

R Kelompok 8

# R STUDIO

Pengenalan Software dan  
Studi Kasus

Read more →

ANALISIS DATA PENELITIAN KUANTITATIF

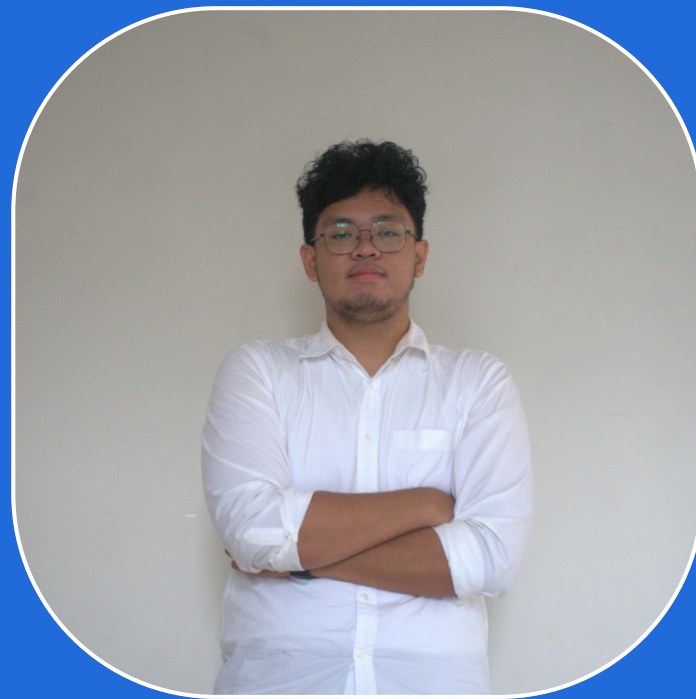


# Meet our team



**T.A.S Al Ayyubi Sind**

NIM: 235150407111067



**Revan Adli Pratama**

NIM: 235150400111052



**Fayyadh Radhwa F**

NIM:235150401111058



**M Nabil Fatahilah**

NIM: 235150401111063



## ABOUT R STUDIO

# R Studio

RStudio adalah IDE untuk bahasa R yang digunakan dalam analisis data, statistik, dan machine learning. Fitur utamanya mencakup code editor, console interaktif, data viewer, plot viewer, dan package manager. RStudio mempermudah pemrosesan dan visualisasi data secara efisien.

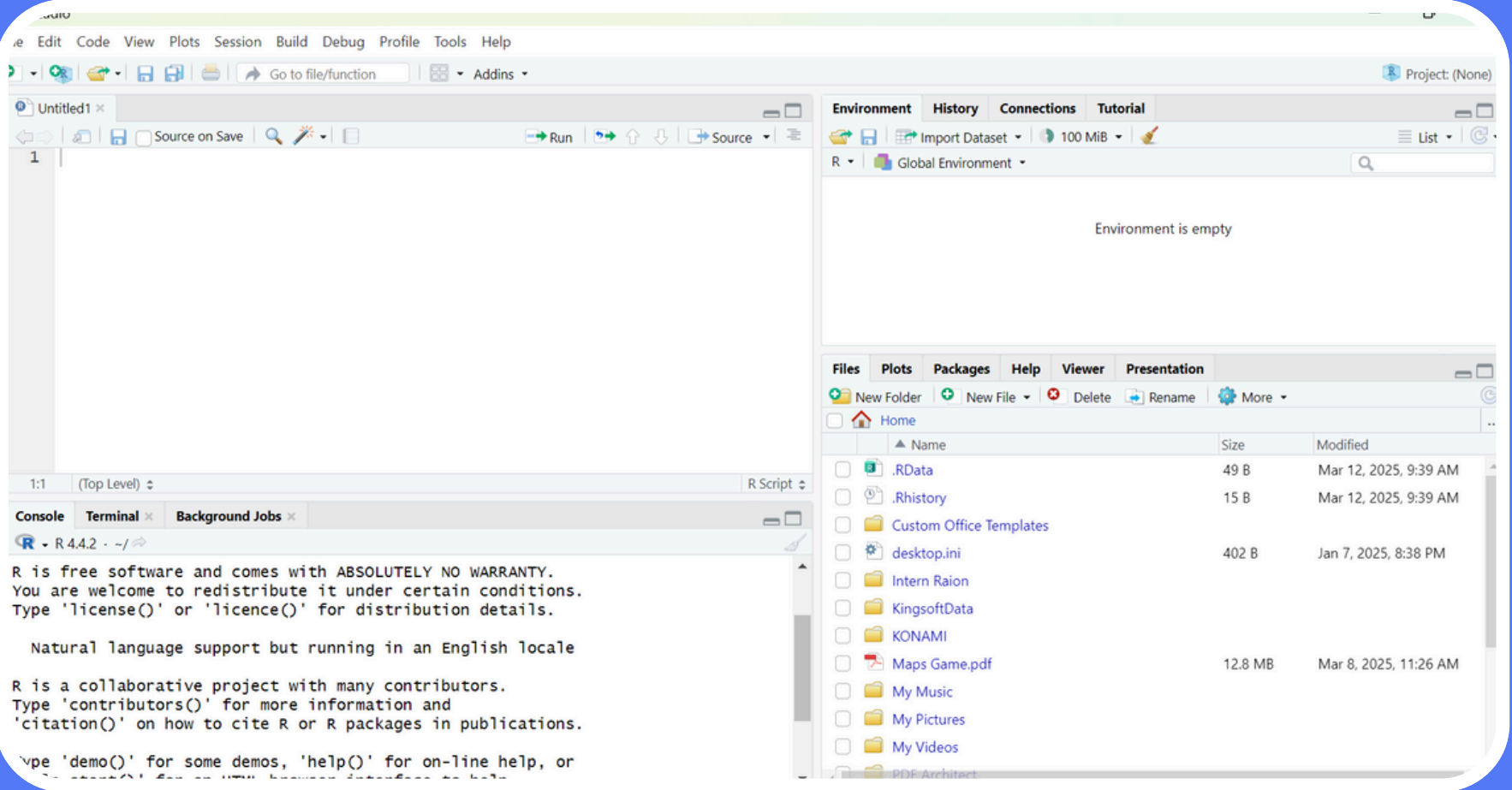
### Keunggulan

- **Open-source & Gratis**
- **Dukungan Statistik & Data Science Kuat** – Banyak paket untuk analisis data, machine learning, dan visualisasi.
- **Integrasi dengan R Markdown** – Mudah membuat laporan interaktif.

### Kekurangan

- **Kurva Belajar Curam**
- **Kurang Optimal untuk Big Data**
- **Tergantung pada R** – Tidak bisa menjalankan bahasa lain seperti Python secara langsung tanpa tambahan ekstensi.

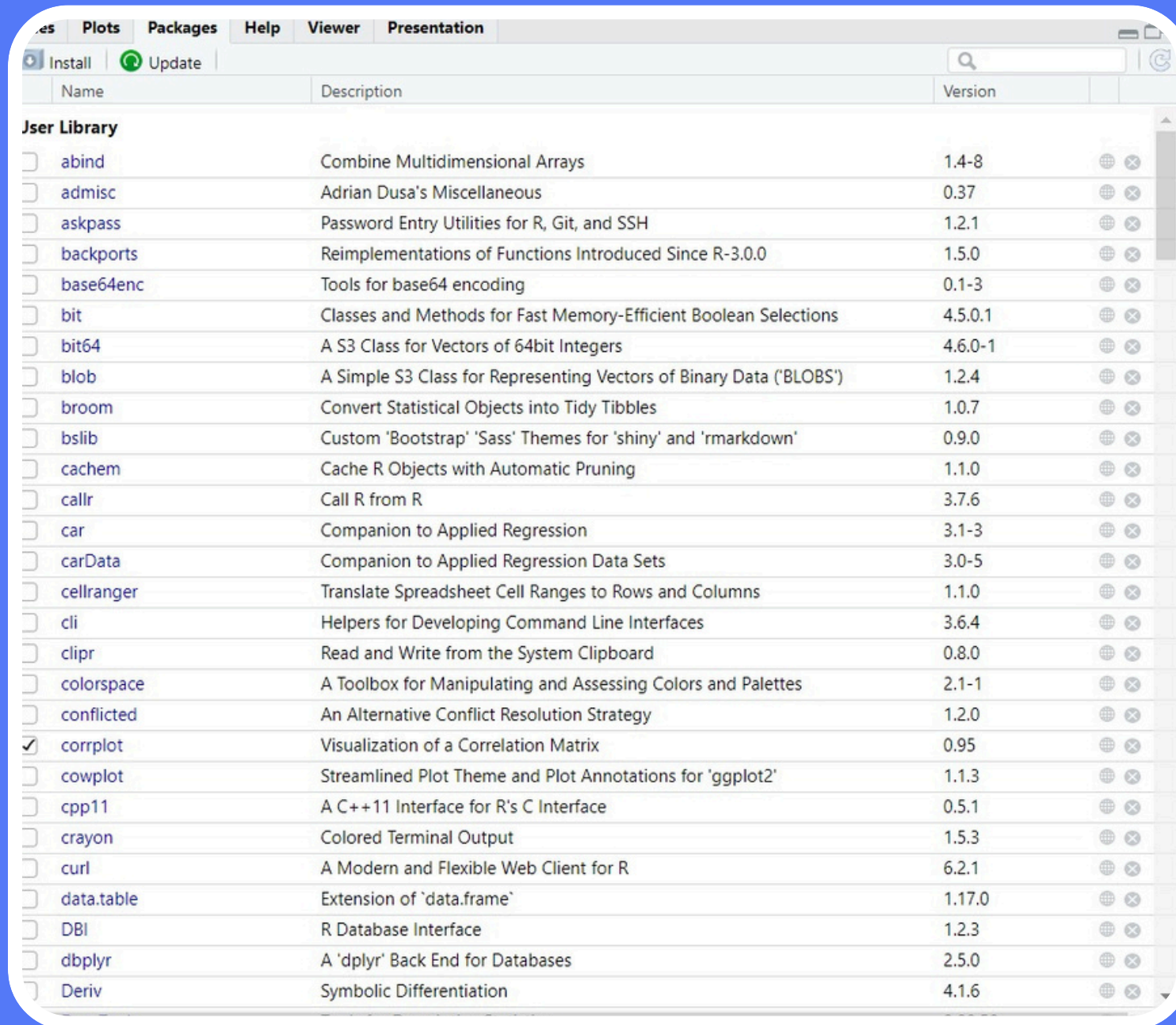




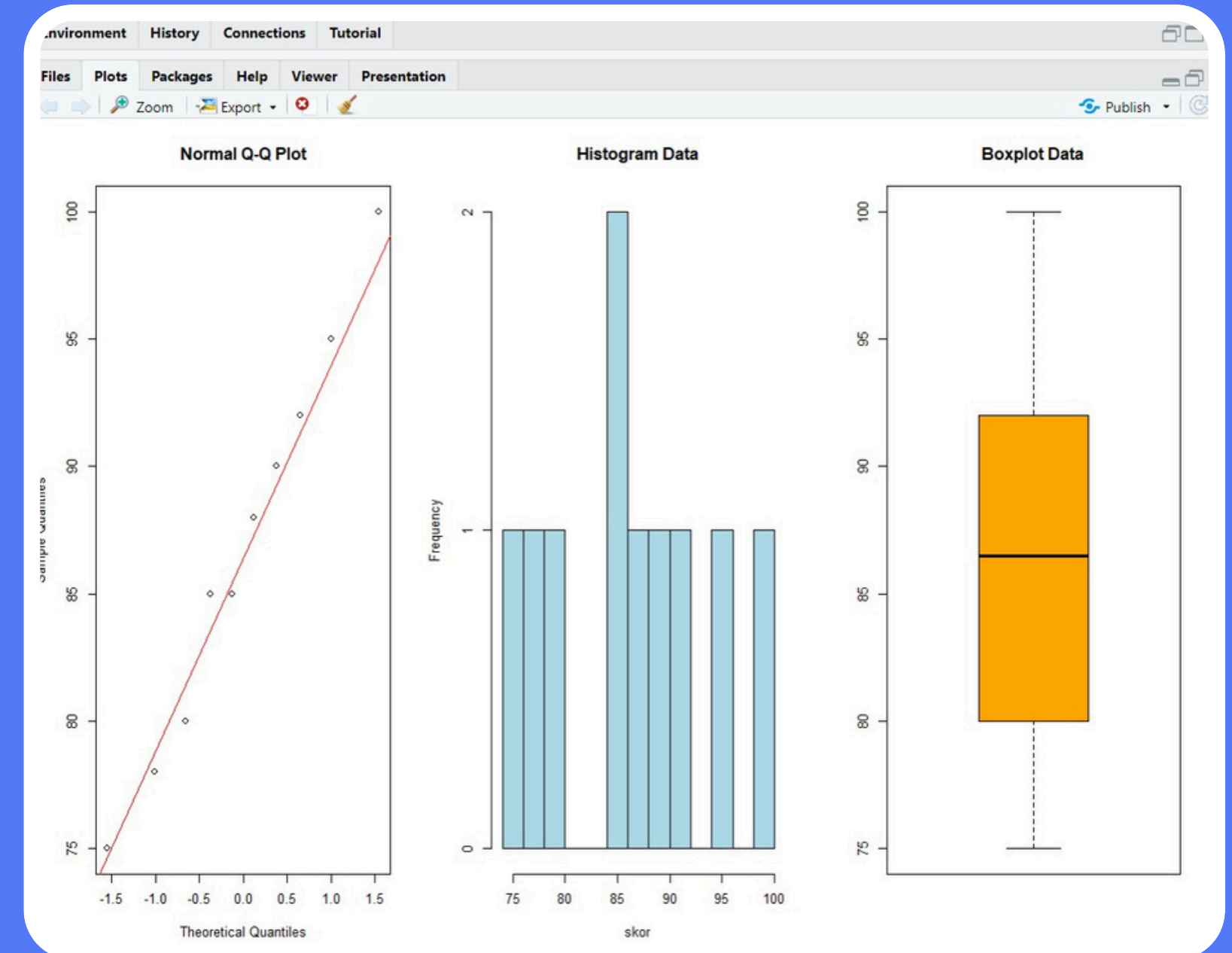
HALAMAN AWAL

The screenshot shows the R Studio interface with the 'Viewer' pane displaying a dataset. The dataset has 26 rows and 14 columns. The columns are: age, sex, cp, trestbps, chol, fbs, restecg, thalachh, exang, oldpeak, slope, ca, thal, and target. The data represents heart disease records. The first row is: 1, 63, 1, 3, 145, 233, 1, 0, 150, 0, 2.3, 0, 0, 1. The last row is: 26, 71, 0, 1, 160, 302, 0, 1, 162, 0, 0.4, 2, 2, 1. The status bar at the bottom indicates 'Showing 1 to 26 of 1,888 entries, 14 total columns'.

