

Men and past smokers from ontario are potentially more physically active than women and non-smokers*

Women are engaging in lesser physical activities than men. Furthermore, individuals who smoked before are more likely to engage in exercise. Consumption of stimulant-related drugs could have also increased as well, resulting in increased overall physical activities in Ontario.

Chris Yong Hong Sen

November 26, 2024

This study explores factors influencing physical activity levels in Ontario, focusing on demographic and substance abuse. Our findings reveal that women engage in less physical activity than men, while individuals who have smoked in the past and those consuming stimulant-related drugs are more likely to exercise. Additionally, age, highest educational attainment, and health region also play significant roles in determining physical activity levels. These insights highlight the complex interplay of gender, lifestyle behaviors, and sociodemographic factors, providing valuable information for designing targeted public health interventions to promote physical activity.

1 Introduction

Overview paragraph

Estimand paragraph The estimand would be the actual effect of socio-demographic factors, substance abuse, and health factors on how much time is spent on strenuous activities among individuals in Ontario who engage in regular exercise.

Results paragraph

Why it matters paragraph

*Code and data are available at: https://github.com/Monoji77/Alcohol_Use_Ontario.

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2....

2 Data

2.1 Context of dataset

The dataset chosen is obtained from Computing in the Humanities and Social Sciences/CHASS at the university of Toronto (Data Centre, Faculty of Arts & Science, University of Toronto 2018), specifically the Canadian Community Health Survey (CCHS) annual component 2017-2018 which was collected by Statistics Canada (2018). The goal of the CCHS survey is to collect comprehensive and reliable health-related data on the Canadian population to support health surveillance, research, and program evaluation. It aims to provide timely, accessible, and flexible data for monitoring health trends, studying small populations and rare characteristics, and addressing emerging health issues, ultimately helping improve the health and well-being of Canadians.

Other datasets explored include sources from World Health Organisation and ParticipACTION. World Health Organisation provides aggregated data informing certain indicators of interest such as obesity levels and alcohol consumption. This can be found in the data section for Canada (World Health Organization 2024). Also, ParticipACTION provides key statistics regarding proportion of adult population in Canada meeting the national guidelines of 150 minutes of moderate to vigorous exercise, as well as how sedentary Canadians are (ParticipACTION 2024). However, the availability of raw individual level data is absent from these sources. Although the aggregated data gives us a sensing about general Canadian health, we are not able to focus our attention to the Ontario population. The dataset chosen from Canadian Community Health Survey (CCHS) annual component is ideal as it contains individual level data, where we are able to obtain the province the respondents are from.

2.2 Original Dataset

The dataset provides 113,290 individual response from Canadians all throughout the country. Surveys were conducted to collect health-related data from respondents. The relevant data collected includes...

- Demographics and Identity: Age, gender, marital status, citizenship, Aboriginal identity, visible minority status, immigrant status, language (mother tongue, spoken at home, first official), sexual orientation.
- Health Status and Conditions: Chronic diseases, neurological conditions, mental health indicators, perceived physical and mental health, pain, discomfort, activity limitations, and disabilities.

- Health Behaviors and Lifestyle: Smoking, drinking, cannabis use, physical activity, helmet use, dietary practices (e.g., fruits, vegetables), breastfeeding, and BMI.
- Health Care Access and Utilization: Contact with health professionals, access to health services, immunizations (e.g., flu shots), cancer screenings (e.g., mammograms, pap smears), waiting times, and satisfaction with services.
- Socio-Economic Factors: Educational attainment, work activity, household income, living arrangements, presence of children in the household, food insecurity.
- Life Quality and Perceptions: Life satisfaction, life stress, sense of belonging, health-adjusted life expectancy.
- Specialized Health Indicators: Two-week disability days, quality and ratings of health care services, functional health status.
- Specific Population Measures: Aboriginal group, immigrant duration, visible minority, minority health disparities.
- Geographic and Household Data: Geographic location, number of persons or households, household demographics.

This dataset is highly suitable in determining both quantitative factors such as c qualitative factors

2.3 Measurement

Our goal is to explore the actual effect of socio-demographic factors, substance abuse, and health factors on how much time is spent on exercise among individuals in Ontario who engage in regular exercise. The instrument of measurement would be surveys. Through the cross-sectional survey conducted by CCHS, we are able to get the number of hours spent on vigorous exercise per individual, our response variable.

According to the questionnaire, respondents were asked in a phone survey to indicate whether they had engaged in any strenuous activity that lasted at least 10 minutes and made them sweat and breathe a little harder in the past 7 days. Afterwards, they were asked to indicate the total time spent on those activities. To address the concern of recall bias where an individual may not recall the specific timing in engaging with exercise, surveyors sectioned vigorous exercise into commuting, sports, work and volunteering. Then within each section, surveyors asked which day (Monday to Sunday) in the past week had they engaged in such exercise that made them sweat and breathe a little harder. By doing so, the activities individuals did during the past week would be anchored on the days of the week. Following this would be for the respondents to indicate the total number of hours they engaged in each section of vigorous exercise for the past 7 days.

In addition, the explanatory variables we are interested in were also obtained in the same survey.

For socio-demographic factors possibly affecting time spent on vigorous exercise, the surveyors asked the respondent for their age (DHH_AGE) from 0 years old to 121 years old), sex (DHH_SEX) being either male or female, personal income obtained from tax forms submitted to the Canada Revenue Agency upon consent from respondent, geographical health region (ADM_Q037) that was provided by respondents when they agreed to the phone survey, and highest educational attainment (EHG2_04) ranging from less than high school diploma to master's degree and above.

For substance abuse, respondents were asked how frequent they consumed alcoholic drinks in the past 12 months (ALC_Q015), ranging from less than once per month to everyday. Surveyors also asked respondents for usage of drugs including one-time marijuana or hashish in the past 12 months (DRG_Q005 to DRG_Q075), and whether they have smoked hundred cigarettes in the past 12 months (SMK_Q020). The last 2 questions allowed respondents to indicate a simple yes or no flag to express their answer.

For health factors, questions were asked to determine an individual's body mass index ranging from underweight to obese and perceived life/work stress ranging from not at all stressful to extremely stressful. Additionally, surveyors asked respondents if they have mood disorders (E.g. depression, bipolarism, mania), for which respondents replied with a simple yes or no.

Overall, each entry obtained by respondents assigns quantitative values to each factor and the the response for which we can explore subsequently using linear regression to determine which factors have greater influence in the time spent exercising amongst individuals that are engaged in exercising.

2.4 Response variable: Time spent on vigorous activity in the past 7 days

From the CCHS dataset, we look at paadvmlva variable. This variable is continuous with values ranging from 0 to 9902 representing the number of minutes spent on vigorous activity in the past 7 days. For easier readability, the variable will be renamed as time_spent_vigorous_exercise_7d. As explained in section Section 2.3, this variable measures the time spent on any activities lasting more than 10 minutes that makes an individual sweat more than they normally would. This would include individuals undertaking laborious employment or volunteering efforts, or it could simply include individuals who do cardio and/or strength training.

