World Happiness Data Analysis

Mario Castro, Isaac Nicolaus Braganza, Gaby Gonzalez, Claudio Rodriguez, Xiaoguang Liang, and Riya Luthra

12/13/2019

Introduction

It is often said that happiness is a priceless commodity; that money, resources, and material goods does not equate to happiness, but is that case in the real world? Every year, an independent group of researchers publish the "World Happiness Report" which reports for each country a Life Ladder score; a 1-10 scale of how happy a country is. Along with the Life Ladder score, the World Happiness Report also publishes social statistics on each country, varying from social support to generosity. The typical method used to record many of these variables include answers to yes or no questions which are then convereted to either 0 or 1. What our group seeks to discover is the various social and economic conditions that lead to better happiness scores and how does this relate to countries' economic development and wealth. The first question we would like address the affect of social support systems within a country and how they affect their overall happiness (Life.Ladder). To follow up this question we will also be observing the unique case of the Scandiavian peninsula, whose nations comprise some of the happiest among the world in terms of Life Ladder scores. Based on a recent article published by the Times magazine titled "I Led One of the Happiest Countries in the World. Here's What Other Democracies Could Learn From Our Model," we wonder how their democratic quality scores stacks up in comparison to the rest of the world, and we will question if the average democratic quality score is vastly different from the rest of the world. If so, could their high happiness scores be a result of how they operate their democracy?

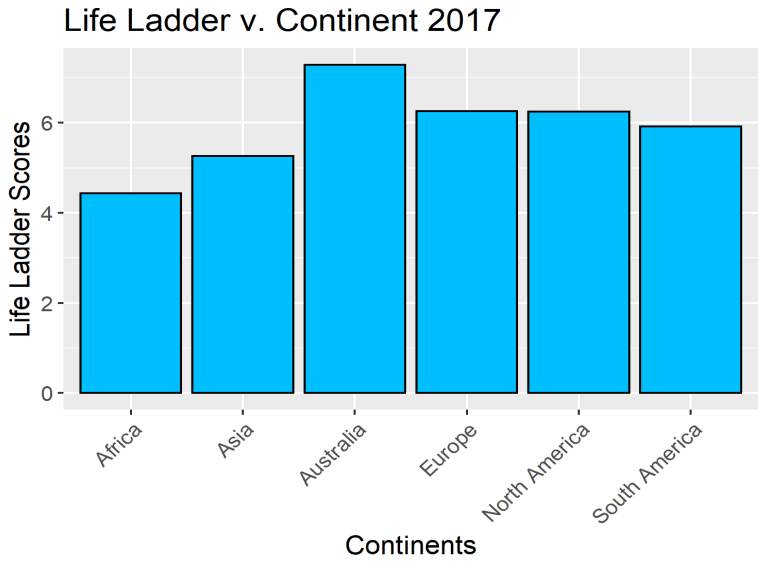
A question that arose during the beginning of the project was if it was possible to categorize the different countries by continents to see if there was some underlying trends and correlations between them. Our group went ahead and categorized each country into respective continents.We were able to find a website that showed us how to categorize each country into continent and we used this method to do so.

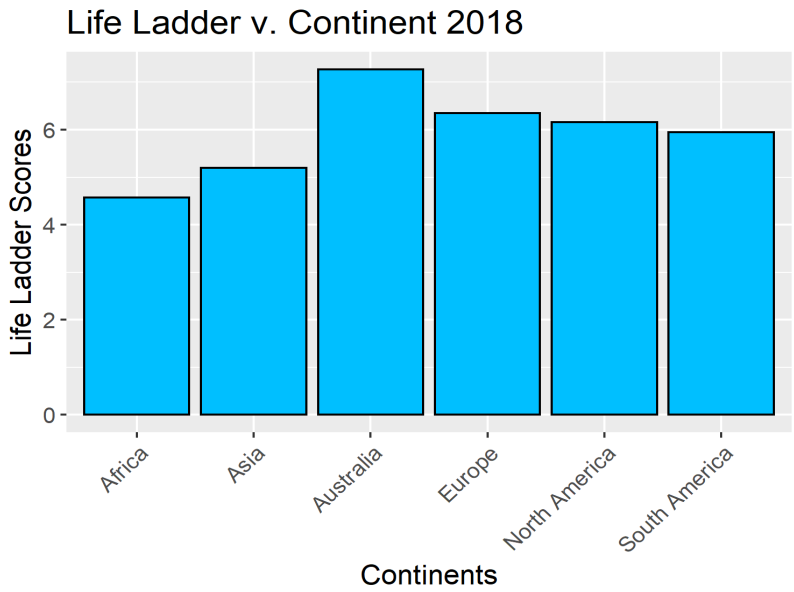
EDA: General/Individual Plot

Note: All articles and outside research will be referenced in the reference section at the end of report

1: Change in Life Ladder Scores from 2017 to 2018

Question: How have the life ladder scores changed in different continents from 2017 to 2018? What continents had the largest change in life ladder scores and what could be the possible reasons?
Graph(s): Bar Graphs
Group Member: Gaby Gonzalez





Response:

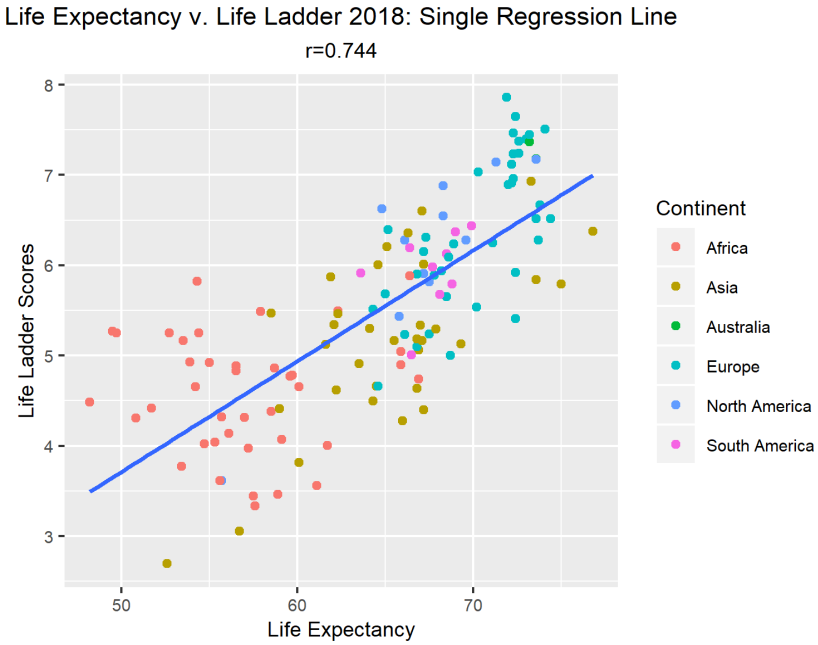
Life ladder is a measure of world happiness that asks people to judge their life on a scale of 0 to 10, from the worst to the best it could be. This is essentially a person's life satisfaction, their overall happiness. We are trying to understand happiness better, what factors affect it, and how it is located geographically. In order to visualize this, we first parted the data into continents in order to have a cleaner data set. I created bar graphs for the 2017 and 2018 averages of the life ladder scores, separated by continent. This enabled me to clearly see the differences between continents' averages, and by changing the graph from year to year I was able to see the actual change that happened. My hypothesis is that continents with a higher percentage of first world countries will do better in average life ladder scores overall, as well as regress less to the mean when we see the change from year to year, since being a first world country is dependent on various factors like the country's GDP, life expectancy, literacy rates, etc., which I believe should be indicative of a person's life satisfaction.

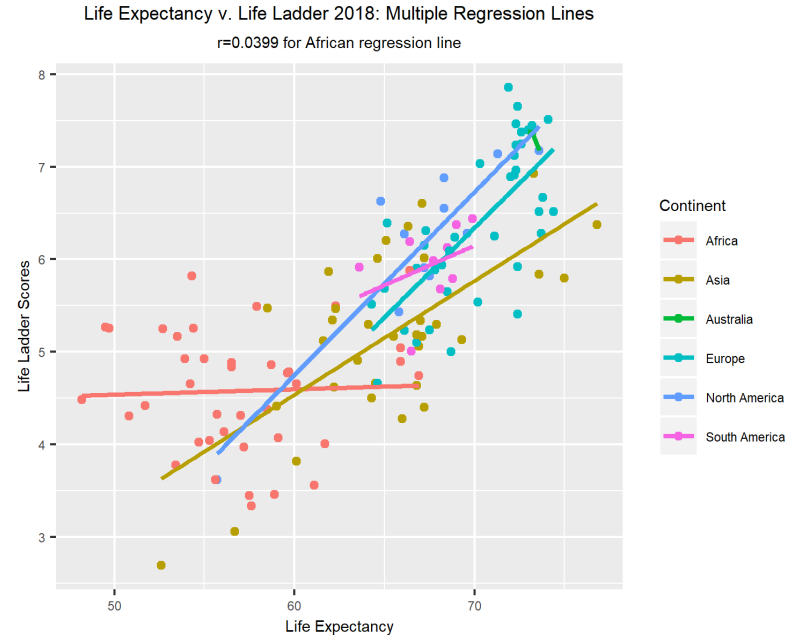
From 2017 to 2018 the data did not change drastically for each continent, since it is only a one-year difference. The 2017 average for all continents was 5.9.Australia still maintained the highest life ladder score, only decreasing by 0.018 from 7.29 average points. This could be because Australia is only measured by two first world countries, and as a result the continent has the lowest standard deviation in its countries' 2017 scores, at 0.049. This continent barely regressed to the mean due to its 100% First World countries. Asia was strange one, because it is below average in its life ladder scores at 5.26, but went down by 0.06 points, which I believe is because of the high standard deviation on from its countries' 2017 scores at and SD of 0.984, the highest of all. I would consider this an outlier because of its extreme SD. Africa had the lowest score in comparison to all the other continents (4.44), and it was the continent that improved the most, by increasing 0.13 average points. As a continent that was noticeably below average, it fits the phenomenon of regression to the mean, and it also has no First World countries in it. While North America and Europe were almost tied at second place in the life ladder, at 6.254 and 6.262 respectively, North America followed the regression to the mean by decreasing its average life ladder score by 0.099. Europe, on the other hand, increased its score almost by this same amount (0.09). This could be explained by the fact that North America only has two (around 15%) First World countries, Canada and the United States, while in Europe more than 50% of the countries are First World Countries. South America (0 First World Countries) stayed around the mean. In conclusion, continents with a higher percentage of First World countries were on average more content with life, and they did better in the year to year change, although most above and below average continents regressed to the mean in their life ladders.

2: Healthy Life Expectancy v. Life Ladder

Question: How does a Healthy Life expectancy correlate with a persons ranking of their happiness? Is country with a higher life expectancy on average happier than those with low life expectancy? If so, what has allowed those countries to have such high life expectancies and could those same methods be applied to other countries that need it?

