Introduction To Version Control

Ben Schmidt

2019-01-17

To manage problem sets and assignments, we'll be using a piece of software called "git." This is software for managing different versions of files: it is most oftened used for code, but many cultural heritage organizations also use it to manage data.

For example, the foreign relations of the United States diplomatic files are kept on github at https://github.com/HistoryAtState/frus.

Git

Git does two things.

1. It maintains a history of changes to any files. Ordinarily, if you change a line in a file of data, there is no record of it. A git repository stores your changes–but it also keeps the *old versions* in a format so that you can always reset to any other saved version.
2. It maintains a set of rules for reconciling changes between two different saved versions of a file. If you edit some lines and I edit some other lines, it can merge the two edits together into a new file to ensure that we can keep working together.

Unlike something like Google Docs, which does much of the same things, git works well across whole directory systems and in the context of RStudio. Also unlike Google docs, git asks you to be proactive about when to sync.

(There are some other features that make it especially popular, such as the ability to have two different versions of a project proceed simultaneously, that we won't get into now.)

A **repository** is the name for a git collection. It will take up a folder on your hard drive, and can have many items.

Installing Git

git is a program that you can download and install following instructions here: https://git-scm.com/downloads.

Github

Github is a website–recently bought by Microsoft–that hosts git repositories. You can connect to and download from it using the git program.

This will make it easier to keep track of all the files from the course.

Go to `File: New Project`: select "Create from version control." And then choose "Clone a project from a Git repository."

You will then be asked for a remote location. Choose the one for this site: https://github.com/HumanitiesDataAnalysis/HDA19

You can then use the 'git' tab in the upper right hand corner of RStudio to download all files and assignments for this class. The 'Pull' button will update your local copy.

"Push" would let you upload things to the course web site. If this is necessary, let me know: for this class, it's safe to have things be a one-way street.

Be sure to check for updates **each week** before working on the problems.