IMD CA 1

Liam Glennie

November 18, 2021

1 CA Summary

In this study, we'll use visualisations to analyse the data from the World Happiness Report from 2015 to 2019. The main objective of this study is to identify which aspects of life are most related to happiness and to describe how each country's happiness has evolved over these years.

2 Background

Most people want to be happy or, at least, get to a point in life where they are satisfied with their level of happiness. So, hopefully, we can achieve results from this study that can explain why certain countries are happier than others and whether we can learn from these results in order to obtain a higher level of happiness.

Now, most of the data we use is from the World Happiness Report which is an organization that studies and analyses happiness around the world, trying to find the main aspects that influence it. However, each year's report goes into further detail on certain concrete topics instead of the main aspects that we evaluate in this study.

3 Data sets

For this study, we used a total of 7 data sets. However, two of them didn't get included in the final dashboards and the other 5 were fused into one for visualisation purposes because they had the same attributes but corresponded to different years.

- Country Data (Not used for final dashboards)
- GDP (Not used for final dashboards)
- Final Happiness Data (includes the data from 2015-2019)
  - 2015
  - 2016
  - 2017
  - 2018
  - 2019

4 Seven Stages

4.1 Acquire

The GDP and the Final Happiness Data was obtained from Kaggle and the Country Data was obtained from Excel. If you have geographical data, such as countries or states, you can obtain information from Excel's data interpreter.

4.2 Parse

In this step, we cleaned the data and removed any country that wasn't part of every year data set or the GDP data set. So, if there was a country that was only in the data sets 2015, 2016, 2019, it was removed. For this, we used Excel's *Conditional Formatting* tool to highlight the unique values. There were countries that had different names over the data sets, therefore we changed those instances to have a common name. We also renamed the columns and ordered them in terms of interest.

4.3 Filter

In the filter step, we removed any columns that weren't of interest or weren't part of all the data sets that composed the Final Happiness Data. Initially, we removed the GPD attribute in the data sets 2015, 2016, 2017, 2018 and 2019 but after realising our mistake, we inserted back in.

We also created a new data set called Final Happiness Data which groups all the data from the data sets 2015, 2016, 2017, 2018 and 2019. This was done so we could have an attribute called "Year" that could be used in Tableau to create pages and to be able to create an area chart to visualize the evolution of a country's happiness score over the years 2015-2019.

4.4 Mine

In terms of mining, 2 attributes weren't available in all the happiness data sets, such as a country's happiness score rank and their "Dystopia Residual". Therefore, we removed them because they could be calculated in Tableau.

The rank is a calculation field that takes as input the sum of "Happiness Score" and ranks each country. We used the RANK_UNIQUE function to avoid having 2 countries with the same score being ranked in the same position. If we had used the normal rank function, we could have had 2 or more countries in the same row of the "Countries Ranked By Happiness" table. Here's an example of the problem:

85	Morocco	Azerbaijan	5.208	5.208
86				
87	Lebanon		5.197	

Figure 1: Country Ranked Table using RANK instead of RANK_UNIQUE

The "Dystopia Residual" is how each country to an imaginary country called Dystopia which is an imaginary country with the worst scores for each of the attributes. In summary, Dystopia is the realistically least happy country that could exist. The way it's calculated is by subtracting from "Happiness Score" the sum of "Family", "GDP per capita", "Life Expectancy", "Freedom", "Trust" and "Generosity". In the workbook, you'll find a "My Happiness Score" calculated field which is just the sum that we mentioned before. The reason why this calculated field exists is that at one point, we were going to use a different happiness score that was based on the attributes that we did have access to so that the sum of the attributes was equal to the happiness score.

However, after further evaluation, the ranking changed using the new happiness score and it seemed like we were incorrectly manipulating the results. Therefore, we decided to calculate the "Dystopia Residual" to complete the set of attributes related to happiness.

4.5 Represent

We decided to divide this step into 3 parts:

4.5.1 Visualisations used in the main dashboard

- Map

  With countries being one of the main attributes of the data sets, it seemed fairly obvious to include a map of the world where a country has a different colour depending on its happiness

score. Therefore, the happiest country will be coloured in with dark green, the country with the median happiness score will be a golden colour and the least happy country will be dark red. Furthermore, when you click on a country, you'll be redirected to the secondary dashboard with specific data for that country.

- Scatter plots

  As one of the objectives of this study was to investigate which aspect of life has the most impact on happiness, we decided to use scatter plots to represent this. Each circle of a scatter plot represent a country and the colour of the circle depends on the country's happiness score. In the final dashboard, we included the plots of the 4 attributes most correlated to the happiness score.

- Country Ranked Table

  This is a table that shows the countries ranked by their happiness score. So, there are 3 columns, the country's rank, the country's name and its score in the form of a bar. This bar has the happiness score as its label and its colour also depends on it.

