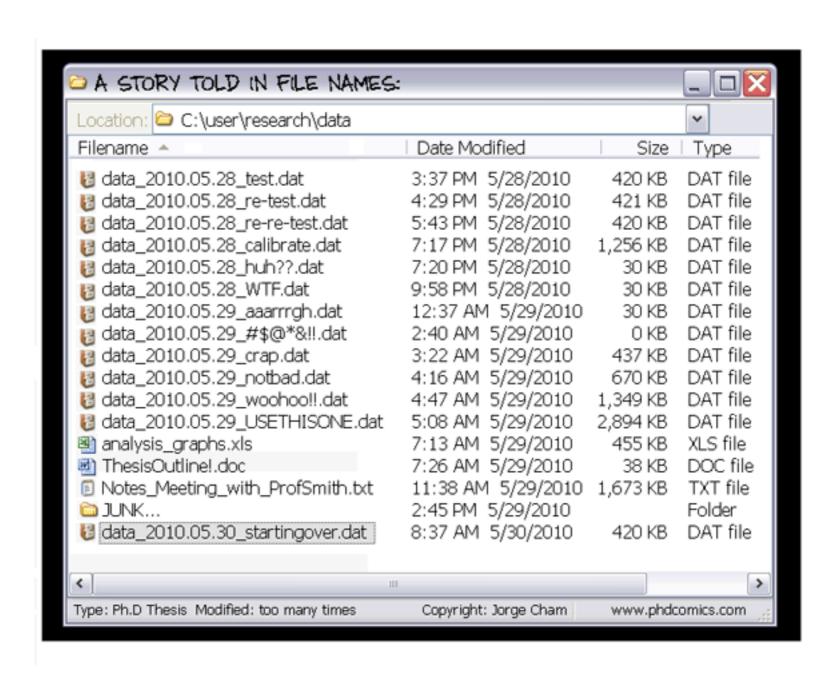
Collaborating with git/github

Rike-Benjamin Schuppner @debilski // http://debilski.de

Motivation: A story told in filenames



http://www.phdcomics.com/comics/archive.php?comicid=1323

Why version control?

- Keeps all historical versions for easy tracking and reference
- Allows you to quickly checkpoint new ideas (and dump them if they fail)
- Benefits team collaboration
- Improves our efficiency
- Can be used as 'data centre' to make build automation possible and testing easy

Creating a project

- Initializing
 - \$ git init.
 Initialized empty Git repository in /home/myself/Project01/.git/
- Cloning an existing project
 - \$ git clone <u>https://github.com/ASPP/GeoSim-Fall-School</u>

Configuration

- Your identity:
 - \$ git config --global user.name "Your Name Comes Here" \$ git config --global user.email you@yourdomain.com
- Setting up your editor:
 - \$ git config --global core.editor vim
- Checking your settings:
 - \$ git config --list

.gitignore

- Each repository (rather: each folder) can contain a .gitignore file
- Files and file patterns you do not want to track belong there

```
• .pyc
build/
.o
.DS Store
```

Git also excludes them from git status automatically

Git concepts

- Working directory
 - All the files you created in the directory
- Staging area (also called 'index')
 - Intermediate place to put files before committing them
- Master (commit), HEAD
 - Immutable snapshot with complete history, which can be shared with other people

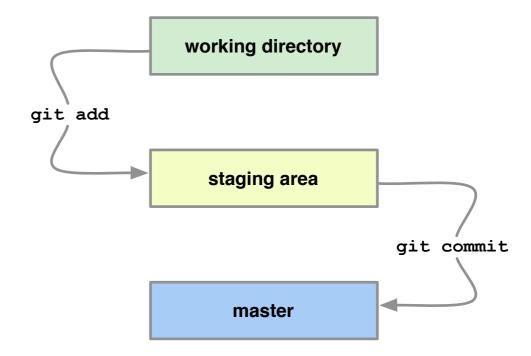
working directory

staging area

master

Using the staging area

- Add files to staging area with
 - \$ git add <filename>
- Commit the whole staging area using
 - \$ git commit
 - \$ git commit -m "I committed those files"



Git commit

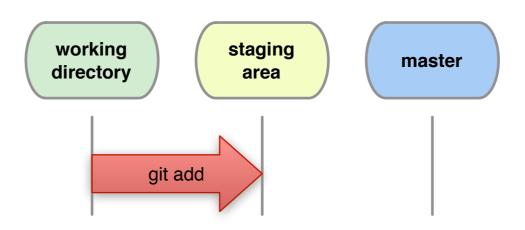
- Committing a version generates a hash calculated from:
 - Name of committer, current date
 - SHA of (one or more) parent commits
 - SHA of data tree
 - eg: 2a9a3c876b344cb74807edd6340b7bcd930014a8

- This means that everybody who has a commit with the same hash knows:
 - It is from the same person
 - It has the same history
 - It contains the same data
- Sharing data is simply checking that the most recent hash is the same

git add <filename>

adds a file to the index

- For adding all tracked (and changed) files (and removing deleted ones):
 - git add -u
- For adding interactively:
 - git add -p



git rm <filename>

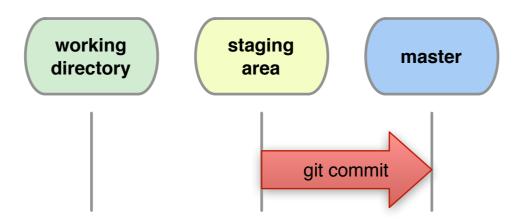
removes a file from the index

- For moving/renaming files, there is also
 - \$ git mv oldname newname
 - (Internally it is the same as doing git rm + git add)

git commit [-m "Commit message"]

adds a file to the index

- Creates a new commit with the current index
- Moves the current HEAD branch to the new commit



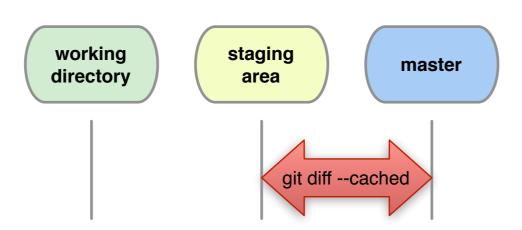
git diff

Shows what has been changed in working (and not yet added to the index)

```
diff --git a/Monday/Introduction-to-Plotting.ipynb b/Monday/Introduction-to-Plotting.ipynb
index 14fd585..d4b4fa7 100644
--- a/Monday/Introduction-to-Plotting.ipynb
+++ b/Monday/Introduction-to-Plotting.ipynb
00 -1,7 +1,7 00
 "metadata": {
 "name": "",
- "signature": "sha256:61bfab1750c869858224ad5cb43995bf3327bcaa6cf0890a542bbf6623a26806"
+ "signature": "sha256:00e69344cf858d41422342f6ec2179efe0482194a2dc8dec8bc7d3ec42d488d2"
 }, "nbformat": 3,
                                   working
                                                     staging
                                                                        master
                                   directory
                                                       area
                                             git diff
```

git diff --cached

Shows what is to be committed to HEAD (eg. shows the index)



git log shows the history

commit 2a9a3c876b344cb74807edd6340b7bcd930014a8

Author: Rike-Benjamin Schuppner < rikebs@debilski.de>

Date: Mon Oct 6 09:07:07 2014 +0200

ENH: Updates.

commit d3cd330f259617aa7a9b59bb07f2bf15129bd4a1

Author: Rike-Benjamin Schuppner < rikebs@debilski.de>

Date: Fri Oct 3 14:47:44 2014 +0200

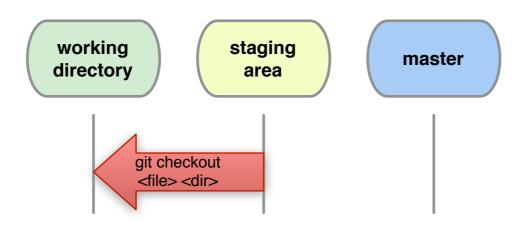
NF: Introduction to Plotting.

\$ git log --oneline --graph

git checkout -- <filename>

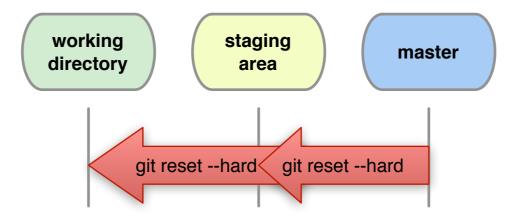
resets a file to the version in index

- Can also retrieve a file from a specific commit
 - \$ git checkout COMMIT -- <filename>



git reset current HEAD

- You can discard all changes in the working directory and index with
 - \$ git reset --hard
- Always check git status before doing so
- git reset can also be used to move around the HEAD branch (see help)



git clean

remove untracked files from working tree

- Always make a dry-run before cleaning
 - \$ git clean -n
- -x option also removes files from .gitignore
- -X only removes files listed in .gitignore

