



Introduction to Python

Designed for New Coders

Housekeeping for Virtual Delivery

- In case of technical problems:
 - Something wrong on my end (e.g. power outage), I will send you an email.
 - Something wrong on your end, please send me a text message. 508-769-6446

Before we begin...

- There is very little straight lecture.
- Most of the time we will be writing code
- We will take frequent breaks. Most breaks will be attached to a workshop.
- Ask questions at any time. If you ask via audio, please let me know who you are.
- Feel free to ask via chat.

A little about me

■ Experience:

- 25+ years consulting and training experience
- Extensive work with “big data” and analytics

■ Education

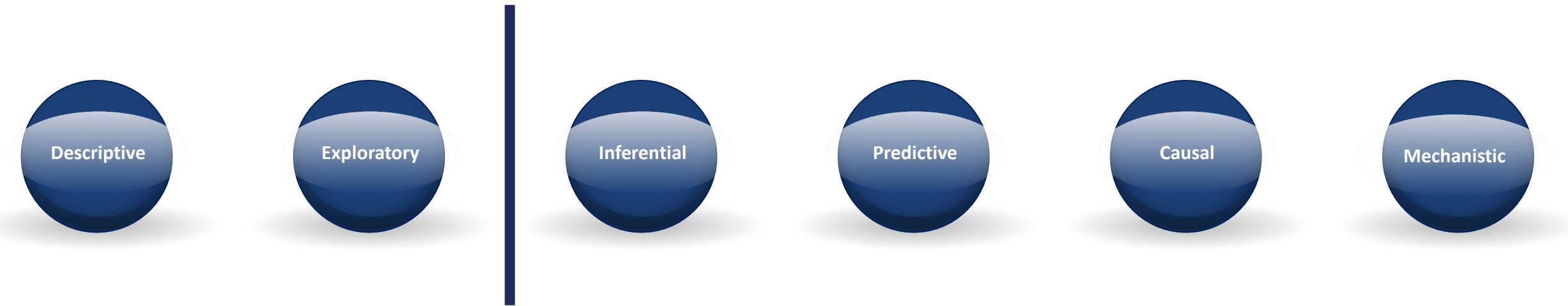
- Ed. M., Technology, Innovation & Education, Harvard University
- PhD Candidate, Education Policy University of Massachusetts, Amherst

Our goals for the course are:

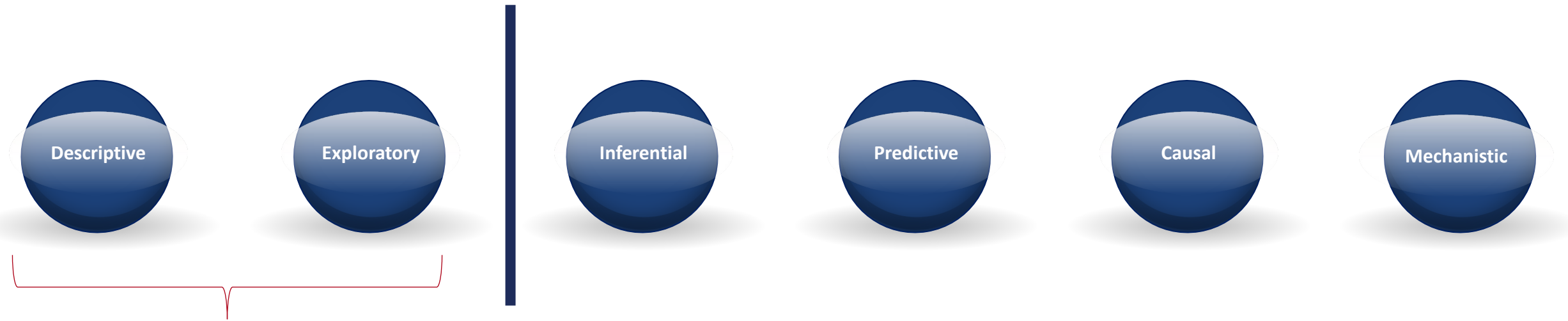
- Be able to use:
 - the core python library
 - Numpy
 - Pandas
 - Matplotlib
 - Seaborn
- Be able to conduct a basic exploratory data analysis with visualizations using the libraries listed above.

