

Applied Data Analysis for Atmospheric Sciences Using Python

November 2018

Outline



Python



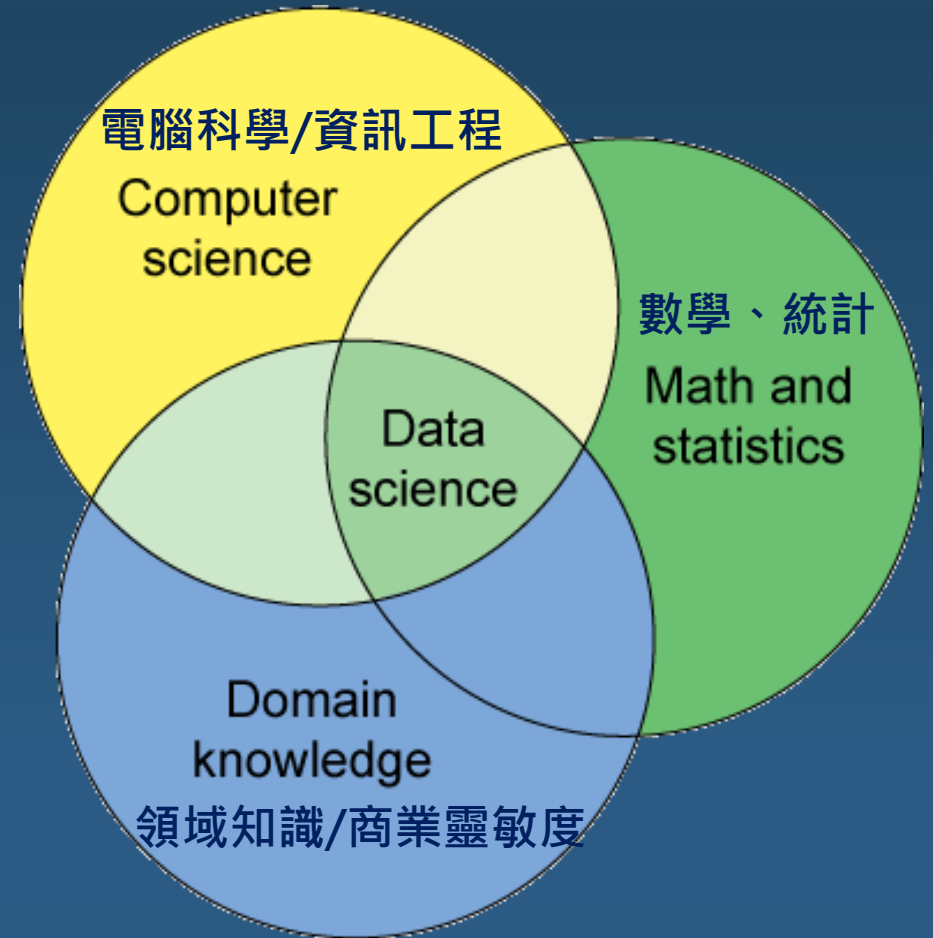
Data
Analysis



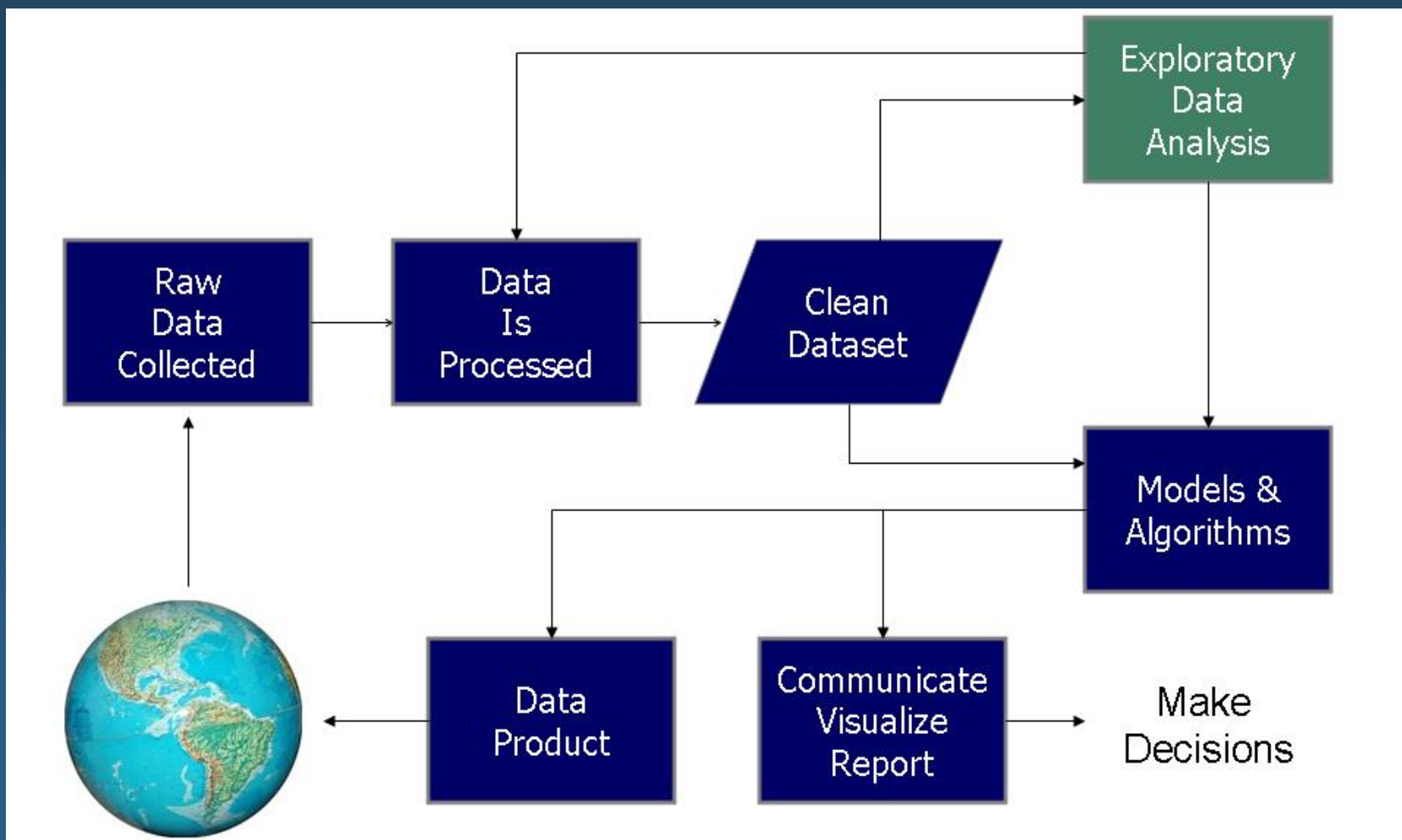
Atmospheric
Sciences

What is Data Science?

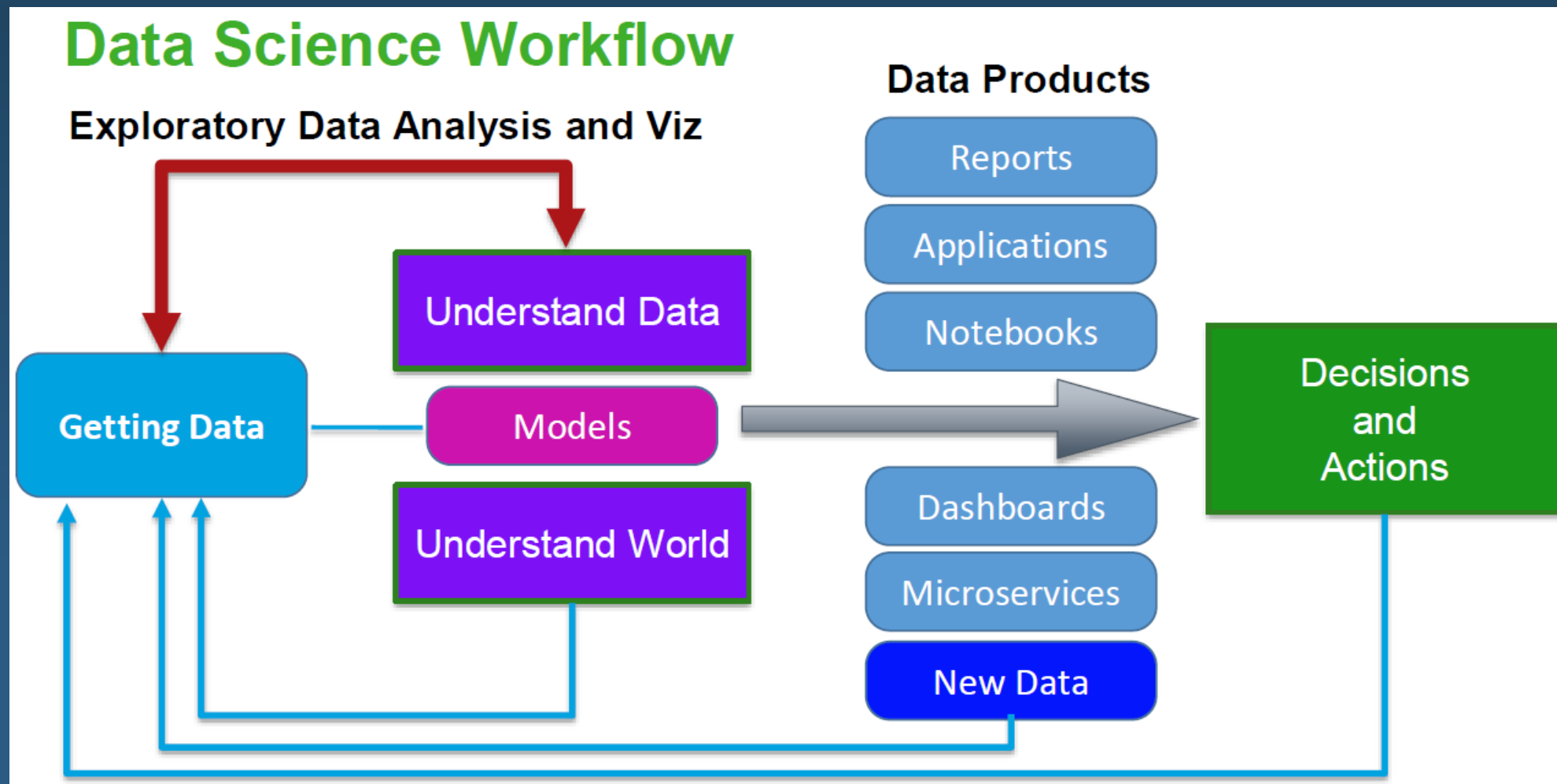
- It is not totally new
 - Statistics
 - Computer Science
 - Domain Knowledge
- It is not a new name for old stuff
- Data science is a “data-driven” approach for decision making



Data Science Process (資料科學流程)



Data Science Process (資料科學流程)



