DIT852
Assignment 1
Erik Rosvall 960523
Christoffer Wikner 931012

1.

a) The dataset life expectancy and the dataset GDP per capita are both downloaded from www.ourworldindata.org. We as that authors had an underlying assumption that whilst not linear GDP per capita would be a strong correlation to life expectancy.
As we interpret the question as seeing if the correlation exists as a function of modern economies rather than of all time we decided to use a dataset for GDP per capita that started in 1990 and went forward to 2019. We went through the data to see that values weren't missing and we decided to remove region-based stats and stick to countries as we saw this as a narrowing of the data which would be better for answering the question. We merged the life expectancy that set through using amongst other things year so the data and vectors were equally long.

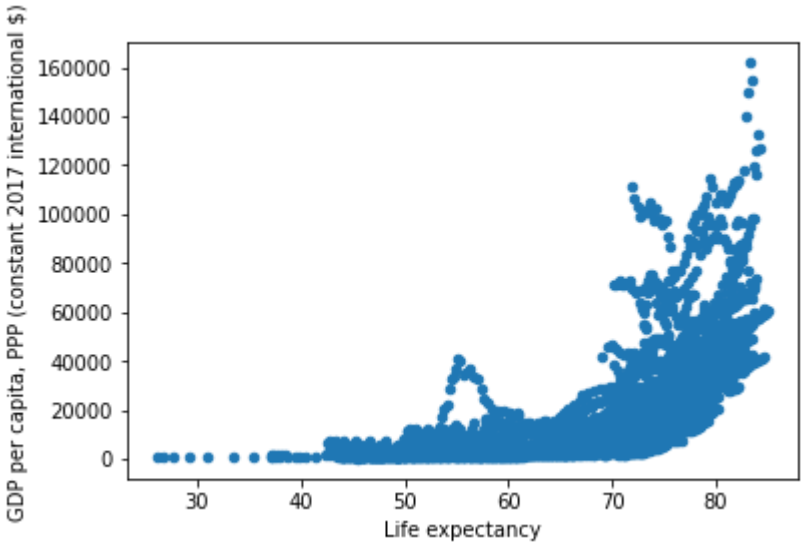

Figure 1. GDP per capita on the y-axis and life expectancy on the x-axis

b) The assumption made in question 1 a) about there being a strong correlation between GDP per capita and life expectancy seemed to be validated. However, it is notable that there are several countries with extremely high GDP per capita in the range of 80-120 thousand which still however in the mid-'70s as the average for life expectancy. This contrasted to a considerable amount of countries having a GDP per capita of circa 20-40 thousand yet still surpassing the aforementioned countries >5 years.

A possible explanation is that there is a minimum GDP per capita needed for basic human needs which greatly increase life expectancies such as food, shelter, and clean

water. Thereafter the correlation doesn't become linear, a possible reason why it doesn't increase as linearly is that while a country might have a high GDP per capita it might still have a very small amount of resources and a small population, this position doesn't make cumulative tax incomes enough to strengthen infrastructure. A country that has 1k people with 100k GDP per capita at a 30 % tax bracket would bring in 30000000 in taxes. This whilst a country with 30k GDP per capita but a population of 100k would bring in 1000000000. These infrastructure changes would be needed to provide for health care, education, and other things that are most likely important for life expectancy.

c) We decided to remove entities that were not countries but rather amalgamations of several different countries such as the European Union. We initially thought of dropping all countries which had not to code since that was the case for most however exceptions such as the world with the code OWID_WRL made this a non-valid strategy and thus did manual data removal by finding the non-countries in the data-set. We also went through the data looking for missing values on countries as well as looking at countries seeing if any year of their GDP per capita or life expectancy jumped out which could happen with misplaced or missing commas i.e 18.000 becoming 1800

   For question 2 of the assignment, we did the following cleaning. We cleaned all the country data sets after we merged it with the life expectancy data set. This made sure not to include years from life expectancy that didn't have a match with the other data..

d) Life expectancy mean = 68.74444781818181
   Life expectancy standard deviation = 9.639146337043176
   Mean + 1*standard deviation = 78.38359415522498
   Countries with a life expectancy greater than 78.38359415522498

