Insights of World Happiness

Rui Song

Xiao Wan

Xiaoyu Zhang

Instructor: Feng Mai

Stevens Institute of Technology

BIA-658-C

2017 Spring

Contents

1. Project Introduction

2. Data Resource

3. Network Construction --- Data Processing

4. Network Construction --- Visualization

5. Network Analysis

6. Community Analysis

7. Conclusions

8. Recommendations

9. References

**1. Project Introduction**

Are you happy? We all have asked ourselves at some moment in our life. Down to the bottom, we all only live once, and happiness matters to all people. And it is worthwhile to explore what are the components that make people happy.

Saints, especially in ancient China. Always educate people that money could not buy happiness. Ancient Chinese poetries have written thousands of poems describing the enjoyment of living a frugal life. Similar philosophy also exists in other ancient civilizations. But is it the case? Our idea is that individuals cannot avoid making a judgement based on feelings and emotions vary amount individuals, different groups, people from different cultures. In short, you won't expect everyone to enjoy the lifestyle of Buddhas.

In our project, we will capture the underlying pattern of what defines a happiness country and understand how the different factors are contributing to the happiness in different patterns of the country.

**2. Data Resource**

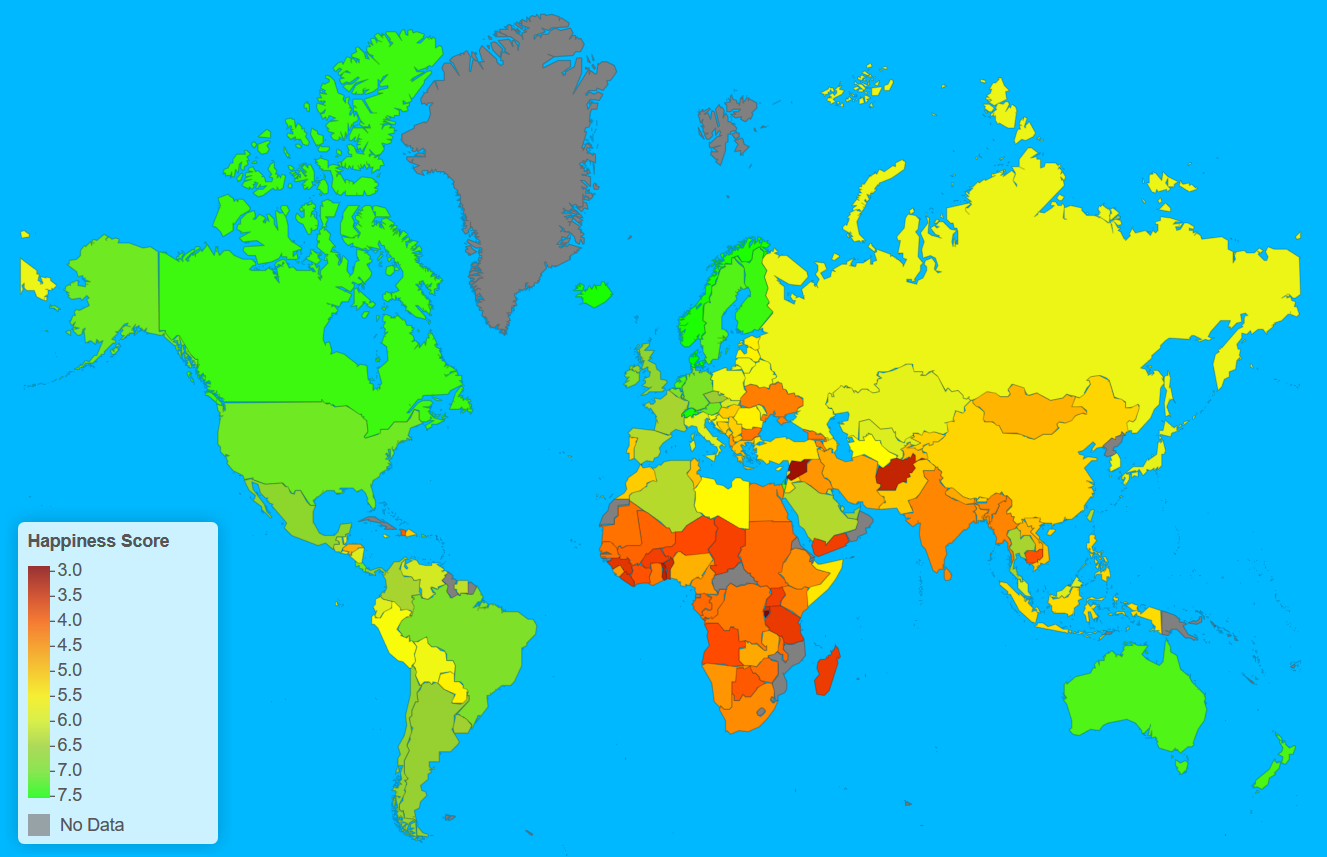


Figure 1. Sample raw data of World Happiness Report (2016)

Luckily, today we are living in a data-driven world. Since 2012, United Nations Sustainable Development Solutions Network have been publishing the World Happiness Report of UN member countries – landmark survey of the state of global happiness. The initiative of which was to measure the happiness of their people and to use it to help guide their public policies.

