

IS240 Lab 3: Modul Grafik Statistika Deskriptif

Prodi Sistem Informasi

1 INFO

1.1 Tujuan Pembelajaran

- Mahasiswa bisa membuat grafik untuk statistika deskriptif dengan R.

1.2 Materi Praktikum

- Diagram Batang (Bar Chart)
- Diagram Pareto
- Histogram
- Density Plot
- Diagram Acak (Scatter Plot)
- Boxplot

1.3 Cheat Sheets

Lab 3

- 2data-visualization_ggplot2-cheatsheet.pdf
- 2qplot_cheatsheet.pdf
- 1Syntaxcomparison-cheatsheet.pdf,
- 2RTutorialTTH2018DescriptiveStatistics.pdf,

Lab 1 dan 2

- 2data-transformation_dplyr-cheatsheet.pdf,
- 2data-import-cheatsheet,
- 1base-r-cheatsheet.pdf dan
- 3Baggott-refcard-v2.pdf.

2 ERROR MESSAGE

2.1 there is no package called ‘blahblah’

Bila anda mendapatkan error message **Error in library(blahblah) : there is no package called ‘blah-blah’** pada saat anda mengetikkan

```
library(blahblah)
```

maka package tersebut belum terpasang pada sistem anda.
Solusi: pasang package.

```
install.packages("blahblah")  
#atau  
install.packages("blahblah", dependencies = TRUE)
```

2.2 Error in read_csv("blahblah.csv") : could not find function "read_csv"

```
mydata <- read_csv("blahblah.csv")
```

Solusi: panggil library blahblah.

```
library(blahblah) #panggil package blahblah  
mydata <- read_csv("blahblah.csv")
```

2.3 Error: path does not exist: 'blahblah.xlsx'

Bila anda mendapatkan error message **Error: path does not exist: 'blahblah.xlsx'** pada saat anda mengetikkan

```
readxl::read_excel("blahblah.xlsx", sheet = "blah1")
```

Solusi

- Periksa apakah *path ke working directory* anda pada pane **Console** sudah sama dengan path di pane **File**.
- Periksa nama file. Sistem penamaan pada R bersifat case sensitive.

2.4 Tips

Secara umum, bila anda dapat menggoogle error message dengan mengetikkan pada Internet browser

```
r "error message blahblah"
```

3 SETUP

3.1 Panggil semua package yang diperlukan.

```
library(readxl) #membaca file Excel
```

```
## Warning: package 'readxl' was built under R version 3.6.2
```

```
library(moments) #skewness dan kurtosis  
library(nortest) #uji normalitas  
library(qualityTools) #diagram Pareto
```

```
## Warning: package 'qualityTools' was built under R version 3.6.2
```

```
## Warning: package 'Rsolnp' was built under R version 3.6.2
```

```
library(ggplot2) #grafik dgn package ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.6.2
```

```
library(dplyr) #untuk filter
```

```
## Warning: package 'dplyr' was built under R version 3.6.2
```

```
library(magrittr) #untuk piping
```

```
## Warning: package 'magrittr' was built under R version 3.6.2
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.6.2
```

```
## Warning: package 'tibble' was built under R version 3.6.2
```

```
## Warning: package 'tidyr' was built under R version 3.6.2
```

```
## Warning: package 'readr' was built under R version 3.6.2
```

```
## Warning: package 'purrr' was built under R version 3.6.2
```

```
## Warning: package 'stringr' was built under R version 3.6.2
```

```
## Warning: package 'forcats' was built under R version 3.6.2
```

Package readr, ggplot2, dplyr, readr dan magrittr dapat sekaligus dipanggil dengan menggunakan perintah

```
library(tidyverse)
```

3.2 Impor Data

File dengan extension

- .csv (comma separated value)
- .txt
- .dat
- .tsv (tab separated value)

dapat diimpor dengan fungsi `read.csv()` dari base R.

Hal-hal yang perlu diperhatikan saat impor data: - delimiter: comma atau tab atau spasi? - header: ada atau tidak ada - path file: path dituliskan dengan tanda slash / - baris kosong (empty row) - decimal point untuk penulisan angka: . (sistem Amerika) atau , (sistem Indonesia)

Ingat untuk menyimpan data yang diimpor ke dalam obyek di R agar data tersebut dapat anda akses.

3.3 File Teks

```
diamond <- read.csv("Diamonds.csv", header = TRUE, sep = ",")
#atau
diamond <- read.csv("Diamonds.csv")
#baca help read.csv untuk default value dari tiap-tiap argument dan option.

str(diamond) #cek struktur data
```

```
## 'data.frame': 308 obs. of 6 variables:
## $ IDNO : int 1 2 3 4 5 6 7 8 9 10 ...
## $ WEIGHT : num 0.3 0.3 0.3 0.3 0.31 0.31 0.31 0.31 0.31 0.31 ...
## $ COLOR : Factor w/ 6 levels "D","E","F","G",...: 1 2 4 4 1 2 3 4 5 6 ...
## $ CLARITY: Factor w/ 5 levels "IF","VS1","VS2",...: 3 2 4 2 2 2 2 5 3 2 ...
## $ RATER : Factor w/ 3 levels "GIA","HRD","IGI": 1 1 1 1 1 1 1 1 1 1 ...
## $ PRICE : int 1302 1510 1510 1260 1641 1555 1427 1427 1126 1126 ...
```

```
head(diamond, 3) #tampilkan beberapa baris pertama data
```

```
## IDNO WEIGHT COLOR CLARITY RATER PRICE
## 1 1 0.3 D VS2 GIA 1302
## 2 2 0.3 E VS1 GIA 1510
## 3 3 0.3 G VVS1 GIA 1510
```

3.4 File Excel

File Excel memerlukan package khusus. Anda dapat menggunakan package `readxl` atau `openxlsx`.

```
diamond <- readxl::read_excel("lab2data.xlsx", sheet = "Diamonds")
str(diamond)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame': 308 obs. of 5 variables:
## $ CARAT : num 0.3 0.3 0.3 0.3 0.31 0.31 0.31 0.31 0.31 0.31 ...
## $ COLOR : chr "D" "E" "G" "G" ...
## $ CLARITY: chr "VS2" "VS1" "VVS1" "VS1" ...
## $ CERT : chr "GIA" "GIA" "GIA" "GIA" ...
## $ PRICE : num 1302 1510 1510 1260 1641 ...
```

```
head(diamond, 3)
```

```
## # A tibble: 3 x 5
## CARAT COLOR CLARITY CERT PRICE
## <dbl> <chr> <chr> <chr> <dbl>
## 1 0.3 D VS2 GIA 1302
## 2 0.3 E VS1 GIA 1510
## 3 0.3 G VVS1 GIA 1510
```

