Group Project Paper I

C. Morgenstern, E. Weldon, *Data Science students at Lewis University*

***Abstract*—This project paper was written for the group assignment of week 2/3 as part of the course 70-530 Data Visualization at Lewis University in fall term 2, 2019. It is a report that summarizes results from the analysis of three chosen data sets using the Tableau Desktop 2019.3 software. The data sets were taken from the United Nations Development Programme (UNDP) website [1] and deal with the unemployment, gross domestic product and life expectancy data of different countries. This report aims at establishing meaningful relationships between the different metrics.**

I. Description of Data Sets

The data sets were taken from the United Nations Development Programme (UNDP) website [1] which has published human development data from 1990 until 2017. We decided on three data sets: gross domestic product (GDP) per capita (2011 PPP $), unemployment, total (% of labour force) and life expectancy at birth (years). The data sets were downloaded as comma separated files and explored and processed using Python and its Pandas library. Each data set contains values for the mentioned indicators of GDP, unemployment rate and life expectancy distributed over a certain period of time for different countries. Additionally, each country is assigned a Human Development Index (HDI) Rank (2017) leading to a ranking of countries from 1 to 189. The data is organized in columns with the headings HDI Rank (2017), Country and followed by numerous columns naming the different years.

II. Data Preprocessing

The exploratory data analysis showed that all three data sets had additional columns without any content in between columns with values. Also, there were rows and columns with missing values in all three data sets. We used Python and its Pandas library to load, explore and clean the data sets. This involved in the first step the removal of unnamed columns (Unnamed: ) without any data using a filter operation and Python´s regex library. In the next step we checked if there are further missing values in the data sets and dropped rows to exclude those countries. In order to merge the three data sets and to conduct meaningful analysis, we inspected the different columns representing the years where the data have been recorded. All three data sets start at the beginning of the 90s but had different yearly distributions. The following years: 1995, 2000, 2005, 2010, 2011, 2012, 2013, 2014, 2015, 2016 and 2017 were present in all three data sets and thus we deleted columns not included in this list from the other tables. Following these cleaning steps, the three tables were merged using Python´s concatenation function concat. The resulting joined data frame was exported as comma separated file to serve as basis for generating data visualizations in Tableau software.

III. Data Analysis Process

The cleaned and merged table comprising the data of 189 countries over a time period of 22 years (from 1995 until 2018) and representing values of unemployment, GDP and life expectancy were imported into Tableau Desktop 2019.3 using the Connect pane and the “To a file” – “Text file” selection. Individual graphics were designed on the sheet pages taken into account the learnings of week 2 and 3 and then exported as .png for incorporation into this paper.

The questions that should be answered by analyzing this data set are as follows:

* How do different countries rank according to the Human Development Index 2017?
* What are the wealthiest economies according to GDP in 2017?
* How do unemployment rates compare in 2017?
* Is there an improvement in life expectancy over time for different countries?
* Is there a correlation between GDP and life expectancy i.e. are lower income countries prone to a lower life expectancy or are higher income countries advantageous in this regard?

IV. Results

*A. Analysis of HDI Ranking*

The Human Development Index (HDI) has been implemented by the United Nations Development Programme in order to assess the development of a country that is not based on sole economic growth. It is a summary measure that takes into account the following dimensions of human development: life expectancy, education and per capita income. The HDI value represents the geometric mean of normalized indices for each of the three dimensions [2]. Countries rank higher in this list, and thus have a higher HDI if the lifespan is higher, the educational level is higher and the gross national income per capita is higher. Figure 1 lists the top and bottom 10 countries according to their HDI ranking in 2017. In the top 10, we find mostly European countries, with Norway in the lead, but also the Asian city states Hong Kong and Singapore as well as Australia. The bottom 10 countries are all African countries suggesting that lifespan, educational level and gross national income is the lowest there (see Fig. 1).

