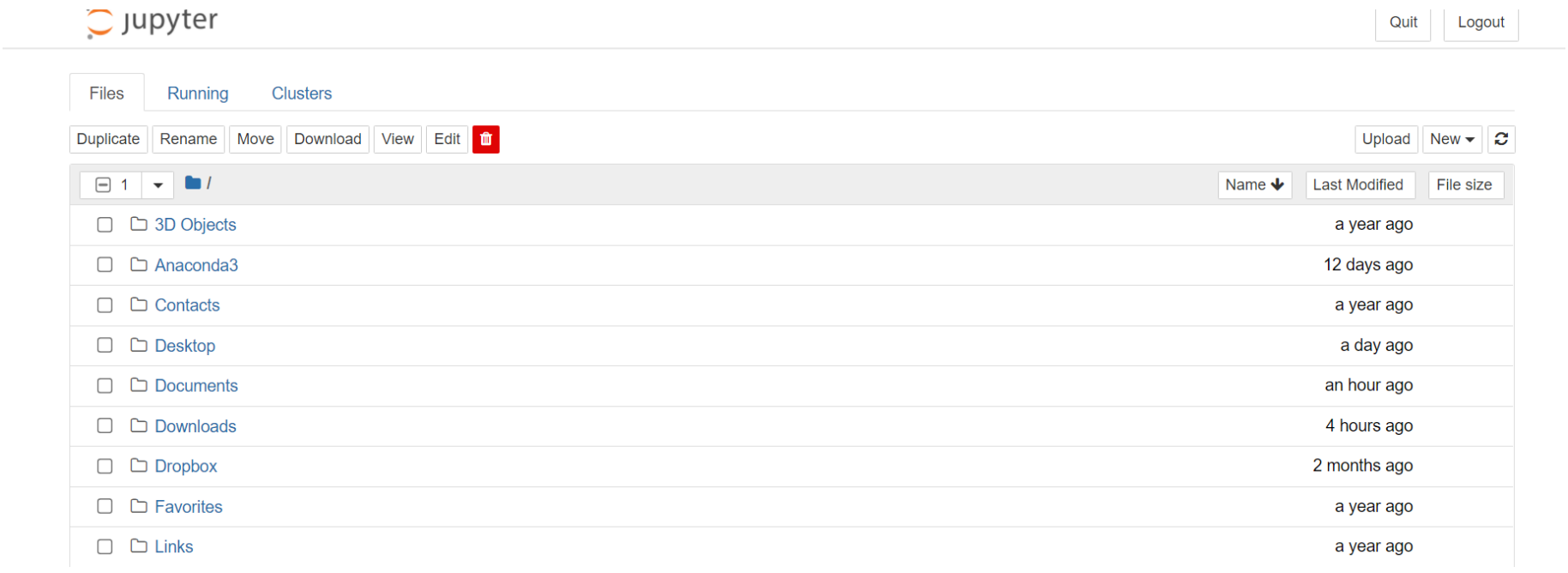
Some Python Basics

Mon Sept 13

Running Python from JuPyTeR

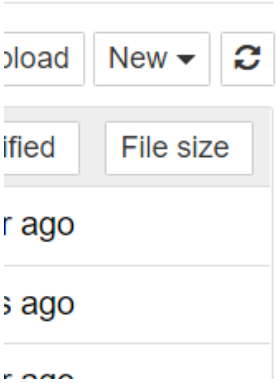
- You've downloaded JuPyTer
- The 'Py' in JuPyTer stands for Python
- Click on JuPyTeR icon to start the program
- The JuPyTer web icon looks like this:

After a few other screens gets you to this



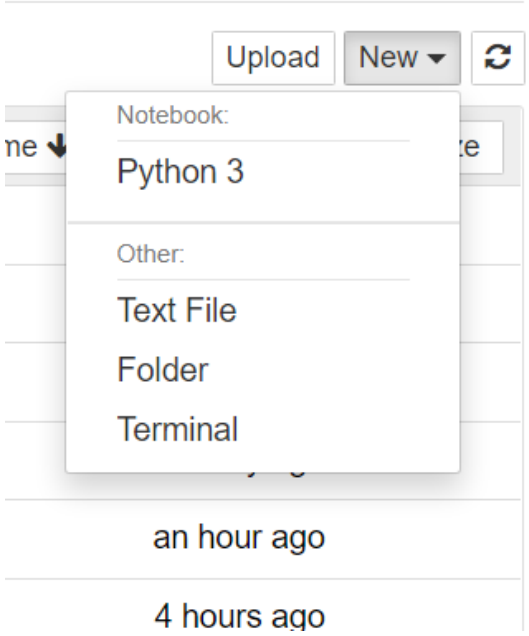
Now you need to start a Python Session

- Click on the “New” button in the upper right of the screen as shown here:

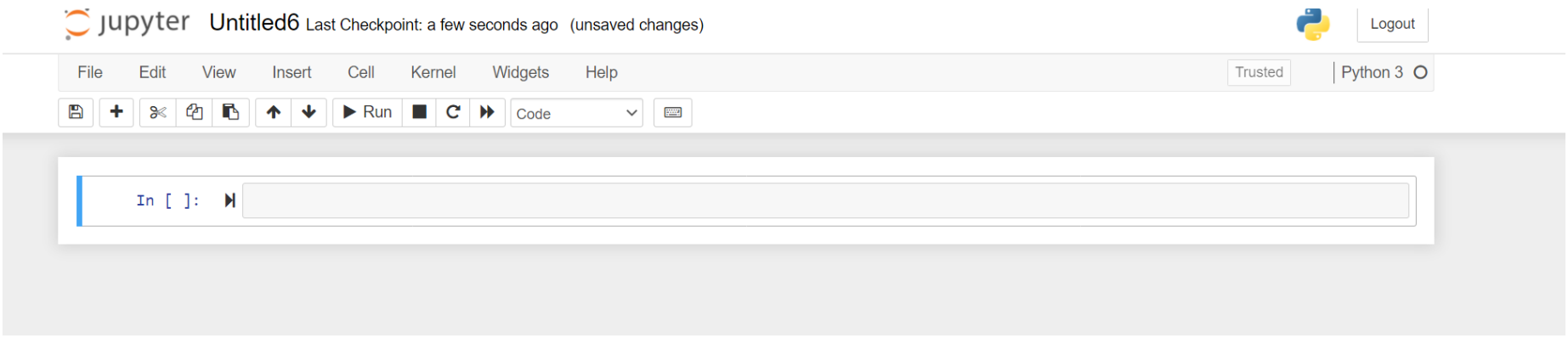


Clicking on “New”

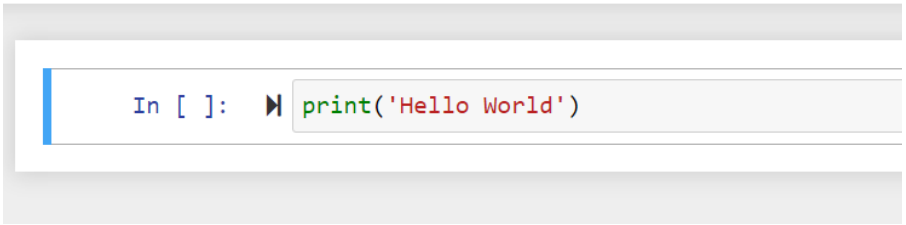
- When you click on “New” it will a menu shows up which includes Python 3
- Go to the Python 3

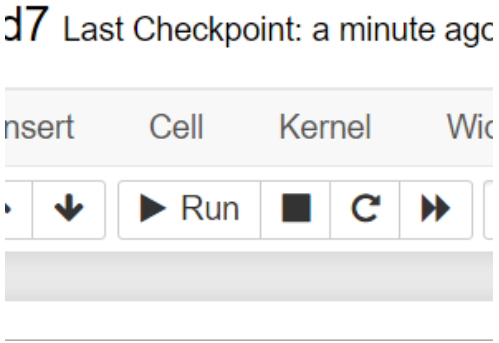


This gets you to a Python Screen

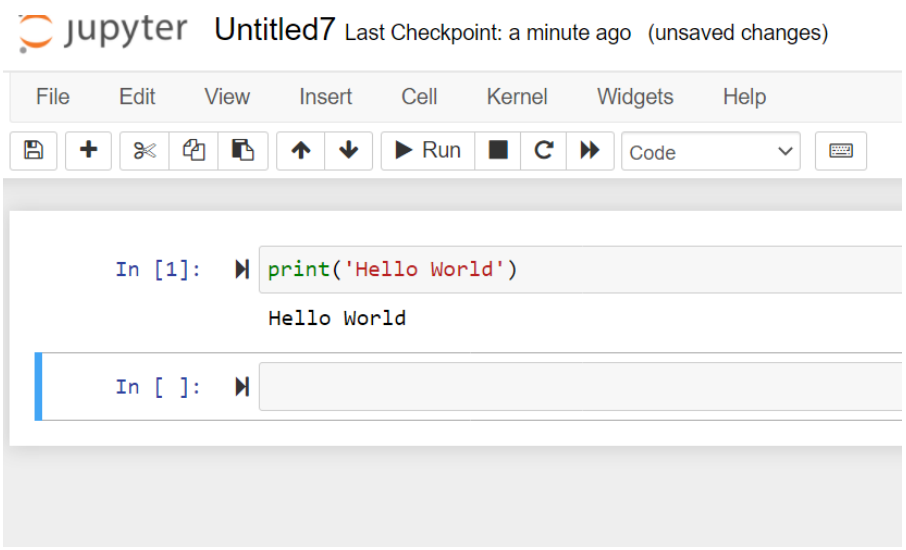


Simplest Program: Enter then hit the “Run” key





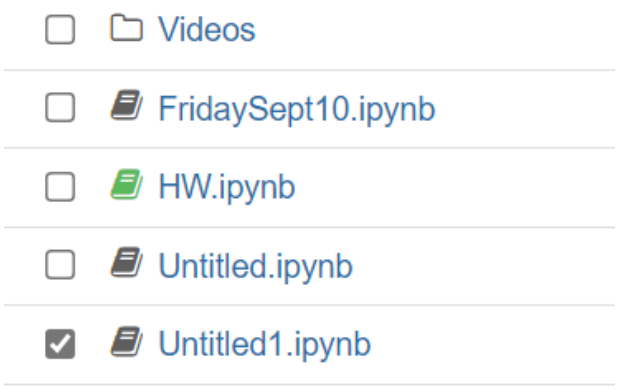
Program runs; then we save it



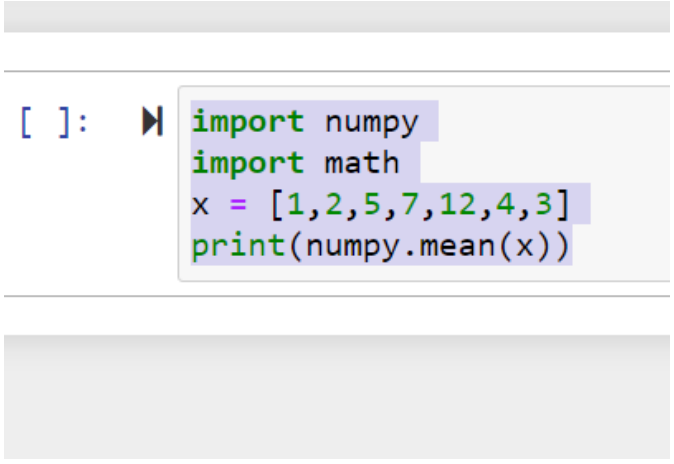
- Then go to File
- Save as
- HW
- It will be saved as the HW notebook

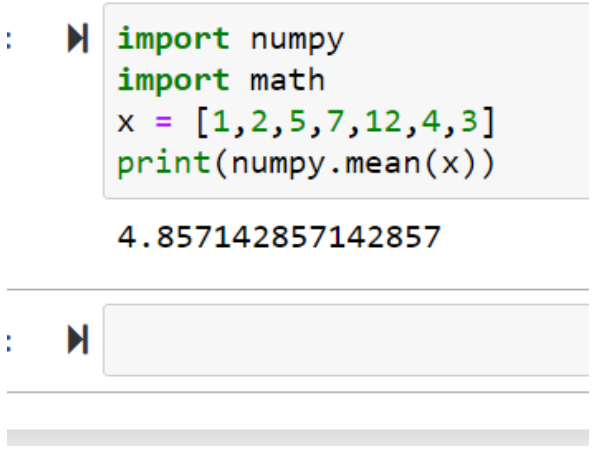
Go back to the Home Page JuPyTeR page

- It shows up as HW.ipynb

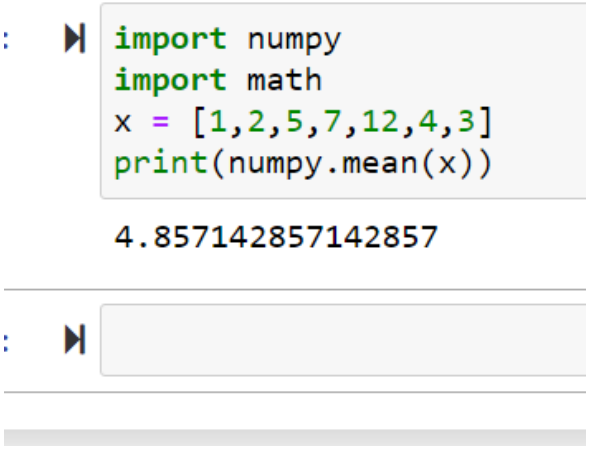


More interesting program; Run to get:





Python Libraries



```
import numpy
import math
x = [1,2,5,7,12,4,3]
print(numpy.mean(x))
```

```
4.857142857142857
```

- What is this doing?
- Import Libraries
- ‘Numpy’ is the Python library (written in C or C++) that implements “linear algebra” operations on vectors and matrices
- ‘Math’ is the Python library that does, well,….

Math in action

```
n [ ]:
import math
print(math.pi)
print(math.e)
print(math.exp(1))
print(math.sin(math.pi))
```

```
import math
print(math.pi)
print(math.e)
print(math.exp(1))
print(math.sin(math.pi))
```

```
3.141592653589793
2.718281828459045
2.718281828459045
1.2246467991473532e-16
```

Common error and saving characters

```
import math
print(sin(pi))
```

```
---------------------------------------------
NameError                                 Tra
<ipython-input-4-53a5d637372f> in <module>
      1 import math
----> 2 print(sin(pi))

NameError: name 'sin' is not defined
```

```
import math
print(pi)
```

```
---------------------------------------------
NameError                                 Tra
<ipython-input-5-6945cc70914a> in <module>
      1 import math
----> 2 print(pi)

NameError: name 'pi' is not defined
```

```
import math as m
print(m.pi)
print(m.e)
print(m.exp(1))
print(m.sin(math.pi))
```

```
3.141592653589793
2.718281828459045
2.718281828459045
1.2246467991473532e-16
```

Downloading Data from a flat '.csv' file

```
import pandas as pd
data = pd.read_csv('C:/Users/mdavison/Desktop/DS1000/textbook_datasets_csv/chapter_01/eg01-02majors.csv')
print(data)
```

```
       Field of Study  Percent
0      ArtsHumanities      8.8
1  BiologicalSciences     15.5
2            Business     13.8
3           Education      4.4
4         Engineering     11.5
5          HealthProf     11.7
6             MathComp      6.2
7    PhysicalSciences      2.7
8     Social Sciences     11.0
9               Other     13.1
```

Important fact

data = pd.read_csv('C:/Users/mdavison/Desktop/DS1000/textbook_datasets_csv/chapter_01/eg01-02majors.csv')

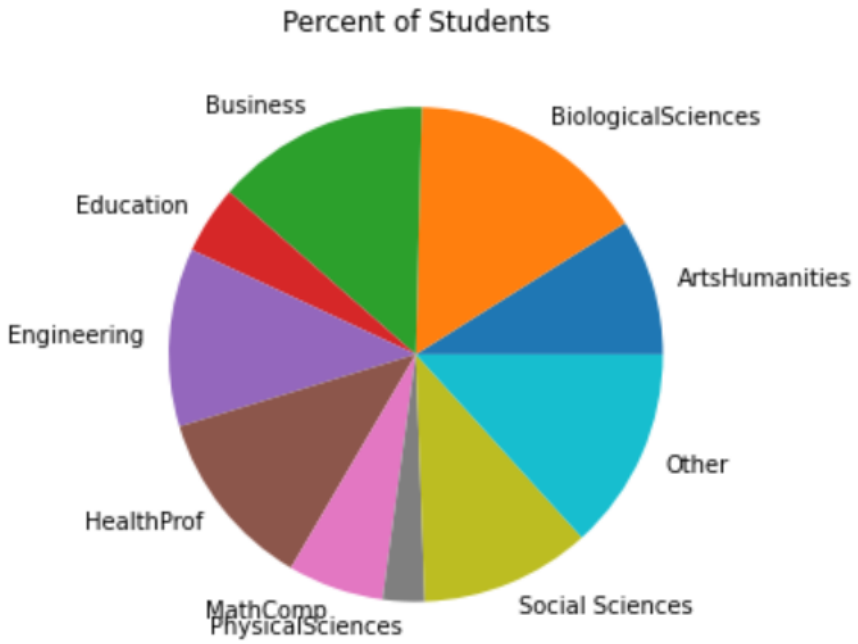
Only works on my machine. You will have to modify that to show where the same data is stored on your own machine.

What is pandas?

- It is a python package that deals with “panel data”
- It allows us to upload 3 kinds of data:
- 1 dimensional (‘Series’) (think stock prices over time(.
- 2 dimensional (tables; called ‘frames’ here) (think a table of data like the one we just saw, but maybe with > 2 columns)
- 3 dimensional (‘panels’) think measurements of height and weight for many babies taken over many times).

Making Pie Chart (same data as earlier table)

```
import matplotlib.pyplot as plt
fig = df.plot.pie(y='Percent',figsize=(5,5),labels=df['Field of Study'],
                  legend = False,title="Percent of Students",ylabel='')
plt.show(fig)
```



Downloading data from Web

- Begin by installing python datareader
- Command is pip install pandas_datareader
- Now suppose we want to download stock prices from Yahoo Finance (free and you don’t need to register).

Code to download Apple Stock Prices

```
import pandas_datareader.data as web
import pandas as pd
import datetime as dt
import matplotlib.pyplot as plt
plt.style.use('ggplot')
#start of the data
start = dt.datetime(2010,1,1)

#last data point to download
end =dt.datetime(2020,1,1)

# name of the stock symbol
symbol = 'AAPL' ###using Apple as an example

##source of the data
source = 'yahoo'

# pass in the arguments above to pandas datareader
data = web.DataReader(symbol, source, start, end)

##view first 5 rows of the data
data.head()
```

Output

Out[1]:

Date	High	Low	Open	Close	Volume	Adj Close
2010-01-04	7.660714	7.585000	7.622500	7.643214	493729600.0	6.562591
2010-01-05	7.699643	7.616071	7.664286	7.656429	601904800.0	6.573935
2010-01-06	7.686786	7.526786	7.656429	7.534643	552160000.0	6.469369
2010-01-07	7.571429	7.466071	7.562500	7.520714	477131200.0	6.457407
2010-01-08	7.571429	7.466429	7.510714	7.570714	447610800.0	6.500339