

# Comparison of Social Factors Versus per capita GDP

Sophie Smith

Saratoga Springs High School

### Acknowledgement

I would like to thank my Science Research teachers, Ms. Lohnes, Mr. Robinson and Ms. Perkins, for their assistance and encouragement with my project and during class, my parents, for their support, as well as my mentor, Dr. Necip Doganaksoy, for his enthusiasm for my project and his guidance during my research.

### Abstract

Gross Domestic Product (GDP) is a measure of the final domestic goods produced over one year in a given country. GDP reflects the production capabilities of a country, often signifying their economic sophistication. Economic prosperity is believably associated with social progress, stimulating the analysis of GDP versus social factors, as seen in this study. The variables compared to per capita GDP throughout this study were life expectancy, the percentage of females in the labor force, the percentage of women in the national parliament, internet users, mobile cellular subscriptions, high-tech exports, electric power consumption, access to improved water sources and primary education enrollment. Using accessible data from 2003, 2008, and 2013, the associations between these variables and GDP along with the variables which demonstrate the most differentiation among these country clusters were analyzed. Using the techniques of clustering and partitioning, the groupings of similar countries remained fairly consistent over the years. Throughout this analysis, GDP was consistently associated with internet users and mobile cellular subscriptions for the selected countries.

## Table of Contents

Abstract.....	2
Acknowledgement.....	2
Introduction.....	5
Statement of Purpose.....	7
Hypothesis.....	7
Methods and Materials.....	7
Results.....	10
Discussion and Conclusion.....	19
Appendix.....	21
References.....	24

## Lists of Figures and Tables

Figure 1: Variables.....	6
Figure 2: Data File.....	8
Figure 3: 2013 Matrix Scatter Plot.....	10

Figure 4: 2008 Matrix Scatter Plot.....	10
Figure 5: 2003 Matrix Scatter Plot.....	11
Figure 6: Cluster Number Chart.....	11
Figure 7: Plot of Cluster Groups.....	12
Figure 8: 2013 Cluster Map.....	13
Figure 9: 2013 Loess Scatter Plot.....	14
Figure 10: 2008 Loess Scatter Plot.....	14
Figure 11: 2003 Loess Scatter Plot.....	15
Figure 12: 2013 Partitioning Tree.....	16
Figure 13: 2013 GDP Prediction.....	16
Figure 14: 2008 Partitioning Tree.....	17
Figure 15: 2008 GDP Prediction.....	17
Figure 16: 2003 Partitioning Tree.....	18
Figure 17: 2003 GDP Prediction.....	18
Figure A: 2013 Summary Statistic Script.....	21
Figure B: 2013 Cluster Script.....	22
Figure C: 2013 Partitioning Script.....	22
Figure D: 2013 Map Cluster Script.....	23

Figure E: 2013 Cluster Map Interpretation.....	23
--	----

## Introduction

Gross Domestic Product (GDP) represents the total dollar value of the goods and services produced over a specific time period. For this study, it was one year. GDP displays the production measures of an economy through accounting for the final produced goods by a nation. While displaying a fairly accurate representation of the wealth and economic sophistication of a nation, GDP fails to measure the well-being of citizens in a country. This occurs because GDP minimizes the weighted importance of leisure activity expenses in its measurement, instead solely displaying the economic performance of these nations. To compensate for the limitations of GDP, the United Nations produced the Human Development Index (HDI) in order to accurately represent the development of countries in retrospect to each other. The main purpose of the HDI is to “emphasize that expanding human choices should be the ultimate criteria for assessing development results.” This is accomplished through the development of this index, based on variables including life expectancy, expected years of schooling, gross national income, homicide rate, mobile phone subscriptions, and carbon dioxide emissions [6]. Another index, the Gender Development Index (GNI), was developed by the United Nations to gain a more representative understanding of gender gaps in human development achievements through an analysis of variables regarding health, knowledge and living standards [3]. With these various altered methods of demonstrating social progress throughout a nation, along with claims to the limited accuracy of GDP relation to societal development, through this study a specific index was not used, however, the associations between the selected variables and the per capita GDP of a selected population of countries were

compared to determine any relationship between social progress and GDP for these countries three distinct years.

The variables compared to GDP within this study were divided into four main categories: infrastructure, technology, education and social development. See Figure 1 for an overview of these categories and the variables associated with each category.

<b>INFRASTRUCTURE</b> - Percentage of the population with access to improved water source - Electric power consumption (kWh per Capita)	<b>SOCIAL DEVELOPMENT</b> - Life expectancy - Percentage of females in the labor force - Percentage of females in the National Parliament
<b>TECHNOLOGY</b> - Internet users (per 100 people) - Mobile cellular subscriptions (per 100 people) - High-technology exports (% of total exports)	<b>EDUCATION</b> - Percentage of individuals enrolled in primary education

**Figure 1** The variables used in this study

Life expectancy was included to serve as a standard factor, enhancing clustering and comparisons regarding GDP. The social variables compared against per capita GDP were the percentage of females in the labor force and the percentage of women participating in the country's national parliament. With the evident gender disparities across the globe, these variables were selected to account for adjustments in the participation of women in these sectors and compare those deviations to per capita GDP. The technological variables included internet users (per 100 people), mobile cellular subscriptions (per 100 people), and high-technology exports. Internet users represented the portion of the population who have access to the internet at their homes while mobile cellular subscriptions refer to "the number of subscriptions to a public mobile telephone service." [10] High-technology exports more closely resemble the production tendencies of a country as they are "products which require significant resources and research in order to develop and produce." [9] These products consist of developments in the computer, aerospace, pharmaceutical industries, along with others representing similar natures.

As for infrastructure, this section included the percentage of the population of a country with access to an improved water source, quantified as a source of water protected from outside contamination, and also electric power consumption (kWh per capita). [7] Lastly, education was accounted for with the incorporation of the percentage of individuals, in the appropriate age category, enrolled in primary education. Primary education refers to the first years of compulsory education, the term being often synonymous with elementary schooling. [2]

### Statement of Purpose

This study will describe the methods and results in determining associations between the previously discussed variables and per capita GDP in 2003, 2008, and 2013. The groupings of similar countries by these variables were discussed, along with the accuracy of predicting per capita GDP by means of partitioning, and the comparison of the distinct variables in differentiating per capita GDP across countries over the selected years.

### Hypothesis

Based on previous literature read, it was predicted that as the values for social development variables increased, so would the per capita GDP of the corresponding countries. Also, primary education enrollment was predicted to be most closely related to per capita GDP.

