DSCT Assignment 2 Report

By Zheng Yi Tao

Note: Refer to the code for bigger version of the graphs. All the graphs are labelled with Figure 1, 2, 3 etc.…

Data Preparation

For the first part of this assignment, there was some data tidying and cleaning to be done.

Initial Cleaning:

1) Used sapply function to check for any missing records. This dataset (WEF_GlobalCompetitivenessReport_2014-15) does not have missing records.
2) Next, I removed the rows of data that will not be used for this assignment to reduce the size of the data. The rows were removed by its Series.code value.
3) I found that some records has values "<0.1" which is a non-numeric value and I have decided to change it to 0.
    a. Rationale is that these values were found under HIV Prevalence (4.05) and is represented as a percentage. Thus, would not make sense if the percentage is negative.

Using Tidyr Package:

1) The goal of tidying these dataset is to make sure that all attributes of the dataset are represented in the columns and not the row – while also vice versa – that the value have to be in the rows and not the columns.
2) The function "gather" is used to bring all countries from multiple columns into just one columns called "Country" and the values to one column called "Elements".
3) Next, using "spread", the column Attribute is split into "Rank" and "Value".
4) These steps has reduced the number of columns and made the data more readable.

More Cleaning:

1) All the values has been converted to 2 decimal places.
2) Series.code and Edition columns are changed to factor.
3) Edition has its factor levels renamed to one year (e.g.: 2006) instead of a range of year (e.g.: 2006-2007).
4) Rank has been changed to numeric from character.

Question 1

What are the trends that can be witnessed in the domestic and foreign market size indices?

There were 2 line graph created where one is the domestic market size indices and another is the foreign market size indices.

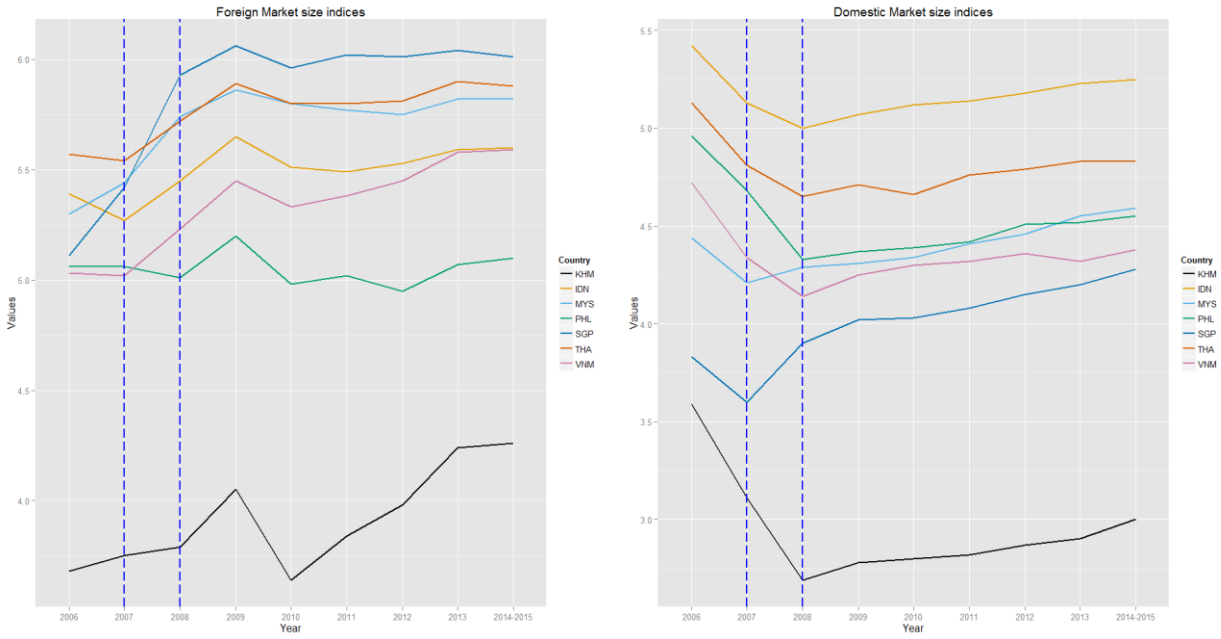

Figure 0-1.1

Figure 1.1 shows the comparison between both markets, more specifically between the year 2007 and 2008 as marked by the 2 dashed lines on each line graph.

