



Intro to Git and GitHub

What is Git?

Version Control for **local machines**

Used by software programmers and data scientists to keep track of changes to their files

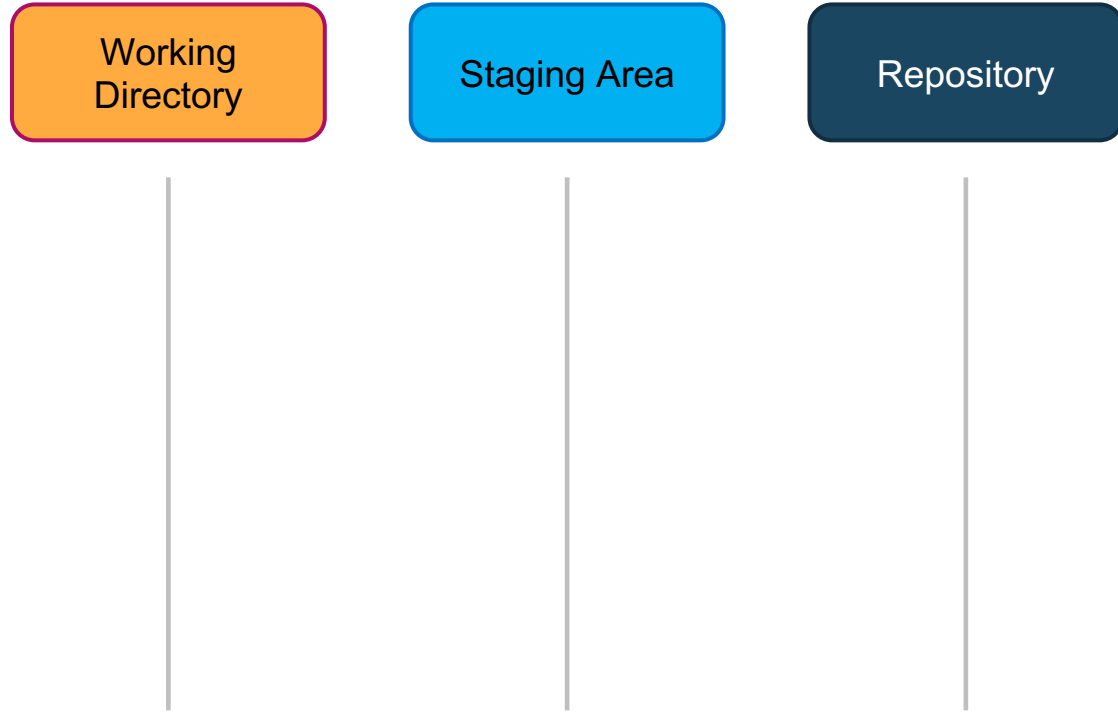


Chapter 1: Git Repositories

Learning Objectives

- Explain and describe the three phases of code tracking in Git, including:
 - Working Directory
 - Staging Area
 - Repository
- Explain how the command ``git status`` works

Git has three phases of code tracking



The working directory

The working directory area is our “normal” file system. This tracks the code as it exists on our computer right now.

If we’ve made changes, then the working directory sees all of those changes.

It doesn’t track any history. It just says, “this is how my code looks right now.”



Working
Directory

A diagram consisting of an orange rounded rectangle with a dark red border, containing the text 'Working Directory'. A thin vertical grey line extends downwards from the bottom center of the rectangle.

Git status

The repository is where all of our checkpoints live.

Run the command ``git status`` to display all the files that have been modified in the working directory since the last time changes were saved.



Repository



Chapter 2: Git Commits

Learning Objectives

- Recognize that Git repos are changed using **commits**
- Explain the two steps when git commits happen including:
 - git **add**
 - git **commit**

The staging area

The staging area is how we prepare for a **commit**.

Before making a commit, we'll tell git that we want certain files to be “staged” for the **commit**.

In non-git terms, we're telling git that we want it to check this file to see if it's changed since the last checkpoint.

A blue rounded rectangle labeled "Staging Area" is connected by a vertical grey line to the text "In non-git terms, we're telling git that we want it to check this file to see if it's changed since the last checkpoint." in the text block below.

Staging Area

The repository

Every time we make a **commit**, all of those changes are pushed into the repository.

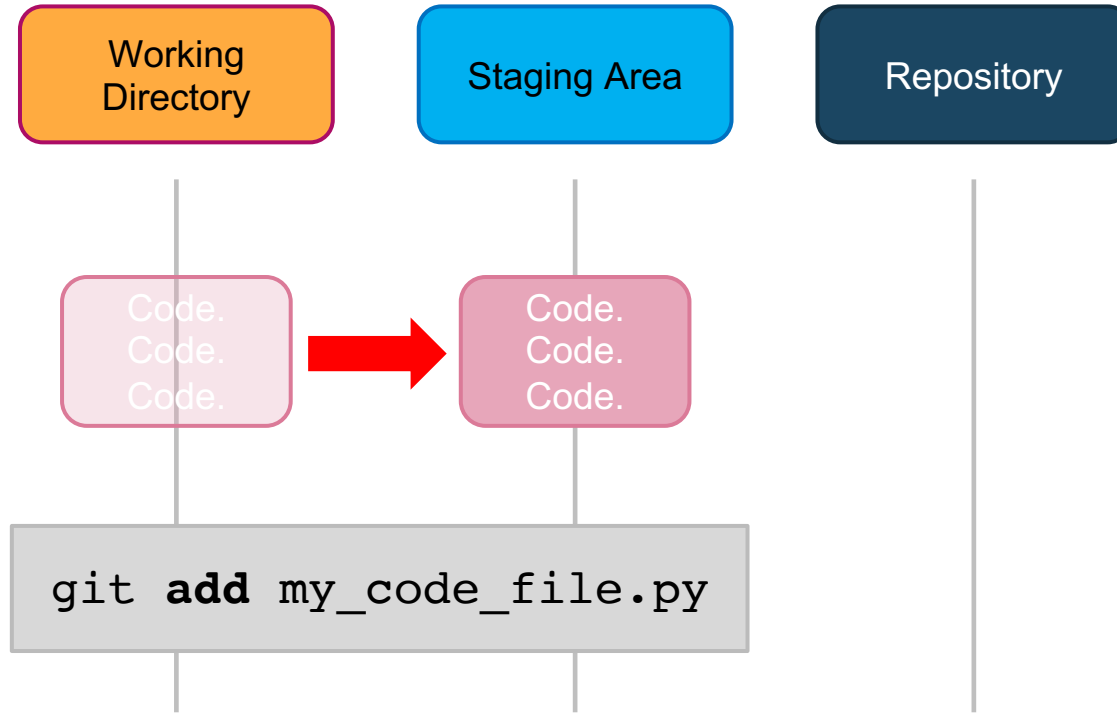
The repository is the holder of all knowledge about each version of a project.

Until the code is **committed** to the repository, no checkpoints have been made.

