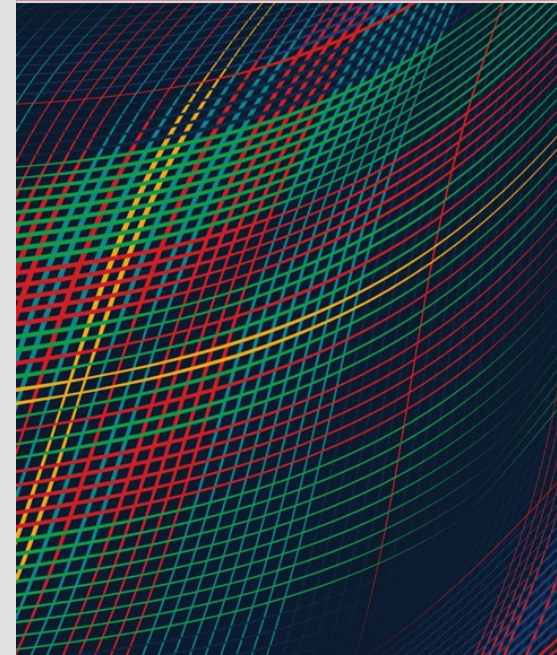


# Google Colab: Environment Setup for the 2023 InfoSecWorld Cybersecurity Data Science Workshop

**SEPTEMBER 25-27, 2023**

Clarence Worrell  
Senior Data Scientist



# Document Markings

Copyright 2023 Carnegie Mellon University.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

References herein to any specific commercial product, process, or service by trade name, trade mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by Carnegie Mellon University or its Software Engineering Institute.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other use. Requests for permission should be directed to the Software Engineering Institute at [permission@sei.cmu.edu](mailto:permission@sei.cmu.edu).

DM23-0837

# Environment options

- **Option 1: Google Colab (recommended)**

- Online alternative to using a local Anaconda installation
- Free data analysis and machine learning tool
- Write and execute python code in a browser
- Mix rich text, coding, and output into a well-formatted PDF report
- No installations required
- Requires a gmail account
- <https://colab.research.google.com>

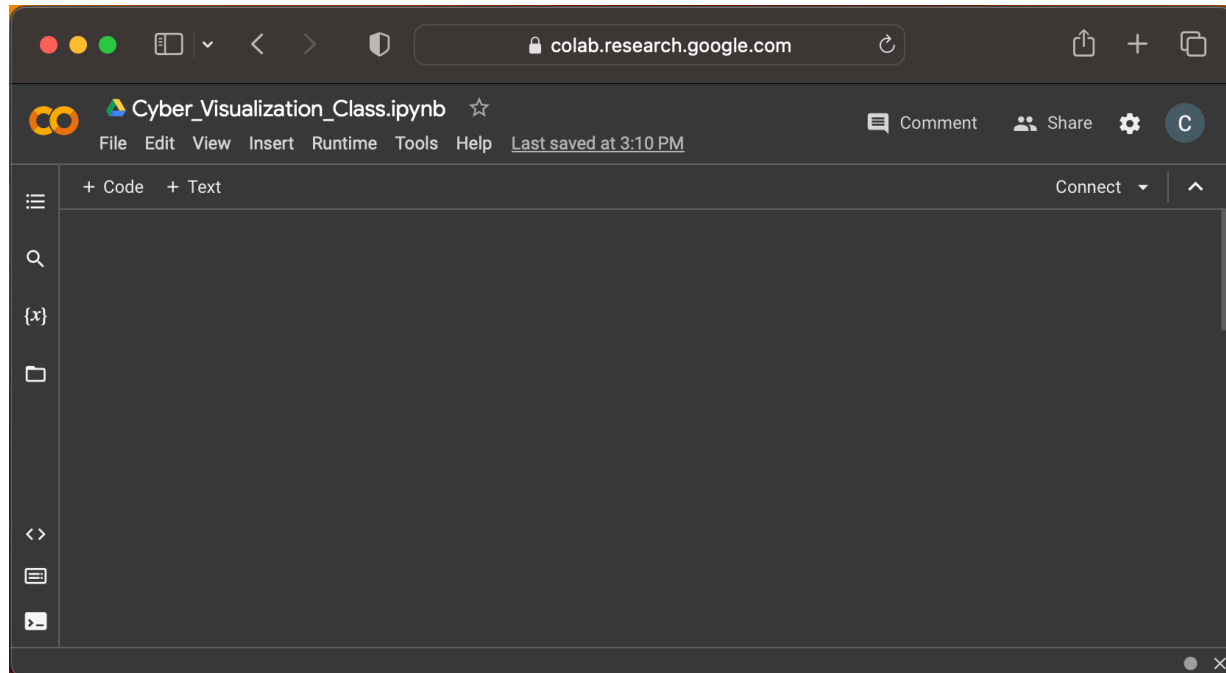
- **Option 2: Student-preferred python environment**

