Version control and reproducible research

Adapted from materials provided by Devin Silvia

WHY ARE WE HERE?

WHY ARE WE HERE?

- Scientific research is becoming increasingly dependent on computers
- Even outside of computational and data science, a great deal of research and other types of work depend on writing and running code
- Poorly written, undocumented code without a version history is **not useful** and makes *reproducibility* **difficult**

HOW DO WE FIX THIS?

- Write quality code
- Use version control! (e.g. git, mercurial)

WRITING QUALITY CODE

BECAUSE YOU REALLY SHOULD.

Discussion: what does it mean to **write quality code**?

"Always code as if the person who ends up maintaining your code will be a violent psychopath who knows where you live."

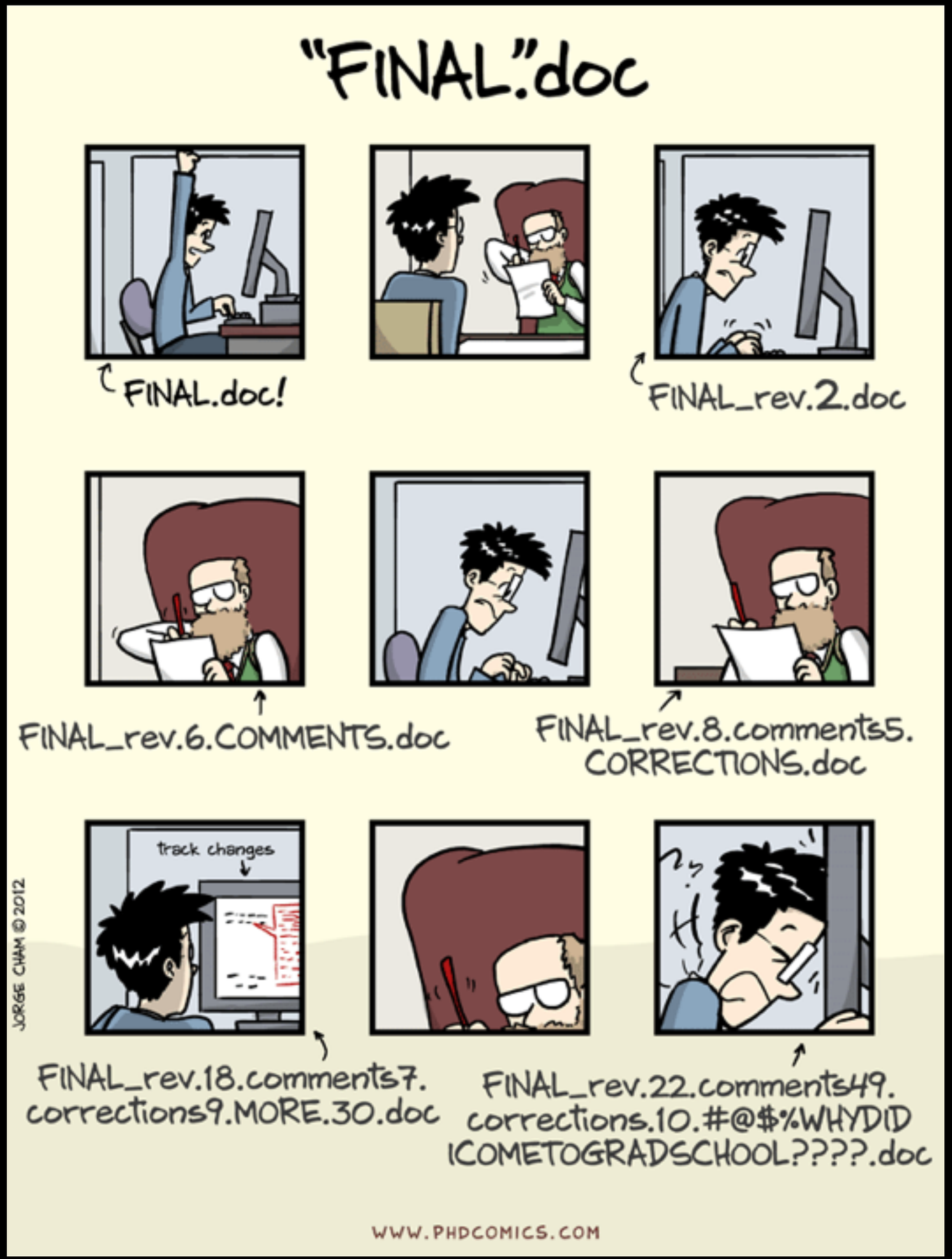
–JOHN F. WOODS (C++ PROGRAMMER)

NEVER LOSE CODE AGAIN WITH...

