

# Data Analysis Workflow

---

Kamarul Imran Musa

Assoc Prof in Epidemiology and Statistics

Public Health Physician

Fellow of the American College of  
Epidemiology

[drkamarul@usm.my](mailto:drkamarul@usm.my)



# The analysis workflow for

Read

Explore

Assume

Analyze

Assess

Present

# Read

- Structured vs Unstructured
- Tabular vs Non-tabular
- Examples:
  - Spreadsheet
  - Statistical data
  - Database
  - Images
  - Sound



# Explore data

## Graphical

---

One variable

---

Two variables

---

More than two variables

---

Package

---

Function

## Tabular

- Overall
- Stratified
- Package
- Function



# Assumption

- Population being studied
- Sample size
- Data collection
- Variables collected
- Distribution of variable of interest
- Confounders, mediators
- Colliders



# Analyze

## **Confirmatory or exploratory**

- Confirmatory
- Exploratory

## **Comparison or relationship**

- Comparison
- Relationship

# Assess

---

- Check assumption
- If assumption violated?



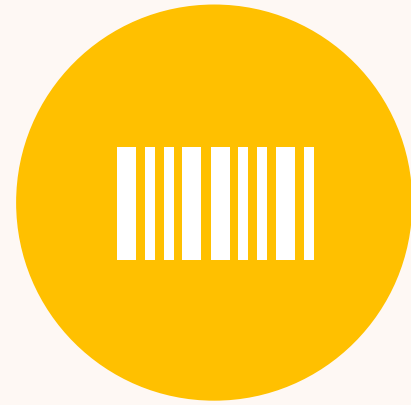
# Present



RESULTS



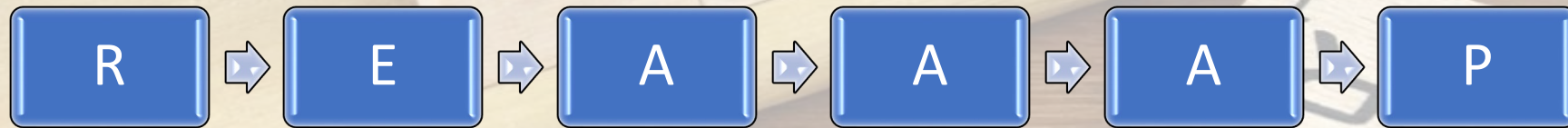
DATA



CODES



# Steps in the workflow



- Think
  - Objectives
  - Packages
  - Functions