Useful tricks

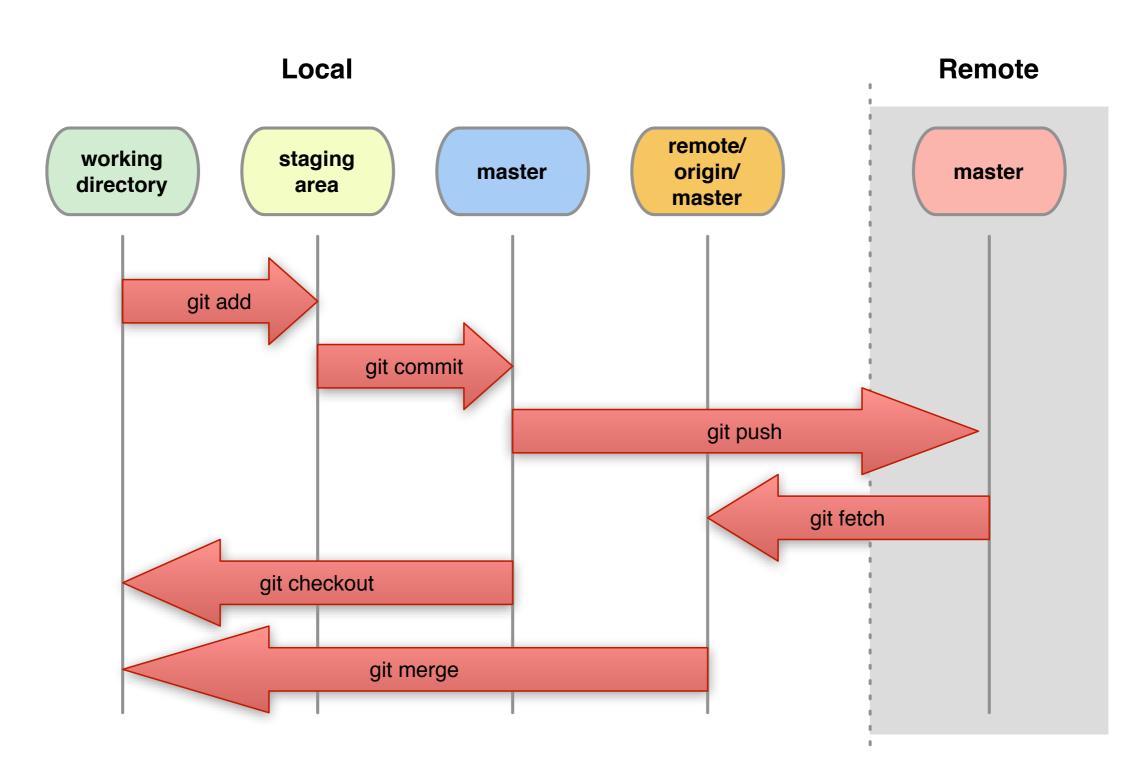
examine your repository

- Search for a string in all tracked files:
 - \$ git grep PATTERN
- Search the history for string addition/deletion:
 - \$ git log -Spattern
- Find out who did something stupid/awesome
 - \$ git blame

Branches and remotes

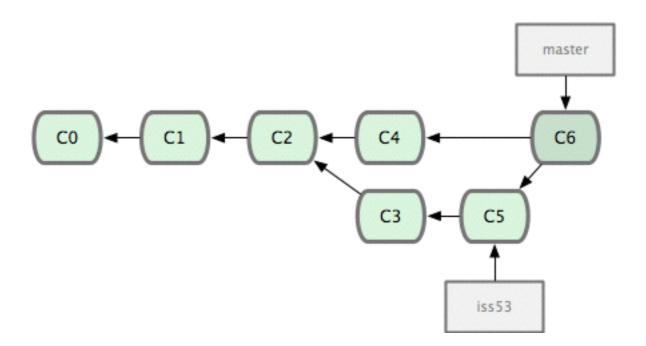
working non-linear and collaborating with others

Command overview



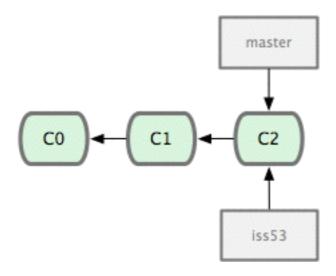
Branches

History is no linear process



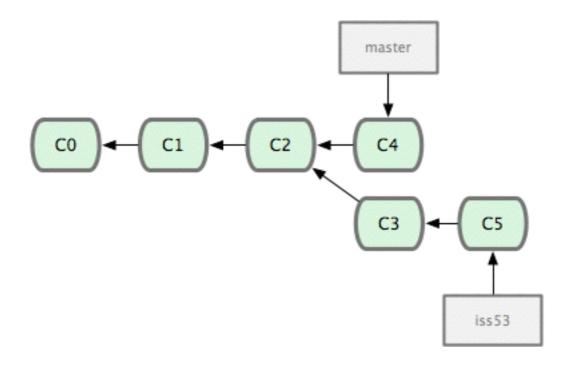
git branch
branchname>

Creates a new branch



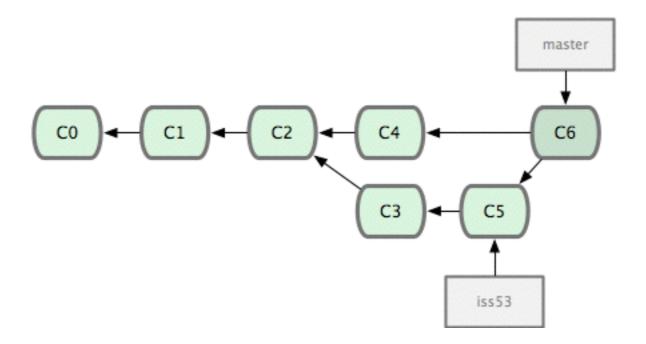
git checkout
branchname>

Switches to another branch



git merge
hranchname>

Merges the changes of another branch



resolving merge conflicts (1)

