Sample Code

Blair Cha

23 April 2020

Contents

This is the sample code for my Statistics capstone project, `Investigating the Causes of Suicide Across Nations: A Mixed-Effect Model Approach.` I use a mixed-effect model with the random slope effect to conduct longitudinal data analysis. It is composed of five steps shown above.

The fundamental part of the data comes from My World in Data (Ritchie, Roser, and Ortiz-Ospina 2015). Based on the literature, I obtain additional variables that could explain suicide rates from Gapminder.com ("Data," n.d.), the United Nations Development Program ("Human Development Reports," n.d.), World Bank ("World Development Indicators," n.d.), and the World Health Organization ("Suicide," n.d.).

The final report for the project is the writing sample in my application.

Step 1. Data Cleaning

Install Relevant Libraries

```
library(readr)
library(dplyr)
library(ggplot2)
library(tidyverse)
library(lubridate)
library(countrycode)
library(stringr)
library(Hmisc)
library(xtable)
library(memisc)
```

Download the .xlsx and .csv files into your device, and import data.

```
BMI0 <- read_csv("NCD_BMI_25A.csv")
Suicide_Rates <- read_csv("suicide-death-rates.csv")
Gini0 <- readxl::read_xlsx("Gini.xlsx")
Mental_disorder0 <- readxl::read_xlsx("Mental disorder.xlsx")
colnames(Mental_disorder0)[colnames(Mental_disorder0)=="Country"] <- "country"
colnames(Mental_disorder0)[colnames(Mental_disorder0)=="Year"] <- "year"

Gender_Ratio0 <- readxl::read_xlsx("Gender Ratio.xlsx")
Unemployment0 <- readxl::read_xlsx("Unemployment.xlsx")
CO2 <- readxl::read_xlsx("CO2.xlsx")
Adolescent <- readxl::read_xlsx("Adolescent fertility.xlsx")
Inflation <- readxl::read_xlsx("Inflation.xlsx")
Freedom <- readxl::read_xlsx("Freedom.xlsx")
Rural <- readxl::read_xlsx("Rural population.xlsx")
BirthRate <- readxl::read_xlsx("Birth rate.xlsx")
Military <- readxl::read_xlsx("Military expenditure.xlsx")
GDP_capita <- readxl::read_xlsx("GDP_capita.xlsx")
```

Convert each dataset into a longer format with pivot_longer.

```
BMI <- BMI0 %>%
  pivot_longer(cols = -country,
               names_to = "year",
               values_to="BMI") %>%
    mutate(year = as.numeric(year))

Gini <- Gini0 %>%
  pivot_longer(cols = -country,
               names_to = "year",
               values_to="Gini") %>%
    mutate(year = as.numeric(year))

Mental_disorder <- Mental_disorder0 %>%
  mutate(year = as.numeric(year))

Gender_Ratio <- Gender_Ratio0 %>%
  pivot_longer(cols = -country,
               names_to = "year",
               values_to="Gender_ratio") %>%
  mutate(year=as.numeric(year))

Unemployment <- Unemployment0 %>%
  pivot_longer(cols = -country,
               names_to = "year",
               values_to="Unemployment") %>%
  mutate(year = as.numeric(year))%>%
  drop_na()

CO2 <- CO2 %>%
  pivot_longer(cols = -country,
               names_to = "year",
```

```
               values_to="CO2") %>%
  mutate(year = as.numeric(year))

Inflation <- Inflation %>%
  pivot_longer(cols = -country,
               names_to = "year",
               values_to="Inflation") %>%
  mutate(year = as.numeric(year))%>%
  drop_na()

Adolescent <- Adolescent %>%
  pivot_longer(cols = -country,
               names_to = "year",
               values_to="Adolescent") %>%
  mutate(year = as.numeric(year))

Freedom <- Freedom %>%
  pivot_longer(cols = -country,
               names_to = "year",
               values_to="Freedom") %>%
  mutate(year = as.numeric(year))

BirthRate <- BirthRate %>%
  pivot_longer(cols = -country,
               names_to = "year",
               values_to="BirthRate") %>%
  mutate(year = as.numeric(year))

Military <- Military %>%
  pivot_longer(cols = -country,
               names_to = "year",
               values_to="Military") %>%
  mutate(year = as.numeric(year))

Rural <- Rural%>%
  pivot_longer(cols = -country,
               names_to = "year",
               values_to="Rural") %>%
  mutate(year = as.numeric(year))

GDP_capita <- GDP_capita%>%
  pivot_longer(cols = -country,
               names_to = "year",
               values_to="GDP_capita") %>%
  mutate(year = as.numeric(year))
```

Join all datasets

```
Suicide_Rates <- Suicide_Rates %>%
  inner_join(Gini, by = c("year","country")) %>%
  inner_join(Unemployment, by = c("year","country")) %>%
  inner_join(Gender_Ratio, by = c("year","country")) %>%
  inner_join(BMI, by = c("year","country")) %>%
```

```
  inner_join(Mental_disorder, by = c("year","country")) %>%
  inner_join(CO2, by = c("year","country")) %>%
  inner_join(Inflation, by = c("year","country")) %>%
  inner_join(Adolescent, by = c("year","country")) %>%
  inner_join(Rural, by = c("year","country")) %>%
  inner_join(Freedom, by = c("year","country")) %>%
  inner_join(Military, by = c("year","country")) %>%
  inner_join(BirthRate, by = c("year","country")) %>%
  inner_join(GDP_capita, by = c("year","country")) %>%
  mutate(scale_GDP_capita = GDP_capita/1000) %>%
  mutate(scale_CO2 = CO2/1000)
```

Other final changes to the dataset: deleting and renaming columns, cleaning rows in a column, and KNN imputation.

```
# Delete column called, "Code."
Suicide_Rates$Code <- NULL


# Rename the Prevalence of Mental and Substance Use Disorder variable to "disorder."
colnames(Suicide_Rates)[
  colnames(Suicide_Rates)=="Prevalence - Mental and substance use disorders - Sex: Both - Age: Age-stanc
  ] <- "disorder"


# Delete the range of BMI in brackets "[,]" in every row to leave just average BMI rates.
Suicide_Rates <- Suicide_Rates %>%
  mutate(BMI = str_replace(BMI, " \\[.*\\]", "")) %>%
  mutate(BMI = as.numeric(BMI))


# Add the "continent" varible that indicates which continent each country belongs to.
newdata <- data.frame(country=Suicide_Rates$country)
newdata$continent <- countrycode(sourcevar = Suicide_Rates$country,
                                 origin= "country.name",
                                 destination= "continent")

newdata <- newdata %>%
  dplyr::select(continent, country) %>%
  distinct()

Suicide_Rates <- Suicide_Rates %>%
  inner_join(newdata, by="country")


# Using the VIM package, fill in missing data in all columns
# except for the variables after "-."
library(VIM)
Suicide <- Suicide_Rates %>%
  dplyr::select(-continent, -Suicide, -Gender_ratio, -GDP_capita, -CO2) %>%
  VIM::kNN(imp_var = FALSE) %>%
  mutate(Suicide = Suicide_Rates$Suicide) %>%
  mutate(continent = Suicide_Rates$continent)
```

Save cleaned dataset as "Suicide.Rda."

```
save(Suicide, file="Suicide.Rda")
```

Step 2. Fitting Mixed Effects Models

Review of Mixed Effects Models

A linear mixed effects model in terms of the outcome vector can be written as,

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i$$

$$= \text{Fixed effects} + \text{Random effects} + \text{Error}$$

where $\mathbf{X}_i$ is the design matrix for the fixed effects, $\mathbf{Z}_i$ is the design matrix for the random effects (a subset of columns of $\mathbf{X}_i$), $\boldsymbol{\epsilon}_i \sim N(0, \boldsymbol{\Sigma})$, $\mathbf{b}_i \sim N(0, \mathbf{G})$, and $\mathbf{b}_i$ and $\boldsymbol{\epsilon}_i$ are independent.

