The Attribution of Climate to Local People's Mental Health

Peilin Chen

2020/12/22

Abstract

Mental illness has become a leading death cause globally, which affects individuals in different forms at distinct life stages. Besides economic biological, economic, and social determinants, there is a consistent investigation of other potential driving factors of mental issues. The effect of climate and its change on individuals' mental health is one factor that researchers have been focusing on these years. This study explored the dataset of General Social Survey on Time-stress and Well-being, conducted in 2010. The propensity score matching technique is utilized to assign the treatment of living in the Atlantic region compared to living in Ontario to the non-randomized dataset by matching scores. Then, another logistic regression model is implemented to look at the causal relationship between self-rated mental health and climate, controlling for age, sex, education, marital status, income, and total duration for active sports. The model results have shown that living in the Atlantic region where the climate is older, income, and being married has a positive impact on enhancing mental status while being separated negatively.

Keywords

Causal Inference, Propensity Score Matching, Climate, Mental Health

Introduction

A recent study has estimated that 792 million people lived with a mental health disorder, which is slightly more than one in ten people globally in 2017 (Ritchie and Roser, 2018). According to the World Health Organization, of the 870 million people living in the European Region, at any one time, about 100 million people are estimated to suffer from anxiety and depression and over 40 million people from other mental disorders. Suicide, one of the most tragic mental disorder results, has become a leading and under-estimated death cause of young adults. These striking and aggressively growing numbers and facts have informed us repeatedly that mental health is currently one of the most critical social issues, and it can affect human beings in different life stages in distinct forms at various levels. Unfortunately, the persistent spread of COVID-19, along with the resulting isolation and loss of income, has further exacerbated mental illness. One arresting data is that in late June, 40% of U.S. adults reported struggling with mental health or substance use (Solomon-Maynard, 2020).

When it comes to the factors that can potentially cause mental disorders, the most discussed categories are biological, psychological, behavioral, social, and economic determinants. What are some other hidden relevant driving factors of mental illness? People have observed that areas with a higher latitude are more likely to have a higher suicide rate than those in lower latitude areas. The significance of the impact of climate on residents' mental status might have been underestimated as there is relatively limited analysis of observational data in this field. To better understand the topic and draw a solid causal inference, this study will utilize propensity score matching methodology to explore the causal link between climate and mental status and to explore any potential policies or advancements that local governments can implement to assist the rising mental health issues and avoid tragic consequences in the cities or countries locating in the higher latitude.

The following analysis is based on 2014 GSS data, and the variables are selected sophisticatedly. In the Methodology Section (Section 2), each variable will be explained and described delicately. Then, the Model Section (Section 3), where the model used for propensity score matching analysis will be illustrated, followed by the Result section (Section 4), representing all the results and associated interpretations. A more in-depth discussion and next-step suggestions will lastly be stated.

Methodolody

Data

Age refers to respondents' age when the survey was conducted, which ranges from 15-year-old to over 75-year-old. The selected variable divides age into ten groups. To reduce unnecessary complexity in the model, the average of each age group is calculated to represent the respondents who fall into that group. For instance, for the age group of 15-year-old to 24-year-old, an average of 20-year-old is taken for further analysis. The average age group of over 75-year-old is assumed to be 80-year-old. Table 1 below indicates that most of the respondents are between 40 to 60-year-old, and the overall distribution is approximately normal distributed with a mean of 49.79 and middle 50% falls within 40 to 60-year-old.

Table 1: Age Background of the Sample Respondents

reduced_data$age	n
20	607
30	823
40	1101
50	1283
60	1310
70	819
80	495

Income refers to respondents' annual income in Canadian Dollars before taxes and deductions from all sources during the year ending December 31, 2009. Like the age variable, income has divided into 12 groups, and the average of each income group is calculated to represent the annual personal income of the respondents who fall into that group. For example, for the income group of 5,000 to 9,999, an average of 7,500 is taken for further analysis. The average age group of over 100,000 is assumed to be 200,000. Figure 1 below shows that the income is right-skewed, with most of the respondents have an annual pre-tax income below 10,000. The right-skewness can also be illustrated through the median of 35,000 is far less than the mean of 45,759.

Sex is a categorical variable which takes on two values: female and male. In the original data, female is assigned the value of 2, and male is assigned the value of 1. For analyzing purpose, this variable is converted to a dummy variable which assigns value of 1 to male and value of 0 to female. Table 2 below shows that approximately 55.9% of respondents are females.

