INFO 2950 hw0

4 September 2020

Please submit your completed `hw0.ipynb` file by 10 September 2020, 11:59 PM EST on CMS.

Welcome

Welcome to INFO 2950! This week is just about getting everyone set up and ready to learn.

By next discussion, you need to know how to open, edit, and save Jupyter Notebook (`.ipynb`) files on your computer.

If you need help getting set up beyond discussion, please make sure to drop into the study hall when it is staffed so that you can get individualized help from a TA. Study hall is *always* staffed the hour right after each discussion, so it's a great time to go in and get started on your homework with classmates. You're also welcome to use the study hall with classmates anytime, staffed or not!

If you cannot or prefer not to use the study hall, you can also post on Campuswire (entry code 4334).

Virtually everything you need to know about this course can be found in the Student Handbook.

Weekly discussion agenda

Discussions will typically be structured as follows:

1. A brief opportunity to **ask questions** about the course content or logistics *(about 5 mins)*
2. A **group discussion** about a topic related to data science, first in small groups via breakout rooms, then all together as a class *(about 20 mins)*
3. A **guided tour of the latest hw** (released Friday mornings), with an opportunity to ask clarifying questions *(about 25 mins)*. If no homework is released that Friday, we will instead use this time to discuss final project-related matters.

Today's discussion question

What do you think it takes to perform "good" data science?

Think about various definitions of "good" here. Think about your conception of "data science".

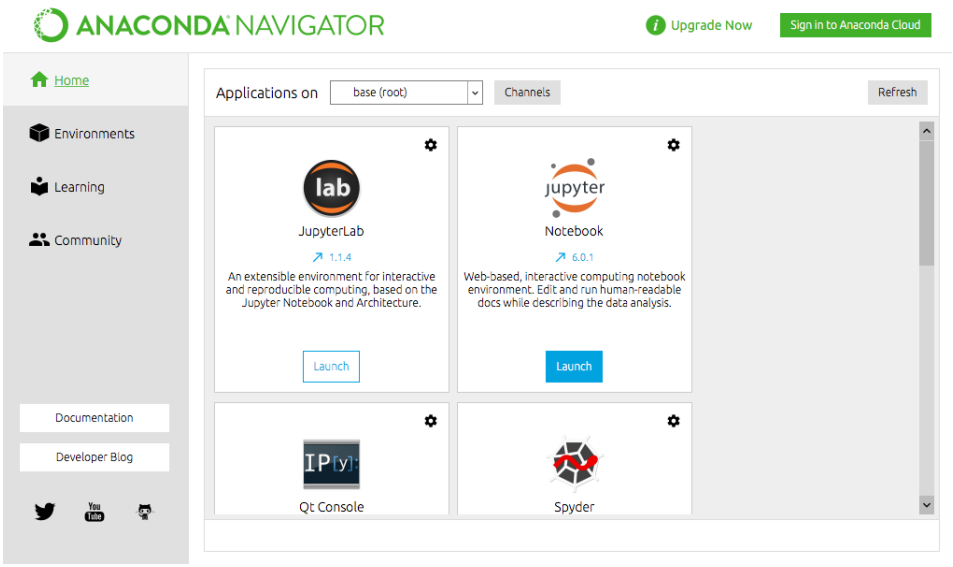
We'll start the discussion with 3-4 people per breakout room. While you're in there, add your thoughts to the Google Doc linked by your TA in the Zoom chat. We will come back together to go over and refine our thoughts as a big group.

