Happiness Report

Siddharth Bhatia, Audrey Wang, Sanjana Subramanyan, and Leo Ding

Table of Contents

Big Picture

As a society, the United States has always emphasized our economic growth and establishing ourselves as a global leader. But what is all of that for? Shouldn't we prioritize the happiness of the people? In terms of citizen happiness, Costa Rica is ranked 15th out of 153 countries, but their people's income and GDP per capita are not even close to being on the top. How about Bhutan, a country that ranks Gross Domestic Happiness so much higher than GDP that they enshrine it in their constitution? Through the United Nations World Happiness Report, we investigate the factors that contribute to a country's happiness to see what is actually vital for the people of this world.

Dataset

For the research, we used data from the World Happiness Report, obtained from Kaggle. The dataset is significant because it tracks a measure of a country's progress for policy makers, as well as it being a point of comparison for how well a country is doing compared to its neighbors. The data ranges from 2015 to 2019 and contains information on 158 countries. The dataset contains happiness data in the form of a "Happiness Score", a measure which asks respondents to rate their lives from 0 to 10. In addition, the dataset includes country statistics on GDP, Freedoms, Life Expectancy, Generosity, Trust, Corruption and, in some cases, Social Support. In the context of the dataset, these factors are estimated to form life evaluations.

The World Happiness Report (WHR) is published yearly by the United Nations Sustainable Development Solutions Network, using data primarily from Gallup World Poll. The WHR's purpose is to rank countries and compare happiness data to quality of life factors. The Kaggle dataset was published by the UN, not a third party. It was last updated in 2019, to add the 2019 data. We also used data from 2020, obtained directly from the World Happiness Report website. The 2020 report contained all of the columns found in previous years, which made it easy to compare.

Data Sample (2015 and 2020)

Country	Region	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	Trust (Government Corruption)	Generosity
Switzerland	Western Europe	1	7.587	0.03411	1.39651	1.34951	0.94143	0.66557	0.41978	0.29678
Iceland	Western Europe	2	7.561	0.04884	1.30232	1.40223	0.94784	0.62877	0.14145	0.4363
Denmark	Western Europe	3	7.527	0.03328	1.32548	1.36058	0.87464	0.64938	0.48357	0.34139
Norway	Western Europe	4	7.522	0.0388	1.459	1.33095	0.88521	0.66973	0.36503	0.34699
Canada	North America	5	7.427	0.03553	1.32629	1.32261	0.90563	0.63297	0.32957	0.45811
Finland	Western Europe	6	7.406	0.0314	1.29025	1.31826	0.88911	0.64169	0.41372	0.23351
Netherlands	Western Europe	7	7.378	0.02799	1.32944	1.28017	0.89284	0.61576	0.31814	0.4761
Sweden	Western Europe	8	7.364	0.03157	1.33171	1.28907	0.91087	0.6598	0.43844	0.36262
New Zealand	Australia and New Zealand	9	7.286	0.03371	1.25018	1.31967	0.90837	0.63938	0.42922	0.47501
Australia	Australia and New Zealand	10	7.284	0.04083	1.33358	1.30923	0.93156	0.65124	0.35637	0.43562

Dataset (2015-2019): https://www.kaggle.com/unsdsn/world-happiness

2020									
Rank	Country	Region	Score	Logged GDP	Social Support	Health (Life Expectancy)	Freedom	Generosity	Corruption
1	Finland	Western Europe	7.808700085	10.63926697	0.954329729	71.9008255	0.949172199	-0.059482019	0.195444584
2	Denmark	Western Europe	7.645599842	10.77400112	0.955990791	72.40250397	0.951444268	0.066201776	0.168489456
3	Switzerland	Western Europe	7.559899807	10.97993279	0.942846596	74.10244751	0.921336651	0.105911039	0.303728431
4	Iceland	Western Europe	7.504499912	10.77255917	0.974669576	73	0.948891878	0.246944219	0.711709738
5	Norway	Western Europe	7.487999916	11.08780384	0.952486575	73.20078278	0.955750287	0.13453263	0.263218194
6	Netherlands	Western Europe	7.448900223	10.81271172	0.93913883	72.30091858	0.908547819	0.20761244	0.364717126
7	Sweden	Western Europe	7.353499889	10.75879383	0.926311195	72.60076904	0.939144194	0.111614622	0.250880182
8	New Zealand	North America and ANZ	7.299600124	10.50094318	0.949118972	73.20262909	0.936217487	0.191598073	0.22113885
9	Austria	Western Europe	7.294199944	10.7428236	0.928045869	73.00250244	0.899989486	0.085429221	0.499954879
10	Luxembourg	Western Europe	7.237500191	11.45068073	0.906912208	72.59999847	0.90563643	-0.004620588	0.367084295

2020 Data:
https://happiness-report.s3.amazonaws.com/2020/WHR20_DataForTable2.1.xls

World Happiness Report website: https://worldhappiness.report

From this data sample of the top 10 countries from 2015 and 2020, the names of the columns, formatting, and values were varied.

Questions

To assist with our research, we listed out a group of questions that would be answered by the end of the report using our analysis:

1. **Are there general regions/continents of the world that are happier than others?**
2. **What factors seem the most crucial in contributing to the overall happiness of a country? Which ones aren't as significant?**
3. **Can we predict happiness? How accurately? What's the best method?**
4. **Which countries' happiness doesn't seem explainable by the data? What unforeseeable factors could be involved?**
5. **Are there trends in world happiness? Is the world getting happier or sadder?**
6. **Which countries' happiness have changed the most over time? Does the data explain the shift?**
7. **If you look at the table from 2019, 5 out of the 10 top countries are in Northern Europe, why is this and what did they do right to maintain the happiness of their people?**

Data Cleaning

The World Happiness Report had yet to be cleaned. There were several missing values that were either NA's or zeroes in the datasets from 2015-2019. We converted all the zeros to NA's so R would know how to handle them. The 2020 data was already clean.

Despite the 2015-2019 data being from the same dataset, there were many inconsistencies with columns. One issue was column names. For example, the 2017 data had a column called "Freedom", but the same column was called "Freedom to make life choices" in 2019. Data from 2017 and before had a column about Family, while data after that year did not. The Social Support column was only included starting in 2018. These shifts could be due to the fact that the poll's questions changed slightly over time. To further complicate things, 2020 data used "logged GDP per capita", while the other years had GDP per capita — the two measures were not comparable. Also, the 2020 data measured Health by life expectancy, while other years used a different method for determining Health. Besides that, some years included regional data (eg: Finland categorized in Western Europe), while others didn't. In Excel, we manually added region data to the years which didn't have it and made columns consistent before we imported the data to R.

Moreover, not all countries had data for all years. The country Lesotho wasn't present in 2016. There were other countries that appeared in the data only once, meaning changes in their happiness level over time could not be analyzed. so we removed Trinidad & Tobago, Northern Cyprus, North Macedonia, Gambia, Hong Kong, Maldives, Suriname, and more. In total, we removed 11 data points from the original 158.

```
countryDiff <- countryDiff %>%
  filter(!is.na(Change))
```

R code for filtering out data where we couldn't determine change over time.
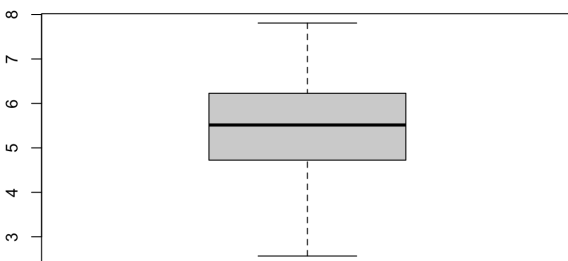
Exploratory Analysis

We used the 2020 data for exploratory analysis, because it was the most recent. The first thing we did was understand how the scores were distributed.

**Happiness Score Distribution**
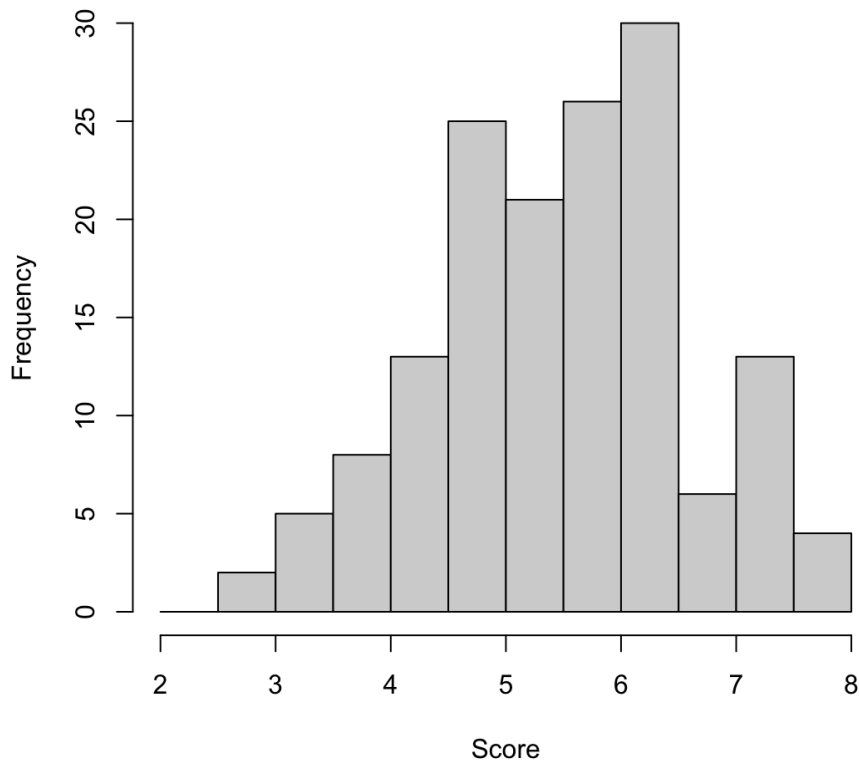
As we can tell, this histogram tells us most scores are in the range 4.5-6.5. There are very few in the high 7's and few below 4.

