Country Analysis by GDP and Income Group for 2012

James Hosker

March 22, 2017

Introduction

This report examines data from the World Bank in 2012 on Gross Domestic Product (GDP) combined with country specific Federal Statistics by country. A summary of our methods and analyses are below:

I. We successfully merge and clean two datasets from the World Bank: 2012 GDP data of 190 countries and 2012 Federal Statistics of 235 countries. This resulting dataset for GDP vs. Federal Statistics provides data on 190 countries spanning several income groups for low, middle, upper and high income countries. For more information on the raw datasets, see the References section at the end of this report.

II. We rank the 2012 cleaned dataset by increasing GDP in USD millions and break it out into income group categories for each country: low income; lower middle income; upper middle income; high income: nonOECD; and high income: OECD. OECD is the classification from the Organization for Economic Cooperation and Development (OECD), for more information see OECD website link.

    1. In merging the datasets, we provide how many country code IDs matched.
    2. We show as an example the 13th country in GDP rank after sorting each country by GDP in USD millions in ascending order.
    3. For the high income: OECD and nonOECD country income groups, we compared the mean and summary statistics for GDP rankings and GDP in USD millions.
    4. We plot GDP for all of the countries and plot GDPs by country income group. In addition, we provide summary statistics GDP ranks by country income groups.
    5. We separate the country GDP ranks into 5 separate quantile groups, making a table of increasing GDPs with country income groups. In addition, we determine how many countries are lower middle income but among the 38 nations with highest GDPs globally. Perhaps these countries will have income group upgrades in the future, if GDP continues to remain high and expand.

III. Finally, we provide our conclusions and recommendations for future work.

Folder Description

The URL with the code and our analysis is on the GitHub website: https://github.com/jjhosker/CaseStudy1. It contains the following important files:

1. README: This file provides more information in detail of each of the files in the GitHub directory provided above.
2. RMakefile_CaseStudy1.R: This file can be found in the Analysis sub-directory under this GitHub repository. It is the main R makefile that takes the cleaned data and performs all the analysis required for this project and in this report. In cleaning the data, we read in the csv files and in some cases re-name variables and reset data types. In addition, we remove NAs and blanks in the data. Finally, we merge the data by the unique country code in both datasets and create and output one cleaned and merged dataset (GDPFEDclean).

3. Makefile_GatherData.R: This file can be found in the Analysis/Data sub-directory under this GitHub repository. This R makefile installs all required R-libraries and then loads the R-libraries. Second, this R makefile downloads in the csv data files from the World Bank websites using URLs, reads the downloaded csv files, creates the raw data, and creates the cleaned datasets. Finally, this R makefile merges the GDP with the Federal Statistical dataset and creates one a clean data set. This makefile then saves all these dataset to the Analysis/Data sub-directory.

4. PlotFunctions_CaseStudy1.R: In the Anlaysis subdirectory of this GitHub repository, this R Makefile loads plot functions to plot Figures 1 and 2 used in RMakefile_CaseStudy1.R and CaseStudy1PDF.Rmd.

5. There are three data files created below: two raw datsets downloaded from the World Bank URLs and one cleaned dataset. In addition, we provide a text file (SessionsInfo.txt) containing all the R session information including the current R version and all the libraries and packages being utilized. All of the files below can be found in the Analysis/Data sub-directory under this GitHub repository.
   - Raw: FEDSTATS_Country.csv
   - Raw: GDP_Country.csv
   - Cleaned: GDPFEDclean.csv
   - R version: SessionInfo.txt

6. Finally, the CaseStudy1PDF.Rmd file runs the main RMakefile_CaseStudy1.R and extracts code to generate this "CaseStudy1PDF.pdf"" report. This report uses code that that is extracted from the main RMakefile_CaseStudy1.R and then run.
   - To use this Rmd file to create a PDF, the user has to install MikTex (the non-basic version). The URL with instructions to install MikTex is MikText Install website.

Running the R-Code

Before beginning please review the README.md file in GitHub website https://github.com/jjhosker/CaseStudy1. In short, download and clone the GitHub repository and the sub-directories. RMakefile_CaseStudy1.R should be executed in the Analysis sub-directory of this GitHub repository; and CaseStudy1PDF.Rmd can be executed from this GitHub repository.

1. In the Analysis sub-directory in this repository, execute the main R source code RMakefile_CaseStudy1.R to load all the libraries, read the raw data, create the cleaned data and perform the analysis for this client study. The csv files will be created in the same directory as the project that you created on your computer. In addition, we save the R session information and all the libraries and packages being utilized.
   - In the Analysis sub-directory in this repository, the RMakefile_CaseStudy.R executes the makefile Makefile_GatherData.R, which installs the packages and load the libraries. A message indicating that the libraries loaded may appear after executing the R code that follows. If a failure occurs, please check the function "packagelibrary.check" in the Makefile_GatherData.R and/or the SessionInfo.txt file, both found in the Analysis/Data sub-directory.

