**Term Project – Final Submission Report**

Akiko Koyama

Dec 5, 2022



## Choice of Data

There are two data sets chosen for this project:

1. WHO Life Expectancy 2000 – 2016

Kaggle, <https://www.kaggle.com/datasets/mmattson/who-national-life-expectancy>

This dataset used is based on indicators published by World Health Organization (WHO) and United Nations Educational, Scientific and Cultural Organization (UNESCO) for years 2000-2016 for 183 countries​. It includes 32 features, most of which are continuous data expect for few categorical data of country and region. The features used in the analysis for this project are as follows:

* life\_expect: Life expectancy at birth (years)
* infant\_mort: Death rate up to 1 (%)
* age5-19 thinness: Prevalence of thinness among children and adolescents, BMI < -2 standard deviation (%)
* age5-19 obesity: Prevalence of obesity among children and adolescents, BMI > +2 standard deviation (%)
* measles, polio, and diphtheria: immunization coverage among 1 year old (%)
* basic\_water: Population using at least basic drinking water services (%)
* gghe-d: Domestic government general health expenditure as percentage of gross domestic product (GDP)

1. World Happiness Index 2008 -2021

World Happiness Report, <https://worldhappiness.report/>

The report is published by Sustainable Development Solutions Network, which is an NGO established by United Nations in response to UN’s resolution to measure each country’s happiness so as to utilize the data in its public policy. The happiness score is based on respondents’ ratings of their own lives from over 150 countries based on the following six variables:

* real GDP per capita
* social support
* healthy life expectancy
* freedom to make life choices
* generosity
* perceptions of corruption

The score ranges from 0 to 10. The first report was published in 2012 and since then the report is published annually on UN’s International Day of Happiness which is the 20th of March.

## Analysis Questions

Life expectancy and happiness score explores the quality of life from biological and psychological aspects. The aim of the project is to explore some global patterns and trends, as well as any relationships that exists between the two indexes. Specifically, the questions posed are:

1. What are the factors impacting the life expectancy and happiness of score of a country?
2. What are the global patterns and trends?
3. Is there a correlation between happiness score and life expectancy?
4. How is Canada performing in terms of life expectancy and happiness score?

## Story – Global Life Expectancy and Happiness Score

Graphical user interface, application

Description automatically generated Map

Description automatically generatedGraphical user interface, chart, application, line chart

Description automatically generated

Graphical user interface

Description automatically generatedGraphical user interface

Description automatically generated with medium confidence Chart, bar chart

Description automatically generated

Chart, bubble chart

Description automatically generated Graphical user interface, application

Description automatically generated Chart, funnel chart

Description automatically generated

The Story is comprised of the 6 dashboards and one worksheet as well as the cover and ending pages. The story starts by providing the big picture, visualizing the global pattern and relationships between life expectancy and happiness score. Then we look closer into the factors that affects life expectancy followed by investigation into the happiness score data and ending with a case study of Canada.

1. *Global Comparison – Life Expectancy and Happiness Score.*

**Map, scatter chart

Description automatically generated**

This story point illustrates the high correlation between life expectancy and happiness score. The two filled maps with color encoding indicating the degree of life expectancy and happiness score indicates a similar pattern where African region is least happy with lowest life expectancy and the OECD countries being happier with higher life expectancy. The scatter plot with happiness score in the y axis and life expectancy in the x axis indicates a linear relationship where countries with higher life expectancy tend to have higher happiness score. The filters enable the user to filter according to the region as well as by country and clicking on the icons enable the user to access the data sources.

1. *Life Expectancy – Global Trend.*

Chart, line chart

Description automatically generated

The line chart titled “Life Expectancy – Global Trend” indicates rising life expectancy across all regions. Coinciding with this trend, the average vaccination rates of polio, diphtheria and measles are also increasing globally which is one of the reasons behind the improving trend of life expectancy. The bottom chart “Percentage Difference in Life Expectancy y.o.y” looks into some of the dips observed in the Life Expectancy Trend chart. As described in the annotated points, the sudden changes in life expectancy observed in some countries are mostly due to natural disasters or civil wars.

