Anaconda

Anaconda Python

- Particular Distribution of Python and R Programming Languages
  - Used primarily for scientific computing
  - Includes many data-science packages out the box
  - Simplifies package management
  - Used ubiquitously in industry
- Python or R?
  - One is not better than the other
  - Both great languages
  - Python
    - More libraries
    - Used more in industry
    - And academia (for data science related research)

Installing Anaconda Python

- MAC OS:
  - https://repo.anaconda.com/archive/Anaconda3-2021.11-MacOSX-x86_64.pkg
- Windows
  - https://repo.anaconda.com/archive/Anaconda3-2021.11-Windows-x86_64.exe
- Linux
  - https://repo.anaconda.com/archive/Anaconda3-2021.11-Linux-x86_64.sh
- If you're using another architecture that's not x86 64-bit:
  - https://www.anaconda.com/products/individual

Anaconda Python

```
(base) jarvis@Kelechis-MBP CSC-DataScienceExample % conda -h
usage: conda [-h] [-V] command ...

conda is a tool for managing and deploying applications, environments and packages.

Options:

positional arguments:
  command
    clean        Remove unused packages and caches.
    compare      Compare packages between conda environments.
    config       Modify configuration values in .condarc. This is modeled after the git config command.
                 Writes to the user .condarc file (/Users/jarvis/.condarc) by default.
    create       Create a new conda environment from a list of specified packages.
    help         Displays a list of available conda commands and their help strings.
    info         Display information about current conda install.
    init         Initialize conda for shell interaction. [Experimental]
    install      Installs a list of packages into a specified conda environment.
    list         List linked packages in a conda environment.
    package      Low-level conda package utility. (EXPERIMENTAL)
    remove       Remove a list of packages from a specified conda environment.
    uninstall    Alias for conda remove.
    run          Run an executable in a conda environment. [Experimental]
    search       Search for packages and display associated information. The input is a MatchSpec, a
                 query language for conda packages. See examples below.
    update       Updates conda packages to the latest compatible version.
    upgrade      Alias for conda update.
```

Anaconda Python

- Virtual Environments
  - https://docs.conda.io/projects/conda/en/latest/user-guide/tasks/manage-environments.html
- Resolves Version Conflicts Well

```
(base) jarvis@Kelechis-MBP CSC-405-605_Fall_2021 % conda install geopandas

Collecting package metadata (current_repodata.json): done
Solving environment: failed with initial frozen solve. Retrying with flexible solve.
Solving environment: failed with repodata from current_repodata.json, will retry with next repodata source.
Collecting package metadata (repodata.json): done
Solving environment: failed with initial frozen solve. Retrying with flexible solve.
Solving environment: -
Found conflicts! Looking for incompatible packages.
This can take several minutes.  Press CTRL-C to abort.
Examining conflict for scipy anaconda dask bottleneck networkx numexpr bkcharts h5py astropy tifffile numpy mkl-service matplotlib-base mkl_random pandas statsmodels \
```

Anaconda Python

- To Create virtual environment
  - conda env create …
- To create virtual environment from a file
  - conda env create –f {file_name}.yml
  - Course Environment on GitHub
    - “…/CSC-405-605_Spring_2022/environment_CSC605.yml”

```
jarvis@Kelechis-MBP CSC-405-605_Sring_2022 % conda env create -f environment_CSC605.yml
```

Anaconda Python

- To use the environment
  - conda activate {Environment}
- To stop using the environment
  - conda deactivate {Environment}
- Environment

```
(base) jarvis@Kelechis-MBP CSC-405-605_Sring_2022 % conda activate CSC605
(CSC605) jarvis@Kelechis-MBP CSC-405-605_Sring_2022 % conda deactivate
(base) jarvis@Kelechis-MBP CSC-405-605_Sring_2022 %
```

PIP

- What is PIP?
  - Package Manager to install Python Libraries
  - Traditionally used when not using Anaconda Python
- Pip documentation
  - https://pip.pypa.io/en/stable/getting-started/
  - Or pip –h

```
(base) jarvis@Kelechis-MacBook-Pro ~ % pip -h

Usage:
  pip <command> [options]

Commands:
  install                     Install packages.
  download                    Download packages.
  uninstall                   Uninstall packages.
  freeze                      Output installed packages in requirements format.
  list                        List installed packages.
  show                        Show information about installed packages.
  check                       Verify installed packages have compatible dependencies.
  config                      Manage local and global configuration.
  search                      Search PyPI for packages.
  cache                       Inspect and manage pip's wheel cache.
  wheel                       Build wheels from your requirements.
  hash                        Compute hashes of package archives.
  completion                  A helper command used for command completion.
  debug                       Show information useful for debugging.
  help                        Show help for commands.
```