2. To generate this report, execute CaseStudy1PDF.Rmd in this GitHub repository. Please note the required installations of software to generate the PDF is detailed in the prior section, Folder Description.

```
#####################################################
## Execute Main RMakefile_CaseStudy1.R Makefile
#####################################################
source("Analysis/RMakefile_CaseStudy1.R")
```

```
## Loading required package: devtools

## Loading required package: repmis

## Loading required package: downloader

##
## Attaching package: 'downloader'

## The following object is masked from 'package:devtools':
##
##     source_url

## Loading required package: ggplot2

## Loading required package: xtable

## Loading required package: kableExtra

## Loading required package: knitr

## Session info -------------------------------------------------------------

## Packages -----------------------------------------------------------------
```

Five Key or Major Points/Questions:

Key Point #1: Country Code IDs that Match & Variables Names

The R code below prints the necessary information around the IDs that match and the variables that we are using. For more details, see R-code section Key Point/Question #1 in RMakefile_CaseStudy1.R.

```
print(paste("Number of country code IDs that match: ",
 sum(!is.na(GDPFEDclean$GDP_MillUSD)),"."))
```

```
## [1] "Number of country code IDs that match:  190 ."
```

```
print(paste("Number of country code IDs that do not match: ",
 sum(is.na(GDPFEDclean$GDP_MillUSD)),"."))
```

```
## [1] "Number of country code IDs that do not match:  45 ."
```

```
print(paste("Total Number of Variables: ",length(GDPFEDclean),"."))
```

```
## [1] "Total Number of Variables:  7 ."
```

```
cat(paste("Variable names are:", toString(variable.names(GDPFEDclean)[1:4]),"\n",
          "                    ",
          toString(variable.names(GDPFEDclean)[5:7]),"."))
```

```
## Variable names are: CountryCode, GDPRank, Economy, GDP_MillUSD
##                     Long.Name, Income.Group, Region .
```

1. The number of country codes IDs that match between the 2012 World Bank GDP and Federal Statistic datasets is 190. There are 190 entries in the World Bank GDP dataset and 235 in the Federal Statistic dataset. Therefore, 45 do not match and are generally small countries or countries that have little GDP, like North Korea (PRK) even though South Korea is part of those countries that do match. In addition, there are some countries with two entries such as Samoa and American Samoa, in which only Samoa is in both datasets.

2. We can see that the out merged dataset that we use 7 major variables for this analysis: CountryCode, GDPRank, Economy, GDP_MillUSD, Long.Name, Income.Group and Region.

Key Point #2: GPD & GDP Rank by Country

The R code below finds the 13th lowest ranked country by GDP (ranked lowest to highest). For more details, see R-code section Key Point #2 in RMakefile_CaseStudy1.R.

```
## find 13th entry into the data frame after sorting
print(paste("13th Ranked Country by GDP is",
 GDPFEDsort$CountryCode[13], ", which is ", GDPFEDsort$Economy[13],"."))
```

```
## [1] "13th Ranked Country by GDP is KNA , which is  St. Kitts and Nevis ."
```

We sort the data by ascending GDP from (lowest to highest) and we can find the country with the lowest to highest GDP. For example, we find that the 13th country from the country with the lowest GDP is St. Kitts and Nevis (KNA) with a 2012 GDP of $767 Million.

Key Point #3: Mean Stats for High Income OECD & non-OECD Countries

The R code below finds the mean GDP rank of two income groups. For more details, see R-code section Key Point #3 in RMakefile_CaseStudy1.R.

```
cat(paste("The mean GDP rank of the High Income:  OECD income group is",
    round(MeanRankOECD,1),"\n       & the GDP is in",
    round(MeanGDPOECD,1),"USD Millions"))
```

```
## The mean GDP rank of the High Income:  OECD income group is 33
##        & the GDP is in 1483917.1 USD Millions
```

```
cat(paste("The mean GDP rank of the High Income: nonOECD income group is",
    round(MeanRanknonOECD,1),"\n      & the GDP is in",
    round(MeanGDPnonOECD,1),"USD Millions"))
```

```
## The mean GDP rank of the High Income: nonOECD income group is 91.9
##       & the GDP is in 104349.8 USD Millions
```