3.5 Package

Data dari default package R dapat langsung digunakan.

```
mtcars
```

```
##          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
## Merc 280C      17.8   6  167.6 123 3.92 3.440 18.90 1 0   4    4
## Merc 450SE     16.4   8  275.8 180 3.07 4.070 17.40 0 0   3    3
## Merc 450SL     17.3   8  275.8 180 3.07 3.730 17.60 0 0   3    3
## Merc 450SLC    15.2   8  275.8 180 3.07 3.780 18.00 0 0   3    3
## Cadillac Fleetwood 10.4   8  472.0 205 2.93 5.250 17.98 0 0   3    4
## Lincoln Continental 10.4   8  460.0 215 3.00 5.424 17.82 0 0   3    4
## Chrysler Imperial 14.7   8  440.0 230 3.23 5.345 17.42 0 0   3    4
## 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
## Toyota Corona  21.5   4  120.1  97 3.70 2.465 20.01 1 0   3    1
## Dodge Challenger 15.5   8  318.0 150 2.76 3.520 16.87 0 0   3    2
## AMC Javelin    15.2   8  304.0 150 3.15 3.435 17.30 0 0   3    2
## Camaro Z28     13.3   8  350.0 245 3.73 3.840 15.41 0 0   3    4
## Pontiac Firebird 19.2   8  400.0 175 3.08 3.845 17.05 0 0   3    2
## Fiat X1-9      27.3   4   79.0  66 4.08 1.935 18.90 1 1   4    1
## Porsche 914-2  26.0   4  120.3  91 4.43 2.140 16.70 0 1   5    2
## Lotus Europa   30.4   4   95.1 113 3.77 1.513 16.90 1 1   5    2
## Ford Pantera L  15.8   8  351.0 264 4.22 3.170 14.50 0 1   5    4
```

```
## Ferrari Dino      19.7   6 145.0 175 3.62 2.770 15.50 0 1   5   6
## Maserati Bora     15.0   8 301.0 335 3.54 3.570 14.60 0 1   5   8
## Volvo 142E       21.4   4 121.0 109 4.11 2.780 18.60 1 1   4   2
```

Menampilkan daftar data pada package

```
basedata <- data() #data pada base R
basedata

dplyrdata <- data(package = "dplyr") #data pada package dplyr
dplyrdata
```

3.6 Tipe Data

- Data kategorikal yang dituliskan dengan angka atau data kategorikal dengan level perlu diubah menjadi tipe **factor** dalam R.
- Data kategorikal seperti NamaPelanggan pada data Customer sebaiknya tetap disimpan sebagai tipe **character**, bukan tipe faktor.

```
str(mtcars)
```

```
## 'data.frame':   32 obs. of  11 variables:
## $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
## $ cyl : num   6  6  4  6  8  6  8  4  4  6 ...
## $ disp: num  160 160 108 258 360 ...
## $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
## $ drat: num   3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
## $ wt  : num   2.62 2.88 2.32 3.21 3.44 ...
## $ qsec: num   16.5 17 18.6 19.4 17 ...
## $ vs  : num   0  0  1  1  0  1  0  1  1  1 ...
## $ am  : num   1  1  1  0  0  0  0  0  0  0 ...
## $ gear: num   4  4  4  3  3  3  3  4  4  4 ...
## $ carb: num   4  4  1  1  2  1  4  2  2  4 ...
```

```
head(mtcars, 3)
```

```
##           mpg cyl disp  hp drat   wt  qsec vs am gear carb
## Mazda RX4    21.0   6  160 110 3.90 2.620 16.46 0  1    4    4
## Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02 0  1    4    4
## Datsun 710    22.8   4  108  93 3.85 2.320 18.61 1  1    4    1
```

```
str(diamond)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':   308 obs. of  5 variables:
## $ CARAT : num  0.3 0.3 0.3 0.3 0.31 0.31 0.31 0.31 0.31 0.31 ...
## $ COLOR : chr  "D" "E" "G" "G" ...
## $ CLARITY: chr  "VS2" "VS1" "VVS1" "VS1" ...
## $ CERT : chr  "GIA" "GIA" "GIA" "GIA" ...
## $ PRICE : num  1302 1510 1510 1260 1641 ...
```

```
head(diamond, 3)
```

```
## # A tibble: 3 x 5
##   CARAT COLOR CLARITY CERT  PRICE
##   <dbl> <chr> <chr>   <chr> <dbl>
## 1  0.3 D     VS2     GIA    1302
## 2  0.3 E     VS1     GIA    1510
## 3  0.3 G     VVS1    GIA    1510
```

Pada contoh data mtcars, variabel am (tipe transmisi) masih berupa data numerik. Gunakan fungsi *as.factor()* untuk merubah tipe data am, cyl, vs, gear, carb menjadi tipe factor.

```
mtcars$am <- as.factor(mtcars$am)
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$gear <- as.factor(mtcars$gear)
mtcars$carb <- as.factor(mtcars$carb)
str(mtcars)
```

```
## 'data.frame':   32 obs. of  11 variables:
## $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
## $ cyl : Factor w/ 3 levels "4","6","8": 2 2 1 2 3 2 3 1 1 2 ...
## $ disp: num  160 160 108 258 360 ...
## $ hp : num  110 110 93 110 175 105 245 62 95 123 ...
## $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
## $ wt : num  2.62 2.88 2.32 3.21 3.44 ...
## $ qsec: num  16.5 17 18.6 19.4 17 ...
## $ vs : Factor w/ 2 levels "0","1": 1 1 2 2 1 2 1 2 2 2 ...
## $ am : Factor w/ 2 levels "0","1": 2 2 2 1 1 1 1 1 1 1 ...
## $ gear: Factor w/ 3 levels "3","4","5": 2 2 2 1 1 1 1 2 2 2 ...
## $ carb: Factor w/ 6 levels "1","2","3","4",...: 4 4 1 1 2 1 4 2 2 4 ...
```

R secara otomatis akan mengurutkan level menurut abjad. Bila anda ingin merubahnya, gunakan fungsi *level()*.

```
table(diamond$CLARITY)
```

```
##
##   IF   VS1   VS2  VVS1  VVS2
##   44   81   53   52   78
```

```
diamond$CLARITY <- as.factor(diamond$CLARITY)
levels(diamond$CLARITY) <- c("IF", "VVS1", "VVS2", "VS1", "VS2")
levels(diamond$CLARITY)
```

```
## [1] "IF"    "VVS1" "VVS2" "VS1"   "VS2"
```

Informasi lebih lengkap mengenai diamond 4C dapat dibaca di laman [Blue Nile](#)

