

TUGAS BESAR ADS - Analisis Statistik Deskriptif IPK, Jam Belajar, dan Karakteristik Mahasiswa ITERA

KELOMPOK 1

2025-11-18

1. Persiapan Lingkungan (Environment Setup)

Pada tahap ini, saya akan memuat semua library yang dibutuhkan dan mendefinisikan fungsi-fungsi statistik agar lebih mudah

```
# =====MEMUAT LIBRARY=====

library(ggplot2)
library(dplyr)
```

```
##
## 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
```

```
library(readr)

# =====DEFINISI FUNGSI=====

# 1. Menghitung Skewness
skew_manual <- function(x) {
  x <- x[!is.na(x)]
  m <- mean(x)
  s <- sd(x)
  mean(((x - m) / s)^3)
}

# 2. Interpretasi
interpret_skewness <- function(value) {
  if (is.na(value)) return("Data tidak cukup untuk menghitung skewness.")

  if (value > 1) {
```

```

    return("Distribusi sangat condong ke kanan (right-skewed kuat).")
  } else if (value > 0.5) {
    return("Distribusi condong ke kanan dengan tingkat moderat.")
  } else if (value > 0) {
    return("Distribusi sedikit condong ke kanan.")
  } else if (value == 0) {
    return("Distribusi simetris.")
  } else if (value > -0.5) {
    return("Distribusi sedikit condong ke kiri.")
  } else if (value > -1) {
    return("Distribusi condong ke kiri dengan tingkat moderat.")
  } else {
    return("Distribusi sangat condong ke kiri (left-skewed kuat).")
  }
}

```

2. Data Loading & Preprocessing

membaca dataset dan melakukan pembersihan data (Data Cleansing) sekaligus untuk semua kolom yang bermasalah. Ini memastikan data siap digunakan untuk visualisasi apa pun.

```

# Load Data
data <- read.csv("Dataset/dataset.csv")

# --- DATA CLEANSING ---

# A. Pembersihan Kolom IPK Terakhir
# Mengganti karakter non-numerik dengan NA, memperbaiki desimal, dan filter range logis (0-4)
data$IPK.Terakhir..[data$IPK.Terakhir.. %in% c("-", ".", ",", "", " ")] <- NA
data$IPK.Terakhir.. <- gsub(",", ".", data$IPK.Terakhir..)
data$IPK.Terakhir.. <- as.numeric(data$IPK.Terakhir..)

```

Warning: NAs introduced by coercion

```

data$IPK.Terakhir..[data$IPK.Terakhir.. > 4 | data$IPK.Terakhir.. == 0] <- NA

# B. Pembersihan Kolom Jam Belajar
# Rename kolom, perbaikan desimal, dan filter range logis (0-60 jam/minggu)
names(data)[names(data) == "Rata.rata.belajar.perminggu..dalam.jam.."] <- "BelajarJam"
data$BelajarJam <- gsub(",", ".", data$BelajarJam)
data$BelajarJam <- as.numeric(data$BelajarJam)

```

Warning: NAs introduced by coercion

```

data$BelajarJam[data$BelajarJam < 0 | data$BelajarJam > 60] <- NA

# C. Pembersihan Kolom Anggota Keluarga
# Mengambil hanya angka, mengubah ke integer, batas wajar max 12
data$Jumlah.Anggota.Keluarga.. <- as.character(data$Jumlah.Anggota.Keluarga..)
data$Jumlah.Anggota.Keluarga.. <- gsub("[^0-9]", "", data$Jumlah.Anggota.Keluarga..)

