

# World Bank Data Analysis

*Constantin Rusu, Ignacio Carracedo, Nestor Gallo, Avi Subramaniam*

*October 24, 2016*



## INTRODUCTION

The World Bank is an international organization that aims to defeat poverty in the world (specifically in countries with low to medium income) through means of training, development and loan offerings to countries so that they may “help countries help themselves”[1]. Started in 1944 as an initiative after World War II to rebuild a torn Europe (where the International Bank for Reconstruction and Development was also started), the World Bank currently offers several financial products and services such as:

- low interest loans
- low interest credits
- grants for developing countries
- financing through trust fund partnerships

The organization focuses on delivering measurable results and in order to do that it leads a powerful initiative for open development. This consists of free access to tools and data that spans over more than 50 years for several macroeconomic and microeconomic factors for countries from all over the globe.

## OBJECTIVE(S)

The world has seen some major shifts in the landscapes of politics, economy and society in the past 10 to 20 years. From the tragic events of September 11 in 2001 and the financial crisis of 2008 to the increasing gap in income equality and the rise of the internet as a universal digital medium, the world that has rebuilt after the second World War is anything if not a dynamic one. Social-economic interconnectedness (also known as economic globalization) has allowed for these changes to more

easily scatter and disseminate all over the world so much so that it is now perfectly reasonable to expect a piece of news one night to result in protests being organized in several places around the world by the break of dawn the following day.

Forces of change such as the ones presented so far are, however, unexpected and hard to predict. However, if there is one force of world change that has been relatively consistent since the beginning of mankind it would have to be: population growth. Data from the Population Reference Bureau [2] has attempted to model an answer for the question “How many people have ever lived on Earth?” and started by mapping out the human population (as a best estimate) at a few checkpoint moments throughout human history as can be seen below. (the first column represents the year, the second column represent the estimated population on Earth at that point in time, the third column represents the estimate number of births per one thousand people, the fourth column represents the number of births between the current in-line year and the previous benchmark point)

##	Year	Population	Births.per.1.000	Births.Between.Benchmarks
## 1	50,000 B.C.	2	-	-
## 2	8000 B.C.	5,000,000	80	1,137,789,769
## 3	1 A.D.	300,000,000	80	46,025,332,354
## 4	1200	450,000,000	60	26,591,343,000
## 5	1650	500,000,000	60	12,782,002,453
## 6	1750	795,000,000	50	3,171,931,513
## 7	1850	1,265,000,000	40	4,046,240,009
## 8	1900	1,656,000,000	40	2,900,237,856
## 9	1950	2,516,000,000	31-38	3,390,198,215
## 10	1995	5,760,000,000	31	5,427,305,000
## 11	2011	6,987,000,000	23	2,130,327,622

What is clear from this table is that the population on Earth seems to be headed in an upwards direction and it therefore poses the question: With the certainty of human population growing, how has this impacted the world we live in?

The goal of this paper is to study population growth and analyze how this growth impacts other variables such as mortality rate, arable land, CO2 emissions, energy use, GDP growth, etc. as well as researching the evolution of such factors across several countries through time. It is safe to assume that not all countries are impacted by population growth in the same way but, if anything, this creates an opportunity for digging into an interesting topic and revealing insightful results.

Other than studying population growth we also look to answer questions that include:

- Can the growth in CO2 emissions be explained by population growth? If the information is analyzed per capita among certain locations, how do these locations or areas differ in CO2 emissions?
- How did farmland (as a percentage of land in each country) change over the years and what is the impact population growth has brought in this aspect?
- What are the countries that present consistent negative GDP growth in the past years?
- How does the population of 0 to 14 year olds vary against the population aged between 15 and 64? How do populations metrics shift and is there a noticeable trend?

- To what extent is mortality rate correlated with CO2 emissions or the use of renewable energy?

# DATASET DESCRIPTION

As previously mentioned, The World Bank offers easy access to its publicly available data sets that contain information on a wide array of country indicators for each year where collection of that data was possible. These indicators cover different economic, political, geographical and social metrics that help paint a picture of countries all over the world.

Considering the aforementioned objectives and scope, for this report we select 15 indicators. These are:

- % Access sanitation facilities
- % Agricultural land
- % Population 0 to 14 years old
- % Population 15 to 64 years old
- % Rural population
- % Urban population
- Population growth
- Population total
- % Renewable electricity output
- CO2 emissions
- Energy use
- GDP
- GDP annual growth
- Surface area
- Mortality rate

We create a single data set using the following process:

- Merge all variables of interest into a single dataset.
- The missing values have been reduced, for this we have eliminated years if the country data was scarce.

In total, we obtain a dataset with 40,200 rows, on which the analysis is done.

# DATASET PREPARATION

To get the final data set we create scripts that do the following:

- Reading the `csv` files of all the 15 indicators
- Renaming variables
- Converting year into a column
- Concatenating all 15 indicators into one `data.frame`
- Checking for missing values
- Removing missing values (years and/or countries) to end up with a `data.frame` with no missing values

First we load the WDI, readr and dplyr libraries:

```
library(WDI)
library(readr)
library(dplyr)
library(tidyr)
```

The descriptions of these packages are as follows:

- `readr` - Read flat/tabular text files from disk (or a connection)
- `dplyr` - A Grammar of Data Manipulation. A fast, consistent tool for working with data frame like objects, both in memory and out of memory.
- `tidyr` - An evolution of 'reshape2'. It's designed specifically for data tidying (not general reshaping or aggregating) and works well with 'dplyr' data pipelines.

Next, each one of the indicators of interest are read into a data frame from the downloaded CSV files.

```
setwd("C:/Users/Ross DeWitt/Desktop")
san.df = read_csv("./data/%_access_sanitation_facilities_90_15.csv")
ag.df = read_csv("./data/%_agricultural_land_61_13.csv")
pop0.df = read_csv("./data/%_population_0_14_60_15.csv")
pop15.df = read_csv("./data/%_population_15_64_60_15.csv")
poprur.df= read_csv("./data/%_rural_population_60_15.csv")
popur.df= read_csv("./data/%_urban_population_60_15.csv")
co2.df= read_csv("./data/CO2_emissions_60_11.csv")
kmsqr.df= read_csv("./data/km_square_61_15.csv")
popgrowth.df= read_csv("./data/population_growth_60_15.csv")
poptotal.df= read_csv("./data/population_total_60_15.csv")
elecoutput.df= read_csv("./data/renew_elec_output_90_12.csv")
mort5.df= read_csv("./data/mortality_less5_rate.csv")
energy.df= read_csv("./data/energy_use_kg_oil.csv")
gdp_growth.df= read_csv("./data/GDP_growth_annual%.csv")
gdp.df= read_csv("./data/GDP.csv")
```

All data frames are now put into a list.

```
all.list <- list(san.df, ag.df, pop0.df, pop15.df, poprur.df, popur.df, co2.df, kmsqr.df,
                popgrowth.df, poptotal.df, elecoutput.df, mort5.df, energy.df,
                gdp_growth.df, gdp.df)
```

All data frames are put into a single `data.frame` called `wdi.df`.

```
wdi.df = data.frame()
```

We then loop the list to set year and value as a column, rename important variables and bind everything together.

```

for (aux.df in all.list){
  # Rename variables
  aux.df <- rename(aux.df, country.name = `Country Name`, country.code = `Country Code`, indicator.name = `Indicator Name`, indicator.code = `Indicator Code`)
  # "Gather" the year variables and values into a pair of variables: `year` and `value`
  aux.df <- gather(aux.df, year, value, -country.name, -country.code, -indicator.name, -indicator.code)
  # Print number of rows to check
  print(nrow(aux.df))
  # Concatenate
  wdi.df <- rbind(aux.df, wdi.df)
}

```

Convert the `indicator.code`, `country.code` variables into factors and the `year` variable into an ordered factor:

```

wdi.df$indicator.code = factor(wdi.df$indicator.code)
wdi.df$country.code = factor(wdi.df$country.code)
wdi.df$year = factor(wdi.df$year, ordered=TRUE)

```

We check final data.frame.

```
str(wdi.df)
```

```

## Classes 'tbl_df', 'tbl' and 'data.frame':    225720 obs. of  6 variables:
##  $ country.name   : chr  "Aruba" "Andorra" "Afghanistan" "Angola" ...
##  $ country.code   : Factor w/ 264 levels "ABW","AFG","AGO",...: 1 5 2 3 4 6 7
##  $ indicator.name: chr  "GDP (current US$)" "GDP (current US$)" "GDP (current US$)" "GDP (current US$)" ...
##  $ indicator.code: Factor w/ 15 levels "AG.LND.AGRI.ZS",...: 6 6 6 6 6 6 6 6 6 6 ...
##  $ year           : Ord.factor w/ 57 levels "1960"<"1961"<...: 1 1 1 1 1 1 1 1 1 1 ...
##  $ value          : chr  NA NA "537777811.911111" NA ...

```

The second part of the script removes a set of years and countries in order to have a final `data.frame` with no missing values. After studying the data we crafted a final `data.frame` that has 134 countries and 20 years (1992 -2011).

We remove column `year=x61` and `country.code=INX` with 100% missing values.

```

wdi.df <- filter(wdi.df, year!="X61")
wdi.df <- filter(wdi.df, country.code!="INX")

```

Next, we transfer the data into a new data frame named 'wdi.dfv2' for 20 years and 134 countries without any missing values.

```
wdi.dfv2 <- filter(wdi.df, year>1991& year<2012)
countries.mv <- wdi.dfv2 %>% group_by(country.code) %>% summarize(count=n(),
cnt.missing=sum(is.na(value)), pct.missing= 100*cnt.missing/count) %>%
  arrange(country.code) %>% select(country.code, pct.missing) %>% arrange(des
c(pct.missing))
```

We keep countries with 0% missing values and check for % missing values for each year and each country.

```
country_keep = (filter(countries.mv, pct.missing == 0))$country.code
wdi.dfv2 <- filter(wdi.dfv2, country.code %in% country_keep)
wdi.dfv2 %>% group_by(year) %>% summarize(count=n(), cnt.missing=sum(is.na(va
lue)), pct.missing= 100*cnt.missing/count) %>%
  arrange(year) %>% select(year, pct.missing) %>% print(n=100)
wdi.dfv2 %>% group_by(country.code) %>% summarize(count=n(), cnt.missing=sum
(is.na(value)), pct.missing= 100*cnt.missing/count) %>%
  arrange(country.code) %>% select(country.code, pct.missing) %>% arrange(des
c(pct.missing)) %>% print(n=250)
```

We save the final data frame wdi.dfv2 as a CSV file and then check the structure.

```
wdi.dfv2$country.code = factor(wdi.dfv2$country.code) #refactor
wdi.dfv2$year = factor(wdi.dfv2$year, ordered=TRUE) #refactor
write.csv(file="./data/wdi.df.csv", x=wdi.dfv2)
str(wdi.dfv2)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    40200 obs. of  6 variables:
## $ country.name : chr  "Angola" "Albania" "Arab World" "United Arab Emirate
s" ...
## $ country.code : Factor w/ 134 levels "AGO","ALB","ARB",...: 1 2 3 4 5 6 7
8 9 10 ...
## $ indicator.name: chr  "GDP (current US$)" "GDP (current US$)" "GDP (curren
t US$)" "GDP (current US$)" ...
## $ indicator.code: Factor w/ 15 levels "AG.LND.AGRI.ZS",...: 6 6 6 6 6 6 6 6
6 6 ...
## $ year          : Ord.factor w/ 20 levels "1992"<"1993"<...: 1 1 1 1 1 1 1
1 1 1 ...
## $ value         : chr  "10341398740.4707" "709452583.880319" "470374932070.
412" "54239171887.769" ...
```

## VARIABLE SUMMARIES

Now that we have a clean dataset to work with that contains the data we deem of interest we load it

from the file we saved previously and drop the column `index`.

```
wdi.df = read_csv("./data/wdi.df.csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
## Parsed with column specification:
## cols(
##   X1 = col_integer(),
##   country.name = col_character(),
##   country.code = col_character(),
##   indicator.name = col_character(),
##   indicator.code = col_character(),
##   year = col_integer(),
##   value = col_double()
## )
```

```
drops='X1'
wdi.df = wdi.df[, !(names(wdi.df) %in% drops)]
options(scipen=999) # disable scientific notation. Easier to read
```

First, let's check the number of unique values for each of our columns (year, code, country):

```
wdi.df %>% summarise(number_of_unique_years = n_distinct(year))
wdi.df %>% summarise(number_of_unique_ind = n_distinct(indicator.code))
wdi.df %>% summarise(number_of_unique_countries = n_distinct(country.code))
```

Next, we'll calculate the mean, median, standard deviation, minimum, and maximum of each indicator code. For clarity we'll also add the indicator name:

```
# get values summary for every indicator (mean, median, min, max, sd)
options(dplyr.width = Inf)
wdi.df %>% group_by(indicator.name, indicator.code) %>% summarise(mean=round(me
an(value), 2), median=round(median(value), 2), sd=round(sd(value), 2),
                                                                    max=round(max
(value), 2), min=round(min(value), 2)) %>% print(n=)
```

```
## Source: local data frame [15 x 7]
## Groups: indicator.name [?]
```

	indicator.name
	<chr>
## 1	Agricultural land (% of land area)
## 2	CO2 emissions (metric tons per capita)
## 3	Energy use (kg of oil equivalent per capita)
## 4	GDP (current US\$)
## 5	GDP growth (annual %)
## 6	Improved sanitation facilities (% of population with access)
## 7	Mortality rate, under-5 (per 1,000 live births)
## 8	Population ages 0-14 (% of total)
## 9	Population ages 15-64 (% of total)
## 10	Population growth (annual %)
## 11	Population, total
## 12	Renewable electricity output (% of total electricity output)
## 13	Rural population (% of total population)
## 14	Surface area (sq. km)
## 15	Urban population (% of total)

	indicator.code	mean	median	sd
	<chr>	<dbl>	<dbl>	<dbl>
## 1	AG.LND.AGRI.ZS	41.81	41.44	18.47
## 2	EN.ATM.CO2E.PC	4.55	2.56	5.02
## 3	EG.USE.PCAP.KG.OE	2023.20	1032.08	2248.32
## 4	NY.GDP.MKTP.CD	2353318971840.02	146213939568.08	6733032505830.54
## 5	NY.GDP.MKTP.KD.ZG	3.79	3.95	4.38
## 6	SH.STA.ACSN	69.15	75.49	28.11
## 7	SH.DYN.MORT	50.03	32.10	48.44
## 8	SP.POP.0014.TO.ZS	30.44	30.43	9.89
## 9	SP.POP.1564.TO.ZS	62.09	63.32	6.24
## 10	SP.POP.GROW	1.48	1.38	1.26
## 11	SP.POP.TOTL	467235649.28	29620939.00	1062067152.23
## 12	EG.ELC.RNEW.ZS	34.17	21.16	31.52
## 13	SP.RUR.TOTL.ZS	44.01	44.06	20.75
## 14	AG.SRF.TOTL.K2	9579820.75	754353.00	20516466.52
## 15	SP.URB.TOTL.IN.ZS	55.99	55.94	20.75

	max	min
	<dbl>	<dbl>
## 1	85.49	1.00
## 2	36.90	0.03
## 3	18157.60	119.41
## 4	72843138849548.91	709452583.88
## 5	33.74	-44.90
## 6	100.00	2.70
## 7	232.40	2.30
## 8	51.31	13.22
## 9	85.96	45.92



```
## 10          15.03          -2.66
## 11      7006907989.08      261057.00
## 12          100.00           0.00
## 13          90.42           0.00
## 14      134327660.80        680.00
## 15          100.00           9.59
```

We'll calculate the same statistics for each indicator code as above but this time we'll show the statistics for the first 5 years :

```
# get values summary for every year/indicator (mean, median, min, max, sd)
wdi.df %>% group_by(year, indicator.code ) %>% summarise(mean=round(mean(value), 2), median=round(median(value), 2), sd=round(sd(value), 2), max=round(max(value), 2), min=round(min(value), 2)) -> summary.year.code.df
for (y in unique(summary.year.code.df$year)[1:5]){summary.year.code.df %>% filter(year == y) %>% print() }
```

```
## Source: local data frame [15 x 7]
## Groups: year [1]
##
##   year   indicator.code      mean      median
##   <int>      <chr>      <dbl>      <dbl>
## 1  1992   AG.LND.AGRI.ZS      41.44      41.48
## 2  1992   AG.SRF.TOTL.K2  9587984.94  754353.00
## 3  1992   EG.ELC.RNEW.ZS      36.25      22.75
## 4  1992  EG.USE.PCAP.KG.OE    1905.57     991.40
## 5  1992   EN.ATM.CO2E.PC       4.46       2.30
## 6  1992   NY.GDP.MKTP.CD 1376563035542.83 62469186959.29
## 7  1992  NY.GDP.MKTP.KD.ZG       0.84       1.73
## 8  1992    SH.DYN.MORT      65.76      52.25
## 9  1992    SH.STA.ACSN      64.58      69.50
## 10 1992  SP.POP.0014.TO.ZS      33.67      35.37
## 11 1992  SP.POP.1564.TO.ZS      59.70      59.84
## 12 1992    SP.POP.GROW       1.80       1.78
## 13 1992    SP.POP.TOTL  407096348.54  25448099.50
## 14 1992    SP.RUR.TOTL.ZS      47.34      47.68
## 15 1992  SP.URB.TOTL.IN.ZS      52.66      52.32
##
##           sd           max           min
##           <dbl>         <dbl>         <dbl>
## 1           18.44          85.28          1.79
## 2    20603671.26    134160122.40    680.00
## 3           33.01         100.00          0.00
## 4    2087.12       11073.20     119.41
## 5           5.13          28.79          0.06
## 6  4050878585794.08 25374818246063.10 709452583.88
## 7           8.29          18.67     -44.90
## 8           57.38         232.40          5.90
## 9           30.01         100.00          2.70
## 10          9.78          51.31         15.77
## 11          6.20          72.57         45.92
## 12          1.23           5.64         -1.79
## 13   926475657.80   5453473436.00   261057.00
## 14          21.31          90.42          0.00
## 15          21.31         100.00          9.59
## Source: local data frame [15 x 7]
## Groups: year [1]
##
##   year   indicator.code      mean      median
##   <int>      <chr>      <dbl>      <dbl>
## 1  1993   AG.LND.AGRI.ZS      41.54      41.43
## 2  1993   AG.SRF.TOTL.K2  9578340.61  754353.00
## 3  1993   EG.ELC.RNEW.ZS      37.14      22.89
## 4  1993  EG.USE.PCAP.KG.OE    1900.01     944.61
## 5  1993   EN.ATM.CO2E.PC       4.50       2.28
## 6  1993   NY.GDP.MKTP.CD 1395593592012.71 66410016205.71
```

```

## 7 1993 NY.GDP.MKTP.KD.ZG 1.64 2.37
## 8 1993 SH.DYN.MORT 64.37 49.80
## 9 1993 SH.STA.ACSN 65.00 70.23
## 10 1993 SP.POP.0014.TO.ZS 33.39 35.03
## 11 1993 SP.POP.1564.TO.ZS 59.89 60.04
## 12 1993 SP.POP.GROW 1.73 1.74
## 13 1993 SP.POP.TOTL 413613478.82 26069537.50
## 14 1993 SP.RUR.TOTL.ZS 46.99 46.92
## 15 1993 SP.URB.TOTL.IN.ZS 53.01 53.08
## sd max min
## <dbl> <dbl> <dbl>
## 1 18.61 85.30 1.79
## 2 20586462.28 134042632.40 680.00
## 3 33.57 99.96 0.00
## 4 2120.27 11153.20 123.34
## 5 5.44 31.01 0.06
## 6 4089376471264.65 25805707315784.60 768401665.82
## 7 6.20 13.94 -29.30
## 8 56.80 227.70 5.70
## 9 29.81 100.00 2.80
## 10 9.74 51.02 15.48
## 11 6.12 72.10 46.08
## 12 1.25 5.54 -2.34
## 13 941254362.44 5537776461.00 263725.00
## 14 21.25 90.00 0.00
## 15 21.25 100.00 10.01
## Source: local data frame [15 x 7]
## Groups: year [1]
##
## year indicator.code mean median
## <int> <chr> <dbl> <dbl>
## 1 1994 AG.LND.AGRI.ZS 41.57 41.06
## 2 1994 AG.SRF.TOTL.K2 9577772.90 754353.00
## 3 1994 EG.ELC.RNEW.ZS 36.58 24.14
## 4 1994 EG.USE.PCAP.KG.OE 1903.16 963.94
## 5 1994 EN.ATM.CO2E.PC 4.50 2.31
## 6 1994 NY.GDP.MKTP.CD 1497585447128.82 78186608054.25
## 7 1994 NY.GDP.MKTP.KD.ZG 2.70 3.20
## 8 1994 SH.DYN.MORT 63.03 47.40
## 9 1994 SH.STA.ACSN 65.51 70.93
## 10 1994 SP.POP.0014.TO.ZS 33.09 34.58
## 11 1994 SP.POP.1564.TO.ZS 60.11 60.33
## 12 1994 SP.POP.GROW 1.65 1.72
## 13 1994 SP.POP.TOTL 420070547.52 26677874.50
## 14 1994 SP.RUR.TOTL.ZS 46.65 46.21
## 15 1994 SP.URB.TOTL.IN.ZS 53.35 53.79
## sd max min
## <dbl> <dbl> <dbl>
## 1 18.62 85.33 1.79

```

```
## 2      20585349.48      134035719.40      680.00
## 3          33.18          99.98          0.00
## 4      2172.95      11640.06      126.08
## 5          5.51          32.67          0.05
## 6  4378981199037.68 27706373746783.00 925817089.48
## 7          5.74          14.98      -30.90
## 8          56.26          225.50          5.20
## 9          29.59          100.00          2.90
## 10         9.69          50.64          15.23
## 11         6.04          71.67          46.32
## 12         1.24          5.06          -2.39
## 13    955937643.86    5621146521.00    266021.00
## 14         21.20          89.56          0.00
## 15         21.20          100.00          10.44
## Source: local data frame [15 x 7]
## Groups: year [1]
##
##      year      indicator.code      mean      median
##   <int>      <chr>      <dbl>      <dbl>
## 1  1995      AG.LND.AGRI.ZS      41.61      41.28
## 2  1995      AG.SRF.TOTL.K2    9577774.47    754353.00
## 3  1995      EG.ELC.RNEW.ZS      35.85      22.20
## 4  1995  EG.USE.PCAP.KG.OE    1903.63    953.99
## 5  1995      EN.ATM.CO2E.PC      4.40       2.34
## 6  1995      NY.GDP.MKTP.CD 1663301829980.82 96238031022.10
## 7  1995  NY.GDP.MKTP.KD.ZG      3.79       3.84
## 8  1995      SH.DYN.MORT      61.58      44.70
## 9  1995      SH.STA.ACSN      65.99      71.40
## 10 1995  SP.POP.0014.TO.ZS      32.76      33.98
## 11 1995  SP.POP.1564.TO.ZS      60.36      60.70
## 12 1995      SP.POP.GROW      1.58       1.67
## 13 1995      SP.POP.TOTL    426558251.75    27267462.00
## 14 1995      SP.RUR.TOTL.ZS      46.32      46.10
## 15 1995  SP.URB.TOTL.IN.ZS      53.68      53.90
##
##      sd      max      min
##   <dbl>      <dbl>      <dbl>
## 1      18.62      85.35      1.79
## 2    20585351.37    134035749.40    680.00
## 3      32.76      99.82       0.00
## 4    2153.56    11779.77    134.26
## 5       5.20     30.06       0.04
## 6  4836343001380.10 30658944469580.70 1309382885.33
## 7       3.57     13.30     -10.40
## 8      55.53     224.80       4.80
## 9      29.38     100.00       3.00
## 10      9.63     50.14     15.01
## 11      5.97     72.32     46.66
## 12      1.21      4.88     -2.66
## 13    970659003.66    5705843054.00    267468.00
```

```
## 14      21.14      89.11      0.00
## 15      21.14     100.00     10.89
## Source: local data frame [15 x 7]
## Groups: year [1]
##
##      year      indicator.code      mean      median
##    <int>      <chr>      <dbl>      <dbl>
## 1  1996      AG.LND.AGRI.ZS      41.69      40.86
## 2  1996      AG.SRF.TOTL.K2     9577778.73     754353.00
## 3  1996      EG.ELC.RNEW.ZS      35.81      22.63
## 4  1996      EG.USE.PCAP.KG.OE     1946.18      941.87
## 5  1996      EN.ATM.CO2E.PC        4.40        2.41
## 6  1996      NY.GDP.MKTP.CD 1711134882363.49 105314548999.29
## 7  1996      NY.GDP.MKTP.KD.ZG        4.65        4.53
## 8  1996      SH.DYN.MORT        60.07      42.87
## 9  1996      SH.STA.ACSN        66.46      71.69
## 10 1996      SP.POP.0014.TO.ZS      32.45      33.63
## 11 1996      SP.POP.1564.TO.ZS      60.57      61.18
## 12 1996      SP.POP.GROW         1.52        1.61
## 13 1996      SP.POP.TOTL     432958363.43     27836459.50
## 14 1996      SP.RUR.TOTL.ZS        46.00      45.63
## 15 1996      SP.URB.TOTL.IN.ZS      54.00      54.37
##
##      sd      max      min
##    <dbl>      <dbl>      <dbl>
## 1      18.60      85.37      1.79
## 2    20585359.81    134035829.40     680.00
## 3      32.86      99.97      0.00
## 4    2214.67    12674.14    132.20
## 5       4.89     26.94      0.06
## 6  4904507840082.99 31324994778851.40 1345719435.21
## 7       3.47     26.85     -5.20
## 8     54.71    224.00      4.50
## 9     29.17    100.00      3.00
## 10      9.63     50.10     14.83
## 11      5.94     72.66     46.79
## 12      1.14      4.88     -2.52
## 13    985256646.66    5788596142.00    268916.00
## 14      21.07     88.63      0.00
## 15      21.07    100.00     11.37
```

And now for every country:

```
# get values summary for every country/indicator (mean, median, min, max, sd)
wdi.df %>% group_by(country.name, indicator.code) %>% summarise(mean=round(mean
(value), 2), median=round(median(value), 2), sd=round(sd(value), 2),
                                                                    max=round(max
(value), 2), min=round(min(value), 2)) -> summary.country.code.df
for (c in unique(summary.country.code.df$country.name)[1:5]){summary.country.co
de.df %>% filter(country.name == c) %>% print() }
```

```
## Source: local data frame [15 x 7]
## Groups: country.name [1]
##
##   country.name  indicator.code      mean      median
##   <chr>         <chr>          <dbl>      <dbl>
## 1 Albania      AG.LND.AGRI.ZS      41.65      41.35
## 2 Albania      AG.SRF.TOTL.K2    28750.00    28750.00
## 3 Albania      EG.ELC.RNEW.ZS      97.46      97.29
## 4 Albania      EG.USE.PCAP.KG.OE    583.30     613.84
## 5 Albania      EN.ATM.CO2E.PC       1.09       1.14
## 6 Albania      NY.GDP.MKTP.CD 6046483215.79 4247918725.98
## 7 Albania      NY.GDP.MKTP.KD.ZG       5.46       5.84
## 8 Albania      SH.DYN.MORT      25.24      24.25
## 9 Albania      SH.STA.ACSN      84.59      84.60
## 10 Albania     SP.POP.0014.TO.ZS      28.22      29.28
## 11 Albania     SP.POP.1564.TO.ZS     63.81      63.24
## 12 Albania     SP.POP.GROW      -0.59      -0.62
## 13 Albania     SP.POP.TOTL    3067835.90   3055591.50
## 14 Albania     SP.RUR.TOTL.ZS      55.93      57.03
## 15 Albania     SP.URB.TOTL.IN.ZS     44.07      42.97
##
##           sd           max           min
##           <dbl>         <dbl>         <dbl>
## 1           1.16          43.84          39.31
## 2           0.00        28750.00      28750.00
## 3           1.53         100.00         95.18
## 4        134.18         770.34        384.59
## 5           0.40           1.61           0.49
## 6 4189124041.65 12890867538.53 709452583.88
## 7           5.77          13.50        -10.84
## 8           6.88          37.30         16.00
## 9           4.44          91.50         78.30
## 10          4.14          32.85         20.61
## 11          2.46          68.16         60.90
## 12          0.16          -0.28         -0.94
## 13 107545.38    3247039.00    2904780.00
## 14          5.09          62.75         46.75
## 15          5.09          53.25         37.25
## Source: local data frame [15 x 7]
## Groups: country.name [1]
##
##   country.name  indicator.code      mean      median
##   <chr>         <chr>          <dbl>      <dbl>
## 1 Algeria      AG.LND.AGRI.ZS      16.92      16.78
## 2 Algeria      AG.SRF.TOTL.K2    2381740.00    2381740.00
## 3 Algeria      EG.ELC.RNEW.ZS       0.78       0.80
## 4 Algeria      EG.USE.PCAP.KG.OE    938.64     894.10
## 5 Algeria      EN.ATM.CO2E.PC       3.08       3.03
## 6 Algeria      NY.GDP.MKTP.CD 85916133071.68 55775174454.63
```

```

## 7      Algeria NY.GDP.MKTP.KD.ZG          3.08          3.30
## 8      Algeria      SH.DYN.MORT          36.73          38.35
## 9      Algeria      SH.STA.ACSN          83.98          84.15
## 10     Algeria SP.POP.0014.TO.ZS          33.36          32.51
## 11     Algeria SP.POP.1564.TO.ZS          62.13          62.95
## 12     Algeria      SP.POP.GROW           1.62           1.55
## 13     Algeria      SP.POP.TOTL    31842900.10    31790353.50
## 14     Algeria      SP.RUR.TOTL.ZS          38.95          38.89
## 15     Algeria SP.URB.TOTL.IN.ZS          61.05          61.11
##              sd              max              min
##              <dbl>              <dbl>              <dbl>
## 1              0.37              17.38              16.32
## 2              0.00              2381740.00          2381740.00
## 3              0.42              1.82              0.21
## 4              117.88            1151.45            798.13
## 5              0.24              3.53              2.67
## 6    50087145375.42 200013050828.17 41764052457.88
## 7              2.20              7.20             -2.10
## 8              5.94              44.90            26.60
## 9              1.98              86.80            80.70
## 10             5.31              42.13            27.19
## 11             4.59              67.27            54.42
## 12             0.32              2.33             1.26
## 13    2792964.43    36717132.00    27180921.00
## 14             4.57              46.34            31.79
## 15             4.57              68.21            53.66
## Source: local data frame [15 x 7]
## Groups: country.name [1]
##
##   country.name  indicator.code      mean      median
##   <chr>         <chr>          <dbl>      <dbl>
## 1      Angola   AG.LND.AGRI.ZS      46.25      46.12
## 2      Angola   AG.SRF.TOTL.K2  1246700.00  1246700.00
## 3      Angola   EG.ELC.RNEW.ZS      76.58      76.99
## 4      Angola   EG.USE.PCAP.KG.OE   533.90     511.75
## 5      Angola   EN.ATM.CO2E.PC        0.86        0.78
## 6      Angola   NY.GDP.MKTP.CD 30236964440.63 11459124481.60
## 7      Angola   NY.GDP.MKTP.KD.ZG        5.65        4.73
## 8      Angola   SH.DYN.MORT      209.34     212.90
## 9      Angola   SH.STA.ACSN       34.43      34.15
## 10     Angola   SP.POP.0014.TO.ZS      48.02      48.00
## 11     Angola   SP.POP.1564.TO.ZS      49.56      49.55
## 12     Angola   SP.POP.GROW         3.24        3.32
## 13     Angola   SP.POP.TOTL   16284211.00  15836243.50
## 14     Angola   SP.RUR.TOTL.ZS       66.34      66.48
## 15     Angola   SP.URB.TOTL.IN.ZS      33.66      33.52
##              sd              max              min
##              <dbl>              <dbl>              <dbl>
## 1              0.33              47.32            45.96

```



```
## 2      0.00      1246700.00      1246700.00
## 3      11.04      93.75      62.09
## 4      49.26      638.64      483.20
## 5      0.38      1.43      0.31
## 6 32267171793.61 104115923082.74 4967585281.90
## 7      10.52      22.59      -24.70
## 8      15.49      226.00      177.30
## 9      7.93      47.60      23.10
## 10     0.15      48.27      47.75
## 11     0.12      49.78      49.36
## 12     0.25      3.56      2.76
## 13     3142553.72 21942296.00 11848971.00
## 14      4.38      73.13      59.10
## 15      4.38      40.90      26.87
## Source: local data frame [15 x 7]
## Groups: country.name [1]
##
##   country.name  indicator.code      mean      median
##   <chr>        <chr>          <dbl>      <dbl>
## 1 Arab World  AG.LND.AGRI.ZS      39.44      39.62
## 2 Arab World  AG.SRF.TOTL.K2 13750084.32 13781771.00
## 3 Arab World  EG.ELC.RNEW.ZS      4.55      4.58
## 4 Arab World  EG.USE.PCAP.KG.OE 1417.92     1368.51
## 5 Arab World  EN.ATM.CO2E.PC      3.88      3.73
## 6 Arab World  NY.GDP.MKTP.CD 1053066504425.37 730970977331.24
## 7 Arab World  NY.GDP.MKTP.KD.ZG      4.56      4.74
## 8 Arab World  SH.DYN.MORT      57.91     57.74
## 9 Arab World  SH.STA.ACSN      73.16     73.20
## 10 Arab World SP.POP.0014.TO.ZS    37.49     37.34
## 11 Arab World SP.POP.1564.TO.ZS    58.47     58.55
## 12 Arab World SP.POP.GROW      2.32      2.27
## 13 Arab World SP.POP.TOTL 293186770.60 290347194.50
## 14 Arab World SP.RUR.TOTL.ZS     46.39     46.51
## 15 Arab World SP.URB.TOTL.IN.ZS    53.61     53.49
##
##           sd           max           min
##           <dbl>         <dbl>         <dbl>
## 1           1.09          40.12          35.28
## 2        140579.88      13781821.00      13152835.50
## 3           0.69           5.77           3.40
## 4        192.01       1774.16       1148.14
## 5           0.49           4.76           3.17
## 6 638594721929.94 2497944765462.02 470374932070.41
## 7           1.98           9.74           1.50
## 8          10.50          75.34          41.52
## 9           3.71          78.68          67.52
## 10          3.03          42.35          33.32
## 11          2.91          62.50          53.89
## 12          0.27           3.14           1.86
## 13      39196969.53      361031820.00      232417246.00
```

```
## 14          1.77          49.21          43.37
## 15          1.77          56.63          50.79
## Source: local data frame [15 x 7]
## Groups: country.name [1]
##
##   country.name  indicator.code      mean      median
##   <chr>         <chr>          <dbl>      <dbl>
## 1   Argentina  AG.LND.AGRI.ZS      49.01      47.01
## 2   Argentina  AG.SRF.TOTL.K2  2780400.00  2780400.00
## 3   Argentina  EG.ELC.RNEW.ZS      34.58      34.20
## 4   Argentina  EG.USE.PCAP.KG.OE  1679.47    1655.64
## 5   Argentina  EN.ATM.CO2E.PC       3.93       3.80
## 6   Argentina  NY.GDP.MKTP.CD 285065226839.51 270423250000.00
## 7   Argentina  NY.GDP.MKTP.KD.ZG       4.16       5.87
## 8   Argentina  SH.DYN.MORT      19.45      19.45
## 9   Argentina  SH.STA.ACSN      91.86      91.90
## 10  Argentina  SP.POP.0014.TO.ZS     27.72      27.63
## 11  Argentina  SP.POP.1564.TO.ZS     62.34      62.37
## 12  Argentina  SP.POP.GROW       1.14       1.11
## 13  Argentina  SP.POP.TOTL  37682604.65  37680489.00
## 14  Argentina  SP.RUR.TOTL.ZS      10.60      10.58
## 15  Argentina  SP.URB.TOTL.IN.ZS     89.40      89.42
##
##           sd           max           min
##           <dbl>         <dbl>         <dbl>
## 1           2.82          54.18          46.68
## 2           0.00      2780400.00    2780400.00
## 3           5.58          43.56          26.03
## 4        152.87      1924.37      1469.01
## 5           0.44           4.74           3.25
## 6  105857432025.05  557890203658.12  97724513456.48
## 7           5.99          11.94         -10.89
## 8           3.60          25.70          14.20
## 9           2.15          95.20          88.30
## 10          1.34          30.18          25.70
## 11          1.00          63.80          60.55
## 12          0.11           1.38           1.03
## 13    2468969.97    41655616.00   33655149.00
## 14          1.12          12.46           8.87
## 15          1.12          91.13          87.54
```

## References

[1] Investopedia, <http://www.investopedia.com/articles/03/042303.asp>  
(<http://www.investopedia.com/articles/03/042303.asp>)

[2] Population Reference Bureau,  
<http://www.prb.org/Publications/Articles/2002/HowManyPeopleHaveEverLivedonEarth.aspx>  
(<http://www.prb.org/Publications/Articles/2002/HowManyPeopleHaveEverLivedonEarth.aspx>)