STAT 209 Version Control

June 15, 2021

Colin Reimer Dawson

1/15

What is Version Control?

A version control system...

- records the history of changes to your code
- facilitates collaboration by propagating changes among users
- enables reversion to an earlier project state

Why Version Control?

A STORY TOLD IN FILE NAMES:			
Location: 😂 C:\user\research\data			~
Filename 🔺	Date Modified	Size	Туре
関 data_2010.05.28_test.dat	3:37 PM 5/28/2010	420 KB	DAT file
🚦 data_2010.05.28_re-test.dat	4:29 PM 5/28/2010	421 KB	DAT file
🔀 data_2010.05.28_re-re-test.dat	5:43 PM 5/28/2010	420 KB	DAT file
🚦 data_2010.05.28_calibrate.dat	7:17 PM 5/28/2010	1,256 KB	DAT file
🔋 data_2010.05.28_huh??.dat	7:20 PM 5/28/2010	30 KB	DAT file
🚦 data_2010.05.28_WTF.dat	9:58 PM 5/28/2010	30 KB	DAT file
🚦 data_2010.05.29_aaarrrgh.dat	12:37 AM 5/29/2010	30 KB	DAT file
🚦 data_2010.05.29_#\$@*&!!.dat	2:40 AM 5/29/2010	0 KB	DAT file
🔋 data_2010.05.29_crap.dat	3:22 AM 5/29/2010	437 KB	DAT file
🚦 data_2010.05.29_notbad.dat	4:16 AM 5/29/2010	670 KB	DAT file
🚦 data_2010.05.29_woohoo!!.dat	4:47 AM 5/29/2010	1,349 KB	DAT file
🚦 data_2010.05.29_USETHISONE.dat	5:08 AM 5/29/2010	2,894 KB	DAT file
🕙 analysis_graphs.xls	7:13 AM 5/29/2010	455 KB	XLS file
ThesisOutline!.doc	7:26 AM 5/29/2010	38 KB	DOC file
Notes_Meeting_with_ProfSmith.txt	11:38 AM 5/29/2010	1,673 KB	TXT file
🗀 JUNK	2:45 PM 5/29/2010		Folder
🚦 data_2010.05.30_startingover.dat	8:37 AM 5/30/2010	420 KB	DAT file
<			>
Type: Ph.D Thesis Modified: too many times	Copyright: Jorge Cham	www.phde	comics.com 🦽

3/15

Version Control Systems



- Concurrent Versions Systems (CVS) since 1986
- Subversion (svn) since 2000
- Git since 2005

Version Control Systems



- Concurrent Versions Systems (CVS) since 1986
- Subversion (svn) since 2000
- Git since 2005
- Various others

How does it work?



Key Concept: Snapshots

- A record of what the files in your project look like at any given time
- You decide when to take a snapshot (called "committing" or "making a commit")
- The version control system allows you to revisit earlier snapshots

Key Concept: Commit

- As a verb, to create a new snapshot
 - "I committed my code"
- As a noun, the update from one snapshot to the next
 - "I made a commit"

What's in a Commit?

- 1. A reference to the previous commit ("parent commit")
- 2. A record of changes since the last commit
- A unique "hash code" identifier (a long string of letters and numbers like 2ff78d5ebaf48f43f7de26d1bcae52714fa23549)

Key Concept: Repository

- A collection of files in a project, along with a version history of those files
 - Consists of all commits
- "repo" for short

Centralized version control



Git is Distributed

- In a distributed model, each user possesses a full (local) repository
- Changes can be propagated or **pushed** from a local repo to a **remote** repo on a server ("in the cloud")
- Changes can be fetched or **pulled** from the remote repo to the local repo
- GitHub is one place where git repos can be hosted remotely

Distributed version control



The Main Git Verbs

clone	Copy a remote repo to your computer
pull	"Download" changes from the remote repo to
	your local repo
add	Register changes to some files to be commit-
	ted at the next commit
commit	Take a snapshot of your working directory
	and register the state in your local repo
push	"Upload" new commits from your local repo
	to the remote repo

Git in RStudio

- RStudio is integrated with Git and provides a graphical interface for it
- You can use the command line if you prefer, by opening a Shell in RStudio, but I will guide you through the graphical approach
- Interactive tutorials for the standard CLI at
 - try.github.io
 - DataCamp
 - Numerous other places