



Rensselaer

why not change the world?®

Lab 1: Beginning to work with data: Data Frames, examining variables, plotting distributions

Ahmed Eleish

ITWS-4600/ITWS-6600/CSCI-4600 MGMT4600/6600/ BCBP4600

Lab 1, September 11th, 2026

Tetherless World Constellation
Rensselaer Polytechnic Institute



Lab tasks

- Reading data in R
- Exploring variables
- Comparing distributions
- Statistical tests



Matlab/R/scipy-numpy Reference

<http://hyperpolyglot.org/numerical-analysis>

Exercises – importing data

- Rstudio

- read in csv file (two ways to do this) - filename.csv
- Read in excel file (directly or by csv convert) - filename.xlsx
- Plot some variables

- Also for other datasets, enter these in the R command window panel or cmd line

```
> data()
```

```
> help(data)
```

Files

Here -> <https://rpi.box.com/s/zpifgwdbu5c3lwooqthfqtq5v9w5q5v9>

- R_stats_functions.R
- epi_results_2024_pop_gdp.csv
- epi2024variables2024-12-11.csv

Exercise 1: Starting out

- Run the code in the provided script
- Make sure all commands run successfully
- Get familiar with reading files, printing variables/summaries, plotting figures
- Examine plots and outputs

Exercise 2: Exploring Variables

Choose two other variables in the dataset (aside from EPI & MHP) and print/plot the following for both variables:

- Variable summaries
- Variable boxplots
- Histograms with overlaid theoretical probability distributions
- ECDF plots
- QQ plots of each variable against the normal distribution
- QQ plot of the 2 variables against each other
- Normality statistical tests for each variable
- A statistical test for whether the variables having identical distributions

Upload your code to LMS

- Upload your code script (.R file) for exercise 2

Next Week:

September 15th – Linear models, kNN, Decision Trees

September 18th – lab 2

Have a great weekend!

