Github: https://github.com/ReneStander/BMG_Intro_R_Python

Introduction to Python Programming

Dr René Stander & Dr Paul J van Staden

Department of Statistics, University of Pretoria

29 & 30 September 2025

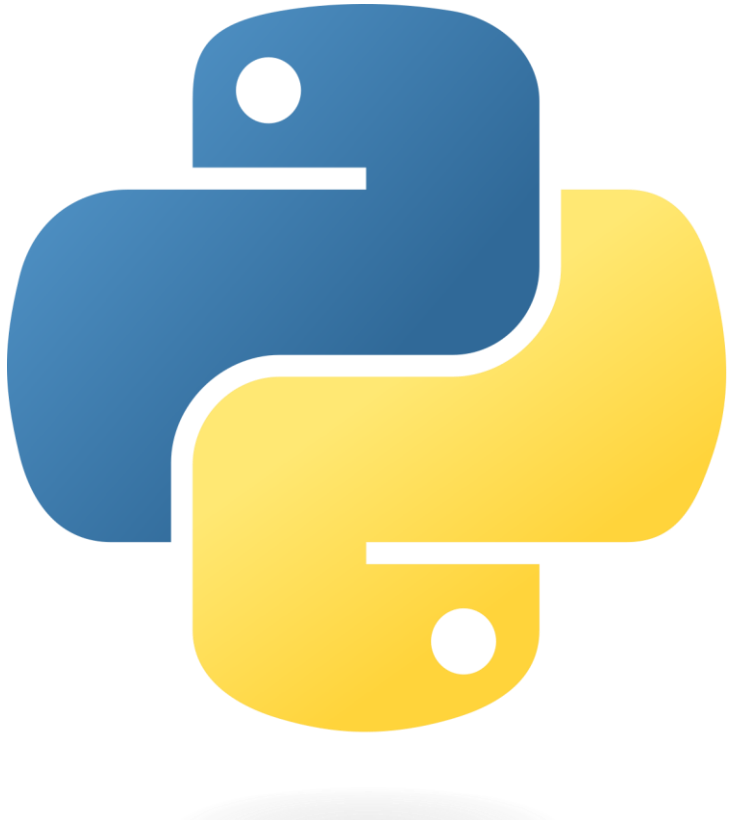
BILL & MELINDA GATES foundation

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

saMRC advancinglife

Outline

1. Introduction to Python
2. Basics of the programming language
3. Data manipulation
4. Data visualization
5. Basic statistical analysis



Program

09:00 – 10:30	Session 1
10:30 – 10:45	Break
10:45 – 12:00	Session 2
12:00 – 13:00	Lunch
13:00 – 14:30	Session 3
14:30 – 14:45	Break
14:45 – 16:00	Session 4

Resources

- Downey, A.B., 2012. *Think python*. O'Reilly Media, Inc. https://allendowney.github.io/ThinkPython/#

- Adhikari, A., DeNero, J. and Wagner, D., 2022. *Computational and Inferential Thinking: The Foundations of Data Science*, Second edition, University of California, Berkeley. https://inferentialthinking.com/

Introduction to Python

Why Python?

- Mature programming language.
- Has excellent properties for newbie programmers.
- Currently one of the most flexible programming languages.
- Has a large ecosystem of libraries making it easy for Data Scientists.

