

# An Introduction to Python

Pritam Dalal

# The Anaconda Distribution

- ▶ In this class we will be using the *Anaconda* distribution of Python.
- ▶ A *distribution* is a curated collection of packages.
- ▶ In the case of Anaconda, the packages are curated in the sense that they are all relevant to scientific use-cases, and they are all ensured to work together.

# Installing The Anaconda Distribution

1. Google “anaconda distribution”.
2. Click on the first search result.
3. Download the 3.7 version for your platform (Windows, MacOSX).
4. Follow the installation instructions.
5. We'll have a general discussion as folks install the software.

# A Bit More On Distributions

- ▶ You never hear of a *distribution* of Windows or Mac OSX.
- ▶ On the other hand, there are many different distributions (distros) of Linux, an open source operating system.
- ▶ If you try to install Python, and then all the data science packages separately, it will take time and might not work.
- ▶ The company Continuum Analytics, bundles together Python, as well as all the major science related packages into a distribution called *Anaconda*.
  - ▶ Free to use, but Continuum charges for support. (*freemium*)

# A Brief Overview

- ▶ Python: the programming language.
- ▶ Python 2 vs 3: a while back there was a major revision of the language (from 2 to 3) that made a lot of 2 obsolete. So now there is a rift in the Python world. This won't affect us, and we will use 3.
- ▶ SciPy: a collection of packages related to scientific computing and data analysis.
  - ▶ NumPy: matrix computations
  - ▶ SciPy: also a package, optimization
  - ▶ IPython: interactive wrapper around Python
  - ▶ Pandas: analysis dataframes and timeseries
  - ▶ Matplotlib: data visualization
  - ▶ Jupyter: a notebook interface for IPython

# A Brief Overview

- ▶ Python at its core is a scripting language and an object oriented application development framework (think C#).
  - ▶ In Python, you will interact with object which are instances of classes. The objects will have properties and methods.
  - ▶ In R, at our level of use, there is pretty much data and functions.
- ▶ SciPy turns Python into a scientific computing framework much like R and Matlab.
- ▶ In this class, we are mainly going to use the SciPy ecosystem of packages for the purposes of financial data analysis.

# Where Do You Write Code

- ▶ In this class, we'll mainly be writing code in Jupyter Notebooks.
- ▶ PyCharm is a integrated development environment that we may also use for more involved projects, depending on our needs.
- ▶ There are lots of different alternative IDEs for Python. As compared to R/RStudio, none of the Python IDEs seem as dominant.

# Self-Study Resources

- ▶ *Python for Data Analysis* - Wes McKinney (pdf of first edition is freely available online).
- ▶ *Python Data Science Handbook* - Jake VanderPlas (there is a free online version).
- ▶ *Automate the Boring Stuff* - Al Sweigert (freely available online, there is also a Udemy course).
- ▶ *Python for Data Science and Machine Learning Bootcamp* - this is a Udemy course - it's fine but the examples are boring and the instructor's voice is a little bit grating.
- ▶ My tutorials that will be available on the website.