

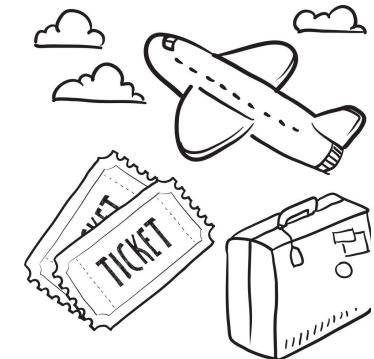
World Happiness

Barbara Fagundez

Motivation

As a young person with not many obligations such as paying a mortgage and having my own family, I'd like to move abroad for a couple of years.

Happiness is something I value a lot and I think our environment plays a huge role in that.



I was curious to see which countries are the “happiest”. Of course, happiness is very subjective but this dataset could be a good place to start.

Dataset

Happiness score based off of answers by individual respondents, where each respondent asked to rate their own life on a scale of 0-10, 10 being the best life.

Respondents were asked how each of the following 6 factors contributes to increasing quality of life: 'GDP per capita','Social support', 'Healthy life expectancy', 'Freedom to make life choices', 'Generosity', and 'Perceptions of corruption'.

The dataset is a pandas dataframe

	Year	Score	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption
0	2015	5.375734	0.846137	0.991046	0.630259	0.428615	0.237296	0.143422
1	2016	5.382185	0.953880	0.793621	0.557619	0.370994	0.242635	0.137624
2	2017	5.354019	0.984718	1.188898	0.551341	0.408786	0.246883	0.123120
3	2018	5.366897	0.883677	1.216058	0.596877	0.455606	0.180974	0.112000
4	2019	5.407096	0.905147	1.208814	0.725244	0.392571	0.184846	0.110603

Python Libraries

Using the libraries pandas, matplotlib.pyplot, seaborn, folium, and json, each country's happiness score and how much they value each of the 6 factors was visualized and analyzed.

Plotting grouped histograms was difficult with matplotlib, so Seaborn was used instead. It adds to what matplotlib has to offer, but makes it easier to work with pandas data structures.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import folium
import json

import warnings
warnings.filterwarnings('ignore')

%matplotlib inline
```

Methods

1) The data from different years...

- a) ...had different column names, so the columns had to be renamed to be consistent with each other

```
# Tables do not have the same column names, so we need to fix that
new_names = ['Country', 'Happiness Rank', 'Happiness Score', 'Economy (GDP per Capita)', 'Family',
             'Health (Life Expectancy)', 'Freedom', 'Generosity', 'Trust (Government Corruption)',
             'Dystopia Residual']

data_2015.columns = new_names
data_2016.columns = new_names
data_2017.columns = new_names
```

- b) ...was not in the same file, so a new dataframe was created to merge all the data

```
# Finally, concatenating all data
data = pd.concat([old_data, new_data], axis=0)
```

- c) ...was grouped by year and the mean of each factor was calculated across all years (2015-2019)

```
# Group the data by year and average the factors
grouped = data.groupby(by = 'Year')[['Score', 'GDP per capita',
                                         'Social support', 'Healthy life expectancy',
                                         'Freedom to make life choices', 'Generosity',
                                         'Perceptions of corruption']].mean().reset_index()
```

Methods

2) Histograms plotted with seaborn

- a) Created grouped histogram to show how important each factor was compared to other years.
Grouped by 6 factors, with average score on the y-axis

```
plt.figure(figsize = (16, 9))

ax = sns.barplot(x = grouped[grouped['Factor'] != 'Score']['Factor'], y = grouped['Avg_value'],
                  hue = grouped['Year'])

plt.title("Difference in Factors - Then and Now - ", fontsize = 25)
plt.xlabel("Factor", fontsize = 20)
plt.ylabel("Average Score", fontsize = 20)
plt.xticks(fontsize = 15)
plt.yticks(fontsize = 15)
plt.legend(fontsize = 15)

ax.set_xticklabels(['Money','Family', 'Health', 'Freedom', 'Generosity', 'Trust']);
```

- b) Plotted Country vs. Happiness Score for the top 10 happiest and bottom 10 least happy countries

```
# Average top 10 most happy countries
country_score_avg = data[data['Year']==2019].groupby(by = ['Country or region'])['Score'].mean().reset_index()
table = country_score_avg.sort_values(by = 'Score', ascending = False)
top10 = table.head(10)
```

```
# Average top 10 least happy countries
table2 = country_score_avg.sort_values(by = 'Score', ascending = True)
bottom10 = table2.head(10)
```

