**SILESIAN UNIVERSITY OF TECHNOLOGY**

Faculty of Automatic Control, Electronics and Computer Science

Une image contenant croquis, dessin, Dessin au trait, clipart

Description générée automatiquement

**Data Science**

**Analyzing Happiness Levels**

*Authors:*

*Mustapha Terkmane*

*Zakarya Louahchi*

*Kendrick Choong*

1. **Abstract**

Understanding the determinants of happiness is crucial for promoting well-being and informing policy decisions. In this study, we investigate the relationship between perceived happiness and various socio-economic factors using a comprehensive dataset from multiple countries. Our analysis draws upon data from the "[Awesome EDA] 2021 Happiness & Population" dataset sourced from Kaggle, which includes key variables such as GDP per capita, social support, healthy life expectancy, and perceptions of corruption.

Utilizing regression analysis techniques, we explore how these socio-economic factors contribute to individuals' perceived happiness levels while controlling for demographic variables such as age, gender, and education. Our analysis considers a wide range of factors, including economic prosperity, social cohesion, healthcare access, and political stability.

Preliminary findings suggest that factors such as GDP per capita, social support, and healthy life expectancy are significant predictors of happiness across different countries. Additionally, we examine the impact of income inequality, access to healthcare, and environmental quality on subjective well-being.

Implications of our findings for public policy and societal well-being are discussed, highlighting the importance of addressing socio-economic disparities and promoting factors that contribute to happiness and quality of life. This study contributes to the growing body of research on happiness and provides insights into the complex interplay between socio-economic factors and subjective well-being.

1. **Introduction**

Happiness is one of the most important values [1] that has always been associated with pleasure, optimism and hope [2]. In fact, happiness is an inner state that is derived from human judgment and experience [3,4]. Eysenck believed that happiness is a collection of fun, a combination of maximum positive effect and the least negative emotion without pain. Happiness and vitality are the most essential human innate desires and psychological needs. So happiness and its determinants, has been considered as an important subject [5]. It is important to understand that happiness creates energy, enthusiasm, vitality, movement and dynamism. It also protects humans against stress and problems [6] along with physical and mental health [7]. This means that often happy people are less sick [8] live longer, have less stress and depression than unhappy ones [2]. In fact, happy people have stable marriages, higher incomes and have more creative ideas in comparison to others [9]. Because of the importance of happiness in human life, psychologists, sociologists, economists, scholars and researchers have investigated the factors affecting human happiness. Subsequently, various theories have been discussed in this field. So, the present study aimed to evaluate the individual, socioeconomic and social factors as main contributing factors on happiness.

1. **Objective**

This study examines the influences on subjective well-being in congruence with the emerging field of happiness economics. The field developed when economists observed that increases in per-capita income had minimal effects on the average happiness levels of a country. Applying this phenomenon to micro economic theory challenges and expands the neoclassical premise for utility curves which neglects other influences like social comparison and non-pecuniary factors [10]. By examining the relationship between variables such as GDP per capita, social support, and healthy life expectancy, we aim to identify key drivers of happiness and assess their impact across diverse populations.

1. **Literature Review**

Previous research on happiness has yielded diverse perspectives and findings, shedding light on the complex interplay between subjective well-being and socio-economic factors. Oswald (2006) highlighted the misconception of equating economic growth with happiness, noting that industrial countries, despite their economic prosperity, have not experienced a corresponding increase in happiness levels. Maddison (1991) echoed this sentiment, emphasizing the disconnect between GDP growth and happiness improvement.

Easterlin's seminal work (1995) challenged the notion that increased income directly translates to greater happiness, citing evidence from longitudinal studies across European countries. Similarly, Easterlin et al. (2011) underscored the long-term contradiction between happiness and income growth, suggesting that the pursuit of economic prosperity alone does not guarantee subjective well-being.

Eckersley (2009) emphasized the importance of societal welfare in fostering individual happiness, advocating for a holistic approach to happiness measurement. Guo and Hu (2011) identified an inverse relationship between happiness and economic variables such as unemployment and inflation, corroborating previous research findings.

Baumeister et al. (2013) explored the distinction between happiness as a life state and happiness as meaning, underscoring the role of giving in fostering happiness. Furthermore, recent studies have delved into the influence of technology and consumption behavior on happiness, suggesting nuanced relationships between material wealth and subjective well-being (Rossi Kamal & Choong Hong, 2015; Zining Peng & Maolin Ye, 2015).

Shoval & Morag (2017) investigated the relationship between intergroup interactions and happiness, underscoring the relevance of social dynamics in subjective well-being assessment.

1. **Sources of Happiness**

*5.1 Social Factors*

Social factors deeply affect people’s well being and happiness. The factors that can be considered as sources of happiness and show the quality of life are: health, education, unemployment and marital status.

* **Health**

For many decades, many analyses presented the interaction between health and happiness. Good health will make people feel more cheerful and cause greater contentment in their daily life. However, having bad physical or mental health leads to more restrictions in people’s daily activities and causes a pessimistic view of life.

In the United States and Europe, governments believe in the strong link between individual happiness and good quality of health care services received. People’s satisfaction is playing an important role in improving the quality of health care reforms.

* **Education**

Happiness and education are strongly connected. A good education contributes significantly to personal and collective happiness. It is generally admitted that education improves people’s lives in many aspects. Education enhances people’s lives as higher educational attainment is linked to better career paths and is also believed to enhance outcomes in other life domains, such as health and relationships. Education is considered as the most important activity in modern man life. In many countries it is one of the biggest items of public spending.

* **Unemployment**

Happiness and work was the subject of various empirical economic studies. Jobs are very important for sustaining individual’s living, family and health which are the main elements shaping people’s happiness Accordingly, many studies stated a negative effect of unemployment on happiness

* **Marital Status**

In a study of multiple regression analysis conducted by Steven Stack and J. Ross Eshleman (1998), they found that marriage increases happiness equally among men and women and this positive impact is due to three facts7. First, marriage provides financial satisfaction as married people combine two incomes and may enjoy a higher standard of living. Second, it leads to the improvement of health through the support of encouraging partners to follow a medical treatment in case of illness, quitting bad behaviors such as drinking and smoking and helping spouses to follow a healthy diet. Finally, marriage provides greater emotional support which refers to being esteemed, cared about and valued as a person.

*5.2 Economical Factors*

* **GDP**

Economic growth is one of the main objectives of any economy. High economic growth rates and low economic growth contribute to economic development.

* **Infrastructures**

Infrastructure is the backbone of many economic, social and political life activities in societies. Countries that do not care about infrastructure will be affected by the decline of its growth and development.

*5.3 Political Factors*

* **Freedom**

Freedom, often associated with ideals of equality and brotherhood, plays a significant role in shaping individual happiness and societal well-being. It encompasses the ability to make choices without undue constraints in economic, political, and personal spheres. Previous research suggests a positive correlation between freedom and happiness, particularly in affluent nations. However, there are differing perspectives on the impact of freedom on happiness.

Some conservative views caution against the potential negative consequences of excessive freedom, arguing that individuals may lack the insight to make optimal decisions, thereby threatening essential societal institutions. Conversely, other schools of thought posit that economic freedom, in particular, enhances human well-being, while the relationship with political freedom varies based on contextual factors such as societal maturity.

* **Corruption**

Corruption is generally considered as an important factor which defines the quality of countries’ governance.

Living in a country reigned by corruption and discriminatory institutions can lead to unhappiness. This is illustrated by a report released by the United Nations Human Development program about the case of Bosnia and Herzegovina in 2002. The reported survey showed that 70% of people in these countries strongly believe that their local authorities and international aid organizations are corrupt.

1. **Data and Methodology**

*6.1 Data Sources*

To create a comprehensive data source for the analysis of happiness across multiple demographic factors, years, and countries, we had to combine multiple data sources together. A majority of the happiness scores for each year came from Kaggle datasets created and uploaded by Ajaypal Singh [11-13].

The Kaggle datasets also had some demographic data, but a majority of the data for demographics were taken from the US Census Bureau's International Database [15]. Things like population, GDP, population by sex, age demographics, etc. were all present in the census data.

These two overarching sources of information had to then be combined based on a key. So the country name and the year were combined to create a unique key for each record to be used in the visualization. Originally the census data only had 2 digit country codes so another data set from Data Hub’s website was used to help decode the country code to join the data sources [14].

*6.2 Cleaning*

The following steps outline how we went about cleaning, merging, and shaping the data to be used to answer our research question.

1. Parse all data files from text/csv files and input the data into data frames in Python.
   1. Most clean csv files were picked up fined using Pandas libraries with other text files like the census data needing a slight change in the function to read the pipe delimiter.
2. Clean each respective data frame/data source.
   1. Mean imputation to get rid of missing values.
   2. Removing big outliers from the data that could skew analysis.
   3. Limiting the years from each data source for relevancy.
   4. Renaming, adding, or dropping columns necessary/unnecessary columns.
3. Append and code the census data with country names.
   1. The census data lacked the full country name and the WHR datasets all contained country name instead of a country code. So we chose to convert the country code to be country name so the data could be joined together.
4. Merge the data sets on country name and year.
   1. We had to create a key containing the country code and the year so both the census data and the WHR data could be combined together for analysis.
   2. The country code data and the census data was joined on a left join using census data as the left data set to capture all records in the census data.
   3. Then an inner join was performed with the newly created key for both the census data and the WHR data so only matching records would be in the final data set also removing records with possible incomplete data for happiness or that didn’t exist in either data set.

*6.3 Variables*

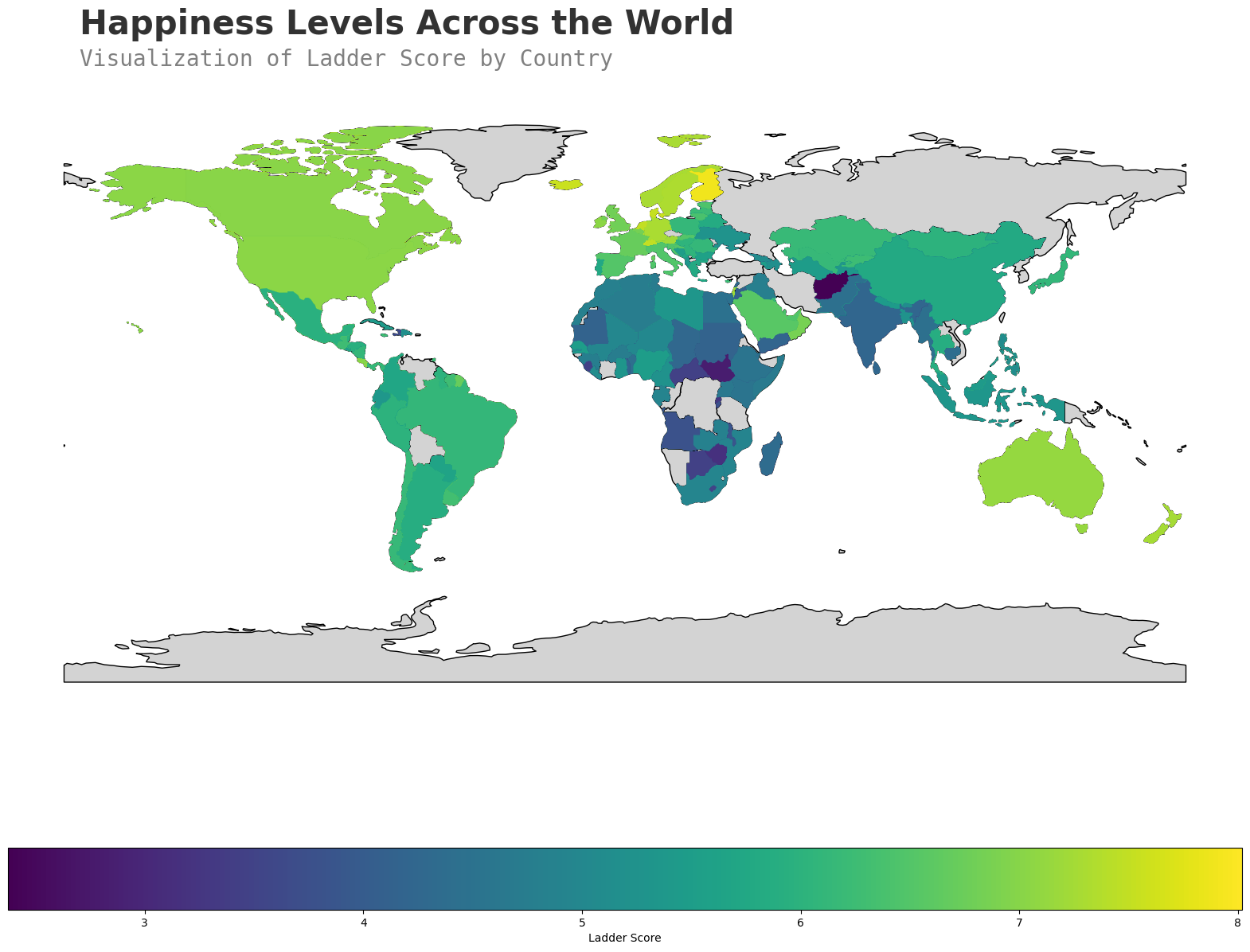
| **Variable Name** | **Description** |
| --- | --- |
| Happiness Score | A nationally representative average of individuals rating their own lives on a 0 to 10 scale with 0 being the absolute worst life they could have and 10 being the absolute best life they could have. |
| Population | The midyear population counts for a specific country. |
| Healthy Life Expectancy | Healthy life expectancies at birth are based on the data extracted from the World Health Organization’s (WHO) Global Health Observatory data repository (Last updated: 2020-12-04). |
| GDP Per Capita | The statistics of GDP per capita (variable name gdp) in purchasing power parity  (PPP) at constant 2017 international dollar prices are from World Development  Indicators. |
| Social Support | The national average of the binary responses 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?” |

*6.4 Descriptive Statistics*

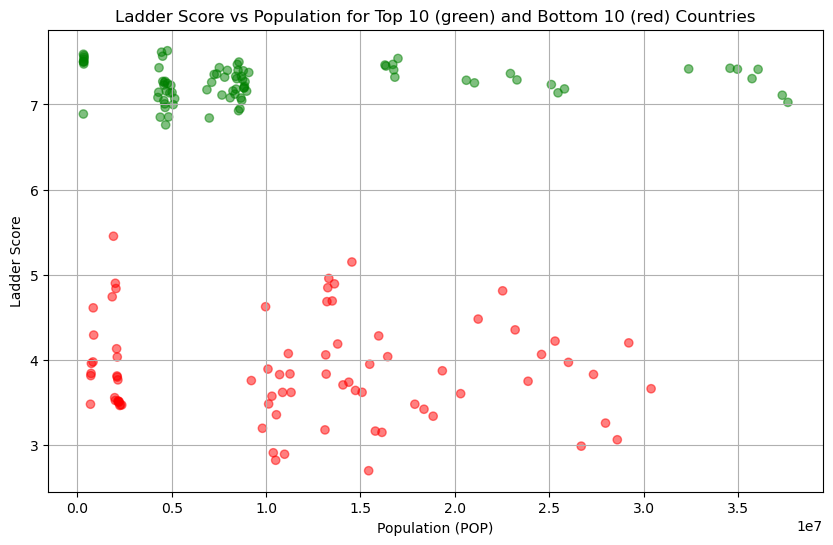
In the dataset, there are a total of 122 variables, 1837 records, 145 countries, and a date range of 2005-2021. The table below is the descriptive statistics of some important variables.

|  | Ladder Score | Logged GDP per Capita | Social Support | Healthy Life Expectancy |
| --- | --- | --- | --- | --- |
| Mean | 5.51 | 9.40 | 0.82 | 63.67 |
| Std | 1.14 | 1.12 | 0.12 | 7.56 |
| Min | 2.38 | 6.64 | 0.29 | 32.30 |
| Max | 8.02 | 11.65 | 0.99 | 77.1 |
| 25% | 4.64 | 8.46 | 0.75 | 58.97 |
| 50% | 5.44 | 9.54 | 0.85 | 65.5 |
| 75% | 6.38 | 10.41 | 0.91 | 69.0 |

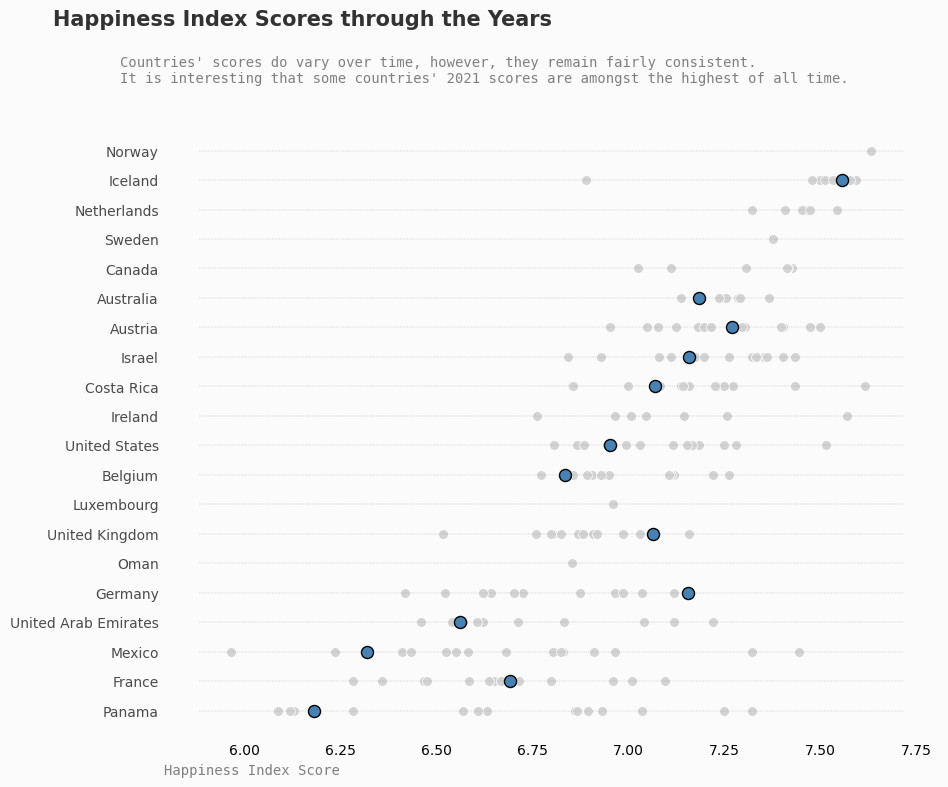
*6.5 Visual Analysis*



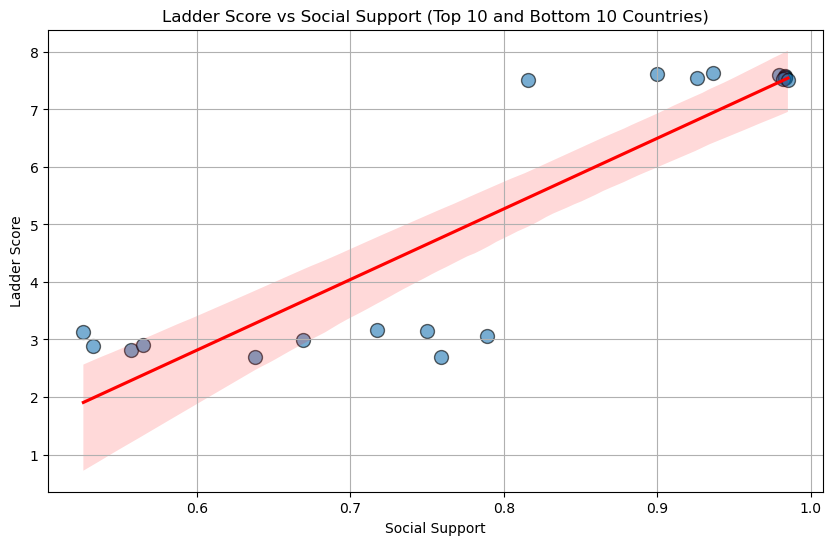
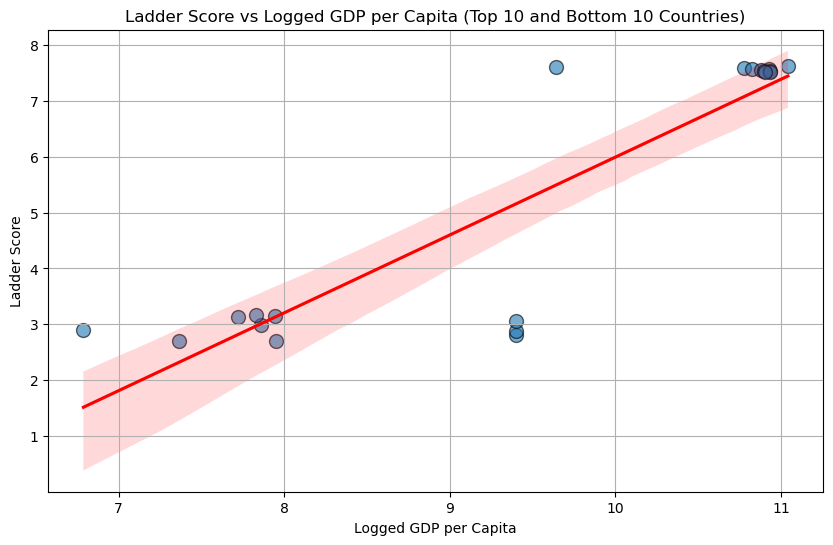
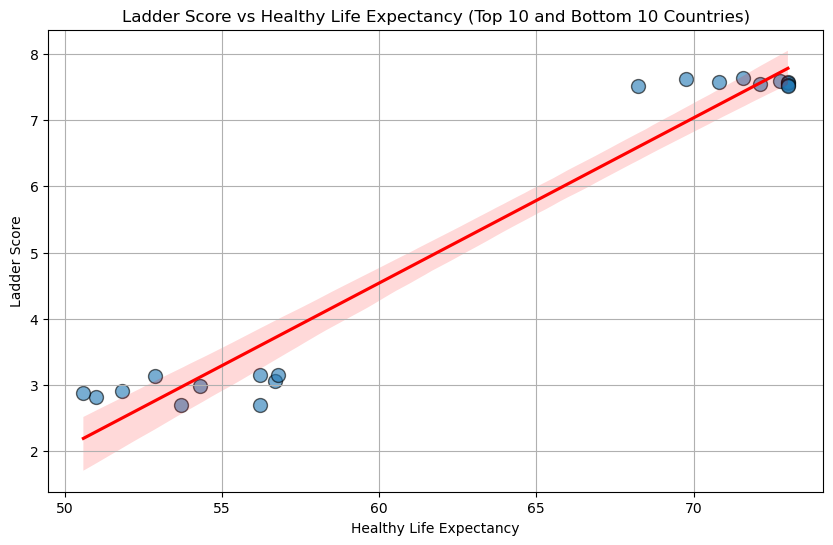
Looking at the figure above, graphing the happiness by country in a geographical way helps identify which regions are happiest vs. not. We see that the happiest regions are in North America and Europe with some of the unhappiest countries residing in Africa. Understanding what these regions share in common could help predict how happy their citizens are which could be reasons ranging from weather all the way to culture.



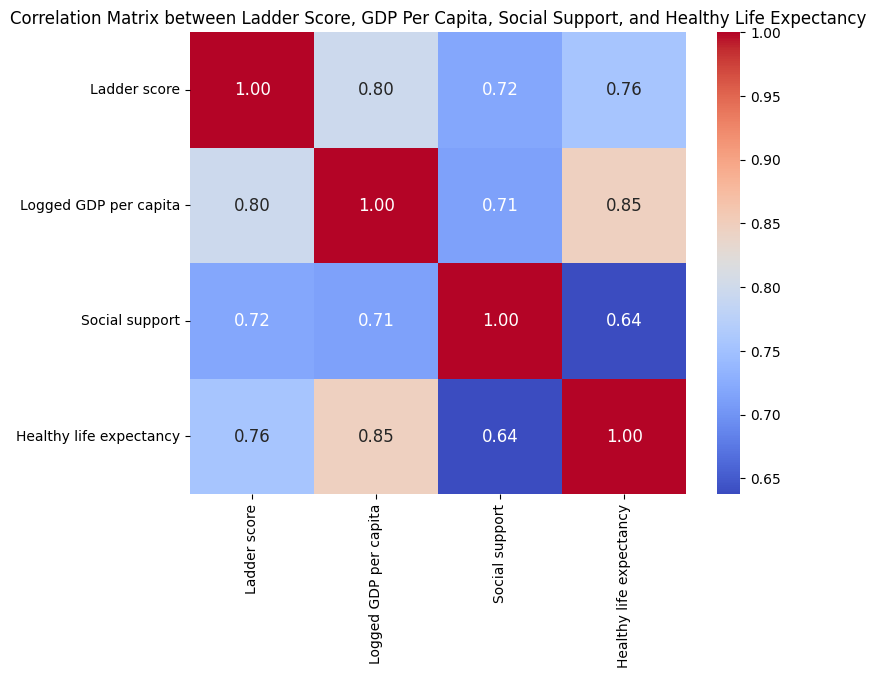
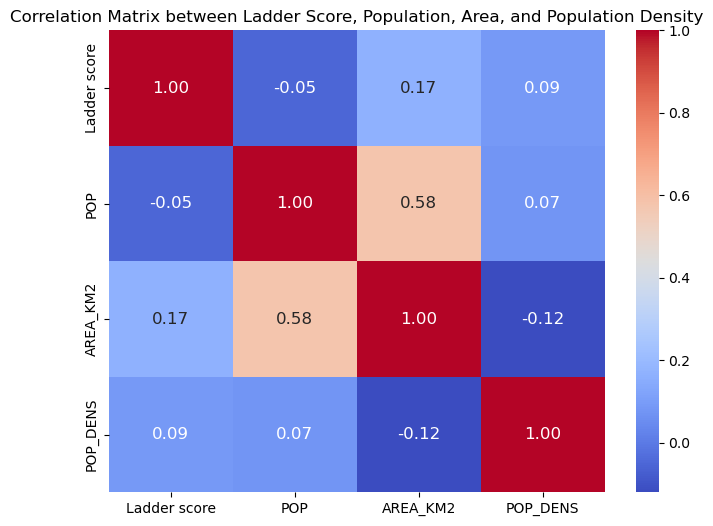
Examining population and happiness, there isn’t an apparent correlation in the data. We see happy countries that have a lot of citizens and happy countries with not a lot of citizens. Having more or less people could lead to issues of overcrowding, but we see many countries that are equally happy with more and less people. An additional question that could be added to the surveys that go out for the WHR could include perceptions of overcrowdedness. The more people feel the competition for resources the more stressed out they could be in their day to day lives.



Performing an analysis on the happiness through the years for countries, not every country is consistently happy year over year like Iceland. Some countries have had much happier years and much sadder years in the dataset. For some countries, 2021 was a relatively good year and for some it was on the lower end of their band despite it being the start of the COVID-19 pandemic.



Graphing happiness against major variables such as GDP, Healthy Life Expectancy, and Social Support, we se that the more quality of life aspects a country has, the happier it is. GDP per capita and Healthy Life Expectancy is tightly grouped compared to Social Support. So even if a country has higher perceived social support, it might still not be overall happier. This could indicate some cultural differences having a big influence on happiness as some cultures promote more social support between family members than others.



Taking another look at population, there doesn’t seem to be any strong correlation between the ladder score and the number of people in a country. There is more positive correlation between other aspects of a country in relation to the ladder score such as the GDP, the social support, etc. Could there be even more elements to happiness rooted in the actual physical space of a country? Would this be more relevant in future happiness studies as the world continues to get smaller and smaller with our ever growing population?

*6.6 Statistical Analysis*

Also looking at some statistical analysis on the variables, a multiple linear regression analysis was performed on happiness, GDP, social support, and healthy life expectancy. Our goal with this analysis was to quantify the relationship between these variables, determine if they were important determiners of happiness, and to test how well the model itself does at representing and predicting data.

Some of the key analytical takeaways were that:

* Higher GDP per capita is associated with higher happiness.
* A strong social support greatly enhances happiness.
* Longer healthy life expectancy boosts happiness.
* All variables had P-values < 0.0001, indicating high statistical significance.

| **Variable** | **Coeff** | **P-val** |
| --- | --- | --- |
| Logged GDP per capita | 0.3693 | <0.0001 |
| Social Support | 2.7211 | <0.0001 |
| Healthy Life Expectancy | 0.0383 | <0.0001 |

1. **Conclusion**

After analyzing the data, there are 3 main conclusions that can be derived from visualization of happiness data against other demographic data.

1. Some quality of life aspects in a country seem to correlate positively with happiness such as life expectancy, GDP per capita, and social support.
2. Other demographics like population or anything related to the land mass of a country has little correlation with the reported happiness.
3. Some countries have consistently been happy whereas other countries fluctuate in being happier or sadder over the years.

Future areas for research on this topic would be to look closer at some cultural differences in countries and how that might impact happiness. Another area for research could also be with a high longevity study on population and happiness as the world becomes more and more crowded with increased birth rates and longer life spans.

Happiness is still such an abstract concept that is hard to define without the subject opinions of each individual human being. As we gather more and more data to perhaps quantity the behaviors and subjective opinions into objective truths, one can only hope that we use that data to better the living situation of each and every place.

**References:**

*Literature Review*

[1] Kim YS. Perception of Happiness among health-related university students:q methodological approach. Indian Journal of Science and Technology. 2015;8:161–68. [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Indian+Journal+of+Science+and+Technology&title=Perception+of+Happiness+among+health-related+university+students:q+methodological+approach&author=YS+Kim&volume=8&publication_year=2015&pages=161-68&)]

[2] Amir Kafi M, Zare B. The impact of social capital on happiness (case study:Kerman city) The Socio Cultural Strategy Journal. 2012;2(5):41–77. [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=The+Socio+Cultural+Strategy+Journal&title=The+impact+of+social+capital+on+happiness+(case+study:Kerman+city)&author=Kafi+M+Amir&author=B+Zare&volume=2&issue=5&publication_year=2012&pages=41-77&)]

[3] Haghighi J, Shokrkon H, Shehni Yeilagh M, Neisi AK. The relation of five personality factors and feeling of happiness undergraduate at the University of Shahid Chamran. Journal of Education and Psychology of Shahid Chamran University. 2006;13(3):163–88. [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Journal+of+Education+and+Psychology+of+Shahid+Chamran+University&title=The+relation+of+five+personality+factors+and+feeling+of+happiness+undergraduate+at+the+University+of+Shahid+Chamran&author=J+Haghighi&author=H+Shokrkon&author=Yeilagh+M+Shehni&author=AK+Neisi&volume=13&issue=3&publication_year=2006&pages=163-88&)]

[4] Uchida Y, Oishi S. The happiness of individuals and the collective. Japanese Psychological Research. 2016;58(1):125–41. [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Japanese+Psychological+Research&title=The+happiness+of+individuals+and+the+collective&author=Y+Uchida&author=S+Oishi&volume=58&issue=1&publication_year=2016&pages=125-41&)]

[5] Mehrdadi A, Sadeghian S, Direkvand-Moghadam A, Hashemian A. Factors affecting happiness:a cross-sectional study in the Iranian youth. J Clin Diagn Res. 2016;10(5):VC01–VC03. [PMC free article] [PubMed] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=J+Clin+Diagn+Res&title=Factors+affecting+happiness:a+cross-sectional+study+in+the+Iranian+youth&author=A+Mehrdadi&author=S+Sadeghian&author=A+Direkvand-Moghadam&author=A+Hashemian&volume=10&issue=5&publication_year=2016&pages=VC01-VC03&)]

[6] Niazazari K, Safari SH. A study of the Relationship between personality characteristics and happiness among students of Azad University, Azadshahr Branch. Quarterly Journal of Educational Psychology. 2011;2(3):65–75. [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Quarterly+Journal+of+Educational+Psychology&title=A+study+of+the+Relationship+between+personality+characteristics+and+happiness+among+students+of+Azad+University,+Azadshahr+Branch&author=K+Niazazari&author=SH+Safari&volume=2&issue=3&publication_year=2011&pages=65-75&)]

[7] Siadatian SH, Ghamarani A, Yaghobian F. The meta-analysis of the effectiveness of psychological interventions on the Iranians’feeling of happiness. Developmental Pschology:Iranian Psychologists. 2013;10(37):61–69. [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Developmental+Pschology:Iranian+Psychologists&title=The+meta-analysis+of+the+effectiveness+of+psychological+interventions+on+the+Iranians%E2%80%99feeling+of+happiness&author=SH+Siadatian&author=A+Ghamarani&author=F+Yaghobian&volume=10&issue=37&publication_year=2013&pages=61-69&)]

[8] Adullah YA, Zulkifli FL. Concepts and Theories of Happiness of Population in Urban Neighbourhoods. 6th Asia-Pacific International Conference on Environment-Behaviour Studies, Barcelona School of Architecture (ETSAB), Barcelona 2015, Spain, 31 Aug.- 05 Sep [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=6th+Asia-Pacific+International+Conference+on+Environment-Behaviour+Studies,+Barcelona+School+of+Architecture+(ETSAB),+Barcelona+2015,+Spain,+31+Aug.-+05+Sep&title=Concepts+and+Theories+of+Happiness+of+Population+in+Urban+Neighbourhoods&author=YA+Adullah&author=FL+Zulkifli&)]

[9] Lyubomirsky S, Layous L. How Do Simple Positive Activities Increase Well-Being? Current Directions in Psychological Science. 2013;22(1):57–62. [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Current+Directions+in+Psychological+Science&title=How+Do+Simple+Positive+Activities+Increase+Well-Being?&author=S+Lyubomirsky&author=L+Layous&volume=22&issue=1&publication_year=2013&pages=57-62&)]

[10] Elizabeth K. Hancock, Assessing Happiness: How Economic Factors Measure Up, Illinois Wesleyan University, 2013. [[Google Scholar](https://core.ac.uk/download/pdf/59230012.pdf)]

*Data Sources*

[11] Singh A. (2021). *World Happiness Report* (2021) [Data Set]. Kaggle. [[Link](https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2021)]

[12] Singh A. (2022). *World Happiness Report* (2022) [Data Set]. Kaggle. [[Link](https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2022)]

[13] Singh A. (2023). *World Happiness Report* (2023) [Data Set]. Kaggle. [[Link](https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2023)]

[14] Data Hub. List of all countries with their 2 digit codes (ISO 3166-1) [Data Set]. [[Link](https://datahub.io/core/country-list)]

[15] United States Census Bureau. International Database (Time Series: various year - 2100) [Data Set]. [[Link](https://www.census.gov/data/developers/data-sets/international-database.html)]