

Mini Project 1

Khushal Giri, Vaishnavi Srinivasan, Ayesha Bhimdiwala and Harsh Mehta

2/1/2018

Introduction:

Gapminder Foundation (Gapminder.org) is a non-profit venture that promotes sustainable global development and achievement of the United Nations Millennium Development Goals. It uses statistics and social, economic and environmental development information at local, national and global levels. [Source: Wikipedia]

Gapminder package contains values for life expectancy, GDP per capita, and population, every five years, from 1952 to 2007 for each of 142 countries from five continents. [Source: CRAN.R]

We will be addressing the question -

Does increase in life expectancy since World War 2 be largely explained by increases in GDP per capita?

To validate if such a relationship exists, we focus on performing data analysis on the explanatory variables using three questions.

Question 1: GDP and life expectancy in 2007

We begin the analysis by looking at trend between GDP and life expectancy in 2007 for all continents.

Looking at the linear model plots (attached in Appendix), we found that it doesn't capture the trend of GDP per Capita vs Life Expectancy for all continents. For example, most of the African countries have a GDP of less than 2500\$ and very few countries with GDP greater than 2500\$.

Now we use loess model, a nonparametric method, that fits a smooth line through a timeplot or scatter plot to help asses the relationship between variables and foresee trends. It is used when we have noisy data, sparse data points or weak interrelationships that interfere with fitting a line of best fit. We will also be performing log transformation on the data to improve the fit.

From the graphs, with respect to GDP and life expectancy in 2007, we can see that:

- Africa has a linear increase with more countries having lower GDP and lower life expectancy and few countries with high GDP and life expectancy.
- America's trends are best described by a loess curve, with most countries having high GDP and life expectancy (centered around 72).
- Europe also has a linear increase in GDP and life expectancy - low GDP countries have low life expectancy compared to countries with high GDP. There are a few outliers - few countries have higher GDP but have comparatively lower life expectancy compared with countries having the same GDP.
- Asia is also best described by a loess curve, with countries being spread out on the curve. It has countries with low, medium and high GDP and life expectancy, with more data cluttered around the tails.

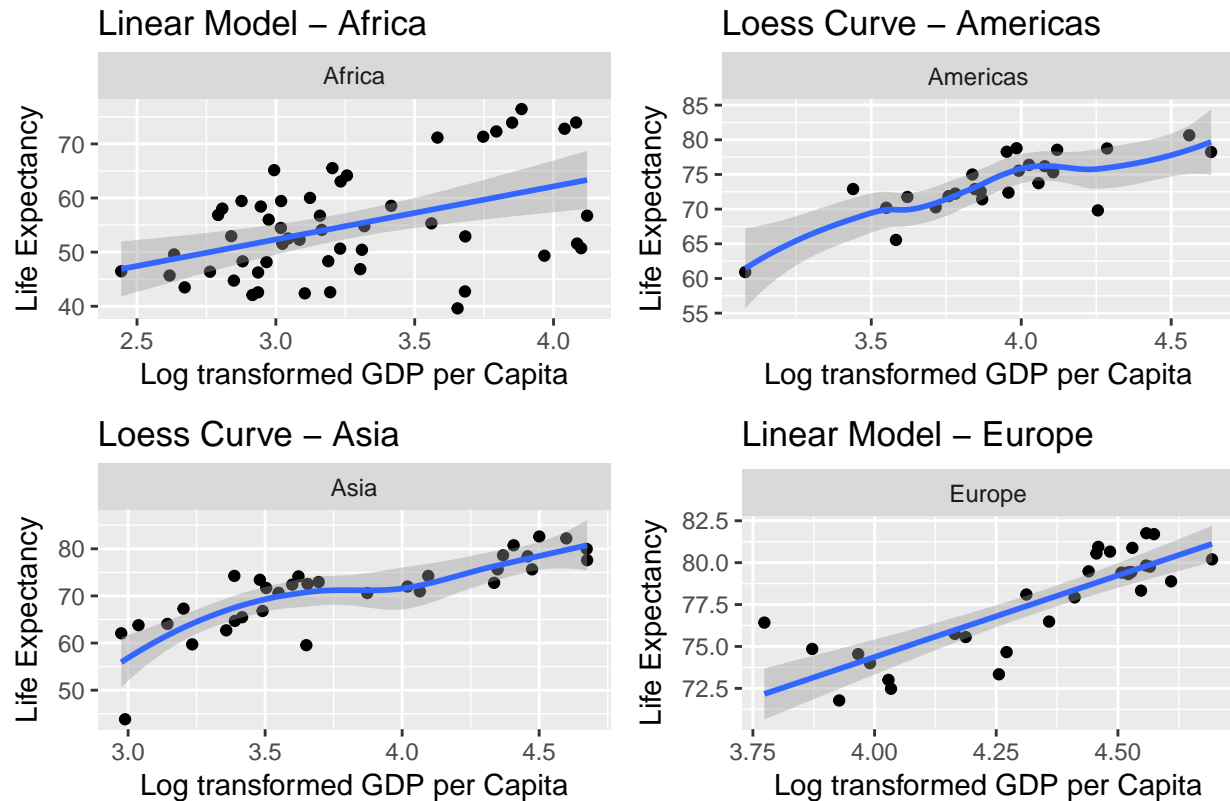
- We did not consider Oceania as it has only two data points and it was not possible to fit any models for that continent.

Even though we find linear relationship between GDP and life expectancy, there a number of additional parameters to be considered for estimating life expectancy like health-care available, social factors, etc.

