Does life expectancy depend on GDP?: A case study taking WDI data of India

Chandra Adhikari

Orientation:

- **Dataset Used:** World Development Indicators(WDI)
- **Data Source:** https://www.kaggle.com/worldbank/world-development-indicators
- **Objective:**

- Study of the life expectancy at birth and GDP per capita in India.
- Compare the life expectancy trend in India with that of the USA.
- Are there any correlation between the life expectancy and GDP per capita in India.

Motivation

❑ Why GDP per capita vs. life expectancy?

- Increase in GDP per capita improves the quality of life. People can access physical facilities such as health, education etc. more easily than before. If the GDP decreases over time just positive trend is expected. I want to see whether this rationale actually hold in the WDI dataset.

❑ Why I choose India?

- India is one of the densely populated country in the world with current population of about 1.35 billions (about 1/5 of the entire population of the world). The society in India is very diverse. Thus studying the data of India may lead to reasonable result and give us a better understanding.

Research Questions

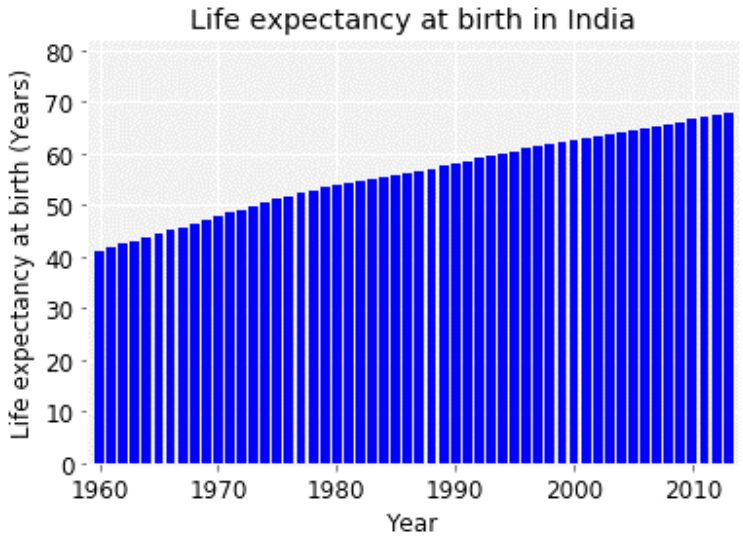
Using the World Development Indicators (WDI) dataset, I aim to answer the following questions:

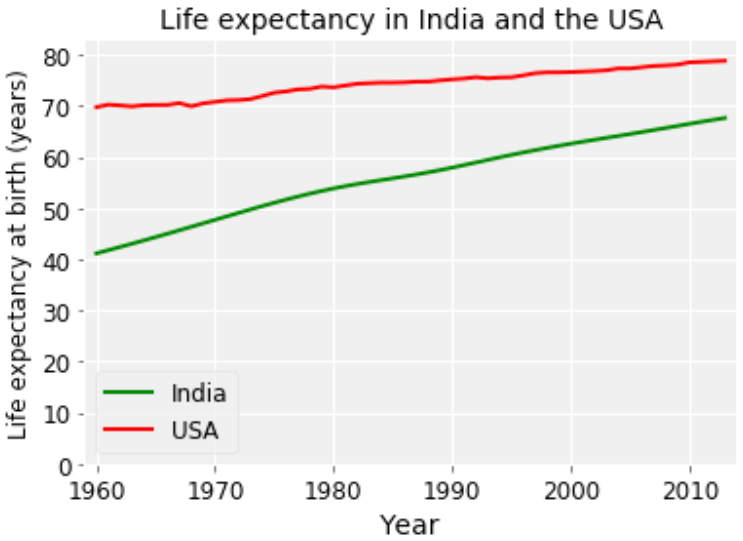
- How did the life expectancy at birth change over time in India from 1960 to 2014?
- Did the United States also have the same trend?

  Why did I compare with the trend in India with that of USA? → I believe that comparing a finding with some other reasonable reference always adds values of the finding. As the course has been designed and delivered from a University in the USA, I find the USA a reasonable reference.

