

STAT 4355 Final

John Kenney, Kevin Shaik, Keerthi Srilakshmidaran

Problem 1

```
library(dplyr)
library(tidyR)
df <- read.csv("C:/Users/jkenn/OneDrive/Documents/R/Stat4355HW/final/life1.csv")
full_data <- read.csv("C:/Users/jkenn/OneDrive/Documents/R/Stat4355HW/final/life Expectancy Data.csv")
# head(temp)
full_data <- dplyr::select(full_data, -(Status), -(Adult.Mortality:infant.deaths),
  -(percentage.expenditure), -(under.five.deaths), -(Diphtheria), -(Population),
  -(thinness..1.19.years:thinness.5.9.years))
# head(full_data) nrow(full_data) colSums(is.na(full_data))
full_data <- na.omit(full_data)
# full_data <- temp[,1:2]
rownames(full_data) <- 1:nrow(full_data)

# nrow(full_data) write.csv(full_data, 'Life1NoNAIndexedCorrectly.csv')
df <- df[, 3:13] #select(df, -(X:Year))
# colnames(df)[19:20] <- c('thick', 'thin') df <- select(df,
# -(Country>Status),-(thick:thin))

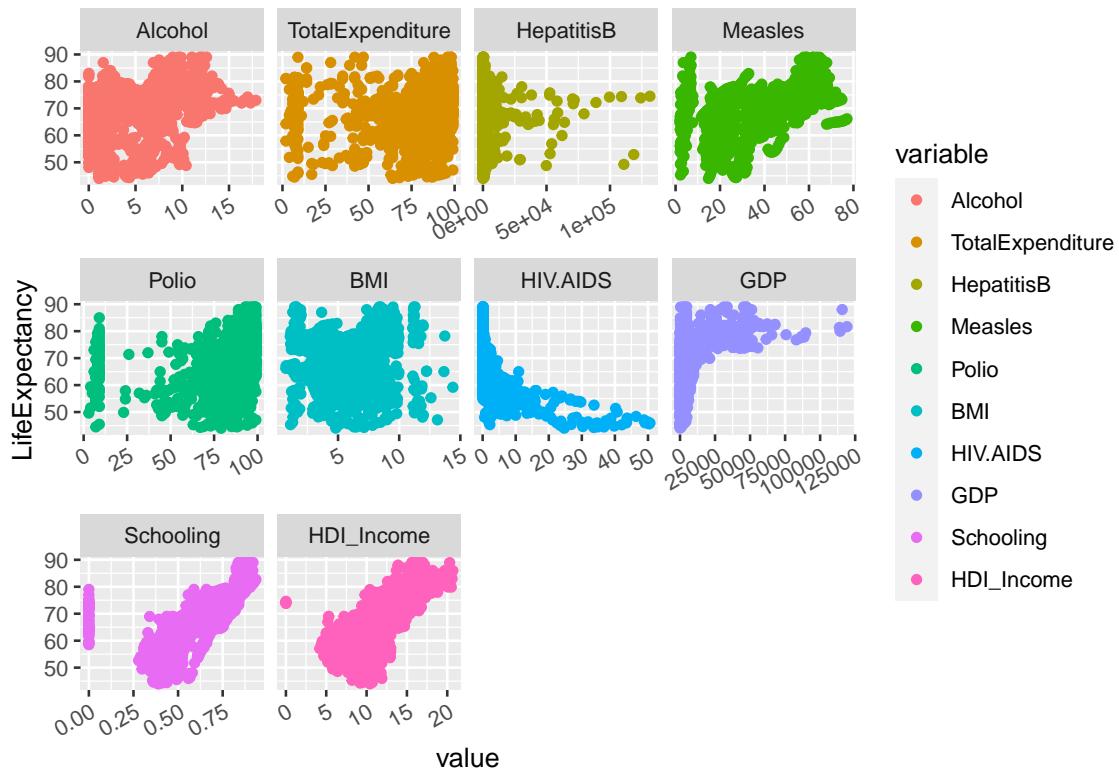
# head(df)

df <- df %>% relocate(LifeExpectancy, .after = last_col())
df <- df %>% drop_na("LifeExpectancy")
# colSums(is.na(df))
df <- na.omit(df)
# head(df) indexer <- df[,1] indexer <- data.frame( 'index to smaller data' =
# 1:nrow(df), 'index to full data' = df[,1]) View(indexer) nrow(df) head(df) df <-
# df[,2:12] rownames(df) <- 1:nrow(df) View(df) View(full_data)
# colSums(is.na(df))

library(reshape2)

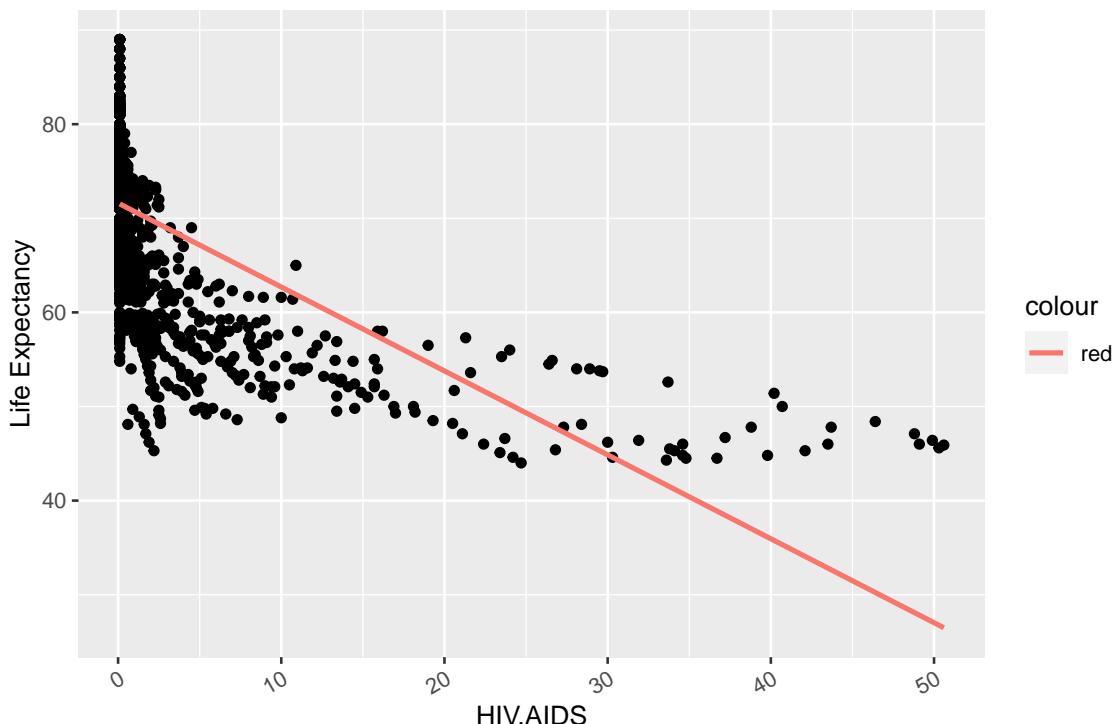
## Warning: package 'reshape2' was built under R version 4.0.4

library(ggplot2)
df2 <- melt(df[, 1:11], id.vars = "LifeExpectancy")
ggplot(df2) + geom_jitter(aes(value, LifeExpectancy, colour = variable), ) + facet_wrap(~variable,
  scales = "free_x") + theme(axis.text.x = element_text(angle = 30, hjust = 1))
```

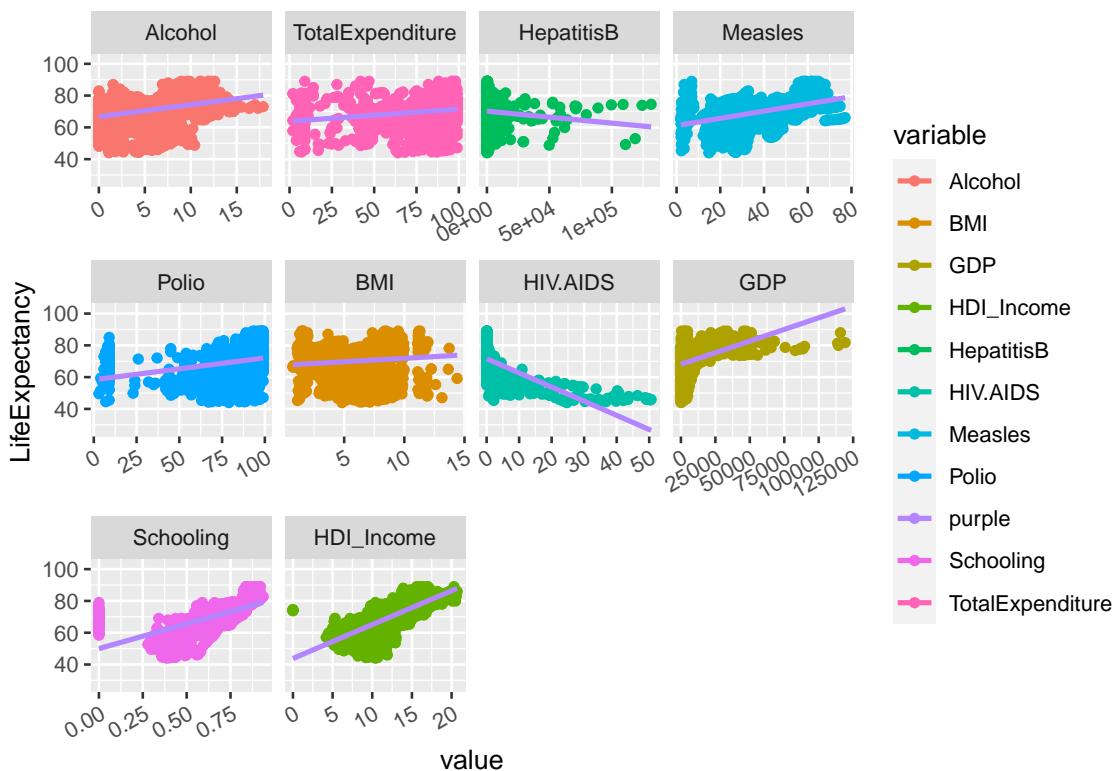


```
library(reshape2)
library(ggplot2)
ggplot(df, aes(x = df[, 7], y = df[, 11])) + geom_point() + geom_smooth(aes(x = df[, 7], y = df[, 11], col = "red"), method = lm, se = FALSE) + theme(axis.text.x = element_text(angle = 90, hjust = 1)) + xlab("HIV.AIDS") + ylab("Life Expectancy") + ggtitle("Life Expectancy vs HIV.AIDS")
```

Life Expectancy vs HIV.AIDS



```
ggplot(df2) + geom_jitter(aes(value, LifeExpectancy, colour = variable), ) + geom_smooth(aes(value, LifeExpectancy, colour = "purple"), method = lm, se = FALSE) + facet_wrap(~variable, scales = "free_x") + theme(axis.text.x = element_text(angle = 30, hjust = 1))
```



```

full <- lm(LifeExpectancy ~ ., data = df)
summary(full)

## 
## Call:
## lm(formula = LifeExpectancy ~ ., data = df)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -17.3650  -2.5226   0.1088   2.5485  23.4122 
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 4.728e+01 6.098e-01 77.522 < 2e-16 ***
## Alcohol     -2.902e-02 3.011e-02 -0.964 0.33527  
## TotalExpenditure 3.067e-03 4.477e-03 0.685 0.49338  
## HepatitisB   4.096e-06 1.046e-05 0.392 0.69534  
## Measles      4.782e-02 5.956e-03 8.029 1.74e-15 ***
## Polio        2.932e-02 5.326e-03 5.504 4.23e-08 *** 
## BMI          1.209e-01 4.384e-02 2.758 0.00588 **  
## HIV.AIDS    -6.407e-01 1.791e-02 -35.774 < 2e-16 *** 
## GDP          7.997e-05 8.563e-06 9.339 < 2e-16 *** 
## Schooling    1.020e+01 8.035e-01 12.692 < 2e-16 *** 
## HDI_Income   9.426e-01 5.910e-02 15.948 < 2e-16 *** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 4.189 on 1842 degrees of freedom
## Multiple R-squared:  0.7647, Adjusted R-squared:  0.7634 
## F-statistic: 598.5 on 10 and 1842 DF,  p-value: < 2.2e-16

