Appendix

Vinh Nghiem To

11/16/2020

```
# Subset PWT dataset to be between 2011-2017
pwt = read_xlsx("PWT.xlsx")
pwt = subset(pwt, year >= 2011)
pwt = subset(pwt, year <= 2017)

# Format the FDI dataset to match with PWT
FDI = read.csv("FDI_stacked.csv")
names(FDI)[1] = "countrycode"

# Merging the dataset
final_df = merge(FDI, pwt)
```

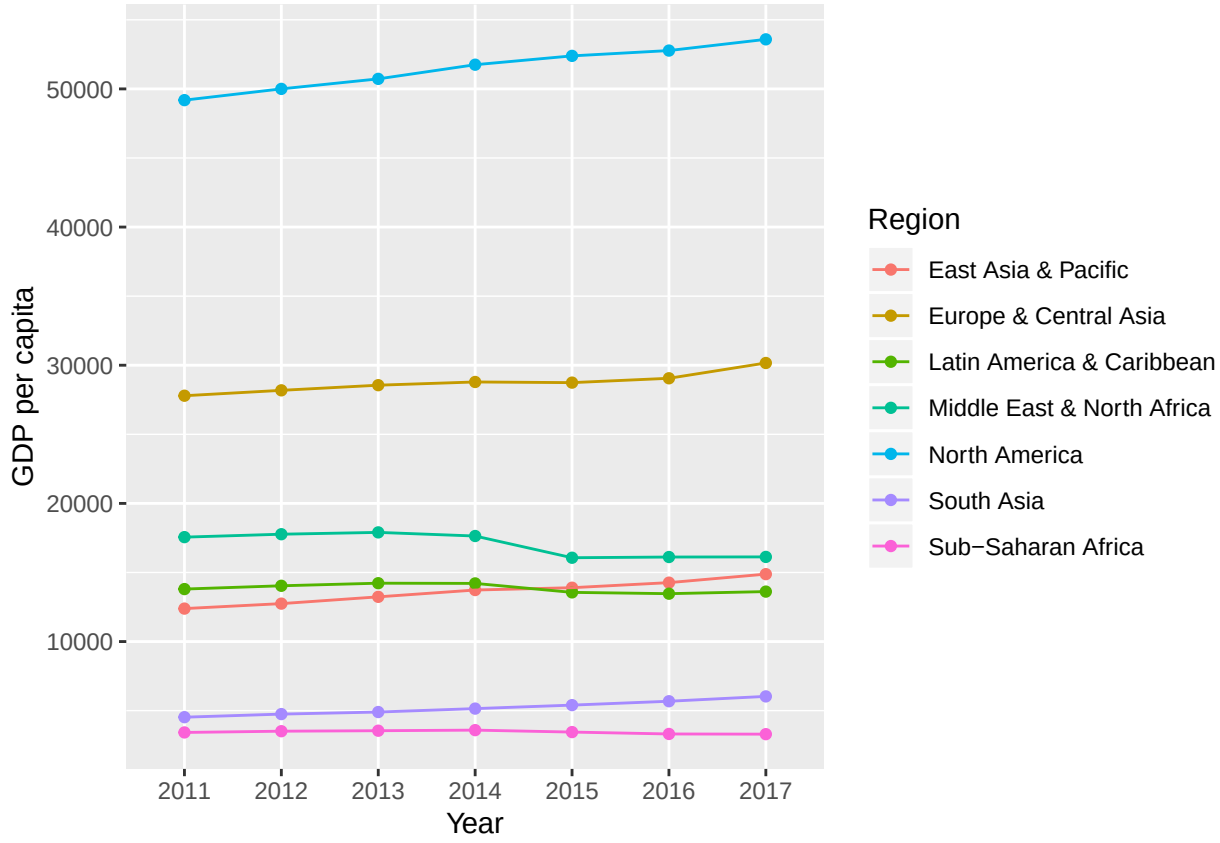
This meta file consists of information about the Region and Income Group of each country for further grouping analysis.

```
meta = read.csv("Metadata_Country.csv")
names(meta)[1] = "countrycode"
#meta = select(meta, c("countrycode", "Region", "IncomeGroup"))

final_df = merge(meta, final_df)
```

EDA

Scatter plot of GDP per capita grouped by Region. GDP per capita is a more accurate and concise measurement than only cumulative GDP.

```
GDP = final_df %>% group_by(year, Region) %>% summarise(GDP_pc=sum(rgdpo)/sum(pop), HC=sum(hc, na.rm = T

ggplot(data = GDP, mapping = aes(x=as.factor(year), y=GDP_pc, color=Region, group=Region)) + geom_point
```




```
FDI_year = final_df %>% group_by(year) %>% summarise(FDI=sum(FDI, na.rm = T), GDP=sum(rgdpo))

# Compute GDP growth rate
FDI_year = FDI_year %>% mutate(GDP_growth=(GDP-lag(GDP))/lag(GDP)*100)

# Compute FDI growth rate
FDI_year = FDI_year %>% mutate(FDI_growth=(FDI-lag(FDI))/lag(FDI)*100)
```

