Visualization of World Happiness Report from 2015 to 2021

Ming Xian, Silvey Yu, Mingqian Zheng

{mx565,ly1164,mz2392}@gmail.com

May 15, 2022

1 Overview

How happy are people around the world? What are the happiest and unhappiest countries? Is there a pattern in where they are located? What factors affect the level of happiness in that country? What's the trend of so-called happiness scores throughout the time?

In this project, we answer the questions above by visualizing the World Happiness Report data during 2015 to 2021, illustrating a global picture (Figure 1, Figure 2, Figure 3) of individual happiness and a time trend for the countries involved.

Here is a link to the data and our source code: https://github.com/SilvesterYu/DATS-SHU235-Information-Visualization-Final-Project.

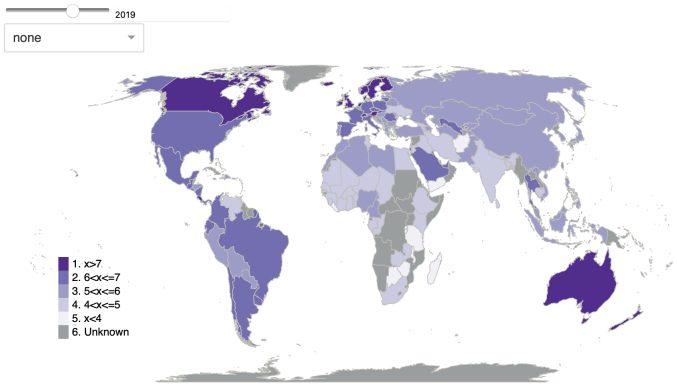

Figure 1: The world map of happiness score for 2019

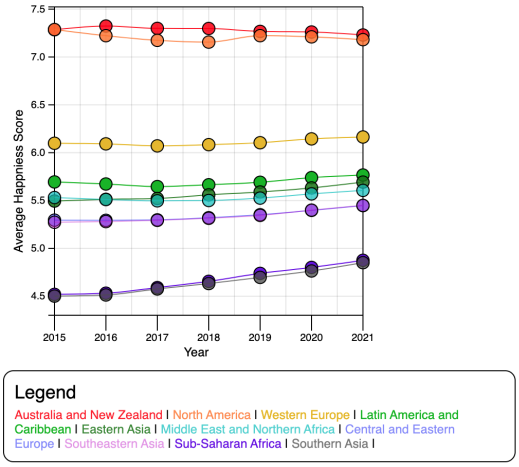

Figure 2: The line chart of the continent's average happiness score

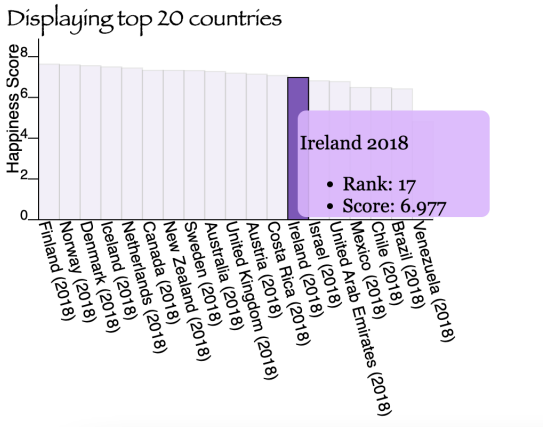

Figure 3: The bar chart of the top 20 countries' happiness score

2 Data

2.1 Data Description

We used a Kaggle dataset from the World Happiness Report project, an on-going project that surveys people from all over the world for their happiness levels, and uses econometrics methods to analyse how much social, economic, and medical factors contributed to the happiness levels. The dataset covers most countries in the world from the year 2015-2022, and the 8 main variables are as below.

- The "Happiness score" (i.e. Ladder score in 2020-2021) is a populated-weighted average score on a scale from 0 to 10. Its standard deviation is also recorded.
- There are 6 other columns indicating how many different socioeconomic factors contributed to the happiness score: GDP per capita, healthy life expectancy, social support, freedom to make life choices, generosity and corruption perception.
- There is a 7th factor "Dystopia". It represents the lowest national averages for each key variable and is, along with residual error, used as a regression benchmark. Adding it with the other 6 factors will produce the "Happiness score".

2.2 Data Preprocessing

As the raw dataset is unorganized and unsuitable for our visualization analyses, we spent a considerable amount of time on data cleaning and merging. The steps are as below.

- According to the original definition, the addition of all the 7 component variables should equal the happiness score. The 2022 data failed to meet the criteria. So we excluded year 2022.
- Countries vary a little from year to year. To maintain consistent comparisons across time, we kept countries that appear every year from 2015 to 2021, using the VLOOKUP function of Excel. The number of countries is 135. The happiness rank for each country was manually re-assigned after excluding some countries.
- The variables and countries are named differently each year. We unified the countries names across the yearly tables.
- For records without "region" attribute, we match them according to the that in other years using the VLOOKUP function of Excel.
- For easier comparison, we transformed all variables to 4 decimal places.

3 Goals and Tasks

Mainly introduced in a scenario-based application, with explanations using abstract visualization languages

We provide users a quick overview of how happy each country is and what aspects of their country make them happy. When a user logs on the system, they can choose a specific year of using the slide bar. On our geomap, the user can see at first glance each country's happiness level by different color saturation and their geographical distribution. Hovering the mouse over the country, a pie chart will show aspects such as economy, healthcare, or freedom, etc., and how much they contributed to the happiness score. The tooltip also includes the country's happiness rank and score.

To the right of the geomap is a line chart demonstrating average happiness scores over 2015-2021, each line representing a continent. User can compare the continents and see the trends. When the mouse is on any point on the lines, a tooltip will show the specific number of that point, countries in that region will be highlighted on the geomap.

A drop down menu allows users to choose top xx (e.g. top 5, top 10) happiest countries. The corresponding scores by country will be shown in a bar chart below by descending order. Sliding the slide bar, different top X countries each year will be highlighted on the geomap.

4 Description of Visualization

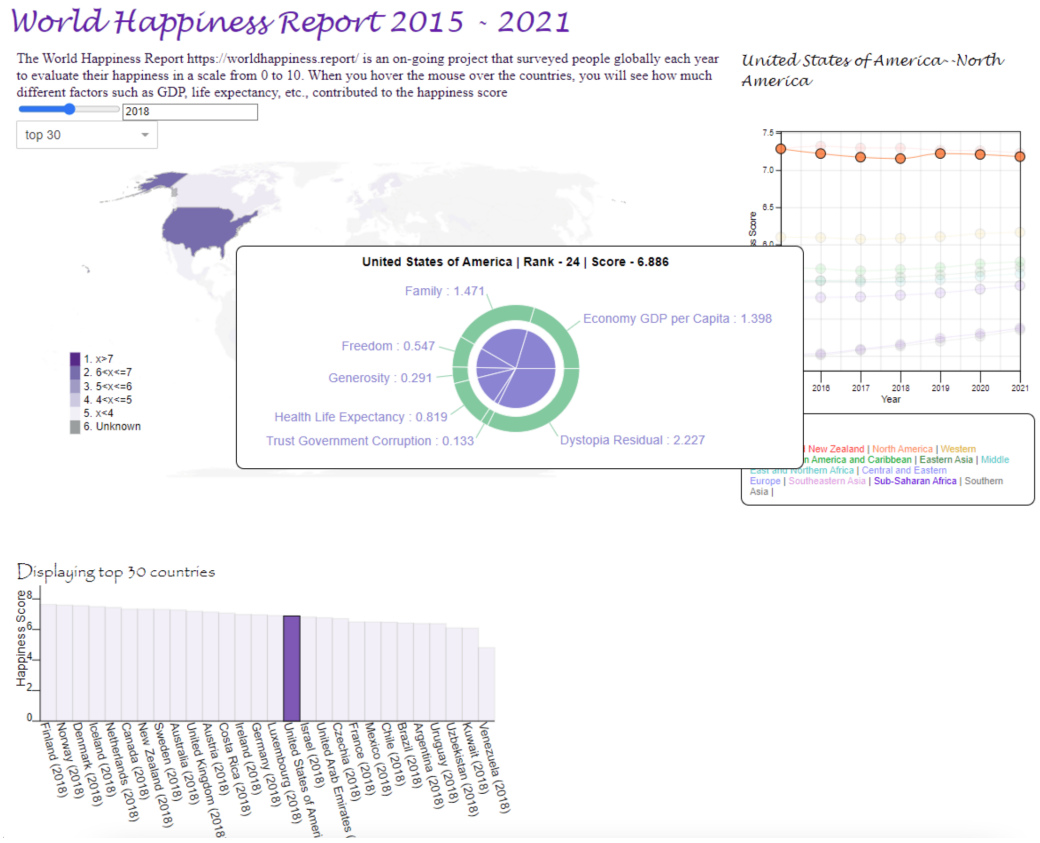

Figure 4: The overview of the linked views

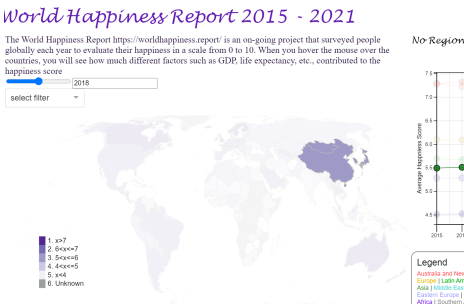

Figure 5: Interaction between line chart and geomap

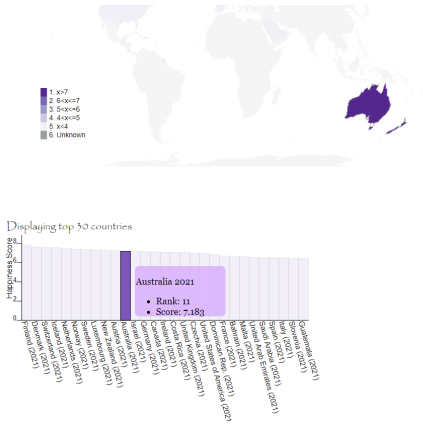

Figure 6: Interaction between bar chart and geomap

A **line chart** of each region's happiness trend during 2015 and 2021 gives users a static overview of the global happiness trend. We use line mark with amplified points with hue channel. Each line with a distinctive color represents a region and the nodes along it refer to the average happiness score in each year.

A **world map/geomap** is filled with color indicating happiness level of each country at the year specified by the slide bar. We use the saturation channel to encode the happiness level where higher scores correspond to greater saturation.

A **bar chart** shows the happiness scores of top X countries in descending order at a certain year collectively determined by the drop-down bar and slide bar.

4.1 User Interaction

4.1.1 Selection

The **slide bar** at the top left corner allows user to select a specific year of happiness scores used in the world map. The **drop-down bar** below is designed for the section of top 5, 10, etc. happiest countries, and it will make other countries half-opaque to emphasize the filtered countries.

4.1.2 MouseOver

- **Geomap:** When the mouse hovers over a country on the map, a **MouseOver** the country will be highlighted on the map, along with the corresponding region line in line chart and the bar in bar chart. A **tooltip** will also appear with (1) country name (2) country's happiness level (3) country's happiness rank (4) **pie chart** showing the composition of the happiness score at that year. Users can see what factors contributed most to the happiness level, and how much they each contributed. The pie chart is created using the Recharts library.
- **Line chart:** When the mouse is over a point of a specific line (for example that of Eastern Asia), a **tooltip** will appear, next to the selected point, displaying the region name, happiness score of that point, and the year of that point. the corresponding region (in this case, countries in Eastern Asia) on the map will be highlighted and other regions will be half opaque.
- **Bar chart:** When the mouse hovers on a bar, the selected bar and the area of that country in the world map will be highlighted. A **tooltip** will appear next to the selected bar displaying the country name, rank and score.

5 Reflection

5.1 Project Development

We chose this topic from 3 drafted projects because of its many usage scenarios and data availability. Nowadays, people have increased options in where to live and are more aware of mental well-being. Thus, we believe it is meaningful to visualize and compare the happiness levels time-wise and world-wise.

We timely adjusted our schedule as we found the original dataset needs much pre-processing. We explored several matching and cleaning methods using Excel, Python and Stata. During the coding process, we made reflections on each completed feature to make sure it is useful and clear.

The coding took a considerable amount of time. Development details are as follows: **Geomap** has many linked interactions and layers of selection. The color scheme is purple, considering other colors like red or green might not be friendly to people with color vision deficiency. **Line chart** is coded from scratch using React and d3. An auto scaling according to min, max values of the lines is used, so we are showing only the effective range (from 4.5 to 7.5), and thus we have less clutter, lines are more apart and the trend is more visible. A Javascript color selector is used when selecting the colors for the lines. **Bar chart** has a flexible length for the x axis according to the number of top X countries selected. **Tooltip** contains a pie chart, so it is coded using the pie chart component in Recharts for it provides simple and beautiful solutions to bar charts.

5.2 Changes in Goals

- As explained in previous sections, our visualization samples only cover the year 2015-2021 without 2022 due to its data inaccuracy.
- We did not include a line for a specific country's happiness score to compare with its continent. This is because this feature of comparing the country with its continent does not provide useful information to the audience. A user can first decide a satisfactory continent based on the line chart, and narrow their options to see a country in that continent by hovering the mouse on the world map. If they want to see this country's score, they can simply see it by the tooltip of the world map.
- For the drop-down bar, we did not include the option of selecting the countries with the lowest world ranking as planned. We believe our audience is more interested in seeing the top happiest countries when he chooses to use our platform. If he really wants to know which countries have lower rankings, he can have a rough idea by looking at the shallowest areas in the world map.