When branches do not mix

\$ git merge iss53
 Auto-merging index.html
 CONFLICT (content): Merge conflict in index.html
 Automatic merge failed; fix conflicts and then commit the result.

resolving merge conflicts (2)

When branches do not mix

```
• $ git status
   index.html: needs merge
   # On branch master
   # Changed but not updated:
       (use "git add <file>..." to update what will be committed)
       (use "git checkout -- <file>..." to discard changes in working dir
   #
       unmerged: index.html
   #
```

resolving merge conflicts (3)

When branches do not mix

```
    $ cat index.html
    <<l><<l><<l><<le>HEAD:index.html
    div id="footer">contact : email.support@github.com
    div id="footer"> please contact us at <a href="support@github.com">support@github.com</a></div>
    >>>> iss53:index.html
```

resolving merge conflicts (4)

When branches do not mix

- TODO list:
 - Resolve conflicts
 - Add file to index
 - Commit
- If you prefer one specific version:
 - \$ git checkout COMMIT -- filename

git remote -v

shows/manages the remote repositories

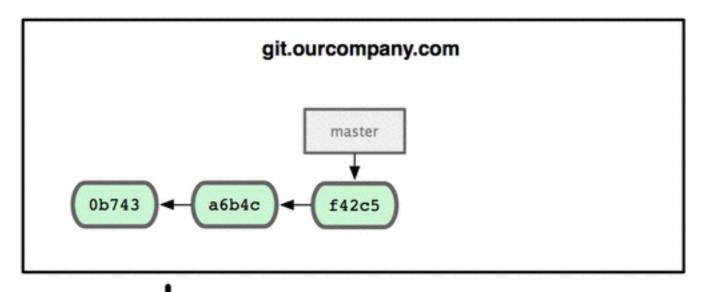
- You can have more than one remote repository! Usually origin refers to the remote (but for git it's just a name).
- \$ git remote -v
 origin git@github.com:ASPP/GeoSim-Fall-School.git
- For adding and removing remotes:
 - \$ git remote add NAME REPO \$ git remote rm NAME REPO

Remote repositories

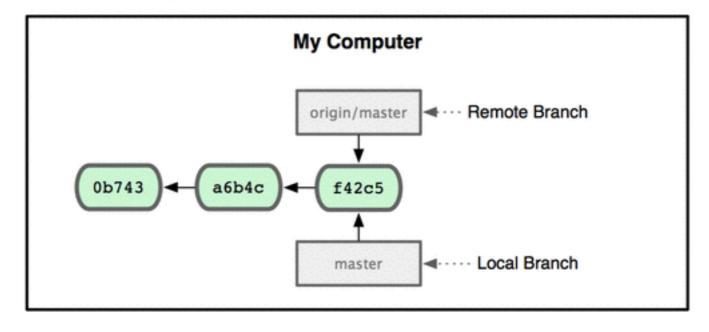
- Set up empty repository on remote server
 - remote \$ git init --bare Project01.git
- Add remote to local repo (here named origin)
 - local \$ git remote add origin myself@remote:/Project01.git
- Push everything
 - local \$ git push origin master

Working with others

Cloning another repository

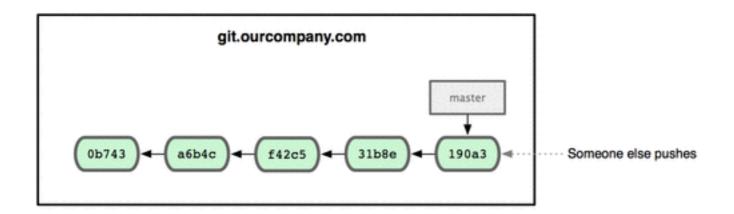


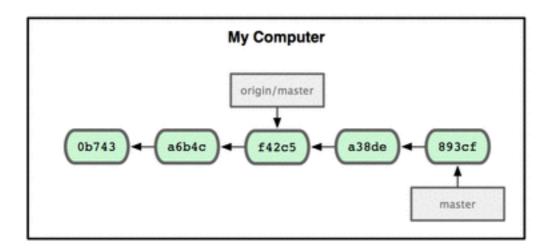
git clone schacon@git.ourcompany.com:project.git



