Happiness Scores For Global Citizens

Lynn Fan, Annie Lin, Leona Lu, Eric Young
Recording URL: https://duke.box.com/s/apahfiniquqco8ekv9dicyvr7uc9lqca

Research Questions

1. How are parameters such as GDP per capita and life expectancy related to happiness scores of citizens among all countries?
2. Do developed countries overall have a different set of statistically significant parameters than developing countries?
3. Is there any country experiencing significant increase/decrease in happiness score over the duration of the Report, and which parameters are the most significant in affecting the changes?

Relevance and Motivation:
By reflecting the impact of big events and government policies on the happiness of populations, the report allows policymakers and researchers to evaluate past policies and to make informed future political decisions.

Data Sources

World Happiness Report

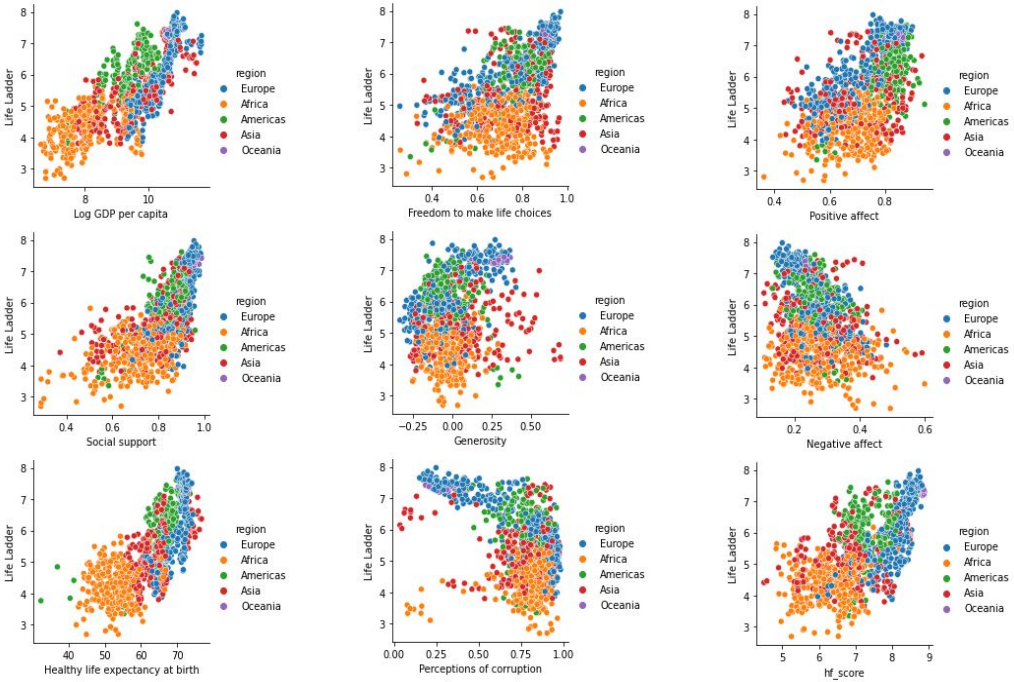
- Obtained from Kaggle
- 5 csv files that detail happiness scores and rankings of 155 countries in 2015, 2016, 2017, 2018, and 2019 respectively
- Data columns representative of 6 factors - economic production, social support, life expectancy, freedom, absence of corruption, and generosity

The Human Freedom Index

- Developed to measure economic freedom metrics of different countries, such as the freedom to trade or to use sound money
- Would the human freedom index, calculated by another research group, help explain happiness scores?

1. How are parameters such as GDP per capita and life expectancy related to happiness scores of citizens among all countries?

Exploratory Data Analysis



Model Results

	coef
Social support	2.018753
Freedom to make life choices	1.002960
Generosity	0.904757
Log GDP per capita	0.536794
Perceptions of corruption	-0.723674

MSE: 0.36

R^2 score: 0.73

For social support to decrease by 1 unit (a range of 0 to 1), with other variables remaining the same, Happiness score is predicted to decrease by 2.018753 (a range of 0 to 5).