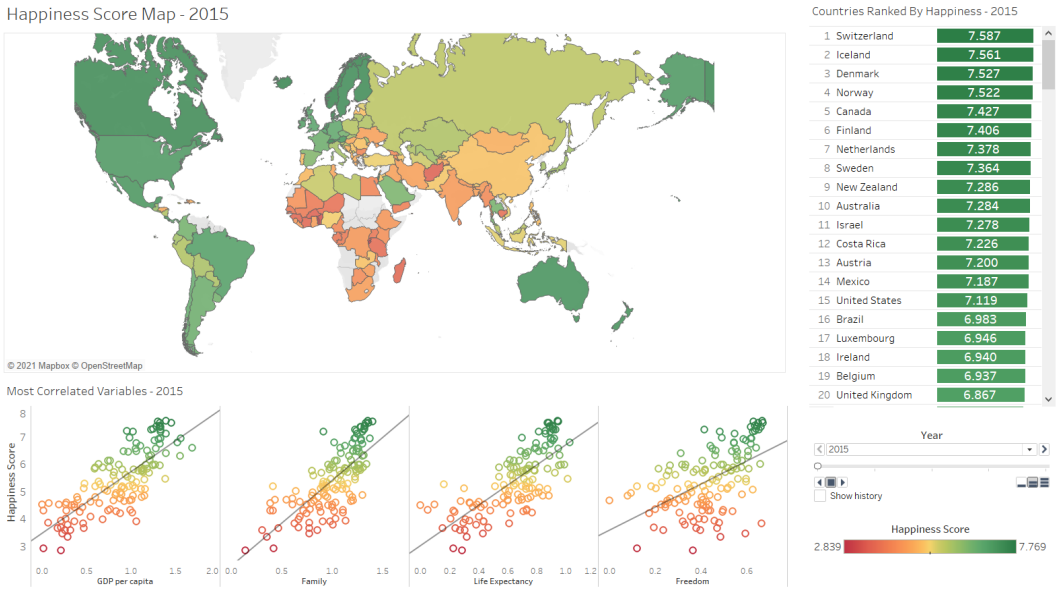

Figure 2: Main Dashboard

4.5.2 Visualisations used in secondary the dashboard

- Happiness Breakdown Bar Chart

  This bar chart aims to break down the happiness score, showing how much of an effect each attribute has on it. The first six attributes are ordered in terms of their impact on happiness. It seems that most countries, that most countries value highly their family, GDP and life expectancy or health. Finally, "Dystopia Residual" doesn't follow this rule and is coloured in a darker colour to bring it user's attention because we believe it's an attribute that not many people would take into account.

- Happiness Evolution area plot

  In order to visualise the evolution of the happiness score for each country over the years 2015-2019, we believe that the area plot is the most appropriate for this goal.

- Flag

This is an improvised plot that has as a column the country name. Each country has a custom shape associated with it which is its flag. To be able to do this, we have created a new shape folder in MyTableauRepository.

- Definitions

  We felt the need to include a text box with the meaning each attribute had for the World Happiness Report. Without this, the visualisation could be interpreted differently and possibly wrong.

