A study of finding the factors that influence people's satisfaction with their life

Yujing Hua, Tingyi Li, Kaicheng Huang, Yehao Zhengs

Oct 19, 202e

Abstract

In nowadays, more and more people suffer from depression and lose confidence in life. According to the survey, people's life satisfaction is only 70 to 80 percent and this is the main reason why some people hurt themselves or even commit suicide. Thus, we wanted to look at what factors might affect their satisfaction with life. In this report, we create a multi-linear model to find the relationship between felling-live and age, number of children, gender, marital status, education, income, and number of hours they worked. From the model, we surprisedly found that high income will not influence people's feelings about their life, in contrast, marriage is the main factor that can cause depression for people.

Introduction

We aim to find the factors which are related to people's satisfaction with their life, and how significant they are. First we abstract data from GSS and select some variables we mentioned above to create a model and plots to find out how does each variable we chose influences feeling-live specifically. Thus, we pay most of our attention to the field of relationship between felling-live and age, the number of children, gender, marital status, education, income, and the number of hours they worked respectively, try to figure out the degree of association of each variable to feeling-live. Our final analysis will be used to predict what kind of group of people are likely to lose confidence in their life and bring depression to themselves.

Data

Data cleaning

Data overview

```
## Rows: 20,602
## Columns: 10
## $ caseid                   <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,...
## $ age                      <dbl> 52.7, 51.1, 63.6, 80.0, 28.0, 63.0, 58.8, ...
## $ total_children           <int> 1, 5, 5, 1, 0, 2, 2, 7, 0, 1, 0, 0, 4, 2, ...
## $ felling_live             <int> 8, 10, 8, 10, 8, 9, 4, 10, 8, 5, 10, 6, 6,...
## $ gender                   <chr> "Female", "Male", "Female", "Female", "Mal...
## $ marital_status           <chr> "Single, never married", "Married", "Marri...
## $ education                <chr> "High school diploma or a high school equi...
## $ self_rated_mental_health <chr> "Excellent", "Good", "Good", "Very good", ...
## $ income_respondent        <chr> "$25,000 to $49,999", "Less than $25,000",...
## $ average_hours_worked     <chr> "30.0 to 40.0 hours", "50.1 hours and more...
```

In the data, the target population is all persons 15 years of age and older in Canada contains residents of the Yukon, Northwest Territories, and Nunavut and full-time residents of institutions. (1) Besides, the survey

frame includes lists of telephone numbers in use (both landline and cellular) available to Statistics Canada from various sources (telephone companies, Census of population, etc.); and the Address Register (AR): List of all dwellings within the ten provinces.(2)The data was sampled by stratification such that divide ten provinces into strata, and then do the sampling in each stratum.

In this dataset, we have 20602 observations and all of them are collected by telephone interview, recorders use centralized telephone facilities in five of Statistics Canada's regional offices in Halifax, Sherbrooke, Sturgeon Falls, Winnipeg, and Edmonton from 9:00 a.m. to 9:30 p.m. on weekdays and I from 10:00 a.m. to 5:00 p.m. on Saturdays and 1:00 p.m. to 9:00 p.m. on Sundays. (3)The member of households is selected randomly, if the interviewer refused or cancel the census, they will reschedule the time. However, the response rate of 2017 GSS is only 52.4%, so they decide to consider the households of 'cell phone only ' as part of the frame.(4) Overall, the sample size we have in 2017 is 20000 but the number of interviewed people is 20602.(5)Investigators also use Address Register to the group together all telephone numbers associated with the same valid address. About 86% of available telephone numbers were linked to the AR.(6)

The method that investigators use to collect data is easier to get the data immediately, and it is clear for an investigator that which person they are interviewing now, therefore they can get accurate and useful data. However, there are some drawbacks such that it is time and money-waste, and this receives a low response rate due to refusing of phone call.

The data has 20602 rows and 10 columns and contains four numerical variables which are caseid, age, totoal_children, felling_live, and six categorical variables which are marital_status, Education, self_rated_mental_health, income_respondent, and average_hours_worked. In addition, there are some very similar variables we do not use in this dataset which are self-rated mental health number of marriages, ever marriage, and income of the family. We do not choose these variables because the income of response include family and mental is a part of health.

Model

```
##
## Call:
## lm(formula = gss$felling_live ~ gss$age + gss$total_children +
##     gss$gender + gss$marital_status + gss$education + gss$income_respondent +
##     gss$average_hours_worked)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.6226 -0.6841  0.1891  1.2992  2.9303
##
## Coefficients:
##                                                                          Estimate
## (Intercept)                                                             6.6788646
## gss$age                                                                 0.0026525
## gss$total_children                                                      0.0292512
## gss$genderMale                                                         -0.1623594
## gss$marital_statusLiving common-law                                     0.5927529
## gss$marital_statusMarried                                               0.7641697
## gss$marital_statusNA                                                    0.3952519
## gss$marital_statusSeparated                                            -0.3349420
## gss$marital_statusSingle, never married                                 0.1441055
## gss$marital_statusWidowed                                               0.2270798
## gss$educationCollege, CEGEP or other non-university certificate or di... -0.0205697
## gss$educationHigh school diploma or a high school equivalency certificate -0.0066000
## gss$educationLess than high school diploma or its equivalent            0.1419430
## gss$educationNA                                                         0.0895922
## gss$educationTrade certificate or diploma                              -0.0027984
```

```
## gss$educationUniversity certificate or diploma below the bachelor's level  0.0518088
## gss$educationUniversity certificate, diploma or degree above the bach...  -0.0263861
## gss$income_respondent$125,000 and more                                      0.1640057
## gss$income_respondent$25,000 to $49,999                                    -0.1629191
## gss$income_respondent$50,000 to $74,999                                    -0.0746242
## gss$income_respondent$75,000 to $99,999                                     0.0046952
## gss$income_respondentLess than $25,000                                     -0.3137045
## gss$average_hours_worked0.1 to 29.9 hours                                   1.0609860
## gss$average_hours_worked30.0 to 40.0 hours                                  0.9905347
## gss$average_hours_worked40.1 to 50.0 hours                                  1.0950564
## gss$average_hours_worked50.1 hours and more                                 1.1056277
## gss$average_hours_workedDon't know                                          0.9042458
## gss$average_hours_workedRefusal                                             1.1436991
## gss$average_hours_workedValid skip                                          0.9301295
##                                                                            Std. Error
## (Intercept)                                                                 0.3450232
## gss$age                                                                     0.0009225
## gss$total_children                                                          0.0091219
## gss$genderMale                                                              0.0240616
## gss$marital_statusLiving common-law                                         0.0537214
## gss$marital_statusMarried                                                   0.0422605
## gss$marital_statusNA                                                        0.6545152
## gss$marital_statusSeparated                                                 0.0747291
## gss$marital_statusSingle, never married                                     0.0499396
## gss$marital_statusWidowed                                                   0.0550158
## gss$educationCollege, CEGEP or other non-university certificate or di...    0.0358940
## gss$educationHigh school diploma or a high school equivalency certificate   0.0362421
## gss$educationLess than high school diploma or its equivalent                0.0427537
## gss$educationNA                                                             0.1006731
## gss$educationTrade certificate or diploma                                   0.0500851
## gss$educationUniversity certificate or diploma below the bachelor's level  0.0649946
## gss$educationUniversity certificate, diploma or degree above the bach...    0.0460253
## gss$income_respondent$125,000 and more                                      0.0774320
## gss$income_respondent$25,000 to $49,999                                     0.0603570
## gss$income_respondent$50,000 to $74,999                                     0.0613681
## gss$income_respondent$75,000 to $99,999                                     0.0656935
## gss$income_respondentLess than $25,000                                      0.0625210
## gss$average_hours_worked0.1 to 29.9 hours                                   0.3349463
## gss$average_hours_worked30.0 to 40.0 hours                                  0.3339195
## gss$average_hours_worked40.1 to 50.0 hours                                  0.3360340
## gss$average_hours_worked50.1 hours and more                                 0.3374596
## gss$average_hours_workedDon't know                                          0.3438981
## gss$average_hours_workedRefusal                                             0.4052760
## gss$average_hours_workedValid skip                                          0.3340654
##                                                                            t value
## (Intercept)                                                                 19.358
## gss$age                                                                      2.875
## gss$total_children                                                           3.207
## gss$genderMale                                                              -6.748
## gss$marital_statusLiving common-law                                         11.034
## gss$marital_statusMarried                                                   18.082
## gss$marital_statusNA                                                         0.604
## gss$marital_statusSeparated                                                 -4.482
## gss$marital_statusSingle, never married                                      2.886
```

```
## gss$marital_statusWidowed                                                 4.128
## gss$educationCollege, CEGEP or other non-university certificate or di...  -0.573
## gss$educationHigh school diploma or a high school equivalency certificate -0.182
## gss$educationLess than high school diploma or its equivalent               3.320
## gss$educationNA                                                            0.890
## gss$educationTrade certificate or diploma                                 -0.056
## gss$educationUniversity certificate or diploma below the bachelor's level  0.797
## gss$educationUniversity certificate, diploma or degree above the bach...  -0.573
## gss$income_respondent$125,000 and more                                     2.118
## gss$income_respondent$25,000 to $49,999                                   -2.699
## gss$income_respondent$50,000 to $74,999                                   -1.216
## gss$income_respondent$75,000 to $99,999                                    0.071
## gss$income_respondentLess than $25,000                                    -5.018
## gss$average_hours_worked0.1 to 29.9 hours                                  3.168
## gss$average_hours_worked30.0 to 40.0 hours                                 2.966
## gss$average_hours_worked40.1 to 50.0 hours                                 3.259
## gss$average_hours_worked50.1 hours and more                                3.276
## gss$average_hours_workedDon't know                                         2.629
## gss$average_hours_workedRefusal                                            2.822
## gss$average_hours_workedValid skip                                         2.784
##                                                                         Pr(>|t|)
## (Intercept)                                                              < 2e-16
## gss$age                                                                 0.004041
## gss$total_children                                                      0.001345
## gss$genderMale                                                          1.54e-11
## gss$marital_statusLiving common-law                                      < 2e-16
## gss$marital_statusMarried                                                < 2e-16
## gss$marital_statusNA                                                    0.545927
## gss$marital_statusSeparated                                             7.43e-06
## gss$marital_statusSingle, never married                                 0.003911
## gss$marital_statusWidowed                                               3.68e-05
## gss$educationCollege, CEGEP or other non-university certificate or di... 0.566604
## gss$educationHigh school diploma or a high school equivalency certificate 0.855499
## gss$educationLess than high school diploma or its equivalent            0.000902
## gss$educationNA                                                         0.373513
## gss$educationTrade certificate or diploma                               0.955443
## gss$educationUniversity certificate or diploma below the bachelor's level 0.425388
## gss$educationUniversity certificate, diploma or degree above the bach... 0.566451
## gss$income_respondent$125,000 and more                                  0.034182
## gss$income_respondent$25,000 to $49,999                                 0.006955
## gss$income_respondent$50,000 to $74,999                                 0.223996
## gss$income_respondent$75,000 to $99,999                                 0.943023
## gss$income_respondentLess than $25,000                                  5.28e-07
## gss$average_hours_worked0.1 to 29.9 hours                               0.001539
## gss$average_hours_worked30.0 to 40.0 hours                              0.003017
## gss$average_hours_worked40.1 to 50.0 hours                              0.001121
## gss$average_hours_worked50.1 hours and more                             0.001053
## gss$average_hours_workedDon't know                                      0.008560
## gss$average_hours_workedRefusal                                         0.004777
## gss$average_hours_workedValid skip                                      0.005370
##
## (Intercept)                                                             ***
## gss$age                                                                 **
## gss$total_children                                                      **
```

```
## gss$genderMale                                                          ***
## gss$marital_statusLiving common-law                                     ***
## gss$marital_statusMarried                                               ***
## gss$marital_statusNA
## gss$marital_statusSeparated                                             ***
## gss$marital_statusSingle, never married                                 **
## gss$marital_statusWidowed                                               ***
## gss$educationCollege, CEGEP or other non-university certificate or di...
## gss$educationHigh school diploma or a high school equivalency certificate
## gss$educationLess than high school diploma or its equivalent            ***
## gss$educationNA
## gss$educationTrade certificate or diploma
## gss$educationUniversity certificate or diploma below the bachelor's level
## gss$educationUniversity certificate, diploma or degree above the bach...
## gss$income_respondent$125,000 and more                                  *
## gss$income_respondent$25,000 to $49,999                                 **
## gss$income_respondent$50,000 to $74,999
## gss$income_respondent$75,000 to $99,999
## gss$income_respondentLess than $25,000                                  ***
## gss$average_hours_worked0.1 to 29.9 hours                               **
## gss$average_hours_worked30.0 to 40.0 hours                              **
## gss$average_hours_worked40.1 to 50.0 hours                              **
## gss$average_hours_worked50.1 hours and more                             **
## gss$average_hours_workedDon't know                                      **
## gss$average_hours_workedRefusal                                         **
## gss$average_hours_workedValid skip                                      **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.597 on 20283 degrees of freedom
##   (290 observations deleted due to missingness)
## Multiple R-squared:  0.05808,    Adjusted R-squared:  0.05678
## F-statistic: 44.67 on 28 and 20283 DF,  p-value: < 2.2e-16
```

In this section, we build a multi-linear model by using age, total_children, gender,marital_status, education,income_respondent, and average_hours_worked to predict felling_live. One thing that should be noticed is that all of the data are collected in the provinces in Canada, so what we conclude and predict from this model is only suitable and appropriate for people in Canada. Based on the model, we predict the value of felling_live by formula: $felling_live = \beta_0 + \beta_1 age + \beta_2 total_children + \beta_3 gender \beta_4 marital_status + \beta_5 education + \beta_6 income_respondent + \beta_7 average_hours_worked$, where beta_0 is the value of predicted felling_live when we set every other variable to be zero, in MLR $beta_1$ to $beta_7$ are slope coefficient for each explanatory variable, such that when we keep other variables constant, what would be the average change of felling live when one variable change for one unit.

In null hypothesis testing, the p-value is the probability of obtaining test results as results actually observed under the assumption that the null hypothesis is correct. In our dataset, the p-value indicates whether the variable is significant to the model. If the p-value is small ($p < 0.05$), it means the variable is significant to the model. In our case, age total_children, genderMale, arital_statusLiving common-law, marital_statusMarried, marital_statusSeparated, marital_statusSingle, never married, marital_statusWidowed, etc. While, the big p-value ($p > 0{,}05$) indicates that the variable does not significant to the model which shows that we can remove this variable, such as ducationHigh school diploma or a high school equivalency certificate, educationTrade certificate or diploma, income_respondent75,000 to 99,999. From these, we can notice that education and high income do not affect feelings about life.