git fetch origin

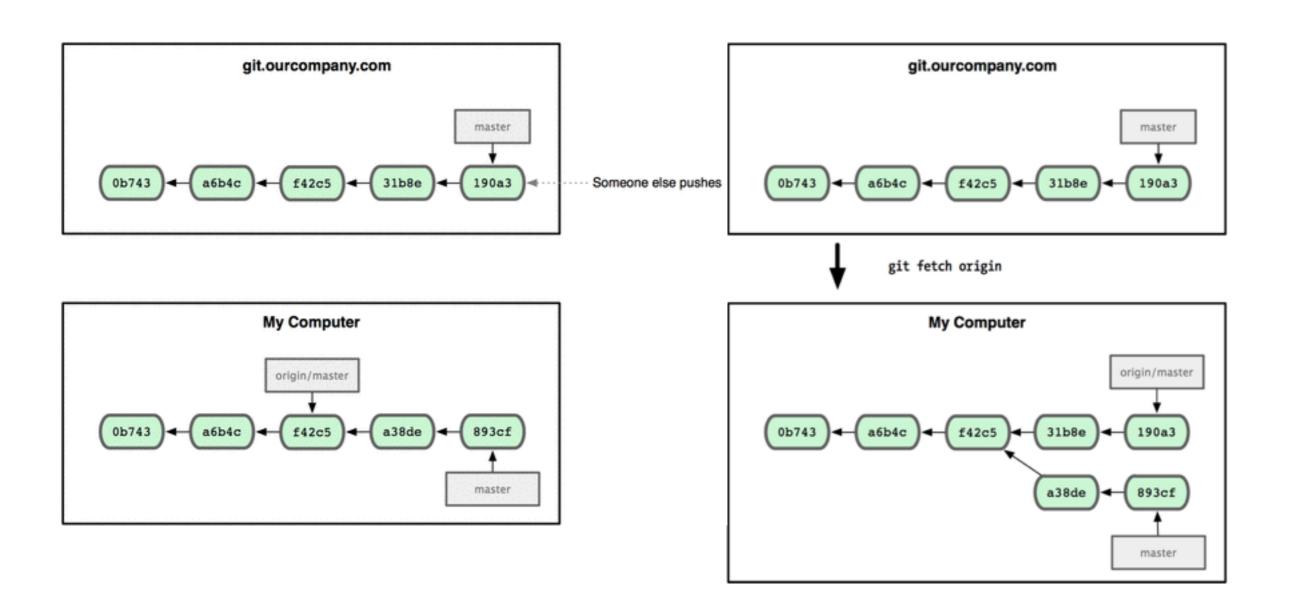
Fetch commits from a remote repository





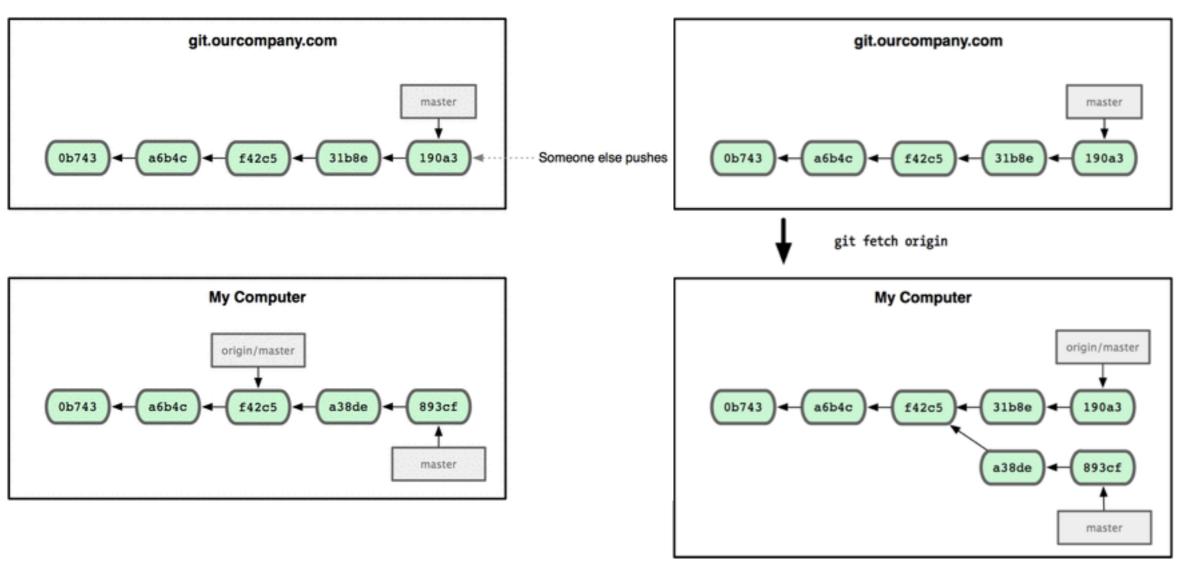
git fetch origin

Fetch commits from a remote repository



git fetch origin

Fetch commits from a remote repository



Does not change local commits. Need to merge manually.

git pull = git fetch && git merge

- Be careful with this one
- Fine to use when you want to update your code to upstream and do not have local changes
- If you do have local changes, you might prefer merging manually (sometimes)

git push

move your changes to a remote repository

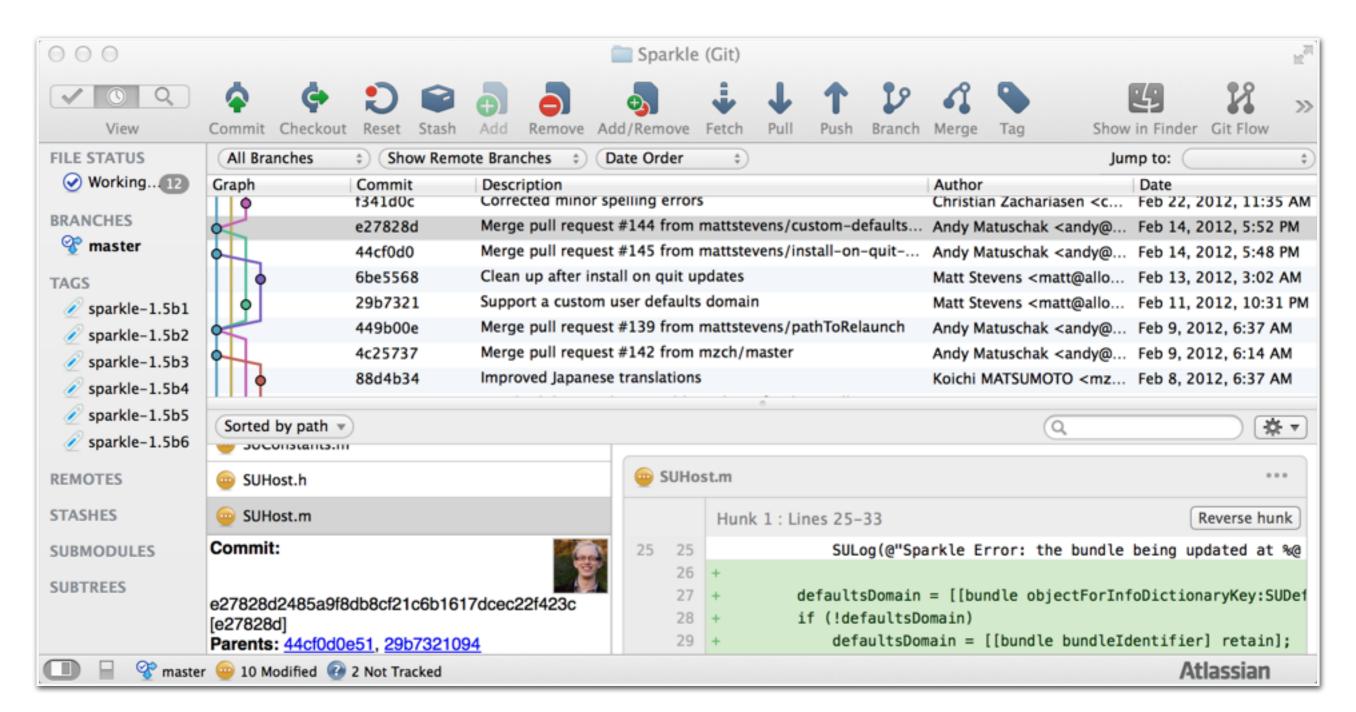
- \$ git push REMOTE BRANCH
- If the associated branch of the remote has been updated, your push will be rejected
- In this case, do a git pull (or fetch + merge) first and then do a new push
- Remote branches will only be moved forward, unless you force them to a specific commit (this makes it harder to rewrite history in a remote repo)

UI Helpers

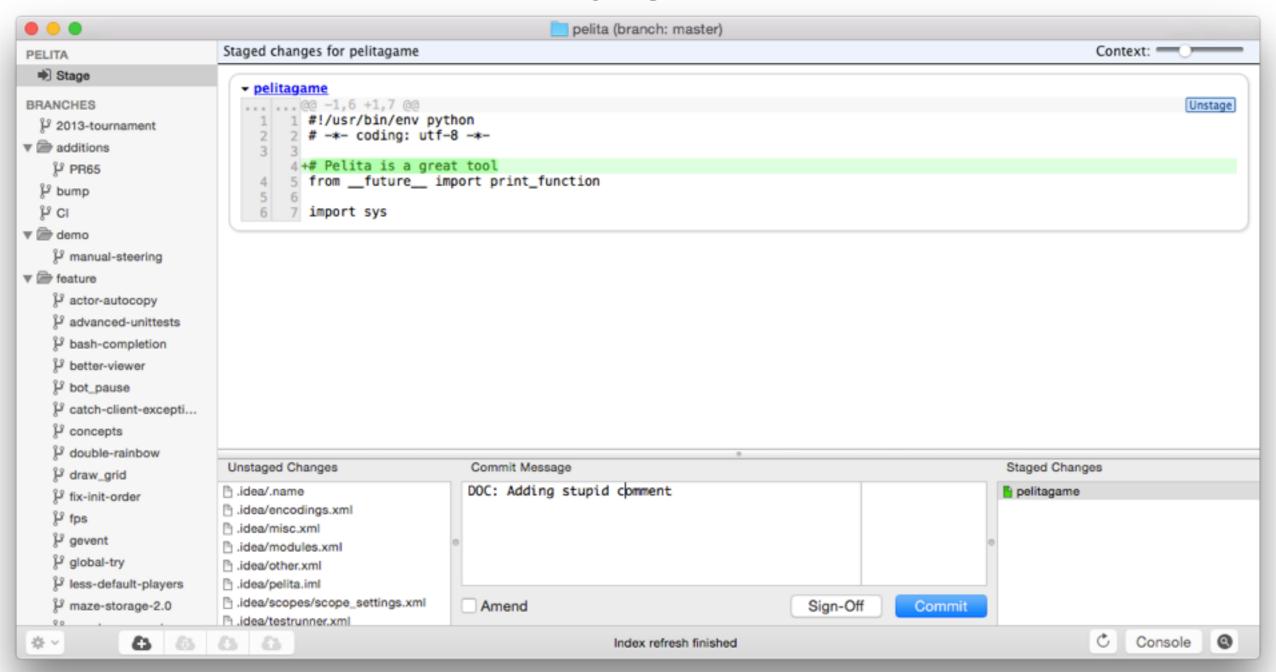
Sometimes there's a better way than using the command line

