"Git" started with version control

Cecile Murray, Data Scientist

U.S. Census Bureau, Business Development Staff (ERD)

February 3, 2022

United States® Census Bureau

This is a form of version control

Three goals for this talk

- Describe what version control software is and why it's useful

- Understand conceptual workflow and core vocabulary associated with Git and GitHub

- See these concepts in action in a live demo

No command line: we'll use GitHub Desktop

What is version control software?

- "Version control systems are software tools that help software teams manage changes to source code over time." [source]

- "Version control is a system that records changes to a file or set of files over time so that you can recall specific versions later." [source]

Why are version control systems useful for lone developers?

- Maintain a complete history of code development where changes are documented at each step

- Back up your files regularly as an automatic side effect of development

- Work on code on multiple computers or servers

Why are version control systems useful for collaboration?

- Maintain a complete history of code development where changes are documented at each step

- Back up your files regularly as an automatic side effect of development

- Work on code on multiple computers or servers

- Collaborate without worrying about overwriting someone else's work

- Integrate multiple people's work across multiple script files mostly automatically

- Review, document, and resolve any conflicting changes in a systematic way

A few key terms

- **Repository**, aka "repo" = your project folder + all version history
  - **Local** = full copy of repo on your computer/where you do your work
  - **Remote** = full copy hosted elsewhere (e.g. GitHub) where others can see and access your work
- **Commit** = a snapshot of the changes to all tracked files in your repo at a specified point
- **Push** = bring remote repo files up to date with local repo files
- **Pull** = bring local repo files up to date with remote repo files
- **Branch** = an independent line of development within a repo

Git, GitHub, GitLab??

- Git = the version control technology itself
- GitHub = a public, free website to host remote repos + access associated project management tools
- GitLab = very similar to GitHub, but hosted internally inside your organization's firewall instead of living on the internet

Version control with Git

Source: https://git-scm.com/book/en/v2/Getting-Started-What-is-Git%3F
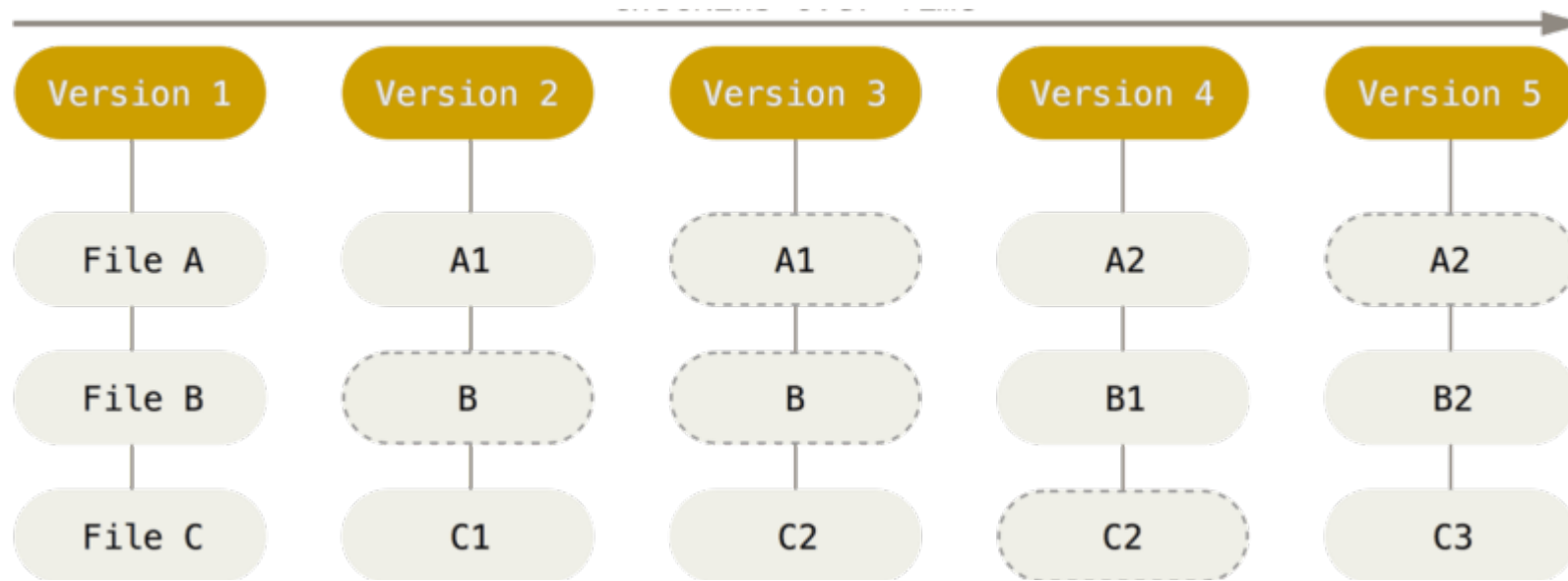
A simple version history could look like this