Table 2: Sex Background of the Sample Respondents

sex	n
Female	3601
Male	2837

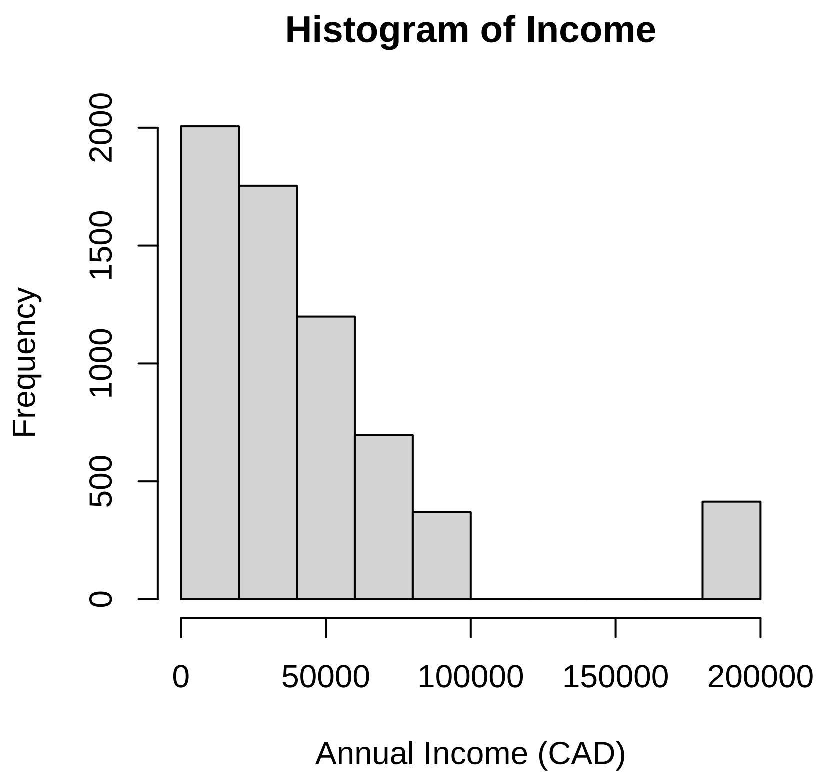

Figure 1: Histogram of Annual Income

Region is a categorical variable, which contains five values: “Atlantic region”, “Ontario”, “Quebec”, “Prairie region”, and “British Columbia”. Only respondents living in the “Atlantic region” and “Ontario” will be kept since this analysis investigates the causal link between climate and mental health. According to the climate data organization website, the annual average temperature in the Atlantic region is 5.3 °C, whereas the annual average temperature in Ontario is 17.9 °C. Therefore, living in the Atlantic region will be the treatment group, and living in Ontario will be the control group. The details regarding the construction of experiments will be explained in the Model Section. Table 3 illustrates that approximately 43.3% of respondents’ residences are in the Atlantic region.

Table 3: Region of Residence of the Sample Respondents

region	n
Atlantic region	2788
Ontario	3650

Education is another categorical variable, which indicates the highest level of degree achieved by the respondents. Answers of “Not Stated” and “Don’t Know” are removed for analyzing purposes. According to Table 4, the respondents with “Diploma/certificate from community college or trade/technical” take up the majority of the sample, following by respondents with “Doctorate/masters/bachelor’s degree” being the second largest group.

Table 4: Education Background of the Sample Respondents

education	n
Diploma/certificate from community college or trade/technical	1862
Doctorate/masters/bachelor’s degree	1713
High school diploma	925
Some secondary/elementary/no schooling	1063
Some university/community college	875

Marital Status is also a categorical variable. Answers of “Not Stated” and “Don’t Know” are removed for analyzing purposes. According to Table 5, the respondents with “Married” take up the majority (52.4%) of the sample, following by respondents with “Single (Never married)” being the second largest group (21.8%).

Table 5: Marital Status of the Sample Respondents

marital_status	n
Divorced	478
Living common-law	439
Married	3373
Separated	223
Single (Never married)	1401
Widowed	524

Total duration (in minutes) for active sports is a numerical variable ranging from 0 to maximum 930 minutes. Not surprisingly, this variable’s distribution is extremely right-skewed, with most of the values below 50 minutes. The right skewness can also be demonstrated through a median of 0 minutes that is far less than the mean of 27.65 minutes.