HALAMAN DATA



HALAMAN INSTALL PACKAGE



HALAMAN STATISIK

# Uji Kasus Data

## 01 uji normalitas

Menentukan apakah data yang digunakan dalam penelitian berdistribusi normal atau tidak.

## 02 uji validitas

Memastikan bahwa instrumen atau kuesioner benar-benar mengukur apa yang seharusnya diukur.

## 03 uji realibilitas

Menilai konsistensi suatu instrumen atau kuesioner saat diuji ulang.

Kami melakukan pengujian dataset untuk menganalisis hubungan antara 4 variabel yang ada dengan risiko serangan jantung. Pengujian ini mencakup uji normalitas menggunakan Kolmogorov-Smirnov untuk menentukan distribusi data. Hasil uji normalitas akan menentukan metode analisis yang digunakan, baik parametrik maupun non-parametrik.



The background is a solid blue color. In the center, there are several overlapping, organic, blob-like shapes in a lighter shade of blue. These shapes are irregular and fluid, creating a modern, abstract design.

**“Uji Variabel yang berpengaruh  
terhadap Serangan Jantung”**



# Sebelum melakukan Pengujian:

```
{r}
# Install package yang diperlukan
install.packages("psych")    # Untuk uji validitas
install.packages("ltm")      # Untuk uji reliabilitas
install.packages("nortest")  # Untuk KS test
install.packages("ggplot2")
install.packages("ggpubr")
install.packages("corrplot")
```

## INSTALL PACKAGE

```
{r}
# Load library
library(psych)
library(ltm)
library(nortest)
library(ggplot2)
library(ggpubr)
library(corrplot)
```

## LOAD LIBRARY

```
{r}
# 1. Baca dataset
data <- read.csv("C:/Users/victu/Downloads/cleaned_merged_heart_dataset.csv")
View(data)
```

```
{r}
# 2. Cek nama kolom yang tersedia di dataset
cat("Nama kolom dalam dataset:\n")
print(names(data))
```

Nama kolom dalam dataset:

|     |         |           |         |            |        |          |           |            |
|-----|---------|-----------|---------|------------|--------|----------|-----------|------------|
| [1] | "age"   | "sex"     | "cp"    | "trestbps" | "chol" | "fbs"    | "restecg" | "thalachh" |
| [9] | "exang" | "oldpeak" | "slope" | "ca"       | "thal" | "target" |           |            |

```
{r}
# 3. Pilih hanya kolom numerik kontinu yang sesuai (pastikan kolom ini ada)
selected_cols <- c("age", "trestbps", "chol", "oldpeak")

# Pastikan semua kolom yang dipilih ada dalam dataset
selected_cols <- selected_cols[selected_cols %in% names(data)]

# Jika tidak ada kolom yang cocok, hentikan program
if (length(selected_cols) == 0) {
  stop("Tidak ada kolom numerik yang cocok ditemukan dalam dataset. Periksa kembali nama kolom!")
}
```

## LOAD DATA



# Uji Normalitas Data



## Uji Kolmogorov-Smirnov

Uji normalitas dilakukan untuk menentukan apakah data dari masing-masing variabel mengikuti distribusi normal atau tidak.

## Metode statistik yang digunakan

- **Jika data berdistribusi normal** → Gunakan uji parametrik
- **Jika data tidak berdistribusi normal** → Gunakan uji non-parametrik

## Hasil yang didapatkan

Hasil uji normalitas menunjukkan bahwa semua variabel (age, trestbps, chol, dan oldpeak) **tidak berdistribusi normal**.

Hal ini terlihat dari **nilai p-value yang sangat kecil (di bawah 0.05)**, sehingga hipotesis nol ( $H_0$ ) yang menyatakan bahwa data berdistribusi normal ditolak.

