Tools We Will be Accessing

Ensuring you have the right tool for each

// FLATIRON SCHOOL

Objectives

- Navigate to and within IllumiDesk
- Use Anaconda, Jupyter and VS Code for data science
- Understand the relationship between Anaconda, Jupyter, Git and Python

Tools of the Trade

- Python
- Anaconda
- Jupyter
- Visual Studio Code (VSC)
- Git
- Github

// FLATIRON SCHOOL

What is *Python?*

- A **coding language** used extensively by data science

// FLATIRON SCHOOL

Python

“Python, named after the British comedy group Monty Python, is an interpreted, interactive, object-oriented programming language.
Its flexibility allows it to do many things, both big and small.

Python can be used to write simple programs,

but it also possesses the full power required to create complex, large-scale enterprise solutions.” - Derrick Kearney

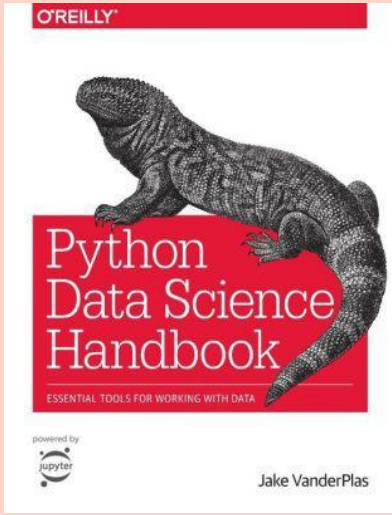
Python is an Object Oriented Programming language, however, unlike Java...

// FLATIRON SCHOOL

Python for Data Science

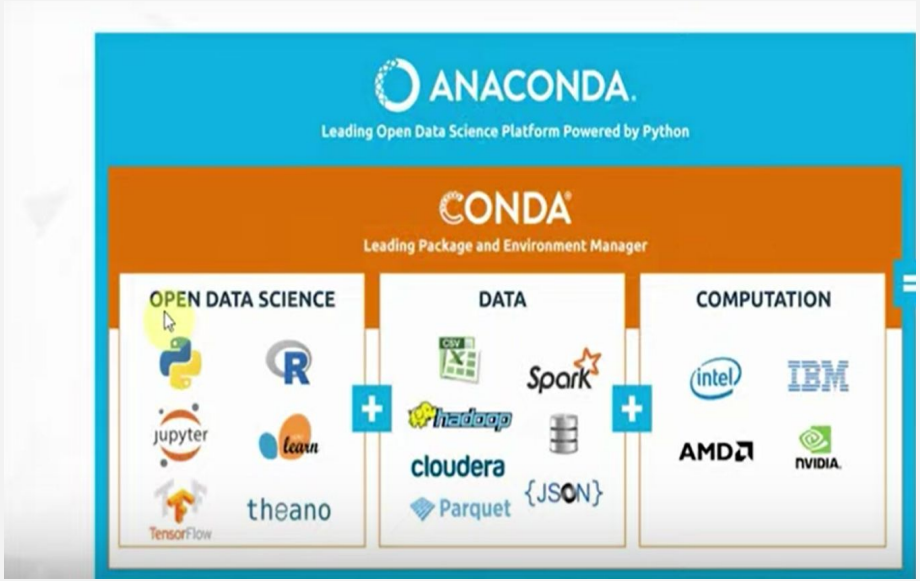
“The usefulness of Python for data science stems primarily from the large and active ecosystem of third-party packages:

- *NumPy* for manipulation of homogeneous array-based data;
- *Pandas* for manipulation of heterogeneous and labeled data;
- *SciPy* for common scientific computing tasks;
- *Matplotlib* for publication-quality visualizations;
- *[Jupyter]* for interactive execution and sharing of code;
- *Scikit-Learn* for machine learning, and many more tools...”
  - Jake VanderPlas



