Setting up Your System with Anaconda and Git.

Tools for working with python on your local machine.

// FLATIRON SCHOOL

Objectives for today:

- Recap common Command Line Interface (CLI) commands.
- Review Anaconda Installation (MacOS & Windows)
  - Phase 2 Appendix Module: Setting up a Professional Data Science environment
- Discuss Anaconda, Jupyter notebooks and other code editors
- Talk about Git version control and Github

What is ***Anaconda?***

“The open-source Anaconda Distribution is the easiest way to perform Python/R data science and machine learning on Linux, Windows, and Mac OS X. With over 15 million users worldwide, it is the industry standard for...enabling individual data scientists to:

- *Quickly download 1,500+ Python/R data science packages*
- *Manage libraries, dependencies, and environments with Conda”*

— *Anaconda Distribution*
- Package List

- Conda is an open source package management system and environment management system that runs on Windows, macOS and Linux.
- Conda quickly installs, runs and updates packages and their dependencies.
- Conda easily creates, saves, loads and switches between environments on your local computer.
- You'll create conda environments to share, collaborate on, and reproduce projects with specific versions of particular packages.
- Source: Conda Documentation + Managing Environments Documentation + conda cheat sheet

What is ***Jupyter?***

Jupyter

- Project Jupyter exists to develop open-source software, open-standards, and services for interactive computing across dozens of programming languages.

- Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text.
  - Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

- JupyterLab is a next-generation web-based user interface

- **Is included in the Anaconda software distribution**

Installing the Git & Anaconda Software

Go to Canvas and review installation assignment in appendix

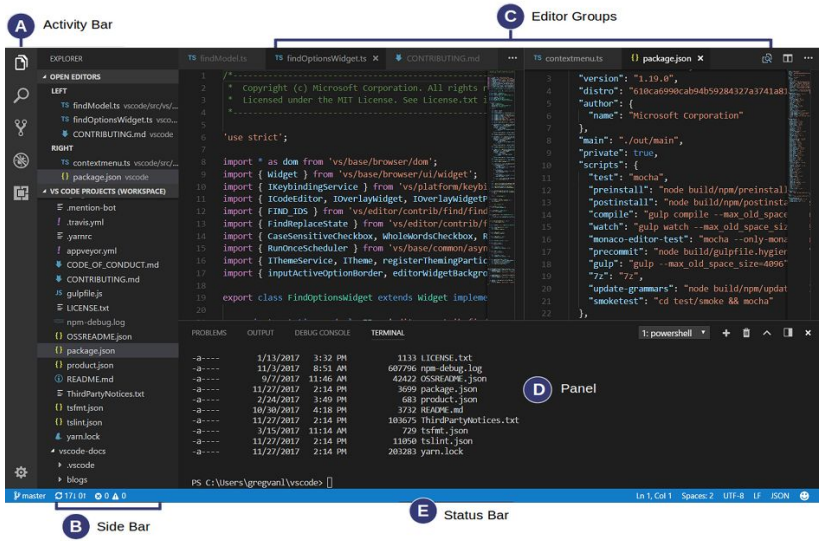
MacOS

Windows

What is ***Visual Studio Code***?

Visual Studio (VS) Code

- Visual Studio Code is an open-source text editor created by Microsoft
- Navigate directory structure, make/remove files, and direct access to the Terminal/Command Line
- Allows you to write text files (`.py`, `README.md`, etc.) and recently, VS Code allows you to edit Jupyter Notebooks directly
- Easy to switch between conda environments and lint code