R-square is a statistical measure that measures how many variables can be explained by the regression model.

Where 0 indicates that no variable can be explained by the model while 1 indicates that the outcome can be explained from the regression model. In our model, R-square is 0.05808 which means there are 5.808% variation that can be explained by the model.

check model

```
## [1] 76700.14
```

```
## [1] 76937.71
```

```
## [1] 0.05678009
```

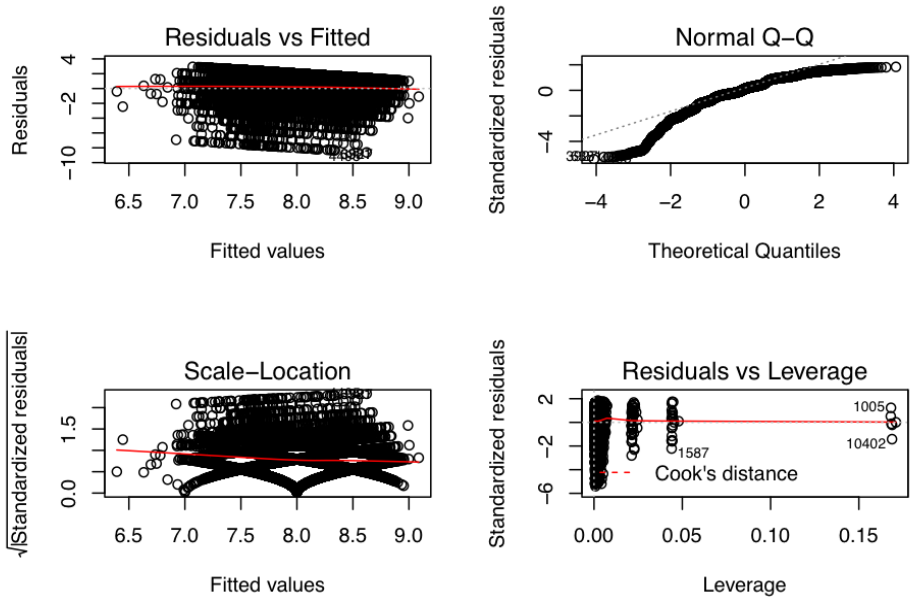
```
##    12   13   20   27   35   55   77   82  102  106  107  136  145
##    12   13   20   27   35   55   77   82  100  104  105  132  141
##   180  201  208  228  235  265  266  317  321  337  395  421  452
##   176  196  203  223  230  259  260  309  313  329  387  413  444
##   469  474  505  513  537  538  556  558  570  636  642  661  678
##   460  465  495  503  527  528  546  548  560  626  632  651  667
##   682  704  735  763  775  779  807  818  826  890  897  911  957
##   671  693  724  752  764  768  796  806  814  878  885  899  945
##   998 1004 1005 1007 1009 1037 1056 1076 1097 1122 1139 1159 1161
##   985  991  992  994  996 1023 1042 1062 1083 1107 1124 1144 1146
##  1198 1200 1223 1261 1281 1284 1289 1308 1325 1331 1351 1429 1444
##  1182 1184 1207 1245 1265 1268 1273 1292 1308 1314 1334 1411 1425
##  1461 1478 1494 1505 1513 1518 1547 1572 1585 1586 1587 1590 1592
##  1442 1459 1475 1486 1493 1498 1526 1550 1563 1564 1565 1568 1570
##  1594 1595 1610 1611 1616 1624 1643 1690 1695 1750 1792 1798 1821
##  1572 1573 1588 1589 1594 1602 1619 1665 1670 1724 1766 1772 1794
##  1822 1847 1914 1955 1965 1986 2004 2011 2017 2060 2100 2112 2146
##  1795 1820 1885 1924 1934 1955 1972 1979 1985 2028 2068 2079 2113
##  2169 2211 2242 2243 2296 2306 2327 2356 2385 2397 2404 2425 2436
##  2134 2176 2207 2208 2261 2271 2292 2321 2349 2361 2368 2388 2398
##  2458 2466 2495 2502 2509 2555 2594 2671 2678 2690 2727 2728 2732
##  2420 2428 2457 2464 2471 2515 2554 2629 2636 2647 2684 2685 2689
##  2752 2757 2771 2774 2777 2799 2837 2908 2944 3022 3059 3090 3116
##  2709 2714 2728 2731 2734 2754 2792 2862 2897 2975 3010 3041 3067
##  3185 3207 3225 3258 3271 3278 3280 3312 3324 3407 3425 3460 3462
##  3135 3157 3175 3208 3221 3228 3230 3262 3273 3356 3374 3408 3410
##  3471 3475 3503 3537 3543 3545 3599 3642 3678 3680 3699 3702 3705
##  3419 3423 3451 3485 3491 3493 3544 3587 3623 3625 3643 3646 3649
##  3722 3743 3746 3765 3801 3813 3838 3841 3863 3875 3891 3907 3953
##  3666 3687 3690 3709 3745 3757 3781 3784 3806 3818 3834 3850 3896
##  3967 3988 4001 4007 4019 4052 4090 4108 4131 4162 4176 4180 4197
##  3910 3931 3944 3950 3962 3994 4032 4050 4073 4103 4117 4121 4138
##  4233 4257 4258 4269 4316 4324 4334 4351 4441 4455 4525 4535 4561
##  4173 4196 4197 4208 4255 4263 4273 4290 4378 4392 4461 4471 4496
##  4577 4609 4627 4630 4665 4685 4707 4716 4742 4759 4765 4818 4820
##  4512 4543 4560 4563 4598 4618 4640 4649 4675 4691 4697 4750 4752
##  4841 4845 4851 4950 4971 5014 5106 5111 5115 5116 5126 5175 5208
##  4772 4776 4782 4880 4900 4943 5034 5039 5043 5044 5054 5103 5136
##  5246 5388 5423 5424 5429 5464 5478 5485 5529 5556 5557 5561 5562
##  5174 5311 5345 5346 5351 5386 5400 5406 5450 5477 5478 5482 5483
##  5581 5584 5605 5631 5638 5653 5682 5684 5685 5694 5716 5738 5745
##  5502 5505 5526 5552 5559 5574 5603 5605 5606 5615 5636 5658 5665
##  5764 5767 5781 5794 5821 5836 5844 5913 5923 5943 5999 6007 6021
```

```
##  5683  5686  5700  5713  5738  5753  5761  5830  5840  5860  5915  5923  5937
##  6036  6114  6199  6218  6225  6226  6312  6316  6364  6386  6389  6403  6407
##  5952  6028  6113  6132  6139  6140  6225  6229  6276  6297  6300  6314  6318
##  6435  6470  6492  6518  6520  6529  6571  6608  6647  6651  6659  6676  6737
##  6346  6379  6401  6427  6429  6438  6480  6516  6555  6559  6567  6584  6644
##  6757  6766  6790  6900  6938  7005  7012  7025  7078  7101  7128  7138  7241
##  6663  6672  6696  6805  6842  6909  6916  6929  6982  7004  7031  7041  7143
##  7263  7264  7277  7344  7373  7384  7394  7420  7423  7431  7468  7474  7484
##  7165  7166  7179  7246  7275  7286  7296  7322  7325  7333  7370  7376  7386
##  7488  7492  7493  7499  7507  7532  7545  7585  7596  7613  7682  7717  7731
##  7390  7394  7395  7401  7409  7434  7447  7486  7497  7514  7581  7616  7630
##  7740  7746  7755  7779  7805  7818  7822  7826  7842  7920  7938  7948  7951
##  7638  7644  7653  7677  7703  7716  7720  7724  7740  7815  7831  7841  7844
##  7964  7971  7991  8027  8036  8059  8084  8085  8129  8141  8149  8156  8166
##  7857  7864  7884  7920  7929  7952  7975  7976  8018  8030  8038  8045  8055
##  8169  8170  8171  8202  8210  8217  8278  8295  8328  8340  8442  8452  8463
##  8058  8059  8060  8091  8099  8106  8167  8184  8216  8228  8329  8339  8350
##  8470  8516  8523  8530  8575  8592  8627  8665  8748  8776  8784  8809  8875
##  8356  8401  8407  8414  8459  8476  8511  8548  8629  8655  8663  8688  8753
##  8916  8944  8950  8956  8961  9008  9087  9088  9137  9146  9204  9226  9296
##  8793  8820  8826  8832  8837  8882  8960  8961  9008  9017  9075  9097  9165
##  9322  9332  9410  9457  9502  9504  9508  9523  9546  9585  9591  9612  9614
##  9191  9201  9277  9322  9366  9368  9372  9387  9410  9447  9453  9474  9476
##  9626  9715  9755  9757  9827  9829  9860  9872  9893  9895  9910  9946  9947
##  9488  9574  9614  9616  9686  9688  9719  9730  9751  9753  9768  9804  9805
## 10020 10035 10040 10045 10073 10086 10139 10157 10174 10195 10233 10234 10249
##  9877  9891  9896  9901  9929  9942  9995 10013 10030 10051 10088 10089 10103
## 10254 10260 10262 10286 10345 10377 10385 10397 10402 10419 10429 10435 10469
## 10108 10114 10116 10139 10196 10228 10236 10248 10253 10270 10280 10286 10319
## 10477 10484 10545 10558 10570 10677 10679 10683 10684 10716 10744 10756 10771
## 10327 10334 10393 10405 10417 10523 10525 10529 10530 10562 10590 10602 10617
## 10783 10873 10920 10935 10965 10980 10983 10988 11014 11099 11136 11138 11151
## 10629 10717 10763 10777 10806 10821 10824 10829 10855 10938 10975 10977 10989
## 11161 11184 11240 11282 11294 11323 11335 11336 11406 11419 11453 11477 11532
## 10999 11022 11077 11119 11131 11159 11171 11172 11241 11254 11288 11312 11365
## 11540 11600 11617 11633 11640 11643 11659 11726 11737 11751 11809 11820 11835
## 11373 11431 11448 11464 11471 11474 11489 