// FLATIRON SCHOOL

What is Anaconda



// FLATIRON SCHOOL

“The open-source Anaconda Distribution

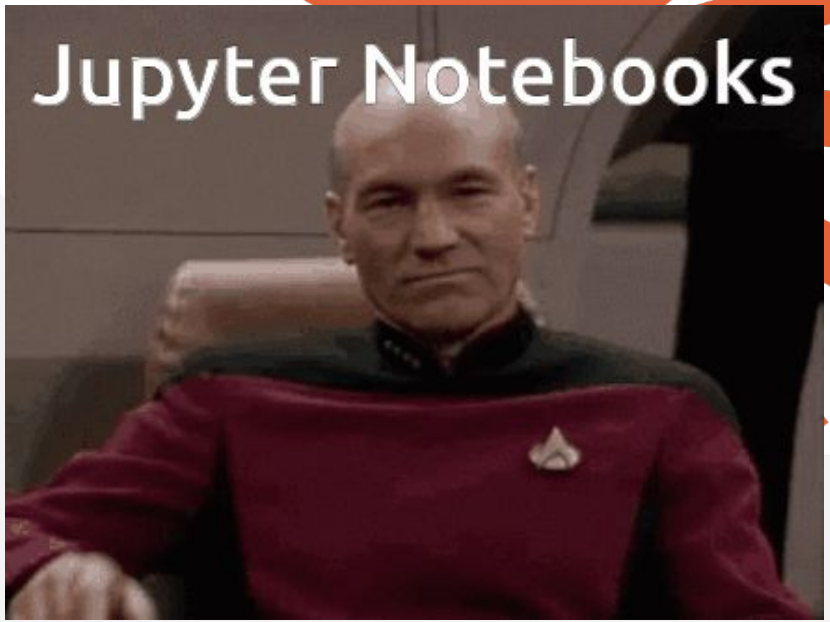
- ***Package Management System***
- Easiest way to perform **Python/R data science** on Linux, Windows, and Mac OS X.
- Over **15 million users** worldwide
- **Industry standard** for...enabling data scientists to:
  - *Quickly download 1,500+ Python/R data science packages*
  - *Manage libraries, dependencies, and environments with Conda*”

— *Anaconda Distribution*-Packge List

- 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

- 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

What is Visual Studio Code

// FLATIRON SCHOOL

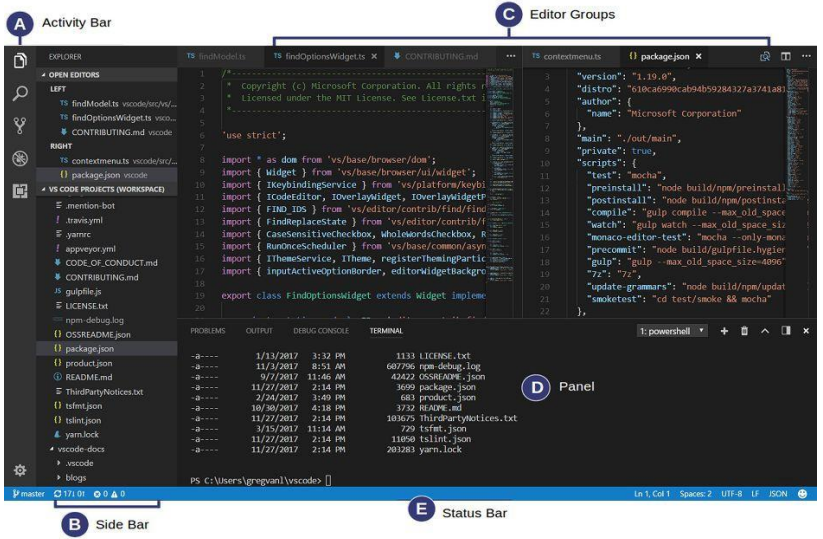
Visual Studio (VS) Code

- 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

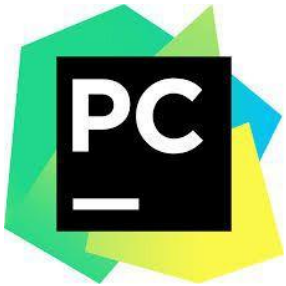
- Allows you to write text files (`.py`, `README.md`, etc.) and recently, [VS Code allows you to edit Jupyter Notebooks directly]

- Easy to switch between conda environments



// FLATIRON SCHOOL

Choose the tools that work for You



// FLATIRON SCHOOL

What is Git?

// FLATIRON SCHOOL

What is 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.

// FLATIRON SCHOOL

What is Github?

