Raspberry Jelly Jam
James Smith, Joelle Lum, Ryan Siu

APCS1 pd5
1/10/17

Final Project Proposal

Terminal CSV Tool

Description: Our CSV Tool is a tool used on the command line to work with and create csv data. This tool allows a user to sort columns, perform inner and outer joins on two csv files or columns, search for words or numbers in a csv, selection of specific columns for output, and perform mathematical operations on columns such as addition, subtraction, multiplication, division, and operations relating to statistical analysis. This tool will also allow for a user to print out csv's in a nice ASCII format (see below). The idea behind this project is to give a user a quick and easy way to work with CSVs through the terminal so they can also utilize the functionality and convenience of terminal commands.

CSVArray class

CSVArray is a data structure class that will be used for all operations in our CSVTool. It's data structure is that of a 2D ArrayList of Objects where each row will be an ArrayList. An instance will be created by providing a filename for the constructor when instantiating. In order to read the data from the CSV file, we will use Scanner. We will have a type-deciding algorithm that will create an object to represent each of the elements in the CSV. The advantage of doing this is that we can easily decide what can be performed on a column and each individual element because we will be utilizing the wrapper classes of primitives(ex. Integer) and Strings. This class will also have the ability to convert itself back into a csv format. The use of this is so that after the user is done manipulating it, he/she can call that method and transform a CSVArray instance into a csv format that the user can then do whatever he/she wishes.

CSVGeneral

Contains the basic methods: get, set, number of columns/rows, delete, sort, inner join, outer join, search, add column/row, and pretty printing.

CSVMath

Contains the basic mathematical operation methods: Addition, subtraction, multiplication, division.

CSVStat

Contains the statistics methods: Stat summary, box and whisker, mean, standard deviation, correlation.

CSVTool class

CSVTool imports CSVGeneral, CSVMath, and CSVStat along with the CSVArray class. It also includes a help method that prints out all of the commands available for the user. The true point of this class lies in the main method where it will parse arguments and handle them to create an instance of CSVArray and use the correct methods to accomplish what the user wants.

Methods of CSVTool

Help

- Will print out a help page that will have each method and a quick summary of what it does
- Will pop up if you call it explicitly or if you compile without any arguments
    - java CSVTool.java file.csv --help
    - java CSVTool.java file.csv

Methods of CSVGeneral

Get

- Returns the value at a cell given input coordinates
    - java CSVTool.java file.csv --get_cell 2 3
- Returns a column of a csv
    - java CSVTool.java file.csv --get_col 5
- Returns the row of a csv
    - java CSVTool.java file.csv --get_row 5

Set

- Changes the value at a cell given input coordinates and the new value
    - java CSVTool.java file.csv --set 2 3 “bob”
- Sets the column at a given column number to a new column provided that the new column is in csv format
- Sets the row at a given row number to a new row provided that the new column is in csv format

Number of Columns/Rows

- Returns the max number of rows or columns in the CSV
    - java CSVTool.java file.csv --num_rows
    - java CSVTool.java file.csv --num_cols

Add Column

- Accepts a CSV column and adds it to a given CSV file and outputs the new CSV
- Can be added at a specific position or will default to the end
    - java CSVTool.java file.csv --add_col $(java CSVTool.java file.csv --get_col 5)

Add Row

- Accepts a CSV row and adds it to a given CSV file and outputs the new CSV
- Can be added at a specific position or will default to the end
    - java CSVTool.java file.csv --add_row $(java CSVTool.java file.csv --get_row 5) 2
    - java CSVTool.java file.csv --add_row $(java CSVTool.java file.csv --get_row 5)

Delete

- Accepts a coordinate and changes the value at that cell to null
    - java CSVTool.java file.csv --delete_cell 3 4

- Accepts a row number and shifts all the rows below it up 1
    - java CSVTool.java file.csv --delete_row 2
- Accepts a column number and shifts all the columns to the right of it left one
    - java CSVTool.java file.csv --delete_col 2

Sort - Allows a user to sort columns in ascending or descending order. Sorts strings alphabetically and by capitalization.

- Chooses a sorting method depending on the sortedness of the data
- Sortedness can be represented as the reciprocal of number of non-descending sequences
- Use a threshold value (if sortedness is above threshold, then use insertion sort, otherwise use selection sort)
    - java CSVTool.java file.csv --sort_col 3

Inner Join

- Takes in two columns from the same or different CSV files and outputs a column that is the intersection of the two columns.
    - java CSVTool.java file.csv --inner_join 2 3

Outer Join

- Takes in two columns from the same or different CSV files and outputs a column that is the union of two columns
    - java CSVTool.java file.csv --outer_join 2 3

Search

- Will output the first instance of the input it finds
- Takes in a digit or string and outputs the entire row that has the given input in them.
    - java CSVTool.java file.csv --search_row apple
- Takes in a digit or string and outputs the coordinates that have the given input in them.
    - java CSVTool.java file.csv --search_cell “juice”
- Takes in a digit or string and outputs the column that has the given input
    - java CSVTool.java file.csv --search_col 2.317

Pretty Printing

- Outputs a CSV in an organized ASCII based grid (see below for an example)
- java CSVTool.java file.csv --pretty_print

Methods of CSVMath:

- Addition: Will output the sum of 2 columns (adds across rows)
    - If given only a single column it will output the sum of the column
    - java CSVTool.java file.csv --add_col 3 5
- Subtraction: Will output the difference of two columns
    - java CSVTool.java file.csv --subtract_col 3 5

- Multiplication: Will output a column that is the outcome of multiplying across the rows of 2 columns.
    - java CSVTool.java file.csv --multiply_col 3 5
- Division: Will output a column that is the outcome of dividing across the rows of 2 columns.
    - java CSVTool.java file.csv --divide_col 3 5

Methods of CSVStat:

- 5 Number Summary: Outputs the minimum, 1st quartile, median, 3rd quartile and maximum of a column
    - java CSVTool.java file.csv --stat_summary 4
    - Box and whisker graph
        - java CSVTool.java file.csv --box_whisker 5
- Average: Outputs the average of a column
    - java CSVTool.java file.csv --mean 2
- Standard Deviation: Outputs the standard deviation of a column
    - java CSVTool.java file.csv --stddev 7
- Correlation between two data sets, as specified by the user
    - java CSVTool.java file.csv --correlation 5

MVP(In no specific order)

- Functional CSVArray class
- argument parsing and handling the arguments in CSVTool
- Get
- Set
- Add Column/Row
- Delete

Stretch

Continually implement the other methods until the project is due.

UX: The user should have basic knowledge of terminal commands and abilities to fully utilize this program. A description of this program (its methods and how to use them) will be in the readme file.

TC:

- Sorting
- Searching
- ArrayList
- Wrapper class
- 2d arrays
- Scanner

prettyPrint output: