Assignment #1: Charts and plots

Due on September 9th, before midnight

Goals and Requirements:

The goal of this assignment is to get familiar with the principles of generating effective graphical representations (specifically, charts and plots) for storytelling.

You will be given a few data sets to choose from (see below). You are required to choose TWO from the given options to finish the specified tasks. It is up to you to choose the tools to generate the appropriate plots and charts. But you need to specify which tool you use to finish this assignment in your report.

The submission of this assignment is in the form of a report (a single PDF file) with the generated plots/charts and corresponding descriptions (see below).

You should submit your report via Blackboard before the deadline.

Options of the Data:

Please choose TWO from the following options. The meaning of attributes of those data should be straightforward based on their names.

- Custom data
- Education data
- Population data
- Pollution data
- Titanic data

Tasks:

1. Writing question (20 points)

Please provide one example for scientific data AND another example for information data. Describe why you think the data you select is a scientific (or information) data. DO NOT USE THE EXAMPLES FROM THE CLASS LECTURE. IF YOU DO SO, YOU WILL NOT RECEIVE CREDIT FOR THIS QUESTION.

2. Plots/charts generation (50 points)

Choose the proper types of plots for each data set you select and generate the corresponding plots. Depending on the data sets and the attributes/columns they store, you may need multiple plots for each data sets. If you choose to use certain plots (e.g., line plots vs. histogram or scatter plots,
etc.) for the data set you select, **please explain why you think these are the proper plot types for a data set** (Hint: it depends on what you want to reveal).

**Note that, you should produce two sets of plots, one for each data set.**

Please exercise the principles we introduced in the class. **Each plot should have proper title, axis labels, caption (different from title), etc.**

3. **Knowledge discovery (30 points)**

Based on the background of the data and the individual plots generated in Task 2, **summarize your discovery**, and answer the following questions as much as you can:

- What information/pattern/structure/correlation you found from each data set you select?
- Which plot type(s) most effectively reveals the information/pattern/structure/correlation found above?
- Any possible (hidden) information that the existing plots cannot reveal (i.e., requires the design of new type of plots/charts or additional information/data source)?

**Grading rubric:**

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Total points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
</tr>
<tr>
<td><strong>Python-based plots (bonus)</strong></td>
<td><strong>5</strong></td>
</tr>
</tbody>
</table>

**Suggestions:**

Although any plotting tool can be used to complete this assignment, it is recommended to use Python Matplotlib or Seaborn to produce charts and plots for this assignment. Students who use Python + Matplotlib/Seaborn to produce plots for this assignment will receive 5 extra points (student needs to submit their Jupyter notebook to prove the use of Python).

The following book is a good reference to generate plots/charts using Python.

*AI Publishing, Data Visualization with Python for Beginners, Visualize Your Data Using Pandas, Matplotlib, and Seaborn.*

There are also plenty of resources online for producing nice plots using Python.