final_project

Young Kyung Kim

11/11/2019

Summary

The research question is if there is an inverted-U shaped relationship between GDP and suicide rate for developed nations. I used a hierarchical model controlling for nations level for random intercept and a random slope to evaluate whether there is an inverted-U shaped relationship between GDP and suicide rate. Only about 67% of the developed nations, which are top half in GDP among OECD nations, had an inverted-U shaped relationship.

Introduction

Suicide is a significant problem for nations. In an economic perspective, suicide is the loss of labour. Furthermore, suicide is the tip of the iceberg of mental health issues. There are a lot of researches about how mental health issues impact the efficiency of workers. In this case, suicide and mental health might be an obstacle for nations from developing further. Despite such importance, suicide hasn't been studied as much, especially in economics.

This research explores the relationship between economic development and the suicide rate. One may assume that the suicide rate will decrease as the economy develops, especially for developing nations. For nations struggling with starvations, more output will not only save people's lives but also improve well-being. Yet, for a certain stage of economic development, development doesn't seem to enhance well-being as the suicide rate increase. Looking at some of nations' history of suicide rate and GDP per capita, they seemed to have an inverted-U shape.

My hypothesis is that from rapid economic development, a lot of social issues arises such as inequality in income or workplace. For example, the industrial revolution did bring wealth to a nation, but it generated a lot of social issues. Due to those social issues, well being will decrease, and the suicide rate will increase till the government start to notice the seriousness of the problem. As the government implement a lot of policy to solve social and suicide issues, the suicide rate will start to decrease. These trends will make inverted-U shape relationship between economic development and suicide rate. In this research, I am going to explore whether OECD nations or developed nations that went to rapid development have an inverted-U shape relationship in the suicide rate and GDP per capita.

Data

I used dataset from OECD, which provides data on various indicators as health and economic for each nations that are part of the members.

The dictionary of features included in this study are as follows:

Table 1: Data Dictionary

Variable	Description
GDP	Gross Domestic Product
Lifeexp	Life expectancy
country	country
suiciderate	number of suicide per 100,000
year	year
healthspending	The final consumption of health care goods and services
fertility	fertility
unemployment	the number of unemployed people as a percentage of the labour force

In the data cleaning process, I found GINI, education, and low-pay variable to have a lot of missing data, which was more than half of the data sets. I excluded those variables from dataset. I also excluded the population variable as the information is included in the suicide rate variable.

Health spending and unemployment had 602 and 860 missing data for each. I split the dataset into two. One including health spending and unemployment variables, which has 882 observations. One excluding health spending and unemployment variables, which has 1716 observations.

We started EDA by looking at the distribution of the response variable, which is a suicide rate and found that they are not normally distributed. The response variable seems to have rightly-skewed for both data set; therefore, I applied square root transformation, and the result is shown in the Appendix A1.

Then, I made the scatterplot and linear smooth curve to see if there is an inverted-U shape relationship between suicide rate and economic development, which is shown in the Appendix A2. Each point in the scatterplot represents suicide and GDP for a year. The graph is divided into four by each nations GDP from highest to lowest. Interestingly the developed nations seemed to have inverted-U shape relationship. However, less-developed nations don't seem to have an inverted-U shape relationship. Half of them have decreasing trends, while the other half doesn't have any trend at all. From the scatterplots, economic development may have a different effect on the suicide rate by various nations, so the hierarchical model seemed adequate for usage. I will have a random slope and intercepts by countries on GDP.

Model

Before building the model, I standardized all the variables except the suicide rate, which has been already transformed using a square root. I build two hierarchical models. One with dataset excluding health spending and unemployment variables. Another one with the dataset, including health spending and unemployment variables.

The final model parameter for dataset excluding health spending and unemployment variables are below with AIC value.

term	estimate	std.error	statistic	conf.low	conf.high	group
(Intercept)	3.60085	0.16288	22.10741	3.28161	3.92008	fixed
poly(GDP, 2)1	-3.45276	2.40563	-1.43529	-8.16770	1.26217	fixed
poly(GDP, 2)2	1.92370	4.15878	0.46256	-6.22737	10.07476	fixed
year	0.02518	0.04453	0.56546	-0.06209	0.11245	fixed
Life_exp	-0.20929	0.03506	-5.96971	-0.27800	-0.14058	fixed
fertility	-0.20579	0.01905	-10.80442	-0.24312	-0.16846	fixed
sd_(Intercept).country	1.03602	NA	NA	NA	NA	country
sd_poly(GDP, 2)1.country	12.66020	NA	NA	NA	NA	country
sd_poly(GDP, 2)2.country	24.66371	NA	NA	NA	NA	country

term	estimate	std.error	statistic	conf.low	conf.high	group
cor_(Intercept).poly(GDP, 2)1.country	-0.83324	NA	NA	NA	NA	country
cor_(Intercept).poly(GDP, 2)2.country	0.23419	NA	NA	NA	NA	country
cor_poly(GDP, 2)1.poly(GDP, 2)2.country	-0.22731	NA	NA	NA	NA	country
sd_Observation.Residual	0.23878	NA	NA	NA	NA	Residual

```
## [1] 476.8867
```

The quadratic term in GDP is not significant at 0.05 significance level for the model.

The final model parameter for dataset including health spending and unemployment variables are below with AIC value

term	estimate	std.error	statistic	conf.low	conf.high	group
(Intercept)	3.74999	0.13749	27.27457	3.48051	4.01946	fixed
poly(GDP, 2)1	-2.25079	2.02190	-1.11320	-6.21364	1.71207	fixed
poly(GDP, 2)2	2.16350	2.84324	0.76093	-3.40916	7.73615	fixed
year	-0.15144	0.06514	-2.32477	-0.27911	-0.02376	fixed
Life_exp	-0.01341	0.05896	-0.22752	-0.12897	0.10214	fixed
fertility	-0.09778	0.02287	-4.27540	-0.14260	-0.05295	fixed
health_spending	0.02692	0.04440	0.60618	-0.06011	0.11395	fixed
unemployment	0.06216	0.01291	4.81449	0.03685	0.08746	fixed
sd_(Intercept).country	0.80496	NA	NA	NA	NA	country
sd_poly(GDP, 2)1.country	8.63966	NA	NA	NA	NA	country
sd_poly(GDP, 2)2.country	15.05265	NA	NA	NA	NA	country
cor_(Intercept).poly(GDP, 2)1.country	-0.87776	NA	NA	NA	NA	country
cor_(Intercept).poly(GDP, 2)2.country	0.03307	NA	NA	NA	NA	country
cor_poly(GDP, 2)1.poly(GDP, 2)2.country	0.06272	NA	NA	NA	NA	country
sd_Observation.Residual	0.18586	NA	NA	NA	NA	Residual

```
## [1] -50.06934
```

The quadratic term in GDP is not significant at 0.05 significance level for the model. Both model did not have quadratic term in GDP as significant. It might be due to as the model had a lot developing nations data, which might not have inverted-U shaped relationship. I will draw the prediction of the each nations from the model to see which nations have the inverted-U shaped relationship after assessing the model.

Model Assessment

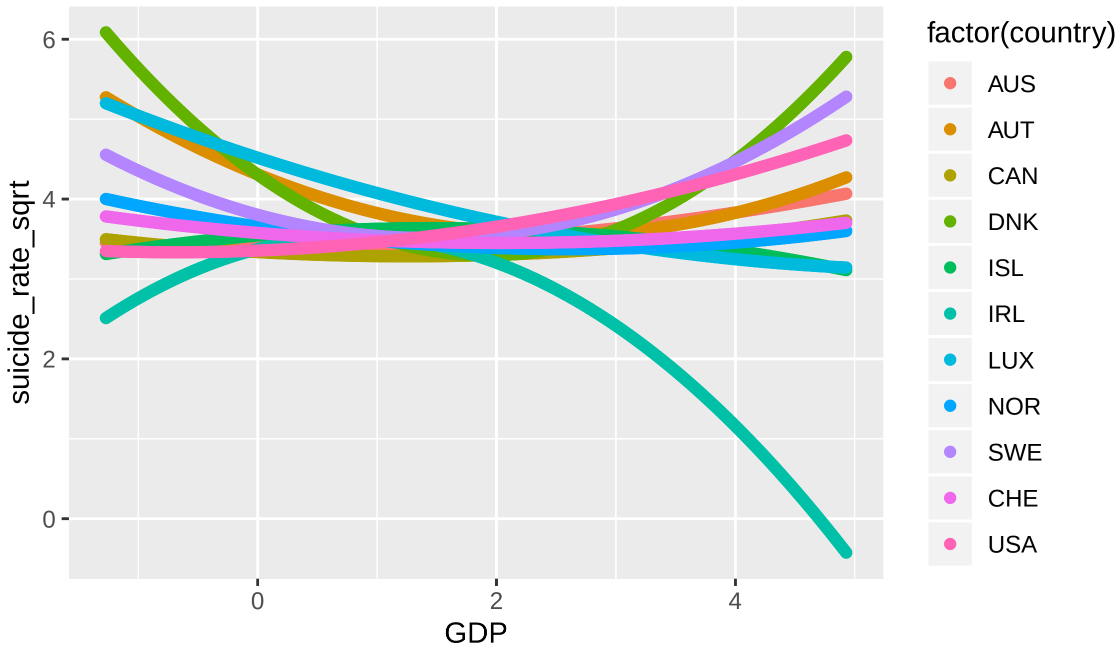
Looking at the graph, it seems like residuals have a constant variance for both models. Furthermore, the error appears to be independent of each other as it doesn't show any clear pattern in the scatterplot for both models. Looking at the VIF, multicolinearity seems to be fine. The normality assumption of both models fails from looking at the graph. Furthermore, both model seems to have autocorrelation issues. (Appendix A3)

Results

To see whether there is an inverted-U shape relationship between GDP and suicide rate, I plotted the predicted value for both models.

Below is the predicted graph for the hierarchical model made from dataset excluding health spending and unemployment variables. I chose the 11 most developed nations in the graph. Predicted graph for other countries is in the Appendix A4.

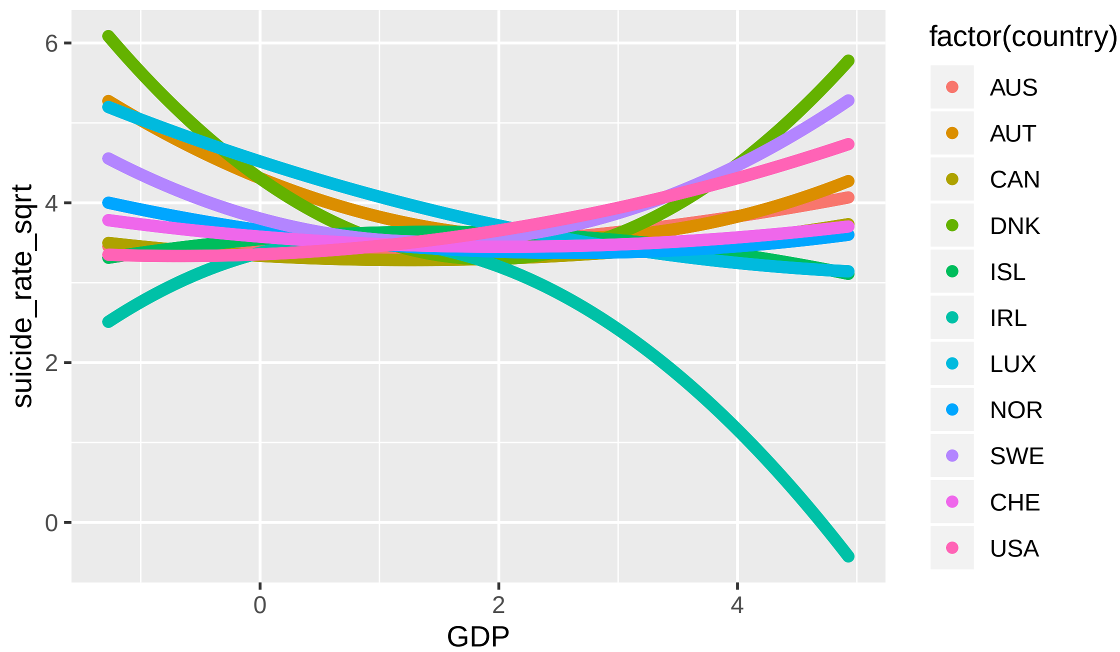
Among 11 most developed nations, Austria, Israel, Sweden and the United States of America did not have inverted-U shaped relationship. All other countries had an inverted-U shaped relationship. Among 21 most developed nations, about seven nations did not have an inverted-U shaped relationship. All other nations had an inverted-U shaped relationship.



Below is the predicted graph for the hierarchical model made from a dataset, including health spending and unemployment variables. I chose the 11 most developed nations in the graph. Predicted graph for other nations is in the Appendix A5.

Among the 11 most developed nations, Ireland and Israel had an inverted-U shaped relationship. All other nations did not have the inverted-U shaped relationship. Among 21 most developed nations, about five nations had an inverted-U shaped relationship. All other nations did not have the inverted-U shaped relationship.

Compared to the hierarchical model made from dataset excluding health spending and unemployment variables, this model had a significantly lower number of nations with an inverted-U shaped relationship. One of the hypothesis is that such phenomena happened due to deleting half of the data points to include health spending and unemployment variables. Maybe it takes a certain period for inverted-U shape relationship to happen and dataset, including health spending and unemployment variables, did not have enough data points for the relationship to form. I need to do further study to support or reject this hypothesis.



