How to gain Happiness in Life: The Effect of Age and Self-evaluated Health Condition on Life Satisfaction

18/10/2020

Authors: Jiaheng Li(lijiahe5, 1003825088), Heye Liu(liuheye1, 1004864694), Yan Wang(wang1154, 1004990183), Yuxin Yang(yangyu77, 1004834673)

Abstract

This project focuses on how people's ages and their physical health conditions could affect their life happiness based on the Canadian General Social Survey. Statistical models such as linear regressions, boxplot, are used in the analyses. The results lead to the conclusion that both age and health conditions have influences on people's life happiness.

Introduction

With the rapid modern city developments, people have enjoyed a higher quality of life. More convenient transportations, communication equipment, and more diverse forms of entertainment are provided for individuals to gain pleasure and comfort. Meanwile, it seems that people also live a life with a faster pace than ever. Many people live a two-point life, having to find a balance of work and family. There is very little time left for relaxations. This can either directly or indirectly put many people's physical health in risk. A lot of people are starting to ask themselves: is it worth sacrificing one's physical health for living a more comfortable life? We are also part of the busy crowd and have been thinking about this question.

We would like to find the relationship between self-evaluated health conditions and life satisfaction. Given that the health condition may be age-related, we would also add age as a factor to our analysis. The Canadian General Social Survey recorded information about the basic living conditions of Canadian people including their age and self-rated health conditions, and how they feel about their current life. In the following part of the report we will introduce our selected data, the statistical models that help us visualize the relationship between the explanatory variables age and self-reported health condition (Xs) and the response variable, which is the score of "feelings about life as a whole". Eventually we will give a conclusion based on our analysis result, explaining how people's age and health conditions may bring whether positive or negative influences to their life satisfaction. More detailed information about the variables and models will be further explained in the sections after.

Data

Source

We chose to analyze the General social survey data on Family (cycle 31), 2017 from the U of T library, provided by Statistics Canada under the terms of the Data Liberation Initiative (DLI) license.

Target population

The target population of this survey includes all persons 15 years of age and older in Canada.

Frame

The survey used the redesigned GSS frame, integrating data from sources of both landline and cellular phone numbers available to Statistics Canada and the Address Register.

Smapling methodology and Sample

The sample population was obtained with stratification. Provinces and many Census Metropolitan Areas (CMAs) were considered as separate stratum, and each record in the survey frame was assigned to a stratum. Then a simple random sample without replacement of records was performed in each stratum. Any household with no telephone or at least one person 15 years of age or older was excluded from the survey. Respondents were then randomly selected from eligible households and participated in the survey through telephone interview. All the telephone interviewing was scheduled approximately 9:00 a.m. to 9:30 p.m. Mondays to Fridays, 10:00 a.m. to 5:00 p.m. on Saturdays and 1:00 p.m. to 9:00 p.m. on Sundays. All respondents who initially refused to participate were recontacted up to two more times to encourage their participation. The eventual responding rate was 52.4%.

Key features, strenths and weakness

The survey collected a large amount of data for each selected respondent as well as some information about each member of the respondent's household, including age, sex, family origins, conjugal history, the respondent's children, financial condition, labor market activity, subjective well-being and feelings about life as a whole.

In order to avoid non-thinking answers, this survey asked questions of the same topic in multiple ways. This methodology ensured the truethness of each reponse and eventually the accuracy of the data set. The large amount of data also enables statistician to explore various topic regarding the personal and fimily life situation of Canadians.

The survey excluded the households without telephone during the data collecting process, which might cause the samples to be misrepresentative for the most impoverished part of the Canadian families.

