

Happiness Analysis

Group 2: Shen, O'Donnell, Ingram, Gloss, Scerbo

Agenda

- Data Introduction
 - Data Introduction
 - Data Cleaning and Preprocessing
 - Research Questions
- Visualizations and Results
 - Region
 - US
 - Overall Variables
- Regression Model and Unit Testing
- Conclusion

Data Introduction

Cavan Ingram

Data Introduction

- Data obtained from Gallup World Poll from 2015-2022
 - Happiness scores are based on answers to Cantril Ladder question, which asks people to rate their life on a scale from 0-10
 - Economy, Family, Trust, Freedom, Health, and Generosity are columns that each measure the extent to which each factor contributes to the overall happiness in each country
 - These do not have a direct impact on the happiness score, but help explain why some countries rank higher than others (Dystopia residual accounts for any over/under explained average)
 - Happiness Rank ordered the happiness scores for each country (highest score being 1) for each year
 - Region gave a location of where to find a given country

Data cleaning and Preprocessing

- Deal with missing data
 - For missing Happiness Rank values, we computed the happiness rank from the happiness score
 - For missing Region values, we created a country region dictionary and filled in the missing data
 - Added a Year column for each file
- Uniform the format
 - Unify the data column name for each of the 8 files
 - Unify the data format for each column
- Append 8 year data files together to be ready for future analysis

Research Questions

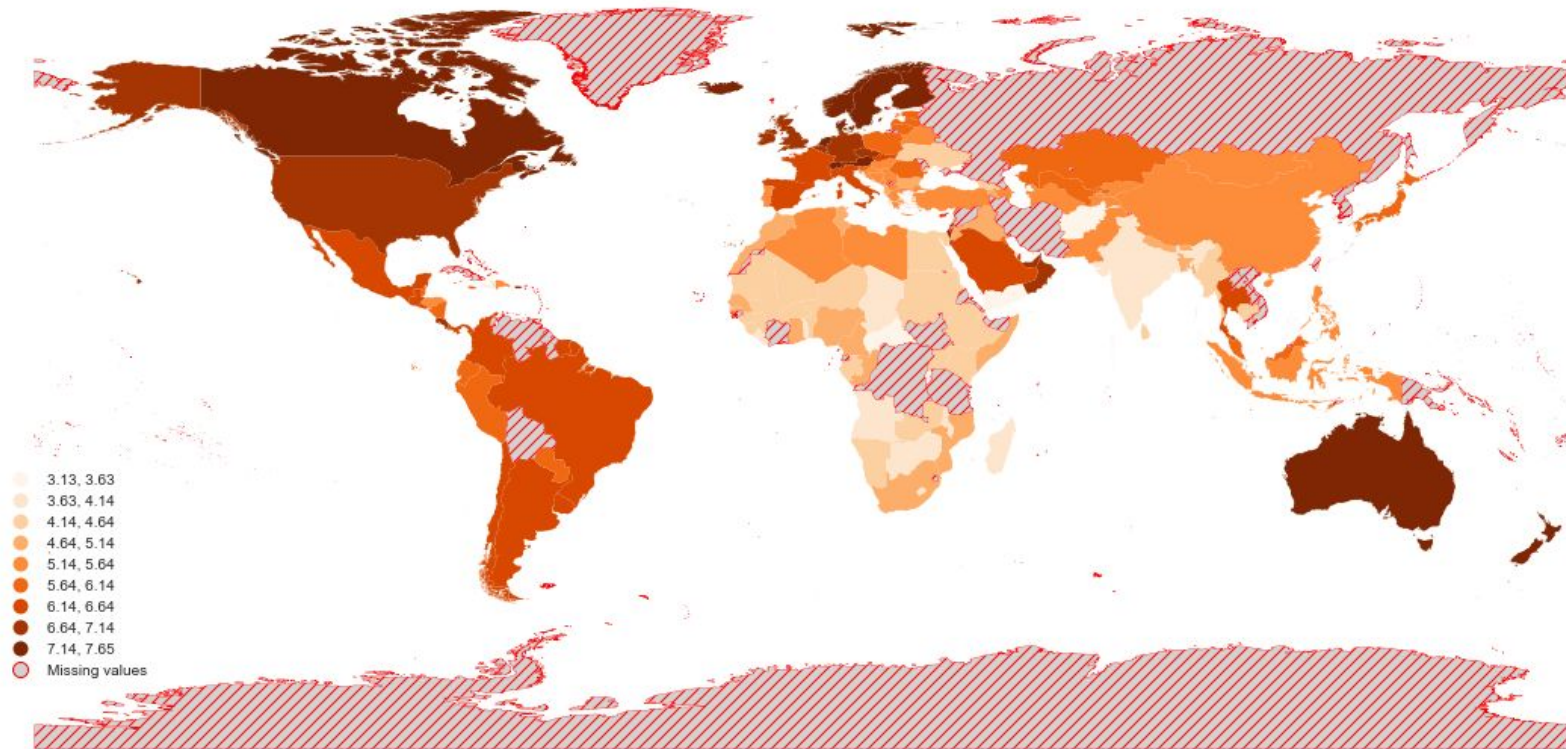
- What contributes to country happiness?
 - What countries or regions globally have the highest happiness score
 - How is the US happiness score compared to other countries
 - What features in the dataset relate to happiness score the most
 - Based on the features that have the highest correlation to the happiness score, are they good predictors for a linear regression model

