

CS4802: Final Project

Process Book

Alexander Bell, Brian Earl, Haley Hauptfeld, Pooja Patel

Table Of Contents:

Section 1: Overview and Motivation	2
Section 1.1: Project Overview	2
Section 1.2: Motivation	2
Section 2: Related Work	2
Section 2.1: OECD World Happiness Data Set	2
Section 2.2: Alexander Bastidas's World Happiness Visualization	2
Section 2.3: Weighted Scatter Plot of World Happiness Report	3
Section 3: Questions	3
Section 3.1: What questions are we trying to answer?	3
Section 3.2: How did these questions evolve over the course of the project?	3
Section 3.3: What new questions did you consider in the course of your analysis?	3
Section 4: Data	4
Section 4.1: Data Source	4
Section 4.2: Scraping Method	4
Section 4.3: Data Cleanup	4
Section 5: Exploratory Data Analysis	4
Section 5.1: What visualizations did we use to initially look at our data?	4
Section 5.2: What insights did we gain from those visualizations?	5
Section 5.3: How did these insights inform our design?	6
Section 6: Design Evolution	6
Section 6.1: What are the different visualizations we considered? Justify the design decisions you made using the perceptual and design principles you learned in the course.	6
Section 6.2: Did we deviate from our proposal?	9

Section 7: Implementation	9
Section 7.1: Intent of Visualization Implementation	9
Section 7.2: Functionality of Visualization Implementation	10
Section 8: Evaluation	12
Section 8.1: What did we learn about the data from our visualizations?	12
Section 8.2: How did we answer our questions?	12
Section 8.3: How well does our visualization work?	13
Section 8.4: Future Improvements	13
Section 9: References	13

Section 1: Overview and Motivation

Section 1.1: Project Overview

Our project is focused on the different measurements of happiness from countries across the world. We focused on measuring world happiness reports from different countries based upon different factors including economy (GDP per Capita), Family, Health (Life Expectancy), Freedom, Trust (Government Corruption), and Generosity.

Section 1.2: Motivation

We were motivated to focus on world happiness because when our group was deciding on a topic of data to visualize, we agreed that world happiness was something that interested all team members. It's interesting to see what countries are happier compared to others, especially when deciding which factors to consider in the data that we decided to keep in our visualization. At an initial observation, happiness seems like something that is difficult to measure using numbers. However, the data that we displayed are all important factors when considering what makes people happy.

Section 2: Related Work

Section 2.1: OECD World Happiness Data Set

One related work that we viewed was [the data](#) regarding a better life index from the *Organization for Economic Co-Operation and Development*. This data was interesting because it provided several different factors regarding what makes a better life for people. This relates to our work because a better life correlates to how happy people are in different areas of the world. This source offered data across ~30 different countries. Yet, it covered topics that we didn't consider before visualizing our data for world happiness. This is important because it helps us realize that there are an immense amount of factors that decide what makes people "happy".

Section 2.2: Alexander Bastidas's World Happiness Visualization

Another related work that we viewed is [a visualization](#) of the world happiness data created by Alexander Bastidas. This visualization depicts data in a linked view between a global 3D choropleth map, and a scatter plot with a line of best fit displaying the correlation between the economic production and the happiness of a specific country. When a country is hovered over on the choropleth map, the scatterplot visualizes the data for that specific country. This is related to

our project because it gave us an example of a choropleth map that visualizes data based on happiness. It also gave us the idea to make a linked view between a choropleth map and data for a specific country in some format based upon which country is hovered over on the choropleth map.

Section 2.3: Weighted Scatter Plot of World Happiness Report

A third related work we looked at is [a weighted scatter plot of the world happiness report](#). This visualization shows the correlation between the happiness index and five different factors related to happiness including GDP Per Capita, Social Support, Freedom, Generosity, and Life Expectancy. The user can toggle between viewing the correlation of happiness to each of these five factors. This work is related to our project because it showed which topics in life are important to consider when measuring happiness.

Section 3: Questions

Section 3.1: What questions are we trying to answer?

The questions we're trying to answer are as follows:

- What countries are the happiest countries in the world?
- Why are some countries happier than others?
- What factors are most important when deciding if a country is happy?
- What data best represents happiness?

Section 3.2: How did these questions evolve over the course of the project?

These questions evolved by having us process the data. Through processing the data, we started to question the following:

- What dataset should we use between the dataset we found on Kaggle versus the dataset from the OECD?
- Which columns from our selected dataset are most important when deciding what happiness is best measured by?
- What techniques should we use when visualizing our data to display comparisons of happiness across different countries?

Section 3.3: What new questions did you consider in the course of your analysis?

Some new questions we considered in the course of our analysis are :

- How are scores calculated for different factors, such as ‘Family’ or ‘Health’?
- What does the happiness factor, ‘Perceptions of Corruption’ mean? How was it calculated?

Section 4: Data

Section 4.1: Data Source

The data was sourced from Kaggle.com. The original data can be found at <https://www.kaggle.com/unsdsn/world-happiness>. We also looked into other sources on world happiness data, however, we decided that this data set had the most clear data and had the benefit of multiple years worth of data at once.

Section 4.2: Scraping Method

We gathered the data by downloading each year’s CSV file from Kaggle.

Section 4.3: Data Cleanup

We cleaned up the data by removing some columns from the data, since there were discontinuities between years. For example, the year 2015 had data on “Dystopia Residual”, which was not present in 2018’s data. We also changed the order in which columns appear for some years in order to be consistent. Lastly, in case it was needed, we also created one larger CSV from each other year combined, in the case that it would be easier later.

Section 5: Exploratory Data Analysis

Section 5.1: What visualizations did we use to initially look at our data?

In order to get a sense of the data we first looked at the data in a tabular form. We were able to identify the values in the data according to the column labels. To better understand the data, research was done to understand the column headings and how they were calculated. A sample of the data can be seen below.

Table 1: Sample of happiness data for 2015

Country	Year	Happiness Rank	Happiness Score	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	Trust (Government Corruption)	Generosity
Switzerland	2015	1	7.587	1.39651	1.34951	0.94143	0.66557	0.41978	0.29678

Iceland	2015	2	7.561	1.30232	1.40223	0.94784	0.62877	0.14145	0.4363
Denmark	2015	3	7.527	1.32548	1.36058	0.87464	0.64938	0.48357	0.34139
...

One visualization we did to explore the data was a bar chart of the Happiness Score for each country. The Happiness Score is the sum of the other column values (Generosity, Trust, Freedom, Health, Family, and Economy). We created the chart to get a sense of where each country stands relative to another. The chart can be seen below.

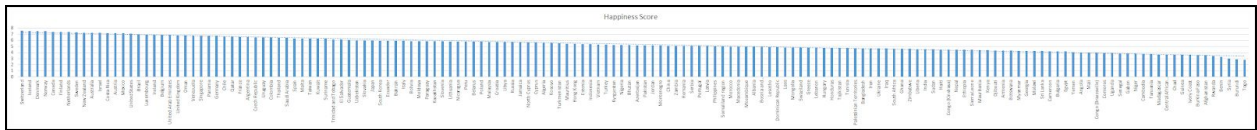


Figure 1: Bar chart of each countries happiness score (2015)

Another visualization we did was a stacked line graph to see the rankings of the different countries against one another and to see how each component of happiness aligned. The chart can be seen below.

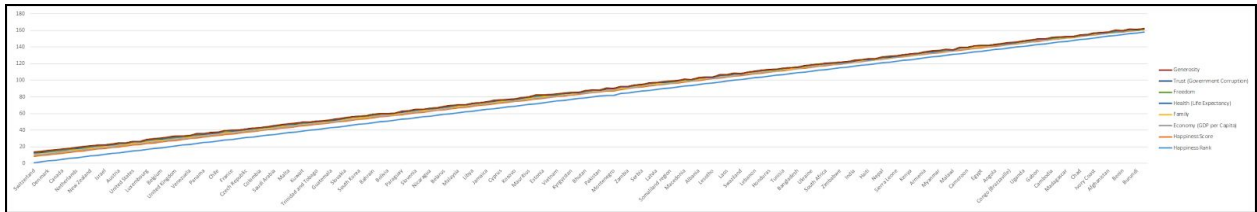


Figure 2: Line graph of happiness ranking (2015)

The final visualization created during the data exploration stage was a breakdown of each country's happiness values. We constructed this data in a stacked bar chart to visualize the various values in an easier to understand way. The chart can be seen below.

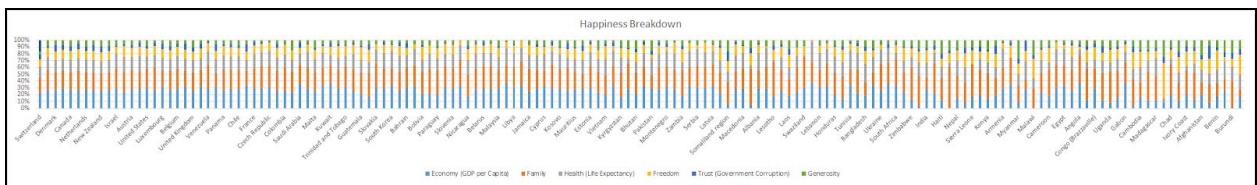


Figure 3: Stacked bar chart of happiness factors (2015)

Ultimately, the visualizations we created during the exploration phase allowed us to understand the data more comprehensively.

Section 5.2: What insights did we gain from those visualizations?

From the three visualizations and the tabular data we were able to identify the components that combine to calculate the happiness score for each country. We were also able to order the countries from “most happiest” to “least happiest”. One of the insights from the data and the

initial visualizations was that it was hard to visually compare neighboring countries or regions because the data was not organized by region but rather by rank.

Section 5.3: How did these insights inform our design?

The insights we gained from our exploration of the data informed our design by the need to visualize the data in a more understandable way. The dataset identifies each country and the happiness rank which prompted us from the beginning to consider visualizing the data in the context of a map.

Section 6: Design Evolution

Section 6.1: What are the different visualizations we considered? Justify the design decisions you made using the perceptual and design principles you learned in the course.

During our design process we considered several routes to visualize the data. One of the initial design ideas was a choropleth map that would encode the data by coloring in a country based on a scale. Since the data was already aggregated by country, it made logical sense to visualize the data in the form of a map. We also liked the idea that when a user hovers over or clicks on a country that specific stats relevant to that country would appear. Below is an initial sketch of the map concept.

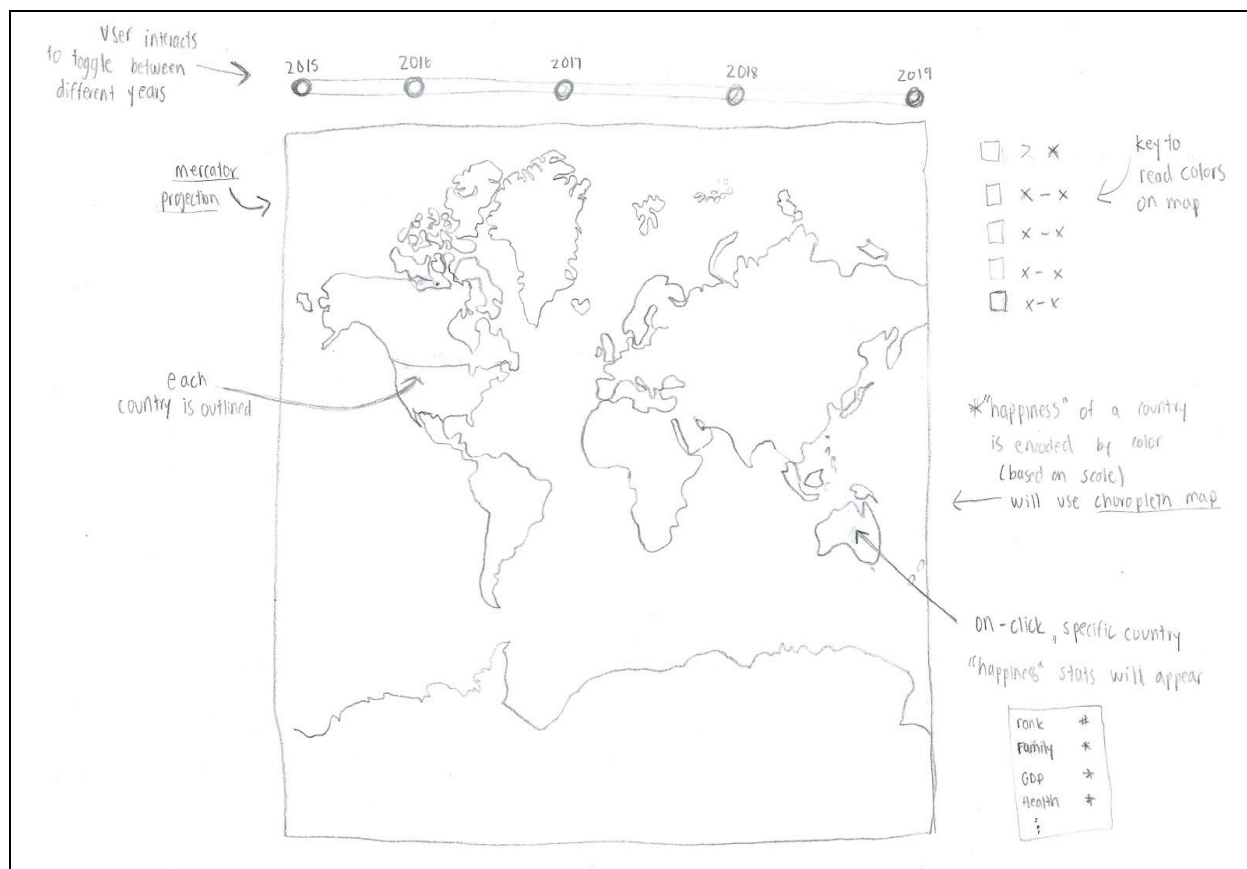


Figure 4: Initial sketch of map concept

Our design then evolved by considering a choropleth map to visualize the data. By using a choropleth map, one country could easily be compared to another country based on the color scale. From our textbook, “a choropleth map shows a quantitative attribute encoded as color over regions delimited as area marks, where the shape of each region is determined by using given geometry”. Based on this definition, the choropleth map would be an appropriate choice to visualize the data. Below is an initial sketch of the choropleth map concept.

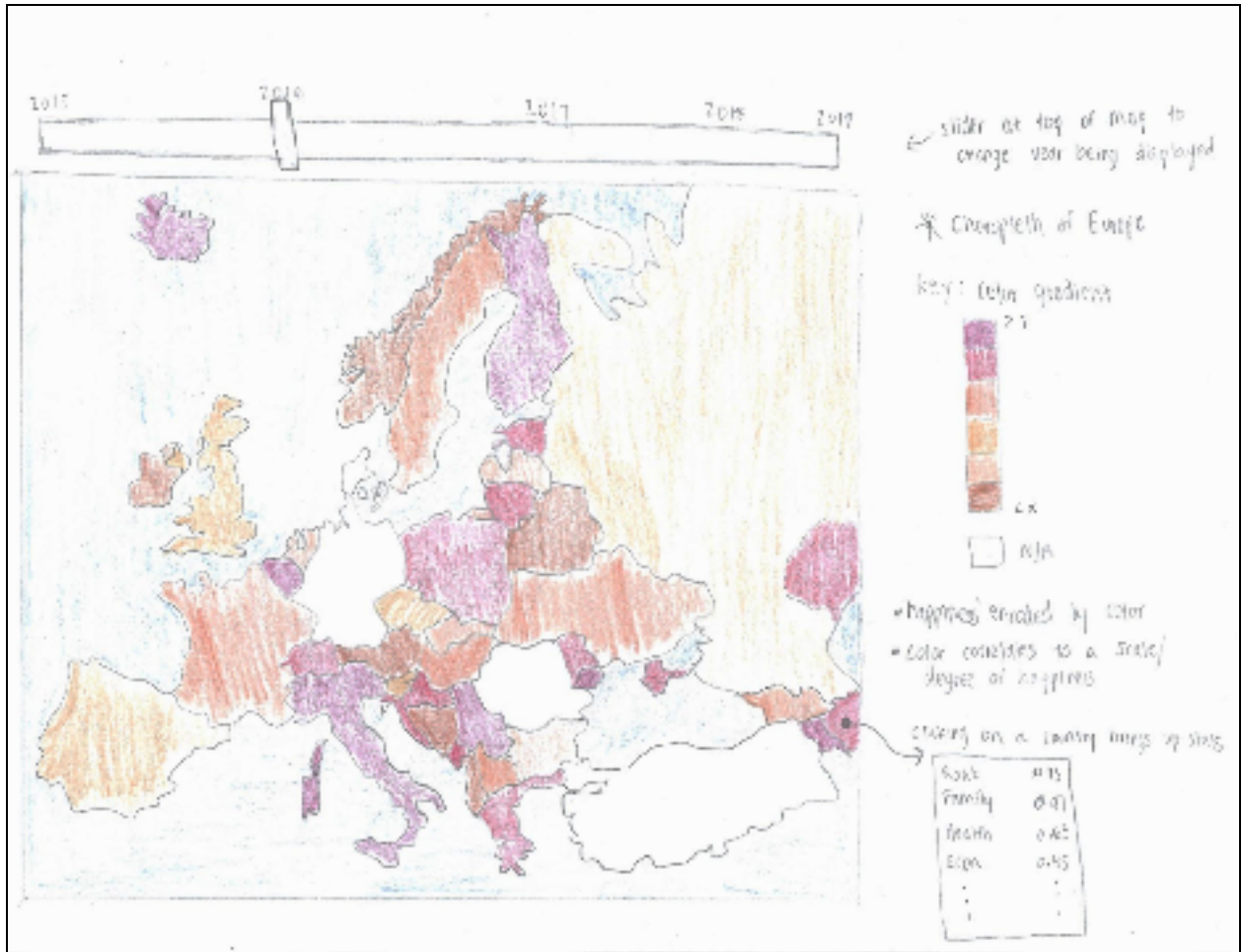


Figure 5: Sketch of choropleth map with timeline and table concept

In addition to the previous sketches, we did a general sketch of what the webpage could look like. At the top of the webpage would be the title, followed by a description of the project, data, and links to the screencast and this report. A major portion of the webpage could then be the map. One potential visualization we thought about was doing a radar map that would appear when a user clicks on a country. The radar map would have several points around for each of the happiness columns (Generosity, Trust, Freedom, Health, Family, and Economy), and the inner circles would be percentages. The sketch of what the webpage could look like can be seen below.

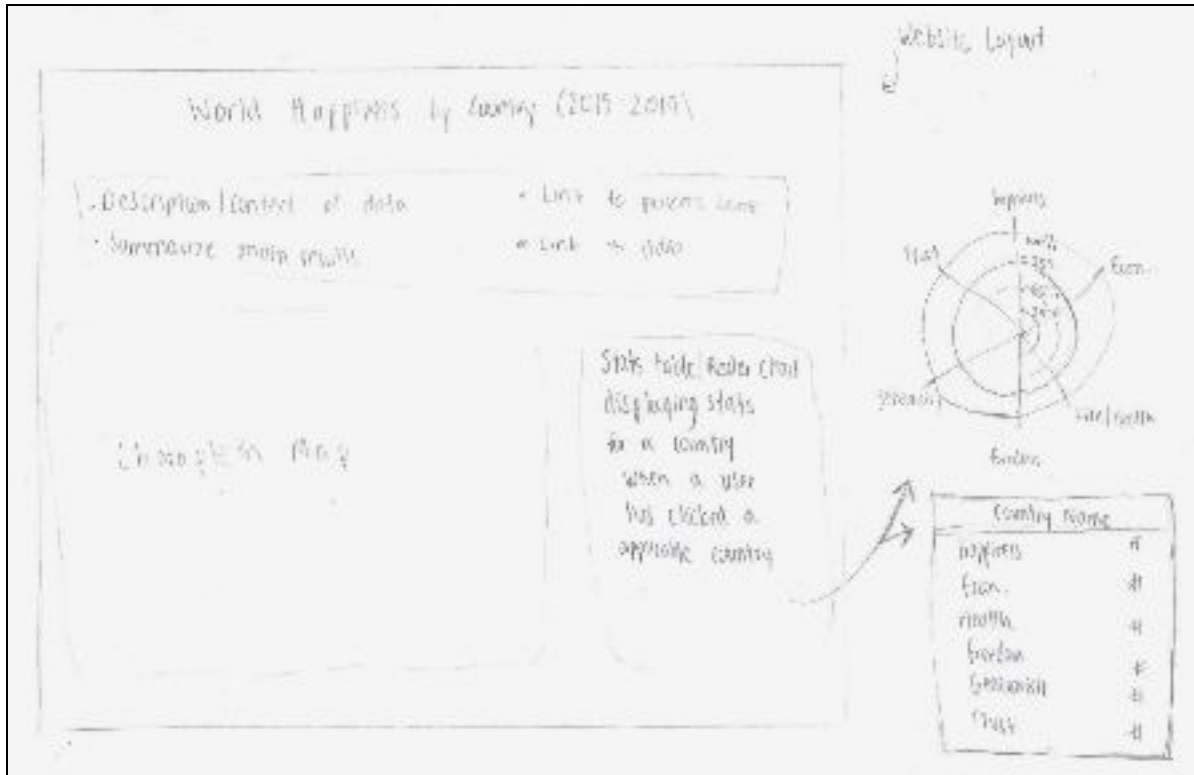


Figure 6: Initial sketch of webpage layout

Ultimately, our sketches provided us an overview and a prototype to base our work off of. The design principles we can see from the sketches include encoding quantitative attributes with color. Using given geometry to show regions as area marks. In one of the sketches we use a color scale to distinguish the varying levels of happiness. In practice the colors chosen would be designed to be monotonically increasing luminance. Where the multiple hues are ordered according to their luminance from highest to lowest.

Section 6.2: Did we deviate from our proposal?

In our original proposal we stated we intended to display the data through the use of maps, specifically choropleth maps to display the happiness in different countries. Thus, we did not deviate from our initial proposal. We decided to go forward with using the happiness by country dataset sourced from Kaggle and then implementing a choropleth map.

Section 7: Implementation

Section 7.1: Intent of Visualization Implementation

We intended for our visualization to be implemented by doing a choropleth map in D3. We began by understanding and implementing a [basic D3 choropleth map](#). The basic choropleth map

takes in geojson data as input. Moreover, there is a common id between the geospatial file and the dataset. The choropleth map also utilizes a color scale to differentiate different regions in the map. The basic D3 choropleth map can be seen below.

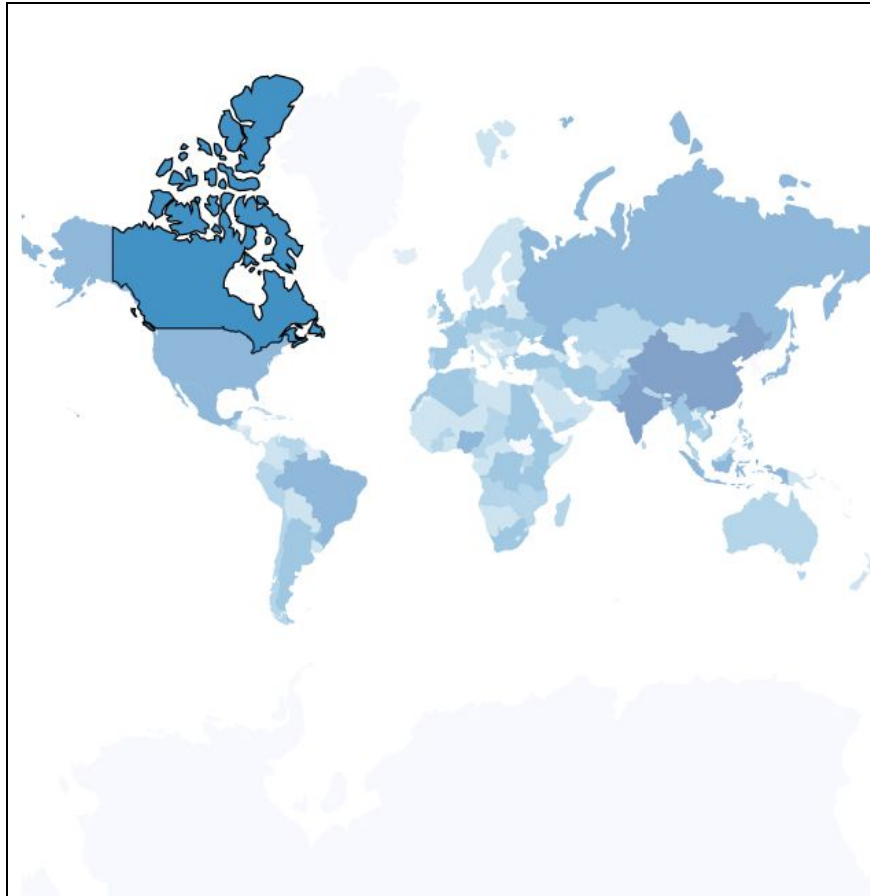


Figure 7: Basic D3 choropleth map

Section 7.2: Functionality of Visualization Implementation

We built our visualization by doing the following:

1. Loading d3.js
 - a. d3.v4
 - b. d3-scale chromatic
 - c. d3-geo-projection
2. Creating table with *Category*, *Max Value*, and *Value*
 - a. Where:
 - i. *Category* is ‘Economy’, ‘Family’, ‘Health’, ‘Freedom’, ‘Trust’, and ‘Generosity’
 - ii. *Max Value* is the maximum value for that category based on the data
 - iii. *Value* is the value for that category based on the country that a user is hovering over

3. Creating the map SVG
 - a. Creation of width and height parameters
4. Creation of the map projection
 - a. `geoPath()` for the path
 - b. `geoMercator` for the projection
5. Data and creating a color scale
 - a. Assign data to `d3.map`
 - b. `d3.scaleThreshold`
 - c. `schemeBlues[7]`
6. Creating the radar chart SVG
 - a. Creation of width and height parameters
7. Creation of the radar chart
 - a. tooltip
 - b. Circle and line creations for diagram
8. Loading the data
 - a. `geojson` data
 - b. `csv` data
 - c. Add the country code to the new `csv`
9. Populate table
 - a. tooltip
 - b. Data for a country displays on hover
 - c. If data is not available, then “NO DATA” is displayed
 - d. Parse and normalize data
 - e. Display the max for each category
10. Drawing and coloring the map
 - a. Draws a black outline over a country on-hover
 - b. Each country is drawn
 - c. The color of each country is set, if a country does not exist on record, null will be returned

For more context, every year (2015-2019) on file for a country is accessible as an array that is placed into the object for each country. For example the image in the figure below is what would be available when a user would mouse-over the United States.

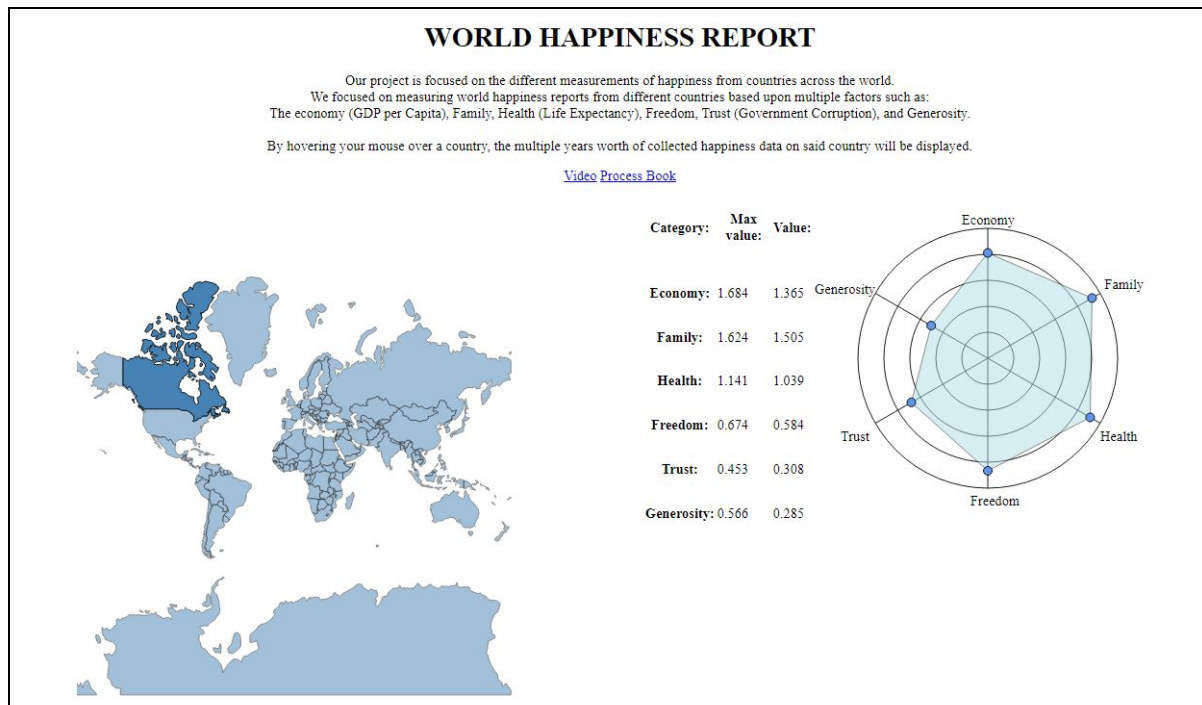
```

(5) [{"Country": "United States", Year: "2015", Happiness_Rank: "15", Happiness_Score: "7.119", Econom...
  ▶ 0: {Country: "United States", Year: "2015", Happiness_Rank: "15", Happiness_Score: "7.119", Econom...
  ▶ 1: {Country: "United States", Year: "2016", Happiness_Rank: "13", Happiness_Score: "7.104", Econom...
  ▶ 2: {Country: "United States", Year: "2017", Happiness_Rank: "14", Happiness_Score: "6.993000031", ...
  ▶ 3: {Country: "United States", Year: "2018", Happiness_Rank: "18", Happiness_Score: "6.886", Econom...
  ▶ 4: {Country: "United States", Year: "2019", Happiness_Rank: "19", Happiness_Score: "6.892", Econom...
  length: 5
  __proto__: Array(0)

```

Additionally, those countries not represented in the dataset or include no entries would return null instead of an array.

Below is a screenshot of our implementation of the choropleth map and radar chart:



Section 8: Evaluation

Section 8.1: What did we learn about the data from our visualizations?

From our visualizations, we learned that happiness varies greatly between regions. Furthermore, the variability of the different factors shares more insight and perspective to the national variations in happiness. From the choropleth map, we can see which countries exist in the dataset and how those countries compare to others in the same region. From the radar chart, we can see the breakdown of factors that represent each country and how they can quantifiably explain why some countries rank higher than others.

Section 8.2: How did we answer our questions?

We decided that the best data/factors to represent happiness are economy (GDP per Capita), Family, Health (Life Expectancy), Freedom, Trust (Government Corruption), and Generosity. According to the data that we decided to collect and visualize, we concluded that some of the happiest countries in the world are China, India, the United States of America, and Brazil. These countries were considered to be the happiest because they had significantly higher GDP per capita and life expectancy when compared to other countries. We decided that the techniques used to answer this data should be in the form of a choropleth map with a linked view

to a radial map of the six factors we chose that displays data for the country that is being hovered upon in the choropleth map.

Section 8.3: How well does our visualization work?

Our visualization works well because it shows the data in an interesting format and ties it with an interactive choropleth map. It is fun to look at the data and find which countries have higher scores than others. The data we used was not intuitive in terms of what the data means, and what scale the data is on. We believe we made it easier to comprehend the data for most people. Sadly, we were unable to show all of the data available to us. We only used the latest data from each country, and we had to cut some columns out from some years to make sure data was compatible across years. However, our visualization is not able to give insights as to how happy a certain individual is within a country.

Section 8.4: Future Improvements

In the future there are a few things we change about our visualization to improve the user experience. The main factor we could improve is discoverability which could be improved by adding new options to allow the user to explore the data. One such feature would be to allow the user to filter the data by year. As it stands now, we only display the collective data from a series of years. By adding this feature, users would be able to see how a country's overall happiness has changed over time. With this data, another feature could be added to take the history of a country's data and construct a series of timeline line charts, with one for each data point. This would be used to better illustrate how each individual data point has changed over the years. Another feature that would improve the visualization would be a filtering option in which the user could filter the scope of countries down to ones who have data that fall into an inputted range. This would allow the user to discover countries that fall into specific categories such as, countries with a low family rating or countries that have a high generosity rating. With additions like these, our visualization would be a great tool for discovering how happy people living in certain countries are and how their happiness has changed overtime.

Section 9: References

- https://www.d3-graph-gallery.com/graph/choropleth_basic.html
- <https://www.kaggle.com/unsdsn/world-happiness>
- <https://stats.oecd.org/Index.aspx?DataSetCode=BLI>
- https://www.d3-graph-gallery.com/graph/choropleth_hover_effect.html
- https://learning.oreilly.com/library/view/visualization-analysis-and/9781466508910/K14708_C010.xhtml
- <http://www.alexanderbastidasfry.com/happy/>
- https://mrkchoi.github.io/WHR_data_visualization/