STAT 19000 Project 1

Topics: Setting up your computer and going on a scavenger hunt through Scholar

Read the Motivation / Context / Scope discussion, called "Why are we here?", in the Blackboard site.

ThinLinc

Before doing anything else, we will get you setup for the work we will be doing all semester. We will connect to the Scholar cluster in Research Computing, which is where the data and our programs live. You will need to download and install the ThinLinc client to your desktop. Alternatively, you can use this website to connect each time: https://desktop.scholar.rcac.purdue.edu

A key upside of the ThinLinc Client is that it is stable and allows copy-and-paste to work with your computer.

Download the ThinLinc Client files:

- For Windows users, the installer is here:

https://www.cendio.com/downloads/clients/tl-4.10.0-client-windows.exe

- For Mac users, the installer is here:

https://www.cendio.com/downloads/clients/tl-4.10.0__6068-client-macos.iso

- For UNIX users, choose the appropriate installer:

https://www.cendio.com/thinlinc/download

Setting up the ThinLinc Client:

Here are the settings to use inside the ThinLinc Client:

- Server: `desktop.scholar.rcac.purdue.edu`
- Username: your Purdue username without the @purdue.edu – for instance, Dr Ward's username is: `mdw`
- Password: your regular Purdue Career account password. (This is not BoilerKey.)
- Options: Click on "Options" and when the window opens, choose the "Screen" tab, and do the following:
    - choose the "Resize remote session to the local window"
    - do NOT choose Full screen mode, and do NOT enable full screen mode over all monitors
- If you do accidentally get stuck in full screen mode, the F8 key will help you to escape.
- NOTE: The very first time that you log onto Scholar, you will have an option of "use default config" or "one empty panel". PLEASE choose the "use default config".

Question 1: Scavenger Hunt on Scholar

Try the following *inside* the Scholar environment.

1a. Find Firefox (either in the Applications menu or in the Dock at the bottom of the screen)

1b. Find the Text Editor (in the Applications menu)

1c. Find the LibreOffice Writer (in the Applications menu). It looks a lot like Microsoft Word, and it is mostly compatible with Word. We will not need this too much during the semester.

1d. Find the Document Viewer (in the Applications menu). It looks/acts a lot like Adobe Reader for pdf files.

1e. Find the Terminal Emulator (which we just call "the Terminal"; in the Applications menu or in the Dock).

Question 2: Getting Familiar with Storage on Scholar

2a. Find the File Manager (in the Applications menu or in the Dock; it looks like a filing cabinet).

2b. Open the File Manager and click on the File System. You have basically three places to store files:

In your **home directory**, you can store files and the files are backed up. You have a friendly little home-shaped icon in the left menu bar, to access your home directory. Dr Ward's is called: `/home/mdw`

Yours will have a similar name. A user named Jessica Gilbert might have home directory: `/home/gilber58`

In your home directory, you have a limited amount of space (only a few gigabytes) but this is where you should store your projects.

In your **scratch directory**, you can put a nearly unlimited amount of files. They are not backed up, and they could be deleted after they have not been touched for many weeks, but it is a useful sandbox, especially for huge files.

For instance, Dr Ward's scratch directory is: `/scratch/scholar/mdw`

and Jess's scratch directory is: `/scratch/scholar/gilber58`

In the **class directory**, we will have the data to be used in projects. The data cannot be modified there, but it can be read, and from this directory, you can copy the data to your scratch directory, or just work with the data in its current location, as long as you do not change it.

Question 3: Investigating some Data on Scholar

You ONLY need to submit your answers to Questions 3a, 3b, 3c, 3d, 3e for this project. You submit them at this URL: https://classroom.github.com/a/e2NZc7Ho using the instructions found in the GitHub Classroom instructions folder on Blackboard.

3a. Some airline data is located at: `/class/datamine/data/flights` Roughly how much data in stored in this directory? We do not need an exact count (yet). Just give a sense: several kilobytes? megabytes? gigabytes? petabytes?

3b. Find the map files for the taxi data (these are jpg files). How many files are there, in that maps directory?

3c. Roughly how much data is stored (altogether) in the directory about **yellow** taxi cabs?

3d. For which years do we have election data?

3e. For which cities in California do we have AirBnB data?