Factors affecting feelings of life

Brendan Kan, Hengguang Zhou, Linhao Liu

Oct 19, 2020

Code and data supporting this analysis is available at: https://github.com/HengguangZhou/sta304A2

Abstract

We analyzed how different fields give contributions to respondent's life satisfaction (feelings_life column in data). We applied logistic regression for five fields as input variables and tried to predict whether the respondent felt satisfied or not by checking whether feelings_life greater than or equal to 5. We found that religion and income are two main factors in evaluating life. At the same time, the education level does not contribute a lot. Also, people who live with a partner (no matter married or just living common-law) usually have a more positive feeling of their life.

Introduction

GSS(General Social Survey) data is obtained from 20602 respondents, including all non-institutionalized persons 15 years of age and older, living in the 10 provinces of Canada.

In this paper, we focus on five attributes in this detailed dataset: population center (living in urban or rural), marital status, income, education and religion participation. We choose them based on our following assumptions.

On the one hand, various research has claimed higher income more or less promotes people's feeling of happiness[1, 2]. People with higher salaries are able to buy expensive products such as cars, luxuries, new technology goods, and they are capable of enjoying high-quality services. A wealthy person has the ability to satisfy all his or her needs in time. Also, suppose you have a higher salary than someone else. In that case, it usually means you play a more important role in modern society, and it brings superiority complex and satisfaction.

On the other hand, religion can also bring a higher level of happiness [3]. Religious activities help the devouts build a broad social network. The more often you visit the church (religious participation), the more often you meet with priests and friends. Social relationships are also the main reason we believe why having a spouse or partner is important. Furthermore, spirituality gives meaning to life, enlightenment, inner peace and inspiration.

In our initial assumption, education is closely related to social status, income and range of knowledge. An educated person has a broader range of knowledge, can view things more thoroughly, and has a better understanding of what is going on. So they should be able to find problems in life more easily and solve them. However, the result contradicts our hypothesis.

Instead of using score 1 to 10, we want to use a binary indicator to show whether a respondent has a sense of happiness or not. So, we define feelings_life $>= 5$ as true and otherwise false. In this situation, logistic regression is the best model to use. [4] Logistic regression is suitable when the outcome of interest is binary.

In the **Data** section, we filtered needed attributes while keeping all respondents in the dataset. Also, we define feeling_binary as {true: feeling_life $>= 5$, false: feeling_life < 5}. In the **Model** section, we

explained what model we used in detail, and we presented the result in the **Result** section. **Discussion**, **Weakness** and **Next Steps** discussed what we found, what we can improve and what we can do for further investigation.

Data

The data we selected includes five different columns from the GSS 2017 data [5]. It consists of the following columns from the GSS data: the satisfaction of life for each respondent (feelings_life) as the dependent variable, and the marital status, annual income for the respondent(income_respondent), education of the respondent, and the frequency of the participation in religious activities (religious_participation) for the independent variables.

[6] Population: The population included all persons 15 years of age and older in Canada, excluding: 1. Residents of the Yukon, Northwest Territories, and Nunavut; and 2. Full-time residents of institutions. Frame: The survey frame was created using two different components: Lists of telephone numbers in use (both landline and cellular) available to Statistics Canada from various sources (telephone companies, Census of population, etc.); The Address Register (AR): List of all dwellings within the ten provinces. Each record in the survey frame was assigned to a stratum within its province. Sample: A simple random sample without replacement of records was next performed in each stratum.

The following is the first few lines of the data before the transformation:

```
##   feelings_life        marital_status  income_respondent religion_participation
## 1             8 Single, never married $25,000 to $49,999   Once or twice a year
## 2            10               Married  Less than $25,000             Don't know
## 3             8               Married $25,000 to $49,999   At least once a week
## 4            10               Married $50,000 to $74,999             Not at all
## 5             8   Living common-law   Less than $25,000             Not at all
## 6             9               Married  Less than $25,000             Not at all
##                                                    education
## 1 High school diploma or a high school equivalency certificate
## 2                             Trade certificate or diploma
## 3                Bachelor's degree (e.g. B.A., B.Sc., LL.B.)
## 4 High school diploma or a high school equivalency certificate
## 5  College, CEGEP or other non-university certificate or di...
## 6 High school diploma or a high school equivalency certificate
```

Since we decided to analyze using a logistic regression model, we first extracted the data we want from the GSS data. Then we transformed the feeling_life from a 1 to 10 numeric categorical variable into a binary variable, depending on whether the value of feelings_life is less than 5 or not, with less than 5 being unsatisfied with life and greater/equal to 5 being satisfied with life.

