A Linear Regression on Global Suicide Rates and Country Income, Region, Age, and Sex in 2019*

An analysis of global suicide rates in 2019 using multiple linear regression

Hyuk Jang

June 2, 2024

Suicide has become a main cause of death globally over the past decades. This creates curiosity in what immutable factors may be affecting suicide rates. This paper looks into the regression between global suicide rates and sex, age, country income, and region. Our results show that there is a weak correlation to age and region/country income affect suicide rates more. Then further analysis is done by studying what factors generate such tendencies.

Table of contents

*Code and data are available at: https://github.com/anggimude/Health.

1 Introduction

Suicide is defined as the act of intentionally causing one's own death. The main risk factors of suicide are mental disorders, physical disorders, and substance abuse. In particular there is a significant correlation between suicidal behaviour and mental health. Most suicides are related to any sort of mental disorder ranging from trauma, personality, anxiety, depression, substance abuse, and so forth. In 2019, 1.3% of all deaths resulted from suicide and is the fourth leading cause of deaths in 15~29 years old. This paper will dive deeper into suicide rates and its correlation between country income, age, gender, and region in 2019. The most recent year with relevant data is for 2019, from the WHO(Organization (2021b)) and since we are only looking into one year, we look at the global trends and world bank income group trends to get a more thorough understanding of the trends of suicide rates of 2019. Because the main goal of the paper is to observe the correlation between suicide rates and country income and region, the countries are classified based on the world bank income group; with four income groups - low, lower-middle, upper-middle, and high. Low-income economies is defined as a country with a gross national income less than $1135, and in the manner lower-middle is between the range $1136 to $4465, upper-middle in the range of $4455 to $13845 and high is $13846 or more. Region is defined based on the continents; Asia, Africa, North America, South America, and Europe. The total number of countries is 183 which are all mapped to different regions and income groups.

This paper will produce a multiple linear regression to analyze the regression coefficients of each factor. The results of the model will be checked by doing a 95% confident interval on the regression coefficients. In addition, we will be able to study the trends of the factors of suicide. We want to analyze the change in demographics for the different age groups, country income, region, and gender. The data was obtained from the WHO database under the mental health category(World Health Organization (2021)). R Core Team (2023) was used to clean the raw data to what we use for modelling in other to write an analysis of the intended study. The estimand of this paper are the intercept and coefficients of each predictors which correspond to region/income/sex and the different age categories.

This paper has 4 sections in total not including the introduction. In Section 2 we look at the data that used to carry out the reports including tables and graphs of cleaned data that will be used for the models and the summary statistics. In the next section, we discuss about the models that will be used to analyze our cleaned data, how it is set up and the justifications of it. Next we display and examine the results obtained from the models including tables of the model summaries which helps us make predictions. Lastly, we make final discussions of our results and research based on each cause and dive into some weaknesses that our paper has. In addition, we explore some next steps we or anyone else interested is willing to take after reading this paper.

2 Data

2.1 Raw Data

The data used in this paper is derived from WHO(Organization (2021a)), and WHO data repository(Organization (2021b)) and was downloaded from the data available in the links. WHO(Organization (2021a)) provides data of age-standardized suicide rates per 100,000 population for 183 countries for the years 2000-2019. WHO data repository provides data of crude suicide rates per 100,000 population for the same 183 countries for the year 2019. Since the only data available data for crude suicide rates depending on age is 2019, we only look into the year 2019 for the age-standardized dataset as well. The data the two tables include are the 183 countries, sex (both, male, female), age groups of 85+, 75-84, 65-74, 55-64, 45-54, 35-44, 25-34, 15-24, and age standardized suicide rates. The age-standardized suicide rate is used as the dependent variable for our model as we are interested in the correlation between the independent variables which are gender, age group, country income, and region.

The cleaning, testing, and modelling of the data for this paper was done through R (R Core Team 2023) with the aid of the following packages: tidyverse (Wickham et al. 2019), dplyr (Wickham et al. 2023), rstanarm (Goodrich et al. 2020), ggplot2 (Wickham 2016), modelsummary (Arel-Bundock 2022), kableExtra (Zhu 2021), arrow(Richardson et al. 2024), tidybayes(Kay 2023).

2.2 Cleaned Data

The goal of the cleaning process of this paper is to create a table including income group(high, upper middle, lower middle, low), sex(both, male, female), region(Asia, Europe, North America, South America, Africa, Oceania) as the rows, and age-standardized suicide rates, and the 10 year age group suicide rates. This is so that we can create a model which will provide us the correlation of global suicide rates of 2019 against income group/sex/region and age groups. To do this we merge the two raw data tables we have downloaded. Before we merge the tables, WHO includes the age-standardized suicide rates as a form of the actual value[lower limit, upper limit]. When we create our models or graphs, the interval is not necessary, so we first remove the interval so that each column only consists of the actual rate. After merging and doing some minor cleaning like removing repeated rows and changing names of columns and rows, the table will have country, sex, age-standardized suicide rates, and the 10 year age groups. We must also include the region and income group of each data so we create a mapping so that we country will have additional columns of continent and world bank income group. This was done based on the information provided from (Bank (2024)) and (Worldpopulationreview.com (2024)). Since we are looking into the global suicide rates and how the independent variables affect it, we must calculate the averages of the age-standardized suicide rates and age groups depending on sex, income group, region. The table with age-standardized suicide rates and suicide rates for the 10 year age groups depending on the 183 countries is available in Section A.