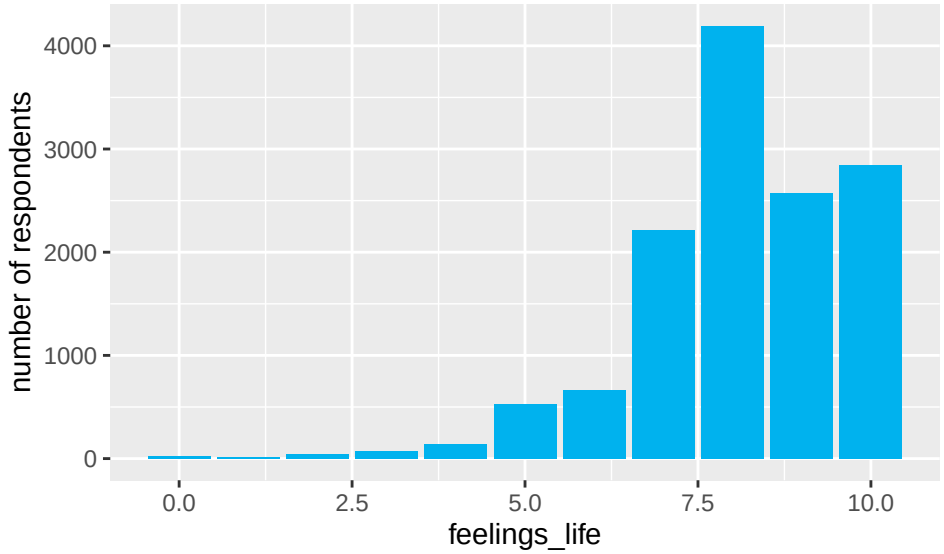
Selected variables

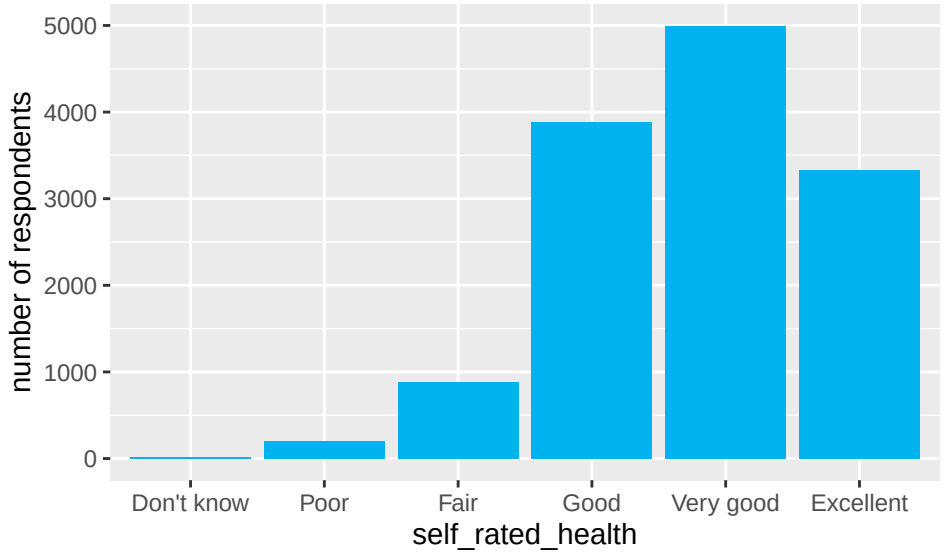
We downloaded the raw csv dataset for STATA, cleaned and prepared it in R with the gss_cleaning file provided ahead alongside the problem set question. The `gss_cleaning` file converts the raw dataset into a dictionary type dataset `gss.csv` with renamed variables. 20602 respondents in total and 81 variables were recorded in the data set `gss.csv`.

