

Household Income Associated with Awareness of Canada’s Food Guide (2007): A Logistic Regression and Propensity Score Analysis of the Accessibility and Effects of CFG

Jeffrey Li 1005454919

12/4/2020

ABSTRACT

The purpose of this analysis was to build a model from the 2017-2018 Canadian Community Health Survey to explore the factors which limited or enhanced Canadian’s exposure to Canada’s Food Guide. In addition, supplementary analysis including propensity score matching (PSM) was employed to evaluate whether biased information from the food guide regarding fruit juice actually influenced Canadian consumption and dietary habits. Furthermore, CFG awareness-to-usage conversion rate was also examined. The methods employed to prepare the data for modeling included a cleaning and removal of NA measurements, establishing broader age brackets, isolating the observations of interest, and eliminating categorical variables from the data set which contained only 1 level. The logistic model was built using stepwise backwards selection and the final model included variables age, sex, education, country of birth, and family income. The model has an area under the curve (AUC-ROC) of 0.88 and a prediction error of 1.077413. Overall, this model, though surprising because access to health services, unmet health care needs, and food security were not significant, informs the government that access to the food guide is more so affected by age, sex, education, country of birth, and income. PSM revealed that food guide awareness did not significantly influence fruit juice consumption and the awareness-to-usage conversion rate was approximately 59%. These results confirm that more effective policy is need to ensure more Canadians actively access Canada’s Food Guide and adhere to its guidelines. Code and data supporting this analysis can be found at (<https://github.com/jeffxinli/STA304>).

Keywords: Human, nutrition, public health, food guide, Canada, healthcare, logistic regression, propensity score, CCHS, Canadian Community Health Survey.

INTRODUCTION

Human nutrition is the science which explores the links between all aspects of human health and nutrition, often identifying important nutrients and their roles in not only preventing disease, but also in maintaining and improving health. While poor nutrition is often associated with poverty, especially in developing countries, advancements in scientific literature highlight that no human population is invulnerable to improper nutrition. In Canada, cardiovascular disease (CVD), diabetes, and kidney disease are among the top leading causes of death alongside respiratory disease, influenza, and Alzheimer’s (StatsCan, 2011). Nutrition research, both observational studies and clinical trials, have elucidated the associations between nutrition and diseases like CVD as well as the nutritional causes and biomarkers of disease. More novel evidence even suggests the importance of proper nutrition in maintaining cognitive function in older adults and treating Alzheimer’s. Given the expansive scientific evidence suggesting the importance of proper nutrition on health, the government of Canada aims to encourage healthy dietary habits within its population through the official food guide.

The history of Canadian food guides is complicated; the first version released in 1942 was called “Canada’s Official Food Rules” and was more focused on wartime nutrition and limitations (Government of Canada, 2007). Subsequent versions brought slight improvements; however, through the years the guide was not only

adjusted for new technologies in food processing and storage but also in emerging evidence. The most well-known food guide today is that from 2007 called “2007 Eating Well with Canada’s Food Guide”. This guide was by far the most comprehensive at the time with updated nutrient reference values, revision of scientific evidence, consultations, and the considerations of the evolution of Canada’s food supply and the lifestyles of its people (Government of Canada, 2007). This guide, since its publication, has been marketed across Canada through multiple campaigns and schools are pushed to familiarize students with the guide. However, in January 2019, the food guide was revised. In fact, the 2007 guide was found to be scientifically inaccurate and biased towards many Canadian industries like the meat and dairy industries. Other information that were debunked include the substitution of fruits with juice and fruit snacks as well as the claim that all proteins are equally healthy (Government of Canada, 2019). The new guide, unbiased and verified by nutrition experts, attempts to provide Canadians with the most up-to-date and accurate information.

Nonetheless, despite advancements in scientific literature and parallel public health policies, the number of Canadians suffering from diseases linked to nutrition is deeply concerning. From the high incidence of obesity and heart disease to the prevalence of diabetes, this increase in disease among Canadians will bring unprecedented pressure on Canadian health care systems. This brings experts, scientists, and public health officials to question the effectiveness of the food guide. That is, are Canadians aware it exists, if so, are they adhering to it and how well? This study aims to evaluate through a binomial logistic model built using the 2017-2018 Canadian Community Health Survey, the people who are more or less likely to know about the food guide. Furthermore, supplemental analyses will briefly evaluate the effects of the food guide on the dietary habits of Canadians as well as the awareness-to-use conversion rate, defined as the proportion of those who have seen or heard of the food guide and actually use it. The methodology section will detail the survey and its data as well as the models used in the analysis. Indeed, the survey was conducted in 2017-2018 and reported on the 2007 guide; however, given the establishment and popularity of the 2007 guide through the past decade and a half, the results should provide an accurate estimation of the characteristics of the people who are more likely to be uninformed or heavily affected. Then, the findings of the analysis will be presented in the results section followed by a discussion of the analysis including a summary, conclusions, and limitations and next steps. Ultimately, this study would be able to help the Government of Canada direct marketing, education, and outreach resources to more vulnerable groups. Common variables such as age and gender will be considered alongside variables such as access to health care and food security. Through this study, the government will have a better understanding of the people it needs to help and the shortcomings of the 2007 CFG as well as identify the main potential barriers between the 2019 Canada’s Food Guide and Canadians.

METHODOLOGY

Canadian Community Health Survey

The data for this study was from the 2017-2018 Canadian Community Health Survey a cross sectional survey which collects information related to health status, health care, and other determinants of health in the Canadian population. The survey is offered every two years with the objectives of (1) supporting health surveillance programs; (2) providing a single data source for research on populations and rare characteristics; (3) providing information accessible to all community researchers; and (4) creating a flexible instrument that can gather information to address emerging issues relating to population health. The target population of the CCHS is the Canadian population aged 12 years and older living in the provinces and territories. The sampling frame, though, differs since the CCHS officially excludes those living in Aboriginal settlements, full-time members of the Canadian Forces, youth aged 12-17 in foster homes, institutionalized populations, and those living in the Quebec health regions of Nunavik and Terres Cries de la Baie James. Nonetheless, the CCHS is still representative of the Canadian population as these excluded groups represent less than 3% of the target population. Finally, the CCHS aims to achieve a sample size of 130 000 respondents. They recruit their respondents using a multi-stage sample allocation strategy. Overall, to ensure fair distribution according to the health regions and provinces, their sampling is weighted to the size of each population. Overall, there are 1051 variables measured; however, while some variables like age and sex are directly measured through the questionnaire, other variables like “Unmet health care needs” are derived from health care needs variables asked directly through the survey. These variables are intended to broadly cover population health aspects. While they are useful in visualizing the general state of factors, for example health care needs, the

meaningfulness of the variable is reduced since we do not actually know the magnitude of the seriousness of the unmet health care needs. Nonetheless, in a study looking to broadly examine food guide knowledge and use, these variables may be very informative of general trends. A general weaknesses of this survey are the large non-responses to certain questions; certain variables, like those regarding nutrition and CFG have around 1-3 thousand responses out of a sample of over 100 000 observations.