### Methods and Materials

To obtain the variables listed above, the open source data published by the World Bank was used. The World Bank is a coalition consisting of the International Bank for Reconstruction and Development, the International Development Association, the International Finance Corporation, the Multilateral Investment Guarantee Agency, and the International Centre for Settlement of Investment Disputes. The World Bank focuses on raising awareness about

inequality and ending poverty through bringing awareness to issues. Their website (data.worldbank.org) lists indicators or variables categorized in subsections, e.g. Agriculture and Rural Development, Climate Change, Economy and Growth, Education, Environment, Gender, Health, Infrastructure, Poverty, Science and Technology, Social Development, and other related sections. The variables were selected based on accessibility of data for many countries across a wide span of years while choosing a representative selection of indicators. The three years used (2003, 2008, and 2013) were selected because there is a 5-year spread between the years and because of their availability of data. The farther back in time for which data were sought, the fewer variable values available for each country due to less documentation. From there, the data were exported from The World Bank as .csv files and compiled them in an Excel document (Figure 2). To narrow the quantity of countries for the analysis, countries with missing values were deleted and non-country entities were eliminated. See the Appendix for a list of all the countries included within the study.

1		Per Capita G	Life Expect	Females in Labor f	Internet Us	Mobile Cellular Sub	High Tech Expo	Women in Natic	Electric Power	Improved Water	Primary Educ
2	Albania	4412.34558	77.640463	41.17663517	57.2	116.1572081	0.513330673	17.9	2531.888539	95.1	95.54573822
3	United Ar	42831.0891	77.195634	13.00612812	88	171.8737993	3.177578231	17.5	10904.45132	100	87.55709839
4	Australia	67652.6832	82.14878	45.38381602	83	106.8434974	12.90999462	26	10133.85918	100	97.35562897
5	Azerbaijan	7811.62142	70.690585	48.88199057	58.7	107.6134072	13.42318223	16	2092.536521	75.3	93.5072937
6	Belgium	46622.468	80.587805	45.73660163	82.1702	110.9031799	11.48752156	38	7966.686748	100	98.87512207
7	Bulgaria	7656.63851	74.860976	46.572502	53.0615	145.1876887	7.960635886	24.6	4639.706323	99	95.17063904
8	Bosnia and	4748.04069	76.276927	38.01695034	57.79	91.09530785	2.301306161	21.4	3219.012364	99.7	97.87940216
9	Belarus	7722.12335	72.470732	48.68706275	54.17	118.786176	4.399741122	26.6	3648.320304	99.1	92.84977722
10	Brazil	12071.778	74.122439	43.70370815	51.04	135.3050481	9.63455019	8.6	2529.30371	86.2	90.35398102

**Figure 2** Screenshot of Excel file displaying counties versus values per country for selected variables

The main goals for the data analysis were to determine the existence of associations between per capita GDP and the selected variables, along with determining the similarities of countries based on these variables, the most prominent factors in developing distinct divisions among the countries through partitioning, and the prediction of per capita GDP versus actual value of GDP based on the partitioning results. To analyze this data, R, a language for statistical



computing, was used (<https://cran.r-project.org/>). The main uses of R consisted of three subsections: summary statistics and exploratory analysis, clustering, and partitioning (the scripts seen in the Appendix). Figure A, within the appendix, was developed to set the working directory in R to the folder consisting of the compiled spreadsheet to access and read the file and produce summary statistics, i.e. mean, median, minimum, maximum, and quartiles, and a matrix scatter plot comparing both the independent variables to per capita GDP, and to each other. From there the existence of confounding variables could be determined from strong slopes in the independent variable scatter plots.

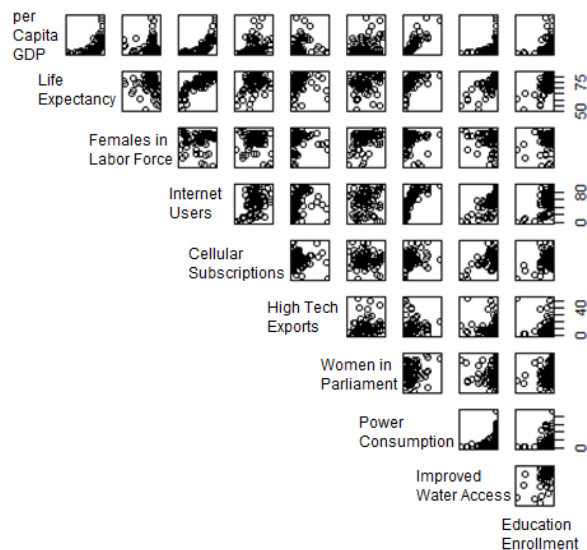
The next portion of the analysis consisted of using clustering techniques to group the countries based upon similarities in their independent variables (seen in Figure B). K-Means Clustering functions by dividing the observations into k groups, dependent on minimizing the distance between the observed values and the cluster center. With clustering, there is no dependent variable, as the set of objects from the study is grouped based on similarities amongst their independent variables. An example of this is seen with houses, commonly grouped together based on area, bedrooms, bathrooms, and other independent features.

Next, partitioning was used to divide the dataset into mutually exclusive groupings based on the input values. Partitioning uses the dependent variable and organizes the input samples into groups based on separate groups of similar outputs. The process breaks down the output into a number of “branches,” each regarding a separate variable. The output of this process is the production of a Classification and Regression Tree from which the inputs which most distinctly separate the observations into groups will be shown. The loess portion of the following code works to develop separate scatter plots for each input versus per capita GDP on which cluster averages were plotted. Loess is a method to obtain a connective line between an outcome and a

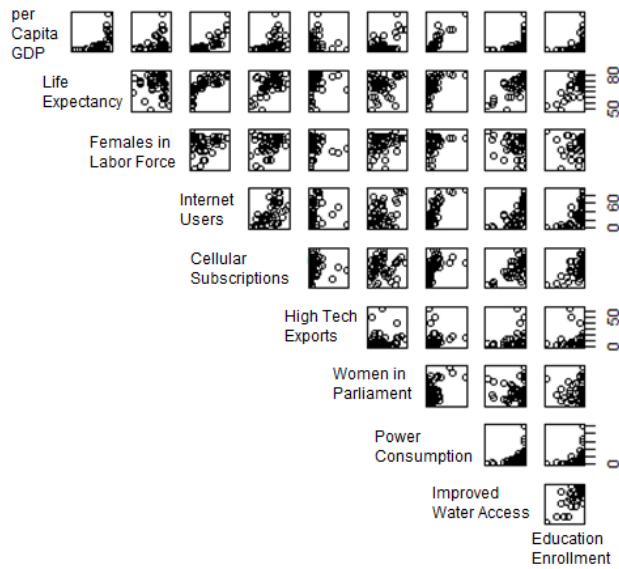
distinct number of predictor points, ultimately developing a prediction model (seen in Figures C and D).

