GDP-Aanalytics

Oral Fredric

June 21, 2016

# Analyze the GDP and Educational data, Categorize based on Income Groups

### GDP Ranking Data

#### Year to year changes in the nominal level of output or income of an economy are affected by a combination of forces: real growth, price inflation, and exchange rates. Changes in any of the three can affect an economy's relative size and, therefore, its ranking in comparison to other economies. Of the rankings presented here, nominal GDP, perhaps the most familiar measure of aggregate economic activity, is most subject to price and exchange rate effects. Rankings are based on available data only.

### Education Statistics Data

#### The World Bank EdStats All Indicator Query holds around 3,000 internationally comparable indicators that describe education access, progression, completion, literacy, teachers, population, and expenditures. The indicators cover the education cycle from pre-primary to vocational and tertiary education.

### Download GPD Ranking Data

url <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FGDP.csv"  
gdpfile <- file.path(getwd(), "GDP.csv")  
download.file(url, gdpfile)

### Tidying the GDP Ranking Data

library(data.table)  
GDPdata <- data.table(read.csv(gdpfile))  
head(GDPdata)

## X Gross.domestic.product.2012 X.1 X.2 X.3 X.4 X.5  
## 1: NA NA  
## 2: NA (millions of NA  
## 3: Ranking NA Economy US dollars) NA  
## 4: NA NA  
## 5: USA 1 NA United States 16,244,600 NA  
## 6: CHN 2 NA China 8,227,103 NA  
## X.6 X.7 X.8  
## 1: NA NA NA  
## 2: NA NA NA  
## 3: NA NA NA  
## 4: NA NA NA  
## 5: NA NA NA  
## 6: NA NA NA

GDPdata <- data.table(read.csv(gdpfile, skip=4, nrows=215))  
head(GDPdata)

## X X.1 X.2 X.3 X.4 X.5 X.6 X.7 X.8 X.9  
## 1: USA 1 NA United States 16,244,600 NA NA NA NA  
## 2: CHN 2 NA China 8,227,103 NA NA NA NA  
## 3: JPN 3 NA Japan 5,959,718 NA NA NA NA  
## 4: DEU 4 NA Germany 3,428,131 NA NA NA NA  
## 5: FRA 5 NA France 2,612,878 NA NA NA NA  
## 6: GBR 6 NA United Kingdom 2,471,784 NA NA NA NA

GDPdata <- GDPdata[X != ""]  
GDPdata <- GDPdata[, list(X, X.1, X.3, X.4)]  
setnames(GDPdata, c("X", "X.1", "X.3", "X.4"), c("CountryCode", "rankingGDP", "Long.Name", "gdp"))  
head(GDPdata)

## CountryCode rankingGDP Long.Name gdp  
## 1: USA 1 United States 16,244,600   
## 2: CHN 2 China 8,227,103   
## 3: JPN 3 Japan 5,959,718   
## 4: DEU 4 Germany 3,428,131   
## 5: FRA 5 France 2,612,878   
## 6: GBR 6 United Kingdom 2,471,784

### Download Education Statistics Data

url <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FEDSTATS\_Country.csv"  
edfile <- file.path(getwd(), "EDSTATS\_Country.csv")  
download.file(url, edfile)

### Reading Education Statistics Data

EDdata <- data.table(read.csv(edfile))  
head(EDdata)

