



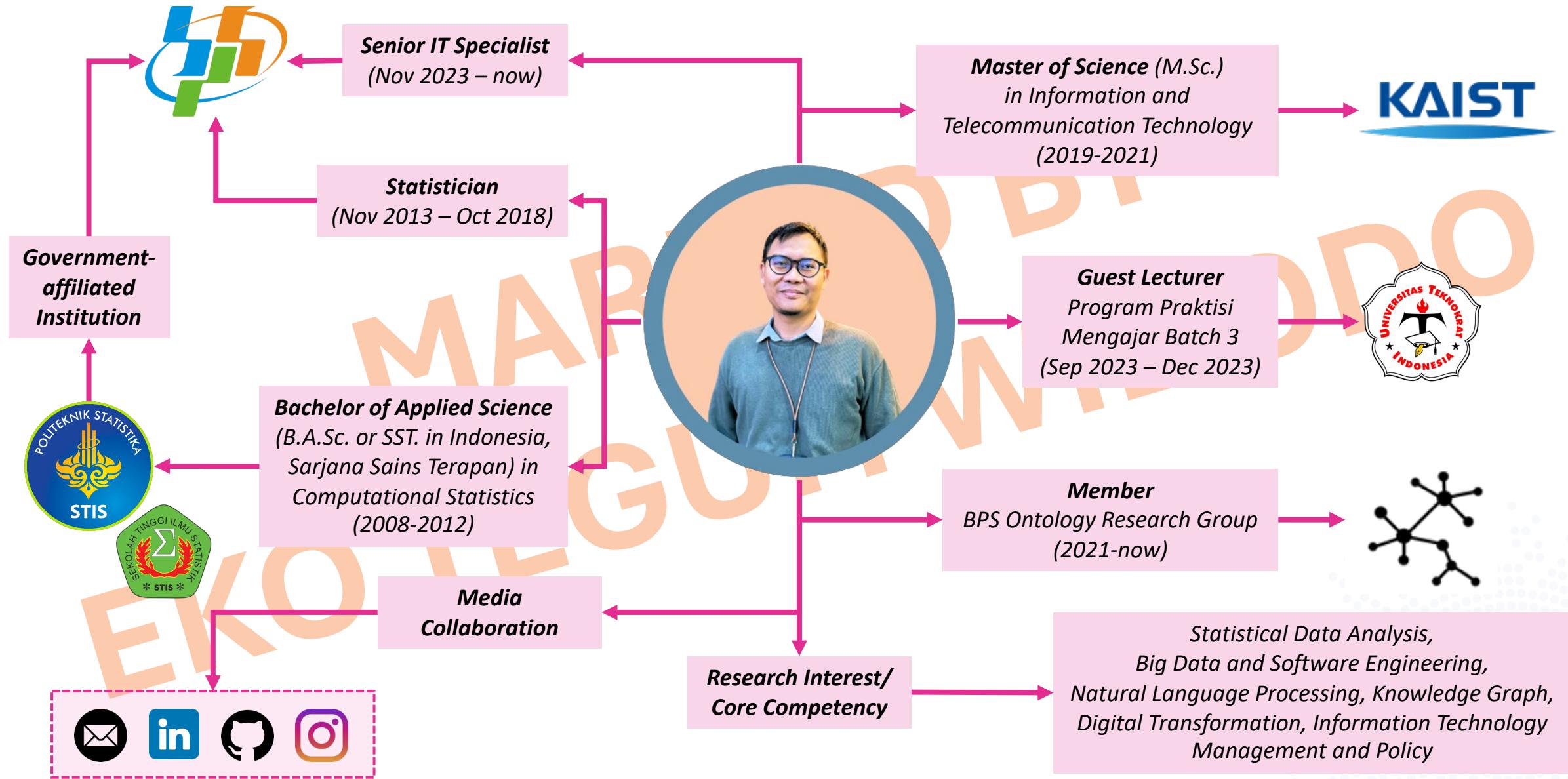
PENGOLAHAN DATA STATISTIK DENGAN R

PROGRAM PRAKTIKI MENGAJAR

BATCH 3 TAHUN 2023

TIM DOSEN

1. **Putri Sukma Dewi**, S.Pd., M.Pd. (*Dosen Pengampu*)
 2. **Eko Teguh Widodo**, SST., M.Sc. (*Dosen Praktisi*)





PENGENALAN R

- Software Pengolahan Data
- R dan Kenapa memilih R untuk Pengolahan Data Statistik
- Arsitektur R
- Prasyarat untuk Memulai R
- Basics R
- Intermediate and Advanced R



PRAKTIK

- Live Coding dengan R

MARKED BY
EKOTEGUH WIDODO



ANALISIS DATA STATISTIK

PENGENALAN R



Pertimbangan:

- ✓ Input data besar
- ✓ Repetisi pengolahan
- ✓ Kebutuhan olah data secara cepat
- ✓ Kebutuhan ketepatan olah data
- ✓ Pengolahan data kompleks: Big Data

EKO

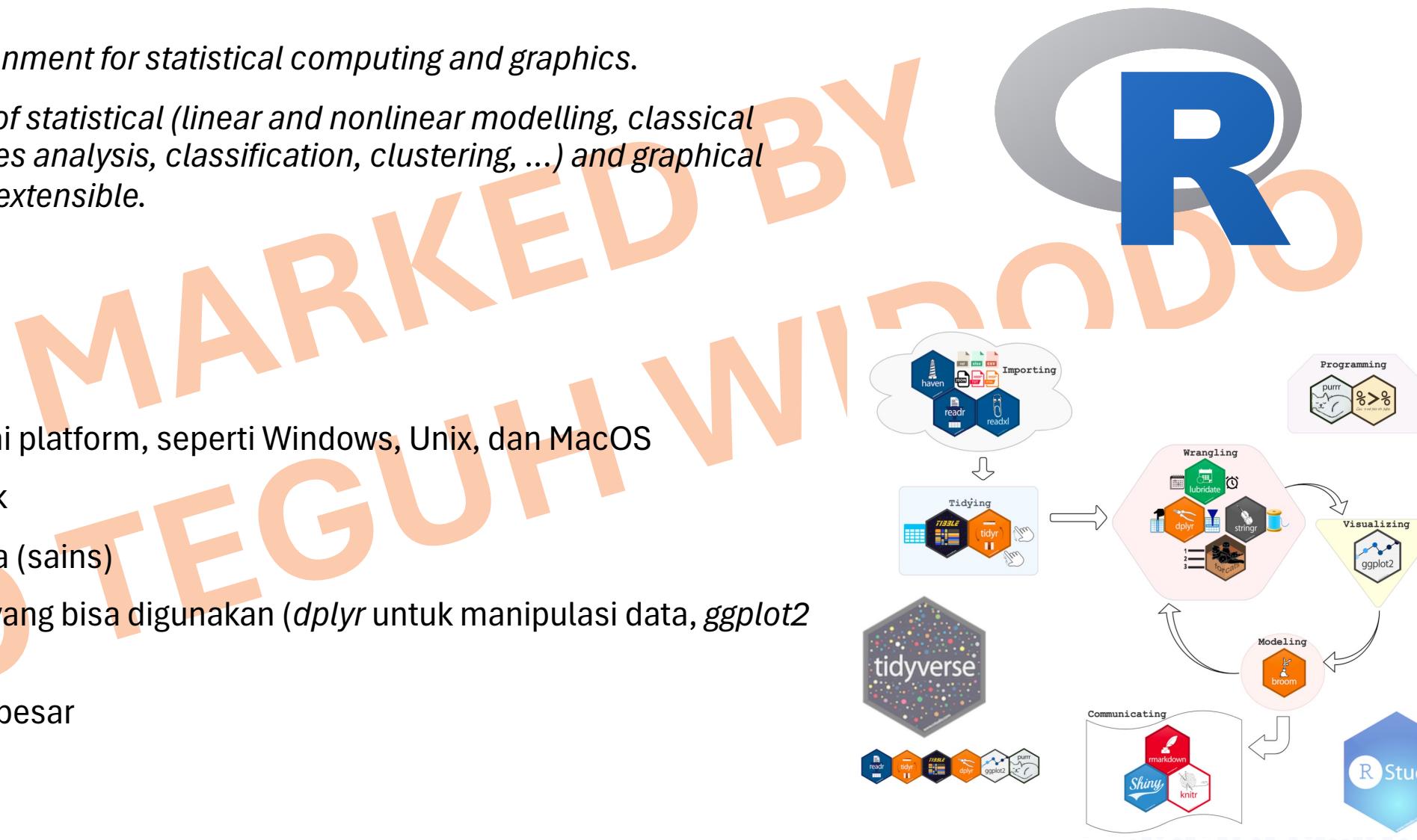


Apa itu R?

- ✓ *R is a language and environment for statistical computing and graphics.*
- ✓ *R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible.*

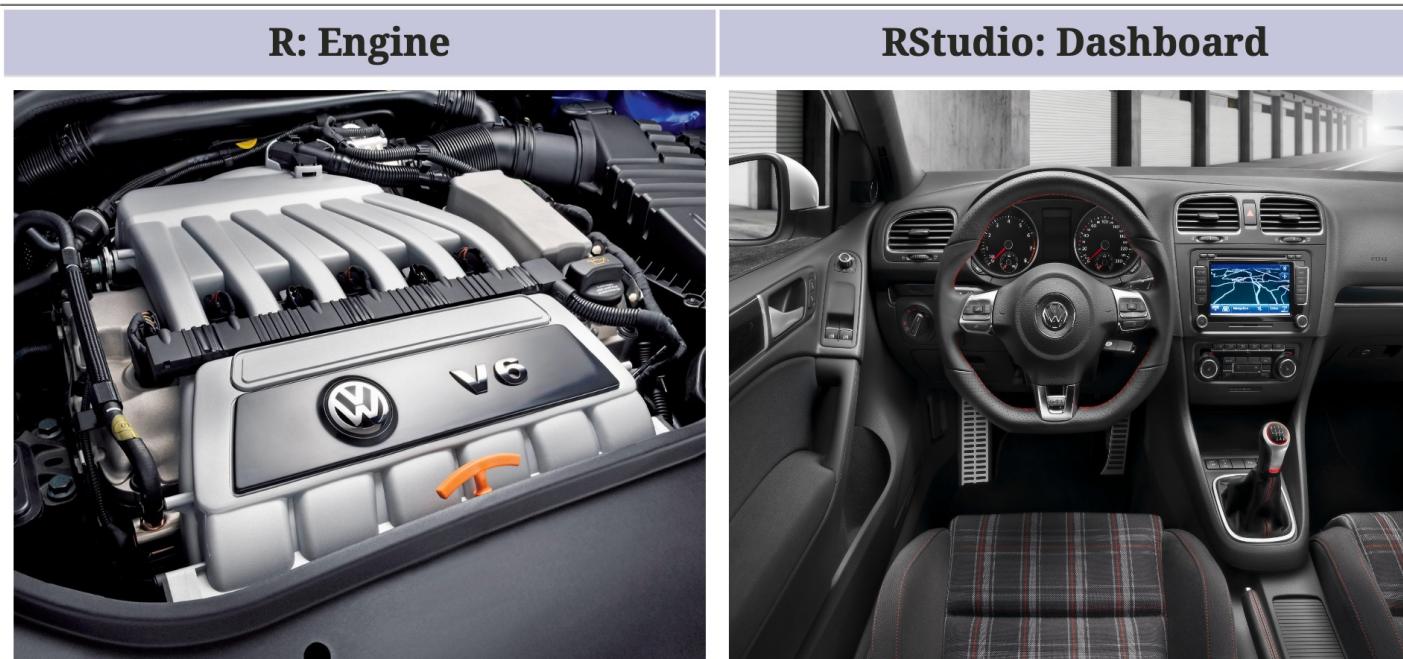
Kenapa memilih R?

- ✓ It's *FREE!*
- ✓ Dapat berjalan di berbagai platform, seperti Windows, Unix, dan MacOS
- ✓ R dibangun untuk Statistik
- ✓ Populer untuk bidang data (sains)
- ✓ Beragam jenis packages yang bisa digunakan (*dplyr* untuk manipulasi data, *ggplot2* untuk visualisasi, dsb)
- ✓ Memiliki komunitas yang besar