## Results

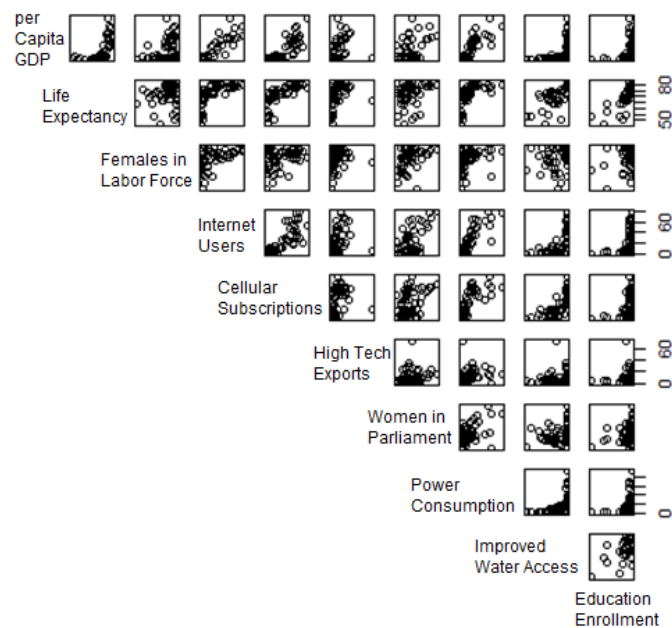
The initial step of analysis was the production of summary statistics for each year. Using the data compiled in Excel, the summary script specified above was run on it for each year (2003, 2008, and 2013). The data values of summary statistics are used for the comparison of data set values, and the matrix scatter plot allows for a greater understanding of the relationship between input and output variables.



**Figure 3** Matrix scatter plot for 2013

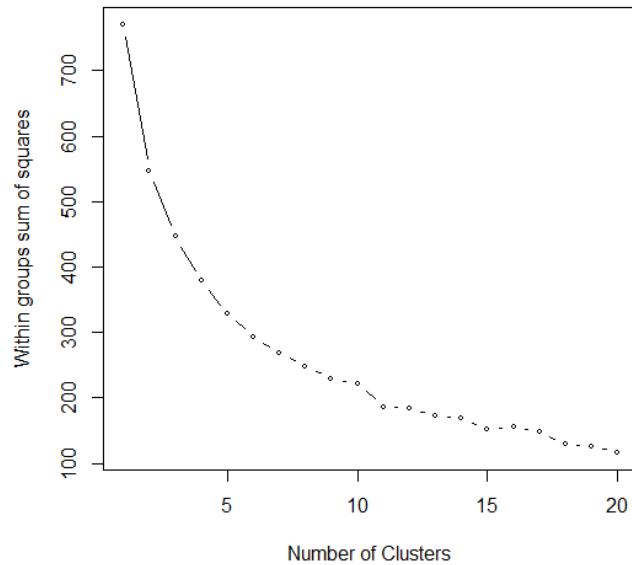


**Figure 4** Matrix scatter plot for 2008



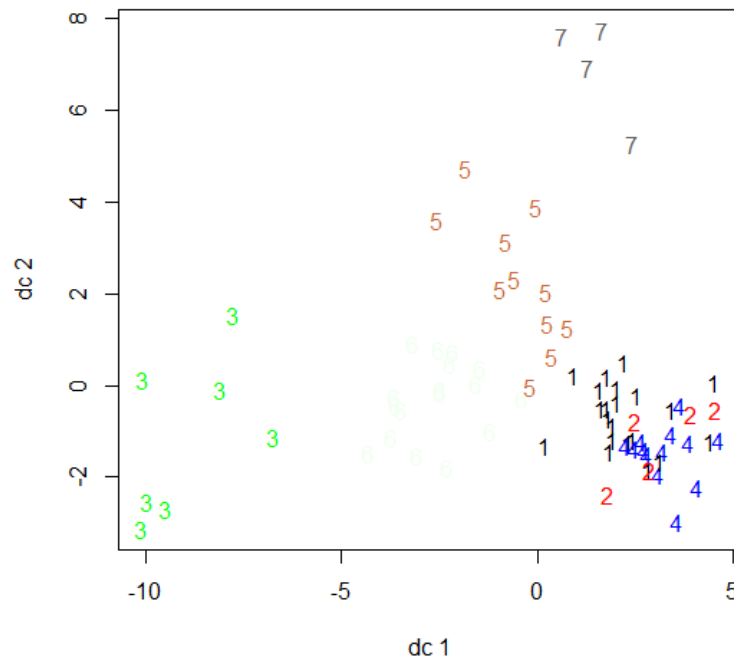
**Figure 5** Matrix scatter plot for 2003

The script for this section (Figure B) inputs the file for analysis, sets the ideal number of clusters for use, separates the data into this number of clusters and organizes those clusters, also producing diagrams relating to them. The following diagram was used to determine the ideal number of clusters.



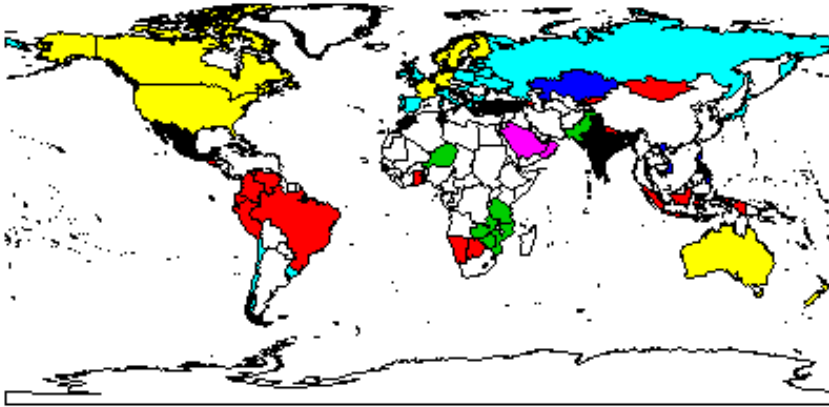
**Figure 6** Determination of ideal number of clusters

The ideal number of clusters is seen where cluster variability decreases sharply, approximately 7 for this example from 2013 (see Figure 6). Seven was experimentally determined as appropriate through developing the clusters and examining the number of countries within each cluster, i.e. to make sure all clusters contain greater than one country and to analyze the spread of the plots, in Figure 7, ensuring distance from separate cluster data values. From this point, 7 was concluded as an appropriate number of clusters for 2013 and used this parameter for the analysis of 2003 and 2008 to ensure minimal variability amongst the different years.