## CountryCode Long.Name Income.Group  
## 1: ABW Aruba High income: nonOECD  
## 2: ADO Principality of Andorra High income: nonOECD  
## 3: AFG Islamic State of Afghanistan Low income  
## 4: AGO People's Republic of Angola Lower middle income  
## 5: ALB Republic of Albania Upper middle income  
## 6: ARE United Arab Emirates High income: nonOECD  
## Region Lending.category Other.groups Currency.Unit  
## 1: Latin America & Caribbean Aruban florin  
## 2: Europe & Central Asia Euro  
## 3: South Asia IDA HIPC Afghan afghani  
## 4: Sub-Saharan Africa IDA Angolan kwanza  
## 5: Europe & Central Asia IBRD Albanian lek  
## 6: Middle East & North Africa U.A.E. dirham  
## Latest.population.census Latest.household.survey  
## 1: 2000   
## 2: Register based   
## 3: 1979 MICS, 2003  
## 4: 1970 MICS, 2001, MIS, 2006/07  
## 5: 2001 MICS, 2005  
## 6: 2005   
## Special.Notes  
## 1:   
## 2:   
## 3: Fiscal year end: March 20; reporting period for national accounts data: FY.  
## 4:   
## 5:   
## 6:   
## National.accounts.base.year National.accounts.reference.year  
## 1: 1995 NA  
## 2: NA  
## 3: 2002/2003 NA  
## 4: 1997 NA  
## 5: 1996  
## 6: 1995 NA  
## System.of.National.Accounts SNA.price.valuation  
## 1: NA   
## 2: NA   
## 3: NA VAB  
## 4: NA VAP  
## 5: 1993 VAB  
## 6: NA VAB  
## Alternative.conversion.factor PPP.survey.year  
## 1: NA  
## 2: NA  
## 3: NA  
## 4: 1991-96 2005  
## 5: 2005  
## 6: NA  
## Balance.of.Payments.Manual.in.use External.debt.Reporting.status  
## 1:   
## 2:   
## 3: Actual  
## 4: BPM5 Actual  
## 5: BPM5 Actual  
## 6: BPM4   
## System.of.trade Government.Accounting.concept  
## 1: Special   
## 2: General   
## 3: General Consolidated  
## 4: Special   
## 5: General Consolidated  
## 6: General Consolidated  
## IMF.data.dissemination.standard  
## 1:   
## 2:   
## 3: GDDS  
## 4: GDDS  
## 5: GDDS  
## 6: GDDS  
## Source.of.most.recent.Income.and.expenditure.data  
## 1:   
## 2:   
## 3:   
## 4: IHS, 2000  
## 5: LSMS, 2005  
## 6:   
## Vital.registration.complete Latest.agricultural.census  
## 1:   
## 2: Yes   
## 3:   
## 4: 1964-65  
## 5: Yes 1998  
## 6: 1998  
## Latest.industrial.data Latest.trade.data Latest.water.withdrawal.data  
## 1: NA 2008 NA  
## 2: NA 2006 NA  
## 3: NA 2008 2000  
## 4: NA 1991 2000  
## 5: 2005 2008 2000  
## 6: NA 2008 2005  
## X2.alpha.code WB.2.code Table.Name Short.Name  
## 1: AW AW Aruba Aruba  
## 2: AD AD Andorra Andorra  
## 3: AF AF Afghanistan Afghanistan  
## 4: AO AO Angola Angola  
## 5: AL AL Albania Albania  
## 6: AE AE United Arab Emirates United Arab Emirates

### Match the Data based on CountyCode and Count How many IDs match

GDPmergeED <- merge(GDPdata, EDdata, all=TRUE, by=c("CountryCode"))  
head(GDPmergeED)

## CountryCode rankingGDP Long.Name.x gdp  
## 1: ABW 161 Aruba 2,584   
## 2: ADO NA Andorra ..  
## 3: AFG 105 Afghanistan 20,497   
## 4: AGO 60 Angola 114,147   
## 5: ALB 125 Albania 12,648   
## 6: ARE 32 United Arab Emirates 348,595   
## Long.Name.y Income.Group  
## 1: Aruba High income: nonOECD  
## 2: Principality of Andorra High income: nonOECD  
## 3: Islamic State of Afghanistan Low income  
## 4: People's Republic of Angola Lower middle income  
## 5: Republic of Albania Upper middle income  
## 6: United Arab Emirates High income: nonOECD  
## Region Lending.category Other.groups Currency.Unit  
## 1: Latin America & Caribbean Aruban florin  
## 2: Europe & Central Asia Euro  
## 3: South Asia IDA HIPC Afghan afghani  
## 4: Sub-Saharan Africa IDA Angolan kwanza  
## 5: Europe & Central Asia IBRD Albanian lek  
## 6: Middle East & North Africa U.A.E. dirham  
## Latest.population.census Latest.household.survey  
## 1: 2000   
## 2: Register based   
## 3: 1979 MICS, 2003  
## 4: 1970 MICS, 2001, MIS, 2006/07  
## 5: 2001 MICS, 2005  
## 6: 2005   
## Special.Notes  
## 1:   
## 2:   
## 3: Fiscal year end: March 20; reporting period for national accounts data: FY.  
## 4:   
## 5:   
## 6:   
## National.accounts.base.year National.accounts.reference.year  
## 1: 1995 NA  
## 2: NA  
## 3: 2002/2003 NA  
## 4: 1997 NA  
## 5: 1996  
## 6: 1995 NA  
## System.of.National.Accounts SNA.price.valuation  
## 1: NA   
## 2: NA   
## 3: NA VAB  
## 4: NA VAP  
## 5: 1993 VAB  
## 6: NA VAB  
## Alternative.conversion.factor PPP.survey.year  
## 1: NA  
## 2: NA  
## 3: NA  
## 4: 1991-96 2005  
## 5: 2005  
## 6: NA  
## Balance.of.Payments.Manual.in.use External.debt.Reporting.status  
## 1:   
## 2:   
## 3: Actual  
## 4: BPM5 Actual  
## 5: BPM5 Actual  
## 6: BPM4   
## System.of.trade Government.Accounting.concept  
## 1: Special   
## 2: General   
## 3: General Consolidated  
## 4: Special   
## 5: General Consolidated  
## 6: General Consolidated  
## IMF.data.dissemination.standard  
## 1:   
## 2:   
## 3: GDDS  
## 4: GDDS  
## 5: GDDS  
## 6: GDDS  
## Source.of.most.recent.Income.and.expenditure.data  
## 1:   
## 2:   
## 3:   
## 4: IHS, 2000  
## 5: LSMS, 2005  
## 6:   
## Vital.registration.complete Latest.agricultural.census  
## 1:   
## 2: Yes   
## 3:   
## 4: 1964-65  
## 5: Yes 1998  
## 6: 1998  
## Latest.industrial.data Latest.trade.data Latest.water.withdrawal.data  
## 1: NA 2008 NA  
## 2: NA 2006 NA  
## 3: NA 2008 2000  
## 4: NA 1991 2000  
## 5: 2005 2008 2000  
## 6: NA 2008 2005  
## X2.alpha.code WB.2.code Table.Name Short.Name  
## 1: AW AW Aruba Aruba  
## 2: AD AD Andorra Andorra  
## 3: AF AF Afghanistan Afghanistan  
## 4: AO AO Angola Angola  
## 5: AL AL Albania Albania  
## 6: AE AE United Arab Emirates United Arab Emirates

str(GDPmergeED)

## Classes 'data.table' and 'data.frame': 238 obs. of 34 variables:  
## $ CountryCode : Factor w/ 239 levels "","ABW","ADO",..: 2 3 4 5 6 7 8 9 10 11 ...  
## $ rankingGDP : int 161 NA 105 60 125 32 26 133 NA 172 ...  
## $ Long.Name.x : Factor w/ 215 levels "","Afghanistan",..: 11 6 2 7 3 203 9 10 5 8 ...  
## $ gdp : Factor w/ 191 levels ""," 1,008 "," 1,129 ",..: 65 191 68 24 26 109 134 186 191 5 ...  
## $ Long.Name.y : Factor w/ 234 levels "American Samoa",..: 5 104 57 99 109 226 4 110 1 2 ...  
## $ Income.Group : Factor w/ 6 levels "","High income: nonOECD",..: 2 2 4 5 6 2 6 5 6 6 ...  
## $ Region : Factor w/ 8 levels "","East Asia & Pacific",..: 4 3 7 8 3 5 4 3 2 4 ...  
## $ Lending.category : Factor w/ 4 levels "","Blend","IBRD",..: 1 1 4 4 3 1 3 2 1 3 ...  
## $ Other.groups : Factor w/ 3 levels "","Euro area",..: 1 1 3 1 1 1 1 1 1 1 ...  
## $ Currency.Unit : Factor w/ 155 levels "","Afghan afghani",..: 8 49 2 5 3 144 6 7 145 44 ...  
## $ Latest.population.census : Factor w/ 28 levels "","1970","1979",..: 17 28 3 2 18 22 18 18 17 18 ...  
## $ Latest.household.survey : Factor w/ 56 levels "","CPS (monthly)",..: 1 1 39 38 40 1 1 16 1 1 ...  
## $ Special.Notes : Factor w/ 70 levels "","A simple multiplier is used to convert the national currencies of EMU members to euros. The following irrevocable euro conversi"| \_\_truncated\_\_,..: 1 1 27 1 1 1 1 1 1 63 ...  
## $ National.accounts.base.year : Factor w/ 44 levels "","1954","1973",..: 25 1 38 28 1 25 22 1 1 18 ...  
## $ National.accounts.reference.year : int NA NA NA NA 1996 NA NA 1996 NA NA ...  
## $ System.of.National.Accounts : int NA NA NA NA 1993 NA 1993 1993 NA NA ...  
## $ SNA.price.valuation : Factor w/ 3 levels "","VAB","VAP": 1 1 2 3 2 2 2 2 1 2 ...  
## $ Alternative.conversion.factor : Factor w/ 33 levels "","1960-85","1965-84",..: 1 1 1 24 1 1 6 21 1 1 ...  
## $ PPP.survey.year : int NA NA NA 2005 2005 NA 2005 2005 NA NA ...  
## $ Balance.of.Payments.Manual.in.use : Factor w/ 3 levels "","BPM4","BPM5": 1 1 1 3 3 2 3 3 1 3 ...  
## $ External.debt.Reporting.status : Factor w/ 4 levels "","Actual","Estimate",..: 1 1 2 2 2 1 2 2 1 1 ...  
## $ System.of.trade : Factor w/ 3 levels "","General","Special": 3 2 2 3 2 2 3 3 1 2 ...  
## $ Government.Accounting.concept : Factor w/ 3 levels "","Budgetary",..: 1 1 3 1 3 3 3 3 1 1 ...  
## $ IMF.data.dissemination.standard : Factor w/ 3 levels "","GDDS","SDDS": 1 1 2 2 2 2 3 3 1 2 ...  
## $ Source.of.most.recent.Income.and.expenditure.data: Factor w/ 77 levels "","1-2-3, 2005-06",..: 1 1 1 35 66 1 45 46 1 1 ...  
## $ Vital.registration.complete : Factor w/ 2 levels "","Yes": 1 2 1 1 2 1 2 2 2 2 ...  
## $ Latest.agricultural.census : Factor w/ 45 levels "","1960","1964-65",..: 1 1 1 3 32 32 41 1 1 1 ...  
## $ Latest.industrial.data : int NA NA NA NA 2005 NA 2001 NA NA NA ...  
## $ Latest.trade.data : int 2008 2006 2008 1991 2008 2008 2008 2008 NA 2007 ...  
## $ Latest.water.withdrawal.data : int NA NA 2000 2000 2000 2005 2000 2000 NA 1990 ...  
## $ X2.alpha.code : Factor w/ 208 levels "","AD","AE","AF",..: 13 2 4 8 6 3 9 7 10 5 ...  
## $ WB.2.code : Factor w/ 209 levels "","AD","AE","AF",..: 13 2 4 8 6 3 9 7 10 5 ...  
## $ Table.Name : Factor w/ 234 levels "Afghanistan",..: 10 5 1 6 2 220 8 9 4 7 ...  
## $ Short.Name : Factor w/ 234 levels "Afghanistan",..: 10 5 1 6 2 220 8 9 4 7 ...  
## - attr(\*, ".internal.selfref")=<externalptr>   
## - attr(\*, "sorted")= chr "CountryCode"