Before we get into Python, let's talk about data analysis.

Types of analyses



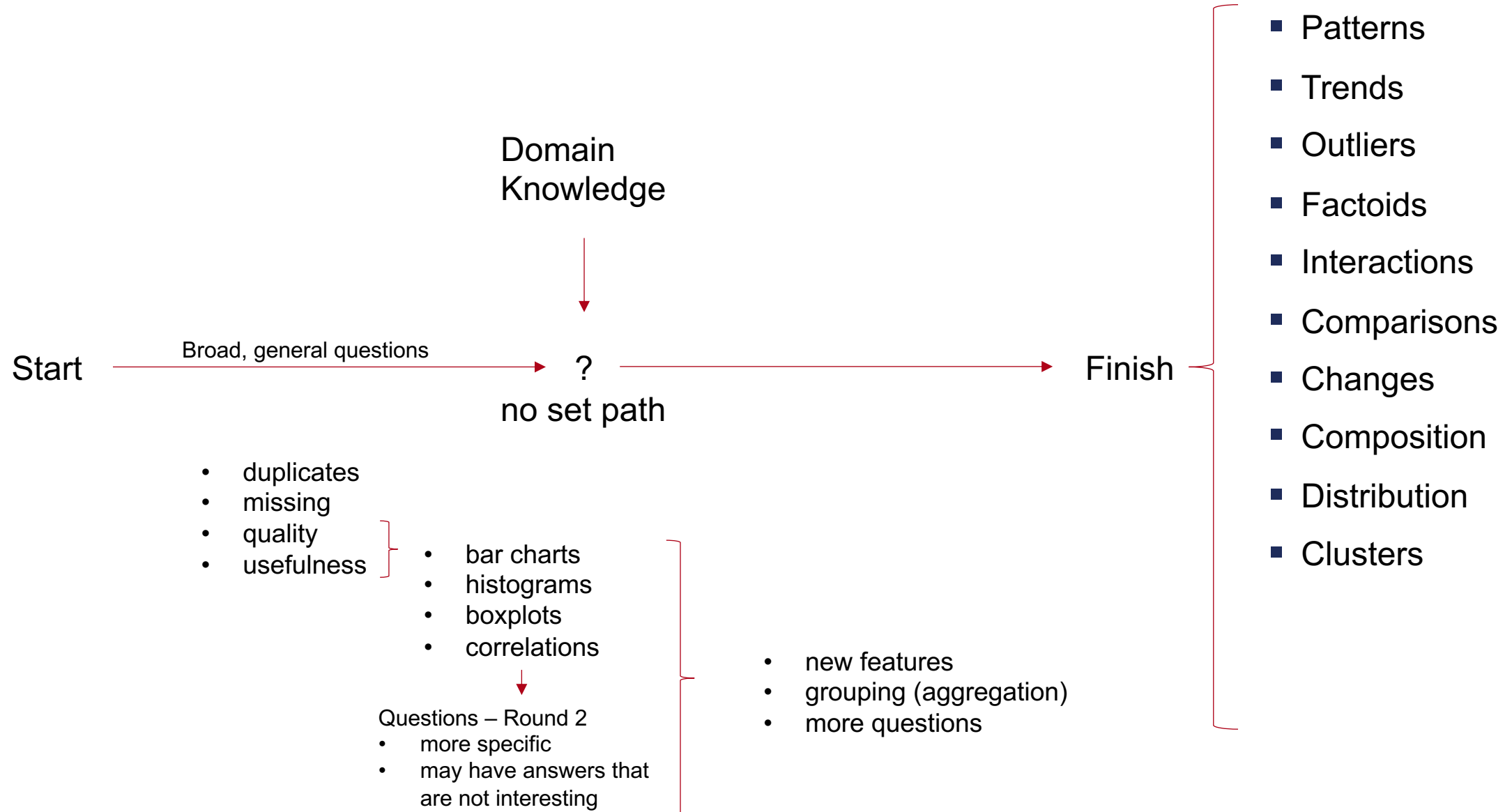
Types of analyses



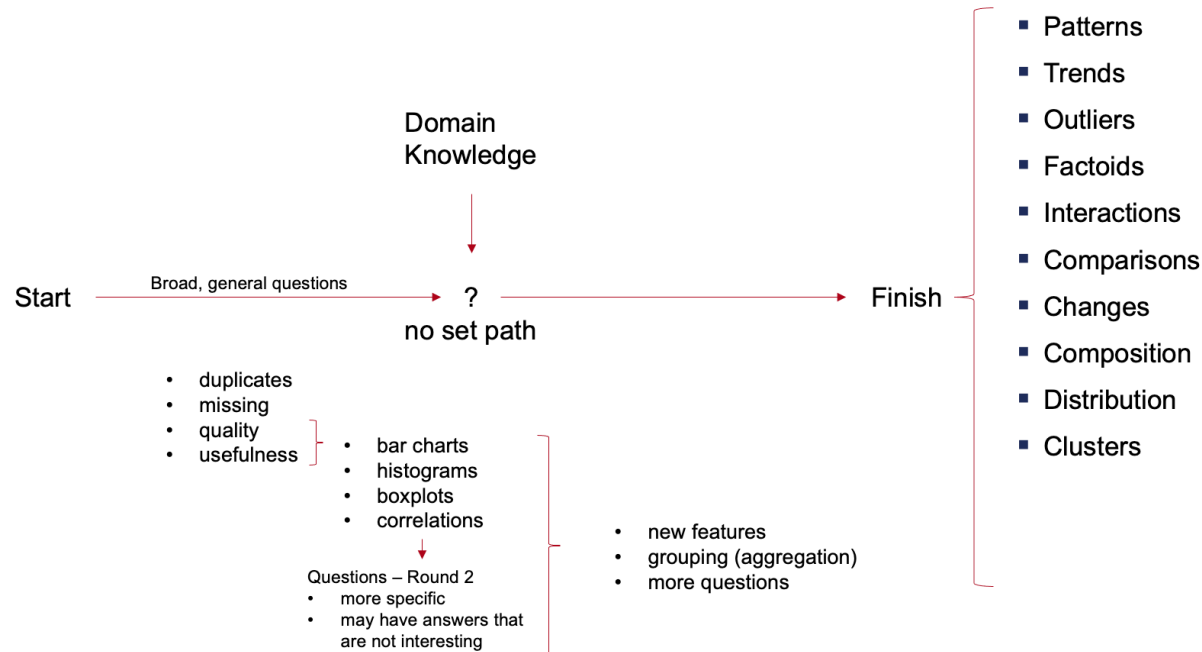
” Here’s a dataset. Tell me something interesting.”

Is there a process to go from ‘tell me something...’ to ‘here is what we found’?

Looking for insights



Approach to learning Python



1.A little coding background.

2.A high-level pass to learn a lot of basic functionality.

3.Expanding into more complex functionality.

- ☐ 0 - Working in Google CoLab.ipynb
- ☐ 1 - Python Basics.ipynb
- ☐ 10 - Pandas-datetime.ipynb
- ☐ 11 - Pandas-Merge-Join-Concat.ipynb
- ☐ 12 - Pandas-Reshape.ipynb
- ☐ 2 - NumPy.ipynb
- ☐ 3 - Matplotlib_plt.ipynb
- ☐ 4 - Seaborn.ipynb
- ☐ 5 - Pandas-Intro.ipynb
- ☐ 6 - Pandas-IngestData.ipynb
- ☐ 7 - Pandas-Indexes.ipynb
- ☐ 8 - Pandas-inspect_clean.ipynb
- ☐ 9 - Pandas-features and groupby.ipynb

A large, dark blue circle with a lighter blue border is positioned in the top-left corner of the slide. It is partially cut off by the left and top edges.