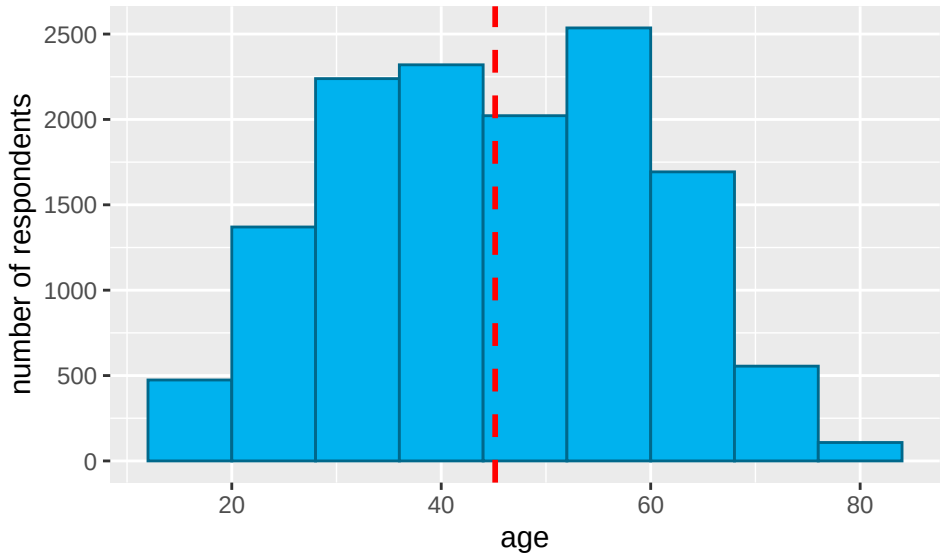
In order to analyze the effect of personal life situation on the life satisfaction score, we select the variables below:

- `age`: the age of respondent.
- `self_rated_health`: self rated health of respondent, ranging from excellent to poor, including five valid levels in total;
- `feelings_life`: feelings about life as a whole using a scale of 0 to 10, where 0 means "very dissatisfied" and 10 means "very satisfied".

We didn't pick the variable `self_rated_mental_health` since over 60% of the respondents reated themselves as either "excellent" or "very good", which made this variable less valuable than the self-rated health. We would also like to keep the model straightforward for the general audience to understand, so we select the self-rated health for further analysis.








```
summary(dataset$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   15.00   37.30   54.20   52.19   66.78   80.00
```

Models

In order to determine the relationship between the age of people, their self rated health and feelings to life, we use the variable "age", "self_rated_health", and "feelings_life". Since the "self_rated_health" is a categorical variable, we assume each state as a predictor observation to "feelings_life". Although age is collected as numeric variable, we separate the age to different interval, like age20-40, age40-60, age60-80, each interval is a predictor observation to "feelings_life". Our hypothesis is the health affects more to feelings of life compare to age. The model is:

$$feelings.life = \beta_0 + \beta_1 X_1 \ + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_a + \beta_7 X_b + \beta_8 X_c + \epsilon_i$$

$feelings.life$ represents the index from 0-10 collected from General social surveys year 2017.\
X_1:self rated health is excellent
X_2:self rated health is fair
X_3:self rated health is good
X_4:self rated health is poor
X_5:self rated health is very good
X_a:age interval from 20-40
X_b:age interval from 40-60
X_c:age interval from 60-80
ϵ_i:error term in the model

$$\beta_0 = 7.25805, \beta_1 = 1.52382, \beta_2 = -0.20491, \beta_3 = 0.52931, \beta_4 = -1.41613$$
$$\beta_5 = 1.03851, \beta_6 = -0.18829, \beta_7 = -0.03153, \beta_8 = 0.33590$$
$$\hat{feelings.life} = 7.25805 + 1.52382 X_1 \ + -0.20491 X_2 + 0.52931 X_3 - 1.41613 X_4$$
$$+1.03851 X_5 - 0.18829 X_a - 0.03153 X_b + 0.33590 X_c + \epsilon_i$$

When all predictors is zero, the model represents the mean feelings to life in the suvey year 2017.
When X_1 increase by one unit, the average value of feelings to life increase 1.52382, given other predictors hold constant.
When X_2 increase by one unit, the average value of feelings to life decrease 0.20491, given other predictors hold constant.
When X_3 increase by one unit, the average value of feelings to life increase 0.52931, given other predictors hold constant.
When X_4 increase by one unit, the average value of feelings to life decrease 1.41613, given other predictors hold constant.
When X_5 increase by one unit, the average value of feelings to life increase 1.03851, given other predictors hold constant.
When X_a increase by one unit, the average value of feelings to life decrease 0.18829, given other predictors hold constant.
When X_b increase by one unit, the average value of feelings to life decrease 0.03153, given other predictors hold constant.
When X_c increase by one unit, the average value of feelings to life increase 0.33590, given other predictors hold constant.

