

# Visual Analytics Tool China Social Survey Data - Happiness

Song Jiashun, Su Yu, Xu Yang, Instructor: Tin Seong KAM

**Abstract**—Happiness is a very important topic all around the world nowadays, especially in China, the country which is becoming more and more focus on people's happiness because of the development. Happiness can be influenced by a lot of factors, such as health, education and income. Also, happiness can show different features in different regions. Our objective is to get the most important factors to happiness in China. Therefore, we will use several types of graphs such as likert scale, bubble plot and mapping to analyze the result of a survey on Chinese people's happiness in 2015 and consider different provinces in China.

**Index Terms**—China, happiness, factor.

## 1 INTRODUCTION

Nowadays, China is developing more and more rapidly, and people's life quality has been improved a lot. Since the basic material life of people can be generally fulfilled now, the next step is to improve people's spiritual life. For this reason, the country is becoming more and more focus on people's happiness.

Every a few years, China will make a survey on people's happiness, including a lot of questions of many different fields, such as health, education and income. Also, the survey will directly ask people about their score of happiness, based on their feelings.

Therefore, we think happiness is a very important topic in China, and we want to make a research about which factors are most important to happiness, using the most recent survey result in 2015.

## 2 EXPOSITION

### 2.1 Objective & Methodology

Our objective is to see the influence of different factors on happiness and get the most important factors.

In order to get the factors, we will use several types of graphs, including exploratory data analysis, multivariate matrix analysis, Likert scale, bubble plot, choropleth mapping and cluster analysis to analyse the happiness survey result in China in 2015 and show the results by different provinces.

We will consider the happiness score given by the respondents themselves as the main target and choose several factors which we think have large influence on happiness to see the importance of them. Besides, when choosing the question results which can represent factors, we will choose subjective questions first, because they can directly show the real feelings of people.

### 2.2 Subsection One

Our data come from the result of <Chinese General Social Survey 2015 Annually Survey> done by Renmin University of China, which is basically about people's life. It has many questions about nearly all fields. Because there are too many questions and results, we only choose the ones we need to analyse.

Except for the basic results such as the number of respondents, their gender and age, and the happiness scores, for the happiness factors, we mainly focus on health, depression, equity, class, peer status, income, relaxing, socializing and learning. All of them have subjective questions which can show people's real status and logic scores. Pay attention, in the survey, the higher the scores of peer status and income ability are, the worse the situation is.

In addition, Xinjiang, Xizang and Hainan does not have any data, so we will not consider them when making graphs.

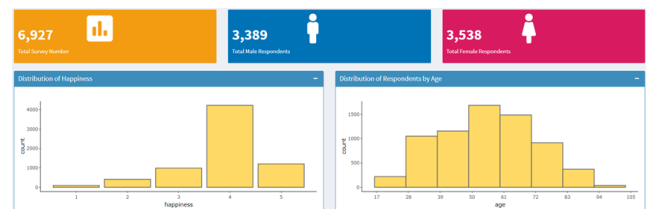
## 3 DISCUSSION

### 3.1 Basic Features of Respondents

As shown in the graph 3.1, our data has 6927 results. Among them, 3389 results are from males and 3538 results are from females.

Totally, the most people give their happiness scores as 4. Most people's happiness levels are between 3 and 5. It is a really good situation.

For all respondents, the age groups are between 17 and 105. Most people are between 28 years old and 94 years old. Namely, our data has nearly all age groups, so the data is very representative.



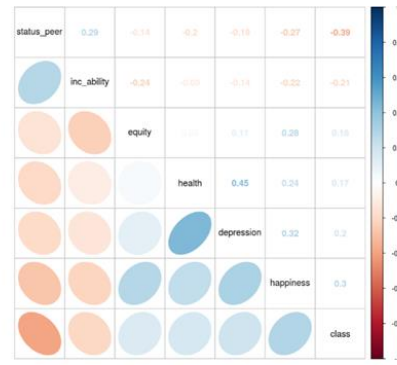
### 3.2 Distribution of Six Factors in Different Provinces

As shown in the graph 3.2.1, for the whole nation of China, the most people give health, depression and equity scores of 4, namely they think themselves are healthy, rarely feel depressed and think the society is fairer now. For the class, the most people think they are at the 5th level currently. For their socioeconomic status compared to peers, the most people give score of 2, namely think the status is almost the same. For the income ability, the most people give the score of 2, namely think their current income is reasonable. In a word, the general status of people's life is nice in China.

However, there are still many people who give negative results. In health field, many people think they are less healthy (score of 2), also in equity field, many people think the society is more unfair (score of 2). Besides, many people also think their socioeconomic status is lower than peers (score of 3) and their current income is unreasonable (score of 3). What's more, most people think they are at the 1st to 6th levels, which are not high.