11555 11566 11580 11637 11648 11663
## 11876 11972 12002 12029 12035 12077 12088 12098 12106 12140 12199 12203 12225
## 11703 11797 11827 11853 11859 11901 11912 11922 11930 11964 12022 12026 12048
## 12244 12268 12290 12304 12351 12383 12412 12424 12446 12451 12453 12458 12486
## 12067 12091 12113 12127 12172 12203 12232 12244 12265 12270 12272 12277 12305
## 12516 12520 12562 12571 12576 12580 12583 12615 12662 12663 12722 12755 12791
## 12335 12339 12380 12389 12394 12398 12401 12433 12479 12480 12539 12572 12607
## 12834 12841 12878 12998 13002 13028 13037 13060 13088 13110 13139 13140 13165
## 12650 12657 12694 12813 12817 12842 12851 12874 12901 12922 12951 12952 12977
## 13175 13180 13190 13258 13266 13290 13306 13342 13359 13395 13423 13458 13481
## 12987 12992 13001 13068 13076 13100 13116 13151 13168 13203 13231 13266 13289
## 13494 13509 13542 13564 13599 13600 13606 13637 13659 13660 13674 13679 13753
## 13302 13317 13349 13371 13406 13407 13413 13444 13464 13465 13479 13484 13556
## 13826 13832 13856 13858 13881 13898 13900 13923 13925 13934 13944 14029 14041
## 13628 13634 13658 13660 13683 13700 13702 13725 13727 13736 13746 13830 13841
## 14055 14059 14070 14078 14085 14129 14142 14157 14187 14312 14324 14339 14386
## 13855 13859 13870 13878 13885 13929 13942 13956 13986 14110 14122 14137 14183
## 14402 14425 14430 14435 14453 14463 14473 14497 14504 14528 14529 14535 14544
```

```
## 14198 14221 14226 14230 14247 14257 14266 14289 14296 14320 14321 14327 14336
## 14582 14592 14593 14628 14631 14640 14675 14680 14708 14745 14799 14894 14913
## 14374 14383 14384 14419 14422 14430 14464 14469 14497 14533 14586 14680 14699
## 14936 14939 14941 14951 14959 14967 14980 15015 15028 15034 15038 15039 15061
## 14722 14725 14727 14737 14745 14753 14766 14801 14814 14820 14824 14825 14847
## 15075 15079 15085 15107 15120 15121 15130 15134 15138 15164 15167 15203 15251
## 14861 14865 14871 14893 14906 14907 14916 14919 14923 14949 14952 14988 15035
## 15258 15321 15322 15347 15391 15463 15480 15502 15517 15528 15531 15532 15559
## 15042 15104 15105 15130 15173 15244 15261 15283 15298 15309 15312 15313 15339
## 15586 15616 15634 15652 15656 15661 15676 15703 15742 15782 15860 15873 15879
## 15365 15394 15412 15428 15432 15437 15452 15478 15517 15557 15634 15647 15653
## 15941 15951 15957 15958 15976 15988 16003 16007 16066 16098 16101 16108 16154
## 15715 15725 15731 15732 15749 15761 15776 15780 15839 15871 15874 15881 15926
## 16170 16206 16215 16228 16231 16236 16273 16277 16285 16286 16298 16300 16317
## 15941 15977 15986 15998 16001 16006 16043 16047 16055 16056 16068 16070 16086
## 16337 16362 16372 16417 16423 16445 16455 16463 16472 16494 16500 16517 16525
## 16106 16131 16141 16186 16192 16214 16224 16232 16241 16263 16269 16285 16293
## 16551 16555 16558 16578 16584 16607 16613 16647 16659 16682 16707 16740 16766
## 16319 16323 16326 16345 16351 16374 16380 16413 16424 16447 16471 16504 16530
## 16775 16797 16835 16853 16925 16941 16975 17016 17027 17032 17037 17043 17070
## 16539 16561 16599 16617 16689 16705 16738 16776 16787 16792 16797 16803 16830
## 17093 17124 17130 17132 17134 17163 17168 17178 17248 17263 17296 17362 17381
## 16853 16884 16890 16892 16894 16922 16927 16937 17007 17022 17055 17120 17139
## 17395 17556 17646 17649 17659 17676 17712 17718 17755 17797 17885 17894 17928
## 17153 17309 17398 17401 17410 17427 17461 17467 17504 17544 17631 17640 17674
## 17940 17966 17977 17997 17998 18020 18031 18039 18041 18061 18083 18087 18091
## 17686 17712 17723 17743 17744 17766 17777 17784 17785 17805 17827 17830 17834
## 18110 18130 18141 18150 18153 18154 18184 18208 18217 18220 18235 18251 18256
## 17852 17871 17881 17889 17892 17893 17923 17946 17955 17958 17973 17989 17994
## 18261 18277 18283 18287 18290 18333 18341 18352 18357 18358 18379 18383 18404
## 17999 18014 18020 18024 18027 18070 18078 18089 18094 18095 18116 18120 18141
## 18437 18459 18462 18471 18511 18617 18624 18640 18693 18720 18757 18802 18813
## 18174 18196 18199 18208 18246 18350 18357 18373 18426 18453 18490 18535 18546
## 18815 18827 18832 18841 18884 18900 18948 18990 19034 19053 19065 19113 19137
## 18548 18558 18563 18572 18615 18631 18678 18720 18764 18783 18795 18842 18865
## 19139 19152 19167 19169 19202 19205 19206 19242 19259 19282 19286 19288 19293
## 18867 18879 18894 18896 18929 18932 18933 18968 18985 19008 19012 19014 19019
## 19301 19372 19388 19431 19523 19533 19563 19568 19606 19609 19648 19682 19730
## 19027 19096 19112 19154 19246 19256 19286 19291 19329 19332 19370 19404 19451
## 19756 19777 19783 19822 19882 19891 19945 19973 19994 20030 20057 20079 20114
## 19477 19498 19504 19543 19602 19611 19665 19693 19713 19747 19773 19794 19829
## 20127 20153 20173 20183 20186 20192 20205 20230 20234 20241 20248 20253 20257
## 19841 19867 19886 19896 19899 19905 19918 19943 19947 19954 19961 19966 19970
## 20259 20300 20315 20331 20338 20339 20358 20428 20429 20436 20507 20514 20520
## 19972 20013 20028 20043 20050 20051 20069 20139 20140 20147 20218 20225 20230
## 20540 20559
## 20250 20269

##    99   139   173   277   327   337   515   795   824   827  1342  1357  1545
##    97   135   169   271   319   329   505   784   812   815  1325  1340  1524
##  1602  1930  2415  2438  2453  2585  2660  2966  3677  3927  4416  4495  4544
##  1580  1900  2379  2400  2415  2545  2618  2919  3622  3870  4353  4432  4480
##  4588  5186  5343  5457  5481  5510  5529  6085  6184  6292  6381  6444  6505
##  4523  5114  5269  5379  5403  5431  5450  5999  6098  6206  6292  6355  6414
```

```
##  7608  7755  7812  7829  7949  7980  8150  8297  8487  8508  8531  8652  8700
##  7509  7653  7710  7727  7842  7873  8039  8186  8373  8393  8415  8536  8582
##  8780  8894  9176  9372  9378  9483  9904 10162 10201 10498 10517 10940 11076
##  8659  8771  9047  9240  9246  9347  9762 10018 10057 10347 10366 10781 10916
## 11487 11639 11848 11974 12095 12153 12233 12602 12907 12916 12996 13152 13571
## 11322 11470 11676 11799 11919 11977 12056 12420 12723 12732 12811 12964 13378
## 13784 13811 13833 14433 14944 15079 15218 15251 15532 15939 16083 16581 16821
## 13587 13614 13635 14229 14730 14865 15003 15035 15313 15713 15856 16348 16585
## 16954 17124 17445 17457 17486 17684 17804 18081 18203 18625 18851 19235 19852
## 16718 16884 17201 17213 17242 17435 17551 17825 17941 18358 18582 18961 19572
## 19909 20487
## 19629 20198
```