Looking at the “Life Expectancy vs GDP 2007” graph (next page), we can see that life expectancy is increasing as GDP is increasing in each country of the continent. We can see that all the countries have an additive shift - the lines are parallel to one another with different mean.

```
## `geom_smooth()` using method = 'loess'
## `geom_smooth()` using method = 'loess'
```

Log10(GDP) vs LifeExpectancy for each continent



```
## The model captures 20.38938 % of variance in the data for Africa in 2007.
```

```
##
```

```
## The model captures 73.44411 % of variance in the data for Americas in 2007.
```

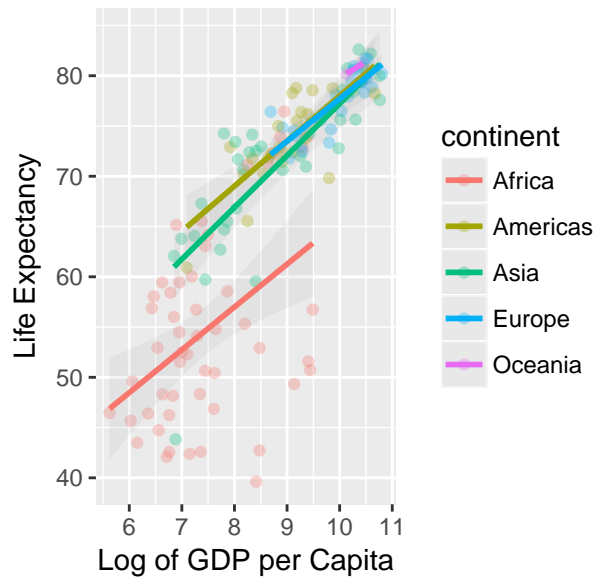
```
##
```

```
## The model captures 69.50029 % of variance in the data for Asia in 2007.
```

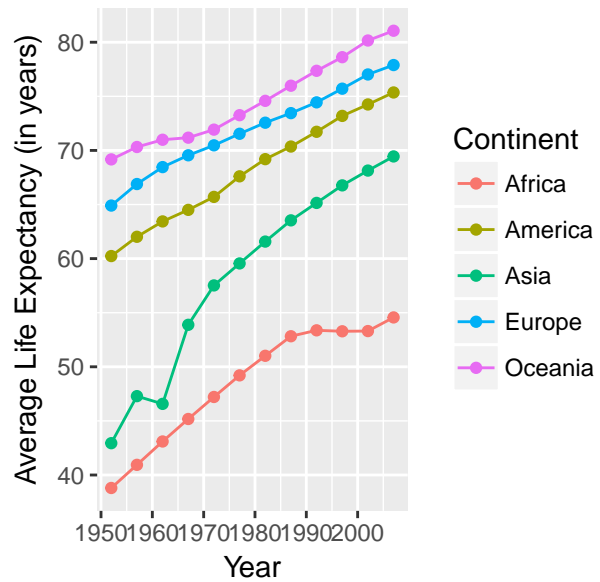
```
##
```

```
## The model captures 69.86601 % of variance in the data for Europe in 2007.
```

Life Expectancy vs GDP – 2007



Life Expectancy Over Time



Question 2: Life expectancy over time by continent

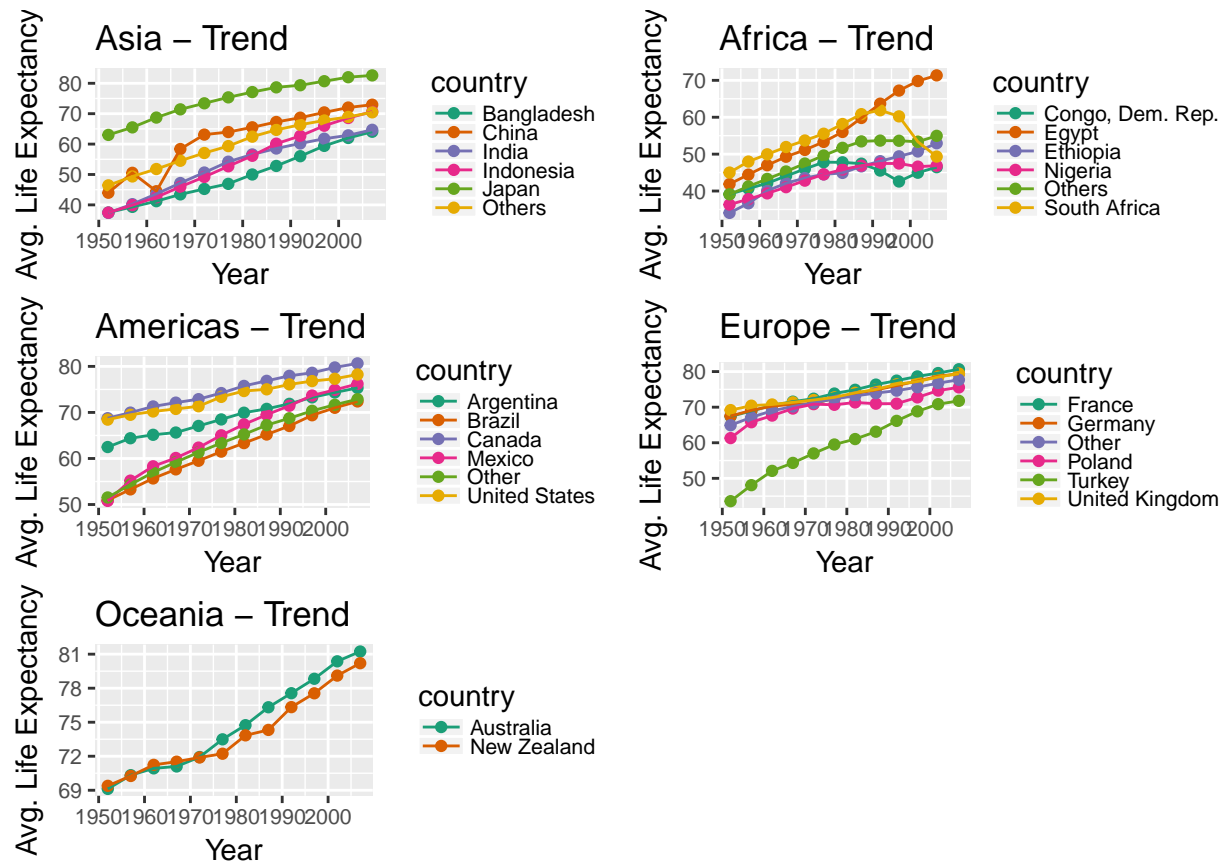
Next we start analysing life expectancy over time for each continent. We begin the analysis by calculating the weighted average life expectancy and then use this data for plotting the average life expectancy against each year.

By observing the “Life Expectancy Over Time” graph (above), we deduce that average life expectancy increases linearly over the years for all the continents - with Oceania having the highest and African countries having the lowest average life expectancy from 1952 to 2007.

Oceania, Europe and America have a head start with higher life expectancy at 1952. Asia, with life expectancy at around 40 in 1952, has the fastest growth rate. It nears America's average life expectancy - starting with a gap of 18 years at 1952 to a gap of 5 years in 2007.

Africa has the slowest growth rate with average life expectancy starting at below 40 years and reaching 55 years near 2007. Even though the growth is linear till the 1990's, it drops a little near the 1990's.