We can use a box plot to corroborate: half of the countries have scores between 4.75 and 6.25. For comparison, the US has a score of 6.8, which places us in the middle of the fourth quarter.

Next, we plotted the scores on a scatter plot.
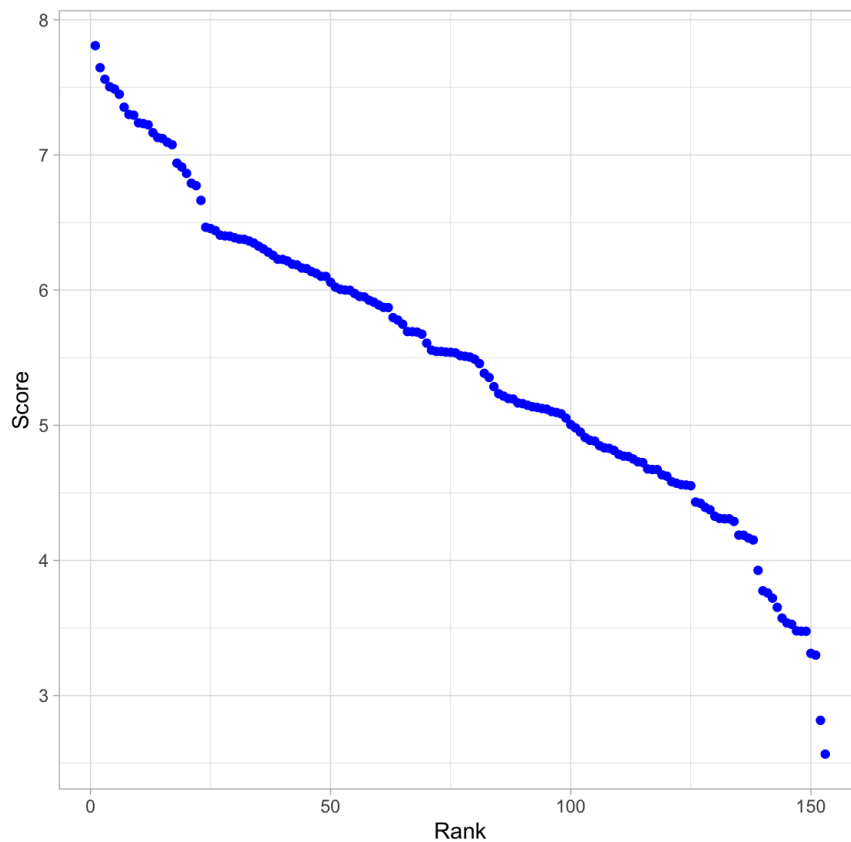
A scatter plot is useful because we can instantly see scores that are far lower or higher than the rest. Cross-referencing with the data, we can see Afghanistan and South Sudan are by far the unhappiest countries, while Finland outperforms everyone else. Although some data points are more distant, there aren't any outliers. The data isn't arranged in a straight line: there is a huge uptrend and downtrend at the beginning and ends of the ranks, respectively. The more gradual slope in the middle section could be due to a group of countries on both extremes, some "really happy", and some "really sad". It makes sense that Afghanistan and South Sudan are the only countries under 3. Both are going through a rough patch in their history, which would rank them especially low, at least for this year.

Lastly, we graphed each variable in relation to the score to see what kind of relationship there was.

Social Support vs Score
R-squared: 0.585226157776264.
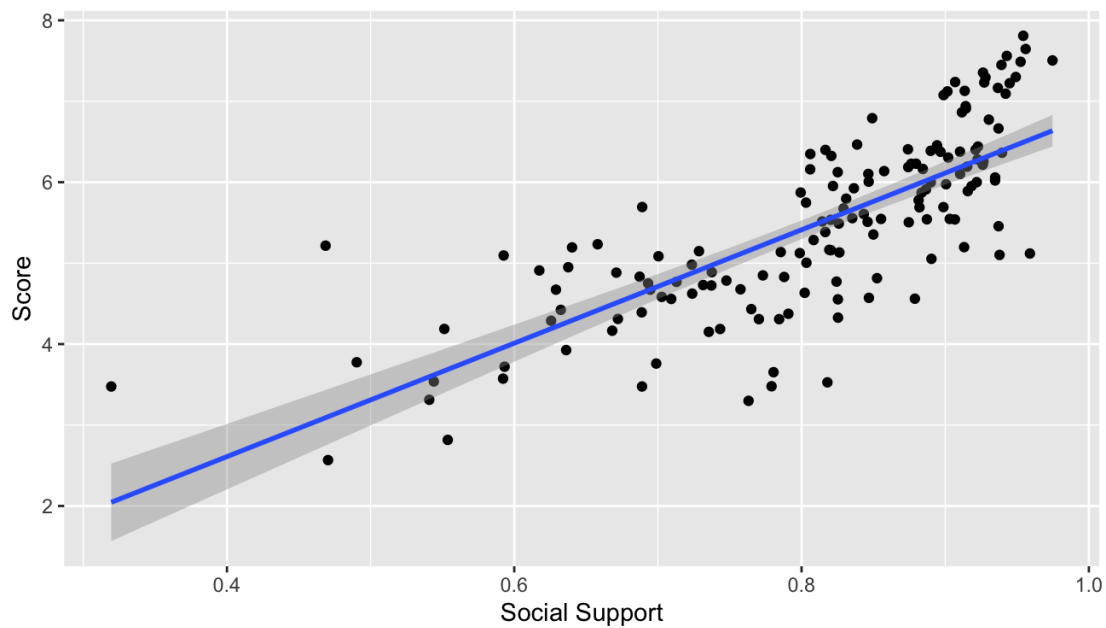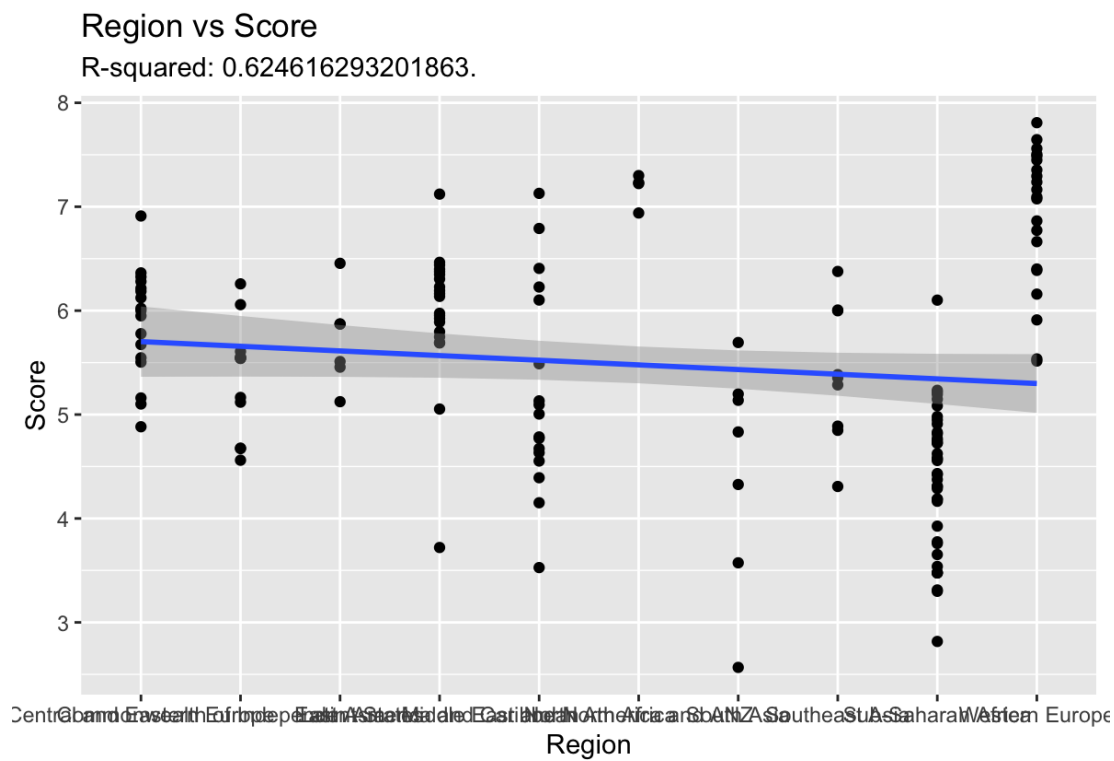

Corruption vs Score
R-squared: 0.174979146024174.