4 VARIABEL KATEGORIKAL (NONNUMERIK)

4.1 Bar Chart: Distribusi Frekuensi Variabel Kategorikal

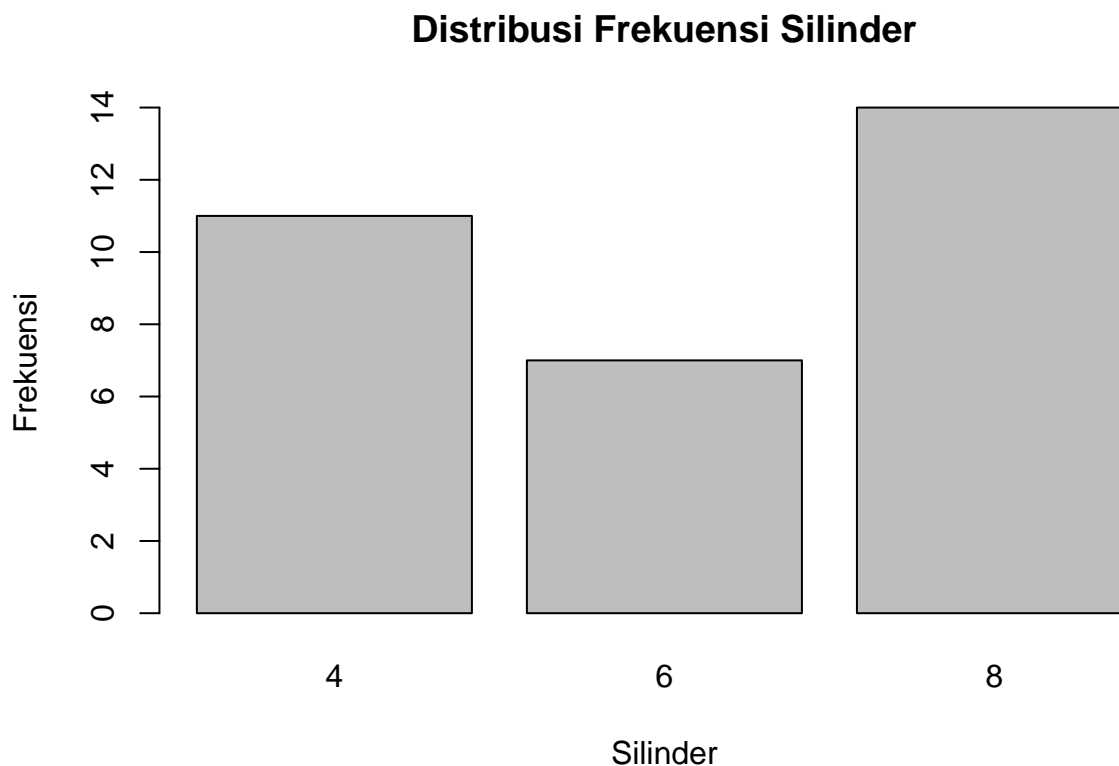
Diagram batang (bar chart) digunakan untuk menggambarkan tabulasi frekuensi sebuah variabel kategorikal. Sintak bar chart memiliki format minimum

```
barplot(datadalambentuktabel, ...)
```

```
(tab <- table(mtcars$cyl)) #Siapkan tabel untuk input data ke dalam fungsi barplot.
```

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

```
#base R  
judul <- "Distribusi Frekuensi Silinder"  
sbX <- "Silinder"  
sbY <- "Frekuensi"  
legenda <- c(4,6,8) #Silinder mobil pada data mtcars memiliki 4, 6 atau 8 silinder.  
#legenda <- rownames(mtcars$cyl)  
  
barplot(tab, main = judul, names.arg = legenda, xlab = sbX, ylab = sbY)
```



Gunakan shortcut rownames atau colnames bila kita tidak ingin mengetikkan label satu persatu.

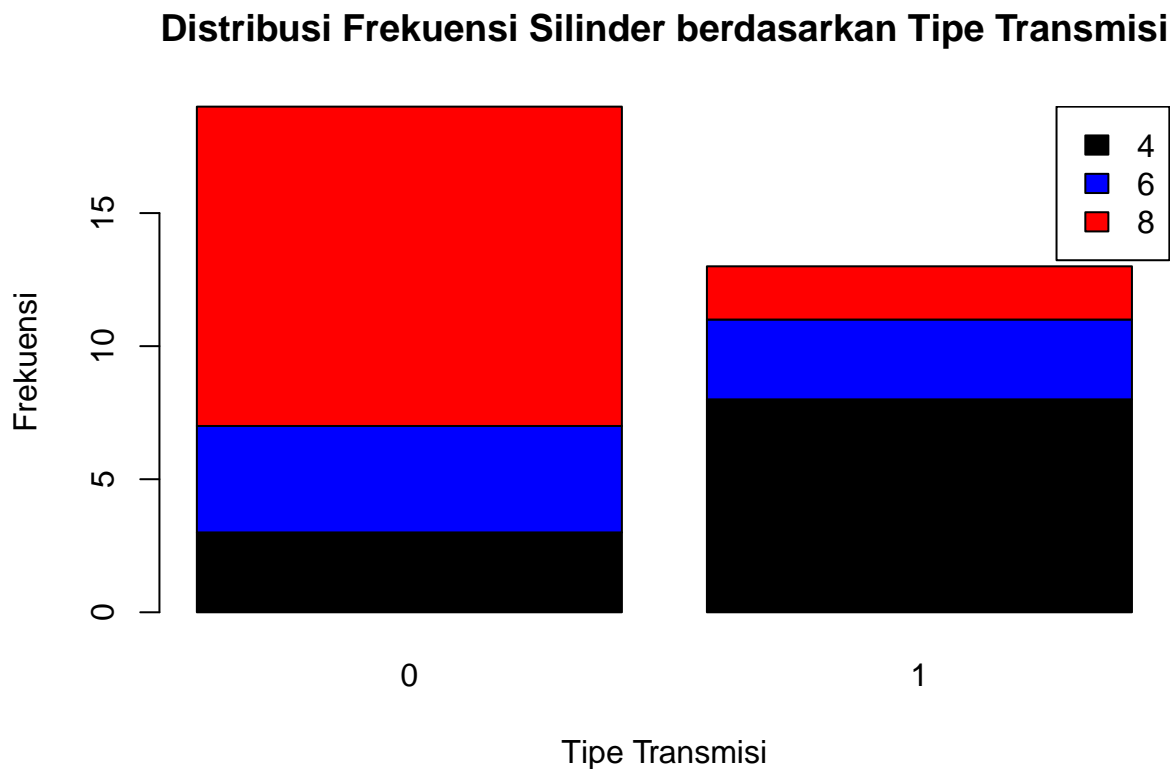
4.2 Stacked Bar Chart

Buatlah Stacked Bar Plot untuk variabel cyl dan am.

```
(tab <- table(mtcars$cyl, mtcars$am))

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

#base R
judul <- "Distribusi Frekuensi Silinder berdasarkan Tipe Transmisi"
sbX <- "Tipe Transmisi"
sbY <- "Frekuensi"
warna <- c("black", "blue", "red")
# legenda <- c(4, 6, 8)
legenda <- rownames(tab)
barplot(tab, main = judul, xlab = sbX, ylab = sbY, col = warna)
legend("topright", legend = legenda, fill = warna)
```