// FLATIRON SCHOOL

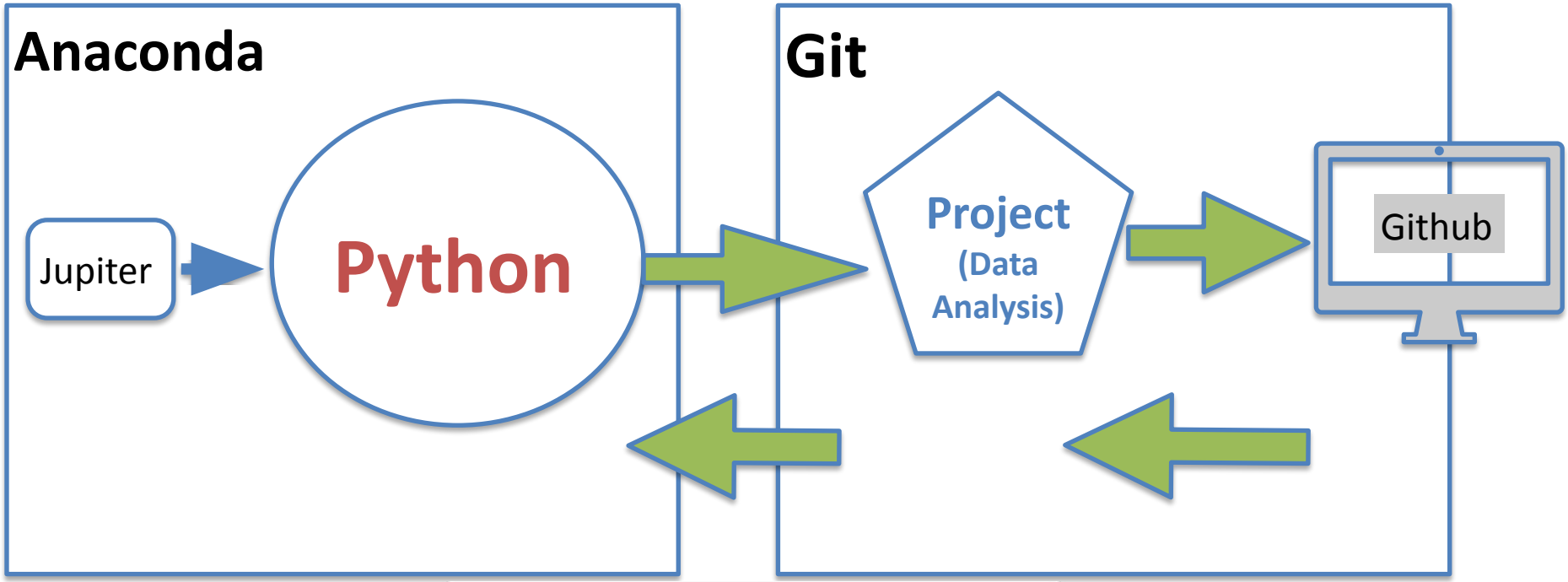
GitHub

- 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

// FLATIRON SCHOOL



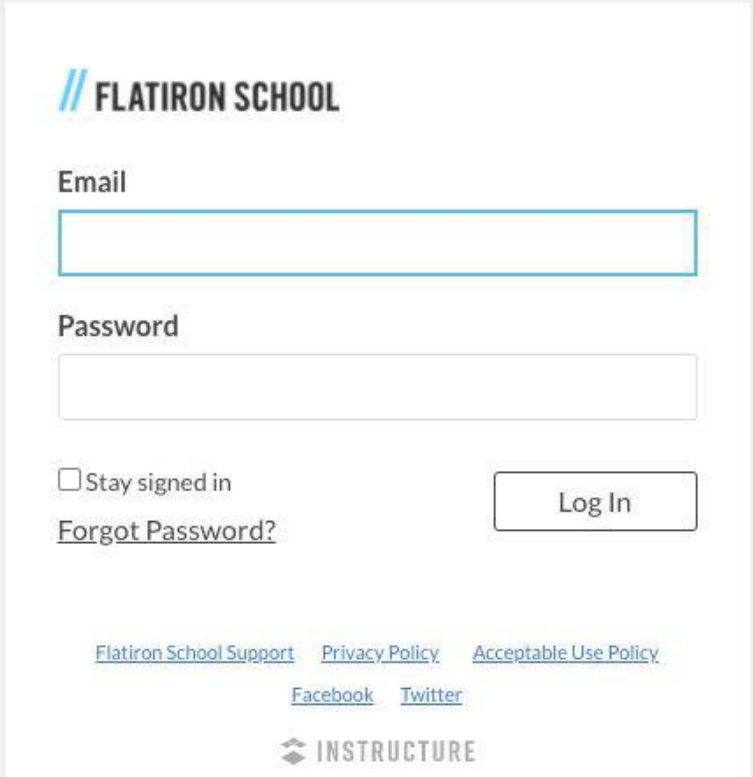
Work Flow

// FLATIRON SCHOOL

CANVAS

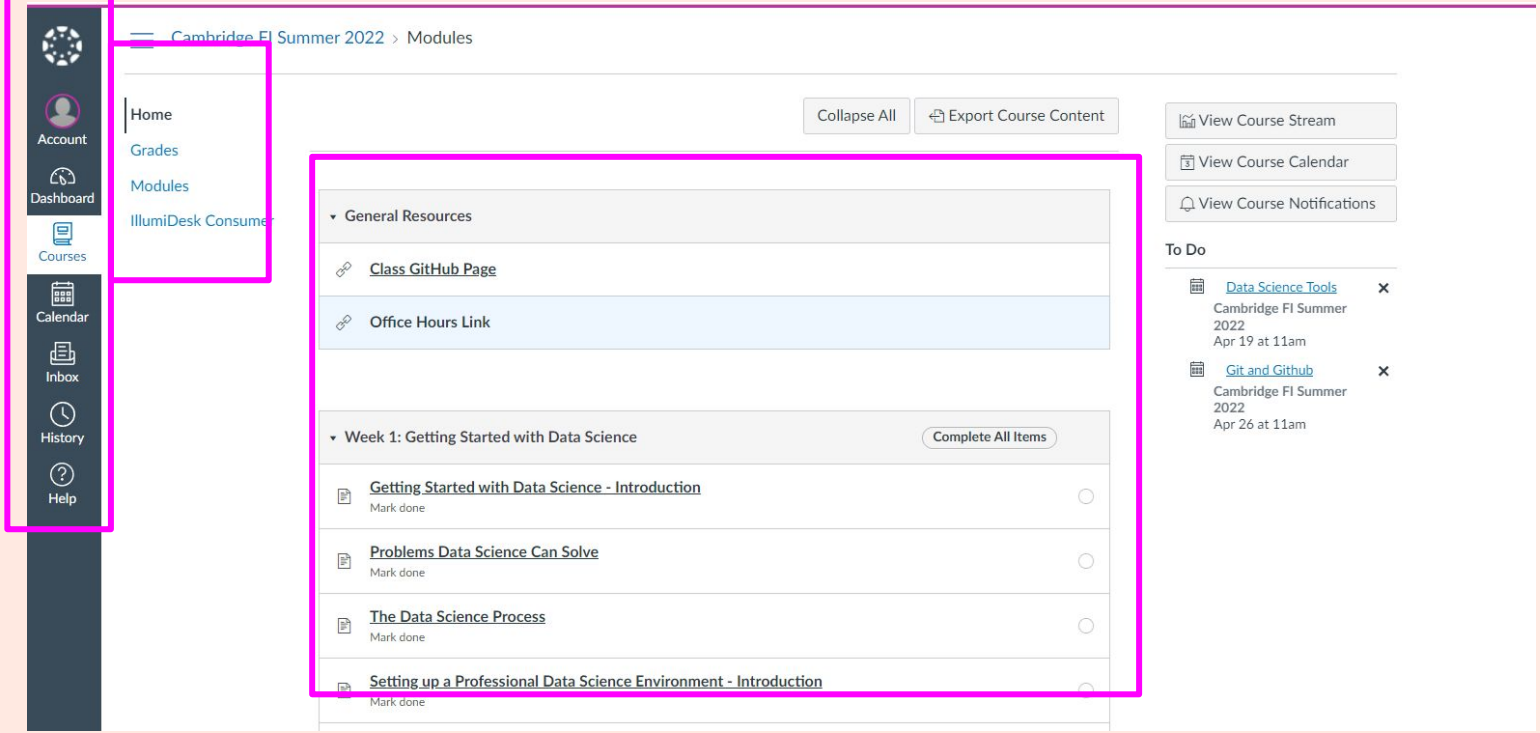
// FLATIRON SCHOOL

CANVAS



// FLATIRON SCHOOL

