Investigating Canadian sleeping trends*

Pascal Lee Slew

26 April 2022

Abstract

Sleep has become a luxury in modern times. There is only a limited time for so many things to do. In this paper, we used the 2017-2018 CCHS to focus on the sleeping trends and the factors affecting them. We found that a significant number of canadians have difficulty sleeping. There are also noticeable differences in sleeping hours among age groups which denote potentially some sort of social trend. The analysis of sleeping data is important to get an insight of the current situation and identify ways to help Canadians live a better life.

keywords: sleep, canadian community health survey, sleep deficiency, age

1 Introduction

Sleep is important for our health. In today's world, there are so much things we have or want to do but we only have 24 hours a day. Often, people reduce their sleeping hours for other priorities with the belief that the benefits outweigh the costs. This is particularly the case for many university students who are willing to sacrifice their sleep to study or complete assignments. Failing to get enough sleep over time can lead to other health complications. Therefore, having an idea of the current sleep habits of Canadians is important to formulate policies and implement programs to improve the situation of Canadians.

The Canadian Community Health Survey (CCHS) has been set up with the objective of collecting health-related data on Canadians. With the collected data, analysis can be conducted to identify any worrying trends and relay crucial information to the stakeholders.

The Canadian Community Health Survey (CCHS), a joint initiative of Health Canada, the Public Health Agency of Canada, Statistics Canada and Canadian Institute for Health Information (CIHI), provides information on a broad range of topics around Canadians health lifestyles.

In this paper, we seek to investigate the sleeping trends of Canadians. We will be looking at the demographics and other factors such as having a baby or drinking first. Then, we will proceed by building a linear mixed model with sleeping hours as the response and age group, baby, diabetes, cholesterol among other predictors. We found that ...

The remaining part of the paper is divided as follows: Section 2 explains where our data comes from and gives a general idea of the variables present in the dataset. Section 3 shows the model that explains how Canadian sleeping is influenced. Section 4 explains the key findings of the data analysis and model. Section 5 expands on what is found and why it is important. Section 5.4 contains supplementary graphs that support the arguments in the discussion section.

*Code and data are available at: https://github.com/Pascal-304/canadian-sleep-2018.

Table 1: An overview of key variables of the dataset

Age	Marital Status	Sex	Sleep hours	Perceived Health	Perceived Mental Health
Age between 70 and 74	Widowed/Divorced/Separated	Male	6 hours to less than 7 hours	Poor	Good
Age between 50 and 54	Widowed/Divorced/Separated	Female	5 hours to less than 6 hours	Very good	Very good
Age between 30 and 34	Common-law	Female	7 hours to less than 8 hours	Excellent	Excellent
Age between 45 and 49	Common-law	Female	9 hours to less than 10 hours	Very good	Very good
Age between 75 and 79	Widowed/Divorced/Separated	Female	7 hours to less than 8 hours	Good	Good
Age between 40 and 44	Married	Female	7 hours to less than 8 hours	Very good	Very good
Age between 45 and 49	Married	Male	6 hours to less than 7 hours	Very good	Good
Age between 12 and 14	Single	Female	8 hours to less than 9 hours	Very good	Good
Age between 60 and 64	Widowed/Divorced/Separated	Female	7 hours to less than 8 hours	Fair	Excellent
Age between 35 and 39	Common-law	Male	7 hours to less than 8 hours	Very good	Very good

2 Data

2.1 Data Source

This paper uses a subset of the Canadian Community Health Survey 2017-2018 data [citation of source].

In paper we run our analysis in `R` (R Core Team 2020).

2.2 Methodology

The Canadian Community Health Survey is a national-level survey carried out in Canada. It was conducted by

The survey frame respondents eligible for the survey were of age 12 and over in the ten provinces.

The target sample size was about 130,000 respondents to obtain reliable estimates for each Health regions. A minimum of 500 respondents is needed for each HR for attaining a reasonable level of data quality. A restriction of maximum sampling fraction of 1 out of 20 dwellings was imposed to prevent sampling too many dwellings in smaller regions.

The CCHS sample is selected using different frames according to the age group. For the adult population, the sample of households is selected from an area frame while for the young population, a list frame is used to select persons.

The area frame used by the Canadian Labour Force Survey(LFS) is used as a sampling frame for the adult population. It uses a two-stage sample design. In the first stage, a sample of primary sampling units(PSUs), corresponding to geographical regions called clusters, is selected. In each selected PSU, a sample of dwelling is drawn at the second-stage. The clusters are formed of 100 to 600 dwellings generally.

For the youth population, a list frame created from the Canadian Child Benefit (CCB) files is used. The files contain address and phone numbers which are used to conduct phone interviews.

The initial sample size for each frame is collected over a period of 3 months 4 times.

2.3 Key features

The dataset used has been

This survey contains a total of 1,051 variables and 113,290 observations.Since the survey has

This report will focus on a subset of variables that will be used to analyze the relationships of the number of hours usually spent sleeping with different variables such as perceived mental health, life satisfaction, and work stress. Table 1 shows a subset of key variables that will be discussed in this paper.

