

Rotman

**TD
Management
Data and
Analytics Lab**



Rotman School of Management
UNIVERSITY OF TORONTO

What is Git?

Git is an active, open source, version control system

-developed by Linus Torvalds

GitHub is a hosting service for Git

-subsidiary of Microsoft since 2018

-200 million repositories

-83 million developers

Name	2015	2017	2018	2022
Git	69.3%	69.2%	87.2%	93.9%
Subversion	36.9%	9.1%	16.1%	5.2%
TFVC	12.2%	7.3%	10.9%	[ii]
Mercurial	7.9%	1.9%	3.6%	1.1%
CVS	4.2%	[ii]	[ii]	[ii]
Perforce	3.3%	[ii]	[ii]	[ii]
VSS	[ii]	0.6%	[ii]	[ii]
ClearCase	[ii]	0.4%	[ii]	[ii]
Zip file backups	[ii]	2.0%	7.9%	[ii]
Raw network sharing	[ii]	1.7%	7.9%	[ii]
Other	5.8%	3.0%	[ii]	[ii]
None	9.3%	4.8%	4.8%	4.3%

Why Use GitHub? Or “Mistake I Have Made”

1. Code Sharing is Now Standard Practice at Journals
2. Different Filenames Break Dependencies
3. You Can Explore New Ideas Without Fear
4. You Can Fix Bugs (and other people can too!)

Code Sharing is Now Standard Practice (and Good Practice!)

Management Science Policy for Data and Code Disclosure

Effective June 1, 2019

Download [Data and Code Disclosure Form](#)

A fundamental principle of the scientific method is replication: the validity of a research finding requires that it can be reproduced by other researchers. The intent of the Data and Code Disclosure policy is to assure the availability of the material necessary to replicate the research published in the journal.

A secondary benefit of this policy is to advance the research in the fields covered by the journal. Inevitably, the sharing of data and codes will be of value to the relevant research community, allowing them to leverage this prior work in their own pursuits. This sharing should increase the rate of scientific progress and impact.

So you will need some platform to distribute your code. Start now, not later. This isn't just "red-tape" to address replication. It increases your impact.

“Filename Method” Break Dependencies

- artificial_data_experiment_2.nb.html
- artificial_data_experiment_2.Rmd
- artificial_data_experiment_3.nb.html
- artificial_data_experiment_3.Rmd
- artificial_data_experiment.knit.md
- artificial_data_experiment.nb.html
- artificial_data_experiment.Rmd
- artificial_data_experiment.utf8.md
- backup code
- choice_data_debug.R
- choice_data_estimation_V1.R
- choice_data_estimation_V2.R
- choice_data_estimation_V5.R
- choice_data_estimation_V6.R
- choice_data_estimation_V7.R

“Filename Method” Break Dependencies

The screenshot displays a file explorer interface with three main sections:

- Left Panel:** A list of files and folders. Each item has a small icon to its left and a green checkmark to its right. The items are:
 - artificial_data_experiment_2.nb.html
 - artificial_data_experiment_2.Rmd
 - artificial_data_experiment_3.nb.html
 - artificial_data_experiment_3.Rmd
 - artificial_data_experiment.knit.md
 - artificial_data_experiment.nb.html
 - artificial_data_experiment.Rmd
 - artificial_data_experiment.utf8.md
 - backup code (highlighted)
 - choice_data_debug.R
 - choice_data_estimation_V1.R
 - choice_data_estimation_V2.R
 - choice_data_estimation_V5.R
 - choice data estimation V6.R
- Middle Panel:** A blue header bar containing the text "April 1 2021" and a right-pointing arrow.
- Right Panel:** A list of files, each with a small icon to its left:
 - choice_data_estimation_V5.R
 - likelihood_functions.cpp
 - likelihood_functions.R

