

# Life Expectancy Data Exploration & Analysis

Roy Minato

<https://royminato.com/>

Life expectancy refers to the number of years a person or other biological organism is expected to live for. While this may seem like a straightforward definition, it is important to recognize that life expectancy for a particular year estimates the age a person born in that same year would expect to live for if the average age of death in their environment did not change over their lifetime.<sup>1</sup> So while many people might interpret life expectancy to be a predictor of the future, it more suited to be thought of as a reflection of current conditions in a person's environment. With this clarification in mind, what this paper aims to explore is how life expectancy varies across multiple geographic locations and what relationships exist between life expectancy and several factors of interest.

---

<sup>1</sup> Esteban Ortiz-Ospina, "'Life Expectancy' – What Does This Actually Mean?," Our World in Data, August 28, 2017, <https://ourworldindata.org/life-expectancy-how-is-it-calculated-and-how-should-it-be-interpreted>.

## **GitHub Project Repository:**

<https://github.com/RoyMinato/Life-Expectency-Study>

## **Data Sources & Dataset Creation:**

The dataset used throughout my exploration and analysis is a custom dataset I created using Python to web scrape multiple tables from several sources which I then joined into a single comprehensive table. The dataset and code written to create it is available as a Jupyter Notebook file in this projects GitHub repository. All visuals used in this report are also available in an R markdown file in the same repository.

The original data tables used include data from: the UN Development Program which includes life expectancy estimates for each country, a table of countries and their respective continents and regions as designated from the UN Statistics Division, infant and under-five mortality rates by country from the World Bank, and global peace index scores for each country from the Institute for Economics & Peace. Each source used is concerned with 2019 data for consistency.

## **Dataset Sample:**

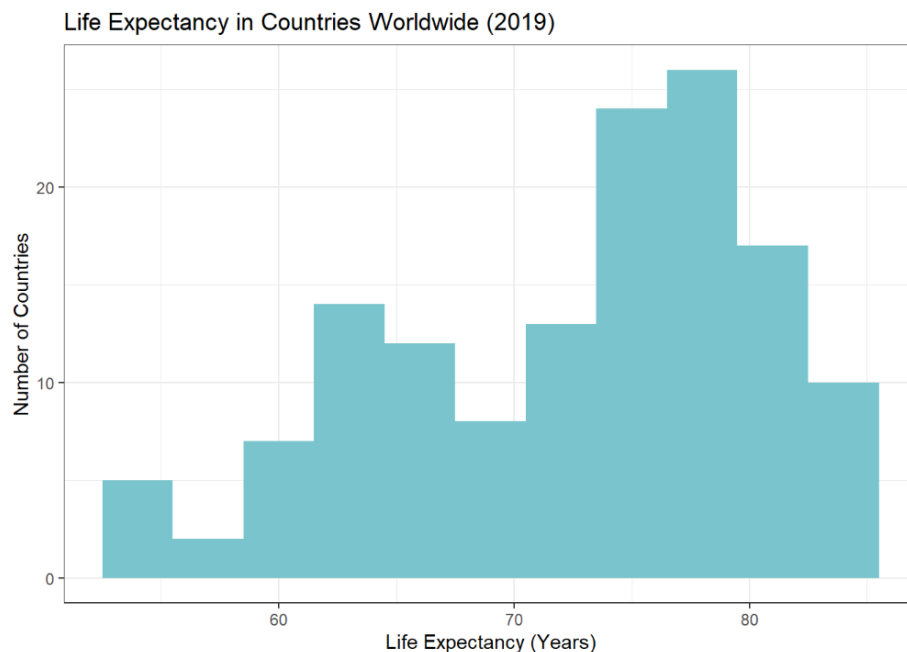
	Country	All	M	F	Gender life gap	Region	Continent	U5 Mortality Rate	GPI Score
0	Japan	84.6	81.5	87.7	6.2	Eastern Asia	Asia	2.5	1.369
1	Italy	84.0	81.9	86.1	4.2	Southern Europe	Europe	3.1	1.754
2	Switzerland	83.8	81.9	85.6	3.7	Western Europe	Europe	4.0	1.375
3	Singapore	83.6	81.5	85.7	4.2	South-eastern Asia	Asia	2.5	1.347
4	Spain	83.5	80.8	86.2	5.4	Southern Europe	Europe	3.1	1.699

In the sample provided above, the variable “All” refers to the life expectancy for the entire population of a country. “U5 Mortality Rate” is the infant and under-five years of age mortality rate per one-thousand births and will be used as a metric to convey general healthcare quality of a country. “GPI Score” stands for global peace index score and is meant to provide an indication of the level of peacefulness in a country. GPI scores range from a scale of one to five where a lower score means a higher level of peacefulness.

