World Happiness Comparative Visualization

CS573 Final Process Book

Jasmine Duerk and Joseph Yuen

Table of Contents

Document Structure

To best relay our process during our final project, we break up the document into 3 main sections:

1. Project Overview
2. Process
3. Final Product

In the Project Overview, we discuss our proposed goals, motivations, questions, and initial analysis of the data. The Process section contains a series of entries on our work after the initial analysis of the data as described in the Project Overview. These entries consist of our major milestones, findings, pivots, and necessary tasks. Lastly, the Final Product reviews the end project's features, evolution, final conclusions, and team evaluation. We also include our website link to the project in this section.

Project Overview

Goals

At the start of the project, we proposed in our prospectus to:

1. develop a data visualization using d3 to
2. effectively spread awareness and knowledge about the world happiness score to users
3. through a combination of easy-to-understand idioms that allow for data consumption and comparison.

Motivation

Like many students in need of a dataset, we went to Kaggle and found the World Happiness Report data. The team felt that the science of happiness was an interesting topic as there seemed to be a lack of happiness due to the COVID pandemic. We also wanted to accomplish these goals because we observed many World Happiness Report visualizations were rather unapproachable to the uninformed user. We wanted to create an easy to understand visualization that would allow people unfamiliar with the World Happiness Report to learn about the various factors that contribute to happiness, how they differ amongst each other, and how they differ among countries. Lastly, we were inspired by The Refugee Project visualization that displayed the movement of Refugees over decades through a networked choropleth global map. We hoped to represent geographical data at a similar level of quality.

Related Work and Exploratory Data Analysis

As discussed in the previous paragraph, we observed various happiness related research, websites, and visualizations. Below are our more notable findings.

In the official World Happiness Report, we learned about how the final score is calculated and what each contributing variable represents. It also extensively discusses citizens' perceptions of happiness across the globe, as well as how happiness is affected by social, urban, and natural environments.

We interacted with the OECD Better Life Index visualization that measures the quality of life using flower-like glyphs. The team found the glyphs to be slightly difficult to read but appreciated the filtering options and exploratory nature of the program. Although the OECD does not use the same calculation as the World Happiness Report Score, the two frameworks can be represented similarly in expression.

We also discovered *An information visualization application case to understand the World Happiness Report* that analyzed various happiness related works. We learned how certain types of bar graphs achieve greater usability for comparing contributing factors between countries and what users look for when analyzing a rankings list. This paper also gave us insight on how most happiness visualizations consist of maps with supplemental bar graphs. The paper also discussed the concept of clustering countries into groups with similar statistics.

Questions

As we crafted our prospectus, we had questions that inspired our goals, process, and end product.

1. How is happiness calculated?
2. What factors contribute to happiness and how much do they contribute?
3. How do these factors relate to one another?
4. How do these factors differ in other countries?
5. How do these factors differ in groups of countries to other groups?

Data

We found the World Happiness Report data from 2015-2019 on Kaggle. Using Python and Google Sheets, we found that the data had differing layouts and column names from year to year. Even though we did not end up using multiple yearly data in our end product, this data cleaning process forced us to further investigate how the score is calculated and allowed for easier data analysis. For more specific notes on how we cleaned the data, please see our Data Cleaning Entry in the Process section.

Initial Design

Before we turned in our prospectus, the team created sketches and figma boards to visualize our end product. These initial designs centered around an interactive map and was based on the UI layout of The Refugee Project and related readings. Due to the geographic nature of our dataset, we planned on implementing an interactive global map that allows users to discover countries' happiness scores by selecting countries that are colored differently depending on their score (see Figure 1-1). In addition, we wanted to implement a tooltip on hover over a country to provide a quick preview of the country's data (see Figure 1-2). To allow users to examine the factors that contribute to happiness, we added filters that would isolate certain factors and determine how countries rank on that specific factor (see Figure 1-3). We also drew a ranking list with representative bar graphs for the given filter to allow for a non-geographical comparison of countries (see Figure 1-4). Our dataset contains records from 2015 to 2019, so we wanted to make an interactive bar graph timeline to view a country's score over time (see Figure 1-5). The ideas was that if a user selects a year that is not currently selected, the application will update the map and rankings for that selected year similarly to a filter (see Figure 1-6).

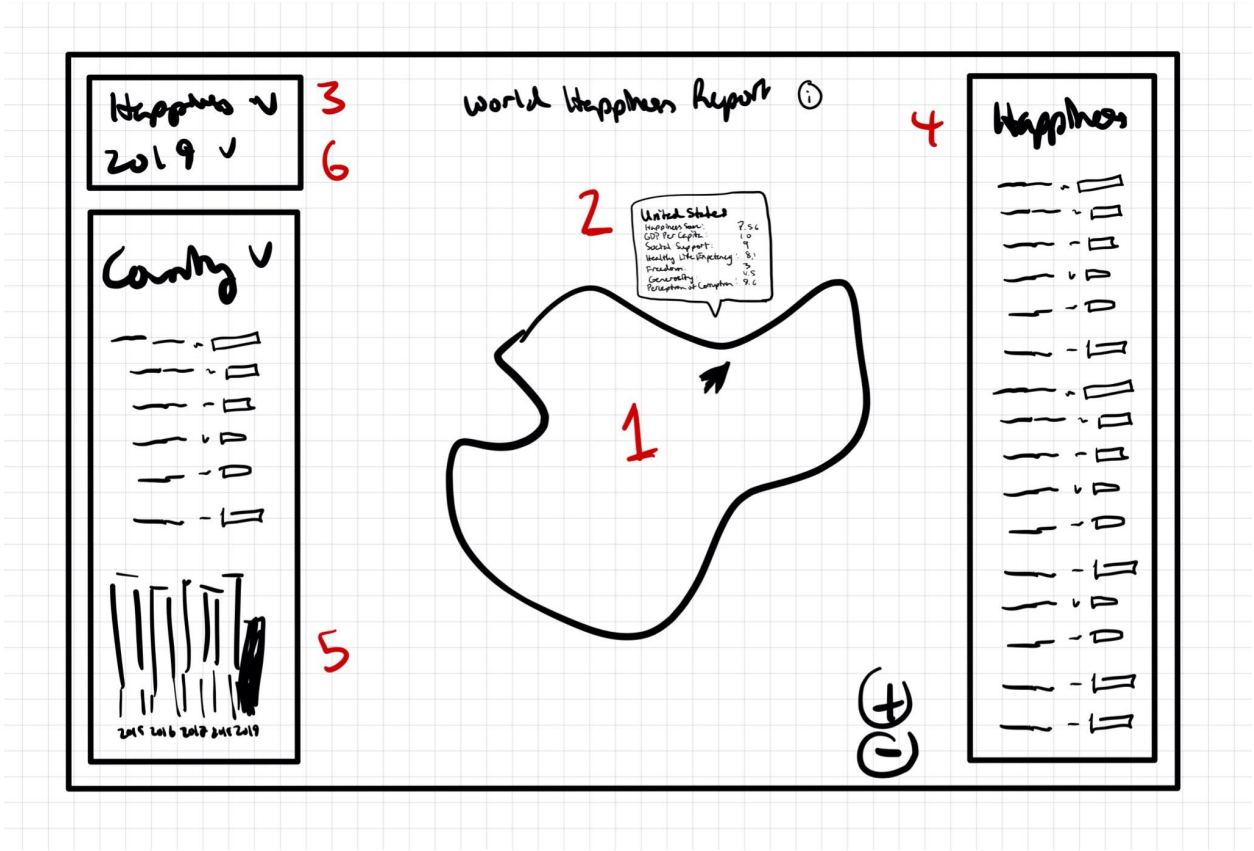
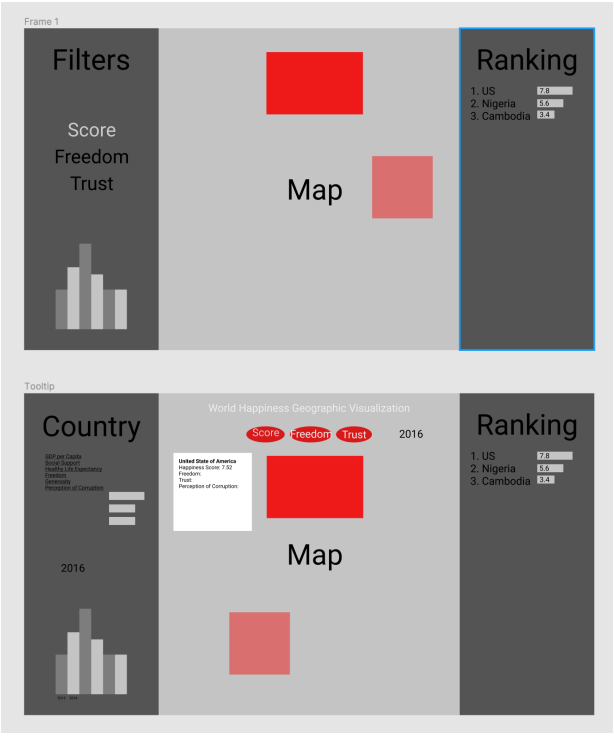