# Uji Normalitas Data

```
{r}
# 4. Ambil data hanya dari kolom yang tersedia
numeric_data <- data[selected_cols]

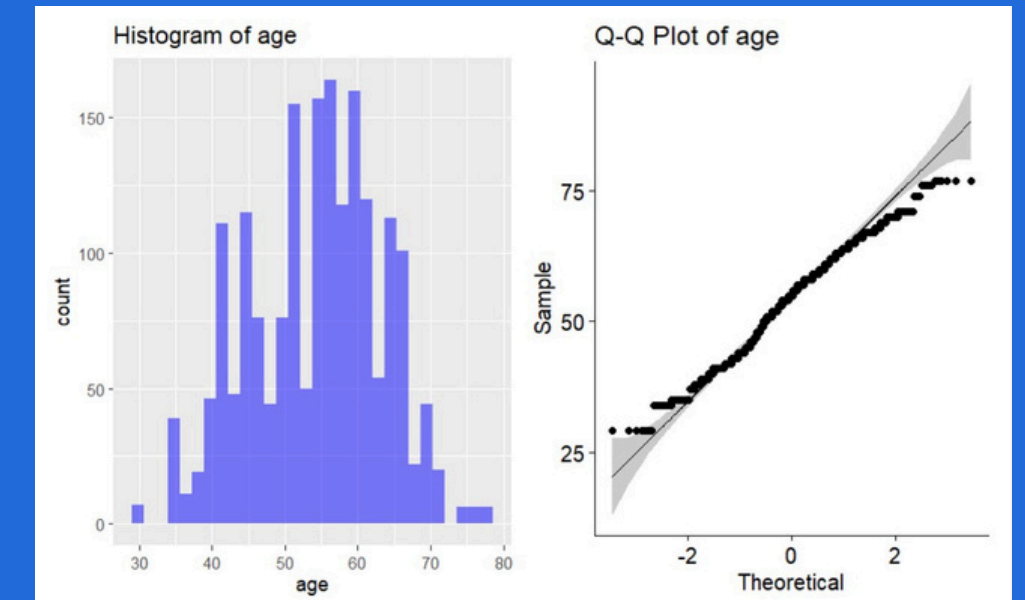
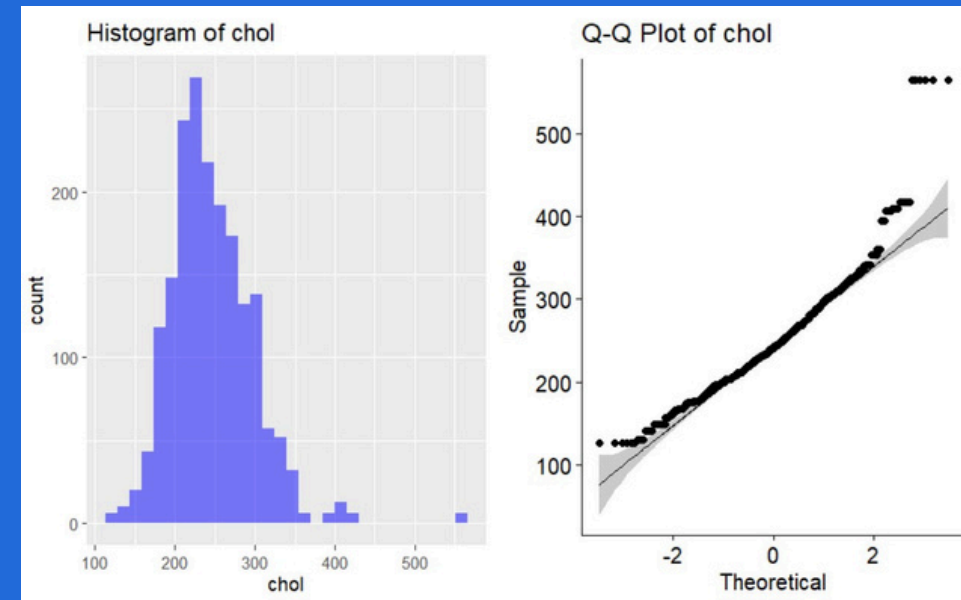
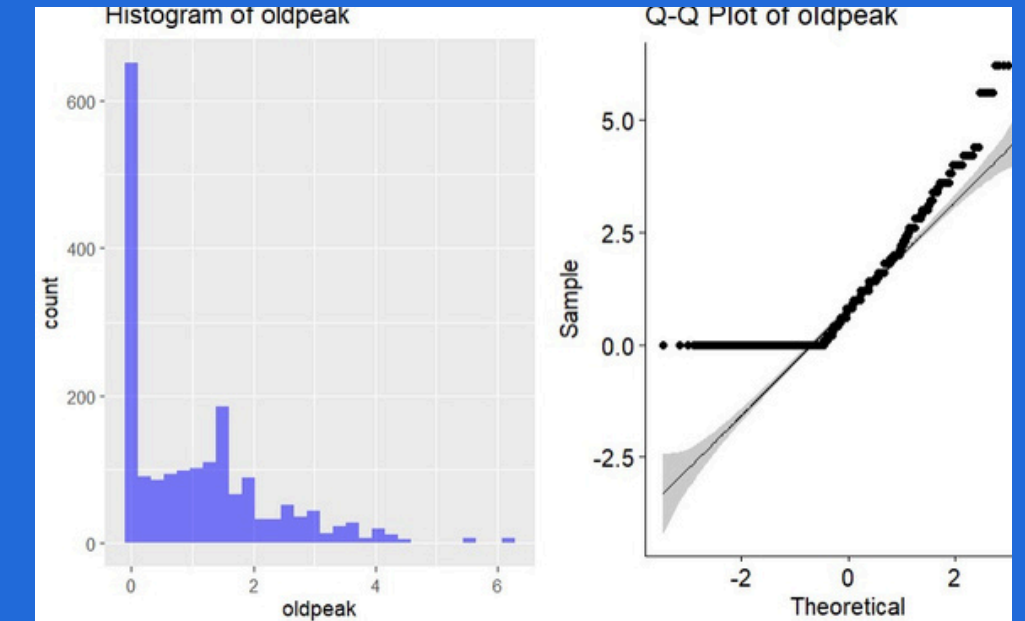
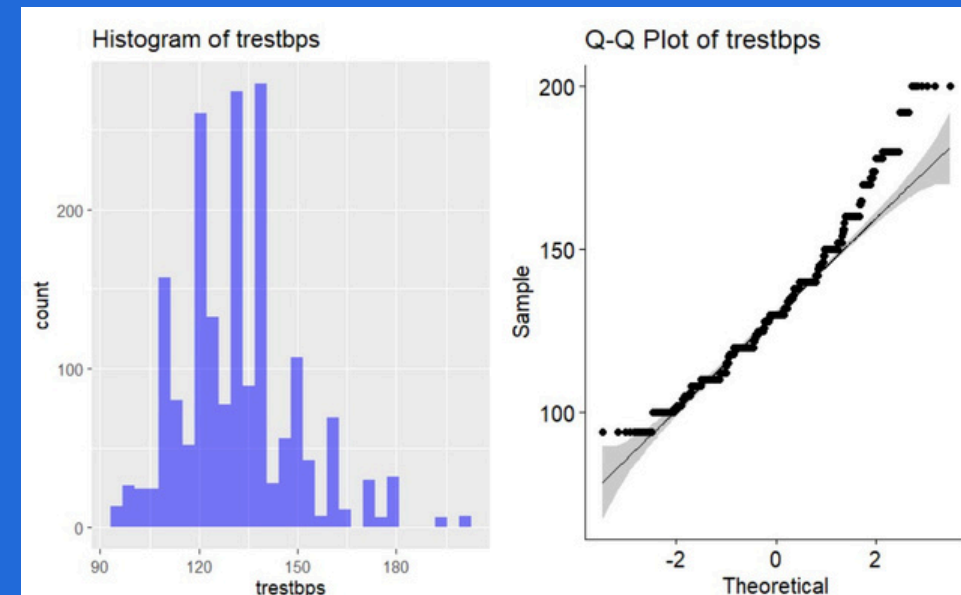
{r}
# 5. Uji Normalitas (Kolmogorov-Smirnov & Shapiro-Wilk)
cat("\n--- Uji Normalitas ---\n")
for (col in selected_cols) {
  cat("\nKolom:", col)

  # Kolmogorov-Smirnov Test
  ks_test <- ks.test(numeric_data[[col]], "pnorm", mean=mean(numeric_data[[col]], na.rm=TRUE), sd=sd(
    numeric_data[[col]], na.rm=TRUE))
  cat("\n KS Test p-value:", ks_test$p.value, ifelse(ks_test$p.value > 0.05, "✅ Normal", "❌ Tidak Normal")
  )
}

--- Uji Normalitas ---

Kolom: agewarning: ties should not be present for the one-sample Kolmogorov-Smirnov test
KS Test p-value: 4.174638e-09 ❌ Tidak Normal
Kolom: trestbpswarning: ties should not be present for the one-sample Kolmogorov-Smirnov test
KS Test p-value: 9.750427e-18 ❌ Tidak Normal
Kolom: cholwarning: ties should not be present for the one-sample Kolmogorov-Smirnov test
KS Test p-value: 2.555428e-05 ❌ Tidak Normal
Kolom: oldpeakwarning: ties should not be present for the one-sample Kolmogorov-Smirnov test
KS Test p-value: 8.260403e-55 ❌ Tidak Normal
```

## Uji Kolmogorov-Smirnov



# Uji Validitas Data



## Uji Validitas Spearman

Uji validitas dilakukan untuk mengevaluasi hubungan antara masing-masing variabel (age, trestbps, chol, dan oldpeak)

## Jenis Validitas yang Diuji

Uji validitas dilakukan untuk mengevaluasi hubungan antara masing-masing variabel (age, trestbps, chol, dan oldpeak)

## Hasil yang didapatkan

- Age memiliki korelasi tertinggi dengan semua variabel lain, terutama dengan trestbps (0.28) dan oldpeak (0.26).
- Trestbps dan chol memiliki korelasi rendah (0.14), menunjukkan bahwa tekanan darah dan kolesterol tidak selalu berkaitan erat.
- Oldpeak memiliki korelasi lemah dengan chol (0.05), sehingga kadar kolesterol mungkin bukan faktor utama dalam menentukan nilai oldpeak.
- Secara umum, tidak ada hubungan yang sangat kuat antara variabel-variabel ini, tetapi usia tampaknya menjadi faktor yang paling berpengaruh dalam menentukan tekanan darah, kolesterol, dan oldpeak.



