

Workshop: *Basic R and More*
Jackson Heart Study Training and Education Center
Summer Research Institute II
University of Mississippi Medical Center (UMMC)
Jackson, MS
07/14/2023 - 07/15/2023

Dr. Philip Turk
Professor of Data Science
Chair, Department of Data Science
John D. Bower School of Population Health
Email: pturk@umc.edu
Cell: (304) 376-5377

- Workshop Type: Five-hour workshop on an introduction to R
- Timeline: 3:30 PM - 5:30 PM, Friday, July 14; 2:30 PM - 5:30 PM, Saturday, July 15
- Facilitators: Dr. Ben Walker (bwalker5@umc.edu), Assistant Professor of Population Health Science; Dr. Xiaoqian Zhu (xzhu3@umc.edu), Biostatistician IV, Department of Data Science
- Workshop Description
 - Abstract: In this five-hour workshop, participants will be introduced to the basics of the R programming language and RStudio. We will go through types of data structures, how to access data, and how to manipulate, reshape, and visualize data with R's `tidyverse` packages. In addition, we will cover working with classes of data commonly seen in the healthcare sector; namely, strings, dates, and times. Finally, we will end the workshop by understanding how to save and communicate our results reproducibly with R scripts. **Participants are strongly encouraged to use their computers to work through the examples demonstrated.**
 - Outline: Introduction to R; Vectors, Matrices, and Data Frames; Importing Data; Using `dplyr`; Data Reshaping; Using `ggplot2`; Working with Strings, Dates, and Times; R Scripts
 - Intended Level of Participant: This workshop is designed for individuals who have little to no experience with R and RStudio.
 - Prerequisite: It is assumed that the participant has some prior experience with data management or analysis through tools such as Excel. It is also assumed that each participant has R and RStudio already installed on their computer. If you do not, then instructions are given at the end of this handout so that you can install it. Please feel free to contact me if you run into installation problems.
 - Objective: The objective for this workshop is to help attendees learn how to use R software and the RStudio Integrated Development Environment in their

current workflows. Attendees will learn how to use data engineering to obtain and wrangle data. After this workshop, one should have an understanding of basic R coding/programming fundamentals, as well as methods in R for describing and visualizing data. Attendees will also learn how to save and share analysis results with other end users through the concept of reproducible research.

- Relevance to Professional Development

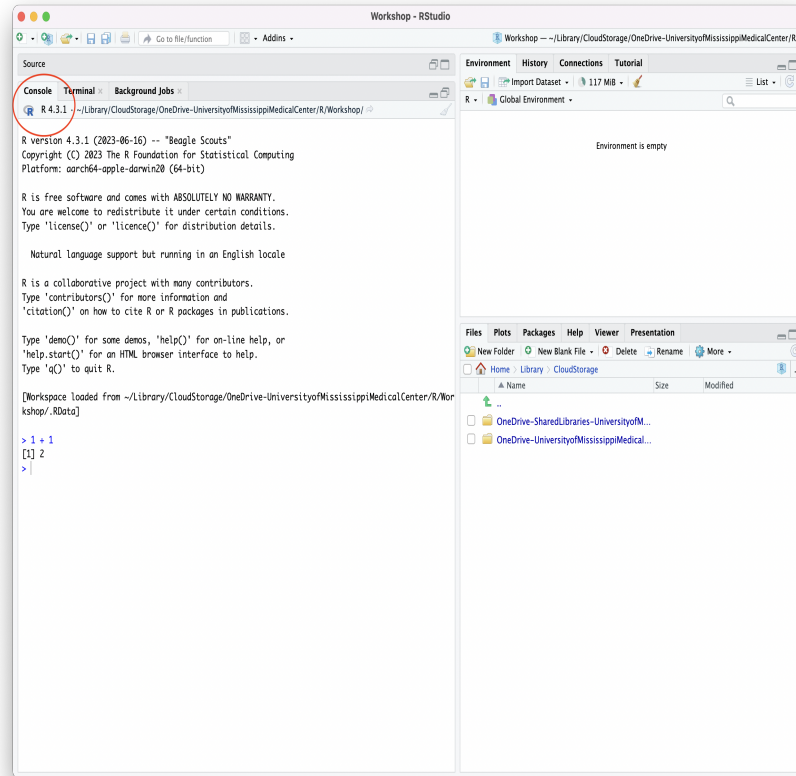
Times are changing and the continued rise of data science needs will create 11.5 million job openings by 2026 (U.S. Bureau of Labor Statistics). It is no longer sufficient to only be able to generate graphs and simple summary statistics in Excel. What with the rise of AI and its cohorts deep learning and machine learning, it is critical that organizations pivot towards a culture of data analytics. As a cornerstone, this requires a basic understanding of current, yet evolving, statistical softwares packages.

By learning R, attendees will understand a valuable technical skill after this workshop that they will be able to immediately apply to their research. Through the use of R scripts, they will be able to conduct reproducible research. Because both R and RStudio are open source and therefore free, implementation will save organizations money spent on costly site licenses for proprietary software.

- Software Packages: Both R and RStudio will be used in this workshop. As mentioned, both R and RStudio are open source and therefore free.
- Setting up R and RStudio: **Steps 1 - 4 need to be done prior to the workshop.** Having your computer set up correctly before the workshop maximizes your enjoyment and what you will get out of the experience. If you have any problems setting up your computer, then please contact either one of the facilitators or me well before the workshop.

- Step 1: Download and install R (<https://cran.r-project.org>)
 - * Click on “Download R for Windows” and then “base” for Windows users, or “Download R for macOS” for Mac users.
 - * Click on “Download R 4.3.1 for Windows” for Windows users, or “R-4.3.1.pkg” for Mac users (n.b., Mac users must specify the chip in their computer).
 - * When the download is complete, click on the downloaded .exe or .dmg file, and finish the installation using the setup wizard.
- Step 2: Download and install RStudio Desktop (<https://www.rstudio.com/products/rstudio/download/#download>)
 - * Click on “RSTUDIO-2023.06.1-524.exe” for Windows users, or “RSTUDIO-2023.06.1-524.dmg” for Mac users. When the download is complete, click on the downloaded .exe or .dmg file, and finish the installation using the setup wizard.
- Step 3: Check that everything is running
Click on the RStudio icon on your desktop or within your applications folder. What you see should resemble the figure that follows. The R Console is where

R code is executed and output is shown. You should see the R version on the top-left corner of the Console (circled in red). At the prompt `>`, try typing `1 + 1` and press Enter. You should see the number 2 show up. If so, R is running successfully within RStudio.



- Step 4: Install packages necessary for workshop

If you are comfortable with the RStudio environment, then install the R packages **tidyverse** and **pander**. To install an R package, type at the command line

`> install.packages("<the package's name>")`. R will download the package from a centralized repository, so you'll need to be connected to the internet. If you have network issues with security and firewalls while on campus, then try using public Wi-Fi.