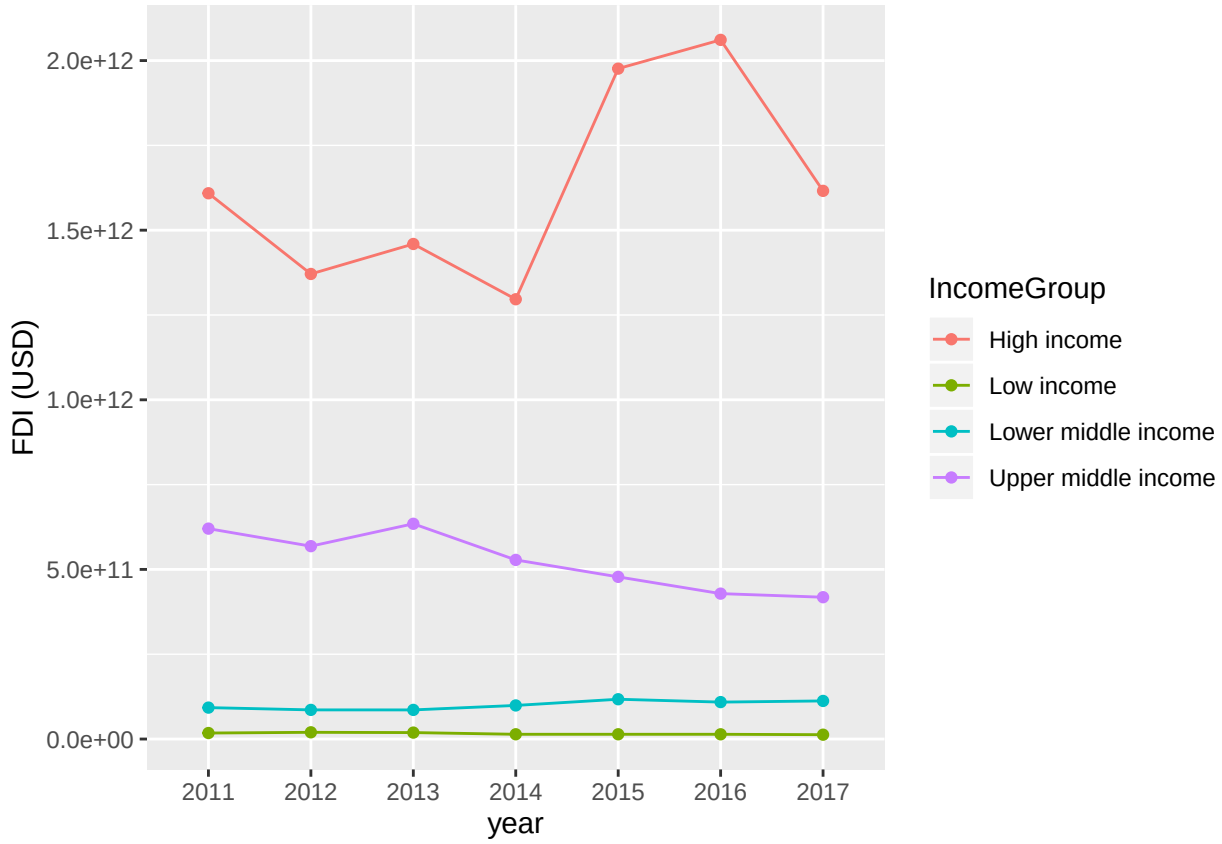
Graph 2 y-axis plot with the left y-axis and the red line represent total FDI annually while the right y-axis and the blue line represent GDP growth rate. The code runs in R script but not R markdown, as it keeps warning "calling par(new=TRUE) with no plot". Thus, I am commenting out my code below.

```
# par(mar = c(5, 5, 3, 5))
# plot(x=FDI_year$year, y=FDI_year$FDI, type = "l", xlab = "year", ylab = "FDI (USD)", col="red", main =
# par(new = TRUE)
# plot(x=FDI_year$year, y=FDI_year$GDP_growth, type = "l", xaxt = "n", yaxt = "n",ylab = "", xlab = "",
# axis(side = 4)
# mtext("GDP growth rate (%)", side=4, line = 3)
# legend("topleft", c("GDP growth rate", "FDI"),col = c("blue", "red"), lty = c(1, 1), cex = 0.8)
```

Below is **Figure A1**. We have looked at cumulative FDI and GDP growth by Regions, the followings are those grouped by Income Group.

```
GDP_income = final_df %>% group_by(year, IncomeGroup) %>% summarise(FDI=sum(FDI,na.rm = T), GDP=sum(rgdp
```

```
ggplot(data = GDP_income, mapping = aes(x=as.factor(year), y=FDI, color=IncomeGroup, group=IncomeGroup)
```



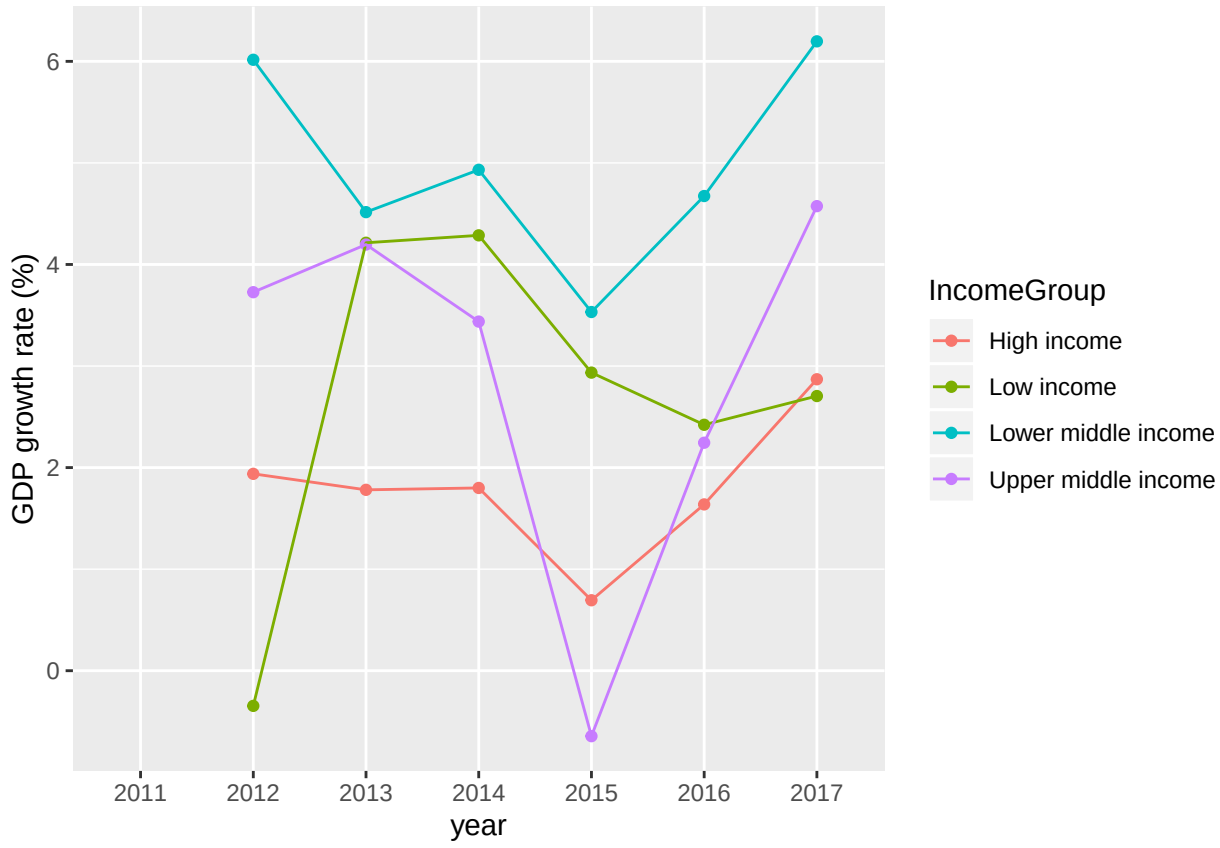
Below is **Figure A2**.

```
GDP_income = GDP_income %>% group_by(IncomeGroup) %>% mutate(GDP_growth=(GDP-lag(GDP))/lag(GDP)*100)
```

```
ggplot(data = GDP_income, mapping = aes(x=as.factor(year), y=GDP_growth, color=IncomeGroup, group=Income
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```