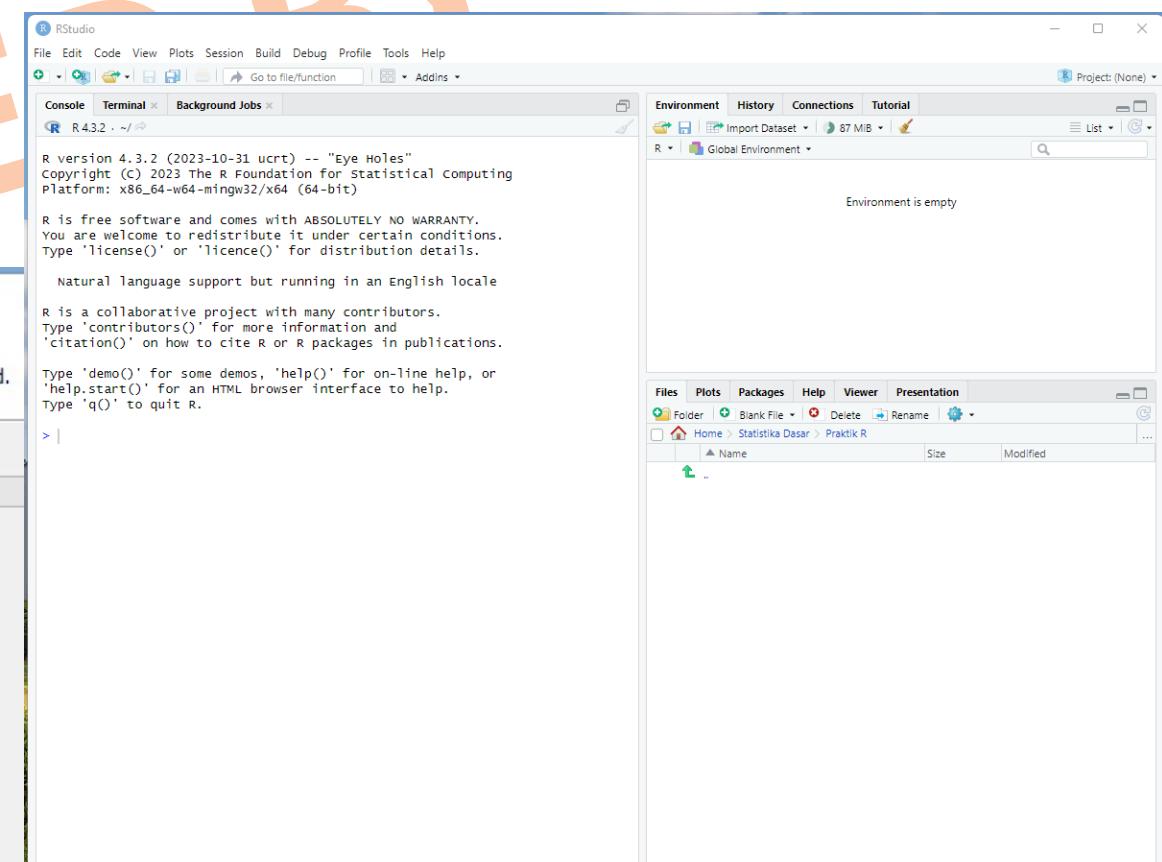
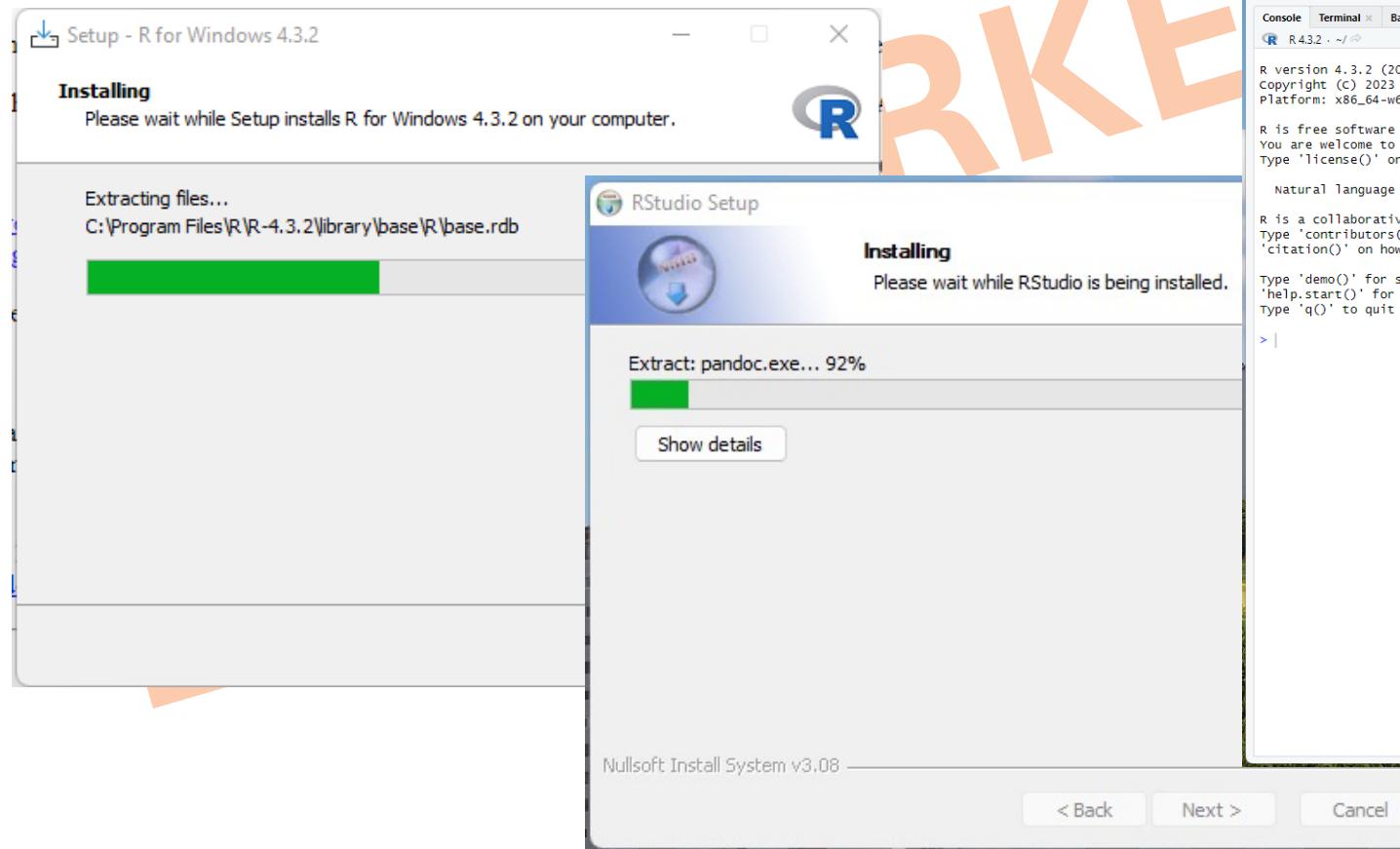
- ✓ R itu seperti **mesin (engine)** sebuah mobil.
- ✓ **RStudio** itu seperti dashboard dari mobil tersebut → *Integrated Development Environment (IDE)*



• Unduh R dan R Studio (Posit)

- Software R: <https://cran.rstudio.com>
- Software R Studio (Posit): <https://posit.co/download/rstudio-desktop>

• Instalasi R dan R Studio

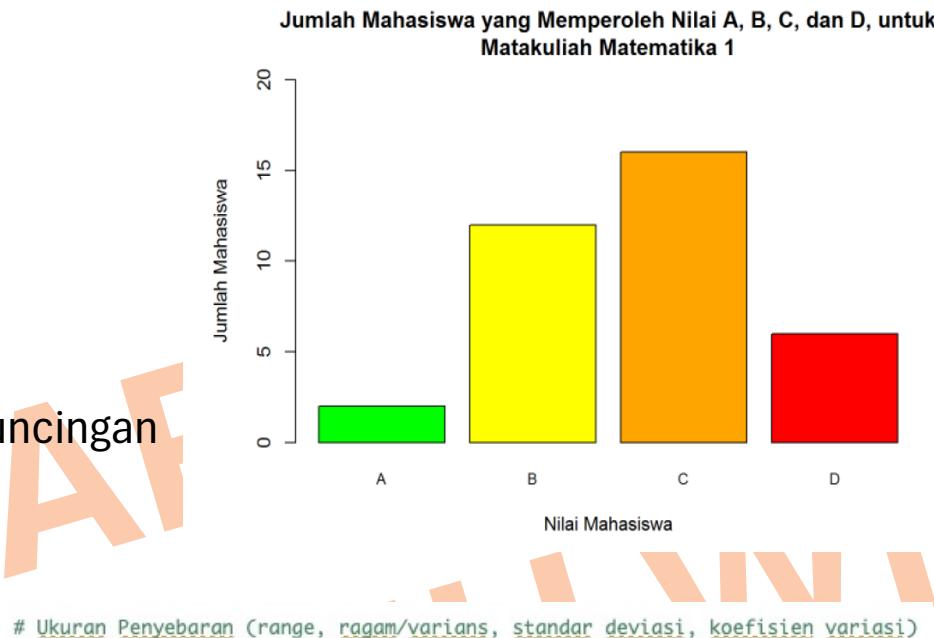


- Variabel
 - Tipe Data
 - Bilangan bulat (integer)
 - Bilangan decimal (double)
 - Karakter (string)
 - Boolean (true/false)
 - ...
 - Operator
 - Aritmatika
 - Logika
 - Perbandingan
 - Penugasan
 - Percabangan
 - 2 cabang
 - Lebih dari 2 cabang
 - Perulangan
 - For
 - While
 - Fungsi
- MARKED BY**
EKO TEGUH WIDODO

- Visualisasi Data
- Ukuran Statistik
 - Ukuran Pemusatan
 - Ukuran Penyebaran
 - Ukuran Lokasi
 - Ukuran Kemencengan dan Keruncingan
 - ...
- Korelasi
- Regresi
- Hipotesis
- ...

MAF

EKOTI

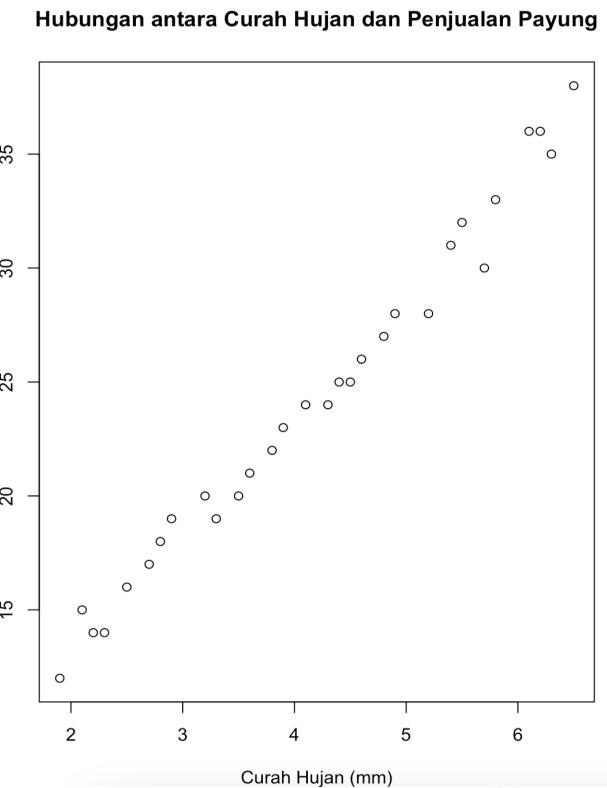


```
range_tinggi_badan <- max(tinggi_badan) - min(tinggi_badan)
range_tinggi_badan

varians_tinggi_badan <- var(tinggi_badan)
varians_tinggi_badan

stdev_tinggi_badan <- sd(tinggi_badan)
stdev_tinggi_badan

koef_variasi_tinggi_badan <- stdev_tinggi_badan / mean_tinggi_badan * 100
koef_variasi_tinggi_badan
```



```
Call:
lm(formula = data_hujan_payung$penjualan_payung ~ data_hujan_payung$curah_hujan)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.23880 -0.54954 -0.06048  0.66911  1.68151 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 2.6033    0.5589   4.658 7.06e-05 ***
data_hujan_payung$curah_hujan 5.1992    0.1274  40.808 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9567 on 28 degrees of freedom
Multiple R-squared:  0.9835, Adjusted R-squared:  0.9829 
F-statistic: 1665 on 1 and 28 DF,  p-value: < 2.2e-16
```



R Programming for Statistical Data Analysis

LIVE CODING R



Open the Link:

<https://github.com/ekotwidodo/UTI-Workshop-Pengolahan-Data-R>

Folder Materi:
Slides/

Folder Praktik:
Practices/

Folder Dataset:
Datasets/

EKO TEGUH WIDODO
MARKED BY



lpdp
lembaga pengelola dana pendidikan

Kampus
Merdeka
INDONESIA JAYA

PRAKTIKI
MENGAJAR



TERIMA KASIH

EKO

Link Materials:

<https://github.com/ekotwidodo/UTI-Workshop-Pengolahan-Data-R>

Requests for Collaboration:





R Coding Documentation for Workshop

LAMPIRAN

```
nama <- "teguh" # _nama tidak diizinkan
nama_depan <- "eko teguh"
nama.belakang <- "widodo"
nama10 <- "nama 10"
.alamat <- "bandar lampung" # tapi .100alamat tidak diizinkan
TINGGI_BADAN <- 175
for <- "for" # Reserved words
&number <- 10
```

```
# memasukkan data
tinggi_badan <- c(175, 173, 171, 160, 168, 168, 172)
# membaca data
tinggi_badan
tinggi_badan[3]
tinggi_badan[c(1,3,6)]
tinggi_badan[3:6]
# mengubah nilai data
tinggi_badan[5] = 169
tinggi_badan
# menghapus nilai data
tinggi_badan_hapus = tinggi_badan[-2]
tinggi_badan_hapus
tinggi_badan_hapus_dua = tinggi_badan[c(-1, -4)]
tinggi_badan_hapus_dua
```

```
bilangan_1 <- 10
bilangan_2 <- 5

# Fungsi cat() digunakan untuk mencetak sesuatu yang ingin ditampilkan
# Tanda "\n" digunakan untuk pindah baris

# Operasi Penjumlahan
hasil <- bilangan_1 + bilangan_2
cat("Hasil Penjumlahan : ", hasil, "\n")

# Operasi Pengurangan
hasil <- bilangan_1 - bilangan_2
cat("Hasil Pengurangan : ", hasil, "\n")

# Operasi Perkalian
hasil <- bilangan_1 * bilangan_2
cat("Hasil Perkalian : ", hasil, "\n")

# Operasi Pembagian
hasil <- bilangan_1 / bilangan_2
cat("Hasil Pembagian : ", hasil, "\n")

# Operasi Perpangkatan
hasil <- bilangan_1 ^ bilangan_2
cat("Hasil Perpangkatan : ", hasil, "\n")

# Operasi Modulus
hasil <- bilangan_1 %% bilangan_2
cat("Hasil Modulus : ", hasil, "\n")
```

```
nilai_1 <- 10  
nilai_2 <- 5
```

Fungsi cat() digunakan untuk mencetak sesuatu yang ingin ditampilkan

```
cat("nilai_1 > nilai_2 = " ,nilai_1 > nilai_2, "\n")  
cat("nilai_1 < nilai_2 = " ,nilai_1 < nilai_2, "\n")  
cat("nilai_1 >= nilai_2 = " ,nilai_1 >= nilai_2, "\n")  
cat("nilai_1 <= nilai_2 = " ,nilai_1 <= nilai_2, "\n")  
cat("nilai_1 == nilai_2 = " ,nilai_1 == nilai_2, "\n")  
cat("nilai_1 != nilai_2 = " ,nilai_1 != nilai_2, "\n")
```

```
nilai_1 <- TRUE  
nilai_2 <- FALSE  
nilai_3 <- TRUE
```

Logika AND

```
cat("nilai_1 and nilai_2 = ",nilai_1 & nilai_2, "\n")  
cat("nilai_1 and nilai_3 = ",nilai_1 & nilai_3, "\n")
```

Logika OR

```
cat("nilai_1 and nilai_2 = ",nilai_1 || nilai_2, "\n")  
cat("nilai_1 and nilai_3 = ",nilai_1 || nilai_3, "\n")
```

Logika NOT

```
cat("not nilai_1 = ",!nilai_1, "\n")  
cat("not nilai_2 = ",!nilai_2, "\n")
```

