Visualizing Country Data for the General Public

The assignment

Visualization design

Visualization has become a tool both for the exploration of data and for the presentation of analysed data to end users. In this assignment you take the role of a data visualization designer who has been asked to design visualization that supports exploration, comparison, and analysis of global country data. The visualization is to be part of a project aimed to engage the general public in data exploration and understanding of differences between countries worldwide, and it should be accessible as an HTML page.

To enable exploration, the visualization is expected to be interactive and include a minimum of three views that display different aspects of the data, including at least 7 data indicators/attributes to enable broader understanding of the data. The visualization should support gaining of statistical insight from the data through representation of analytical/statistical patterns of relevance.

There are several constraints you will have to take into consideration in your design:

- You will typically not be there to explain the visualization, so it must be entirely standalone.
- Most, if not all, of the users will not understand statistical methods or mathematics.
- You don't know what technical equipment your users have or how they intend to use the visualization, they may print or fax a screenshot of their findings, it should hence work both on screen and on paper.

The final multiple views visualization (i.e. dashboard) should fit on a standard sized desktop screen with a maximum resolution of 1920x1080 and be presented as an HTML page.

The data

The data you have been given for this task is a slightly modified version of the Kaggle Global Country Information Dataset 2023 (https://www.kaggle.com/datasets/nelgiriyewithana/countries-of-the-world-2023), which provides global data at country level on, e.g., demographic statistics, economic indicators, environmental factors, health metrics, educational statistics, etc.

The dataset contains measures for 195 countries, across 36 attributes and indicators. The first three attributes include country identifiers (Country name, country id, and the 2-letter county abbreviation), which is followed by Latitude and Longitude data for the country. The remaining 31 attributes contain data on a range of factors and indicators, and you are expected to visualize at least 7 of these attributes. You are free to choose which attributes you want to visualize but will be expected to motivate the choice. You are also welcome to generate new attributes from the existing data, if this has justifiable benefit to analysis/visualization. Some data cleaning may be needed as the dataset include a small number of missing values.

Use of software tools

You **must** use Python and the Vega-Altair visualization library for the visualization design and for generating the HTML version of the visualization (you will learn how to save as html file in the week 4 practical). You can also use the Pandas and NumPy libraries for data management and analysis, as well as Vega datasets to access additional data if needed.

Additionally, you are allowed to use any data analysis software or programming language (i.e., Excel, R, etc.) for data wrangling/cleaning prior to visualization, and are welcome to load a wrangled version of the dataset in your Python code.

Your HTML visualization and code must work on a NUIT supported computer in the teaching laboratories.

Report and reflective evaluation

You are expected to submit a report in the form of a filled-out table proforma, where you describe and justify your visualization design, based on the task and visualization theory and guidelines, using academic references where appropriate.

As a final step of the reporting, you are expected to reflect on how your visualization meet three of the visualization heuristics of Wall et al. (2019), and how you could modify the visualization to better meet these heuristics. You do not have to implement anything for this and should not carry out an evaluation with other participants but instead reflect yourself and fill in the relevant part of the submission table.

The three heuristics to evaluate based on are:

- The visualization facilitates answering questions about the data.
- The visualization provides opportunities for serendipitous discoveries.
- The visualization provides a big picture perspective of the data.

Note: marking of the reflective evaluation is **not** based on if your visualization meets the heuristic criteria, but on your understanding of how the heuristic is applied and could be met. Hence, not meeting a heuristic criterion but having a good suggestion of how you could meet it may be marked equally high as meeting the heuristic.

What to submit

Submission on Canvas of a zip file including:

1. The visualization as a web page (as .html)
2. A technical report, using the provided grid template (as .pdf)
3. Your python code (as .py or .ipynb)
4. The dataset loaded by the Python code

The coursework deadline is 6th December 16:00

The marking scheme

To gain marks, you must demonstrate your application of visualization skills, techniques and theory against the following marking scheme. For full marks, it is expected that your submission goes above and beyond merely fulfilling the basic requirements, and that you are making good use of academic literature.

Visualization design	**Mark**
Pre-processing: Appropriateness of data pre-processing and justification of attributes to visualize	/6
Fit to task: does the visualization support exploration, comparison and understanding of differences between countries?	/8
Fit to user: is the visualization self-explanatory and appropriate for the intended user?	/8
Analysis: does the visualization present analytical/statistical patterns of relevance.	/6
Visualization design: Appropriate use of visual channels.	/8
Visualization design: Appropriate use of Gestalt theory and visualization design principles.	/8
Visualization design: Appropriate use of colour.	/8
Visualization design: Appropriateness and functionality of interactive features.	/8
Visualization design: Appropriateness and design and coordination of multiple views.	/8
Visualization design: Appropriate use of language and text in the visualization.	/6
Technical aspects: reliability of operation, fit on 1920x1080 screen.	/6
Total for visualization	**/80**
Evaluation	
Relevant reflection on how the visualization meets or does not meet the three heuristics.	/6
Relevant suggestions on how the visualization could be improved to better meet the three heuristics.	/6
Total for evaluation	**/12**
Reporting	
Logical content and structure, range and quality of references used	/8
Total	**/100**