Lab 1
Setup & basic data stats

09/06/2022

Jeroen Olieslager
Richard-John Lin

Lab logistics

- Office Hours
  - Jeroen Olieslager : CDS 763 Friday 2:30PM - 3:30PM
  - Richard-John Lin : CDS 244 Friday 2:30PM - 3:30PM

- Labs are on Tuesdays, assignments are due **Monday the week after at 6pm**
  - So you have 6.5 days to complete them. Most of you will be able to finish the bulk of the work in lab however
  - 2 worst scores dropped

- Class policy:

  “Policies

  Students should try to solve problems on their own first. If stuck, one can discuss homework questions with colleagues, but should write up the final solution individually. Any violation will be penalized with a zero score for the assignment and referred to the DGS. Credit should be explicitly given for any external code use.

  There will be no special dispensations about late submissions for any reason. Instead, you have the option to drop the worst scores from the final tally (use that option wisely). Late submission penalties: 20% points off for each day of delay.”

Installing conda

Install miniconda

https://docs.conda.io/en/latest/miniconda.html

Or Install Anaconda

https://www.anaconda.com/products/distribution

To run a Jupyter Notebook, you need both ‘jupyter’ and ‘ipykernel’ packages

Useful conda commands

conda create -n {env_name} : Create an environment {env_name}

conda create -n {env_name} python==3.9 “pkg1>=2.1” pkg2

conda env list : List all created environments

conda env remove -n {env_name} : Remove the environment {env_name}

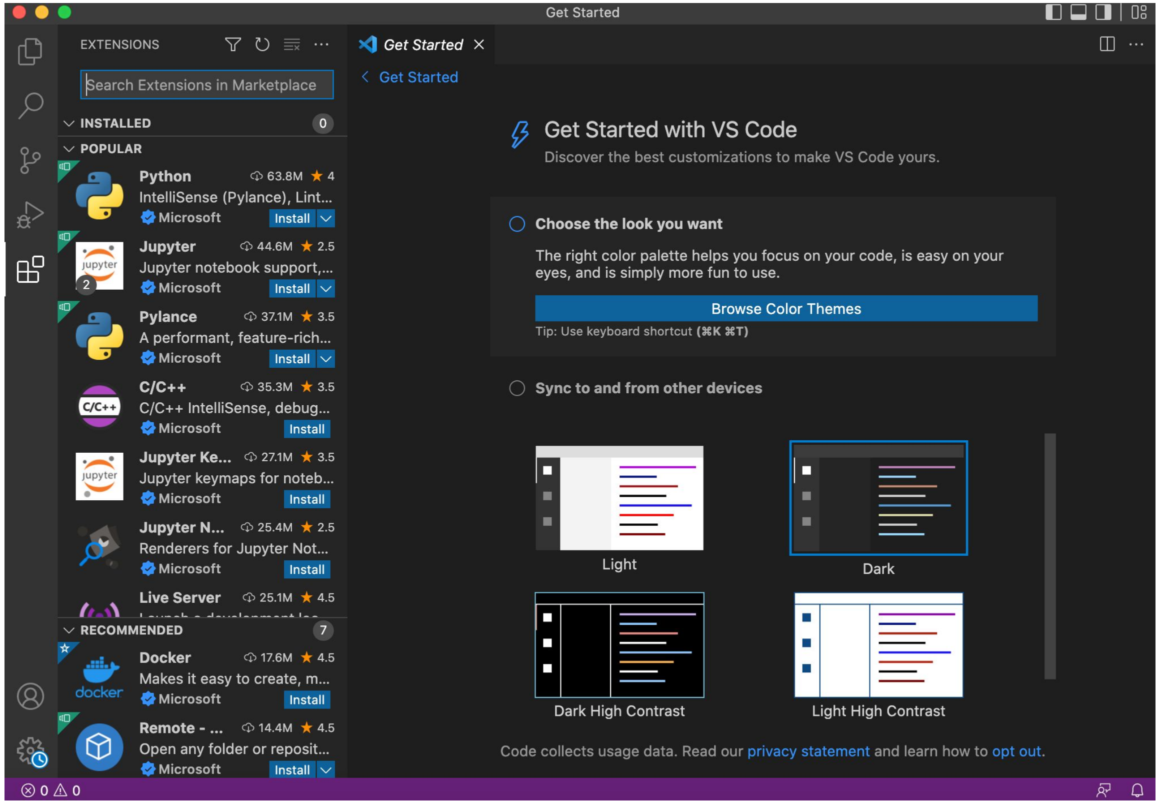
conda activate {env_name} : Activate the environment {env_name}