We use “<Continent_Name> - Trend” graphs to display some of the most populous countries in each continent. These countries best describe how they affect average life expectancy over the years in that continent.



Oceania

- There is a steady growth till 1975, after which the growth is faster.

Americas

- United States and Canada have the highest population with very high average life expectancy, between 70 to 80 years.
- Brazil, Mexico and Argentina are the next highly populated countries having a relatively lower life expectancy, between 50 to 70.
- The average value of these five countries put together represent the average life expectancy of Americas.

Europe

- There is a very light dip in the life expectancy that starts around 1980.
- The fall of Soviet Union around 1980's reduced the life expectancy of Eastern European countries (eg. Poland) due to poor economy and lack of good health care facilities.
- We also observe that even though there is an increase in the average life expectancy in Eastern European bloc after 1995, it is comparatively lower to central and western europe.

Asia

- Since, India and China share approximately 60% of the population in Asia, most of the anomalies can be explained using them.
- China and India had a life expectancy of around 40 during these years, bringing down the life expectancy of Asia to around 45, even though Japan has a average life expectancy of 65 and rest of the countries at around 52.
- There is a drop in average life expectancy between the years 1955 and 1965 in the “Life Expectancy Over Time” graph.
 - China suffered from a great famine which claimed an estimated 30 million lives.
 - It also reduced birth rate from 1955 till 1967. This was followed by steady growth in life expectancy starting from 1972.
- India’s life expectancy was linearly growing. But compared to other populated countries, it was low. Also, it starts flattening at around 1990’s. During the late 1990’s and early 2000’s India lost around 2.5 million to AIDS, reducing its average life expectancy.
- At 2007, we can see that India and China have a life expectancy at 60 and 68 respectively which marks Asia’s life expectancy is around 65’s years.

Africa

- Life expectancy is growing linearly till 1990 after which drops a little at 1990 and raises at 2002.
- There are numerous reasons for this drop. Some of them are described below,
 - AIDS and tuberculosis epidemic which hit South Africa in late 90’s.
 - It was estimated that South Africa had lost at least 17 years of average life expectancy to the AIDS epidemic by 2000-2005.
 - Younger age groups were affected by AIDS more and lost their lives quite early (nearly a quarter million deaths were caused in Ethiopia and Nigeria.)
 - The neonatal death rate (due to poor health-care facilities) was severely high in some African countries (South Africa, Nigeria).
 - Regional politics in Nigeria around 1990’s also contributed to higher percentage of deaths, bringing down Africa’s life expectancy.

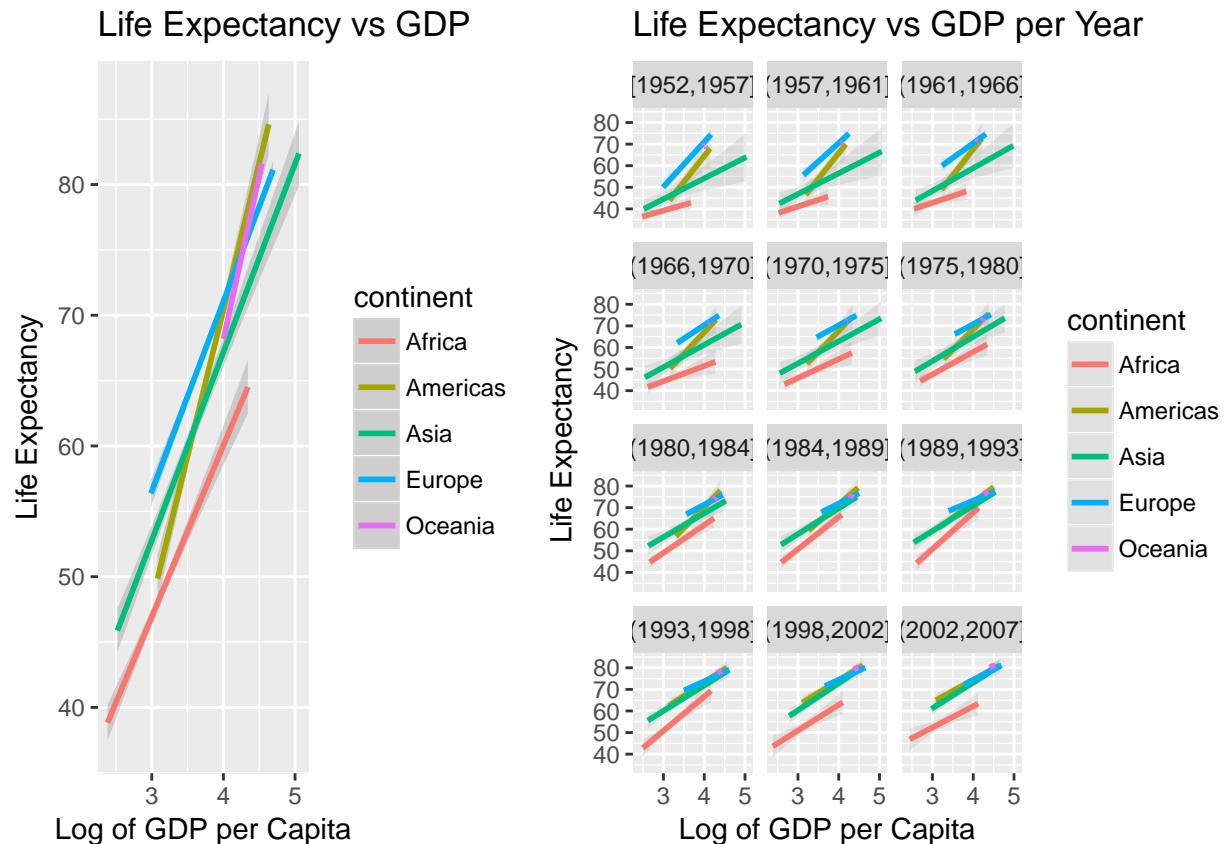
Question 3: Changes in the relationship between GDP and life expectancy over time

Now we start analysing the relation between GDP and life expectancy over time. GDP per capita is log transformed to ease data analysis. To understand the relationship, we plot GDP vs life expectancy for each continent over the years.

The life expectancy per continent is increasing linearly as GDP increases. This can mean that as countries gain more money, they were investing in better health care facilities, better infrastructure, etc that promotes good health and in turn higher life expectancy.

