

Apply logistic regression Model to analyze the number of children someone has

Junming Zhang, Hairong Sun

Saturday, October 17, 2020

Code and data supporting this analysis is available at: [here \(click “here”\)](#)

Abstract

Children play an important role in family relationship, at the same time, birth rate also influences the future of a community because they will be the labour force of the community, which has an impact on the economical and social development, and even existence of the community. Therefore, it is necessary to have an indicator to suggest what factors influence if someone may have many children, and so the government can make policies to control birth rate based on the indicator. To generate such an indicator, we built a logistic regression model with the dataset **Canadian General Social Survey (GSS)** (citation 5) of year 2017, which predicts the possibility of many children (more than 2 children) one may have with some attributes (total_children, marital_status, education, partner_education, income_family, self-rated_health and self-rated_mental_health) we searched before which may help. We found that the probability of having many children has positive relation with mental health condition, and have negative relation with education background and physical health condition.

Introduction

Our goal is to find a model to predict the possibility one may have more than 2 children based on attributes that may have strong effects and test how strong are these attributes correlated to the number of children one may have. To create such a model, we use the dataset **Canadian General Social Survey (GSS)** from year 2017, which has attributes that may contribute to our research goals, like total_children, marital_status, education, partner_education, income_family, self-rated_health, and self-rated_mental_health, and we built our model with these attributes. The number of children is the attribute we are interested in. Education and partner_education are important because people who are enrolled in tertiary education tend to postpone their marriage and have fewer children (citation 1). We consider income_family because income_family is a mirror of the economic development, which is related to birth rate, for instance, economic depression may mean low fertility (Pobric & Robinson, 2015). We look into marital_status, because the type of partnership may contribute to the birth rate, for example, those who get married may have more children than those who cohabit (Martinez, Daniels, & Chandra, 2012). Finally but still important, the health status impacts the number of children one may have, for example, countries with higher HDI (Human Development Index, which involves life expectancy, education, and per capita income) may have lower fertility rate, which is reflected by the citation 4, countries with lower HDI have higher fertility rate and vice versa. Since health is related to HDI for life expectancy, we put the self-rated_health and self-rated_mental_health into the model. Therefore, in the modeling section, we create a logistic regression model with these attributes to predict the possibility one may have many (more than 2) children and analyze how is our prediction related to these attributes, and how strong are our interests and these attributes correlated. The results are delivered by the graphs between the probability of having many children and each attribute. It is useful because the correlation is straightforward through the delivery of graphs. This allows a policymaker to conclude the significant attributes that influence the number of having more than 2 children. Then the government could hand on policy/incentives to control the birth rate. In the discussion part, we summarized the analysis and explained the dataset. Weakness and next step parts indicate the limitation and space for the progress of this analysis.

Data

The dataset is obtained from **Canadian General Social Survey (GSS)** of year 2017, it contains all the attributes I listed in the Introduction section that are used to build the model. To make the dataset, they use a questionnaire and interview the respondents on phone call (Beaupré, 2020). A brief outline of the questionnaire is following (Beaupré, 2020):

- Entry component (respondent's date of birth)
- Family origins
- Leaving the parental home
- Conjugal history
- Intentions and reasons to form a union
- Respondent's children
- Fertility intentions
- Maternity/parental leave
- Organization and decision making within the household
- Arrangements and financial support after a separation/divorce
- Labour market new and education
- Health and subjective well-being
- Characteristics of respondent's dwelling
- Characteristics of respondent of spouse/partner

The questionnaire to build the dataset was delivered by telephone (Beaupré, 2020), and this questionnaire was helpful because it covers many details on the personal conditions of the respondent, like health and education, for which there is significant proof that influences the fertility rate. However, the previous nationalities of respondents are ignored, which may also introduce errors in the result, because some countries may have special cultures and religions that affect the fertility rate. There are also pros and cons for collecting data by phone. The benefit is that, since the most people have their own telephones today, so it is easy to connect and the data can be collected with lower costs. However, some people may not respond to the phone call, which leads to the non-participation error.

There are 81 variables/attributes and 20602 observations in the dataset. The variables generally cover many aspects about the living conditions and the personal conditions of the interviewee, which may suggest our interest, and we tested some of them which are possibly helpful according to the documents and references we found, and investigate the correlation between the variables in our scope and our interest. And also, the dataset has a large number of observations with respect to the place where the data were collected, and thus this makes the results (in Canada) can be found from the dataset more representative. However, since the dataset is only limited to one country (Canada), the variables do not reflect other factors may also have impacts but not suitable for just one country, like policy, war or peace, natural conditions, and so hard to reflect worldwide facts.

The data are collected with the stratified random sampling (simple random sampling without replacement in the stratum) method (Beaupré, 2020), a probability sampling approach. The target population for the dataset 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 (Beaupré, 2020). The frame of the survey is 1. Lists of telephone numbers in use (both landline and cellular) available to Statistics Canada from various sources (telephone companies, Census of population, etc.); and 2. The Address Register (AR): List of all dwellings within the ten provinces. The probability sampling method (collection approach for this dataset) decreases errors like generalization and more representative for the whole population. However, there are some drawbacks of the dataset from both non-sampling error. The non-sampling error is mainly

from (partial or total) non-participation. This is handled by adjusting the weights to less for non-participation cases (Beaupré, 2020), and in our implementation, we removed all rows with NA in the columns we needed to build the model. One similar variable we did not choose for this analysis is `income_respondent` because having a child is more a family decision. Therefore, choose `income_family` to represent the financial situation seems more reasonable.

