CS 503 – Data Visualization

Final project: Exploratory Data Analysis

1. Introduction

In this final project, you will design four visualizations techniques for a challenging dataset and provide a rigorous rationale for your design choices.

2. Tasks

The dataset is about the Human Development report in 2015. In fact, there are six main factors for the Human Development that play an important role to distinguish among first world countries, countries under development and third world countries. These factors include gender development, gender inequality, historical index, human development, inequality adjusted, and multidimensional poverty. The dataset is composed of six files where each one of them represent a factor. In addition, a technical report is available for further explanation.

Step 1. Choose only two files that you prefer to study and pose two initial questions that you would like to answer as you did in the assignment 2 and in the assignment 3. You have the right to merge the two files in a one file. However, you need to justify any preprocessing of your dataset.

Step 2. Assess the fitness of the data for answering your question.

Inspect the data--it is invariably helpful to first look at the raw values. Does the data seem appropriate for answering your question? If not, you may need to start the process over. If so, does the data need to be reformatted or cleaned prior to analysis? Perform any steps necessary to get the data into shape prior to visual analysis.

Step 3. Design four visualization techniques that you believe effectively communicates the data and provide a short write-up (no more than 4 paragraphs) describing your design. This design must be implemented in Python. Illustration with interactive visualization library Plotly and the design of multidimensional visualization technique will be an asset. While you must use the data set given, note that you are free to transform the data as you see fit.

Your chart image should be interpretable without recourse to your short write-up. Do not forget to include title, axis labels or legends as needed!

As different visualizations can emphasize different aspects of a data set, you should document what aspects of the data you are attempting to communicate effectively. In short, what story (or stories) are you trying to tell? Just as important, also note which aspects of the data might be obscured or down-played due to your visualization design. In your write-up, you should provide a rigorous rationale for your design decisions. Document the visual encodings you used and why they are appropriate for the data. These decisions include the choice of visualization type, size, color, scale, and other visual elements, as well as the use of sorting or other data transformations.

3. Submission

You must submit the *pdf* file of your report. In fact, it must contain the illustration of your visualization technique in term of images (four images that illustrates four visualization techniques) and a rigorous explanation. Indeed, your explanation must justify which one of the four visualization techniques will help the reader to give a relevant interpretation. In addition, you have to submit your *.py* or *.ipynb* if you prefer using *Jupyter Notebook*. ***Please, do not submit your assignment in .zip or .rar files.*** If you do a transformation to the dataset, you will have to mention it on your report and to submit it.