conda deactivate : Deactivate the current environment

conda install pkg1 pkg2 : Install pkg1, pkg2 inside current active environment

conda remove pkg1 pkg2 : Remove pkg1, pkg2 from current active environment

Installing VSCode



Get Started

EXTENSIONS

Search Extensions in Marketplace

INSTALLED 0

POPULAR

Python 63.8M ★ 4
IntelliSense (Pylance), Lint...
Microsoft Install

Jupyter 44.6M ★ 2.5
Jupyter notebook support,...
Microsoft Install

Pylance 37.1M ★ 3.5
A performant, feature-rich...
Microsoft Install

C/C++ 35.3M ★ 3.5
C/C++ IntelliSense, debug...
Microsoft Install

Jupyter Ke... 27.1M ★ 3.5
Jupyter keymaps for noteb...
Microsoft Install

Jupyter N... 25.4M ★ 2.5
Renderers for Jupyter Not...
Microsoft Install

Live Server 25.1M ★ 4.5

RECOMMENDED 7

Docker 17.6M ★ 4.5
Makes it easy to create, m...
Microsoft Install

Remote - ... 14.4M ★ 4.5
Open any folder or reposit...
Microsoft Install

Get Started

Get Started with VS Code

Discover the best customizations to make VS Code yours.

Choose the look you want

The right color palette helps you focus on your code, is easy on your eyes, and is simply more fun to use.

Browse Color Themes

Tip: Use keyboard shortcut (⌘K ⌘T)

Sync to and from other devices

Code collects usage data. Read our privacy statement and learn how to opt out.

⊗ 0 ⚠ 0

Starting a jupyter notebook in vscode

Open the command palette with

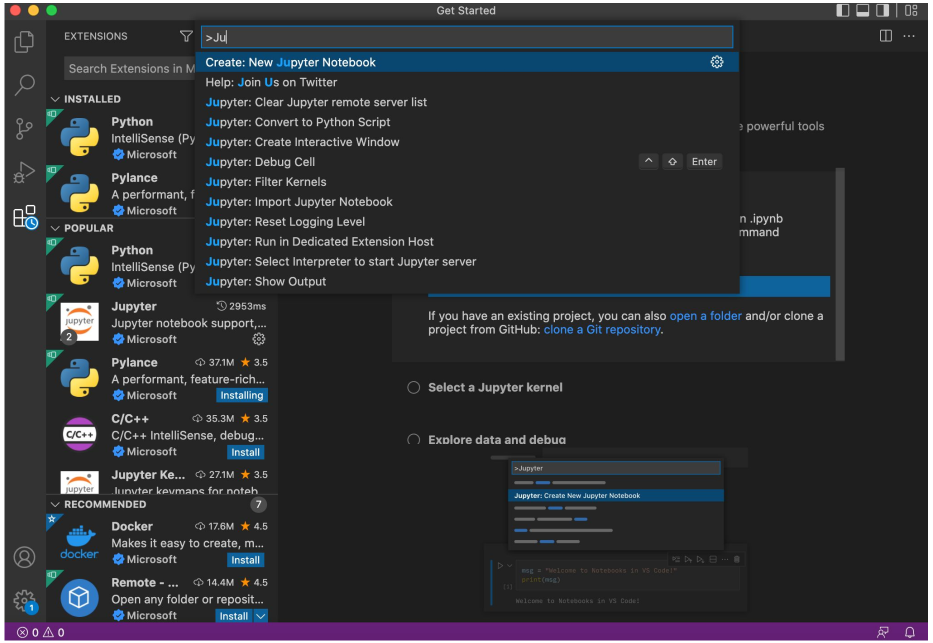
Cmd + Shift + P

Search for :

"Create: New Jupyter Notebook"

Recommended to save the notebook immediately to avoid workspace issues

Do not forget the ".ipynb" extension at the end of the name !