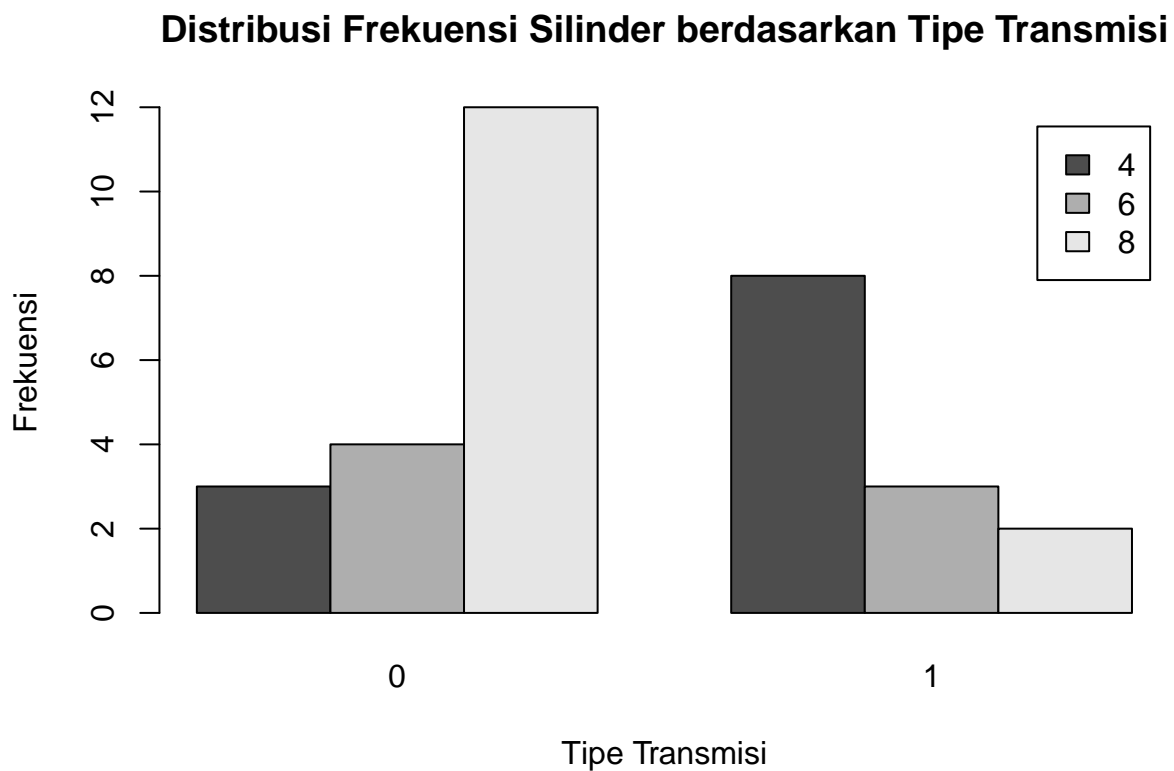
4.3 Side-by-side Bar Chart

Side-by-side bar chart dihasilkan dengan menggunakan opsi **beside**.

```
#base R
(tab <- table(mtcars$cyl, mtcars$am))
```

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

```
barplot(tab, beside = TRUE, main = judul, xlab = sbX, ylab = sbY, legend = legenda)
```



4.4 Diagram Pareto

Diagram Pareto merupakan gabungan antara diagram batang dengan ogive. Diagram ini digunakan untuk menunjukkan kategori-kategori dengan frekuensi terbanyak. Kategori-kategori yang penting adalah kategori-kategori sampai dengan jumlah persentase kumulatif 80%.

```
(tab <- table(mtcars$cyl))
qualityTools::paretoChart(tab) #dari package qualityTools

(tab <- table(diamond$clarity))
paretoChart(tab)
```

5 VARIABEL NUMERIK

5.1 Histogram: Distribusi Data Numerik (1 Variable)

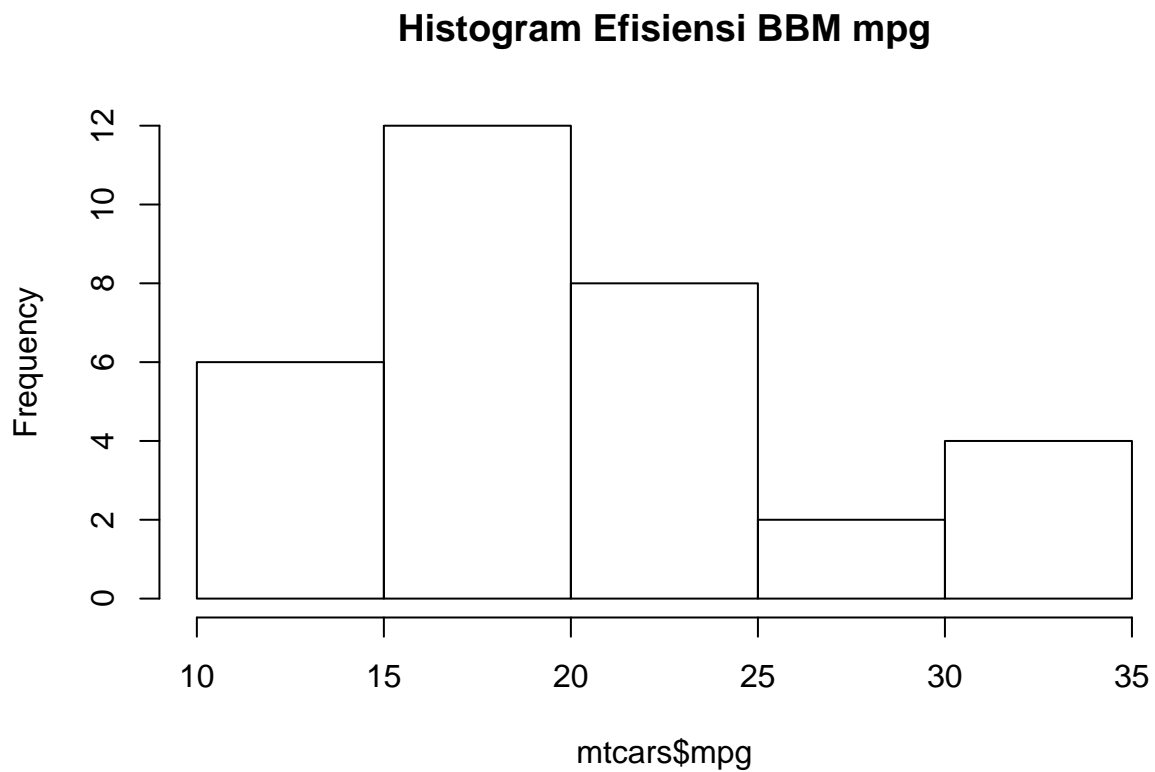
```
moments::skewness(mtcars[,c("mpg", "hp")])
```

```
##      mpg      hp  
## 0.6404399 0.7614356
```

```
moments::kurtosis(mtcars[,c("mpg", "hp")])
```