Data

The data retrieved for this study is only a subset of the 2017-2018 CCHS comprising of 21 of the 1051 variables. The variables selected were hypothesized to relate or be associated to food guide use/awareness. Though variables like “eat leafy green vegetables” were available and probably have high association with food guide use/awareness, they were not selected due to the meaningfulness of the variable. That is knowing that food guide knowledge is associated with leafy vegetable consumption is not very helpful in identifying those with greater risk of decreased awareness of the food guide. In sum, the subset data extracted from the CCHS contains 113 290 observations of 21 variables. The raw data is plotted to visualize the general trends and distributions in the raw data (Appendix 1). The data was slightly adjusted for the plots as the variable names are coded, however, the data itself was left untouched. From the plots, it becomes clear that many responses regarding the food guide are NA and the sample size for analysis will be smaller than expected. In addition, age and gender seem well distributed as well as income. It is noteworthy that most respondents were born in Canada and report food security. However, it will be interesting to verify whether those who report no to having heard or seen Canada’s food guide generally fit into either of these highly skewed groups. First, the coded variable names were changed to reflect the factors measured (e.g. DODIS changed to Distress). Then, only “heard or seen Canadian food guide” was kept as the dependent variable. Variables relating to food guide use were kept in the data set for supplementary analysis of those who are aware of Canada’s food guide. Subsequently, NA responses were removed and variables “height”, “weight”, and “culture/race” were completely removed because they contained too many NA responses. Furthermore, the age categories were broadened to 15-24yrs (Young Adult), 25-39yrs (Adult), 40-59yrs (Middle Age), and 60 and over (Old). Finally, each variable was analyzed to ensure that responses existed in all levels of the variable. Interestingly, access to health services and unmet health care needs only contained ‘no’ responses. Similarly, Distress only contained yes and Depression and Health insurance only contained no. As a result, these variables were not included in our model; however, they retain some significance in terms of meaningful inferences, therefore, they will be further examined in the discussion section. The final cleaned data for our model analysis contained 836 observations of 9 variables. The characteristics of those responses are presented in Table 1 below.

Table 1: Baseline Characteristics Table

	Yes	No	p	test
n	746	90		
Sex = Female (%)	430 (57.6)	26 (28.9)	<0.001	
Age (%)			<0.001	
Age between 12 and 14	37 (5.0)	21 (23.3)		
Young Adult	81 (10.9)	7 (7.8)		
Adult	172 (23.1)	16 (17.8)		
Middle Age	244 (32.7)	15 (16.7)		
Old	212 (28.4)	31 (34.4)		
Education (%)			<0.001	
Less than secondary school graduation	44 (5.9)	18 (20.0)		
Secondary school graduation, no post-secondary education	111 (14.9)	13 (14.4)		
Post-secondary certificate diploma or univ degree	591 (79.2)	59 (65.6)		
Birth_country = Other (%)	70 (9.4)	33 (36.7)	<0.001	
Worried_food_run_out (%)			0.758	
Often true	23 (3.1)	2 (2.2)		
Sometimes true	68 (9.1)	10 (11.1)		

	Yes	No	p	test
Never true	655 (87.8)	78 (86.7)	0.294	
Household_food_security_status (%)				
Food secure	660 (88.5)	78 (86.7)		
Moderately food insecure	54 (7.2)	10 (11.1)		
Severely food insecure	32 (4.3)	2 (2.2)	<0.001	
Household_income (%)				
No income or less than \$20,000	41 (5.5)	9 (10.0)		
\$20,000 to \$39,999	72 (9.7)	21 (23.3)		
\$40,000 to \$59,999	85 (11.4)	12 (13.3)		
\$60,000 to \$79,999	90 (12.1)	2 (2.2)		
\$80,000 or more	458 (61.4)	46 (51.1)		

As shown in Table 1, most respondents are highly educated Canadian born individuals aged 25 and over. Other noteworthy characteristics are the income status with 60% earning more than 80 000\$ household income, and food security variables with most individuals reporting food security. Given the great disparity between income, education, and food security groups, it will be interesting to see if food guide awareness is associated with greater or lower levels of these variables. Data for propensity score matching called ‘data_matched’ is a subset of the data used for the main logistic regression analysis.

Models

Logistic Regression Analysis

This model was created in RStudio. Given the categorical nature of the data, meaning that the response variable “heard or seen of Canadian food guide” is qualitative and measured as ‘yes’ or ‘no’, a binary logistic regression model was employed. This logistic regression model allows predictions of categorical variable levels instead of continuous quantitative measurements for which linear regression models may be more appropriate. To build this model, an initial full model containing all the pre-selected variables was created. An initial chi-square test was conducted between the response variable ‘heard or seen Canadian food guide’ and each independent variable to evaluate the significance of each variable (details in Appendix 2). This was for visualization purposes and no variable was removed for insignificant p-values. However, it was noticeable that the two food security variables were statistically insignificant p-value > 0.1. Subsequently, stepwise backwards selection was employed to simplify the model and remove any variables that did not contribute any meaningful information. Specifically, the significance of each variable was evaluated using the Chi-square ANOVA tests detailed in Appendix 3. ANOVA stands for analysis of variance and is used to examine the differences between two models (Wiley & Wiley, 2019). Indeed, it evaluates whether the new model without the variable of interest is any different from the base model. This comparison determines whether the variable of interest actually contributes any additional information to the model. Due to a small sample size, to avoid any Type II errors, a significance level of 0.1 was used instead of the conventional 0.05. Through this process a final model containing Sex, Age, Education, birth_country, and Household_income as the predictor variables was created. The formula for the model is as follows:

$$\begin{aligned}
\hat{y} = \log\left(\frac{\Pr(\text{Heard_or_seen_Canadian_food_guide} = 1)}{\Pr(\text{Heard_or_seen_Canadian_food_guide} = 0)}\right) = \\
2.289 - 1.265 * Sex_{Female} - 2.421 * Age_{Young_Adult} - 2.370 * Age_{Adult} \\
- 2.956 * Age_{Middle_Age} - 2.591 * Age_{Old} \\
- 1.398 * Education_{Secondary_school_graduation_no_post-secondary_education} \\
- 1.697 * Education_{Post-secondary_certificate_diploma_or_univ_degree} \\
+ 2.379 * Birth_country_{Other} + 0.243 * Household_income_{20,000_to_39,999} \\
- 0.502 * Household_income_{40,000_to_59,999} \\
- 2.807 * Household_income_{60,000_to_79,999} \\
- 0.861 * Household_income_{80,000_or_more}
\end{aligned}$$

The formula shows each predictors' levels' coefficient - also detailed in Appendix 4. The \hat{y} is the log odds of a specific individual having heard or seen Canada's food guide. A predictor's coefficient indicates that when all other factors are kept constant, the log odds of the people who have heard or seen Canada's food guide changes by its coefficient. For example, when all factors are kept constant and we compare a person's whose age is between 15 and 24 (Young adult) to someone aged between 25 and 39, the person aged between 25 and 39 (Adult) will have a higher log odds of having heard or seen Canada's food guide because the coefficient is -2.2809 greater than that of the young adult level (-2.4080). The greater the coefficient the more heavily outcome is positively affected (more likely to have heard or seen Canada's food guide), while negative coefficients have the opposite effect.

Model diagnostics included verifying the AUC-ROC of the model as well as the prediction error. The area under the curve (AUC) and the curve called receiving operating characteristics (ROC) were employed to assess the performance of the model (Wiley & Wiley, 2019, Steen, 2020). This specific instrument is created for instances where predictors are ordinal, a type of categorical data. This tool was used without a validation using training/test data split because the initial sample size was greatly reduced and no section of the data could be spared for testing. As detailed in Appendix 5, the AUC-ROC of the final model is 0.88. Finally, a prediction error was calculated as shown in Appendix 6 and found to be 1.077413.