1. *Life Expectancy – Access to Drinking Water.*

Graphical user interface

Description automatically generated

Researching into the factors affecting life expectancy led to a finding of the study published in 1999 that indicates strong relationship between having drinking water access and mortality, which is included in the dashboard for reference as a web page object. On the left is the visualization this relationship with our data. The color encoding indicates life expectancy, and the size of the bubble indicates access to drinking water. As the study suggests there is a high correlation between the two that the countries with lower life expectancy have lower rate of access to drinking water which is why the red bubbles are smaller in size.

1. *Life Expectancy – Distribution*

Chart, bar chart

Description automatically generated

The two bullet graphs on the top row illustrates the prevalence of obesity and thinness in childhood. The orange bars indicate the obesity level of each country between age of 5 – 19 and the reference line indicates the thinness for the same age group. The left graph is the bottom 10 countries, and the right is the top 10 countries in terms of life expectancy. For the bottom 10 bullet graph, the thinness is prevalent and exceeding the obesity rate in all countries, whereas for the top 10 countries, obesity is prevalent in all countries and exceeding the thinness rate. This is indicative of the trend that as countries get wealthier, the life expectancy increases but so does the obesity rate, which is a problem seen in developed countries where there is enough wealth to feed, or often overfeed the population. On the bottom row, the histogram visualizes global distribution of life expectancy with added feature of government health expenditure using the color encoding. The countries falling under the bin with higher life expectancy have higher government health expenditure, indicating the vital role, the government plays in improving general health of the public and therefore leading to longer life span. Lastly the box-and-whisker plot shows the dispersion and the range of infant mortality rate for each region. Africa has the highest infant mortality rate which is likely to be impacting the lower life expectancy of the region. The filters provided enable the user to look at the data at a more granular level according to the region or the year.

1. *World Happiness*

Graphical user interface, chart, application, Excel, bar chart, pie chart

Description automatically generated

This story point moves on to the exploration of the Happiness Score data. The “Top <N> Happy Nations in Year <selected year>” displays the global happiness ranking. The color indicates the region and ranking can be adjusted using the Top N parameter as well as the year which can be selected using the single value slider filter. Two points to note is that regardless of the selected year, European nations are prevalent in the top ranked countries, and that Canada used to be ranked higher (1st in 2013) in the past. The pie chart visualizes the split of the happiness score according to the contributing factors in 2021, where GDP per capita and social support are the top two contributors amounting to 62% in total. The histogram indicates the distribution of Happiness Score to provide further insights.

1. *Happiness Score vs GDP per Capita*

Chart, bubble chart

Description automatically generated

As noted earlier, GDP per capita or wealth plays a vital role in people’s happiness. This visualization explores the relationship between the wealth and happiness. The size of the bubble represents the GDP per capita of a country and the color indicates the happiness level. The countries are sorted in order of happiness score, so the happiest countries are concentrated in the center. As the visualization demonstrates, the countries in the middle are of larger bubble size indicating that the happier countries are also richer. But there is also indication that after a certain point the wealth does not linearly contribute to the happiness level as the happiest countries such as Denmark and Finland are not the richest. There are many other richer countries with lower happiness score. However, there are no dark blue *and* small sized bubble meaning that all the poorer countries are not as happy, so wealth of a country does play a significant role in bringing the country’s happiness level to a certain level.

1. *Case Study – Canada.*

Chart

Description automatically generated

Last story point looks into where Canada is and how it has performed in terms of life expectancy and happiness score. The top two bar charts indicate how Canada is well positioned in terms of life expectancy. Canadians are the longest living among nations in Americas region, averaging at 81 years which is fairly big difference compared with the United States which has an average life expectancy of 78 years. The trend line also shows improvement in life expectancy over the years accompanied by declining infant mortality rate. However, Canada’s happiness score has been on a declining trend. Although still relatively high at 10th position in 2021, there is a clear indication of a downward trend. One interesting observation looking at the change over the years in the contributing factors to its happiness score is the social support factor which showed a sharp uplift in 2021. Social support is the national average of the binary responses (0=no, 1=yes) to the question “If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?”, and the increase in significance of this factor could be a trend caused by COVID that people are putting more emphasis on its community and relationships to get through the pandemic.

