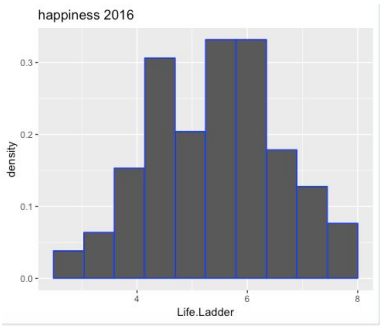
World Happiness Report: Data Analysis

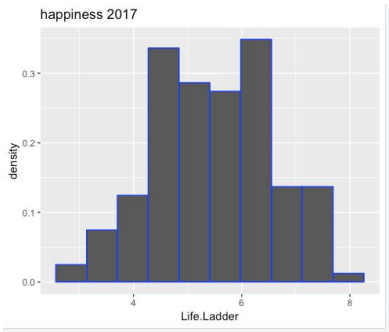
I/ Introduction

a) Background questions

How happy are people today? What are the factors that have a strong influence on happiness? Are there notable differences in happiness levels between different countries and between different periods? What accounts for these differences? Can happiness be even accurately measured? These are not only questions that social scientists have sought to answer for years, but also important questions that affect all of us. This paper is going to try to answer them, looking at the effect specific variables have on happiness and at the geographical and period differences that appear in the dataset extracted from the World Happiness Report[1]. From Ortiz-Ospina and Max Roser's study "Happiness and Life Satisfaction"[2] and general beliefs, one can expect richer countries to be happier than poorer ones. Moreover, we expect important differences in happiness levels between countries, due to different levels of GDP, of health quality, of inequality, etc.

b) Study of the Life Ladder





The mean of the Life Ladder, considered as a happiness score for the World Happiness Report has increased from 5.399 to 5.486 in 2017, an increase of 1.6% in one year. However the spread, which was 4.967 in 2016, has increased to 5.126. While the maximum value largely increase in 1 year, the minimum value decrease. Thus, even if it seem that the happiness score increase overall, the inequalities are also increasing, and the less happy countries become less and less happy.

According to the Life Ladder data, we can see that the 5 worst ranked countries in 2016 were the Central African Republic, South Sudan, Tanzania, Rwanda and Haïti, whereas the 5 worst ranked countries in 2017 were Afghanistan, South Sudan, Yemen, Tanzania and Malawi. These countries belong to impoverished regions of the world (mainly African continent, least wealthy countries of Western Asia and the Antillaise country of Haiti), characterised by a particularly low GDP and little democratic quality. On the other hand, we can see that the 5 best ranked countries in 2016 and 2017 were Iceland, Netherlands, Denmark, Norway, Finland and Switzerland, Iceland, Norway, Denmark and Finland respectively. These are all European countries, all northern European except Switzerland. It is utmost particularly interesting that given the existing variability of the ranking, particular world regions characterize the extreme positions on the rankings.

Hypothesis testing: We have noticed that countries with a log GDP per capita higher than 10.4 had on average a higher Life Ladder score in the 2017 data. Is this due to chance variation? In order to test this hypothesis, we will use the two box testing model. We will suppose that the data on the dataset is a sample of all the countries that exist in the Universe.
H0: Null hypothesis: a) This variation is due to chance. b) Both populations have the same average scores, diff=0
H1: Alternative hypothesis: a) The variation is not due to chance. b) Countries which have a log GDP per capita larger than 10.4 have a higher average Life Ladder score.

[1] http://worldhappiness.report/ed/2018/
[2] https://ourworldindata.org/happiness-and-life-satisfaction

Box A: 2017 Life Ladder scores for countries with have a log GDP per capita higher than 10.4 in the dataset. There are 30 countries in the data set, which will be used as a sample, and we'll suppose the box contains a very large (>300) number of countries which have a log GDP per capita higher than 10.4.
Observed average = 6.399141 SDbox=1.247951 SEavg1=0.2241386

Box B: 2017 Life Ladder scores for countries with have a log GDP per capita lower than 10.4 in the dataset 2016. There are 111 countries in the data set, which will be used as a sample, and we'll suppose the box contains a very large (>1110) number of countries which have a log GDP per capita lower than 10.4
Observed average = 5.228073 SDbox= 0.9135201 SEavg2=0.08670749