Below, Table 1 is a table of the averages of suicide rates of age standardized and 10 year age groups depending on region, country income, and sex. The rows include global, male, female, lower income, lower-middle income, upper-middle, high income, Asia, Europe, North America, South America, Africa, and Oceania. We observe the average suicide rates of each of these variables for age-standardized and the 10 year age groups. The most noticeable trend from Table 1 is the difference in suicide rates in male and female. While female had a suicide rate of 4.54, male's was 16.14 and the global average being 10.09 exhibiting significance of further research. Within the income groups, we see that the lower income groups tend to have a higher average suicide rate compared to the higher income countries. The lower income groups have an average of around 11.50 while the comparatively higher income is 9.00. The difference is not significant as the global average is 10.00, but it is worth looking into these trends in Section 5. Comparing the data of each region, we can notice that Asia, North America, and Europe are below the average, on the other hand, South America, Oceania, and Africa have suicide rates above the average. For the regions, the deviation from the global average is larger than income groups, but smaller than sex. Here we can notice the similarities with the income group data. For example, a large proportion of African and South American countries lie in the lower and lower-middle income bracket which has a comparably higher suicide rate, and it shows that South America and Africa have a suicide rate of 11.44 and 13.40 respectively.

Figure 1 and Figure 2 shows the Table 1 as a scatter plot. The graph is divided into two to improve clarity and legibility. The split allows for easier viewing of labels, axis, and point, ensuring a better understanding of the data. The scatter plot is build for an analysis of the age group trends. We can observe that the older age groups, Age 85+ and Age 75-84 have the highest suicide rates. Especially the 85+ age group is an outlier in every region/income/sex group. We can also notice that the younger the age, the lower suicide rates tends to be. Some notable exceptions are Oceania's high suicide rate among the ages 15-44. Unlike the other region/income/sex groups, Oceania's younger generation and 85+ has the highest suicide rates. In addition, the older the age group, we notice more extreme suicide rates that goes above 100 per 100,000 population. This occurs three times all of which are in 85+ of age, and male, lower income, and Africa displays such results. The x-axis variables other than Africa, male, and lower income, the data points tend to be gathered up around the average of each variable, meaning that the standard deviation of each region/income group/sex is small.

Table 1: Global Suicide Rates by Age, Region, Sex, and Income Group

Region/Income/Sex	Age-standardized suicide rates (per 100 000 population)	Age 85+	Age 75-84	Age 65-74	Age 55-64	Age 45-54	Age 35-44	Age 25-34	Age 15-24
Global	10.09	59.32	27.47	20.90	17.56	15.46	12.90	10.80	7.79
Male	16.14	121.59	48.42	34.47	28.35	24.87	20.75	17.01	11.33

Region/Income/Sex	Age-standardized suicide rates (per 100 000 population)	Age 85+	Age 75-84	Age 65-74	Age 55-64	Age 45-54	Age 35-44	Age 25-34	Age 15-24
Female	4.54	27.26	13.56	10.12	7.91	6.40	5.12	4.42	4.06
Lower Income	11.08	131.75	50.75	37.67	25.74	17.13	9.64	6.68	4.54
Lower Middle	11.97	75.86	31.40	24.30	19.63	18.29	16.23	12.71	9.54
Upper Middle	8.25	40.10	20.62	15.29	13.09	11.97	11.01	9.87	6.95
High Income	9.39	25.35	18.55	14.50	15.60	15.00	12.88	11.71	8.38
Asia	6.92	28.96	15.51	11.13	9.54	9.28	9.42	9.10	7.10
Europe	9.73	32.25	22.08	17.31	17.89	16.80	13.20	10.55	7.10
North America	5.60	14.08	10.96	8.76	8.30	7.42	7.34	7.39	5.84
South America	11.44	34.30	21.25	17.77	16.79	14.37	14.70	16.32	13.03
Africa	13.40	131.11	50.86	37.99	27.77	22.81	15.69	9.88	5.74
Oceania	15.17	43.67	17.93	15.25	15.86	14.77	21.15	24.73	22.62

2.3 Basic Summary Statistics

Table 2 is a representation of an summary of the suicide rates of the different age groups. The summary includes the minimum, maximum, mean, standard deviation, variance, and number of observations. The observations is 13 as the summary is created from Table 1, and we have 13 rows representing region/income group/sex. It is shown that the age-standardized suicide rate has a mean of 10, with a minimum of 5, and a maximum of 16. Again it is noticeable that the older the age, suicide rate is higher. Moreover, the variance also gets larger, the older the age group. For example, because age 85+ group has a range of 14 to 132, and a mean of 59 which leads to a extreme variance of 1806 as the data is skewed off the global averages. As the younger the age group, the summary becomes more similar to the global average. The age groups from 15-74 are clustered around a mean of 15, and the ages 75~ consists of means higher than the previous group.

3 Model

The goal of the Bayesian model is to incorporate prior knowledge, such as insights from previous studies or analyses, into the selection of the model. In this paper, we use multiple linear