Propensity Score Regression Analysis

While the main logistic regression analysis focused on factors predicting awareness of Canada's Food Guide, the purpose of the PSM is to determine whether those who have heard or seen CFG have marked differences in their dietary habits. Essentially, this measures to a certain extent the effectiveness of the food guide as well as how well Canadians adhere to its suggestions. Based on the available data, the assigned treatment was 'heard or seen Canada's food guide' and the outcome observed was 'fruit juice consumption'. As mentioned previously, the 2007 CFG wrongfully suggested to Canadians that fruit juice could replace actual fruits and vegetables. Therefore, proper adherence to the food guide could suggest people to consume plenty fruits and vegetables which would result in people increasing their juice consumption. The analysis lies on the assumptions that propensity score matching can adjust for the differences between those who are not aware of CFG and those who are; therefore, any differences in fruit juice consumption can be attributed to the lack of or knowledge regarding the food guide effectively highlighting the 'true treatment effect'. Overall, there were 180 matches in the data. The propensity regression model is as follows:

$$\begin{aligned}
\hat{y} = & 0.485 + 0.226 * \text{Hear_or_seen_Canadian_food_guide}_{no} + 0.104 * \text{Sex}_{Female} \\
& + 0.714 * \text{Age}_{Young_adult} + 0.144 * \text{Age}_{Adult} + 0.016 * \text{Age}_{Middle_age} + 0.0254 * \text{Age}_{Old} \\
& - 0.109 * \text{Education}_{Secondary_school_graduation_no_post-secondary_education} \\
& - 0.213 * \text{Education}_{Post-secondary_certificate_diploma_or_univ_degree} \\
& - 0.023 * \text{Birth_country}_{Other} - 0.164 * \text{Household_income}_{20,000_to_39,999} \\
& + 0.485 * \text{Household_income}_{40,000_to_59,999} - 0.215 * \text{Household_income}_{60,000_to_79,999} \\
& - 0.311 * \text{Household_income}_{80,000_or_more}
\end{aligned}$$

The summarized output from this formula will detail the significance of the predictors. Specifically, \hat{y} is the predicted amount of fruit juice consumed while the coefficients of the variables indicate the magnitude and direction of the variables' effects on juice consumption. Essentially, should 'heard or seen Canada's food guide' be significant then it could be reasonably inferred that knowledge of the food guide caused individuals to consume fruit juice differently.

RESULTS

The model summary (Appendix 2) details the coefficients of each variable. Women are less likely to be aware of the food guide with a coefficient of -1.2994 compared to men. Additionally, while the coefficients seem similar between age groups, it appears as though Adults (aged between 25 and 39) are most likely to know about the guide. Furthermore, those born outside of Canada have a much higher chance of being aware of the guide with a coefficient of 2.3558. Finally, the overall trend suggests that those in households with lower total income are also more likely to have heard or seen of the food guide with coefficients of 0.2701 and -0.5151 for income brackets 20 000 to 39 999 and 40 000 to 59 999 respectively. An additional analysis was run using Canada Food Guide usage variables and results are presented in the following pie chart. Of those who report using Canada's Food Guide, 7% use it to choose their food; 7% to determine the quality of food needed; 8% to plan meals or groceries; 15% to assess how well they and their families are eating; and 62% report more than one of the above reasons. This data and information will also be useful in the discussion on the effectiveness of the food guide as well as in the creation of more effective outreach strategies. Moreover, through the data cleaning process of the analysis, it was also determined that in the subset of those who responded either 'yes' or 'no' to having heard or seen Canada's Food Guide, all reported no difficulty accessing health services; no unmet health needs; yes to experiencing some form of distress; and no depression. Subsequently in the PSM, 'heard or seen Canada's Food Guide' was found to not be a significant predictor of fruit juice consumption with a p-value > 0.05 (Appendix 7). Finally, it is also important to consider that of those who reported to having seen or heard of Canada's Food Guide, approximately 59% also report using the food guide (calculations presented in Appendix 8).

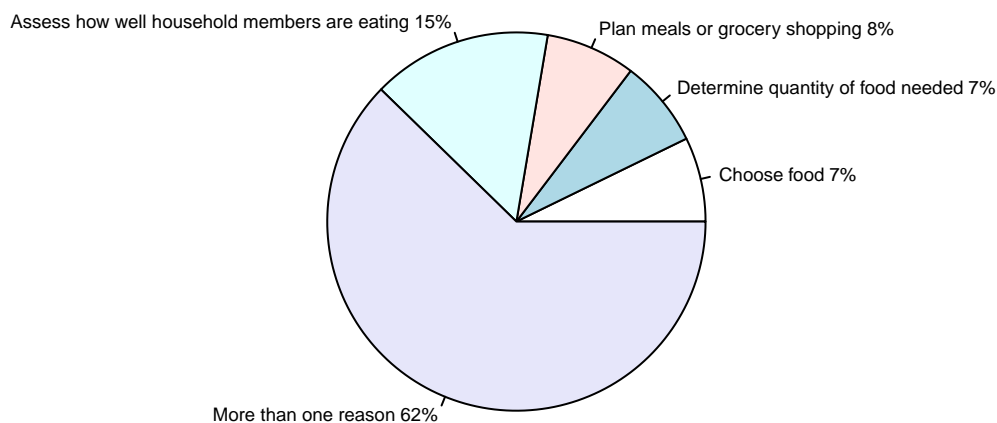


Figure 1: Distributions of Canadian's Food Guide Usage

DISCUSSION

Summary

Firstly, the survey data was retrieved as a subset of the 2017-2018 CCHS. The initial variables of interest were numerous and included factors relating to mental health, food security, access to healthcare as well as general demographic information like income, age, and sex. The variables relating to food guide use and dietary habits were interesting and saved for supplemental analysis. While observations with NA as the response were eliminated, insignificant variables or variables with too few response levels were also eliminated. Subsequently, the age category was simplified into 4 broader age brackets. An initial model was constructed, each predictor's significance was examined, and then the model was simplified using stepwise backwards selection. The final model included age, sex, household income, country of birth, and education variable and appendix 1 shows the coefficients of the model. Additionally, supplemental analyses included PSM, which evaluated the relationship between awareness of CFG and increased consumption of fruit juice; a brief calculation of the food guide usage; and an examination of CFG's awareness-to-usage rate. The association between CFG awareness and fruit juice consumption was not significant in propensity score regression. Finally, the proportion of those who heard or seen the food guide and actually used the food guide was calculated (Appendix 6) and Of those who claimed to use the food guide, a pie chart, which presented how the guide is generally used, was generated.

Conclusions

When it comes to factors predicting awareness of Canada's Food Guide, through the logistic regression it is suggested that lower household income and being born outside of Canada are factors associated with increased likelihood of being aware of CFG. Variations between age groups, gender, and education attainment while present are not as great in magnitude. Nonetheless, it is surprising that being born outside of Canada and earning lower household income are associated with CFG awareness. A possibility explaining that being born outside of Canada is associated with increased awareness is that those arriving from foreign countries have greater incentive and motivation to learn about Canada and assimilate into the country. On the other hand, lower income families may be incentivized by the health benefits of food guide diets; that is, they may not be able to afford the resources to undergo various health treatments and are therefore more cautious of their dietary habits. Furthermore, in the data cleaning process, various variables like access to health services, unmet health care needs, distress, depression were found to only contain one level of response; that is, all respondents regardless of whether they were aware or not of CFG had that same answer for that category. This is important to consider because factors that do not affect awareness are eliminated from consideration. Indeed, since all responses for unmet health care needs and for difficulty accessing health care were 'no', then access and awareness to CFG is not inhibited by any limitations of the healthcare system. Similarly, distress and depression also do not factor into awareness of the food guide. Based on this information, it is suggested that food guide access is not generally restricted, after all it is a free online resource; however, populations with incentive - financial or social - express increased awareness. From this logisitic regression model analysis, it then becomes evident that there are no apparent factors limiting awareness regarding CFG; however, groups can be pushed to educate themselves. The PSM and the first part of the supplemental analysis illustrated that awareness of CFG did not cause any changes in dietary habits. This suggests a low adherence rate to the food guide. The exact reasons, though, for the low adherence rate are unclear. In the second part of the supplemental analysis, it was observed that most of those who did utilize the guide did so for many reasons. This suggests that when adherence is active, it is consistent and relevant to people's lifestyles. Finally, the awareness-to-usage rate of the guide was calculated and found to be 59%, slightly greater than a coin toss. These results together suggest that (1) people can be incentivized financially and socially to educate themselves; (2) the food guide has a low adherence rate; and (3) when people do adhere to the guide, they do so consistently and for many different reasons. Given the suggestions 1,2, and 3, it becomes evident that the food guide must become not only more relevant to Canadians individually, which can increase interest and adherence, but also more practical and easier to adhere to. Based on these findings, future outreach programs of the CFG should aim to convince individuals and families of its practicality in the context of every culture's habits as well as the immense health benefits of adopting evidence-based dietary practices. Moreover, the guide should also implement broader and easier-to-follow suggestions as well as personalized suggestions for different groups of people. These adjustments

