

Using Git, GitHub Desktop, and GitHub for novice and experienced users

Yi Liu and Tom Palmer
1st December 2023

Learning outcomes

- Novice users: be able to make and commit into a repo on GitHub
- Experienced users: how to avoid some common pitfalls
- Understand that Git and GitHub can be used in different ways for different projects
- Understand that Git provides certain features and GitHub provides additional features
- Be able to use a graphical Git GUI, e.g., GitHub Desktop, VS Code extension
- Be able to use Git, GitHub Desktop/VS Code and GitHub with confidence
- Be able to competently contribute to other Git repositories

Why Git

"FINAL".doc



FINAL.doc!



FINAL_rev.2.doc



FINAL_rev.6.COMMENTS.doc



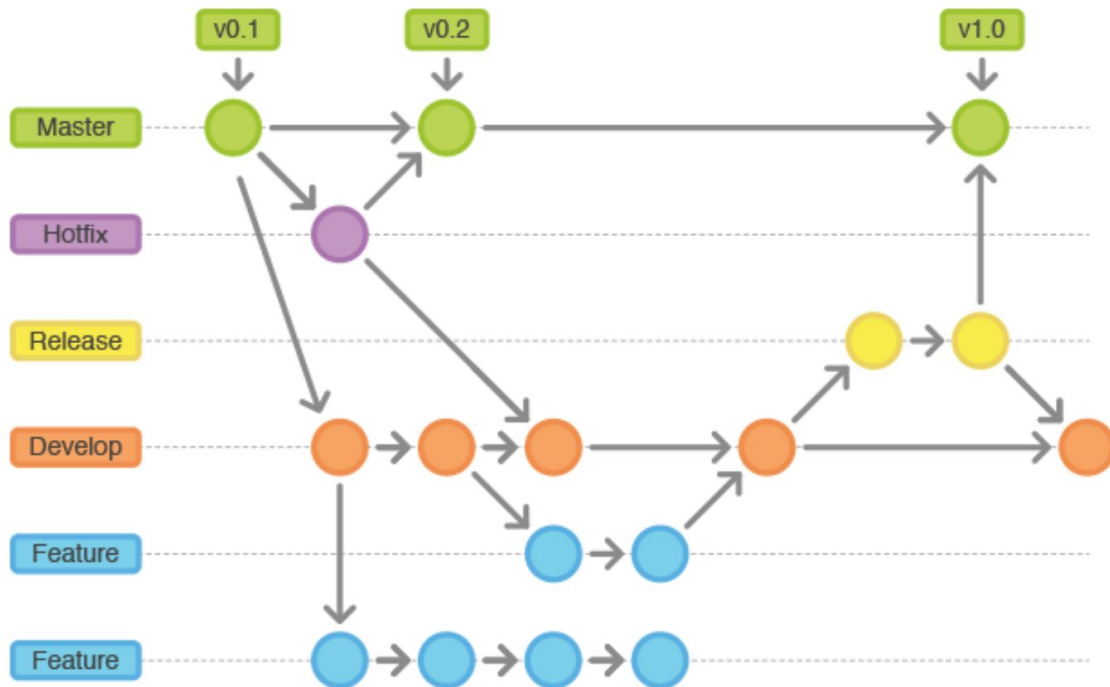
FINAL_rev.8.comments5.
CORRECTIONS.doc



FINAL_rev.18.comments7.
corrections9.MORE.30.doc



FINAL_rev.22.comments49.
corrections.10. #@\$%WHYDID
ICOMETOGRADSCHOOL?????.doc



THIS IS GIT. IT TRACKS COLLABORATIVE WORK
ON PROJECTS THROUGH A BEAUTIFUL
DISTRIBUTED GRAPH THEORY TREE MODEL.

COOL. HOW DO WE USE IT?

NO IDEA. JUST MEMORIZE THESE SHELL
COMMANDS AND TYPE THEM TO SYNC UP.
IF YOU GET ERRORS, SAVE YOUR WORK
ELSEWHERE, DELETE THE PROJECT,
AND DOWNLOAD A FRESH COPY.



Installation links

- GitHub Desktop <https://desktop.github.com/>
- Git for Windows <https://git-scm.com/download/win> , <https://gitforwindows.org/>
- macOS install with homebrew: `brew install git`

3 researcher use cases

- Helena
- Des
- Dom

Helena

- Wants to make code/supplementary material available online for her journal article
- Does not need either Git nor GitHub Desktop installed on computer
- Doesn't need her code folders initialised as Git repos
- Simply drag and drop files from computer into GitHub web interface, commit message: "Add files via upload"
- Recommend turning off GitHub repository features such as Issues, Releases, Discussions, etc (through About pane | settings). If someone wants to contact her they can email



my-test-repo

Public



Pin



Unwatch

1



main ▾



1 branch



0 tags

Go to file

Add file ▾

<> Code ▾



remlapmot Add files via upload

4

Create new file

Upload files



2 commits



README.md

Initial commit

16 minutes ago

Des

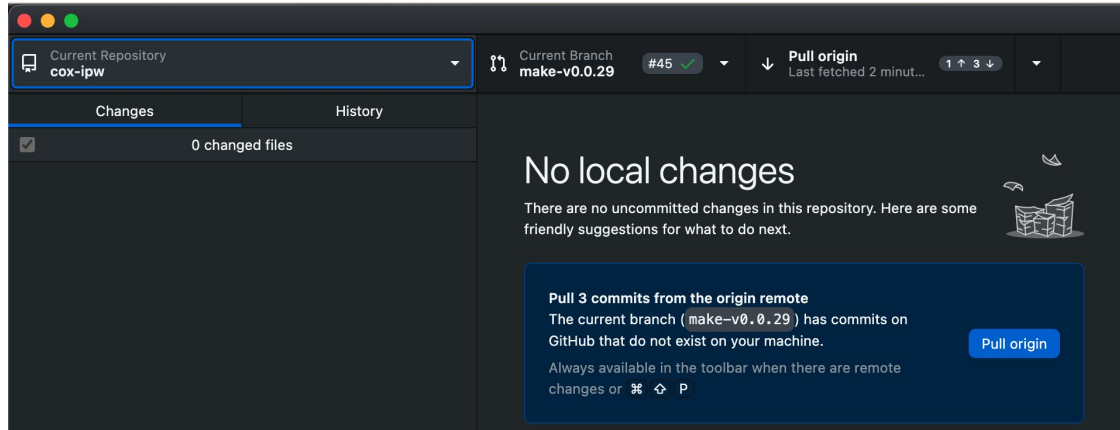
- Coding an R package
- Essentially working alone, occasionally might receive a pull request from a supervisor/interested user
- Needs GitHub Desktop installed (which has a version of Git within it)
- Needs project folders initialised as Git repos
- Needs to take care to write meaningful commit messages
- Does it really matter if he always works on the main branch?
- Question: Why might you want to send a PR to your own repo even if you are the only contributor?
- Helpful to make Releases in the GitHub repo (and/or use Git tags)

Dom

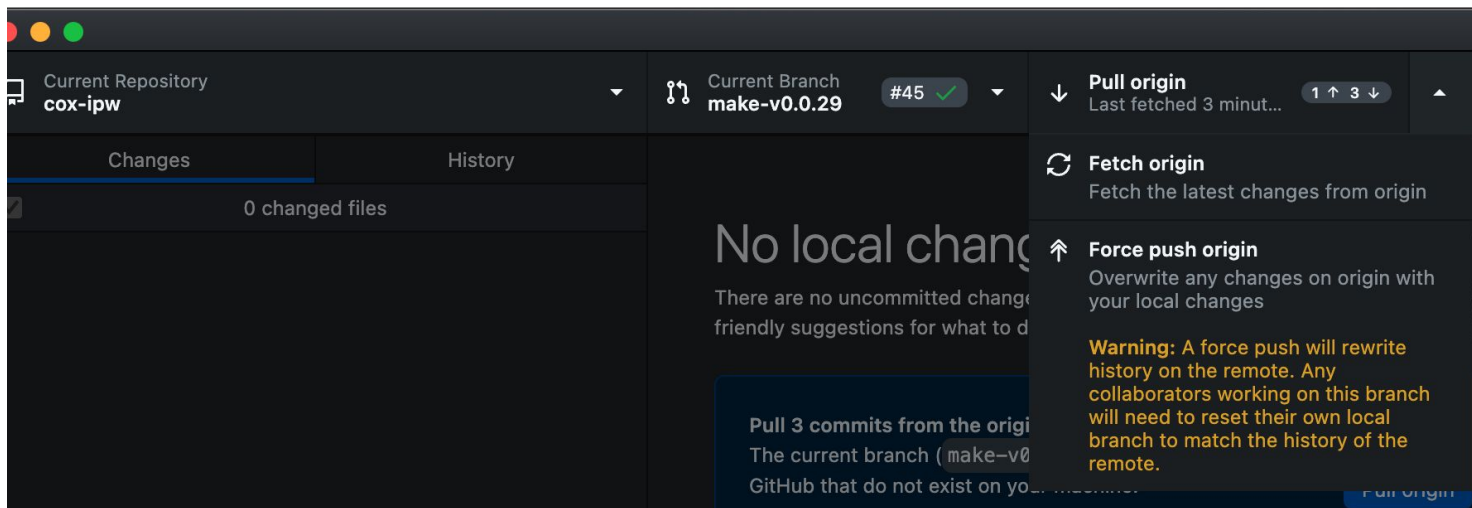
- Coding R packages and additionally wants to send pull requests to other people's projects on GitHub
- Needs both Git and GitHub Desktop installed
- Needs to understand
 - How to fork a repo
 - Never send a PR from his own main/master/default branch - why??
 - Needs to understand how to rebase a branch
 - When is it acceptable/required to force push (i.e., rewrite Git history) a branch??
 - Will need to understand some Git Terminal commands (`git cmdname`)
 - Needs to know how to resolve Git merge conflicts