By using this variable, we are assuming that the routine of an individual from the past 7 days is indicative of their general routine beyond the past 7 days. This variable may not account for individuals who only began vigorous sweating routine within the past 7 days and were sedentary before, which may not be helpful if included in subsequent model fitting.

Since we are interested in individuals that are already engaging in strenuous activities, we can filter `time_spent_vigorous_exercise_7d` to exclude 0. After removing such entries, we obtain the following summary statistics as shown in Table 1. The median time spent on vigorous activities the past 7 days would be 300 minutes, which averages out to 43 minutes per day.

Table 1: Summarised statistics for time spent on vigorous exercise in the past 7 days

Min.	1st Qu.	Median	3rd Qu.	Max.
10	135	300	600	9902

Interestingly, the maximum time spent on vigorous activities the past 7 days was 9902 minutes, and this averages to 1415 minutes per day, which is equivalent to 23.5 hours in a day. This is very unlikely as this suggests that an individual is constantly engaging in physical strenuous activity without sleeping. Logically, this suggests the possibility of outlier values. By observing Figure 1, there are many outlier observations for this variable. Roughly 10% of the observed data are considered outliers as they fall outside 1.5 times the interquartile range from the 25% and 75% quartile values. We would be performing log transformation which would be discussed in section (INSERT SECTION HERE PLEASE DONT FORGET) which would address these outlier values.

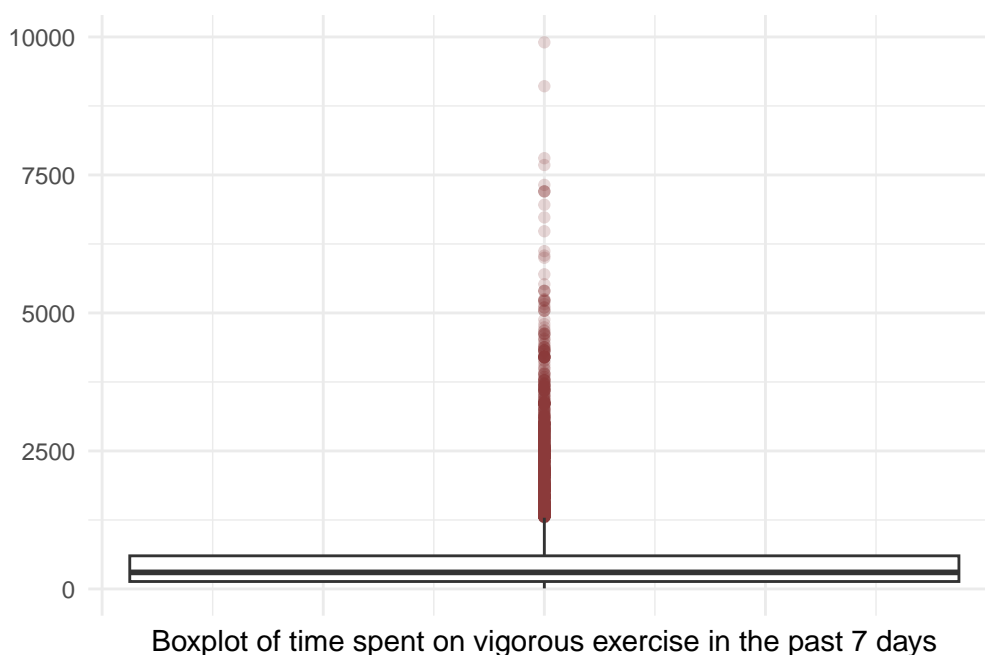


Figure 1: Boxplot of time spent on vigorous exercise in the past 7 days

Subsequently, we would have to remove observations that are very unlikely as they do not

represent the general population and could skew the distribution of time spent on vigorous exercise.

By observing the histogram generated in Figure 2, we obtain a severely right skewed response histogram where most observations lie between 0 to 500 minutes. For the building of linear models in subsequent sections of the paper, this would not be ideal as linear regression assumes a normal distribution of the response variable, suggesting that a transformation would be required to fit a linear regression model.

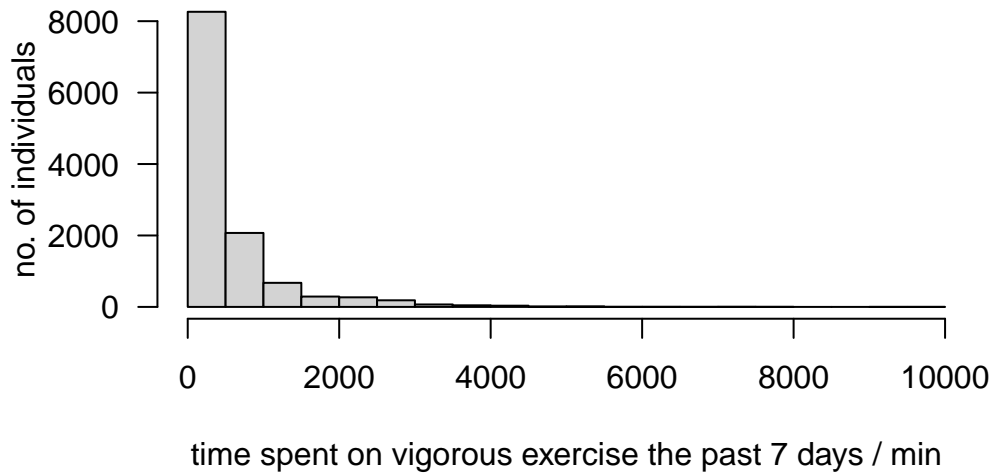


Figure 2: Histogram of time spent on vigorous exercise the past 7 days

2.5 Interested explanatory variables

There are a total of 13 predictors of interest that is seemingly related to the time spent in vigorous exercise an exercising individual gets.

2.5.1 Variable 1: Age

This is a variable of interest because age is seemingly correlated to The age groups of respondents were obtained and they exist in bins that spans ages 18 years to 74 years old.

- 3: 18-19 years old
- 4: 20-24 years old

- 5: 25-29 years old
- 6: 30-34 years old
- 7: 35-39 years old
- 8: 40-44 years old
- 9: 45-49 years old
- 10: 50-54 years old
- 11: 55-59 years old
- 12: 60-64 years old
- 13: 65-69 years old
- 14: 70-74 years old

From Figure 3, we can immediately see that most observations fall within age groups 30 to 34 years old (bin 6), and lesser observations fall within age groups 70-74 years old (bin 14).

