

Version Control: What and Why

Andy Choens, MSW

Office of Quality and Patient Safety

May 20, 2015

Outline

Introduction

Informal Version Control

Formal Version Control

Questions?

Dafynitions

- ▶ **Version Control:** *the management of changes to documents, computer programs, large web sites, and other collections of information.¹ It has many names.*
 - ▶ *Revision Control*
 - ▶ *Source Control*
 - ▶ *Renaming Files*
 - ▶ *VC*
- ▶ **Git:** *a distributed revision control system with an emphasis on speed, data integrity, and support for distributed, non-linear workflows.²*
- ▶ **GitLab:** *a web-based Git repository manager with wiki and issue tracking features. GitLab is similar to GitHub, but GitLab has an open source version, unlike GitHub.³*

¹https://en.wikipedia.org/wiki/Revision_control

²[https://en.wikipedia.org/wiki/Git_\(software\)](https://en.wikipedia.org/wiki/Git_(software))

³<https://en.wikipedia.org/wiki/GitLab>

Git and R:

- ▶ VC and git are heavily used by the R community.
- ▶ You can use git for SAS development as easily as R.
- ▶ Or Excel
- ▶ Or a Presentation

After The Show:

Today:

- ▶ What and Why.
- ▶ Not How.
- ▶ Git isn't rocket science. . . or epidemiology.

On Your Own:

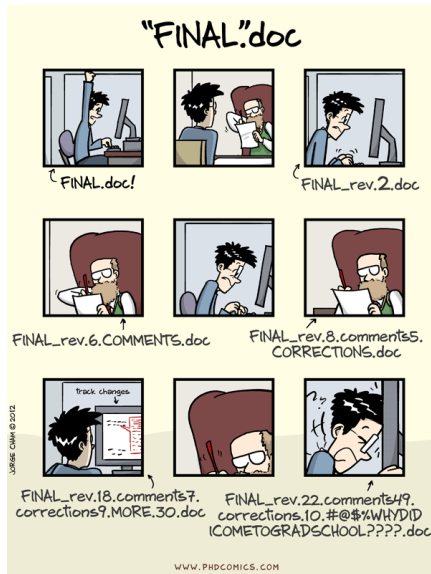
- ▶ <https://msysgit.github.io/>
- ▶ <http://sourceforge.net/projects/gitextensions/>
BLOCKED!
- ▶ <http://doc.gitlab.com/ce/>
- ▶ <https://help.github.com/>

This presentation can be found at:

- ▶ <https://github.com/choens/git-for-research-scientists>

When Version Control Matters

- ▶ **Q:** *Is all this relevant to me?*
- ▶ **A:** Yes.



Informal Version Control: Examples

Woefully incomplete list of examples:

- ▶ File-name spaghetti.
- ▶ Track Changes In Word, Excel, etc.
- ▶ DropBox, SpiderOak
- ▶ Time Machine in OSX
- ▶ Who else has a different USB drive for every day of the week?

Informal Version Control: Examples 2

Advantages

- ▶ Sometimes, it is built in.
 - ▶ Word
 - ▶ SAS
- ▶ Easily implemented.
- ▶ You already know how.

Disadvantages

- ▶ Not equally available.
 - ▶ PowerPoint
 - ▶ Data revisions?
- ▶ Only starts simple.
- ▶ I hope everyone agrees how.
- ▶ Multiple collaborators?
 - ▶ Who changed the code?
 - ▶ When?
 - ▶ Why?
 - ▶ Which version do I have?
 - ▶ Where is last year's code?
 - ▶ What is branching?

Informal Version Control: Fail

Q: Which SAS file generated the graph given to the Medicaid Director last fall?

- ▶ big_project/2014.05.27.run.sas
- ▶ big_project/2014.05.27.re_run.sas
- ▶ big_project/2014.05.27.re_re_run.sas
- ▶ big_project/2014.05.28.use_this_one.sas
- ▶ big_project/2014.05.29.huh.sas
- ▶ big_project/2014.05.30.aaaargh.sas
- ▶ big_project/2014.05.31.not_bad.sas
- ▶ big_project/2014.05.31.woohoo.sas

Welcome To Informal Version Control.

Informal Version Control: FAILS

Have you or a colleague ever

- ▶ Wasted all morning waiting for the shared drive to come back on-line?
- ▶ Been unable to reproduce the *really* important graph given to the Medicaid Director last fall, because the code has been lost (overwritten)?
- ▶ Struggled to maintain two parallel code bases?
- ▶ Been unsure if the code you are using is the most current version? If it isn't, what are the changes and where are they in the code?

