

COVID-19 DATA VISUALIZATION PROJECT

CW2 (F20DV)

Abdullah Sayed (H00320650)

Submission date: April 13th

Video Link: <https://youtu.be/5lVK_NADz6c>

1 Data Handling

The first step in this project was to come up with an idea and source the data for the topic. After significant deliberation I decided to explore a couple of routes and their availability of data about ended up settling on the impact of government spending on the education levels within a population and the crime rates within the country. For this, I derived data from 3 separate sources. The sources the World Development Indicators from the World Bank, The UNESCO Institute for Statistics as well as a crowd-sourced global database for consumer prices, crime rates, quality of health care, among other statistics. One of the challenges of procuring the data is trying to find the variables that are correct to represent the story that you want to explore. The first one I chose was for my government expense. I made sure to find one that included all government expenses to make sure I have the best chance of uncovering relationships as all government spending should in theory effect all of the country's stats. For education it was tough to choose. I chose to go with the cumulative percent of the population that has at least a bachelors degree. This is due to the fact that I was under the impression that this would indicate a country that valued higher education and therefore lead to a more advanced adult population in terms of education. The final variable was a crime index based on several surveys which I intended to use as the perceived danger of different places. I downloaded the information and uploaded it to my GitHub. Then I used the d3.csv in conjunction with the permalink to load the data. I combined all the data and for the countries that don't have all 3 values they'd be removed. This give me a comprehensive data set with all the values that I required. When reading the data in, I set it to take the newest data of the past 3 years in order to get a larger data set.

2 Features

Unlike last CW, I have created a more of a web page with my visualizations that explores different things. At first there is a Correlogram that starts off by exploring the initial Pearson coefficient between the variables and reveals that both the education and the percent of a countries GDP spent actually relates to the crime rate. This makes sense as intuitively a country that is better educated and has better spending would be a better society in which crime isn't as widespread. Keeping that in mind, the next section is all about comparing how much the government spend vs its effet on the education levels as well as crime. These 2 scatter plots have a bi-directional brushing feature that allows you to select points on one graph and it shows you the corresponding points in the other graph. As the x axis is the same this allows you to view the data with a restricted y axis value and its projection on the other graph. For example, you can try too see for all countries where crime is high, what does their education look like. To further explore this, once the data is selected you can calculate and animate a linear regressor line to show you how they relate, The R-squared is also calculated and shows you objectively how well of a relationship there is. The visualization then goes to combine all the variables into a contour plot. This is to overlay the crime index over the expense vs education chart. with the regions weight being calculated by their crime index we can show how crime rates vary across different combinations of expense and education levels, revealing insights that might not be immediately from the previous scatter plots. Furthermore there are sliders that sallow the user to manipulate the contour parameters that allow them to explore and gain insights to the data in a way that makes sense to them instead of being predetermined by the programmer before hand. This adds a level of flexibility and individual ability to learn something from the data.

3 Task Completion

We are required to assess if we have passed the objectives in Task 1 which will be done by analyzing the criteria from the prompt and seeing if I fulfilled them.

" You need to design and implement an interactive web-based visualisation application using D3 on a theme of your choice. The coursework should be unique to each student (including the report, code and any data sources). Any test data or resources used for the project, should be included when you submit the final report/code. The unique visualisation dashboard you develop must be sufficiently complex, interactive and animated while containing multiple different graphical output. Your implementation must utilise multiple data sources, which are manipulated to communicate an in-depth coherent visualisation story. Appropriate data analytics techniques must be implemented. "

Below I have extracted the key points mentioned alongside their completion

Criteria	Status
Sufficiently complex	✓
Interactive	✓
Animated	✓
Multiple different graphical output	✓
Multiple data sources	✓
In-depth coherent visualisation story	✓
Data analytics techniques	✓

4 Requirements

In terms of meeting core requirements, I have met C1 as I have created a web-based application using D3 version 7. C2 has also been met as my transitions between the regressor moving provides the user of confirmation of the new calculation and how the change in selection impacted the result. My visualization being intuitive to use C3 has also been satisfied. To consider accessibility whilst designing my visualization, I chose to maintain a color contrast level of over 4.5:1 ratio between the foreground content and the background content. As for the requirements, A1 through A3 have all been satisfied as can be seen in the running of my code and its structure. A4 is satisfied by me having a correlogram, scatter plots, and 2d density plot. A5 is also satisfied although in a slightly different manner. As my data didn't have a positive and negative side per say, I showed both sides on how the data is correlated and also how the data isn't sufficient enough to be classified as the sole driving factor. Part of this uses linear regressors so that satisfies A6. A7 is also achieved in my project through the implementation of multiple sufficiently complex, interactive and animated graphical outputs.