Limitation

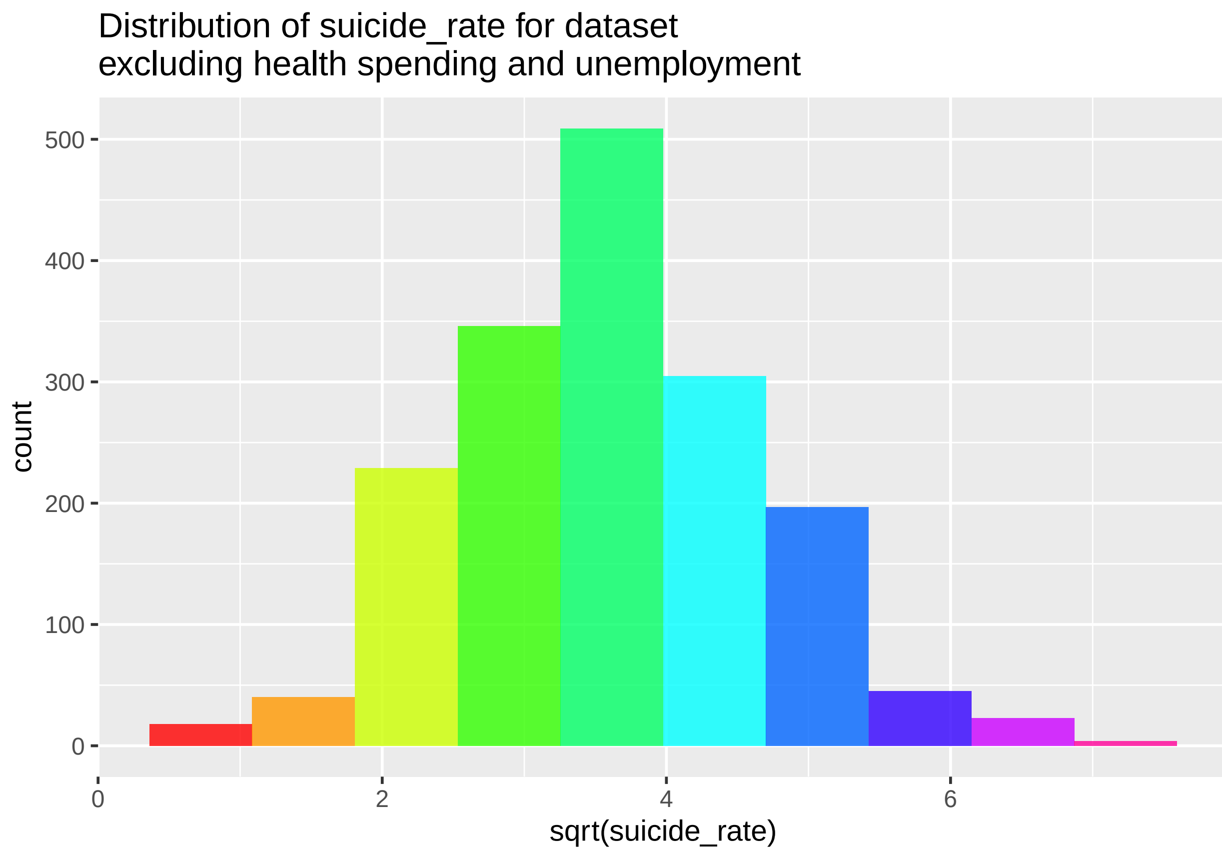
There had been some limitations from our model, especially from lack of data. Looking at the developed nations that did not have an inverted-U shaped relationship, some countries seemed to pass the peak already. For example, Britain had a decreasing trend. Britain economic development peak is around the 19th century, but we didn't have any data on economic development or suicide rate during that period. However, from reading literature at that period, one wouldn't be surprised if they had a high suicide rate as living conditions in the cities and towns were miserable, especially for workers. If there is a way for us to get or calculate the economic development or suicide rate for the 19th century, then it would be interesting if there is an inverted-U shaped relationship. There are also some limitations on whether our data on suicide is accurate enough. As suicide is considered taboo in most of the culture, it wouldn't be surprising to find out whether suicide is underreported. For example, South Korea used to have a suicide rate reported by the police and government. Police had a higher suicide rate compared to a government report. For the police report, the police decide whether certain death is suicide or not. For government report, a family member of deceased one would report whether it was suicide or not. Family member tends to falsely report that it was an accident whether than it was a suicide, as it is considered shameful for some family. Thus, there was a more than 20% difference in the suicide rate between those two reports. Such problem could arise in a lot of nations. If there is a way for us to consider the sensitivity of suicide for each culture or nations to calculate accurate suicide rate, then it might help us to identify whether there is an inverted-U shaped relationship.

Conclusion

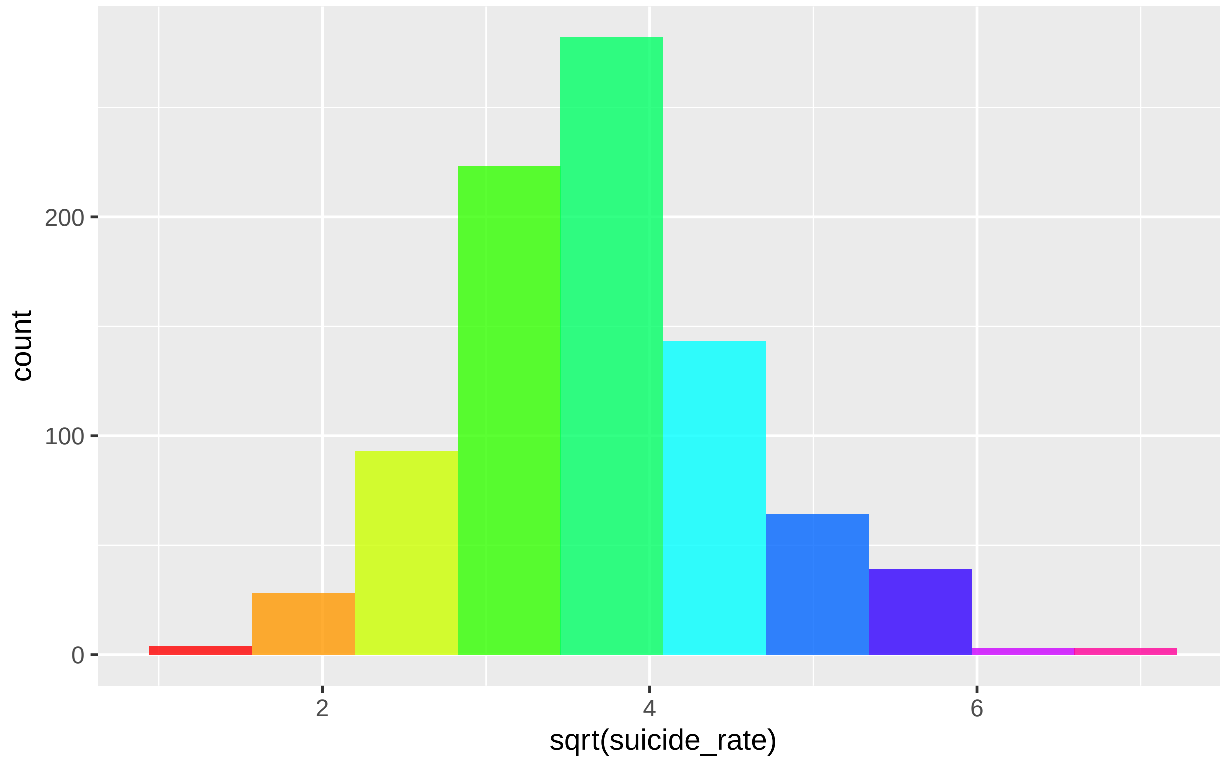
This research explored the relationship between suicide and economic development. I examined whether there is an Inverted-U shaped relationship between suicide rate and economic growth. There doesn't seem to be an inverted-U shaped relationship for all the OECD nations. When we only look at the data on most wealthy OECD nations, we did notice 67% of nations had an inverted-U shaped relationship, especially with the model built from a lot of data points. Other remaining wealthy nations did not have inverted-U shaped relationship due to the limitations described above, or the relationship might not exist in the first place. Further research is needed.

Appendix

Appendix A1

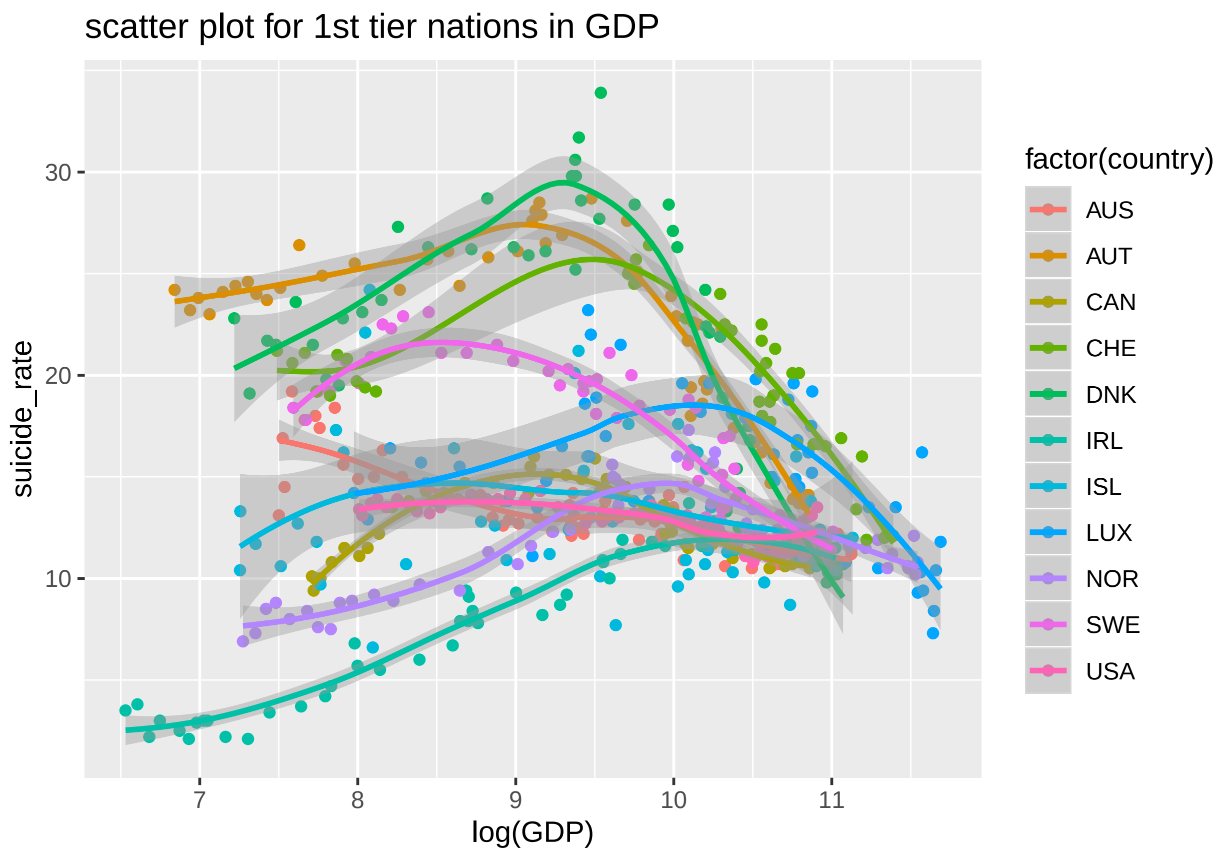


Distribution of suicide_rate for dataset
including health spending and unemployment

