Version control with git and GitHub

#### Karl Broman

Biostatistics & Medical Informatics, UW-Madison

kbroman.org github.com/kbroman @kwbroman Course web: kbroman.org/AdvData

Slides prepared with Sam Younkin

#### "FINAL".doc















FINAL\_rev.6.COMMENTS.doc

FINAL\_rev.8.comments5. CORRECTIONS.doc





FINAL\_rev.18.comments7. FINAL\_rev.22.comments49. corrections9.MORE.30.doc corrections.10.#@\$XWHYDD ICOMETOGRADSCHOOL????.doc

WWW. PHDCOMICS. COM

http://www.phdcomics.com/comics/archive.php?comicid=1531

## Methods for tracking versions

- Don't keep track
- Save numbered zip files
- ► Formal version control

# Suppose it stops working...

### Don't keep track

- good luck!

#### Save numbered zip files

- Unzip versions and diff

#### Formal version control

- Easy to study changes back in time
- Easy to jump back and test

### Why use formal version control?

- History of changes
- Able to go back
- No worries about breaking things that work
- ► Merging changes from multiple people

### Example repository

#### PUBLIC 📗 kbroman / Talk\_MAGIC

@ Unwatch - 1 ★ Star 0 ₽ Fork 0

Talk for MAGIC workshop in Cambridge, UK, 12 June 2013 - Edit

🕻 🕼 branch: m	aster - Talk_MAGIC / D		D D d Desueste
an atha aireadifa tha	and the descent a shall be the Descellar		11 million readness
khromen sufficie	public domain sum in the Headewe	latast commit f1777af192	00 Wiki
Figs	Add crazy table from preCC paper	4 months ago	4- Pulse
Perl	Add lines_of_code_by_version.csv to repository	4 months ago	di Graphs
R R	Another fix regarding map expansion in 8-way RIL by selfing at $k\!\!=\!\!0$	4 months ago	17 Network
gitignore	Add lines_of_code_by_version.csv to repository	4 months ago	- mundh
Makefile	Revise Readme to link to version for web	4 months ago	X Settings
ReadMe.md	Greatly simplify the public domain stuff in the ReadMe	15 days ago	
magic.tex	Fix two slight bugs in slides:	4 months ago	https://github.com/
ReadMe.md			You can done with HTTPS, SSH, or Subversion. @
			Clone in Desktop
Talk fo	r MAGIC Workshop in Cambrido	ie. UK	Ownload ZIP
		,-,	

To the extent possible under law, Karl Broman has waived all copyright and related or neighboring rights to "MAGIC design and other topics". This work is published from: United States.



# Example repository

	Pueuc La kbroman / Talk_MAGIC	@ Unwatch + 1 ★ Star 0 D Fork 0
	Talk for MAGIC workshop in Cambridge, UK, 12 June 2013 - Edit	
	○ 97 commits // 1 branch	1 contributor      ① Issues
Greatly simplify the	public domain stuff in the ReadMe	
kbroman authore	ed 15 days ago	latest commit f1777ef192 🔂
Figs	Add crazy table from preCC paper	4 months ago
Perl	Add lines_of_code_by_version.csv to repository	4 months ago
🖿 R	Another fix regarding map expansion in 8-way RIL by	selfing at k=0 4 months ago
.gitignore	Add lines_of_code_by_version.csv to repository	4 months ago
Makefile	Revise Readme to link to version for web	4 months ago
ReadMe.md	Greatly simplify the public domain stuff in the ReadMe	15 days ago
magic.tex	Fix two slight bugs in slides:	4 months ago

rights to "MAGIC design and other topics". This work is published from: United States.



### Example history

branch: master - Talk_MAGIC / Commits		
Sep 27, 2013		
Greatly simplify the public domain stuff in the ReadMe kbroman suffored 16 days ago		f1777ef192 (*) Browse code
Fix url in ReadMe.md file kbroman suthored 15 days ago		o6515023f9 (+ Browse code
Jun 17, 2013		
Another fix regarding map expansion in 8-way RIL by selfing at k=0 kbroman suffored 4 months ago		28os482f2c (+ Browse code
Fix two slight bugs in slides:	19-	S1d4oa9ceb (*) Browse code
<ul> <li>Slight repair to definition of 3-pt coincidence kbromen subored 4 months ago</li> </ul>		
Jun 10, 2013		
Change one page number		e0e0608515 •
Manual and a manual allocation of the		Browse code
Add missing paren kbroman authored 4 months ago		Browse code f4975dee6e (* Browse code
Add missing paren     Addroid a whore a drorthe ago     Addroid a who is     Atrona authoral 4 months ago     Atrona authoral 4 months ago		Browse code f4975dee6e • Browse code 886f20f008 • Browse code
Add missing permin Add has a horized in months ago who's a who is tamana a horized in months ago who's a who is tamana a horized in months ago		Browse code 14975dee6e • Browse code 886f20f098 • Browse code e6fbf2f647 • Browse code
Add nearing period Add nearing		Browse code f4975dee6e • Browse code 886f20f008 • Browse code e6fbf2f647 • Browse code 4edf3e8476 • Browse code
		Browse code f4975dee6e • Browse code 886f20f408 • Browse code e6fbf2f647 • Browse code 4edf3e8876 • Browse code 14ebb1eeb5 • Browse code