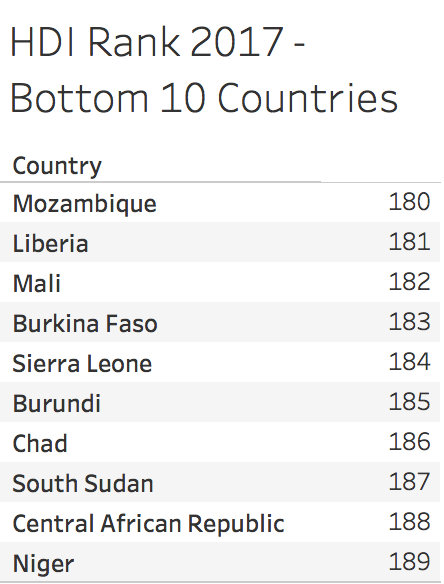
**

Fig. 1. Top 10 and bottom 10 countries according to HDI Rank (2017).

*B. Gross Domestic Product*

Gross Domestic Product (GDP) is the total monetary or market value of all the finished goods and services produced within a country within a specific time period. The value is a broad measure of the overall domestic production and thus it is often seen as a major indicator of a country's economic health [3]. The “healthiest” and most productive economies have high values of GDP and rank in the top of the HDI Rank. A colored geographic map marks out the big players in this field with the dark blue countries being closer to the top (see Fig. 2). The leading economies according to the GDP per capita in 2017 are European and Asian countries such as Qatar, Luxembourg, and Singapore.

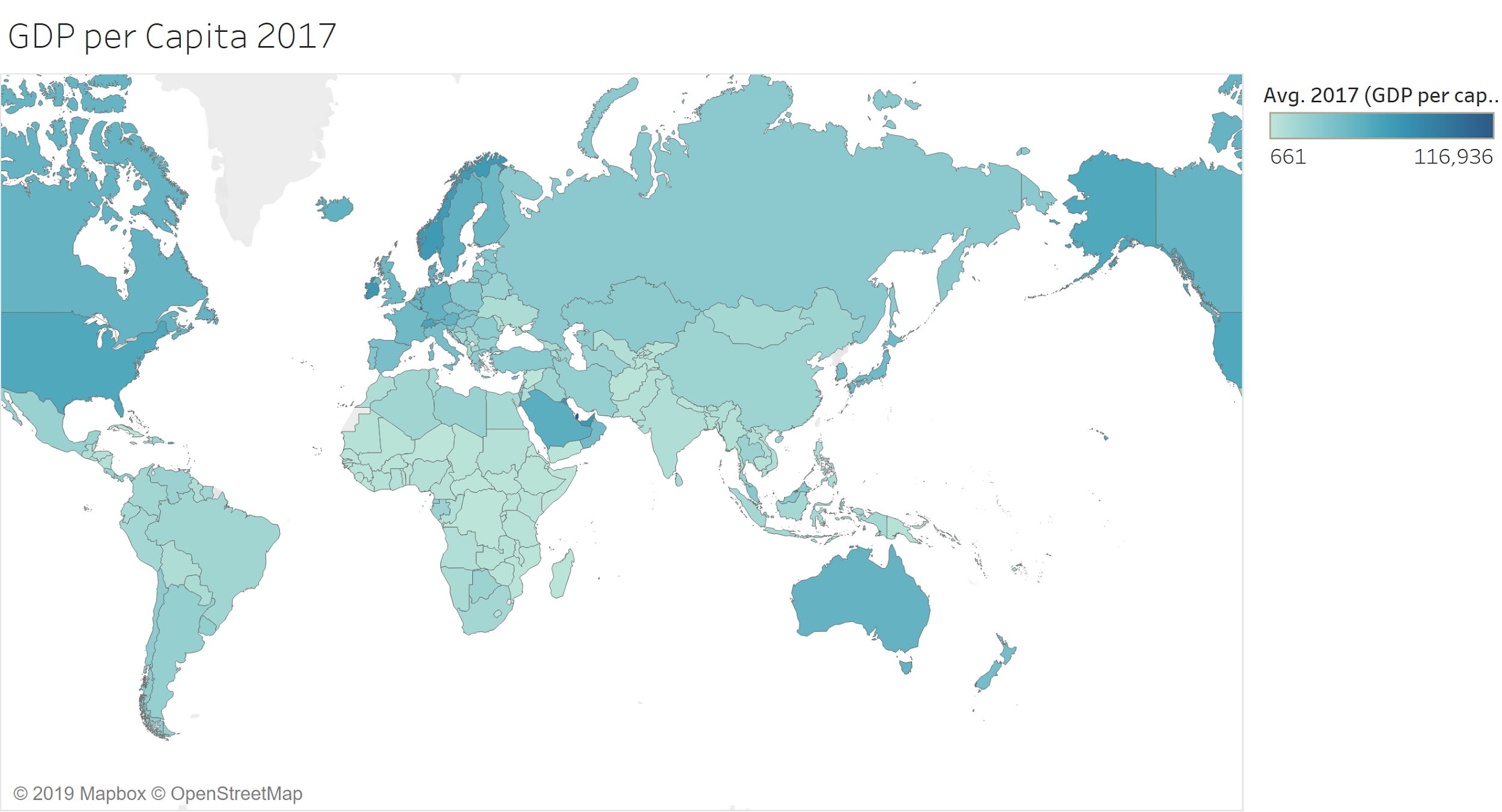
**

Fig. 2. Geographic map showing the country's GDP per capita for 2017. The intensity of the blue color indicates the average value of GDP with darker colors indicating wealthier economies.

*C. The Burden of Unemployment*

Unemployment is a burden, not only for the individual but also for the society and the country. We visualized the unemployment rate for the year 2017 using a geographic map and bubbles reflecting this value in terms of their size. Thus, countries with a higher unemployment rate have a bigger bubble as compared to countries with a lower unemployment rate (see Fig. 3). As seen in previous analyses, the trend continues in this graphic and shows African countries with high unemployment rates, whereas European and North American countries have smaller rates.

**

Fig. 3. Unemployment rate for 2017. The size of the circle is proportional to the unemployment rate i.e. countries with a small unemployment rate have a small spot.

*D. Life Expectancy over Time*

The statistical measure of the average life span of an organism is termed life expectancy. It takes into account the year the individual is born, the current age as well as other demographic factors like gender. Whereas in Bronze Age the life expectancy at birth was 26 years, in 2010 this value has risen to 67.2 years [4]. However, not all countries benefit from an ever-increasing average life span. We were interested to investigate the life expectancy at birth of different countries between 1995 and 2017, the start and end of our data set to visualize this point (Fig. 4 and 5). As seen in Figures 4 and 5, the average life expectancy is increasing between 1995 and 2017, both in the top 10 as well as in the bottom 10 countries.

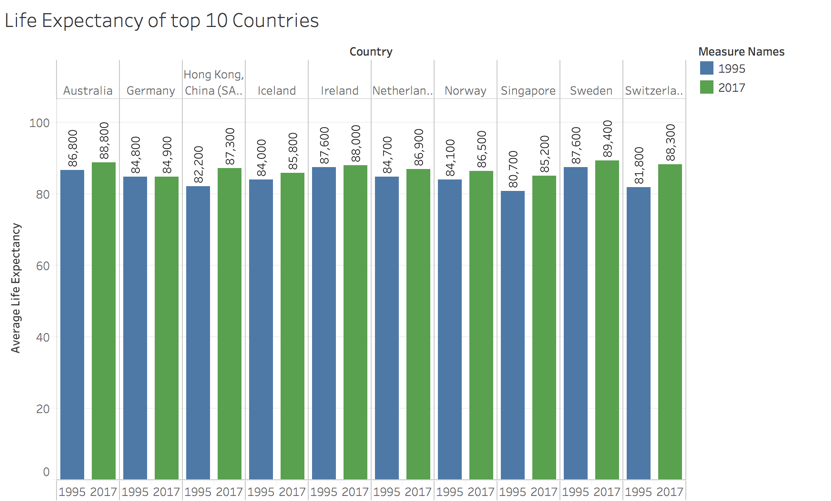


Fig. 4. Predicted life expectancy at birth in 1995 and 2017 in the top 10 countries as ranked by HDI 2017.