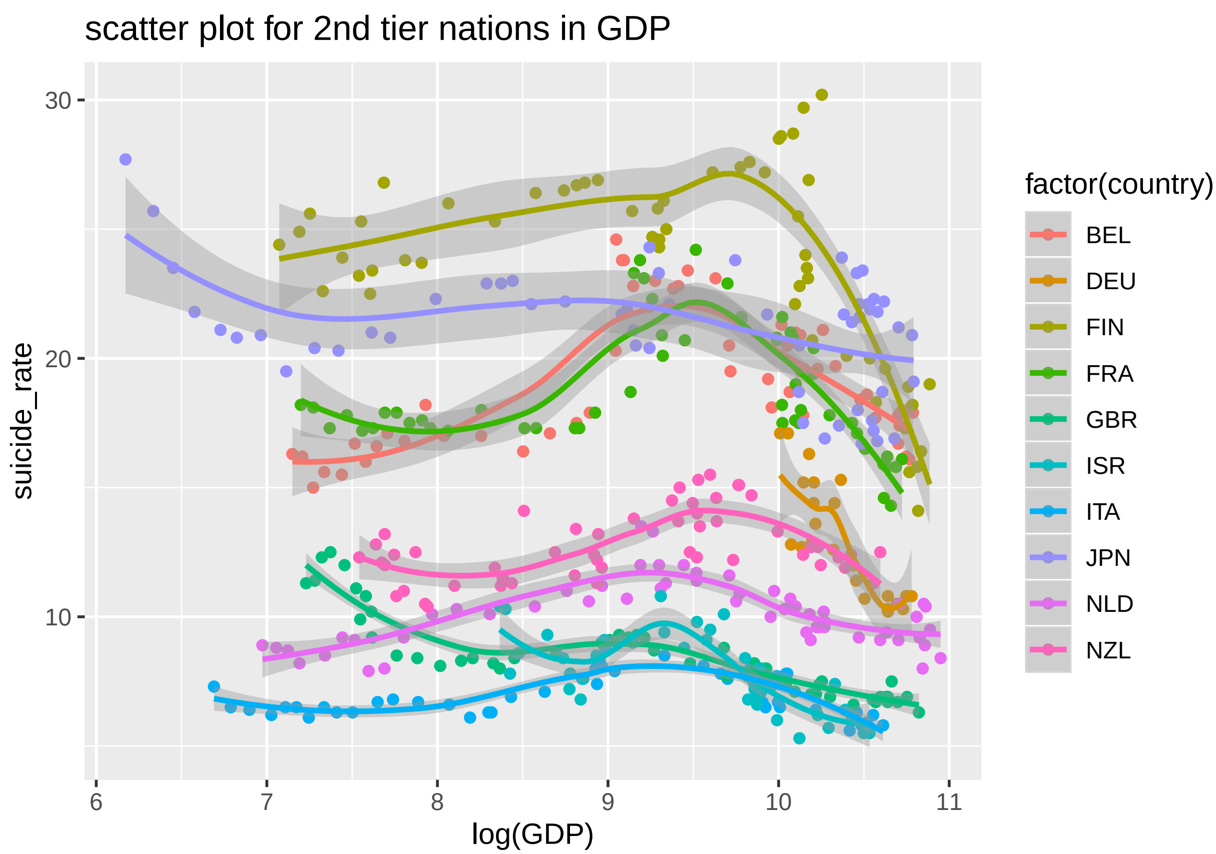


Appendix A2

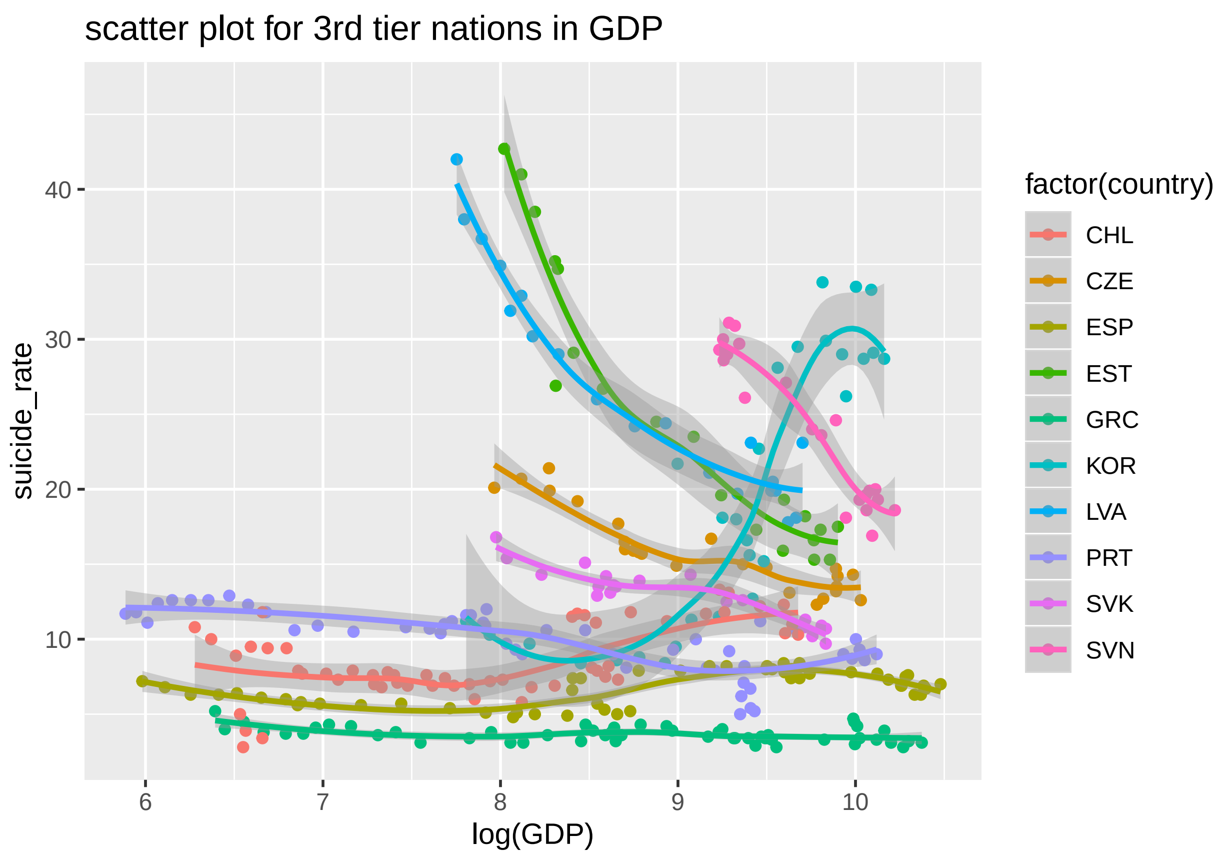
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



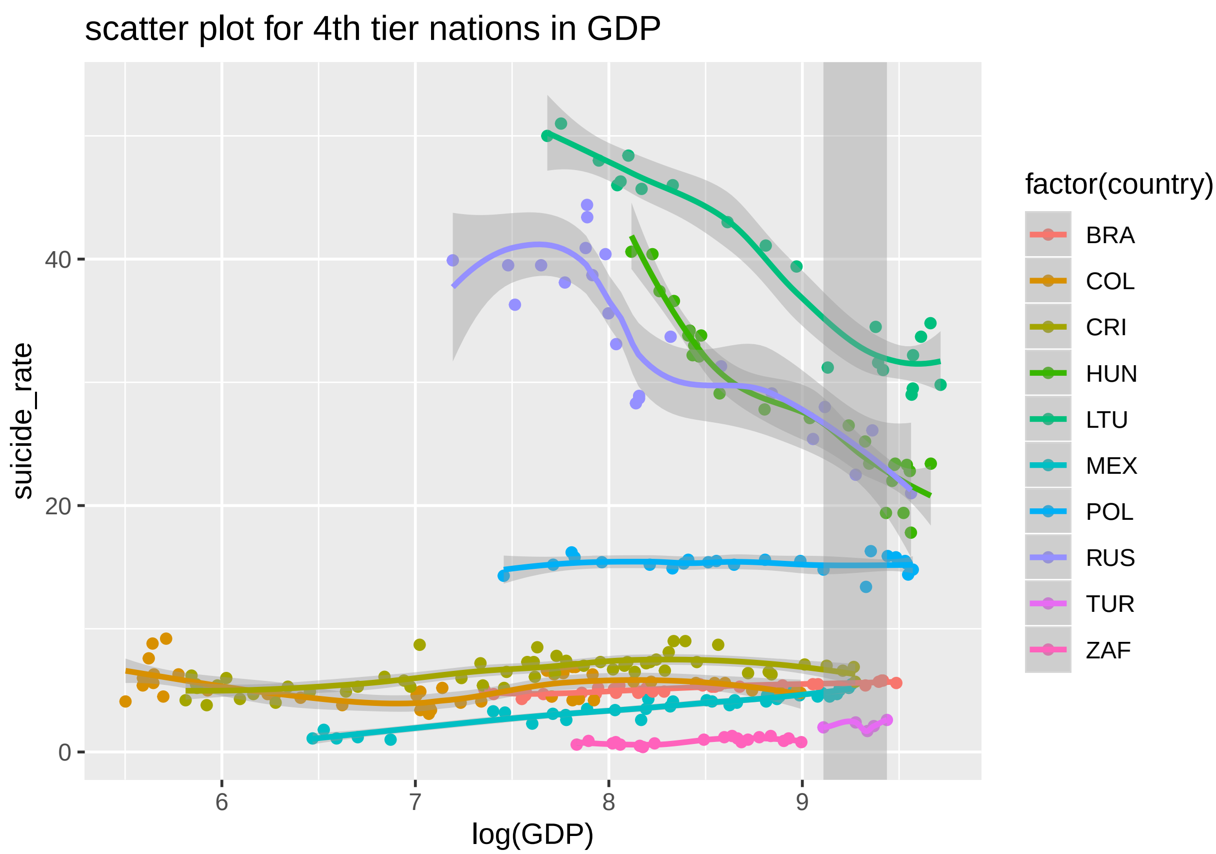
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

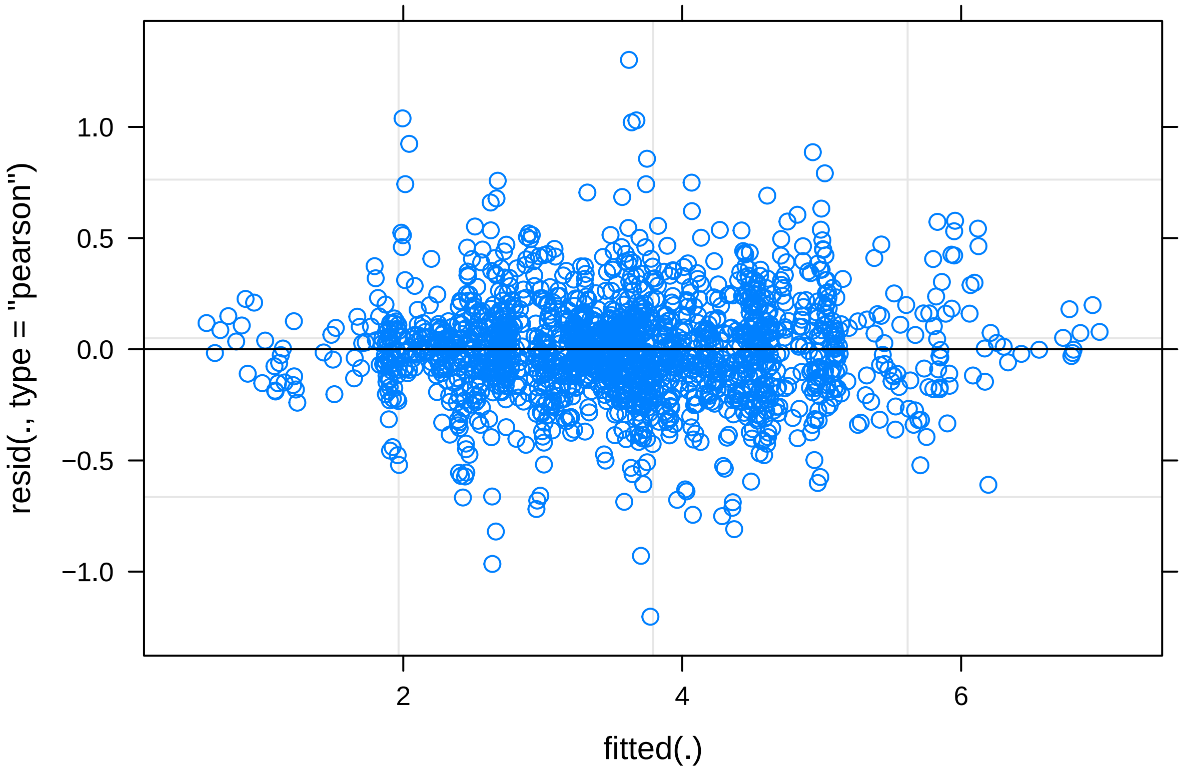


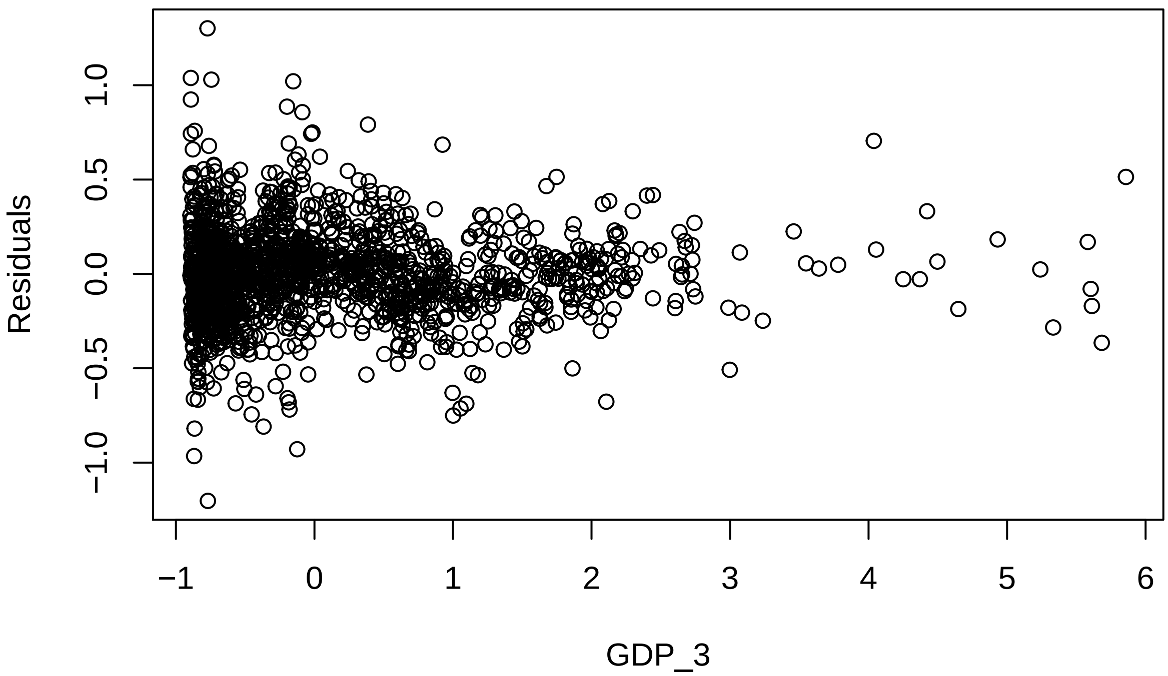
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

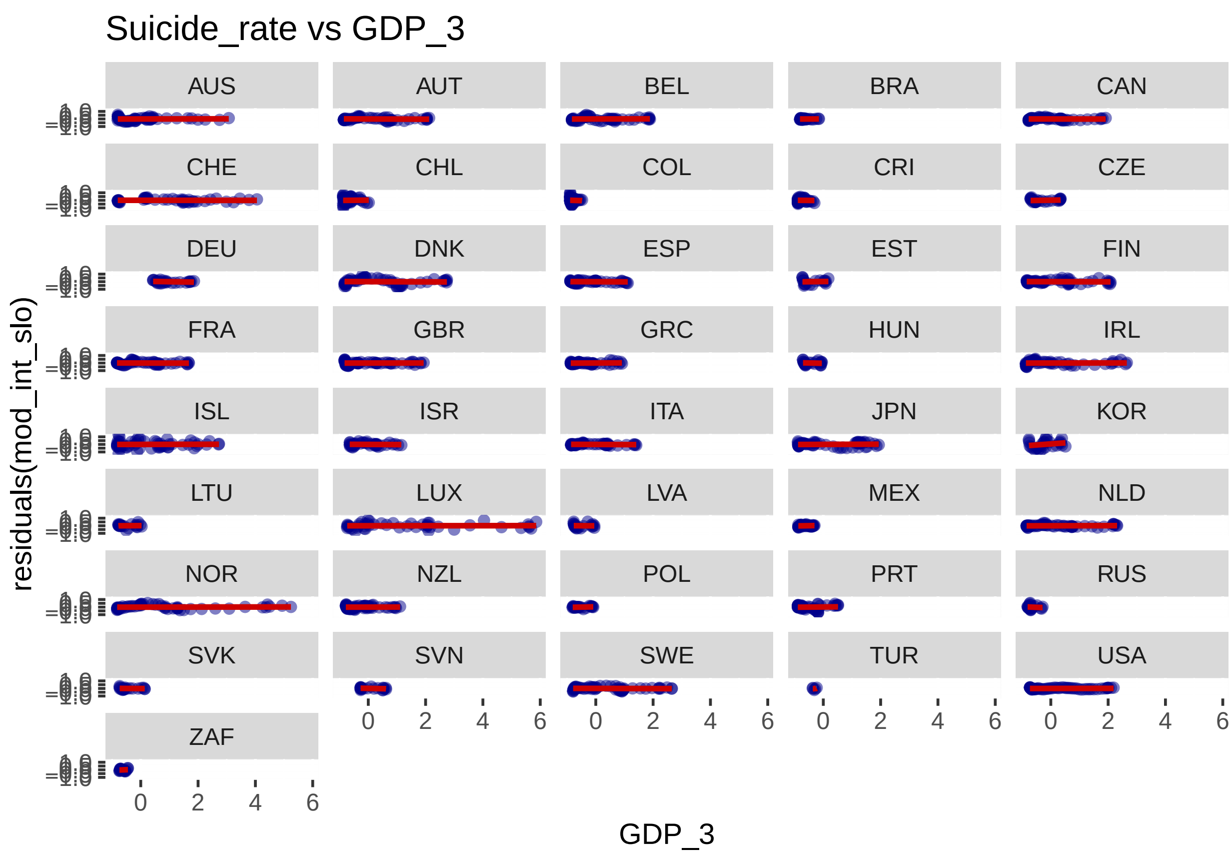


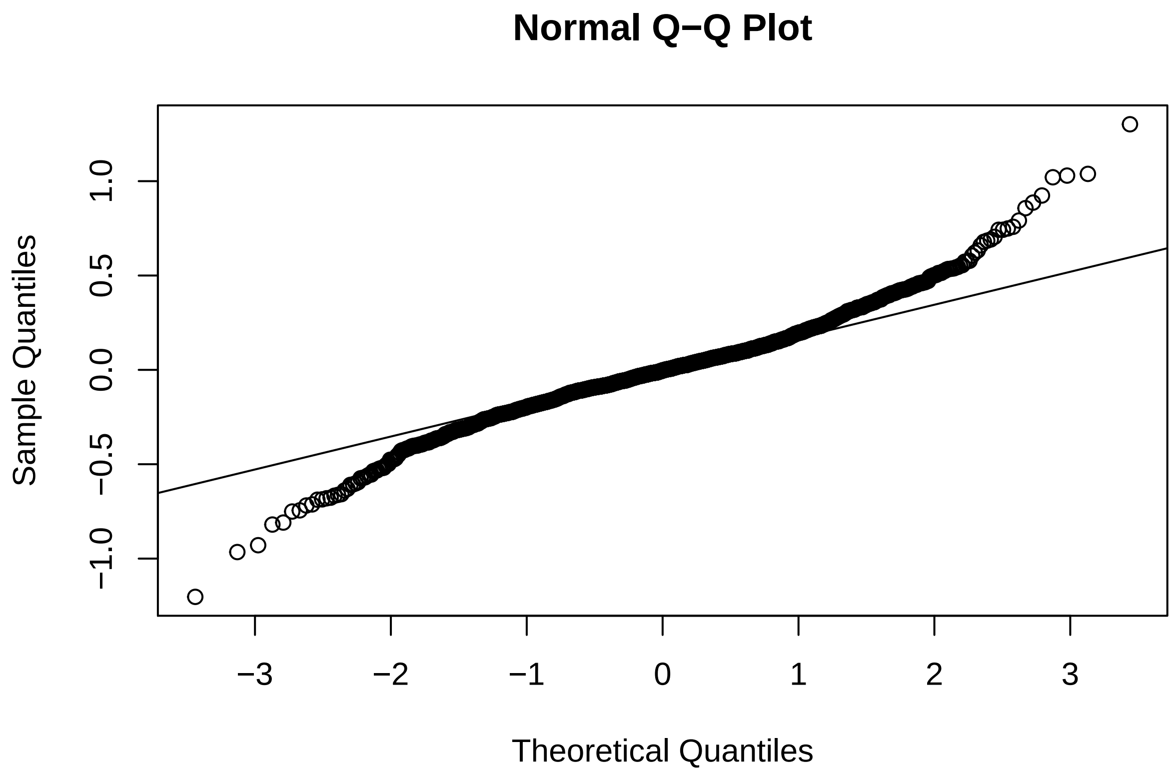
Appendix A3

Model Assessment for dataset excluding health spending and unemployment

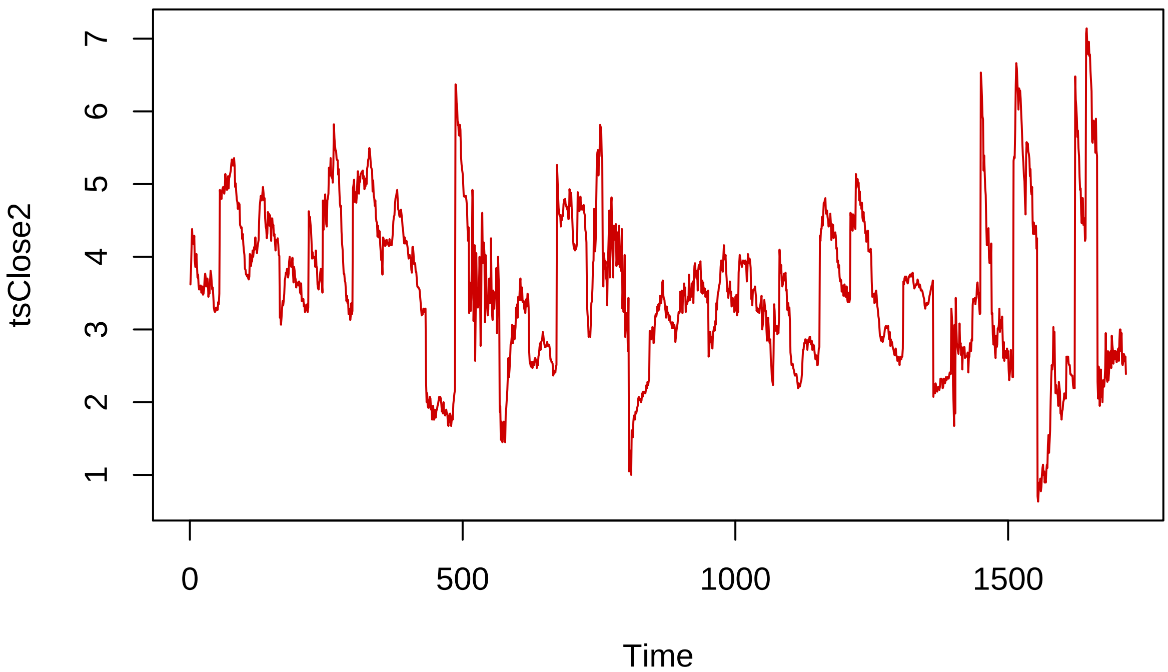



Residuals
1.0
0.5
0.0
−0.5
−1.0
−1
0
1
2
3
4
5
6
GDP_3

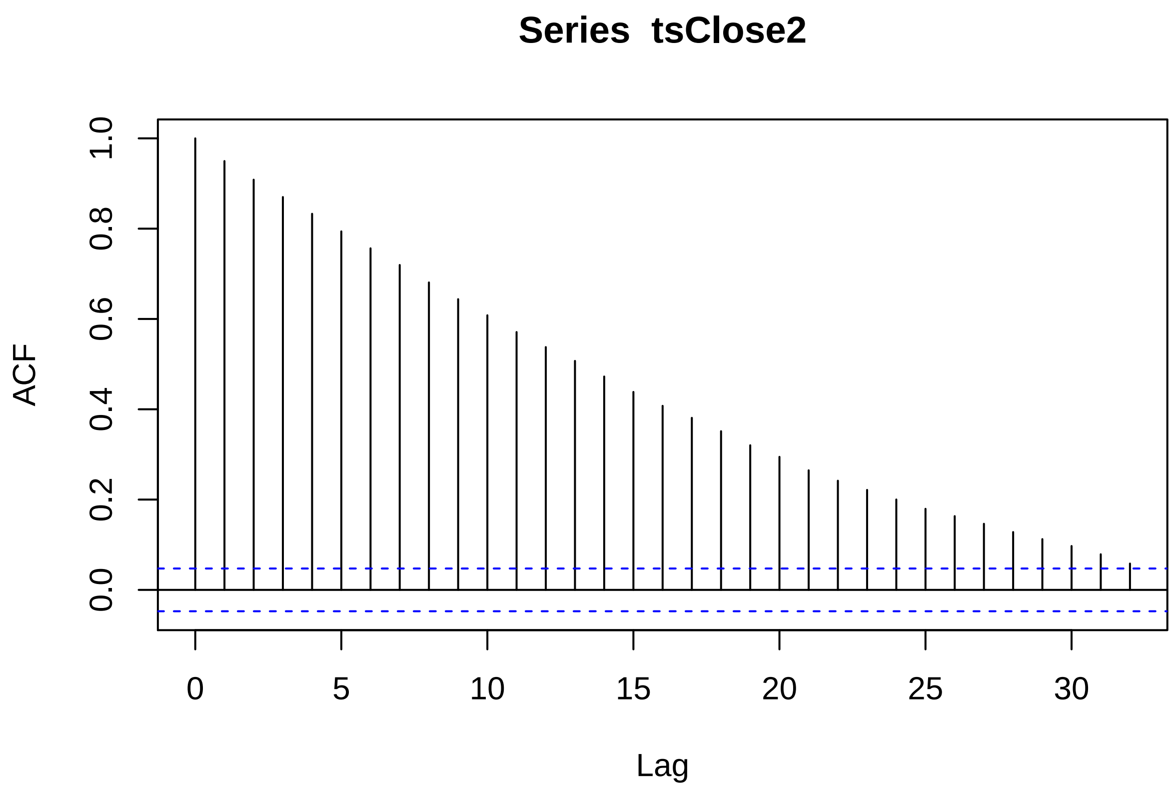

Suicide_rate vs GDP_3
AUS
AUT
BEL
BRA
CAN
CHE
CHL
COL
CRI
CZE
DEU
DNK
ESP
EST
FIN
FRA
GBR
GRC
HUN
IRL
ISL
ISR
ITA
JPN
KOR
LTU
LUX
LVA
MEX
NLD
NOR
NZL
POL
PRT
RUS
SVK
SVN
SWE
TUR
USA
ZAF
residuals(mod_int_slo)
GDP_3
0
2
4
6

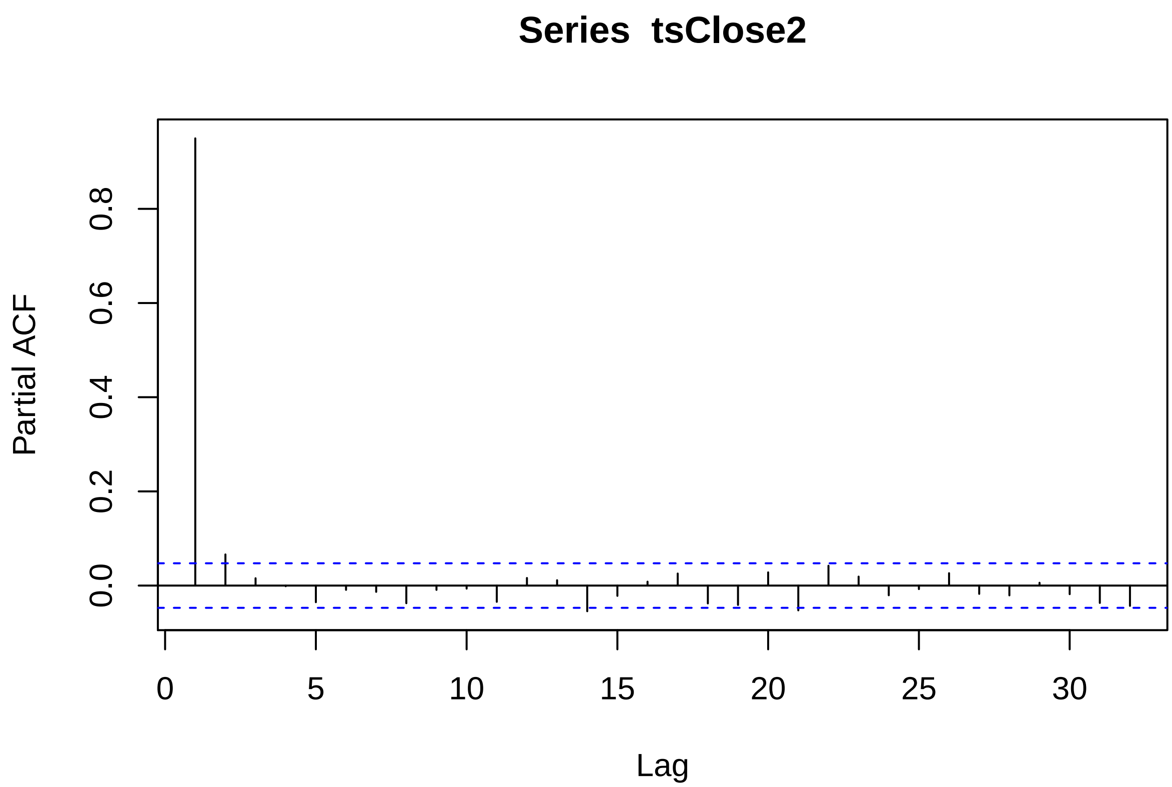



```
##                    GVIF Df GVIF^(1/(2*Df))
## poly(GDP, 2) 1.178233  2        1.041856
## year         3.936442  1        1.984047
## Life_exp     3.545867  1        1.883047
## fertility    2.248451  1        1.499483
```