Get Started

EXTENSIONS

Search Extensions in M

>Ju

Create: New Jupyter Notebook
Help: Join Us on Twitter
Jupyter: Clear Jupyter remote server list
Jupyter: Convert to Python Script
Jupyter: Create Interactive Window
Jupyter: Debug Cell
Jupyter: Filter Kernels
Jupyter: Import Jupyter Notebook
Jupyter: Reset Logging Level
Jupyter: Run in Dedicated Extension Host
Jupyter: Select Interpreter to start Jupyter server
Jupyter: Show Output

INSTALLED
Python — IntelliSense (Py — Microsoft
Pylance — A performant, f — Microsoft

POPULAR
Python — IntelliSense (Py — Microsoft
Jupyter — 2953ms — Jupyter notebook support,... — Microsoft
Pylance — 37.1M ★ 3.5 — A performant, feature-rich... — Microsoft — Installing
C/C++ — 35.3M ★ 3.5 — C/C++ IntelliSense, debug... — Microsoft — Install
Jupyter Ke... — 27.1M ★ 3.5 — Jupyter keymaps for noteb

RECOMMENDED 7
Docker — 17.6M ★ 4.5 — Makes it easy to create, m... — Microsoft — Install
Remote - ... — 14.4M ★ 4.5 — Open any folder or reposit... — Microsoft — Install

e powerful tools

n .ipynb
mmand

If you have an existing project, you can also open a folder and/or clone a project from GitHub: clone a Git repository.

Select a Jupyter kernel

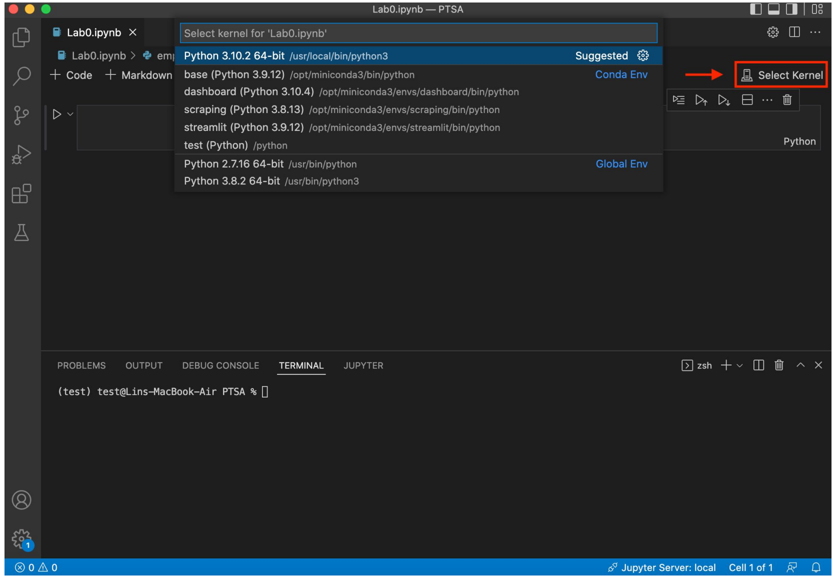
Explore data and debug

>Jupyter
Jupyter: Create New Jupyter Notebook

```
msg = "Welcome to Notebooks in VS Code!"
print(msg)
Welcome to Notebooks in VS Code!
```

Selecting a Python environment

Jupyter notebook kernel



Lab0.ipynb — PTSA

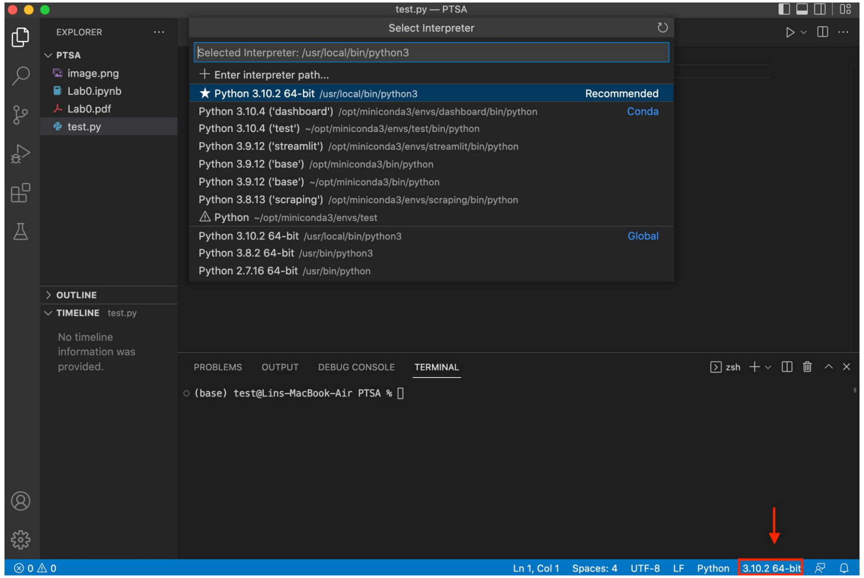
Select kernel for 'Lab0.ipynb'