Main

Each circle
is a **commit**
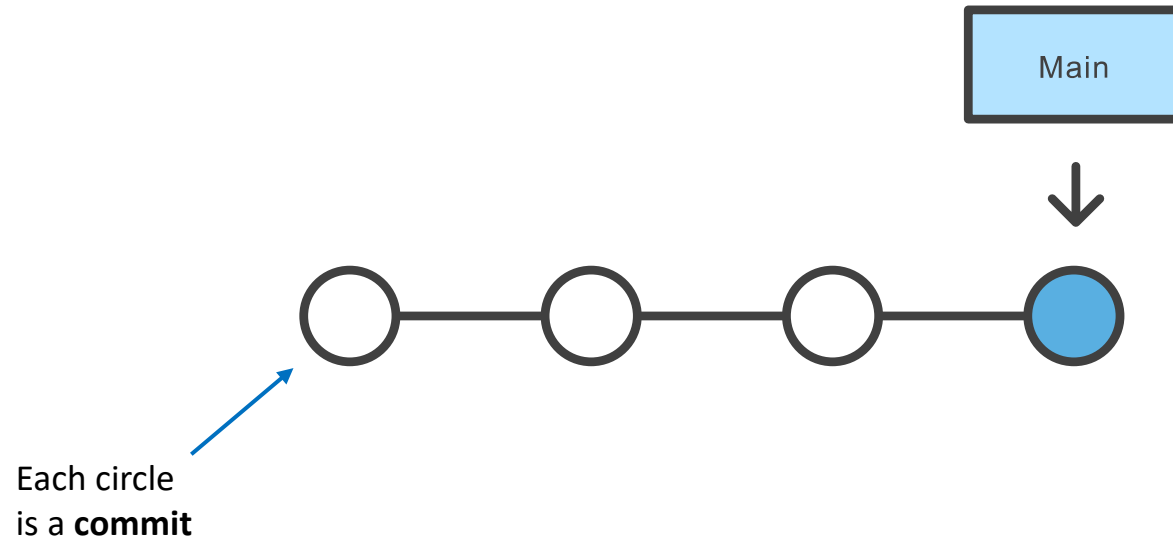
Git can track multiple version histories at once