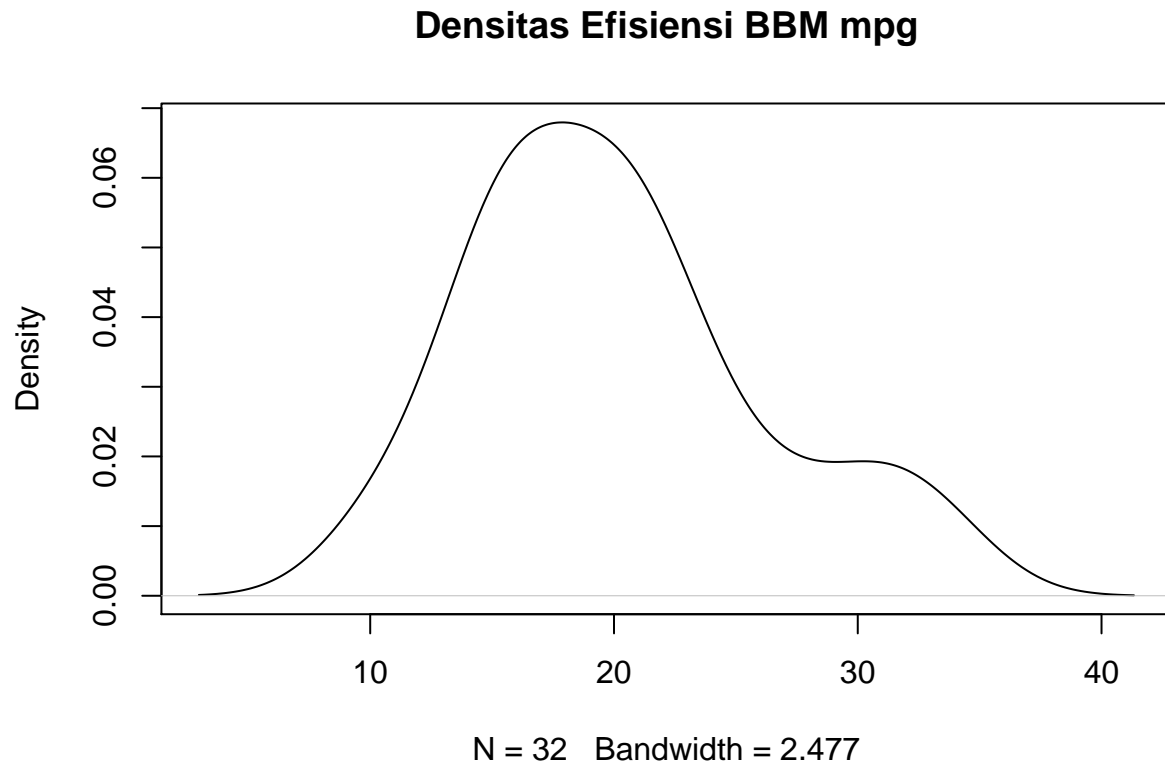
```
##      mpg      hp  
## 2.799467 3.052233
```

```
sbX <- "mil/ galon"  
judul <- "Histogram Efisiensi BBM mpg"  
hist(mtcars$mpg, main = judul)
```

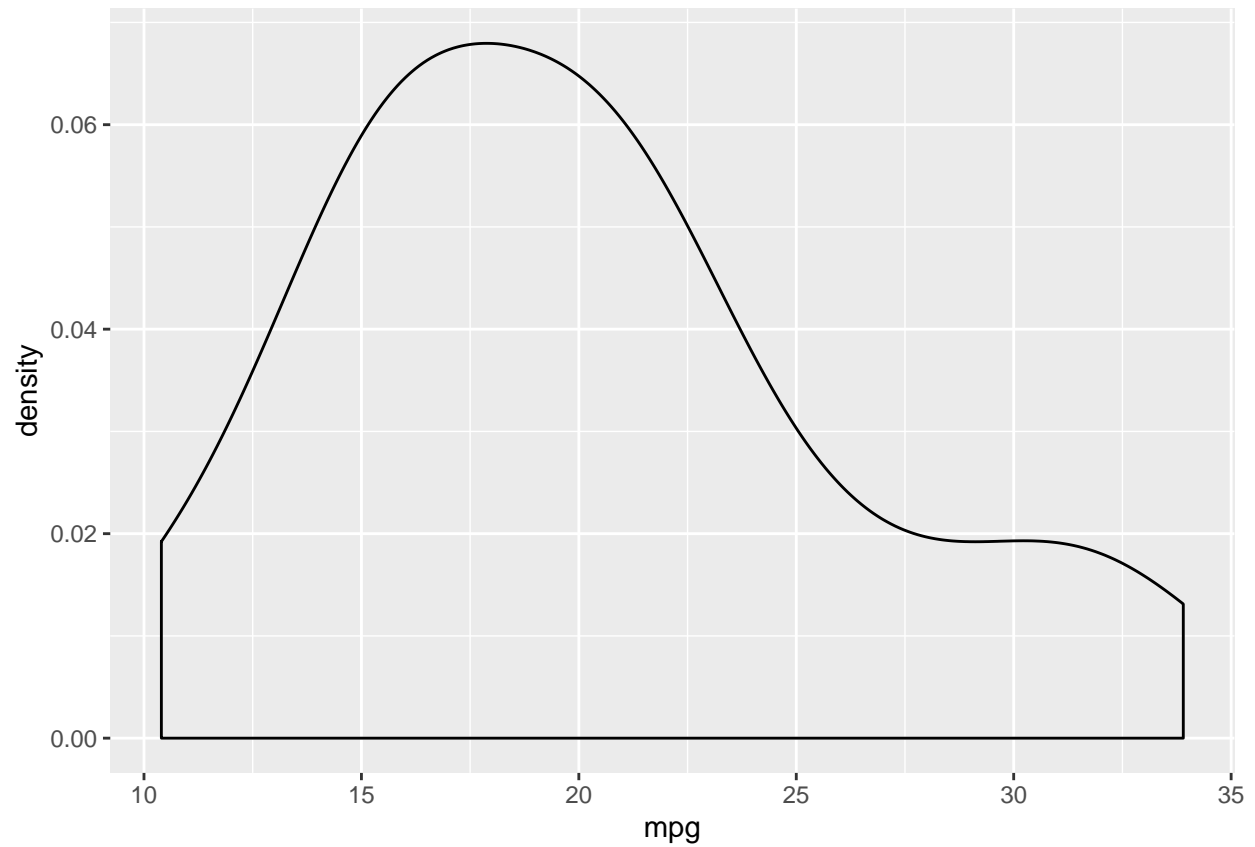


5.2 Density Plot: Distribusi Data Numerik (1 Variable)

```
#base R  
plot(density(mtcars$mpg), main = "Densitas Efisiensi BBM mpg")
```



```
#ggplot2  
mtcars %>% ggplot(aes(x = mpg)) +  
  geom_density()
```

