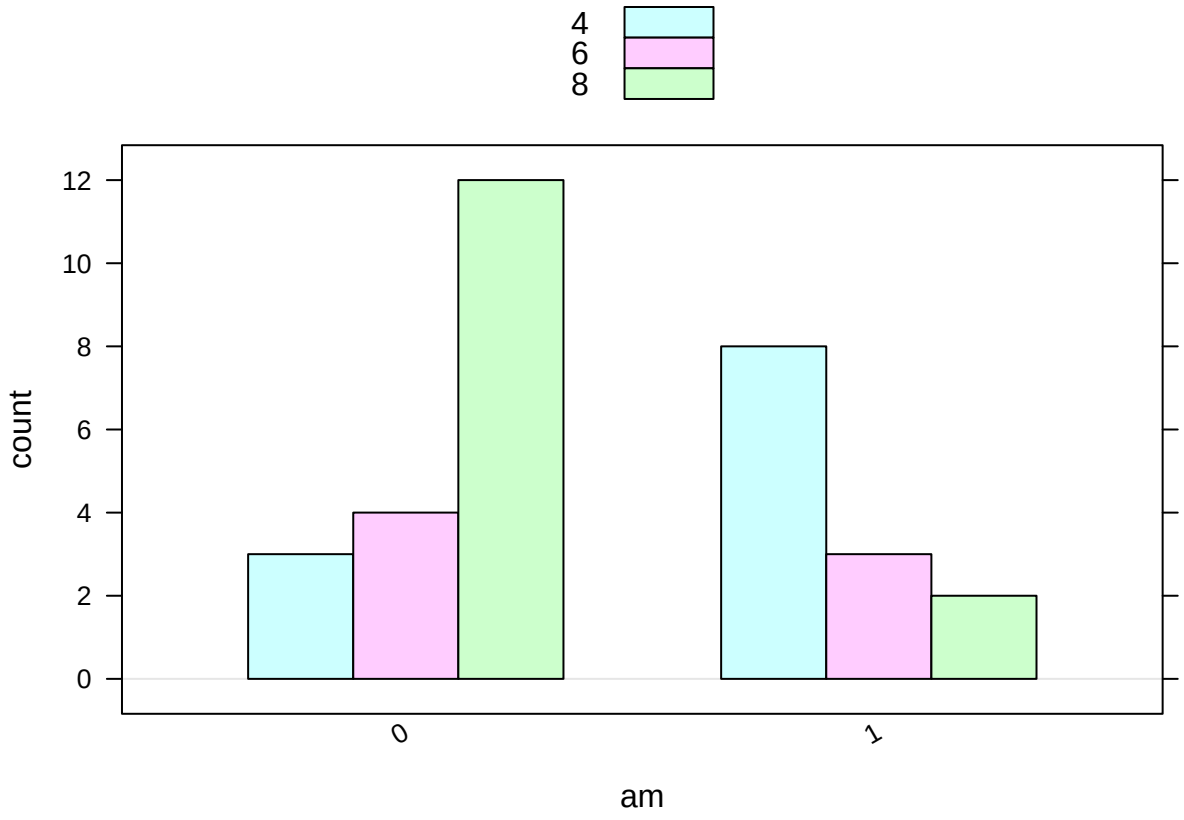
```
## Warning: package 'bindrcpp' was built under R version 3.5.1
```



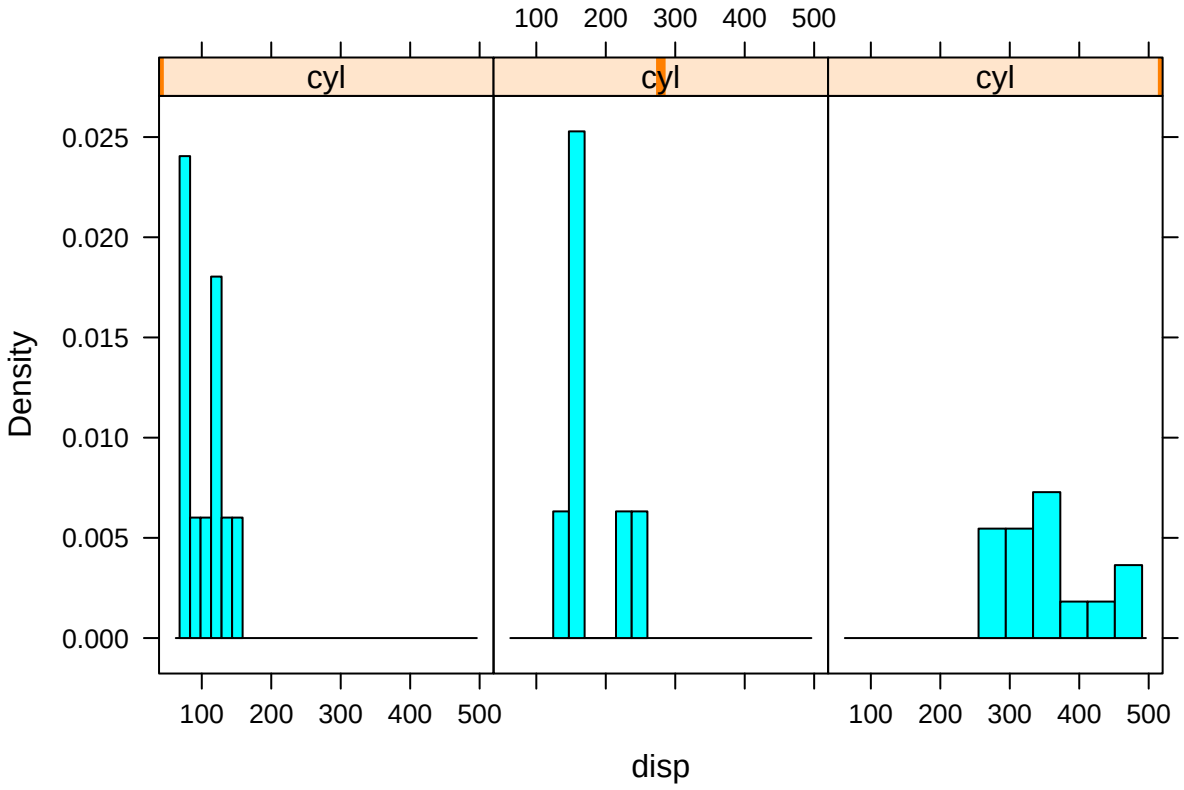

```
xyplot(mpg~disp,data = mtcars)
```



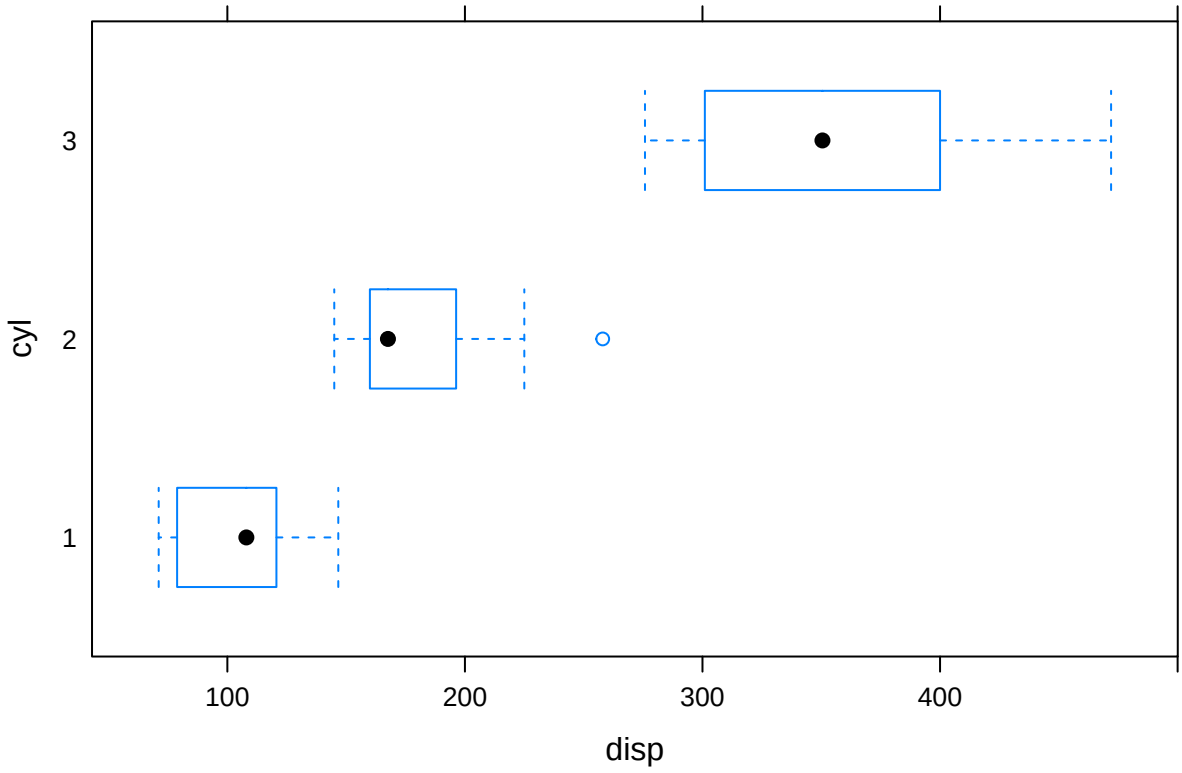

```
# two categorical variables

bargraph(~am,data=mtcars,groups = cyl)
```




```
histogram(~disp|cyl,data = mtcars)
```




```
bwplot(cyl~disp,data = mtcars)
```




```
# =====tidyverse syntax====
mtcars%>% summarize(mean(mpg))
```

```
##   mean(mpg)
## 1  20.09062
```

```
#   mean(mpg)
# 1  20.09062
# one categorical variable:
mtcars%>%group_by(cyl)%>%summarize(n())
```

```
## # A tibble: 3 x 2
##     cyl `n()`
##   <dbl> <int>
## 1     4    11
## 2     6     7
## 3     8    14
```

```
# two categorical variablse
mtcars%>%group_by(cyl,am)%>%summarize(n())
```