(Graph 3.2.1)

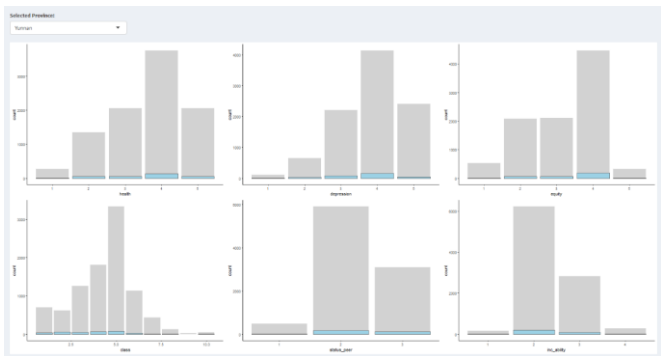


(Graph 3.3.1)

### 3.2.1 Compare Different Provinces



(Graph 3.2.2)



(Graph 3.2.3)

In developed regions such as Shanghai (Graph 3.2.2), much less people think they are less healthy, and most people think they are at the 2nd to 7th levels, which are higher than the average levels.

However, in some less developed regions such as Yunnan (Graph 3.2.3), the status is almost like the nation status, and more people think they at the 1st to 5th levels, which may be even worse.

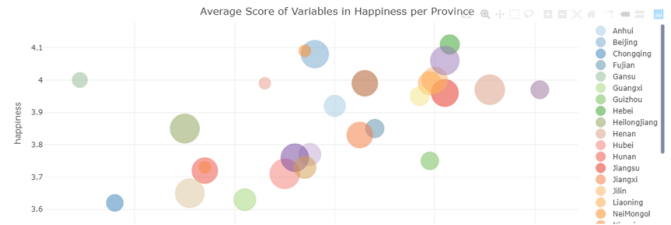
Therefore, in conclusion, the status between different regions may be very different because of the unbalanced development, and this difference will cause different happiness status.

### 3.3 Correlation between Happiness and Different Factors

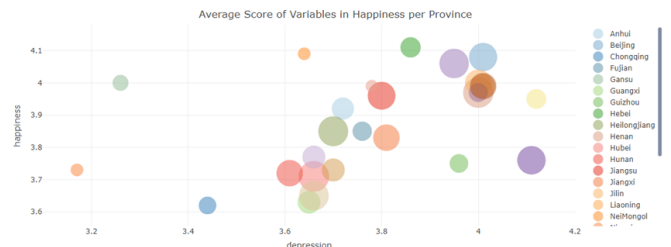
As shown in the graph 3.3.1, happiness has weak correlation with every factor, because the scores are all logic scores. For this reason, we cannot judge which factors are more important. However, we can see happiness has positive relationship with equity, health, depression and class, and has negative relationship with peer status and income ability, which is the same as the situation. We will do further research about the correlation between happiness and the other factors in the next part.

### 3.4 Further Research on Correlation

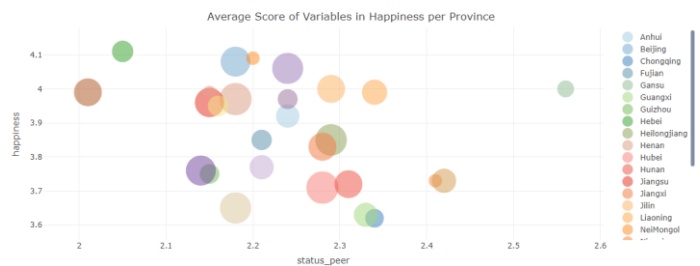
In this part, we use bubble plot to see the correlation between happiness and the other factors, using the average scores of all factors in different regions. We also add three new factors: relaxing, socializing and learning.



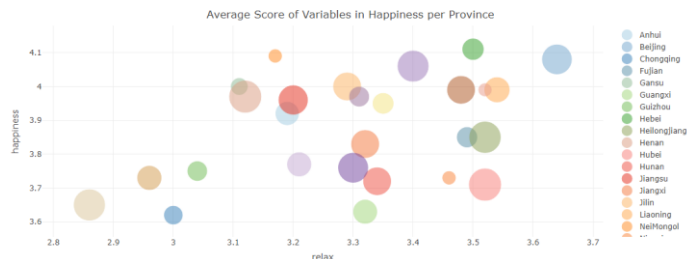
(Graph 3.4.1)



(Graph 3.4.2)