- Python 3.10.2 64-bit /usr/local/bin/python3 — Suggested
- base (Python 3.9.12) /opt/miniconda3/bin/python — Conda Env
- dashboard (Python 3.10.4) /opt/miniconda3/envs/dashboard/bin/python
- scraping (Python 3.8.13) /opt/miniconda3/envs/scraping/bin/python
- streamlit (Python 3.9.12) /opt/miniconda3/envs/streamlit/bin/python
- test (Python) /python
- Python 2.7.16 64-bit /usr/bin/python — Global Env
- Python 3.8.2 64-bit /usr/bin/python3

Select Kernel

```
(test) test@Lins-MacBook-Air PTSA %
```

Script



test.py — PTSA

Select Interpreter

Selected Interpreter: /usr/local/bin/python3

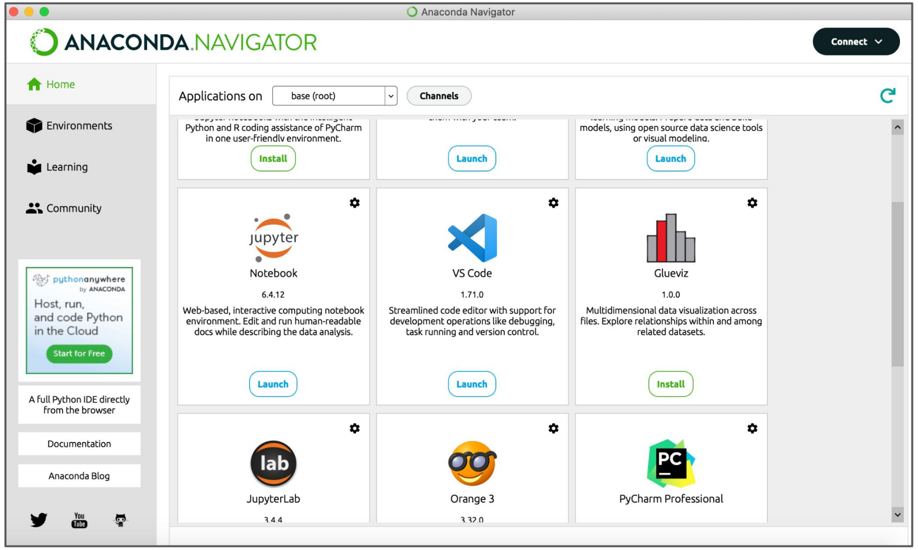
- + Enter interpreter path...
- ★ Python 3.10.2 64-bit /usr/local/bin/python3 — Recommended
- Python 3.10.4 ('dashboard') /opt/miniconda3/envs/dashboard/bin/python — Conda
- Python 3.10.4 ('test') ~/opt/miniconda3/envs/test/bin/python
- Python 3.9.12 ('streamlit') /opt/miniconda3/envs/streamlit/bin/python
- Python 3.9.12 ('base') /opt/miniconda3/bin/python
- Python 3.9.12 ('base') ~/opt/miniconda3/bin/python
- Python 3.8.13 ('scraping') /opt/miniconda3/envs/scraping/bin/python
- ⚠ Python ~/opt/miniconda3/envs/test
- Python 3.10.2 64-bit /usr/local/bin/python3 — Global
- Python 3.8.2 64-bit /usr/bin/python3
- Python 2.7.16 64-bit /usr/bin/python

```
(base) test@Lins-MacBook-Air PTSA %
```