Outline

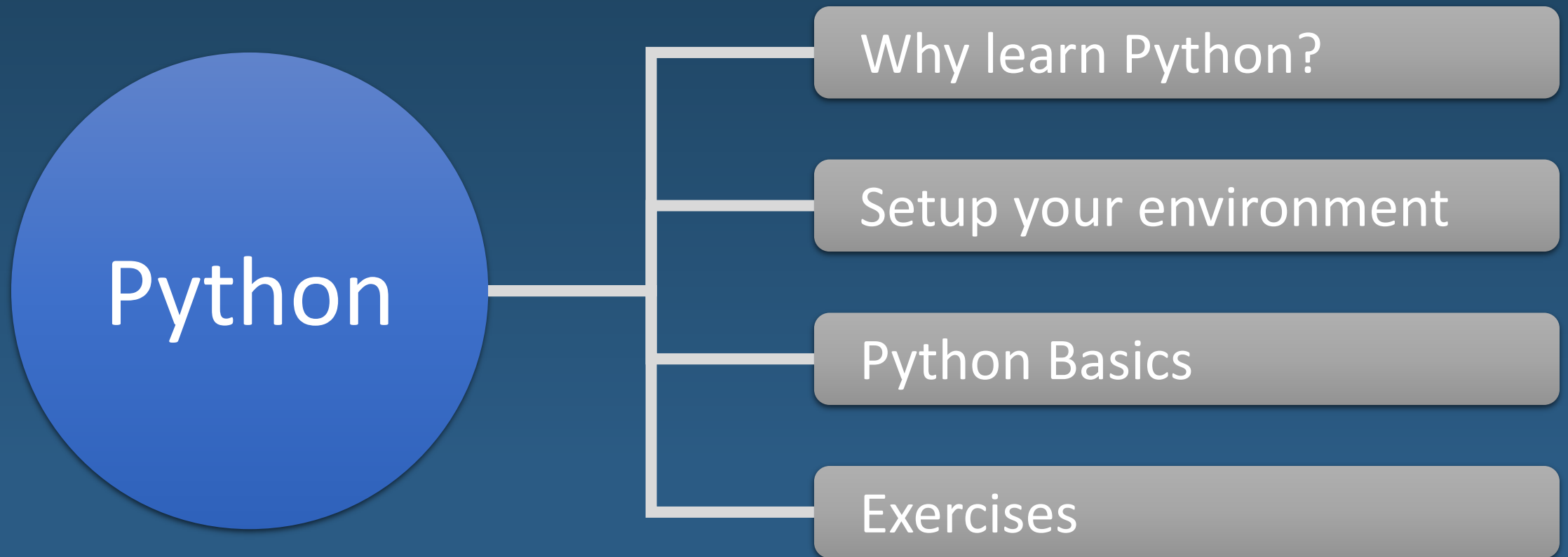


Python

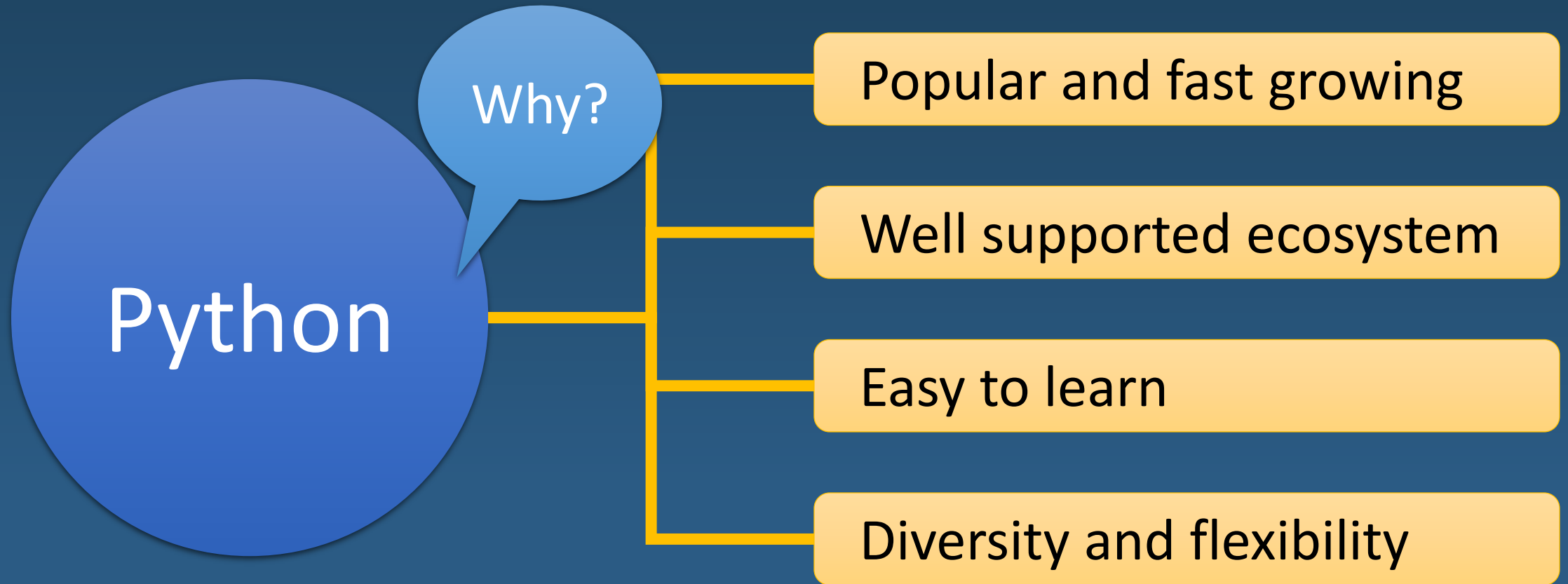
Data
Analysis

Atmospheric
Sciences

Introduction to Python



Why Python?



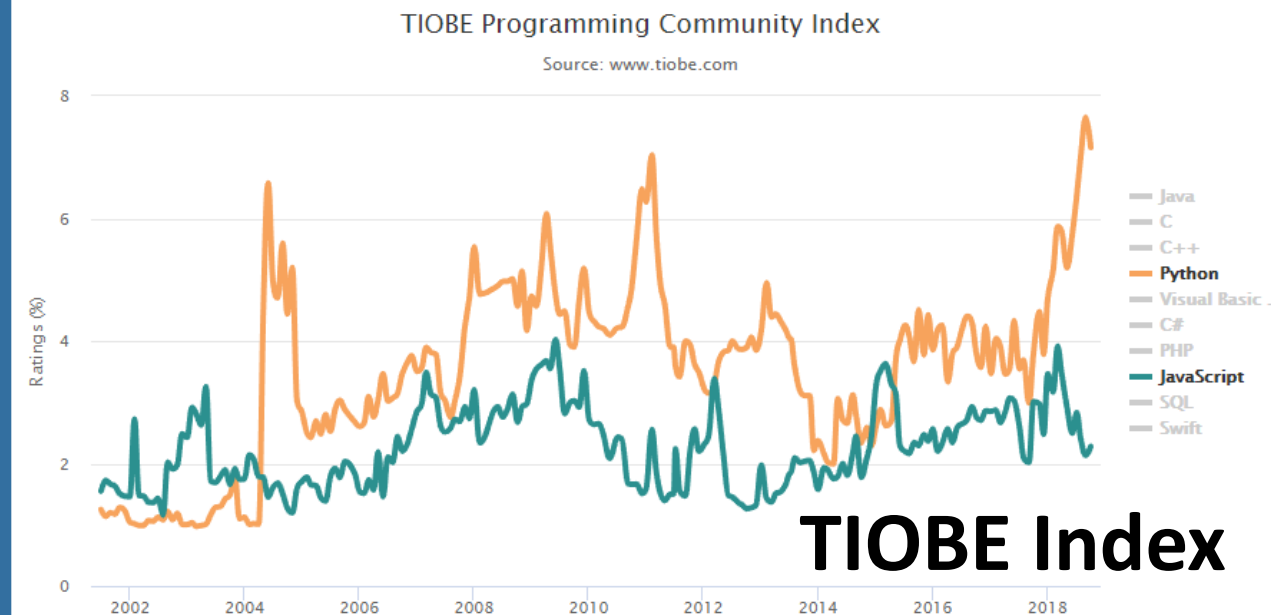
The 2018 Top Programming Languages

Language Rank	Types	Spectrum Ranking
1. Python	🌐 🖥️ 📱	100.0
2. C++	📱 🖥️ 📱	99.7
3. Java	🌐 📱 🖥️	97.5
4. C	📱 🖥️ 📱	96.7
5. C#	🌐 📱 🖥️	89.4
6. PHP	🌐	84.9
7. R	🖥️	82.9
8. JavaScript	🌐 📱	82.6
9. Go	🌐 🖥️	76.4
10. Assembly	📱	74.1

IEEE Spectrum

<https://spectrum.ieee.org/at-work/innovation/the-2018-top-programming-languages>

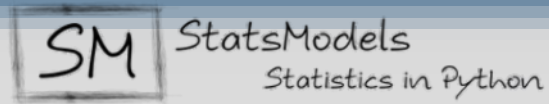
Oct 2018	Oct 2017	Change	Programming Language	Ratings	Change
1	1		Java	17.801%	+5.37%
2	2		C	15.376%	+7.00%
3	3		C++	7.593%	+2.59%
4	5	⬆️	Python	7.156%	+3.35%
5	8	⬆️	Visual Basic .NET	5.884%	+3.15%



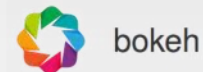
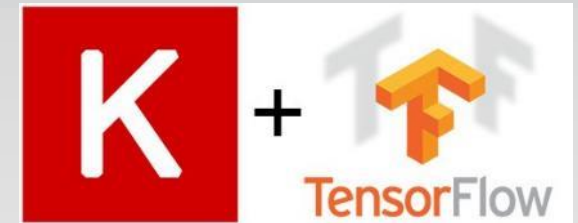
TIOBE Index

