Mini Project 1 - Life Expectancy (Gapminder)

Khusaal 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 mainly address the following question - Does increase in life expectancy since World War 2 be largely explained by increases in GDP per capita?

To answer this, we focus on performing data analysis on the explanatory variables year (ranges from 1952 to 2007 in increments of 5 years), lifeExp (Life Expectancy in years), pop (population), and gdpPercap (GDP per capita [US$, inflation-adjusted]).

We divide our analysis in three sections - GDP and life expectancy in 2007, Life expectancy over time by continent, and Changes in the relationship between GDP and life expectancy over time.

Section 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 time plot or scatterplot to help assess 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.

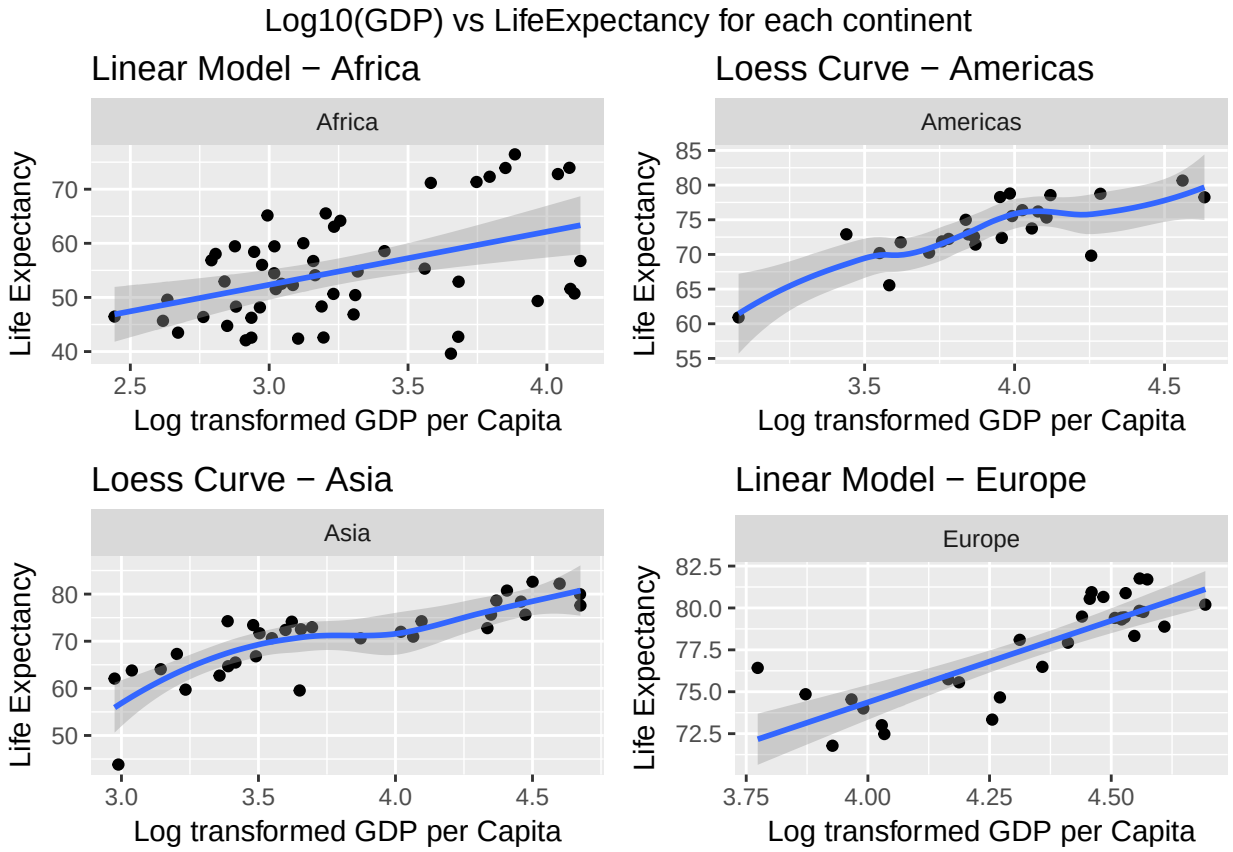
From the graphs, with respect to GDP and life expectancy in 2007, we can see that:

- 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 a 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'
```

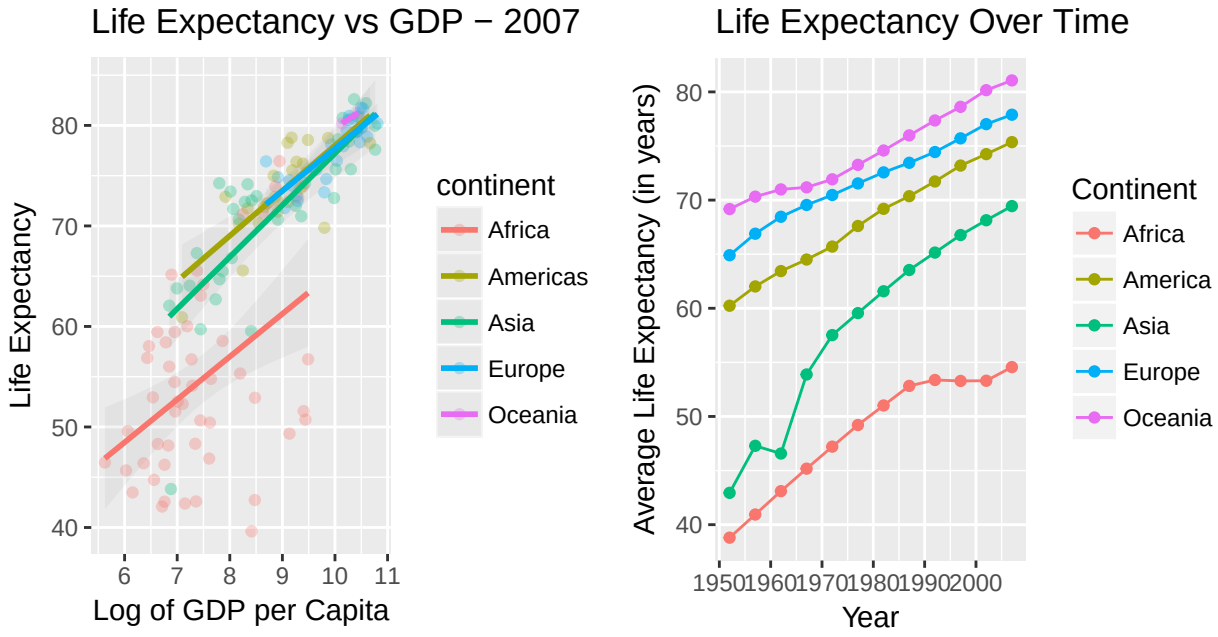