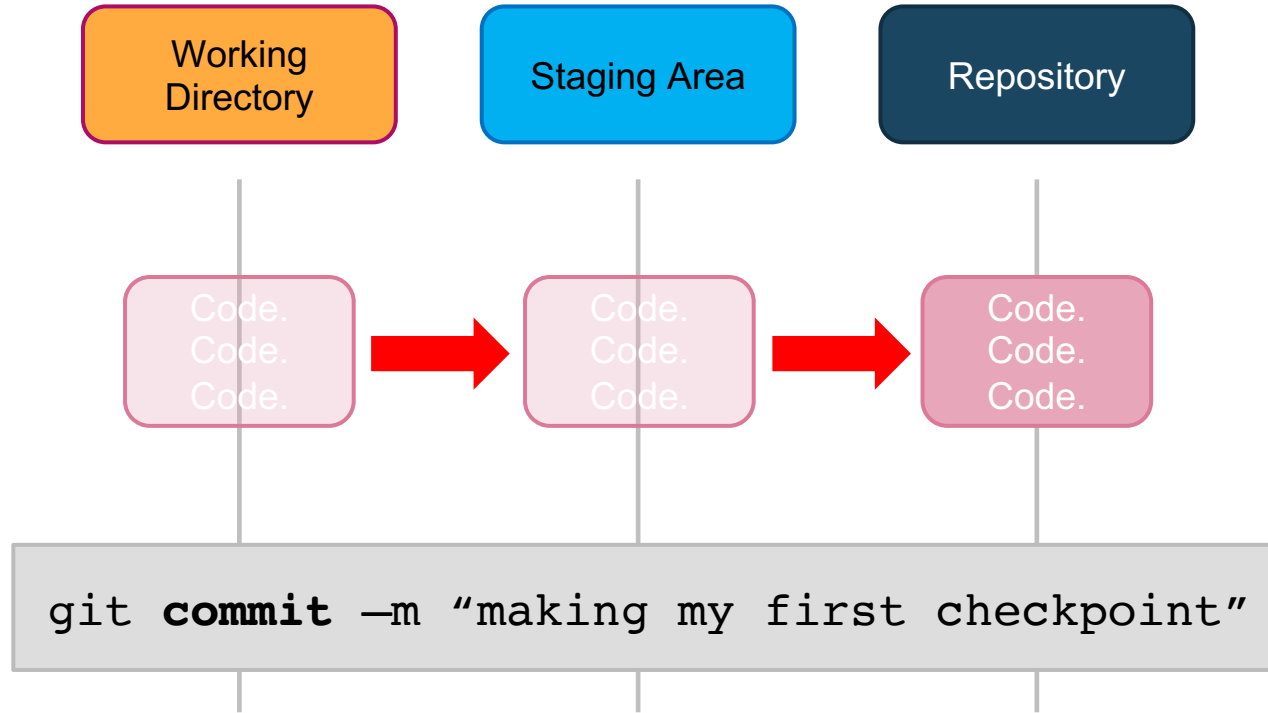


Repository

Step 1: Add files to staging area



Step 2: Commit all files in staging area



Git commit messages should be clear

```
git commit -m "making my first checkpoint"
```

Commit messages should:

- Clearly state what's changed since the last commit
- Tell what bugs were fixed, and what new code was added

Git commit messages should be clear

```
git commit -m "making my first checkpoint"
```

If we remove the **-m** "message here" part from the previous example, we will be taken to a text editor to make our commit messages instead and can make it more specific.

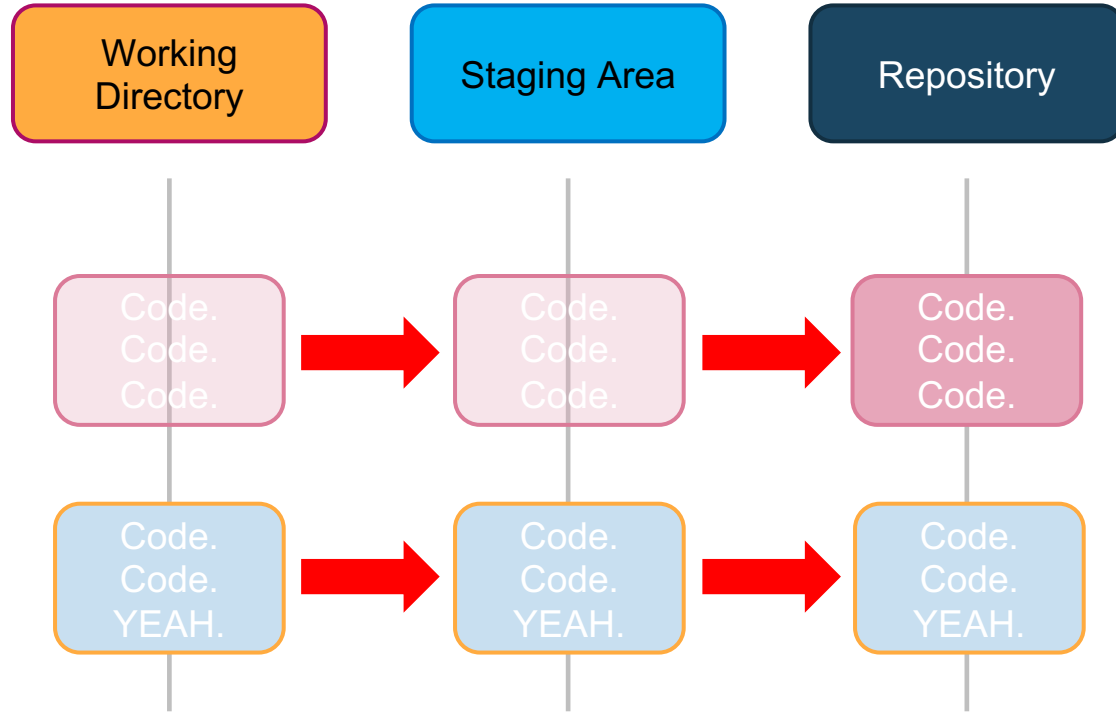


Chapter 3: Git Reset

Learning Objectives

- Describe how to revert to a previous Git stage using Git **reset**
- Explain what **git log** can be used for

How can we revert to a previous commit?



Use Git Log

```
zachariahmill@metis-zachariahmill~/Documents/PythonProjects$ git log
commit 8861cc02c62df5946d760a0618d349f6b4481f64 (HEAD -> master, origin/master)
Author: zwmiller <za[REDACTED].com>
Date:   Sun Sep 2 11:29:57 2018 -0500

    breakout network analysis to own repo

commit 128fc8819d8433971302cc94f5f3db7af08d4c51
Author: zwmiller <za[REDACTED].com>
Date:   Sun Sep 2 11:23:46 2018 -0500

    refactor code, fix sorting bug, redraw demo, add testing and test data

commit f27dbf83405f876643ffef270f11a1a21ab672a9
Author: zwmiller <za[REDACTED].com>
Date:   Fri Aug 31 14:30:39 2018 -0500

    add docstrings

commit 1dcaad09c98c3f8db2732db134b2e2a7a65ec681
Author: zwmiller <za[REDACTED].com>
Date:   Fri Aug 31 14:22:11 2018 -0500

    refactor plot making code
```

