This is a guide to help you set up Python, Jupyter notebook, and other packages you will need on your system to engage in class lab activities. *You are NOT required to use Python for your assignments and projects.* However, because Python seems to be the most commonly used language for text mining/NLP these days, we will focus on it for practice.

Download and install Anaconda environment for Python 3.8

https://www.anaconda.com/download/

Anaconda environment will help you manage and deploy packages/libraries more efficiently.

To get started with conda, see Conda Getting Started guide

Create new anaconda environment for this class called *textmining*

```
conda create --name textmining python=3.8.8
```

Activate environment

```
source activate textmining
```

Check version (should be 3.8.8)

```
python --version
```

Install packages

Be sure to install these specific versions so that everyone in the class is using the same environment and debugging is easier.

```
conda install nb_conda=2.2.1
conda install nltk=3.6.2
conda install -c conda-forge spacy=3.0
conda install scikit-learn=0.24.2
conda install pandas=1.3.1
conda install matplotlib=3.4.2
```

Install spaCy English model

```
python -m spacy download en_core_web_sm
```

Install wordcloud for visual representation of the corpus

```
conda install -c conda-forge wordcloud
```

Use Jupyter notebooks

Most class exercises will be provided to you as Jupyter notebooks. To open a Jupyter notebook in this setup, open up the terminal and navigate to the folder containing the notebook; then activate the textmining environment to access these libraries and start up the notebook:

```
source activate textmining
jupyter notebook
```

If you're new to Jupyter notebooks, you can check out a tutorial here:

- [Jupyter notebook tutorial](#)

This guide has been adapted from https://github.com/dbamman/anlp21/blob/main/0.setup.