```
## The model captures  20.38938 % of variance in the data for Africa in 2007.
```

```
##
## 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.
```



Section 2: Life expectancy over time by continent

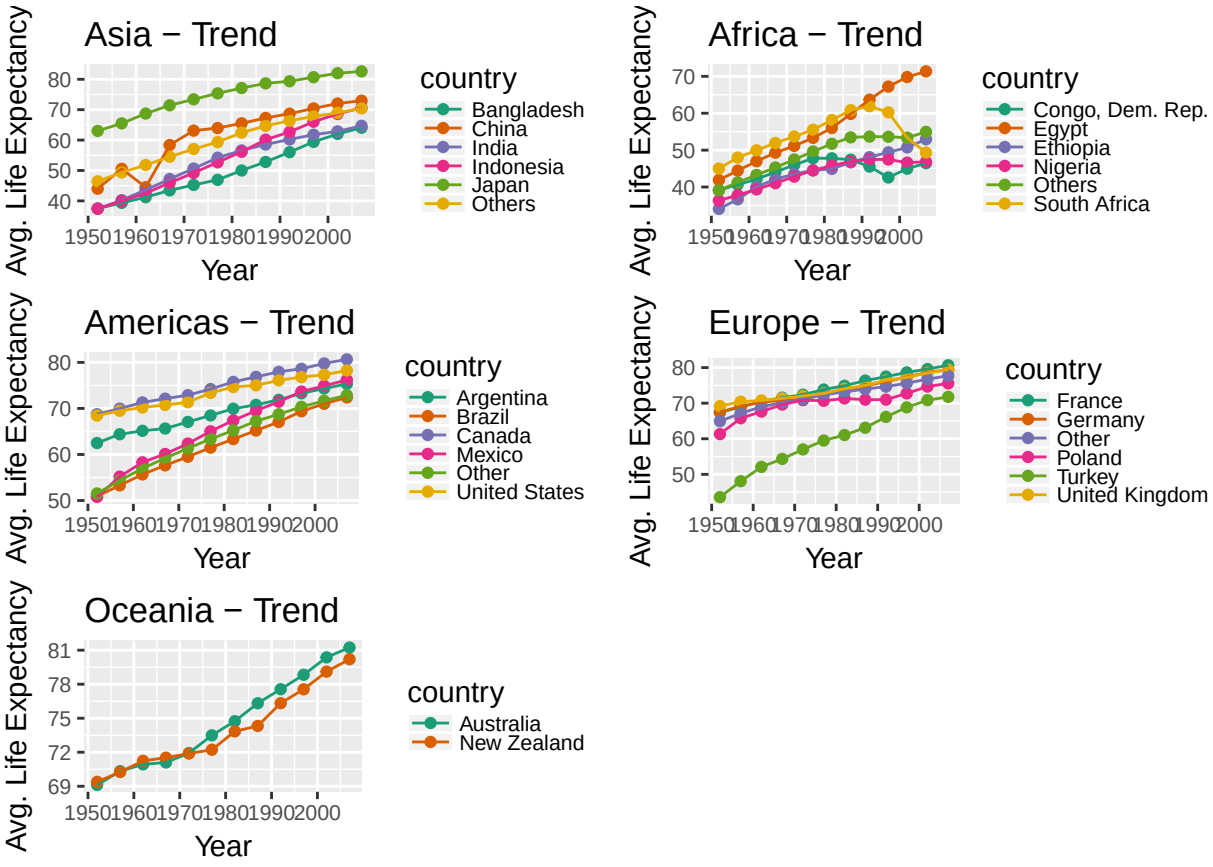
Next, we start analyzing life expectancy over time for each continent. We begin the analysis by calculating the weighted average life expectancy and then use this data for plotting the average life expectancy against each year.

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 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 1990's, it drops a till 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:

- The 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 represents the average life expectancy of Americas.

Europe:

- There is a very slight 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 a poor economy and lack of good healthcare 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 in 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 an 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 the 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.
- The are numerous reason 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 quiet 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 a higher percentage of deaths, bringing down Africa's life expectancy.

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

Now we start analyzing the relation between GDP and life expectancy over time. We begin by plotting GDP vs life expectancy for each continent and then look at how these variables have changed over the years. GDP per capita is log transformed to ease our analysis.

We get a linear model of log10 GDP per capita as an explanatory variable and life expectancy as a response variable for each continent. Therefore, the life expectancy in each continent is growing linearly by 10 units increase in GDP. This could be a good indicator to substantiate the claim that, as countries gain more money, they invest in better healthcare facilities.

During the initial years (1952 - 1957), it is interesting to note that all continents have a huge difference in their respective life expectancies. For example, Africa's life expectancy is about 40 years which is the lowest among all continents whereas Oceania life expectancy is about 70 years which is the highest among all continents.

Over the years they are racing against each other to either catch-up or outperform. Americas has the steepest rise, while Asia and Europe have a normal rise.

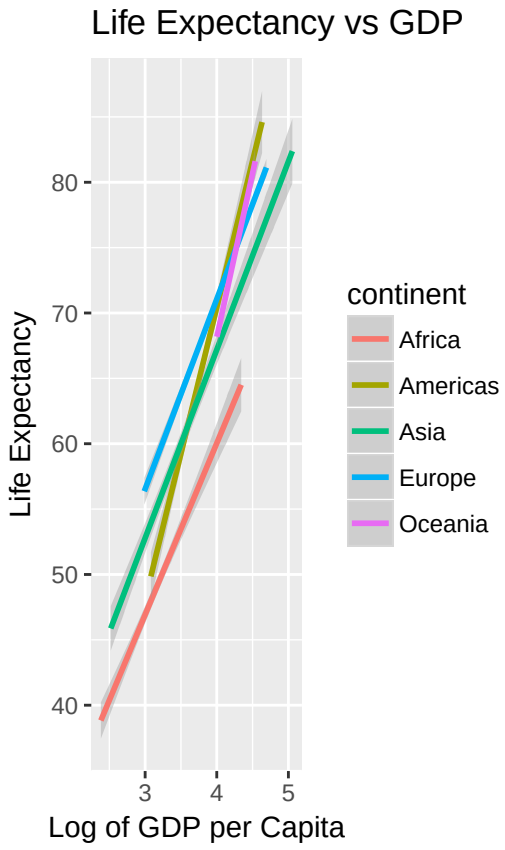
As explained before, the major contribution to Americas came from the USA and Canada which were working towards better healthcare policies and medical research. The steep rise of Americas can thus be explained due to better healthcare facilities which led to an increase in life expectancy that resulted in a better GDP per capita.

