MSDS 6306 401 Case Study 1

Yao Yao

March 4, 2017

Introduction:

Import the required libraries needed for the gathering and analysis procedures

```
#require library(doumloader), needed for gathering
#require library(ggplot2), needed for plotting
#require library(reshape2), needed for creating a quantile table
```

==Gather GDP and Educational data==

Directory set up and confirmation of files and location

```
#setwd(".\\Data") # set as data directory
getwd()
```

```
## [1] "C:/Users/Yao/Documents/GitHub/DDS-HW8/Data"
```

```
list.files()
```

```
##  [1] "EducationalWeb.csv"
##  [2] "EducationRaw.csv"
##  [3] "GDPData.csv"
##  [4] "GDPRaw.csv"
##  [5] "GDPWeb.csv"
##  [6] "HINonOECD.csv"
##  [7] "HIOECD.csv"
##  [8] "IncomeSub.csv"
##  [9] "LowerMiddleTop38.csv"
## [10] "MergeData1.csv"
## [11] "MergeData2.csv"
## [12] "NegGDP.csv"
## [13] "Quantiles.csv"
## [14] "Yao Yao MSDS 6306 401 Case Study1 Makefile.txt"
## [15] "Yao Yao MSDS 6306 401 Case Study1.R"
## [16] "Yao Yao MSDS 6306 401 Case Study1.Rmd"
## [17] "Yao_Yao_MSDS_6306_401_Case_Study1.Rmd"
## [18] "Yao_Yao_MSDS_6306_401_Case_Study1_files"
```

Download files via internet, read files into csv into respective destination files

```
library(downloader)
download('https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FGDP.csv', destfile='GDPWeb.csv')
download('https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FEDSTATS_Country.csv', destfile='Educati
list.files() #make sure the files are there
```

```
##  [1] "EducationalWeb.csv"
##  [2] "EducationRaw.csv"
##  [3] "GDPData.csv"
##  [4] "GDPRaw.csv"
##  [5] "GDPWeb.csv"
##  [6] "HINonOECD.csv"
```

```
##  [7] "HIOECD.csv"
##  [8] "IncomeSub.csv"
##  [9] "LowerMiddleTop38.csv"
## [10] "MergeData1.csv"
## [11] "MergeData2.csv"
## [12] "NegGDP.csv"
## [13] "Quantiles.csv"
## [14] "Yao Yao MSDS 6306 401 Case Study1 Makefile.txt"
## [15] "Yao Yao MSDS 6306 401 Case Study1.R"
## [16] "Yao Yao MSDS 6306 401 Case Study1.Rmd"
## [17] "Yao_Yao_MSDS_6306_401_Case_Study1.Rmd"
## [18] "Yao_Yao_MSDS_6306_401_Case_Study1_files"
```

==clean GDP Data==

Import csv into raw GDP file and examine attributes

```
GDPRaw <- read.csv('GDPWeb.csv',stringsAsFactors = FALSE, header = TRUE)
str(GDPRaw)
```

```
## 'data.frame':    330 obs. of  10 variables:
##  $ X                           : chr  "" "" "" "" ...
##  $ Gross.domestic.product.2012: chr  "" "" "Ranking" "" ...
##  $ X.1                         : logi  NA NA NA NA NA NA ...
##  $ X.2                         : chr  "" "" "Economy" "" ...
##  $ X.3                         : chr  "" "(millions of" "US dollars)" "" ...
##  $ X.4                         : chr  "" "" "" "" ...
##  $ X.5                         : logi  NA NA NA NA NA NA ...
##  $ X.6                         : logi  NA NA NA NA NA NA ...
##  $ X.7                         : logi  NA NA NA NA NA NA ...
##  $ X.8                         : logi  NA NA NA NA NA NA ...
```

Eliminate header rows, rows without GDP, and unused columns Reset column count and examine data

```
GDPData <- GDPRaw[5:194,1:5]
rownames(GDPData) <- seq(length=nrow(GDPData))
head(GDPData)
```

```
##     X Gross.domestic.product.2012 X.1            X.2          X.3
## 1 USA                           1  NA  United States  16,244,600
## 2 CHN                           2  NA          China   8,227,103
## 3 JPN                           3  NA          Japan   5,959,718
## 4 DEU                           4  NA        Germany   3,428,131
## 5 FRA                           5  NA         France   2,612,878
## 6 GBR                           6  NA United Kingdom   2,471,784
```

```
tail(GDPData)
```

```
##       X Gross.domestic.product.2012 X.1                   X.2   X.3
## 185 FSM                         185  NA Micronesia, Fed. Sts.  326
## 186 STP                         186  NA São Tomé and Principe  263
## 187 PLW                         187  NA                Palau  228
## 188 MHL                         188  NA      Marshall Islands  182
## 189 KIR                         189  NA              Kiribati  175
## 190 TUV                         190  NA                Tuvalu   40
```

Retitle columns and remove unused columns

```
colnames(GDPData) <- c("CountryCode","Ranking","x","Economy","US Dollars (millions)")
head(GDPData)
```

```
##   CountryCode Ranking  x        Economy US Dollars (millions)
## 1         USA       1 NA  United States           16,244,600
## 2         CHN       2 NA          China            8,227,103
## 3         JPN       3 NA          Japan            5,959,718
## 4         DEU       4 NA        Germany            3,428,131
## 5         FRA       5 NA         France            2,612,878
## 6         GBR       6 NA United Kingdom            2,471,784
```

```
GDPData <- GDPData[,c("CountryCode","Ranking","Economy","US Dollars (millions)")]
head(GDPData)
```

```
##   CountryCode Ranking        Economy US Dollars (millions)
## 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
```

Set the ranking as integer and GDP as numeric for later analysis, recheck attributes

```
GDPData$Ranking <- as.integer(GDPData$Ranking)
GDPData$`US Dollars (millions)` <- as.numeric(gsub(",", "", GDPData$`US Dollars (millions)`))
str(GDPData)
```

```
## 'data.frame':    190 obs. of  4 variables:
##  $ CountryCode          : chr  "USA" "CHN" "JPN" "DEU" ...
##  $ Ranking              : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Economy              : chr  "United States" "China" "Japan" "Germany" ...
##  $ US Dollars (millions): num  16244600 8227103 5959718 3428131 2612878 ...
```

```
dim(GDPData)
```

```
## [1] 190   4
```