Commit
IDs



Use Git Log

```
zachariahmill@metis-zachariahmill~/Documents/PythonProjects$ git log
commit 8861cc02c62df5946d760a0618d349f6b4481f64 (HEAD -> master, origin/master)
Author: zwmiller <za[REDACTED].com>
Date:   Sun Sep 2 11:29:57 2018 -0500
```

breakout network analysis to own repo

```
commit 128fc8819d8433971302cc94f5f3db7af08d4c51
Author: zwmiller <za[REDACTED].com>
Date:   Sun Sep 2 11:23:46 2018 -0500
```

refactor code, fix sorting bug, redraw demo, add testing and test data

```
commit f27dbf83405f876643ffef270f11a1a21ab672a9
Author: zwmiller <za[REDACTED].com>
Date:   Fri Aug 31 14:30:39 2018 -0500
```

add docstrings

```
commit 1dcaad09c98c3f8db2732db134b2e2a7a65ec681
Author: zwmiller <za[REDACTED].com>
Date:   Fri Aug 31 14:22:11 2018 -0500
```

refactor plot making code

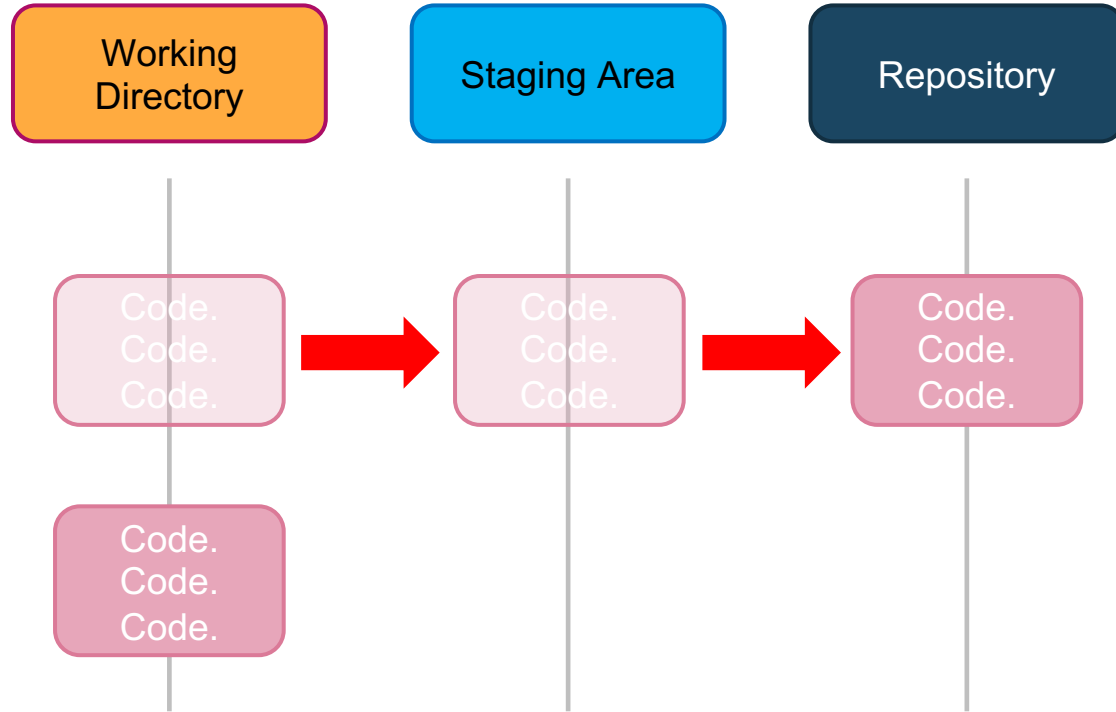
Commit
Messages

A diagram consisting of four red arrows pointing from the commit messages to the text 'Commit Messages'. The arrows originate from the commit messages: 'breakout network analysis to own repo', 'refactor code, fix sorting bug, redraw demo, add testing and test data', 'add docstrings', and 'refactor plot making code'. They all converge towards the 'Commit Messages' text on the right side of the image.

Use Git Reset

```
git reset 128fc8819d8433971302cc94f5f3db7af08d4c51
```

After Git Reset





Chapter 4: GitHub

Learning Objectives

- Recognize that GitHub extends Git online
- Recognize the components of a GitHub repository
- Explain how to import a GitHub repository onto a local machine using the **git clone** command

What is GitHub?