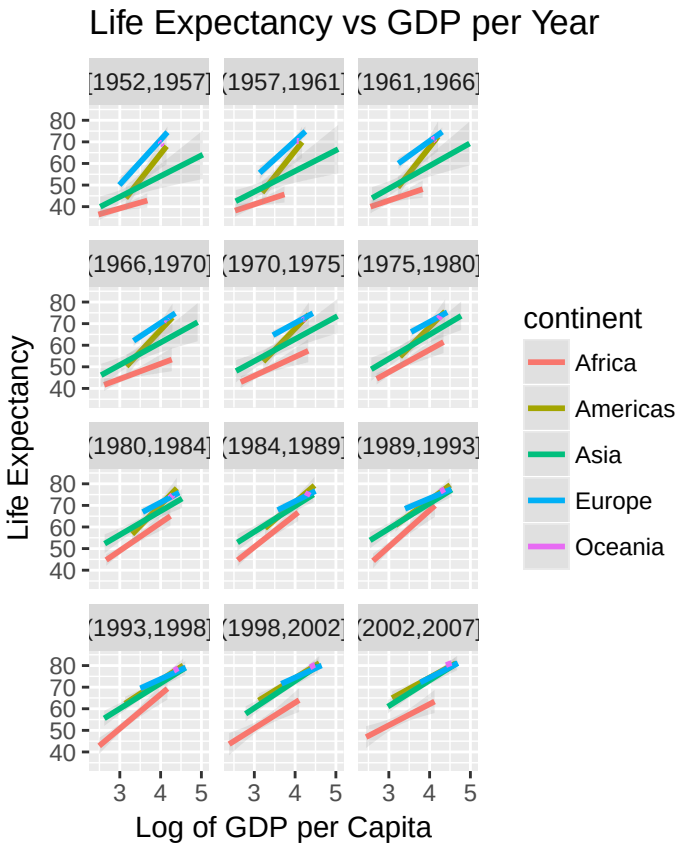
Initially, Europe has a steep rise but then it becomes somewhat constant. This is because after World War II, both the USA and Soviet were competing in Cold War and making efforts to rise but after the Soviet collapsed, there was an economic burden in eastern and central Europe. To support our claim, we found the following -

"Between 1970 and the end of the 1980s, life expectancy at birth in the former communist countries of CEE (Czech Republic, Hungary, Poland, and Slovakia), Russia and the Baltic states (Estonia, Latvia and Lithuania) stagnated or declined. This led to an increasing gap between them and Western European countries as the latter steadily improved. However, within a few years of the collapse of the Berlin wall in 1989, life expectancy started to steadily increase." [8]