ED BY
indodo

```
ujian_mtk <- c(95, 100, 80, 77, 96, 67, 65, 30, 24, 38, 78, 100, 88, 79, 64)  
  
# Mendapatkan nilai ujian yang lebih besar dari 75  
ujian_mtk[ujian_mtk>75]  
  
# Mendapatkan nilai ujian yang nilainya 60 ke bawah untuk remedial  
ujian_mtk[ujian_mtk<=60]  
  
# Mendapatkan nilai ujian antara 80-100  
ujian_mtk[ujian_mtk>=80 & ujian_mtk<=100]
```

```
hujan <- "tidak"

if (hujan == "tidak") {
  print("pergi ke kampus")
} else {
  print("tidak pergi ke kampus")
}
```

```
ujian_statistik <- readline(prompt = "Masukkan hasil ujian Statistika Dasar: ")

if (ujian_statistik > 90) {
  print("Luar biasa bagus!")
} else if (ujian_statistik > 80) {
  print("Sudah bagus")
} else if (ujian_statistik > 70) {
  print("Lumayan bagus")
} else if (ujian_statistik > 60) {
  print("Cukup bagus")
} else {
  print("Belajar lagi ya!")
}
```

```
for (i in 1:10) {
  print(i)
}

# Perulangan mencetak nama mahasiswa

mahasiswa <- list("eko", "teguh", "widodo")

for (i in mahasiswa) {
  print(i)
}
```

```
counter = 4

while (counter > 0) {
  print(counter)
  counter <- counter - 1
}
```

```
mahasiswa <- list("eko", "teguh", "widodo", "putri", "sukma", "dewi")

for (nama in mahasiswa) {
  if (nama == "putri") {
    cat(nama, "akhirnya ketemu")
    break
  }
}
```

```
say_hello_param <- function(nama) {  
  cat("Halo,", nama , "! ", "Selamat Datang di Perkuliahan Statistika Dasar\n")  
}  
  
say_hello_param("Teguh")  
say_hello_param("Putri")
```

MARKED BY EKO TEGO

```
ujian_mtk <- c(95, 100, 80, 77, 96, 67, 65, 30, 24, 38, 78, 100, 88, 79, 64)  
  
rata_rata <- function(x)  
{  
  total <- 0  
  for (i in 1 : length(x)) {  
    total <- total + x[i]  
  }  
  
  return(total/length(x))  
}  
  
rata_rata(ujian_mtk)
```

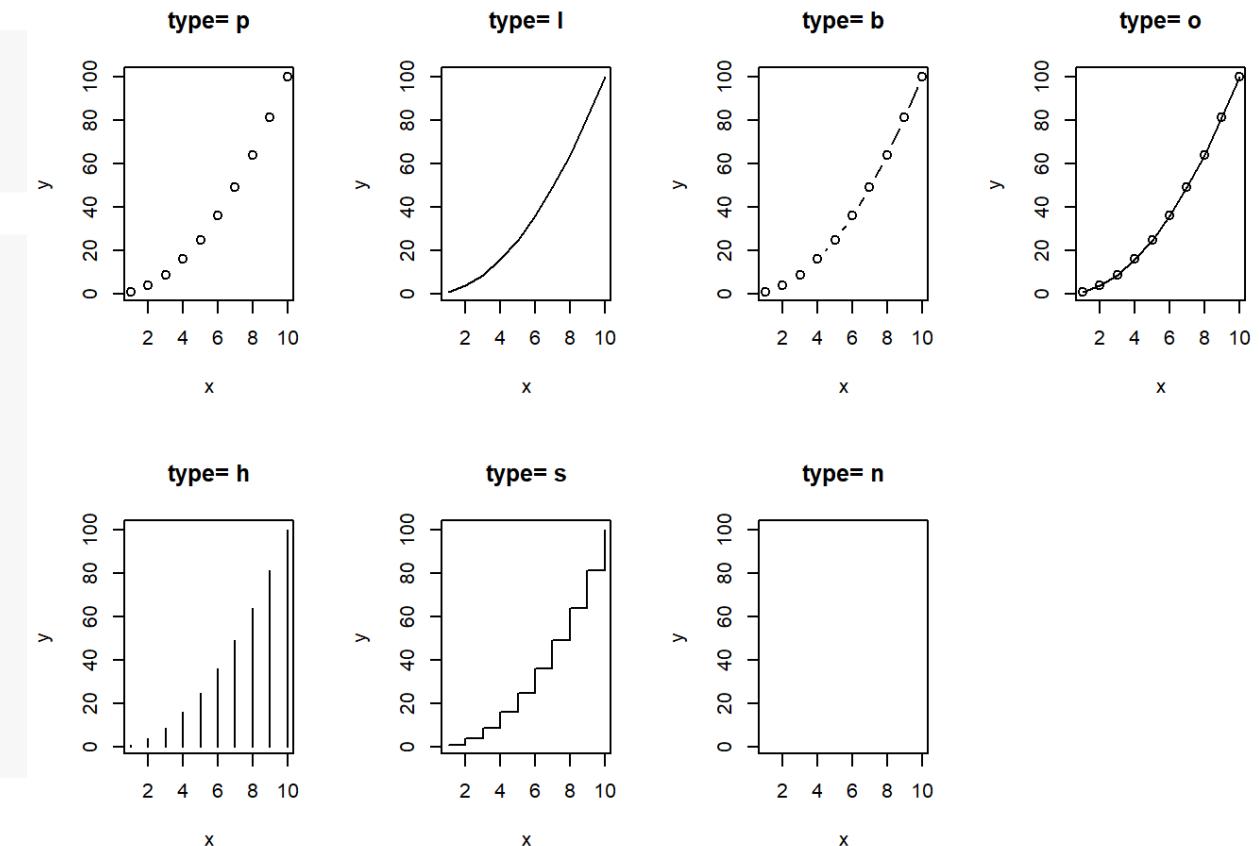
✓ Packages for Data Visualization?

- *plot*

```
# membuat vektor data
x <- c(1:10); y <- x^2
```

```
# membagi jendela grafik menjadi 2 baris dan 4 kolom
par(mfrow=c(2,4))
```

```
# Loop
type <- c("p","l","b","o","h","s","n")
for (i in type){
  plot(x,y, type= i,
    main= paste("type=", i))
}
```



✓ Packages for Data Visualization?