VERSION CONTROL

WHAT IS IT?

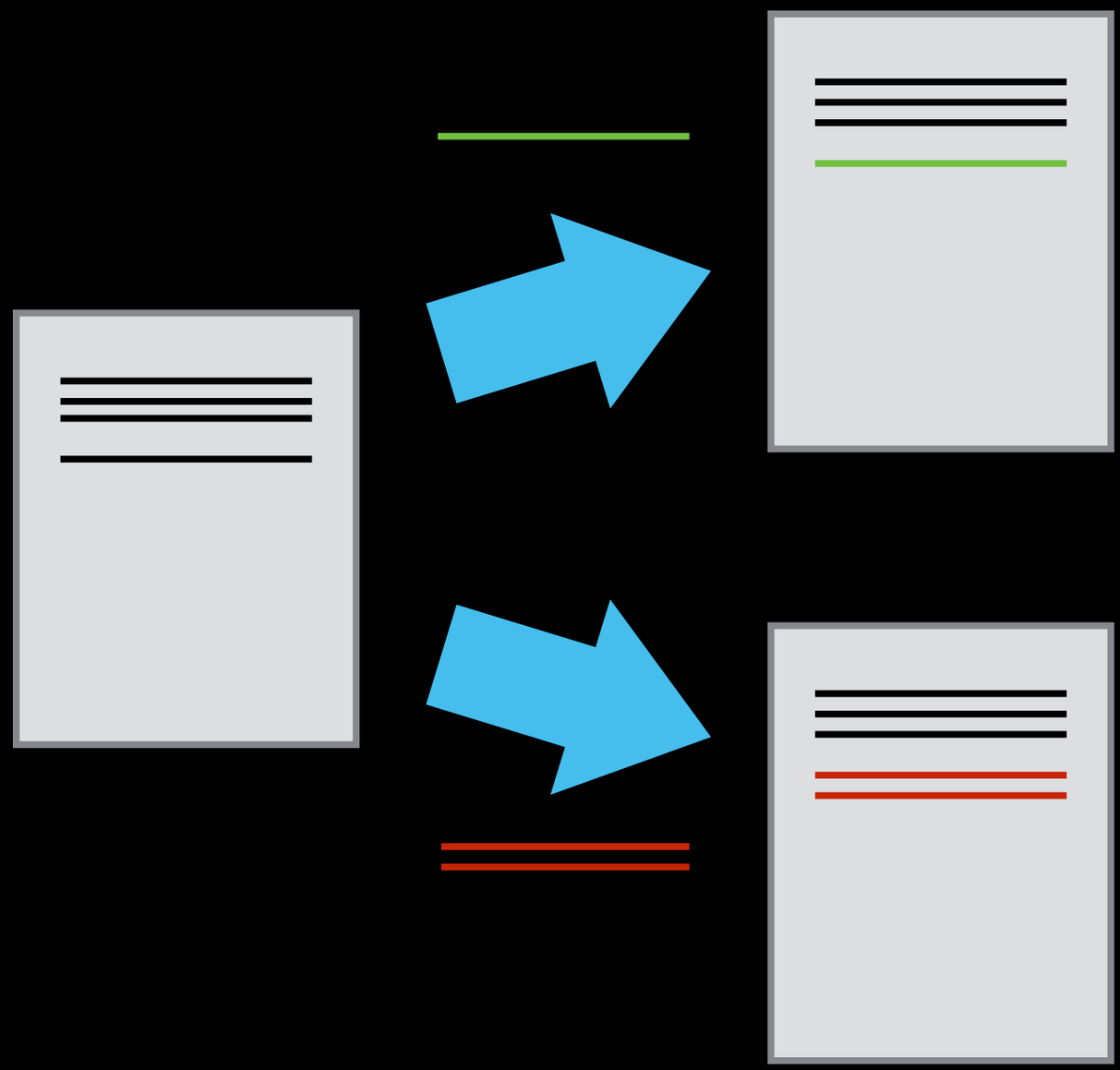
It's not this:



WHAT IS IT?

