Lab 2

In this lab please write a simple character, word, and line count function and test it on a couple of files

Some useful background material:

The [Purdue Data Mine Examples Book (https://the-examples-book.com/book/introduction)](https://the-examples-book.com/book/introduction) contains many useful chapters on data science. While they have not been directly designed for this class, they may be useful. You will not need to use scholar to perform the exercises of this class so don't worry about that part. Here is a direct link to the [Python chapter. (https://the-examples-book.com/book/python/introduction)](https://the-examples-book.com/book/python/introduction)

Additional useful links for Python include:

[Python 3.9.4 documentation (https://docs.python.org/3/)](https://docs.python.org/3/)

The [Python Package Index (https://pypi.org/)](https://pypi.org/) (This contains many of the useful Python "add-on" packages such as the math package)

The [Numpy Package (https://numpy.org/)](https://numpy.org/) (This contains specialized array (vector and matrix) routines. Numpy stands for "Numerical Python")

United States Department of Agriculture: [Quick Stats (https://quickstats.nass.usda.gov/)](https://quickstats.nass.usda.gov/) (The USDA's National Ag Statistics Service -- go here and familiarize yourself with the available data)

In [ ]:
```python
# Import packages ...
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
```

A Python Function to count characters, words, and lines in a file

Takes a file located in the current working directory. Reads it. And counts these items printing the numbers to standard output ...

```python
In [ ]:    # Skeleton of the function to write ...

           def PoorMansWordCount(file_name):

               fin = # open the given file name for reading only

               # initialize three variables to zero for holding the numbers of charac

               Nchar = 0
               Nword = 0
               Nlines = 0

               for line in fin:
                   Nlines += 1 # Everytime we read a line, we count it ...
                   Nchar += # increment by the length of the line you just read ...
                   Nword += # increment of the length of the list you get by splittin

               print(f'Number of lines in {file_name} is {Nlines}.')
               print(f'Number of words in {file_name} is {Nword}.')
               print(f'Number of characters in {file_name} is {Nchar}.')

               return None
```

```python
In [2]:    #  Skeleton of the function to write ...

           def PoorMansWordCount(file_name):

               fin = open(file_name, 'r')

               # initialize three variables to zero for holding the numbers of charac

               Nchar = 0
               Nword = 0
               Nlines = 0

               for line in fin:
                   Nlines += 1 # Everytime we read a line, we count it ...
                   Nchar += len(line.replace('\n',''))
                   Nword += len(line.replace('\n','').split())

               print(f'Number of lines in {file_name} is {Nlines}.')
               print(f'Number of words in {file_name} is {Nword}.')
               print(f'Number of characters in {file_name} is {Nchar}.')

               return None
```

```python
In [3]:    # Test your solution on the PLimerick file ...

           PoorMansWordCount('Data/PLimerick')
```

```
Number of lines in Data/PLimerick is 8.
Number of words in Data/PLimerick is 42.
Number of characters in Data/PLimerick is 184.
```

```
In [4]:  ▶  # Test your solution on President Obama's First Inagural Speech

            PoorMansWordCount('Data/ObamaFirstInagural.tex')
```

Number of lines in Data/ObamaFirstInagural.tex is 183.
Number of words in Data/ObamaFirstInagural.tex is 2423.
Number of characters in Data/ObamaFirstInagural.tex is 13366.

```
In [ ]:  ▶
```

```
In [ ]:  ▶
```