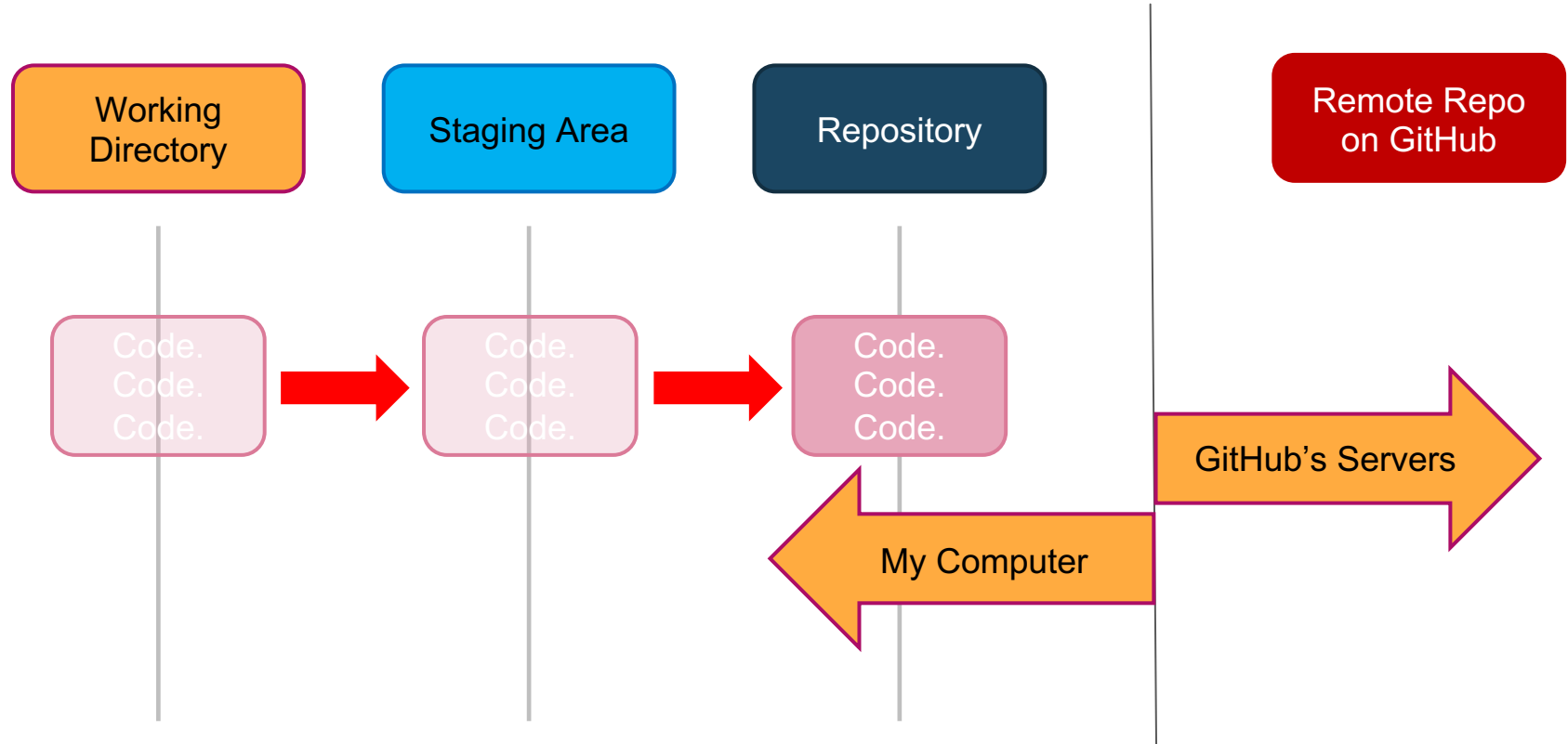
Online Version Control

Back up Git local machine changes

Allow for shared collaboration



GitHub extends Git online



GitHub Example: Scikit-learn

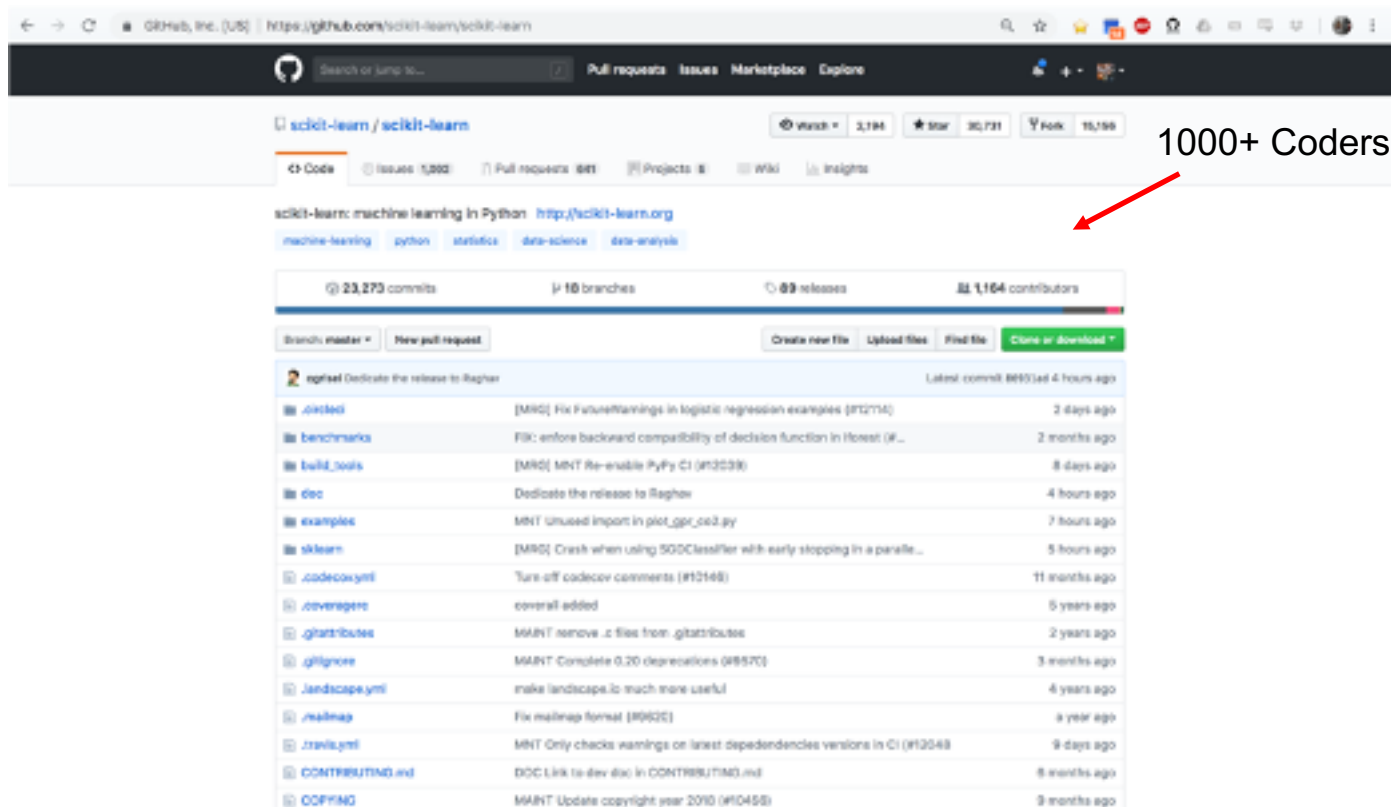
The screenshot shows the GitHub repository page for `scikit-learn/scikit-learn`. The repository is a Python machine learning library. It has 2,194 forks, 30,731 stars, and 15,156 forks. The repository is described as "machine learning in Python" and includes tags for "machine-learning", "python", "statistics", "data-science", and "data-analysis".

Key statistics shown include 23,273 commits, 10 branches, 89 releases, and 1,164 contributors. The repository is currently on the `master` branch. A "New pull request" button is visible.

A list of recent commits is displayed, including:

- `fixes`: Fix FutureWarnings in logistic regression examples (#12714) - 2 days ago
- `benchmarks`: Fix: enforce backward compatibility of decision function in forest (#12714) - 2 months ago
- `build_tools`: MNT Re-enable PyPy CI (#12538) - 8 days ago
- `doc`: Dedicate the release to Raghu - 4 hours ago
- `examples`: MNT unused import in plot_gpr_coe2.py - 7 hours ago
- `sklearn`: [MNT] Crash when using SGDClassifier with early stopping in a paralle... - 5 hours ago
- `codecov.yml`: Turn off codecov comments (#12148) - 11 months ago
- `coveragers`: coverall added - 5 years ago
- `gitattributes`: MANT remove .z files from .gitattributes - 2 years ago
- `gitignore`: MANT Complete 0.20 deprecations (#8570) - 3 months ago
- `landscape.yml`: make landscape.io much more useful - 4 years ago
- `mailmap`: Fix mailmap format (#9632) - a year ago
- `travis.yml`: MNT Only checks warnings on latest dependencies versions in CI (#12548) - 9 days ago
- `CONTRIBUTING.md`: DOC link to dev doc in CONTRIBUTING.md - 8 months ago
- `COPYING`: MANT Update copyright year 2010 (#10458) - 9 months ago

Scikit-learn has many contributors



The screenshot shows the GitHub repository for scikit-learn/scikit-learn. The repository has 3,194 forks, 30,731 stars, and 15,198 forks. The repository description is "scikit-learn: machine learning in Python" with the URL "http://scikit-learn.org". The repository statistics show 23,273 commits, 10 branches, 69 releases, and 1,164 contributors. A red arrow points to the "1,164 contributors" text, which is accompanied by the text "1000+ Coders".

scikit-learn: machine learning in Python <http://scikit-learn.org>

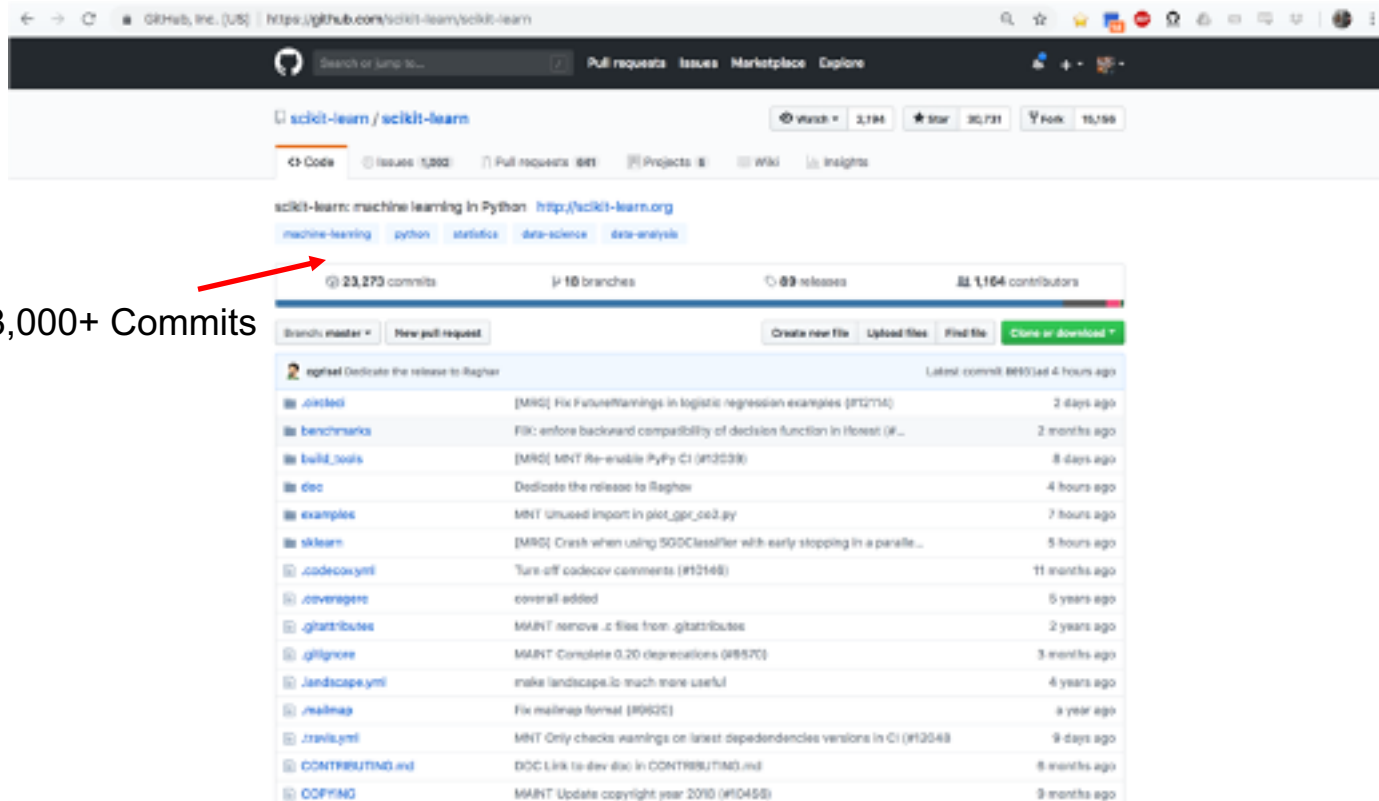
machine-learning python statistics data-science data-analysis