Graph(s): Scatter plot, Regression Line
Group Member: Mario Castro





Response:

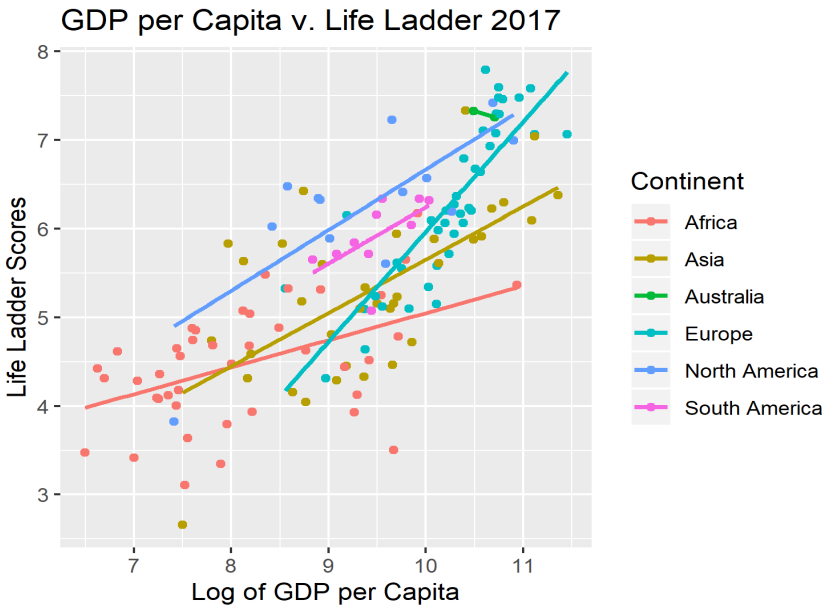
Looking through the data set given for world happiness I became interested in seeing if there was a correlation between a country's life expectancy and how happy their people viewed themselves. There was no surprise to see that there was a pretty strong correlation between to the two factors. The r coefficient supports this statement as it calculated to be about 0.74. After graphing the scatter plot, I added a single regression line for all the countries in the first plot. This is the blue line seen in the first graph. The positive slope

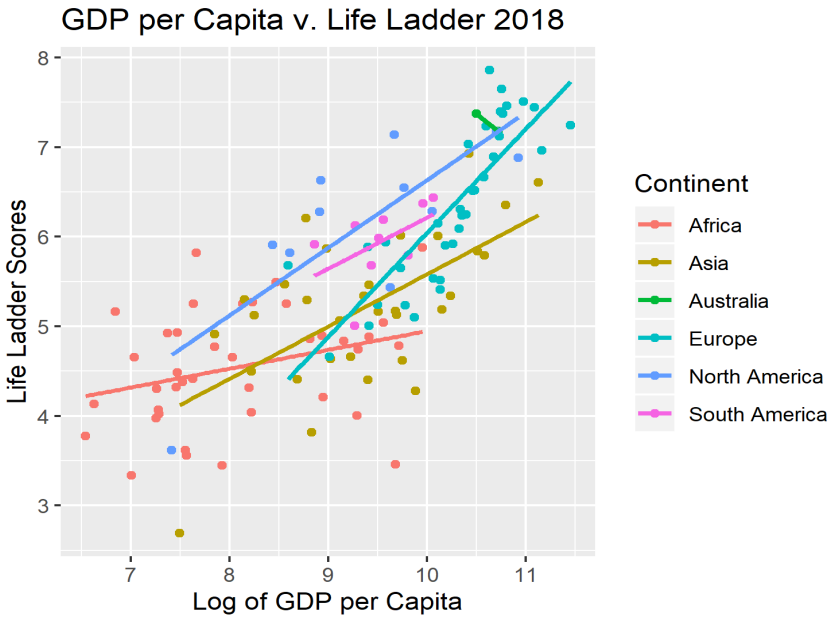
shows how as country's life expectancy increases, their life ladder score (happiness) generally increases too. One thing to note from this graph is that it is mainly the countries from Europe that have a high life expectancy and high life ladder scores when compared to other countries from Africa and Asia. This lead me to wonder why. Looking at various articles online I found one written by Michael Nedelman for CNN in 2018. Within it he stated that many European countries have seen that "maternal and infant mortality are down in recent years." This gives some insight as to why European countries rank so high in life expectancy when compared to those in Africa. Many African countries lack the technology and economic resources needed to reduce the deaths involved in the process child delivery.

After plotting the single regression line I wanted to see if I could find positive correlations between the individual countries in their own continent. I replotted the same scatter plot but this time I added a different regression line for each continent. This is the second scatter plot that is shown. While I was looking at all the individual regression lines I saw that almost all of them had a positive correlation. All except for the regression line for Africa. I was surprised to see that the regression line was almost horizontal. I then proceeded to calculate the r coeffecient and got that it was nearly 0 (r=0.04). This means that there really wasn't any correlation between the two variables when it came down to the African countries. Why was this so? I began to look online for some possible explanations/inferences to this. In an article posted by Angelle B. Kwemo in the online blog, Brookings, she states how "as a region, Africa accounts for around 20 percent of U.S. aid." This makes sense as most of the African countries do not have the resources to support themselves. However, in a seperate article posted by the UN, it is stated that most of Africa's wealth was "increasingly concentrated in a few hands. Disappointingly, 10 of the world's 19 most unequal countries are in sub-Saharan Africa." The inference that can be made from this is that the unequal distribution of economic and technological resources within Africa leads to different life expectancies for each individual country. This is because not all of them will have access to the US aid which allows for modern medical resources to be implemented in processes such as child delivery. This may be why there is no correlation between life expectancy and the life ladder scores in African countries. Many of those African countries become accustomed to a short life span so their happiness does not rely on how long they live.