### Example commit

#### PUBLIC 🛄 kbroman / Talk\_MAGIC

641 641 \item

642 642 No interference { \color{myblue} = 1 }

#### @ Unwatch + 1 ★ Star 0 \$7 Fork 0

ix	two s	light bugs in slides: Browse code	le
- ;	8-way	RIL by selfing: map expansion = 1 at $k=0$	
	Sligh	t renair to definition of 3-nt coincidence	
	Jergn	e repair eo derinieton or o pe contendence	
2 ma	aster		
<u>d</u>	kbrom	an authorad 4 months ago 1 parent e0e0608 commit 51d4aa9ceb104bbf26e0cbe105a5c7f8dc02a8	332
Sho	wing 2	changed files with 5 additions and 3 deletions. Show Diff Stat	its
6 🔳		R/map_expansion_func.R	9
		@@ -25,8 +25,10 @@ mesibA4 <- function(k)	
25	25	**********	
26	26	# Eight-way	
27			
		- d = (((1)(2)))(k-2)	
	28	+neself8 <- function(k) {	
	2.9	+ $if(k=0)$ return(1)	
	30	$+ 4 - (((1)/(2)))^{(k-2)}$	
	31	+)	
30	32		
31	33	mesibX8 <- function(k)	
	34	((14)/(3)) - (((30 + 14*sqrt(5))/(15))) * (((1+sqrt(5))/(4)))^k - (((30 - 14*sqrt(5))/(15))) * (((1	1-s(
2 🔳	•	magic.tex View file @ 5164aa9	9
		00 -636,7 +636,7 00	
\$36	636		
537	637	<pre>\hspace{20mm} {\color{myblue} = \$Pr(rec'n in 23 \</pre>	
538	638	\ \text{rec'n in 12}) /	
\$39		- Pr(\text{rec'n in 12}))\$}	
		+ Pr(\text{rec'n in 23})}\$}	

# What is git?

#### Formal version control system

- Developed by Linus Torvalds (developer of Linux)
  - used to manage the source code for Linux
- Tracks any content (but mostly plain text files)
  - source code
  - data analysis projects
  - manuscripts
  - websites
  - presentations

# Why use git?

#### It's fast

- You don't need access to a server
- Amazingly good at merging simultaneous changes
- ► Everyone's using it

### What is GitHub?

- ► A home for git repositories
- Interface for exploring git repositories
- Real open source
  - immediate, easy access to the code
- ► Like facebook for programmers
- ► Free 2-year "Pro" account for students
  - education.github.com
- ► (Bitbucket.org is an alternative)
  - free private repositories

# Why use GitHub?

It takes care of the server aspects of git

### ► Graphical user interface for git

- Exploring code and its history
- Tracking issues

#### ► Facilitates:

- Learning from others
- Seeing what people are up to
- Contributing to others' code

#### Lowers the barrier to collaboration

"There's a typo in your documentation." vs.
 "Here's a correction for your documentation."

#### Change some files

#### See what you've changed

- git status git diff git log
- Indicate what changes to save git add

### Commit to those changes

git commit

#### Change some files

#### See what you've changed

- git status git diff git log
- Indicate what changes to save git add

#### Commit to those changes git commit

#### Push the changes to GitHub git push

### Change some files

- See what you've changed
  - git status git diff git log
- Indicate what changes to save git add
- ► Commit to those changes git commit
- Push the changes to GitHub git push
- Pull changes from your collaborator git pull

### Change some files

#### See what you've changed

- git status git diff git log
- Indicate what changes to save git add
- ► Commit to those changes git commit
- Push the changes to GitHub git push
- Pull changes from your collaborator
  - git fetch git merge

# Initialize repository

Create (and cd to) a working directory

- For example, ~/Docs/Talks/Graphs
- Initialize it to be a git repository
  - git init
  - Creates subdirectory ~/Docs/Talks/Graphs/.git

```
$ mkdir ~/Docs/Talks/Graphs
$ cd ~/Docs/Talks/Graphs
$ git init
Initialized empty Git repository in ~/Docs/Talks/Graphs/.git/
```

### Produce content

#### ► Create a README.md file

```
## Talk on "How to display data badly"
```

```
These are slides for a talk that I give as often as possible, because it's fun.
```

```
This was inspired by Howard Wainer's article, whose title I
stole: H Wainer (1984) How to display data badly.
American Statistician 38:137-147
```

```
A recent PDF is
[here](
http://www.biostat.wisc.edu/~kbroman/talks/graphs2013.pdf).
```