```
plt.figure(figsize = (16, 9))

sns.barplot(y = top10['Country or region'], x = top10['Score'])

plt.title("Top 10 Happiest Countries in 2019", fontsize = 25)
plt.xlabel("Happiness Score", fontsize = 20)
plt.ylabel("Country", fontsize = 20)
plt.xticks(fontsize = 15)
plt.yticks(fontsize = 15);
```

Methods

- 3) Global choropleth map of happiness score plotted using Folium for 2019
- a) A json file of world countries and a .csv file of country codes were uploaded
 - b) The two files were merged with the happiness score and country name data, and once the country codes were added, the full name of the country was dropped

```
#Merge json and country code CSV files together to create the data to display on the map
data_to_plot = pd.merge(left= country_codes[['Alpha-3 code', 'Country or region']],
                       right= country_score_avg[['Score', 'Country or region']],
                       on = ['Country or region'],
                       how='inner')
data_to_plot.drop(labels = 'Country or region', axis = 1, inplace = True)
```

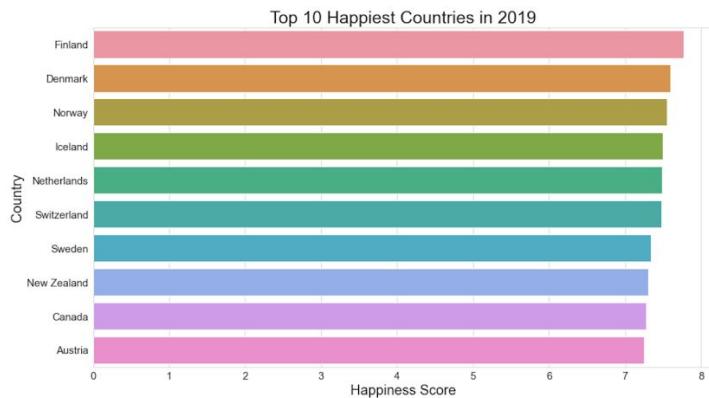
```
#Creating the map using Folium
my_map = folium.Map(location=[10, 6], zoom_start=1.49)

my_map.choropleth(geo_data=country_geo, data=data_to_plot,
                  name='choropleth',
                  columns=['Alpha-3 code', 'Score'],
                  key_on='feature.id',
                  fill_color='BuPu', fill_opacity=0.5, line_opacity=0.2,
                  nan_fill_color="white",
                  nan_fill_opacity=0.5,
                  highlight=True,
                  legend_name='Average Happiness Indicator')

