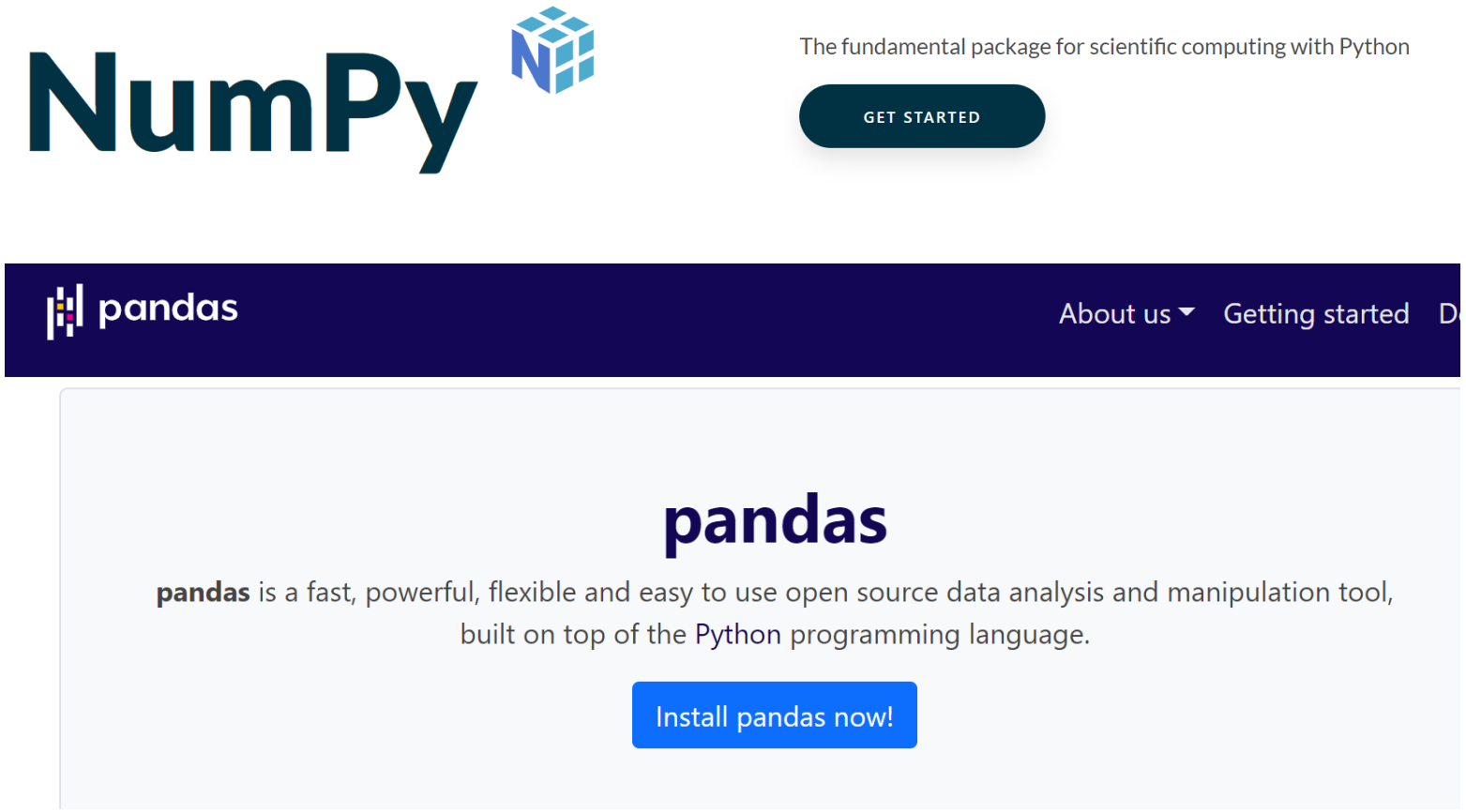
Course Software

- **Python** in general-purpose programming language.
  - Easy to learn
  - Supports multiple programming paradigms
  - Extensible: Growing number of modules and libraries
  - Active open-source community

- **Anaconda** is a packaged compilation of Python along with a core suite of libraries that cover the basics of data science (also has R and other programming libraries).

- **Spyder** is an integrated development environment (IDE) that comes with Anaconda. It's basically a nice front-end for Python (Spyder = Scientific Python Development Environment), giving you a console, a scripting window, a graphics window, and a Python workspace, among other options.