We think there are some potential drawbacks to this data. For example, we transformed the data for feelings_life from a 1 to 10 discrete numeric categorical variable into a binary variable based on whether the value is less than 5 or not. This might affect the result by amplifying the effect the predictor variables have on the final output since even if the respondents are only borderline satisfied, they would still be considered satisfied with life after the transformation.

Another one would be the fact that we only analyzed the relationship between the satisfaction for life and frequency of religion participation, but there are several more columns from the GSS 2017 data, such as religious importance and whether an individual respondent has an affiliation with one or more religion(s) or not. These drawbacks might potentially affect the results we obtained from performing the logistic regression.

Model

We make the assumption that the feeling is in linear relationships with the other factors in which we are interested, thus we use a logistics model to describe the relationship between the independent variable(feeling_binary) and other factors.

$$Y_{feeling} = \beta_0 + \beta_1 X_{education} + \beta_2 X_{religion} + \beta_3 X_{marital\_status} + \beta_4 X_{income}$$
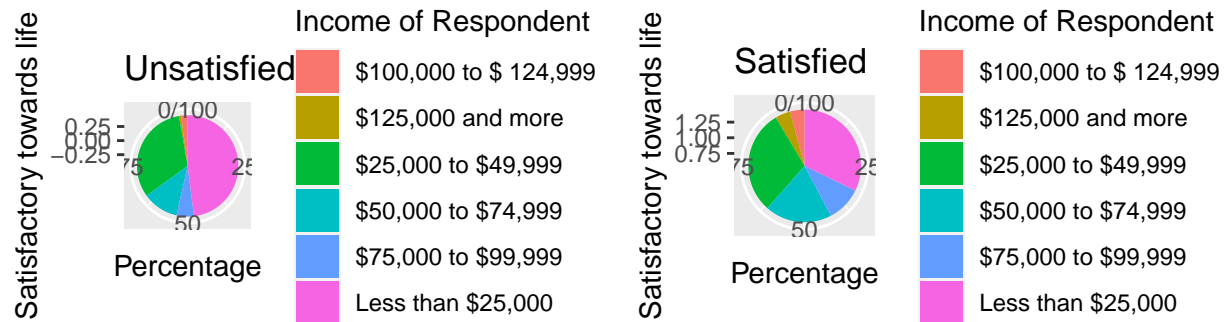
Here $Y_{feeling}$ represents the binary indicator of satisfaction for life. $X_{education}$ represents the level of education, $X_{religion}$ represents religion_participation, $X_{marital\_status}$ represents marital_status, and X_{income} represents the column income_respondent. We implement this equation in R using the survey-weighted generalised linear model svyglm() function. One alternative is to use multinomial logistic regression. However, we choose to use binary logistic regression because of the binary nature of our topic: satisfied or unsatisfied.

```
##                             GVIF Df GVIF^(1/(2*Df))
## education_              1.585641  6        1.039163
## religion_participation_ 1.237402  5        1.021530
## marital_status_         1.423647  5        1.035953
## income_respondent_      1.535173  5        1.043796
```
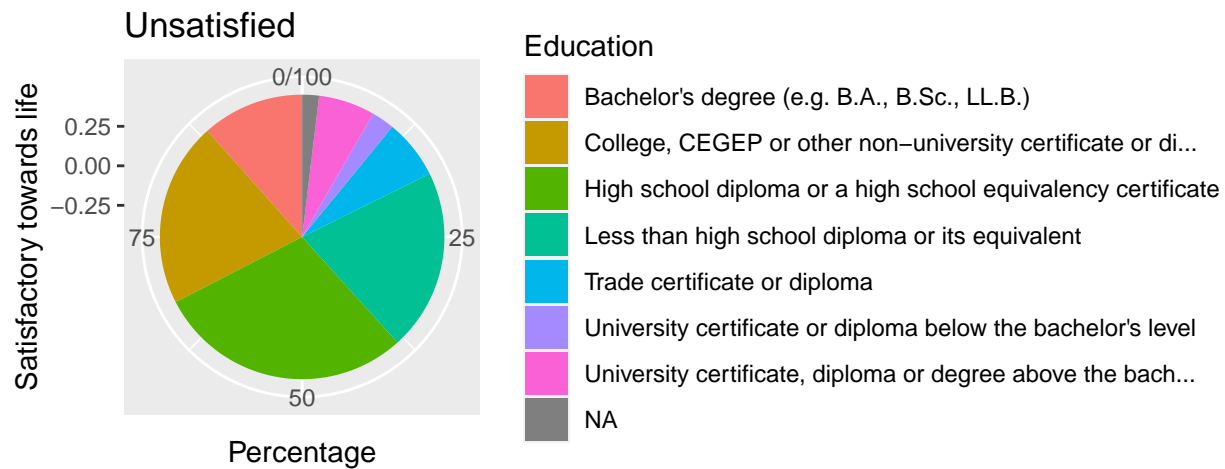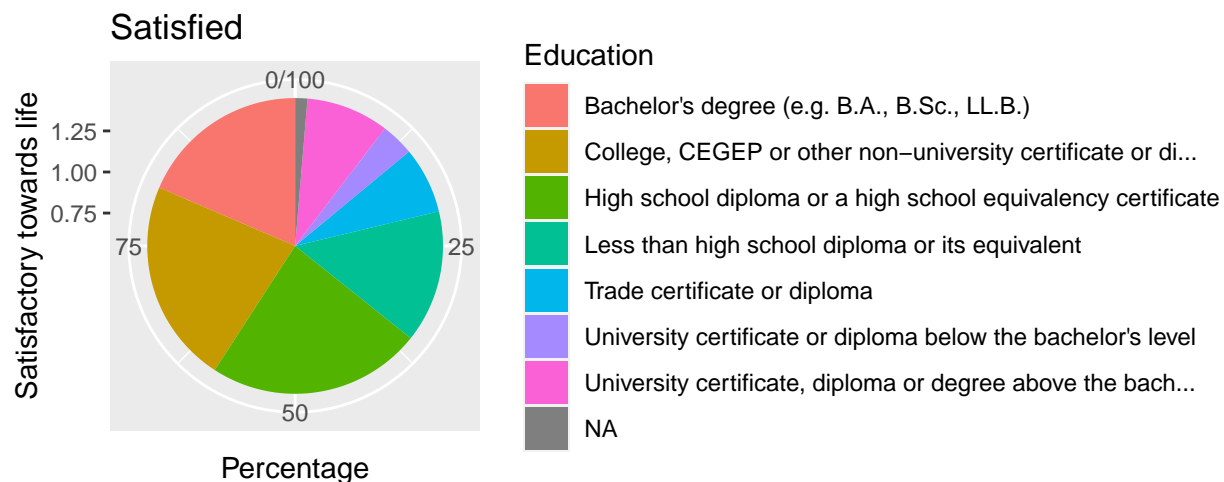
Results
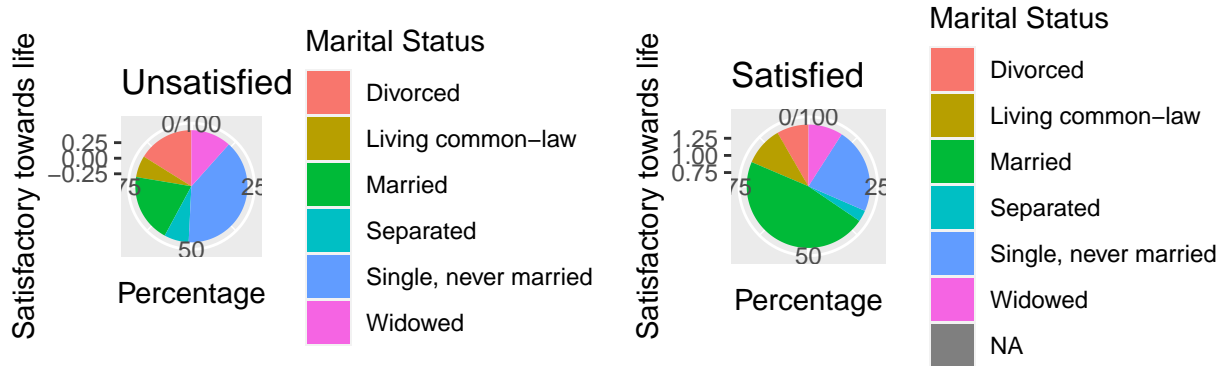
From the resulting pie chart, we can see that the percentage of people that are married and are satisfied with their life is higher than all other marital statuses. In comparison, the percentage of people that are single and unsatisfied with their life are higher than others.

Unsatisfied

Satisfactory towards life

0/100

0.25
0.00
−0.25

75

25

50

Percentage

Income of Respondent

- $100,000 to $ 124,999
- $125,000 and more
- $25,000 to $49,999
- $50,000 to $74,999
- $75,000 to $99,999
- Less than $25,000

Satisfied

Satisfactory towards life

0/100

1.25
1.00
0.75

75

25

50

Percentage

Income of Respondent

- $100,000 to $ 124,999
- $125,000 and more
- $25,000 to $49,999
- $50,000 to $74,999
- $75,000 to $99,999
- Less than $25,000

We can see that for the resulting pie chart, almost half the people that are unsatisfied with their life have an income of less than 25000 annually. For the people that are satisfied with their lives, we can see that the people who make less than 25000 are considerably less compared to the percentage of people that are unsatisfied with life. In contrast, the percentage of people who earn between 50000 to 74999 is considerably more.

Unsatisfied

Satisfactory towards life

Percentage

Education

- Bachelor's degree (e.g. B.A., B.Sc., LL.B.)
- College, CEGEP or other non–university certificate or di...
- High school diploma or a high school equivalency certificate
- Less than high school diploma or its equivalent
- Trade certificate or diploma
- University certificate or diploma below the bachelor's level
- University certificate, diploma or degree above the bach...
- NA

For the education of the people who are unsatisfied with their life, the majority of people have an education level of college, high school, or less than a high school diploma.

Satisfied

Education

- Bachelor's degree (e.g. B.A., B.Sc., LL.B.)
- College, CEGEP or other non–university certificate or di...
- High school diploma or a high school equivalency certificate
- Less than high school diploma or its equivalent
- Trade certificate or diploma
- University certificate or diploma below the bachelor's level
- University certificate, diploma or degree above the bach...
- NA

We can see that for the people who are satisfied with their lives, the majority of them also have a college, high school, or less than a high school diploma. However, one difference between this chart and the chart for the people that are unsatisfied is that the percentage of people who hold an education degree of either a bachelor or higher than bachelor is considerably higher within the group of people who are satisfied with their life.
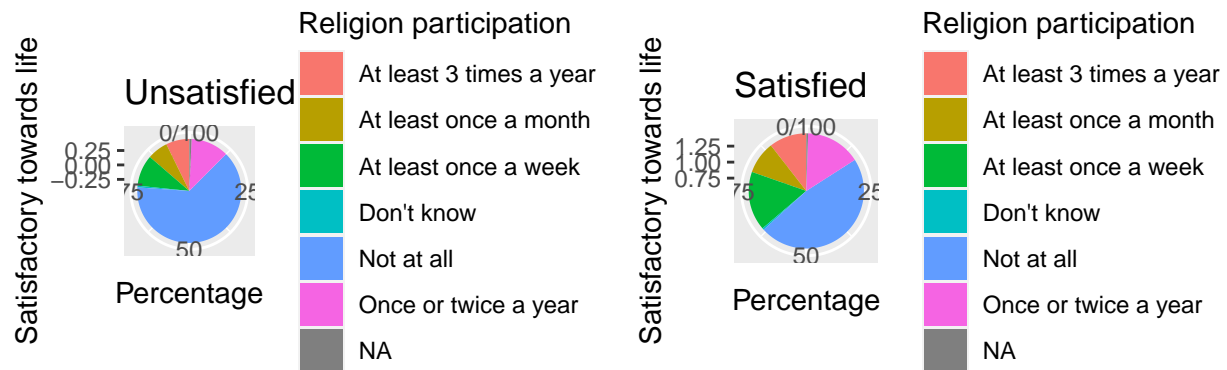
For the religious participation bar plot, we can see that the percentage of people who do not participate in a religion practice at all is considerably higher in the group of people who are unsatisfied with their life compared to the group of people who are satisfied with their life. Also, the percentage of people who practice religion at least a few times during the year is much higher in the group of people who are satisfied with their lives.

```
## # A tibble: 22 x 5
##    term                                 estimate std.error statistic  p.value
##    <chr>                                   <dbl>     <dbl>     <dbl>     <dbl>
##  1 (Intercept)                              4.03     0.367     11.0   7.06e-28
##  2 education_College, CEGEP or other non-~ -0.206    0.158     -1.31   1.91e- 1
##  3 education_High school diploma or a hig~ -0.273    0.154     -1.77   7.62e- 2
##  4 education_Less than high school diplom~ -0.292    0.169     -1.73   8.41e- 2
##  5 education_Trade certificate or diploma  -0.206    0.209     -0.990  3.22e- 1
##  6 education_University certificate or di~ -0.0926   0.292     -0.317  7.51e- 1
##  7 education_University certificate, dipl~ -0.283    0.208     -1.36   1.74e- 1
##  8 religion_participation_At least once a~  0.0373   0.232      0.161  8.72e- 1
##  9 religion_participation_At least once a~  0.152    0.211      0.722  4.71e- 1
## 10 religion_participation_Don't know       -0.442    0.621     -0.712  4.77e- 1
## 11 religion_participation_Not at all       -0.601    0.168     -3.57   3.54e- 4
## 12 religion_participation_Once or twice a~ -0.118    0.201     -0.588  5.56e- 1
## 13 marital_status_Living common-law         1.24     0.207      6.00   2.01e- 9
## 14 marital_status_Married                   1.40     0.146      9.57   1.21e-21
## 15 marital_status_Separated                -0.239    0.194     -1.23   2.20e- 1
## 16 marital_status_Single, never married     0.182    0.130      1.40   1.60e- 1
## 17 marital_status_Widowed                   0.363    0.172      2.11   3.46e- 2
## 18 income_respondent_$125,000 and more      1.27     0.655      1.93   5.32e- 2
```

```
## 19 income_respondent_$25,000 to $49,999      -0.669      0.315    -2.12  3.39e- 2
## 20 income_respondent_$50,000 to $74,999      -0.161      0.330    -0.489 6.25e- 1
## 21 income_respondent_$75,000 to $99,999      -0.182      0.354    -0.513 6.08e- 1
## 22 income_respondent_Less than $25,000       -0.878      0.316    -2.77  5.53e- 3
```

As we can see in the table above, using a 97.5% confidence interval, we found that the religion_participation, marital_status, and income_respondent are closely related to satisfaction. Notably, living common-law, married has a positive impact on happiness, while not participating in religion and having an income less than 25,000 would decrease satisfaction. On the other hand, we cannot find any significant evidence that supports the relationship between education and satisfaction.

Discussion

Surprisingly, we found that education level does not have much influence on one's satisfaction.

Another thing worth noting is that only extremes of these factors have a significant impact on satisfaction. By extremes, we mean "having income less than 25000" or "No religion participation at all." Generally speaking, we do not have significant support for the impact of having middle-range income or a moderate frequency in religion participation to satisfaction.

Weaknesses

One weakness of this dataset is that most of the attributes are categorical and thus difficult to analyze.

A possible weakness of our model is using satisfaction level as a binary variable oversimplifies the model and loses information. Simultaneously, we do not have a particularly effective approach for choosing the threshold value while it could be influential to the result.

Next Steps

It might be interesting to analyze the correlation between different factors in the future. A surprising result we got is that education is not related to satisfaction. From our perspective, education is likely to have a strong relation to income(or no?), but income is related to satisfaction while education is not. Studying the correlation between them could help us understand their relation to satisfaction better.

Furthermore, there are multiple attributes related to one field in the dataset. For example, income_family and income_respondent are related to income, religion_has_affiliation, religion_importance and religion_participation are all related to religion. We can further investigate these related attributes for more findings.

References

- 1.Daniel Kahneman and Angus Deaton. High income improves evaluation of life but not emotional well-being. Retrieved August 4, 2010, from https://www.pnas.org/content/pnas/107/38/16489.full.pdf
- 2.Ada Ferrer-i-Carbonell. Income and well-being: an empirical analysis of the comparison income effect. Retrieved from http://darp.lse.ac.uk/papersDB/Ferrer-i-Carbonell_%28JPubE05%29.pdf
- 3.: Religion's Relationship to Happiness, Civic Engagement and Health Around the World. Retreved from https://www.pewforum.org/2019/01/31/religions-relationship-to-happiness-civic-engagement-and-health-around-the-world/
- 4: STA304 - Logistic Regression Intro, Samantha-Jo Caetano. Retrieved from week 4 course material.
- 5: Statistics Canada. (2020, April). GSS 2017 Dataset, Retrieved from http://dc.chass.utoronto.ca/myaccess.html

- 6: Statistics Canada. (2020, April). GSS 2017: User Guide for the Public Use Microdata File (PUMF). Retrieved from https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/sdaweb/dli2/gss/gss31/gss31/more_doc/GSS31_User_Guide.pdf
- 7: Government of Canada, Statistics Canada. (2020, September 29). Population estimates, quarterly. Retrieved October 17, 2020, from https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000901
- 8: Table format code from https://piazza.com/class/kevvm6h64jo5hh?cid=255
- 9: Cowplot package in R. Retrieved from https://cran.r-project.org/web/packages/cowplot/index.html