- How did the GDP per capita change over time in India from 1960 to 2014?
- Was there any correlation between the life expectancy and GDP per capita in India?

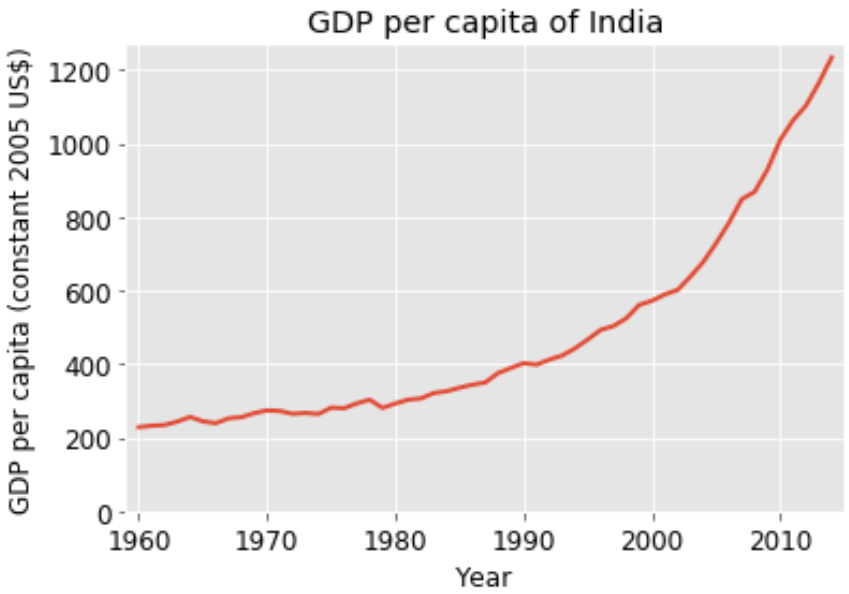
Findings: Life expectancy

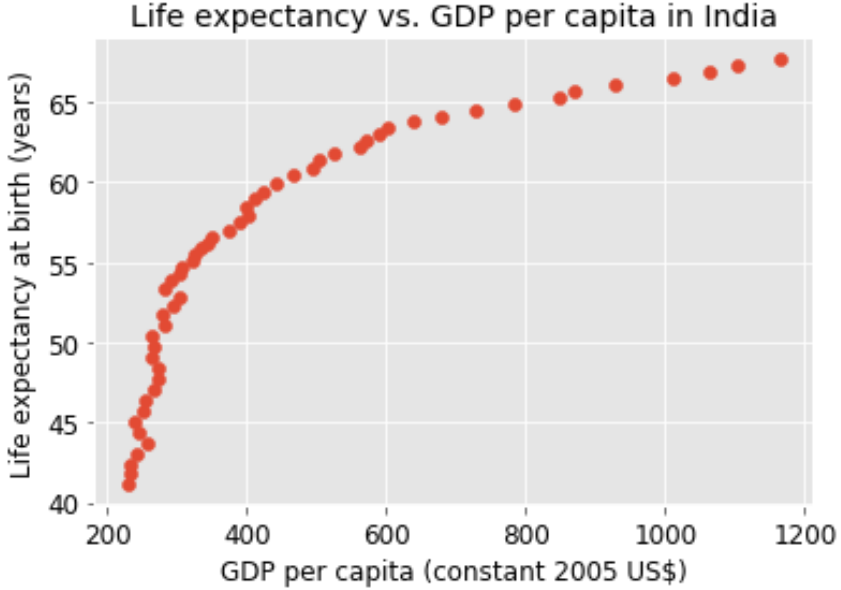




- Life expectancy at birth was about 40 years in India in 1960, which consistently increased with no drop.
- Life expectancy in the USA was about 70 years in 1960, which is approximately what it was in India in 2013.
- Life expectancy is higher in the USA than India in each year in the time frame of our data.
- Life expectancy does increase in both country (although there is slight drop in the USA in 1968), but the rate of increase in life expectancy is faster in India than in USA in the chosen time frame.

