


Assignment Overview

This assignment focuses on the design, implementation and testing of Python programs to process data files to allow for queries about the measles inoculation rate over time around the world, as described below.

Learning Goals

1. Understand how to process csv file information
2. Practice using an external toolkit to support major functionality (visualization)
3. Understand the relationship between methods for processing real world data and the kinds of conclusions that can be drawn.

Assignment Specification

The World Health Organization (WHO) compiles data about immunization levels around the world. The file named “[measles.csv](#) ” contains data about the level of measles vaccinations in various countries over time.

Each line of the file contains the following information

- Country (50 characters)
- Income Level (6 characters)
- Percent Vaccinated (3 characters)
- Region (25 characters)
- Year (4 characters)

The “Country” field contains the name of the country.

The “Income Level” field identifies the category assigned to that country by the World Bank:

WB_LI low income

WB_LMI lower middle income

WB_UMI upper middle income

WB_HI high income

The “Percent Vaccinated” field contains an integer number representing the percentage of children in that country who have received measles vaccine by the age of one.

The “Year” field contains the year for which the data was compiled.

Part 1: Importing and Querying the Data

1. The program will copy selected lines from “measles.csv” into a file selected by the user.
2. a) The program will always read from “measles.csv” (it will not prompt the user for the name of the input file). If it is unable to open that file, the program will halt with an error message.
3. b) The program will prompt the user for the name of the output file. If that file does not exist, the program will create it and continue. If that file does exist, the program will discard the current contents of the file and continue (it will overwrite it).
4. c) The program will prompt the user to enter a year, and will then prompt the user to enter an income level. The income level must be one of the characters in the set {1, 2, 3, 4}, where 1 corresponds to “low income”, 2 corresponds to “lower middle income”, 3 corresponds to “upper middle income” and 4 corresponds to “high income”.

The program will copy all data of “measles.csv” selected by the user’s response. A column is selected if the user’s response matches the Year field. A row (country) is selected if the income level matches. All data (in either/both rows and columns) are selected if the user’s response is “all” or “ALL”.

For example, a line whose Year field contains “1987” would be selected by any of the following user responses: “1987”, “all”, “ALL”.

1. d) The output file created by the program will have the same format as the input file.
Note that when the user selects all lines, the output file will be identical to the input file.
2. The program will identify all records in the input file which match the user’s criteria for year and income level, and the program will display a report with the following information to the console:

The count of records in the input file that match the user’s criteria

The average percentage for those records (displayed with one fractional digit)

The country with the lowest percentage for those records

The country with the highest percentage for those records

The name of the country and the percent of children vaccinated will be displayed for the last two items (lowest percentage and highest percentage).

4. The program will display appropriate messages to inform the user about any errors.

Part 2: Visualization

Use a toolkit to graph the data. Save and include a good quality view of the image; include it in the essay in part 3.

Part 3: Reflection Essay

1. Discuss what you learned and compare and contrast the kinds of questions can be asked and answered by the data queries vs. the visualization. Examples: what limitations are introduced by the data elements (e.g. World Bank income levels)? what granularity of data is provided by the visualization?
2. Explore an ethical issue introduced by the data and explained by the data. Examples: you could choose a specific country and research why measles vaccine rates developed as they did (e.g. effects of colonization, funding, political opposition, etc.); you could investigate the role of the World Bank and World Health Organization and how they make these measurements and what purposes they serve.

Each part should take between 750-1000 words. Standard academic formatting and citations should be used (e.g. in-text citations using APA format; a final References list should be provided).

Assignment Notes

1. A simple approach to determining the maximal value in a set of values is described below.

Initialize the current maximal value to something smaller than all valid values

For each value in the set:

If the value is larger than the current maximal value, save it as the current maximal value

A similar procedure can be used to determine the minimal value in a set of values.