```
## Warning: Removed 4 rows containing missing values (geom_path).
```

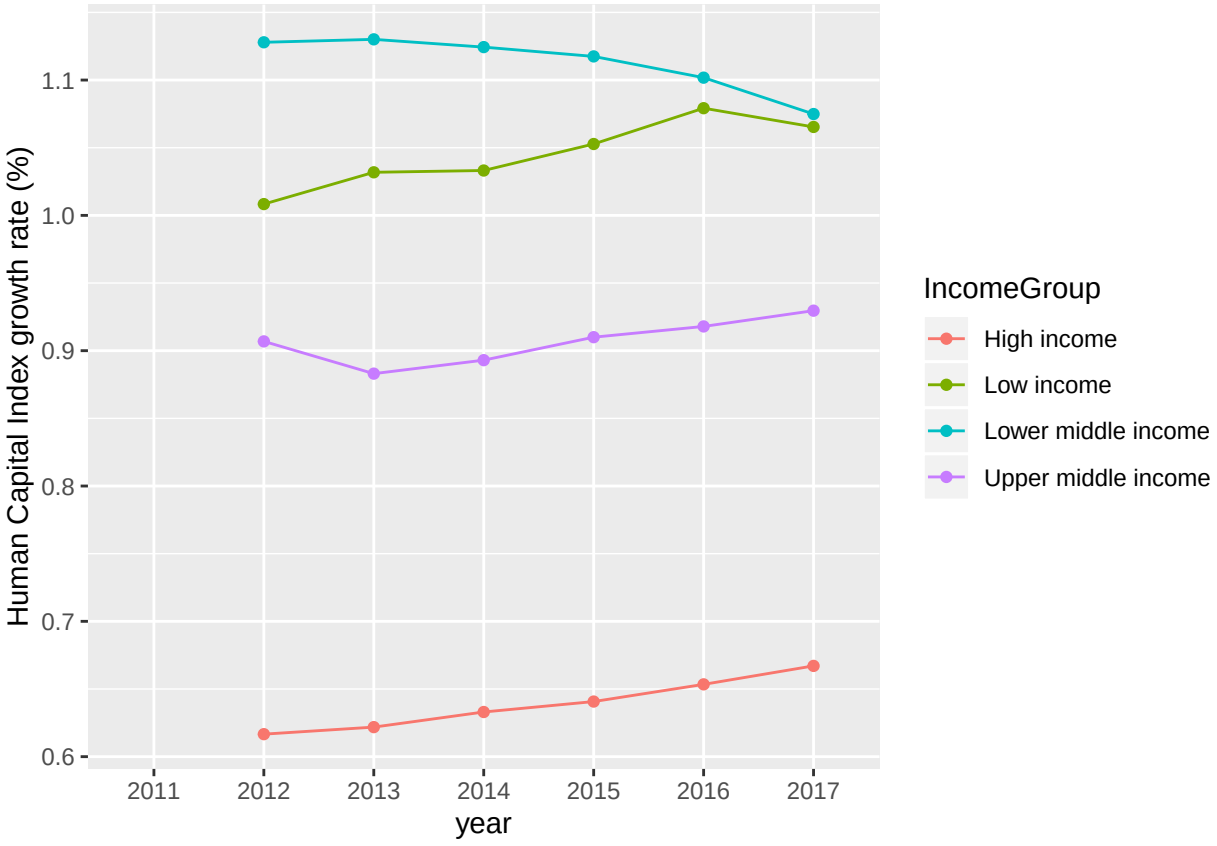


These are the plots of human capital rate of change by Region and Income Group.

```
HC = final_df %>% group_by(year, IncomeGroup) %>% summarise(HC=sum(hc, na.rm = T))
HC = HC %>% group_by(IncomeGroup) %>% mutate(HC_growth=(HC-lag(HC))/lag(HC)*100)
ggplot(data = HC, mapping = aes(x=as.factor(year), y=HC_growth, color=IncomeGroup, group=IncomeGroup)) +
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```