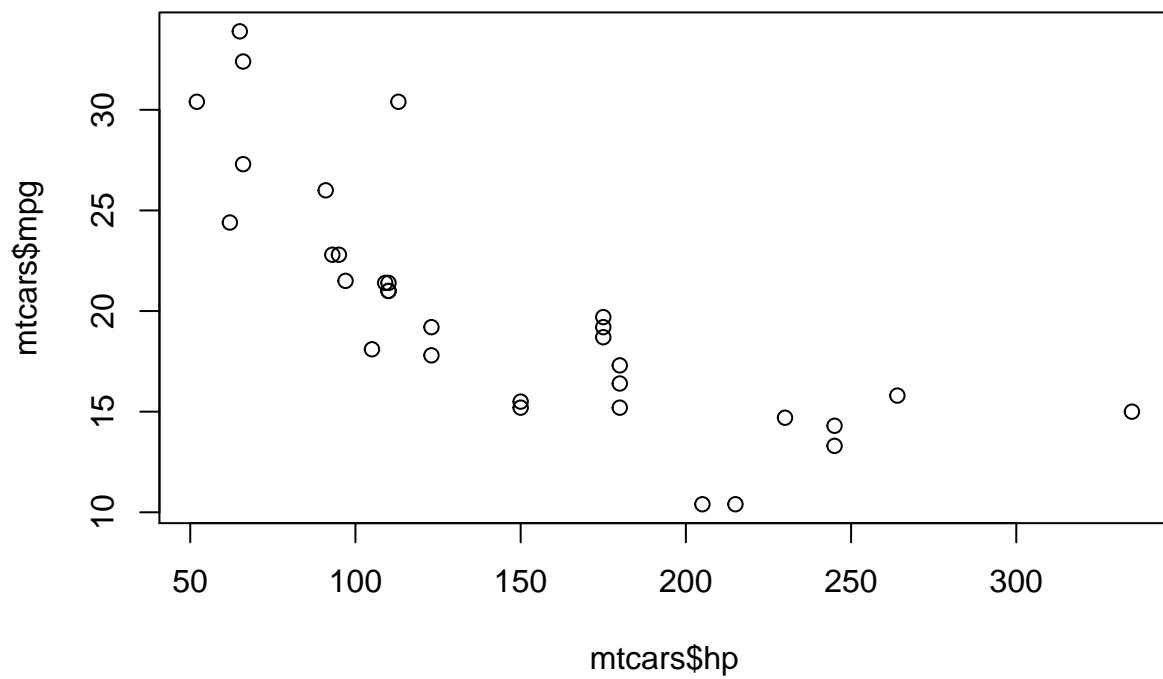


5.3 Scatter Plot: Hubungan antara 2 variabel numerik

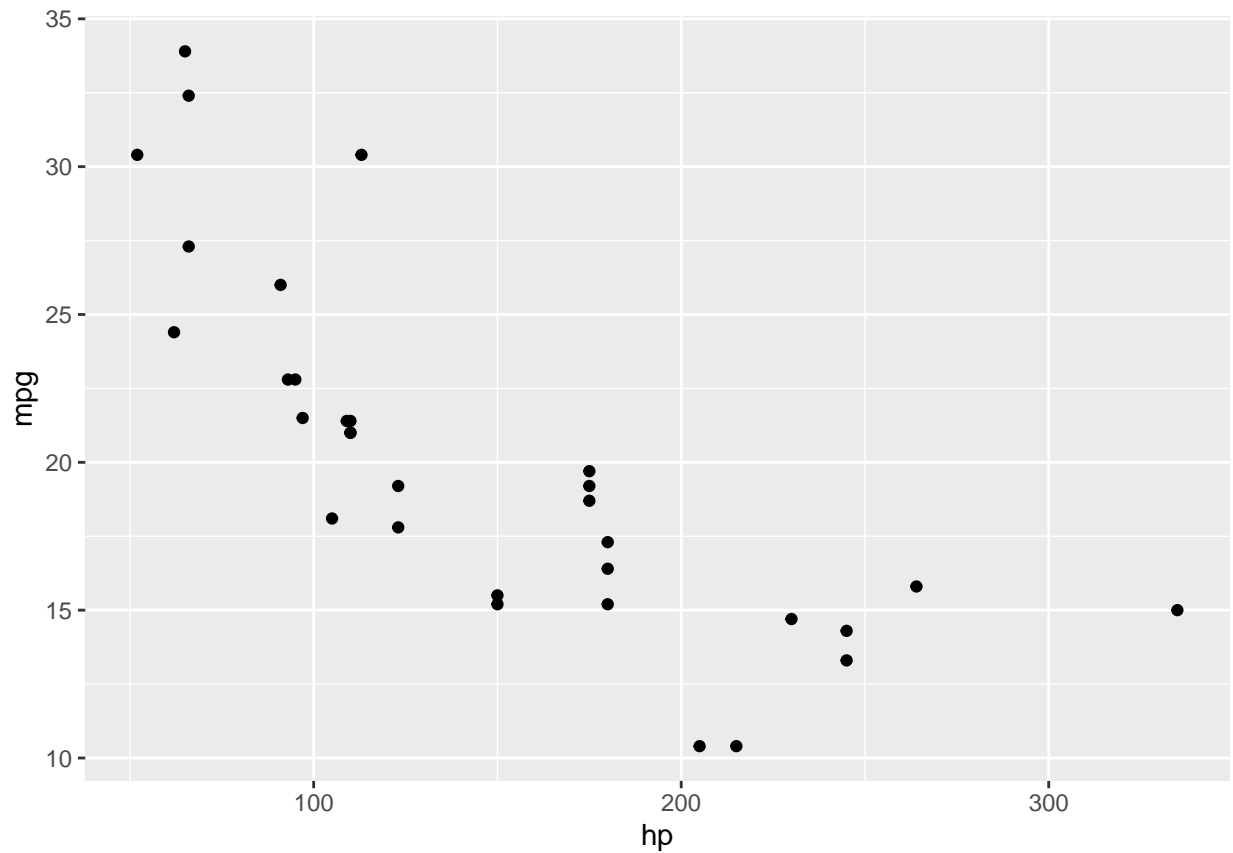
```
cor(mtcars$hp, mtcars$mpg, method = "pearson")
```

```
## [1] -0.7761684
```

```
#base R  
plot(mtcars$hp, mtcars$mpg)
```