To work with data, you want to be able to work with the python libraries of Numpy and Pandas



Installing libraries

INSTALLING NUMPY

On this page

The only prerequisite for installing NumPy is Python itself. If you don't have Python yet and want the simplest way to get started, we recommend you use the Anaconda Distribution - it includes Python, NumPy, and many other commonly used packages for scientific computing and data science.

NumPy can be installed with `conda`, with `pip`, with a package manager on macOS and Linux, or from source. For more detailed instructions, consult our Python and NumPy installation guide below.

CONDA

If you use `conda`, you can install NumPy from the `defaults` or `conda-forge` channels:

```
# Best practice, use an environment rather than install in the base env
conda create -n my-env
conda activate my-env
# If you want to install from conda-forge
conda config --env --add channels conda-forge
# The actual install command
conda install numpy
```

PIP

If you use `pip`, you can install NumPy with:

```
pip install numpy
```

Python 3.11 (64-bit)

```
Python 3.11.1 (tags/v3.11.1:a7a450f, Dec  6 2022, 19:58:39) [MSC v.1934 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> pip install numpy
  File "<stdin>", line 1
    pip install numpy
        ^^^^^^^
SyntaxError: invalid syntax
>>> help(pip)
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'pip' is not defined. Did you mean: 'zip'?
>>> pip install numpy
  File "<stdin>", line 1
    pip install numpy
        ^^^^^^^
SyntaxError: invalid syntax
>>> pip3 install numpy
  File "<stdin>", line 1
    pip3 install numpy
         ^^^^^^^
SyntaxError: invalid syntax
>>> import nummpy as np
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
ModuleNotFoundError: No module named 'nummpy'
>>> import numpy as np
>>> import pandas as pd
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
ModuleNotFoundError: No module named 'pandas'
>>>
```

Install the library on a prompt or command line for Python to use.

Then you can import it into your Python programming shell or IDE (JupyterLab, Spyder, PyCharm, etc…) to be able to use.

C:\WINDOWS\system32\cmd.exe

```
Microsoft Windows [Version 10.0.22000.1455]
(c) Microsoft Corporation. All rights reserved.

C:\Users\carol>pip install numpy
Collecting numpy
  Downloading numpy-1.24.1-cp311-cp311-win_amd64.whl (14.8 MB)
     ---------------------------------------- 14.8/14.8 MB 59.4 MB/s eta 0:00:00
Installing collected packages: numpy
Successfully installed numpy-1.24.1

C:\Users\carol>pip install pandas
Collecting pandas
  Downloading pandas-1.5.2-cp311-cp311-win_amd64.whl (10.3 MB)
     ---------------------------------------- 10.3/10.3 MB 12.6 MB/s eta 0:00:00
Collecting python-dateutil>=2.8.1
  Downloading python_dateutil-2.8.2-py2.py3-none-any.whl (247 kB)
     ---------------------------------------- 247.7/247.7 kB 14.8 MB/s eta 0:00:00
Collecting pytz>=2020.1
  Downloading pytz-2022.7-py2.py3-none-any.whl (499 kB)
     ---------------------------------------- 499.4/499.4 kB ? eta 0:00:00
Requirement already satisfied: numpy>=1.21.0 in c:\users\carol\appdata\local\programs\p
 (from pandas) (1.24.1)
Collecting six>=1.5
  Downloading six-1.16.0-py2.py3-none-any.whl (11 kB)
Installing collected packages: pytz, six, python-dateutil, pandas
Successfully installed pandas-1.5.2 python-dateutil-2.8.2 pytz-2022.7 six-1.16.0
```

To import a DataFrame from the local directory:

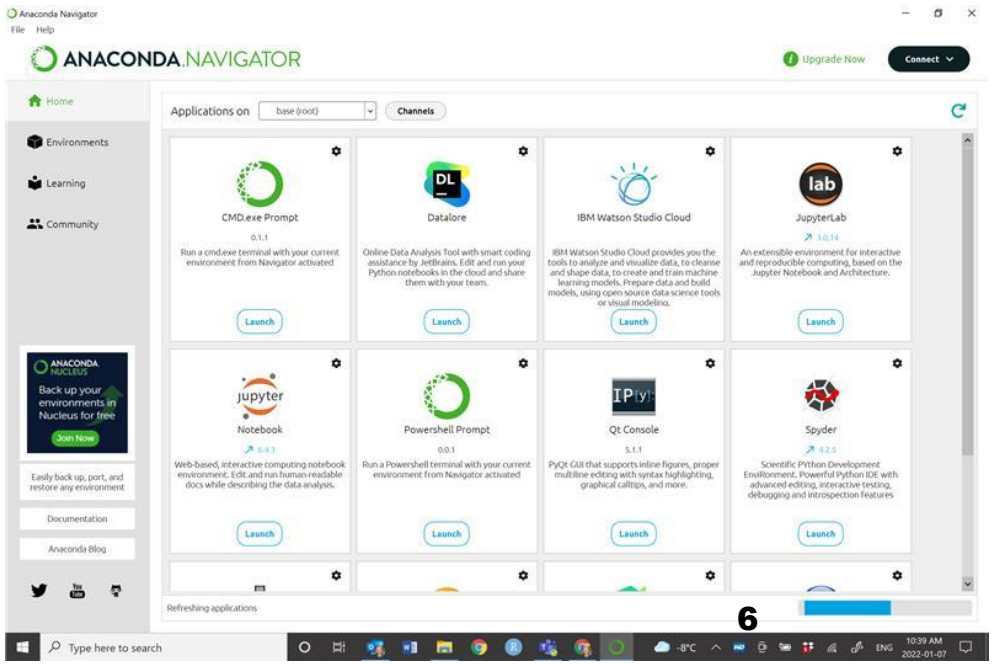
```
>>> import os
>>> os.getcwd()
'C:\\Users\\carol\\AppData\\Local\\Programs\\Python\\Python311'
>>> all_boards = pd.read_csv('six-point-board-thickness.csv')
>>> print(all_boards)
             Date.Time  Pos1  Pos2  Pos3  Pos4  Pos5  Pos6
0      2010-02-18 3:04  1761  1739  1758  1677  1684  1692
1      2010-02-18 3:37  1801  1688  1753  1741  1692  1675
2      2010-02-18 3:37  1697  1682  1663  1671  1685  1651
3      2010-02-18 3:37  1679  1712  1672  1703  1683  1674
4      2010-02-18 3:37  1699  1688  1699  1678  1688  1705
...                ...   ...   ...   ...   ...   ...   ...
4995  2010-02-18 13:15  1690  1701  1690  1694  1735  1695
4996  2010-02-18 13:15  1703  1674  1666  1694  1659  1728
4997  2010-02-18 13:16  1657  1667  1675  1654  1648  1609
4998  2010-02-18 13:16  1746  1717  1638  1723  1703  1706
4999  2010-02-18 13:16  1668  1680  1668  1669  1651  1629

[5000 rows x 7 columns]
>>>
```

Course Software

- The Anaconda distribution also comes with other useful utilities, such as the package management tool Conda.

- Another utility is **Juypter Notebooks and JuypterLab:** JupyterLab is the latest web-based interactive development environment for notebooks, code, and data**.**



Lecture2_MeanExampleA.ipy | Tutorial1KEYReadAndExplore | Lecture3_5NumberSummary

Markdown | Python 3 (ipykernel)

1 Read the "six-point-board-thickness.csv" file into Python. You will need to import some packages to use. Take a look through the dataframe to review.

[98]:
```
import numpy as np
import pandas as pd
```

[99]:
```
all_boards = pd.read_csv('six-point-board-thickness.csv')
print(all_boards)
```

```
            Date.Time  Pos1  Pos2  Pos3  Pos4  Pos5  Pos6
0      2010-02-18 3:04  1761  1739  1758  1677  1684  1692
1      2010-02-18 3:37  1801  1688  1753  1741  1692  1675
2      2010-02-18 3:37  1697  1682  1663  1671  1685  1651
3      2010-02-18 3:37  1679  1712  1672  1703  1683  1674
4      2010-02-18 3:37  1699  1688  1699  1678  1688  1705
...                ...   ...   ...   ...   ...   ...   ...
4995  2010-02-18 13:15  1690  1701  1690  1694  1735  1695
4996  2010-02-18 13:15  1703  1674  1666  1694  1659  1728
4997  2010-02-18 13:16  1657  1667  1675  1654  1648  1609
4998  2010-02-18 13:16  1746  1717  1638  1723  1703  1706
4999  2010-02-18 13:16  1668  1680  1668  1669  1651  1629
```