How to Build Your Own Project

The Essentials

1) Git (and Github)
2) IDE
3) Programming Language, Required Libraries and Virt Env.
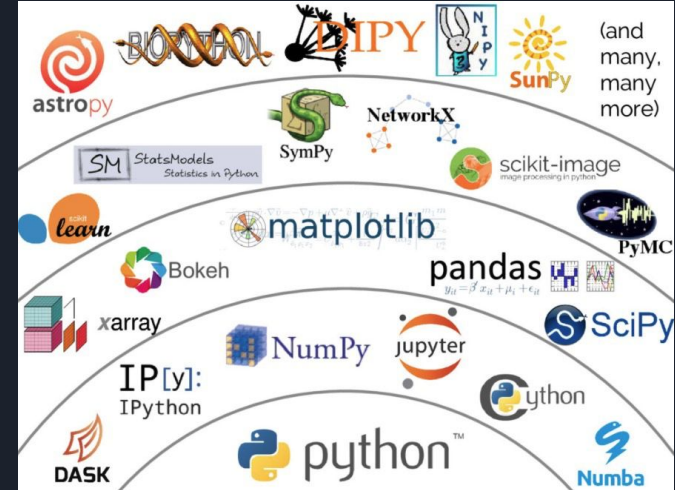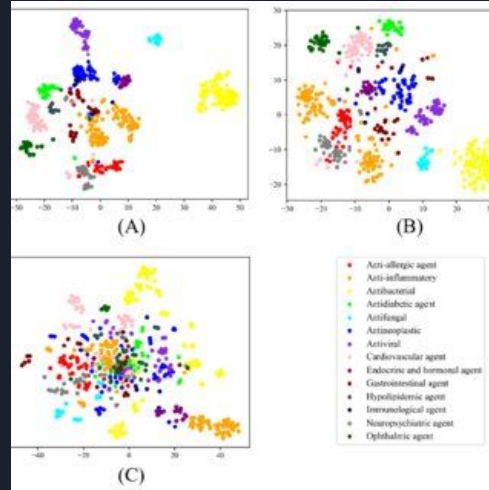
Things We're Going To Download

Mandatory

- Python (Or your language of choice)
- Pip
- Git and Github Desktop
- An IDE (VS Code)

Project Dependent

- Anaconda (for Python projects)
- Cuda (If you have a GPU)
- Git Bash

Git and Github

- Git allows for version control
  - If you mess up on a version you can go back to the previous one
- Github is a repository for Git
  - Can store multiple projects
  - Also allows you to collaborate with others
- Some basic git commands:
  - *git commit -m "message"* - Prepare to commit changes to the repository
  - *git push* - pushes changes to repository
  - *git clone /path/to/repository* - allows you to clone another repository onto your device
  - More commands: https://confluence.atlassian.com/bitbucketserver/basic-git-commands-776639767.html

What's an IDE?

- Stands for "Integrated Development Environment"
- Is a place where you can create, run, and test your code
  - Often has convenient shortcuts/tools that you can use for your projects
- The one you use depends on your preferences and what you plan to work on.
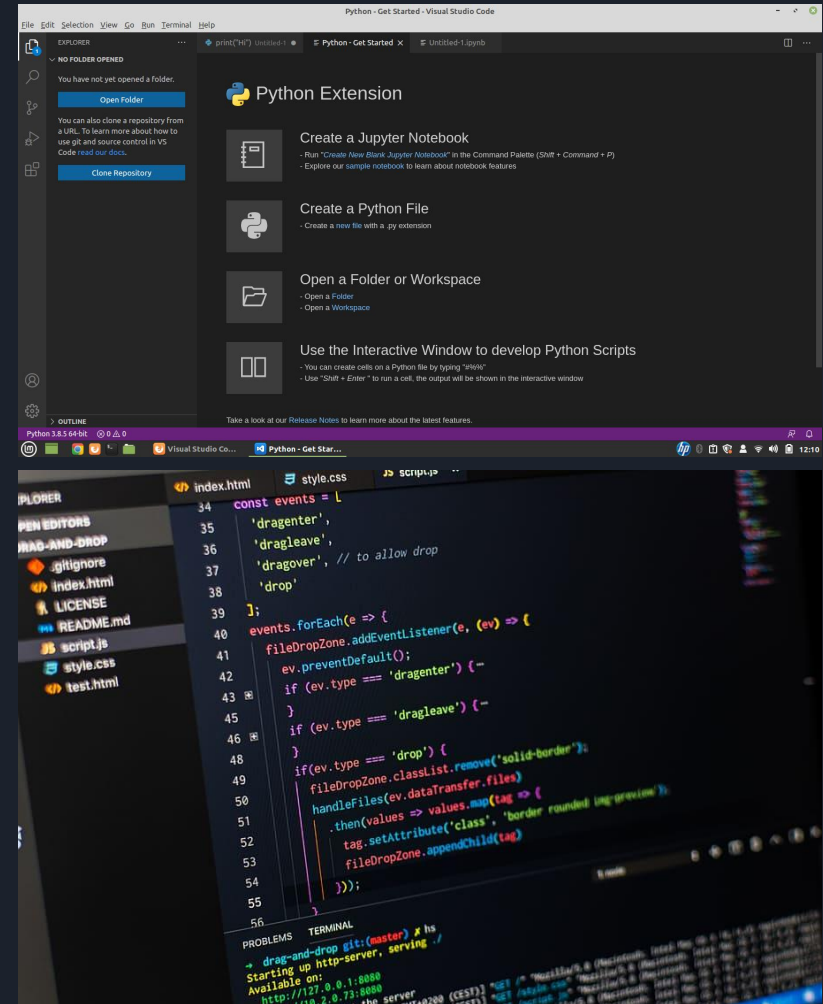
Some Popular IDEs:

Visual Studio:

- Made by Microsoft
- More robust version of intellisense
- Made for collaboration
- Very popular

Xcode/Android Studio:

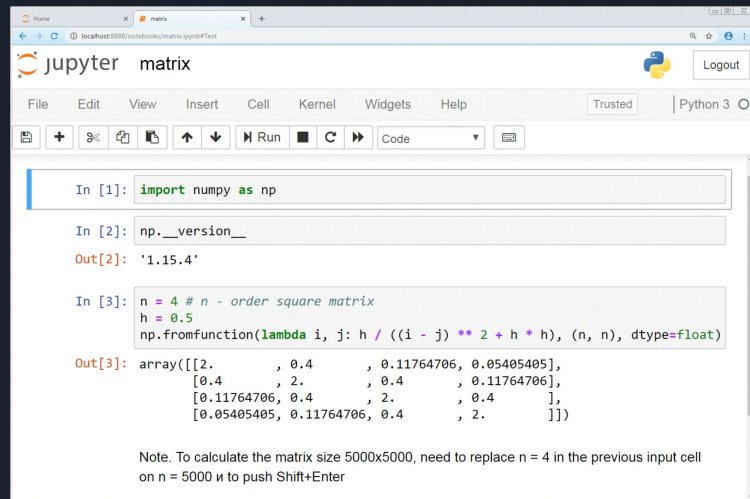- Meant for IOS and Android development respectively

Visual Studio Code:

- Also Made by Microsoft
- Technically a text editor, but it extremely customizable due to a large amount of extensions available

Programming Languages + Required Libraries

- IDEs can allow different programming languages to be used
  - Different programming languages fit different needs
- Data science is often done in python
- You can often integrate needed functionalities that aren't present in original language by importing other libraries
  - Ex: numpy, math
- You can also include other modules for data analysis, machine learning, etc.
  - Ex: Keras, SciKit, Jupyter Notebook

Pip and Installation

To install packages you will not a package manager, the most popular being pip:

1. You will need Python installed, which can be found here: [Python Install](#)
2. Once downloaded, go to your terminal, and input the following:

```
python -m ensurepip --upgrade
```

    a. On windows replace 'python' with 'py'

3. Now to install any package, just use the following:

```
pip install packagename
```

Creating a Virtual Environment

Why?

- CLEAN!
    - Avoid messing up with libraries for other projects
- Makes it easier for your products to be reproduced

pip install pipenv

pipenv install

pipenv shell

Tutorial

- Download Visual Studio Code and Github
- Clone the repository here:
  https://github.com/couchsnail/ds3_w24_byo
- You can fill out the Notebook yourself!
  - A sample of what real data science projects are like

Further Resources

A more in-depth tutorial to Git and Github:
https://product.hubspot.com/blog/git-and-github-tutorial-for-beginners

How to use JupyterNotebooks on vscode:
https://code.visualstudio.com/docs/datascience/data-science-tutorial