Cjache Kang, Vatsal Bajaj, Vedant Kajaria, Nimit Baid, Bansari Rao
Shobhana M. Stoyanov
Group 42

Stagnated Happiness

INTRODUCTION

Decades of trying to place a cardinal value on happiness resulted in a report by the UN—The World Happiness Report. The World Happiness Report is a landmark survey of 100+ countries, focusing on the changes in happiness around the world. The report collects data on variables such as GDP per capita, healthy life expectancy and Generosity. As illustrated in Table 1, we have segmented these covariates under three broad topics: social, economic and political indicators.

Table 1: Categorization of Covariates under three topics.

Social Indicators	Political Indicators	Economic Indicators
Generosity	Perceptions of Corruption	Log GDP per Capita
Social Support	Confidence in national government	GINI Index
Freedom to make life choices	Democratic Quality	GINI index of household income
Generosity	Delivery Quality	Healthy Life Expectancy at birth
Positive / Negative Effect		

This report focuses on life ladder as an umbrella indicator of happiness that we compare to other social, political and economic indicators. The first question that we explore is whether world happiness has increased from 2016 to 2017. Our hypothesis test concludes that this is not the case. The second question we look at is whether the GINI index of household income correlates with our happiness indicator. Our regressions conclude that there is a negative relation between the two, which illustrates a positive relation between between equal distribution of income and world happiness.

EXPLORATORY DATA ANALYSIS

As a preliminary analysis, we observed the relations of the life ladder with two political, social and economic indicators each as illustrated in Figure 1. We assumed that the World Happiness Report conducted surveys on subjects who were people of various countries. Moreover, to measure outcomes, the survey utilized number scales. For example, the life ladder, or the happiness score, represents the national average response between 0 - 10 to a survey question inquiring about overall happiness. Log GDP per capita represents the average income in an economy adjusted for different exchange rates. GINI of household income is a number between 0 to 1 showcasing the equality of distribution of income within an economy. Countries with a GINI coefficient closer to 0 are more equitable. Social support represented binary responses to surveys about whether people have a strong support system. Freedom to make life choices and Perceptions of corruption are national averages of the survey responses to related questions. Confidence in national government similarly show survey responses. This scores for all three of these are between 0 and 1.

Figure 1: Scatter plots of different variables against Life Ladder to indicate relationships
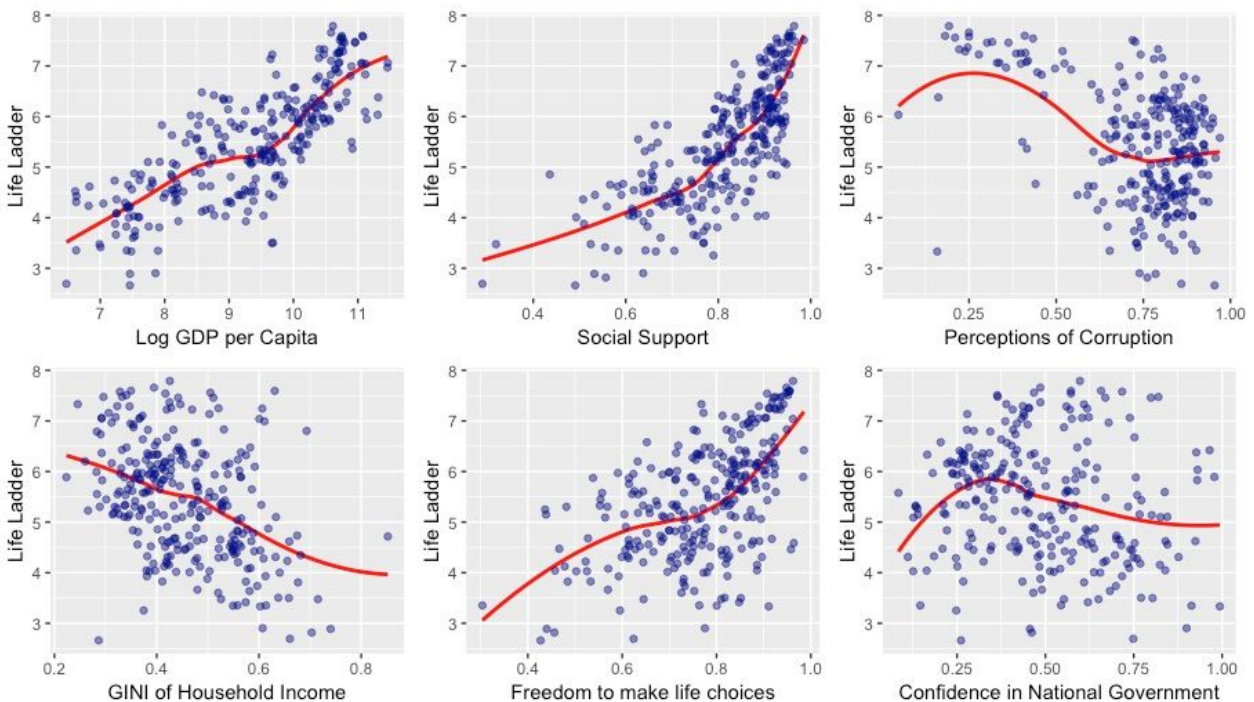
We observed a positive relationship between the Log GDP per capita and Life Ladder. This illustrates that increases in average income per person in an economy were correlated with higher scores on the life ladder. As expected, this positive relation was also the case for social support and freedom to make life choices. Perceptions of corruption and confidence in national government, however, did not show linear relationships at all. The possibly negative relationship between the GINI of Household Income and Life Ladder, however, is striking since it correlates equality of distribution of income with world happiness.