Python

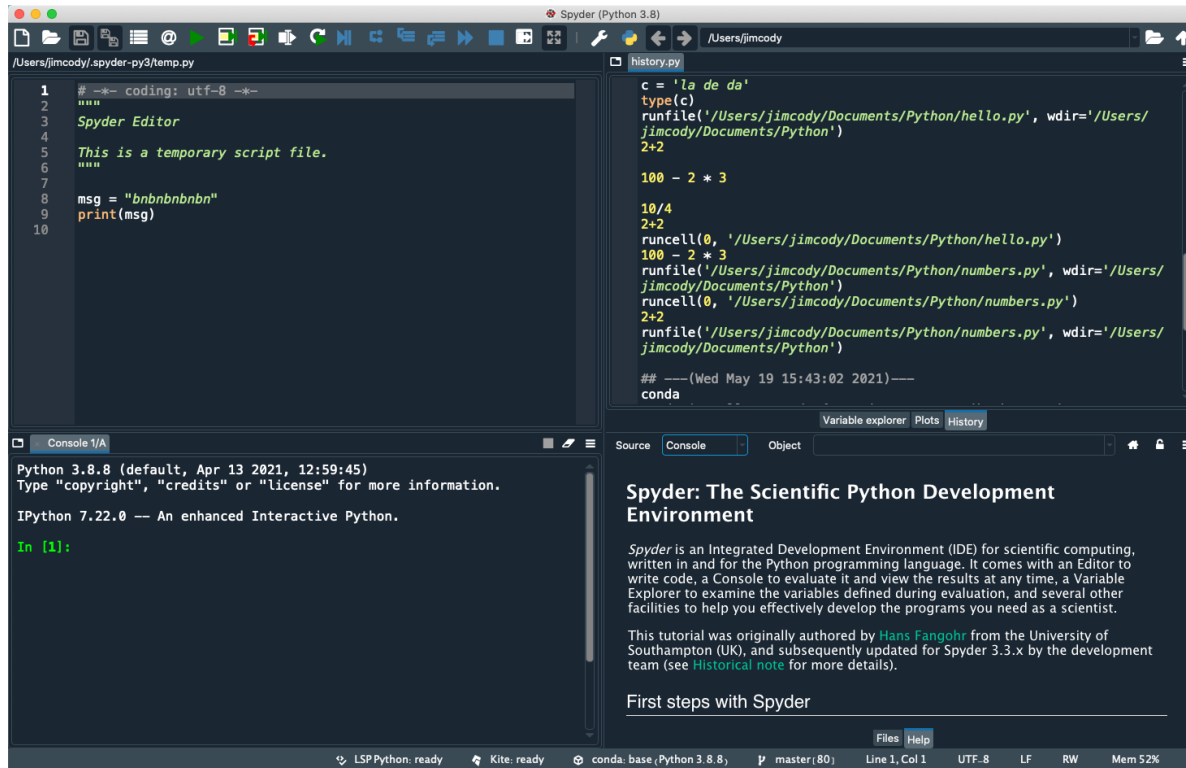
What is Python?

- **Python** is a high-level, interpreted scripting language developed in the late 1980's.

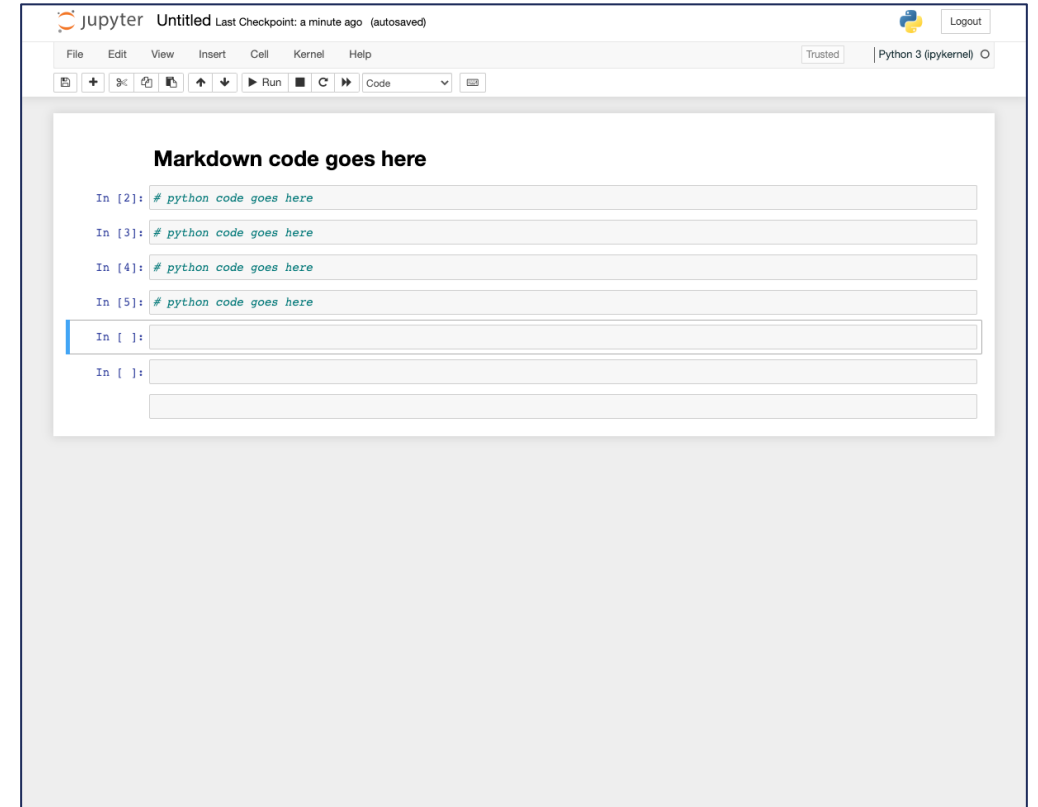
A scripting language is a programming language that is interpreted. It is translated into machine code when the code is run, rather than beforehand.