A scenario

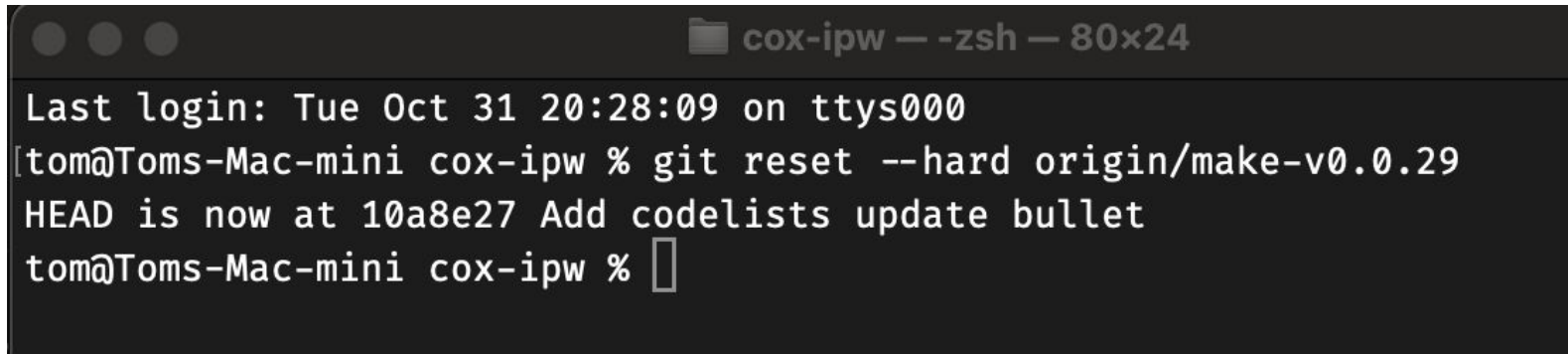
- Dom starts work on his home PC on a new branch (pushes it up to GitHub)
- Goes to work, works on the branch on his work PC but at some point force pushes to the branch (up to GitHub)
- When he goes home, how does he get the updated version of his branch?
- At home in GitHub Desktop in top right box he will see both arrows - be careful when you see these!!



- (For show) Expanding the dropdown arrow - in this box **do not** select either option

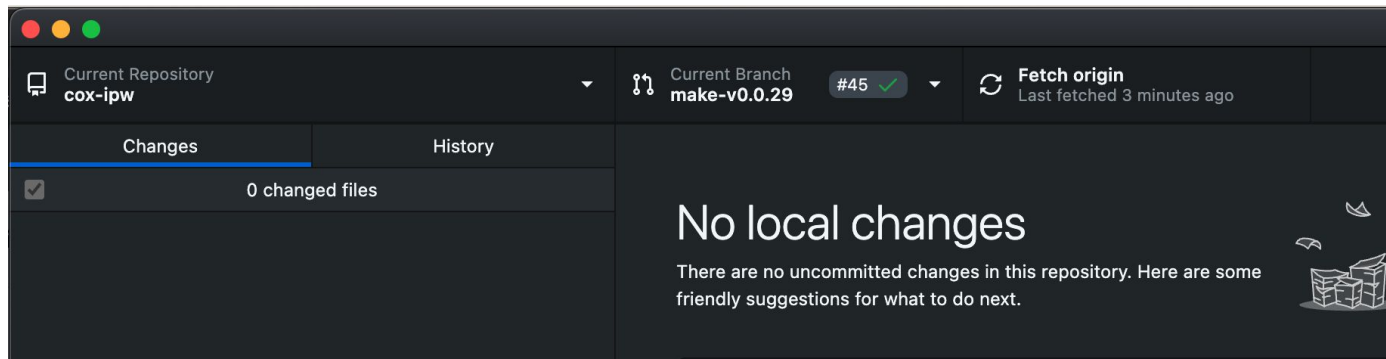


- git reset is your friend

A terminal window with a dark background and light gray text. The window title bar at the top shows three window control buttons (red, yellow, green) on the left, followed by a folder icon and the text "cox-ipw — -zsh — 80x24". The terminal content shows a login message, a git reset command being executed, and the resulting commit hash and message. The prompt is currently empty with a cursor.

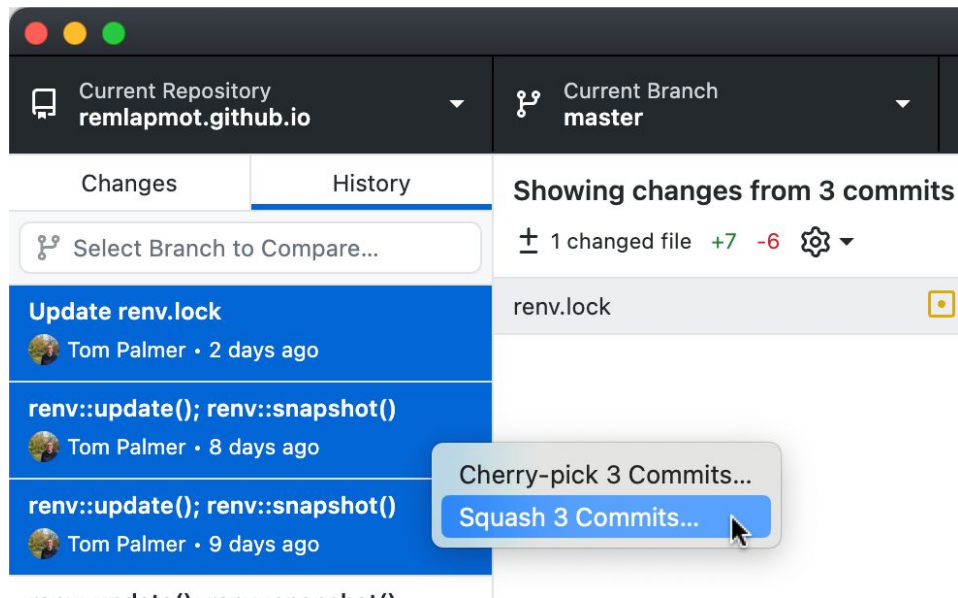
```
Last login: Tue Oct 31 20:28:09 on ttys000
[tom@Toms-Mac-mini cox-ipw % git reset --hard origin/make-v0.0.29
HEAD is now at 10a8e27 Add codelists update bullet
tom@Toms-Mac-mini cox-ipw %
```