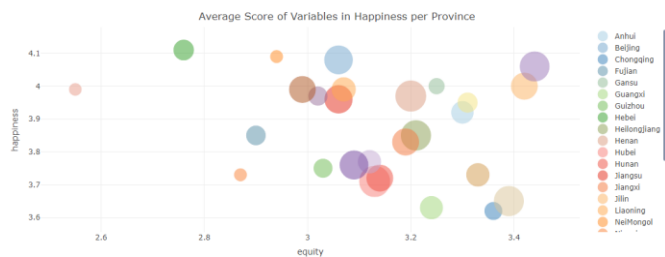


(Graph 3.4.3)



(Graph 3.4.4)

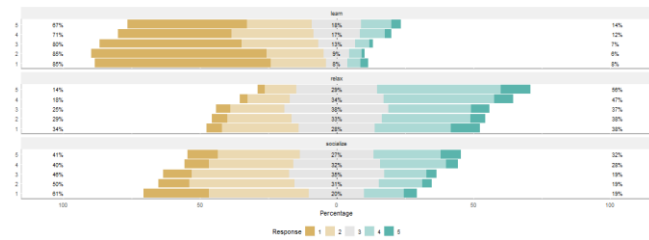
As shown in the graph 3.4.1, 3.4.2, 3.4.3, 3.4.4, health, depression, peer status and relaxing have stronger correlation with happiness. Namely, if one person is healthier, feels depressed more rarely, has better socioeconomic status compared to peers, and relax more often, he or she will feel happier.



(Graph 3.4.5)

For the other factors, such as equity shown in the graph 3.4.5, we cannot clearly see the trend, so we do not consider that they have strong correlation with happiness.

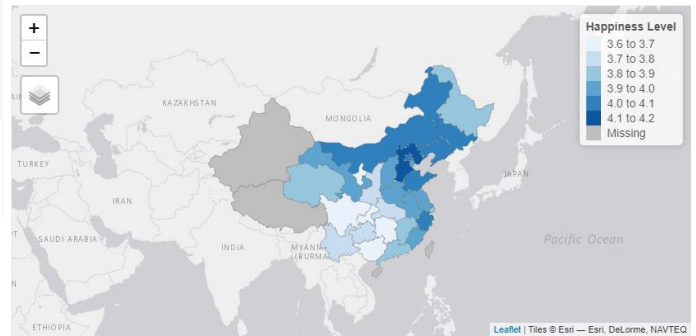
### 3.5 Distribution of Three Factors.



(Graph 3.5.1)

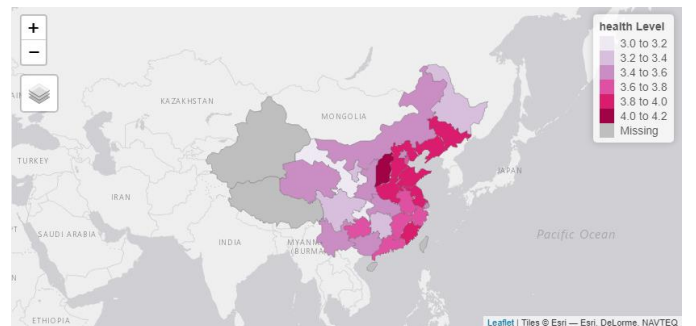
As shown in the graph 3.5.1, for learning, the most people think they never learn (score of 1), and for relaxing and socializing, the most people give the scores between 2 to 5, namely there are many people rarely, sometimes or often relax or socialize. Compare learning and socializing, people seem have more time to relax. Besides, the more learning, relaxing and socializing people have, the happier they are. This trend is most obvious in relaxing.

### 3.6 Geographical Distribution of Six Factors



(Graph 3.6.1)

As shown in the graph 3.6.1, people have higher happiness level in north China compared to the people in south China. Besides, the average happiness level of China is higher than 3.6, which is high, so the health level in China is nice. What's more, the happiness level of Hebei is the highest in China, then Nei Mongol, Jilin, Beijing and Shandong.



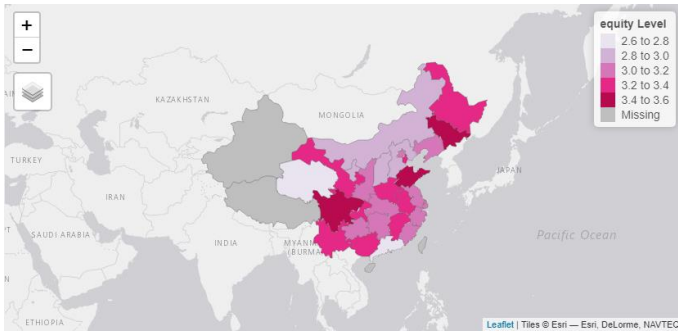
(Graph 3.6.2)

As shown in the graph 3.6.2, people in east China the highest health level. The average health level is higher than 3.0, which is also normal.