```

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.0.5
```

```
vif(full)
```

	Alcohol	TotalExpenditure	HepatitisB	Measles
##	1.526155	1.321104	1.049843	1.468164
##	Polio	BMI	HIV.AIDS	GDP
##	1.410379	1.112646	1.107714	1.277588
##	Schooling	HDI_Income		
##	2.474129	3.008093		

```
# View(trainer)
df2 <- dplyr::select(df, -(Alcohol), -(HepatitisB), -(TotalExpenditure))

reducedmodel <- lm(LifeExpectancy ~ ., data = df2)
summary(reducedmodel)
```

```
## 
## Call:
```

```

## lm(formula = LifeExpectancy ~ ., data = df2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.2465 -2.5331  0.1145  2.5688 23.0888
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.758e+01 5.582e-01 85.236 < 2e-16 ***
## Measles     4.767e-02 5.914e-03 8.061 1.35e-15 ***
## Polio       3.081e-02 4.790e-03 6.431 1.61e-10 ***
## BMI         1.122e-01 4.291e-02 2.615 0.009 **
## HIV.AIDS    -6.434e-01 1.776e-02 -36.219 < 2e-16 ***
## GDP          7.882e-05 8.506e-06 9.266 < 2e-16 ***
## Schooling   1.019e+01 8.013e-01 12.714 < 2e-16 ***
## HDI_Income   9.242e-01 5.584e-02 16.551 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.188 on 1845 degrees of freedom
## Multiple R-squared: 0.7645, Adjusted R-squared: 0.7636
## F-statistic: 855.5 on 7 and 1845 DF, p-value: < 2.2e-16

```

```

library(car)
vif(reducedmodel)

```

```

##      Measles      Polio       BMI    HIV.AIDS       GDP  Schooling HDI_Income
## 1.448845 1.141842 1.066612 1.090649 1.261795 2.462386 2.687011

```

```

anova(reducedmodel, full)

```

```

## Analysis of Variance Table
##
## Model 1: LifeExpectancy ~ Measles + Polio + BMI + HIV.AIDS + GDP + Schooling +
##           HDI_Income
## Model 2: LifeExpectancy ~ Alcohol + TotalExpenditure + HepatitisB + Measles +
##           Polio + BMI + HIV.AIDS + GDP + Schooling + HDI_Income
##   Res.Df   RSS Df Sum of Sq    F Pr(>F)
## 1    1845 32355
## 2    1842 32327  3    27.461 0.5216 0.6675

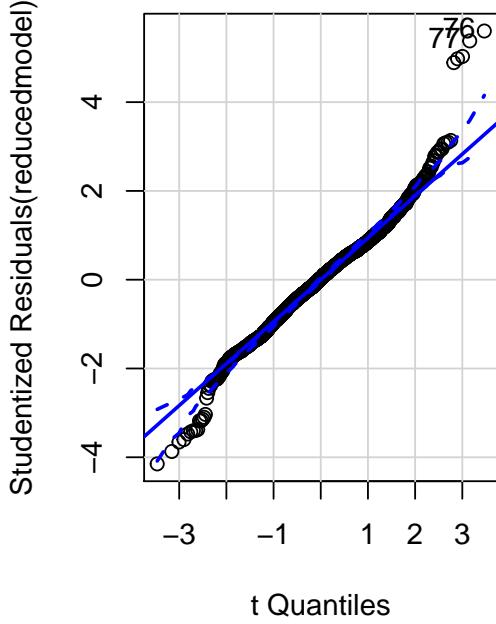
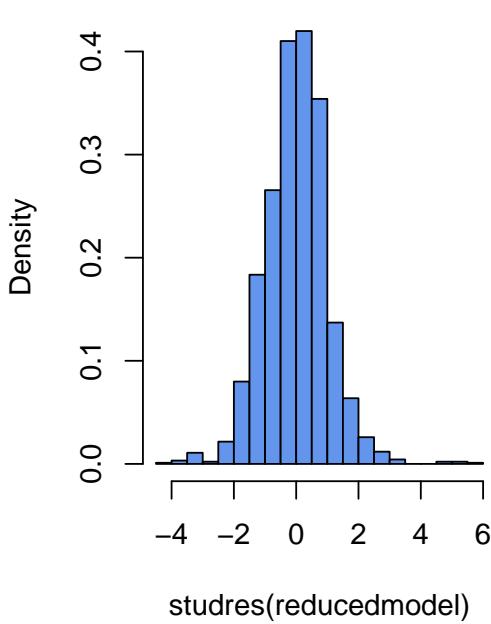
```

```

# QQ plot approach
library(MASS)
library(car)
par(mfrow = c(1, 2))
hist(studres(reducedmodel), breaks = 25, freq = F, col = "cornflowerblue", cex.axis = 1,
     cex.lab = 1, cex.main = 1)
qqPlot(reducedmodel) + title("QQ plot of StuResiduals vs t-Quantiles", cex.main = 1)

```

Histogram of studies(reducedmodel) QQ plot of StuResiduals vs t-Quantile



```

## integer(0)

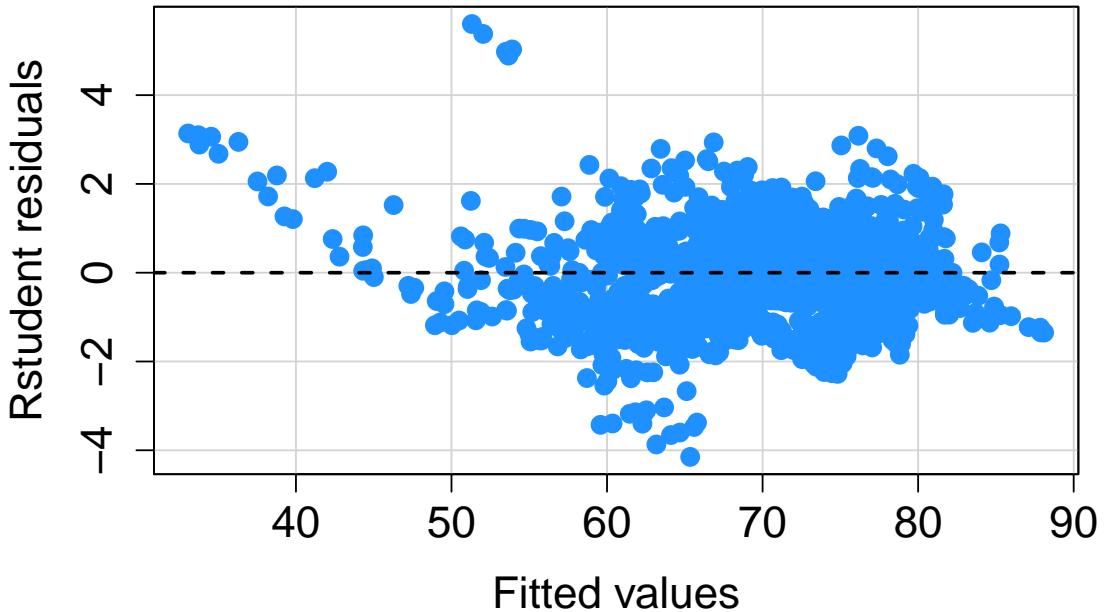
full_data[62:63, ]

##          Country Year Life.expectancy Alcohol Hepatitis.B Measles BMI
## 62 Antigua and Barbuda 2004      74.4    7.28      97     0 4.8
## 63 Antigua and Barbuda 2003      74.2    7.16      99     0 4.1
##   Polio Total.expenditure HIV.AIDS      GDP Income.composition.of.resources
## 62      97           4.21      0.1 1352.837                  0
## 63      99           4.53      0.1 9739.826                  0
##   Schooling
## 62      0
## 63      0

residualPlot(reducedmodel, type = "rstudent", quadratic = F, col = "dodgerblue",
             pch = 16, cex = 1.5, cex.axis = 1.5, cex.lab = 1.5) + title("Rstudent Residuals vs Fitted values")

```

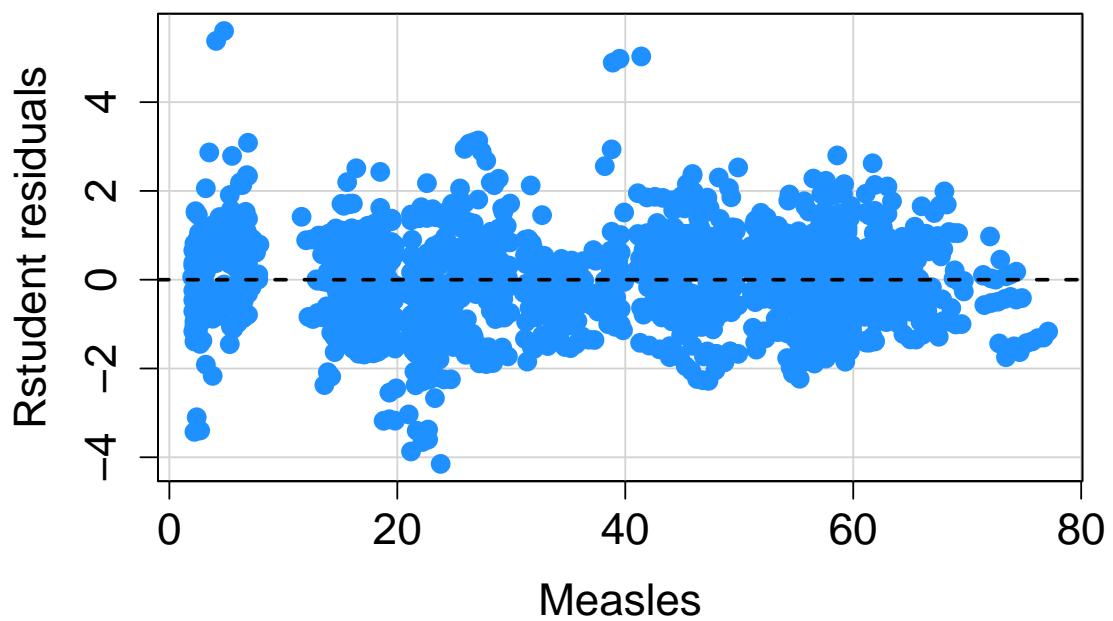
Rstudent Residuals vs Fitted values



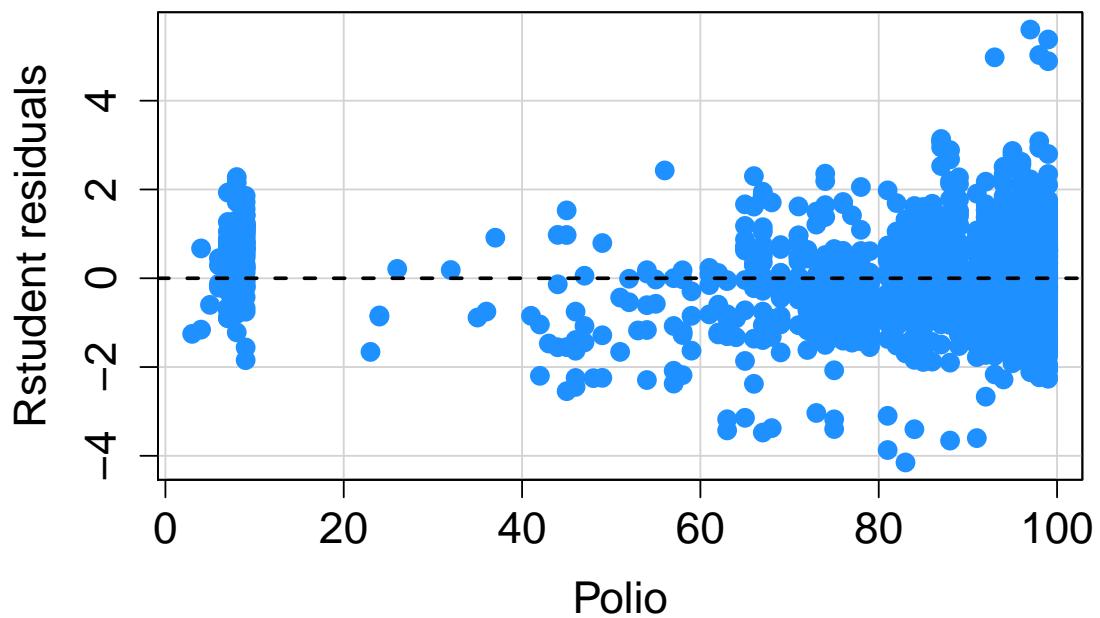
```
## numeric(0)

