World happiness Process book

Shadi Naguib

Niccolò Stefanini

Thomas Rowlands

1. Introduction

Although often neglected, the research of happiness represents a critical goal of human mankind. A happy human is empirically more creative, energetic and feels healthier both physically and emotionally. Research on the key factors of happiness has historically been carried out by philosophers and writers, as happiness was thought of as an abstract, transcendental concept based solely on the mindset of an individual. Would a data-driven approach be able to provide us with new insights? This document will detail the journey of the Happiviz visualization - from conception to publication - and give insight into how and why key design and implementation decisions were made.

2. Data selection

The project began with the proposal of several candidate data sets, each of which were evaluated and debated upon before a final decision was made. Candidates included data on world passport strength, historical football results, and search engine trends, but the winner was a set displaying the happiness levels of over 150 countries worldwide.

Beginning in 2012, research was conducted on world happiness levels by asking the question "How would you rate your happiness on a scale of 0 to 10, where 10 is the happiest?" worldwide. The results, which are published annually on March 20th (World Happiness Day), are summarized on Kaggle, where they are merged together with additional country data, to form the complete data set.

The World Happiness Report Dataset includes one .csv file for each year from 2015 through 2019 (the 2020 report was recently published as a pdf, but has not made its way into the dataset yet), and ranks 156 countries based on their surveyed happiness score, along with additional data on several topics, including the economy (GDP), health (life expectancy), perceived freedom, trust in the government, and generosity.

Overall ra...	Country o...	Score	GDP per ...	Social su...	Healthy li...	Freedom ...	Generosity	Perceptio...
1	Finland	7.769	1.340	1.587	0.986	0.596	0.153	0.393
2	Denmark	7.600	1.383	1.573	0.996	0.592	0.252	0.410
3	Norway	7.554	1.488	1.582	1.028	0.603	0.271	0.341
4	Iceland	7.494	1.380	1.624	1.026	0.591	0.354	0.118
5	Netherlands	7.488	1.396	1.522	0.999	0.557	0.322	0.298
6	Switzerland	7.480	1.452	1.526	1.052	0.572	0.263	0.343
7	Sweden	7.343	1.387	1.487	1.009	0.574	0.267	0.373
8	New Zealand	7.307	1.303	1.557	1.026	0.585	0.330	0.380
9	Canada	7.278	1.365	1.505	1.039	0.584	0.285	0.308
10	Austria	7.246	1.376	1.475	1.016	0.532	0.244	0.226

Figure 2.1: Raw data sample from the 2019 data set.

The happiness dataset was attractive for several reasons. First and foremost, it was cleanly formatted and ready-to-use, allowing the focus of the project to be on running detailed analysis and creating compelling visualizations from the data, with no time being wasted on web scraping or API set-up. Furthermore, the data was well-defined across a primary metric (happiness), and several auxiliary metrics (GDP, life expectancy, etc.), and could be divided up in many ways (by country, by year, etc.), lending itself to many potentially cool visualizations.

3. Data Preprocessing

Although the dataset was generally well-formatted, there was nevertheless some preprocessing that needed to be done before the visualization design and implementation could take place.

The dataset was comprised of five separate csv files, so the first preprocessing task was to merge them all into one, and add a year column to the newly amalgamated file. Having all the relevant data in one file would make general analysis easier, and simplify the frontend code by requiring only one csv file to be loaded per visualization.

Next, names of countries had to be made consistent across every year. There were several instances where the same country was referred to by a different name in different years. This was the result of either inconsistent data logging, or a geopolitical event leading to the official name change of a country (e.g. Macedonia being renamed to North Macedonia in 2019).

```
pop = pop.set_index("country")

popToChange = {
     'Congo, Dem. Rep.': 'Congo (Kinshasa)',
     'Congo, Rep.': 'Congo (Brazzaville)',
     'Palestine' : 'Palestinian Territories',
     'Lao' : 'Laos',
     "Cote d'Ivoire": 'Ivory Coast',
     'North Macedonia': 'Macedonia',
     'Slovak Republic': 'Slovakia',
     'North Cyprus' : 'Cyprus'
}

for csv_format in popToChange.keys():
   pop.rename(index={csv_format:popToChange[csv_format]},inplace=True)
```

Figure 3.1: Script to clean and giving consistency to country names.

With the data now merged and cleaned, the final preprocessing task was to generate some preliminary visualizations, in order to gain early intuition on what certain visuals would look like. This was done using the Python library Plotly. The aim of these visuals was not to be aesthetically pleasing, but rather to gain insight on how the various metrics related to one another, and on how certain visuals would look if fully built out using JavaScript and d3.js.