Here, we **branched.**

Some Feature

Commits on the new branch

Commits on the existing branch

Main
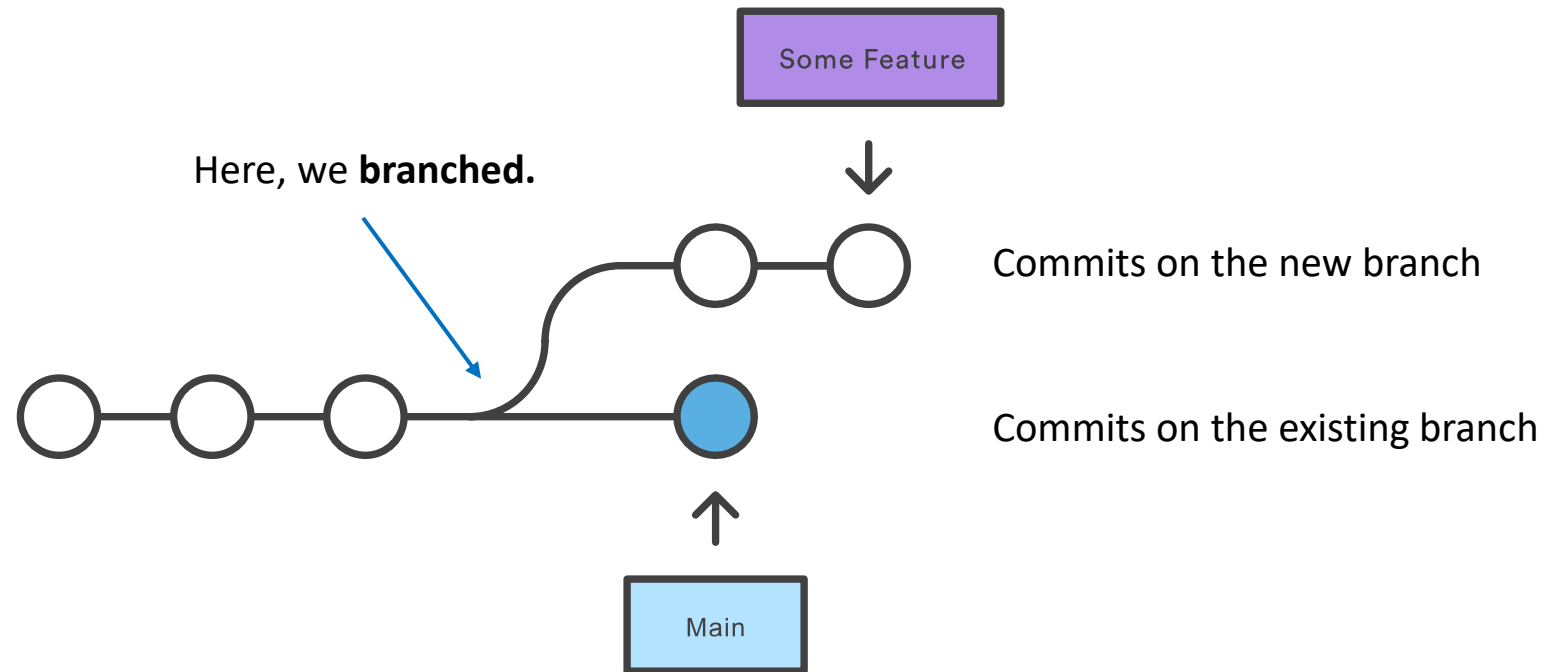
Git helps you merge changes between versions

