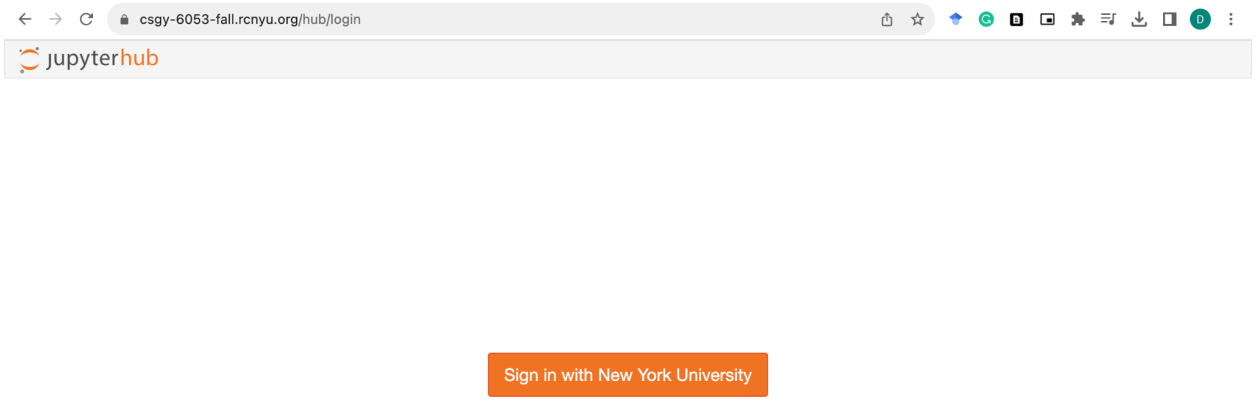
Foundations of Data Science - Homework Instructions

Introduction
The instructions below are written with the goal of getting you comfortable working with the process for completing assignments in the course this semester. While the instructions below have some specifics to the first assignment, you can expect these instructions to generalize to the remaining assignments for the semester. Be sure to read this document in its entirety. **Important: Do not install any additional libraries in your JupyterHub environment. All of the libraries that you need for this course have been installed. Installing additional libraries may lead to problems using JupyterHub and you will be responsible for fixing your environment.**

Accessing the Assignment
As described in the first lecture, we will be using JupyterHub as provided by NYU High Performance Computing (HPC). You can access the JupyterHub environment for this class by visiting the following url:

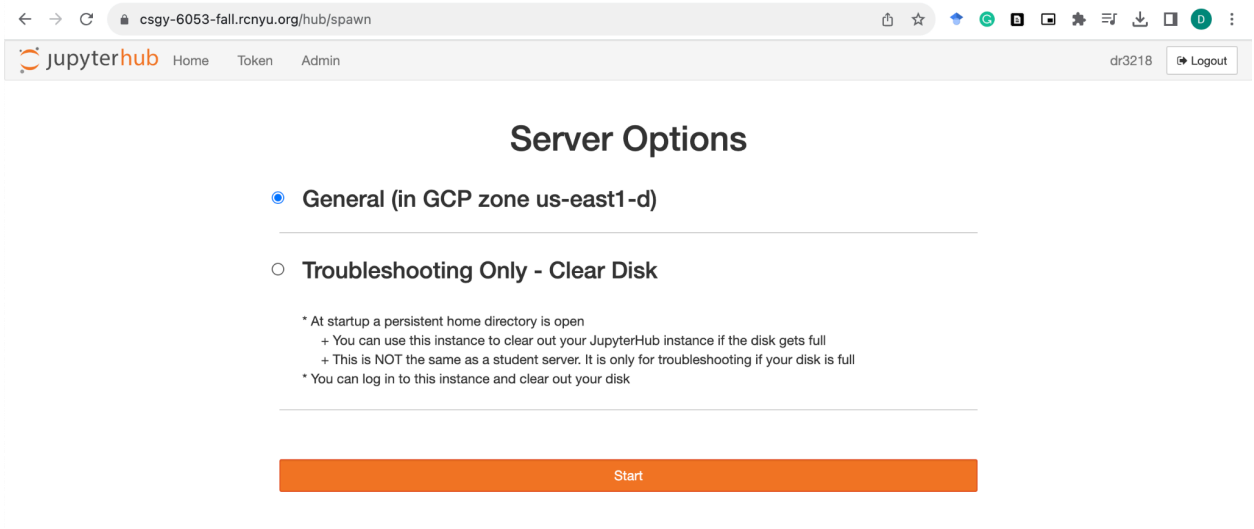
https://csgy-6053-fall.rcnyu.org/

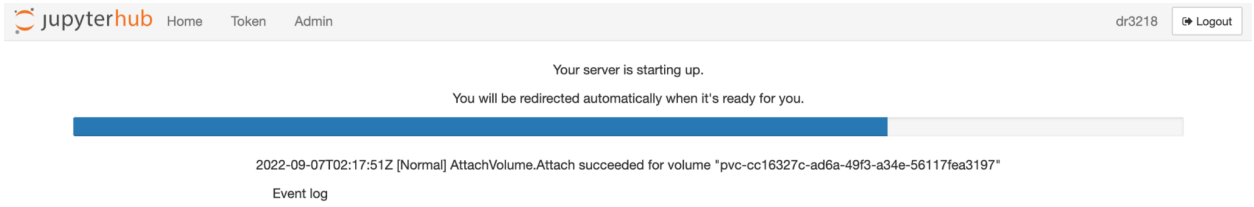
You will be presented with a sign-in page similar to the one below when visiting this url:



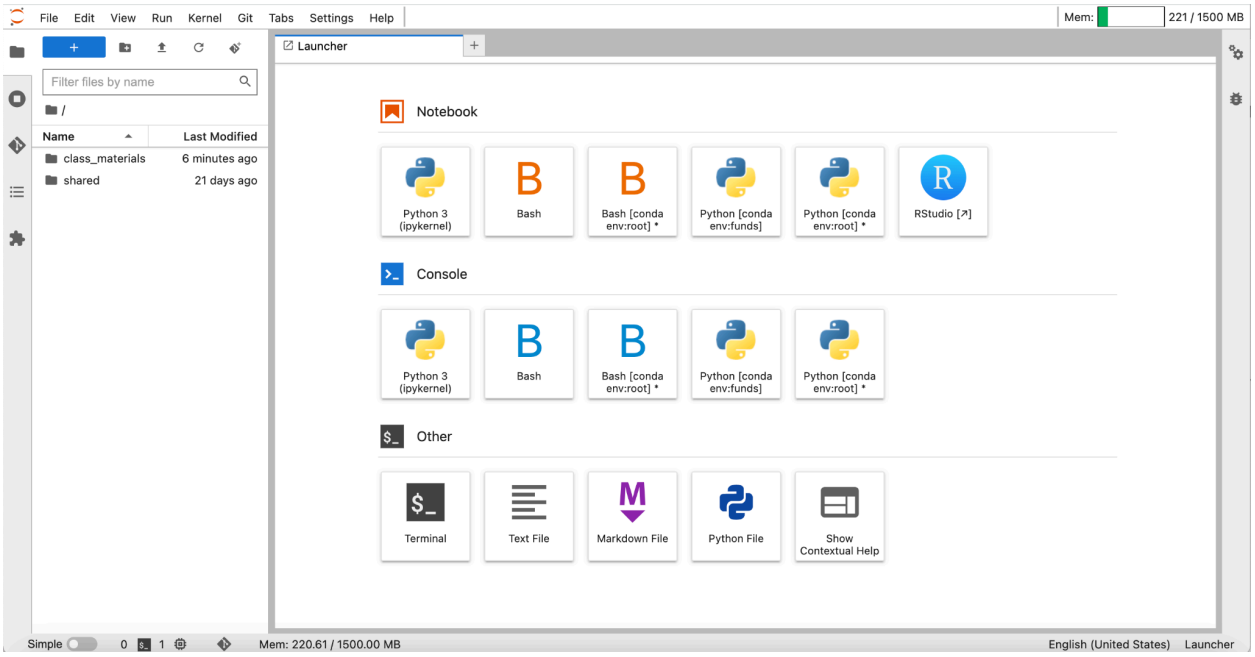
Click on the orange button to sign in with your NetID using the normal process. **Don't use any other Google accounts that you might have access to as you have only been granted access to the site under your NYU NetID**. You will be taken to the following page after entering your NYU credentials.