<https://www.tiobe.com/tiobe-index/>

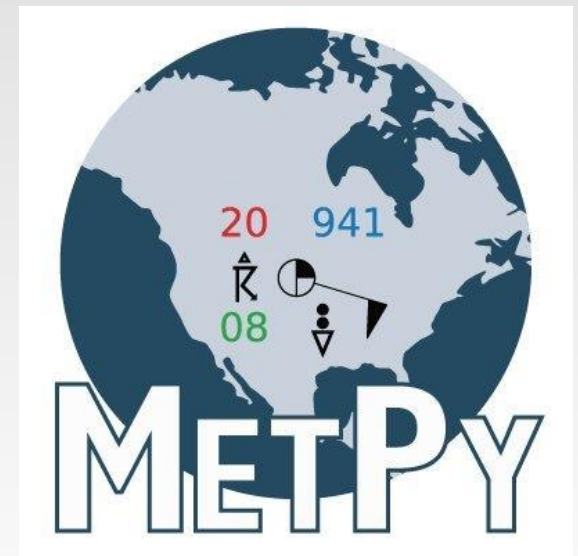
Python Ecosystem for Science



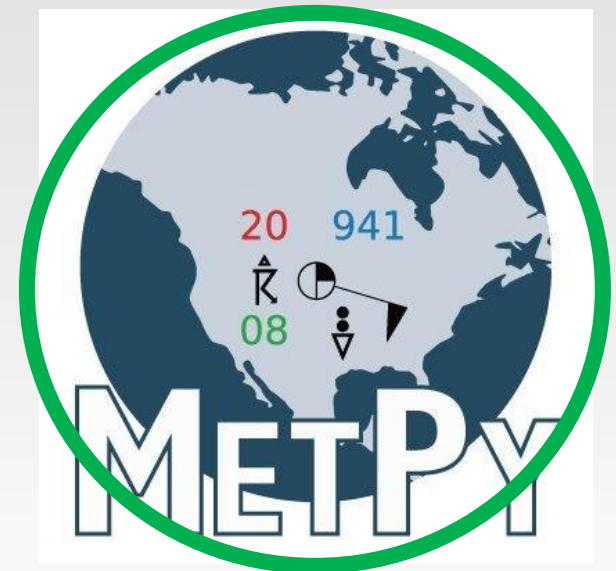
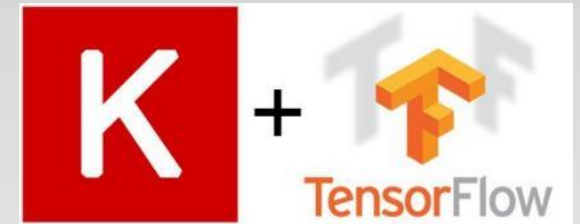
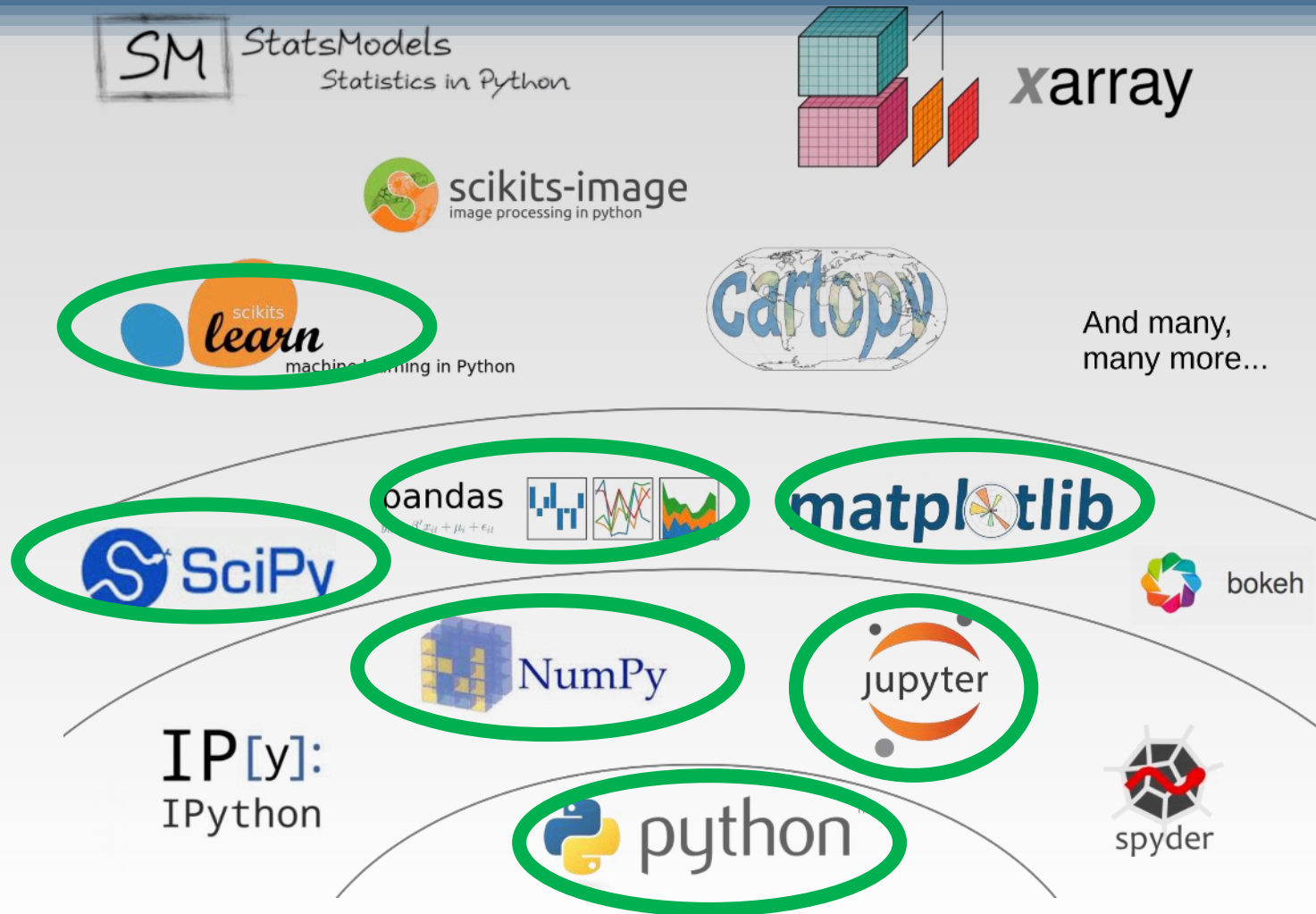
And many,
many more...