Two main working environments

IDE Environments & scripts



jupyter notebooks



jupyter notebook on Google CoLab

The screenshot shows a Google Colab notebook titled "0 - Working in Google CoLab.ipynb". The interface includes a top menu bar with options like File, Edit, View, Insert, Runtime, Tools, and Help. Below the menu, there are tabs for "+ Code" and "+ Text", and a "Copy to Drive" button. On the right, there are status indicators for RAM and Disk usage, and an "Editing" mode button.

The notebook content is organized into sections:

- Attach Your Google Drive**
You will have to do this everytime you open a notebook.
A code cell is shown with the following code:

```
[1] 1 from google.colab import drive
    2 drive.mount('/content/drive')
```
- Upload outbreaks to your google drive**
 - From within the notebook on CoLab:
 - Drive > MyDrive > Press the three dots and select 'New Folder'
 - Name the folder Intro
 - Go to the Intro folder and upload **outbreaks2.csv**
- Read from your Google Drive**
A code cell is shown with the following code:

```
[ ] 1 import pandas as pd
```

In the files directory:

 - go to the file you want to load
 - click on the three dots and select 'Copy path'
 - paste the path into the read statement
Another code cell is shown with the following code:

```
[ ] 1 vac = pd.read_csv('/content/drive/MyDrive/Intro/outbreaks2.csv')
    2 vac.head()
```
- Read from Github**
A code cell is shown with the following code:

```
[ ] 1 latuda = pd.read_csv('https://raw.githubusercontent.com/jimcody2014/nlp_cdc/main/data/latuda.csv')
    2 latuda.head()
```

A few pandas basics

<https://pandas.pydata.org/docs/reference/frame.html>

```
latuda = pd.read_csv('https://raw.githubusercontent.com/jimcody2014/Intro2022/main/data/latuda.csv')
latuda.head()
```

Dataframe

	drugName	condition	review	rating	date	usefulCount	← column index
0	Latuda	Bipolar Disorde	"I have had great experience so far with Latud...	8	20-Feb-12	39	
1	Latuda	Bipolar Disorde	"I've taken a lot of medications. I was prescr...	3	28-Oct-16	25	
2	Latuda	Bipolar Disorde	"I was deeply suicidal for 15 years with repet...	9	25-Dec-15	33	
3	Latuda	Bipolar Disorde	"I have been taking Latuda 80 mg for 8 months ...	3	23-Apr-13	34	
4	Latuda	Schizophrenia	"We've heard of the Latuda 20-40 mg restlessne...	9	5-Feb-16	15	

dataframe attributes

df.columns
df.shape
df.axes
df.size
....

dataframe methods

df.head()
df.tail()
df.groupby()
df.mean()
df.drop_duplicates()
....

Some terminology

■ Module

- collection of functions
- has a file extension `.py`

■ Package

- collection of related modules

■ Library

- a loose umbrella term

All packages are modules.
Not all modules are packages.

`matplotlib.pyplot.boxplot()`

package
(sometimes
called a library;
sometimes
called a module)

module
(sometimes called
a submodule)

function