CANVAS



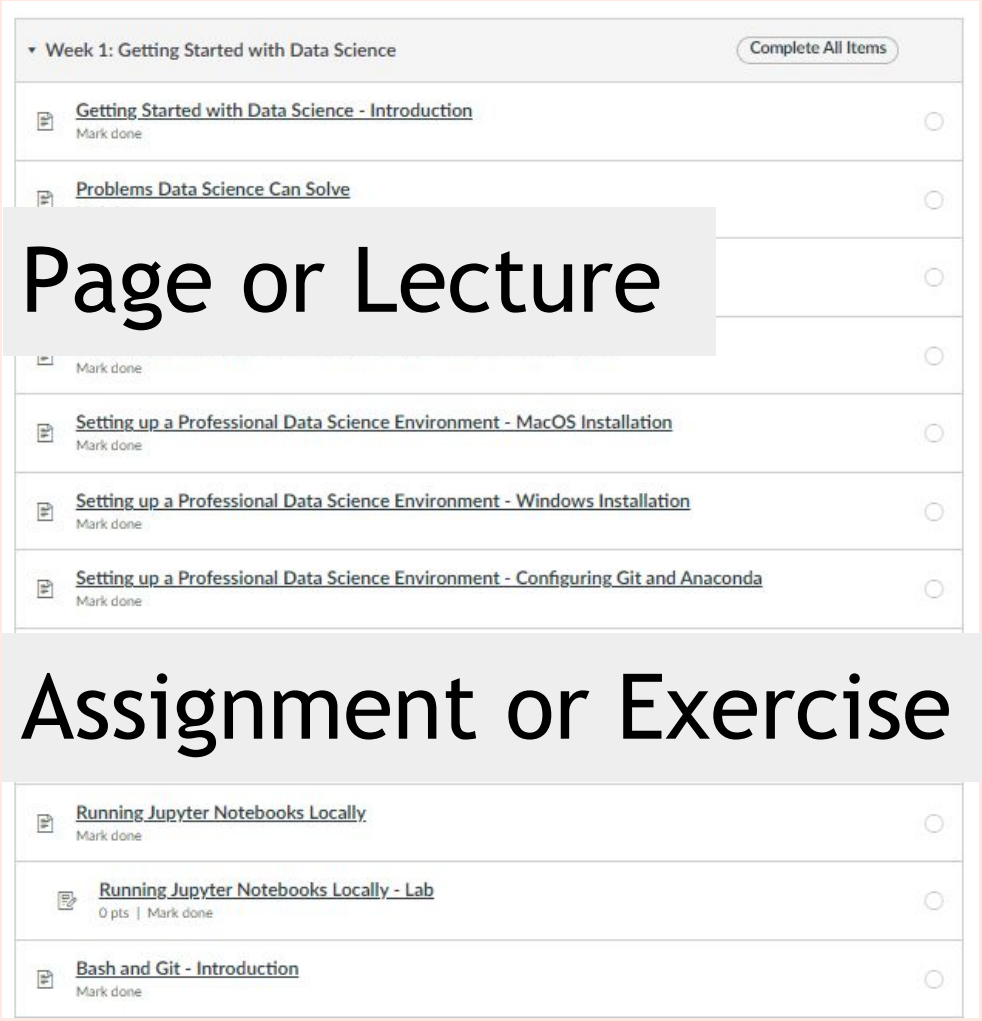
CANVAS

// FLATIRON SCHOOL

CANVAS

Page or Lecture

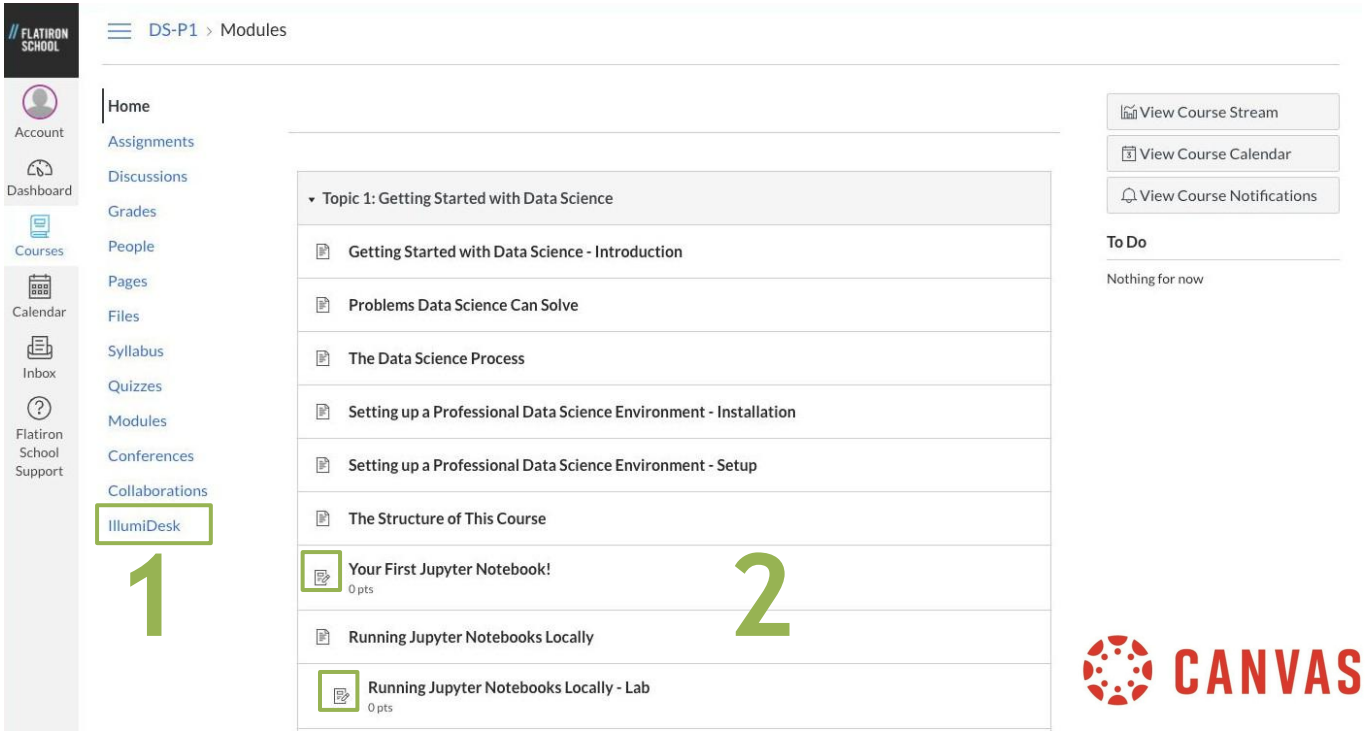
Assignment or Exercise



CANVAS

// FLATIRON SCHOOL

Two ways to access



IllumiDesk link

Either blue button will direct you to your Jupyter environment

