Basic R and More
Wake Forest Virtual Short Course
July 9, 2021

Dr. Philip Turk
Director, Biostatistics and Data Science
Center for Outcomes Research and Evaluation (CORE), Atrium Health
Email: Philip.Turk@atriumhealth.org
Cell: (304) 376-5377

- Course Type: Full-day (8 hours) short course on an introduction to R
- Timeline: 8:00 AM - 12:00 PM; 1:00 PM - 5:00 PM; Friday, July 9
- Course Description
  - Abstract: In this one-day course, 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. Basic programming concepts will be introduced. 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 course by understanding how to communicate our results reproducibly with R scripts and R Markdown. **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`; Basic Programming; Working with Strings, Dates, and Times; R Scripts; R Markdown
  - Intended Level of Participant: This course is designed for individuals who have little to no experience with R and RStudio.
  - Prerequisite: It is assumed that the participant has prior experience with data management or analysis through tools such as SQL or SAS. 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 course 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 course, one should have an understanding of basic R programming fundamentals, as well as methods in R for describing and visualizing data. Attendees will also learn how to 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, including Atrium Health, pivot towards a culture of data analytics. As a cornerstone, this requires a basic understanding of current, yet evolving, statistical softwares packages.

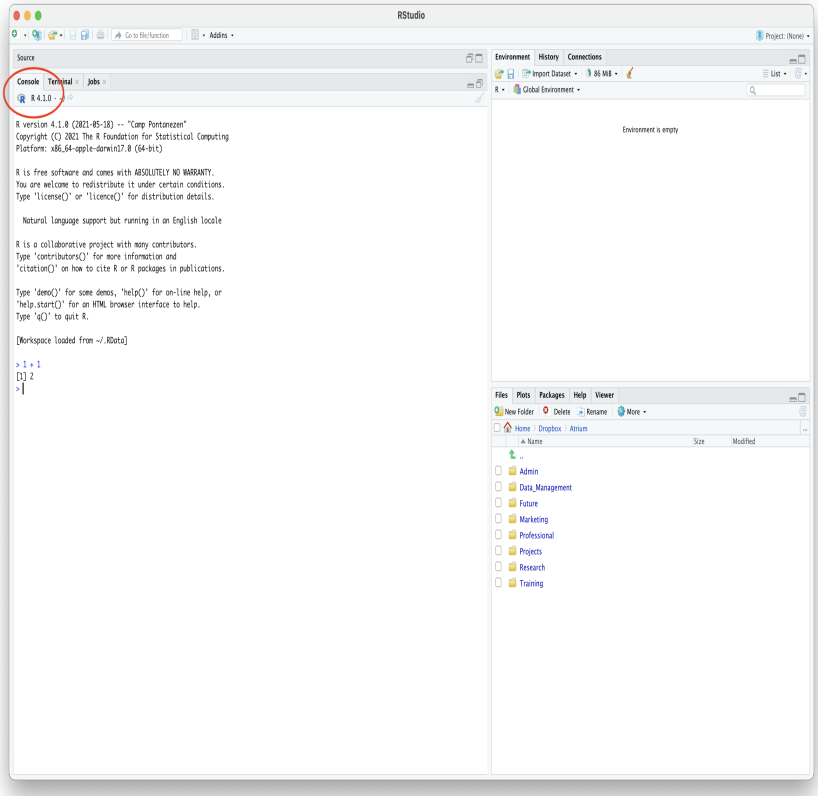
  By learning R, attendees will understand a valuable technical skill after this course that they will be able to immediately apply to their jobs. Through the use of R Markdown, they will be able to conduct reproducible research and more efficiently write routine reports. Because both R and RStudio are open source and therefore free, implementation will save Atrium Health money spent on costly site licenses for proprietary software.

- Software Packages: Both R and RStudio will be used in this course. As mentioned, both R and RStudio are open source and therefore free.

- Setting up R and RStudio

  - Step 1: Download and install R (https://cran.r-project.org)
    - Click on "Download R for Windows" for Windows users, or "Download R for macOS" for Mac users
    - Click on "base" for Windows users, or "R-4.1.0.pkg" for Mac users
    - Click on "Download R 4.1.0 for Windows". 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-1.4.1717.exe" for Windows users, or "RStudio-1.4.1717.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, then please see your IT group to do these installs.