Looking at the Domestic Market, there was a significant drop in the market size indices for most of the countries except Singapore and Malaysia between 2007 and 2008. The Foreign Market on the other hand shows an increase in the market size indices except Philippines.

This shows that in the period between 2007 and 2008, there has been a focus shift from Domestic Market (importing of goods and services) to Foreign Market (exporting of goods and services).

One country to note is Singapore who not only have both increase of market size on importing and exporting, there seems to be a sharp increase greater than other countries listed here in exporting. This has brought Singapore up to the highest market size indices in just the span of one year. Also, for the increase of import rather than decrease, could be because of the lack of natural resources in Singapore that there is still a high demand for the importing goods and services.

Vietnam seems to have much smaller market sizes as compared to the rest of the countries listed here.

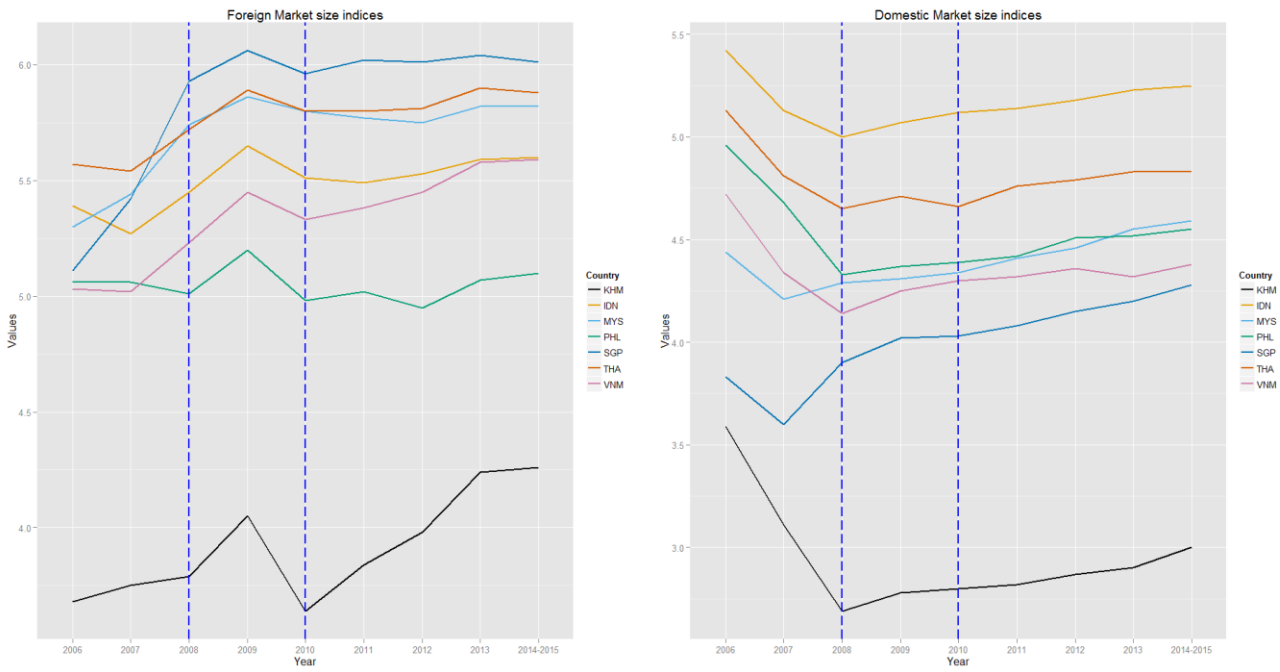

Figure 1.0-2

Figure 1.2 shows the same graph but with the dashed line placed between the year 2008 and 2010.

As seen here, for both Domestic and Foreign market, there was a steady increase from 2008 to 2009. However, after 2009, the market sizes for both markets dropped especially for the Foreign Market which dropped more significantly. This seems to be related to the stock market crash that happened in 2009 which causes such a sharp drop in the market size.

However, looking from the year 2010 onwards to 2015, there has been an increase for both markets with Vietnam increasing the most in the foreign market. This shows the recovery of both markets for all the countries during these 5 years.

Question 2

How does the quality of the education system affect the GDP?

For this question, both series “GDP (US$ Billions)” and “GDP per capita” were used to be compared to see if it affects “Quality of the Education System”.

There was an **initial problem** with using both the GDPs because of the highly skewed values of the data. For example in GDP (US$ Billions), the values can range from smallest $10.8 billion to $511765 billion. There was a possibility also, that the highly skewed data can be due to either formatting error or just input error. Looking at Cambodia in the GDP (US$ Billions) series, the values dropped from $11182 billion to $10.8 billion in the span of 1 year from 2009 to 2010 as shown below (Table 1):

Edition	Series.code	Series	Country	Rank	Value
2008	0.01	GDP (US$ billions)	KHM	107	8604.00
2009	0.01	GDP (US$ billions)	KHM	109	11182.00
2010	0.01	GDP (US$ billions)	KHM	108	10.80
2011	0.01	GDP (US$ billions)	KHM	111	11.63
2012	0.01	GDP (US$ billions)	KHM	108	12.86
2013	0.01	GDP (US$ billions)	KHM	110	14.24

Table 1

This same issue also applies to other countries especially between 2009 and 2010 thus making the data questionable. I have then decided to use Rank instead of Value for both series on GDP as data.

Consistency of GDP per capital and GDP (US$ Billions)