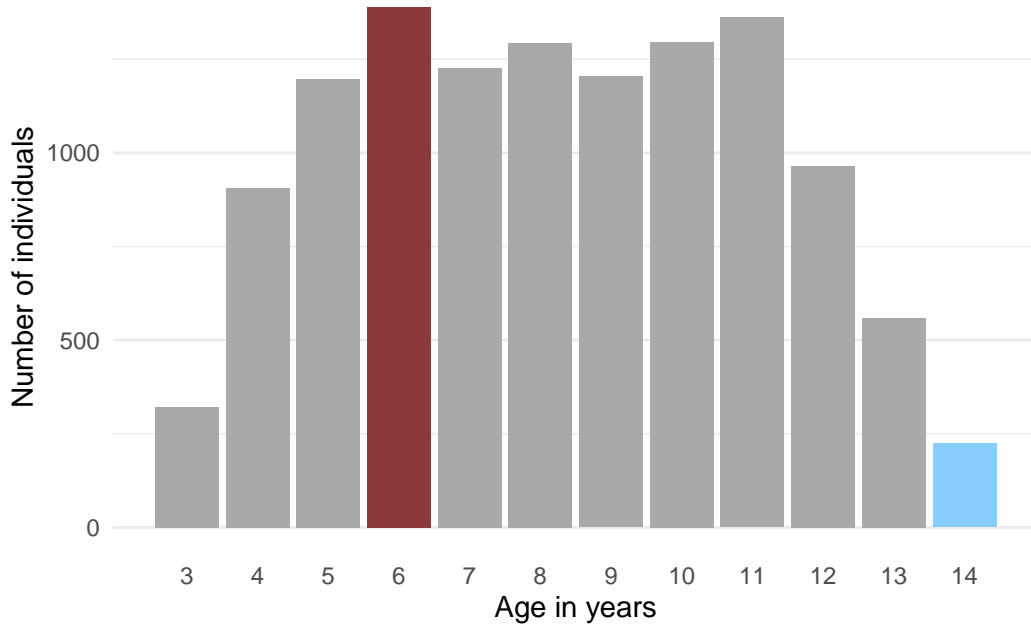


Figure 3: Histogram of time spent on vigorous exercise the past 7 days

From Table 2 and Figure 4, we are able to see that the the median age bin belongs to age group 40-44 years old. We are able to see that there are no outlier values from Figure 4, which means that the survey has obtained an acceptable number of respondents across all age bins.

Table 2: Summarised statistics of respondent ages

Min.	1st Qu.	Median	3rd Qu.	Max.
3	6	8	11	14

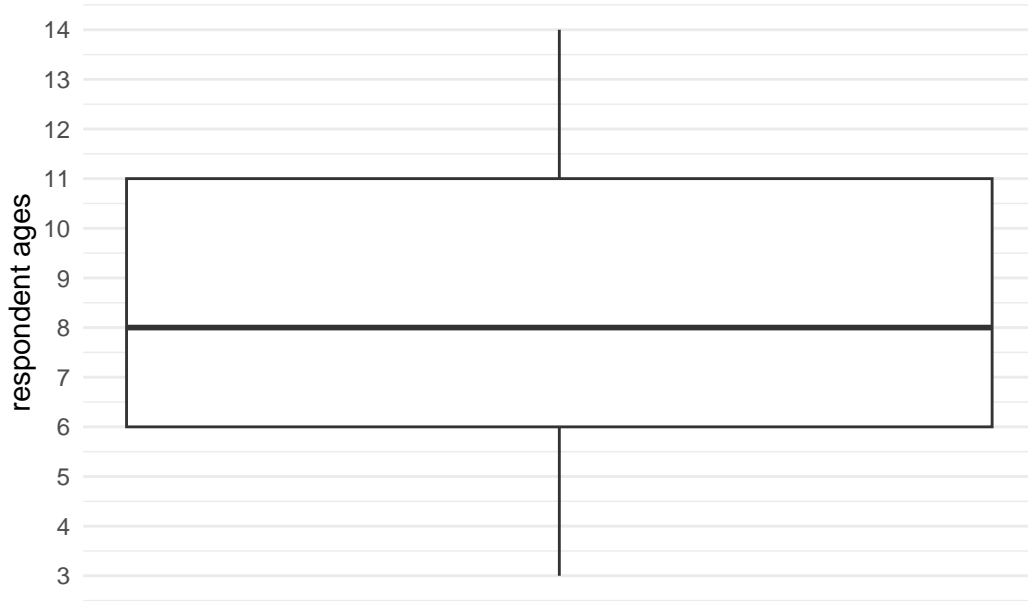


Figure 4: Boxplot of respondent ages

2.5.2 Variable 2: Sex

Sex is a possible influential predictor for total time spent on vigorous activity. Historically, men are associated with exercise and involved in more labour intensive roles such as construction which involves heavy lifting. According to World Health Organisation (2024), women are less active than men by 5% since 2000. However, with the advent of female only establishments such as female gyms, and 32% increase in memberships at fitness and health clubs from 2010 to 2019, it would be interesting to see if the sex of an individual still affects the time spent on vigorous activities among physically active individuals.

According to Figure 5, there is roughly an equal proportion of sexes from our dataset, with slightly more males (6082 respondents) than females (5859 respondents). Since this was a health survey conducted over phone call, there is an equal chance of a respondent to be male or female, which is roughly reflected in Figure 5.

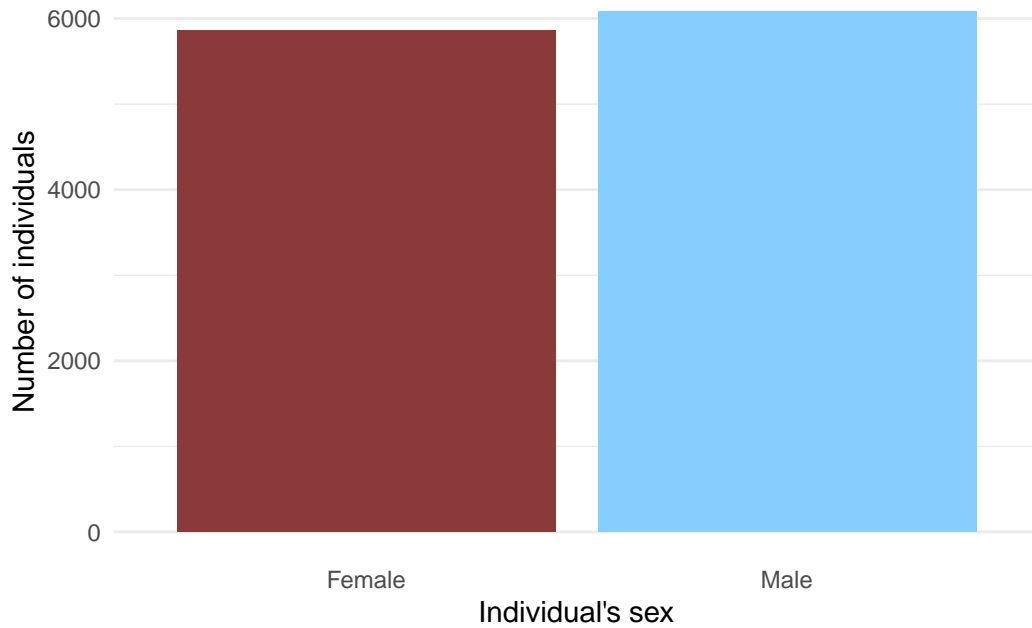


Figure 5: Barplot of sex

2.5.3 Variable 3: Highest level of education

In 2023, Kari et al. (2020) has done a study to show that highest educational attainment may be a leading factor for physical activity. Therefore, it would be worth exploring the effects of highest levels of education with the time spent on vigorous activities among the Ontario residents.

According to Table 3, respondents who did not graduate secondary school account for only 5% of the population, and respondents who have at least a bachelor's degree account for 72% of the population. According to Figure 6, there seems to be class imbalance for this variable.

However, as seen from post-secondary institution enrollment data from Statista Research Department (2024), there has been an increase in total number of Canadians enrolled in post secondary education from 2000 to 2020. Accounting for population changes from 2000 (The Daily, Statistics Canada 2000) to 2022 (Statistics Canada 2023), there is still a real increase in the number of real enrollments into post secondary education. Therefore, it is very possible that this is representative of the Ontario's population where majority of the population has at least post secondary education.

Table 3: Proportion of respondents grouped by highest educational level

highest_educational_attainment	proportion_of_respondents
less than sec sch grad	5%
sec sch grad	23%
post sec certification and above	72%

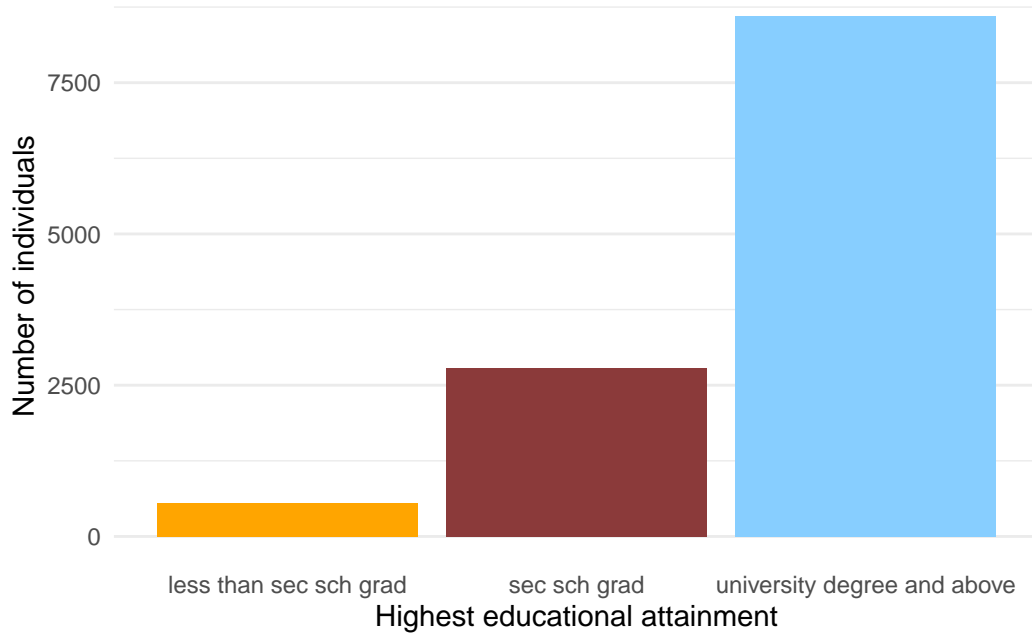


Figure 6: Barplot of highest_educational_attainment

2.5.4 Variable 4: Health Region

Health region is a categorical variable which tells us the geographical location an individual belongs to defined by Provincial Ministries of Health. It could be possible that certain health regions have more funding that are projected towards increasing awareness towards having an active lifestyle. Therefore, it is worth exploring the relationship between health region and time spent on vigorous activities.

There are a total of 34 different health regions in the Ontario province. They are encoded using 5 digits. Kindly refer to the appendix (INSERT APPENDIX HERE PLEASE DONT FORGET) to refer to all the various health regions. Figure 7 shows us the various health regions in Ontario. The blue regions shows us health regions for which lesser respondents belong to while red regions show us health regions for which more respondents belong to.

Highest occurring health regions

- 35951: City of Ottawa Health Unit
- 35953: Peel Regional Health Unit
- 35960: Simcoe Muskoka District Health Unit
- 35970: York Regional Health Unit
- 35995: City of Toronto Health Unit *(highest number of observations)

Lowest occurring health regions

- 35931: Elgin-St Thomas Health Unit
- 35935: Haliburton, Kawartha, Pine Ridge District Health Unit
- 35940: Chatham-Kent Health Unit
- 35942: Lambton Health Unit
- 35956: Porcupine Health Unit *(lowest number of observations)

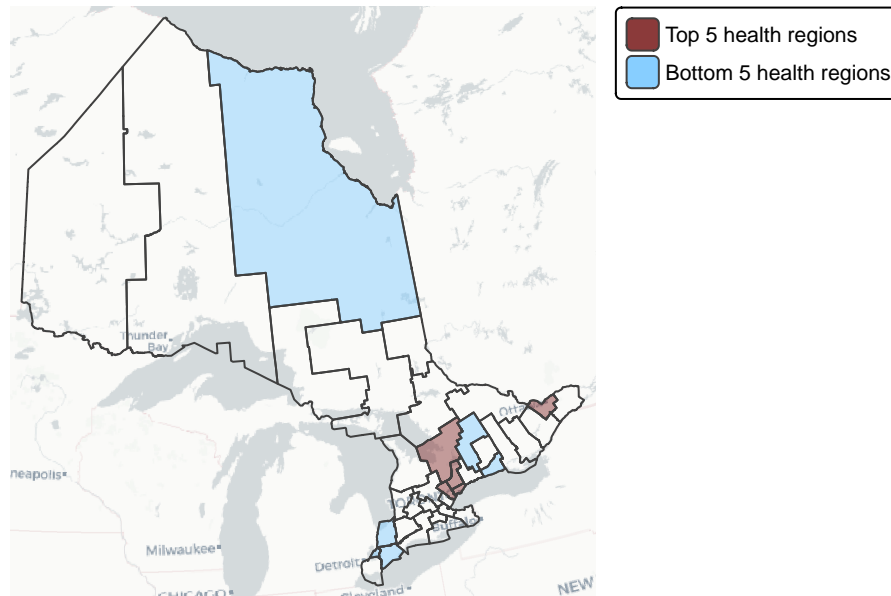


Figure 7: Map of top five and bottom five occurring Ontario health

Overall, Figure 8 shows us the actual number of observations belonging to the top five and bottom 5 occurring health regions. We can see that most respondents come from the City of Ontario Health Unit (Health_Region = 35995) and lesser respondents belong to Porcupine Health Unit (Health_Region = 35956).

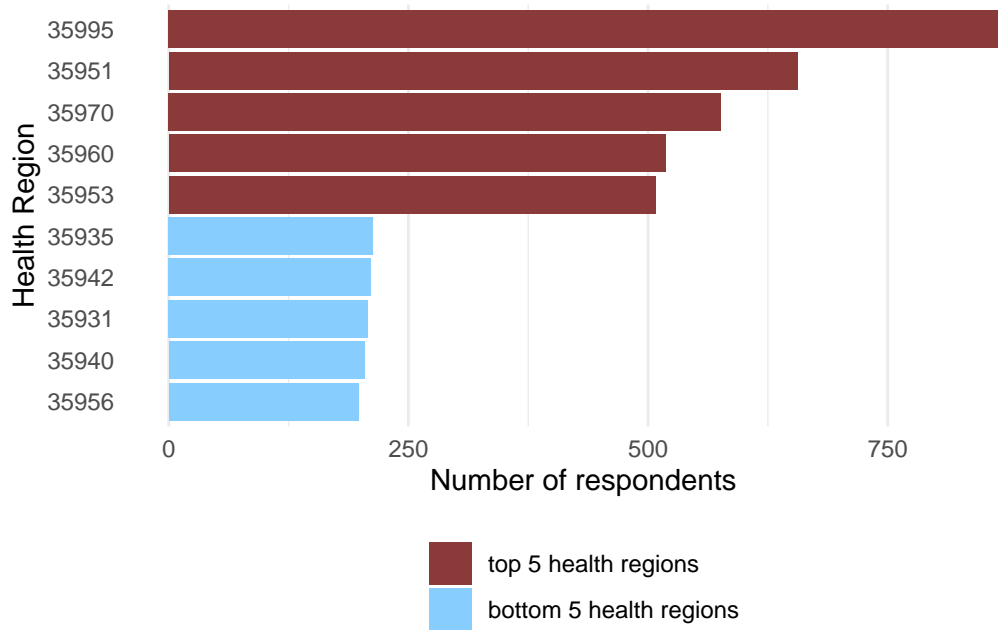


Figure 8: Barplot of top and bottom five occurring health regions

2.5.5 Variable 5: Personal Income

In the dataset, personal income (INCDGPER) is a categorical variable that places an individual's annual income into a range of income values. Below summarises the encoded number and the corresponding meaning of each encoded number. All income values are measured per year.

- 1: No income or income loss
- 2: Less than \$20,000
- 3: \$20,000 to \$39,999
- 4: \$40,000 to \$59,999
- 5: \$60,000 to \$79,999
- 6: \$80,000 or more

According to Figure 9, most of our respondents (25% of the sample) in Ontario are earning between \$20,000 to \$39,000 as part of their annual personal income. It is also worth mentioning that there are very few respondents who are not earning any income in that year. They account for less than 0.5% of the sample. In the 2018 annual wages report provided by Statistics Canada (2020), the annual personal income reported from tax filing in Ontario that year was \$39,510. Therefore, the sampling population's personal income data is representative of Ontario individuals.

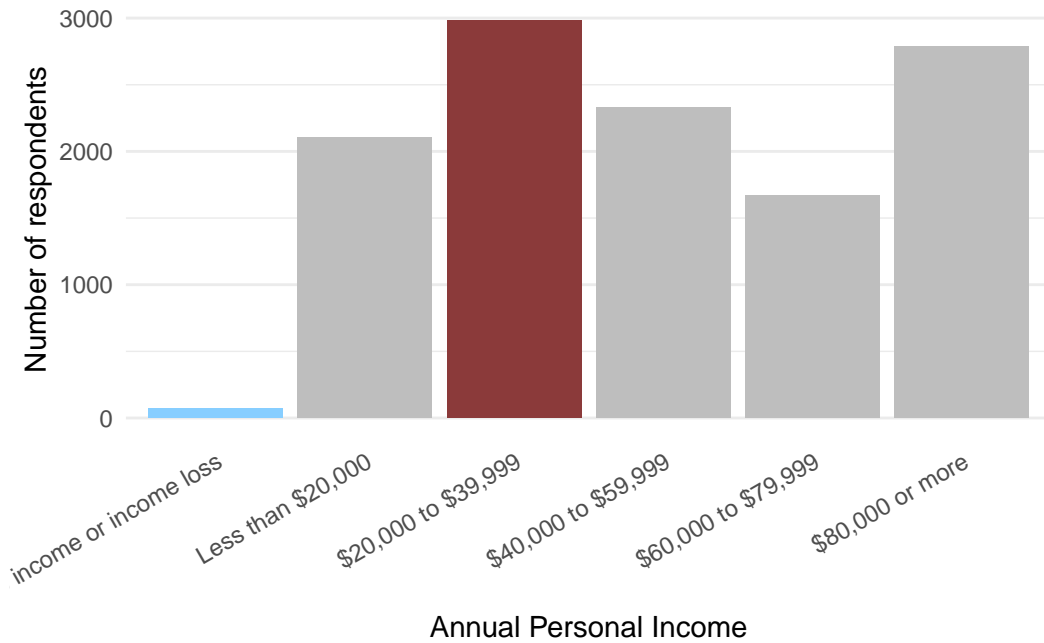


Figure 9: Barplot of personal income

2.5.6 Variable 6: Frequency of alcohol consumption

This variable is a categorical variable with an implied order.

- 1: Less than once a month
- 2: Once a month
- 3: two to three times a month
- 4: once a week
- 5: two to three times a week
- 6: four to six times a week
- 7: every day

From Figure 10, we are able to see that most respondents from this dataset drink about 2-3 times per week (2973 respondents) as seen from the red bar. Also, we are seeing fewer respondents drinking about 4-6 times per week (887 respondents) as seen from the blue bar.

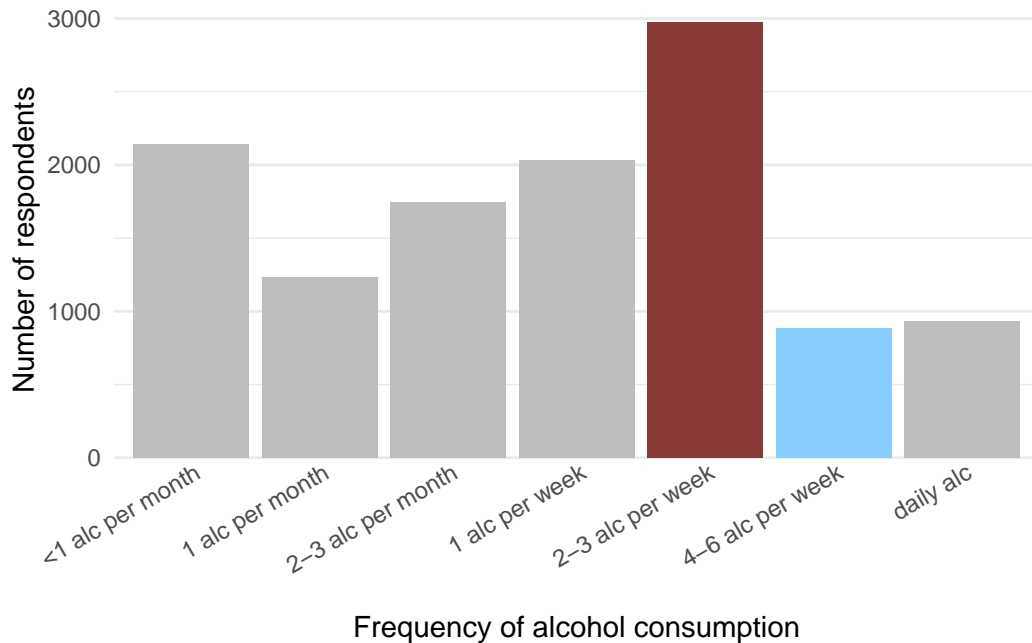


Figure 10: Barplot of frequency of alcohol consumption

2.5.7 Variable 7: Illicit drug use the past 12 months

This is a binary variable which indicates whether an individual has used drugs in the past 12 months prior to the interview. Drugs include marijuana, hashish, cocaine, amphetamines, speed, methamphetamines, crystal meth, ecstasy, hallucinogens, and sniffing glue, gasoline or other solvents. If they have attempted any of these drugs, this variable would be flagged as '1' and '0' otherwise.

It would be interesting to see if drug consumption has an effect on exercising. Brellenthin and Lee (2018) exerts that individuals who are more active tend to have lower substance abuse, suggesting a correlation between these two. Having this variable would be interesting to explore the relationship of an individuals drug usage and their time spent on exercise.

Figure 11 shows us the number of respondents who have (or haven't) consumed drugs 12 months prior to their interviews. We can see that majority of respondents have not consumed drugs in the given timeframe. However, from Table 4, we can see that there are still a number of respondents who had consumed drugs in the past 12 months, particularly a fifth of the sampling population. According to Canadian Mental Health Association (2024), they estimate that about 21% of Canadians, or 6 million people, will meet the criteria for substance use in their lifetime, and this is inclusive of alcohol and that there is a growing opioid crisis in Ontario.

Therefore the obtained sample is somewhat representative of usage of drugs among the Ontario population.

Table 4: Summary of illicit drug use the past 12 months

illicit_drug_use	num_of_respondents	prop_of_population
never consumed drugs	9328	78%
consumed at least once	2613	22%

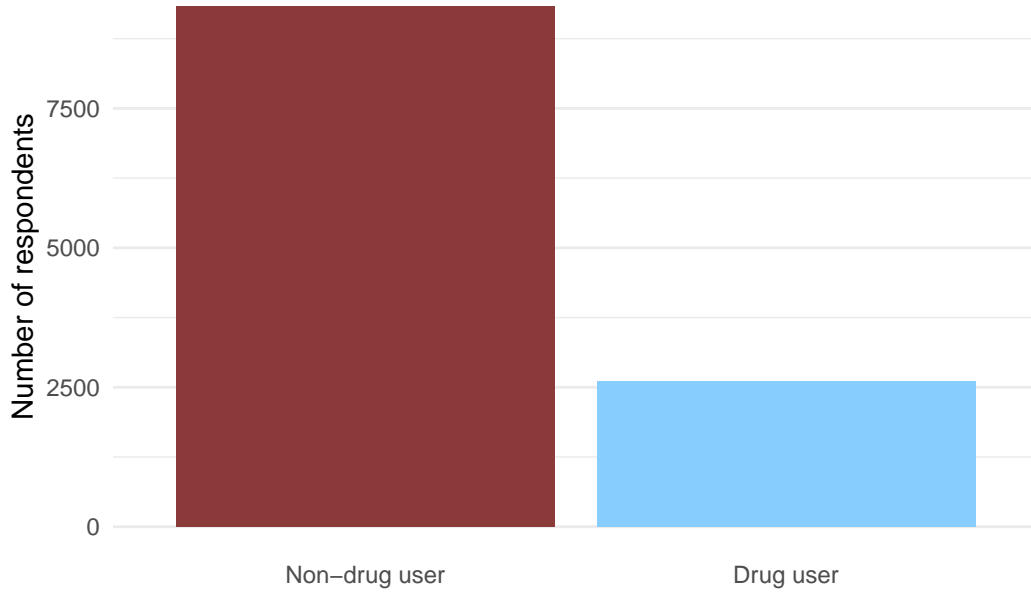


Figure 11: Barplot of illicit drug use the past 12 months

2.5.8 Variable 8: Cigarette consumption

Respondents were asked whether they had smoked more than 100 cigarettes in their lifetimes. This variable is therefore binary, with '1' representing respondents that meet the criteria and '0' otherwise.

It is well known that excessive smoking has negative health impacts on individuals. A cross sectional study conducted by Heydari et al. (2015) found that smokers are less likely to be physically active compared to non-smokers. Hence, it would be interesting to discover if there is an effect of smoking more than 100 cigarettes on the time spent on exercise.

Figure 12 shows us that within the sample, the visual difference between the number of respondents who smoked more than 100 cigarettes (blue column) and respondents who smoked lesser than 100 cigarettes are not that big. From `?tbl-barplot-cigarettes`, 57% of respondents have smoked less than 100 cigarettes while the remaining 43% of respondents smoked more than 100 cigarettes. As the classes in this variable is almost split half,

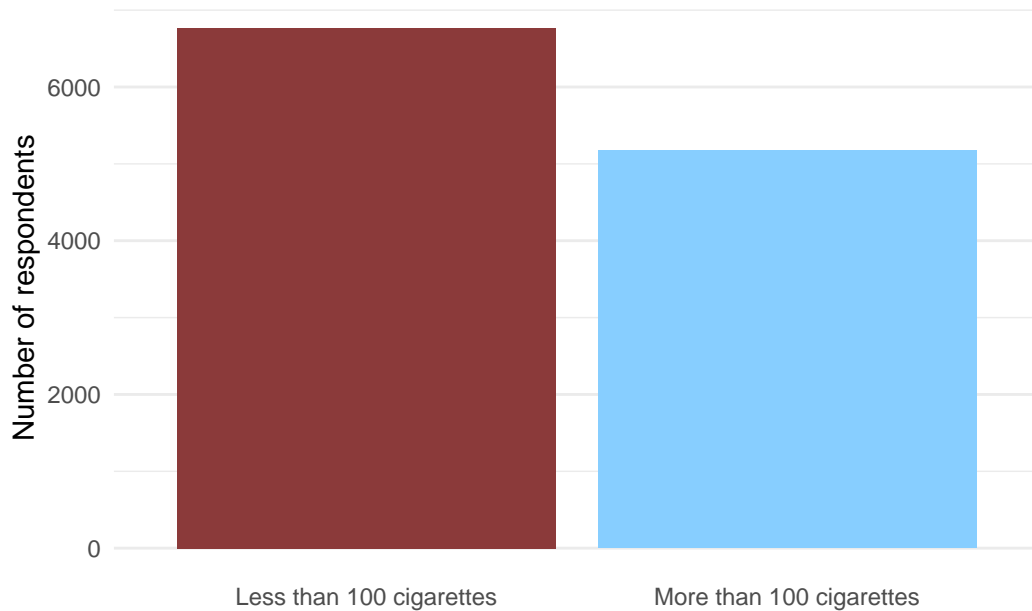


Figure 12: Barplot of total cigarettes smoked in a respondent's lifetime

Figure 13 shows us cigarette consumption by age bins among our sample population. As respondent age increases from the first bin (18 to 19 year old) to the fourth bin (30 to 34 year old), we can see the number of respondents who have smoked more than 100 cigarettes in their lifetime increased from less than 125 respondents to more than 500 respondents (about 3 times more respondents between the two age groups).

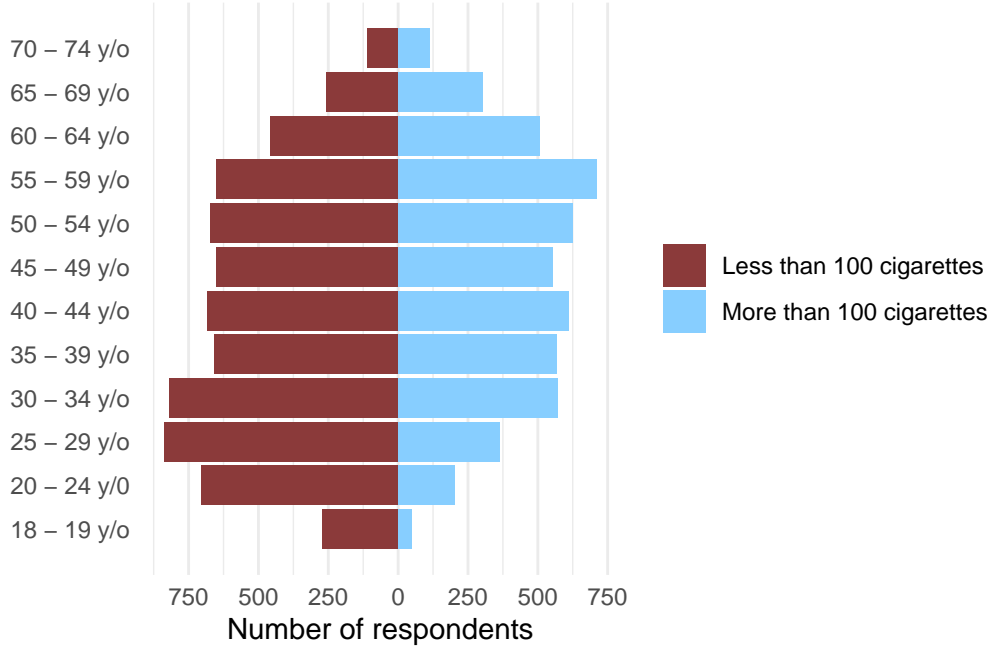


Figure 13

3 Model

The goal of our model is to explore which factors affect time spent on physical activities. Therefore a simple exploratory linear regression would be the chosen modelling algorithm to determine it. A full model where our response (`time_spent_on_physical_activities`) would be fitted on all our interested variables outlined in `?@sec-data-interestedvariables` (INSERT THE SECTION HERE LATER PLEASE DONT FORGET). Background details and diagnostics are included in [Appendix B](#).

3.1 Model set-up

Let k be the individual from CCHS dataset. Define y_k as the time spent on physical activities for individual k . Then Age_k , Sex_k , $Educ_k$ and Alc_k represents individual k 's age, sex, highest educational attainment, and frequency of alcohol consumption the past 12 months respectively. Also, $Drug_k$ and $Cigar_k$ would be indicator explanatory variables derived from the binary variables in our dataset representing if individual k had consumed drugs in the past 12 months and if he/she has smoked 100 cigars in their lifetime respectively. Finally, $HealthRegion_k = 35953$ and $HealthRegion_k = 35970$ are also indicator explanatory variables derived from the categorical variable `health_region` in our dataset, which represents if individual k belongs to

Health Region 35953 (Peel Regional Health Unit) and if he/she belongs to Health Region 35970 (York Regional Health Unit). Note that the reference with respect to Health Region would be individuals that are not in both health region 35953 and 35970.

Therefore the final model would be...

$$y_k = \beta_0 + \beta_1 Age_k + \beta_2 Sex_k + \beta_3 Educ_k + \beta_4 Alc_k + \beta_5 I * (Drug_k) + \beta_6 I * (Cigar_k) + \beta_7 I * (HealthRegion_k = 35953) + \beta_8 * I(HealthRegion_k = 35970) + \epsilon_k$$

$$y_k | \mu_k, \sigma^2 \sim \text{Normal}(\mu_k, \sigma^2) \quad (1)$$

$$\begin{aligned} \mu_k = & \beta_0 + \beta_1 Age_k + \beta_2 Sex_k + \beta_3 Educ_k + \beta_4 Alc_k + \\ & \beta_5 I * (Drug_k) + \beta_6 I * (Cigar_k) + \\ & \beta_7 I * (HealthRegion_k = 35953) + \\ & \beta_8 * I(HealthRegion_k = 35970) \end{aligned} \quad (2)$$

$$\epsilon_k \sim \text{Normal}(0, \sigma^2) \quad (3)$$

We run the model in R (R Core Team 2023) using the **stats** package which is a base library offered by the R Core Team, specifically the **lm()** function to create our simple linear regression. The default parameters of the **lm()** function would be used.

3.1.1 Model justification

Initially, we fitted a full model containing every variable of interest. Since health region is a categorical variable with no implied order, To handle health region, we will be creating dummy variables for each health region, specifically $I * (HealthRegion_k = i)$ for $i \in health_region$, and the reference chosen for health region would be the $health_region = 35926$ (The District of Algoma HU). We define $Income_k$ to be the the personal income of individual k in our dataset.

Then, for $8 \leq c \leq 40$ and $i \in health_region$, our original full model was defined as...

$$y_k = \beta_0 + \beta_1 Age_k + \beta_2 Sex_k + \beta_3 Educ_k + \beta_4 Income_k + \beta_5 Alc_k + \beta_6 I * (Drug_k) + \beta_7 I * (Cigar_k) + \beta_c I * (HealthRegion = i) + \epsilon_k$$

We then conducted model reduction using backward feature elimination using Bayesian Information Criterion (BIC) as our scoring metric to obtain the least number of predictors that minimises the residual errors from the reduced model. This model reduction was performed using **stats** package from the R Core Team (2023) as well. The resulting BIC value was obtained from the **lme4** package (Bates et al. 2015).

4 Results

The final model with our estimated coefficients are as follows

$$y_k = 5.84 - 0.03Age_k + 0.23Sex_k - 0.08Educ_k + 0.04Alc_k + 0.15I * (Drug_k) + \\ 0.12I * (Cigar_k) + \beta_c I * (HealthRegion = i) + \epsilon_k$$

Our results are summarized in Table [5](#).