Model

Here is the number of observations for each stratum (since the data are collected by stratified sampling without replacement), the stratum was divided based on the province the interviewee lived (Beaupré, 2020), we built our logistic regression model based on the stratification below:

```
## # A tibble: 10 x 2
## # Groups:   province [10]
##   province      n
##   <fct>      <int>
## 1 Alberta      1064
## 2 British Columbia 1490
## 3 Manitoba       708
## 4 New Brunswick   772
## 5 Newfoundland and Labrador 702
## 6 Nova Scotia     837
## 7 Ontario      3313
## 8 Prince Edward Island 421
## 9 Quebec       2191
## 10 Saskatchewan   675
##
## Call:
## svyglm(formula = if_many ~ as.factor(marital_status) + as.factor(education) +
##   as.factor(partner_education) + as.factor(income_family) +
##   as.factor(self_rated_health) + as.factor(self_rated_mental_health),
##   design = gss.design, family = "binomial")
##
## Weighted Residuals:
##   Min      1Q  Median      3Q      Max
## -1.5352 -0.6442 -0.5368  1.0985  6.0155
##
## Coefficients:
##
## (Intercept)                                Estimate
## as.factor(marital_status)Living common-law    -0.43755
## as.factor(marital_status)Married              0.11963
## as.factor(marital_status)Separated            0.08655
## as.factor(marital_status)Single, never married -1.78663
## as.factor(marital_status)Widowed              0.15065
## as.factor(education)College, CEGEP or other non-university certificate or di... 0.10964
## as.factor(education)High school diploma or a high school equivalency certificate 0.32493
## as.factor(education)Less than high school diploma or its equivalent 0.63463
## as.factor(education)Trade certificate or diploma 0.26572
## as.factor(education)University certificate or diploma below the bachelor's level 0.15623
## as.factor(education)University certificate, diploma or degree above the bach... 0.01973
## as.factor(partner_education)College, CEGEP or other non-university certificate or d... 0.08019
## as.factor(partner_education)High school diploma or a high school equivalency certi... 0.27359
## as.factor(partner_education)Less than high school diploma or its equivalent 0.49913
```

## as.factor(partner_education)Trade certificate or diploma	0.17458
## as.factor(partner_education)University certificate or diploma below the bachelor's level	0.15677
## as.factor(partner_education)University certificate, diploma or degree above the ba...	-0.18488
## as.factor(income_family)\$125,000 and more	0.09295
## as.factor(income_family)\$25,000 to \$49,999	0.04431
## as.factor(income_family)\$50,000 to \$74,999	-0.08042
## as.factor(income_family)\$75,000 to \$99,999	-0.10151
## as.factor(income_family)Less than \$25,000	-0.35354
## as.factor(selfRatedHealth)Excellent	-1.00447
## as.factor(selfRatedHealth)Fair	-0.74889
## as.factor(selfRatedHealth)Good	-0.91246
## as.factor(selfRatedHealth)Poor	-0.57634
## as.factor(selfRatedHealth)Very good	-1.01712
## as.factor(selfRatedMentalHealth)Excellent	1.11938
## as.factor(selfRatedMentalHealth)Fair	1.01850
## as.factor(selfRatedMentalHealth)Good	1.15656
## as.factor(selfRatedMentalHealth)Poor	0.72011
## as.factor(selfRatedMentalHealth)Very good	1.04924
##	Std. Error
## (Intercept)	0.80190
## as.factor(marital_status)Living common-law	0.15624
## as.factor(marital_status)Married	0.14709
## as.factor(marital_status)Separated	0.27844
## as.factor(marital_status)Single, never married	0.21058
## as.factor(marital_status)Widowed	0.28506
## as.factor(education)College, CEGEP or other non-university certificate or di...	0.06728
## as.factor(education)High school diploma or a high school equivalency certificate	0.06961
## as.factor(education)Less than high school diploma or its equivalent	0.08758
## as.factor(education)Trade certificate or diploma	0.08999
## as.factor(education)University certificate or diploma below the bachelor's level	0.11604
## as.factor(education)University certificate, diploma or degree above the bach...	0.08362
## as.factor(partner_education)College, CEGEP or other non-university certificate or d...	0.06902
## as.factor(partner_education)High school diploma or a high school equivalency certi...	0.06823
## as.factor(partner_education)Less than high school diploma or its equivalent	0.08684
## as.factor(partner_education)Trade certificate or diploma	0.09231
## as.factor(partner_education)University certificate or diploma below the bachelor's level	0.11428
## as.factor(partner_education)University certificate, diploma or degree above the ba...	0.08616
## as.factor(income_family)\$125,000 and more	0.06704
## as.factor(income_family)\$25,000 to \$49,999	0.07913
## as.factor(income_family)\$50,000 to \$74,999	0.07434
## as.factor(income_family)\$75,000 to \$99,999	0.07511
## as.factor(income_family)Less than \$25,000	0.13440
## as.factor(selfRatedHealth)Excellent	0.47467
## as.factor(selfRatedHealth)Fair	0.47697
## as.factor(selfRatedHealth)Good	0.47342
## as.factor(selfRatedHealth)Poor	0.48759
## as.factor(selfRatedHealth)Very good	0.47333
## as.factor(selfRatedMentalHealth)Excellent	0.64966
## as.factor(selfRatedMentalHealth)Fair	0.65492
## as.factor(selfRatedMentalHealth)Good	0.64903
## as.factor(selfRatedMentalHealth)Poor	0.69056
## as.factor(selfRatedMentalHealth)Very good	0.64929
##	t value
## (Intercept)	-1.771