- Students are welcome to use their own preferred python environment
- For example, a Jupyter notebook using an Anaconda environment will work for the exercise
- You will need capability to install new packages, such as umap-learn

- We show how to set up Google Colab for the workshop in the following slides

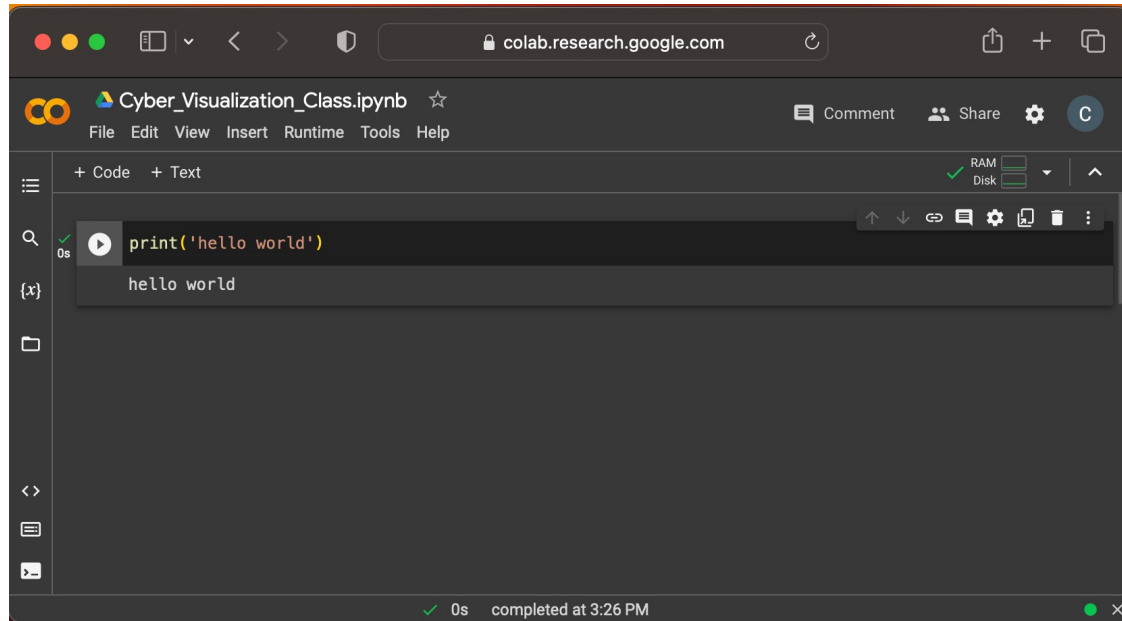
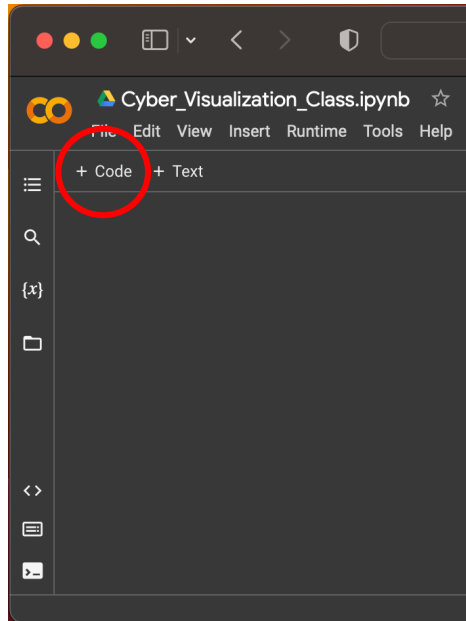
# Colab: Getting started

- Sign into your gmail account in your browser
- Go to <https://colab.research.google.com>
- Click “File → New Notebook”



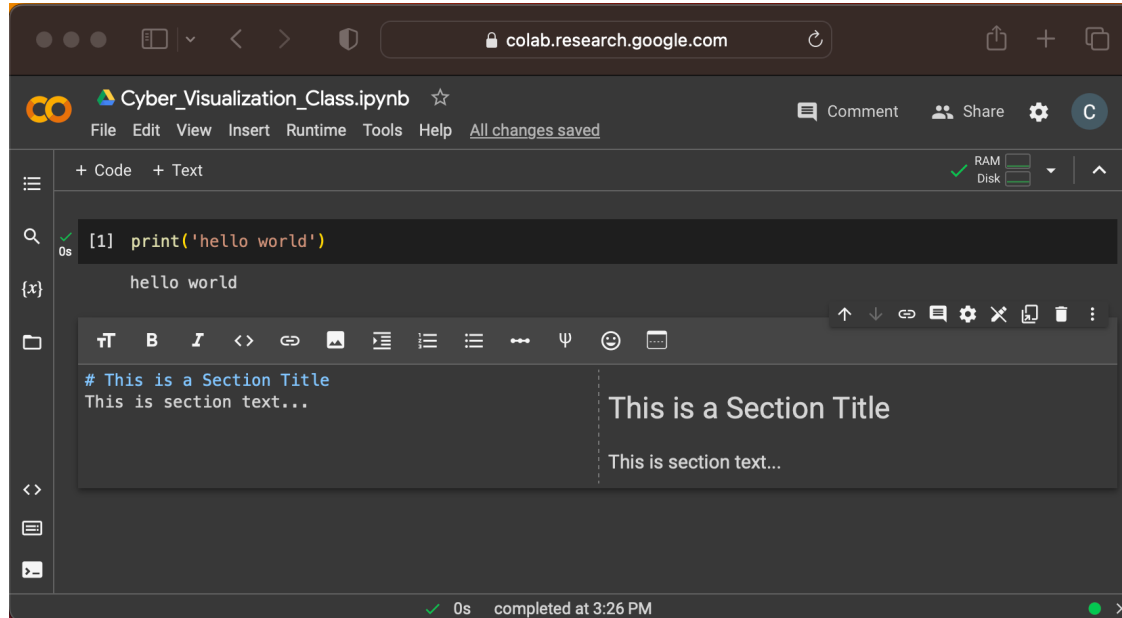
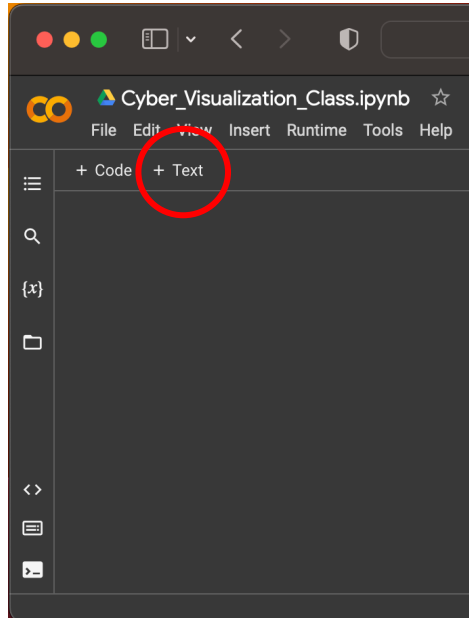
# Colab: Getting started (cont.)

- Click “+ Code” to create a code cell
- Type: `print('hello world')`
- Press: CNTL-Enter, or press the “play button”, to execute the cell



# Colab: Getting started (cont.)

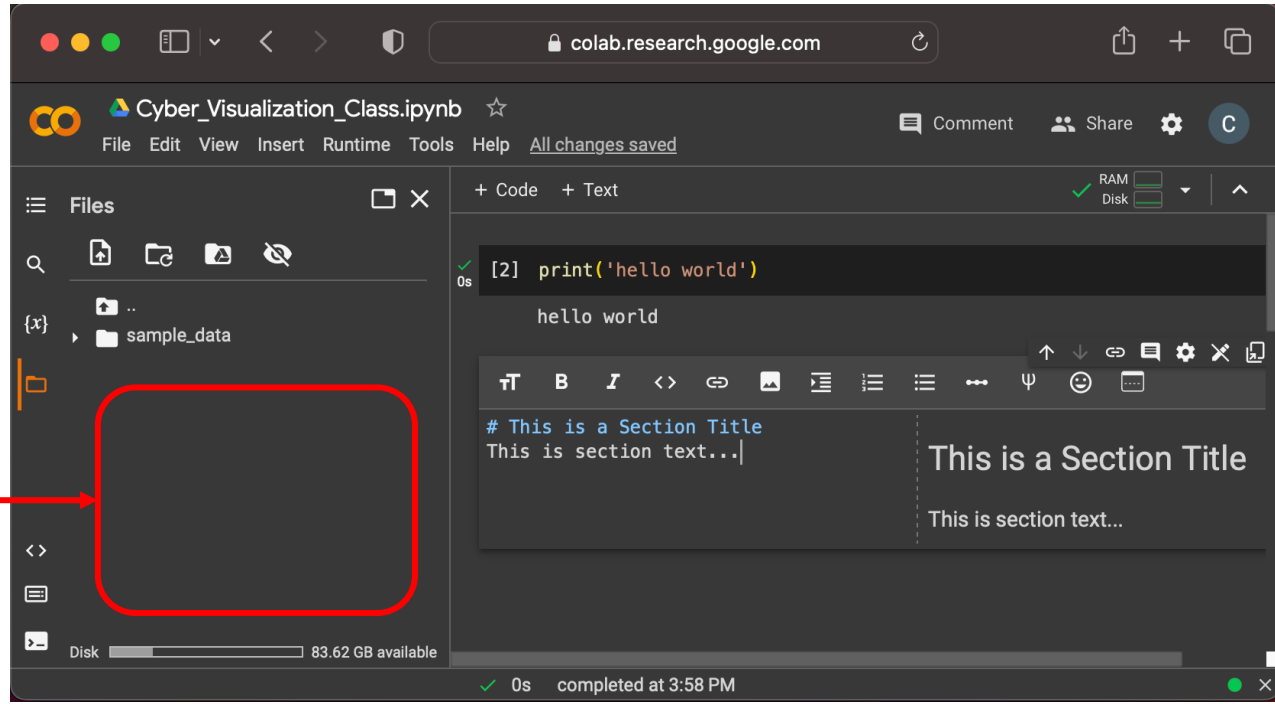
- Click “+ Text” to create a text cell
- Allows adding formatted text throughout your notebook to explain the analysis
- Use # (hashtag) for section titles