In the reports, leading experts in several fields--economics, psychology, survey analysis, national statistics, and more—describe how measurements of well-being can be used effectively to assess the progress of nations. [1]



Figure 2. Sample raw data of World Happiness Report (2016)

Data is collected from people in over 150 countries. Each variable measured reveals a populated-weighted average score on a scale running from 0 to 10 that is tracked over time and compared to other countries. These variables currently include real GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption. Each country is also compared against a hypothetical nation called Dystopia. Dystopia represents the lowest national averages for each key variable and is, along with the residual error, used as a regression benchmark. [1]

Also, to enrich our model from visual presentation's perspective, we also explored other resources, mostly from Wikipedia (corresponding entry). From which we gained geological information – latitude and longitude, raw population and GDP per capital. The integrated table was as follows.



Figure 3. Sample raw data of Wikipedia

**3. Network Construction – data processing**

3.1 overall process

In raw data, each country was evaluated and given a weighted average score for each of the six attributes, the higher the score is, the better a country performance on that particular field.

Our goal was to measure the similarity between any two countries for each attribute. The natural intuition was to calculate the difference between two countries. For instance, from table 2.2, we can see that Argentina scored 1.15 in term of Economy while Australia scored 1.44. The difference is therefore 0.29. So when constructing a pair-wise comparison matrix, the cell that corresponds to the relation between Argentina and Australia should be 0.29. However, the problem is, the result of this intuitive approach will lead to reversed strongness. Argentina against itself end up 0, but clearly, this "connection" should be the strongest.

Then the next step is to reverse the score. Before that, we made an assumption that all attributes are equally important – the weight is the same (from later network analysis, we know that this does not hold true). So we scaled it the score from a 10 to 1. 10 means the strongest connection while 1 represents the weakest connection.



Figure 4. Sample of raw data



Figure 5. Sample of adjacent matrix

The upper procedure needed to go through for six times, each time an adjacent matrix corresponding to an attribute was generated. Since at this step, the goal was to gain an overall network that takes all attributes into consideration. Through addition, we incorporate the effect of all the six criterions into one single integrated adjacent matrix.

To construct the network. Clearly, a balance between data representation and visualization effect must be made, attempts to include all edges failed due to one simple fact – when elements increases, a graph become too dense that human eyes cannot efficiently detect patterns if any. The combination of 156 \* 155 is simply uninterpretable, therefore after rounds of trials, we found that to only include the first five strongest connection is the clearest solution.

Figure 6. Sample of filter edges

This process was done automatically through the script in the left blue box. (To decrease the scale, the weight of each connection was divided by 10).

**4. Network Construction – visualization**

Now we have got the adjacent matrix that contains all the nodes – 156 countries that are members of UN, as well as 156 \* 5 filtered edges between countries. The next step is to use R I graph package to construct the network.

“Undirected” – a collection has no starting point and end point.

“Diag = False” – self-loop connection is avoided in our paragraph.

4.1 Community Analysis

Countries were categorized as developing countries and developed countries. But developing countries varies extremely differently. Nigeria and China are both developing countries, but at least for most people, China is regarded as a better country to live.