Visualizations and Results

Claire Gloss, Qian Shen, & Lauren O'Donnell

Choropleth Map

Average Happiness Score by Country



Five Highest and Lowest Countries by Year

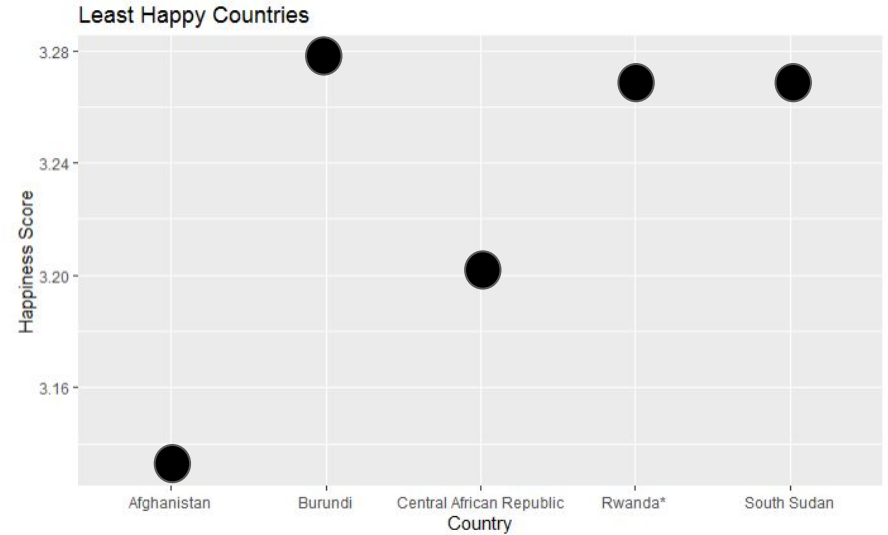
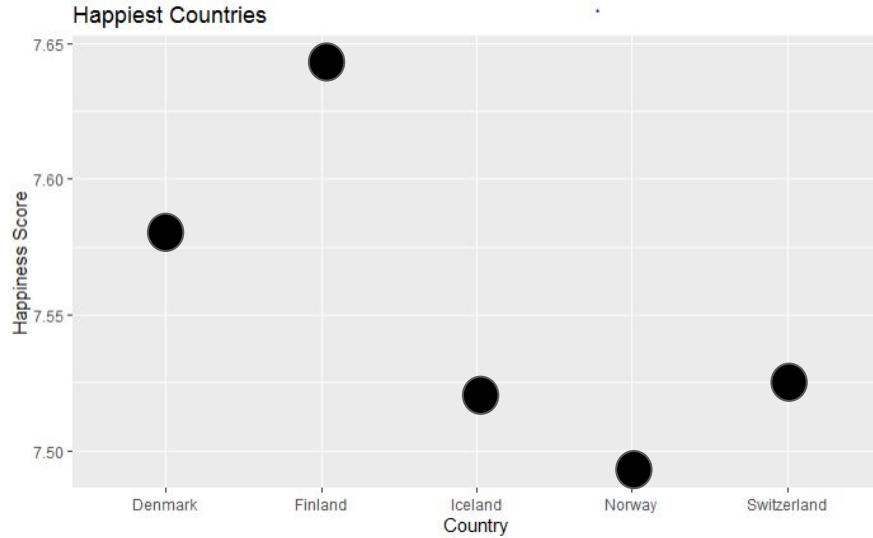
Rank	2015	2016	2017	2018	2019	2020	2021	2022
1	Switzerland	Denmark	Norway	Finland	Finland	Finland	Finland	Finland
2	Iceland	Switzerland	Denmark	Norway	Denmark	Denmark	Denmark	Denmark
3	Denmark	Iceland	Iceland	Denmark	Norway	Switzerland	Switzerland	Iceland
4	Norway	Norway	Switzerland	Iceland	Iceland	Iceland	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