IP[y]:
IPython



Python Ecosystem for Science



Setup your Python Environment



<http://140.112.67.93/downloads/>



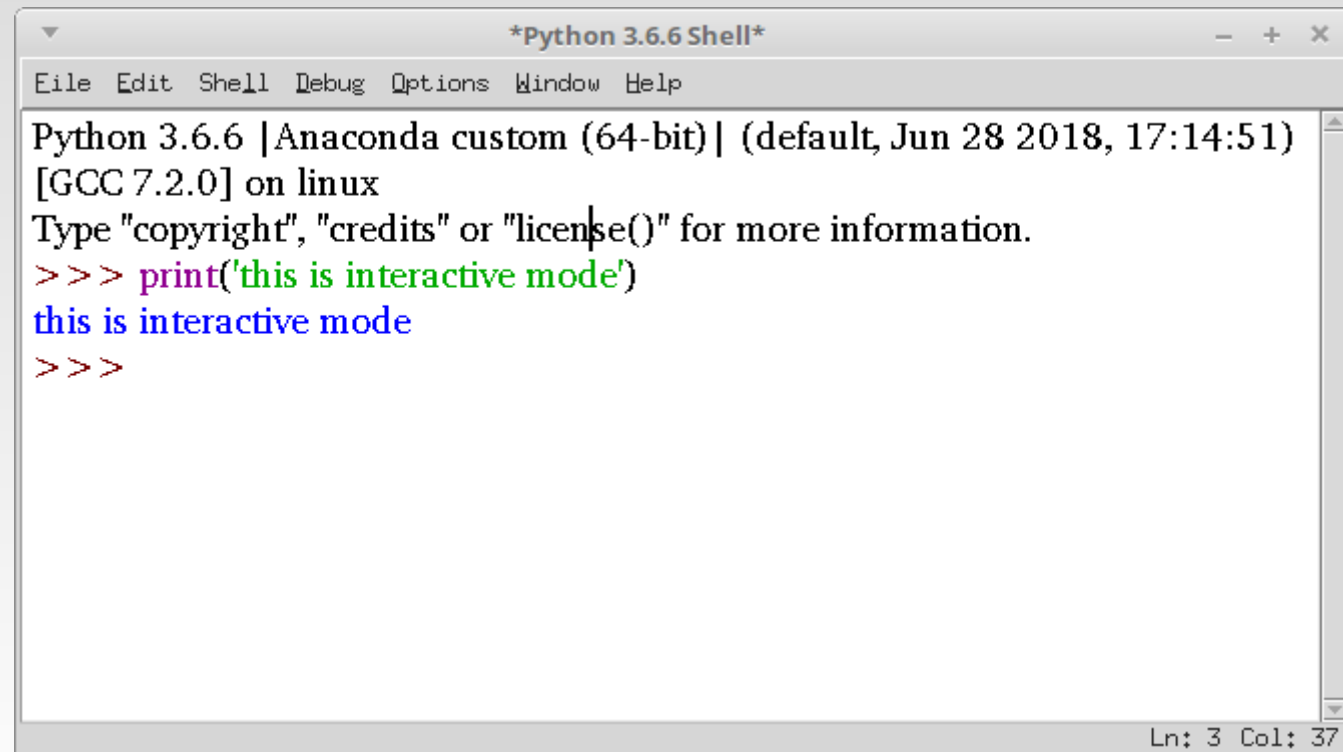
- Anaconda is a **free and open source** distribution of the **Python** ... for **data science** and machine learning ... to simplify **package_management** ...

NumPy	SciPy	Pandas	Scikit-learn	Jupyter/IPython	
Numba	Matplotlib	Spyder	TensorFlow	Cython	Theano
Scikit-image	NLTK	Dask	Caffe	dplyr	shiny
ggplot2	tidyr	caret	PySpark	& 1000+ packages	

CONDA®

Python in Interactive Mode

- Interactive mode (a.k.a. shell mode) is great for quickly and conveniently running single lines or blocks of code.



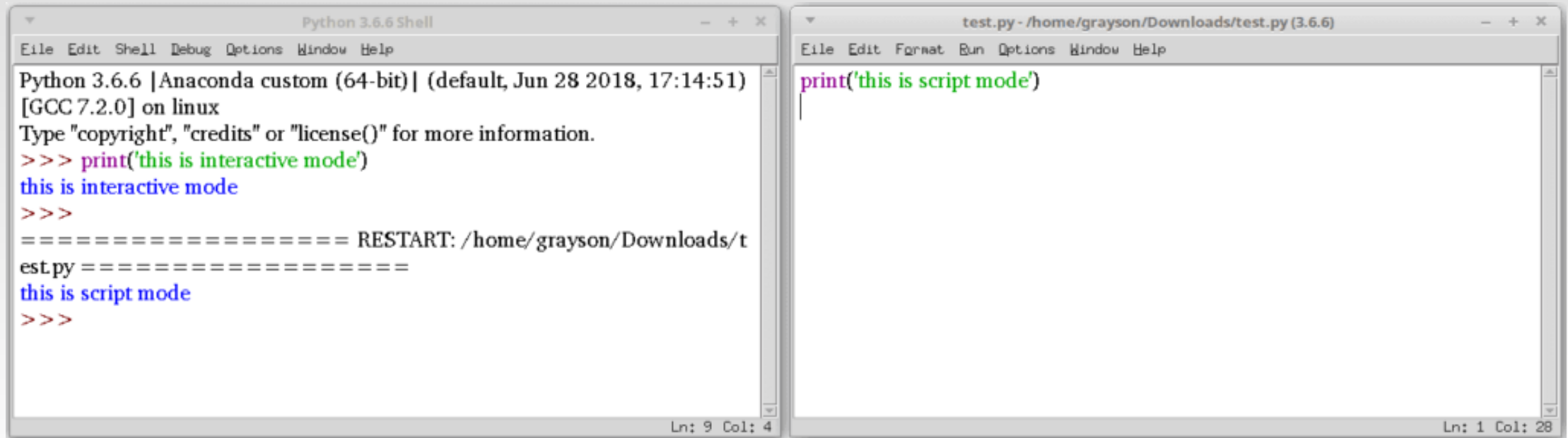
The screenshot shows a window titled '*Python 3.6.6 Shell*' with a menu bar containing 'File', 'Edit', 'Shell', 'Debug', 'Options', 'Window', and 'Help'. The main text area displays the following content:

```
Python 3.6.6 |Anaconda custom (64-bit)| (default, Jun 28 2018, 17:14:51)
[GCC 7.2.0] on linux
Type "copyright", "credits" or "license()" for more information.
>>> print('this is interactive mode')
this is interactive mode
>>>
```

The status bar at the bottom right indicates 'Ln: 3 Col: 37'.

Python in Script Mode

- Script mode (a.k.a. program mode) is suitable for working with more than a few lines of code, or you're ready to write an actual program.



```
Python 3.6.6 Shell
File Edit Shell Debug Options Window Help
Python 3.6.6 |Anaconda custom (64-bit)| (default, Jun 28 2018, 17:14:51)
[GCC 7.2.0] on linux
Type "copyright", "credits" or "license()" for more information.
>>> print('this is interactive mode')
this is interactive mode
>>>
===== RESTART: /home/grayson/Downloads/t
est.py =====
this is script mode
>>>
Ln: 9 Col: 4
```

```
test.py - /home/grayson/Downloads/test.py (3.6.6)
File Edit Format Run Options Window Help
print('this is script mode')
Ln: 1 Col: 28
```

Install Extra Packages from Anaconda

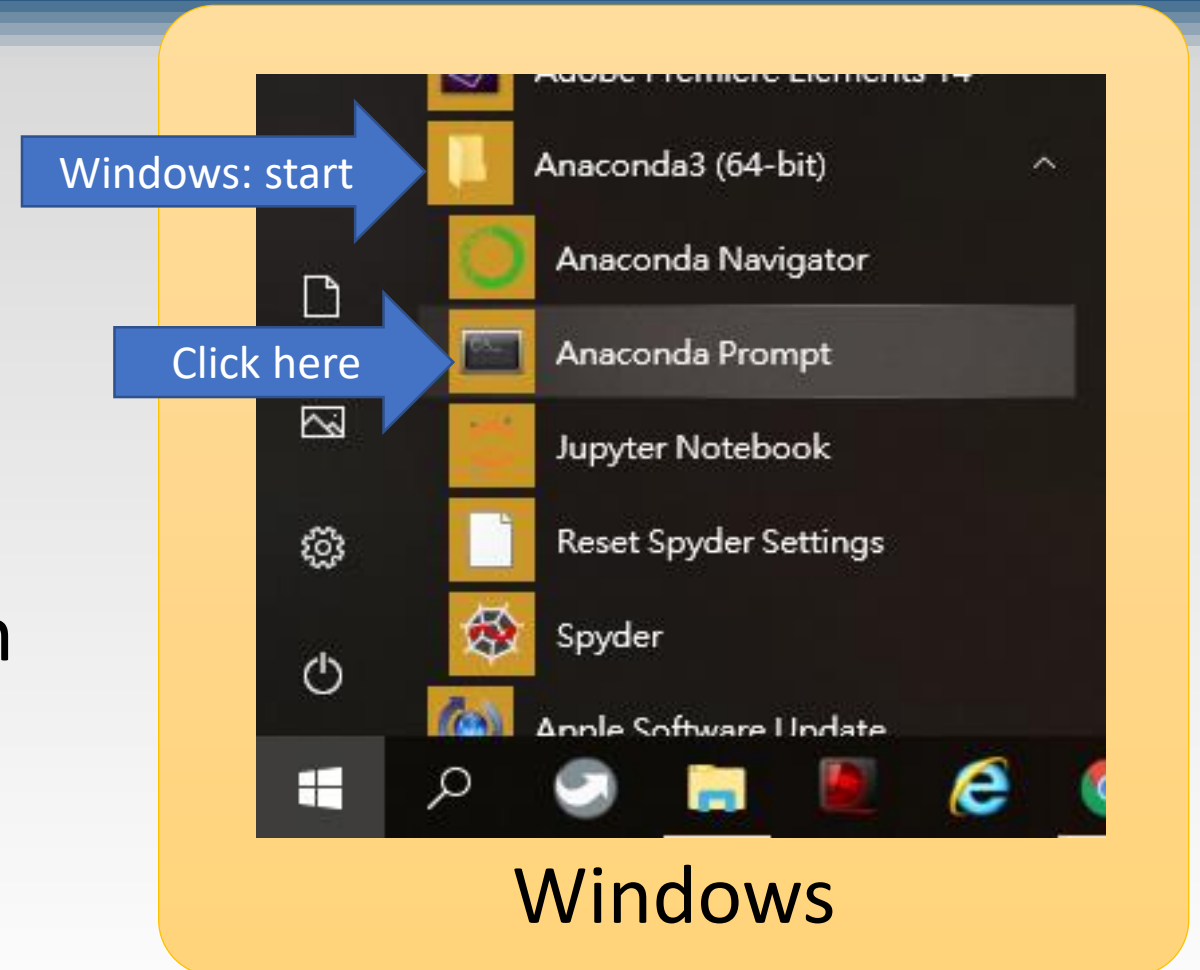
- Two options:

1. Anaconda Navigator

- Package management in **Environments** tab

2. Start Anaconda Prompt

- Package management with **pip** command



Install Extra Packages from Anaconda

- Update pip

```
> python -m pip install --upgrade pip
```

- Install pandas

```
> pip install pandas
```

- Install matplotlib

```
> pip install matplotlib
```

- Install metpy

```
> pip install metpy
```

```
> pip install netcdf4
```



- **GitHub Inc.** is a web-based hosting service for version control using Git.

<https://github.com/>

A screenshot of the GitHub homepage. The top navigation bar is dark blue with the GitHub logo, links for Features, Business, Explore, Marketplace, and Pricing, a search bar labeled "Search GitHub", and links for "Sign in" and "Sign up". The main content area has a dark background with the text "Built for developers" in large white font. On the right side, there is a white login/signup form with fields for Username, Email, and Password. The Username field contains the placeholder text "Pick a username". The Email field contains the placeholder text "you@example.com". The Password field is empty.

Username

Pick a username

Email

you@example.com


Password

Download course material from Github

- Go to the course materials on github

https://github.com/tingsyo/course_py4as

<https://goo.gl/Jh8zSz>

- Click  at upper right corner to make your own copy of the files.
- Download or clone the repository to your own computer by clicking



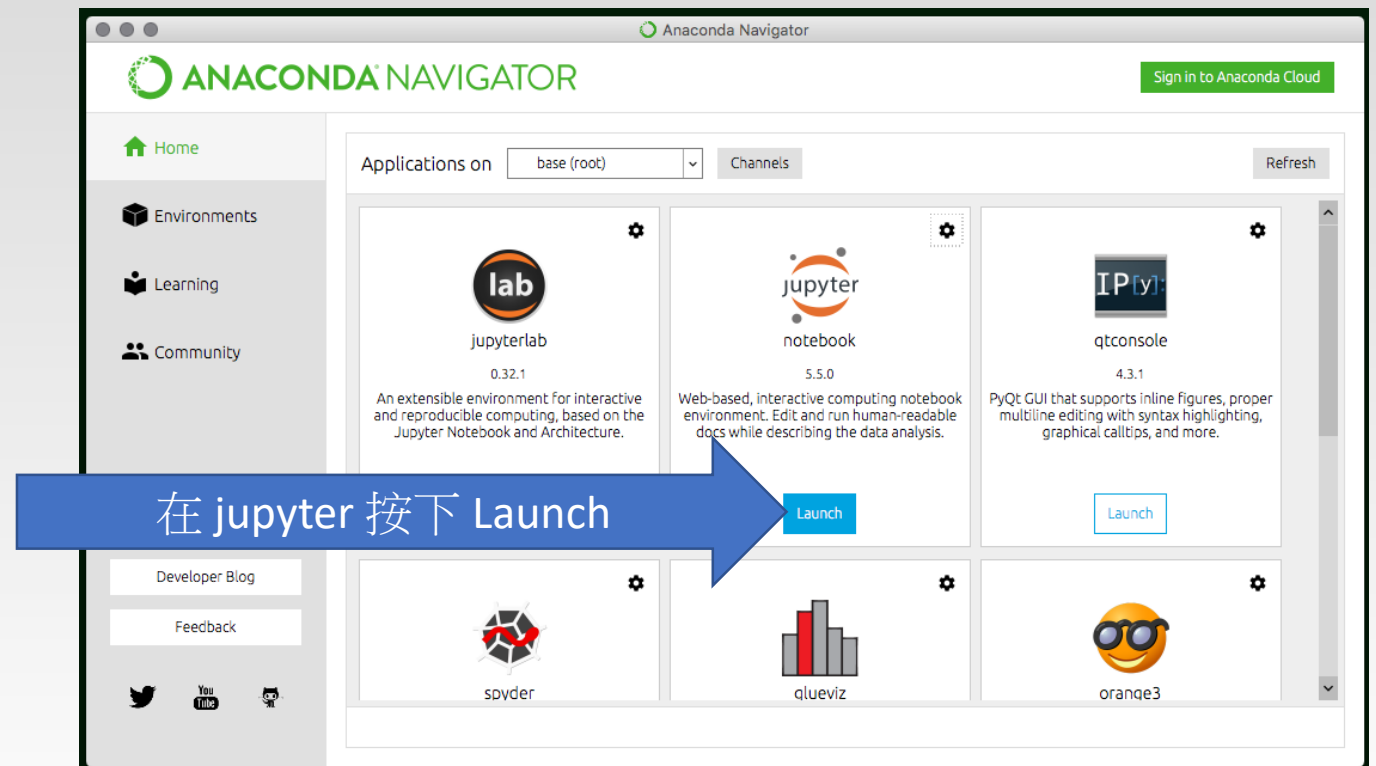
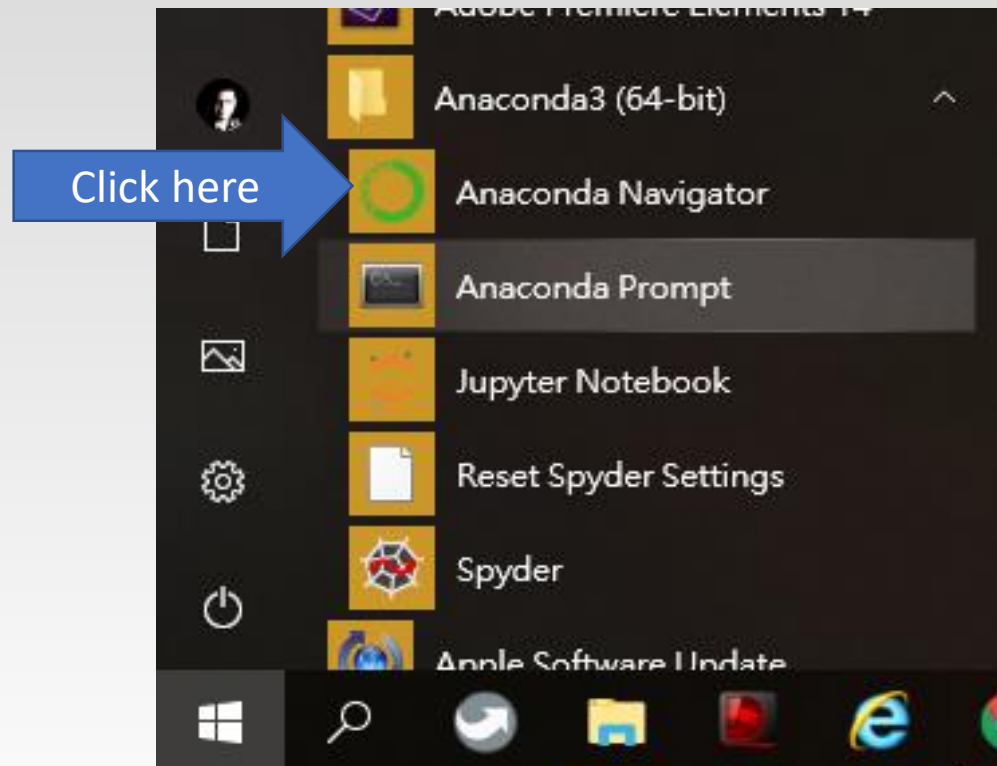