By looking through time, GDP and life expectancy also linearly increase. An interesting information to note is that even though all continents had a wide gap between life expectancy in the initial years(1952), over time (from 1993) they are racing to catch up to each other. America has the steepest rise, with Asia and

Europe having a normal rise. Asia has the widest rise from a very low GDP and life expectancy to very high values. Africa being the only continent to show a very slow minimal rise both in terms of GDP and life expectancy. But we also need to note that there is a wider standard deviation towards the ends, especially in Asia, though over the years it tends to reduce.



Conclusion:

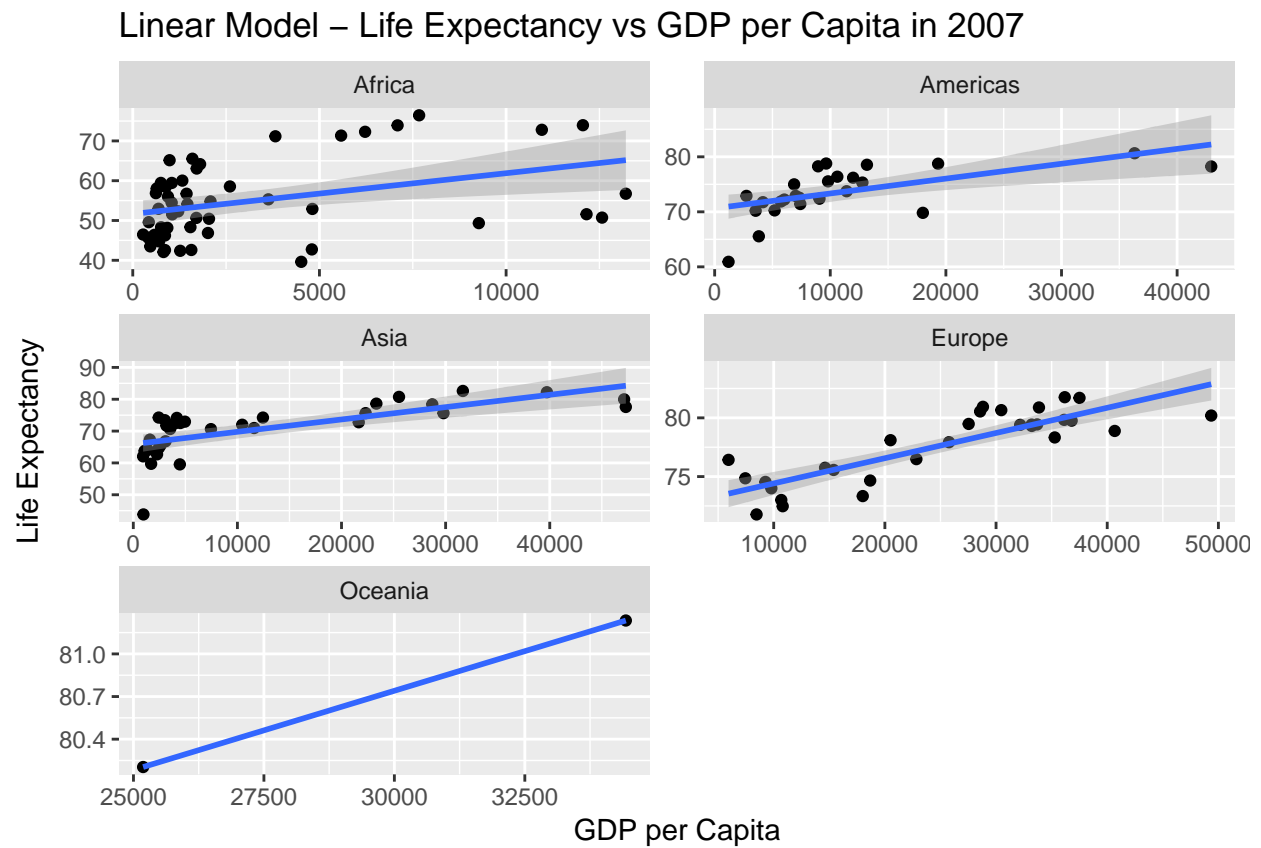
Even though looking at the graphs we can conclude that through the years, as GDP increases so does life expectancy, but it would be unwise. Based on standard deviation bands, we can see that different countries have different life expectancy even though they have nearly the same GDP. Also to note are endogenous factors such as regional politics, natural disasters and epidemics and how they would have disabled a nation's progress, such as Africa (even though it holds a majority of the world's natural resources (30%) and is considered as the world's 10 fastest-growing economies, it has the slowest growth rate).

References:

- <http://www.pilibrary.com/articles1/political%20experiences%20in%20nigeria.htm>
- <http://www.aljazeera.com/indepth/interactive/2016/10/mapping-africa-natural-resources-161020075811145.html>
- <https://www.un.org/press/en/2001/aids18.doc.htm>
- http://www.who.int/gho/mortality_burden_disease/life_tables/situation_trends_text/en/
- <https://www.iol.co.za/news/south-africa/western-cape/four-reasons-for-sas-low-life-expectancy-1798106>

