ASDS Code Camp – Tuesday Assignment

<u>Installing Python</u>

Whereas R is a programming language used for statistics, Python is a general-purpose programming language which can be used to do many things. It shares some similarities with R, but at root it is more of a "programmer's" language, in which statistics and data analysis capabilities are "bolted on" through attaching additional libraries. The most important libraries for data science are NumPy, pandas and matplotlib.

Like R, Python is completely free. Unlike R, however, there is no single solution for setting up Python. Because we will be using Python for data science, we will use a distribution known as Anaconda, which is specifically designed for this purpose, and contains the most important data science libraries. As well as the language itself, Anaconda also comes with additional "add-ons" which help with the practicalities of doing data science in Python. As with R Studio, the individual user edition of Anaconda is free. To download Anaconda:

1. Go to https://www.anaconda.com/products/individual

2. In the box headed "Anaconda Distribution", click the green "download" box (this should automatically download the correct version for your system).

3. Follow the instructions. (For Windows systems, a pop-up box should appear to download an .exe file. You should click on the "save" box to download the file, which should then execute automatically.)

<u>Using Anaconda</u>

The Anaconda distribution comes with a "navigator" programme, which is essentially a launch screen for the different applications, environments and learning resources associated with Anaconda. It can be accessed via the Anaconda Navigator shortcut, which should be either in your start menu under "Anaconda" (Windows), or your Applications folder (Mac).

Anaconda also provides an IDE (integrated development environment), rather like R Studio for R. It is called Spyder, and can be accessed from the Navigator, or directly through the shortcut in the Anaconda folder.

As well as Spyder, Anaconda is primarily known for Jupyter Notebook. This is an "interactive computing notebook environment", which runs through your web browser. It is similar to an IDE, though more lightweight in terms of its capabilities. It is particularly useful as a format for recording and sharing data analysis, rather like a virtual research notebook. Like Spyder, it can be accessed either through the Navigator or its own shortcut in the Anaconda folder. We will work with Jupyter Notebook during Friday's class.