- GitHub Desktop is happy again

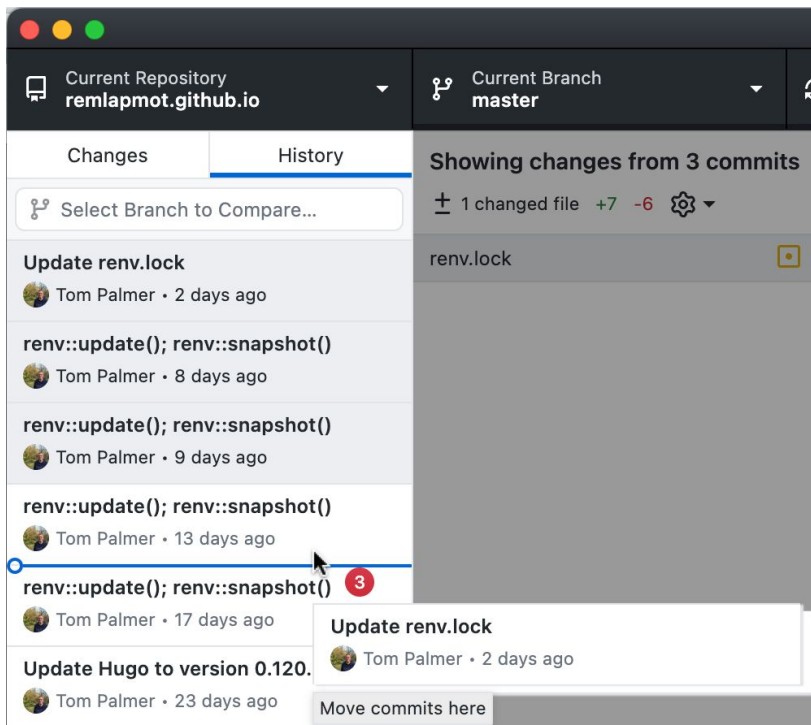


Squashing and reordering commits

- Select commits and reorder/squash
- Occasionally GitHub Desktop can hang and fail to complete such operations
- In which case open a terminal in your repo and run: `git rebase --abort`



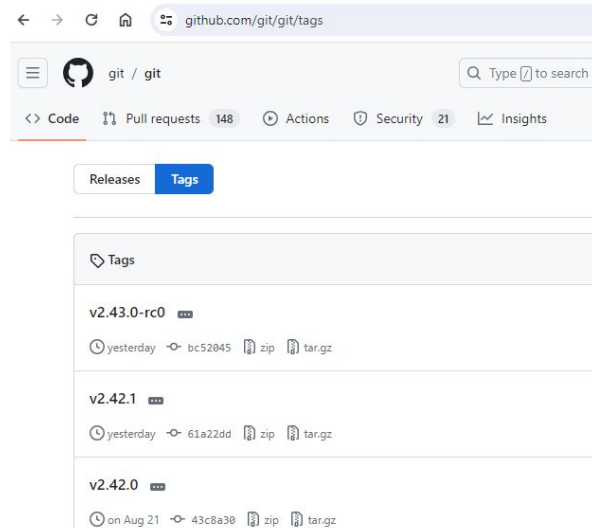
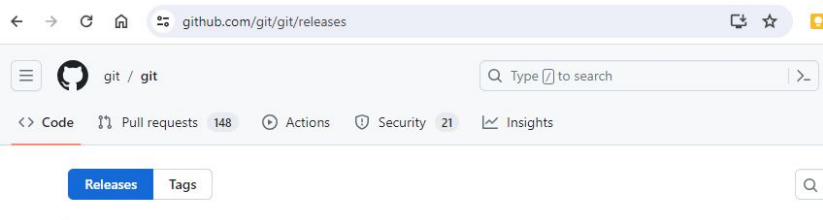
Reordering example (select commits | drag and drop where you want)



Git features versus GitHub features

- GitHub provides a lot of features on top of Git features
- E.g., Git has the concept of tags (for a version release), GitHub extends this with its Releases
 - This is why a lot of projects which are mailing list based only use Git features, e.g., Git itself

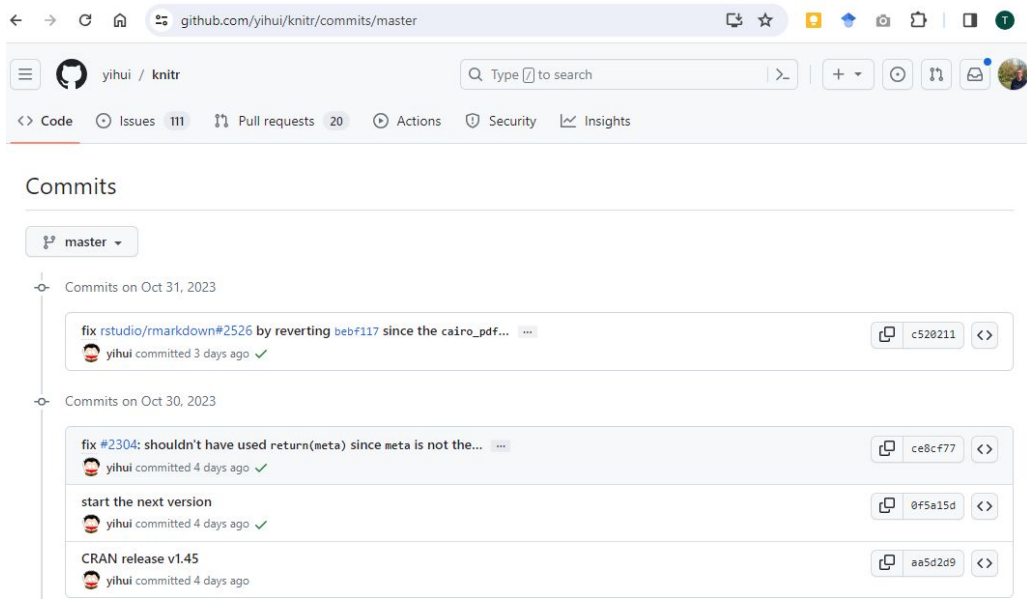
<https://github.com/git/git/>



My R package is on CRAN - what's my workflow?

- Once you have made a release to CRAN and release/tag in your repo you can continue committing on the default branch e.g.,

<https://github.com/yihui/knitr>



My R package is not on CRAN - what's my workflow?

- Users are (likely) installing your package from the latest commit on the main branch, e.g., `remotes::install_github("TwoSampleMR")`
- Should you declare a new version with each commit onto main/master??
- We could instruct users to install the latest release, i.e.,

```
remotes::install_github("MRCIEU/TwoSampleMR@*release")
```

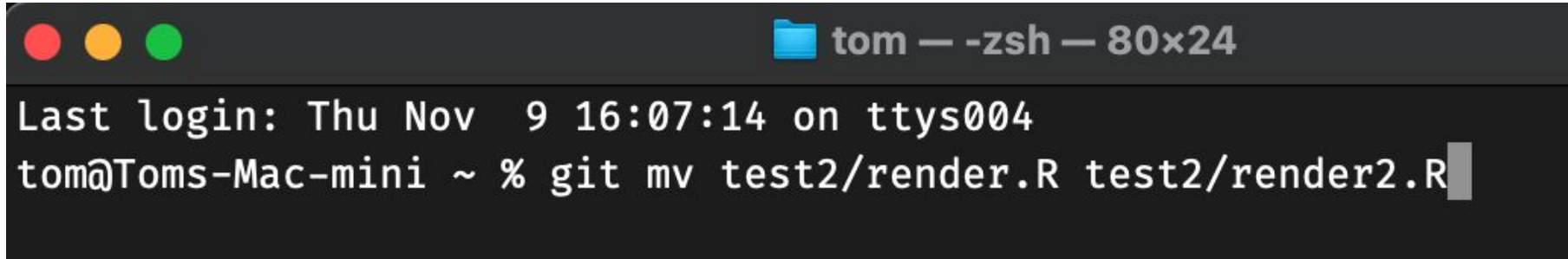
Under what circumstances do merge conflicts occur?

<https://docs.github.com/en/pull-requests/collaborating-with-pull-requests/addressing-merge-conflicts/about-merge-conflicts>

- Editing same line of same file (or editing a file that someone else deletes)

Renaming a file to retain its Git history: git mv

- Do not just commit in deletion of old file and creation of new one
- `git mv existing-filename new-filename`

A terminal window with a dark background. The title bar shows three colored circles (red, yellow, green) on the left and a folder icon followed by the text "tom — -zsh — 80x24" on the right. The terminal content shows the last login time and the command being entered.

```
Last login: Thu Nov  9 16:07:14 on ttys004
tom@Toms-Mac-mini ~ % git mv test2/render.R test2/render2.R
```

Gotcha: Windows (CRLF) versus macOS/Linux/Unix (LF) line endings

- Secretly Git wants text files to have LF line endings, and this is how they are stored on GitHub
- Include a .gitattributes file in your repo to be safe

```
# Auto detect text files and perform LF normalization
* text=auto
```

- <https://grpehr.github.io/training/01-08-common-errors.html#no-content-changes-found>
- Be careful if all text files in a repo show as changed - give away you might be about to make this mistake

Perhaps the best worst error/warning message

You are in 'detached HEAD' state. You can look around, make experimental changes and commit them, and you can discard any commits you make in this state without impacting any branches by switching back to a branch.