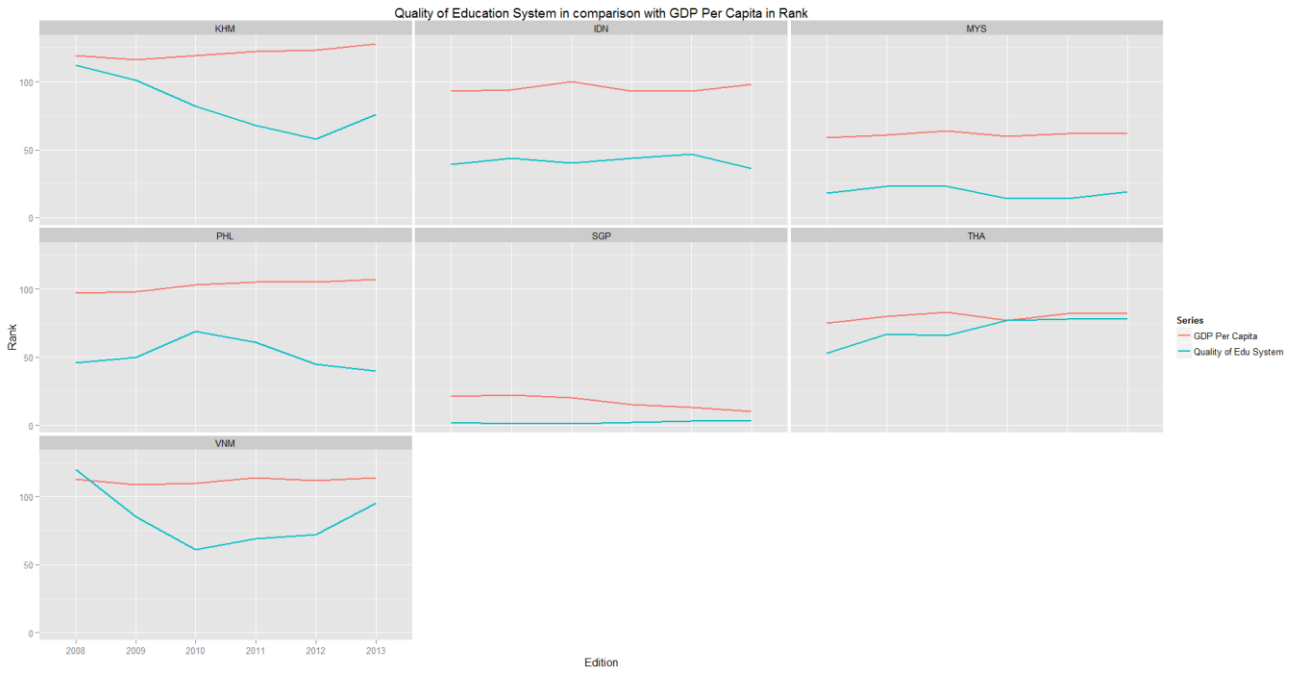

Figure 2.1

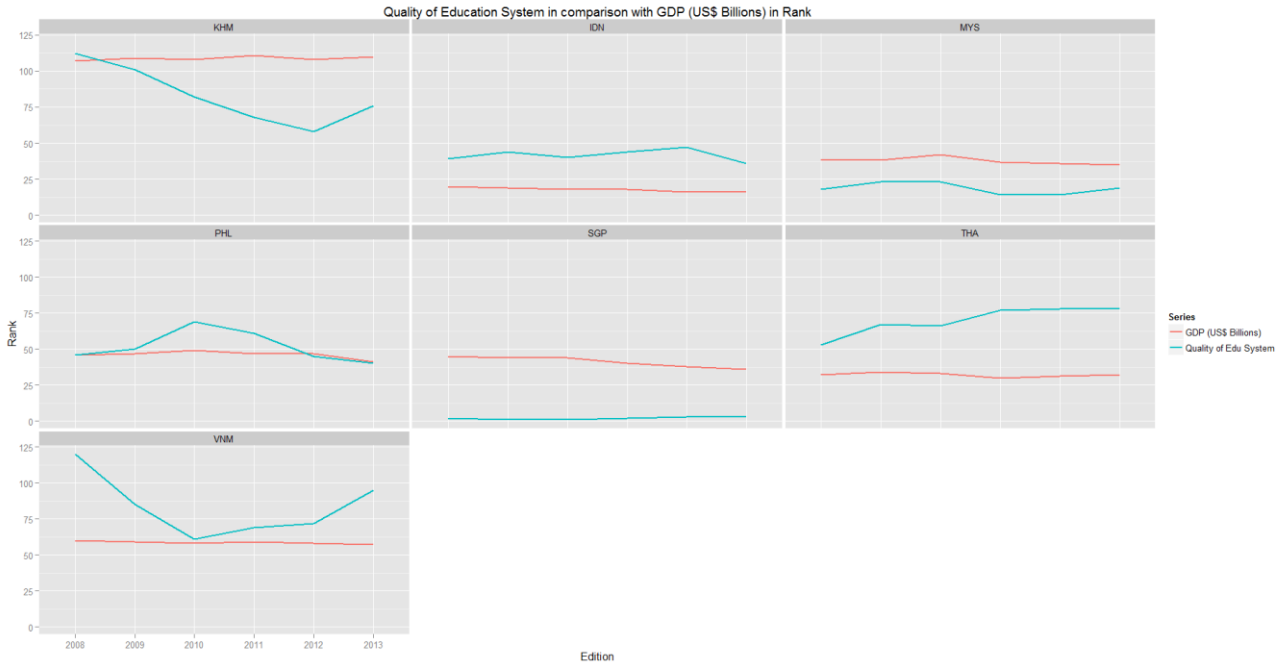

Figure 2.2

From both Figure 2.1 and 2.2, we can see that the GDP either per capita or in US$ billions are consistent with the ranking in the sense that the rankings does not fluctuate too much. This can be seen with the near horizontal red line showing the values quite consistent. On the other hand, we can see that the quality of the Education System (green line) fluctuates very much for certain countries especially Vietnam, Philippines as well as Cambodia.

While the GDP remains consistent, the quality of education system fluctuates a lot. This shows that there are little correlation between GDP and quality of the education system.

Comparison of quality of education and GDP among the countries

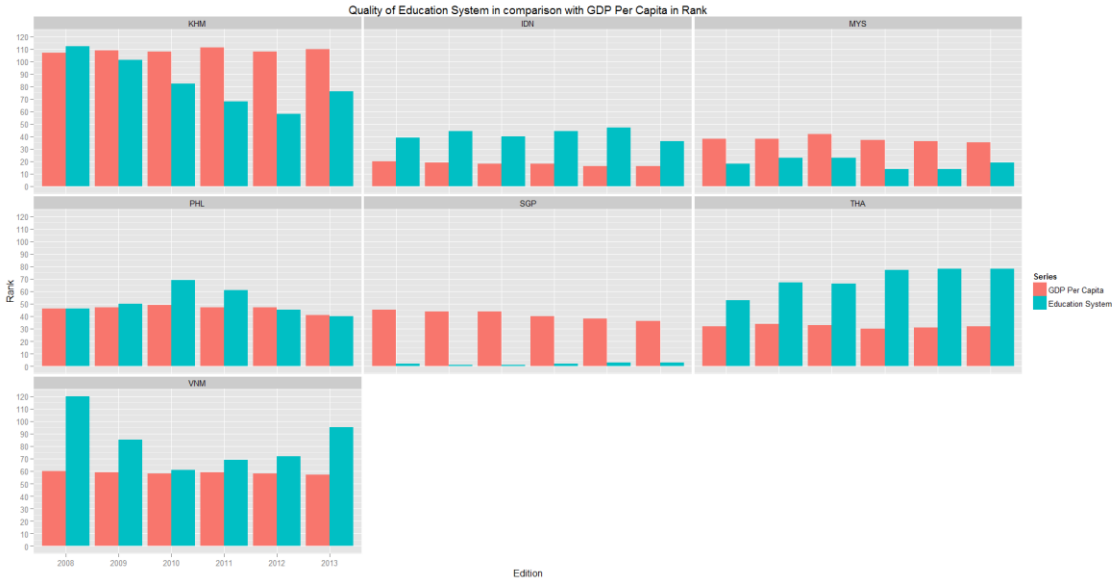

Figure 2.3

Figure 2.4

From Figure 2.3 and 2.4, we can see that Singapore has the highest quality of education (lower rank means higher quality). There were not much trend in Figure 2.3 other than to show that Indonesia has the highest GDP per capita while Cambodia has the lowest GDP per capita.

Figure 2.4 on the other hand, shows that Singapore has the highest GDP (US$ Billions) as well as the highest quality in education system. While Cambodia and Vietnam has the lowest GDP (US$ Billions), their quality of education are similarly lower than the other countries, even though not very significantly.

Conclusion

Due to the limitation of only using Rank as a measurement of whether a country is having high or low GDP and quality of education system, the comparison might not be fully reliable. This is because a drop in rank of one country does not necessarily means they are doing worst, but it could rather mean that another country is actually doing much better than this country.

However with the results, we can see that there are no obvious correlation between GDP and the quality of education as seen in Figure 2.1 and 2.2. As for Figure 2.4, there was some consistency where country with high GDP have high quality of education, we cannot seems to find a convincing trend of the opposite, that the country with low GDP have low quality of education.

Question 3

Does lower level of education lead to more HIV prevalence for year 2006 and 2015?

