exercise.M2.exercise.1 => Python Library

0.Problem summary

For this problem

Objectives:

1. **Import and manipulate a .csv file**
2. **Assess your Python Programming Skills**

=> Other assignments are more challenging. Use this to assess your skills.
=> Attempt to solve the problems without searching for online assistance.
=> Prepare questions for class discussion to help source additional tools.

Codebook and data files

a. = article (news, journal)	c. = cheatsheet code. = .py or .ipynb	g = graphic
howTo. = explanandum	py.M. exercise or assignment python file	r = reading

File Name	Purpose\Description
https://github.com/cosc-526/cosc.526.home.page/blob/main/code_notebook_cosc_526.ipynb **save your own copy!**	Course Codebook in Jupyter Notebook name = code.notebook.cosc.526.ipynb
https://github.com/cosc-526/cosc.526.home.page/tree/data_shakespeare_corpus_text_name	Data: all shakespeare text files
	i) Kaggle data homepage ii) grab an api key from this page if using that method to import data

note.1: the codebook is formatted differently and below highlights expected outcomes.

note.2: the instructions below are an overview and more details are in the notebook.

note.3: perform your work in the notebook, and export and submit as a .pdf file.

Problem 0 - Import, inspect, and view descriptive statistics

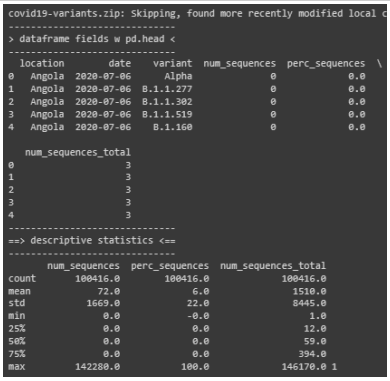
Import data and view descriptive statistics with the pandas library.

- grab data from **Github** URL, .csv. or kaggle api

Problem.1 - Description =>

The United S

Task.0 - Expected outcome:



```
covid19-variants.zip: Skipping, found more recently modified local c
------------------------------
> dataframe fields w pd.head <
------------------------------
  location        date    variant  num_sequences  perc_sequences  \
0   Angola  2020-07-06      Alpha              0             0.0
1   Angola  2020-07-06  B.1.1.277              0             0.0
2   Angola  2020-07-06  B.1.1.302              0             0.0
3   Angola  2020-07-06  B.1.1.519              0             0.0
4   Angola  2020-07-06    B.1.160              0             0.0

   num_sequences_total
0                    3
1                    3
2                    3
3                    3
4                    3
------------------------------
==> descriptive statistics <==
------------------------------
       num_sequences  perc_sequences  num_sequences_total
count       100416.0        100416.0             100416.0
mean            72.0             6.0               1510.0
std           1669.0            22.0               8445.0
min              0.0            -0.0                  1.0
25%              0.0             0.0                 12.0
50%              0.0             0.0                 59.0
75%              0.0             0.0                394.0
max         142280.0           100.0             146170.0 1
```

Problem.2 - Description =>

Which v

Task.0 - Expected outcome:

Problem.3 - Description =>

Which cou

Task.0 - Expected outcome:

Problem.4 - Description =>

Task.0 - Expected outcome:

Problem.5 - Description =>

Determine e

Task.0 - Expected outcome:

Problem.6 - Description =>

Determine

Task.0 - Expected outcome:

Additional resources

- https://github.com/cosc-526/cosc.526.home.page
- Jupyter Community Forum

Additional resources

- need help? Jupyter Community Forum

10. Additional resources

- Anaconda for windows
- Install scientific packages.
- Anaconda installation documentation.
- Jupyter Notebook documentation (including get started guides).
- Jupyter Discourse Forum.
    - Search here for tips, tricks, and solutions.
- Python Package Index (pypi)