would not only make the guide more attractive and interesting but also increase adherence. Interestingly, many of these changes have already been adopted by the updated 2019 guide; that is, the guide is simpler, offered in many languages, caters to all cultures, and encourages easy daily changes in habits that improve diet and health. Indeed, the food categories were abandoned and now Canadians are encouraged to follow a general guideline when selecting foods (e.g. eat more vegetables, decrease sugar, have meals with loved ones, drink more water, move towards plant-based proteins, etc.). However, outreach can still be improved and Canadians should be pushed to learn of the health consequences of improper diets. Additionally, in the age of Internet, many still choose to find information from non-reputable sources regarding diet and lifestyle improvements which can not only negatively impact population health, but also decrease traffic towards governmental resources. To address this, government of Canada should increase its online presence through larger more comprehensive marketing programs.

Limitations and Next Steps

In this analysis, a major limitation is the analyzable sample size. Indeed, while the survey collected responses from over 110 000 individuals, after removal of NA responses from the selected variables, only approximately 800 observations remained. While the sample size was statistically sufficient for analysis, a larger sample size may increase the confidence in the model. Furthermore, a reduced sample size also drastically reduced the number of variables that could be included in the model, thereby reducing the number of factors that could be considered; however, this was slightly compensated for by a discussion of variables like access to health services, unmet health needs, distress, and depression. The PSM was also further limited by the reduced sample size. Matching for propensity score further reduced the sample reducing the confidence in the results. Additionally, a general limitation of logistic regression models are the defined levels of each categorical variable considered. That is, age was broadened to four conventional categories; however, these categories remain arbitrary and consequently the meaningfulness of our model and results are reduced. In terms of PSM, the analysis is limited due to only adjusting for observed covariates (many hidden factors could be affecting the model) and the fact that the sample was so small that likely there was no distributional balance of the observations. All these factors, including bias, reduce the meaning and confidence of the results. Further, another limitation in the context of the purpose of the study is that the 2017-2018 CCHS could only measure food guide responses regarding the 2007 guide. This information is effective for predicting possible outreach issues of the 2019 guide as well as inform improvements however, future studies should focus on the limitations and factors affecting awareness of the 2019 food guide. The data should become available in the 2019-2020 CCHS which should publish within the next year. Also, future studies should aim to increase sample size and responses regarding CFG questions. This would increase the generalizability of the findings, ensure distributional balance when performing PSM, and allow for the analysis of more variables in the models (Caetano, 2020); all of which lead to more precise conclusions and better informed policies. Finally, in the same context, other studies could examine food guide knowledge and adherence with surveys specifically designed to gather information regarding knowledge of the food guide and respondents' characteristics and dietary habits.

REFERENCES

1. “The 10 leading causes of death, 2011.” StatsCan. 2011. <https://www150.statcan.gc.ca/n1/pub/82-625-x/2014001/article/11896-eng.htm>
2. “History of Canada’s Food Guides from 1942 to 2007.” Government of Canada. 2007. <https://www.canada.ca/en/health-canada/services/canada-food-guide/about/history-food-guide.html>
3. “Canada’s Food Guide.” Government of Canada. 2019. <https://food-guide.canada.ca/en/>
4. Wiley, M., & Wiley, J.F. (2019) Advanced R statistical programming and data models: Analysis, machine learning, and visualization. New York: APRESS.
5. “Understanding the ROC Curve and AUC.” Doug Steen. TowardsDataScience. 2020. <https://towardsdatascience.com/understanding-the-roc-curve-and-auc-dd4f9a192ecb>
6. Caetano, SJ. (2020) Propensity Score Matching - Theory. [Video]. Retrieved from <https://web.microsoftstream.com/video/53989ba9-82ec-4fcb-96bf-56d2e3cc3f9c>
7. Caetano, SJ.(2020) Propensity Score Matching - R Code. [Video]. Retrieved from <https://web.microsoftstream.com/video/55020107-a6a6-4b01-b926-957735d62c33>

