

COP3530 Project 1 – Array Searches and Sorts

Due Date: Friday, 9/17/2021 11:59 PM

Submit the zipped Eclipse project including at least Project1.java, State.java and States1.csv. The zip file should be named <your last name>_Project1.zip (for example, Liu_Project1.zip). The States1.csv file contains information about all 50 States in the USA. For each State, in the CSV are its name, capitol, region, US House seats and population (according to the 2020 Census found in *Wikipedia* at

https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States_by_population), numbers of COVID-19 total cases and deaths (as of Sept 2, 2021, according to *Centers for Disease Control and Prevention* at <https://covid.cdc.gov/covid-data-tracker/>), median household income (according to the 2020 *World Population Review* at <https://worldpopulationreview.com/state-rankings/median-household-income-by-state>), and violent crime rates (according to the 2019 *FBI report* at <https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019>, did not find 2020 FBI data for this).

Your programs should be well documented in the format of doc comments in Java. Detailed formats are found at <http://www.oracle.com/technetwork/articles/java/index-137868.html>.

Requirements:

1. Create a class named **State** that will store information about a State and provide methods to get, and set the data, and compare the States by several fields.
 - a. Fields: name, capitol, region, US House seats, population, COVID-19 cases, COVID-19 deaths, median household income, and violent crime rate.
 - b. Constructor
 - c. Get and set methods for each field
 - d. Method to print a State object
2. Create a class named **Project1** that will:
 - a. Prompt user to enter the name of the CSV file, e.g., States1.csv, as input to the system.
 - b. Parse States1.csv and create an array of State objects containing the data in the file. (Note: do not change States1.csv file at all.)
 - c. Repetitively print to the user a menu of the following options:
 - 1) Print a States report
 - 2) Sort by **name** (alphabetically using **Bubble** sort)
 - 3) Sort by **COVID-19 case fatality rate** (CFR) (ascendingly using **Selection** sort): CFR is the ratio of deaths over cases.
 - 4) Sort by **median household income** (ascendingly using **Insertion** sort)
 - 5) Find and print a State for a given name (using **binary search** if the data is sorted by name; **sequential search**, if not)
 - 6) Print **Spearman's ρ** correlation matrix in the following format

| | | |
|--------------------|-------------------------|--------------------|
| | Median Household Income | Violent Crime Rate |
| COVID-19 Case Rate | X_1 | X_2 |

COP3530 Project 1 – Array Searches and Sorts

| | | |
|------------------------|----------------|----------------|
| COVID-19 Death Rate | X ₃ | X ₄ |
|------------------------|----------------|----------------|

(**Case rate** is the number of cases per 100,000 population. That is, case rate of a state is equal to its cases divided by its population and then times 100,000. **Death rate** is defined similarly. **Violent Crime Rate** from our dataset already is per 100,000. The Spearman's ρ correlation between two fields is defined next section. X's are the actual values your program computed, printed 4 digits after decimal point.)

7) Quit

- d. Project1 should have main and methods for each option 1-6. Do check for validity of user inputs.
- e. The State report in option 1 should be in this form:

| Name | MHI | VCR | CFR | Case Rate | Death Rate |
|---------|-------|-------|----------|-----------|------------|
| Alabama | 50536 | 510.8 | 0.017373 | 14126.05 | 245.41 |
| ... | | | | | |

(MHI is shorthand for Median Household Income, VCR for Violent Crime Rate, and CFR for Case Fatality Rate. Note: whenever printed, CFR should be printed 6 digits after the decimal point, Case Rate 2, and Death Rate 2.)

- f. The State report in option 5 should be in this form:

```
Name:      Alabama
MHI:       50536
VCR:       510.8
CFR:       0.017373
Case Rate: 14126.05
Death Rate: 245.41
```

(Note: CFR should be printed 6 digits after the decimal point, Case Rate 2, and Death Rate 2.)

Spearman's ρ Correlation

One way to tell the correlation between two rankings, whether they are positively or negatively related, of the same set of n elements is Spearman's ρ . Let R_1 and R_2 be two **ascending** rankings of the same set of n elements. Then the Spearman's ρ between R_1 and R_2 is

$$1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

In the formula, d_i is the rank difference of element i in the two rankings R_1 and R_2 .