==Clean Educational Data==

Import raw education data from csv and examine attributes

```
EducationRaw <- read.csv('EducationalWeb.csv',stringsAsFactors = FALSE, header = TRUE)
str(EducationRaw)
```

```
## 'data.frame':    234 obs. of  31 variables:
##  $ CountryCode                                : chr  "ABW" "ADO" "AFG" "AGO" ...
##  $ Long.Name                                  : chr  "Aruba" "Principality of Andorra" "Islami
##  $ Income.Group                               : chr  "High income: nonOECD" "High income: nonO
##  $ Region                                     : chr  "Latin America & Caribbean" "Europe & Cen
##  $ Lending.category                           : chr  "" "" "IDA" "IDA" ...
##  $ Other.groups                               : chr  "" "" "HIPC" "" ...
##  $ Currency.Unit                              : chr  "Aruban florin" "Euro" "Afghan afghani" "A
##  $ Latest.population.census                   : chr  "2000" "Register based" "1979" "1970" ...
##  $ Latest.household.survey                    : chr  "" "" "MICS, 2003" "MICS, 2001, MIS, 2006
##  $ Special.Notes                              : chr  "" "" "Fiscal year end: March 20; reporti
##  $ National.accounts.base.year                : chr  "1995" "" "2002/2003" "1997" ...
##  $ 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                                : chr  "" "" "VAB" "VAP" ...
##  $ Alternative.conversion.factor                      : chr  "" "" "" "1991-96" ...
##  $ PPP.survey.year                                    : int  NA NA NA 2005 2005 NA 2005 2005 NA NA ...
##  $ Balance.of.Payments.Manual.in.use                  : chr  "" "" "" "BPM5" ...
##  $ External.debt.Reporting.status                     : chr  "" "" "Actual" "Actual" ...
##  $ System.of.trade                                    : chr  "Special" "General" "General" "Special" .
##  $ Government.Accounting.concept                      : chr  "" "" "Consolidated" "" ...
##  $ IMF.data.dissemination.standard                    : chr  "" "" "GDDS" "GDDS" ...
##  $ Source.of.most.recent.Income.and.expenditure.data: chr  "" "" "" "IHS, 2000" ...
##  $ Vital.registration.complete                        : chr  "" "Yes" "" "" ...
##  $ Latest.agricultural.census                         : chr  "" "" "" "1964-65" ...
##  $ 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
##  $ Latest.water.withdrawal.data                       : int  NA NA 2000 2000 2000 2005 2000 2000 NA 199
##  $ X2.alpha.code                                      : chr  "AW" "AD" "AF" "AO" ...
##  $ WB.2.code                                          : chr  "AW" "AD" "AF" "AO" ...
##  $ Table.Name                                         : chr  "Aruba" "Andorra" "Afghanistan" "Angola"
##  $ Short.Name                                         : chr  "Aruba" "Andorra" "Afghanistan" "Angola"
```

```
dim(EducationRaw)
```

```
## [1] 234  31
```

==Merge Education and GDP data==

Merge all columns from raw GDP data and Education based by country code Save into raw merge file and examine attributes

```
MergeData1 <- merge(x = GDPData, y = EducationRaw, by ='CountryCode', all=TRUE)
head(MergeData1)
```

```
##   CountryCode Ranking              Economy US Dollars (millions)
## 1         ABW     161                Aruba                  2584
## 2         ADO      NA                 <NA>                    NA
## 3         AFG     105          Afghanistan                 20497
## 4         AGO      60               Angola                114147
## 5         ALB     125              Albania                 12648
## 6         ARE      32 United Arab Emirates                348595
##                       Long.Name         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
```

```
tail(MergeData1)
```

```
##     CountryCode Ranking          Economy US Dollars (millions)
## 230         WSM     181            Samoa                   684
## 231         YEM      90      Yemen, Rep.                 35646
## 232         ZAF      28     South Africa                384313
## 233         ZAR     112 Congo, Dem. Rep.                 17204
## 234         ZMB     104           Zambia                 20678
## 235         ZWE     134         Zimbabwe                  9802
##                            Long.Name        Income.Group
## 230                           Samoa Lower middle income
## 231                 Republic of Yemen Lower middle income
## 232         Republic of South Africa Upper middle income
## 233 Democratic Republic of the Congo          Low income
## 234                Republic of Zambia          Low income
## 235              Republic of Zimbabwe          Low income
##                         Region Lending.category Other.groups
## 230        East Asia & Pacific              IDA
## 231 Middle East & North Africa              IDA
## 232         Sub-Saharan Africa             IBRD
## 233         Sub-Saharan Africa              IDA         HIPC
## 234         Sub-Saharan Africa              IDA         HIPC
## 235         Sub-Saharan Africa            Blend
##          Currency.Unit Latest.population.census Latest.household.survey
## 230        Samoan tala                    2006
## 231        Yemeni rial                    2004              MICS, 2006
## 232 South African rand                    2001               DHS, 2003
## 233    Congolese franc                    1984                DHS 2007
```

```
## 234     Zambian kwacha                      2000              DHS, 2007
## 235    Zimbabwe dollar                      2002            DHS, 2005/06
##                                                                Special.Notes
## 230
## 231
## 232 Fiscal year end: March 31; reporting period for national accounts data: CY.
## 233
## 234
## 235  Fiscal year end: June 30; reporting period for national accounts data: CY.
##     National.accounts.base.year National.accounts.reference.year
## 230                        2002                                NA
## 231                        1990                                NA
## 232                        2000                                NA
## 233                        1987                                NA
## 234                        1994                                NA
## 235                        1990                                NA
##     System.of.National.Accounts SNA.price.valuation
## 230                          NA                 VAB
## 231                          NA                 VAP
## 232                        1993                 VAB
## 233                        1993                 VAB
## 234                          NA                 VAB
## 235                          NA                 VAB
##     Alternative.conversion.factor PPP.survey.year
## 230                                           NA
## 231                       1990-96            2005
## 232                                          2005
## 233                       1999-01            2005
## 234                       1990-92            2005
## 235                     1991, 1998           2005
##     Balance.of.Payments.Manual.in.use External.debt.Reporting.status
## 230                              BPM5                    Preliminary
## 231                              BPM5                         Actual
## 232                              BPM5                    Preliminary
## 233                              BPM5                       Estimate
## 234                              BPM5                    Preliminary
## 235                              BPM5                         Actual
##     System.of.trade Government.Accounting.concept
## 230         General
## 231         General                     Budgetary
## 232         General                  Consolidated
## 233         Special                  Consolidated
## 234         General                     Budgetary
## 235         General                  Consolidated
##     IMF.data.dissemination.standard
## 230
## 231                            GDDS
## 232                            SDDS
## 233                            GDDS
## 234                            GDDS
## 235                            GDDS
##     Source.of.most.recent.Income.and.expenditure.data
## 230
## 231                                    ES/BS, 2005
```

```
## 232                                    ES/BS, 2000
## 233                                  1-2-3, 2005-06
## 234                                    IHS, 2004-05
## 235
##     Vital.registration.complete Latest.agricultural.census
## 230                                                  1999
## 231                                                  2002
## 232                                                  2000
## 233                                                  1990
## 234                                                  1990
## 235                                                  1960
##     Latest.industrial.data Latest.trade.data Latest.water.withdrawal.data
## 230                     NA              2008                           NA
## 231                   2005              2008                         2000
## 232                   2005              2008                         2000
## 233                     NA              1986                         2000
## 234                     NA              2008                         2000
## 235                   1995              2008                         2002
##     X2.alpha.code WB.2.code       Table.Name      Short.Name
## 230            WS        WS            Samoa           Samoa
## 231            YE        RY      Yemen, Rep.           Yemen
## 232            ZA        ZA     South Africa    South Africa
## 233            CD        ZR Congo, Dem. Rep. Dem. Rep. Congo
## 234            ZM        ZM           Zambia          Zambia
## 235            ZW        ZW         Zimbabwe        Zimbabwe
```

```
dim(MergeData1)
```

```
## [1] 235  34
```