AIC is an constant estimate measures relative distance between the unknown likelihood function and the fitting model likelihood function. Thus, a lower AIC means a model is considered to be closer to the truth. In our model, AIC is 76700.14. The lower the BIC, the more likely the model is to be a true model, because the BIC is an estimate of a function of posterior probabilities that the model is true in a particular Bayesian setting.In our model, the value of BIC is 76937.71. Also, there are some leverage points and outliers of our model. In addition, since the VIF of each variable is lower than 5, our model does not have multicollinearity.

Results

Based on the model, we can find the relationship between felling_live and all of the other variables. For example, we can see age, total_children, and gender_male all have a positive beta such that they all have a positive relationship for felling life which means no matter how old they are, how many children they have and being a male will not bring a negative influence for their life. Also, we found something surprising which is not the same as what we usually think of in our life. We noticed that people with less than a high school diploma and people who have a university certificate diploma below the bachelor's level has a positive relationship while people who went to university and has a higher diploma has a negative relationship with felling live. We may conclude that people who have higher education level is easier to lose confidence in their

life compared to those who have a lower education level. In addition, we also found something that matches our real-life assumption, for an instant, people who have high-income feels more satisfied with their life than these who have low salaries and people who are married have a positive felling about their life compare to those who are separated.

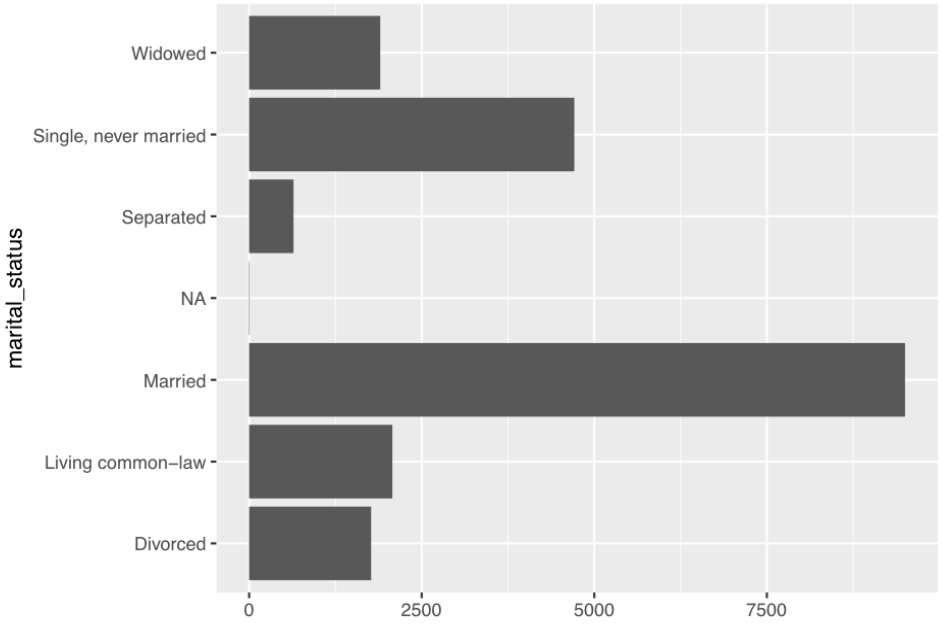
To summary our model, we found that variables of genderMale, marital_statusLiving common Law, marital_statusMarries,marital_statusSeparate,marital_statusWindowed, education than high school diploma or its equivalent and income _respondentLess than $25000 are the most significant factors that influence the felling live while income_respondent$125000 and more is the least significant one.