   'Albania', 'Australia', 'Austria', 'Barbados', 'Belgium' Bermuda', 'Canada', 'Cayman Islands', 'Chile', 'Costa Rica' 'Croatia', 'Curacao', 'Cyprus', 'Czechia', 'Denmark', 'Estonia', 'Finland', 'France', 'Germany', 'Greece', 'Hong Kong', 'Iceland', 'Ireland', 'Israel', 'Italy', 'Japan', 'Lebanon', 'Luxembourg', 'Macao', 'Maldives', 'Malta', 'Netherlands', 'New Zealand', 'Norway', 'Panama', 'Poland', 'Portugal', 'Puerto Rico', 'Qatar' 'San Marino', 'Singapore', 'Sint Maarten (Dutch part)', 'Slovenia', 'South Korea', 'Spain', 'Sweden', 'Switzerland' 'Turks and Caicos Islands', 'United Kingdom', 'United States

e) The mean of life expectancy + 1 standard deviation is what we consider a high life expectancy
   Life expectancy mean = 68.74444781818181
   Life expectancy standard deviation = 9.639146337043176

Mean + 1*standard deviation = 78.38359415522498

The mean of GDP Per Capita * 0.85 is what we consider a weak economy.
GDP per capita mean = 17890.146722229907
17890.146722229907 * 0.85 = 15206.62471389542

Results: Albania', 'Barbados', 'Lebanon'

f) We define a high GDP as the mean + standard_deviation
Mean = 256975907393.24448
Standard Deviation = 1013933233004.3193

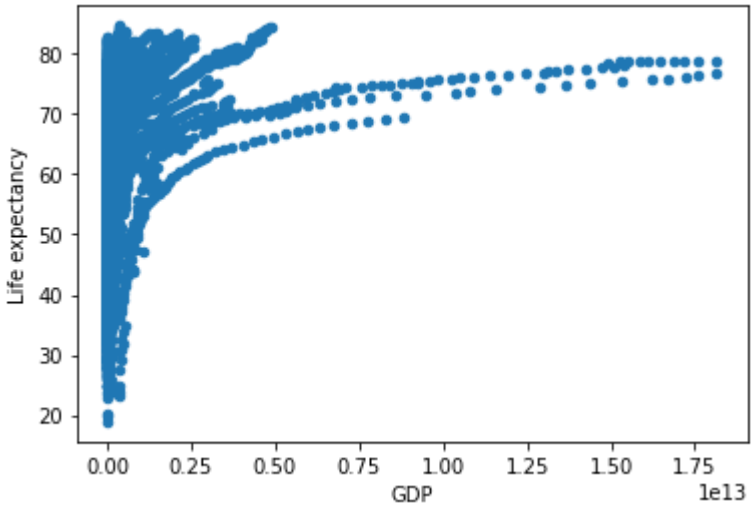

Figure 2. Life expectancy at the y-axis and GDP at the x-axis

Based on what we consider a strong economy with the mean + standard deviation every country which has a GDP of >= the threshold does not have a high life expectancy based on the previous metric of high being >= 78.38359415522498.

There is a very strong correlation as evident by figure 2 between GDP and life expectancy however the question asked in 1 f) is “Does every strong economy (normally indicated by GDP) have a high life expectancy?” and based on the framework of which we use to determine a strong economy and a high life expectancy to answer whether every country with a high GDP has a high life expectancy the answer is no. It’s an extremely strong correlation but it is not *every* country.

g)

