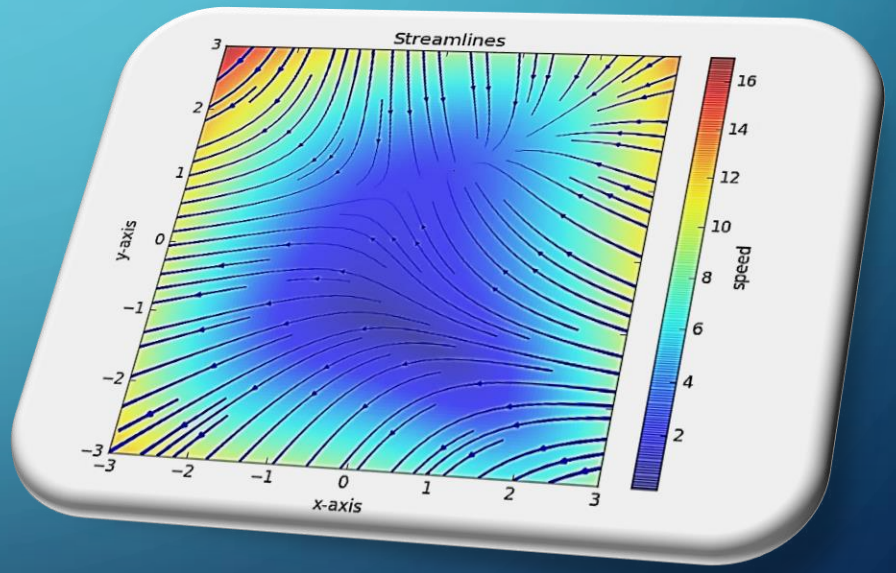
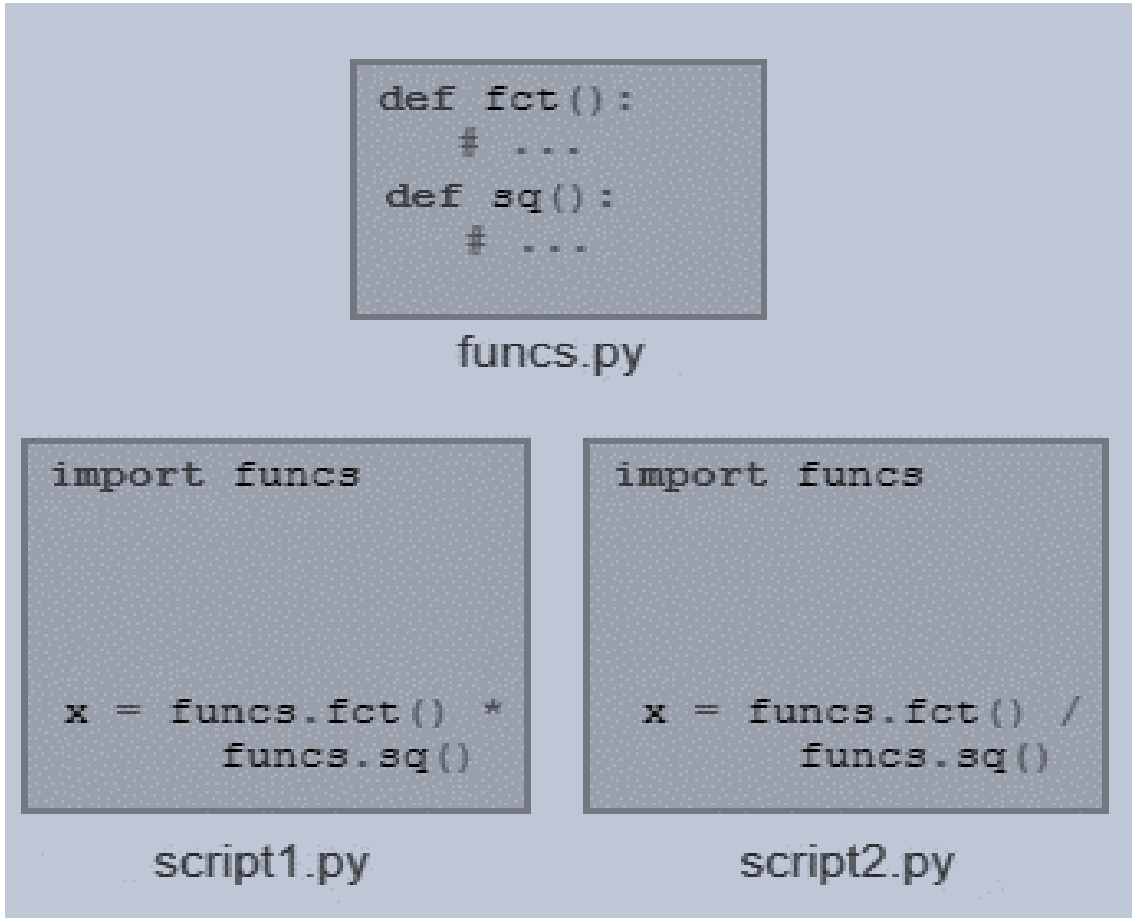
MODULES