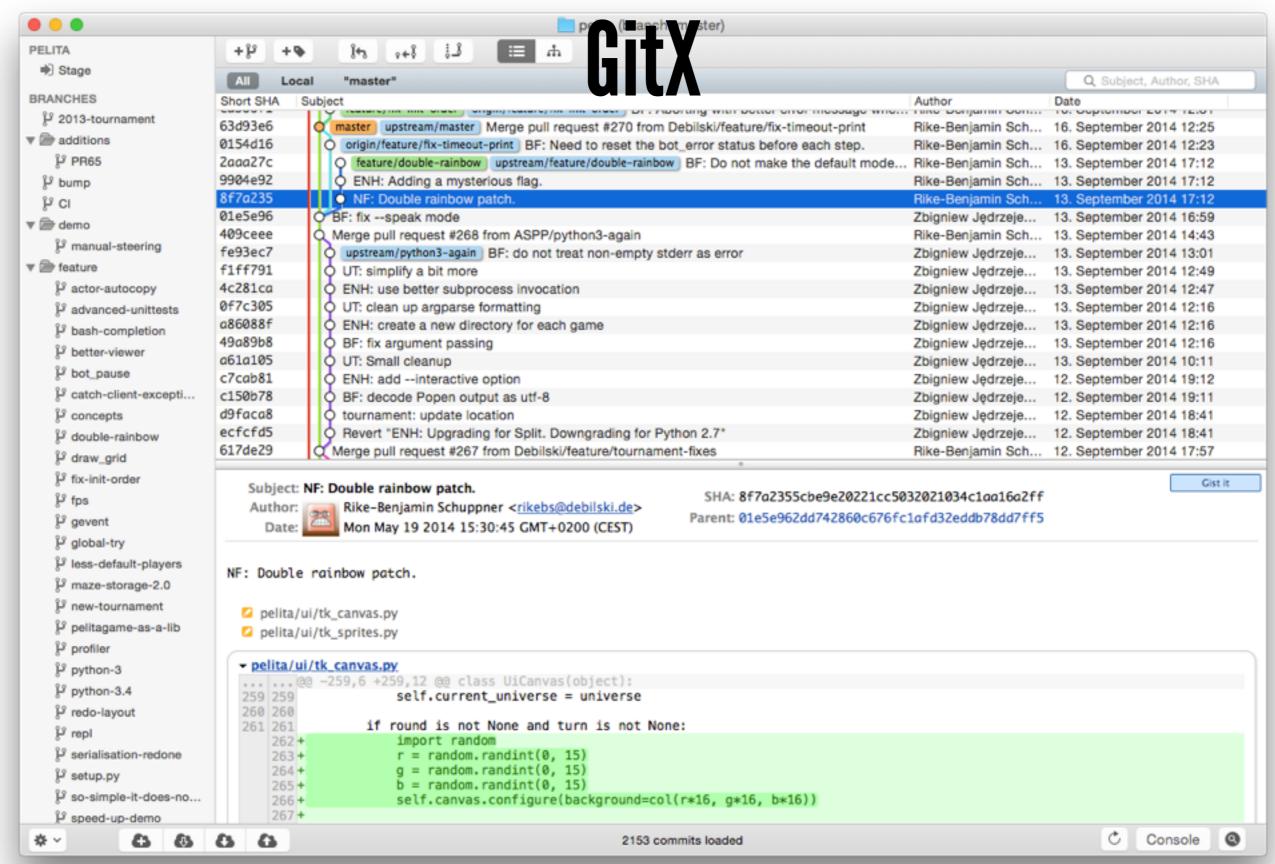
- In general it is very useful to understand the basic command line commands and also a bit of git internals
- A GUI may never catch all corner cases of what can happen
- Some GUIs look like they need the same amount of learning time than the CLI
- However, GUIs are very useful in two cases:
 - Staging chunks of code (à la git add -p)
 - Inspecting history

