HKUST
VISLAB

COMP 4462
Data Visualization Tutorial

Leo Yu Ho, Lo
Ming Yao

Tuesday 5 March, 2019

https://bit.ly/vis-t04

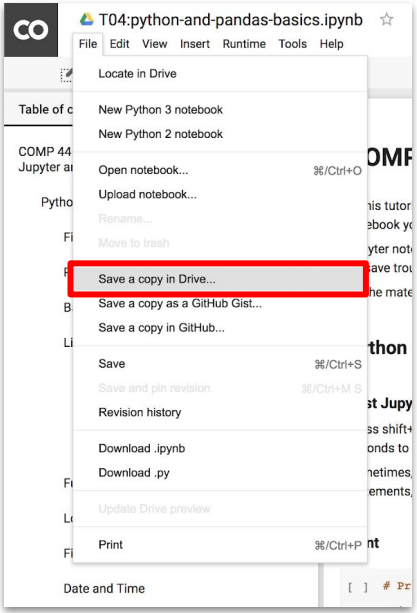
Python, Jupyter Notebook and Google Colab

- 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!

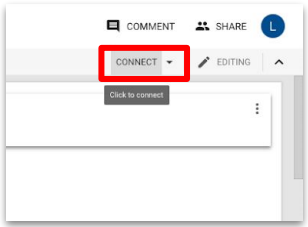
Google Colab

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

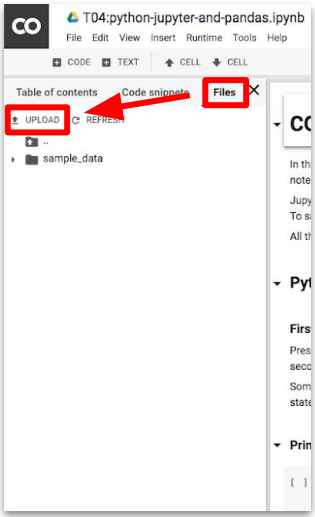
3. Make a copy



4. Connect



5. Upload dataset



Python / Jupyter Notebook Basics

- 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
- %run command

Lab exercise

- Tasks
  - Open this Google Colab notebook, make a copy and connect
  - Read through “Python and Jupyter Basics”
  - Download Pokemon dataset from GitHub
  - Upload and read dataset using pandas
  - Print the loaded dataframe
  - Sort and print the dataframe by “base_total”
  - Plot with “.plot()” function of pandas dataframe
- Optional
  - Explore the Pokemon dataset, we will revisit it again later

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
  - Next Tutorial: Pandas and Python visualizations

Next tutorial

Pandas and Python visualizations

- We will use Google Colab again
- Pandas
- Visualization libraries
  - Matplotlib / Seaborn
  - Altair