for (i in 1:7) {
  temp <- paste(c("Rstudent Residuals vs ", names(df2)[i]), collapse = " ")
  residualPlot(reducedmodel, type = "rstudent", quadratic = F, col = "dodgerblue",
    pch = 16, cex = 1.5, cex.axis = 1.5, cex.lab = 1.5, var = names(df2)[i]) +
    title(temp)
}
```

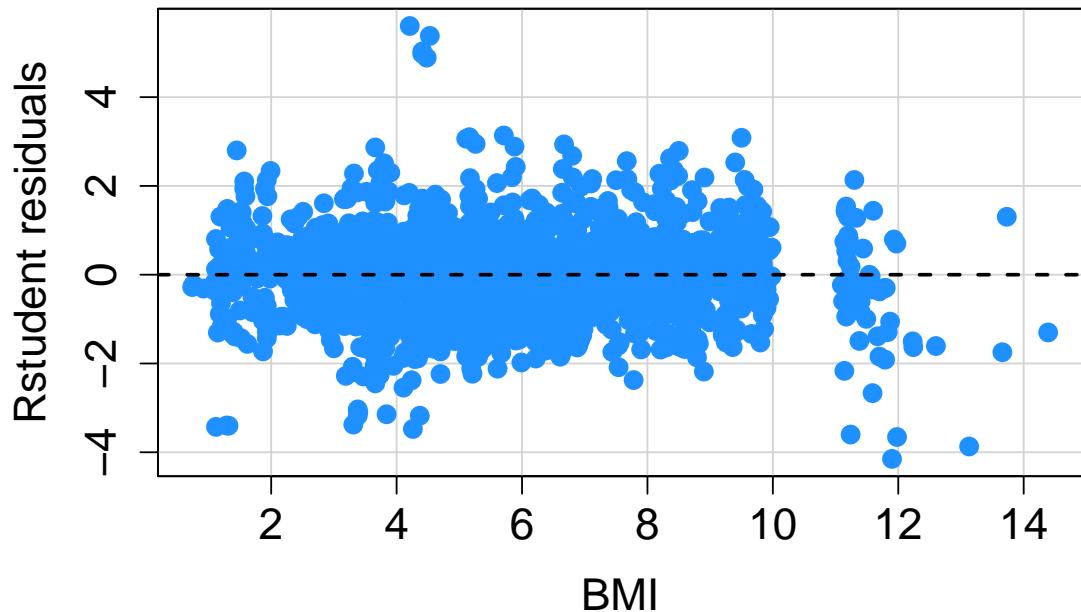
Rstudent Residuals vs Measles



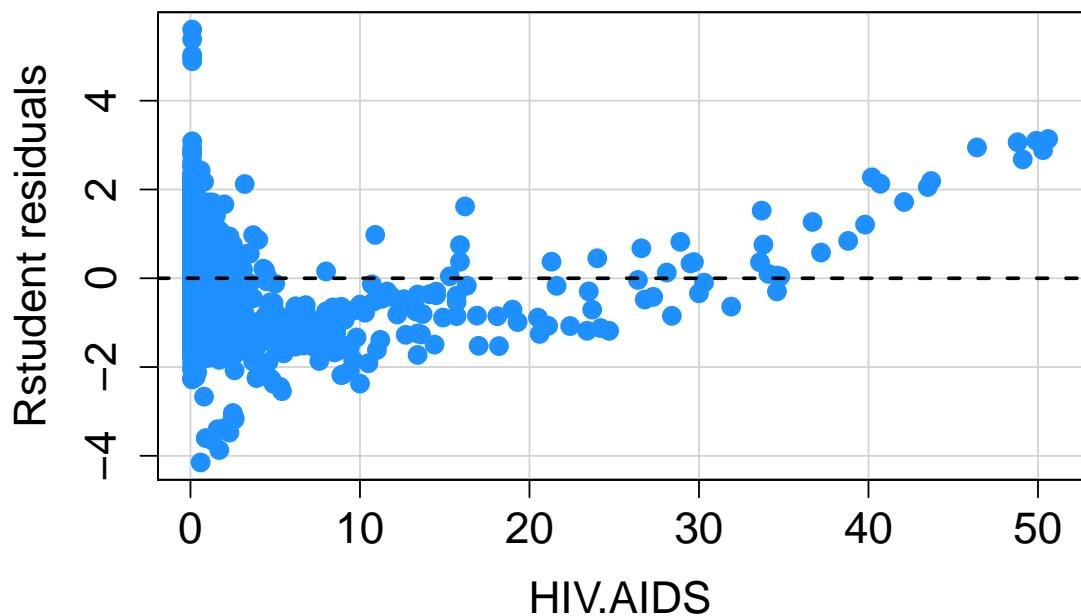
Rstudent Residuals vs Polio



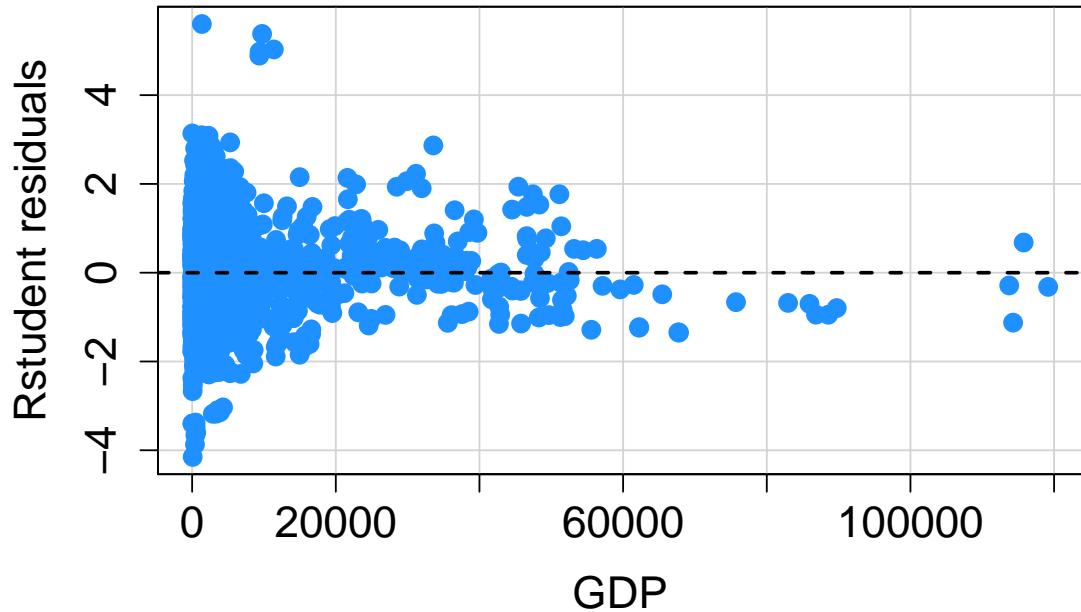
Rstudent Residuals vs BMI



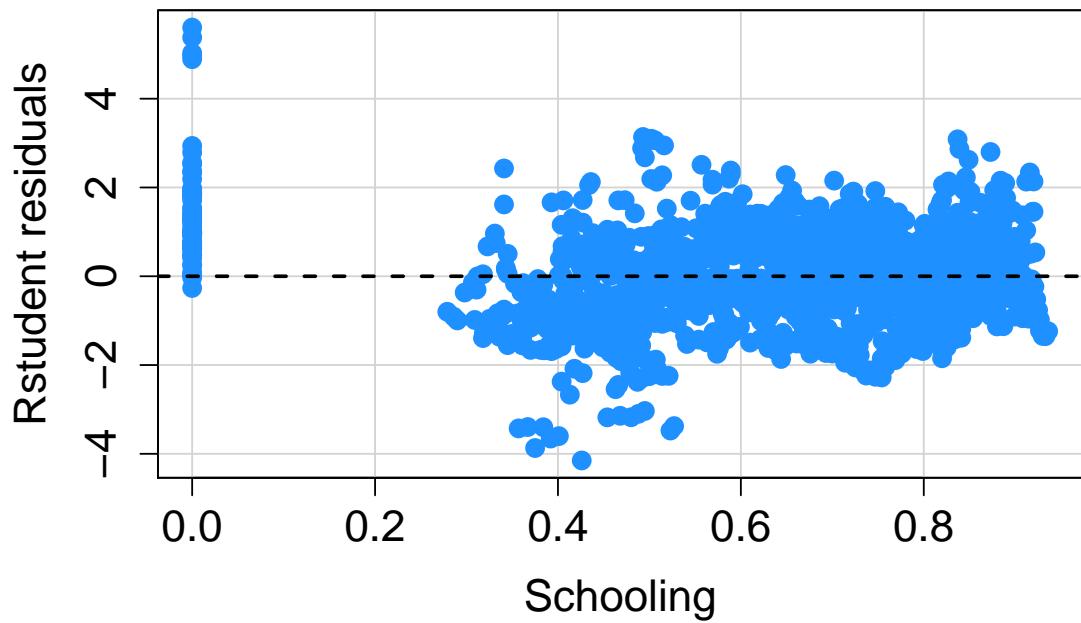
Rstudent Residuals vs HIV.AIDS



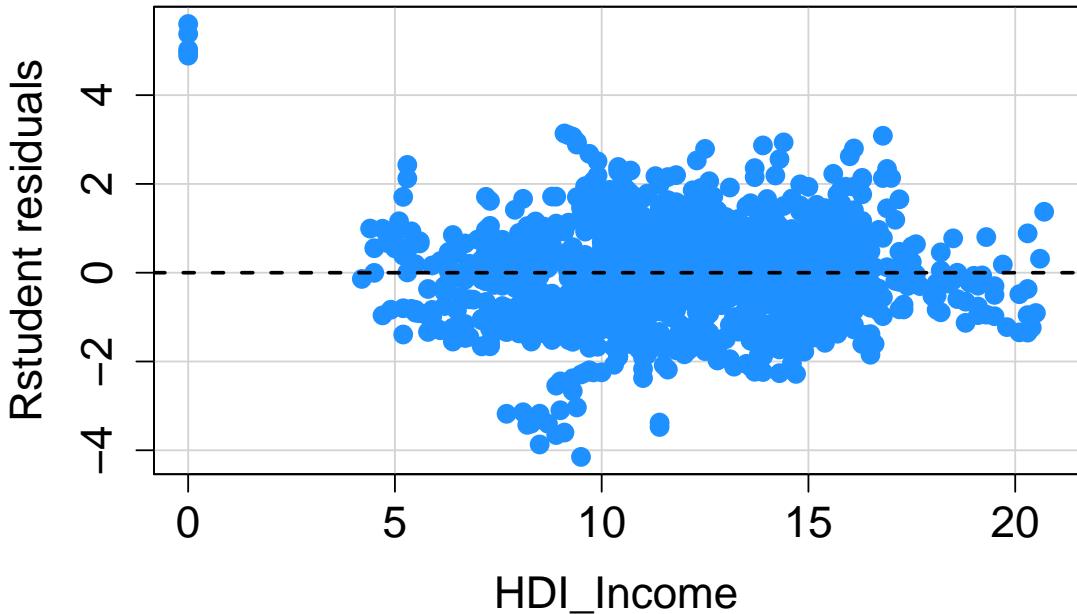
Rstudent Residuals vs GDP



Rstudent Residuals vs Schooling



Rstudent Residuals vs HDI_Income

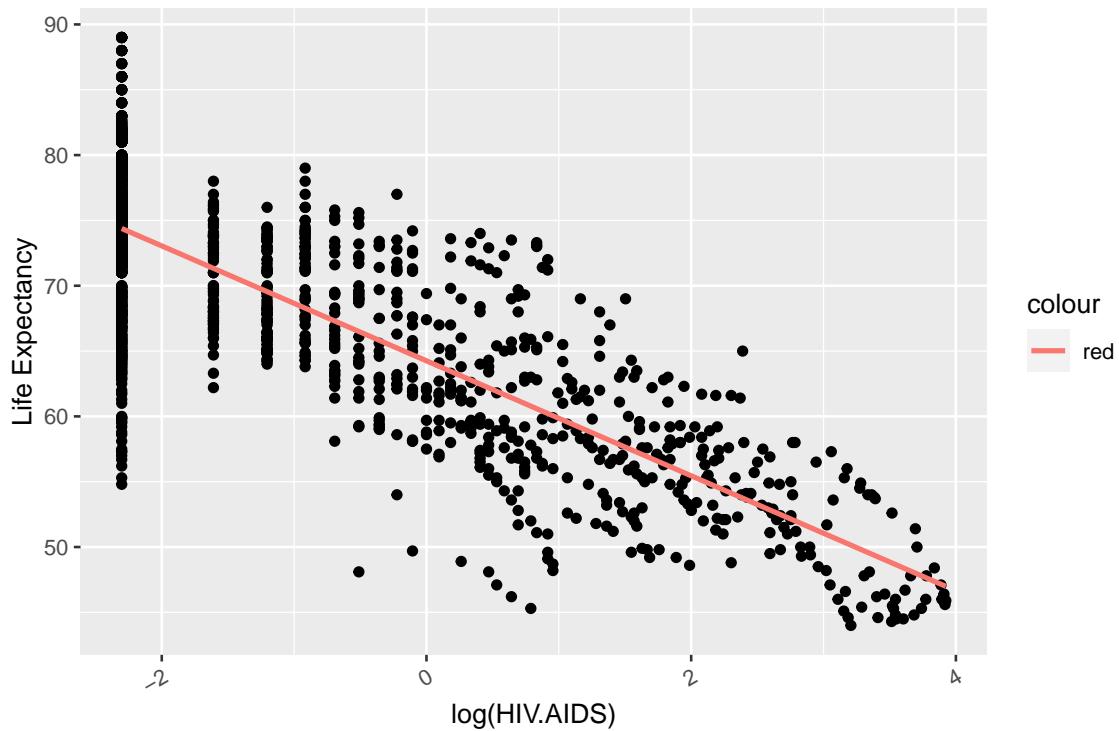


Transformed data by taking the log of the predictor HIV.AIDS

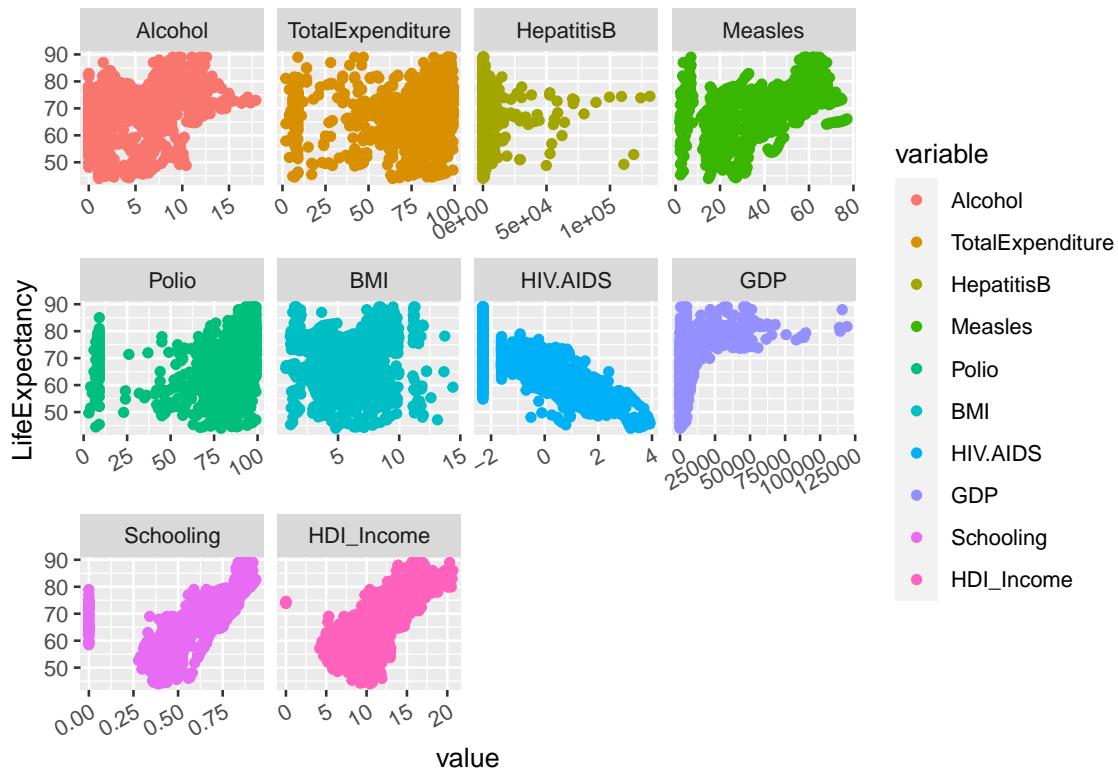
```
df3 <- df
df3[, 7] <- log(df3[, 7])
full_data_Transformed <- full_data
full_data_Transformed[, 10] <- log(full_data_Transformed[, 10])
```