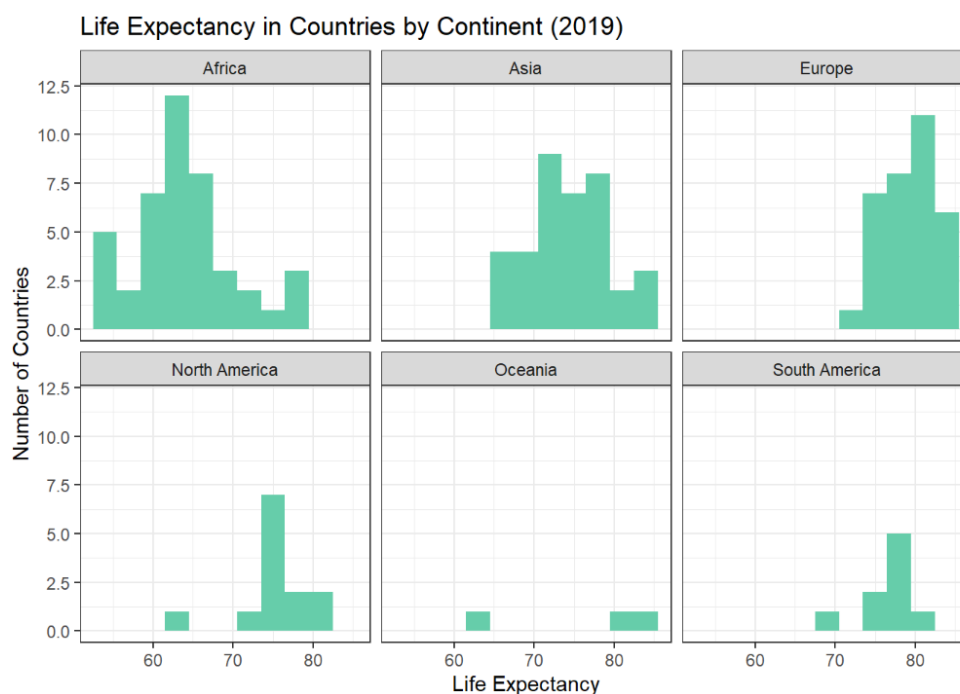
## **Data Exploration and Analysis:**

Before looking at how different geographic locations vary in life expectancy, a logical starting point would be to get an idea of how life expectancy is distributed

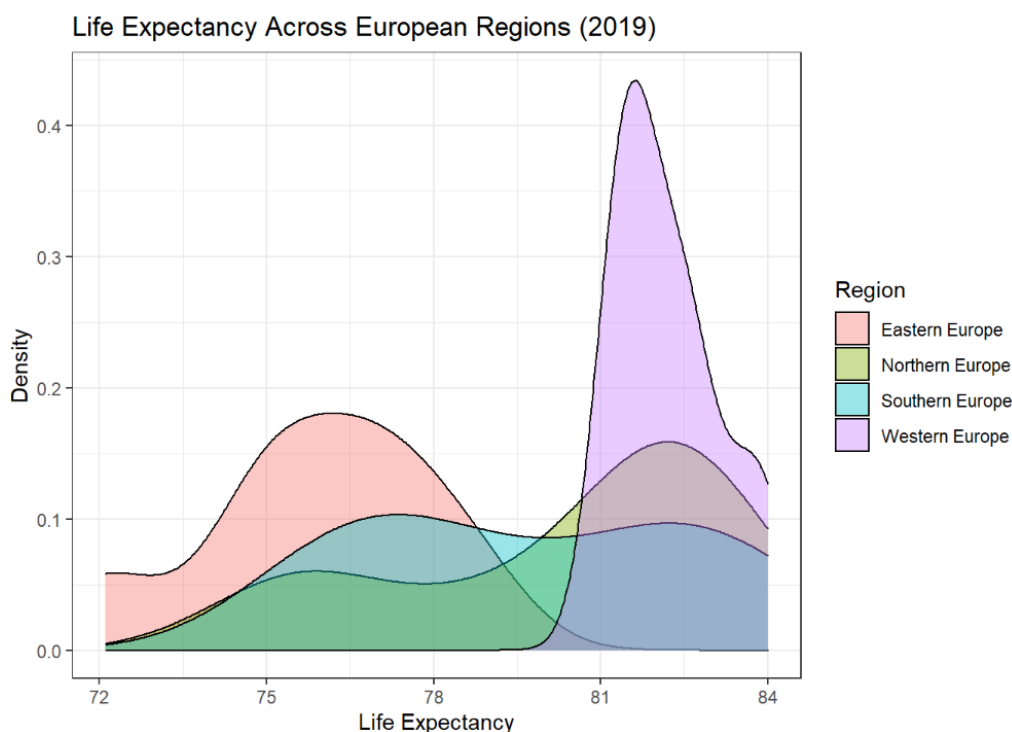
worldwide with no particular attention to region. This will also allow us to understand how findings in specific regions match up with worldwide trends.



This histogram shows that most countries tend to have a life expectancy of at least 70 years however, we can also see that a considerable number of countries have much lower life expectancies in the range of 60-65 years with a few countries being even lower. More information can be found when we shift our perspective to life expectancy based on continent.

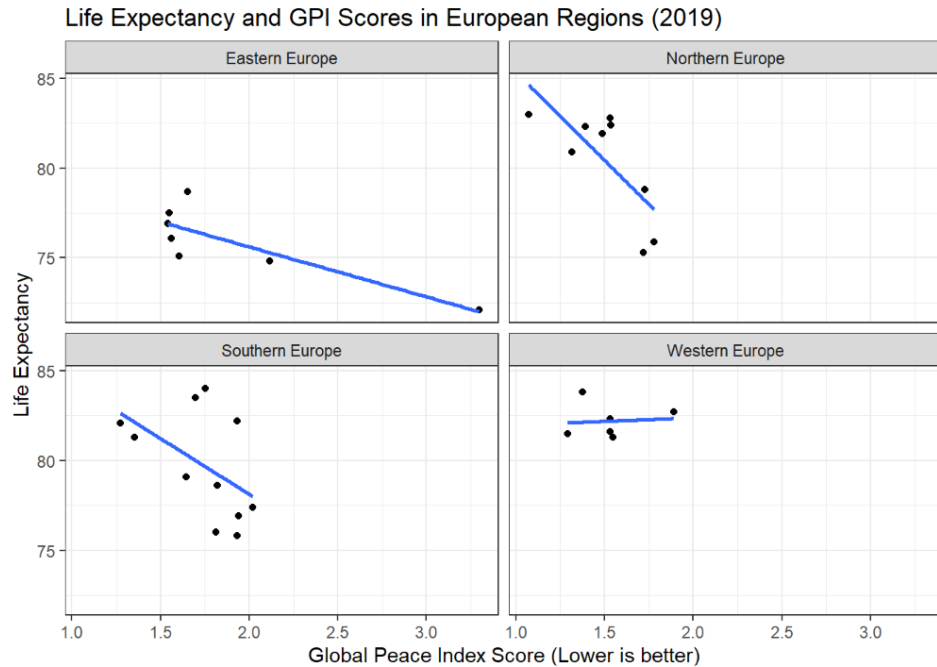


Already we can begin to see some very interesting trends and differences among geographical locations at the continent level. For example, we can see that life expectancy varies the most in Africa where it can range from 53 to 80 years but seems to have a median life expectancy in the low 60's. Conversely, Europe seems to have a much smaller range from 70 to 85 years with a median around the high 70's or low 80's. Just as the process of taking a more granular view from a worldwide perspective already presented valuable information, we can further focus our attention on intracontinental trends to better understand life expectancy in countries within the same continent and analyze why they might differ.



Focusing on the above density plot, we can see that although Europe has the highest median life expectancy among continents, the distribution of life expectancy across European regions is not evenly distributed. This lack of equality of life expectancy in Europe is most pronounced between the east and west. Another interesting observation is the very low variation of life expectancy in Western European countries compared to countries in the other three regions which all display considerably greater variation.