APPENDIX

Appendix 1

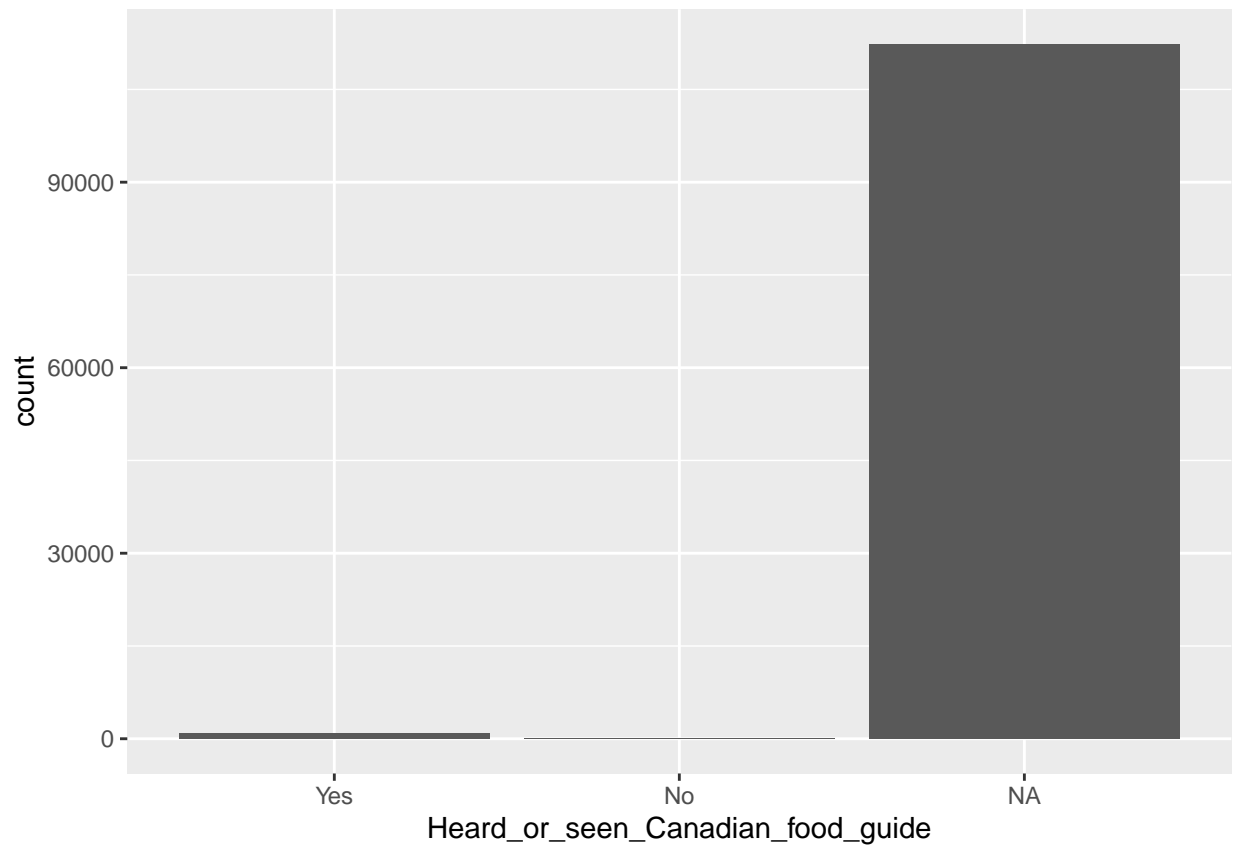


Figure 2: Distribution of People Having Heard or Seen CFG

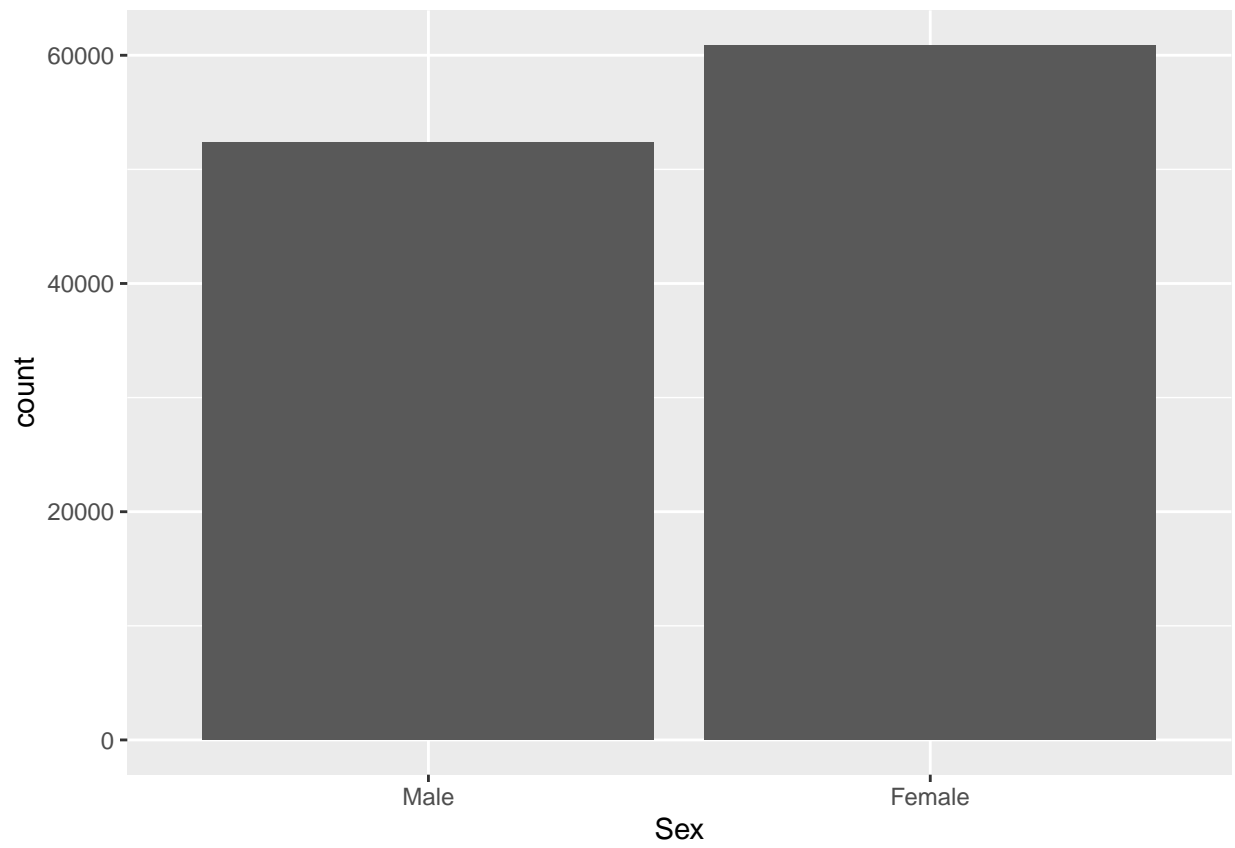


Figure 3: Distribution of Gender