MATPLOTLIB, PANDAS AND NUMPY



Module Introduction

- A module is a file containing Python statements and definitions that can be used by other Python sources.
- A programmer can then import the module and use its functions.



Python Interpreter

- When importing a module, the interpreter checks for the availability of the module.
- The interpreter first checks for a matching **built-in** module. Python comes with many built-in modules like Tkinter and Turtle.
- If the module is found in the Python path, it is imported.

Importing Specific Names

- A Python standard library module that contains a number of algorithms.
- Some of the functions in different modules have the same names.
- To select which function to use, use the ‘from’ import keyword:

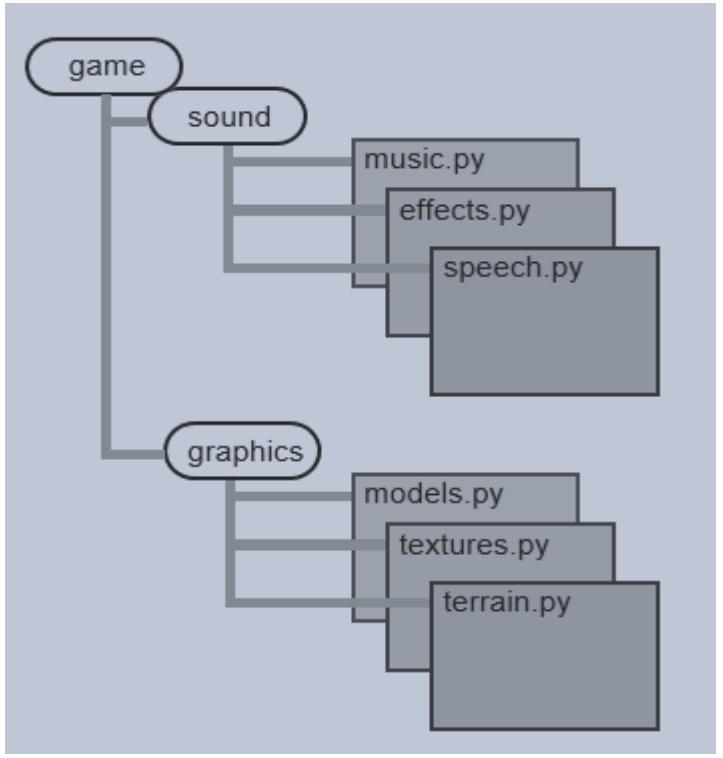
```
from tkinter import messagebox
```

Packages

- An entire directory of modules can be imported all at once.
- A ***package*** is a directory that, when imported, gives access to all of the modules stored in the directory.

```
import sound

sound.music.play_sound()
sound.effects.play_whoosh()
# ...
```

```
import game

game.sound.speech.talk()
game.graphics.terrain.draw_gnd()
# ...
```