```
library(reshape2)
library(ggplot2)
ggplot(df, aes(x = df3[, 7], y = df3[, 11])) + geom_point() + geom_smooth(aes(x = df3[, 7], y = df3[, 11], col = "red"), method = lm, se = FALSE) + theme(axis.text.x = element_text(angle = 90))
  hjust = 1)) + xlab("log(HIV.AIDS)") + ylab("Life Expectancy") + ggtitle("Life Expectancy vs log(HIV.AIDS)"))
```

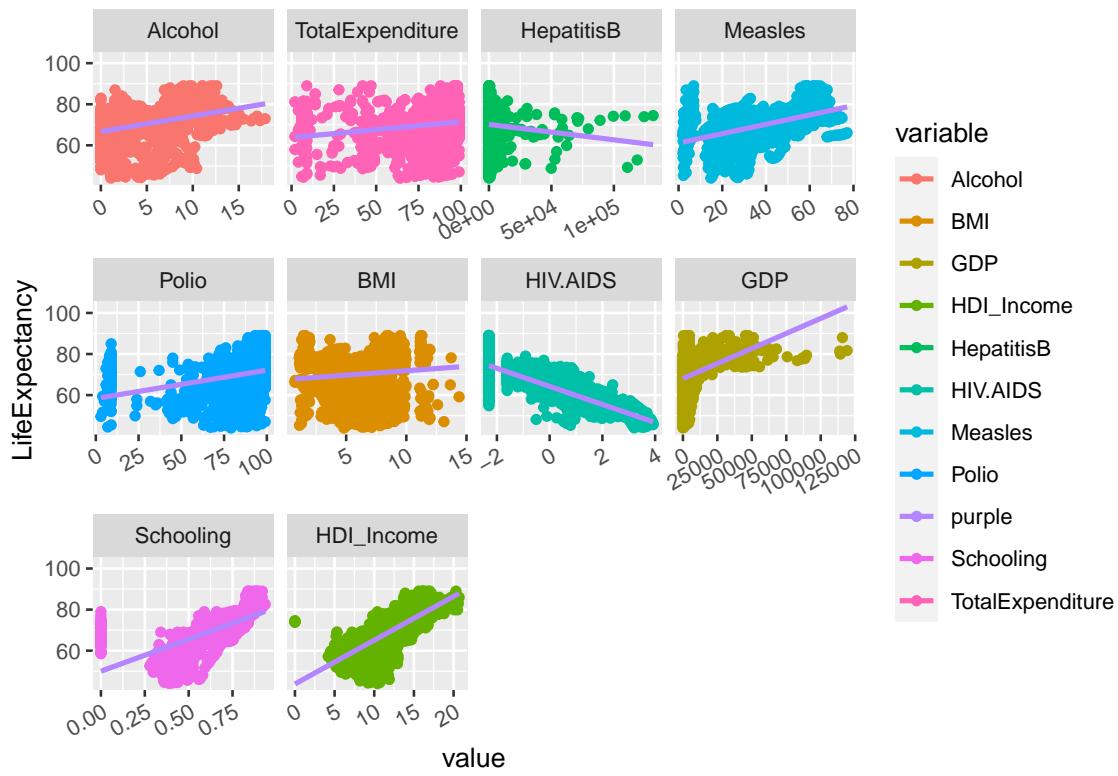
Life Expectancy vs log(HIV.AIDS)



```
library(reshape2)
library(ggplot2)
df2 <- melt(df3[, 1:11], id.vars = "LifeExpectancy")
ggplot(df2 + geom_jitter(aes(value, LifeExpectancy, colour = variable), ) + facet_wrap(~variable,
  scales = "free_x") + theme(axis.text.x = element_text(angle = 30, hjust = 1))
```



```
ggplot(df2) + geom_jitter(aes(value, LifeExpectancy, colour = variable), ) + geom_smooth(aes(value, LifeExpectancy, colour = "purple"), method = lm, se = FALSE) + facet_wrap(~variable, scales = "free_x") + theme(axis.text.x = element_text(angle = 30, hjust = 1))
```



```

full2 <- lm(LifeExpectancy ~ ., data = df3)
summary(full2)

##
## Call:
## lm(formula = LifeExpectancy ~ ., data = df3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -16.0151 -2.0282 -0.0508  2.0987 14.3073 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 5.022e+01 5.621e-01  89.351 < 2e-16 ***
## Alcohol      7.008e-02 2.732e-02   2.565  0.01039 *  
## TotalExpenditure -1.133e-03 4.026e-03  -0.281  0.77841  
## HepatitisB    -1.070e-05 9.410e-06  -1.138  0.25540  
## Measles       1.699e-02 5.452e-03   3.116  0.00186 ** 
## Polio         1.441e-02 4.803e-03   3.000  0.00274 ** 
## BMI           1.756e-01 3.947e-02   4.450  9.1e-06 *** 
## HIV.AIDS     -3.107e+00 6.904e-02  -45.005 < 2e-16 *** 
## GDP           7.056e-05 7.699e-06   9.166 < 2e-16 *** 
## Schooling     9.401e+00 7.229e-01  13.004 < 2e-16 *** 
## HDI_Income    4.955e-01 5.456e-02   9.082 < 2e-16 *** 
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.764 on 1842 degrees of freedom
## Multiple R-squared:  0.81, Adjusted R-squared:  0.809 
## F-statistic: 785.5 on 10 and 1842 DF, p-value: < 2.2e-16

```

```

library(car)
vif(full2)

```

	Alcohol	TotalExpenditure	HepatitisB	Measles
##	1.556616	1.323582	1.053105	1.524246
##	Polio	BMI	HIV.AIDS	GDP
##	1.420775	1.117049	1.536178	1.279349
##	Schooling	HDI_Income		
##	2.480993	3.175616		

```

# View(trainer)
df4 <- dplyr::select(df3, -(HepatitisB), -(TotalExpenditure))

reducedmodel2 <- lm(LifeExpectancy ~ ., data = df4)
summary(reducedmodel2)

```

```

##
## Call:
## lm(formula = LifeExpectancy ~ ., data = df4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -16.0151 -2.0282 -0.0508  2.0987 14.3073 
##
```

