Intro to Python for Data Analysis

December 10, 2019

Brooke Luetgert, PhD

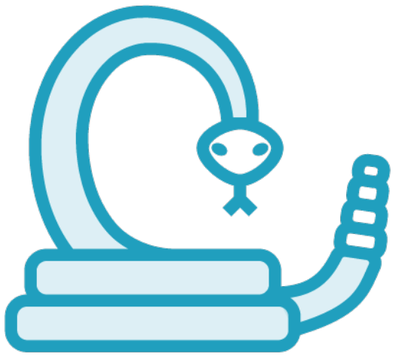
THE UNIVERSITY OF CHICAGO
Office of Research and National Laboratories Research Computing Center

Plan for Today

We will explore Python and several libraries for basic data analysis and visualization.

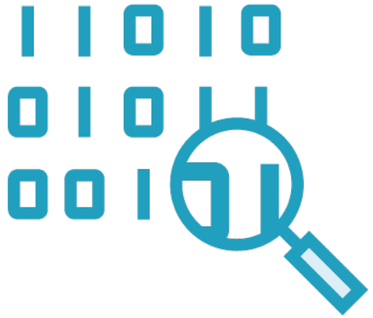
1. Download Anaconda
2. Look at Jupyter Notebook and Markdown
3. Build some basic Python syntax and introduce libraries
4. Import data into a data frame with Pandas
5. Plot data with Matplotlib and export