## as.factor(marital_status)Living common-law	-2.801
## as.factor(marital_status)Married	0.813
## as.factor(marital_status)Separated	0.311
## as.factor(marital_status)Single, never married	-8.484
## as.factor(marital_status)Widowed	0.528
## as.factor(education)College, CEGEP or other non-university certificate or di...	1.630
## as.factor(education)High school diploma or a high school equivalency certificate	4.668
## as.factor(education)Less than high school diploma or its equivalent	7.246
## as.factor(education)Trade certificate or diploma	2.953
## as.factor(education)University certificate or diploma below the bachelor's level	1.346
## as.factor(education)University certificate, diploma or degree above the bach...	0.236
## as.factor(partner_education)College, CEGEP or other non-university certificate or d...	1.162
## as.factor(partner_education)High school diploma or a high school equivalency certi...	4.010
## as.factor(partner_education)Less than high school diploma or its equivalent	5.748
## as.factor(partner_education)Trade certificate or diploma	1.891
## as.factor(partner_education)University certificate or diploma below the bachelor's level	1.372
## as.factor(partner_education)University certificate, diploma or degree above the ba...	-2.146
## as.factor(income_family)\$125,000 and more	1.387
## as.factor(income_family)\$25,000 to \$49,999	0.560
## as.factor(income_family)\$50,000 to \$74,999	-1.082
## as.factor(income_family)\$75,000 to \$99,999	-1.351
## as.factor(income_family)Less than \$25,000	-2.630
## as.factor(selfRatedHealth)Excellent	-2.116
## as.factor(selfRatedHealth)Fair	-1.570
## as.factor(selfRatedHealth)Good	-1.927
## as.factor(selfRatedHealth)Poor	-1.182
## as.factor(selfRatedHealth)Very good	-2.149
## as.factor(selfRatedMentalHealth)Excellent	1.723
## as.factor(selfRatedMentalHealth)Fair	1.555
## as.factor(selfRatedMentalHealth)Good	1.782
## as.factor(selfRatedMentalHealth)Poor	1.043
## as.factor(selfRatedMentalHealth)Very good	1.616
##	Pr(> t)
## (Intercept)	0.07660
## as.factor(marital_status)Living common-law	0.00511
## as.factor(marital_status)Married	0.41604
## as.factor(marital_status)Separated	0.75592
## as.factor(marital_status)Single, never married	< 2e-16
## as.factor(marital_status)Widowed	0.59718
## as.factor(education)College, CEGEP or other non-university certificate or di...	0.10319
## as.factor(education)High school diploma or a high school equivalency certificate	3.08e-06
## as.factor(education)Less than high school diploma or its equivalent	4.55e-13
## as.factor(education)Trade certificate or diploma	0.00316
## as.factor(education)University certificate or diploma below the bachelor's level	0.17820
## as.factor(education)University certificate, diploma or degree above the bach...	0.81348
## as.factor(partner_education)College, CEGEP or other non-university certificate or d...	0.24537
## as.factor(partner_education)High school diploma or a high school equivalency certi...	6.11e-05
## as.factor(partner_education)Less than high school diploma or its equivalent	9.27e-09
## as.factor(partner_education)Trade certificate or diploma	0.05863
## as.factor(partner_education)University certificate or diploma below the bachelor's level	0.17012
## as.factor(partner_education)University certificate, diploma or degree above the ba...	0.03192
## as.factor(income_family)\$125,000 and more	0.16560
## as.factor(income_family)\$25,000 to \$49,999	0.57555
## as.factor(income_family)\$50,000 to \$74,999	0.27938

```

## as.factor(income_family)$75,000 to $99,999 0.17660
## as.factor(income_family)Less than $25,000 0.00854
## as.factor(selfRated_health)Excellent 0.03435
## as.factor(selfRated_health)Fair 0.11642
## as.factor(selfRated_health)Good 0.05396
## as.factor(selfRated_health)Poor 0.23722
## as.factor(selfRated_health)Very good 0.03167
## as.factor(selfRated_mental_health)Excellent 0.08491
## as.factor(selfRated_mental_health)Fair 0.11993
## as.factor(selfRated_mental_health)Good 0.07477
## as.factor(selfRated_mental_health)Poor 0.29706
## as.factor(selfRated_mental_health)Very good 0.10612
##
## (Intercept) .
## as.factor(marital_status)Living common-law **
## as.factor(marital_status)Married
## as.factor(marital_status)Separated
## as.factor(marital_status)Single, never married ***
## as.factor(marital_status)Widowed
## as.factor(education)College, CEGEP or other non-university certificate or di...
## as.factor(education)High school diploma or a high school equivalency certificate ***
## as.factor(education)Less than high school diploma or its equivalent ***
## as.factor(education)Trade certificate or diploma **
## as.factor(education)University certificate or diploma below the bachelor's level
## as.factor(education)University certificate, diploma or degree above the bach...
## as.factor(partner_education)College, CEGEP or other non-university certificate or d...
## as.factor(partner_education)High school diploma or a high school equivalency certi... ***
## as.factor(partner_education)Less than high school diploma or its equivalent ***
## as.factor(partner_education)Trade certificate or diploma .
## as.factor(partner_education)University certificate or diploma below the bachelor's level
## as.factor(partner_education)University certificate, diploma or degree above the ba... *
## as.factor(income_family)$125,000 and more
## as.factor(income_family)$25,000 to $49,999
## as.factor(income_family)$50,000 to $74,999
## as.factor(income_family)$75,000 to $99,999
## as.factor(income_family)Less than $25,000 **
## as.factor(selfRated_health)Excellent *
## as.factor(selfRated_health)Fair
## as.factor(selfRated_health)Good .
## as.factor(selfRated_health)Poor
## as.factor(selfRated_health)Very good *
## as.factor(selfRated_mental_health)Excellent .
## as.factor(selfRated_mental_health)Fair
## as.factor(selfRated_mental_health)Good .
## as.factor(selfRated_mental_health)Poor
## as.factor(selfRated_mental_health)Very good
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.998 on 12131 degrees of freedom
## Multiple R-squared:  0.001652, Adjusted R-squared:  -0.001722
## F-statistic: 0.6274 on 32 and 12131 DF, p-value: 0.9498