Generosity vs Score
R-squared: 0.00476695332676614.
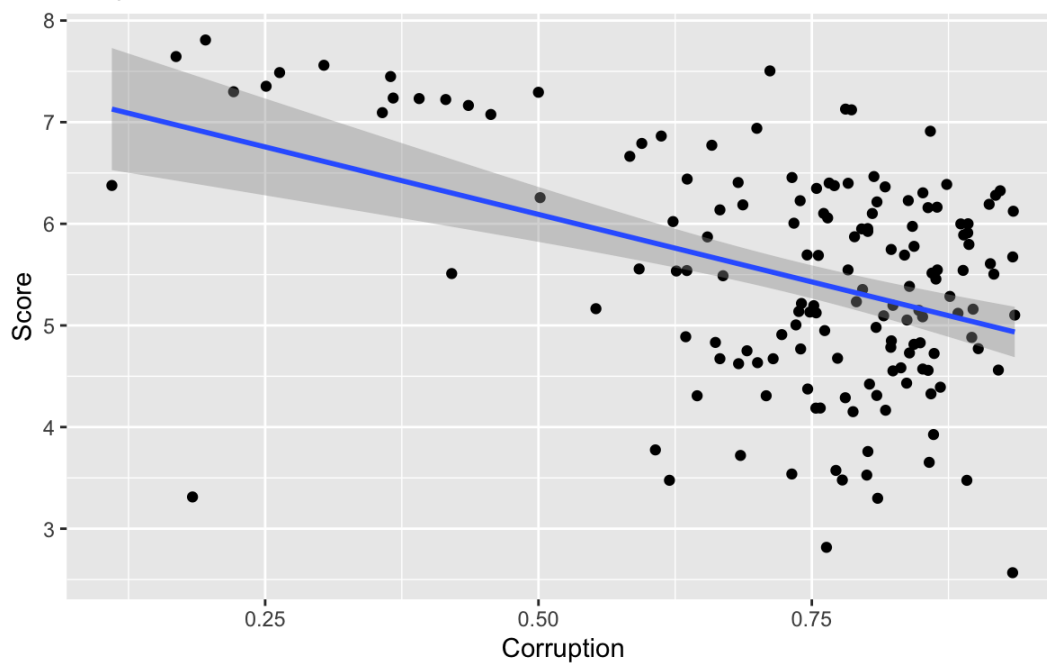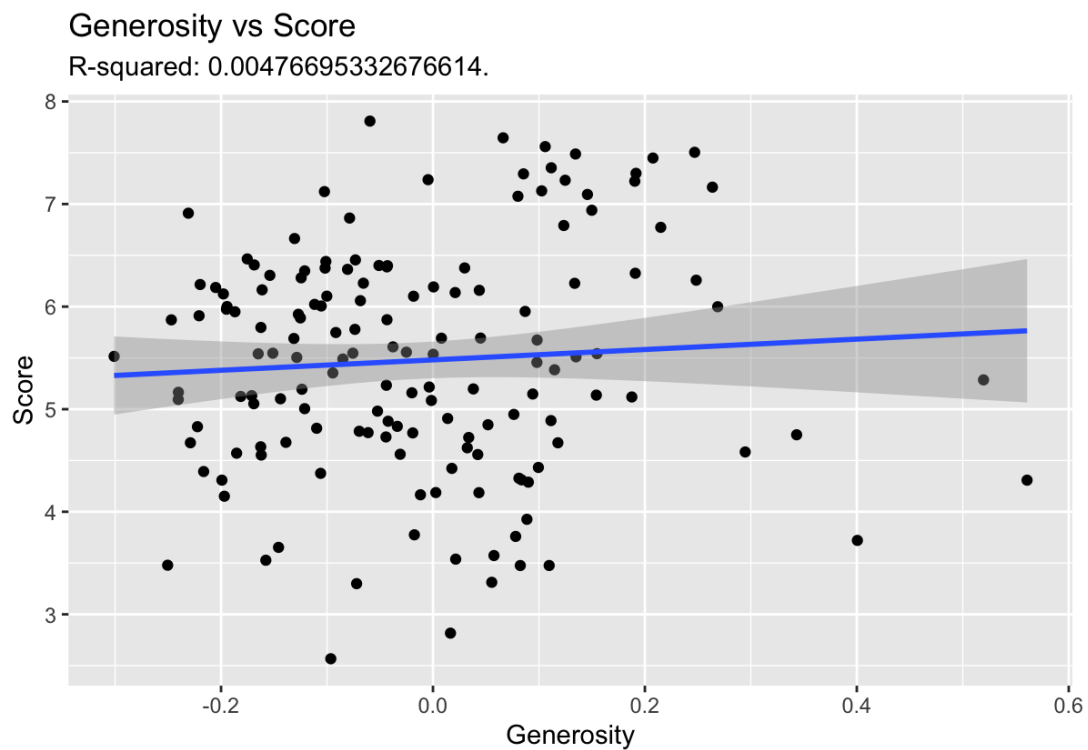
Freedom vs Score
R-squared: 0.348804559998511.
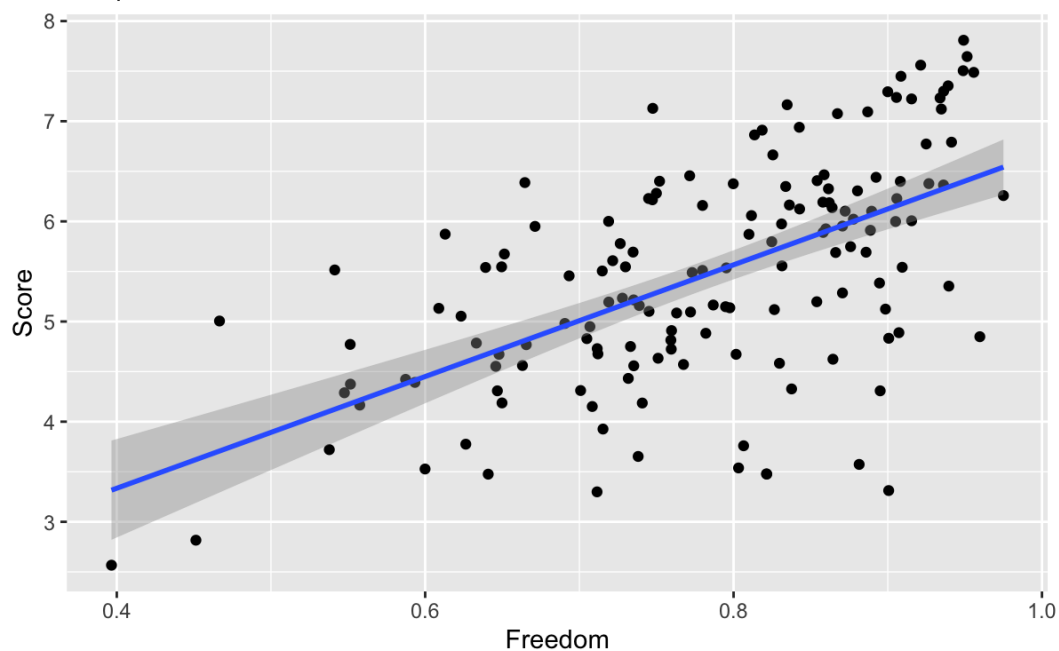
Health (Life Expectancy) vs Score
R-squared: 0.593387186441932.

Region vs Score
R-squared: 0.624616293201863.

Logged GDP vs Score
R-squared: 0.601205461342584.
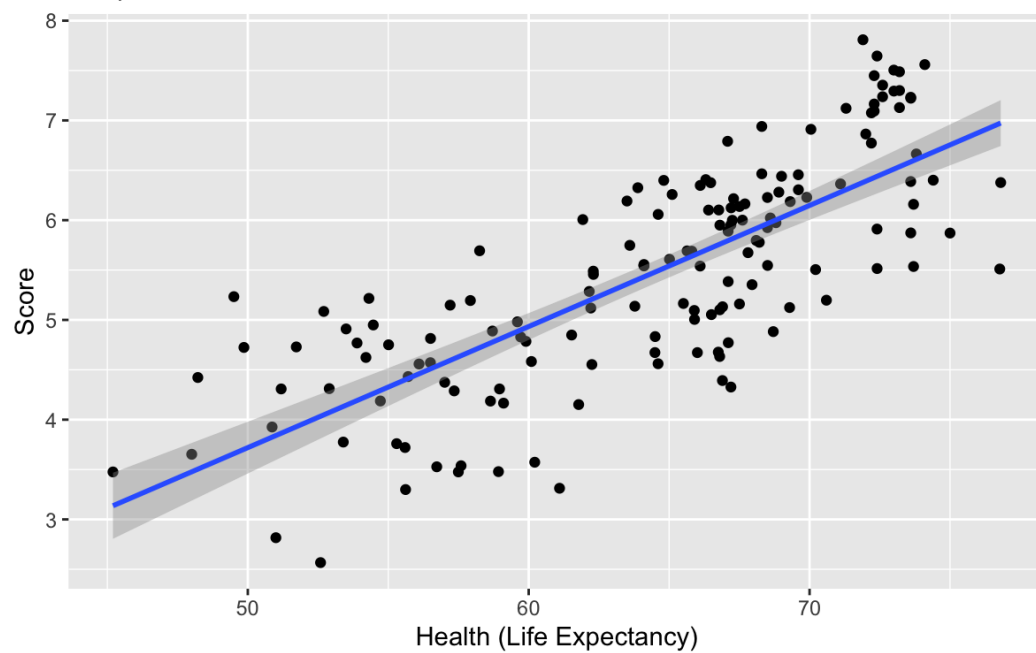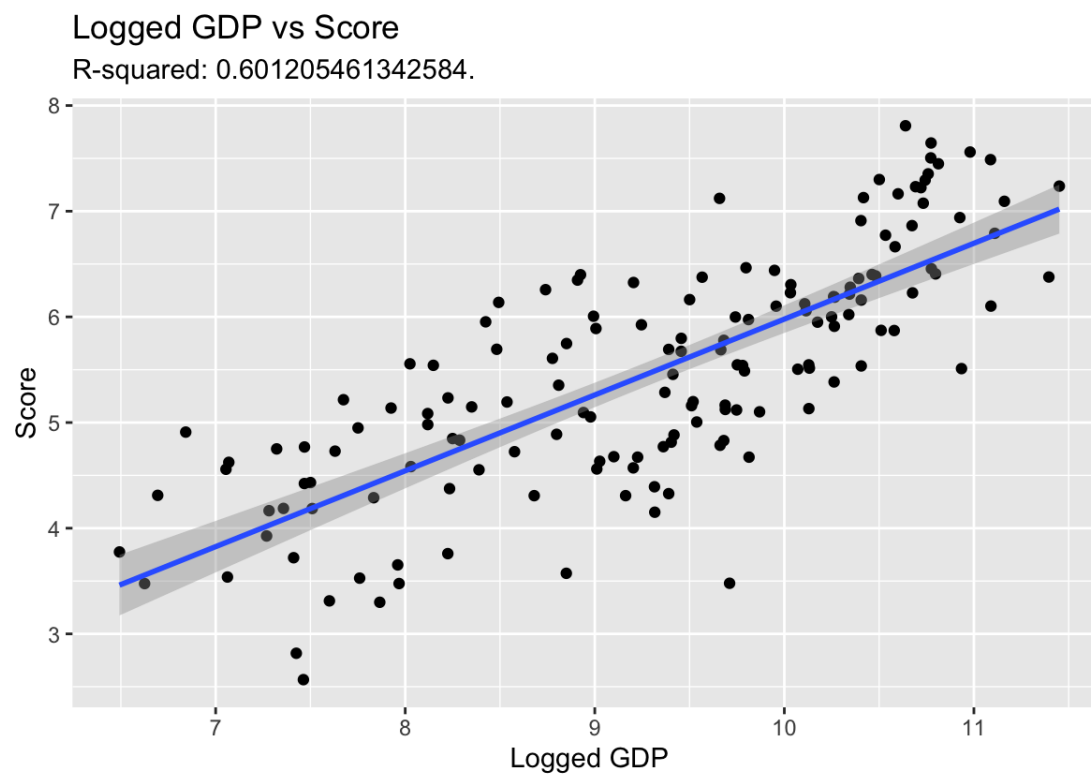
Data Reduction