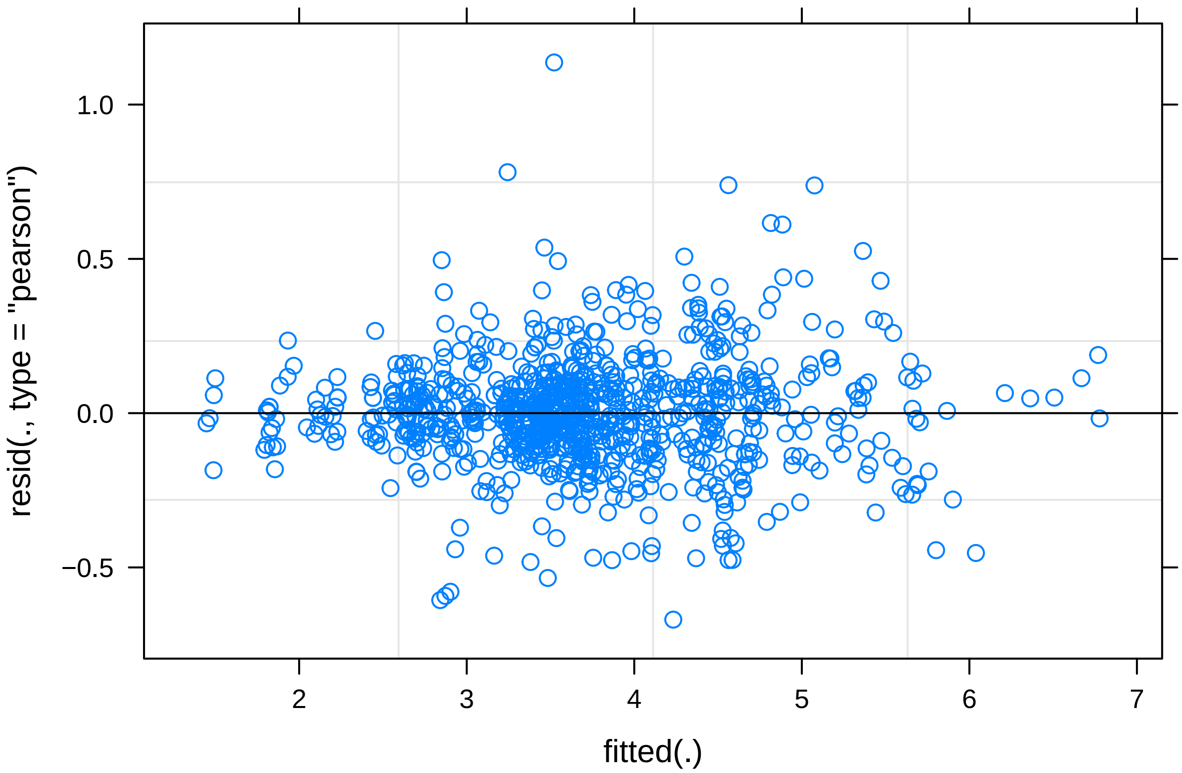


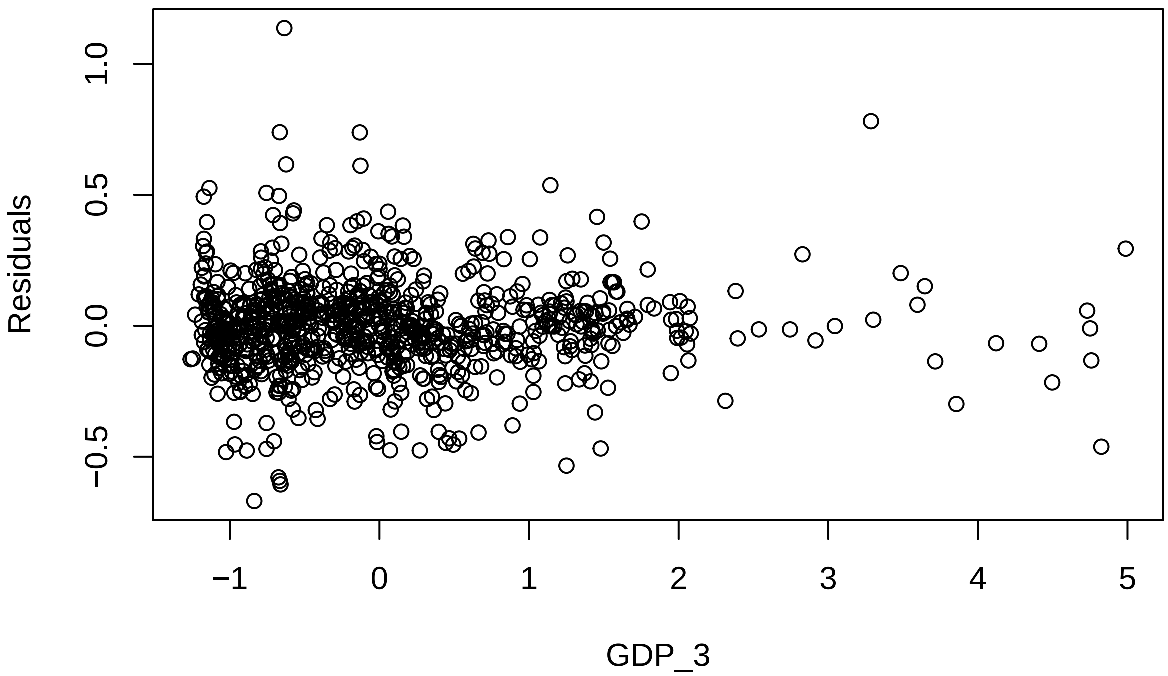
tsClose2
1
2
3
4
5
6
7
0
500
1000
1500
Time


Series tsClose2
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

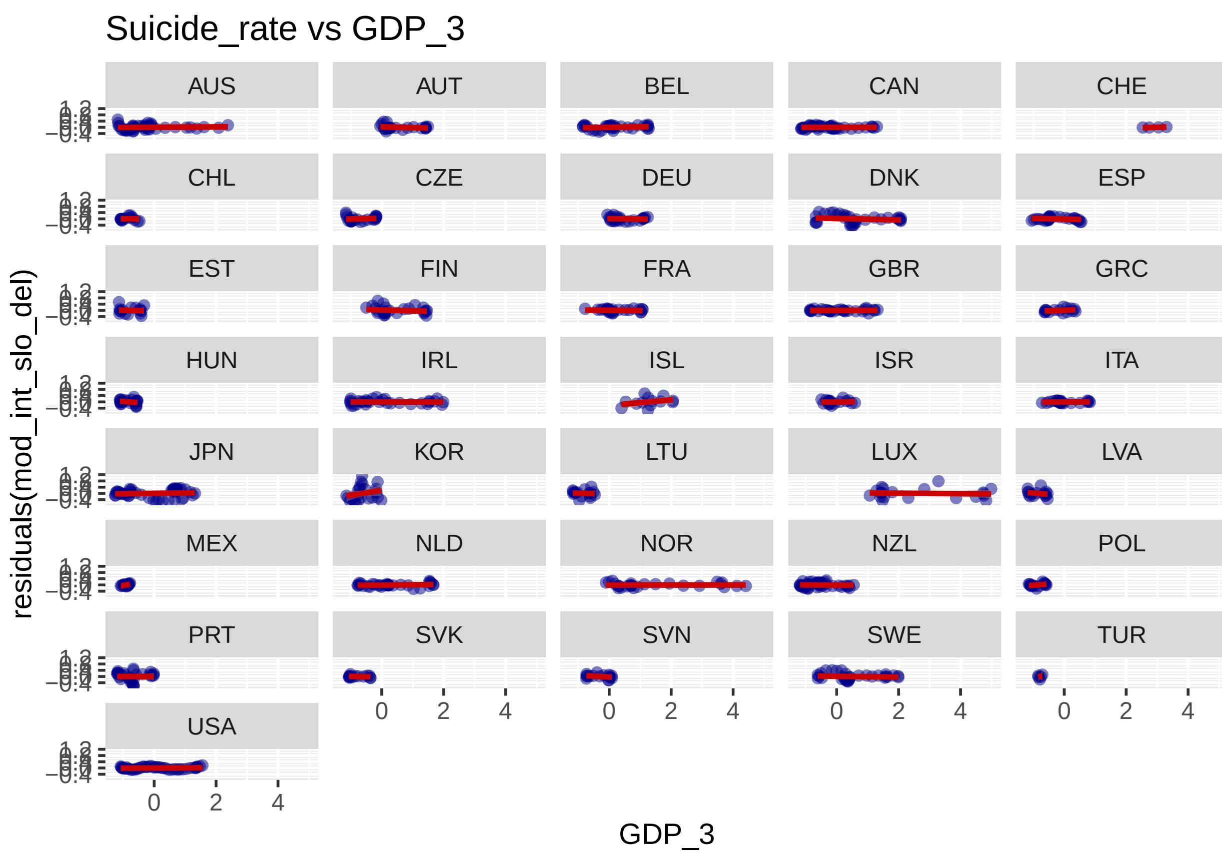

Series tsClose2
Partial ACF
0.0
0.2
0.4
0.6
0.8
0
5
10
15
20
25
30
Lag

Model Assessment for dataset including health spending and unemployment

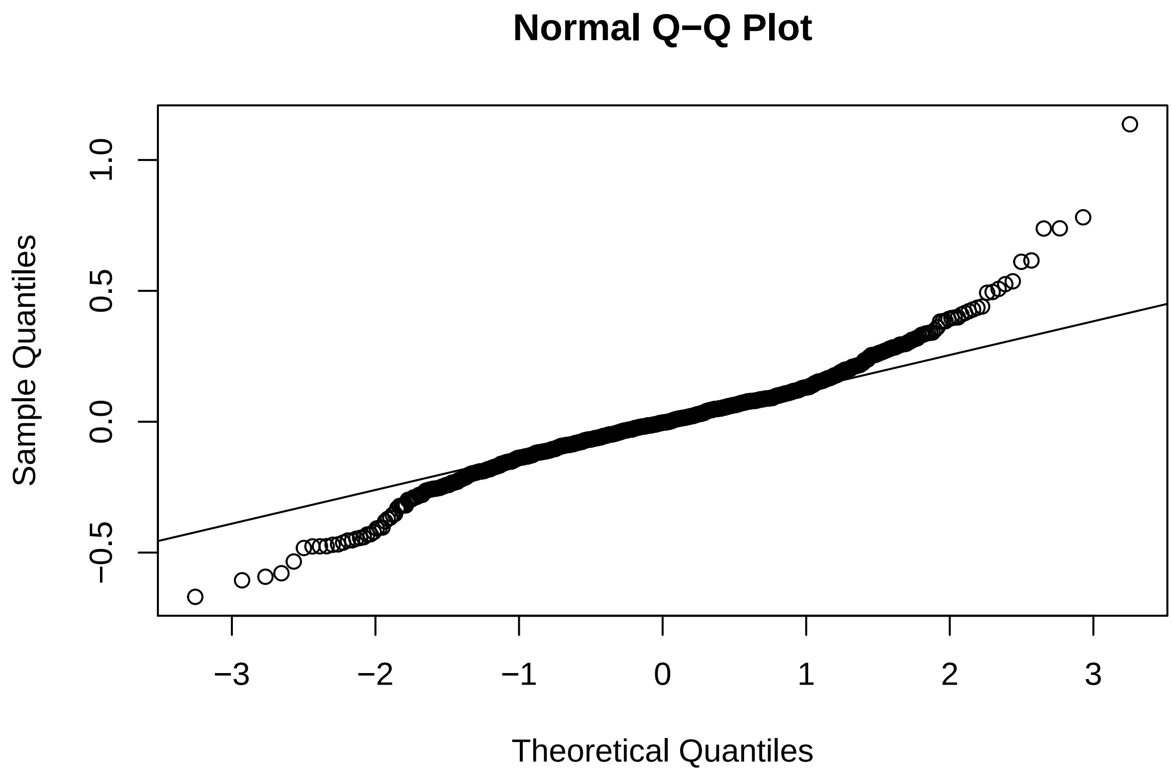




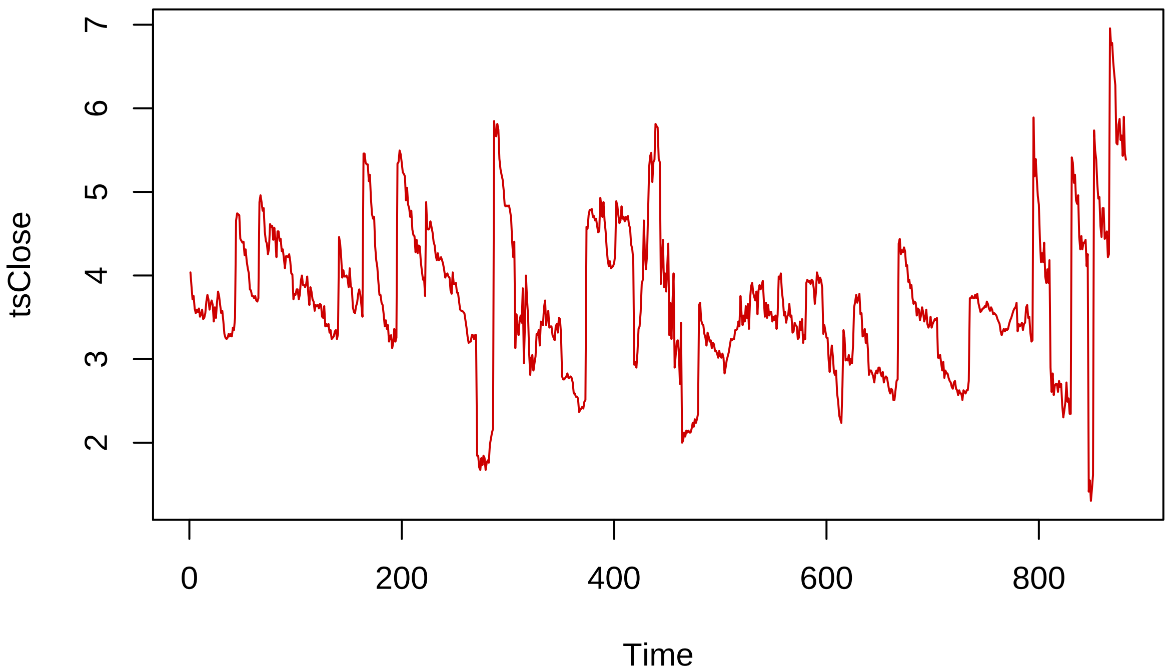
Residuals
1.0
0.5
0.0
−0.5
−1
0
1
2
3
4
5
GDP_3