```

## -16.0173 -2.0172 -0.0585 2.0699 14.3770
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.009e+01 5.407e-01 92.634 < 2e-16 ***
## Alcohol     6.887e-02 2.723e-02  2.529 0.01151 *
## Measles     1.770e-02 5.412e-03  3.270 0.00109 **
## Polio       1.401e-02 4.335e-03  3.232 0.00125 **
## BMI         1.781e-01 3.937e-02  4.525 6.44e-06 ***
## HIV.AIDS   -3.101e+00 6.878e-02 -45.091 < 2e-16 ***
## GDP         7.097e-05 7.687e-06  9.233 < 2e-16 ***
## Schooling    9.359e+00 7.219e-01 12.965 < 2e-16 ***
## HDI_Income   4.995e-01 5.439e-02  9.184 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.763 on 1844 degrees of freedom
## Multiple R-squared: 0.8099, Adjusted R-squared: 0.8091
## F-statistic: 982 on 8 and 1844 DF, p-value: < 2.2e-16

```

LifeExpectancy = 5.009e+01 + (6.887e-02)Alcohol + (1.770e-02)Measles + (1.401e-02)Polio + (1.781e-01)BMI - (3.101)HIV/AIDS + (7.097e-05)GDP + (9.359)Schooling + (4.995e-01)HDI_Income

```

library(car)
vif(reducedmodel2)

```

	Alcohol	Measles	Polio	BMI	HIV.AIDS	GDP	Schooling
##	1.546804	1.502020	1.157750	1.112086	1.525205	1.275908	2.474571
## HDI_Income							
##	3.157052						

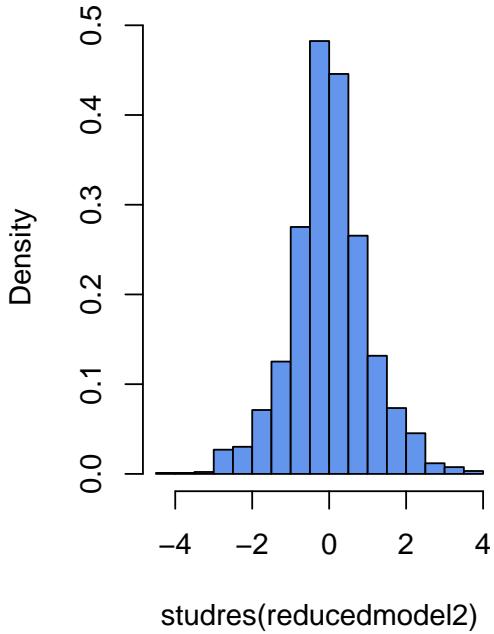
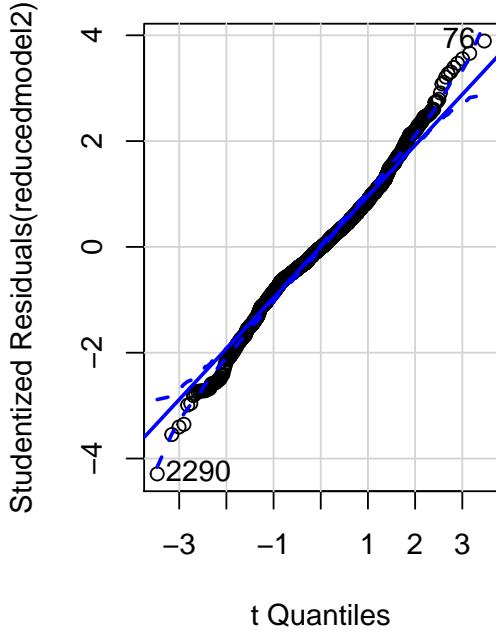
```
anova(reducedmodel2, full2)
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
##	1	1844	26113			
##	2	1842	26095	2	18.698	0.66 0.517

```

# QQ plot approach
library(MASS)
library(car)
par(mfrow = c(1, 2))
hist(studres(reducedmodel2), breaks = 25, freq = F, col = "cornflowerblue", cex.axis = 1,
      cex.lab = 1, cex.main = 1)
qqPlot(reducedmodel2) + title("QQ plot of StuResiduals vs tQuantiles", cex.main = 1)

```

Histogram of studres(reducedmodel2)**QQ plot of StuResiduals vs tQuantile**

```

## integer(0)

full_data_Transformed[62, ]

##           Country Year Life.expectancy Alcohol Hepatitis.B Measles BMI
## 62  Antigua and Barbuda 2004          74.4    7.28      97     0 4.8
##   Polio Total.expenditure HIV.AIDS      GDP Income.composition.of.resources
## 62    97             4.21 -2.302585 1352.837
##   Schooling
## 62      0

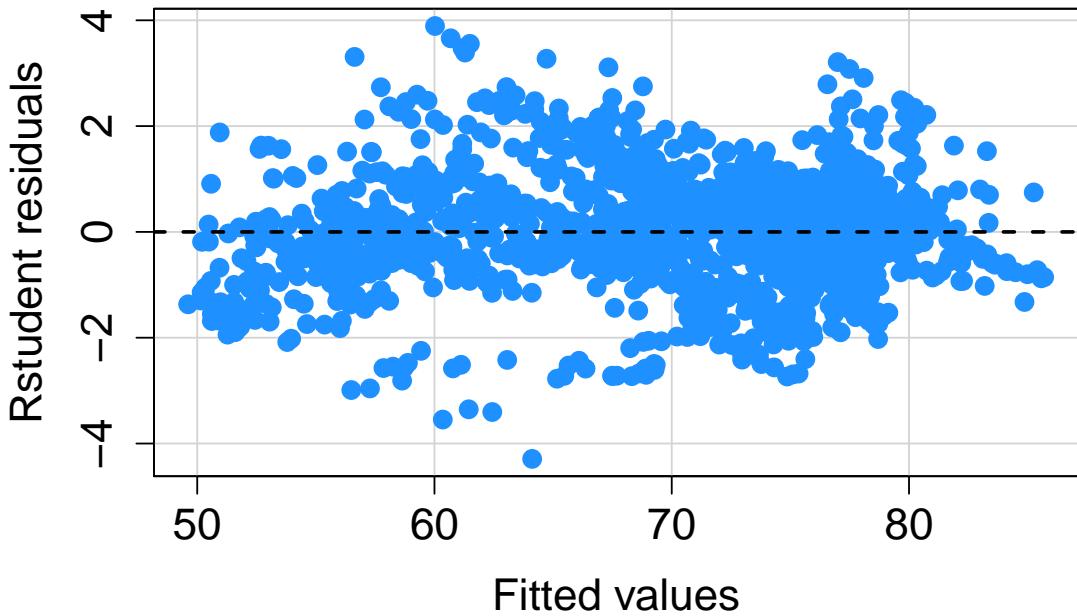
full_data_Transformed[1513, ]

##           Country Year Life.expectancy Alcohol Hepatitis.B Measles   BMI Polio
## 1513 Sierra Leone 2014          48.1    0.01      83 1006 23.8   83
##   Total.expenditure HIV.AIDS      GDP Income.composition.of.resources
## 1513            11.9 -0.5108256 78.43948
##   Schooling
## 1513      9.5

residualPlot(reducedmodel2, type = "rstudent", quadratic = F, col = "dodgerblue",
             pch = 16, cex = 1.5, cex.axis = 1.5, cex.lab = 1.5) + title("Rstudent Residuals vs Fitted values")

```

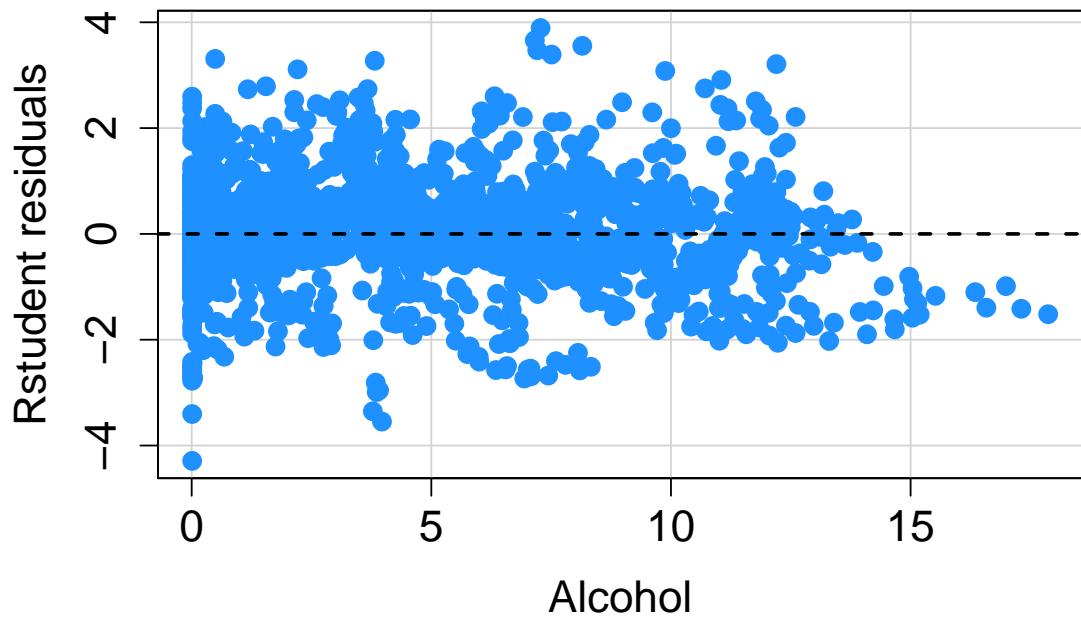
Rstudent Residuals vs Fitted values



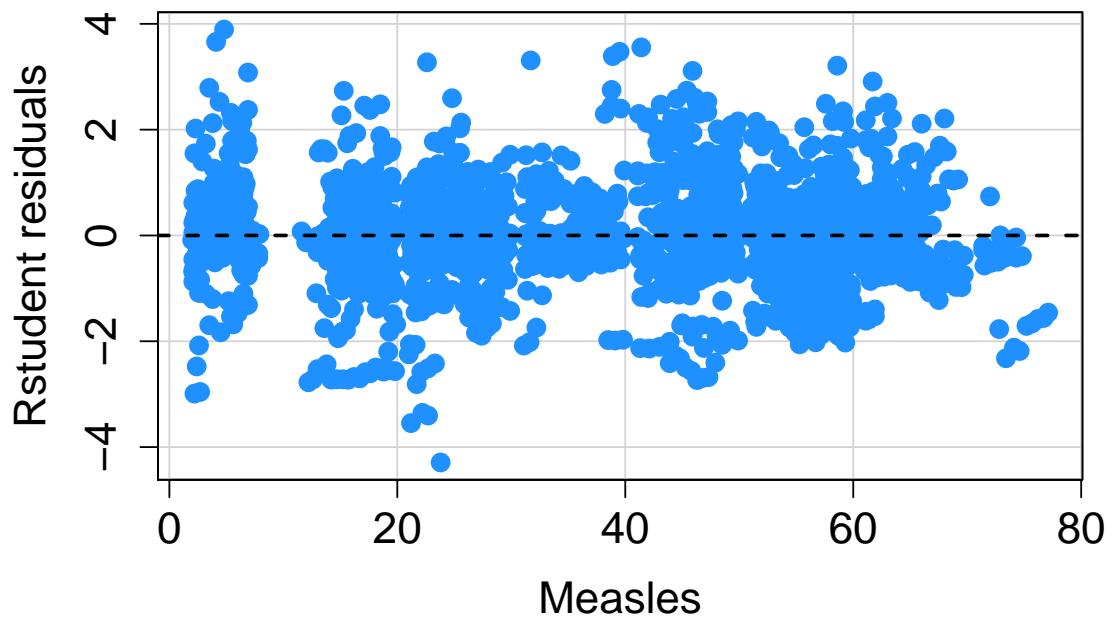
```
## numeric(0)

for (i in 1:8) {
  temp <- paste(c("Rstudent Residuals vs ", names(df4)[i]), collapse = " ")
  residualPlot(reducedmodel2, type = "rstudent", quadratic = F, col = "dodgerblue",
    pch = 16, cex = 1.5, cex.axis = 1.5, cex.lab = 1.5, var = names(df4)[i]) +
    title(temp)
}
```