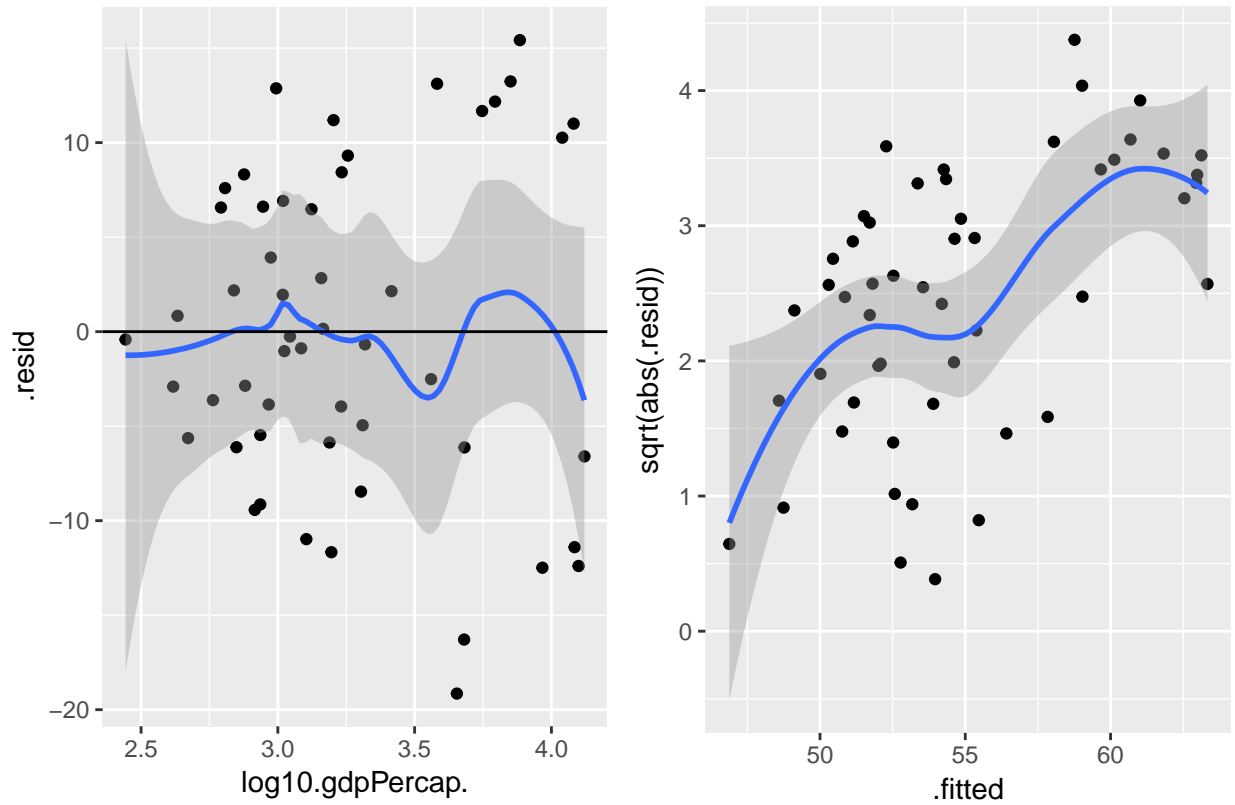
- https://en.wikipedia.org/wiki/Great_Leap_Forward
- <http://www.statisticshowto.com/lowess-smoothing/>

Appendix:



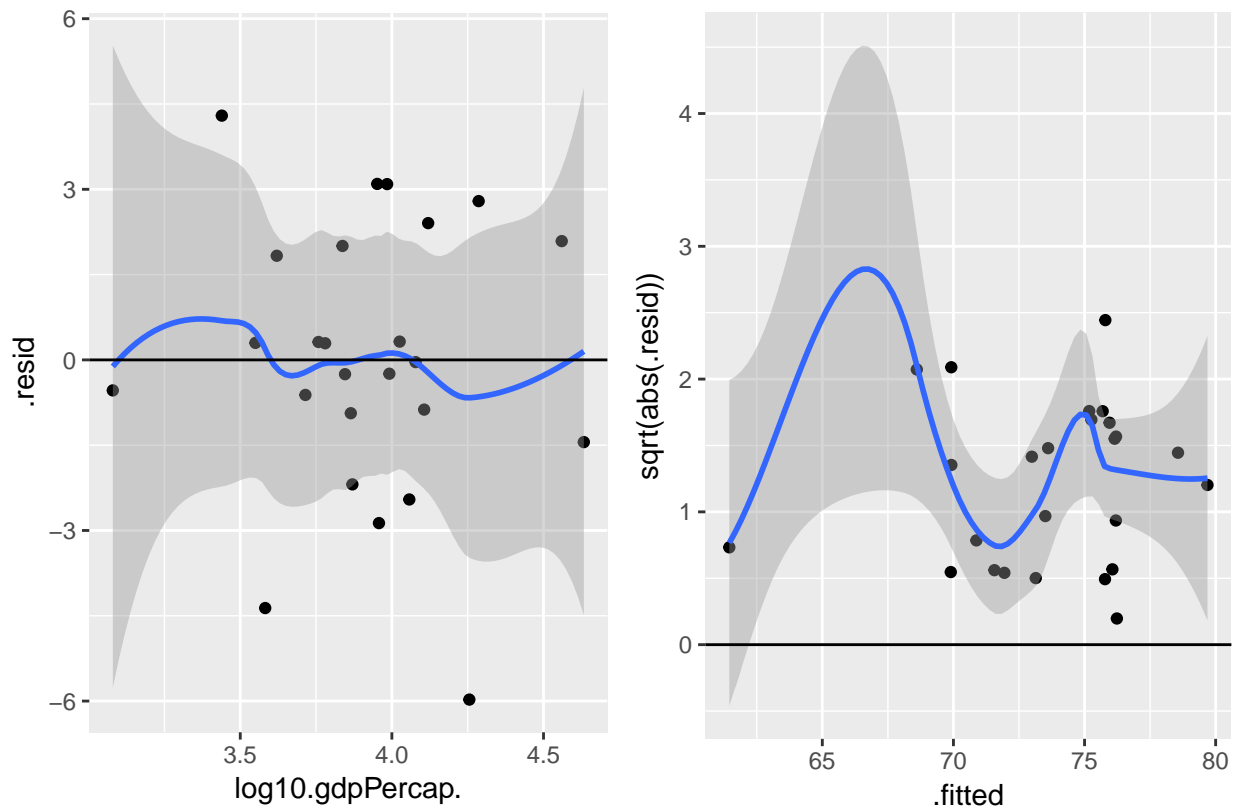
```
## `geom_smooth()` using method = 'loess'
## `geom_smooth()` using method = 'loess'
```

