World Happiness Report

Happyr (Group 2): Cavan Ingram, Claire Gloss, Dominic Scerbo, Lauren O'Donnell, & Qian Shen
May 4, 2022

Abstract

Happiness scores have been determined by country based on the Gallup World Poll for the past eight years. To better understand these happiness scores, we investigated what contributes to the happiness score. We analyzed the countries and regions with the best and worst happiness scores, and specifically analyzed the United States's ratings to determine how it compared to other nations. We also investigated the relationships between the contributing factors to the happiness score: economy, health, freedom, family, trust in the government, and generosity. We tested the data for a possible linear relationship between the variables and happiness score. In our tests and observations, we determined the happiest nations resided in Northern Europe and the least happy nations reside in sub-Saharan Africa. We also found that the United States is consistently in the top 15 happiest nations, and is increasingly happy with each passing year. Overall, we found economy, health, freedom, and family to have a high correlation with happiness score and fit a regression equation to predict the happiness scores.

Data Introduction

Data Description

The data used for our analysis of Country Happiness was obtained from the Gallup World Poll from 2015 to 2022. The Happiness scores were based on answers to the Cantril Ladder question, which asked individuals to rate their life on scale from 0 to 10. Data was polled early in the reporting year.

The survey provided a list of variables that could factor into overall happiness. These factors were ranked by the individuals surveyed from most the most influential to least in relation to their overall happiness. The factors included economy, trust, freedom, family, and generosity. These factors did not contribute to the overall happiness score, but were utilized by the survey as potential sources of explanation for country happiness. The dataset also included a dystopia residual, which accounts for any over/under explained averages.

The data ranked countries based on their happiness score (highest score being 1) for each year and divided by region of the world.

We hoped to investigate this data to determine what factors contributed most to overall country happiness. We further divided our question of overall happiness into four additional questions as analysis progressed:

- What countries or regions globally have the highest happiness scores?

- What does the US happiness score look like and how is it compared to other countries?
- What features in the data contribute to happiness the most?

Are the determined features a good predictor of happiness?

Data Processing & Methodology

The data was originally presented in eight separate data files; one file for each year observed. The data were processed to account for missing values. Happiness rank was computed via the happiness score. Missing region values were completed using a region dictionary and a year column was added to each year's data set.

The column names for each data file were standardized along with the data format for each column before appending all eight files into one dataset.

Results

Analysis of World Regions

We first investigated happiness scores across the globe. We created a heat map, Figure 1, that showed the average happiness scores by country. This revealed the happiest countries are Canada, Australia, and Northern European countries. It also showed the least happy nations were generally located in Africa and South Asia.

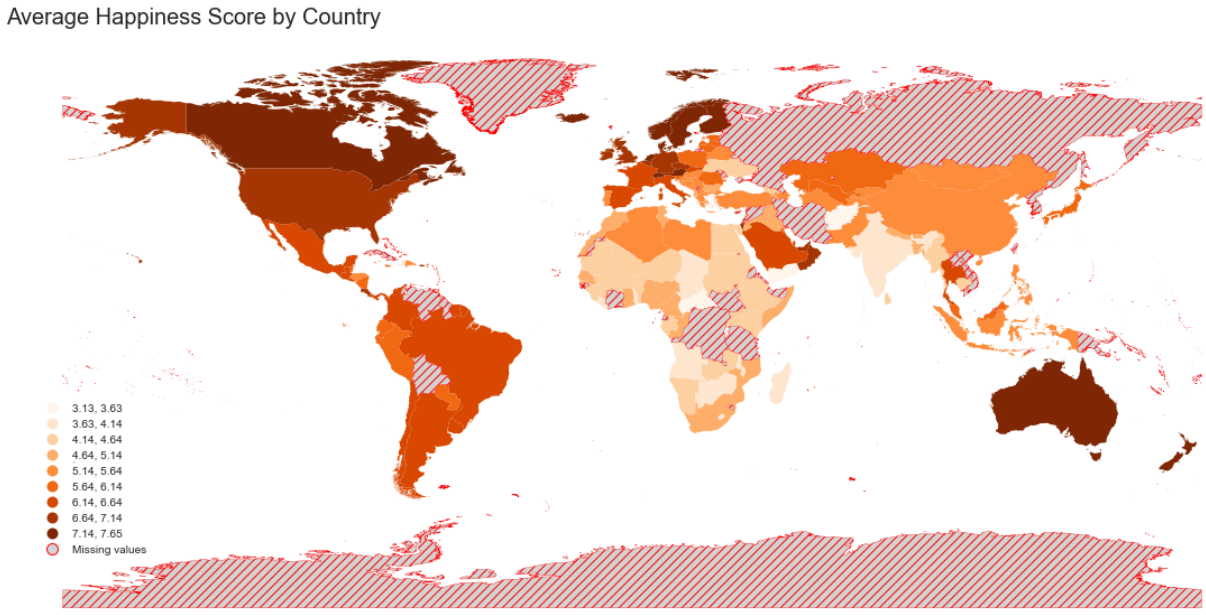

Figure 1

To identify the extremes of the happiness scores, we looked at which countries were ranked the top and bottom five countries. Consistent with the map above, Northern European countries were consistently in the top five happiest countries. Similarly, Sub-Saharan African nations

frequented the chart in the bottom five. Figure 2 shows the five happiest and five least happy countries by year. Unique nations were highlighted to better visualize the data's consistency over the years.