Exploring Multicollinearity

	variables	VIF
0	Log GDP per capita	376.941248
1	Social support	104.364910
2	Healthy life expectancy at birth	298.922636
3	Freedom to make life choices	38.531594
4	Generosity	1.208316
5	Perceptions of corruption	11.286572
6	hf_score	146.905392

	variables	VIF
0	Log GDP per capita	109.251525
1	Social support	103.346769
2	Freedom to make life choices	35.637514
3	Generosity	1.206848
4	Perceptions of corruption	10.926971

Limitation and Future Work

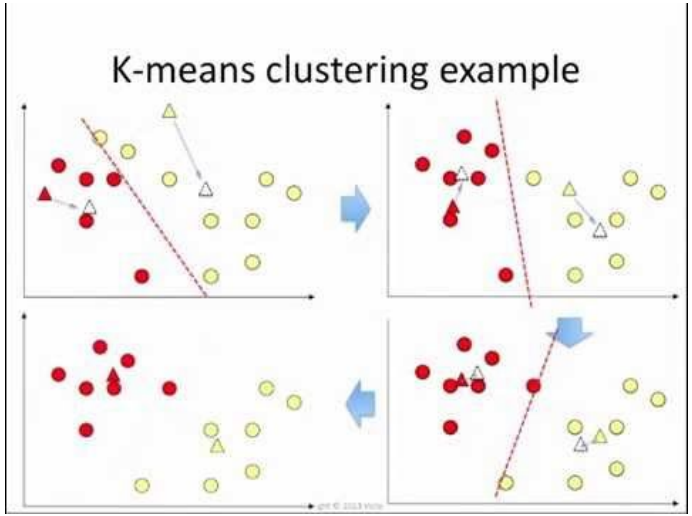
In the first question, multicollinearity problems still persist in the model after the model selector excludes two variables from the full set of variables. More work could be done to mitigate these issues, potentially by looking to more detailed variables and exploring why these variables are highly correlated to each other. In addition, a linear regression model might not capture higher-order variable relationships. Due to the time constraints, we did not use other models and compare their robustness with each other, and a linear regression model might not be the most suitable model for the research problem after trying more models.

2. Do developed countries overall have a different set of statistically significant parameters than developing countries?

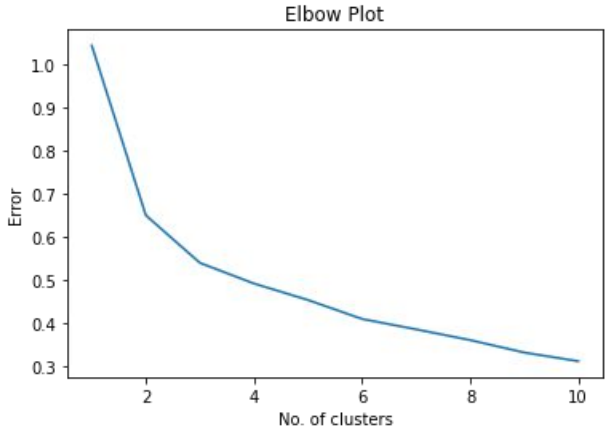
Methodology

K-Means Clustering (& Elbow Method)