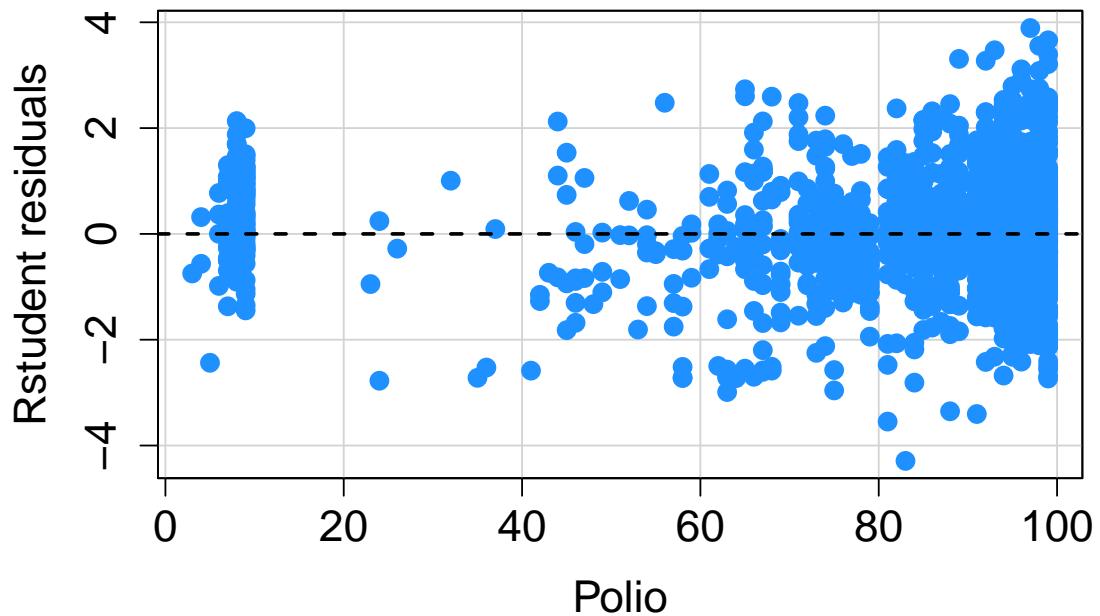
Rstudent Residuals vs Alcohol



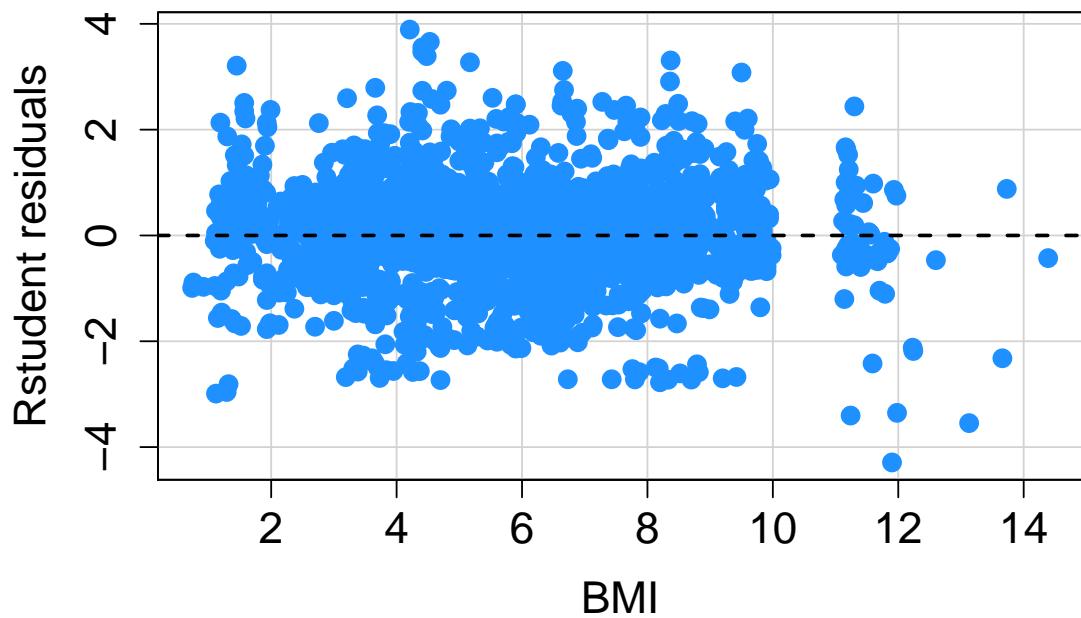
Rstudent Residuals vs Measles



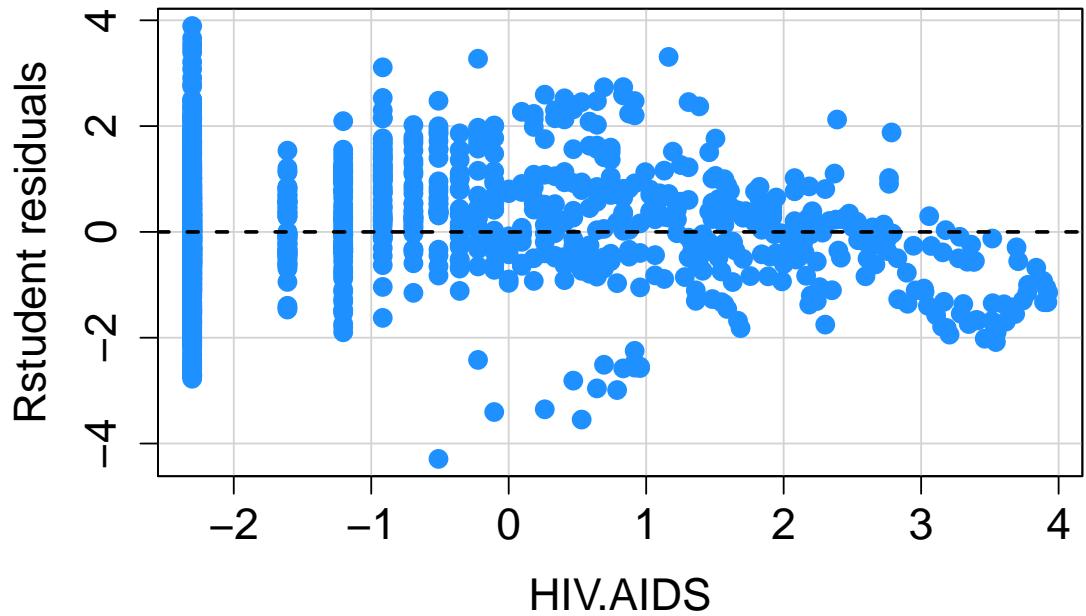
Rstudent Residuals vs Polio



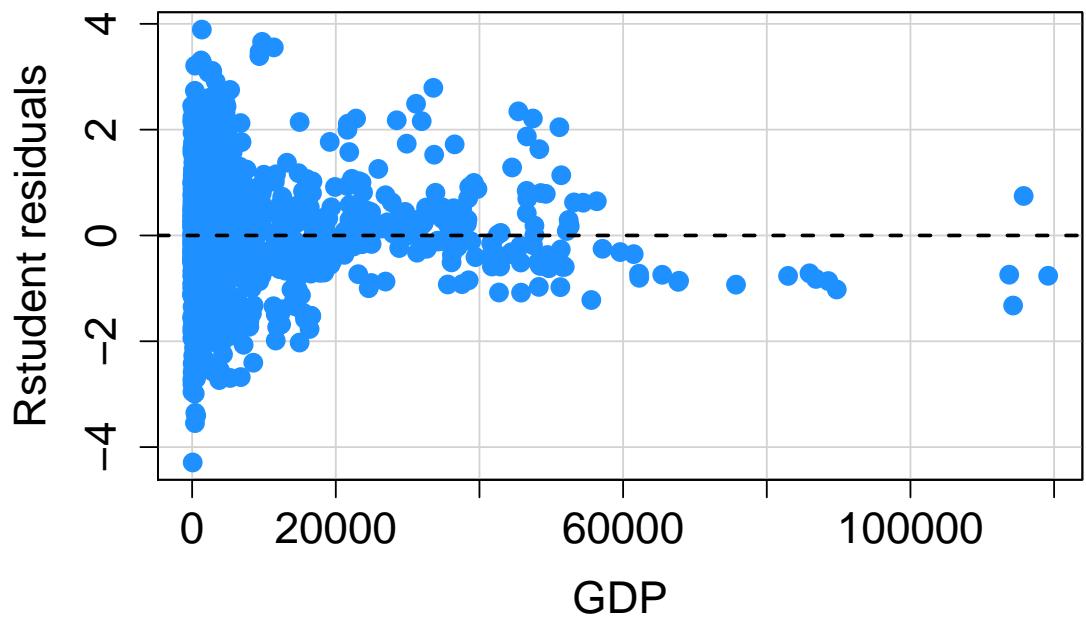
Rstudent Residuals vs BMI



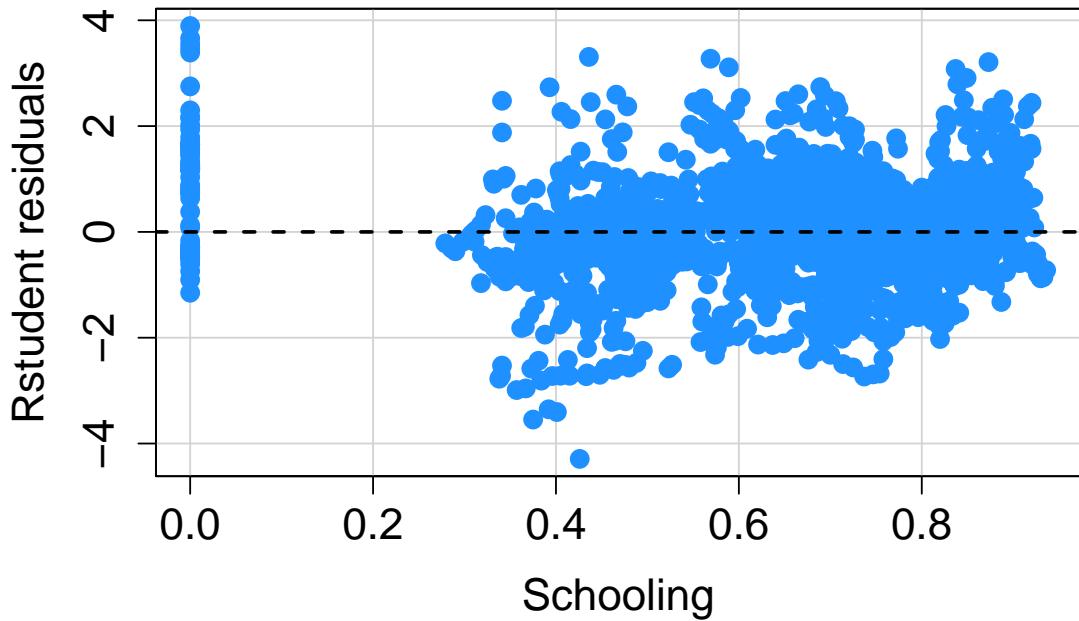
Rstudent Residuals vs HIV.AIDS



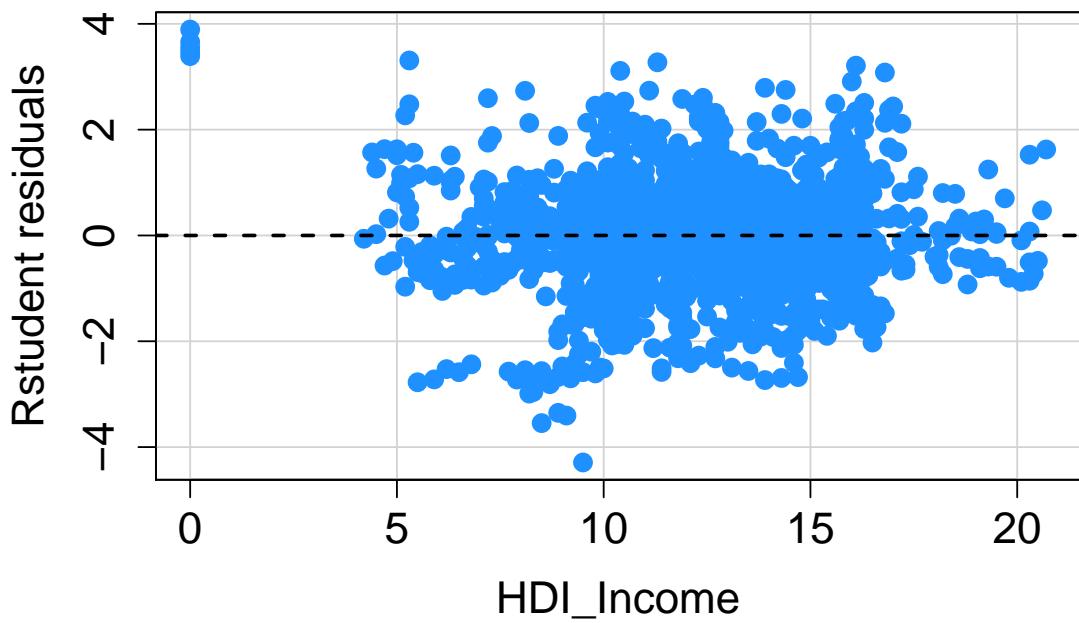
Rstudent Residuals vs GDP



Rstudent Residuals vs Schooling



Rstudent Residuals vs HDI_Income



```
library(MASS)
library(printr)
```

```
## Warning: package 'printr' was built under R version 4.0.5
```

