
INST452: Health Data Analytics

Module 2 Reproducible Research Principles and Pipelines

Dr. Nikki Sigalo

What is reproducible research and why is it important?

- Reproducible research
 - Research is reproducible when others can replicate (REPRODUCE) the results of a scientific study given only the original data, code, and documentation
- There are two basic reasons to be concerned about making your research reproducible:
 - To show evidence of the correctness of your results
 - Provide audience with the code and data that demonstrates the decisions made as results were generated
 - Makes it easier for others to determine whether results are reliable
 - To enable others to make use of our methods and results

What are the different kinds of reproducible research?

Reproducibility	Information Provided About
<i>Computational</i>	... code, software, hardware and implementation details
<i>Empirical</i>	... non-computational empirical scientific experiments and observations. Make data freely available, as well as details of how the data was collected.
<i>Statistical</i>	... the choice of statistical tests, model parameters, threshold values, etc.

How can software tools make our research more reproducible?

There are several software tools that help make research more reproducible:

- Version control
 - These tools enable you to keep a record of file changes over time, so specific versions can be recalled later
 - [Github](#)
- Literate programming
 - These tools enable writing and publishing self-contained documents that include narrative and code used to generate both text and graphical results.
 - [knitr](#) in [RStudio](#)

Version Control

What is Version Control?

A system that records changes to a file or set of files over time so that you can recall specific versions later.

Knowledge Check

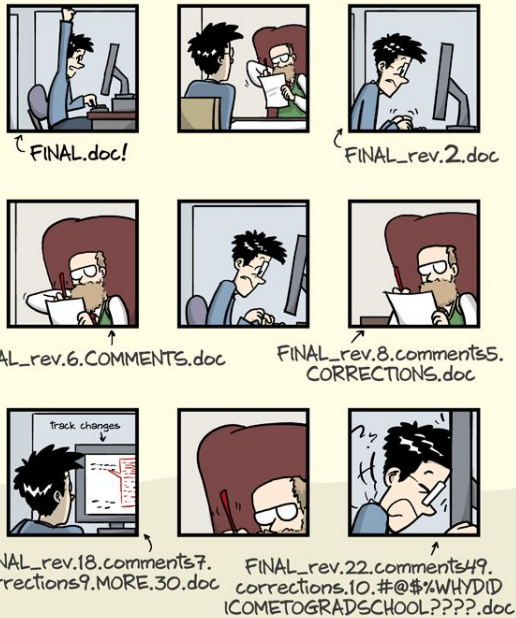
- ▶ Navigate to Socrative.com
- ▶ Click “Login” in the top right-hand corner
- ▶ Click “Student Login”
- ▶ Enter INST452 for the Room Name and click “JOIN”
- ▶ Enter your full name (LastName, FirstName) and click “DONE”
- ▶ i.e. Sigalo, Nekabari

Question:

How do you keep track of changes to your code/work/analyses?

Bad!

"FINAL".doc



Better?

```
2023-10-14_manuscript.doc
2023-10-30_manuscript.doc
2023-11-05_manuscript_intititalRyanEdits.doc
2023-11-10_manuscript.doc
2023-11-11_manuscript.doc
2023-11-15_manuscript.doc
2023-11-30_manuscript.doc
2023-12-01_manuscript.doc
2023-12-02_manuscript_PNASsubmitted.doc
2024-01-03_manuscript_PLOSsubmitted.doc
```

Best!



How does a version control system work?

- Version control systems start with a base version of the document and then save just the changes you made at each step of the way



Literate Programming

What is literate programming?

- A **literate program** is an ordinary computer program that has been cut up and rearranged to make it easier for a human to understand
- Uses a notebook interface to weave together narrative text and code to produce nicely formatted output

What is an RMarkdown File?

A versatile document format used for literate programming that combines the power of R programming language with the simplicity of Markdown formatting. Here's what it entails:

1. Markdown Syntax
2. Embedded R Code
3. Dynamic Output
4. Multiple Output Formats
5. **Reproducibility**

```
Getting_Started_RMarkdown_Blog... * x  Untitled.Rmd x
← →  ABC  🔍  Knit  ⚙️
1  ---
2  title: "Untitled"
3  author: "Dataquest"
4  date: "July 2020"
5  output: html_document
6  ---
7
8  ```{r setup, include=FALSE}
9  knitr::opts_chunk$set(echo = TRUE)
10 ```
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting
15 using R Markdown see <http://rmarkdown.rstudio.com>.
16
17 When you click the Knit button a document will be generated
18 within the document. You can embed an R code chunk like this:
19
20 ```{r cars}
21 summary(cars)
22 ```
23
24 ## Including Plots
25
26 You can also embed plots, for example:
27
28 ```{r pressure, echo=FALSE}
29 plot(pressure)
30 ```
31
32 Note that the `echo = FALSE` parameter was added to the code chunk
33
34 6:4  # Untitled
```

Knit Button

Untitled.html Open in Browser Find

Untitled

Dataquest
July 2020

R Markdown

This is an R Markdown document. Markdown is a simple formatting on using R Markdown see <http://rmarkdown.rstudio.com>.


When you click the **Knit** button a document will be generated that in within the document. You can embed an R code chunk like this:

```
summary(cars)
```

	speed	dist
## Min.	: 4.0	Min. : 2.00
## 1st Qu.:	12.0	1st Qu.: 26.00
## Median	:15.0	Median : 36.00
## Mean	:15.4	Mean : 42.98
## 3rd Qu.:	19.0	3rd Qu.: 56.00
## Max.	:25.0	Max. :120.00

Including Plots

You can also embed plots, for example:



Knowledge Check

- ▶ Navigate to Socrative.com
- ▶ Click “Login” in the top right-hand corner
- ▶ Click “Student Login”
- ▶ Enter INST452 for the Room Name and click “JOIN”
- ▶ Enter your full name (LastName, FirstName) and click “DONE”
- ▶ i.e. Sigalo, Nekabari



10 MINUTE BREAK

Live Coding - Getting Started with R Markdown

In-Class Exercise:

Literate Programming