3: Log of GDP v. Life Ladder

Question: Does higher GDP per capita lead to high life ladder? Which continent has the highest correlation between the two variables? Which continent has the lowest? How have the correlation changed from 2017 to 2018?
Graph(s): Scatter plot with regression lines
Group Member: Xiaoguang Liang





Response:

It seems intuitive that the wealthier people are, they are more likely to be happy. In this plot, we use the whr_alternate_2018 data set, which the log function was used in the given data, to make the scatter points more homoscedastic. Then, we plot log gdp per capita against life ladder acrossed all nations which are categorized into continents to see the relationship between the two variables using a regression line and the correlation coefficient. Then, I will use the data set from 2017 to calculate the correlation coefficient to compare with the ones from 2018.

As we can see from the scatter plot, most of continents, except for Australia, which only has two countries in the data set, if we look at the regression line, they all present a positive

correlation between GDP per capita and life ladder scores. What we find here is that there is a trend that as the GDP per capita increases, life ladder tends to increase. Now let's look at the overall correlation coefficient, which results in about 0.76, a pretty strong indication about their relationship.

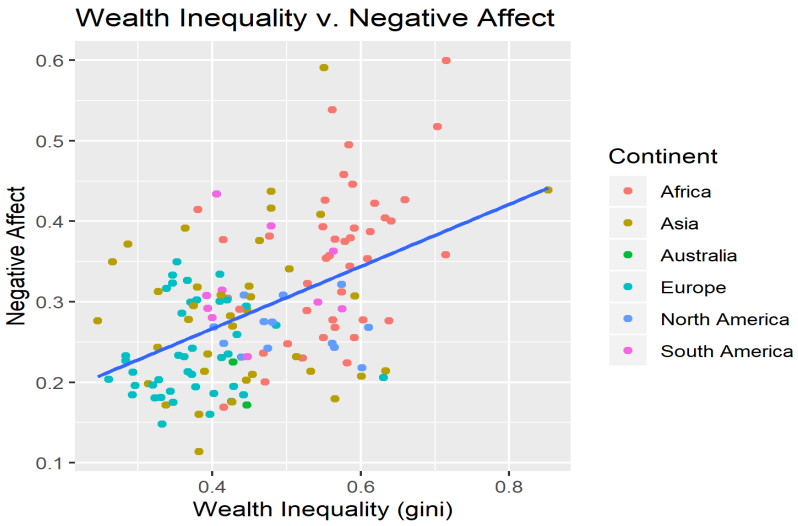
Let's take a look at the correlation coefficent of each individual continent and see which one has the highest and lowest correlation between GDP per capita and life ladder scores. Africa: 0.30, Europe: 0.82, Asia: 0.61, South America: 0.50, North America: 0.77, Australia: -1. Notice that Australia has the strongest negative correlation between the two variables. Although it is tempting to make the conclusion that as GDP per capita increases in Asutralia, the life ladder score decreases, when we look at the the data set whr_alternate_2018, we can see that there is only two countries categorized under Australia, Australia and New Zealand, with very similar data. Based on the number of countries, it is not fair to make such a conclusion. We need to have more country's data to test the hypothesis. Besides the result of Australia, the continent which has the highest correlation is Europe, and the continent which has the lowest correlation is Africa. What causes the difference in the correlation coefficient? Could it be the economic development, or could it be the culture that one believes wealth bring happiness more than the other? Unfortunately, we do not have an exact answer from this scatter plot or the correlation coefficient. The scatter plot and coefficient only tell us that there is a trend that increase in GDP per capita corresponds to higher life ladder scores, and such a relationship is stonger in Europe than in Africa. They, however, do not tell us what causes the difference.

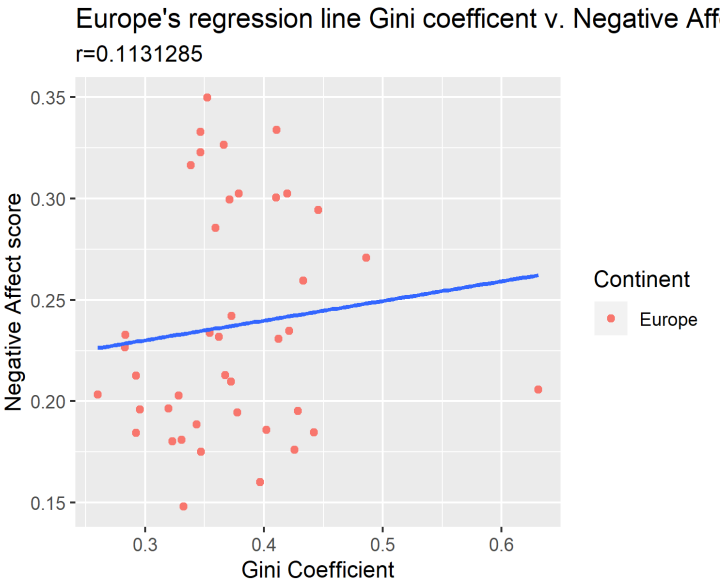
The correlation coefficients from 2017 are overall: 0.77, Africa, 0.48, Europe 0.83, Asia 0.64, South America 0.60, North America 0.75, Australia -1. Therefore, compared to the results from 2017, in 2018 there has been a small decrease in the overall correlation between GPD per capita and life ladder scores; Africa, Europe, Asia, and South America, have also experienced a decrease on the tendcy; North America is the only one that has experienced an increase in the relationship; while Australia, which again only has two countries under the category, has remained unchanged.