1. Initialization
2. Cluster Assignment
3. Centroid Movement



Results



Group 0	Group 1	Group 2
'Finland', 'Denmark', 'Norway', 'Iceland', 'Netherlands', 'Switzerland', 'Sweden', 'New Zealand', 'Canada', 'Austria', 'Australia', 'Israel', 'Luxembourg', 'United Kingdom', 'Ireland', 'Germany', 'Belgium', 'United States', 'Czech Republic', 'United Arab Emirates', 'Malta', 'Mexico', 'France', 'Taiwan', 'Chile', 'Saudi Arabia', 'Qatar', 'Spain', 'Panama', 'Brazil', 'Uruguay', 'Singapore', 'Italy', 'Bahrain', 'Slovakia', 'Trinidad & Tobago', 'Poland', 'Lithuania', 'Colombia', 'Slovenia', 'Argentina', 'Romania', 'Cyprus', 'Kuwait', 'Thailand', 'Latvia', 'South Korea', 'Estonia', 'Mauritius', 'Japan', 'Kazakhstan', 'Hungary', 'Northern Cyprus', 'Peru', 'Portugal', 'Russia', 'Serbia', 'Libya', 'Montenegro', 'Croatia', 'Hong Kong', 'Dominican Republic', 'Bosnia and Herzegovina', 'Turkey', 'Malaysia', 'Belarus', 'Greece', 'Mongolia', 'North Macedonia', 'Turkmenistan', 'Algeria', 'Morocco', 'Azerbaijan', 'Lebanon', 'China', 'Bulgaria', 'Jordan', 'Gabon', 'South Africa', 'Albania', 'Venezuela', 'Namibia', 'Armenia', 'Iran', 'Georgia', 'Tunisia', 'Iraq', 'Sri Lanka', 'Ukraine', 'Egypt', 'India', 'Botswana'	'Pakistan', 'Tajikistan', 'Nigeria', 'Cameroon', 'Ghana', 'Ivory Coast', 'Benin', 'Congo (Brazzaville)', 'Somalia', 'Niger', 'Burkina Faso', 'Guinea', 'Gambia', 'Mozambique', 'Congo (Kinshasa)', 'Sierra Leone', 'Chad', 'Togo', 'Liberia', 'Comoros', 'Burundi', 'Haiti', 'Syria', 'Malawi', 'Afghanistan', 'Central African Republic', 'South Sudan'	'Costa Rica', 'Guatemala', 'El Salvador', 'Uzbekistan', 'Nicaragua', 'Kosovo', 'Ecuador', 'Jamaica', 'Honduras', 'Bolivia', 'Paraguay', 'Philippines', 'Moldova', 'Kyrgyzstan', 'Indonesia', 'Vietnam', 'Bhutan', 'Nepal', 'Laos', 'Cambodia', 'Palestinian Territories', 'Senegal', 'Kenya', 'Mauritania', 'Bangladesh', 'Mali', 'Myanmar', 'Ethiopia', 'Swaziland', 'Uganda', 'Zambia', 'Madagascar', 'Lesotho', 'Zimbabwe', 'Yemen', 'Rwanda', 'Tanzania'
Score Mean: 5.901663043478262 Score SD: 0.9505040456558449	Score Mean: 4.298307692307692 Score SD: 0.7077769514598165	Score Mean: 4.960351351351351 Score SD: 0.9591030217249644
GDP Mean: 1.179945652173913 GDP SD: 0.2003505820854184	GDP Mean: 0.3277307692307692 GDP SD: 0.19867049598853714	GDP Mean: 0.6332702702702702 GDP SD: 0.2004677838441539
Social Support Mean: 1.3495760869565216 Social Support SD: 0.1969450504745606	Social Support Mean: 0.7398461538461538 Social Support SD: 0.2482849932432939	Social Support Mean: 1.1909999999999998 Social Support SD: 0.17562921630987718
Life Expectancy Mean: 0.8675760869565217 Life Expectancy SD: 0.14198095659933416	Life Expectancy Mean: 0.38799999999999996 Life Expectancy SD: 0.12155973523650515	Life Expectancy Mean: 0.621297297297973 Life Expectancy SD: 0.19428842656061143
Generosity Mean: 0.16431521739130434 Generosity SD: 0.09228805676036805	Generosity Mean: 0.2185769230769231 Generosity SD: 0.05972319036568308	Generosity Mean: 0.21137837837837833 Generosity SD: 0.10839750454676038
Corruption Mean: 0.12126086956521738 Corruption SD: 0.10903488368459618	Corruption Mean: 0.09626923076923079 Corruption SD: 0.05333894355354522	Corruption Mean: 0.09605405405405405 Corruption SD: 0.07149104874830284

Limitations & Future Work

Limitations:

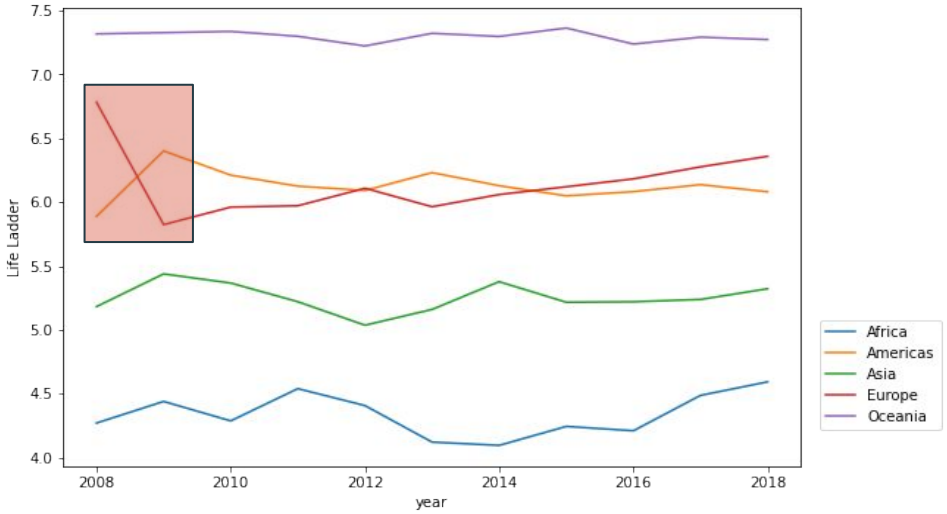
1. Subjective interpretation of the “elbow method”
2. Randomness of results based on initial centroid assignments

Future Work:

Conduct literature review on why developed countries tend to have a higher perception of corruption but lower generosity level compared to developing countries

3. Is there any country experiencing significant changes in happiness score over the duration of the Report?

While Oceania is stable, Europe has the most drastic change of Happiness Score from 2008 to 2018



Before 2015, America had the second highest average life ladder and Europe had the third highest, but Europe's average life ladder exceeds that of Americas in 2008. Asia has the fourth highest average life ladder, and Africa has the lowest.

While Oceania's happiness rating remained stable through 2008 to 2018. Europe has displayed drastic change in the year of 2008. It is also interesting to observe that trends of Happiness rating are similar between Asia and America, being the opposite direction of Europe.

What cause the drastic change of happiness score in Europe from 2008 to 2009?

Perceptions of Corruption plays an significant role in happiness score for both 2008 and 2009's report

Linear Regression Result for 2008

	coef	std err	t	P>|t|	[0.025	0.975]
x1	0.9362	0.834	1.123	0.294	−0.987	2.859
x2	−1.5350	3.239	−0.474	0.648	−9.004	5.934
x3	0.1035	0.079	1.313	0.226	−0.078	0.285
x4	2.8737	2.015	1.426	0.192	−1.774	7.521
x5	0.8174	0.602	1.359	0.211	−0.570	2.205
x6	−1.7783	0.998	−1.781	0.113	−4.080	0.524
x7	−2.9000	2.700	−1.074	0.314	−9.126	3.326
x8	1.4716	3.026	0.486	0.640	−5.506	8.449
x9	−0.9945	0.436	−2.279	0.052	−2.001	0.012

Omnibus:	0.524	Durbin-Watson:	2.474
Prob(Omnibus):	0.770	Jarque-Bera (JB):	0.579
Skew:	0.329	Prob(JB):	0.749
Kurtosis:	2.381	Cond. No.	4.26e+03

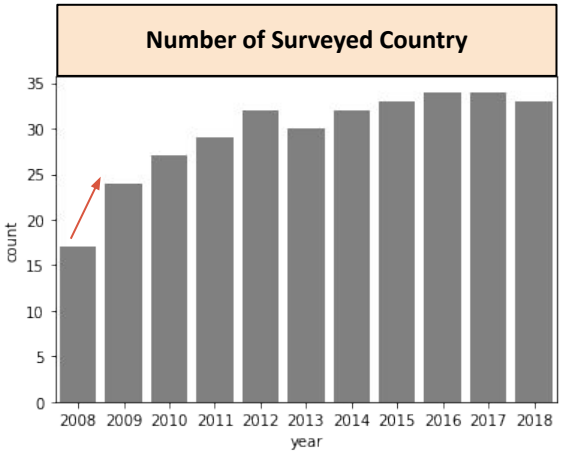
Linear Regression Result for 2009

	coef	std err	t	P>|t|	[0.025	0.975]
x1	−0.1874	0.276	−0.680	0.507	−0.775	0.400
x2	4.0678	2.006	2.028	0.061	−0.208	8.343
x3	0.2001	0.055	3.651	0.002	0.083	0.317
x4	0.1174	0.710	0.165	0.871	−1.396	1.630
x5	1.0833	0.684	1.585	0.134	−0.374	2.540
x6	−1.9433	0.507	−3.830	0.002	−3.025	−0.862
x7	−3.4264	2.224	−1.541	0.144	−8.167	1.314
x8	−4.9891	1.334	−3.740	0.002	−7.833	−2.146
x9	−0.5343	0.268	−1.995	0.064	−1.105	0.036

Omnibus:	0.508	Durbin-Watson:	2.727
Prob(Omnibus):	0.776	Jarque-Bera (JB):	0.606
Skew:	0.149	Prob(JB):	0.739
Kurtosis:	2.281	Cond. No.	3.14e+03

Log GDP per capita	Social Support	Healthy Life expectancy at Birth	Freedom to make life choices	Generosity	Perceptions of corruption	Positive affect	Negative affect	hf score
x1	x2	x3	x4	x5	x6	x7	x8	x9

Limitation: the report of 2008 centered on Northern and Western Europe; the report of 2009 centered on Eastern Europe



In 2009, the report include more Eastern European countries:

"Bosnia and Herzegovina", "Croatia", "Finland" ,"Germany", "Greece", "Hungary","Luxembourg","Montenegro","North Macedonia" , "Poland", "Romania", "Serbia","Slovenia", "Switzerland".

And the report leaves out certain Western European countries:

"Austria", "Belgium", "Finland", "Iceland", Netherland", "Norway","Portugal".

Category Mean	Freedom to make life choices	Generosity	Healthy life expectancy at birth	hf_score	Log GDP per capita	Negative affect	Perceptions of corruption	Positive affect	Social support
2008	0.789824	0.023941	69.445882	8.2865	10.607118	0.206294	0.639529	0.736118	0.924
2009	0.650708	-0.045	68.180375	8.0313	10.228	0.274875	0.771958	0.672583	0.870417

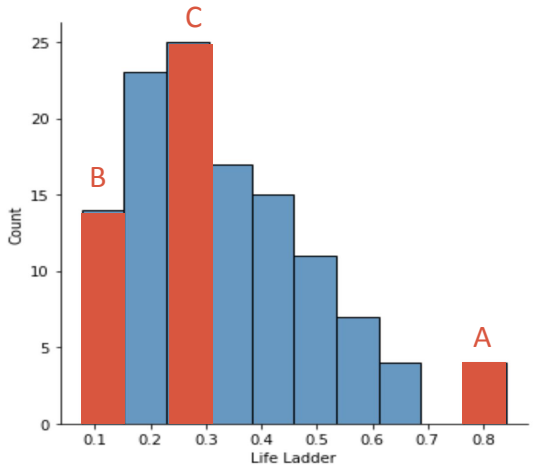
Method: $\sigma \rightarrow \Delta$

- We used *standard deviation of Life Ladder* (σcLL) to model change in Life Ladder;
- used *standard deviation of each parameter* (σc_params) to model change in each parameter;