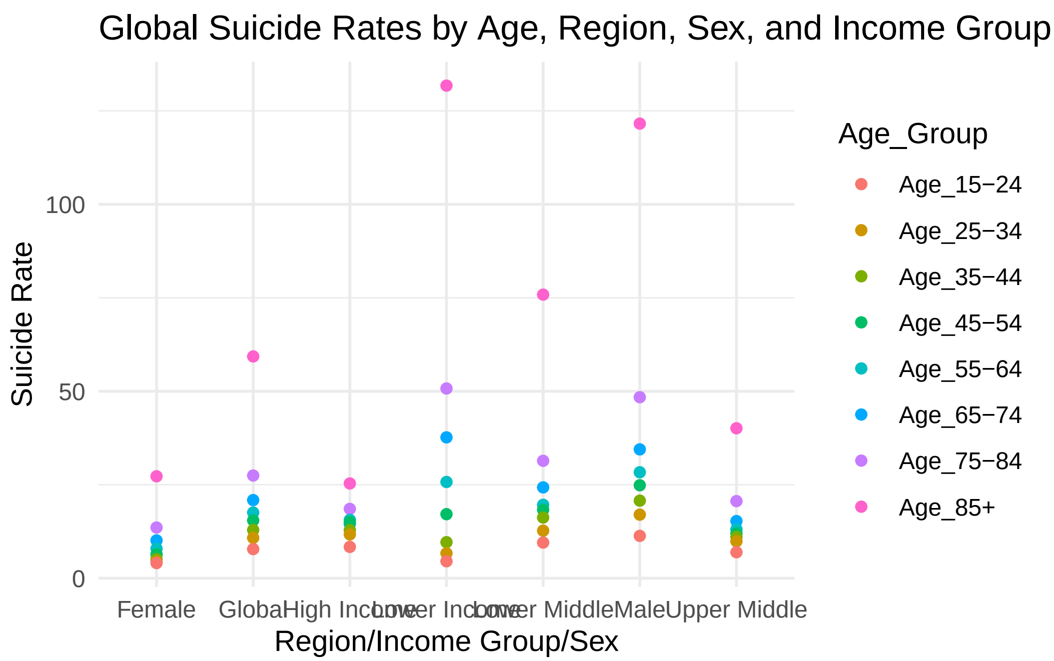

Figure 1: Global Suicide Rates by Age, Region, Sex, and Income Group - 1

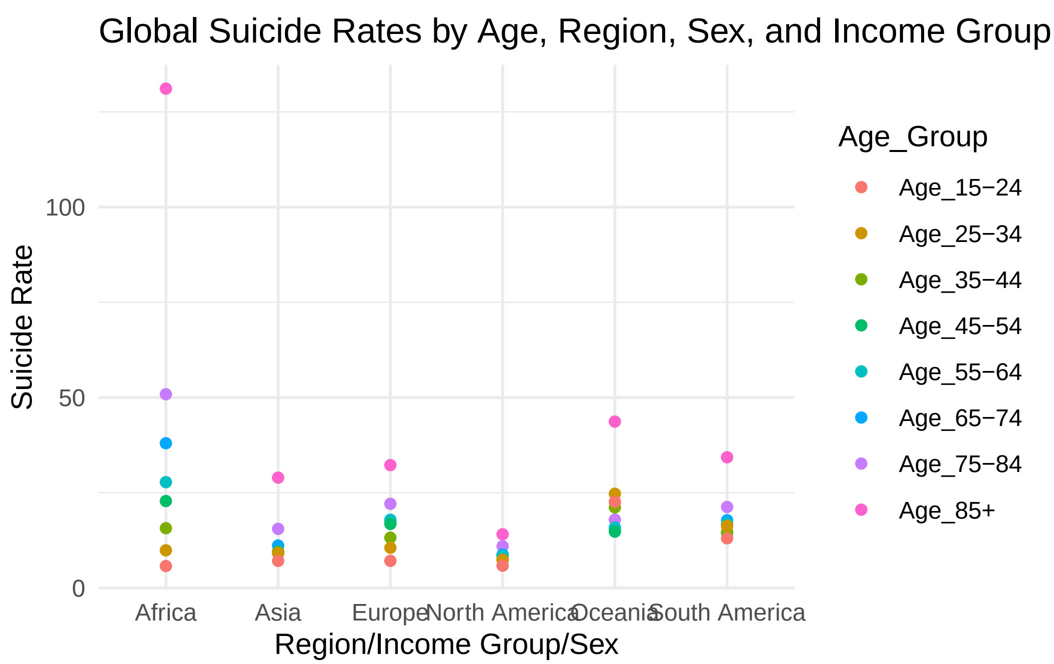

Figure 2: Global Suicide Rates by Age, Region, Sex, and Income Group - 2

Table 2: Summary statistics of suicide rates of different age groups

	Min	Mean	Max	SD	Var	N
Age-standardized suicide rates (per 100 000 population)	5	10	16	3	12	13
Age_85+	14	59	132	42	1806	13
Age_75-84	11	27	51	14	203	13
Age_65-74	9	20	38	10	104	13
Age_55-64	8	17	28	7	47	13
Age_45-54	6	15	25	5	29	13
Age_35-44	5	13	21	5	22	13
Age_25-34	4	12	25	5	28	13
Age_15-24	4	9	23	5	24	13

regression because this model is commonly applied when predicting a continuous outcome variable based on multiple predictor variables. The normal (Gaussian) distribution is effective when used for modeling scenarios where the residuals of the data are assumed to be independent and normally distributed around the regression line. Table 1 captures how suicide rate changes depending on various demographic segments. Given that our response variable, suicide rate is continuous and follows a normal distribution as we have seen from Figure 1 and Figure 2, each region/income group/sex demographic consists of data points clustered around its mean. Thus, the Gaussian family model from the Bayesian framework allows a flexible approach to understanding the relationships between the continuous outcome (age-standardized suicide rates(per 100,000 population)) and predictors (region/income/sex, age categories) with the benefits of integrating prior knowledge and quantifying uncertainty around estimates.

3.1 Model set-up

The multiple linear regression models utilized in this paper is run on R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2020). We use the default priors from `rstanarm`. The response variable y_i is defined as an individual observation of 'Age-standardized suicide rates (per 100 000 population)' for each combination of the explanatory variables. The intercept α represents the intercept of our linear model. It is the expected value of y_i when all predictors are held at reference level. Each β_j is the coefficient associated with each predictor in our model. This would correspond to the 'Region/Income/Sex', 'Age 85+', 'Age 75-84', ... , 'Age 15-24'. β_j tells us the change in the expected suicide rate per unit change in the predictor assuming all else is held constant. We have specified that both α and β_j will have normal priors with a mean of 0 and a standard deviation of 2.5. With this, the `rstanarm` model will internally adjust the priors based on the scale of the predictors to help with the convergence and effectiveness. The error term σ is the standard deviation of the residual errors in our model. By default, the `rstanarm` uses the student-t distribution with 3 degrees of freedom for the standard deviation.

$$
\begin{aligned}
y_i \mid \mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma^2) && (1)\\
\mu_i &= \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} && (2)\\
\alpha &\sim \text{Normal}(0, 2.5) && (3)\\
\beta_j &\sim \text{Normal}(0, 2.5) \quad \text{for } j = 1, 2, \ldots, p && (4)\\
\sigma &\sim \text{Exponential}(1) && (5)
\end{aligned}
$$

3.1.1 Model justification

Based on our inference from Section 2, we expect a stronger positive coefficient in male, and the relatively lower income groups. Because there is a higher suicide rate among males and lower income and lower-middle income groups, we can expect that when there is an increase in the suicide rates of these predictors, age-standardized suicide rate is likely to increase as well. Female and the higher income countries have lower suicide rates than the global average in which helps us conclude that an increase in suicide rates of these predictors isn't likely to increase the global average rate leading us to expect a negative coefficient. The results of the model will let us verify whether our expectations are true or false.

4 Results

4.1 Overview of model results

Our results are summarized in Table 3. We are primarily interested how global age standardized suicide rate depends on each of its predictors. The multiple linear model provides us with the estimates for the intercept and coefficients for the predictors which are male, female, global, lower income, lower-middle, upper-middle, high income, Asia, Europe, North America, South America, Africa, Oceania, age 85+, age 75-84, age 65-74, age 55-64, age 45-54, age 35-44, age 25-34, age 15-24. The intercept represents the estimate of age-standardized suicide rate when the predictors are zero. The coefficients represent the additional suicide rates per 100,000 population associated with each predictor. The model results will display the estimates - posterior means or medians for each coefficient including the intercept, uncertainty measures - credible intervals. The output values of each predictor is the regression coefficient, meaning how much the outcome is expected to increase or decrease with a one unit increase in the suicide rate of that predictor, holding all else constant. The value in the brackets represent the Median Absolute Deviation of the posterior distributions of the coefficients. It conveys the dispersion around the median of each coefficient's posterior distribution, exhibiting how spread the distributions are. Num.Obs represents the number of observations made in the model. R2 is the R-squared value which is the proportion of variance in the dependent variable that can be explained by the independent variable. The R2 adj is the adjusted R squared which accounts for the number of predictors used. Log.lik is the log-likelihood which gives us an idea of the likelihood of the data, higher is better, but this is typically used for comparison between models. ELPD and ELPD s.e. explains the log predictive density and its standard error. The ELPD measures the sum of the log predictive densities for each observation, used for model comparison. LOOIC is an acronym for leave-one-out information criterion in which a lower value indicates a model with better out-of-sample predictive performance. WAIC stands for Watanabe-Akaike information criterion which is another measure of good fit; lower values are better fit. RMSE is the room mean squared error measuring the model's predictive performance where the lower values mean more accurate predicts.

4.2 Multiple linear regression results

Table 3 is a table of the results from our model. The intercept is the baseline value which in our case is 3.07. Asia shows a negative coefficient of -1.45, suggesting lower values compared to the baseline, with a large standard error of 12.58. This can be interpreted as the predictor increases, the outcome tends to increase. Conversely, a negative coefficient indicates a decrease. Europe has a negligible negative effect of -0.11 on the outcome with a large standard error of 11.24. Female has a more substantial negative impact of -2.07 and the standard error of 16.91 defines considerable variability in the estimate. Global, high income, and lower income groups show minimal negative coefficients suggesting slight decreases from the baseline. Lower middle income presents almost a neutral effect of 0.01 which is surprising. Male shows a positive relation of 0.91 compared to the baseline. North American and Oceania have more notable negative and positive coefficients respectively in which we can observe significant regional variations. In general, age category shows an increasing trend in coefficients as age decreases. Age groups of 55~ has a very small coefficient, stipulating minor increases compared to the baseline, all with small standard errors. The younger age groups from 15-54 suggesting that younger ages are associated with increasing values of the age standardized suicide rate. With the model having a R-squared value of 0.791 is a good level of explanatory power supporting the accuracy of the results. The adjusted R-squared value is at 0.596, which is lower than the R-squared value which may be because the number of observations is only 13 not the whole dataset before getting the averages. The RMSE is at 0.64 telling us the average magnitude of the model's prediction errors and the value indicates better predictive accuracy. Thus, by ignoring values like LOOIC and WAIC as we don't have another model to compare it to and analyzing other results, we can conclude that the model is performing well on the dataset we have cleaned. Further evaluation of the model results will be done in Section B.

Table 3: Multiiple linear model summary

	Gaussian(Normal)
(Intercept)	3.07
	(26.07)
Region/Income/Sex‘Asia	−1.45
	(12.58)
Region/Income/Sex‘Europe	−0.11
	(11.24)
Region/Income/Sex‘Female	−2.07
	(16.91)
Region/Income/Sex‘Global	−0.49
	(7.91)
Region/Income/Sex‘High Income	−0.35
	(11.26)
Region/Income/Sex‘Lower Income	−0.53
	(10.81)
Region/Income/Sex‘Lower Middle	0.01
	(7.81)
Region/Income/Sex‘Male	0.91
	(10.74)
Region/Income/Sex‘North America	−1.70
	(14.63)
Region/Income/Sex‘Oceania	0.99
	(22.82)
Region/Income/Sex‘South America	0.08
	(11.91)
Region/Income/Sex‘Upper Middle	−1.02
	(10.47)
Age_85+	0.01
	(0.17)
Age_75-84	0.02
	(0.54)
Age_65-74	0.04
	(0.73)
Age_55-64	0.07
	(1.16)
Age_45-54	0.07
	(1.37)
Age_35-44	0.11
	(1.59)
Age_25-34	0.12
	(1.42)
Age_15-24	0.13
	(1.50)
Num.Obs.	13
R2	0.791
R2 Adj.	0.596
Log.Lik.	−26.819
ELPD	−39.0
ELPD s.e.	0.3
LOOIC	77.9
LOOIC s.e.	0.7
WAIC	73.7
RMSE	0.64