Rank	2015	2016	2017	2018	2019	2020	2021	2022
1	Switzerland	Denmark	Norway	Finland	Finland	Finland	Finland	Finland
2	Iceland	Switzerland	Denmark	Norway	Denmark	Denmark	Denmark	Denmark
3	Denmark	Iceland	Iceland	Denmark	Norway	Switzerland	Switzerland	Iceland
4	Norway	Norway	Switzerland	Iceland	Iceland	Iceland	Iceland	Switzerland
5	Canada	Finland	Finland	Switzerland	Netherlands	Norway	Netherlands	Netherlands
-5	Rwanda	Benin	Rwanda	Yemen	Rwanda	Central African Republic	Lesotho	Botswana
-4	Benin	Afghanistan	Syria	Tanzania	Tanzania	Rwanda	Botswana	Rwanda
-3	Syria	Togo	Tanzania	South Sudan	Afghanistan	Zimbabwe	Rwanda	Zimbabwe
-2	Burundi	Syria	Burundi	Central African Republic	Central African Republic	South Sudan	Zimbabwe	Lebanon
-1	Togo	Burundi	Central African Republic	Burundi	South Sudan	Afghanistan	Afghanistan	Afghanistan

Figure 2

Since the countries were relatively consistently ranked across the years, we averaged their scores and observed the overall happiest and least happy countries in the world. We compared the scores to the mean happiness score for all countries, 5.3665. In Figure 3, the happiness score for the happiest country, Finland, is 142% of the average country's happiness score. Similarly, the happiness score of the least happy country, Afghanistan, is 58% of the average score. The top five nations were averaged together, as well as the bottom five nations, and comparisons were made between them. The happiness score for the happiest nations were about 230% of the least happy countries' score, and the happiness scores for the least happy countries were about 43% of the happiest countries' score.

Rank	Countries	Happiness Scores	Delta to Mean Happiness	Delta to Average Happiest Countries	Delta to Avg Least Happy Countries
1	Finland	7.6451	142.46%	101.22%	236.69%
2	Denmark	7.5789	141.22%	100.35%	234.64%
3	Switzerland	7.5250	140.22%	99.63%	232.97%
4	Iceland	7.5213	140.15%	99.58%	232.86%
5	Norway	7.4938	139.64%	99.22%	232.00%
-5	Burundi	3.2779	61.08%	43.40%	101.48%
-4	South Sudan	3.2693	60.92%	43.29%	101.22%
-3	Rwanda	3.2680	60.90%	43.27%	101.18%
-2	Central African Republic	3.2026	59.68%	42.40%	99.15%
-1	Afghanistan	3.1322	58.37%	41.47%	96.97%

Mean Happiness Score
5.3665

Figure 3

We then investigated the regions and their happiness rankings. We ranked the top and bottom four regions' average happiness scores across the eight years. This analysis interestingly revealed the region that is home to the top five happiest countries was not contained in the results. Figure 4 displays Australia/New Zealand as the happiest region. Based on this analysis, the Western European countries may have been better divided into smaller regions. However, the least happy regions were not surprising and aligned to the least happy nations.

Rank	Regions	Happiness Scores	Delta to Mean Happiness	Delta to Average Happiest Regions	Delta to Avg Least Happy Regions
1	Australia and New Zealand	7.3043	128.39%	103.81%	157.29%
2	North America	7.2635	127.68%	103.23%	156.41%
3	North America and ANZ	6.8373	120.18%	97.17%	147.23%
4	Western Europe	6.7400	118.47%	95.79%	145.14%
-4	Middle East and North Africa	5.2379	92.07%	74.44%	112.79%
-3	Southern Asia	4.5721	80.37%	64.98%	98.45%
-2	South Asia	4.4515	78.25%	63.26%	95.86%
-1	Sub-Saharan Africa	4.3140	75.83%	61.31%	92.90%

Mean Happiness Score
5.6890

Figure 4

Analysis of the United States

The United States was consistently in the top 20 happiest countries between 2015 and 2022. As Figure 5 shows, America's ranking started to trend down in 2018 but improved in 2022, returning to 16th place. We also observed from the table that Finland has been the world's happiest country for 5 consecutive years.

