



WORLD HAPPINESS

Raymond Bell

George Tomaras

Shaunghnessy Robertson

Introduction:

Happiness is essential to who we are, it can be summed up as "the extent to which an individual views the overall quality of their life as a whole positively". A quote that stood out to us is "Happiness depends upon ourselves". Does our happiness depend on ourselves, or are there other things that may or may not be under our control that decide it?

Inspiration:

Our group's shared interest in how we evaluate happiness served as the source of inspiration for our project. The World Happiness Report describes how measurements of well-being can be used to measure a country's progress in a useful way. The studies examine the current state of happiness in the world and demonstrate how the personal and regional variations in happiness are explained by the new science of happiness.

Hypothesis:

Our team's objective is to determine the main causes of changes in country rankings or scores between the 2015 and 2019 reports and why changes occur annually. This led to the Null hypothesis that there is no difference between the extent to which each of the key factors impacts the happiness score. The alternative hypothesis is that there is a difference between the extent to which each of these factors impacts happiness score.

Data:

Our research was done using data from Kaggle from The World Happiness Report, a landmark survey of the state of global happiness that ranks 156 countries by how happy their citizens perceive themselves to be. The study provides five separate CSV files with data based on five different factors including Economy (GDP per Capita), Health (Life Expectancy), Generosity, Social Support, and Freedom. The total Happiness Score for each country was calculated by adding together each of these factors.

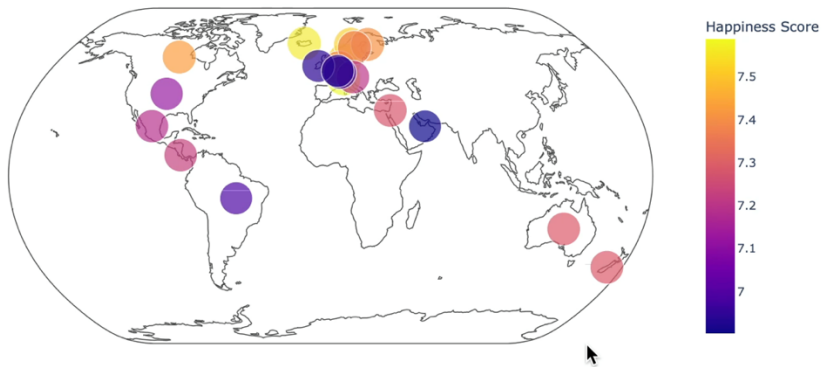
Data Cleaning:

Our data needed to be cleaned and organized before we could start our research. To compare the various column names, we initially imported the CSV data from each year. For consistency, we then had to rename various columns for each year's data frame. Additionally, we had to add a column to each data frame for the year. Afterward, we were able to merge all six data frames into one data frame.

Happiness Score:

Are certain countries happier?

Happiness Score for Top 20 Countries

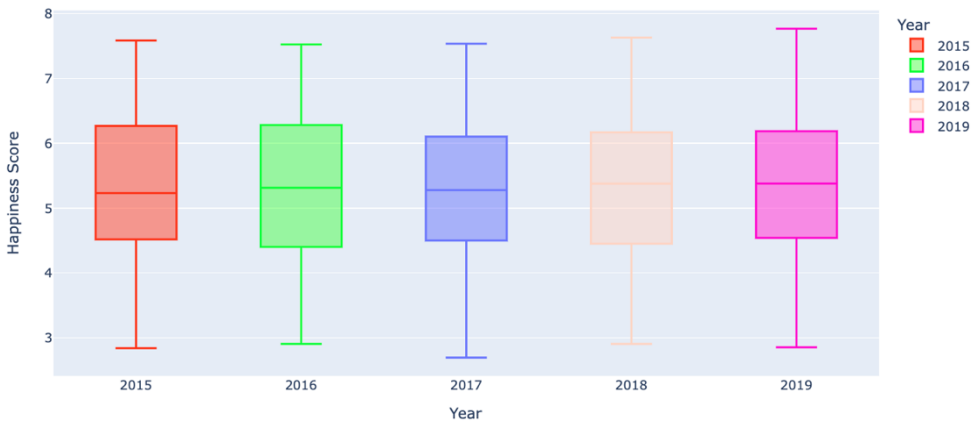


The above map illustrates the top 20 countries with the highest happiness score from 2015 – 2019. Switzerland and Iceland ranked the highest with scores between 7.561 and 7.587.

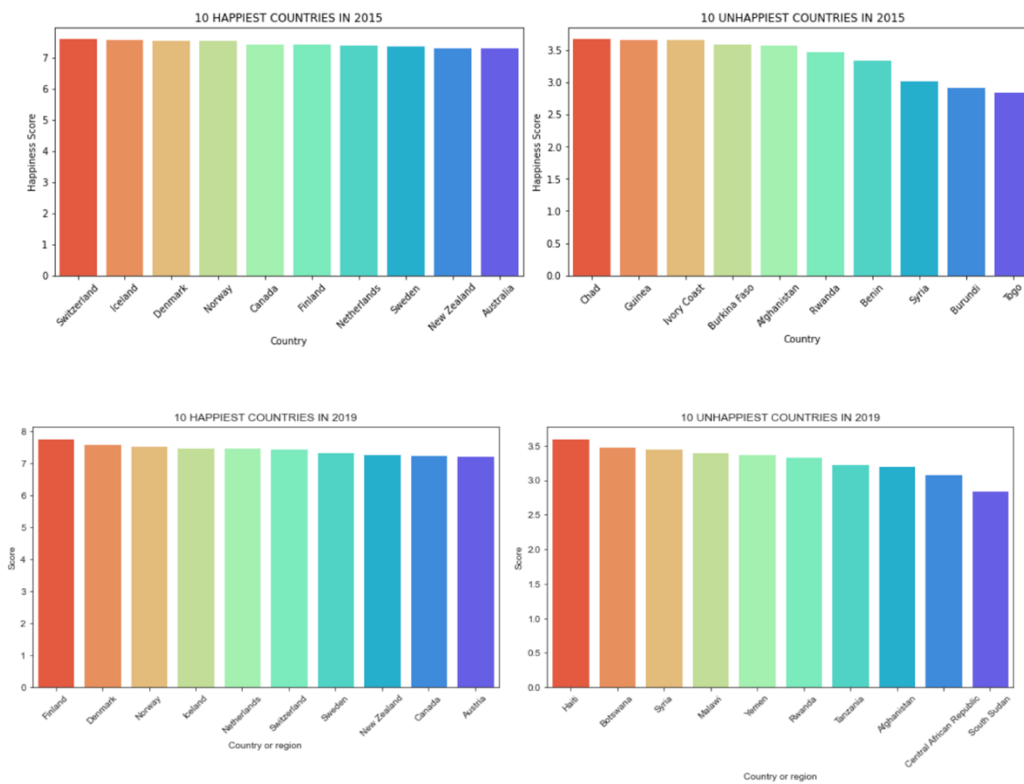


The above histogram illustrates a normal distribution of the happiness score across the years 2015 – 2019.