Averaged Happiness Level by Country

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

Plot Comparison for Happiness Scores

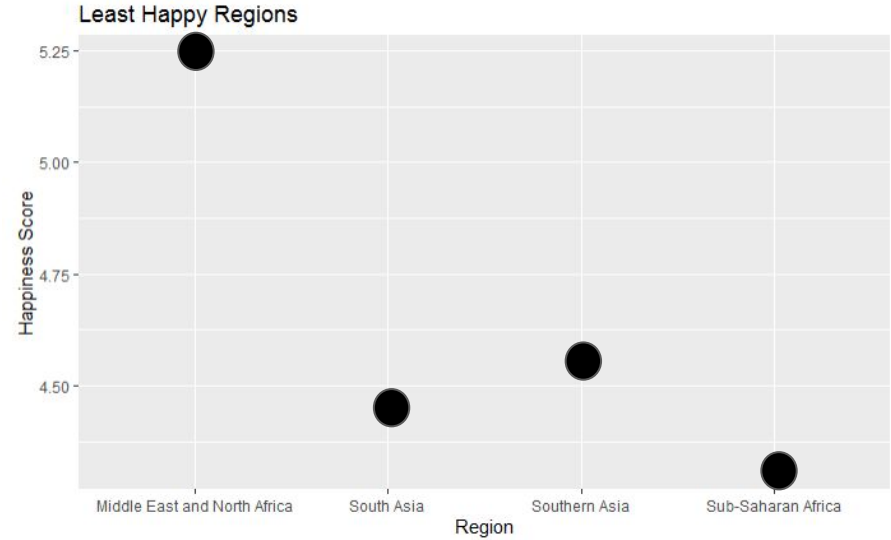


Averaged Happiness Level by Region

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

Plot Comparison for Happiness Scores

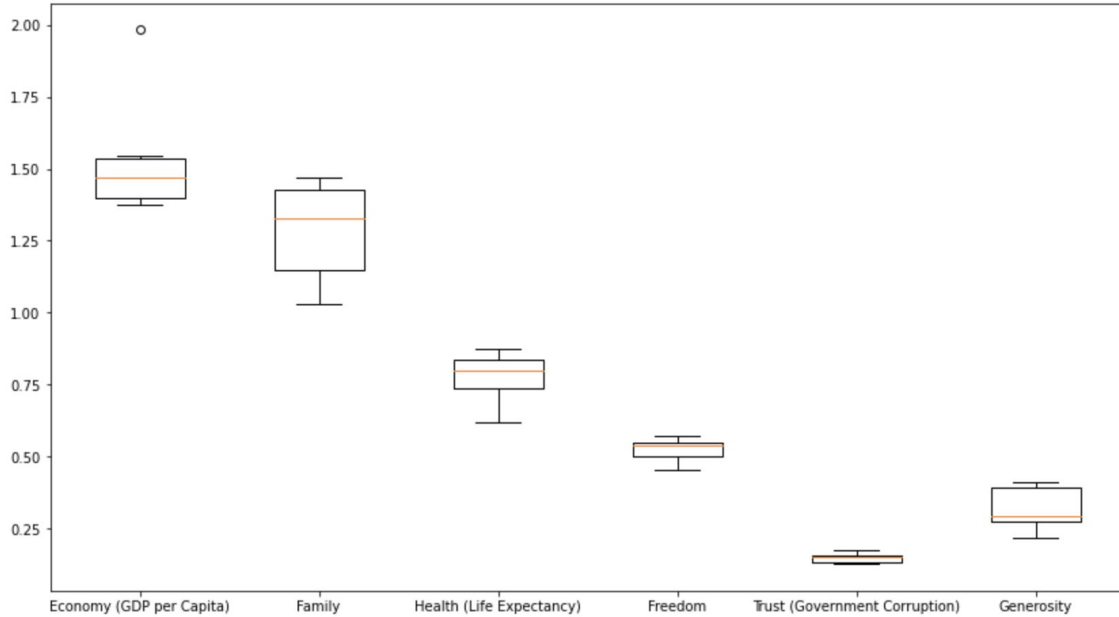


Deep dive into America

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	ited Arab Emirat	Czech Republic	Belgium	Belgium	France