5 Discussion

5.1 Insights into discrepancies in sex

In the previous sections we have noticed unusual trends; the discrepancy in sex. Table 2 explains that there is almost 12 people more who suicide in the male population per 100,000 population. Some of the explanations to the high rates of suicide in male population is from unwillingness to engage in health help-seeking, lack of gender-sensitive mental health services, impulsivity, alcohol and drug use, and use of lethal means. However, it is known that female tend to have higher rates of depression diagnoses and more suicide attempts than male does. The gender paradox in suicide explains such abnormal trends. While females have suicide thoughts more often, males die by suicide more frequently. This is because the methods females use are less immediately lethal. Male tend to use more violent methods such as hanging, carbon monoxide poisoning, or firearm use while females tend to rely on drug overdose and hanging. This makes it more likely for males to die from suicidal attempts compared to females. Though, this may not be the only reason for such discrepancies. Other explanations may include the pressure of gender roles. Traditional male gender roles highlight attributes like strength, independence, willingness to take risks, and financial success and the pressure of society may halt males from seeking help and support. For example, in the traditional male gender role, men are the ones who earn and maintain the money of a family. When men become vulnerable to job loss or financial difficulties, the significant stress can contribute to suicidal thoughts. Moreover, men are more likely to abuse alcohol and drug as a coping mechanism for mental health issues. This may be the factors that lead to the abnormal trends we see from our analysis. Thus, it important for governments to note that different approaches must be used to prevent suicide due to this contrasting data.

5.2 Looking into discrepancies among income groups

From the summary and analysis of the Table 2 we have seen that South America, Oceania, and Africa has a majority of countries in low and lower-middle income groups. This explains a correlation between income groups and regions as the relatively lower income countries have higher suicide rates leading to trends in the region predictors. Some of the main reasons to this discrepancy is economic stability and unemployment, access to healthcare, means of suicide, preventative policies. Higher income countries typically have more stable economies and lower levels of poverty which may affect mental health positively or at least not negatively. Conversely, in lower income economies, with high unemployment rates and financial instability can increase stress levels and depressive symptoms leading to higher suicide rates. In general, higher income countries have better healthcare infrastructure and mental health services. This can lead to more effective identification and treatment of depression or any mental health disorders leading to suicidal behavior. The availability and commodity of means of suicide can vary by income level and region. For example, in South East Asia where agriculture is prominent, it may be easier to find pesticides compared to wealthier countries. While, in higher income countries, strict gun laws and regulations can lower the risk of suicide, decreasing suicide rate as a result. Income gaps and regional differences may increase the risk of suicide as it these types of means may be more accessible to the public due to the lack of government regulation. Not only this, higher income countries may have more resources to create public advertisements. This includes public education, suicide hotlines, and any other support system like counseling. These resources may not be prevalent in lower income countries as it requires human resources, money, and time for all this to be achieved. Thus, these reasons may explain the reason why we saw discrepancies among income groups and regions from Table 3.

5.3 Further analysis on unemployment and suicide rates

To look deeper into the correlation between unemployment and suicide rates we have created Figure 3 using data from the IMF(Fund (2021)). The red points of the graph shows the unemployment rate and the bar chart displays suicide rates for each country. For simplicity, we have removed the countries that are not common in both datasets. From the previous sections the data discussed the relationship between unemployment and suicidal behavior from the income group discrepancies. By observing the graph, it is difficult to notice any correlation as there is no obvious global trends. However, there still seems to be a small correlation as some data points exhibits that a small size of the countries do have a weak correlation between suicide rates and unemployment. For example, in the graph, countries like Argentina, Azerbaijan, Belize, Chile, China, and so forth show very small differences between suicide rates and unemployment and show correlations. This implies that there must be at least a very small correlation between unemployment and suicide rates but looking at the global data, though these trends may be negligible on a global level, as there is a plethora of factors that account for suicide rates. Despite Figure 3 does not display significant results like we expected, there is importance in looking into such data. However, for individual countries to analyze such correlations among causes of suicide as on a smaller scale, for specific countries, the instability of economy may be a leading cause of suicide. Thus, when governments build policies for suicide prevention they may look into the data of diverse factors that affect suicide rates and utilize elimination methods to find out which factors weigh more in accounting for suicide rates.

5.4 Weaknesses and next steps

There are some weaknesses that occur throughout the paper that are crucial to note. One is that the data we used for modelling only has an observation number of 13. This is because what we are studying requires the data to be a mean depending on the mapping we create. Because the number of observations is small, the model results may not be accurate enough due to the low iterations it has. This can be improved if there is a method to model the tables in such a way that we can maintain the original dataset so that the observations would be the same size as the number of countries in each category instead of 13. Another weakness