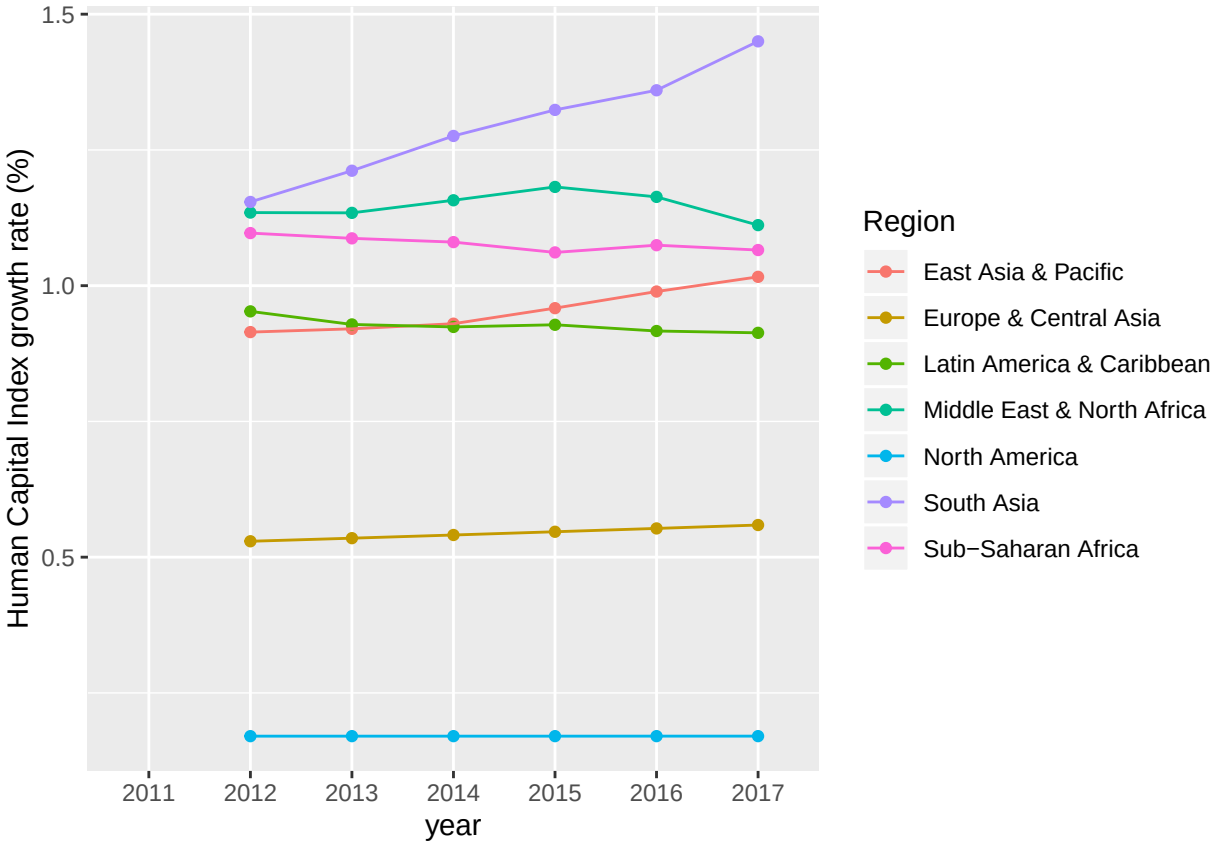
```
## Warning: Removed 4 rows containing missing values (geom_path).
```




```
HC_region = final_df %>% group_by(year, Region) %>% summarise(HC=sum(hc, na.rm = T))
HC_region = HC_region %>% group_by(Region) %>% mutate(HC_growth=(HC-lag(HC))/lag(HC)*100)
ggplot(data = HC_region, mapping = aes(x=as.factor(year), y=HC_growth, color=Region, group=Region)) + g
```

```
## Warning: Removed 7 rows containing missing values (geom_point).
```

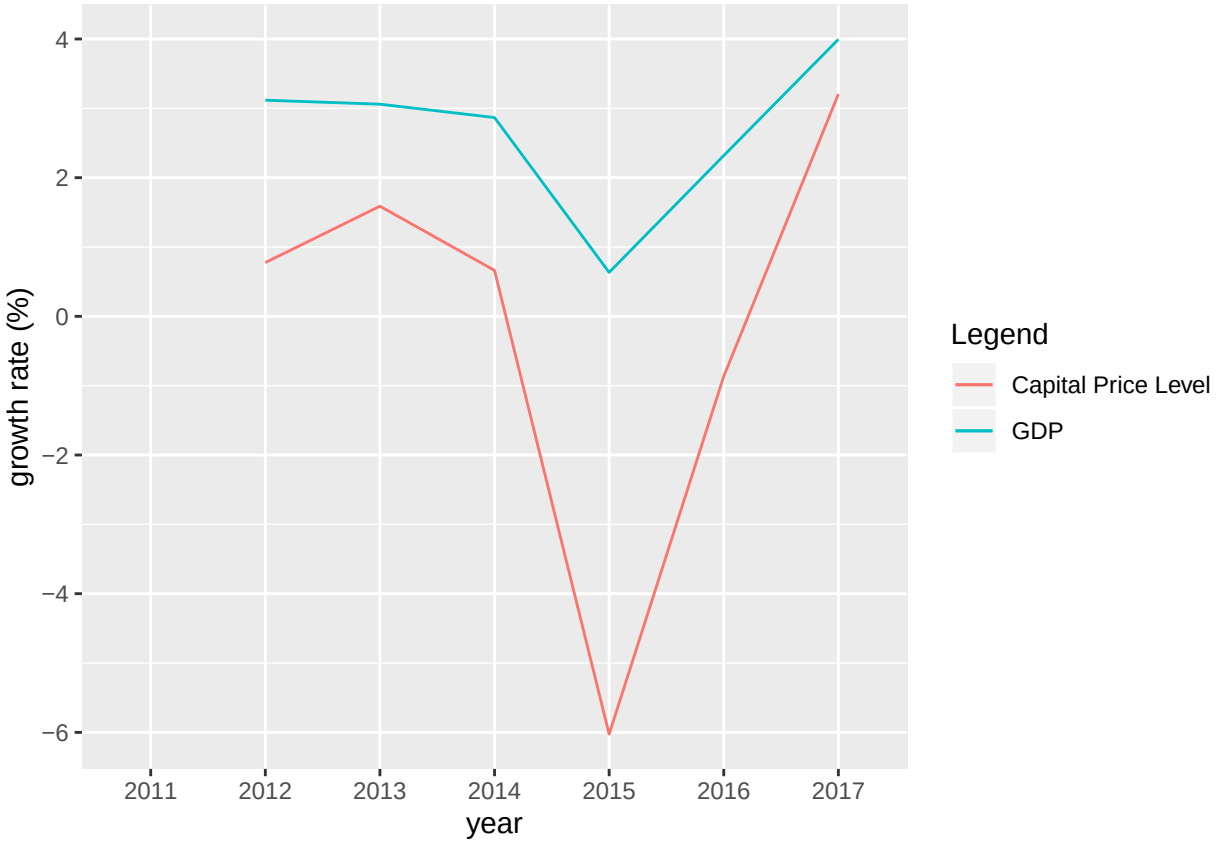
```
## Warning: Removed 7 rows containing missing values (geom_path).
```



