LATIS R Workshop Series
Summer 2021

R Script Reproducibility Framework

Directory and Files	**Yes / No / Maybe? (explain if necessary)**
Is it clear where the working directory should be set (e.g., where the data is coming from)?	
Are file paths used for data or output relative to the working directory (rather than containing absolute paths)? ● E.g., does the file names in commands such as "read.csv()" include the entire file path? Or only the specific sub-directory and file name? ● Are the names of any subdirectories self-explanatory?	
Is it clear what data files are read into with script?	
Script Organization	**Yes / No / Maybe? (explain if necessary)**
Are required packages loaded or noted at the beginning of the script ● Are any non-CRAN packages (e.g., github, self-created, etc) documented with their source location?	
Are headings or clear comments used throughout?	
If multiple scripts are included, is it clear which order they need to be run and which (if any) dependencies there are between them?	
Describing the Script	**Yes / No / Maybe? (explain if necessary)**
Are all data import, cleaning, analysis, and output steps clearly described by the script?	
Are object names used in the script self-explanatory?	
Are **decisions** behind data cleaning, analysis, and other scripts well documented within the code as annotations, or as a reproducible report (e.g. R markdown (*.Rmd))?	
Running the Script	
Do all lines of the script run without error? ● Helpful hint: try doing this on a different computer or have a friend try to run your script.	

This table was created by LATIS based on the Reproducibility Framework prepared by Shahira Khair, Data Curation Librarian, University of Victoria, Sandra Sawchuk, Librarian, Mount Saint Vincent University, and Qian Zhang, CLIR Postdoctoral Follow, University of Waterloo.

Does all output match what you expect? ● Does the R markdown (if applicable) output match previous knit reports? ● Do any figures or tables (if applicable) produce the same results? ● Do any numbers in the comments (if applicable) match the output?	
If any randomization is used, is set.seed() included?	
If the script takes significant compute power or time, is this described in comments? ● If analysis requires multiple cores, is that specified in comments? ● If lines take a long time to run, is that noted or checks/progress bar printed?	
Can the script be run in its entirety at once? ● E.g., no manual steps are required within the script	
Document Software & Packages	**Yes / No / Maybe? (explain if necessary)**
Is the specific version of R specified? ● Hint: use sessionInfo() ○ R Markdown - in visible code chunk ○ R Script - paste or sink to a text file	
Are versions of packages specified? ● Hint: use sessionInfo() ○ R Markdown - in visible code chunk ○ R Script - paste or sink to a text file	
Version Control	**Yes / No / Maybe? (explain if necessary)**
If there are multiple versions of an R script, is it clear which is the most up to date version? ● Hint: Use GitHub ● Or Hint: Use version of R by tacking on date at the end of the script-- YYYYMMDD format	
If GitHub or another publicly available version control repository is used, is a link to the repository included in the documentation?	

This table was created by LATIS based on the [Reproducibility Framework] prepared by Shahira Khair, Data Curation Librarian, University of Victoria, Sandra Sawchuk, Librarian, Mount Saint Vincent University, and Qian Zhang, CLIR Postdoctoral Follow, University of Waterloo.