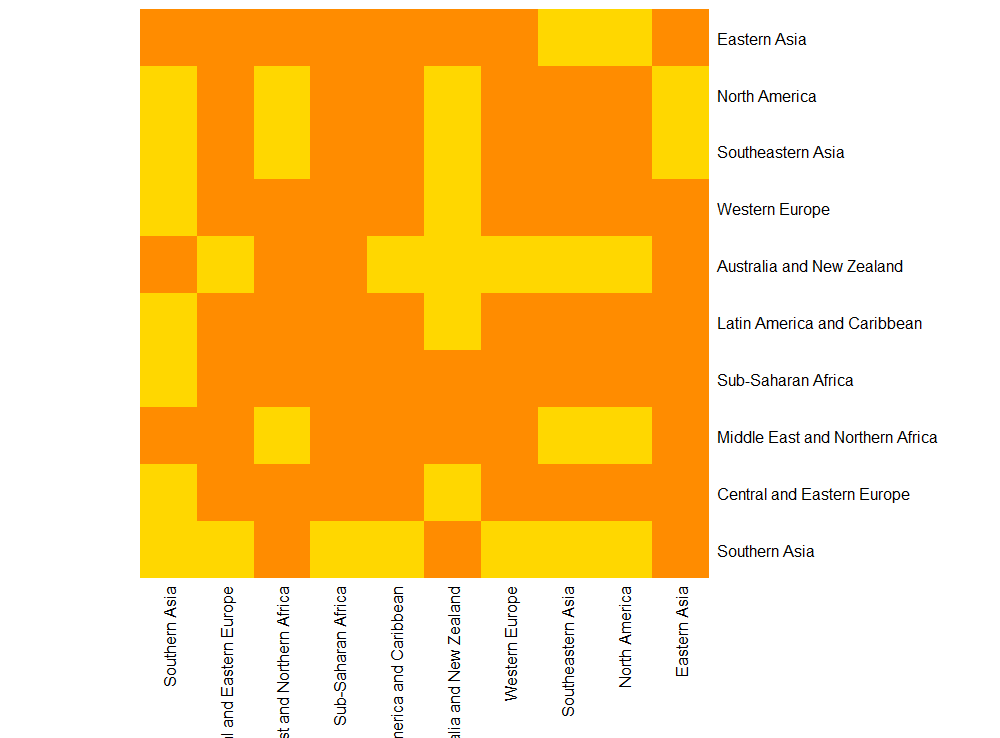


Figure 7. Region heat map

4.1.1 Heat map

A heat map (or heat map) is a graphical representation of data where the individual values contained in a matrix are represented as colors.

Instead of setting each country to stand for an element, we compressed the country network to a region network. Figure 6 shows a clear pattern that regions clusters into four groups.

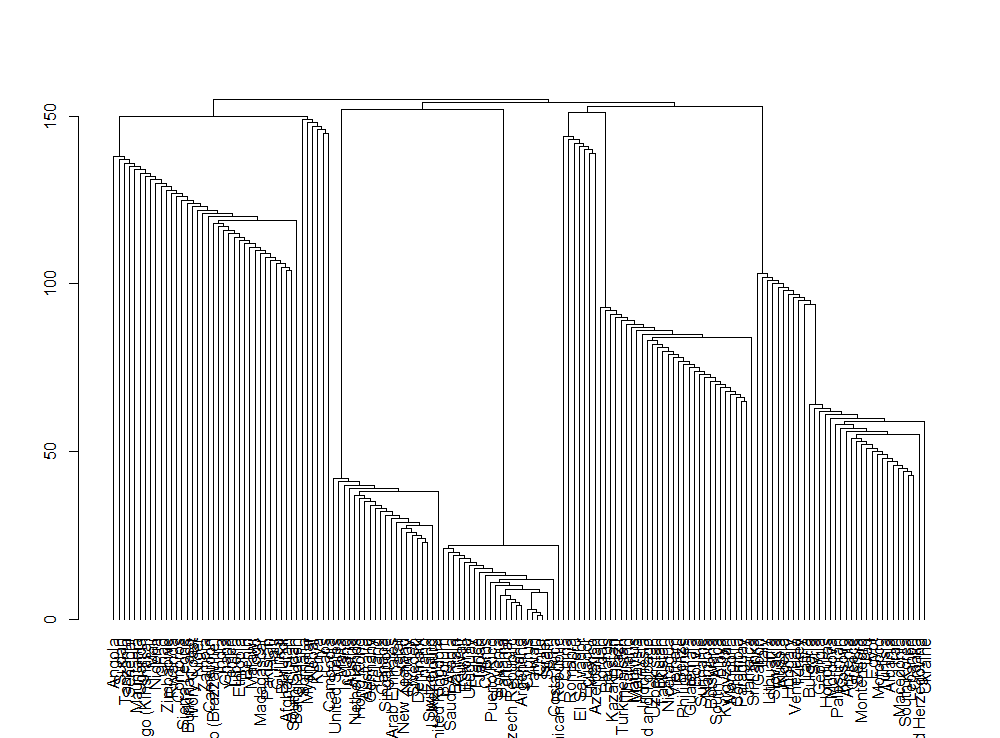
4.1.2 Hierarchical clustering

Figure 8. Hierarchical clustering

In data mining and statistics, hierarchical clustering (also called hierarchical cluster analysis or HCA) is a method of cluster analysis which seeks to build a hierarchy of clusters. Strategies for hierarchical clustering fall into two types

Agglomerative: This is a "bottom-up" approach: each observation starts in its cluster, and pairs of clusters are merged as one moves up the hierarchy.

Divisive: This is a "top-down" approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy. [2]

The above graph represents an agglomerative approach – fast greedy. This function tries to find dense subgraph, also called communities in graphs via directly optimizing a modularity score. [3]

4.1.3 Conclusion

Both heat map and hierarchical clustering analysis demonstrate that four groups are most likely the optimal way to categorize the 156 countries.