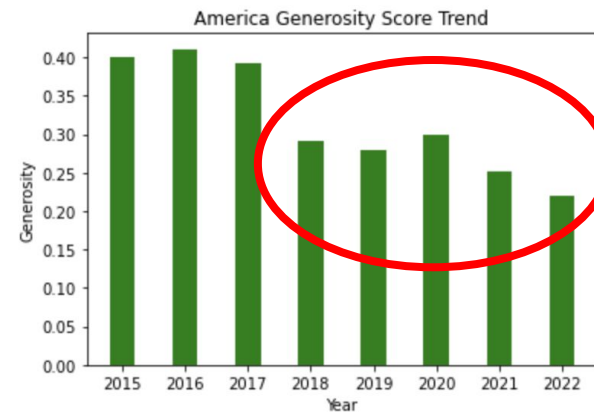
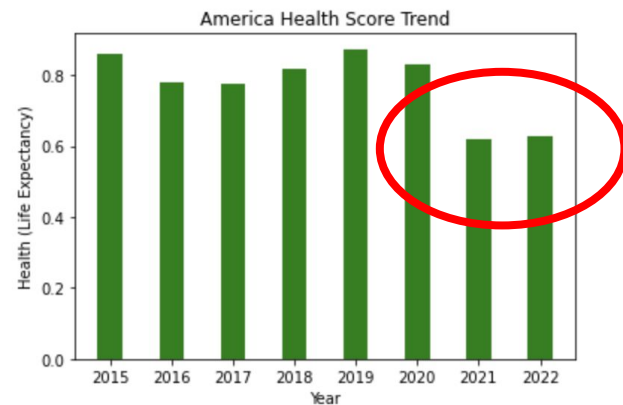
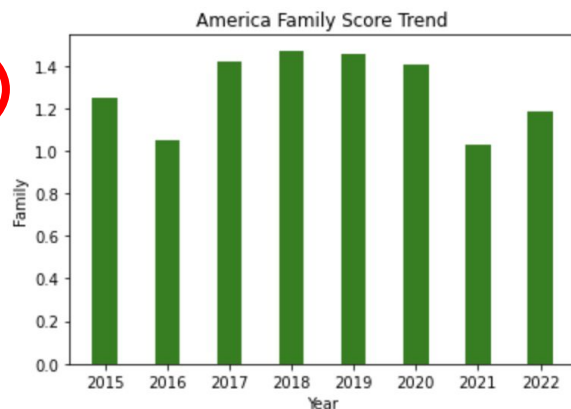
America's happiness ranking trended down in the past few years and stayed 16th in 2022

Contributors of America's Happiness Score



- Economy and Family are main contributors to America's Happiness score
- Trust contributes the least to the overall happiness score

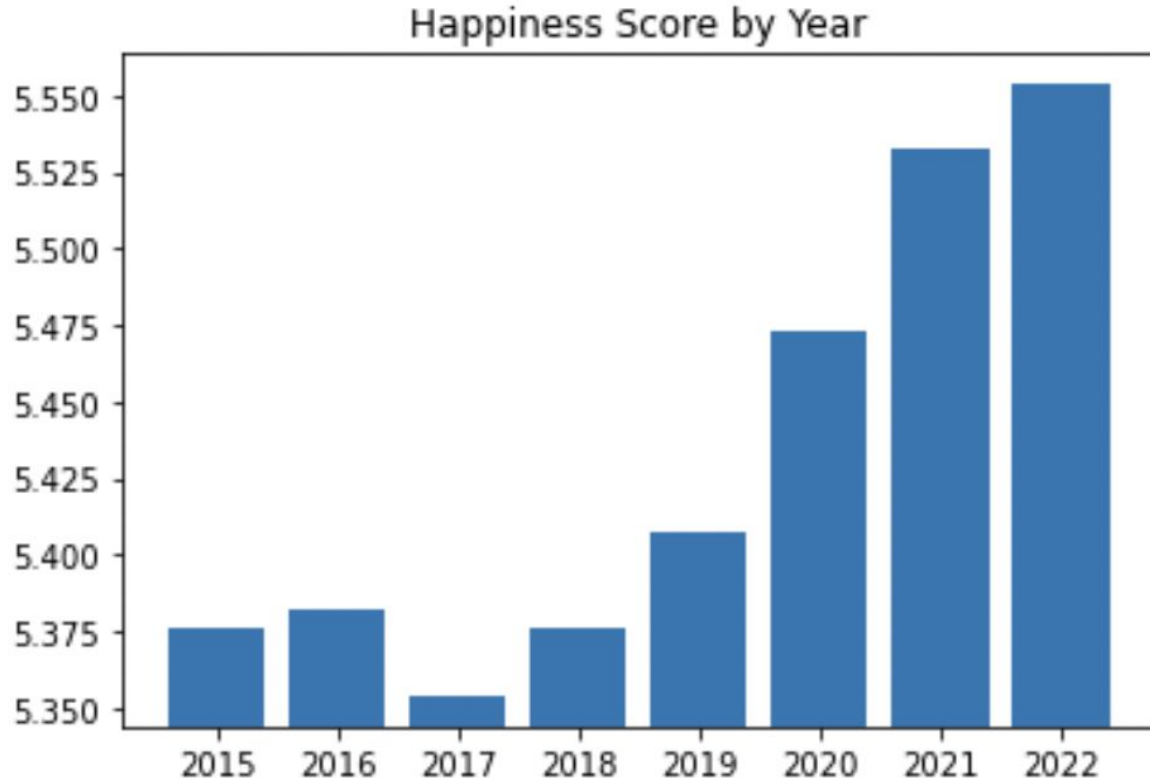
Trend analysis of America's happiness score factors



