Rensselaer

why not change the world?®

Lab exercises: 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 12th, 2025

Tetherless World Constellation
Rensselaer Polytechnic Institute

Lab tasks

- Data Frames in R
- Exploring variables
- Fitting distributions

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.xls
  - 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/7q5chie4ig43uonqy2yf66kyvivyg9p9

- Lab1_code_sample.R
- epi_results_2024_pop_gdp.csv

Exercise 1: Starting out

- Run the code in the lab 1 code sample scripts
- 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 overlayed 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
- Statistical test for the variables having identical distributions

Push your Lab code to Github

• Please create a folder for Lab1 in your Github repository and push your code/plots.

• Upload your code script (.R file) and a docx/pdf with the plots/outputs

- Plots can be copied from the Export button in RStudio

Next Class:

September 16th – Introduction to analytical methods, data mining, machine learning

Thanks!

Have a great weekend!!!