Pwani University, Introduction to R Workshop (2024)

July 8th – 12th

# About the workshop

In this workshop, you will be introduced to the R and RStudio data analysis platforms. R is a simple programming language that enables the effective handling of data while providing excellent graphical support. R studio is a tool that provides a user-friendly environment for working with R. The course gives an overview of the basics of R and RStudio. The practical sessions will cover reading/importing data into the R environment, accessing and installing R packages, and writing annotated R code for data manipulation, analysis, and making graphical plots.

### Course prerequisites:

* Basic statistics: Took an introductory statistics course at the undergraduate level.
* Basic computing knowledge, such as, installing and uninstalling software.

### Computing requirements:

* A laptop with at least 8GB RAM. .
* Participants will download R and RStudio into their personal machines.
* Instructions of how to download R and RStudio are given below.

### Download R:

* Go to the R download site: https://cran.r-project.org/
* Click on the link, which matches your operating system.
* Download the current base version of R (R 4.4.1).
* When you run the executable file, accept the default setting, (click next throughout).
* Once you are done, you have R on your machine (look out for a blue R).

### Download RStudio:

* Go to https://posit.co/download/rstudio-desktop/ and download R-Studio for Desktop.
* Follow the prompts for your operating system (Mac/Windows).
* Once you finish your download, the icon for R studio is a white “R” in a blue circle.

**DAY 1 (8TH JULY 2024)**

**THEME:** Background, getting started, and fundamentals of R programming.

**OBJECTIVE:** Gain basic knowledge of the R ecosystem, understand its strengths and limitations compared to other alternatives.

|  |  |  |  |
| --- | --- | --- | --- |
| **Time** | **Description** | **Tutor** | **Assisted by** |
| 08:30 | Arrival at the training computer room |  |  |
| 08:45 | Introductions and course outline | Osman/Mark |  |
| 09:05 | Introduction to R and R studio | Mark | Esther |
| 09:25 | Setting up and working with R and R Studio | Mark | Esther |
| 10:05 | R Syntax, data types, and data structures | Mark | Esther / Ken |
| 10:55 | BREAK | | |
| 11:15 | R functions, arguments, and packages | Mark | Esther / Ken |
| 11:50 | File management in R (Working directory and R studio projects) | Mark | Esther / Ken |
| 12:35 | Writing an R script - good coding practices. | Mark | Esther / Ken |
| 13:00 | LUNCH | | |
| 14:00 | Hands-on Practice | Mark | Esther / Ken |
| 16:30 | Day summary and Resources | Mark |  |

**DAY 2 (9TH JULY 2024)**

**THEME:** Data manipulation.

**OBJECTIVE:** To learn how to transform raw data into a usable format for analysis.

|  |  |  |  |
| --- | --- | --- | --- |
| **Time** | **Description** | **Tutor** | **Assisted by** |
| 08:30 | Day 1 review and Q&A | Mark |  |
| 8:45 | Data Wrangling using **Tidyverse** (Part 1):   * Tidyverse ecosystem * Importing data into R * Exploring data frames * Creating new variables | Ken | Mark / Boniface |
| 09:45 | Data Wrangling using **Tidyverse** (Part 2):   * Subsetting data * Reshaping data * Merging two data frames * Exporting data from R | Ken | Mark / Boniface |
| 10:55 | BREAK | | |
| 11:55 | Data Wrangling using **Tidyverse** (Part 3):   * Working with Factors * Working with Dates and Times * Working with Strings/Text | Ken | Mark / Boniface |
| 13:00 | LUNCH | | |
| 14:00 | Hands-on Practice | Ken | Mark / Boniface |
| 16:30 | Day summary and Resources | Ken |  |

**DAY 3 (10TH JULY 2024)**

**THEME:** Data visualization

**OBJECTIVE:** To learn how to create informative and effective visual representations of data.

|  |  |  |  |
| --- | --- | --- | --- |
| **Time** | **Description** | **Tutor** | **Assisted by** |
| 08:30 | Day 2 review and Q&A | Ken |  |
| 08:45 | Visualizing Data in R:   * The grammar of graphics * Plot types * Guidelines for making good plots * Making plots using **ggplot2** | Mark | Ken / Nelson |
| 10:55 | BREAK | | |
| 11:15 | Visualizing Data in R (continued):   * Making plots using **ggplot2** | Mark | Ken / Nelson |
| 13:00 | LUNCH | | |
| 14:00 | Hands-on Practice | Mark | Ken / Nelson |
| 16:00 | Day summary and Resources | Mark |  |

**DAY 4 (11TH JULY 2024)**

**THEME:** Summary statistics

**OBJECTIVE:** Learn how to calculate and interpret descriptive statistics in R.

|  |  |  |  |
| --- | --- | --- | --- |
| **Time** | **Description** | **Tutor** | **Assisted by** |
| 08:30 | Day 3 review and Q&A | Mark |  |
| 08:45 | Explore and summarize data using R (part 1):   * Summary statistics for continuous variables (mean, median, mode) * Summary statistics for categorical variables (e.g. proportions) | Ken | Mark / Nelson |
| 10:55 | BREAK | | |
| 11:15 | Explore and summarize data using R (part 2):   * Comparison of means * Comparisons of proportions | Ken | Mark / Nelson |
| 13:00 | LUNCH | | |
| 14:00 | Hands-on Practice | Ken | Mark / Nelson |
| 16:30 | Day summary and Resources | Ken |  |

**DAY 5 (12TH JULY 2024)**

**THEME:** Exploring relationships between variables

**OBJECTIVE:** To identify patterns and associations in data.

|  |  |  |  |
| --- | --- | --- | --- |
| **Time** | **Description** | **Tutor** | **Assisted by** |
| 08:30 | Day 4 review and Q&A | Ken |  |
| 08:45 | Explore the relationship between two continuous variables:   * Scatter plots * Correlation analysis * Linear regression | Mark | Ken / Esther |
| 10:55 | BREAK | | |
| 11:15 | Explore the relationship between two categorical variables:   * Two-by-two tables * Chi-Square Test | Mark | Ken / Esther |
| 13:00 | LUNCH | | |
| 14:00 | Hands-on Practice | Mark | Ken / Esther |
| 16:00 | Day summary and Resources | Mark |  |
| 16:30 | Closing remarks | Osman / Ken / Mark |  |