# Happiness Before And After COVID-19 : Poster

Course Title: DH 100 Theory and Methods
Instructor: Dr, Adam Anderson
Student: Alicia Zhou Zhu
Date: 06/18/2021

Happiness is the state of the mind

Happiness Tokyo Japan Black And White Street Photography


We chose happiness

## World Happiness Report (2005 - 2020)

| | Country name | year | Ladder score | Log GDP per capita | Social support | Healthy life expectancy | Freedom to make life choices | Generosity | Perceptions of corruption | Regional indicator |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Afghanistan | 2008 | 3.724 | 7.370 | 0.451 | 50.80 | 0.718 | 0.168 | 0.882 | South Asia |
| 1 | Afghanistan | 2009 | 4.402 | 7.540 | 0.552 | 51.20 | 0.679 | 0.190 | 0.850 | South Asia |
| 2 | Afghanistan | 2010 | 4.758 | 7.647 | 0.539 | 51.60 | 0.600 | 0.121 | 0.707 | South Asia |
| 3 | Afghanistan | 2011 | 3.832 | 7.620 | 0.521 | 51.92 | 0.496 | 0.162 | 0.731 | South Asia |
| 4 | Afghanistan | 2012 | 3.783 | 7.705 | 0.521 | 52.24 | 0.531 | 0.236 | 0.776 | South Asia |

## Descriptions

The World Happiness Report relies on data collected from the Gallup World Poll surveys, an annual survey that started in 2005 and covers supposedly 99 per cent of the world's population. In this project, I will be working with the datasets spanning from 2005 to 2021. 2021 is the most recent one with the happenings of COVID-19, while the previous years, 2005 to 2020 are before.

Our main outcome is dependent on the comparison of the main life evaluation question asked in the poll, the so-called *Cantril ladder* (in the dataset it is represented as Ladder score). The basis of the Cantril ladder basically tells respondents to imagine themselves on a ladder with steps numbered from zero (the bottom) to ten (the top) where zero is the worst and ten being the best possible life. As one can tell, Cantril ladder is an evaluative measure of subjective well-being.

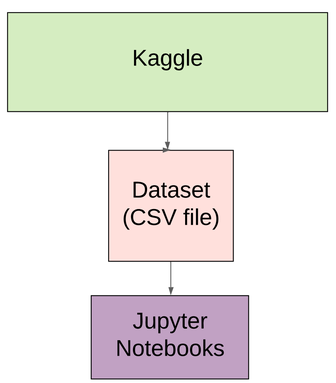
In addition to the Ladder score, there are other factors to explain life evaluations. The columns demonstrate happiness scores through six factors: economic production (log GDP per capita), social support, life expectancy, freedom, absence of corruption, and generosity. They have been deemed as contributing variables to the impact of happiness and of importance within research literature, as it explains national-level differences in life evaluations.

Economic production column is self-explanatory, it is the log GDP per capita. Social support is based on the national average response (Yes = 1, No = 0) to the question: "If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?". Life expectancy is constructed based on data from the World Health Organization (WHO) and World Development Indicators (WDI). Freedom is determined by the freedom to make life choices. It takes the national average of the binary responses to the question: "Are you satisfied or dissatisfied with your freedom to choose what you do with your life?". Generosity is the residual of regressing the national average responses to the question "Have you donated money to a charity in the past month?" on GDP per capita. Absence of corruption are the average of binary answers to two questions: "Is corruption widespread throughout the government or not?" and "Is corruption widespread within businesses or not?" Where data for government corruption are missing, the perception of business corruption is used as the overall corruption-perception measure.

## Methods & Tools

Kaggle was used as the tool to obtain the dataset. Kaggle is an online crowd-sourced platform for data scientists to find and publish data sets, work with others, and to learn and explore predictive analytics problems.

To do the data analysis, I have decided to use Jupyter Notebook as my means of creating graphs through code. Jupyter Notebook is a server-client application that allows editing and running Notebook documents. The Notebook documents contain computer code (like Python) and rich text elements (such as paragraphs, equations, links, etc.). It can be executed locally (requiring no internet access) and it includes the packages I am going to use such as "pandas", "matplotlib", "numpy", and "seaborn". Pandas is used to import the dataset (CSV file) and for data analysis. Matplotlib and Seaborn are being used for graphing the necessary data charts. Numpy is for data manipulation and other miscellaneous usage.




## Research Question:

**How is happiness after the spread of COVID-19 compared to the previous years?**

## Sub-Questions:

1) What does happiness per year look like?
2) What are the happiness levels per region in comparison to before and after COVID-19?
3) What are the differing factors between a happy and a less happycountry?
4) How does the world's subjective well-being fare after the COVID-19 pandemic (in comparison to before)?

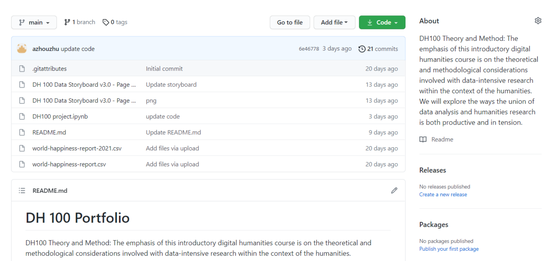
## Potential Results

The hypothetical outcome is that the happiness levels should have decreased worldwide due to the pandemic. The comparison of the tables in other years to the one relevant to the pandemic should display lower happiness levels and other categories that correlate with unhappiness.

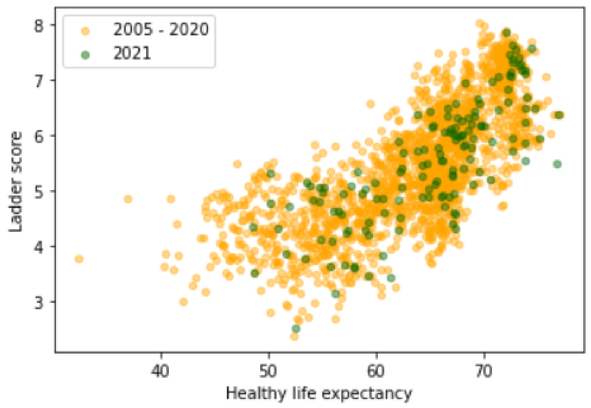
## GitHub Repository

Within my GitHub repository, all the datasets and code used to produce the graphs will be provided. There will also be additional resources included in the repository to better understand my project such as presentation slides and a video walkthrough..

Click here or copy and paste the link below:
https://github.com/azhouzhu/DH100_Summer2021



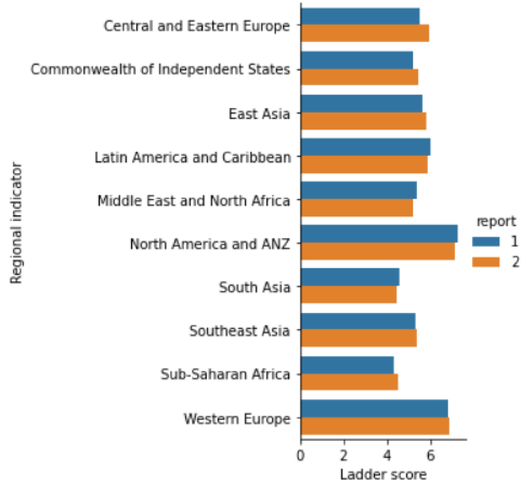
## Correlation between Life Satisfaction and Life Expectancy




Healthy life expectancy is the average life expectancy expected at birth from each country. The Ladder score indicates the average life satisfaction score of each country. Thus, showing what happiness per year is in accordance with healthy life expectancy.

The yellow dots represent the life satisfaction rating of all the countries throughout the years from 2005 to 2020. The green dots are data from 2021 to see a contrast between years. In this plot, there is also a strong positive correlation with a healthy life and life satisfaction. It seems that people tend to be happier when they are healthy and feel as though they have a longer lifespan. Thus, a possible factor that affects happiness is life expectancy.

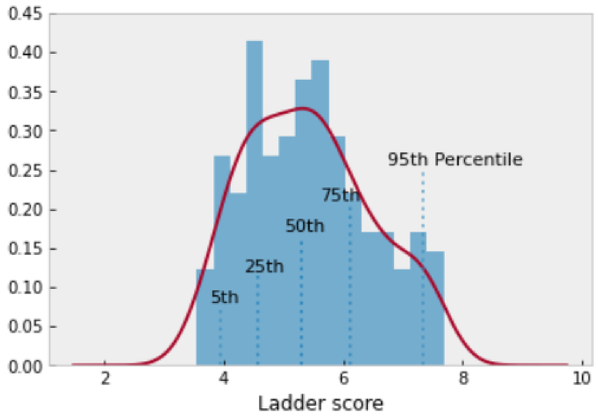
## Life Satisfaction sorted by region between years 2005 - 2021



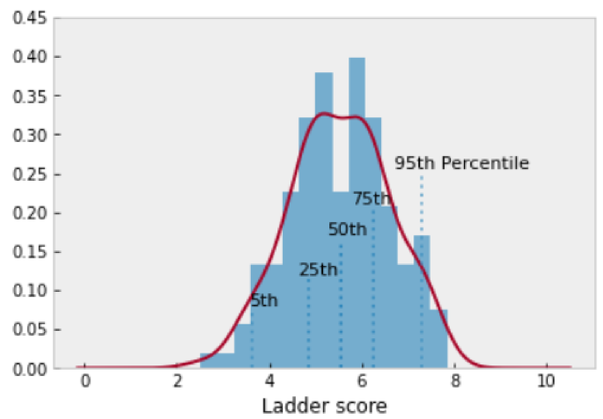

Report 1 represents the ladder scores of years 2005 to 2020. Report 2 represents the ladder scores of the year 2021. The blue bar or report 1 represents life satisfaction before COVID-19 happened. The orange bar or report 2 represents life satisfaction after the pandemic.

The visualizations demonstrate that there was no significant change between the mean ladder score of the years 2005-2020 compared to 2021. If anything, some regions demonstrate that their ladder score in the year of 2021 increased. And, the correlation graph shows higher scores in 2021.

## Life Satisfaction histogram with percentile rankings between years 2005 - 2020




## Life Satisfaction histogram with percentile rankings between years 2021




The red line represents the Kernel Density Estimation (KDE). KDE is a way to estimate the probability density function (PDF) of the random variable that “underlies” the data. KDE is a means of data smoothing. In simple terms, KDE is a method for density estimation. If the red line shows to be at a higher value, then at that point it is more dense. The lower it is, the less dense.

The dashed lines signify the placement of where each percentile ranking is labeled. The percentile rankings demonstrate where the majority of the population opinion of their life satisfaction lies. At the 95th percentile, 95 per cent of the population rated their ladder score to below that point. The rest of the 5 per cent exceeds that point.

The histogram above is right-skewed, most values cluster on the left. This signifies the mean is typically less than the median. In the plot, you can see that the peak lies below the 25th percentile. Simply put, most countries are not satisfied with their life (or not happy).

The histogram above is double-peaked (also known as bimodal). This just means there are two peaks. We can see the first peak is 4.8 - 5 and the second peak is around 5.8 - 6.1. Most countries' ladder score (life satisfaction rating) averages falls into those two bins.

## Jupyter Notebook Code

Correlation between Life Satisfaction and Life Expectancy

```
ax = report1.plot.scatter(x = 'Healthy life expectancy', y = 'Ladder score', color = 'Orange', label = '2005 - 2020')
report2.plot.scatter(x = 'Healthy life expectancy', y = 'Ladder score', color = 'DarkGreen', ax = ax, label = '2021')
```

Life Satisfaction sorted by region between years 2005 - 2021

```
r1_region_ladder = report1_meanByRegion[['Regional indicator', 'Ladder score']]
r2_region_ladder = report2_meanByRegion[['Regional indicator', 'Ladder score']]
r1_r2_meanByRegion = pd.concat([r1_region_ladder, r2_region_ladder])
r1_r2_meanByRegion['report'] = (len(r1_region_ladder)*(1,) + len(r2_region_ladder)*(2,))
r1_r2_meanByRegion.reset_index(inplace = True)
r1_r2_meanByRegion
sns.factorplot(x='Ladder score', y='Regional indicator', hue='report', kind='bar', data=r1_r2_meanByRegion)
```

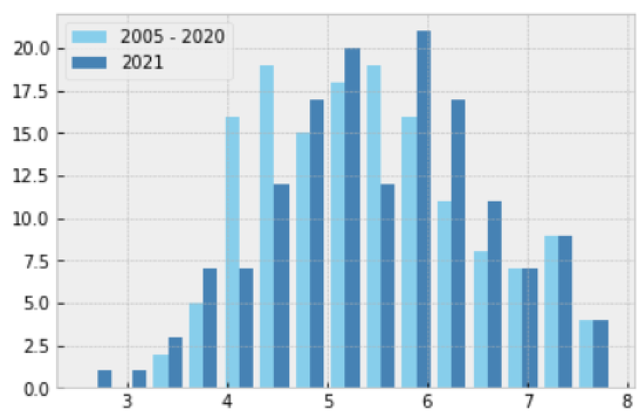
Life Satisfaction histogram comparison (2005 - 2021)

```
ladScore1_arr = ladScore1.to_numpy()
ladScore2_arr = ladScore2.to_numpy()

plt.hist([ladScore1_arr, ladScore2_arr], bins = 15, label = ['2005 - 2021', '2021'], color = ['skyblue', 'steelblue'])
plt.legend(loc='upper left')
```

**For more code, refer to my GitHub. (page 1)**

## Life Satisfaction histogram comparison (2005 - 2021)




The year 2021 (after the COVID-19 spread) shows that there is lower ladder scores (life satisfaction rating). With the first two bins of the histogram having no data comparison to before COVID-19 spread. The first three bins is shown to be the same before and after. However, it seems like there has been a shift for most countries to score higher than pre-COVID-19. With 2021, the most found score feel between 5.8 and 6.0 bin. Whereas 2005 to 2020 peaked between scores 4.3 to 4.5. Showing that the average is to likely have improved after the spread of COVID-19.

## Plot Overview

The Ladder score rank is ranked from highest Ladder score to lowest ladder score. In other words, Ladder score rank 1 would indicate the happiest country in accordance with the *Cantril Ladder*.

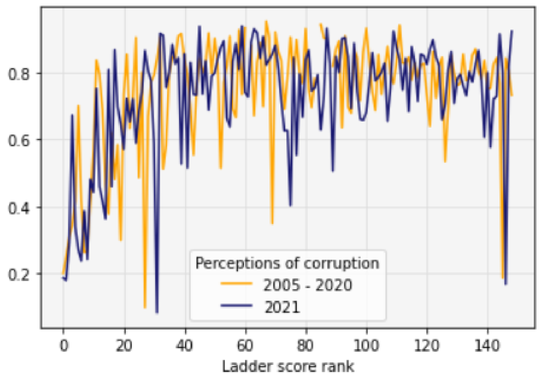
For the years of 2005 to 2020, the data taken for the line plots is the average of every country within those years. The idea is to see 2005 to 2020 as one entity, before the pandemic started.

For the year 2021, it is just the dataset of that year alone. This dataset represents the after effects of  COVID-19.

The line plots being analyzed in this section includes: Generosity, Perception of Corruption, Log GDP per capita, Freedom to make life choices, and Social support. The comparisons are in comparison of the split view we just mentioned, all data before (2005 to 2020) as one line and the after (2021) as the second line.

I have made the charts for the 2021 line plot to be more prominent than the dataset 2005 to 2020 to see the changes.

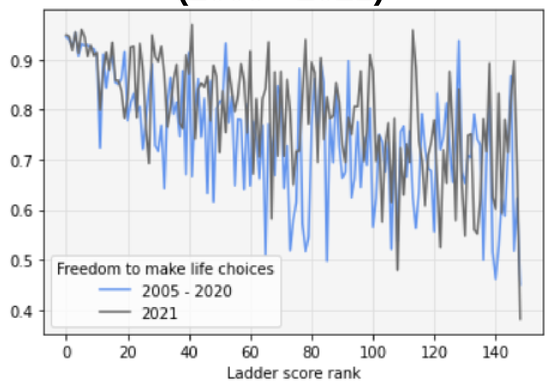
## Perceptions of Corruption comparison (2005 - 2021)




The plot demonstrates the first rank to have a low perception of corruption, however, it is not the lowest. As you progress to lower ranks, the levels of perceptions of corruption grows.

The claim would be that a happier country has lower perception of corruption. However, interestingly enough, there is two significant dips in the plot, one near the top ranks and one very close to the last ranks. This line plot does not prove if perceptions of corruption has an affect on life satisfaction.

## Freedom comparison (2005 - 2021)




The plot shows that there is an association to happiness to high freedom of choice and unhappiness to no low freedom of choice. It is also apparent that there has been an increase of freedom of choice after (2021) in comparison to before (2005-2020)

The higher ranks have more freedom to make life choices. Meaning, they have more options to choose from. While lower ranks must have restrictions in life choices. This could be in relation on how the cultural and political structure of these countries.

## Interpreting the results

From the visualizations, happiness before and after has not changed in a drastic manner whether positive or negative. In some regions, they have become happier while others have lowered slightly. However, overall there is no significant change in happiness after the spread of COVID-19.

The factors to happiness have also not demonstrated major changes in the before and after data. The factors, however, help explain the ladder score average of each country and how they are associated.

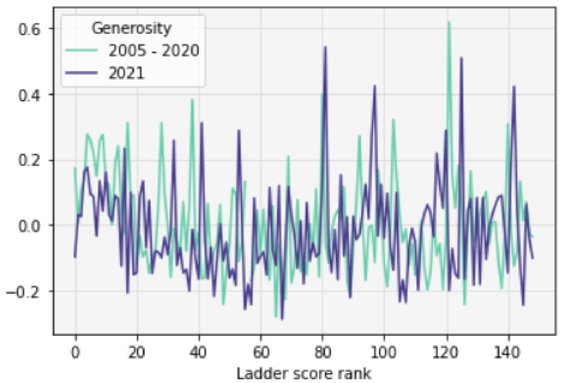
## Conclusions

From my research, worldwide issues such as pandemics are not directly associated  with the happiness levels of the world. Analyzing the life satisfaction rating (ladder score) by country has produced no concrete evidence of happiness being affected by COVID-19. If there are significant changes, it is not within the factors I have chosen to study.

## Next Steps / Future Work

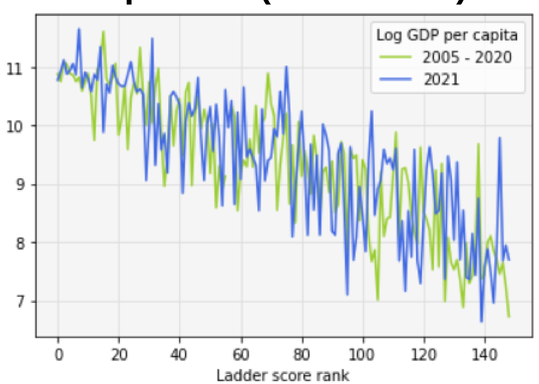
The dataset "World Happiness Report" can potentially be used for more than analyzing life satisfaction. It has the potential to be used to further research how economics of different countries affect life expectancy. Through a common pattern from the research, it could lead to understanding why some countries underperform and determine if it could be a causation to life satisfaction scores from each country.

## Generosity comparison (2005 - 2021)




The plot shows that the most happy country (1st rank) is not the most generous. If anything, the generosity of the most happy country diminished after the spread of COVID-19. On the other hand, countries who scored lower on life satisfaction demonstrates to be more generous. You can see that peak generosity for both before and after are countries with lower life satisfaction.

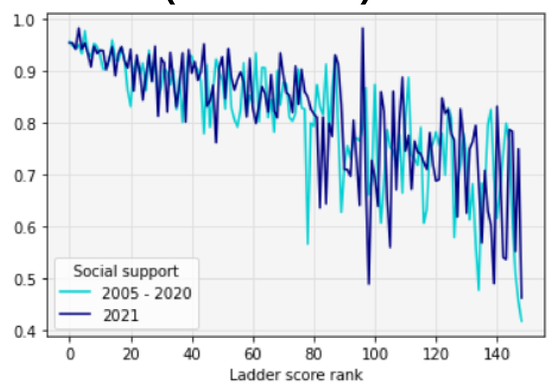
## Log GDP per capita comparison (2005 - 2021)




There is an association to happiness and higher log GDP per capita. As the ranks get lower, you can see that the economic production is decreasing. In comparison to before and after COVID-19, some lower ranking countries log GDP per capita went up.

It seems that there are factors causing the increase in log GDP per capita in the lower ranks.

## Social Support comparison (2005 - 2021)




There is an association to happiness and higher social support. The contrast of the highest and lowest rank makes it apparent that social supprot may contribute to life satisfaction ratings.

Before COVID-19, there is a steady decrease in correlation to the plot as it decreases rank. However, after, the fluctuations within social support increased. This plot shows the highest peak is within the year 2021 .

## Work Cited / References


Dataset:
https://www.kaggle.com/ajaypalsinghlo/world-happiness-report-2021)

Helpful Links:
https://worldhappiness.report/ed/2021/
https://pandas.pydata.org/pandas-docs/stable/user_guide/index.html
https://towardsdatascience.com/take-your-histograms-to-the-next-level-using-matplotlib-5f093ad7b9d3
https://pandas.pydata.org/pandas-docs/version/0.23/generated/pandas.DataFrame.plot.html
https://www.gnyha.org/wp-content/uploads/2020/06/World-Happiness-Report.pdf
https://worldhappiness.report/ed/2020/cities-and-happiness-a-global-ranking-and-analysis/