June 4th, 2021

EPFL

Personality Traits Across the World

COM-480: Data Visualization Process Book

EPFL, Spring 2021

Group name: "Slackers"

Furkan Karakaş
Alexandre Luster
Barış Sevilmiş

Path

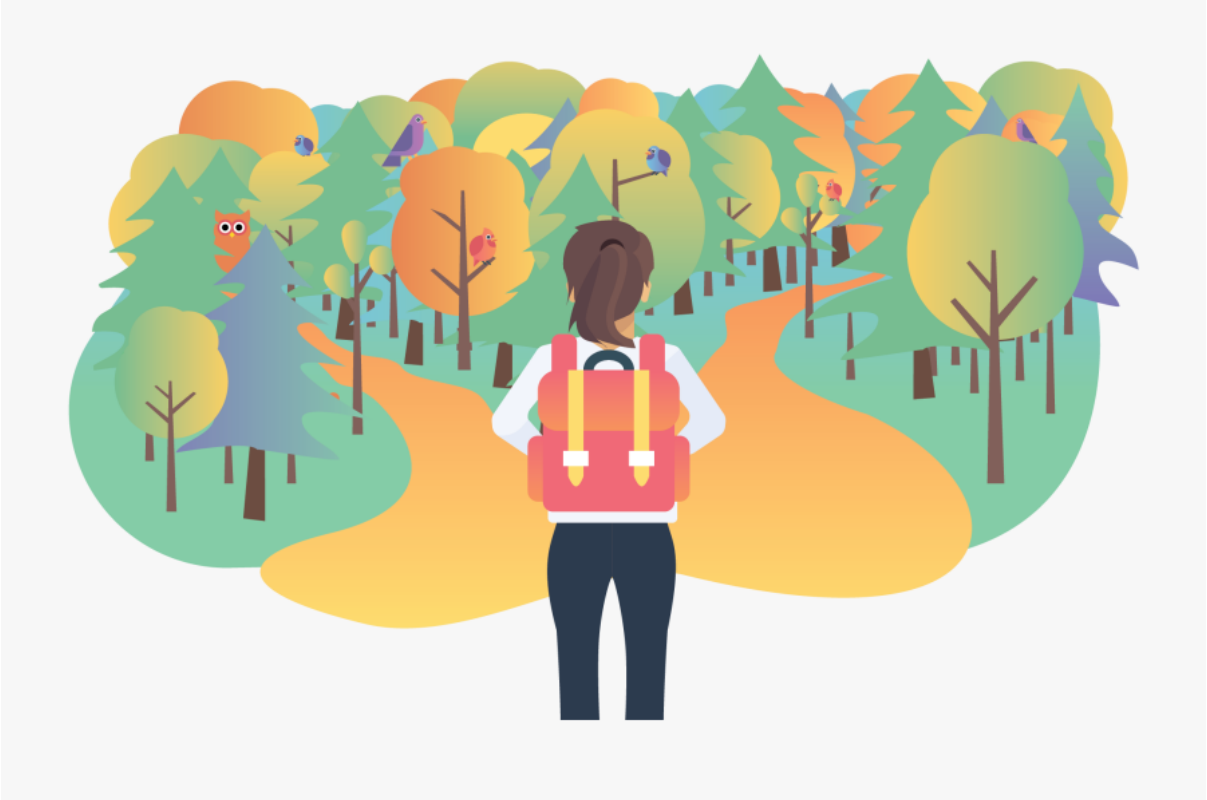

Figure 1: Alternative paths

There exist many alternative paths to design and implement a website, in which the visualization of a specific dataset can take place. HTML, CSS and JavaScript are bread-and-butter in a web development project. Since the aim of this course is to be accustomed with these programming languages, we did not want to use heavy back-end or front-end frameworks such as Node.js, React.js, Angular.js etc. Other than that, there exist many JavaScript libraries to choose from, and some of the choices are crucial depending on the data.

