**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) Preform 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 Schedule:**

Week 1:

Day 1:

* + Intro to R/Rstudio IDE
    - Intro to R and RStudio environment
  + Let’s get you setup. Time will be set aside to offer one-on-one help to install, open and navigating R/Rstudio
    - 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.
    - Optional interactive R learning with “SWIRL” package

Day 2:

* + Why R/Studio (30-45min)—May be of general interest, including managers and management who do not code.
  + R as a Calculator
  + Variable Assignment

Day 3:

* + - Subsetting/querying -1 hr
    - Functions (in Supporting Info)
    - Exercises and practice

**Lunch 12 – 12:45**

Day 4:

* + 3\_Intro to Data Frames
    - Subsetting and querying data -30 min
    - Base plots, basic data visualizations -1 hrs
    - Exercises – 1 hr

Day 5:

* + 3\_Intro to Data Frames continued
    - review material
    - Installing and loading packages – 30 min
    - Exercises and review

Day 6

* + Review (30 min)
  + Intro to Advanced Data Wrangling (getting down with dpyr)
    - select, filter, summarise, mutate, arrange 1.5 hrs

Day 7

* + Pipe Operations 30 min
  + Joining and merging datasets

Day 8

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

Beyond 4 weeks: Even\_More\_R

-Importing different types of data

-Fun with Loops

-Writing fancy functions