Thus,

$$\mathbf{Y}_i \sim \mathcal{N}(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^T + \boldsymbol{\Sigma})$$

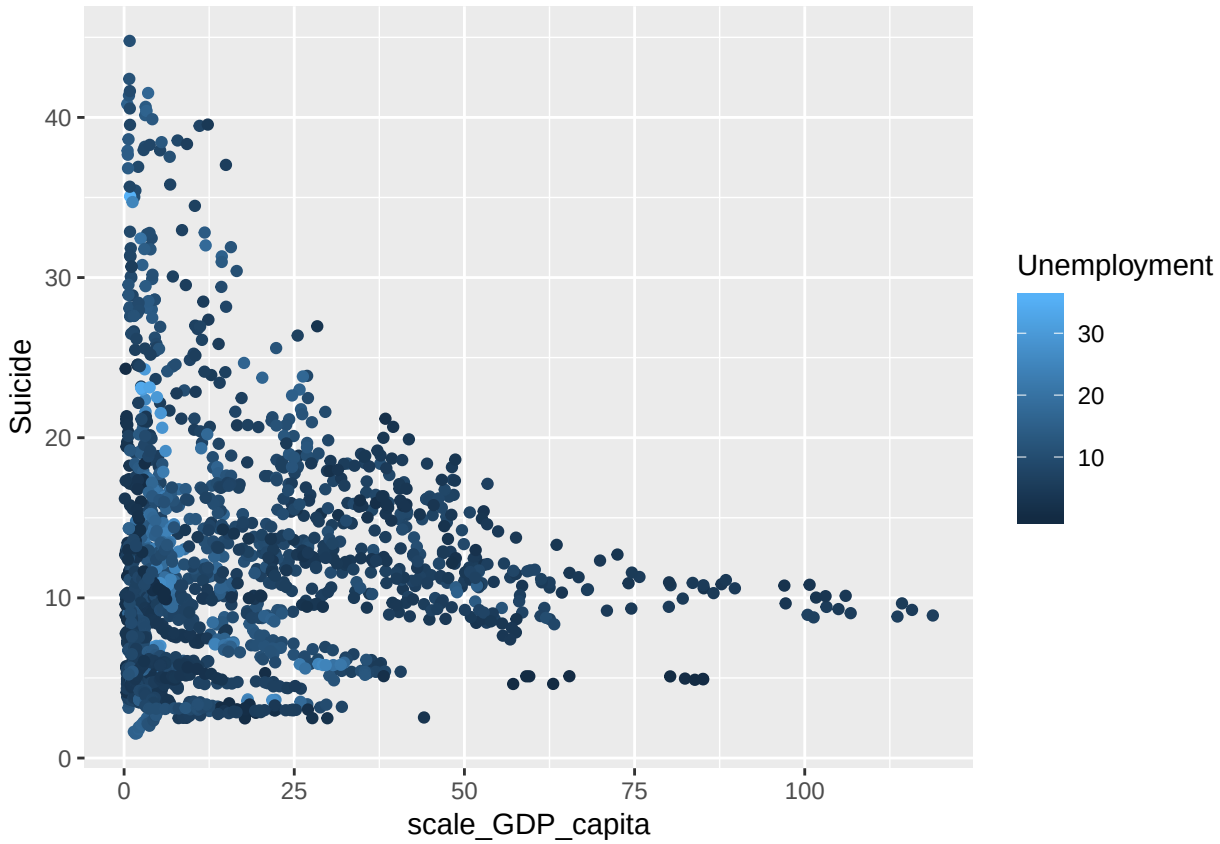
Load previously cleaned dataset and relevant libraries.

```
load("Suicide.Rda")
require(dplyr)
library(lme4) # Package for Mixed Models
```

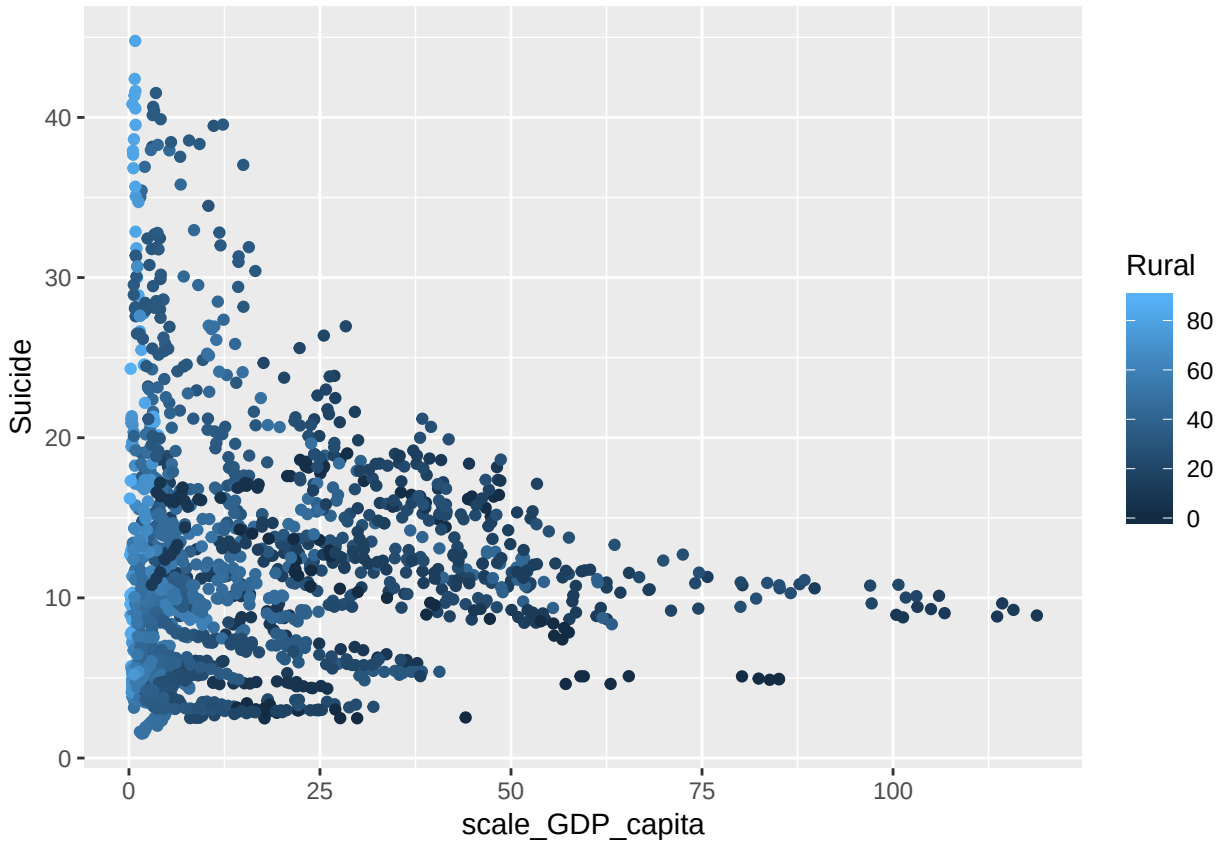
Generate scatterplots to assess the existence of any interaction between variables.

There seems to be no clear indication of an interaction effect where one variable affects another's relationship with the response variable, Suicide. Therefore, I do not include any interaction variables in my final model.

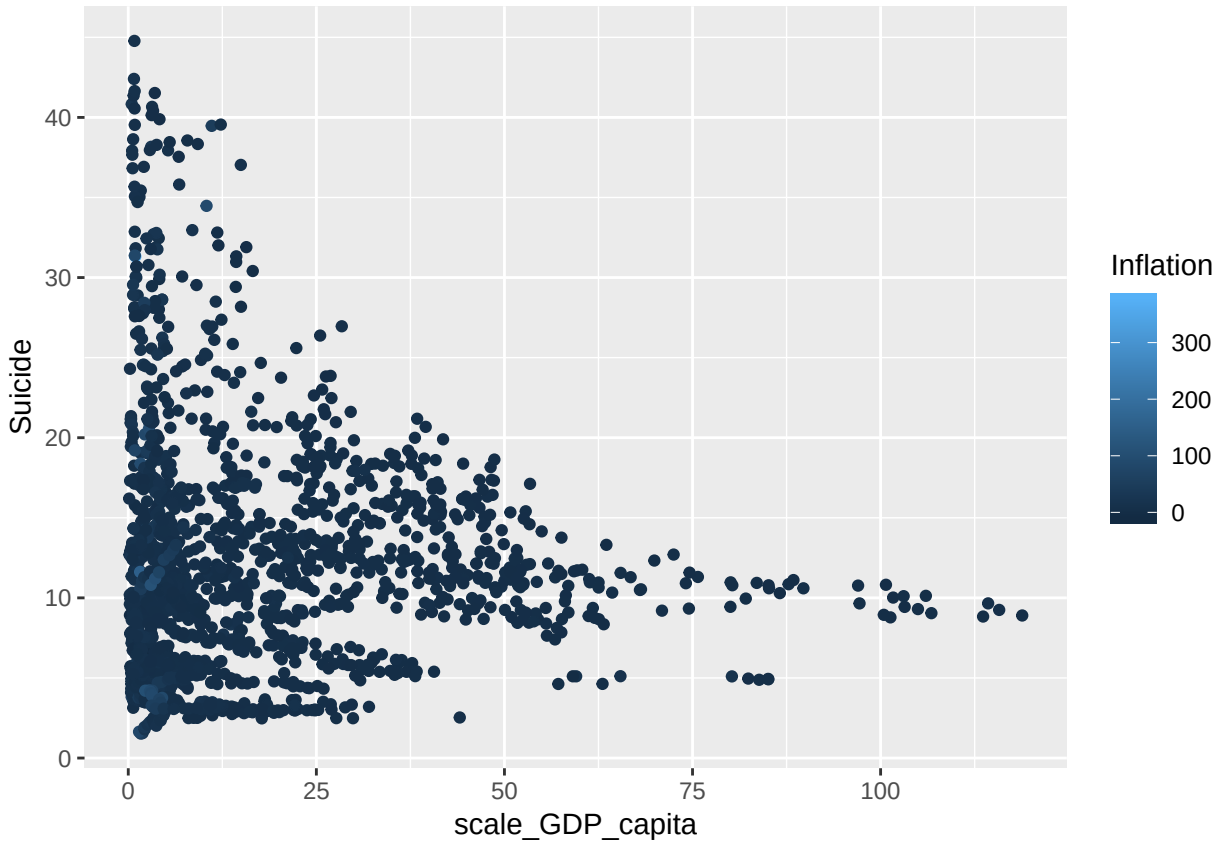
```
ggplot(Suicide, aes(x=scale_GDP_capita, y=Suicide, color=Unemployment)) + geom_point()
```



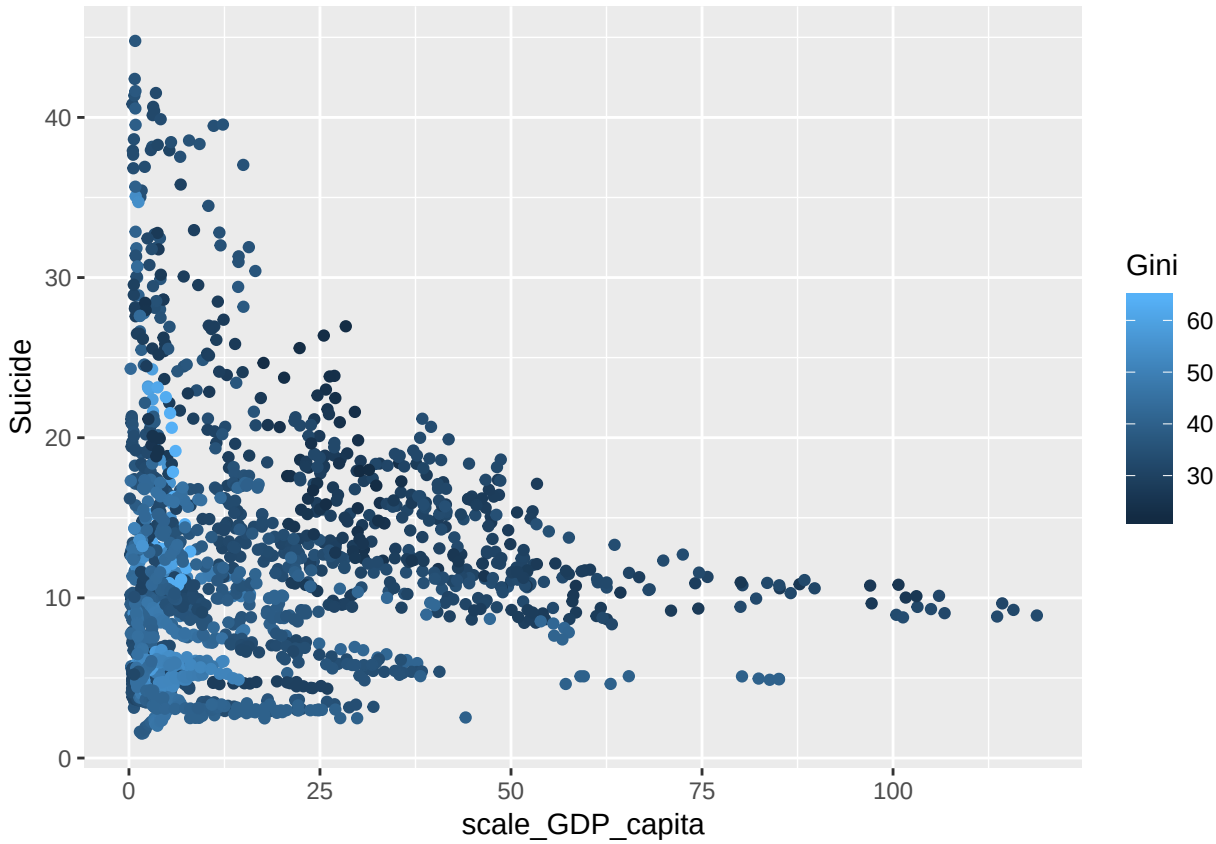

```
ggplot(Suicide, aes(x=scale_GDP_capita, y=Suicide, color=Rural)) + geom_point()
```



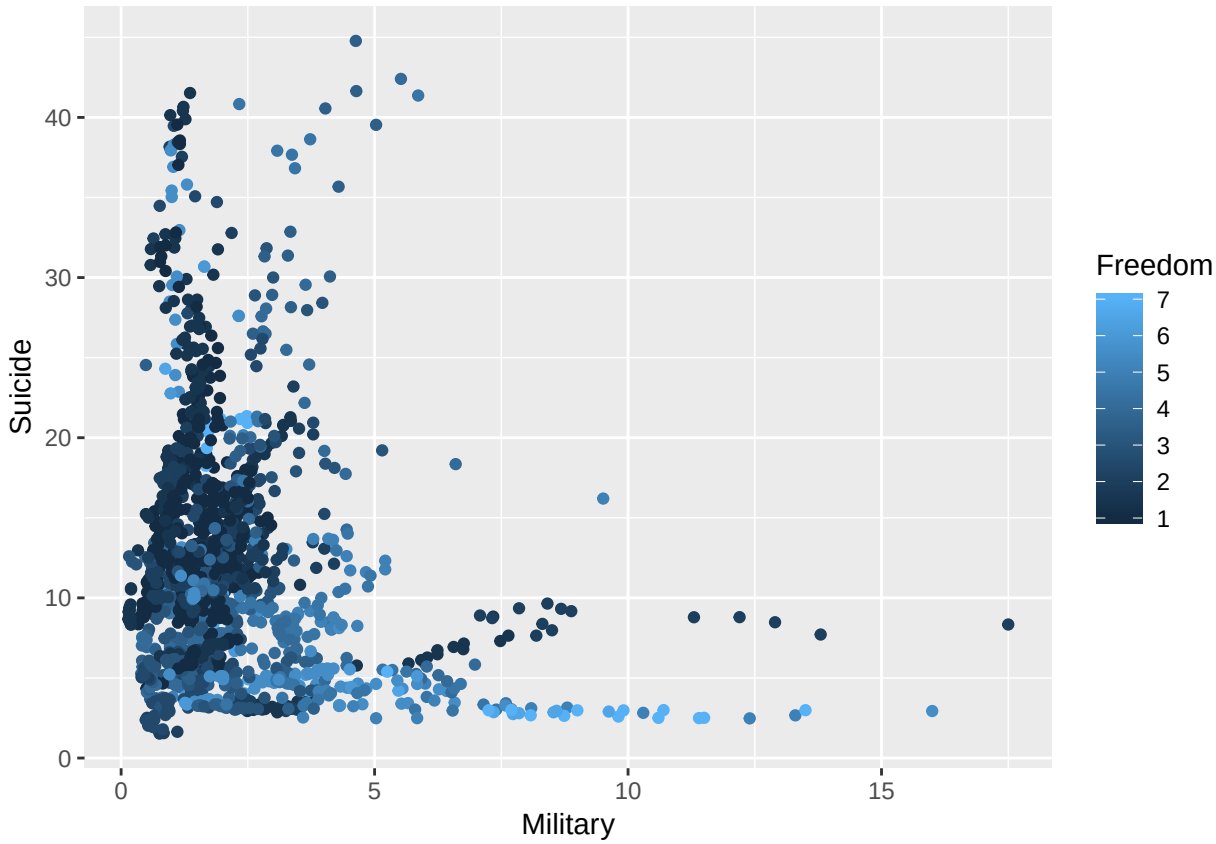

```
ggplot(Suicide, aes(x=scale_GDP_capita, y=Suicide, color=Inflation)) + geom_point()
```



