🧪 Practical Class in R: Measuring Correlation

📚 Topic: Correlation Analysis using R

🎯 Learning Outcomes:

By the end of this session, students will be able to:

- Import and inspect data in R
- Create scatter plots for pairs of numeric variables
- Compute Pearson and Spearman correlation coefficients
- Interpret the results and distinguish between linear and monotonic relationships

---

📁 Dataset:

Download the dataset:
📥 X correlation_dataset_200_rows.xlsx

🛠️ Practical Steps

🧩 1. Load Required Packages

```
# Install only if not already installed
install.packages("readxl")
install.packages("ggplot2")
install.packages("GGally")

# Load libraries
library(readxl)
library(ggplot2)
library(GGally)
```

📥 2. Import Dataset

```
# Read Excel file
data <- read_excel("path/to/your/correlation_dataset_200_rows.xlsx")

# View structure
str(data)

# Optional: check column names
colnames(data)
```

📈 3. Create Scatter Plots

Create scatter plots for selected pairs:

```
# Example: Hours_Studied vs Test_Score
ggplot(data, aes(x = Hours_Studied, y = Test_Score)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "Scatterplot: Hours Studied vs Test Score")
```

✅ Do the same for:

- `Hours_Studied` vs `Stress_Level`
- `Hours_Studied` vs `Coffee_Cups`
- `Hours_Studied` vs `Random_Noise`
- `Test_Score` vs `Commuter_time`

```
# Create a pair plot for all variables
GGally::ggpairs(data)
```

🧮 4. Calculate Pearson and Spearman Correlation

```
# Pearson correlation
cor(data$Hours_Studied, data$Test_Score, method = "pearson")

# Spearman correlation
cor(data$Hours_Studied, data$Test_Score, method = "spearman")
```

✅ Repeat for other pairs:

🧠 Discussion Questions

- Which variables show a strong linear relationship?
- Which pair is better captured by Spearman's correlation than Pearson's?
- Are there any variable pairs that show no apparent relationship?