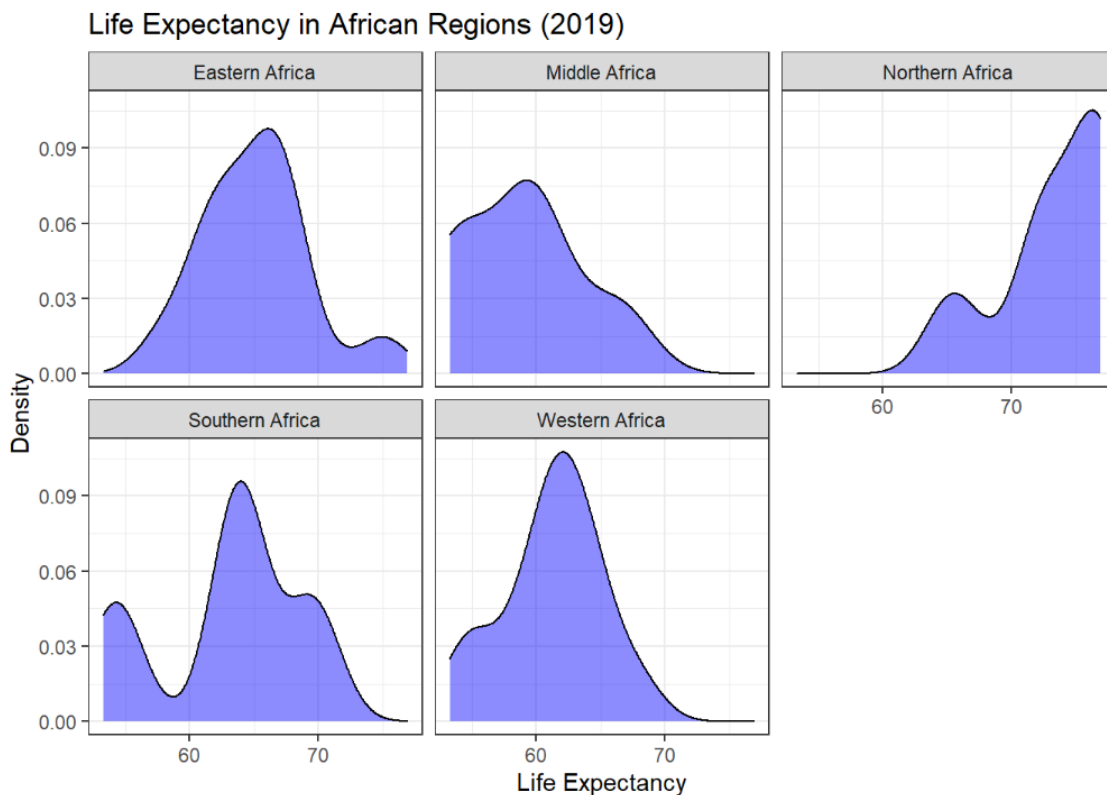
Continuing with an established idea of how life expectancy can differ between and within continents, we can further our understanding of life expectancy in different countries by analyzing how different national factors can affect the life expectancy in a country. These factors will relate to but not necessarily fully represent a nation's general quality of healthcare and its state of peacefulness.