**5. Network Analysis**

5.1 Glimpse of worldwide happiness

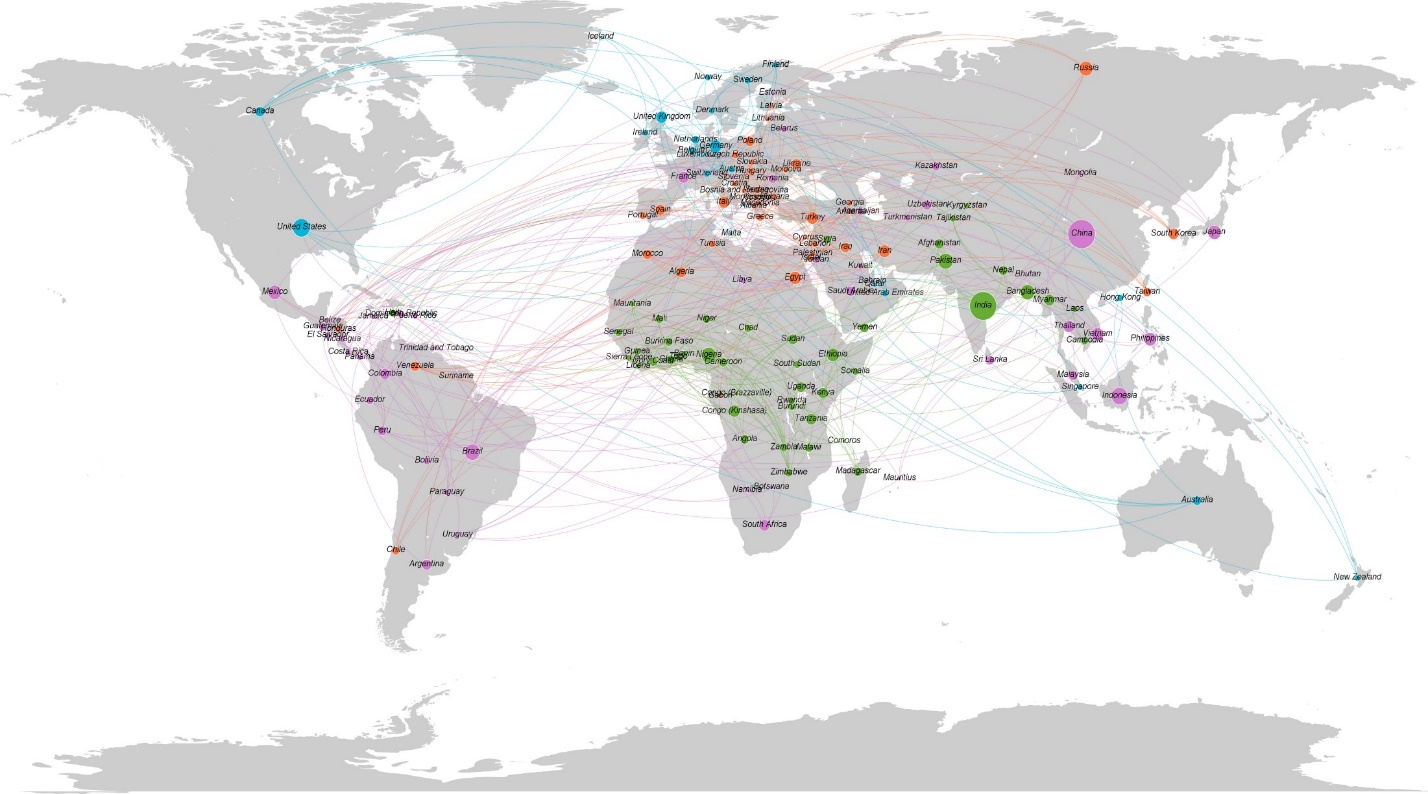


Figure 9. World happiness map

5.1.1 Approach

In section 3, we have already constructed the basic network file using R studio. The next step is to visualize it, Gephi is one of the most powerful network visualization tool available so instead of doing it in R, we used Gephi to gain better visualization result.

As was described before, we have gained the geological as well as population data from Wikipedia. All of which were used in this graph – latitude, and longitude were used to spread the nodes based on its corresponding location on the world map; the population was used to define the size of each node, the bigger the population is, the bigger the population is. But due to the apparent scale difference between the people of countries, log function was used to flat the size.

5.1.2 Interpretation

Now it's obvious to see both the distribution of four categories of countries and also how many people lives in those countries. Following are some observations

1. most of happiest countries spread among we Europe, Northern America and Australia and New Zealand.
2. Most of the unfortunate countries clusters in Sub-Saharan Africa.
3. Former Soviet Union countries were still categorized in the same group
4. In a sense, the similarity between U.S.A and Canada is low (no connection), so is Mexico and America.
5. Also, countries around Mediterranean Sea also show high similarity.
6. China and India are the two most populated countries, but they don't show another similarity