Technologies Used

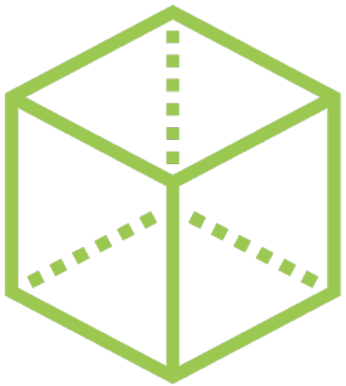


Python

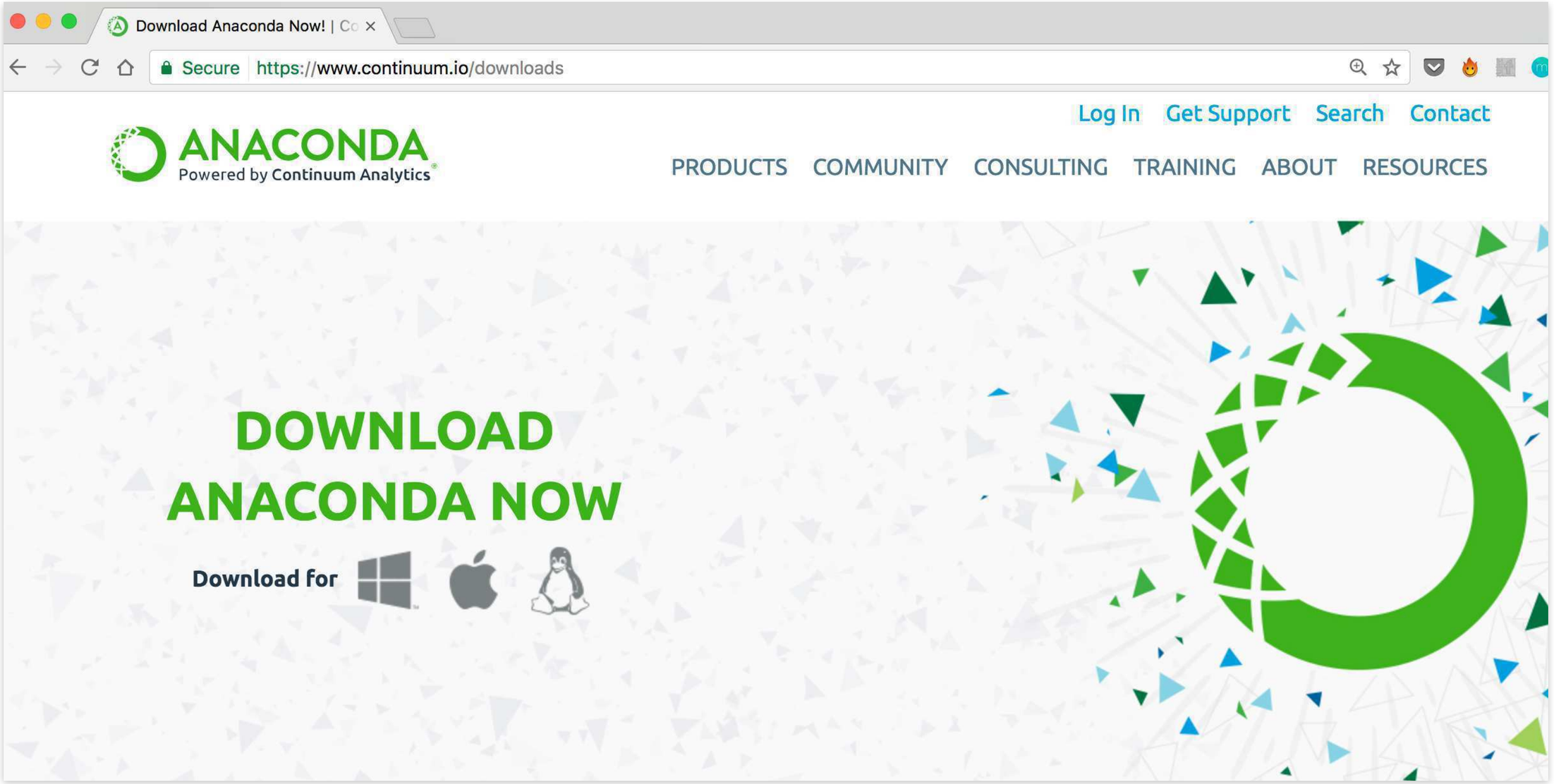
Matplotlib

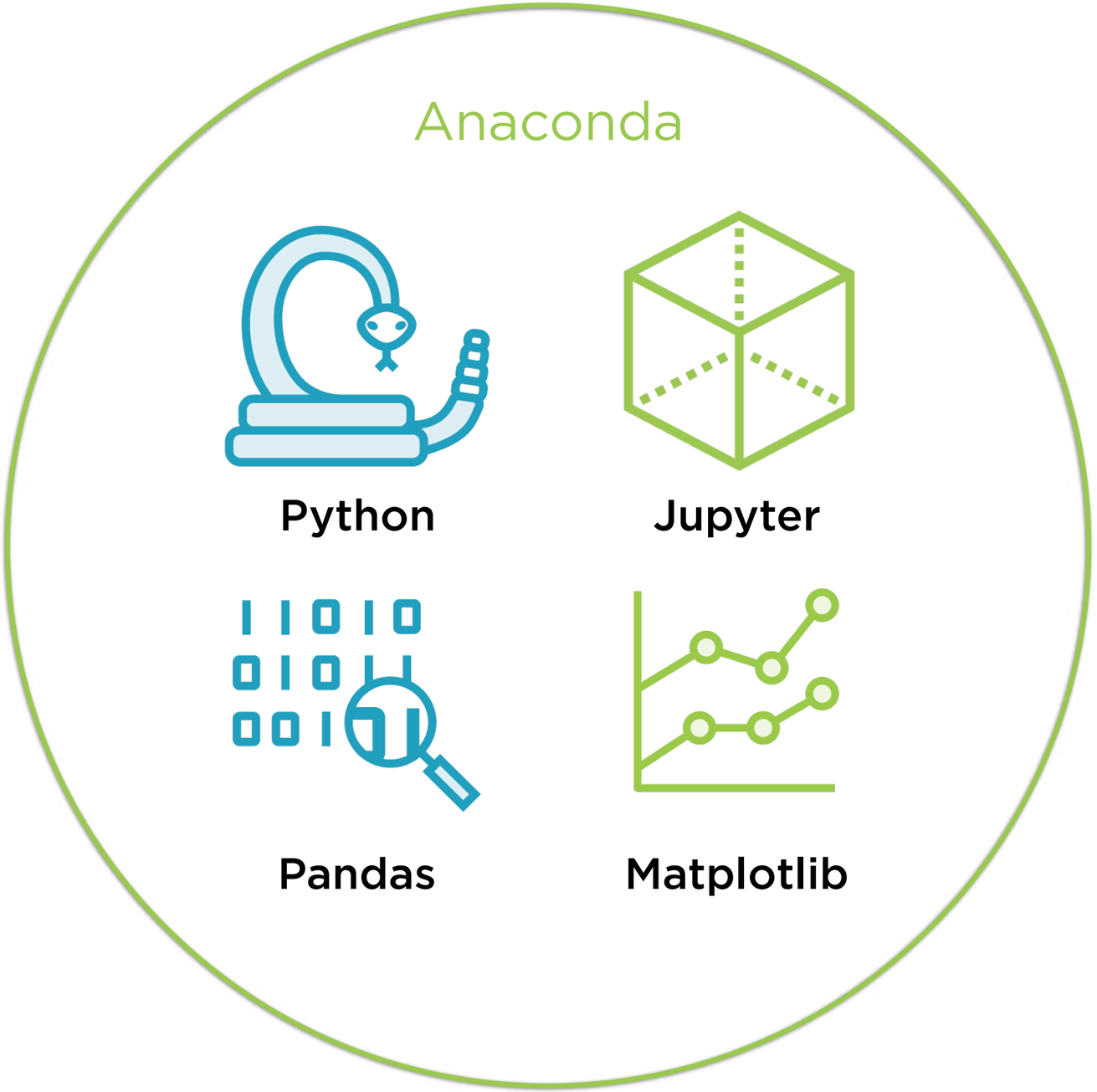


Pandas



Jupyter


Download Anaconda Now! | Co
Secure
https://www.continuum.io/downloads
Log In
Get Support
Search
Contact
ANACONDA
Powered by Continuum Analytics
PRODUCTS
COMMUNITY
CONSULTING
TRAINING
ABOUT
RESOURCES
DOWNLOAD
ANACONDA NOW
Download for



No Separate
Installation Required

Essential Libraries

NumPy- tools for fast reading and writing array-based datasets, linear algebra, transformations and random number generator

pandas- adds flexible data manipulation capabilities of spreadsheets and relational databases with indexing, reshaping, aggregation, time-series and merging functionalities

matplotlib-most popular library for plots and 2D visualizations

Jupyter-web-based code notebooks allowing integration of Markdown and HTML for the creation of rich docs with code and text. Also useful for visualization and de-bugging.

Why Python?

Web integration

- Python integrates exceptionally well into web APIs and interactive interfaces

Resolves 2-language problem

- Python can be used for both research and production, eliminating need for Java and C engineers

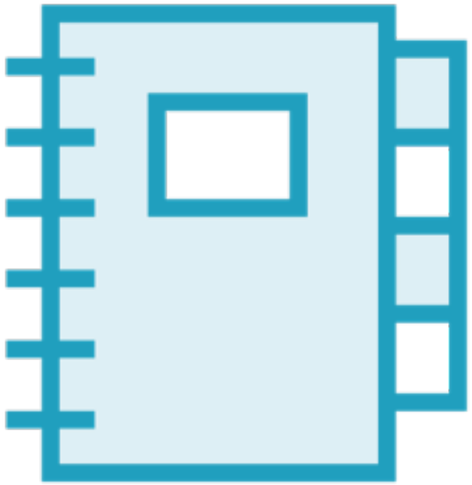
OS independent, large user base, extensive libraries

Using Jupyter for the First Time

What is Jupyter?

