University of Sri Jayewardenepura
Faculty of Applied Scineces
Department of Statistics

Batch: 2016/2017   **Year: 2020**   **Semester: First Semester**

Course Unit: STA 326 2.0 Programming and Data Analysis with R

Type of the course unit: Core for special degree students/ Optional for others

Pre-Requisites:
STA 114 2.0 Probability and Distribution Theory I, STA 123 2.0 Probability and Distribution Theory II, STA 124 1.5 Data Analysis I, STA 213 2.0 Statistical Inference, STA 226 1.5 Data Analysis II

Workload:
Minimum total expected workload to achieve the learning outcomes for this unit is 100 hours per semester typically comprising a mixture of scheduled learning activities, independent study and 26 hours of lectures. Independent study may include associated readings, assessment and preparation for scheduled activities.

Course Objective(s):

- To introduce how to program efficiently in R.
- To provide an in-depth, more advanced coverage of data processing, visualisation and analysis in the R programming environment.

Course Contents:

1. R basics: Objects in R, Data types, Operations, Installing packages, Control structures, Piping

2. Writing functions in R

3. Data analysis with the tidyverse

   3.1. Data import and export

   3.2. Data wrangling: Tidy data principles, Reshaping data into tidy form, Data transformation

   3.3. Data visualization: The grammar of graphics

   3.4. Statistical modelling and inference

   3.5. Communication: Dynamic reproducible reporting

Learning Outcomes: At the end of this course, the student should be able to:

- navigate the R integrated development environment (IDE) R Studio
- execute basic arithmetic operations in R
- define data classes, object attriutes, data structures in R
- describe differences in different data structures
- write user-defined functions to solve a given problem in R
- solve fundamental error problems and bugs in R programs
- describe the principles of tidyverse programming
- use Tidyverse packages in data science workflow
- import external data into R for data processing and statistical analysis.
- organise complex, messy, data into the most convenient form for analysis or reporting.
- create data graphics using the ggplot2 package
- select effective visualisation and modelling approaches to understand relationships between variables, and make decisions with data
- interpret the results of analysis and communicate these to a broad audience.
- Evaluate several functions that perform the same task and explain which one is better and why?
- Revise functions/programs to improve the readability and computational efficiency

Method of Assessment:

- Continuous assessment: 20%
- Final examination: 80%

Recommended Readings:

- Course website: Everything you want to know about the course, and everything you will need for the course (links to weekly reading, tutorials and lecture materials) will be posted at hellor.netlify.com
  Author: Thiyanga Talagala

- Title: R for Data Science
  Author(s): Hadley Wickham and Garrett Grolemund
  Publisher: O'REILLY
  This book is available online for free. Visit https://r4ds.had.co.nz/

- Title: Advanced R
  Author(s): Hadley Wickham
  Publisher: Chapman & Hall/CRC
  This book is available online for free. Visit https://adv-r.hadley.nz/

Lecturer in Charge:

  Name: Dr Thiyanga Talagala

  Email: ttalagala@sjp.ac.lk

Note: The course can be dropped on or before 21 February 2020.