

Modern Data Science Methods for Educational Research

R for Data Analysis in Educational Research

อ.ดร.ประภาศิริ รัชประภาพรกุล

ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

February 11, 2023

สารบัญ I

1. ประเภทของไฟล์ข้อมูล

2. Flat Files

3. MS Excels Files

4. SPSS and other Statistical program Files

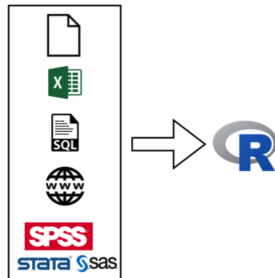


1. ประเภทของไฟล์ข้อมูล

ประเภทของไฟล์ข้อมูล

R สามารถดำเนินการร่วมกับไฟล์ข้อมูลที่หลากหลาย

- ▶ Flat Files
- ▶ MS Excel Files
- ▶ Statistical softwares
- ▶ Databases
- ▶ Internet



2. Flat Files

Flat Files

- ▶ เป็นไฟล์ข้อมูลที่จัดเก็บโดยไม่มีโครงสร้างตาราง และบันทึกอยู่ในรูปแบบไฟล์ข้อความ (text files)
- ▶ Flat file จำแนกได้หลายประเภท ขึ้นอยู่กับลักษณะของตัวคั่นข้อมูล (separator) เช่น CSV, TSV

| | | |
|---|---------------------------|--------------------------------|
| 1 | gender,MathScore,EngScore | << ชื่อตัวแปร (variable names) |
| 2 | M,33,80 | } ข้อมูล (data) |
| 3 | M,57,20 | |
| 4 | F,76,80 | |
| 5 | M,80,85 | |
| 6 | F,92,95 | |

| | | |
|---|---------------------------|--------------------------------|
| 1 | gender MathScore EngScore | << ชื่อตัวแปร (variable names) |
| 2 | Male 33 80 | } ข้อมูล (data) |
| 3 | Male 57 20 | |
| 4 | Female 76 80 | |
| 5 | Male 80 85 | |
| 6 | Female 92 95 | |

Figure 1: ที่มา : สิวะโชติ ศรีสุทธธียากร (2564)

Importing CSV Files

การนำข้อมูล .csv เข้าสู่โปรแกรมสามารถใช้ฟังก์ชัน `read.csv()` โดยมีรูปแบบของคำสั่งดังนี้

```
1 read.csv(file = "file.csv",  
2         header = TRUE,  
3         stringsAsFactors = TRUE)
```

อาร์กิวเมนต์สำคัญของ `read.csv()` ได้แก่

- ▶ `file`
- ▶ `header`
- ▶ `stringsFactors`

กิจกรรม : นำเข้า CSV file

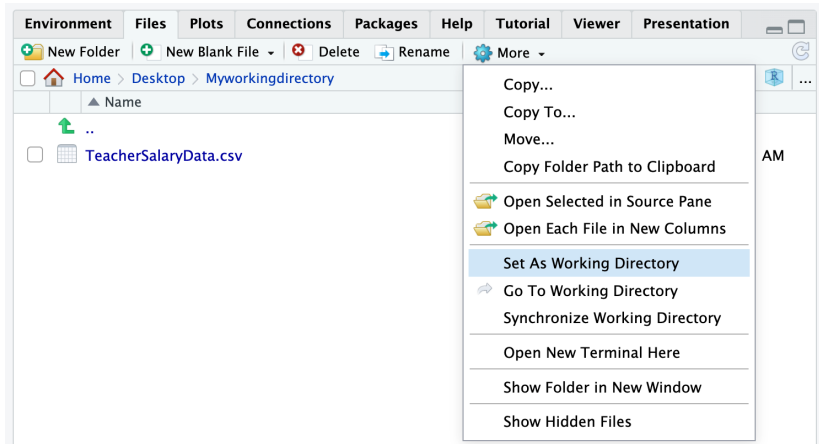
1. ดาวน์โหลด TeacherSalaryData.csv ← คลิกเลย
2. นำเข้าไฟล์ข้อมูล TeacherSalaryData.csv โดยพิมพ์คำสั่งต่อไปนี้

```
1 mydata <- read.csv(file = "TeacherSalaryData.csv",  
2                       header = TRUE,  
3                       stringAsFactors = TRUE)
```

3. สังเกตผลลัพธ์ที่ได้ ท่านสามารถนำไฟล์ข้อมูลดังกล่าวเข้าสู่ R ได้หรือไม่
ถ้าไม่ท่านคิดว่าเกิดปัญหาอะไร

กิจกรรม : การกำหนด Working Directory

การจะนำไฟล์ข้อมูลเข้าใน R ไฟล์ข้อมูลดังกล่าวจะต้องบันทึกอยู่ใน working directory ของ R



กิจกรรม : นำเข้า CSV file (revisited)