SEdiff= $\sqrt{SE1 * SE1 + SE2 * SE2}$ =0.2403254

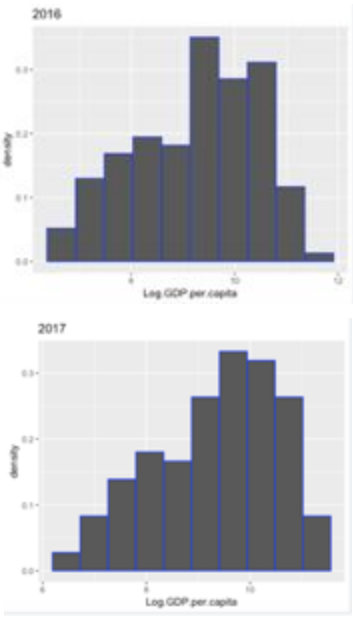
Test statistic: z=(Observed difference - Expected difference)/SEdiff = 4.872842

P-value: 5.50021e-07

Conclusion: a) We can reject the null hypothesis at 1% level
b) The difference in Life Ladder scores between countries which have a log GDP per capita higher than 10.4 and those which have a lower one is not due to chance. Countries with a log GDP p.c. higher than 10.4 are happier than the ones with lower GDP: this could mean that money makes happiness, or it could be due to something else.

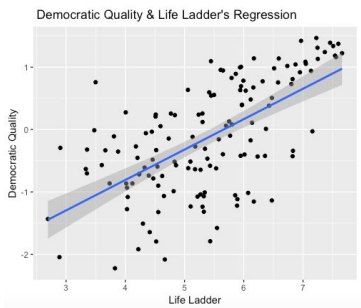
II/ Exploratory Statistics

- **Log GDP per capita in PPP (Hugo Savoyat)**



Through the summary statistics we see that he mean of Log GDP per capita index increased between 2016 and 2017. It is 9.274 in 2016 and 9.341 in 2017.This increase can appear similar to the increase in the mean of the Life Ladder. In 2016 the minimum value was 6.5 and the maximum value was 11.5, whereas in 2017, the minimum value was 6.6 and the maximum value still 11.5. The spread has been reduced by 2% in 1 year, thanks to an improvement of GDP per capita in purchasing power parity (PPP) in the poorest countries. We can first observe that in 1 year, there has been a shift to the right that translate an increase in GDP per capita in PPP in some countries. The GDP per capita in PPP in the poorest countries seems to have increase as the left bin has been reduced by at list a factor 2. Moreover, a lot of countries seem to have join the wealthiest countries, as the proportion of countries having a Log GDP per capita in PPP above 11 has been multiplied by approximately 7. Nevertheless, the proportion of very happy countries plummeted between 2016 and 2017. The summary statistics may forecast a positive correlation between GDP per capita in PPP and happiness, but it seems relevant to see if the countries at the extremes have the same linkage.

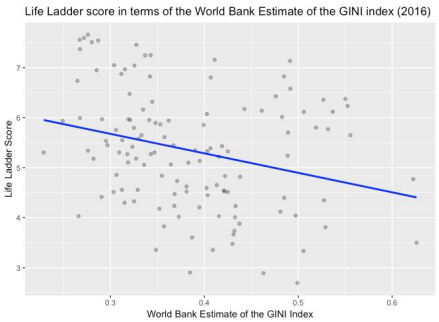
- **Democratic Quality & Confidence in National Government (Thomas Carrié)**



We decided to gather those 2 indicators as the 2nd one is related to the first: Democratic Quality encompasses governments' effectiveness through its delivery effectiveness criterions. Thus, we should expect a correlation between the two