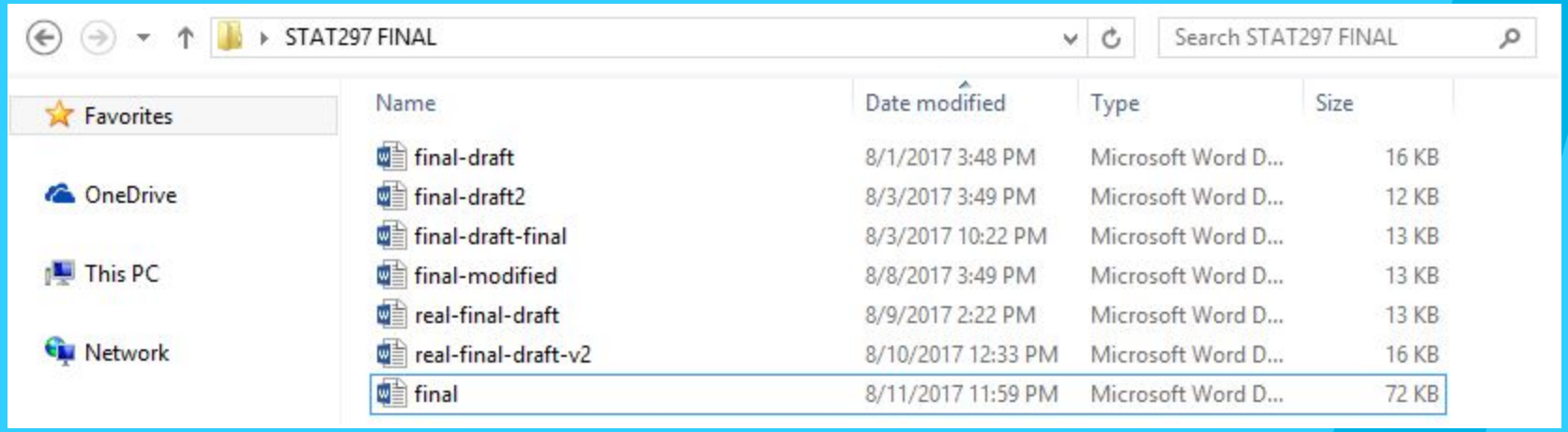
Choose the tools that work for you:

What is ***Git***?

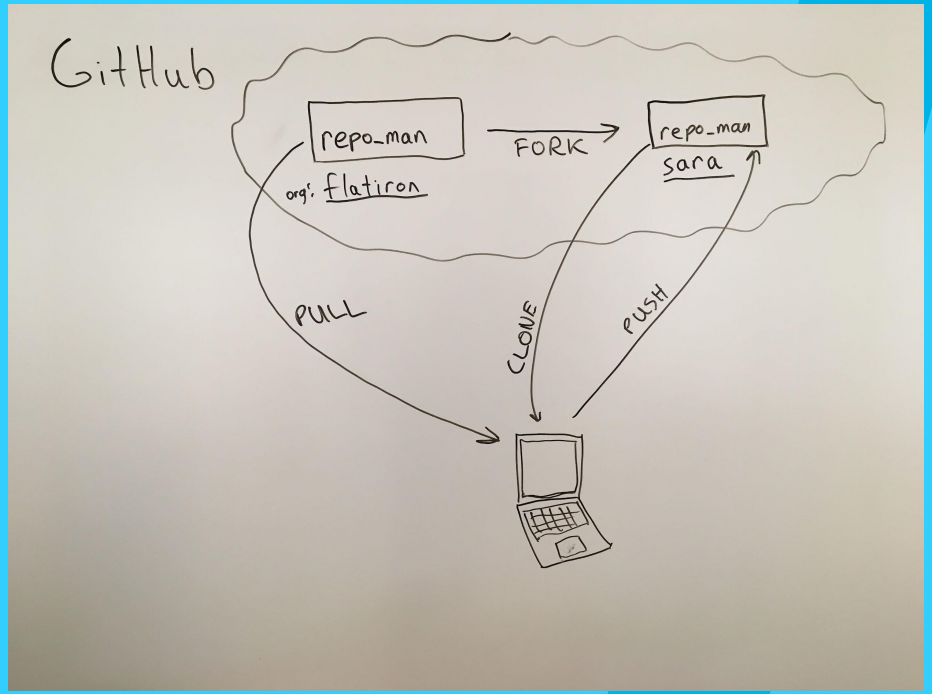
Git

- Git is a version control system.
- It’s a way of keeping track of all the changes made across your project.
- Think of it like “track changes” in Word - but with the ability to track changes across multiple documents.

How many of you have seen a folder like this?



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



What is ***GitHub***?

GitHub

- GitHub is a free software platform that hosts over 40 million developers code
- You'll primarily use GitHub to collaborate with others, document your projects, and build your portfolio to showcase your abilities as a data scientist
- You can also use GitHub for any of the following tasks:
  - Code hosting
  - Code review
  - Project management
  - Team management
  - Documentation

Advantages of GitHub

Employers look for comfort using git
A “green” robust github commit history
Content accessible after the program
It is what you will be using in the real world
Built for collaboration

Implementation in Industry?

As an analyst:

- Creating / maintaining visualization dashboards (Python, Tableau, PowerBI)
- Hosting your portfolio online / reporting for your organization online.

As a Data Scientist/Engineer:

- Similar to analysts with dashboards and portfolio.
- Storage and development of Machine learning models or pipelines

Personally:

- Store any projects and their version history online. Remember Python is a versatile language! (Web development, app development, Data etc.)