Standard Modules

Module name	Description
datetime	Creation and editing of dates and times objects
random	Functions for working with random numbers
copy	Create complete copies of objects
time	Get the current time, convert time zones, sleep for a number of seconds
math	Mathematical functions
os	Operating system informational and management helpers
sys	System specific environment or configuration helpers
pdb	The Python interactive debugger
urllib	URL handling functions, such as requesting web pages

Installing Modules

pip is the package installer for Python. You can use pip to install packages from the Python Package Index and other indexes.

python -m pip install pip

1 python -m pip install numpy

2 python -m pip install scipy

3 python -m pip install matplotlib

Numpy

NumPy is the fundamental package for scientific computing with Python. It contains among other things:

- a powerful N-dimensional array object
- sophisticated (broadcasting) functions
- tools for integrating C/C++ and Fortran code
- useful linear algebra, Fourier transform, and random number capabilities

Scipy

SciPy (pronounced “Sigh Pie”) is a Python-based ecosystem of open-source software for mathematics, science, and engineering. In particular, these are some of the core packages:

NumPy:	Base N-dimensional array package
SciPy library:	Fundamental library for scientific computing
Matplotlib:	Comprehensive 2D Plotting
Ipython:	Enhanced Interactive Console
Sympy:	Symbolic mathematics
Pandas:	Data structures & analysis

