The Characteristics of People Who Commit Suicide.

Cara You

Math 4840-904
Independent Study
Dr. French
05/08/2019

Introduction

Background and motivation

From the World Health Organization (WHO) data, on average, 800 000 people die due to suicide every year. Suicide is the second leading cause of death among 15–29-year-olds. Suicide does not just occur in high-income countries but is a global phenomenon in all regions of the world. Yet, only 38 countries report having a national strategy for suicide prevention. The Eastern Europe and East Asia regions have the highest suicide rate worldwide. The region with the lowest suicide rate is the Caribbean, followed by the Middle East (WHO).

This paper aims at discovering the important candidates, and measuring their impacts for suicide count in order to make suggestions on suicide prevention in the future. There is a strong correlation between suicide rates and Gross domestic product (GPD) per capita (Fountoulakis ,2014). In fact, in 2016, over 79% of global suicides occurred in low- and middle-income countries. Furthermore, males were between 1.7 and 2.1 times more likely to commit suicide than females in 2015 (Värnik, P, March 2012). The WHO states that suicide rates have increased by 60% globally in the last 45 years (2018). Based on this knowledge, we predict that higher GDP is associated with lower suicide count. We predict that males will commit suicide of a rate double that of females. We predict that there is an upward trend for suicide rate over time, meaning that for every additional year, suicide rate increases.

To study our research question, we use a database called Suicide Rates Overview 1985 to 2016 from the World Health Organization (2018), obtained from Kaggle (Rusty 2018, December 01). The data compares socio-economic information with suicide rates by year and country. The compiled set of data was built to find signals correlated with increased suicide rates among different cohorts globally, across the socio-economic spectrum, with a goal of inspiring suicide prevention.

There are 12 variables in the raw data: country, year, sex, age, suicide_no, population, suicides/ 100k pop, county-year, HDI for year, gdp_for_year ($), gdp_per_capita ($), and generation. The variable **country** indicates which country the suicide count represents. The variable **year** (ranging from 1985 to 2016) indicates which year the suicide count represents. The variable **sex** (sexmale or sexfemale) indicates the suicide count for males or females. A**ge** (including 15-24 years, 25-34 years, 35-54 years, 55-74 years, and 75+ years) indicates the age group the suicide count represents. **Suicide_no** indicates the number of suicides in a country in a

particular year within a typical age group. **Population** is the population of the age group and sex group of a country in a year. **Suicides/ 100k pop** is the standardized suicide rate in terms of number of suicides per 100,000 people in that year. **Country-year** is the combination of country and year in a format like "Albania1985". **HDI for year** is the Human Development Index for the year. It is a statistical composite index of life expectancy, education, and per capita income indicators. **Gdp_for_year** indicates the GDP in US dollars of the country for the year in which the suicide count represents. **Generation** (including G.I. Generation, Silent, Boomers, Generation X, millennials, and Generation Z) is based on age group. The "Greatest Generation" (or GI Generation) who were born in 1924 or earlier, lived through the Great Depression and then fought in World War II. The "silent generation" are those born from 1925 to 1945 – so called because they were raised during a period of war and economic depression, thus more cautious than their parents. The "baby boomers" came next from 1945 to 1964, the result of an increase in births following the end of World War II. After the baby boomers came "Generation X", from around 1965 to 1976. The label reflected the counterculture of a rebellious generation, distrustful of the establishment and keen to find their own voice. The cohort known as millennials were born around the millennium, from the early 1980s to the mid-1990s or early 2000s. The group experiences a variety of unique social and economic conditions. "Generation Z" is the current name for the cohort born from the mid-1990s to mid-2000s, who has grown up in a hyper connected world.

Data Exploration

Even before fitting a model to the data, we can plot the response against the predictors to explore the data. We begin by constructing a box plot of suicide rate per 100k people by sex in Figure 1.

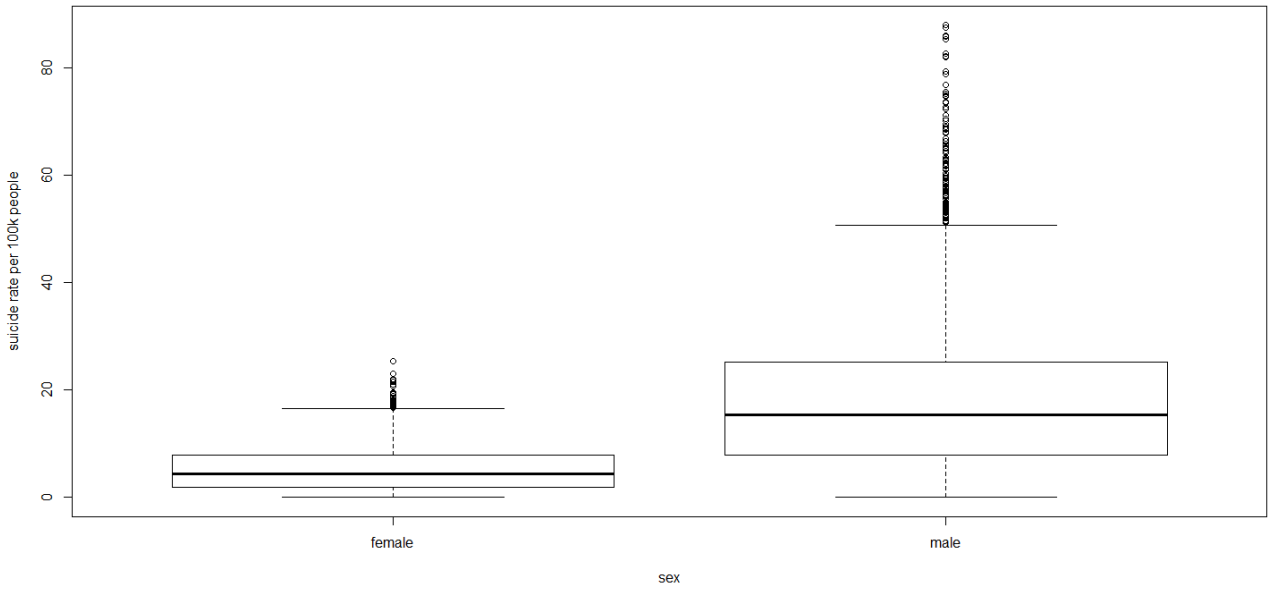


Figure 1: box plots of the suicide rate per 100k people stratified by sex

From the graph, males typically have larger values of suicide rate than females. The box plots for males and females have longer upper whiskers than lower whiskers, meaning that both sexes have suicide rates more varied among the most positive quartile group, and both sexes have suicide rates that are very similar for the least positive quartile group.

Next, we construct box plots for suicide rate per 100k people by age and generation in Figure 2 and 3.

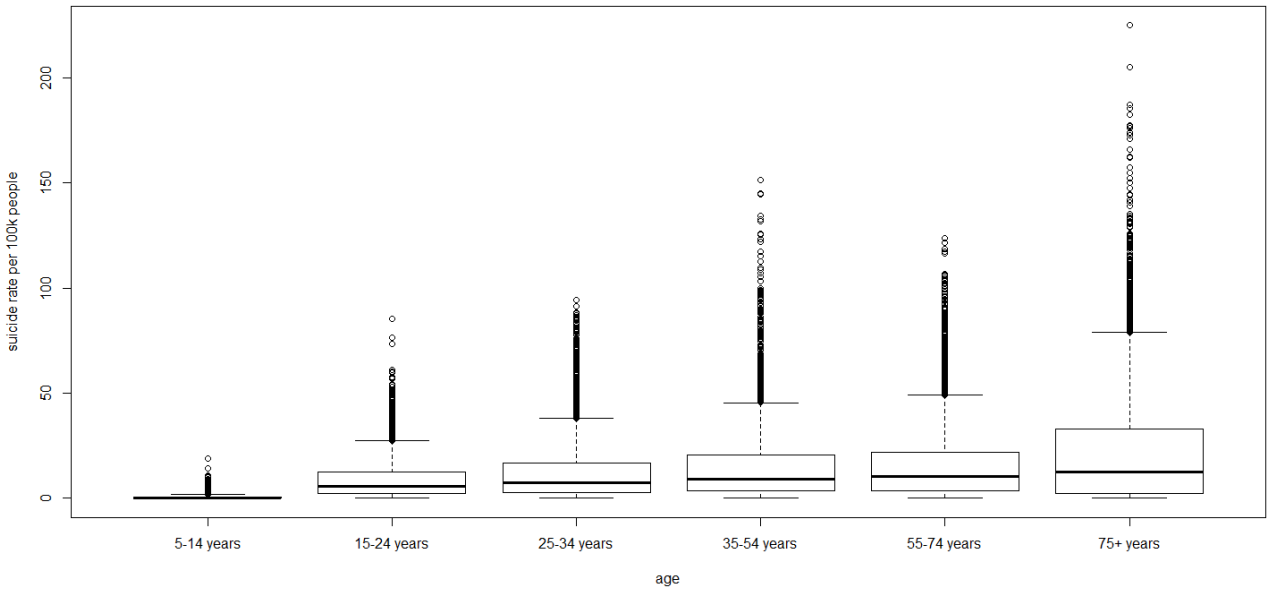


Figure 2: box plots of the suicide rate per 100k people stratified by age

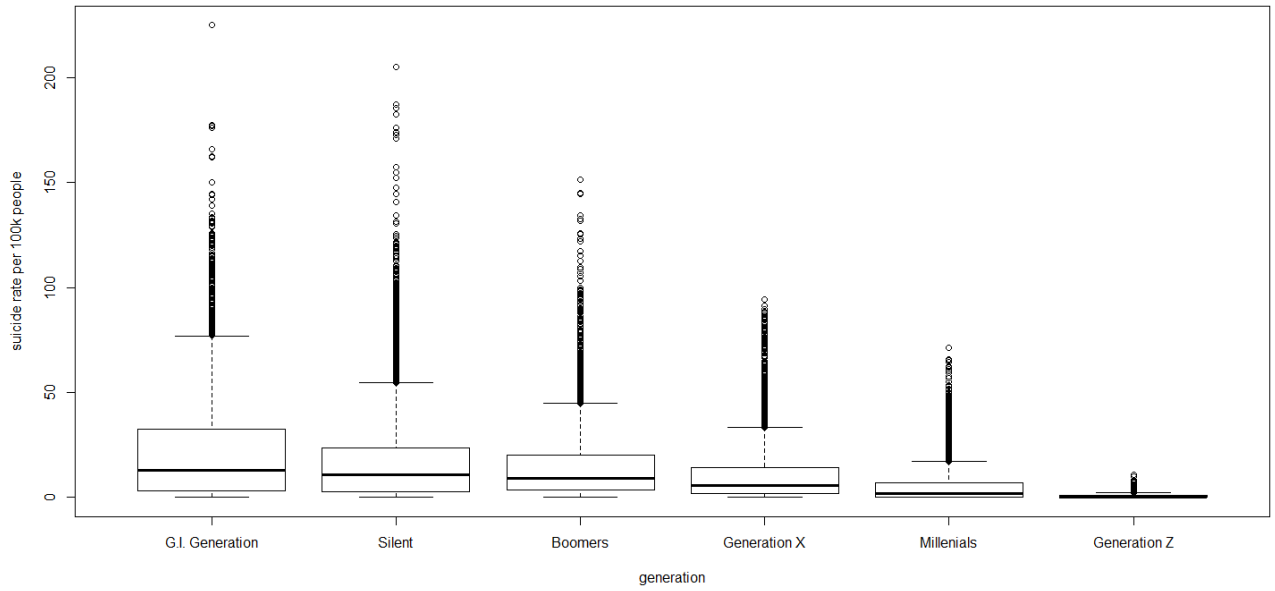


Figure 3: box plots of the suicide rate per 100k people stratified by generation

From the age group plot in Figure 2, in general, higher age is associated with higher suicide rates. People older than 75 have the most varied suicide rate in the second to third quartile and in the upper whisker among the other age groups. They also have the highest amount of extremely large suicide rates compared to the other age groups. Compared to younger age groups, this group is possibly suffering from more physical illness and psychological obstacles, giving them greater motivation to commit suicide. The second most varied suicide rate age group is the 35 to 54 years group. Compared to younger age groups, this group is probably facing pressure from both family and career. 5 to 14 years kids have similar low suicide rate, which makes sense because they have less access to and awareness of suicide.