The 2020 data had quite a few columns that we didn't need.

We pared it down so it consisted of the same columns as the rest of the data.

Rank	Country	Region	Score	Logged GDP	Social Suppo	Health (Life I	Freedom	Generosity	Corruption	Dystopia + residual
1	Finland	Western Eur	7.80870009	10.639267	0.95432973	71.9008255	0.9491722	-0.059482	0.19544458	2.76283503
2	Denmark	Western Eur	7.64559984	10.7740011	0.95599079	72.402504	0.95144427	0.06620178	0.16848946	2.43274069
3	Switzerland	Western Eur	7.55989981	10.9799328	0.9428466	74.1024475	0.92133665	0.10591104	0.30372843	2.35026741
4	Iceland	Western Eur	7.50449991	10.7725592	0.97466958	73	0.94889188	0.24694422	0.71170974	2.46068811
5	Norway	Western Eur	7.48799992	11.0878038	0.95248658	73.2007828	0.95575029	0.13453263	0.26321819	2.1682663
6	Netherlands	Western Eur	7.44890022	10.8127117	0.93913883	72.3009186	0.90854782	0.20761244	0.36471713	2.35211706
7	Sweden	Western Eur	7.35349989	10.7587938	0.9263112	72.600769	0.93914419	0.11161462	0.25088018	2.24629927
8	New Zealand	North Ameri	7.29960012	10.5009432	0.94911897	73.2026291	0.93621749	0.19159807	0.22113885	2.12810803
9	Austria	Western Eur	7.29419994	10.7428236	0.92804587	73.0025024	0.89998949	0.08542922	0.49995488	2.39844608
10	Luxembourg	Western Eur	7.23750019	11.4506807	0.90691221	72.5999985	0.90563643	-0.0046206	0.3670843	2.15370035
11	Canada	North Ameri	7.23210001	10.6923685	0.92717665	73.6016007	0.93391323	0.12477077	0.39084336	2.19526911
12	Australia	North Ameri	7.22279978	10.7205963	0.94485462	73.604538	0.91543174	0.19046016	0.41516912	2.12980413
13	United Kingd	Western Eur	7.16450024	10.6001349	0.93668282	72.3016052	0.83474368	0.26373249	0.43591592	2.23672199
14	Israel	Middle East	7.12860012	10.4175253	0.91357124	73.2002564	0.74758059	0.10259372	0.78085506	2.7133584
15	Costa Rica	Latin Americ	7.12139988	9.6581564	0.90154594	71.2998505	0.93473893	-0.1024297	0.78633243	2.95313501
16	Ireland	Western Eur	7.09369993	11.1609783	0.94208157	72.3007889	0.88698298	0.14578497	0.35718417	1.94387817
17	Germany	Western Eur	7.07579994	10.7328186	0.89887446	72.2020187	0.86737108	0.08017939	0.45642203	2.29524946
18	United State	North Ameri	6.93959999	10.9257689	0.91421902	68.2994995	0.84261954	0.14989173	0.69971532	2.34412479
19	Czech Repub	Central and E	6.91090012	10.4041634	0.91443068	70.0479355	0.81853735	-0.2308615	0.85844624	2.79680824
20	Belgium	Western Eur	6.86350012	10.6736393	0.91163254	72.001648	0.81358206	-0.0786912	0.61229795	2.34862661
21	United Arab	Middle East	6.79080001	11.1099987	0.84918112	67.0827866	0.94134551	0.12345085	0.59450221	2.16696596

The other tables had some statistical columns that we removed. 2015 data had a "Standard Error" column, 2016 had lower and upper confidence intervals, and 2017 data had low and high whiskers. We also removed the dystopia + residual

Standard Error	Lower Confidence Interval	Upper Confidence Interval	Whisker.high	Whisker.low
0.03411	7.46	7.592	7.59444482058287	7.47955553799868
0.04884	7.428	7.59	7.58172806486487	7.46227160707116

column, because it wasn't necessary for our research.

Analysis

Are there regions of the world that are happier than others?

Methodology

To uncover the relationship between Region and Happiness, we created box plots for scores of every region then compared them side-by-side. We found that this format was the best way to make comparisons between different regions. Box plots display a lot of information, so comparing them together is extremely powerful.

Happiness by Region
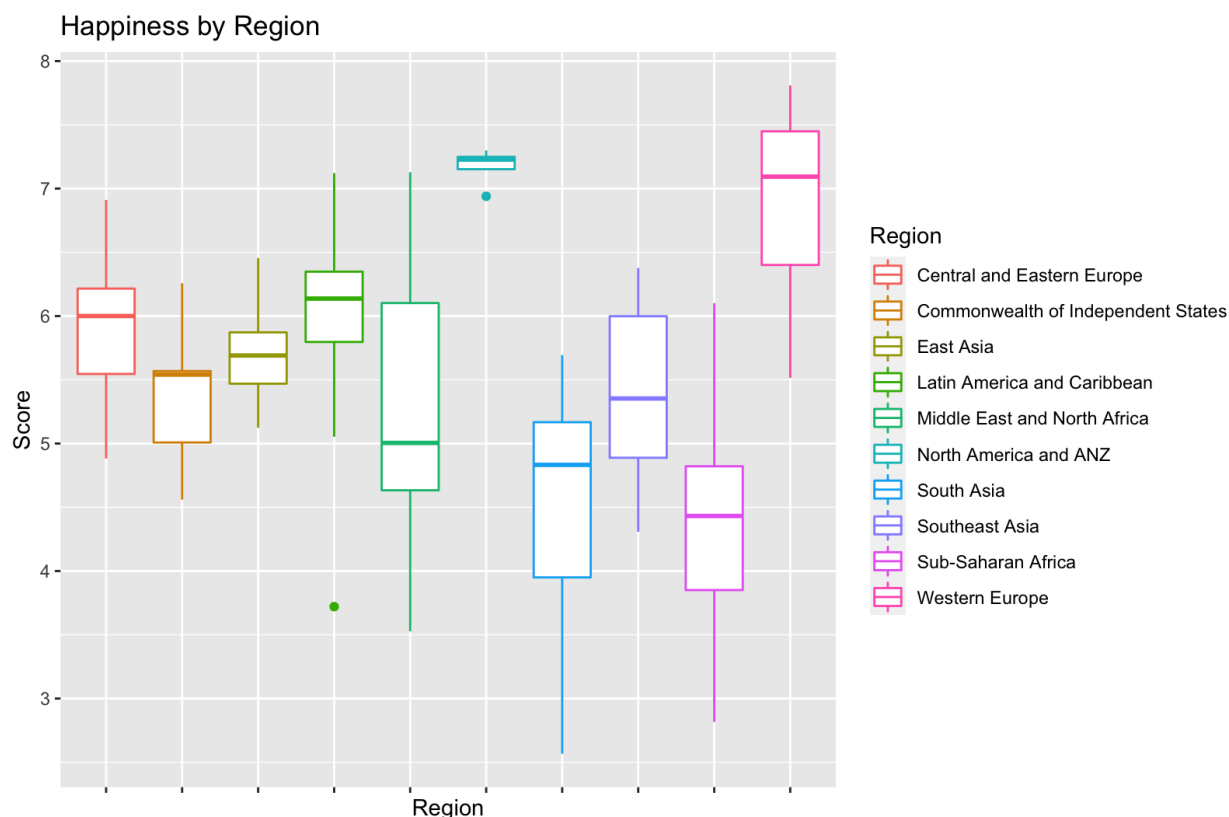
Countries are divided by the United Nations into 10 distinct regions. With this box plot, we can conclude that North America, Australia, and New Zealand is the happiest region overall. Their minimum, Q1, and median are higher than any other region. Also, they have the least variability, possibly due to the small size of the region. While North America and ANZ are doing very well, Western Europe still contains the happiest countries in the world — despite there being more countries in Western Europe than North America, their top quarter of countries are still happier than anywhere else! Many

of the happiest countries, such as Finland, are in Western Europe. 9 out of 10 of the happiest countries are located in the European Union.