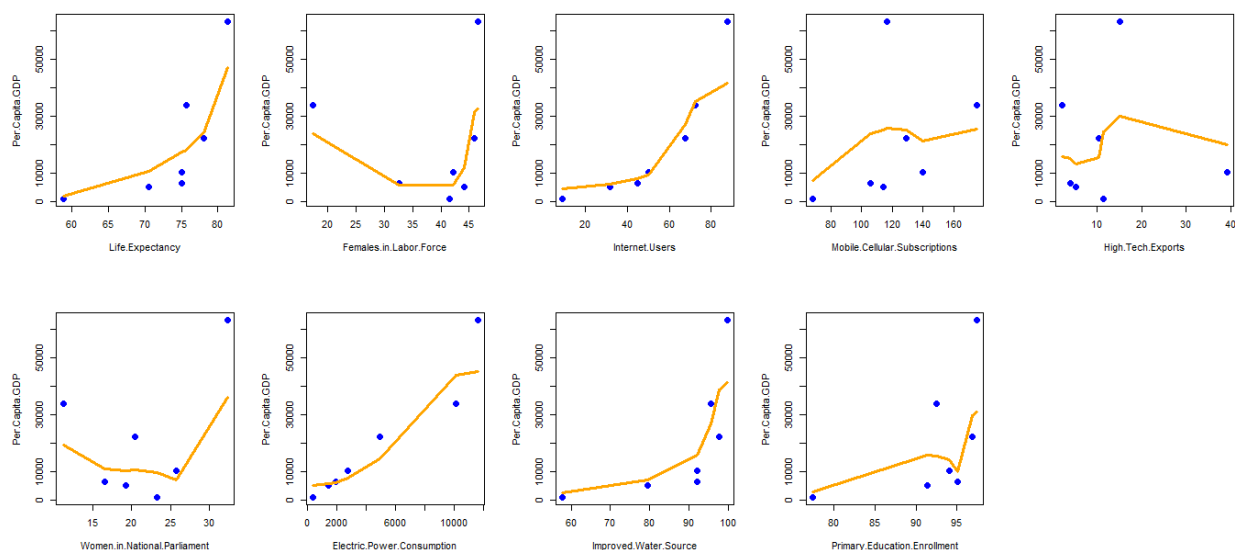
**Figure 7** Plot of variability and grouping of clusters for 2013

To summarize the clustering results over the years, a plot of the world map highlighted by color with each color representing the clusters developed by the code was developed. Yellow represents the countries with the greatest per capita GDP and a tendency of the highest social development. The progression of colors was yellow, purple, light blue, dark blue, green, red to black, where red represented the countries with the lowest societal development and per capita GDP. As displayed in this graph, the cluster groupings stayed relatively similar across the years, with the main adjustment resulting from the change in number of countries used due to the lack of variable data for certain entities. See the Appendix for an in-depth explanation of 2013 country clusters.

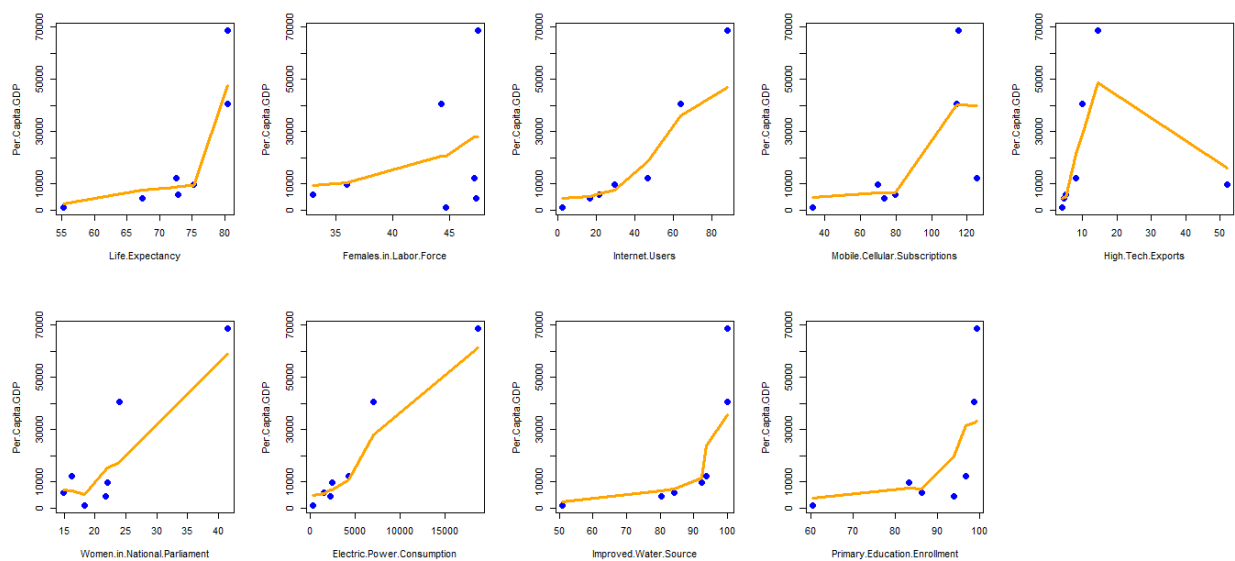


**Figure 8** Cluster map for 2013

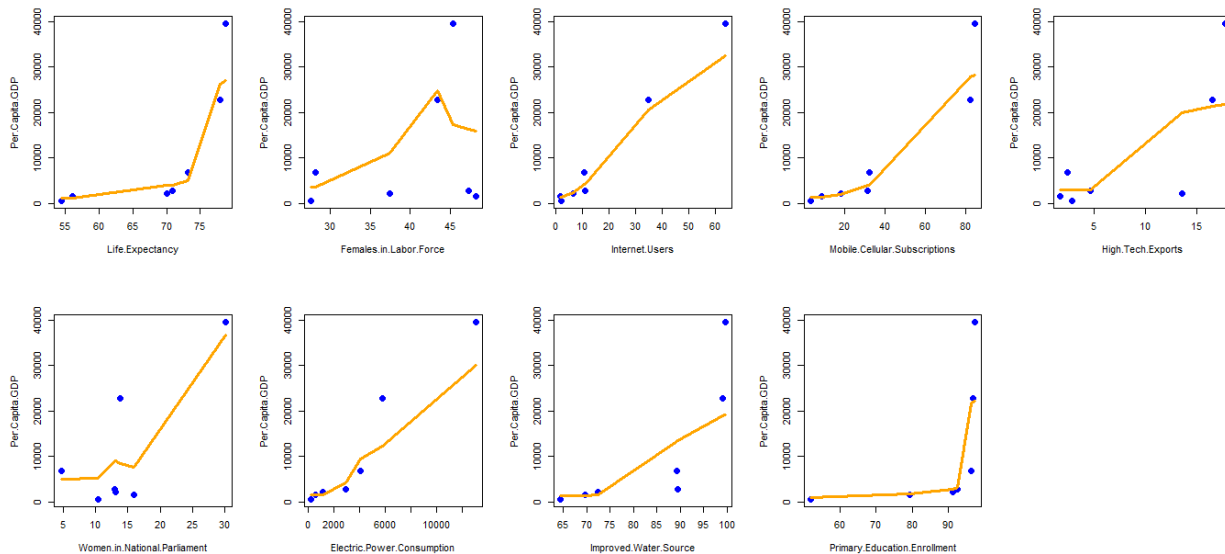
The next part of the research focused on the association between the independent variables and per capita GDP. In order to gain insight into this, the cluster averages of the variables were utilized. The results are shown in Figures 9, 10 and 11 for 2003, 2008, and 2013, respectively. This step was accomplished using loess smooth curve fitting techniques. This method works by plotting cluster averages for each individual independent variable and comparing the cluster averages, developing a predictive line as a result. The results, inserted below, demonstrate the cluster average values in comparison to one another, for each variable, and relative patterns in cluster values. The clusters and overall performance of the indicators remained relatively constant over the years, however, one major adjustment in the graphs was seen with women in the labor force. The year 2013 appeared to have a drop in the female participants in the labor force, whereas, 2008 and 2003 either demonstrated an increase in value or a relatively constant slope.