Considering the generation plot in Figure 3, in general, older generations are associated with larger suicide rates among countries across years. The GI Generation have the most varied suicide rate in the second to third quartile and in the upper whisker among the other generation groups. The Silent Generation have the most amount of extremely large suicide rate compared to the other generation groups.

We then plot time series of suicide rate per 100 thousand people for various countries, and here we only show four as an example. In Figure 4 we see time series plots of suicide rate per 100k people for four countries: Japan, United States, Finland, and Brazil.

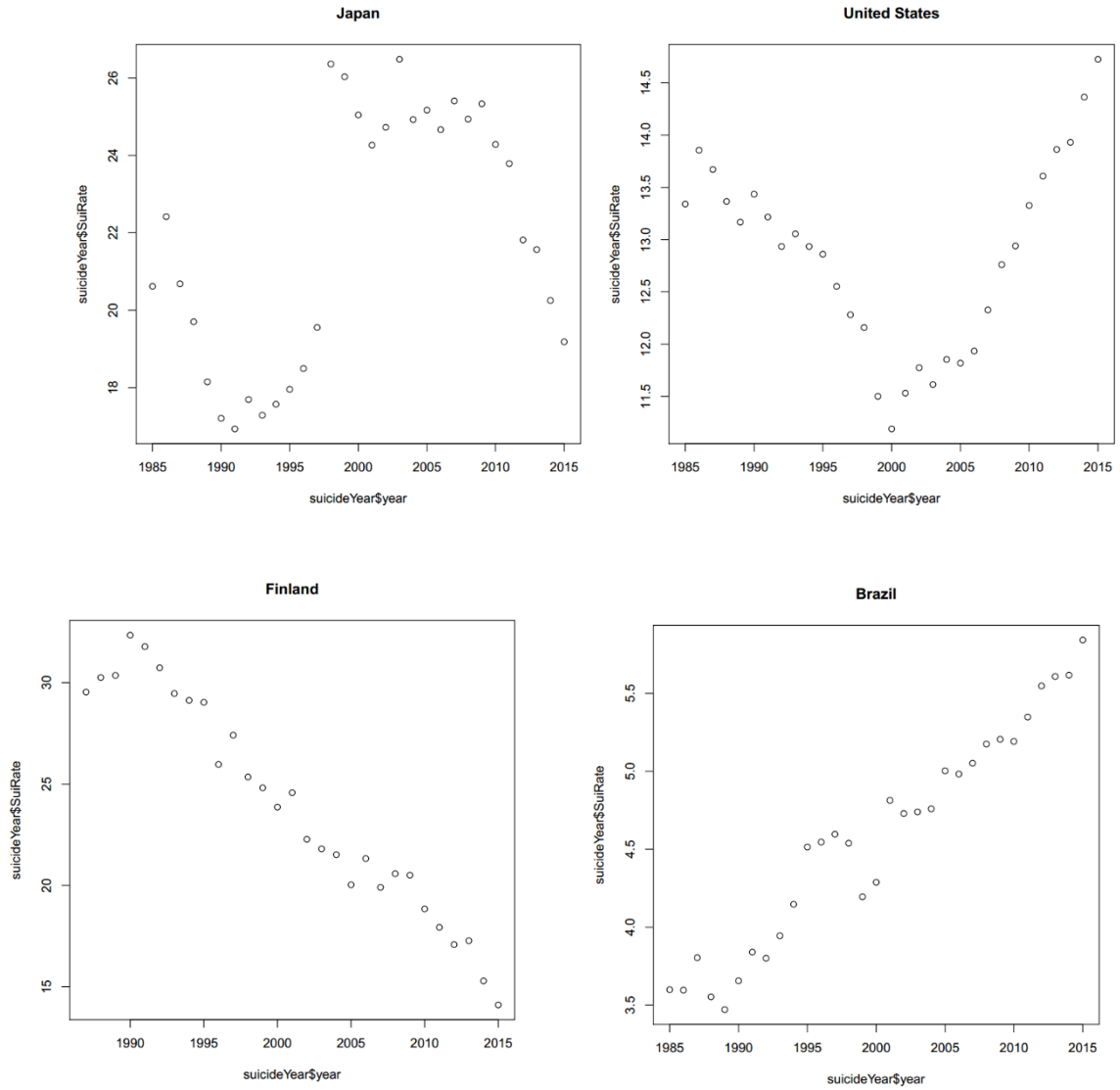


Figure 4: Time series plots of suicide rate per 100k people for four countries

Argentina, Brazil, Greece, Ireland, Philippine, Poland, Portugal, Spain, Thailand, Ukraine, and Uruguay have an upward trend. Austria, Denmark, Estonia, Finland, France, Hungary, Kiribati, Luxembourg, Norway, Sweden, and Switzerland have a downward trend. Cuba, Germany, Netherlands, United Kingdom, and the United States have a downward then upward trend. Belarus, Bulgaria, Czech Republic, Italy, Japan, Kazakhstan, Romania, and the Russian Federation have an upward then downward trend. Since different countries have different patterns of suicide rate over years, altogether, the suicide count may not have a strong relationship with GDP.

Next, we plot suicide rate per 100k people stratified by GDP for the year and log (GDP) for the year in Figure 5.

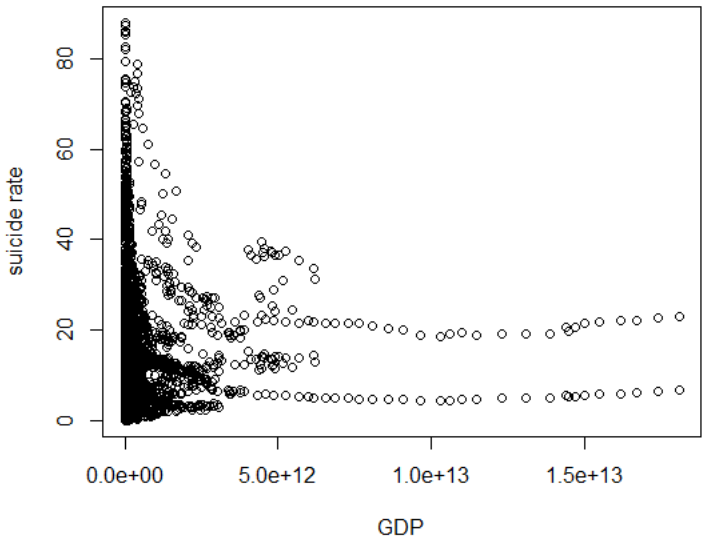


Figure 5: suicide rate per 100k people stratified by GDP

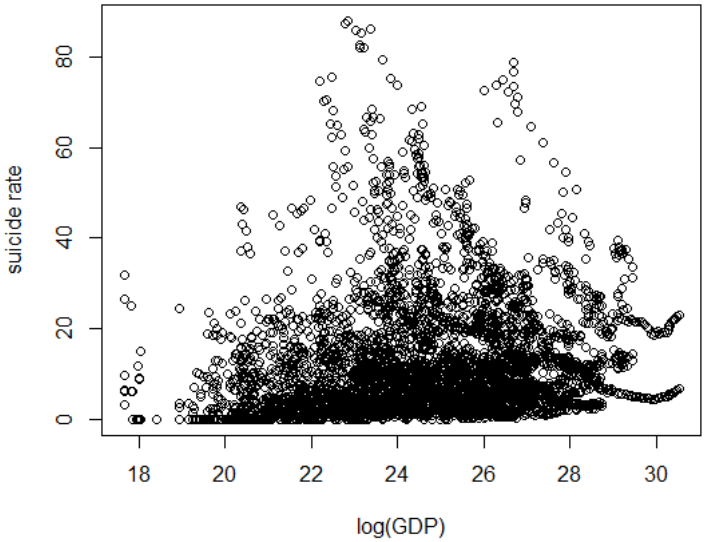


Figure 5: suicide rate per 100k people stratified by log(GDP)

We see that the scale of the GDP variable is not friendly to display. We consider the log transformation of GDP for the predictor. We now see perhaps a weak linear relationship between suicide rate and log (GDP) suggesting that no further transformation of GDP is necessary.

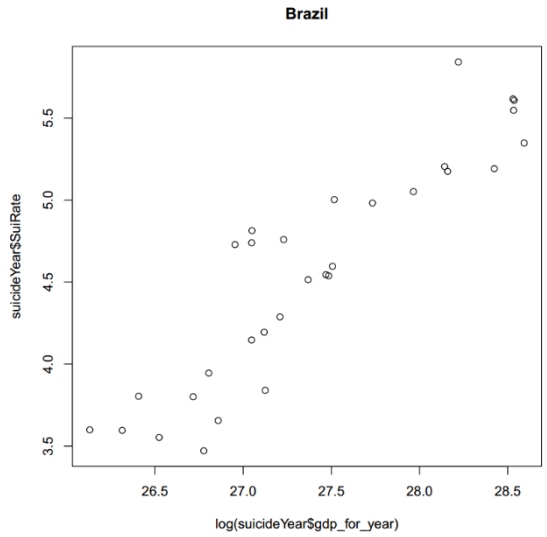


Figure 6: suicide rate per year versus log (GDP) for Brazil

In Figure 6, we examine the suicide rate per year versus log (GDP) for Brazil.

Brazil is used as an example to demonstrate the similarity between suicide rate versus year and suicide rate versus GDP for different countries. For most countries, the time series for log (GDP) is similar to their time series for country. This is probably due to the relatively steady increase in GDP in most countries.

Methods and Results

GLM for Poisson data

A typical linear model is not appropriate for the suicide data for a couple of reasons. Firstly, the suicide data are count data; therefore, the response variable is not continuous. Secondly, the variance of the suicide data is unlikely to be constant, as count data typically has larger variability for larger counts.

Generalized linear models (GLM) cover both situations by allowing for response variables that have non-normal distributions rather than simply normal distributions, and for a function of the response variable, the link function, to vary linearly with the predicted values. The Poisson distribution is typical for the count of occurrences in a fixed amount of time; therefore, the generalized linear model for Poisson data is suitable for the suicide data.

There are three components of a GLM: the random component, the linear predictor, and the link function. The random component consists of a response variable y with independent observations $y_1, y_2, \dots, y_n$. The random component must be a member of an exponential family, which includes the Poisson distribution. The link function transforms the mean of the response and connects it to the linear predictors. Let η_i denote the linear combination of predictors: $\eta_i = \sum_{j=1}^{p} \beta_j x_{ij}$. Let μ_i denote the mean of y_i and g the link function. Then $g(\mu_i) = \eta_i$. For instance, a log linear model using the log-link function is often used for a Poisson response, $\log(\mu_i) = \log(n_i \lambda_i) = \eta_i = \sum_{j=1}^{p} \beta_j x_{ij}$, where η_i denotes the population of region i and λ_i is the suicide rate of observation i. Log (n_i) is just a known constant additive term to the linear predictor. In our case, we connect the suicide rate, $\frac{suicied\ count}{population}$, to the linear predictors: $\log\left(\frac{E(suicide\ count)}{population}\right) = X\beta$, so, $\log(E(suicide\ count)) = \log(population) + X\beta$. In order to get the suicide rate, we adjust for the population size by adding the term offset(log(population)) in the implementation regression in R.

Assumptions for the suicide data from the Poisson distribution

GLM makes several assumptions. The data must be independent. The response must be a member of an exponential family distribution with mean, μ_i. There should be no existence of high leverage outliers to influence the fit of the model to the data. There should be large sample approximations for GLM to do maximum

likelihood estimation (MLE) instead of ordinary least square (OLS) to estimate the parameters. The suicide count data is from a Poisson distribution, which belongs to the exponential family of distributions. For the rest of the assumptions we need to check them in the following sessions one by one.

Assessing collinearity and performing variable selection

Initially there were 12 variables in the raw data: country, year, sex, age, suicide_no, population, suicides/ 100k pop, county-year, HDI for year, GDP ($), GDP_per_capita ($), and generation. Since we miss the data of **HDI** for 2/3 of the countries, it is not suitable to be included in a model aimed at prediction for all countries. We are examining the suicide count using the Poisson distribution; therefore, **suicides/ 100k pop** (per 100 thousand population) is not useful and is dropped from the model. As **suicide_no** is from multiple age groups and different sex groups, we also process the data by grouping all age group and both sex within the same year together in order to get the sum of the suicide count of a country in a year. Since we decided to include both regressors country and year, we drop the country-year variable. We are interested in suicide count, which is related to the size of the population; therefore, we keep the regressor **GDP** ($) instead of **GDP_per_capita** ($). Generation is literally similar to age and we drop both of them from the model because they are categorical regressors. But we still use them when exploring our data using boxplots in the introduction. There are 5 variables remaining: country, year, sex, population (as offset term in the model), and **GDP**. The magnitude for **GDP** is very large; therefore, a log transformation is used for the regressor **GDP**. We use an interaction variable country*year in order to examine the additional year effect on each country. Notice that the variable country*year automatically includes two variables: country and year with respect to model hierarchy. In this paper log represents the natural logarithm.

After selecting a subset of the available variables to begin analysis, multicollinearity checks were necessary to prevent serious problems in the model fit and interpretation of the results. We dropped 8 countries that have perfect multicollinearity between the country*year variable and year variable: Bosnia and Herzegovina, Cabo Ver, Dominica, Macau, United Arab Emirates, Saint Kitts and Nevis, Saint Vincent and Grenadines, and Mongolia. They all have less than 10 years' data for suicide count, meaning that there may be too little variation in year for the country so that year can be linearly predicted from the predictor country*year with a substantial degree of accuracy. To check for collinearity amongst potential regressors for the model, variance inflation factors are usually estimated and compared for numerical regressors. Since we only include

two numeric regressors: year, and log (GDP), we check the correlation between them instead. The correlation between log (GDP) and year is 0.2047205, which is not too big. Thus, there is no serious collinearity problem between GDP and year.

Akaike's Information Criterion (AIC) is an information-based criterion for variable selection. The model with the lowest AIC is favored, which indicates the model with variables including sex, country * year, and log (GDP). We also try to figure out the simplest model using BIC, which give us the exact same result as we get from AIC criteria. The results of the AIC and BIC are in the appendix.

Since deviance measures how closely a model's predictions are to the observed outcomes, one might consider using it as the basis for a goodness of fit test of a given model. Comparing nested models' deviances is equivalent to a profile log-likelihood ratio test of the hypothesis that the extra parameters in the more complex model are all zero. We can test for the significance of the regression coefficients by comparing the full model with sex, country, year, country * year, log (GDP), to different null models using the Analysis of Deviance (ANOD) method. We compute and compare analysis of deviance tables for the full model to the reduced model with log (GDP); the reduced model with sex and country; the reduced model with sex and year; and the reduced model with sex, country, and year. The results of the ANOD test are in the appendix. For models with known Poisson fits, we use the ANOD test with chi-squared test. Because all of the residual deviances of the reduced models are substantially larger than the full model, the full model is preferred. We reject the null hypothesis that country and year are useless in predicting the suicide count and conclude that the regression coefficients of country, year, and country*year are all significantly different from 0. Together with the results from AIC and BIC, we feel confident that the model with sex, country * year, log (GDP), and offset(log(population) is the best model.

$$\log\big(E(SuiCount)\big) = \widehat{\beta_0} + \widehat{\beta_1} * male + \widehat{\beta_2} * year + \hat{\beta}_{country} * country + \widehat{\beta_4}\log(gdp\ for\ year) + \hat{\beta}_{country,year} * country * year + offset(log(population))$$

Checking model structure

We now examine how well the model structure affects the observed data using several techniques. Marginal Model Plot consists of the plot of y vs x_j for each quantitative, non-interactive regressor; it consists of a plot of the fitted values for each observation of regressor j, which is the red model line in the plot; it consists

of a weighted average of nearby responses y versus x_j. It compares the marginal relationship between the response and each regressor while ignoring the other regressors in the model.

Added-Variable Plots display the residuals of y from the model with all regressors except x_j versus the residuals of x_j regressed on all regressors except x_j. It helps to isolate the impact of regressor x_j on the response y after accounting for the effect of the other regressors in the model.

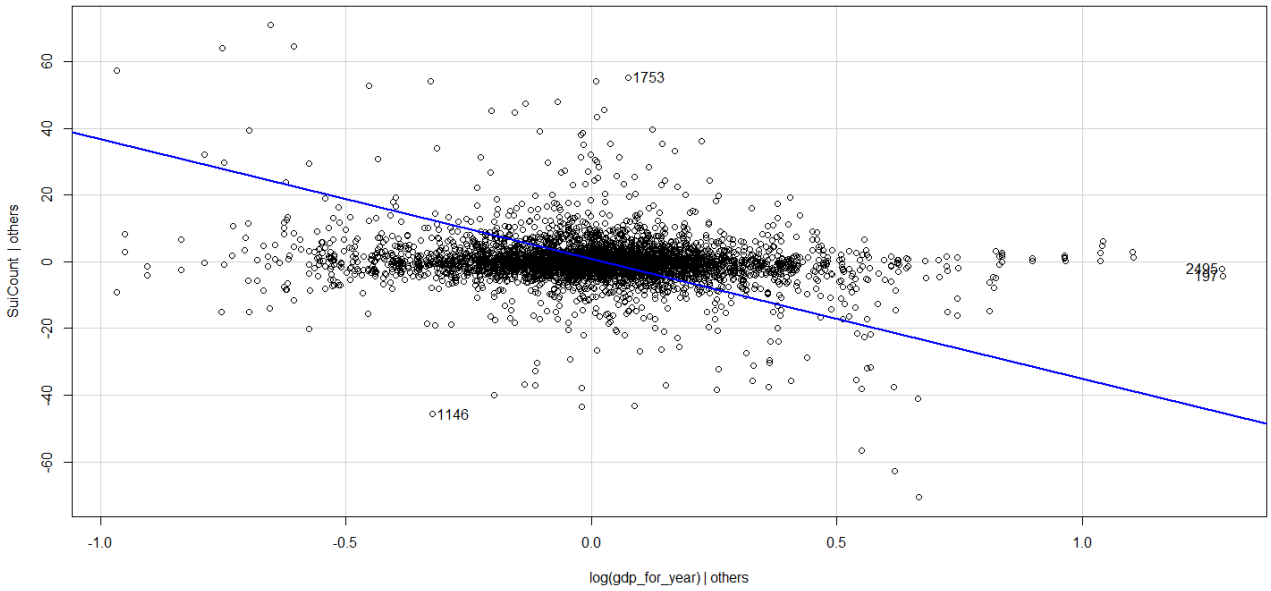


Figure 8: Added-Variables Plot of suicide rate by log (GDP)

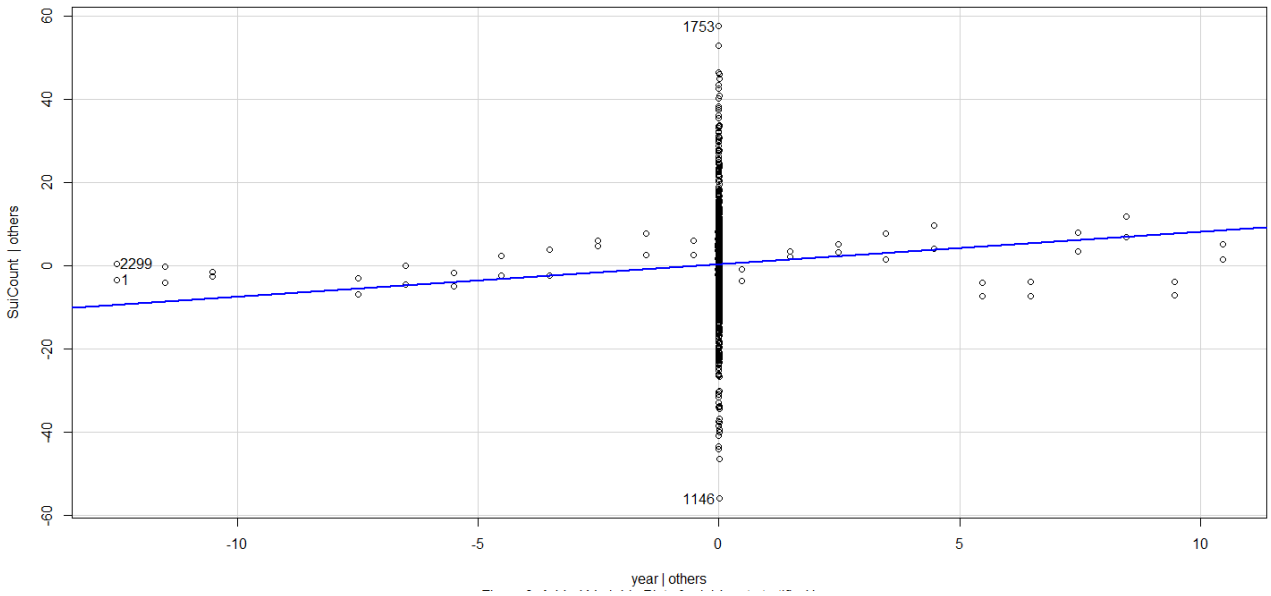


Figure 9: Added-Variable Plot of suicide rate by year

The added variable plots represent the impact of a regressor on the response that is not explained by the other regressors. The added variable plot for log (GDP) indicates that there is a weak negative relationship between suicide count and log (GDP). The added variable plot for the variable year has a flat band of points, meaning that there is a weak relationship between suicide count and year, after accounting for the other regressors. There is no curve in the points or dramatic change in the structure of the points that would indicate a nonlinear problem with the structural component of the model from either plot.

For a linear model, we might consider a transformation of the response, but this is usually impractical for a GLM since it would change the assumed distribution of the response. We might also consider a change to the link function, but often this is undesirable since there are few choices of link function that lead to easily interpretable models. It is best if a change in the choice of predictors or transformations on these predictors can be made since this involves the least disruption to the GLM. But here one of the only two numeric predictors log (GDP) has already been transformed, so no further transformations are made here.

Checking error assumptions & checking influential observations

By the fact that we are making fewer distributional assumptions in Poisson regression, so there is no need to inspect residuals for normality, skewness, or heteroskedasticity. But issues of outliers and influential observations are just as relevant for Poisson regression as they are for linear regression. As with standard linear models, it is important to check the adequacy of the assumptions that support the GLM and the Poisson distribution.

We first check for residuals. Residuals represent the difference between the data and the model and are essential to explore the adequacy of the model. The Pearson residual is comparable to the standardized residuals used for linear models and is defined as: $r_p = \frac{\mathrm{y} - \hat{\mu}}{\sqrt{\mathrm{V}(\hat{\mu})}}$. From the results, "Russian Federation 2009 male" and "Kazakhstan 1993 male" have the largest two absolute Pearson residuals 59.57529 and -54.30594. The model does not fit these two observations very well. For GLMs, we must decide on the appropriate scale for the fitted values using the plot of residuals against fitted values. Usually, it is better to plot the linear predictors $\hat{\eta}$ rather than the predicted responses $\hat{\mu}$ (Faraway, Julian James, 2016).

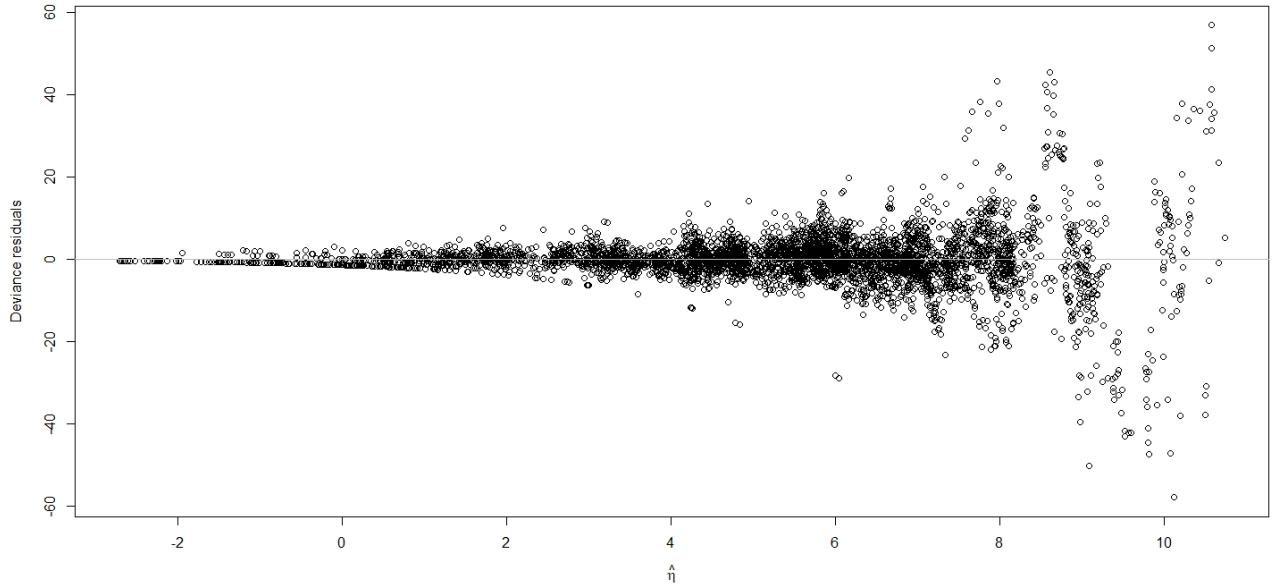


Figure 10: Deviance Residual versus predicted regressors

From Figure 10, there is no obvious non-linear relationship between residuals and fitted values. But the plot has a non-constant thickness, with the variability growing larger as the linear predictor grows larger. Count data are often over-dispersed. Poisson distributed data is intrinsically integer-valued, which makes sense for count data. For a Poisson model, the mean of response equals to its variance. That is, variance increases as the mean increases. As a result, ordinary raw residuals ($r_i = y_i - \widehat{\mu_i}$) should have a spread that increases with fitted values. However, Pearson residuals are residuals subtracted by the mean of the response and then divided by the square root of the variance according to the model ($r_i^P = \frac{y_i - \widehat{\mu_i}}{\sqrt{\widehat{\mu_i}}}$ for a Poisson model). This means that if the model is correct, the Pearson residuals should have constant spread. Thus, we plot the Pearson residual plot below, which looks better than the Deviance residual plot, but it still doesn't have constant variances.

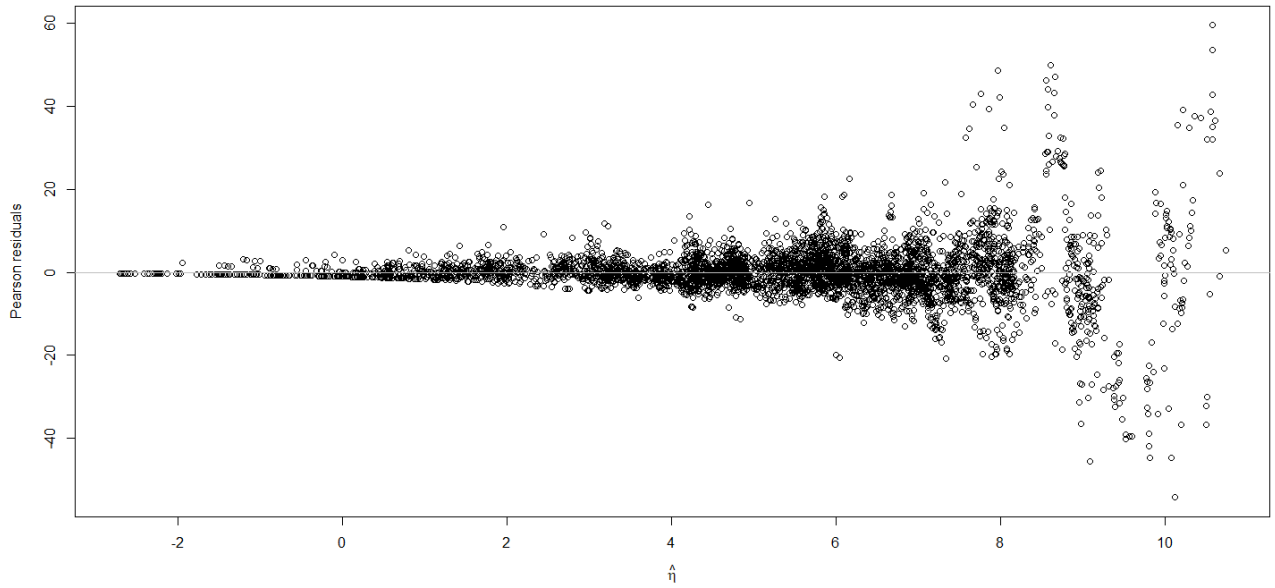


Figure 11: Pearson Residual versus predicted regressors

A violation of this assumption would prompt a change in the model. As discussed in the previous checking model structure section, we might not consider a transformation of the response, nor a change to the link function. It is best if a change in the choice of predictors or transformations on these predictors can be made since this involves the least disruption to the GLM. But here one of the only two numeric predictors log (GDP) has already been transformed, so no further transformation is made here.

We now check for autocorrelation among the counts. We consider a plot of successive residuals in Figure 12.

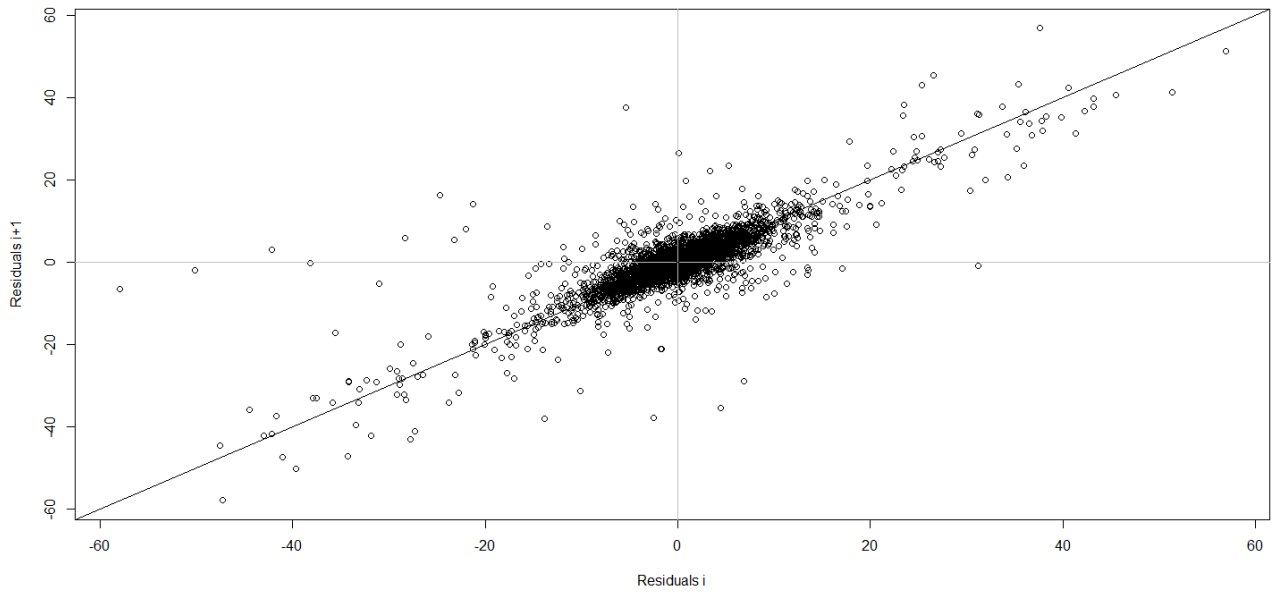


Figure 12: Successive Residual Plot

From Figure 12, there is a positive linear trend, which suggests positive serial correlation between residuals. Application of the Durbin-Watson test to the residuals yields similar results. There is very strong evidence of positive serial autocorrelation in our residuals. Since the suicide count data is time-ordered data, measured yearly from 1985 to 2016 with respect to different countries, it forms a time series. For a time-series, the previous year observation can have an influence on the current observation, which may be the cause of the autocorrelation in our residuals. The results of the Durbin-Watson test are in the appendix.

