Project 3_nlk322

Neil Klenk

May 1, 2016

Project 3

Introduction

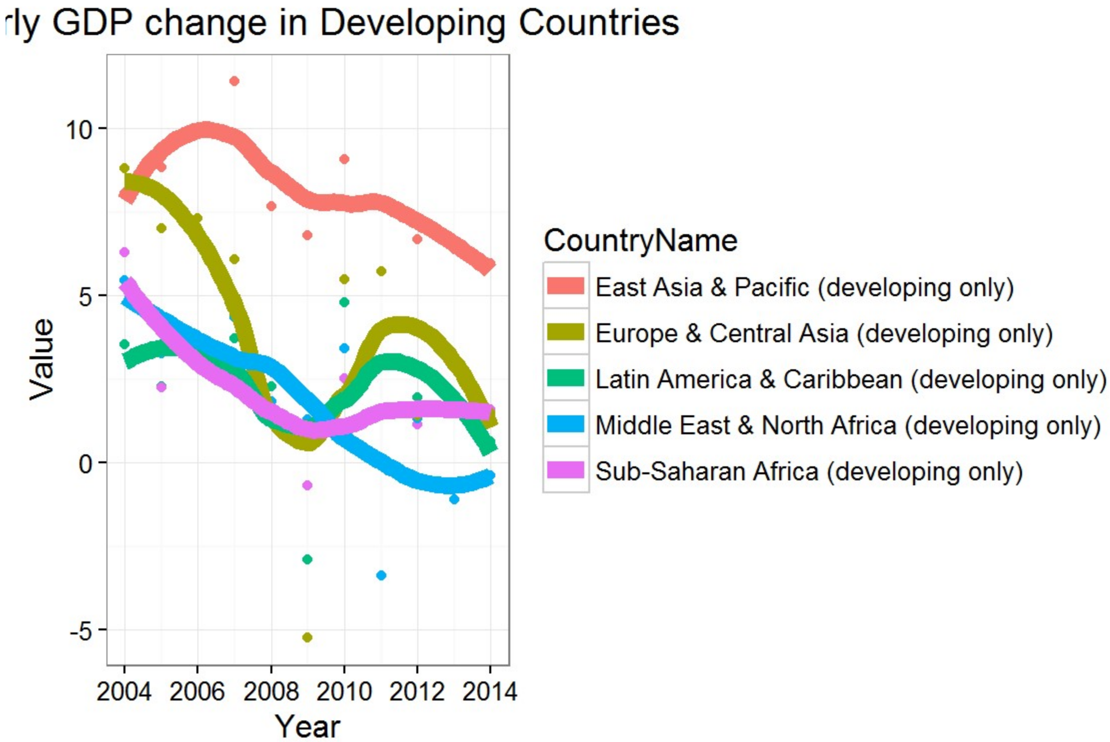
For this project I will be using the World Development Indicators dataset located at https://www.kaggle.com/worldbank/world-development-indicators (https://www.kaggle.com/worldbank/world-development-indicators). This dataset contains over a thousand annual indicators of economic development from hundreds of countries around the world, supplied by the world bank. This is a tidy dataset with each row containg an observation.

```
 Indicators <- read.csv("~/Spring 2016/Bioinformatics/Projects/Project 3/world-
development-indicators-release-2016-01-28-06-31-53/world-development-indicator
s/Indicators.csv")
```

Question 1:

How are the GDP growth trends from 2004 to 2014 different for developing countries worldwide?"

```
#sepparating larger Indicators dataset into 5 smaller sets, each contining a Co
untryCode of interest, and only the relevant GDP infomration.
EAP_GDP <- Indicators %>% filter(CountryCode == "EAP") %>% filter(IndicatorCod
e == "NY.GDP.PCAP.KD.ZG")

ECA_GDP <- Indicators %>% filter(CountryCode == "ECA") %>% filter(IndicatorCod
e == "NY.GDP.PCAP.KD.ZG")

LAC_GDP <- Indicators %>% filter(CountryCode == "LAC") %>% filter(IndicatorCod
e == "NY.GDP.PCAP.KD.ZG")

MNA_GDP <- Indicators %>% filter(CountryCode == "MNA") %>% filter(IndicatorCod
e == "NY.GDP.PCAP.KD.ZG")

SSA_GDP <- Indicators %>% filter(CountryCode == "SSA") %>% filter(IndicatorCod
e == "NY.GDP.PCAP.KD.ZG")

#Combining all of the GDP information into a single dataframe
My_GDP <- union(EAP_GDP, ECA_GDP)
My_GDP <- union(My_GDP, LAC_GDP)
My_GDP <- union(My_GDP, MNA_GDP)
My_GDP <- union(My_GDP, SSA_GDP)

#Filtering for timespan of interest
My_GDP_I <- My_GDP %>% filter(2003 < Year, Year < 2015)

#Plotting The yearly change in GDP by country
GDP_Plot <- My_GDP_I %>% ggplot(aes(x = Year, y = Value, group = CountryCode, c
olor = CountryName)) + geom_point() + geom_smooth(se = FALSE, size = 3) + ggtit
le("Yearly GDP change in Developing Countries")

GDP_Plot
```



This question required the comparison of GDP growth over time from multiple different groups. A scatterplot with lines of best fit was choosen because it would give the reader a good sense of both the volatility of this measure, while giving them a way to easily visualize overall trends of the dfferent groups. The overall “Indicators” dataset was filtered by country code and to only include the information about the GDP change each year. All of the resulting datasets were then joined using union() to create a new single dataset with all if the neede information. Value was then plotted against the year in a scatterplot with geom_smooth() providing a smoothed conditional mean (loess method). Each developing region recieved its own color for easy differentiation.

By directly comparing the change in GDP over time, it is easy to see any deviation between the regions. The smoothed conditional mean gives a better look at the overall trend of the GDP, which is especially volatile in developing regions. The loess method was appropriate to use because it deterines the value at point x from the points neghiboring it, weighted by their distance, curbing some of the volatility.

The East Asia & Pacific region is the most dramatically different, staying very positive over the plotted timeframe. The rest of the regions approach 0 change in GDP around 2009, and again in 2014. Of those that approached 0, the Middle East and North African region was the only to not experience a period of growth prior to falling back to 0.

From 2004 to 2014, Developing regions around the world Experienced varied changes in their GDP. While regions like Europe and Central Asia had GDP growth patterns nearly idential to Latin America and Carribean, East Asia and Pacific region can be seen to have much better GDP growth than any of the other regions. The Middle East and North Africa region followed the overall downward trend of the others, but missed the upward swing the others got starting in 2009.

Question 2:

Should the developing countries be focusing on Taxes on International Trade, Interest Payments, Business Extent of Disclosure Index, and New Business Density to help increase GDP growth?

```
##########East Asia & Pacific Region###########
#Filtering out unecessary info, sepparating data set into different Indicator codes, and rearranging resulting table
EAP_GDP_reg_0 <- Indicators %>% filter(CountryCode == "EAP") %>% filter(IndicatorCode == "NY.GDP.PCAP.KD.ZG") %>% filter(2003 < Year, Year < 2015) %>% group_by(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
EAP_GDP_reg_1 <- Indicators %>% filter(CountryCode == "EAP") %>% filter(IndicatorCode == "GC.TAX.INTT.RV.ZS") %>% filter(2003 < Year, Year < 2015) %>% group_by(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
EAP_GDP_reg_2 <- Indicators %>% filter(CountryCode == "EAP") %>% filter(IndicatorCode == "GC.XPN.INTP.RV.ZS") %>% filter(2003 < Year, Year < 2015) %>% group_by(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
EAP_GDP_reg_3 <- Indicators %>% filter(CountryCode == "EAP") %>% filter(IndicatorCode == "IC.BUS.DISC.XQ") %>% filter(2003 < Year, Year < 2015) %>% group_by(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
EAP_GDP_reg_4 <- Indicators %>% filter(CountryCode == "EAP") %>% filter(IndicatorCode == "IC.BUS.NDNS.ZS") %>% filter(2003 < Year, Year < 2015) %>% group_by(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)

#Combining all the resulting data sets into a larger one for the linear regression
EAP_GDP_reg <- full_join(EAP_GDP_reg_0, EAP_GDP_reg_1)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```
EAP_GDP_reg <- full_join(EAP_GDP_reg, EAP_GDP_reg_2)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```
EAP_GDP_reg <- full_join(EAP_GDP_reg, EAP_GDP_reg_3)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```
EAP_GDP_reg <- full_join(EAP_GDP_reg, EAP_GDP_reg_4)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```
#imputing the missing data
EAP_Imp_temp <- EAP_GDP_reg %>% mice(m=5, method = "pmm")
```

```
##
##   iter imp variable
##    1    1   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    1    2   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    1    3   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    1    4   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    1    5   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    2    1   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    2    2   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    2    3   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    2    4   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    2    5   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    3    1   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    3    2   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    3    3   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    3    4   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    3    5   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    4    1   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    4    2   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    4    3   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    4    4   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    4    5   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    5    1   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    5    2   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
##    5    3   GC.TAX.INTT.RV.ZS   GC.XPN.INTP.RV.ZS   IC.BUS.DISC.XQ   IC.BUS.NDNS.Z
S
```

```
##   5   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   5   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
```

```
#Performing Linear Regression
EAP_glm.out <- with(EAP_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.TAX.INTT.RV.ZS + G
C.XPN.INTP.RV.ZS + IC.BUS.DISC.XQ + IC.BUS.NDNS.ZS))

summary(pool(EAP_glm.out))
```

```
##                           est         se          t       df  Pr(>|t|)
## (Intercept)       19.4935347 17.2682150  1.1288680 3.023274 0.3404989
## GC.TAX.INTT.RV.ZS  0.1018030  0.8359156  0.1217862 4.389337 0.9084434
## GC.XPN.INTP.RV.ZS  0.3346189  0.5866409  0.5703983 3.578169 0.6022944
## IC.BUS.DISC.XQ    -3.3357329  4.1402134 -0.8056911 3.296340 0.4744859
## IC.BUS.NDNS.ZS     0.2463052  1.6358771  0.1505646 3.618867 0.8883324
##                       lo 95     hi 95 nmis       fmi    lambda
## (Intercept)       -35.223076 74.210146   NA 0.5348579 0.3036322
## GC.TAX.INTT.RV.ZS  -2.140094  2.343700    3 0.3116208 0.0561611
## GC.XPN.INTP.RV.ZS  -1.372753  2.041991    5 0.4457997 0.2036940
## IC.BUS.DISC.XQ    -15.866331  9.194866    1 0.4909517 0.2539833
## IC.BUS.NDNS.ZS     -4.490633  4.983244    1 0.4392683 0.1964677
```

```
#removing worst predictor
EAP_glm.out <- with(EAP_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.TAX.INTT.RV.ZS + I
C.BUS.DISC.XQ + IC.BUS.NDNS.ZS))

summary(pool(EAP_glm.out))
```

```
##                           est         se          t       df  Pr(>|t|)
## (Intercept)       24.8162444 12.3287725  2.0128723 4.082600 0.1129990
## GC.TAX.INTT.RV.ZS -0.1840360  0.6954935 -0.2646121 5.025346 0.8018185
## IC.BUS.DISC.XQ    -3.8843653  3.5311317 -1.1000341 4.461266 0.3271012
## IC.BUS.NDNS.ZS     0.4268284  1.4337488  0.2977009 4.873185 0.7782037
##                       lo 95     hi 95 nmis       fmi     lambda
## (Intercept)        -9.142510 58.774999   NA 0.4470995 0.22953355
## GC.TAX.INTT.RV.ZS  -1.969150  1.601078    3 0.3188847 0.09280127
## IC.BUS.DISC.XQ    -13.301240  5.532510    1 0.3962366 0.17512909
## IC.BUS.NDNS.ZS     -3.287775  4.141432    1 0.3401342 0.11542955
```