D3 library was among the first of our choices, since it provides very useful world map visualizations, as well as interactions such as brushing and transitions. Our map choice was based on a few criteria, which are mentioned in the Challenges & Design Decisions section. In short, our priority was to display the entire content of the dataset at once with a visually appealing map.

Crossfilter library was used to filter the data. We offer two filtering operations that can take place simultaneously and interactively. The first is filtering by age, allowing the users to choose an age range using a brush. Then, the world map will be colored only

according to the data within this age range. Filtering by sex is the second option, in which male, female or both are the choices. The map will again be colored using only the data corresponding to the chosen sex(es).

TopoJSON library was used to encode the topology of countries in our project. This library allows the geometries of countries, the lines between them and their coordinates to be stored efficiently in the same file. Once the raw data has been loaded, it provides an API such as `topoJSONdata.objects.countries` to efficiently display the countries on the world map.

D3-geo-projection provides many different functions to project world coordinates into their projected coordinates, such as the Winkel tripel projection which we ended up using.

Challenges & Design Decisions

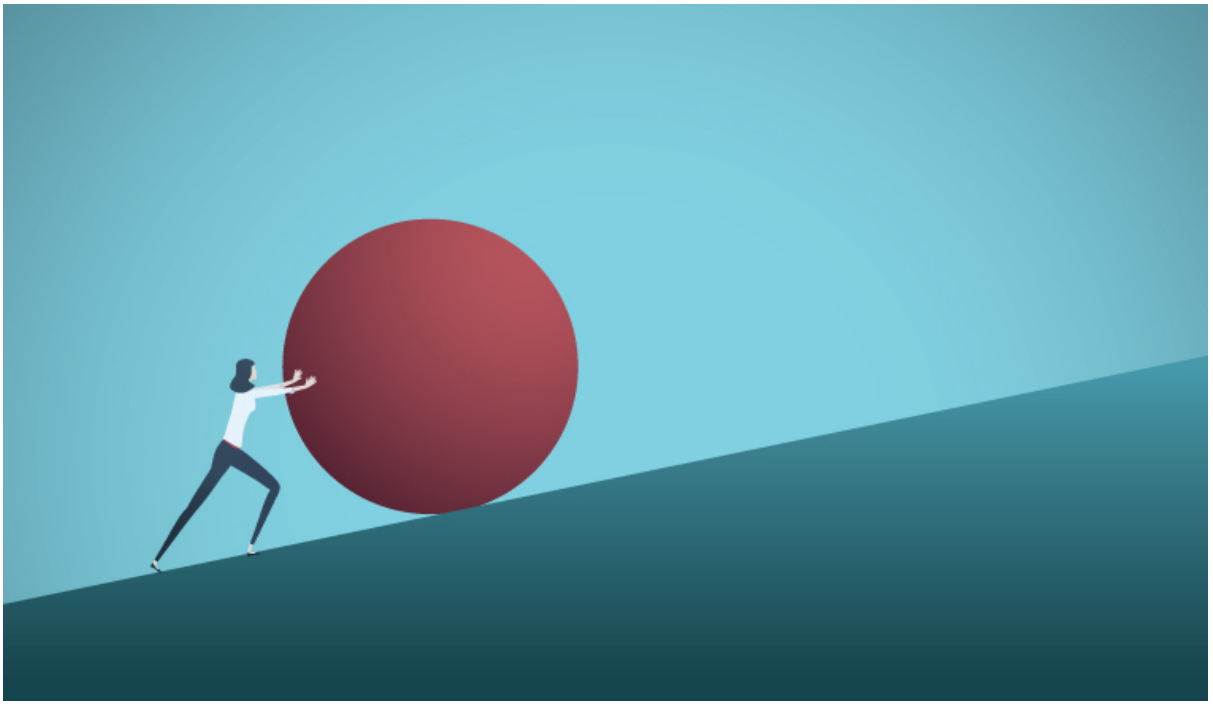

Figure 2: Challenges on the path

We faced some challenges and design decisions during the implementation:

Normalizing the scores

Instead of having a fixed color scale, we fix a relative color scale based on the current selection. The mean score for the select trait is computed for each country and

normalized to the range [0,1]. This is important so that differences between countries are more visible. For example, we wanted the definition of "high" openness to be the highest mean score, not the improbable score of 1. This is also important because "low" and "high" carries a different meaning for each trait: in the dataset, the mean score for openness is 0.734, but only 0.574 for neuroticism. Hence, it is not really a good idea to put them on the same scale.