Since labor and capital are two of the most important factors in producing goods, I would like to analyze how capital price change related with GDP growth rate.

```
Capital = final_df %>% group_by(year) %>% summarise(PK=sum(pl_i, na.rm = T), GDP=sum(rgdpo))

Capital = Capital %>% mutate(PK_growth=(PK-lag(PK))/lag(PK)*100)

Capital = Capital %>% mutate(GDP_growth=(GDP-lag(GDP))/lag(GDP)*100)

ggplot(data = Capital, aes(x=as.factor(year), group=1)) + geom_line(aes(y=PK_growth, color="Capital Pri
```

```
## Warning: Removed 1 rows containing missing values (geom_path).

## Warning: Removed 1 rows containing missing values (geom_path).
```



The followings are the regression equations of GDP per capita by FDI, and GDP growth rate by capital price change. The result is to test whether there are statistically significant relationship between the variables. Below is **Table A1**.

```
summary(lm(GDP_pc~FDI, data = GDP))
```

```
##
## Call:
## lm(formula = GDP_pc ~ FDI, data = GDP)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -12320  -8963  -5576   3593  32885
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.155e+04  2.856e+03   4.043 0.000194 ***
## FDI         2.315e-08  6.364e-09   3.638 0.000681 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13830 on 47 degrees of freedom
## Multiple R-squared:  0.2197, Adjusted R-squared:  0.2031
## F-statistic: 13.23 on 1 and 47 DF,  p-value: 0.0006814
```

Below is **Table A2**.

```r
summary(lm(GDP_growth~PK_growth, data = Capital))
```

```
##
## Call:
## lm(formula = GDP_growth ~ PK_growth, data = Capital)
##
## Residuals:
##       2       3       4       5       6       7
##  0.13988 -0.20402 -0.07079  0.05381 -0.08031  0.16143
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.70385    0.06511   41.53 2.01e-06 ***
## PK_growth    0.35245    0.02236   15.77 9.46e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1594 on 4 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.9842, Adjusted R-squared:  0.9802
## F-statistic: 248.6 on 1 and 4 DF,  p-value: 9.457e-05
```

Results and Discussion

Below is the summary of linear regression, stepwise approach and residual plots for each year.

```r
# This part shows the results of year 2015
GDP_2015 = subset(final_df, year==2015)
GDP_2015 = GDP_2015 %>% group_by(country) %>% mutate(GDP_pc=rgdpo/pop)
reg_2015 = lm(log(GDP_pc)~hc+FDI+Region+IncomeGroup, data = GDP_2015)
summary(reg_2015)
```

```
##
## Call:
## lm(formula = log(GDP_pc) ~ hc + FDI + Region + IncomeGroup, data = GDP_2015)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.86798 -0.25307  0.00491  0.22401  1.09911
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       9.703e+00  3.089e-01  31.407  < 2e-16 ***
## hc                                2.536e-01  9.506e-02   2.668  0.00861 **
## FDI                               1.216e-12  6.591e-13   1.844  0.06742 .
## RegionEurope & Central Asia      -1.177e-01  1.155e-01  -1.019  0.31004
## RegionLatin America & Caribbean  -3.545e-01  1.242e-01  -2.853  0.00503 **
## RegionMiddle East & North Africa  2.779e-01  1.375e-01   2.022  0.04528 *
## RegionNorth America              -2.217e-01  3.387e-01  -0.654  0.51403
## RegionSouth Asia                 -1.515e-01  1.922e-01  -0.788  0.43213
## RegionSub-Saharan Africa         -4.273e-01  1.346e-01  -3.175  0.00187 **
```

```
## IncomeGroupLow income            -2.423e+00  1.645e-01 -14.730  < 2e-16 ***
## IncomeGroupLower middle income   -1.563e+00  1.165e-01 -13.409  < 2e-16 ***
## IncomeGroupUpper middle income   -7.868e-01  9.832e-02  -8.003  5.9e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.386 on 130 degrees of freedom
##   (37 observations deleted due to missingness)
## Multiple R-squared:  0.9044, Adjusted R-squared:  0.8963
## F-statistic: 111.8 on 11 and 130 DF,  p-value: < 2.2e-16
```

```
# Because my regression model has factorvariables and hence is a constant. The stepwise regression in R
#stepAIC(lm(GDP_pc~hc+FDI+Region+IncomeGroup+FDI*Region+FDI*IncomeGroup+hc*Region+hc*IncomeGroup, data
```

```
exp(reg_2015$coefficients)
```

```
##                       (Intercept)                                hc
##                      1.636930e+04                      1.288622e+00
##                               FDI      RegionEurope & Central Asia
##                      1.000000e+00                      8.889556e-01
##  RegionLatin America & Caribbean RegionMiddle East & North Africa
##                      7.015119e-01                      1.320324e+00
##              RegionNorth America                  RegionSouth Asia
##                      8.011875e-01                      8.594570e-01
##         RegionSub-Saharan Africa             IncomeGroupLow income
##                      6.522538e-01                      8.864701e-02
##   IncomeGroupLower middle income   IncomeGroupUpper middle income
##                      2.095780e-01                      4.552901e-01
```

```
exp(confint(reg_2015))
```