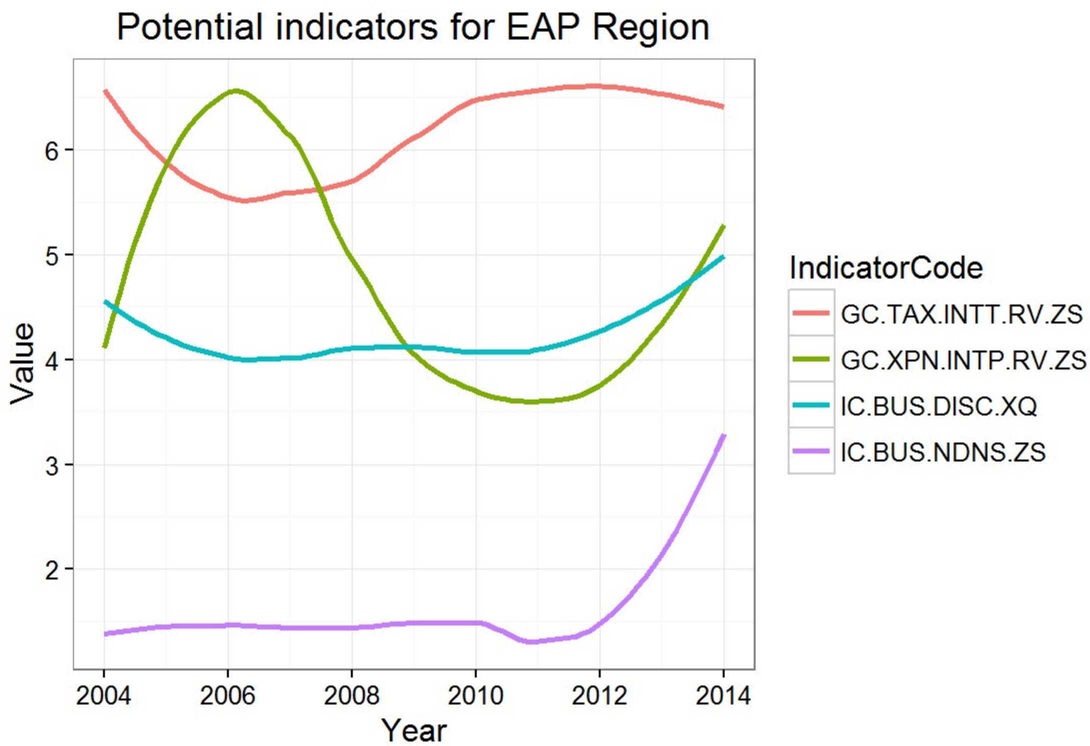
```
#removing worst predictor
EAP_glm.out <- with(EAP_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.TAX.INTT.RV.ZS + I
C.BUS.DISC.XQ))

summary(pool(EAP_glm.out))
```

```
##                          est       se          t       df   Pr(>|t|)
## (Intercept)       21.7818248 7.208893  3.0215215 5.427263 0.02647022
## GC.TAX.INTT.RV.ZS -0.2317202 0.608759 -0.3806436 5.324570 0.71818440
## IC.BUS.DISC.XQ    -2.9137211 1.497925 -1.9451720 6.172853 0.09837334
##                       lo 95      hi 95 nmis       fmi     lambda
## (Intercept)        3.681089 39.8825607   NA 0.3487545 0.14610350
## GC.TAX.INTT.RV.ZS -1.768325  1.3048842    3 0.3606055 0.15841164
## IC.BUS.DISC.XQ    -6.554275  0.7268325    1 0.2593565 0.05284357
```