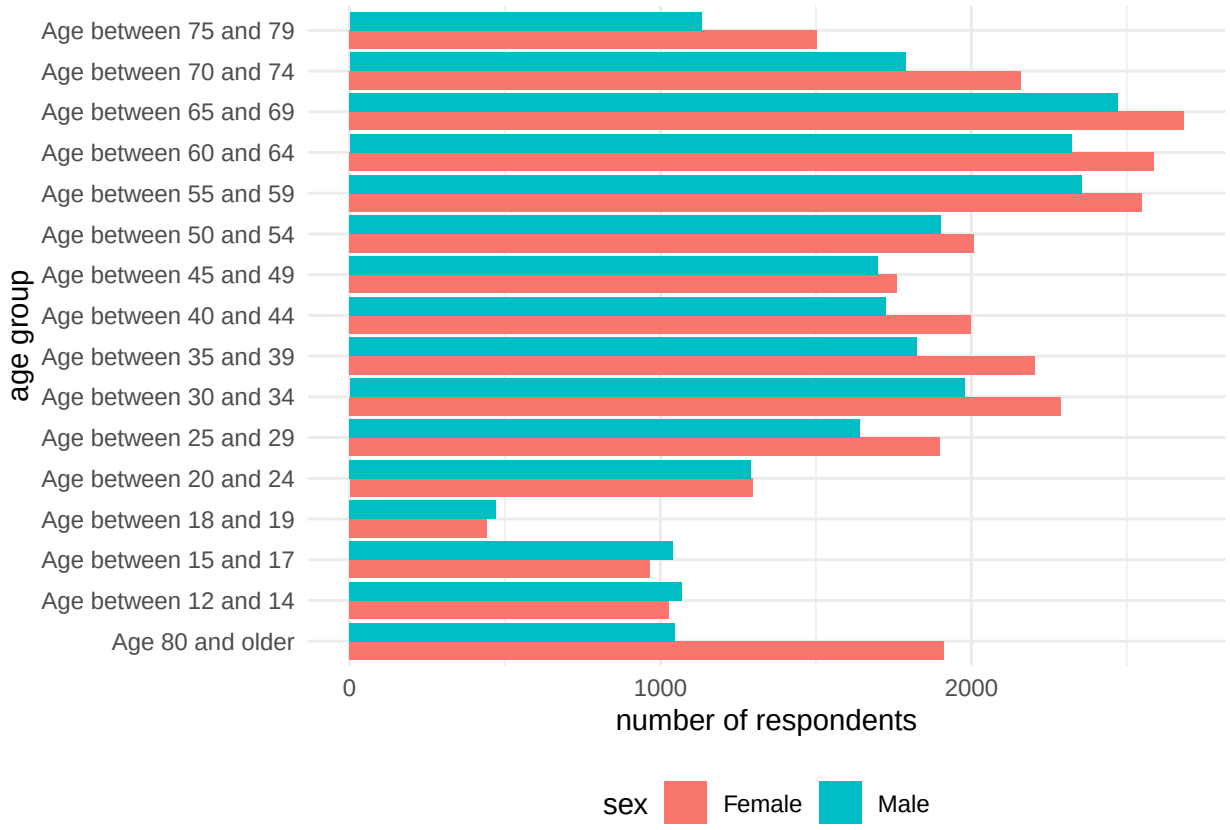

Figure 1: Distribution of survey respondents by age group and sex

Figure 1 shows the age distribution of the respondents. For all age groups, there seems to be a higher number of female adult respondents than male respondents. The converse is true for the youth population; there is higher male respondents than female. About 53.7% of respondents are female and 46.3% are male.