5.2 Clustered network

Figure 10. World happiness map

5.2.1 Happiness axis

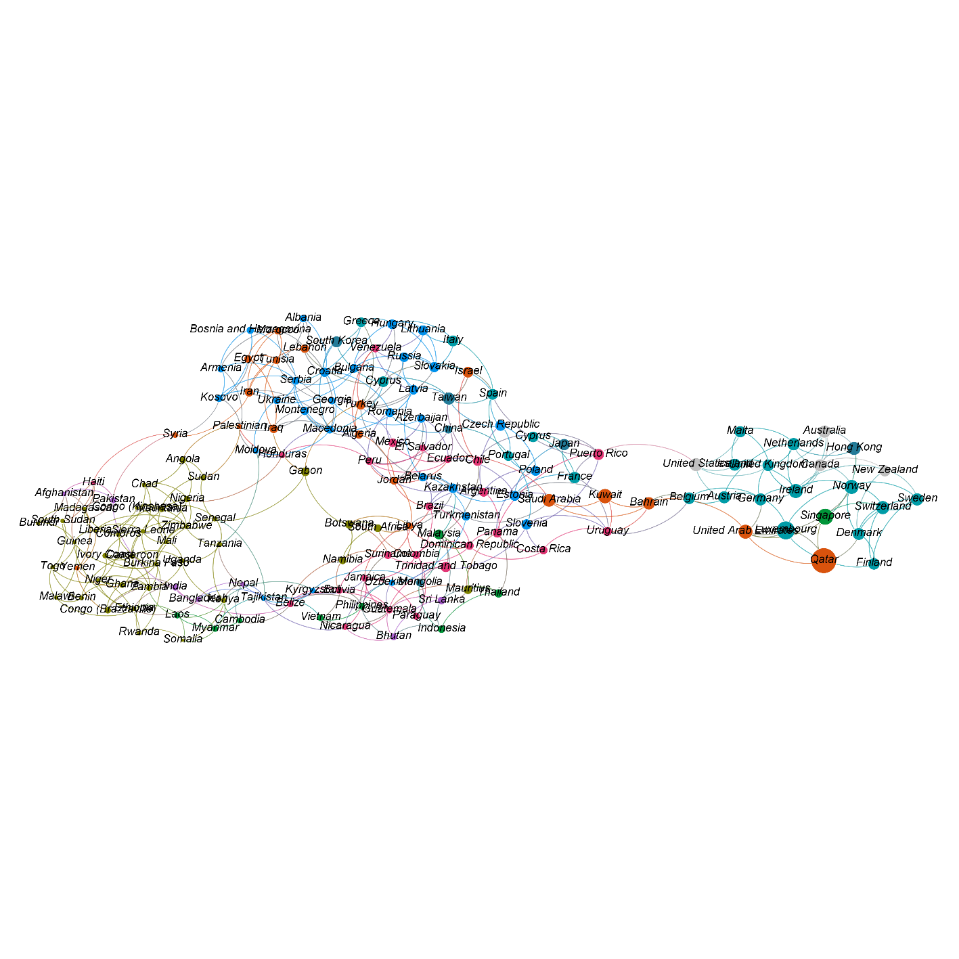
From left to right, the happiness score increases, the size of each node is proportional to GDP per capital, which strongly indicates that happiness is strongly affected by income.

Figure 11. Region distribution

5.2.2 Region Distribution

Compare to the upper picture; we can see that the unhappiest countries centralized in Sub-Saharan Arica. While the happiest countries mostly locate in Western Europe. Meanwhile, in the middle part, one part is dominated by Central and Eastern Europe countries

**6. Community Analysis**

6.1 Region & Freedom

Now we consolidated the big graph by region, to get a clear view of the similarity between regions (Figure 12). Here are some interesting observations:

1. A strong triangle connects Central and Eastern Europe, Latin America and the Caribbean, Middle East and Northern Africa, which indicates that countries in these three regions are quite alike. This also fits our intuition.
2. Western Europe has connections to every region, weak or strong, except Sub-Saharan Africa and Southern Asia while these two regions show high similarity.
3. Southern Asia is quite distinct from Eastern Asia, which shares the same continent but with no link between them.

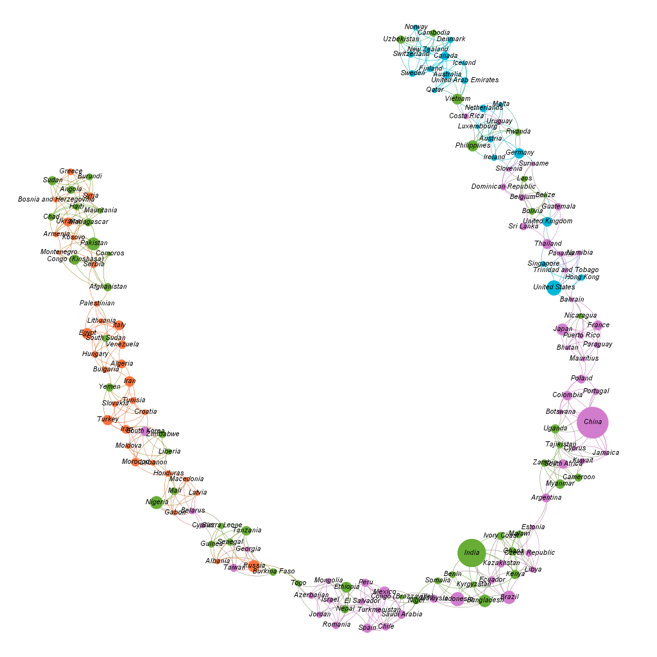
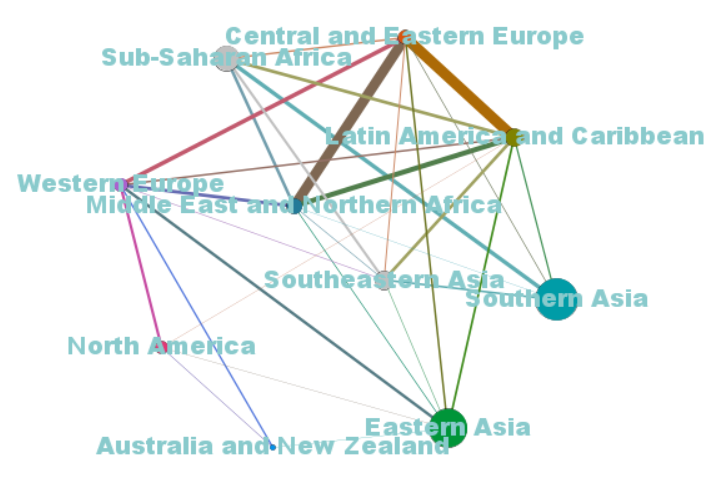


Figure 12. Region similarity Figure 13. Happiness & Freedom

On the other hand, we try to explore the network pertain to freedom score independently (Figure 13). Again, we can see a high correlation between happiness and freedom. Freedom can be affected by local social customs, traditional opinions and common values, political atmosphere, and religion constraints. For the countries at the similar economy level, the freedom score results in the difference of happiness score.

6.2 Patterns

Given the clustering result, we try to discover the patterns hidden behind each group of countries. Happy countries resemble each other, while unhappy ones have their problems, and we want to figure out the problems they are facing with. After construction of radar charts (Figure 14) using the average of each group, we can see clear variance and somewhat similarity between them.

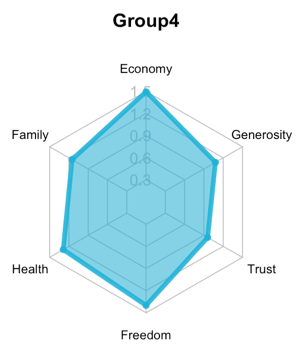
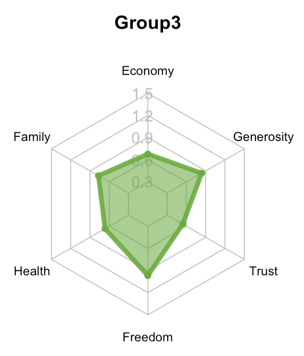
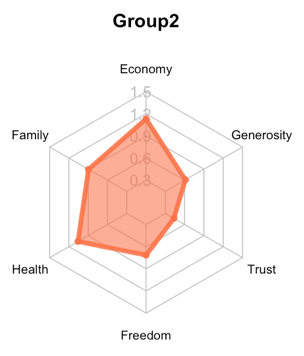
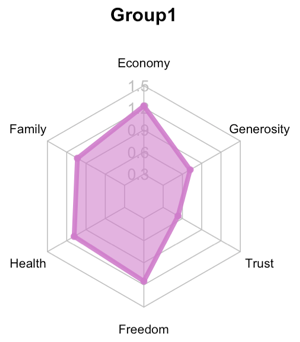


Figure 14. Radar charts of each cluster

On six features, we can see large span in economy level and freedom level. At the same time, some common states appear in other dimensions, such as high average family scores, close health score for most countries (except Group 3), low average score of trust in government, and slightly difference in generosity score (which stand for the strength of social support from government benefits, charity institution and community assistance).

On groups, more comparison and detailed analysis can be employed. We choose the most distinguished two groups – group 3 & group 4 – and take two specific countries as an illustration in next step[4].

6.3 Example of Group Four

The characteristic properties of group 4 include highest average GDP per capital, longest average life expectancy, high average freedom and generosity score, and good trust in government. Representative countries and corresponding regions are the United States in Northern America, Sweden in Western Europe, New Zealand in Australia and New Zealand.

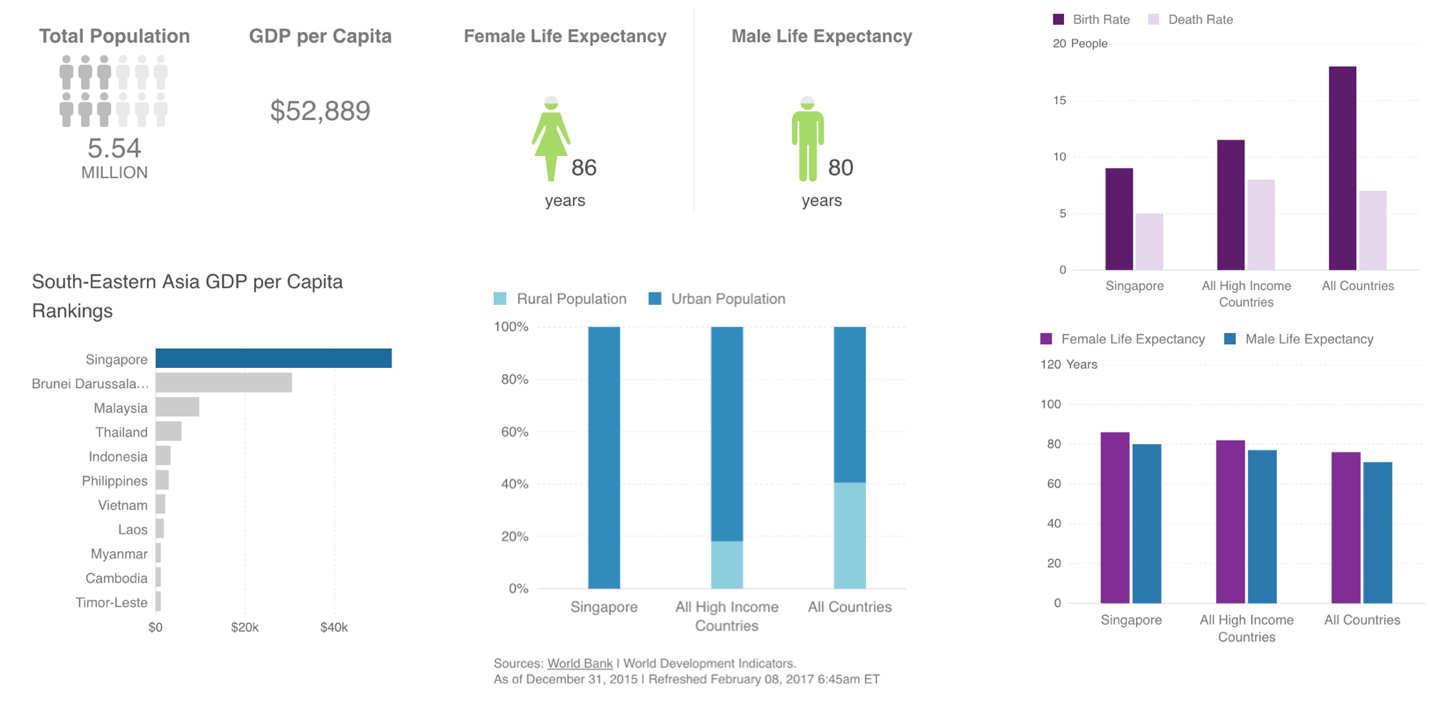


Figure 15. Statistics of Singapore

Here we choose Singapore as an example (Figure 15) because its most similar countries are U.S., Norway, and H.K., however, locates in Southeastern Asia instead of any representative regions. Singapore has a republic government, only 5.54 million population but ranks the first place in all 11 South-Eastern Asia countries regarding GDP per capita. It is almost entirely urbanized and benefits from the low unemployment rate. Residents in Singapore have a life expectancy of 86 years for female and 80 years for male, which are even longer than average level of all high-income countries. On the other hand, it also has very low birth rate and death rate, which are lower than the average of all high-income countries.

6.4 Example of Group three

The characteristic properties of group 3 include lowest average GDP per capital, shortest average life expectancy, high average freedom and generosity score, and bad trust in government. Representative countries and corresponding regions are Nepal in Southern Asia, Laos in Southeastern Asia, Sudan in Sub-Saharan Africa.

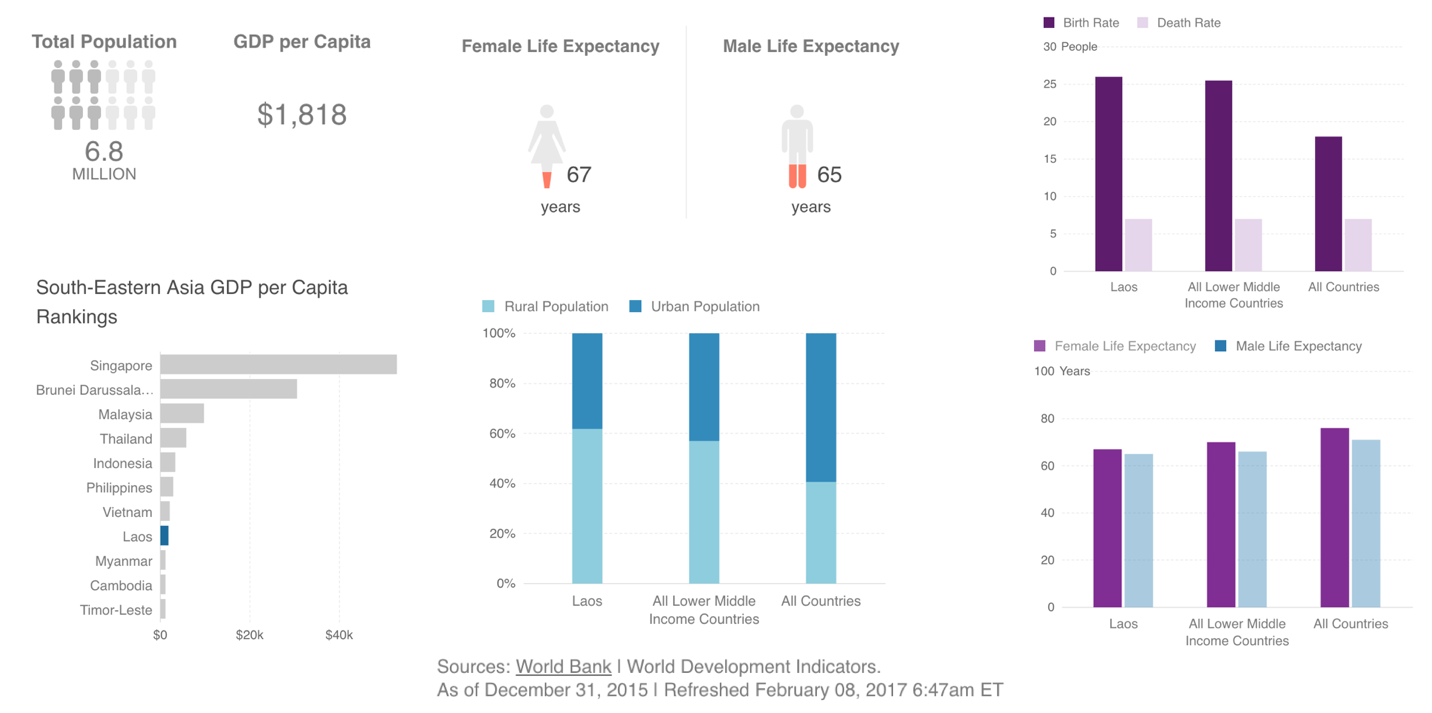


Figure 16. Statistics of Laos

We choose Laos as an example (Figure 15), which also locates in Southeastern Asia but is quietly different from Singapore. It is similar to Somalia and Myanmar, both developing countries. Laos is a communist country with 6.8 million population and ranks the 8th place in all 11 South-Eastern Asia countries regarding GDP per capita. It has a low GDP which however is growing rapidly, and has an extremely low unemployment rate. Its urbanization rate is lower than the average of all lower middle-income countries. Residents’ life expectancy is quite low – 67 years for female and 65 years for male, which are even shorter than average level of all lower middle-income countries. On the other hand, it also has much higher birth rate and almost the same death rate compared to the average of all countries.

The gap in almost all respects between these two example countries makes this comparison cruel, but it is also representative of the unbalance states between countries all over the world. We should take it as a reminder of our responsibility to eliminate this phenomenon for people who are still struggling to live a peaceful and meaningful life.

**7. Conclusions**

Now we can answer the questions we mentioned at the beginning. Through our community analysis, we get four clusters. The four clusters have apparently regional attribute. For six factors, economy(GDP Per Captial) has the most significant effect on happiness than Freedom and Health other factors.

When it comes to Group, we can find the distribution of each group on six factors was significantly different. In each cluster, there is some exception when taking geographical factors into account.

**8. Recommendations**

In this project, we just explored the connection between happiness and freedom. In a further step, we will think more about other factors like happiness and family, happiness and trust to see if there are other connections we haven’t find yet.

Another recommendation is about the community analysis of group. In this time we compared the representative country in the perspective of region, characteristic and national system. Next step we will taking more factors into accounts such as religion and social culture.

**9. References**

[1] World Happiness Report. In *Wikipedia.* Retrieved May 1st, 2017, from

<https://en.wikipedia.org/wiki/World_Happiness_Report>

[2] Hierarchical clustering. In *Wikipedia.* Retrieved May 1st, 2017, from

<https://en.wikipedia.org/wiki/Hierarchical_clustering>

[3] R graph manual pages. In *graph.* Retrieved May 1st, 2017, from

<http://igraph.org/r/doc/cluster_fast_greedy.html>

[4] Countries’ data and information. Retrieved May 2nd, 2017, from

<https://country-facts.findthedata.com/l/106/Singapore>

<https://country-facts.findthedata.com/l/101/Laos>