Remove merged columns based on lack of value in GDP, country, GDP ranking, or income group Examine attributes and reset column count based on country code

```
MergeData2<-MergeData1[rowSums(is.na(MergeData1[,2:5]))==FALSE,]
head(MergeData2)
```

```
##   CountryCode Ranking              Economy US Dollars (millions)
## 1         ABW     161                Aruba                  2584
## 3         AFG     105          Afghanistan                 20497
## 4         AGO      60               Angola                114147
## 5         ALB     125              Albania                 12648
## 6         ARE      32 United Arab Emirates                348595
## 7         ARG      26            Argentina                475502
##                       Long.Name         Income.Group
## 1                         Aruba 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
## 7           Argentine Republic  Upper middle income
##                         Region Lending.category Other.groups  Currency.Unit
## 1  Latin America & Caribbean                                    Aruban florin
## 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
```

```
## 7  Latin America & Caribbean             IBRD              Argentine peso
##   Latest.population.census  Latest.household.survey
## 1                     2000
## 3                     1979               MICS, 2003
## 4                     1970 MICS, 2001, MIS, 2006/07
## 5                     2001               MICS, 2005
## 6                     2005
## 7                     2001
##                                                                Special.Notes
## 1
## 3 Fiscal year end: March 20; reporting period for national accounts data: FY.
## 4
## 5
## 6
## 7
##   National.accounts.base.year National.accounts.reference.year
## 1                        1995                                NA
## 3                   2002/2003                                NA
## 4                        1997                                NA
## 5                                                          1996
## 6                        1995                                NA
## 7                        1993                                NA
##   System.of.National.Accounts SNA.price.valuation
## 1                          NA
## 3                          NA                VAB
## 4                          NA                VAP
## 5                        1993                VAB
## 6                          NA                VAB
## 7                        1993                VAB
##   Alternative.conversion.factor PPP.survey.year
## 1                                            NA
## 3                                            NA
## 4                       1991-96            2005
## 5                                          2005
## 6                                            NA
## 7                       1971-84            2005
##   Balance.of.Payments.Manual.in.use External.debt.Reporting.status
## 1
## 3                                                          Actual
## 4                             BPM5                         Actual
## 5                             BPM5                         Actual
## 6                             BPM4
## 7                             BPM5                         Actual
##   System.of.trade Government.Accounting.concept
## 1         Special
## 3         General                  Consolidated
## 4         Special
## 5         General                  Consolidated
## 6         General                  Consolidated
## 7         Special                  Consolidated
##   IMF.data.dissemination.standard
## 1
## 3                            GDDS
## 4                            GDDS
```

```
## 5                            GDDS
## 6                            GDDS
## 7                            SDDS
##   Source.of.most.recent.Income.and.expenditure.data
## 1
## 3
## 4                                       IHS, 2000
## 5                                      LSMS, 2005
## 6
## 7                                       IHS, 2006
##   Vital.registration.complete Latest.agricultural.census
## 1
## 3
## 4                                           1964-65
## 5                         Yes                  1998
## 6                                              1998
## 7                         Yes                  2002
##   Latest.industrial.data Latest.trade.data Latest.water.withdrawal.data
## 1                     NA              2008                           NA
## 3                     NA              2008                         2000
## 4                     NA              1991                         2000
## 5                   2005              2008                         2000
## 6                     NA              2008                         2005
## 7                   2001              2008                         2000
##   X2.alpha.code WB.2.code           Table.Name           Short.Name
## 1            AW        AW                Aruba                Aruba
## 3            AF        AF          Afghanistan          Afghanistan
## 4            AO        AO               Angola               Angola
## 5            AL        AL              Albania              Albania
## 6            AE        AE United Arab Emirates United Arab Emirates
## 7            AR        AR            Argentina            Argentina
```

```
tail(MergeData2)
```

```
##     CountryCode Ranking          Economy US Dollars (millions)
## 230         WSM     181            Samoa                   684
## 231         YEM      90      Yemen, Rep.                 35646
## 232         ZAF      28     South Africa                384313
## 233         ZAR     112 Congo, Dem. Rep.                 17204
## 234         ZMB     104           Zambia                 20678
## 235         ZWE     134         Zimbabwe                  9802
##                            Long.Name        Income.Group
## 230                           Samoa Lower middle income
## 231                Republic of Yemen Lower middle income
## 232         Republic of South Africa Upper middle income
## 233 Democratic Republic of the Congo          Low income
## 234                Republic of Zambia          Low income
## 235              Republic of Zimbabwe          Low income
##                         Region Lending.category Other.groups
## 230        East Asia & Pacific              IDA
## 231 Middle East & North Africa              IDA
## 232         Sub-Saharan Africa             IBRD
## 233         Sub-Saharan Africa              IDA         HIPC
## 234         Sub-Saharan Africa              IDA         HIPC
## 235         Sub-Saharan Africa            Blend
```

```
##          Currency.Unit Latest.population.census Latest.household.survey
## 230        Samoan tala                     2006
## 231        Yemeni rial                     2004              MICS, 2006
## 232 South African rand                     2001               DHS, 2003
## 233    Congolese franc                     1984                DHS 2007
## 234     Zambian kwacha                     2000               DHS, 2007
## 235    Zimbabwe dollar                     2002             DHS, 2005/06
##                                                                Special.Notes
## 230
## 231
## 232 Fiscal year end: March 31; reporting period for national accounts data: CY.
## 233
## 234
## 235  Fiscal year end: June 30; reporting period for national accounts data: CY.
##     National.accounts.base.year National.accounts.reference.year
## 230                        2002                                NA
## 231                        1990                                NA
## 232                        2000                                NA
## 233                        1987                                NA
## 234                        1994                                NA
## 235                        1990                                NA
##     System.of.National.Accounts SNA.price.valuation
## 230                          NA                VAB
## 231                          NA                VAP
## 232                        1993                VAB
## 233                        1993                VAB
## 234                          NA                VAB
## 235                          NA                VAB
##     Alternative.conversion.factor PPP.survey.year
## 230                                           NA
## 231                       1990-96            2005
## 232                                          2005
## 233                       1999-01            2005
## 234                       1990-92            2005
## 235                    1991, 1998            2005
##     Balance.of.Payments.Manual.in.use External.debt.Reporting.status
## 230                              BPM5                    Preliminary
## 231                              BPM5                         Actual
## 232                              BPM5                    Preliminary
## 233                              BPM5                       Estimate
## 234                              BPM5                    Preliminary
## 235                              BPM5                         Actual
##     System.of.trade Government.Accounting.concept
## 230         General
## 231         General                    Budgetary
## 232         General                 Consolidated
## 233         Special                 Consolidated
## 234         General                    Budgetary
## 235         General                 Consolidated
##     IMF.data.dissemination.standard
## 230
## 231                            GDDS
## 232                            SDDS
## 233                            GDDS
```

```
## 234                             GDDS
## 235                             GDDS
##     Source.of.most.recent.Income.and.expenditure.data
## 230
## 231                                     ES/BS, 2005
## 232                                     ES/BS, 2000
## 233                                  1-2-3, 2005-06
## 234                                    IHS, 2004-05
## 235
##     Vital.registration.complete Latest.agricultural.census
## 230                                                   1999
## 231                                                   2002
## 232                                                   2000
## 233                                                   1990
## 234                                                   1990
## 235                                                   1960
##     Latest.industrial.data Latest.trade.data Latest.water.withdrawal.data
## 230                     NA              2008                           NA
## 231                   2005              2008                         2000
## 232                   2005              2008                         2000
## 233                     NA              1986                         2000
## 234                     NA              2008                         2000
## 235                   1995              2008                         2002
##     X2.alpha.code WB.2.code       Table.Name      Short.Name
## 230            WS        WS            Samoa           Samoa
## 231            YE        RY      Yemen, Rep.           Yemen
## 232            ZA        ZA     South Africa    South Africa
## 233            CD        ZR Congo, Dem. Rep. Dem. Rep. Congo
## 234            ZM        ZM           Zambia          Zambia
## 235            ZW        ZW         Zimbabwe        Zimbabwe
```

```
dim(MergeData2)
```

```
## [1] 189  34
```

```
rownames(MergeData2) <- seq(length=nrow(MergeData2))
```

Export data frames into csv file in the data directory to keep track of edits and merges

```
write.csv(MergeData1, "MergeData1.csv")
write.csv(MergeData2, "MergeData2.csv")
write.csv(GDPRaw, "GDPRaw.csv")
write.csv(GDPData, "GDPData.csv")
write.csv(EducationRaw, "EducationRaw.csv")
write.csv(GDPData, "GDPData.csv")
```

==Analysis to answer questions==

Extract the number of rows from origianl merged raw data, Track the number of cumulative matched rows, number of N/A values in Rankings, Economies, GDP, Income groups, and cumulative unmatched rows

```
print(paste0("Total Number of Rows in Merged Data: ", nrow(MergeData1)))
```

```
## [1] "Total Number of Rows in Merged Data: 235"
```

```
Matches<-sum(is.na(MergeData1$`US Dollars (millions)`) == FALSE & is.na(MergeData1$Income.Group) == FALS
print(paste0("Number of Matched Rows: ", Matches))
```

```
## [1] "Number of Matched Rows: 189"
```

```
NARanking<-sum(is.na(MergeData1$Ranking) == TRUE)
print(paste0("Number of N/A Rankings: ", NARanking))
```

```
## [1] "Number of N/A Rankings: 45"
```

```
NAEconomy<-sum(is.na(MergeData1$Economy) == TRUE)
print(paste0("Number of N/A Economies: ", NAEconomy))
```

```
## [1] "Number of N/A Economies: 45"
```

```
NAGDP<-sum(is.na(MergeData1$`US Dollars (millions)`) == TRUE)
print(paste0("Number of N/A GDPs: ", NAGDP))
```

```
## [1] "Number of N/A GDPs: 45"
```

```
NAIncomeGroup<-sum(is.na(MergeData1$Income.Group) == TRUE)
print(paste0("Number of N/A Income Groups: ", NAIncomeGroup))
```

```
## [1] "Number of N/A Income Groups: 1"
```

```
NATotal<-sum(is.na(MergeData1$Ranking) == TRUE| is.na(MergeData1$Economy) == TRUE | MergeData1$`US Dolla
print(paste0("Total Number of Unmatched Rows: ", NATotal))
```

```
## [1] "Total Number of Unmatched Rows: 46"
```

```
print(paste0("Number of Rows in Merged Data without N/A values: ", nrow(MergeData2)))
```

```
## [1] "Number of Rows in Merged Data without N/A values: 189"
```

Question 0) Include code to count the number of missing values for each variable used in the analysis:

For each of the variables utilized in the analysis, there are 45 N/A values for rankings, economies, and GDP. There is one country with a N/A value for income groups and is also eliminated, which results in a total of 46 unmatched rows

Question 1) Merge the data based on the country shortcode. How many of the IDs match?

After merging the data by country shortcode and eliminating pertinent rows with N/A values, 189 countries has GDP, educational, and ranking values matching with income group.

Rank the merged data by ascending GDP ranking, examine attributes and export dataset

```
NegGDP <- MergeData2[order(MergeData2$`US Dollars (millions)`),]
head(NegGDP)
```

```
##     CountryCode Ranking               Economy US Dollars (millions)
## 173         TUV     190                Tuvalu                    40
## 92          KIR     189              Kiribati                   175
## 113         MHL     188      Marshall Islands                   182
## 137         PLW     187                 Palau                   228
## 155         STP     186 São Tomé and Principe                   263
## 59          FSM     185 Micronesia, Fed. Sts.                   326
##                                      Long.Name        Income.Group
## 173                                     Tuvalu Lower middle income
## 92                         Republic of Kiribati Lower middle income
## 113             Republic of the Marshall Islands Lower middle income
## 137                            Republic of Palau Upper middle income
## 155 Democratic Republic of São Tomé and Principe Lower middle income
## 59                Federated States of Micronesia Lower middle income
##                  Region Lending.category Other.groups
```

```
## 173 East Asia & Pacific
## 92  East Asia & Pacific              IDA
## 113 East Asia & Pacific             IBRD
## 137 East Asia & Pacific             IBRD
## 155  Sub-Saharan Africa              IDA         HIPC
## 59  East Asia & Pacific             IBRD
##                   Currency.Unit Latest.population.census
## 173           Australian dollar
## 92            Australian dollar                     2005
## 113                U.S. dollar                     1999
## 137                U.S. dollar                     2005
## 155 São Tomé and Principe dobra                     2001
## 59                 U.S. dollar                     2000
##     Latest.household.survey
## 173
## 92
## 113
## 137
## 155
## 59
##                                                                                Special.Notes
## 173
## 92  The government statistical office has revised national accounts data for 1970-2008.
## 113
## 137
## 155
## 59  The government statistical office has revised national accounts data for 1995-2008.
##     National.accounts.base.year National.accounts.reference.year
## 173                                                           NA
## 92                        1991                                NA
## 113                       1991                                NA
## 137                       1995                                NA
## 155                       2001                                NA
## 59                        1998                                NA
##     System.of.National.Accounts SNA.price.valuation
## 173                          NA
## 92                           NA                VAB
## 113                          NA                VAB
## 137                          NA                VAB
## 155                          NA                VAP
## 59                           NA                VAB
##     Alternative.conversion.factor PPP.survey.year
## 173                                           NA
## 92                                            NA
## 113                                           NA
## 137                                           NA
## 155                                         2005
## 59                                            NA
##     Balance.of.Payments.Manual.in.use External.debt.Reporting.status
## 173
## 92
## 113
## 137
## 155                                                      Preliminary
```

```
## 59
##     System.of.trade Government.Accounting.concept
## 173
## 92          General
## 113
## 137
## 155         Special
## 59
##     IMF.data.dissemination.standard
## 173
## 92                            GDDS
## 113
## 137
## 155                           GDDS
## 59
##     Source.of.most.recent.Income.and.expenditure.data
## 173
## 92
## 113
## 137
## 155                                        PS 2000-01
## 59
##     Vital.registration.complete Latest.agricultural.census
## 173
## 92
## 113
## 137                         Yes
## 155
## 59
##     Latest.industrial.data Latest.trade.data Latest.water.withdrawal.data
## 173                     NA                NA                           NA
## 92                      NA              2005                           NA
## 113                     NA                NA                           NA
## 137                     NA                NA                           NA
## 155                     NA              2008                           NA
## 59                      NA                NA                           NA
##     X2.alpha.code WB.2.code             Table.Name            Short.Name
## 173            TV        TV                 Tuvalu                Tuvalu
## 92             KI        KI               Kiribati              Kiribati
## 113            MH        MH      Marshall Islands      Marshall Islands
## 137            PW        PW                 Palau                 Palau
## 155            ST        ST São Tomé and Principe São Tomé and Principe
## 59             FM        FM Micronesia, Fed. Sts.            Micronesia
```

```
str(NegGDP)
```

```
## 'data.frame':    189 obs. of  34 variables:
##  $ CountryCode                                  : chr  "TUV" "KIR" "MHL" "PLW" ...
##  $ Ranking                                      : int  190 189 188 187 186 185 184 183 182 181 .
##  $ Economy                                      : chr  "Tuvalu" "Kiribati" "Marshall Islands" "Pa
##  $ US Dollars (millions)                        : num  40 175 182 228 263 326 472 480 596 684 ..
##  $ Long.Name                                    : chr  "Tuvalu" "Republic of Kiribati" "Republic
##  $ Income.Group                                 : chr  "Lower middle income" "Lower middle incom
##  $ Region                                       : chr  "East Asia & Pacific" "East Asia & Pacifi
##  $ Lending.category                             : chr  "" "IDA" "IBRD" "IBRD" ...
```

```
##  $ Other.groups                                       : chr  "" "" "" "" ...
##  $ Currency.Unit                                      : chr  "Australian dollar" "Australian dollar" "
##  $ Latest.population.census                           : chr  "" "2005" "1999" "2005" ...
##  $ Latest.household.survey                            : chr  "" "" "" "" ...
##  $ Special.Notes                                      : chr  "" "The government statistical office has
##  $ National.accounts.base.year                        : chr  "" "1991" "1991" "1995" ...
##  $ National.accounts.reference.year                   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ System.of.National.Accounts                        : int  NA NA NA NA NA NA NA 1993 NA NA ...
##  $ SNA.price.valuation                                : chr  "" "VAB" "VAB" "VAB" ...
##  $ Alternative.conversion.factor                      : chr  "" "" "" "" ...
##  $ PPP.survey.year                                    : int  NA NA NA NA 2005 NA NA NA 2005 NA ...
##  $ Balance.of.Payments.Manual.in.use                  : chr  "" "" "" "" ...
##  $ External.debt.Reporting.status                     : chr  "" "" "" "" ...
##  $ System.of.trade                                    : chr  "" "General" "" "" ...
##  $ Government.Accounting.concept                      : chr  "" "" "" "" ...
##  $ IMF.data.dissemination.standard                    : chr  "" "GDDS" "" "" ...
##  $ Source.of.most.recent.Income.and.expenditure.data: chr  "" "" "" "" ...
##  $ Vital.registration.complete                        : chr  "" "" "" "Yes" ...
##  $ Latest.agricultural.census                         : chr  "" "" "" "" ...
##  $ Latest.industrial.data                             : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Latest.trade.data                                  : int  NA 2005 NA NA 2008 NA 2007 2008 2007 2008
##  $ Latest.water.withdrawal.data                       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ X2.alpha.code                                      : chr  "TV" "KI" "MH" "PW" ...
##  $ WB.2.code                                          : chr  "TV" "KI" "MH" "PW" ...
##  $ Table.Name                                         : chr  "Tuvalu" "Kiribati" "Marshall Islands" "Pa
##  $ Short.Name                                         : chr  "Tuvalu" "Kiribati" "Marshall Islands" "Pa
```

```
write.csv(NegGDP, "NegGDP.csv")
```

Code to find 13th country with the ascending GDP. More code to show that there is a tie between St. Kitts and Grenada at 12th place, which results the alphabetical order to dictate St. Kitts at 13th place in ranking.

```
country13NegGDP<-NegGDP[13,3]
print(paste0("The 13th country in ascending order by GDP is: ", country13NegGDP))
```

```
## [1] "The 13th country in ascending order by GDP is: St. Kitts and Nevis"
```

```
NegGDP[12:13,c(2, 3, 4, 6)]
```

```
##    Ranking             Economy US Dollars (millions)        Income.Group
## 69     178             Grenada                   767 Upper middle income
## 93     178 St. Kitts and Nevis                   767 Upper middle income
```

Question 2) Sort the data frame in ascending order by GDP (so United States is last). What is the 13th country in the resulting data frame?

