```
.file/github  ==> exercise.M1.exercise.1
.institution  ==> University of Tennessee
     .course  ==> COSC.526 Intro. to Data Mining
```

exercise.M1.exercise.1

0. Working with .csv and .tsv files

In this problem we will explore reading in and parsing delimiter-separated values stored in files. We will start with comma-separated values and then move on to tab-separated values.

0.1 Files

File Name	Purpose\Description
exercise.M.1.excercise.1.pdf	
code_notebook_cosc_526.ipynb	

1.Problem.1 Comma-Separated Values (CSV)

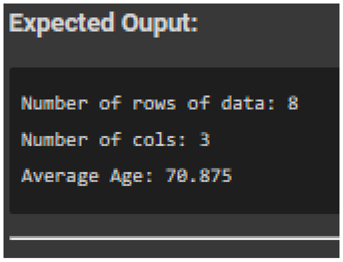
Problem 1A: Comma-Separated Values (CSV)

From Wikipedia: In computing, a comma-separated values (CSV) file stores tabular data(numbers and text) in plain text. Each line of the file is a data record. Each record consists of one or more fields, separated by commas. The use of the comma as a field separator is the source of the name for this file format.

If you considered the CSV file a matrix, each line would represent a row and each comma would represent a column. In the provided CSV file, the first row consists of a header that "names" each column.

1. Count (and print) the number of rows of data (header is excluded) in the csv file
2. Count (and print) the number of columns of data in the csv file
3. Calculate (and print) the average of the values that are in the "age" column
4. Assume each age in the file is an integer, but the average should be calculated as a float

Problem.1A - Expected outcome:



Expected Ouput:

```
Number of rows of data: 8
Number of cols: 3
Average Age: 70.875
```

References:

1. open
2. readlines
3. list comprehension
4. rstrip
5. split
6. splice
7. "more on lists"
8. len
9. int
10. format

2. Problem Tab-Separated Values (TSV)

A tab-separated values (TSV) file is a simple text format for storing data in a tabular structure, e.g., database table or spreadsheet data, and a way of exchanging information between databases.

- Each record in the table is one line of the text file.
- Each field value of a record is separated from the next by a tab character.
- The TSV format is thus a type of the more general delimiter-separated values format.
- In this problem, repeat the analyses performed in the previous problem, but for the provided tab-delimited file.

Note: The order of the columns has changed. If you hardcoded the position of the "age" column, think about how you can generalize the parse_delimited_file function to work for any delimited file with an "age" column.

Convert the unicode-formatted names into ascii-formated names.
Save the names out to a file named data-ascii.txt (one name per line).

3. Problem 3

8.1 Sub sction 1.1

7. Additional resources

- **course repository:** https://github.com/cosc-526/cosc.526.home.page
- **quality help:** Jupyter Community Forum