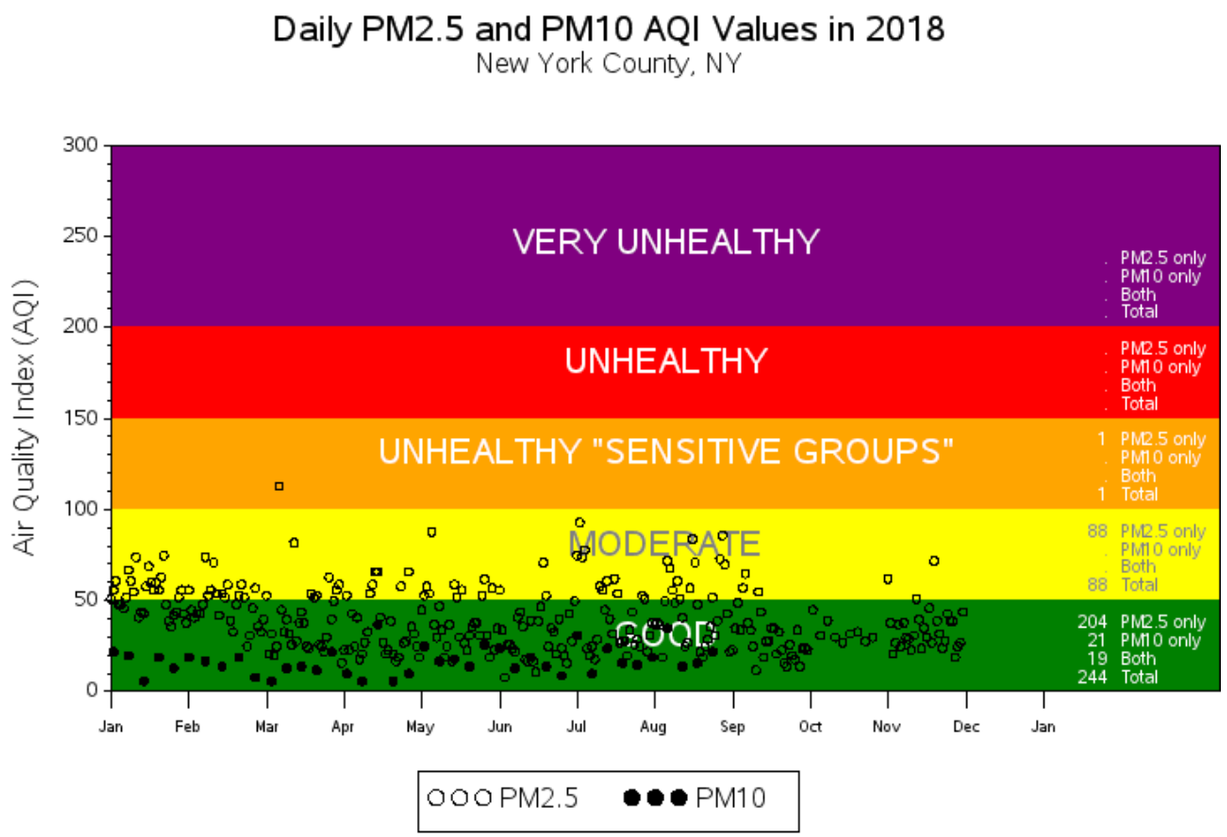
EASC2410 Lecture 1

Data Analysis and Get Start with Python

Dr. Binzheng Zhang
Department of Earth Sciences

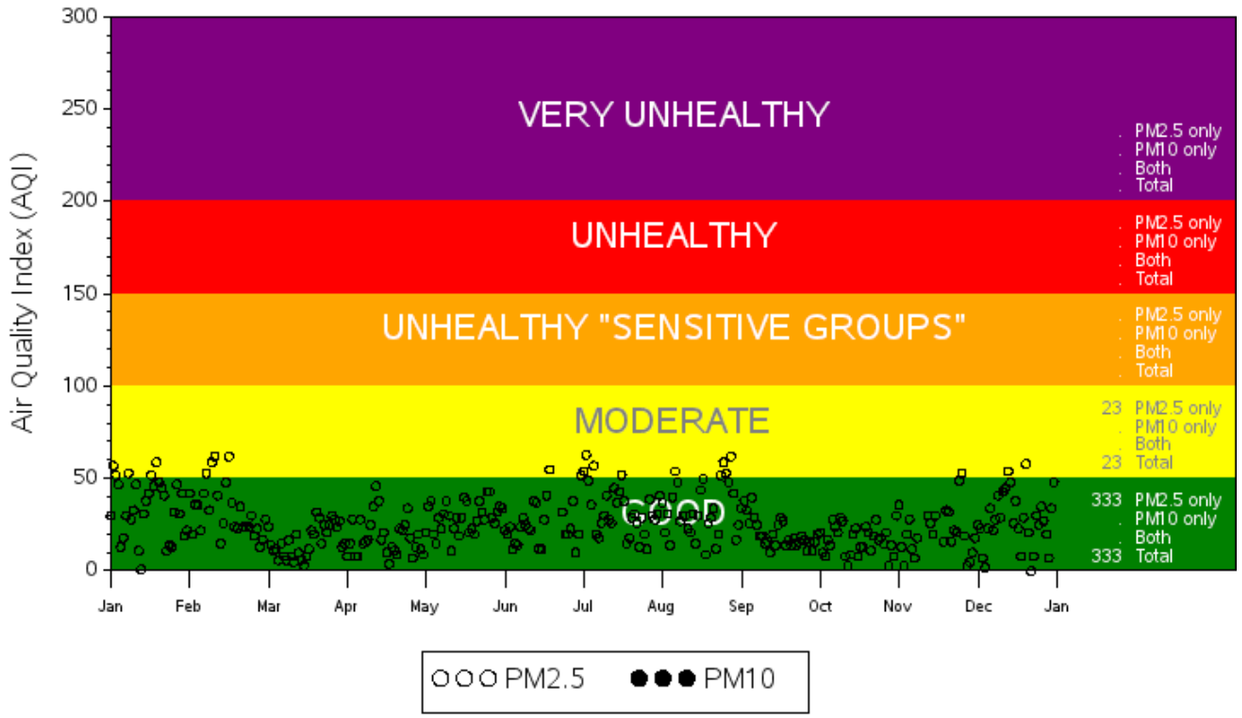
Data Analysis in Earth Sciences

Air Quality Data from EPA



Source: U.S. EPA AirData <https://www.epa.gov/air-data>
Generated: January 11, 2019

New York City, NY



Source: U.S. EPA AirData <https://www.epa.gov/air-data>
Generated: January 11, 2019