EXPLORER
PTSA
image.png
Lab0.ipynb
Lab0.pdf
test.py
OUTLINE
TIMELINE test.py
No timeline information was provided.
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL
zsh
⊗ 0 ⚠ 0
Ln 1, Col 1 Spaces: 4 UTF-8 LF Python 3.10.2 64-bit

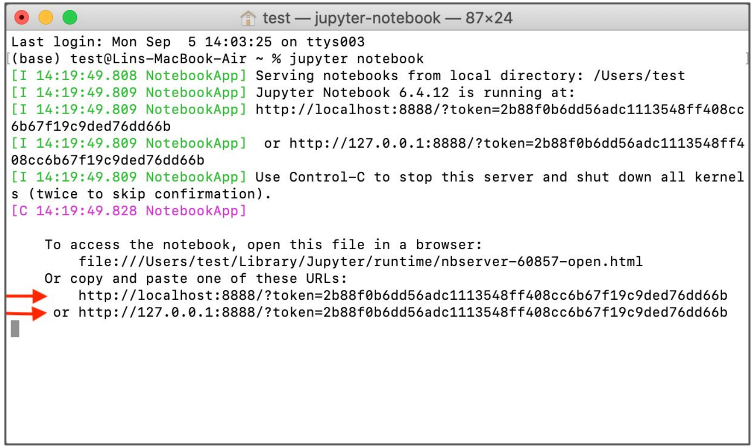
Alternatively : Run Jupyter

From Anaconda Navigator



Anaconda Navigator
ANACONDA.NAVIGATOR
Connect
Home
Environments
Learning
Community
Applications on
base (root)
Channels
Python and R coding assistance of PyCharm in one user-friendly environment.
Install
Launch
models, using open source data science tools or visual modeling.
Launch
Jupyter Notebook
6.4.12
Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.
Launch
VS Code
1.71.0
Streamlined code editor with support for development operations like debugging, task running and version control.
Launch
Glueviz
1.0.0
Multidimensional data visualization across files. Explore relationships within and among related datasets.
Install
pythonanywhere by ANACONDA
Host, run, and code Python in the Cloud
Start for Free
A full Python IDE directly from the browser
Documentation
Anaconda Blog
JupyterLab
3.4.4
Orange 3
3.32.0
PyCharm Professional

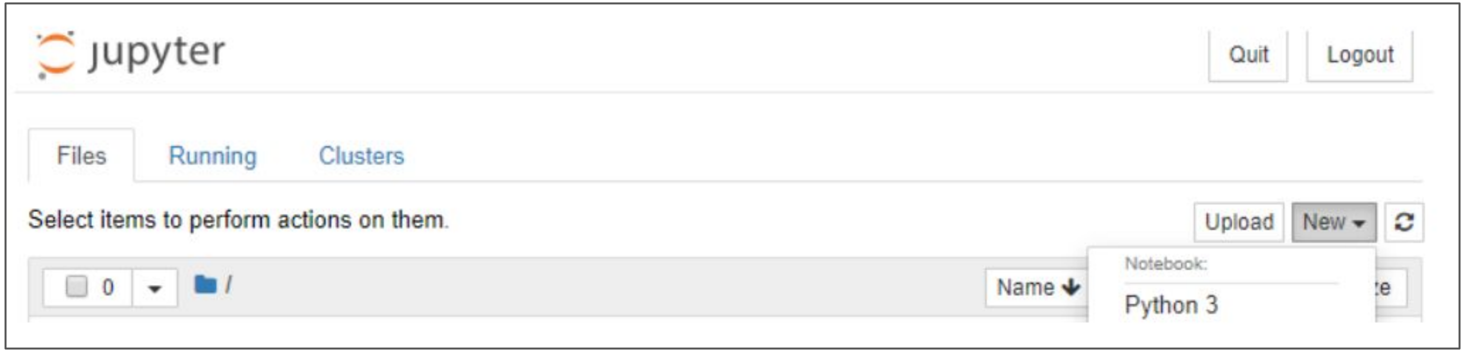
From terminal



```
Last login: Mon Sep  5 14:03:25 on ttys003
(base) test@Lins-MacBook-Air ~ % jupyter notebook
[I 14:19:49.808 NotebookApp] Serving notebooks from local directory: /Users/test
[I 14:19:49.809 NotebookApp] Jupyter Notebook 6.4.12 is running at:
[I 14:19:49.809 NotebookApp] http://localhost:8888/?token=2b88f0b6dd56adc1113548ff408cc
6b67f19c9ded76dd66b
[I 14:19:49.809 NotebookApp]  or http://127.0.0.1:8888/?token=2b88f0b6dd56adc1113548ff4
08cc6b67f19c9ded76dd66b
[I 14:19:49.809 NotebookApp] Use Control-C to stop this server and shut down all kernel
s (twice to skip confirmation).
[C 14:19:49.828 NotebookApp]

    To access the notebook, open this file in a browser:
        file:///Users/test/Library/Jupyter/runtime/nbserver-60857-open.html
    Or copy and paste one of these URLs:
        http://localhost:8888/?token=2b88f0b6dd56adc1113548ff408cc6b67f19c9ded76dd66b
     or http://127.0.0.1:8888/?token=2b88f0b6dd56adc1113548ff408cc6b67f19c9ded76dd66b
```

