Fundamentals of Statistics and Data Science Workshop Prerequisites

Prior to the workshop, you must be familiar working with Jupyter notebooks and have your own Python 3.x (along with pandas, numpy, matplotlib, seaborn, statsmodels, scipy, & Jupyter) and R environments installed.

Information is provided below for this process if you have not already completed it. If you are already comfortable working with Jupyter notebooks and have your own Python 3.x and R environments with required libraries, you do not need to complete these steps.
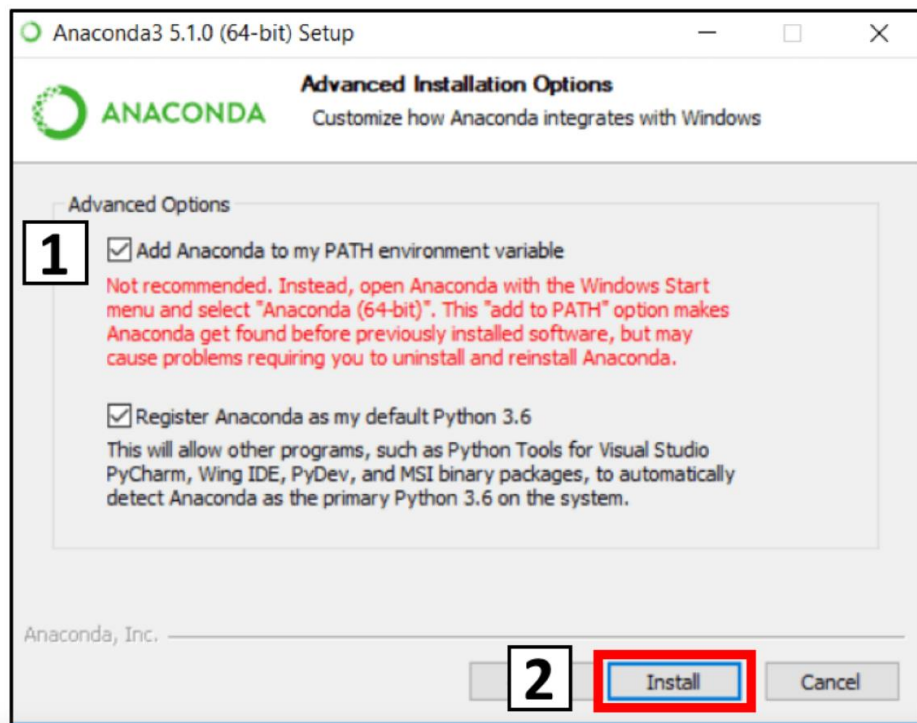
Background

Jupyter notebook is an open-source web application that allows you to create and share documents that contain live code, text, and visualizations. It supports many programming languages including Python and R.

Anaconda distribution is an open source distribution that allows you to quickly install and manage many data science libraries and IDEs. We will use it to install Python, R, Jupyter notebooks and related libraries.

Install Jupyter Notebook and Python 3.x with Anaconda

1. Download and install Anaconda: https://www.anaconda.com/download/
   a. Make sure to select the 3.x version versus the 2.x version
2. If you are an Administrator on your machine, add the path directly during the installation by checking the "Add Anaconda to my PATH environment variable." It is not the default option since many people may not have administrative privileges, but you should be an administrator on your OSIsoft machine. This will allow you to run python and anaconda commands from the command prompt and PowerShell.

   a.
3. Otherwise, you can set the path yourself manually afterwards using the command prompt with the SETX PATH command as shown in this video segment.

4. For the other choices, the defaults or your preferences should be fine. Keep in mind this may take an hour to fully install.

Install R Essentials Through Anaconda

5. To install the R packages using Anaconda, follow these steps:
    a. Open an administrative command prompt or PowerShell window
    b. Type the following command in the prompt:
        conda install -c r r-essentials
    c. After the environment is solved, you will need to enter Y to continue
    d. This will install the packages that are most commonly used in R. Again, keep in mind this may take some time to fully install.

Learn How to Use Jupyter Notebooks with your new Anaconda Installation

6. Read and follow along with Page 1 of the Jupyter Notebook Tutorial (You do not need to do the analysis shown on Page 2): https://www.kdnuggets.com/2018/05/jupyter-notebook-beginners-tutorial.html