Figure 1. Initial Sketch



Figma Experimentation

Process

Professor Feedback - 4/25

On April 25th, we received feedback from Professor Harrison on our initial prospectus. Here are our notes from the discussion that followed:

- We need to go beyond the map to make an interesting visualization
    - Maps for the world happiness index have already been done many times so we should try to create something more novel
    - Perhaps look at flower glyphs
        - Example: http://www.oecdbetterlifeindex.org/#/11114231311
    - Looking for correlations between different factors could have some interesting results
        - Correlation heat map?
    - There could also be interesting data that contrasts between the different years

Data Cleaning - 4/27

We cleaned our data to make it easy to analyze for interesting correlations or other items of note. We deleted the dystopia residual factor from the earlier years, as it was removed in 2019 (or hidden as a factor even if it still affected the happiness score). The dystopia residual is the unexplained happiness or unhappiness of each country, added to the happiness rating of the hypothetical unhappiest country. While it is useful when calculating the score, it is not extremely relevant on its own. Additionally, we renamed each year's groups so that they were the same as the 2019 dataset.

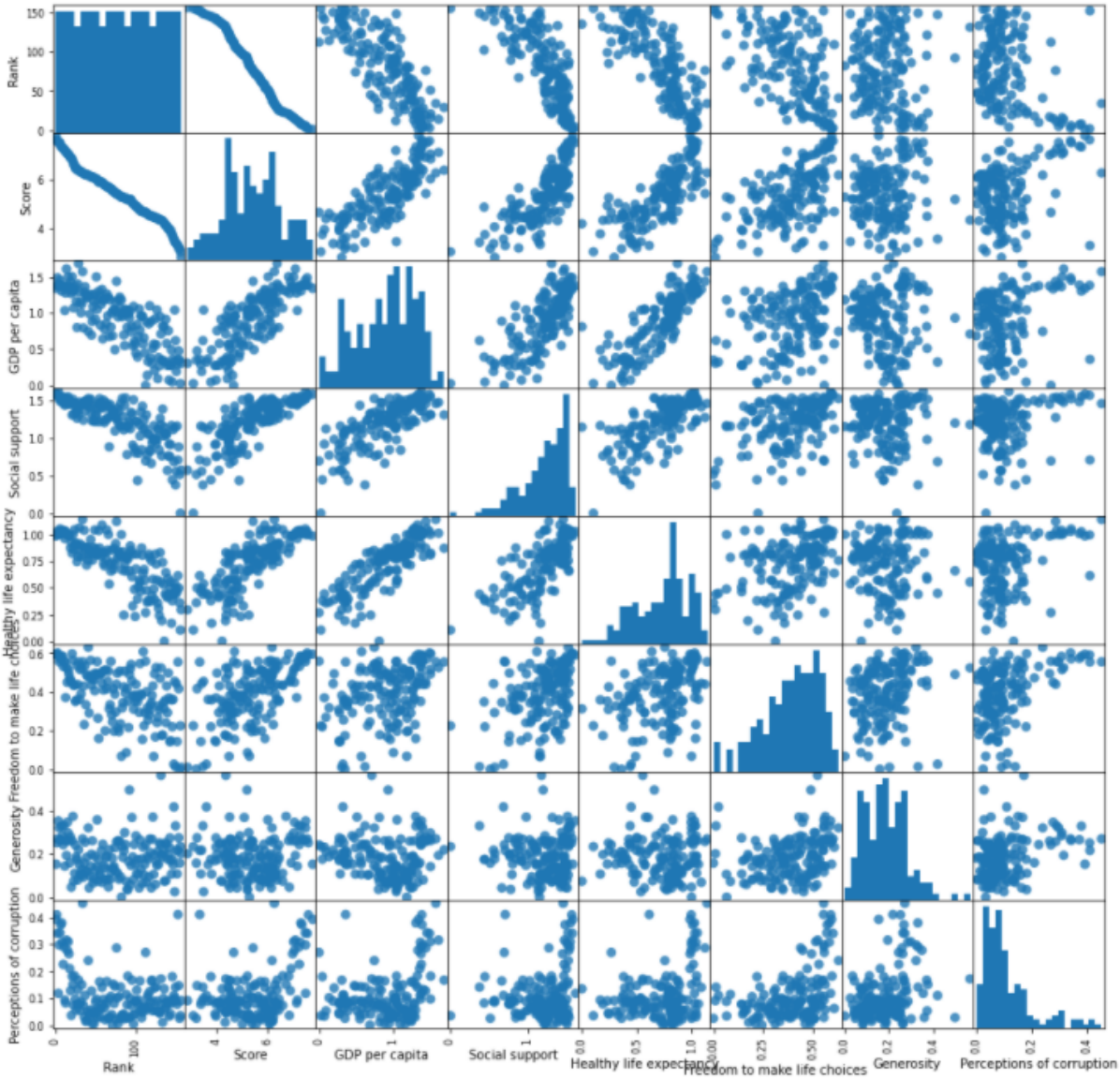
With our cleaned data, we plan to investigate correlations between the different factors that contribute to overall happiness score. We are still trying to determine what kind of visualization to make, but it could be interesting to show the correlations if we find any.