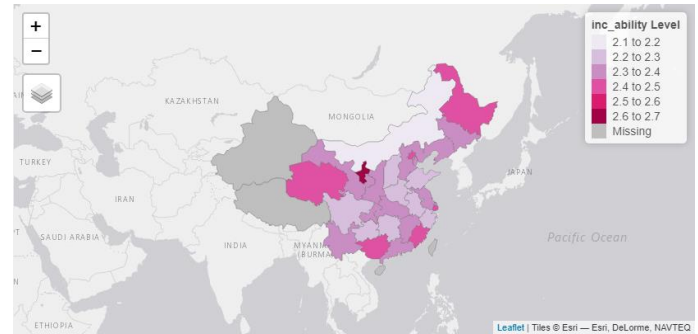
(Graph 3.6.3)

As shown in the graph 3.6.3, people in east China the highest depression level, namely the people there feel depressed most rarely, especially the people in Liaoning, Beijing, Tianjin, Shanxi, Henan, Shanghai and Zhejiang, which are all more developed regions compared to the other regions in China. Totally, the average depression level is also higher than 3.0, which is a normal level.



(Graph 3.6.4)

As shown in the graph 3.6.4, Shandong and Jilin seem have the highest equity level, then Heilongjiang and Southwest China. The total average level is higher than 2.6, which is still acceptable. Because of the fast but unbalanced development, the equity situation seems not good enough in China.



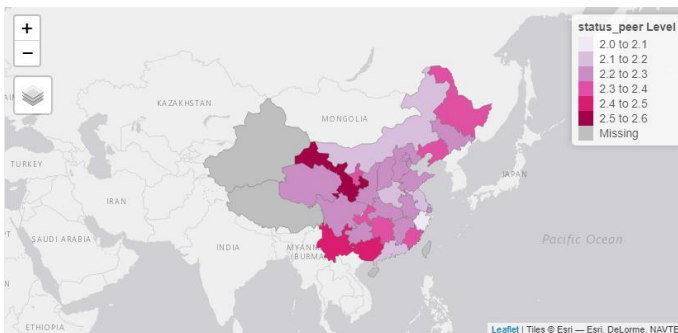
(Graph 3.6.7)

As shown in the graph 3.6.7, the income situation is the best in Shanxi and Guangdong, and worst in Nei Mongol and Gansu, namely the most people in Nei Mongol and Gansu think their incomes are not reasonable. Totally, the average level in China should be between 2.1 and 2.7, which is a bad situation that means more people think their incomes are unreasonable.



(Graph 3.6.5)

As shown in the graph 3.6.5, the average class levels the people think they are at in Nei Mongol, Jiangsu, Shanghai, Anhui and Zhejiang are between 4.6 and 4.8, the highest in China. The people in these regions must have good confidence, because of less pressure in life or better development. However, the total average level is between 3.5 and 5.5, and the top score is 10, so most people in China still do not have a good class level.

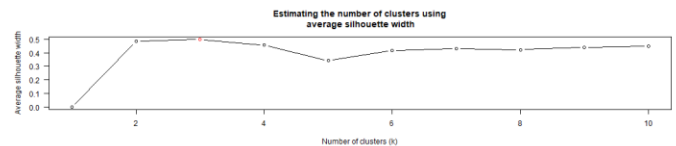
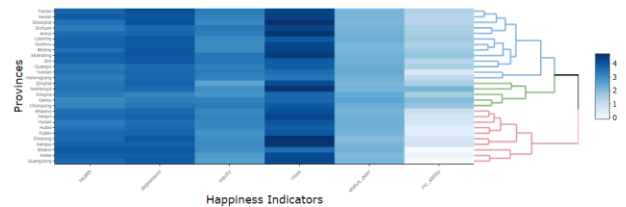


(Graph 3.6.6)

As shown in the graph 3.6.6, the situation of peer status is better in north, east and central China. The situations the worst in Gansu. Totally, the average level is between 2.0 and 2.6, which is high. This means most people think their socioeconomic status is lower than their peers. Also, the unbalanced development causes this situation, and most people do not have enough confidence on their socioeconomic status.

According to all the above graphs, there is no decisive factor on happiness, and the factors may influence happiness together. Hebei and Beijing have the highest happiness level, and nearly all situations of the other factors are good in these two regions. For the other regions, they all have some situations which are not good enough, thus the happiness levels are lower.

### 3.7 Geographical Distribution of Six Factors



(Graph 3.7.1)

As shown in the graph 3.7.1, we choose three clusters. According to the result, we consider the first cluster as developed regions, the second cluster as developing regions and the third cluster as less developed regions. In the developed regions, most situations of factors are good enough and in the developing factors, many situations of factors are not good. Also as shown in the above maps, the regions in cluster 1 have the higher happiness level than the regions in cluster 2 and 3, and the regions in cluster 2 have the lowest happiness levels. It is reasonable that if a region is more developed, the happiness level there should be higher.

## 4 CONCLUSION

According to the discussion part, all factors: health, depression, equity, class, peer status, income, relaxing, socializing and learning can to some degree influence happiness. In these factors, the most important ones are health, depression, peer status and relaxing

because they have clearer trend with happiness, namely we consider that they have stronger correlation with happiness.

Besides, we can divide China regions into three clusters: developed regions such as Beijing and Shanghai, less developed regions such as Zhejiang and Guangdong, and developing regions such as Nei Mongol and Gansu. Generally, those developed regions have the highest happiness levels and the best situations of factors, and those developing regions have the lowest happiness levels and the worst situations of factors.

Totally, the happiness level and situations of factors in China are good enough now, especially compared to a decade ago. China is developing really fast. If the government wants to improve the happiness level of people, it can focus on those less developed regions, then those developing regions. Some regions should develop first, then they can lead other regions to develop. Also, the government should first focus on people's physical and mental health and try to improve the income of people. As for relaxing, we think the government cannot do many things because the press comes from the fast developing economic and this development is very necessary. After improving these factors, the government should consider what to do for other factors. For example, it can distribute resources more fairly to improve the society equity.

## 5 FUTURE WORK

We will add more possible factors into our research, in order to make the research more effective. Based on more factors, we will do regression for happiness, so that we can see the relationship between happiness and those factors more clearly and be able to predict happiness of a region if given the scores of factors.

## ACKNOWLEDGMENTS

The authors wish to thank Prof Tin Seong Kam. Thanks to his patient guide from the beginning to the ending of our project and also many good advices for conducting the project and visualization of data.

## REFERENCES

- [1] [\\*https://worldhappiness.report/](https://worldhappiness.report/) World Happiness Report 2020]
- [2] [\\*https://shiny.rstudio.com/articles/layout-guide.html](https://shiny.rstudio.com/articles/layout-guide.html) Shiny Application layout guide Allaire]
- [3] [\\*https://shiny.rstudio.com/reference/shiny/latest/observeEvent.html](https://shiny.rstudio.com/reference/shiny/latest/observeEvent.html) Event handler R Shiny]
- [4] [\\*https://www.freecodecamp.org/news/build-your-first-web-app-dashboard-using-shiny-and-r-ec433c9f3f6c/](https://www.freecodecamp.org/news/build-your-first-web-app-dashboard-using-shiny-and-r-ec433c9f3f6c/) Build your first web app dashboard using Shiny and R]
- [5] [\\*https://db.rstudio.com/best-practices/dashboards/](https://db.rstudio.com/best-practices/dashboards/) Enterprise-ready dashboards]
- [6] [\\*http://rstudio-pubs-static.s3.amazonaws.com/5312\\_98fc1aba2d5740dd849a5ab797cc2c8d.html](http://rstudio-pubs-static.s3.amazonaws.com/5312_98fc1aba2d5740dd849a5ab797cc2c8d.html) ggplot2 Reference and Examples (Part 2) - Colors]
- [7] [\\*https://rpubs.com/tskam/likert](https://rpubs.com/tskam/likert) Hands-On Exercise 8: Diverging Stacked Bar Chart, Dr. Kam Tin Seong]
- [8] [\\*https://rpubs.com/tskam/heatmaps](https://rpubs.com/tskam/heatmaps) Hands-on Exercise 8: Heatmap Visualization with R, Dr. Kam Tin Seong]
- [9] [\\*https://rpubs.com/tskam/Corrgram](https://rpubs.com/tskam/Corrgram) Hands-On Exercise 8: Visualizing Correlation Matrix, Dr. Kam Tin Seong]
- [10] [\\*https://rpubs.com/tskam/Choropleth\\_Mapping](https://rpubs.com/tskam/Choropleth_Mapping) Hands-on Exercise 10: Choropleth Mapping with R, Dr. Kam Tin Seong]
- [11] [\\*https://rpubs.com/tskam/Proportional\\_Symbol\\_Map](https://rpubs.com/tskam/Proportional_Symbol_Map) Hands-on Exercise 10: Mapping Geospatial Point Data with R, Dr. Kam Tin Seong]