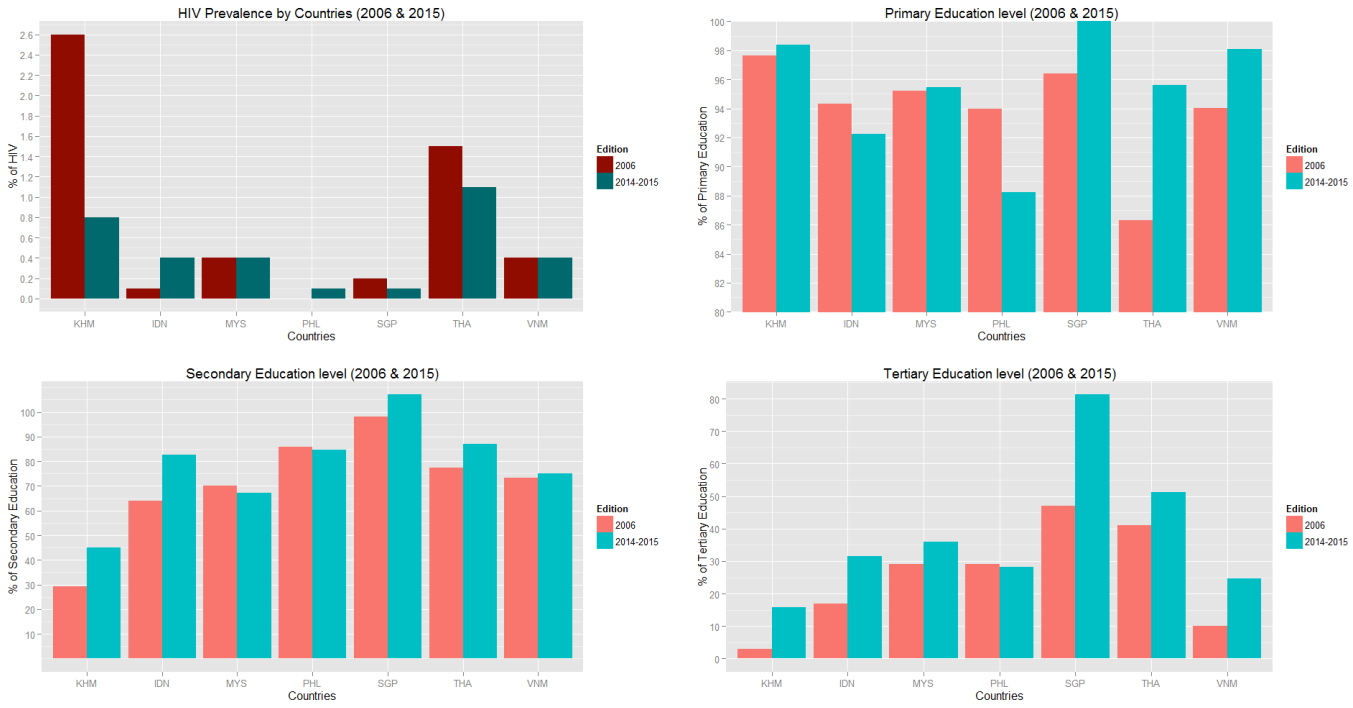

Figure 3.1

Figure 3.1 shows the bar plot of the percentage of HIV Prevalence as well as the bar plots of all the different level of educations. These plots compares between the year 2006 and 2015 among the different countries.

At first look, Cambodia have the highest percentage of HIV Prevalence. There was a sharp drop over the span of 9 years. As we can see from the education levels, there was an increase from 2006-2015. This shows that the raising of education level will result in the drop of HIV Prevalence.

The same result applies to other countries like Malaysia, Singapore, Thailand and Vietnam where the increase of education levels leads to the decrease in HIV Prevalence. Similarly in Philippines, but the other way round, it shows that they have a decrease in education levels which leads to the increase in HIV Prevalence.

Indonesia however, shows that the increase in education levels leads also to the increase in HIV Prevalence. This seems to be the only one inconsistent with the rest. But looking at Primary Education level alone, we see that there is actually a drop in percentage of people who study Primary education. This unique example could show that Primary education has more correlation to HIV Prevalence than other education levels.

So in **summary**, there is a consistency that more people in the higher level of education leads to lower percentage of HIV Prevalence while Primary education level seems to be more correlated than other education level with HIV Prevalence.

Question 4

Does irregular payments and bribes as well as diversion of public funds lead to lower / higher reliability of police services for the last 5 years?

Data Preparation

For the data used was from 2010 to 2015 because the question mentioned the “last 5 years”.

As for the irregular payments of bribes and diversion of public funds, I have decided to combine both together using their mean value to create a one to one comparison with the high reliability of police services. This can be done because both irregular payments and bribes together with diversion of public funds uses the same scale of 1-7 where 1 is the worst and 7 is the best. This can make the data comparison between 2 sets of data easier as compared to 3 sets of data.