Asia has the broadest rise from a very low GDP (of about 800 GDP per Capita) and life expectancy (of about 45 years) to very high values of (100,000 GDP per Capita) and life expectancy (of 82 years).

Africa is 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:

Looking at the graphs we can conclude that over the years, there is a linear relationship between the logarithmic scale of GDP per capita and life expectancy. But it would be unwise to arrive at such a conclusion because based on standard deviation bands in the graphs, we can see that different countries have

different life expectancy even though they have nearly the same GDP. Therefore, we should also consider the endogenous factors such as regional politics, natural disasters, and epidemics that have disabled the progress of a nation. For example, **Africa** holds 30% of the world's natural resources and is considered as the world's 10 fastest-growing economies, but has the slowest growth rate for life expectancy and **Eastern Europe** notwithstanding the consequences of the collapse of communism shows how abrupt political, economic and social change which could have serious adverse effects on population health.

References:

[1] http://www.pilibrary.com/articles1/political%20experiences%20in%20nigeria.htm

[2] http://www.aljazeera.com/indepth/interactive/2016/10/mapping-africa-natural-resources-161020075811145.html

[3] https://www.un.org/press/en/2001/aids18.doc.htm

[4] http://www.who.int/gho/mortality_burden_disease/life_tables/situation_trends_text/en/

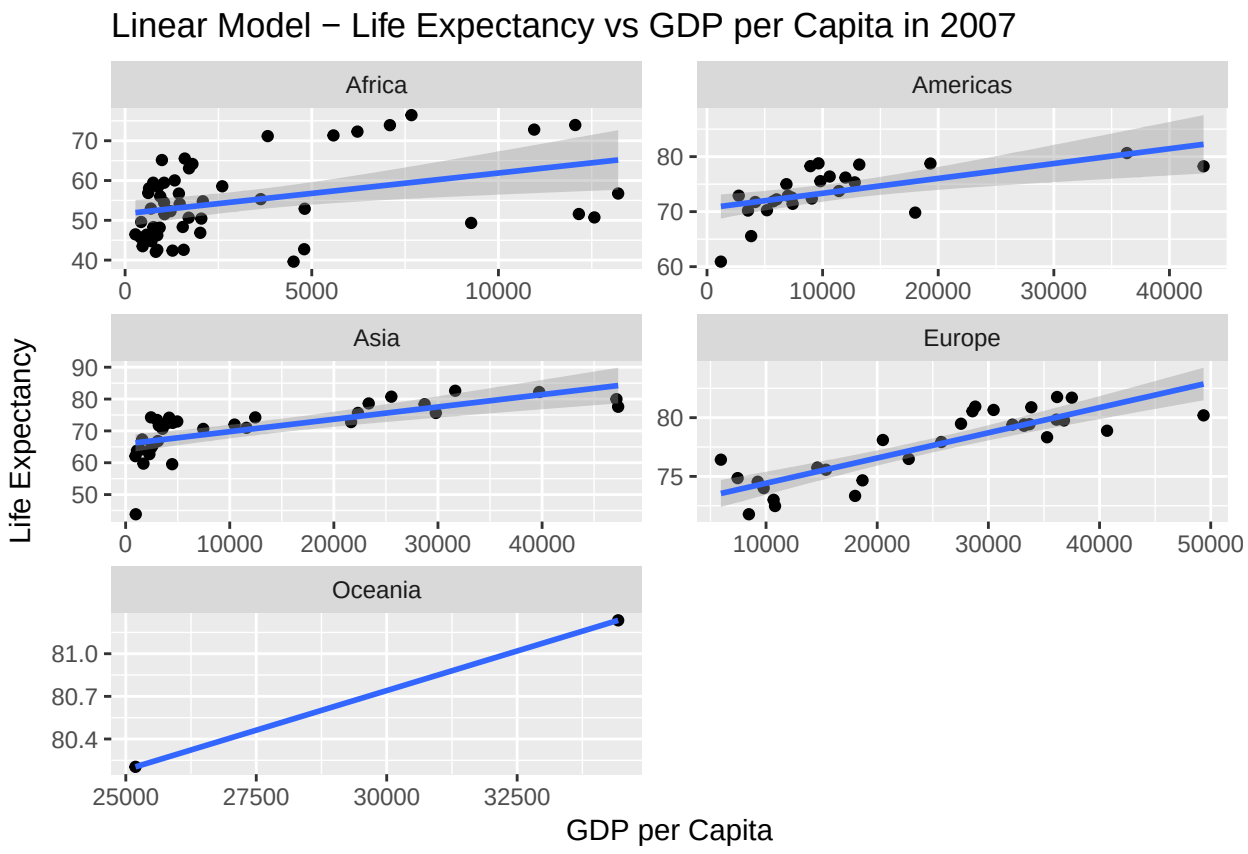
[5] https://www.iol.co.za/news/south-africa/western-cape/four-reasons-for-sas-low-life-expectancy-1798106

[6] https://en.wikipedia.org/wiki/Great_Leap_Forward