Happiness Score From 2015 To 2019



The distribution between the five years has not altered significantly, as shown by the boxplots above. We can observe that the median happiness score for 2019 is greater than it was from 2015 to 2018. Also, the values for the interquartile range for the year 2017 are slightly shorter than those for the other years.



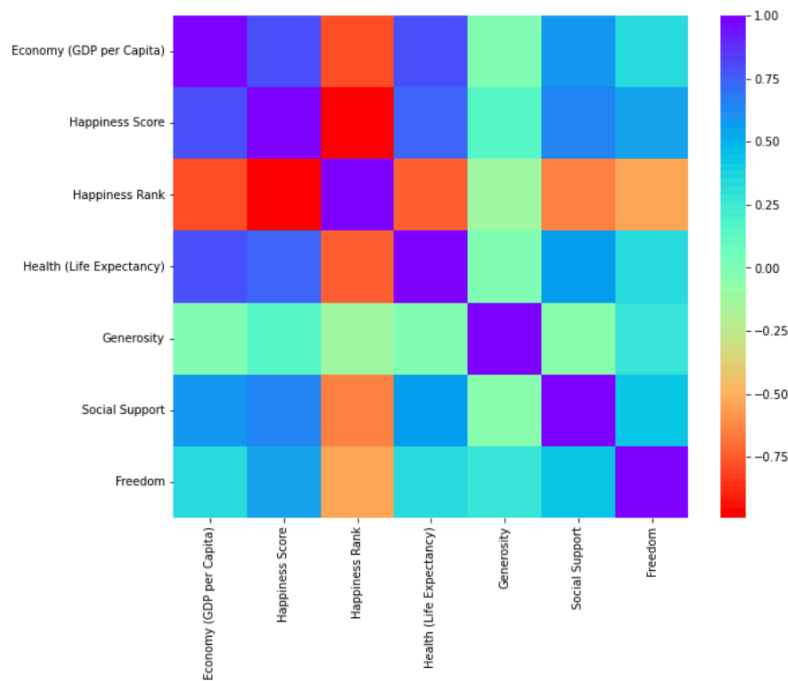
We decided to go a little deeper and look at the significant differences between countries that repeatedly appeared at the top and bottom based on their Happiness Score. The bar graphs above show a comparison of the top 10 and bottom 10

happiness rankings between 2015 and 2019. The top 3 countries in 2015 are Switzerland, Iceland, and Denmark and the bottom 3 are Syria, Burundi, and Togo, however, those aren't the same top 3 countries in 2019. Finland, Denmark, and Norway are the top 3 countries in 2019 while Afghanistan, the Central African Republic, and South Sudan are the bottom 3.

I want to point out how Switzerland was #1 in 2015 and dropped to #6 in 2019. That could be because of a few different major events that took place in Switzerland like The Jura conflict flaring up again in November of 2018. After a 17-month investigation to decide if the town of Moutier should join the town of Jura the government of Bern decided to declare the result of the vote invalid, which could have had an impact on happiness. Also, health insurance costs continued to increase in 2019, and the cost of healthcare increased by 4.2% each year. Since the averages are so close, minor adjustments may cause the rankings to change from year to year.

Correlation:

The correlations are plotted as a color-coded heatmap with a reference scale for easy reading. This led to the following plot:

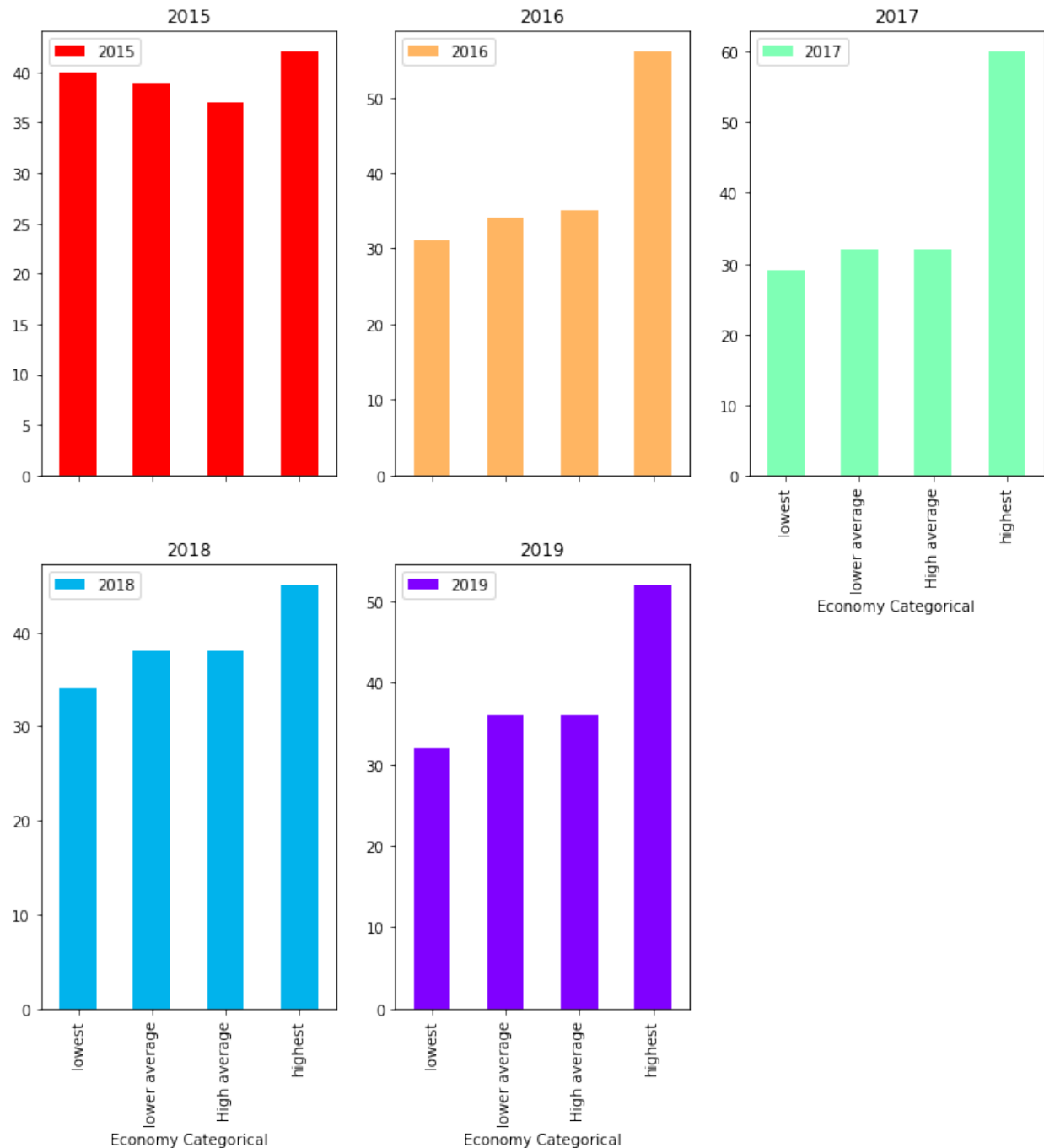


The scale on the right shows that darker plots between the features indicate a significant linear correlation. Positive linear correlation is represented by purple, whereas negative linear correlation is represented by red. Surprisingly, there is no negative correlation between any of the key features. This suggests that raising one predictor variable does not cause a decrease in the other contributing factors.

Generosity has the weakest to no correlation with the other variables, whereas GDP, social support, and health also have medium associations with them.

Economy (GDP per Capita):

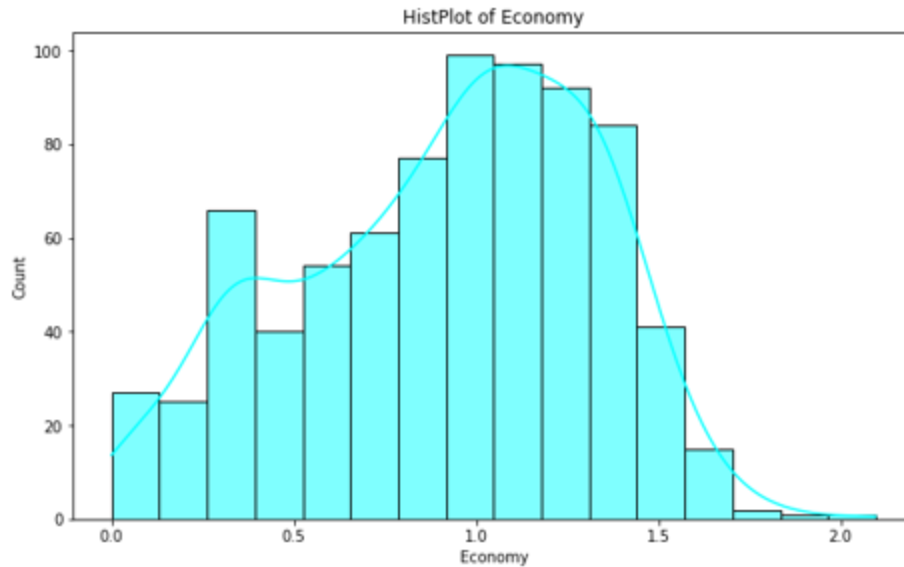
Are larger, more developed countries happier?



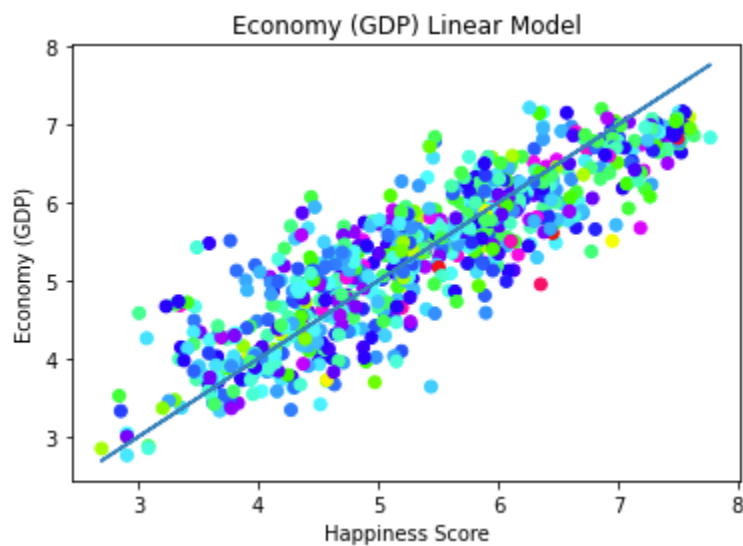
The above graphs illustrate the average economic scores between the years 2015 - 2019. As you can see, in comparison to the other years, 2015 had the highest averages across all categories followed by a sharp decline in the results in 2016. A global economic impact occurred in 2016. A sell-off in China's stock market of roughly 7% occurred between January 4 and January 7, which swiftly sent markets down all across the world. China's stock market then declined by around 18% between January 4 and January 15. Shockwaves from the Chinese stock market collapse struck US and European markets, among others. It was not until 2018 that we began to see an increase in the averages. Today, China plays a large role in how the global economy runs due to the swings in international financial markets brought on by the collapse of the Chinese stock market.



Visually, we can see the positive correlation between the scatterplots above.



The range of GDP per capita is 2.0, with 0.0 as the minimum and 2.0 as the maximum. Intuitively, countries with stronger economies have lower inflation and more ease for their residents. Higher living standards in nations with stronger economies may increase happiness among their population.





We decided to investigate the variable Economy in more detail to determine how strong the relationships are. We first created a scatter plot for the variable related to the happiness score. In the scatterplot above, visually, we can observe the positive correlations between them.

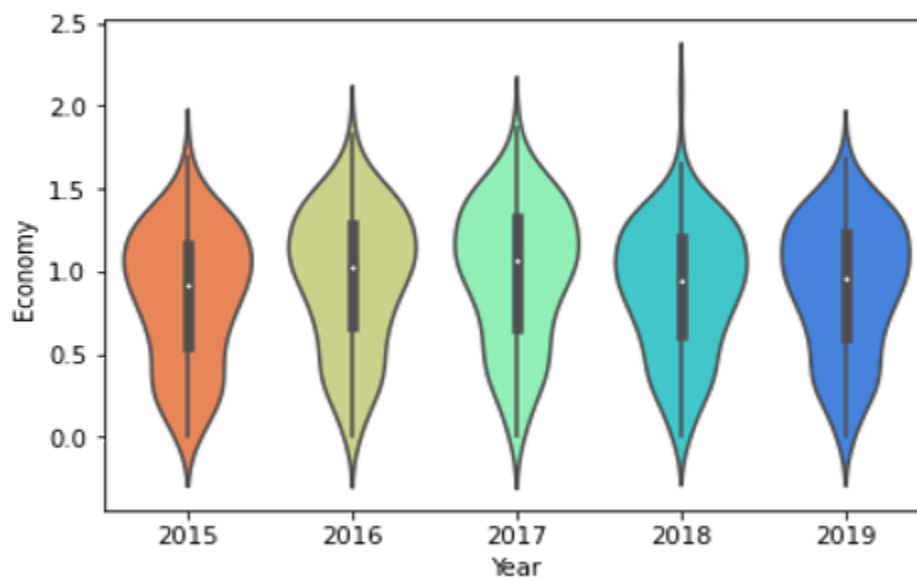
We ran a stats model on all of the variables from the World Happiness Report vs Happiness Score. The table below shows that we were given an Adjusted R-squared of 0.758, as a result.

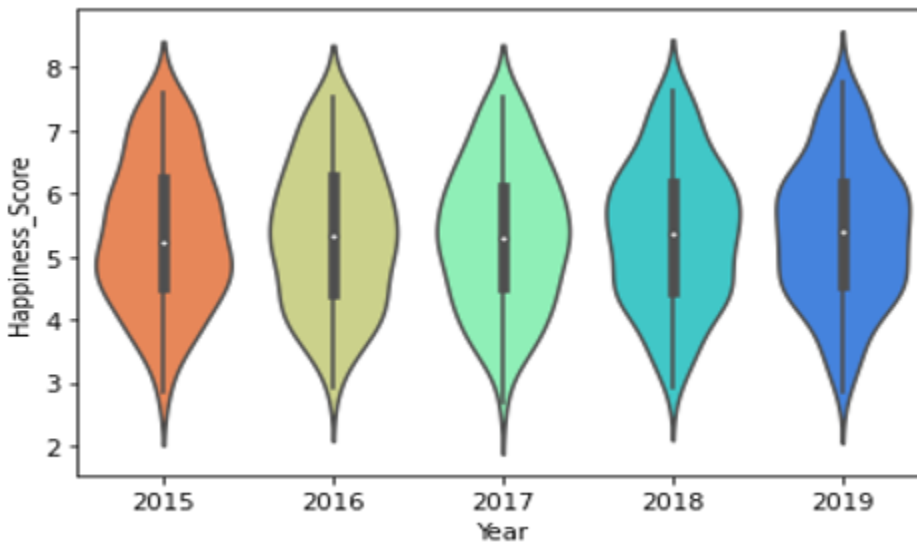
Dep. Variable:	Happiness Score	R-squared:	0.760			
Model:	OLS	Adj. R-squared:	0.758			
Method:	Least Squares	F-statistic:	490.4			
Date:	Sun, 13 Nov 2022	Prob (F-statistic):	2.48e-237			
Time:	22:01:53	Log-Likelihood:	-645.59			
No. Observations:	782	AIC:	1303.			
Df Residuals:	776	BIC:	1331.			
Df Model:	5					
Covariance Type:	nonrobust					
	coef	std err	t	P> t 	[0.025	0.975]
	const	2.1577	0.080	26.857	0.000	2.000 2.315
Economy (GDP per Capita)	1.2113	0.082	14.822	0.000	1.051	1.372
Health (Life Expectancy)	1.0053	0.133	7.584	0.000	0.745	1.265
Generosity	0.7442	0.173	4.310	0.000	0.405	1.083
Social Support	0.5862	0.080	7.326	0.000	0.429	0.743
Freedom	1.7057	0.154	11.106	0.000	1.404	2.007
Omnibus:	11.325	Durbin-Watson:	1.460			
Prob(Omnibus):	0.003	Jarque-Bera (JB):	11.942			
Skew:	-0.246	Prob(JB):	0.00255			
Kurtosis:	3.354	Cond. No.	18.8			

Next, we ran stats models to determine how strong the correlations were with economy being the target variable. In the stats model below, we got an adjusted r-squared of 0.0622.

Dep. Variable:	Economy	R-squared:	0.623			
Model:	OLS	Adj. R-squared:	0.622			
Method:	Least Squares	F-statistic:	1289.			
Date:	Sun, 13 Nov 2022	Prob (F-statistic):	2.21e-167			
Time:	21:24:56	Log-Likelihood:	-25.398			
No. Observations:	782	AIC:	54.80			
Df Residuals:	780	BIC:	64.12			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	-0.6178	0.044	-14.153	0.000	-0.704	-0.532
Happiness_Score	0.2852	0.008	35.900	0.000	0.270	0.301
Omnibus:	2.721	Durbin-Watson:	1.932			
Prob(Omnibus):	0.257	Jarque-Bera (JB):	2.557			
Skew:	-0.129	Prob(JB):	0.278			
Kurtosis:	3.107	Cond. No.	27.7			

T-Tests





The results for the above t-tests are statistic=-0.7276658661569179, pvalue=0.4788025567726826.

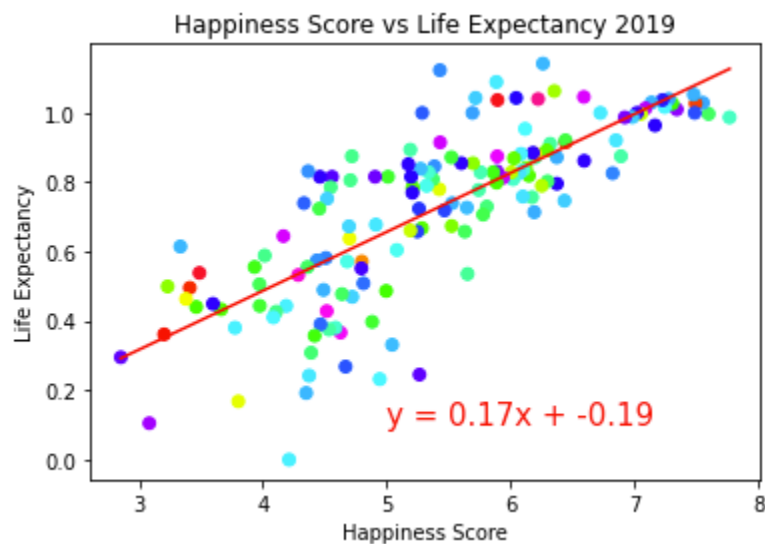
Health Life (Expectancy)

Are countries with longer life expectancies happier?

Since 2013, more global interest has exploded surrounding “well-being”, “quality of life”, “life satisfaction”, and other similar topics. As a result, countless journal articles and publications have been written to give a broader definition and deeper meaning to the subject matter. Some of these publications have even criticized the Happiness Score, which has been around since 2012, arguing that economy and health (life expectancy) introduces a degree of bias into the global ranking system. However, these two of five factors are not only factors with significant correlation coefficients, but also plausible quantifiable measures.



The scatter plots for 2015 above and 2019 below both shows positive correlations between Happiness Score and Health (Life Expectancy).



In our study, we first examined the degree of influence that Health (Life Expectancy) had on Happiness Score as a feature in our model. However, we went further to create a model making Health (Life Expectancy) the target to determine the correlation of the other features. The output from this OLS Regression model is shown below.

OLS Regression Results

Dep. Variable:	Health (Life Expectancy)	R-squared:	0.662
Model:	OLS	Adj. R-squared:	0.659
Method:	Least Squares	F-statistic:	303.3
Date:	Sun, 13 Nov 2022	Prob (F-statistic):	8.97e-180
Time:	20:36:12	Log-Likelihood:	403.85
No. Observations:	782	AIC:	-795.7
Df Residuals:	776	BIC:	-767.7
Df Model:	5		
Covariance Type:	nonrobust		

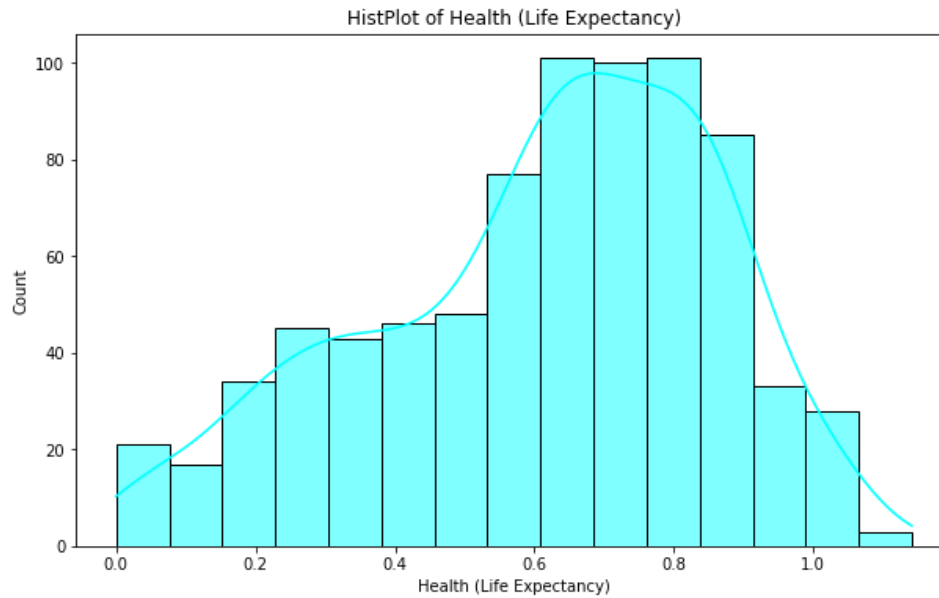
	coef	std err	t	P> t	[0.025	0.975]
const	-0.0815	0.029	-2.808	0.005	-0.138	-0.025
Economy (GDP per Capita)	0.3015	0.022	13.936	0.000	0.259	0.344
Happiness Score	0.0687	0.009	7.584	0.000	0.051	0.086
Generosity	-0.0221	0.046	-0.484	0.629	-0.112	0.068
Social Support	0.0720	0.021	3.353	0.001	0.030	0.114
Freedom	-0.0593	0.043	-1.374	0.170	-0.144	0.025

Omnibus:	42.548	Durbin-Watson:	1.533
Prob(Omnibus):	0.000	Jarque-Bera (JB):	51.880
Skew:	-0.522	Prob(JB):	5.42e-12
Kurtosis:	3.708	Cond. No.	56.5

The histogram depicts left skew in our data concerning health (life expectancy). This table to the left also captures the negative or left skew of -0.522 . This left skew is due to some of the African countries with health (life expectancy) indexes at or below zero. Depending on the number and magnitude of such outliers, the mean will be less than the median measurement, as the average is pulled downward. Although the boxplot for 2019 detects an extremely low outlier, the overall mean is higher than the other years. This is because the health (life expectancy) for all countries, even the bottom 10, increases relative to the previous three-year decline, except for Swaziland.

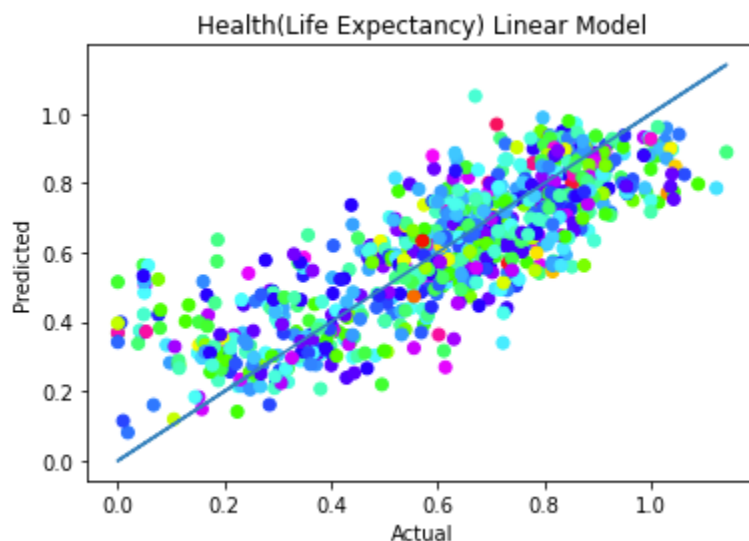
The histogram below illustrates this left skew and kurtosis with a distribution that could be considered not only non-normal but bi-modal as well. One of the implications of a bi-modal distribution is the potential presence of two distinct populations. This makes sense as the developing countries vis-a-vis the developed countries of the world are “worlds apart”. Because of this, it would be statistically appropriate to deploy non-parametric testing on this particular slice of our data set.

However, we didn’t conduct such a test because Happiness Score is our target and primary objective rather than Health (Life Expectancy). Here, we simply wanted to examine the outcome when we made Health (Life Expectancy) our target while declaring the other five factors as features.

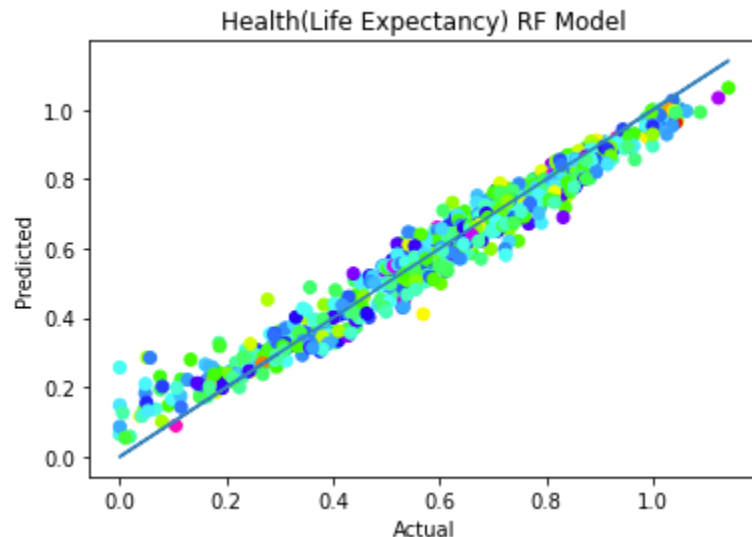


This branch of our analysis shows that Economy (GDP per Capita) has a correlation coefficient of 0.305, the highest of the given features. The heatmap above also shows the strongest correlation with Economy (GDP per Capita) when using Health (Life Expectancy) as the index or point of reference. Here we derive the r-squared value of 0.662.

The scatterplot shows a pattern that is fairly evenly distributed between .4 and .95. However, there is an indication of overprediction below .4 and underprediction above .95. Although the data fails to translate these indices in terms of years of life, the model suggests that countries with lower health (life expectancy) saw shorter life spans than predicted and vice-versa.



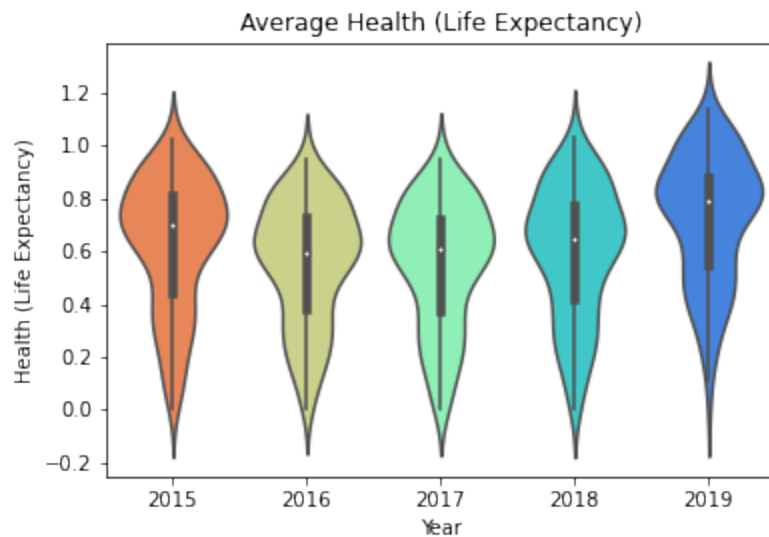
As expected, the scatter plot below shows a better fit, indicating a better prediction model. Notice that we can still observe the same patterns occurring at the same points along the x and y-axis.



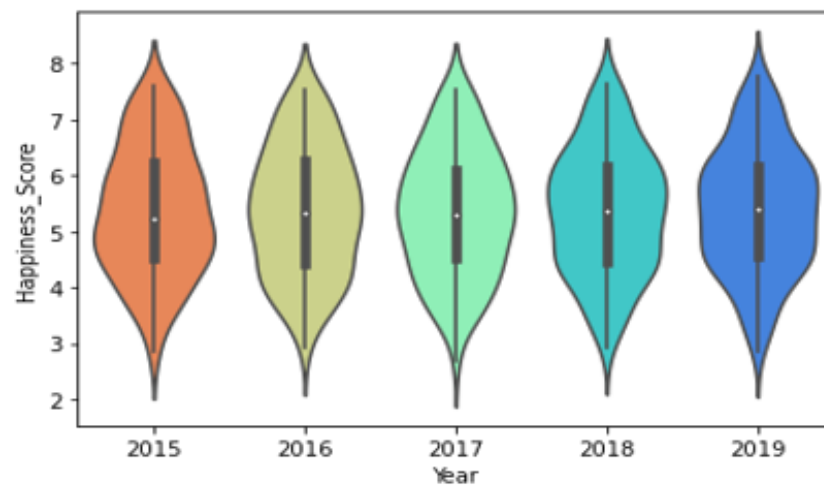
T-Tests

Before conducting a t-test on the top ten countries concerning Health (Life Expectancy) for 2015 and 2019, we checked to determine whether there was equal variation for those two years. The variation in the top ten for 2015 was 0.0005510992472727277, while the variation in the top ten for 2019 was 0.00040825454545454594. At this point, we set the variances to equal in the `stat.ttest` formula. The result yielded a t-statistic of -11.794638789113031, including a P-value of $1.8433429079763735e-10$. Given more time, we would repeat this process for countries within the bottom ten for Health (Life Expectancy).

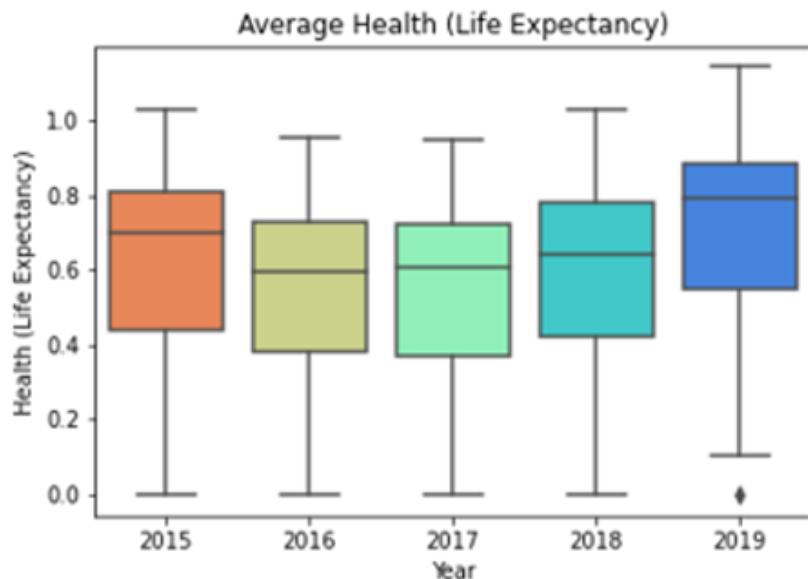
Our violin chart below suggests that it would also prove beneficial to examine mean differences in Health (Life Expectancy) between 2019 and 2016. Depending on the variation results found here, our next step would be to conduct an ANOVA to closer compare inter and intra-variation across multiple years.



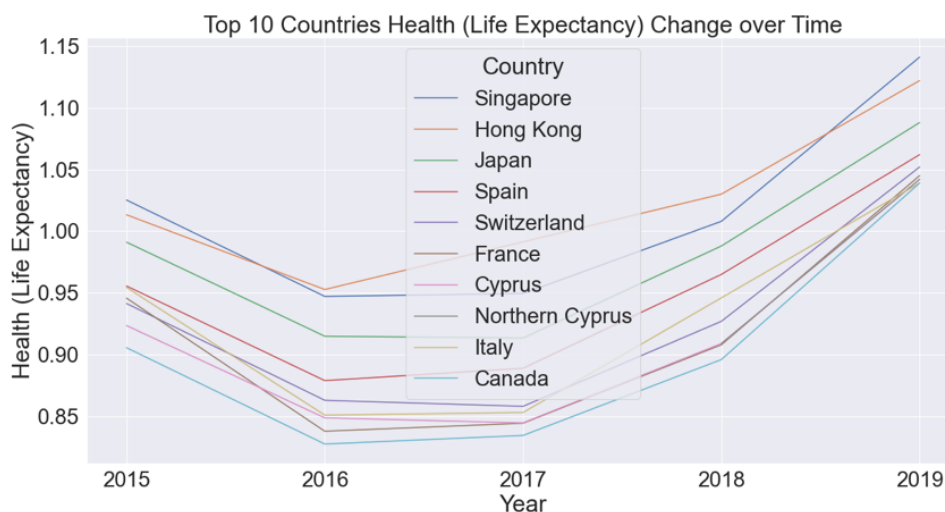
In this exploratory data analysis, we seek to either reject or fail to reject the H_0 that there are no differences between the extent to which these five factors impact happiness score. Thus far, we have shown that economy (GDP per Capita) and Health (Life Expectancy) has strong correlation coefficients and are weighted more in the model. However, here we see that happiness scores, on average, remain near constant across these five years. Health (Life Expectancy), on the other hand, shows a three-year decline during this period. This suggests that although there is a correlation, there is no case of causation concerning Health (Life Expectancy) and Happiness Scores.



Although the boxplot below shows an extreme outlier for 2019, it also suggests that our decision to evaluate 2015 and 2019 was sound, as those years had means that were substantially different from 2016, 2017, and 2018's top ten countries in terms of Health (Life Expectancy). We can observe from the boxplot below that the means for those three years are almost equal.



Change in Health (Life Expectancy) over time

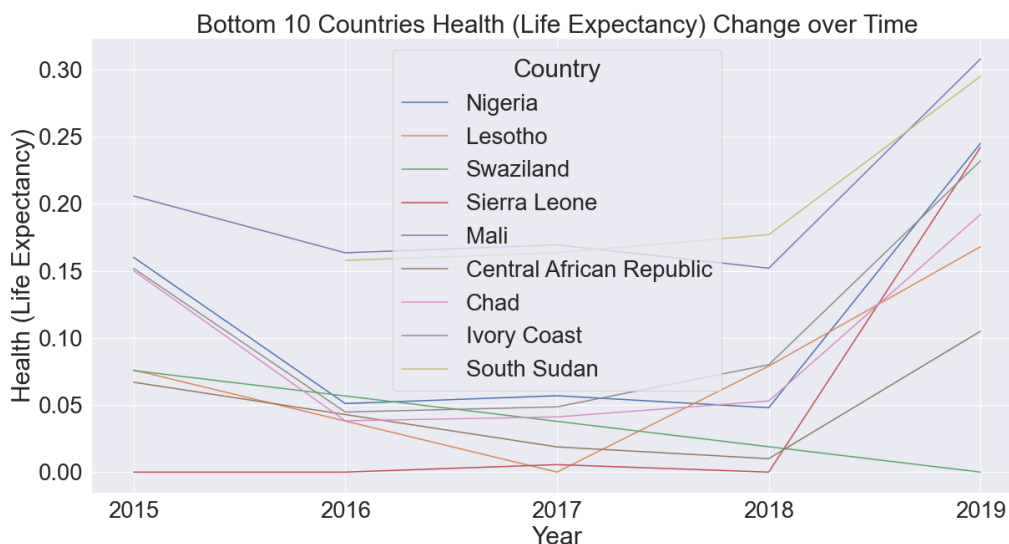


In the same way that we noticed the irony regarding a near-constant happiness score despite declines in health (life expectancy), we also notice a similar parallel concerning health (life expectancy) versus economy (GDP per Capita). For example, Singapore, a southeast Asian country, is listed consistently among the top ten countries across all five years. With a population of roughly 5.7 million and a global economic ranking of 38, Singapore is proof that health (life expectancy) is driven by more than just a robust economy.