Installing Python and setting up Jupyter Notebooks

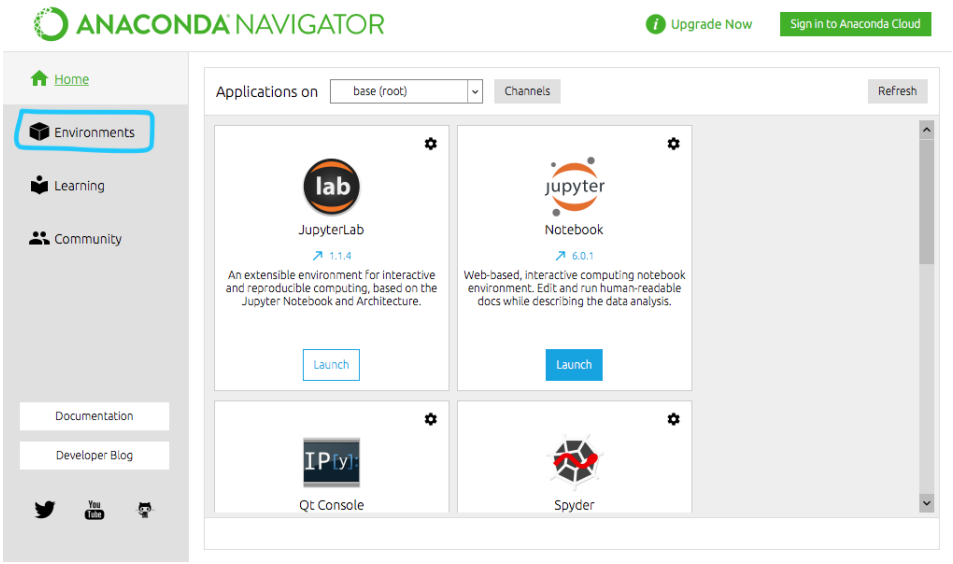
In this course, we will only interact with Python through Jupyter Notebooks. Jupyter Notebooks are files with the extension `.ipynb`; they're documents where you can interactively execute code and also write text around the code, making it perfect for working on homework or presenting project reports.

Even if you already have Python 3 installed on your system, please go through the steps below. We want to make sure everyone in the class has a consistent Python installation to avoid situations that are difficult to troubleshoot.

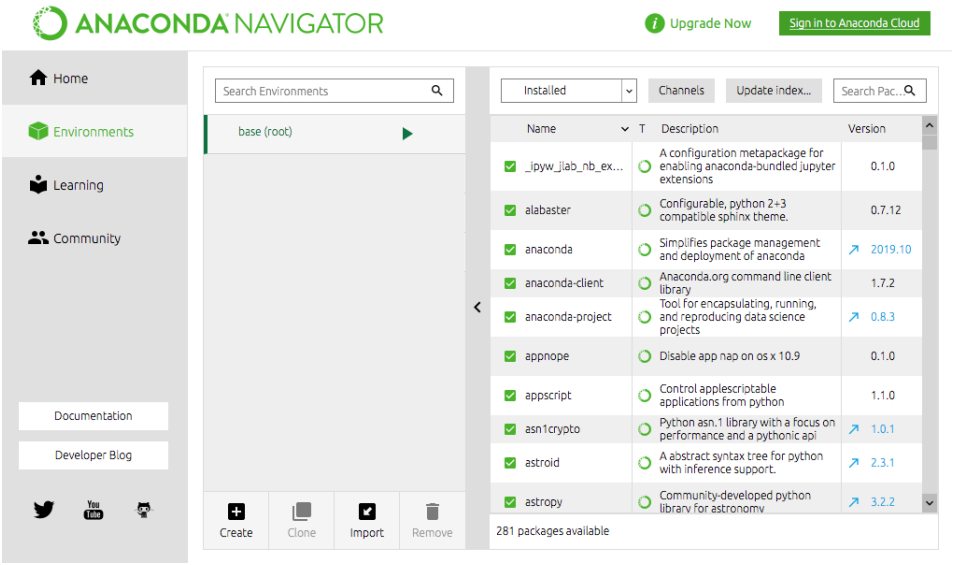
1. Anaconda is a Python installer and package manager that will help us get Python and important packages installed in a (hopefully) painless way. Start by following these instructions to **install Anaconda** for your operating system. If you already have Anaconda installed, please be sure to update to the latest version.
2. **Open the Anaconda Navigator app**. The main app window should look something like this:



3. In the menu on the left, **select the "Environments" tab**:



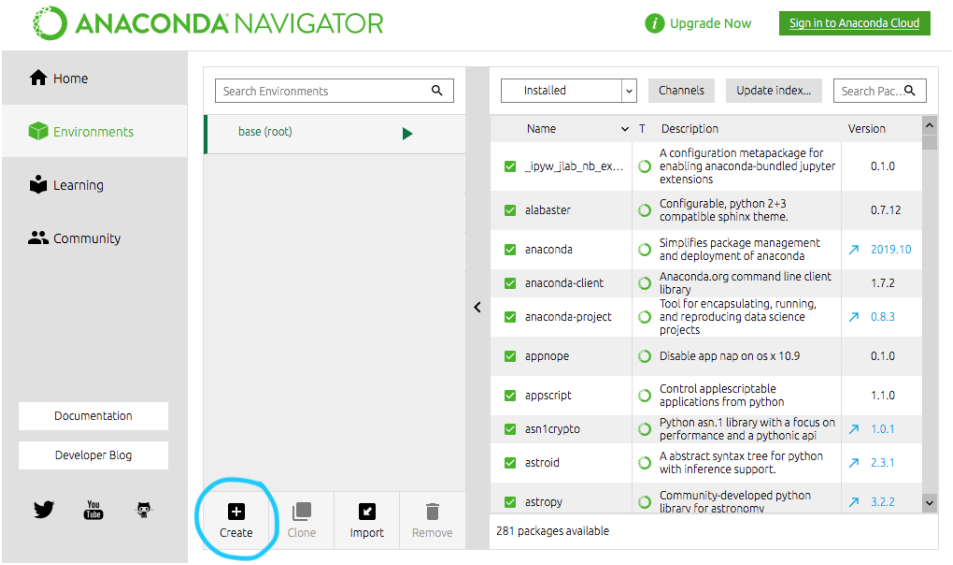
You'll be taken to a window that looks like this:



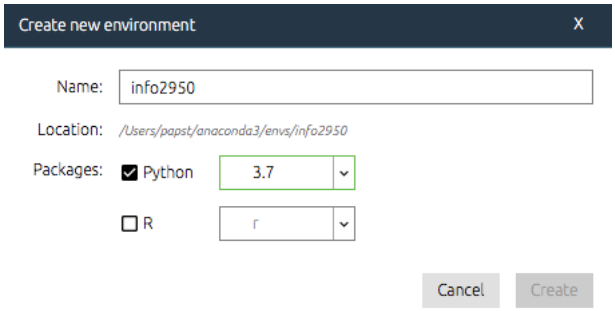
You will notice there is already one "base (root)" environment. We're going to set up a new environment for our course-specific installation of Python to make sure that everyone's setup is consistent, whether or not you've installed Python on your computer before.

Basically, we're going to make an `info2950` "box" into which we will put all of our course software tools. When you want to access these tools, you have to tell you computer to look into the box for them (it will not know where they are by default). More on this later.

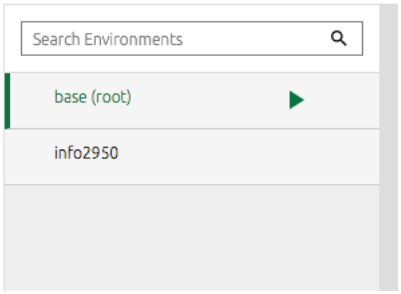
4. **Click the "Create" button** to make a new environment:



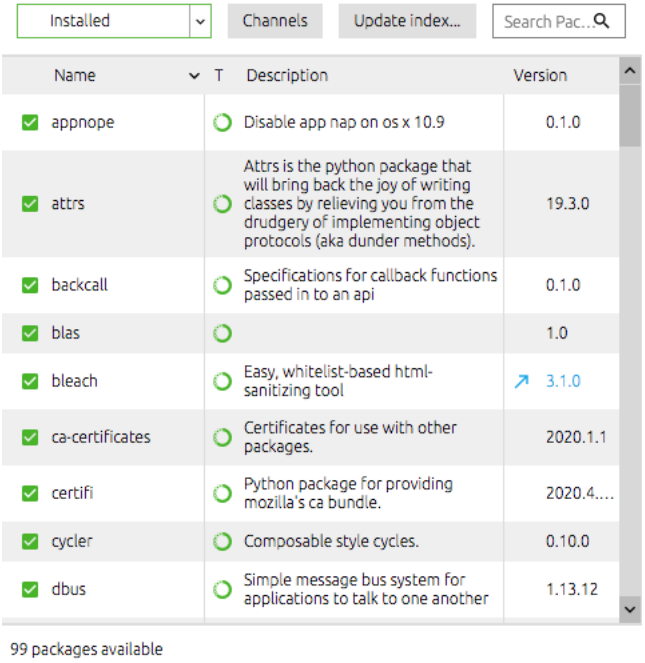
A dialogue box will open. Fill it in like this:



Click create. You should now see a new environment called “info2950” in your environments list:



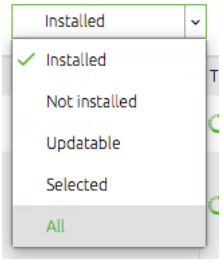
Click the “info2950” environment name. You should notice that window pane on the right refreshes:



This is the list of packages currently installed in your “info2950” environment. **Since we set up a *new* environment, it may not have any Python packages you may have installed for other courses.** Instead, it will have whichever default packages Anaconda thinks most users want in a new environment.

5. For our course, we need to **install a few packages** that don't come in a new Anaconda environment by default, so let's install them now.

   In the top left of the packages pane, there is a dropdown menu that reads "Installed". From this menu, choose "All":



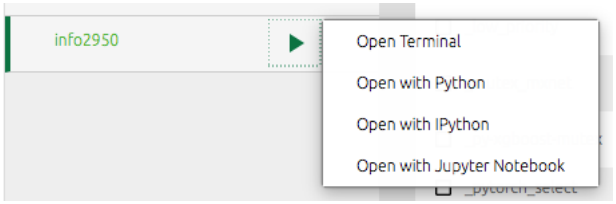
   Using the search bar in the top right of the packages pane, search for the `jupyter` package. You may have to click the "Update index..." button to refresh the package list. You will see a number of search results, but look for the one that whose package name is just `jupyter`; its description should read "Jupyter metapackage. install all the jupyter components in one go". Click the checkbox next to the package name and then click the "Apply" button in the bottom right of the packages pane (only appears when at least one package is selected). This will install the `jupyter` package—it may take a few moments.

   Repeat the above process to also install `pandas` and `matplotlib`.

   In the future, if you install a package while a notebook is running, you will need to restart the notebook before it can access the new package contents.

6. Now we just need to know how to **launch a Jupyter Notebook file**, because unfortunately it's a little more complicated than just double clicking the file...

   To open Jupyter Notebooks within the context of our course-specific install, we need to launch Jupyter from within our "info2950" environment. Select the "info2950" environment from the environment list. You should see a triangular button appear by the environment name. Click this button and you will see a menu pop up:



   Select "Open with Jupyter Notebook". You'll see a Terminal window pop up, followed by a new browser tab that takes you to a list of files on your system. By default, this will be in your computer's "home" directory. You can now navigate to wherever the Jupyter notebook you downloaded lives on your computer, and then double click it (in the Jupyter browser) to open.

hw0 Jupyter Notebook

Use the instructions above to open the `hw0.ipynb` notebook you downloaded from CMS in the "info2950" environment and follow the instructions in the notebook.

To submit your homework, simply upload your `hw0.ipynb` to CMS. **Before submitting, make sure all cells have been executed so the results are visible; TAs will not execute your code when grading.**