```

We used R studio to run the model. In this model, we continued our analysis by building a logistic regression

model to predict the odds of having more than 2 children in Canada using variables marital status, respondents and partners' education, family income, self rated physical and mental health. Since the data is collected from Canada in 2017, the scope of application of my research is within Canada. Using the data selected, we build a logistic regression model with the following formula:

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 \text{marital_status_Living_common_law} + \beta_2 \text{marital_status_Married} + \dots + \beta_{23} \text{self_rated_health_Excellent} + \dots + \beta_{32} \text{self_rated_mental_health_Very_Good}$$

p is the probability of having many children. (p/1-p) is the odds of having more than 2 children in Canada. β_0 to β_{32} are coefficients and associate with each variable. Variables from as.factor(marital_status)Living common-law to as.factor(self_rated_mental_health)Very good are dummy variables. There are only 2 possible value range: 1 or 0 for dummy variables. For example, when a person is married, the marital_status_Married = 1. Otherwise equals to 0. Values shown above.

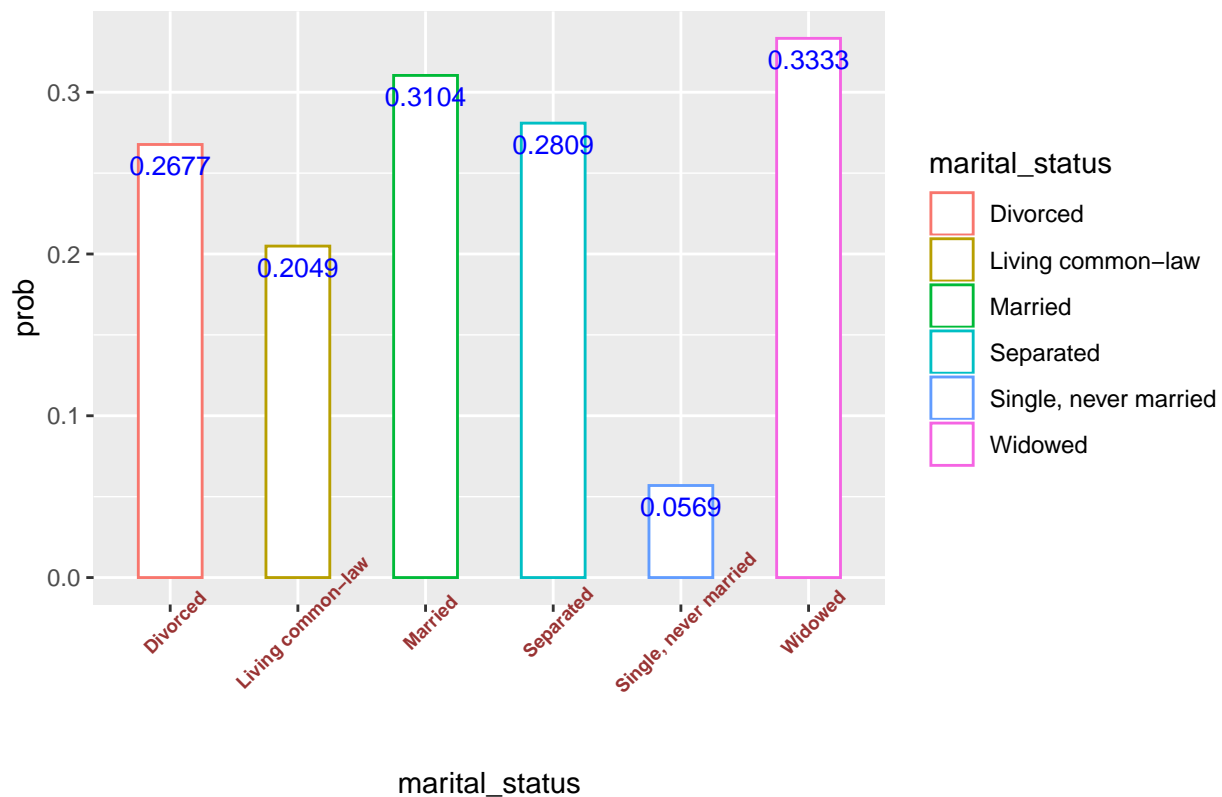
When the marital status = divorced, education = Bachelor's degree (e.g. B.A., B.Sc., LL.B.), partner_education = Bachelor's degree (e.g. B.A., B.Sc., LL.B.), family_income = \$100,000 to \$124,999, self_rated_health = Don't know, self_rated_mental_health = Don't know, the log odds of having >2 children equals to -1.42008.

Keep other conditions unchanged, when the marital_status changed from divorced to Living common-law, the log odds of having more than 2 children will be decrease by 0.43755. This rule also applies to other dummy variables.

Results

##	marital_status	have_many	total	prob
## 1	Divorced	68	254	0.2677
## 2	Living common-law	382	1864	0.2049
## 3	Married	2824	9097	0.3104
## 4	Separated	25	89	0.2809
## 5	Single, never married	45	791	0.0569
## 6	Widowed	26	78	0.3333