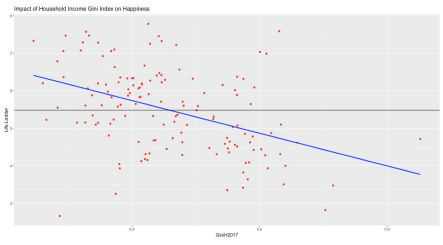
indicators. This intuition is false, as the correlation between the 2 indicators is only of -0.16, which is not statistically significant. Taken separately, both indicators are revealing interesting situations for different reasons: on the one hand, confidence in national government has a correlation of -0.15 with the Life Ladder index, thus trusting or not your national government does not seem to influence your happiness. On the other hand, democratic quality has a significant correlation with the happiness index (0.63). However, this is clearly weaker than expected regarding the central role of democracy in human happiness, especially according to the former president of India, the world's largest democracy, A.P.J. Abdul Kalam, for whom happiness is a central question in a democracy, as he stated: "in a democracy, the well-being, individuality and happiness of every citizen is important for the overall prosperity, peace and happiness of the nation."[3] However, this vision seems contradicted by the number of outliers in the regression.

Gini World Bank (Alba Nouet)



According to the data, we can observe that there is quite little correlation (-0.2805007 in 2016 and -0.2662538 in 2017) between the World Bank Estimate of GINI index and The Life Ladder scores. The negative correlation implies that the greater the GINI index, the smaller the happiness, but the small absolute value of the correlation coefficient means that this association is only true in broad terms, which we can see with the large scattering of the data on the plots. On top of this, we can observe both on the 2016 and 2017 data which have a GINI index smaller than the mean seem to follow more closely a steeper regression line. For the 2016 data, the correlation between the World Bank's estimate of the GINI index of the countries for which the index is smaller than the mean for the dataset, and the Life Ladder is -0.3946803. Considering that the GINI index measures the extent to which the distribution of income among individuals or households within an economy deviates from a perfectly equal distribution, as it measures the area between the Lorenz curve and a hypothetical line of absolute equality, expressed as a percentage of the maximum area under the line, the more egalitarian a country is, the happier it tends to be, and this association is stronger for countries with high equality. On top of this, the study of this variable is particularly interesting, since it shows the limits of linear regression; instead, the data seems to be better fit by another type of regression curve, a polynomial one, for example.

- **Gini Household (Martin Lepercq)**



Given that the mean as well as the SD of the Gini household income index are very close in 2016 and in 2017, we will run the regression focusing only on 2017 data.

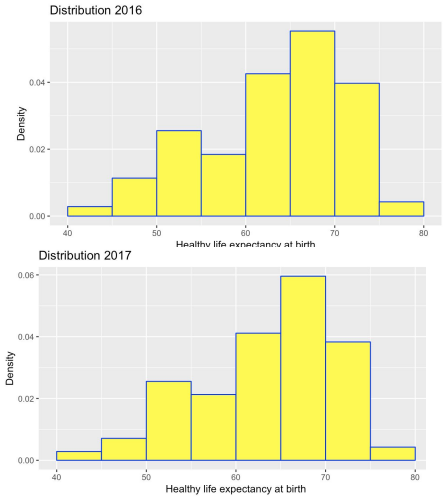
In this regression the intercept cuts the Y axis at 7.484, and the slope of the SD line is equal to -4.358.

Consequently it seems that increasing equality among households leads to increasing happiness.

Plugging the happiness average we notice that happy societies are equal societies. Nevertheless happiest countries are not necessarily the most egalitarian. For instance the 18 happiest countries have a household income Gini index going from 0.1 to 0.6 points. As well as for the 10 saddest countries where the household Gini index starts from 0.1 to 0.7. Consequently the relation we observed, seems to be true for countries around the average, however for extreme values the relation between equality and happiness is vaguer. As the cloud of points shows, the regression between the Gini index for households and happiness is not that strong. We observe many outliers and the point are not well-gathered around the regression line. Indeed, for 2017 data, the correlation factor is equal -0.43.