```
## # A tibble: 6 x 3
## # Groups:   cyl [?]
##     cyl    am `n()`
##   <dbl> <dbl> <int>
## 1     4     0     3
## 2     4     1     8
## 3     6     0     4
```

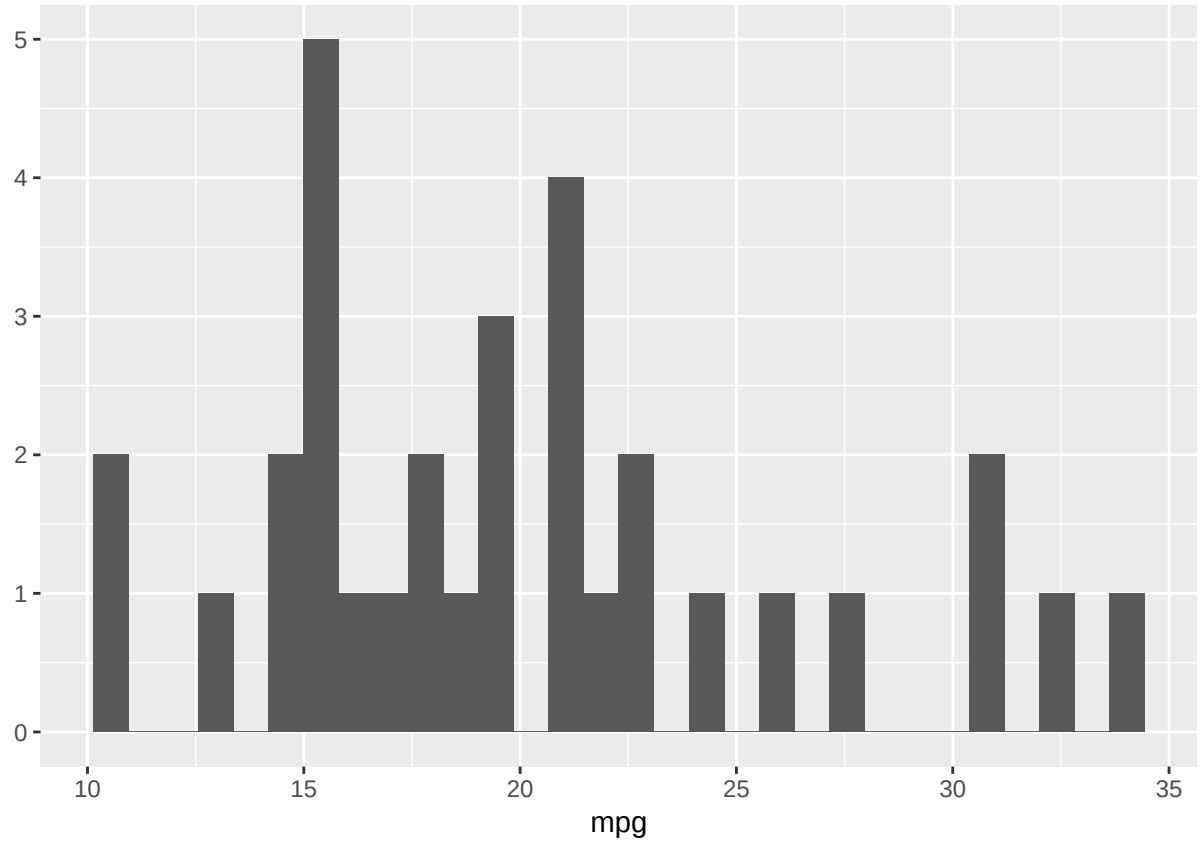
```
## 4     6     1     3
## 5     8     0    12
## 6     8     1     2
```