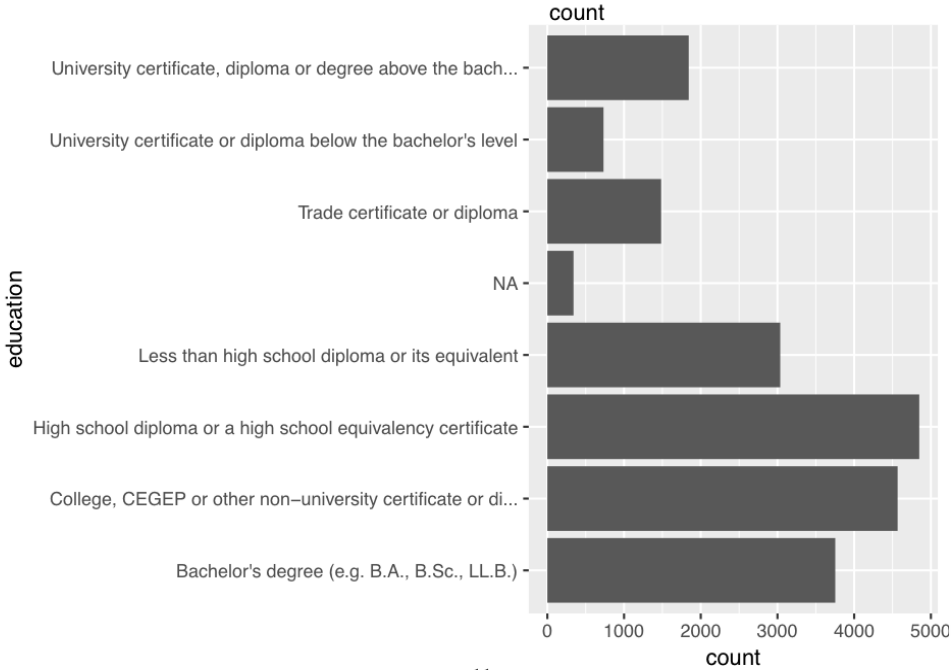
We also conclude some aspects about our model, from figure 1 which is a residuals plot that is used to visually assess wether our model fit the data, we can there is no pattern for the points, so we can see we have a good model that fit the data we have. From figure 2, we can see that this is not the normality of the errors such that we are not following the assumption of running a linear model and this assumption is not valid for our model inference. In addition, we can finger out whether it is a constant variance from Figure 3. We notice that it has a horizontal line and no pattern. Thus, we assume that error has constant variance. Our assumption is valid and model inference should also be valid. In addition, From Figure 4, We should pay more attention to the influential points which are outside to cook's distance in order to get a better result, because these points have a large effect on the parameter estimates.

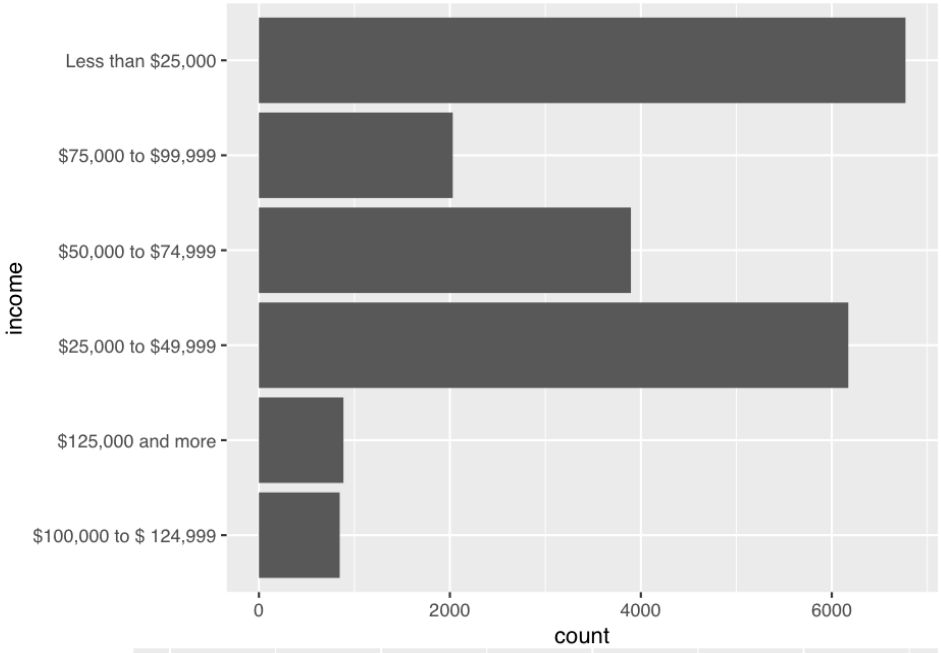
Note: figure 1 is a residuals plot, figure 2 is Normal Q-Q plot, figure 3 is Scale-Location plot, figure 4 is Residuals vs. Leverage plot.