GUIS



GitX





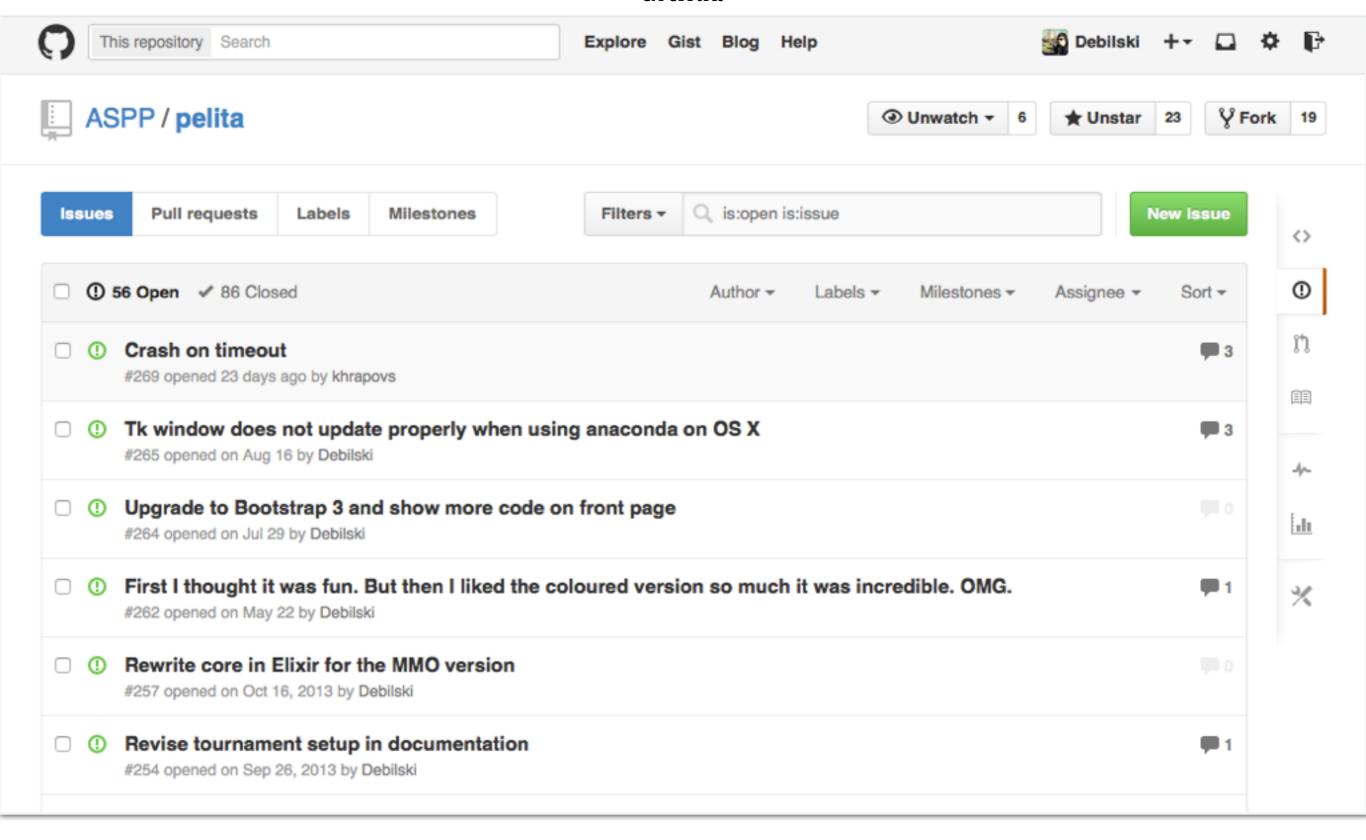
Collaborating with Github

- In principle, no external service is needed for using git
- Teams can have a 'blessed' repo on a server and everybody with ssh/unix access can write to it
- As simple as it gets. But eg. notification hooks must be made manually
- Services such as Github/Bitbucket (*) provide additional features for collaboration
 - (* or self-hosted instances à la GitLab/Gitbucket)

GitHub

- Individuals or teams own a repository on the server
- Collaborators fork the repository and submit a Pull Request (PR) when they have made changes
- In principle everybody with a GitHub account can suggest changes
- PRs can be discussed prior merging
- PRs will be merged with their complete history
- PR model is useful even if all members of a team have direct write access to main repo

GitHub



Advanced GitHub

- GitHub provides integration with https://travis-ci.org for automated testing/compiling of projects
- This can make sure that your code also runs in a 'standard environment'
- You'll get notified when a build breaks the test suite
- In principle doable with plain git hooks but usually no-one bothers



Help me, please

- Homepage: http://git-scm.com
- Reference to common commands: http://gitref.org
- Pro Git: http://git-scm.com/book

 (especially, if you're interested in the internals of git)
- \$ git help

Excuses for not using git

- The other people in my group surely are to stupid to understand git, why should I
 bother with it, then?
- My collaborators always send me whole files by email, so git does not work with me.

No excuses for not using git

- The other people in my group surely are to stupid to understand git, why should I
 bother with it, then?
- My collaborators always send me whole files by email, so git does not work with me.
- There is nothing stopping you from using git yourself and let other people do their work manually.