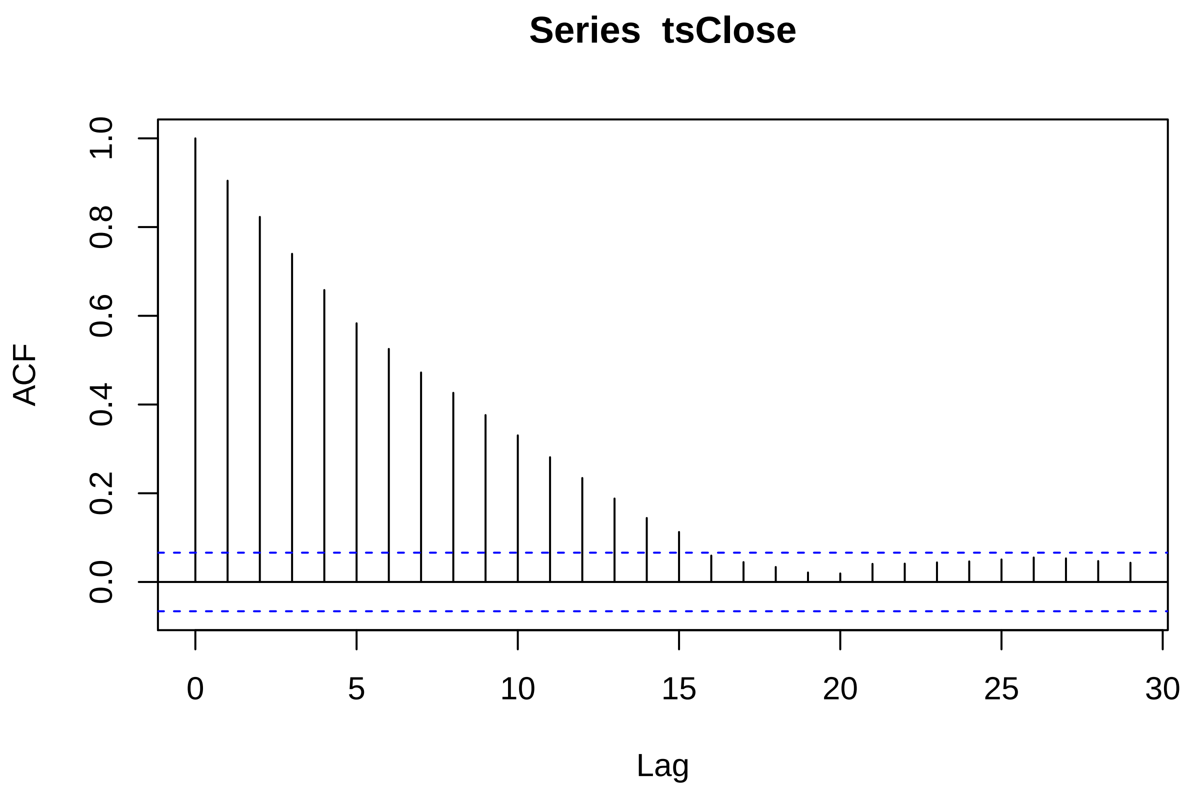


Suicide_rate vs GDP_3
AUS
AUT
BEL
CAN
CHE
CHL
CZE
DEU
DNK
ESP
EST
FIN
FRA
GBR
GRC
HUN
IRL
ISL
ISR
ITA
JPN
KOR
LTU
LUX
LVA
MEX
NLD
NOR
NZL
POL
PRT
SVK
SVN
SWE
TUR
USA
residuals(mod_int_slo_del)
GDP_3
0
2
4

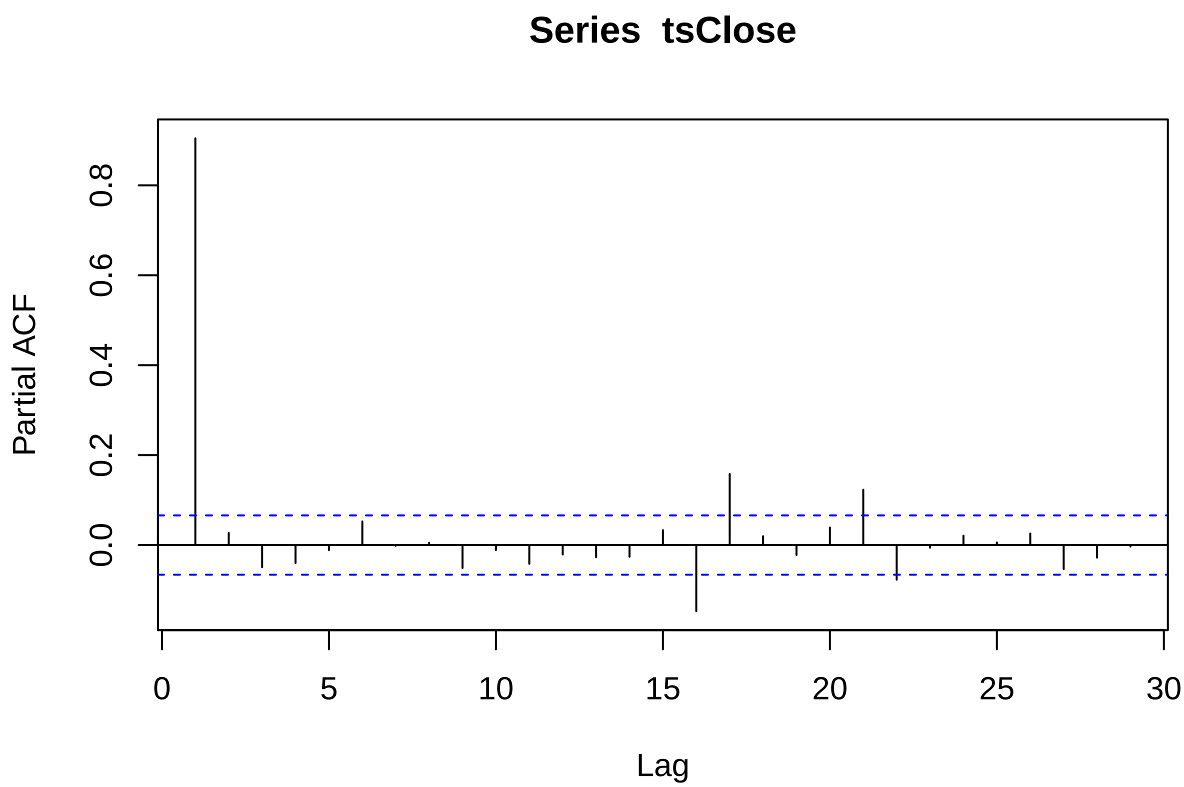



```
##                     GVIF Df GVIF^(1/(2*Df))
## poly(GDP, 2)    1.550002  2        1.115792
## year            8.017740  1        2.831561
## Life_exp        6.148262  1        2.479569
## fertility       1.182697  1        1.087519
## health_spending 3.417633  1        1.848684
## unemployment    1.285902  1        1.133976
```

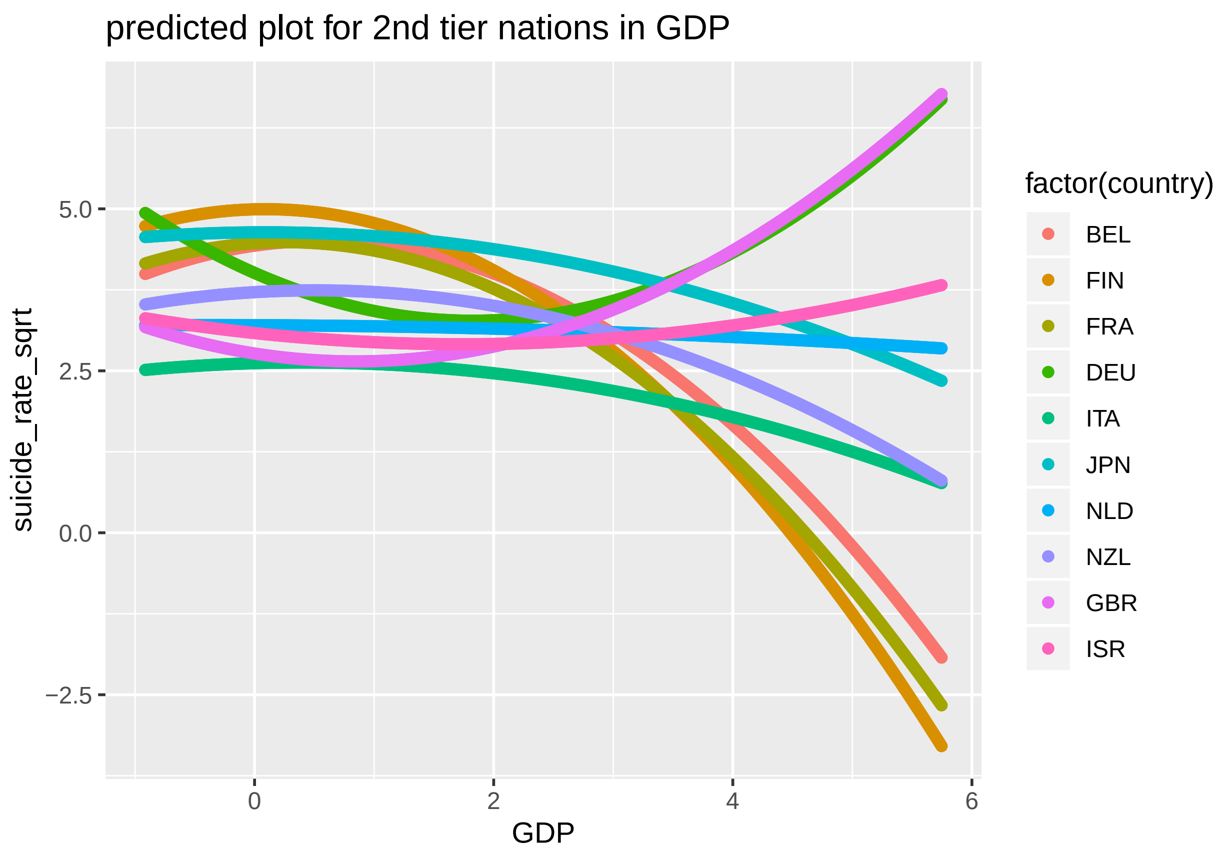

tsClose
7
6
5
4
3
2
0
200
400
600
800
Time



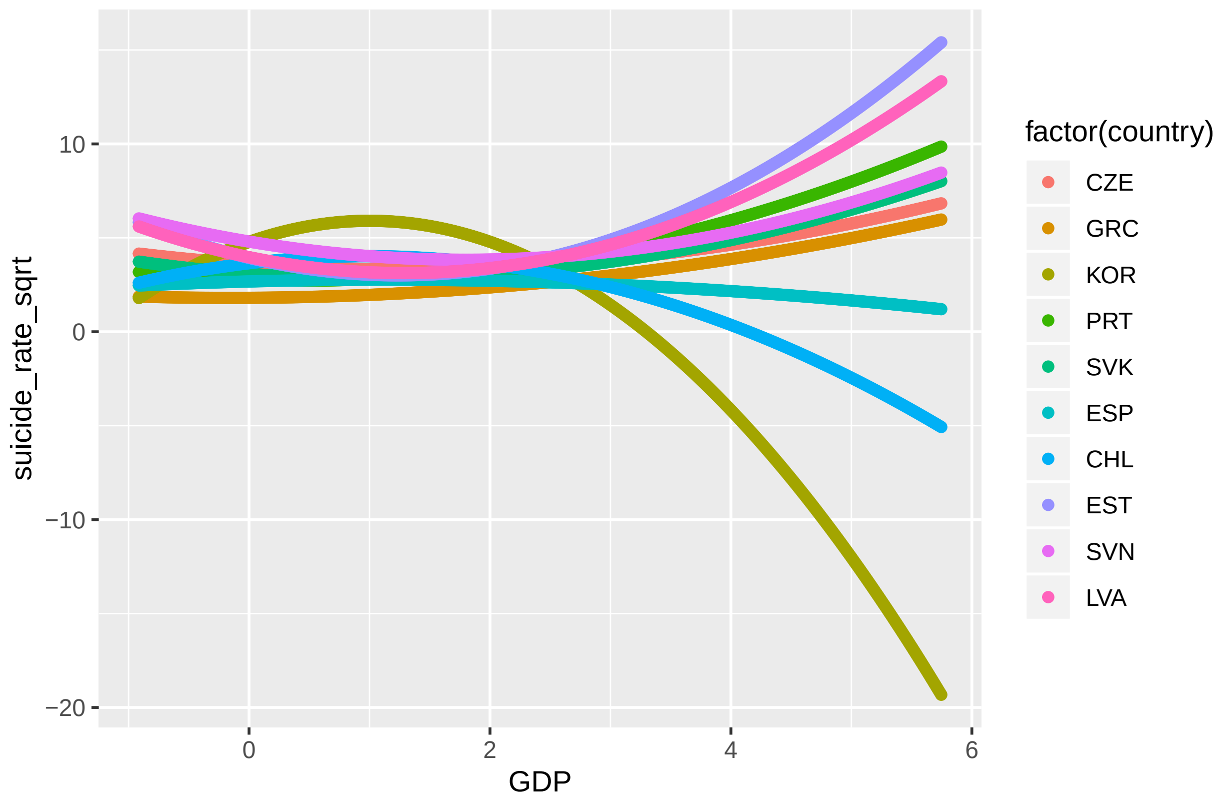
Series tsClose
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