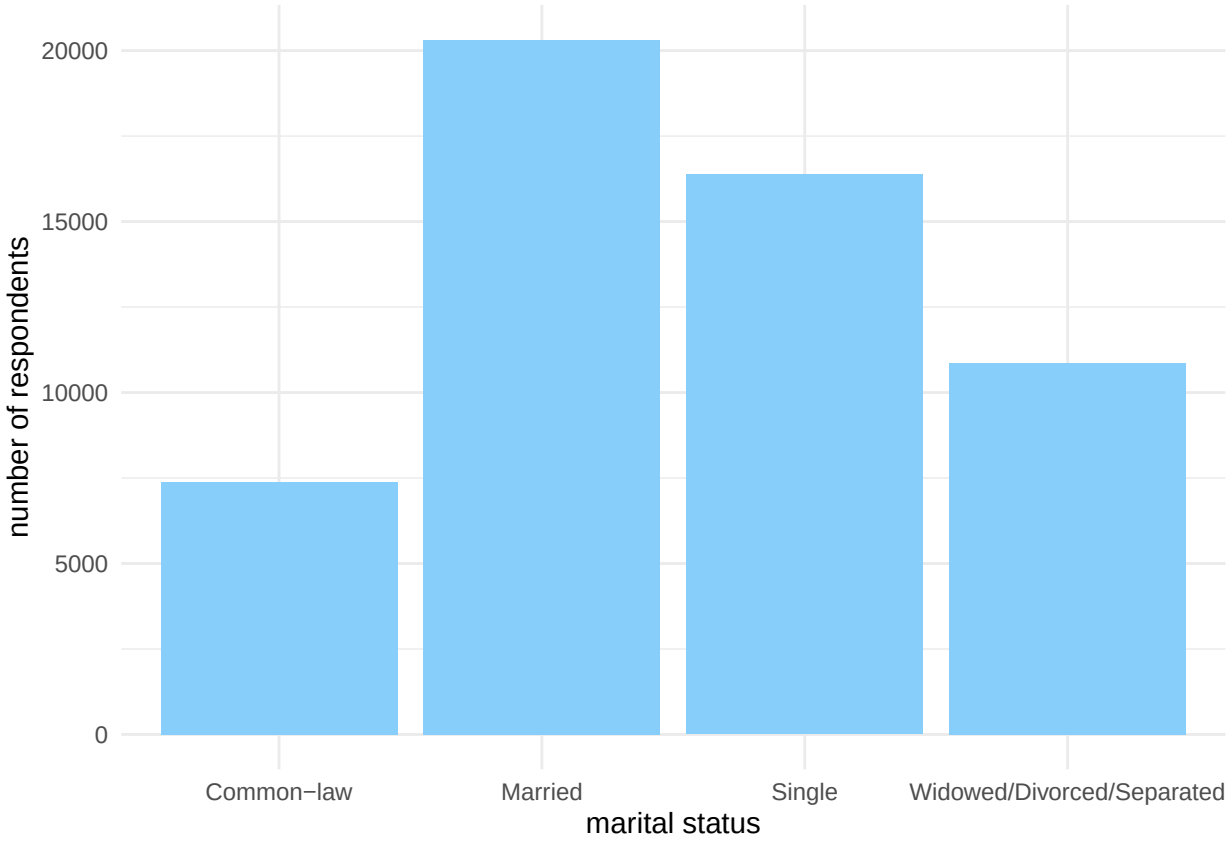

Figure 2: Marital status of respondents

Figure 2 shows the marital status distribution of the respondents. Most respondents are married (), followed by single. A fairly high number of respondents are widowed, divorced or separated while being in a common-law relationship is the least preferred.

Figure 3 shows the number of hours usually spent sleeping. Most respondents tend to sleep between 7 to 8 hours, followed by 6 to 7 hours and 8 to 9 hours. A fairly high number of people sleep more than 9 hours on a regular basis. It is also worthwhile to note that X% usually sleep for less than 5 hours. This suggests that we should take a closer look at the reasons why it is the case.

Figure 4 illustrates the distribution of the number of hours spent sleeping for different perceived health. For those who perceived their health as "Excellent", "Very Good" and "Good", we observe that the distribution of the number of sleeping hours to be fairly similar with 7 hours to less than 8 hours as the mode. For those who perceived their health as "Fair", the distribution of their sleeping hours seems to be close to uniform. We can see that a higher proportion of respondents answered to have lower sleeping hours. Similarly for those who perceived their health as "Poor", the distribution is almost uniform with most people sleeping less than 5 hours on a regular basis. This suggests that how people perceive their health may have an influence over the number of hours they usually sleep among other factors.

Figure 5 shows the distribution of the number of hours usually spent sleeping for different perceived mental health. For those who perceived their mental health as "Excellent", "Very Good" and "Good", the distribution of sleeping hours are similar. The higher people perceived their mental health, the lower the proportion