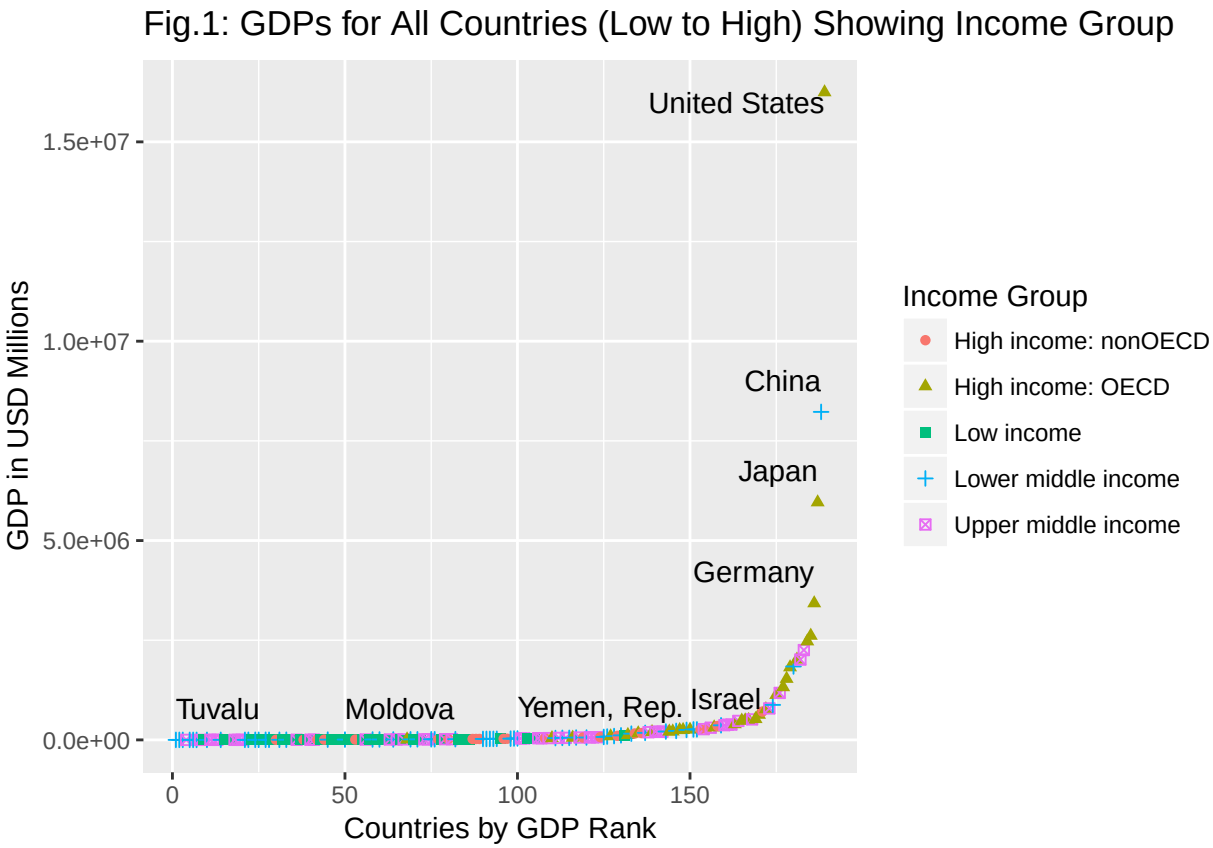
1. GDP Rank give highest GDP at 1 and lowest at 190.
2. Here we analyze the data and find that the mean GDP Rankfor the high income: OECD countries is 32.97 (higher) and the mean GDP for the high income: OECD countries is $1,483,917 million.
3. Here we analyze the data and find that the mean GDP Rank for the high income: non-OECD countries is lower at 91.97 (lower) and the mean GDP for the high income: non-OECD countries is $104,349.8 million.

Key Point #4: Graphs & Summary Stats on Country GDPs vs. Income Groups

Here we provide three graphs and one table for GDP and GDP vs. Income Group. GDP in USD Millions is ranked from lowest to highest. One country gets eliminated since South Sudan (SSD) has an NA for country income group but has a valid GDP of 10,200 USD Million and valid GDP rank. The number of countries with both GDP and income group reduces to 189 from 190.

This R code below provides a plot for GDP by country income groups. It also provides a summary table of statistics for each income group by GDP ranks. For more details, see R-code section Key Point #4 in RMakefile_CaseStudy1.R.

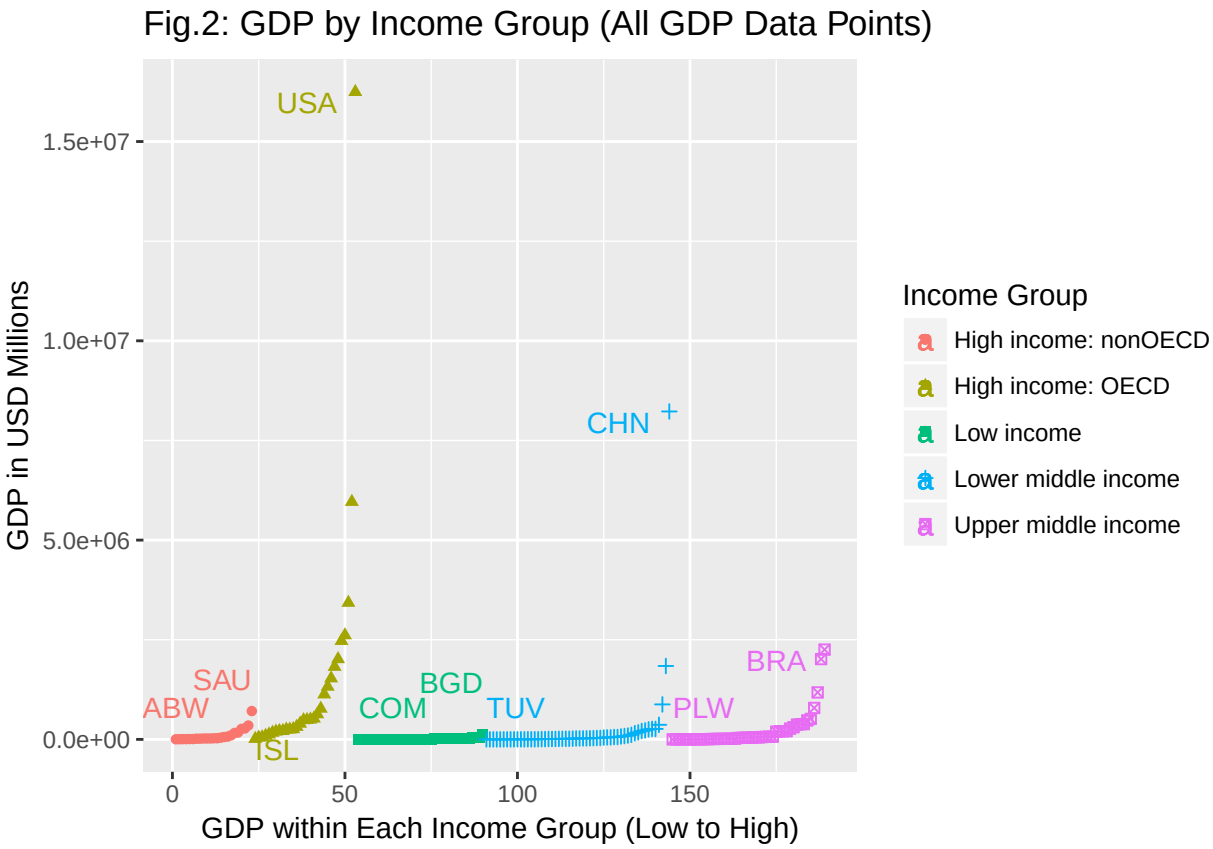
```
## Plot all GDPs with income group classifications
PlotFig1()
```



