CSCI 203 FINAL PROJECT - PART 2
Fall 2022
YOUR ANALYSIS OF A HEAT-RELATED HEALTH ISSUE

Introduction

In Part 1 of the Final Project, you wrote code for organizing and visualizing data. For Part 2 of the Final Project, you will:

- pose a question on a heat or heat-related health issue,
- select a dataset corresponding to your area of inquiry,
- analyze the data with Python to answer your question, and
- illustrate your results using Matplotlib.

You may use one or both of the large datasets provided for Part 1 of the Final Project or select a different dataset at this website using the content label: Heat & Heat-related Illness

https://ephtracking.cdc.gov/DataExplorer/

Details on the available data are given at this websites for:

- Emergency Department Visits for Heat and Heat-related Illnesses
- Historical Temperature and Heat Index
- Hospitalizations for Heat and Heat-related Illnesses
- Mortality from Heat and Heat-related Illnesses
- Projected Temperature & Heat
- Vulnerability & Preparedness

Use the pull-down menu under the Indicator label to see explanations of the data for the above topics.

Look at the csv file with the hospitalization data that you were given for Part 1. We only looked at how the total yearly hospitalizations varied for one state at a time. The provided file has data to answer other questions:

- How do yearly hospitalizations vary by reported gender?
- How do the yearly hospitalizations vary by age of the patient?
- Do hospitalizations vary by state? How would you compare different states? Should you look at different geographical regions (northeast, southeast, midwest, northwest, or southwest) for comparisons?

Using some of the other datasets available at https://ephtracking.cdc.gov/DataExplorer/, you could answer other questions such as: How do the hospitalizations normalized for population density vary by state? Instead of looking at hospitalizations, you could ask questions about how mortality or emergency-department visits for heat and heat-related illnesses vary by some variable.

The amount of available data is vast, as are the many questions that you could ask. The idea is that you are using your new skills to investigate a question that is meaningful to you. The only restriction is that at least one of your datasets must have data on a heat or heat-related health issue.

YOUR PROPOSAL - DUE by 11 pm on Tuesday, Nov 29, 2022

Use the template on Moodle to: (1) formulate a question about a heat or heat-related health issue and (2) thoughtfully consider the steps needed to complete your analysis.
When you complete your proposal, upload the document to Moodle as a pdf. Your proposal should be broken into the following sections.

- **Overview and Motivation** What question(s) do you wish to answer? What led you to this question and/or why do you find this question interesting? (A paragraph is sufficient.) You may use questions suggested above or come up with your own.
- **Dataset(s)** What dataset are you using to answer your questions and where did you find the data? Describe the dataset and **include link(s) to the source(s)**. If using the dataset we already provided, specify what parts of that dataset.
    - Select data that align with the overarching question(s) that you are planning to answer.
    - Include a preliminary count of the number of measurements in your dataset. Ideally, *the total number of cells in your spreadsheet or pieces of data that you analyze should be at least 10K*. (For reference, the file hospitalizations.csv has 5,741 rows. Each row has 6 columns with data. So the number of cells would be 5,741 * 6 or 34,446.) Smaller datasets may be considered if you can provide a compelling argument for using a particular dataset.
- **Analysis:** How will you analyze your data? There are many potential analyses of data such as comparing maximum values, minimum values, means, or running averages. Like any good analysis, think of an interesting question first. Then, determine which kinds of analysis will help you answer that question. List what functions you will use. You need at least four functions such as:
    - A function to read data from a file.
    - A function to organize and/or perform some operation(s) on the data.
    - A function to visualize your data.
    - A function (`main`) to run your analysis.
- **A description of your visualization.** Provide as much detail as possible such as the type of plot and how you will label the axes. Feel free to include a sketch of your proposed visualization. You will probably want to flip through existing plots on Matplotlib to get a sense of what's possible: matplotlib examples. Feel free to learn a new way to plot or show your results.
- **Challenges that you may face in completing the project.** Do you see any aspect of your analysis and visualization that may cause difficulties?

Communication is critical in computer science, as it is in any other major. Meaningful feedback on your proposal will be impossible if you do not provide a clear and concise description of your proposed work.

YOUR PRESENTATION

During the last lab (December 6, 2022), you will share the results of your analysis with the class. To do this, prepare three Google slides and share these slides with your lab instructor when you submit your final project.
The first slide should include:

- The question that you answered
- A description of your data and the source
- A description of your analysis

The second slide should show the visualization of your results.
The third slide should include any conclusions and/or recommendations for further study.

FINAL PROJECT PART 2 GRADING RUBRIC

Points will be allotted for Part 2 of the Final Project as shown in Table 1.

Table 1. Point Allotment for Part 2 of the Final Project

Points	Task
10	Project Proposal and dataset selection
3	Function to upload data
8	Function(s) to analyze data
8	Function(s) to visualize data
4	Competition of something more than uploading a single data set and plotting that data set. These points could be earned for something interesting in your selection, analysis, and/or visualization of the data.
4	Function to run analysis (a main function or the equivalent)
4	Presentation of results including three Google slides