From ascending GDP, country #13 is St. Kitts and Nevis in the resulting data frame NegGDP. Technically, St. Kitts and Grenada are tied at 12th in ascending GDP and further ascending alphabetical sorting makes St. Kitts appear at 13th and Grenada at 12th place.

Assign a subset of High Income OECD countries from income group and find the mean of their GDP rank

```
HIOECD <- MergeData2[ which(MergeData2$Income.Group=='High income: OECD'), ]
head(HIOECD)
```

```
##    CountryCode Ranking        Economy US Dollars (millions)
## 9          AUS      12      Australia                1532408
## 10         AUT      27        Austria                 394708
```

```
## 13         BEL      25        Belgium                483262
## 31         CAN      11         Canada               1821424
## 32         CHE      20    Switzerland                631173
## 44         CZE      51 Czech Republic                196446
##                    Long.Name      Income.Group                Region
## 9  Commonwealth of Australia High income: OECD   East Asia & Pacific
## 10       Republic of Austria High income: OECD Europe & Central Asia
## 13        Kingdom of Belgium High income: OECD Europe & Central Asia
## 31                    Canada High income: OECD         North America
## 32               Switzerland High income: OECD Europe & Central Asia
## 44            Czech Republic High income: OECD Europe & Central Asia
##    Lending.category Other.groups     Currency.Unit
## 9                                 Australian dollar
## 10                    Euro area              Euro
## 13                    Euro area              Euro
## 31                                 Canadian dollar
## 32                                     Swiss franc
## 44                                    Czech koruna
##    Latest.population.census Latest.household.survey
## 9                      2006
## 10                     2001
## 13                     2001
## 31                     2006
## 32                     2000
## 44                     2001              RHS, 1993
##
## 9
## 10 A simple multiplier is used to convert the national currencies of EMU members to euros. The follo
## 13      A simple multiplier is used to convert the national currencies of EMU members to euros. The
## 31
## 32
## 44
##    National.accounts.base.year National.accounts.reference.year
## 9                                                          2007
## 10                        2000                               NA
## 13                        2000                               NA
## 31                        2000                               NA
## 32                        2000                               NA
## 44                        2000                             1995
##    System.of.National.Accounts SNA.price.valuation
## 9                         1993                VAB
## 10                        1993                VAB
## 13                        1993                VAB
## 31                        1993                VAB
## 32                          NA                VAB
## 44                        1993                VAB
##    Alternative.conversion.factor PPP.survey.year
## 9                                           2005
## 10                                          2005
## 13                                          2005
## 31                                          2005
## 32                                          2005
## 44                                          2005
##    Balance.of.Payments.Manual.in.use External.debt.Reporting.status
```

```
## 9                              BPM5
## 10                             BPM5
## 13                             BPM5
## 31                             BPM5
## 32                             BPM5
## 44                             BPM5
##    System.of.trade Government.Accounting.concept
## 9          General                  Consolidated
## 10         Special                  Consolidated
## 13         Special                  Consolidated
## 31         General                  Consolidated
## 32         Special                  Consolidated
## 44         General                  Consolidated
##    IMF.data.dissemination.standard
## 9                             SDDS
## 10                            SDDS
## 13                            SDDS
## 31                            SDDS
## 32                            SDDS
## 44                            SDDS
##    Source.of.most.recent.Income.and.expenditure.data
## 9                                         ES/BS, 1994
## 10                                            IS 2000
## 13                                          IHS, 2000
## 31                                          LFS, 2000
## 32                                        ES/BS, 2000
## 44                                            IS 1996
##    Vital.registration.complete     Latest.agricultural.census
## 9                          Yes                           2001
## 10                         Yes                      1999-2000
## 13                         Yes 1999-2000 (conducted annually)
## 31                         Yes                      1996/2001
## 32                         Yes                           2000
## 44                         Yes                           2000
##    Latest.industrial.data Latest.trade.data Latest.water.withdrawal.data
## 9                    2004              2008                         2000
## 10                   2004              2008                         2000
## 13                   2004              2008                           NA
## 31                   2001              2008                         2000
## 32                     NA              2008                         2000
## 44                   2005              2008                         2000
##    X2.alpha.code WB.2.code     Table.Name     Short.Name
## 9             AU        AU      Australia      Australia
## 10            AT        AT        Austria        Austria
## 13            BE        BE        Belgium        Belgium
## 31            CA        CA         Canada         Canada
## 32            CH        CH    Switzerland    Switzerland
## 44            CZ        CZ Czech Republic Czech Republic
```

```
write.csv(HIOECD, "HIOECD.csv")
HAvgGDPRank<- mean(HIOECD$Ranking)
print(paste0("The average GDP ranking of high income, OECD countries is: ", round(HAvgGDPRank, digits =
```

```
## [1] "The average GDP ranking of high income, OECD countries is: 32.97"
```