1. Figure 1 is the GDP of all Countries is shown first from Lowest to Highest GDP. Each country income group is highlighted in a different shape and color. We also label the countries with the lowest, the 50th, the 100th, the 150th and the four top GDP ranks.

This R code below provides a plot for GDP by country income groups separately. For more details, see R-code section Key Point #4 in RMakefile_CaseStudy1.R.

```
## Plot GDP ranks of each income group separately and label lowest to highest by country code
PlotFig2()
```

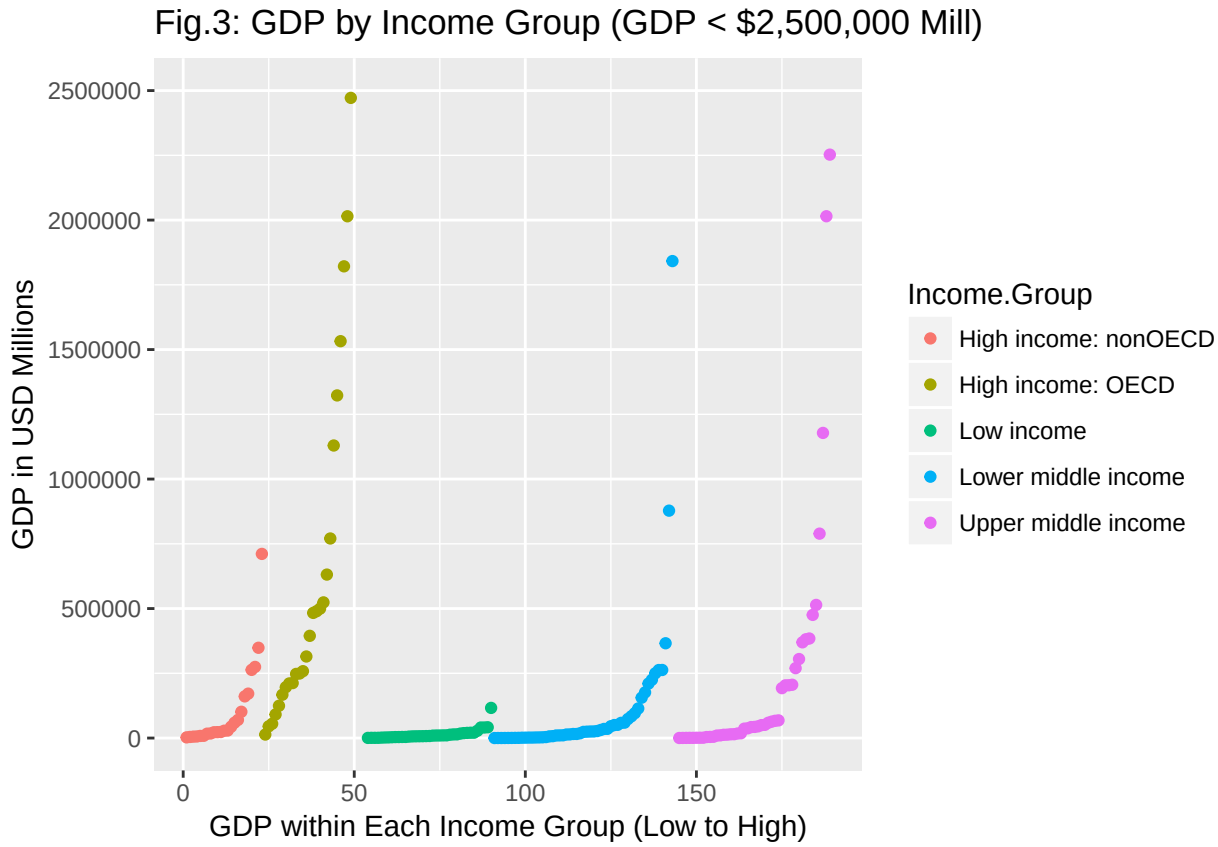


2. Figure 2 is the GDP of countries by income group from lowest to highest GDP within each income group in order to get a clearer picture of each income group individually. In addition, we provide a label for the lowest and highest GDP of each income group by its country code.