Applications developed with Python you may know…

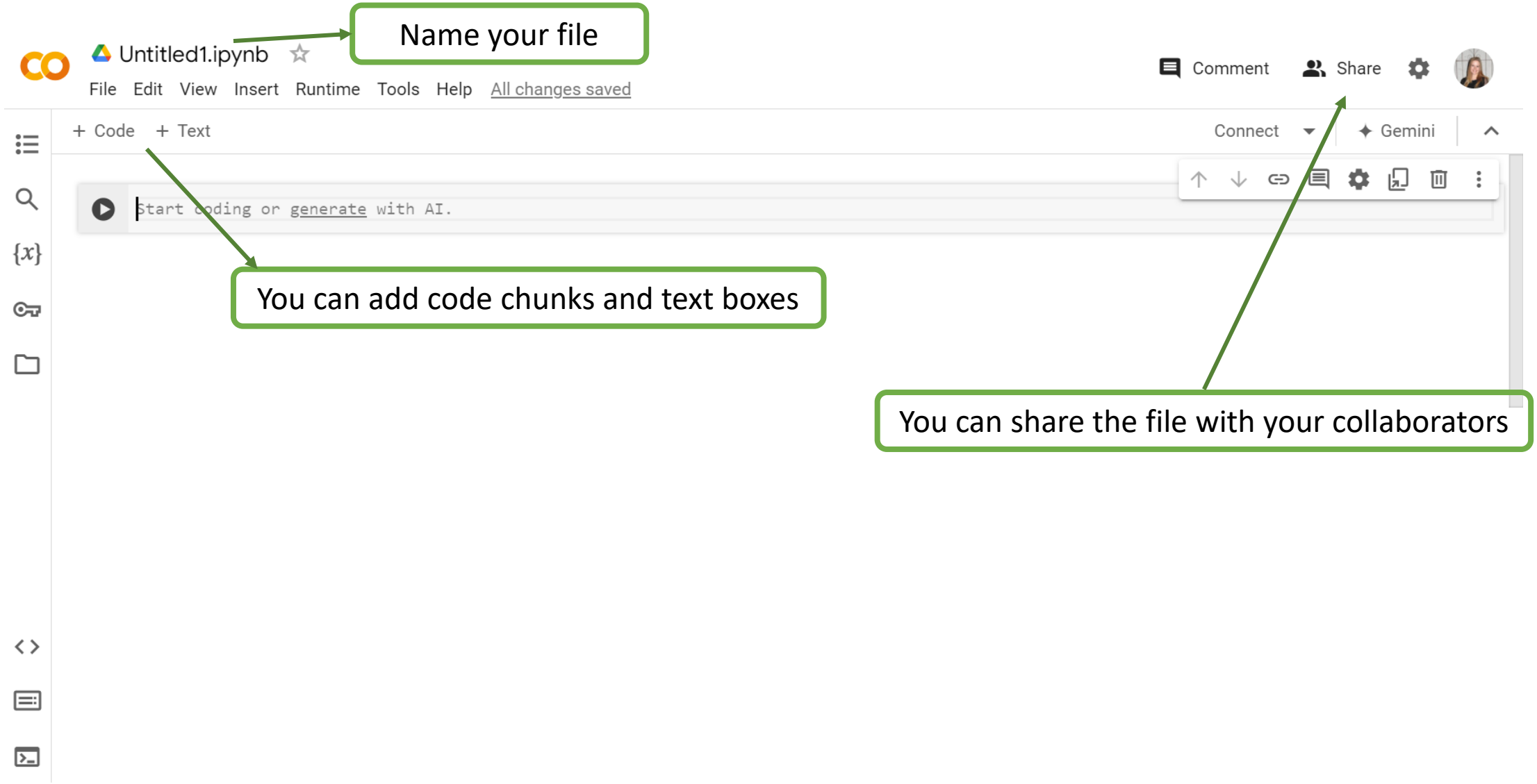
Running Python in Google Colab

Google Colab

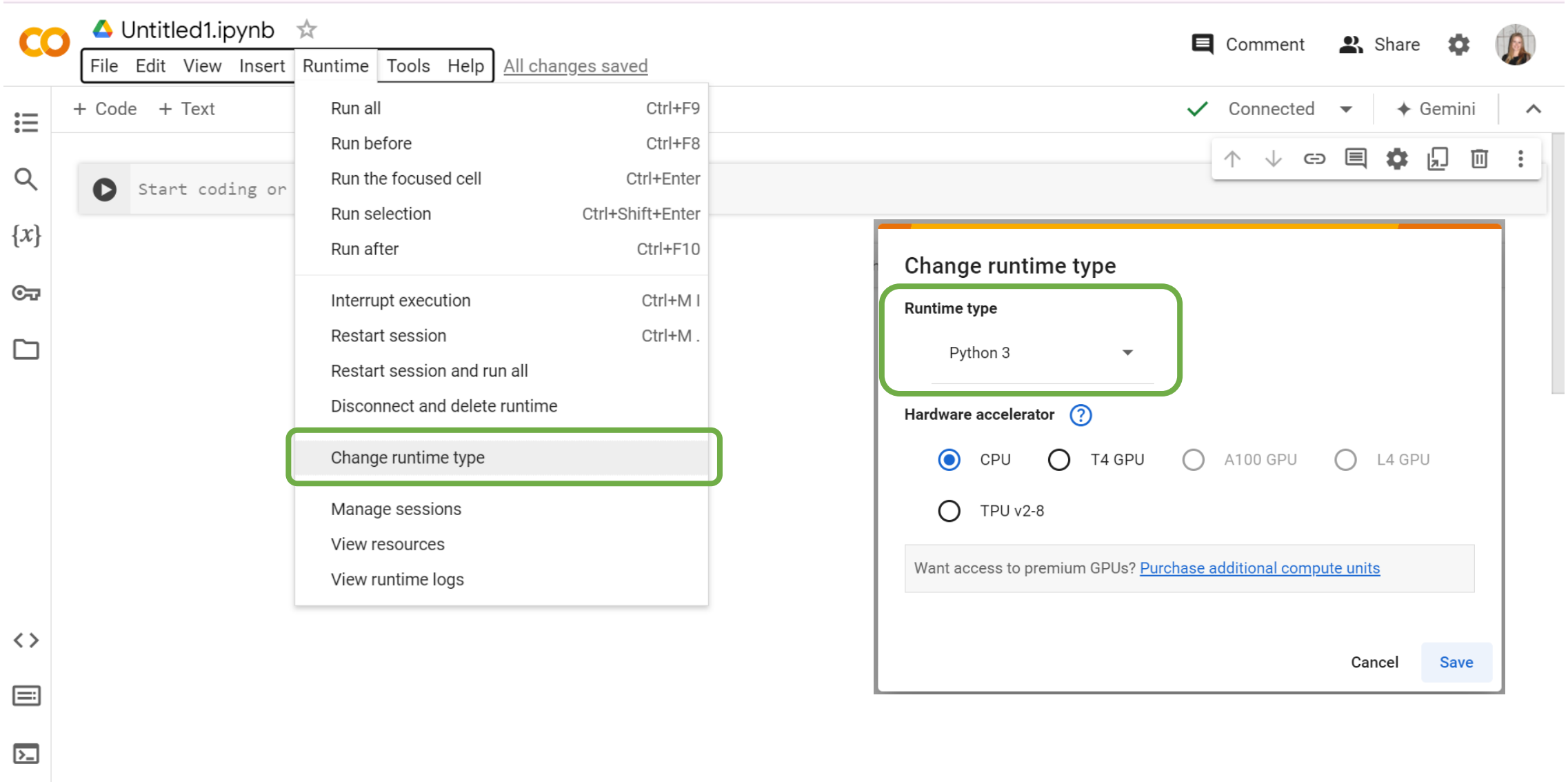
- Hosted **Jupyter Notebook** service.
- Required **no** hardware or software to be installed on your device.
- Provides free access to computing resources such as GPUs.

