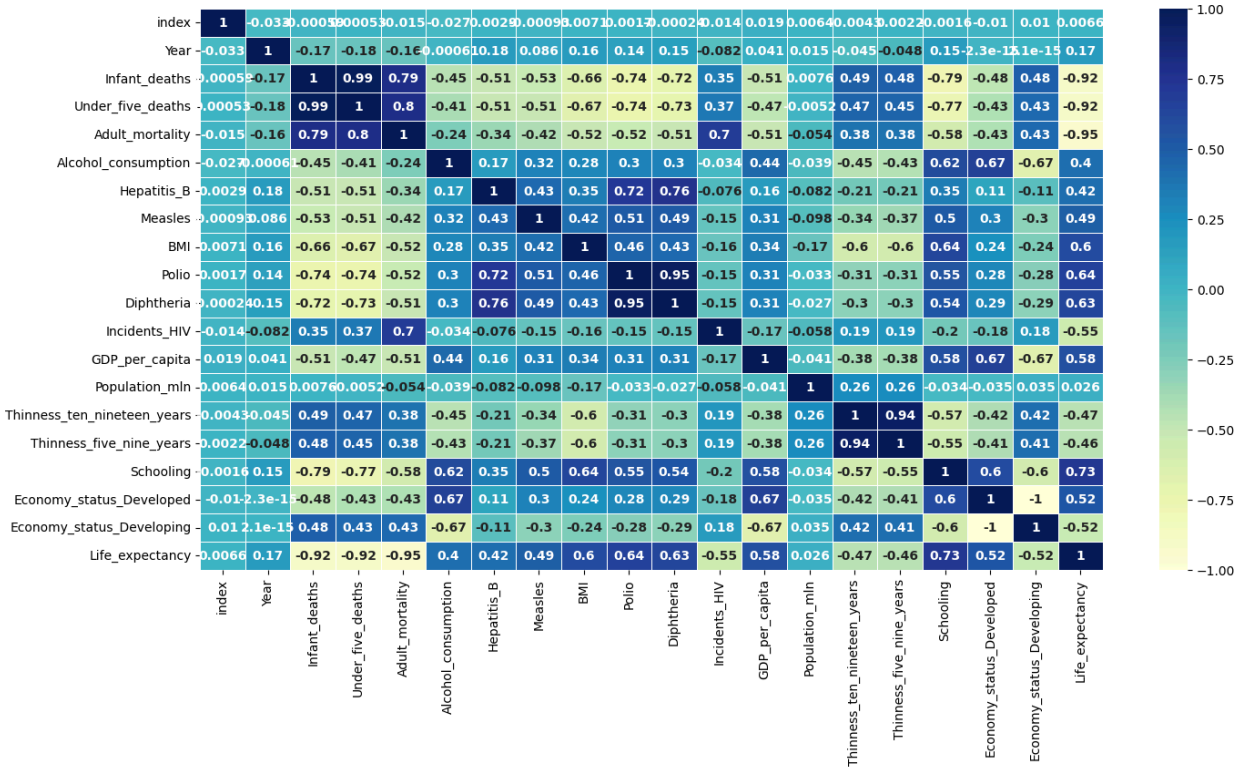
Appendix - A Contrarian View

Contact: Frank Miceli

There was some discussion that was not fully resolved. Below are a few thoughts that propose an alternative view showing that GDP per capita is not the best indicator of Life expectancy.

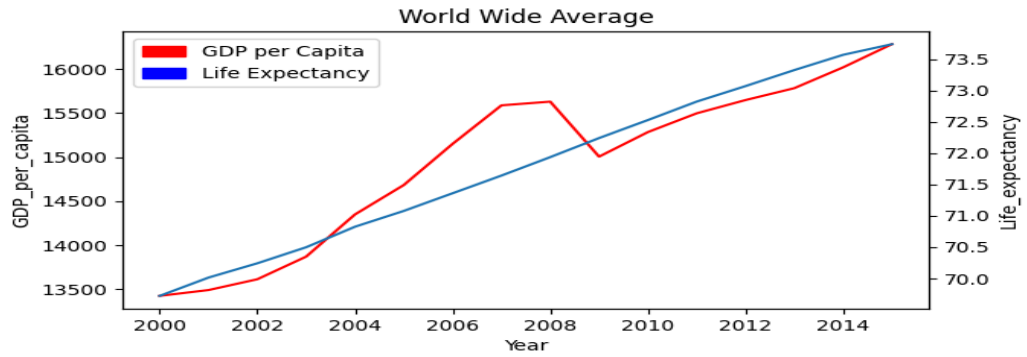
Without overloading the reader, first please review the correlation matrix of the data set.



While I am not discounting the original findings discussed above, what is not mentioned is that the highest correlation of life expectancy is a negative relationship of a combination of Infant_deaths, (-0.92), Under_five_deaths (-0.92) and Adult_mortality (-0.95).

The correlation of Life_expectancy and GDP_per_capita is much less at + 0.58.

Exploring the original conclusion, a world wide visual representation is provided as well as each region's relationship of Life_expectancy and GDP_per_capita per region from 2000 to 2015.



Definitions are key for proceeding further.

	In Kaggle DataSet	In Referenced DataSet	Estimate or Observed Value
Life_expectancy	Average life expectancy of both genders in different years from 2010 to 2015	Life Expectancy at birth (years)	Estimated
GDP_per_capita	GDP per capita in USD		Observed
Infant_deaths	deaths per 1000 population		Observed
Under_five_deaths	deaths per 1000 population		Observed
Adult_mortality	deaths per 1000 population		Observed

Quote from WorldBank.org

"The statistic "Life expectancy at birth" actually refers to the average number of years a newborn is expected to live *if mortality patterns at the time of its birth remain constant in the future"*

"**Life expectancy at birth** is the total person-years lived beyond exact age 0 divided by the number of newborns"

As I understand the process of calculating life expectancy, actuarial tables using a probability of dying and the current mortality rates result in the life expectancy estimate.

By comparison, does an observed increase or decrease of GDP_per_capita impact Life expectancy estimates?

GDP_per_capita is a country's GDP divided by its population. GDP = C + I + G + (X-M). That is, GDP = Consumer Spending + Business Investment + Government Spending + (Exports - Imports).

Governments will allocate funds to its population for health care, based on its policies. But allocation of GDP to Health Care is unknown in this dataset. It would be best to include these proportions in future analysis.

Below are a few excerpts from an OLS of various relations:

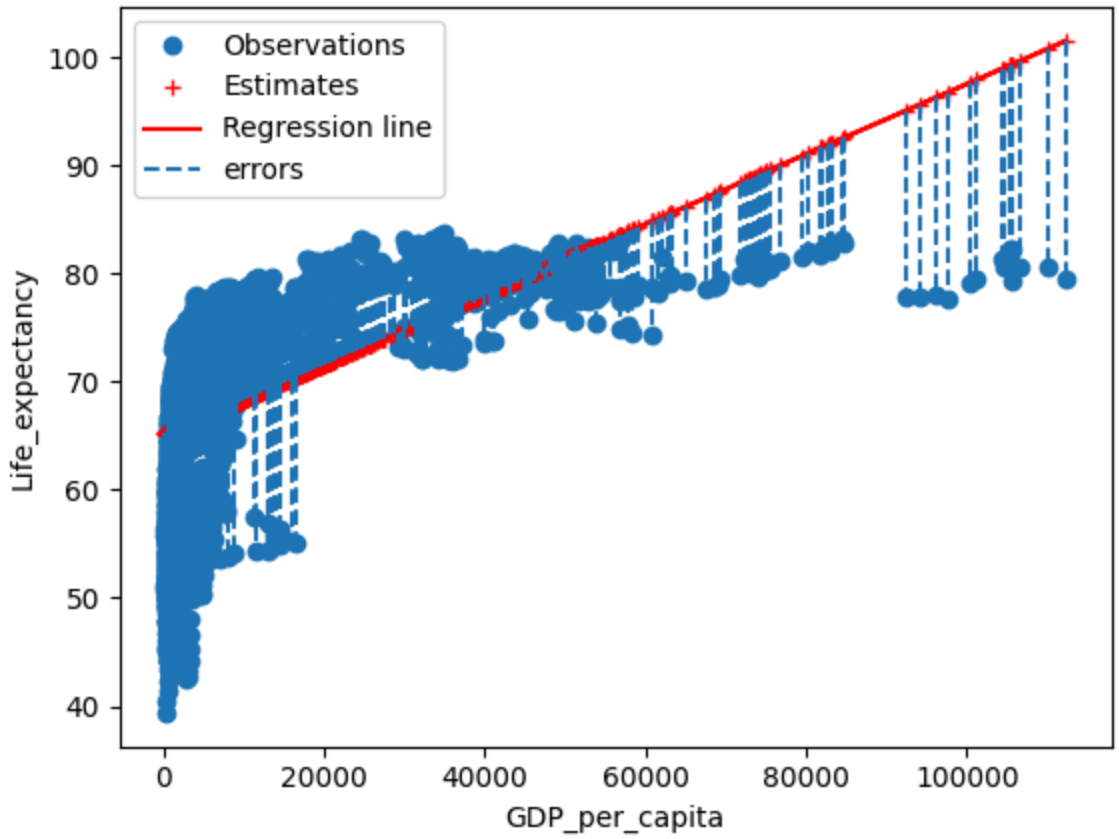
First, OLS of Life_expectancy vs GDP_per_capita