```
# # A tibble: 6 x 3
# # Groups:   cyl [?]
# cyl    am `n()`
# <dbl> <dbl> <int>
#   1     4     0     3
# 2     4     1     8
# 3     6     0     4
# 4     6     1     3
# 5     8     0    12
# 6     8     1     2

# one continuous, one categorical
mtcars%>%group_by(cyl)%>%summarize(mean(mpg))
```

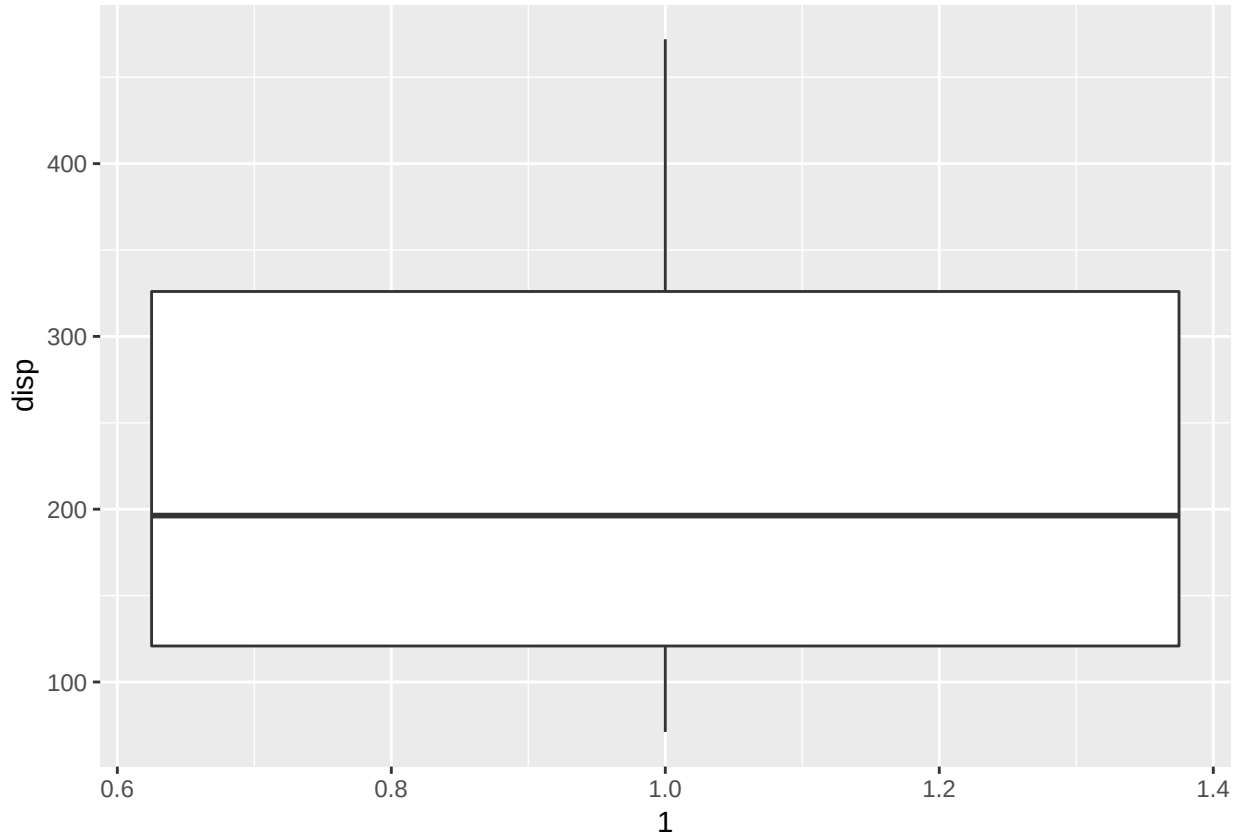
```