the possibility to have many kids (> 2) for each marital status

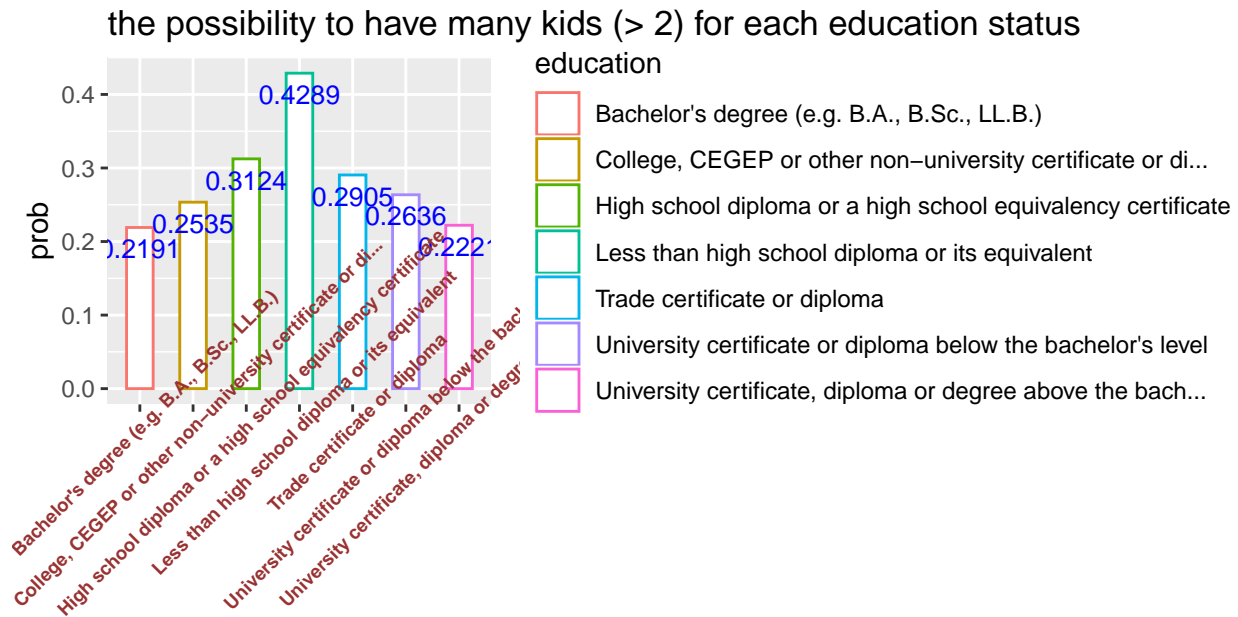


We analyzed the relationship between the probability of having more than 2 kids in Canada and each attribute respectively. In this graph, based on marital status, we can see that widowed people have the highest probability while people who never married have the lowest probability of parenting more than 2 children. The probability is ranked as below:

Single, never married < Living common-law < Divorced < Separated < Married < Widowed

The calculation of probability is the amount of (>2 kids) cases/total cases within one situation.

```
##                                     education have_many
## 1      Bachelor's degree (e.g. B.A., B.Sc., LL.B.)      583
## 2 College, CEGEP or other non-university certificate or di... 720
## 3 High school diploma or a high school equivalency certificate 831
## 4      Less than high school diploma or its equivalent    537
## 5                                     Trade certificate or diploma 278
## 6 University certificate or diploma below the bachelor's level 131
## 7 University certificate, diploma or degree above the bach... 290
## total prob
## 1 2661 0.2191
## 2 2840 0.2535
## 3 2660 0.3124
## 4 1252 0.4289
## 5  957 0.2905
## 6  497 0.2636
## 7 1306 0.2221
```

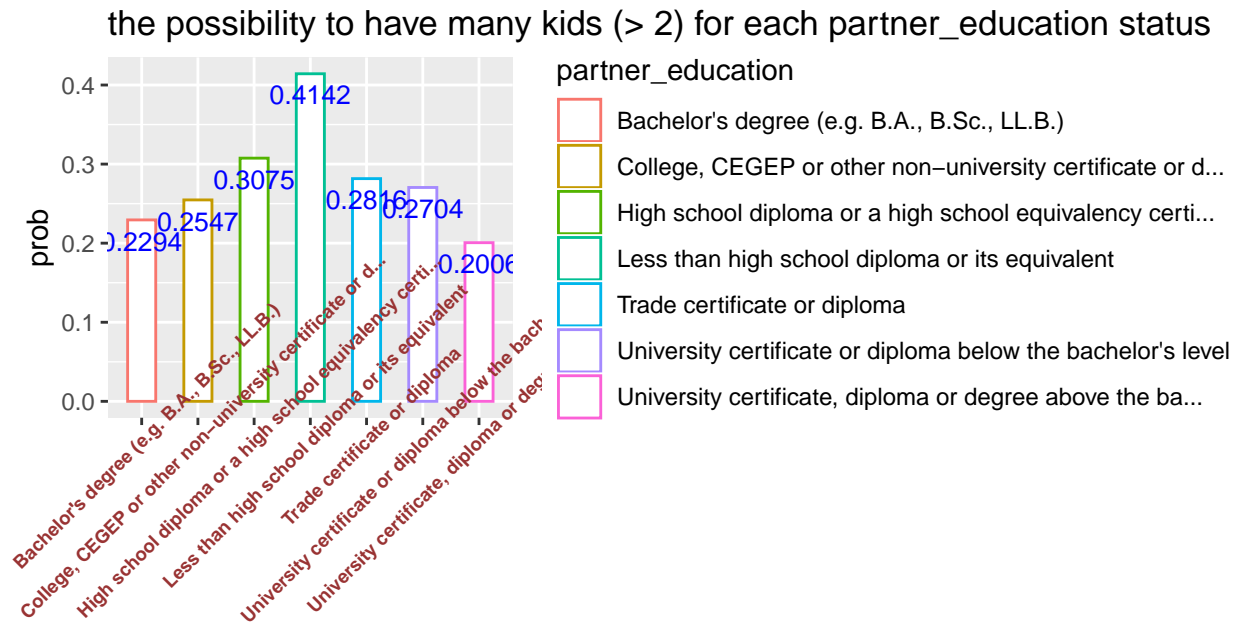
education

In the education section, the probability of having more children for people who received less than high school diploma is surprisingly doubled the amount of higher education background groups(people who received a Bachelor's degree/ University certificate). The probability is ranked as below:

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

##	partner_education	have_many
## 1	Bachelor's degree (e.g. B.A., B.Sc., LL.B.)	576
## 2	College, CEGEP or other non-university certificate or d...	657
## 3	High school diploma or a high school equivalency certi...	944
## 4	Less than high school diploma or its equivalent	541
## 5	Trade certificate or diploma	254
## 6	University certificate or diploma below the bachelor's level	139
## 7	University certificate, diploma or degree above the ba...	259

##	total	prob
## 1	2511	0.2294
## 2	2579	0.2547
## 3	3070	0.3075
## 4	1306	0.4142
## 5	902	0.2816
## 6	514	0.2704
## 7	1291	0.2006



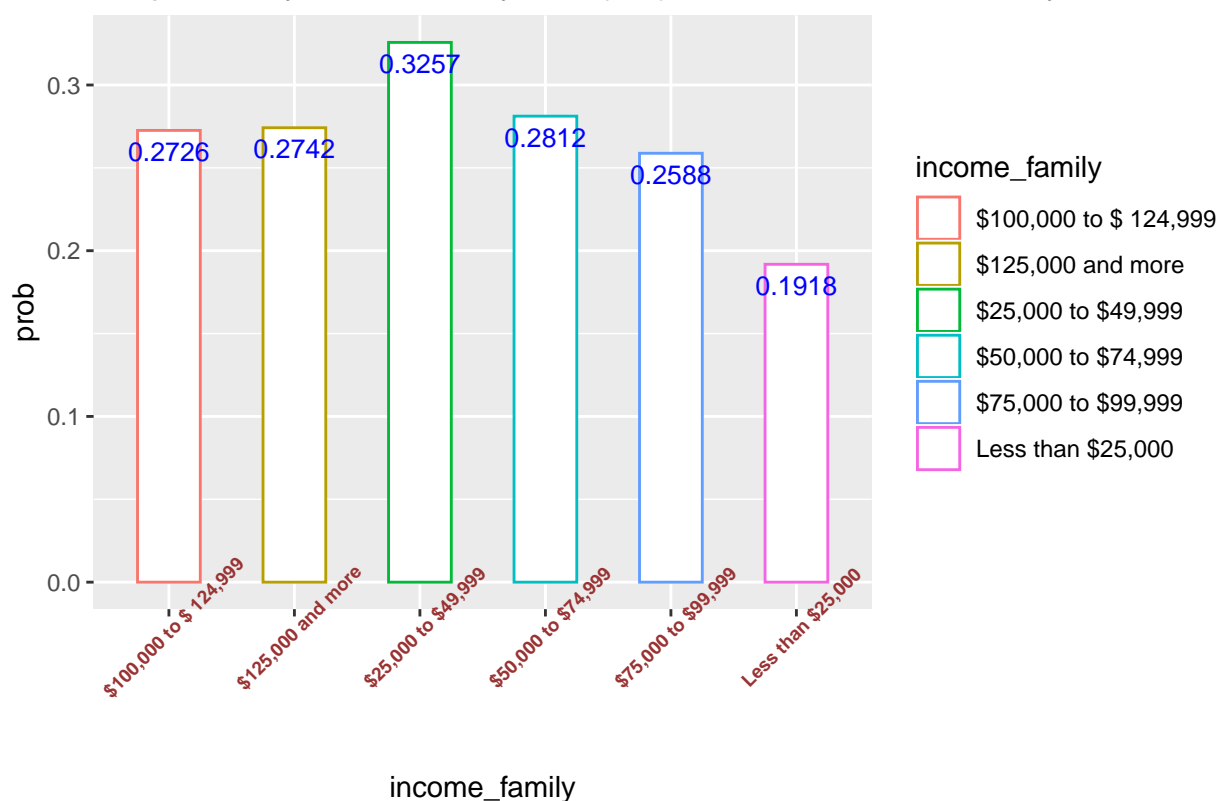
partner_education

In the partner_education section, the probability of having more children for people who received less than high school diploma is almost doubled the amount of higher education background groups(people who received a Bachelor's degree/ University certificate) as the last graph.The probability is ranked as below:

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

##	income_family	have_many	total	prob
## 1	\$100,000 to \$ 124,999	467	1713	0.2726
## 2	\$125,000 and more	1037	3782	0.2742
## 3	\$25,000 to \$49,999	595	1827	0.3257
## 4	\$50,000 to \$74,999	624	2219	0.2812
## 5	\$75,000 to \$99,999	549	2121	0.2588
## 6	Less than \$25,000	98	511	0.1918

