

Git and GitLab/GitHub

A Basic Guide to Version Control and Collaboration



Luke Hannan

20 February 2021

Outline

◆ About Git and GitHub/GitLab

? Why Git?

🔗 Key Terminology

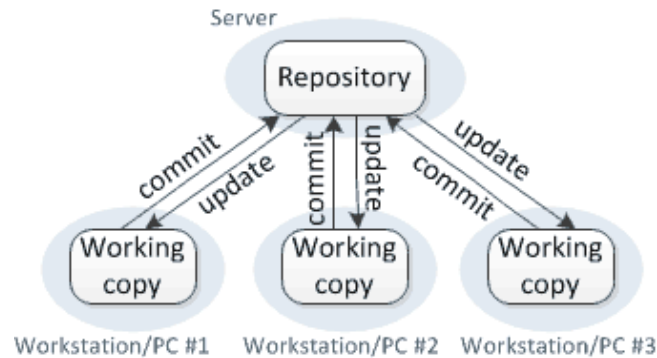
{ } How to Git Together

About Git

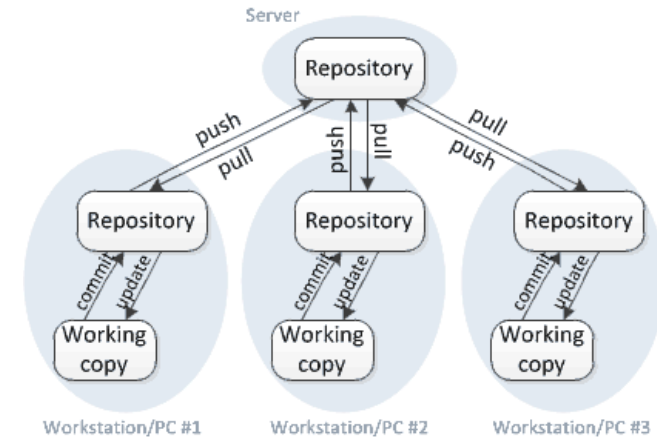
Git is a Distributed Version Control System (DVCS) designed to track changes in code.

It was designed to allow multiple collaborators to work on the same project; and to allow all collaborators to have the entire project's history on their local device without connecting to a central server.

Centralized version control



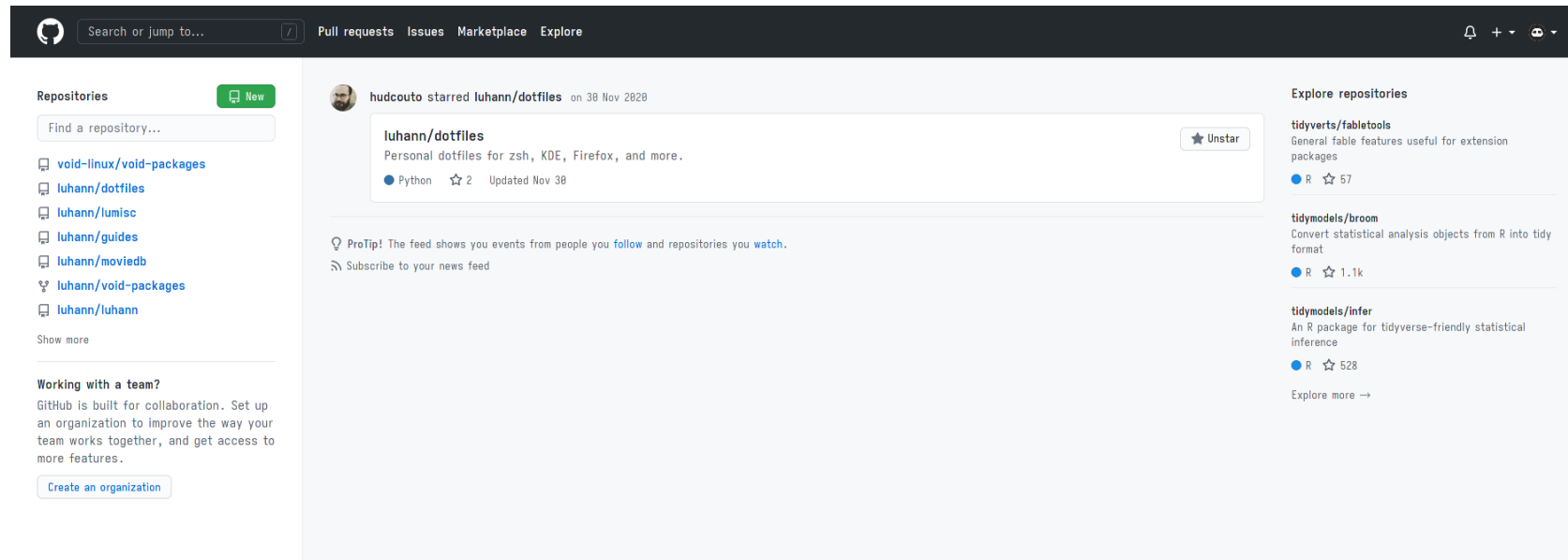
Distributed version control



Source: <https://homes.cs.washington.edu/~mernst/advice/version-control.html>

About GitHub and GitLab

- GitHub is a **web-based** service for storing, sharing, and collaborating on projects managed with Git.
- GitHub makes collaborating with other people much easier than using git alone.
- GitHub also provides an easy to use website for collaboration.



Why use Git?

- **Fast:** No network overhead, unlike centralized version control software.
- **Distributed:** Each user has the entire project history on their local machine.
- **Local:** You don't need internet to use it. You only need internet if you want to share your code with others (e.g., using GitLab).
- **Branches:** Keep your coding experiments separate from code that is already working.

Why use Git for *Research*?

Git is useful for researchers in a number ways. It can help:

- Protect yourself from accidental deletion:

```
# It's easy to make mistakes like these  
rm -rf *  
rm -rf ./
```

- If you accidentally **overwrote** important code to test something new.
- Perhaps you copied your code file to test new things:

```
# This isn't necessarily bad, but it's easy to lose track of all the changes that you make  
cp working_code.R new_function_test.R
```

- Finally, Git also helps others to collaborate on your **code**.