Mental health is a discrete numerical variable ranging from 1 to 5 inclusively. The value of 1 standing for

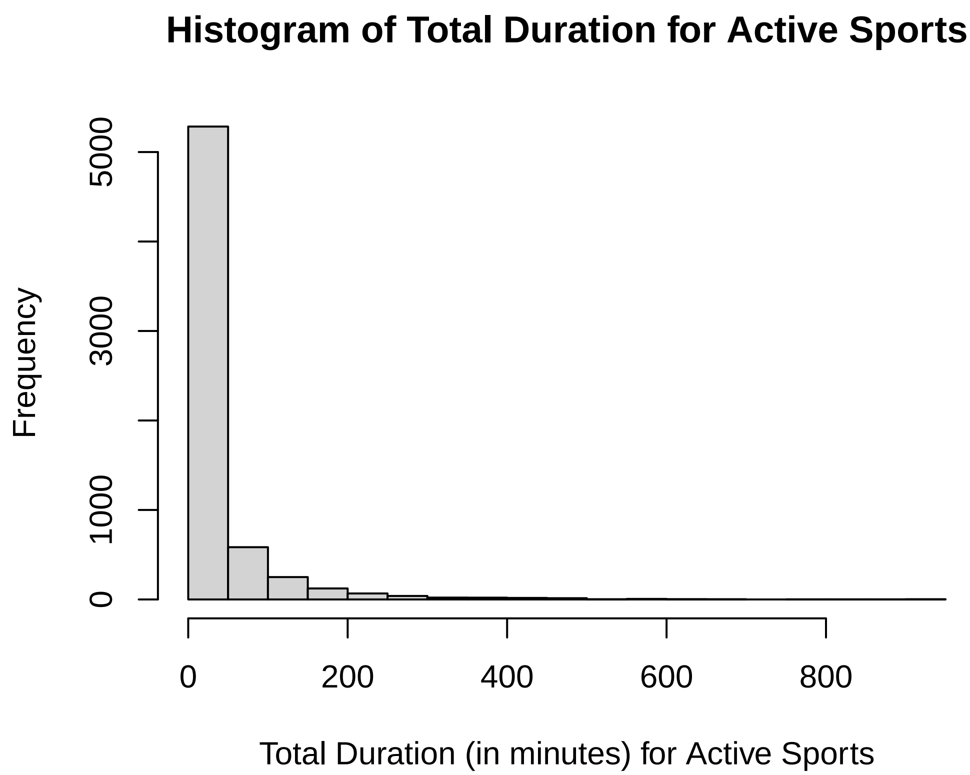

Figure 2: Histogram of Total Duration for Active Sports

"Excellent"; the value of 2 standing for "Very good"; the value of 3 standing for "Good"; the value of 4 standing for "Fair"; and the value of 5 standing for "Poor". Answers of "Not Stated" and "Don't Know" are removed for analyzing purposes. Table 6 below shows that most of the respondents have relatively good mental status, and approximately 1.7% respondents have poor mental health.

Table 6: Mental Health Status of the Sample Respondents

mental_health	n
1	1500
2	2404
3	1918
4	509
5	107

In summary, there are four discrete numerical variables, two dummy variable (including the treatment and control group indicator), and two categorical variables.

Model

Choice of Propensity Score Matching Propensity score matching is a commonly used statistical technique, grounded in the Rubin counterfactual framework, aiming to control for self-selection bias and extend causal inference into non-randomized studies (Rosenbaum & Rubin, 1983). That being said, such a statistical technique can help researchers to assign the treatment to the non-randomized dataset by matching scores. Several matching methods are available, including kernel, nearest neighbor, radius, stratification and so on. This study will implement the matching method of the nearest neighbor.

Since this study is based on the GSS, which is a national census and aims to investigate the causal link between whether living in a region with colder and harsher climate has a negative impact on local residents' mental health, propensity score matching would be helpful to assign treatment to observations for further analysis.

To be more specific on the mechanism of propensity score matching. Assume Treatment A is a binary variable that determines whether an observation receives treatment or not. A=1 for treated observations and A=0 for control observations. The scores are essentially the probability for an observation to be in the treatment group given pre-treatment characteristics. Therefore, a probit or logit model will be utilized for conducting propensity score matching. Assume x variables that may affect the likelihood of being assigned into the treated group; the mathematical notation is shown below:

$$P(A = 1|x) = E(A|x)$$

Then, using nearest neighbor to match observations from treated and control groups based on their propensity scores. Specifically, if two observations receive very similar scores, which means that the probability for both of them to receive the treatment given the pre-treatment characteristics x will be similar, then one observation will be allocated to the treated group with the other one being allocated to control group. In reality, this technique is used for many reasons; one of them is cost-saving and efficiency.