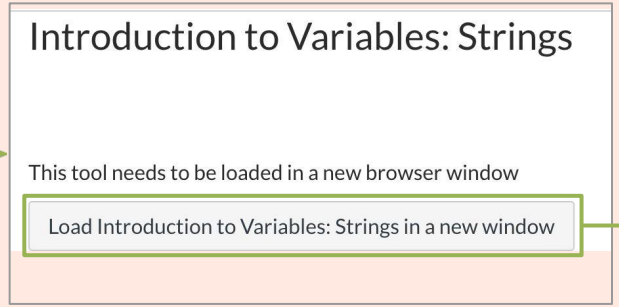
First time you access IllumiDesk in a Canvas session, the you will need to start your server at the Control Panel

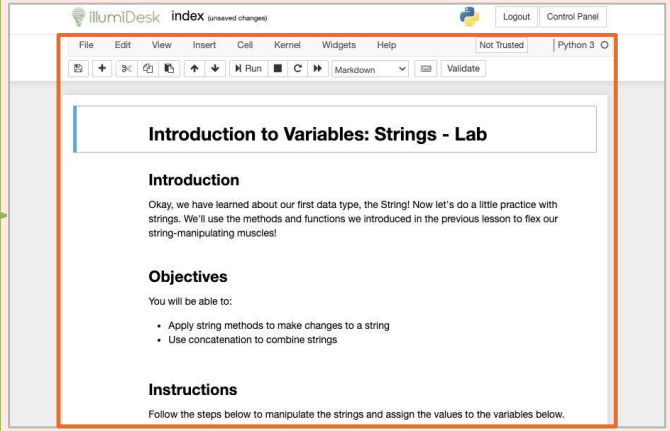
After starting the server the Control Panel has multiple options

The Control Panel button navigates back to the Control Panel, also known as “Home”