Comparison of both series

Figure 4.1

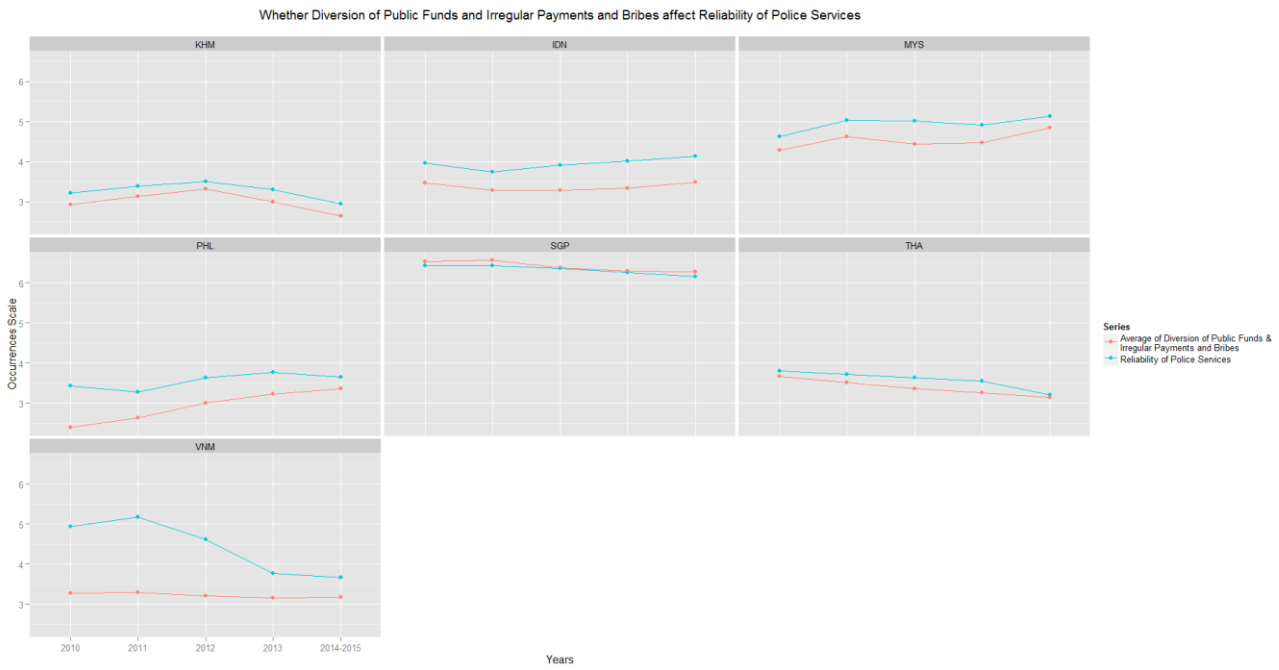

Figure 4.2

Both Figure 4.1 and 4.2 shows that there is a consistency that the having a high score irregular payments, bribes and diversion of public funds (green line) leads to the high score of reliability of the police services (red line) especially for Singapore. And similar for the opposite as seen in Thailand. Vietnam seems to be much more different than from the rest, that the despite having higher score (which means better) of irregular payments, bribes and diversion of public funds in 2010 and 2011, the reliability of the police services still remains low. It remained low even until 2015 as the green line gets lower (worst).

This shows that consistently, the data are correlated to each other, that even for Vietnam, a consistent low reliability of the police services will end up getting worst in terms of irregular payments and bribes as well as the diversion of public funds over time.

Question 5

How has inflation and the soundness of banks affected the average income per person for each country?

For the data of the Soundness of Bank, it is measured with the score 1 to 7 where 1 is worst and 7 is better. Where 1 means that the banks may require recapitalization and 7 means the banks are generally healthy with sound balance sheets.

Inflation on the other hand, has the highest score of 7 if the inflation percentage falls within the range of 0.5 to 2.9 percent. When the percentage gets further away from the range, the score of 7 will start to decrease depending on the distance away from the range. This is shown in Figure 5.1 below.