**Figure 9** Loess scatter plots for 2013



**Figure 10** Loess scatter plots for 2008



**Figure 11** Loess scatter plots for 2003

Partitioning analysis was used to study the relationship between the independent variables and per capita GDP. The output of partitioning was a classification tree, demonstrating which variables are most effective at dividing the countries into groups. A scatter plot comparing predicted per capita GDP versus actual per capita GDP was also obtained. These plots can be used to compare variable values over the years. The black lines represent confidence interval meaning that 95% of the time the average per capita GDP for the entire population was between those intervals. The plots for all three years had  $R^2$  values between 0.83 and 0.87 reflecting that the variation in per capita GDP was accounted for between 83% and 87% of the time.



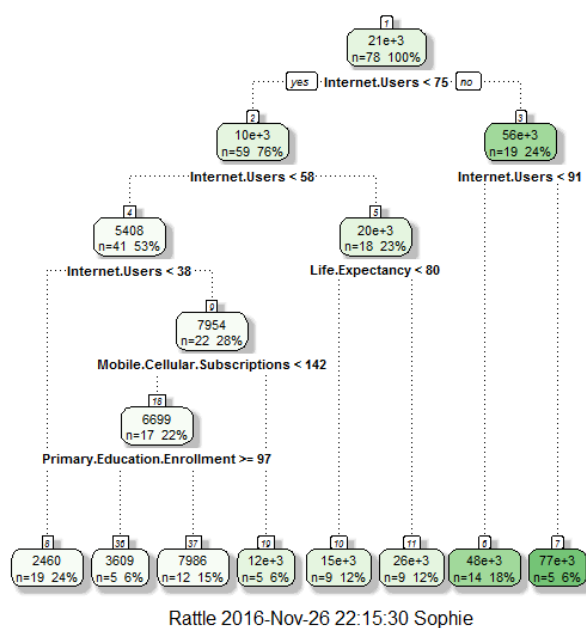


Figure 12 Partitioning tree for 2013

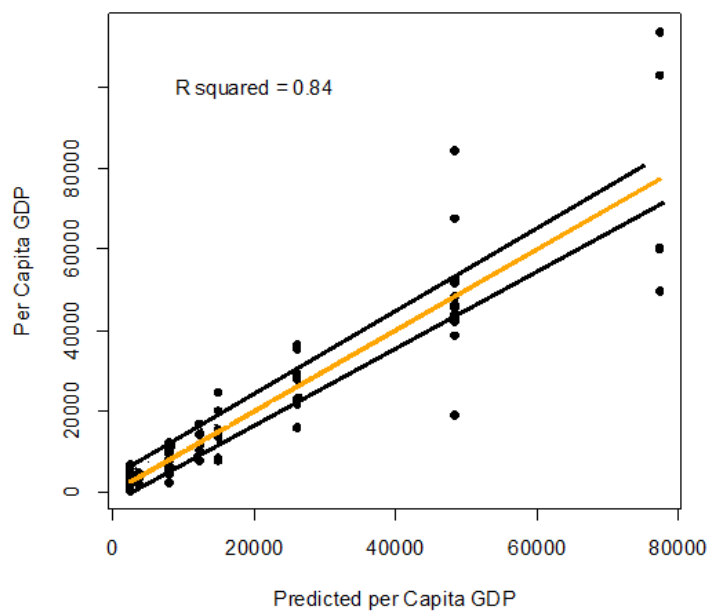


Figure 13 Predicted per capita GDP vs. Actual for 2013

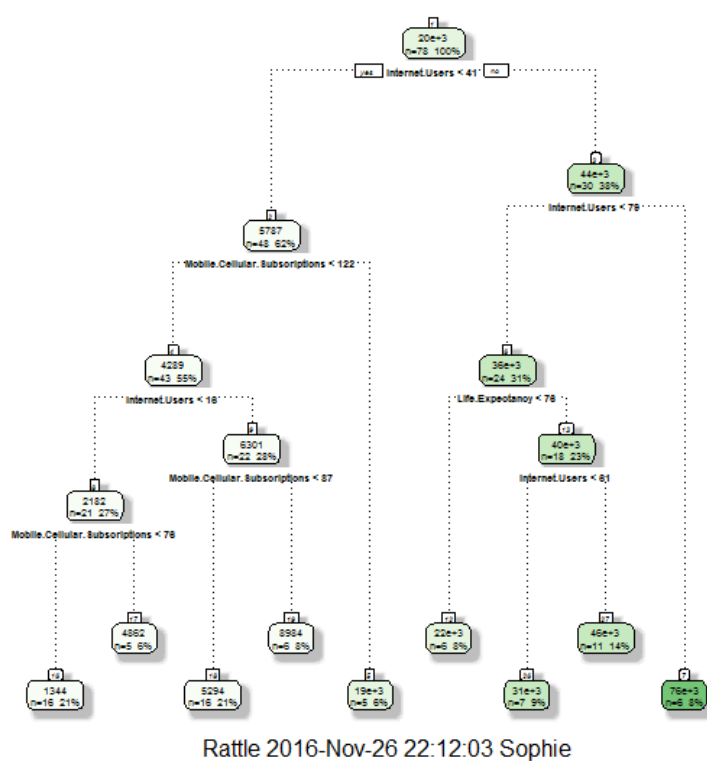


Figure 14 Partitioning tree for 2008

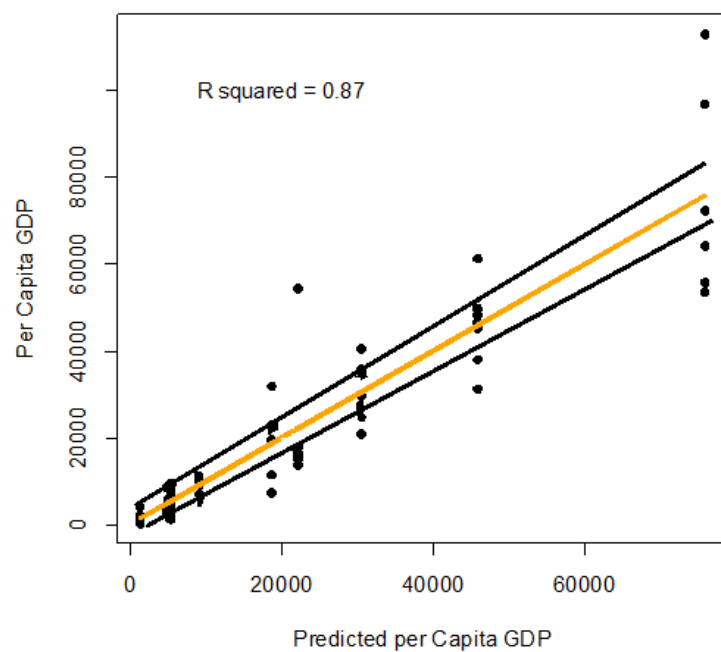


Figure 15 Predicted per capita GDP vs. Actual for 2008

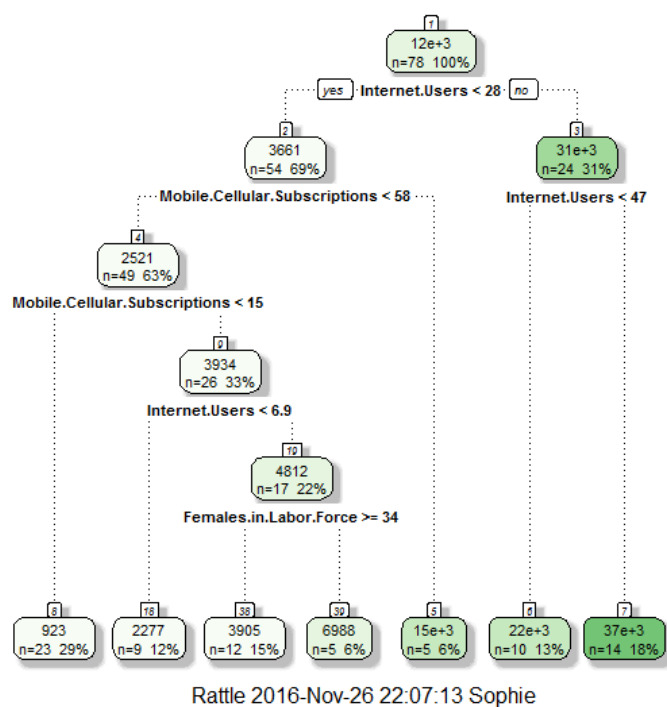


Figure 16 Partitioning tree for 2003

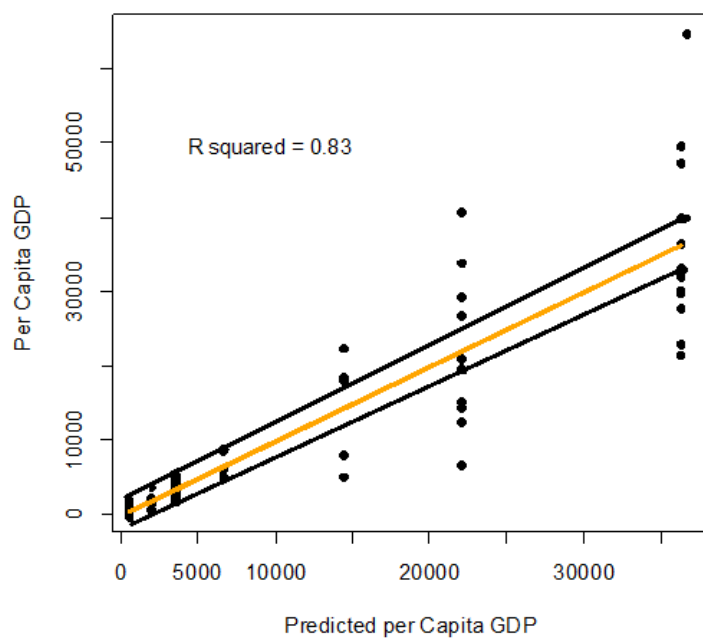


Figure 17 Predicted per capita GDP vs. Actual for 2003

## Discussion and Conclusion

The matrix scatter plot is more interpretable than the summary statistics, which display average variable values per country. The first row of the scatter plots shows the association between per capita GDP (y-axis) and the input variables (x-axis). An examination of these graphs reveals that some variables are highly associated with GDP, e.g. power consumption, while others do not display any discernible associations, e.g. high tech exports. This plot is also used to reveal any associations between the independent x-variables. Certain variables show dependency, e.g. internet use and power consumption. Such confounding between variables poses challenges for interpreting the results of social studies. Possibilities of confounding variables for 2003 and 2013 are internet users and electric power consumption, while 2008 is electric power consumption, internet users, and mobile cellular subscriptions. The presented plots also did not demonstrate any outliers. Overall, although matrix scatter plots demonstrate the relationships between many variables, clustering and partitioning are more effective in exploring associations in a more quantified manner.