America's happiness scoregap to Finland

	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

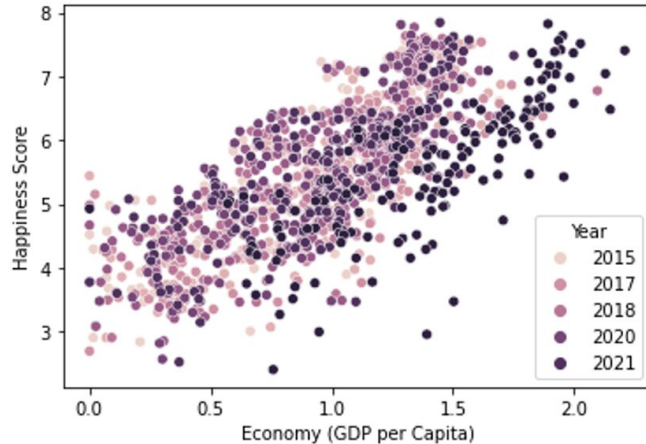
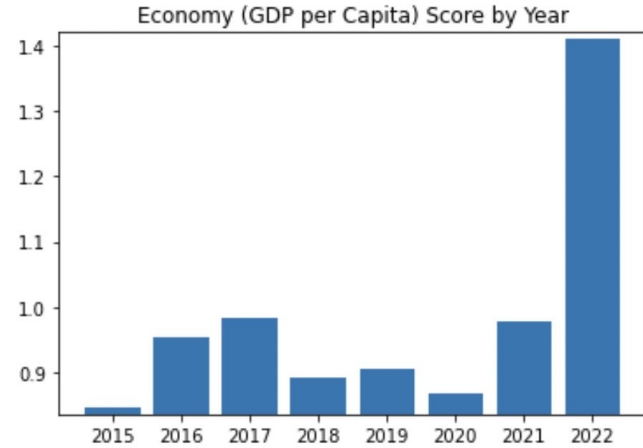
Overall Happiness by Year



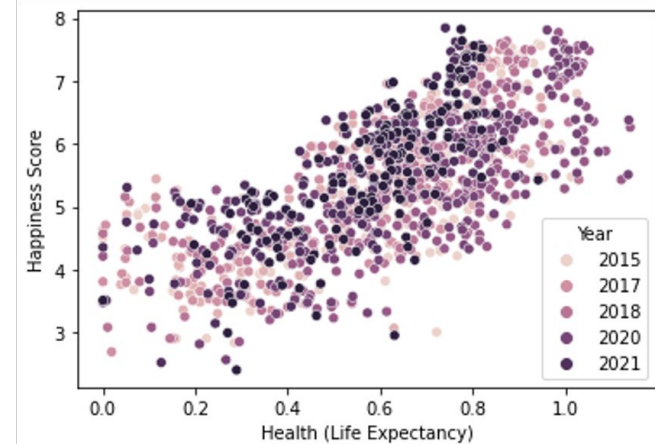
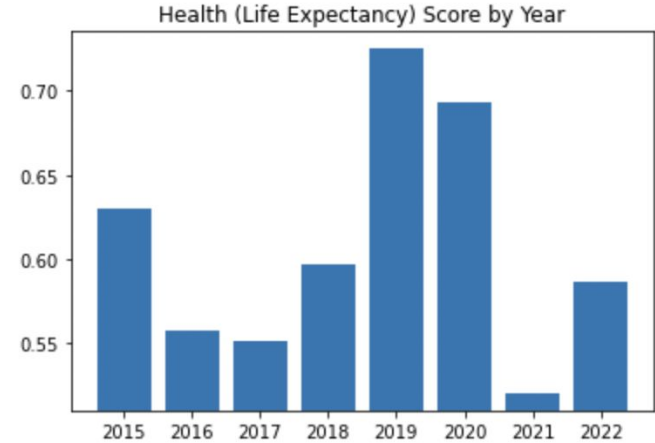
World Events:

- 2015 Chinese Market Crash
- 2016 Brexit Vote
- 2017 Rohingya Crisis and US Withdrawal from Paris Climate Agreement & Trans-Pacific Agreement
- 2021 COVID
- 2022 Return from COVID