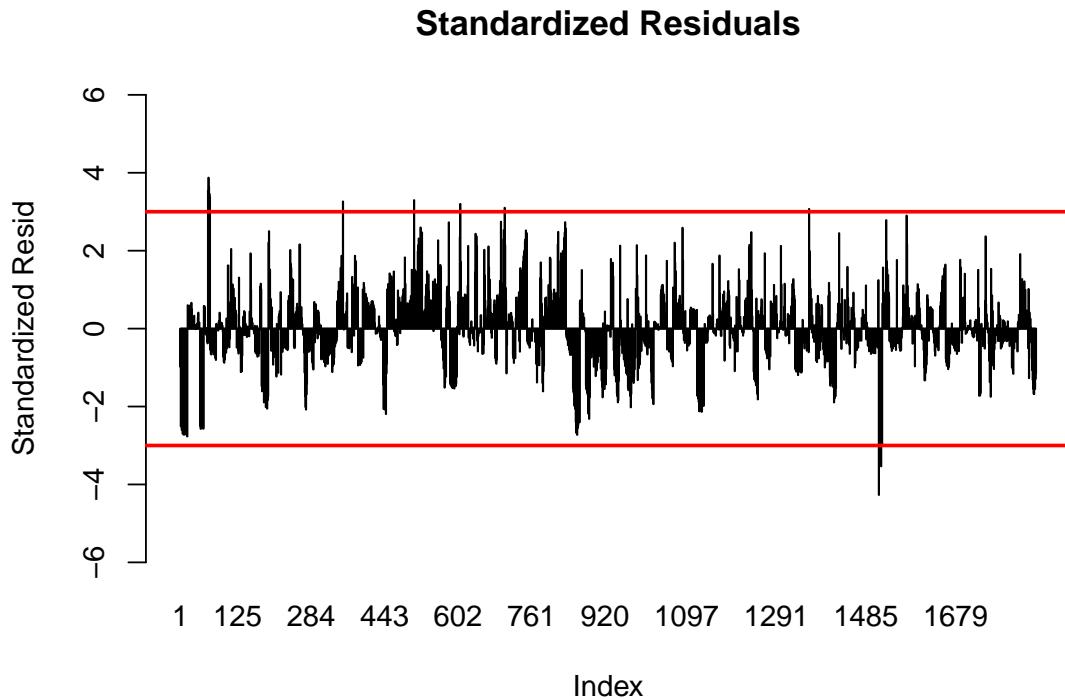
```

stdresiduals <- stdres(reducedmodel2)
range(stdresiduals)

## [1] -4.270536  3.876834

barplot(height = stdresiduals, names.arg = 1:nrow(df4), main = "Standardized Residuals",
       xlab = "Index", ylab = "Standardized Resid", ylim = c(-6, 6))
# Add cutoff values. Either 2 or 3 can be chosen.
abline(h = 3, col = "Red", lwd = 2)
abline(h = -3, col = "Red", lwd = 2)

```



```

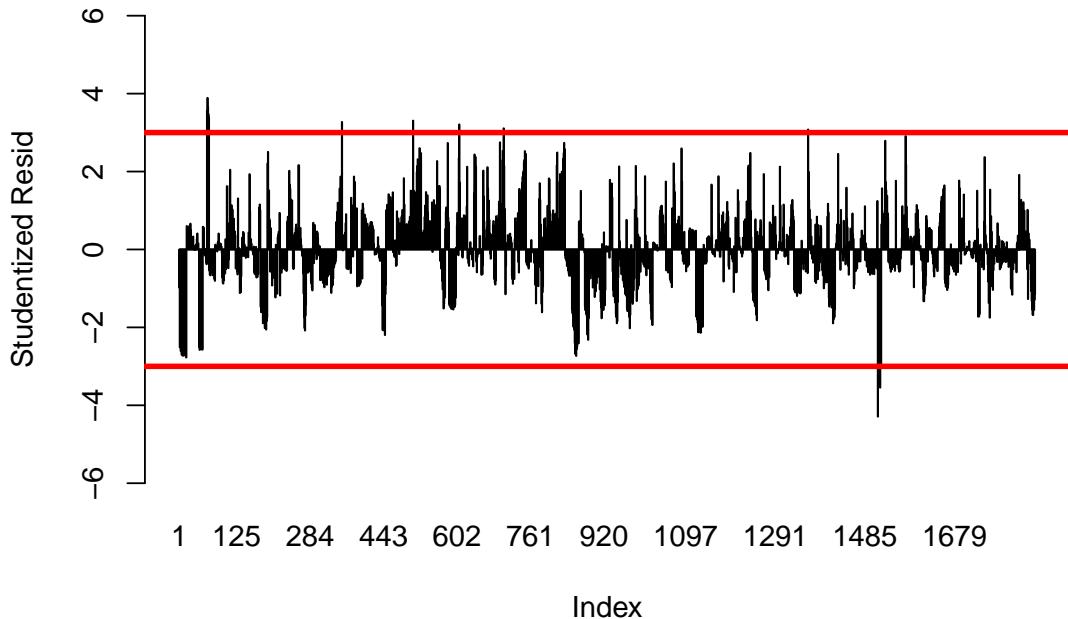
studresiduals <- studres(reducedmodel2)
range(studresiduals)

## [1] -4.290649  3.891675

barplot(height = studresiduals, names.arg = 1:nrow(df4), main = "Studentized Residuals",
       xlab = "Index", ylab = "Studentized Resid", ylim = c(-6, 6))
# Add cutoff values. Either 2 or 3 can be chosen.
abline(h = 3, col = "Red", lwd = 3)
abline(h = -3, col = "Red", lwd = 3)

```

Studentized Residuals

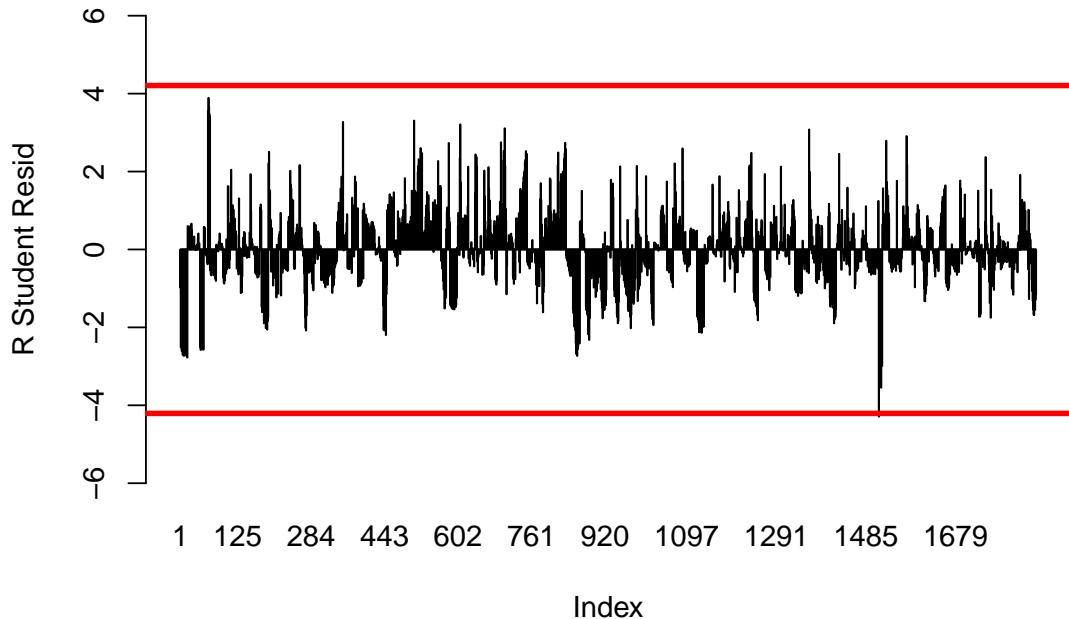


```
RStudent <- rstudent(reducedmodel2)
range(RStudent)

## [1] -4.290649  3.891675

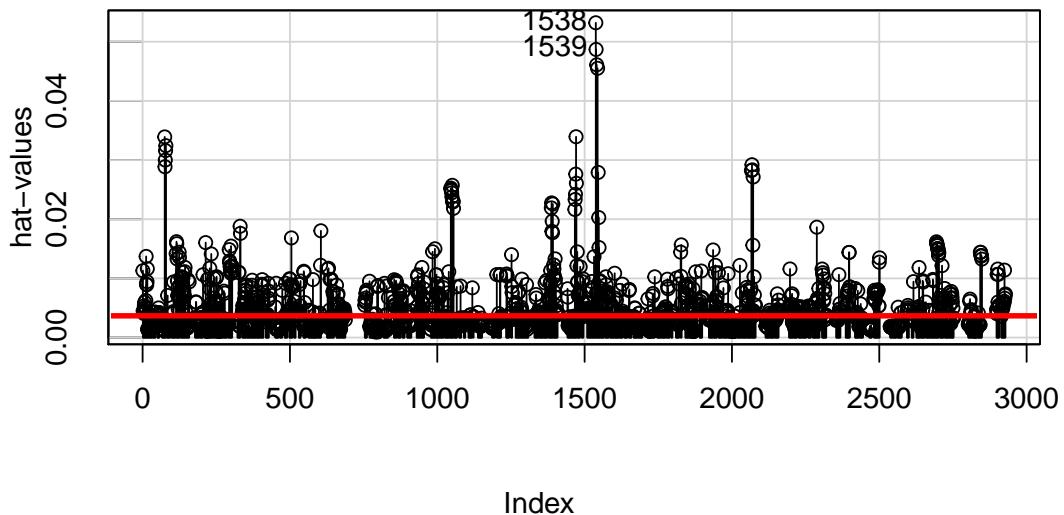
barplot(height = RStudent, names.arg = 1:nrow(df4), main = "R Student Residuals",
       xlab = "Index", ylab = "R Student Resid", ylim = c(-6, 6))
cor.level <- 0.05/(2 * nrow(df4))
cor.qt <- qt(cor.level, nrow(df4) - ncol(df4) - 1, lower.tail = F)
abline(h = cor.qt, col = "Red", lwd = 3)
abline(h = -cor.qt, col = "Red", lwd = 3)
```