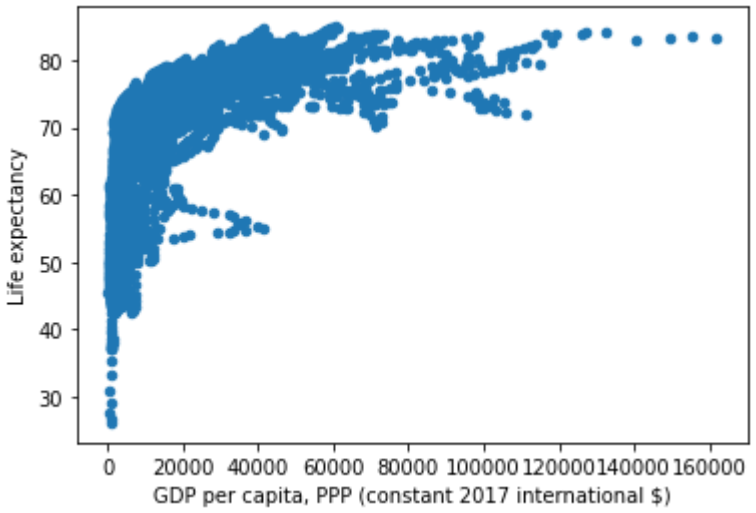

Figure 3. Life expectancy at the y-axis and GDP per capita at the x-axis

We define a high GDP per capita as the mean + 2*standard_deviateions
17890.146722229907 + 2 * 20381.59218542165 = 58653.331093073204

As with GDP, GDP per capita seems highly correlated to life expectancy but to answer the question of whether *every* country with a strong economy (GDP per capita mean + 2 standard deviations) has a high life expectancy the answer is no. As the graph illustrates, the intersection of the y and x-axis at 58653.33109 and life exp is about 72, 73 years which is lower than the previously stated standard of >= 78.38359415522498.

So the answer is with GDP per capita as with GDP based on the parameters we set up to define what is a strong economy and a high life expectancy *every* country with a strong economy does not have a high life expectancy.

2.

a) For this analysis, we decided to investigate the correlation between GDP per capita, Intentional homicides per capita, government expenditure on education, total (% of government expenditure).

It's interesting to investigate whether a country is investing more into education (not in raw numbers but rather as % of government expenditure) and its impact on GDP per capita and the intentional homicide rate per capita. We acknowledge that it's possible and a potential problem that a country might have spent a large % of their government expenditure on education but effects in the data on either GDP per capita or homicides per capita thus not giving an accurate level of correlation.

b)

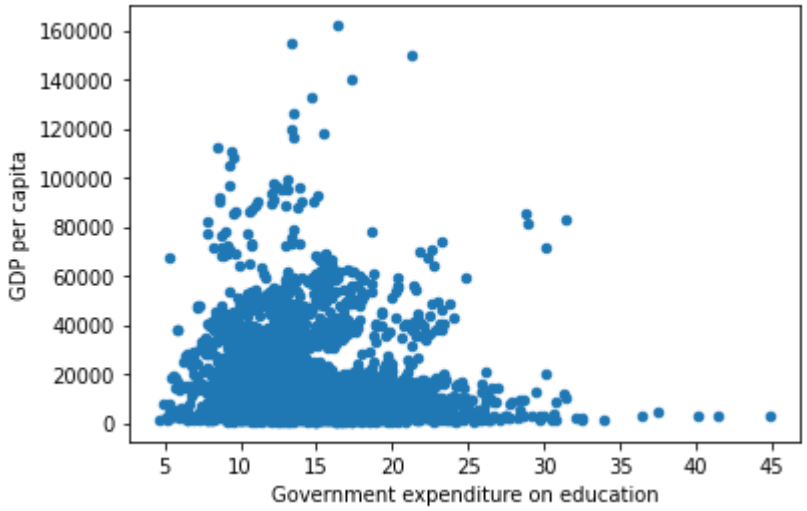

Figure 4. GDP per capita at the y-axis and education as a % of total government expenditure on the x-axis

Figure 4 shows the relationship between education as a % of total government expenditure and GDP per capita. Education as a % of total government expenditure is measured from 0 to 100 %. From the figure, it can be assumed that there is not a very strong correlation between these 2 variables. A possible explanation could be that whilst a country might have a high % total government expenditure being education, that is still a very small amount either in total or compared to the country's population. This as some countries only have extremely basic infrastructure such as schools, hospitals, and police and more basic functions and do not provide for example free health-care, a large welfare state, and other things that make the governments budget larger by default and making education a smaller % of that expenditure but still provides a very strong educational system.

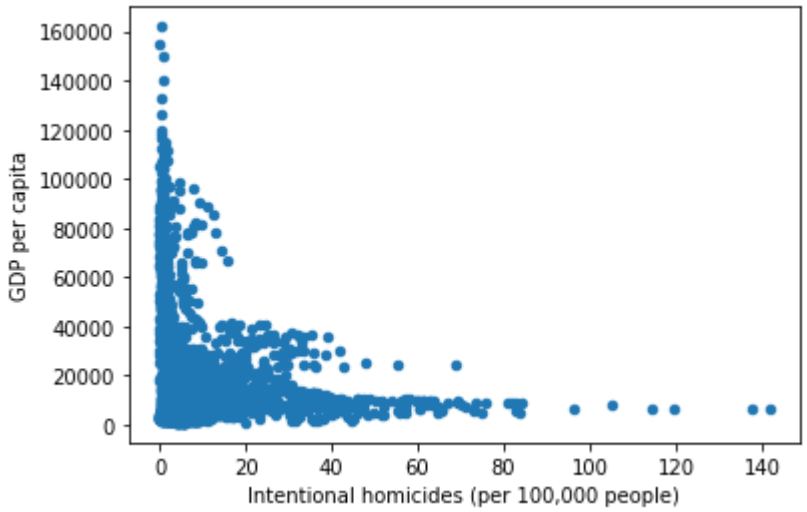

Figure 5. GDP per capita at the y-axis and Intentional homicides (per 100.000) at the x-axis

Figure 5 shows the relation between GDP per capita and intentional homicides per capita. GDP per capita and intentional homicides per capita is not measured in a range of 0-100 or any similar metric to the previous example but rather just as in the name per capita. Per capita is a strong indicator when evaluating a country's rate of education, wealth, and other means as with pure GDP a country might have a high GDP but still be very poor because it has an enormous population.

Unlike figure 4, figure 5 provides us with great insight into the strength of the relation between GDP per capita and homicides per capita. All countries with homicides per capita being >= 40 have a GDP per capita of <= 40.000. All the countries with homicides of >= 80 have a GDP per capita of <= 25.000. The inverse is also true where all countries with a GDP per capita of >= 100.000 have a homicide rate of <= 20. From the data above there seems to be enough to say that there is a strong correlation between a higher GDP per capita and a lower homicide rate per capita and the inverse relation is also true. This could warrant further exploration.

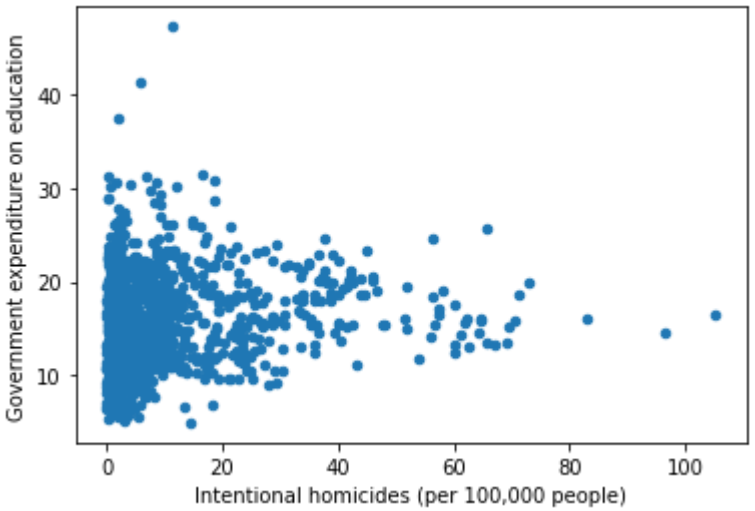

Figure 6. Education as a % of total government expenditure and intentional homicides (per 100 000) at the x-axis

Figure 5 shows the relation between intentional homicides per capita (100.000) and education as a % of total government expenditure. When viewing the chart there seems to be some level of correlation as all of the countries with homicides per capita being >= 50 have education as a % of total government expenditure <= 30 %. However, there is a large collection of data points in the lower half of the left corner where education as a % of total government expenditure moves between 5 to 20 % whilst having seemingly little effect on the homicide rate per capita. Several countries which invest almost nothing as a % of government expenditure on education still have a very low homicide rate per capita. This could as previously explained however be that despite the % being low it is still a very large amount in absolute numbers.