# Uji Validitas Data

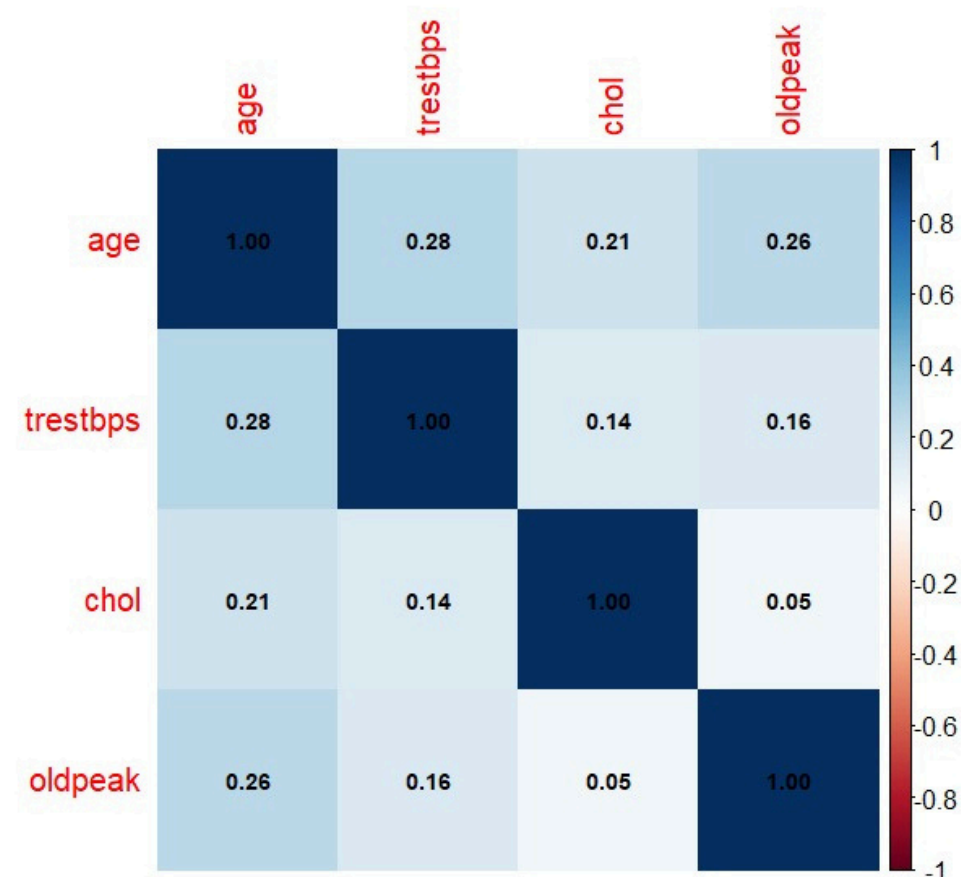
```
{r}
# 6. Pilih hanya kolom numerik untuk Uji Validitas
selected_cols <- c("age", "trestbps", "chol", "oldpeak")

# Ambil data hanya dari kolom yang tersedia
selected_cols <- selected_cols[selected_cols %in% names(data)]
numeric_data <- data[selected_cols]
```

Uji validitas Spearman

```
{r}
# Hitung korelasi Spearman
cor_matrix <- cor(data[selected_cols], method = "spearman")

# Plot heatmap korelasi
corrplot(cor_matrix, method = "color", addCoef.col = "black", number.cex = 0.7)
```



Heat Map Uji Validitas

# Analisis Korelasi antar Variabel

## 01 Korelasi Age (Umur) dengan Variabel Lain

**trestbps (0.28)** → Semakin tua seseorang, semakin tinggi tekanan darah istirahatnya.

**chol (0.21)** → Usia lebih tua cenderung memiliki kadar kolesterol lebih tinggi.

**oldpeak (0.26)** → Semakin tua seseorang, semakin besar kemungkinan mengalami depresi segmen ST saat stres.

## 02 Korelasi Trestbps dengan Variabel Lain

**chol (0.14)** → Orang dengan tekanan darah tinggi cenderung memiliki kadar kolesterol yang lebih tinggi.  
**oldpeak (0.16)** → Semakin tinggi tekanan darah, semakin tinggi oldpeak, yang bisa menunjukkan kondisi jantung kurang baik.

## 03 Korelasi Chol (Kolesterol) dengan Variabel Lain

- **oldpeak (0.05)** → Hubungan sangat lemah, menunjukkan bahwa kadar kolesterol tidak terlalu berkaitan dengan oldpeak dalam dataset ini.

## 04 Korelasi Oldpeak (Depresi ST) dengan Variabel Lain

- Hubungan dengan semua variabel relatif kecil → oldpeak memiliki korelasi tertinggi dengan age (0.26), yang menunjukkan bahwa semakin tua seseorang, semakin besar kemungkinan memiliki nilai oldpeak lebih tinggi

# Uji Realibilitas Data



## Cronbach's Alpha

Hasil uji reliabilitas menunjukkan bahwa Cronbach's Alpha = 0.207, yang jauh di bawah standar reliabilitas  $\geq 0.7$ .

## Hasil yang didapatkan

Nilai yang sangat rendah menunjukkan bahwa data tidak reliabel, artinya konsistensi internal antar-variabel dalam mengukur konstruk yang sama masih sangat lemah.

Salah satu kemungkinan penyebabnya adalah adanya variabel yang kurang berkaitan satu sama lain atau jumlah kategori yang tidak cukup untuk menghasilkan distribusi yang stabil.



# Uji Realibilitas Data

```
{r}
# 9. Hitung Cronbach's Alpha
cat("\n--- Uji Reliabilitas (Cronbach's Alpha) ---\n")
alpha_result <- alpha(reliability_data)
cat("\nCronbach's Alpha:", alpha_result$total$raw_alpha)

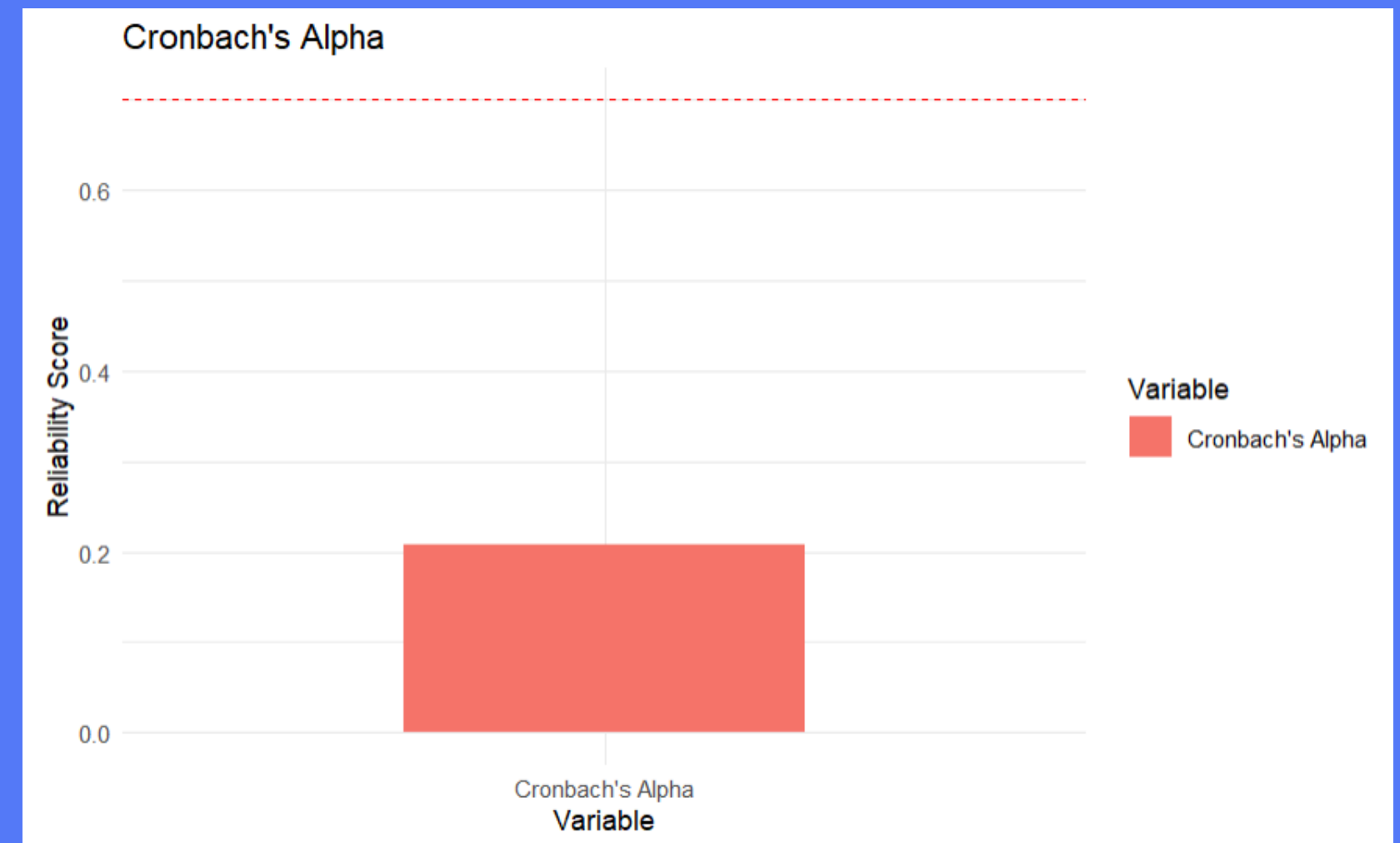
# 10. Interpretasi Hasil
if (alpha_result$total$raw_alpha >= 0.7) {
  cat("\n✅ Reliabel (Cronbach's Alpha ≥ 0.7)")
} else {
  cat("\n❌ Tidak Reliabel (Cronbach's Alpha < 0.7)")
}

--- Uji Reliabilitas (Cronbach's Alpha) ---
Number of categories should be increased in order to count frequencies.

Cronbach's Alpha: 0.2070289
❌ Tidak Reliabel (Cronbach's Alpha < 0.7)
```

```
{r}
# Buat Data Frame
alpha_data <- data.frame(
  Variable = "Cronbach's Alpha",
  Value = 0.207
)

# Plot Bar Chart
ggplot(alpha_data, aes(x = Variable, y = Value, fill = Variable)) +
  geom_bar(stat = "identity", width = 0.5) +
  geom_hline(yintercept = 0.7, linetype="dashed", color="red") +
  ggtitle("Cronbach's Alpha") +
  ylab("Reliability Score") +
  theme_minimal()
```



