

Intro to R/RStudio

April 2020

R has become one of the most commonly used tools across many science and social science disciplines for data manipulation, statistical analysis and creating publication quality data visualizations. Although the use of R at STATCAN is relatively recent, there is already a seasoned R-users group and many teams use this tool to help speed research.

The goal of this course is to provide novice R users a solid foundation in the best practices for data wrangling and analysis, while emphasizing strategies to ensure reproducibility. Learning a new programming language can be difficult. Even if you have coding experience, a new language introduces a new environment, styles, and syntax and can make you feel like a beginner again. There will be lots of opportunity for participants to move through exercises at their own pace and one-on-one assistance will be available throughout the course.

This course is intended for novice R users and analytics team leads/managers who'd like to learn how R/RStudio can be leveraged by their team. Throughout the course material, extra resources that can support further learning are highlighted.

Learning Objectives:

At the end of the two day session you be able to:

- 1) Perform basic data queries using dplyr and baseR
- 2) Perform data manipulations and summaries
- 3) Merge and append datasets
- 4) Produce fast plots for data exploration
- 5) Produce publication quality plots using ggplot2

Course Material:

https://github.com/MagB/R_Course1

System Requirements

Software should be available on a computer that does not require a STACAN network connection to access and run the software. If you are on a VDI only setup, then you may be limited to a personal machine.

Download R/Rstudio:

<https://rstudio.com/>

<https://www.r-project.org/>

Access R/RStudio through a browser (no installation of software needed)

<https://rstudio.cloud/>

Course Schedule:

Day 1:

- Intro to R/Rstudio IDE
- Finding course material
- Tips on using SLACK and Zoom
- Let's get you setup. Time will be set aside to offer one-on-one help to install, open and navigating R/Rstudio

Day 2:

- Why R/Studio (30-45min)—May be of general interest
- R as a Calculator
- Variable Assignment

Day 3:

- Subsetting/querying vectors
- 3_ Intro to Data Frames
- Exercises and practice
- Optional: interactive R learning with “SWIRL” package

Day 4:

- Base plots, basic data visualizations
- Installing and loading packages
- Importing data from excel, sas, csv

Day 5:

- Review material and exercises
- Introduction to the dplyr package
- select, filter, summarise, mutate

Day 6

- The piping operator: combining multiple data wrangling steps
- exercises

Day 7

- Review
- Joining and merging datasets

Day 8

- Intro to ggplot2 (fast, easy, publication quality data viz)
- Exercises

Day 9

- Review & Special requests