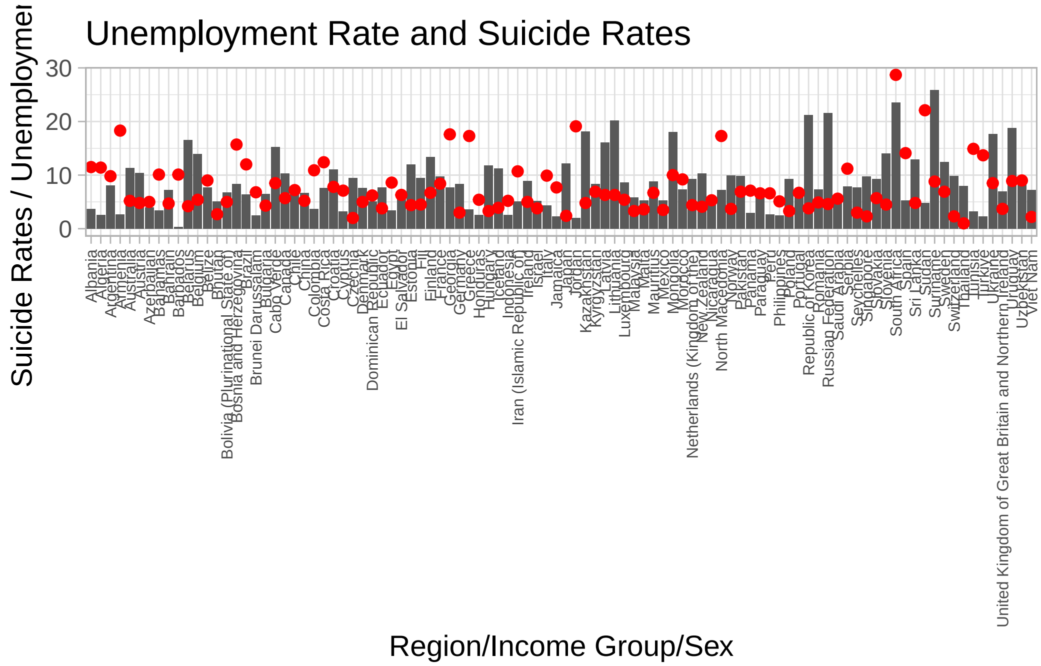

Figure 3: Unemployment and suicide rates

occurs from the models as well. As we have mentioned in Section 4.2, we only have a single multiple linear model for our research which puts us on a disadvantage as we can't compare the posterior predictive checks, LOOIC, WAIC, and more. This stops us from concluding our results with 100% confidence because it may appear to be that there exists a better fit model for the dataset. The lack of usage of a variety of models halts us from getting a deeper and more accurate understanding of the study. Lastly, another weakness may arise form the generalization of the data. This paper looks into the global suicide trends depending on the predictors we have chosen. This is done by downloading suicide rate data by country from the WHO. Because the raw data downloaded despite the accuracy is still an estimate of suicide rates per 100,000 population, and the averages calculated for each predictor is an average of an estimate. This can lead to some minor calculation errors leading to potentially different results. In addition, the values we find for each predictor may not be the best representation of the data from predictors as there may be some countries missing from the dataset which should have been calculated. For future studies, it would be suggested to use a more variety of models such as the multilevel modelling to increase the accuracy of the models by cross validation or comparing the results from other models as well. Further insight can be made like we did for Section 5.3. For example, one can research the correlation between data like social support systems, life satisfaction, social media usage and more. Studying deeper into such correlations will let us notice which factors influences suicidal behavior and suicide rates. Nowadays as suicide rates increase and become a significant cause of death globally, it is crucial for governments and policy makers to know what is the leading cause of suicide from each country/region/income group/sex to create different preventative policies to create a healthier society for the citizens.

Appendix

A Additional data details

Table 4: Suicide rates for different countries

Country	Age-standardized suicide rates	Income Group	Region
Afghanistan	6.0	lower_income	Asia
Albania	3.7	upper_middle	Europe
Algeria	2.6	lower_middle	Africa
Angola	12.6	lower_middle	Africa
Antigua and Barbuda	0.3	high_income	North America
Argentina	8.1	upper_middle	South America
Armenia	2.7	upper_middle	Europe
Australia	11.3	high_income	Oceania
Austria	10.4	high_income	Europe
Azerbaijan	4.0	upper_middle	Europe
Bahamas	3.4	high_income	North America
Bahrain	7.2	high_income	Asia
Bangladesh	3.9	lower_middle	Asia
Barbados	0.3	high_income	North America
Belarus	16.5	upper_middle	Europe
Belgium	13.9	high_income	Europe
Belize	7.7	upper_middle	North America
Benin	12.7	lower_middle	Africa
Bhutan	5.1	lower_middle	Asia
Bolivia (Plurinational State of)	6.8	lower_middle	South America
Bosnia and Herzegovina	8.3	upper_middle	Europe
Botswana	20.2	upper_middle	Africa
Brazil	6.4	upper_middle	South America
Brunei Darussalam	2.5	high_income	Asia
Bulgaria	6.5	upper_middle	Europe
Burkina Faso	14.4	lower_income	Africa

Burundi	12.1	lower_income	Africa
Cote d'Ivoire	15.7	lower_middle	Africa
Cabo Verde	15.2	lower_middle	Africa
Cambodia	5.5	lower_middle	Asia
Cameroon	15.9	lower_middle	Africa
Canada	10.3	high_income	North America
Central African Republic	23.0	lower_income	Africa
Chad	13.2	lower_income	Africa
Chile	8.0	high_income	South America
China	6.7	upper_middle	Asia
Colombia	3.7	upper_middle	South America
Comoros	8.5	lower_middle	Africa
Congo	11.6	lower_middle	Africa
Costa Rica	7.6	upper_middle	North America
Croatia	11.0	high_income	Europe
Cuba	10.2	upper_middle	North America
Cyprus	3.2	high_income	Europe
Czechia	9.5	high_income	Europe
Democratic People's Republic of Korea	8.2	lower_income	Asia
Democratic Republic of the Congo	12.4	lower_income	Africa
Denmark	7.6	high_income	Europe
Djibouti	11.9	lower_middle	Africa
Dominican Republic	5.1	upper_middle	North America
Ecuador	7.7	upper_middle	South America
Egypt	3.4	lower_middle	Africa
El Salvador	6.1	upper_middle	North America
Equatorial Guinea	13.5	upper_middle	Africa
Eritrea	17.3	lower_income	Africa
Estonia	12.0	high_income	Europe
Eswatini	40.5	lower_middle	Africa
Ethiopia	9.5	lower_income	Africa
Fiji	9.5	upper_middle	Oceania
Finland	13.4	high_income	Europe