Findings: Life expectancy vs. GDP per capita





- GDP per capita of India increased from about 200 US dollar to about 1200 US dollar from 1960 to 2013.
- The GDP per capita of India increased sharply after 1990.
- Correlation of life expectancy with GDP per capita ≈ 0.85, which implies that the life expectancy seems to be strongly correlated with the GDP per capita.

Conclusions

- Both the life expectancy and the GDP per capita increased in India from 1960 to 2013.
- The life expectancy in USA also increased from 1960 to 2013, but the rate in which the life expectancy increased is higher in India.
- The life expectancy in the USA was always larger than in India in the chosen time frame.
- Focusing on the data of India, we came to a conclusion that the life expectancy at birth and the GDP per capita are strong correlated.

References

I followed the lecture videos a number of times, worked with the exercise, used hints the Professor provided at the end of a video and explored the data set. I worked in the dataset as discussed in the lectures. The lectures are so thorough that I did not need an extra source to google.

Thank you for your time

The following is the pdf version of the Jupitor notebook I played with.

Chandra_Mini_Project_Week_06_final

June 23, 2020

1 Does Life Expectancy Depend on GDP?: A Study Taking WDI Data of India

Chandra Adhikari

June 21, 2020

North Carolina, USA

1.0.1 Objective:

Study of the life expectancy at birth and GDP per capita in India.

Compare the life expectancy in India trend with that of the USA.

Are there any correlation between the life expectancy and GDP per capita in India.

Data Source: https://www.kaggle.com/worldbank/world-development-indicators

Folder: 'world-development-indicators'

1.0.2 Import packages

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib import style
```

1.0.3 Initial exploration of the Dataset

[2]:
```
df_indicator = pd.read_csv('./world-development-indicators/Indicators.csv') #␣
↪load the data
df_indicator.shape # shape of the data
```

[2]:
```
(5656458, 6)
```

[3]:
```
df_indicator[:10] # display the first 10 rows
```

[3]:
```
  CountryName CountryCode                                      IndicatorName  \
0  Arab World         ARB  Adolescent fertility rate (births per 1,000 wo…
```

```
1  Arab World         ARB  Age dependency ratio (% of working-age populat…
2  Arab World         ARB  Age dependency ratio, old (% of working-age po…
3  Arab World         ARB  Age dependency ratio, young (% of working-age …
4  Arab World         ARB        Arms exports (SIPRI trend indicator values)
5  Arab World         ARB        Arms imports (SIPRI trend indicator values)
6  Arab World         ARB               Birth rate, crude (per 1,000 people)
7  Arab World         ARB                                    CO2 emissions (kt)
8  Arab World         ARB             CO2 emissions (metric tons per capita)
9  Arab World         ARB  CO2 emissions from gaseous fuel consumption (%…
```

	IndicatorCode	Year	Value
0	SP.ADO.TFRT	1960	1.335609e+02
1	SP.POP.DPND	1960	8.779760e+01
2	SP.POP.DPND.OL	1960	6.634579e+00
3	SP.POP.DPND.YG	1960	8.102333e+01
4	MS.MIL.XPRT.KD	1960	3.000000e+06
5	MS.MIL.MPRT.KD	1960	5.380000e+08
6	SP.DYN.CBRT.IN	1960	4.769789e+01
7	EN.ATM.CO2E.KT	1960	5.956399e+04
8	EN.ATM.CO2E.PC	1960	6.439635e-01
9	EN.ATM.CO2E.GF.ZS	1960	5.041292e+00

Country name and country code are the same, the year is also the same, but indicators are many. I would like to see on the GDP per capita and the life expentancy. Try to see a few more rows to see more the IndicatorName.

1.0.4 Let us make sure we have the data of India:

[4]:
```
df_indicator['CountryCode'].str.contains("IND").any()
```

[4]:
```
True
```

[5]:
```
df_indicator['CountryName'].str.contains("India").any()
```

[5]:
```
True
```

1.0.5 How many years of data do we have ?

[6]:
```
# How many years of data do we have ?
years = df_indicator['Year'].unique().tolist()
len(years)
```

[6]:
```
56
```

1.0.6 What's the range of years?

[7]:
```
print(min(years)," to ",max(years))
```