The standard error for each predictor is very small.

Next we use a p-value test to find the correlation between self rated_health_excellent and feelings to life. We assume the hypothesis test is there is no correlation between rated_health_excellent and feelings to life, and the alternative test is there is correlation.

$$H_0 : \beta_1 = 0, H_a : \beta_1 \neq 0$$

$$pvalue = 5.37e - 12$$

Use the significance level of 5%, the p-value is much smaller than the significance level, so we reject the hypothesis test. The p-value test support the alternative test that there is correlation between self_rated_health_excellent and feelings_life. We repeat testing the p-value for each predictor and intercept. The result supports X_1,X_3,X_4,X_5,X_a,X_c,and $\beta_0$has relation to feelings to life.

The r-squared in model1 is 0.163,it indicates the security does not generally follow the movements of the index.

```
## # A tibble: 9 x 7
##   term                estimate std_error statistic p_value lower_ci upper_ci
##   <chr>                  <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept              7.26      0.227    32.0     0        6.81     7.70
## 2 self_rated_health1    -1.42      0.226    -6.26    0       -1.86    -0.973
## 3 self_rated_health2    -0.205     0.222    -0.923   0.356   -0.64     0.23
## 4 self_rated_health3     0.529     0.22      2.40    0.016    0.097    0.961
## 5 self_rated_health4     1.04      0.22      4.71    0        0.607    1.47
## 6 self_rated_health5     1.52      0.221     6.9     0        1.09     1.96
## 7 age20-40              -0.188     0.06     -3.11    0.002   -0.307   -0.07
## 8 age40-60              -0.032     0.06     -0.527   0.598   -0.149    0.086
## 9 age60-80               0.336     0.06      5.63    0        0.219    0.453
```

Results

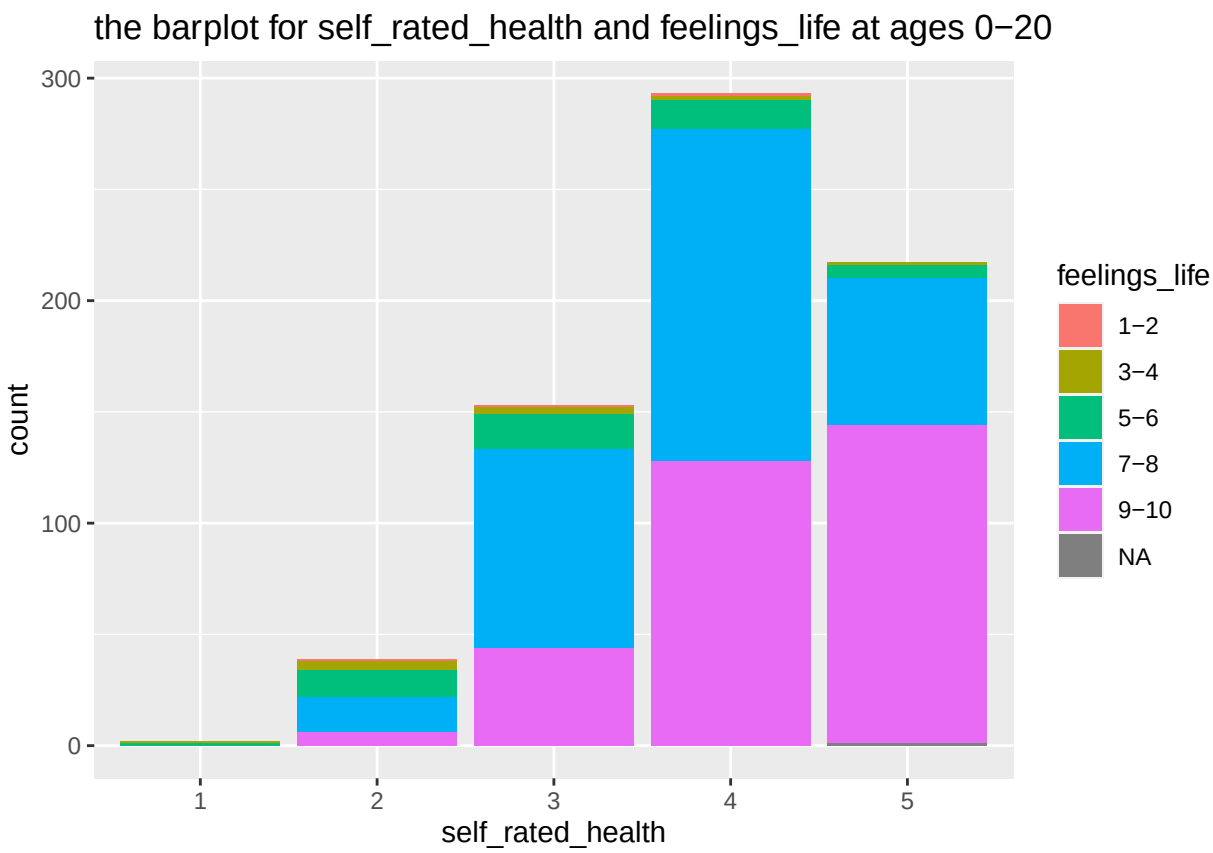
Firstly, we create a multiple linear model for feelings_life with self_rated_health vs ages.

