**Intro to R/RStudio**

Methodology—February 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:**

Morning

Day 1:

* + Get one-on-one help opening and navigating R/Rstudio (optional 8:30 am)
  + 1\_ Why R/Studio (9 a.m. approx. 30 min)
  + 2\_Intro to R/Rstudio IDE
    - Intro to Rstudio environment (20 min)
    - R as a Calculator – 30min
    - Variable Assignment -30min

**Break 10:30**

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

**Lunch 12 – 12:45**

Afternoon

* + 3\_Intro to Data Frames
    - Subsetting and querying data -30 min
    - Base plots, basic data visualizations -1 hrs
    - Installing and loading packages – 30 min
    - Exercises – 1 hr
* Day 2

Morning

* + 9:00 a.m. Day 1 Review (30 min)
  + 4\_ Advanced Data Wrangling
    - select, filter, summarise, mutate, arrange 1.5 hrs
    - **Break at 10 :30**
  + Pipe Operations 30 min
  + Joining and merging datasets
  + Lunch 12 – 12:45

Afternoon

* + 5\_Intro to ggplot2 (1.5hr)
  + Exercises (1hr)
  + 6\_Even\_More\_R (time permitting)
    - Importing different types of data Fun with Loops
    - Optional interactive R learning with “SWIRL” package

**Pre-requirements:**

1. Please ensure that R and RStudio are installed on your machine:
   1. To do so, submit an SRM as soon as possible requesting the software be installed (NetA preferred).
2. Follow steps to create a .Rprofile file
   1. On Net A Confluence: https://confluence.statcan.ca/display/RUG/How+to+use+Artifactory