On the contrary, Sub-Saharan Africa has the unhappiest median and there is a lot of variability within the region. South Asia comes close as well, with the lowest score pertaining in that region, but their median is still higher than Sub-Saharan Africa. From this graph, we can conclude that a country's region affects their happiness, but there is definitely more to the story.

What factors seem the most crucial in contributing to the overall happiness of a country? Which ones aren't as significant?

One of the most important questions we had was which factors were the most meaningful for determining the happiness of a country. For lawmakers, knowing which elements affect their country's happiness can help guide public policy. GDP per capita is the most important factor in determining happiness, followed by Social Support. The least meaningful factor is generosity.

Methodology

We used all available methods to investigate what affects a country's happiness.

The first method was linear regression. Using regression, we found the R-squared of each attribute and the score. We can see that Region accounts for the most variation in score, followed by GDP, Health, and Social Support close behind. Region also was well represented by the linear regression model; we were surprised to discover that Region had the highest correlation, as it isn't a number. Interestingly, Freedom, Corruption, and Generosity, things that we would associate with happiness, have the least correlation. Unfortunately, none of the factors have a high enough R-squared value to be an accurate predictor on their own.

```
  Attribute                  `Linear R-Squared`
  <chr>                                   <dbl>
1 Region                                  0.625
2 Logged GDP                              0.601
3 Health (Life Expectancy)                0.593
4 Social Support                          0.585
5 Freedom                                 0.349
6 Corruption                              0.175
7 Generosity                            0.00477
```

Another way we can see which attributes matter most is by having the computer generate a decision tree.
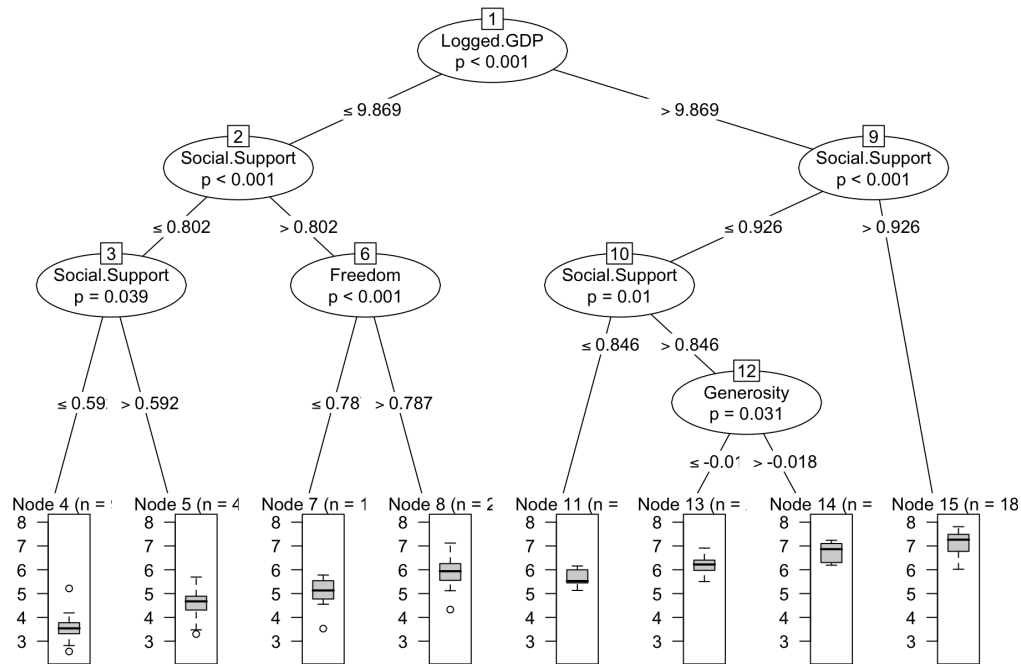
The algorithm found that the most crucial component in determining happiness is GDP per capita, then Social Support. The decision tree differs greatly from the linear model: even though health had one of the strongest linear associations, it didn't appear at all on the decision tree. Also, the tree didn't factor in a country's region, even though it had the strongest linear association.

The third way we can find how much attributes matter is through a random forest.

	Attribute	IncNodePurity
1	Social.Support	42.666217
2	Health	38.402206
3	Logged.GDP	35.851910
4	Region	30.538199
5	Freedom	18.033554
6	Corruption	11.567810
7	Generosity	5.592944

When creating a random forest model, it calculates which variables are most relevant to the output. The model found Social Support to be the most important factor in determining happiness, followed by Health and GDP.

Summary of attributes deemed most important factors

Linear Regression	Decision Tree	Random Forest
Region	GDP per capita	Social Support
GDP per capita	Social Support	Health
Health (Life Expectancy)	Freedom	GDP per capita

While all three methods produced slightly different results, they all agree that Social Support and GDP per capita are some of the most important factors. It was stunning to see that Generosity, Freedom, and Corruption were the least important factors.

Can we predict happiness? How accurately? What's the best method?

Happiness is hard to predict. After testing many different methods, we decided using SVM produced the most accurate results. With SVM, a country's happiness score can be predicted to within half a point 75% of the time.

Methodology

We used three different computer-generated models to predict happiness. The factors we inputted into the model were Social Support, Health, GDP per capita, Region, Freedom, Corruption, and Generosity. Since we weren't doing classification, we couldn't use a simple confusion matrix to calculate the accuracy of the model. Instead, we calculated the difference between each prediction and it's actual value, then evaluated if the difference was under 0.5. If it was, we regarded the prediction as correct. Otherwise, the prediction was wrong.

First, we tried using a decision tree to predict happiness.

Decision Tree Accuracy
68.627% of countries were predicted within 0.5 of their happiness score.
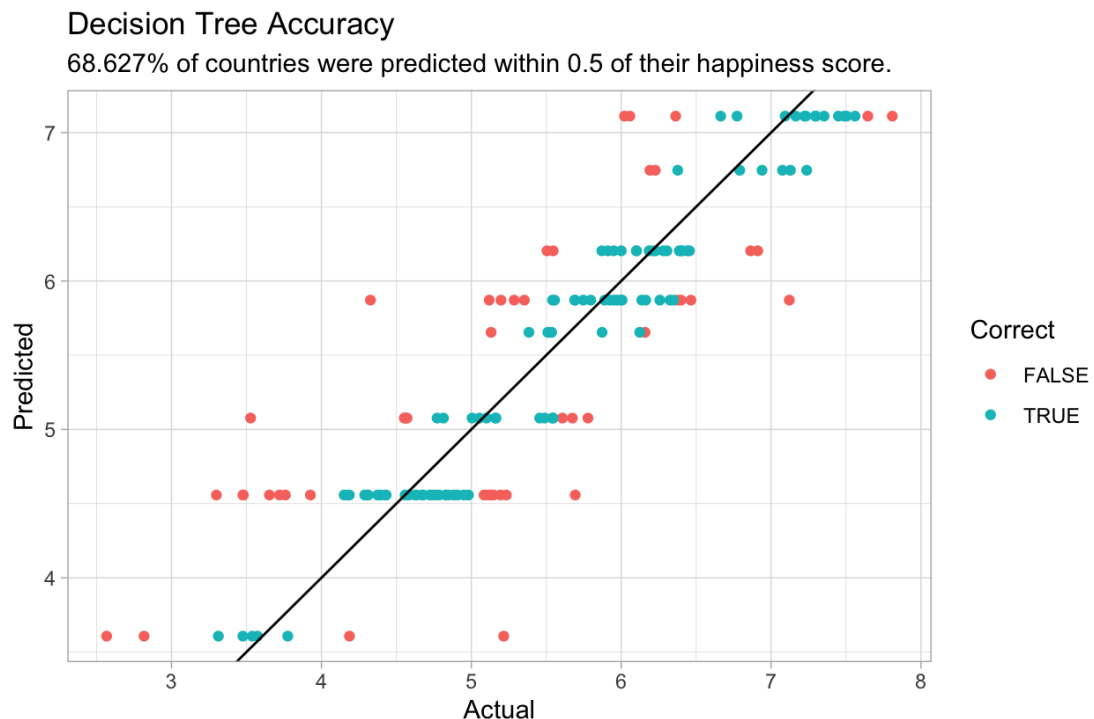
Only 69% of countries were predicted within 0.5 of their happiness score by the decision tree model. The model isn't accurate, but it was interesting to see how the decision tree reached its conclusions.
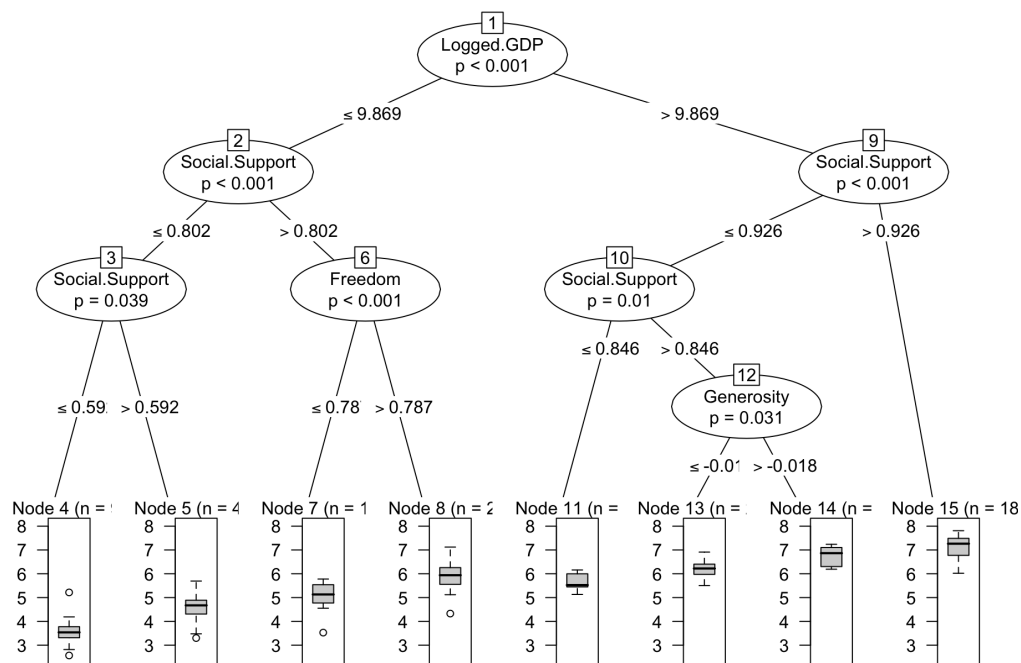
Measures of health, freedom, and region didn't appear at all in the decision tree. The model was too basic to predict such a complicated value.

Next, we ran the data through a Random Forest model.

Random Forest Accuracy

70.588% of countries were predicted within 0.5 of their happiness score.
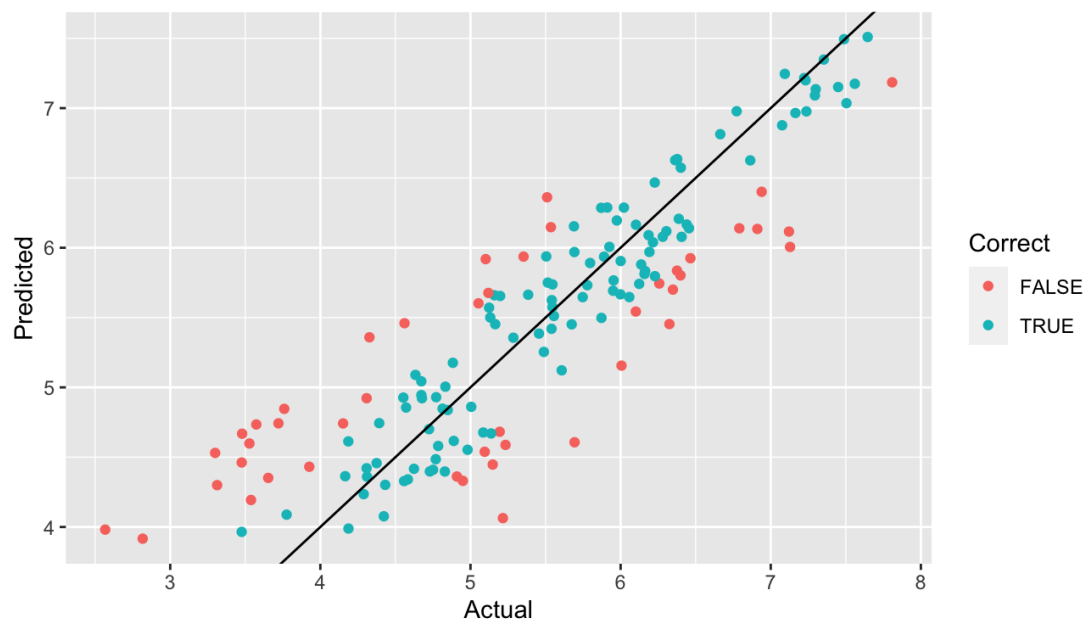
Random forest's precision is slightly greater than decision trees, resting at about 71%. This makes sense because random forests are more capable at handling complicated data.

rf
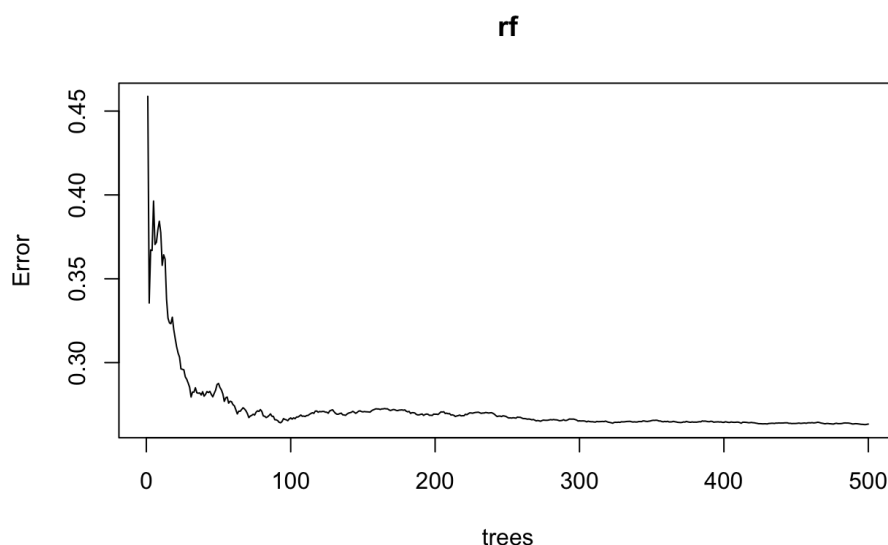
Looking at the error graph for Random Forest, we can see that 130 is a good number of trees: the elbow of the line ends at about 100, so just after that seems like a

good spot. After multiple tests, we concluded that using 130 trees has approximately the same accuracy as using 500.

While the 71% accuracy for random forest is more reliable than decision trees, it can still be higher.

SVM Accuracy
74.51% of countries were predicted within 0.5 of their happiness score.

The SVM model beats out both decision trees and random forests at an accuracy rate of 75%. SVM works in regression and it separates the data points into hyperplanes. Again, the SVM accuracy is just slightly higher than random forests, but still not high enough for it to be useful in real life.

Which countries' happiness doesn't seem explainable by the data? What unforeseeable factors could be involved?

Costa Rica, a Latin American country, sits comfortably in the low 10s on the World Happiness Report. However, it doesn't go unnoticed that their GDP per capita is lower than the countries with similar overall happiness ranking. Some unforeseeable factors that pertain to Costa Rica's happiness lies in their

prominence working on environmental protection, sustainability, and biodiversity. They strongly care about their environment, and unlike the United States, they don't spend any money on their army, instead, they put their efforts towards being carbon neutral and converting plain lands into forests.

Another unforeseeable factor is population density; too crowded or not enough people can affect the social aspect of people's lives and consequently, their happiness. Finland's stellar education system seems to determine the happiness of the youth population.

Are there trends in world happiness? Is the world getting happier or sadder?

Based on the data, it appears that the world is getting (slightly) happier!

Methodology

The first thing we did was compare how the country's happiness has changed from 2015 to 2020. The median change is at almost 0, indicating that there has been almost no shift in world happiness. Because the median splits the data in half, it also tells us that in the last 5 years, half of countries have become happier, and the other half have become sadder.

Country Happiness Changes from 2015–2020
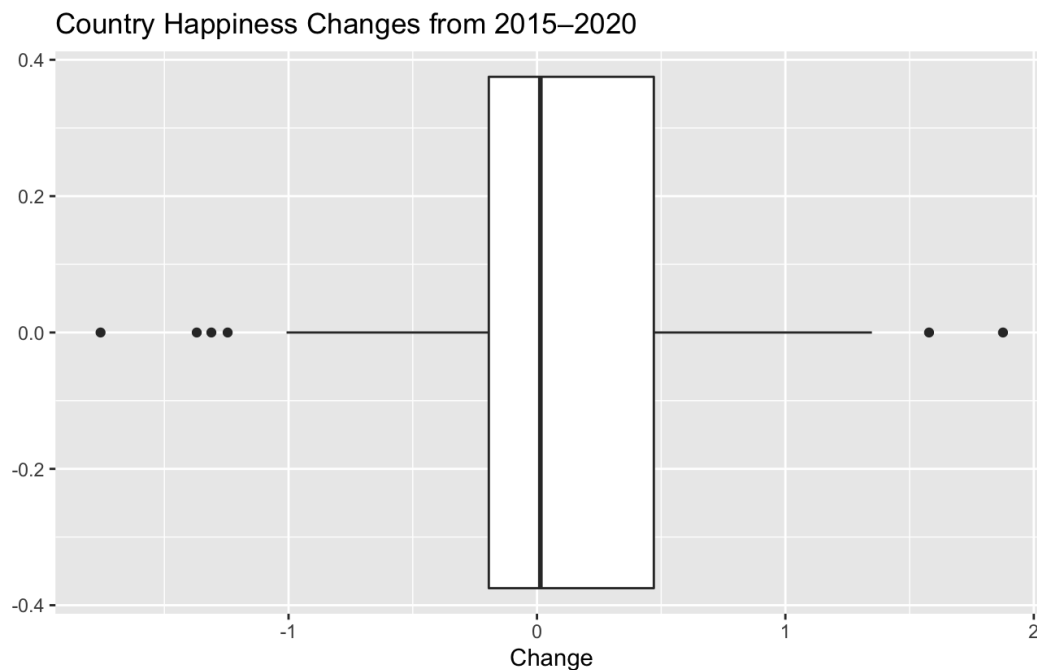
Next, we took the average of every country's happiness for each year, to create a world average. Graphing the data, we can see that there is an upward trend. However, the trend is small: mean world happiness has only increased by 0.1 from 2015 to 2020.
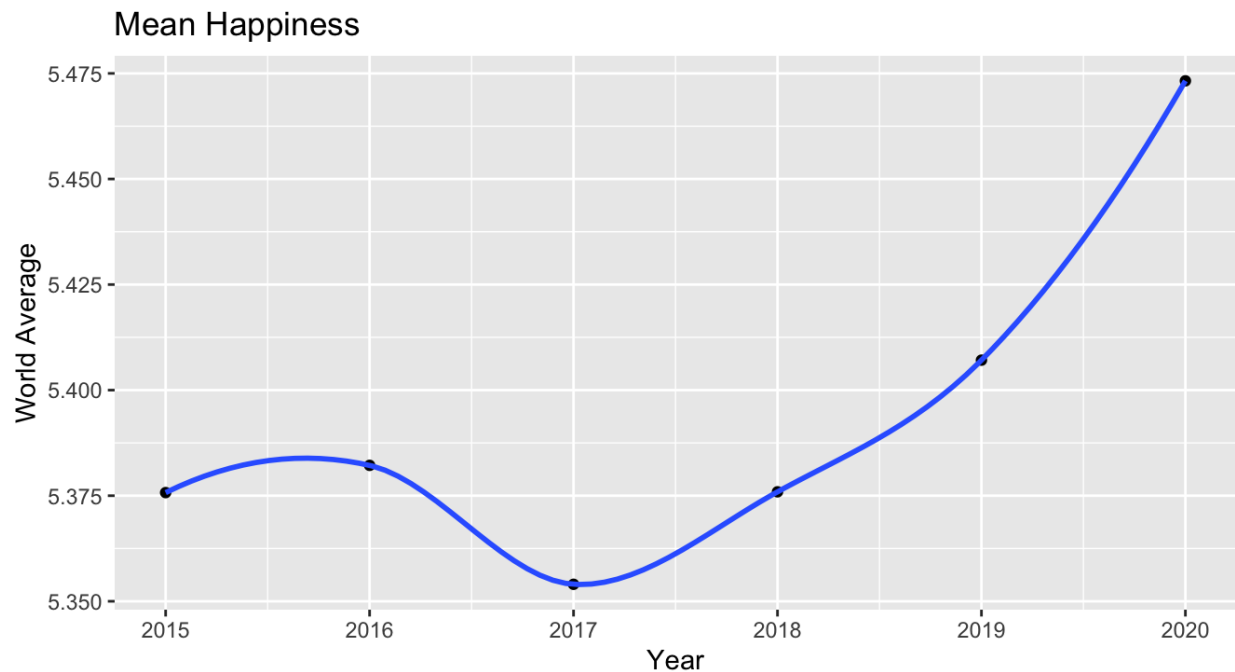
Mean Happiness

Happiness corresponds with worry, stress, and anxiety, which were greater in 2017, according to the Gallup Global emotions poll, which surveys people's daily emotions in more than 145 countries. This could be due to factors such as the #MeToo movement, Brexit, and former president Trump's inauguration. The continual rise in happiness levels into 2020 was shocking, especially looking back at the devastations and domino effect the COVID-19 pandemic had on the world. Nevertheless, for many individuals, having shelter-in-place granted many people time to reflect and practice gratitude.

We considered using boxplots for different years to show trends, but they were complicated and it was hard to understand what was going on.
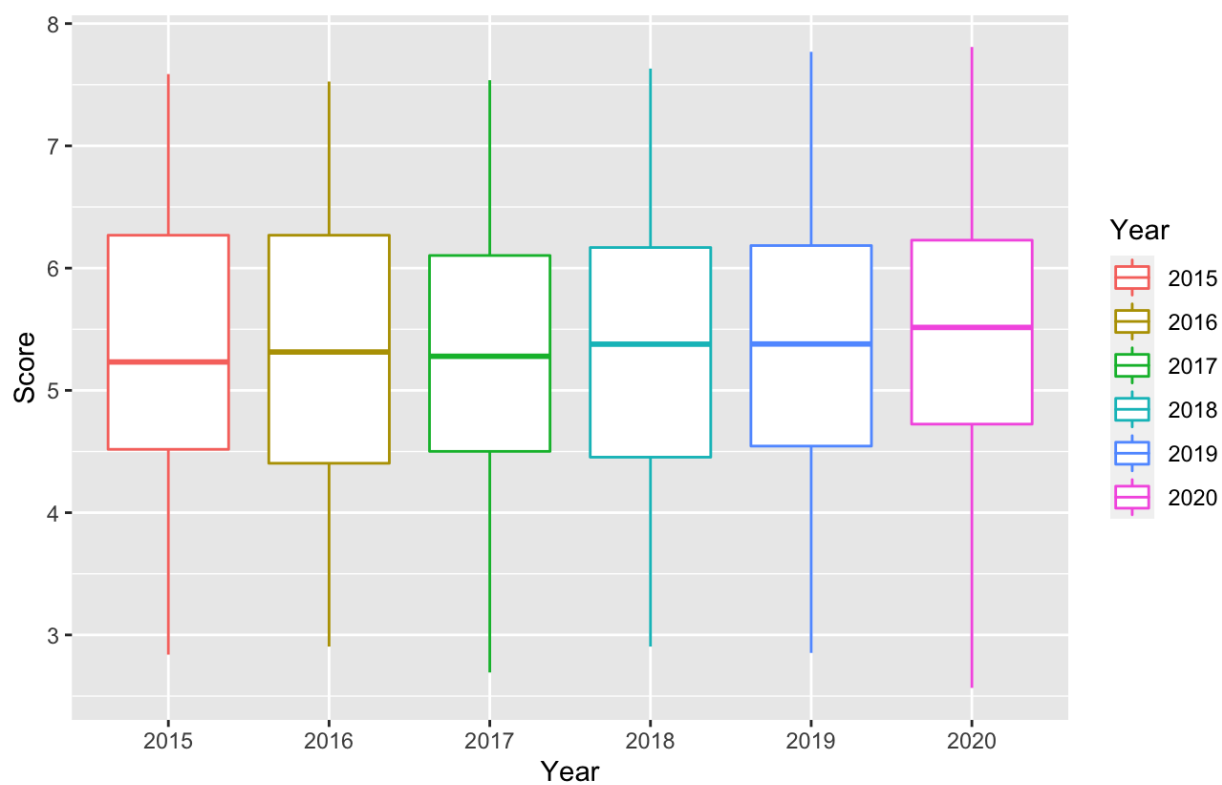
Which countries' happiness have changed the most over time? Does the data explain the shift?

Country Happiness Change from 2015-2020
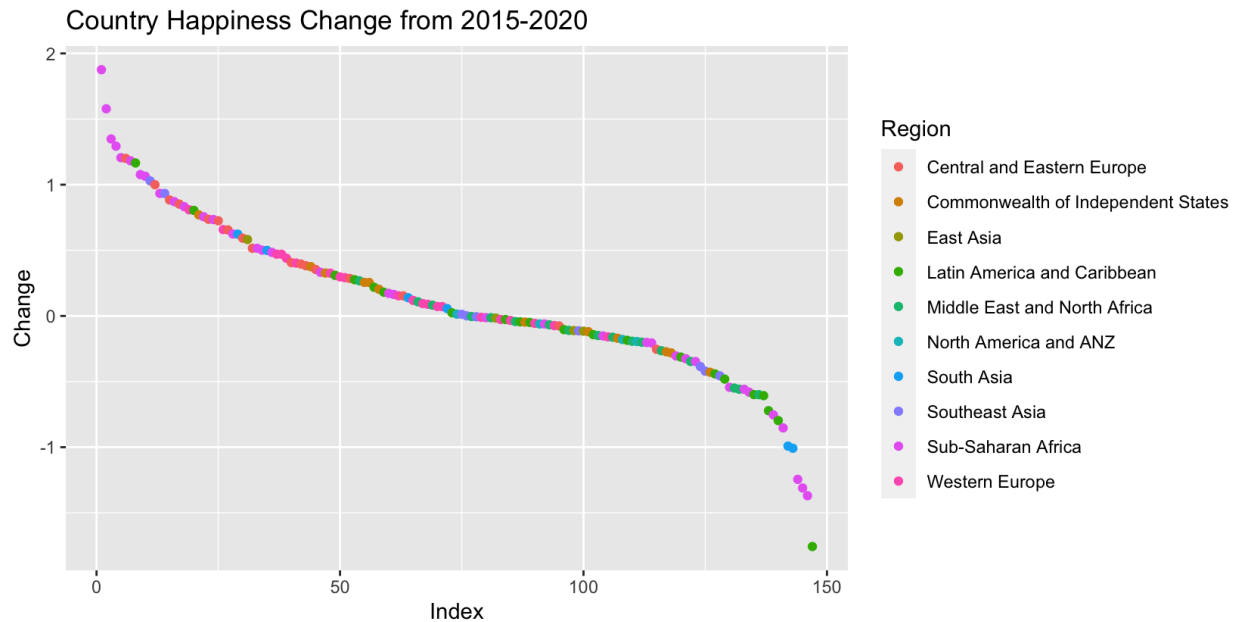
On the positive end of the spectrum, Benin and Ivory Coast have changed the most, with their happiness scores increasing by 1.88 and 1.58 respectively. On the other end, Venezuela and Zambia's happiness scores have gone down by 1.76 and 1.36.

Many of the countries with drastic happiness level changes over time were in Sub-Saharan Africa, as shown by the bar graph below.

Methodology

We decided to use a bar chart to display which countries had the largest difference in happiness levels. Since we discovered that region does matter when it comes to happiness, we displayed which regions contained the countries with the greatest shifts.

Top 10 Most Changed Countries
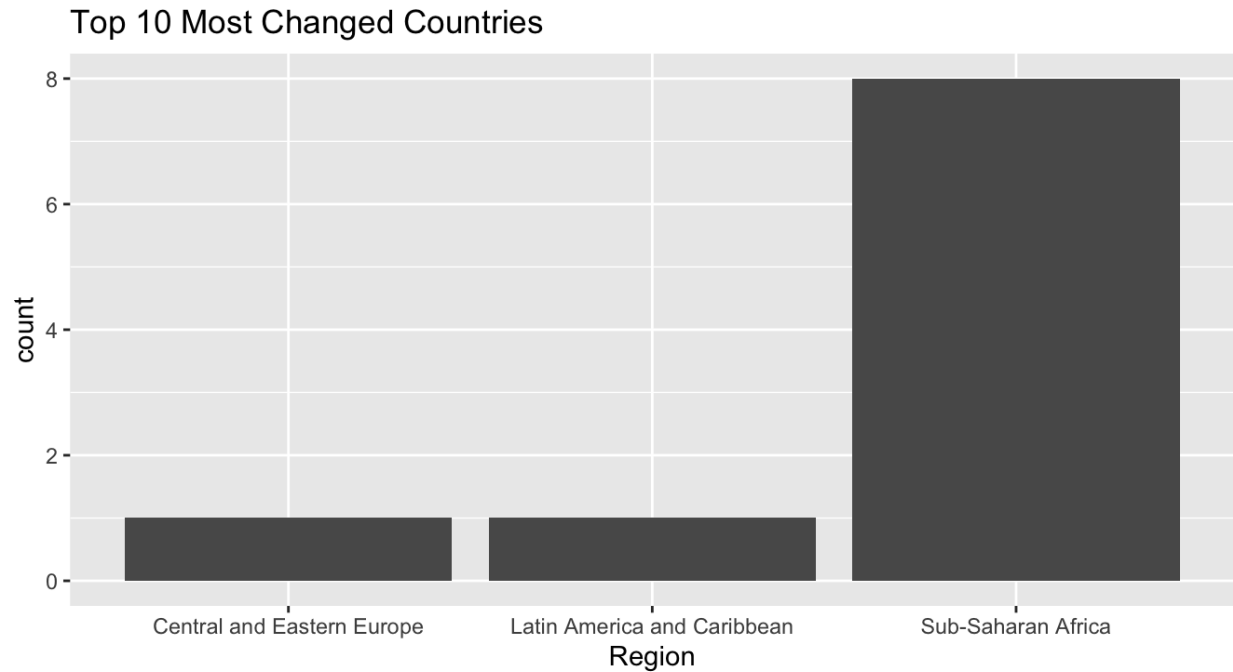
Graphing the changes of all countries by region, we can see that Sub-Saharan Africa has by far the largest range. The Sub-Saharan amassed eight of the top ten most changed countries, while Central/ Eastern Europe and Latin America/ Carribean each had one.

Noticing the pattern, we graphed all happiness changes by region to see if there was any correlation.

Happiness Change by Region, 2015-2020
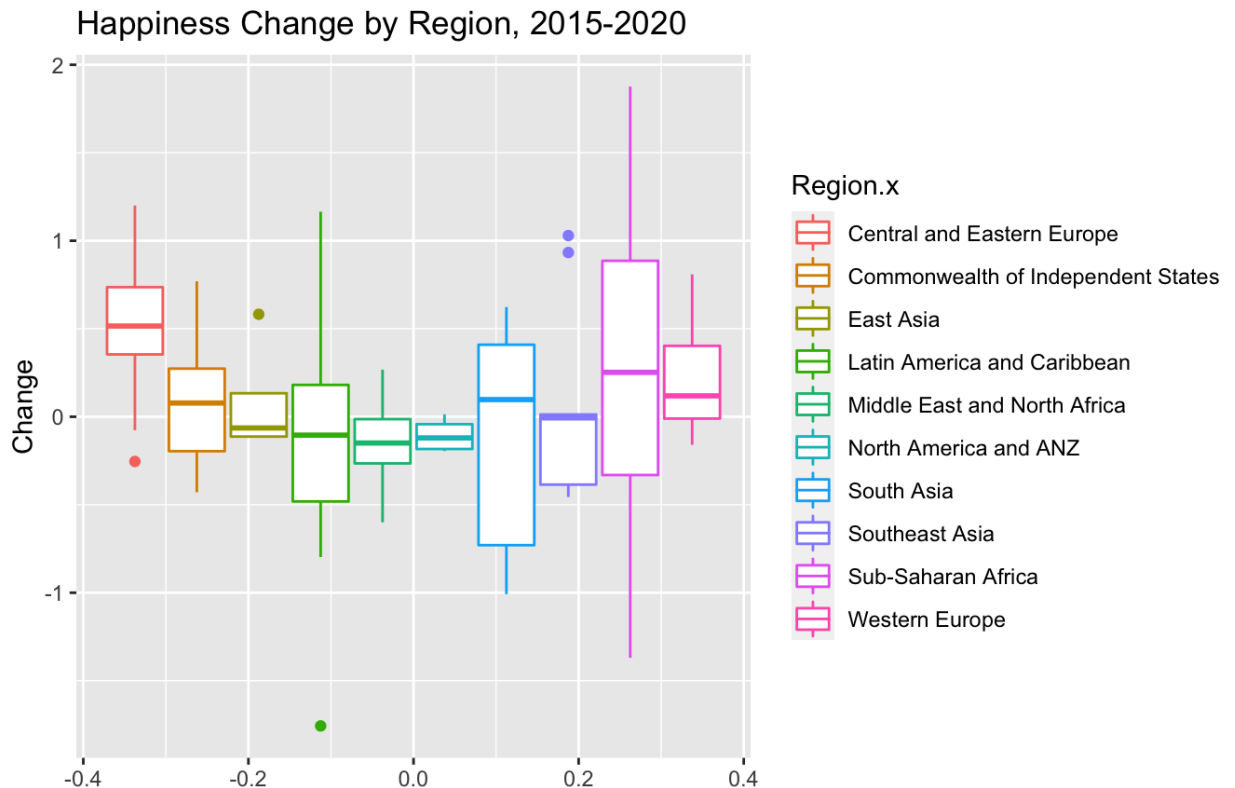
With just one quick scan of the graph, a few things are quickly noticed. Regions such as Sub-Saharan Africa and Latin America have great fluctuations in happiness, compared to regions in Europe and North America (the two happiest regions). Further research shows that regions with greater change tend to contain a larger number of developing nations, whose government systems may not be as stable and less efficient.

But seeing all this change is also a positive! You can see the median of the Sub-Saharan African group is slightly above a 0, which means there is a sign of possible positive development in the region. Central and Eastern Europe poses the greatest positive change, at around +0.5. The Middle East and North Africa have a slight depression in happiness levels, going below neutral. The Sub-Saharan Africa and South Asia regions have an IQR spanning one, whereas North America/ ANZ and East Asia have an IQR of a quarter, showing the difference between the regions.

If you look at the table from 2019, 5 out of the 10 top countries are in Northern Europe. Why is this? What did they do right to maintain the happiness of their people?

- The answer to this is more or less like Finland's case. The governments of these countries prioritize balance, which means not doing "too much" of any one thing. Their goal is not to make everyone billionaires, but to make sure that everyone is living well and no one is left behind. Even their prison system is like that. Prisoners in Norway enjoy free libraries and can take online STEM courses, possibly getting let out early as a reward for learning. According to the Bureau of Labor Statistics, a "full-time" workweek in Denmark is typically only 37 hours spread over the course of five days. The average American works seven hours more than a Dane in a week.

Results

Certain regions are happier — location is a crucial component in deciding a country's happiness.

It's hard to predict happiness, but it isn't impossible. There are numerous factors that contribute to the overall score.

To combine the factors together and look at how they directly affect a country's happiness, let's compare 3 countries: one happy, one neutral, and one sad.

First, Finland, which is part of Western Europe. They succeed because they have built a society with trust in their government and their peers. It is not about everyone being rich and living extravagant lifestyles, but about everyone being fulfilled where they are.

Based on private individuals, Finland has a fairly stable wealth, social network, and trust with their institution. People are not necessarily lit with bursts of joy every day, rather they are content with their life. Life satisfaction is what determines Finland's happiness, not the everyday level of optimism. Finland's quality education system is also a significant factor in the happiness of the country. Their system ensures that the learning environment is equal and positive for all students. Teachers are well regarded and well paid. Finland places an emphasis on the arts and encourages a variety of learning materials.

Next, let us look at a country not far away, Russia, or the Commonwealth of Independent States. Although their systems are very different and the two places have certainly different circumstances, we can learn from how their society is run. To start off, nearly half of pre-tax national income goes to Russia's top 10%, and wealth inequality is

even more drastic, with the richest 10 percent of Russians owning 87 percent of the country's wealth. This makes it one of the world's biggest wealth gaps of a major country. Russia is also ranked 124 on the democracy index, which ranks countries based on the state of democracy.

Lastly, we look at Haiti, which is part of Latin America. They rank 147 on the world happiness index, which is almost at the end. Unlike the other countries, Haiti does not have a stable supply of food and is still recovering from an earthquake that happened over a decade ago. They are the poorest country in the Western Hemisphere, with a 40% unemployment rate, half a million orphans, 48% literacy rate, 37% safe drinking water, 95% deforestation, and frequent hurricanes. If you examine the factors, Haiti's numbers are terribly low.

To summarize, it seems like the happiest countries are excelling in nearly all facets, while countries in the neutral zone have some positive aspects and downfalls. And in the case of the least happiest countries, it is common for them to be struggling in all factors.

Reproducibility

Our code is fully reproducible! By downloading all the datasets from 2015-2020, then following our data cleaning steps, you can run our code!

Our cleaned data is available here: https://drive.google.com/drive/folders/1TggJz0Bytt4gCNaAde3wygmv9w39yMI1?usp=sharing