COSC 526 - "Start Here" Document

January 22, 2021

This document outlines how to install and use most of the programming tools that we will be using this semester. Completing all of the steps in this document is a prerequisite to starting the first assignment.

---

2 Git & GitHub

In this class, we will be using git & GitHub to distribute & collect assignments as well as other class materials (e.g., slides, code, and datasets). For this lecture, we will only be covering how to download repositories from GitHub (called cloning in git terminology). Next lecture we will go more in depth into both git and GitHub, including how to submit your assignments.

2.0.1 GitHub Account

The first thing you need is a GitHub account. If you don't already have one, go to the GitHub Sign Up page and create a free account. Once you have made your account, send your GitHub username to the email address written on the board. Once you have been given access by the class instructor, you should be able to view this document on GitHub (2020S-COSC526 > Assignment01 > StartHere.ipynb).

Note: if you add your utk.edu email address to your GitHub account, you will eligible for the GitHub student pack, which includes tons of free stuff (e.g., free premium GitHub account with unlimited private repos, free AWS credits, and free cloud hosting).

2.0.2 Git

Now that you have a GitHub account, you need git. - If you are using Windows or Mac, you can download the GitHub Desktop application to use git via a GUI. - You can find GitHub's installation help here. - Once you have downloaded and installed GitHub Desktop, go ahead and login with your new account. - If you are using Linux, you must install the command-line interface (CLI) of git (if you don't know how to do this for your linux distro, please ask for assistance).

2.0.3 Class GitHub Repository

Using either the GitHub Desktop app or the git CLI, clone the class repo: https://github.com/CISC879-BigData/2020S-COSC526. - Step-by-step instructions for cloning with GitHub Desktop can be found here.

---

3 Anaconda

In this class, we will be extensively using Python (specifically Python 3.7) as well as many Python packages (e.g., jupyter notebooks, pyspark, numpy, scipy, scikit-learn, and pandas). To ease the installation/use of these programming tools and ensure that the entire class is using the same environment, we will all be installing and using Anaconda. Anaconda is a Python distribution that includes many popular packages by default and makes installing additional packages easy. You can find more information about Anaconda here.

3.0.1 Installation

To install Anaconda (Distribution, not Enterprise or Cloud), follow the installation instructions for your operating system: - Windows - Mac - Linux

---

4 Jupyter Notebooks

In this class, all of your assignments will be done in Jupyter Notebooks. Jupyter Notebooks allow you to create and share documents that contain live code, equations, visualizations and narrative text. We will give you a Jupyter Notebook that contains the problems that you must solve (in fact the document you are reading right now is a Jupyter Notebook). Some of the problems we give you will already contain some template code that you can build off of.

Note: For you own work beyond this class, JupyterLab is a great alternative to Jupyter Notebooks for portable code that can be edited and executed via a web-browser.
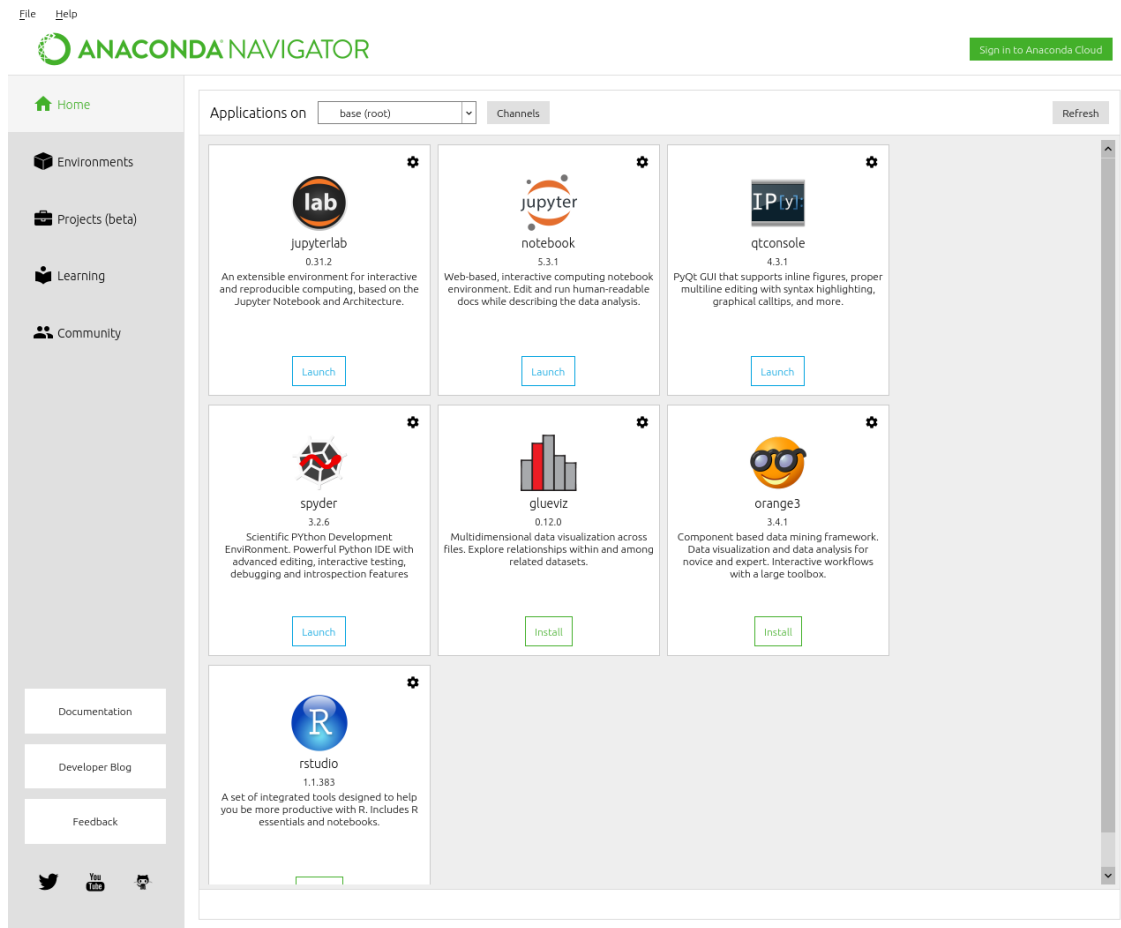
4.0.1 CLI Usage