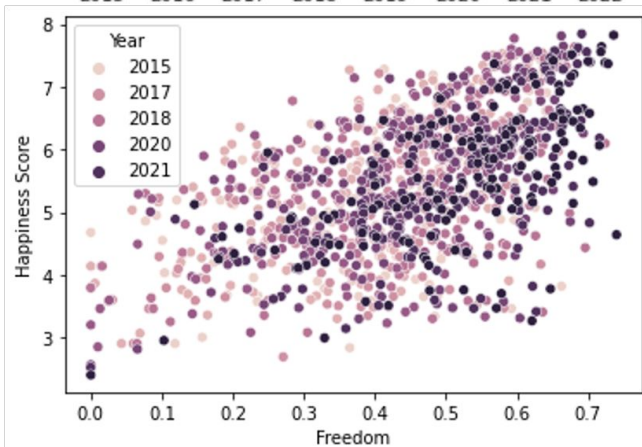
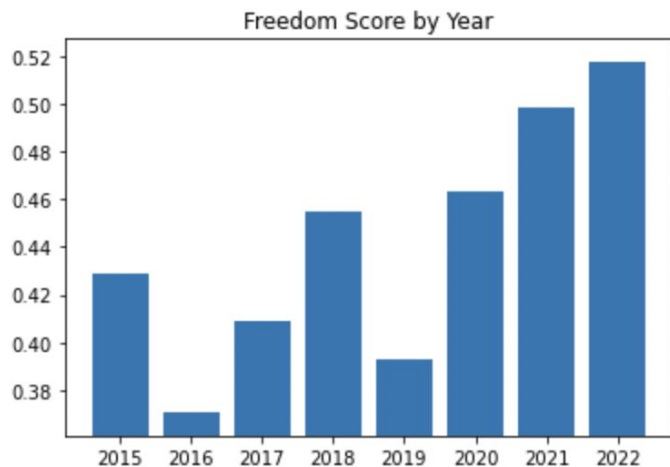
Economy (GDP per Capita)



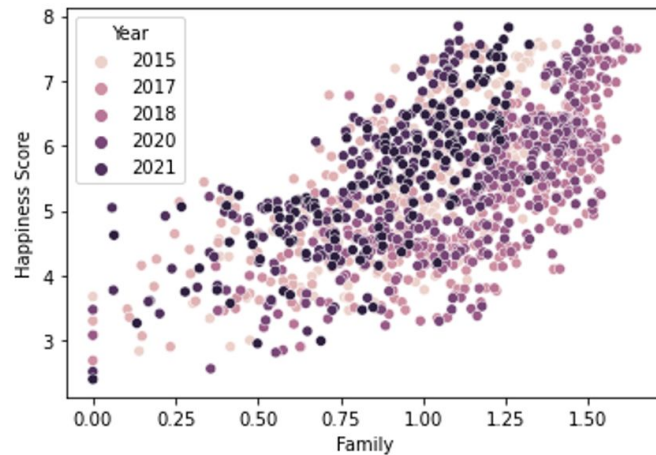
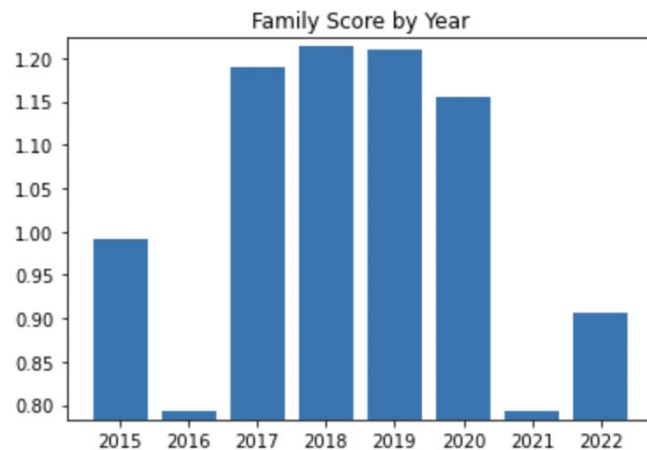
Health (Life Expectancy)



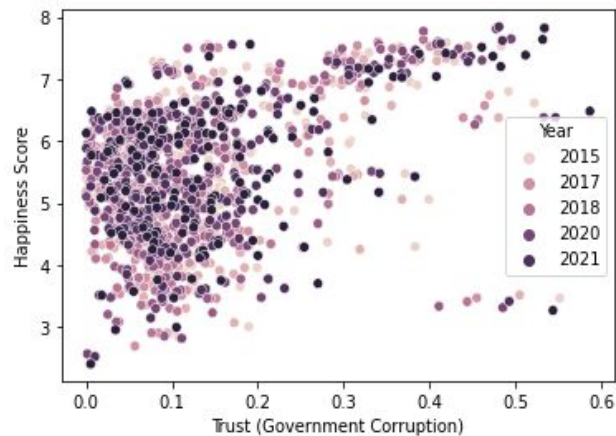
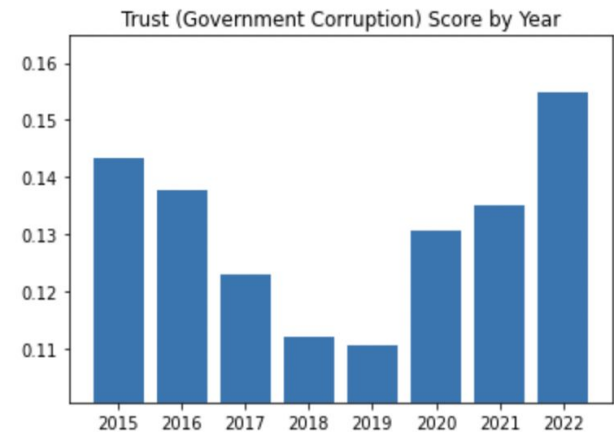
Freedom



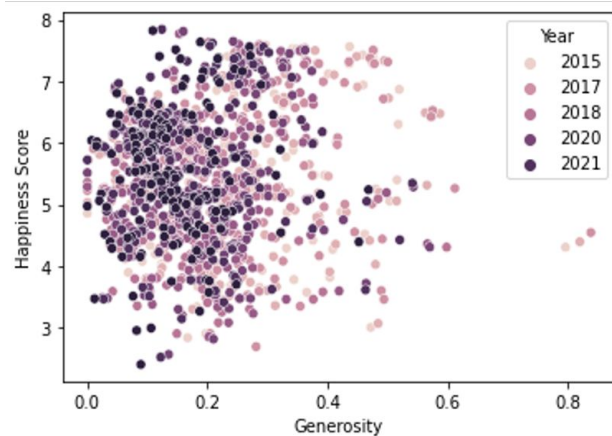
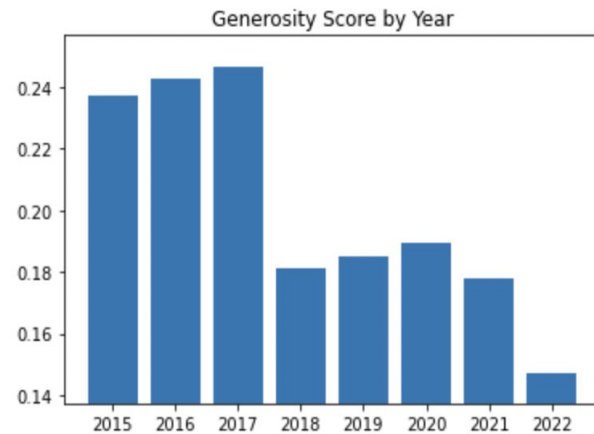
Family



Trust (Government Corruption)



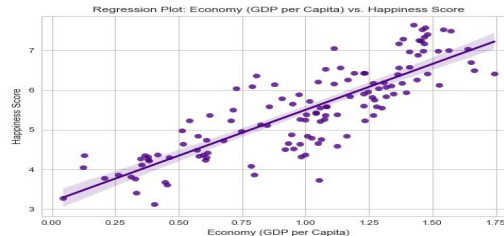
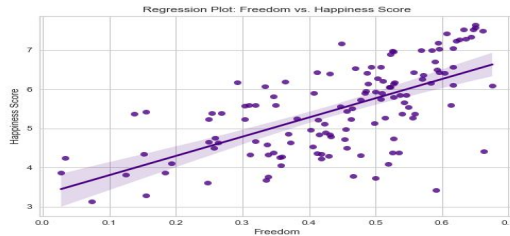
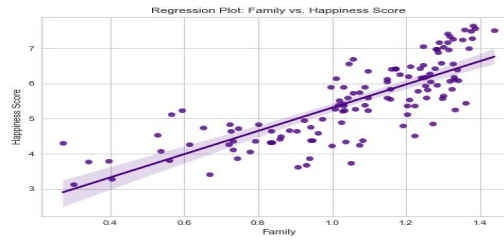
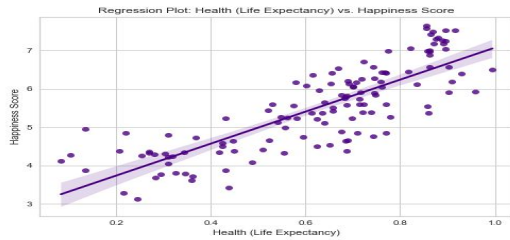
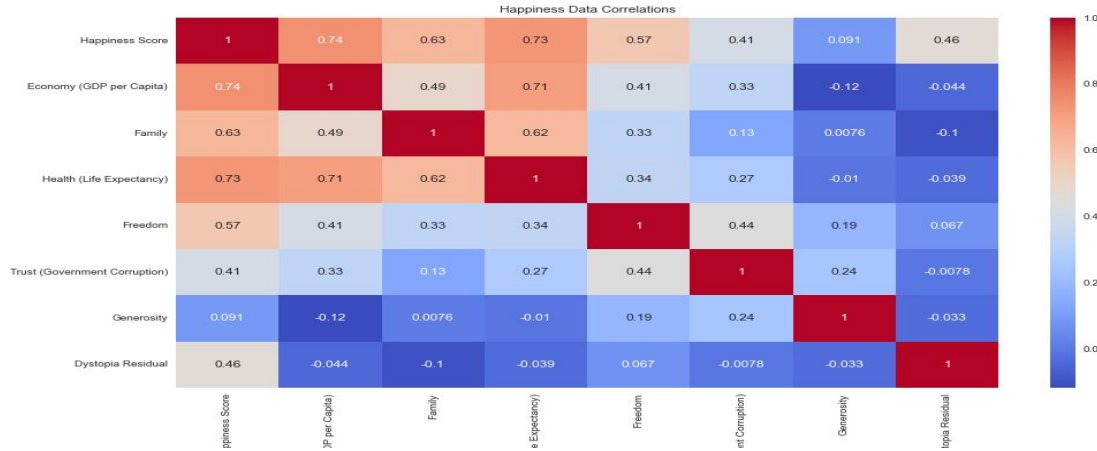
Generosity



