This project will examine datasets available at Gapminder. To be more specific, it will take a closer look on the life expectancy of the population from different countries and the influences from other variables. It will also take a look on the development of these variables and the goal is to find trends which are related to a higher life expectancy.

What is Gapminder? "Gapminder is an independent Swedish foundation with no political, religious or economic affiliations. Gapminder is a fact tank, not a think tank. Gapminder fights devastating misconceptions about global development." (https://www.gapminder.org/about-gapminder/)

Posing Questions
The following analysis will take a look on the following questions:

How did the world poplation changed over time?

How did the life expectancy and the income per person changed over time?

Is the life expectancy somehow related to the income?

With the actual shape of the data, we could easily compare different years of the given data, which we should remember for later. On the other hand the data is not tidy enough to make good visializations or calculations. The next step will handle this problem by melting the dataframe, so that each year will be an own row. The dataframe will also be sorted by the country name and year.

Although it's sometimes good to have each metric in one dataframe, it is needed for this analysis to create a dataframe with multiple metrics, because we are looking for relationships between this data. Therefore, we have to merge the melted dataframe together into one. Since the size of the population will be considered in most parts of the analysis, it will be used as "starting dataframe" for the merging process

The dataframe was successfully melted and merged, so that we now have two metrics in one dataframe. While the population has 45705 rows, the life expectancy 'only' has 40437, which means, that there already is some missing data, which we have to handle later. Using the library "missingno" is a good and quick way to visualize missing data.

Looking at the info of the newly created dataframe, we can see that there is alot more missing data than before. Also the "year" column is a string, which should be transformed into an integer, because we will create bins based on the years later. Also the income_per_person column should be transformed into an integer. The sugar_cons is in gramm, so it can also be transformed into an integer.