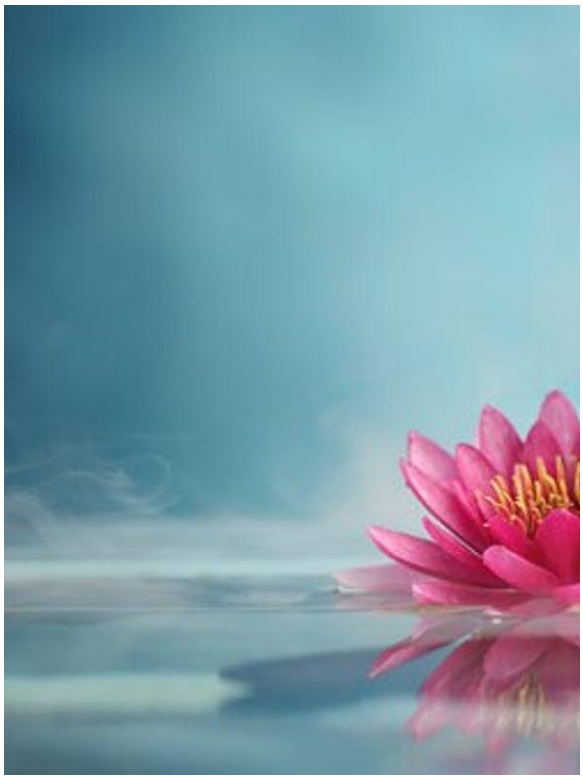


Traits of World Happiness

HOW GEOGRAPHY, ECONOMICS, HEALTH, EDUCATION, AND DEMOGRAPHICS IMPACT HAPPINESS AROUND THE WORLD.

by Ahasan Hossain, Chris Lichliter, David Obembe, and Jack Tambert

BACKGROUND

We've all heard the adage "You can't buy happiness. However, when measuring a country's development, the most used metric is GDP. If the ultimate purpose of a country is to provide for its citizens, then should happiness be a better indicator of success? America's founding fathers believed that "the pursuit of happiness" is an "unalienable right." In 2008, Bhutan went further, creating the metric "Gross National Happiness (GNH)," and amended their constitution to "promote conditions that will enable the pursuit of GNH."[1]

If happiness is a better indicator of a country's success. Then it is important to explore the different factors in what makes a country's citizens happy. The following report will explore data concerning each country's geography, economics, health, education, and demographics, and their relationships with happiness.

HYPOTHESIS

We believe that countries that are wealthier, healthier, more educated, and have more beaches, woods, and mountains to play in, will be happier.

SOURCES OF DATA

We collected data from two sources. The first source comes from Kaggle, containing five years of data from the World Happiness Report.[2] This report employs the Gallup World Poll to survey people from around the world about happiness. Subjects are asked "How would you rate your happiness on a scale of 0 to 10 where 10 is the happiest?"

The next data set to compare against a country's happiness comes from the CIA World Factbook. The CIA World Factbook uses multiple sources to produce an almanac of data for 260 countries and entities. In 2017, Ian Coleman converted this text into a .json file.[3]

DATA CLEANING AND EXPLORATION

The first problem between combining these data set concerned their timepoints. The happiness data had annual scores from 2015-2019. The Factbook data easily available was

---

[1] Ura, D. K., Alkire, S., & Zangmo, T. (2011). The Gross National Happiness Index of Bhutan: Method and Illustrative Results. Thimphu: The Centre for Bhutan Studies.

[2] Sustainable Development Solutions Network. (2019, November 27). World Happiness Report. Retrieved May 1, 2020, from https://www.kaggle.com/unsdsn/world-happiness

[3] Coleman, I. (2017, July 6). Exploring the CIA World Factbook. Retrieved May 1, 2020, from https://iancoleman.io/exploring-the-cia-world-factbook/

primarily from 2019. From this we used two different methods of exploration, trends of happiness over time, and the five-year average.

To explore how happiness changes over time, we performed a groupby to select data by country name. Each country was then plotted on a line graph.