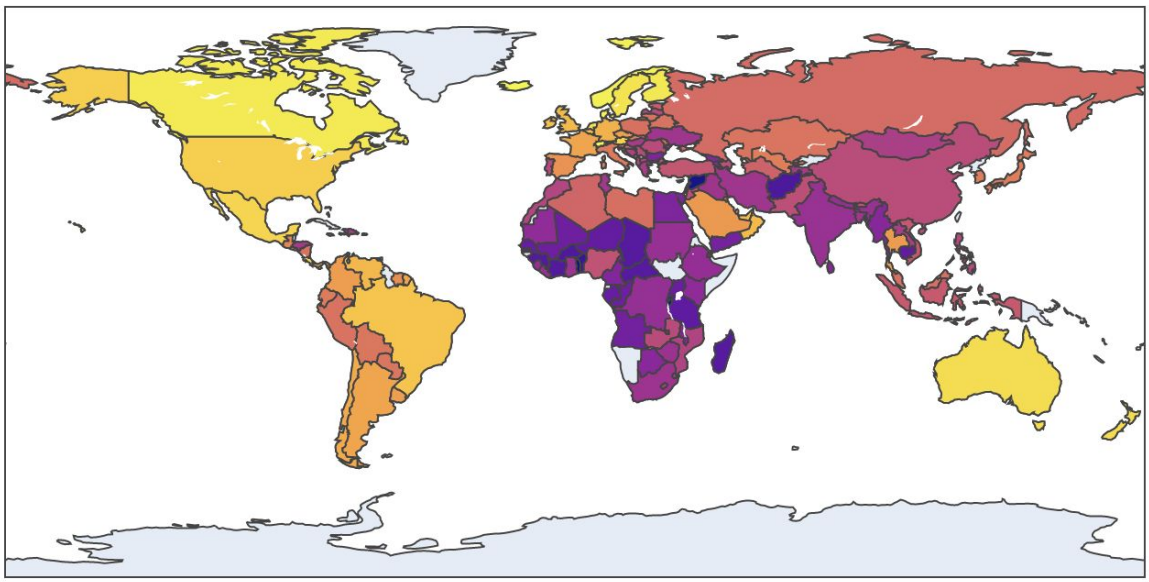

Figure 3.2: Choropleth of happiness value for 2015 data set.

Figure 3.3: Scatter plot comparing GDP per capita to happiness value for 2015 data set.

With the data cleaned and basic visuals generated, the project was then free to enter the design phase.

4. Project Design

Coming up with a design to best display the data set was one of the biggest challenges of the project. The data set was rich enough that the number of possible visualizations was endless. The goal was to come up with three to four powerful visuals that would be easy to use, and give the user insight into the data unique from the others. With this in mind, three components were designed to form the minimal viable product.

Component 1: Summary World Map

The first visual the users would see, the map would summarize the dataset by colouring countries according to their happiness level, as well as other sub-metrics. Users would be able to interact with this visual by changing the year of the data. The map was chosen because it is a vibrant and colourful component that will 'hook' the user into wanting to explore further.

Component 2: Scatter Plot (Happiness vs. Sub-Metrics)

A series of scatter plots would be made to allow users to examine the correlation between a country's happiness, and several sub-metrics, such as GDP, perceived freedom, and level of corruption. The visual would help users better understand the most important factors of a happy country. The scatter was chosen because

examining the factors that contribute towards a happy country was one of the key motivations of the project.

Component 3: Bar Graph (Score and Sub-Metrics of Each Country)

A large bar graph would give users a look into the breakdown of each country's happiness score. The value of each sub-metric will be displayed as a portion of the bar, with the total bar representing the country's overall happiness score (the sum of all sub-metrics). The bar graph was chosen because it neatly displays the overall happiness score and all sub-metrics in an easy to consume manner, and makes the data of each country equally accessible (a good contrast to the map, where it is harder to access the data of geographically tiny countries).