## Apa Itu Cronbach's Alpha?

Cronbach's Alpha adalah metode untuk mengukur reliabilitas atau konsistensi internal dari sekumpulan item dalam suatu instrumen penelitian (misalnya kuesioner atau skala pengukuran).

- Nilai Cronbach's Alpha berkisar antara 0 hingga 1:  $\geq 0.9$  → Sangat reliabel (Excellent) ✓
- 0.7 – 0.9 → Reliabel (Good) ✓
- 0.6 – 0.7 → Cukup reliabel (Acceptable) ⚠
- $< 0.6$  → Tidak reliabel ✗

## Interpretasi Hasil Plot

- Cronbach's Alpha = 0.207 ✗ → Tidak reliabel (sangat rendah).
- Kesimpulan: Instrumen pengukuran ini tidak cukup konsisten dalam mengukur konsep yang ingin diuji.
- Perlu revisi atau penyesuaian terhadap variabel atau item dalam pengukuran agar lebih koheren.

## Mengapa Nilainya Rendah? dan Solusi yang Bisa Dilakukan

Mengapa Nilainya Rendah?

- Item dalam skala tidak berkorelasi kuat → Artinya, pertanyaan atau indikator yang digunakan mungkin tidak saling mendukung dalam mengukur satu konsep yang sama.
- Data mungkin memiliki noise atau variasi besar → Misalnya, responden memberikan jawaban yang tidak terstruktur.

Solusi yang Bisa Dilakukan

1. Tambahkan lebih banyak item dalam skala agar lebih representatif.
2. Periksa apakah ada pertanyaan yang tidak relevan atau tidak memiliki hubungan yang jelas dengan variabel lainnya.
3. Gunakan metode validasi lain, seperti factor analysis (analisis faktor) untuk mengidentifikasi kelemahan dalam item skala.
4. Tingkatkan jumlah responden agar distribusi data lebih stabil.

# Uji Mann-Whitney



## Uji Mann-Whitney

Uji Mann-Whitney U (atau Wilcoxon Rank-Sum Test) adalah uji statistik non-parametrik yang digunakan untuk, Membandingkan median dua kelompok independen (tidak berpasangan).

## Jenis Validitas yang Diuji

Uji Mann-Whitney menguji Validitas Statistik (Kesimpulan), Validitas Konstruksi (Konsep), Validitas Eksternal (Generalisasi)

## Hasil yang didapatkan

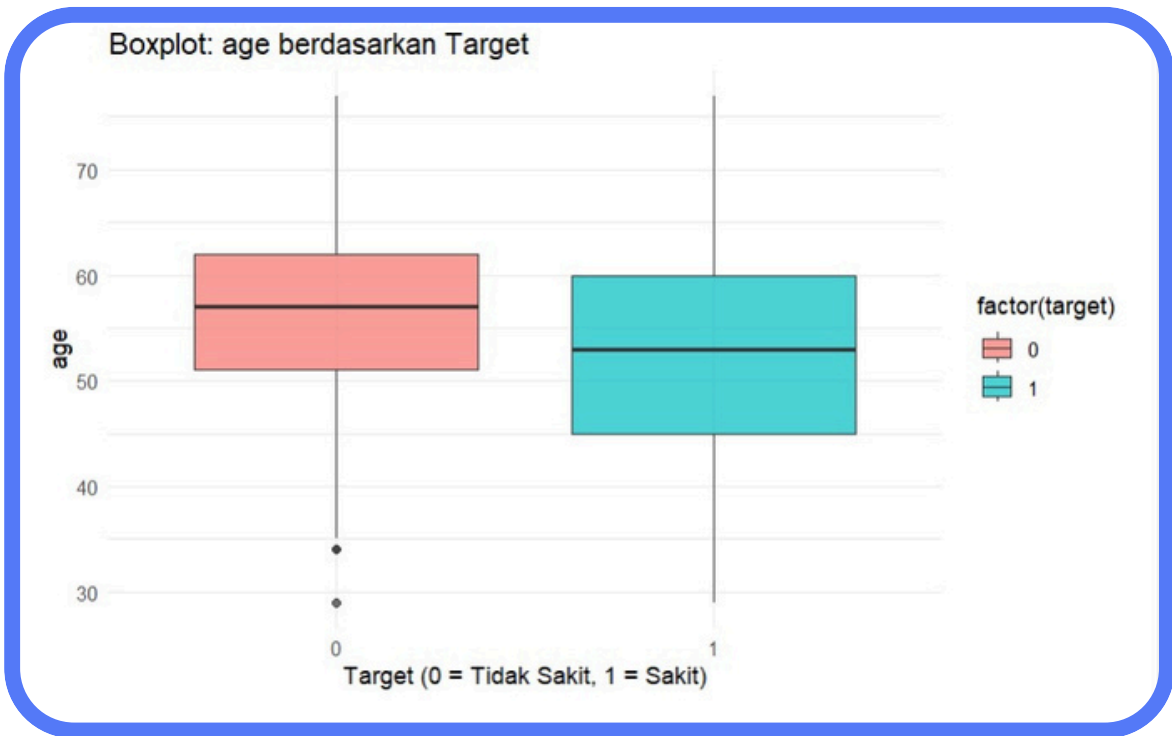
Dari analisis ini terlihat bahwa:

- Usia merupakan faktor pembeda yang signifikan
- Kolesterol tidak menunjukkan perbedaan berarti antara kelompok
- Hasil oldpeak yang kontraindikasi memerlukan investigasi lebih lanjut
- Beberapa variabel seperti tekanan darah tidak menunjukkan perbedaan median yang jelas

Analisis ini membantu memahami karakteristik pasien dan faktor-faktor yang mungkin berhubungan dengan kondisi sakit jantung dalam dataset ini.