Write

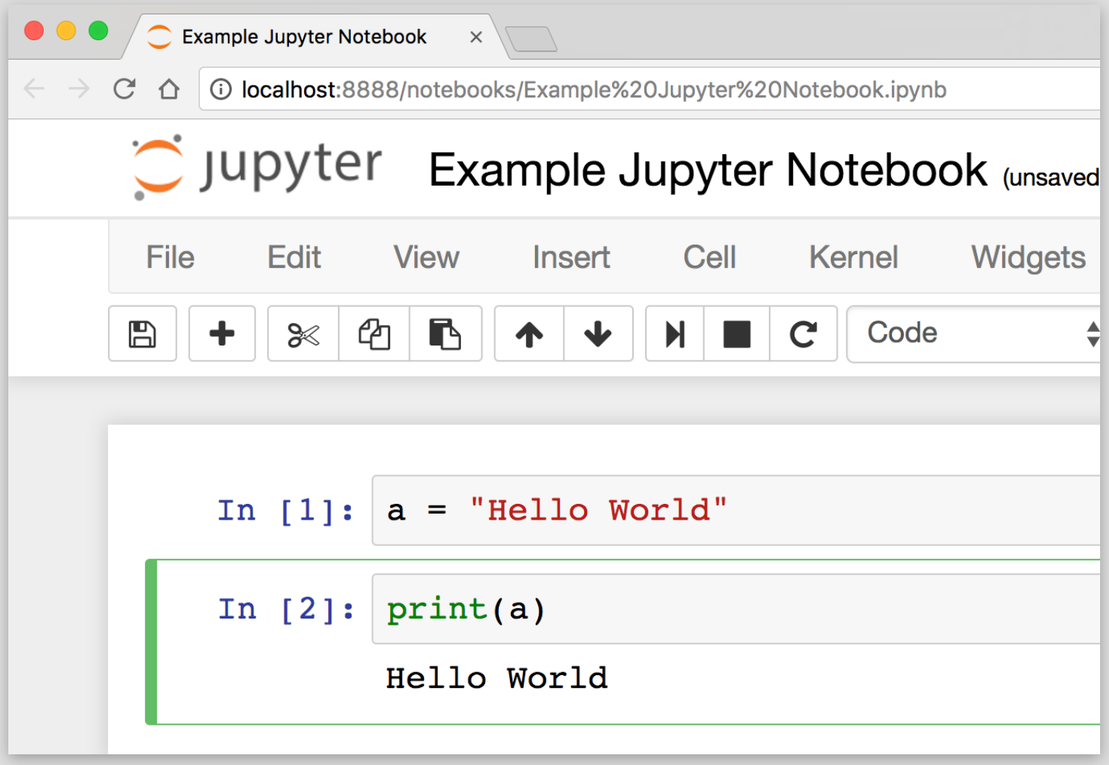
Test



Organize

Browser Interface

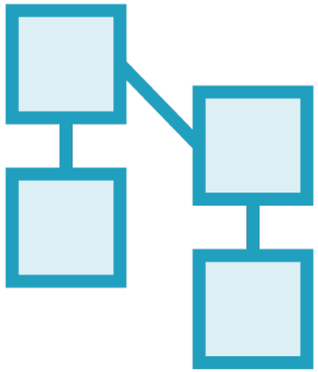
Python Interpreter



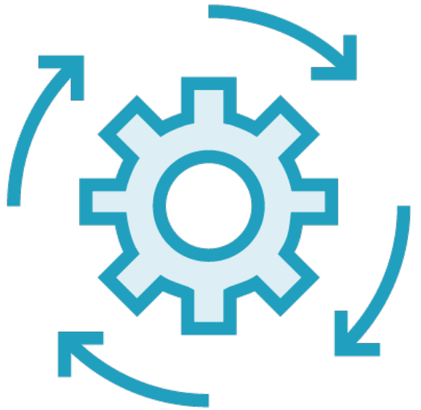
Using Pandas for the First Time

What is Pandas?

Import



Organize



Process

Demo: Pandas

Import pandas

- use pandas to import csv

Data frame created

- csv file is now neatly imported

Pandas transformations

- Practice recoding and pulling sub-frames from data

Using Matplotlib for the First Time

Why Matplotlib?