If you want to create a new branch to retain commits you create, you may do so (now or later) by using -c with the switch command. Example:

```
git switch -c <new-branch-name>
```

Or undo this operation with:

```
git switch -
```

Turn off this advice by setting config variable advice.detachedHead to false

How can we improve the commits in this PR?

<https://github.com/MRCIEU/TwoSampleMR/pull/419>

added steiger sensitivity for unmeasured confounding #419



explodecomputer wants to merge 1 commit into `master` from `418-add-extra-mr-steiger-sensitivity` 



Conversation 0



Commits 1



Checks 8



Files changed 8



Commits on Jun 13, 2023

added steiger sensitivity for unmeasured confounding



explodecomputer committed on Jun 13 

- Try to make sure each commit does a specific thing - so that it could be reverted later if necessary
- (Arguably this should be 3 PRs)
 - the addition of the new Steiger code
 - ggplot2 size to linewidth argument
 - mr_keep fix

GitHub message ... but couldn't find those commits

Replace Gitpod with GitHub Codespaces #1398



Open

StevenMaude wants to merge 9 commits into `main` from `steve/replace-gitpod-with-github-codespaces`



Conversation 3

Commits 9



Checks 2



Files changed 27



We went looking everywhere, but couldn't find those commits.

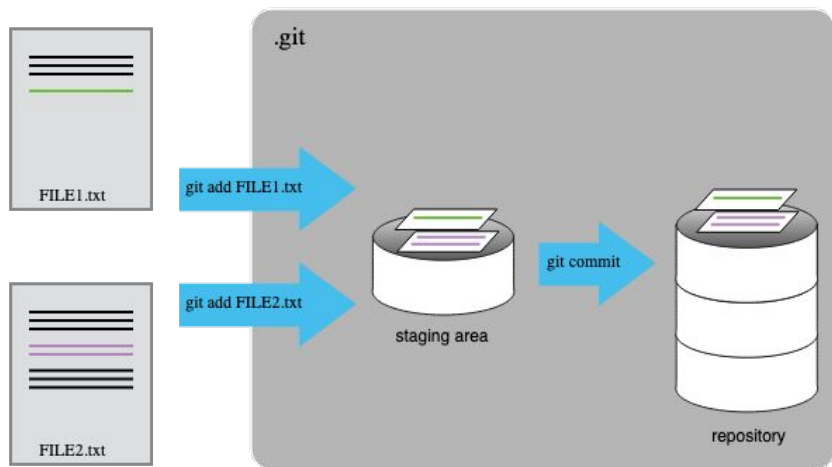
Sometimes commits can disappear after a force-push. Head back to the [latest changes here](#).

Empty subdirectories ... don't show up in Git

- Git tracks file contents
- If a new subdirectory of your repo is empty it will not show as a change in Git (nor in GitHub Desktop)
- If you want to keep a subdirectory add a blank text file named `.gitkeep` and check that in

git concepts: working tree

A git repository (or repo) is a directory tracked by git, and the version control information is stored under the `./git` directory. This external directory is called the **working tree**.



```
XMCF7HJ8F4 in ~/local-projects/TwoSampleMR
> tree -L 1 -l -a
.
├── .Rbuildignore
├── .git
├── .gitattributes
├── .github
├── .gitignore
├── CITATION.cff
├── DESCRIPTION
├── LICENSE
├── LICENSE.md
├── NAMESPACE
├── NEWS.md
├── R
├── README.Rmd
├── README.md
├── TwoSampleMR.Rproj
├── _pkgdown.yml
├── index.md
├── inst
├── man
├── pkgdown
├── tests
└── vignettes

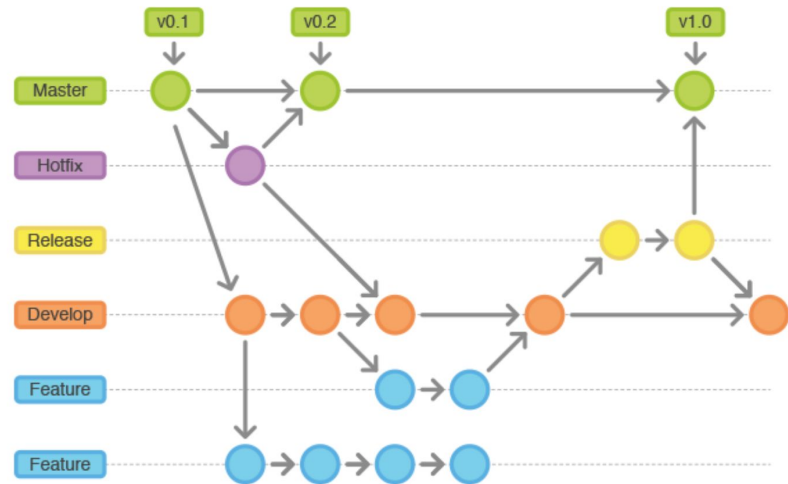
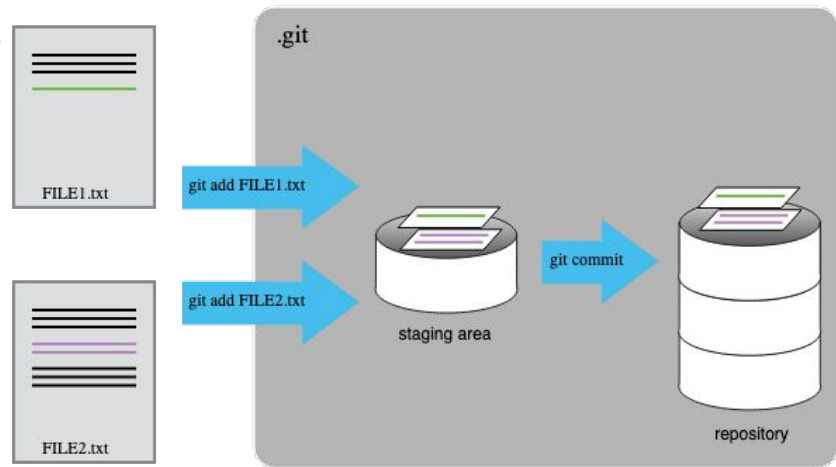
9 directories, 14 files
```

git concepts: commits, branches

Your source version history consists of **commits** as basic units

A **branch** is a collection of commits.

As a convention, for a big repo, extensive changes should be first done in their own branches, e.g. a branch involving development for a big feature. After the development reaches a good level of maturity, this branch gets **merged** back to the default **master** / **main** branch.



git concepts: the **remote** repository

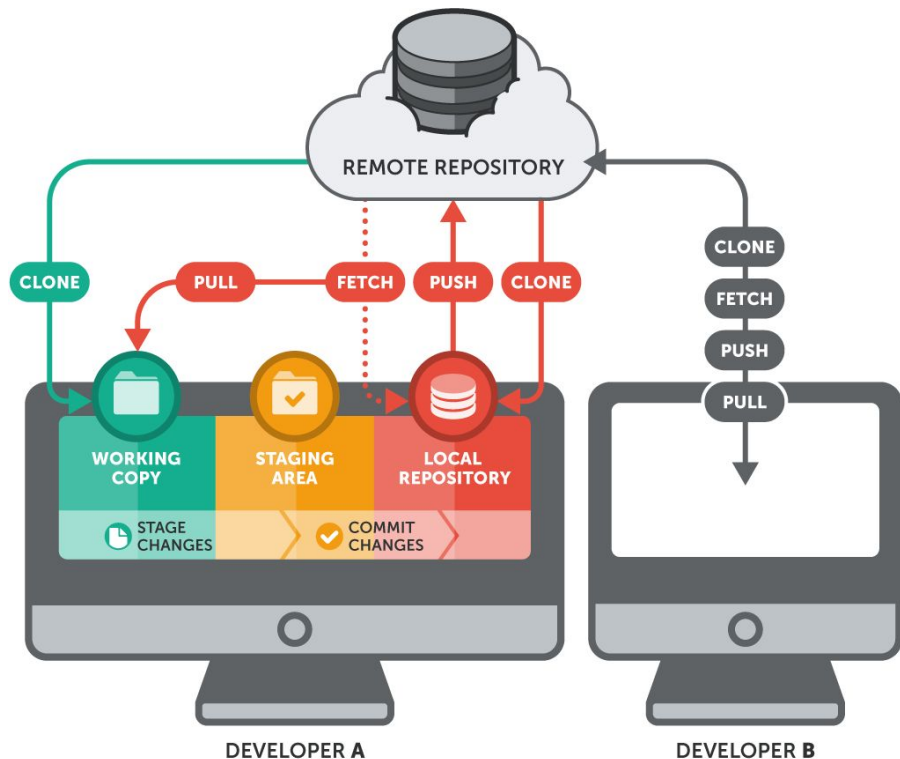
git is a *decentralised* system -- your **local** repo contains all source history already.