Alternatively : Select environment



Select items to perform actions on them.

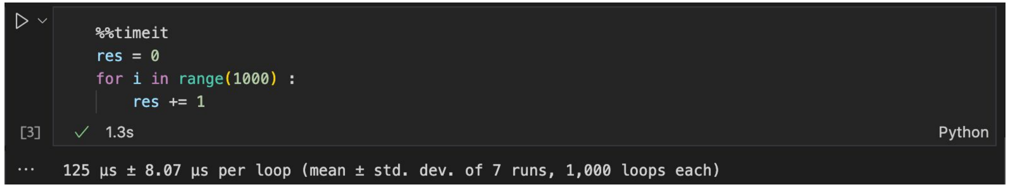
Useful Jupyter commands

% : inline magic

%% : cell magic

→ time : Execution time of a cell

→ timeit : Mean / Std time of a cell



```
%%timeit
res = 0
for i in range(1000) :
    res += 1
```

[3] ✓ 1.3s Python

```
125 µs ± 8.07 µs per loop (mean ± std. dev. of 7 runs, 1,000 loops each)
```

→ %autoreload 2 : Automatically reload imports from extensions / scripts

→ %history [-g] : Display command executed in current / all sessions

! : Run the following expression in a terminal e.g. !pip install pkg

LaTeX installation

Linux : TeXLive (https://www.tug.org/texlive/)

macOS : MacTeX (https://www.tug.org/mactex/)

Windows : MikTeX (https://miktex.org/download)

BasicTeX (lightweight version of MacTeX)

For MacOS users with limited storage, I recommend using BasicTeX

To install a LaTeX package : (sudo) tlmgr install pkg

To look for possible distributions : (sudo) tlmgr search --global --file {filename}.sty

If using LaTeX with a GUI program, you need to compile successively

LaTeX (x2), BibTeX, LaTeX

Markdown and LaTeX

If in path variables, LaTeX should be detected.

Even without LaTeX installed, you should be able to write simple expressions.

Inline expression : $ expression $

Centered expression : $$ expression $$

It is possible to write LaTeX environments directly in markdown, e.g.

```
\begin{equation}
    expression
\end{equation}
```

Converting to PDF

Using ‘nbconvert’

‘nbconvert’ should be installed if ‘jupyter’ and ‘ipykernel’ are installed.

```
jupyter-nbconvert --to pdf {notebook}
```

From Jupyter

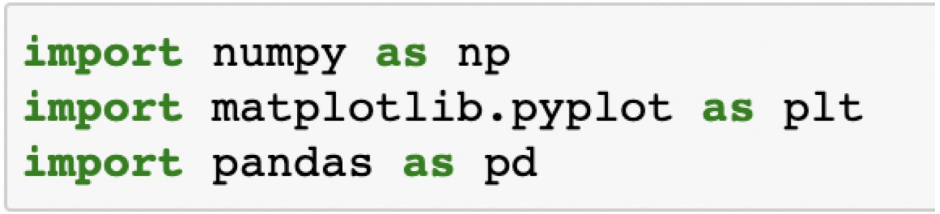
File > Download as > PDF via LaTeX

You might want to double check that the outputs are not truncated.