```
1960  to  2015
```

1.0.7 Lets pick a country and an indicator to explore: Life expentancy

[8]:
```
# select Life expectancy for India
hist_indicator = 'Life expectancy at birth, total \(years\)'
hist_country = 'IND'

m_LE = df_indicator['IndicatorName'].str.contains(hist_indicator)
m_CC = df_indicator['CountryCode'].str.contains(hist_country)

# df_LE is just those indicators matching the USA for country code and Life␣
 ↪expectancy at birth, total (years) over time.
df_LE= df_indicator[m_LE & m_CC]
```

[9]:
```
df_LE.head()
```

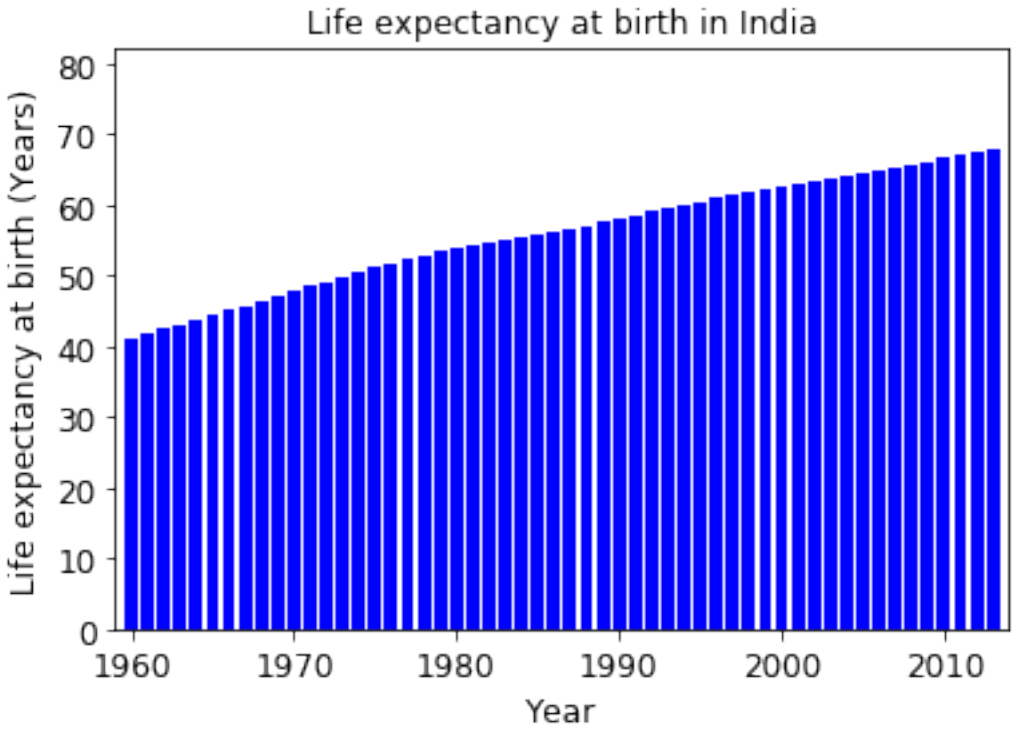
[9]:

	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
11684	India	IND	Life expectancy at birth, total (years)	SP.DYN.LE00.IN	1960	41.171951
36635	India	IND	Life expectancy at birth, total (years)	SP.DYN.LE00.IN	1961	41.790488
64177	India	IND	Life expectancy at birth, total (years)	SP.DYN.LE00.IN	1962	42.417415
92622	India	IND	Life expectancy at birth, total (years)	SP.DYN.LE00.IN	1963	43.052732
121419	India	IND	Life expectancy at birth, total (years)	SP.DYN.LE00.IN	1964	43.698415

1.0.8 How life expectancy have changed over time?

[10]:
```
# get the years
years =df_LE['Year'].values
# get the values
LifeExp = df_LE['Value'].values

# create
plt.xlabel('Year',size=12, color='k')
plt.ylabel('Life expectancy at birth (Years)',size=12, color='k')
plt.bar(years,LifeExp,color='b')
plt.xticks(fontsize=12,color='k')
```