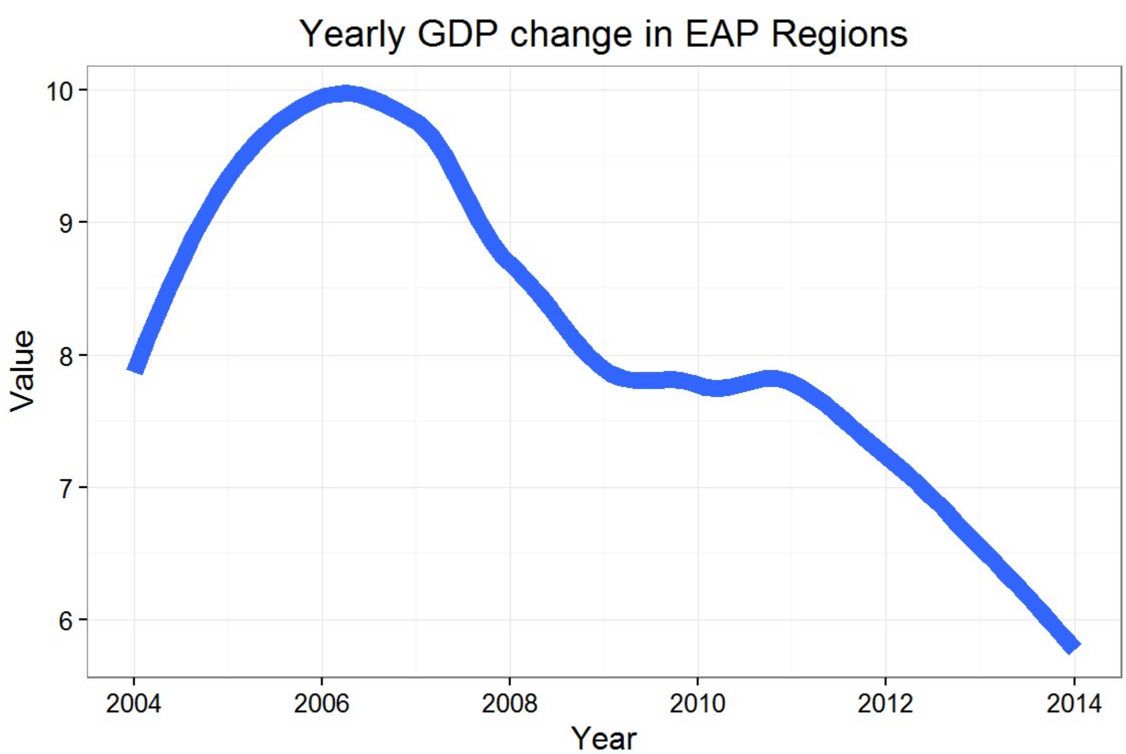
```
#removing worst predictor
EAP_glm.out <- with(EAP_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ IC.BUS.DISC.XQ ))

summary(pool(EAP_glm.out))
```

```
##                      est       se         t       df    Pr(>|t|)     lo 95
## (Intercept)    20.267950 5.838133  3.471649 7.171742 0.009992726  6.529694
## IC.BUS.DISC.XQ -2.889077 1.381155 -2.091784 7.116155 0.074125494 -6.144214
##                    hi 95 nmis       fmi     lambda
## (Intercept)    34.0062056   NA 0.2294750 0.04089220
## IC.BUS.DISC.XQ  0.3660597    1 0.2357128 0.04737555
```

```
#Creating a dataset that includes the imputed values
EAP_Imp <- complete(EAP_Imp_temp, 1)

#Rearranging the table
EAP_GDP_regfix <- EAP_Imp %>% gather("IndicatorCode", "Value", 4:8)

#Setting Margins
par(mar=c(6,6,5,3))


#Plotting the variables and the overall GDP curve
ggplot(EAP_GDP_regfix %>% filter(IndicatorCode != "NY.GDP.PCAP.KD.ZG"), aes(x
= Year, y = Value, group = IndicatorCode, color = IndicatorCode)) + geom_smooth
(se = FALSE, size = 1) + ggtitle("Potential indicators for EAP Region")
```




```
EAP_GDP_Plot <- My_GDP_I %>% filter(CountryCode == "EAP") %>% ggplot(aes(x = Year, y = Value, group = CountryCode)) + geom_smooth(se = FALSE, size = 3) + ggtitle("Yearly GDP change in EAP Regions")

EAP_GDP_Plot
```



```r
##########Europe and Central Asia Region##########
#Filtering out unecessary info, sepparating data set into different Indicator c
odes, and rearranging resulting table
ECA_GDP_reg_0 <- Indicators %>% filter(CountryCode == "ECA") %>% filter(Indicat
orCode == "NY.GDP.PCAP.KD.ZG") %>% filter(2003 < Year, Year < 2015) %>% group_b
y(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
ECA_GDP_reg_1 <- Indicators %>% filter(CountryCode == "ECA") %>% filter(Indicat
orCode == "GC.TAX.INTT.RV.ZS") %>% filter(2003 < Year, Year < 2015) %>% group_b
y(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
ECA_GDP_reg_2 <- Indicators %>% filter(CountryCode == "ECA") %>% filter(Indicat
orCode == "GC.XPN.INTP.RV.ZS") %>% filter(2003 < Year, Year < 2015) %>% group_b
y(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
ECA_GDP_reg_3 <- Indicators %>% filter(CountryCode == "ECA") %>% filter(Indicat
orCode == "IC.BUS.DISC.XQ") %>% filter(2003 < Year, Year < 2015) %>% group_by(Y
ear) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
ECA_GDP_reg_4 <- Indicators %>% filter(CountryCode == "ECA") %>% filter(Indicat
orCode == "IC.BUS.NDNS.ZS") %>% filter(2003 < Year, Year < 2015) %>% group_by(Y
ear) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)

#Combining all the resulting data sets into a larger one for the linear regress
ion
ECA_GDP_reg <- full_join(ECA_GDP_reg_0, ECA_GDP_reg_1)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```r
ECA_GDP_reg <- full_join(ECA_GDP_reg, ECA_GDP_reg_2)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```r
ECA_GDP_reg <- full_join(ECA_GDP_reg, ECA_GDP_reg_3)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```r
ECA_GDP_reg <- full_join(ECA_GDP_reg, ECA_GDP_reg_4)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```r
#imputing the missing data
ECA_Imp_temp <- ECA_GDP_reg %>% mice(m=5, method = "pmm")
```

```
##
##  iter imp variable
##   1   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   1   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   1   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   1   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   1   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   2   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   2   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   2   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   2   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   2   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   3   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   3   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   3   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   3   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   3   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   4   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   4   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   4   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   4   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   4   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   5   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   5   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   5   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   5   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   5   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
```

```
#performing the linear regression
ECA_glm.out <- with(ECA_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.TAX.INTT.RV.ZS + G
C.XPN.INTP.RV.ZS + IC.BUS.DISC.XQ + IC.BUS.NDNS.ZS))

summary(pool(ECA_glm.out))
```

```
##                          est        se          t       df  Pr(>|t|)
## (Intercept)        2.4122405 16.096580  0.1498604 4.430367 0.8874536
## GC.TAX.INTT.RV.ZS  1.0353827  1.739328  0.5952773 4.265399 0.5818116
## GC.XPN.INTP.RV.ZS  2.3295778  1.829609  1.2732654 2.716182 0.3009737
## IC.BUS.DISC.XQ    -1.8642043  2.088765 -0.8924912 3.561961 0.4283078
## IC.BUS.NDNS.ZS     0.8912273  2.360268  0.3775958 3.479722 0.7276105
##                       lo 95     hi 95 nmis       fmi     lambda
## (Intercept)       -40.617878 45.442359   NA 0.3043829 0.04818761
## GC.TAX.INTT.RV.ZS  -3.677542  5.748308    3 0.3330707 0.07974539
## GC.XPN.INTP.RV.ZS  -3.852849  8.512004    3 0.5847952 0.36133747
## IC.BUS.DISC.XQ     -7.955936  4.227527    1 0.4483991 0.20657289
## IC.BUS.NDNS.ZS     -6.067035  7.849489    0 0.4615782 0.22119645
```

```
#removing the worst predictor
ECA_glm.out <- with(ECA_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.TAX.INTT.RV.ZS + G
C.XPN.INTP.RV.ZS + IC.BUS.DISC.XQ))

summary(pool(ECA_glm.out))
```

```
##                          est        se          t       df  Pr(>|t|)
## (Intercept)         4.542163 14.312738  0.3173511 5.287157 0.7631296
## GC.TAX.INTT.RV.ZS  1.117766  1.536455  0.7274968 5.487746 0.4967691
## GC.XPN.INTP.RV.ZS  2.052644  1.447652  1.4179127 4.134094 0.2269886
## IC.BUS.DISC.XQ    -1.813926  1.905105 -0.9521394 4.452099 0.3898058
##                        lo 95     hi 95 nmis       fmi     lambda
## (Intercept)       -31.656916 40.741242   NA 0.2811084 0.05242272
## GC.TAX.INTT.RV.ZS  -2.728533  4.964066    3 0.2505636 0.01953219
## GC.XPN.INTP.RV.ZS  -1.915785  6.021073    3 0.4401850 0.22210763
## IC.BUS.DISC.XQ     -6.898102  3.270250    1 0.3974713 0.17644490
```

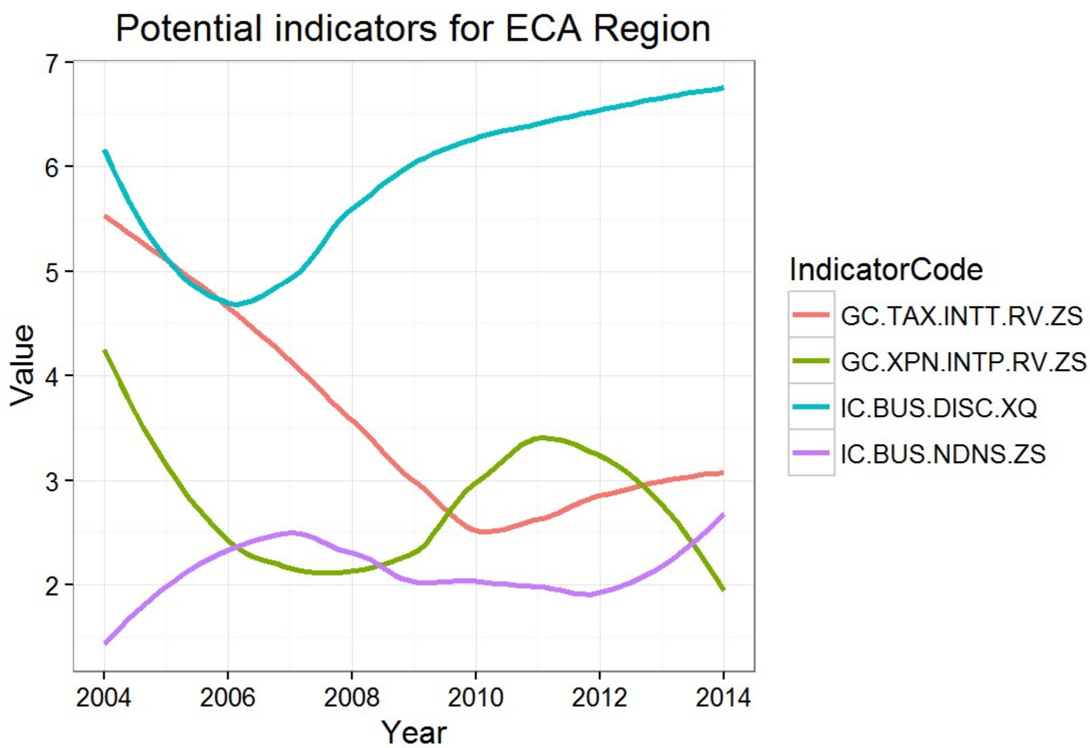
```
#removing the worst predictor
ECA_glm.out <- with(ECA_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.XPN.INTP.RV.ZS + I
C.BUS.DISC.XQ))

summary(pool(ECA_glm.out))
```

```
##                        est       se         t       df  Pr(>|t|)     lo 95
## (Intercept)       11.10493 7.877537  1.409696 6.394086 0.2053667 -7.886359
## GC.XPN.INTP.RV.ZS  2.30418 1.483524  1.553180 4.109687 0.1934614 -1.771748
## IC.BUS.DISC.XQ    -2.39601 1.313564 -1.824053 5.653555 0.1209652 -5.658540
##                        hi 95 nmis       fmi     lambda
## (Intercept)       30.0962246   NA 0.2304880 0.02234547
## GC.XPN.INTP.RV.ZS  6.3801087    3 0.5008865 0.30552677
## IC.BUS.DISC.XQ     0.8665192    1 0.3223923 0.11870944
```