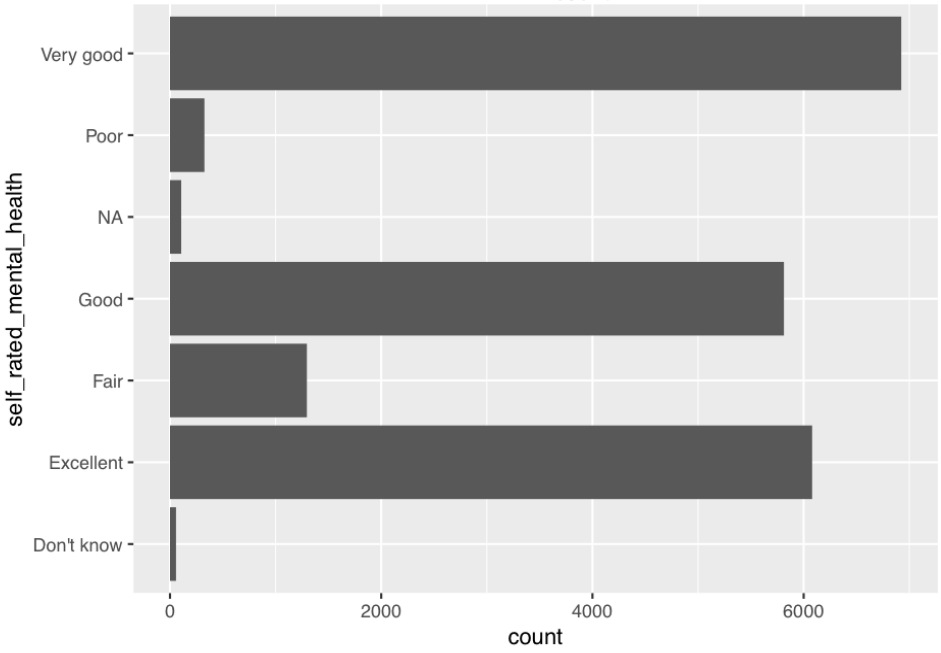
Discussion

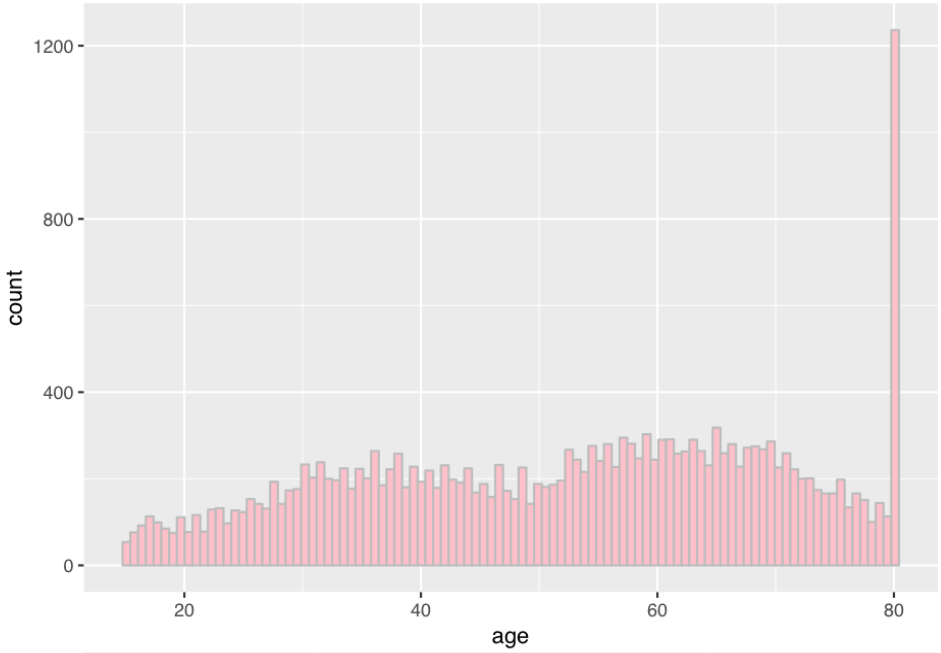
Plots of each variable

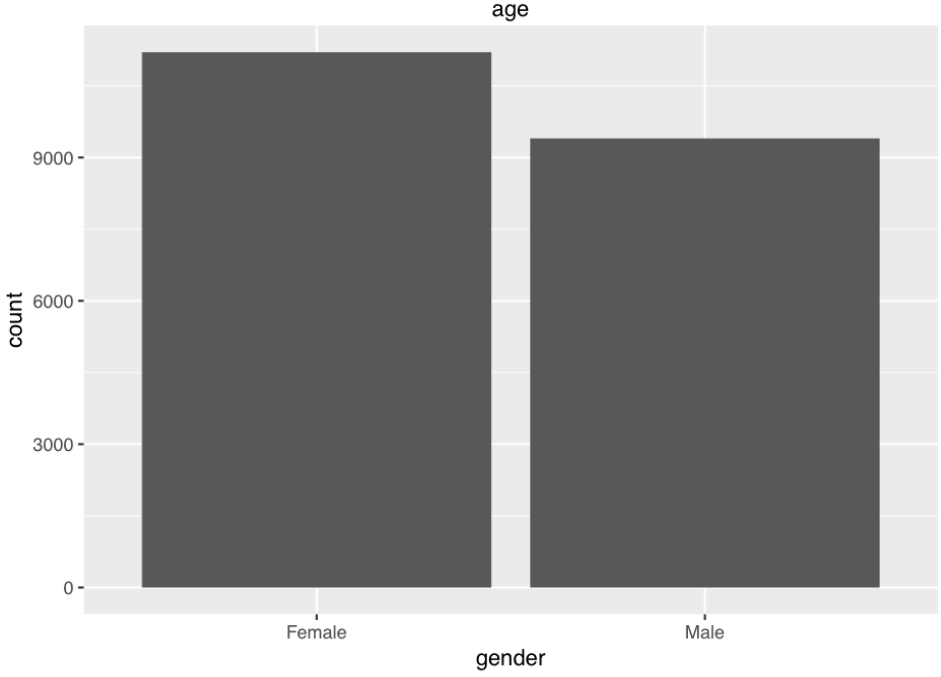




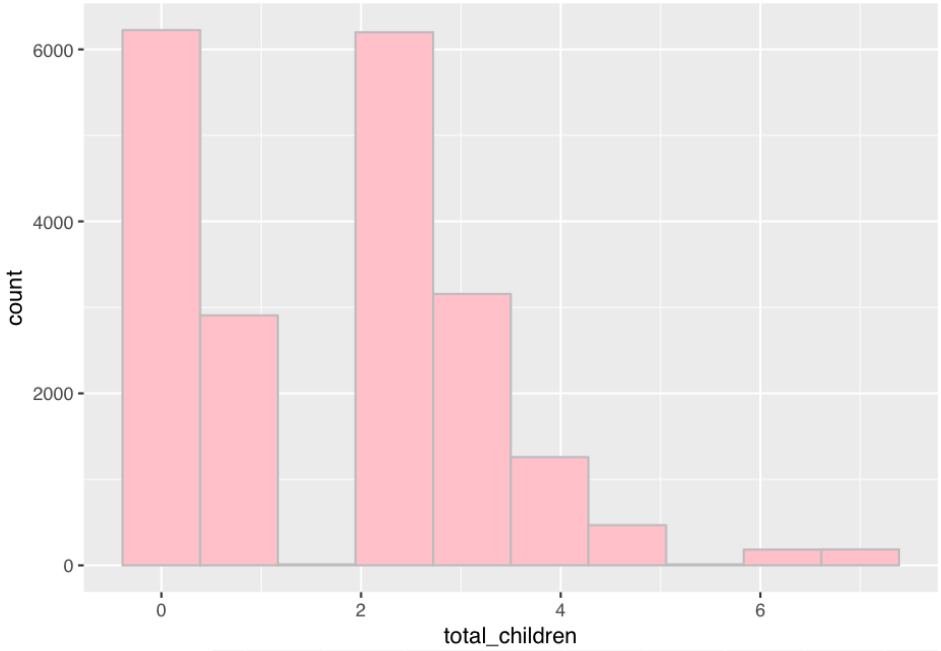

Less than $25,000
$75,000 to $99,999
$50,000 to $74,999
$25,000 to $49,999
$125,000 and more
$100,000 to $ 124,999
income
0
2000
4000
6000
count

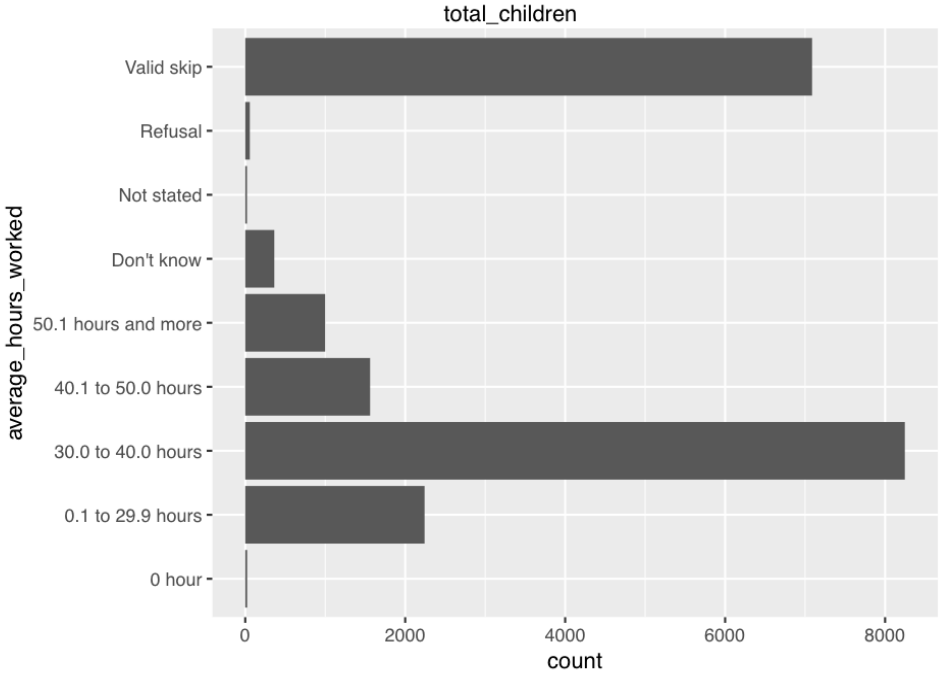

Very good
Poor
NA
Good
Fair
Excellent
Don't know
self_rated_mental_health
0
2000
4000
6000
count

