

# *Setting up a Professional Data Science Environment*

Ensuring you have the right tool for each project



# Agenda

1. **What is Python?**
2. **What is Anaconda?**
3. **What is Jupyter?**
4. **What is Visual Studio Code?**
5. **What is Git?**
6. **What is GitHub?**

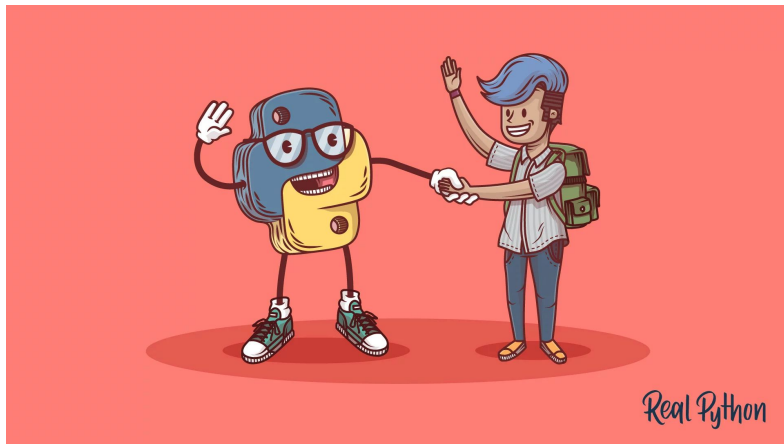
# 1. What is *Python*?

# 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 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](#)



Jake VanderPlas

## 2. 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](#)



# Conda

- 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](#)





### 3. 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
- Share notebooks using [nbviewer](#)



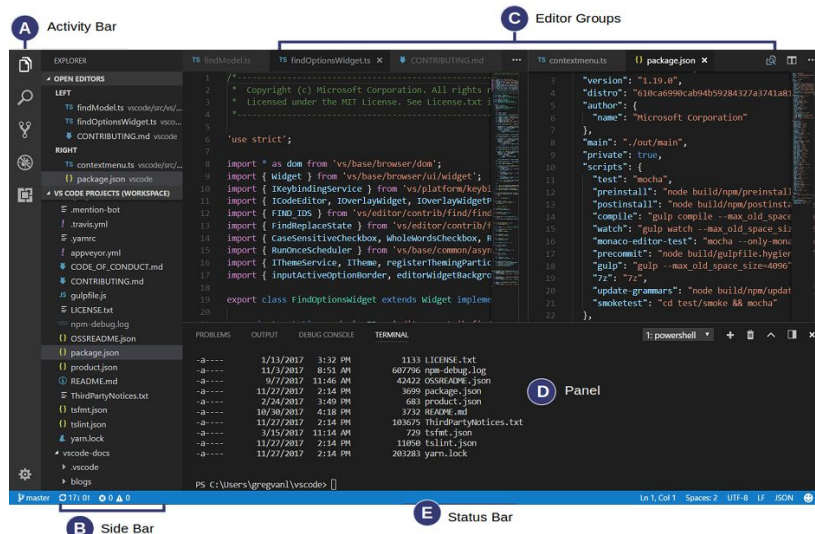
## 4. What is ***Visual Studio Code***?

---



# 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



## 5. 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.



## 6. 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





# Putting it All Together

- See the following post to download nearly all of the tools mentioned in this presentation:

<https://learn.co/tracks/module-1-data-science-career-2-1/intro-to-data-with-python-and-sql/section-01-getting-started-with-data-science/setting-up-a-professional-data-science-environment-setup>

- For VS Code, run the following in your Terminal:

```
brew cask install visual-studio-code
```