Africa – Residual Plot & Fitted vs Residual Plot



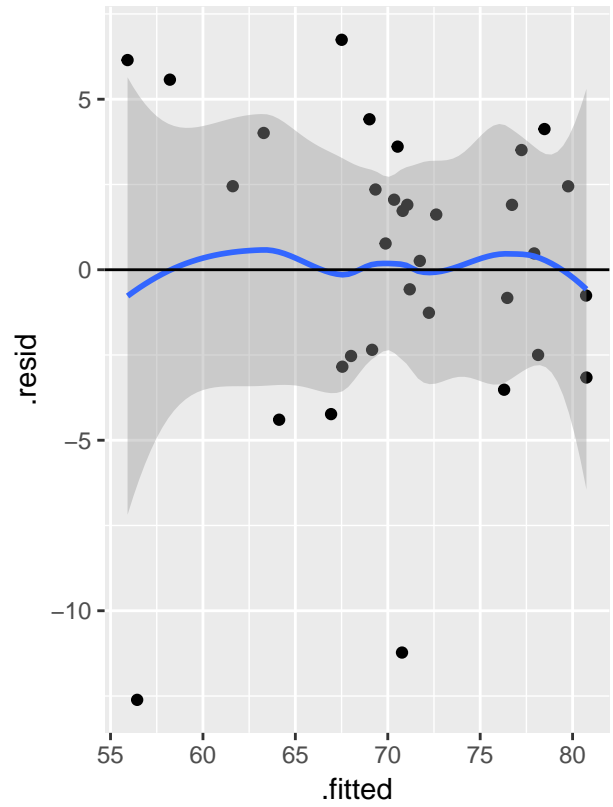
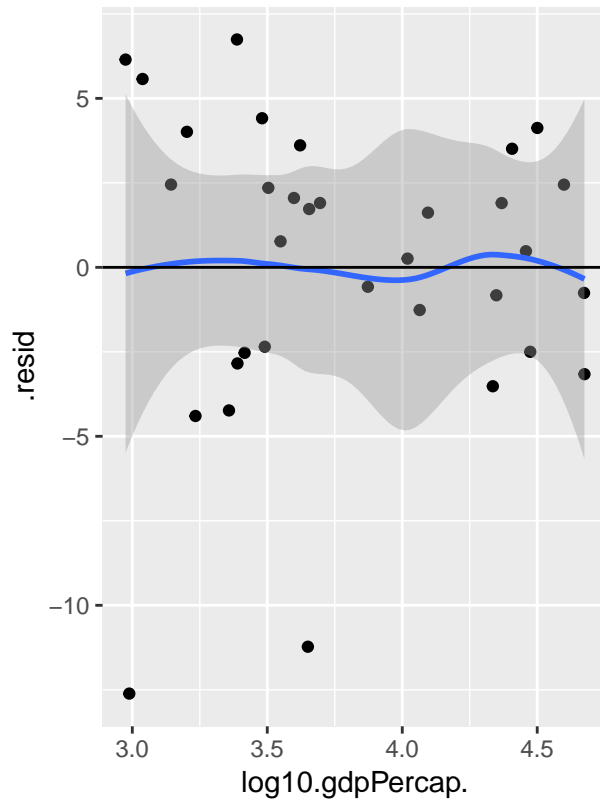
```
## `geom_smooth()` using method = 'loess'
## `geom_smooth()` using method = 'loess'
```


Americas – Residual Plot & Fitted vs Residual Plot



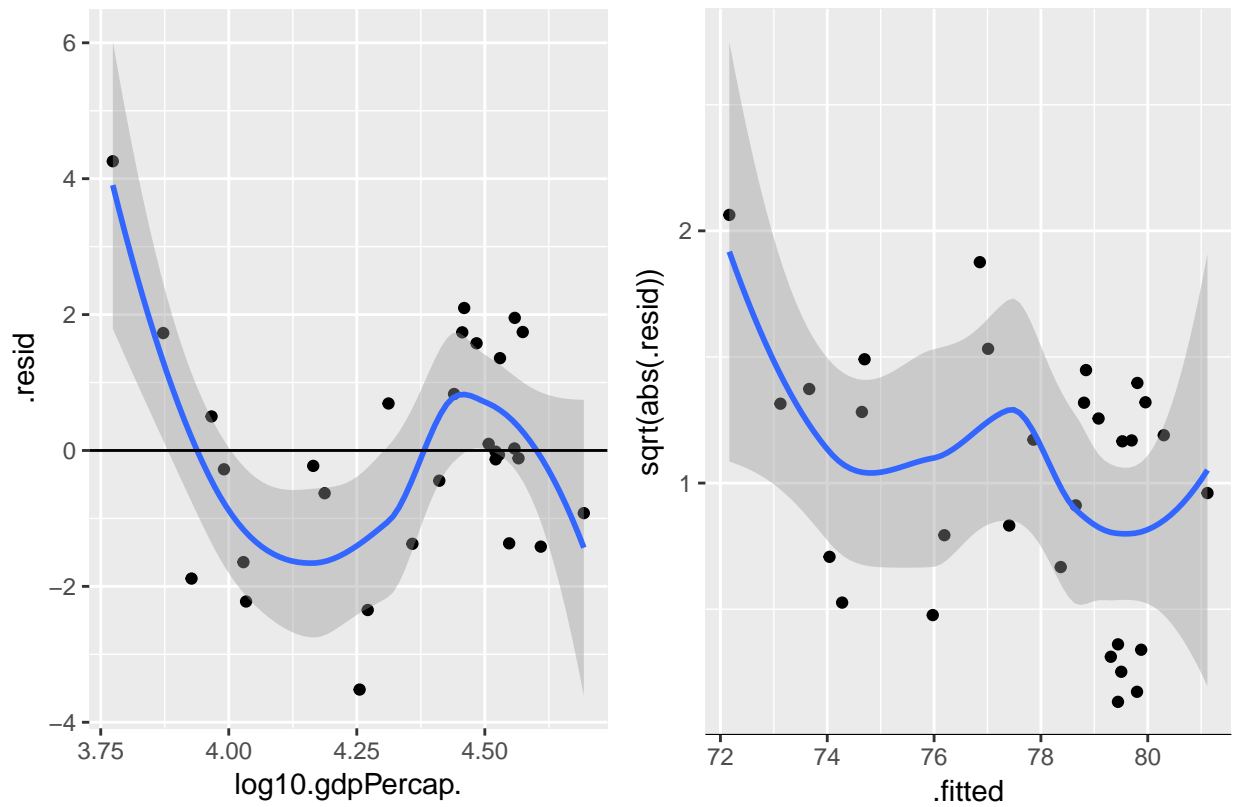
```
## `geom_smooth()` using method = 'loess'
## `geom_smooth()` using method = 'loess'
```

Asia – Residual Plot & Fitted vs Residual Plot



```
## `geom_smooth()` using method = 'loess'
## `geom_smooth()` using method = 'loess'
```