## # A tibble: 3 x 2
##     cyl `mean(mpg)`
##   <dbl>       <dbl>
## 1     4        26.7
## 2     6        19.7
## 3     8        15.1
```

```
# # # A tibble: 3 x 2
# cyl `mean(mpg)`
#      <dbl>       <dbl>
# 1     4        26.7
# 2     6        19.7
# 3     8        15.1

# ---ggplot plotting----
# one continous variable
qplot(x=mpg,data = mtcars,geom='histogram')
```

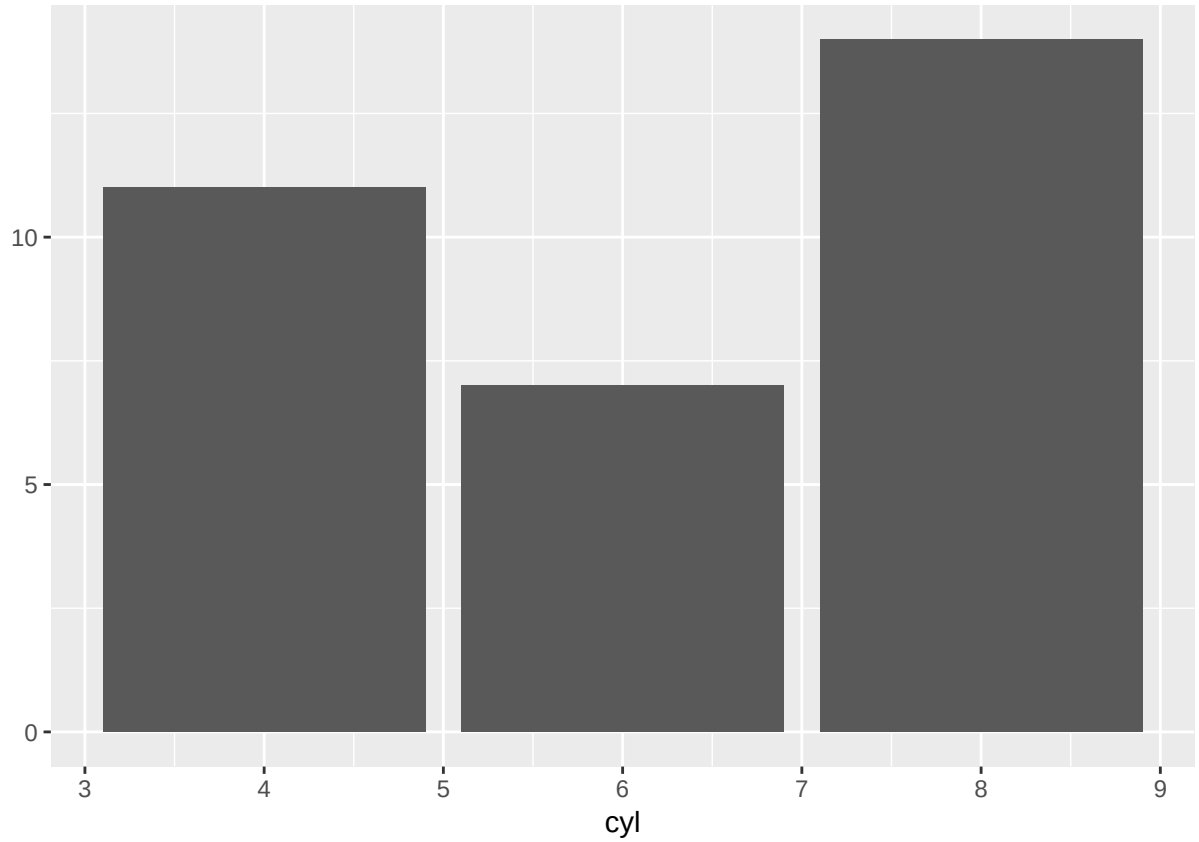
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