<https://tortoisegit.org/download/>



```
/usr/bin/ruby -e "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/master/install)"
```

Use Jupyter Notebook



Use Jupyter Notebook

The image shows a Windows Start menu on the left and an Anaconda Prompt window on the right. A blue arrow labeled "Click here" points to the "Anaconda Prompt" icon in the Start menu. Another blue arrow labeled "移動到工作目錄" (Move to work directory) points to the command `>>cd usr\Dropbox\work\201809\py4as` in the Anaconda Prompt. A third blue arrow labeled "輸入 jupyter notebook" (Enter jupyter notebook) points to the command `>jupyter notebook` at the bottom of the Anaconda Prompt.

Click here

移動到工作目錄

Anaconda Prompt - jupyter notebook

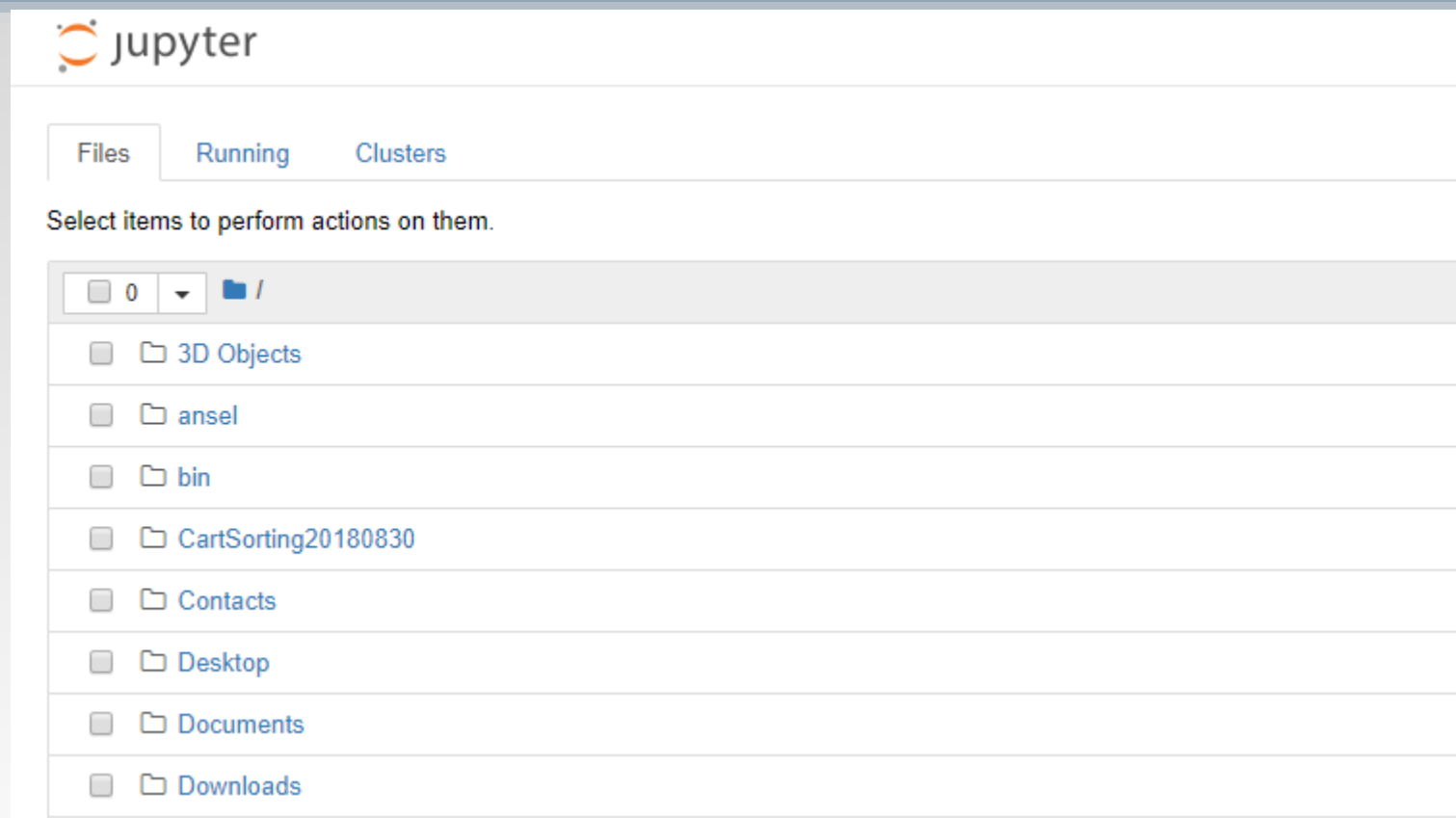
```
(base) D:\work>c:
(base) C:\usr\Dropbox\work\201809\py4as>>cd usr\Dropbox\work\201809\py4as
(base) C:\usr\Dropbox\work\201809\py4as>dir
磁碟區 C 中的磁碟是 Windows
磁碟區序號: FE93-01D3

C:\usr\Dropbox\work\201809\py4as 的目錄

2018/11/01 上午 12:50 <DIR> .
2018/11/01 上午 12:50 <DIR> ..
2018/11/01 上午 12:50 <DIR> course_py4as
2018/10/22 下午 05:46 <DIR> NCDR_r_programming
2018/11/01 上午 12:46 <DIR> pre_course_doc
2018/10/30 上午 10:03 <DIR> ref
2018/10/31 下午 02:31 <DIR> workspace
0 個檔案
7 個目錄 883,393,744,896 位元組可用

>jupyter notebook
```

Jupyter Notebook as a Web Service



Next ▶



python™

Basics

Course Objectives



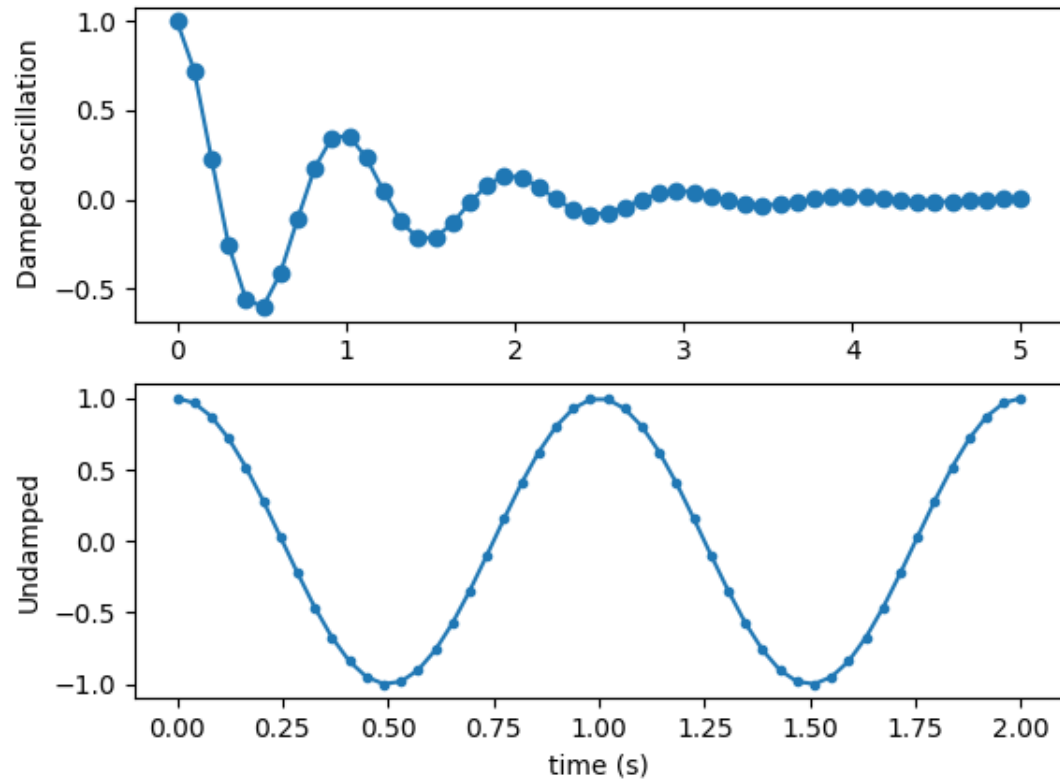
Python

Data
Analysis

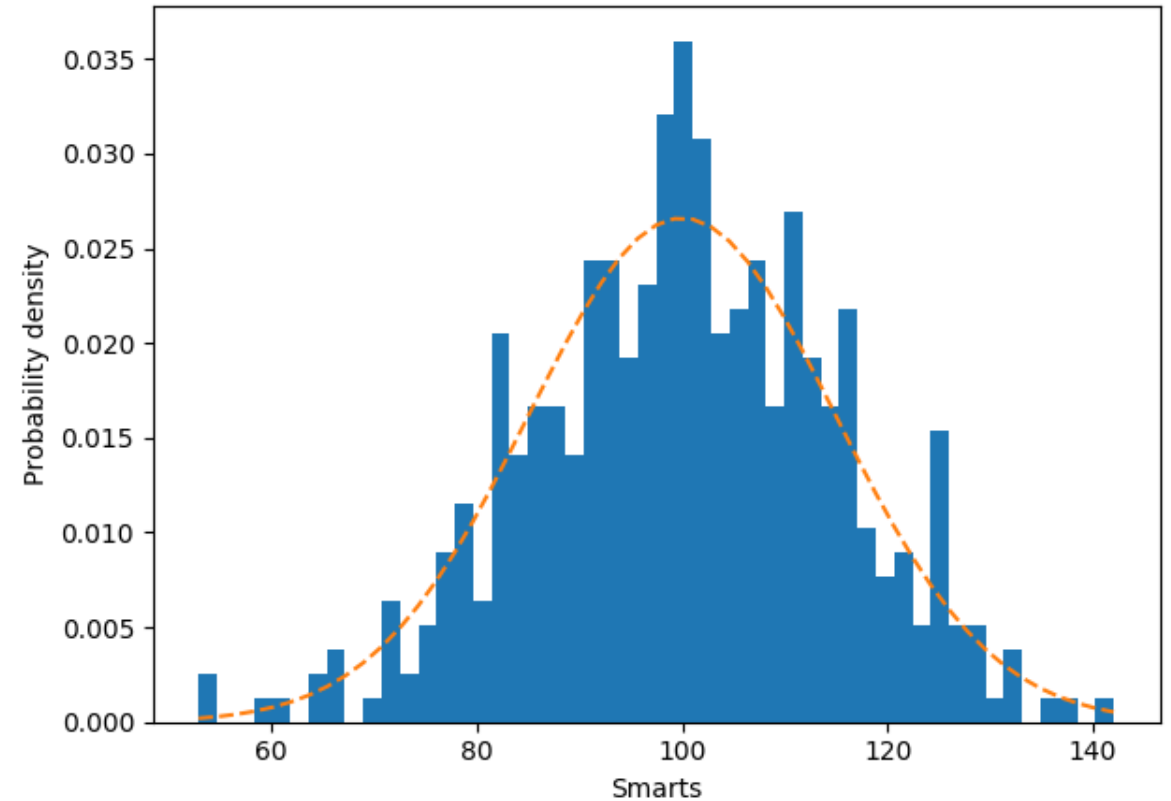
Atmospheric
Sciences

Course Objectives

A tale of 2 subplots



Histogram of IQ: $\mu = 100$, $\sigma = 15$



Course Objectives