Nevertheless you can pair your local repo with an **upstream** remote repo to use it as a

- Backup
- *Centralised* source for collaboration between multiple colleagues, or yourself across multiple machines

Your local repo can be synced with a remote via **git pull (fetch & rebase)** and **git push**.

GitHub is one of the platforms to host your remote repos.



Common git operations

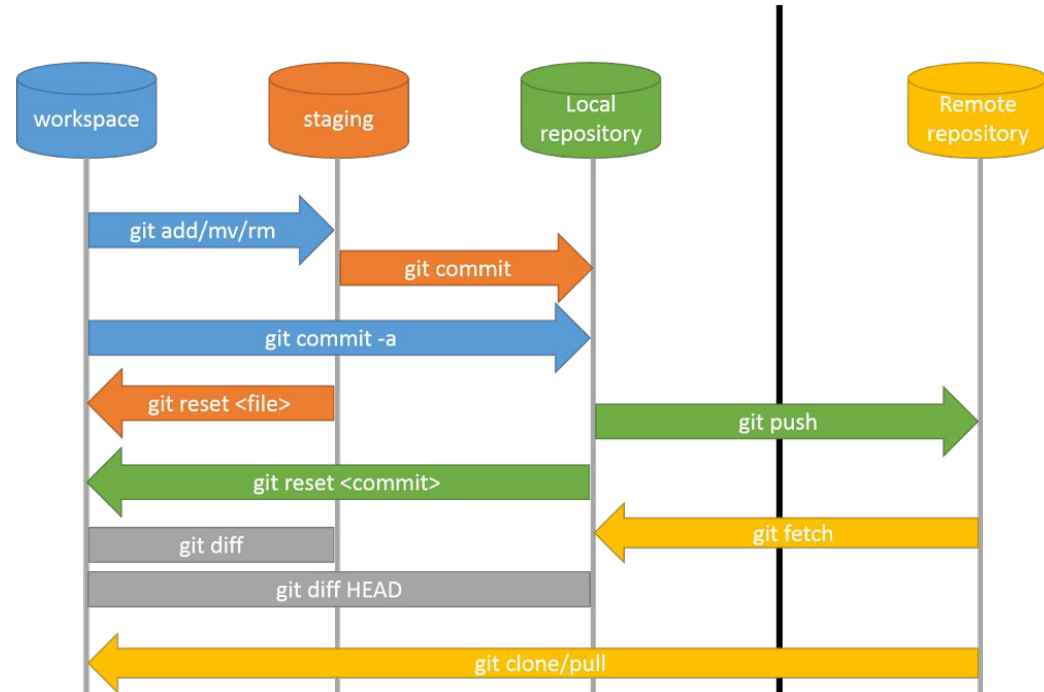
- View git history
 - repo
 - file
- See what has been changed (git diff)
- Try make some changes (git add, git commit)
- Reset those changes (git reset)!
 - NB: git revert is a different command
- Branches
- Remote

Tips for starters:


1. Billions of tutorials on git out there
2. Use a graphical client for those git commands

Here I use the Visual Studio Code text editor as my git client.



Extra extension: Git History





https://github.com/YiLiu6240/TwoSampleMR-test-git


 **TwoSampleMR-test-git** Public


forked from [MRCIEU/TwoSampleMR](#)


 Pin  Watch 0


 master


 26 branches

 43 tags

 Go to file

 Add file


 Code


 **Your master branch isn't protected**


Protect this branch from force pushing or deletion, or require status checks before merging. [Learn more](#)









Protect this branch

This branch is up to date with MRCIEU/TwoSampleMR:master.

 Contribute

 Sync fork

 remlapmot Merge pull request [MRCIEU#456](#) from remlapmot/tweaks-oct-2... cbd8e6 2 weeks ago 1,459 commits


	.github	Add write permission to GITHUB_TOKEN	2 weeks ago
	R	Fix MR-Base URL to use https	2 weeks ago
	inst	Fix http to https for ebl.ac.uk	2 weeks ago
	man	devtools::document()	2 weeks ago
	pkgdown	Set font-size to 14px relative to 1rem equals 16px	7 months ago
	tests	Make tests non-empty (in testthat meaning)	7 months ago
	vignettes	Add url to HMG article	2 weeks ago
	.Rbuildignore	Ignore README.html	2 weeks ago

Create a new fork

A *fork* is a copy of a repository. Forking a repository allows you to freely experiment with changes without affecting the original project. [View existing forks.](#)

Required fields are marked with an asterisk (*).

Owner *

 YiLiu6240

Repository name *

TwoSampleMR-test-git

✔ TwoSampleMR-test-git is available.

By default, forks are named the same as their upstream repository. You can customize the name to distinguish it further.

Description (optional)

R package for performing 2-sample MR using MR-Base database

☐ Copy the **master** branch only

Contribute back to MRCIEU/TwoSampleMR by adding your own branch. [Learn more.](#)

 You are creating a fork in your personal account.

Create fork

git history

The screenshot displays the Visual Studio Code interface with the 'TwoSampleMR' project open. The left sidebar shows the Explorer view with the following structure:

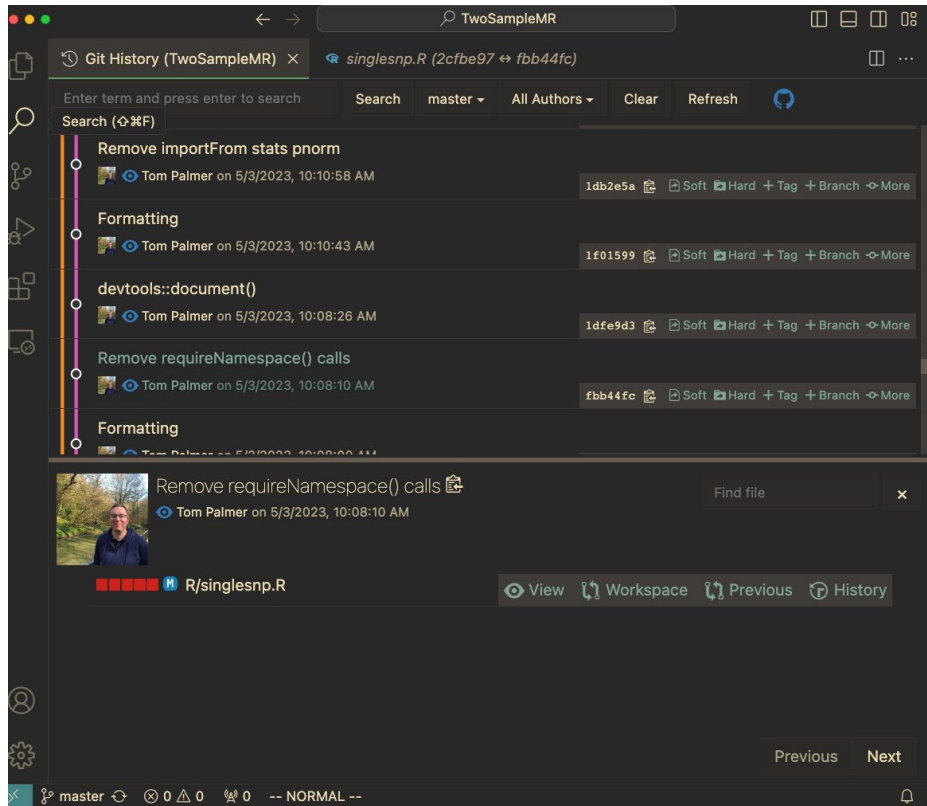
- OPEN EDITORS
 - Git History (TwoSampleMR)
- TWOSAMPLEMR
 - .github
 - inst
 - man
 - pkgdown
 - R
 - tests
 - vignettes
 - ! _pkgdown.yml
 - ! .gitattributes
 - ! .gitignore
 - ! .Rbuildignore
 - ! CITATION.cff
 - ! DESCRIPTION
 - ! index.md
- OUTLINE
- TIMELINE
 - The active editor cannot provide timeline information.
- COMMIT

