**Review of Pilot Course: Intro to R/RStudio - Feb 13 & 14, 2020.**

**Instructor:** Magdalena Bartkowska, PhD (analyst in ISTD) [magdalena.bartkowska@canada.ca](mailto:magdalena.bartkowska@canada.ca)

**Summary**

The pilot course “Intro to R/RStudio” was developed in partnership between Methodology (Yanick Beaucage and Guillaume Miranda) and CHSP (M Bartkowska, supervisor François Couture). Deirdre Hennessey from the Health Research Division and Steve Martin from Producer Prices Division reviewed the course content and syllabus in January 2020.

Generally, the course seemed well received. Participants were from ISTD (Investment Science and Technology Division), PPD (Producer Pricing Division), Census and Methodology.

Three changes are recommended if the course were to become a regular offering:

1. Increase the number of exercises on day 1
2. Dedicate more time to data merges, joins and binds
3. Create a stand along module for basic best practices for building reproducible workflows:
   1. How to track version of R and packages
   2. How to set and change global and project environments
   3. Documenting code, README files and folder/file structure]
4. Review course for typos and flow of information, and consider creating course notes.

**Goals of Participants**

We started the first day with a round table introduction, where participants were asked, “What would you need to learn to say that spending the next two days learning R was a success for you?” The most common answer was simply learning how to navigate basic data manipulations and general feel of the language.

All, but the one participant from Census, expressed that they have already experienced barriers to their work by not having a more solid understanding and ability to use R. In most cases, participants either work on a team where specific models and packages from R were being used by their team.

**Recommendations for updating and improving the pilot course:**

1) More exercises:

In general, more exercises on the first day of learning is recommended. This can be achieved either by writing more examples/practice questions or by using the interactive R training library called “SWIRL”. I highly recommend using the swirl library. This would require either 1) adding the basic R programming course to the artifactory repo on Net A, and/or 2) setting aside 25 min during the course to help participants load and install the swirl library and course content.

On the second day of the course, participants who were working through the exercise material quickly were shown how to access and load the R programming course from swirl. The participants who tired this package all said they really enjoyed it. This also allowed, students with stronger programming skills to dive into more advanced topics like building functions and loops on their own, while students with more basic programming were able to continue moving at their pace.

2) More content on data merging/joining and comparing different sets

For most participants, merging or comparing datasets is one of the most common and frequent steps taken in their work. In the current format of the course, only a short module (about 35min + 30 min exercise) was dedicated to this topic. More complex examples and increased number of exercises would help participants build a deeper comfort with these tasks.

3) Create a stand-alone module on reproducibility.

Throughout the course elements of establishing best practices for achieving reproducibility were discussed. Having a stand-alone module, however, would help ensure that any teacher of the course would cover this content. This would also help students to find that material if they were to use the course content as reference material.

4) Improve reference material for the course

The R scripts and documentation within them are the main course reference material currently available for this course. The written material should be reviewed for grammar/spelling and to ensure that wording of instructions/explanations is as clear as possible.