Data Analysis with R
Syllabus
22/03/2024

Instructor: Gentian Gashi

Format: Online
Contact Information: ggashi@worldbank.org

Course Dates: April 17, 2024 – June 17, 2024

Course Time: Wednesdays, 13:30 – 15:15

Course Description
Welcome to the Data Analysis with R course! In this syllabus, you'll find an overview of the topics we'll cover throughout the program, the learning outcomes you can expect to achieve by the end of the course, and pointers to additional resources.

This course is designed to equip participants with the essential skills and knowledge needed to proficiently use R for data analysis tasks in economic and financial contexts. Through a combination of lectures, hands-on exercises, and practical examples, you'll learn how to leverage R to import, manipulate, visualize, and analyze data effectively.

Learning Outcomes
1. Understand the fundamentals of R programming, including data types, variables, basic operations, and functions.
2. Utilize advanced data manipulation techniques with the dplyr and tidyverse packages to clean, transform, and prepare data for analysis.
3. Create impactful data visualizations using ggplot2 to communicate insights and trends effectively.
4. Implement version control using Git within the RStudio environment for project management and collaboration.
5. Complete a simple in-course data analysis project by the end of the course, and optionally complete an end-to-end data analysis project based on a homework dataset.

Course Structure
- Duration: 10 lectures, each approximately 1.5 hours in length with a 15-minute pause in the middle.
- Format: Each lecture will consist of a combination of lecture-style teaching, hands-on exercises, and discussions.

Course Outline

Lecture 1: Set-up and Introduction to R
- Introduction to course and expected outcomes
- Overview of R and RStudio
- Introduction to R basics
  - Syntax, arithmetic, variables, data types, vectors, indexing and slicing
- Introduction to R matrices and data frames
- Introduction to lists

Lecture 2: (R) Programming Basics
- Logical operators

- Conditional if-else statements
- Loops
- Functions

Lecture 3: Data Manipulation with R (Part 1)
- Data input and output with R
  - Excel, CSV, Stata, SPSS data files
- Introduction to R packages
- Introduction to *dplyr*
- Working with pipes for improved code readability
- Grouping and summarizing data with
- Practical examples and exercises

Lecture 4: Data Manipulation with R (Part 2)
- Introduction to the *tidyverse*
- Working with strings and dates
- Altering data types
- Real-world examples

Lecture 5: Data Visualization with *ggplot2* (Part 1)
- Introduction to *ggplot2*
- Basic plotting techniques
- Customizing plot aesthetics and themes
- Practical exercises on creating various types of plots

Lecture 6a: Introduction to Version Control with Git
- Overview of version control and its importance
- Introduction to Git and GitHub
- Setting up a Git repository within RStudio
- Basic Git commands for version control

Lecture 6b: Creating Your First Data Analysis Project
- Project creation and version control
- Introduction to dataset and analysis questions
- Introduction to a code-based analysis workflow
- Introduction to optional homework assignment

Lecture 7: Project Analysis (Part 1)
- Reviewing analysis questions
- Data loading
- Data cleaning
- Data transformation
- Discussion

Lecture 8: Project Analysis (Part 2)
- Reviewing analysis questions
- Data preparation for analysis
- Data analysis

Lecture 9: Project Analysis (Part 3)
- Reviewing analysis questions
- Data preparation for visualization
- Graph customization
- Graph export

Lecture 10: Final Project Presentations, Review and Wrap-Up
- Project finalization
- Project and course review
- Feedback and discussion on project outcomes
- Course wrap-up and next steps for further learning

Additional Resources

In addition to the course lectures and materials provided, participants are encouraged to explore the following resources to further enhance their learning experience:

1. RStudio Cheatsheets:
   - RStudio's cheatsheets (https://posit.co/resources/cheatsheets/) offer quick references and tips for using R, RStudio, and several packages effectively.
2. Online Learning Sources:
   - Posit (Youtube, https://www.youtube.com/@PositPBC)
   - Stack Overflow (https://stackoverflow.com/)
   - From Data to Viz (https://www.data-to-viz.com/)
   - R Graph Gallery (https://r-graph-gallery.com/)
   - 1 dataset 100 visualizations (https://100.datavizproject.com/)
   - ggplot2 extenstions (https://exts.ggplot2.tidyverse.org/gallery/)
   - tidyverse (https://www.tidyverse.org/packages/)
   - #tidytuesday (https://github.com/rfordatascience/tidytuesday)
   - Cedric Scherer (https://www.cedricscherer.com/top/dataviz/)
   - Albert Rapp (https://www.youtube.com/@rappa753/videos)
   - R for the Rest of Us (https://www.youtube.com/@rfortherestofus)
3. Books:
   - "R for Data Science (2e)" by Hadley Wickham, Mine Çetinkaya-Rundel, Garrett Grolemund (https://r4ds.hadley.nz/)
   - "R Graphics Cookbook, 2nd edition" by Winston Chang (https://r-graphics.org/)
   - "ggplot2: Elegant Graphics for Data Analysis (3e)" by Handley Wickham (https://ggplot2-book.org/)
   - "An Introduction to Statistical Learning with applications in R" by Gareth James et al. (https://www.statlearning.com/)
   - "Storytelling with Data" by Cole Nussbaumer Knaflic (https://www.storytellingwithdata.com/books)
   - "Applied Econometrics with R" by Christian Kleiber and Achim Zeileis (https://link.springer.com/book/10.1007/978-0-387-77318-6)
   - "Forecasting: Principles and Practice, 3rd edition" by Rob J Hyndman and George Athanasopoulos (https://otexts.com/fpp3/)
4. Online Tutorials:
   - DataCamp (https://www.datacamp.com/tracks/data-analyst-with-r)
   - Udemy (https://www.udemy.com/course/data-science-and-machine-learning-bootcamp-with-r/?couponCode=ST14MT32124)
   - Coursera (https://www.coursera.org/professional-certificates/google-data-analytics)