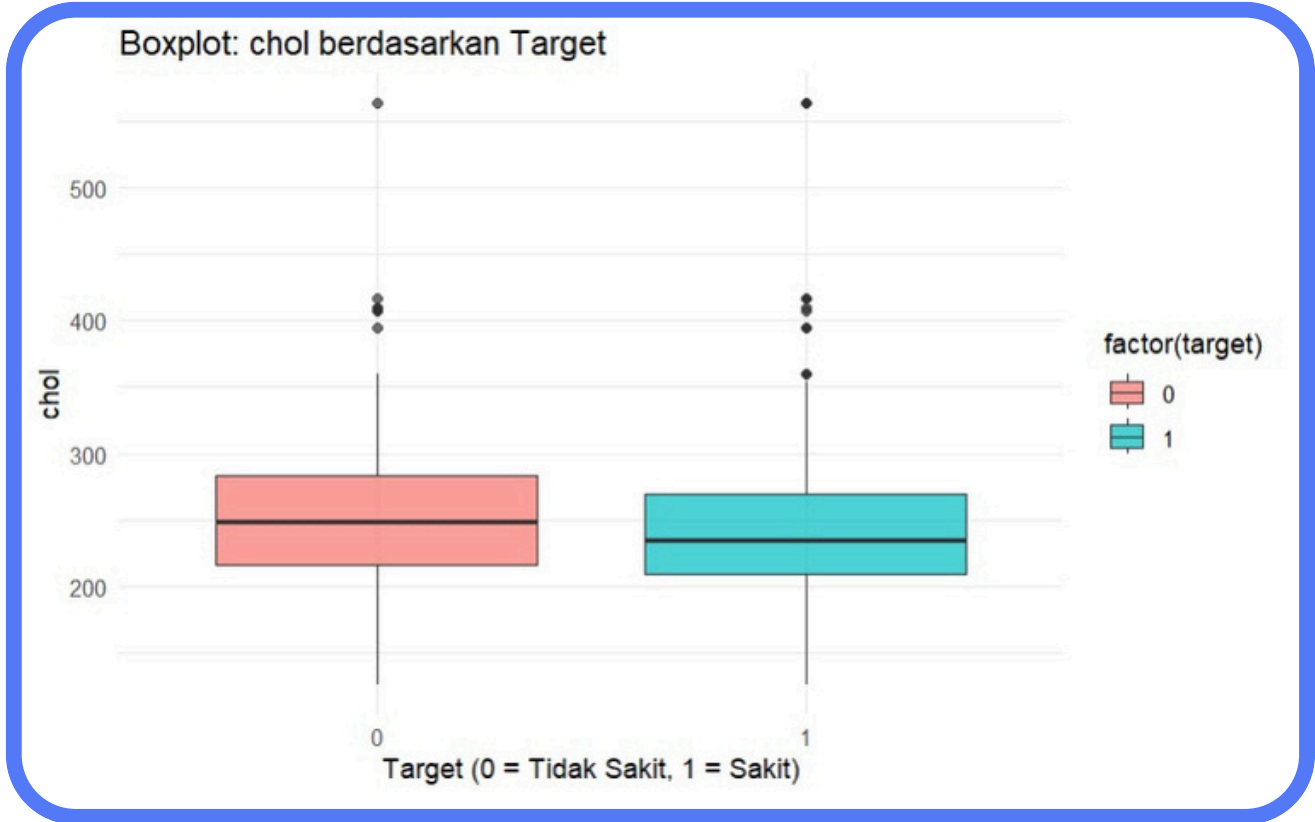
# Uji Mann-Whitney



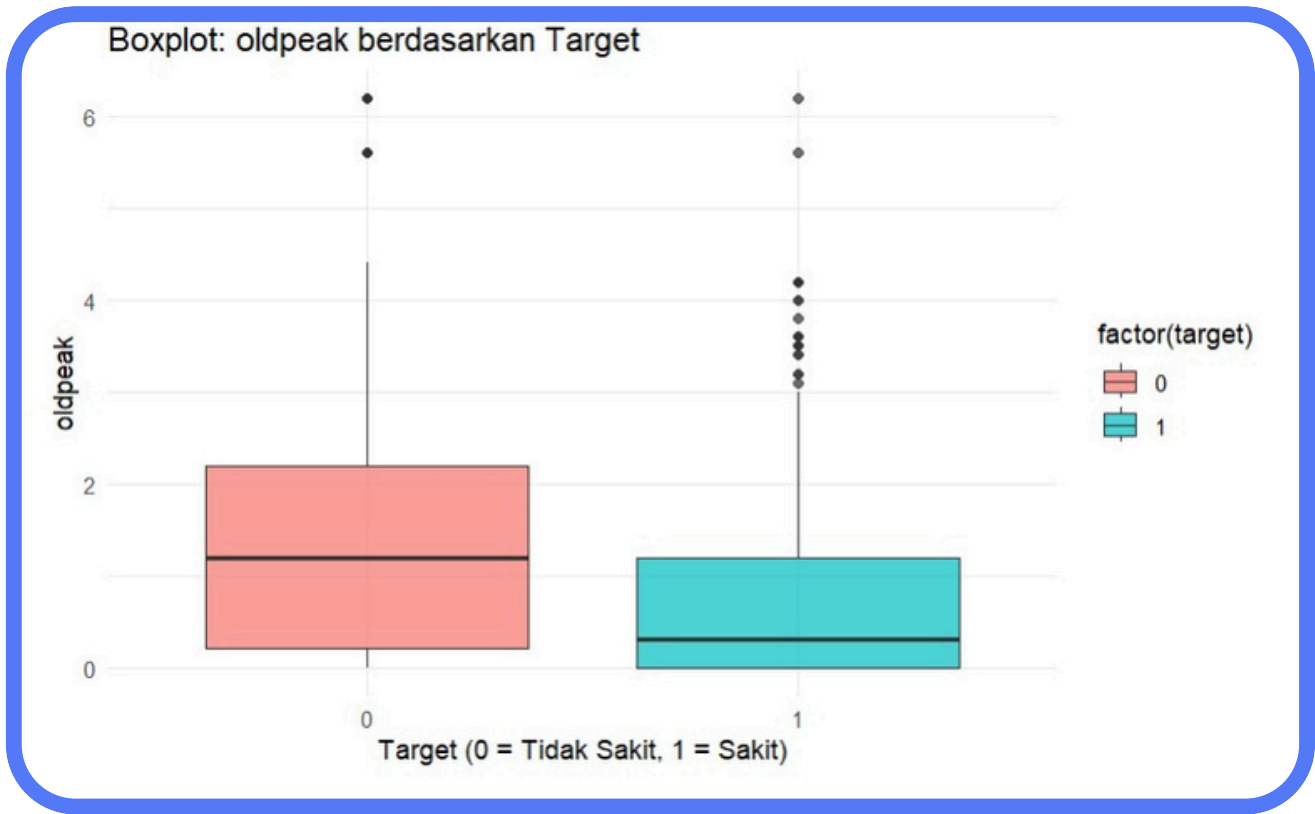
Age Target



Trestbps target



chol target



oldpeak target

# Analisis Boxplot dan Uji Mann-Whitney

## 01 Usia (Age)

- Boxplot menunjukkan:
  - Kelompok tidak sakit (0) memiliki median usia lebih tinggi
  - Kelompok sakit (1) cenderung lebih muda
- Uji Mann-Whitney:
  - p-value = 0.023 ( $< 0.05$ )
  - Kesimpulan: Ada perbedaan signifikan usia antara yang sakit dan tidak sakit

