C-Innova: Ciencia de Data Taller

Instrucciones de Installacion y Mas informacion sobre data

**\*\* Seguir en orden por favor \*\***

1. Install **Python3**
   1. **Link:** [**https://www.python.org/downloads/**](https://www.python.org/downloads/)
      1. **Download latest version!** (
   2. What is python?
      1. Allows us to program -> Programming: process of encoding a rule or process into a notation (in this case: a programming language) to be executed by a computer
         1. Python: The language/notation we have selected that allows us to communicate and express what we want the computer to do
   3. **Command line terminal** 
      1. What is a command line terminal?
         1. A window/user interface to type commands and communicate with your computer’s operating system (to carry out activities), such as running python files
      2. If you have an Apple MAC laptop, you don’t need to install Git Bash, just use your computer’s “Terminal”
      3. **If you don’t have an Apple MAC, Install Git Bash** 
         1. What is Git Bash?
         2. It’s a command line terminal for **unix commands** (so you can use the same commands as someone who would be typing in an Apple MAC terminal)
   4. **Pip3**
      1. What is pip?
         1. Pip is a package management system used to install and manage software packages written in Python
      2. Libraries you can now install because you have Pip3!
         1. **Numpy**
            1. Instructions:

Open your command line terminal (MAC: “Terminal”, other: “Git Bash”)

Type in: pip3 install numpy

* + - * 1. What is Numpy?

Library for python programming language to support large multidimensional arrays and matrices (data structures) and collection for high level mathematical functions to operate + manipulate these arrays and matrices

* + - * 1. Our purpose: It’s a library depended on by Pandas, Matplotlib etc; numerical purposes and handling numbers/data structures
      1. **Pandas** (depends on Numpy, so make sure you install Numpy first)
         1. Instructions:

Open your terminal

Type in: pip3 install pandas

* + - * 1. What is Pandas?

It’s not a bear.

Software library written for the python programming language for data manipulation/editing and analysis

The python functions written to make up the library offer data structures and operations to manipulate numerical tables and time series (series of data points indexed in time order/how things change over time)

The name Pandas was inspired by “panel data” – econometrics term for multidimensional structured data sets

* + - * 1. Our purpose: We’ll be using this to plot figures, collecting/pulling in data, cleaning up data, and analyzing data for visualization!
      1. **Matplotlib**
         1. Instructions:

Open your terminal

Type in: pip3 install matplotlib

* + - * 1. What is matplotlib?

Plotting library for the python programming language and it’s a numerical extension of Numpy

* + - * 1. Our purpose: We’ll be using this to plot figures and visualize data!
    1. **Jupyter Notebook**
       1. Instructions:
          1. Link for reading: <https://jupyter.readthedocs.io/en/latest/install.html>

Requires Python! You should have already installed this.

* + - * 1. Open terminal
        2. Type in: pip3 install --upgrade pip
        3. Type in: pip3 install jupyter
        4. Running the jupyter notebook

Link for more reading: https://jupyter.readthedocs.io/en/latest/running.html#running - jupyter notebook

What your terminal should look like:

$ jupyter notebook [I 08:58:24.417 NotebookApp]

Serving notebooks from local directory: /Users/catherine [I 08:58:24.417 NotebookApp] 0 active kernels [I 08:58:24.417 NotebookApp]

The Jupyter Notebook is running at: http://localhost:8888/ [I 08:58:24.417 NotebookApp]

Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).

It will then open your default web browser to this URL.

When the notebook opens in your browser, you will see the Notebook Dashboard, which will show a list of the notebooks, files, and subdirectories in the directory where the notebook server was started. Most of the time, you will wish to start a notebook server in the highest level directory containing notebooks. Often this will be your home directory.

* + - 1. What is Jupyter Notebook?
         1. Open source web application that allows you to create and share documents that contain live code, equations, visualizations, and explanatory text
      2. Our purpose:
         1. To provide instructions to code for you to revisit
         2. For you to take your own notes from activities we do
  1. Other data visualization/analytics languages:
     1. R (can be integrated with Tableau, a data visualization/business intelligence software we’ll talk more about)
     2. Matlab

1. **Install a text editor to write code** 
   1. Recommended text editor: **Sublime**
      1. Other options you can explore: Atom, EMacs, Vim/Vi
      2. **Sublime text editor directions:** 
         1. Link: <http://www.sublimetext.com/3>
            1. OS X (10.7 or later is required)
            2. Windows - also available as a portable version
            3. Windows 64 bit - also available as a portable version
            4. Linux repos - also available as a 64 bit or 32 bit tarball
2. **Tableau**
   1. Data visualization software – drag and drop features; user friendly for visualizing data; use server, excel, csv and other sources of data
   2. Other similar software: Spotfire, SAS, etc
   3. We’ll load in CSV for testing purposes
      1. Can integrate with R and Python to affect data visualizations for advanced analytics – good connection/transition