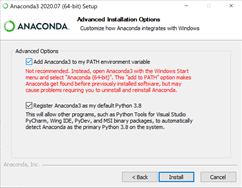
### **TASK AID: How Download, Install, and Use Anaconda, Jupyter Notebooks, and Python**

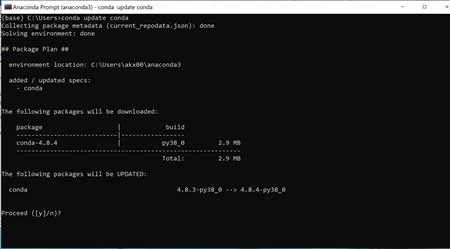
Students are encouraged to install Python via the Anaconda distribution since it includes a pre-configured collection of the most commonly used packages they'll encounter throughout the program, such as matplotlib, numpy, pandas, and scikit-learn. Detailed instructions are as follows.

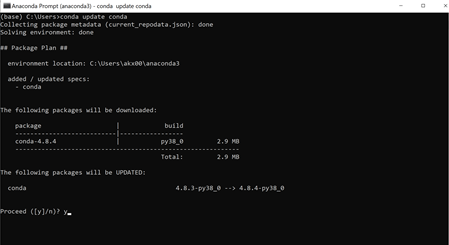
1. [Download](https://www.anaconda.com/download/) and Install Anaconda.
   1. Overview
      * Windows Users: Please use the standard Graphical User Interface (GUI) installation (link with instructions below). A GUI will allow you to proceed with the install via the typical point-and-click interface.
      * Mac Users: We recommend first trying to install Anaconda on your computer using a command line interface (CLI) application called Terminal (which executes a programming language called Bash). If you have any issues installing via the CLI, please instead use the Graphical User Interface installation (refer to links and notes below). A GUI will allow you to proceed with the install via the typical point-and-click interface.
   2. [MacOS Users - Bash Installation Notes](https://docs.anaconda.com/anaconda/install/mac-os/#using-the-command-line-install)
      * Please note that the Anaconda installation instructions recommend verifying data integrity with SHA-256. The instructions on the page for this are not completely clear. We recommend handling this as a separate/optional step during class, and referencing [this resource instead](https://conda.io/projects/conda/en/latest/user-guide/install/download.html#cryptographic-hash-verification).
   3. [MacOS Users - GUI Installation Notes](https://docs.anaconda.com/anaconda/install/mac-os/)
   4. [Windows Users - GUI Installation Notes](https://docs.anaconda.com/anaconda/install/windows/)
      * NOTE: If using the GUI installer, recommend choosing both advanced options, to avoid issues with class exercises:



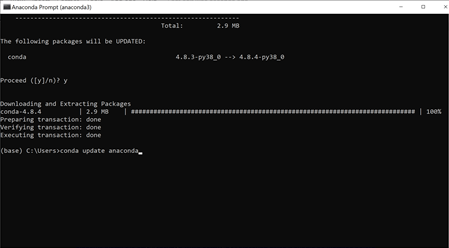
1. Once installed, use Anaconda's package manager, Conda, to verify that you have the most current updates. A package manager is an important tool that automates the process of installing, updating, and removing packages.
   1. Open Anaconda Prompt



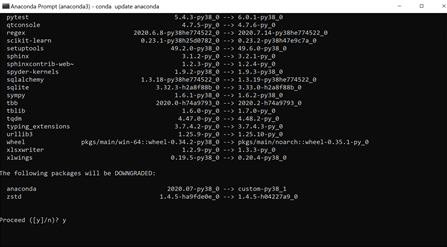
* 1. Type “**conda update conda**” and Hit Enter. It may take a moment to run.
  2. Review, type “**y**” and Hit Enter. You will know if it successfully executed if it proceeds to “Downloading and Extracting Packages”, completes this, then returns to the command line.



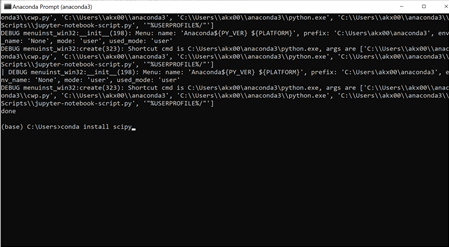
* 1. Then type “**conda update anaconda**” and Hit Enter. It may take a moment to run.