https://colab.google/notebooks/

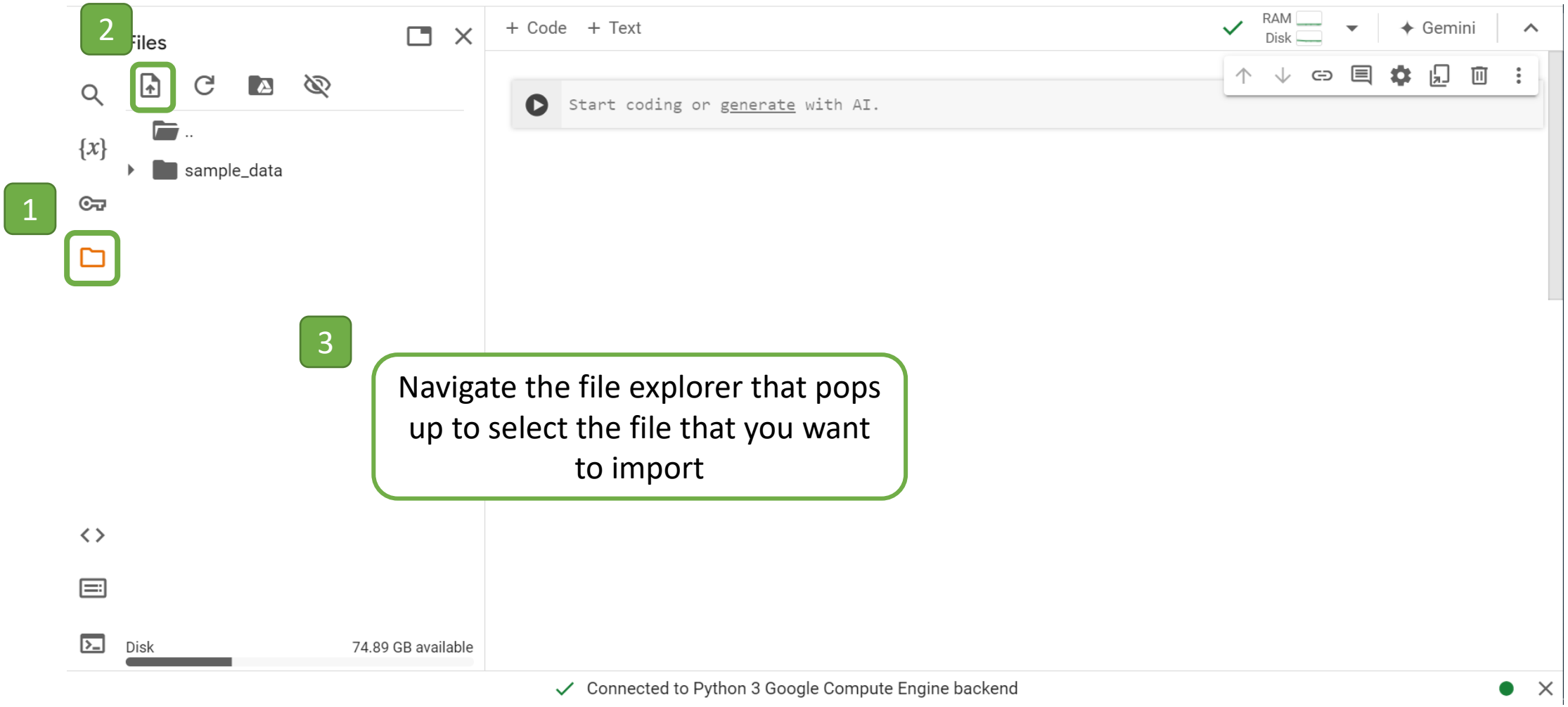
Running Python in Google Colab



Running Python in Google Colab



Running Python in Google Colab



Installing Python

1. Download Anaconda navigator. https://www.anaconda.com/download

2. Access the Python IDE, **Spyder**, through the Anaconda navigator.

Fundamental Python Libraries for Data Scientists

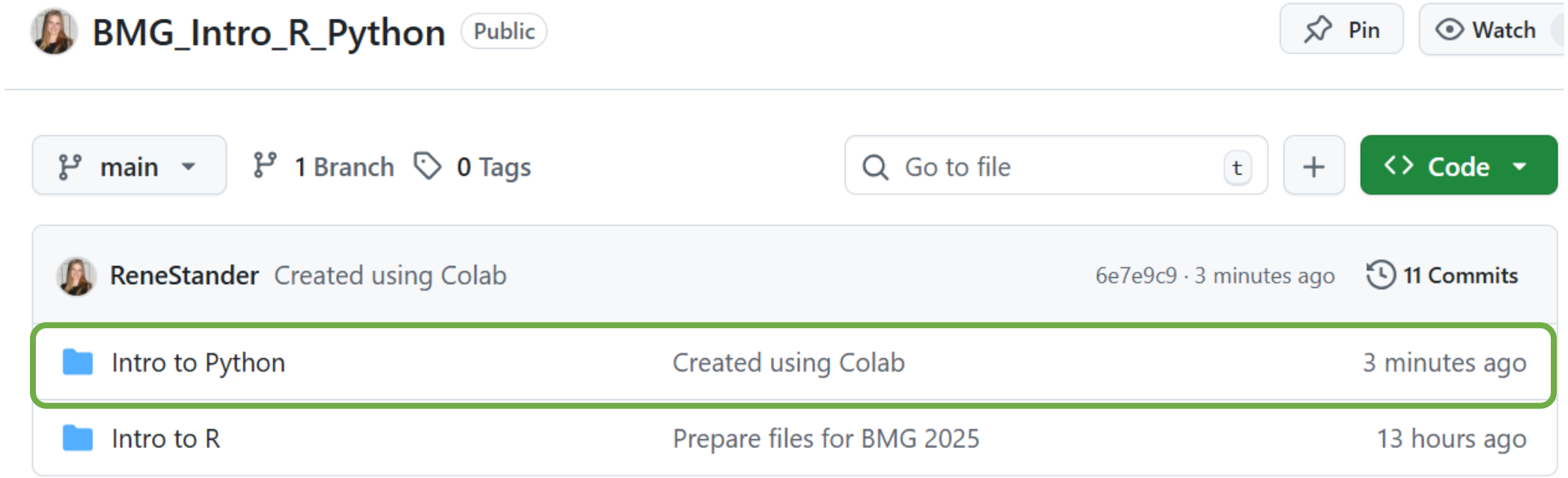
- **NumPy:** Provides support for multidimensional **arrays with basic operations on them** and useful linear algebra functions.
- **SciPy:** Provides a collection of **numerical algorithms** and domain-specific toolboxes, including signal processing, optimization and statistics.
- **Matplotlib:** Enables **data visualization**.
- **Pandas:** Provides high performance **data structures** and data analysis tools.
- **Scikit-Learn:** Offers simple an efficient tools for **common tasks in data analysis** such as classification, regression, clustering, and many more…

Coding scripts for this course

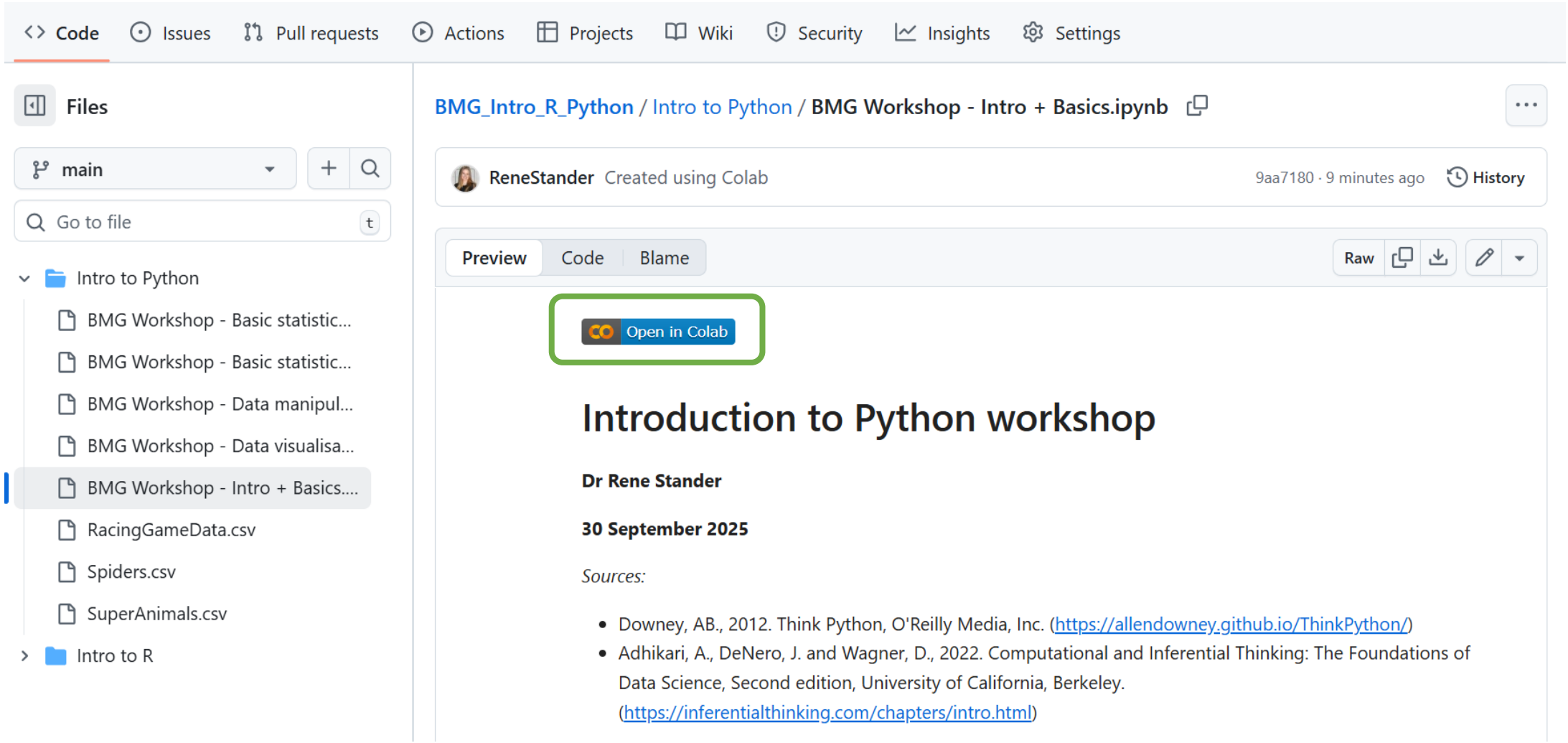
Github:

https://github.com/ReneStander/BMG_Intro_R_Python

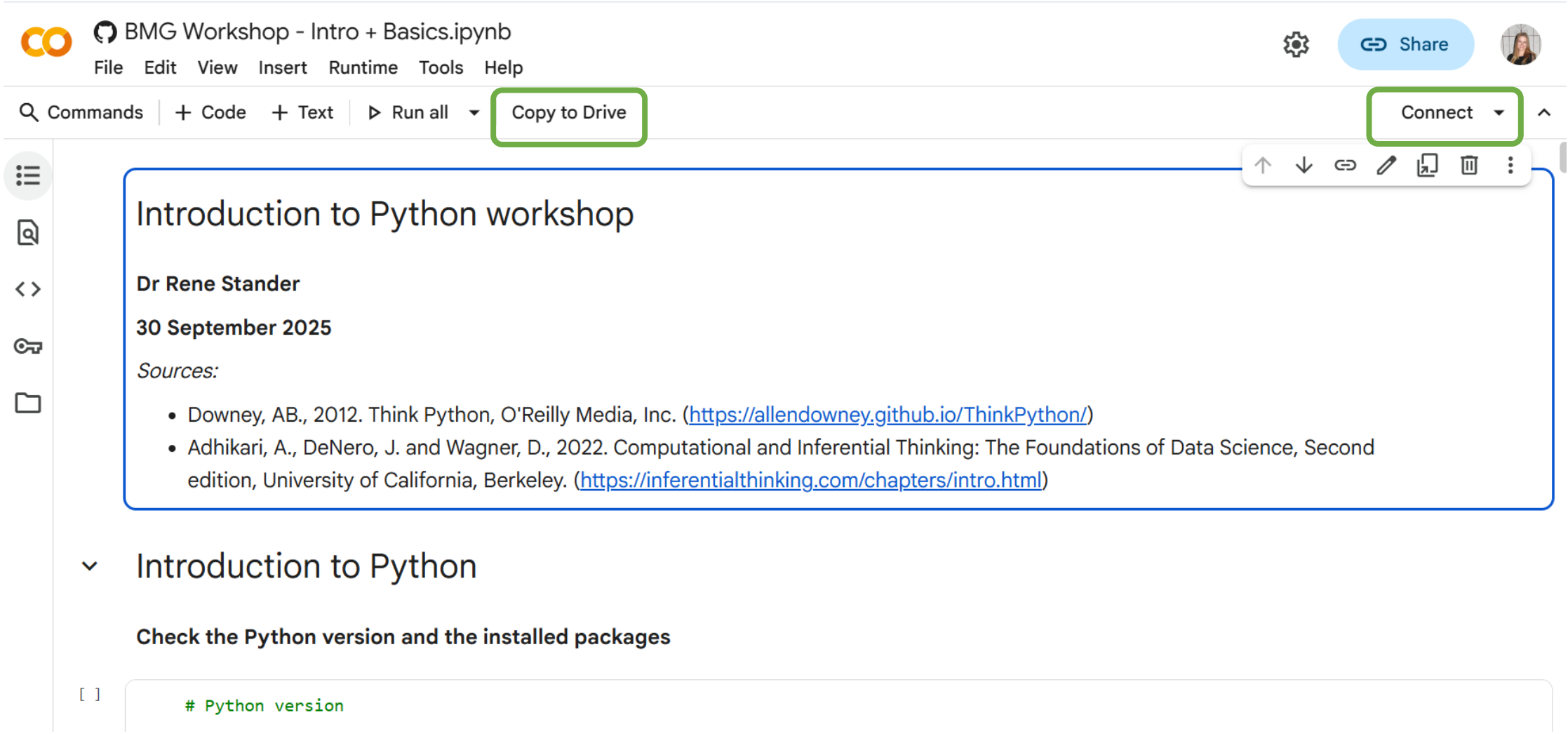
Coding scripts for this course



Coding scripts for this course



Coding scripts for this course



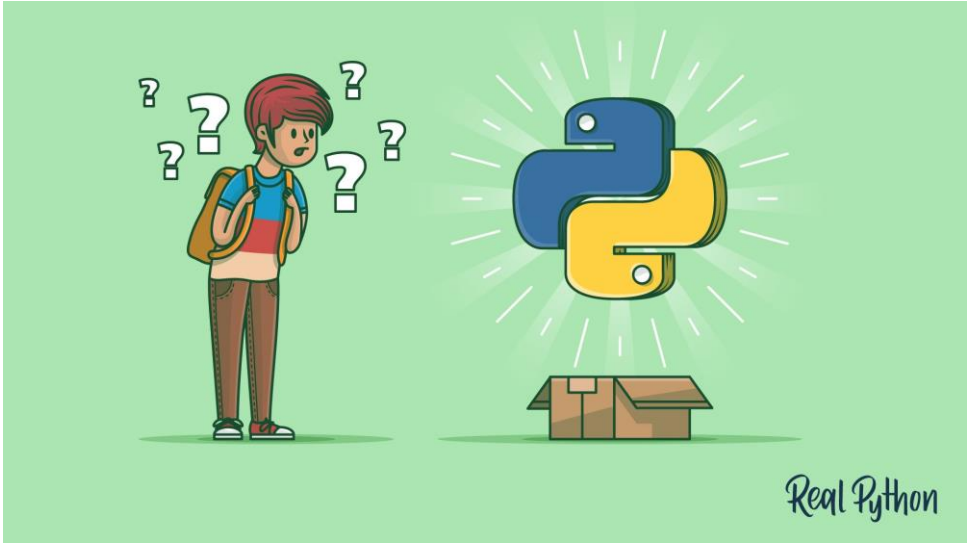
Intro + Basics

Open the notebook named **`BMG Workshop – Intro + Basics.ipynb`** in Google Colab

This part of the work does not require any additional data sets

Github:

https://github.com/ReneStander/BMG_Intro_R_Python



Data manipulation

Open the notebook named **`BMG Workshop – Data manipulation.ipynb`** in Google Colab

To run the code in this notebook, please upload the following data sets into Google Colab:

- **`RacingGameData.csv`**

Github:

https://github.com/ReneStander/BMG_Intro_R_Python

Data visualisation

Open the notebook named **`BMG Workshop – Data visualisation.ipynb`** in Google Colab

To run the code in this notebook, please upload the following data sets into Google Colab:

- **`SuperAnimals.csv`**
- **`Spiders.csv`**

Github:

https://github.com/ReneStander/BMG_Intro_R_Python

Data visualisation

Open the notebook named **`BMG Workshop – Data visualisation.ipynb`** in Google Colab

To run the code in this notebook, please upload the following data sets into Google Colab:

- **`SuperAnimals.csv`**
- **`Spiders.csv`**

Github:

https://github.com/ReneStander/BMG_Intro_R_Python

Statistical analysis – Categorical variables

Open the notebook named **`BMG Workshop – Basic statistical analysis – Categorical variables.ipynb`** in Google Colab

This part of the work does not require any additional data sets

Github:

https://github.com/ReneStander/BMG_Intro_R_Python

Statistical analysis – Categorical variables

Open the notebook named **`BMG Workshop – Basic statistical analysis – Numerical variables.ipynb`** in Google Colab

To run the code in this notebook, please upload the following data sets into Google Colab:

- **`SuperAnimals.csv`**
- **`Spiders.csv`**

Github:

https://github.com/ReneStander/BMG_Intro_R_Python