

The ecological cost of social progress

HarvardX Data Science Capstone Own Project

Codrin Kruijne

2019-06-04

This report has been prepared for the [HarvardX Data Science program](#) Capstone course final project submission and consists of three files of original work that can be found [on GitHub](#) (with all commits): this [report in PDF](#), this [report in R markdown](#) and an [R markdown script file](#) with all the code (also [knitted to PDF with results](#) for convenience). Results are read in from data objects stored locally by the script (also saved in the GitHub repository), that were the output of the script file. We were limited in our visualisations by the requirement to publish to PDF; interactive visualisations by publishing to HTML would have been more insightful.

Summary

Humankind uses diverse progress indicators, most currently Growth Domestic Product (GDP) to gauge economic growth. It is becoming more apparent that this is a limited and myopic view and alternative measures have been emerging over the last years, like the United Nations Human Development Index. I wanted to have a closer look at two concepts that follow; the Social Progress Index (SPI) and the Happy Planet index (HPI), which take into account the state of our ecosystems. This report describes merging the last available data for both from the year 2016, applying clustering to see how certain countries group, and having a look at their differences. Resulting clusters show that social progress seems to come at the cost of ecological systems that cannot sustain levels of accompanied consumption.

Analysis

Raw data

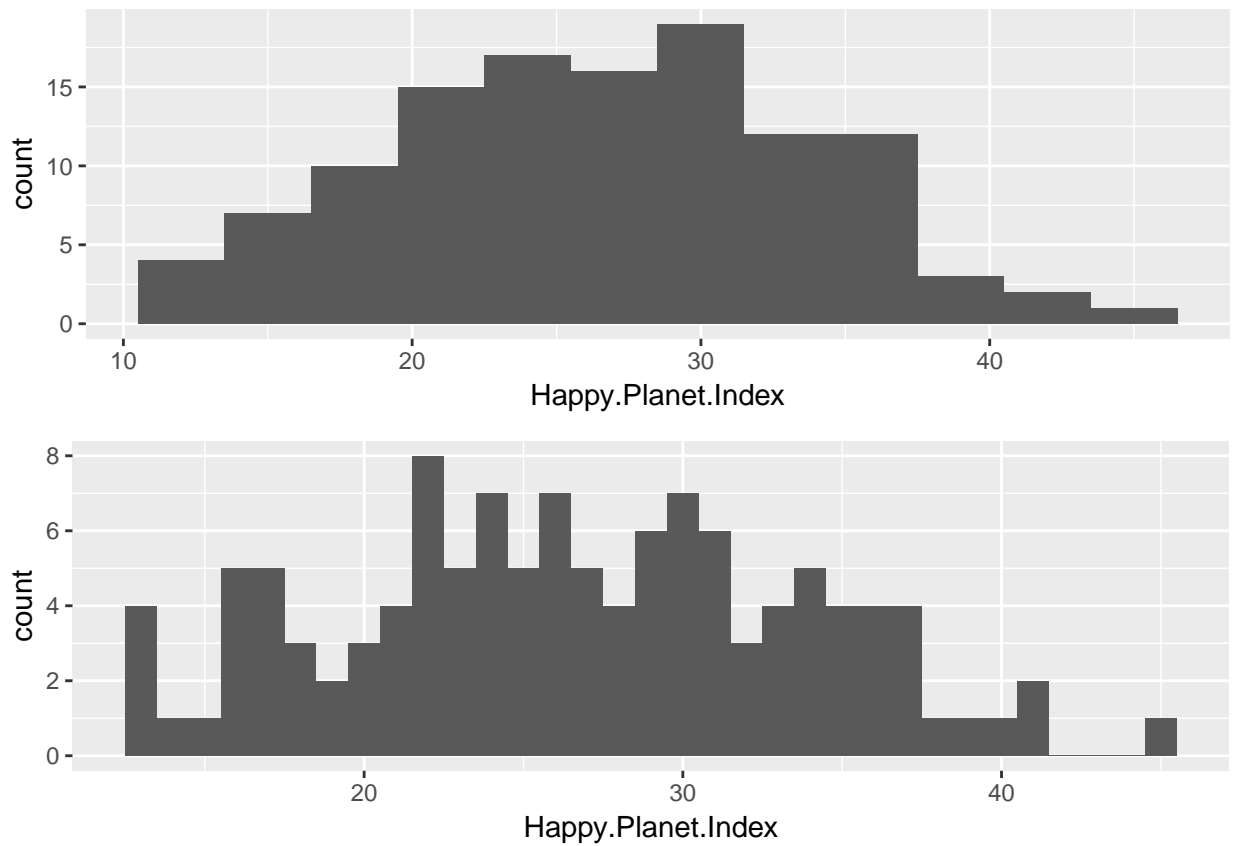
Data was downloaded from the official websites as Excel files and the last common year, 2016 was read in. HPI data in short looks at the amount of happy life years people in a country achieve, divided by the ecological footprint or 'ecosystem costs' to obtain that to balance means and ends. SPI data includes more diverse indicators with the main categories being Basic Human needs, like health and safety, Foundations for Wellbeing, like access to knowledge and quality environment and, last, Opportunity, with freedom, rights and inclusiveness. Links to the websites can be found in references.

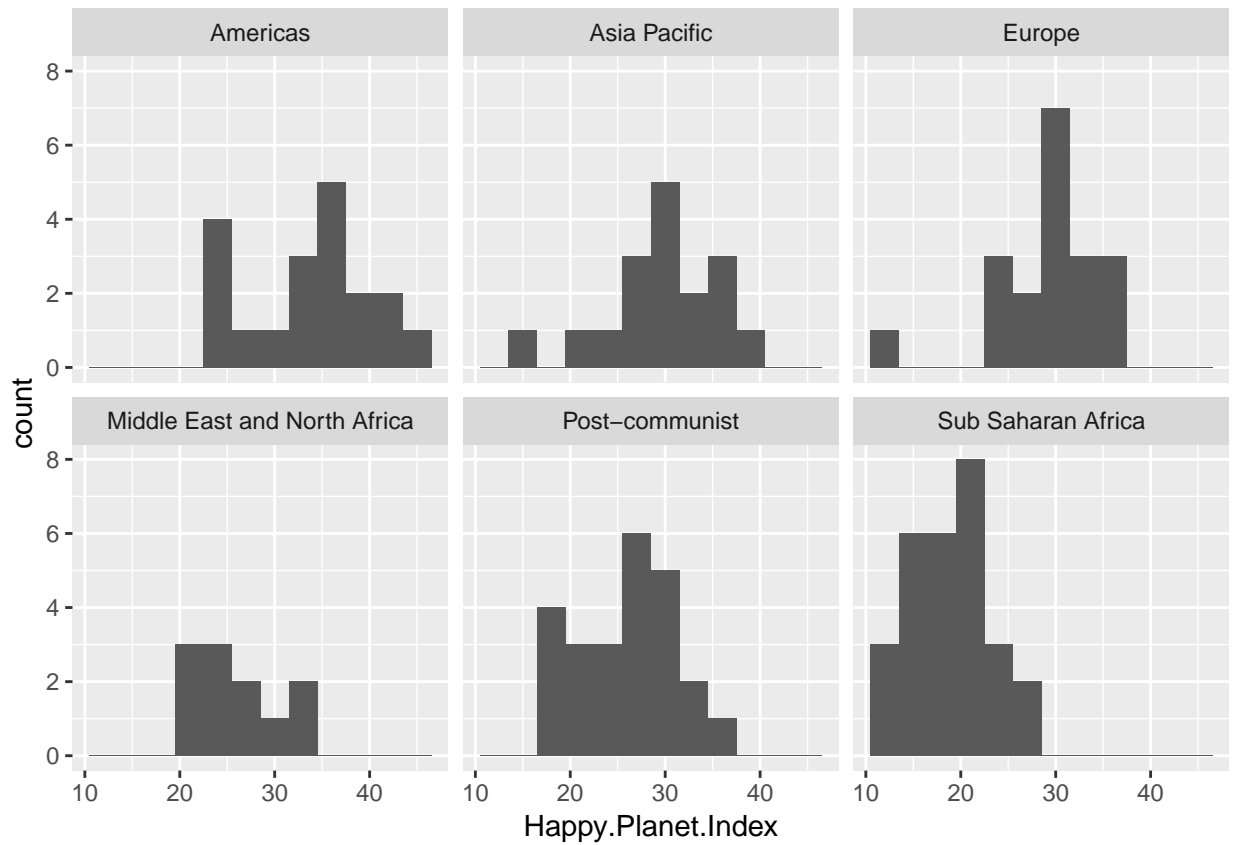
Data preparation

Data from HPI and SPI were joined by Country name. Missing values imputed by mean values. All numeric variables were scaled in preparation for clustering. This leaves us with 118 observations of 87 variables. We explore the data next.

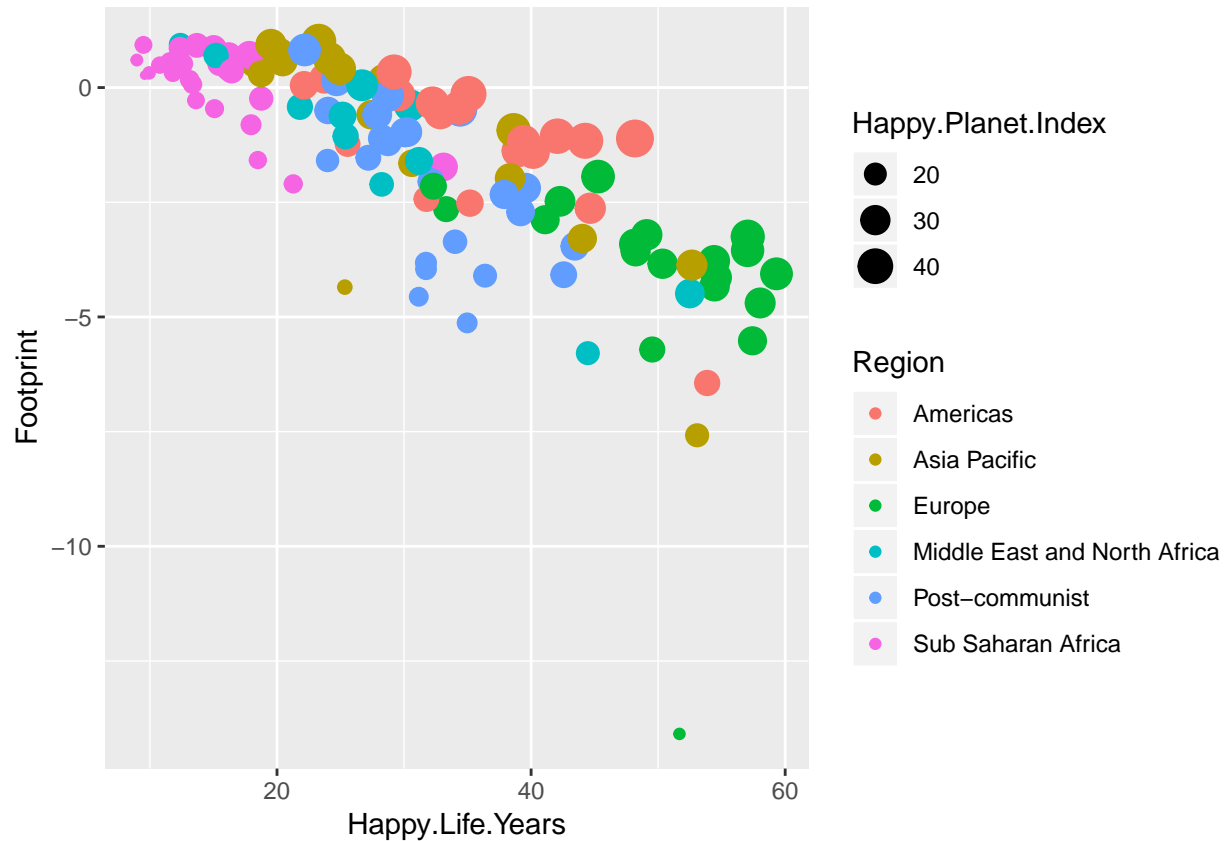
Data exploration

To get an initial view of the data we plot histograms of HPI scores with different bin widths and a scatter plot with major variables.





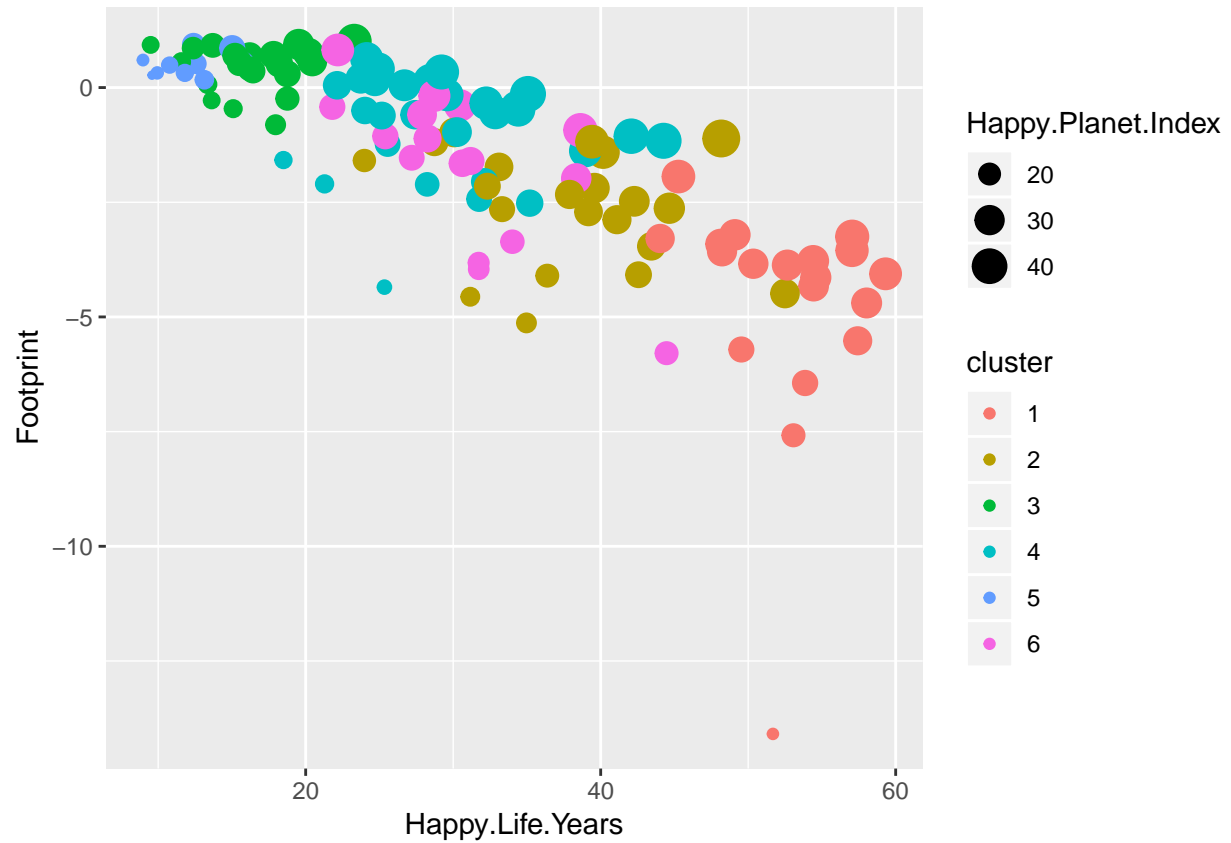
Looking at all HPI scores the histogram seems to show a somewhat normal distribution. The HPI data is already labeled with regions and separating them clearly shows differences, most notably contrasts like the low scores in Sub Saharan Africa and the distributed, yet high scores of the Americas.



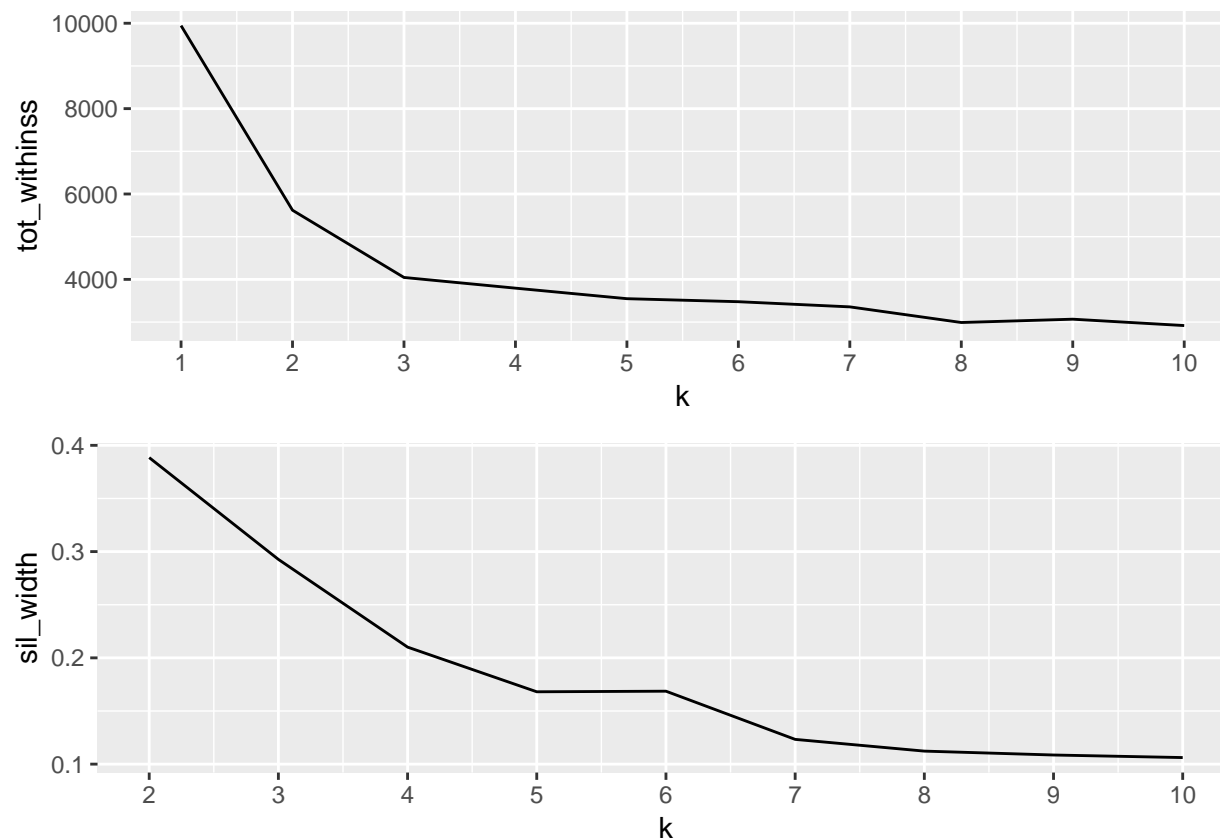
The scatterplot shows HPI on the x-axis and the effective footprint on the y-axis. Footprint data represents how big an area of ecosystems is needed (in “Global Hectares per person”) to realise the consumption levels in that country (corrected for many things). 1.73 is the theoretical amount our world would be able to sustain. I chose to show the difference on the y-axis; a positive value means countries consume within planetary boundaries; a negative value shows a consumption rate requires more ecosystem capacity than is available. Here you see differences and the underlying values with notably Sub Saharan africa and Europe at opposite ends. As we have more data from the SPI, I wanted to explore what clustering based on all variables beyond region would yield for insights.

Clustering modeling

As the data provides six regions, we use that for the amount of initial clusters, after which we will calculate the most appropriate amount of clusters. We used K-means clustering.



These clusters calculated on all variables, not just Region, differ quite a bit from the geographical region clusters. Lets calculate the optimal amount of clusters using the Elbow method and Silhouette width.

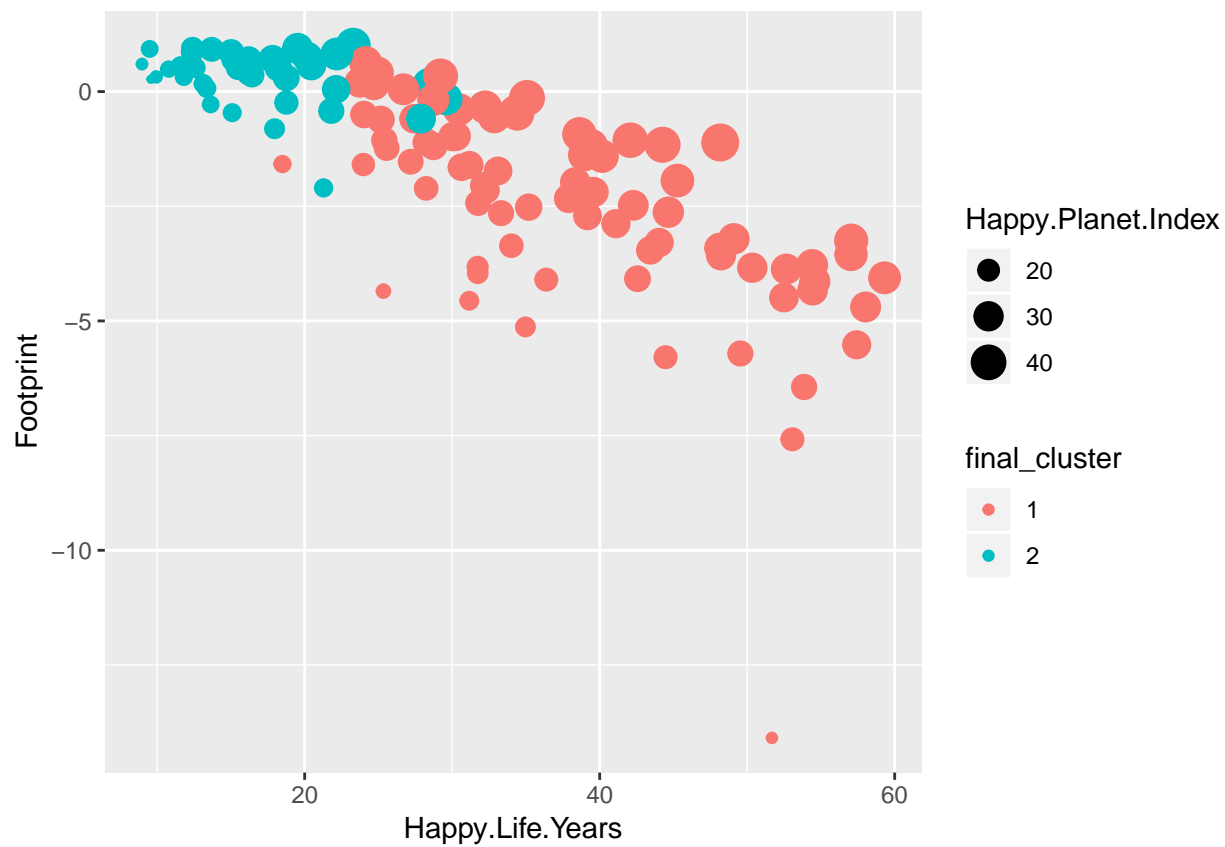


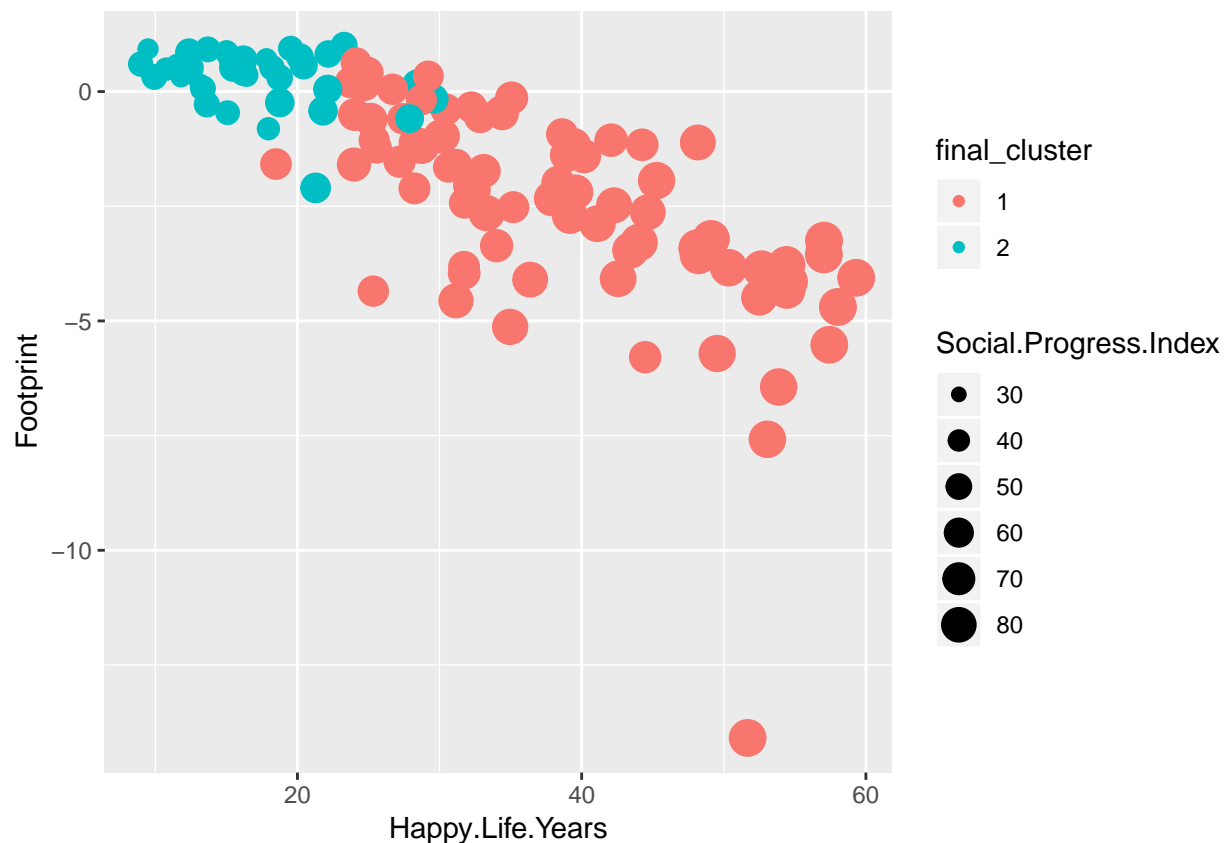
Fitting several cluster models (up to $k = 10$) shows two clusters seems to be the best separation. At $k = 2$ we see the elbow bend and the average silhouette width is highest when data is separated into two clusters.

Results

Visualising clustered data

The figure below show the data with clusters identified by color and size representing first HPI and then SPI scores.

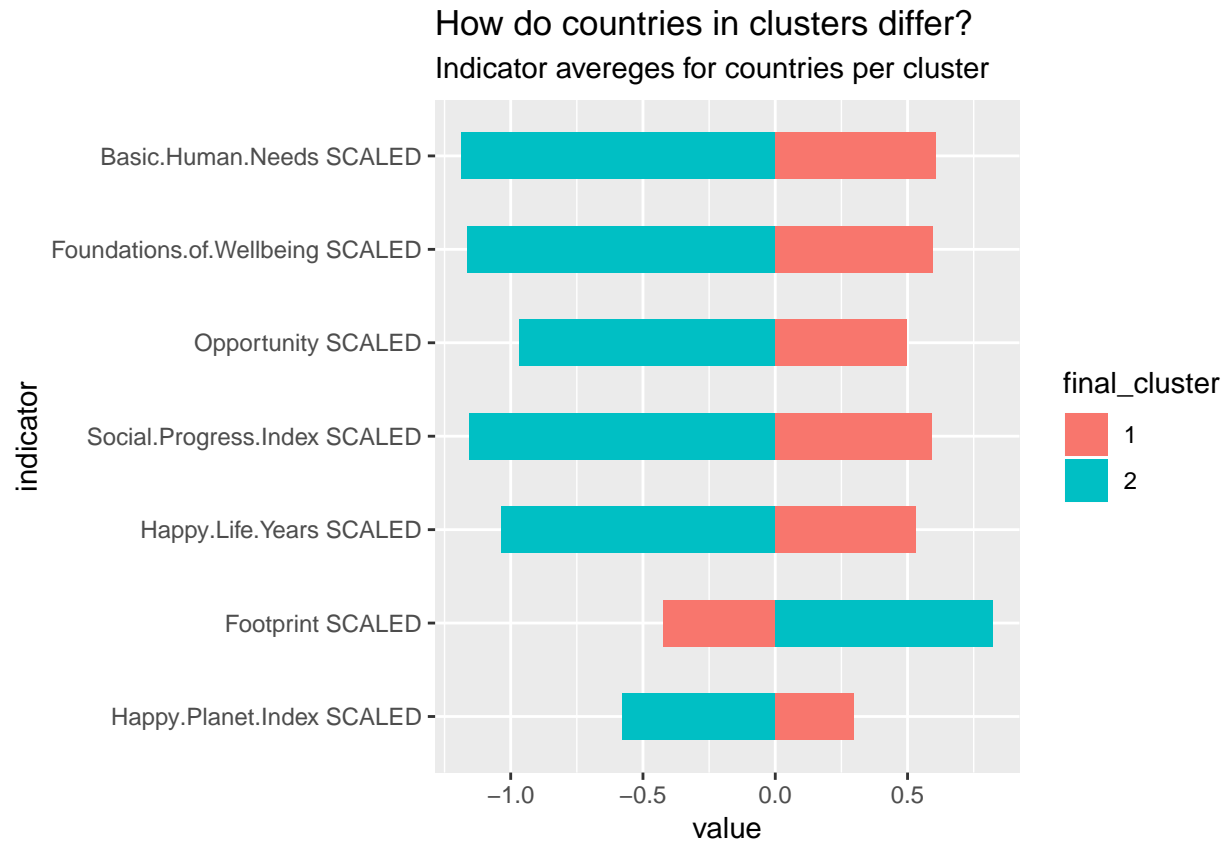




The first cluster seems to have most countries with a positive footprint, meaning their consumption can be sustained, while the second cluster has increasingly Happy Life Years and more negative footprint. Using HPI as size for datapoints accentuates that the lower happy life years and the higher negative footprint, the less is HPI score, whereas SPI score visually seems similar putting an emphasis on social progress.

Cluster analysis

To better understand the differences in countries we plot the average scores per cluster on the major components of both the HPI and SPI indicators.



We can see a clear distinction between clusters on average scores of a select number of indicators. Countries in cluster 1 score above average except for their environmental footprint; they are doing well socially, but at the cost of our ecological systems. Countries in cluster 2 score well below averages, except for their footprint.

Conclusions

Although the results here show relative performance of clusters, a major concern is the high number of countries with absolute negative net footprint; meaning we are consuming not only the services or results of our ecosystems, but the production capacity itself, which will lead to increasing footprint pressure. Further analysis into all measures may provide additional insights into differences in these clusters. Particularly interesting would be to explore countries with relatively high social progress and positive net footprint.

References

1. [How is the Happy Planet Index calculated?](#)
2. [Social Progress Imperative](#)

Appendix

Cluster.1	Cluster.1		
Albania	Latvia		
Algeria	Lebanon		
Argentina	Lithuania		
Armenia	Luxembourg		
Australia	Macedonia		
Austria	Malaysia		
Belarus	Mauritius		
Belgium	Mexico		
Bhutan	Mongolia	Cluster.2	Cluster.2
Bolivia	Montenegro	Afghanistan	Lesotho
Brazil	Morocco	Bangladesh	Liberia
Bulgaria	Netherlands	Benin	Malawi
Canada	New Zealand	Botswana	Mauritania
Chile	Nicaragua	Burkina Faso	Mozambique
China	Norway	Burundi	Myanmar
Colombia	Oman	Cambodia	Nepal
Costa Rica	Panama	Cameroon	Niger
Croatia	Paraguay	Chad	Nigeria
Cyprus	Peru	Comoros	Pakistan
Czech Republic	Philippines	Djibouti	Rwanda
Denmark	Poland	Egypt	Senegal
Dominican Republic	Portugal	Ethiopia	Sierra Leone
Ecuador	Romania	Ghana	Swaziland
El Salvador	Russia	Guatemala	Tajikistan
Estonia	Serbia	Guinea	Tanzania
Finland	Slovakia	Honduras	Togo
France	Slovenia	India	Uzbekistan
Georgia	South Africa	Indonesia	Yemen
Germany	Spain	Kenya	Zimbabwe
Greece	Sri Lanka		
Hungary	Suriname		
Iceland	Sweden		
Iran	Switzerland		
Ireland	Thailand		
Israel	Tunisia		
Italy	Turkey		
Japan	Ukraine		
Kazakhstan	United Kingdom		
Kyrgyzstan	Uruguay		