US Score	7.12	7.10	6.99	6.89	6.89	6.94	6.95	6.98
Rank	2015	2016	2017	2018	2019	2020	2021	2022
1	Switzerland	Denmark	Norway	Finland	Finland	Finland	Finland	Finland
2	Iceland	Switzerland	Denmark	Norway	Denmark	Denmark	Denmark	Denmark
3	Denmark	Iceland	Iceland	Denmark	Norway	Switzerland	Switzerland	Iceland
4	Norway	Norway	Switzerland	Iceland	Iceland	Iceland	Iceland	Switzerland
5	Canada	Finland	Finland	Switzerland	Netherlands	Norway	Netherlands	Netherlands
6	Finland	Canada	Netherlands	Netherlands	Switzerland	Netherlands	Norway	Luxembourg*
7	Netherlands	Netherlands	Canada	Canada	Sweden	Sweden	Sweden	Sweden
8	Sweden	New Zealand	New Zealand	New Zealand	New Zealand	New Zealand	Luxembourg	Norway
9	New Zealand	Australia	Sweden	Sweden	Canada	Austria	New Zealand	Israel
10	Australia	Sweden	Australia	Australia	Austria	Luxembourg	Austria	New Zealand
11	Israel	Israel	Israel	United Kingdom	Australia	Canada	Australia	Austria
12	Costa Rica	Austria	Costa Rica	Austria	Costa Rica	Australia	Israel	Australia
13	Austria	United States	Austria	Costa Rica	Israel	United Kingdom	Germany	Ireland
14	Mexico	Costa Rica	United States	Ireland	Luxembourg	Israel	Canada	Germany
15	United States	Puerto Rico	Ireland	Germany	United Kingdom	Costa Rica	Ireland	Canada
16	Brazil	Germany	Germany	Belgium	Ireland	Ireland	Costa Rica	United States
17	Luxembourg	Brazil	Belgium	Luxembourg	Germany	Germany	United Kingdom	United Kingdom
18	Ireland	Belgium	Luxembourg	United States	Belgium	United States	Czech Republic	Czechia
19	Belgium	Ireland	United Kingdom	Israel	United States	Czech Republic	United States	Belgium
20	United Arab Emirates	Luxembourg	Chile	ıited Arab Emirat	Czech Republic	Belgium	Belgium	France

Figure 5

To understand what contributed most to America's happiness score, we examined the data in a boxplot, separated by the different factors. It was clear that the Economy and Family were the key contributors for America's happiness score, while Trust (of the government) contributed the least.

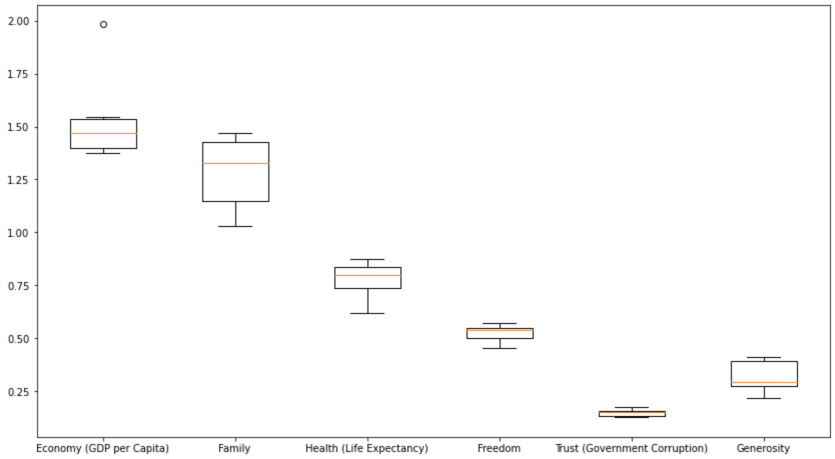

Figure 6

We then conducted trend analysis for each of the drivers and found:

- The economy rating had a significant increase in 2022, likely due to a big comeback after the pandemic. This also drove America's overall happiness ranking back to 16th place in 2022.
- America's heath scores were low due to Covid impact during 2021 and 2022
- Interestingly, considered as one of the most generous countries in the world, America's generosity score started to trend down beginning in 2018. That may be because, though the economy is strong, America's household debt has been increasing in the past few years and that might have negative impact on the generosity score.