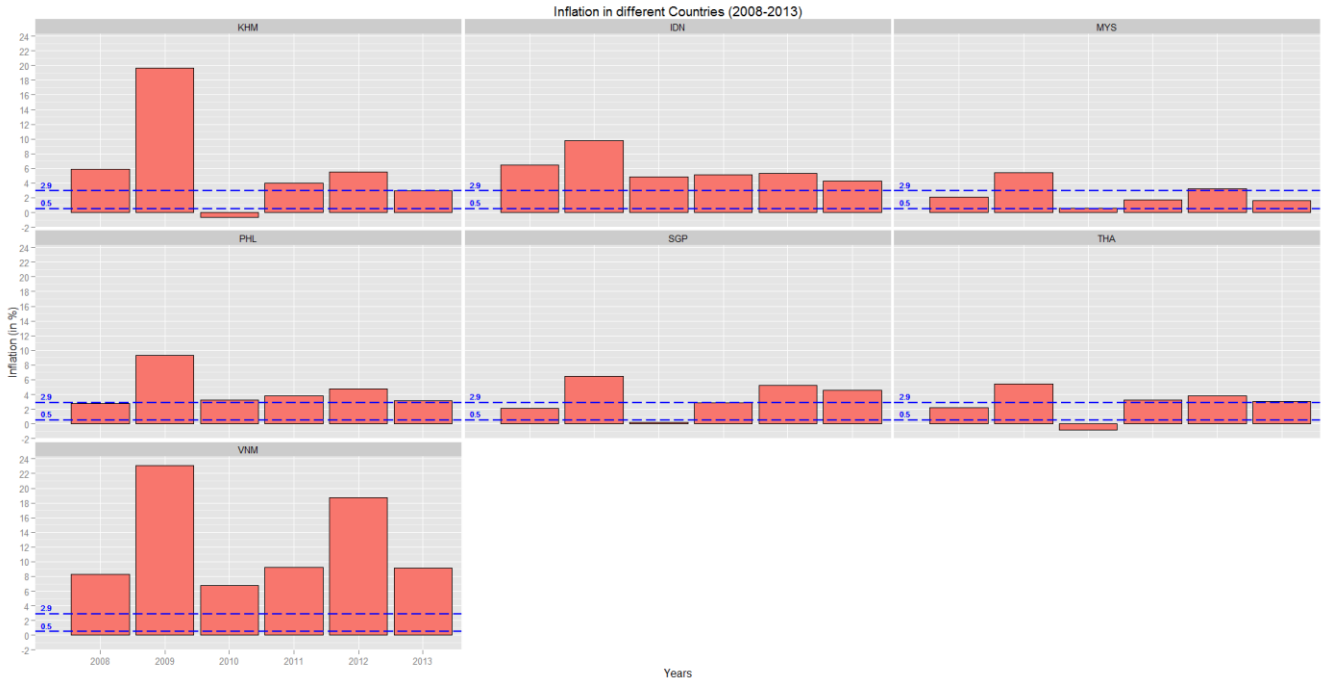

Figure 5.1

From Figure 5.1, we can see the 2 dashed line representing the range of 0.5 to 2.9 percent and that many of the countries do not remain within the range for more than a year. Some has even fallen below the range. But one outstanding observation of this plot is that in the year of 2009, there has been a surge in inflation rate for all the countries where Vietnam was the hardest hit, followed by Cambodia. This could be due to the crash in stock market in 2009.

Figure 5.2

Figure 5.2 shows the comparison between inflation and the GDP between countries. With the same reason in question 1, GDP per capita has been changed to Rank instead of Value, which means that a fair comparison would require inflation to be rank as well. Because rank was used, there is no significant changes in the year 2009 as when using values. This is because the crash in stock market affected all the countries together.

From the bar plot we can see that GDP per capita has little correlation with the inflation rate ranking of the countries. For example, despite the fluctuation of the inflation rate ranking between the countries, the GDP per capita remains steady with only a small value of increase or decrease over the span of the couple of years.

For example in Singapore, during the year 2012 and 2013, there has been a surge in inflation rate ranking but the GDP per capita remained. Another example is Cambodia in the year 2010, there was a huge drop in inflation rate but the GDP per capita still remained.

Figure 5.3

Next, Figure 5.3 shows the comparison between the soundness of Bank as well as the GDP per capita of the countries. There seems to be a trend where the lower rank of the soundness of Bank leads to the lower of rank in the GDP per capita as seen especially in Singapore. While Vietnam shows the opposite for higher rank. However, other countries like Philippines, Thailand and Malaysia showed that the low rank of the soundness of bank does not necessarily mean the lower rank of GDP per capita. Thus shows that the trend even though exist, is not a strong correlation.

Conclusion

In conclusion, both Inflation and the Soundness of Bank does not affect the average income per person in the country even though the soundness of bank has some influence, but are not strong enough. This could be due to the problem of calculating average income per person, where those who are extremely rich does not get affected much by the inflation but yet pulled up the average income of the entire country, making it not a suitable gauge of a country's wealth.