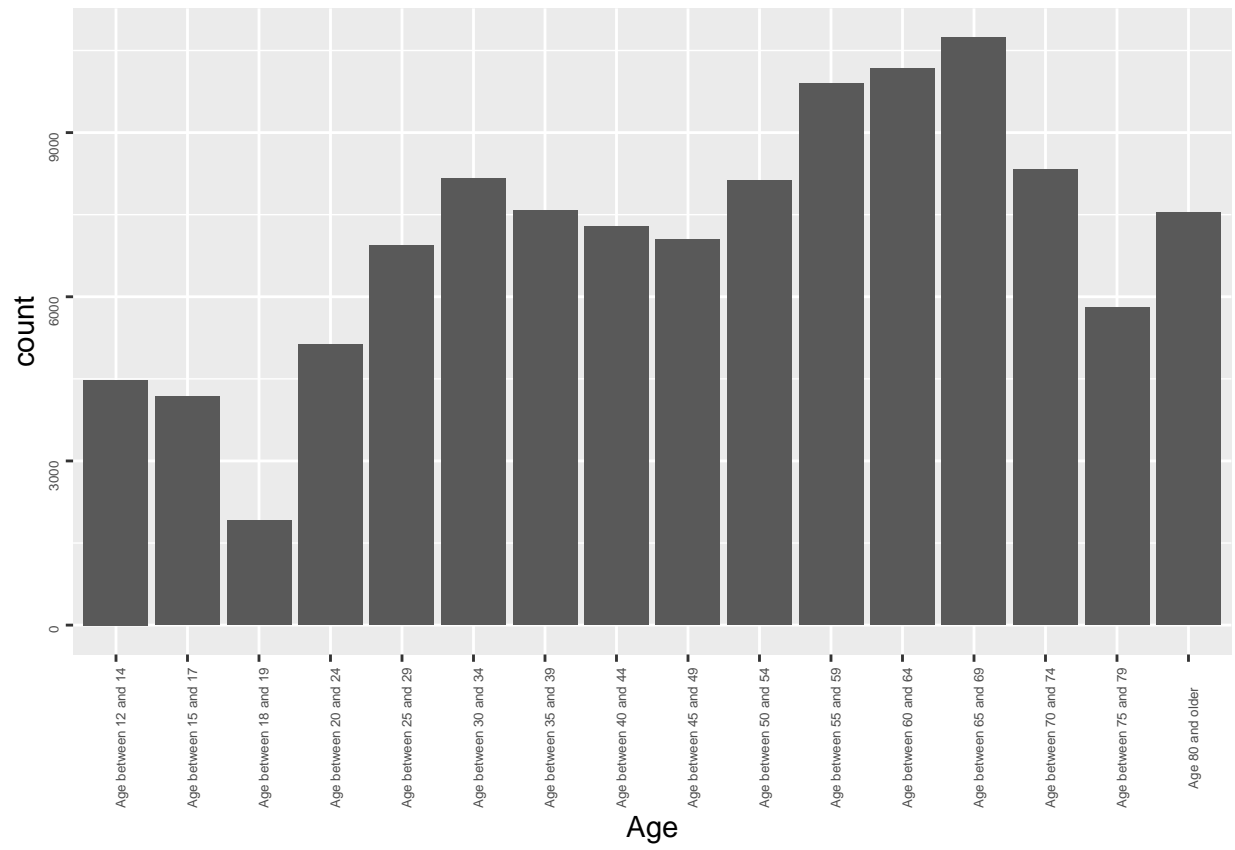


Figure 4: Distribution of Age

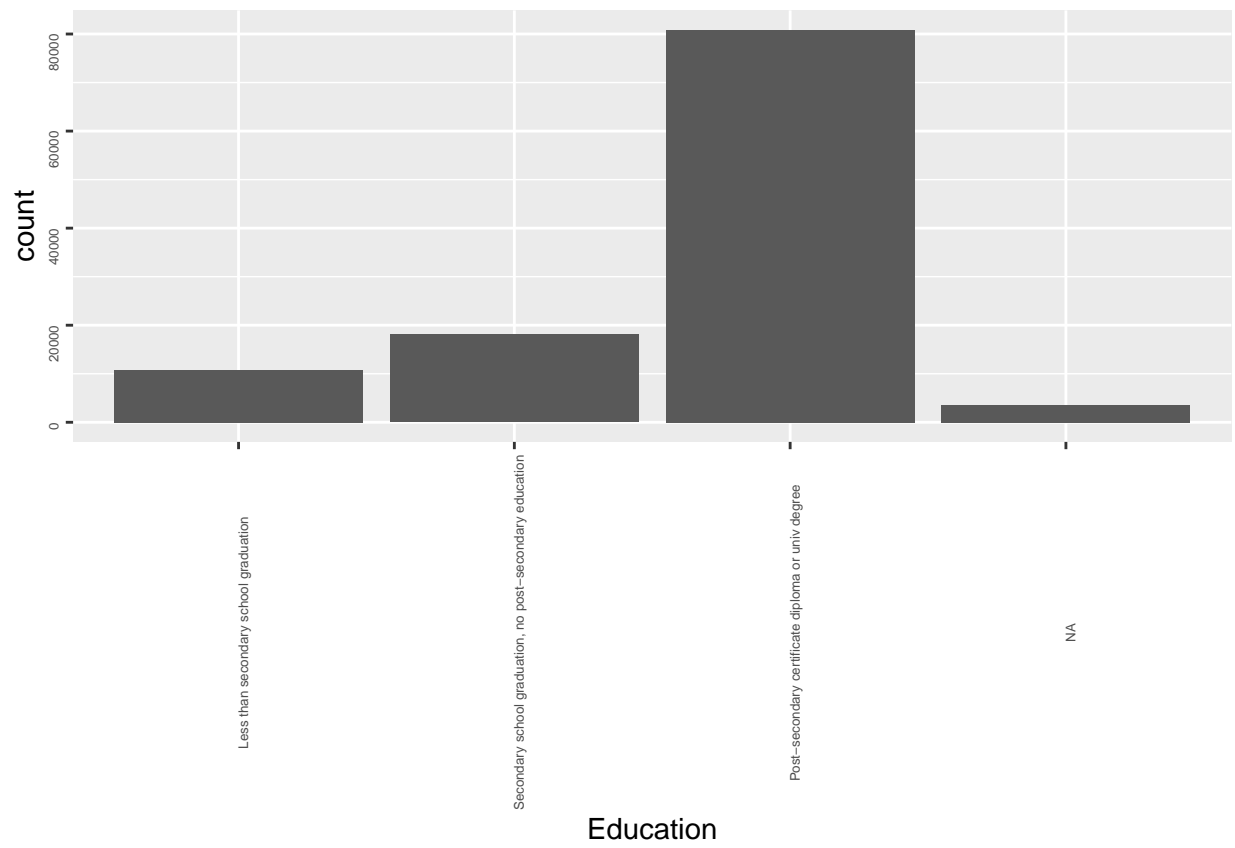


Figure 5: Distribution of Education Attainment

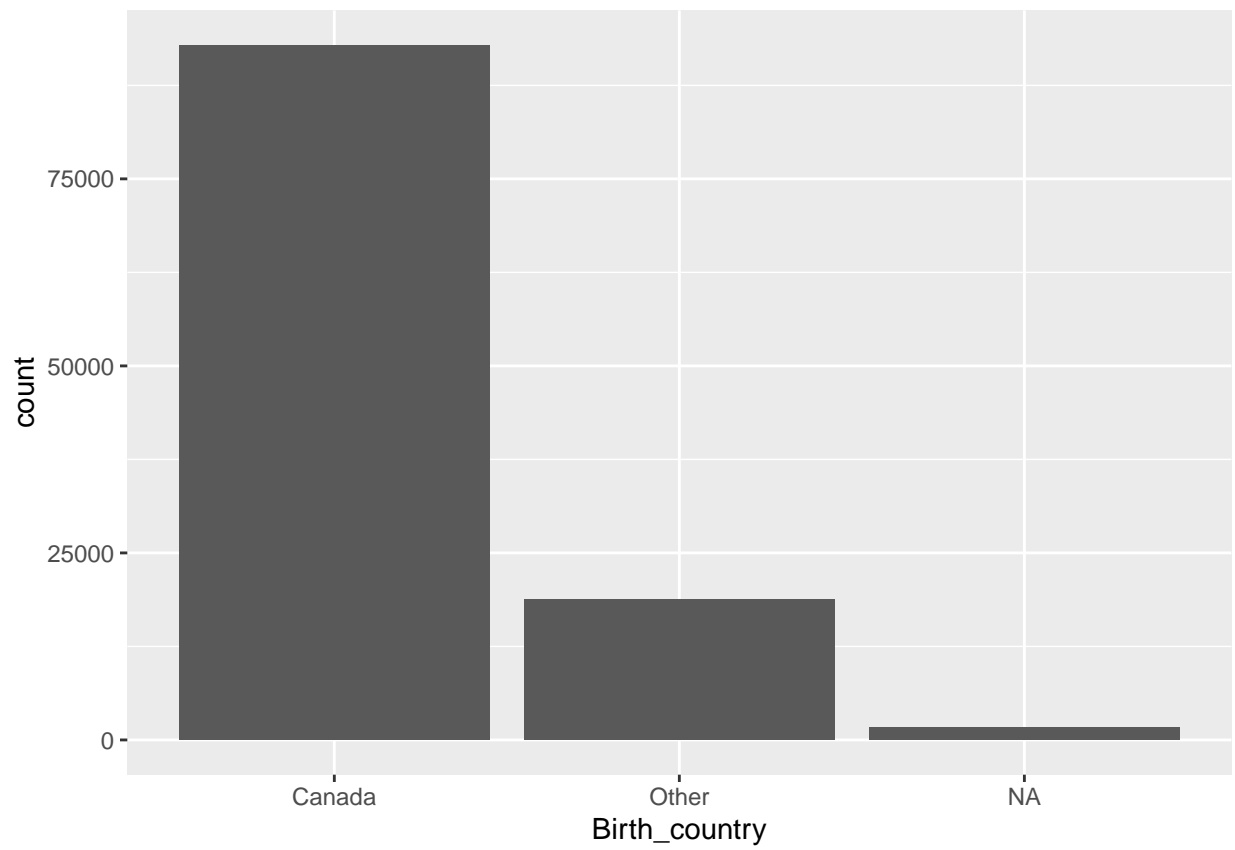


Figure 6: Distribution of Country of Birth

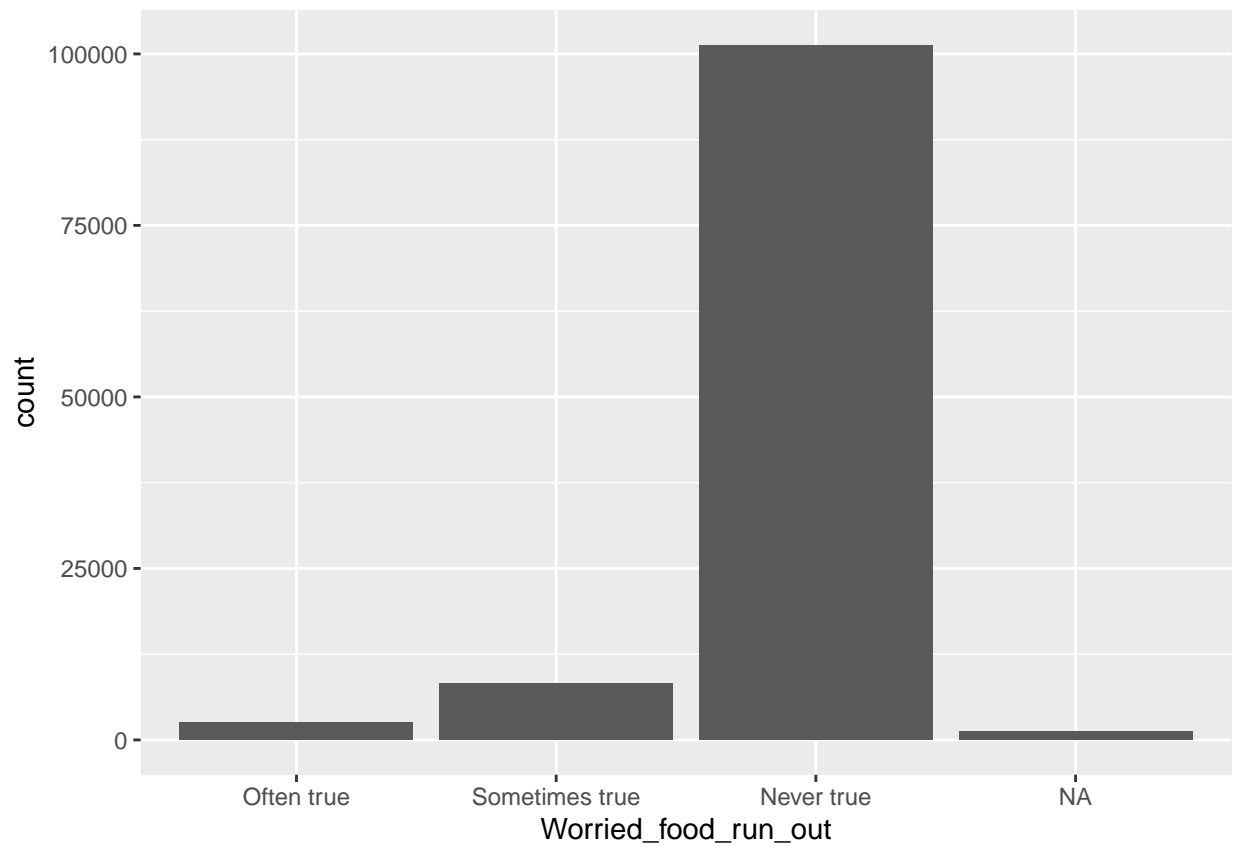


Figure 7: Distribution of Those Who Have Been Worried Food Would Run Out

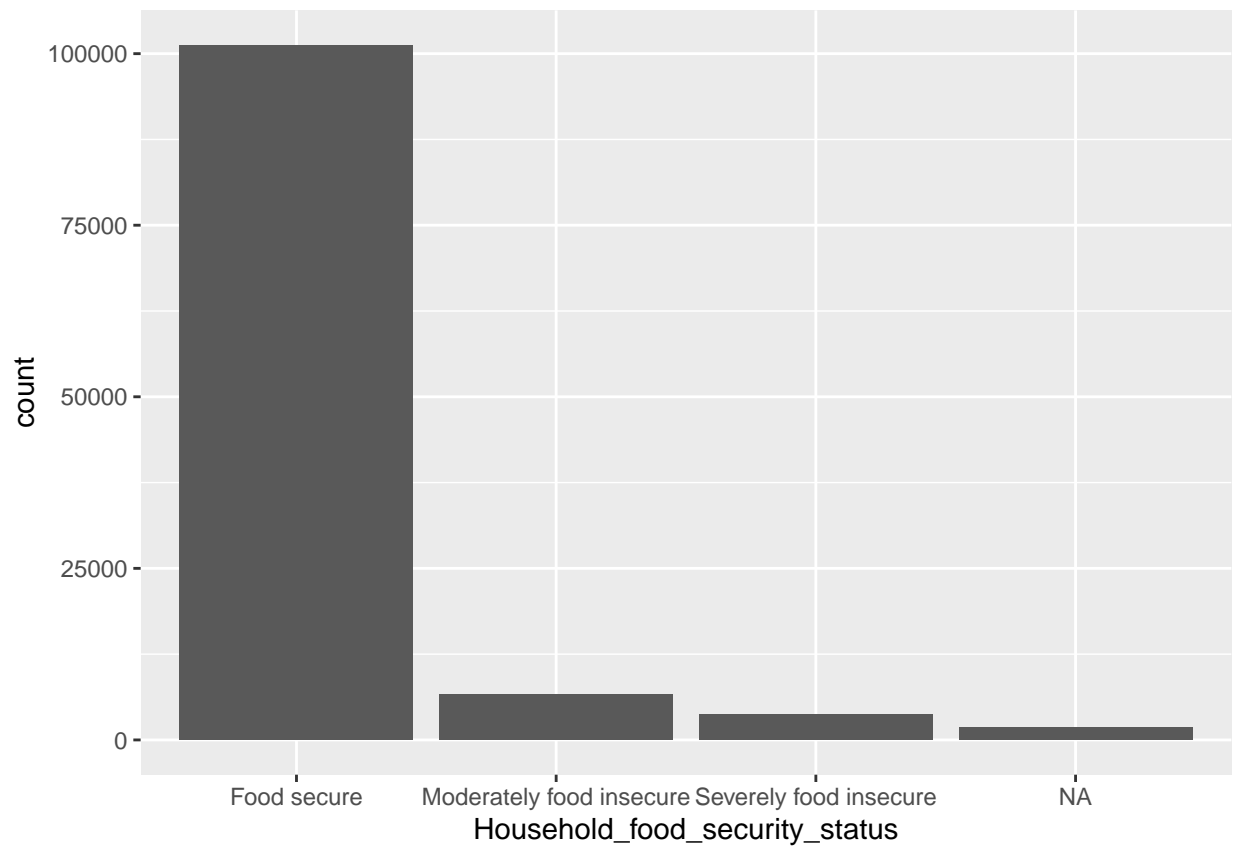
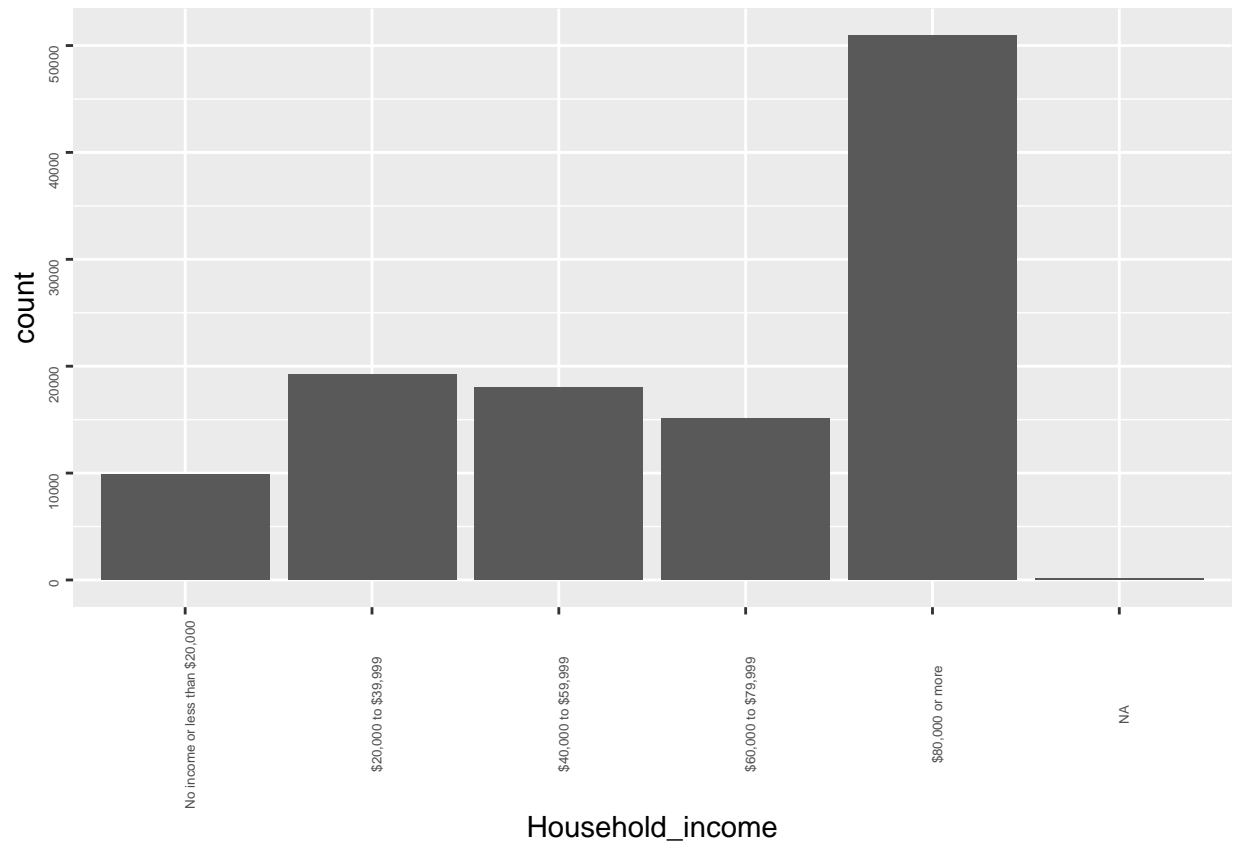