the possibility to have many kids (> 2) for each interval of family income

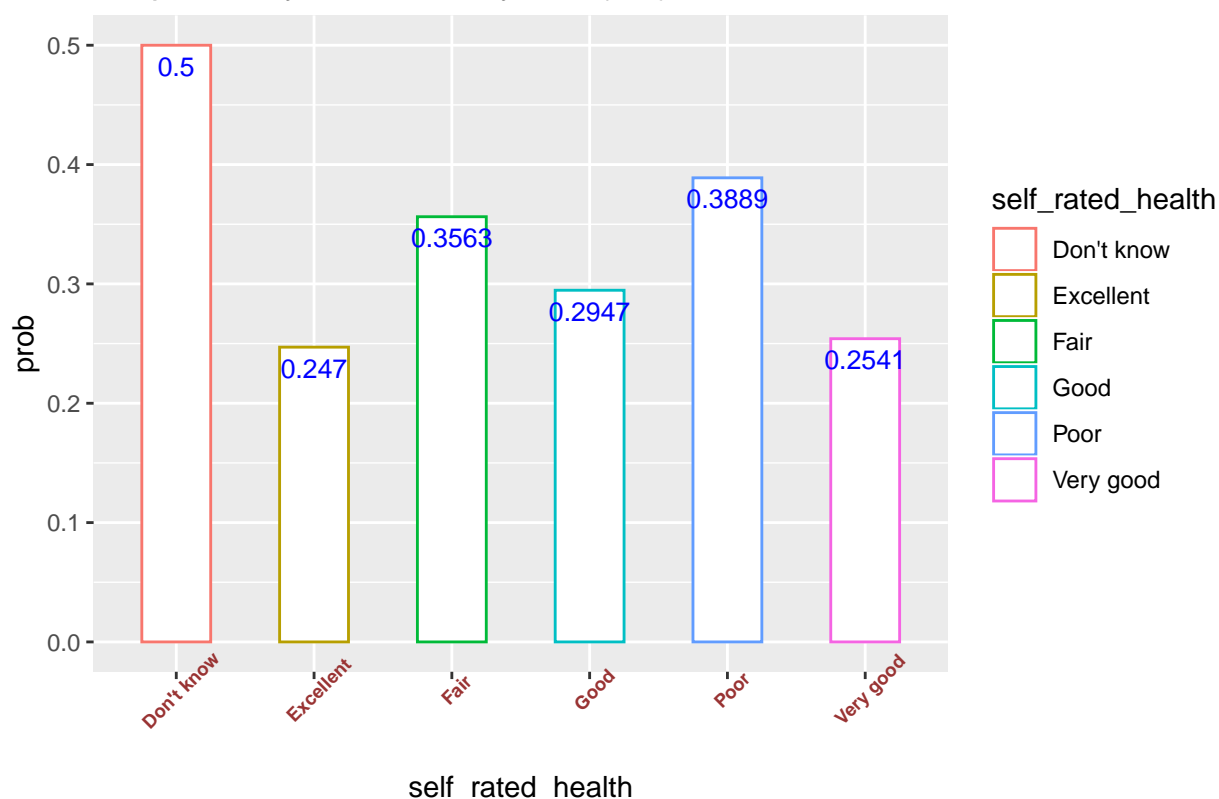


In the family income section, the variation between different groups is not as huge as education-related graphs. People who have a family income of \$25,000 to \$49,999 have the highest probability of parenting more than 2 kids. The probability is ranked as below:

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

##	selfRatedHealth	haveMany	total	prob
## 1	Don't know	10	20	0.5000
## 2	Excellent	692	2802	0.2470
## 3	Fair	362	1016	0.3563
## 4	Good	1065	3614	0.2947
## 5	Poor	119	306	0.3889
## 6	Very good	1122	4415	0.2541

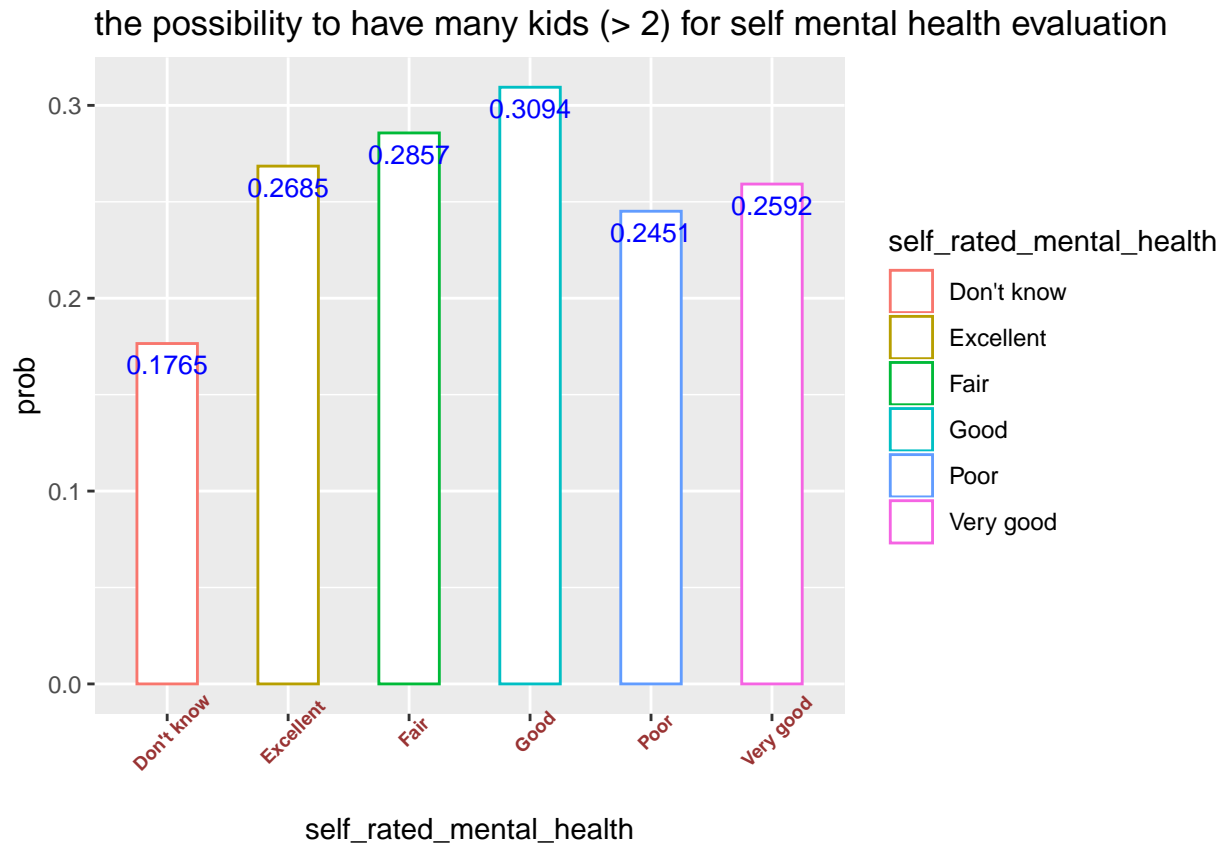
the possibility to have many kids (> 2) for self health evaluation