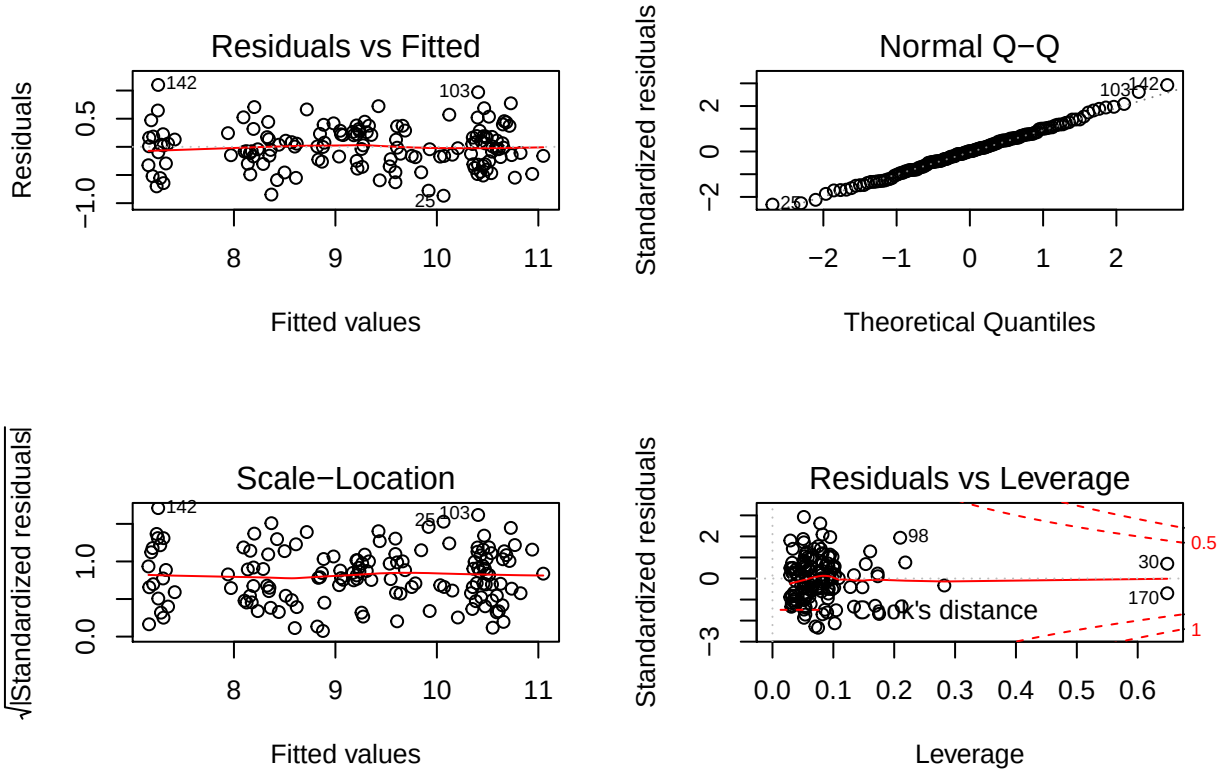
```
##                                       2.5 %       97.5 %
## (Intercept)                      8.883483e+03 3.016315e+04
## hc                               1.067705e+00 1.555249e+00
## FDI                              1.000000e+00 1.000000e+00
## RegionEurope & Central Asia      7.073626e-01 1.117167e+00
## RegionLatin America & Caribbean  5.486396e-01 8.969803e-01
## RegionMiddle East & North Africa 1.005951e+00 1.732942e+00
## RegionNorth America              4.099096e-01 1.565958e+00
## RegionSouth Asia                 5.876044e-01 1.257081e+00
## RegionSub-Saharan Africa         4.997817e-01 8.512416e-01
## IncomeGroupLow income            6.402174e-02 1.227441e-01
## IncomeGroupLower middle income   1.664236e-01 2.639226e-01
## IncomeGroupUpper middle income   3.748136e-01 5.530456e-01
```

```
# Nested f-test
complex = lm(GDP_pc~hc+FDI+Region+IncomeGroup+FDI*Region+FDI*IncomeGroup+hc*Region+hc*IncomeGroup, data
nested = lm(log(GDP_pc)~hc+FDI+Region+IncomeGroup, data = GDP_2015)
anova(complex, nested, test='Chisq')
```

```
## Warning in anova.lmlist(object, ...): models with response '"log(GDP_pc)"'
## removed because response differs from model 1
```

```
## Analysis of Variance Table
##
## Response: GDP_pc
##                   Df     Sum Sq    Mean Sq  F value    Pr(>F)
## hc                 1 2.0769e+10 2.0769e+10 173.9862 < 2.2e-16 ***
## FDI                1 1.6041e+09 1.6041e+09  13.4381 0.0003772 ***
## Region             6 3.8385e+09 6.3974e+08   5.3592 6.682e-05 ***
## IncomeGroup        3 1.1670e+10 3.8901e+09  32.5880 2.973e-15 ***
## FDI:Region         6 5.8727e+08 9.7879e+07   0.8199 0.5566973
## FDI:IncomeGroup    3 1.3909e+08 4.6364e+07   0.3884 0.7615703
## hc:Region          5 2.7789e+08 5.5578e+07   0.4656 0.8011618
## hc:IncomeGroup     3 2.0504e+08 6.8347e+07   0.5725 0.6342185
## Residuals        113 1.3489e+10 1.1937e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
par(mfrow=c(2,2))
plot(reg_2015)
```




```r
# This part shows the results of year 2016
GDP_2016 = subset(final_df, year==2016)
GDP_2016 = GDP_2016 %>% group_by(country) %>% mutate(GDP_pc=rgdpo/pop)
reg_2016 = lm(log(GDP_pc)~hc+FDI+Region+IncomeGroup, data = GDP_2016)
summary(reg_2016)
```

```
##
```

```
## Call:
## lm(formula = log(GDP_pc) ~ hc + FDI + Region + IncomeGroup, data = GDP_2016)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.90034 -0.28641  0.04095  0.21253  1.12294
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       9.653e+00  3.080e-01  31.344  < 2e-16 ***
## hc                                2.703e-01  9.445e-02   2.862 0.004914 **
## FDI                               8.624e-13  7.012e-13   1.230 0.220952
## RegionEurope & Central Asia      -1.046e-01  1.157e-01  -0.904 0.367511
## RegionLatin America & Caribbean  -3.670e-01  1.247e-01  -2.943 0.003846 **
## RegionMiddle East & North Africa  2.473e-01  1.378e-01   1.795 0.075002 .
## RegionNorth America              -1.049e-01  3.337e-01  -0.314 0.753691
## RegionSouth Asia                 -1.624e-01  1.932e-01  -0.841 0.402100
## RegionSub-Saharan Africa         -4.599e-01  1.349e-01  -3.409 0.000869 ***
## IncomeGroupLow income            -2.376e+00  1.647e-01 -14.425  < 2e-16 ***
## IncomeGroupLower middle income   -1.519e+00  1.167e-01 -13.012  < 2e-16 ***
## IncomeGroupUpper middle income   -7.679e-01  9.891e-02  -7.763 2.16e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3886 on 130 degrees of freedom
##   (37 observations deleted due to missingness)
## Multiple R-squared:  0.9032, Adjusted R-squared:  0.8951
## F-statistic: 110.3 on 11 and 130 DF,  p-value: < 2.2e-16
```

```
exp(reg_2016$coefficients)
```

```
##                      (Intercept)                                hc
##                     1.556721e+04                      1.310323e+00
##                              FDI       RegionEurope & Central Asia
##                     1.000000e+00                      9.006580e-01
##  RegionLatin America & Caribbean RegionMiddle East & North Africa
##                     6.927947e-01                      1.280515e+00
##               RegionNorth America                   RegionSouth Asia
##                     9.003849e-01                      8.501283e-01
##         RegionSub-Saharan Africa             IncomeGroupLow income
##                     6.313346e-01                      9.290708e-02
##   IncomeGroupLower middle income   IncomeGroupUpper middle income
##                     2.190252e-01                      4.639826e-01
```

```
exp(confint(reg_2016))
```

```
##                                       2.5 %       97.5 %
## (Intercept)                     8.464537e+03 2.862979e+04
## hc                              1.086997e+00 1.579531e+00
## FDI                             1.000000e+00 1.000000e+00
## RegionEurope & Central Asia     7.163878e-01 1.132326e+00
## RegionLatin America & Caribbean 5.413360e-01 8.866295e-01
## RegionMiddle East & North Africa 9.750327e-01 1.681707e+00
```

```
## RegionNorth America              4.652656e-01 1.742430e+00
## RegionSouth Asia                 5.801355e-01 1.245775e+00
## RegionSub-Saharan Africa         4.834369e-01 8.244786e-01
## IncomeGroupLow income            6.706857e-02 1.287000e-01
## IncomeGroupLower middle income   1.738691e-01 2.759091e-01
## IncomeGroupUpper middle income   3.815185e-01 5.642711e-01
```