Python

- Wide array of libraries for majority of data science applications
  - Numerical (NumPy, Pandas)
  - Scientific (SciPy)
  - Machine Learning (Sci-Kit, Tensorflow, PyTorch, OpenAI, NLTK)
  - Visualization (Matplotlib, Plotly, GeoPandas)
  - Big Data (PySpark)
- Easy to learn
- Excellent Documentation
- Thriving user base in comparison to R
- Preferred language of choice for most Data Science positions

IPython

- IPython provides a rich architecture for interactive computing
- Power Interactive shells (terminal and Qt-based)
- A browser-based notebook (Jupyter Notebook) *with support for code, rich testing, mathematical expressions, inline plots, and other rich media
- Support for interactive data visualization
- Flexible, embeddable interpreter to load into your own projects
- Easy to use, high performance tools for parallel computing

Ipython (Jupyter) Notebook

- Runs code in a web browser
- Stored in Json format
- Allows for code and text (Markdown)
- Has a debugger
- Has Checkpoints
- Sharing / co-editing is a lot easier
- Has access to all python libraries (locally imported and system installed)
- All assignments will use jupyter notebooks

JupyterLab

- An extension of Jupyter Notebooks
  - Web-based lightweight IDE for notebooks, code, and data
  - Flexible interface
  - 3rd party extensions

Jupyter

- To Run JupyterLab locally
  - jupyter lab

```
(base) jarvis@Kelechis-MBP CSC-405-605_Sring_2022 % conda activate CSC605
(CSC605) jarvis@Kelechis-MBP CSC-405-605_Sring_2022 % jupyter lab
[I 2022-01-15 22:38:48.716 ServerApp] jupyterlab | extension was successfully linked.
[I 2022-01-15 22:38:49.219 ServerApp] nbclassic | extension was successfully linked.
[I 2022-01-15 22:38:49.266 ServerApp] The port 8888 is already in use, trying another port.
[I 2022-01-15 22:38:49.266 ServerApp] The port 8889 is already in use, trying another port.
[I 2022-01-15 22:38:49.275 LabApp] JupyterLab extension loaded from /Users/jarvis/opt/anaconda3/envs/CSC605/lib/python3.8/site-packages/jup
yterlab
[I 2022-01-15 22:38:49.275 LabApp] JupyterLab application directory is /Users/jarvis/opt/anaconda3/envs/CSC605/share/jupyter/lab
[I 2022-01-15 22:38:49.279 ServerApp] jupyterlab | extension was successfully loaded.
[I 2022-01-15 22:38:49.285 ServerApp] nbclassic | extension was successfully loaded.
[I 2022-01-15 22:38:49.286 ServerApp] Serving notebooks from local directory: /Users/jarvis/GoogleDrive/Work/UNCG/CSC-405-605_Sring_2022
[I 2022-01-15 22:38:49.286 ServerApp] Jupyter Server 1.4.1 is running at:
[I 2022-01-15 22:38:49.286 ServerApp] http://localhost:8890/lab?token=14bd80f90ebb84dfa57207ea8b306bbe99d956e29fb0d7b0
[I 2022-01-15 22:38:49.286 ServerApp]  or http://127.0.0.1:8890/lab?token=14bd80f90ebb84dfa57207ea8b306bbe99d956e29fb0d7b0
[I 2022-01-15 22:38:49.286 ServerApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 2022-01-15 22:38:49.290 ServerApp]

    To access the server, open this file in a browser:
        file:///Users/jarvis/Library/Jupyter/runtime/jpserver-55132-open.html
    Or copy and paste one of these URLs:
        http://localhost:8890/lab?token=14bd80f90ebb84dfa57207ea8b306bbe99d956e29fb0d7b0
     or http://127.0.0.1:8890/lab?token=14bd80f90ebb84dfa57207ea8b306bbe99d956e29fb0d7b0
```

Cloud Resources

- Microsoft Azure – UNCG
  - https://kangaroo.uncg.edu
  - Login with Spartan ID
  - May show security exception, just confirm it and proceed
  - May take a few minutes to startup
- Google CoLab
  - https://colab.research.google.com/notebooks/welcome.ipynb
  - Links to google drive for data
- Both have limitations
  - Using your own laptop/desktop gives you more control
- Use whatever makes you comfortable
  - **Support only for local installations using class conda environment**

IPython Tutorial