// FLATIRON SCHOOL

IllumiDesk Through Assignment links





Jupyter Notebook

CANVAS

illumiDesk

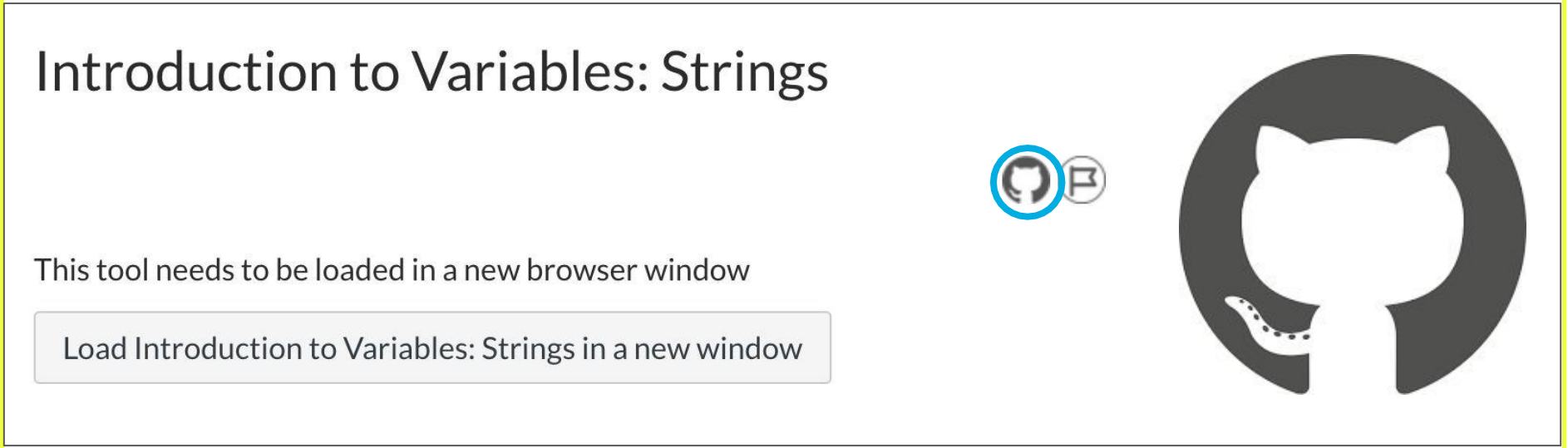
// FLATIRON SCHOOL

// FLATIRON SCHOOL

Advantages of DataCamp for Workspaces

Ease of use
No environment issues
Fully integrated into Canvas
Basically Jupyter Notebook