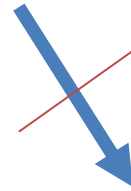
“Filename Method” Breaks Dependencies

Using Filename Method



simulate_choices.R

calc_loglikelihood.R



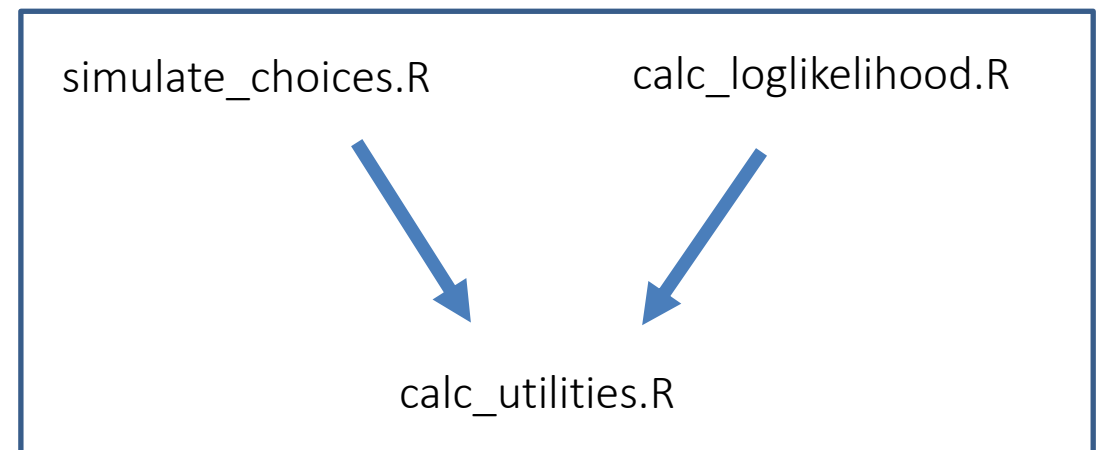
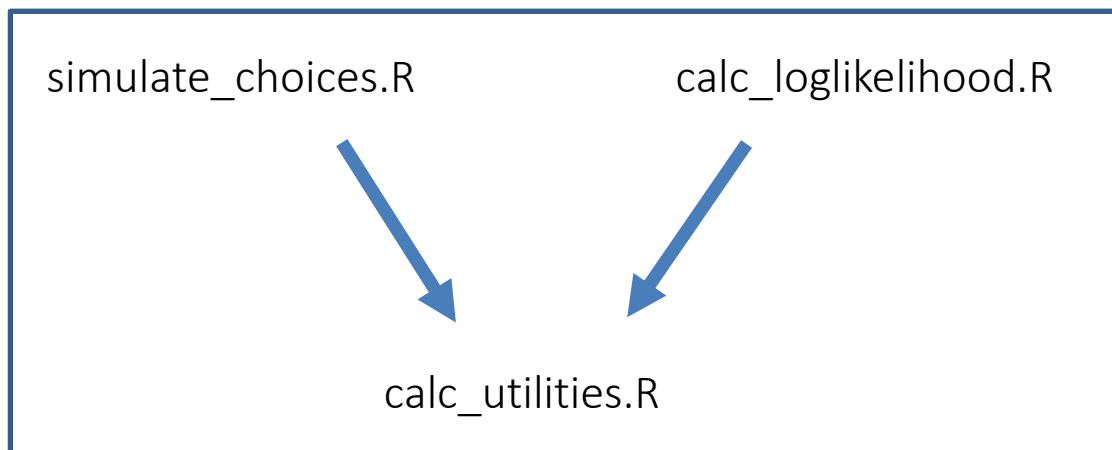
calc_utilities.R

Sometimes will give an error
but can lead to “hidden” bugs

Using Git

Branch 1

Branch 2




You Can Explore New Ideas Without Fear

Default branch

`main`  Updated 4 months ago by matthew-osborne

Stale branches

`last_period_doubt_clean_code`  Updated 2 years ago by matthew-osborne

`contemporaneous_doubt_stock`  Updated 2 years ago by matthew-osborne

`random_coefficients`  Updated 17 months ago by matthew-osborne

You Can Fix Bugs (and others can too!)

DrugowitschLab / dm Public Edit Pins Watch 3 Fork 7 Starred 14

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#)

master 1 branch 5 tags Go to file Add file Code

jdruugo Merge pull request #6 from UofT-Neuroecon-1/master 24902cd on Dec 14, 2020 52 commits

matlab	Fix bug with function call in matlab	2 years ago
python	Added Python interface for diffusion model sampling	8 years ago
src	Fixes #5	3 years ago
LICENSE	Licensed code under New BSD License	9 years ago
README.md	Update README.md to link to the correct papers	7 years ago

About
C++ Diffusion model toolset with Python and Matlab interfaces

- Readme
- View license
- 14 stars
- 3 watching
- 7 forks

Why Use GitHub? Or “Mistakes I Have Made”

1. Code Sharing is Now Standard Practice at Journals
2. Different Filenames Break Dependencies
3. You Can Explore New Ideas Without Fear
4. You Can Fix Bugs (and other people can too!)

Version control matters most at the END of a project
- it is an investment, but one that will pay off

TD MDAL Research Roundtables

Dates:

Thursday, January 26, 2023 – Finance theme

Tuesday, March 21, 2023 – Mixed research theme

Highlight TD MDAL research grant recipients and their projects in the areas of data analytics, machine learning and artificial intelligence.