Table 5: Linear model of time spent on physically taxing activities based on explanatory variables

	Time spent on physical activity
(Intercept)	5.842
	(0.063)
num_alc_drunk_12m	0.038
	(0.006)
age	−0.033
	(0.004)
sex	0.232
	(0.021)
illicit_drug_use	0.155
	(0.027)
highest_educational_attainment	−0.087
	(0.019)
smoked_hundred_cigarettes	0.127
	(0.022)
health_region_35953	−0.233
	(0.051)
health_region_35970	−0.174
	(0.048)
Num.Obs.	11 941
R2	0.039
R2 Adj.	0.038
AIC	36 767.1
BIC	36 841.0
Log.Lik.	−18 373.547
RMSE	1.13

5 Discussion

5.1 Males are more likely to engage in physical activities

For the explanatory variable **sex**, the corresponding coefficient from the fitted linear model is 0.232. This means that if an individual is a male, this increases the time spent on physical activities by 26%. While we were expecting increase in the overall engagement in physical activity among women as explained in Section 2.5.2, men are still more likely to engage in physical activities than women. In 2018, Global News Canada reported that 31% of women and 25% of men have insufficient physical activity (Global News, a division of Corus Entertainment Inc. Corus News. 2018). Therefore, our linear regression model discovery was not surprising.

In fact, in a study published in the Journal of American College of Cardiology, they discovered that women who were physically active would experience greater health benefits than men who were also physically active (Ji et al. 2024). Therefore, this means that more can be done to address the gender gap in the engagement of physical activities.

5.2 Ontario residents who have smoked before might be more likely to engage in physical activities.

From elementary school, Ontario students are exposed to health education where schools would educate about the harms of smoking. It is well known that smoking has negative effects on an individual's health. Therefore it is unexpected that the linear regression model showed a positive coefficient for the indicator variable for 'having smoked 100 cigarettes' in an individual's lifetime.

The linear model is telling us that if an individual has smoked 100 cigarettes in their lifetime, they are 13% more likely to spend more time being physically active than an individual who has not smoked more than 100 cigarettes in the same timeframe. While it may seem contradicting to what we believe, there may be merit to this value.

A team from the school of Public Health Sciences at University of Waterloo had explored patterns and trends for tobacco use in Canada. What they found was that 70% of Canadians who had smoked before have quit as of 2020 Reid JL and Hammond D and Burkhalter R and Rynard VL (2022). While smoking may have been affected an individual's physical capabilities, it could have also increased awareness regarding the importance of their health. Since most Canadians who had a history of smoking had quit smoking, it is very possible that they have greater health conscious. Coupled with increased government support for programs with the sole purpose of increasing physical activeness in Ontario (Public Health Agency of Canada 2024), it is no longer surprising why individuals who smoked 100 cigarettes in their lifetime are on average 13% more likely to engage in physical activities.

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Third discussion point

5.4 Possible increase in stimulant use (about drugs)

5.5 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

In `?@fig-ppcheckandposteriorvsprior-1` we implement a posterior predictive check. This shows...

In `?@fig-ppcheckandposteriorvsprior-2` we compare the posterior with the prior. This shows...

Examining how the model fits, and is affected by, the data

B.2 Diagnostics

`?@fig-stanareyouokay-1` is a trace plot. It shows... This suggests...

`?@fig-stanareyouokay-2` is a Rhat plot. It shows... This suggests...

Checking the convergence of the MCMC algorithm

References

- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software* 67 (1): 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Brellenthin, Angelique G, and Duck-chul Lee. 2018. “Physical Activity and the Development of Substance Use Disorders: Current Knowledge and Future Directions.” *Prog Prev Med (N Y)* 3 (3): e0018. <https://doi.org/10.1097/pp9.0000000000000018>.
- Canadian Mental Health Association. 2024. *Substance Use and Addiction*. https://ontario.cmha.ca/addiction-and-substance-use-and-addiction/#_edn5.
- Data Centre, Faculty of Arts & Science, University of Toronto. 2018. *Canadian Community Health Survey (CCHS)*. <https://sda-artsci-utoronto-ca.myaccess.library.utoronto.ca/index.html/sda.htm>.
- Global News, a division of Corus Entertainment Inc. Corus News. 2018. *Nearly One Third of Canadian Women Don’t Get Enough Physical Activity — Worse Than Men*. <https://globalnews.ca/news/4428978/women-physical-activity-gender-gap/>.
- Heydari, Gholamreza, Mostafa Hosseini, Mahmoud Yousefifard, Hadi Asady, Masoud Baikpour, and Atena Barat. 2015. “Smoking and Physical Activity in Healthy Adults: A Cross-Sectional Study in Tehran.” *Prog Prev Med (N Y)* 14 (4): 238–45.
- Ji, Hongwei, Martha Gulati, Tzu Yu Huang, Alan C. Kwan, David Ouyang, Joseph E. Ebinger, Kaitlin Casaletto, Kerrie L. Moreau, Hicham Skali, and Susan Cheng. 2024. “Sex Differences in Association of Physical Activity with All-Cause and Cardiovascular Mortality.” *Journal of the American College of Cardiology* 83 (8): 783–93. <https://doi.org/https://doi.org/10.1016/j.jacc.2023.12.019>.
- Kari, Jaana T, Jutta Viinikainen, Petri Böckerman, Tuija H Tammelin, Niina Pitkänen, Terho Lehtimäki, Katja Pakkala, Mirja Hirvensalo, Olli T Raitakari, and Jaakko Pehkonen. 2020. “Education Leads to a More Physically Active Lifestyle: Evidence Based on Mendelian Randomization.” *Scand J Med Sci Sports* 30 (7): 1194–204. <https://doi.org/10.1111/sms.13653>.
- ParticipACTION. 2024. *Key Statistics*. <https://www.participaction.com/the-science/key-facts-and-stats/>.
- Public Health Agency of Canada. 2024. *Government of Canada Invests in Programming to Promote Physical Activity and Healthy Living*. <https://www.canada.ca/en/public-health/news/2024/06/government-of-canada-invests-in-programming-to-promote-physical-activity-and-healthy-living.html>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Reid JL and Hammond D and Burkhalter R and Rynard VL. 2022. *Tobacco Use in Canada: Patterns and Trends*. <https://uwaterloo.ca/tobacco-use-canada/tobacco-use-canada-patterns-and-trends>.
- Statista Research Department. 2024. *Number of Students Enrolled in Postsecondary Institutions in Canada from 2000 to 2022*. <https://www.statista.com/statistics/447739/>

- [enrollment-of-postsecondary-students-in-canada/](#).
- Statistics Canada. 2018. *Canadian Community Health Survey - Annual Component (CCHS)*. <https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=329241>.
- . 2020. *Table 2: Median Wages, Salaries and Commissions, by Province and Territory*. <https://www150.statcan.gc.ca/n1/daily-quotidien/200218/t002c-eng.htm>.
- . 2023. *Canada's Population Estimates: Record-High Population Growth in 2022*. <https://www150.statcan.gc.ca/n1/daily-quotidien/230322/dq230322f-eng.htm>.
- The Daily, Statistics Canada. 2000. *Population Estimates*. <https://www150.statcan.gc.ca/n1/daily-quotidien/000926/dq000926a-eng.htm>.
- World Health Organisation. 2024. *Physical Activity*. <https://www.who.int/news-room/fact-sheets/detail/physical-activity>.
- World Health Organization. 2024. *Canada, Health Data Overview for Canada*. <https://data.who.int/countries/124>.