# Import data

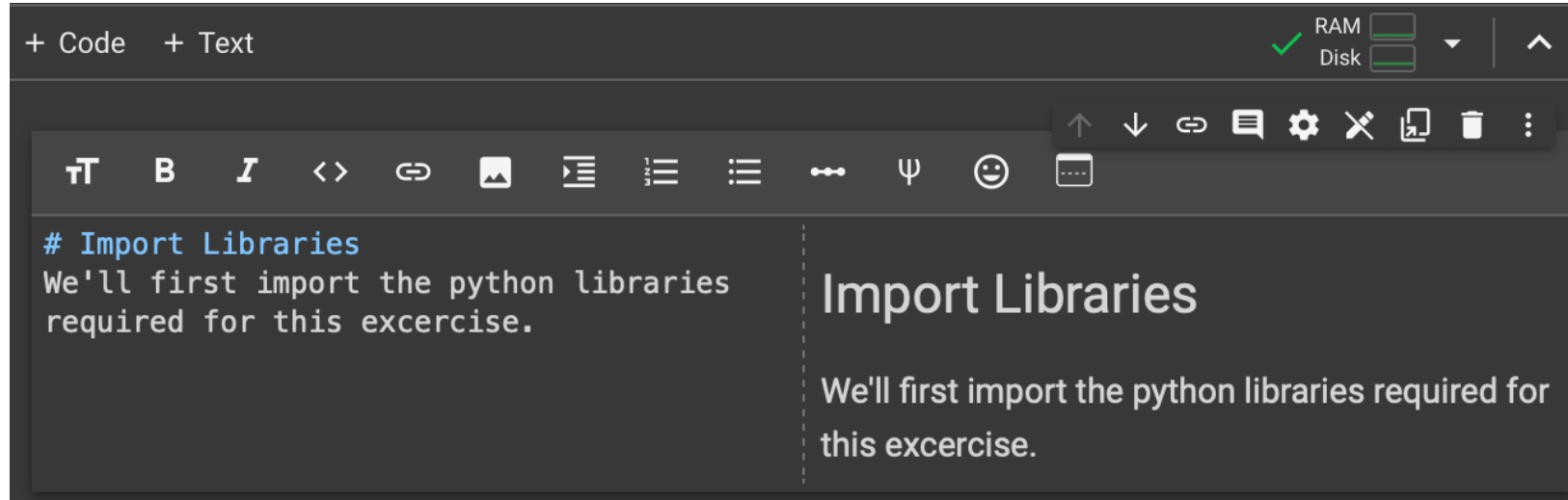
- Import `data.csv` into Colab

Drag-and-drop the .csv file into this area. It will take a moment to upload.



# Import libraries

- Click “+ Text” to add a new text cell
- Add some formatted text to indicate that we'll first import libraries





# Import libraries (cont.)

- Click “+ Code” to add a new code cell
- Add the following code and press CNTL-Enter to execute the cell

## ▼ Import Libraries

We'll first import the python libraries required for this exercise.

```
✓ 0s [5] import matplotlib.pyplot as plt  
      import pandas as pd  
      import seaborn as sns
```

# Setup complete

- You've now completed setup and testing of Google Colab for the the workshop.

# Workshop instructors



**Tom Scanlon**  
Technical Manager



**Clarence Worrell**  
Senior Data Scientist

**For additional information:**

[info@sei.cmu.edu](mailto:info@sei.cmu.edu)