### Incorporate into repository



► Stage the changes using git add

\$ git add README.md

### Incorporate into repository



► Now commit using git commit

```
$ git commit -m "Initial commit of README.md file"
[master (root-commit) 32c9d01] Initial commit of README.md file
1 file changed, 14 insertions(+)
create mode 100644 README.md
```

- ▶ The -m argument allows one to enter a message
- ▶ Without -m. git will spawn a text editor
- Use a meaningful message
- Message can have multiple lines, but make 1st line an overview

### A few points on commits

- ► Use frequent, small commits
- Don't get out of sync with your collaborators
- Commit the sources, not the derived files (R code not images)
- ► Use a .gitignore file to indicate files to be ignored

```
*~
manuscript.pdf
Figs/*.pdf
.RData
.RHistory
*.Rout
*.aux
*.log
*.out
```

# Using git on an existing project

- ▶ git init
- ► Set up .gitignore file
- git status (did you miss any?)
- ▶ git add . (or name files individually)
- git status (did you miss any?)
- ▶ git commit

### Removing/moving files

For files that are being tracked by git:

Use git rm instead of just rm Use git mv instead of just mv

```
$ git rm myfile
$ git mv myfile newname
$ git mv myfile SubDir/
$ git commit
```

### First use of git

```
$ git config --global user.name "Jane Doe"
$ git config --global user.email "janedoe@wisc.edu"
$ git config --global color.ui true
$ git config --global core.editor emacs
$ git config --global core.excludesfile ~/.gitignore_global
```

## Set up GitHub repository

Get a GitHub account

- Click the "Create a new repo" button
- ► Give it a name and description
- Click the "Create repository" button
- Back at the command line:

git remote add origin https://github.com/username/repo
git push -u origin master

### Set up GitHub repository



# Set up GitHub repository



### Configuration file

#### Part of a .git/config file:

```
[remote "origin"]
    url = https://github.com/kbroman/qtl.git
    fetch = +refs/heads/*:refs/remotes/origin/*
[branch "master"]
    remote = origin
    merge = refs/heads/master
[remote "brian"]
    url = git://github.com/byandell/qtl.git
    fetch = +refs/heads/*:refs/remotes/brian/*
```

### Destroy it and start over

### ► Why?

- You commit something you shouldn't have (large and/or private)
- You are embarrassed by your repository's history
- You can't figure out the mess you've made
- Pick the repository you like and destroy the other one
  - For example, get your local directory in the state you like and destroy everything else

#### Local repository

- If you delete the .git subdirectory, it'll no longer be a git repository

#### GitHub repository

- Go to the settings for the repository and head down to the Danger Zone

# Branching and merging

Use branches to test out new features without breaking the working code.

- git branch devel git branch git checkout devel
- When you're happy with the work, merge it back into your master branch.
  - git checkout master git merge devel

### Issues and pull requests

### Problem with or suggestion for someone's code?

- Point it out as an Issue
- Even better: Provide a fix
  - Fork
  - Clone
  - Modify
  - Commit
  - Push
  - Submit a Pull Request

## Suggest a change to a repo

► Go to the repository:

http://github.com/someone/repo

► Fork the repository Click the "Fork" button

Clone your version of it

git clone https://github.com/username/repo

- Change things locally, git add, git commit
- Push your changes to your GitHub repository git push
- ► Go to *your* GitHub repository
- Click "Pull Requests" and "New pull request"

# Pulling a friend's changes

#### Add a connection

git remote add friend git://github.com/friend/repo

#### If you trust them, just pull the changes

git pull friend master

Alternatively, fetch the changes, test them, and then merge them.

git fetch friend master git branch -a git checkout remotes/friend/master git checkout -b friend git checkout master git merge friend

#### Push them back to your GitHub repo

git push

### Merge conflicts

#### Sometimes after git pull friend master

```
Auto-merging README.md
CONFLICT (content): Merge conflict in README.md
Automatic merge failed; fix conflicts and then commit the result.
```

#### Inside the file you'll see:

<<<<<< HEAD A line in my file. ====== A line in my friend's file >>>>>> 031389f2cd2acde08e32f0beb084b2f7c3257fff

Edit, add, commit, push, submit pull request.

# git/GitHub with RStudio

🗯 RStudio File Edit Code View Plots Session Project Build	Tools Window Help	
000	Import Dataset 🔹 🕨	~/Projects/RStu
Q. •	Install Packages Check for Package Updates	Ð
	Version Control	Commit 企業M
R version 3.0.2 (2013-09-25) "Frisb	Pull Branches Push Branch	
copyright (C) 2013 The K Foundation to	History	
Platform: x86_64-apple-darwin10.8.0 (6	Project Setup	
R is free software and comes with ABSO You are welcome to redistribute it und Type 'license()' or 'licence()' for di	LUTELY NO WARRANT er certain conditi stribution details	(. ions. 5.

Open source means everyone can see my stupid mistakes.

Version control means everyone can see every stupid mistake l've ever made.