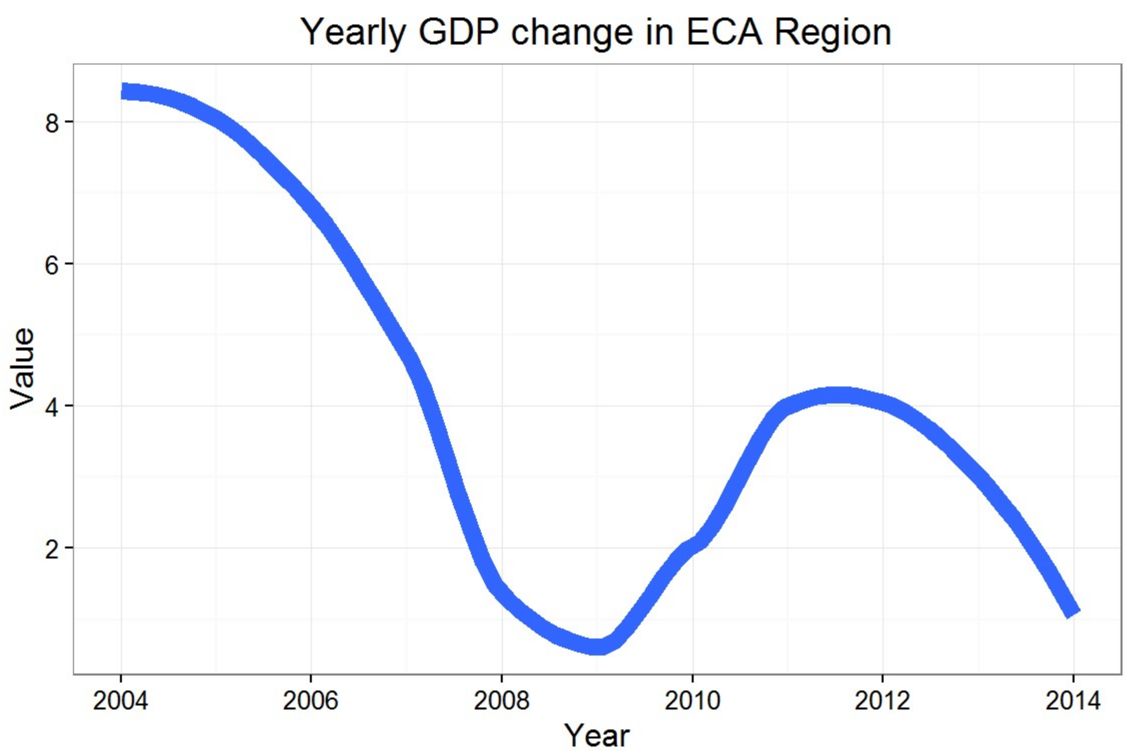
```
#removing the worst predictor
ECA_glm.out <- with(ECA_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ IC.BUS.DISC.XQ))

summary(pool(ECA_glm.out))
```

```
##                      est       se         t       df   Pr(>|t|)     lo 95
## (Intercept)    15.733247 8.181538  1.923018 6.747865 0.09745877 -3.760526
## IC.BUS.DISC.XQ -2.015437 1.396933 -1.442759 6.616007 0.19470900 -5.357923
##                    hi 95 nmis       fmi     lambda
## (Intercept)    35.227020   NA 0.2753519 0.08829441
## IC.BUS.DISC.XQ  1.327049    1 0.2890260 0.10232088
```

```
#Creating a dataset that includes the imputed values
ECA_Imp <- complete(ECA_Imp_temp, 1)

#Rearranging the table
ECA_GDP_regfix <- ECA_Imp %>% gather("IndicatorCode", "Value", 4:8)

#Plotting the variables and the overall GDP curve
ggplot(ECA_GDP_regfix %>% filter(IndicatorCode != "NY.GDP.PCAP.KD.ZG"), aes(x
= Year, y = Value, group = IndicatorCode, color = IndicatorCode)) + geom_smooth
(se = FALSE, size = 1) + ggtitle("Potential indicators for ECA Region")
```




```
ECA_GDP_Plot <- My_GDP_I %>% filter(CountryCode == "ECA") %>% ggplot(aes(x = Ye
ar, y = Value, group = CountryCode)) + geom_smooth(se = FALSE, size = 3) + ggti
tle("Yearly GDP change in ECA Region")

ECA_GDP_Plot
```




```
##########Latin America and Carribean Region##########
#Filtering out unecessary info, sepparating data set into different Indicator c
odes, and rearranging resulting table
LAC_GDP_reg_0 <- Indicators %>% filter(CountryCode == "LAC") %>% filter(Indicat
orCode  == "NY.GDP.PCAP.KD.ZG") %>% filter(2003 < Year, Year < 2015) %>% group_
by(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
LAC_GDP_reg_1 <- Indicators %>% filter(CountryCode == "LAC") %>% filter(Indicat
orCode  == "GC.TAX.INTT.RV.ZS") %>% filter(2003 < Year, Year < 2015) %>% group_
by(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
LAC_GDP_reg_2 <- Indicators %>% filter(CountryCode == "LAC") %>% filter(Indicat
orCode  == "GC.XPN.INTP.RV.ZS") %>% filter(2003 < Year, Year < 2015) %>% group_
by(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
LAC_GDP_reg_3 <- Indicators %>% filter(CountryCode == "LAC") %>% filter(Indicat
orCode  == "IC.BUS.DISC.XQ") %>% filter(2003 < Year, Year < 2015) %>% group_by
(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
LAC_GDP_reg_4 <- Indicators %>% filter(CountryCode == "LAC") %>% filter(Indicat
orCode  == "IC.BUS.NDNS.ZS") %>% filter(2003 < Year, Year < 2015) %>% group_by
(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)

#Combining all the resulting data sets into a larger one for the linear regress
ion
LAC_GDP_reg <- full_join(LAC_GDP_reg_0, LAC_GDP_reg_1)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```r
LAC_GDP_reg <- full_join(LAC_GDP_reg, LAC_GDP_reg_2)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```r
LAC_GDP_reg <- full_join(LAC_GDP_reg, LAC_GDP_reg_3)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```r
LAC_GDP_reg <- full_join(LAC_GDP_reg, LAC_GDP_reg_4)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```r
#imputing the missing data
LAC_Imp_temp <- LAC_GDP_reg %>% mice(m=5, method = "pmm")
```

```
##
##  iter imp variable
##   1   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   1   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   1   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   1   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   1   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   2   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   2   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   2   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   2   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   2   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   3   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   3   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   3   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   3   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   3   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   4   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   4   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   4   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   4   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   4   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   5   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   5   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   5   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
```

```
##   5   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   5   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
```

```
#performing the linear regression
LAC_glm.out <- with(LAC_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.TAX.INTT.RV.ZS + G
C.XPN.INTP.RV.ZS + IC.BUS.DISC.XQ + IC.BUS.NDNS.ZS))

summary(pool(LAC_glm.out))
```

```
##                           est         se           t       df  Pr(>|t|)
## (Intercept)       18.2233051 40.1892617  0.45343717 3.587566 0.6762884
## GC.TAX.INTT.RV.ZS  0.8773693  1.2210536  0.71853459 4.448633 0.5083710
## GC.XPN.INTP.RV.ZS -0.5931967  0.7449721 -0.79626703 3.929785 0.4712076
## IC.BUS.DISC.XQ    -3.9655013  8.9851538 -0.44133927 3.590237 0.6842261
## IC.BUS.NDNS.ZS    -0.1200536  2.1856539 -0.05492799 3.567657 0.9591315
##                        lo 95      hi 95 nmis       fmi    lambda
## (Intercept)       -98.605740 135.052350   NA 0.4442922 0.2020253
## GC.TAX.INTT.RV.ZS  -2.382148   4.136887    2 0.3011356 0.0446072
## GC.XPN.INTP.RV.ZS  -2.676227   1.489834    2 0.3889884 0.1411027
## IC.BUS.DISC.XQ    -30.076320  22.145318    1 0.4438637 0.2015510
## IC.BUS.NDNS.ZS     -6.489722   6.249615    1 0.4474857 0.2055611
```

```
#removing the worst predictor
LAC_glm.out <- with(LAC_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.TAX.INTT.RV.ZS + G
C.XPN.INTP.RV.ZS + IC.BUS.DISC.XQ))

summary(pool(LAC_glm.out))
```

```
##                           est         se          t       df  Pr(>|t|)
## (Intercept)       15.3337022 20.9036477  0.7335419 4.777535 0.4976172
## GC.TAX.INTT.RV.ZS  0.8736640  1.0397182  0.8402893 5.556444 0.4354042
## GC.XPN.INTP.RV.ZS -0.5160741  0.5564272 -0.9274782 5.053400 0.3958182
## IC.BUS.DISC.XQ    -3.4585762  5.3031734 -0.6521711 4.642181 0.5451905
##                        lo 95      hi 95 nmis       fmi      lambda
## (Intercept)       -39.162213 69.8296176   NA 0.3533097 0.129445956
## GC.TAX.INTT.RV.ZS  -1.720484  3.4678125    2 0.2396392 0.007696179
## GC.XPN.INTP.RV.ZS  -1.941882  0.9097333    2 0.3149193 0.088573621
## IC.BUS.DISC.XQ    -17.413012 10.4958592    1 0.3717770 0.149089017
```