Take one example. Let A, B, C, D be four states. Say we are to compute the Spearman's ρ between Case Rate and Income for these four states. First, we would sort to get a ranking of them according to Case Rate, say, $A < C < D < B$, and sort to get a ranking of them per Income, say, $B < D < A < C$. Then, we can use the above formula to calculate ρ between these two rankings:

1. Compute all d_i 's: $d_A = 1-3 = -2$, $d_B = 4-1 = 3$, $d_C = 2-4 = -2$, $d_D = 3-2 = 1$.
2. Compute the sum of all d_i^2 's: $(-2)^2 + 3^2 + (-2)^2 + 1^2 = 18$
3. Compute ρ : $1 - \frac{6 \cdot 18}{4 \cdot (4^2 - 1)} = -0.8$

COP3530 Project 1 – Array Searches and Sorts

De facto, ρ 's value is guaranteed to be between -1 and 1, where -1 indicates two rankings exactly opposite, and 1 exactly the same. Hence, when ρ is positive (negative), the two rankings are positively (negatively, resp.) related. If ρ is 0, the two rankings are not related at all. In our example, since $\rho = -0.8$, it indicates the two rankings based on Case Rate and Income are negatively related: Case Rate increases as Income drops.

To build the matrix for option 6, your program needs to compute 4 Spearman's rho values between Case Rate and MHI, between Case Rate and VCR, between Death Rate and MHI, and between Death Rate and VCR.

Provide comments in this form for the **State** class:

Comments for the class:

```
/**
 * Detailed description of the class.
 *
 * @author <your name>
 * @version <date you last changed the class>
 */
```

Public method comments:

```
/**
 * Description of the purpose of the method, the meaning of the
 * input parameters (if any) and the meaning of the return values
 * (if any).
 *
 * @param parameter description of the parameter (one for each)
 * @return description of the return value
 */
```

Provide comments in this form for the **Project1** class.

Comments for the class:

```
/**
 * COP 3530: Project 1 - Array Searches and Sorts
 * <p>
 * Description of the class using as many lines as needed
 * with <p> between paragraphs. Including descriptions of the
 * input required and output generated.
 *
 * @author <your name>
 * @version <the date you last modified the program>
 */
```

```
public class Project1
{
```

Public method comments:

```
/**
 * Description of the purpose of the method, the meaning of the
 * input parameters (if any) and the meaning of the return values
 * (if any).
```

COP3530 Project 1 – Array Searches and Sorts

```
*  
* @param parameter description of the parameter (one for each)  
* @return description of the return value  
*/
```

Example Outputs:

COP3530 Project 1
Instructor: Xudong Liu

Array Searches and Sorts
Enter the file name: States1.csv

There were 50 records read.

```
1. Print a States report  
2. Sort by Name  
3. Sort by Case Fatality Rate  
4. Sort by Median Household Income  
5. Find and print a given State  
6. Print Spearman's rho matrix  
7. Quit  
Enter your choice: 1
```

| Name | MHI | VCR | CFR | Case Rate | Death Rate |
|-------------|-------|-------|----------|-----------|------------|
| Mississippi | 54970 | 404.9 | 0.014075 | 11814.71 | 166.29 |
| ... | | | | | |

```
1. Print a States report  
2. Sort by Name  
3. Sort by Case Fatality Rate  
4. Sort by Median Household Income  
5. Find and print a given State  
6. Print Spearman's rho matrix  
7. Quit  
Enter your choice: 3
```

States sorted by COVID-19 Case Fatality Rate.

```
1. Print a States report  
2. Sort by Name  
3. Sort by Case Fatality Rate  
4. Sort by Median Household Income  
5. Find and print a given State  
6. Print Spearman's rho matrix  
7. Quit  
Enter your choice: 1
```

| Name | MHI | VCR | CFR | Case Rate | Death Rate |
|--------|-------|-------|----------|-----------|------------|
| Alaska | 77640 | 867.1 | 0.005096 | 11638.81 | 59.31 |
| ... | | | | | |

```
1. Print a States report
```

COP3530 Project 1 – Array Searches and Sorts

2. Sort by Name
3. Sort by Case Fatality Rate
4. Sort by Median Household Income
5. Find and print a given State
6. Print Spearman's rho matrix
7. Quit

Enter your choice: **5**

Enter State name: **Florida**
Sequential search

Name: Florida
MHI: 55660
VCR: 378.4
CFR: 0.013949
Case Rate: 15280.88
Death Rate: 213.15

1. Print a States report
2. Sort by Name
3. Sort by Case Fatality Rate
4. Sort by Median Household Income
5. Find and print a given State
6. Print Spearman's rho matrix
7. Quit

Enter your choice: **5**

Enter State name: **Guam**
Sequential search

Error: Guam not found

1. Print a States report
2. Sort by Name
3. Sort by Case Fatality Rate
4. Sort by Median Household Income
5. Find and print a given State
6. Print Spearman's rho matrix
7. Quit

Enter your choice: **2**

States sorted by Name.