Switzerland is another country that is cited among the top ten countries regarding health (life expectancy). With a global economic ranking of 20, Switzerland, like Singapore, is

proof that the link between economy and health (life expectancy) is less than straightforward. Many other factors, such as diet, lifestyle, work-life balance, etc., affect this measure.

It is reasonable to expect that countries with higher national incomes would also have higher life expectancies, but Switzerland and Singapore are examples of countries “punching above their weight”. Health care is the first recourse for resolving differences related to health (life expectancy), but socioeconomic and political determinants must be factored in to better understand these phenomena. These determinants range from education to wage rate policies. Determining which conditions or policies are responsible for this outcome is critical to improving population health globally (Freeman et al., 2020).



Somalia and Swaziland are two African countries that have been ranked among the bottom ten countries during this time. According to our data, Swaziland is the only country with a health (life expectancy) trending downward, as of 2019. This small country of 6,500 sq. Miles, bordered by Mozambique and South Africa, continue to combat a wide range of communicable and non-communicable diseases. At 27 percent, Swaziland, now known as Eswatini, has the highest prevalence of HIV/AIDS in the world. With roughly 60 percent of its population living below the poverty level, the health (life expectancy) of its citizens is 59 years. Diabetes, Tuberculosis, Stroke, and Malaria are diseases listed among the top ten causes of mortality, according to the World Health Organization (2016).

Somalia is stricken with poverty and suffers from poor infrastructure, malnutrition, poor sanitation, and unclear drinking water to name a few. In 2018, the life expectancy for

men in was 53, while it was 57 years for women. For the past two decades, Somalis citizens have been displaced due to warfare and civil unrest.

This exacerbates an already bad situation given the deaths attributed to cancer, cholera, infant and maternal mortality, and a host of other communicable diseases. Children 5 years old and younger suffer from diarrheal diseases, measles, and pneumonia. Data from UNICEF suggests that there is a 13% chance that a child will die before the age of five. Millions of dollars in humanitarian assistance and foreign aid are poured into this country. According to the US State Department, the United States provided over \$400M during Fiscal Year 2021 to the people of Somalia. The World Health Organization along with the U.N. Refugee Agency continue to supply humanitarian aid to this country, but corruption and governmental instability have been known problems for years.

Health (Life Expectancy) - Global Gaps

Despite scientific breakthroughs, disparities still exist within and between countries around the world, in terms of healthcare systems. Whether we look at tuberculosis in sub-Saharan Africa or Alzheimer's in Asia, an increasingly elderly population is challenging these systems in new ways (Amini et.al., 2019). Now, it is more about cost than capability. Rising costs continue to strain access to health care, even in developed countries. Cost is integral to the global healthcare ecosystem such that it drives to the core of the five segments. These five segments are demographic, financial, operational, innovation, and regulator. Moreover, these segments can either hike or reduce costs.

In 2014, the Economist Intelligence Unit (EIU) projected that healthcare would increase by 4.3% during 2015 – 2019, on average. Generally, costs have continued to climb. The idea that this increase is linked solely to administrative waste, soaring insurance premiums, and expensive care for chronic diseases is a fallacy. In many instances, these cost hikes are accounted for by research & development and progressive technological innovations aimed toward curing illnesses that are currently incurable.

Consider how wearable medical devices, telehealth, electronic records, mobile health clinics, and similar creations facilitate rapid remote diagnostics. These are game-changing inventions even in places like Africa and India where infrastructure may still lag behind top economies of the world. 80% of the antiretroviral medicines needed by the 7.6 million Africans living with HIV are imported because the sources of supply and requisite manufacturing capacity are lacking (London, 1978).

So, what else is happening to cope with this rise in costs? Many countries have implemented various versions of universal healthcare systems. These systems may be

public, private, or hybrid. This cost and competition-based marketplace, which seeks to lower the cost of healthcare, including prescription drugs, has produced lots of mergers and acquisitions. In 2016, the number of administrative programs in Mexico was slashed by 52 percent, and five of its medical institutions, responsible for the health services of 90 percent of its population were reduced to three.

Conclusion:

X Null Hypothesis: There is no difference between the extent to which each of these factors impacts happiness score: economy, health & life expectancy, generosity, social support, and freedom.

√ Alternative Hypothesis: There is a difference between the extent to which each of these factors impacts happiness score: economy, health & life expectancy, generosity, social support, and freedom

REFERENCES

Freeman, Toby, et al. "Why Do Some Countries Do Better or Worse in Life Expectancy Relative to Income? An Analysis of Brazil, Ethiopia, and the United States of America." *International Journal for Equity in Health*, vol. 19, no. 1, 2020, pp. 202–202, <https://doi.org/10.1186/s12939-020-01315-z>.

"Private Sector Rallies Around Transforming Healthcare in Africa." *New African* (London. 1978), no. 591, 2019, pp. 48–51.

Amini, Maedeh, et al. "Years Lived with Disability Due to Alzheimer's Disease and Other Dementias in Asian and North African Countries: A Trend Analysis." *Journal of Epidemiology and Global Health*, vol. 9, no. 1, 2019, pp. 29–35, <https://doi.org/10.2991/jegh.k.190305.002>.

"World Happiness Report | Kaggle." [Online]. <https://www.kaggle.com/datasets/unsdsn/world-happiness>. [Accessed: 23-Nov-2022].