[3] A. P. J. Abdul Kalam Quotes. BrainyQuote.com, BrainyMedia Inc, 2018. https://www.brainyquote.com/quotes/a_p_j_abdul_kalam_589740, accessed December 5, 2018

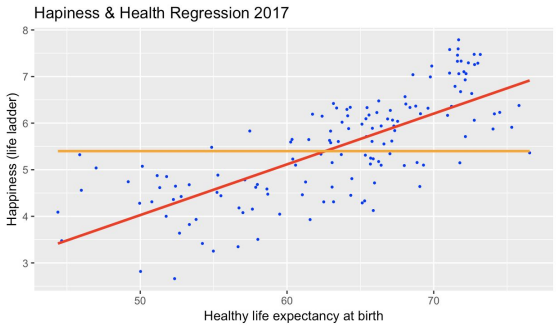
- **Healthy Life Expectancy at Birth (Oscar Chaix)**



The mean and median of the healthy life expectancy at birth have risen from 2016 to 2017, from about 63.15 to 63.40 years and 64.9 to 65.1 years respectively. This can be explained by improving medical services and sanitation worldwide. The spread of healthy life expectancy at birth is relatively large, going from minimums of about 44 years to maximums of about 76 years. The gap between rich and poor countries concerning health is thus especially wide. The standard deviation for the 2016 distribution is in fact about 7.8 years, and the standard deviation for the 2017 distribution is about 7.6 years. The distribution has therefore not only slightly shifted to the right in a year, but the spread has also shrinked, as the different SDs attest. Both distributions are left-tailed and do not follow the normal curve, confidence intervals are therefore irrelevant for the health variable. However, as we will see in part III, the correlation between health and happiness is particularly interesting.

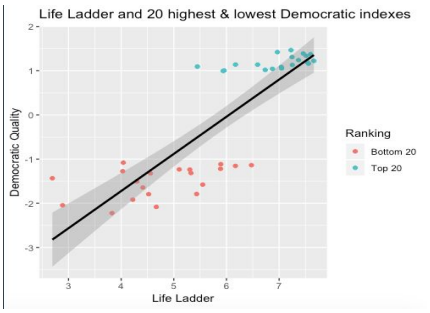
III/ Variables that influence Life Ladder

Healthy Life Expectancy at Birth



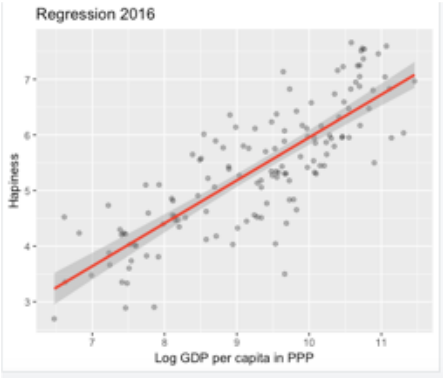
The correlation we find between life expectancy at birth and life ladder is of about 0.78 for 2016 and 0.74 for 2017. Being relatively similar, we will simply analyse 2017's correlation and regression line. The correlation looks strong and positive: the regression line's slope is about 0.109, and the cloud of data is mostly concentrated around it and upward sloping. It therefore seems that with increasing healthy life expectancy at birth, sample populations are happier. Points are especially tightly clustered above the orange line, which corresponds to life ladder's average of about 5.5. This shows the relation is even stronger for countries with higher healthy life expectancy and happiness levels, as most residuals above the orange line and beyond 60 years of healthy life expectancy are small and largely positive. In fact, the "happiest" countries are mostly concentrated after the 70 years of healthy life expectancy threshold, with a mean life ladder of about 6.7 as opposed to about 5.1 for preceding countries. Noteworthy, also, is the fact that the six countries with the highest life expectancy at birth have lower life ladder scores than the countries directly preceding them, which seems to show that past some point, increasing healthy life expectancy does not increase happiness. This is similar for the ten countries with the lowest healthy life expectancy at birth: most score better than the succeeding points and they do not follow the regression line. Computing the correlation between health and happiness for these extreme countries, the bottom 10 countries and top 6 countries in terms of healthy life expectancy at birth, therefore shows correlations coefficients of about -0.04 and -0.47 respectively. The correlation we at first found thus does not apply for countries at both extreme, on the contrary we find negative relationships at these levels. Finally, it is important to note that, even though the correlation is strong, healthy life expectancy at birth is itself very highly correlated to GDP, with an r of about 0.86 in 2017. The strong positive association we witness between health and happiness is therefore mostly the indirect result of a high correlation between wealth and health.

Democratic Quality (Data available only for 2016).