Key Words to Understanding Git

Key Terminology for using Git and GitLab*

Repository is where you keep all the files you want to track. The convention is usually create a repository per project (aka repo).

Commit is an object that holds information about a particular change. Think of **commits** as the snapshots of your work.

HEAD refers to the most recent commit on the current branch.

Branch is the name for a separate line of development, with its own history.

* We will go over these terms again later.

Key Terminology for using Git and GitLab*

Pull: When using an online repo (like GitLab), Pull will download all changes from the online repo and automatically merge them into your current project.

Push: The opposite of pull, git push will transfer all changes you made on your machine to the online repo.

Fork: Forking is a GitLab thing (not git) and allows you to copy a repo to your account and make changes without it affecting the original.

Merge: Copy your fork back to the original repo and add your changes to it.

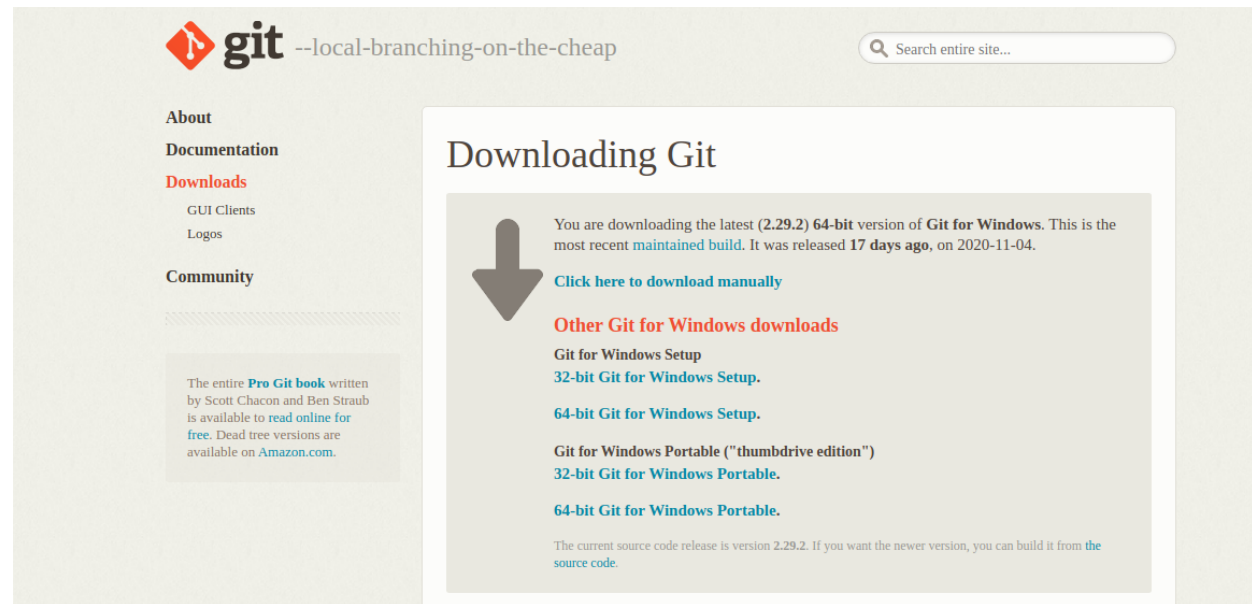
* All of the terms in this slide refer to GitHub/GitLab terms and can be ignored if you are not using them.

How to install Git

How to install Git

On Windows

- Go to the [windows download page on the Git website](#)
- Download the version of Git you want:

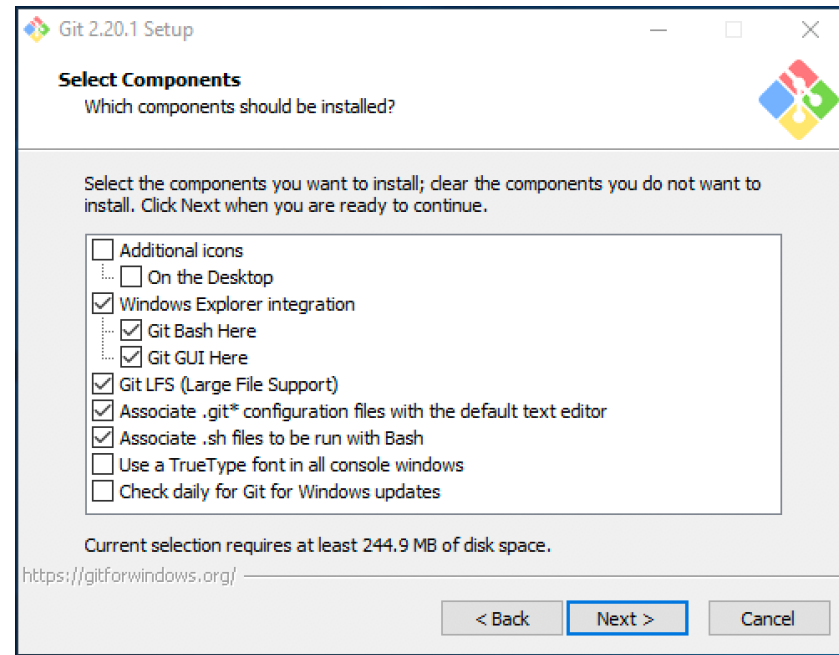


Source: <https://www.linode.com/docs/guides/how-to-install-git-on-linux-mac-and-windows/>

How to install Git

On Windows

- After running installer, select the components you want to install (if you are unsure, go ahead with the default selection):