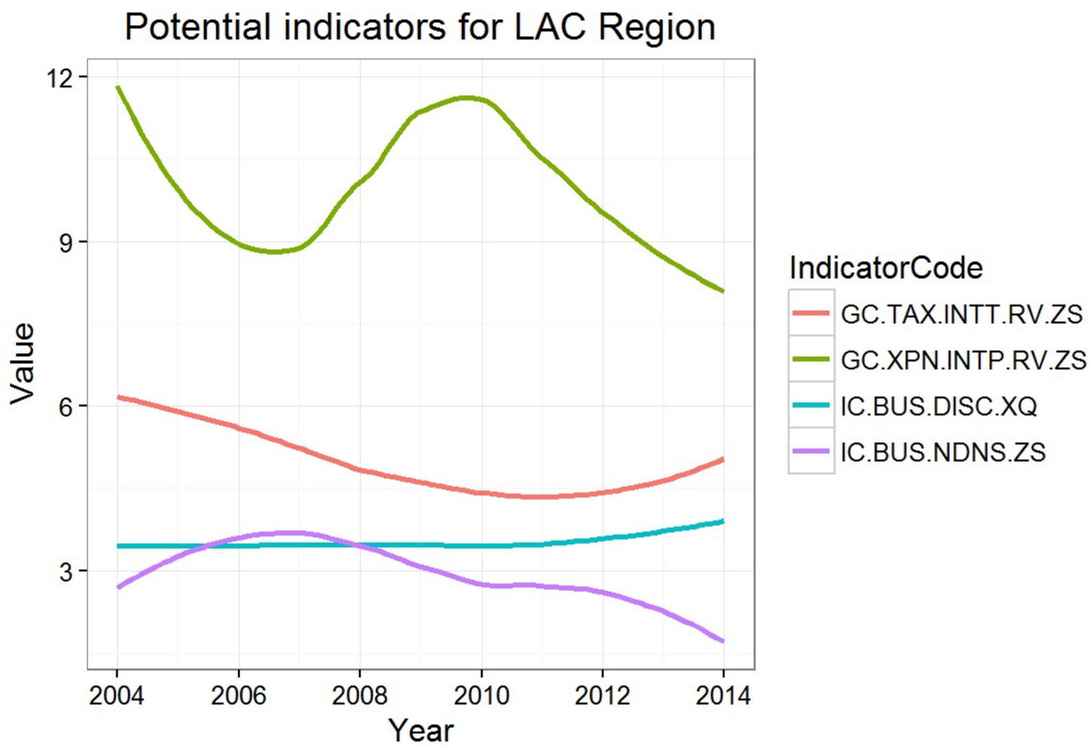
```r
#removing the worst predictor
LAC_glm.out <- with(LAC_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.TAX.INTT.RV.ZS + G
C.XPN.INTP.RV.ZS))

summary(pool(LAC_glm.out))
```

```
##                           est        se          t       df  Pr(>|t|)
## (Intercept)         1.7012724 7.5184511  0.2262796 5.221300 0.8295771
## GC.TAX.INTT.RV.ZS   0.9343004 0.9931785  0.9407175 5.592284 0.3856850
## GC.XPN.INTP.RV.ZS  -0.4092627 0.5266733 -0.7770712 5.017856 0.4721421
##                        lo 95      hi 95 nmis       fmi    lambda
## (Intercept)       -17.381720 20.7842647   NA 0.3724777 0.1707442
## GC.TAX.INTT.RV.ZS  -1.539522  3.4081224    2 0.3295721 0.1261743
## GC.XPN.INTP.RV.ZS  -1.761672  0.9431464    2 0.3957928 0.1949880
```