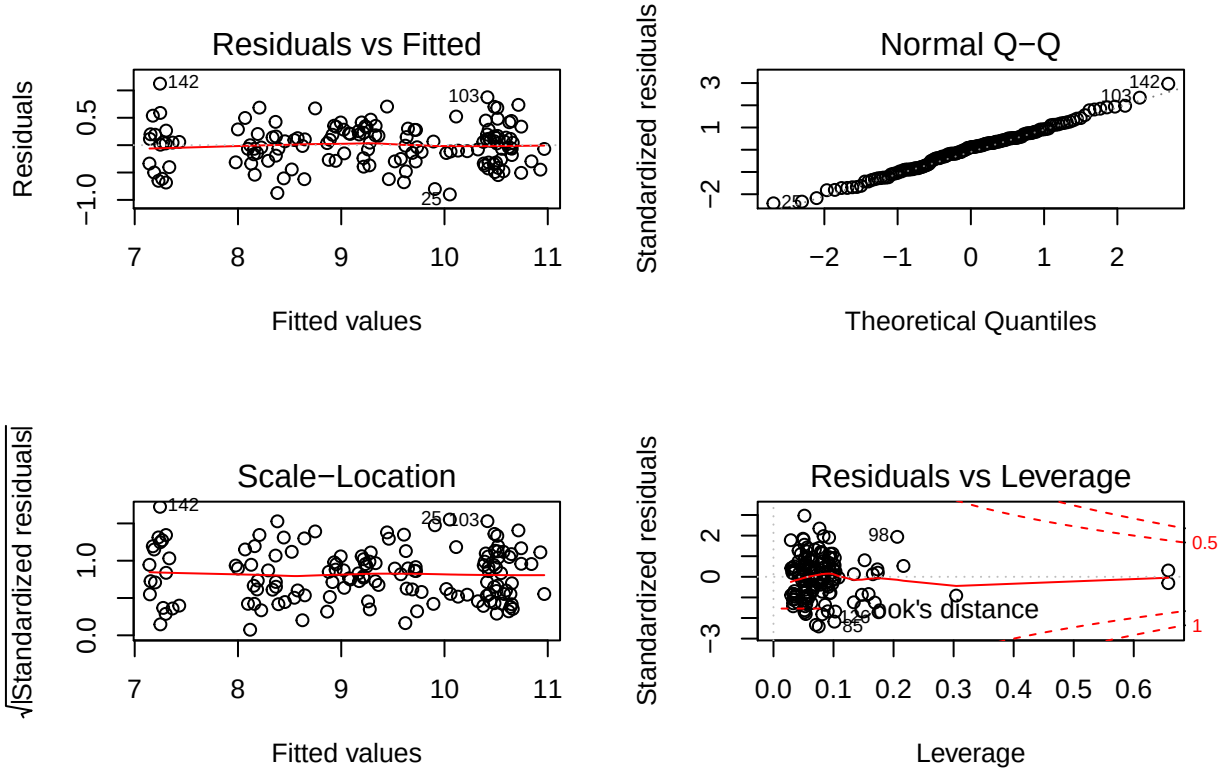
```
# Because my regression model has factorvariables and hence is a constant. The stepwise regression in R
#stepAIC(lm(GDP_pc~hc+FDI+Region+IncomeGroup+FDI*Region+FDI*IncomeGroup+hc*Region+hc*IncomeGroup, data
```

```
# Nested f-test
complex = lm(GDP_pc~hc+FDI+Region+IncomeGroup+FDI*Region+FDI*IncomeGroup+hc*Region+hc*IncomeGroup, data
nested = lm(log(GDP_pc)~hc+FDI+Region+IncomeGroup, data = GDP_2016)
anova(complex, nested, test='Chisq')
```

```
## Warning in anova.lmlist(object, ...): models with response '"log(GDP_pc)"'
## removed because response differs from model 1
```

```
## Analysis of Variance Table
##
## Response: GDP_pc
##                   Df     Sum Sq    Mean Sq  F value    Pr(>F)
## hc                 1 2.1387e+10 2.1387e+10 183.9349 < 2.2e-16 ***
## FDI                1 1.0016e+09 1.0016e+09   8.6139 0.0040418 **
## Region             6 3.5285e+09 5.8809e+08   5.0577 0.0001249 ***
## IncomeGroup        3 1.1172e+10 3.7241e+09  32.0280 4.653e-15 ***
## FDI:Region         6 1.0155e+08 1.6925e+07   0.1456 0.9895899
## FDI:IncomeGroup    3 8.2010e+07 2.7337e+07   0.2351 0.8717416
## hc:Region          5 3.6263e+08 7.2526e+07   0.6237 0.6819564
## hc:IncomeGroup     3 6.0077e+07 2.0026e+07   0.1722 0.9149693
## Residuals        113 1.3139e+10 1.1628e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
par(mfrow=c(2,2))
plot(reg_2016)
```




```
# This part shows the results of year 2017
GDP_2017 = subset(final_df, year==2017)
GDP_2017 = GDP_2017 %>% group_by(country) %>% mutate(GDP_pc=rgdpo/pop)
reg_2017 = lm(log(GDP_pc)~hc+FDI+Region+IncomeGroup, data = GDP_2017)
summary(reg_2017)
```

```
##
## Call:
## lm(formula = log(GDP_pc) ~ hc + FDI + Region + IncomeGroup, data = GDP_2017)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9221 -0.2647  0.0436  0.2234  1.1369
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       9.659e+00  3.016e-01  32.028  < 2e-16 ***
## hc                                2.690e-01  9.178e-02   2.931 0.003998 **
## FDI                               1.422e-12  8.899e-13   1.598 0.112511
## RegionEurope & Central Asia      -7.447e-02  1.146e-01  -0.650 0.517066
## RegionLatin America & Caribbean  -3.651e-01  1.239e-01  -2.948 0.003796 **
## RegionMiddle East & North Africa  2.288e-01  1.367e-01   1.674 0.096526 .
## RegionNorth America              -1.528e-01  3.256e-01  -0.469 0.639755
## RegionSouth Asia                 -1.381e-01  1.907e-01  -0.724 0.470269
## RegionSub-Saharan Africa         -4.629e-01  1.336e-01  -3.465 0.000719 ***
## IncomeGroupLow income            -2.375e+00  1.619e-01 -14.673  < 2e-16 ***
```

```
## IncomeGroupLower middle income   -1.505e+00  1.148e-01 -13.108  < 2e-16 ***
## IncomeGroupUpper middle income   -7.716e-01  9.743e-02  -7.920 9.27e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3838 on 130 degrees of freedom
##   (37 observations deleted due to missingness)
## Multiple R-squared:  0.9063, Adjusted R-squared:  0.8984
## F-statistic: 114.4 on 11 and 130 DF,  p-value: < 2.2e-16
```

```
exp(reg_2017$coefficients)