Regression model and unit test

Dominic Scerbo

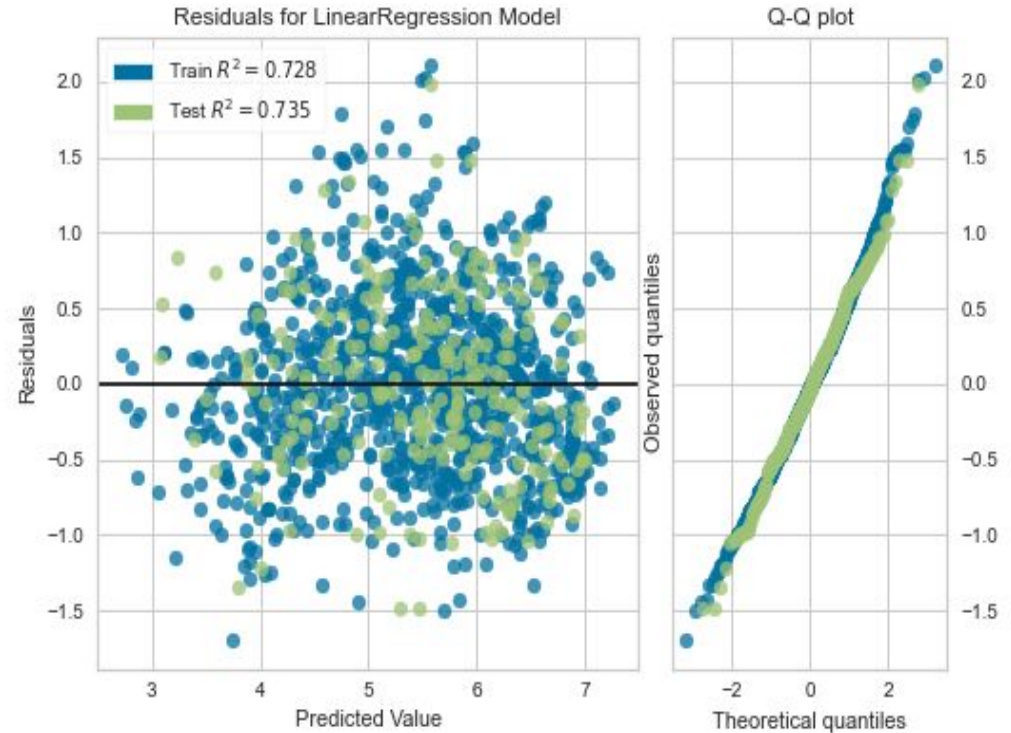
Feature Correlations



- Strong positive correlation to Happiness Score
 - Economy
 - Health
 - Family
 - Freedom
- A Multiple Linear Regression Model could be an appropriate model for this dataset
- Strong positive correlations between the identified variables may lead to overfitting
- Independently, each predictor has a strong linear relationship to the response variable, Happiness Score.
- No signs of concerning outliers or points of interest at the extremes of the axes

Multiple Linear Regression Model

- Sklearn - LinearRegression
- $y = 5.405849946589013 + 0.2861 \cdot \text{Health} + 0.2519 \cdot \text{Family} + 0.3704 \cdot \text{Economy} + 0.2976 \cdot \text{Freedom}$
- The model meets the regression assumptions and is a good predictor of happiness score
 - The residual plot shows the variance of residuals has roughly a mean of 0
 - QQ shows the normality assumption of the errors are met
 - R squared does not indicate overfitting
- R squared (train): 0.73
- R squared (test): 0.74
- Mean Absolute Error: 0.4566
- Mean Square Error: 0.3256
- Root Mean Square Error: 0.5706



Unit testing

- **Geometry Test**

- Ensure the geometry column in shapefile is loaded as a geometry object in order to generate Choropleth plot using geopandas
- The geometry field was the correct object type

- **Country Info Test**

- Validate the pycountry package returns the expected country based on 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_10m_admin_0_countries.shp')
        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
```

Conclusion

Cavan Ingram

Conclusion

- North America, Europe, and Australia have higher overall happiness scores
- Africa and the Middle East have the lowest scores
- The US has strong happiness scores every year (top 20)
- The overall happiness score seems to be increasing year over year
- Economy, health, family, and freedom seem to have a high correlation with the happiness score
- A regression model seems to be fairly accurate in predicting a happiness score

Questions?