



AI in Finance Big Data Analytics

Foundations of AI in Finance Big Data Analytics with Python

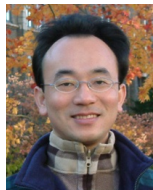
1081AIFBDA06

TLVXM2A (M2449) (8497) (Fall 2019)

(MBA, DBETKU) (3 Credits, Required) [Full English Course]

(Master's Program in Digital Business and Economics)

Tue, 2, 3, 4, (9:10-12:00) (B1012)



Min-Yuh Day, Ph.D.

Associate Professor

Department of Information Management

Tamkang University

<http://mail.tku.edu.tw/myday>

2019-10-22



Course Schedule (1/2)



Tamkang
University

Week	Date	Subject/Topics
1	2019/09/10	Course Orientation on AI in Finance Big Data Analytics
2	2019/09/17	AI in FinTech: Financial Services Innovation and Application
3	2019/09/24	ABC: AI, Big Data, Cloud Computing
4	2019/10/01	Business Models of Fintech
5	2019/10/08	Event Studies in Finance
6	2019/10/15	Case Study on AI in Finance Big Data Analytics I
7	2019/10/22	Foundations of AI in Finance Big Data Analytics with Python
8	2019/10/29	Case Study on Financial Industry Practice I
9	2019/11/05	Quantitative Investing with Pandas in Python

Course Schedule (2/2)



Tamkang
University

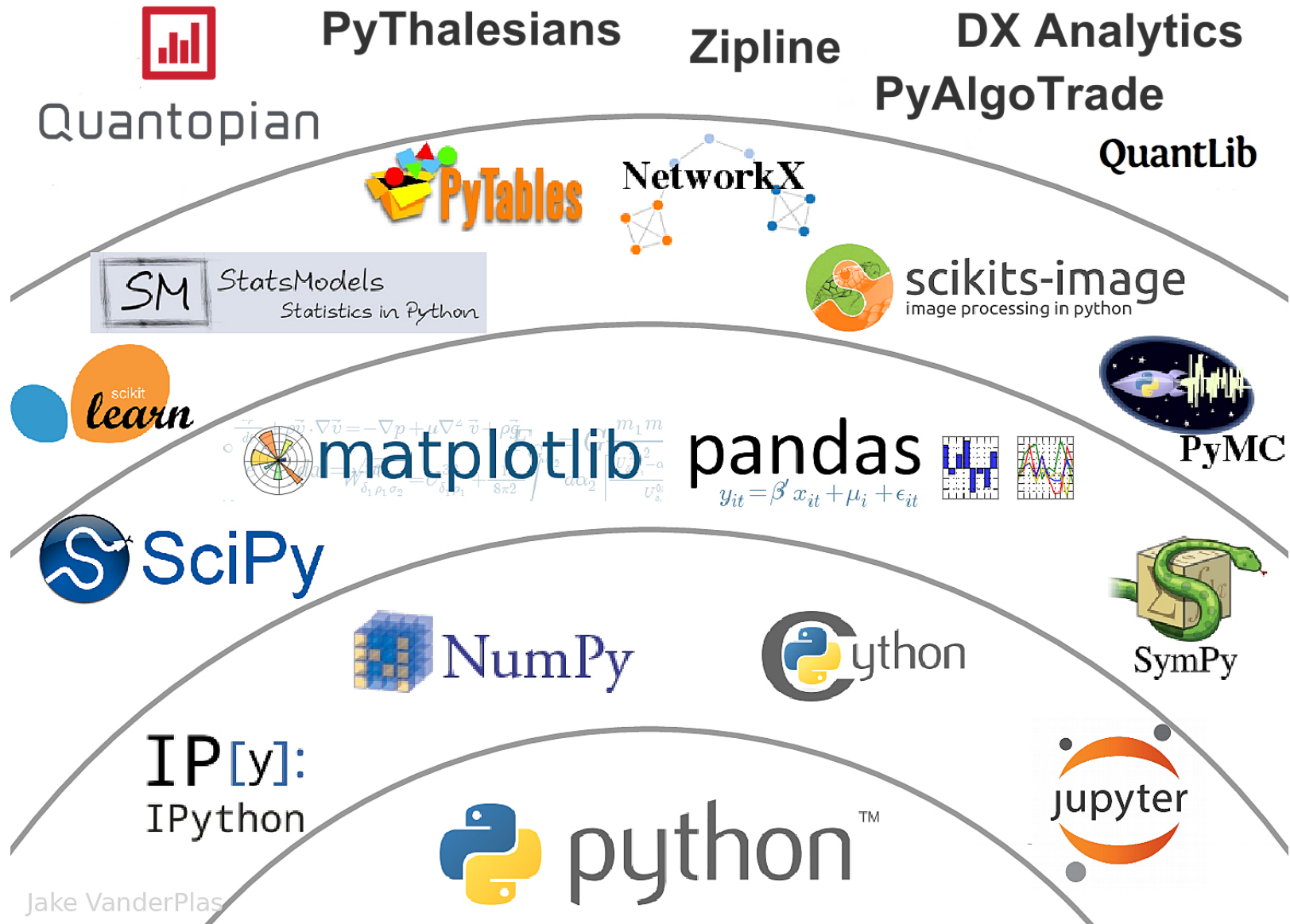
Week	Date	Subject/Topics
10	2019/11/12	Midterm Project Report
11	2019/11/19	Machine Learning in Finance Application with Scikit-Learn In Python
12	2019/11/26	Deep Learning for Financial Time Series Forecasting with TensorFlow I
13	2019/12/03	Case Study on AI in Finance Big Data Analytics II
14	2019/12/10	Deep Learning for Financial Time Series Forecasting with TensorFlow II
15	2019/12/17	Case Study on Financial Industry Practice II
16	2019/12/24	Deep Learning for Financial Time Series Forecasting with TensorFlow III
17	2019/12/31	Final Project Presentation I
18	2020/01/07	Final Project Presentation II

Foundations of AI in Finance Big Data Analytics with Python

Outline

- **Foundations of AI in Finance Big Data Analytics with Python**
 - **Python**
 - Programming language
 - **Numpy**
 - Scientific computing
 - **Pandas**
 - Data structures and data analysis tools

The Quant Finance PyData Stack



AAPL





Python

Python is an
interpreted,
object-oriented,
high-level
programming language
with
dynamic semantics.

Google Colab

The screenshot shows the Google Colaboratory web interface. At the top, the browser address bar displays the URL <https://colab.research.google.com/notebooks/welcome.ipynb>. The Colab logo and name are in the top left, and a 'SHARE' button with a user profile icon is in the top right. A menu bar includes 'File', 'Edit', 'View', 'Insert', 'Runtime', 'Tools', and 'Help'. Below this is a toolbar with icons for 'CODE', 'TEXT', 'CELL' (up and down arrows), and 'COPY TO DRIVE'. On the left side, a 'Table of contents' sidebar is open, listing sections like 'Getting Started', 'Highlighted Features', 'TensorFlow execution', 'GitHub', 'Visualization', 'Forms', 'Examples', and 'Local runtime support'. The main content area features a 'Welcome to Colaboratory!' message with the Colab logo and a brief description. Below this is a 'Getting Started' section with a list of links. Further down, there are expandable sections for 'Highlighted Features' and 'TensorFlow execution'. The 'Highlighted Features' section includes a 'Seedbank' link. The 'TensorFlow execution' section contains a text description and a matrix addition example.

Table of contents

- Getting Started
- Highlighted Features
 - TensorFlow execution
- GitHub
- Visualization
- Forms
- Examples
- Local runtime support

SECTION

Welcome to Colaboratory!

Colaboratory is a free Jupyter notebook environment that requires no setup and runs entirely in the cloud. See our [FAQ](#) for more info.

Getting Started

- [Overview of Colaboratory](#)
- [Loading and saving data: Local files, Drive, Sheets, Google Cloud Storage](#)
- [Importing libraries and installing dependencies](#)
- [Using Google Cloud BigQuery](#)
- [Forms, Charts, Markdown, & Widgets](#)
- [TensorFlow with GPU](#)
- [Machine Learning Crash Course: Intro to Pandas & First Steps with TensorFlow](#)

Highlighted Features

Seedbank

Looking for Colab notebooks to learn from? Check out [Seedbank](#), a place to discover interactive machine learning examples.

TensorFlow execution

Colaboratory allows you to execute TensorFlow code in your browser with a single click. The example below adds two matrices.

$$\begin{bmatrix} 1. & 1. & 1. \end{bmatrix} + \begin{bmatrix} 1. & 2. & 3. \end{bmatrix} = \begin{bmatrix} 2. & 3. & 4. \end{bmatrix}$$

Python in Google Colab

<https://colab.research.google.com/drive/1FEG6DnGvwfUbeo4zJ1zTunjMqf2RkCrT>

python101.ipynb - Collaborator x +

← → ↺ <https://colab.research.google.com/drive/1FEG6DnGvwfUbeo4zJ1zTunjMqf2RkCrT?authuser=2#scrollTo=wsh36fLxDKC3> ☆ 📷 🗺️ ⚙️

python101.ipynb ☆

File Edit View Insert Runtime Tools Help

+ CODE + TEXT ↑ CELL ↓ CELL

✓ CONNECTED ✎ EDITING ^

```
1 # Future Value
2 pv = 100
3 r = 0.1
4 n = 7
5 fv = pv * ((1 + (r)) ** n)
6 print(round(fv, 2))
```

194.87

```
[11] 1 amount = 100
2 interest = 10 #10% = 0.01 * 10
3 years = 7
4
5 future_value = amount * ((1 + (0.01 * interest)) ** years)
6 print(round(future_value, 2))
```

194.87

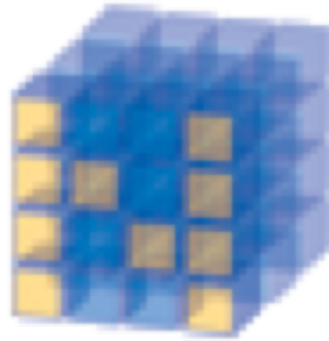
```
[12] 1 # Python Function def
2 def getfv(pv, r, n):
3     fv = pv * ((1 + (r)) ** n)
4     return fv
5 fv = getfv(100, 0.1, 7)
6 print(round(fv, 2))
```

194.87

```
[13] 1 # Python if else
2 score = 80
3 if score >=60 :
4     print("Pass")
5 else:
6     print("Fail").
```

Pass

NumPy



NumPy

Base

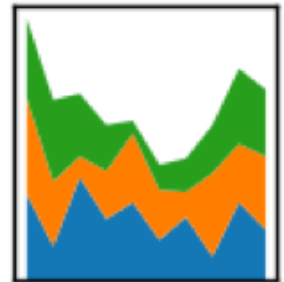
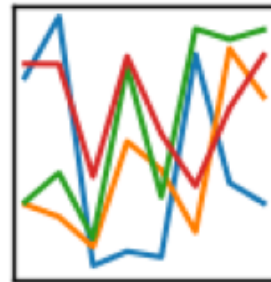
N-dimensional array
package

Python matplotlib *matplotlib*

Python Pandas

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



Iris flower data set

setosa



versicolor



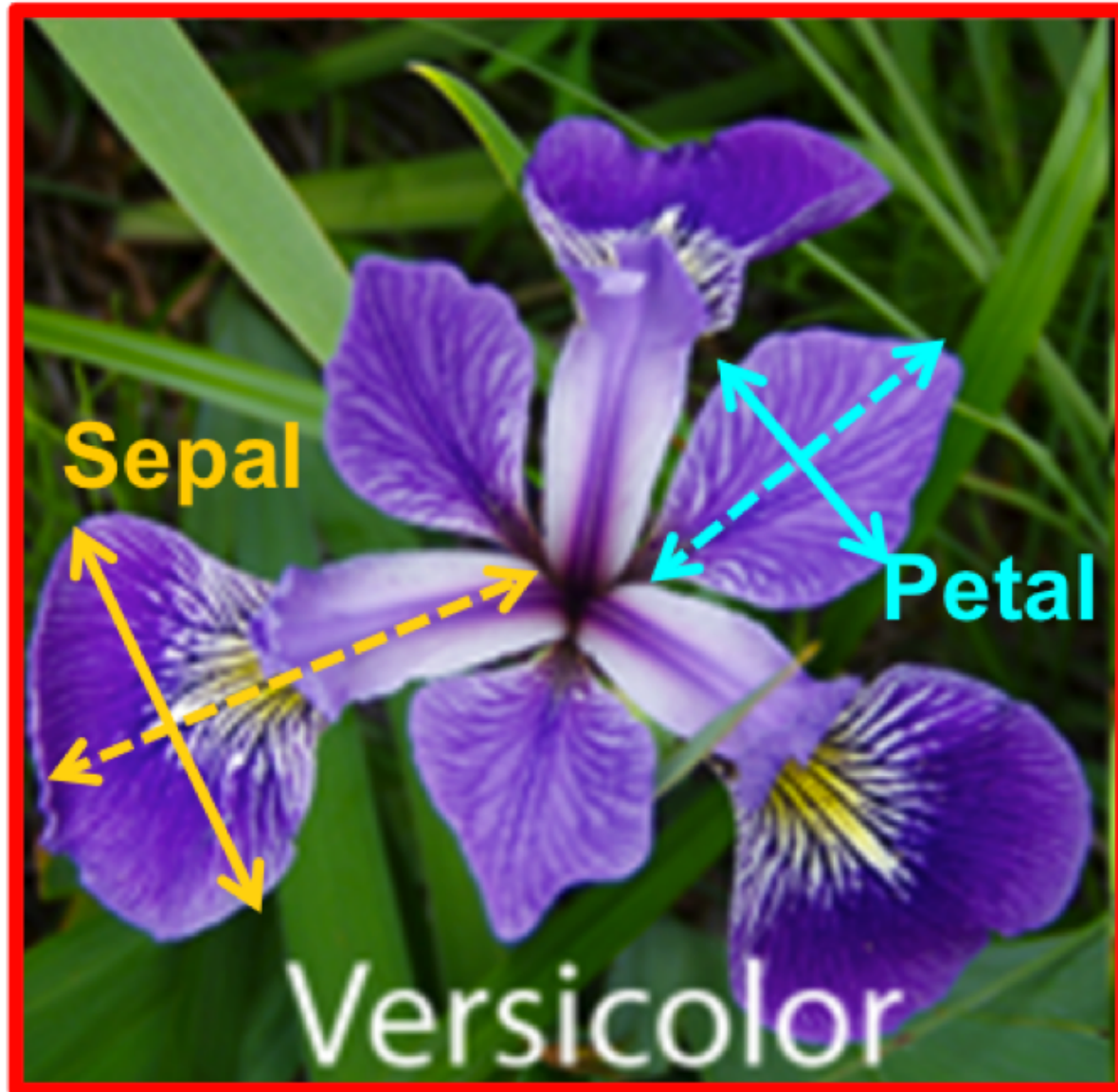
virginica



Source: https://en.wikipedia.org/wiki/Iris_flower_data_set

Source: <http://suruchifialoke.com/2016-10-13-machine-learning-tutorial-iris-classification/>