1. Print a States report
2. Sort by Name
3. Sort by Case Fatality Rate
4. Sort by Median Household Income
5. Find and print a given State
6. Print Spearman's rho matrix
7. Quit

Enter your choice: **1**

| Name | MHI | VCR | CFR | Case Rate | Death Rate |
|---------|-------|-------|----------|-----------|------------|
| Alabama | 50536 | 510.8 | 0.017373 | 14126.05 | 245.41 |
| . . . | | | | | |

1. Print a States report
2. Sort by Name
3. Sort by Case Fatality Rate

COP3530 Project 1 – Array Searches and Sorts

4. Sort by Median Household Income
5. Find and print a given State
6. Print Spearman's rho matrix
7. Quit

Enter your choice: **5**

Enter State name: **Kentucky**

Binary search

Name: Kentucky
MHI: 50589
VCR: 217.1
CFR: 0.013361
Case Rate: 12916.18
Death Rate: 172.58

1. Print a States report
2. Sort by Name
3. Sort by Case Fatality Rate
4. Sort by Median Household Income
5. Find and print a given State
6. Print Spearman's rho matrix
7. Quit

Enter your choice: **4**

States sorted by Median Household Income.

1. Print a States report
2. Sort by Name
3. Sort by Case Fatality Rate
4. Sort by Median Household Income
5. Find and print a given State
6. Print Spearman's rho matrix
7. Quit

Enter your choice: **1**

| Name | MHI | VCR | CFR | Case Rate | Death Rate |
|-------------|-------|-------|----------|-----------|------------|
| Mississippi | 45081 | 277.9 | 0.019181 | 14977.01 | 287.27 |
| . . . | | | | | |

1. Print a States report
2. Sort by Name
3. Sort by Case Fatality Rate
4. Sort by Median Household Income
5. Find and print a given State
6. Print Spearman's rho matrix
7. Quit

Enter your choice: **6**

| | | |
|------------|---------|--------|
| | MHI | VCR |
| Case Rate | -0.4... | 0.3... |
| Death Rate | X | X |

1. Print a States report
2. Sort by Name

COP3530 Project 1 – Array Searches and Sorts

3. Sort by Case Fatality Rate
4. Sort by Median Household Income
5. Find and print a given State
6. Print Spearman's rho matrix
7. Quit

Enter your choice: **23**

Invalid choice enter 1-7: **0**

Invalid choice enter 1-7: **A**

Invalid choice enter 1-7: **7**

Have a good day!

COP3530 Project 1 – Array Searches and Sorts

Potential FAQ That May Help You:

1. Where should I put the CSV file in the project directory?

The CSV file can be put anywhere in the project folder. I suggest you directly put it inside project root directory, the project folder created for you by Eclipse. This way you may enter the name, rather than the path, of the file at run time.

2. Can I use ArrayList instead of arrays for Project 1?

In general, data structures you are asked to use or implement in all projects are those you studied in our course. So when you use arrays for project 1, they will be arrays, not ArrayList in java.util package.

3. How do I parse the CSV file to get the objects into an array?

There is one example program provided in one video on Canvas. Refer to it to see how fields in a CSV file can be extracted. Conceptually, to get the objects into an array, you'd create an array first, and, for each line of the file, create one object using the tokens and fill it into the array.

4. Why am I getting InputMismatchExceptions when calling nextInt?

Check the delimiters that you used and make sure it includes new line ('\n') and it does not include any space. Also, make sure you are not trying to parse and first header line.

5. What size of the object array should I use to create it, because it is unknown how many records there are in the input file?

You may assume there is a maximum number of records in the file and stick to that. You also may first read the file to get an idea of how many records. Or, you may start assuming there is a certain number of records and use that. During reading the file, if that is not enough, create a new array doubling the size, move things from the old array to the new array, and start using the new array. When this new array becomes insufficient, repeat this process. Either way is fine for this project.

6. How do I calculate the Spearman's ρ ?

There is an example above and it should give you the idea how rho is calculated on a rather small scale which you can generate to the bigger scale for the project. The computation of rho takes two rankings, so you may need extra arrays to work on besides the array you created while parsing the file. These extra arrays could be copies of it.

7. How precise should I present the Spearman's ρ ?

As stated in the project document, I ask you to print 4 digits after the decimal point.

8. Can I create classes besides Project1.java and State.java?

Yes.