Assign a subset of High Income nonOECD countries from income group and find the mean of their GDP rank

```
HINonOECD <- MergeData2[ which(MergeData2$Income.Group=='High income: nonOECD'), ]
head(HINonOECD)
```

```
##    CountryCode Ranking              Economy US Dollars (millions)
## 1          ABW     161                Aruba                  2584
## 5          ARE      32 United Arab Emirates                348595
## 18         BHR      93              Bahrain                 29044
## 19         BHS     138         Bahamas, The                  8149
## 23         BMU     149              Bermuda                  5474
## 26         BRB     153             Barbados                  4225
##                       Long.Name         Income.Group
## 1                         Aruba High income: nonOECD
## 5          United Arab Emirates High income: nonOECD
## 18           Kingdom of Bahrain High income: nonOECD
## 19 Commonwealth of The Bahamas High income: nonOECD
## 23                 The Bermudas High income: nonOECD
## 26                     Barbados High income: nonOECD
##                        Region Lending.category Other.groups
## 1   Latin America & Caribbean
## 5  Middle East & North Africa
## 18 Middle East & North Africa
## 19  Latin America & Caribbean
## 23              North America
## 26  Latin America & Caribbean
##      Currency.Unit Latest.population.census Latest.household.survey
## 1    Aruban florin                    2000
## 5    U.A.E. dirham                    2005
## 18  Bahraini dinar                    2001
## 19 Bahamian dollar                    2000
## 23  Bermuda dollar                    2000
## 26 Barbados dollar                    2000
##                                                                                Special.Notes
## 1
## 5
## 18
## 19 The government has revised national accounts data for 1997-2007. The new base year is 2006.
## 23                    The Statistical Office has revised national accounts data for 1996-2007.
## 26
##    National.accounts.base.year National.accounts.reference.year
## 1                         1995                               NA
## 5                         1995                               NA
## 18                        1985                               NA
## 19                        2006                               NA
## 23                        1996                               NA
## 26                        1974                               NA
##    System.of.National.Accounts SNA.price.valuation
## 1                           NA
## 5                           NA                VAB
## 18                          NA                VAP
## 19                        1993                VAB
## 23                          NA                VAB
## 26                          NA                VAB
##    Alternative.conversion.factor PPP.survey.year
```

```
## 1                                           NA
## 5                                           NA
## 18                                        2005
## 19                                          NA
## 23                                          NA
## 26                                          NA
##    Balance.of.Payments.Manual.in.use External.debt.Reporting.status
## 1
## 5                               BPM4
## 18                              BPM5
## 19                              BPM5
## 23
## 26                              BPM5
##    System.of.trade Government.Accounting.concept
## 1          Special
## 5          General                  Consolidated
## 18         General                  Consolidated
## 19         General                     Budgetary
## 23
## 26         General                  Consolidated
##    IMF.data.dissemination.standard
## 1
## 5                             GDDS
## 18                            GDDS
## 19                            GDDS
## 23
## 26                            GDDS
##    Source.of.most.recent.Income.and.expenditure.data
## 1
## 5
## 18
## 19
## 23
## 26
##    Vital.registration.complete Latest.agricultural.census
## 1
## 5                                                   1998
## 18                        Yes
## 19
## 23                        Yes
## 26                        Yes
##    Latest.industrial.data Latest.trade.data Latest.water.withdrawal.data
## 1                      NA              2008                           NA
## 5                      NA              2008                         2005
## 18                     NA              2007                         2003
## 19                   1997              2008                           NA
## 23                     NA              2008                           NA
## 26                     NA              2008                         2000
##    X2.alpha.code WB.2.code           Table.Name           Short.Name
## 1             AW        AW                Aruba                Aruba
## 5             AE        AE United Arab Emirates United Arab Emirates
## 18            BH        BH              Bahrain              Bahrain
## 19            BS        BS         Bahamas, The          The Bahamas
## 23            BM        BM              Bermuda              Bermuda
```

```
## 26            BB        BB             Barbados             Barbados
```

```
write.csv(HINonOECD, "HINonOECD.csv")
NAvgGDPRank<- mean(HINonOECD$Ranking)
print(paste0("The average GDP ranking of high income, nonOECD countries is: ", round(NAvgGDPRank, digits
```

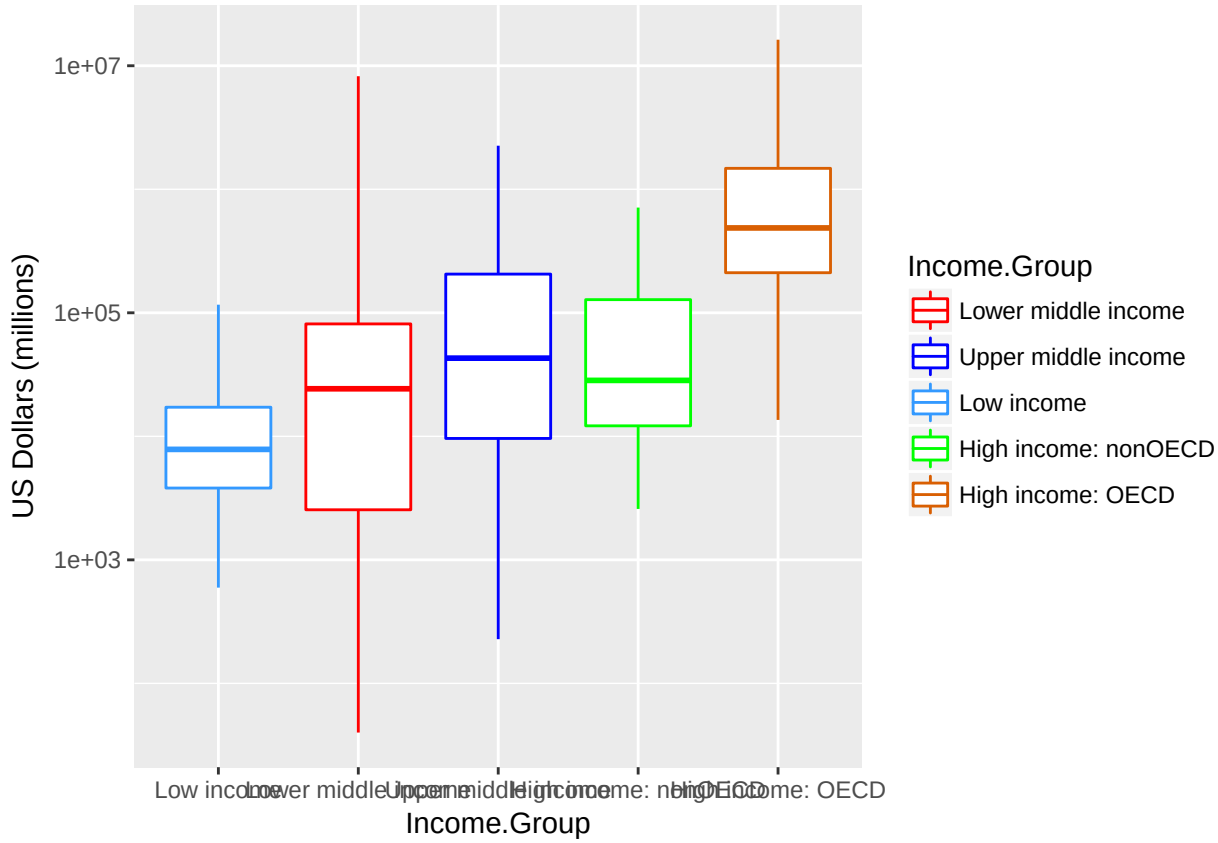
```
## [1] "The average GDP ranking of high income, nonOECD countries is: 91.91"
```

Question 3) What are the average GDP rankings for the "High income: OECD" and "High income: nonOECD" groups?