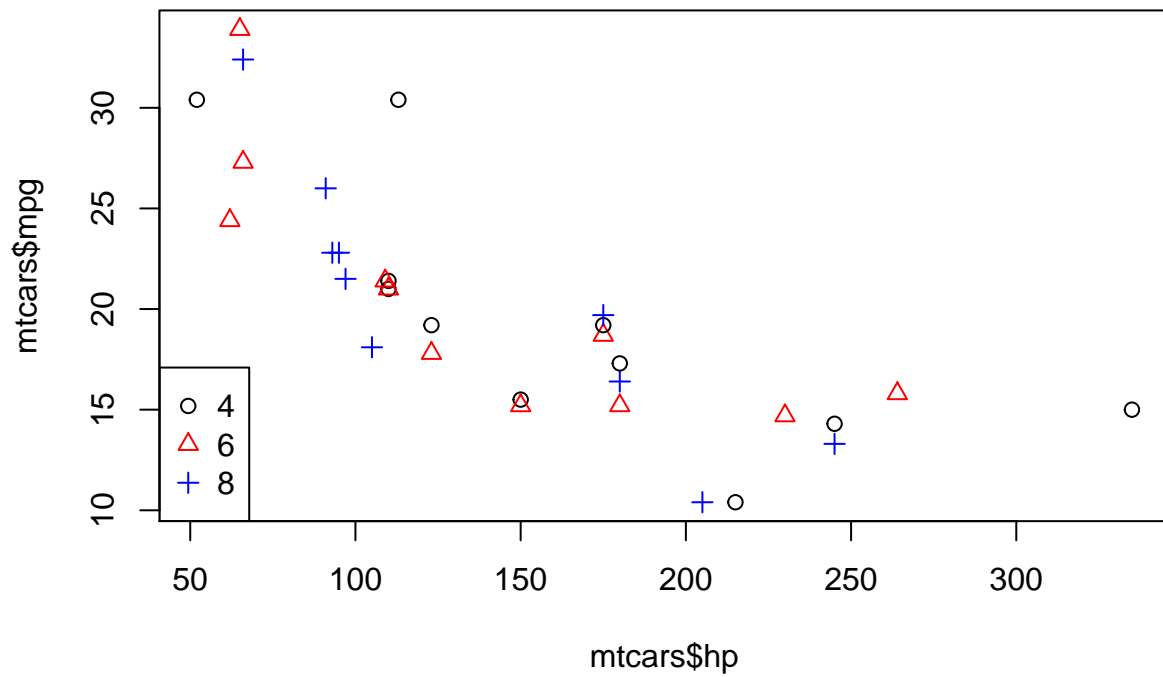


```
#ggplot2
mtcars %>% ggplot(aes(x = hp, y = mpg)) + geom_point()
```



5.4 Scatter Plot untuk Silinder 1 dan 2

```
#base R
plot(mtcars$hp, mtcars$mpg, col = c("black", "red", "blue"), pch = c(1,2,3))
legend("bottomleft", c("4", "6", "8"), pch = c(1,2,3), col = c("black", "red", "blue") )
```

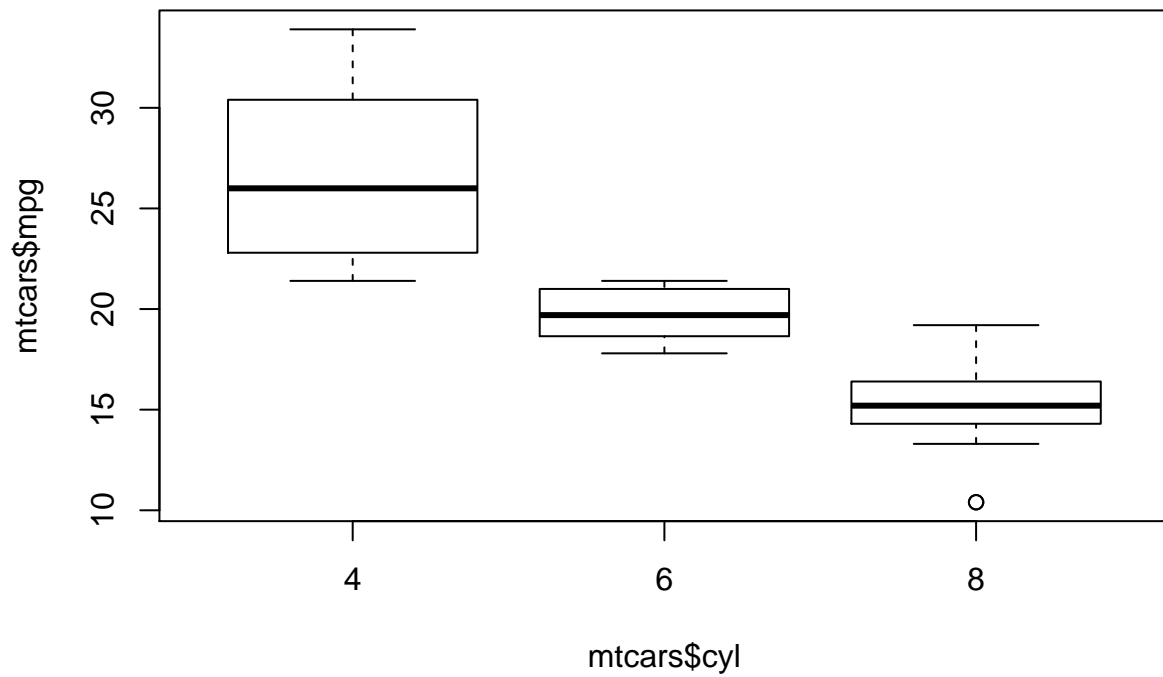


6 VARIABEL NUMERIK DAN KATEGORIKAL

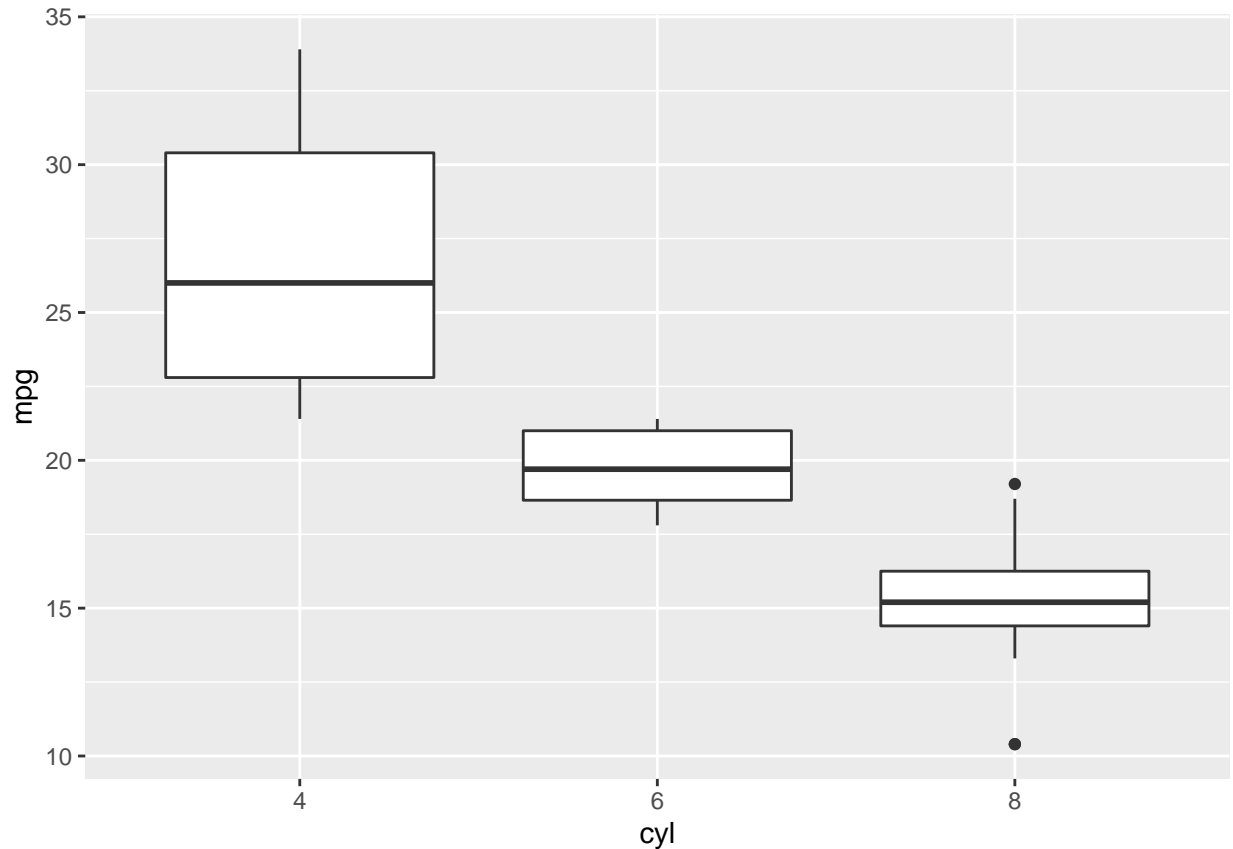
6.1 Boxplot: Distribusi Variable Numerik Berdasarkan Variabel Kategorikal

```
#base R
boxplot(mtcars$mpg ~ mtcars$cyl, main = "Distribusi mpg berdasarkan silinder")
```


Distribusi mpg berdasarkan silinder



```
#ggplot2  
mtcars %>% ggplot(aes(x = cyl, y = mpg)) +  
  geom_boxplot()
```



7 PLOT OPTIONS

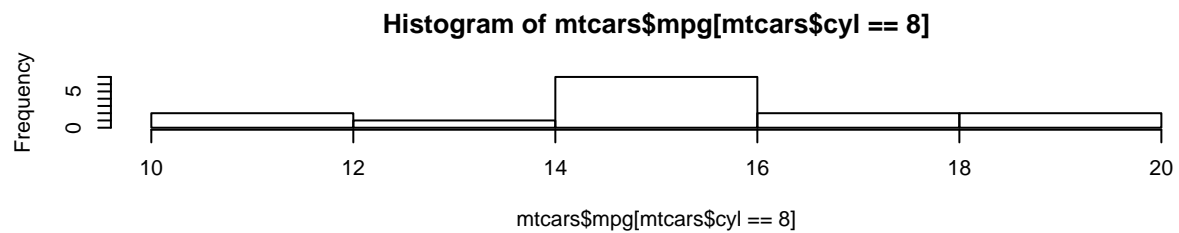
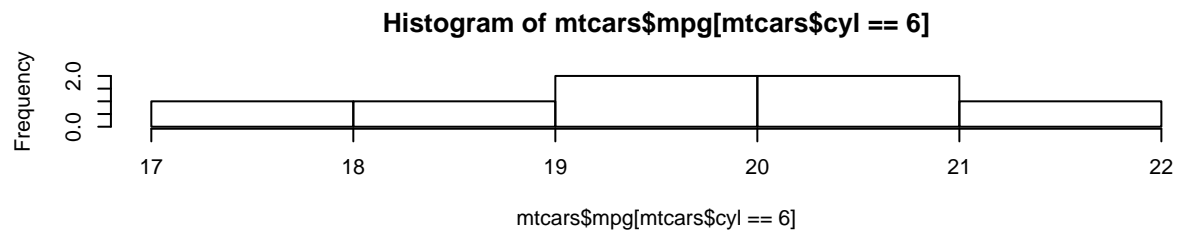
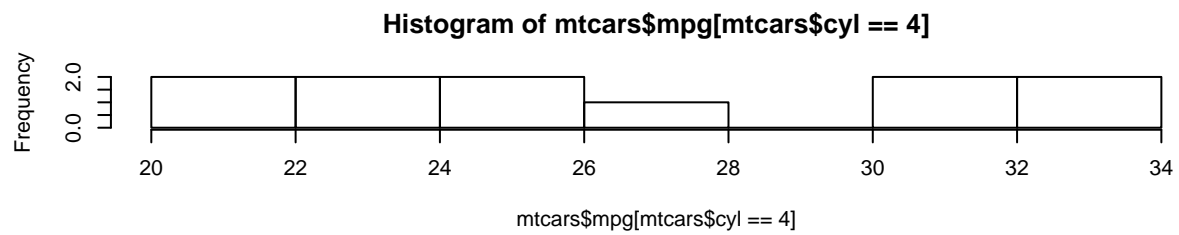
Fungsi `plot()` memiliki banyak option seperti *main*, *sub*, *xlab*, *col*, *pch*, *pcx*, *par* dan *linetype*. Baca help dengan mengetikkan

```
help(plot)
#atau
?plot
```

pada konsol.

7.1 Multiple Rows or Columns

```
#base R
par(mfrow = c(3,1)) #Siapkan layout untuk 3 plot dengan susunan 3 baris dan 1 kolom.
hist(mtcars$mpg[mtcars$cyl == 4])
hist(mtcars$mpg[mtcars$cyl == 6])
hist(mtcars$mpg[mtcars$cyl == 8])
```



```
par(mfrow = c(1,1)) #Kembalikan layout ke default layout c(1,1).

#ggplot2
mtcars %>%
  ggplot(aes(x = mpg)) +
  geom_histogram() +
  facet_grid(cyl ~.)
```

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