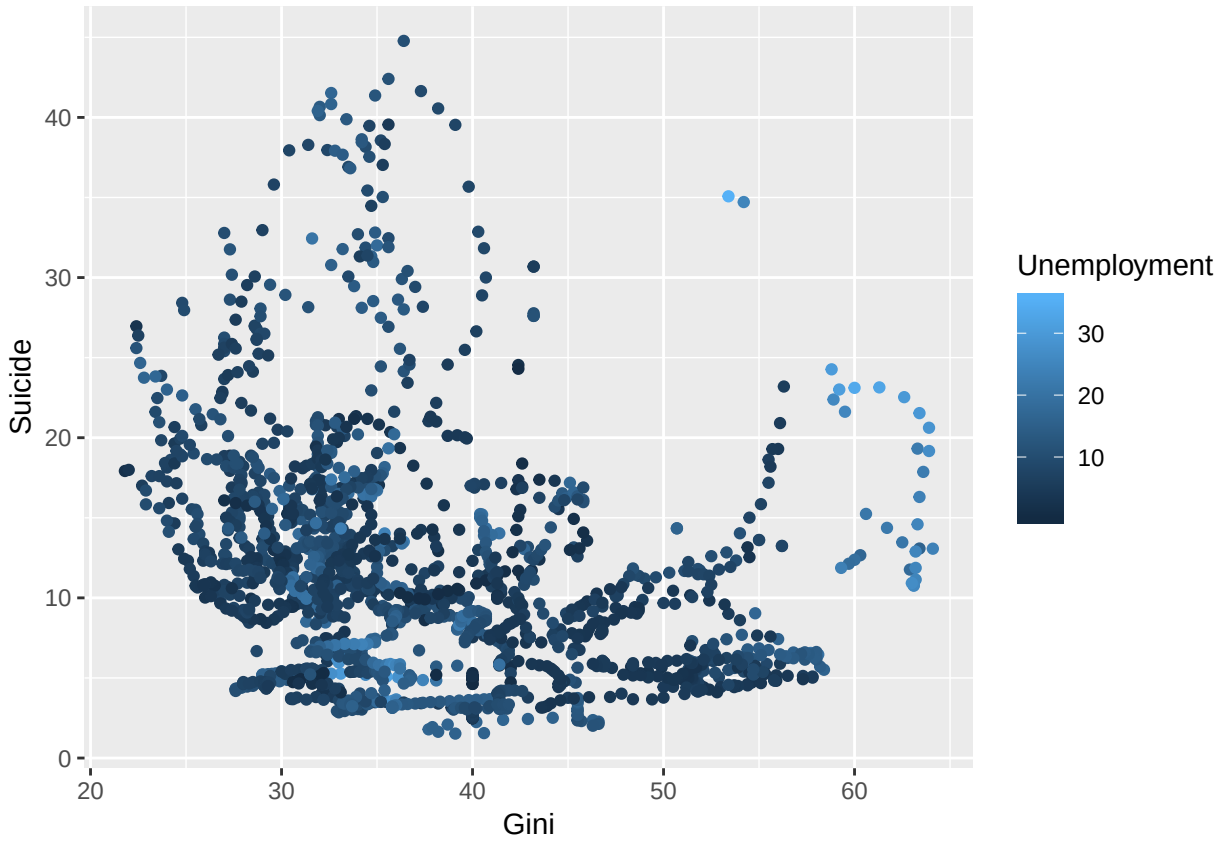

```
ggplot(Suicide, aes(x=scale_GDP_capita, y=Suicide, color=Gini)) + geom_point()
```



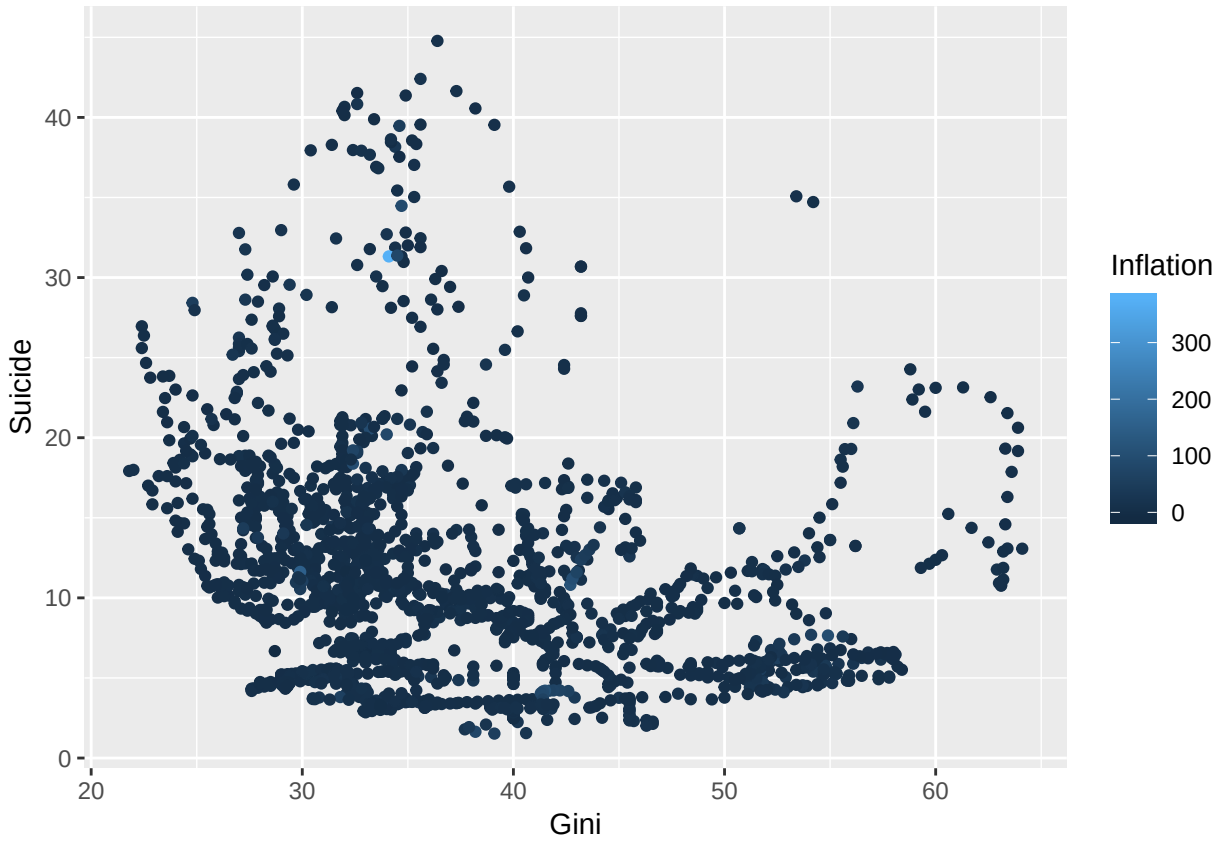

```
ggplot(Suicide, aes(x=Military, y=Suicide, color=Freedom)) + geom_point()
```




```
ggplot(Suicide, aes(x=Gini, y=Suicide, color=Unemployment)) + geom_point()
```




```
ggplot(Suicide, aes(x=Gini, y=Suicide, color=Inflation)) + geom_point()
```



Fitting Candidate Models

I exclude variables "proportion of education attainment at the primary school level" and "gender ratio" because they contain too many NAs and would lead to inaccurate results.

I first include all my explanatory variables in the model and allow all of them to have a random slope.

```
mod1 = lmer(Suicide ~
              Unemployment + BMI + disorder + scale_CO2 + Adolescent + Rural + BirthRate +
              scale_GDP_capita + Gini + Inflation + Freedom + Military +
              (Unemployment + BMI + disorder + scale_CO2 + Adolescent +
                Rural+ BirthRate + scale_GDP_capita + Gini + Inflation +
                Freedom + Military|country),
            data = Suicide, REML = FALSE)
summary(mod1)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Suicide ~ Unemployment + BMI + disorder + scale_CO2 + Adolescent +
##     Rural + BirthRate + scale_GDP_capita + Gini + Inflation +
##     Freedom + Military + (Unemployment + BMI + disorder + scale_CO2 +
##     Adolescent + Rural + BirthRate + scale_GDP_capita + Gini +
##     Inflation + Freedom + Military | country)
##    Data: Suicide
##
##      AIC      BIC   logLik deviance df.resid
##   4397.1   4969.5  -2093.6   4187.1     1616
##
## Scaled residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -6.9297 -0.3161 -0.0066  0.2732  6.9909
##
## Random effects:
##  Groups   Name             Variance  Std.Dev. Corr
##  country  (Intercept)      5.662e-01 0.752493
##           Unemployment     1.026e-02 0.101270  0.95
##           BMI              2.475e-01 0.497476 -0.04 -0.04
##           disorder         1.195e+01 3.457118 -0.26 -0.27 -0.89
##           scale_CO2        7.129e-05 0.008443  0.76  0.54 -0.13 -0.04
##           Adolescent       7.377e-02 0.271598 -0.55 -0.41  0.21 -0.07 -0.71
##           Rural            1.905e-01 0.436422  0.45  0.21  0.21 -0.36  0.75
##           BirthRate        4.228e-01 0.650256  0.21  0.27  0.40 -0.51 -0.06
##           scale_GDP_capita 6.330e-04 0.025160 -0.15 -0.20 -0.42  0.34  0.05
##           Gini             1.116e-01 0.334008  0.22  0.27  0.10 -0.35 -0.03
##           Inflation        9.325e-04 0.030537 -0.13 -0.18 -0.23  0.12  0.05
##           Freedom          6.242e-01 0.790078  0.00  0.05  0.11 -0.06 -0.12
##           Military         4.688e-01 0.684714  0.45  0.42  0.21 -0.44  0.29
##  Residual                  1.753e-01 0.418711
##
##
##
##
##
##
##
##  -0.48
##  -0.40  0.03
##   0.24  0.24 -0.78
##   0.20 -0.17  0.11 -0.11
##  -0.11  0.45 -0.19  0.66 -0.29
##  -0.02 -0.37  0.20 -0.44  0.11 -0.81
##   0.27  0.40 -0.13  0.20  0.47  0.19 -0.51
##
## Number of obs: 1721, groups:  country, 121
##
## Fixed effects:
##                    Estimate Std. Error t value
## (Intercept)      26.6905744  5.0010667   5.337
## Unemployment      0.0191715  0.0150367   1.275
## BMI              -0.3927392  0.0640068  -6.136
## disorder          0.8531684  0.4880111   1.748
## scale_CO2         0.0003941  0.0015558   0.253
## Adolescent        0.0302878  0.0317799   0.953
## Rural            -0.1222530  0.0625803  -1.954
## BirthRate        -0.1524909  0.0839962  -1.815
## scale_GDP_capita -0.0042812  0.0053181  -0.805
## Gini             -0.0461042  0.0447443  -1.030
## Inflation         0.0005642  0.0051542   0.109
## Freedom           0.0048054  0.1138919   0.042
## Military          0.1236945  0.0990567   1.249
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

Looking at the Random effects result, I remove random slopes that have a small standard deviation: Inflation, Unemployment, scale_CO2, and scale_GDP_capita. This is because having a small standard deviation means the relationship between the number of suicide and the explainatory variable does not differ much compared to other countries.

Looking at the Fixed effects result, I remove variables with high standard error, Freedom and Military. At this point, I decide to keep disorder in my model, since despite the large standard error, the standard deviation of the random slope distribution is quite high. This means that the impact of disorder on suicide varies a lot among countries. Thus, I conclude that adding disorder in my model would help explain a legitimate proportion of the variance of the original data set.

I fit the following models with random slopes BMI, disorder, and BirthRate, the variables with the highest standard deviation from mod1.