We decided to dig a bit more in the correlation between life ladder & Democratic quality and thus selected the 20 countries with the highest Democratic Quality index, and the 20 lowest ones. Interestingly, we now find a correlation of 0.84, which is 0.19 points higher than the overall correlation, turning it into a strongly significant one. Continuing in this approach by now selecting the 21st to 40th higher & lower Democratic Quality indexes, the correlation is slightly above the whole dataset's one, at 0.63. Even more striking, by doing so with ranks from the 41st to the 60th, the correlation is now only of 0.29, which is not significant anymore. A tentative explanation of this football shaped phenomenon would be the sensitiveness to democracy: countries with highly efficient democracies tend to be the ones in which citizens are cherishing the most this regime, acknowledging the benefits they owe to their regime. In the opposite way, countries with the lowest democratic qualities could be the one suffering the most from the absence of democracy, having a strong impact on their happiness. The countries with average scores would be simply less aware of the benefits they owe to their democracies, as their quality is not poor enough to strike them, but not good enough either to content them. However, countries with high democratic scores tend to be the richest ones (in terms of GDP, 0.62 correlation), so it could be also a simple collateral effect of richness on happiness. However, countries as Saudi Arabia or Bahrain, outliers of the bottom 20 regression, are also showing that some countries can be extremely wealthy & have a good Life Ladder index while having a terrible democratic score.

GDP per Capita



In general, both years show strong positive correlation between GDP per capita in purchasing power parity (PPP) and happiness. Despite concluding an association between wealth and happiness, we might state that being wealthy, on average, as a citizen of a country is highly correlated to the overall average score of happiness in that country. However, when we study the 10 wealthiest countries in GDP per capita in 2016, we find that the correlation between GDP per capita in purchasing power parity (PPP) and happiness is 0.04. This is a very low positive correlation, that shows that in the 10 wealthiest countries of the world, other factors have to be taken in account for happiness. For example, Kuwait, the third wealthiest country in 2016 (in terms of GDP per capita in PPP), only have a Life Ladder score of 5.95, which is about average. Nevertheless, Kuwait has a poor democratic quality index (-0.42), that can explain a lower Life Ladder than others of the wealthiest countries of the world, despite a third position in the ranking of the wealthiest countries in the world in GDP per capita in PPP.On the other hand, in 2016, in the 10 less wealthy countries, the correlation between between GDP per capita in purchasing power parity (PPP) and happiness is 0.49. Compared to the correlation in the wealthiest countries we can see that, among the less wealthy countries, GDP per capita is more highly correlated with happiness. However, the correlation is still smaller that the correlation computed for all the countries. We might see wealth as a happiness threshold. Indeed, when we look at the top 10 wealthiest countries, all, except 1, their life ladder is above the average, which is about 5.45 (mean of the 2016 average and 2017 average), both in 2016 and in 2017. Furthermore, in the groups made of the 10 less wealthy countries in 2016 and in 2017, none of them has a happiness score above the average of 5.45.

IV/ Countries that have an evolution in their Life Ladder

Studying the data frame, we discovered that the largest increase in Life Ladder score from 2016 to 2017 was in Guinea, with an increase of 1.270868, which is a 35.3% increase. In order to try to understand which factors contributed to this significant increase in happiness, we will study factors which significantly changed from one year to the other. We can observe a 0.25% increase in log GDP per capita, 6.1% decrease in social support, 0.96% increase

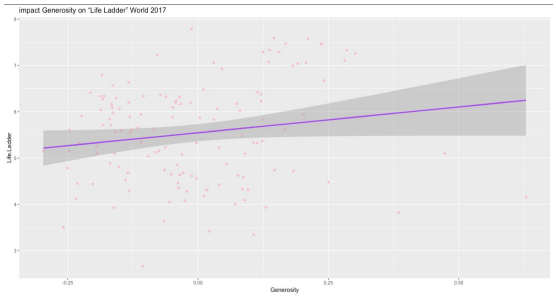
in health life expectancy at birth, 1.7% increase in freedom to make life choices, 225.1% increase in generosity, a 6.6% decrease in perception of corruption, 37.4% increase in confidence in national government and a 18.9% increase in GINI of household income. The most significant changes reported in the data set are therefore the important increase in generosity and the increase in confidence in national government.

