CS 161B: Programming and Problem Solving II

Assignment 5: Fun Data Analysis

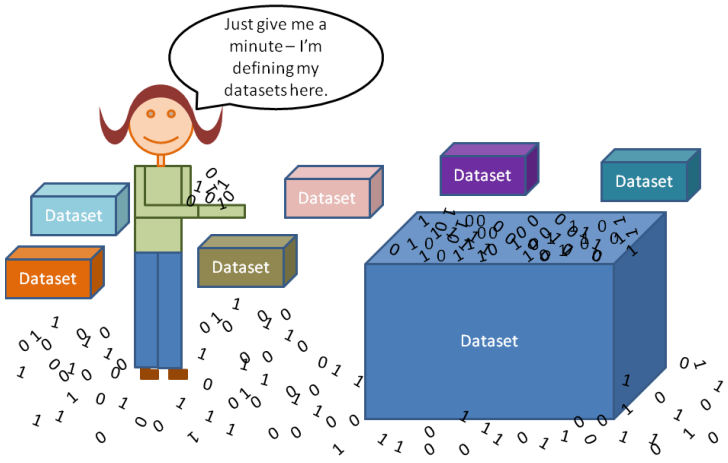
Academic Integrity

You may NOT, under any circumstances, begin a programming assignment by looking for completed code on StackOverflow or Chegg or any such website, which you can claim as your own. Please check out the Student Code of Conduct at PCC.

The only way to learn to code is to do it yourself. The assignments will be built from examples during the lectures, so ask for clarification during class if something seems confusing. If you start with code from another source and just change the variable names or other content to make it look original, you will receive a zero on the assignment.

I may ask you to explain your assignment verbally. If you cannot satisfactorily explain what your code does, and answer questions about why you wrote it in a particular way, then you should also expect a zero.

Data analysis is important in business **to understand problems facing an organization, and to explore data in meaningful ways**. Data in itself is merely facts and figures. Data analysis organizes, interprets, structures and presents the data into useful information that provides context for the data.



Purpose

The purpose of this assignment is to read from a text file, and do some data analysis. You must store your data into parallel arrays. Pick your own data from the list below under Task - you can have any dataset that interests you. Please remember to cite your Sources. Look at the examples in Zybooks, 12.5 File Input/Output and Parallel Arrays and 12.6 Assignment Sample and Video.

After completing this assignment you will be able to:

- Open and read input from a file
- Create parallel arrays that hold related pieces of information
- Create and process two-dimensional arrays

Task

- ❏ **Before you get started:**
  - ❏ **Check out Sample Assignment A05 - Algorithmic Design document**
  - ❏ **Check out Sample Assignment A05 - sampleA05.cpp and items.txt**
- ❏ There is a Unix Lab component to this assignment that must be completed as well. Instructions to this can be found in the Linux Lab 1 document. Watch this Lab 1 video and follow along.
- ❏ Here is a list of data sets to pick from. **You must pick from one of these data sets.** You don't have to use all columns and rows. You can pick a total of 4 columns - at least 1 char array and 2 numbers. Select only 10 to 30 rows of data. It is easier to test with a small set of data.
  - ❏ Salaries.csv
  - ❏ Automobile.csv
  - ❏ Disease.csv
  - ❏ Sleep_and_lifestyle.csv
  - ❏ Spotify_world.csv
  - ❏ cybersecurity.csv
- ❏ Make sure your data set meets the following requirements:
  - ❏ Do not use any of the data files used in zyBooks or my examples (like the cereal data, videos, songs, items and quantities).
  - ❏ You must have at least one column of strings (like movie names, item names etc) like my sample **items.txt**, and 2 columns of numbers. The numbers can be doubles or ints. **Do not use this sample and just change column names.**
  - ❏ Keep your dataset simple - otherwise you will end up with too many parallel arrays that will make it very complicated to write code for.
  - ❏ You must use char arrays to store the strings and depending on the datatypes of the numbers, you can create single or 2-dimensional arrays.

- ❏ Your data must be separated by semicolons or some delimiters other than space.
- ❏ **Check out the sample A05 to get an idea of what you need for your dataset.**
- ❏ Upload your data file with the rest of the files for submission.

❏ Open the Algorithmic Design Document, make a copy, and follow the steps to create your algorithm.

❏ You must express your algorithm as **pseudocode** or a **flowchart.**

❏ Your program must have the following functions:

- ❏ **Write a function to open the text file created from your dataset and check to make sure it opens. If it does not open, the program exits.**
- ❏ **Write a function to print the lists.**
- ❏ **Write a function to read from the arrays and do some data analysis. This could be comparing data such as finding the highest value or lowest value for a column.**
- ❏ **Write a function to read from the arrays and do some summation analysis. Find the sum or average of a column of data, or rows of data as shown below in the second example.**

❏ Then close the file and end the program.

Criteria for Success

❏ **The below sample run is an example - your output will depend on your data set. You may not use this example!**

```
Occupation Name          Employed  Automation  Percent %
_______________          ________  __________  _________
Administrative           23081     13849       60.0017%
Agriculture              1060      594         56.0377%
Arts Entertainment       2773      555         20.0144%
Business                 8067      1129        13.9953%
Computer                 4419      1635        36.9993%
Construction             6813      3407        50.0073%
Education                9427      1697        18.0015%
Engineering              2601      494         18.9927%
Facilities Care          5905      2893        48.9924%
Food Service             13206     10697       81.0011%
Health Practitioner      8752      2888        32.9982%
Health Support           4316      1726        39.9907%
Legal                    1283      488         38.0359%
Maintenance              5654      1187        20.994%
Management               9533      2193        23.0043%
Personal Care            6420      2183        34.0031%
Production               9357      7592        81.1371%
Protective               3506      1262        35.9954%
Sales                    15748     6772        43.0023%
```

```
Science                   1300      416         32    %
Social Service            2571      566         22.0148%
Transportation            10274     5651        55.0029%

Highest/Lowest Occupations Susceptible to Automation:
Production has the highest share (81%)
Business has the lowest share (14%)
```

- ❏ Check out Sample Assignment A05 - Algorithmic Design document
- ❏ Check out Sample Assignment A05 - sampleA05.cpp and items.txt
- ❏ Complete zyBooks section **12. CS161B: File Input Output** activities.
- ❏ Complete the Linux Lab 1 document. Watch this Lab 1 video and follow along.
- ❏ **Follow these Coding Construct Requirements:**
  - ❏ Upload the text file you used with your program.
  - ❏ Do not use any vectors or containers for this program. Use only the concepts we have learned so far.
  - ❏ You may not use the String class - you must use C-strings or char arrays.
  - ❏ Do not use any goto statements. You may not use any breaks or return statements inside any loops - you are allowed to use breaks inside a switch statement.
  - ❏ All input data must be validated. For example, you should not accept any characters for numerical data types.
  - ❏ Must have at least 4 functions - one to load the data from the file, one to do comparison analysis, one to do summation analysis, and one function to print the data.
  - ❏ You must use function prototypes and write your main() function at the top of your program, followed by the other function definitions.
  - ❏ Print a welcome and goodbye message.
- ❏ Complete all sections of your Algorithmic Design Document.
- ❏ Include **pseudocode** in part d of the design document.
- ❏ Please open and compare your work with the grading rubric before submitting.
- ❏ Remember to follow all style guidelines.
- ❏ Download your Algorithmic Design Document as a PDF (File -> Download -> PDF), rename it to `a05.pdf`, and upload it to the D2L assignment by **Wednesday**.
- ❏ Upload your data file `(yourfile.txt)`, `a05.cpp` C++ source file and `lab1.txt` to the D2L assignment by **Sunday**.
- ❏ Do your own work. Consult the syllabus for more information about academic integrity.

Additional Support

- ❏ Check out Sample Assignment A05 - Algorithmic Design document
- ❏ Check out Sample Assignment A05 - sampleA05.cpp and items.txt
- ❏ Post a question for the instructor in the Ask Questions! area of the Course Lobby.