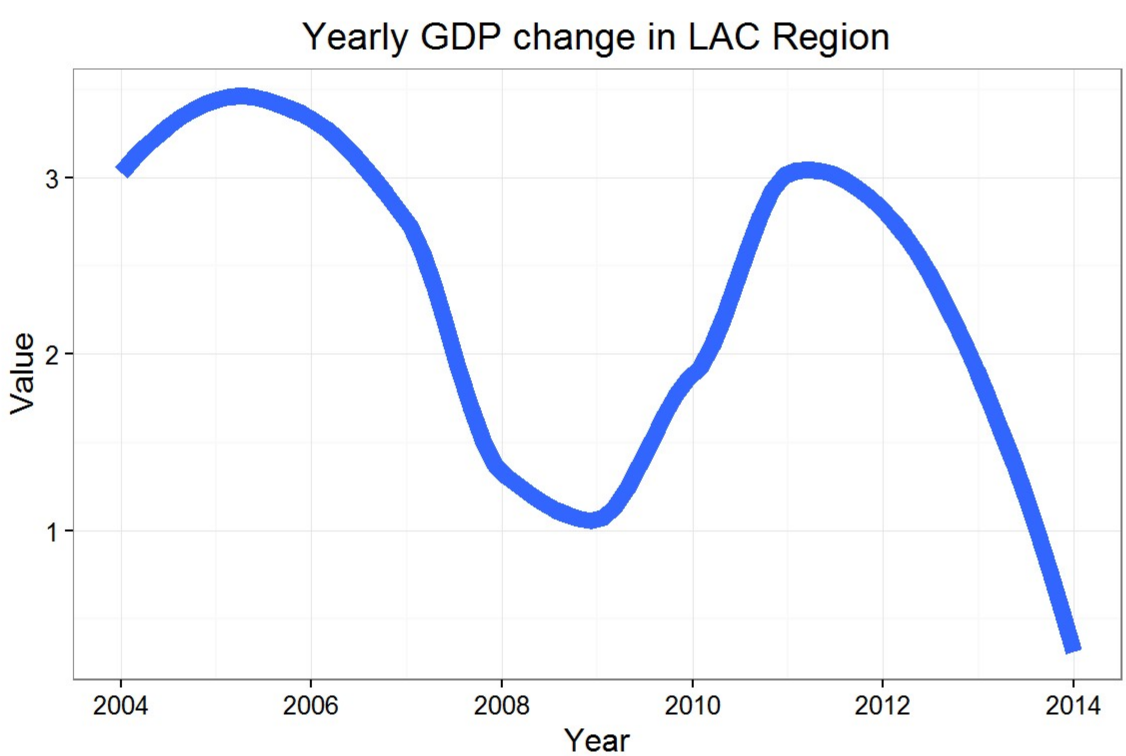
```r
#removing the worst predictor
LAC_glm.out <- with(LAC_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.XPN.INTP.RV.ZS))

summary(pool(LAC_glm.out))
```

```
##                           est        se          t       df  Pr(>|t|)
## (Intercept)         6.4965291 5.3563404  1.2128671 5.765386 0.2725238
## GC.XPN.INTP.RV.ZS  -0.4177505 0.5217503 -0.8006714 5.891252 0.4543904
##                       lo 95      hi 95 nmis       fmi    lambda
## (Intercept)       -6.740341 19.7333992   NA 0.3743382 0.1893786
## GC.XPN.INTP.RV.ZS -1.700162  0.8646608    2 0.3618596 0.1766565
```

```r
#Creating a dataset that includes the imputed values
LAC_Imp <- complete(LAC_Imp_temp, 1)

#Rearranging the table
LAC_GDP_regfix <- LAC_Imp %>% gather("IndicatorCode", "Value", 4:8)

#Plotting the variables and the overall GDP curve
ggplot(LAC_GDP_regfix %>% filter(IndicatorCode != "NY.GDP.PCAP.KD.ZG"), aes(x
= Year, y = Value, group = IndicatorCode, color = IndicatorCode)) + geom_smooth
(se = FALSE, size = 1) + ggtitle("Potential indicators for LAC Region")
```




```
LAC_GDP_Plot <- My_GDP_I %>% filter(CountryCode == "LAC") %>% ggplot(aes(x = Ye
ar, y = Value, group = CountryCode)) + geom_smooth(se = FALSE, size = 3) + ggti
tle("Yearly GDP change in LAC Region")

LAC_GDP_Plot
```



```
##########Middle East and North Africa Region##########
#Filtering out unecessary info, sepparating data set into different Indicator c
odes, and rearranging resulting table
MNA_GDP_reg_0 <- Indicators %>% filter(CountryCode == "MNA") %>% filter(Indicat
orCode == "NY.GDP.PCAP.KD.ZG") %>% filter(2003 < Year, Year < 2015) %>%  group_
by(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
MNA_GDP_reg_1 <- Indicators %>% filter(CountryCode == "MNA") %>% filter(Indicat
orCode == "GC.TAX.INTT.RV.ZS") %>% filter(2003 < Year, Year < 2015) %>% group_b
y(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
MNA_GDP_reg_2 <- Indicators %>% filter(CountryCode == "MNA") %>% filter(Indicat
orCode == "GC.XPN.INTP.RV.ZS") %>% filter(2003 < Year, Year < 2015) %>% group_b
y(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
MNA_GDP_reg_3 <- Indicators %>% filter(CountryCode == "MNA") %>% filter(Indicat
orCode == "IC.BUS.DISC.XQ") %>% filter(2003 < Year, Year < 2015) %>% group_by(Y
ear) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
MNA_GDP_reg_4 <- Indicators %>% filter(CountryCode == "MNA") %>% filter(Indicat
orCode == "IC.BUS.NDNS.ZS") %>% filter(2003 < Year, Year < 2015) %>% group_by(Y
ear) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)

#Combining all the resulting data sets into a larger one for the linear regress
ion
MNA_GDP_reg <- full_join(MNA_GDP_reg_0, MNA_GDP_reg_1)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```
MNA_GDP_reg <- full_join(MNA_GDP_reg, MNA_GDP_reg_2)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```
MNA_GDP_reg <- full_join(MNA_GDP_reg, MNA_GDP_reg_3)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```
MNA_GDP_reg <- full_join(MNA_GDP_reg, MNA_GDP_reg_4)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```
#imputing the missing data
MNA_Imp_temp <- MNA_GDP_reg %>% mice(m=5, method = "pmm")
```

```
##
##  iter imp variable
##   1   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   1   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   1   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   1   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   1   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   2   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   2   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   2   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   2   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   2   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   3   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   3   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   3   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   3   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   3   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   4   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   4   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   4   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   4   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   4   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   5   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   5   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   5   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   5   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
##   5   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ
```

```
#performing the linear regression
MNA_glm.out <- with(MNA_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.TAX.INTT.RV.ZS + G
C.XPN.INTP.RV.ZS + IC.BUS.DISC.XQ + IC.BUS.NDNS.ZS))

summary(pool(MNA_glm.out))
```

```
##                             est         se           t       df  Pr(>|t|)
## (Intercept)        9.24313312 17.9262462  0.51562011 3.913397 0.6338776
## GC.TAX.INTT.RV.ZS  1.53676585  1.2310340  1.24835374 3.522891 0.2884308
## GC.XPN.INTP.RV.ZS  0.01791864  0.7594955  0.02359282 3.975661 0.9823140
## IC.BUS.DISC.XQ    -3.04959199  3.0067571 -1.01424620 4.275734 0.3643731
## IC.BUS.NDNS.ZS    -2.76468220  3.0677236 -0.90121620 2.821968 0.4376800
##                       lo 95     hi 95 nmis       fmi     lambda
## (Intercept)       -40.96551 59.451777   NA 0.3916637 0.14404026
## GC.TAX.INTT.RV.ZS  -2.07166  5.145192    3 0.4546620 0.21351626
## GC.XPN.INTP.RV.ZS  -2.09588  2.131717    3 0.3814783 0.13285943
## IC.BUS.DISC.XQ    -11.18960  5.090415    1 0.3313032 0.07780408
## IC.BUS.NDNS.ZS    -12.88541  7.356042    0 0.5675069 0.34118734
```

```
#removing worst predictor
MNA_glm.out <- with(MNA_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.TAX.INTT.RV.ZS + I
C.BUS.DISC.XQ + IC.BUS.NDNS.ZS))

summary(pool(MNA_glm.out))
```

```
##                          est         se          t       df  Pr(>|t|)
## (Intercept)         9.807033 16.3477053  0.5999027 4.569269 0.5770682
## GC.TAX.INTT.RV.ZS  1.510208  0.8275021  1.8250203 4.016367 0.1417545
## IC.BUS.DISC.XQ    -3.105125  2.6773327 -1.1597830 5.187962 0.2967136
## IC.BUS.NDNS.ZS    -2.769969  2.7226948 -1.0173630 2.713542 0.3909551
##                         lo 95     hi 95 nmis       fmi     lambda
## (Intercept)       -33.4359975 53.050063   NA 0.3816591 0.15960450
## GC.TAX.INTT.RV.ZS  -0.7836183  3.804034    3 0.4560004 0.23911060
## IC.BUS.DISC.XQ     -9.9131014  3.702851    1 0.2956439 0.06799028
## IC.BUS.NDNS.ZS    -11.9759750  6.436037    0 0.6370476 0.44157251
```

```
#removing worst predictor
MNA_glm.out <- with(MNA_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.TAX.INTT.RV.ZS + I
C.BUS.NDNS.ZS))

summary(pool(MNA_glm.out))
```

```
##                          est        se         t       df   Pr(>|t|)
## (Intercept)       -7.075903 5.9950447 -1.180292 3.202437 0.31807237
## GC.TAX.INTT.RV.ZS  1.804485 0.8025395  2.248469 4.003661 0.08773264
## IC.BUS.NDNS.ZS    -3.783126 3.1439154 -1.203317 1.871751 0.35904945
##                         lo 95     hi 95 nmis       fmi    lambda
## (Intercept)       -25.4903474 11.338541   NA 0.6118181 0.4270767
## GC.TAX.INTT.RV.ZS  -0.4229183  4.031889    3 0.5134520 0.3189752
## IC.BUS.NDNS.ZS    -18.2385031 10.672252    0 0.7906970 0.6449302
```

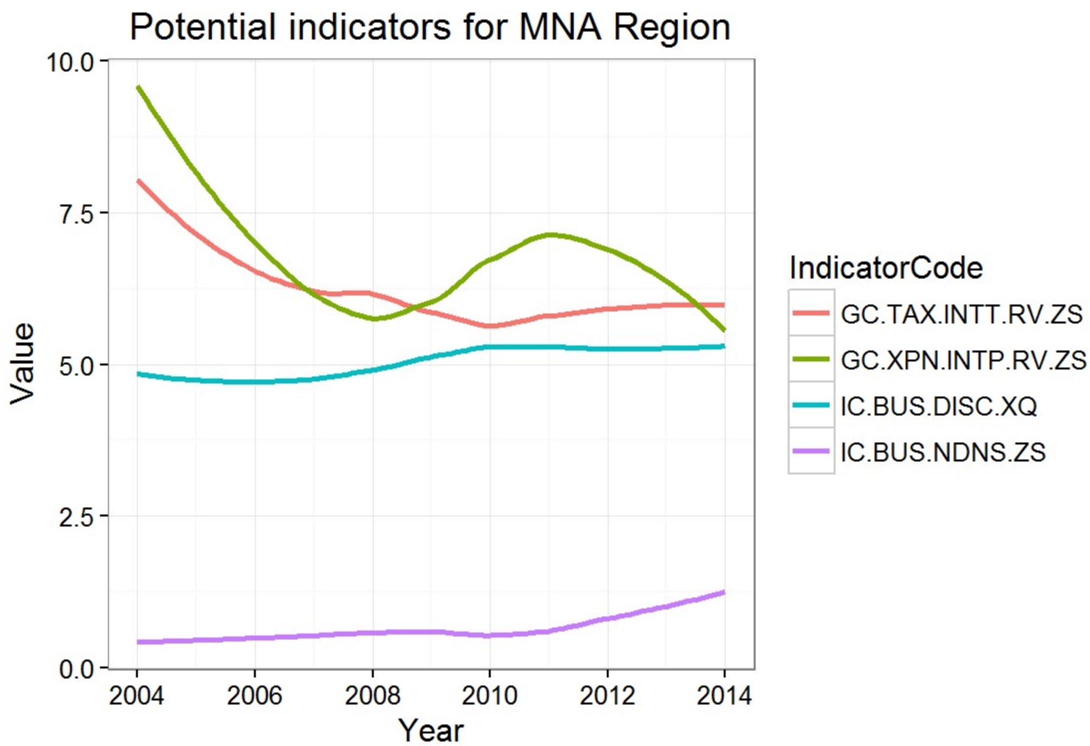
```
#removing worst predictor
MNA_glm.out <- with(MNA_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.TAX.INTT.RV.ZS + I
C.BUS.NDNS.ZS))

summary(pool(MNA_glm.out))
```

```
##                          est        se         t       df   Pr(>|t|)
## (Intercept)       -7.075903 5.9950447 -1.180292 3.202437 0.31807237
## GC.TAX.INTT.RV.ZS  1.804485 0.8025395  2.248469 4.003661 0.08773264
## IC.BUS.NDNS.ZS    -3.783126 3.1439154 -1.203317 1.871751 0.35904945
##                          lo 95     hi 95 nmis       fmi    lambda
## (Intercept)       -25.4903474 11.338541   NA 0.6118181 0.4270767
## GC.TAX.INTT.RV.ZS  -0.4229183  4.031889    3 0.5134520 0.3189752
## IC.BUS.NDNS.ZS    -18.2385031 10.672252    0 0.7906970 0.6449302
```