- *Ggplot2 → plot, data, aesthetic, layer*

```
library(ggplot2)

# Menggunakan package openxlsx
library(openxlsx)

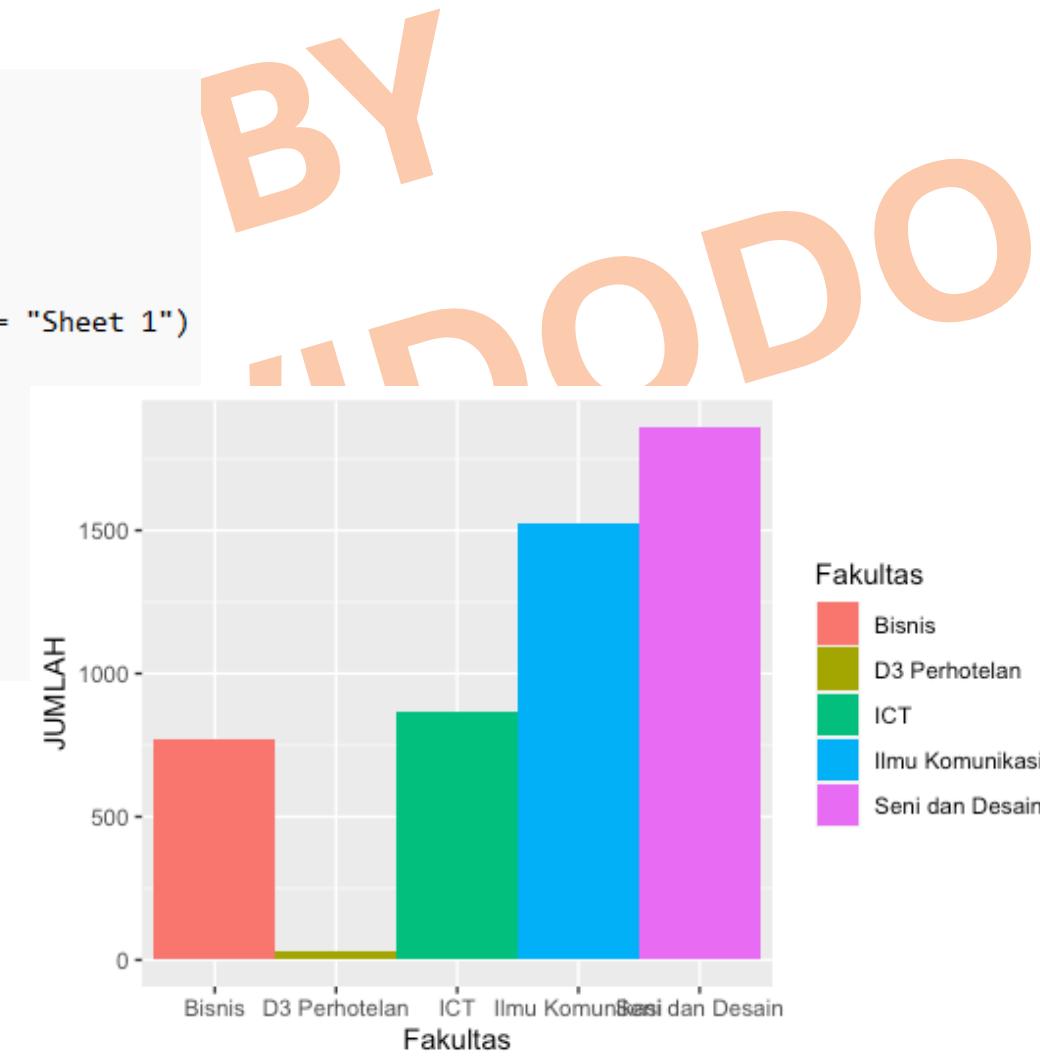
# Membaca file mahasiswa.xlsx
mahasiswa <- read.xlsx("https://academy.dqlab.id/dataset/mahasiswa.xlsx", sheet = "Sheet 1")

# Membuat kanvas
gambar <- ggplot(mahasiswa, aes(x=Fakultas, y=JUMLAH, fill=Fakultas))

# Menambahkan objek bar chart, simpan kembali sebagai variable gambar
gambar <- gambar + geom_bar(width=1, stat="identity")

# Menggambar grafik
gambar
```

EKO IFE



✓ Packages for Data Visualization?

- *Ggplot2 → plot, data, aesthetic, layer*

Menghitung Jumlah Data by Fakultas

```
summaryByFakultas <- aggregate(x=mahasiswa$JUMLAH, by=list(Kategori=mahasiswa$Fakultas, Tahun=mahasiswa$ANGKATAN),  
summaryByFakultas <- setNames(summaryByFakultas, c("Fakultas", "Tahun", "JumlahMahasiswa"))  
summaryByFakultas
```

```
summaryByFakultas$Tahun = as.factor(summaryByFakultas$Tahun)
```

Menampilkan Grafik

```
gambar2 <- ggplot(summaryByFakultas, aes(x=Fakultas, y=JumlahMahasiswa)) +  
  geom_bar(stat = "identity", aes(fill = Tahun), width=0.8, position = position_dodge(width=0.8)) +  
  theme_classic()
```

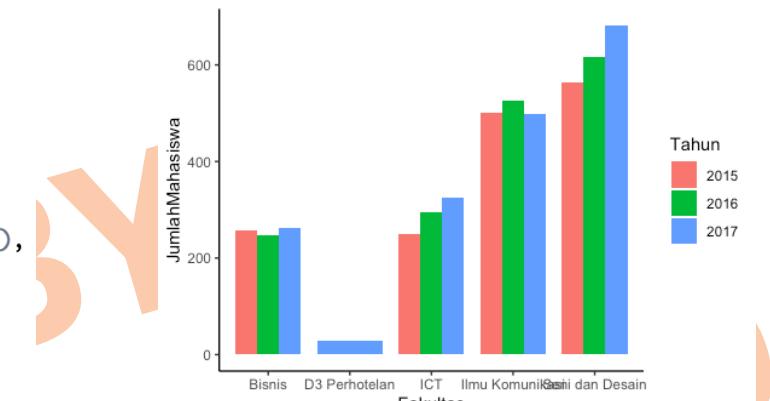
```
gambar2
```

Menampilkan data dengan PieChart

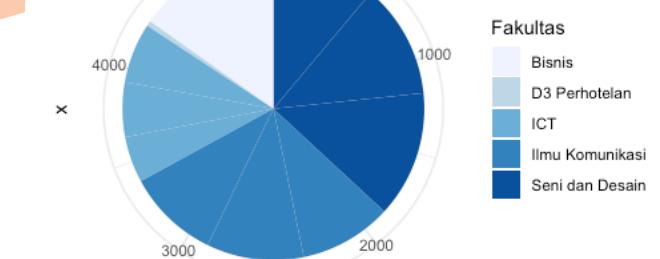
```
piechart <- ggplot(summaryByFakultas, aes(x="", y=JumlahMahasiswa, fill=Fakultas))+ geom_bar(width = 1, stat = "identity")  
piechart <- piechart + coord_polar("y", start=0)  
piechart <- piechart + ggtitle("Disribusi Mahasiswa per Fakultas")  
piechart <- piechart + scale_fill_brewer(palette="Blues")+ theme_minimal()  
piechart <- piechart + guides(fill=guide_legend(title="Fakultas"))  
piechart <- piechart + ylab("Jumlah Mahasiswa")  
piechart
```

Filter Data

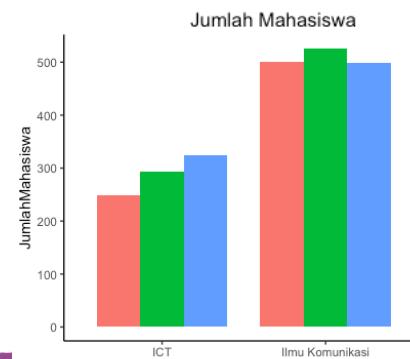
```
summaryByFakultas[summaryByFakultas$Fakultas %in% c("ICT", "Ilmu Komunikasi"),]  
gambar3 <- ggplot(summaryByFakultas[summaryByFakultas$Fakultas %in% c("ICT", "Ilmu Komunikasi"),], aes(x=Fakultas, y=JumlahMahasiswa)) +  
  geom_bar(stat = "identity", aes(fill = Tahun), width=0.8, position = position_dodge(width=0.8)) +  
  theme_classic()  
gambar3
```



Disribusi Mahasiswa per Fakultas



Fakultas
Bisnis
D3 Perhotelan
ICT
Ilmu Komunikasi
Seni dan Desain



Tahun
2015
2016
2017

✓ R untuk Ukuran Statistik?

- *Ukuran Pemusatan (mean, median, modus)*

```
# Data tinggi badan mahasiswa Universitas Y
```

```
tinggi_badan <- c(175, 178, 190, 180, 160, 155, 157, 167, 164, 153,  
156, 167, 170, 169, 170, 167, 170, 181, 183, 167,  
167, 178, 177, 166, 155, 167, 187, 159, 161, 160,  
189, 177, 171, 170, 167, 165, 168, 170, 178, 181,  
162, 163, 163, 171, 173, 163, 164, 161, 160, 172,  
172, 173, 171, 170, 161, 162, 162, 164, 181, 182,  
172, 173, 172, 170, 158, 159, 173, 171, 151, 158,  
177, 181, 151, 159, 178, 180, 182, 174, 182, 180)
```

```
# Mengetahui banyaknya data
```

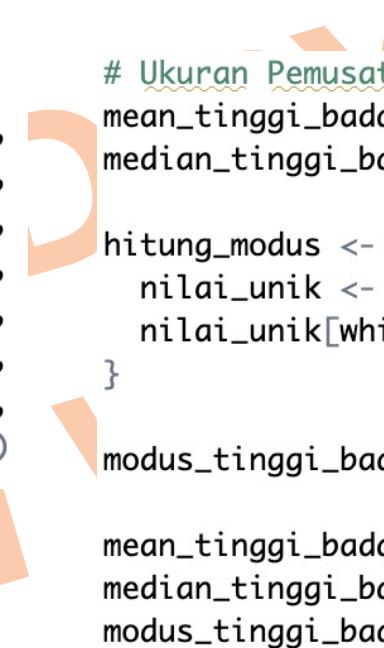
```
banyak_data <- length(tinggi_badan)
```

```
# Visualisasi Tinggi Badan
```

```
hist(tinggi_badan, main = "Histogram Tinggi Badan")
```

```
plot(tinggi_badan, main = "Plot Tinggi Badan")
```

```
boxplot(tinggi_badan, main = "Boxplot Tinggi Badan")
```



```
# Ukuran Pemusatan (mean, median, modus)
```

```
mean_tinggi_badan <- mean(tinggi_badan)  
median_tinggi_badan <- median(tinggi_badan)
```

```
hitung_modus <- function(x) {  
  nilai_unik <- unique(x)  
  nilai_unik[which.max(tabulate(match(x, nilai_unik)))]  
}
```

```
modus_tinggi_badan <- hitung_modus(tinggi_badan)
```

```
mean_tinggi_badan  
median_tinggi_badan  
modus_tinggi_badan
```

✓ R untuk Ukuran Statistik?

- *Ukuran lokasi (persentil, desil, kuartil)*

```
# Ukuran Lokasi
```

```
Q1_tinggi_badan <- quantile(tinggi_badan, probs = seq(1/4, 0))
Q1_tinggi_badan

Q3_tinggi_badan <- quantile(tinggi_badan, probs = seq(3/4, 0))
Q3_tinggi_badan

D3_tinggi_badan <- quantile(tinggi_badan, probs = seq(3/10, 0))
D3_tinggi_badan

D8_tinggi_badan <- quantile(tinggi_badan, probs = seq(8/10, 0))
D8_tinggi_badan

P21_tinggi_badan <- quantile(tinggi_badan, probs = seq(21/100, 0))
P21_tinggi_badan

P45_tinggi_badan <- quantile(tinggi_badan, probs = seq(45/100, 0))
P45_tinggi_badan

P93_tinggi_badan <- quantile(tinggi_badan, probs = seq(93/100, 0))
P93_tinggi_badan
```

ED BY
HWIDODO

✓ R untuk Ukuran Statistik?

- *Ukuran penyebaran (range, ragam/varians, standar deviasi/simpangan baku, dll)*

```
# Ukuran Penyebaran (range, ragam/varians, standar deviasi, koefisien variasi)

range_tinggi_badan <- max(tinggi_badan) - min(tinggi_badan)
range_tinggi_badan

varians_tinggi_badan <- var(tinggi_badan)
varians_tinggi_badan

stdev_tinggi_badan <- sd(tinggi_badan)
stdev_tinggi_badan

koef_variasi_tinggi_badan <- stdev_tinggi_badan / mean_tinggi_badan * 100
koef_variasi_tinggi_badan
```

- *Ukuran kemencengan dan keruncingan*
- *Penghitungan ukuran statistik dengan packages*

```
# Packages untuk memudahkan penghitungan ukuran statistik

install.packages("psych")
library(psych)

describe(tinggi_badan)

install.packages("pastecs")
library(pastecs)

stat.desc(tinggi_badan)
```

```
# Ukuran Kemencengan dan Keruncingan

install.packages("moments")
library(moments)

skewness_tinggi_badan <- skewness(tinggi_badan)
skewness_tinggi_badan

# Jika nilai skewness_tinggi_badan < 0 atau negatif -> kurva cenderung condong ke kiri (kurva negatif)
# Jika nilai skewness_tinggi_badan > 0 atau positif -> kurva cenderung condong ke kanan (kurva positif)
# Jika nilai skewness_tinggi_badan mendekati nol atau 0 -> kurva cenderung simetris

kurtosis_tinggi_badan <- kurtosis(tinggi_badan)
kurtosis_tinggi_badan

# Jika kurtosis_tinggi_badan < 3, platikurtis -> kurva cenderung datar dan puncak tidak terlalu tinggi
# Jika kurtosis_tinggi_badan = 3, mesokurtis -> kurva normal (tidak terlalu tajam dan datar)
# Jika kurtosis_tinggi_badan > 3, leptokurtis -> kurva cenderung terlihat lancip dan tinggi

# Cara lain menghitung skewness dan kurtosis dengan package e1071

install.packages("e1071")
library(e1071)
skewness_tinggi_badan2 <- skewness(tinggi_badan, type = 2) # type=2 menggunakan moment based formula
kurtosis_tinggi_badan2 <- kurtosis(tinggi_badan, type = 2)

# Visualisasi Skewness dan Kurtosis
hist(tinggi_badan, prob = TRUE)
lines(density(tinggi_badan), col = "red")
```

✓ Membaca RAW data excel dengan R

```
# Membaca raw data excel
install.packages("readxl") # install packages readxl

library(readxl) # load library readxl

data_ipm_hls <- read_excel(file.choose()) # memunculkan popup untuk memilih file excel
data_ipm_hls # menampilkan data

View(data_ipm_hls)
```

	Kabupaten/Kota	IPM	HLS
1	A1	67.25	12.27
2	A2	69.25	12.56
3	A3	66.23	11.51
4	A4	63.15	10.70
5	A5	65.67	12.07
6	A6	70.12	13.03
7	A7	65.20	11.34
8	A8	66.35	12.00
9	A9	70.34	11.65

Melakukan pengecekan normalitas ($z = \bar{x} / \text{simp_baku}$) -> UJI NORMALITAS

```
install.packages("nortest") # install packages nortest
```

```
library(nortest) # load library nortest
```

Shapiro-Wilk normality test

```
shapiro.test(data_ipm_hls$IPM)
```

p-value = 0.06165, > 5% -> 0.05 --> variabel IPM berdistribusi normal

```
shapiro.test(data_ipm_hls$HLS)
```

p-value = 0.06981, > 5% --> variabel HLS berdistribusi normal

nODO

END

> shapiro.test(data_ipm_hls\$IPM)
Shapiro-Wilk normality test
data: data_ipm_hls\$IPM
W = 0.94138, p-value = 0.06165

> shapiro.test(data_ipm_hls\$HLS)
Shapiro-Wilk normality test

data: data_ipm_hls\$HLS
W = 0.94315, p-value = 0.06981

✓ Menghitung korelasi

Melakukan pengujian KORELASI - PEARSON

```
cor.test(data_ipm_hls$IPM, data_ipm_hls$HLS, method = 'pearson')
```

nilai korelasi = 0.893947 -> korelasi kuat, positif -> searah = HLS naik, IPM naik

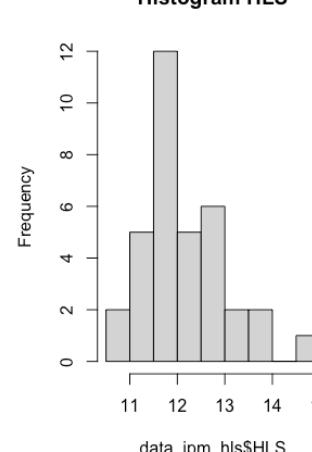
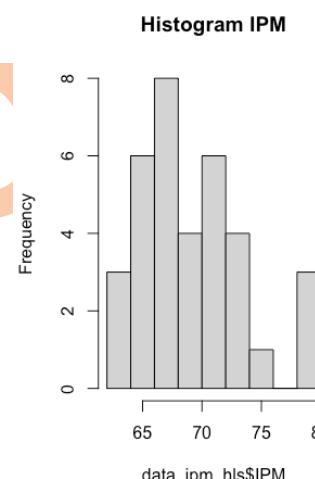
```
> cor.test(data_ipm_hls$IPM, data_ipm_hls$HLS, method = 'pearson')
```

Pearson's product-moment correlation

```
data: data_ipm_hls$IPM and data_ipm_hls$HLS
t = 11.458, df = 33, p-value = 4.816e-13
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7986102 0.9455189
sample estimates:
```

```
cor
0.893947
```

EKC



RKL

Visualisasikan

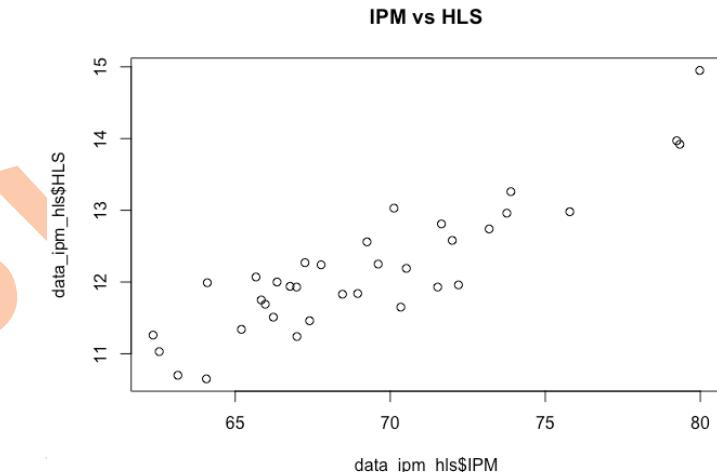
```
plot(data_ipm_hls$IPM, data_ipm_hls$HLS, main = "IPM vs HLS")
```

par membagi tampilan menjadi sekian kolom

```
par(mfrow = c(1, 2))
```

```
hist(data_ipm_hls$IPM, main = "Histogram IPM")
```

```
hist(data_ipm_hls$HLS, main = "Histogram HLS")
```



B

3



✓ Menghitung regresi

```
# REGRESI LINIER
# Menggunakan data curah hujan dan penjualan payung

# 1. Membaca data dari file excel
data_hujan_payung <- read_excel(file.choose())
data_hujan_payung

# Rename nama kolom menjadi curah_hujan dan penjualan_payung
colnames(data_hujan_payung)[colnames(data_hujan_payung) == "Curah Hujan (X)"] = "curah_hujan"
colnames(data_hujan_payung)[colnames(data_hujan_payung) == "Penjualan Payung (Y)"] = "penjualan_payung"
data_hujan_payung

# 2. Visualisasi data
plot(data_hujan_payung$curah_hujan, data_hujan_payung$penjualan_payung,
      main = "Hubungan antara Curah Hujan dan Penjualan Payung",
      xlab = "Curah Hujan (mm)", ylab = "Penjualan Payung")

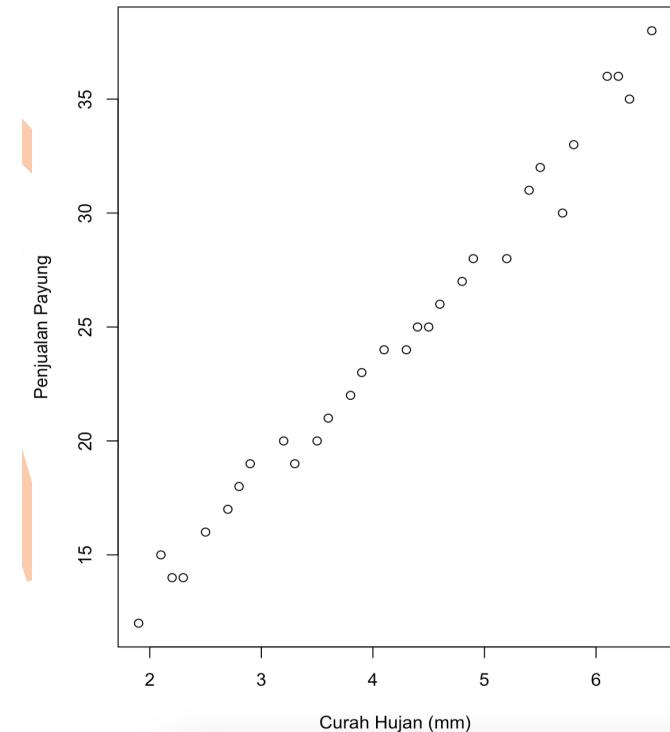
# 3. Uji Normalitas
shapiro.test(data_hujan_payung$curah_hujan)
shapiro.test(data_hujan_payung$penjualan_payung)
# --> kedua variabel p-value > 5% -> distribusi normal

# 4. Uji Korelasi
cor.test(data_hujan_payung$curah_hujan, data_hujan_payung$penjualan_payung, method = 'pearson')
# --> korelasinya 99% (sangat kuat) dan searah -> semakin tinggi curah hujan, penjualan payung semakin tinggi

# 5. Regresi Linier Sederhana
model_regresi <- lm(data_hujan_payung$penjualan_payung ~ data_hujan_payung$curah_hujan)
summary(model_regresi)
```



Hubungan antara Curah Hujan dan Penjualan Payung



```
Call:
lm(formula = data_hujan_payung$penjualan_payung ~ data_hujan_payung$curah_hujan)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.23880 -0.54954 -0.06048  0.66911  1.68151 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.6033    0.5589   4.658 7.06e-05 ***
data_hujan_payung$curah_hujan  5.1992    0.1274  40.808 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9567 on 28 degrees of freedom
Multiple R-squared:  0.9835,    Adjusted R-squared:  0.9829 
F-statistic: 1665 on 1 and 28 DF,  p-value: < 2.2e-16
```