Some Feature

Here, we **branched.**

Here, we **merged.**

Main

https://www.atlassian.com/git/tutorials/using-branches/git-merge
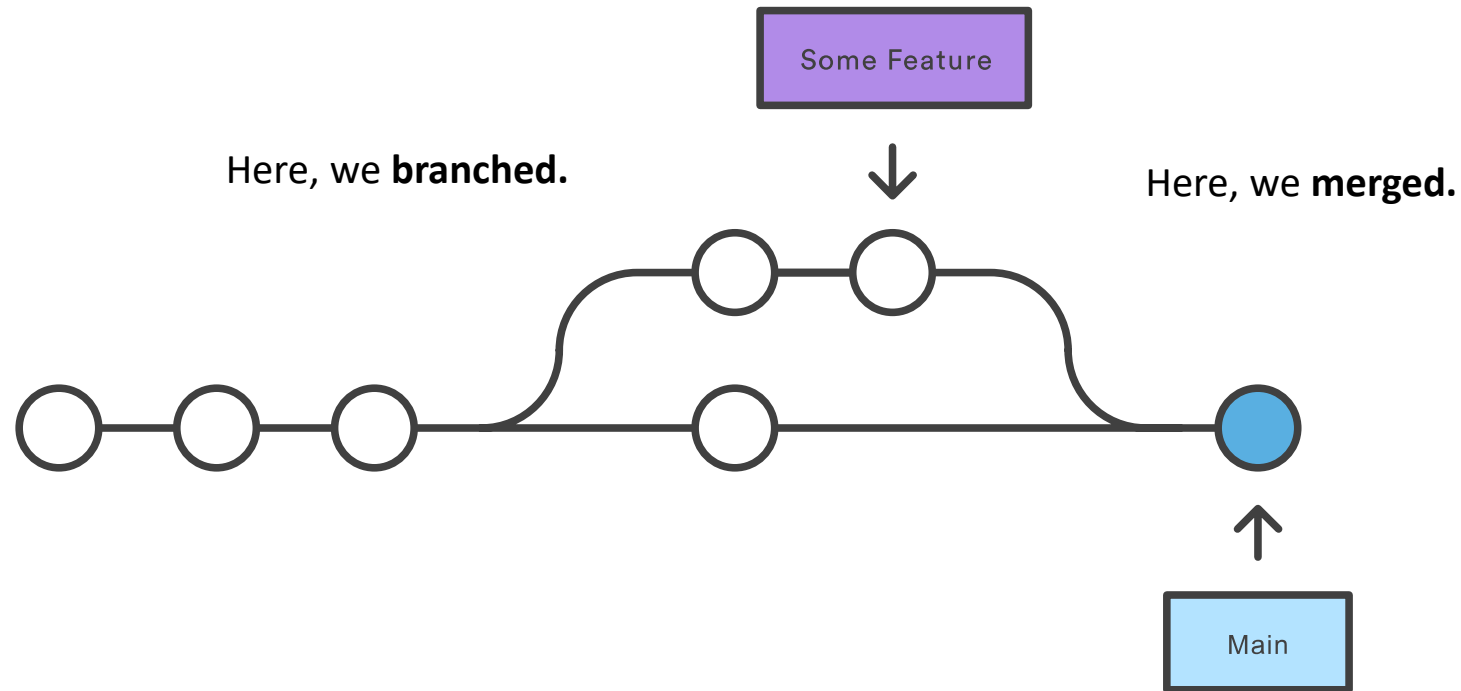
Time to practice!

Best practices

- Pull before you start your work

- Commit moderately frequently at sensible break points

- Don't forget to push your commits

- Do not commit:
  - Data, especially sensitive data
  - Large files
  - Binary files, e.g. MS Office files

Which history do you prefer?

Name	Date modified
SLP_workingpaper	3/31/2021 18:06
SLP_workingpaper_v1	12/15/2020 12:13
SLP_workingpaper_v2	1/5/2021 17:42
SLP_workingpaper_v3	1/11/2021 9:30
SLP_workingpaper_v4	1/12/2021 11:01
SLP_workingpaper_v5_T13	1/29/2021 15:23
SLP_workingpaper_v6_T13	2/1/2021 14:00
SLP_workingpaper_v7_T13	2/9/2021 12:54
SLP_workingpaper_v8	2/16/2021 22:52
SLP_workingpaper_v8_JLreview	2/17/2021 10:46
SLP_workingpaper_v9	2/25/2021 11:17
SLP_workingpaper_v10	3/4/2021 11:30
SLP_workingpaper_v10_carla	3/10/2021 11:02
SLP_workingpaper_v11	3/11/2021 9:44
SLP_workingpaper_v12	3/11/2021 15:18
SLP_workingpaper_v13	3/31/2021 18:05

master ⌄   stat_lang_pro

19 Jan, 2021 2 commits

syrupy bug fix
Cecile Mcwilliams Murray (CENSUS/ERD CTR) authored 6 months ago

alter syrupy.py to use psutil
Cecile Mcwilliams Murray (CENSUS/ERD CTR) authored 6 months ago

09 Dec, 2020 4 commits

add new plots from blog draft
Cecile Mcwilliams Murray (CENSUS/ERD CTR) authored 8 months ago

fix merge conflict in make_plots
Cecile Mcwilliams Murray (CENSUS/ERD CTR) authored 8 months ago

begin adding csv results to plots
Cecile Mcwilliams Murray (CENSUS/ERD CTR) authored 8 months ago

latest on csv runs
Cecile Mcwilliams Murray authored 8 months ago

13 Nov, 2020 3 commits

Thank you! Questions?

Repo with demo materials: https://github.com/census-bds/intro-git

Feel free to reach out by email: cecile.m.murray@census.gov