## Final Note

The analysis demonstrates how the happiness level of a country is closely related to the life expectancy. The high correlation between life expectancy and happiness score does not indicate direct cause of the other but as it can be seen from how the life expectancy is driven by factors such as the vaccination rates, hygienic water supply, government expenditure in health care and infant mortality rate, that depends heavily on national wealth to implement the social security and infrastructure required, it can be inferred that wealthier a country, the longer the life expectancy, and happier the people. And the analysis highlights how Africa is left behind both in terms of life expectancy and happiness score. But there is also indication that the happiness score and GDP per capita is not a linear relationship but of a logarithmic one where after a certain point, the magnitude of the contribution of wealth to happiness starts to decline and there are more complexities involved in the relationship between life expectancy, wealth, and happiness score which requires further study. Lastly, one of the limitations of the analysis is the lack of more recent data for life expectancy, however the same time span is used when comparing with the happiness index and the relationship is unlikely to have changed in the past few years which makes it sufficient to evaluate the trend and pattern for the purpose of the project.

## APPENDIX

## Visualizations

1. *Global Life Expectancy Map for 2016*

The life expectancy in year 2016 is depicted using the filled map visualization tool. The colors indicate the life expectancy years for each country. Dark blue countries have higher life expectancy and those in red have lower life expectancy. This clearly indicates where the countries with low life expectancies are located, and that they are mostly in Africa region. Red color was chosen to indicate low life expectancy to intuitively signal that Africa is left behind and that this is where the problem lies in terms of life expectancy.

Map

Description automatically generated

1. *Happiness Score vs Life Expectancy*

This visualization plots the Happiness Scores in the y axis and the life expectancies in the x axis to see the correlation between the two factors. As you can see, as life expectancy increases so does the happiness score. This indicates that there is a strong correlation between happiness and life expectancy. The data points are colored according to the region, and you can see that lower left corner is mostly red which are African nations. This again indicates that most African nations have both low life expectancy and low happiness score.

Chart, scatter chart

Description automatically generated

1. *Access to Drinking Water*

This packed bubble visualization looks into the relationship between life expectancy and the drinking water supply for every nation. As with the filled map visualization the color indicates the life expectancy, red being low expectancy and blue meaning higher life expectancy. The size of the bubble represents the magnitude of basic drinking water supply available in each country. Smaller bubble signify that the basic water supply is limited for that country. This visualization neatly shows that the most outer layer bubbles are small and red, meaning that those countries with lower life expectancy tend to have lower supply of basic drinking water as well. This was another attempt to investigate what factors impact the life expectancy.

Chart, bubble chart

Description automatically generated

1. *Contributing Factors to Happiness Score*

This visualization demonstrates the split of factors contributing to the Happiness Score index. The pie chart is based on global average which indicates that the GDP per capita makes the greatest contribution to the Happiness Score at 38%, followed by social support at 24%. Having the sense of freedom to make life choices and healthy life expectancy have also been key factors in determining the sense of happiness in any country. The “healthy life expectancy” is slightly different measure compared with the life expectancy discussed in the first visualization that the former focuses on the years a person remains fit to perform day to day activities rather than purely in biological sense of mortality.

Chart, pie chart

Description automatically generated

## Trend Line

*Trendline 1 – Canada’s Happiness Score since 2005.*This chart represents the historical Happiness Score of Canada from 2005 to 2001. First data is filtered to only display that of Canada. Although there have been fluctuations, the Happiness Score index is on declining trend as indicated by the trend line.

Chart, line chart

Description automatically generated

*Trendline-2 Happiness Factors Trend in Canada.* The line graphs represent the change in the six factors that explains the Happiness Score of Canada since 2014. The average score of the six factors is included as a reference line and the three lines that are above the average are GDP per capita, Social Support and Healthy Life Expectancy. This is in line with the visualization performed in Assignment 1 (pie chart - 4. Contributing Factors to Happiness Score) which demonstrated that the top three contributors are those three listed above. However, the only contributor that indicates an upward trend is the GDP per capita and remaining factors are on decreasing trend.