Iris Classification



iris.data

<https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data>

5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
5.4,3.7,1.5,0.2,Iris-setosa
4.8,3.4,1.6,0.2,Iris-setosa
4.8,3.0,1.4,0.1,Iris-setosa
4.3,3.0,1.1,0.1,Iris-setosa
5.8,4.0,1.2,0.2,Iris-setosa
5.7,4.4,1.5,0.4,Iris-setosa
5.4,3.9,1.3,0.4,Iris-setosa
5.1,3.5,1.4,0.3,Iris-setosa
5.7,3.8,1.7,0.3,Iris-setosa
5.1,3.8,1.5,0.3,Iris-setosa
5.4,3.4,1.7,0.2,Iris-setosa
5.1,3.7,1.5,0.4,Iris-setosa
4.6,3.6,1.0,0.2,Iris-setosa
5.1,3.3,1.7,0.5,Iris-setosa
4.8,3.4,1.9,0.2,Iris-setosa
5.0,3.0,1.6,0.2,Iris-setosa
5.0,3.4,1.6,0.4,Iris-setosa

setosa



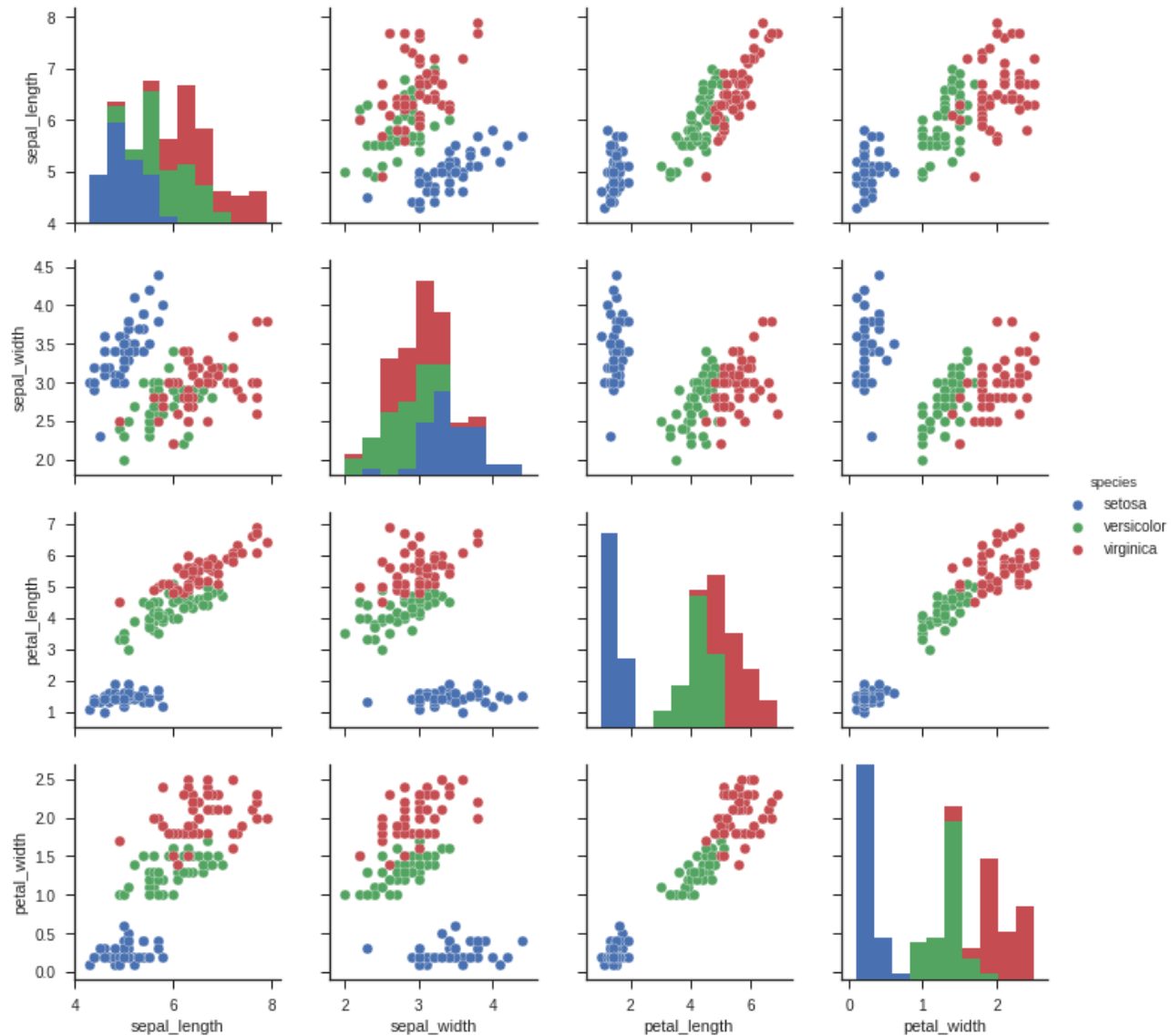
virginica



versicolor



Iris Data Visualization



Connect Google Colab in Google Drive

The screenshot shows the Google Drive web interface. The browser tab is 'My Drive - Google Drive' and the URL is 'https://drive.google.com/drive/u/2/my-drive'. The Drive logo and a search bar are at the top. On the left, the 'New' button is highlighted with a red dashed box. A dropdown menu is open, showing options like 'New folder...', 'Upload files...', 'Upload folder...', 'Google Docs', 'Google Sheets', 'Google Slides', and 'More'. The 'More' option is also highlighted with a red dashed box. A second dropdown menu is open from 'More', listing 'Google Forms', 'Google Drawings', 'Google My Maps', 'Google Sites', and 'Connect more apps'. The 'Connect more apps' option is highlighted with a red dashed box. The main area shows 'My Drive' and 'Quick Access' sections. The bottom of the page has a storage status bar and a 'Get Backup and Sync for Mac' notification.

My Drive - Google Drive

https://drive.google.com/drive/u/2/my-drive

Drive

Search Drive

My Drive

Quick Access

New

My Drive

Computers

Shared with me

Recent

Starred

Trash

Backups

Storage

0 bytes of 15 GB used

UPGRADE STORAGE

Get Backup and Sync for Mac

New folder...

Upload files...

Upload folder...

Google Docs

Google Sheets

Google Slides

More

Google Forms

Google Drawings

Google My Maps

Google Sites

Connect more apps

Name ↑

Google Colab

My Drive - Google Drive x +

https://drive.google.com/drive/u/2/my-drive

Drive

Search Drive

Connect apps to Drive

All ▾ colab x

 ZIP Extractor Extract ZIP files to Google Drive Extraction complete. View extracted files Share Extract another  Test.zip ZIP Extractor 307,585 users	 LUMIN PDF The fast and simple PDF Viewer    Lumin PDF - Beautiful PDF Editor 289,310 users	 cloudconvert CloudConvert 373,161 users
 Sejda Merge PDF - Split PDF - Sejda.com ★★★★★ (1106)	 DocHub Edit, Send & Sign PDFs DocHub - Edit and Sign PDF Docu... 2,131,600 users	 Google Forms Google Forms 4,803,614 users

Get Backup and Sync for Mac

Access anywhere
Every file is online whenever.

Share easily
Give others access to any file or folder.

Google Colab

My Drive - Google Drive x +

https://drive.google.com/drive/u/2/my-drive

Drive

Search Drive

New

My Drive

Computers

Shared with me

Recent

Starred

Trash

Backups

Storage

0 bytes of 15 GB used

[UPGRADE STORAGE](#)

Get Backup and Sync for Mac

Access anywhere

Share easily

Connect apps to Drive

All colab



Colaboratory
offered by <https://colab.research.google.com>
A data analysis tool that combines code, output, and descriptive text into one collaborative document.

[+ CONNECT](#)

Productivity
★★★★★ (195)

Name ↑

Connect Colaboratory to Google Drive

The screenshot shows the Google Drive web interface. A dialog box titled "Connect apps to Drive" is open, displaying a list of apps. The search filter "colab" is entered in the search bar. The Colaboratory app is listed with a green checkmark, indicating it is connected. A confirmation message states: "Colaboratory was connected to Google Drive." Below this message, there is a checkbox labeled "Make Colaboratory the default app for files it can open" which is checked. An "OK" button is visible at the bottom right of the dialog box. The background shows the Google Drive sidebar with options like "New", "My Drive", "Computers", "Shared with me", "Recent", "Starred", "Trash", "Backups", and "Storage". The storage usage is shown as "0 bytes of 15 GB used" with a link to "UPGRADE STORAGE".

My Drive - Google Drive

https://drive.google.com/drive/u/2/my-drive

Drive

Search Drive

New

My Drive

Computers

Shared with me

Recent

Starred

Trash

Backups

Storage

0 bytes of 15 GB used

[UPGRADE STORAGE](#)

Get Backup and Sync for Mac

Access anywhere

Share easily

Connect apps to Drive

colab

Colaboratory was connected to Google Drive.

☒ Make Colaboratory the default app for files it can open

OK

Google Colab

The image shows the Google Drive web interface. At the top, the browser tab is labeled 'My Drive - Google Drive' and the address bar shows the URL 'https://drive.google.com/drive/u/2/my-drive'. The Drive logo and a search bar are visible. On the left sidebar, the 'New' button is highlighted with a red dashed box. A dropdown menu is open, showing options like 'New folder...', 'Upload files...', 'Upload folder...', 'Google Docs', 'Google Sheets', 'Google Slides', and 'More'. The 'More' option is also highlighted with a red dashed box. A second dropdown menu is open from 'More', listing 'Google Forms', 'Google Drawings', 'Google My Maps', 'Google Sites', and 'Colaboratory'. The 'Colaboratory' option is highlighted with a red dashed box. At the bottom left, there is a notification for 'Get Backup and Sync for Mac'.

My Drive - Google Drive

https://drive.google.com/drive/u/2/my-drive

Drive

Search Drive

My Drive

Quick Access

New

My Drive

Computers

Shared with me

Recent

Starred

Trash

Backups

Storage

0 bytes of 15 GB used

UPGRADE STORAGE

Get Backup and Sync for Mac

New folder...

Upload files...

Upload folder...

Google Docs

Google Sheets

Google Slides

More

Google Forms

Google Drawings

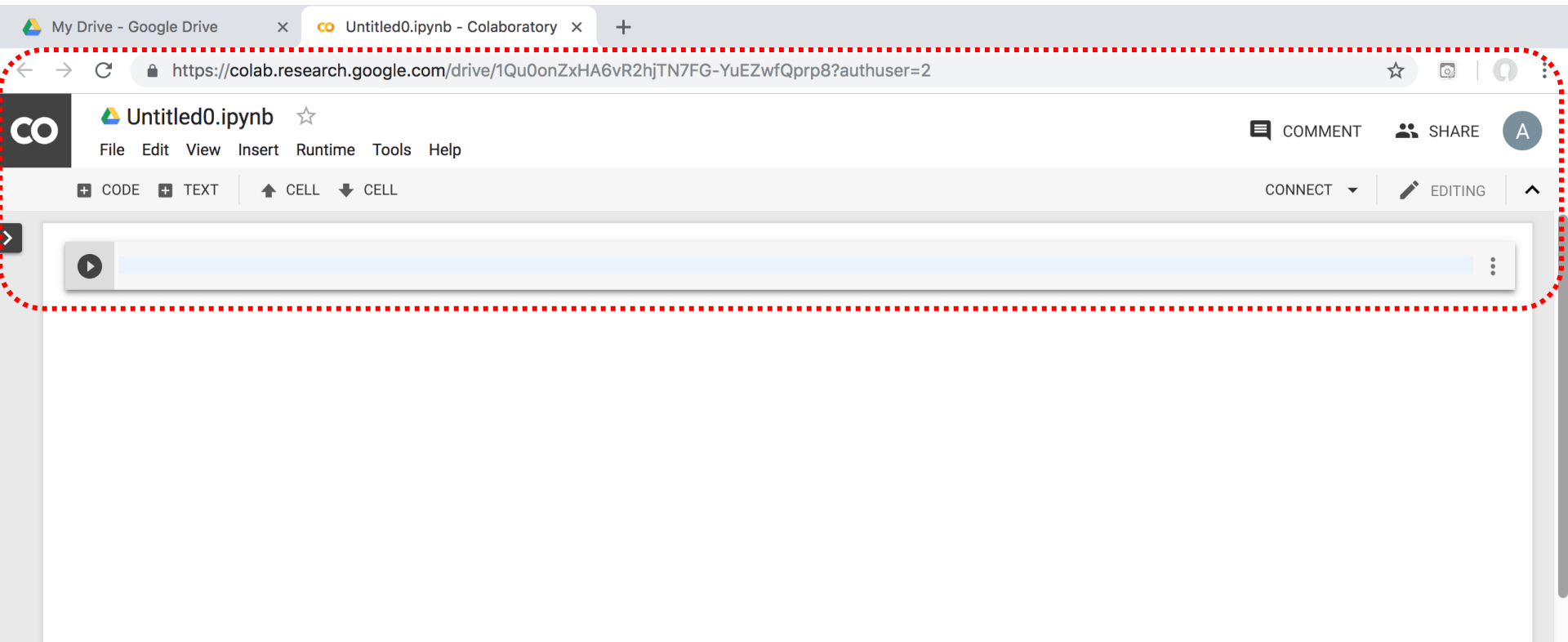
Google My Maps

Google Sites

Colaboratory

Connect more apps

Google Colab



Google Colab

The screenshot shows the Google Colab web interface. At the top, there are two browser tabs: 'My Drive - Google Drive' and 'Untitled0.ipynb - Colaboratory'. The address bar shows the URL: <https://colab.research.google.com/drive/1Qu0onZxHA6vR2hjTN7FG-YuEZwfQprp8?authuser=2>. The main header area includes the Google Colab logo, the file name 'Untitled0.ipynb', and a star icon. Below this is a menu bar with 'File', 'Edit', 'View', 'Insert', 'Runtime', 'Tools', and 'Help'. The 'Runtime' menu is open, showing a list of options: 'Run all' (⌘/Ctrl+F9), 'Run before' (⌘/Ctrl+F8), 'Run the focused cell' (⌘/Ctrl+Enter), 'Run selection' (⌘/Ctrl+Shift+Enter), 'Run after' (⌘/Ctrl+F10), 'Interrupt execution' (⌘/Ctrl+M I), 'Restart runtime...' (⌘/Ctrl+M .), 'Restart and run all...', 'Reset all runtimes...', 'Change runtime type', and 'Manage sessions'. The 'Runtime' menu item in the top bar and the 'Change runtime type' option in the dropdown are highlighted with red dashed boxes. On the right side of the interface, there are buttons for 'COMMENT', 'SHARE', and a user profile icon. Below these are 'CONNECT' and 'EDITING' buttons. The main workspace area is visible in the background, showing a code editor with a play button icon on the left.

Run Jupyter Notebook Python3 GPU Google Colab

The screenshot shows the Google Colab web interface. At the top, the browser tabs include 'My Drive - Google Drive' and 'Untitled0.ipynb - Colaboratory'. The address bar shows the URL: <https://colab.research.google.com/drive/1Qu0onZxHA6vR2hjTN7FG-YuEZwfQprp8?authuser=2>. The Colab logo is on the left, and the menu bar includes 'File', 'Edit', 'View', 'Insert', 'Runtime', 'Tools', and 'Help'. On the right, there are buttons for 'COMMENT', 'SHARE', and a user profile icon. Below the menu bar, there are tabs for '+ CODE', '+ TEXT', and buttons for '↑ CELL' and '↓ CELL'. The main area is a Jupyter notebook editor. A 'Notebook settings' dialog box is open in the center, featuring two dropdown menus: 'Runtime type' set to 'Python 3' and 'Hardware accelerator' set to 'GPU'. Both dropdowns are highlighted with red dashed boxes. A blue question mark icon is next to the 'Hardware accelerator' dropdown. Below these is a checkbox labeled 'Omit code cell output when saving this notebook', which is currently unchecked. At the bottom right of the dialog are 'CANCEL' and 'SAVE' buttons.

My Drive - Google Drive x Untitled0.ipynb - Colaboratory x +

← → ↻ <https://colab.research.google.com/drive/1Qu0onZxHA6vR2hjTN7FG-YuEZwfQprp8?authuser=2> ☆ 📷 🔊 ⋮

co Untitled0.ipynb ☆

File Edit View Insert Runtime Tools Help

+ CODE + TEXT ↑ CELL ↓ CELL

CONNECT ▾ EDITING ^

>

Notebook settings

Runtime type
Python 3 ▾

Hardware accelerator
GPU ▾ ⓘ

☐ Omit code cell output when saving this notebook

CANCEL SAVE

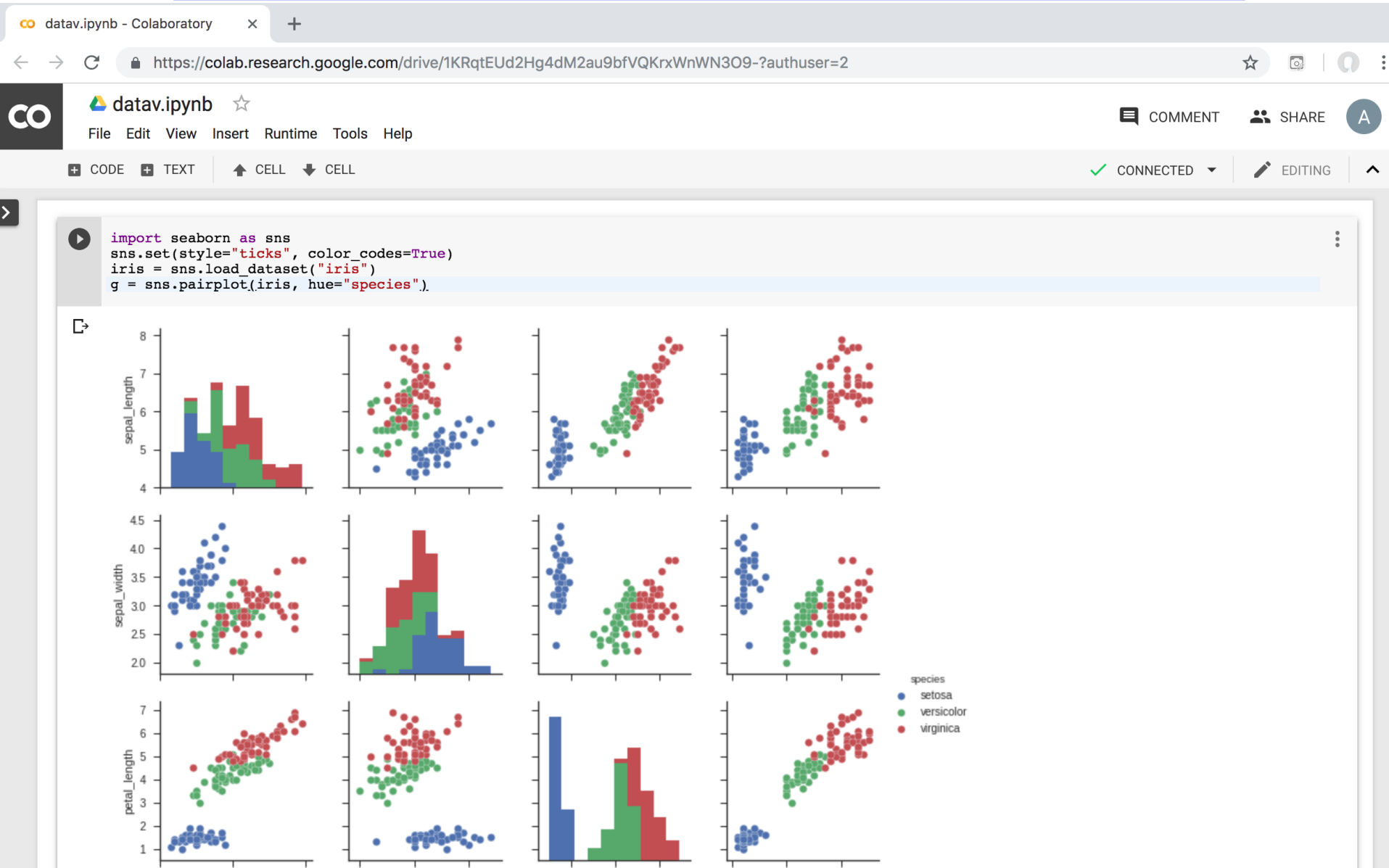
Google Colab Python Hello World

```
print('Hello World')
```



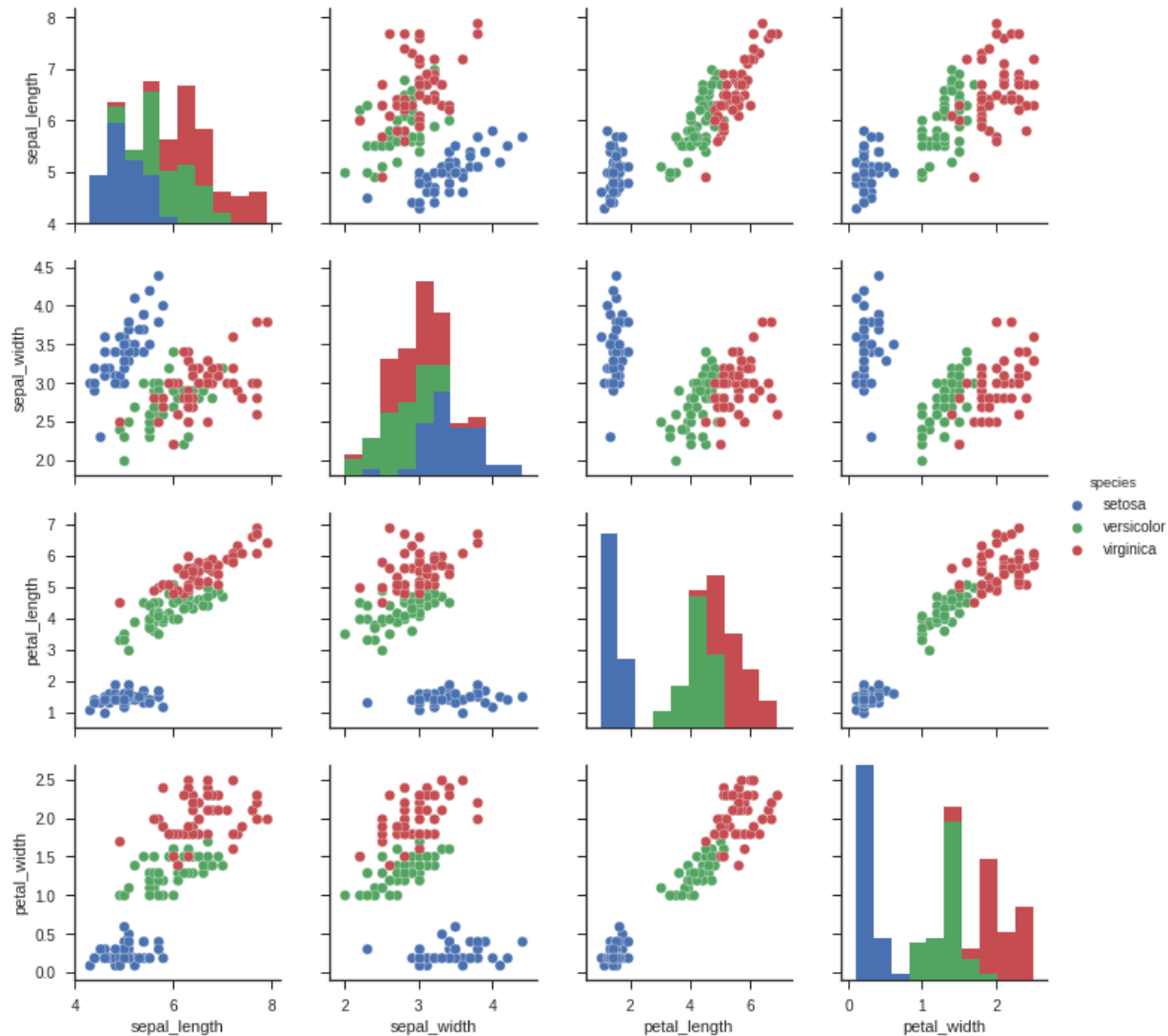
Data Visualization in Google Colab

<https://colab.research.google.com/drive/1KRqtEUd2Hg4dM2au9bfVQKrxWnWN3O9->



Source: <https://seaborn.pydata.org/generated/seaborn.pairplot.html>


```
import seaborn as sns
sns.set(style="ticks", color_codes=True)
iris = sns.load_dataset("iris")
g = sns.pairplot(iris, hue="species")
```



```
import numpy as np
import pandas as pd
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
from pandas.plotting import scatter_matrix

# Load dataset
url = "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"
names = ['sepal-length', 'sepal-width', 'petal-length', 'petal-width', 'class']
df = pd.read_csv(url, names=names)

print(df.head(10))
print(df.tail(10))
print(df.describe())
print(df.info())
print(df.shape)
print(df.groupby('class').size())

plt.rcParams["figure.figsize"] = (10,8)
df.plot(kind='box', subplots=True, layout=(2,2), sharex=False, sharey=False)
plt.show()

df.hist()
plt.show()

scatter_matrix(df)
plt.show()

sns.pairplot(df, hue="class", size=2)
```

```
import numpy as np
import pandas as pd
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
from pandas.plotting import scatter_matrix
```

```
# Import Libraries
import numpy as np
import pandas as pd
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
from pandas.plotting import scatter_matrix
print('imported')
```

imported

```
url = "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"
names = ['sepal-length', 'sepal-width', 'petal-length', 'petal-width', 'class']
df = pd.read_csv(url, names=names)
print(df.head(10))
```

```
# Load dataset
```

```
url = "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"
names = ['sepal-length', 'sepal-width', 'petal-length', 'petal-width', 'class']
df = pd.read_csv(url, names=names)
print(df.head(10)).
```

	sepal-length	sepal-width	petal-length	petal-width	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
5	5.4	3.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa
8	4.4	2.9	1.4	0.2	Iris-setosa
9	4.9	3.1	1.5	0.1	Iris-setosa

df.tail(10)

```
print(df.tail(10))
```

	sepal-length	sepal-width	petal-length	petal-width	class
140	6.7	3.1	5.6	2.4	Iris-virginica
141	6.9	3.1	5.1	2.3	Iris-virginica
142	5.8	2.7	5.1	1.9	Iris-virginica
143	6.8	3.2	5.9	2.3	Iris-virginica
144	6.7	3.3	5.7	2.5	Iris-virginica
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

df.describe()

```
print(df.describe())
```

	sepal-length	sepal-width	petal-length	petal-width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

```
print(df.info())  
print(df.shape)
```

```
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 150 entries, 0 to 149  
Data columns (total 5 columns):  
sepal-length      150 non-null float64  
sepal-width       150 non-null float64  
petal-length      150 non-null float64  
petal-width       150 non-null float64  
class             150 non-null object  
dtypes: float64(4), object(1)  
memory usage: 5.9+ KB  
None
```

```
print(df.shape)
```

```
(150, 5)
```

```
df.groupby( 'class' ).size()
```

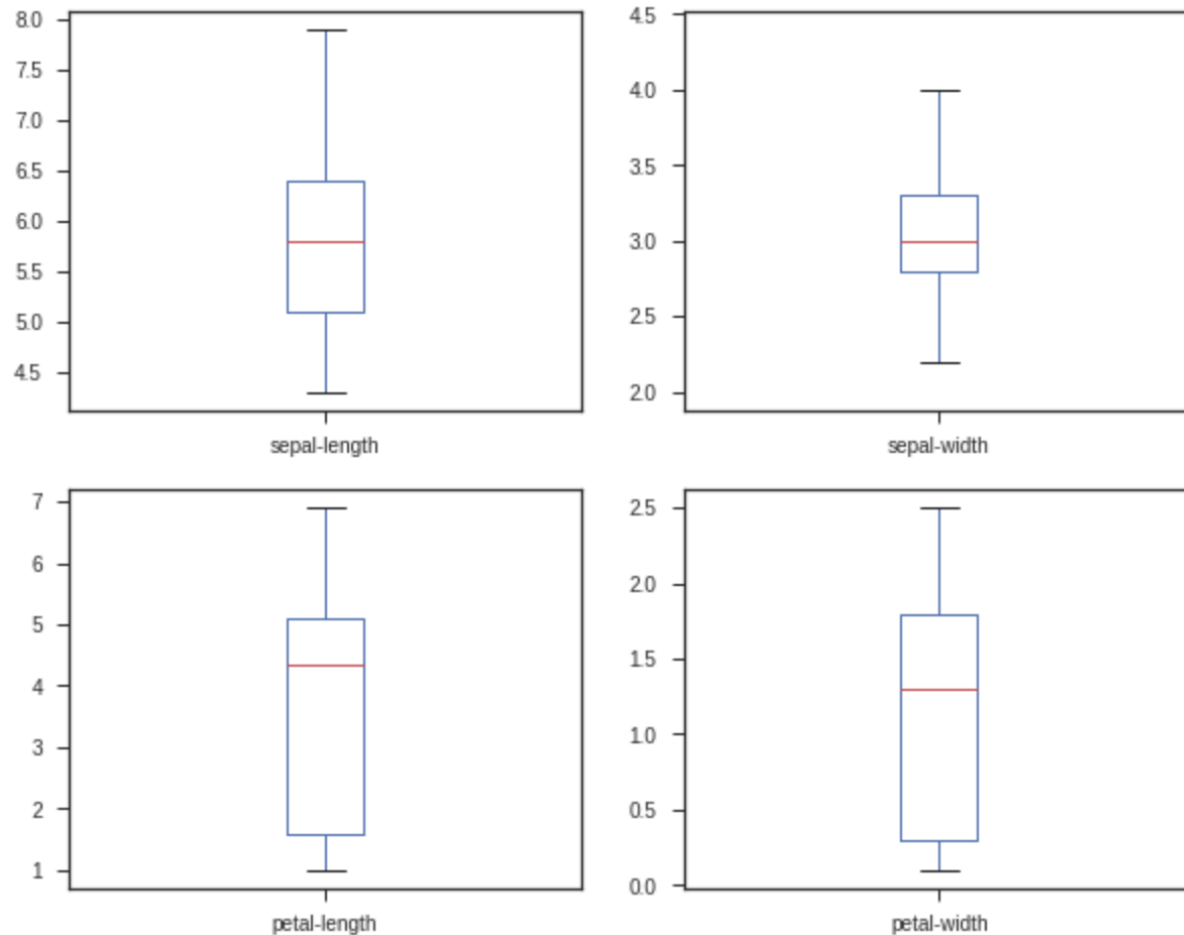
```
print(df.groupby( 'class' ).size())
```

```
class
Iris-setosa      50
Iris-versicolor  50
Iris-virginica   50
dtype: int64
```