Model Construction The basic statistical characteristics are shown below:

$$Y_i \sim Binomial(N_i, p_i)$$

The entire model construction contains three main steps.

Step 1: Implement a logit regression model to compute propensity score for each observation i. Propensity score is the probability for an observation to receive treatment, where Y_i=1.

$$log(\frac{p_i}{1-p_i}) = logit(p_i) = X_i\beta = \beta_0+\beta_1 x_{age}+\beta_2 x_{education}+\beta_3 x_{income}+\beta_4 x_{maritalstatus}+\beta_5 x_{sex}+\beta_6 x_{sportstime}$$

Where,
- Y_i is the whether the observation is assigned to the treatment group or control group (Y_i=1 if in the treatment group, otherwise Y_i=0) - N_i is the total sample size - p_i is the probability for an observation to receive treatment - X_i is the vector of the independent variable (i.e. age, education, income, sex, total duration for active sports, marital status) - β is the coefficient vector corresponding to each independent variable

Step 2: Utilize arm package and its matching function to find the closest propensity scores of those not treated.

Step 3: Implement a logit regression model to examine the effect of being treated on mental health status. In this case, examining the effect of living in a region with a colder and harsher climate on local people's mental health status. Mental health, which ranges from 1 to 5 initially, has converted to 1 and 0 only. Value of 1 has mental illness, which was 5 (i.e. self-rated as bad mental health) in the original dataset. Value of 0 is having a good mental health, which was 1 to 4 (i.e. self-rated as excellent, very good, good, and fair mental health).

$$log(\frac{p_i}{1-p_i}) = \beta_0 + \beta_1 x_{age} + \beta_2 x_{education} + \beta_3 x_{income} + \beta_4 x_{maritalstatus} + \beta_5 x_{sex} + \beta_6 x_{sportstime} + \beta_7 x_{treated}$$

Where, p_i is the probability for an individual to have bad mental health

Results

Table 7: Coefficient Table of Regression Linear Model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.74570	0.56986	4.81822	0.00000
age	-0.00250	0.00754	-0.33203	0.73987
educationDoctorate/masters/bachelor's degree	0.15377	0.37431	0.41082	0.68120
educationHigh school diploma	-0.48701	0.31175	-1.56216	0.11825
educationSome secondary/elementary/no schooling	-0.30007	0.31007	-0.96777	0.33316
educationSome university/community college	-0.50193	0.33842	-1.48315	0.13804
sex	0.16757	0.22892	0.73201	0.46416
income	0.00002	0.00001	3.53070	0.00041
region	0.58307	0.21404	2.72414	0.00645
active_sports_time	-0.00018	0.00154	-0.11397	0.90926
marital_statusLiving common-law	0.04341	0.41980	0.10341	0.91764
marital_statusMarried	1.27028	0.34060	3.72959	0.00019
marital_statusSeparated	-0.79245	0.40282	-1.96727	0.04915
marital_statusSingle (Never married)	0.72790	0.41088	1.77155	0.07647
marital_statusWidowed	0.87999	0.46577	1.88931	0.05885

```
## Waiting for profiling to be done...
```

Table 8: 95% Confidence Interval of Logistic Regression Model Coefficients

	2.5 %	97.5 %
(Intercept)	1.64080	3.87771
age	-0.01712	0.01248
educationDoctorate/masters/bachelor's degree	-0.55818	0.92455
educationHigh school diploma	-1.09943	0.13057
educationSome secondary/elementary/no schooling	-0.91405	0.30730
educationSome university/community college	-1.15851	0.17799
sex	-0.27351	0.62676
income	0.00001	0.00003
region	0.16932	1.01119
active_sports_time	-0.00275	0.00335
marital_statusLiving common-law	-0.77997	0.88174
marital_statusMarried	0.57711	1.92336
marital_statusSeparated	-1.58686	0.00810
marital_statusSingle (Never married)	-0.08716	1.53217
marital_statusWidowed	-0.01384	1.83689

As indicated in the summarized table above, income is one of the explanatory variables that are statistically significant at 1% significance level. Income has a positive impact on mental health that individuals with higher income are more likely to have better mental status. Specifically, for one Canadian dollar increase in annual pre-tax income, the respondents' log odds of having a good mental status would increase by 0.00002. To make it a more realistic interpretation, for 1,000 CAD increase in annual pre-tax income, the respondents' log odds of having a good mental status would increase by 0.02.

