

# Lab 01: Introduction to R for Statistics

**Due date:** Wednesday, January 22, 2025 submitted as Word document to Canvas **Lab01** link


This lab counts 8 % toward your total grade.

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

- a) Import data into R
- b) Understand the basics of working with data frames.
- c) Learn basic R commands for data manipulation and exploration.
- d) Perform summary statistics
- e) Create basic statistical graphs.

**Format of answer:** Submit your answers as a **Word document** with graphs and verbal descriptions, properly labeled in the task sequence, with answers in **red text** and only relevant content included

## Task 1: Setting Up Your Environment (0.5 pts)

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

## Task 2: Importing Data (1.5 pts)

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

- a) Use function **read.csv()** to import **economic\_indicators.csv** file and assign it to an object named **economics**. (0.5 pts)
- b) Use function **read\_excel()** from library **readxl** to import **free\_wifi\_locations.xls** file and assign it to an object named **wifi**. (0.5 pts)
- c) Make a screenshot of **GLOBAL ENVIRONMENT** to display all 2 data-frames. (0.5 pts)

## Task 3: Data-Frame Basics (3 pts)

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.

- d) Access **unemp\_rate** and **labor\_force\_part\_rate** columns. (1 pts)
- a) 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)
- b) Apply the statement

```
economics[order(economics$diff_unemp_labor, decreasing =  
TRUE),c('Year','Month')]
```

What is this statement doing? (0.5 pts)

- c) Use **summary()** to see the summary information of the **wifi** data-frame. (0.5 pts)
- d) Describe the summary information for **OID\_** and **neighborhood\_id**, and explain why they are different? (0.5 pts)

## Task 4: Plot basics (3 pts)

Boxplot analysis

- a) 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) (0.5 pts)
- b) Apply below statement:

```
boxplot(logan_intl_flights ~ Month, data = economics)
```

what insights can we gather about seasonal trends in international flights from grouped boxplot? (0.5 pts)

Regression line analysis

- c) 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? (1 pts)

What does the regression line added to the plot tell us about the relationship between **logan\_intl\_flights** and **Time**? (1 pts)