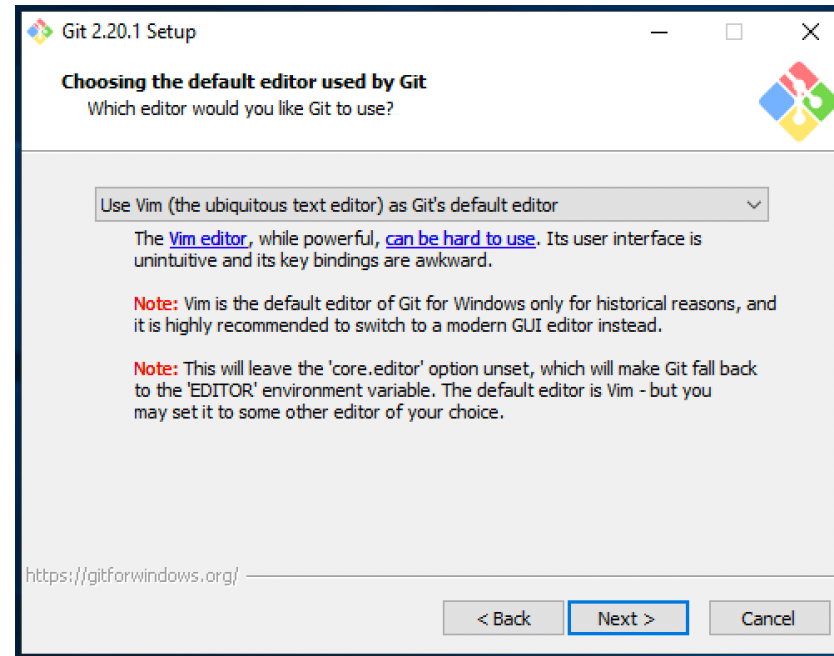


Source: <https://www.linode.com/docs/guides/how-to-install-git-on-linux-mac-and-windows/>

How to install Git

On Windows

- Choose the default editor for Git, I recommend you change this to your editor of choice:

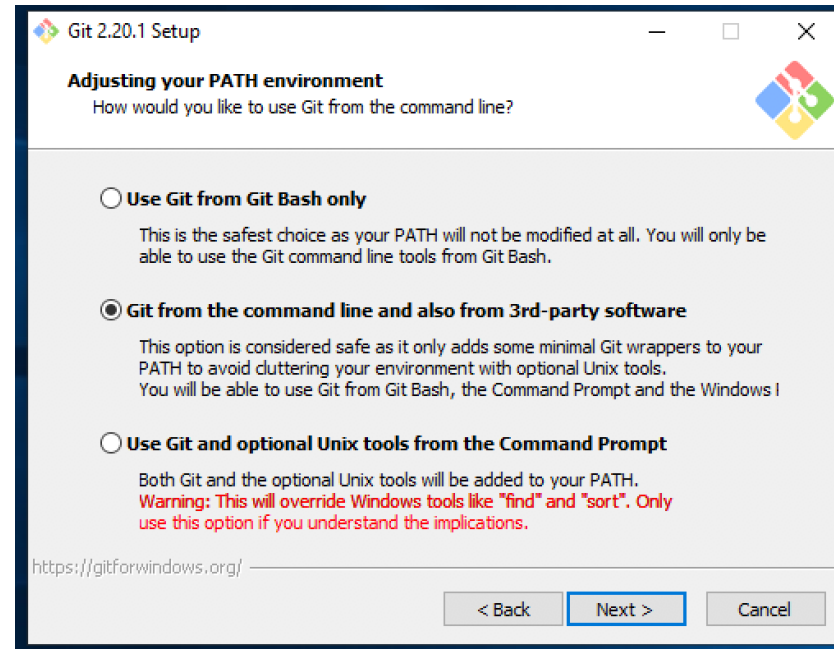


Source: <https://www.linode.com/docs/guides/how-to-install-git-on-linux-mac-and-windows/>

How to install Git

On Windows

- Choose how to use Git, the middle option is the safest:

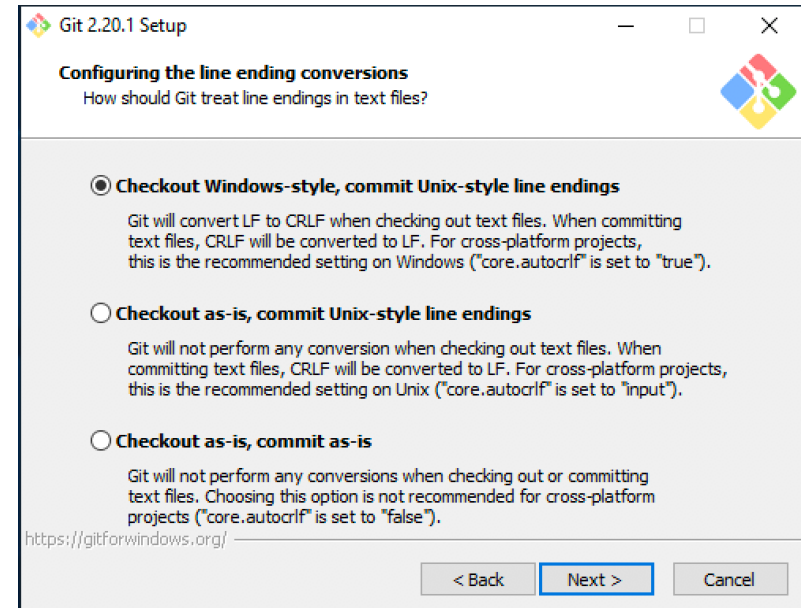
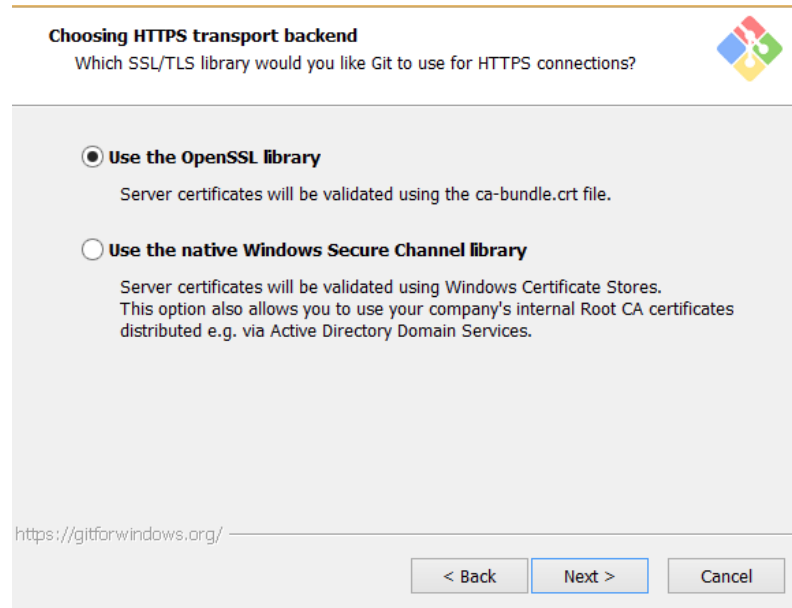