Every lesson *with* code is *stored* on *GitHub*

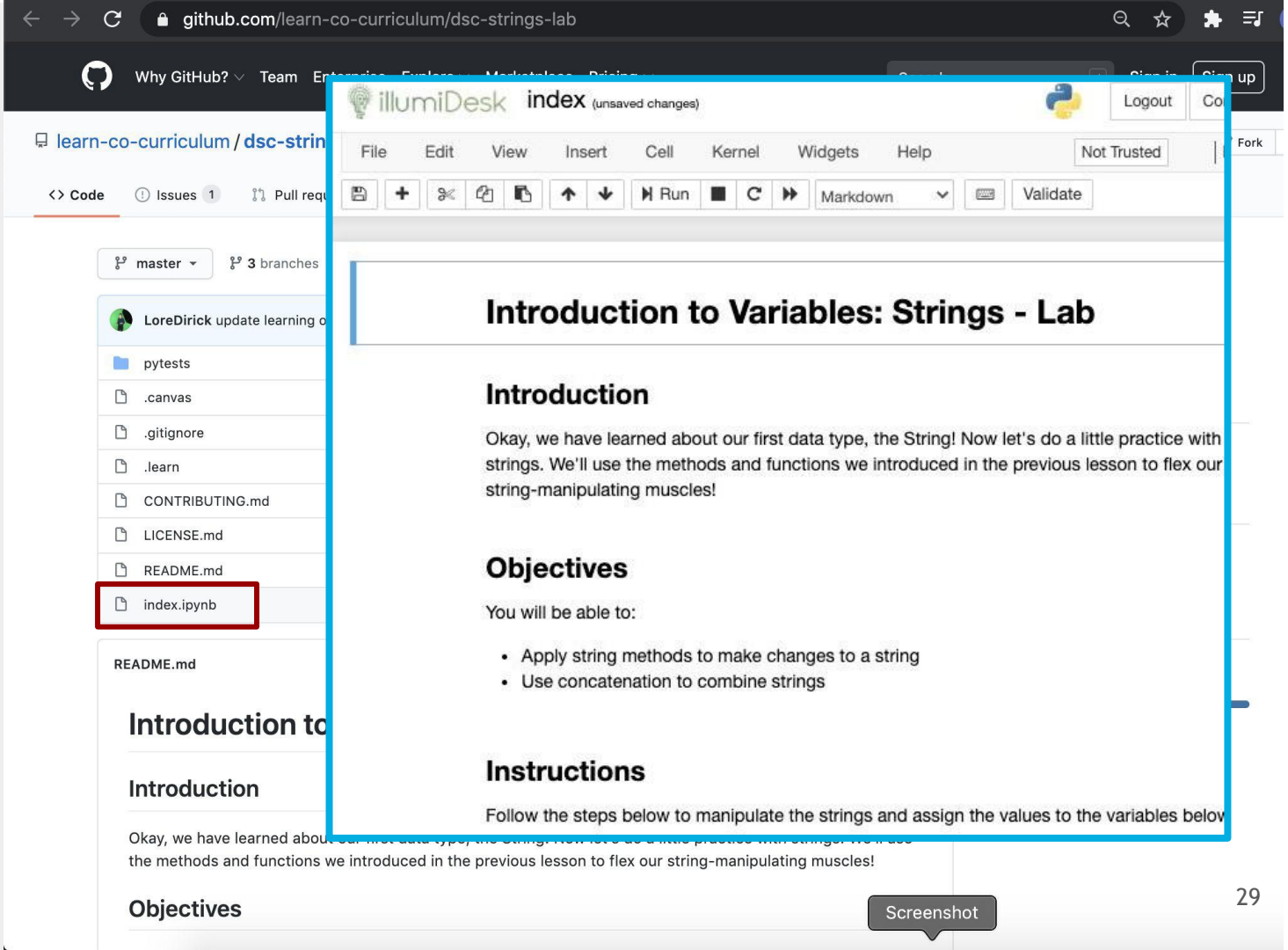


The GitHub logo in Canvas will navigate to the lesson's GitHub repository

// FLATIRON SCHOOL

You fork and clone the source material from GitHub to your own machine.