Another statistically significant independent variable is "married". Being married has a positive impact on health status, which is statistically supported at 1% significance level. In other words, being married can enhance the likelihood of being mentally healthy. To be specific, being married would increase the respondents' log odds of having decent mental health by 1.27. On the contrary, factor "separated" has an inverse effect on mental health that being separated undermines the probability of being mentally healthy. To be specific, being separated would decrease the respondents' log odds of having good mental health by 0.792, which is statistically significant at 10% level.

In contrast to the initial expectation, living in the Atlantic region, where climate is colder and harsher, turns out to positively impact residents' mental status. Specifically, compared to living in Ontario, living in the Atlantic region would increase respondents' log odds of being mentally healthy by 0.583, which is statistically supported at 5% significance level.

Other than marital status and income, the factor of education, sex, age, and total duration (in minutes) for active sports are not statistically significant at even 10% level. This indicates that these variables have a limited impact on respondents' probability of having good mental health.

Discussion

Summary

By far, to investigate the causal relationship between climate and mental health, explanatory variables of age, income, sex, marital status, highest education degree achieved, and total duration for active sports are selected to construct the treated and control group. Before implementing propensity score matching, the variable region is converted to have a value of 1 if the respondents live in the Atlantic region in Canada and a value of 0 if the respondents live in Ontario of Canada. Living in the Atlantic region is the treatment. Propensity scores are computed by utilizing a logistic regression model, with region being the binary variable.

Then, the nearest neighbor approach of matching is adopted to find the closest scores of the ones that were not treated to each one that was treated. After assigning the treatment to non-randomized dataset by matching scores, another logistic regression model is conducted to see the effects of living in a region with a colder and harsher climate on respondents' mental health status. Moreover, the results of models indicate that higher income and being married would increase respondents' probability of having good mental health while living a separated marital status would decrease respondents' probability of having good mental health. Surprisingly, according to the models implemented and presented above, living in the Atlantic region turns out to have a positive impact on respondents' mental health, rather than making them more likely to be depressed.

Conclusions

In the regression model results, income is a statistically significant factor that would enhance a person's mental health. In fact, mental health is shaped largely by the social and economic environment in which people inhabit (WHO, 2015). There are abundant researches have demonstrated that poverty's negative impact on ones' mental health. For instance, England research has shown that children and adults from the lowest quintile (20%) of household income are three times more likely to have common mental health problems and nine times as likely to have the psychotic disorder (WHO, 2015). It is reasonable to investigate such phenomenon as poverty predisposes individuals to social deprivation. The poor are more likely to encounter unfair and unfriendly treatments, have less time for sports and hobbies, experience a sense of insecure, and so on. All these pieces combine together is likely to cause anxiety and depression for a person.

Besides income, an economic determinant, marital status is one of the social determinants that could potentially affect ones' mental health. Recent studies have underlined the association between mental illness and coming from fractured, dysfunctional, and fatherless families. Poor attachment and family discord affect the timing of the onset of puberty, which in turn, could contribute to conflict with parents, low self-esteem and associations with deviant peers (WHO, 2015). A recent study from Good Childhood Inquiry has found that children with separated, single or step-parents are 50 percent more likely to suffer from mental illness (WHO, 2015). Children, adults who have a marital status of being separated, are also likely to result in a similar adverse consequence. The model above has also helped highlight the negative impact of having a separated marital status on mental health. On the contrary, the model above has instead indicated that being married, on average, has a positive effect on respondents' mental status. In addition to families and marital status, some other social determinants could be health care adequacy, different kinds of discrimination, and employment conditions.