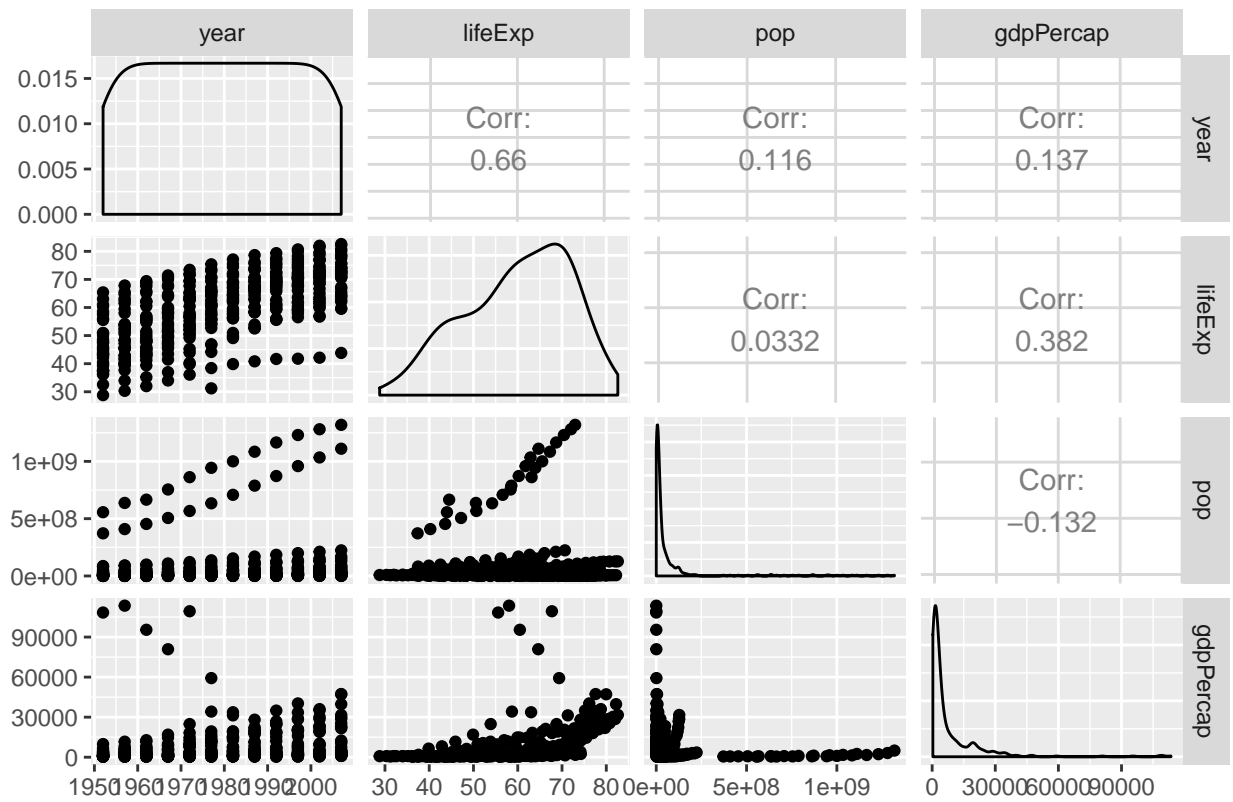
Europe – Residual Plot & Fitted vs Residual Plot



The scatter plots describing the correlation between GDP, life expectancy, time and population per continent are given below.

