R for Data Cleaning

Facilitation Guide - 85 minutes

Description

In this module, participants will be introduced to R: a scripting language useful for reproducing and/or sharing workflows. They will learn to navigate R-Studio, and learn how to read and write an R-script that can input data from a CSV, reshape it,and output a cleaned CSV file. Participants will also be provided with additional resources, suggested applications, and the chance to test common functions in the presence of an instructor for troubleshooting purposes.

Learning Objectives

Participants will…

- Learn what R is and how it can be used
- Become familiar with R-Studio
- Learn how to read and write an R-script

Materials

Presentation
Pre and Post Assessment
Optional: Virtual Question Parking Lot (i.e. Jamboard)

Facilitator Guide	
Welcome and Getting Started (Slides 1-2) - Introduce yourself and your role. Share the purpose of the session and discuss the Learning Objectives. 5 Minutes	Tips & Tricks ● Key Ideas to Highlight ○ This is going to be fun! You'll learn a lot. ● Resources to Share ○ Pre-Assessment
● We will introduce R and what it is used for ○ R is a scripting language for reproducing/sharing workflows	

Take a few minutes to go through the pre-assessment and see what you know. If you're unsure/have questions about the content, those are great places to start when engaging with the presentation.

Consider – Using a virtual tool like Jamboard to serve as a question parking lot for participants to share questions that come to mind throughout the module.

Introduction to R (Slides 3-12) - In this section, we'll provide participants with a foundational explanation of what R and R-Studio can do. 10 Minutes	Tips & Tricks ● Key Ideas to Highlight ○ Tidy vs Messy data ○ Gathering and Spreading ● Resources to Share

- We will…
  - Introduce "R" and "R-Studio" and their uses.
  - Define data structure in terms of observations (records) and variables (attributes)
    - Highlight difference between tidy and messy data structure
    - Why we clean data
  - Explain the phrase "Split-Apply-Combine"
  - Explain how data can be reshaped with Gathering and Spreading

R-Studio Setup (Slides 13-23) - In this section, we'll cover the installation process and how to create a project. Participants will learn how to navigate around the R-Studio interface and project directory, the importance of mapping and planning, and where to find help. 15 Minutes	Tips & Tricks ● Key Ideas to Highlight ○ How to create a project ○ Mapping and Planning ○ Where to find help ● Resources to Share ○ Help Desk email ○ Handouts of links/resources

- We will…
  - Help participants install R and R-Studio
    - Provide a link to CheatSheet packages and explain what they can do

- Guide participants through starting a project
    - Have participants save a Test to their desktop
    - Navigate to R Script and the Source Editor pane
    - How to navigate the R-Studio interface
    - How to navigate your project directory
- Show participants how to change the background theme
- Explain the purpose of Mapping and Planning
    - How to find your functions and corresponding packages

Script Setup (Slides 24-34) - In this section, we'll cover basic terminology and guide participants through setting up a script. We will also offer tips and present a few style guides that participants may find helpful. 15 Minutes	Tips & Tricks ● Key Ideas to Highlight ○ Basic terminology ○ Setting up a script ● Resources to Share ○ Style guide links

- We will…
    - Define Objects, Functions, Arguments, and Comments
        - Define Data Frames and Vectors, Packages and Library
    - Offer tips and Style Guide resources
    - Guide participants through setting up a script
        - Explain first 4 lines
        - Installing and loading tidyr package
        - How to manually set a working directory (but explain why this step may not be needed)
        - How to save your script
        - How to run parts of your script

Importing, Viewing, and Reshaping Data (Slides 35-47) - In this section, we'll guide participants through the process of importing data into R and ways to reshape that data. We will highlight important functions and make relevant CheatSheets available. 20 Minutes	Tips & Tricks ● Key Ideas to Highlight ○ read.csv and strings as factors ○ rawLoans ○ Gathering and Spreading ○ Subsetting ● Resources to Share ○ CheatSheet links
● We will… ○ Guide participants through importing data into R ■ Discuss the function read.csv and talk about strings as factors ■ Navigate Environment pane and explain rawLoans ○ Explore the data preview with the Filter button, Sort icons, and search bar ○ Guide participants through reshaping data ■ Examples of Gathering and Spreading ■ Explain subsetting	
Exporting Data & Extras (Slides 48-59) - In this section, we'll cover the process of exporting cleaned data out of R. We will allow time for participants to test out functions so the instructor can help troubleshoot errors. We will also introduce Git as a way to track changes in code, and provide additional links and resources for participants. 20 Minutes	Tips & Tricks ● Key Ideas to Highlight ○ Exporting your CSV ○ Unmangle your data ○ Git ● Resources to Share ○ Post Assessment ○ Links to additional resources
● We will… ○ Guide participants through exporting data ■ Explaining how to fix data that is mangled in Excel ■ Let participants try out other functions, and troubleshoot unexpected results ○ Introduce participants to Git	

        - Offer link for installation and a CheatSheet link
    - Provide additional resources

- **Questions?** - Check to see if your audience has any questions about the presentation. If used, refer back to the virtual question parking lot (i.e. Jamboard) to answer any last questions.

- **Thank you!** – Give yourself a round of applause! Share link to post-assessment for the final Check for Understanding.