Another reason why we wanted to use **relative** scores rather than **absolute** ones is that in the world map, the mean scores of different countries are usually close to each other. Hence, using an absolute scale over the full [0,1] range would make every country look the same with a similar color.

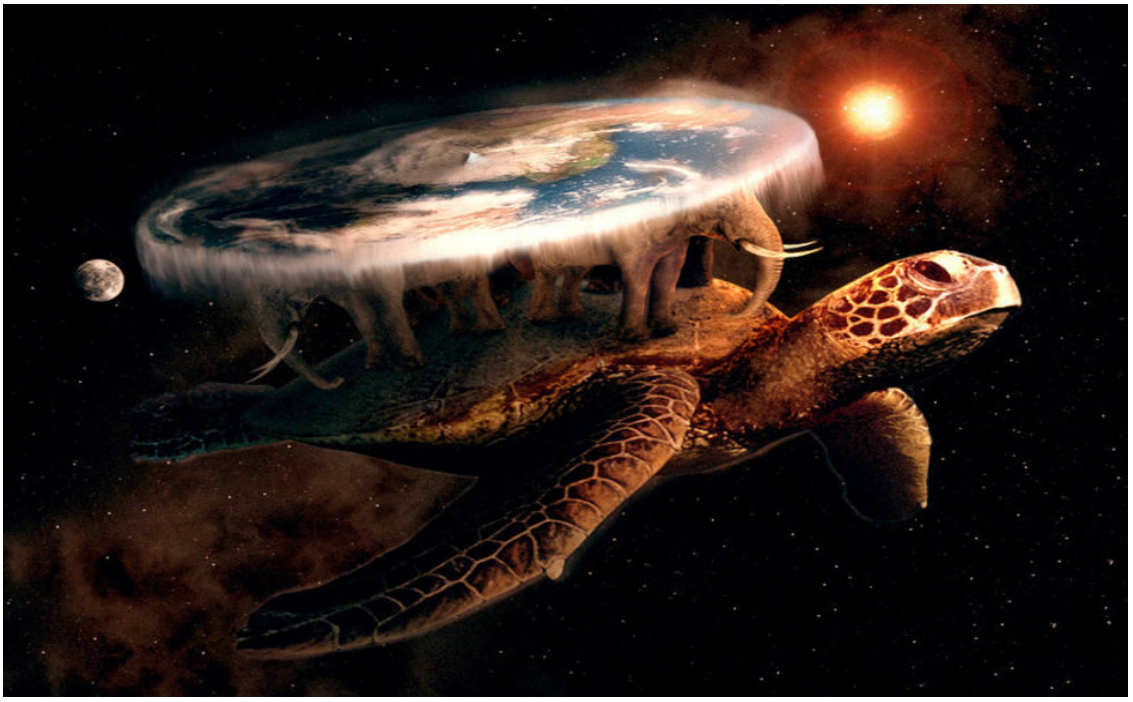

Figure 3: World map design

Map selection

Various types of map can be chosen, and 3D options are available as well. Our choice was to use the Winkel tripel projection on a 2D map. Although a 3D map provides a fancier visualization, it would require the user to interact with the map through rotation, which may decrease the user satisfaction. In addition, a 3D map could not provide a simultaneous view of all countries, making visual comparison much more difficult. The choice of the Winkel tripel projection is motivated by its resemblance to the elliptical shape, which indeed is a good fit for the image that we utilize in the

background. In addition, it minimizes area, direction and distance distortion, giving a representative view of the world.