The second method of exploration proved more difficult. First, we used groupby to select each country, and take the mean of its happiness score over time. To retrieve data from the factbook, we manually went through the directory from a test country to find paths for types of data deemed worth exploring. These paths were defined as lists. We then looped through all countries to append data to each list. These lists and countries were then converted into a .csv file.

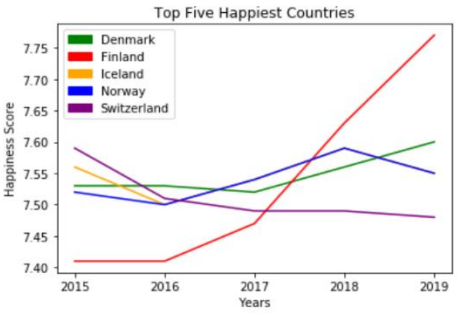
To merge these data, we first had to clean the formatting of country names. This required to change all letters to lower case in the happiness data and change underscores to spaces in the factbook data. An inner join was used for the merge in order to exclude countries not in both data sets.

With the data clean and in the same data frame, we could then explore how happiness relates to different variables. This was done with scatter plots and calculating linear regressions.

ANALYSIS

CHANGES OVER TIME

When looking at individual countries and their change in happiness over time, they tend to not vary greatly. Of the top five countries, the largest change over the five-year period belonged to Finland. Their rating went from 7.4 to 7.75 (Fig 1). This is a total increase of less than 5%. It is unknown what caused this increase, but we can speculate after looking at how happiness correlates to factors from the Factbook, and outside research on any changes in Finnish society.



CORRELATION TO HAPPINESS

While we discovered that no individual variable had a very strong relationship, there was a group that had moderate relationships, and three that could be considered somewhat strong. Three of the five strongest relationships are related to infrastructure (Internet, electricity, and sanitation access). It is also important to note that geography seems to be a poor indicator of happiness. Of our measures, only coastline as a percent of total borders had a weak relationship (corr coeff of 0.32). The other geographic measures have no meaningful correlation with happiness.

VARIABLE	Corr. Coef.
Dependency Ratio	-0.64
Fertility Rate	-0.60
Obesity Rate	0.55
Median Age	0.55
Clean Water Access	0.59
Literacy	0.61
Years of School	0.62
Sanitation Access	0.66
PPPpC	0.68
Internet Access	0.70
Life Expectancy	0.70
Electricity Access	0.70

PREDICTION MODELING

While the Happiness Report included data from 152 countries and entities, the World Factbook has data for 260 entities. Using the combined data, we were able to create a multilinear regression model to predict the unknown happiness scores from other countries. This model was created with the Statsmodel library for Python. Our model used the following variables:

- Electricity Access
- PPPpC
- Gross National Savings
- Unemployment
- Forest as a Percent of Total Land
- Sex Ratio (M:F)
- Dependency Ratio
- Median Age

This model has a R^2 of 0.67 and an adjusted R^2 of 0.64. When plotting our model's predicted values against actual happiness scores, the model appears to underpredict scores as countries get happier, and over predict scores as countries get sadder.

To test this model with country not included in the World Happiness Report, we looked at Barbados. Inputting data from the Factbook for our model, we predict that Barbados should have a happiness score of 5.31. Comme ci comme ça.

CONCLUSIONS

When looking at ways to improve one's country, it is prudent to know what is most effective. In an ideal world, a state could work to improve all these metrics. However, in times of austerity, it is vital to know what projects are most cost effective. Our study shows that the strongest correlation is between happiness and utility access. Internet, Electricity, and Sanitation are among the best correlating variables. Education also seems to be somewhat related. Surprisingly, while PPP was a decent indicator of happiness, unemployment is not. This could suggest that a state's wealth is a better predictor than an individual's wealth. Maybe South Africa should focus on improving ESKOM's regular power outages before tackling unemployment.

LIMITATIONS AND FUTURE ANALYSIS

When dealing with subjective metrics like happiness, there is greater room for bias in survey design. We also do not know how the samples were weighted to represent the larger population. A better understanding of the study design would help identify any unintended biases.

Our study only includes the most recent data from the CIA World Factbook. While this is a tremendous resource, it limited our ability to explore why happiness changes over time. Collecting data from previous editions of the Factbook, or other sources would help better understand those changes.

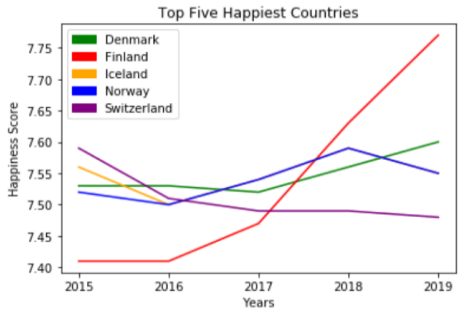
With enough time, it would be interesting to explore similar countries. A large number of countries that have 100% access to electricity, internet, clean water, and sanitation. Knowing what drives happiness in those countries may not be the same as less developed countries.

With greater access to resources, it would be interesting to see what relationship weather has with happiness. While that might not provide insight to state management, it could be useful in deciding where to start a psychiatry practice.

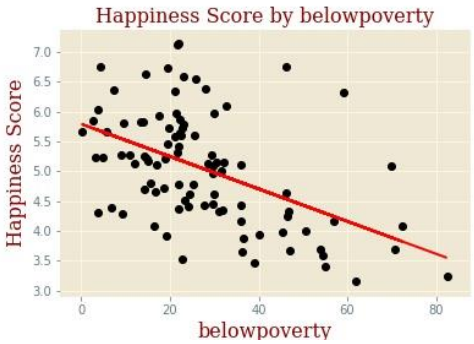
On a personal level, a better understanding of statistics would likely enable the creation of a more accurate prediction model.

APPENDIX

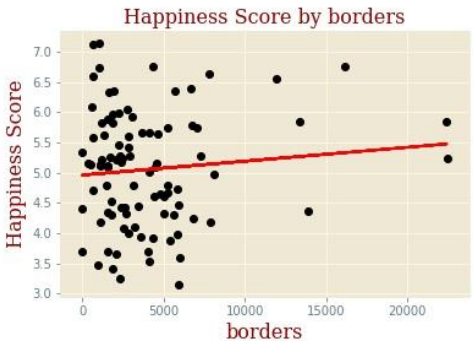
Happiest Countries Over Time



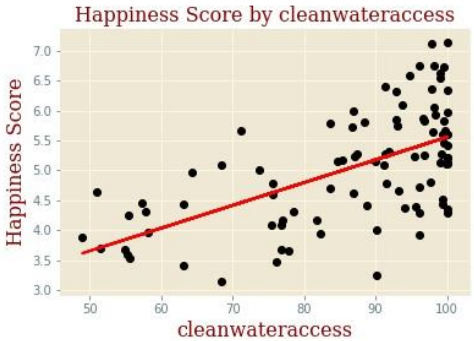
Happiness vs Poverty Rate



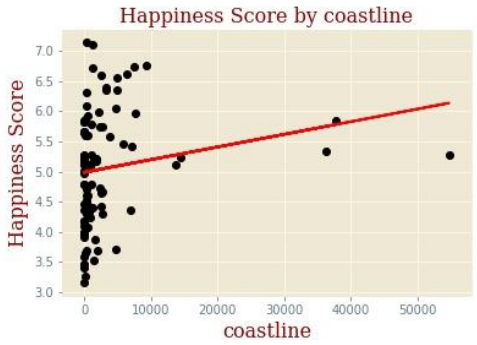
Happiness vs Border Length



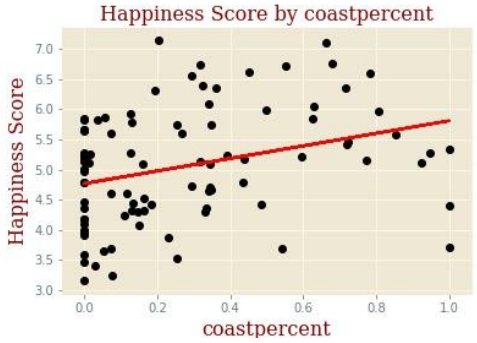
Happiness vs Access to Clean Water



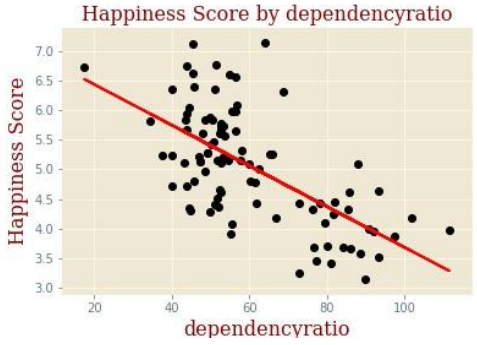
Happiness vs Total Coastline



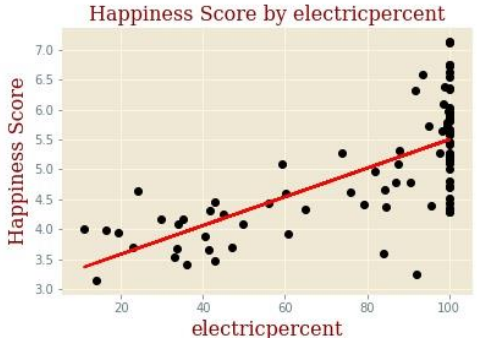
Happiness vs Coast as percent of Borders



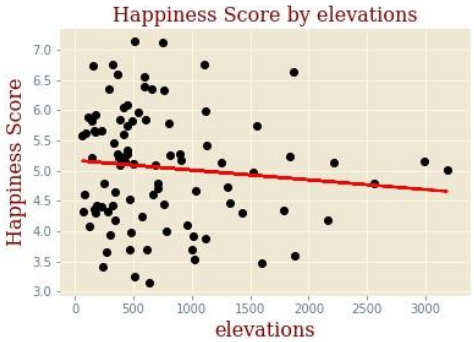
Happiness vs Dependency Ratio



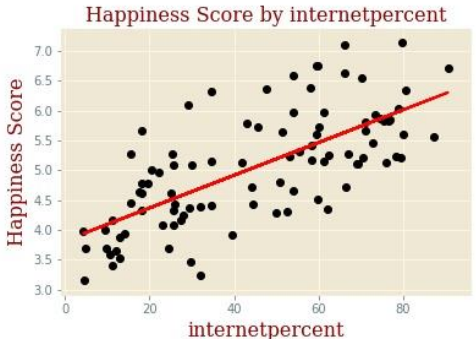
Happiness vs Electricity Access (%)



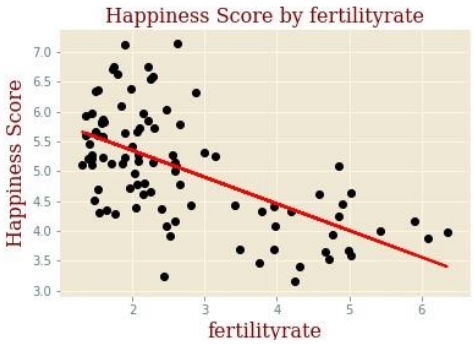
Happiness by Mean Elevation



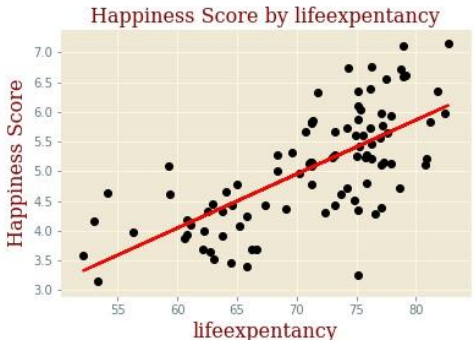
Happiness by Internet Access (%)



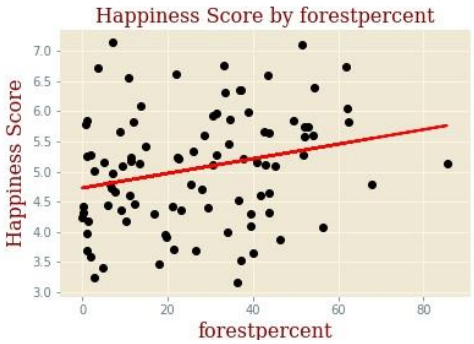
Happiness by Fertility Rate



Happiness by Life Expectancy

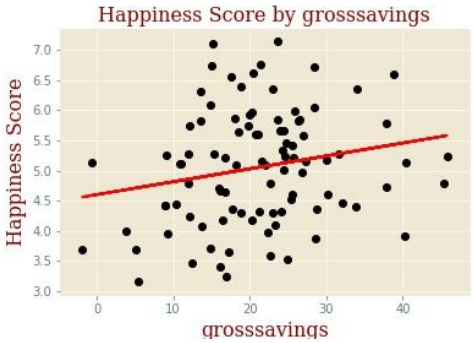


Happiness by Forest as a % of Total Land

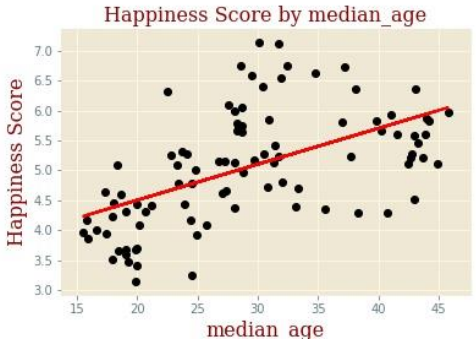


Happiness by Literacy

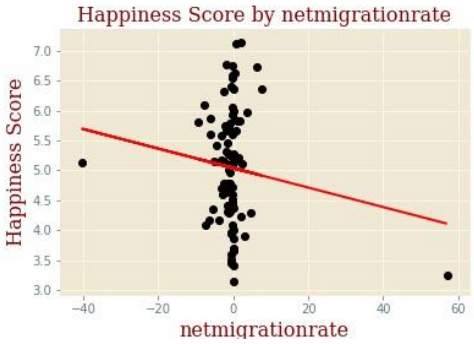
Happiness by Gross National Savings



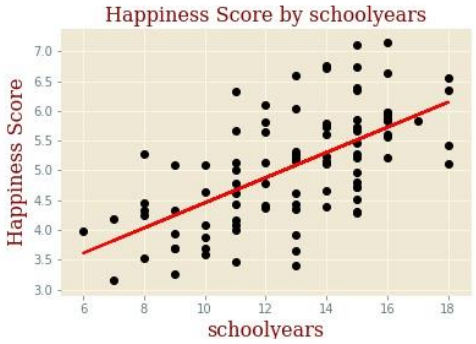
Happiness by Median Age



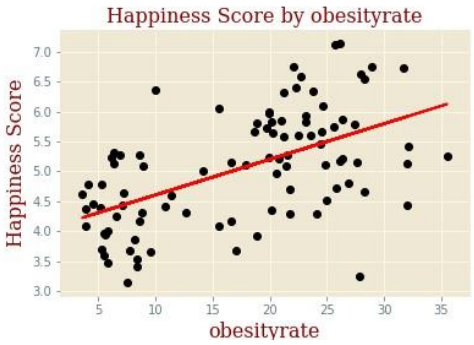
Happiness by Net Migration Rate



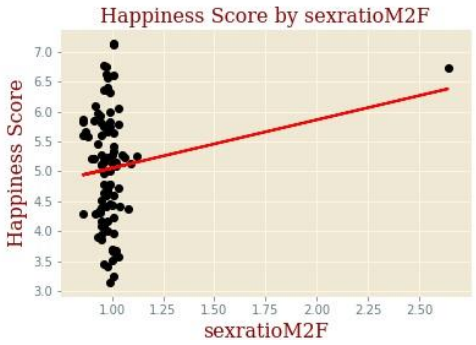
Happiness by Length of School Life



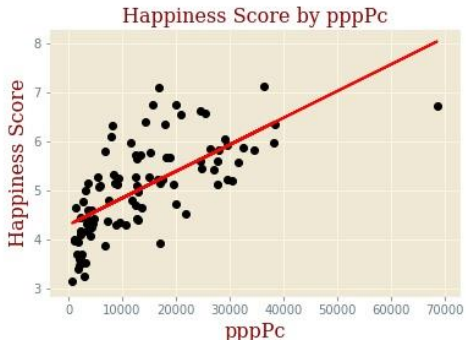
Happiness by Obesity Rate



Happiness by Sex Ratio (M:F)



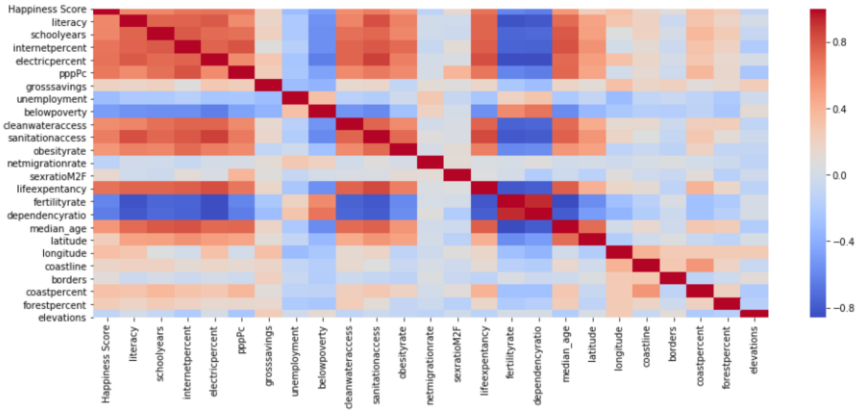
Happiness by PPPpC



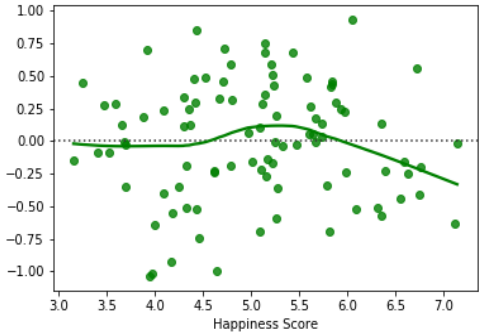
Happiness by Unemployment Rate

Happiness by Sanitation Access (%)

Correlation Heatmap



Residuals



Dep. Variable:	Happiness Score	**R-squared:**	0.670
Model:	OLS	**Adj. R-squared:**	0.635
Method:	Least Squares	**F-statistic:**	18.97
Date:	Thu, 14 May 2020	**Prob (F-statistic):**	7.35e-17
Time:	10:08:00	**Log-Likelihood:**	-75.847
No. Observations:	94	**AIC:**	171.7
Df Residuals:	84	**BIC:**	197.1
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	6.4467	1.414	4.559	0.000	3.634	9.259
electricpercent	0.0159	0.004	3.564	0.001	0.007	0.025
pppPc	5.735e-05	1.03e-05	5.576	0.000	3.69e-05	7.78e-05
grosssavings	-0.0095	0.007	-1.320	0.190	-0.024	0.005
unemployment	-0.0078	0.006	-1.375	0.173	-0.019	0.003
cleanwateraccess	0.0026	0.007	0.370	0.712	-0.011	0.016
sexratioM2F	-1.0254	0.485	-2.116	0.037	-1.989	-0.062
dependencyratio	-0.0124	0.008	-1.516	0.133	-0.029	0.004
median_age	-0.0625	0.017	-3.752	0.000	-0.096	-0.029
forestpercent	0.0063	0.003	1.897	0.061	-0.000	0.013

Omnibus:	3.607	**Durbin-Watson:**	1.392
Prob(Omnibus):	0.165	**Jarque-Bera (JB):**	2.272
Skew:	0.166	**Prob(JB):**	0.321
Kurtosis:	2.314	**Cond. No.**	4.45e+05