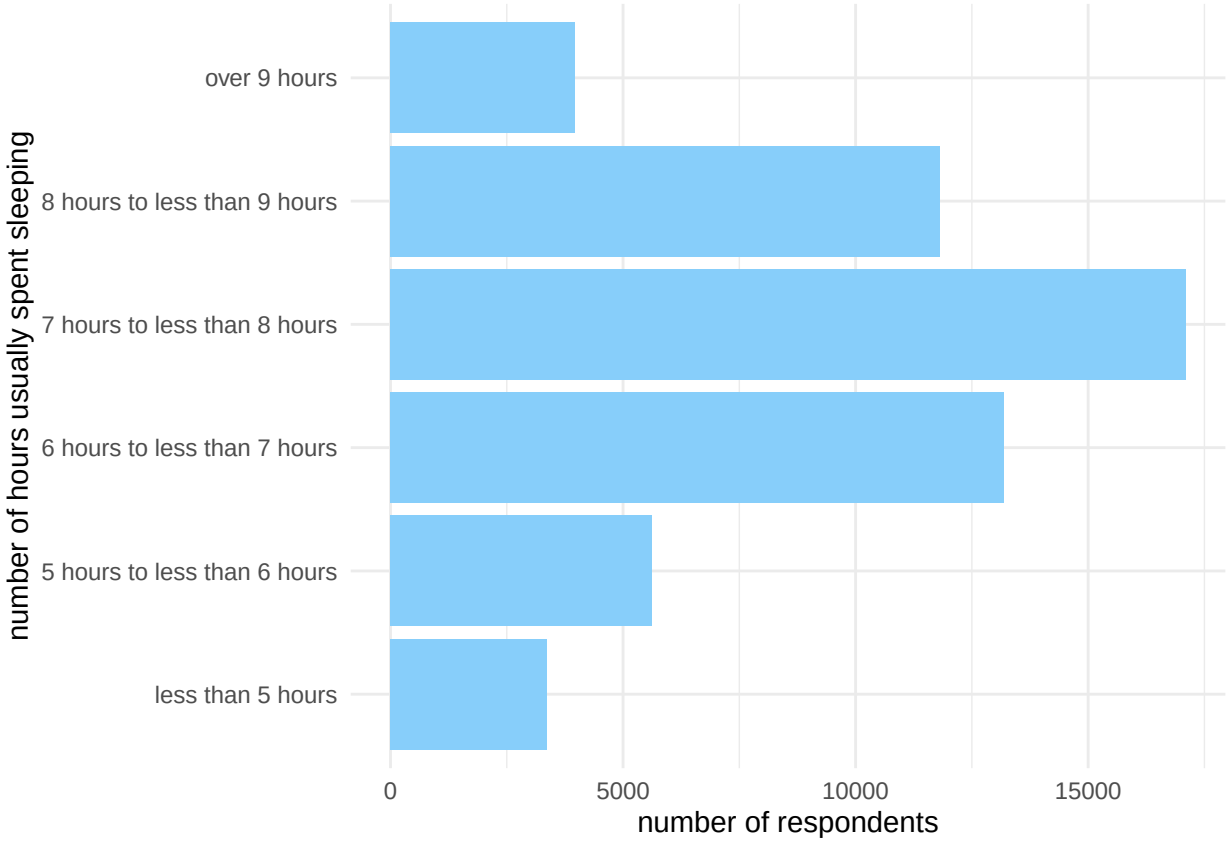

Figure 3: Distribution of the hours of sleep of respondents

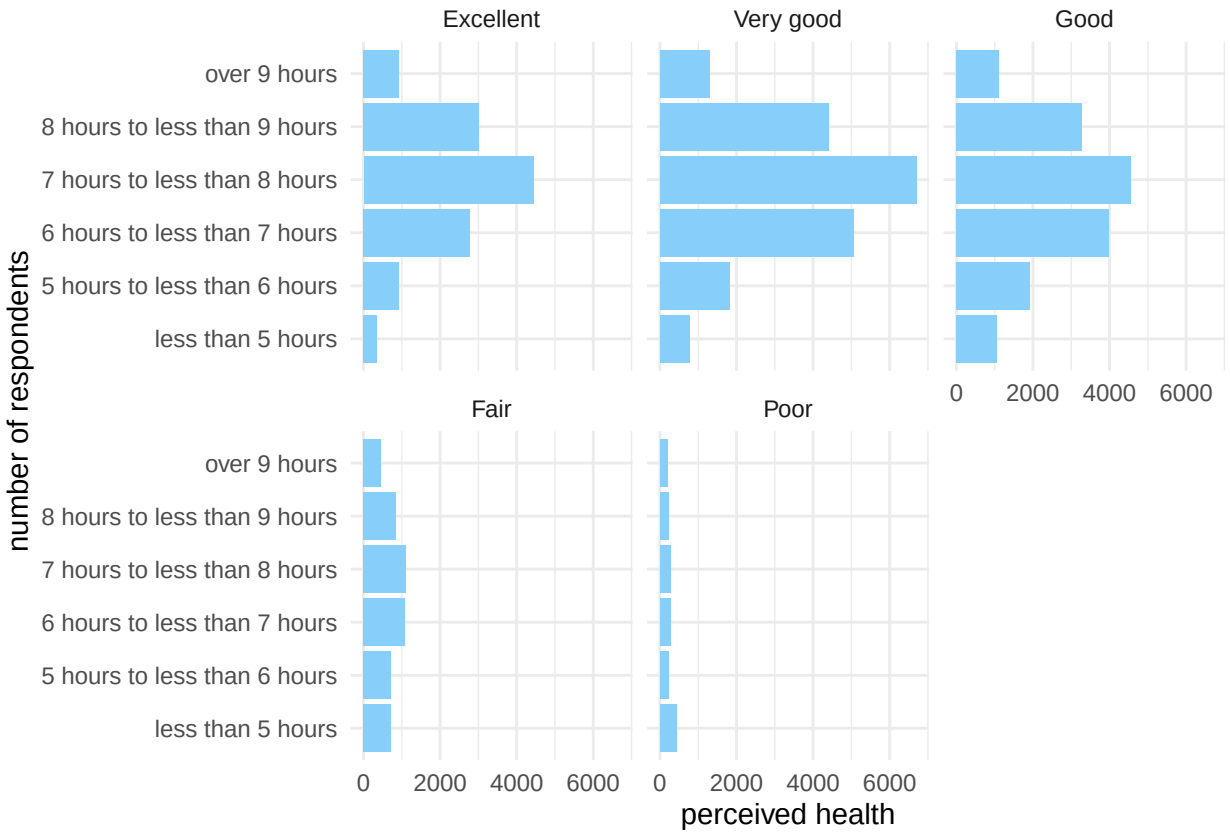

Figure 4: Comparison of the distribution of respondent based on their perceived health for sleeping hours

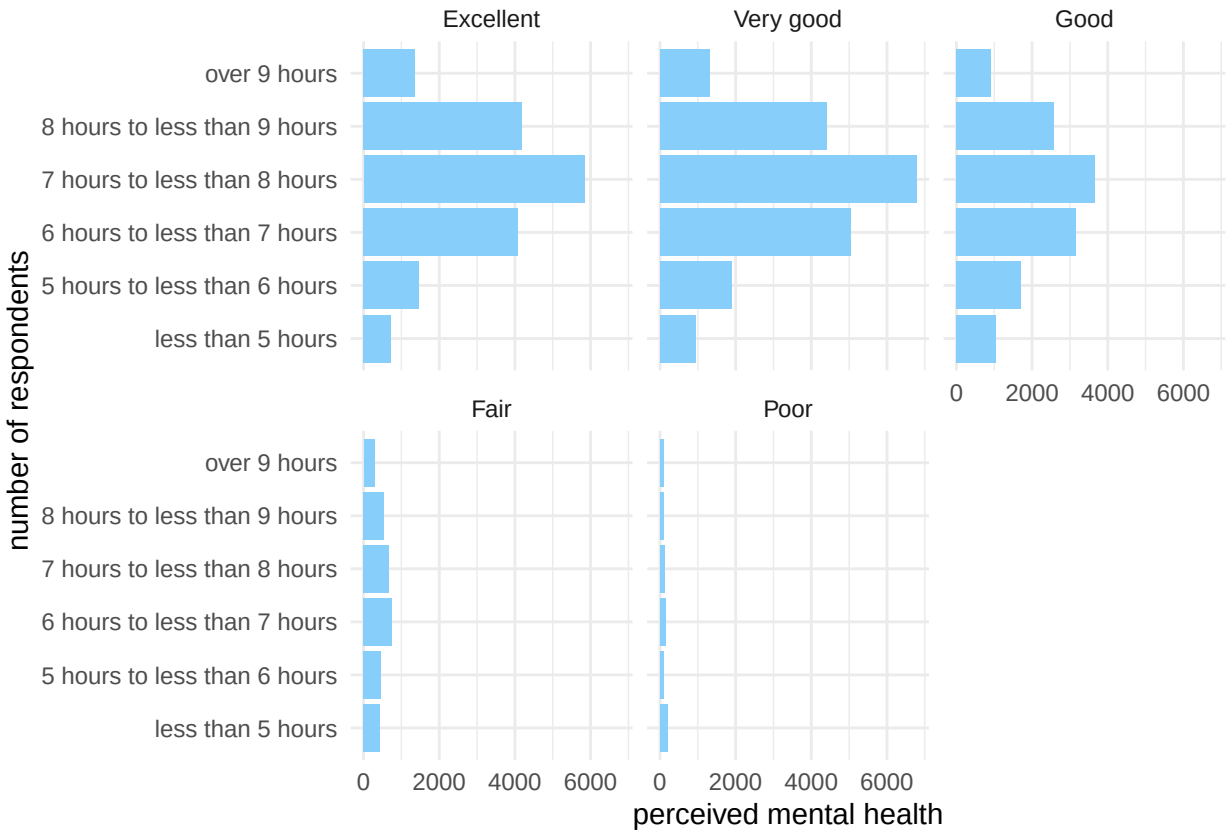

Figure 5: Comparison of the distribution of respondents' sleeping hours based on their perceived mental health

of people sleeping too less. For "Fair" and "Poor", the proportion of respondents who reported lower sleeping hours is the highest. Again, the perception of mental health seems to have some effect on the hours spent sleeping.

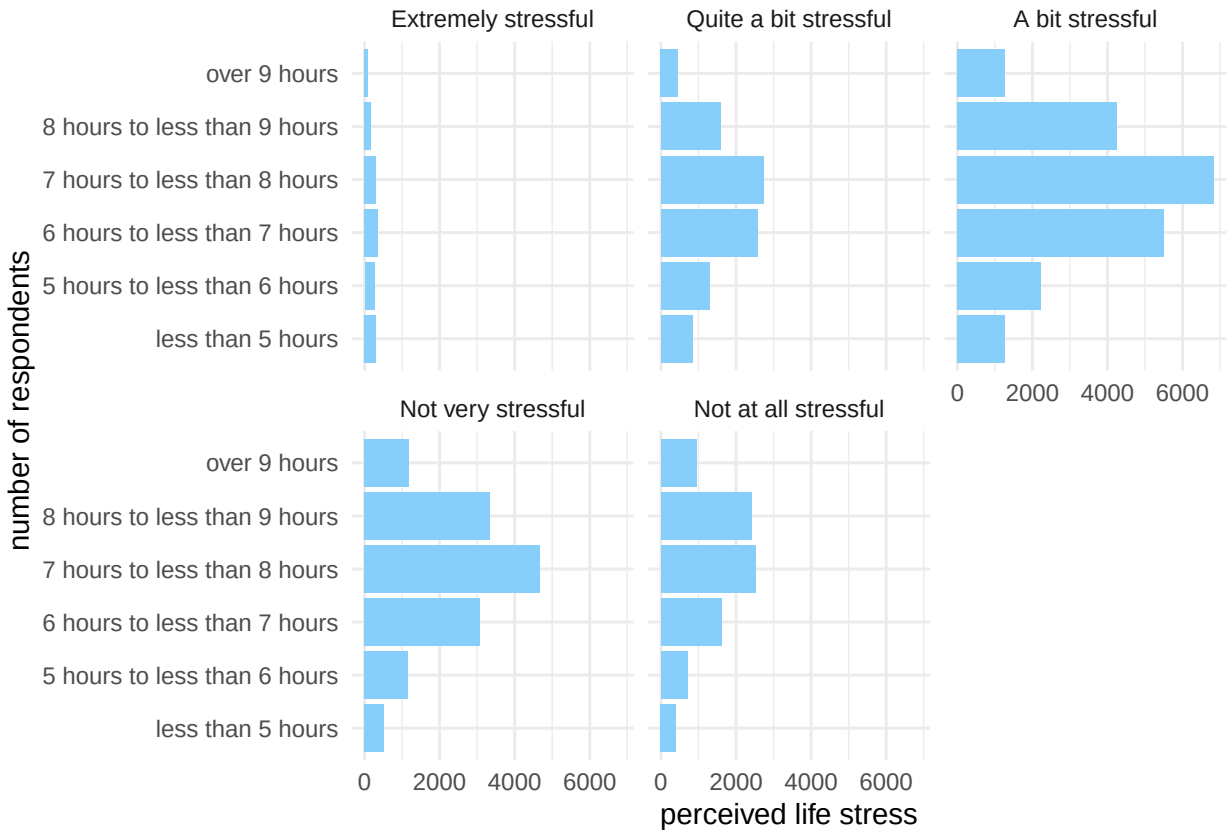

Figure **??** shows how the number of hours usually spent sleeping is distributed for how the respondents perceived their stress in life. Except for those who feel that life is extremely stressful, the distributions of the number of hours spent sleeping are similar. The less stress people feel, the higher the concentration of people sleeping adequately. The proportion of people sleeping 8 to 9 hours is highest for those who feel life is not at all stressful.

Figure 6 shows the comparison of the number of hours of sleep based on perceived work stress. Most people feel that their work is a bit stressful or not very stressful. Regardless of how they feel their work stress, people tend to sleep between 7 to 8 hours of sleep. It seems that work stress does not really have a noticeable effect on the number of hours usually spent sleeping.

```
## Warning: Unknown levels in `f`: 11
```

Figure 7 compares the number of hours slept based on life satisfaction. Life satisfaction is on a scale of 0-10, with 0 as being not satisfied at all and 10 as being fully satisfied with their life. Most people reported a 9 or 10 which is great.

Figure 8 shows the income distribution of the CCHS respondents. Most respondents earned between 20,000 dollars to 39,999 dollars, followed by those in the less than $20,000 income group. It is worth to note that most selected respondents in the Health regions are not financially stable. A small number of respondents (700) had no income or income loss; they should consists of mostly of youth.

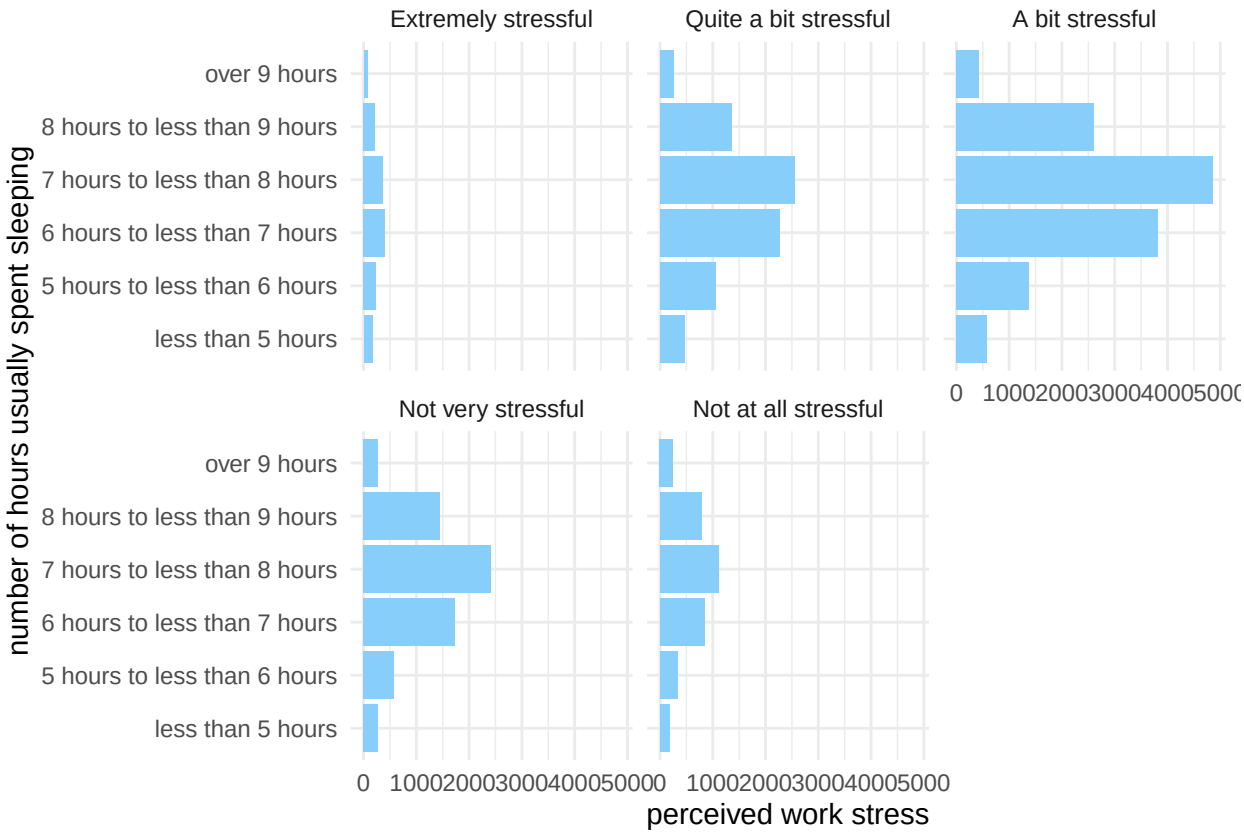

Figure 6: Comparison of the distribution of respondents based on their perceived stress at work

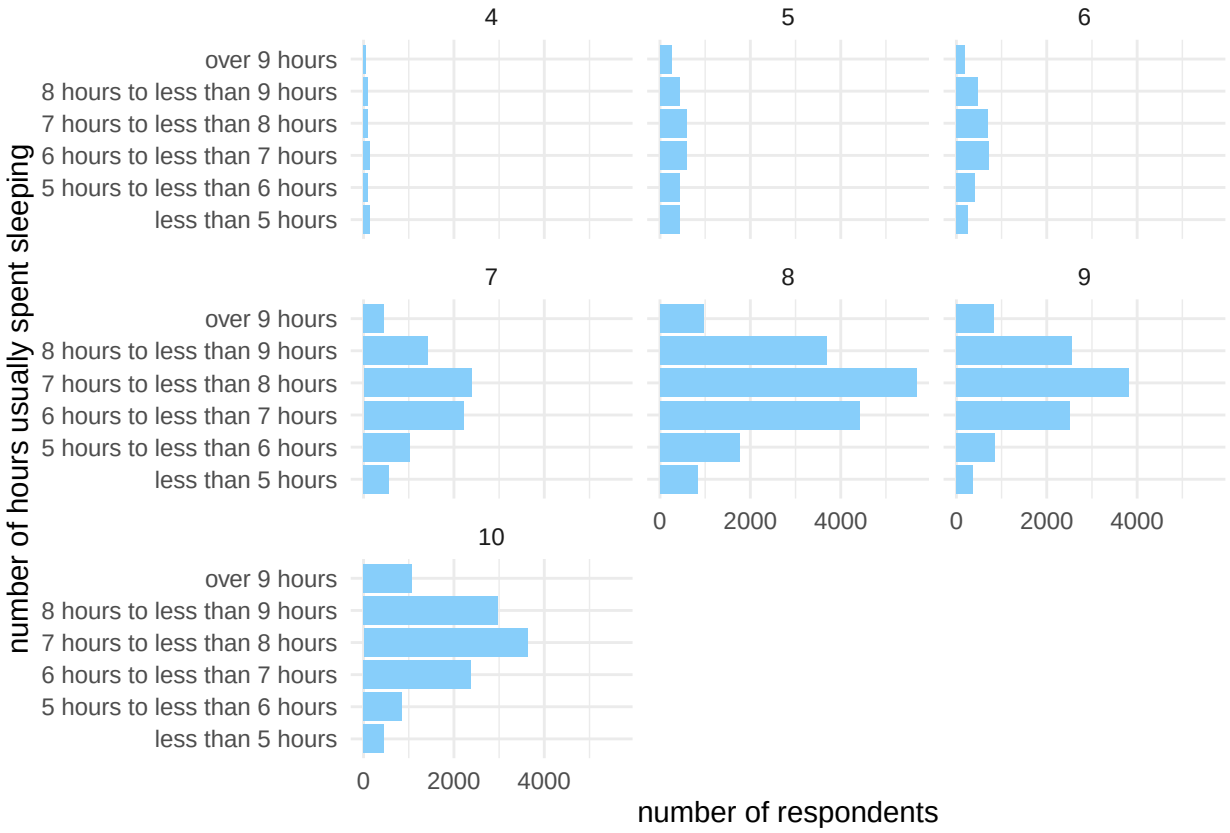

Figure 7: Comparing the number of hours usually spent sleeping against perceived life satisfaction

