

# Introduction to Data Science

## Data Science Essentials

---



# Goals for today

- **conda environments**
- **Create a choropleth using *geopandas***



# Review last session coding tasks

**week3\_review** notebook



# conda

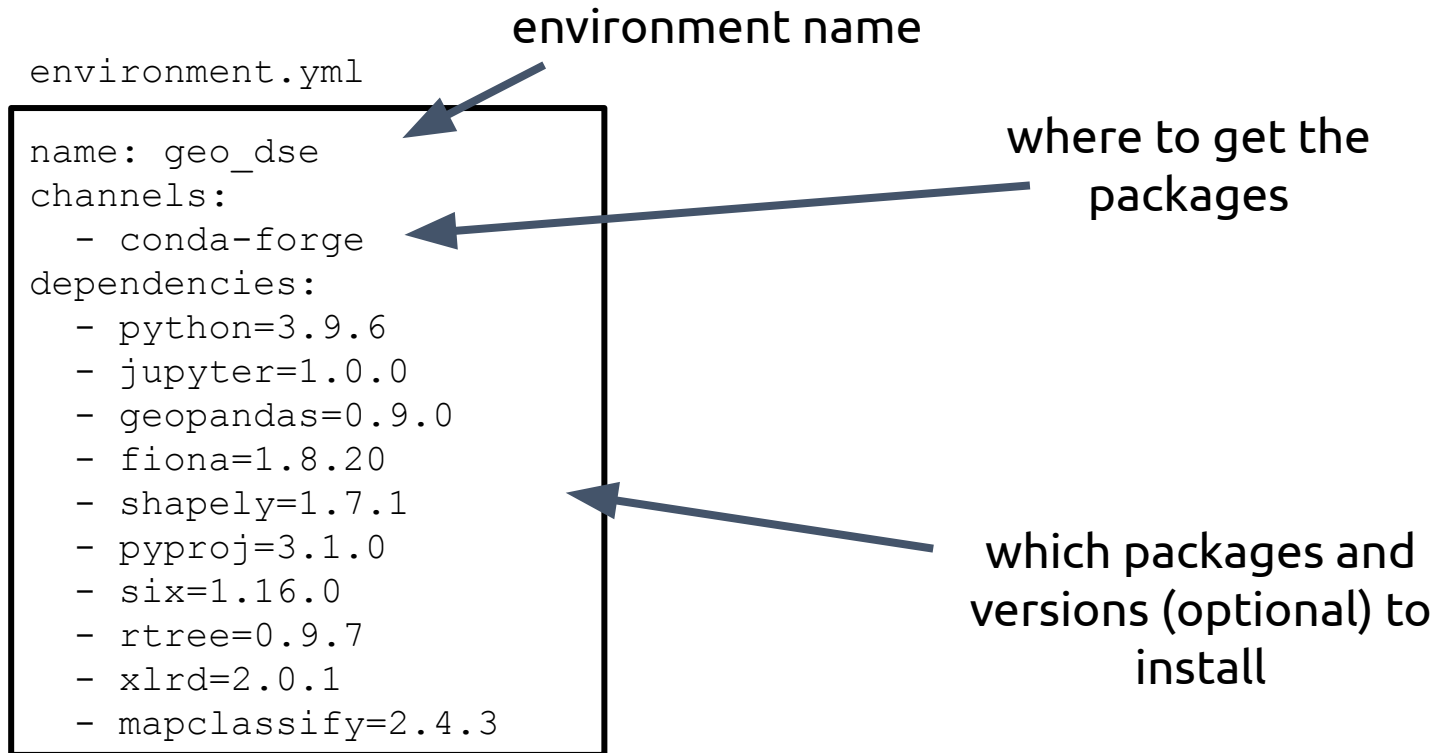
conda is not just a package manager, but is also an *environment* manager, meaning that you can create separate environments containing files, packages, their dependencies, and their own versions of the Python interpreter.

This serves two purposes:

1. Isolates your projects
2. Makes it easier to share your work and allows for reproducibility.

# conda Environments

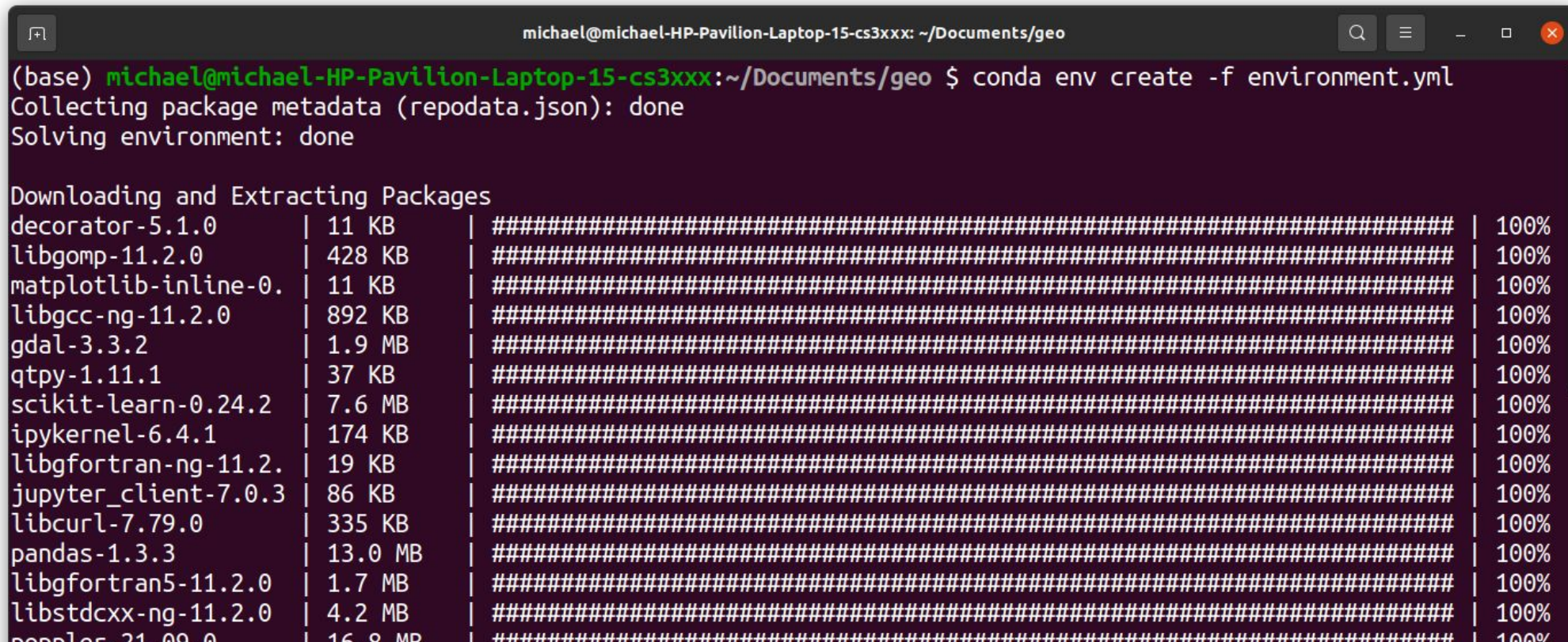
We're going to create an environment from a YAML file which lists which packages and versions should be installed.



# conda Environments

To create an environment from a YAML file, run this from the folder containing the environment file:

```
$ conda env create -f environment.yml
```

A terminal window with a dark purple background. The title bar shows the user 'michael' on a 'michael-HP-Pavilion-Laptop-15-cs3xxx' machine, in the directory '~/Documents/geo'. The terminal output shows the command being executed, followed by status messages for metadata collection and environment solving. A progress table for downloading and extracting packages is displayed, listing various scientific and data science packages with their sizes and progress percentages. The window includes standard Linux window controls (minimize, maximize, close) and a search icon.

```
(base) michael@michael-HP-Pavilion-Laptop-15-cs3xxx:~/Documents/geo $ conda env create -f environment.yml
Collecting package metadata (repodata.json): done
Solving environment: done

Downloading and Extracting Packages
decorator-5.1.0           | 11 KB | ##### | 100%
libgomp-11.2.0            | 428 KB | ##### | 100%
matplotlib-inline-0.     | 11 KB | ##### | 100%
libgcc-ng-11.2.0         | 892 KB | ##### | 100%
gdal-3.3.2               | 1.9 MB | ##### | 100%
qtpy-1.11.1              | 37 KB | ##### | 100%
scikit-learn-0.24.2      | 7.6 MB | ##### | 100%
ipykernel-6.4.1          | 174 KB | ##### | 100%
libgfortran-ng-11.2.     | 19 KB | ##### | 100%
jupyter_client-7.0.3     | 86 KB | ##### | 100%
libcurl-7.79.0           | 335 KB | ##### | 100%
pandas-1.3.3             | 13.0 MB | ##### | 100%
libgfortran5-11.2.0      | 1.7 MB | ##### | 100%
libstdcxx-ng-11.2.0      | 4.2 MB | ##### | 100%
poppler-21.09.0          | 16.8 MB | ##### | 100%
```

# conda Environments

You can see a list of all of your conda environments along with the one currently active by typing

```
$ conda env list
```

```
(base) michael@michael-HP-Pavilion-Laptop-15-cs3xxx:~/Documents/geo $ conda env list
# conda environments:
#
base                * /home/michael/anaconda3
abtesting           /home/michael/anaconda3/envs/abtesting
bayes               /home/michael/anaconda3/envs/bayes
ft_vectors          /home/michael/anaconda3/envs/ft_vectors
geo_dse             /home/michael/anaconda3/envs/geo_dse
/home/michael/.local/share/r-miniconda
/home/michael/.local/share/r-miniconda/envs/pandas
/home/michael/.local/share/r-miniconda/envs/r-reticulate
```

This will display a \* next to the active environment.



# conda Environments

To switch environments, type

```
$ conda activate <environment name>
```

```
(base) michael@michael-HP-Pavilion-Laptop-15-cs3xxx:~/Documents/geo $ conda activate geo_dse
(geo_dse) michael@michael-HP-Pavilion-Laptop-15-cs3xxx:~/Documents/geo $ conda env list
# conda environments:
#
base                    /home/michael/.local/share/r-miniconda
abtesting               /home/michael/.local/share/r-miniconda/envs/pandas
bayes                   /home/michael/.local/share/r-miniconda/envs/r-reticulate
ft_vectors              /home/michael/anaconda3
geo_dse                 * /home/michael/anaconda3/envs/geo_dse
```

If you need to return to the base environment, type

```
$ conda deactivate
```



# conda Environments

Then (as long as your current environment includes jupyter), you can launch jupyter in the current environment by typing

```
$ jupyter notebook
```

```
(geospatial) michael@michael-HP-Pavilion-Laptop-15-cs3xxx ~ $ jupyter notebook
[I 10:24:09.366 NotebookApp] The port 8888 is already in use, trying another port.
[I 10:24:09.371 NotebookApp] Serving notebooks from local directory: /home/michael
[I 10:24:09.371 NotebookApp] The Jupyter Notebook is running at:
[I 10:24:09.371 NotebookApp] http://localhost:8889/?token=dc72fd8e78a41802290ae2c77e5c8b143c37adab3c7ef971
[I 10:24:09.371 NotebookApp] or http://127.0.0.1:8889/?token=dc72fd8e78a41802290ae2c77e5c8b143c37adab3c7ef971
[I 10:24:09.371 NotebookApp] Use Control-C to stop this server and shut down all
```

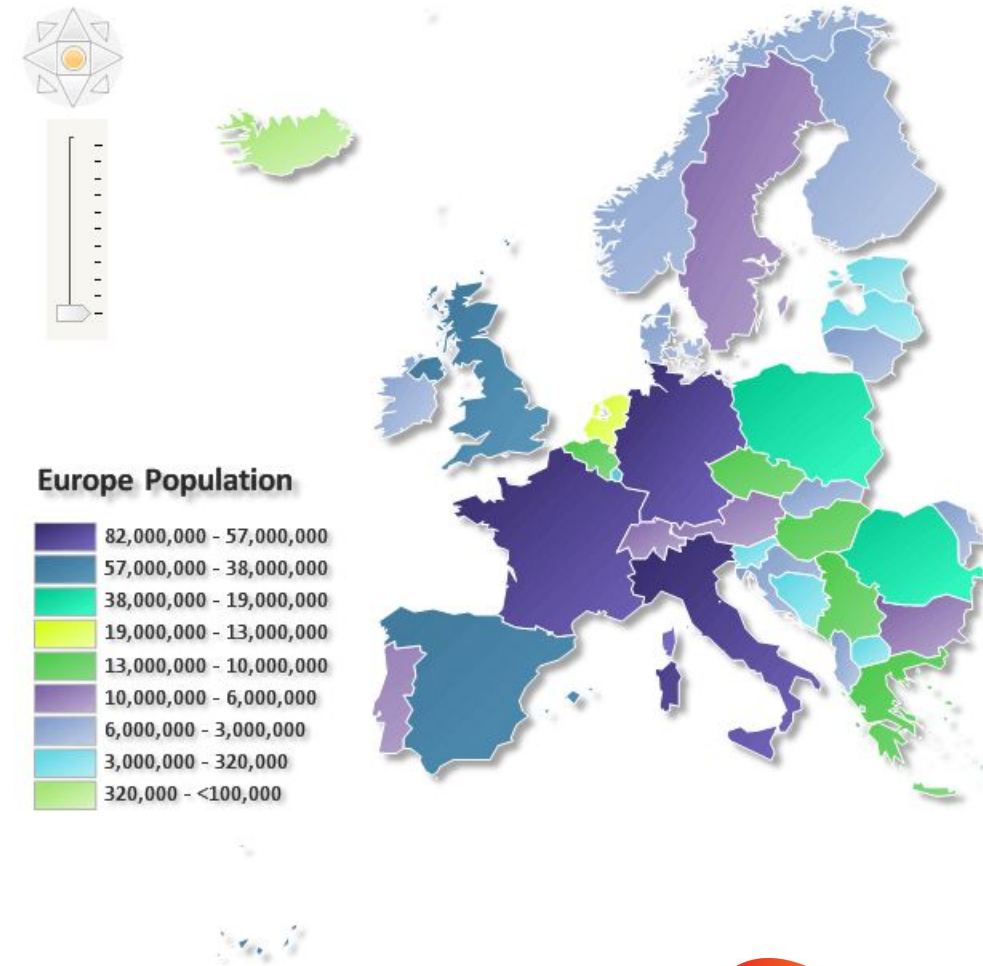
# Choropleths

A choropleth is a map where areas are colored or shaded according to the value of some aggregate statistic for that area (eg. average income, population density, unemployment rate, etc.)

We will create choropleths in Python by using the *geopandas* library, which you will most likely need to install.

To install geopandas, open the Terminal (Mac) or Anaconda Prompt (Windows) and type

\$ `conda install geopandas -c conda-forge`



# Building a choropleth

**Choropleth\_Tutorial** notebook



# Questions?