Easy to use

Most common

Demo: Matplotlib

Import matplotlib

- use library to plot data

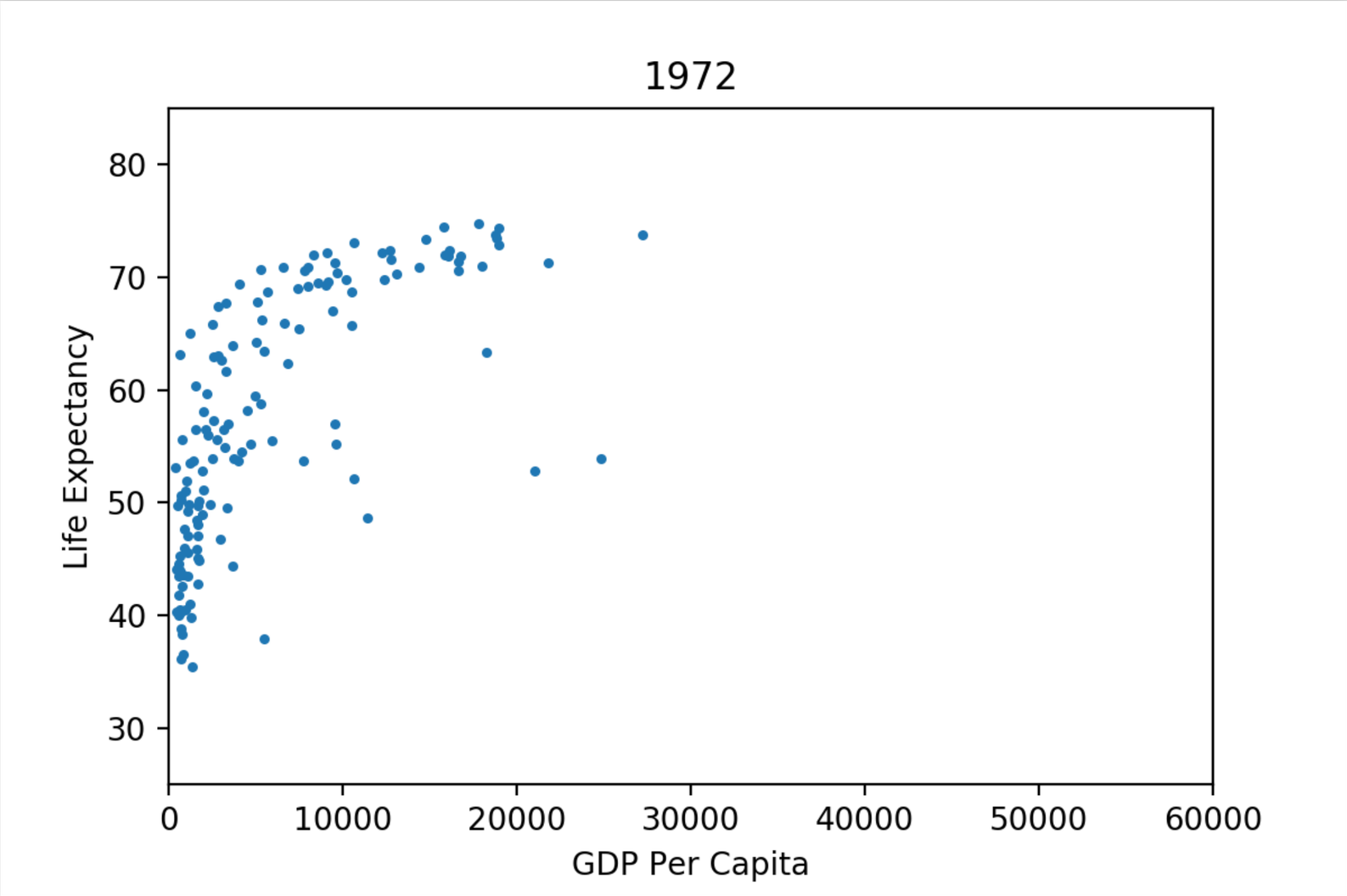
Plot alternatives

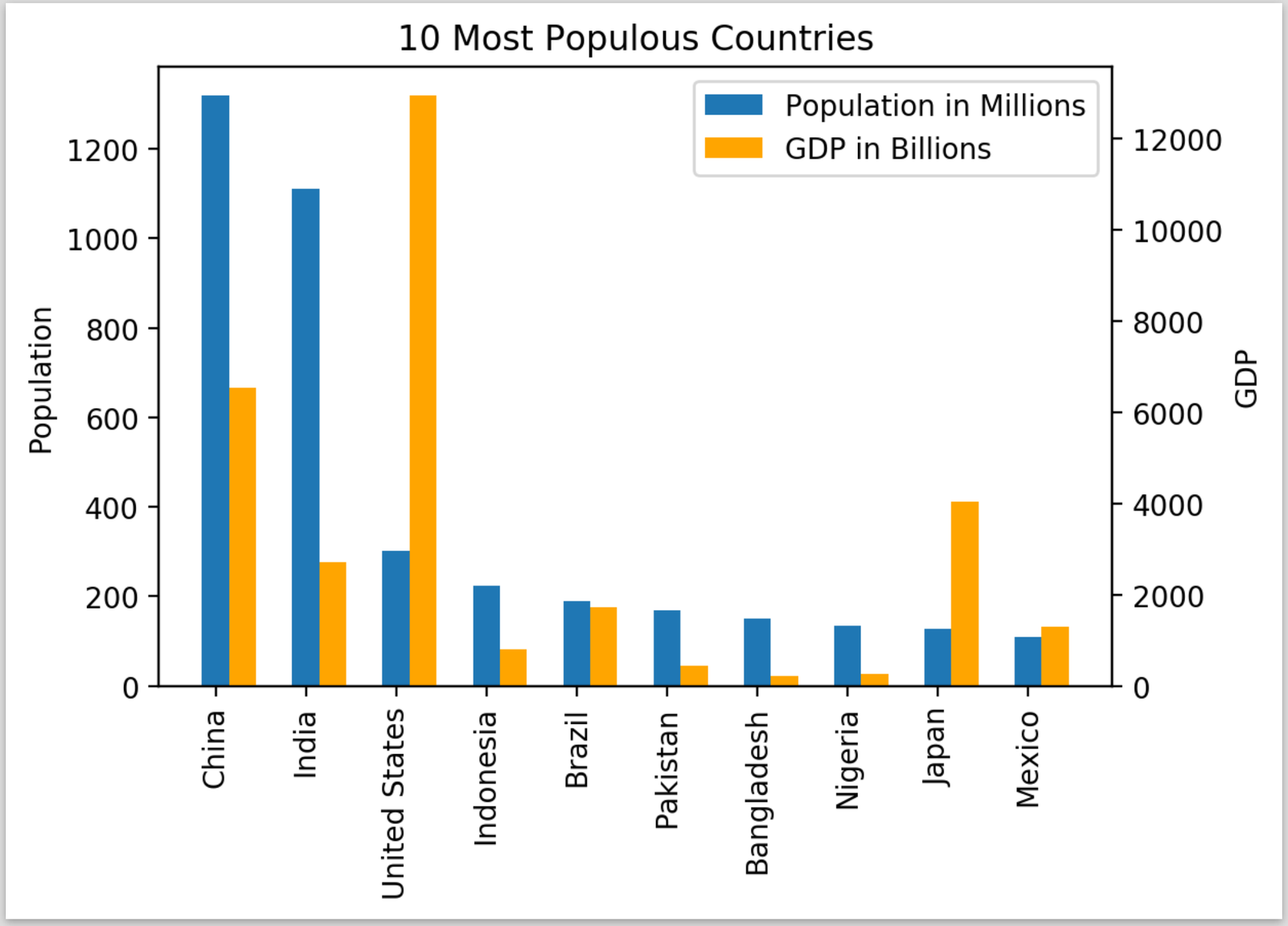
- Bar chart, scatter plot, line plot, help()

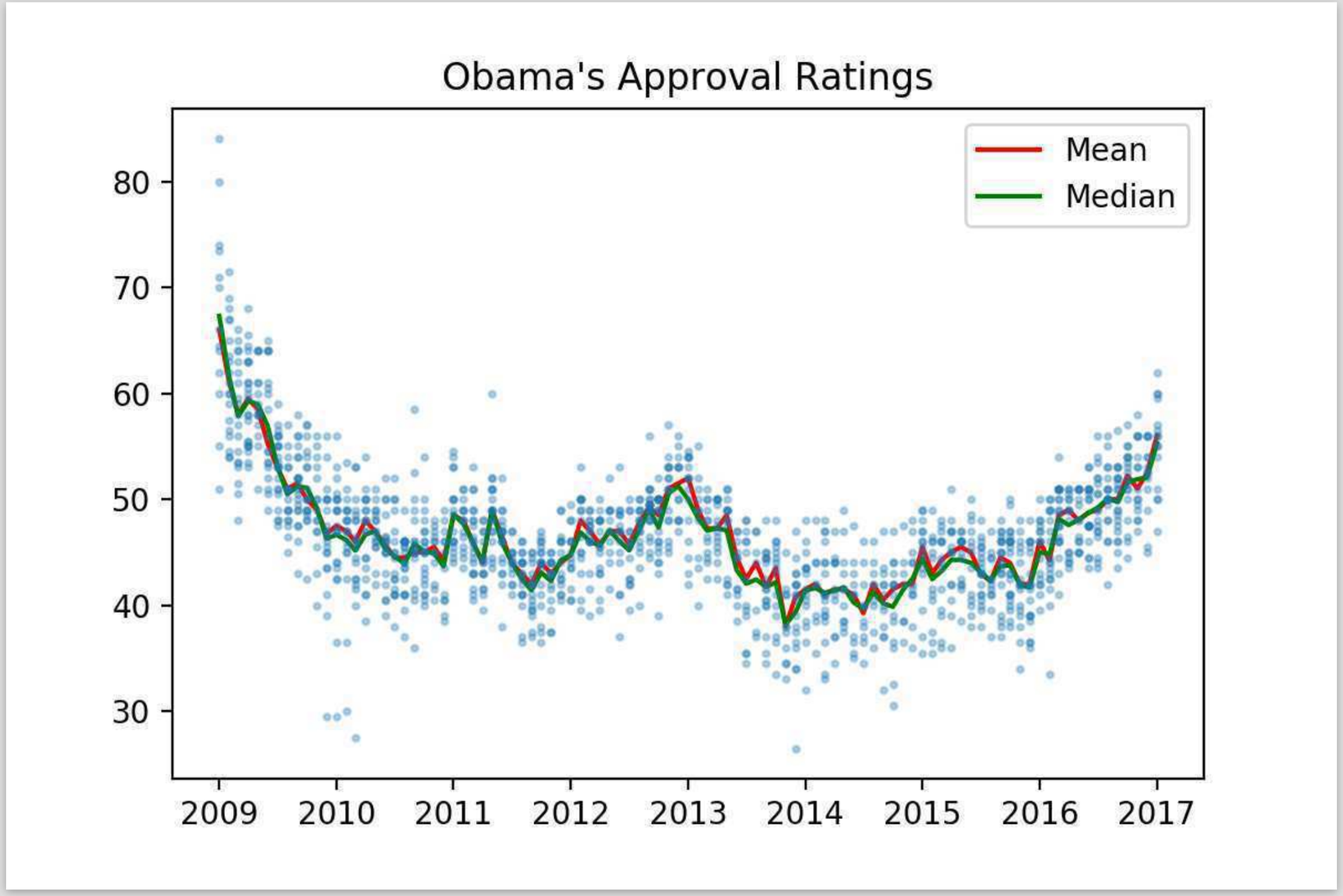
Label plots

- use matplotlib to create titles, label axes, create legend

Examples from matplotlib




10 Most Populous Countries
Population in Millions
GDP in Billions
Population
GDP
0
200
400
600
800
1000
1200
0
2000
4000
6000
8000
10000
12000
China
India
United States
Indonesia
Brazil
Pakistan
Bangladesh
Nigeria
Japan
Mexico


Obama's Approval Ratings
Mean
Median
80
70
60
50
40
30
2009
2010
2011
2012
2013
2014
2015
2016
2017

Where are we located?

The RCC is a unit in the Office of Research and National Laboratories

Regenstein Room 216:
Walk-in lab and Help Desk

Call:(773)795-2667

Central office:5607 S. Drexel Ave
Data center: 6045 S. Kenwood Ave