The main panel shows the 'Git History (TwoSampleMR)' view. It includes a search bar and filters for 'master', 'All Authors', 'Clear', and 'Refresh'. The commit history is listed as follows:

Commit Hash	Author	Date	Message
951b34a	Tom Palmer	5/3/2023, 7:03:10 AM	devtools::document()
41f4b96	Tom Palmer	5/3/2023, 7:03:01 AM	Formatting
21d0885	Tom Palmer	5/3/2023, 7:00:18 AM	devtools::document()
b8f6dd4	Tom Palmer	5/3/2023, 7:00:09 AM	Remove importFrom stats
9217fdc	Tom Palmer	5/3/2023, 6:55:32 AM	devtools::document()
02be3b9	Tom Palmer	5/3/2023, 6:55:13 AM	Formatting
5eaa799	Tom Palmer	5/3/2023, 6:55:04 AM	Remove importFrom stats
a4c8eb3	Tom Palmer	5/3/2023, 6:45:00 AM	devtools::document()
907cfb8	Tom Palmer	5/3/2023, 6:44:49 AM	Replace {code}{link{}} with []
			devtools::document()

The status bar at the bottom indicates the current branch is 'master' and shows 0 changes, 0 conflicts, and 0 errors.

git diff



The screenshot shows the Git GUI interface for the repository 'TwoSampleMR'. The 'Git History' tab is active, displaying a list of commits. The search bar at the top contains 'singlesnp.R (2cfbe97 ↔ fbb44fc)'. The commit list shows several entries, including 'Remove importFrom stats pnorm', 'Formatting', 'devtools::document()', 'Remove requireNamespace() calls', and 'Formatting'. The commit 'Remove requireNamespace() calls' by Tom Palmer on 5/3/2023, 10:08:10 AM is selected, showing a diff for 'R/singlesnp.R'. The diff shows changes to the 'requireNamespace()' function calls. The bottom status bar indicates the current branch is 'master' and the diff is 'NORMAL'.

Git History (TwoSampleMR) x singlesnp.R (2cfbe97 ↔ fbb44fc)

Enter term and press enter to search Search master All Authors Clear Refresh

Search (⌘F)

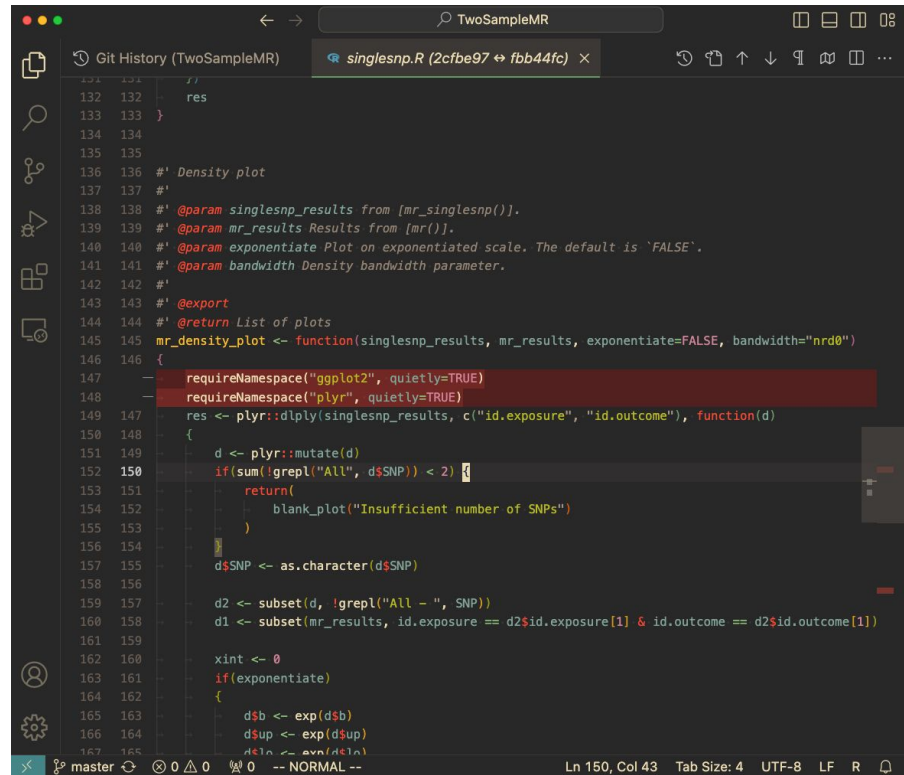
- Remove importFrom stats pnorm
Tom Palmer on 5/3/2023, 10:10:58 AM 1db2e5a [Soft] [Hard] + Tag + Branch < More
- Formatting
Tom Palmer on 5/3/2023, 10:10:43 AM 1f01599 [Soft] [Hard] + Tag + Branch < More
- devtools::document()
Tom Palmer on 5/3/2023, 10:08:26 AM 1dfe9d3 [Soft] [Hard] + Tag + Branch < More
- Remove requireNamespace() calls
Tom Palmer on 5/3/2023, 10:08:10 AM fbb44fc [Soft] [Hard] + Tag + Branch < More
- Formatting
Tom Palmer on 5/3/2023, 10:08:10 AM

Remove requireNamespace() calls
Tom Palmer on 5/3/2023, 10:08:10 AM Find file x

R/singlesnp.R View Workspace Previous History

Previous Next

master 0 0 0 -- NORMAL --



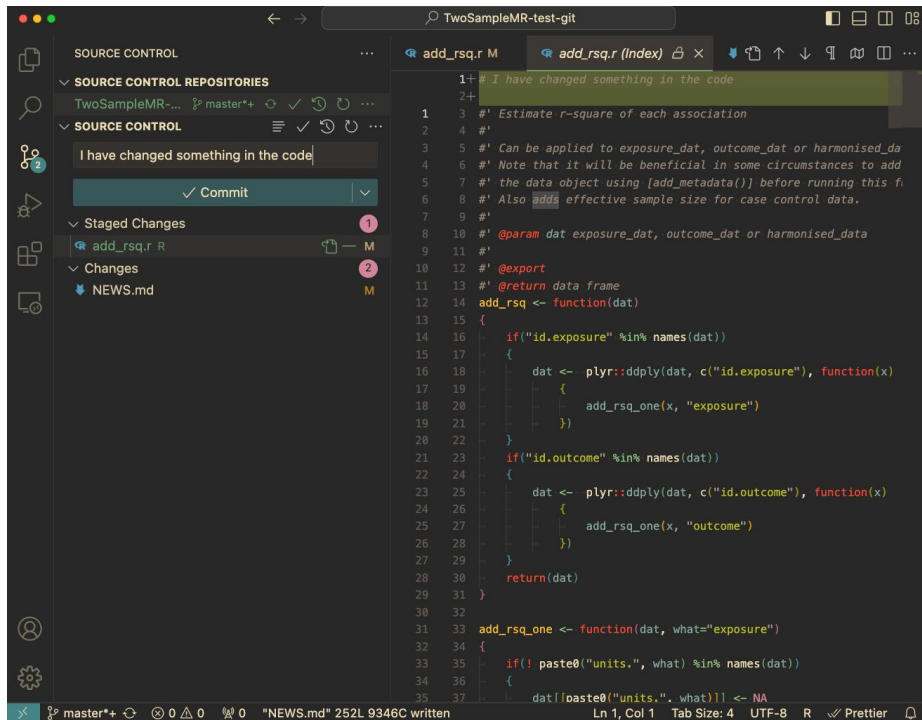
The screenshot shows the Git GUI interface for the repository 'TwoSampleMR', displaying the diff for the commit 'Remove requireNamespace() calls'. The 'Git History' tab is active, and the search bar contains 'singlesnp.R (2cfbe97 ↔ fbb44fc)'. The diff shows changes to the 'requireNamespace()' function calls in the 'R/singlesnp.R' file. The diff is highlighted in red, indicating the changes. The bottom status bar indicates the current branch is 'master' and the diff is 'NORMAL'.