You can use ‘\’ to write code on multiple lines.

The first notebook

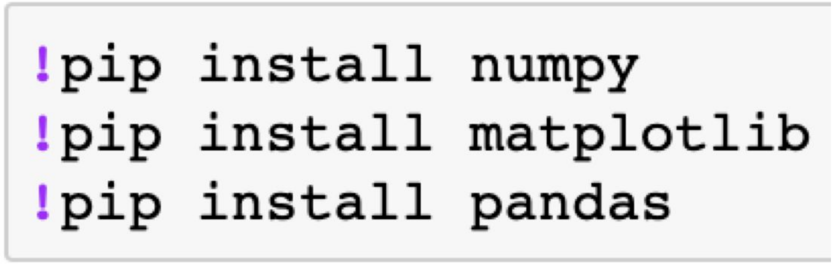
Installing packages



```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

Press `[shift]` + `[return]` to run code block

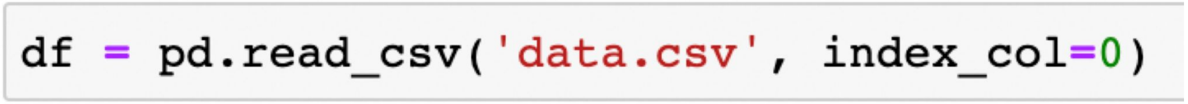
If packages not installed, install using:



```
!pip install numpy
!pip install matplotlib
!pip install pandas
```

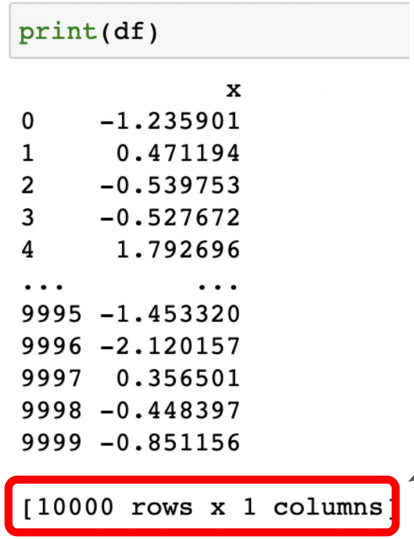
Loading data

Use pandas to load data into “DataFrame”



```
df = pd.read_csv('data.csv', index_col=0)
```

Inspect data using “print”



```
print(df)
```

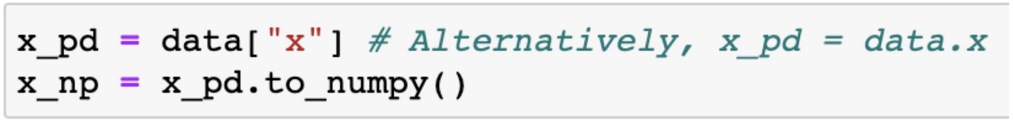
```
             x
0    -1.235901
1     0.471194
2    -0.539753
3    -0.527672
4     1.792696
...        ...
9995 -1.453320
9996 -2.120157
9997  0.356501
9998 -0.448397
9999 -0.851156

[10000 rows x 1 columns]
```

“shape” of DataFrame

Loading data

Extract column of interest (`x` column)



```
x_pd = data["x"] # Alternatively, x_pd = data.x
x_np = x_pd.to_numpy()
```

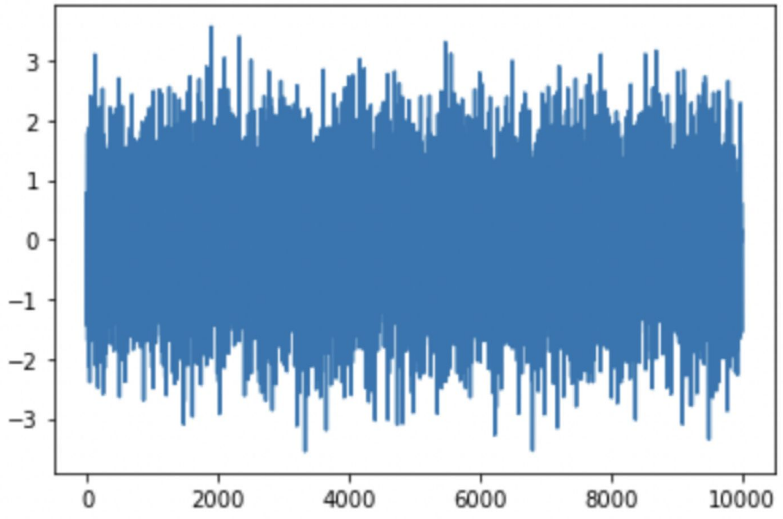
We want to work with `numpy` arrays (fast data structures), so need to convert from `pandas` objects to `numpy` using "`.to_numpy()`"

Plotting

Use `matplotlib.pyplot` to display data ([tutorial]())