Histogram: σcLL

We categorized countries into 3 groups

- Group A: countries with largest σcLL (group A)
- Group B: countries with smallest σcLL (group B)
- Group C: countries sharing the most common range of σcLL




```
group_a
```

	Country name	Life Ladder
45	Guinea	0.779655
65	Liberia	0.789411
2	Angola	0.815286
13	Benin	0.840790

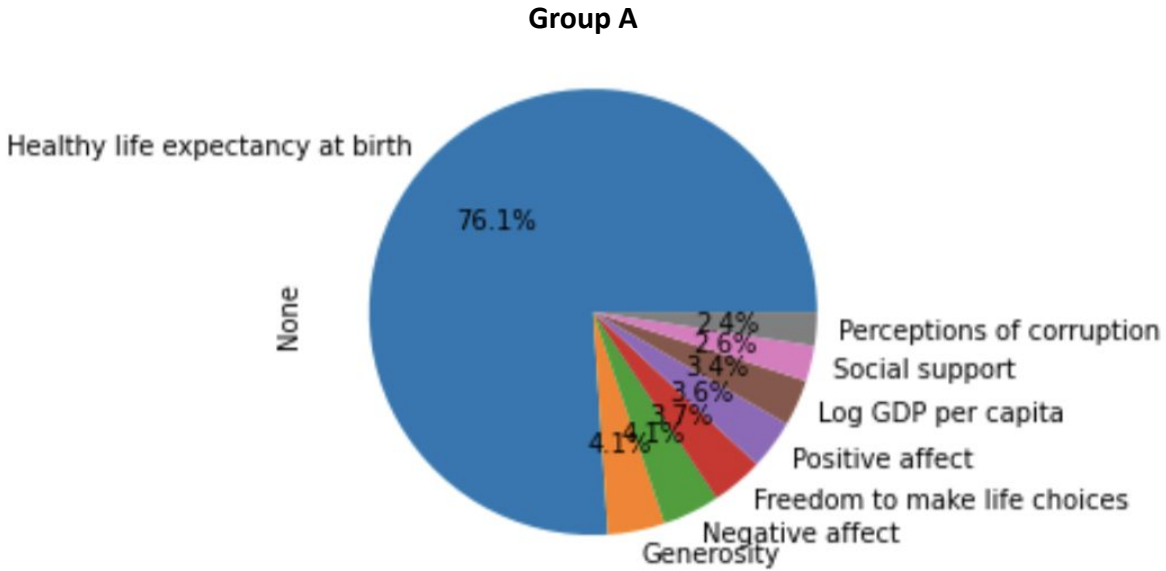
```
group_b.head()
```

	Country name	Life Ladder
85	New Zealand	0.076092
5	Australia	0.088326
90	Norway	0.090442
84	Netherlands	0.099135
11	Belgium	0.105270

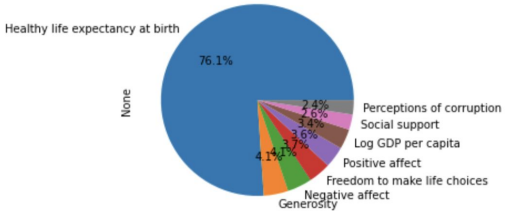
```
group_c.head()
```

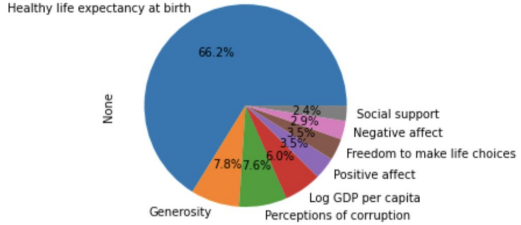
	Country name	Life Ladder
38	France	0.229625
93	Paraguay	0.229815
1	Algeria	0.232318
80	Mozambique	0.247990
4	Armenia	0.248054