The groupings of countries within clusters remained relatively similar over the years, despite some notable differences. Northern Europe and other major countries, such as the United States and Australia (when included in the data) were continually clustered with each other based on social factors. Some noticeable differences in clusters were with the Middle East, Southeastern Asia and Africa (more specifically: Tanzania, Mozambique, Ghana, and Turkey). These adjustments can be interpreted to represent changes in the variable values of these countries or other influential ones which affected the organization of similar values. The clustering analysis shows that countries can be grouped in a meaningful manner based on variables beyond the ones used to measure economic progress.

The main goal of partitioning was to determine the input variables that create the most distinction among generated groups based on GDP, therefore, symbolizing the most division in GDP by country. The 2013 partitioning was mainly dependent on internet users and both 2008 and 2003 on internet users and mobile cellular subscriptions. These inputs stayed relatively consistent across the years, possibly signifying an association between per capita GDP and internet users and mobile cellular subscriptions

The final portion of data analysis was seen through the plot comparing predicted per capita GDP versus actual per capita GDP. Over all three years, there was a relatively consistent linear pattern between the predicted GDP and corresponding actual value demonstrating a fairly accurate GDP prediction by means of the partitioning classification tree via input values.

The main limitation encountered within this analysis was the difference in country quantities per analyzed year. With 2013 beginning as the base year, the countries and variables were selected to ensure the complete coverage of each value in this data set. When comparing this to 2003 and 2008, however, certain countries used in the 2013 analysis were missing values in other years and, therefore, had to be omitted. These differences could cause adjustments in the accuracy of the results.

Overall, to group countries, or other entities, based upon a single value can be misleading as it does not take into account other characteristics of such entity, such as social or technological developments. In order to account for the societal characteristics of countries, clustering techniques were used with variables of economic, social, technology, and infrastructure natures, analyzing ten variables altogether. The cluster analysis results, displayed as groupings of similar countries, aligned with the original beliefs of country associations. Through the partitioning of the data, the association between the selected variables and the per

capita GDP of the countries studied was examined. Resulting from this analysis, internet users and mobile cellular subscriptions were consistently associated with per capita GDP over the years. Although, the variable of mobile cellular subscriptions was not dominant across all three selected years, only 2003 and 2008. It is interesting to note that similar observations were concluded by researchers looking into primarily the relationship between per capita GDP and internet usages. Out of the 9 variables relating to GDP within the study, internet users resulted as the most dominant one.

## Appendix

```
# ATTACH DATA
setwd("C:/Users/Sophie/Desktop/SR Analysis")
data<-read.csv(file="2013FinalNoPopulation.csv")
attach(data)
data=na.omit(data)
# PERFORM EDA
summary(data)
pairs(data)
```

**Figure A** R Script for attaching and summarizing data for 2013

```
# CLUSTER ANALYSIS
rm(list=ls())
graphics.off()
input.file<-"2013FinalNoPopulation.csv"
no.of.clusters=7
print.file="Y"
setwd("C:/Users/Sophie/Desktop/SR Analysis")
library(lattice)
# READ DATA
data<-read.csv(input.file,na.strings=c("NA","NaN", " "),header=T)
attach(data)
# DATA PRE-PROCESSING
nomissingdata <- na.omit(data) # listwise deletion of missing
numericdata <-nomissingdata[sapply(data, is.numeric)]
scaleddata <- scale(numericdata) # standardize variables
# FAST EDA
summary(data)
NROW(data);NROW(na.omit(data))
sapply(data, function(x) sum(is.na(x)))
dev.new()
par(mfrow=c(3,5))
for (i in 1:dim(numericdata)[2])
{hist(numericdata[,i],xlab=names(numericdata)[i],main=NULL)}
# AID TO DETERMINE NUMBER OF CLUSTERS
maxclusters=min(20,NROW(nomissingdata)-1)
wss <- (nrow(scaleddata)-1)*sum(apply(scaleddata,2,var))
for (i in 2:maxclusters) wss[i] <- sum(kmeans(scaleddata,
centers=i)$withinss)
dev.new()
plot(1:maxclusters, wss, type="b", xlab="Number of Clusters", ylab="Within groups sum of
squares",cex=0.5)
# K-MEANS CLUSTER ANALYSIS
```

```

set.seed(1111)
fit <- kmeans(scaleddata, no.of.clusters) # cluster solution
dev.new()
library(fpc)
plotcluster(scaleddata, fit$cluster)
# PARALLEL PLOT
dev.new()
scaleddata <- data.frame(scaleddata, fit$cluster)
parallelplot(~scaleddata[1:dim(scaleddata)[2]-1] | as.factor(fit.cluster), scaleddata)
# MATRIX SCATTER PLOT
dev.new()
numericdata <- data.frame(numericdata, fit$cluster)
pairs(numericdata[, -
dim(numericdata)[2]], col=as.factor(numericdata$fit.cluster), lower.panel=NULL, pch=16, cex=2)
# DATA FRAME WITH CLUSTER INFORMATION
nomissingdata<-data.frame(nomissingdata, fit$cluster)
ordered<-nomissingdata[with(nomissingdata, order(nomissingdata$fit.cluster)),]
# GET CLUSTER MEANS
aggregate(numericdata, by=list(fit$cluster), FUN=mean)
# PLOT CLUSTER MEANS BY EACH VARIABLE
a<-data.frame(aggregate(numericdata, by=list(fit$cluster), FUN=mean))
dev.new()
par(mfrow=c(3,5))
for (i in 2:as.numeric(dim(a)[2]-1)) {dotchart(a[,i], main=names(a[i]), col=a[,1], pch=16,
cex=.8)}
# WRITE CLUSTER RESULTS INTO .CSV FILE
if (print.file=="Y") write.csv(nomissingdata, paste(input.file, " ", no.of.clusters, "
clusters", ".csv")) # clustered data in the original order
if (print.file=="Y") write.csv(ordered, paste(input.file, " ", no.of.clusters, " clusters
ordered", ".csv")) # clustered data ordered by clusters

```

**Figure B** Cluster script for 2013

```

par(mfrow=c(2,5))
lw1 <- loess(Per.Capita.GDP ~Life.Expectancy, data=a, family="gaussian", span=.75, degree=0)
plot(Per.Capita.GDP ~Life.Expectancy, data=a, pch=19, cex=1.5, col="blue")
j <- order(a$Life.Expectancy)
lines(a$Life.Expectancy[j], lw1$fitted[j], col="orange", lwd=3)
lw1 <- loess(Per.Capita.GDP ~Females.in.Labor.Force, data=a, family="gaussian", span=.75,
degree=0)
plot(Per.Capita.GDP ~Females.in.Labor.Force, data=a, pch=19, cex=1.5, col="blue")
j <- order(a$Females.in.Labor.Force)
lines(a$Females.in.Labor.Force[j], lw1$fitted[j], col="orange", lwd=3)
rm(list=ls()); options(scipen = 999)
# Read data
data<-read.csv("2013FinalNoPopulation.csv", header = TRUE, na.strings=c("NA", "NaN", " "))
attach(data)
# Partitioning
library(RGtk2)
library(rpart)
library(rattle)
library(rpart.plot)
library(RColorBrewer)
# PARTITIONING ANALYSIS: Final Tree
per.capita.GDP.fit<-
rpart(Per.Capita.GDP~+Life.Expectancy+Females.in.Labor.Force+Internet.Users+
Mobile.Cellular.Subscriptions+High.Tech.Exports+Women.in.National.Parliament+Improved.Water
.Source+
Primary.Education.Enrollment
,minbucket=5, cp=0.001);
dev.new(); fancyRpartPlot(per.capita.GDP.fit); dev.new()
data=cbind(data, data.frame(predict(per.capita.GDP.fit)))
plot( data$predict.per.capita.GDP.fit., data$Per.Capita.GDP, xlab="Predicted per capita
GDP", ylab="Per capita GDP", pch=19)
lines(data$predict.per.capita.GDP.fit., fitted(lm(data$Per.Capita.GDP~data$predict.per.capit
a.GDP.fit.)), col="orange", lwd=3)
dev.new()