We now consider identification of influential observations. For a linear model, $\hat{y}$ = Hy, where H = X $(X^T\mathrm{X})^{-1}X^T$ is the hat matrix that projects the data onto the fitted values. The leverages h_i are given by the diagonal of H and represent the potential of the point to influence the fit. They are solely a function of X and whether they are in fact influential will also depend on y. Leverages are different for GLMs. The IRWLS algorithm used to fit the GLM uses weights, w, that are not user-assigned. The weights do affect the leverage. We form a matrix W = diag(w) and the hat matrix is: H=$W^{\frac{1}{2}}\mathrm{X}(X^T\mathrm{WX})^{-1}X^TW^{\frac{1}{2}}$. Studentized residuals are a way to calculate $\hat{y}_{(i)}$, where the (i) means that $\hat{y}_{(i)}$ is estimated without using the ith observation when fitting the model. If $y_i - \hat{y}_i$ is large, then case i is an outlier ((Faraway, Julian James, 2016)).

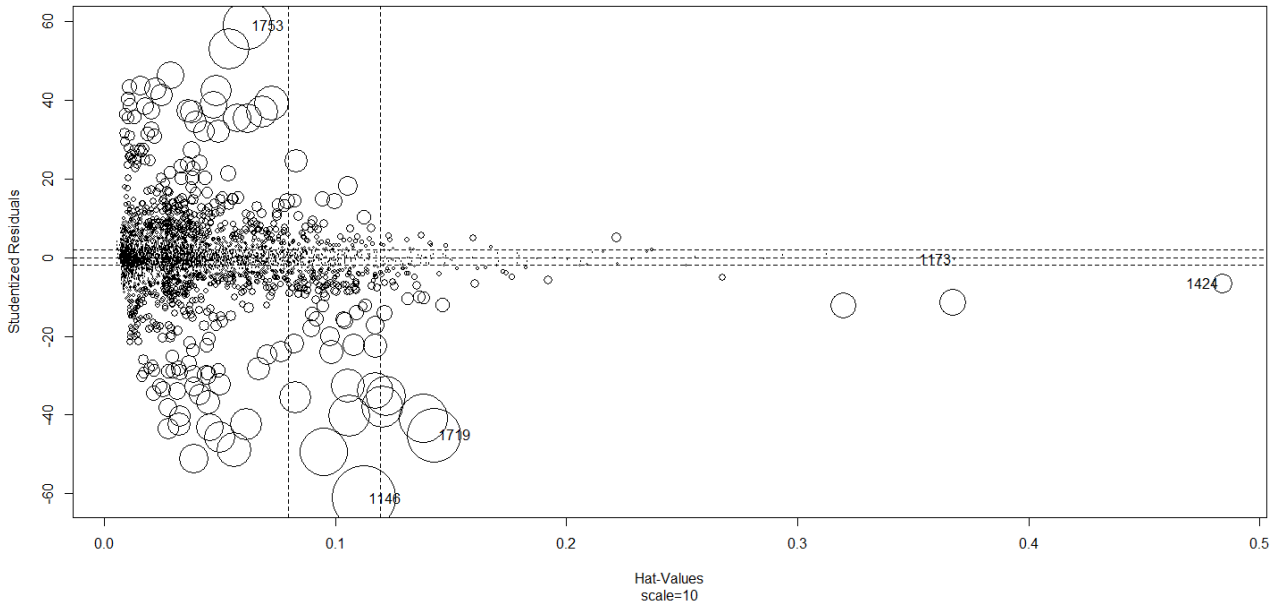


Figure17: Influence Plot

Leverage or studentized residuals alone only measure the potential to affect the fit whereas the combination of the two, as measures of influence more directly assess the effect of each case on the fit. An influential observation is one whose removal from the dataset would cause a large change in the fitted model. An influential observation typically has one of the properties of being an outlier or have a large leverage value. For the purpose of detecting influential observations, it is better to use a half-normal plot that compares the sorted absolute residuals and the quantiles of the half-normal distribution. Studentized residuals test for outliers, while hat-values test for leverage (Faraway, Julian James, 2015). An influence plot compares the studentized residuals to the hat values. From Figure 17, observation 1146, 1719, 1753, 1173 and 1424 have unusually large residuals and hat values, which are "Kazakhstan1993 male", "Romania2003 male", "Russian Federation2009 male", "Kiribati1995 male", and "Netherlands1990 male", respectively. "Netherlands1990 male" has a large hat-value, which means that it has large leverage. Therefore, it has extremely large values of predictors (X). "Kazakhstan1993 male" has large studentized residual which means that its response (Y) does not match the pattern in the rest of the data, and it could be an outlier. All the five observations are candidates of being influential observations. The results of the influence plot statistics are in the appendix.

After dropping the observations with countries Romania, Kazakhstan, Russian Federation, Kiribati, and Netherlands, the coefficient for sex only decrease by 7%; but the coefficient for the regressor year decrease by around 25%, and GDP decreased by around 32%, meaning that these five countries are influential observations. However, as part of the population of interest, these five countries are kept in our model. The results of the change in coefficients are in the appendix.

Then we check for the link function assumption. The link function is a fundamental assumption of the GLM. After eliminating violations of the assumptions that are more easily rectified such as outliers or transformations of the predictors, we can check the link assumption that the link function transforms the mean of the response and connects it to the linear predictors. Because the default Poisson GLM uses a log link which we need to take into account, we plot the linearized response z against the linear predictor $\hat{\eta}$: $\hat{z} = \hat{\eta} + (y - \hat{\mu})\frac{d\hat{\eta}}{d\hat{\mu}}$ (Faraway, Julian James, 2015).

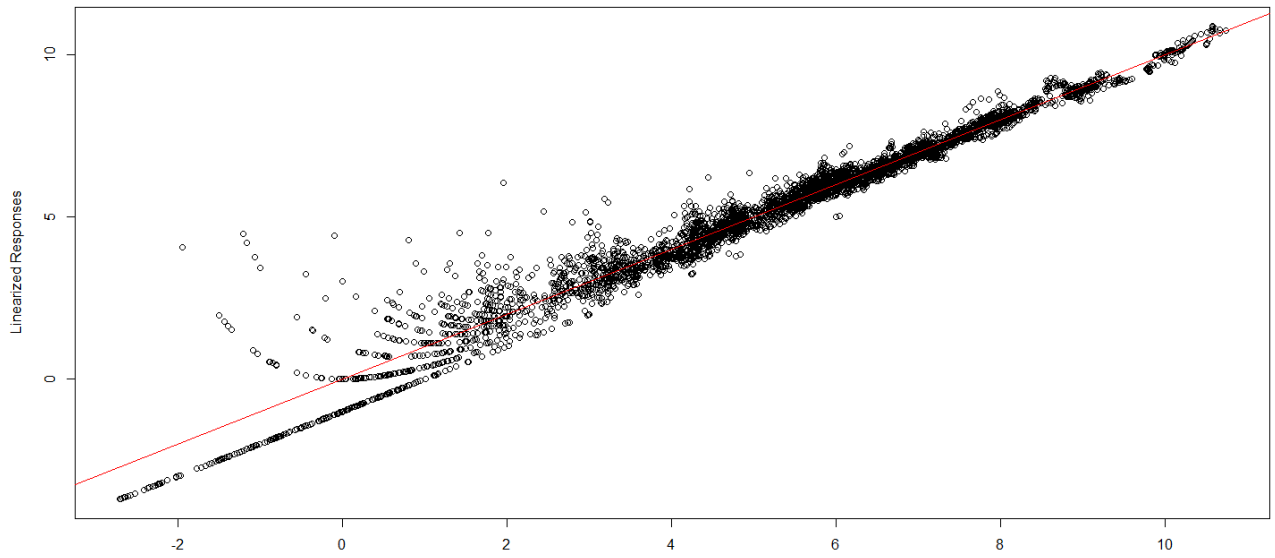


Figure 18: Lenearized Response vs linear predictors

From Figure 18, the transformed response in terms of the log link function is the linearized response. The response transformed by the link function has a better linear relationship with the linear predictor when the linear predictors have larger values. There are some nonlinear patterns between the transformed response and the predictors when linear predictors have small values. This may be an indication that the link function does not fit the data very well with small linear predictors values.

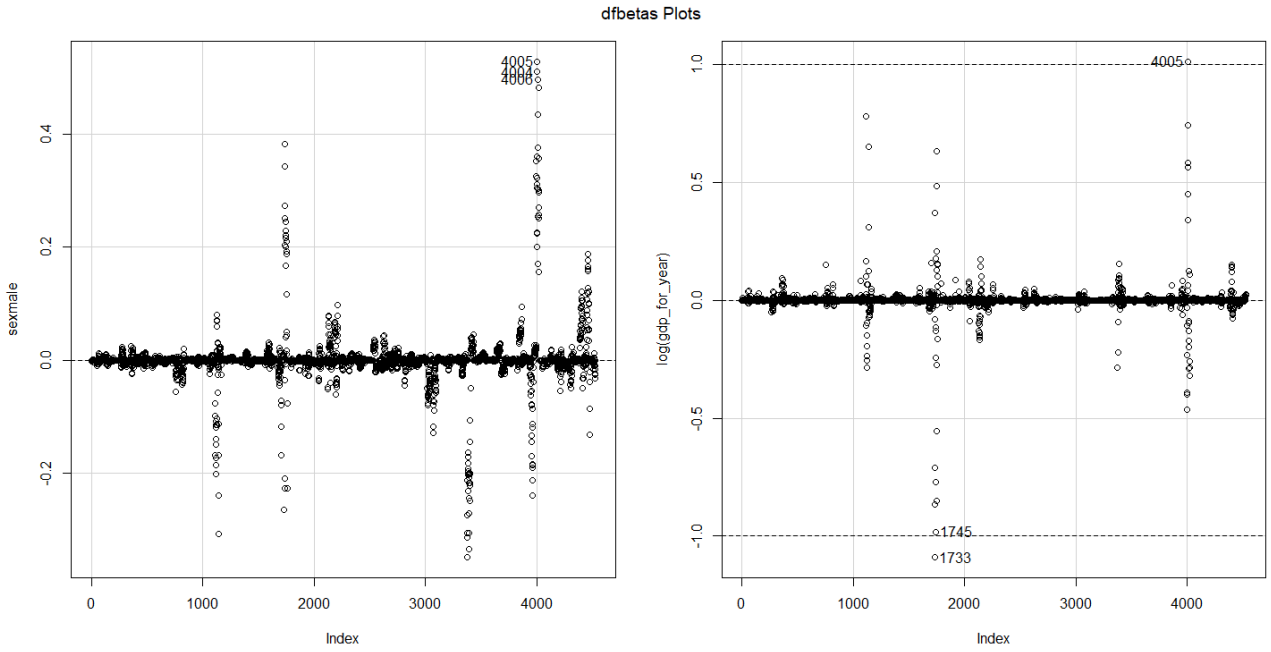
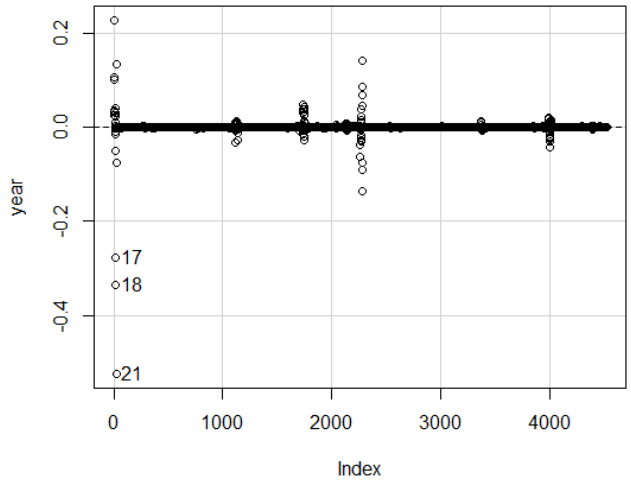


Figure 19: dfbetas plots



The dfbetas plot displays the effect on coefficients of predictors by deleting each observation in turn, standardized by a deleted estimate of the coefficient standard error. It can also give us candidates of influential observations. From Figure 19, observations 4004, 4005, 4006 are influential for the regressor sex, which are Russian Federation1998 female, Russian Federation 1999 female, and Russian Federation 2000 female respectively. Observations 1745,1733, and 4005 are influential for the regressor log (GDO), which are Russian Federation2001 male, Russian Federation1989 male, and Russian Federation1999 female respectively. Observations 17, 18, and 21 are influential for the regressor year. which are Albania2005 male, Albania2006 male, and Albania2009 male respectively.

Interpretation

Now that we have arrived at a final model, we can observe the coefficients for the regressor variables and provide an interpretation of our findings. The complete fitted model is available in the Appendix since the model spans multiple pages due to the large number of countries.

Coefficient	*Estimated Value*	*P-value*
Intercept	-71.99	< 2e-16
Sexmale	1.271	< 2e-16
Year	0.03308	< 2e-16
Log (GDP)	-0.2410	< 2e-16

The basic form of the fitted model is:

$$\log\big(E(SuiCount)\big) = \widehat{\beta_0} + \widehat{\beta_1} * male + \widehat{\beta_2} * year + \hat{\beta}_{country} * country + \widehat{\beta_4}\log(gdp\ for\ year) + \hat{\beta}_{country,year} * country * year + offset(log(population))$$

To make it easier to interpret, we raise both sides of the equation to the power of e: $E(SuiCount) = e^{\widehat{\beta_0}} * e^{\widehat{\beta_1} male} * e^{\widehat{\beta_2} year} * e^{\hat{\beta}_{country}} * e^{\widehat{\beta_4}\log(gdp\ for\ year)} * e^{\hat{\beta}_{country,year} * country * year} * e^{offset(log(population))}$

Therefore, in terms of the regressor, sex, with other factors being constant, we estimate that males have $e^{1.271}$, or 3.5644 times the suicide count of females. Since Albania is the reference level country by default, when we are examining Albania, country equals 0; otherwise, country equals 1. To analyze the regressor year, with other factors being constant, we estimate that for Albania, each additional year is associated with $e^{3.308e-02}$, or 1.0313 times more suicides than the previous year. For other countries, with other factors being constant, each additional year is associated with $e^{\widehat{\beta_2}} * e^{\hat{\beta}_{country,year} * country * year}$ more suicides than the previous year. For each additional year, in comparison with Albania, a country is associated with an extra $e^{\hat{\beta}_{country,year} * country * year}$ multiplicative effect in expected suicide count. For instance, for Norway, each additional year is associated with $e^{3.308e-02} * e^{-3.147e-02}$, or 1.0016 times more suicides than the previous year, meaning an 0.16% increase in suicide count. In comparison to Albania, for Norway, each additional year is associated with $e^{-3.147e-02}$, or

0.9690 times additional effect of change in expected suicide count, meaning that a 3.10% decrease in suicide count related to Albania. When interpreting the regressor log (GDP), with other factors being constant, a 1 unit increase in log (GDP) in a country is associated with a $2.410e-01$, or 0.241 multiplicative change in suicide counts. These interpretations assume we can hold other variables constant, which is not the case.

Since our model contains a categorical interaction variable, country*year, we will interpret Denmark males and United States females as examples, to illustrate the result.

$$E(Denmark_{SuiCount}) = e^{-7.199e+01} * e^{1.271e+00} * e^{(3.308e-02)(2015)} * e^{1.039e+02} * e^{(-2.410e-01)(\log(301000000000))}) * e^{(-5.071e-02)(2015)} * e^{\log(2711902)} \approx e^{6.1000} \approx 446$$

$$E(United\ States_{SuiCount}) = e^{-7.199e+01} * e^{(3.308e-02)(2015)} * e^{4.525e+01} * e^{(-2.410e-01)\log((18120700000000))} * e^{(-2.099e-02)(2015)} * e^{\log(148120000)} \approx e^{9.077617} \approx 8757$$

On average, we estimate that Denmark males in year 2015 have 446 suicides or 16.45 suicides per 10,000 people. On average, we estimate that the United States females in year 2015 have 8,757 counts or 5.91 suicides per 10,000 people.

All the coefficients used in the calculation including intercept, sex, year, log (GDP), countryNorway: year, countryDenmark, countryDenmark: year, countryUnited States, countryUnited States: year are statistically significant. Since we don't have the data of R-square of our model, we compute correlation between response y and the fitted value of response y. The correlation is 0.9863999, which indicates that, in general, our model fits the data very well.

Conclusions

From figure 2, in general, higher age is associated with higher suicide rates worldwide. People older than 75 and people from 35 to 54 years old have the most varied suicide rate in the upper 50% quantile among the other age groups. They also have the most amount of extremely large suicide rate compared to the other age groups. Males have 2.5644 times more suicides than females. A 1% decrease in GDP in a country is associated with 0.00241 more suicides than the previous year. This is a very small scale, which makes sense because higher GDP is assumed to be associated with more developed mental care and medical care system, which might help preventing people from committing suicide. But correlation between suicide rates and GDP per capita "does not

support there being a clear causal relationship between the current economic crisis and an increase in the suicide rate" (Fountoulakis, 2014). Also, since different countries have different patterns of suicide rate over years, altogether, the suicide count may not have a strong relationship with GDP.

The result of the relationship between suicide count and GDP coincides with our prediction that higher GDP is associated with lower suicide rate. This makes sense as a higher GDP country is usually associated with better health care and mental care systems. Suicide results from many sociocultural factors such as unemployment too. This could also explain why GDP is negatively associated with suicide count. The result of the relationship between suicide count and sex coincides with our prediction that males have about twice as large a suicide rate than females. According to the WHO, females tend to have higher rates of reported nonfatal suicidal behaviors, while males have a much higher rate of completed suicides, which demonstrates why males have two times higher suicide count than females. The result of the relationship between suicide count and year coincides with our prediction that suicide count increase by year.

According to the WHO, clinical depression is an especially common cause to suicide. Substance abuse and severe physical disease or infirmity are also recognized as causes. Policy makers should come up with national strategy involving prevention, diagnosis, and treatment on committing suicides. Prevention involves restriction of access to common methods of suicide. What's more, from the WHO, "school-based interventions involving crisis management, self-esteem enhancement and the development of coping skills and healthy decision making have been demonstrated to reduce the risk of suicide among the youth." For people aged 35 to 54 years old, adequate treatment of depression, alcohol and substance abuse can also help reducing suicide rates. For people above 75 years old, more efforts surrounding community-based care as well as educational programs for primary health care providers on the identification and treatment of late-life depression can help lowering suicide rates.

Improvement & Extension

The first improvement that could be made in our model is that we only have 5 predictors: sex, year, GDP, country, and country*year. Among them we only have two numeric predictors: GDP and year. In this case, there may be omitted variable bias, which may influence the accuracy of the coefficients and the variances

of the coefficients. Future model building may consider including more numeric predictors such as Gini Coefficient (wealth distribution among a population), HDI (measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard of living), and Engel's Coefficient (the proportion of income spent on food). One caveat is to use a predictor that is applicable to most or all of the observations; otherwise, the predictor is not representative for the data.

The second improvement that is needed is the autocorrelation problem in our data. Residuals are autocorrelated in time. It is not surprising to see autocorrelation in our model because as a time series, the suicide count is measured yearly, and previous observation can influence current observation in a time series. The proxies such as GDP, sex, and country can only partially predict the suicide count and we would naturally expect some carryover effect from one year to the next. Adding more valid numeric predictors as mentioned in the previous paragraph might help with the autocorrelation problem, but it is hard to be eliminate from a time series data. Moreover, we could also include data on whether a country have a suicide prevention program. This is beneficial when we are trying to analyze the influence of the suicide prevention program.

The third improvement is related to the weak relationship between suicide count and regressors including sex, year, and GDP. Suicide count is subject to macro-economic trends, sociocultural condition and many other factors that are different across countries. Since different countries have different patterns of suicide rate for males and females over years. Altogether, suicide count may not have a strong relationship with GDP globally. Suicide count may not have a strong relationship with sex, nor year globally, either.

The fourth improvement is that we could include age as one of our regressors. From Figure 2, different age groups have different variability of suicide rate. It might be informative to regress suicide count on different age groups too. We did not include age in our model because we don't want to have too much categorical regressors in our model. But we failed to capture the different patterns in different age groups at the same time. Generation is similar to age, in future models we could add either one of them into our model.

The fifth improvement lies in that instead of analyzing on each country, we could group countries into different continents and analyze. This could be beneficial, for different continents may share some similarities in their suicide count. We could also focus on a specific age group, for instance, people who are above 75 years old, instead of analyzing all six age groups together. This can be helpful when we are making a policy for a specific age group.

References

Faraway, Julian James. Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models, Second Edition. Vol. Second edition, Chapman and Hall/CRC, 2016.

Faraway, Julian James. Linear Models with R. Chapman & Hall/CRC, 2015.

Fountoulakis, Konstantinos N., et al. “Relationship of Suicide Rates to Economic Variables in Europe: 2000–2011.” British Journal of Psychiatry, vol. 205, no. 6, 2014, pp. 486–496., doi:10.1192/bjp.bp.114.147454.

“Poisson Regression for Rates.” RPubs, rpubs.com/kaz_yos/poisson.

Prabhakaran, Selva. “Time Series Analysis.” r-Statistics.co, r-statistics.co/Time-Series-Analysis-With-R.html.

“R: Analysis of Deviance for Generalized Linear Model Fits”, stat.ethz.ch/R-manual/R-devel/library/stats/html/anova.glm.html.

“Residuals in Poisson Regression.” Cross Validated, stats.stackexchange.com/questions/99052/residuals-in-poisson-regression.

Rusty. (2018, December 01). Suicide Rates Overview 1985 to 2016. Retrieved from https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016/version/1

United Nations Development Program. (2018). Human development index (HDI). Retrieved from http://hdr.undp.org/en/indicators/137506

Värnik, P (March 2012). "Suicide in the world". International Journal of Environmental Research and Public Health. 9 (3): 760–71. doi:10.3390/ijerph9030760. PMC 3367275. PMID 22690161.

World Bank. (2018). World development indicators: GDP (current US$) by country:1985 to 2016. Retrieved from http://databank.worldbank.org/data/source/world-development-indicators#

[Szamil]. (2017). Suicide in the Twenty-First Century [dataset]. Retrieved from https://www.kaggle.com/szamil/suicide-in-the-twenty-first-century/notebook

WHO (2002). "Self-directed violence" (PDF). www.who.int.

World Health Organization. (2018). Suicide prevention. Retrieved from

http://www.who.int/mental_health/suicide-prevention/en/

“6.1 - Introduction to Generalized Linear Models.” 6.1 - Introduction to Generalized Linear Models | STAT 504, newonlinecourses.science.psu.edu/stat504/node/216/.