```
mod2 = lmer(Suicide ~  BMI + disorder + Adolescent + Rural + BirthRate  + Inflation +
               Gini +
                (BMI + disorder + BirthRate|country),
                 data = Suicide, REML = FALSE)
summary(mod2)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## Suicide ~ BMI + disorder + Adolescent + Rural + BirthRate + Inflation +
##     Gini + (BMI + disorder + BirthRate | country)
##    Data: Suicide
##
##      AIC      BIC   logLik deviance df.resid
##   5274.4   5378.0  -2618.2   5236.4     1702
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.8799 -0.3348 -0.0032  0.2646  7.9755
##
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
##  country  (Intercept) 9489.0937 97.4120
##           BMI            0.4577  0.6765  -0.17
##           disorder      65.9247  8.1194  -0.90 -0.25
##           BirthRate      0.9044  0.9510  -0.22  0.53 -0.08
##  Residual                0.4276  0.6539
## Number of obs: 1721, groups:  country, 121
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 22.143443  12.842606   1.724
## BMI         -0.480048   0.080427  -5.969
## disorder     1.371765   1.011020   1.357
## Adolescent   0.067145   0.014829   4.528
## Rural       -0.169653   0.040766  -4.162
## BirthRate   -0.270086   0.110753  -2.439
## Inflation    0.010296   0.001791   5.750
## Gini         0.044750   0.020352   2.199
##
## Correlation of Fixed Effects:
##             (Intr) BMI    disrdr Adlscn Rural  BrthRt Infltn
## BMI         -0.274
```

```
## disorder   -0.893 -0.145
## Adolescent -0.034  0.095  0.003
## Rural      -0.204  0.290  0.010  0.028
## BirthRate  -0.184  0.499 -0.096 -0.235 -0.063
## Inflation  -0.047  0.026  0.038 -0.092  0.020  0.000
## Gini       -0.077  0.027  0.002 -0.229  0.040  0.090  0.059
## convergence code: 0
## Model failed to converge with max|grad| = 0.0141675 (tol = 0.002, component 1)
```

From the Information Criteria BIC, I discover that including the Gini variable generates higher BIC values, indicating the model is overfit and penalized. Therefore, I fit another model, mod3 without Gini.

```
mod3 = lmer(Suicide ~  BMI + disorder + Adolescent + Rural + BirthRate  + Inflation +
               (BMI + disorder + BirthRate|country),
                data = Suicide, REML = FALSE)
summary(mod3)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## Suicide ~ BMI + disorder + Adolescent + Rural + BirthRate + Inflation +
##     (BMI + disorder + BirthRate | country)
##    Data: Suicide
##
##      AIC      BIC   logLik deviance df.resid
##   5277.2   5375.3  -2620.6   5241.2     1703
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.8754 -0.3358 -0.0009  0.2572  8.0176
##
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
##  country  (Intercept) 9176.2133 95.7926
##           BMI            0.4518  0.6721  -0.17
##           disorder      63.8466  7.9904  -0.90 -0.25
##           BirthRate      0.9068  0.9522  -0.24  0.53 -0.07
##  Residual                0.4302  0.6559
## Number of obs: 1721, groups:  country, 121
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 24.21533   12.65680   1.913
## BMI         -0.48416    0.07994  -6.057
## disorder     1.37270    0.99907   1.374
## Adolescent   0.07465    0.01444   5.171
## Rural       -0.17201    0.04072  -4.225
## BirthRate   -0.29403    0.11027  -2.666
## Inflation    0.01009    0.00179   5.636
##
## Correlation of Fixed Effects:
##             (Intr) BMI    disrdr Adlscn Rural  BrthRt
## BMI         -0.273
## disorder    -0.894 -0.146
## Adolescent -0.051  0.104  0.002
## Rural      -0.202  0.290  0.009  0.038
```

```
## BirthRate  -0.187  0.499 -0.089 -0.222 -0.067
## Inflation  -0.043  0.024  0.039 -0.083  0.018 -0.005
## convergence code: 0
## Model failed to converge with max|grad| = 0.185822 (tol = 0.002, component 1)
```

```
BIC(mod2)
```

```
## [1] 5377.958
```

```
BIC(mod3)
```

```
## [1] 5375.323
```

According to the Fixed effects result, all variables pass the hypothesis test by having t-values greater than 2, except for disorder. I keep disorder in my model for reasons stated above.

mod3 seems like the most optimal model so far. Before making any decision, I generate another model with gdp per capita based on my literature review that suicide is a combined result of mental illness, major life events, and economic conditions (Berk, Dodd, and Henry 2006).

```
mod4 = lmer(Suicide ~  BMI + disorder + Adolescent + Rural + BirthRate + Inflation +
              scale_GDP_capita +
                (BMI + disorder + BirthRate|country),
                 data = Suicide, REML = FALSE)
summary(mod4)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## Suicide ~ BMI + disorder + Adolescent + Rural + BirthRate + Inflation +
##     scale_GDP_capita + (BMI + disorder + BirthRate | country)
##    Data: Suicide
##
##      AIC      BIC   logLik deviance df.resid
##   5279.2   5382.7  -2620.6   5241.2     1702
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.8771 -0.3349 -0.0009  0.2579  8.0246
##
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
##  country  (Intercept) 9550.9960 97.7292
##           BMI            0.4545  0.6742  -0.17
##           disorder      65.9576  8.1214  -0.90 -0.24
##           BirthRate      0.9081  0.9529  -0.24  0.53 -0.07
##  Residual                0.4290  0.6550
## Number of obs: 1721, groups:  country, 121
##
## Fixed effects:
##                    Estimate Std. Error t value
## (Intercept)      24.1473729 12.8461903   1.880
## BMI              -0.4846174  0.0802909  -6.036
## disorder          1.3800729  1.0127333   1.363
## Adolescent        0.0747439  0.0144568   5.170
## Rural            -0.1725391  0.0409093  -4.218
## BirthRate        -0.2938607  0.1106257  -2.656
## Inflation         0.0100718  0.0017895   5.628
```

```
## scale_GDP_capita  0.0001775  0.0051265   0.035
##
## Correlation of Fixed Effects:
##             (Intr) BMI    disrdr Adlscn Rural  BrthRt Infltn
## BMI         -0.272
## disorder    -0.897 -0.141
## Adolescent  -0.052  0.102  0.003
## Rural       -0.201  0.285  0.012  0.040
## BirthRate   -0.188  0.501 -0.085 -0.222 -0.071
## Inflation   -0.042  0.025  0.038 -0.083  0.017 -0.005
## scl_GDP_cpt -0.032 -0.050  0.036  0.027  0.075 -0.061 -0.014
## convergence code: 0
## Model failed to converge with max|grad| = 0.00514339 (tol = 0.002, component 1)
```

Step 3. Hypothesis Testing and Information Criteria

Likelihood Ratio Test

I compare mod2 and mod3 with the likelihood ratio test. Suppose we have two models with the same random effects and covariance models. In this case, the full model is mod2, and the nested is mod3.

- Full model: M_f has p columns in $\mathbf{X}_i$ and thus p fixed effects $\beta_0, ..., \beta_{p-1}$.
- Nested model: M_n has k columns such that $k < p$ and $p - k$ fixed effects $\beta_l = 0$.

If we are using maximum likelihood estimation, it makes sense to compare the likelihood of two models.

H_0 : nested model is true, H_A : full model is true

If we take the ratio of the likelihoods from the nested model and full model and plug in the maximum likelihood estimators, then we have another statistic.

$$D = -2\log\left(\frac{L_n(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}})}{L_f(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}})}\right) = -2\log(L_n(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}})) + 2\log(L_f(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}}))$$

The sampling distribution of this statistic is approximatley **chi-squared** with degrees of freedom equal to the **difference in the number of parameters between the two models**.

```
anova(mod3, mod4)
```

```
## Data: Suicide
## Models:
## mod3: Suicide ~ BMI + disorder + Adolescent + Rural + BirthRate + Inflation +
## mod3:     (BMI + disorder + BirthRate | country)
## mod4: Suicide ~ BMI + disorder + Adolescent + Rural + BirthRate + Inflation +
## mod4:     scale_GDP_capita + (BMI + disorder + BirthRate | country)
##      Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## mod3 18 5277.2 5375.3 -2620.6   5241.2
## mod4 19 5279.2 5382.7 -2620.6   5241.2 0.0554      1     0.8139
```

Since the p-value, 0.81, is not at a statistically significant level, I do not reject the null hypothesis that the smaller model is correct. Thus, I favor the nested model, mod3.

Hypothesis Testing for Coefficients

To test whether multiple slopes are zero or a linear combination of slopes is zero, $H_0 : \mathbf{L}\boldsymbol{\beta} = 0$, I calculate a Wald statistic,

$$W^2 = (\mathbf{L}\hat{\boldsymbol{\beta}})^T(\mathbf{L}\widehat{Cov}(\hat{\boldsymbol{\beta}})\mathbf{L}^T)^{-1}(\mathbf{L}\hat{\boldsymbol{\beta}})$$

Then I assume the sampling distribution is approximately χ^2 with df = # of rows of $\mathbf{L}$ to calculate p-values (as long as n is large).

For the mixed effects model, mod3, here is the sample code.

```
b_mod3 = fixef(mod3)
W_mod3 = vcov(mod3)

L = matrix(c(0,0,0,0,0,0,1), nrow=1)

L%*%b_mod3
```

```
##              [,1]
## [1,] 0.01008806
```

```
(se = sqrt(diag(L%*%W_mod3%*%t(L)))) ##Robust SE for Lb
```

```
## [1] 0.001789866
```

```
## 95% Confidence Interval (using Asymptotic Normality)
L%*%b_mod3 - 1.96*se
```

```
##               [,1]
## [1,] 0.006579918
```

```
L%*%b_mod3 + 1.96*se
```

```
##              [,1]
## [1,] 0.01359619
```

```
## Hypothesis Testing
w2 <- as.numeric( t(L%*%b_mod3) %*% solve(L %*% W_mod3 %*% t(L))%*% (L%*%b_mod3))
## should be approximately chi squared

1 - pchisq(w2, df = nrow(L)) #p-value
```

```
## [1] 1.738366e-08
```

The confidence interval, [0.00658, 0.0136] does not contain a 0, meaning that the statistic is significantly different from 0 at the 0.05 level.

Also, the p-value is lower than 0.05. This means I reject the null hypothesis that multiple slopes are zero or a linear combination of slopes is zero. Therefore, the relationship between the response and independent variables are significant in model 3.

Information Criteria for Choosing Fixed Effects

```
BIC(mod1)
```

```
## [1] 4969.465
```

```
BIC(mod2)
```

```
## [1] 5377.958
```

```
BIC(mod3)
```

```
## [1] 5375.323
```

```
BIC(mod4)
```

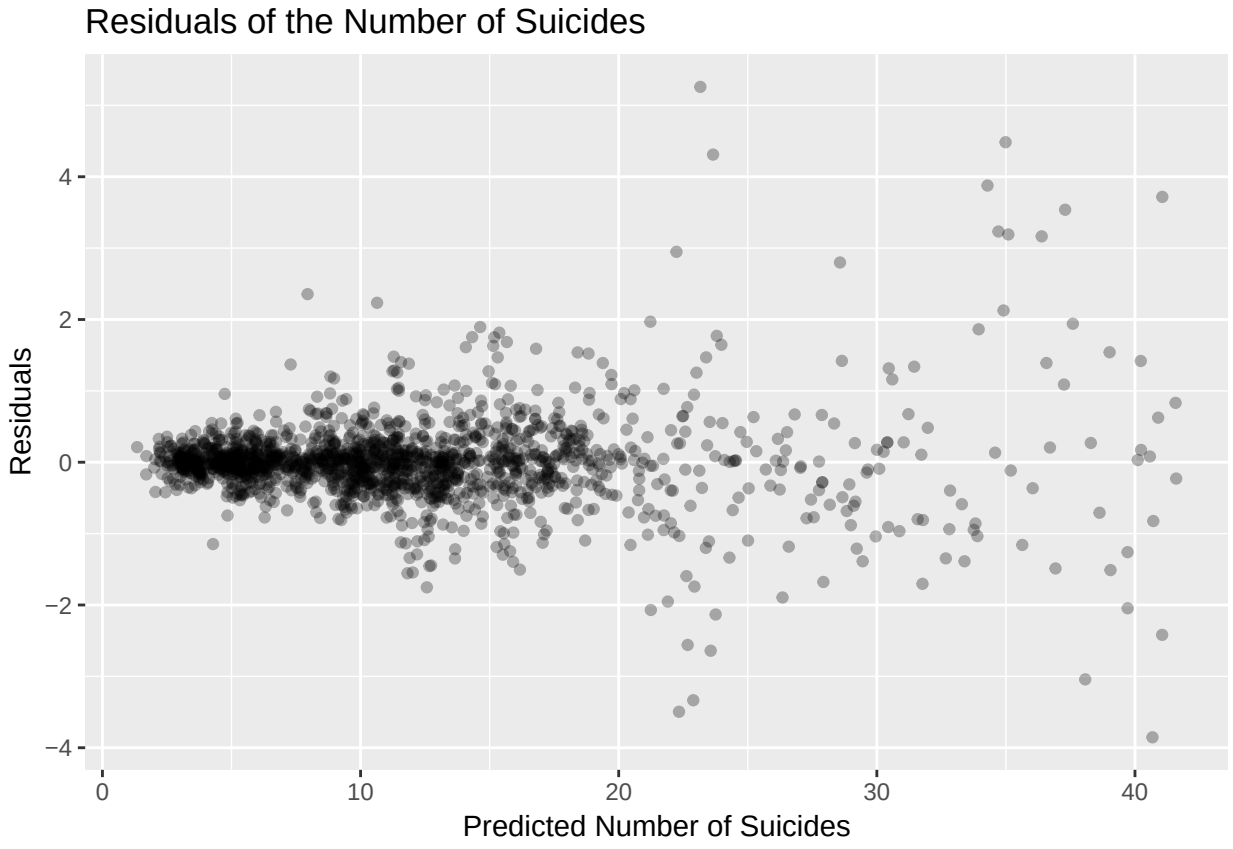
```
## [1] 5382.718
```

From the BIC output of the four models, I can see that model 1 has the lowest BIC (4969). However, model 1 was just used to estimate the trend of variables in a high level. Comparing the BIC values among models 2, 3, and 4, I find that model 3 has the lowest value. BIC uses the penalty term of the number of parameters to prevent overfitting. This makes model 3 the best model.

Step 4. Diagnostics

Residual Plot

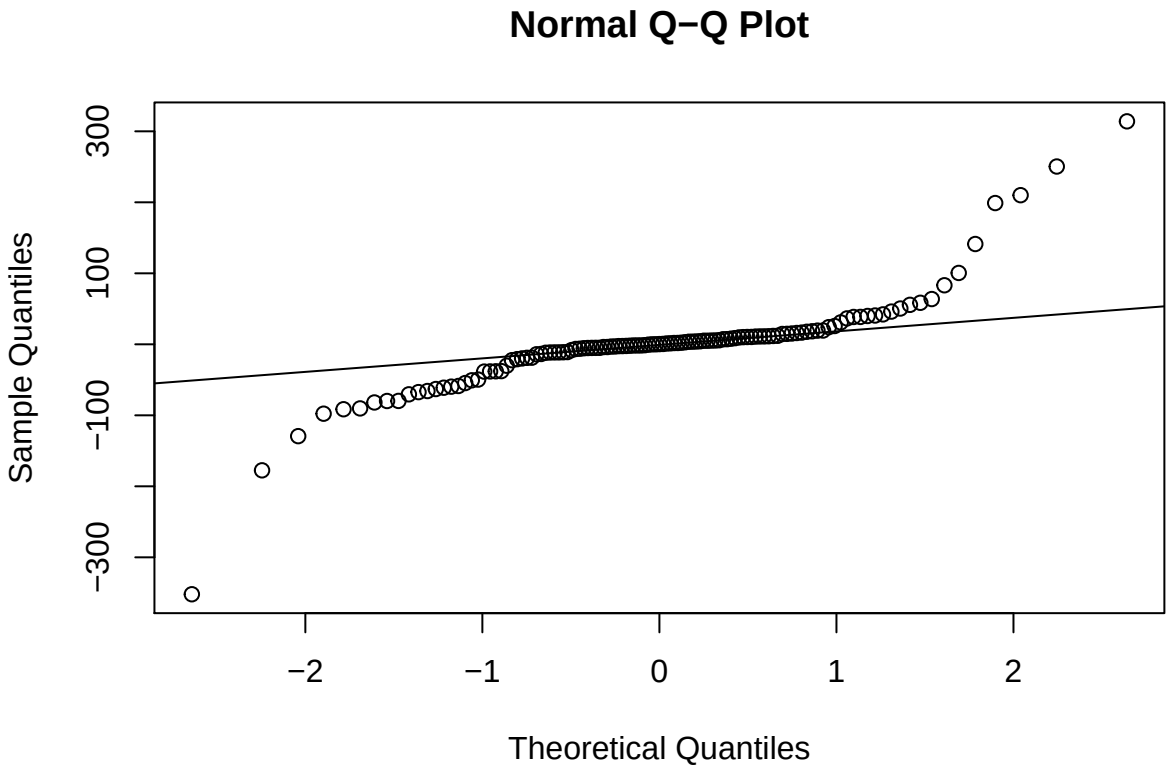
I check the residual plot and find out that all the residuals are scattered closely around zero. This shows that there are no systematic over or under estimations.



Q-Q Plot

In order to see if the random effects distribution is Normal, I also check the Q-Q plot. Even though it is not perfectly normally distributed, I still find a large proportion of my observations are normally distributed except for some boundary values. Thus, I state that my model does not have systematic over or under predictions.

```
A = ranef(mod3)$country$`(Intercept)`
qqnorm(A)
qqline(A)
```



Extreme Random Intercept Estimates

Checking the distribution of the predicted random effects and arranging by decreasing intercept, I find the countries with the highest and lowest number of suicides compared to the majority of the countries in the middle of the normal distribution bell curve. Thailand, Japan, Estonia, Sri Lanka, and Ukraine all have significantly higher intercepts, or number of average suicides, compared to the rest of the world. South Africa, Kazakhstan, Philippines, Bulgaria, and Denmark have significantly lower number of average suicides compared to the rest of the countries. These countries lie on the tail of the normal bell curve of the distribution of predicted random effects.

```
RE = ranef(mod3)$country

RE %>%
  mutate(countryname = row.names(RE)) %>%
  arrange(desc(`(Intercept)`)) # countries with the most to least suicide rates
```

```
##       (Intercept)           BMI      disorder     BirthRate
## 1    314.02845323 -7.983564e-01 -2.680273e+01  1.8925683109
## 2    250.43020583  7.510258e-01 -2.262163e+01 -0.1367786844
## 3    210.04528226 -2.772332e+00 -3.193403e+00 -1.1319136077
## 4    198.96649512 -3.152623e+00 -1.829549e+00 -5.9596992692
## 5    141.26615487  1.283242e-01 -8.992676e+00 -2.1581487206
## 6    100.50940796  5.623301e-01 -7.335363e+00 -1.2782292989
## 7     83.15349203 -6.226375e-02 -7.917027e+00  0.3521841340
## 8     63.73961840 -1.126851e+00  3.350997e+00 -1.2129056333
## 9     58.50376897 -4.136607e-02 -3.866268e+00  0.2880504667
## 10    55.59570824  4.174110e-01 -4.610037e+00  0.0499632224
## 11    50.40883851  2.257161e-01 -4.326809e+00 -0.2174110153
## 12    46.09939653  1.014374e-01 -2.998898e+00 -0.4093147741
## 13    42.16762737 -3.781355e-02 -2.992196e+00 -0.4041219849
```

```
## 14    40.59424918 -7.538592e-01  2.008330e+00 -0.9916608821
## 15    39.91696859  3.095809e-01 -5.471470e+00  0.0432240709
## 16    38.59125624  1.049758e-02 -2.353364e+00  0.0284587625
## 17    38.52998683 -1.514695e-01 -2.409019e+00 -0.2030572961
## 18    36.46751227 -2.160171e-01 -1.341912e+00  0.0999690124
## 19    30.90731893  1.968869e-01 -3.548165e+00  0.0567807841
## 20    25.76531800 -1.556974e-01 -8.828970e-01 -0.6280548912
## 21    24.15483574 -2.735540e-01 -4.278557e-01  0.0478768550
## 22    19.45093223  3.186376e-01 -2.922897e+00 -0.0024145329
## 23    19.29377140 -3.396380e-02 -1.123578e+00 -0.2940018908
## 24    18.38181397  2.408107e-01 -2.426505e+00  0.0682733623
## 25    17.67729876  5.247457e-01 -2.635477e+00 -0.5828477672
## 26    16.51894298  9.148220e-02 -1.513206e+00  0.1782263651
## 27    15.92537734  2.012911e-01 -2.508787e+00  0.1244114213
## 28    15.22179987  5.990131e-01 -4.657303e+00  0.3092216641
## 29    14.75619874  4.386051e-01 -3.609175e+00  0.3371654881
## 30    14.56786591  3.360128e-01 -2.707802e+00 -0.2459518855
## 31    12.04691612 -4.931765e-01  8.934693e-01 -0.0087099874
## 32    11.75333167  1.835100e-01 -2.923780e+00  0.5370795792
## 33    11.34707366  4.655929e-02 -1.097914e+00 -0.3130594747
## 34    11.17330297  5.144410e-02 -1.108170e+00 -0.1699163273
## 35    11.08056292  8.336100e-02 -1.191476e+00 -0.0541705793
## 36    10.91242513  2.341895e-01 -2.161132e+00  0.0367851647
## 37    10.36752575 -7.044811e-03 -8.780665e-01 -0.0200449700
## 38    10.29021390  3.505061e-02 -9.798063e-01 -0.1870163073
## 39    10.10770295  5.770146e-02 -2.291041e+00  0.1281058276
## 40     9.70130603  1.217427e-01 -1.401515e+00  0.0017303307
## 41     8.70614176  1.606640e-01 -1.237646e+00  0.0367872590
## 42     7.94801985  3.441800e-01 -2.129786e+00  0.9076184627
## 43     7.26039352  9.659717e-02 -1.084645e+00 -0.0228535870
## 44     7.15284034 -1.654201e-02 -5.726745e-01 -0.0285179209
## 45     5.75252182  1.142520e-02 -4.850516e-01 -0.2051825590
## 46     5.43040214  3.535315e-01 -2.138036e+00  0.1911473035
## 47     5.27613999  2.759200e-02 -5.500294e-01  0.1139670096
## 48     5.04991727 -5.324932e-03 -4.210891e-01 -0.1788171293
## 49     5.04396639  7.312828e-02 -8.000577e-01 -0.1082828954
## 50     4.28025109  3.877065e-02 -4.411983e-01 -0.2406740743
## 51     4.00448920  3.188455e-01 -1.755004e+00  0.1329245103
## 52     3.73335028  2.179112e-01 -1.397366e+00  0.0296808500
## 53     3.63363539  1.726821e-02 -2.925331e-01 -0.2210353360
## 54     2.92472875  6.042007e-02 -1.918564e+00  1.0669501858
## 55     2.27794502  8.448112e-03 -2.211243e-01 -0.0268258724
## 56     1.89852982  1.866477e-01 -7.824853e-01 -0.1585217640
## 57     1.82045579  1.726631e-02 -2.337701e-01 -0.1131053915
## 58     1.38740899  2.068071e-02 -2.103189e-01 -0.0425566400
## 59     1.18795748  2.257681e-02 -2.076804e-01 -0.0616219499
## 60     0.53868922  3.118125e-02 -1.391470e-01 -0.0494364821
## 61     0.32443436 -1.911304e-03 -1.751690e-02  0.0192447147
## 62     0.03936573  2.800097e-03 -1.516701e-02 -0.0127233535
## 63    -0.01522775  4.717750e-06  9.538984e-04  0.0023931222
## 64    -0.46650865  2.712342e-02 -1.008392e-01 -0.1356075170
## 65    -1.21043621  3.853419e-02  7.609262e-03  0.0225962755
## 66    -1.54410798  2.018060e-01 -8.055187e-01  0.1453946508
## 67    -1.59512084  5.099515e-01 -2.463221e+00  0.2531496534
```

```
## 68    -1.61840514  2.805596e-01 -1.823862e+00  0.4178000034
## 69    -1.98160531  3.093285e-03  1.498141e-01  0.0126569814
## 70    -2.28804226  1.382092e-02  9.171166e-02 -0.0813865437
## 71    -2.46026641 -4.148679e-03  2.244207e-01 -0.1101821065
## 72    -2.53138516  3.829936e-01 -1.630077e+00 -0.0008803681
## 73    -3.03850117  2.709976e-03  2.525839e-01  0.0309446505
## 74    -3.19345200  2.153706e-01  1.083604e-01  0.1606925905
## 75    -3.82457776 -2.231979e-02  4.132784e-01  0.0754076393
## 76    -3.86351475  6.327899e-03  3.257984e-01 -0.2163867391
## 77    -4.92854802  3.512568e-01 -9.024468e-01  0.0997973528
## 78    -4.96648785  4.018786e-01 -2.339403e+00  0.6825114258
## 79    -5.05056623 -1.252193e-02  5.390927e-01 -0.1196745736
## 80    -5.20420936  2.050680e-02 -4.646931e-01  0.7895753286
## 81    -5.50485968  2.356078e-01 -9.367108e-01  0.1164880802
## 82    -6.22357518 -2.761467e-01 -4.355334e-01  0.4776649448
## 83    -6.54328476 -2.324885e-02  3.983179e-01  0.0452505380
## 84    -8.13171788  1.048529e+00 -5.075283e+00  0.8494962567
## 85   -10.90430400 -5.025455e-01  2.504553e+00  0.4107894657
## 86   -11.18636016  4.757034e-01 -7.596566e-01 -0.2025440845
## 87   -11.36673181  3.237412e-01 -5.942718e-01 -0.0904583478
## 88   -11.40697532  2.567787e-01 -2.246265e-01  0.0162367171
## 89   -11.50158483 -9.215439e-02  1.371311e+00  0.1568162432
## 90   -11.89643901  2.448454e-01 -2.837744e-01  0.0282163752
## 91   -13.61906637  4.557337e-02  9.802844e-01  0.0857738341
## 92   -14.01601952 -2.256408e-01  2.167966e+00 -0.0030403709
## 93   -18.82855143  1.205925e-01  7.836279e-01  0.0471580784
## 94   -19.03078227  2.163531e-01  4.664717e-01  0.2153627684
## 95   -19.84238296  2.409552e-01  8.757407e-02 -0.0124251472
## 96   -20.90678339  2.443636e-01  4.067408e-01 -0.0125912259
## 97   -22.34673040  2.540872e-01  5.458588e-01 -0.0394947016
## 98   -30.14982927  2.932998e-01  1.248916e+00 -0.0594606940
## 99   -37.74751006  7.705130e-01 -2.507148e+00  1.9457908478
## 100  -38.01123137  1.017242e-01  1.960841e+00  0.7942397403
## 101  -38.18880725 -3.987897e-01  6.284715e+00 -0.6288923612
## 102  -38.41415521 -8.497058e-02  3.838734e+00  0.1025357996
## 103  -49.64651294  3.552774e-01  2.152618e+00  0.4427458377
## 104  -50.67824911 -1.278144e+00  1.106918e+01  0.7136767356
## 105  -54.54977297  5.849620e-01 -5.918394e-01  1.2355379187
## 106  -58.70231760 -5.060081e-01  7.927622e+00  1.1184179629
## 107  -59.63861602  2.167342e-01  3.537941e+00 -0.1196006478
## 108  -61.12319732  1.033249e-01  3.108110e+00  0.6472506803
## 109  -62.96571589 -2.309373e-03  5.982899e+00  0.4601738376
## 110  -65.76817126  1.219974e-01  5.326565e+00  0.3311194719
## 111  -67.30852136 -1.613433e-01  7.606462e+00 -0.9757090897
## 112  -70.45297853  2.733187e-01  3.131539e+00  0.5644038550
## 113  -79.73714634 -2.901074e-01  5.766786e+00  1.2455839887
## 114  -79.81549376  4.449349e-01  2.816691e+00  0.3442278884
## 115  -81.90793869 -3.451231e-02  5.203768e+00  0.4151310187
## 116  -90.26360679  3.082372e-01  5.357622e+00  0.1636474876
## 117  -91.45696084 -5.436047e-01  8.885061e+00 -0.9169431530
## 118  -97.71013652 -6.423315e-02  9.536807e+00 -0.3013019535
## 119 -129.33920364 -5.133561e-01  1.574414e+01 -1.6287839683
## 120 -177.46962847 -1.926655e+00  2.456792e+01  1.0787166402
## 121 -352.00503067 -9.517662e-01  3.191462e+01 -0.0888194750
```

```
##                countryname
## 1                 Thailand
## 2                    Japan
## 3                  Estonia
## 4                Sri Lanka
## 5                  Ukraine
## 6                  Uruguay
## 7                Mauritius
## 8                Lithuania
## 9                   Norway
## 10               Australia
## 11                    Fiji
## 12                 Belgium
## 13                Pakistan
## 14                Slovenia
## 15                  Panama
## 16                Mongolia
## 17                   India
## 18                  Serbia
## 19                  Brazil
## 20                Portugal
## 21                  France
## 22                  Turkey
## 23             Afghanistan
## 24                   Spain
## 25                  Poland
## 26              Montenegro
## 27                   Nepal
## 28                  Kuwait
## 29                   Malta
## 30                  Mexico
## 31             Switzerland
## 32              Azerbaijan
## 33                    Peru
## 34                 Nigeria
## 35               Nicaragua
## 36                   Qatar
## 37                 Lebanon
## 38             Netherlands
## 39              Bangladesh
## 40                    Oman
## 41                 Armenia
## 42       Papua New Guinea
## 43     Dominican Republic
## 44                 Bahrain
## 45                  Angola
## 46                 Tunisia
## 47              Tajikistan
## 48                   Gabon
## 49              Madagascar
## 50                    Chad
## 51                 Algeria
## 52                   Italy
## 53                   Niger
```

```
## 54                Zimbabwe
## 55                  Uganda
## 56             New Zealand
## 57                 Liberia
## 58                   Kenya
## 59                  Malawi
## 60                   Benin
## 61                 Senegal
## 62            Burkina Faso
## 63                Cameroon
## 64            Sierra Leone
## 65    United Arab Emirates
## 66                 Albania
## 67                  Greece
## 68                Malaysia
## 69                   Sudan
## 70                   Ghana
## 71              Mauritania
## 72            Saudi Arabia
## 73                   Haiti
## 74                 Belarus
## 75                 Burundi
## 76                    Mali
## 77                 Romania
## 78                Ethiopia
## 79                  Zambia
## 80                  Sweden
## 81                Cambodia
## 82               Singapore
## 83                    Togo
## 84                 Jamaica
## 85                 Finland
## 86  Bosnia and Herzegovina
## 87                Paraguay
## 88                Honduras
## 89                  Rwanda
## 90              Seychelles
## 91                   Libya
## 92                    Iraq
## 93                 Myanmar
## 94                Botswana
## 95               Indonesia
## 96                 Morocco
## 97                  Canada
## 98                Colombia
## 99                   Chile
## 100                Austria
## 101                Namibia
## 102                 Guyana
## 103            El Salvador
## 104                 Latvia
## 105              Guatemala
## 106                Hungary
## 107    Trinidad and Tobago
```

```
## 108               Germany
## 109               Croatia
## 110               Lesotho
## 111               Ecuador
## 112               Ireland
## 113                 China
## 114                Jordan
## 115            Luxembourg
## 116                Israel
## 117               Denmark
## 118              Bulgaria
## 119           Philippines
## 120            Kazakhstan
## 121          South Africa
```

Step 5. Conclusion and Limitations

Conclusion

In order to select the best mixed effect, random slope model to explain my longitudinal data, I first fit various linear mixed effect models, then assessed the best-fitting model with information criteria and various hypothesis tests.

My final model, mod3 is shown below:

$$Suicide_{ij} = \beta_0 + \beta_1 BMI_{ij} + \beta_2 Disorder_{ij} + \beta_3 Adolescent_{ij} + \beta_4 Rural_{ij} + \beta_5 Birthrate_{ij}$$
$$+\beta_6 Inflation_{ij} + b_{ij}(BMI_{ij} + Disorder_{ij} + Birthrate_{ij}) + \epsilon_{ij}$$

I welcome any opportunity to discuss my project or statistical programming skills.

Limitations

There are a few limitations in my analysis. First, my estimates of coefficients would become biased if my model is missing important explanatory variables. Furthermore, the decision of not including any interaction and non-linear terms is made based on the scatter plots between variables in a global level. If some patterns only occur in the country level, I would miss information about interacting variables on the global level. In the future, I plan to generate visualizations in the country level and assess the need to add interactions and non-linear terms in my model.

Besides the issues pertaining to model and variable selection, I also face the challenge of lacking observations. I only obtain asymptotically unbiased parameter estimations due to my limited number of observations. The estimates of the variance of random effects are also biased with finite observations. The distribution with finite observations is only asymptotically normal instead of normal. In the future, if more data is available, I will increase my sample size as large as possible so that as $n \to \infty$, my estimates and the variance of estimates would be more unbiased, and the sampling distribution would become more normal.

References

Berk, Michael, Seetal Dodd, and Margaret Henry. 2006. "The Effect of Macroeconomic Variables on Suicide." *Psychological Medicine* 36 (2). Cambridge University Press: 181–89.

"Data." n.d. *Gapminder*. https://www.gapminder.org/data/.

"Human Development Reports." n.d. / *Human Development Reports*. http://hdr.undp.org/en/indicators/137506.

Ritchie, Hannah, Max Roser, and Esteban Ortiz-Ospina. 2015. "Suicide." *Our World in Data*. https://ourworldindata.org/suicide.

"Suicide." n.d. *World Health Organization*. World Health Organization. http://www.who.int/mental_health/suicide-prevention/en/.

"World Development Indicators." n.d. *DataBank*. http://databank.worldbank.org/data/source/world-development-indicators#.