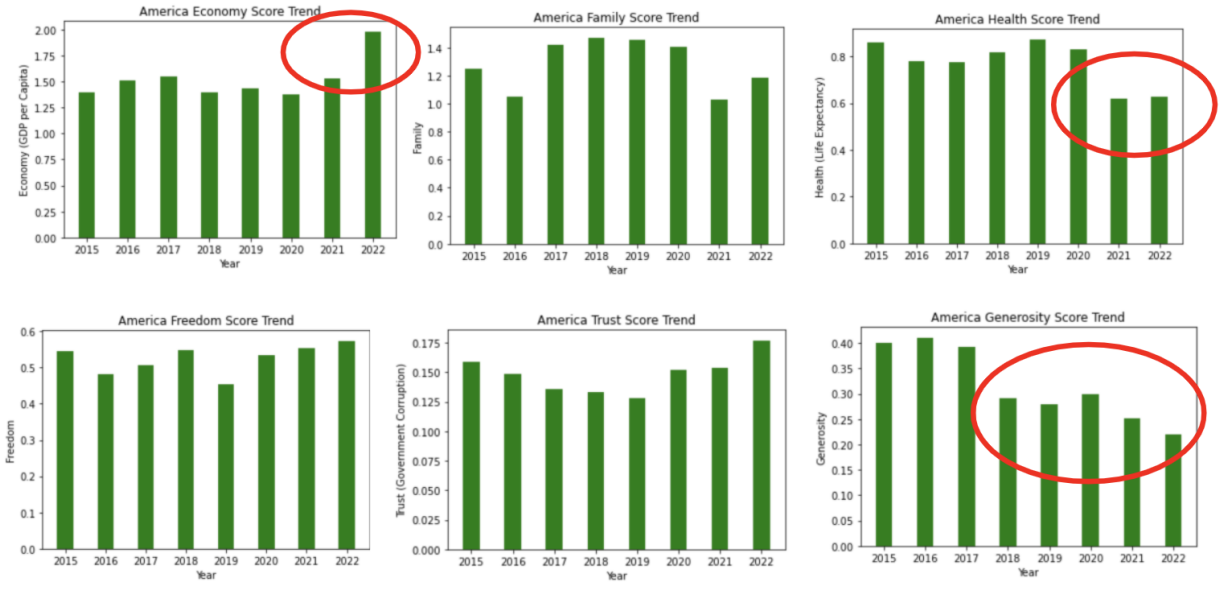

Figure 7

Compared to the world's happiest country Finland, America is lagged behind mostly in Trust score, followed by freedom and health score.

	Year 2022		
	Finland	United States	Difference
Rank	1	16	
Generosity	0.109	0.22	-0.111
Family	1.258	1.182	0.076
Health	0.775	0.628	0.147
Trust of Government	0.534	0.177	0.357
Freedom	0.736	0.574	0.162

Figure 8

Variable Relationships

To continue exploratory data analysis, we turned to the relationship(s) between the variables and the happiness score for each country over time.

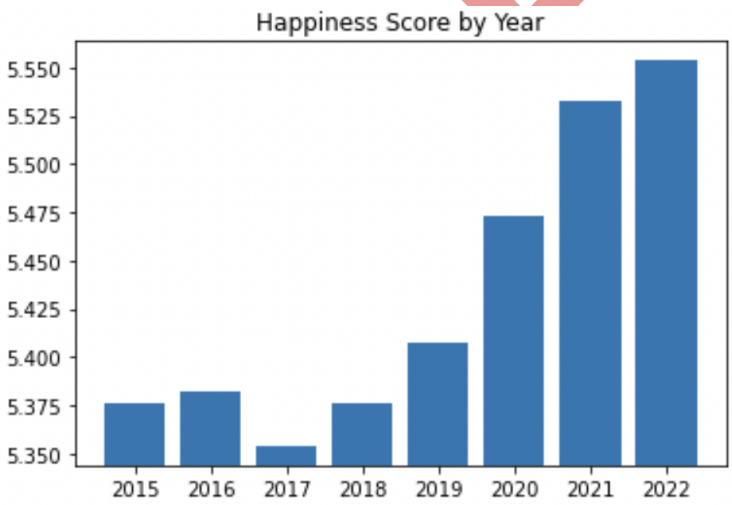

Figure 9

Figure 9 displays the overall world happiness score year over year. The data revealed the happiness scores overall generally increase, with the exception of 2017. As we continued to investigate individual variables, it became clear that world events and local events had dramatic affects on happiness scores. For example, in 2016 the world saw Great Britain vote to leave the EU, potentially resulting in a low happiness score in 2017.

To determine the relationship between each variable and happiness, we plotted each variable against happiness as well as constructed a bar chart to display the change over time of the ranked importance of the happiness factor.

Economy (GDP per Capita)

The impact of economy on happiness appeared to have the strongest linear relationship, based on the scatterplot seen in Figure 11. Countries who ranked economy as most important, seemed to have higher happiness scores. However, year over year, economy was not consistent in its rating, as seen in Figure 10. The variance could be a result of world events. For example, in 2015 the Chinese stock market crashed, in 2017 the stock markets began to rise, 2020 saw the beginnings of COVID-19, and 2022 saw the start of recovery from the COVID-19 lockdowns.

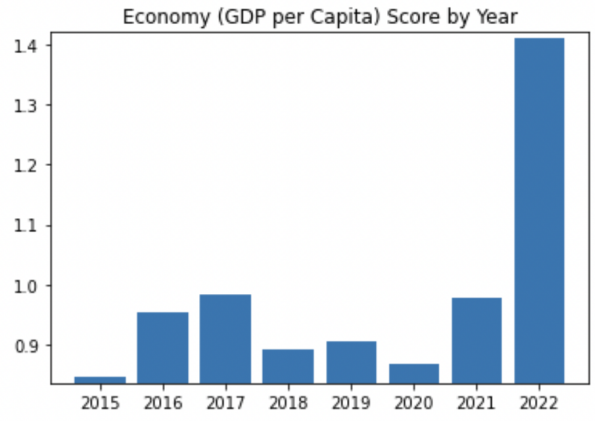

Figure 10

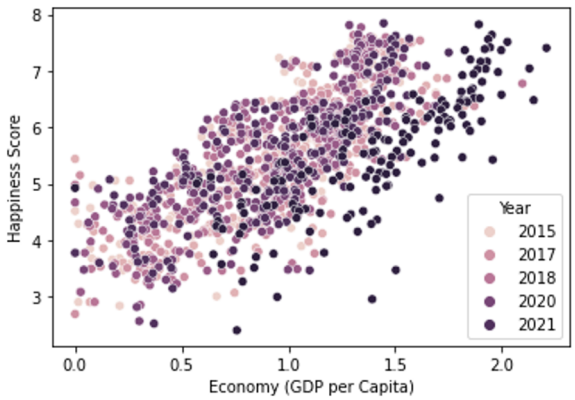

Figure 11

Health (Life Expectancy)

Health also appeared to have a linear relationship with the happiness score. Like economy, health also saw variance over the years, 2021 seeing a steep drop likely due to the COVID-19 worldwide pandemic.

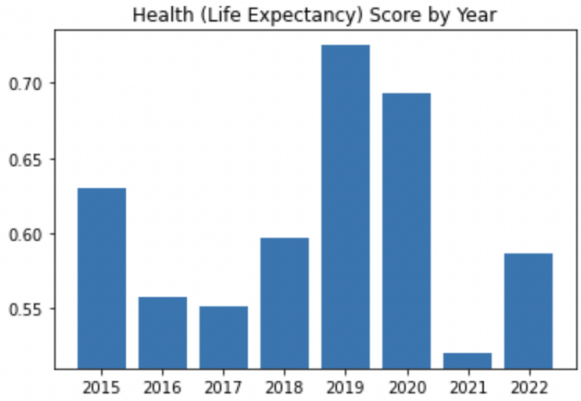

Figure 12

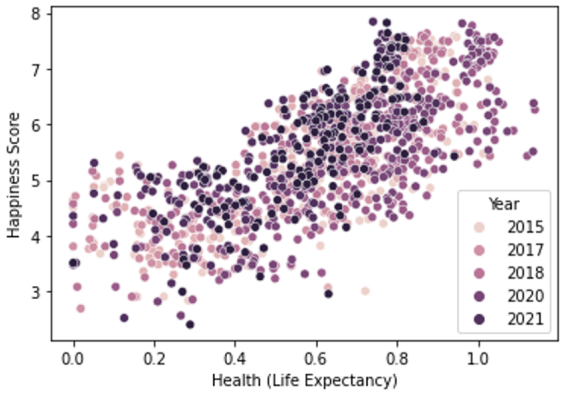

Figure 13

Freedom

Freedom also appeared to have a relatively linear relationship with happiness, though potentially weaker than that of economy and health. The data suggested variance year over year, again potentially due to world events. For instance, 2016 saw the Brexit vote, Russian

interference with an election, North Korean missile tests, and other impeachments/ interferences. 2019 saw many government protests worldwide.

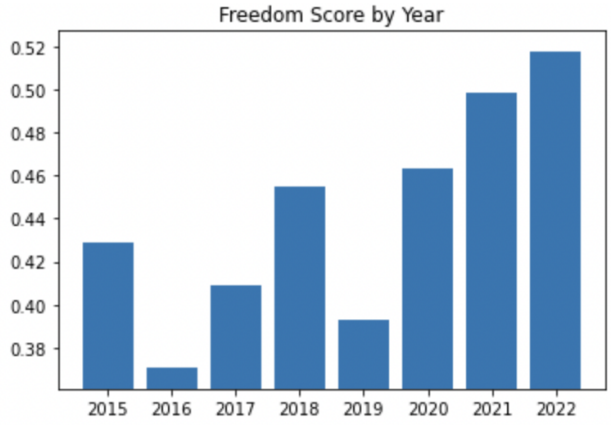

Figure 14

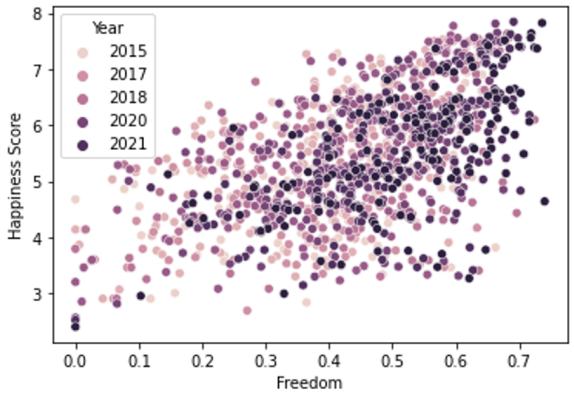

Figure 15

Family

Family's relationship with happiness score was also observed to be potentially linear. The data over the year suggested a bellshaped curve, seeing a significant drop in 2021 which could potentially be related to families quarantining together in 2021.

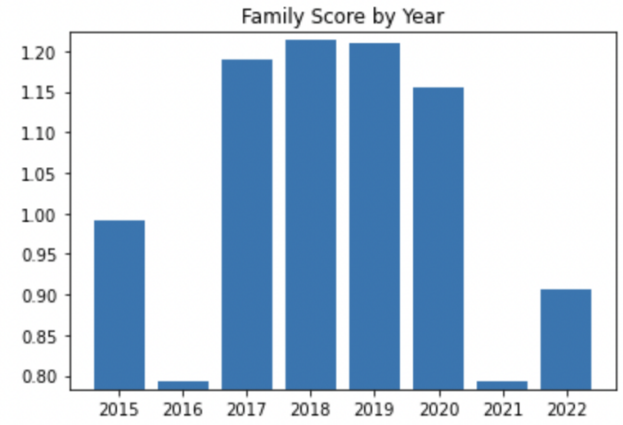

Figure 16

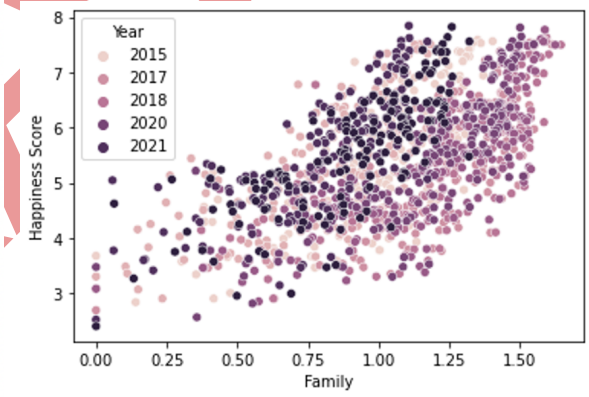

Figure 17

Trust (Government Corruption)

Trust was the first variable to have a clear non-linear relationship with the happiness score. Instead, trust appeared to have a more logarithmic shape, as seen in Figure 19. Interestingly, trust appeared to have an inverse relationship to the bellshaped curve of family.

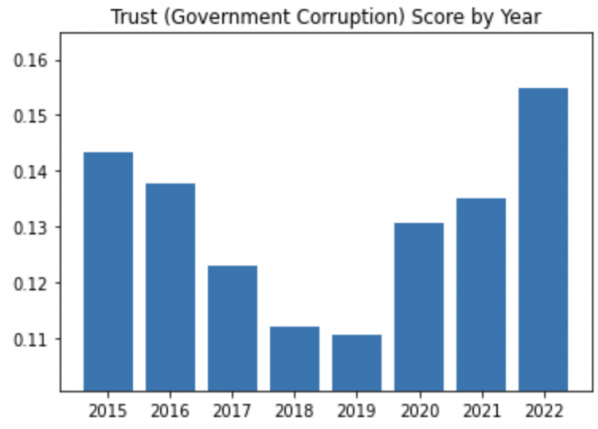

Figure 18

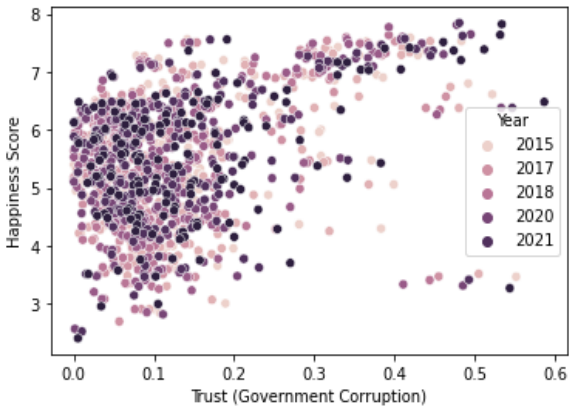

Figure 19

Generosity

Likewise to trust, generosity also did not appear to have a linear relationship with happiness. Several of the years experienced major drop-offs in the importance of generosity to country happiness. Some potentially influential events include the many worldwide protests against governments in 2018 and 2019, the Me Too Movement, global warming's sharp increase, and an increase in mass shootings.

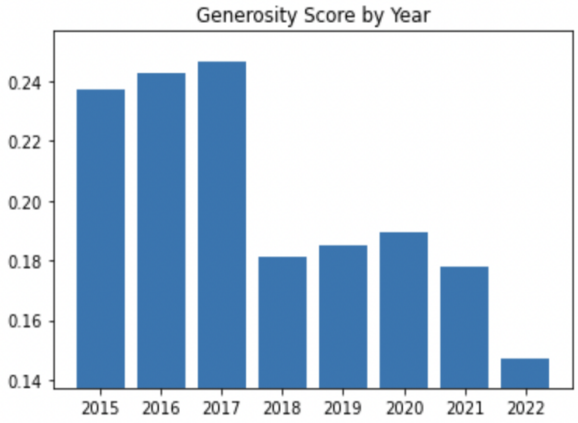

Figure 20

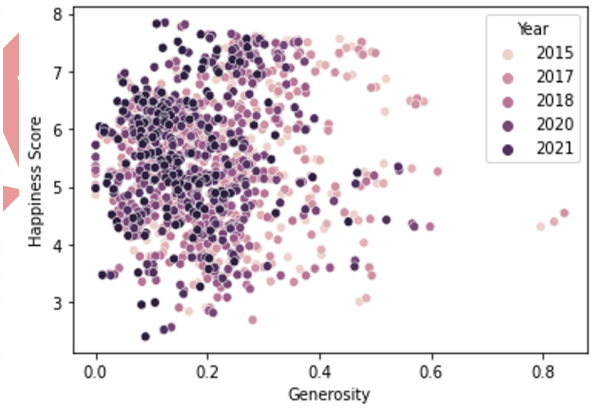

Figure 21

Linear Regression

Based on the exploratory data analysis findings of the individual variables, we explored if a linear regression could be used to predict country happiness scores.

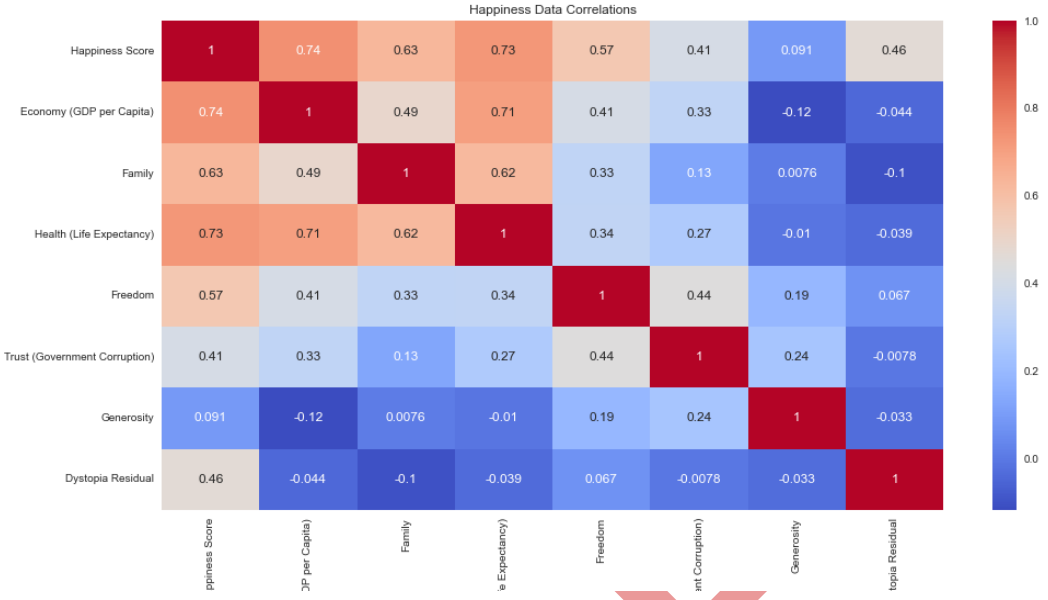

Figure 22

Using a heat map, Figure 22, we confirmed our initial findings that economy, health, freedom, and family have a strong positive correlation with country happiness scores. This indicated a multiple linear regression may be an appropriate model for this dataset.

We individually plotted the variables against the happiness score along with a regression line and visually determined there were no clear outliers or concerning points of interest.

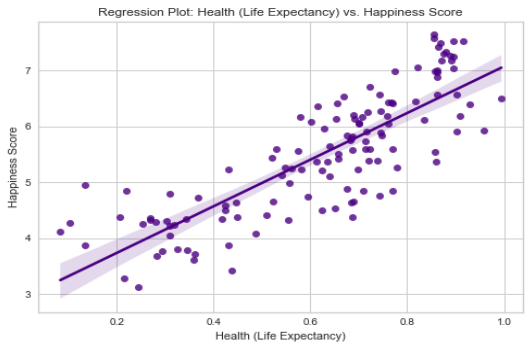

Figure 23

Figure 24

Figure 25

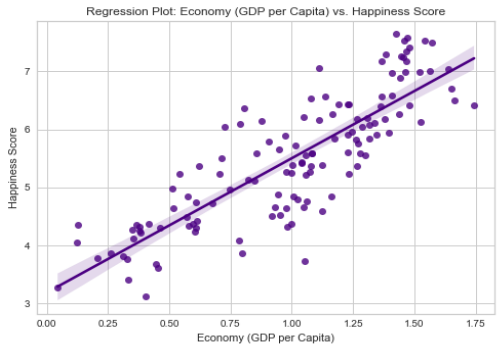

Figure 26

We then checked the assumptions of linear regression with a residual and normal probability plot. All assumptions were met.

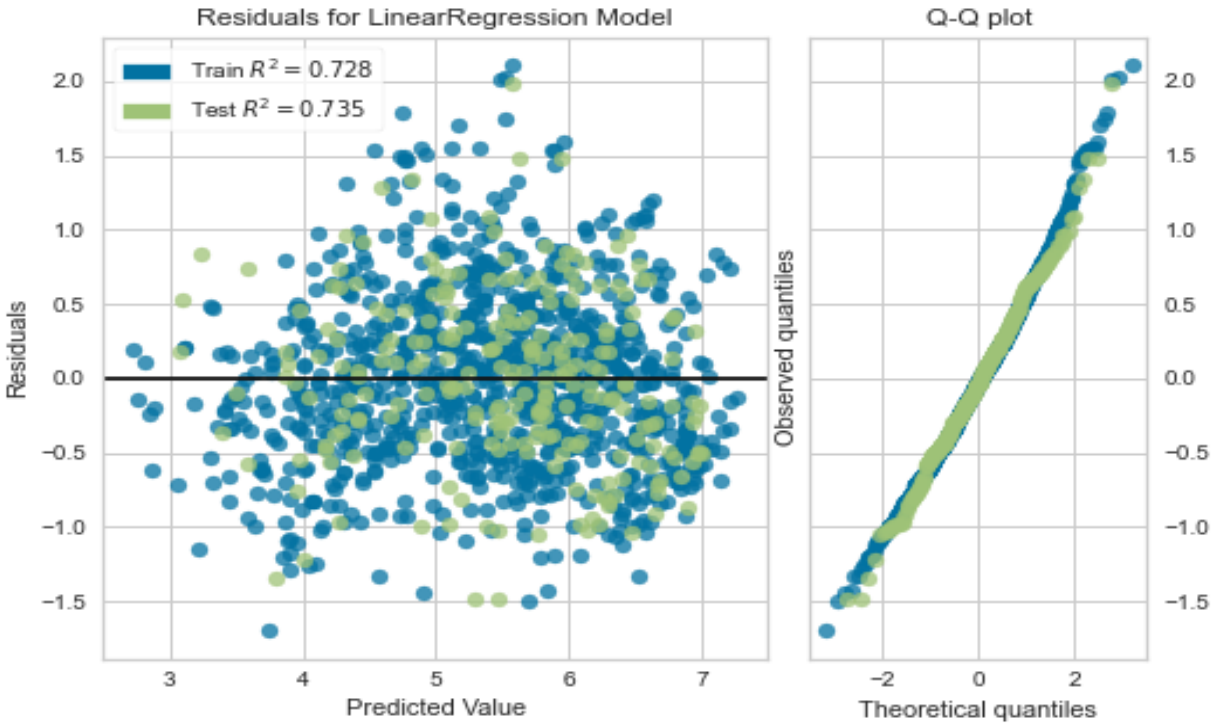

Figure 27

We concluded the linear regression model of this dataset to be:

$$y = 5.4058499 + 0.2861x_1 + 0.2519x_2 + 0.3704x_3 + 0.2976x_4,$$

where x_1 represents the health rating, x_2 represents family, x_3 represents economy, and x_4 represents freedom.

Unit Testing

During our analysis process, we conducted several unit tests.

We conducted a geometry test, seen in Figure 28, to ensure the geometry column in the shapefile loaded as a geometry object to generate the Choropeth plot (Figure 1) using geopandas. The geometry file was the correct object type.

Figure 28 also displays the country info test conducted to validate the pycountry package returned the expected country based on the alpha_3 country code. The pycountry information aligned with the expected inputs/outputs.

```
import os
import unittest
import pycountry
import geopandas as gpd

class GDFTest(unittest.TestCase):

    def test_for_geometry(self):
        fp = os.path.join(os.path.dirname(__file__), r'data\ne_10m_admin_0_countries\ne_10
        test_gdf = gpd.read_file(fp)[['geometry']].to_crs('EPSG:4326')
        sample_gdf = gpd.GeoDataFrame({'geometry': []}, crs="EPSG:4326")
        self.assertEqual(test_gdf['geometry'].dtype, sample_gdf['geometry'].dtype)

    def test_country_info(self):
        country_dat = {'Germany': 'DEU', 'United States': 'USA', 'Italy': 'ITA', 'Australia': 'AUS'}
        for country_name, country_code in country_dat.items():
            result = pycountry.countries.get(alpha_3=country_code)
            result.name
            self.assertTrue(country_name==result.name, f'Expected name {country_name} does not match {result.name}')

if __name__ == '__main__':
    unittest.main()
```

```
!python unittests.py

--------------------------------
Ran 2 tests in 0.699s

OK
```

Figure 28

Additional testing was conducted to ensure the function used to create a year-specific data frame ran correctly. The first test checked that the data frames which were outputted from the function matched those created manually. We specifically tested the data frame for the year 2015's accuracy. The test passed, indicating the function-made data frame was equivalent to the manually created data frame. The second test checked that the function identified the out-of-range years correctly. We confirmed that an error was returned when a year was entered into the function that fell outside of the years of the data set. To conserve runtime, years just outside of this range were tested instead of all years outside of the range. Each of these years passed, indicating that the function will only work for years within the range 2015-2022.

```
##Create a function for the below code so that can meet the unit test requirement
df_yr<- function(year){
  if(year %in% seq(2015,2022,1)){
  df%>%
    filter(Year==year)%>%
    arrange(desc(Happiness.Score))
  }
  else{
    stop("year not in data set. ",.call=F)
  }

}

df2015<-df_yr(2015)
df2016<-df_yr(2016)
df2017<-df_yr(2017)
df2018<-df_yr(2018)
df2019<-df_yr(2019)
df2020<-df_yr(2020)
df2021<-df_yr(2021)
df2022<-df_yr(2022)

##Test that the function found data frame matches a manually found data frame
df2015_expected<-df%>%
  filter(Year==2015)%>%
  arrange(desc(Happiness.Score))

test_that("DF Correctness", expect_equal(df2015,df2015_expected))

##Test that the function gives an error when the year entered is not in the data frame
for(x in seq(2010,2030,1)){
  if(x %in% seq(2015,2022,1)){

  }
  else{
    test_that("DF Year",expect_error(df_yr(x)))
  }
}
```

Figure 29

Conclusion

We concluded the strongest indicators of country happiness are economy, health, freedom, and family. Through analysis, we determined the North American, European, and Australian regions had an overall higher happiness score while the African and Middle Eastern regions had the lowest scores. The United States's happiness score consistently ranked in the top 20 between 2015 and 2022 for country happiness and the overall happiness score worldwise seems to be increasing year over year. Finally, we determined our regression model is fairly accurate in predicting the happiness score of a country.