23,273 commits 10 branches 69 releases 1,164 contributors

branch: master New pull request Create new file Upload files Pin file Clone or download

commit	message	latest commit
agf1ed	[MRG] Fix FutureWarnings in logistic regression examples (#12714)	2 days ago
benchmarks	Fix: enforce backward compatibility of decision function in forest (#...	2 months ago
build_tools	[MRG] MNT Re-enable PyPy CI (#12538)	8 days ago
doc	Dedicate the release to Raghu	4 hours ago
examples	MNT unused import in plot_gpr_coe2.py	7 hours ago
sklearn	[MRG] Crash when using SGDClassifier with early stopping in a paralle...	5 hours ago
.codecov.yml	Turn off codecov comments (#12148)	11 months ago
.coveragerc	coverall added	5 years ago
.gitattributes	MANT remove .c files from .gitattributes	2 years ago
.gitignore	MANT Complete 0.20 deprecations (#9570)	3 months ago
landscape.yml	make landscape.io much more useful	4 years ago
.mailmap	Fix mailmap format (#9622)	a year ago
.travis.yml	MNT Only checks warnings on latest dependencies versions in CI (#12048)	9 days ago
CONTRIBUTING.md	DOC Link to dev doc in CONTRIBUTING.md	8 months ago
COPYING	MANT Update copyright year 2010 (#10455)	9 months ago

GitHub keeps track of all contributors' commits

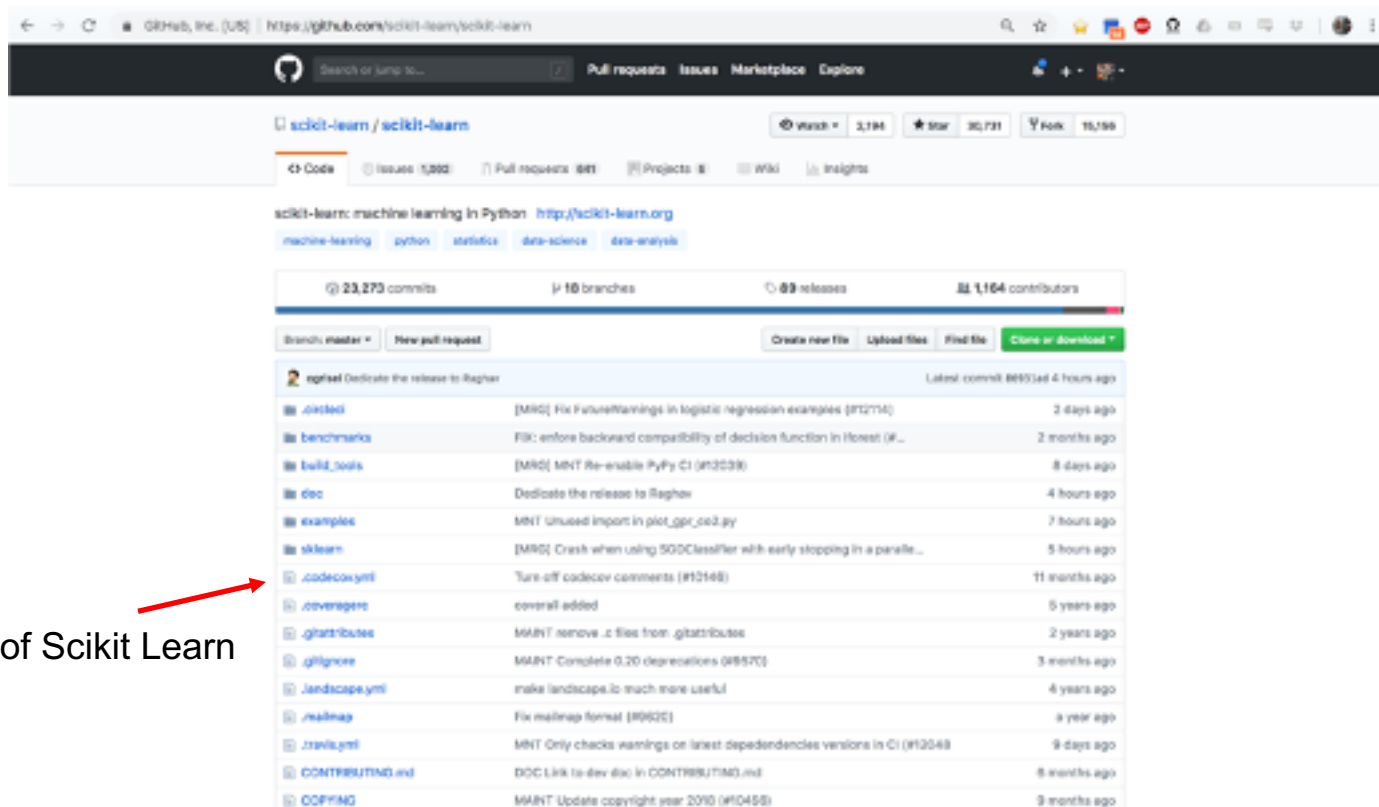


The screenshot shows the GitHub repository page for `scikit-learn/scikit-learn`. The repository has 2,194 stars, 30,731 forks, and 15,166 pulls. It is described as "machine learning in Python" with a link to <http://scikit-learn.org>. The repository statistics show 23,273 commits, 10 branches, 69 releases, and 1,164 contributors. A red arrow points to the "23,273 commits" text. Below the statistics, a table lists recent commits with their titles, commit hashes, and timestamps.