Source: <https://www.linode.com/docs/guides/how-to-install-git-on-linux-mac-and-windows/>

How to install Git

On Windows

- The rest of the default options are fine to select:



Source: <https://www.linode.com/docs/guides/how-to-install-git-on-linux-mac-and-windows/>

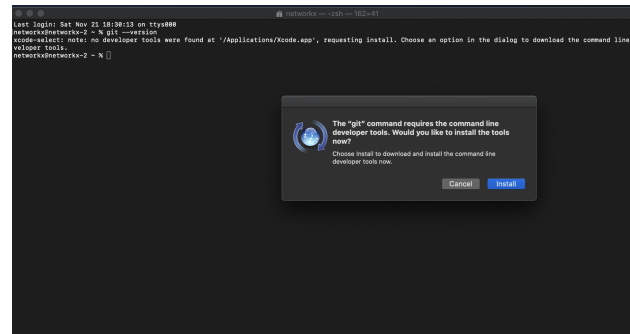
How to install Git

On Mac

- Check if Git is already installed on your Mac using:

```
# In terminal type  
git --version
```

- If Git isn't installed on your computer, the terminal prompts you with the following message:



- Click on Install

Source: <https://www.linode.com/docs/guides/how-to-install-git-on-linux-mac-and-windows/>

How to install Git

On Linux (Ubuntu)

- In a terminal, update first:

```
# In terminal type  
sudo apt update
```

- Then install Git:

```
# In terminal type  
sudo apt install git
```

- Check if Git installed correctly:

```
# In terminal type  
git --version
```

How to use Git

How to use Git and GitLab when you are collaborating with others

Setting Up Git

This is only done once

- Git needs to know who you are to assign authorship to your commits.

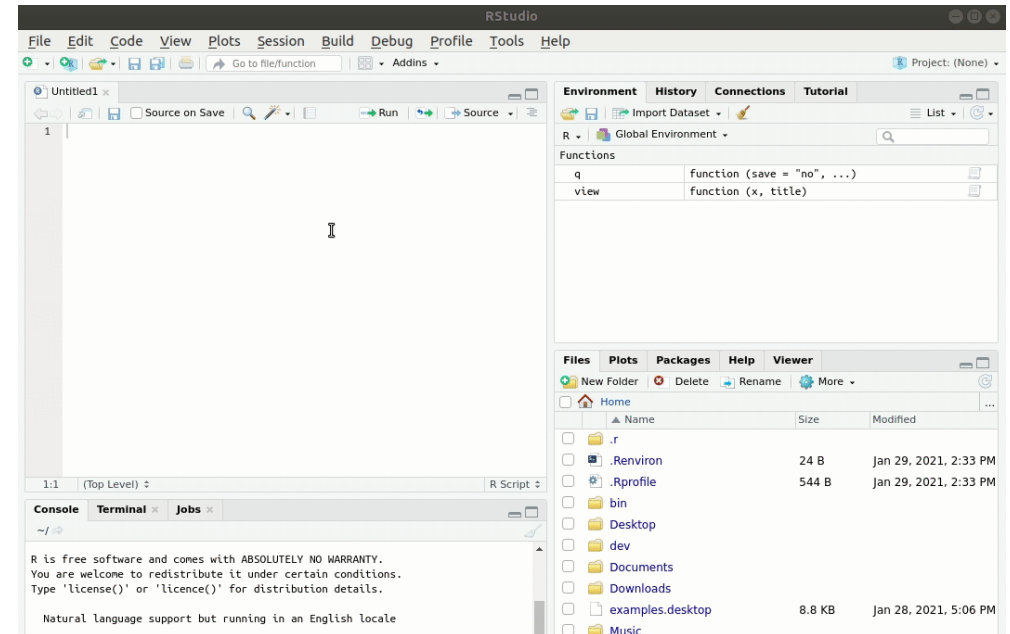
```
# Tell git your name  
git config --global user.name "Dr. Henry Jekyll"  
  
# Tell git your email  
git config --global user.email jekyll@example.com
```

Typical Git Workflow

Creating a Repository

- First we initialize the repository. This creates an empty repository that allows us to add files and folders to be tracked by git.

```
# Move to your project directory  
cd new_project  
# Initialize repository  
# This is only done once at the start  
# of each project  
git init .  
  
# At any point you can see that status of  
# your git repository  
git status
```



Typical Git Workflow

Adding Files

- Once we have created our new repo, we need to tell git which files to track.