The correlation coefficients are a powerful tool. They allow us to have some predictive power. If a continent increases its GPD and GDP per capita, through the correlation coefficient, and the average and standard deviation of GDP per capita and life ladder scores, we can predict how much life ladder scores of the continent will tend to increase.

4: Inequality (Gini) v. Negative Affect

Question: Is there a relationship between wealth inequality and how people percieve their day?
Graph(s): Scatter plot
Group Member: Isaac Nicolaus Braganza





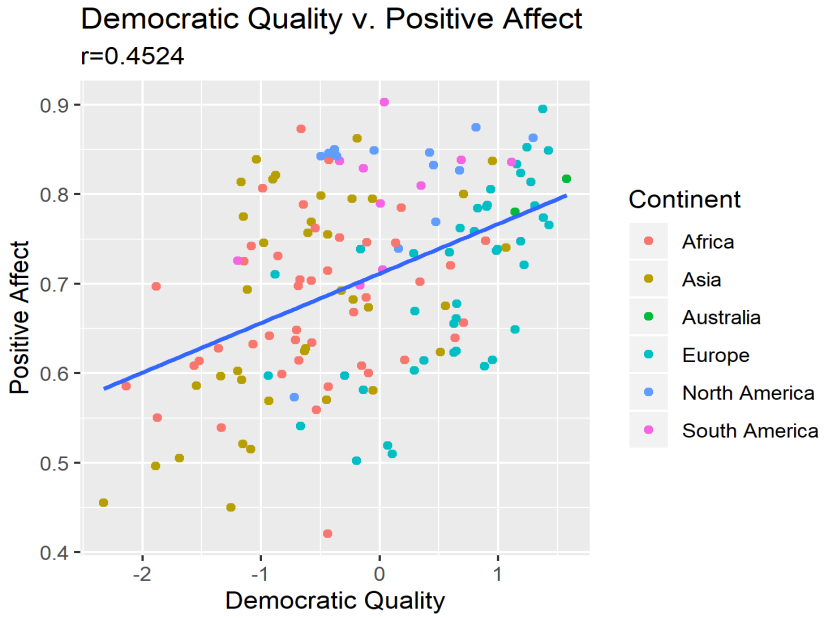
Response:

There is a moderate correlaton between the level of wealth inequality and negative affect as the r coefficient was around 0.48; negative affect is an average score that generalizes if people of the nation experienced common negative emotions yesterday; i.e anger, sorrow, and worry. What I could assume is that based on this correlation and graph, the more wealth inequality there is in a nation, the more negative emotions people experience day by day. Thus, we can assume that since the correlation is stronger in countries with higher wealth inequality we can extrapolate two things, one is that the affects of wealth inequality take a toll on peoples minds, and likely this could be because people are more conginzant of the level of inequality they experience, since it may seem more obvious. Secondly, since the correlation appears weaker when the the gini coefficient is lower, this could be because people overall tend to not notice inequality as much if it isn't as stark or prevalent, thus such inequalities may not have a strong of an impact on peoples lives if its not noticable by

their standards. Some limitations of this result would include the possibility of outliers that may be weighing down the correlation. This relationship could be explored further by investigating continents with lower levels of wealth inequality in general, Europe on the original scatterplot tends to be clustered on the lower end of gini coefficients, by analyzing Europe we find that the correlation is very weak with a r coefficient of 0.11, which is very suprising since generally I would expect highly educated societies to be more aware of wealth inequality, and if it is stronger, it would affect them more. My hypothesis is that since Europe is already a generally developed and wealthy continent, they may be more indifferent to wealth inequality to some degree as the presence of many social safety nets in these countries gives most citizens a basic level of health care, income, and education, so even wealth distribution is unequal, the idea that everyone has their basic need met and a little extra would make an indifferent society. Good healthcare, and cheap education, all in all leads to less worry.

5: Democratic Quality v. Positive Affect

Question: Does having a Democratic society or government in power directly affect your overall happiness? Does it improve your quality of life and is there correlation? Why do coutries with democratic societies generally have happier people?
Graph(s): Scatter plots
Group Member: Claudio Rodriguez



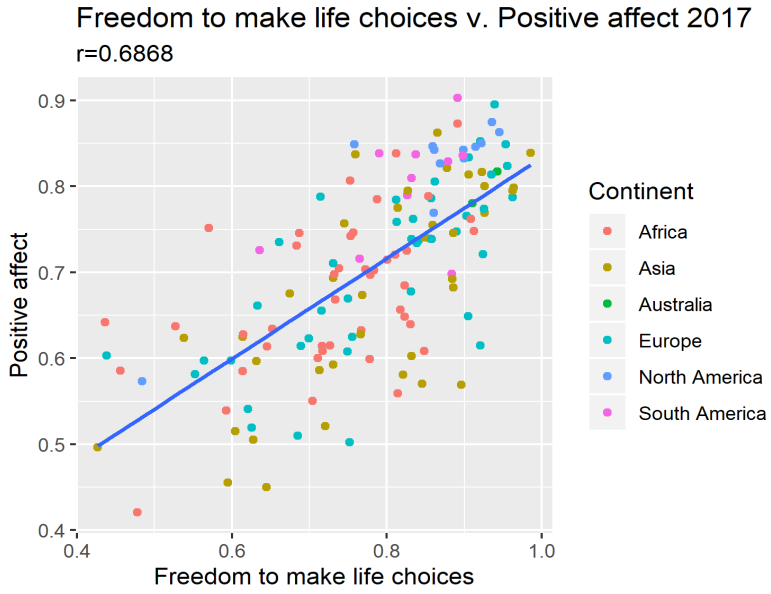
Response:

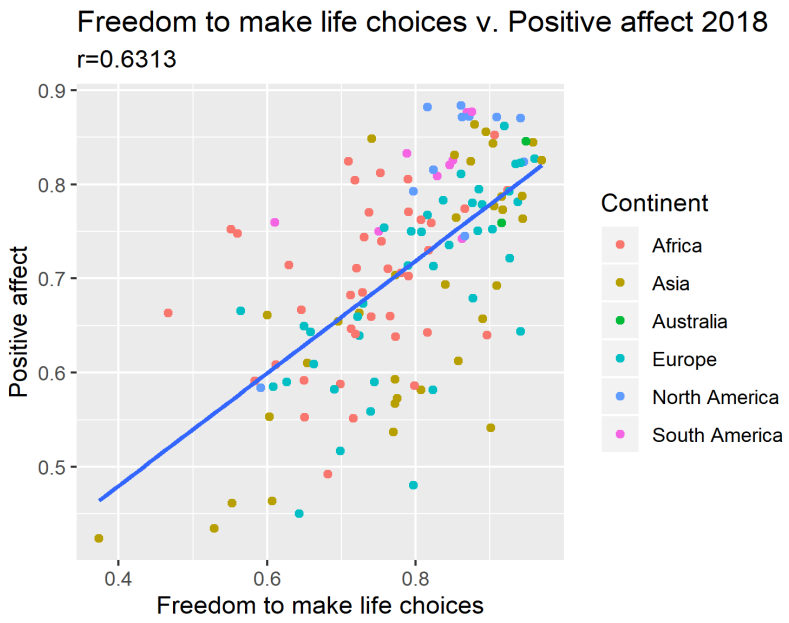
To start of I decided to make a scatter plot in order to see if there was correlation and if the plot came out to be Homoscedastic. On the X axis I plotted Democratic Quality and on the Y axis I then plotted Positive effect. This represents how government can impact your overall happiness. If your countries point was in the upper right-hand corner this means that you are happy with life and the democratic quality had a positive impact and vice versa if your

point were to be on the lower left-hand corner this means that your generally not happy with life and your government. Having a Democratic party in office or a country with democratic views does directly affect your overall happiness. I was able to do a scatter plot to see if there was a direct correlation between Positive affect and Democratic quality. When I did this, it seemed homoscedastic meaning it was football-shaped. There was no straight direct positive trending line, but it seemed to have a positive trending correlation. The correlation coefficient was .4524347. This means that there's a moderate correlation between positive affect and democratic quality. This could potentially mean that there is a big difference in what type of government is in power and has an impact on your overall happiness. This could potentially mean that depending on what sort of government is in power it can improve your quality of life, mental health and how happy you are daily, but it would not be the primary factor in doing so. Since there is a moderate correlation this means that you are more likely to laugh, life, and overall be a happier person but since it is not a direct strong positive correlation this means that it is not the only factor in determining if you are happy or not. Democratic countries generally have happier people because they have the right to vote, effective participation, enlightened understanding, citizen control of the agenda and be inclusive to all citizens. This can all be a correlation to have a happier individual because it gives you a sense of freedom and makes you feel as if you have more control and power in your life.

6: Freedom to Make Life Choices v. Positive Affect

Question: How does freedom of choice impact their positive affect?
Graph(s): Scatter plots
Group Member: Riya Luthra





Response:

The two plots above show the correlation between a person's freedom of choice, which represents the satisfaction individuals feel about their freedom to choose their life decisions, and positive affect, which represents how much the person felt that they laughed or enjoyed their time the day before. Within the plot the data of all continents is plotted with different colors. The first graph represents the data from 2017 and the second plot has the data from 2018. I wanted to test both years separately to see if there were any major changes. However, the correlation coefficient for 2017 is pretty close to the correlation coefficient for 2018 related to the relationship between freedom of choice and positive affect. Both graphs have a positive slope and strong correlation thus there is a positive association between freedom of choice and positive affect. The correlation coefficient for 2017 was 0.68 and in 2018 it was 0.63. Although there was a slight decrease in the correlation coefficient in 2018, both plots show a strong positive linear relationship in the two variables.