However, one big surprise of the model conducted above is that living in a region with a colder climate does not have an adverse impact on respondents' mental health. On the contrary, living in the Atlantic region is beneficial for one's mental health, controlling all other variables constant. What are the reasons behind this striking result? It could be the extent of climate change, but not the variability in temperatures, which plays a role in one's mental status. Differing from climate change, variability in temperatures is recognized as a risk factor of heat-related deaths, with a robust association between the proportion of hot days and mortality rates (IRP, 2014). Therefore such deaths have resulted from physical illness but not mental illness. Nevertheless, previous studies completed by Australian researchers have reported that there are higher levels of psychological distress and hopelessness among people exposed to environmental degradation and disruptions to agriculture as a result of persistent drought and weather disasters (IRP, 2014). A comparison of the recent climate change in the Atlantic region and Ontario can potentially explain the positive effect of living in the Atlantic region on respondents' mental health. According to National Geographic, the temperature has raised an average 1.6 degrees Celsius between 1948 and 2013, and the annual precipitation has enhanced over 200 mm. Yet the climate change is more severe in Ontario that the average temperature in Ontario has increased by approximately 1.5 degrees Celsius, but in an annual basis. In the most recent decades, Ontario has a higher frequency of suffering from severe weather events such as record-breaking storms, floods, droughts and heat waves (Gough & Anderson & Herod, 2016). Thus, knowing that Ontario has encountered a more aggressive climate change in the recent decades compared to Atlantic region, a deeper insight into the positive "estimate coefficient" of treatment indicator can be generated.

Weakness & Next Steps

The entire study is based on the data from 2010 GSS study. The first weakness was the low response rate. According to the user guidebook provided by Statistics Canada, the overall response rate was only 52.9%, even though a few techniques had been used to prevent non-response such as rearranging appointments to call back when the timing of the first call was inconvenient, and recontacting up to two times or elaborating the importance of the survey. The relatively low response rate was a major source of non-sampling errors in the survey, potentially devaluing the results. Moreover, the extent of non-response varied from partial non-response to total non-response, for which the non-response would be discussed in detail below. A survey's response rate was long viewed as a significant survey quality indicator, and a low response rate could give rise to sampling bias. Furthermore, primarily because of the non-response, the actual sample size of 33,127 records was approximately 5,000 less than the target size of 39,674, adding another layer of risk of non-representative bias caused by the smaller than expected sample size.

Besides total non-response, partial non-response should be regarded as another weakness needed to be discussed with attention. Such non-response occurred when the respondent did not understand or misinterpreted a question, refused to answer a question, or could not recall the requested information. All the above scenarios would lead to a missing value, yet with distinct degrees of bias and different population groups being under- or over-represented. Some of the possible approaches to mitigate the disadvantages of missing value would be discussed in the following section.

The next weakness came from the sampling methodology used by General Social Survey (GSS). GSS used a stratified design, with significant differences in sampling fractions between strata. As a result, some areas were over-represented in the sample, while some were relatively under-represented. With the non-response issue, we argued above; such a sampling method made the sample even less representative.

The causal relationship between climate and mental health requires further investigation. Two potential reasonable future steps can be taken into considerations. Firstly, the current model has a limited number of explanatory variables to be chosen from. Besides age, sex,income, education, marital status, and total duration for active sports, there could have been other reasonable and significant factors, allowing a more in-depth insight into the topic. Thus, our current model could have drawn more specific and valuable conclusions if other variables were involved in computing the propensity score matching. In addition, the technique of post-stratification can also be utilized for the last logistic regression model, which examines the probability of a respondent having good mental health given included explanatory variables. By doing so, a more accurate and reliable result can be obtained as the size of each group, which compromises the entire sample, is being considered and weighted.

References

- Ritchie, H. (2018, January 20). Mental Health. Our World in Data. https://ourworldindata.org/mental-health
- 10 Surprising Mental Health Statistics From 2020. (2020, November 5). Mental Health First Aid. https://www.mentalhealthfirstaid.org/external/2020/11/10-surprising-mental-health-statistics-from-2020/#:%7E:text=In%20late%20June%2C%2040%25%20of,nation%20about%20%24210.5%20billion%20annually
- The Ontario Climate Change and Health Toolkit - Ministry Reports - Publications - Public Information - MOHLTC. (2016). Ontario Ministry of Health. http://www.health.gov.on.ca/en/common/ministry/publications/reports/climate_change_toolkit/climate_change_toolkit.aspx
- Story Map Journal. (2020). ESRI. https://www.arcgis.com/apps/MapJournal/index.html?appid=ae5635f46af9438bb96713d665fbad5c#:%7E:text=Environmental%20Impacts&text=The%20environmental%20impact%20of%20climate,has%20increased%20over%20200%20mm
- Determinants of Mental Health. (2015). World Health Organization. https://www.researchgate.net/publication/282399773_Determinants_of_Mental_Health
- Bourque F, Willox AC. Climate change: the next challenge for public mental health? Int Rev Psychiatry. 2014 Aug;26(4):415-22. doi: 10.3109/09540261.2014.925851. PMID: 25137107.