Grafton, NH

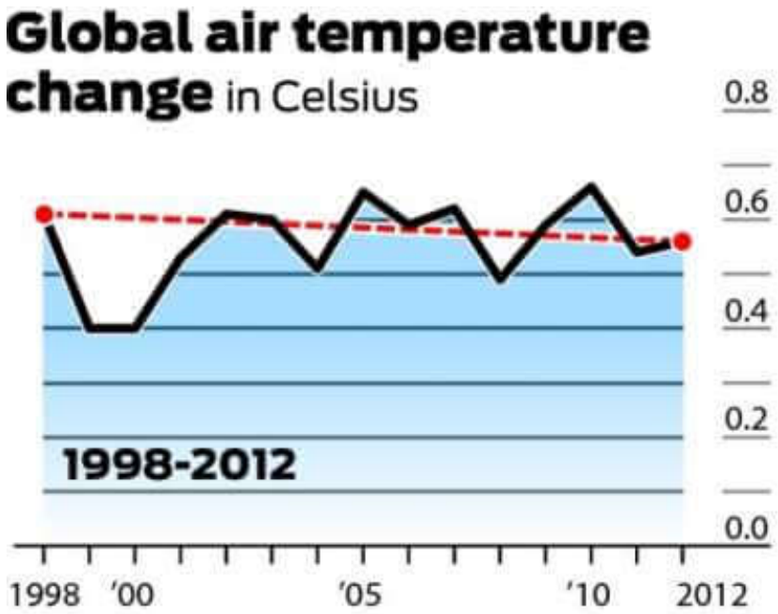
DARTMOUTH
250

Data Analysis in Earth Sciences

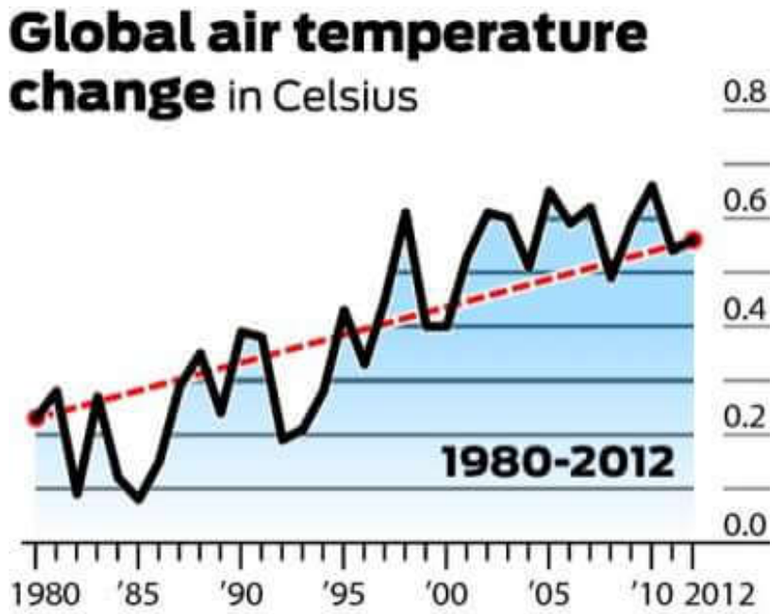
Earthquakes Data from USGS

Misleading Data Analysis

Global temperature change

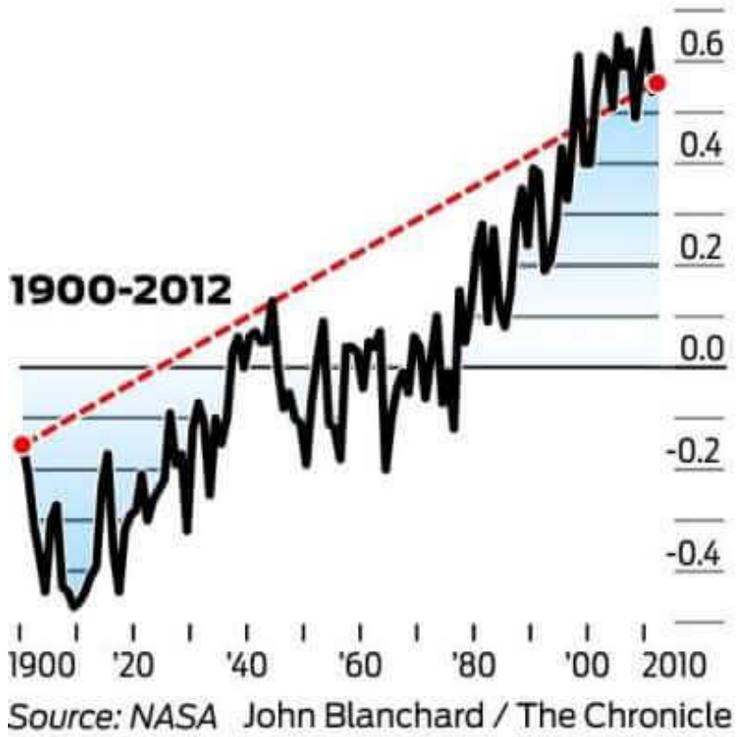


If you look at the temperature changes between 1998 and 2012, you may drawn the conclusion that the global air temperature is not increasing - in fact it's decreasing slightly



But, if you look at a longer time scale between 1980 and 2012, it is clearly to show that the global air temperature increases about 0.4 degrees in Celsius. That's because the atmosphere system has a much longer time scale of variation than 12 years

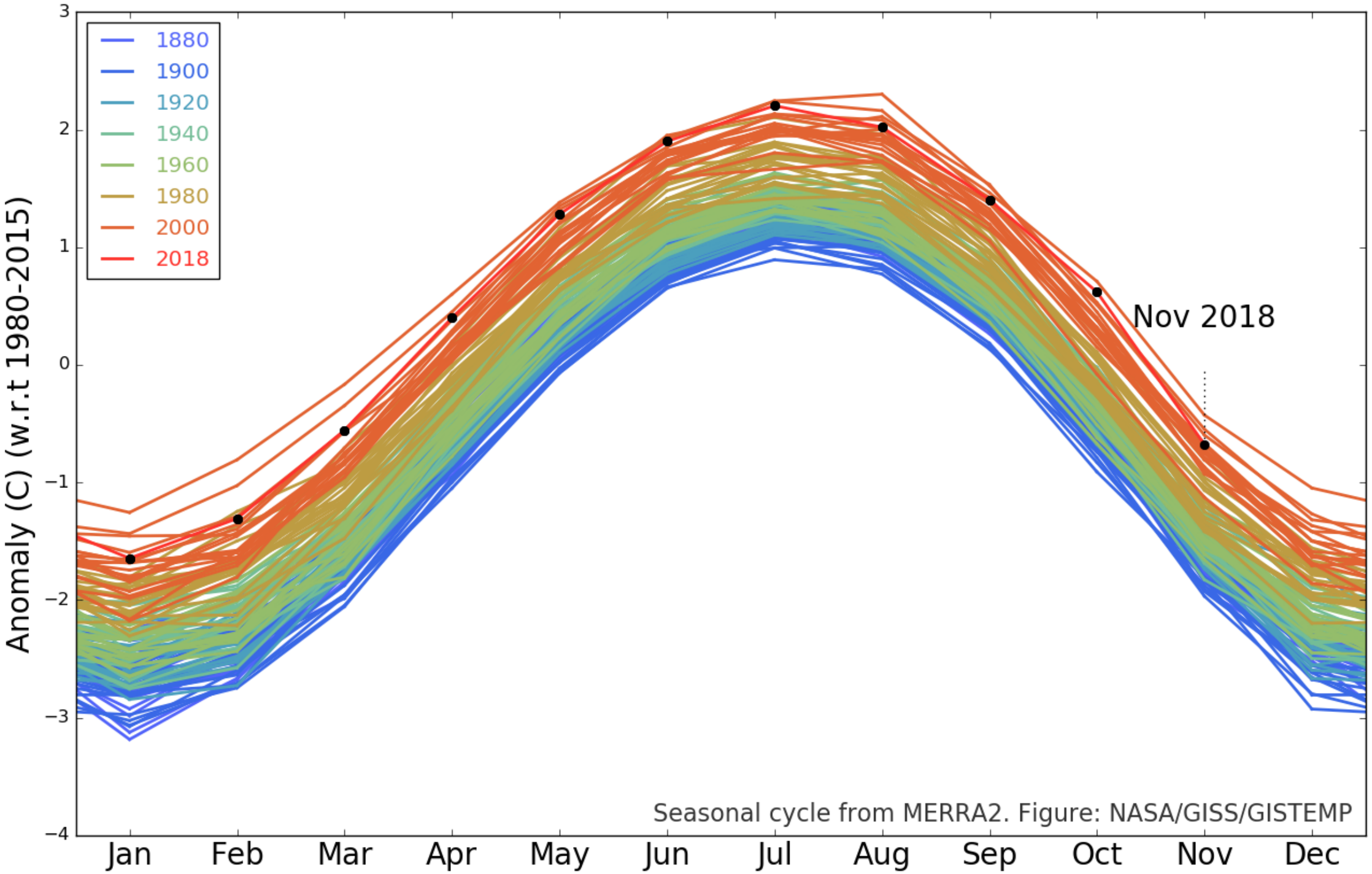
now if you look even further in time, it is very illustrative that we are facing a global temperature increases almost 100 years ago.