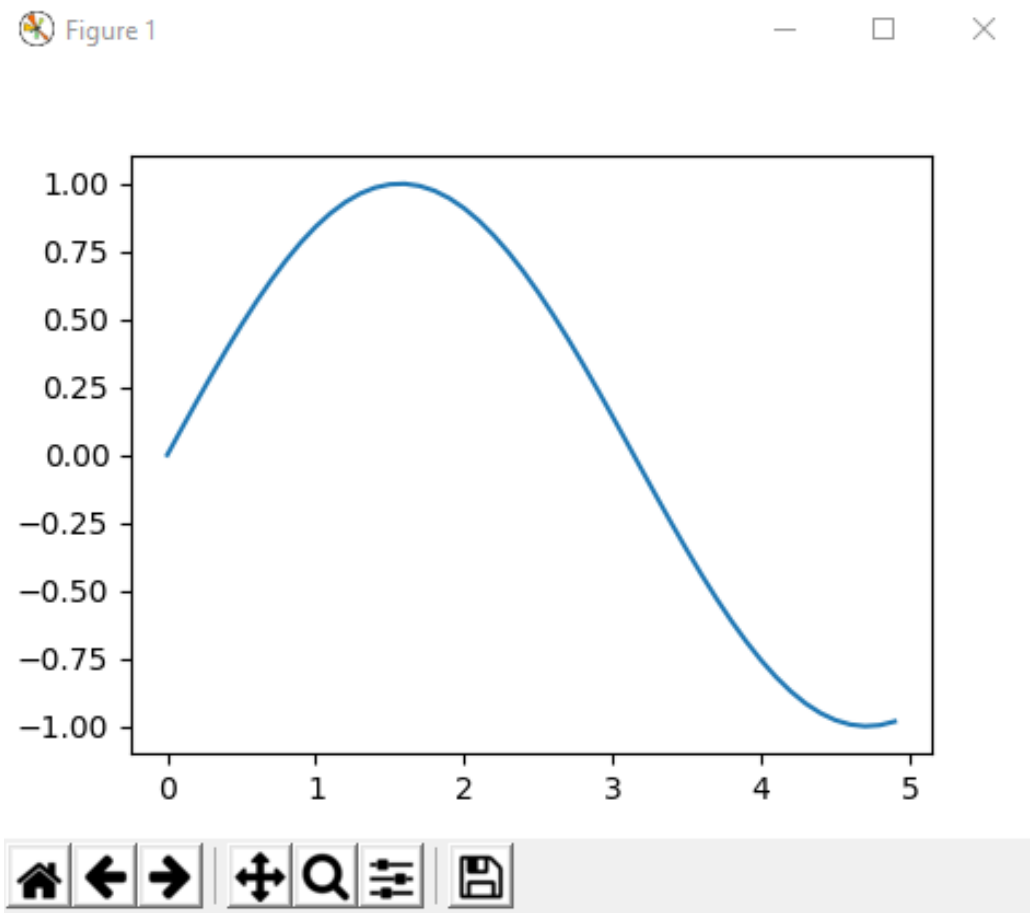
Matplotlib

- Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms.

- Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, error charts, scatterplots, etc., with just a few lines of code.