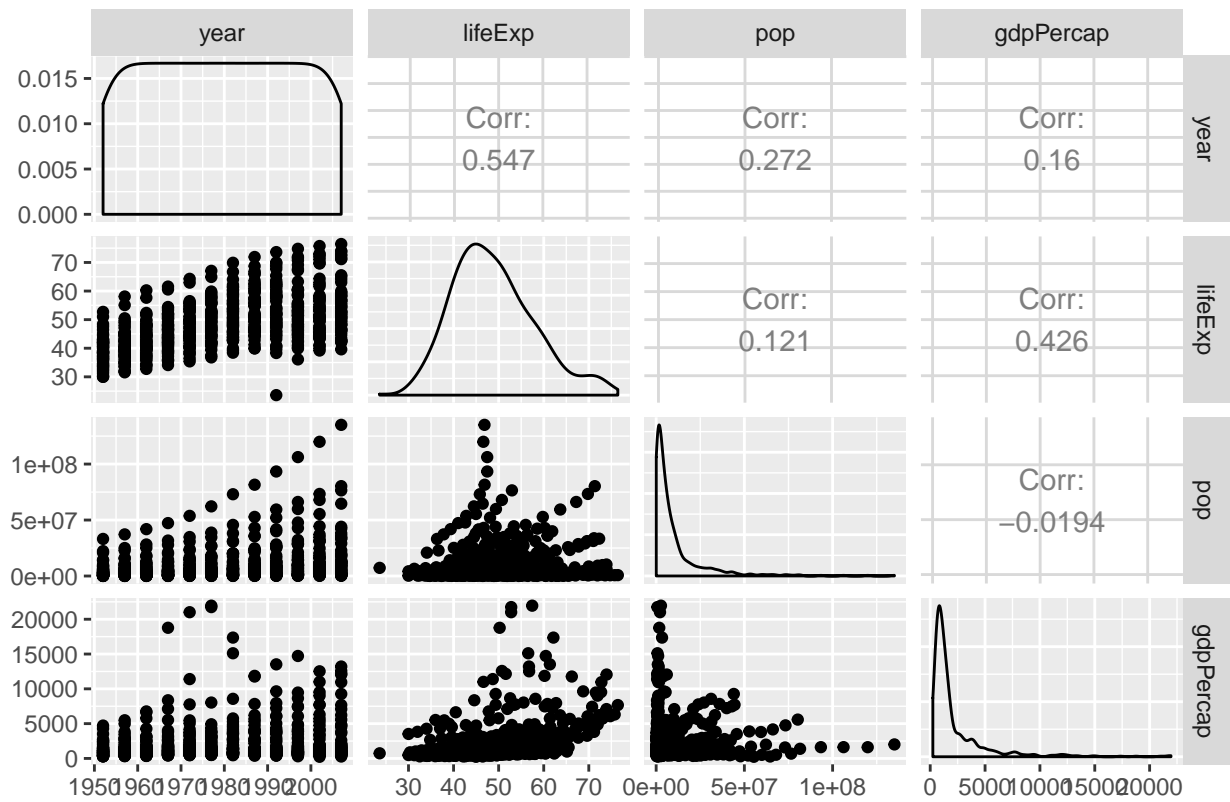
Asia

Asia – Scatter Plot between GDP, Life Expectancy, Time and Population



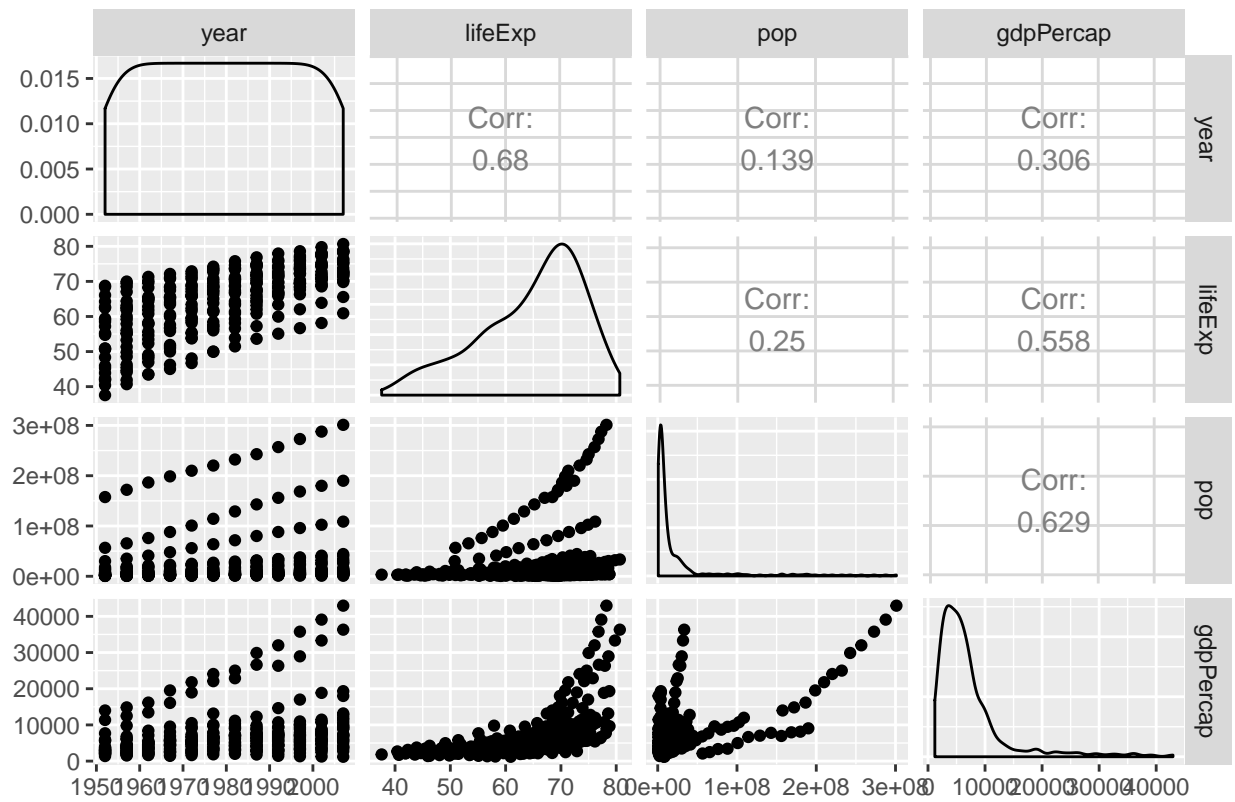
Africa

Africa – Scatter Plot between GDP, Life Expectancy, Time and Population



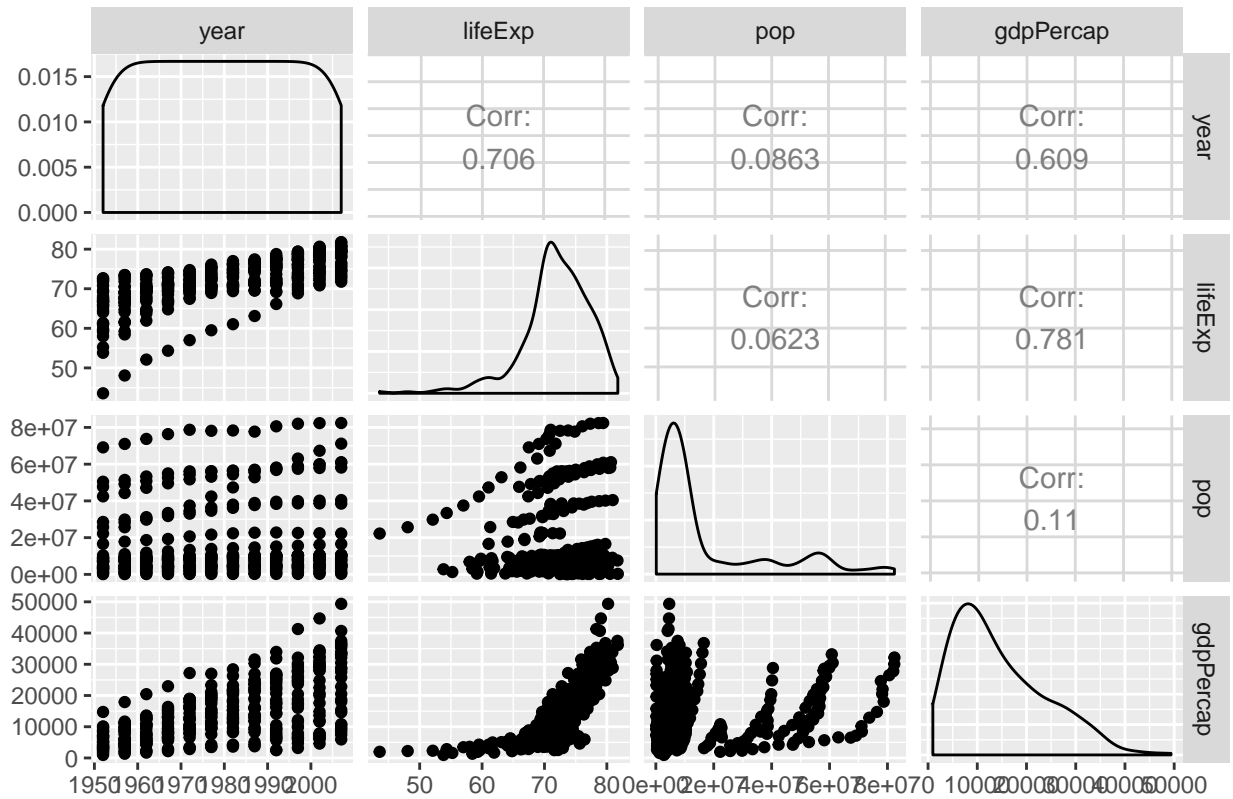
Americas

America – Scatter Plot between GDP, Life Expectancy, Time and Population



Europes

Europe – Scatter Plot between GDP, Life Expectancy, Time and Population



Oceania

Oceania – Scatter Plot between GDP, Life Expectancy, Time and Population