Traits

Traits in this project are absolute and users would have to know about these traits for the project to become interesting. Instead of putting a link for users to read direct descriptions, or to provide long explanations directly, pages with corresponding trait images in background are provided. These pages contain some famous quotes relevant to these traits as well as short dictionary definitions to enable users to understand the basics of these traits. The quotes may evoke interest in users towards the subject in addition to the descriptive backgrounds.

Sample questions

Lastly, some sample questions were listed on the last page for the users to better understand the dataset. This part is quite simplistic, and it contains 12 example questions used to assess the respondents' personality, among which 6 had a positive effect on the corresponding trait, and 6 others a negative impact. For each question, respondents answer on a five-level scale (disagree, mostly disagree, neutral, mostly agree, agree).

Instead of providing a background as in for the traits descriptions, a simple white background is utilized, because there is no common theme. Each group of questions is accompanied by an image to form a visual understanding of the topic as well.

Changes from the First Milestone

In the first milestone, we mostly did data visualization in the Jupyter notebook in order to gain insightful information from the dataset. We were not sure which aspects of the data to include in our final product since we do not have a large dataset as the number of respondents in most countries is relatively small (70% of the respondents are from the United States).

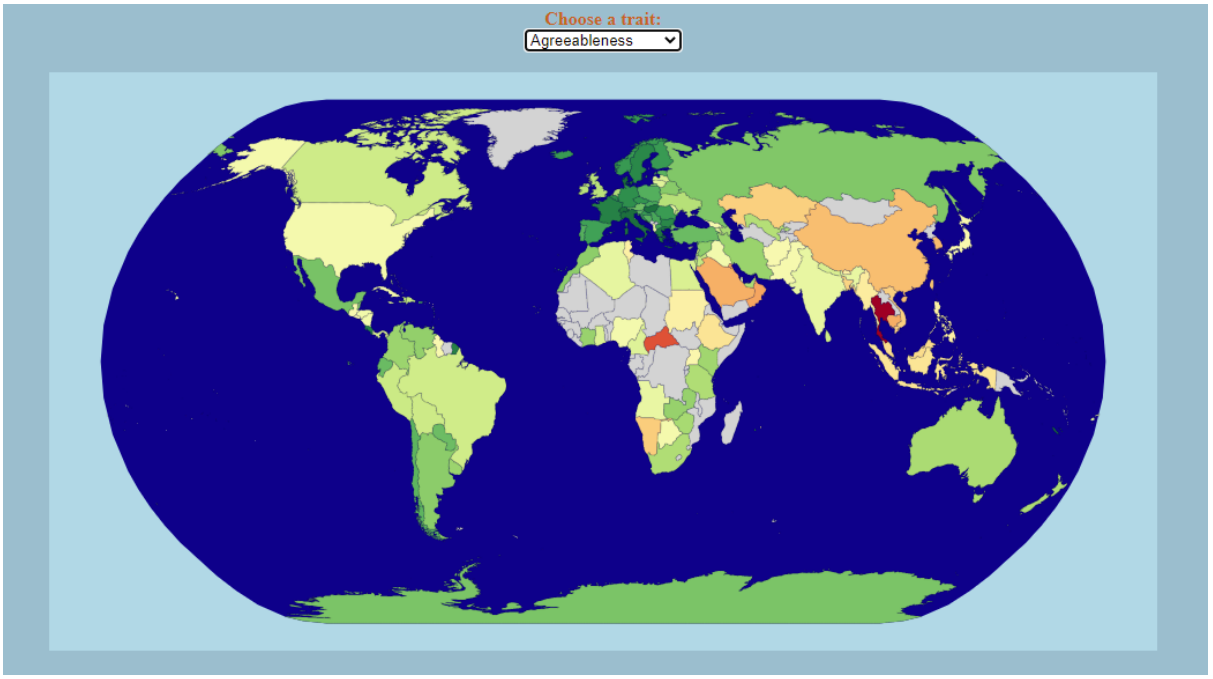

Figure 4: Initial layout of the website

In the first milestone, we wanted to include histograms to visualize the distribution of the score of a specific trait within a country. We added this histogram for discrete bins for selected combinations of traits. However, as we mentioned above, this distribution will not be very useful for the countries with few participants since the number of respondents are quite low. Other than this histogram, we decided to include bar charts where we plot the average scores of the traits with respect to four age bins "teenagers (younger than 20)", "young adults (between 20 and 39)", "middle-aged adults (between 40 and 59)" and "seniors (60 and over)". We also had in mind to put the histogram with an arrow right next to the country. However, we decided to include it in its own section so as not to crowd the main section too much.

To mitigate the effect of low participation from certain countries, we decided to add a minimum number of observations input from the user of the website. We allow the user to enforce a minimum number of observations, so that the countries below that threshold are not taken into account (shown in gray).

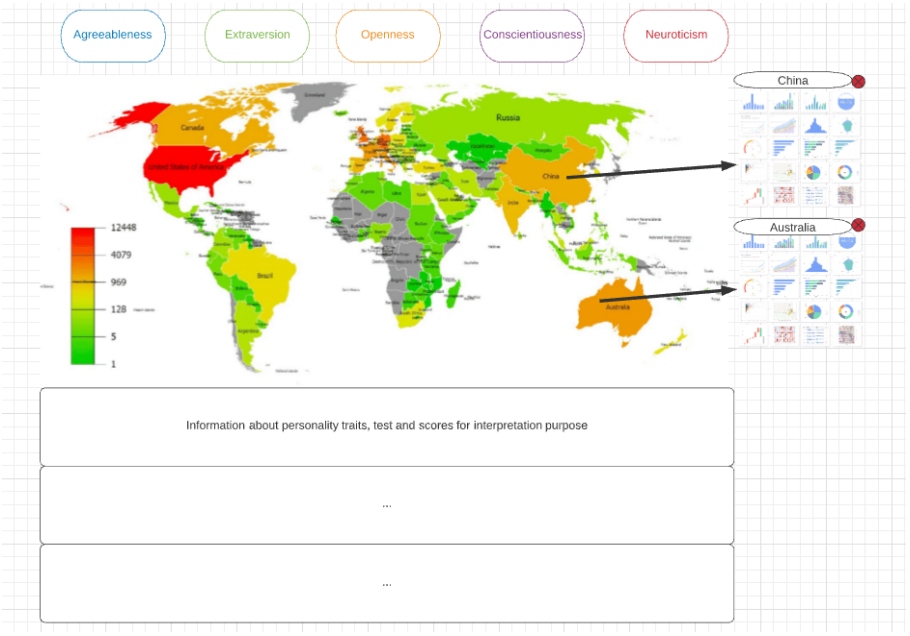

Figure 5: The initial design idea of the project

Peer Assessment

Figure 6: Peer Review

We were constantly in communication with each other during the implementation of the project. Whenever we faced some problems on the implementation side or we were not sure which API of the D3.js framework we were supposed to use in the website, we were able to consult each other so that other team members could point out useful documentation or tutorials on the web how the work could be done. In terms of design, we laid out different ideas on the table which would look better and more appealing to the user of the end product. We discussed several possible solutions, and we chose the ones that would appeal the most to the end product users. In this project, we have five major tasks to be completed:

1. integration of the world map to the raw CSV data in D3.js framework
2. implementation of the bar chart in the D3.js framework
3. making the website more appealing to the end users by writing Cascading Style Sheets (CSS)
4. capturing the events such as click, change, mouse hover, etc. of the HTML elements in JavaScript and defining the appropriate logic
5. preparing the process book and a short video

Alexandre was responsible for the first two points, and **Barış** mostly took care of doing the third part, whereas **Furkan** was focusing on the fourth part above. Nevertheless, these tasks cannot be easily separated from each other. For example, **Furkan** also worked on correcting minor mistakes in the world map and bar charts. In general, there was involvement from every team member in every part of the project, and we were able to manage this process effectively.

For the fifth part in the list of tasks, everyone worked on preparing the final process book and our team member **Furkan** recorded the short, two-minute-long video.