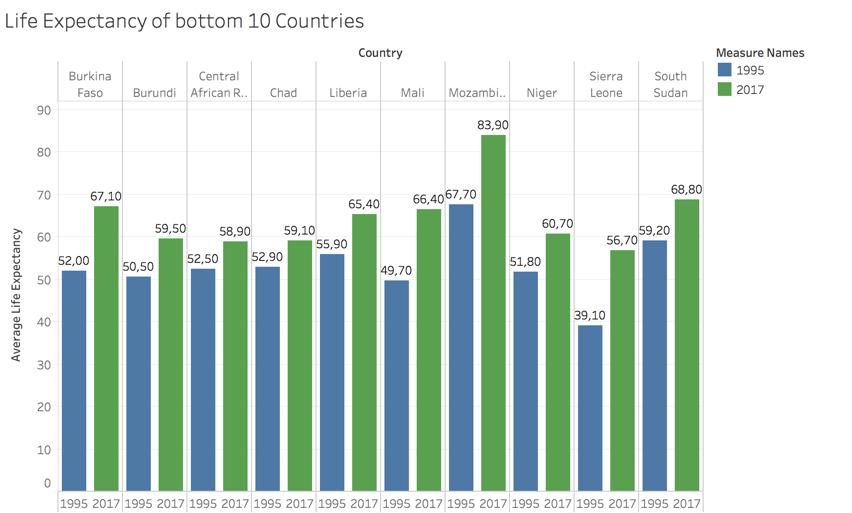


Fig. 5. Predicted life expectancy at birth in 1995 and 2017 in the bottom 10 countries as ranked by HDI 2017.

Whereas life expectancy in the top 10 countries hasn´t dramatically changed in these 22 years and is on average 85 years, the life expectancy in the bottom 10 countries has increased dramatically by about 15 years from 1995 to 2017. People in Mozambique are the only country in this list that have a life expectancy with 83 years and that compares to the top 10 countries. Dramatic increases in life expectancy have also seen people living in Mali and Sierra Leone with 17 and 16 added years of life, respectively.

*E. Analysis of GDP vs. Life Expectancy at Birth*

Life expectancy at birth in this data set is an average value assigned to each country per year which describes the age from birth to which the average person is expected to live to. Since the assumption is that the healthiest and most productive economies have high values of GDP, the following correlation matrix in Figure 6 shows the relationship between GDP and life expectancy at birth. Observing either the trend lines in the upper right quadrant or the lower left quadrant provides the reader with an R2 value of 0.431547. This signifies that these correlation models account for a small positive correlation between GDP and life expectancy, but the value is not strong enough to suggest that GDP directly indicates a longer life expectancy.