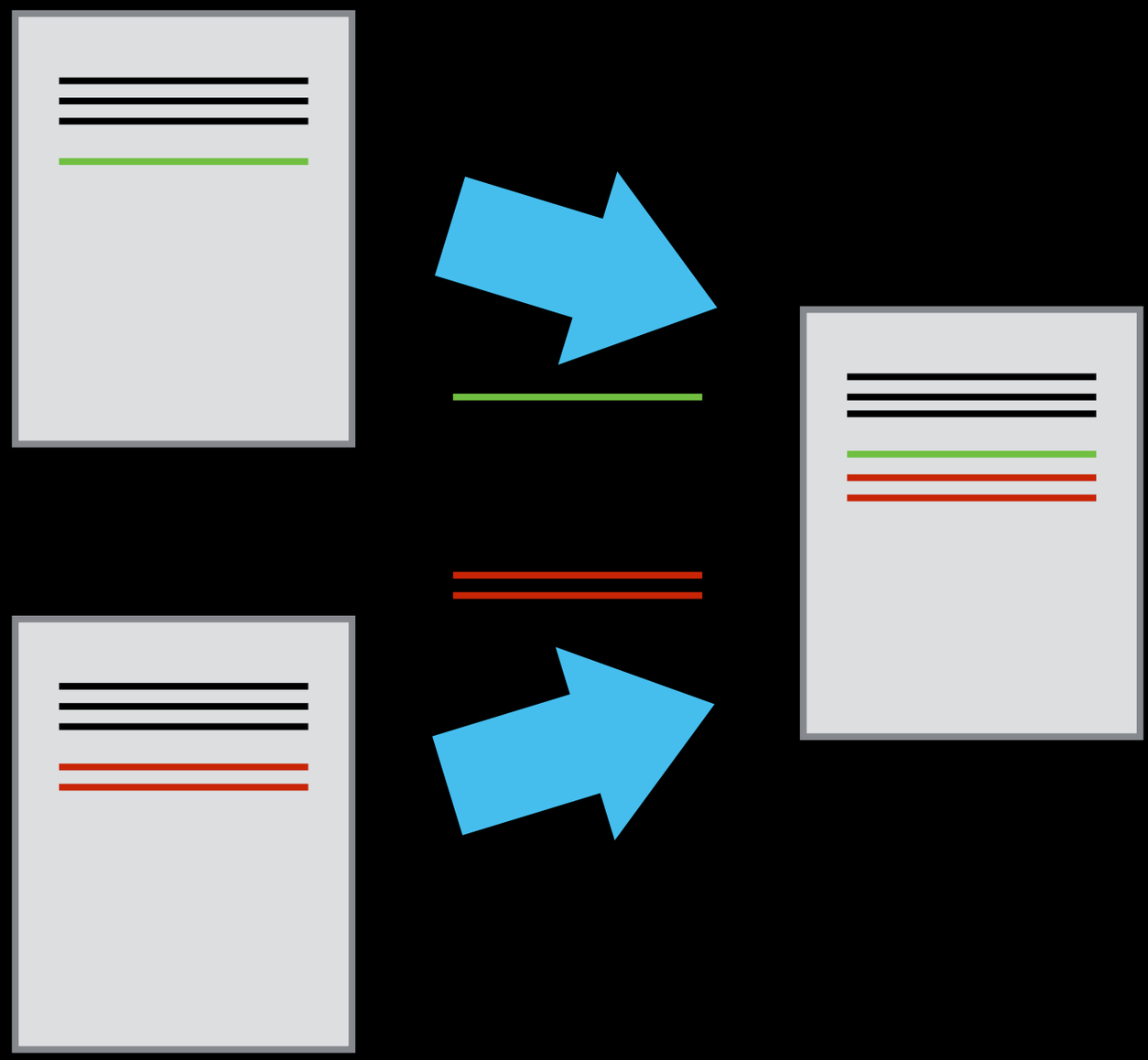
Automated version tracking.
Every change is a new commit.

WHAT IS IT?