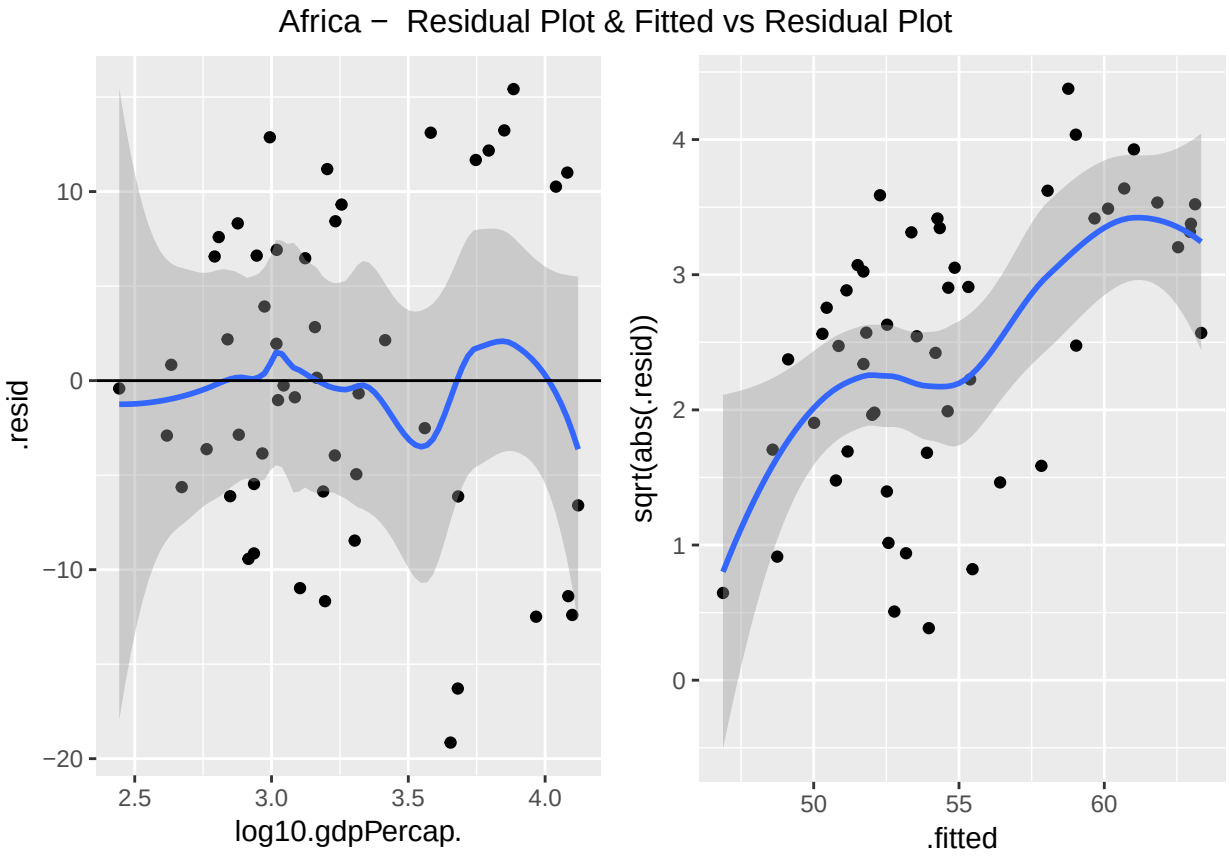
[7] http://www.statisticshowto.com/lowess-smoothing/

[8] https://academic.oup.com/ije/article/40/2/271/735545

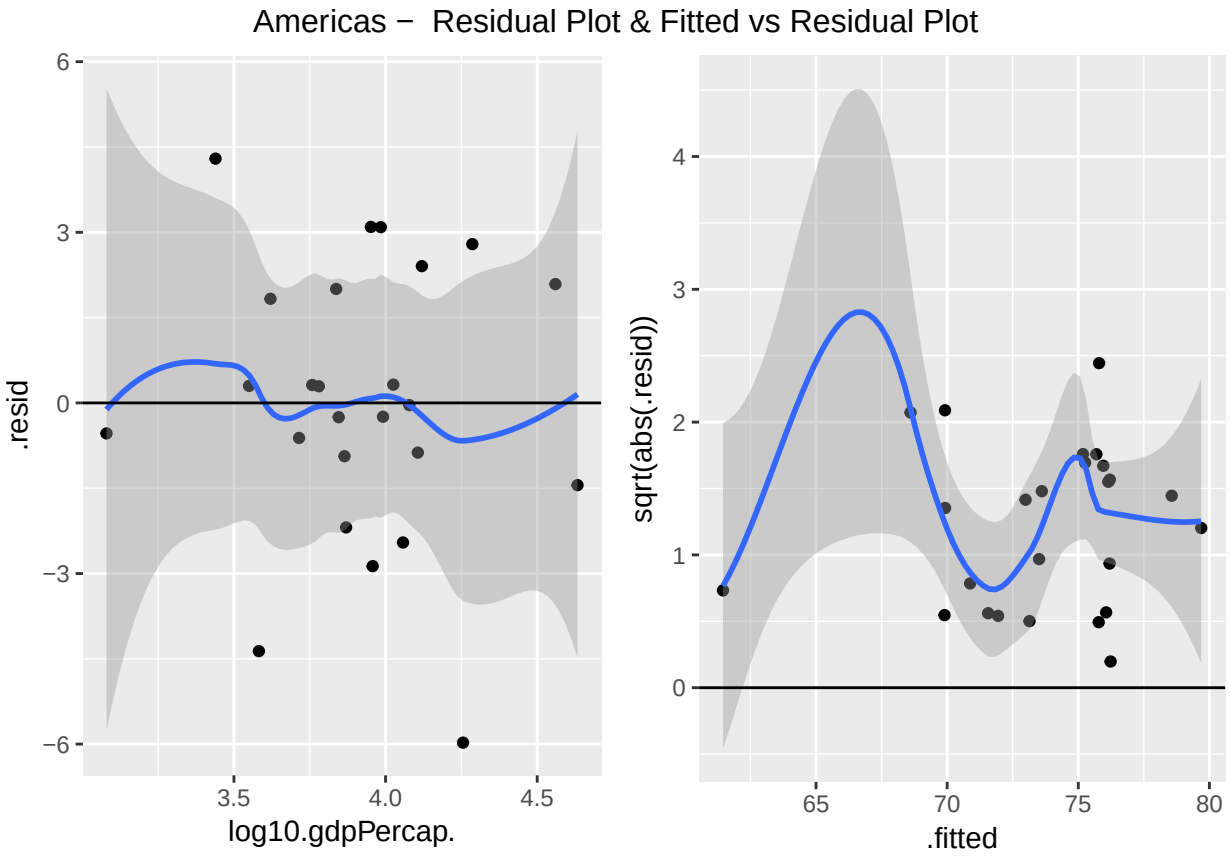
Appendix:



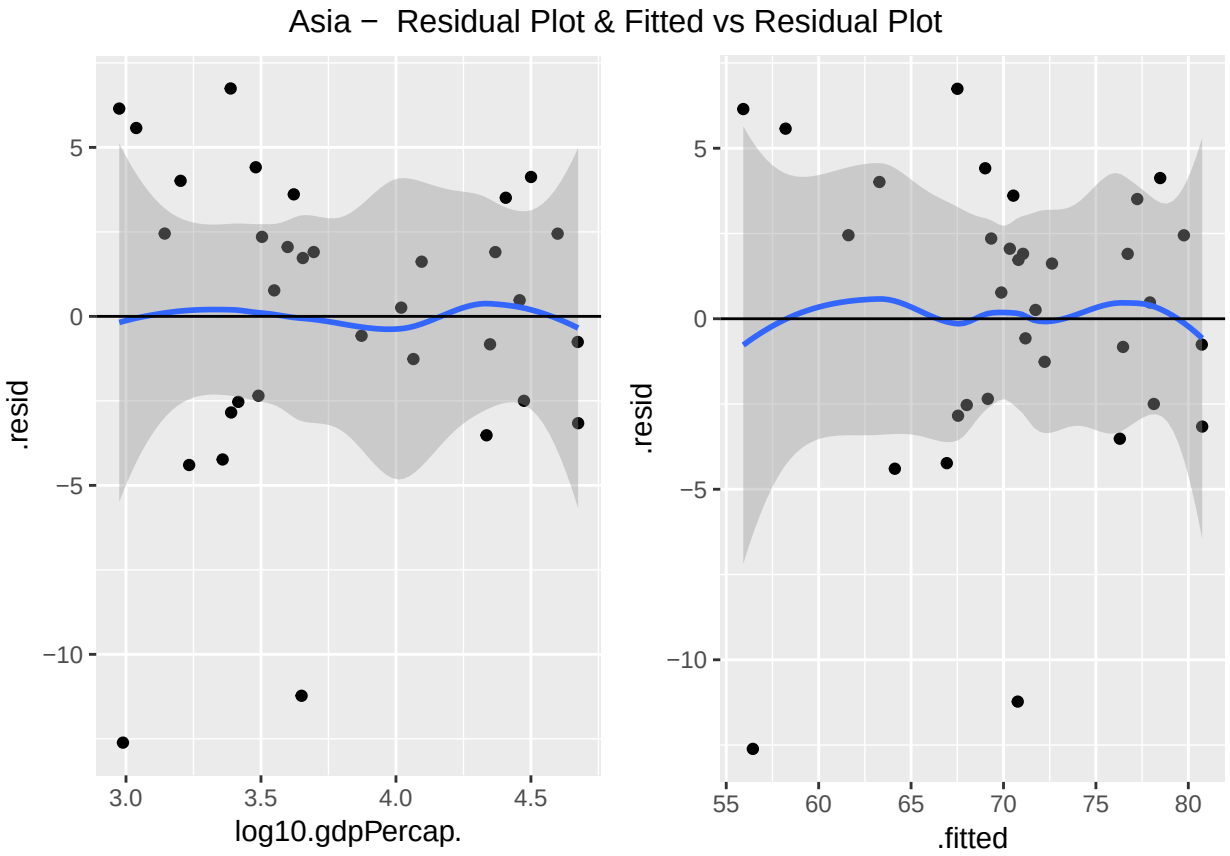
```
## `geom_smooth()` using method = 'loess'
## `geom_smooth()` using method = 'loess'
```



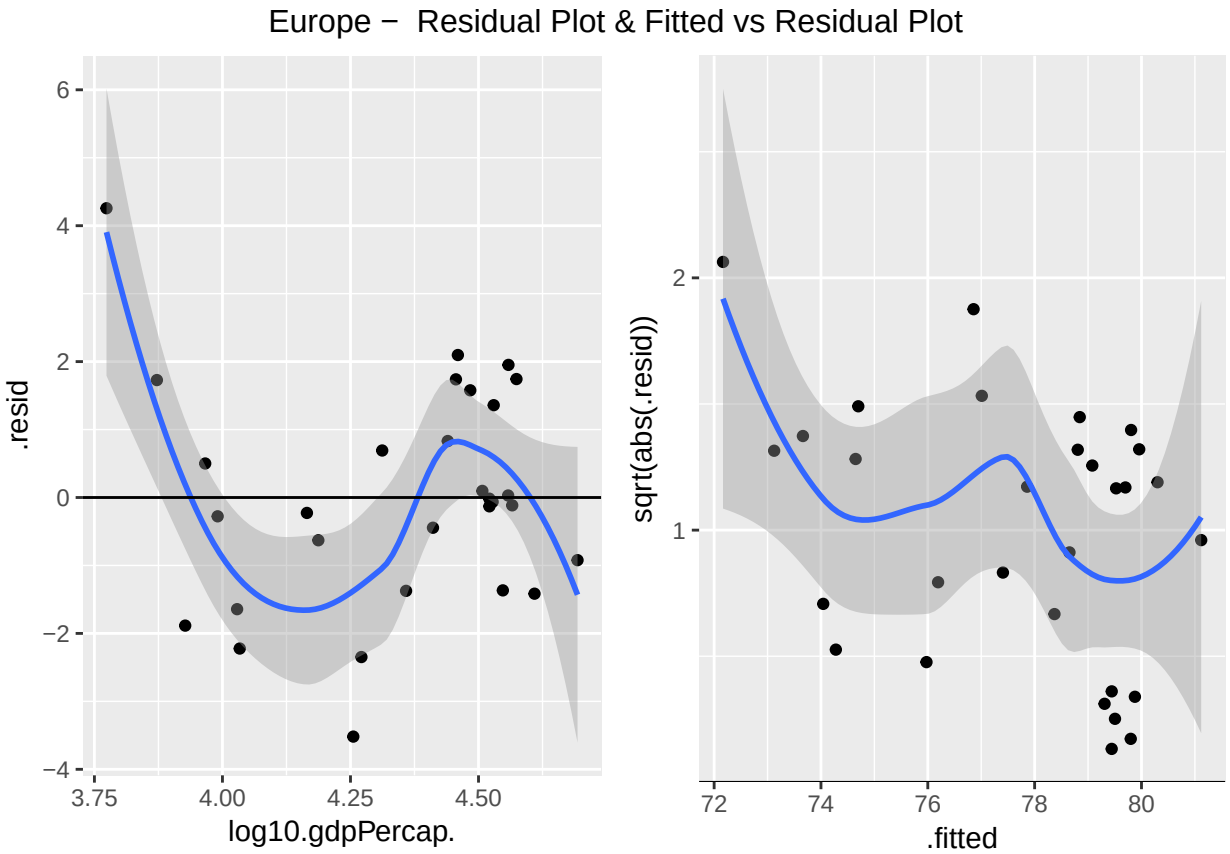

```
## `geom_smooth()` using method = 'loess'
## `geom_smooth()` using method = 'loess'
```




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

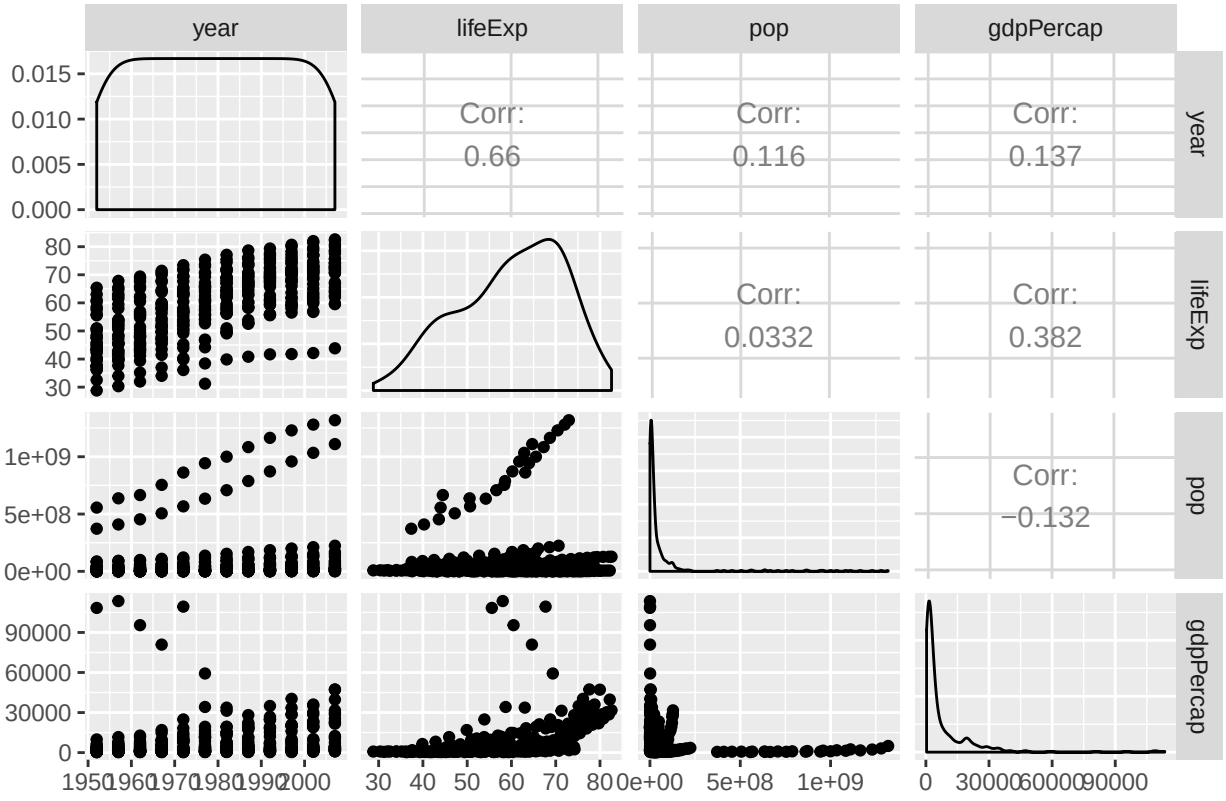



```
## `geom_smooth()` using method = 'loess'
## `geom_smooth()` using method = 'loess'
```

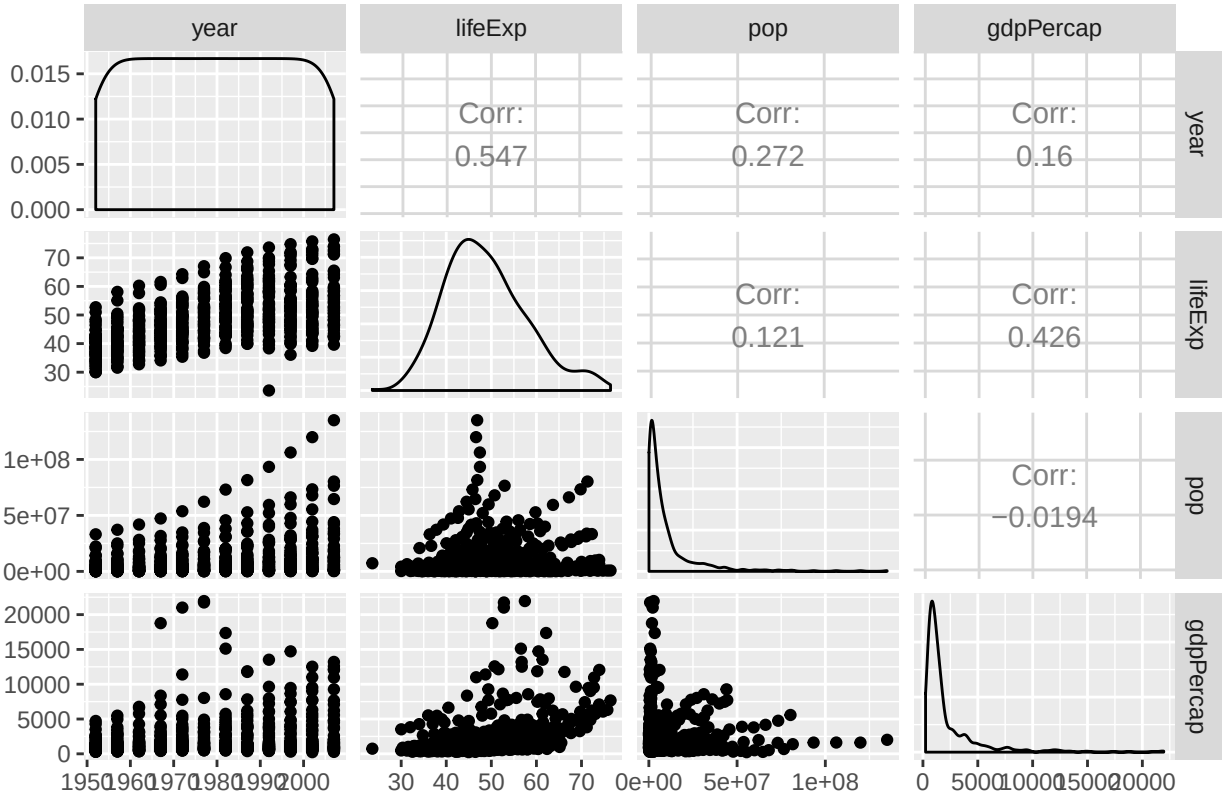


The scatter plots describing the correlation between GDP, life expectancy, time and population per continent are given below.

