Description of Case Study 1

Introduction

Gross Domestic Product or GDP, measures the amount of goods and material that is produced by a country. This value is measured over the course of a year or quarter and is typically used as an indicator of the wealth of a nation. The amount people earned or spent are usually used in GDP calculations. While GDP is an excellent indicator of the overall productivity of a nation, it is possible that it falls short as a measure of quality of life, education, and overall intrinsic wealth. This study seeks to measure GDP against other measures such as income groups. We will seek to find out if there exists a disparity between GDP and income groups. This study will shed a light on whether the productivity of a nation is a good indication of the overall wealth of the individuals belonging to that nation.

Objectives

Throughout the course of this study, a series of 5 tasks and questions will be addressed. These questions are listed below.

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

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

What are the average GDP rankings for the “High income: OECD” and “High income: nonOECD” groups?

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

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

Downloading and cleaning the data

Step 1: Install and load necessary packages:

To start, load downloader, repmis and RCurl. Other packages will be added as necessary

Download and create tidy data sets

The steps for creating tidy data are listed below.

First the data is saved as CSV files. Next, the CSV’s are loaded into a data frame. Once the data is in a data frame, remove missing rows and columns. Next, it is wise to rename the columns using meaningful names.

It is important to scan your columns and make sure the data types are in the format needed for the analysis you intend to run. Convert any character data types to numeric and factors as you see fit. In the analysis below, I converted both Economy and Rank to numeric fields. Notice how I removed the commas from the Economy attributes prior to converting the data. The commas were being factored into the data conversion and produced incorrect results when converting to numeric data.

Source URL’s are <https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FGDP.csv>

<https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FEDSTATS_Country.csv>

Load education data into a data frame named educraw

The education data set has 234 rows and 31 columns. For our analysis, we do not need all 31 columns. To proceed with the analysis, the education data will be subset into a new dataframe called educ that contains only 2 columns. One column contains the country code that will be used to merge the data set with the gdp data set. The other column, “Income.Group”, will be used to create bins of data further along in our analysis.

A note about NA attributes

Though the education data set has a plethora of NA’s, we will not remove them using the complete.cases or na.omit statements. By omiting the extraneous columns in the data frame, we have effectively removed NA values without removing valuable rows that will be needed to merge our data later on in our analysis.

#Analysis

##Question 1:

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

##Answer:

The total number of records returned in the joined data set is 189. This number matches the total number of records in the gdp data frame and we thus have a perfect match using the country codes. A 100% join match indicates that our method for removing NA's by omitting extraneous columns was successful.

##Question 2:

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

country in the resulting data frame?

##Answer: St. Kitts and Nevis

To obtain this answer, it is necessary to convert the Economy variable to a numeric data type and then sort ascending so that the highest GDP is last in the list. Once sorted, the 13th record is St. Kitts and Nevis.

##Question 3:

What are the average GDP rankings for the "High income: OECD" and "High income:

nonOECD" groups?

##Answer:

Average Rank of High Income OECD group: 32.9667

Average Rank of High Income nonOECD group: 91.91304

Note: To obtain this answer, the data was subsetted into 2 data frames called HighIncome\_OECD and HighIncome\_nonOECD. Data containing only records with Income.Group equal to HighIncome: OECD was subset or queried and stored in the HighIncome\_OECD data frame. Data containing only records with Income.Group equal to HighIncome: nonOECD was subset or queried and stored in the HighIncome\_nonOECD data frame. Subsequently, the mean was taken of both data frames to obtain the answers 32.9667 and 91.91304.

##Question 4:

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

##Answer:

Looking at the graph below, it is apparent that there are 4 distinct outliers with very high GDP's. The GDP numbers drop sharply until around the 25th ranked country. The numbers continue to drop steadily until around the 50th ranked country and then the numbers stay relatively flat-lined until the last country.

The logical deduction that can be made from this graph is that there are only a few very wealthy nations in the world and about 20 fairly wealthy nations, however, the vast majority of nations have an extremely low GDP compared to the wealthy nations.

The inferences that could be made about this analysis is that the HighIncome: OECD have a much better GDP ranking than the HighIncome: nonOECD group. In fact, the average rank of HighIncome: OECD is nearly 3 times as better as HighIncome: nonOECD.

##Question 5:

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

How many countries are Lower middle income but among the 38 nations with highest

GDP?

##Answer:

5 Countries are in the lower middle income group and also in the highest ranked GDP.

The logical conclusion that can be made regarding this analysis is that even though a country has a high GDP, it does not guarantee that people dwelling in these nations have a high income.

## Conclusion:

The data in this study and analysis suggests that while GDP is a good indicator of the wealth of a country, it can overlook other indications of wealth, such as overall income. As seen in question 5, 5 of the highest ranked GDP nations are also in the lower middle income group. This suggests that the money earned or spent within these nations is not reflected upon the income of the general public. Further analysis is needed to determine causes for this apparent anomaly.

Another observation that can be made from the analysis in this study is that wealth amongst the nations of the world is in the hands of a minority of nations. The majority of nations have a very low GDP compared to the highest ranked GDP nations.

## References:

The World Bank Website: Description of income groups. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

Investopedia. What is GDP and why is it so important to economists and investors?

http://www.investopedia.com/ask/answers/199.asp