Set Up Git Repo - 4/27

We set up a repository on Github so that we can collaborate on code virtually. In this way, we can send code to each other at any time, and modify each other's work.

Correlation Low Fidelity Prototype - 4/31

In Python, we investigated for correlations between the factors on the 2019 dataset. We determined that there were indeed correlations between three of the factors: GDP per Capita, Healthy Life Expectancy, and Social Support.

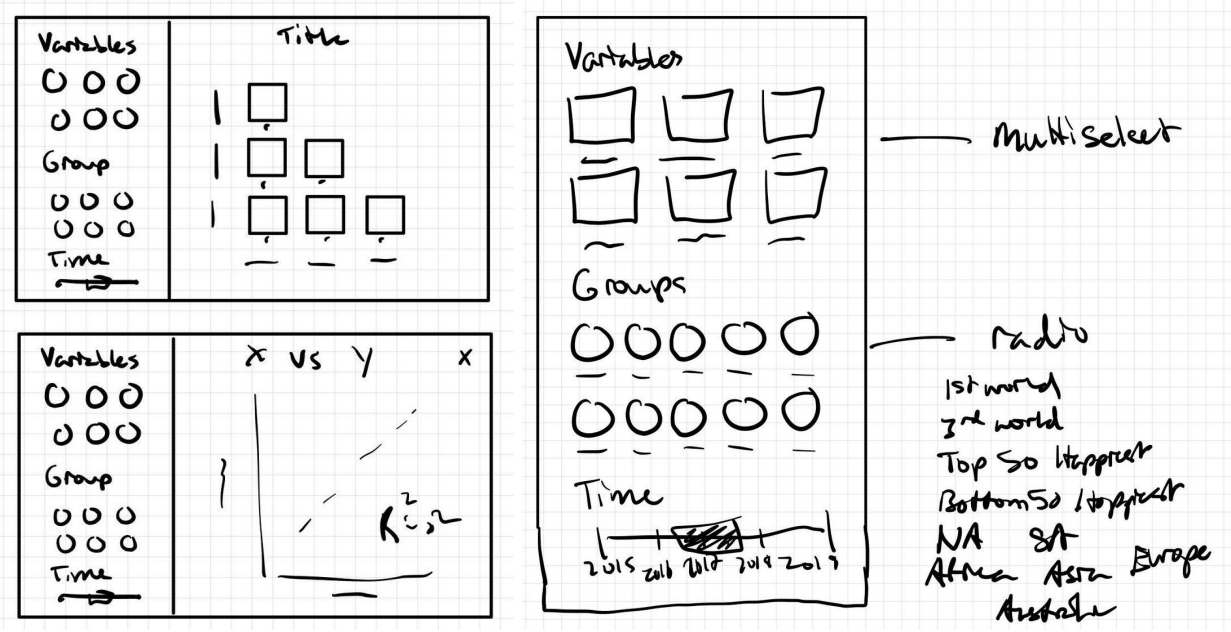

Rank
Score
GDP per capita
Social support
Healthy life expectancy
Freedom to make life choices
Generosity
Perceptions of corruption

```
       GDP per capita  Social support  Healthy life expectancy  \
count      156.000000      156.000000               156.000000
mean         0.905147        1.208814                 0.725244
std          0.398389        0.299191                 0.242124
min          0.000000        0.000000                 0.000000
25%          0.602750        1.055750                 0.547750
50%          0.960000        1.271500                 0.789000
75%          1.232500        1.452500                 0.881750
max          1.684000        1.624000                 1.141000

       Freedom to make life choices  Generosity  Perceptions of corruption
count                    156.000000  156.000000                 156.000000
mean                       0.392571    0.184846                   0.110603
std                        0.143289    0.095254                   0.094538
min                        0.000000    0.000000                   0.000000
25%                        0.308000    0.108750                   0.047000
50%                        0.417000    0.177500                   0.085500
75%                        0.507250    0.248250                   0.141250
max                        0.631000    0.566000                   0.453000

Correlation Coefficients:
[[ 1.          0.75490573  0.83546212  0.37907907 -0.07966231  0.29891985]
 [ 0.75490573  1.          0.71900946  0.44733316 -0.04812645  0.18189946]
 [ 0.83546212  0.71900946  1.          0.39039478 -0.02951086  0.29528281]
 [ 0.37907907  0.44733316  0.39039478  1.          0.26974181  0.43884331]
 [-0.07966231 -0.04812645 -0.02951086  0.26974181  1.          0.32653754]
 [ 0.29891985  0.18189946  0.29528281  0.43884331  0.32653754  1.        ]]
```

Based on this preliminary data analysis, we created a basic layout for a potential correlation visualization.



In this prototype, the variables on the left are a multi-select toggle, allowing the user to select which ones to analyze. The group can only be selected one at a time and only calculates the correlation based on the countries in the group (for example, the countries in Europe). There will also be a timeline slider so

that different years can be chosen. The base visualization would show all the correlations between every variable (as seen in the Python graphs above). The user could then click on individual graphs and it will be enlarged with more information.

Correlation Investigation - 5/1

After creating the low fidelity prototype, we conducted a deeper investigation to see if there were any other interesting correlations. The countries were sorted into groups based on continent, into developed, transition, and developing, and into top and bottom 50 countries based on happiness score. Then, we tested the correlation between the variables in each group to see if there was any variation or any new correlations that were not there before. However, we observed no significant or noteworthy correlations besides the original three to warrant a visualization. Because of this, we are considering different ways to compare aspects of the dataset.

Professor Feedback & Comparative Investigation - 5/4

Based on previous conversations and professor instruction, we pivoted to look for a new and interesting direction for our visualization. We thought of doing a side-by-side group comparison tool that would allow the user to compare the happiness scores and respective variables of individual countries as well as the pre-defined groups we created previously.

When looking at the comparisons between the various groups on the 2019 data, there were big differences between the averages as shown in the table below.

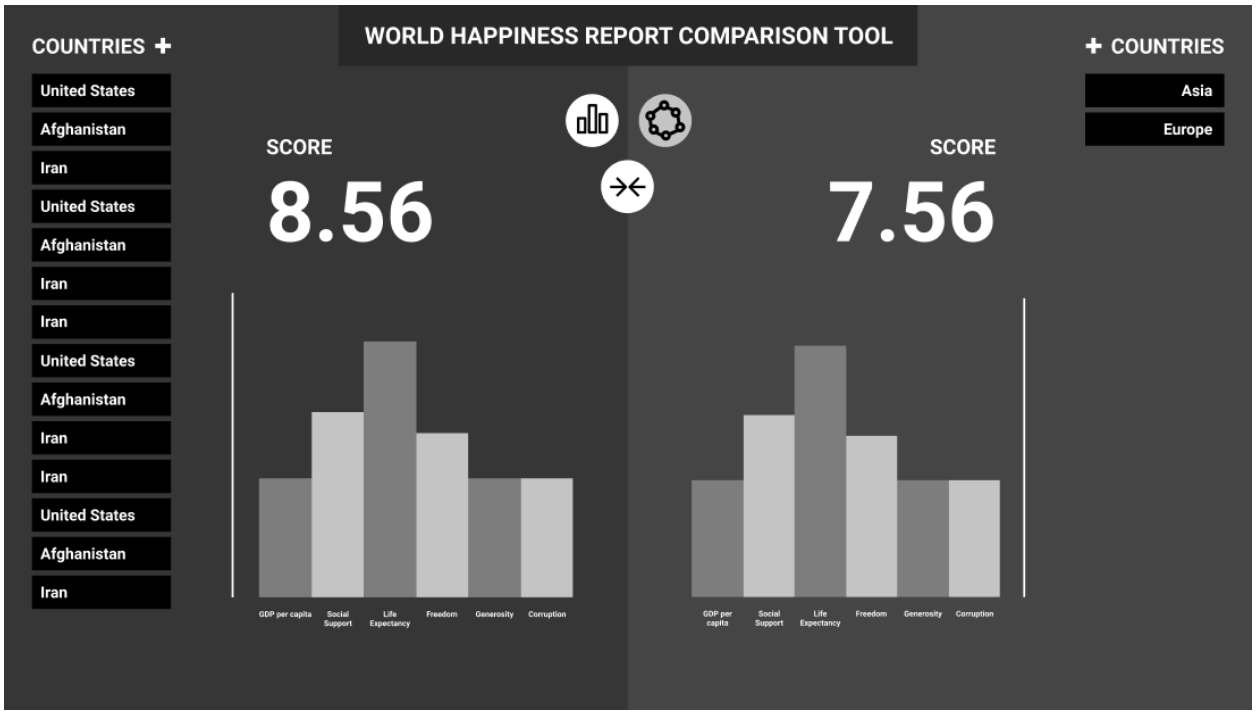
	Score	GDP per Capita	Social Support	Life Expectancy	Freedom	Generosity	Corruption
All countries	5.41	0.91	1.21	0.73	0.39	0.18	0.11
Top 50	6.67	1.25	1.45	0.93	0.49	0.20	0.16
Bottom 50	4.15	0.51	0.93	0.50	0.31	0.19	0.10
North America	6.15	0.95	1.33	0.82	0.46	0.19	0.10
South America	5.92	0.98	1.38	0.83	0.43	0.12	0.07
Europe	6.26	1.22	1.42	0.93	0.41	0.17	0.13
Asia	5.24	0.97	1.20	0.76	0.41	0.21	0.11
Africa	4.37	0.50	0.93	0.44	0.33	0.18	0.09
Oceania	7.27	1.34	1.55	1.03	0.57	0.33	0.34

Developed	6.62	1.31	1.47	0.97	0.46	0.19	0.17
Transition	5.27	0.91	1.28	0.78	0.34	0.17	0.08
Developing	5.01	0.76	1.11	0.63	0.38	0.19	0.10

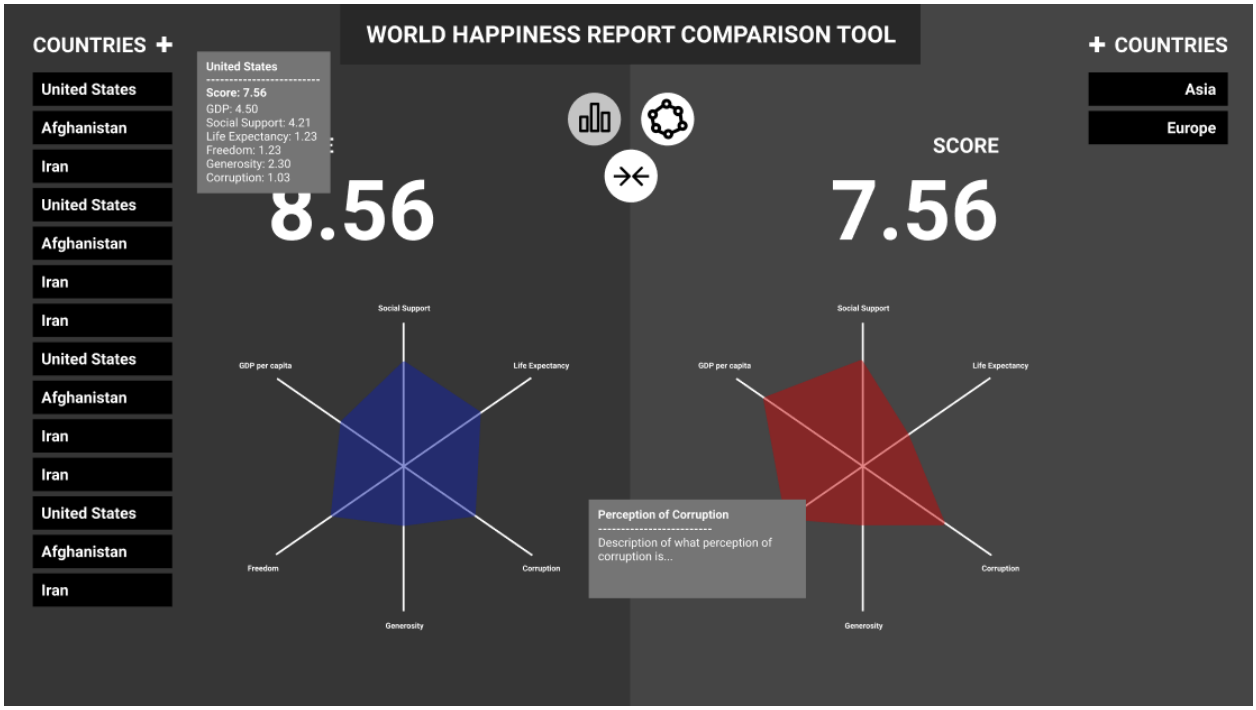
With this information, we can proceed with creating a prototype with a new focus on comparison.

High Fidelity Prototype - 5/6

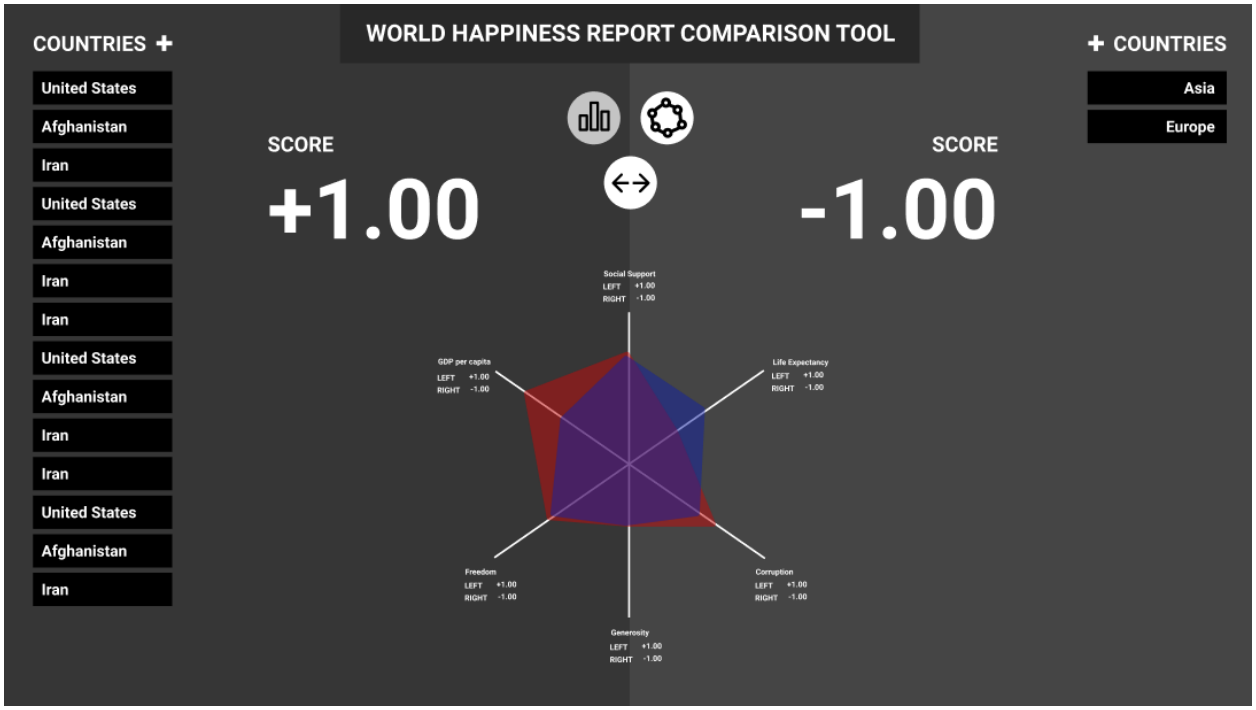
We developed a high fidelity prototype in Figma so that we had a better idea of the visualization we were creating. In this iteration, we have completely strayed from our original plan to do a geospatial visualization, and instead we wanted to display the distributions through a bar graph and radar chart. With these methods, there are two ways to view the same information. The mockup can be seen in the images below.



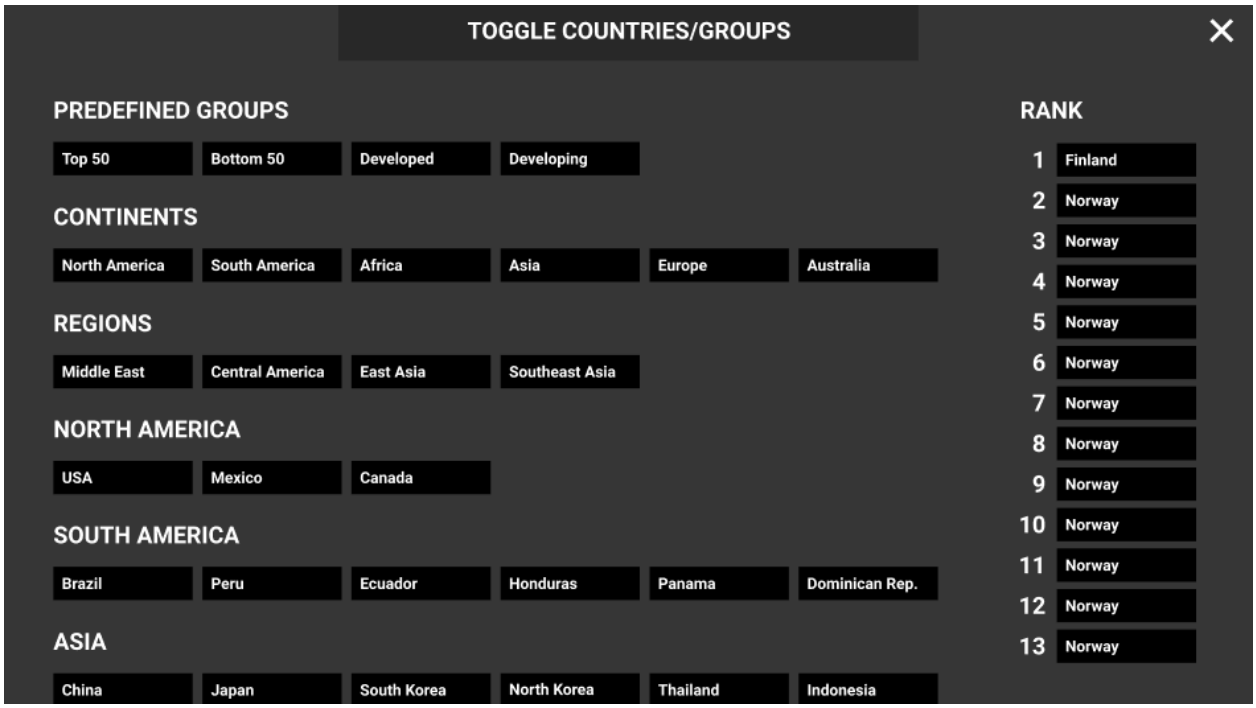
As seen above, there will be two bar charts for different selections of countries that can be compared side by side. The countries (or groups) chosen for each group are displayed in the side bars on the left and right of the charts. The icons in the center allow the user to switch between the bar and radar chart views.



The above image displays the default radar chart view. Additionally, this image shows the possible tooltips we plan to implement when the user hovers over different elements of the visualization. On the top left, the light gray box shows the tooltip when a user hovers over the United States on the country list. It displays the actual values for each factor as well as the country's happiness score. The light gray box in the center shows the tooltip that appears when a user hovers over one of the factors on the chart itself. It explains what the factor actually means so that the user can have a greater understanding of the value and how it is involved in happiness.



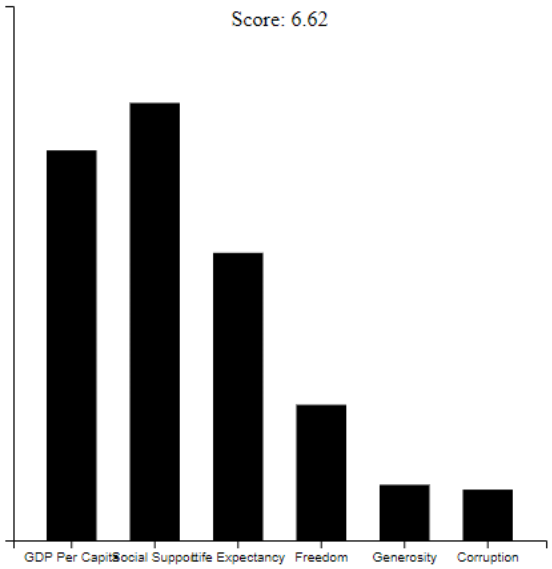
The above view shows the combined radar chart view. The two groups being compared are shown overlapped on the same chart so that it is easier to compare the values with each other.

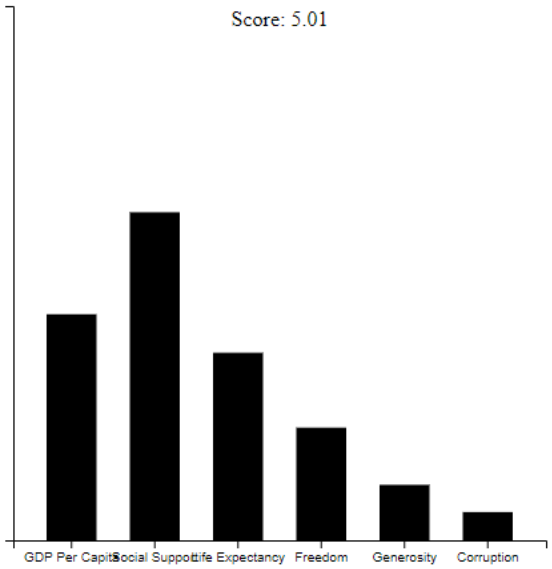


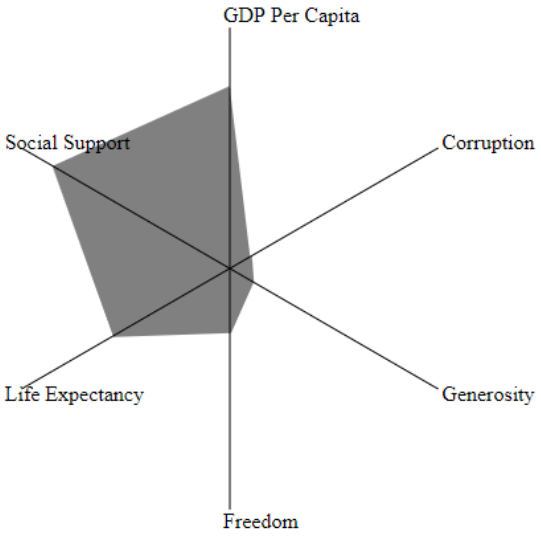
The image above shows the country selection interface. The user can choose from predefined groups or manually select any number of countries. Here, the ranking is also displayed on the right.

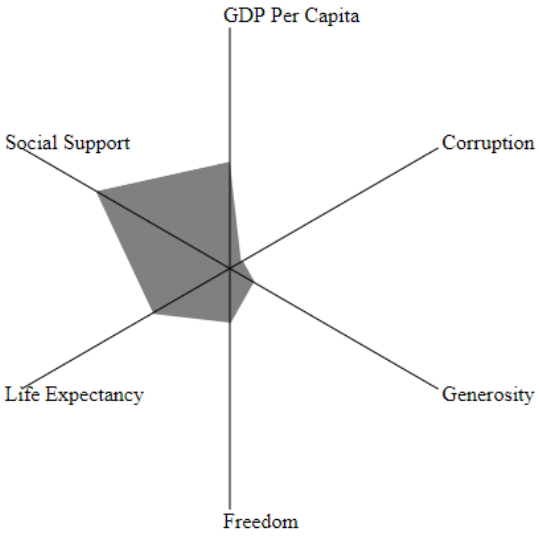
Visualization Prototypes - 5/7

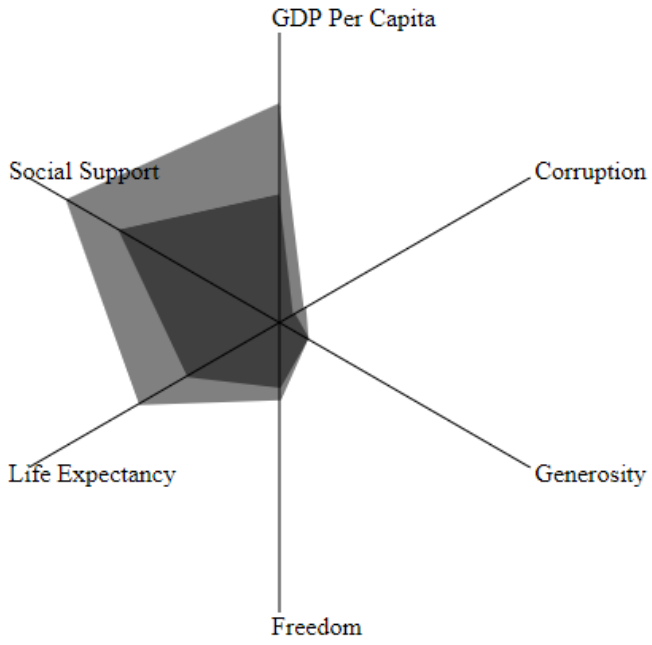
Based on our mockup, we created basic visualizations in d3 to represent the information we wanted to convey. First, we made a function that took in an array of country names and created data structures from the original csv that included the chosen countries, formatted properly for d3. Then, functions were created for each type of chart that took in the datasets generated with the chosen countries, and displayed a graph based on the averages for the score and each factor. These charts can be seen below.











Interface Construction - 5/7

We laid out the home screen and selection screens in css and html using flex boxes. We also used jQuery to do the screen switching animation. The team created custom button icons in Sketch as the icon library did not have a good icon for radar charts.

We used the defined groups of countries to automatically load in all of the toggle buttons at runtime. Toggle buttons were activated using an active class that is given and taken after each press of the button.

By the end of the work session, each screen was laid out although everything was oddly colored for debugging purposes.

Application Integration - 5/9

Integrated visualization prototypes into the rest of the application. We colored the whole application in the gray and orange color scheme, and we recolored the d3 visualizations to white, red, and blue. The team also added reload functions to refresh the visualizations upon entering the home screen or pressing a different vis button.

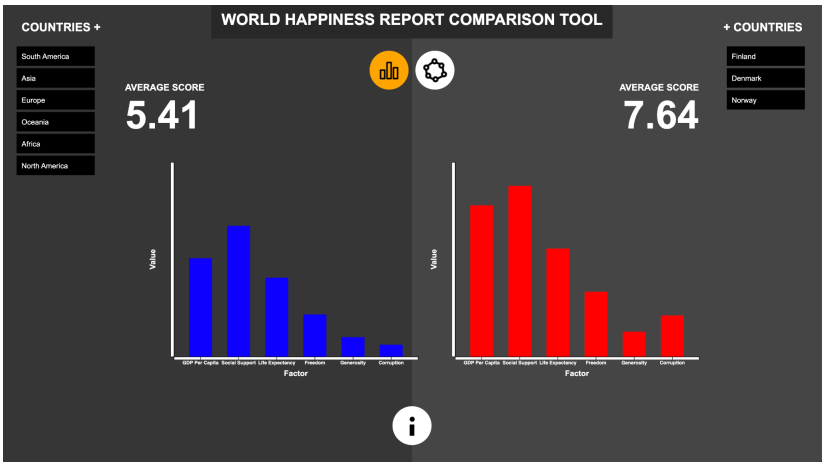
Publish Visualization Application - 5/10

The team checked the vis for bugs, polished the vis interface, recorded the presentation video, finished the process book, and published the website to github.

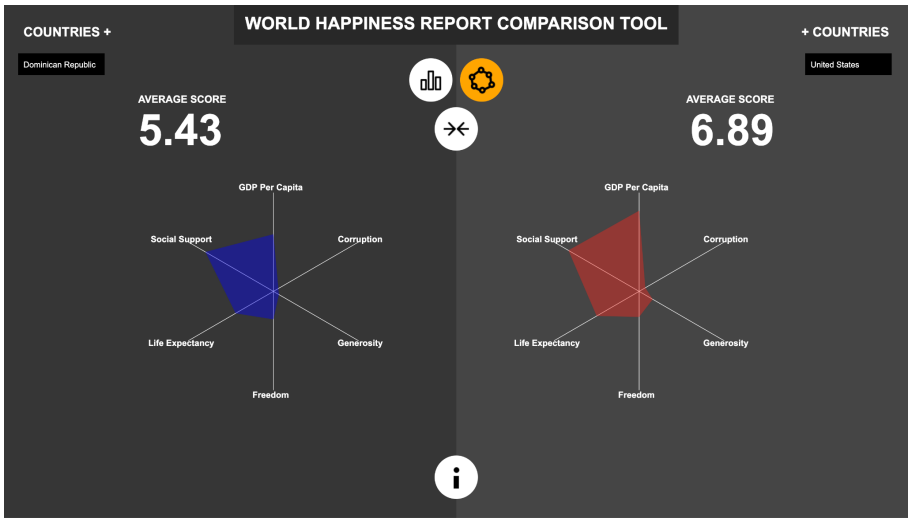
Final Product

Features

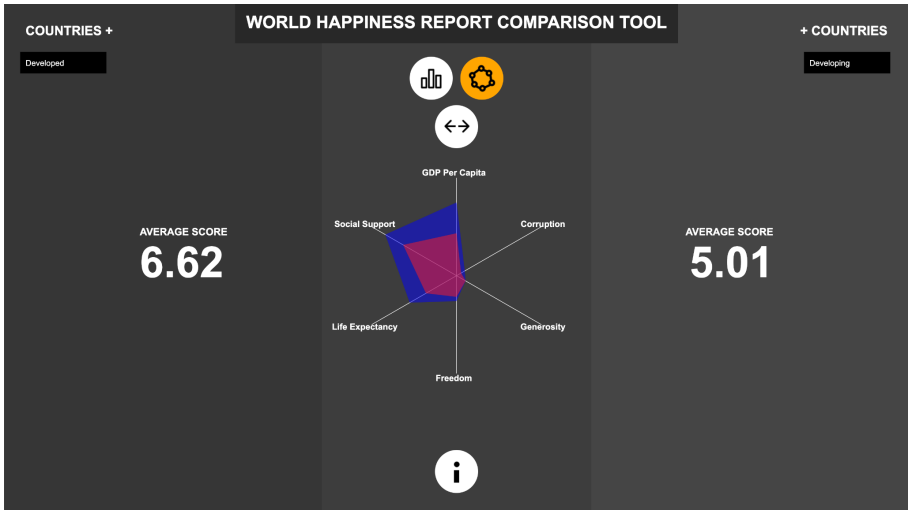
Using this Comparison Tool, users can compare individual countries' breakdown of their happiness scores using the bar graph, radar chart, and overlapping radar chart idioms. Users can also compare predefined and custom groups of countries to find interesting patterns between their averages. If the user wants to further investigate, the user can hover over an individual country and see the raw data behind each. They can also learn more about the factors by hovering over the factor names or by pressing the info button in the bottom center to learn more.



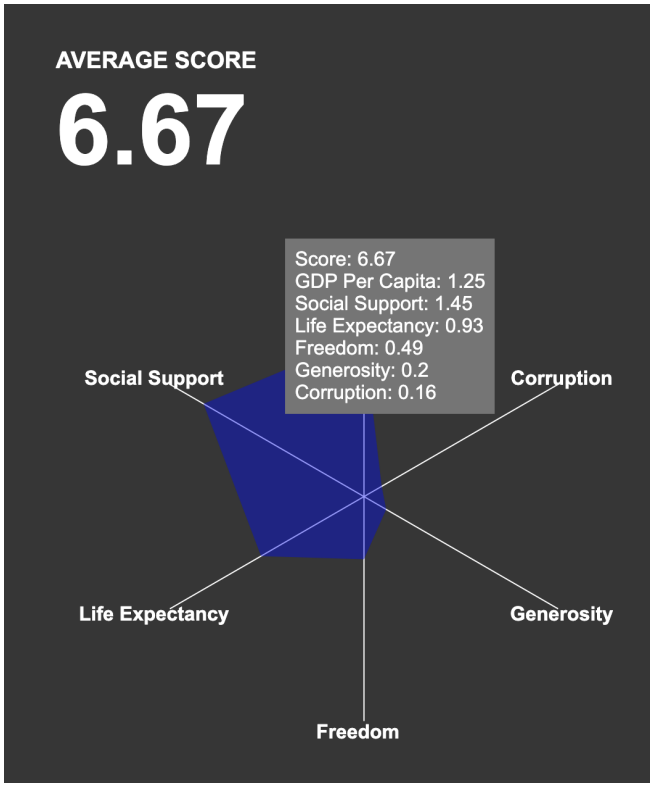
Bar Graph Comparison View



Radar Chart Comparison View

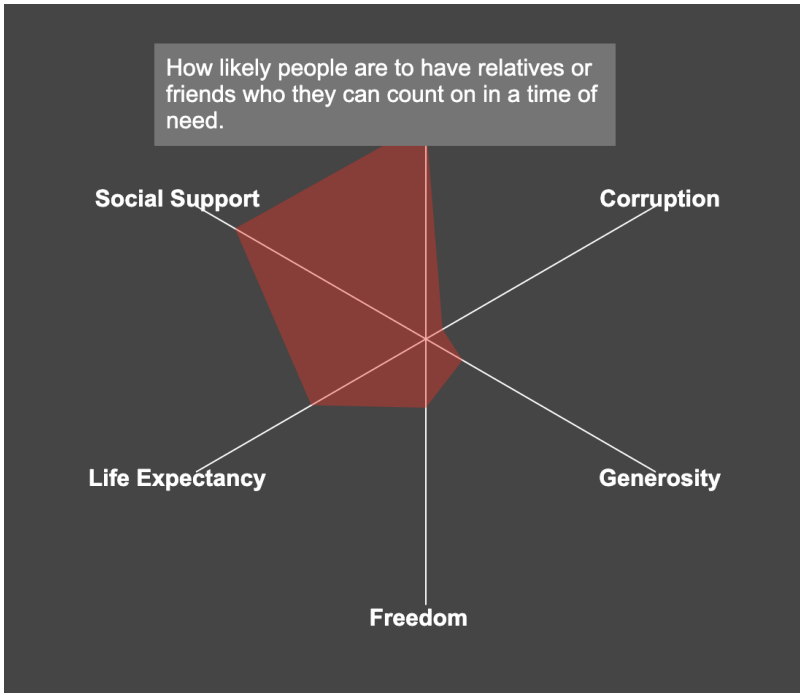


Overlapping Radar Chart Comparison View

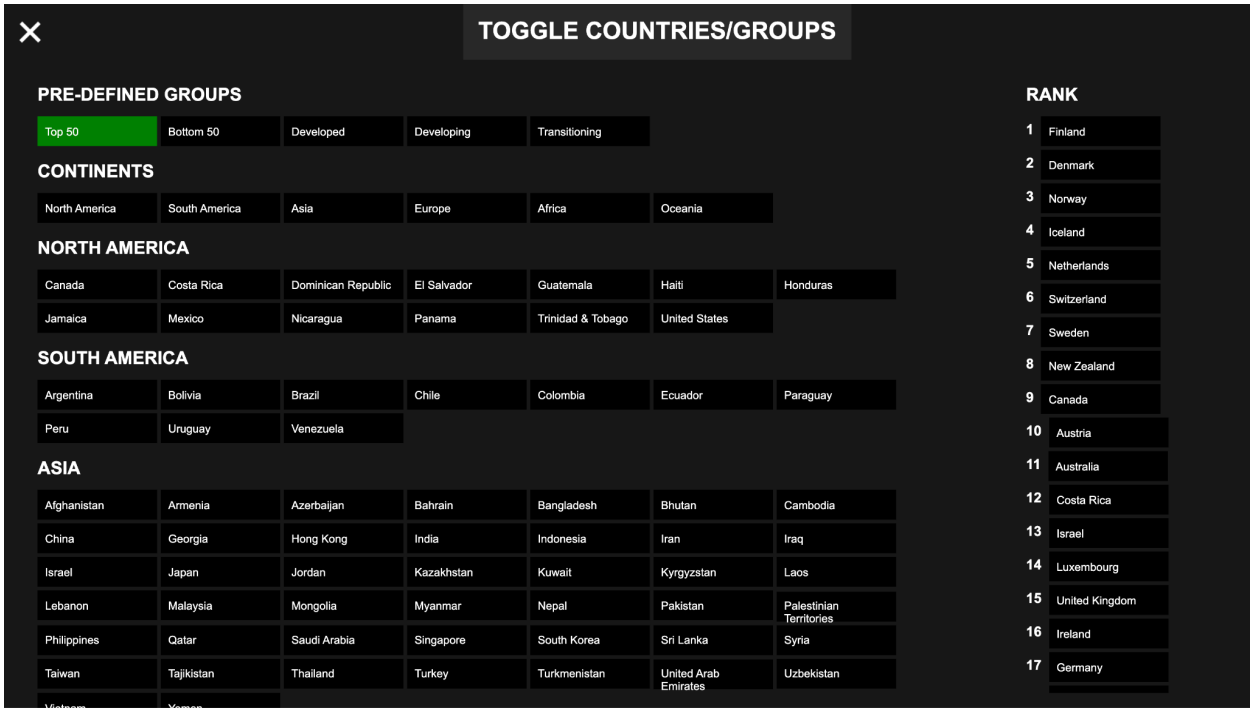


Radar Chart Tooltip

Group Data Tooltip



Factor Description Tooltip



Toggle Country Selection Screen

Home Screen

Reflection on Design Evolution

We initially planned to make a visualization centered around a map as seen in the Project Overview section. This made sense as the data is centered around countries and spatial data, however, we wanted to interact with the data in a new way. As a result, we analyzed correlations between variables and constructed a design based on that as seen in the Process section. Correlations were interesting as we wanted to investigate the relationships between variables and how they differed. To visualize the correlations, we originally thought about using heat maps as they highlighted which correlation was notable, but we also wanted a way to visualize the relationship. We considered multiple scatter plots in a grid fashion with highlighting elements from heat maps to quickly see which correlations were significant and how the data actually looked. In that design, we also wanted to add an exploration feature where a user can select the listed scatter plot and investigate each plot point. Even though the team was very excited about this design, we did not find many notable or differing correlations even across varying groups of countries. Thus, we did not go with this design. Despite this, we realized we had formed groups of countries based on their ranking or geographic region. Even though the groups didn't have any notable correlations, the group averages produced interesting points of comparison. This realization led the team to focus on averaging groups of countries and comparison. Since we discovered that barcharts are one of the most effective idioms for comparison, we included the visualization to compare relative height between groups. Even though we didn't like the OECD flower glyphs as they were slightly difficult to interpret, we knew we could use shape and color for comparison. As a result, we included an overlapping and non overlapping radar charts so that users could recognize the similar shape that all groups have. While we deviated from our proposal's initial visualization, we were still able to accomplish our goal of comparing the contributing factors between each other, different countries, and different groups of countries.

Evaluation - Data Analysis and Conclusion

Using this tool, we were able to make many interesting discoveries about the data. More generally, countries have an average happiness score of 5.41, with the highest being Finland with a score of 7.77, and the lowest being South Sudan with a score of 2.85. The happiest continent is Oceania, with an average happiness score of 7.27. The unhappiest continent is Africa, with an average happiness score of 4.37. More notably, we learned that developed countries had an average happiness score of 0.55 higher than developing countries. The biggest cause for the disparity is GDP per Capita, which has a difference of 0.36. Generosity on the other hand, is on average the same (0.19) for developed and developing countries. Additionally, the raw value for social support is the highest out of all factors, meaning that is the largest contributor to happiness score. The visualization successfully answered our original questions by providing an in-depth look at how each individual factor affects happiness, and how they differ between countries and groups.

Our visualization works well, meeting our original design goals and providing valuable insight about the factors involved in world happiness. That being said, there are areas where we could further improve it. Currently, selecting a predefined group (such as Asia) does not also highlight all the Asian countries. That means that the user can select a group as well as individual countries that are a part of the group, which is not ideal. Selecting a country from the Rank list also does not select the same country from the

general list, meaning a country can be selected twice. In the future, this should be prevented by detecting what countries are selected (either from the Rank list or from a group) and selecting them automatically in every other place that they appear. Additionally, the overlapping radar chart does not explicitly say which color corresponds with which side, so it may be unintuitive to the viewer. This could be improved by adding labels or with more apparent color-coding for each side.

Team Evaluation

Jasmine

Jasmine worked primarily with the dataset, first by cleaning the data with Joe. She then used Python and Google Sheets for preliminary data analysis that allowed us to determine what visualization would be interesting to create. She also created functions that parsed the csv data file and generated new datasets with average calculations based on the selected countries. Finally, she developed the bar and radar charts in d3, that displayed the information from the dataset based on the selected countries.

Joseph

Joseph contributed by cleaning the original dataset with Jasmine by renaming column names and rearranging them for easier data analysis. He designed the lo-fi and hi-fi concept designs in Figma. Lastly, he developed the selection interface, performed css and d3 styling, and integrated the d3 charts into the home screen.