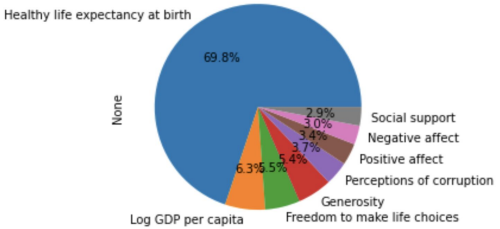
Result: close look



Result







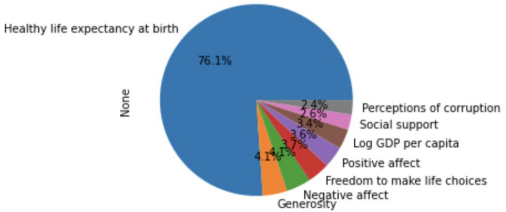
Impact of each parameter in each group
(from the most significant to the least)

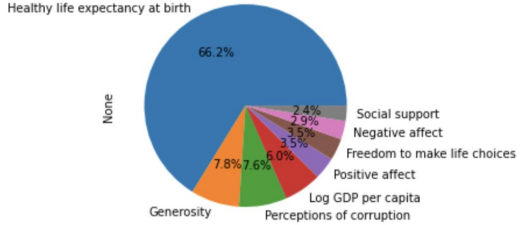
Group A
```
Healthy life expectancy at birth
Generosity
Negative affect
Freedom to make life choices
Positive affect
Log GDP per capita
Social support
Perceptions of corruption
```

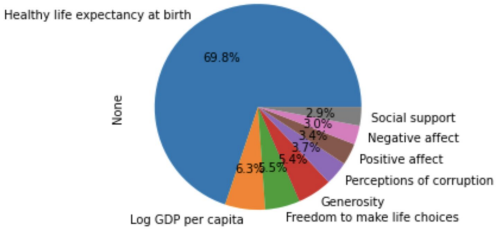
Group B
```
Healthy life expectancy at birth
Generosity
Perceptions of corruption
Log GDP per capita
Positive affect
Freedom to make life choices
Negative affect
Social support
```

Group C
```
Healthy life expectancy at birth
Log GDP per capita
Freedom to make life choices
Generosity
Perceptions of corruption
Positive affect
Negative affect
Social support
```

Result







Impact of each parameter in each group
(from the most significant to the least)

Group A
- Healthy life expectancy at birth
- Generosity
- Negative affect
- Freedom to make life choices
- Positive affect

top 5

- Log GDP per capita
- Social support
- Perceptions of corruption

Group B
- Healthy life expectancy at birth
- Generosity
- Perceptions of corruption
- Log GDP per capita
- Positive affect

top 5

- Freedom to make life choices
- Negative affect
- Social support

Group C
- Healthy life expectancy at birth
- Log GDP per capita
- Freedom to make life choices
- Generosity
- Perceptions of corruption

top 5

- Positive affect
- Negative affect
- Social support

Part 3 Limitation and Future Work

- In the conclusion, we determined *Healthy life expectancy at birth*, *generosity, freedom to make life choices, perceptions of corruption* are the most significant parameters affecting Life Ladder, while *social support* is the least significant parameter. Further literature review is needed to determine the cause.
- In addition, the number of countries in Group C (representing countries sharing the most common range of σcLL) is determined by the bin size of the histogram, so changing the bin size will cause slight changes in which countries are included in this group and in the relative weight of each parameter.