France	9.7	high_income	Europe
Gabon	13.1	upper_middle	Africa
Gambia	9.6	lower_income	Africa
Georgia	7.7	upper_middle	Europe
Germany	8.3	high_income	Europe
Ghana	10.5	lower_middle	Africa
Greece	3.6	high_income	Europe
Grenada	0.6	upper_middle	North America
Guatemala	6.2	upper_middle	North America
Guinea	12.3	lower_middle	Africa
Guinea-Bissau	12.4	lower_income	Africa
Guyana	40.9	high_income	South America
Haiti	11.2	lower_middle	North America
Honduras	2.6	lower_middle	North America
Hungary	11.8	high_income	Europe
Iceland	11.2	high_income	Europe
India	12.9	lower_middle	Asia
Indonesia	2.6	upper_middle	Asia
Iran (Islamic Republic of)	5.1	lower_middle	Asia
Iraq	4.7	upper_middle	Asia
Ireland	8.9	high_income	Europe
Israel	5.2	high_income	Asia
Italy	4.3	high_income	Europe
Jamaica	2.3	upper_middle	North America
Japan	12.2	high_income	Asia
Jordan	2.0	lower_middle	Asia
Kazakhstan	18.1	upper_middle	Europe
Kenya	11.0	lower_middle	Africa
Kiribati	30.6	lower_middle	Oceania
Kuwait	2.7	high_income	Asia
Kyrgyzstan	8.3	lower_middle	Asia
Lao People's Democratic Republic	6.0	lower_middle	Asia
Latvia	16.1	high_income	Europe
Lebanon	2.8	lower_middle	Asia
Lesotho	87.5	lower_middle	Africa
Liberia	7.4	lower_income	Africa

Libya	4.5	upper_middle	Africa
Lithuania	20.2	high_income	Europe
Luxembourg	8.6	high_income	Europe
Madagascar	9.2	lower_income	Africa
Malawi	10.6	lower_income	Africa
Malaysia	5.8	upper_middle	Asia
Maldives	2.8	upper_middle	Asia
Mali	8.0	lower_income	Africa
Malta	5.3	high_income	Europe
Mauritania	5.5	lower_middle	Africa
Mauritius	8.8	upper_middle	Africa
Mexico	5.3	upper_middle	North America
Micronesia (Federated States of)	29.0	lower_middle	Oceania
Mongolia	18.0	lower_middle	Asia
Montenegro	16.2	upper_middle	Europe
Morocco	7.3	lower_middle	Africa
Mozambique	23.2	lower_income	Africa
Myanmar	3.0	lower_middle	Asia
Namibia	13.5	upper_middle	Africa
Nepal	9.8	lower_middle	Asia
Netherlands (Kingdom of the)	9.3	high_income	Europe
New Zealand	10.3	high_income	Oceania
Nicaragua	4.7	lower_middle	North America
Niger	10.1	lower_income	Africa
Nigeria	6.9	lower_middle	Africa
North Macedonia	7.2	upper_middle	Europe
Norway	9.9	high_income	Europe
Oman	4.5	high_income	Asia
Pakistan	9.8	lower_middle	Asia
Panama	2.9	high_income	North America
Papua New Guinea	3.6	lower_middle	Oceania
Paraguay	6.2	upper_middle	South America
Peru	2.7	upper_middle	South America
Philippines	2.5	lower_middle	Asia
Poland	9.3	high_income	Europe
Portugal	7.2	high_income	Europe

Qatar	4.7	high_income	Asia
Republic of Korea	21.2	high_income	Asia
Republic of Moldova	12.2	upper_middle	Europe
Romania	7.3	high_income	Europe
Russian Federation	21.6	upper_middle	Asia
Rwanda	9.5	lower_income	Africa
Saint Lucia	6.9	upper_middle	North America
Saint Vincent and the Grenadines	1.0	upper_middle	North America
Samoa	14.6	lower_middle	Oceania
Sao Tome and Principe	2.2	lower_middle	Africa
Saudi Arabia	5.4	high_income	Asia
Senegal	11.0	lower_middle	Africa
Serbia	7.9	upper_middle	Europe
Seychelles	7.7	high_income	Africa
Sierra Leone	11.3	lower_income	Africa
Singapore	9.7	high_income	Asia
Slovakia	9.3	high_income	Europe
Slovenia	14.0	high_income	Europe
Solomon Islands	17.4	lower_middle	Oceania
Somalia	14.7	lower_income	Africa
South Africa	23.5	upper_middle	Africa
South Sudan	6.7	lower_income	Africa
Spain	5.3	high_income	Europe
Sri Lanka	12.9	lower_middle	Asia
Sudan	4.8	lower_income	Africa
Suriname	25.9	upper_middle	South America
Sweden	12.4	high_income	Europe
Switzerland	9.8	high_income	Europe
Syrian Arab Republic	2.1	lower_income	Asia
Turkiye	2.3	upper_middle	Asia
Tajikistan	5.3	lower_middle	Asia
Thailand	8.0	upper_middle	Asia
Timor-Leste	4.5	lower_middle	Asia

Togo	14.8	lower_income	Africa
Tonga	4.4	upper_middle	Oceania
Trinidad and Tobago	8.3	high_income	North America
Tunisia	3.2	lower_middle	Africa
Turkmenistan	6.1	upper_middle	Asia
Uganda	10.4	lower_income	Africa
Ukraine	17.7	lower_middle	Europe
United Arab Emirates	5.2	high_income	Asia
United Kingdom of Great Britain and Northern Ireland	6.9	high_income	Europe
United Republic of Tanzania	8.2	lower_middle	Africa
United States of America	14.5	high_income	North America
Uruguay	18.8	high_income	South America
Uzbekistan	8.3	lower_middle	Asia
Vanuatu	21.0	lower_middle	Oceania
Venezuela (Bolivarian Republic of)	2.1	upper_middle	South America
Viet Nam	7.2	lower_middle	Asia
Yemen	7.1	lower_income	Asia
Zambia	14.4	lower_middle	Africa
Zimbabwe	23.6	lower_middle	Africa

B Model details

B.1 Credibility interval

Figure 4 is a visualization of the estimated effect of each predictor on the outcome variable. The credibility interval we have created is the range in which we can say the true value of the coefficient lies with 95% credibility.