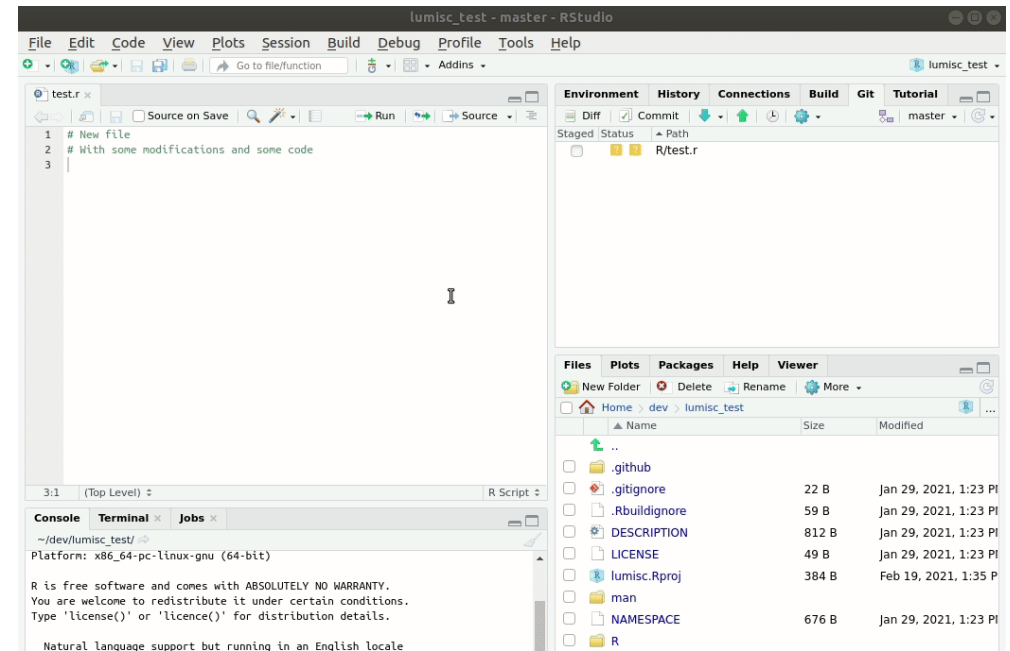
```
# In your project directory
git add R/test.r

# Adding files "stages" them telling git
# we want to track them

# See our added files
git status

# "staged" files need to be "committed" for
# the changes to be permanent
git commit R/test.r

# You can commit files with a message as well
git commit R/test.r --message "An informative
```



Typical Git Workflow

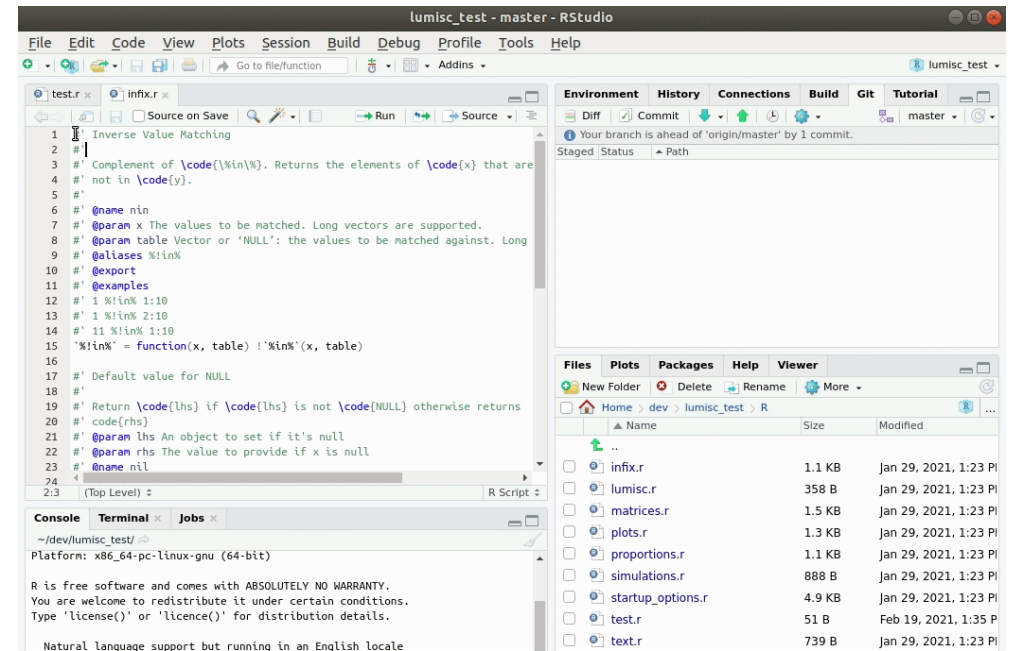
Changing Existing Files

- Once files have been added and committed, we can make changes that are permanently tracked.

```
# First edit or change Files

# See what files have changed
git status

# Commit our newly changed files
git commit R/infix.r
# or with a message
git commit R/infix.r --message "We changed an
```



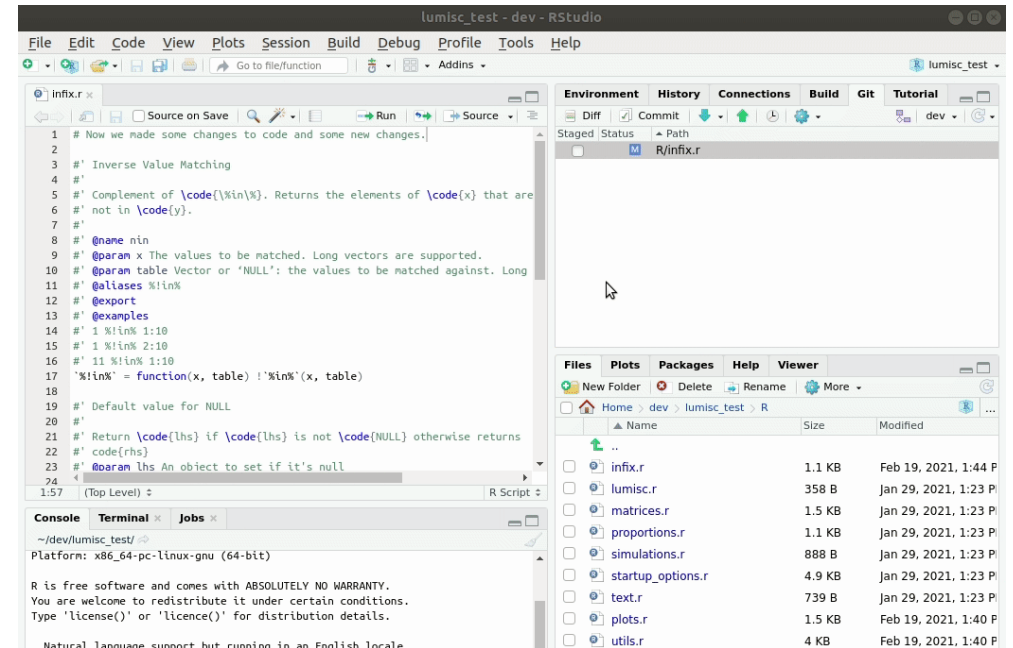
Typical Git Workflow

Reverting Changes

- Perhaps you have **staged** some changes but you don't want them anymore and want to revert the file to its original state.

```
# See any potentially changed Files
git status

# Revert changes to latest commit
git revert R/infix.r
```