sum(!is.na(unique(GDPmergeED$rankingGDP)))

## [1] 189

### Sort the data frame in ascending order by GDP rank and find 13th one

GDPmergeED[order(rankingGDP, decreasing=TRUE), list(CountryCode, Long.Name.x, Long.Name.y, rankingGDP, gdp)][13]

## CountryCode Long.Name.x Long.Name.y rankingGDP gdp  
## 1: KNA St. Kitts and Nevis St. Kitts and Nevis 178 767

### Find the average GDP rankings for the "High income: OECD" and "High income: nonOECD" groups

GDPmergeED[, mean(rankingGDP, na.rm=TRUE), by=Income.Group]

## Income.Group V1  
## 1: High income: nonOECD 91.91304  
## 2: Low income 133.72973  
## 3: Lower middle income 107.70370  
## 4: Upper middle income 92.13333  
## 5: High income: OECD 32.96667  
## 6: NA 131.00000  
## 7: NaN

### GDP ranking into 5 separate quantile groups. Make a table versus Income.Group. How many countries are Lower middle income but among the 38 nations with highest GDP

breaks <- quantile(GDPmergeED$rankingGDP, probs=seq(0, 1, 0.2), na.rm=TRUE)  
GDPmergeED$quantileGDP <- cut(GDPmergeED$rankingGDP, breaks=breaks)  
GDPmergeED[Income.Group == "Lower middle income", .N, by=c("Income.Group", "quantileGDP")]

## Income.Group quantileGDP N  
## 1: Lower middle income (38.8,76.6] 13  
## 2: Lower middle income (114,152] 8  
## 3: Lower middle income (152,190] 16  
## 4: Lower middle income (76.6,114] 12  
## 5: Lower middle income (1,38.8] 5  
## 6: Lower middle income NA 2

### Plot the GDP for all of the countries. Use ggplot2 to color your plot by Income Group

library(ggplot2)  
ggplot(GDPmergeED,aes(CountryCode, gdp)) + geom\_point(aes(color = Income.Group))

