Practice 01: Introduction to R for Statistics

**Objectives:** In this lab, you will practice your skills in

1. Import data into R
2. Understand the basics of working with data frames.
3. Learn basic R commands for data manipulation and exploration.
4. Perform summary statistics
5. Create basic statistical graphs.

# Task 1: Setting Up Your Environment

1. Open RStudio.
2. Create a new R script (File > New File > R Script).
3. Use function **setwd()** to setup working directory. Show your R code for this calculation. (0.5 pts)
4. Click  to save your R document

# Task 2: Importing Data

Import **economic\_indicators.csv** and **free\_wifi\_locations.xls** file using function in R. show your R code for this calculation.

1. Use function **read.csv()** toimport **economic\_indicators.csv** file and assign it to an object named **economics**.
2. Use function **read\_excel()** from library **readxl** to import **free\_wifi\_locations.xls** file and assign it to an object named **wifi**.
3. Make a screenshot of **GLOBAL ENVIRONMENT** to display all 2 data-frames.

economics = read.csv('economic\_indicators.csv')

wifi = readxl::read\_excel('free\_wifi\_locations.xls')

A screenshot of a computer

Description automatically generated

# Task 3: Data-Frame Basics

Economic indicators data include values related to topics such employment, housing and real estate development, covering the period from Jan 2013 and Dec 2019. Show your R code for this calculation.

1. Access **unemp\_rate** and **labor\_force\_part\_rate** columns.
2. Use **labor\_force\_part\_rate** to minus **unemp\_rate** to calculate the difference between these two values and add the new variable **diff\_unemp\_labor** to the **economics** data-frame. (0.5 pts)
3. Apply the statement

**economics[order(economics$diff\_unemp\_labor, decreasing = TRUE),c('Year','Month')]**

What is this statement doing?

1. Use **summary()** to see the summary information of the **wifi** data-frame.
2. Describe the summary information for **OID\_** and **neightborhood\_id**, and explain why they are different?

# Task 4: Plot basics

Boxplot analysis

1. Make a boxplot based on column **logan\_intl\_flights** in economics data-frame. (hint: using boxplot() and input variable is **logan\_intl\_flights** from **economics** data-frame)
2. Apply below statement:

**boxplot(logan\_intl\_flights ~ Month, data = economics)**

what insights can we gather about seasonal trends in international flights from grouped boxplot?

Regression line analysis

1. Apply below statements:

**plot(logan\_intl\_flights~Time, data = economics, type = 'l')**

**abline(lm(logan\_intl\_flights~Time, data=economics))**

how does the trend of **logan\_intl\_flights** change over **Time** based on the first plotted line?

What does the regression line added to the plot tell us about the relationship between **logan\_intl\_flights** and **Time**?