If you are familiar and comfortable working with at the command line, you can launch jupyter by running `jupyter notebook` inside of the clone'd repo.

If you aren't familiar with the command line, go to the next section "GUI Usage".
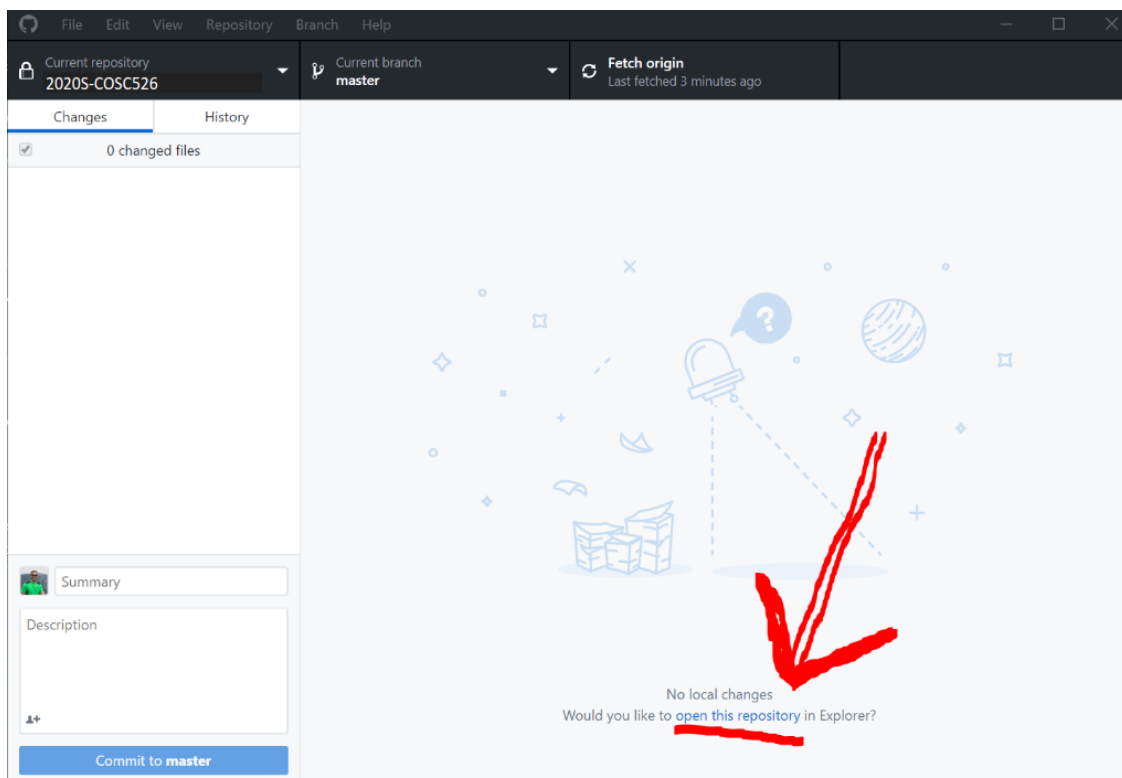
4.0.2 GUI Usage

- Launch the 'Anaconda Navigator' application, which should now be installed on your system:

- From the 'Anaconda Navigator', launch the 'Jupyter Notebook':

- Find and open the notebook titled 'StartHere.ipynb' that you clone'd from GitHub

- If you are having trouble locating the location of your clone'd repo, look here in the GitHub Desktop app:

5 Test Your Installation

At this point, you should no longer be viewing a pdf version of this document. To verify this, try running the code below;–select the code cell and press 'Shift+Enter'. You should see it print (below the cell): "Hello World!" along with the date.

```python
from datetime import datetime
print("Hello World! Right now, it is {}.".format(datetime.today().
 ↪strftime("%c")))
```

If you are able to execute the above code, great! You can move onto the first assignment Assignment01 - Assignment01.ipynb.

If not, please ask for assistance in troubleshooting the problem. You will not be able to complete the assignment from GitHub. You must be running this notebook locally on your machine.