Formal Version Control: The Answer

Gives users tools, practices and community norms that enable more effective collaboration and communication.

Formal Version Control: Wikipedia Example

- ▶ Version Control can be found in lots of places, like Wikipedia⁵.
- ▶ The entire history of EVERY Wikipedia Article is available.
- ▶ Wikipedians place a high priority in understanding how their product changes over time.
- ▶ They have a common system for sharing, tracking and using changes to their product.
- ▶ There is not reason we can't have the same history for the work we do.

⁵https://en.wikipedia.org/wiki/Help:Page_history

Formal Version Control: WINS

Have you or a colleague ever:

- ▶ Wasted all morning waiting for the shared drive to come back on-line?
- ▶ Been unable to reproduce the *really* important graph given to the Medicaid Director last fall, because the code has been lost (overwritten)?
- ▶ Struggled to maintain two parallel code bases?
- ▶ Been unsure if the code you are using is the most current version? If it isn't, what are the changes and where are they?

Let's turn these FAILS into WINS!

Have you or a colleague ever # 1

Fail: Wasted all morning waiting for the shared drive to come back on-line?

Win: Git stores a local copy of the code on YOUR local computer. If the server goes down, you can still work, but you can't collaborate.

Have you or a colleague ever # 2

Fail: Been unable to reproduce the *really* important graph given to the Medicaid Director last fall, because the code has been lost (overwritten)?

Win: Committed changes are a permanent part of the project⁶.

Win: Code state at commit is always reproducible.

Win: Important versions can be tagged⁷.

⁶<http://appdev1/oqps/hh-quarterly-qa-report/commits/master>

⁷<http://appdev1/oqps/hh-quarterly-qa-report/tags>

This Is A Practice, Not A Guarantee

	COMMENT	DATE
○	CREATED MAIN LOOP & TIMING CONTROL	14 HOURS AGO
○	ENABLED CONFIG FILE PARSING	9 HOURS AGO
○	MISC BUGFIXES	5 HOURS AGO
○	CODE ADDITIONS/EDITS	4 HOURS AGO
○	MORE CODE	4 HOURS AGO
○	HERE HAVE CODE	4 HOURS AGO
○	AAAAAAA	3 HOURS AGO
○	ADKFJSLKDFJSDKLFJ	3 HOURS AGO
○	MY HANDS ARE TYPING WORDS	2 HOURS AGO
○	HAAAAAAAAAANDS	2 HOURS AGO

AS A PROJECT DRAGS ON, MY GIT COMMIT MESSAGES GET LESS AND LESS INFORMATIVE.

Have you or a colleague ever # 3

Fail: Struggled to maintain two parallel code bases? (Vertica?)

Win: Branches: You can maintain many parallel code bases easily⁹.

⁹<http://appdev1/axc38/hp-cmart/network/master>

Have you or a colleague ever # 4

Fail: Been unsure if the code you are using is the most current version? If it isn't, what are the changes and where are they?

Win: Git repos contain the full history & status of a project¹⁰.

Win: Tools to sync the local copy to the remote master. (git pull)

¹⁰http://appdev1/oqps/hh-quarterly-qa-report/merge_requests

Git + Gitlab == Project Management

- ▶ Central place to share scientific code and projects
- ▶ 1 repo == 1 project
- ▶ Tracks changes in files over time¹¹
- ▶ Issues / Bug Tracking¹²
- ▶ Tools for collaboration ¹³

¹¹<http://appdev1/oqps/hh-quarterly-qa-report/commits/master>

¹²<http://appdev1/oqps/hh-quarterly-qa-report/issues>

¹³<http://appdev1/care-management/hp-cmart-reports/issues/1>

What Happens Next?

- ▶ This is a different way of working, but anyone can learn it.
- ▶ I am developing some formal training materials with EBCOP.
- ▶ Don't forget the links from the beginning of the presentation.
- ▶ File an ITS ticket to get MSYS Git and Git Extensions.
- ▶ For GitLab: Contact Jeremy Redmond and Avidah Sadaghiani.
- ▶ Questions? Post it to the EBCOP blog and we can discuss.

Questions?

QUESTIONS NOT EVEN 5+ YEARS OF GRAD SCHOOL WILL HELP YOU ANSWER



PHD IN ENVIRONMENTAL ENGINEERING



PHD IN PHYSICS



PHD IN BIOLOGY



PHD IN MECHANICAL ENGINEERING



PHD IN POLITICAL SCIENCE

www.phdcomics.com