Guinea is a Western African country, with an economy mainly dependent on mineral production and agriculture, characterized for being at the core of the Ebola outbreak in 2014. The increase in confidence in the national government reported in the dataset could reflect the arrest of the former Prime Minister of Gabon in August 2016, charged with bribing officials in various African countries, including Guinea, in order to secure mining concessions[4]. We could deduce that this had a significant impact in lowering the confidence in national government at first (characterized by a specially low index in 2016) as they learnt the news. However, the longer-term impact of this arrest and investigation could result in an increased confidence in the country's judiciary institutions. On top of this, Conde's first cabinet held a successful political dialogue between the government and the opposition to address long-standing tensions in August and September 2016 (therefore towards the end of the year), which may have well contributed to some extent to the increased confidence in the national government in 2017. Confidence in national government might could be associated to Guinea's increased score in the happiness index, however, from Thomas' regression above, this factor doesn't seem to play a significant role for other countries' happiness indexes, as the correlation coefficient is only -0.15.

The very significant increase in generosity, which is the residual of regressing the national average of GWP responses to the question "Have you donated money to a charity in the past month?" on GDP per capita, could be associated to the increased Life Ladder score (see regression below)

However, this significant increase in the Guinean reported happiness could well be due to changes in other factors which do not appear listed in this data set, or even by biases related to the way the data was intaken.

Studying the data frame, we discovered that the biggest drop in happiness for a country was in Afghanistan. Between 2016 and 2017, the overall level of happiness dropped by 1.56 points. Indeed in Afghanistan, 2016 was one of the most deadly year in the war against Talibans. An article published in the New York Times in April 2016 and written by Mujib Mashal[5], revealed that the number of deaths caused by pro-government forces increased by 70 % compared to 2015 at the same period.The injuries are particularly affecting children, he also said. This climate creates an environment of mistrust where generosity is very is low . Those informations highlighted the importance of generosity in happiness. Indeed in Afghanistan between 2016 and 2017, the level of generosity dropped from 0.057 to -0.106.



It would be interesting to see if, in the world, generosity is as important as in Afghanistan to predict happiness.

When we regress the variable "generosity" on the outcome variable "Life.Ladder" we discover that overall "generosity" doesn't have an impact on happiness. In 2016 as well as in 2017 the regression doesn't show a clear trend in the cloud of points. Moreover in 2016 and in 2017 the correlation factor is very low. It's equal respectively to 0.12 and 0.16.

In both Afghanistan and Guinea generosity seems to be the starting point of a happy society, however it is not the case everywhere. It reveals a huge limit of our study. We want to isolate a global a determinant of happiness, however cultural and historical differences between countries make this task very hard.

---

[4] Stevenson, Alexandra (2016-08-16). "Bribery Arrest May Expose African Mining Rights Scandal Tied to Och-Ziff". *The New York Times*. ISSN 0362-4331. Retrieved 2017-02-19.

[5] https://www.nytimes.com/2016/04/18/world/middleeast/civilian-casualties-in-afghan-war-are-unabated-in-2016.html?rref=collection%2Ftimestopic%2FAfghanistan&action=click&contentCollection=world&region=stream&module=stream_unit&version=search&contentPlacement=2&pgtype=collection

V - Conclusion and Limits

Thanks to this analysis, it is possible to distinguish some variables that are significantly positively correlated to the happiness score and some that are negatively correlated to the happiness score. Furthermore, some changes in one country's Life Ladder could find significations in changes in some variables. Therefore, this paper supposes a positive influence of wealth, life expectancy, and the democratic characteristics of a political regime on happiness within a country. Indeed, these are the variables that are the most largely correlated to the Life Ladder, which seeks to accurately measure happiness score. However, it might appear too ambitious to determine the factors of happiness, which is a peculiarly subjective variable, thanks to quantitative data. Moreover, this paper can only underline associations between these variables and happiness, but cannot prove causation. Some limits of the regression method can also be highlighted by this data analysis. Regression is interesting to get a broad correlation between two variables, however, when the scatter plots are heteroscedastic the regression loses its relevance for the extreme values. Among other conclusions, this limit of the regression method allows us to state that, despite a very strong correlation between GDP and the Life Ladder, the richest countries are not necessarily the happiest.