Appendix

```
    Deviance Residuals:
    Min       1Q   Median       3Q      Max
-57.997   -2.288   -0.199    2.161   56.911

Coefficients:
                                Estimate Std. Error  z value Pr(>|z|)
(Intercept)                   -7.199e+01  6.891e+00  -10.447  < 2e-16 ***
sexmale                        1.271e+00  9.144e-04 1390.334  < 2e-16 ***
countryAntigua and Barbuda     1.881e+01  6.674e+01    0.282 0.778097
countryArgentina               3.719e+01  6.933e+00    5.364 8.15e-08 ***
countryArmenia                 2.676e+01  9.070e+00    2.951 0.003172 **
countryAruba                   1.007e+02  4.542e+01    2.217 0.026604 *
countryAustralia               4.621e+01  6.938e+00    6.660 2.74e-11 ***
countryAustria                 9.156e+01  6.956e+00   13.163  < 2e-16 ***
countryAzerbaijan              7.134e+01  1.277e+01    5.588 2.29e-08 ***
countryBahamas                 1.958e+01  2.935e+01    0.667 0.504671
countryBahrain                 7.260e+01  1.401e+01    5.181 2.21e-07 ***
countryBarbados                2.123e+02  2.042e+01   10.395  < 2e-16 ***
countryBelarus                 4.765e+01  6.984e+00    6.823 8.94e-12 ***
countryBelgium                 5.948e+01  6.944e+00    8.566  < 2e-16 ***
countryBelize                 -3.155e+01  1.561e+01   -2.021 0.043262 *
countryBrazil                 -1.150e+00  6.904e+00   -0.167 0.867658
countryBulgaria                7.957e+01  7.006e+00   11.358  < 2e-16 ***
countryCanada                  6.135e+01  6.927e+00    8.856  < 2e-16 ***
countryChile                  -3.144e+01  6.983e+00   -4.503 6.71e-06 ***
countryColombia               -5.480e+00  6.960e+00   -0.787 0.431092
countryCosta Rica              7.651e+00  7.486e+00    1.022 0.306810
countryCroatia                 6.753e+01  7.269e+00    9.290  < 2e-16 ***
countryCuba                    9.471e+01  7.037e+00   13.458  < 2e-16 ***
countryCyprus                 -1.584e+02  2.349e+01   -6.745 1.53e-11 ***
countryCzech Republic          5.429e+01  6.995e+00    7.761 8.41e-15 ***
countryDenmark                 1.039e+02  7.351e+00   14.138  < 2e-16 ***
countryEcuador                 9.261e+00  7.076e+00    1.309 0.190612
countryEl Salvador             6.721e+01  7.352e+00    9.142  < 2e-16 ***
countryEstonia                 1.222e+02  7.993e+00   15.284  < 2e-16 ***
countryFiji                    2.463e+01  3.375e+01    0.730 0.465583
countryFinland                 1.026e+02  7.015e+00   14.626  < 2e-16 ***
countryFrance                  7.446e+01  6.900e+00   10.792  < 2e-16 ***
countryGeorgia                 3.407e+01  8.299e+00    4.106 4.03e-05 ***
countryGermany                 8.837e+01  6.908e+00   12.792  < 2e-16 ***
countryGreece                  3.314e+01  7.174e+00    4.620 3.84e-06 ***
countryGrenada                 1.357e+02  3.597e+01    3.771 0.000162 ***
countryGuatemala              -9.542e+00  7.388e+00   -1.291 0.196541
countryGuyana                 -6.741e+01  8.356e+00   -8.068 7.17e-16 ***
countryHungary                 9.023e+01  6.958e+00   12.967  < 2e-16 ***
```

```
countryIceland                  4.782e+01  9.476e+00    5.046 4.51e-07 ***
countryIreland                  7.584e+00  7.183e+00    1.056 0.291022
countryIsrael                   6.095e+01  7.204e+00    8.461  < 2e-16 ***
countryItaly                    6.480e+01  6.915e+00    9.370  < 2e-16 ***
countryJamaica                 -1.317e+02  2.258e+01   -5.834 5.41e-09 ***
countryJapan                    4.144e+01  6.894e+00    6.010 1.86e-09 ***
countryKazakhstan               2.711e+01  6.939e+00    3.907 9.34e-05 ***
countryKiribati                 5.385e+02  1.005e+02    5.360 8.30e-08 ***
countryKuwait                   2.284e+01  1.080e+01    2.115 0.034424 *
countryKyrgyzstan               8.319e+01  7.302e+00   11.392  < 2e-16 ***
countryLatvia                   9.415e+01  7.508e+00   12.541  < 2e-16 ***
countryLithuania                6.620e+01  7.146e+00    9.264  < 2e-16 ***
countryLuxembourg               7.556e+01  8.497e+00    8.892  < 2e-16 ***
countryMaldives                 1.072e+02  1.262e+02    0.849 0.395693
countryMalta                   -3.535e+01  1.188e+01   -2.975 0.002929 **
countryMauritius                6.629e+01  7.720e+00    8.587  < 2e-16 ***
countryMexico                  -1.765e+01  6.922e+00   -2.551 0.010756 *
countryMontenegro              -8.392e+02  4.583e+01  -18.311  < 2e-16 ***
countryNetherlands              4.594e+01  6.954e+00    6.607 3.93e-11 ***
countryNew Zealand              4.989e+01  7.174e+00    6.955 3.53e-12 ***
countryNorway                   6.532e+01  7.110e+00    9.187  < 2e-16 ***
countryPanama                   3.012e+01  7.991e+00    3.769 0.000164 ***
countryParaguay                -4.742e+01  7.925e+00   -5.984 2.18e-09 ***
countryPhilippines             -7.003e+01  7.362e+00   -9.513  < 2e-16 ***
countryPoland                   1.836e+01  6.922e+00    2.652 0.008007 **
countryPortugal                 2.187e+01  7.032e+00    3.111 0.001867 **
countryPuerto Rico              6.453e+01  7.278e+00    8.867  < 2e-16 ***
countryQatar                    2.287e+00  1.996e+01    0.115 0.908764
countryRepublic of Korea       -6.839e+01  6.903e+00   -9.907  < 2e-16 ***
countryRomania                  2.211e+01  6.950e+00    3.181 0.001465 **
countryRussian Federation       7.479e+01  6.891e+00   10.852  < 2e-16 ***
countrySaint Lucia              5.651e+01  1.716e+01    3.293 0.000993 ***
countrySerbia                   5.488e+01  7.314e+00    7.503 6.24e-14 ***
countrySeychelles               4.194e+01  2.463e+01    1.703 0.088588 .
countrySingapore                5.143e+01  7.235e+00    7.108 1.18e-12 ***
countrySlovakia                 6.843e+01  7.382e+00    9.271  < 2e-16 ***
countrySlovenia                 9.257e+01  7.601e+00   12.178  < 2e-16 ***
countrySouth Africa            -4.988e+01  8.049e+00   -6.197 5.76e-10 ***
countrySpain                    3.854e+01  6.923e+00    5.567 2.59e-08 ***
countrySri Lanka                8.356e+01  7.011e+00   11.919  < 2e-16 ***
countrySuriname                -4.508e+01  8.797e+00   -5.124 2.99e-07 ***
countrySweden                   8.063e+01  6.992e+00   11.532  < 2e-16 ***
countrySwitzerland              9.569e+01  7.189e+00   13.310  < 2e-16 ***
countryThailand                 3.609e+01  6.919e+00    5.215 1.84e-07 ***
countryTrinidad and Tobago      2.969e+01  8.046e+00    3.690 0.000224 ***
countryTurkmenistan             7.161e+01  7.412e+00    9.662  < 2e-16 ***
countryUkraine                  6.617e+01  6.904e+00    9.583  < 2e-16 ***
countryUnited Kingdom           5.677e+01  6.914e+00    8.211  < 2e-16 ***
countryUnited States            4.525e+01  6.892e+00    6.565 5.20e-11 ***
countryUruguay                 -4.742e+00  7.177e+00   -0.661 0.508828
countryUzbekistan               5.593e+01  7.048e+00    7.936 2.09e-15 ***
year                            3.308e-02  3.447e-03    9.596  < 2e-16 ***
log(gdp_for_year)              -2.410e-01  1.370e-03 -175.818  < 2e-16 ***
countryAntigua and Barbuda:year -1.048e-02  3.334e-02   -0.314 0.753218
countryArgentina:year          -1.765e-02  3.467e-03   -5.090 3.57e-07 ***
countryArmenia:year            -1.354e-02  4.534e-03   -2.987 0.002817 **
countryAruba:year              -4.993e-02  2.267e-02   -2.203 0.027621 *
countryAustralia:year          -2.182e-02  3.470e-03   -6.289 3.20e-10 ***
countryAustria:year            -4.436e-02  3.479e-03  -12.751  < 2e-16 ***
```

```
countryAzerbaijan:year          -3.601e-02  6.389e-03   -5.637 1.73e-08 ***
countryBahamas:year             -1.015e-02  1.466e-02   -0.693 0.488429
countryBahrain:year             -3.626e-02  6.995e-03   -5.184 2.17e-07 ***
countryBarbados:year            -1.064e-01  1.024e-02  -10.387  < 2e-16 ***
countryBelarus:year             -2.248e-02  3.493e-03   -6.435 1.23e-10 ***
countryBelgium:year             -2.829e-02  3.473e-03   -8.146 3.77e-16 ***
countryBelize:year               1.577e-02  7.789e-03    2.025 0.042849 *
countryBrazil:year               1.385e-03  3.453e-03    0.401 0.688331
countryBulgaria:year            -3.880e-02  3.504e-03  -11.072  < 2e-16 ***
countryCanada:year              -2.933e-02  3.464e-03   -8.466  < 2e-16 ***
countryChile:year                1.660e-02  3.492e-03    4.754 2.00e-06 ***
countryColombia:year             3.322e-03  3.481e-03    0.954 0.339877
countryCosta Rica:year          -3.308e-03  3.743e-03   -0.884 0.376869
countryCroatia:year             -3.257e-02  3.635e-03   -8.961  < 2e-16 ***
countryCuba:year                -4.624e-02  3.519e-03  -13.139  < 2e-16 ***
countryCyprus:year               7.893e-02  1.169e-02    6.753 1.45e-11 ***
countryCzech Republic:year      -2.594e-02  3.498e-03   -7.417 1.20e-13 ***
countryDenmark:year             -5.071e-02  3.676e-03  -13.797  < 2e-16 ***
countryEcuador:year             -4.084e-03  3.539e-03   -1.154 0.248515
countryEl Salvador:year         -3.292e-02  3.676e-03   -8.955  < 2e-16 ***
countryEstonia:year             -5.985e-02  3.996e-03  -14.979  < 2e-16 ***
countryFiji:year                -1.235e-02  1.682e-02   -0.734 0.462763
countryFinland:year             -4.984e-02  3.508e-03  -14.208  < 2e-16 ***
countryFrance:year              -3.559e-02  3.451e-03  -10.315  < 2e-16 ***
countryGeorgia:year             -1.695e-02  4.149e-03   -4.087 4.38e-05 ***
countryGermany:year             -4.264e-02  3.455e-03  -12.342  < 2e-16 ***
countryGreece:year              -1.603e-02  3.587e-03   -4.468 7.89e-06 ***
countryGrenada:year             -6.843e-02  1.800e-02   -3.801 0.000144 ***
countryGuatemala:year            4.844e-03  3.694e-03    1.311 0.189728
countryGuyana:year               3.440e-02  4.176e-03    8.237  < 2e-16 ***
countryHungary:year             -4.362e-02  3.480e-03  -12.536  < 2e-16 ***
countryIceland:year             -2.310e-02  4.737e-03   -4.877 1.08e-06 ***
countryIreland:year             -2.763e-03  3.592e-03   -0.769 0.441838
countryIsrael:year              -2.971e-02  3.603e-03   -8.245  < 2e-16 ***
countryItaly:year               -3.125e-02  3.458e-03   -9.037  < 2e-16 ***
countryJamaica:year              6.491e-02  1.127e-02    5.760 8.41e-09 ***
countryJapan:year               -1.892e-02  3.448e-03   -5.487 4.10e-08 ***
countryKazakhstan:year          -1.221e-02  3.470e-03   -3.518 0.000436 ***
countryKiribati:year            -2.700e-01  5.039e-02   -5.358 8.41e-08 ***
countryKuwait:year              -1.150e-02  5.394e-03   -2.132 0.032971 *
countryKyrgyzstan:year          -4.099e-02  3.652e-03  -11.226  < 2e-16 ***
countryLatvia:year              -4.579e-02  3.754e-03  -12.199  < 2e-16 ***
countryLithuania:year           -3.161e-02  3.573e-03   -8.846  < 2e-16 ***
countryLuxembourg:year          -3.678e-02  4.250e-03   -8.655  < 2e-16 ***
countryMaldives:year            -5.446e-02  6.295e-02   -0.865 0.387004
countryMalta:year                1.790e-02  5.935e-03    3.015 0.002569 **
countryMauritius:year           -3.247e-02  3.861e-03   -8.412  < 2e-16 ***
countryMexico:year               9.502e-03  3.462e-03    2.745 0.006052 **
countryMontenegro:year           4.187e-01  2.283e-02   18.334  < 2e-16 ***
countryNetherlands:year         -2.179e-02  3.478e-03   -6.267 3.69e-10 ***
countryNew Zealand:year         -2.386e-02  3.588e-03   -6.649 2.95e-11 ***
countryNorway:year              -3.147e-02  3.556e-03   -8.850  < 2e-16 ***
countryPanama:year              -1.470e-02  3.994e-03   -3.679 0.000234 ***
countryParaguay:year             2.383e-02  3.961e-03    6.016 1.79e-09 ***
countryPhilippines:year          3.505e-02  3.681e-03    9.523  < 2e-16 ***
countryPoland:year              -7.909e-03  3.462e-03   -2.285 0.022330 *
countryPortugal:year            -9.981e-03  3.517e-03   -2.838 0.004536 **
countryPuerto Rico:year         -3.145e-02  3.640e-03   -8.640  < 2e-16 ***
countryQatar:year               -1.136e-03  9.932e-03   -0.114 0.908903
```

```
countryRepublic of Korea:year    3.561e-02  3.452e-03   10.316  < 2e-16 ***
countryRomania:year             -1.005e-02  3.476e-03   -2.890 0.003850 **
countryRussian Federation:year  -3.560e-02  3.446e-03  -10.330  < 2e-16 ***
countrySaint Lucia:year         -2.814e-02  8.581e-03   -3.280 0.001039 **
countrySerbia:year              -2.633e-02  3.657e-03   -7.201 5.97e-13 ***
countrySeychelles:year          -2.081e-02  1.229e-02   -1.694 0.090332 .
countrySingapore:year           -2.473e-02  3.618e-03   -6.835 8.21e-12 ***
countrySlovakia:year            -3.326e-02  3.691e-03   -9.012  < 2e-16 ***
countrySlovenia:year            -4.495e-02  3.800e-03  -11.829  < 2e-16 ***
countrySouth Africa:year         2.458e-02  4.021e-03    6.112 9.85e-10 ***
countrySpain:year               -1.817e-02  3.462e-03   -5.249 1.53e-07 ***
countrySri Lanka:year           -4.051e-02  3.506e-03  -11.555  < 2e-16 ***
countrySuriname:year             2.324e-02  4.396e-03    5.287 1.24e-07 ***
countrySweden:year              -3.901e-02  3.497e-03  -11.156  < 2e-16 ***
countrySwitzerland:year         -4.640e-02  3.595e-03  -12.908  < 2e-16 ***
countryThailand:year            -1.721e-02  3.461e-03   -4.974 6.56e-07 ***
countryTrinidad and Tobago:year -1.404e-02  4.025e-03   -3.489 0.000484 ***
countryTurkmenistan:year        -3.532e-02  3.707e-03   -9.530  < 2e-16 ***
countryUkraine:year             -3.169e-02  3.453e-03   -9.179  < 2e-16 ***
countryUnited Kingdom:year      -2.721e-02  3.458e-03   -7.869 3.56e-15 ***
countryUnited States:year       -2.099e-02  3.447e-03   -6.090 1.13e-09 ***
countryUruguay:year              3.341e-03  3.589e-03    0.931 0.351987
countryUzbekistan:year          -2.738e-02  3.524e-03   -7.769 7.91e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 5471427  on 4523  degrees of freedom
Residual deviance:  215823  on 4344  degrees of freedom
AIC: 247971

    Number of Fisher Scoring iterations: 7

    > dwtest(gmod7a)

        Durbin-Watson test

    data:  gmod7a
    DW = 0.072859, p-value < 2.2e-16
    alternative hypothesis: true autocorrelation is greater than 0

  suicide2=subset(suicide,country!="Bosnia and Herzegovi" & country!="Cabo Verde"
                  &country!="Dominica" & country!="Macau" & country!="Mongolia" &
                    country!="Nicaragua" &country!="Oman" &country!="Saint Kitts and Nevi
"
                  &country!="Saint Vincent and Gr" &country!="San Marino"
                    &country!="Turkey" &country!="United Arab Emirates")

  gmod1<-glm(SuiCount ~ log(gdp_for_year) +
               offset(log(population)), data = suicide2, family = poisson)
  gmod2<-glm(SuiCount ~ sex + log(gdp_for_year) +
               offset(log(population)), data = suicide2, family = poisson)
  gmod3<-glm(SuiCount ~ year + log(gdp_for_year) +
               offset(log(population)), data = suicide2, family = poisson)
  gmod4<-glm(SuiCount ~ country + log(gdp_for_year) +
               offset(log(population)), data = suicide2, family = poisson)
  gmod5<-glm(SuiCount ~ country * sex + log(gdp_for_year) +
               offset(log(population)), data = suicide2, family = poisson)
```

```
gmod6<-glm(SuiCount ~ country * year + log(gdp_for_year) +
             offset(log(population)), data = suicide2, family = poisson)
gmod7a<-glm(SuiCount ~ sex + country * year + log(gdp_for_year) +
             offset(log(population)), data = suicide2, family = poisson)
gmod8<-glm(SuiCount ~ year + country * sex + log(gdp_for_year) +
             offset(log(population)), data = suicide2, family = poisson)
gmod9<-glm(SuiCount ~ sex + country + log(gdp_for_year) +
             offset(log(population)), data = suicide2, family = poisson)
gmod10<-glm(SuiCount ~ sex + year + log(gdp_for_year) +
              offset(log(population)), data = suicide2, family = poisson)
gmod11<-glm(SuiCount ~ year + country + log(gdp_for_year) +
              offset(log(population)), data = suicide2, family = poisson)
gmod12<-glm(SuiCount ~ sex * year + log(gdp_for_year) +
              offset(log(population)), data = suicide2, family = poisson)
gmod13<-glm(SuiCount ~ country + sex * year + log(gdp_for_year) +
              offset(log(population)), data = suicide2, family = poisson)
gmod14<-glm(SuiCount ~ year + sex + country +log(gdp_for_year) +
                offset(log(population)), data = suicide2, family = poisson)

suicide1 = subset(suicide2, country!="Russian Federation" & country!="Romania"
                  &country!="Kazakhstan" & country!="Netherlands")
gmod7b<-glm(SuiCount ~ sex + country * year + log(gdp_for_year) +
                offset(log(population)), data = suicide1, family = poisson)

  > compareCoefs(gmod7a,gmod7b)
                                Model 1  Model 2
(Intercept)                       -71.99    -57.09
SE                                  6.89      6.85

sexmale                         1.271282  1.182120
SE                              0.000914  0.001011

year                             0.03308   0.02480
SE                               0.00345   0.00343

log(gdp_for_year)               -0.24095  -0.16382
SE                               0.00137   0.00212

  > library(leaps)
> AIC(gmod1)
[1] 5488116
> AIC(gmod2)
[1] 3251382
> AIC(gmod3)
[1] 5479904
> AIC(gmod4)
[1] 2707236
> AIC(gmod5)
[1] 275563.1
> AIC(gmod6)
[1] 2559704
> AIC(gmod7a)
[1] 247970.7
> AIC(gmod8)
[1] 265999.1
> AIC(gmod9)
[1] 394918.3
> AIC(gmod10)
[1] 3243076
```

```
> AIC(gmod11)
[1] 2697298
> AIC(gmod12)
[1] 3240235
> AIC(gmod13)
[1] 381944.4
> AIC(gmod14)
[1] 385269.3

  > library(car)
  Loading required package: carData
> BIC(gmod7a)
[1] 249125.8
> BIC(gmod8)
  [1] 267154.2

> anova(gmod1, gmod7a,test = "Chisq")
Analysis of Deviance Table

Model 1: SuiCount ~ log(gdp_for_year) + offset(log(population))
Model 2: SuiCount ~ sex + country * year + log(gdp_for_year) + offset(log(population))
  Resid. Df Resid. Dev  Df Deviance  Pr(>Chi)
1      4522    5456325
2      4344     215823 178  5240502 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> anova(gmod3, gmod7a,test = "Chisq")
Analysis of Deviance Table

Model 1: SuiCount ~ year + log(gdp_for_year) + offset(log(population))
Model 2: SuiCount ~ sex + country * year + log(gdp_for_year) + offset(log(population))
  Resid. Df Resid. Dev  Df Deviance  Pr(>Chi)
1      4521    5448111
2      4344     215823 177  5232288 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> anova(gmod4, gmod7a,test = "Chisq")
Analysis of Deviance Table

Model 1: SuiCount ~ country + log(gdp_for_year) + offset(log(population))
Model 2: SuiCount ~ sex + country * year + log(gdp_for_year) + offset(log(population))
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      4434    2675269
2      4344     215823 90  2459445 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> anova(gmod11, gmod7a,test = "Chisq")
Analysis of Deviance Table

Model 1: SuiCount ~ year + country + log(gdp_for_year) + offset(log(population))
Model 2: SuiCount ~ sex + country * year + log(gdp_for_year) + offset(log(population))
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      4433    2665328
2      4344     215823 89  2449505 < 2.2e-16 ***
---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> influencePlot(gmod7a, sub="Figure 17: influence plot")
        StudRes        Hat        CookD
```

```
1146 -61.1338848 0.11246433 2.3391893344
1173  -0.4878846 0.36814537 0.0007491796
1424  -6.7151705 0.48389485 0.2157239095
1719 -45.1724410 0.14273637 1.6937400664
1753  58.9329088 0.06191676 1.3873477725
```