Figure 8: Distribution of Food Security



Appendix 2

Table 2: Pvalues for Categorical Variables Associations

	Heard_or_seen_Canadian_food_guide
Sex	0.0000004
Age	0.0000000
Education	0.0000080
Birth_country	0.0000000
Worried_food_run_out	0.7582000
Household_food_security_status	0.2942000
Household_income	0.0000509

Appendix 3

Table 3: Pvalues for Categorical Variables Associations

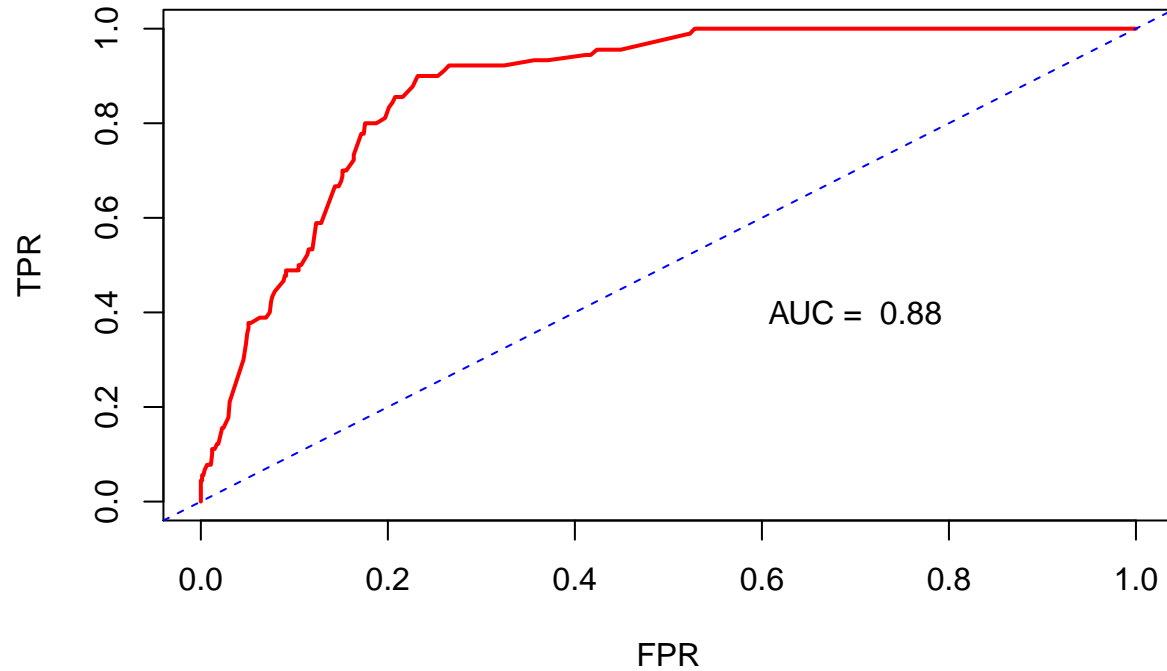
	P.Values
Model delete Sex	0.0000020
Model delete Age	0.0000000
Model delete Education	0.0001064
Model delete Birth_country	0.0000000
Model delete Household_income	0.0000596
Model delete Worried_food_run_out	0.5548000
Model delete Household_food_security_status	0.2163000

Appendix 4

Table 4: Summary Table for Final Model

term	estimate	std.error	statistic	p.value
(Intercept)	2.2886397	0.6123547	3.7374414	0.0001859
SexFemale	-	0.2745621	-	0.0000041
AgeYoung Adult	1.2648304	-	4.6067195	-
AgeAdult	-	0.5455746	-	0.0000091
AgeMiddle Age	2.4207082	-	4.4369886	-
AgeOld	-	0.4371813	-	0.0000001
EducationSecondary school graduation, no post-secondary education	2.3699299	-	5.4209310	-
EducationPost-secondary certificate diploma or univ degree	-	0.4465563	-	0.0000000
Birth_countryOther	2.9558770	-	6.6192703	-
Household_income\$20,000 to \$39,999	-	0.4338856	-	0.0000000
Household_income\$40,000 to \$59,999	2.5912659	-	5.9722325	-
Household_income\$60,000 to \$79,999	-	0.4831848	-	0.0038145
Household_income\$80,000 or more	1.3979040	-	2.8931040	-
	-	0.4062551	-	0.0000294
	1.6974039	-	4.1781723	-
	2.3790269	0.3191996	7.4531023	0.0000000
	0.2434029	0.5005641	0.4862572	0.6267848
	-	0.5470516	-	0.3592239
	0.5015615	-	0.9168450	-
	-	0.8946598	-	0.0017053
	2.8067926	-	3.1372737	-
	-	0.4800580	-	0.0727519
	0.8614085	-	1.7943842	-

Appendix 5



```
## integer(0)
```

Appendix 6

```
#Check predictions
data$logodds_estimate <-
  model_final%>%
  predict(newdata = data)
data$estimate <- exp(data$logodds_estimate)/(1+exp(data$logodds_estimate))
#Prediction error
mean((as.numeric(data$Heard_or_seen_Canadian_food_guide) - data$estimate)^2)
```

```
## [1] 1.077413
```

Appendix 7

	names	model1
		(1)
1	(Intercept)	0.485186012463352
2		(0.663236646438958)
3	Heard_or_seen_Canadian_food_guide0	0.225798405194985
4		(0.346732288595162)
5	SexFemale	0.103767594835128
6		(0.155175240706078)
7	AgeYoung Adult	0.714405428102594 *
8		(0.351269588834526)
9	AgeAdult	0.144296131562644
10		(0.30033041330011)
11	AgeMiddle Age	0.015681531860326
12		(0.301452925784237)
13	AgeOld	0.253968391566211
14		(0.306120434302033)
15	EducationSecondary school graduation, no post-secondary education	-0.108771687491145
16		(0.314037162501393)
17	EducationPost-secondary certificate diploma or univ degree	-0.212987677955884
18		(0.278107555718147)
19	Birth_countryOther	-0.022843224685719
20		(0.232881779573302)
21	Household_income\$20,000 to \$39,999	-0.164351340042309
22		(0.36872601832162)
23	Household_income\$40,000 to \$59,999	0.485079412658192
24		(0.438318630803479)
25	Household_income\$60,000 to \$79,999	-0.215295234324262
26		(0.472159647834527)
27	Household_income\$80,000 or more	-0.311070157915044
28		(0.400837648005642)
1.1	N	180
2.1	R2	0.136153190229606
3.1	logLik	-226.788430839353
4.1	AIC	483.576861678706
.1	*** p < 0.001; ** p < 0.01; * p < 0.05.	

Appendix 8

```
#Proportion of seen/heard who use CFG
a <- sum(data$Heard_or_seen_Canadian_food_guide==1)
#Data_sub contains the number of individuals who reported using CFG
443/a
```

```
## [1] 0.5938338
```