```

```
##                       (Intercept)                                hc
##                      1.566158e+04                      1.308621e+00
##                               FDI      RegionEurope & Central Asia
##                      1.000000e+00                      9.282386e-01
##  RegionLatin America & Caribbean RegionMiddle East & North Africa
##                      6.941177e-01                      1.257053e+00
##               RegionNorth America                  RegionSouth Asia
##                      8.583223e-01                      8.710303e-01
##         RegionSub-Saharan Africa             IncomeGroupLow income
##                      6.294564e-01                      9.300553e-02
##   IncomeGroupLower middle income   IncomeGroupUpper middle income
##                      2.221234e-01                      4.622584e-01
```

```
exp(confint(reg_2017))
```

```
##                                       2.5 %       97.5 %
## (Intercept)                      8.624133e+03 2.844170e+04
## hc                               1.091324e+00 1.569184e+00
## FDI                              1.000000e+00 1.000000e+00
## RegionEurope & Central Asia      7.399001e-01 1.164518e+00
## RegionLatin America & Caribbean  5.432642e-01 8.868603e-01
## RegionMiddle East & North Africa 9.592638e-01 1.647286e+00
## RegionNorth America              4.506633e-01 1.634740e+00
## RegionSouth Asia                 5.973188e-01 1.270166e+00
## RegionSub-Saharan Africa         4.832593e-01 8.198816e-01
## IncomeGroupLow income            6.751924e-02 1.281120e-01
## IncomeGroupLower middle income   1.770010e-01 2.787488e-01
## IncomeGroupUpper middle income   3.812164e-01 5.605289e-01
```

```
# Because my regression model has factorvariables and hence is a constant. The stepwise regression in R
#stepAIC(lm(GDP_pc~hc+FDI+Region+IncomeGroup+FDI*Region+FDI*IncomeGroup+hc*Region+hc*IncomeGroup, data
```

```
# Nested f-test
complex = lm(GDP_pc~hc+FDI+Region+IncomeGroup+FDI*Region+FDI*IncomeGroup+hc*Region+hc*IncomeGroup, data
nested = lm(log(GDP_pc)~hc+FDI+Region+IncomeGroup, data = GDP_2017)
anova(complex, nested, test='Chisq')
```

```
## Warning in anova.lmlist(object, ...): models with response '"log(GDP_pc)"'
## removed because response differs from model 1
```

```
## Analysis of Variance Table
##
## Response: GDP_pc
##                   Df     Sum Sq    Mean Sq  F value    Pr(>F)
## hc                 1 2.2968e+10 2.2968e+10 201.0057 < 2.2e-16 ***
## FDI                1 1.3291e+09 1.3291e+09  11.6318 0.0009000 ***
## Region             6 3.5172e+09 5.8620e+08   5.1303 0.0001074 ***
## IncomeGroup        3 1.1564e+10 3.8546e+09  33.7342 1.202e-15 ***
## FDI:Region         6 2.4573e+08 4.0956e+07   0.3584 0.9036349
## FDI:IncomeGroup    3 1.7763e+08 5.9210e+07   0.5182 0.6706052
## hc:Region          5 4.2323e+08 8.4646e+07   0.7408 0.5944661
## hc:IncomeGroup     3 4.0324e+07 1.3441e+07   0.1176 0.9495910
## Residuals        113 1.2912e+10 1.1426e+08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
par(mfrow=c(2,2))
plot(reg_2017)
```