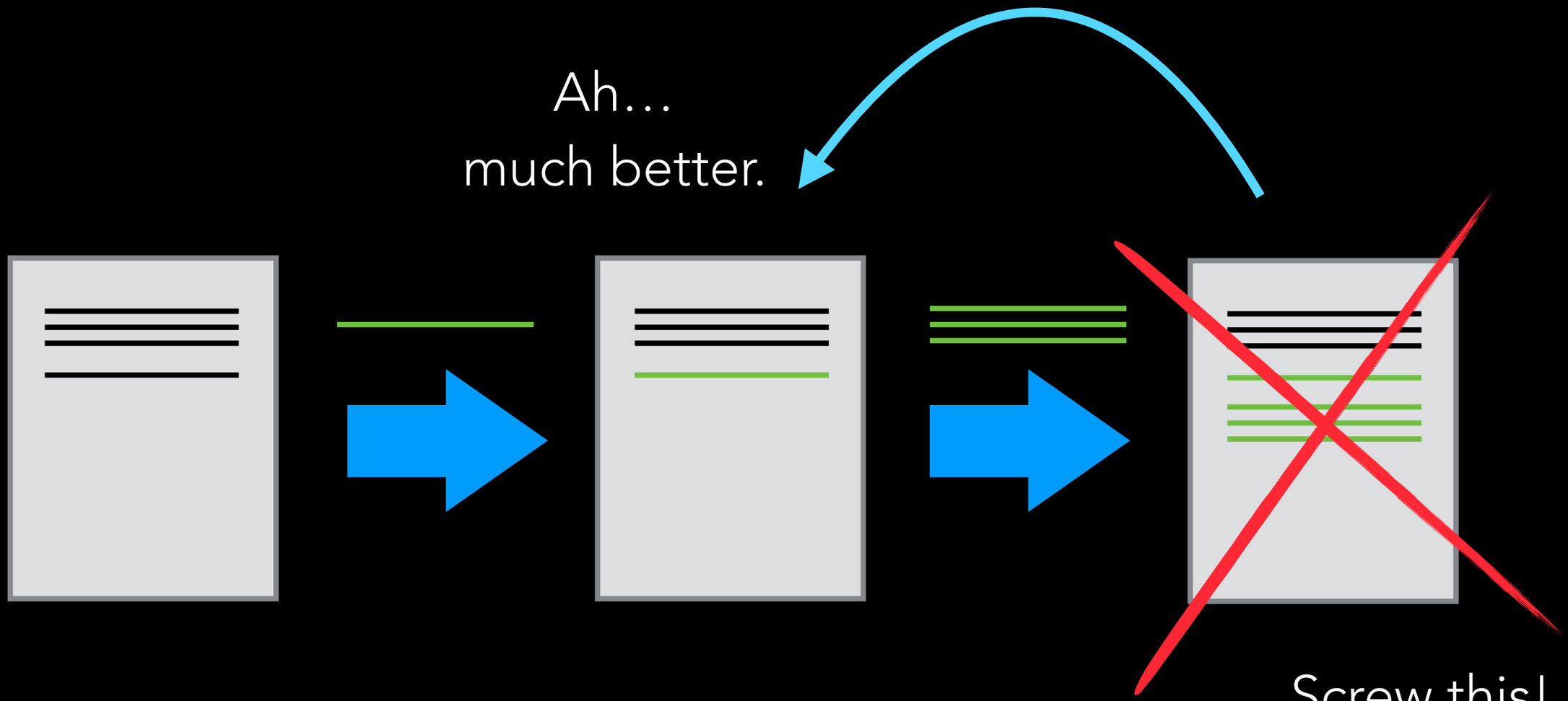
Different people can have different versions.

WHAT IS IT?



Assuming there are no direct conflicts,
you can merge changes.

WHAT IS IT?



And if something goes wrong?
Jump back to a previous version.

Discussion 1: under what circumstances might you use version control in your own work?

Discussion 2: how might you use version control to make your work more reproducible?

LET'S GIVE IT A GO

THINGS ARE ABOUT TO GET INTERACTIVE...

$ git clone https://github.com/bwoshea/cmse_bootcamp_git_tutorial_Fall2017

NOW, TAKE SOME OF YOUR OWN CODE AND SET UP A REPOSITORY