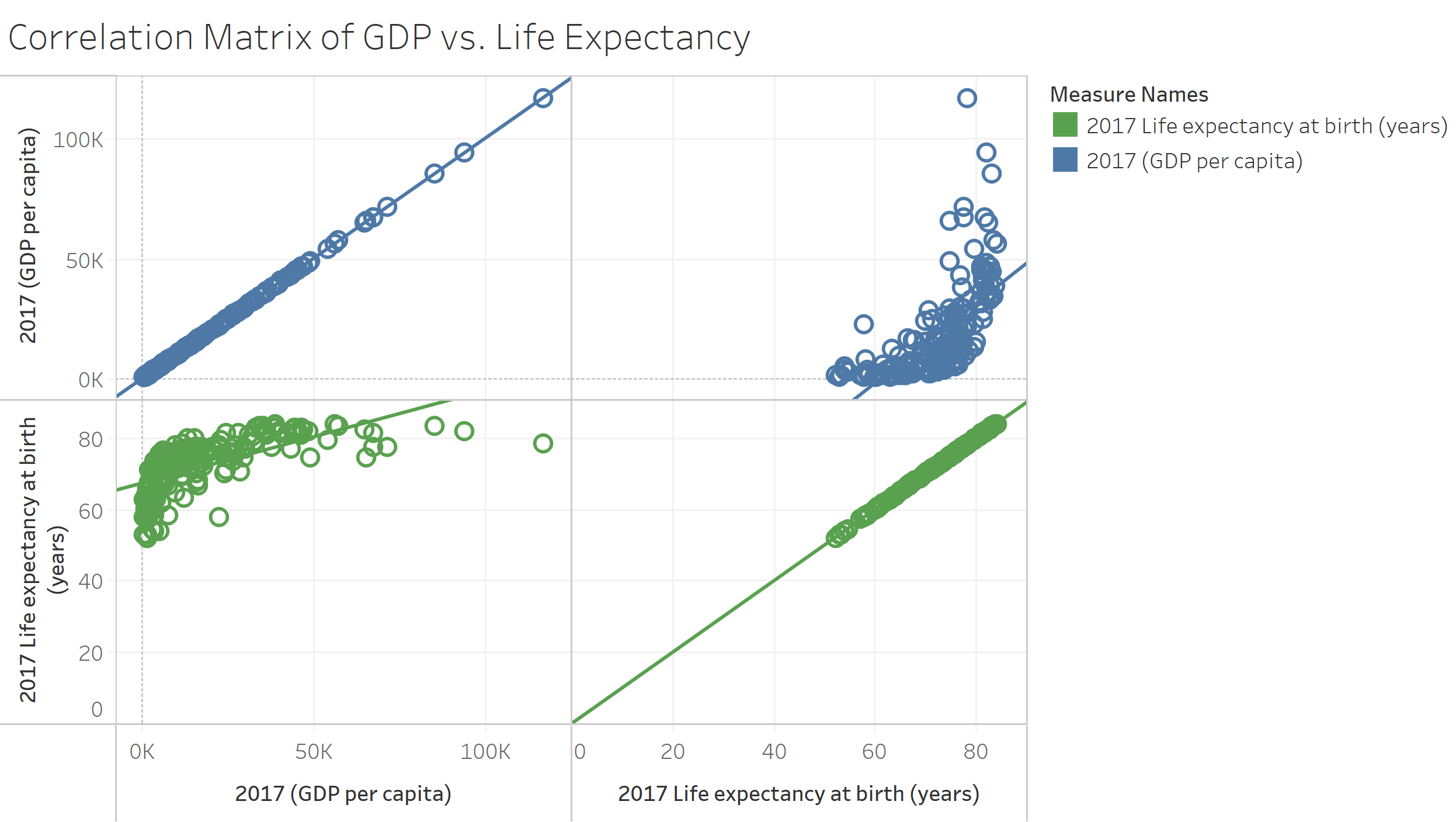


Fig. 6. Correlation matrix of GDP per capita values and life expectancy at birth with associated trend lines.

V. Conclusions

In this paper, we take a group approach to analyze three different data sets and make conclusions about them. We have been able to find interesting data sets, set up a cleaning script in Python as well as merged them and created interesting visualizations in Tableau.

The take home message from the analysis is that there are still countries, mainly African countries, that are lagging behind the major European and American economies with high unemployment rates and low GDP values. Although, life expectancy is increasing in these countries, policies must be taken to keep people in employment. Since there is a relationship between employment and economic growth, these countries must take action to tackle the high unemployment rates to prevent further losses in the country´s gross domestic product.

References

*Weblinks*

[1] <http://hdr.undp.org/en/data>

[2] <http://hdr.undp.org/en/content/human-development-index-hdi>

[3] <https://www.investopedia.com/terms/g/gdp.asp>

[4] <https://en.wikipedia.org/wiki/Life_expectancy>