CS 124 Project 1

Due by 11:59pm ET on Wednesday, February 2nd

For this project, you will:

- choose a data set,
- write a C++ class to store its information,
- read the data into a vector of objects,
- perform a calculation on the vector,
- and write about your project in a report.

Requirements

Data Set

- Your data set must contain at least 1000 entries (rows).
- Each entry must have at least 5 attributes (columns).
- At least one attribute must be unique for each entry in the data set (i.e. no duplicates).
- Keep in mind, you will be using this data set for the rest of the projects this semester.

Class

- Your class must have at least 5 fields where at least one is a numerical type and at least one is a string type.
- You should have constructors, getters, setters, and other methods as appropriate.

Global Functions

- You must create a function that opens your data file and reads the data into a vector of objects of your class.
- You must create a function that will loop through your vector of data and perform some kind of calculation on it.

Design

Consider the following questions:

- What data do you want to use? Where will you get it from?
- How will you store it in a file? Is it in csv format?
- How will you read it in using C++? How will you know that it read in correctly?
- What should you name the class?
- What fields do you need? What are their types? What names make sense?
- Which fields and methods need comments to clarify their meaning?
- What function makes the most sense to calculate on your data?

Some examples of data sets:

- All the courses offered at UVM this semester (there is a csv download on the enrollment page).
- The National Oceanic and Atmospheric Administration (NOAA) collects climate and weather data and allows you to search and download the raw data.
- Find a list of the top 1000 songs of your favorite genre or decade.
- Find the 1000 top-rated movies from your favorite movie rating source.
- Get data about your favorite sports team/teams.

Example functions include:
- The sum, average, min, and/or max of a column.
- The shortest/longest word in a column.
- The number of times a specific value appears in a column.

Test

How can you demonstrate in your code that all your class methods function correctly?
How can you demonstrate in your code that your program successfully reads and stores 1000+ objects from your data file?
How can you demonstrate in your code that your calculation function works correctly?

Report and Submit

You must write a report about your project:
- What each of the 5+ attributes represent.
- Where you got the data from.
- Why you chose that data set.
- How the entries are ordered by default.
- How you know your functions work correctly.
- Why you chose that calculation for your data set.
- **Note: Any code that was not authored by yourself or the instructor must be cited in your report. This includes the use of concepts not taught in lecture.**

You must submit your source files, your data file(s), and your PDF report.

Grading

The project is out of 60 points.

5 pts	Data set satisfies requirements.
5 pts	Program compiles and runs.
5 pts	Code style. Readable, naming style is consistent, comments where appropriate.
10 pts	Class satisfies requirements.
5 pts	You have two global functions as described above.
10 pts	File input works correctly.
5 pts	Data is stored in a vector of 1000+ objects.
5 pts	You test your code to demonstrate that everything works correctly.
10 pts	Report satisfies requirements, is easily readable, and is professional.