***model*:**

```
## # A tibble: 9 x 7
##   term                estimate std_error statistic p_value lower_ci upper_ci
##   <chr>                  <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept              7.26      0.227    32.0     0        6.81     7.70
## 2 self_rated_health1    -1.42      0.226    -6.26    0       -1.86    -0.973
## 3 self_rated_health2    -0.205     0.222    -0.923   0.356   -0.64     0.23
## 4 self_rated_health3     0.529     0.22      2.40    0.016    0.097    0.961
## 5 self_rated_health4     1.04      0.22      4.71    0        0.607    1.47
## 6 self_rated_health5     1.52      0.221     6.9     0        1.09     1.96
## 7 age20-40              -0.188     0.06     -3.11    0.002   -0.307   -0.07
## 8 age40-60              -0.032     0.06     -0.527   0.598   -0.149    0.086
## 9 age60-80               0.336     0.06      5.63    0        0.219    0.453
```

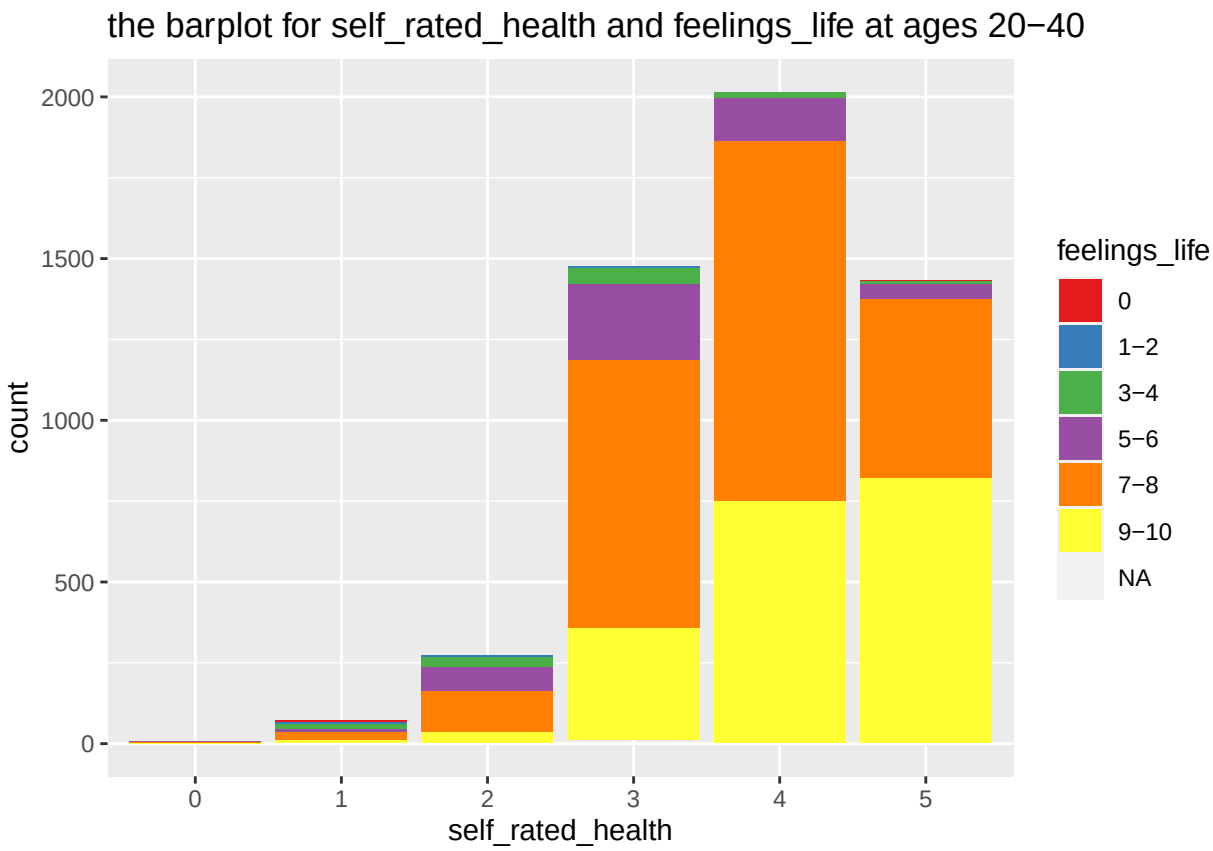
In addition, we make four different bar plots to reflects the impact of physical fitness on people's feeling_life at different ages. In these plots, we use use (1, 2, 3, 4, 5) to correspond to (poor, fair, good, verygood, excellent) in the variable self_rated_health. Moreover, The age groups of these four bar plots are 0-20, 20-40, 40-60 and 60-80 respectively.

***Figure1*:**



For Figure1, we can see that in the ages of 0-20, the more healthy people are, the greater the percentage of those with a happiness index of 9 to 10. From the graph, we can clearly observe that the healthier people are at this age, the happier most of them will be in life.