## 02 Kolesterol (chol)

- Boxplot menunjukkan:
  - Median kolesterol sangat mirip (245-250) antara kedua kelompok
  - Sebaran data juga mirip
  - Ada beberapa outlier tinggi (hingga 560)
- Interpretasi:  
Tidak ada perbedaan signifikan kadar kolesterol total antara yang sakit dan tidak sakit

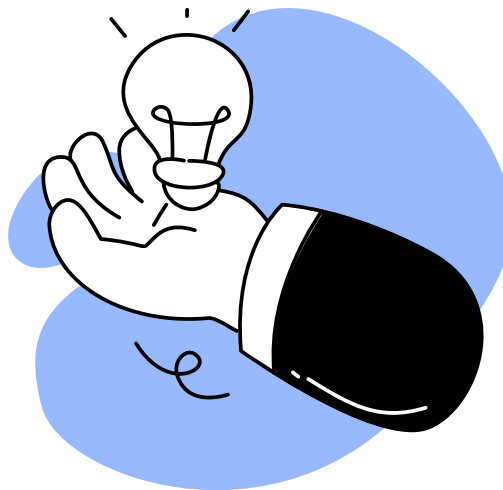
## 03 Oldpeak (depresi ST pada EKG)

- Boxplot menunjukkan:
  - Kelompok tidak sakit (0) memiliki median lebih tinggi (1.4-1.5)
  - Kelompok sakit (1) memiliki median lebih rendah (~0.5)
  - Banyak outlier pada kelompok tidak sakit
- Interpretasi:  
Nilai oldpeak lebih tinggi justru ditemukan pada yang tidak sakit

## 04 Tekanan Darah Istirahat (trestbps)

- Boxplot menunjukkan:
  - Median tekanan darah mirip antara kedua kelompok
  - Banyak outlier di kedua sisi

# Kesimpulan Analisis Studi



## Uji Normalitas

- Semua variabel yang diuji (age, trestbps, chol, oldpeak) tidak berdistribusi normal berdasarkan uji Kolmogorov-Smirnov ( $p\text{-value} < 0.05$ ).
- Hal ini menunjukkan bahwa data tidak mengikuti distribusi normal, sehingga analisis statistik yang digunakan sebaiknya metode non-parametrik.



## Uji Validitas

Age, trestbps, chol, dan oldpeak memiliki korelasi lemah satu sama lain, namun usia cenderung berkaitan dengan tekanan darah, kolesterol, dan depresi ST.



## Uji Realibilitas

- Nilai Cronbach's Alpha sebesar 0.207 menunjukkan bahwa instrumen pengukuran ini tidak reliabel.
- Artinya, indikator yang digunakan dalam pengukuran tidak memiliki konsistensi internal yang cukup baik dan perlu diperbaiki, misalnya dengan menambah jumlah item atau menyusun ulang pertanyaan.



# Kesimpulan Analisis Studi

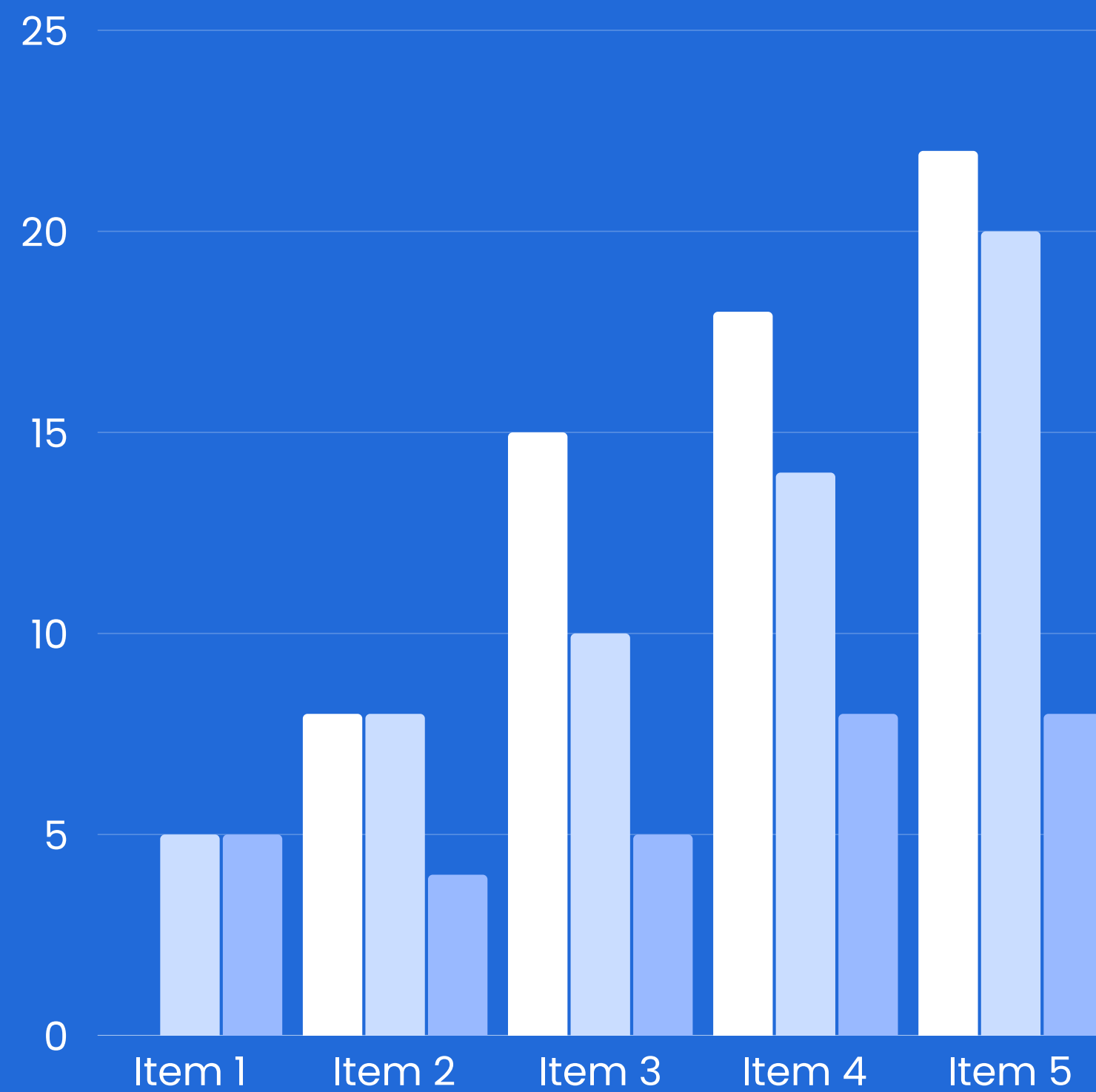


Let mi ting..let mi ting...

## Uji Mann-Whitney

- Usia merupakan faktor pembeda yang signifikan
- Kolesterol tidak menunjukkan perbedaan berarti antara kelompok
- Hasil oldpeak yang kontraindikasi memerlukan investigasi lebih lanjut
- Beberapa variabel seperti tekanan darah tidak menunjukkan perbedaan median yang jelas

Analisis ini membantu memahami karakteristik pasien dan faktor-faktor yang mungkin berhubungan dengan kondisi sakit jantung dalam dataset ini.



THANKS:)

Any Question about R???

Kelompok 8