

# lecture01\_Introduction\_and\_Setup

January 9, 2023

## 1 Topics covered in this course

- The fundamentals of programming in Python
  - Variables and basic data types (int, float, str, bool)
  - Data structures (lists, tuples, dictionaries)
  - Conditional statements (“if/else” else)
  - Iteration (“for” loops)
  - Functions (“def” statement)
  - The “dot” notation (everything in Python is an “object”)
  - Interpreting error messages (almost everything in an error message is useless)
  - Finding answers online (documentation files, Google is your friend, stackoverflow)
- Data processing
  - Data Input/Output (IO)
  - Data cleaning
  - Dates and lags
  - Merging datasets
- Descriptive statistics and data visualization
  - Unconditional (full-sample) statistics and visualization
  - Conditional (subsample) statistics and visualization
- Linear Regression
  - Fundamentals of linear regression
  - Robust linear regression in panel and time-series datasets
- Backtesting investment strategies (academic paper replication)
  - Collecting and cleaning the data
  - Calculating portfolio returns
  - Analyzing portfolio performance

Note that these topics are subject to change, based on how the tempo of the class is going. You will be notified when changes are made.

## 2 Anaconda, Jupyter, and VS Code installation

To install Python, we will actually install a different program, called Anaconda, which comes with Python and many other tools frequently used in Data Science. Anaconda takes a lot more space than Python itself, but the benefit is that it comes with everything you need for this course (and

likely any other course you will ever take on Machine Learning or Data Science). Once you install Anaconda, install Jupyter Notebook and VS Code following the instructions below.

### 2.0.1 Installing Anaconda

1. Go to <https://www.anaconda.com/products/individual>
2. Click “Download” and then select the “64-Bit Graphical Installer” for your operating system
  - I am assuming you all have 64-bit systems
  - If you have a 32-bit Windows system, you also have the option to download a 32-bit Anaconda installer
3. Follow the instructions on the screen
  - For additional instructions on installation, see the “Resources” section below
  - When in doubt, use the default settings suggested by the Anaconda installer

To open Anaconda:

- Windows
  - From the Start menu, click the Anaconda Navigator desktop app
- macOS
  - Open Launchpad, then click the Anaconda Navigator icon

For all of our classes, we will use Jupyter Notebooks, which you can install from inside Anaconda like this:

### 2.0.2 Installing Jupyter Notebook

- On Anaconda Navigator’s Home tab, in the Applications pane on the right, scroll to the Jupyter Notebook (not JupyterLab) tile and click the Install button to install Jupyter Notebook

### 2.0.3 Installing VS Code

This is an IDE (integrated development environment) that can run Jupyter notebooks too. I will be using it during class, but you do not have to (as explained in the following section). I strongly recommend you install it though, since it comes with some very useful functionality when writing code (e.g. autocompletion, integrated terminal etc).

To install VS Code:

- Go to <https://code.visualstudio.com/download>
  - On Windows, click on the “User Installer” “x64” tile under “Windows”
  - On Mac, click on “.zip” “Apple Silicon” if you have an M1 or M2 Mac, or “Intel Chip” if you have an older Mac.
- Follow the installer’s instructions after the download is complete

## 3 How to open the lecture notes

In D2L, click on the “Lectures and Data” tab. You will see a set of subfolders inside that tab. Please replicate that exact subfolder structure somewhere on your computer (say inside a folder called FIN525). Save all the contents of the “data” folder in your “data” folder. Before each

class, save the contents of the “lectureXX” folder for that class into the “lectureXX” folder on your computer where XX stands the lecture number corresponding to that class.

There are (at least) three ways you can open a particular lecture (a Jupyter notebook) on your computer:

- METHOD 1 (recommended):
  - Open VS Code
  - Click “Open Folder” on the welcome page or under the “File” menu
  - Navigate to the folder that contains your lecture notes and hit “Enter”
  - This folder should show up on the right side of the screen. Find the lecture notes you want to open and double click them.
- METHOD 2:
  - Open Anaconda Navigator, find the Jupyter Notebook tile in the pane on the right, and hit “Launch”. This will open a web browser showing a directory structure on your computer. Use that to navigate to the lecture you want to open and click on it.
- METHOD 3:
  - Navigate to the folder containing the lecture you want to open on your computer. Copy the full path to that folder:
    - \* on Windows: double-click your folder and then click the navigation bar at the top of the window
    - \* on macOS: see “Resources” below
  - Open a terminal like this:
    - \* on Windows: from the Start menu, search for and open “Anaconda Prompt”
    - \* on macOS: open Launchpad, then click the Terminal icon
  - In the terminal, type “cd” (without the quotes) and then paste the path to your lectures folder you copied above. Hit enter.
  - In the same terminal, type “jupyter notebook” (without the quotes) and hit enter.
  - This process should have opened a browser showing all the file in your lectures folder. Open a lecture by just clicking on it in this web browser.