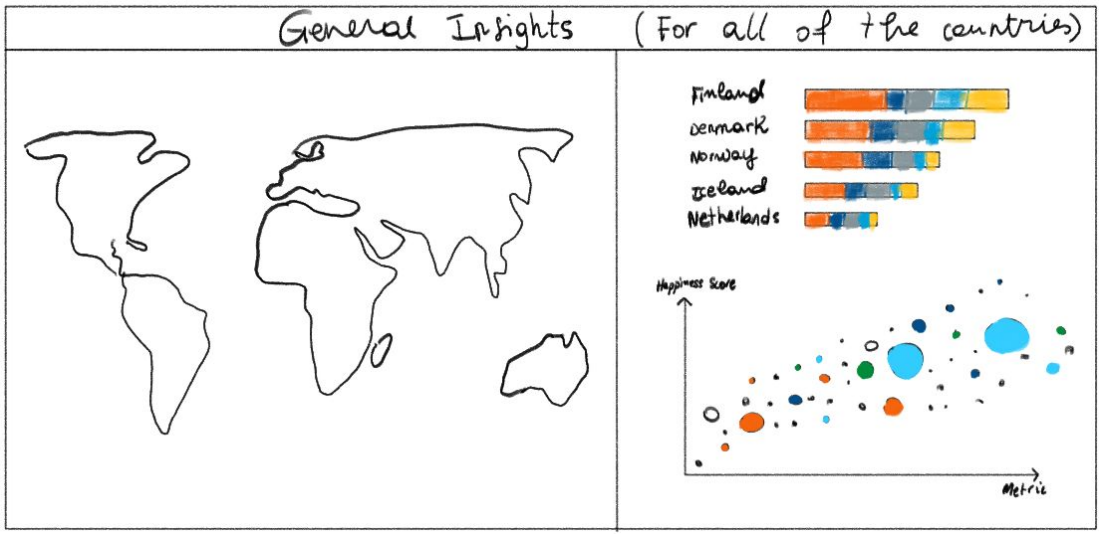

Figure 4.1: Sketch of minimal viable product components.

A few more ambitious components were designed as well, to be added if the timeline permitted it. These included a text-based visualization where the size of each country name would be weighted based on happiness score, as well as a radar chart, which would simultaneously display the value of all a country's sub-metrics and enable direct comparisons between multiple countries.

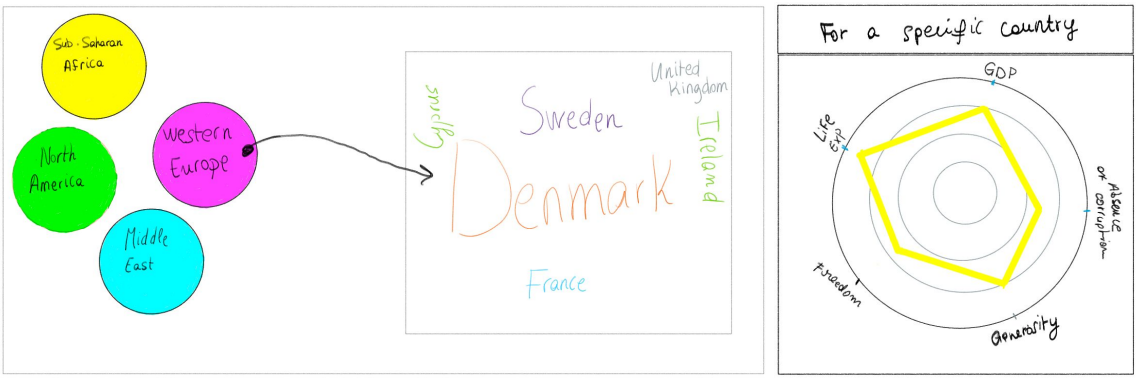

Figure 4.2: Sketch of challenging components, including text-based (left), and radar (right) charts.

After finalizing the designs, the implementation process began.

5. Project implementation

The implementation of the visuals occurred in three distinct stages. Initially, the focus was on setting up the architecture of the project, loading the data into each visual, and displaying it accurately (no matter how ugly it looked).

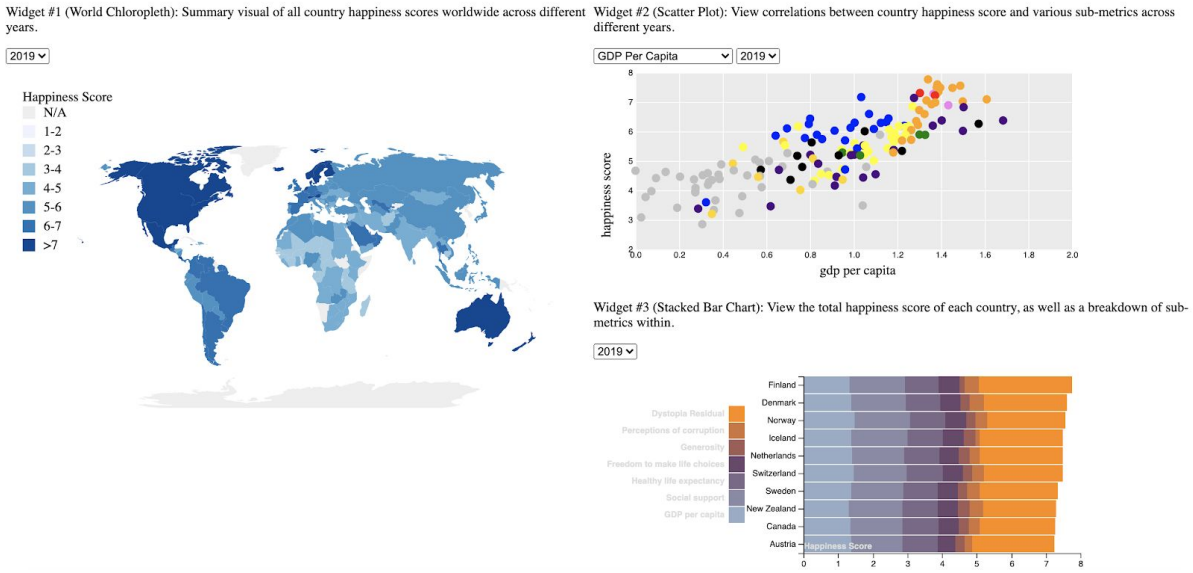

Figure 5.1: Initial skeleton of visualization.