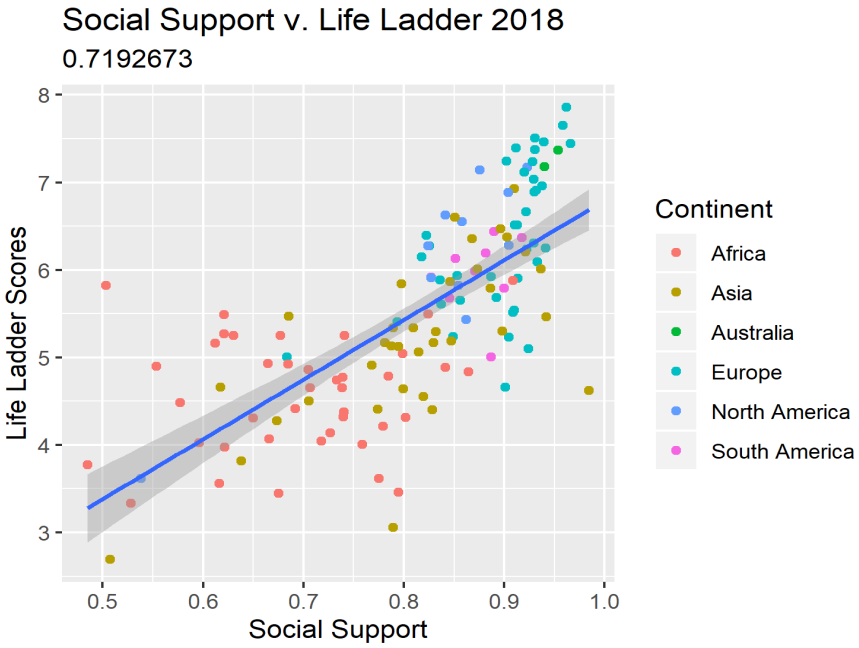
This conclusion is understandable because happiness is an innate trait and having control of your life can help with happiness because people tend to do what makes them happy. Having someone else controlling your life is not always great because no one knows a person's likes and dislikes other than the person themselves. Will Wilkson in his article on Forbes also completed a study on freedom and how that affected the satisfaction of a person with their life. He explained how freedom of choice is an internal idea that allows you to control the external outcomes. Through freedom of choice people are able to express themselves. People show happiness by laughing and feeling that they are having a good time in an activity when they are passionate about the situation.

Data Analysis

Detailed Topic 1: Social Support and its impacts on overall Happiness

Background:
We chose to look at social support for one of our detailed questions because it is a factor that we all come into contact with on a day to day basis. Friends and family help us deal with stress, make better lifestyle choices that keep us strong, and allow us to rebound from health issues and disease more quickly. In college, many students experience their first moments of freedom by living away from their parents so friend become a big part of one's life. We wanted to see how strong of correlation there was between social support and the ladder score which is a score from 0-10 on how a country would rank itself on happiness.

We will begin by creating a scatter plot of social support and the ladder scores for the year of 2018 since it is most recent. To see some further trends in the data we will also color code the countries by continent.



From this graph and the correlation coefficient (r= 0.7651) we can see that there is a strong positive correlation between these two variables. However, could this correlation just be due to chance? To test that we will be performing a hypothesis test on the regression line.
Hypothesis Test for Regression Line
Null:
a) The results from the produced scatter plot are due to chance
b) There is not correlation between the social support and happiness (ladder). r=0
Alternative:
a) The results are not due to chance
b) There is a correlation between the social support and happiness (ladder). r!=0
Calculation of z-statistic and p-value:

We used R and its built in functions to go ahead and calculate the variables needed to find the z-statistic and p-value. We referenced the pdf on bcourses to calculate the SE. The calculated values are listed below:

SD for social support= 0.1163321
SD for life ladder= 1.103461
Slope for the regression line= 6.822566 RMS for the regression line= 0.7666114
SE= 0.5692764

Using these values we were able to find that:

z= 11.98463
p-value (aprox.)= 0

In addition, we also used the lm() function in R to see if our calculated z-value and p-value checked out. We found that calculations were very close to R's own calculation.

Conclusion:
a) Given the z-score of around 12 and a p-value of essentially 0, we can reject the null at 1% significance level
b) Results lie outside the realm of chance
c) There is a clear relationship/correlation between social support and happiness

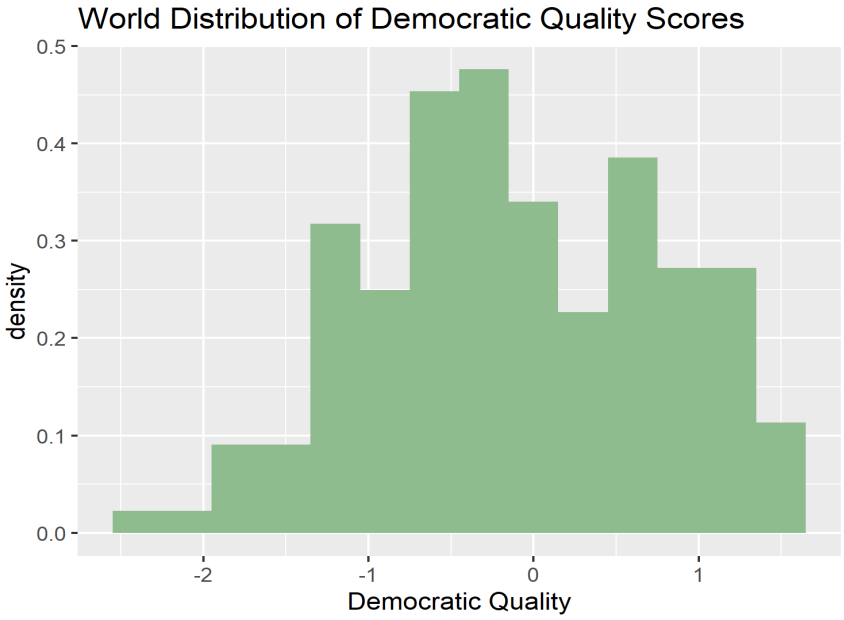
Written Analysis:
There is a strong positive correlation between having social support and one's life ladder score. Individuals who felt they had a lot of social support had greater life satisfaction. In fact, Yang Claire Yang, a sociologist who studies physiological effects of social ties, believes that there is an association between social support and how bodies process stress (Pappas). From our experience, we do feel released of stress when we talk about our problems to the people who support us. Also, we can see that there is a negative correlation between social support and negative effect, which is defined as the average as previous-day measures for worry, sadness, or anger. However, from the graph, we see the trend that the higher the level of social support corresponds to the lower the level of negative affect. Accordingly, Pappas reported that friends help with coping with rejection and tough issues through the examples of children being rejected the classroom and women going through breast cancer. Having social support may help us have a better attitude to face our problems in a more positive way. This is an understandable conclusion because we know having good strong relationships with people result in people feeling less stressed, being more content with themselves and life.