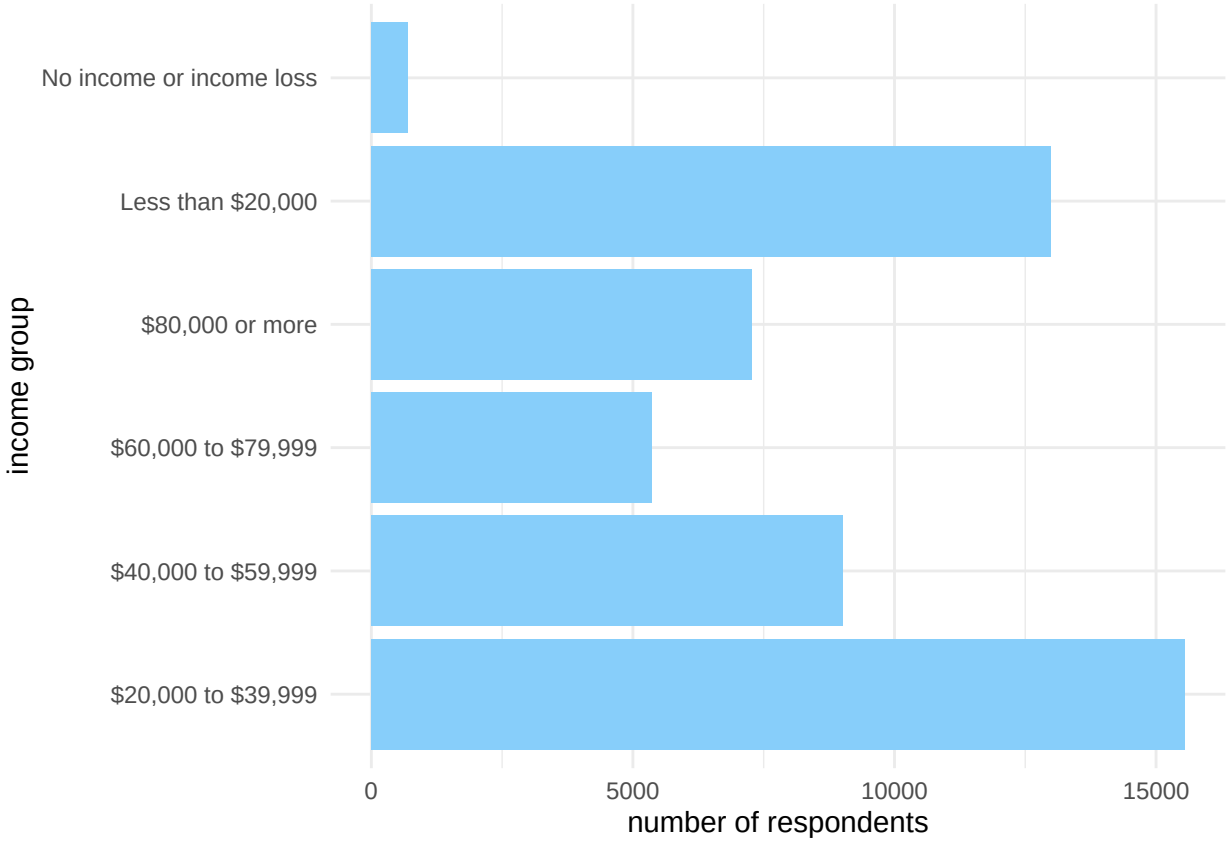

Figure 8: Income groups of respondents

3 Model

3.1 Logistic Regression

This study aims to find whether our variables of interest have an effect on sleeping well or not. We are especially interested in the odds of people sleeping well.

We created a binary variable which takes 1 if the respondent reported sleeping between 7 hours to 9 hours and 0 otherwise. We assumed that sleeping well means 7-9 hours of sleep. It will be the dependent variable.

Binary logistic regression was used due to its efficiency and popularity in assessing dichotomous outcomes, providing measures of appropriateness for each predictor as coefficients and indicating the direction of each association, either positive or negative.

The logistic regression model was fitted in R (R Core Team 2020) using the lme4 r package (Bates et al. 2015). I fitted different possible models based on the variables of interest and compared the models using the drop-in-deviance test. The drop-in-deviance test is used to test the significance of model coefficients. The null hypothesis for the drop-in-deviance test is that the simple model explains the data as well as the complex model while the alternative hypothesis would be that the complex model explains the data better than the simple model. I used a 5% level of significance which means that if a p-value less than 0.05 is obtained, we have strong evidence supporting the complex model being better.

The final model obtained is as follows:

$$log\ p/(1-p) = \beta_0 + \beta_1 age + \beta_2 income + \beta_3 sex + \beta_4 mental\ health + \beta_5 life\ stress + \beta_6 marital\ status$$

where $log\ p/(1-p)$ represents the log-odds of sleeping well and the β_i for i = 1 to 6 represents the estimated coefficients of the predictors.

Finally, we performed a goodness of fit test by comparing the residual deviance (68560) to a χ^2 distribution with 50492 degrees f freedom. We found that our model with age, sex, income, mental health, life stress and marital status has statistically significant evidence of lack-of-fit (p < .001). The lack-of-fit could be the result of many reasons, namely: a) missing covariates, b) outliers, or c) over dispersion.

We still choose to go with the model despite the seemingly lack-of-fit. As the statistician George Box said, all models are wrong, but some are useful.

4 Results

\begin{table}

\caption{MLEs of baseline odds, with 95% confidence intervals}

Variables	Estimated coefficient	2.5%	97.5%
Reference level	0.661	0.567	0.771
mental health: Very Good	0.962	0.921	1.004
mental health: Good	0.764	0.728	0.803
mental health: Fair	0.590	0.542	0.641
mental health: Poor	0.469	0.394	0.555
Male	0.853	0.822	0.885
Age between 18 and 19	1.707	1.450	2.011
Age between 20 and 24	1.538	1.366	1.733
Age between 25 and 29	1.320	1.185	1.470
Age between 30 and 34	1.166	1.052	1.291
Age between 35 and 39	1.241	1.120	1.375
Age between 40 and 44	1.206	1.087	1.338
Age between 45 and 49	1.117	1.006	1.241
Age between 50 and 54	1.069	0.966	1.184
Age between 55 and 59	1.047	0.951	1.154
Age between 60 and 64	1.167	1.061	1.284
Age between 65 and 69	1.167	1.062	1.282
Age between 70 and 74	1.136	1.029	1.253
Age between 75 and 79	1.066	0.957	1.186
income: $40,000 to $59,999	1.074	1.018	1.132
income: $60,000 to $79,999	1.102	1.034	1.175
income: $80,000 or more	1.124	1.060	1.192
income: less than $20,000	0.899	0.856	0.945
No income or income loss	0.876	0.749	1.023
life stress: Quite a bit stressful	1.537	1.363	1.736
life stress: A bit stressful	1.902	1.692	2.140
life stress: Not very stressful	2.415	2.142	2.726
life stress: Not at all stressful	2.494	2.202	2.828
Married	0.895	0.846	0.946
Single	0.820	0.771	0.872
Widowed/Divorced/Separated	0.746	0.699	0.797

\end{table}

Table 4 shows the model coefficients with their 95% confidence intervals. The reference group for the model is female, aged 80 and older, common-law relationship, feel life is extremely stressful, perceived their mental health as “Excellent” and with an income of $20,000 to 39,999.

4.1 Interpretation

We found that the odds of sleeping well are considerably greater for females compared to males. In addition,

5 Discussion

5.1

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Appendix

A Additional details

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.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.