```
plt.yticks(fontsize=12,color='k')
plt.title('Life expectancy at birth in India')
plt.axis([1959, 2014,0,82])
plt.show()
```



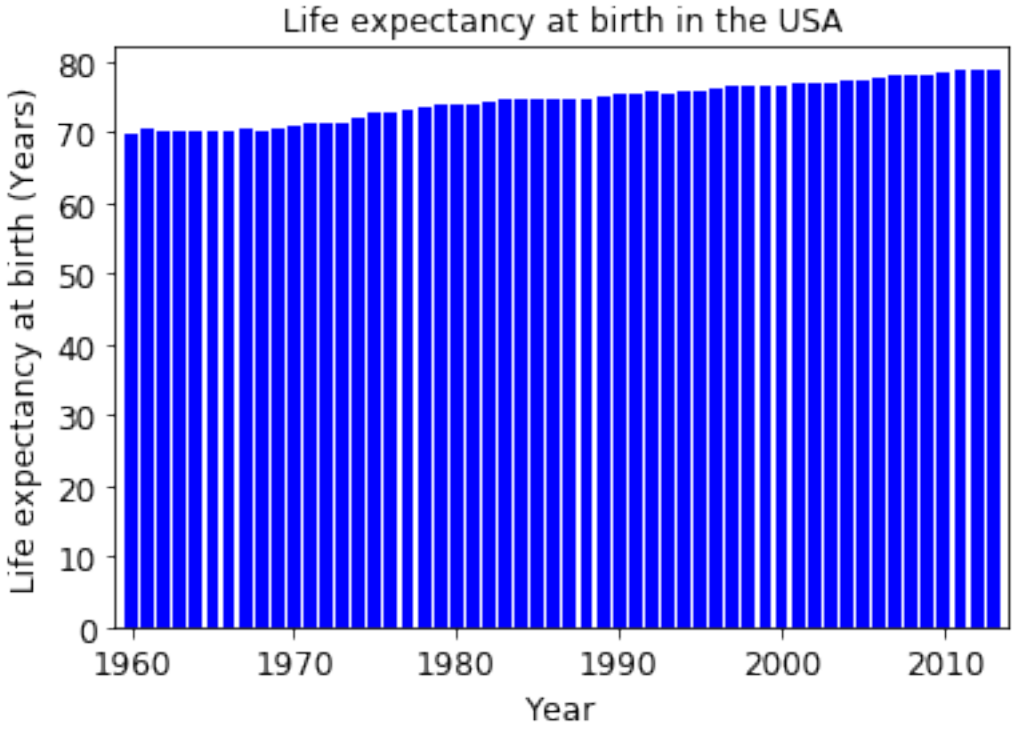
Turns out Life expectancy at birth have been consistently increaseing. From 1960 to 2014, the life expectancy increased consistently from about 40 to 70 years.

1.0.9 How about life expectancy in USA? Does the USA have the same trend?

[11]:
```
# select Life expectancy for the USA
hist_indicator_2 = 'Life expectancy at birth, total \(years\)'
hist_country_2 = 'USA'

m_LE_2 = df_indicator['IndicatorName'].str.contains(hist_indicator_2)
m_CC_2 = df_indicator['CountryCode'].str.contains(hist_country_2)

# df_LE_2 is just those indicators matching the USA for country code and Life␣
 ↪expectancy at birth, total (years) over time.
df_LE_2= df_indicator[m_LE_2 & m_CC_2]
```

[12]:
```
# get the years
years_2 =df_LE_2['Year'].values
```

```
# get the values
LifeExp_2 = df_LE_2['Value'].values

# create
plt.xlabel('Year',size=12, color='k')
plt.ylabel('Life expectancy at birth (Years)',size=12, color='k')
plt.bar(years_2,LifeExp_2,color='b')
plt.xticks(fontsize=12,color='k')
plt.yticks(fontsize=12,color='k')
plt.title('Life expectancy at birth in the USA')
plt.axis([1959, 2014,0,82])
plt.show()
```

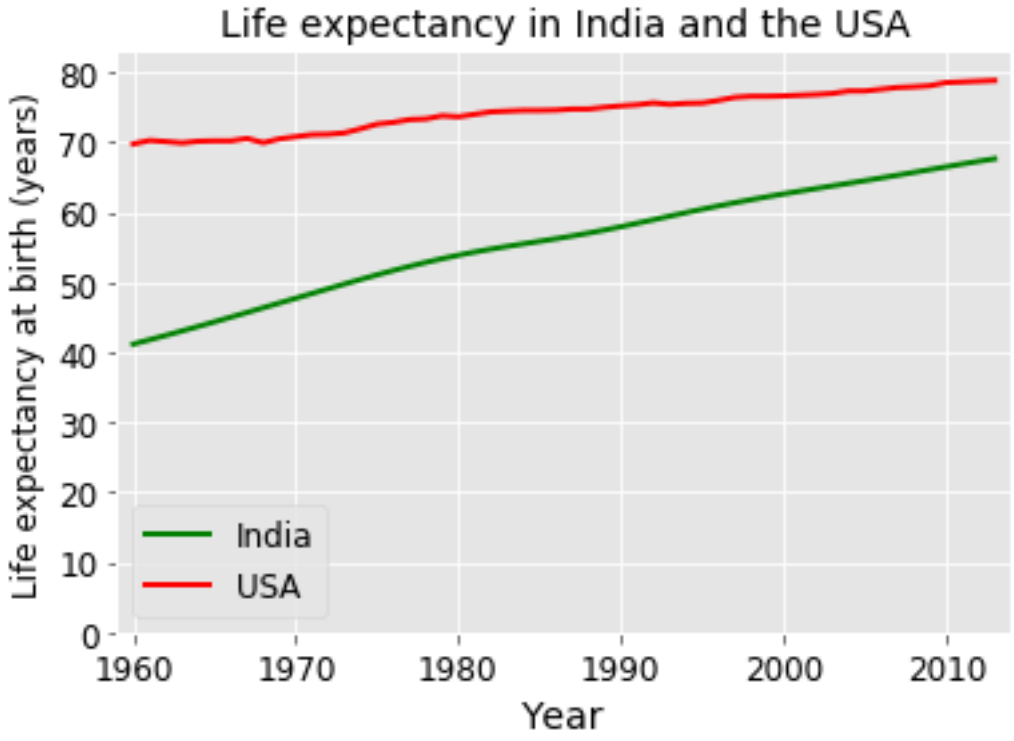


Yes, the life expectancy has increased over the year but the rate of increase in life ecpectancy is not as fast as in India. Indeed, it was about 70 years in the USA in 1960, which was only 40 years in India at that time.

1.0.10 Let's get line plots in the same graph to make a comparision easy.

[13]:
```
# switch to a line plot
style.use('ggplot')
plt.plot(df_LE['Year'].values, df_LE['Value'].values,'g',label="India",␣
 ↪linewidth=2)
```

```
plt.plot(df_LE_2['Year'].values, df_LE_2['Value'].values,'r',label="USA",␣
 ↪linewidth=2)


# Label the axes
plt.xlabel('Year',size=14,color='k')
#plt.ylabel(df_LE['IndicatorName'].iloc[0],size=12)
plt.ylabel('Life expectancy at birth (years)',size=12,color='k')

#label the figure
plt.title('Life expectancy in India and the USA', size=14,color='k')
plt.xticks(fontsize=12,color='k')
plt.yticks(fontsize=12,color='k')

#show legend
plt.legend(fontsize=12)
plt.grid(True, color='w')


# to make more honest, start they y axis at 0
plt.axis([1959, 2014,0,83])

plt.show()
```



1.1 Relationship between Life expectancy and GDP per capita in India

[14]:
```
# select GDP Per capita  for India
hist_indicator_G = 'GDP per capita \(constant 2005'
hist_country_G = 'IND'

mk_G1 = df_indicator['IndicatorName'].str.contains(hist_indicator_G)
mk_G2 = df_indicator['CountryCode'].str.contains(hist_country_G)

gdp_India = df_indicator[mk_G1 & mk_G2]