* 1. Then type “**y**” and Hit Enter. This may then take a moment to run. You will know if it successfully executed if it proceeds to “Downloading and Extracting Packages”, completes this, then returns to the command line.

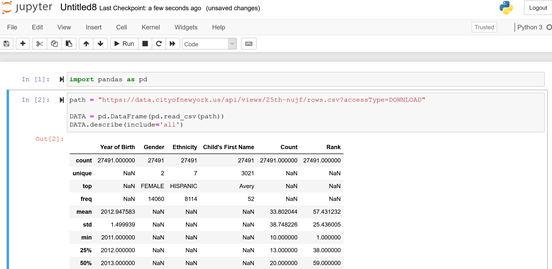


1. While Anaconda takes care of compiler dependencies for common packages, we recommend students install the scipy package while setting up Anaconda since it should be installed before packages that they will need later on in the course, such as Seaborn.
   1. To install, type “**conda install scipy**” and Hit Enter



1. Using Anaconda, Jupyter Notebook, and Python Syntax
   1. **Example:** Install non-standard-library Python packages, like [pandas](https://pandas.pydata.org/pandas-docs/stable/install.html#installing-from-pypi) and [python-Levenshtein](https://pypi.org/project/python-Levenshtein/)
      * In Anaconda Prompt, type “**conda install pandas**” and Hit Enter
   2. **Example:** Opening a Jupyter Notebook from Anaconda Prompt
      * First, recommend navigating to the directory where you are keeping your project files. This will let you flexibly code directories in your Python scripts.
        + To do this, type “**dir**” and Hit Enter to see what your subfolders are.
        + Then type “**cd**” and the name of the subfolder to navigate to that (no quotation marks needed) then Hit Enter.
        + If you need to go back up a level, type “**cd ..**” then Hit Enter.
      * Once you are in the directory you need, type “**jupyter notebook**” and Hit Enter. Jupyter Notebook will then launch.
   3. **Example:** Read in a CSV file: for example, [the most popular baby names in NYC dataset, here](https://catalog.data.gov/dataset/most-popular-baby-names-by-sex-and-mothers-ethnic-group-new-york-city-8c742) -- perhaps, try using the [csv](https://docs.python.org/2/library/csv.html#module-contents) package in addition to [pandas.read\_csv()](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.read_csv.html)
      * In Jupyter Notebook, in the first cell type “**import pandas as pd**”. Then on the menu bar choose Cell -> Cell Type -> Code
      * Then in the menu bar choose Insert -> Insert Cell Below. Copy and paste the following lines into the cell. Make sure this is also the Code Cell Type.

|  |
| --- |
| **path = "https://data.cityofnewyork.us/api/views/25th-nujf/rows.csv?accessType=DOWNLOAD" DATA = pd.DataFrame(pd.read\_csv(path)) DATA.describe(include='all')** |



* 1. Example: Start basic analysis of the CSV using Pandas (for example, find the most popular first letter for baby names for boys vs. girls, or convert the data into a histogram)
     + In your Jupyter Notebook try out some basic pandas syntax, take a look at cheat sheets online such as these to start, and be liberal in your Google searches to get what you need (for example: “pandas substrings” as a Google query)

<https://pandas.pydata.org/Pandas_Cheat_Sheet.pdf>

<https://www.dataquest.io/blog/pandas-cheat-sheet/>

* + - For example:

|  |
| --- |
| **DATA['First Initial']= DATA["Child's First Name"].str[:1] DATA[DATA['Gender'].isin(['FEMALE'])]['First Initial'].value\_counts() DATA[DATA['Gender'].isin(['MALE'])]['First Initial'].value\_counts() pd.crosstab(index=DATA['First Initial'], columns=DATA['Gender'])** |

* 1. Take a bit of time to also review the Conda [user cheat sheet](http://know.continuum.io/rs/387-XNW-688/images/conda-cheatsheet.pdf).