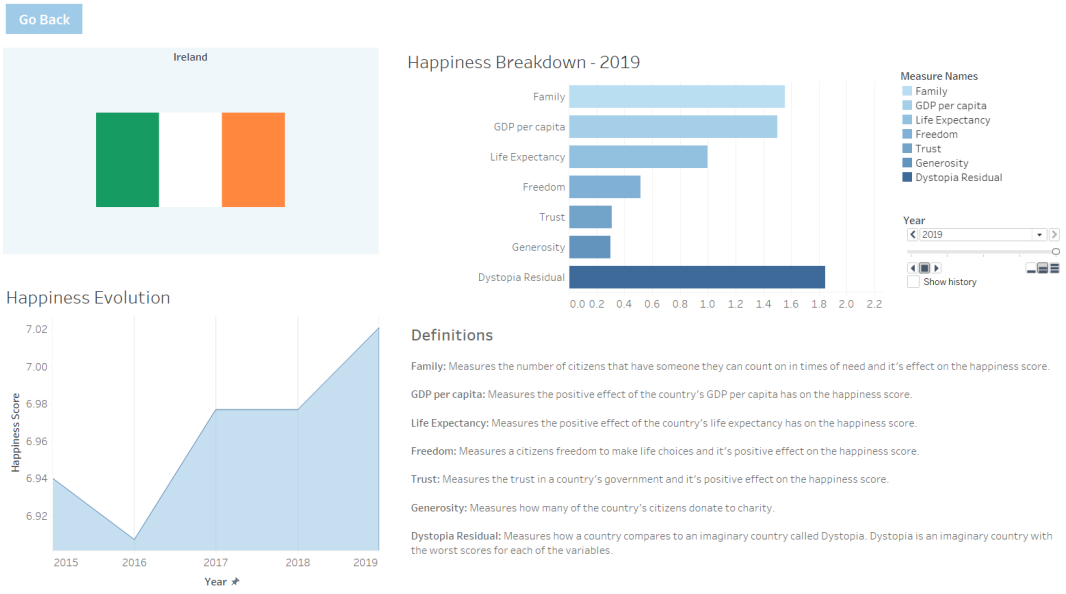

Figure 3: Secondary Dashboard

4.5.3 Visualisations not used in dashboards

- Country Information Table

  This is a table with general information about the country. This visualisation was replaced by the text box with the attribute definitions because we believe that it provides more to context to the user.

- Other scatter plots

  We could have included the scatter plots of all the attributes that have an effect on the happiness score. However, they didn't all fit and some didn't provide any useful information.

Country Data

Area km^2	243,610
Carbon dioxide emissions (kilotons per year)	379,025
Population	66,834,405
Unemployment rate	0
Urban population	55,908,316

Figure 4: Visualisations not used in final dashboards

4.6 Refine

In this step, we changed the names of the sheets so they had a more adequate name, changed the colours to follow each dashboard's colour palette and removed any unnecessary lines from the charts. In the case of the area chart, we changed the starting point to not include zero because it accentuates the difference in the happiness score over the years. Here are the changes:

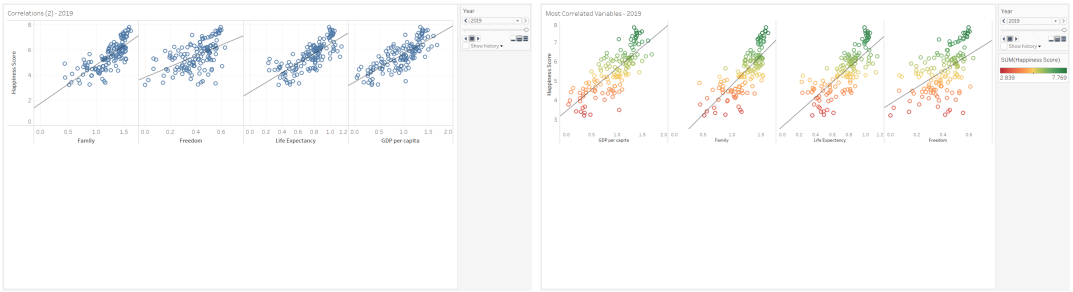

Figure 5: Scatter Plots Before and After

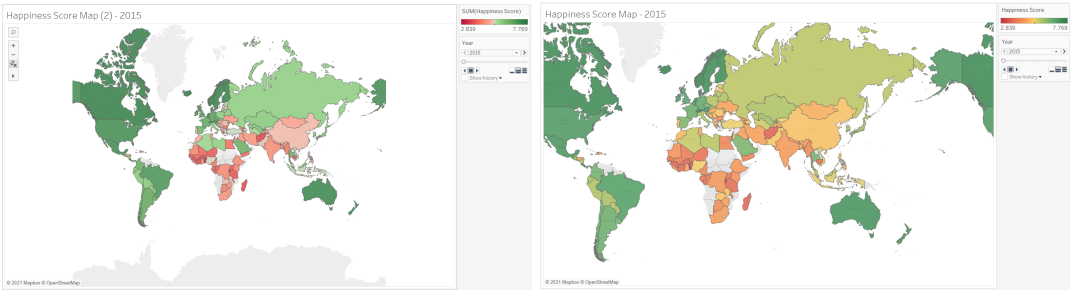

Figure 6: Happiness Score Map Before and After

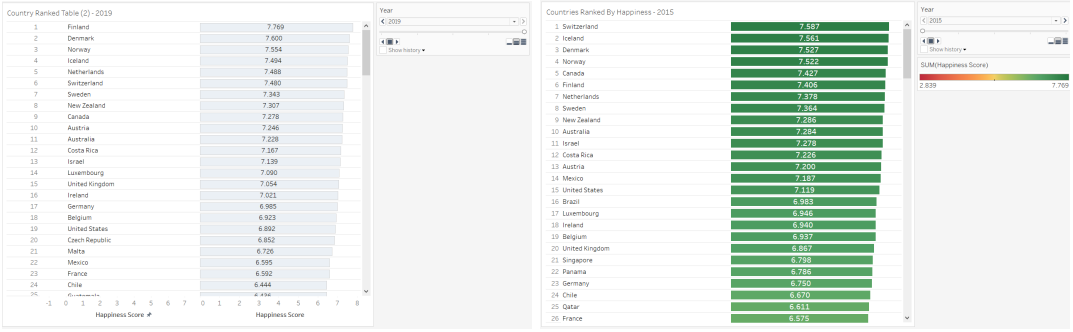

Figure 7: Happiness Ranked Table Before and After

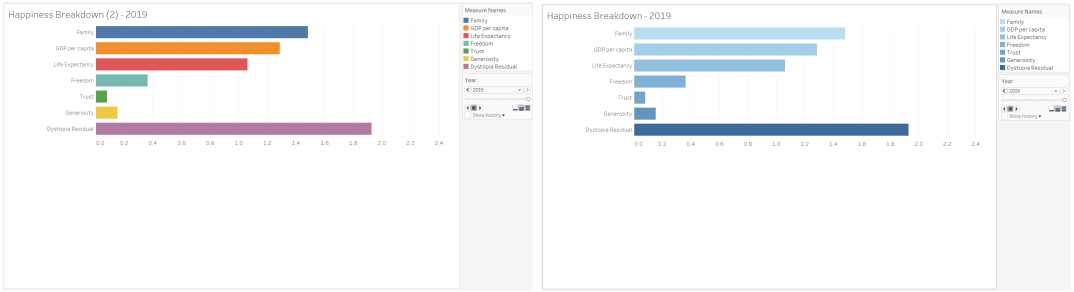

Figure 8: Happiness Breakdown Bar Chart Before and After

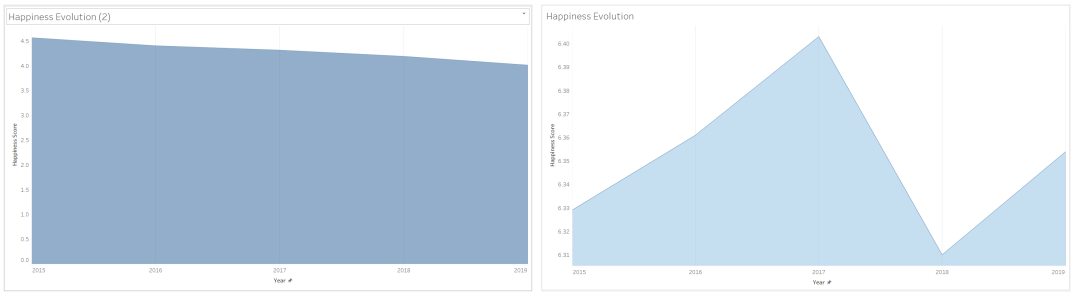

Figure 9: Happiness Evolution Area Chart Before and After

4.7 Interact

In this step, we've given users the ability to change the year the data from on certain visualisation. Also, when you hover over any of the charts, you get the value of the attributes that were involved in building that chart.

5 Problems & Solutions

The main problem we faced at the start was to show the evolution of the happiness score over the years and to use the *Pages* tool within Tableau to be able to filter out a certain year. We couldn't do this because the data sets were separated into different ones for each year. Therefore, we decided to compile one final data set formed by the data from the World Happiness Report over the years 2015-2019 adding an attribute for the years. This way, we could use said attribute in the *Pages* tool in Tableau to achieve what we were looking for. This solution was also used to be able to use an area chart to visualise the happiness evolution.

The second noticeable problem we had was with creating the Country Ranked Table. This is the table that sorts the countries in terms of their happiness score. However, in our initial attempt, the countries wouldn't sort themselves automatically if you changed the year from 2015 to 2018 for example. In the end, we manage to find a solution in the Tableau community that we could adapt to this study in order to achieve the visualisation that we wanted to build.

6 Conclusion

We chose this topic to see if there were any aspects of life that had the most effect on one's happiness. We've found that the most important factors are "GDP per capita", "Family" and "Life Expectancy". This can be seen in the scatter plots and also in the happiness breakdown bar chart.

GDP makes sense because it allows a country to afford certain luxuries. However, some countries that have a high GDP that don't report to be amongst the happiest countries in the world, such as,

China or India. This could be because the country's wealth isn't well divided amongst its citizens, having a reduced group enjoying most of the wealth and a larger group in poverty.

Now, a country's GDP, generally, isn't something you can control. However, the other two aspects that are related to happiness are. Starting with the "Family" attribute, it measures whether you have someone in times of need and its importance. So, this aspect could be improved by fortifying relationships with family members and friends, making sure you have each other when needed. In terms of the "Life Expectancy" attribute, it measures the importance of having a long healthy life. Again, this is an aspect of life that you have certain control over by taking care of yourself both physically and mentally. It seems that by staying healthy, you can live a longer happier life.

Next, there was an interesting attribute in terms of how little it seemed to affect a country's happiness score, "Trust". Now, this measures the citizens trust in their country's government and how it affects their happiness. And even in countries where there is very little to no corruption within the government, its effect was quite low on the happiness score.

In short, a person's happiness depends on where they live, how they take care of themselves and the support of those people that are around them.