#plot gdp_India vs df_LE
```

[15]:
```
gdp_India.head(2)
```

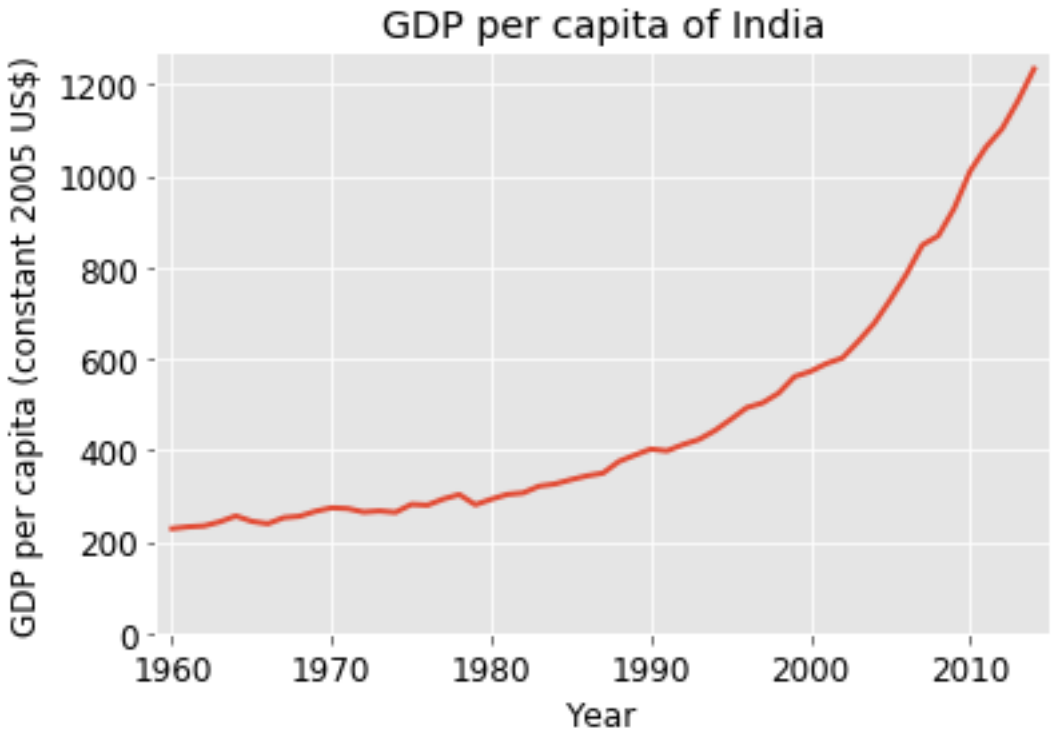
[15]:
```
      CountryName CountryCode                      IndicatorName  \
11616       India         IND  GDP per capita (constant 2005 US$)
36555       India         IND  GDP per capita (constant 2005 US$)

        IndicatorCode  Year       Value
11616  NY.GDP.PCAP.KD  1960  228.304470
36555  NY.GDP.PCAP.KD  1961  232.142053
```

[16]:
```
df_LE.head(2)
```

[16]:
```
      CountryName CountryCode                          IndicatorName  \
11684       India         IND  Life expectancy at birth, total (years)
36635       India         IND  Life expectancy at birth, total (years)

        IndicatorCode  Year      Value
11684  SP.DYN.LE00.IN  1960  41.171951
36635  SP.DYN.LE00.IN  1961  41.790488
```

[17]:
```
# switch to a line plot
plt.plot(gdp_India['Year'].values, gdp_India['Value'].values,linewidth=2)

# Label the axes
plt.xlabel('Year',fontsize=12,color='k')
plt.ylabel(gdp_India['IndicatorName'].iloc[0],fontsize=12,color='k')

#label the figure
plt.title('GDP per capita of India')
plt.xticks(fontsize=12,color='k')
plt.yticks(fontsize=12,color='k')