```
plt.plot(x)
```

```
[<matplotlib.lines.Line2D at 0x7ff48b482460>]
```

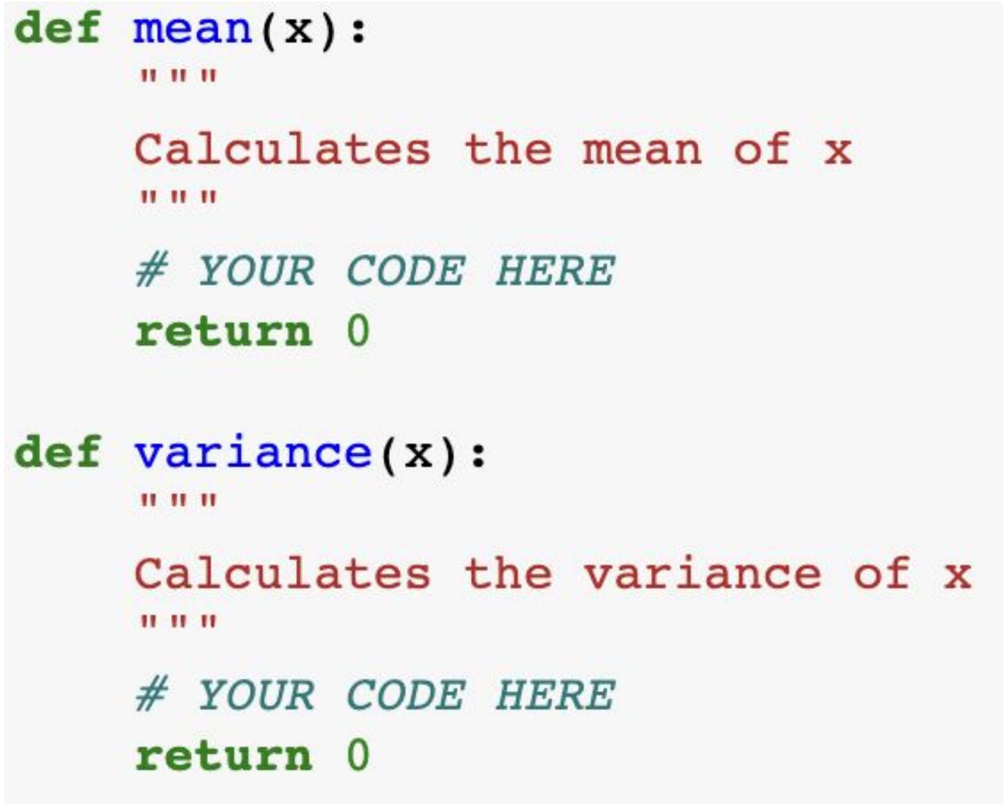


A little competition

The rules:

1. Compute the empirical mean and variance of “`x`” without using `np.mean` or `np.var` (or `np.std` or any other variants on these from different packages)
   - Note, data is <u>**STATIONARY**</u>
2. Use skeleton functions mean and variance in the lab notebook from Brightspace (these validate your functions and calculate execution time)
3. One winner for fastest solution, and one for most creative solution (feel free to use any package or function that DOES NOT directly compute the mean or variance
4. Upload your solutions (notebook, both ipynb and saved as pdf with Latex) to Brightspace

A little competition



```python
def mean(x):
    """
    Calculates the mean of x
    """
    # YOUR CODE HERE
    return 0

def variance(x):
    """
    Calculates the variance of x
    """
    # YOUR CODE HERE
    return 0
```