

# Git for statisticians

*Matt Mulvahill CU Anschutz School of Medicine Dept. of Pediatrics*

*2017-04-25*

## “The lab notebook of the digital world” -Software Carpentry

### Who What Why How

*Topics to cover* - key git commands - Git is the lab notebook - Github is the “dropbox” - 5 key commands - .gitignore - SSH keys - where to put data. HPAA issues,

### Who is version control for?

#### You!

- For everyone - although the details

### Git User’s Survey 2016

- Software developers (81%)
- Researchers (12%)
- Analysts (6%)

### What is version control?

**Version control system (VCS):** software for documenting and tracking changes to files.

- git

Term	Description
git	software/tool for creating and managing repositories
repository	a folder with files tracked by git
remote	an external location where the repository is copied

### Git is currently top version control system

- Largely due to being a ‘distributed’ VCS
- Ubiquitous in software development
- Increasingly seen as a key tool in conducting reproducible statistical research.



Figure 1:

## Git locally

### Basic git basics

In this talk we introduce the basics of - using git through the command line and a GUI (Rstudio or GitKraken)  
- the difference between a git repository and a remote (such as GitHub) - implementation issues specific to statistical analyses.



### .gitignore

## References & Resources

- Git for Humans! by Alice Bartlette
- line endings in git
- Happy Git with R by Jenny Bryan
- Git cheatsheet - all functions by the domains they operate across
- What is Version Control? by Atlassian (BitBucket)
- Large file storage with Git