Data Analysis using Anaconda/Python

Python for Data Analysis

- Data Analysis
  - Manipulating
  - Processing
  - Cleaning
  - Crunching
- Data (structured data)
  - Single dimensional arrays
  - Multidimensional arrays (matrices)
  - Tabular/spreadsheet-like data
    - Relational DBs
    - Tab/comma delimited text files
  - Multiple tables of data related by key columns

Alternatives

- Matlab
- R
- SAS
- Stata

Python Libraries for Analysis

- NumPy
  - Scientific computing
  - Fast, efficient multidim array object & functions
- Pandas
  - Data structures for structured data
  - Reshape, slice&dice, aggregate, select
  - DataFrame object similar to R data.frame
- Matplotlib
  - Plots & 2D visualizations
- Ipython
  - Shell for writing, testing, debugging
  - Notebook
  - GUI console with support for graphics

Python Distributions

3.2. Alternative bundles

Besides the standard CPython distribution, there are modified packages including additional functionality. The following is a list of popular versions and their key features:

ActivePython
: Installer with multi-platform compatibility, documentation, PyWin32

Anaconda
: Popular scientific modules (such as numpy, scipy and pandas) and the `conda` package manager.

Canopy
: A "comprehensive Python analysis environment" with editors and other development tools.

WinPython
: Windows-specific distribution with prebuilt scientific packages and tools for building packages.

Note that these packages may not include the latest versions of Python or other libraries, and are not maintained or supported by the core Python team.

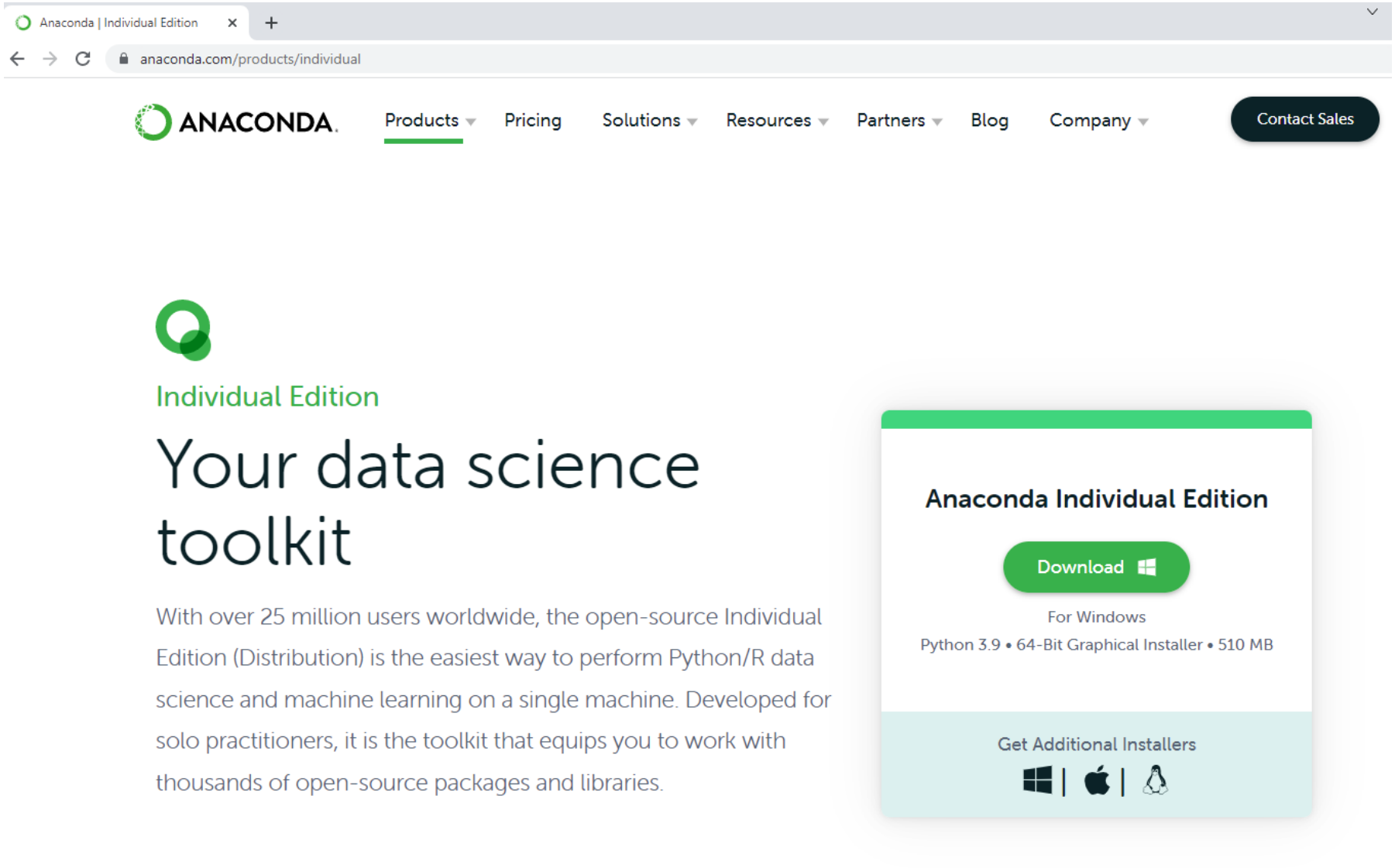
https://docs.python.org/3/using/windows.html

numpy, pandas, scipy

- You can use any distribution of Python that will support numpy, pandas, and scipy
- I will use the Anaconda distribution:
  - Python 3
  - Spyder
  - IPython notebooks

Anaconda Individual Distribution

https://www.anaconda.com/



Anaconda Installation

- When I installed Anaconda, I used these settings (for Windows 10):
  - Install for: Just me (recommended)
  - Destination folder: **C:\Users\rcapra\Anaconda3**
  - Advanced options:
    - Leave unchecked “Add Anaconda to my PATH…”
    - Leave checked “Register Anaconda as my default Python 3.9”
  - At the end of the install, I SKIPPED installing MS VSCode.
- Installation took my computer about 15 minutes
- If asked about network/firewall permissions, I either “Cancelled” or “Blocked” to not allow it.

Anaconda & Spyder

- When finished, start Spyder
- Spyder is an IDE for Python development
  - Includes many useful features, including IPython & debugging