Association of life expectancy with other explanatory variables for different countries from the GapMinder dataset

A **correlation analysis** was conducted on the **GapMinder** dataset to understand the association of 14 explanatory variables (including income per person, alcohol consumption, armed forces rate, breast cancer per 100th, co2 emissions, female employment rate, hiv rate, internet use rate, oil per person, polity score, relectric per person, suicide per 100th, employment rate, urbanization rate) with the variable life expectancy.

After removing the observations with missing values the **Pearson correlation coefficient** is computed. As can be seen from the below results, the variable *internetuserate* has a very strong positive (perfect) correlation with the variable *life expectancy*. The variable *incomeperperson* also has a strong positive correlation with *life expectancy*. The variable *hivrate* is the variable most negatively associated with the variable life expectancy. The variable *suicideper100th* has the least corrleation with life expectancy. Also, the corrsponding p-values(with the null hypothesis that the variables are not corrlelated) are reported. All the variables except hivrate, armedforcesrate, suicideper100th, co2emissions and employrate have **statistically significant** correlations at **5%** level of significance.

Correlation coefficients with the **lifeexpectancy** variable

variables	pearson-r	p-value
hivrate	-0.226524	9.318806e-02
armedforcesrate	-0.116446	3.927408e-01
suicideper100th	0.072865	5.935585e-01
co2emissions	0.143092	2.927631e-01
employrate	0.190514	1.595943e-01
femaleemployrate	0.386966	3.217162e-03
polityscore	0.388039	3.125989e-03
oilperperson	0.441645	6.557597e-04
alcconsumption	0.465096	3.043404e-04
urbanrate	0.628376	2.160189e-07
relectricperperson	0.717309	4.980896e-10
breastcancerper100th	0.745454	4.390452e-11
incomeperperson	0.823008	7.092079e-15
internetuserate	1.000000	0.000000e+00