```

```

data$Jumlah.Anggota.Keluarga..[data$Jumlah.Anggota.Keluarga.. == ""] <- NA
data$Jumlah.Anggota.Keluarga.. <- as.integer(data$Jumlah.Anggota.Keluarga..)
data$Jumlah.Anggota.Keluarga..[data$Jumlah.Anggota.Keluarga.. > 12] <- NA

# D. Pembersihan Kolom Berat Badan
# Hapus "kg", spasi, ambil angka
data$Berat.Badan.. <- as.character(data$Berat.Badan..)
data$Berat.Badan.. <- gsub("kg", "", data$Berat.Badan.., ignore.case = TRUE)
data$Berat.Badan.. <- gsub("[^0-9.]", "", data$Berat.Badan..)
data$Berat.Badan.. <- as.numeric(data$Berat.Badan..)

# E. Pembersihan Kolom Tinggi Badan
# Hapus "cm", filter range logis (120-220 cm)
data$Tinggi.Badan.. <- as.character(data$Tinggi.Badan..)
data$Tinggi.Badan.. <- gsub("cm", "", data$Tinggi.Badan.., ignore.case = TRUE)
data$Tinggi.Badan.. <- gsub("[^0-9.]", "", data$Tinggi.Badan..)
data$Tinggi.Badan.. <- as.numeric(data$Tinggi.Badan..)
data$Tinggi.Badan..[data$Tinggi.Badan.. < 120 | data$Tinggi.Badan.. > 220] <- NA

# Menampilkan Ringkasan Data Bersih
summary(data)

```

```

##      NIM      Program.Studi..  IPK.Terakhir..  Jenis.Kelamin
## Min.   : 1.0    Length:458      Min.   :1.400    Length:458
## 1st Qu.:115.2    Class :character  1st Qu.:3.100    Class :character
## Median :229.5    Mode  :character  Median :3.400    Mode  :character
## Mean   :229.5
## 3rd Qu.:343.8
## Max.   :458.0
##                      NA's   :135
## Tinggi.Badan..  Berat.Badan..  Pendidikan.terakhir  BelajarJam
## Min.   :142.0    Min.   : 33.00      Length:458          Min.   : 0.00
## 1st Qu.:155.0    1st Qu.: 47.00      Class :character     1st Qu.: 3.00
## Median :161.0    Median : 53.00      Mode  :character     Median : 7.00
## Mean   :162.3    Mean   : 56.35
## 3rd Qu.:169.5    3rd Qu.: 62.00
## Max.   :192.0    Max.   :160.00
## NA's   :7        NA's   :2
## Apakah.penerima.beasiswa..
## Length:458
## Class :character
## Mode  :character
##
##
##
## Asal.Daerah..Kota.beserta.Provinsi.....Ex...Padang..Sumatera.Barat.
## Length:458
## Class :character
## Mode  :character
##
##
##

```

```

##
## Pekerjaan.saat.ini..selain.kuliah.
## Length:458
## Class :character
## Mode :character
##
##
##
##
## Akses.Internet..yang.paling.sering.digunakan. Keterlibatan.Organisai
## Length:458 Length:458
## Class :character Class :character
## Mode :character Mode :character
##
##
##
## Uang.Saku.....yang.diberikan.oleh.orangtua.
## Length:458
## Class :character
## Mode :character
##
##
##
##
## Jenis.Tempat.Tinggal.....Tempat.tinggal.saat.ini..
## Length:458
## Class :character
## Mode :character
##
##
##
##
## Jarak.rumah.dari.kampus.ITERA.. Jenis.pekerjaan.Ayah.. Jenis.pekerjaan.Ibu..
## Length:458 Length:458 Length:458
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
##
##
## Pendapatan.Orangtua.. Jumlah.Anggota.Keluarga..
## Length:458 Min. : 2.000
## Class :character 1st Qu.: 4.000
## Mode :character Median : 5.000
## Mean : 4.665
## 3rd Qu.: 5.000
## Max. :10.000
## NA's :4

```

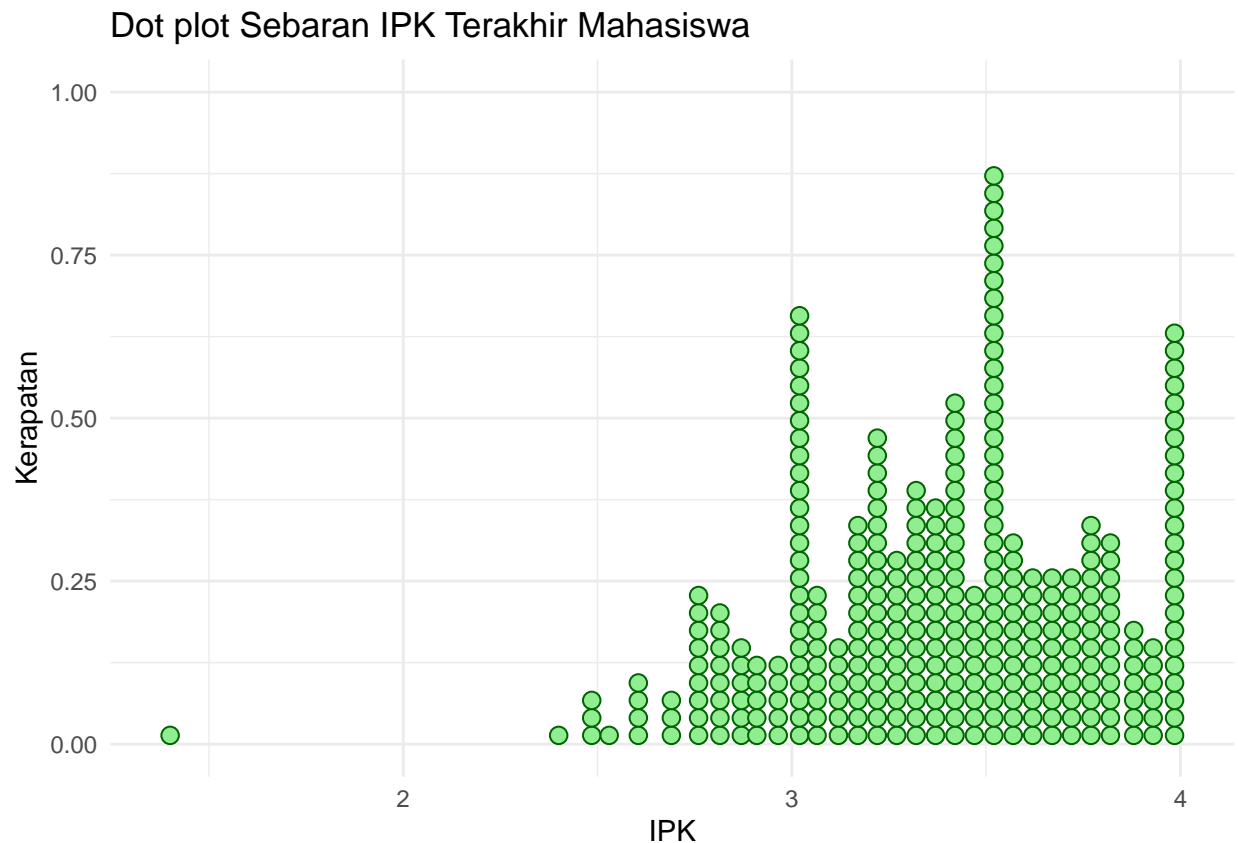
3. Analisis Variabel Akademik

A. Sebaran IPK Terakhir

visualisasi sebaran IPK mahasiswa menggunakan Dot Plot, stem-leaf, dan analisis Outlier menggunakan metode IQR.

```
# 1. Dot Plot
ggplot(data, aes(x = `IPK.Terakhir..`)) +
  geom_dotplot(binwidth = 0.045, fill = "lightgreen", color = "darkgreen") +
  labs(
    title = "Dot plot Sebaran IPK Terakhir Mahasiswa",
    x = "IPK",
    y = "Kerapatan"
  ) +
  theme_minimal()
```

```
## Warning: Removed 135 rows containing missing values or values outside the scale range
## ('stat_bindot()').
```



```
# 2. Stem and Leaf Plot
cat("Stem and Leaf Plot IPK:\n")
```

```
## Stem and Leaf Plot IPK:
```

```
stem(data$`IPK.Terakhir..`, scale = 2)
```

```
##
```

```
## The decimal point is 1 digit(s) to the left of the |
##
## 14 | 0
## 15 |
## 16 |
## 17 |
## 18 |
## 19 |
## 20 |
## 21 |
## 22 |
## 23 |
## 24 | 07
## 25 | 003
## 26 | 00017
## 27 | 01455667888
## 28 | 00000033578999
## 29 | 0001257788
## 30 | 0000000000000001112233334566677778
## 31 | 00344445567777889999
## 32 | 000000000111222444556668889999
## 33 | 000001112333334555567888899999
## 34 | 000000000000122333345777889999
## 35 | 0000000000000000000000000001133444555677888889
## 36 | 00000024445677788889
## 37 | 000002444455566888899999
## 38 | 000011133334667999
## 39 | 014445577
## 40 | 00000000000000000000000000
```

```
# 3. Deteksi Outlier (Metode IQR)
```

```
cat("=====")
```

```
## =====
```

```
cat("\n")
```

```
Q1 <- quantile(data$`IPK.Terakhir..`, 0.25, na.rm = TRUE)
Q3 <- quantile(data$`IPK.Terakhir..`, 0.75, na.rm = TRUE)
IQR_value <- Q3 - Q1
```

```
batas_bawah <- Q1 - (1.5 * IQR_value)
batas_atas <- Q3 + (1.5 * IQR_value)
```

```
cat("\n--- HASIL PERHITUNGAN IQR ---\n")
```

```
##
```

```
## --- HASIL PERHITUNGAN IQR ---
```

```
cat("Kuartil 1 (25%) :", Q1, "\n")
```

```
## Kuartil 1 (25%) : 3.1
```

```

cat("Kuartil 3 (75%) :", Q3, "\n")

## Kuartil 3 (75%) : 3.655

cat("Nilai IQR      :", IQR_value, "\n")

## Nilai IQR      : 0.555

cat("Batas Bawah   :", batas_bawah, "\n")

## Batas Bawah    : 2.2675

cat("Batas Atas    :", batas_atas, "\n")

## Batas Atas     : 4.4875

cat("-----\n")

## -----

data_outlier <- data[which(data$`IPK.Terakhir..` < batas_bawah |
                           data$`IPK.Terakhir..` > batas_atas), ]

if(nrow(data_outlier) > 0) {
  cat("DITEMUKAN", nrow(data_outlier), "DATA OUTLIER:\n")
  print(data_outlier[, c("IPK.Terakhir..")])
} else {
  cat("Data bersih. Tidak ditemukan outlier dengan metode IQR.\n")
}

## DITEMUKAN 1 DATA OUTLIER:
## [1] 1.4

cat("=====\n")

## =====

cat("Ringkasan\n")

## Ringkasan

summary(data$`IPK.Terakhir..`)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##      1.400   3.100   3.400   3.373   3.655   4.000     135

```

B. Distribusi Jam Belajar

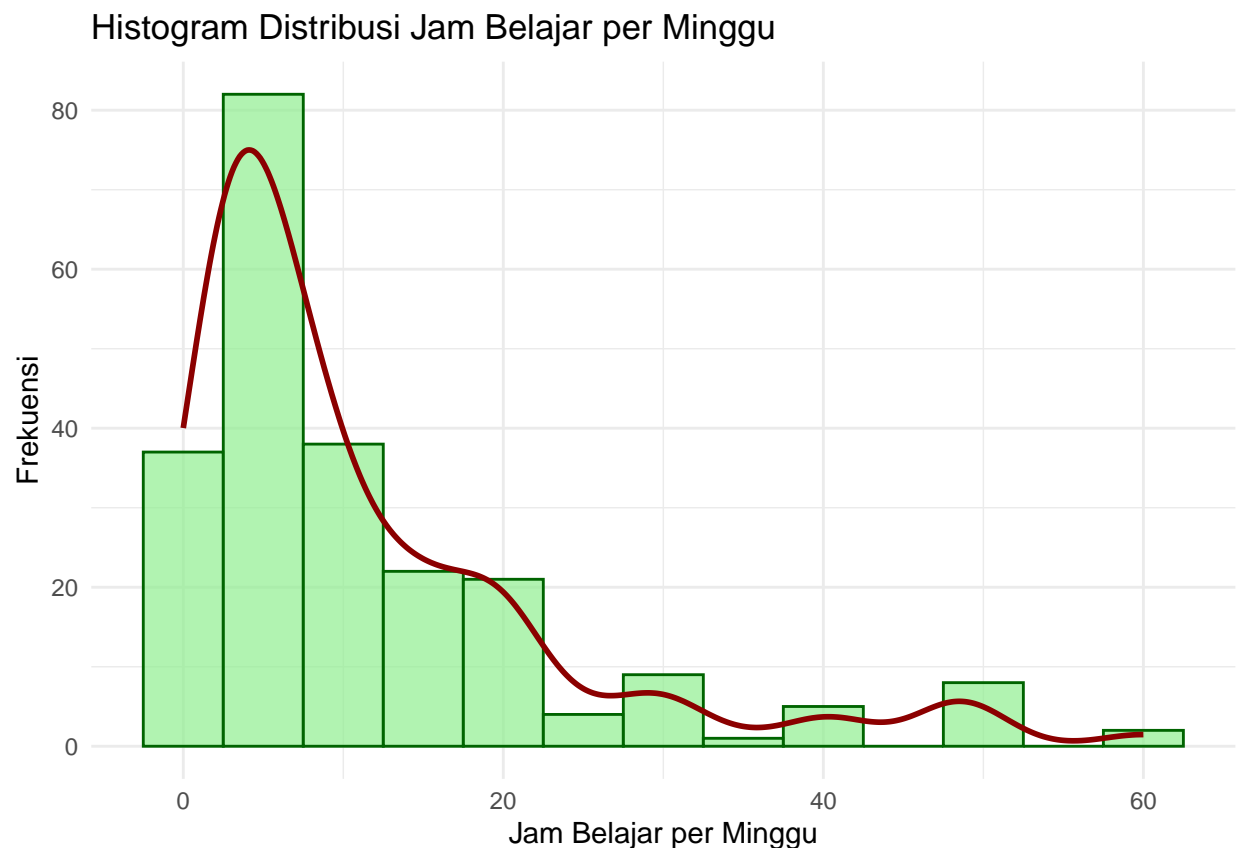
jam belajar mahasiswa per minggu.

```
# 1. Histogram
names(data)[names(data) == "Rata.rata.belajar.perminggu..dalam.jam..."] <- "BelajarJam"
hist_plot <- ggplot(data, aes(x = BelajarJam)) +
  geom_histogram(
    binwidth = 5, # Anda bisa menyesuaikan lebar bin
    fill = "lightgreen",
    color = "darkgreen",
    alpha = 0.7
  ) +
  geom_density(aes(y = after_stat(count) * 5), color = "darkred", linewidth = 1) +
  labs(
    title = "Histogram Distribusi Jam Belajar per Minggu",
    x = "Jam Belajar per Minggu",
    y = "Frekuensi"
  ) +
  theme_minimal()

print(hist_plot)
```

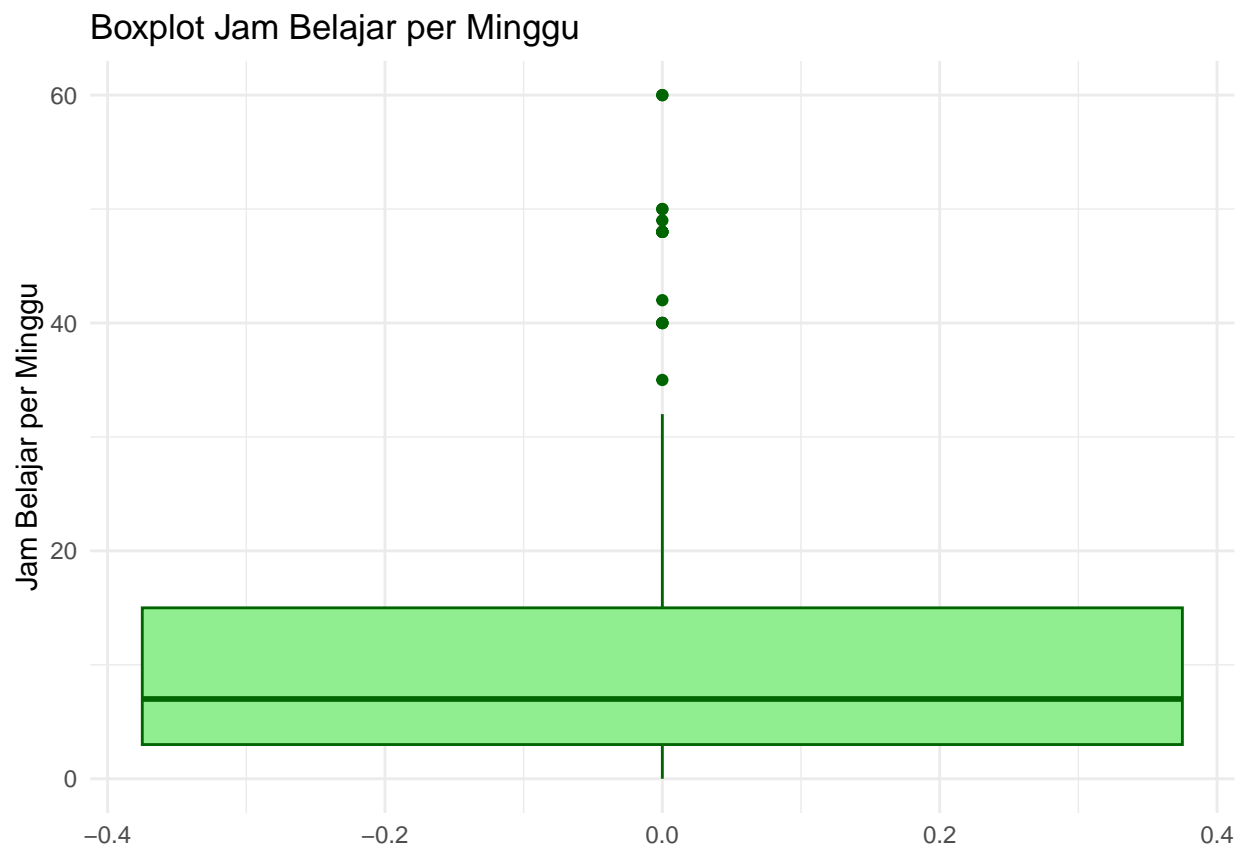
```
## Warning: Removed 229 rows containing non-finite outside the scale range
## ('stat_bin()').
```

```
## Warning: Removed 229 rows containing non-finite outside the scale range
## ('stat_density()').
```




```
# 2. Boxplot
ggplot(data, aes(y = BelajarJam)) +
  geom_boxplot(fill = "lightgreen", color = "darkgreen") +
  labs(
    title = "Boxplot Jam Belajar per Minggu",
    y = "Jam Belajar per Minggu"
  ) +
  theme_minimal()
```

```
## Warning: Removed 229 rows containing non-finite outside the scale range
## ('stat_boxplot()').
```



```
cat("Ringkasan\n")
```

```
## Ringkasan
```

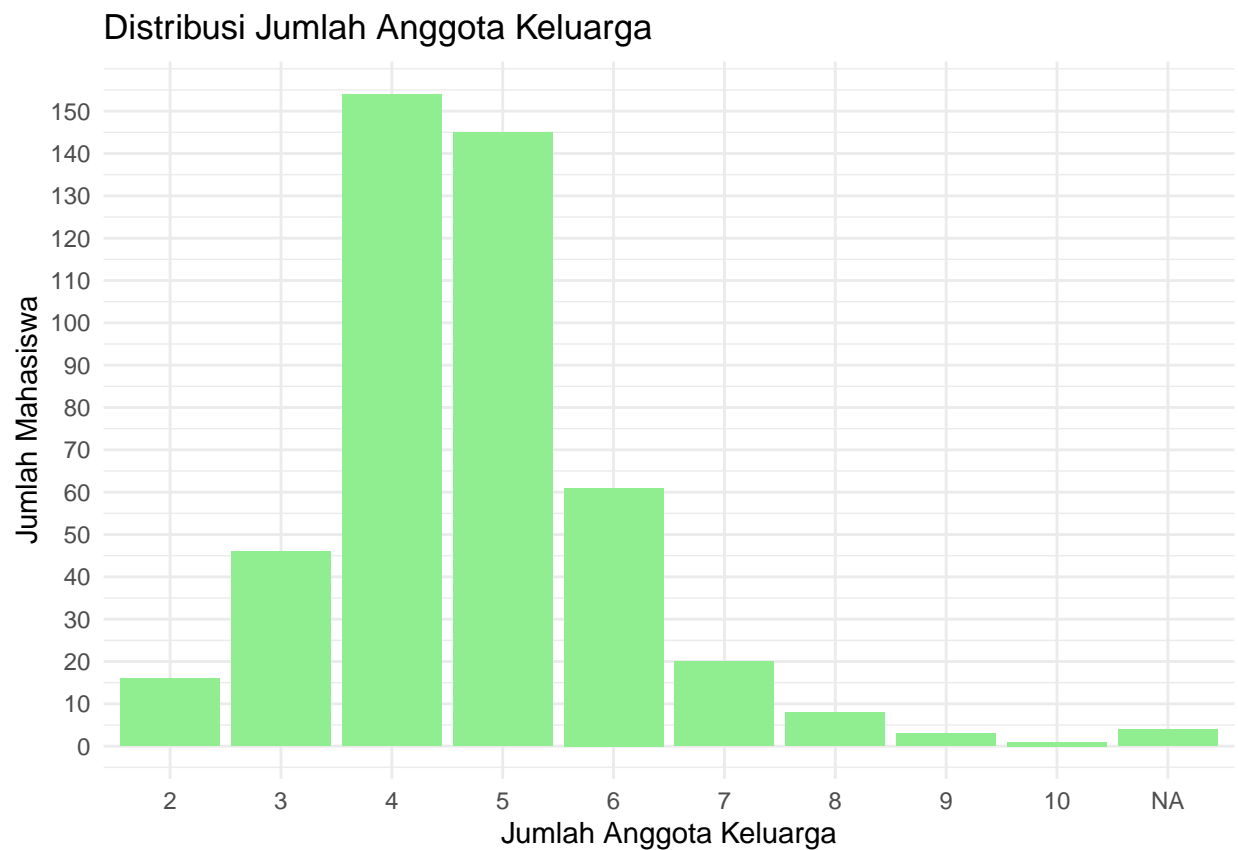
```
summary(data$BelajarJam)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      0.00    3.00    7.00   11.58   15.00   60.00    229
```

4. Analisis Variabel Demografis & Fisik

A. Jumlah Anggota Keluarga

```
ggplot(data, aes(x = factor(Jumlah.Anggota.Keluarga..))) +  
  geom_bar(fill = "lightgreen") +  
  scale_y_continuous(  
    breaks = seq(0, max(table(data$Jumlah.Anggota.Keluarga..)), by = 10)  
  ) +  
  labs(  
    title = "Distribusi Jumlah Anggota Keluarga",  
    x = "Jumlah Anggota Keluarga",  
    y = "Jumlah Mahasiswa"  
  ) +  
  theme_minimal()
```



```
summary(data$`Jumlah.Anggota.Keluarga..`)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	2.000	4.000	5.000	4.665	5.000	10.000	4

B. Analisis Fisik (Berat & Tinggi Badan)

melihat kecendrungan (skewness) dari data fisik mahasiswa.

```
# --- BERAT BADAN ---

cat("Ringkasan berat badan:\n")

## Ringkasan berat badan:

summary(data$Berat.Badan..)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##    33.00  47.00   53.00   56.35  62.00  160.00         2

cat("\n\nRingkasan tinggi badan:\n")

##
##
## Ringkasan tinggi badan:

summary(data$Tinggi.Badan..)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##    142.0   155.0   161.0   162.3   169.5   192.0         7

skew_berat <- skew_manual(data$Berat.Badan..)
cat("\n===== \n")

##
## =====

cat("--- ANALISIS BERAT BADAN ---\n")

## --- ANALISIS BERAT BADAN ---

cat("Nilai Skewness :", round(skew_berat, 4), "\n")

## Nilai Skewness : 1.8729

cat("Kesimpulan      :", interpret_skewness(skew_berat), "\n\n")

## Kesimpulan      : Distribusi sangat condong ke kanan (right-skewed kuat).

# --- TINGGI BADAN ---
skew_tinggi <- skew_manual(data$Tinggi.Badan..)

cat("--- ANALISIS TINGGI BADAN ---\n")

## --- ANALISIS TINGGI BADAN ---
```

```
cat("Nilai Skewness :", round(skew_tinggi, 4), "\n")
```

```
## Nilai Skewness : 0.267
```

```
cat("Kesimpulan      :", interpret_skewness(skew_tinggi), "\n")
```

```
## Kesimpulan       : Distribusi sedikit condong ke kanan.
```