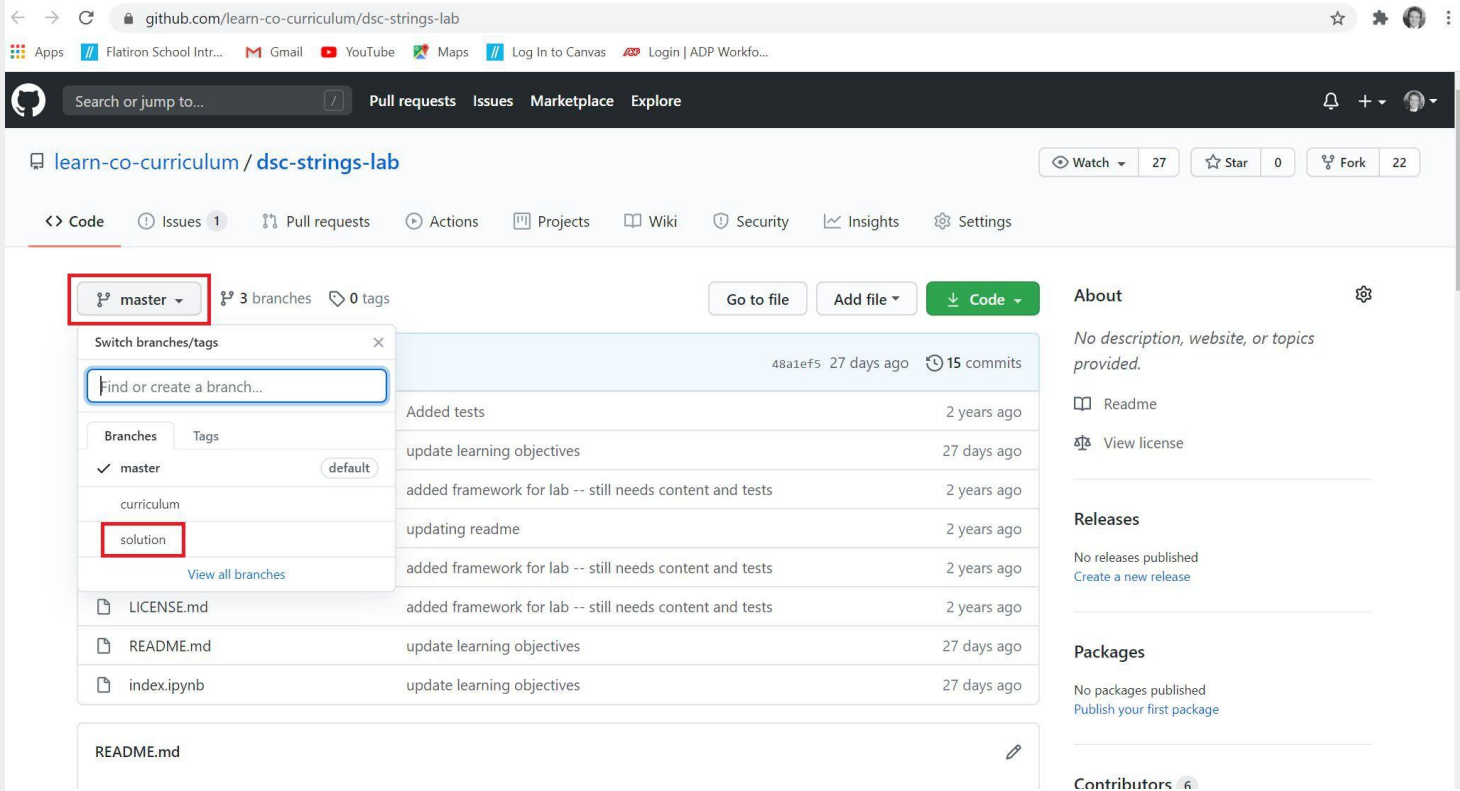
(instructions to come)



// FLATIRON SCHOOL

Lab solutions are on the “solution” branch of each repository.

(we will teach you what that means soon)



Course Structure

- Lectures
- Lessons
- Exercises/Assignments
- Labs
- Quizzes

// FLATIRON SCHOOL

Cambridge FI Summer 2022

Legend: Modules Lessons Assignments Labs - Graded Quizzes Other

- **Week 1: Getting Started with Data Science**
  - Problems Data Science Can Solve
  - The Data Science Process
  - Setting up a Professional Data Science Environment - Introduction
  - Setting up a Professional Data Science Environment - MacOS Installation
  - Setting up a Professional Data Science Environment - Windows Installation
  - Setting up a Professional Data Science Environment - Configuring Git and Anaconda
  - The Structure of This Course
  - Your First Jupyter Notebook!
  - Running Jupyter Notebooks Locally
  - Running Jupyter Notebooks Locally - Lab
  - Bash and Git - Introduction
  - The Bash Shell
  - Getting Started with Data Science - Recap

// FLATIRON SCHOOL

// FLATIRON SCHOOL