By income group, the average GDP rankings for High income: OECD countries is 32.97 and for High income: nonOECD countries is 91.91. High income OECD countries have higher GDP than that of High income nonOECD countries. Higher GDP ranking suggests that high income countries that are open to free world trade and development are more prosperous.

Using ggplot2, individual countries with matching rows in GDP are logarithmically plotted by separately colored income group box plots to show quantile distribution.

```
library(ggplot2)
NegGDP$Income.Group <- factor(NegGDP$Income.Group, levels=c("Low income", "Lower middle income", "Upper
color.codes<-as.character(c("#3399FF", "#FF0000","#0000FF", "#00FF00", "#D95F02"))
ggplot(data = NegGDP, aes(y = `US Dollars (millions)`, x = Income.Group, colour = Income.Group))+
   geom_boxplot() + scale_y_log10() +
  scale_colour_manual(breaks = NegGDP$Income.Group,values = unique(as.character(color.codes)))
```



Question 4) Show the distribution of GDP value for all the countries and color plots by income group. Use ggplot2 to create your plot.

Graphically by boxplot log distribution, it was expected that the median GDP of countries grouped by income

group rose from low income to lower middle income to upper middle income. For high income countries, there is a discrepancy between OECD and nonOECD countries. If the country is high income but does not allow free global trade and development, they have an median GDP lower than that of upper middle income countries and about equivalent to that of lower middle income countries. Otherwise, if the high income country is an OECD member, they countinue the trend of GDP prosperity. In addition, the giant range of counties that fall into the lower middle income category suggests that the distinction of countries by income groups is not soley based on GDP qualities alone.

GDP summary statistics of countries based on income groups

```
tapply(NegGDP$`US Dollars (millions)`, NegGDP$Income.Group, summary)
```

```
## $`Low income`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     596    3814    7843   14410   17200  116400
##
## $`Lower middle income`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      40    2549   24270  256700   81450 8227000
##
## $`Upper middle income`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     228    9613   42940  231800  205800 2253000
##
## $`High income: nonOECD`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2584   12840   28370  104300  131200  711000
##
## $`High income: OECD`
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##    13580   211100   486500  1484000  1480000 16240000
```

Question 5) Provide summary statistics of GDP by income groups.

From the boxplot log distribution of countries' GDP separated by income groups, the quantile distributions were plotted by range, interquantile range, and medians. The summary statistics show that the mean GDP per income group is very different than that of the median, with the mean being 0.8x, 9.5x, 4.3x, 2.6x, and 2x greater than that of the median for their respective income groups by ascending classification.

The range overlap in country GDP further suggests that countries separated by income group was not soley based on GDP. The order of mean GDP by income group is low income, high income: nonOECD, upper middle income, lower middle income, and high income: OECD, which means that there are more factors that dictate how a country is classifed into income groups than GDP alone.

Breaks the GDP rankings into 5 separate quantile groups, with increment of 20%, and writes the quantiles into csv. Negdata is used because factors and levels are defined previously

```
Quantiles<-cut(NegGDP$Ranking, breaks=quantile(NegGDP$Ranking,seq(0, 1, 0.2)))
head(Quantiles)
```

```
## [1] (152,190] (152,190] (152,190] (152,190] (152,190] (152,190]
## Levels: (1,38.6] (38.6,76.2] (76.2,114] (114,152] (152,190]
```

```
write.csv(Quantiles, "Quantiles.csv")
```

Using reshape2, a table shows the number of contries per income group that falls inside their respective 20% quantile groups based on individual GDP ranking

```
library(reshape2)
table(NegGDP$Income.Group, Quantiles)
```

```
##                       Quantiles
##                        (1,38.6] (38.6,76.2] (76.2,114] (114,152] (152,190]
##   Low income                  0           1          9        16        11
##   Lower middle income         5          13         11         9        16
##   Upper middle income        11           9          8         8         9
##   High income: nonOECD        4           5          8         4         2
##   High income: OECD          17          10          1         1         0
```

Question 6a) Cut the GDP ranking into 5 separate quantile groups. Make a table versus Income.Group.

It was expected that low income has more countries that fall inside the higher quantile GDP rankings while that of the higher income: OECD has more that fall inside the lower GDP quantile rankings. Lower middle income has a concentration of counties that fall inside ther higher GDP rankings with some of its countries in the lower quantile GDP rankings. Upper middle income countries has an even distribution of countries in each quantile category while that of high income: nonOECD countries have countries falling in the middle GDP quantile rankings.

Create a subset featuring GDP ranking, country, and Income group and examine attributes

```
IncomeSub <- MergeData2[c(2,3,6)]
IncomeSub<-IncomeSub[order(IncomeSub$Ranking),]
rownames(IncomeSub) <- seq(length=nrow(IncomeSub))
head(IncomeSub)
```

```
##   Ranking        Economy        Income.Group
## 1       1  United States   High income: OECD
## 2       2          China Lower middle income
## 3       3          Japan   High income: OECD
## 4       4        Germany   High income: OECD
## 5       5         France   High income: OECD
## 6       6 United Kingdom   High income: OECD
```

```
write.csv(IncomeSub, "IncomeSub.csv")
```

Find the subset and number of countries that are lower middle in income group and top 38 in GDP ranking. Export new dataset as LowerMiddleTop38.

```
LowerMiddleTop38 <- IncomeSub[which(IncomeSub$Ranking <= 38 & IncomeSub$Income.Group == "Lower middle in
rownames(LowerMiddleTop38) <- seq(length=nrow(LowerMiddleTop38))
LowerMiddleTop38
```

```
##   Ranking          Economy        Income.Group
## 1       2            China Lower middle income
## 2      10            India Lower middle income
## 3      16        Indonesia Lower middle income
## 4      31         Thailand Lower middle income
## 5      38 Egypt, Arab Rep. Lower middle income
```

```
write.csv(LowerMiddleTop38, "LowerMiddleTop38.csv")
print(paste0("The number of lower income countries that are in the top 38 for GDP ranking is: ", nrow(L
```

```
## [1] "The number of lower income countries that are in the top 38 for GDP ranking is: 5"
```

Question 6b) How many countries are Lower middle income but among the 38 nations with highest GDP?

The table shows there are 5 lower income countries that are in the top 38 for GDP ranking.

Conclusion: