Data Visualization Project

Report

Group 19

M20190932 Ana Cláudia Alferes
M20190089 Francisco Carujo Neves
M20190301 Inês Diogo
M20190935 Luís Filipe Pinho

Teachers

Pedro Cabral
Nuno Alpalhão
António Leitão

Index

1. Dataset Description

Context:

The World Happiness Report is an important survey of the state of the happiness in the world and it ranks countries over the years by their happiness levels. The report is important to many organizations, governments and civil society as it has been used to help make decisions and to evaluate the progress of nations.

Content:

The happiness scores and rankings use data from the Gallup World Poll. The scores are based on answers to an evaluation question asked in a poll. In this question, on a scale from 0 to 10, being 0 the worst life possible, and 10 the best, respondents had to rate their current lives. The scores are from representative samples for the years 2015-2019 and use the Gallup weights to make the estimates representative. The columns following the happiness score added together are equal to this value, which makes it possible to observe what factors contribute to the happiness in each country. Only the common columns were used in our study - GDP per Capita, Life Expectancy, Freedom, Government Corruption and Generosity.

2. Choices & Interactions

Inspiration:

It is from general knowledge that depression is considered the main disease of the 21st century. With that in mind, we found this report very interesting, in a way to perceive which are the factors that most contribute to the happiness in each country of the world. Therefore, from the several datasets that caught our attention in the internet, this was the one that aroused curiosity among us, and we decided that this was the choice to go with.

Type of Interaction:

In our dashboard, the user will have the opportunity to visualize what are the key factors that affect the world happiness, with overall views and more detailed ones. The viewers will be able to establish comparisons between years and countries and know what the happiest countries in the world are.

3. Visualization

Data Encoding:

In order to have a clean and interesting dashboard, we aim to provide a varied selection of visualization types. The first type of visualization that was computed was a line graph which shows the evolution of the average worldwide Happiness Score over the years. This graph is the only that is static, and that will not change with the filters applied.

The next type of visualization is a world heatmap, which is also related to the World Happiness Score by country. The bluer the colour, the higher the score, and the redder the colour, the lower the score. This map changes by selecting the year in the filter. There will also be a dedicated space to the Happiest Country in each

year. This country will be presented by its flag, and it will change by the years, because it not the same in every period.

There are also two bar plots. The first bar plot shows the information of the percentage of the countries included in each continent by the Happiness Score quartiles. This bar plot will be possible to filter by year, as the distribution of the percentages change with time. This is interesting because it shows the continents which have the happiest countries and the least happy.

In order to establish a comparison between the weight of some variables in each country happiness score, we plotted a scatter plot, involving every country. In this scatter, we compared the Corruption variable (y) with the Happiness Score (x), the Health (color) and the GDP per capita (size). This scatter will change by year.

Another visualization that can be seen in the dashboard is the correlation matrix between the variables. This matrix is very important, in order to understand the relationship between the variables used to know how the factors explain the Happiness Score. In this matrix, the main goal is to know the correlation between the Happiness Score, and the variables that contribute to its value.

The next visualization is another line graph, but with different characteristics, and a different goal from the previous one. This line graph shows the evolution of the Happiness Score of a selected country along the years, of a selected country in the dropdown (which is better explained in Data Filtering).

The second bar plot shows the distribution of the weights of the key factors that explain the Happiness Score. This graph presents the distribution in five countries: the top two countries with the highest scores, the bottom two countries with the lowest scores, and a selected country from the dropdown.

Data Filtering:

The dataset provided allowed us to explore the main factors in a way of comparison and evolution. Because the data is organised in countries and years, we decided that the main interactions of our dashboard would be regarding these two measures. For that reason, we added a dropdown, for the user to select the pretended country, and visualize the information related to it. In addition, a slider with all the years was computed, from 2015 to 2019, so that the main graphs displayed change their visual for the information regarding the selected year.

4. Technical aspects

The dashboard was programmed in Python language, using JetBrain PyCharm Community Edition, linked to our Anaconda Environment. Regarding the code used, all the graphs that were made, were plotted by using Plotly, and disposed in the dashboard by using Dash. The code will be hosted on a repository on GitHub, which is used to deploy Heroku.

The code used can be found in the repository of the following link: https://github.com/Fran1iZ/Data-Visualization

5. Discussion

Accomplishments:

Overall, the world happiness has increased in the last years, with a small fluctuation in 2017. After a small research, we found some events that might had an impact in this decrease, such as the start of the presidential mandate of Donald Trump, the start of Brexit negotiations, the conflicts in Venezuela, and the increasing alarms about climate change. Nevertheless, this downgrade did not compromise the trend, and later, the average value of Happiness Score increased again.

Regarding the variables, we may observe by the correlation matrix, that the Happiness Score has a very strong positive correlation with the GDP per capita, and Life Expectancy variables. The remaining variables, which are the Government Corruption, the Generosity and the Freedom have a much lower correlation overall.

Another conclusion that we took is that the northern European countries, and Switzerland, appear to be the ones with the highest score, as well to have the highest values for the economic indicators. This means that these factors are those who contribute the most for their happiness.

A special mention must be made to Rwanda, that has one of the lowest Happiness Scores, but stands out in the scatter plot, by having a big value for the Government Corruption. This country is considered one of the most corrupted in the world, and by the graph, we can conclude that the government in charge sets a major role in the people's happiness, by the negative. The countries that appear to have the smallest Happiness Scores are those from Sub Saharan Africa and conflicting countries in the Orient, such as Iran and Afghanistan. These countries do not have high values for the Health variable, meaning that it does not have great impact on their happiness, in contrast with the countries with highest Happiness Score.

Limitations:

Giving that the scores presented by variables are in a regression model, we cannot say that the nominal value of a variables is comparable to the score. For example, we cannot conclude that the highest GDP per capita represents a higher Happiness Score. We can only conclude if the GDP influenced the happiness of the population or not, regardless of the actual value.

The second limitation found in this analysis is that every year has a different set of variables. Some variables maintain over the five years, but others do not, and that influences our work and analysis when we want to compare each year. In addition, some of the variables that maintain, have different names in different years. Since the database is the same and it was made by the same source, we believe that it should have the same name for the variables over the years.

Future work:

This project gives an opportunity to discuss what are the key factors that affect the happiness of several populations, and allows us to conclude that, besides the way of living is different, they all have different values. Some populations give more value to the health in their country, and some give more importance to their gross annual income. This report could be useful to study the behaviour of country populations, and even cluster them in the future. Moreover, this report could be important to people that are going to study or work abroad and want to know more about the personality of the people in the destination country.

6. References

World Happiness Report. Retrieved from: https://www.kaggle.com/unsdsn/world-happiness [Accessed 12th December 2019];

Plotly Chart Studio. Retrieved from: https://chart-studio.plot.ly/feed/#/ [Last accessed 15th January 2020];