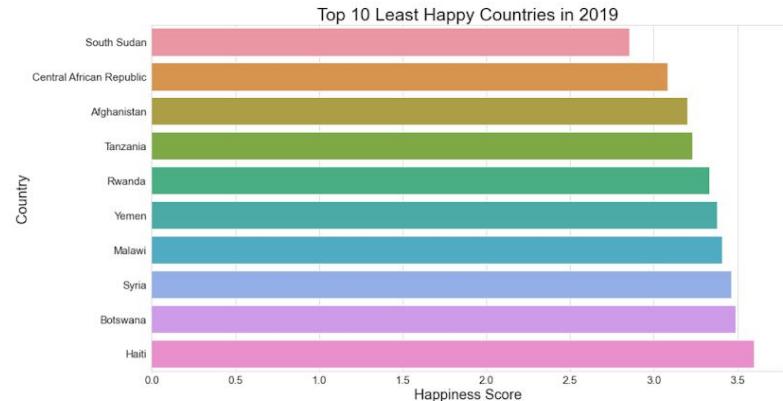
folium.LayerControl().add_to(my_map)
display(my_map)
```

Results

The top 10 happiest countries:

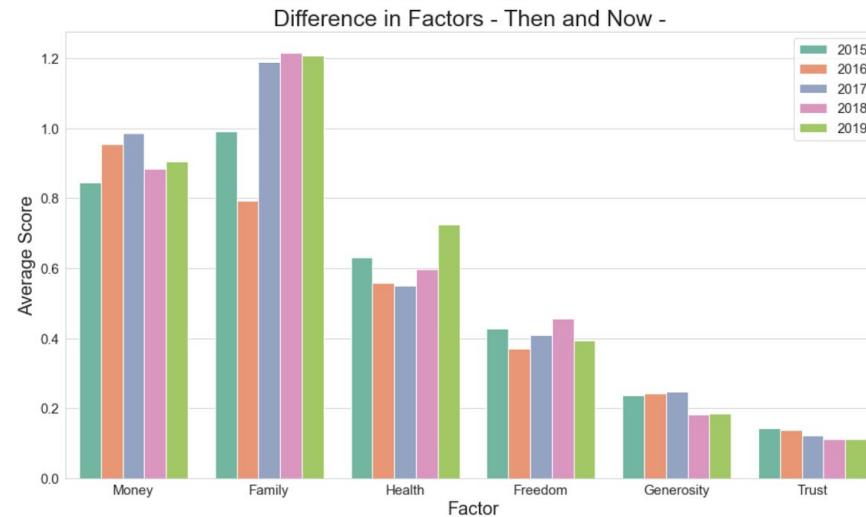


The top 10 least happy countries:



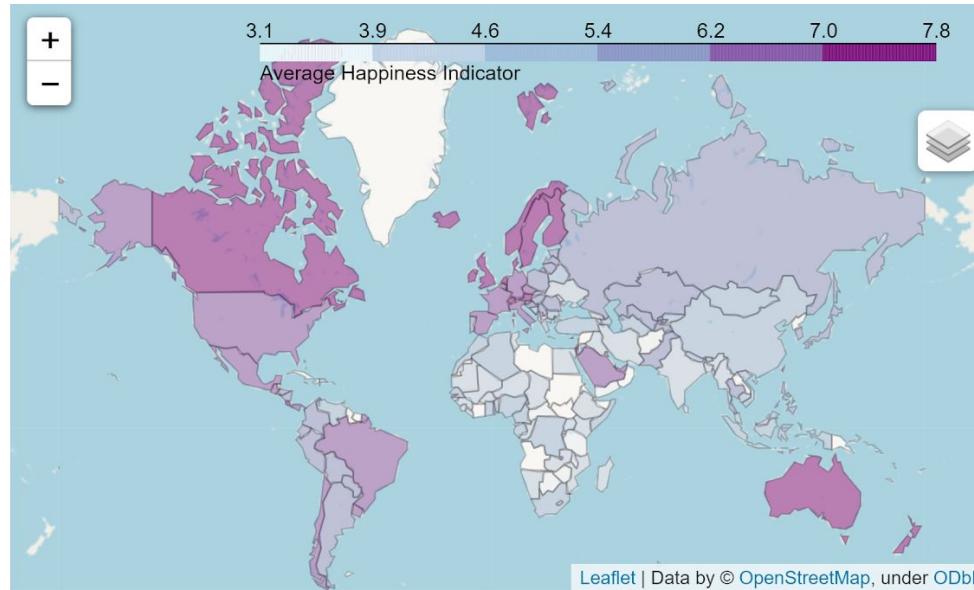
Results

Shows how important each factor is compared to other years.



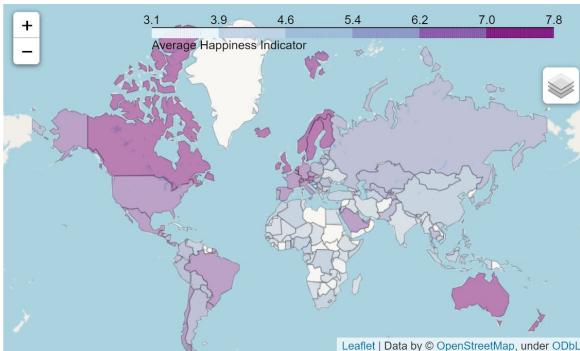
Results

Global choropleth map of happiness score using Folium

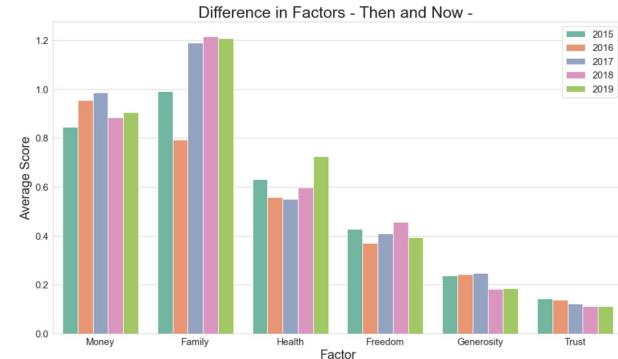


Conclusions

1. The Nordics, the West and Australia are the happiest.
2. East Europe and the majority of Asia is in the middle.
3. The South is the least happiest.



Across all years, social support (family) is valued the most, followed by GDP per capita (money), healthy life expectancy, freedom to make life choices, generosity and perceptions of corruption (trust)



Future Work and Acknowledgements

Future Work: When the 2020 World Happiness Report comes out, it would be interesting to see the differences in factors from previous years, given the effects of the Corona Virus pandemic. My prediction is that the importance score for social support and health will skyrocket.

Acknowledgements: Andrew Tan and Kim Jaroensri