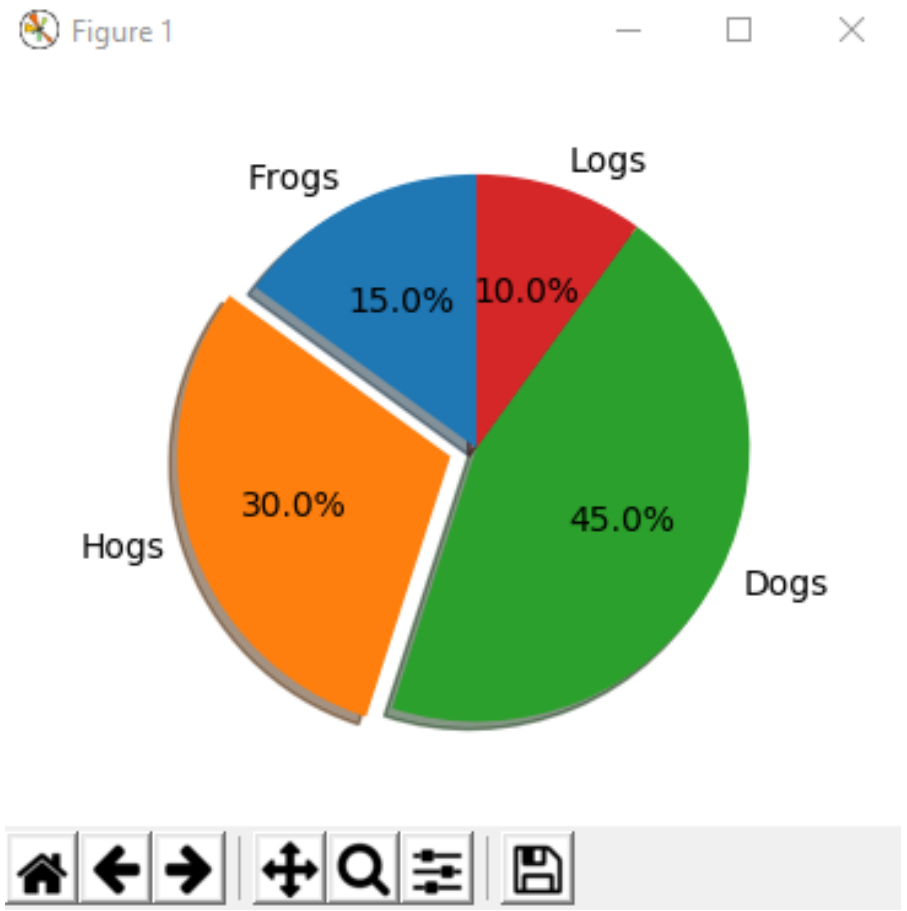
Example 1

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


x = np.arange(0, 5, 0.1)
y = np.sin(x)
plt.plot(x, y)
plt.show()
```



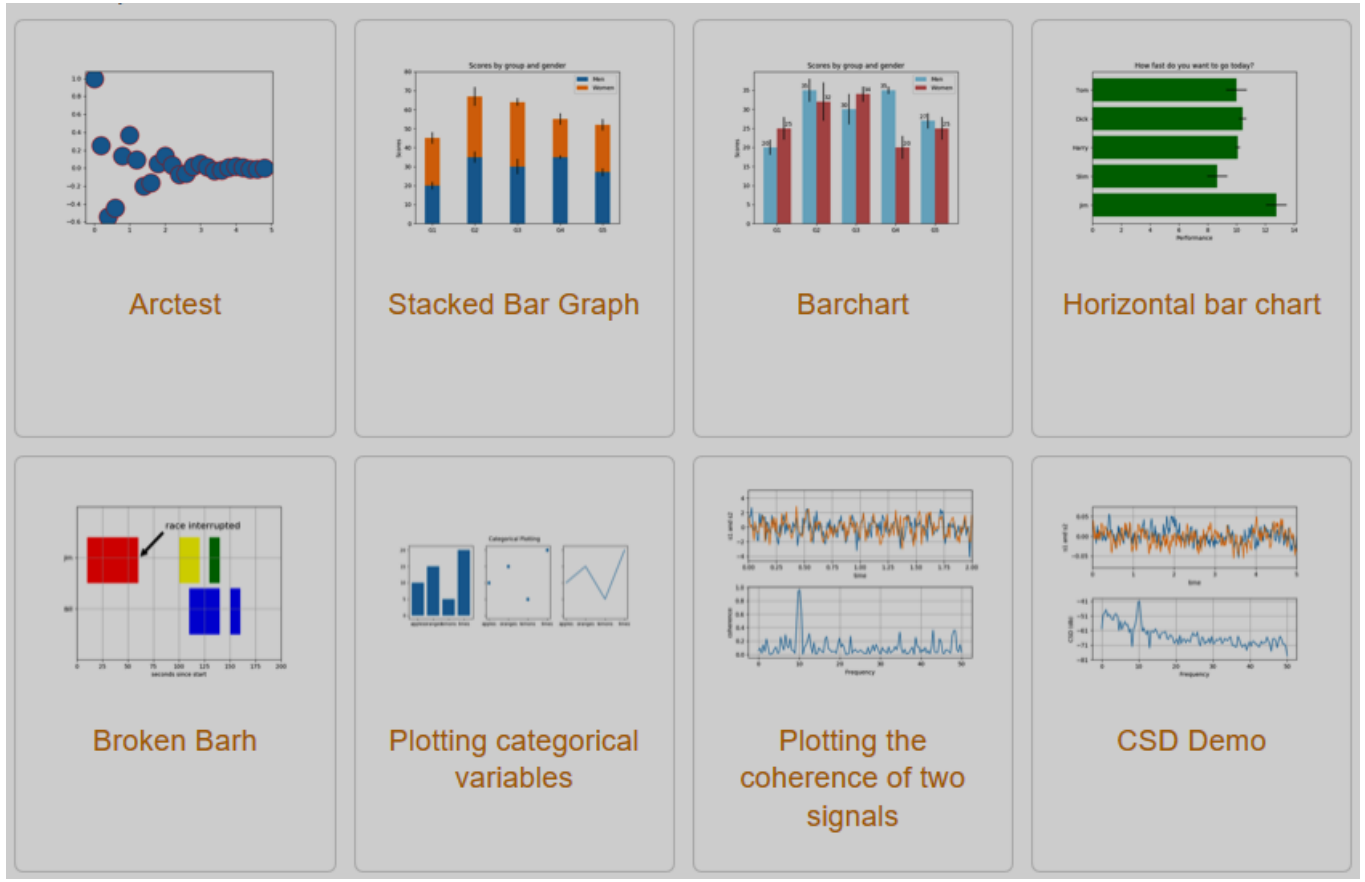
Example