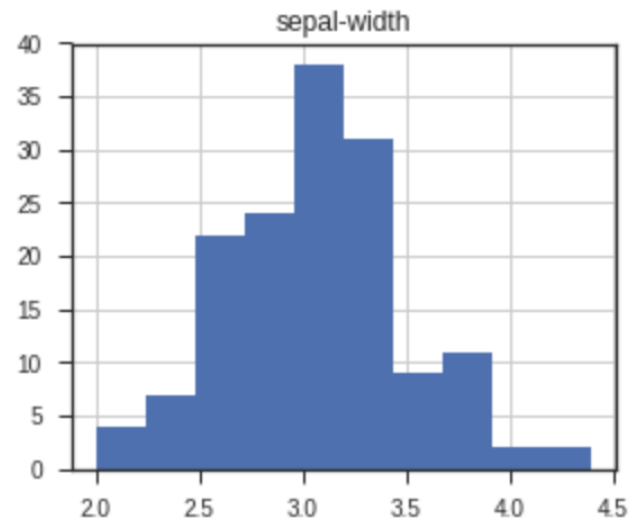
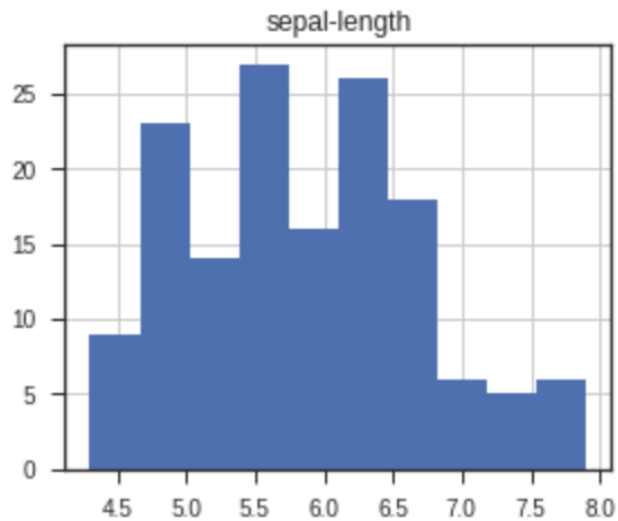
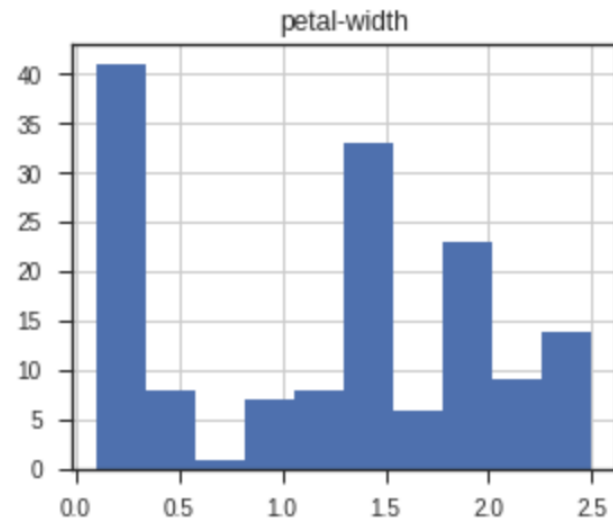
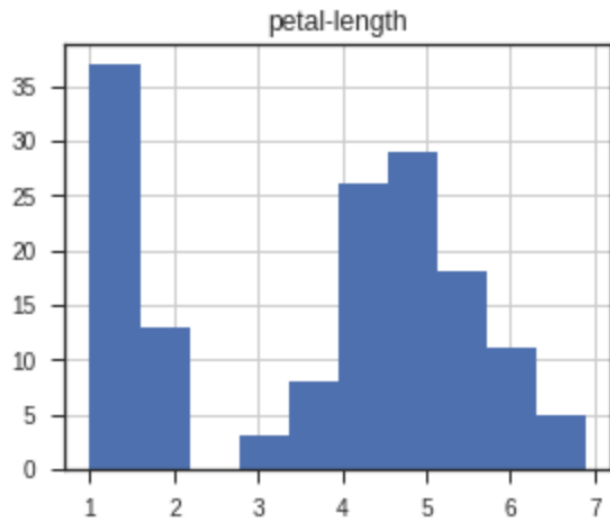
```
plt.rcParams["figure.figsize"] = (10,8)
df.plot(kind='box', subplots=True, layout=(2,2), sharex=False, sharey=False)
plt.show()
```

```
plt.rcParams["figure.figsize"] = (10,8)
df.plot(kind='box', subplots=True, layout=(2,2), sharex=False, sharey=False)
plt.show()
```



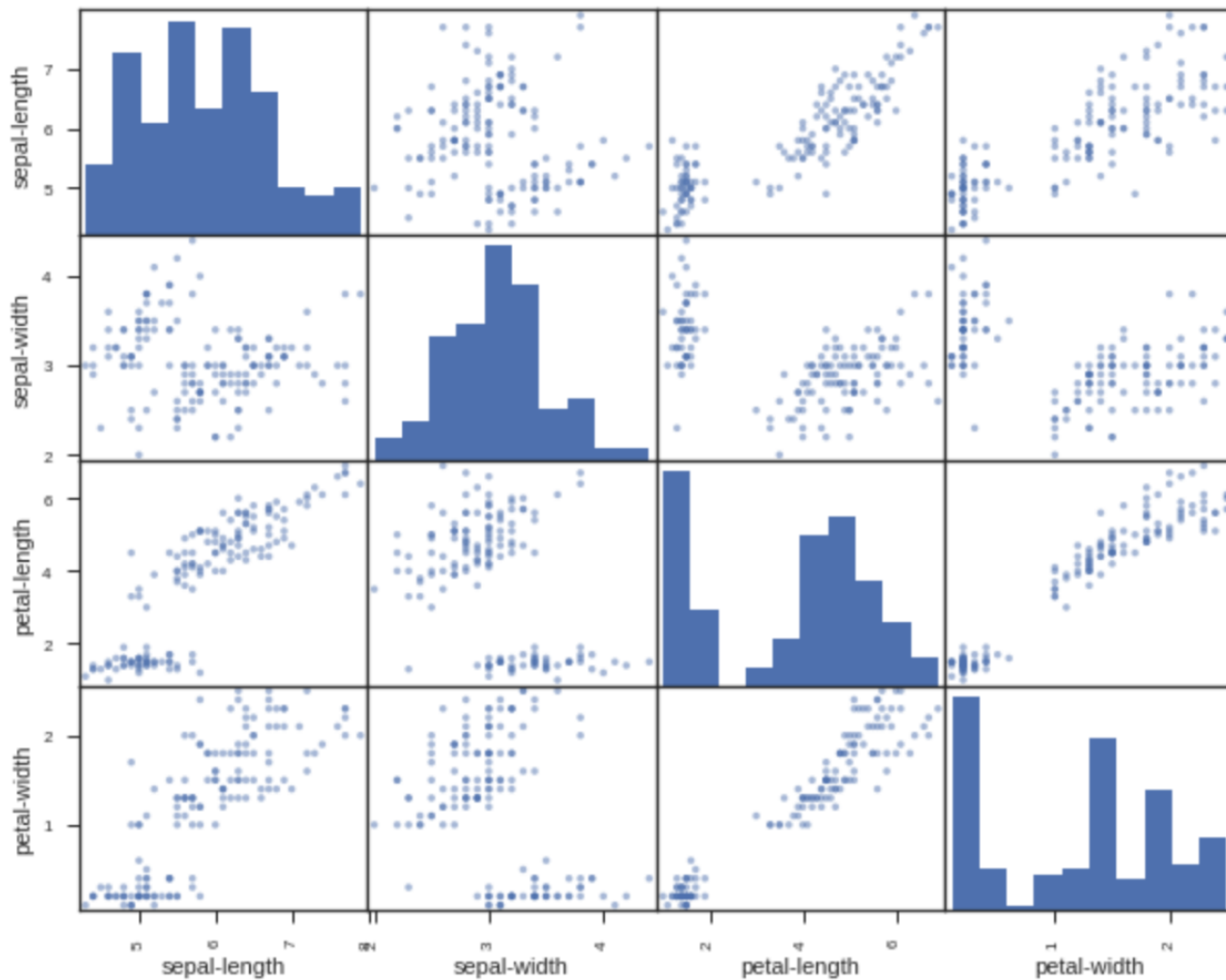
```
df.hist()  
plt.show()
```

```
df.hist()  
plt.show()
```



```
scatter_matrix(df)  
plt.show()
```

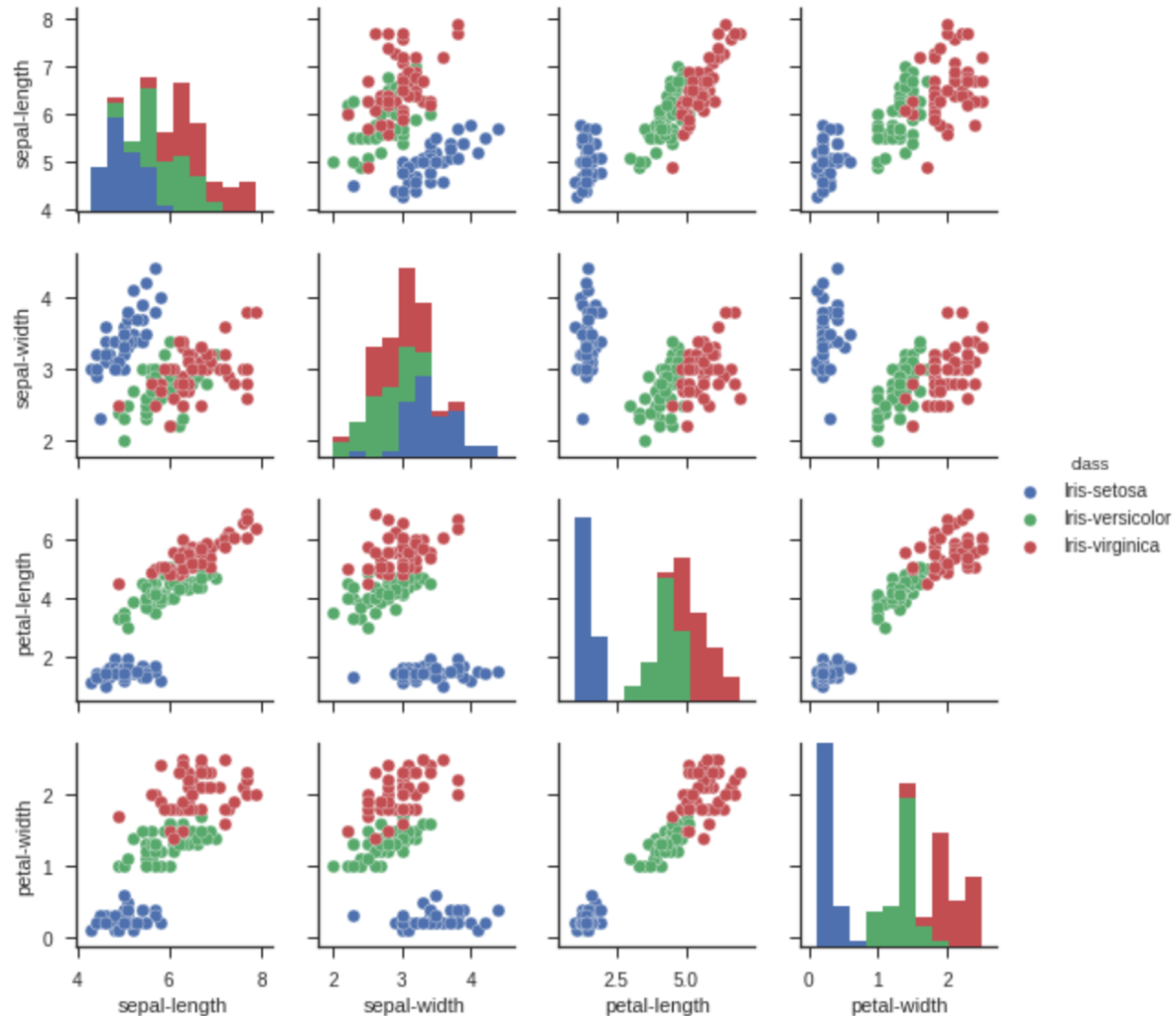
```
scatter_matrix(df)  
plt.show(.)
```



`sns.pairplot(df, hue="class", size=2)`

```
sns.pairplot(df, hue="class", size=2)
```

<seaborn.axisgrid.PairGrid at 0x7f1d21267390>





Anaconda

The Most Popular

Python

Data Science Platform

Download Anaconda



Don't Miss AnacondaCon Apr 8-11 Austin TX!

Download Anaconda Distribution

Version 5.1 | Release Date: February 15, 2018

Download For:   

High-Performance Distribution

Easily install 1,000+ [data science packages](#)

Package Management

Manage packages, dependencies and environments with [conda](#)

Portal to Data Science

Uncover insights in your data and create interactive visualizations



Anaconda 5.1 For macOS Installer

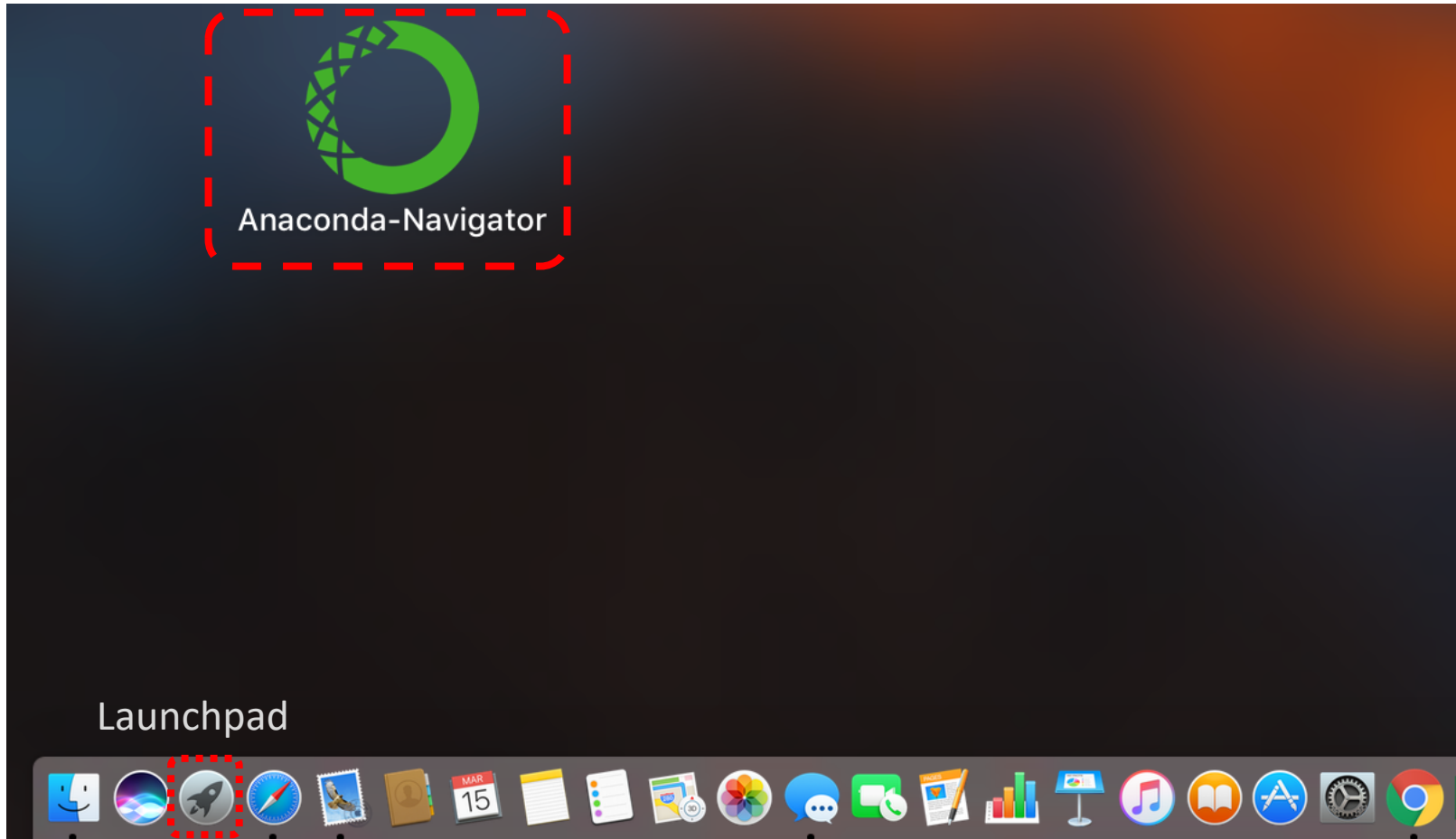
<https://www.anaconda.com/download>



Python

HelloWorld

Anaconda-Navigator



Anaconda Navigator



Home



Environments



Learning



Community

Documentation

Developer Blog

Feedback



Applications on

base (root)

Channels

Refresh



jupyterlab

0.31.5

An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.

Launch



jupyter
notebook

5.4.0

Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

Launch



qtconsole

4.3.1

PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.

Launch



spyder

3.2.6

Scientific PYTHON Development Environment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features

Launch

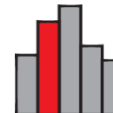


vscode

1.22.2

Streamlined code editor with support for development operations like debugging, task running and version control.

Launch



glueviz

0.12.4

Multidimensional data visualization across files. Explore relationships within and among related datasets.

Install

Jupyter Notebook

The screenshot shows the Jupyter Notebook web interface in a browser window. The address bar displays `localhost:8888/tree/Documents/Data/BDA`. The Jupyter logo and a "Logout" button are in the top right. Below the navigation tabs (Files, Running, Clusters), there are buttons for "Upload", "New", and a refresh icon. A message "Select items to perform actions on them." is shown. The file browser table has a header with "Name" and "Last Modified" columns. The first row shows a folder icon, a selection checkbox, the number "0", and the path `/ Documents / Data / BDA`. The second row shows a folder icon, `..`, and `seconds ago`. A red dashed box highlights the first two rows. The third row contains the text "The notebook list is empty."

Home

localhost:8888/tree/Documents/Data/BDA

jupyter Logout

Files Running Clusters

Select items to perform actions on them.

Upload New ↕ ↺

	Name ↓	Last Modified
☐ 0	/ Documents / Data / BDA	
..		seconds ago
The notebook list is empty.		

Jupyter Notebook

New Python 3

The screenshot shows a web browser window with the Jupyter Notebook interface. The browser tabs are 'Home' and 'Untitled'. The address bar shows 'localhost:8888/tree/Documents/Data/BDA'. The Jupyter logo and 'Logout' button are in the top right. Below the logo are tabs for 'Files', 'Running', and 'Clusters'. A message says 'Select items to perform actions on them.' Below this is a file browser showing the path '/ Documents / Data / BDA'. To the right of the file browser are 'Upload' and 'New' buttons. The 'New' button is highlighted with a red box, and its dropdown menu is open, showing options: 'Notebook:', 'Python 3' (highlighted with a red box), 'Other:', 'Text File', 'Folder', and 'Terminal'.

Home x Untitled x u

localhost:8888/tree/Documents/Data/BDA

jupyter Logout

Files Running Clusters

Select items to perform actions on them.

0 ▾ / Documents / Data / BDA

..

The notebook list is empty.

Upload New ▾

Notebook:

Python 3

Other:

Text File

Folder

Terminal

print("hello, world")

The screenshot shows a web browser window with two tabs: 'Home' and 'HelloWorld'. The address bar displays 'localhost:8888/notebooks/Documents/Data/BDA/HelloWorld.ipynb'. The Jupyter Notebook interface is titled 'jupyter HelloWorld (autosaved)' and includes a 'Logout' button. A menu bar at the top contains 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', 'Widgets', and 'Help'. Below the menu is a toolbar with icons for saving, adding, deleting, copying, pasting, undo, redo, and running. The 'Run' button, represented by a play icon, is circled in red. The main area of the notebook contains a code cell with the text 'In [1]: print("hello, world")' and the output 'hello, world'. The code line is also circled in red. Below the first cell is an empty input prompt 'In []: '.

```
from platform import python_version
print("Python Version:", python_version())
```

The screenshot shows a web browser window with two tabs: 'Home' and 'HelloWorld'. The address bar shows the URL 'localhost:8888/notebooks/Documents/Data/BDA/HelloWorld.ipynb'. The Jupyter Notebook interface is displayed, with the title 'jupyter HelloWorld (unsaved changes)' and a 'Logout' button. The top menu bar includes 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', 'Widgets', and 'Help'. The right side of the menu bar shows 'Trusted', a pencil icon, and 'Python 3'. Below the menu bar is a toolbar with icons for saving, adding, deleting, and running cells. The notebook content area shows two code cells. The first cell, labeled 'In [1]:', contains the code `print("hello, world")` and has outputted 'hello, world'. The second cell, labeled 'In [2]:', contains the code `from platform import python_version` and `print("Python Version:", python_version())` and has outputted 'Python Version: 3.6.5'. A third cell, labeled 'In []:', is currently empty.

```
In [1]: print("hello, world")
hello, world

In [2]: from platform import python_version
print("Python Version:", python_version())
Python Version: 3.6.5

In [ ]:
```



Python

Programming



Python Fiddle

The screenshot shows the Python Fiddle web application in a browser. The browser's address bar displays 'pythonfiddle.com'. The application's header includes a navigation bar with buttons for 'Run', 'Reset', 'Share', 'Import', and 'Login', along with a 'Language' dropdown menu. On the right side of the header, the 'Python Fiddle' logo is displayed next to the 'Python Cloud IDE' logo. Below the header, the main workspace is divided into three sections. On the left is a sidebar with a list of 'Examples' including 'Chaining comparison operators', 'Decorators', 'Creating generators objects', 'Enumerate', 'Function closure', 'Lex tokenizer', 'Step argument in slice operators', 'For Else', 'Verbose regular expressions', 'In-place value swapping', and 'Function argument unpacking'. Below the examples list are buttons for 'Packages' and 'Hotkeys'. The central workspace contains a code editor with two lines of Python code: '1 print("Hello Python Fiddle")' and '2'. To the right of the code editor is a metadata panel with fields for 'Title', 'Description', and 'Tags', each with a corresponding input box. Below the 'Tags' input box is a note: 'A comma-separated list of tags.' and a 'Save' button.

Python Cloud IDE | Python Fiddle x

pythonfiddle.com

Run Reset Share Import Login Language

Python Fiddle Python Cloud IDE

G+1 2.6k

Examples

- [Chaining comparison operators](#)
- [Decorators](#)
- [Creating generators objects](#)
- [Enumerate](#)
- [Function closure](#)
- [Lex tokenizer](#)
- [Step argument in slice operators](#)
- [For Else](#)
- [Verbose regular expressions](#)
- [In-place value swapping](#)
- [Function argument unpacking](#)

Packages Hotkeys

```
1 print("Hello Python Fiddle")
2
```

Title:

Description:

Tags:

A comma-separated list of tags.

Save

Hello Python Fiddle

Text input and output

```
print("Hello World")
```

```
print("Hello World\nThis is a message")
```

```
x = 3  
print(x)
```

```
x = 2  
y = 3  
print(x, ' ', y)
```

```
name = input("Enter a name: ")
```

```
x = int(input("What is x? "))
```

```
x = float(input("Write a number "))
```

Python in Google Colab

<https://colab.research.google.com/drive/1FEG6DnGvwfUbeo4zJ1zTunjMqf2RkCrT>

python101.ipynb - Colaboratory

https://colab.research.google.com/drive/1FEG6DnGvwfUbeo4zJ1zTunjMqf2RkCrT?authuser=2#scrollTo=LSNfVqJO-7-U

python101.ipynb

File Edit View Insert Runtime Tools Help

COMMENT SHARE

CONNECTED EDITING

1 `print("hello, world").`

hello, world

[2] 1 `# comment`
2 `from platform import python_version`
3 `print("Python Version:", python_version())`

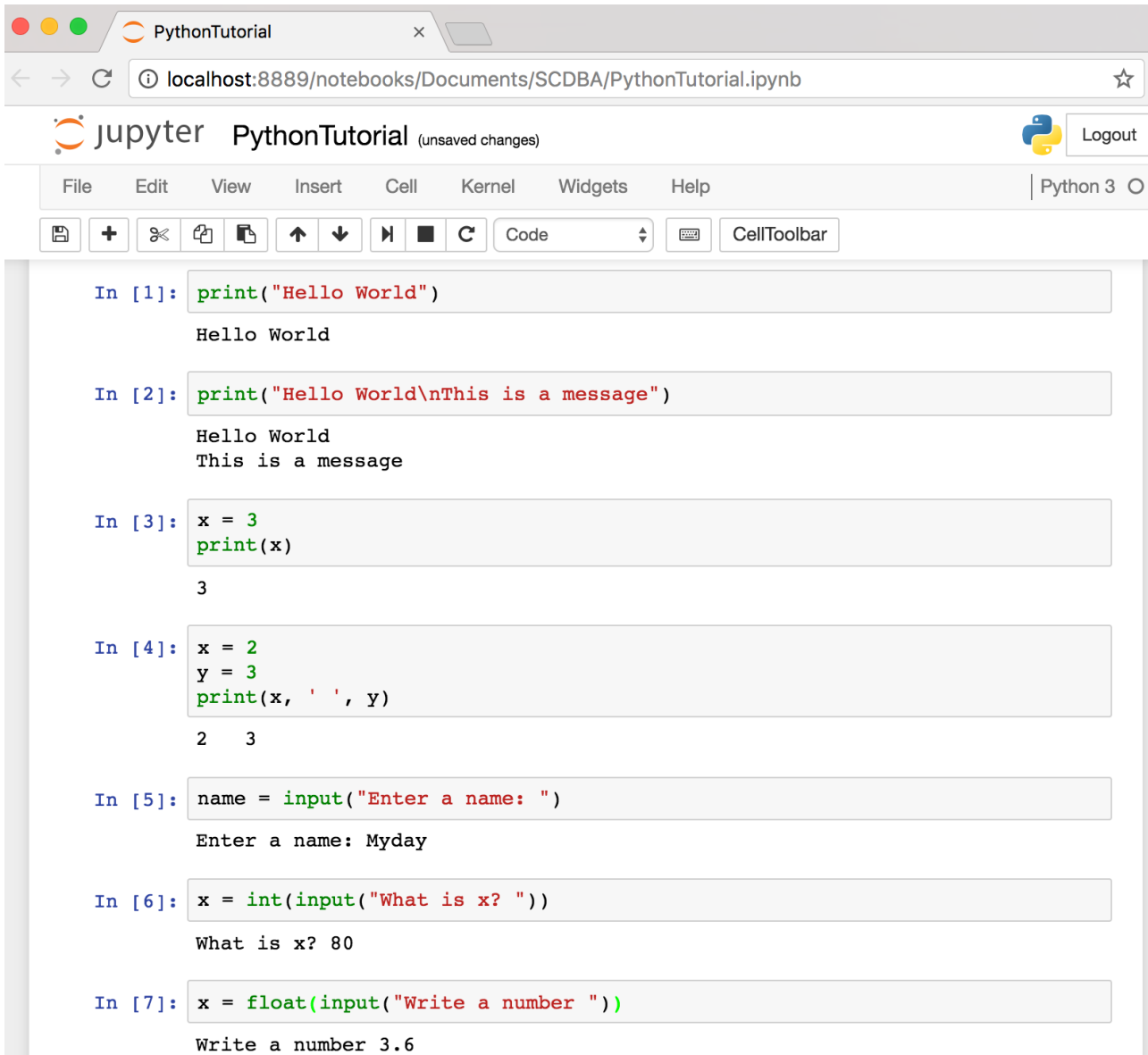
Python Version: 3.6.6

[3] 1 `# https://www.learnpython.org/en/`
2 `# LearnPython.org interactive Python tutorial`
3 `print("Hello World")`
4 `print("Hello World\nThis is a message")`
5 `x = 3`
6 `print(x)`
7 `x = 2`
8 `y = 3`
9 `print(x, ' ', y).`

Hello World
Hello World
This is a message
3
2 3

[4] 1 `# Python Variables`
2 `x = 2`
3 `price = 2.5`
4 `word = 'Hello'`
5
6 `word = 'Hello'`
7 `word = "Hello"`
8 `word = '''Hello'''`

Text input and output



The screenshot shows a Jupyter Notebook window titled "PythonTutorial" with the URL `localhost:8889/notebooks/Documents/SCDBA/PythonTutorial.ipynb`. The interface includes a top bar with the Jupyter logo, the notebook title, and a "Logout" button. Below the top bar is a menu bar with options: File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. A toolbar with various icons for file operations and execution is located below the menu bar. The notebook contains seven code cells, each with input and output:

```
In [1]: print("Hello World")
Hello World

In [2]: print("Hello World\nThis is a message")
Hello World
This is a message

In [3]: x = 3
print(x)
3

In [4]: x = 2
y = 3
print(x, ' ', y)
2  3

In [5]: name = input("Enter a name: ")
Enter a name: Myday

In [6]: x = int(input("What is x? "))
What is x? 80

In [7]: x = float(input("Write a number "))
Write a number 3.6
```

Variables

```
x = 2  
price = 2.5  
word = 'Hello'
```

```
word = 'Hello'  
word = "Hello"  
word = '''Hello'''
```

```
x = 2  
x = x + 1  
x = 5
```

Python Basic Operators

```
print('7 + 2 =', 7 + 2)
print('7 - 2 =', 7 - 2)
print('7 * 2 =', 7 * 2)
print('7 / 2 =', 7 / 2)
print('7 // 2 =', 7 // 2)
print('7 % 2 =', 7 % 2)
print('7 ** 2 =', 7 ** 2)
```

```
print('7 + 2 =', 7 + 2)
print('7 - 2 =', 7 - 2)
print('7 * 2 =', 7 * 2)
print('7 / 2 =', 7 / 2)
print('7 // 2 =', 7 // 2)
print('7 % 2 =', 7 % 2)
print('7 ** 2 =', 7 ** 2)
```

```
7 + 2 = 9
7 - 2 = 5
7 * 2 = 14
7 / 2 = 3.5
7 // 2 = 3
7 % 2 = 1
7 ** 2 = 49
```

BMI Calculator in Python

```
height_cm = float(input("Enter your height in cm: "))  
weight_kg = float(input("Enter your weight in kg: "))  
  
height_m = height_cm/100  
BMI = (weight_kg/(height_m**2))  
  
print("Your BMI is: " + str(round(BMI,1)))
```

BMI Calculator in Python

 jupyter PythonTutorial Last Checkpoint: a minute ago (unsaved changes)

 Logout

File Edit View Insert Cell Kernel Widgets Help

Python 3 

          Code   CellToolbar

```
In [1]: height_cm = float(input("Enter your height in cm: "))
weight_kg = float(input("Enter your weight in kg: "))

height_m = height_cm/100
BMI = (weight_kg/(height_m**2))

print("Your BMI is: " + str(round(BMI,1)))
```

```
Enter your height in cm: 170
Enter your weight in kg: 60
Your BMI is: 20.8
```

```
In [ ]:
```

Future value
of a specified
principal amount,
rate of interest, and
a number of years

Future Value (FV)

```
# How much is your $100 worth after 7 years?  
print(100 * 1.1 ** 7)  
# output = 194.87
```

```
print(100 * 1.1 ** 7)
```

```
194.871710000000012
```

Future Value (FV)

```
pv = 100  
r = 0.1  
n = 7
```

```
fv = pv * ((1 + (r)) ** n)  
print(round(fv, 2))
```

```
pv = 100  
r = 0.1  
n = 7  
  
fv = pv * ((1 + (r)) ** n)  
print(round(fv, 2))
```

194.87

Future Value (FV)

```
amount = 100
interest = 10 #10% = 0.01 * 10
years = 7

future_value = amount * ((1 + (0.01 * interest)) ** years)
print(round(future_value, 2))
```

```
amount = 100
interest = 10 #10% = 0.01 * 10
years = 7

future_value = amount * ((1 + (0.01 * interest)) ** years)
print(round(future_value, 2))
```

194.87

if statements

> greater than
< smaller than
== equals
!= is not

```
score = 80
if score >=60 :
    print("Pass")
else:
    print("Fail")
```

Pass

```
score = 80
if score >=60 :
    print("Pass")
else:
    print("Fail")
```

if elif else

```
score = 90
grade = ""
if score >= 90:
    grade = "A"
elif score >= 80:
    grade = "B"
elif score >= 70:
    grade = "C"
elif score >= 60:
    grade = "D"
else:
    grade = "E"
print(grade)
# grade = "A"
```

A

<http://pythontutor.com/visualize.html>
<https://goo.gl/E6w5ph>

for loops

```
for i in range(1,11):  
    print(i)
```

```
1  
2  
3  
4  
5  
6  
7  
8  
9  
10
```

for loops

```
for i in range(1,10):  
    for j in range(1,10):  
        print(i, ' * ', j, ' = ', i*j)
```

```
9 * 1 = 9  
9 * 2 = 18  
9 * 3 = 27  
9 * 4 = 36  
9 * 5 = 45  
9 * 6 = 54  
9 * 7 = 63  
9 * 8 = 72  
9 * 9 = 81
```

while loops

```
age = 10
```

```
while age < 20:  
    print(age)  
    age = age + 1
```

10

11

12

13

14

15

16

17

18

19

Functions

```
def convertCMtoM(xcm):  
    m = xcm/100  
    return m
```

```
cm = 180  
m = convertCMtoM(cm)  
print(str(m))
```

1.8

Lists

```
x = [60, 70, 80, 90]  
print(len(x))  
print(x[0])  
print(x[1])  
print(x[-1])
```

4

60

70

90

Tuples

A **tuple** in Python is a collection that **cannot be modified**.

A tuple is defined using **parenthesis**.

```
x = (10, 20, 30, 40, 50)
```

```
print(x[0])
```

```
print(x[1])
```

```
print(x[2])
```

```
print(x[-1])
```

10

20

30

50

Dictionary

```
k = { 'EN': 'English', 'FR': 'French' }  
print(k['EN'])
```

Dictionary

'EN' → **'English'**

'FR' → **'French'**

English

Sets

```
animals = {'cat', 'dog'}
```

```
animals = {'cat', 'dog'}
print('cat' in animals)  # Check if an element is in a set; prints "True"
print('fish' in animals) # prints "False"
animals.add('fish')      # Add an element to a set
print('fish' in animals) # Prints "True"
print(len(animals))      # Number of elements in a set; prints "3"
animals.add('cat')       # Adding an element that is already in the set does nothing
print(len(animals))      # Prints "3"
animals.remove('cat')    # Remove an element from a set
print(len(animals))      # Prints "2"
```

```
True
False
True
3
3
2
```

```
animals = {'cat', 'dog'}
print('cat' in animals)
print('fish' in animals)
animals.add('fish')
print('fish' in animals)
print(len(animals))
animals.add('cat')
print(len(animals))
animals.remove('cat')
print(len(animals))
```

File Input / Output

```
with open('myfile.txt', 'w') as file:  
    file.write('Hello World\nThis is Python File Input Output')  
  
with open('myfile.txt', 'r') as file:  
    text = file.read()  
print(text)
```

```
with open('myfile.txt', 'w') as file:  
    file.write('Hello World\nThis is Python File Input Output')
```

```
with open('myfile.txt', 'r') as file:  
    text = file.read()  
print(text)
```

```
Hello World  
This is Python File Input Output
```

```
text
```

```
'Hello World\nThis is Python File Input Output'
```

File Input / Output

```
with open('myfile.txt', 'a+') as file:  
    file.write('\n' + 'New line')
```

```
with open('myfile.txt', 'r') as file:  
    text = file.read()  
print(text)
```

```
with open('myfile.txt', 'a+') as file:  
    file.write('\n' + 'New line')
```

```
with open('myfile.txt', 'r') as file:  
    text = file.read()  
print(text)
```

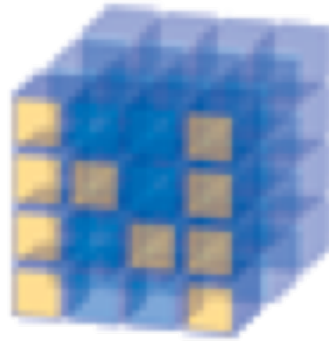
Hello World
This is Python File Input Output
New line

Big Data Analytics
with

Numpy

in Python

NumPy

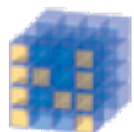


NumPy

Base

N-dimensional array
package

NumPy
is the
fundamental package
for
scientific computing
with Python.



NumPy

NumPy

- NumPy provides a **multidimensional array** object to store homogenous or heterogeneous data; it also provides **optimized functions/methods** to operate on this array object.

NumPy



NumPy

[Scipy.org](http://scipy.org)

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

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

NumPy is licensed under the [BSD license](#), enabling reuse with few restrictions.

Getting Started

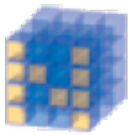
- [Getting NumPy](#)
- [Installing the SciPy Stack](#)
- [NumPy and SciPy documentation page](#)
- [NumPy Tutorial](#)
- [NumPy for MATLAB® Users](#)
- [NumPy functions by category](#)
- [NumPy Mailing List](#)

For more information on the SciPy Stack (for which NumPy provides the fundamental array data structure), see scipy.org.

About NumPy

[License](#)

[Old array packages](#)



NumPy

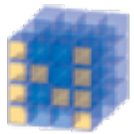
NumPy ndarray

One-dimensional Array (1-D Array)

0	1			n-1
1	2	3	4	5

Two-dimensional Array (2-D Array)

	0	1			n-1
0	1	2	3	4	5
1	6	7	8	9	10
	11	12	13	14	15
m-1	16	17	18	19	20



NumPy

NumPy

```
v = list(range(1, 6))
```

```
v
```

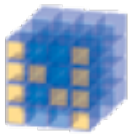
```
2 * v
```

```
import numpy as np
```

```
v = np.arange(1, 6)
```

```
v
```

```
2 * v
```



NumPy

Base
N-dimensional
array package

```
1 v = list(range(1, 6))  
2 v
```

```
[1, 2, 3, 4, 5]
```

```
1 2 * v
```

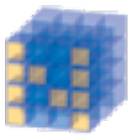
```
[1, 2, 3, 4, 5, 1, 2, 3, 4, 5]
```

```
1 import numpy as np  
2 v = np.arange(1, 6)  
3 v
```

```
array([1, 2, 3, 4, 5])
```

```
1 2 * v
```

```
array([ 2,  4,  6,  8, 10])
```



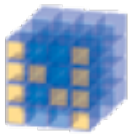
NumPy

NumPy Create Array

```
import numpy as np
a = np.array([1, 2, 3])
b = np.array([4, 5, 6])
c = a * b
c
```

```
import numpy as np
a = np.array([1, 2, 3])
b = np.array([4, 5, 6])
c = a * b
c
```

```
array([ 4, 10, 18])
```

NumPy

NumPy

```
1 import numpy as np
2
3 a = np.zeros((2,2)) # Create an array of all zeros
4 print(a)           # Prints "[[ 0.  0.]
5                     #                [ 0.  0.]]"
6
7 b = np.ones((1,2)) # Create an array of all ones
8 print(b)           # Prints "[[ 1.  1.]]"
9
10 c = np.full((2,2), 7) # Create a constant array
11 print(c)              # Prints "[[ 7.  7.]
12                       #                [ 7.  7.]]"
13
14 d = np.eye(2)         # Create a 2x2 identity matrix
15 print(d)             # Prints "[[ 1.  0.]
16                       #                [ 0.  1.]]"
17
18 e = np.random.random((2,2)) # Create an array filled with random values
19 print(e)               # Might print "[[ 0.91940167  0.08143941]
20                       #                [ 0.68744134  0.87236687]]"
```

```
[[0. 0.]
 [0. 0.]]
[[1. 1.]]
[[7 7]
 [7 7]]
[[1. 0.]
 [0. 1.]]
[[0.66258211 0.65552598]
 [0.00429934 0.21695824]]
```

Numpy Quickstart Tutorial

[SciPy.org](#)[Docs](#)[NumPy v1.13.dev0 Manual](#)[NumPy User Guide](#)[index](#)[next](#)[previous](#)

Quickstart tutorial

Prerequisites

Before reading this tutorial you should know a bit of Python. If you would like to refresh your memory, take a look at the [Python tutorial](#).

If you wish to work the examples in this tutorial, you must also have some software installed on your computer. Please see <http://scipy.org/install.html> for instructions.

The Basics

NumPy's main object is the homogeneous multidimensional array. It is a table of elements (usually numbers), all of the same type, indexed by a tuple of positive integers. In NumPy dimensions are called *axes*. The number of axes is *rank*.

For example, the coordinates of a point in 3D space `[1, 2, 1]` is an array of rank 1, because it has one axis. That axis has a length of 3. In the example pictured below, the array has rank 2 (it is 2-dimensional). The first dimension (axis) has a length of 2, the second dimension has a length of 3.

```
[[ 1., 0., 0.],  
 [ 0., 1., 2.]]
```

NumPy's array class is called `ndarray`. It is also known by the alias `array`. Note that `numpy.array` is not the same as the Standard Python Library class `array.array`, which only handles one-dimensional arrays and offers less functionality. The more important attributes of an `ndarray` object are:

`ndarray.ndim`

the number of axes (dimensions) of the array. In the Python world, the number of dimensions is referred to as *rank*.

`ndarray.shape`

Table Of Contents

- Quickstart tutorial
 - Prerequisites
 - The Basics
 - An example
 - Array Creation
 - Printing Arrays
 - Basic Operations
 - Universal Functions
 - Indexing, Slicing and Iterating
 - Shape Manipulation
 - Changing the shape of an array
 - Stacking together different arrays
 - Splitting one array into several smaller ones
 - Copies and Views
 - No Copy at All
 - View or Shallow Copy
 - Deep Copy
 - Functions and Methods Overview
 - Less Basic
 - Broadcasting rules
 - Fancy indexing and index tricks
 - Indexing with Arrays of

```
import numpy as np
```

```
a = np.arange(15).reshape(3, 5)
```

```
a.shape
```

```
a.ndim
```

```
a.dtype.name
```

```
import numpy as np  
a = np.arange(15).reshape(3, 5)  
a
```

```
array([[ 0,  1,  2,  3,  4],  
       [ 5,  6,  7,  8,  9],  
       [10, 11, 12, 13, 14]])
```

```
print(a.shape)
```

```
(3, 5)
```

```
a.ndim
```

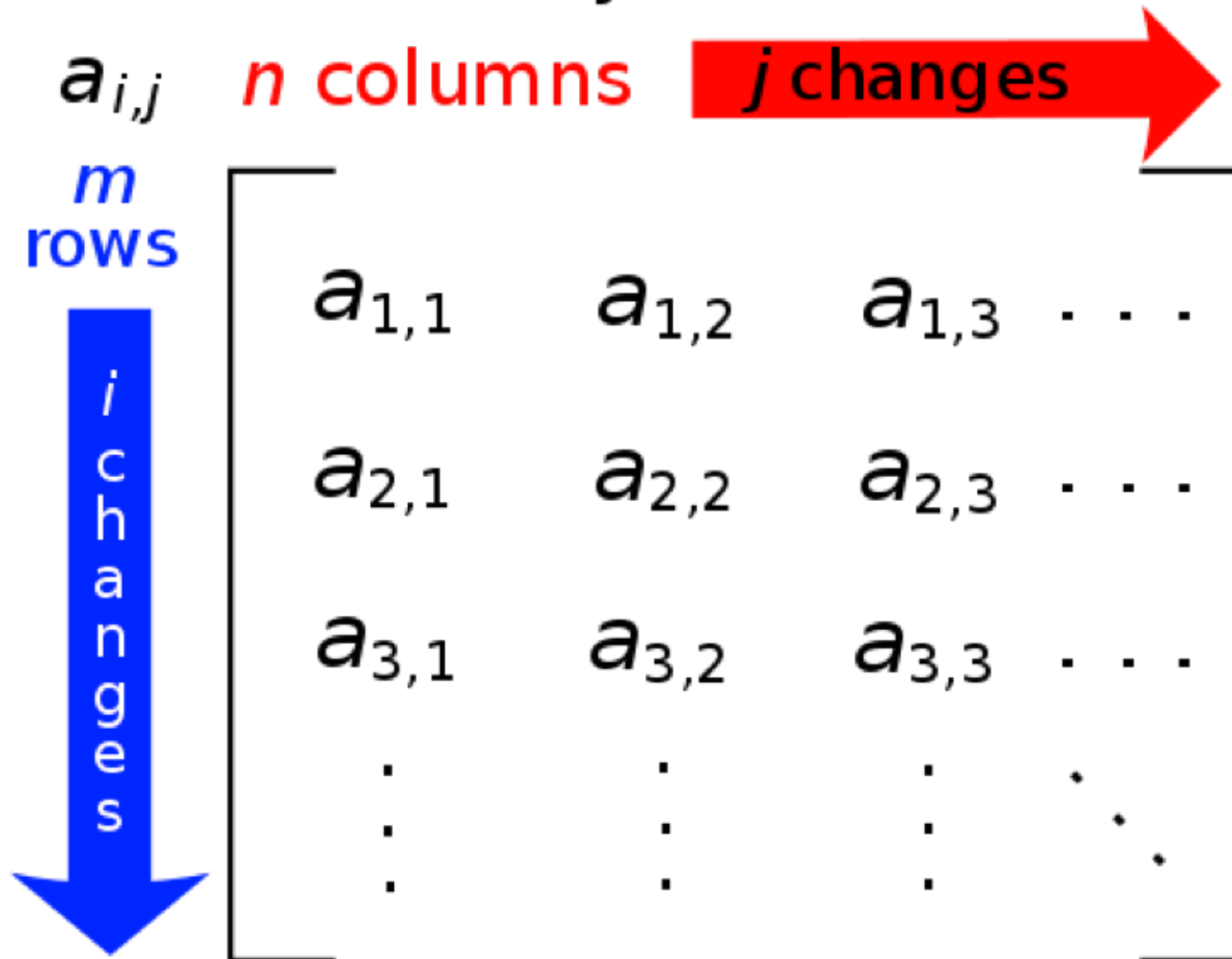
```
2
```

```
a.dtype.name
```

```
'int64'
```

Matrix

m -by- n matrix



NumPy ndarray: Multidimensional Array Object

NumPy ndarray

One-dimensional Array (1-D Array)

0	1			n-1
1	2	3	4	5

Two-dimensional Array (2-D Array)

	0	1			n-1
0	1	2	3	4	5
1	6	7	8	9	10
	11	12	13	14	15
m-1	16	17	18	19	20

```
import numpy as np
a = np.array([1,2,3,4,5])
```

One-dimensional Array (1-D Array)

0	1			n-1
1	2	3	4	5

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

```
array([1, 2, 3, 4, 5])
```

```
a = np.array( [ [1,2,3,4,5],[6,7,8,9,10],[11,12,13,14,15],[16,17,18,19,20] ] )
```

Two-dimensional Array (2-D Array)

	0	1			n-1
0	1	2	3	4	5
1	6	7	8	9	10
	11	12	13	14	15
m-1	16	17	18	19	20

```
a = np.array([[1,2,3,4,5],[6,7,8,9,10],[11,12,13,14,15],[16,17,18,19,20]])  
a
```

```
array([[ 1,  2,  3,  4,  5],  
       [ 6,  7,  8,  9, 10],  
       [11, 12, 13, 14, 15],  
       [16, 17, 18, 19, 20]])
```



```
import numpy as np
a = np.array([[0, 1, 2, 3],
              [10, 11, 12, 13],
              [20, 21, 22, 23]])
a
```

0	1	2	3
10	11	12	13
20	21	22	23

```
a = np.array ([[0, 1, 2, 3], [10, 11, 12, 13], [20, 21, 22, 23]])
```

```
a = np.array([[0, 1, 2, 3], [10, 11, 12, 13], [20, 21, 22, 23]])  
a
```

```
array([[ 0,  1,  2,  3],  
       [10, 11, 12, 13],  
       [20, 21, 22, 23]])
```

```
print(a.ndim)
```

```
2
```

```
print(a.shape)
```

```
(3, 4)
```

0	1	2	3
10	11	12	13
20	21	22	23

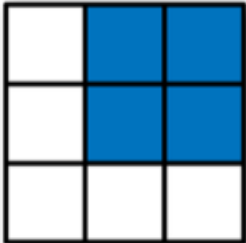
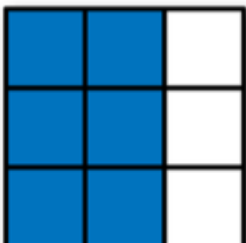
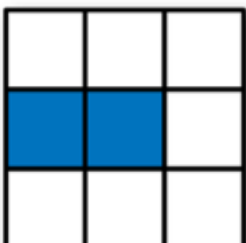
NumPy Basics: Arrays and Vectorized Computation

NumPy Array

axis 1

		0	1	2
axis 0	0	0, 0	0, 1	0, 2
	1	1, 0	1, 1	1, 2
	2	2, 0	2, 1	2, 2

Numpy Array

	Expression	Shape
	<code>arr[:2, 1:]</code>	<code>(2, 2)</code>
	<code>arr[2]</code> <code>arr[2, :]</code> <code>arr[2:, :]</code>	<code>(3,)</code> <code>(3,)</code> <code>(1, 3)</code>
	<code>arr[:, :2]</code>	<code>(3, 2)</code>
	<code>arr[1, :2]</code> <code>arr[1:2, :2]</code>	<code>(2,)</code> <code>(1, 2)</code>

Wes McKinney (2017), "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython", 2nd Edition, O'Reilly Media.

Materials and IPython notebooks for "Python for Data Analysis" by Wes McKinney, published by O'Reilly Media

52 commits

2 branches

0 releases

6 contributors

Branch: 2nd-edition

New pull request

Find file

Clone or download

 betatim committed with wesm Add requirements (#71)

Latest commit ea47998 5 days ago

 datasets	Add Kaggle titanic dataset	5 months ago
 examples	Remove sex column from tips dataset	4 months ago
 .gitignore	Add gitignore	2 years ago
 COPYING	Use MIT license for code examples	a month ago
 README.md	Add launch in Azure Notebooks button (#70)	19 days ago
 appa.ipynb	Make more cells markdown instead of raw	a month ago
 ch02.ipynb	Make more cells markdown instead of raw	a month ago
 ch03.ipynb	Make more cells markdown instead of raw	a month ago
 ch04.ipynb	Convert all notebooks to v4 format	a month ago
 ch05.ipynb	Make more cells markdown instead of raw	a month ago
 ch06.ipynb	Make more cells markdown instead of raw	a month ago
 ch07.ipynb	Convert all notebooks to v4 format	a month ago
 ch08.ipynb	Make more cells markdown instead of raw	a month ago
 ch09.ipynb	Make more cells markdown instead of raw	a month ago
 ch10.ipynb	Make more cells markdown instead of raw	a month ago

<https://github.com/wesm/pydata-book>

Wes McKinney (2017), "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython", 2nd Edition, O'Reilly Media.

wesm / pydata-book

Watch 640 Star 4,031 Fork 4,348

<> Code Issues 2 Pull requests 0 Projects 0 Insights

Branch: 2nd-edition ▾ pydata-book / ch04.ipynb Find file Copy path

wesm Convert all notebooks to v4 format c2780a0 on Sep 27

2 contributors

1857 lines (1856 sloc) 32.6 KB Raw Blame History

NumPy Basics: Arrays and

```
In [ ]: import numpy as np
np.random.seed(12345)
import matplotlib.pyplot as plt
plt.rc('figure', figsize=(10, 6))
np.set_printoptions(precision=4, suppress=True)
```

```
In [ ]: import numpy as np
my_arr = np.arange(1000000)
my_list = list(range(1000000))
```

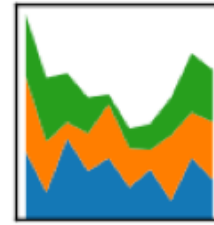
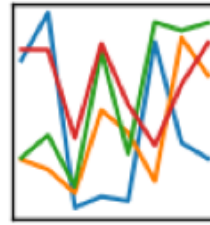
```
In [ ]: %time for _ in range(10): my_arr2 = my_arr * 2
%time for _ in range(10): my_list2 = [x * 2 for x in my_list]
```

The NumPy ndarray: A Multidimensional Array Object

```
In [ ]: import numpy as np
# Generate some random data
data = np.random.randn(2, 3)
data
```

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



Python

Pandas

Python Pandas for Finance

! pip install pandas_datareader

```
1 ! pip install pandas_datareader
```

Collecting pandas_datareader

Downloading https://files.pythonhosted.org/packages/cc/5c/ea5b6dcfd0f55c5fb1e37fb45335ec01ccec199b8a79339137f5ed269e0/pandas_datareader-0.7.0-py3-none-any.whl
100% |██| 112kB 2.7MB/s

Collecting lxml (from pandas_datareader)

Downloading https://files.pythonhosted.org/packages/03/a4/9eea8035fc7c7670e5eab97f34ff2ef0ddd78a491bf96df5accdb0e63f5/lxml-4.2.5-cp36-cp37m-macosx_10_10_x86_64.whl
100% |██| 5.8MB 7.5MB/s

Requirement already satisfied: pandas>=0.19.2 in /usr/local/lib/python3.6/dist-packages (from pandas_datareader) (0.22.0)
Requirement already satisfied: requests>=2.3.0 in /usr/local/lib/python3.6/dist-packages (from pandas_datareader) (2.18.4)
Requirement already satisfied: wrapt in /usr/local/lib/python3.6/dist-packages (from pandas_datareader) (1.10.11)
Requirement already satisfied: python-dateutil>=2 in /usr/local/lib/python3.6/dist-packages (from pandas>=0.19.2->pandas_datareader) (2.6.0)
Requirement already satisfied: numpy>=1.9.0 in /usr/local/lib/python3.6/dist-packages (from pandas>=0.19.2->pandas_datareader) (1.14.6)
Requirement already satisfied: pytz>=2011k in /usr/local/lib/python3.6/dist-packages (from pandas>=0.19.2->pandas_datareader) (2018.5)
Requirement already satisfied: idna<2.7,>=2.5 in /usr/local/lib/python3.6/dist-packages (from requests>=2.3.0->pandas_datareader) (2.6)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in /usr/local/lib/python3.6/dist-packages (from requests>=2.3.0->pandas_datareader) (3.0.4)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.6/dist-packages (from requests>=2.3.0->pandas_datareader) (2018.8.24)
Requirement already satisfied: urllib3<1.23,>=1.21.1 in /usr/local/lib/python3.6/dist-packages (from requests>=2.3.0->pandas_datareader) (1.24.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.6/dist-packages (from python-dateutil>=2->pandas>=0.19.2->pandas_datareader) (1.11.0)
Installing collected packages: lxml, pandas-datareader
Successfully installed lxml-4.2.5 pandas-datareader-0.7.0

conda install pandas-datareader

```
imyday — -bash — 80x24
[iMyday-MacBook-Pro:~ imyday$ conda install pandas-datareader
]
Fetching package metadata .....
Solving package specifications: .

Package plan for installation in environment /Users/imyday/anaconda:

The following NEW packages will be INSTALLED:

    pandas-datareader: 0.2.1-py36_0
    requests-file:     1.4.1-py36_0

Proceed ([y]/n)? y

requests-file- 100% |#####| Time: 0:00:00 1.55 MB/s
pandas-datarea 100% |#####| Time: 0:00:00 409.66 kB/s
[iMyday-MacBook-Pro:~ imyday$ conda list
]
# packages in environment at /Users/imyday/anaconda:
#
_license          1.1                py36_1
alabaster          0.7.9              py36_0
anaconda           4.3.1              np111py36_0
anaconda-client    1.6.0              py36_0
anaconda-navigator 1.5.0              py36_0
anaconda-project   0.4.1              py36_0
```

AAPL



Finance Data from Yahoo Finance

```
# !pip install pandas_datareader
import pandas_datareader.data as web
import datetime as dt
#Read Stock Data from Yahoo Finance
end = dt.datetime(2017, 12, 31)
start = dt.datetime(2016, 1, 1)
df = web.DataReader("AAPL", 'yahoo', start, end)
df.to_csv('AAPL.csv')
df.from_csv('AAPL.csv')
df.tail()
```

```
# !pip install pandas_datareader
import pandas as pd
import pandas_datareader.data as web
import matplotlib.pyplot as plt
import seaborn as sns
import datetime as dt
%matplotlib inline

#Read Stock Data from Yahoo Finance
end = dt.datetime.now()
#start = dt.datetime(end.year-2, end.month,
end.day)
start = dt.datetime(2016, 1, 1)
df = web.DataReader("AAPL", 'yahoo', start, end)
df.to_csv('AAPL.csv')
df.from_csv('AAPL.csv')
df.tail()
```

```
df['Adj Close'].plot(legend=True, figsize=(12, 8), title='AAPL', label='Adj Close')
```



```
plt.figure(figsize=(12,9))
top = plt.subplot2grid((12,9), (0, 0),
rowspan=10, colspan=9)
bottom = plt.subplot2grid((12,9), (10,0),
rowspan=2, colspan=9)
top.plot(df.index, df['Adj Close'],
color='blue') #df.index gives the dates
bottom.bar(df.index, df['Volume'])
```

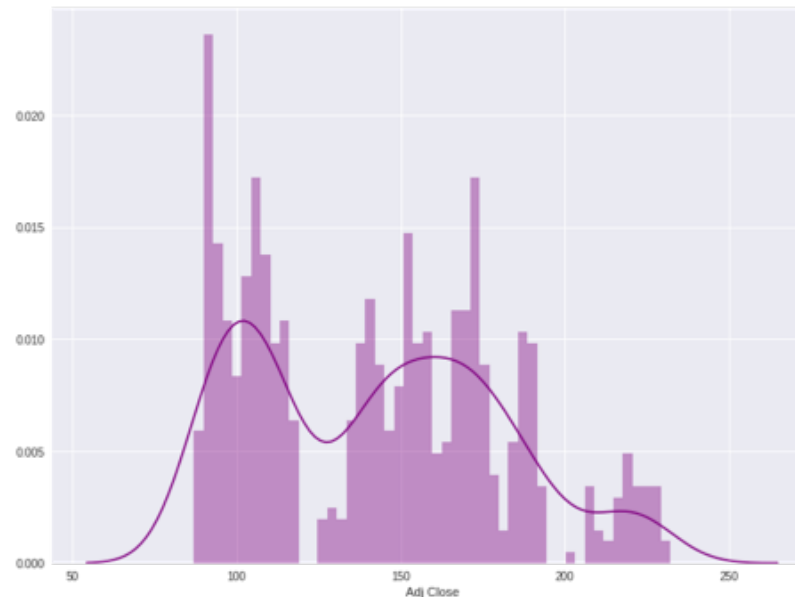


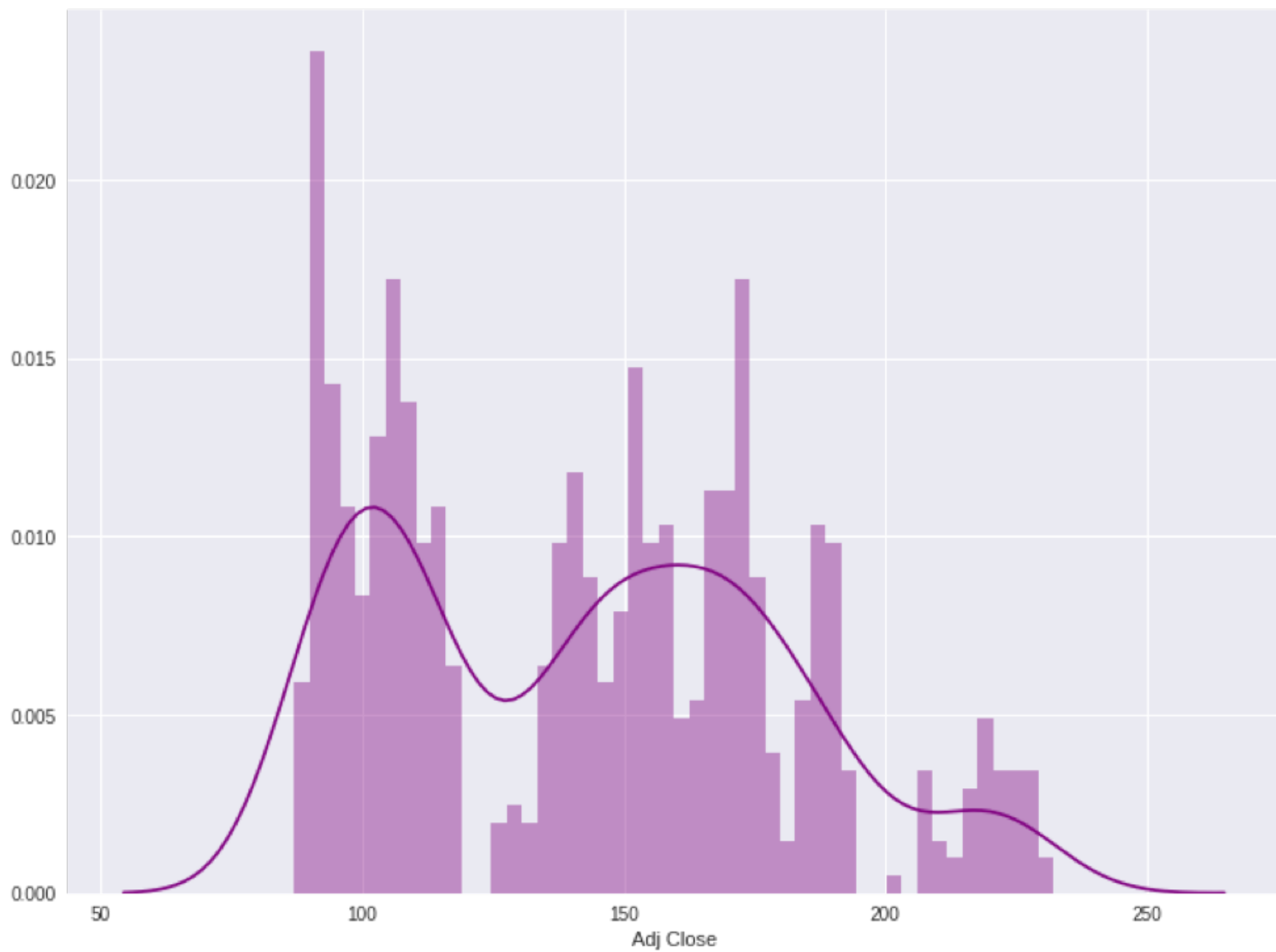
AAPL



```
# set the labels
top.axes.get_xaxis().set_visible(False)
top.set_title('AAPL')
top.set_ylabel('Adj Close')
bottom.set_ylabel('Volume')

plt.figure(figsize=(12,9))
sns.distplot(df['Adj Close'].dropna(),
bins=50, color='purple')
```





```
# simple moving averages
df['MA05'] = df['Adj Close'].rolling(5).mean()
#5 days
df['MA20'] = df['Adj
Close'].rolling(20).mean() #20 days
df['MA60'] = df['Adj
Close'].rolling(60).mean() #60 days
df2 = pd.DataFrame({'Adj Close': df['Adj
Close'], 'MA05': df['MA05'], 'MA20': df['MA20'],
'MA60': df['MA60']})
df2.plot(figsize=(12, 9), legend=True,
title='AAPL')
df2.to_csv('AAPL_MA.csv')
fig = plt.gcf()
fig.set_size_inches(12, 9)
fig.savefig('AAPL_plot.png', dpi=300)
plt.show()
```

AAPL



```

# !pip install pandas_datareader
import pandas as pd
import pandas_datareader.data as web
import matplotlib.pyplot as plt
import seaborn as sns
import datetime as dt
%matplotlib inline

#Read Stock Data from Yahoo Finance
end = dt.datetime.now()
#start = dt.datetime(end.year-2, end.month, end.day)
start = dt.datetime(2016, 1, 1)
df = web.DataReader("AAPL", 'yahoo', start, end)
df.to_csv('AAPL.csv')
df.from_csv('AAPL.csv')
df.tail()

df['Adj Close'].plot(legend=True, figsize=(12, 8), title='AAPL', label='Adj Close')
plt.figure(figsize=(12,9))
top = plt.subplot2grid((12,9), (0, 0), rowspan=10, colspan=9)
bottom = plt.subplot2grid((12,9), (10,0), rowspan=2, colspan=9)
top.plot(df.index, df['Adj Close'], color='blue') #df.index gives the dates
bottom.bar(df.index, df['Volume'])

# set the labels
top.axes.get_xaxis().set_visible(False)
top.set_title('AAPL')
top.set_ylabel('Adj Close')
bottom.set_ylabel('Volume')

plt.figure(figsize=(12,9))
sns.distplot(df['Adj Close'].dropna(), bins=50, color='purple')

# simple moving averages
df['MA05'] = df['Adj Close'].rolling(5).mean() #5 days
df['MA20'] = df['Adj Close'].rolling(20).mean() #20 days
df['MA60'] = df['Adj Close'].rolling(60).mean() #60 days
df2 = pd.DataFrame({'Adj Close': df['Adj Close'], 'MA05': df['MA05'], 'MA20': df['MA20'], 'MA60': df['MA60']})
df2.plot(figsize=(12, 9), legend=True, title='AAPL')
df2.to_csv('AAPL_MA.csv')
fig = plt.gcf()
fig.set_size_inches(12, 9)
fig.savefig('AAPL_plot.png', dpi=300)
plt.show()

```

```

1 # !pip install pandas_datareader
2 import pandas as pd
3 import pandas_datareader.data as web
4 import matplotlib.pyplot as plt
5 import seaborn as sns
6 import datetime as dt
7 %matplotlib inline
8
9 #Read Stock Data from Yahoo Finance
10 end = dt.datetime.now()
11 #start = dt.datetime(end.year-2, end.month, end.day)
12 start = dt.datetime(2016, 1, 1)
13 df = web.DataReader("AAPL", 'yahoo', start, end)
14 df.to_csv('AAPL.csv')
15 df.from_csv('AAPL.csv')
16 df.tail()
17
18 df['Adj Close'].plot(legend=True, figsize=(12, 8), title='AAPL', label='Adj Close')
19 plt.figure(figsize=(12,9))
20 top = plt.subplot2grid((12,9), (0, 0), rowspan=10, colspan=9)
21 bottom = plt.subplot2grid((12,9), (10,0), rowspan=2, colspan=9)
22 top.plot(df.index, df['Adj Close'], color='blue') #df.index gives the dates
23 bottom.bar(df.index, df['Volume'])
24
25 # set the labels
26 top.axes.get_xaxis().set_visible(False)
27 top.set_title('AAPL')
28 top.set_ylabel('Adj Close')
29 bottom.set_ylabel('Volume')
30
31 plt.figure(figsize=(12,9))
32 sns.distplot(df['Adj Close'].dropna(), bins=50, color='purple')
33
34 # simple moving averages
35 df['MA05'] = df['Adj Close'].rolling(5).mean() #5 days
36 df['MA20'] = df['Adj Close'].rolling(20).mean() #20 days
37 df['MA60'] = df['Adj Close'].rolling(60).mean() #60 days
38 df2 = pd.DataFrame({'Adj Close': df['Adj Close'], 'MA05': df['MA05'], 'MA20': df['MA20'], 'MA60': df['MA60']})
39 df2.plot(figsize=(12, 9), legend=True, title='AAPL')
40 df2.to_csv('AAPL_MA.csv')
41 fig = plt.gcf()
42 fig.set_size_inches(12, 9)
43 fig.savefig('AAPL_plot.png', dpi=300)
44 plt.show()

```

Finance Data from Quandl

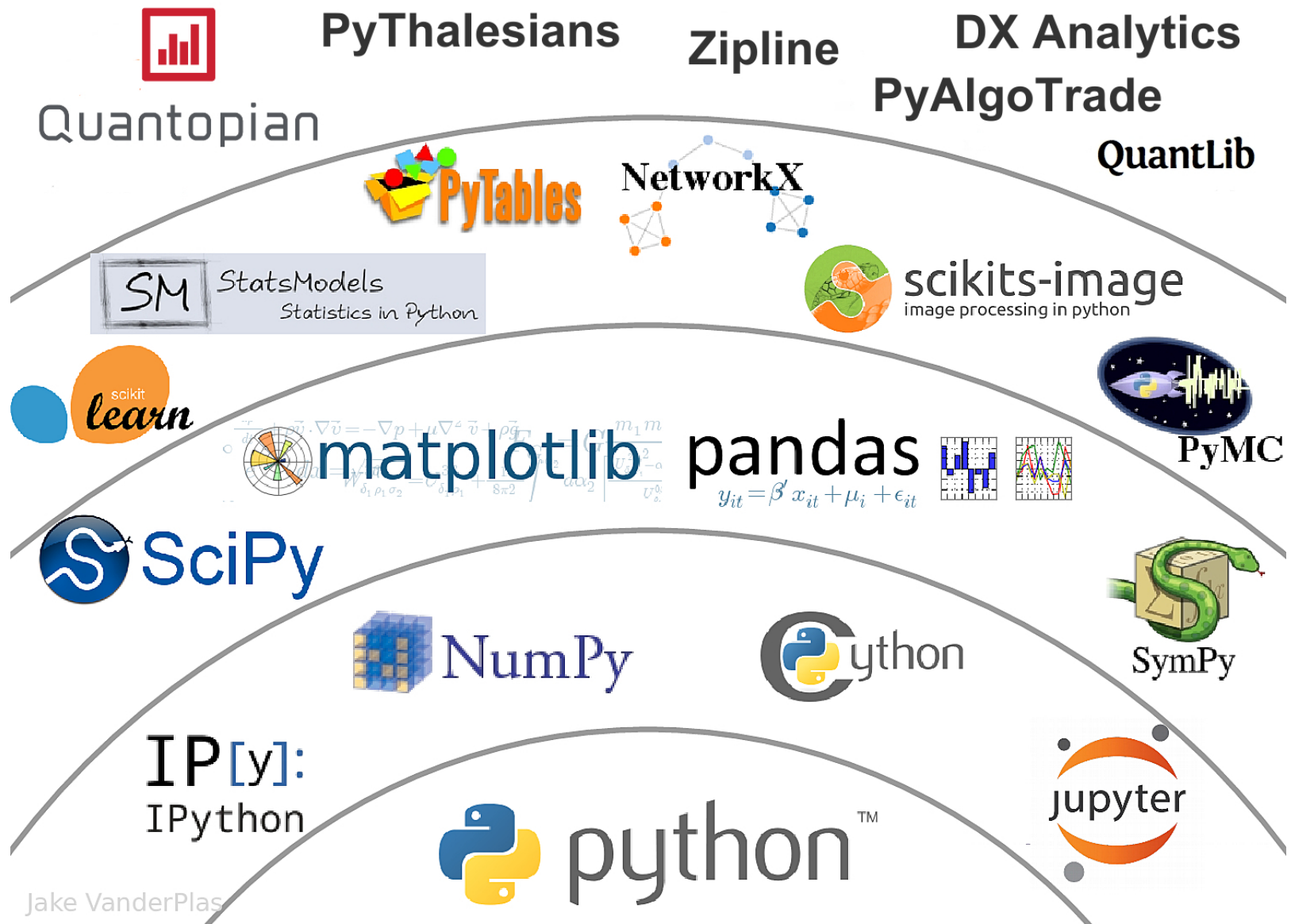
```
# ! pip install quandl
import quandl
# quandl.ApiConfig.api_key = "YOURAPIKEY"
df = quandl.get("WIKI/AAPL", start_date="2016-01-01", end_date="2017-12-31")
df.to_csv('AAPL.csv')
df.from_csv('AAPL.csv')
df.tail()
```

```
1 # ! pip install quandl
2 import quandl
3 # quandl.ApiConfig.api_key = "YOURAPIKEY"
4 df = quandl.get("WIKI/AAPL", start_date="2016-01-01", end_date="2017-12-31")
5 df.to_csv('AAPL.csv')
6 df.from_csv('AAPL.csv')
7 df.tail()
```

/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:5: FutureWarning: from_csv is deprecated. Please use read_csv(...) instead
"""

	Open	High	Low	Close	Volume	Ex-Dividend	Split Ratio	Adj. Open	Adj. High	Adj. Low	Adj. Close	Adj. Volume
Date												
2017-12-22	174.68	175.424	174.500	175.01	16052615.0	0.0	1.0	174.68	175.424	174.500	175.01	16052615.0
2017-12-26	170.80	171.470	169.679	170.57	32968167.0	0.0	1.0	170.80	171.470	169.679	170.57	32968167.0
2017-12-27	170.10	170.780	169.710	170.60	21672062.0	0.0	1.0	170.10	170.780	169.710	170.60	21672062.0
2017-12-28	171.00	171.850	170.480	171.08	15997739.0	0.0	1.0	171.00	171.850	170.480	171.08	15997739.0
2017-12-29	170.52	170.590	169.220	169.23	25643711.0	0.0	1.0	170.52	170.590	169.220	169.23	25643711.0

The Quant Finance PyData Stack



Summary

- **Foundations of AI in Finance Big Data Analytics with Python**
 - **Python**
 - Programming language
 - **Numpy**
 - Scientific computing
 - **Pandas**
 - Data structures and data analysis tools

References

- Wes McKinney (2017), "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython", 2nd Edition, O'Reilly Media.
<https://github.com/wesm/pydata-book>
- Ties de Kok (2017), Learn Python for Research,
<https://github.com/TiesdeKok/LearnPythonforResearch>
- Avinash Jain (2017), Introduction To Python Programming, Udemy,
<https://www.udemy.com/pythonforbeginnersintro/>
- Python Programming, <https://pythonprogramming.net/>
- Python, <https://www.python.org/>
- Python Programming Language, <http://pythonprogramminglanguage.com/>
- Numpy, <http://www.numpy.org/>
- Pandas, <http://pandas.pydata.org/>
- Skikit-learn, <http://scikit-learn.org/>
- Data School (2015), Machine learning in Python with scikit-learn,
<https://www.youtube.com/playlist?list=PL5-da3qGB5ICeMbQuqbbCOQWcS6OYBr5A>
- Jason Brownlee (2016), Your First Machine Learning Project in Python Step-By-Step,
<https://machinelearningmastery.com/machine-learning-in-python-step-by-step/>