



HKUST
VISLAB

HKUST
HCI Initiative

COMP 4462

Data Visualization Tutorial

CHEN Chang
PAN Ziqi

Friday 4 October, 2024

Python, Jupyter Notebook and Google Colab

- Python

- Interpret language (i.e. run line by line, no compilation)
- Dynamic typing (i.e. do not need to declare types like int, float, char*, etc.)
- Easy to pick up, widely used, a lot of learning resources online
- A lot of libraries, fast and powerful
 - Pandas, scikit-learn, matplotlib, altair and a lot of machine learning related libraries

- Jupyter Notebook

- Interactive environment (i.e. you know exactly what it does)
- Design for experimentation
 - Just like a notebook in laboratory, recoding every steps of experiment and results
 - Form hypothesis (setup code) => experiment (run code) => see results (code output) => refine hypothesis (change code) => experiment again (run code again) => results => loop

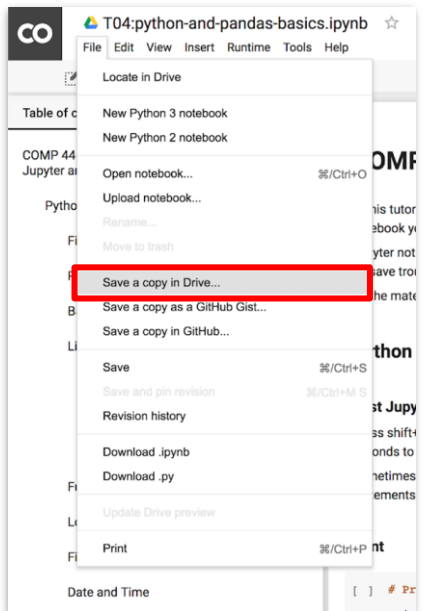
- Google Colab

- A hosted version of Jupyter Notebook
- Provided by Google, free!