Git History (TwoSampleMR) x singlesnp.R (2cfbe97 ↔ fbb44fc)

```
132 132 res
133 133 }
134 134
135 135
136 136 #' Density plot
137 137 #'
138 138 #' @param singlesnp_results from [mr_singlesnp()].
139 139 #' @param mr_results Results from [mr()].
140 140 #' @param exponentiate Plot on exponentiated scale. The default is 'FALSE'.
141 141 #' @param bandwidth Density bandwidth parameter.
142 142 #'
143 143 #' @export
144 144 #' @return List of plots
145 145 mr_density_plot <- function(singlesnp_results, mr_results, exponentiate=FALSE, bandwidth="nrd0")
146 146 {
147 147   requireNamespace("ggplot2", quietly=TRUE)
148 148   requireNamespace("plyr", quietly=TRUE)
149 149   res <- plyr::dply(singlesnp_results, c("id.exposure", "id.outcome"), function(d)
150 150   {
151 151     d <- plyr::mutate(d)
152 152     if(sum(!grepl("All", d$SNP)) < 2) {
153 153       return(
154 154         blank_plot("Insufficient number of SNPs")
155 155       )
156 156     }
157 157     d$SNP <- as.character(d$SNP)
158 158
159 159     d2 <- subset(d, !grepl("All - ", SNP))
160 160     d1 <- subset(mr_results, id.exposure == d2$id.exposure[1] & id.outcome == d2$id.outcome[1])
161 161
162 162     xint <- 0
163 163     if(exponentiate)
164 164     {
165 165       d$b <- exp(d$b)
166 166       d$up <- exp(d$up)
167 167       d$dn <- exp(d$dn)
```

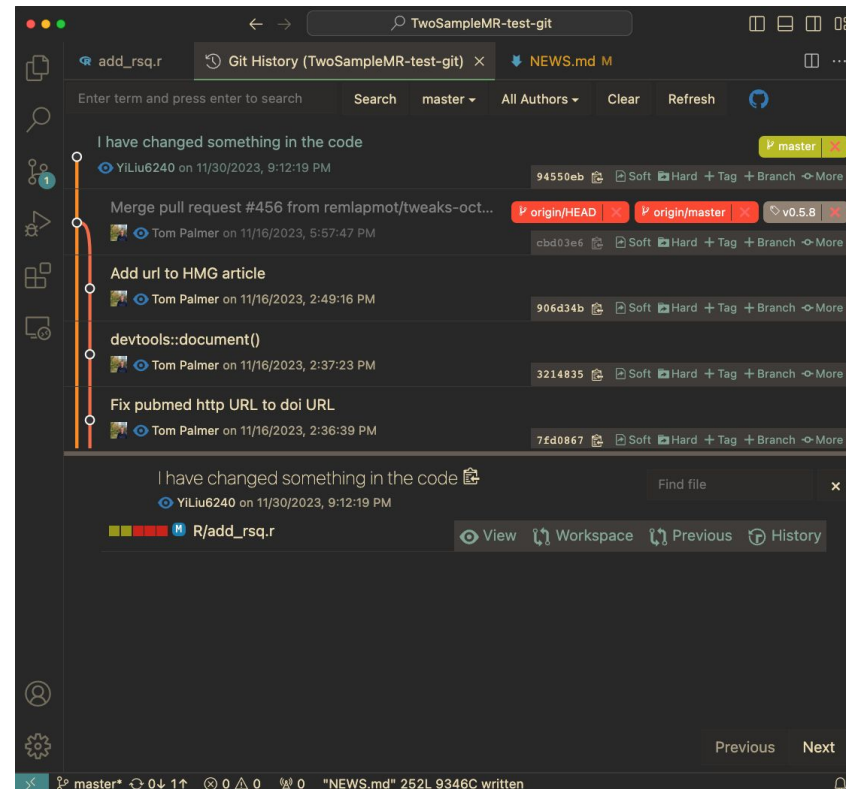
master 0 0 0 -- NORMAL -- Ln 150, Col 43 Tab Size: 4 UTF-8 LF R

Make some changes and commit



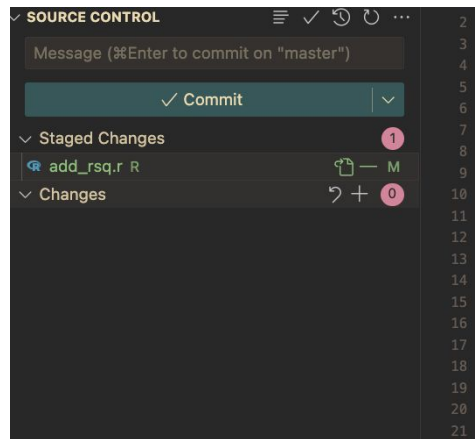
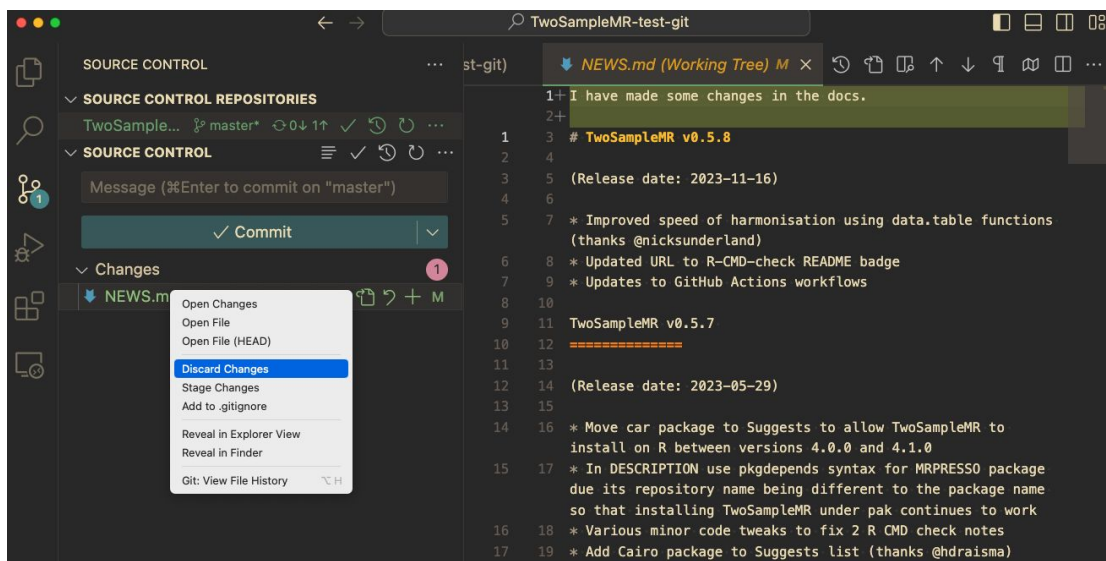
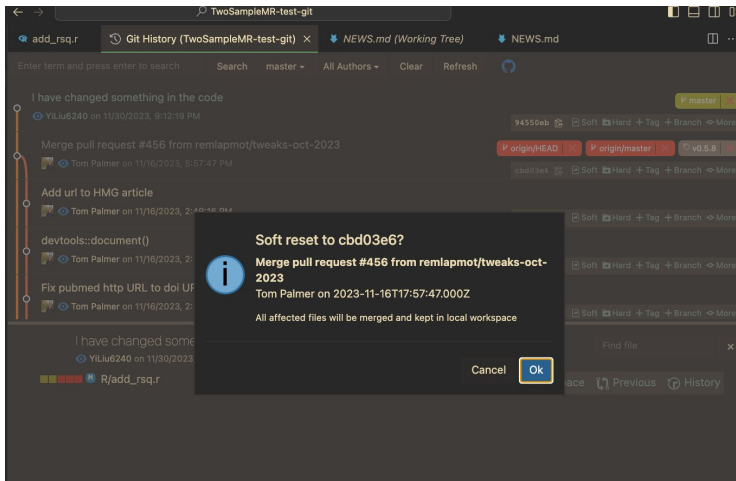
The screenshot shows the VS Code interface with the 'TwoSampleMR-test-git' repository. The 'SOURCE CONTROL' sidebar on the left shows a commit message 'I have changed something in the code' and a 'Commit' button. The main editor displays the R script 'add_rsqr.R' with the following content:

```
1+ # I have changed something in the code
2+
3 #' Estimate r-square of each association
4 #'
5 #' Can be applied to exposure_dat, outcome_dat or harmonised_data
6 #' Note that it will be beneficial in some circumstances to add
7 #' the data object using [add_metadata()] before running this f.
8 #' Also add effective sample size for case control data.
9 #'
10 #' @param dat exposure_dat, outcome_dat or harmonised_data
11 #'
12 #' @export
13 #' @return data frame
14 add_rsqr <- function(dat)
15 {
16   if("id.exposure" %in% names(dat))
17   {
18     dat <- plyr::ddply(dat, c("id.exposure"), function(x)
19     {
20       add_rsqr_one(x, "exposure")
21     })
22   }
23   if("id.outcome" %in% names(dat))
24   {
25     dat <- plyr::ddply(dat, c("id.outcome"), function(x)
26     {
27       add_rsqr_one(x, "outcome")
28     })
29   }
30   return(dat)
31 }
32
33 add_rsqr_one <- function(dat, what="exposure")
34 {
35   if(! paste0("units.", what) %in% names(dat))
36   {
37     dat[[paste0("units.", what)]] <- NA
38   }
39 }
```

