Sustainable happiness an evaluation between pre and post pandemic

Radhu Palliyana

Abstract—In this paper, we study the effects of sustainable happiness on countries pre and post pandemic through visualisation. Countries with high and low ladder score are considered for the study. World happiness index dataset from Kaggle and Happy planet index data set from 2015 to 2020 are merged for the analysis. At first, we analyse how countries with high and low ladder score perform against Happy planet index and ladder score. Then we shift the research to identify the major factors affecting the sustainable happiness of countries. Then the effect of ecological footprint reduction against these factors are analysed. Both pre and post pandemic effects are analysed for this. Lastly, through clustering we analyse post pandemic, if countries are happier if they have any progress in ecological footprint reduction.Through the study we conclude that the ecological footprints reduction improves the sustainable happiness of countries.

1 Problem Statement

The research around understanding the major factors that leads to a happy life have been carried out since ages. World happiness index is calculated using the data from Gallup world poll survey. It considers the following main factors for life evaluation across nations: levels of GDP, life expectancy, generosity, social support, freedom, and corruption. It is evaluated against Dystopia, a hypothetical country that has lowest values for all these factors.[1] For most of the research the major factors are considered to be GDP and wellbeing. But recent studies highlight the importance of sustainable happiness. Happy planet Index on the other hand measures sustainable wellbeing for all. It considers the following major factors for its measurement: Wellbeing, Life expectancy and Ecological Footprint.[2] Through this analysis I aim to answer the following research questions:

1. How does HPI and ladder score differ for worlds happiest and least happy countries.
2. What are the major factors that affected the happiness of countries with high and low ladder score post pandemic?
3. How does the major factor affecting the happiness, perform against ecological footprints pre and post pandemic?
4. How ecological footprints affected ladder score post covid. Does implementation of lockdown have any effect on ecological footprints.

For the analysis on how ecological factors affect the happiness of countries both world happiness report from 2015-2020 and Happy planet index report from 2015- 2020 will be merged. This will help to analyse the sustainable happiness of a country. Thereby answers the research questions too.

2 State of the Art

According to Dr.Catherine O'Brien, "Sustainable happiness is happiness that contributes to individual, community, and global well- being without the exploitation of other people, the environment or future generations"[6]. Over the last decade, there was a lot of initiatives happening worldwide to promote sustainable happiness. As per the report published by Danish ministry it says "It's no longer possible to imagine a future where the pursuit of happiness is not somehow connected to sustainability" [5]. So, it's very important to frame policy and make decisions considering environmental conservation as an important parameter to have a happy future.

The research papers selected varies in their approach in analysing happiness. In which two papers, [7] and [8] check for happiness between countries using clustering analysis and using 2D stacked bar charts. Both these papers use visualization as guide to evaluate the happiness score between countries. The first paper uses happiness score data from Kaggle between 2015 to 2017 whereas the later one concentrates on life evaluation for comparisons of the quality of life for countries between 2005 to 2014.The third research paper [9] analyses countries according to their sustainability state. The study uses HPI 2016 data published by New Economic Foundation. It uses bar charts for visual analysis of data. In this research paper they propose a parameter -free method using Single Value Decomposition.

Through this analysis I will extend the study around happiness by considering ecological footprint as a major factor. For this I will combine world happiness report and Happy planet report to check the important factors which affect sustainable happiness of countries. Also, I will be analysing pre and post covid effects of these factors on ecological footprints. The research will be mainly focus on countries with high and low ladder score. This will help to evaluate the factors affecting happiness of countries in these two categories. Like the study conducted in the research paper [8], I will be using bar charts for visual analysis. From the research paper [7], clustering technique will be adapted to answer the research question. Unlike the research paper, the optimal number of clusters will be selected based on the sum of square.

3 Properties of the Data

World happiness index is calculated by considering the following main factors: GDP, life expectancy, generosity,

social support, freedom, and corruption. Whereas happy planet index is calculated by combining life expectancy with experienced wellbeing and divided by the country's ecological footprint. For the analysis, HPI data published by New Economic foundation from 2015 to 2020 is considered [2]. Data for world happiness report between 2015 to 2020 is obtained from Kaggle.[1] Both these datasets are merged (861 records ,24 features) and cleaned to get the required data variables for analysis as shown in table (1).

Variable name	**Data Type**
Country	Nominal
HPI Rank	Ordinal
Happiness Rank	Ordinal
Ladder score	Interval
Ladder of life (Wellbeing) (0-10)	Interval
Explained by: Social support	Interval
Log GDP per capita	Interval
Health (Life Expectancy)	Interval
Ecological Footprint (g ha)	Interval
HPI	Interval
Family	Interval
Freedom	Interval
Explained by: Generosity	Interval
Dystopia + residual	Interval

Table (1)

Natural log of GDP per capita is used for analysis, as the data is significantly better than GDP per capita. As in Fig (1) Box plots were used to detect the outliers. Missing values are investigated. After merging the datasets, some columns where dropped such as Region and Standard error of ladder score. It detected huge number of null values and we could possibly do not have any insights from them. United Arab Emirates had one missing value for Government Corruption. It was replaced by taking mean of values of the column values from 2015 to 2020 for U.A.E.

Also, Data was missing for the following countries in happy planet index report of 2020. Hence these was removed to have better clarity in analysis between pre and post covid period. The countries removed from analysis due to missing values are: Afghanistan, Armenia, Bhutan, Botswana, Burkina Faso, Burundi, Cambodia, Central African Republic, Chad, Comoros, Congo (Kinshasa), Guinea, Haiti, Hong Kong, Laos, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mozambique, Namibia, Niger, North Macedonia, Rwanda, Sierra Leone, Sri Lanka, Sudan, Taiwan, Togo, Trinidad and Tobago, Uganda, Yemen and Iceland.

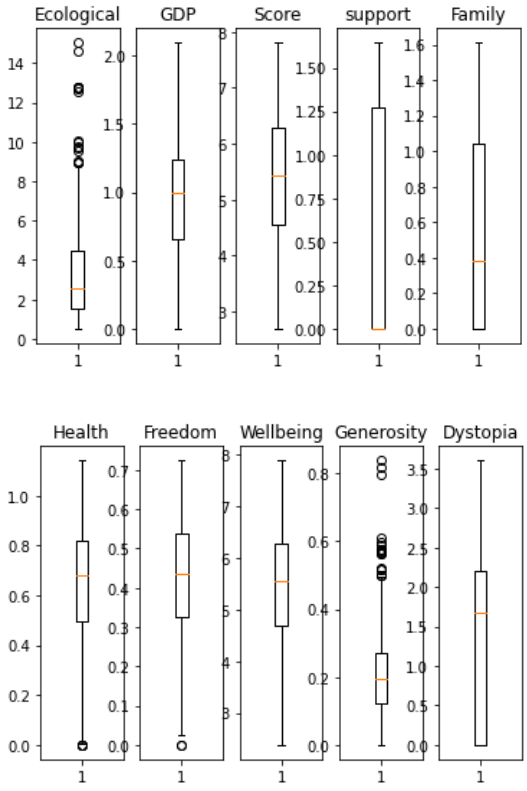

Fig. 1. Outlier detection using Boxplot

4 Analysis

4.1 Approach

In this section we will be discussing the approach followed to answer out the research questions. At first the required datasets need to be merged. Then it should undergo the pre-processing stage of data cleaning. This helps to have accurate models and improved visualization. Here the missing values will be handled and outliers are removed. The least important features will be dropped through feature engineering. Python is used to perform data cleaning. Data pre-processing stage will be revisited to improve the results. Now each research question is analysed in the following way:

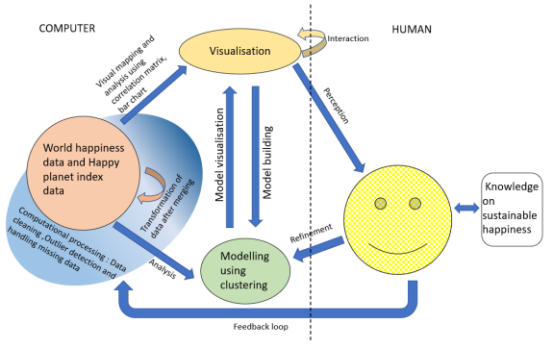

Diagram. 1. Visual Data Exploration. (Idea reference [10])

Approach for question 1:

For the initial analysis of data visual approach will be used. This will help us to compare the results efficiently and ultimately improve our knowledge and decisions. Since we need to compare the ladder score and HPI for countries, the best way to have a visual check of accuracy and trends is by using multiple bar graphs.

Approach for question 2:

For analysing the major factors affected post covid we will first figure out the important features that influence the ladder score. To study the direction of the linear relationship and the strength between the features Pearson correlation coefficient will be used. Using correlation matrix, we can effectively analyse the correlation coefficient between these features visually. And also check if there is any issue of multicollinearity between them.

Approach for question 3:

Correlation matrix will be used to examine if there is any relationship between ecological footprints and the major factors that affect the happiness of countries. The relationship between these variables will be further explored visually using scatter plot. Further for understanding the effects of ecological factors pre and post pandemic we will visually inspect using choropleth map. This way we can study the effect of each feature against one another. We will consider the top happiest and least happy countries for this study.

Approach for question 4:

To evaluate the effects of ecological footprints post covid we use Kmeans clustering technique. The statistical significance of the features will be investigated using the p value. The number of clusters is selected depending on the sum of squares. The countries with high and low ladder score will be considered for the analysis. The clusters will be created on the basis of ecological footprints. Human reasoning is needed to interpret the analysis and choose most favorable number of clusters.

4.2 Process

The analysis is performed as suggested in the previous section for answering the following research questions:

1. How does HPI and ladder score differ for worlds happiest and least happy countries.
2. What are the major factors that affected the happiness of countries with high and low ladder score post pandemic?
3. How does the major factors affecting the happiness, perform against ecological footprints pre and post pandemic?
4. How ecological footprints affected ladder score post covid. Does implementation of lockdown have any effect on ecological footprints.

Process for question 1:

The worlds happiest and least happiest countries are selected based on ladder score. For which we apply filter on countries with a condition set as top 10 countries with high ladder score. Similarly bottom 10 countries with low ladder score is selected. To compare HPI (Happy Planet Index) and ladder score for happiest and least happy country, bar chart is plotted with country, year as column and HPI, Ladder score as rows. Year is chosen as legend. From Fig 2, we could see that after pandemic HPI increased for the most of the countries with high happiness score except Sweden. Sweden has negligible difference in HPI post pandemic. Whereas for ladder score the following countries had shown improvement after the pandemic: Austria, Denmark, Finland, Sweden and Switzerland. The happiest countries which had a dip in ladder score post pandemic are Australia, Canada, Netherland, New Zealand and Norway. When analysed countries with low ladder score, we could see that most of them had improved the happy planet index score except Benin. On the other hand, when we consider the change in ladder score only Benin, Georgia, Ukraine and Tanzania is seen to have some progress. The rest of the countries like Egypt, Ethiopia, India, Myanmar, Zambia and Zimbabwe had a decline in ladder score post pandemic.

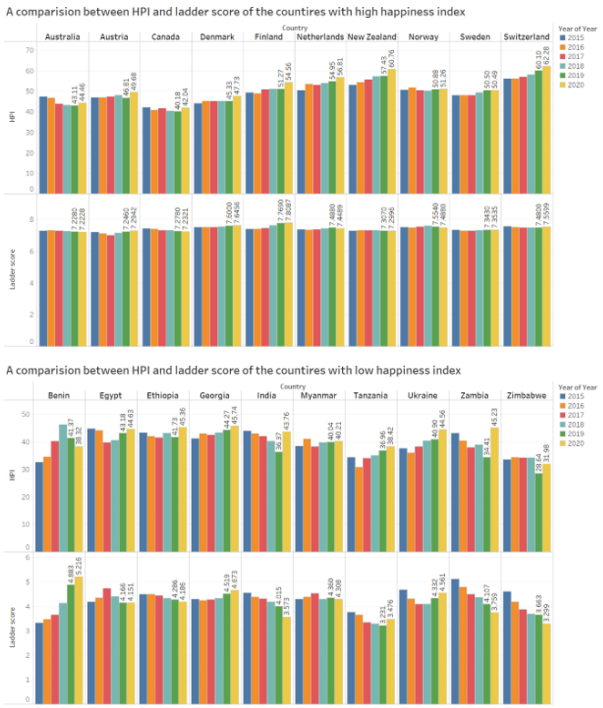

Fig. 2. Bar graph analysing HPI and ladder score

Process for question 2:

After data cleaning, for investigating the major features which affect happiness of a country we constructed a correlation matrix using python. From fig 3, we can see that the features such as happiness rank, Ladder of life (Wellbeing), Log GDP per capita, Health (Life Expectancy) and Ecological Footprint (g ha) are highly correlated to ladder score. It's also observed that there is issue of multicollinearity for Log GDP per capita and Health (Life Expectancy) against ladder score. Pearson correlation coefficient for log GDP per capita against ladder score is 0.72 and for Health (Life Expectancy) against ladder score it is 0.73. For cases involving model prediction one of these features can be ignored to improve accuracy and precision. Thereby build reliable model for prediction. The correlation coefficient for happiness rank, Ladder of life (Wellbeing) and Ecological Footprint (g ha) are -0.85,0.92 and 0.65 respectively.

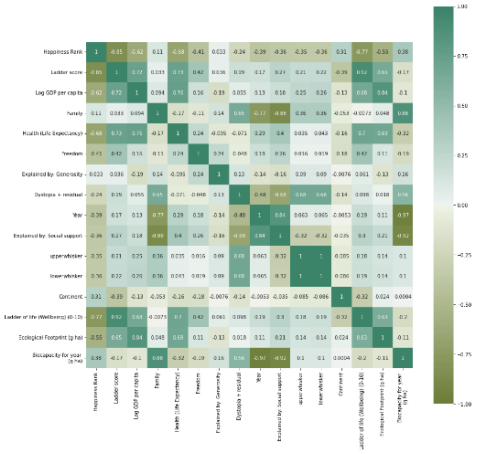

Fig. 3. Correlation matrix

Process for question 3:

The study on the correlation between features reveal that Ladder of life (Wellbeing), Log GDP per capita, Health (Life Expectancy) and Ecological Footprint (g ha) are some of the highly correlated one against ladder score. Hence the effect of ecological footprints is investigated against Ladder of life (Wellbeing), Log GDP per capita and Health (Life Expectancy). In order to understand the linear relationship between these features scatter plot is used. It's seen that features Ladder of life (Wellbeing), Log GDP per capita and Health (Life Expectancy) shows a logarithmic trendline. The P-value is less than 0.0001 which clearly indicates all the features are statistically significant. From Fig 3 its observed that ecological footprint has a strong correlation against log GDP per capita. The correlation coefficient between them is 0.84. The R square value from Table (2) prove this further by showing the proportion of variance in the GDP that is predictable from the Ecological footprints. On the other hand, from fig 3 we see that ecological footprint is less correlated to Health (life expectancy) and Ladder of life (Wellbeing). The correlation coefficient for them is 0.69 and 0.63 respectively.

For further studying the impact of ecological footprint reduction pre and post pandemic, Choropleth map is created. Visualization tool Tableau is used for this. The research is done mainly on countries with high and low ladder score. For having visual clarity in the analysis labels are not overlapped. From the Fig. 5 its transparent that countries with high ladder score such as Australia and Canada had a decline in log GDP per capita, Health (life expectancy) and Ladder of life (Wellbeing) but still shows improvement in ecological footprint reduction. Finland is said to be the happiest country in the world. However, when it comes to ecological footprint reduction Finland is seen to have a higher value than many of the countries with low ladder score. But we could also observe that Finland shows improvement in ecological footprint reduction post pandemic. It is observed that both log GDP per capita and Health (life expectancy) was also reduced in Finland post pandemic. Still the ladder of life (Wellbeing) is shown to have improved for this nation. The countries with poor ladder score such as India, Ethiopia, Tanzania, Zambia and Egypt have seen to have improvement in ladder of life (wellbeing) after lockdown. They also perform well when it comes to ecological footprint reduction. On the other hand, there was a shrink in Health (life expectancy) and log GDP per capita for these countries post pandemic.

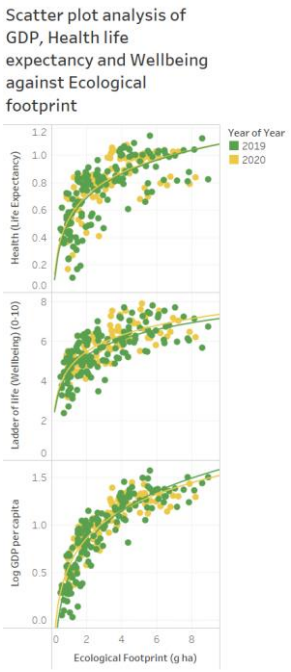

Fig. 4

Year	R square for Health (Life Expectancy)	R squared for Ladder of life (Wellbeing)	R squared for Log GDP per capita
2019	0.573575	0.533482	0.800486
2020	0.523137	0.557306	0.808056

Table (2)

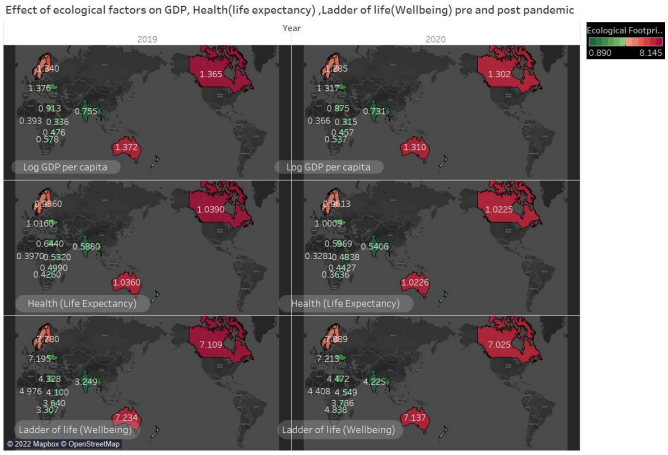

Fig. 5 Choropleth map to analyse effect of ecological footprints

Process for question 4:

Kmeans clustering was performed after completing data pre-processing stage. The variables Ecological footprint and ladder score were considered for clustering. This is to study how well do countries with high and low ladder score perform against ecological footprint post covid. For the analysis countries with high and low ladder score are filtered. Initially four clusters were created. Since sum of square improved when clusters were reduced to three, it was chosen for the study. From choropleth maps in fig 6 its clear that the pattern of cluster pre pandemic and post pandemic had some variance. It's observed that countries with high ladder score such as Canada and Australia had a change in cluster grouping. From the table 3, its evident that this switch in the cluster group is due to the change in ecological footprints. These two countries had reduced its ecological footprint was post covid. From fig 6, we could see that in 2019 the value of ecological footprints was 7.525 and 8.145 respectively for Australia and Canada respectively. In 2020, this got reduced to 6.951 and 7.248 for each of them. Hence, they switched the clusters. From Fig 6, it's also evident that ecological footprints reduced for most of the countries post pandemic. From Table 4 the statistical significance of selected variables can be analysed using the p-value. The p- value is seen to be very low for both ecological footprint and ladder score which indicates that both these features are statistically significant.

Year	Cluster	Number of items	Ecological footprint range
2019	Cluster1	6	3-6
	Cluster2	4	6-8.5
	Cluster3	10	0-3
2020	Cluster1	7	5-8
	Cluster2	3	3-5
	Cluster3	10	0-3

Table (3)

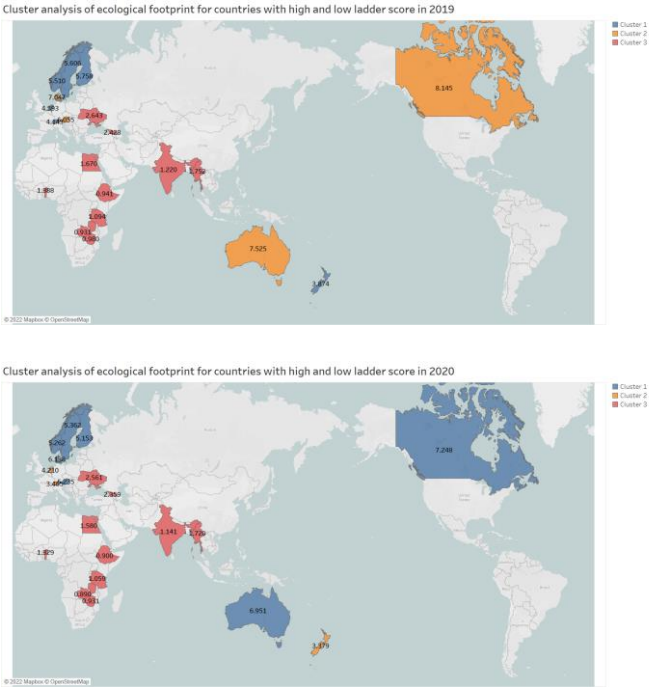

Fig. 6 Choropleth map for Kmean cluster analysis pre and post pandemic

Year	Variable	F-statistic	p-value	Model sum of square	Error sum of square
2019	Sum of Ladder score	8.18	0.003247	2.604	2.706
	Sum of Ecological Footprint (g ha)	7.828	0.003849	2.033	2.208
2020	Sum of Ladder score	7.978	0.0036	2.702	2.879
	Sum of Ecological Footprint (g ha)	7.714	0.004131	2.037	2.244

Table (4)

4.3 Results

To summarise the results, we conclude that through bar graph analysis the happy planet index score improved for most of the countries after pandemic. But there was a slight decrease in ladder score for many after pandemic, except for countries such as Benin, Georgia, Tanzania, Australia, Canada, Netherland and Norway. Through correlation matrix its evident that the major factors affecting the ladder score of countries are happiness rank, Ladder of life (Wellbeing), Log GDP per capita, Health (Life Expectancy) and Ecological Footprint (g ha). Study on the effect of these major factor against Ecological footprints help us to arrive at fact that Ladder of life (Wellbeing) improves when ecological footprint reduces.

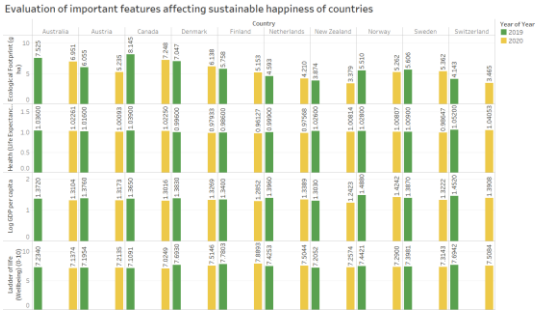

Fig. 7

Fig (7) also shows that neither log GDP per capita nor Health (life expectancy) have much progress when ecological factors are reduced. Through Kmean techniques we could distinguish the countries with respect to ecological footprints. It was observed that Western European countries like Finland shows great progress by reducing the ecological footprints post pandemic. Australia, New Zealand and Canada also shows good improvement. Some Sub- Saharan African and South Asian countries also having good progress in reducing ecological footprint. The notable one among them is India.

5 Critical Reflection

It is interesting to note that countries with high ladder score perform poorly when it comes to ecological footprint. However, using choropleth maps we could see an improvement in this score over year. This clearly highlight that the developed countries are right now realising the importance of sustainable happiness and have initialised further actions and decision making accordingly. Also, this could be an effect of imposed lock down as well. It is also observed that though the countries have a decline in log GDP per capita and Health life expectancy post pandemic, the overall ladder of life (Wellbeing) is improved. Such countries are seen to have reduced its ecological footprint as well. The reduction in Health (life expectancy) and GDP per capita could be an effect of pandemic. However, it helped to improve the overall environment stability. This could be due to the restriction imposed on movement of people. This helped in reducing the environment pollution globally.

Through the visualisation techniques used in the study its evident that the countries with lower ecological footprints tends to have better ladder of life (Wellbeing). Both countries with high and low ladder score was affected due to pandemic. This is clear by seeing the drop in Health (life expectancy) and GDP per capita. From the study we could conclude that all the countries where equally affected by the pandemic but it indirectly helped to reduce ecological footprint due to lockdown and restrictions. The study is restricted around the countries with high and low ladder score. Also, we had to remove many countries from analysis due to missing value. This limit width of the study conducted. Also, it could ignore some exceptions in the data. For example, Countries such as Iceland and Costa Rica had to be excluded from the research due to data missing. These countries had good ladder score and Happy planet index score. For future study, we will try to include all the countries by collecting as much as information from the source data and build a predictive model after applying dimension reduction technique. By applying techniques like PCA we can analyse the important features to be considered for the modelling.

Table of word counts

Problem statement	250/250
State of the art	358/500
Properties of the data	342/500
Analysis: Approach	402/500
Analysis: Process	1279/1500
Analysis: Results	199/200
Critical reflection	360/500

References

[1] Datasets:https://www.kaggle.com/unsdsn/world-happiness
[2] Datasets: https://happyplanetindex.org/
[3] Carlsen, Lars. (2018). Happiness as a sustainability factor. The world happiness index: A posetic-based data analysis. Sustainability Science. 13. 10.1007/s11625-017-0482-9.
[4] Campus, Alba & Porcu, Mariano. (2010). Reconsidering the well-being: the Happy Planet Index and the issue of missing data. Centre for North South Economic Research, University of Cagliari and Sassari, Sardinia, Working Paper CRENoS.
[5] Landes, X., Unger, C., Andsbjerg, K., Frank, K., & Wiking, M. (2015). Sustainable Happiness: Why Waste Prevention May Lead to an Increase in Quality of Life. Danish Ministry of the Environment.
[6] O'Brien, Catherine (2008). "Sustainable Happiness: How Happiness Studies Can Contribute to a More Sustainable Future," Canadian Psychology, 49:4, pp.289-295.
[7] Lee, Angela & Kudbuddin, Ammar & Teh, Phoey. (2019). Data Visualisation: World Happiness at a Glance!. 10.1007/978-3-030-34032-2_6.
[8] Helliwell, John & Huang, Haifang & Wang, Shun. (2015). The Geography of World Happiness.
[9] Bondarchik, Julia & Jabłońska-Sabuka, Matylda & Linnanen, Lassi & Kauranne, Tuomo. (2017). HIP - a Happier Index for the Planet?.
[10] Keim, Daniel & Kohlhammer, Jörn & Ellis, Geoffrey & Mansmann, Florian. (2010). Mastering The Information Age – Solving Problems with Visual Analytics.

Appendices:

Larger figure for Fig (5)