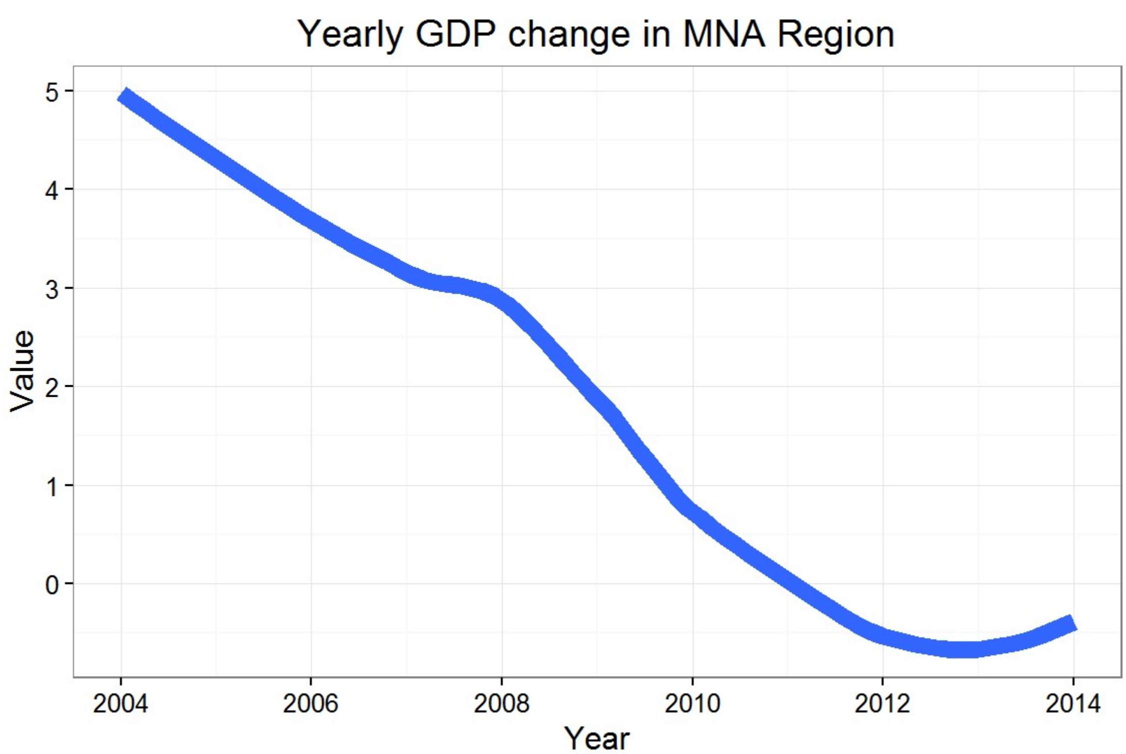
```
#removing worst predictor
MNA_glm.out <- with(MNA_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.TAX.INTT.RV.ZS))

summary(pool(MNA_glm.out))
```

```
##                          est       se         t       df  Pr(>|t|)
## (Intercept)       -9.225147 7.829259 -1.178291 2.017432 0.3589953
## GC.TAX.INTT.RV.ZS  1.759964 1.258348  1.398631 1.877748 0.3042238
##                         lo 95     hi 95 nmis       fmi    lambda
## (Intercept)       -42.634332 24.184038   NA 0.7933280 0.6563426
## GC.TAX.INTT.RV.ZS  -4.006497  7.526425    3 0.8113108 0.6801741
```

```
#Creating a dataset that includes the imputed values
MNA_Imp <- complete(MNA_Imp_temp, 1)

#Rearranging the table
MNA_GDP_regfix <- MNA_Imp %>% gather("IndicatorCode", "Value", 4:8)

#Plotting the variables and the overall GDP curve
ggplot(MNA_GDP_regfix %>% filter(IndicatorCode != "NY.GDP.PCAP.KD.ZG"), aes(x
= Year, y = Value, group = IndicatorCode, color = IndicatorCode)) + geom_smooth
(se = FALSE, size = 1) + ggtitle("Potential indicators for MNA Region")
```




```
MNA_GDP_Plot <- My_GDP_I %>% filter(CountryCode == "MNA") %>% ggplot(aes(x = Ye
ar, y = Value, group = CountryCode)) + geom_smooth(se = FALSE, size = 3) + ggti
tle("Yearly GDP change in MNA Region")

MNA_GDP_Plot
```



```
##########Sub-Saharan Africa Region##########
#Filtering out unecessary info, sepparating data set into different Indicator c
odes, and rearranging resulting table
SSA_GDP_reg_0 <- Indicators %>% filter(CountryCode == "SSA") %>% filter(Indicat
orCode == "NY.GDP.PCAP.KD.ZG") %>% filter(2003 < Year, Year < 2015) %>% group_b
y(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
SSA_GDP_reg_1 <- Indicators %>% filter(CountryCode == "SSA") %>% filter(Indicat
orCode == "GC.TAX.INTT.RV.ZS") %>% filter(2003 < Year, Year < 2015) %>% group_b
y(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
SSA_GDP_reg_2 <- Indicators %>% filter(CountryCode == "SSA") %>% filter(Indicat
orCode == "GC.XPN.INTP.RV.ZS") %>% filter(2003 < Year, Year < 2015) %>% group_b
y(Year) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
SSA_GDP_reg_3 <- Indicators %>% filter(CountryCode == "SSA") %>% filter(Indicat
orCode == "IC.BUS.DISC.XQ") %>% filter(2003 < Year, Year < 2015) %>% group_by(Y
ear) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)
SSA_GDP_reg_4 <- Indicators %>% filter(CountryCode == "SSA") %>% filter(Indicat
orCode == "IC.BUS.NDNS.ZS") %>% filter(2003 < Year, Year < 2015) %>% group_by(Y
ear) %>% spread(IndicatorCode, Value) %>% select(-IndicatorName)

#Combining all the resulting data sets into a larger one for the linear regress
ion
SSA_GDP_reg <- full_join(SSA_GDP_reg_0, SSA_GDP_reg_1)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```
SSA_GDP_reg <- full_join(SSA_GDP_reg, SSA_GDP_reg_2)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```
SSA_GDP_reg <- full_join(SSA_GDP_reg, SSA_GDP_reg_3)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```
SSA_GDP_reg <- full_join(SSA_GDP_reg, SSA_GDP_reg_4)
```

```
## Joining by: c("CountryName", "CountryCode", "Year")
```

```
#imputing the missing data
SSA_Imp_temp <- SSA_GDP_reg %>% mice(m=5, method = "pmm")
```

```
##
##  iter imp variable
##   1   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   1   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   1   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   1   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   1   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   2   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   2   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   2   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   2   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   2   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   3   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   3   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   3   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   3   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   3   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   4   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   4   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   4   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   4   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   4   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   5   1  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   5   2  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   5   3  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
```

```
##   5   4  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
##   5   5  GC.TAX.INTT.RV.ZS  GC.XPN.INTP.RV.ZS  IC.BUS.DISC.XQ  IC.BUS.NDNS.Z
S
```

```
#perfoming the linear regression with filled dataset

SSA_glm.out <- with(SSA_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.TAX.INTT.RV.ZS + G
C.XPN.INTP.RV.ZS + IC.BUS.DISC.XQ + IC.BUS.NDNS.ZS))

summary(pool(SSA_glm.out))
```

```
##                           est         se          t       df  Pr(>|t|)
## (Intercept)       -19.6805981 22.0623313 -0.8920453 4.267750 0.4198164
## GC.TAX.INTT.RV.ZS   0.4685428  0.2626304  1.7840385 4.507823 0.1408576
## GC.XPN.INTP.RV.ZS  -0.1925910  0.4541221 -0.4240951 4.108561 0.6927625
## IC.BUS.DISC.XQ      5.3517175  4.5571420  1.1743583 4.174400 0.3028644
## IC.BUS.NDNS.ZS     -5.7630834  3.5037353 -1.6448398 4.077568 0.1739919
##                         lo 95     hi 95 nmis       fmi     lambda
## (Intercept)       -79.4492062 40.088010   NA 0.3326688 0.07930408
## GC.TAX.INTT.RV.ZS  -0.2293549  1.166440    5 0.2904961 0.03286107
## GC.XPN.INTP.RV.ZS  -1.4404053  1.055223    4 0.3595147 0.10876485
## IC.BUS.DISC.XQ     -7.0951661 17.798601    1 0.3484940 0.09667507
## IC.BUS.NDNS.ZS    -15.4184628  3.892296    1 0.3646678 0.11441718
```

```
#removing worst predictor

SSA_glm.out <- with(SSA_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.TAX.INTT.RV.ZS + I
C.BUS.DISC.XQ + IC.BUS.NDNS.ZS))

summary(pool(SSA_glm.out))
```

```
##                           est         se         t       df  Pr(>|t|)
## (Intercept)       -21.3093978 19.0306806 -1.119739 5.194361 0.3119119
## GC.TAX.INTT.RV.ZS   0.4442532  0.2376116  1.869661 5.343952 0.1167168
## IC.BUS.DISC.XQ      5.2905855  4.1200513  1.284107 4.921918 0.2562252
## IC.BUS.NDNS.ZS     -5.1197835  3.0686536 -1.668414 4.015473 0.1702835
##                         lo 95     hi 95 nmis       fmi     lambda
## (Intercept)       -69.6838494 27.065054   NA 0.2947155 0.06699719
## GC.TAX.INTT.RV.ZS  -0.1549097  1.043416    5 0.2726339 0.04332386
## IC.BUS.DISC.XQ     -5.3510794 15.932250    1 0.3333715 0.10823215
## IC.BUS.NDNS.ZS    -13.6267999  3.387233    1 0.4561207 0.23924015
```