This R code below provides a plot for GDP by country income groups separately capped at 2,500,000 USD million. For more details, see R-code section Key Point #4 in RMakefile_CaseStudy1.R.

```
## Plot GDPs of each income group separately for GDP < $2,500,000 Mill
ggplot(subset(GDPFEDinc, Income.Group %in% c("Low income","Lower middle income",
 "Upper middle income", "High income: nonOECD", "High income: OECD")),
 aes(x=GDPFEDinc$CntryNum, y=GDPFEDinc$GDP_MillUSD, color=Income.Group)) +
 geom_point() +  ggtitle("Fig.3: GDP by Income Group (GDP < $2,500,000 Mill)") +
 labs(x="GDP within Each Income Group (Low to High)", y="GDP in USD Millions") +
 ylim(0,2500000)
```

```
## Warning: Removed 5 rows containing missing values (geom_point).
```



3. Figure 3 is the GDP of countries by income group from lowest to higher GDP within each income group for GDPs <= 2,500,000 USD million OR 2.5 USD trillion. It shows more clearly the distribution of GDPs for each category capping those GDPs that are displayed (removing 5 countries with very high GDPs).

This R code below provides summary statistics table of GDP ranks for each country income group. For more details, see R-code section Key Point #4 in RMakefile_CaseStudy1.R.

```
## Create table of GDP ranks by Country Income Group
kable((sumtable), format="pandoc", align='c',row.names = FALSE,col.names = sumclass,
 caption = "Summary Statistics of GDP Ranks by Income Groups", echo = TRUE)
```

Table 1: Summary Statistics of GDP Ranks by Income Groups

Income.Group	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
Low income	59	112.00	139.0	133.70	156.00	182
Lower middle income	2	64.25	99.5	107.70	162.20	190
Upper middle income	7	48.00	83.0	92.13	135.00	187
High income: nonOECD	19	58.50	94.0	91.91	125.00	161
High income: OECD	1	12.25	24.5	32.97	45.75	122

4. In Table 1, we show summary statistics on the GDP ranks of each income group (highest GDP is ranked 1, lowest GDP is ranked 189). We can see that those classified as low income have the lowest cluster of GDPs (mean GDP rank 133.7). The lower middle income group (mean GDP rank is 107.7) has a few outliers that are larger than the next higher category of the upper income group. However, in general, the lower middle income group has GDPs that are near or slightly higher (lower GDP ranks) than the low income group. The upper middle income (mean GDP rank is 92.13) generally has GDPs higher (lower GDP ranks) than the lower middle income group (excluding the outliers). In general, the 1st quantiles to 3rd quantile follows a distribution in line with our expectations that lower income countries have higher GDP ranks (lower GDPs) than higher income countries (lower GDP rank, higher GDPs)
5. Finally, we see that the high income group of OECD countries (mean GDP rank is 32.97) has the higher GDPs of any income group category. In addition, the high income group with non-OECD countries has a similar GDPs (mean GDP rank 92.17) compared to the upper middle income group (mean GDP rank 92.13).

Key Point #5: Rank GDP by Quintiles & Compare to Income Groups

This R code provides analysis of ranking GDP by quintiles versus income groups. For more details, see R code section Key Point #5 in RMakefile_CaseStudy1.R.

```
## Use kable in knitr to render table of GDP by Quintiles with Income Groups
kable((GDPFEDsort2[,colvars]), format="pandoc", align='c',
 row.names = FALSE, col.names = TableColNames,
 caption = "Summary Data by Quintile Ranking of GDP ($Mill)", echo = TRUE)
```

Table 2: Summary Data by Quintile Ranking of GDP ($Mill)

Quintile	GDP ($Mill)	Income Group	GDP Rank	Country Code	Country
Q1	40	Lower middle income	190	TUV	Tuvalu
Q1	175	Lower middle income	189	KIR	Kiribati
Q1	182	Lower middle income	188	MHL	Marshall Islands
Q1	228	Upper middle income	187	PLW	Palau
Q1	263	Lower middle income	186	STP	São Tomé and Principe
Q1	326	Lower middle income	185	FSM	Micronesia, Fed. Sts.
Q1	472	Lower middle income	184	TON	Tonga
Q1	480	Upper middle income	183	DMA	Dominica
Q1	596	Low income	182	COM	Comoros
Q1	684	Lower middle income	181	WSM	Samoa
Q1	713	Upper middle income	180	VCT	St. Vincent and the Grenadines
Q1	767	Upper middle income	178	GRD	Grenada
Q1	767	Upper middle income	178	KNA	St. Kitts and Nevis
Q1	787	Lower middle income	177	VUT	Vanuatu

Quintile	GDP ($Mill)	Income Group	GDP Rank	Country Code	Country
Q1	822	Low income	176	GNB	Guinea-Bissau
Q1	917	Low income	175	GMB	Gambia, The
Q1	1008	Low income	174	SLB	Solomon Islands
Q1	1129	Upper middle income	173	SYC	Seychelles
Q1	1134	Upper middle income	172	ATG	Antigua and Barbuda
Q1	1239	Upper middle income	171	LCA	St. Lucia
Q1	1293	Lower middle income	170	TMP	Timor-Leste
Q1	1493	Lower middle income	169	BLZ	Belize
Q1	1734	Low income	168	LBR	Liberia
Q1	1780	Lower middle income	167	BTN	Bhutan
Q1	1827	Lower middle income	166	CPV	Cape Verde
Q1	2184	Low income	165	CAF	Central African Republic
Q1	2222	Lower middle income	164	MDV	Maldives
Q1	2448	Lower middle income	163	LSO	Lesotho
Q1	2472	Low income	162	BDI	Burundi
Q1	2584	High income: nonOECD	161	ABW	Aruba
Q1	2851	Lower middle income	160	GUY	Guyana
Q1	3092	Low income	159	ERI	Eritrea
Q1	3744	Lower middle income	158	SWZ	Swaziland
Q1	3796	Low income	157	SLE	Sierra Leone
Q1	3814	Low income	156	TGO	Togo
Q1	3908	Upper middle income	155	FJI	Fiji
Q1	4199	Low income	154	MRT	Mauritania
Q1	4225	High income: nonOECD	153	BRB	Barbados
Q2	4264	Low income	152	MWI	Malawi
Q2	4373	Upper middle income	151	MNE	Montenegro
Q2	5012	Upper middle income	150	SUR	Suriname
Q2	5474	High income: nonOECD	149	BMU	Bermuda
Q2	5632	Low income	148	GIN	Guinea
Q2	6075	High income: nonOECD	147	MCO	Monaco
Q2	6445	Lower middle income	146	KSV	Kosovo
Q2	6475	Low income	145	KGZ	Kyrgyz Republic
Q2	6773	Low income	144	NER	Niger
Q2	6972	Low income	143	TJK	Tajikistan
Q2	7103	Low income	142	RWA	Rwanda
Q2	7253	Lower middle income	141	MDA	Moldova
Q2	7557	Low income	140	BEN	Benin
Q2	7843	Low income	139	HTI	Haiti
Q2	8149	High income: nonOECD	138	BHS	Bahamas, The
Q2	8722	High income: nonOECD	137	MLT	Malta
Q2	9418	Low income	136	LAO	Lao PDR
Q2	9613	Upper middle income	135	MKD	Macedonia, FYR
Q2	9802	Low income	134	ZWE	Zimbabwe
Q2	9951	Lower middle income	133	ARM	Armenia
Q2	9975	Low income	132	MDG	Madagascar
Q2	10271	Lower middle income	130	MNG	Mongolia
Q2	10308	Low income	129	MLI	Mali
Q2	10441	Low income	128	BFA	Burkina Faso
Q2	10486	Upper middle income	127	MUS	Mauritius
Q2	10507	Lower middle income	126	NIC	Nicaragua
Q2	12648	Upper middle income	125	ALB	Albania
Q2	12887	Low income	124	TCD	Chad

Quintile	GDP ($Mill)	Income Group	GDP Rank	Country Code	Country
Q2	13072	Upper middle income	123	NAM	Namibia
Q2	13579	High income: OECD	122	ISL	Iceland
Q2	13678	Lower middle income	121	COG	Congo, Rep.
Q2	14038	Low income	120	KHM	Cambodia
Q2	14046	Lower middle income	119	SEN	Senegal
Q2	14244	Low income	118	MOZ	Mozambique
Q2	14504	Upper middle income	117	BWA	Botswana
Q2	14755	Upper middle income	116	JAM	Jamaica
Q2	15654	Lower middle income	115	PNG	Papua New Guinea
Q2	15747	Lower middle income	114	GEO	Georgia
Q3	16954	High income: nonOECD	113	BRN	Brunei Darussalam
Q3	17204	Low income	112	ZAR	Congo, Dem. Rep.
Q3	17466	Upper middle income	111	BIH	Bosnia and Herzegovina
Q3	17697	High income: nonOECD	110	GNQ	Equatorial Guinea
Q3	18377	Upper middle income	109	GAB	Gabon
Q3	18434	Lower middle income	108	HND	Honduras
Q3	18963	Low income	107	NPL	Nepal
Q3	19881	Low income	106	UGA	Uganda
Q3	20497	Low income	105	AFG	Afghanistan
Q3	20678	Low income	104	ZMB	Zambia
Q3	22390	High income: nonOECD	103	EST	Estonia
Q3	22767	High income: nonOECD	102	CYP	Cyprus
Q3	23320	High income: nonOECD	101	TTO	Trinidad and Tobago
Q3	23864	Lower middle income	100	SLV	El Salvador
Q3	24680	Lower middle income	99	CIV	Côte d'Ivoire
Q3	25322	Lower middle income	98	CMR	Cameroon
Q3	25502	Lower middle income	97	PRY	Paraguay
Q3	27035	Lower middle income	96	BOL	Bolivia
Q3	28242	Low income	95	TZA	Tanzania
Q3	28373	High income: nonOECD	94	LVA	Latvia
Q3	29044	High income: nonOECD	93	BHR	Bahrain
Q3	31015	Lower middle income	92	JOR	Jordan
Q3	35164	Lower middle income	91	TKM	Turkmenistan
Q3	35646	Lower middle income	90	YEM	Yemen, Rep.
Q3	36253	Upper middle income	89	PAN	Panama
Q3	37489	Upper middle income	88	SRB	Serbia
Q3	40697	Low income	87	KEN	Kenya
Q3	40711	Low income	86	GHA	Ghana
Q3	41605	Low income	85	ETH	Ethiopia
Q3	42344	Upper middle income	84	LTU	Lithuania
Q3	42945	Upper middle income	83	LBN	Lebanon
Q3	43582	High income: nonOECD	82	MAC	Macao SAR, China
Q3	45104	Upper middle income	81	CRI	Costa Rica
Q3	45279	High income: OECD	80	SVN	Slovenia
Q3	45662	Lower middle income	79	TUN	Tunisia
Q3	49920	Upper middle income	78	URY	Uruguay
Q3	50234	Lower middle income	77	GTM	Guatemala
Q4	50972	Upper middle income	76	BGR	Bulgaria
Q4	51113	Lower middle income	75	UZB	Uzbekistan
Q4	55178	High income: OECD	74	LUX	Luxembourg
Q4	58769	Lower middle income	73	SDN	Sudan
Q4	59047	Upper middle income	72	DOM	Dominican Republic

Quintile	GDP ($Mill)	Income Group	GDP Rank	Country Code	Country
Q4	59228	High income: nonOECD	71	HRV	Croatia
Q4	59423	Lower middle income	70	LKA	Sri Lanka
Q4	63267	Upper middle income	69	BLR	Belarus
Q4	66605	Upper middle income	68	AZE	Azerbaijan
Q4	68234	Upper middle income	67	CUB	Cuba
Q4	69972	High income: nonOECD	66	OMN	Oman
Q4	73672	Lower middle income	65	SYR	Syrian Arab Republic
Q4	84040	Lower middle income	64	ECU	Ecuador
Q4	91149	High income: OECD	63	SVK	Slovak Republic
Q4	95982	Lower middle income	62	MAR	Morocco
Q4	101496	High income: nonOECD	61	PRI	Puerto Rico
Q4	114147	Lower middle income	60	AGO	Angola
Q4	116355	Low income	59	BGD	Bangladesh
Q4	124600	High income: OECD	58	HUN	Hungary
Q4	155820	Lower middle income	57	VNM	Vietnam
Q4	160913	High income: nonOECD	56	KWT	Kuwait
Q4	167347	High income: OECD	55	NZL	New Zealand
Q4	171476	High income: nonOECD	54	QAT	Qatar
Q4	176309	Lower middle income	53	UKR	Ukraine
Q4	192711	Upper middle income	52	ROM	Romania
Q4	196446	High income: OECD	51	CZE	Czech Republic
Q4	203521	Upper middle income	50	KAZ	Kazakhstan
Q4	203790	Upper middle income	49	PER	Peru
Q4	205789	Upper middle income	48	DZA	Algeria
Q4	210280	Lower middle income	47	IRQ	Iraq
Q4	210771	High income: OECD	46	IRL	Ireland
Q4	212274	High income: OECD	45	PRT	Portugal
Q4	225143	Lower middle income	44	PAK	Pakistan
Q4	247546	High income: OECD	43	FIN	Finland
Q4	249099	High income: OECD	42	GRC	Greece
Q4	250182	Lower middle income	41	PHL	Philippines
Q4	258217	High income: OECD	40	ISR	Israel
Q4	262597	Lower middle income	39	NGA	Nigeria
Q5	262832	Lower middle income	38	EGY	Egypt, Arab Rep.
Q5	263259	High income: nonOECD	37	HKG	Hong Kong SAR, China
Q5	269869	Upper middle income	36	CHL	Chile
Q5	274701	High income: nonOECD	35	SGP	Singapore
Q5	305033	Upper middle income	34	MYS	Malaysia
Q5	314887	High income: OECD	33	DNK	Denmark
Q5	348595	High income: nonOECD	32	ARE	United Arab Emirates
Q5	365966	Lower middle income	31	THA	Thailand
Q5	369606	Upper middle income	30	COL	Colombia
Q5	381286	Upper middle income	29	VEN	Venezuela, RB
Q5	384313	Upper middle income	28	ZAF	South Africa
Q5	394708	High income: OECD	27	AUT	Austria
Q5	475502	Upper middle income	26	ARG	Argentina
Q5	483262	High income: OECD	25	BEL	Belgium
Q5	489795	High income: OECD	24	POL	Poland
Q5	499667	High income: OECD	23	NOR	Norway
Q5	514060	Upper middle income	22	IRN	Iran, Islamic Rep.
Q5	523806	High income: OECD	21	SWE	Sweden
Q5	631173	High income: OECD	20	CHE	Switzerland

Quintile	GDP ($Mill)	Income Group	GDP Rank	Country Code	Country
Q5	711050	High income: nonOECD	19	SAU	Saudi Arabia
Q5	770555	High income: OECD	18	NLD	Netherlands
Q5	789257	Upper middle income	17	TUR	Turkey
Q5	878043	Lower middle income	16	IDN	Indonesia
Q5	1129598	High income: OECD	15	KOR	Korea, Rep.
Q5	1178126	Upper middle income	14	MEX	Mexico
Q5	1322965	High income: OECD	13	ESP	Spain
Q5	1532408	High income: OECD	12	AUS	Australia
Q5	1821424	High income: OECD	11	CAN	Canada
Q5	1841710	Lower middle income	10	IND	India
Q5	2014670	High income: OECD	9	ITA	Italy
Q5	2014775	Upper middle income	8	RUS	Russian Federation
Q5	2252664	Upper middle income	7	BRA	Brazil
Q5	2471784	High income: OECD	6	GBR	United Kingdom
Q5	2612878	High income: OECD	5	FRA	France
Q5	3428131	High income: OECD	4	DEU	Germany
Q5	5959718	High income: OECD	3	JPN	Japan
Q5	8227103	Lower middle income	2	CHN	China
Q5	16244600	High income: OECD	1	USA	United States

1. Table 2 shows the ranking of GDPs from lowest to highest, broken into five Quintiles: Q1, Q2, Q3, Q4 and Q5, where Q1 is the lowest quintile and Q5 is the highest quintile for GDP. One exception is that China is still classified as part of the lower middle income group with the second largest GDP in the world.

This R code provides a summary table of the count of each income group for each GDP quintile. For more details, see R code section Key Point #5 in RMakefile_CaseStudy1.R.

```
## Count for each quintile the number in each income category
kable((sumquint), format="pandoc", align='c',row.names = FALSE,
 ncol.names = varnameclass,
caption = "Summary Count of Income Groups in Each Quintile", echo = TRUE)
```

Table 3: Summary Count of Income Groups in Each Quintile

Income.Group	Q1	Q2	Q3	Q4	Q5
Low income	11	16	9	1	0
Lower middle income	16	9	11	13	5
Upper middle income	9	8	8	9	11
High income: OECD	0	1	1	10	18
High income: nonOECD	2	4	8	5	4

2. In Table 3, we summarize Table 2 by income group counts for each of the quintiles. In general, GDP rankings match income grouping expectations. As GDP increases Q1 to Q5, we see more upper middle income countries and more high income countries in the OECD as well as some high income countries that are non-OECD. Vice-versa as GDP decreases from Q5 to Q1, we see more low income and lower middle income countries.

This R code provides a list of the countries that are in the lower middle income group but are one of the top 38 countries by GDP rank. For more details, see R code section Key Point #5 in RMakefile_CaseStudy1.R.

```
## Find those countries that are in the Lower Middle income but among
## the top 38 nations with the highest GDP (using GDPRank)
Filtered <- GDPFEDsort2[GDPFEDsort2$GDPRank <= 38 &
            GDPFEDsort2$Income.Group == "Lower middle income",
            c("CountryCode","Economy","GDPRank")]
## Render Table of Countries meeting the criteria
kable((Filtered), format="pandoc", align='c',row.names = FALSE,
      col.names = c("CountryCode","Economy","GDPRank"),
      caption = "Countries in Lower Middle Income with GDPRank<=38")
```

Table 4: Countries in Lower Middle Income with GDPRank<=38

CountryCode	Economy	GDPRank
EGY	Egypt, Arab Rep.	38
THA	Thailand	31
IDN	Indonesia	16
IND	India	10
CHN	China	2

3. In Table 4, we find those countries that are in the "Lower Middle income" group that are among the top 38 nations with the highest GDP using the GDP Rank. These five countries may have the potential to have their income group category upgraded in future years.

Conclusions:

In this report, we addressed five major or key points.

- First we merge the datasets of GPD with Federal Statistics from the World Bank and find that 190 country codes match.
- Second, we rank order each country by GDP and find that the 13th country from the lowest GDP is St. Kitts and Nevis (KNA) with a 2012 GDP of 767 USD Million
- Third, we find that the mean GDP Rank for the high income: OECD countries is 32.97 (higher) and the mean GDP is 1,483,917 USD million. In addition, we find that the mean GDP rank for high income: non-OECD countries is lower at 91.97 (lower) and the mean GDP is 104,349.8 USD million.
- Fourth, we find that in general that those classified in the low income group have the lowest GDPs, followed by the lower middle income group with slightly higher GDPs and then followed by the upper middle income group. Finally, we see that the high income group of OECD countries has the higher GDPs of any income group category. The high income group with non-OECD countries has a similar GDPs (mean GDP rank 92.17) compared to the upper middle income group (mean GDP rank 92.13).
- Fifth we show that when we break GDP by rank into five quintiles in ascending order, we generally see that the lower quintiles (Q1) have the lower income and the lower middle income groups, while the upper quintiles (Q5) have the high income group with OECD countries. One exception is that China is still classified as part of the lower middle income group with the second largest GDP in the world.
  - In addition, we found those countries that are in the "Lower Middle income" group and are among the top 38 nations with the highest GDP using the GDP Rank. These five

countries may have the potential to have their income group category upgraded in future years: Egypt (EGY), Thailand (THA), Indonesia (IDN), India (IND) and China (CHN).

Future Work

We have explored much in this dataset but there is still more to analyze. We could analyze GDP or GDP rank to the region the country is in (e.g. Europe & Central Asia, Latin America & Caribbean, East Asia & Pacific, Sub-Saharan Africa, etc.). In addition, we could look at the last census of each country and compare that to GDP or GDP rank. Another comparison, is looking at GDP versus the type of government (consolidated or budgetary) or system of trade (general or specific).

References:

Relevant Data Links/URLs: We look at two 2012 datasets from the following World Bank websites:

1. Country specific GDP Data: World Bank GDP file link
   - For more information on World Bank GDP by country see: World Bank GDP website
2. Country specific Federal Statistical Data: World Bank Federal Stats file link
   - For more information on World Bank Federal Statistics by country see: World Bank Federal Stats website