Detailed Topic 2: Why might the Scandinavian Peninsula be the happiest region in the world?

Background:
Does the Scandinavian Peninsula have better quality democracies or is this just due to chance, if not due to chance, is there a relationship between happiness and Democratic Quality? The Scandinavian Peninsula is unique because its nations are within the 90th percentile in terms of Life Ladder scores. When analyzing this sample, we noticed that one key characteristic of these countries is their high Democratic Quality ratings. Maybe having

a high-quality democracy could have an impact on overall happiness. Note that our inspiration for this topic came from the Time's article "I Led One of the Happiest Countries in the World. Here's What Other Democracies Could Learn From Our Model".

Process: First we must establish which countries fall in the Scandiavian peninsula in a data subset. This is created using R. After that is established, we would like to investigate if the average Democratic Quality score in the Scandiavian region is statistically significant compared to the rest of the world. We will be using a t-test for determining significance as the our sample size is less than 25 and the variable of democratic quality roughly follows the normal curve across the world.



Hypothesis Test Null: a) Differences in averages are due to chance. b) Democratic Quality of the Scandiavian Peninsula is not diffeent from the rest of the world, them being in the 90th percentile of happiness scores is likely not due to Democratic Quality scores.

Alt: a) Differences are real! b) Democratic Quality scores are significantly higher than the rest of the world, and this proof of difference may be a call to further investigate the realtionship between overall happiness of a country and democratic quality.

T-test Calculation ave_ScandiaviaDem.Q=1.315
ave_WorldDem.Q=-0.122
SEQ2=0.04
T-test
(ave_ScandiavianDem.Q-ave_WorldDem.Q)/SEQ2=35.87407

Conclusion:
a) Based on the T-table with a T score of 36 and degrees of freedom of 4 we reject our null hypothesis at the 1% level and our p-value is essentially less than 1% b) Results likely caused by the fact that these countries democracies are operated much differently than the rest of the world, in a way that would lead to such a positive response towards

government. c) We should investigate further the relationship between overall happiness and a good democracy. The first bit of info we needed to find was the averages of the the Scandiavaian Peninsula and the world in the catergory of democratic quality. A T-test would be used since we know that the distribution of democratic quality scores around the world roughly follows the normal distribution, our sample of countries is less than 25 observations, and due to the fact I had to filter out the "NA" out of the data frame in the Democratic Quality scores we can assume we can't find the SD of the box, thus the T-test is a valid and reasonable hypothesis test to use. So we calculated the standard error of the sample, while also using SD+ since this is a T-test. From there, took the difference of the averages of Dem.Quality scores and divided by standard error to get a T score of around 35, and with a T-table and 4 degrees of freedom from our obseravtion count of 5, we concluded that the p-value is close to 0 and thus we reject the Null.

Conclusion:

Happiness can mean different things to different people, but it is a chemical reaction in our brain that can be affected by many different factors. We used positive effect, negative effect, and life ladder scores to assess a person's happiness, so we could get a broader idea of happiness. In our study we noticed that GDP per capita had the strongest positive linear relationship with a 0.83 correlation coefficient, from this we can draw the conclusion that a person's income can play a significant role in their happiness. However, behind the GDP correlation, the next most important factor is social support, or how good we feel about the relationships we have formed with the people around us. This has nothing to do with money, but we realized that so many other factors play a large role in people's satisfaction with their lives. Things such as life expectancy, freedom to make life choices, and democratic quality make us comfortable with the life we lead starting from our physical health (life expectancy), to what we do with our time (freedom to make choices), and lastly, how our government's decisions affect our daily lives (quality in our democracy). All these things relate to how we live our life, and the capability to do what we want to in order to make the most of it. We also found things that impede happiness, like wealth inequality, which, although related to money, ultimately affects the differences in quality of life between people, making relationships harder, and hinders our ability to lead the life we want.

References

1. This was for categorizing the countries into continents
   https://www.kaggle.com/javadzabihihappiness-2017-visualization-prediction
2. Articles used by Mario Castro
   https://www.cnn.com/2018/09/11/health/world-health-organization-europe-report-intl/index.html https://www.brookings.edu/blog/africa-in-focus/2017/04/20/making-africa-great-again-reducing-aid-dependency/ https://www.un.org/africarenewal/magazine/december-2017-march-2018/closing-africa%E2%80%99s-wealth-gap
3. Articles used by Riya Luthra
   https://www.forbes.com/sites/willwilkinson/2011/03/23/happiness-and-freedom/#1db3247bfe5f
4. Articles used by Gabriela Gonzalez
   http://worldpopulationreview.com/countries/first-world-countries/
5. Articles used for the hypothesis test #1
   https://www.livescience.com/53315-how-friendships-are-good-for-your-health.html?fbclid=IwAR1srNG6ncHrHPBrN4HeqL-O-TyQUx9MONpd4v3bTW1b0SxKKxIR_tL-mME
6. Articles used for hypothesis test #2
   https://time.com/5740400/denmark-happiness/?fbclid=IwAR1DHFFiT6l0_thzw6F_e5700ae6sBVr73jMi6UjDXeYBuYxXHzGb3jwKMQ