Next, we explored the life ladder data in detail. We calculated the change in life ladder from 2016 to 2017 and made a histogram to visualize this. The histogram shows how most people's happiness remained stagnant with the life ladder differential from 2016-2017 being zero. There, is however, a right skew if we don't account for the outlier values in Figure 2. This might indicate an increase in happiness from 2016 to 2017. This increase is demonstrated in the box plot showing higher life ladder scores for 2017 since there is a higher median, upper quartile and lower quartile score. This poses the question: was this increase in happiness statistically significant or is it just due to chance variation. Our hypothesis test puts this question to rest.
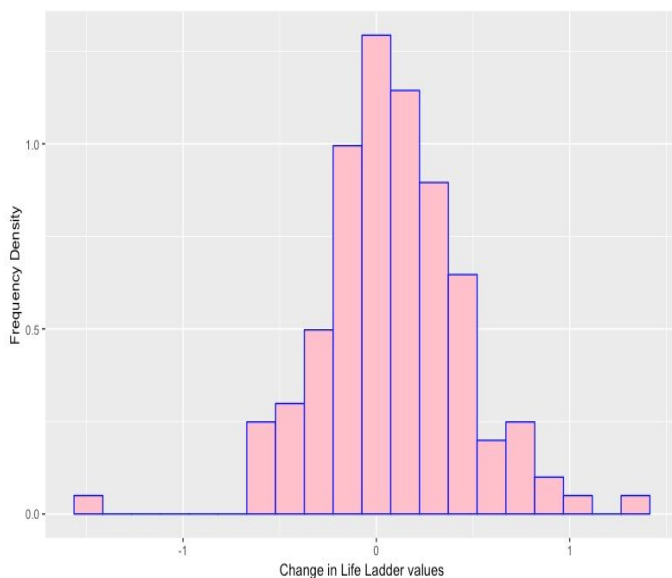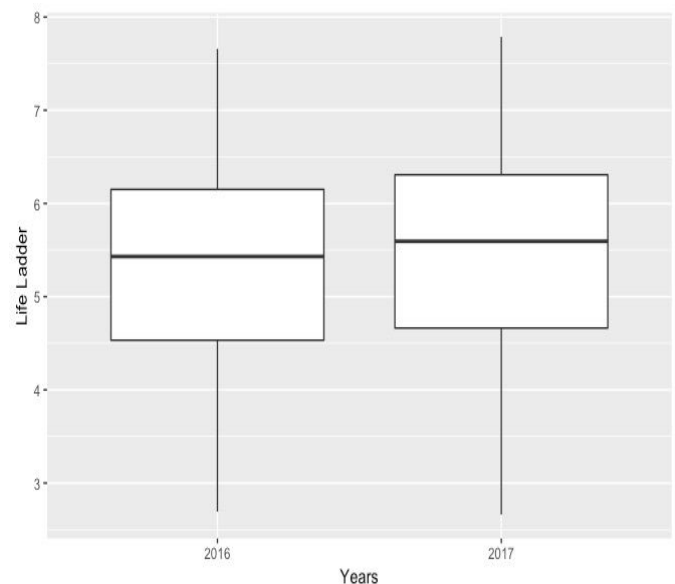
Figure 2: Histogram for change in life ladder

Figure 3: Box plot showing life ladder score

FORMAL DATA ANALYSIS

Assumptions:

We assumed that data was collected using simple random sampling and that there was no non-response bias. Moreover, we assume that the people answered honestly. For the z-test, we are assuming a normal approximation.

Hypothesis Testing

Null Hypothesis: H_0: There has been no increase in happiness from 2016 to 2017 since the difference is only due to chance variation.

 $Happiness_{2017}$ - $Happiness_{2016}$ = 0.

Alternative Hypothesis: H_1: There was a positive change in happiness from 2016 to 2017.

$Happiness_{2017}$ - $Happiness_{2016}$ > 0.

This calls for a one-tailed hypothesis test to check whether the increase in happiness was statistically significant. We had data for 135 countries for both 2016 and 2017 which we used to calculate the mean, standard deviation and standard error to form a z score as can be seen below.[1]

Table 2: Hypothesis Test calculations with mean, SD and SE

	2016	2017
Mean	5.42	5.51
Standard Deviation	1.13	1.13
Standard Error	0.097	0.097

$$SE(2017 - 2016) = \sqrt{SE(2017)^2 + SE(2016)^2} = \sqrt{(0.097)^2 + (0.097)^2} \approx 0.14$$

Now, we perform a one tailed z-test on this data:

$$z = \frac{(Observed\ Value - Expected\ Value)}{Standard\ Error} = \frac{(0.0868 - 0)}{0.137} \approx 0.632$$

---

[1]All values are rounded off to three significant figures

P value ≈ 0.264 ≈ 26.4%

Since the p-value is 26.4%, we accept the null that the change in happiness from 2016 to 2017 is due to chance variation. Therefore, world happiness remained stagnant from 2016 to 2017.

Regression Analysis

World happiness has, therefore, remained stagnant from 2016 to 2017. A second interesting observation from our exploratory data analysis was the negative relationship between GINI coefficient and the life ladder. Our regression shows whether such a relationship actually exists. We noticed a possible negative correlation between the GINI coefficient of Household Income and Life Ladder. Since each country had two data points—one for the year 2016 and one for the year 2017—we did two separate analyses. Using R, we found the correlation to be -0.42. Furthermore, we performed regression analysis on each year for the GINI coefficient. We created residual plots, discovering that because the residual plots did not have a pattern, the scatter plots were, in fact, homoscedastic. Thus, the regression analysis was valid.

The GINI coefficient is a measure of the distribution of income in a given country with a value of 0 meaning that income is equally distributed. Thus, since there is a negative correlation between the GINI coefficient and Happiness, it is interesting to note that as the gap between the rich and poor becomes wider, a given country becomes unhappier.

Figure 4: Regression Lines of GINI of Household Income vs Happiness
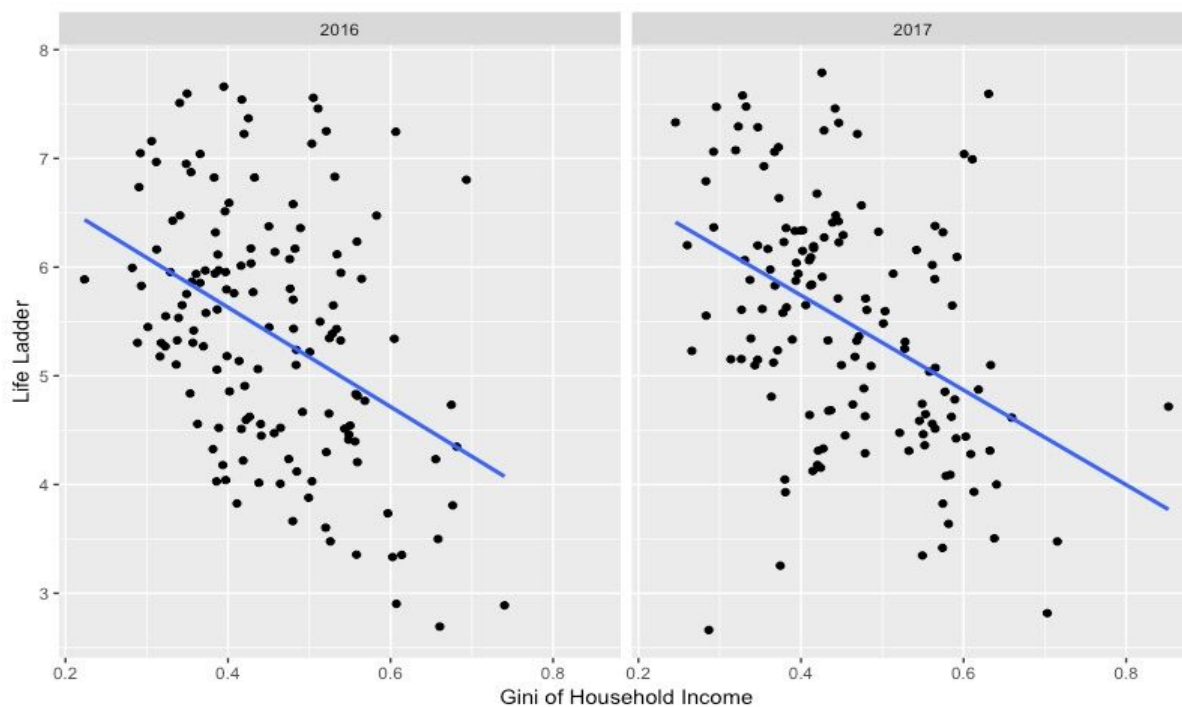
Figure 5: Residual plots of GINI of Household Income vs Happiness - showing a random scatter
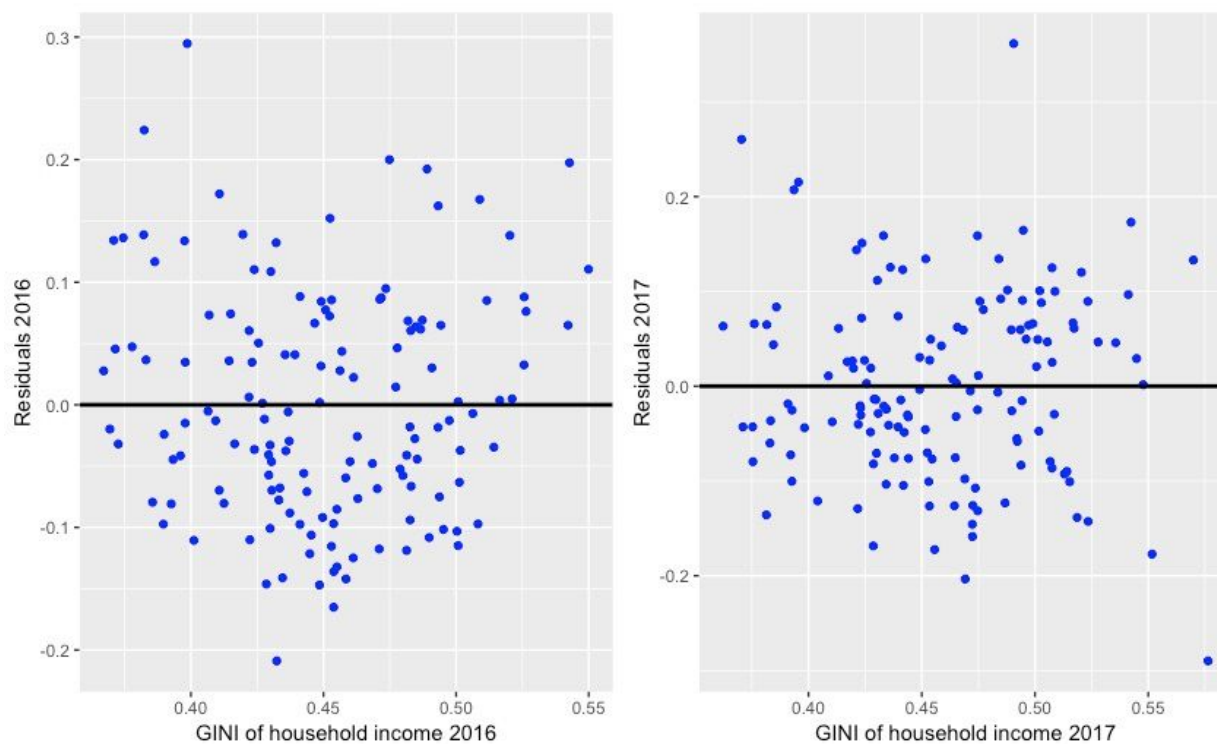
CONCLUSION: FINDINGS AND RESULTS

Through our formal data analysis, we have two integral findings. First, the hypothesis test illustrates that happiness has remained stagnant from 2016 - 2017. We chose the z-test of two independent samples (2016 and 2017) as it fits the hypothesis best. A t-test would be inappropriate due to the large sample size, and chi-square tests is not valid: the test for Goodness of Fit required expected value which is hard to calculate with the given data, and the test for Independence was hard to use due to the design of our analysis.

Second, the regression illustrates that countries with more equitable distribution of incomes have a fairly higher happiness levels since GINI of household income and the life ladder have a fairly negative relation. The homoscedastic scatter plot and post-regression residual plot helped us conclude the appropriateness of Regression Analysis. Hence, a fairly negative correlation was observed in the GINI coefficient and the Level of Happiness. This cannot implicate causality; however, it is a lesson for the world that highly equitable countries are correlated with higher happiness levels.

We made certain key observations: the data wasn't weighted by the population of the countries, placing equal weight on countries with varying population sizes. This can affect the outcome of the happiness indicator. Second, questions to measure Positive Effect and Negative Effect are not devoid of lurking variables, as they ask the responder about their level of happiness or sadness from the previous day, but do not account for the possibility that such levels can vary from day to day and are not representative of general happiness in their lives. Besides, the data available is only for two years, which is a very short period of time for significant changes in factors on the global scale. This data, however, has tremendous scope of research and we only analyze two questions which can be furthered with more research and data analysis.