OLS Regression Results

Dep. Variable:	Life_expectancy	R-squared:	0.340
Model:	OLS	Adj. R-squared:	0.340
Method:	Least Squares	F-statistic:	1474.
Date:	Sun, 11 Aug 2024	Prob (F-statistic):	1.52e-260
Time:	15:18:59	Log-Likelihood:	-9887.4
No. Observations:	2864	AIC:	1.978e+04
Df Residuals:	2862	BIC:	1.979e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	65.1186	0.173	376.787	0.000	64.780	65.457
GDP_per_capita	0.0003	8.43e-06	38.397	0.000	0.000	0.000

Omnibus:	323.832	Durbin-Watson:	2.015
Prob(Omnibus):	0.000	Jarque-Bera (JB):	443.462
Skew:	-0.961	Prob(JB):	5.05e-97
Kurtosis:	3.158	Cond. No.	2.48e+04



Alternative approach, Life Expectancy vs Mortality

where mortality includes Infant_deaths, Under_five_deaths and Adult_mortality

OLS Regression Results

Dep. Variable:	Life_expectancy	R-squared:	0.971
Model:	OLS	Adj. R-squared:	0.971
Method:	Least Squares	F-statistic:	3.220e+04
Date:	Sun, 11 Aug 2024	Prob (F-statistic):	0.00
Time:	20:05:49	Log-Likelihood:	-5400.5
No. Observations:	2864	AIC:	1.081e+04
Df Residuals:	2860	BIC:	1.083e+04
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	82.6150	0.066	1248.176	0.000	82.485	82.745
Infant_deaths	-0.1273	0.006	-19.829	0.000	-0.140	-0.115
Under_five_deaths	-0.0190	0.004	-4.721	0.000	-0.027	-0.011
Adult_mortality	-0.0472	0.000	-108.468	0.000	-0.048	-0.046

```
Omnibus:                        1.934   Durbin-Watson:                   2.026
Prob(Omnibus):                  0.380   Jarque-Bera (JB):                1.862
Skew:                           0.048   Prob(JB):                        0.394
Kurtosis:                       3.080   Cond. No.                         520.
==============================================================================
```

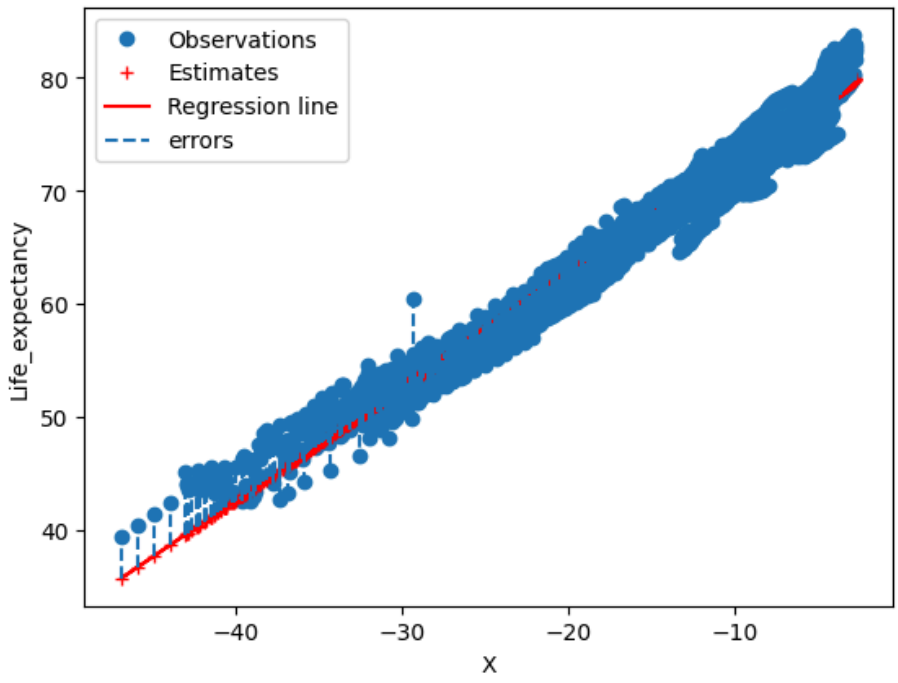
Because this regression is leveraging 3 different independent variables, a summary value called X was created to represent the parameters of the regression residuals in the plot below.

In this case,

```
dfa['X'] = dfa['Infant_deaths'] * reg1.params['Infant_deaths'] +
        dfa['Under_five_deaths']*reg1.params['Under_five_deaths'] +
        dfa['Adult_mortality']*reg1.params['Adult_mortality']
```

where reg1.params =

```
Intercept            82.615036
Infant_deaths        -0.127321
Under_five_deaths    -0.019050
Adult_mortality      -0.047204
dtype: float64
```



Concluding, this alternative is to say that while there is a moderate corelation between GDP and Life Expectancy, there are other factors related to how GDP directly impacts Life Expectancy Estimation such as government spending, and policies of each nation on how and where to fund health that remain unknown. The data promotes a better relationship between Life Expectancy and the mortality metrics of Infant_deaths, Under_five_deaths and Adult_mortality.