Chart, line chart

Description automatically generated

*Trendline3-GDP for Canada since 2005.* This chart depicts Canada’s GDP per capita growth since 2005 and the trend line shows that it is on increasing trend explaining why the GDP factor of the Happiness Score has been on an upward trend as well (refer to Trendline 2). All the trendline charts have used line charts as time series data can be well represented using this format.

Chart, line chart

Description automatically generated

## Sort and Filter – Life Expectancy in Americas (2016)

The horizontal bar chart depicts the life expectancy of countries in the Americas in 2016. The data has been first filtered by region to only show the countries in Americas and then sorted by descending order of life expectancy field value to show the ranking within the region. The horizontal bar chart clearly indicates that Canada ranks first in this region at 82.8 years of life expectancy. United States lags, coming 5th with life expectancy of 78.5 years. Canada and US have been modified to purple and yellow colors to highlight these two countries for comparison. Horizontal bar chart was chosen for this visualization since rankings are often represented using this format.

Chart

Description automatically generated

## Calculated Field – Global Vaccination Rate from 2000 to 2016

The line chart data is the median of the vaccination rates (the percentage of vaccinated children under the age of one) for the three major diseases namely, polio, measles, and diphtheria. This calculated field represents the median vaccination rate for these three vaccines which are amongst the most major disease for children. The vaccination rates are aggregated, and then median is plotted for the years from 2000 to 2016. You can see how the median vaccination rate has been increasing in all the regions which greatly contributed to the increase of overall life expectancy. Median rather than average is used for the calculation in order to eliminate the impact of outliers. Again, this is a time series data and line chart are used to visualize this data to show the trend over the years.

Chart, line chart

Description automatically generated

## Table Calculation – Percentage Different in Life Expectancy year on year

The table calculation used is the percentage difference in life expectancy compared with previous year and I have filtered the countries to display only those that experienced drastic changes. For each country, the year that had a drastic negative change was cross checked with any events that might have affected the population and as you can see, most negative changes in life expectancy are coming from the impact of natural disasters or war. As these are one off events, the life expectancy is back to the normal level the following year, bumping up the percentage difference figure. The line chart is used to visualize this data to depict changes over time and comments were inserted to indicate the event that has impacted the drop.

Chart, line chart

Description automatically generated

## Distribution

*Happiness Score vs Life Expectancy* – the distribution bands of -1 to 1 standard deviations have been applied for both y axis (happiness score) and x axis (life expectancy) to visualize the dispersion of the data.

Chart, scatter chart

Description automatically generated

*Infant Mortality per Region* – Box and whisker plot is used to visualize the distribution of data across different regions. The infant mortality rate refers to the probability of dying between birth and exactly 1 year of age, expressed per 1,000 live births. This plot illustrates how Americas and Europe have the least spread and Africa and Middle Eastern regions have the most spread in infant mortality rate. The average among the region is highest in Africa.

Chart, box and whisker chart

Description automatically generated

*Obesity vs Thinness* – This is a bullet graph where the orange bars indicate the obesity between ages 5 to 19, with added distribution and reference line marking the age 5-19 thinness. The bullet graph is used to make comparison across these variables. The countries are sorted in ascending order of life expectancy, and you can visually see how the obesity increases and thinness decreases as you scroll down the graph.

Graphical user interface, application

Description automatically generated

*Distribution of Global Life Expectancy* – This histogram illustrates the shape of the global distribution of life expectancy and the color encoding indicates the government spending on public health. It illustrates how the largest number of countries fit into the 70-72 years bin and that the countries in the higher life expectancy bins have higher government health expenditure.

Chart, bar chart

Description automatically generated

## Dashboard

*Global Comparison – Life Expectancy and Happiness Score.*This dashboard is to visualize the correlation between life expectancy and happiness score. As you can see, the color patterns of the two color filled maps are quite similar. The red countries are concentrated in the African region, the dark blue in the OECD countries. As shown in the color legend, red indicates low life expectancy or low happiness score and blue indicates higher life expectancy or higher happiness score, meaning that those countries with lower life expectancy tend to have lower happiness score and vice versa. The scatter plot with happiness score in the y axis and life expectancy in the x axis similarly indicate high correlation as seen in the upward linear trend where countries with higher life expectancy tend to have higher happiness score. The filters enable the user to filter according to the region as well as by country and clicking on the icons enable the user to access the data sources.

