Unit 3: Core tools

Github

Lesson 29

Derek Ruths

Overview of unit

Objectives:

- Understanding how data science activities necessitate certain kinds of tools
- Fundamentals with core data science tools

1. Why core tools?
2. Project organization
3. Python
4. Best practice: write CLI tools
5. Best practice: write unit tests
6. Best practice: resource referencing
7. Github
8. Jupyter notebooks
9. Jupyter & statefulness
10. Bokeh
11. Advanced bokeh
12. HW 3

Lesson overview

Objectives

- Use remote version control (e.g., git/github) for your projects
- Don't commit your data

(too big!

Outline

- Why version control?
- Quick git CLI overview
- Handling data
- A word on branching

Why version control?

Backups – you don't want to lose valuable code!

Multi-machine work – desktop, laptop, and cloud?

Collaboration & sharing

- Working in teams
- Sharing with the world

Tracking progress – committing code helps with messages

Quick git CLI overview

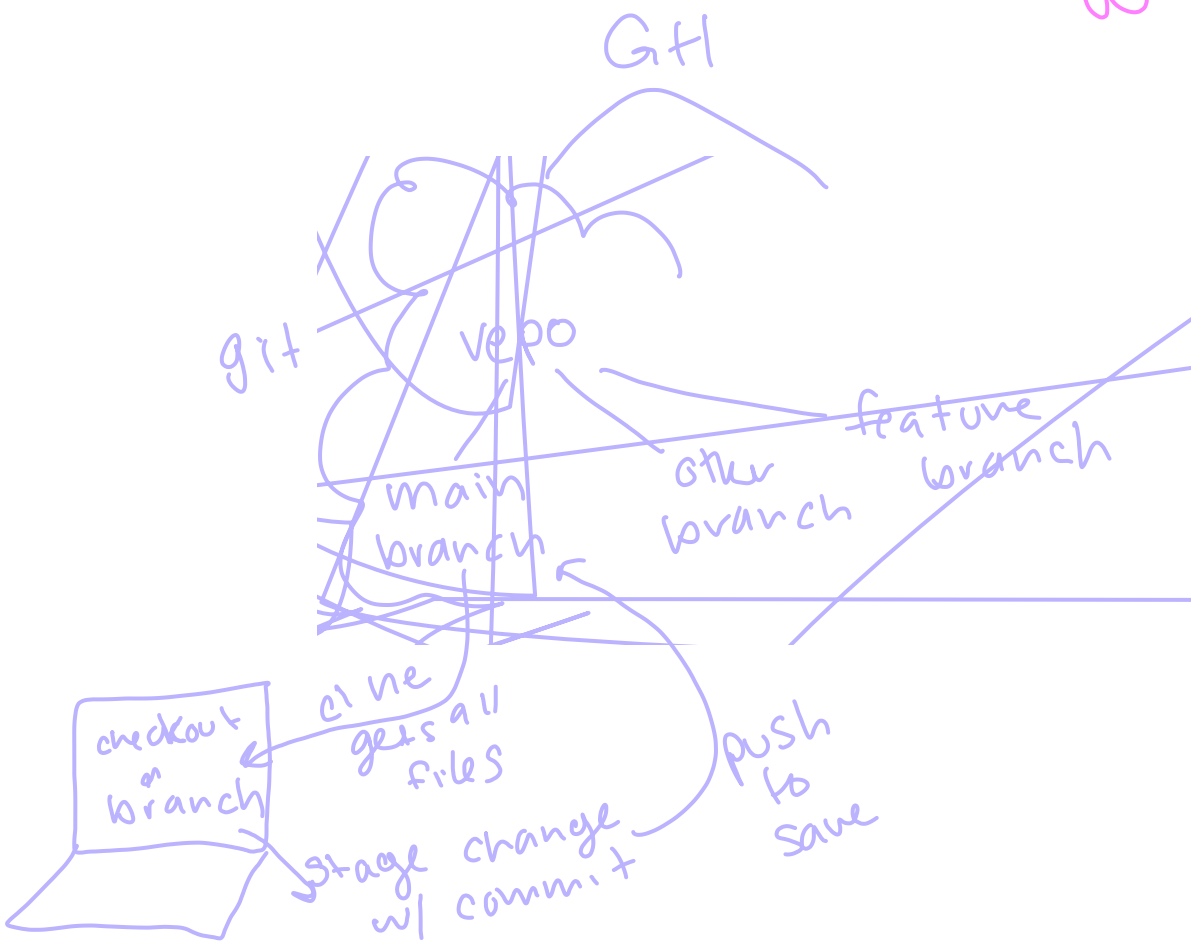
github is a commercial wrapper around git

git clone

git commit

git push

git pull



Best practices

One repository per project

Name projects snazzy, descriptive things

Super large files (like data) shouldn't be checked into a repository

- Use your README to explain how to get there, where to setup, etc…
- Use make or other tools to automate as much as possible

Don't worry too much about branching … it's more of a software development thing.

Lesson wrap-up

Takeaways

- Use version control!
- Git is a good choice…

Up next

- Jupyter & exploratory code