Click the **Start** button. This will start up your server. You may have to wait a few moments for this process to complete (seeing a page similar to the one below):





When the server launches you will see a page similar to the one below.



Navigate the file browser to **class_materials/homework/hw1**. You will see a file named *hw1.ipynb*. This is the file in which you will complete your assignment.

Setting the Kernel

From the menu bar, click on the **Kernel** tab, then **Change Kernel…**. From the drop-down menu select the option **Python [conda env:fnds]** and **click on the checkbox next to *Always start the preferred kernel*.** Then click on the **Select** button to set the kernel.

Starting the Assignment

1. Run the cell at the top of the notebook for initializing Otter-Grader.

```
In [ ]: # Initialize Otter
        import otter
        grader = otter.Notebook("hw1.ipynb")
```

Otter-Grader is a tool to facilitate grading Jupyter notebooks.

- This can be done by first clicking anywhere inside of the cell
- Then, click the **Run** button at the top of the notebook
- When the cell has been run, a number will appear inside of the brackets next to **In**, as demonstrated in the screenshot below

```
In [1]: # Initialize Otter
        import otter
        grader = otter.Notebook("hw1.ipynb")
```

This serves as confirmation that the cell was run successfully. **Please do not edit this cell or else it may impact your submission process and, therefore, your evaluation.**

2. Fill in your name and any group members that you had when completing the assignment.
3. Run the cell above the first question with the following content:

```
In [ ]: # Set random seed and ensure plots are displayed in the notebook
        import numpy as np
        np.random.seed(5)

        %matplotlib inline
```

Do not edit this cell. The cell above will set a random seed so that your results are reproducible and will allow your plots to be displayed properly in the notebook.

One last cell to leave alone: avoid editing anything under the **Submission** section.

Completing the Assignment
The assignment questions are listed below the cell above. Proceed to working on the questions.

In this course, you can expect that your solutions to the questions will consist of both code and text.

- While collaboration on homework assignments is allowed, everyone must submit their own work. **No group submissions are allowed**. **This means that the code and words that you submit for your solutions, must be your own.**
- You must include the code that you have written to answer the questions (not just resulting plots or tables).
- Feel free to utilize whichever Python modules that are available to you (both standard ones and those that have been included in your environment for this course (e.g. NumPy, PyMC, ArviZ, etc))
- You must answer the questions that are posed and not simply use the code to substitute for commentary. **Be sure to answer all parts of the question that have been presented for full credit.**

While a Markdown cell and Python code cell are provided for you, feel free to add additional cells but be sure that you keep code/text for each question in the area designated for it. (**Do not place code relevant to Question 1 in the area designated for Question 3**). If you are unfamiliar with Jupyter notebooks, this tutorial will prove helpful. (You can skip over the part which discusses installation).

Submitting the Assignment
Before running the final cell in the notebook to export the notebook, make sure that you have saved your notebook's contents. This can be done by clicking on the disk icon

in the top left of the notebook.

Run the final cell in the notebook:

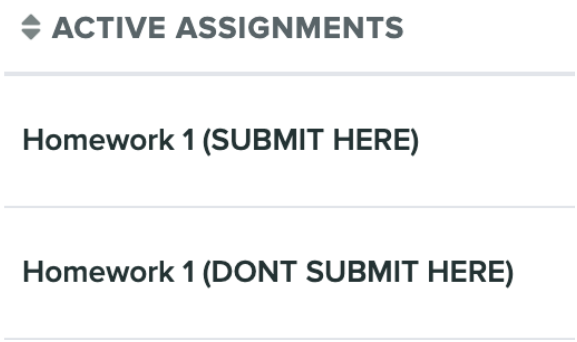
```
In [ ]: # Save your notebook first, then run this cell to export your submission.
        grader.export()
```

This cell will generate a link to download your submission:

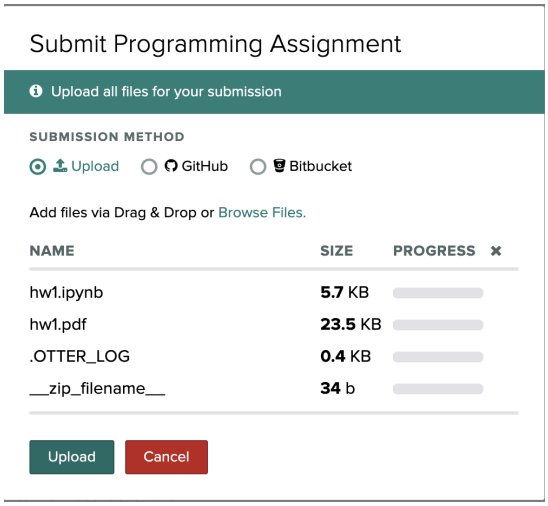
Your submission has been exported. Click here to download the zip file.

Click on that link and take note of where you save the generated zip file. **Be sure to logout out of JupyterHub at this point (and, generally, whenever you are no longer using the server)**.

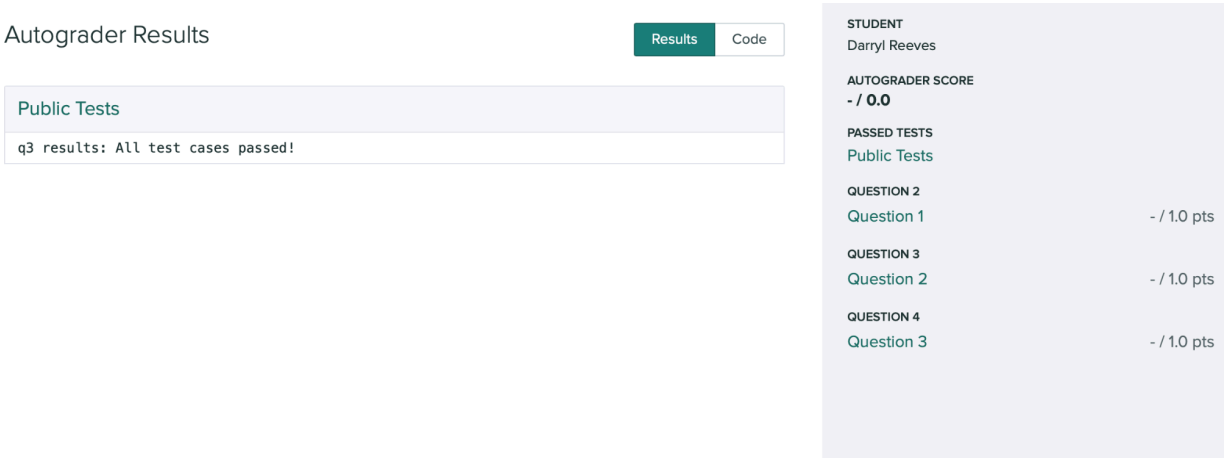
Navigate to Gradescope site for this course via Brightspace (its linked to the front page of the course site). You will see two assignments for each homework assignment. Be sure to submit to the one that has **SUBMIT HERE** in the title.



When submitting your solutions, navigate to where you downloaded the zip archive from JupyterHub and submit the **entire zip archive**. The upload screen will appear as below.



When Gradescope finished processing your file, you should see a screen similar to the one below.



At this point, confirm that you also have a submission for the assignment with (**DONT SUBMIT HERE**) in the title. This submission will be made automatically assuming your manual submission was performed properly.

The last step in the submission process is to tag the pages of your submission for the assignment with (**DONT SUBMIT HERE**). This will ensure that each response in your solution is properly associated with the assignment question being answered. You can find instructions on how to properly perform this tagging process here. To get to the tagging page, click on your submission and find the button **Reselect Pages** at the bottom of the screen. Be sure to do this as part of all of your submissions. **This tagging step must be completed in order to receive full credit on your submission.**

Submission Checklist

- Pages are tagged based on question

- Code is completely readable. (If we can't see the code in the PDF, you will be evaluated as if it was not provided).
- All written commentary must be in Markdown format. **No credit will be given for answers to questions that are written as comments.**
- All code is executed and tables and plots are visible **in the PDF**. Only the PDF will be evaluated, so be sure to check that your PDF represents all of the content that you want evaluated.