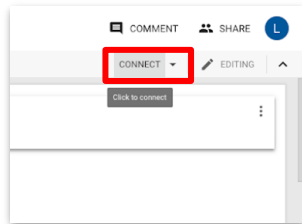
Google Colab

1. Sign in your Google account
2. Go to the [notebook of this tutorial](#)

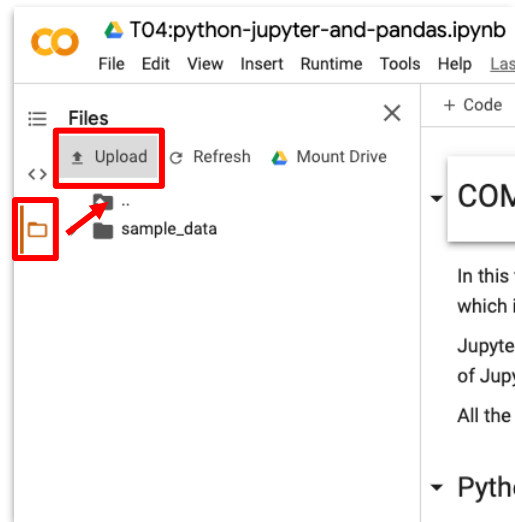
3. Make a copy



4. Connect



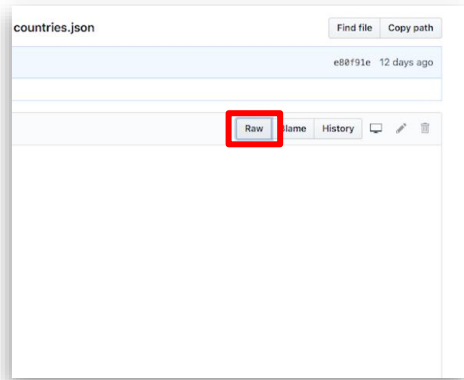
5. Upload dataset



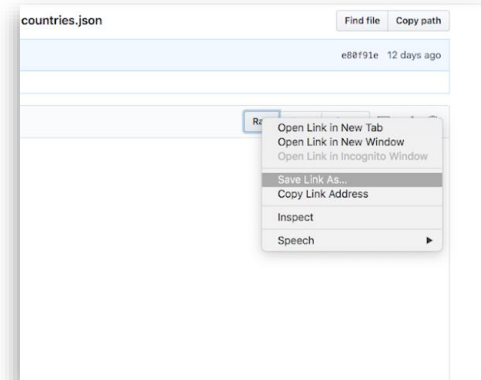
Download dataset from GitHub

1. Go to the [tutorial repository](#)
2. Go to the dataset file you want download, e.g. [pokemon.csv](#)

3. Right click “Raw”

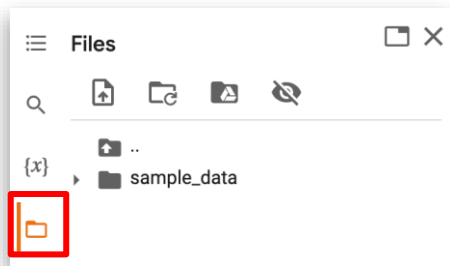


4. Save as file

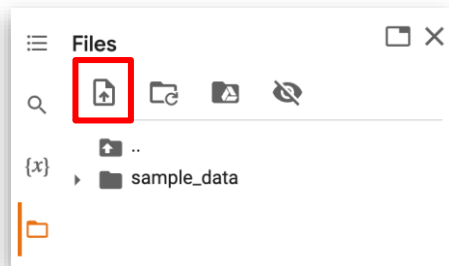


Upload dataset to Colab

1. Switch to “Files”



2. Click the “Upload” button



3. Select the dataset file you want to upload, e.g. [pokemon.csv](#).
4. You may load the dataset with Pandas using `pd.read_csv()`.

Python / Jupyter Notebook Basics

- See the [Jupyter notebook on Google Colab](#)
- Topics:
 - Print statement / comments / shift-enter / Tab completion
 - Data type / string / numeric / booleans / string to int / None
 - List (retrieve / add / remove / combine / sort) / dict / set / tuple / unpack
 - List / set / dict comprehensions
 - Logic flow / loop / range
 - Define function / call function / named arguments / return value(s)
 - Module loading
 - File read / write
 - Dates and times
 - Basic pandas usage
 - Plotting with pandas
 - Scikit-learn with pandas

Lab exercise

- Tasks

- Open [this Google Colab notebook](#), make a copy and connect
- Read through “Python and Jupyter Basics” and fill in the “TODO” cells
- Download Pokemon dataset (pokemon.csv) from [GitHub](#)
- Upload the dataset to Google Colab workspace
- Read dataset using pandas
- Print the loaded dataframe to output
- Sort the dataframe by “base_total” and print to output
- Plot with “.plot()” function of pandas dataframe
- How to submit: **print the whole web page as .pdf and upload to Canvas**
 - Mac: cmd+p
 - Windows: ctrl+p

- Optional

- If you like this tutorial so far, star [our GitHub repository](#) ★ ★ ★ Thank you! ❤️
- Explore the Pokemon dataset, we will revisit it again in later tutorials

More topics on Python / Jupyter / Pandas

- **A lot** more Python language features
 - Lambda, partial argument, spread operator, zip, dict <=> lists, nested comprehensions, namespaces, class and inheritance, variable scope, positional arguments, keyword arguments, generator, iterator
 - Standard libraries
 - Regular Expressions, urllib, itertools, functools
 - Common libraries
 - Joblib, scikit-learn, numpy
 - And **a lot** more...
- **More Jupyter features**
 - JupyterLab, shortcuts, magic commands, extensions
- **More Pandas features**
 - Summary / clean data / handle null values
 - Data selection / filter
 - Sort / groupby / aggregate
 - Join dataframes
 - ...

Next tutorial

Where to find
visualizations and
interesting datasets?

- We will have:
 - Project Proposal Consultation
 - In-Class Exercise Review