***Figure2*:**

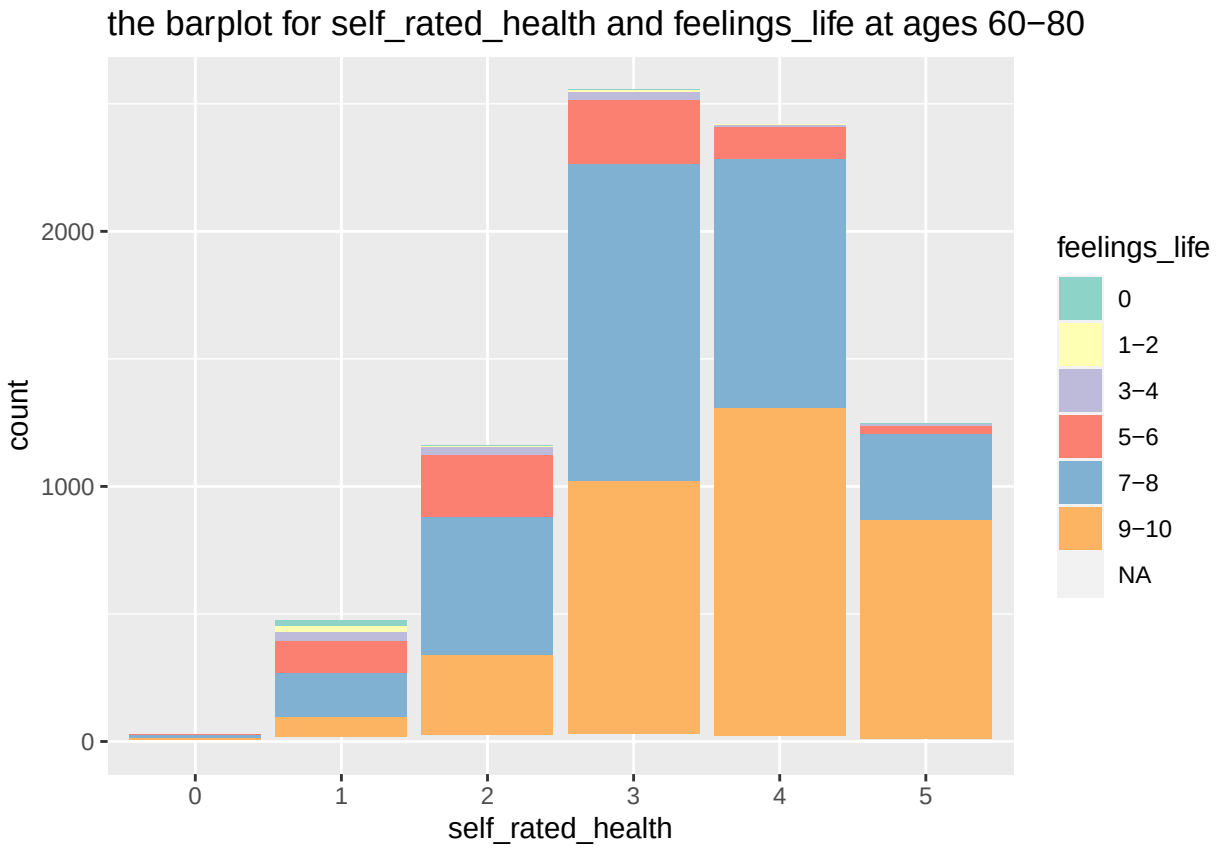


For figrue2, it shows that in the age group of 20-40, There were more participants in this age group than in the 0-20 age group. And when the health status of participants at this age is good, very good or excellent, The proportion of people with a life happiness index less than or equal to 4 is higher than those with the same health status in the age group of 0-20. The reason for this may be that most people in this age group are under a lot of pressure from work, society and family. These kinds of stress may affect their level of feeling life. However, it is the same as the 0-20 age group, at this age group, a person's health status is about good, the happiness index will be high.

***Figure3*:**

For figrue3, we cam make the same conclusion as the age group for 20-40, the healthier people are at this age, the happier most of them will be in life. The proportion of people with different levels of feeling life in different health states is basically the same as that of the 20-40 age group.

Figure4:



For figrue4, we also can get the conclusion that at this age group, a person's health status is about good, the happiness index will be high. There is a difference at this age group that among the people whose health status is "good", the proportion of people with a happiness index of 9-10 is higher than that of the previous three age groups. This difference may be caused by the difference in age. In the 60-80 age group, people's health is generally not as good as before, so they are not too nervous about the decline of their physical health. Moreover, most people in this age group have retired and do not have as much pressure as when they were young. So, the level of feeling life is a little higher than the first three ages.

Discussion

Firstly, from the models' section and results section, we can make a primary conclusion that People's physical health has a greater impact on people's level of feeling life than age has on people. When a person's physical health is good, he will be more likely to have a high level of feeling life. On the contrary, when a person's physical health is not good, it will be difficult for him to have a good level of feeling life. It just like an article named Study suggests health, not wealth, determines happiness, said that "Self-described healthy people are 20 percent happier than average, while "unhealthy" people are 8.25 percent less happy."[1]

The reason for this phenomenon is that physical unhealthy or illness will cause people's lives and psychology to be greatly affected."Physical changes from a disease may affect your appearance. These changes can turn a positive self-image into a poor one. When you don't feel good about yourself, you may prefer to be alone and withdraw from friends and social activities."[2]. Moreover, physical unhealthy and illness "can also influence your ability to work. Morning stiffness decreased range of motion, and other physical limitations may force you to change your work activities and environment. A decreased ability to work may also lead to financial problems."[2] Besides, physical unhealthy and illness also can cause people to have more stress. "Stress can build and can shape your feelings about life. Prolonged stress can lead to frustration, anger, hopelessness, and, at times, depression."[2] Therefore, When your body was not very healthy or sick, it would have a certain impact on your life, psychology, finances, work, etc. And often these factors are the main reference conditions for people to evaluate their level of feeling life.

However, different age groups will also have a little impact on people's level of feeling life. Among these participants, the higher the age, the proportion of their level of feeling life 9-10 is relatively higher than that of other age groups, even if their health status is not very good. In our opinion, the reason why has this kind of difference is that in the 60-80 age group, people's health is generally not as good as before, so they are not too nervous about the decline of their physical health. Moreover, most people in this age group have retired and do not have as much pressure as when they were young. At the same time, Most people already have enough savings at this age. There are children to take care of them, and lovely grandchildren often come to accompany them, so the impact of unhealthy health on them will be greatly reduced by these factors. The article named Why people get happier as they get older also has the same conclusion, "Although as people move towards old age they lose things they treasure — vitality, mental sharpness and looks — they also gain what people spend their lives pursuing: happiness."[3]

Moreover, we also discover that for people aged 0-20, the proportion of people whose level of feeling life is less than 4 is smaller than that of the other three age groups. We think this is because most people in this age group have not yet started work and bear the pressure from society and family. They only need to study and do what they are interested in and do not need to consider others. In the lives of these young people, there are not many factors that affect their level of feeling life, so the proportion of these young people whose level of feeling life is 9-10 is higher than that of the other three age groups.

In general, We conclude from this survey that age and self-health status will affect people's level of feeling life, but the self-health status will have a greater impact on people's level of feeling life. This is also in line with our expectations.

Weaknesses

There are some imperfections in the overall study which were hard to avoid. One important thing to notice is that the health condition we got from the survey was self-rated health condition, which might not be accurate real health condition. Also, people may tend to present themselves in a positive way in a society. This caused a response bias which may affect our results. There exist a lot of NA response in the survey data as well. Which makes our investigations have some blind spots.

Next Steps

We will update our dataset in the future, and strive to get more accurate data. For example, people's health data, we can apply to some government medical structures to get more accurate people's health data. Then, we will expand our sample population and strive to get a conclusion that is more representative of the public. We will try more other statistical models, such as Bayesian and hierarchical model, and find the best model for more detailed analysis

Reference

[1] Mims, C., 2020. Study Suggests Health, Not Wealth, Determines Happiness | Zdnet. [online] ZDNet. Available at: https://www.zdnet.com/article/study-suggests-health-not-wealth-determines-happiness.

[2] Cleveland Clinic. 2020. Living With Chronic Illness. [online] Available at: https://my.clevelandclinic.org/health/articles/4062-chronic-illness.

[3] Medium. 2020. Why People Get Happier As They Get Older. [online] Available at: https://medium.economist.com/why-people-get-happier-as-they-get-older-b5e412e471ed .

[4]RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/

[5]Yihui Xie andJ.J. Allaire and Garrett Grolemund (2018). R Markdown:The Definitive Guide. Chapman and Hall/CRC. URL https://bookdown.org/yihui/rmarkdown.

[6]Alexander, R., & Caetano, S. (2020, October 7). Gss_cleaning.R. Retrieved October 19, 2020

[7]H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

[8]Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

[9]Yihui Xie (2020). tinytex: Helper Functions to Install and Maintain 'TeX Live', and Compile 'LaTeX' Documents. R package version 0.21.