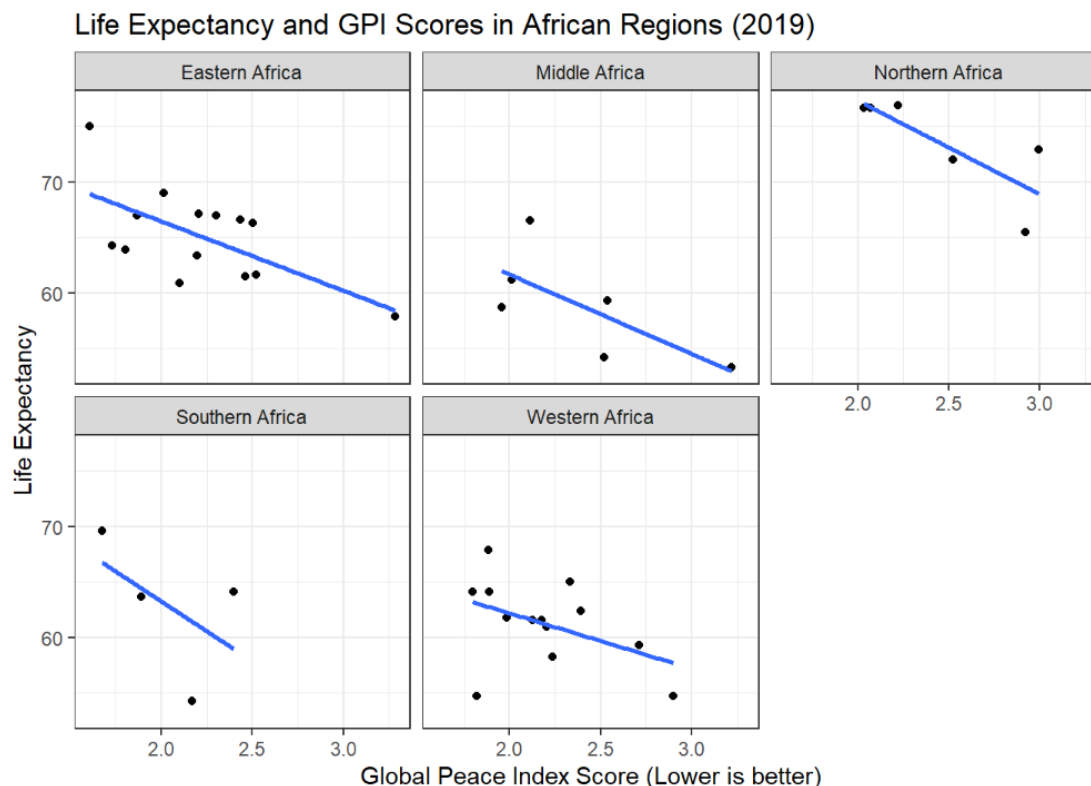
As GPI scores are meant to convey the state of peacefulness in a country, it is not surprising that we can see a clear trend that as GPI scores increase (meaning there is less peace in a country), life expectancy tends to decrease. What is interesting to note though is that GPI scores alone are not necessarily predictive of life expectancy. This is clearly shown in the visual for Southern Europe where there are multiple instances of countries sharing very similar GPI scores but having a great discrepancy in life expectancy. Thus, while we cannot say GPI scores can predict the life expectancy of a country, we can say that there is a general inverse relationship between GPI and life expectancy.



From the above boxplots, we can see that infant mortality rates vary more in eastern and southern Europe than they do in northern and western Europe. What is interesting to note is that while we previously saw Western European countries having the highest life expectancy values in the continent, it also has a higher median mortality rate than southern and northern Europe. Of course, it is important to remember that the general healthcare quality of a nation is much more complicated and not fully represented by a singular metric like infant mortality rate. To gain more information, it would be interesting to investigate another continent, particularly Africa due to having the highest variability in life expectancy.



When looking at individual African regions, what we can see is a clear distinction between Northern Africa and other regions. Most countries in Northern Africa seem to have relatively higher life expectancies than other regions and most of them are concentrated above 70 years while other regions have a very small minority of their countries in the 70 plus range. Countries in other regions seem to have a median life expectancy somewhere in the early to middle 60's and so there is a decent discrepancy between Northern Africa and the other African regions. Looking into GPI factors again with respect to African regions might tell us more about what is going on in the countries that make up those regions as it did with the European regions.



What we can see again in the African regions is the same trend we saw in the European regions when we tried to understand the relationship between life expectancy and the level of peacefulness in a country. Regardless of region or continent, we can make the claim that as the GPI score of a country increases, we can expect to see a decline in life expectancy in that country. Of course, as previously mentioned, we cannot claim that given the GPI score of a country we can reliably predict the life expectancy in that country. This is because there are an extremely high number of factors that go into producing the life expectancy of a country and as we see in this visual, countries with similar GPI scores can have wide discrepancies in life expectancy. Nonetheless, the inverse relationship between life expectancy and GPI score is still present.

### **Conclusion:**

While life expectancy is a very complicated metric in terms of the factors which influence it, this data exploration and analysis was able to understand differences in life expectancy between and within continents. While healthcare quality was something that would be interesting to study with respect to life expectancy, it is difficult to quantify healthcare quality which was what the study attempted to do with infant mortality rates which is only one of a plethora of healthcare factors. GPI scores were more insightful as we could see a clear inverse relationship between GPI scores and life expectancy in a

country which was shown in multiple continents and regions. What is important to recognize though is that countries in the same region with similar GPI scores could have vastly different life expectancies. Again, this is a product of life expectancy being influenced by so many different factors which would require a much larger and more in-depth study to better understand. Some other factors which could provide valuable insight about why life expectancy can differ amongst countries could do with economic related metrics.