IMT 573: Module 1 Lab
Basic Data Analysis in R

Ali Qazi

Due: June 25, 2021

Collaborators: COLLABORATORS' NAMES HERE. Akeel Qazi

Objectives

install.packages("styler")

This lab is your first opportunity, in this course, to write R code for data science. We will practice basic R syntax and take a look at the different data structures that will be used throughout this course. We will practice with built in data in R and even make some preliminary visualizations! The primary objective of this lab is to test out RStudio and RMarkdown – to ensure you have no challenges in getting these tools to work for this course. Please reach out if any questions arise.

Instructions

Before beginning this assignment, please ensure you have access to R and RStudio; this can be on your own personal computer or on the IMT 573 R Studio Cloud.

1. Open the `01_lab_basics.Rmd` and save a copy to your local directory. Supply your solutions to the assignment by editing `01_lab_basics.Rmd`.

2. First, replace the "YOUR NAME HERE" text in the `author:` field with your own full name. Any collaborators must be listed on the top of your assignment.

3. Be sure to include well-documented (e.g. commented) code chucks, figures, and clearly written text chunk explanations as necessary. Any figures should be clearly labeled and appropriately referenced within the text. Be sure that each visualization adds value to your written explanation; avoid redundancy – you do no need four different visualizations of the same pattern.

4. Collaboration on problem sets is fun and useful, and I encourage it, but each student must turn in an individual write-up in their own words as well as code/work that is their own. Regardless of whether you work with others, what you turn in must be your own work; this includes code and interpretation of results. The names of all collaborators must be listed on each assignment. Do not copy-and-paste from other students' responses or code.

5. All materials and resources that you use (with the exception of lecture slides) must be appropriately referenced within your assignment.

6. When you have completed the assignment and have **checked** that your code both runs in the Console and knits correctly when you click `Knit`. When the PDF report is generated rename the knitted PDF file to `lab1_YourLastName_YourFirstName.pdf`, and submit the PDF file on Canvas.

Setup

In this lab you will need, at minimum, the following R packages.

```
# Load standard libraries
library(tidyverse)
library(datasets)
```

Problem 1: Loading and Inspecting Data

In this problem we will use the `faithful` data from the **R datasets** package. Use this data to answer the following questions.

```
data('faithful') # Load data to workspace
```

(a) Where do these data come from?

```
library(help = "datasets")
```

(b) What data type is used to store this data?

```
x <- data(faithful)
```

(c) What are the variables in this dataset? Describe what each variable measures.

```
ls(faithful)
```

```
## [1] "eruptions" "waiting"
```
Hint: There are two variables of interest here.

Problem 2: Basic Computation

Continue to use the `faithful` data to answer these specific questions. summary(faithful)

(a) What is the average, or mean, waiting time between eruptions?

```
mean(faithful$waiting)
```

```
## [1] 70.89706
```

(b) What is the shortest time between eruptions that was observed?

```
min(faithful$eruptions)
```
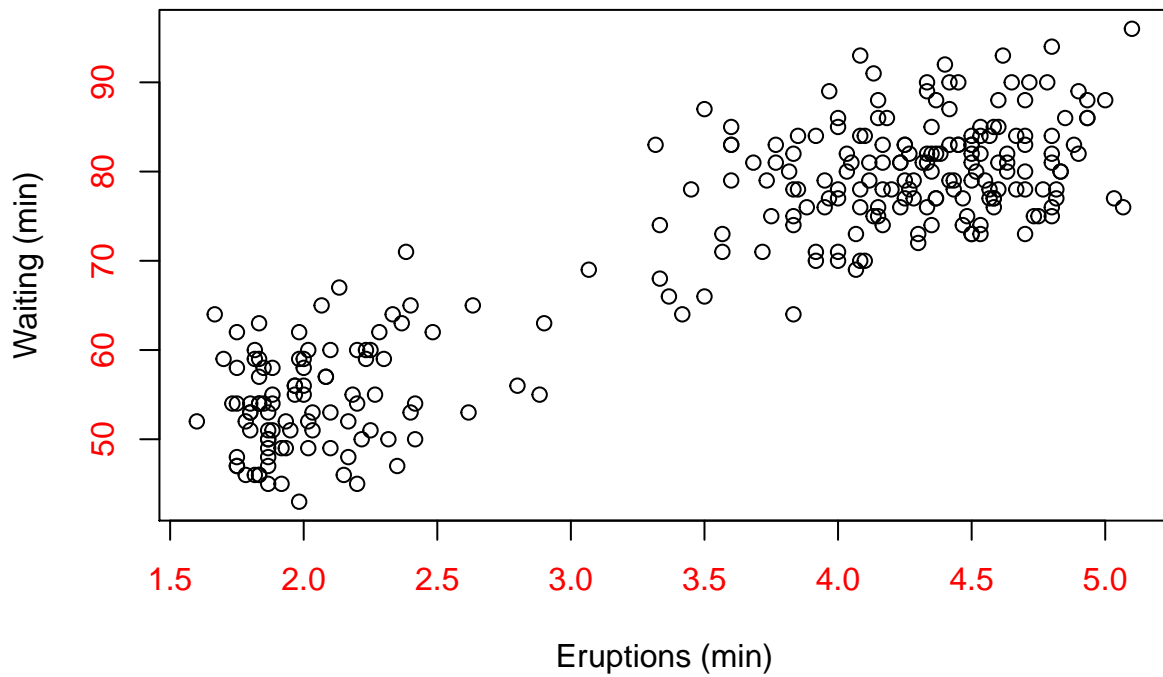
```
## [1] 1.6
```

(c) What was the longest eruption observed?

```
max(faithful$eruptions)
```

```
## [1] 5.1
```

Problem 3: Visualization

```
plot (faithful, xlab = "Eruptions (min)" , ylab = "Waiting (min)" , col.axis = "red")
```


Create a basic visualization of the distribution of eruption times or waiting time between eruptions, your choice.