Chart, scatter chart

Description automatically generated

*Life Expectancy – Global Trend.* The second dashboard looks closer into the life expectancy data visualizing the trend over the years across regions. The line charts on top depict the trend of improving vaccination rates which coincides with the trend of increasing life expectancy. This is a common trend across all regions although few dips in life expectancy is detected in some cases. These dips are usually as a result of natural disasters as seen in the chart at the bottom which illustrates the percentage difference of life expectancy year on year. This chart utilizes table calculation and the countries with sizable percentage change have been filtered.

Chart, line chart

Description automatically generated

*Life Expectancy – Access to Drinking Water.* This dashboard looks into the relationship between access to drinking water and life expectancy. The packed bubbles visualization indicates the importance of having access to hygiene water supply in relation to longer life expectancy as countries concentrated in the center are all dark blue with larger bubble size, indicating high life expectancy and high access to drinking water. The outer layer countries have low life expectancy (red) and low access to drinking water indicated by smaller size of the bubbles. Lastly, a web page which is a study conducted on the relationship between drinking water access and mortality rate is included as a reference.

A picture containing graphical user interface

Description automatically generated

*Life Expectancy – Distribution.* This dashboard looks further into other factors such as obesity, thinness, and government spending on health care in relation to life expectancy, but with a focus on distribution. The two box plots on the top row illustrates the prevalence of obesity and thinness in childhood. The orange bars indicate the obesity level of each country between age of 5 – 19 and the reference line indicates the thinness for the same age group. The left chart is the bottom 10 countries, and the right is the top 10 countries in terms of life expectancy. For the bottom 10 chart, the thinness is prevalent and exceeding obesity rate in all countries, whereas for the top 10 countries obesity is prevalent exceeding the thinness rate. This is indicative of the trend that as countries get wealthier, the life expectancy increases but so does the obesity rate as obesity is a problem in developed countries where there is enough wealth to feed, or often overfeed the population.

On the bottom row, the histogram is a visualization of global distribution of life expectancy with added feature of government health expenditure using the color encoding. The countries falling under the bin with higher life expectancy have higher government health expenditure. Indicating the vital role, the government plays in improving general health of the public and therefore leading to longer life span. Lastly the box-and-whisker plot shows the dispersion and the range of infant mortality rate for each region. Africa has the highest infant mortality rate which is likely to be impacting the lower life expectancy of the region. The filters provided enable the user to look at the data at a more granular level according to the region or the year.

Chart, bar chart

Description automatically generated

*World Happiness.* This dashboard looks deeper into the Happiness Score data. The Top N Happy Nations in Year <selected year> displays the global happiness score ranking. The color indicates the region and ranking can be adjusted using the Top N parameter as well as the year which can be selected using the single value slider filter. Two points to note is that regardless of the selected year, European nations are prevalent in the top ranked countries, and how Canada used to be ranked higher (1st in 2013) in the past. The pie chart visualizes the split of the happiness score according to the contributing factors in 2021, and the histogram indicates the distribution of Happiness Score to provide further insights.

Chart, application, bar chart, pie chart

Description automatically generated

*Case Study – Canada.* This dashboard is looking into one particular country, Canada. Canada specific data and visualizations are chosen to see Canada’s trends in life expectancy and happiness score. Canadians have the longest life expectancy in Americas region. And looking at Canada alone, the life expectancy has constantly been increasing and the infant mortality rate has been decreasing. However, the happiness score is showing trend of gradual decline. Looking further into the factors contributing to Canada’s happiness score, GDP per capita and social support are gaining more impact in the past few years. Social support is the national average of the binary responses (0=no, 1=yes) to the question “If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?”, which could be a trend caused by COVID that people are putting more emphasis on its community and relationships to get through the pandemic.

Chart

Description automatically generated with medium confidence