23,000+ Commits

Commit Hash	Commit Message	Timestamp
8b551ed	Dedicate the release to Raghu	4 hours ago
[MRG]	Fix FutureWarnings in logistic regression examples (#12111)	2 days ago
[MRG]	Fix: enforce backward compatibility of decision function in forest (#12110)	2 months ago
[MRG]	MNT Re-enable PyPy CI (#12030)	8 days ago
	Dedicate the release to Raghu	4 hours ago
	MNT unused import in plot_gpr_co2.py	7 hours ago
[MRG]	Crash when using SGDClassifier with early stopping in a paralle...	5 hours ago
	Turn off codecov comments (#12148)	11 months ago
	coverage added	5 years ago
	MANT remove .x files from .gitattributes	2 years ago
	MANT: Complete 0.20 deprecations (#9570)	3 months ago
	make landscape.io much more useful	4 years ago
	Fix mailing format (#10605)	a year ago
	MNT Only checks warnings on latest dependencies versions in CI (#12048)	9 days ago
	DOC Link to dev doc in CONTRIBUTING.md	8 months ago
	MANT Update copyright year 2010 (#10458)	9 months ago

GitHub stores every file that has been committed



scikit-learn / scikit-learn

forks 2,194 stars 30,731 pulls 15,158

Code Issues 1,202 Pull requests 641 Projects 0 Wiki Insights

scikit-learn: machine learning in Python <http://scikit-learn.org>

machine-learning python statistics data-science data-analysis

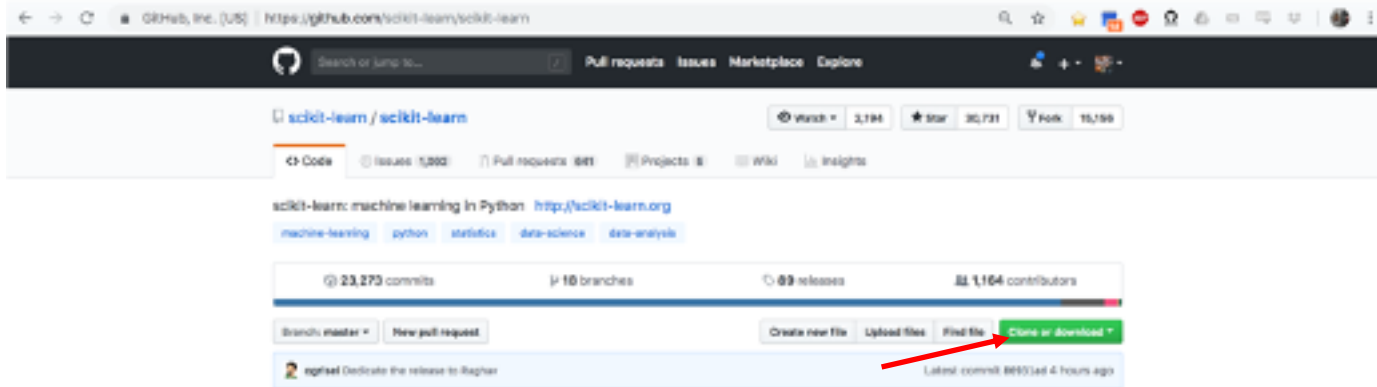
23,273 commits 10 branches 69 releases 1,164 contributors

branch: master New pull request Create new file Upload files Pin file Clone or download

Latest commit: 8f551ad 4 hours ago	
.gitignore [MRG] Fix FutureWarnings in logistic regression examples (#12714)	2 days ago
benchmarks FIX: enforce backward compatibility of decision function in forest (F...	2 months ago
build_tools [MRG] MNT Re-enable PyPy CI (#12538)	8 days ago
doc Dedicate the release to Raghu	4 hours ago
examples MNT unused import in plot_gpr_coe2.py	7 hours ago
sklearn [MRG] Crash when using SGDClassifier with early stopping in a paralle...	5 hours ago
.codecov.yml Turn off codecov comments (#12148)	11 months ago
.coveralls coverall added	5 years ago
.gitattributes MAINT remove .x files from .gitattributes	2 years ago
.gitignore MAINT Complete 0.20 deprecations (#9570)	3 months ago
.landscape.yml make landscape.io much more useful	4 years ago
.mailmap Fix mailmap format (#9622)	a year ago
.travis.yml MNT Only checks warnings on latest dependencies versions in CI (#12048)	9 days ago
CONTRIBUTING.md DOC Link to dev doc in CONTRIBUTING.md	8 months ago
COPYING MAINT Update copyright year 2010 (#10455)	9 months ago

All of Scikit Learn

Git Clone



```
git clone github.com/scikit-learn/scikit
```

<code>.sklearn</code>	[MRG] Crash when using SGDClassifier with early stopping in a paralle...	5 hours ago
<code>.codecov.yml</code>	Turn off codecov comments (#12148)	11 months ago
<code>.coveragerc</code>	coverage added	5 years ago
<code>.gitattributes</code>	MAINT remove .x files from .gitattributes	2 years ago
<code>.gitignore</code>	MAINT Complete 0.20 deprecations (#8570)	3 months ago
<code>.landscape.yml</code>	make landscape.io much more useful	4 years ago
<code>.mailmap</code>	Fix mailmap format (#9505)	a year ago
<code>.travis.yml</code>	MNT Only checks warnings on latest dependencies versions in CI (#12048)	9 days ago
<code>CONTRIBUTING.md</code>	DOC Link to dev doc in CONTRIBUTING.md	8 months ago
<code>COPYING</code>	MAINT Update copyright year 2010 (#10458)	9 months ago



Chapter 5: Git Push and Git Pull

Learning Objectives

- Describe how to move changes from Git to GitHub using **git push**
- Explain how to move changes from GitHub to Git using **git pull**

Syncing between Git and GitHub

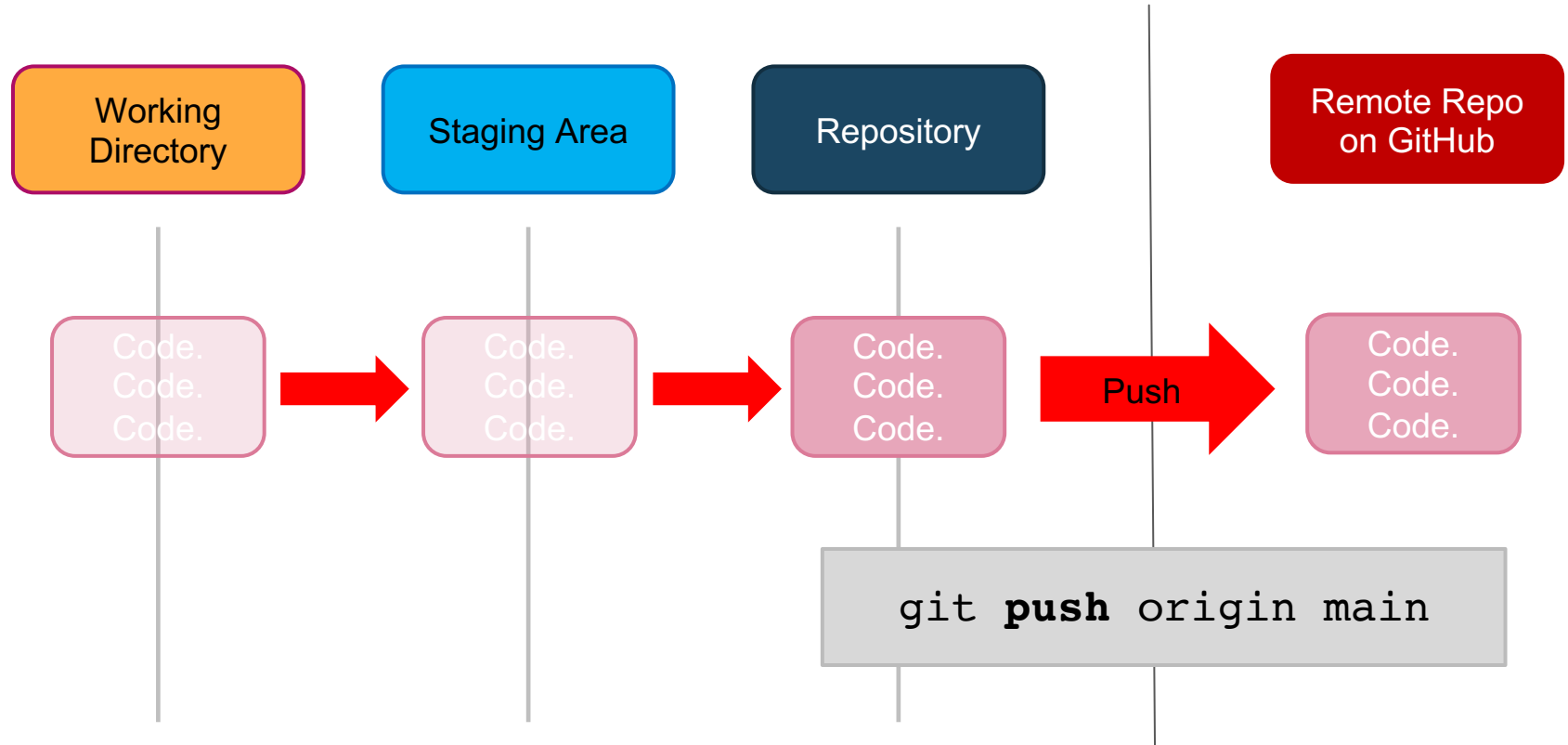
Git push:

“Make the remote version look like my code” and update any files that changed in the last commits

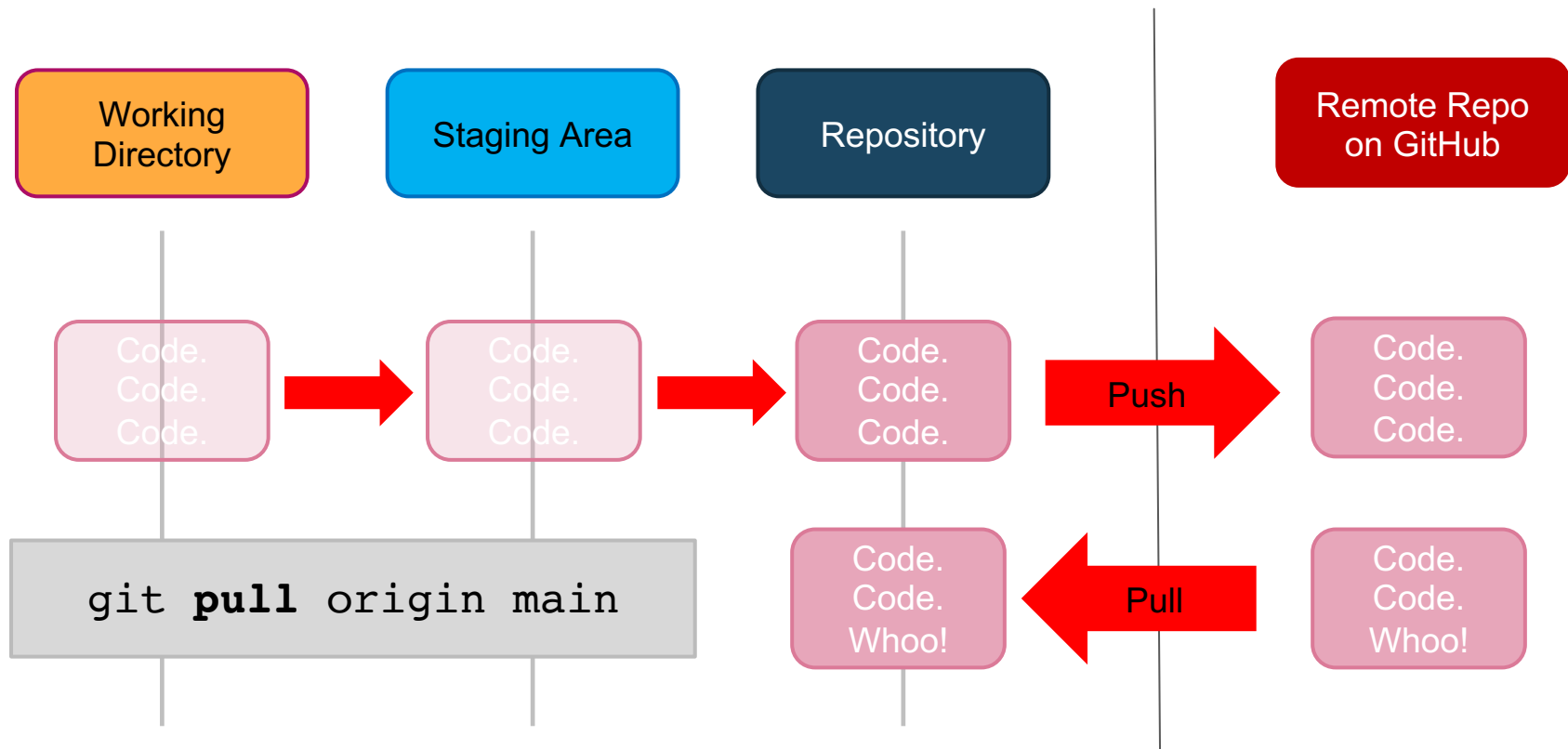
Git pull:

“Bring in any changes that have occurred on the remote version of the repo and make my code match those”

Git Push



Git Push



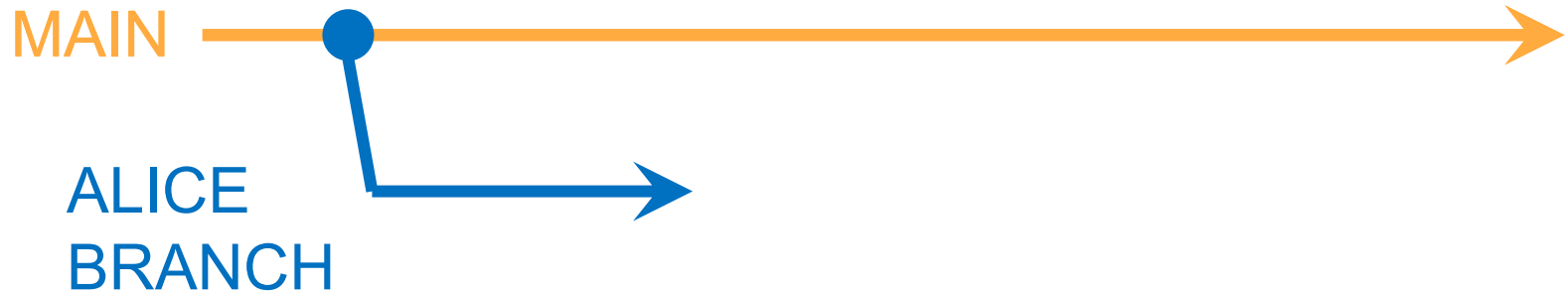


Chapter 6: Git Branches

Learning Objectives

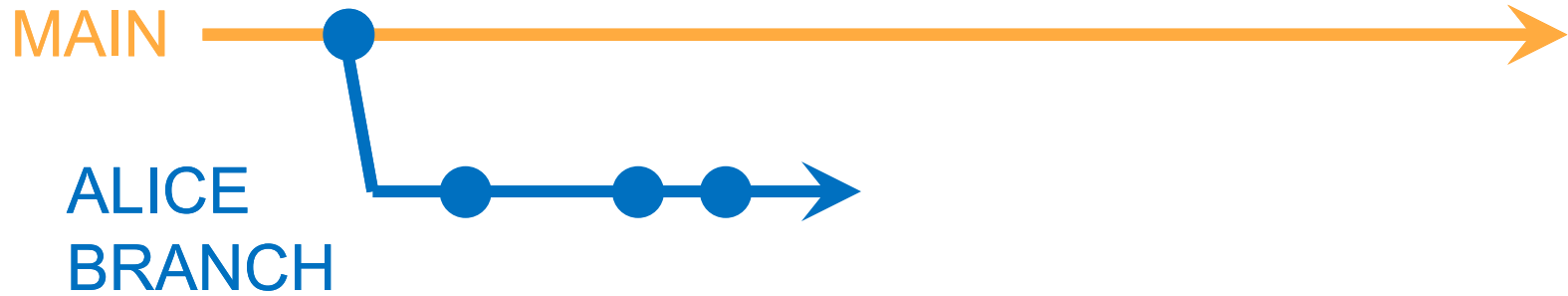
- Recognize the **main** branch of a Git repo
- Create new branches using **git branch**
- Switch between branches using **git checkout**

Git branches allow for individual code changes



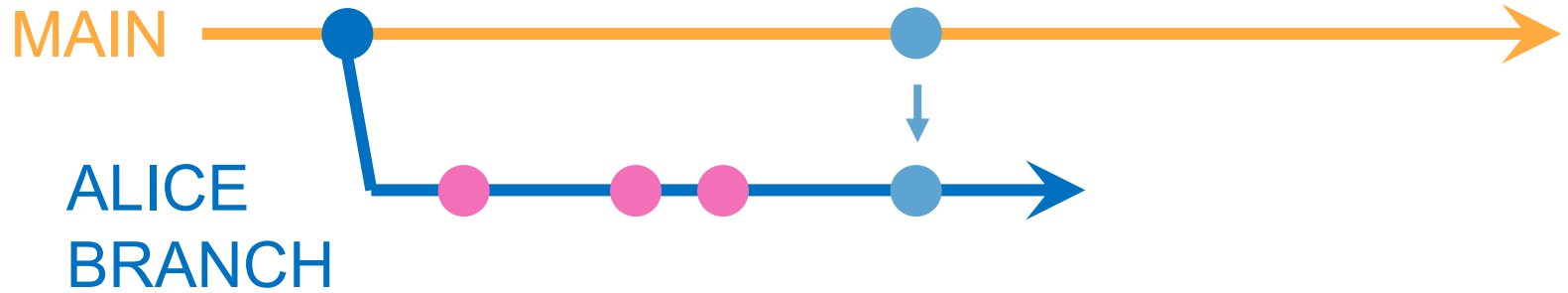
```
git branch alice_branch  
git checkout alice_branch
```

Git branches allow for individual code changes



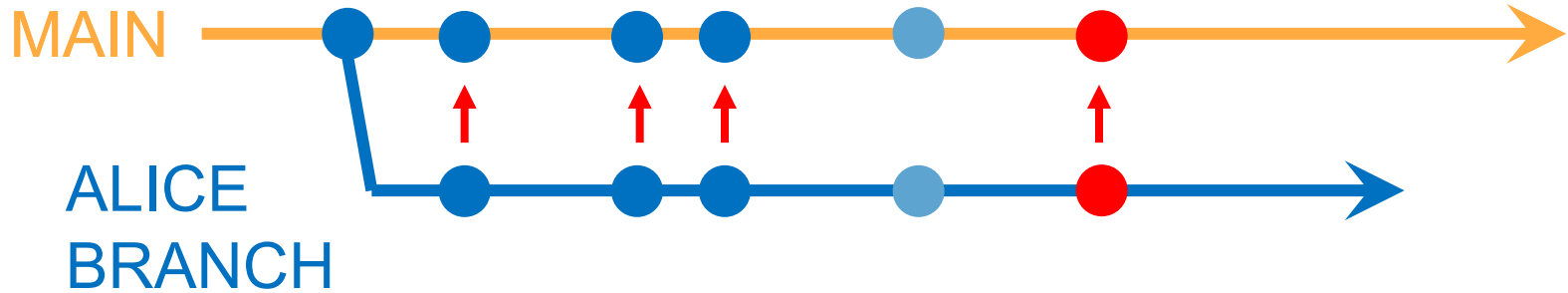
Alice can make whatever changes she wants on her own branch. She can make [commits](#) to her branch.

Git branches allow for individual code changes



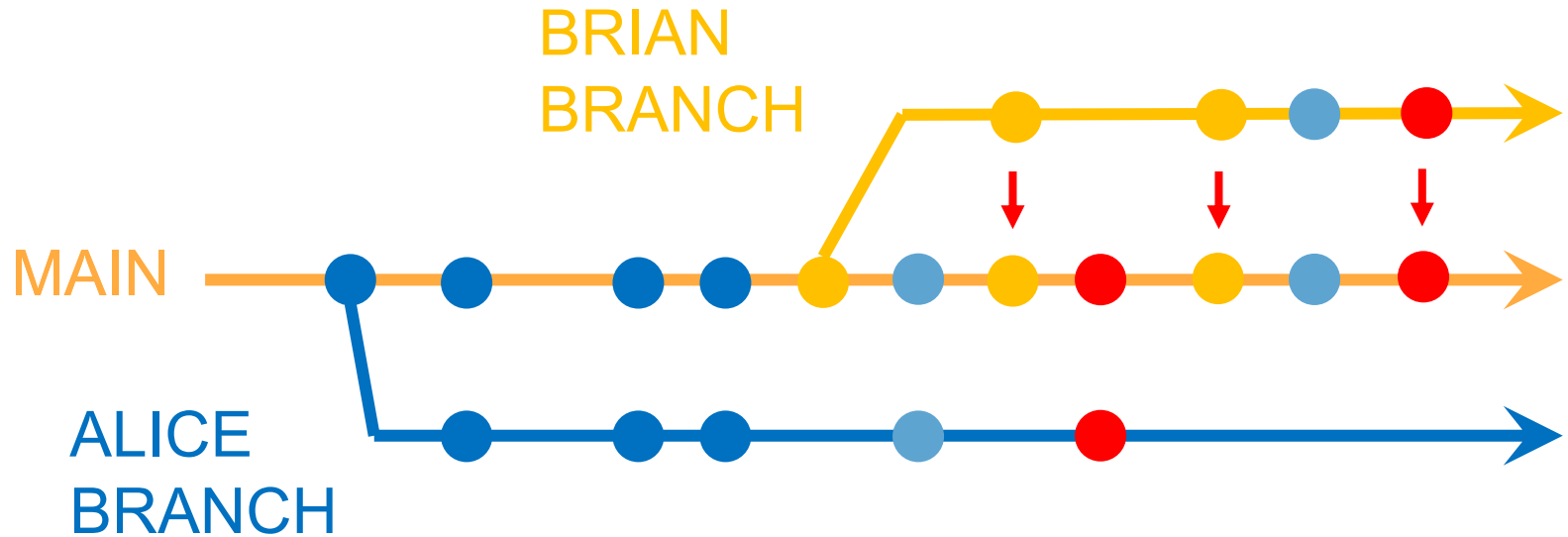
If someone changes the main branch, Alice can keep up with other changes to the main branch using “**merges**”.

Git branches allow for individual code changes



Alice's code changes are added to the main branch, with all of the commit history via a [merge](#).

Git branches support simultaneous code edits



Git branches support simultaneous code edits



Alice's and Brian's changes made it back to the main branch. They were working on different pieces at the same time, but were able to contribute to the project in a cohesive way.