# to make more honest, start they y axis at 0
plt.axis([1959, 2015,0,1270])
```

```
plt.show()
```



GDP per capita increases with years.

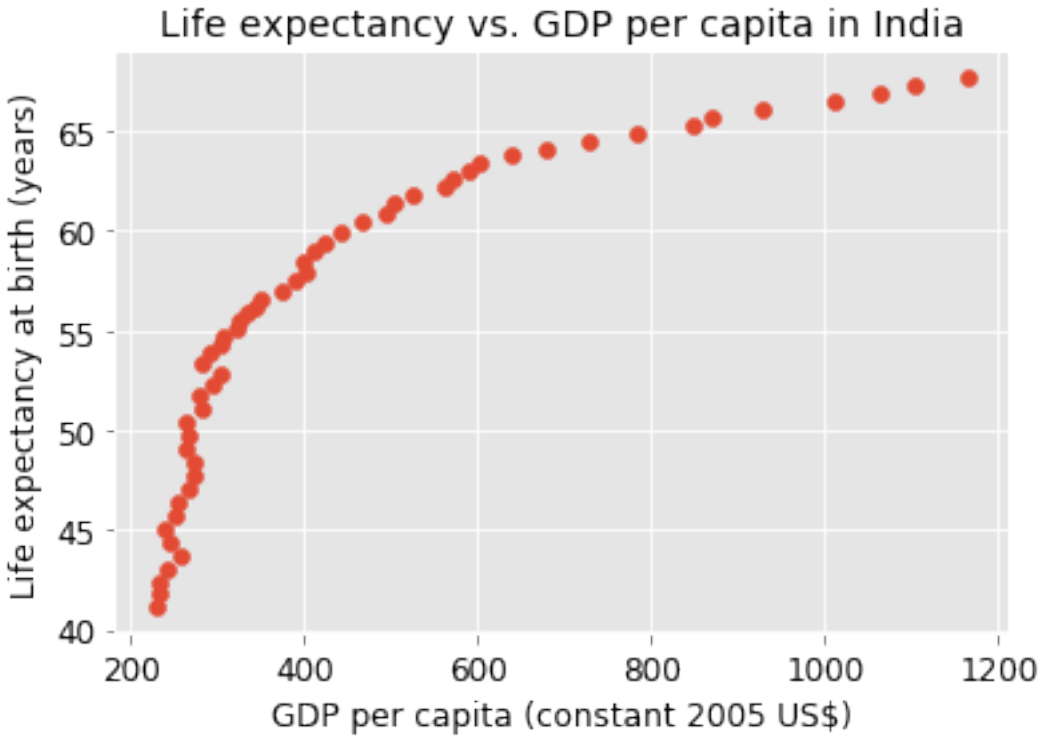
1.1.1 Draw a scatter plot to see corelation between GDP per capita and life expectancy.

[18]:
```
gdp_India_Trunc = gdp_India[gdp_India['Year'] < 2014]
print(len(gdp_India_Trunc))
print(len(df_LE))
```

```
54
54
```

[19]:
```
%matplotlib inline
import matplotlib.pyplot as plt

fig, axis = plt.subplots()
# Grid lines, Xticks, Xlabel, Ylabel

axis.yaxis.grid(True)
```

```
axis.set_title('Life expectancy vs. GDP per capita in India',fontsize=14,␣
↪color='k')
axis.set_xlabel(gdp_India_Trunc['IndicatorName'].iloc[0],fontsize=12, color='k')
axis.set_ylabel('Life expectancy at birth (years)',color='k',fontsize=12)
plt.xticks(fontsize=12,color='k')
plt.yticks(fontsize=12,color='k')

X = gdp_India_Trunc['Value']
Y = df_LE['Value']


axis.scatter(X, Y)
plt.show()
```



This plot shows a good relationship. But how strong is it? We can test this by looking at correlation.

[20]:
```
np.corrcoef(gdp_India_Trunc['Value'],df_LE['Value'])
```

[20]:
```
array([[1.        , 0.84850089],
       [0.84850089, 1.        ]])
```

The correlation of 0.85 is pretty strong.

We come to the End of the Jupiter notebook. **Thank You!**