Typical Git Workflow

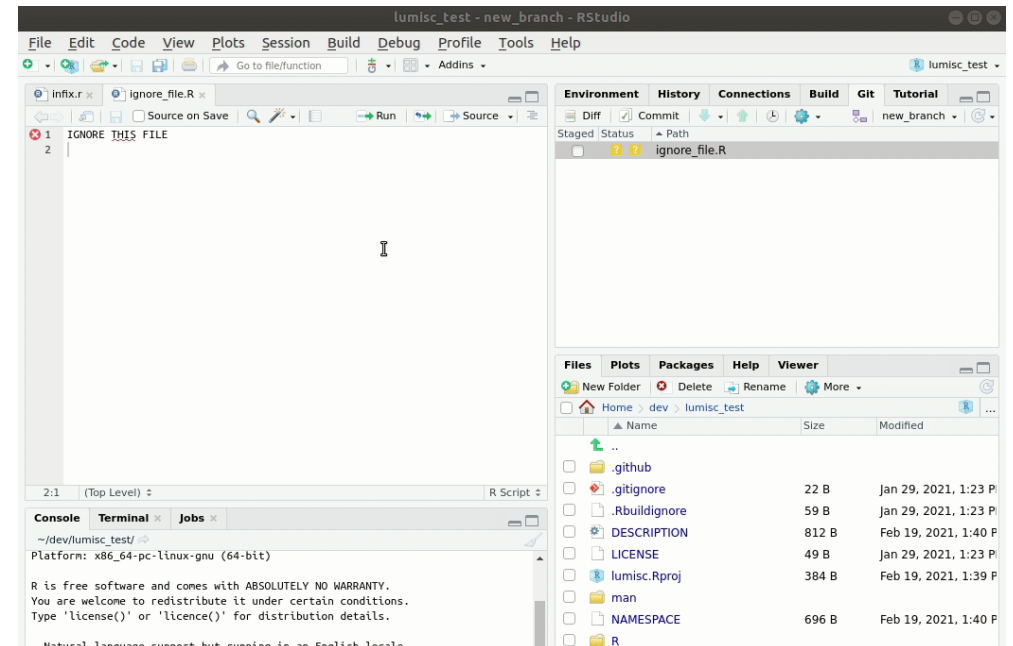
Ignoring Certain Files

- Some files in your project will either be too unwieldy (very large datasets) or not necessary (output files generated by code) for Git to track, but we can tell git to ignore these files with a special file the **.gitignore** file. Ignored files will not be tracked by git at all.

```
# See any potentially changed Files
git status

# Make the .gitignore file in any way
# you choose
echo "ignore_file.R" >> .gitignore

# Make sure to commit
# the .gitignore file
git add .gitignore
git commit .gitignore --message "Add gitignor
```



Typical Git Workflow

Creating a New Branch

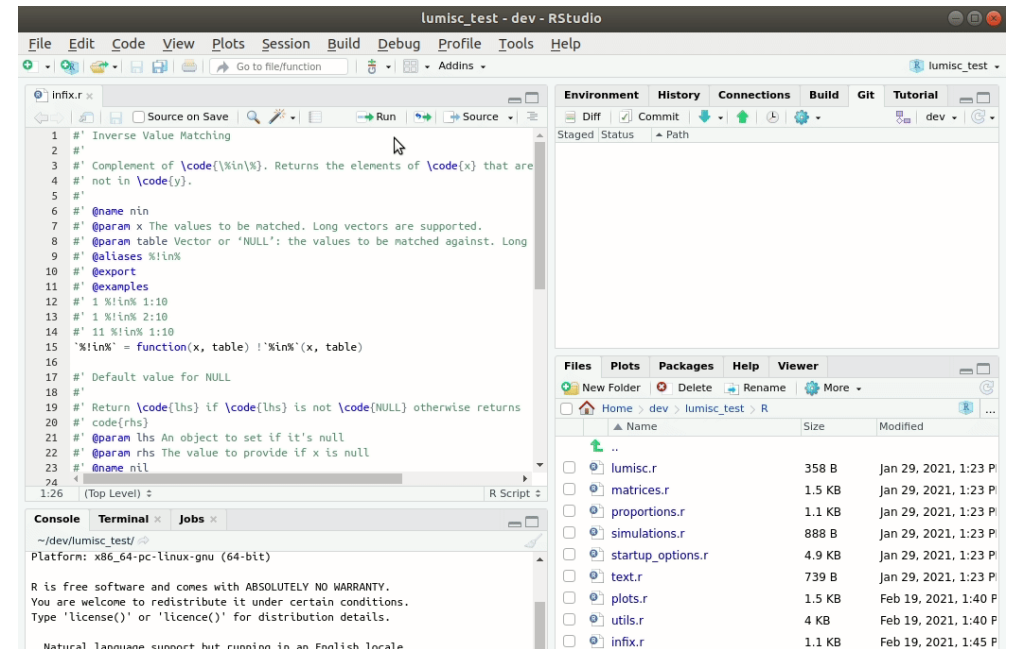
- Git makes it easy for you to test new code or change old code without having to copy/paste or save duplicate files. This is done by creating branches. A branch is an exact copy from the start of the project until the point it was created.

```
# See what files have changed  
git status
```

```
# Commit our newly changed files  
git branch new_branch
```

```
# To see all branches  
git branch -a
```

```
# Once we are done working on our new  
# branch we can delete it  
git branch -d new_branch
```



Typical Git Workflow

Switching Between Branches

- Once we have created between branches we can easily* switch between them.