1. ดาวน์โหลด TeacherSalaryData.csv ← คลิกเลย
2. กำหนด working directory ของ R ให้เป็น folder เดียวกันกับที่บันทึกไฟล์ข้อมูลในข้อ 1.
3. นำเข้าไฟล์ข้อมูล TeacherSalaryData.csv โดยพิมพ์คำสั่งต่อไปนี้

```
1 mydata <- read.csv(file = "TeacherSalaryData.csv",  
2                       header = TRUE,  
3                       stringAsFactors = TRUE)
```

4. ลองพิมพ์คำสั่ง head(mydata) ผลลัพธ์ที่ได้เป็นอย่างไร

3. MS Excels Files

Importing MS Excel files

การนำเข้าไฟล์ข้อมูล .xlsx ของ MS Excel จะต้องติดตั้ง package-readxl เพิ่มเติม การดาวน์โหลดและเรียกใช้ package ดังกล่าวสามารถดำเนินการได้ดังนี้

```
1 install.packages("readxl")  
2 library(readxl)
```

การนำเข้าไฟล์ .xlsx สามารถทำได้โดยใช้ฟังก์ชัน read_excel() ที่มีรูปแบบของคำสั่งดังนี้

```
1 read_excel(file = "file.xlsx",  
2           sheet = 1,  
3           col_names = TRUE,  
4           skip = 0)
```

กิจกรรม : การนำเข้า .xlsx files

1. ดาวน์โหลด Score.xlsx
2. พิมพ์คำสั่ง `excel_sheets("Score.xlsx")` แล้วสังเกตผลลัพธ์ที่ได้
3. นำเข้าไฟล์ข้อมูลจาก sheet ที่ชื่อว่า Score3 ด้วยฟังก์ชัน `read_excel()`
4. ผลลัพธ์ที่ได้ถูกต้องหรือไม่

```
# A tibble: 20 x 5
```

| | ...1 | ...2 | ...3 | ...4 | ...5 |
|---|-------|-------|-------|-------|-------|
| | <chr> | <dbl> | <dbl> | <dbl> | <dbl> |
| 1 | M | 2.7 | 9 | 4 | 17.6 |
| 2 | F | 2.9 | 22 | 3.2 | 23 |
| 3 | M | 3.3 | 12 | 4 | 13.7 |
| 4 | M | 4.2 | 16 | 4.8 | 18 |
| 5 | M | 4.4 | 23 | 4.5 | 14.9 |
| 6 | F | 4.7 | 20 | 5.2 | 25.8 |
| 7 | F | 5.4 | 17 | 2.7 | 13.8 |

กิจกรรม : การนำเข้า .xlsx files (revisited)

จากกิจกรรมที่แล้ว ลองปรับอาร์กิวเมนต์ col_names ให้เป็น

```
1 col_names = c("gender", "interaction", "pretest",  
2             "quiz", "posttest")
```

สังเกตผลลัพธ์ที่ได้

A tibble: 20 x 5

| | gender | interaction | pretest | quiz | posttest |
|---|--------|-------------|---------|-------|----------|
| | <chr> | <dbl> | <dbl> | <dbl> | <dbl> |
| 1 | M | 2.7 | 9 | 4 | 17.6 |
| 2 | F | 2.9 | 22 | 3.2 | 23 |
| 3 | M | 3.3 | 12 | 4 | 13.7 |
| 4 | M | 4.2 | 16 | 4.8 | 18 |
| 5 | M | 4.4 | 23 | 4.5 | 14.9 |
| 6 | F | 4.7 | 20 | 5.2 | 25.8 |

4. SPSS and other Statistical program Files

Haven package



- ▶ SAS: `read_sas()`
- ▶ SPSS: `read_sav()`
- ▶ Stata: `read_dta()`

ในการทำงานเดียวกัน ก่อนใช้ฟังก์ชันใน package-haven จะต้องดาวน์โหลดและเรียกใช้ package ดังนี้

```
1 install.packages("haven")  
2 library(haven)
```

Importing SPSS Files

```
1 # importing AchSolve.sav
2 dat_spss <- read_spss(file = "AchSolve.sav")
```

```
1 head(dat_spss)
```

```
# A tibble: 6 x 3
  ach solve group
  <dbl> <dbl> <dbl+lbl>
1  49.6    1.2 1 [Lecture]
2  43.6    6.3 1 [Lecture]
3  50.3    1.8 1 [Lecture]
4  34.4    5.7 1 [Lecture]
5  56.8    5.9 1 [Lecture]
6  32.8     8  1 [Lecture]
```

```
1 tail(dat_spss)
```

```
# A tibble: 6 x 3
  ach solve group
  <dbl> <dbl> <dbl+lbl>
1  82.7   13.8 2 [PBL]
2  67.4   13.1 2 [PBL]
3  59.5   13.8 2 [PBL]
4  36     11.6 2 [PBL]
5  84.4   12.8 2 [PBL]
6  55.5   11.8 2 [PBL]
```