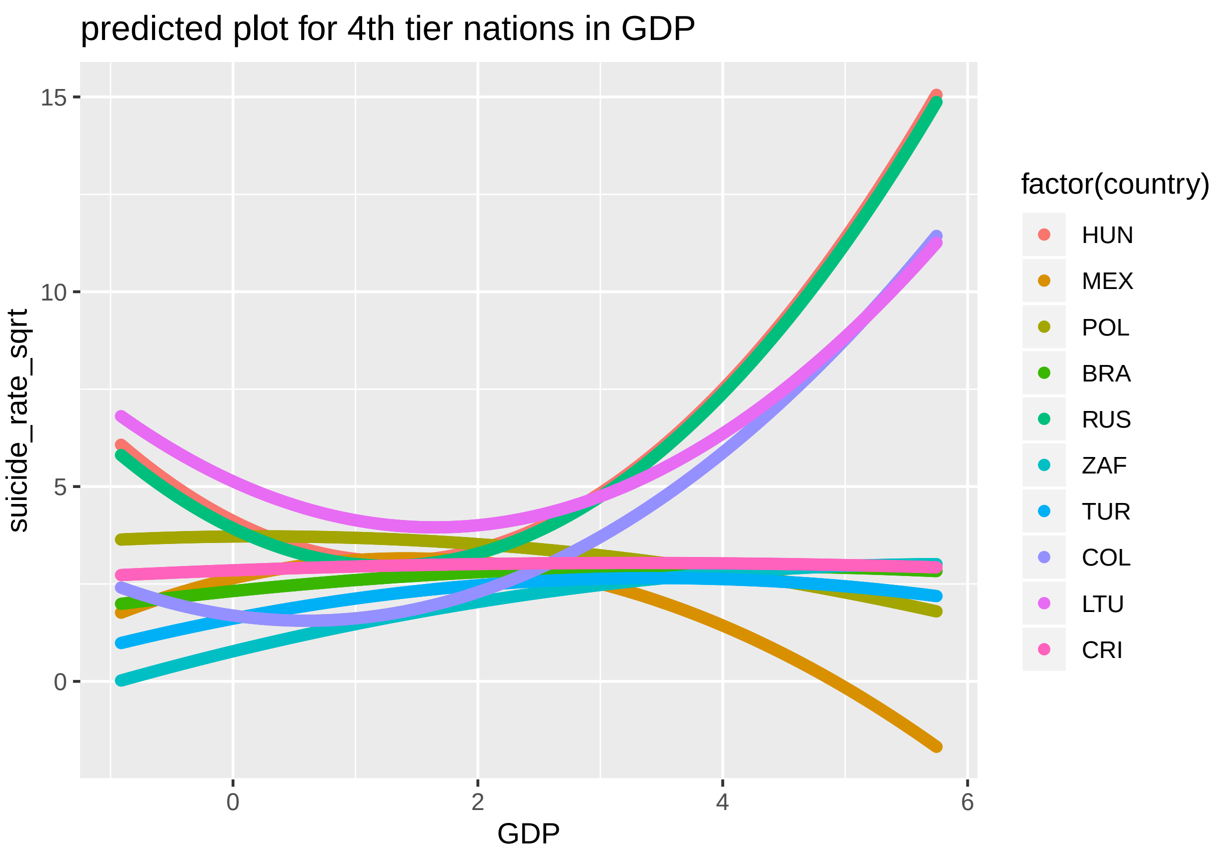

Series tsClose
Partial ACF
0.0
0.2
0.4
0.6
0.8
0
5
10
15
20
25
30
Lag

Appendix A4

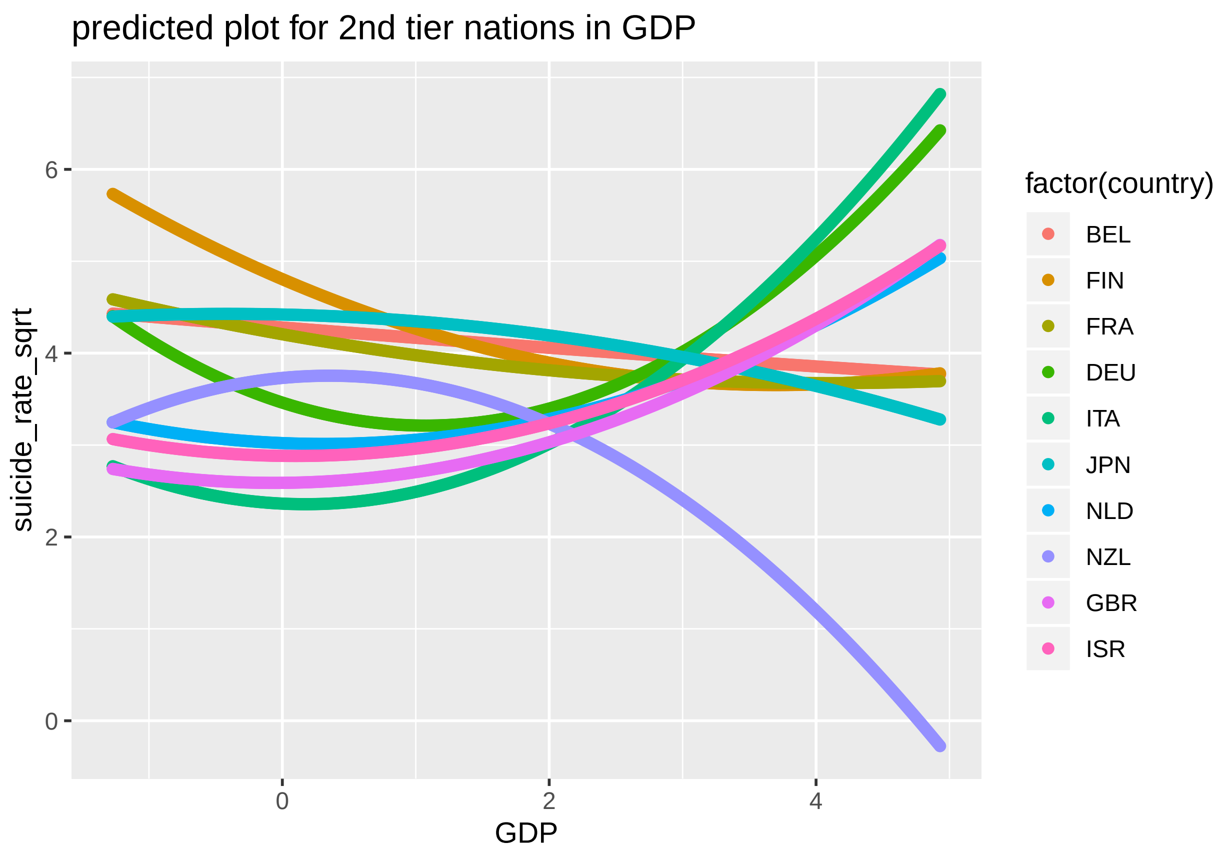

predicted plot for 2nd tier nations in GDP
suicide_rate_sqrt
5.0
2.5
0.0
−2.5
0
2
4
6
GDP
factor(country)
BEL
FIN
FRA
DEU
ITA
JPN
NLD
NZL
GBR
ISR

predicted plot for 3rd tier nations in GDP

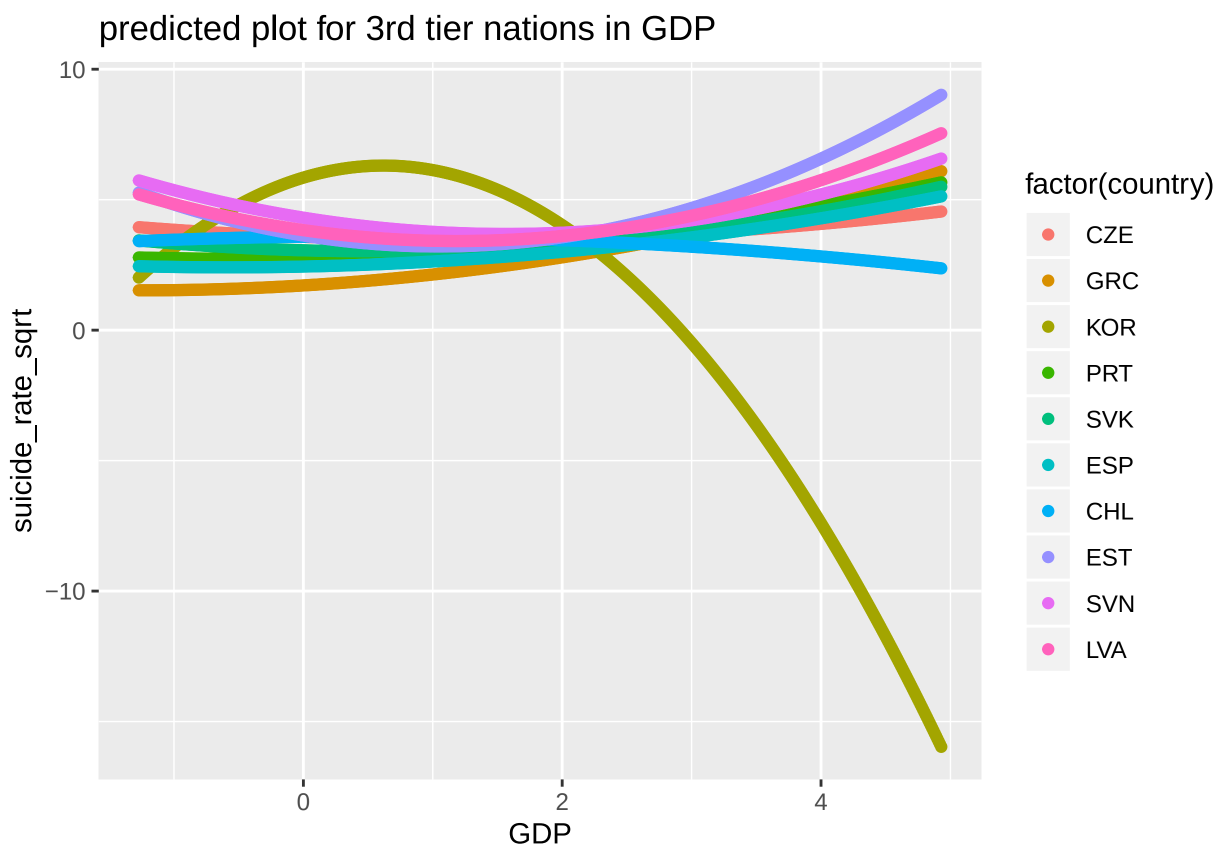


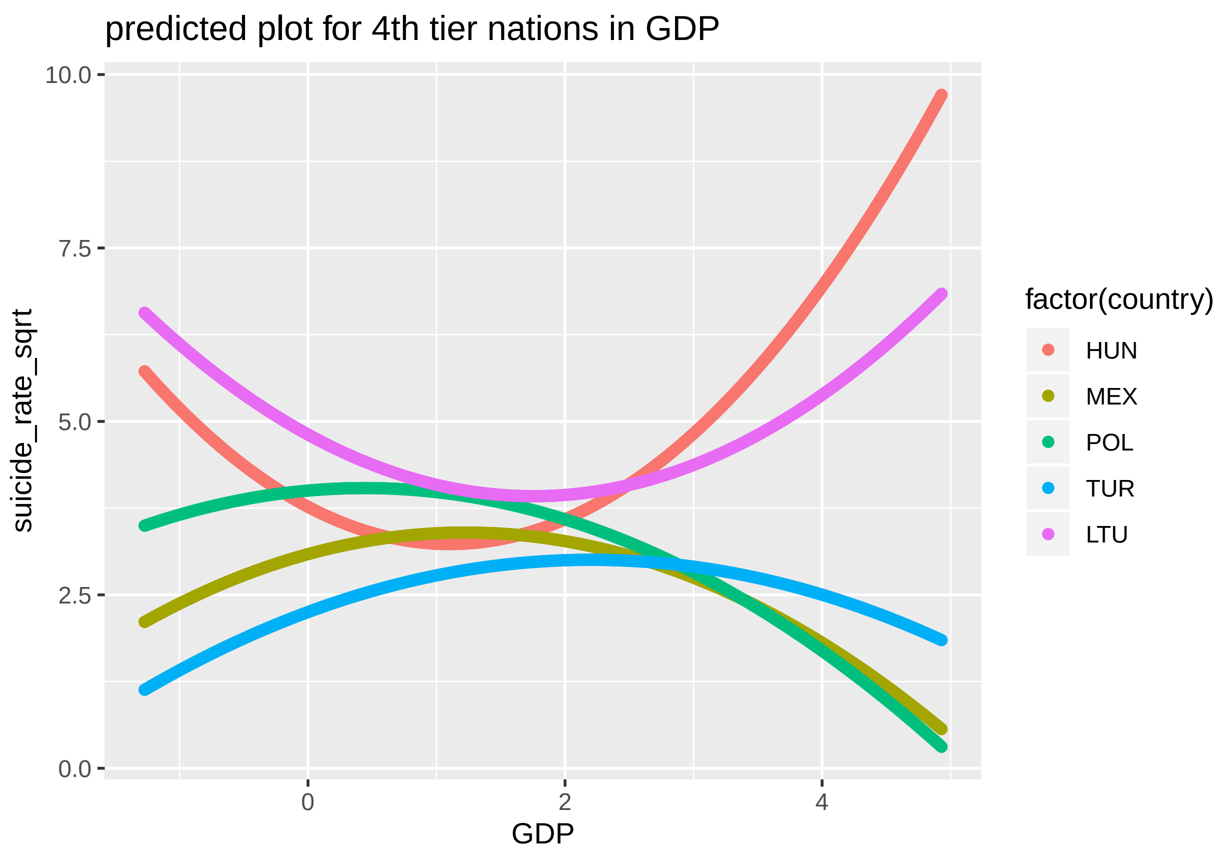

predicted plot for 4th tier nations in GDP
suicide_rate_sqrt
15
10
5
0
0
2
4
6
GDP
factor(country)
HUN
MEX
POL
BRA
RUS
ZAF
TUR
COL
LTU
CRI