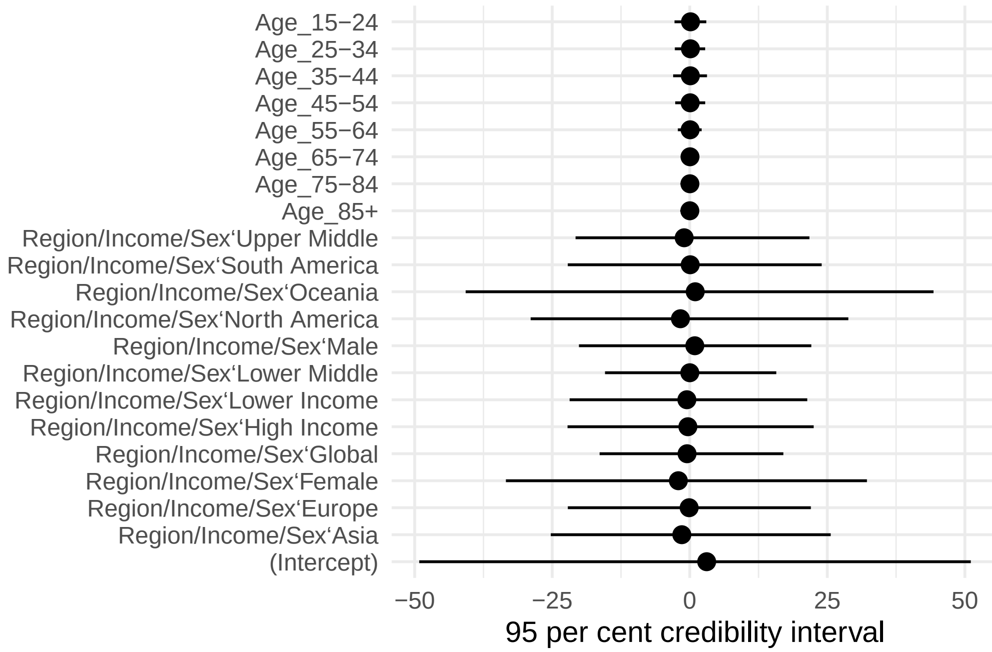

Figure 4: Credibility intervals of the predictors

B.2 Posterior predictive check

In Figure 5 we implement a posterior predictive check. This show the comparison between the simulated data to the actual observed data to assess whether the model is adequate. This shows the comparison for the poisson model and we can assess how well each model fits the observed data. Based on the check we can conclude that the multiple linear model fits the observed data as the trace of the Y plot follows the trends of Yrep.

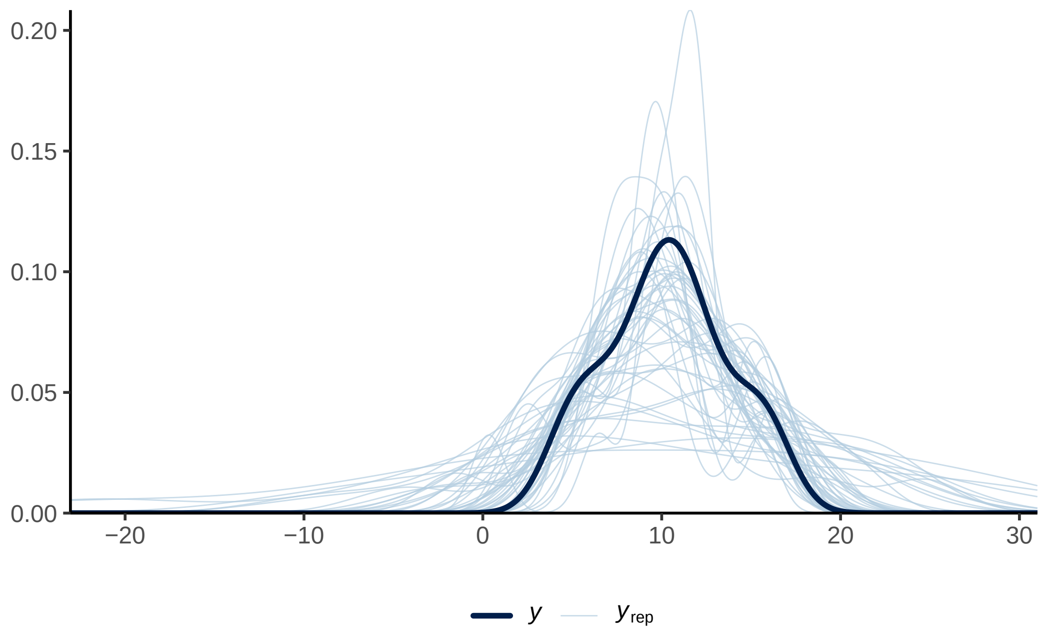

Figure 5: Examining how the model fits, and is affected by, the data

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Bank, World. 2024. "World Bank Country and Lending Groups – World Bank Data Help Desk." https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.

Fund, International Monetary. 2021. "Unemployment Rate Data." https://www.imf.org/external/datamapper/LUR@WEO/OEMDC/ADVEC/WEOWORLD.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2020. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm.

Kay, Matthew. 2023. *tidybayes: Tidy Data and Geoms for Bayesian Models*. https://doi.org/10.5281/zenodo.1308151.

Organization, World Health. 2021a. "GHO | by Category | Suicide Rate Estimates, Age-Standardized - Estimates by Country."

———. 2021b. "GHO | by Category | Suicide Rate Estimates, Crude, 10-Year Age Groups - Estimates by Country." https://doi.org/MINERVA URL TO PAGE.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow.

2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

World Health Organization. 2021. "GHO | by Category | Mental Health." https://doi.org/MINERVA%20URL%20TO%20PAGE.

Worldpopulationreview.com. 2024. "List of Countries by Continent 2024." https://worldpopulationreview.com/country-rankings/list-of-countries-by-continent.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.