```

```

text(x=20000,y=100000, paste("R squared = ",
round(summary(lm(data$Per.Capita.GDP~data$predict.per.capita.GDP.fit.))$r.squared,digits=2)
, sep=""))
lines(data$predict.per.capita.GDP.fit.,fitted(lm(data$Per.Capita.GDP~data$predict.per.capit
a.GDP.fit.)),col="orange",lwd=3)
a <- predict(lm(data$Per.Capita.GDP~data$predict.per.capita.GDP.fit.),
interval="confidence",level=0.99)
lines(data$predict.per.capita.GDP.fit., a[,2], lty=2)
lines(data$predict.per.capita.GDP.fit., a[,3], lty=2)

```

**Figure C** Script for partitioning of 2013, omitted loess portions for rest of variables

```

rm(list=ls())
graphics.off()
library(RColorBrewer)
library(maptools)
data(wrld_simpl)
ddf = read.table(text="
country value
# COPY COUNTRY INFO HERE : 'Cote d'Ivoire' 2      'Moldova' 7 omitted
'Mozambique' 1
'Niger' 1
'United Republic of Tanzania' 1
'Zambia' 1
'Azerbaijan' 2
'Belarus' 2
'Botswana' 2
'Kazakhstan' 2
'Kyrgyzstan' 2
'Namibia' 2
'Peru' 2
# omitted rest
", header=TRUE)
#pal <- colorRampPalette(brewer.pal(5, "Set1"),bias=3)(length(ddf$value))
pal <- palette()
pal <- pal[with(ddf, findInterval(ddf$value, sort(unique(ddf$value))))]
col <- rep(grey(1), length(wrld_simpl@data$NAME))
col[match(ddf$country, wrld_simpl@data$NAME)] <- pal
plot(wrld_simpl, col = col,main="Total")

```

**Figure D** Script for making map clusters for 2013 (omitted similarly formatted country values)

#### Countries used (some omitted depending on years):

Albania United Arab Emirates Australia Azerbaijan Belgium Bulgaria Bosnia and Herzegovina Belarus Brazil Botswana Canada Switzerland Chile Cote d'Ivoire Colombia Costa Rica Cyprus Germany Denmark Dominican Republic Ecuador Spain Estonia Finland France United Kingdom Ghana Greece Guatemala Hungary Indonesia India Ireland Israel Italy Japan Kazakhstan Kyrgyz Republic Kuwait Lebanon Sri Lanka Lithuania Luxembourg Latvia Morocco Moldova Mexico Malta Mongolia Mozambique Mauritius Namibia Niger Norway Nepal New Zealand Oman Pakistan Panama Peru Philippines Poland Portugal Russian Federation Saudi Arabia El Salvador Slovenia Sweden Tunisia Turkey Tanzania Ukraine Uruguay United States Venezuela, RB Vietnam Zambia Zimbabwe

Cluster	Color in Figure 8	Interpretation	Sample Countries
1	Yellow	Most developed, highest GDP	Sweden, Germany, United States
2	Purple	High social development	Saudi Arabia, United Arab Emirates
3	Light Blue	Moderate social development	Slovenia, Ukraine, Japan
4	Dark Blue	Average development and GDP	Vietnam, Malta, Kazakhstan
5	Green	Below average development and GDP	Zimbabwe, Tanzania, Pakistan



6	Red	Low social development and GDP	Brazil, Nepal, Colombia
7	Black	Lowest GDP, least social development	Turkey, Albania, India

Figure E Cluster map interpretation for 2013

### References

1. Amiri, Shahram, and Brian Reif. "Internet Penetration and its Correlation to Gross Domestic Product: An Analysis of the Nordic Countries." *International Journal of Business, Humanities and Technology*, vol. 3, no. 2, Feb. 2013, [webcache.googleusercontent.com/search?q=cache:InuLk6gwFfoJ:www.ijbhtnet.com/journals/Vol\\_3\\_No\\_2\\_February\\_2013/5.pdf+&cd=1&hl=en&ct=clnk&gl=us](http://webcache.googleusercontent.com/search?q=cache:InuLk6gwFfoJ:www.ijbhtnet.com/journals/Vol_3_No_2_February_2013/5.pdf+&cd=1&hl=en&ct=clnk&gl=us). Accessed 6 Dec. 2016.
2. Corsi-Bunker, Antonella. "Guide to the Education System in the United States." *International Student and Scholar Services: University of Minnesota*, [iss.umn.edu/publications/USEducation/4.pdf](http://iss.umn.edu/publications/USEducation/4.pdf). Accessed 6 Dec. 2016.
3. "Gender Development Index." *United Nations Development Programme*, [hdr.undp.org/en/content/gender-development-index-gdi](http://hdr.undp.org/en/content/gender-development-index-gdi). Accessed 6 Dec. 2016.
4. "Global High Tech Exports by Country." *World Atlas*, [www.worldatlas.com/articles/countries-with-the-most-high-tech-exports.html](http://www.worldatlas.com/articles/countries-with-the-most-high-tech-exports.html). Accessed 6 Dec. 2016.
5. "Greater Technology Use Linked to Higher Per capita Income." *Pew Research Center*, 13 Feb. 2014, [www.pewglobal.org/2014/02/13/developing-technology-use/](http://www.pewglobal.org/2014/02/13/developing-technology-use/). Accessed 6 Dec. 2016.
6. "Human Development Report." *United Nations Development Programme*, [hdr.undp.org/en](http://hdr.undp.org/en). Accessed 6 Dec. 2016.

7. "Improved and Unimproved Water Sources and Sanitation Facilities." *Joint Monitoring Programme for Water Supply and Sanitation*, [www.wssinfo.org/definitions-methods/watsan-categories/](http://www.wssinfo.org/definitions-methods/watsan-categories/). Accessed 6 Dec. 2016.
8. "Internet Users." *Internet Live Stats*, [www.internetlivestats.com/internet-users/](http://www.internetlivestats.com/internet-users/). Accessed 6 Dec. 2016.
9. "Millennium Development Goals Indicators." *United Nations Statistics Division*, [mdgs.un.org/unsd/mdg/Metadata.aspx?IndicatorId=0&SeriesId=780](http://mdgs.un.org/unsd/mdg/Metadata.aspx?IndicatorId=0&SeriesId=780). Accessed 6 Dec. 2016.
10. "The strong relationship between per capita income and internet access, smartphone ownership." *Pew Research Center*, [www.pewglobal.org/interactives/tech-gdp/](http://www.pewglobal.org/interactives/tech-gdp/). Accessed 6 Dec. 2016.
11. "Visualizing the Relationship between Internet Usage and GDP per capita." *World Bank Data Viz*, [worldbank.tumblr.com/post/45766793425/visualizing-the-relationship-between-internet?soc\\_src=mail&soc\\_trk=ma](http://worldbank.tumblr.com/post/45766793425/visualizing-the-relationship-between-internet?soc_src=mail&soc_trk=ma). Accessed 6 Dec. 2016.