Data Analysis in Earth Sciences

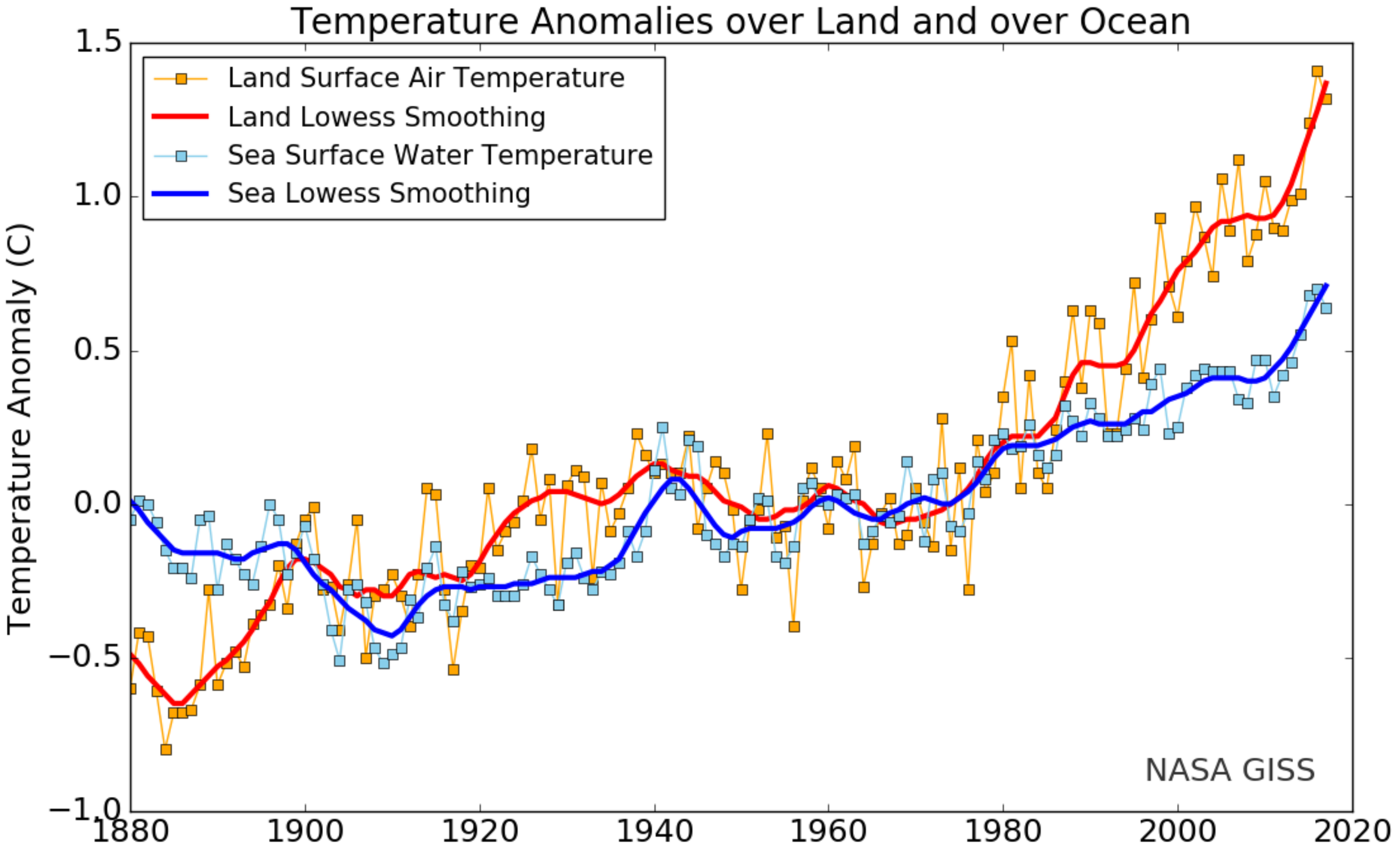
Global Temperature Data from NASA

GISTEMP Seasonal Cycle since 1880

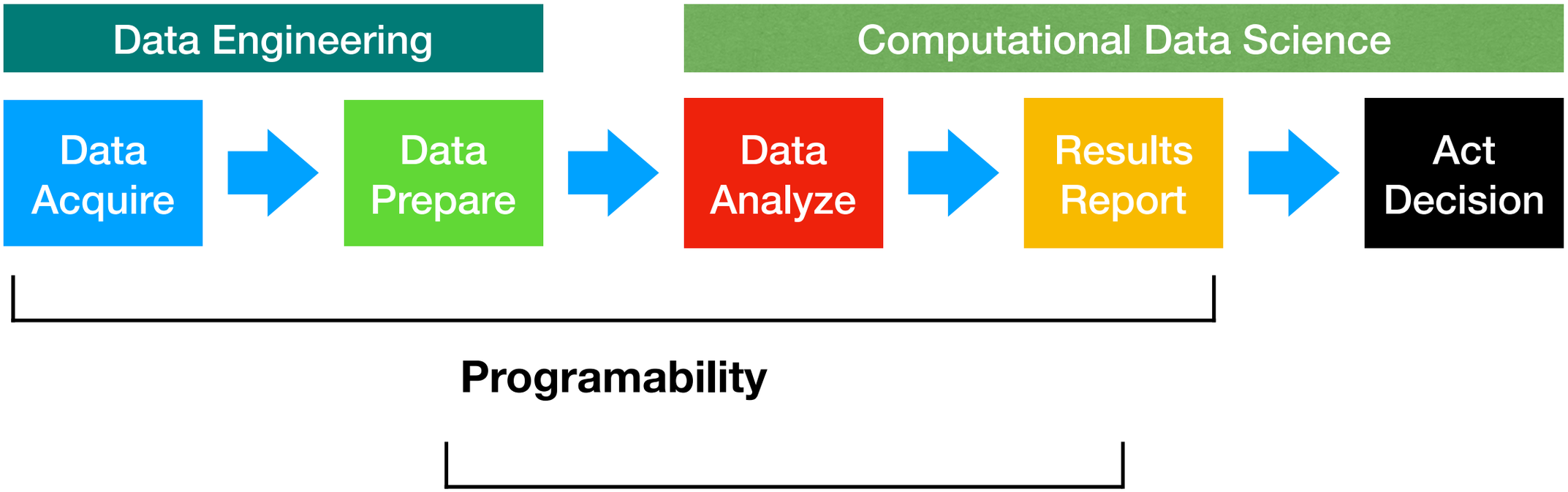


Data Analysis in Earth Sciences

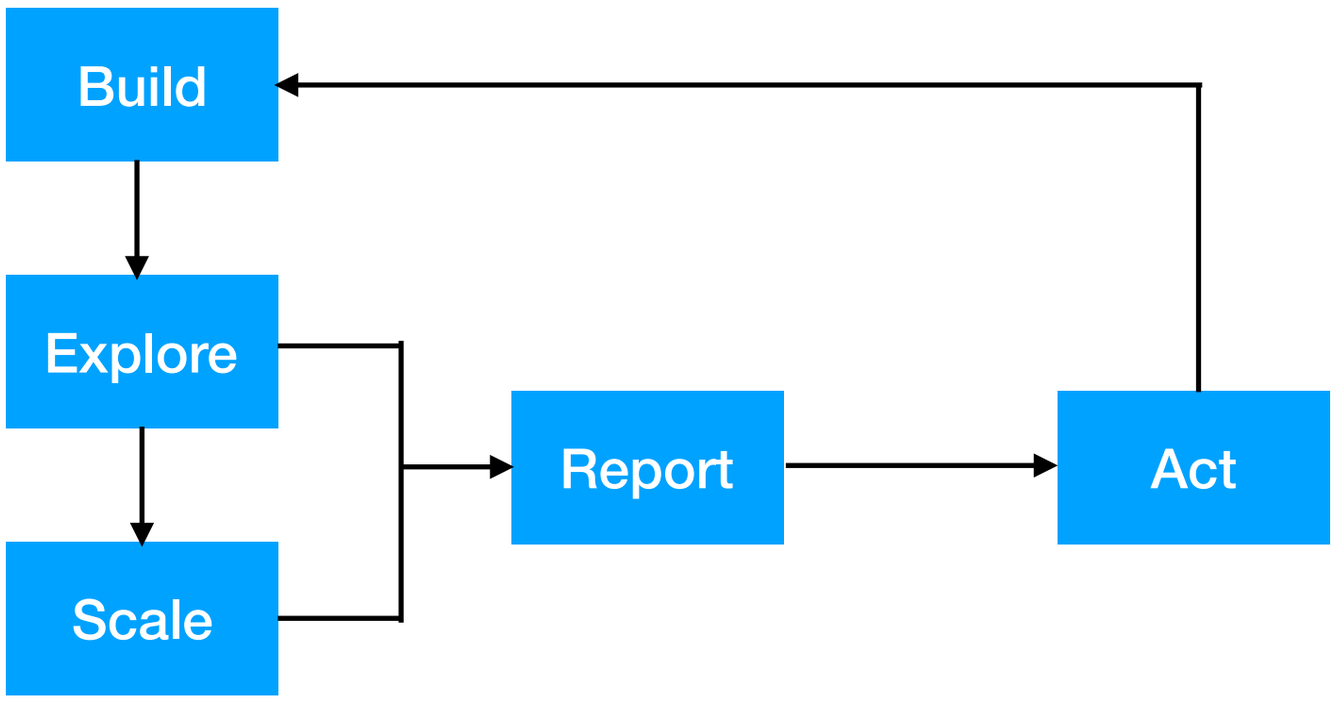
Land and Ocean Temperature Data from NASA



Steps in Data Science



This course

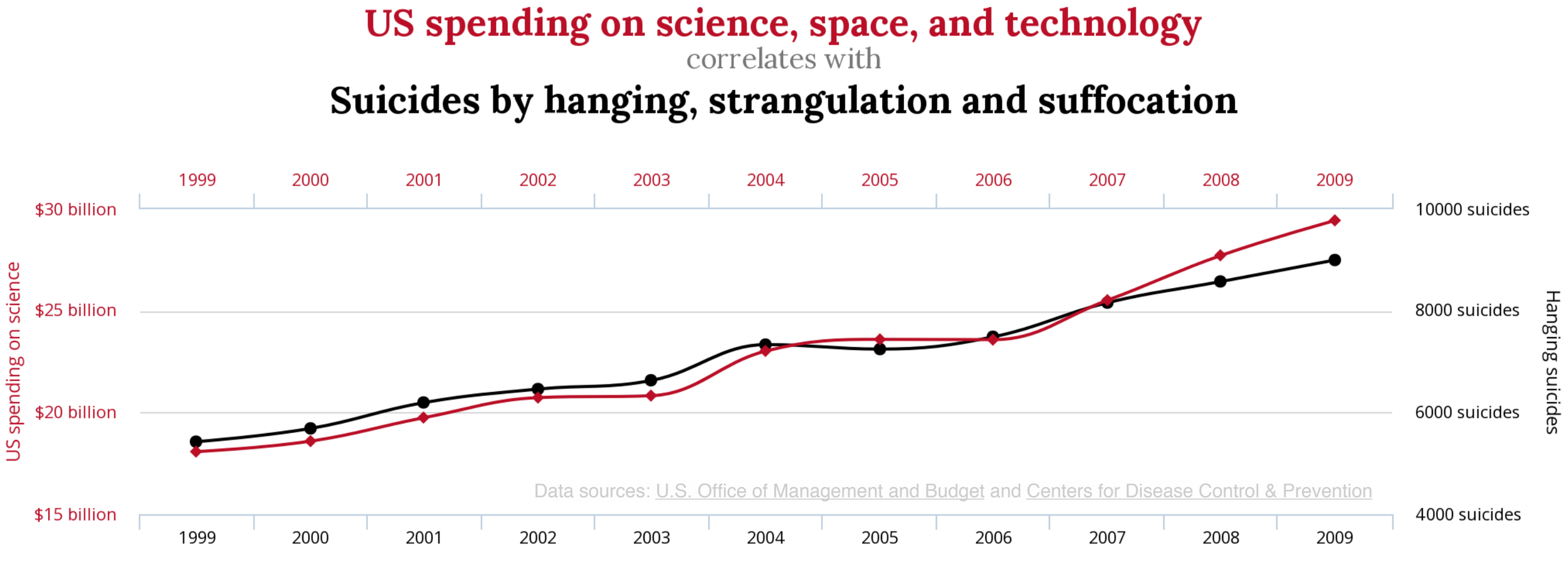


Iterative steps in Data Science

Bad Data Analysis

- Faulty polling
- Flawed correlations
- Data fishing
- Misleading data visualization
- Purposeful and selective bias
- Using percentage change in combination with a small sample size

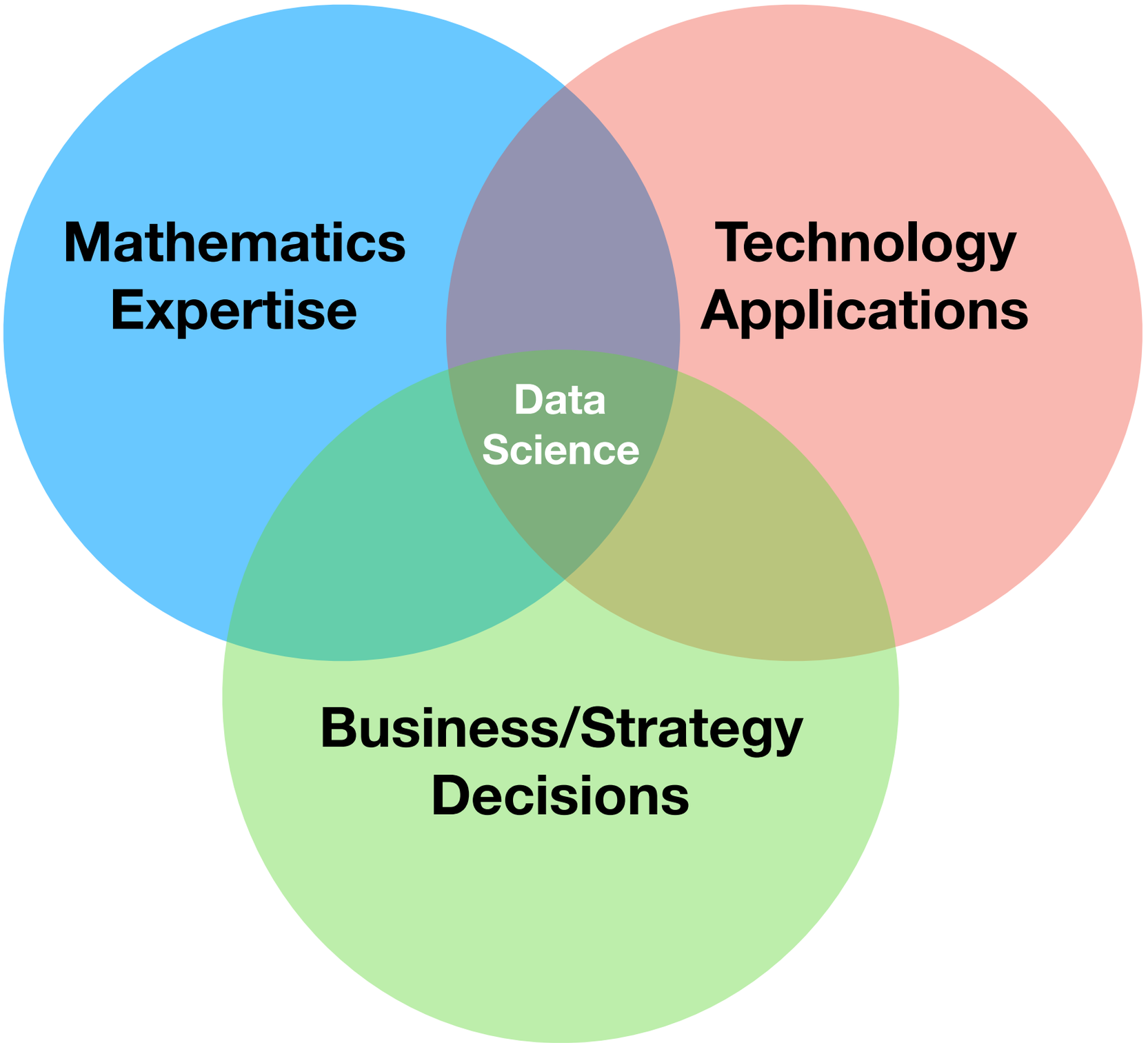
Remember: Correlations are not necessary causalities



What you will learn in this course

- **Basic programing skills for data analysis using Python (numpy, scipy, matplotlib, pandas, basemap, etc., ~ 4 weeks)**
- **Basic statistical skills for handling large amount of data (~ 2 weeks)**
- **Basic skills for performing time-series analysis (~ 2 weeks)**
- **Visualizations skills for making 1-D and 2-D plots including maps (~ 2 weeks)**
- **Learn how to perform basic mathematical modeling using python (~ 2 weeks)**

Data Science



Data Scientists Skills

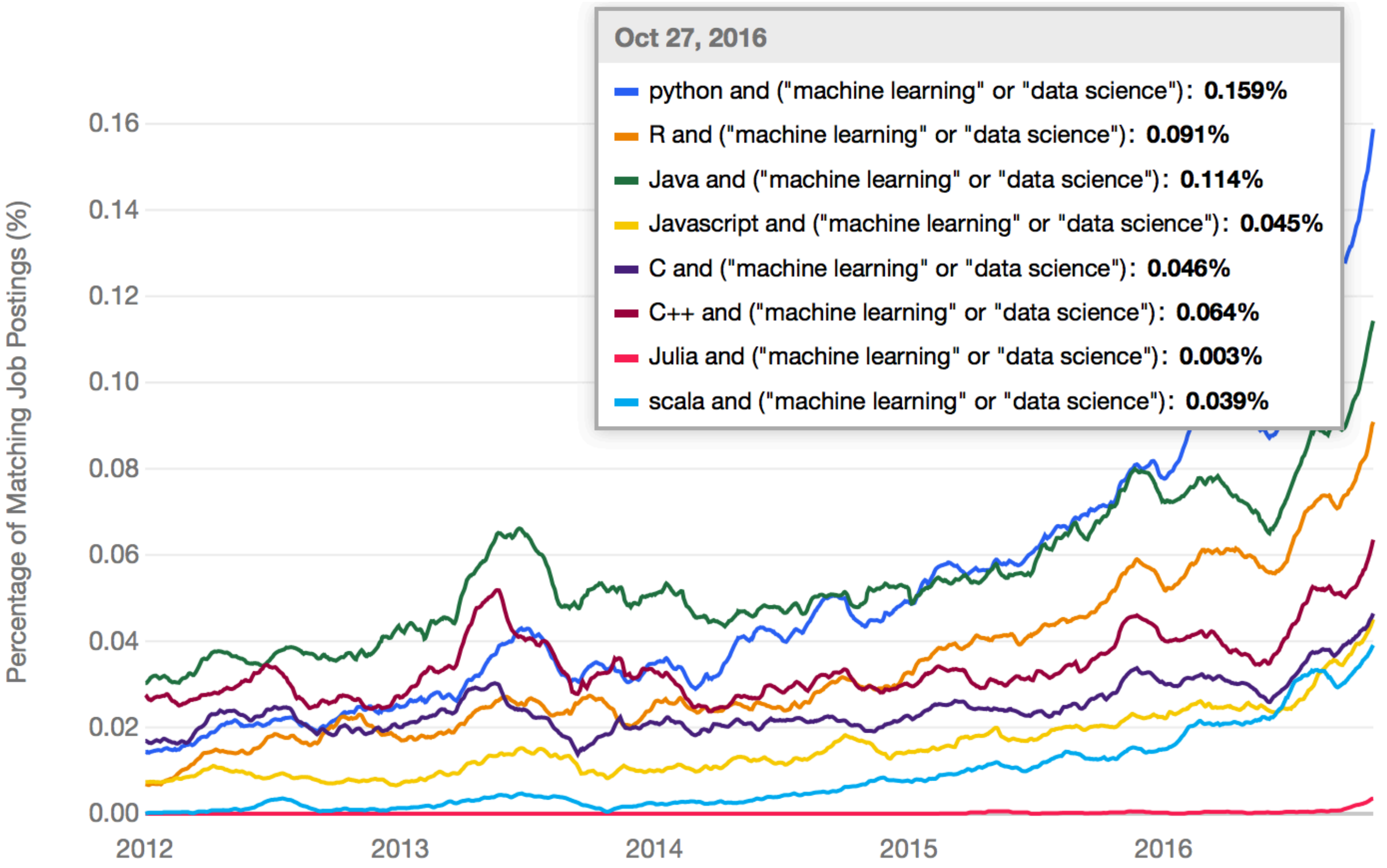
Soft Skills

- Have passion for data
- Relate problems to analytics
- Care about engineering solutions
- exhibit curiosity
- communicate with team players

Hard Skills

- machine learning
- statistical modeling
- relational algebra
- business passion
- problem solving
- data visualization

Why Python for Data Science?



Why Python for Data Science?

- **Easy-to-read and learn (self-learning is straightforward)**
- **Vibrant community**
- **Growing and evolving sets of open-sourced libraries**
  - **Data management**
  - **Analytical processing**
  - **Visualization**
- **Applicable to each step in the data science process**
- **Notebooks!**

Python Basics - the “Hello word” program

```
Hello World.
```

C

```
#include “studio.h”
int main() {
    printf(“Hello World.\n”);
}
```

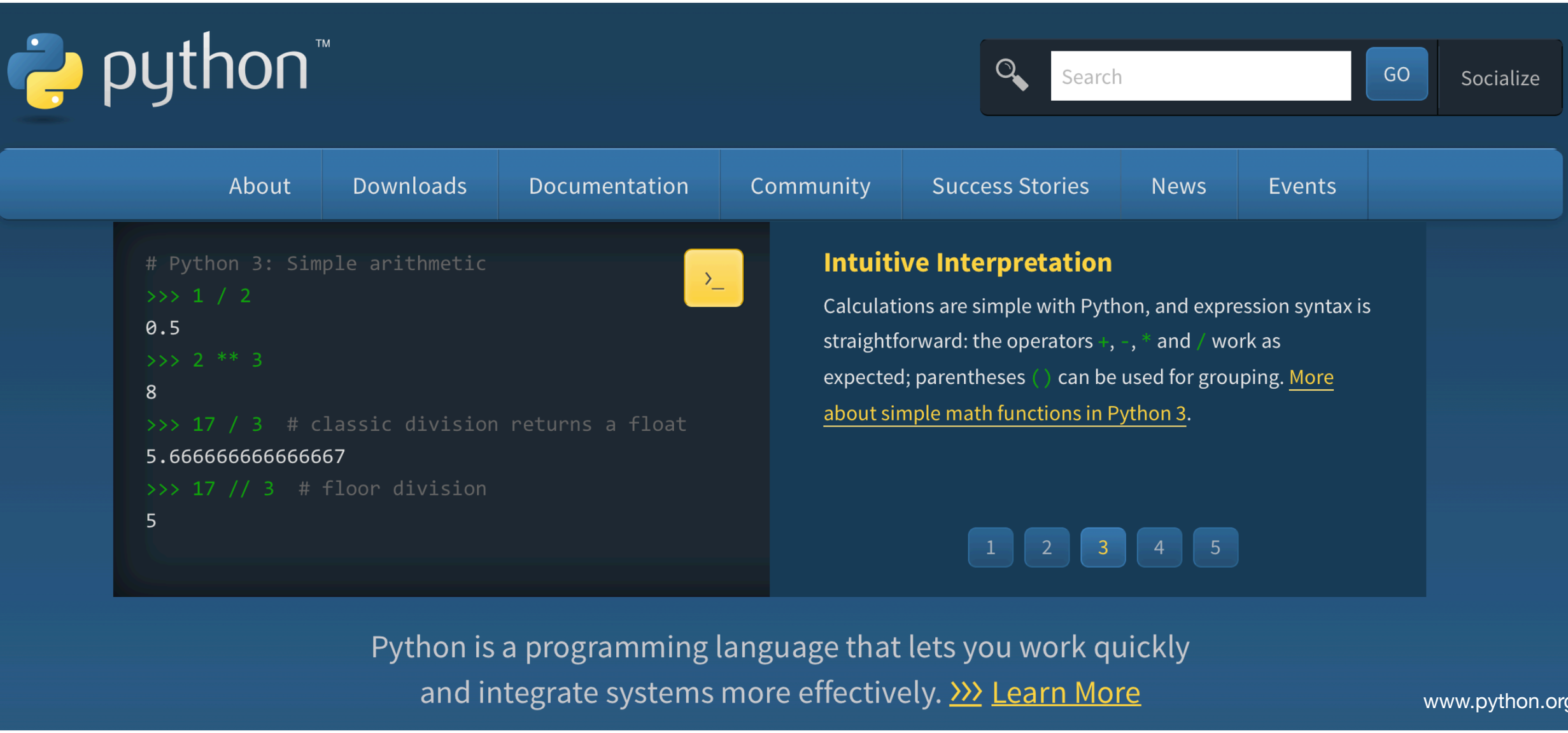
Python

```
print(“Hello world.”)
```

Java

```
public class Hi {
    public static void main (string [] args) {
        System.out.println(“Hello world.”);
    }
}
```

Python Overview



- **Python is powerful and fast**
- **Python plays well with others**
- **Python runs everywhere**
- **Python is high-level, friendly and easy to learn**
- **Python is free**

Python Overview

A couple of useful websites for learning Python:

www.python.org

www.anaconda.com

http://www.numpy.org

https://www.scipy.org

https://matplotlib.org

https://pandas.pydata.org

What we use in this course

https://www.anaconda.com

Documentation Blog Contact

ANACONDA.

What is Anaconda? Products Support Resources About **Downloads**

The Most Popular Python Data Science Platform

Accelerate

Streamline your data science workflows from data ingest through deployment

Connect

Leverage & integrate all your data sources to extract the most value from your data

Empower

Create, collaborate & share with your entire team—from analysts to executives

Anaconda Products

ANACONDA. DISTRIBUTION

The Most Popular Python Data Science Distribution

Download Now

ANACONDA. ENTERPRISE

The AI Enablement Platform for Teams at Scale

Learn More

ANACONDA. TRAINING

Anaconda Data Science Training and Certification

Get Educated

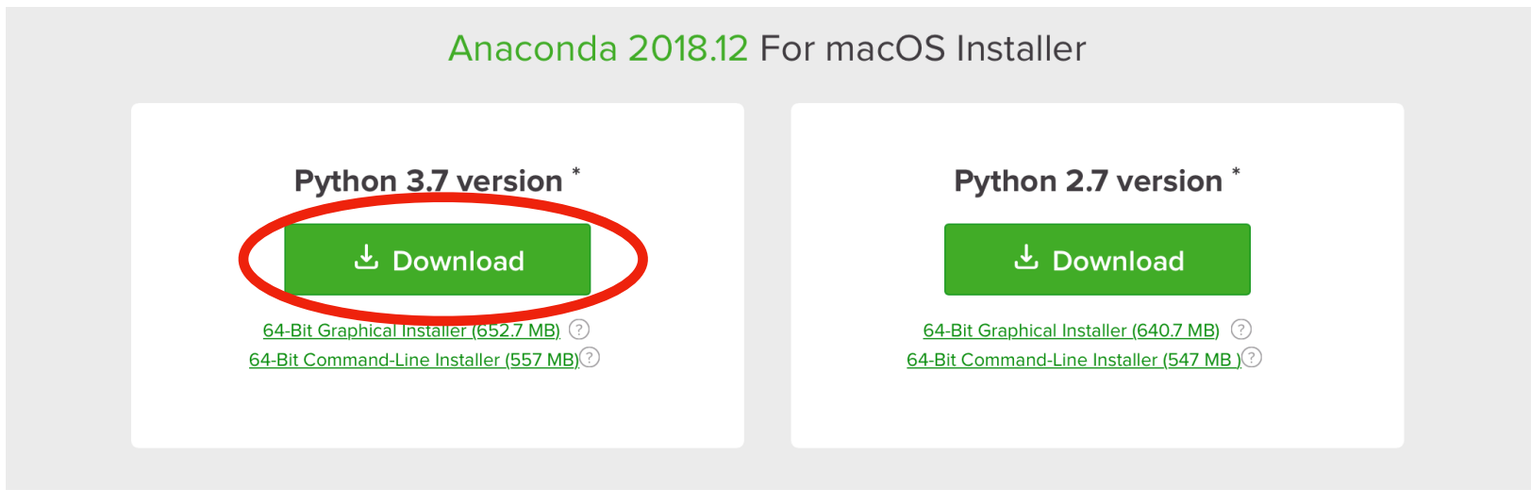
Steps to Get Anaconda and run Jupyter notebook

1. Open your browser, type in www.anaconda.com

2. Click the “Download” button in the top-right corner

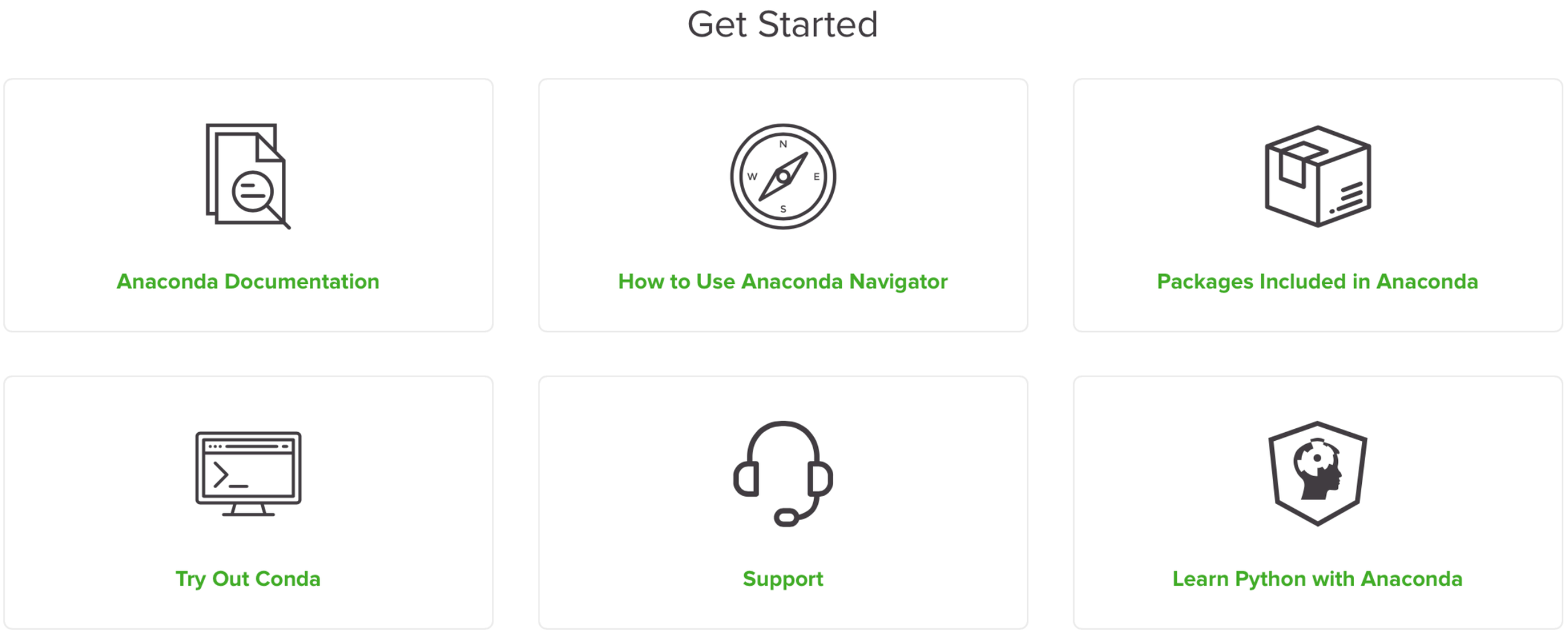
3. Select the appropriate link for your system (Windows, Mac, Linux)

4. Download the graphical installer with the Python 3.7 version



Anaconda Documentation

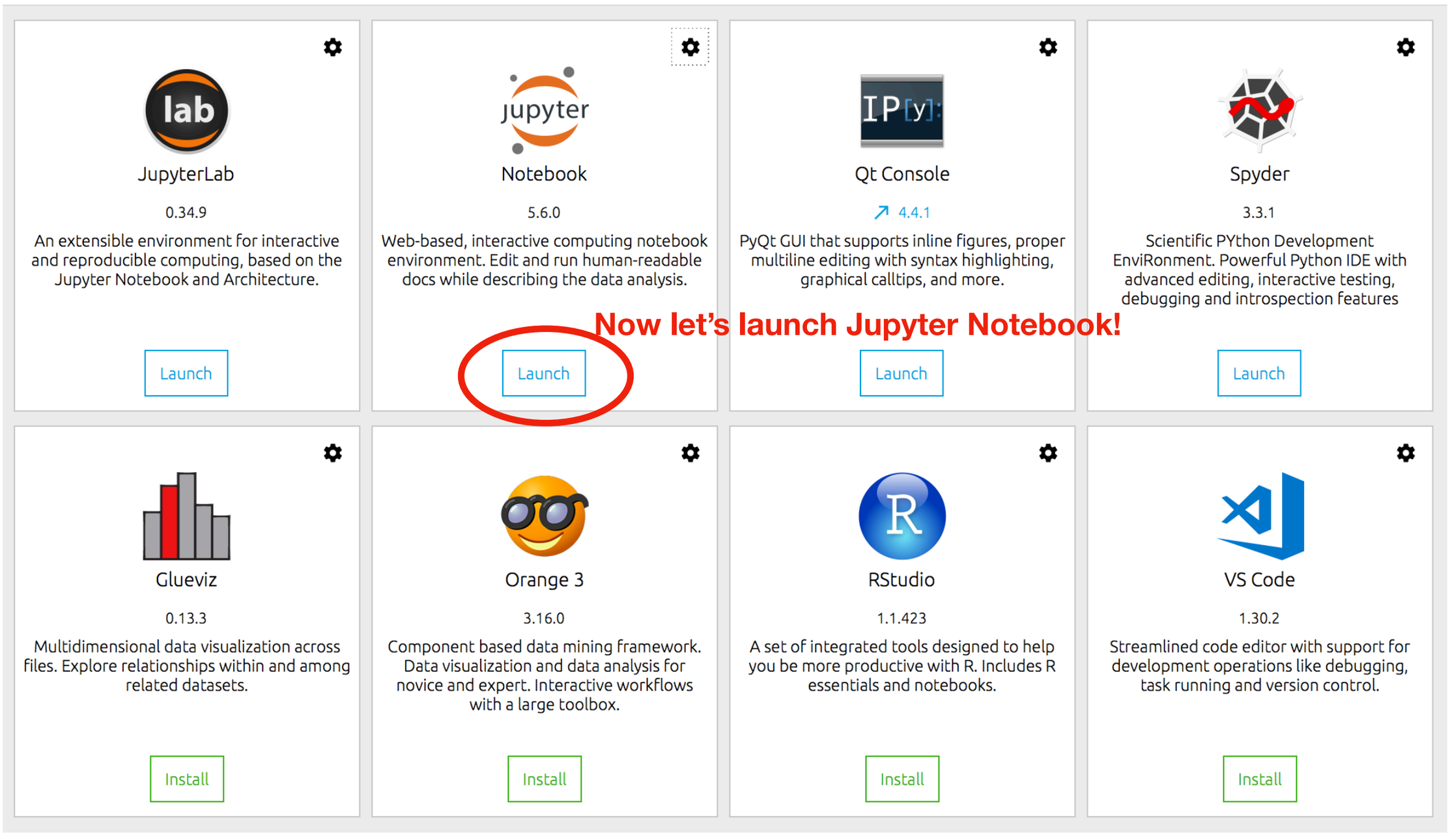
Scroll down to the bottom of the “Anaconda Download” page, here’s the documentation:



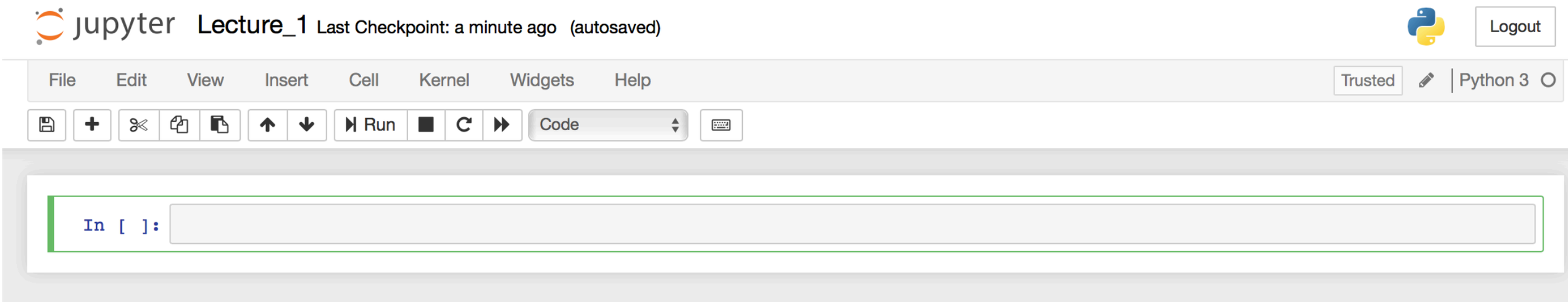
Launch the Anaconda Navigator

Double-Click the Conda symbol:

Here's what you get



Now Jupyter notebook is launched and you're brought to a new browser page looks like this



Here you go! you're all set to enjoy Python - let's try the "hello world" program!

First type print("Hello World")

```
In [ ]: # This is my first Python code
        print("Hello World!")
```

Then click the "Run" bottom (or ctrl+enter)


```
In [ ]: # This is my first Python code
        print("Hello World!")
```

Your code is executed and here's the output:

```
In [3]: # This is my first Python code
        print("Hello World!")
Hello World!
```

Comments, start with a pound sign

Actual codes

Results

Comment your codes!

This is a must

In [ ]:
```
# 1. Comments start with a pound sign "#"
# 2. Comments are not executed by Python
# 3. Basically you can type anything here to explain your program
# 4. A good code usually comes with a fairly large amount of comments
# 5. You are required to comment you code through out the course/homeworks
#    If you codes are not commented, you WILL lose marks on homeworks
```

Note:

- Comments start with the pound sign “#”
- They are ignored by Python (not executed)

What happens if I run the above code?

What happens if I run the following code?

In [5]:
```
# 1. Comments start with a pound sign "#"
# 2. Comments are not executed by Python
# 3. Basically you can type anything here to explain your program
# 4. A good code usually comes with a fairly large amount of comments
# 5. You are required to comment you code through out the course/homeworks
    If you codes are not commented, you WILL lose marks on homeworks
```

Comment your codes!

This is a must

This is what a “bad” code looks like

```
from numpy import *
from matplotlib.pyplot import *
import scipy.integrate as int
import matplotlib.animation as animation

def init():
    quad1.set_array([])
    quad2.set_array([])
    return quad1,quad2

def animate(i):
    global cont
    for c in cont.collections:
        c.remove()
    z = F_calc[:,:,i].T/area.T
    cont = contourf(x,y,z,11,cmap='seismic')
    return cont


def A_dl(s,tilt,phi,r0,r1):
    ct = cos(tilt)
    st = sin(tilt)
    dr = r1-r0
    dx = dr[0]
    dy = dr[1]
    dz = dr[2]
    x = r0[0] + s*dx
    y = r0[1] + s*dy
    z = r0[2] + s*dz
    r = sqrt(x**2+y**2+z**2)
    A_d =( (0.1e1 / r ** 3 * sin(tilt) * sin(phi) * z -
            0.1e1 / r ** 3 * cos(tilt) * y) * dx +
       (-0.1e1 / r ** 3 * sin(tilt) * cos(phi) * z +
        0.1e1 / r ** 3 * cos(tilt) * x) * dy +
       (0.1e1 / r ** 3 * sin(tilt) * cos(phi) * y -
        0.1e1 / r ** 3 * sin(tilt) * sin(phi) * x) * dz)
    return A_d
```

This is what a reasonable code looks like

```
# solar wind calculation
# Author Dr. Kareem Sorathia, Johns Hopkins University

import h5py
import numpy as np
import sys

pW = (1.0e-2)/50.0 #Default THERMAL pressure, nPa
nW = 0.1 #Default density, #/cc
VxW = 300.0 #Default wind, km/s

fOut = "bcwind.h5" # output h5 file name

#Time bounds [hours]
tMin = 0.0 # start time
tMax = 200.0 # stop time
dt = 60.0 #Cadence [s]

SimT = (tMax-tMin)*60.0*60.0
NumT = np.int( np.ceil(SimT/dt)+1 )

print("Generating %d slices, T=[%5.2f,%5.2f]"%(NumT,tMin,tMax)) #info out

T = np.linspace(tMin,tMax,NumT) # time series, in minute
D = np.zeros(NumT) # solar wind density

tWin = 10.0 # Window times without IMF [hr]
for i in range(NumT):
        t = T[i] #Time in hours
        if (t <= tWin):
                D[i] = nW
        else:
                D[i] = nW-1
```

why bad?

You are required to write your code in this way

How to get help - use the “help()” function

For example if we want to know more about the print() function used before, simply type in help(print) in the code line and clicl ctrl+enter:

```
In [1]: help(print)
```

The help function is very useful to the self-learning process, as well as online materials for example you should (probably already are) get used to use something like Google. It is quite possibly that someone has already done something regarding the questions you are asking - there’s no need to figure out everything on your own (just remember to give credit to the source)

Exercise 1-3. Use the Jupyter note book as a printer

```
#  ~~~~~~~~~~~~~~EXERCISE 1 - Getting familiar with Jupyter Notebooks~~~~~~~~~~~~~~~~~~
#
# Creating new notebooks, naming and saving:
#
# To open a new notebook, pull down the 'File' menu and select the 'New Notebook => Python 3' option.
# You can 'tear' the notebook off so it is in a separate window if you like seeing both together.
#
# To rename the notebook, choose 'Rename' under the File menu.
#
# To save it, click on the disk icon on the menu bar.
# Wait until the 'checkpoint created...' message disappears before you try to exit the notebook.
```

```
#  ~~~~~~~~~~~~~~EXERCISE 2 - use the print() function and get help~~~~~~~~~~~~~~~~~~
#
# 1. Copy the following code to your new notebook and execute it, see what happens
# 2. Change the name in the first line to be your own name
# 3. Change "geophysicist" to be your major
# 4. Change the number 3.1415926 to be whatever your favorite food is (a word, with "")
# 5. Move to a new cell, and type in help(print) to see the details of the print() function
#
# NOTE: don't worry about some symbols, we will discuss them in the next lecture

print("My name is Binzheng Zhang")
print("I am a" + " " + "geophysicist")
print("I","like",3.1415926)
print("Now I know Python!")
```

```
#  ~~~~~~~~~~~~~~EXERCISE 3 - Getting help using the help() function~~~~~~~~~~~~~~~~~~
help(print)
```

Exercise 4-5. Use the Jupyter notebook as a calculator

type in the following examples in your Jupyter notebook and get the answers

operation symbol	function	examples	Answers
+	adds	2+3	`5`
-	subtracts	4.2-1.5	`2.7`
*	multiplies	3*91.1	`273.29999999999995`
/	divides	2/3	`0.6666666666666666`
%	gives the remainder	2**10	`1024`
**	raises to the power	4%3	`1`
==	test equality	1==2	`false`
!=	test inequality	1!=2	`true`