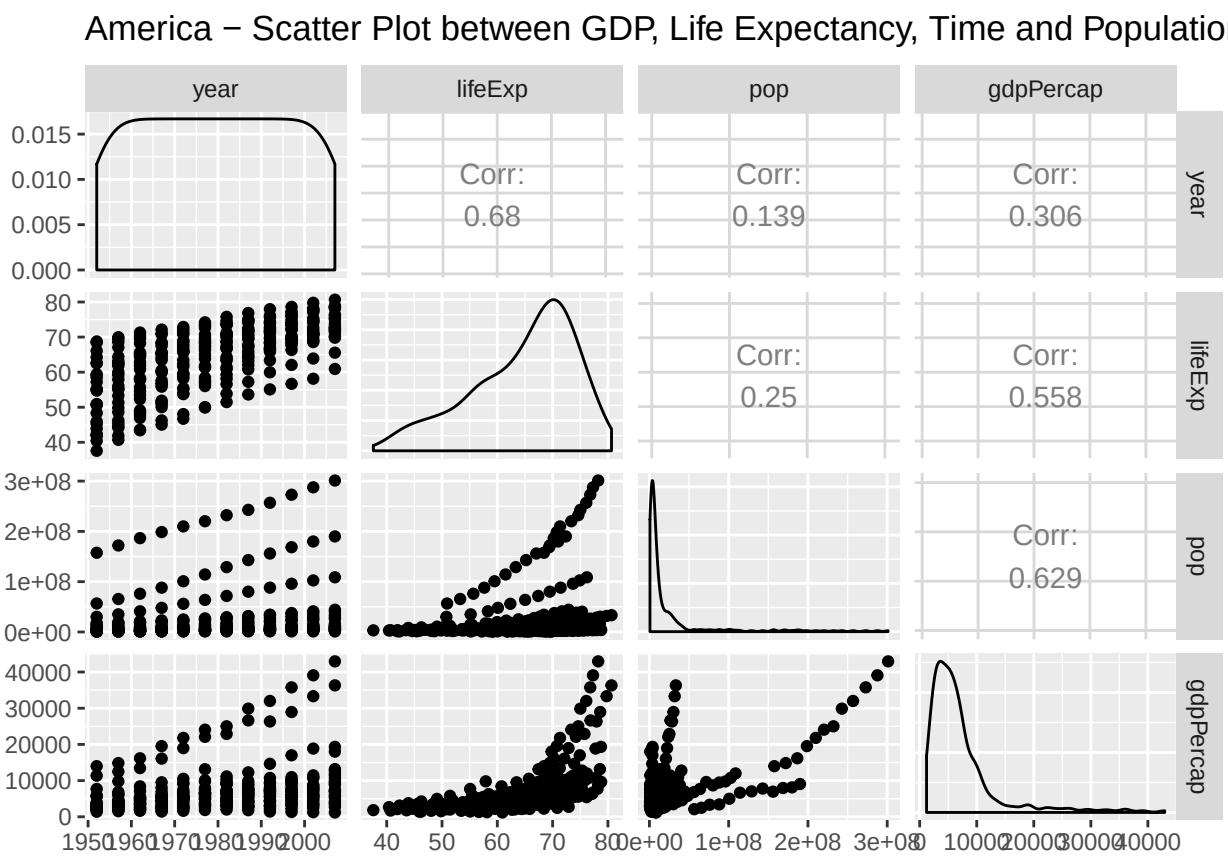
Asia


Asia − Scatter Plot between GDP, Life Expectancy, Time and Population
year
lifeExp
pop
gdpPercap
Corr:
0.66
Corr:
0.116
Corr:
0.137
Corr:
0.0332
Corr:
0.382
Corr:
−0.132
0.015
0.010
0.005
0.000
80
70
60
50
40
30
1e+09
5e+08
0e+00
90000
60000
30000
0
1950
1960
1970
1980
1990
2000
30
40
50
60
70
80
0e+00
5e+08
1e+09
0
30000
60000
90000

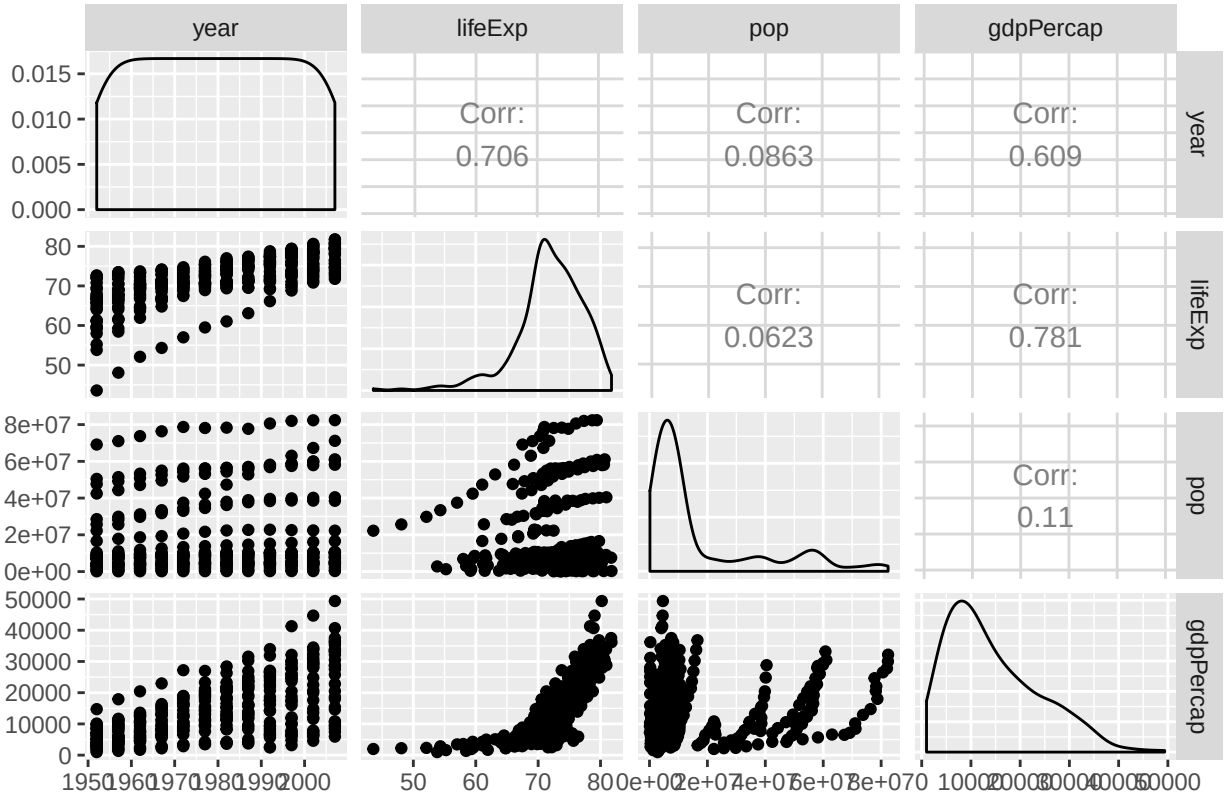
Africa


Africa − Scatter Plot between GDP, Life Expectancy, Time and Population
year
lifeExp
pop
gdpPercap
Corr:
0.547
Corr:
0.272
Corr:
0.16
Corr:
0.121
Corr:
0.426
Corr:
−0.0194
0.015
0.010
0.005
0.000
70
60
50
40
30
1e+08
5e+07
0e+00
20000
15000
10000
5000
0
1950
1960
1970
1980
1990
2000
30
40
50
60
70
0e+00
5e+07
1e+08
0
5000
10000
15000
20000

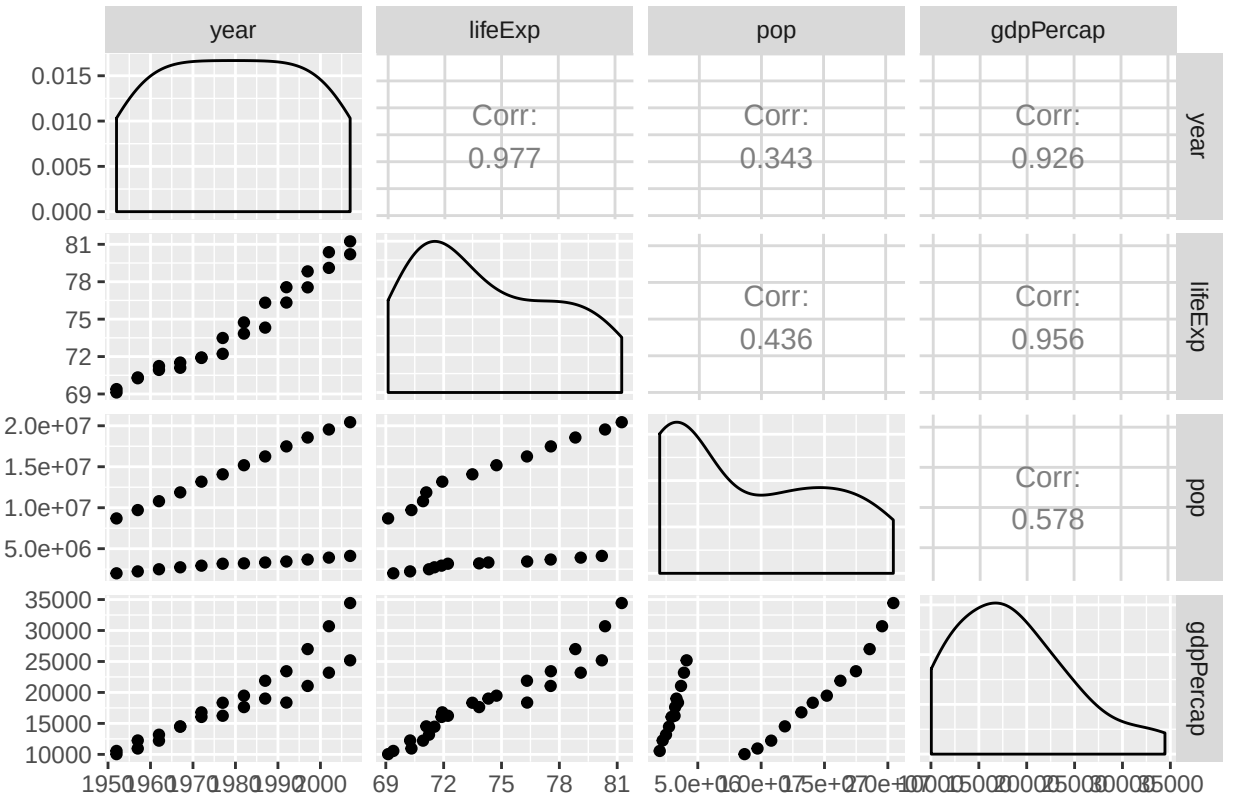
Americas



Europes


Europe − Scatter Plot between GDP, Life Expectancy, Time and Population
year
lifeExp
pop
gdpPercap
Corr:
0.706
Corr:
0.0863
Corr:
0.609
Corr:
0.0623
Corr:
0.781
Corr:
0.11
0.015
0.010
0.005
0.000
80
70
60
50
8e+07
6e+07
4e+07
2e+07
0e+00
50000
40000
30000
20000
10000
0
1950
1960
1970
1980
1990
2000
50
60
70
80
0e+00
2e+07
4e+07
6e+07
8e+07
0
10000
20000
30000
40000
50000

Oceania


Oceania − Scatter Plot between GDP, Life Expectancy, Time and Populatio
year
lifeExp
pop
gdpPercap
Corr:
0.977
Corr:
0.343
Corr:
0.926
Corr:
0.436
Corr:
0.956
Corr:
0.578
0.015
0.010
0.005
0.000
81
78
75
72
69
2.0e+07
1.5e+07
1.0e+07
5.0e+06
35000
30000
25000
20000
15000
10000
1950
1960
1970
1980
1990
2000
69
72
75
78
81
5.0e+06
1e+07
1.5e+07
2e+07
10000
15000
20000
25000
30000
35000