```
import matplotlib.pyplot as plt


labels = 'Frogs', 'Hogs', 'Dogs', 'Logs'
sizes = [15, 30, 45, 10]
explode = (0, 0.1, 0, 0)
fig1, ax1 = plt.subplots()
ax1.pie(sizes, explode=explode, labels=labels,
autopct='%1.1f%%', shadow=True,startangle=90)
ax1.axis('equal')


plt.show()
```



Matplotlib Gallery (PY examples)

https://matplotlib.org/gallery/index.html



NumPy

- NumPy is a python library used for working with arrays.
- It also has functions for working in domain of linear algebra, fourier transform, and matrices.
- NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely.
- NumPy stands for Numerical Python.

NP Arrays

- 1D

```
import numpy as np

arr = np.array([1, 2, 3, 4, 5])
```

- 2D

```
import numpy as np

arr = np.array([[1, 2, 3], [4, 5, 6]])
```

NP Array Dimensions

```
import numpy as np

a = np.array(42)
b = np.array([1, 2, 3, 4, 5])
c = np.array([[1, 2, 3], [4, 5, 6]])
d = np.array([[[1, 2, 3], [4, 5, 6]], [[1, 2, 3], [4, 5, 6]]])

print(a.ndim)       0
print(b.ndim)       1
print(c.ndim)       2
print(d.ndim)       3
```

NumPy Tutorial

https://www.w3schools.com/python/numpy_intro.asp

Pandas Introduction

- Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures.
- The name Pandas is derived from the word Panel Data – an Econometrics from Multidimensional data.

Pandas Data Structures

Pandas deals with the following three data structures –

- Series: 1D labeled homogeneous array, size immutable.
- DataFrame: General 2D labeled, size-mutable tabular structure with potentially heterogeneously typed columns.
- Panel: General 3D labeled, size-mutable array.

Series

- Series is a one-dimensional array like structure with homogeneous data. For example, the following series is a collection of integers 10, 23, 56, …

- Key Points

  Homogeneous data

  Size Immutable

  Values of Data Mutable

pandas.Series(data, index, dtype, copy)

Sr.No	Parameter & Description
1	data data takes various forms like ndarray, list, constants
2	index Index values must be unique and hashable, same length as data. Default np.arrange(n) if no index is passed.
3	dtype dtype is for data type. If None, data type will be inferred
4	copy Copy data. Default False

Series Example

```
#import the pandas library and aliasing as pd
import pandas as pd
import numpy as np
data = np.array(['a','b','c','d’])
s = pd.Series(data)
print(s)
```

```
0 a
1 b
2 c
3 d
```

DataFrame

- DataFrame is a two-dimensional array with heterogeneous data.

Column	Type
Name	String
Age	Integer
Gender	String
Rating	Float

- Key Points
  - Heterogeneous data
  - Size Mutable
  - Data Mutable

pandas.DataFrame(data, index, columns, dtype, copy)

Sr.No	Parameter & Description
1	data data takes various forms like ndarray, series, map, lists, dict, constants and also another DataFrame.
2	index For the row labels, the Index to be used for the resulting frame is Optional Default np.arange(n) if no index is passed.
3	columns For column labels, the optional default syntax is - np.arange(n). This is only true if no index is passed.
4	dtype Data type of each column.
5	copy This command (or whatever it is) is used for copying of data, if the default is False.

DataFrame

```
import pandas as pd
data = [['Alex',10],['Bob',12],['Clarke',13]]
df = pd.DataFrame(data,columns=['Name','Age’])
print(df)
```

```
   Name  Age
0  Alex   10
1  Bob    12
2  Clarke 13
```

Panel (deprecated)

- Panel is a three-dimensional data structure with heterogeneous data. It is hard to represent the panel in graphical representation.
- But a panel can be illustrated as a container of DataFrames.
- Key Points
  - Heterogeneous data
  - Size Mutable
  - Data Mutable

Pandas Tutorial

- https://www.tutorialspoint.com/python_pandas/index.htm