## 4 Installing Python packages

Anaconda will allow you to easily install new Python packages (pieces of code written by someone else). There are two methods to do this (though depending on the package, only one of them might work):

- METHOD 1: using “conda”
  - Open a new terminal as described above
  - Type “conda install” (without quotes) followed by the name(s) of the Python package(s) you want to install. Hit enter.
- METHOD 2: using “pip”
  - Open a new terminal as described above
  - Type “pip install” (without quotes) followed by the name(s) of the Python package(s) you want to install. Hit enter. I recommend using “pip” only if “conda” gives you an error.

**Example:**

A package that we will use almost daily in class, is called “yfinance” and it allows us to download data from Yahoo Finance with just one Python command.

To install “yfinance”, open a new terminal and type:

```
pip install yfinance
```

The reason why we use “pip install” here, rather than “conda install” is because the latter results in some errors for some users and it seems like the pip installation is more stable.

Other packages we will use frequently in class:

- pandas
- numpy
- matplotlib

They all come pre-installed with Anaconda.

## 5 Downloading data from WRDS (OPTIONAL)

Wharton Research Data Services (WRDS) is one of the the most widely used database service in academic finance research. It provides access to many different kinds of data, but for this class, we will only use CRSP and Compustat data:

- CRSP contains stock market data (stock prices, returns, volume, etc.). This is the “crspm.zip” file in the “data” folder on D2L.
- Compustat contains accounting data (all data from the three main financial statements). This is the “crspa.zip” file in the “data” folder on D2L.

I have created a class account for you with WRDS. Please use the information in the Syllabus to log in.

The download instructions below are only here for your reference in case you need to download these datasets later on (e.g. for your summer projects). However, for the purpose of this class, I wanted to make sure everyone will work with the same data (and get the same results) so I downloaded the two datasets mentioned above and uploaded them on D2L. They are under the “Lectures and Data” tab, in the “data” folder. The CRSP monthly file is called “crspm” and the Compustat annual file is called “compa”. I also downloaded the Fama-French risk factor file and called it “ff” on D2L. Please download that as well.

### 5.0.1 Downloading monthly stock-market data from CRSP:

- Home → CRSP → Stock/Security File → Monthly Stock File
- STEP 1: Select start & end dates for your dataset
- STEP 2: Select PERMNO and “Search the entire database”
- STEP 3: Click “Select all” to request all variables, or click on just the ones you want
- STEP 4: Leave everything as is, though, I would choose “gzip” for compression to speed things along (you can extract gzip files on Windows with the “7zip” software available for free on the internet)

### 5.0.2 Downloading annual accounting data from Compustat:

- Home → CRSP → CRSP/Compustat Merged → Fundamentals Annual

- STEP 1: Select start & end dates for your dataset
- STEP 2: Select PERMNO and “Search the entire database”. Leave “Screening Variables” unchanged
- STEP 3: Leave unchanged
- STEP 4: Click “Select all” to request all variables, or click on just the ones you want. Leave “Conditional Statements” unchanged.
- STEP 5: Again, leave unchanged if you don’t mind waiting. Use gzip if you know how to extract compressed files.

## 6 For next class

1. Make sure you have Anaconda, Jupyter Notebook, and VS Code installed and running on your computer before class.
2. Create a folder on your computer which contains the same subfolder structure you see in D2L under “Lectures and Data” (as explained in the “How to open the lecture notes” section above.)
3. Open a Terminal, type the following command and hit enter:
  - `pip install yfinance pandas-datareader statsmodels linearmodels`
4. Open a Terminal, type the following command and hit enter:
  - `conda install -y openpyxl xlrd`

## 7 Resources

- Anaconda installation instructions:
  - <https://docs.anaconda.com/anaconda/install/>
- Getting started with Anaconda:
  - <https://docs.anaconda.com/anaconda/user-guide/getting-started/>
- Installing packages:
  - <https://docs.anaconda.com/anaconda/user-guide/tasks/install-packages/>
- Copy folder path on macOS:
  - <https://www.switchingtomac.com/tutorials/osx/5-ways-to-reveal-the-path-of-a-file-on-macos/>
- WRDS login page
  - <https://wrds-www.wharton.upenn.edu/>