Appendix A5



predicted plot for 2nd tier nations in GDP
suicide_rate_sqrt
6
4
2
0
0
2
4
GDP
factor(country)
BEL
FIN
FRA
DEU
ITA
JPN
NLD
NZL
GBR
ISR


predicted plot for 3rd tier nations in GDP
10
0
−10
suicide_rate_sqrt
0
2
4
GDP
factor(country)
CZE
GRC
KOR
PRT
SVK
ESP
CHL
EST
SVN
LVA


predicted plot for 4th tier nations in GDP
10.0
7.5
5.0
2.5
0.0
suicide_rate_sqrt
0
2
4
GDP
factor(country)
HUN
MEX
POL
TUR
LTU

codes I used

```
### Checking response variable distribution for stat2 dataset
ggplot(stat2,aes(suicide_rate)) +
  geom_histogram(alpha=.8,fill=rainbow(10),bins=10)
ggplot(stat2,aes(sqrt(suicide_rate))) +
  geom_histogram(alpha=.8,fill=rainbow(10),bins=10)

### use years
### be careful GDP_3 current GDP
### autocorrleation
### residuals vs years
### years squared

### Checking response variable distribution for stat2_deleted dataset
ggplot(stat2_deleted,aes(suicide_rate)) +
  geom_histogram(alpha=.8,fill=rainbow(10),bins=10)
ggplot(stat2_deleted,aes(sqrt(suicide_rate))) +
  geom_histogram(alpha=.8,fill=rainbow(10),bins=10)

### making new variable, which transforms the response variable with square root, for each dataset
stat2$suicide_rate_sqrt <- sqrt(stat2$suicide_rate)
stat2_deleted$suicide_rate_sqrt <- sqrt(stat2_deleted$suicide_rate)
#stat2$year <- as.factor(stat2$year)

df <- stat2 %>% group_by(year) %>%
         summarise(suicide_rate=mean(suicide_rate))
df <- stat2 %>% group_by(year)

df2 <- stat2 %>% group_by(year) %>%
         summarise(suicide_rate=mean(suicide_rate),gdp = mean(GDP))

ggplot(df, aes(x = year, y = suicide_rate)) +
    geom_point()+
  geom_smooth()

ggplot(df2, aes(x = log(gdp), y = suicide_rate)) +
    geom_point()+
  geom_smooth()

ggplot(stat2 %>% group_by(year) %>%
         summarise(suicide_rate=mean(suicide_rate)),
       aes(x = year, y = suicide_rate )) +
  geom_point(data=stat2, size=4)
  #geom_line(colour="blue", linetype="11", size=0.3) +
  #geom_point(shape=4, colour="blue", size=3)
```

```
### plots against GDP_3 vs suicide rate for different nations on stat2 dataset
ggplot(stat2[stat2$rank==3,], aes(x = log(GDP), y = suicide_rate)) +
    geom_point(aes(color = factor(country)))+
  geom_smooth(aes(color = factor(country)))
ggplot(stat2[stat2$rank==2,], aes(x = log(GDP), y = suicide_rate)) +
    geom_point(aes(color = factor(country)))+
```

```
  geom_smooth(aes(color = factor(country)))
ggplot(stat2[stat2$rank==1,], aes(x = log(GDP), y = suicide_rate)) +
    geom_point(aes(color = factor(country)))+
  geom_smooth(aes(color = factor(country)))
ggplot(stat2[stat2$rank==0,], aes(x = log(GDP), y = suicide_rate)) +
    geom_point(aes(color = factor(country)))+
  geom_smooth(aes(color = factor(country)))

### plots against GDP_3 vs suicide rate for different nations on stat2_deleted dataset
ggplot(stat2_deleted[stat2_deleted$rank==3,], aes(x = GDP_3, y = suicide_rate)) +
    geom_point(aes(color = factor(country)))+
  geom_smooth(aes(color = factor(country)))
ggplot(stat2_deleted[stat2_deleted$rank==2,], aes(x = GDP_3, y = suicide_rate)) +
    geom_point(aes(color = factor(country)))+
  geom_smooth(aes(color = factor(country)))
ggplot(stat2_deleted[stat2_deleted$rank==1,], aes(x = GDP_3, y = suicide_rate)) +
    geom_point(aes(color = factor(country)))+
  geom_smooth(aes(color = factor(country)))
ggplot(stat2_deleted[stat2_deleted$rank==0,], aes(x = GDP_3, y = suicide_rate)) +
    geom_point(aes(color = factor(country)))+
  geom_smooth(aes(color = factor(country)))

### boxplot on suicide rate vs developed for stat2 dataset
ggplot(stat2, aes(x=devel, y=suicide_rate_sqrt)) +
  geom_boxplot()
### boxplot on suicide rate vs rank for stat2 dataset
ggplot(stat2, aes(x=rank, y=suicide_rate_sqrt)) +
  geom_boxplot()

### boxplot on suicide rate vs developed for stat2_deleted dataset
ggplot(stat2_deleted, aes(x=devel, y=suicide_rate_sqrt)) +
  geom_boxplot()
### boxplot on suicide rate vs rank for stat2_deleted dataset
ggplot(stat2_deleted, aes(x=rank, y=suicide_rate_sqrt)) +
  geom_boxplot()

### check for multicolinearity for stat2 dataset
pairs(suicide_rate_sqrt~ Life_exp + fertility,data=stat2[stat2$rank==3,],col = stat2$country)
pairs(suicide_rate_sqrt~ Life_exp + fertility,data=stat2[stat2$rank==2,],col = stat2$country)
pairs(suicide_rate_sqrt~ Life_exp + fertility,data=stat2[stat2$rank==1,],col = stat2$country)
pairs(suicide_rate_sqrt~ Life_exp + fertility,data=stat2[stat2$rank==0,],col = stat2$country)

### check for multicolinearity for stat2_deleted dataset
pairs(suicide_rate_sqrt~ Life_exp + fertility+health_spending+unemployment,data=stat2_deleted[stat2_dele
pairs(suicide_rate_sqrt~  Life_exp + fertility+health_spending+unemployment,data=stat2_deleted[stat2_del
pairs(suicide_rate_sqrt~ Life_exp + fertility+health_spending+unemployment,data=stat2_deleted[stat2_dele
pairs(suicide_rate_sqrt~ Life_exp + fertility+health_spending+unemployment,data=stat2_deleted[stat2_dele

#xyplot(suicide_rate~ GDP, stat2[stat2$rank==3,],group=country,type = c("p", "smooth"), lwd = 4)
xyplot(fertility~rank, stat2)
xyplot(suicide_rate~ GDP_3, stat2)
xyplot(suicide_rate~ Life_exp, stat2)
xyplot(suicide_rate~ fertility, stat2)
```

```r
xyplot(suicide_rate~ year, stat2)
boxplot(stat2$suicide_rate ~ stat2$devel,
        ylab="Suicide_rate",
        xlab="Developed",
        col=c("red3","green3"))
```

```r
### standarized all the numerical variables
stat3 <- preProcess(stat2, method=c("center", "scale"))
newData <- predict(stat3, stat2)
newData$suicide_rate_sqrt <- stat2$suicide_rate_sqrt
stat3 <- preProcess(stat2_deleted, method=c("center", "scale"))
newData_del <- predict(stat3, stat2_deleted)
newData_del$suicide_rate_sqrt <- stat2_deleted$suicide_rate_sqrt
ggplot(newData,aes(suicide_rate_sqrt)) +
  geom_histogram(alpha=.8,fill=rainbow(10),bins=10)
```

```r
#mod1 <- lmer(suicide_rate_sqrt~ GDP_3 + Population + Life_exp + fertility + rank + (1|country), data =
#summary(mod1)
### buidling two model using stat dataset, one with only random intercept and other one with random int
mod_int <- lmer(suicide_rate_sqrt~ poly(GDP,2) + Life_exp + fertility + rank + (1|country), data = newDa
summary(mod_int)
#confint(mod_int)
#look at the intercepts (and the common slope) for each nations
coef(mod_int)
#these equal the fixed effects plus the random effect
fixef(mod_int)
ranef(mod_int)
  mod_int_slo <- lmer(suicide_rate_sqrt~ poly(GDP,2)+ year+ Life_exp + fertility + (1+poly(GDP,2)|countr
summary(mod_int_slo)
#confint(mod_int_slo)
#look at the intercepts (and the common slope) for each nations
coef(mod_int_slo)
#these equal the fixed effects plus the random effect
fixef(mod_int_slo)
ranef(mod_int_slo)



### buidling two model using stat_delete dataset, one with only random intercept and other one with ran
#mod_int_del <- lmer(suicide_rate_sqrt~ GDP_3 + Population + Life_exp + fertility + rank + (1|country),
#summary(mod1)
mod_int_del <- lmer(suicide_rate_sqrt~ poly(GDP,2) + Life_exp + fertility + devel + health_spending +une
summary(mod_int_del)
#confint(mod_int_del)
#look at the intercepts (and the common slope) for each nations
coef(mod_int_del)
#these equal the fixed effects plus the random effect
fixef(mod_int_del)
ranef(mod_int_del)
mod_int_slo_del <- lmer(suicide_rate_sqrt~ poly(GDP,2)+ year + Life_exp + fertility+rank + health_spendi
summary(mod_int_slo_del)
#confint(mod_int_slo_del)
#look at the intercepts (and the common slope) for each nations
```

```r
coef(mod_int_slo_del)
#these equal the fixed effects plus the random effect
fixef(mod_int_slo_del)
ranef(mod_int_slo_del)

### AIC and BIC for each model
AIC(mod_int)
BIC(mod_int)
AIC(mod_int_slo)
BIC(mod_int_slo)
AIC(mod_int_del)
BIC(mod_int_del)
AIC(mod_int_slo_del)
BIC(mod_int_slo_del)

df <- data.frame("GDP" = seq(min(newData[newData$devel==1,]$GDP),max(newData[newData$devel==1,]$GDP),0.0
df$Life_exp = mean(newData[newData$devel==1,]$Life_exp)
df$year = mean(newData[newData$devel==0,]$year)
df$fertility = mean(newData[newData$devel==1,]$fertility)
df$rank = 3
df$rank <- as.factor(df$rank)
df$country <- "ISR"
df$country <- as.factor(df$country)
df$suicide_rate_sqrt = predict(mod_int_slo, newdata=df)

plot(x=df$GDP, y=df$suicide_rate_sqrt, pch=20, col="grey")
#lines(newData$GDP, predict(mod_int_slo, data=newData), type="l", col="orange1", lwd=2)

df2 <- data.frame("GDP" = seq(min(newData_del[newData_del$devel==1,]$GDP),max(newData_del[newData_del$de
df2$Life_exp = mean(newData_del[newData_del$devel==1,]$Life_exp)
df2$year = mean(newData_del[newData_del$devel==1,]$year)
df2$fertility = mean(newData_del[newData_del$devel==1,]$fertility)
df2$rank = 2
df2$rank <- as.factor(df2$rank)
df2$country <- "ISR"
df2$country <- as.factor(df2$country)
df2$health_spending <- mean(newData_del[newData_del$devel==1,]$health_spending)
df2$unemployment <- mean(newData_del[newData_del$devel==1,]$unemployment)
df2$suicide_rate_sqrt = predict(mod_int_slo_del, newdata=df2)

plot(x=df2$GDP, y=df2$suicide_rate_sqrt, pch=20, col="grey")
#lines(newData$GDP, predict(mod_int_slo, data=newData), type="l", col="orange1", lwd=2)


###
ggplot(newData_del,aes(suicide_rate_sqrt)) +
  geom_histogram(alpha=.8,fill=rainbow(10),bins=10)
ggplot(stat2_deleted,aes(suicide_rate_sqrt)) +
  geom_histogram(alpha=.8,fill=rainbow(10),bins=10)

kable(tidy(mod_int_slo, conf.int = TRUE),digit = 5)
```

```
plot(mod_int_slo)
plot(y = residuals(mod_int_slo), x = newData$GDP_3, xlab= "GDP_3", ylab = "Residuals")
ggplot(newData,aes(x=GDP_3, y=residuals(mod_int_slo))) +
  geom_point(alpha = .5,colour="blue4") +
  geom_smooth(method="lm",col="red3") +
  labs(title="Suicide_rate vs GDP_3") +
  facet_wrap(~country,ncol=5)
qqnorm(residuals(mod_int_slo));qqline(residuals(mod_int_slo))
car::vif(mod_int_slo)

qqnorm(residuals(mod_int_del));qqline(residuals(mod_int_del))

plot(mod_int_slo_del)
plot(y = residuals(mod_int_slo_del), x =newData_del$GDP_3, xlab= "GDP_3", ylab = "Residuals")
plot(y = residuals(mod_int_slo_del), x =newData_del$year, xlab= "year", ylab = "Residuals")
ggplot(newData_del,aes(x=GDP_3, y=residuals(mod_int_slo_del))) +
  geom_point(alpha = .5,colour="blue4") +
  geom_smooth(method="lm",col="red3") +
  labs(title="Suicide_rate vs GDP_3") +
  facet_wrap(~country,ncol=5)
qqnorm(residuals(mod_int_slo_del));qqline(residuals(mod_int_slo_del))
car::vif(mod_int_slo_del)
```