CSE 216 – Programming Abstractions (Spring 2019)

Programming Assignment # 1

In this assignment, you will create a command line based user interface program that shows a particular statistics requested by a user. The program will be developed in Java language using NetBeans as IDE.

The data presented in this assignment is taken from Community Health Status Indicator (CHSI)[[1]](#footnote-1) dataset made available by Center for Disease Control and Prevention, U.S. Department of Health & Human Services. This dataset contains a CSV dataset file LEADINGCAUSESOFDEATH.csv presenting data about leading causes of death in various counties and states of U.S. The other CSV file Dataset\_info.csv describes the columns in a dataset file.

Your tasks are the following:

1. **Preprocessing:** This step involves reading LEADINGCAUSESOFDEATH.csv file and perform automatic preprocessing to remove erroneous data (such as negative numbers). Only include those columns for which description is present in Dataset\_info.csv file. Store processed data columnwise using data structure of your choice. Data for each column can be stored using HashTable<String, ArrayList<?>> where the key represents column name and values represent list of values in that column. You may use Opencsv package (<http://opencsv.sourceforge.net/>) for reading csv files.
2. **Create state-wise and column-wise statistics:** For each of the states and combinedly for all states for each corresponding columns of type PERCENT\_DATA, compute the following statistical parameters[[2]](#footnote-2):
   1. Mean of all counties,
   2. Median of all counties,
   3. Mode of all counties,
   4. Standard Deviation[[3]](#footnote-3), and
   5. Variance.

While computing above parameters, ignore the counties for which data is not available. You may use existing examples[[4]](#footnote-4) for calculating above parameters.

1. **User interaction:** Initially, you will present 5 options (and one exit option) corresponding to above statistical parameters to a user. In case the user doesn’t select Exit, then the program will prompt user to enter column code. In case a proper column code entered, the program will prompt user to select state abbreviation (2 letter code) or ALL (in case data of all states is required). In case the correct state code is entered, the selected statistics for that state is shown. Your program will interact with the user as long as user selects Exit option.

In case user enters the input outside allowed values, an error message should be given and user should be prompted again to enter valid input.

1. **Packaging:** The project is named **CHSIDataProcesing**. Use three different packages in your code – corresponding to **preprocessing**, **statistics** and **interaction**. The package naming follows the convention: edu.sunyk.cse216.<packagename>

The main program will be in **interaction** package with name **userinterface.java**. The main program will take as input path to two dataset files **LEADINGCAUSESOFDEATH.csv** and **Dataset\_info.csv**. After the program is ready, create an executable jar file **CHSIDataProcesing.jar** using the instructions provided here[[5]](#footnote-5). Include all necessary libraries and jar files while creating the executable jar. The program should run on the commandline when executed as follows:

**Cmd prompt>java -jar CHSIDataProcesing.jar <path to LEADINGCAUSESOFDEATH.csv> <path to Dataset\_info.csv>**

**An example of user interaction is shown below:**

Welcome! Select one of the following statistical parameters:

* 1. Mean
  2. Median
  3. Mode
  4. Standard Deviation
  5. Variance
  6. Exit

User selects -> 1

Select parameters of interest:

|  |
| --- |
| 1. A\_Wh\_Comp |
| 1. A\_Bl\_Comp |
| 1. A\_Ot\_Comp |
| 1. A\_Hi\_Comp |
| 1. A\_Wh\_BirthDef |
| 1. A\_Bl\_BirthDef |
| 1. A\_Ot\_BirthDef |
| 1. A\_Hi\_BirthDef |
| 1. B\_Wh\_Injury |
| 1. B\_Bl\_Injury |
| 1. B\_Ot\_Injury |
| 1. B\_Hi\_Injury |
| 1. B\_Wh\_Cancer |
| 1. B\_Bl\_Cancer |
| 1. B\_Bl\_Homicide |
| 1. B\_Hi\_Cancer |
| 1. B\_Hi\_Homicide |
| 1. B\_Ot\_Cancer |
| 1. B\_Ot\_Homicide |
| 1. B\_Wh\_Homicide |
| 1. C\_Bl\_Cancer |
| 1. C\_Bl\_Homicide |
| 1. C\_Bl\_Injury |
| 1. C\_Bl\_Suicide |
| 1. C\_Hi\_Cancer |
| 1. C\_Hi\_Homicide |
| 1. C\_Hi\_Injury |
| 1. C\_Hi\_Suicide |
| 1. C\_Ot\_Cancer |
| 1. C\_Ot\_homicide |
| 1. C\_Ot\_Injury |
| 1. C\_Ot\_Suicide |
| 1. C\_Wh\_Cancer |
| 1. C\_Wh\_Homicide |
| 1. C\_Wh\_Injury |
| 1. C\_Wh\_Suicide |
| 1. D\_Bl\_Cancer |
| 1. D\_Bl\_HeartDis |
| 1. D\_Bl\_HIV |
| 1. D\_Bl\_Homicide |
| 1. D\_Bl\_Injury |
| 1. D\_Bl\_Suicide |
| 1. D\_Hi\_Cancer |
| 1. D\_Hi\_HeartDis |
| 1. D\_Hi\_HIV |
| 1. D\_Hi\_Homicide |
| 1. D\_Hi\_Injury |
| 1. D\_Hi\_Suicide |
| 1. D\_Ot\_Cancer |
| 1. D\_Ot\_HeartDis |
| 1. D\_Ot\_HIV |
| 1. D\_Ot\_Homicide |
| 1. D\_Ot\_Injury |
| 1. D\_Ot\_Suicide |
| 1. D\_Wh\_Cancer |
| 1. D\_Wh\_HeartDis |
| 1. D\_Wh\_HIV |
| 1. D\_Wh\_Homicide |
| 1. D\_Wh\_Injury |
| 1. D\_Wh\_Suicide |
| 1. E\_Bl\_Cancer |
| 1. E\_Bl\_HeartDis |
| 1. E\_Hi\_Cancer |
| 1. E\_Hi\_HeartDis |
| 1. E\_Ot\_Cancer |
| 1. E\_Ot\_HeartDis |
| 1. E\_Wh\_Cancer |
| 1. E\_Wh\_HeartDis |
| 1. F\_Bl\_Cancer |
| 1. F\_Bl\_HeartDis |
| 1. F\_Hi\_Cancer |
| 1. F\_Hi\_HeartDis |
| 1. F\_Ot\_Cancer |
| 1. F\_Ot\_HeartDis |
| 1. F\_Wh\_Cancer |
| 1. F\_Wh\_HeartDis |

User selects -> 1

Select State code:

1. AK
2. AL
3. AR
4. AZ
5. CA
6. CO
7. CT
8. DC
9. DE
10. FL
11. GA
12. HI
13. IA
14. ID
15. IL
16. IN
17. KS
18. KY
19. LA
20. MA
21. MD
22. ME
23. MI
24. MN
25. MO
26. MS
27. MT
28. NC
29. ND
30. NE
31. NH
32. NJ
33. NM
34. NV
35. NY
36. OH
37. OK
38. OR
39. PA
40. RI
41. SC
42. SD
43. TN
44. TX
45. UT
46. VA
47. VT
48. WA
49. WI
50. WV
51. WY

User selects -> 1

Output: The mean value of A\_Wh\_Comp parameter for the state of Alaska is 28.25.

1. **Submission:** Submit your NetBeans project as a Zip file on blackboard. Also submit executable jar file **CHSIDataProcesing.jar** which will be used for testing the project.
2. **Evaluation:** A slot will be announced for evaluating the assignments. During this time the instructor will download and run your program and it will be checked for correctness using a few use-cases. If required, instructor will contact you for clarifications.

**Note:**

**If your code does not compile, it will not be graded.**

**Late submissions will not be accepted under any circumstances.**

**To be safe, always, ALWAYS, prepare to submit ahead of time, not exactly AT last moment!**

**Submission deadline: Monday 25 March, 11:59 PM**

1. Community Health Status Indicators (CHSI) to Combat Obesity, Heart Disease and Cancer, <https://catalog.data.gov/dataset/community-health-status-indicators-chsi-to-combat-obesity-heart-disease-and-cancer>, Last Updated: February 26, 2019. [↑](#footnote-ref-1)
2. Description of statistics formulas, <https://blog.udemy.com/statistics-formula/>. [↑](#footnote-ref-2)
3. Standard deviation and variance, <https://www.mathsisfun.com/data/standard-deviation.html>, <https://www.mathsisfun.com/data/standard-deviation-formulas.html>. [↑](#footnote-ref-3)
4. Java program to calculate standard deviation, <https://www.programiz.com/java-programming/examples/standard-deviation>. [↑](#footnote-ref-4)
5. Creating an executable Jar in NetBeans, http://usarsim.sourceforge.net/wiki/index.php/Creating\_an\_Executable\_Java\_File\_in\_NetBeans. [↑](#footnote-ref-5)