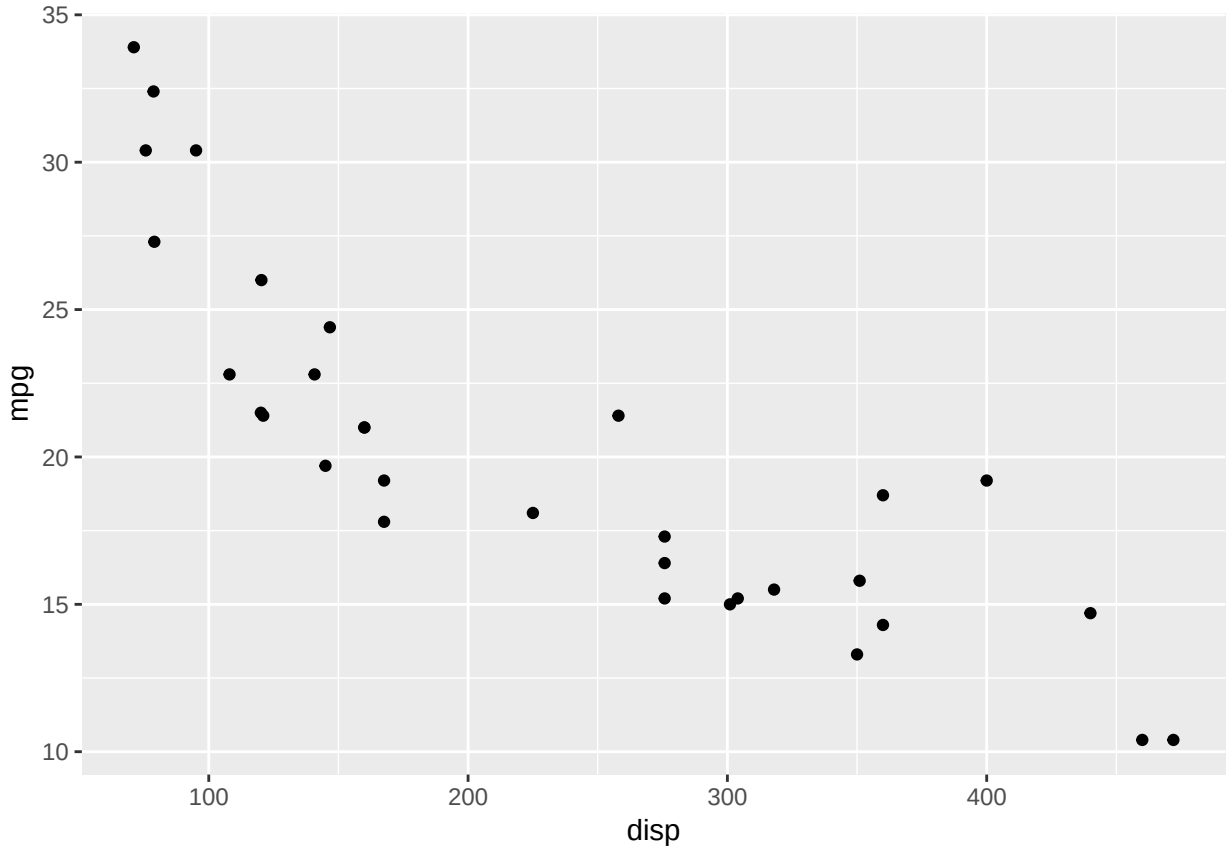

```
qplot(y=disp,x=1,data = mtcars,geom='boxplot')
```



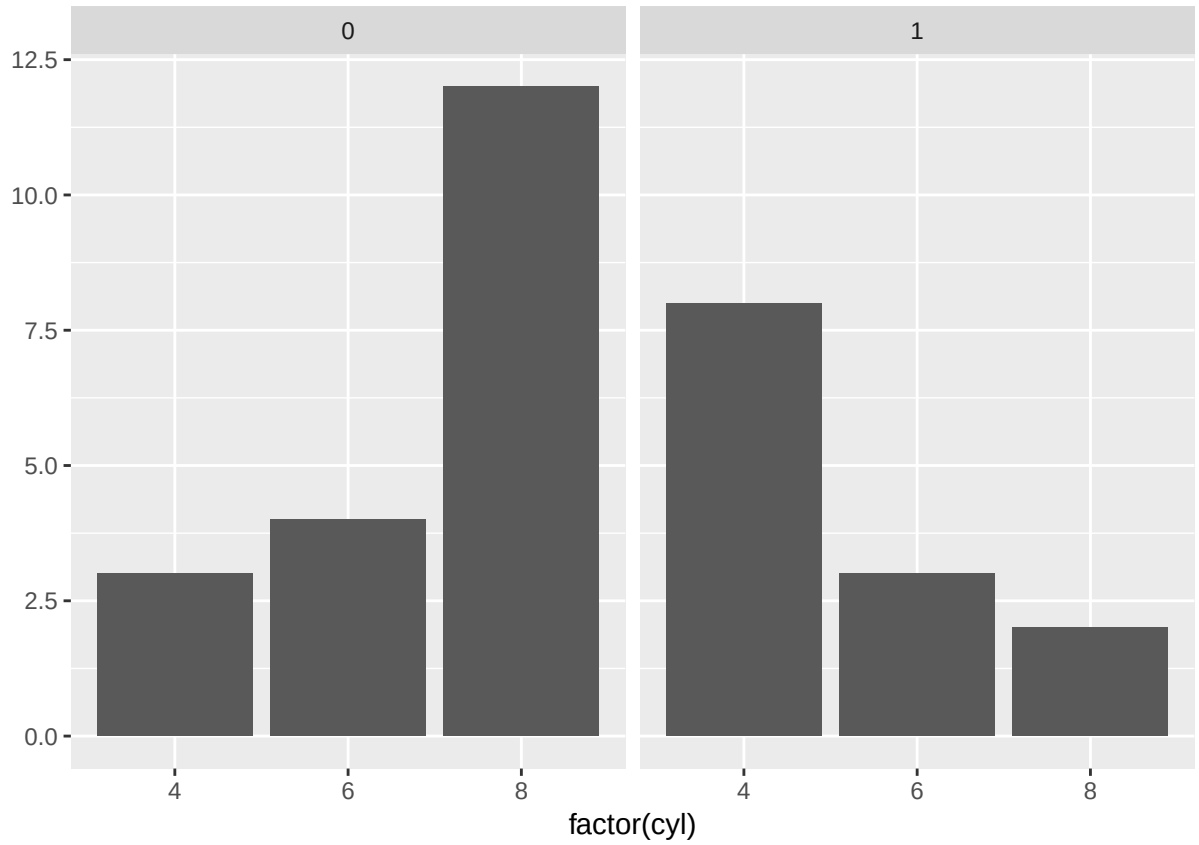

```
# one categorical variable
qplot(x=cyl,data = mtcars,geom='bar')
```




```
# two coninuous variables
qplot(x=disp,y=mpg,data = mtcars,geom = 'point')
```

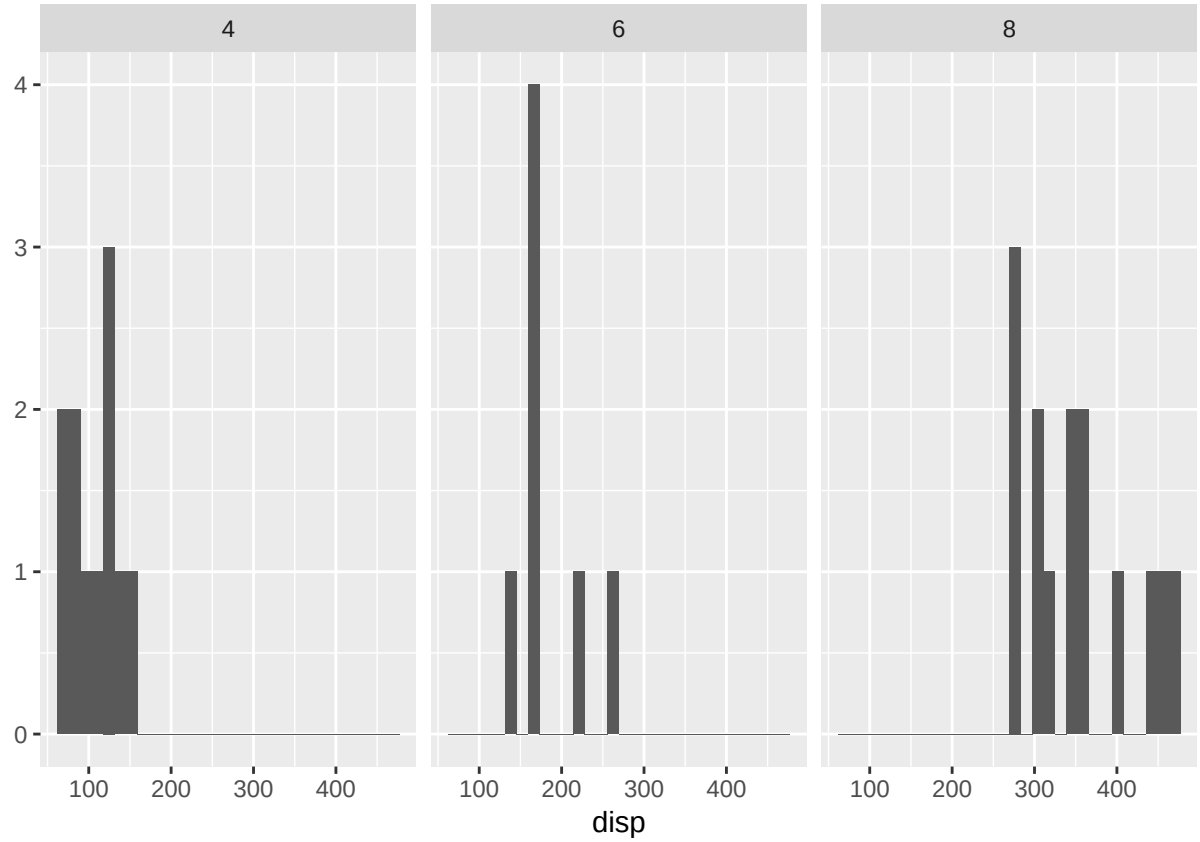



```