In the self-health evaluation section, apart from people who do not know their health condition, the highest probability is 0.3889 (poor), and the lowest 0.247(Excellent). The probability is ranked as below:

Excellent < Very Good < Good < Fair < Poor < Don't know

##	selfRatedMentalHealth	have_many	total	prob
## 1	Don't know	3	17	0.1765
## 2	Excellent	1035	3855	0.2685
## 3	Fair	170	595	0.2857
## 4	Good	1022	3303	0.3094
## 5	Poor	25	102	0.2451
## 6	Very good	1115	4301	0.2592



In the self mental health evaluation section, apart from people who do not know their mental health condition, the highest probability is 0.3094 (Good), and the lowest 0.2451(Poor). The probability is ranked as below:

Don't know < Poor < Very Good < Excellent < Fair < Good

There are several interesting conclusions we found in the result graphs. Firstly, from the perspective of marital status, the probability of having more children is higher for people who once got married. Secondly, from parents' both sides' educational background, people who received the "Less than high school diploma or its equivalent" are at the top of the lists. A pattern shows that the higher the diploma the person receives, the lower the probability of having more children. We deduce that people who received less education may have less knowledge of birth control and planning. In this case, people would have to give birth to more children. Thirdly, the poorer the health condition is, the higher the probability of fertility is and vice versa. Differently, in terms of mental health, the better the mental health is, the higher the probability of having more children. Form this, we can also conclude that parenting children burdens physical conditions but reduce mental stress.

Discussion

This model can provide reference to policy makers, for example, if the birth rate is too high, maybe the government should give an impulse to parenting/ birth control education. This can also provide references to newly married couples. They could know more about the current parenting situation and make their own plans. In large world (all countries), there are cultural and political factors that may influence the statistical result. However, in small world (Canada), Canada does not contain the whole factors. Therefore, the statistical result of Canada is not sufficient to represent the rest of the world.

Weaknesses

The main weakness of this project is the scope of the data. The data we collected are constrained within Canada. In this case, we are not able to introduce a model that is applicable in most countries because each country's economic/cultural/religious backgrounds are different. Moreover, the amount of data is not sufficient to conclude a strong formula. As the **Canadian General Social Survey (GSS)** (citation 5) of the year 2017 indicated that the response rate was 52.4%. Within the received data, there is also a significant amount of NAs existing in the responses.

Next Steps

In the next step, we may collect more data in other countries with different economic conditions and cultural backgrounds, like the country they immigrate from, the religious background to show the results more generally. Also, we can use the principal component analysis to narrow the variables of the model, which are strongly correlated with the interest. It is also worth considering to build a neural network model to make predictions based on our data, since the NN model is more robust to random cases.

References

1. NCHS Pressroom - 1997 Fact Sheet - Mothers Education and Birth Rate. (2009, November 17). Retrieved October 17, 2020, from <https://www.cdc.gov/nchs/pressroom/97facts/edu2birt.htm>
2. Pobric, A., & Robinson, G. M. (2015). Population ageing and low fertility: Recent demographic changes in Bosnia and Herzegovina. *Journal of Population Research*, 32(1), 23-43. doi:10.1007/s12546-014-9141-5
3. Martinez, G., Ph.D, Daniels, K., Ph.D, & Chandra, A., Ph.D. (2012, April 12). Fertility of Men and Women Aged 15–44 Years in the United States: National Survey of Family Growth, 2006–2010. Retrieved October 17, 2020, from <https://www.cdc.gov/nchs/data/nhsr/nhsr051.pdf>
4. Fertility rate, total (births per woman). (n.d.). Retrieved October 17, 2020, from <https://data.worldbank.org/indicator/SP.DYN.TFRT.IN>
5. 2017 General Social Survey (GSS): Families Cycle 31. (2017). Retrieved October 17, 2020, from Statistics Canada.
6. Beaupré, P. (2020). General Social Survey Cycle 31 : Families Public Use Microdata File Documentation and User's Guide. Ottawa, Canada: Authority of the Minister responsible for Statistics Canada. Retrieved October 17, 2020, from https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/sdaweb/dli2/gss/gss31/gss31/more_doc/GSS31_User_Guide.pdf
7. Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
8. T. Lumley (2020) "survey: analysis of complex survey samples". R package version 4.0.
9. T. Lumley (2004) Analysis of complex survey samples. *Journal of Statistical Software* 9(1): 1-19
10. T. Lumley (2010) Complex Surveys: A Guide to Analysis Using R. John Wiley and Sons.
11. User1489975user1489975 1, BenBarnesBenBarnes 17.3k66 gold badges5151 silver badges7070 bronze badges, Mnelmnel 103k2525 gold badges241241 silver badges240240 bronze badges, Amrrsamrrs 5, RnoobRnoob 9031010 silver badges1212 bronze badges, DroneDrone 11111 silver badge44 bronze badges, . . . Luchao QiLuchao Qi 3122 bronze badges. (1961, October 01). Omit rows containing specific column of NA. Retrieved October 18, 2020, from <https://stackoverflow.com/questions/11254524/omit-rows-containing-specific-column-of-na>
12. Ggplot2 barplots : Quick start guide - R software and data visualization. (n.d.). Retrieved October 18, 2020, from <http://www.sthda.com/english/wiki/ggplot2-barplots-quick-start-guide-r-software-and-data-visualization>