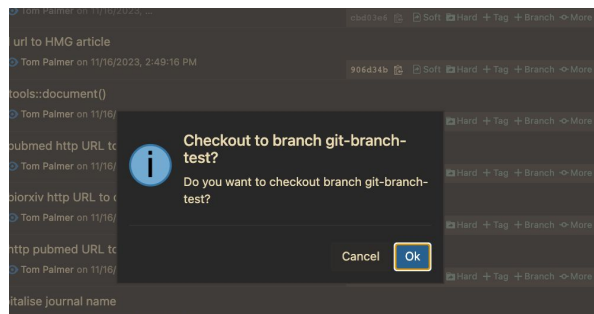
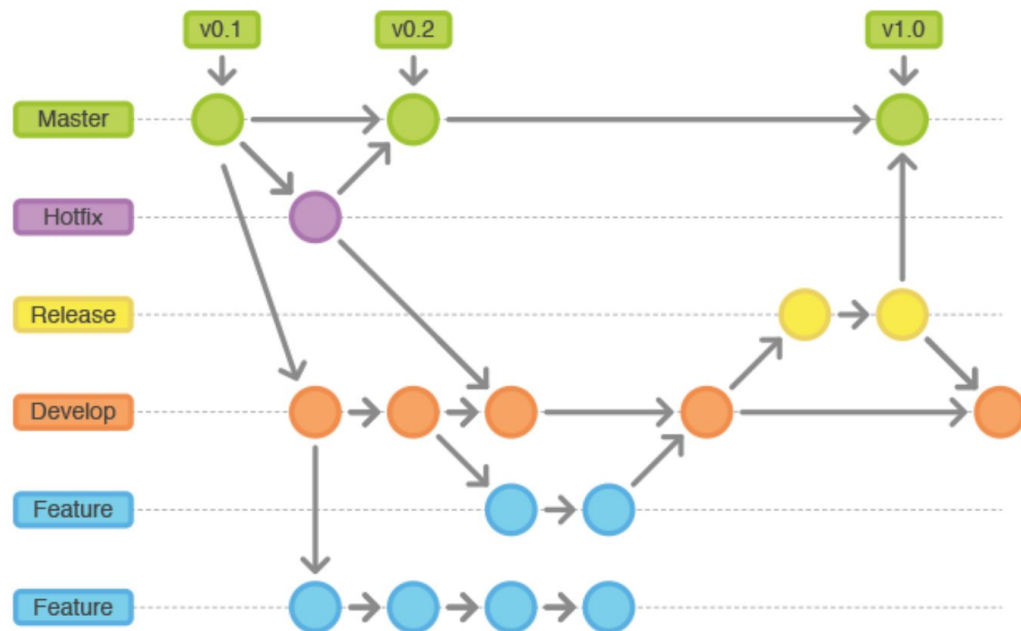
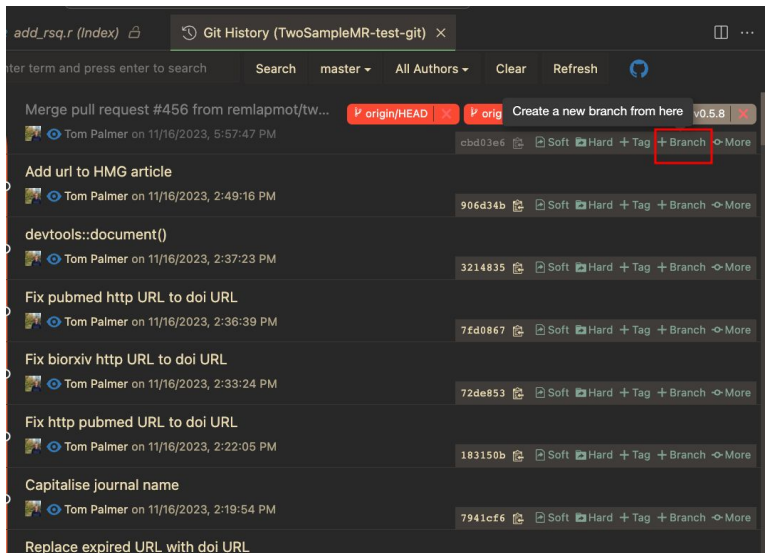


The screenshot shows the VS Code interface with the 'TwoSampleMR-test-git' repository. The 'Git History' panel on the right displays a list of commits. The commit message 'I have changed something in the code' is highlighted. The commit details show the commit hash '94550eb', the author 'YILiu6240', and the date '11/30/2023, 9:12:19 PM'. The commit message is 'I have changed something in the code'. The commit details also show the commit hash '94550eb', the author 'YILiu6240', and the date '11/30/2023, 9:12:19 PM'. The commit message is 'I have changed something in the code'. The commit details also show the commit hash '94550eb', the author 'YILiu6240', and the date '11/30/2023, 9:12:19 PM'. The commit message is 'I have changed something in the code'.

resetting changes



branches



The screenshot shows the RStudio interface with the Source Control panel on the left and a Git History window on the right. The Source Control panel displays a commit message and a 'Commit' button. A context menu is open over the 'Commit' button, showing options like 'Pull', 'Push', 'Clone', 'Checkout to...', 'Fetch', 'Commit', 'Changes', 'Pull, Push', 'Branch', 'Remote', 'Stash', 'Tags', and 'Show Git Output'. The 'Pull, Push' option is highlighted, and a sub-menu is open showing 'Sync', 'Pull', 'Pull (Rebase)', 'Pull from...', 'Push', and 'Push to...'. The 'Push' option is highlighted in the sub-menu. The Git History window shows a list of commits, with the most recent commit selected.

git-branch-test

Commits on Nov 30, 2023

A test commit

Yiliu6240 committed 2 minutes ago

a2388ba

Commits on Nov 16, 2023

Merge pull request [MRClEU#456](#) from remlapmot/tweaks-oct-2023

remlapmot committed 2 weeks ago

Verified
 cb083e6

Add uri to HMG article

remlapmot committed 2 weeks ago

Verified
 986d34b

devtools::document()

remlapmot committed 2 weeks ago

Verified
 3214835

Fix pubmed http URL to doi URL

remlapmot committed 2 weeks ago

Verified
 7fd8867

Fix biorxiv http URL to doi URL

remlapmot committed 2 weeks ago

Verified
 72de8f3

A test commit

git-branch-test

YILiu6240 committed 3 minutes ago

1 parent cbd0366 commit a2380ba

0 comments on commit a2380ba

Lock conversation

Showing 1 changed file with 2 additions and 0 deletions.

Whitespace

Ignore whitespace

Split

Unified

2

R/knit.R

@@ -1,3 +1,5 @@

1 + # Some test changes

2 +

3 #' Knit report using working environment

4 #'

5 #' Warning: It is quite likely that this will be called within an Rmd file

Resources

- Tutorials
 - <https://swcarpentry.github.io/git-novice/>
 - Training material for Electronic Health Records group <https://grpehr.github.io/training/>
- Technical details on Git
 - Linus Torvalds Google Tech Talk <https://youtu.be/4XpnKHJAok8?si=4NNxxajwhbohJ3Wl>
 - Git Rev News https://git.github.io/rev_news/
 - Many excellent blog posts
 - Derrick Stolee <https://github.blog/author/dstolee/>
 - Elijah Newren <https://blog.palantir.com/optimizing-gits-merge-machinery-1-127ceb0ef2a1>
- Git clients
 - GitHub Desktop, Tower, GitKraken (GitLens VSCode Extension)
- ... and a million tutorials about Git on Youtube and other platforms
- Nothing can replace your trial by error – don't be afraid to make mistakes

Summary

- Are you a Helena, Des, or Dom?
- Or maybe you'll have different use cases on different projects
- Any questions/problems working with Git and GitHub?