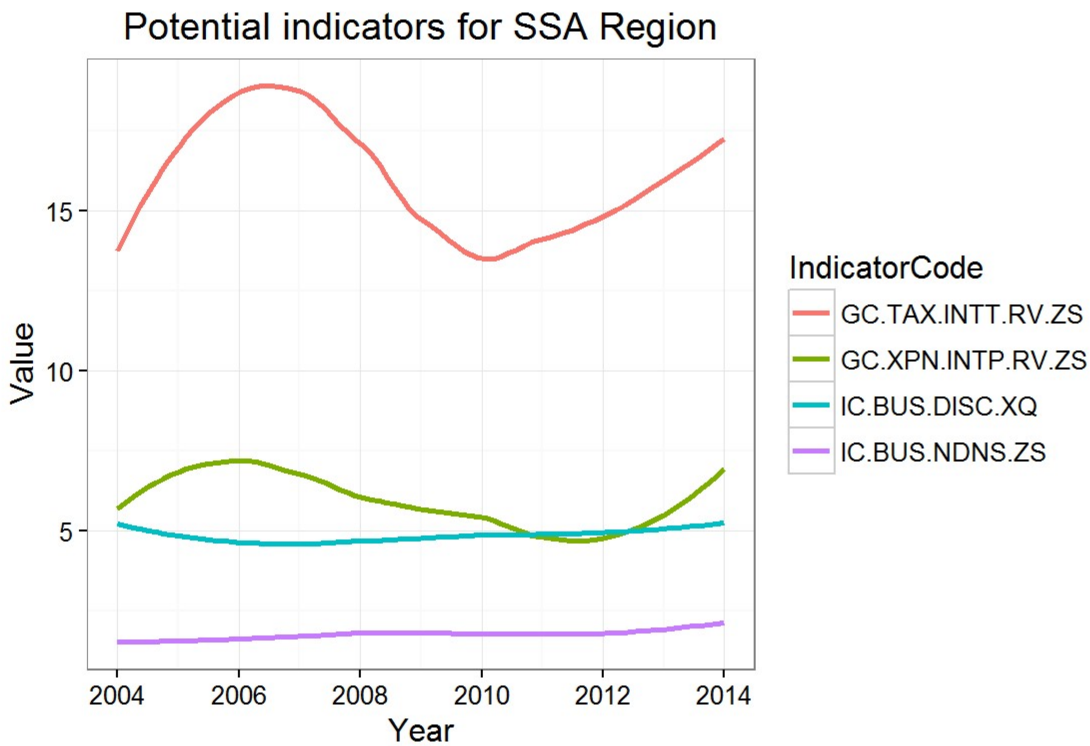
```r
#removing worst predictor

SSA_glm.out <- with(SSA_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ GC.TAX.INTT.RV.ZS +  I
C.BUS.NDNS.ZS))

summary(pool(SSA_glm.out))
```

```
##                           est        se          t       df  Pr(>|t|)
## (Intercept)         3.3027350 5.7664377  0.5727514 4.884197 0.5921680
## GC.TAX.INTT.RV.ZS  0.2313312 0.2238576  1.0333854 4.602907 0.3526311
## IC.BUS.NDNS.ZS    -2.6389968 2.2301728 -1.1833149 6.058927 0.2810250
##                        lo 95      hi 95 nmis       fmi    lambda
## (Intercept)       -11.6267004 18.2321705   NA 0.4110983 0.2109345
## GC.TAX.INTT.RV.ZS  -0.3593703  0.8220328    5 0.4434064 0.2447262
## IC.BUS.NDNS.ZS     -8.0831938  2.8052001    1 0.2736341 0.0678335
```

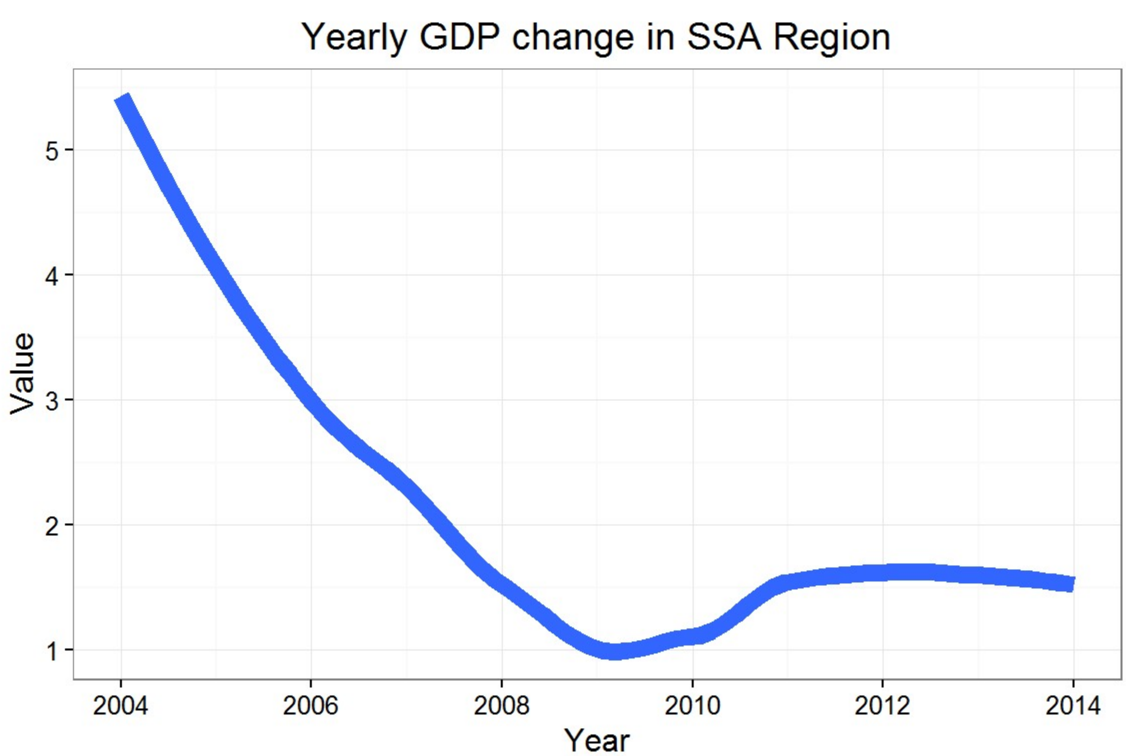
```r
#removing worst predictor

SSA_glm.out <- with(SSA_Imp_temp, lm(NY.GDP.PCAP.KD.ZG ~ IC.BUS.NDNS.ZS))

summary(pool(SSA_glm.out))
```

```
##                       est       se         t       df  Pr(>|t|)     lo 95
## (Intercept)     7.911476 3.858701  2.050295 7.355692 0.0775515 -1.124217
## IC.BUS.NDNS.ZS -3.204311 2.152841 -1.488410 7.346388 0.1782604 -8.246716
##                    hi 95 nmis       fmi     lambda
## (Intercept)    16.947169   NA 0.2081509 0.01861565
## IC.BUS.NDNS.ZS  1.838093    1 0.2092585 0.01977742
```

```r
#Creating a dataset that includes the imputed values
SSA_Imp <- complete(SSA_Imp_temp, 1)

#Rearranging the table
SSA_GDP_regfix <- SSA_Imp %>% gather("IndicatorCode", "Value", 4:8)

#Plotting the variables and the overall GDP curve
ggplot(SSA_GDP_regfix %>% filter(IndicatorCode != "NY.GDP.PCAP.KD.ZG"), aes(x
= Year, y = Value, group = IndicatorCode, color = IndicatorCode)) + geom_smooth
(se = FALSE, size = 1) + ggtitle("Potential indicators for SSA Region")
```




```
SSA_GDP_Plot <- My_GDP_I %>% filter(CountryCode == "SSA") %>% ggplot(aes(x = Ye
ar, y = Value, group = CountryCode)) + geom_smooth(se = FALSE, size = 3) + ggti
tle("Yearly GDP change in SSA Region")

SSA_GDP_Plot
```



In order to determine if any of the suspect factors were actually predictors of the countries change in GDP for that year, a linear regression was performed. The overall “Indicators” dataset was broken up by CountryCode, and irrelavent information wa removed. The remaining tables were then rejoined, and any resulting N/A values were filled with the MICE package. The regression was performed on the resulting dataset. When it was detemined that none of the chosen variables was predictive, the predictor values as well as the GDP change were plotted to visually show a lack of similarity between the predictor and the nonpredictive curves.’

A linear regression is a good way to determine if variable are predictive or not, and further to distinguish which of the variables are predictive. In this case, none of the variables were predictive. Therefore, the plots were made to show how the predictive variables differed from the variable that they were trying to predict.

Comparing the plots of the unpredictive variables and the variable of interest, it can be easily seen that there is no real realtionship between them. This validates the results of the linear regressions.

Developing countries do not need to be focusing on taxes on international trade, interest payments, business extent of disclosure index, or new business density to help increase GDP growth.