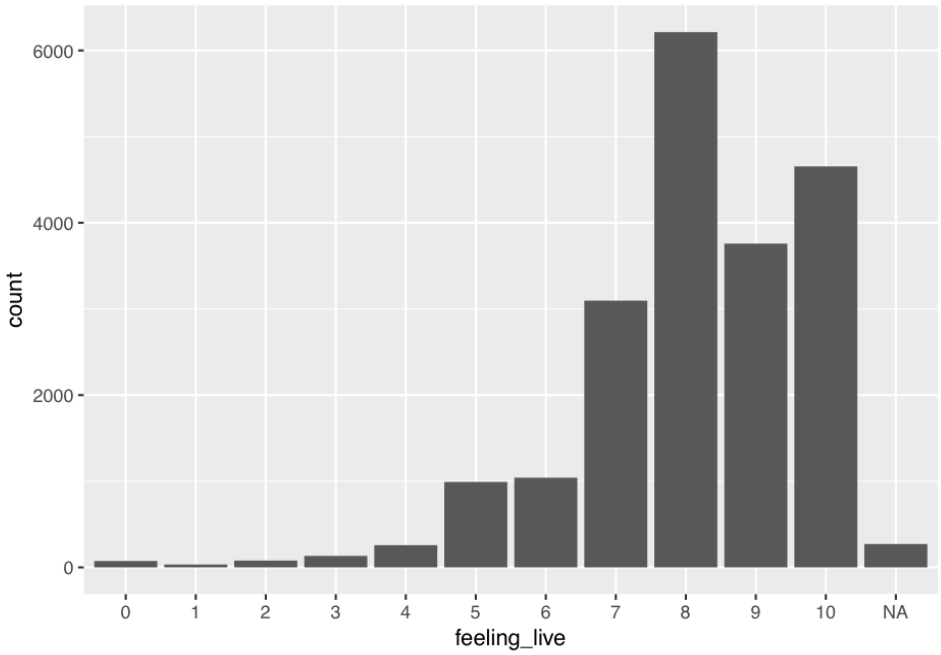





```
## Warning: Removed 19 rows containing non-finite values (stat_bin).
```


6000
4000
2000
0
count
0
2
4
6
total_children


Valid skip
Refusal
Not stated
Don't know
50.1 hours and more
40.1 to 50.0 hours
30.0 to 40.0 hours
0.1 to 29.9 hours
0 hour
average_hours_worked
0
2000
4000
6000
8000
count



-The above plots have showed the characteristics of the respondents. Based on the plots, we see that most of the respondents are married, so from this perspective, we might conclude getting married do bring people more satisfaction in life. Also, our data covers people from most age period, but most of the respondents are around the age of 80, so it might explain why many respondents' income is less than 25000, and hour worked is valid skip. Overall, the plot based on the variable feeling_live showed that the majority of respondents are happy about their lives.

- This data set come from the 2017 General Social Survey (GSS) on the Family. Since the target population of this survey is all non-institutionalized persons 15 years of age and older, living in the 10 provinces of Canada, the data set contains some observations that contain missing values. Therefore, we use mutate function to record these observations as NA to help us analysis the data set better.
- According to the weakness of our data set, as we can see from the age plot, most of responders are over 80 years old. This weakness may effect the accuracy of analysis. In addition, since the target population is under 15 years old, there are some information may missed such as overall felling of life of children. For the approach of sampling, each record in the survey frame was assigned to a stratum within its province, a simple random sample without replacement of records was next performed in each stratum.

Weaknesses

In regards to the weakness in our data set, the range of our sample size is relatively small, which only have five regions, Halifax, Sherbrooke, Sturgeon Falls, Winnipeg, and Edmonton respectively. Therefore, we are not allowed to predict felling-live in other places in Canada, as each place has a different lifestyle and local policies, the factors leading to people's life satisfaction will also be different. Furthermore people interviewed in this data set are not distributed well in every age group, in particular people aged 80 years old occupies the majority of target population. Thus the result we have is more suitable for aged people rather than those

in other age groups. Last but not the least, the data may not be very accurate, because some people may fail to respond while another one may think it is a fraudulent call and do not provide the real data. Also, some people might provide us an unreal data because some of the questions might invade their privacy, such as income. Thus, the data set we got might not be so accurate due to these reasons.

Next Steps

```
## Start:  AIC=19055.18
## gss$felling_live ~ gss$age + gss$total_children + gss$gender +
##     gss$marital_status + gss$education + gss$income_respondent +
##     gss$average_hours_worked
##
##                             Df Sum of Sq   RSS   AIC
## <none>                                   51753 19055
## - gss$age                    1     21.09 51774 19062
## - gss$education              7     55.00 51808 19063
## - gss$total_children         1     26.24 51779 19064
## - gss$average_hours_worked   7     76.98 51830 19071
## - gss$gender                 1    116.17 51869 19099
## - gss$income_respondent      5    238.59 51991 19139
## - gss$marital_status         6   2042.03 53795 19829
```

By using forward selection and backward elimination, we try to improve our data. However, the AIC indicates the model is good enough according to these methods, thus we can not to improve our model through this way. Thus, to improve our dataset, we can take some actions to avoid our weakness in the dataset. First of all, we can increase our sample size by collecting data from other provinces in Canada, and we may add some variables such as welfare policies in our data, thus we can predict a better result for all Canadians. Also, we can interview people from other countries like America or China to make a comparison between them to see how does the factors that influence felling-live vary across countries. Besides, we could change our method to collect data, we should have multiple ways for people to access this research such as online survey or go to different regions send out questionnaires to get more efficient data, thus we can have more people to join the research and get more useful data.

References

1

Kenton, W. (2020, September 21). How Multiple Linear Regression Works. Retrieved October 19, 2020, from https://www.investopedia.com/terms/m/mlr.asp

2

Multiple Linear Regression Analysis. (n.d.). Retrieved October 19, 2020, from http://reliawiki.org/index.php/Multiple_Linear_Regression_Analysis

3

Stastic Canada (2020, April). General Social Survey, Cycle 31 : Families Public Use Microdata File Documentation and User's Guide, fro https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/sdaweb/dli2/gss/gss31/gss31/more_doc/GSS31_User_Guide.pdf

4

Data from, https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/cgi-bin/sda/hsda?harcsda4+gss31

5

```
##
## To cite package 'tidyverse' in publications use:
##
##   Hadley Wickham (2017). tidyverse: Easily Install and Load the
##   'Tidyverse'. R package version 1.2.1.
##   https://CRAN.R-project.org/package=tidyverse
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {tidyverse: Easily Install and Load the 'Tidyverse'},
##     author = {Hadley Wickham},
##     year = {2017},
##     note = {R package version 1.2.1},
##     url = {https://CRAN.R-project.org/package=tidyverse},
##   }
```

```
##
## To cite package 'janitor' in publications use:
##
##   Sam Firke (2020). janitor: Simple Tools for Examining and Cleaning
##   Dirty Data. R package version 2.0.1.
##   https://CRAN.R-project.org/package=janitor
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {janitor: Simple Tools for Examining and Cleaning Dirty Data},
##     author = {Sam Firke},
##     year = {2020},
##     note = {R package version 2.0.1},
##     url = {https://CRAN.R-project.org/package=janitor},
##   }
```

6

DAta cleaning code given by Rohan Alexander and Sam Caetano, rohan.alexander@utoronto.ca