To see all branches

`git branch -a`

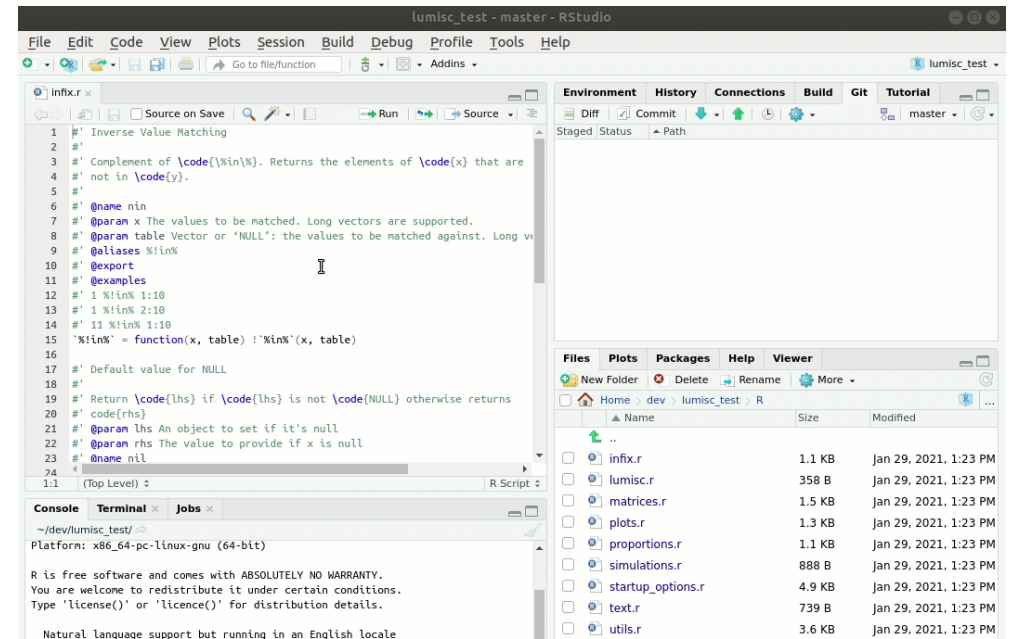
If you have the latest version of git

`git switch new_branch`

If you have an older version of git

Older than git 2.23

`git checkout new_branch`



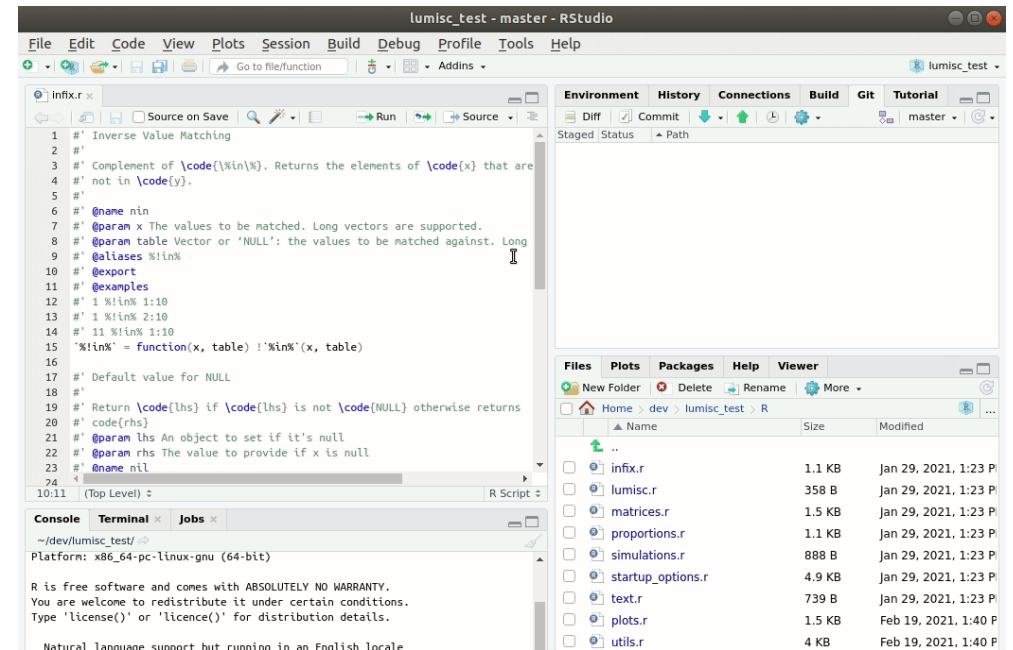
Typical Git Workflow

Viewing Your History

```
# See the entire commit history  
# of your current branch  
git log
```

```
# View some statistics for each commit  
git log --stat
```

```
# Compress history to oneline for  
# each commit for easier reading  
git log --pretty=oneline
```



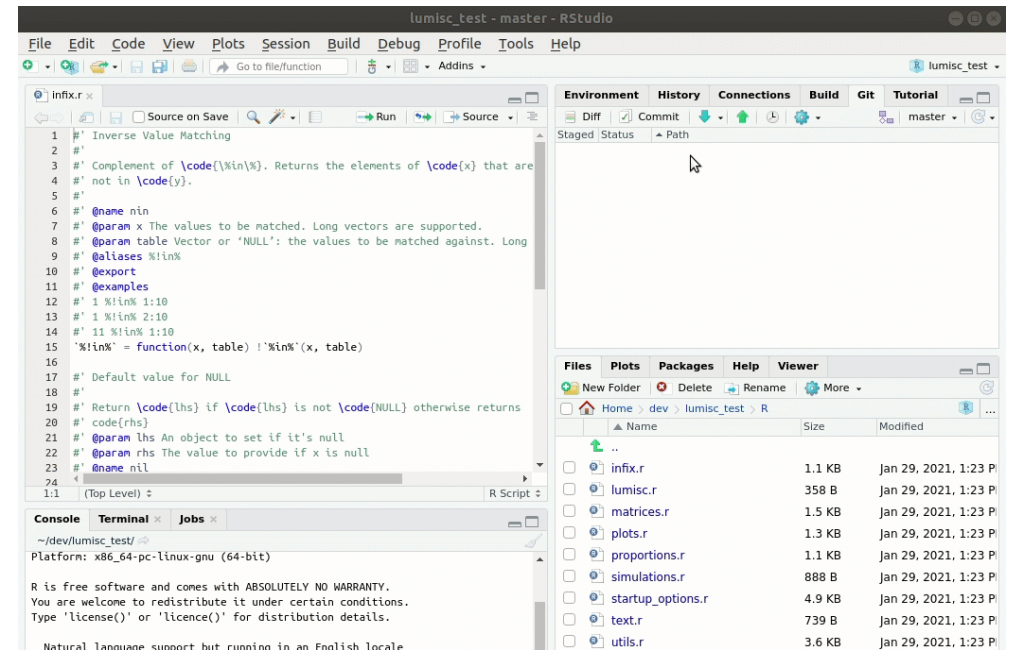
Typical GitHub/GitLab Workflow

Updating Your Local Repository

- This only applies when you are working with a git repository that is stored on some server or cloud service (like **GitHub** or **GitLab**). This is usually done **once per work session** before you start making changes to make sure your local version is the same as the online repository.

```
# Pull the latest changes from the remote  
# branch  
git pull
```

```
# Sometimes git will warn you that you  
# have to pick what to do if there  
# are conflicts between changed files  
# in your local repository and changed files  
# in the remote repository  
git pull --ff-only  
# Will only pull if there are no conflicts  
# otherwise it will error out
```



Typical GitHub/GitLab Workflow

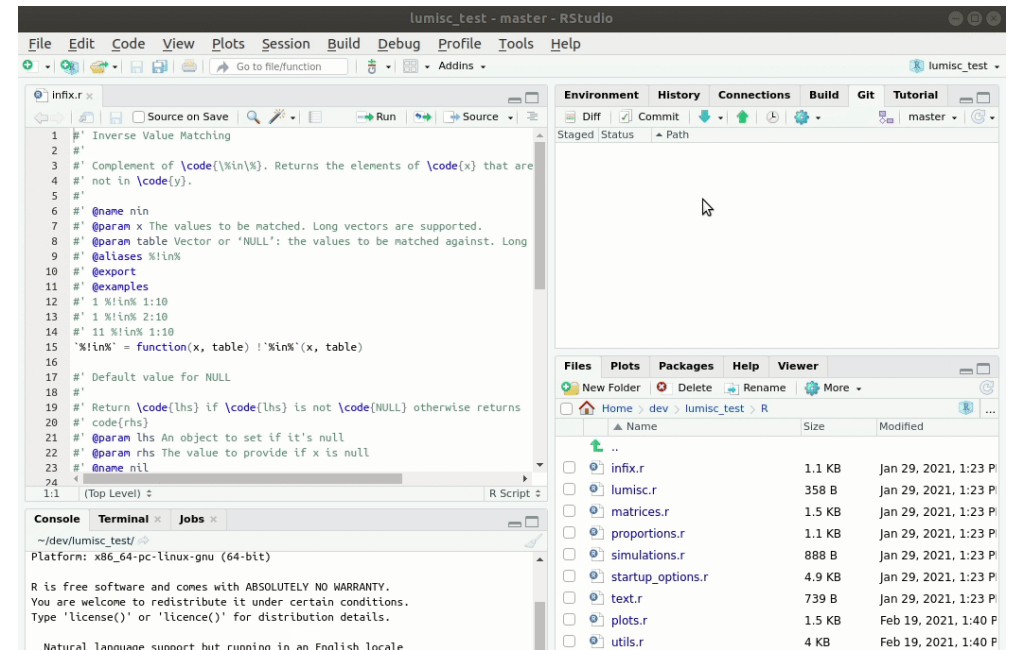
Updating The Remote Repo (i.e GitHub)

- Again this only applies when dealing with a repo stored on a cloud server or GitHub. This is also usually only done **once per work session** at the end of your work session when you want to make sure your changes are recorded on the remote server.

```
# Push the latest changes to the remote
# repo
git push

# If there are conflicts between your
# local repo and the remote repo Git will
# error and give you a chance to
# fix the errors
git push -f

# Will overwrite the remote repo with
# your local changes and is usually
# not recommended
```



Advanced Git

HERE BE DRAGONS

THESE SLIDES ARE ONLY FOR REFERENCE

Advanced Git Tricks

That you should almost never have to use

- Ask me, **Luke**, about any complicated or tricky Git Stuff.
- You can go back to anypoint in your git history.
 - Find the **commit hash** using git log.

```
git checkout "2b1d1b15f46ad71b59"
```

- It can be useful to export/archive all the code in a given git repository.

```
git archive -o filename.zip branch_name
```

End

Thank You for Your Time

Glossary

Version Control: any system that allows you to understand the history of the file and how it has progressed

Git: a version control program which allows you to annotate the changes you make to create an easily traversable system history

Commit: an annotated “snapshot” of the differences made to the system at a given point in time

Local: refers to the computer you’re working on this very minute

Remote: refers to an online location

Repository (repo): a special folder configured with Git superpowers containing all the files pertaining to your project/system

Glossary

Github: takes your local commit history and hosts it remotely so that you can access it from any computer

Pushing: the action of taking local Git commits (and whatever work these encompass) and putting them online on Github

Pulling: the action of taking online Github commits and bringing them into your local machine

Master (branch): the “trunk” of the commit history “tree”; contains all approved content/code

Feature branch: an isolated location, based off of master, where you can write a new piece of work safely before reincorporating said changes back to master

Pull Request: a Github tool that allows users to easily see the changes (the difference or “diff”) that a feature branch is proposing as well as discuss any tweaks that said branch might require before it is merged into master

Merging: the action of taking the commits from a feature branch and adding them to the top of master’s history

Checking out: the action of moving from one branch to another

Extra Resources for Learning

- [Git Basics](#)
- [Git Visualized](#)
- [Interactive Git Tutorial](#)