R Student Residuals

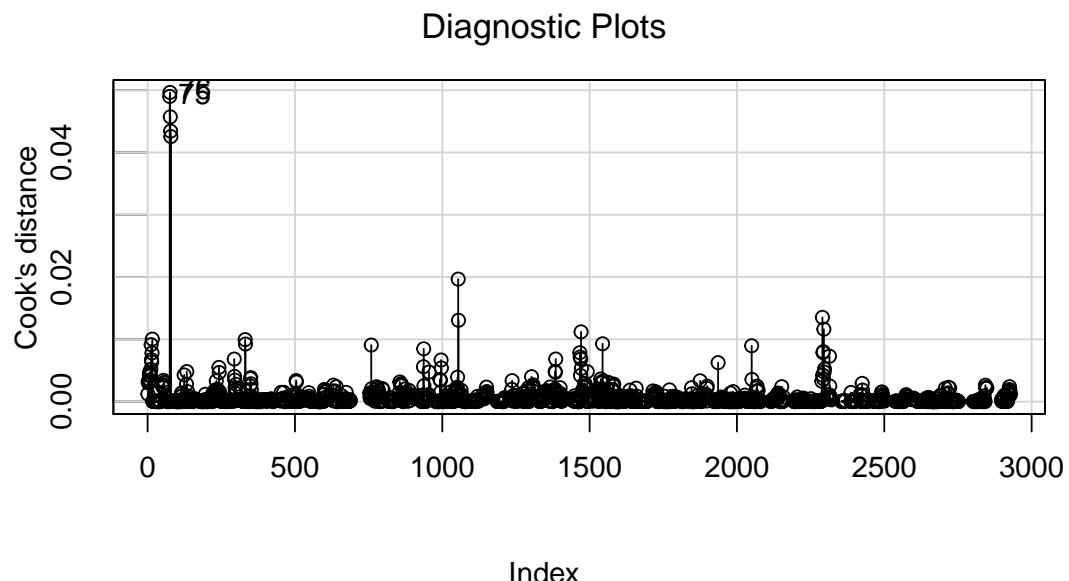


```
influenceIndexPlot(reducedmodel2, vars = c("hat"))
abline(h = 27/1853, col = "Red", lwd = 3)
```

Diagnostic Plots

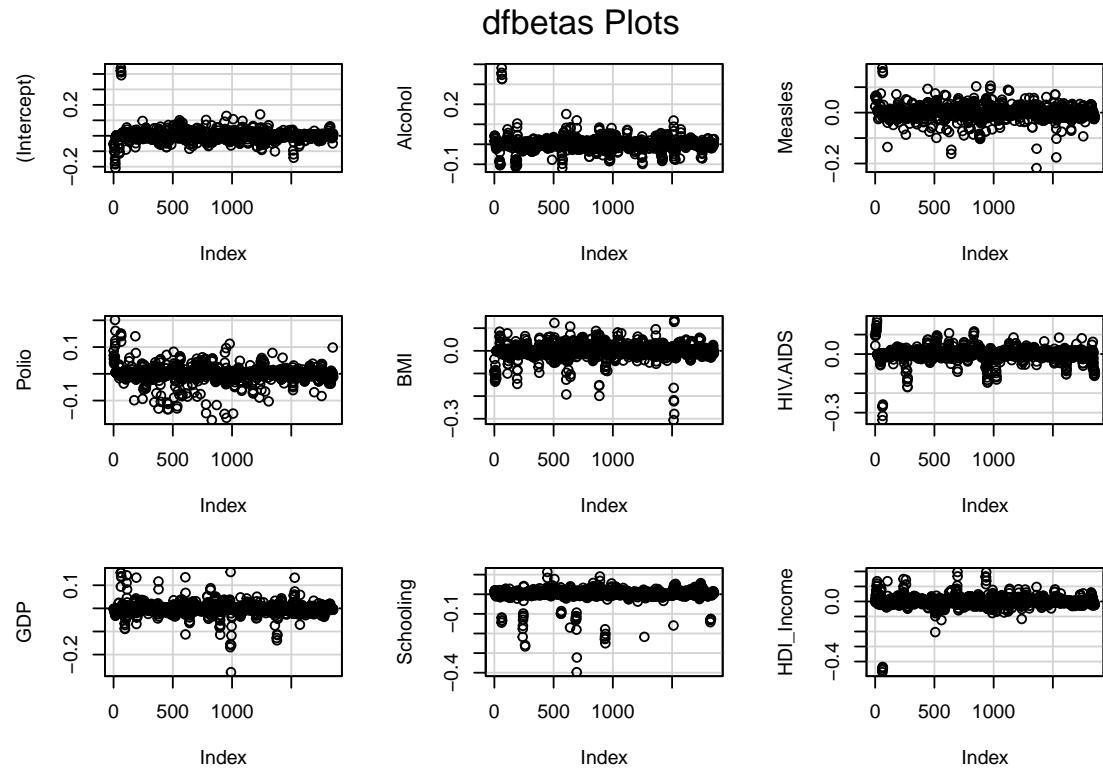


```
influenceIndexPlot(reducedmodel2, vars = c("Cook"))
```



```
myInf <- influence.measures(reducedmodel2)
# summary(myInf)
```

```
library(car)
dfbetasPlots(reducedmodel2, intercept = T)
```



```
# table 1
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 4.0.4
```

```
knitr::kable(plyr::ddply(dplyr::select(full_data_Transformed[which((stdresiduals >
  3 | stdresiduals < -3) == TRUE), ], (Country:Year)), .(Country), summarise, Year = list(sort(unl
  caption = "Potentially Influential Observations of Standardized and Studentized Residual Plots")
```

Table 1: Potentially Influential Observations of Standardized and Studentized Residual Plots

Country	Year
Antigua and Barbuda	2001, 2002, 2003, 2004, 2005
Cabo Verde	2002
Djibouti	2009
France	2007
Guatemala	2008
Portugal	2014
Sierra Leone	2009, 2011, 2012, 2014

```

# table 2
tab <- matrix(c(59, 0, 0, 0, 0, 0, 0, 0, 0, 0, 37, 169), ncol = 13, byrow = TRUE)
colnames(tab) <- c("hat", "cook.d", "dfb.1_", "dfb.Alch", "dfb.Msls", "dfb.Poli",
  "dfb.BMI", "dfb.HIV.", "dfb.GDP", "dfb.Schl", "dfb.HDI_", "dffit", "cov.r")
knitr::kable(tab, caption = "# of Potentially Influential Observations for each Measure")

```

Table 2: # of Potentially Influential Observations for each Measure

hat	cook.d	dfb.1_	dfb.Alch	dfb.Msls	dfb.Poli	dfb.BMI	dfb.HIV.	dfb.GDP	dfb.Schl	dfb.HDI_	dffit	cov.r	
59	0	0	0	0	0	0	0	0	0	0	0	37	169

```

# table 3
knitr::kable(plyr::ddply(dplyr::select(full_data_Transformed[which((hatvalues(reducedmodel2) >
  27/1853) == TRUE), ], (Country:Year)), .(Country), summarise, Year = list(sort(unlist(Year)))), 
  caption = "Potentially Influential Observations of hat Measure")

```

Table 3: Potentially Influential Observations of hat Measure

Country	Year
Antigua and Barbuda	2001, 2002, 2003, 2004, 2005
Australia	2012, 2013
Barbados	2010
Bhutan	2004, 2010
Bosnia and Herzegovina	2004, 2005
Canada	2007
Comoros	2003
Georgia	2000
Grenada	2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010
Kiribati	2000, 2001, 2002, 2003, 2004, 2005
Lebanon	2000, 2001, 2002, 2003, 2004, 2005
Luxembourg	2003, 2004, 2006, 2008, 2011, 2013, 2014
Netherlands	2012
Oman	2000
Qatar	2008, 2009, 2011, 2012, 2014
Seychelles	2000
Turkmenistan	2004, 2005, 2006, 2007, 2008, 2009, 2010

```

# table of diagnostic plot
knitr::kable(plyr::ddply(dplyr::select(full_data_Transformed[61:62, ], (Country:Year)),
  .(Country), summarise, Year = list(sort(unlist(Year)))))

```

Country	Year
Antigua and Barbuda	2004, 2005

```

# table 5
knitr::kable(plyr::ddply(dplyr::select(full_data_Transformed[c(11, 12, 13, 14, 15,
  16, 61, 62, 63, 64, 65, 192, 239, 259, 260, 507, 606, 607, 642, 643, 695, 696,
  886, 933, 934, 936, 937, 938, 991, 1266, 1362, 1513, 1515, 1516, 1518, 1520,
  1529), ], (Country:Year)), .(Country), summarise, Year = list(sort(unlist(Year))))),
caption = "Potentially Influential Observations of DFFIT Measure")

```

Table 5: Potentially Influential Observations of DFFIT Measure

Country	Year
Afghanistan	2000, 2001, 2002, 2003, 2004, 2005
Antigua and Barbuda	2001, 2002, 2003, 2004, 2005
Belgium	2014
Bhutan	2010
Bosnia and Herzegovina	2004, 2005
Djibouti	2009
France	2007, 2008
Germany	2011, 2012
Grenada	2001, 2002
Kiribati	2007
Lebanon	2000, 2001, 2002, 2004, 2005
Luxembourg	2008
Oman	2000
Portugal	2014
Sierra Leone	2007, 2009, 2011, 2012, 2014
Singapore	2006

```

# table 6
knitr::kable(plyr::ddply(dplyr::select(full_data_Transformed[c(2:11, 13:16, 44, 45,
  47:51, 61:65, 94:97, 105, 108, 121, 122, 162, 167, 193, 200, 201, 246:249, 293,
  353, 375, 422, 446, 448, 507, 516, 517, 521:524, 582, 607, 631, 640, 687:694,
  701:704, 747:751, 763, 779, 791, 819, 832:835, 857:865, 888:893, 895, 934:939,
  985, 986, 988, 991, 993, 995, 996, 1071, 1088, 1124, 1128:1130, 1143, 1192, 1202,
  1204, 1232, 1233, 1237, 1239, 1272, 1273, 1362, 1377, 1379, 1380, 1382, 1383,
  1427, 1437, 1512:1520, 1529, 1545, 1562, 1563, 1573, 1620, 1621, 1681, 1734:1742,
  1744, 1750, 1831, 1833, 1835), ], (Country:Year)), .(Country), summarise, Year = list(sort(unlist(
caption = "Potentially Influential Observations of COVRATIO Measure")

```

Table 6: Potentially Influential Observations of COVRATIO Measure

Country	Year
Afghanistan	2000, 2001, 2002, 2003, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014
Angola	2007, 2008, 2009, 2010, 2011, 2013, 2014
Antigua and Barbuda	2001, 2002, 2003, 2004, 2005
Australia	2003, 2011, 2012, 2013, 2014
Austria	2000, 2001, 2014
Bangladesh	2003
Barbados	2010
Belgium	2005, 2006, 2013
Bhutan	2000, 2001, 2002, 2003
Brunei	2012
Darussalam	
Cabo Verde	2002
Canada	2012
Colombia	2014
Comoros	2003, 2005
Djibouti	2009
Dominican Republic	2000, 2001, 2002, 2003, 2007, 2008
Ethiopia	2009
France	2007
Georgia	2008
Germany	2014
Grenada	2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010
Guatemala	2007, 2008, 2009, 2010
Honduras	2000, 2001, 2002, 2003, 2004
Indonesia	2014
Iraq	2013
Ireland	2012
Italy	2004
Jamaica	2003, 2004, 2005, 2006
Kazakhstan	2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008
Kiribati	2000, 2001, 2002, 2003, 2004, 2005
Kuwait	2013
Lebanon	2000, 2001, 2002, 2003, 2004
Lesotho	2014
Luxembourg	2003, 2004, 2006, 2008, 2011, 2013, 2014
Malta	2011
Mauritania	2006

Country	Year
Mongolia	2004, 2005, 2006, 2010
Montenegro	2006
Nepal	2012
Netherlands	2012, 2014
Nicaragua	2002, 2003
Niger	2011, 2013
Pakistan	2008, 2009
Portugal	2014
Qatar	2008, 2009, 2011, 2012, 2014
Rwanda	2008
Samoa	2011
Seychelles	2000
Sierra Leone	2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014
Singapore	2006
Solomon Islands	2005
South Africa	2002, 2003
Spain	2007
Sweden	2011, 2012
Tonga	2005
Turkmenistan	2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010
Uganda	2007, 2013
Zambia	2008, 2010, 2012