International Summer School in Uganda

Syllabus for course: “Introductory Course to R with Applications from Data Analysis”

Lecturers: Dr. Jan-Philipp Kolb
E-mail: Jan-Philipp.Kolb@gesis.org
Homepage: https://github.com/Japhilko

Shelmith Kariuki
E-mail: kariukishelmith@gmail.com
Homepage: https://github.com/Shelmith-Kariuki

Date: September 23-27, 2019
Time: Monday to Friday, 8:30-12:30 and 13:30-17:00

About the Lecturers

Jan-Philipp Kolb is a senior researcher at the Leibniz Institute for Social Sciences (GESIS), working as a survey statistician in the GESIS Panel team. He previously worked as a research assistant in the Economic and Social Statistics Department at the University of Trier teaching sampling techniques and applied statistics using R.

Shelmith Nyagathiri Kariuki is a Senior Data Analyst at Busara Centre for Behavioral Economics. She is also a Zindi Ambassador for Kenya, where she furthers the Zindi mission of building the data science ecosystem in Africa. Shelmith has previously worked as an assistant lecturer in various Kenyan universities, teaching units in Statistics and Actuarial Science. She holds a Bsc in Actuarial Science and Msc in Applied Statistics from JKUAT. Shelmith has extensive experience in data analysis using R and Python and is at the forefront of the AfricaR initiative, striving to achieve improved representation of Africans in the global R community by encouraging, inspiring, and empowering Africans of all genders.

Short Course Description

The open source software package R is free of charge and offers standard data analysis procedures as well as a comprehensive repertoire of highly specialized processes and procedures, even for complex applications. Emphasis in this course will be on methods of graphically-based data analysis as R is particularly suitable for it.

Keywords

R, data import and export, data cleaning, statistical graphics, data analysis, survey data, web scraping

Course Prerequisites

- Basic familiarity with the use of a computer.
- Prior experience with data analysis, basic statistics, and regression. For a refresher in statistics, see http://onlinestatbook.com/2/index.html.
- Experience in dealing with other statistics packages is helpful, but not a requirement.

Target Group

- The course is for people who want to get started with R.
- It might be helpful for people that work with survey data and want to use R as an additional tool.

Course and Learning Objectives

Our workshop provides a hands-on introduction to R and lays the foundations for independently developing your skills in dealing with the programming language R. The participants can expect to receive an overview of the functional scope of R, master the import and export of data, and how to perform basic data analysis in R.
Software and Hardware Requirements
Please bring your own laptops for use in the course and install the following R packages: knitr, Rcmdr, lme4, devtools, ctw, readxl, lattice, xlsx, tibble, haven, foreign, readstat13, org, rio, Hmisc, naniar, memisc, tidyverse, forcats, car, reshape, DT, dplyr, magrittr, jtools, Metrics, visreg, AmesHousing, corrrplot, DAAG, caret, stargazer, faraway, ggplot2, psci, MASS, arm, survey, svyPVpack, mlmRev, vioplot, beanplot, psych, AER, igraph, ggraph, plotly, kknn, maps, dbscan, rvest, rtweet, gapminder, mlbench, purrr, compare, ggmap, leaflet, maptools, raster, rgdal, colorRamps, sp, osmplotr, osmdata, tmap, io, kableExtra, tmaptools before the course.

Long Course Description
Getting started
The first session will cover all preliminary topics. For example, Rstudio is a graphical user interface which makes beginning with R easier. Many of its available features and Add Ins will be explained at the first day. In addition we will cover data import and export as well as the data processing.

Graphics
The foundations (basic graphics) are presented. How to create a boxplot, barplot, density plot etc. Participants will learn how to visualize data using the basic functions, the lattice package and the ggplot2 package.

Basic data analysis
In this part we will give an introduction to basic data analysis. We will use the campus file of the GESIS Panel for those purposes. We will cover basic statistical tests, various regression techniques (linear, logit, lasso and multilevel regression) and clustering methods. In addition we will give a short introduction in how to use the survey package.

More on data analysis
In addition we will give a short introduction in how to use the survey package. In addition some clustering algorithms and their usage in R will be presented. The generation of maps will be presented.

R-programming: Loops, Functions and packages
In this part we plan an introduction to basic programming. Use of the “with” function (instead of “attach”). The basics of debugging and how the functional programming paradigm makes debugging easier will also be covered.

Day-to-day Schedule and Literature

<table>
<thead>
<tr>
<th>Day</th>
<th>Topic(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Getting started with R and Rstudio, how to get help; understanding error messages, how to find out what an error means, data import. Use Case: Data Processing of GESIS Panel data</td>
</tr>
<tr>
<td></td>
<td>Suggested reading: Garrett Grolemund and Hadley Wickham (2016) R for Data Science, <a href="https://r4ds.had.co.nz/">https://r4ds.had.co.nz/</a></td>
</tr>
<tr>
<td>2</td>
<td>Graphics: basic and lattice plots as well as the use of the ggplot2 package.</td>
</tr>
<tr>
<td></td>
<td>Compulsory reading: <a href="https://ggplot2.tidyverse.org/reference/">https://ggplot2.tidyverse.org/reference/</a></td>
</tr>
<tr>
<td>3</td>
<td>Basic data analysis - t-tests, chi-square-tests; General regression (linear, logit, etc.);</td>
</tr>
</tbody>
</table>
**Compulsory reading:**

**Suggested reading:**
https://cran.r-project.org/doc/contrib/Faraway-PRA.pdf

<table>
<thead>
<tr>
<th>4</th>
<th>More data analysis - Using the survey package; clustering (kmeans) and creating maps.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compulsory reading:</td>
<td></td>
</tr>
<tr>
<td>Suggested reading:</td>
<td></td>
</tr>
<tr>
<td><a href="https://geocompr.robinlovelace.net/">https://geocompr.robinlovelace.net/</a></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5</th>
<th>R-programming: Loops, Functions and packages; the advantages of the functional programming paradigm (debugging/parallelization), and recursion.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compulsory reading:</td>
<td></td>
</tr>
<tr>
<td>Suggested reading:</td>
<td></td>
</tr>
<tr>
<td><a href="http://adv-r.had.co.nz/Functional-programming.html">http://adv-r.had.co.nz/Functional-programming.html</a></td>
<td></td>
</tr>
</tbody>
</table>

**Preparatory Reading:**


**Additional Recommended Literature:**