# two categorical vairables
qplot(x=factor(cyl),data=mtcars,geom='bar')+facet_grid(.~am)
```

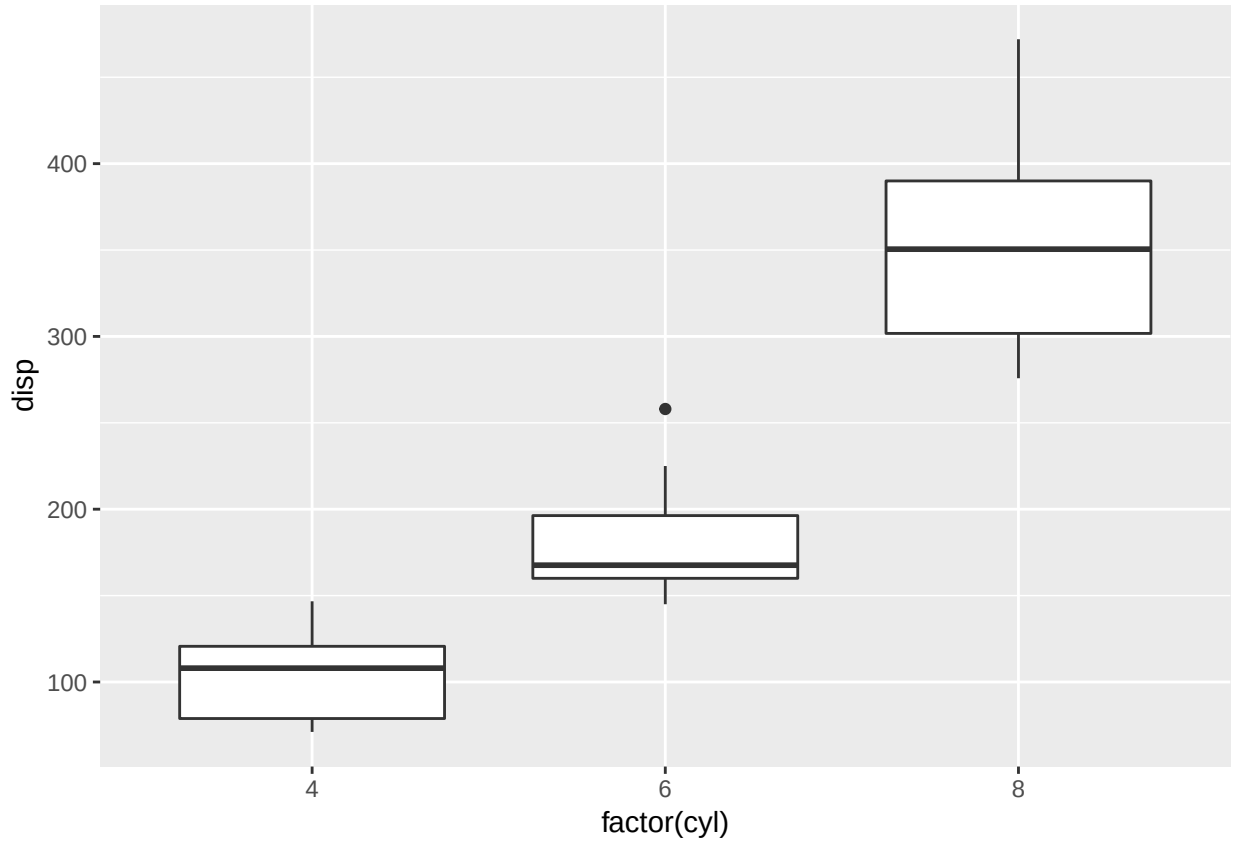



```
qplot(x=disp,data = mtcars,geom = 'histogram')+facet_grid(.~cyl)
```

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




```
qplot(y=disp,x=factor(cyl),data = mtcars,geom = 'boxplot')
```




```
# ---wrangling---
mtcars%>% filter(mpg>30)
```

```
##    mpg cyl disp  hp drat    wt  qsec vs am gear carb efficient
## 1 32.4   4 78.7  66 4.08 2.200 19.47  1  1    4    1      TRUE
## 2 30.4   4 75.7  52 4.93 1.615 18.52  1  1    4    2      TRUE
## 3 33.9   4 71.1  65 4.22 1.835 19.90  1  1    4    1      TRUE
## 4 30.4   4 95.1 113 3.77 1.513 16.90  1  1    5    2      TRUE
```

```
# making a new variable
mtcars=mtcars%>%mutate(efficient=if_else(mpg>30,TRUE,FALSE))
```