Architecture-wise, the goal was to keep things as simple as possible. Each visual was written in its own JavaScript file, and variables within each file were prefixed to avoid scope conflict. Containing each visual in its own file led to a modular architecture

and simple state management, making it easy to change the behaviour of an individual component without inadvertently breaking a neighbouring visual.

After finishing the initial skeleton of the project, the focus turned to user interactions. It was important that each visual be enjoyable for the user to play and interact with, so heavy priority was given to making the visuals communicate data with one another to fully integrate the user experience. The modular architecture of the project was handy here, since files could expose opaque convenience functions to one another, allowing for a set of seamless interactions.

The final step of the implementation was to focus on design details. This included selecting a uniform colour scheme, building custom icons, and making the entire page responsive to all desktop and tablet screen sizes.

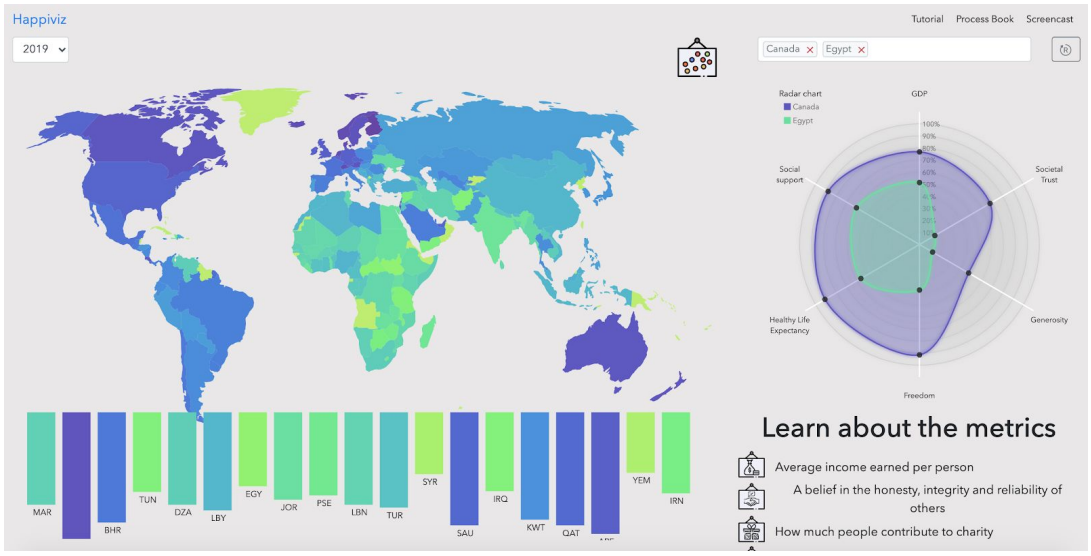

Figure 5.2: Later stages of the visualization, focusing on theming and design.

The final project differed in several ways from the initial design, as an understanding of how to best present the data evolved.

Modification 1: Enlarge scatter plot, and toggle it with world map.

After implementing the radar graph, it became clear that it was a powerful visual, as it allowed for metric-by-metric comparison between several countries. The decision was made for it to take up the entire right panel of the page, pushing the scatter plot to the left, where it would be toggled with the world map. The reasoning for this was twofold - it would allow the radar to be present all the time, and having the scatter in

a larger panel would enable the creation of more interactive features, most notably the ability zoom-in/zoom-out on data.

Modification 2: Change the stacked bar chart from horizontal to inverted-vertical.

After it emerged that the radar would play an important role in the presentation of the data, the bar chart was adapted to become a supporting visual. Switching from a horizontal to a vertical layout fit better with the rest of the page, and inverting the direction of the bars provided a strong visual balance against the map above.

6. Conclusion

Overall, the project was an excellent opportunity to develop a variety of skills (both technical and non-technical) that are extremely relevant in today's marketplace. Highlights included learning Pandas and Plotly (for data preprocessing), gaining design experience, working in JavaScript and d3.js, and further improving teamwork and collaboration skills as part of a group.

7. Peer Assessment

The group worked well together, meeting frequently to brainstorm ideas, divide up tasks, and support one another. Everyone reviewed each other's code, and was always available to assist with technical problems. This was especially satisfying given that the entire project was done remotely, without any in-person interaction.