Determining Factors of Life Expectancy Across Continents and Countries.

Ali Afkhami (30271805)

Evan Losier (30022571)

Luisa Alejandra Sierra Guerra (30261956)

Ruby Nouri Kermani (30261323)

University Of Calgary

Faculty Of Science

Master In Data Science And Analytics

Course: DATA 604

Winter 2025

Introduction

Life expectancy is one of the key indicators of well-being and development of a population. In this project we aim to study and investigate the effect of various broad factors such as **Economy (Ali), Food/Agriculture (Evan), Air Pollutants (Ruby), and Violence (Alejandra)** on life expectancy across different countries and continents. By doing this, we will learn which factors are most correlated and influential in the calculation of life expectancy. The subject is interesting to us because life expectancy can vary a lot around the world and finding out why people in some areas live longer than people in other areas could help us all live longer and healthier lives.

Datasets

Our primary dataset for life expectancy was found on Our World in Data and includes information such as country, year, and life expectancy at birth. The dataset has 4 columns and 21565 rows, and is stored in one csv file. The main field we were interested in was life expectancy at birth, but we used country and year to match the life expectancy data with other datasets. We explored this dataset and other relevant datasets using SQL databases and queries. If any of the individual datasets were missing values for countries or years, we cleaned them by either deleting the null rows or filling the missing values with appropriate alternatives if removing nulls would have been harmful to the analysis. If more datasets were beneficial for the analysis, we added them during later stages of the project.

The datasets used to investigate the economy from Our World In Data were used to explore the effect of economic factor on the life expectancy of the population. The target values were GDP per capita. The datasets used were:

- Economic Growth (4 columns, 7064 rows) - Includes three columns of the countries (as well as their aggregation as continents), the years and GDP per capita.

The datasets used to investigate food and agriculture were all found on Our World in Data and were used to explore the overall topic of life expectancy by comparing how a population's food supply, crop yield, and undernourishment were related to overall life expectancy. We suspected that all the datasets related to food and agriculture were correlated with life expectancy, where countries with high crop yield & food supply, and low malnourishment generally had higher life expectancies. The datasets used were:

- Hunger and Undernourishment (4 columns, 4455 rows) - Includes columns for country, year, and prevalence of undernourishment ("Share of individuals that have a daily food intake that is insufficient to provide the amount of dietary energy required to maintain a normal, active, and healthy life.").

- Crop Yields (3 columns, 14577 rows across multiple datasets) - Includes datasets for different crops (currently corn, wheat, oats, and rice), which have columns of country, year, and crop yield (in units of tonnes per hectare).
- Food Supply (3 columns, 12750 rows) - Includes columns for country, year, and food supply (in units of kcal per capita per day).

The dataset used to investigate air pollutants explored the effect of different types of emissions on life expectancy. We suspected that all different types of pollutants had a negative impact on life expectancy and explored whether any one was more impactful than the others. The dataset used was:

- Emissions of Air Pollutants (11 columns, 63883 rows) - Includes columns for country, year, and 9 columns for different pollutants (ammonia, black carbon, carbon monoxide, methane, nitrogen oxides, nitrous oxide, non-methane volatile organic compounds, organic carbon, and sulfur dioxide, all in units of tonnes).

The datasets used to investigate **violence** explored the effect of different factors that contribute to overall peace and violence of countries and how they might affect total life expectancy. The datasets used are:

- Civilian and combatant deaths in armed conflicts based on where they occurred (6 columns, 7176 rows) - This dataset includes the country, country code, year (ranging from 1989 to 2023), deaths of civilians in ongoing conflicts within a country (conflict type: all), deaths of unknown type in ongoing conflicts within a country (conflict type: all), and deaths of combatants in ongoing conflicts within a country (conflict type: all).
- GPI - The Global Peace Index (GPI) dataset provides GPI information by country from 2008 to 2024. Published by the Institute for Economics and Peace (IEP), this dataset includes three columns: country, year, and the GPI score, with a total of 2,760 rows.
- The Safety and Security Score dataset contains Safety and Security Score information by country from 2008 to 2024, also published by the Institute for Economics and Peace (IEP). It includes three columns: country, year, and the safety and security score, with a total of 2,760 rows.
- The Violent Crime Score dataset contains data on violent crime scores by country from 2008 to 2024, published by the Institute for Economics and Peace (IEP). The dataset includes three columns: country, year, and the violent crime score, with a total of 2,760 rows.
- The Homicide Score dataset includes homicide score information by country from 2008 to 2024, published by the Institute for Economics and Peace (IEP). This dataset has three columns: country, year, and the homicide score, with a total of 2,760 rows.

Guiding Questions

1. Among the various factors such as the Economy, Violence, Air Pollutants, and Agriculture, which have the strongest impact on the life expectancy of the population.
2. How do the correlations vary based on the continents and certain specific countries (Canada, Ireland, Israel and Mauritius)?
3. How correlated are these factors in the world with life expectancy and which one is the strongest?

General Methodology

1. **Data Preparation & Cleaning:** The dataset was preprocessed using Pandas and SQL to ensure consistency and reliability. Column names were standardized, missing values were appropriately handled through imputation or removal, and datasets were merged to create a comprehensive data table. The cleaned data was then uploaded into a SQL database to facilitate efficient querying and analysis.
2. **Feature Engineering:** Key variables were generated to enhance the analytical depth of the study. This included the classification of countries by continent and the creation of additional variables such as total deaths due to conflict and a composite measure for economic and social factors affecting life expectancy. Unique identifiers, such as CYID, were also added to facilitate table joins and relational data analysis.
3. **Exploratory Data Analysis (EDA):** An initial assessment of the dataset was conducted to identify patterns, trends, and potential correlations. Specifically, correlation analyses were performed to examine the relationships between GDP per capita variation, violence-related variables, and life expectancy. The influence of food supply, crop yields, and air pollutants on life expectancy were also explored to provide a multidimensional perspective.
4. **Regional and National-Level Examination:** The study further examined the impact of economic, conflict-related, agricultural, and air pollution related variables on life expectancy at both regional and national levels. This involved analyzing how different continents and specific countries exhibited variations in life expectancy based on economic conditions, levels of violence, agricultural productivity, and presence of air pollutants.

Results

Life Expectancy and The Economy

Table 1, illustrates the top two (table on the left) and bottom two (table on the right) countries in each continent.

Table 1, top two and bottom two countries in each continent based on their GDP in international dollars

Country	GDP	Year	Continent	Rank_Desc
Seychelles	29468.951	2023	Africa	1
Mauritius	26590.424	2023	Africa	2
Singapore	127543.555	2023	Asia	1
Qatar	116159.140	2023	Asia	2
Luxembourg	132846.580	2023	Europe	1
Ireland	114922.390	2023	Europe	2
Bermuda	98041.690	2023	North America	1
Cayman Islands	78109.350	2023	North America	2
Australia	59552.880	2023	Oceania	1
New Zealand	48826.574	2023	Oceania	2
Guyana	49315.160	2023	South America	1
Uruguay	31019.310	2023	South America	2

Country	GDP	Year	Continent	Rank_Asc
Burundi	828.8611	2023	Africa	1
Central African Republic	1135.1362	2023	Africa	2
Afghanistan	1992.4244	2023	Asia	1
East Timor	4328.5080	2023	Asia	2
Kosovo	13642.6770	2023	Europe	1
Moldova	15855.2730	2023	Europe	2
Haiti	2956.4622	2023	North America	1
Honduras	6468.3020	2023	North America	2
Solomon Islands	2534.0276	2023	Oceania	1
Vanuatu	3118.3960	2023	Oceania	2
Bolivia	9843.9700	2023	South America	1
Ecuador	14472.3160	2023	South America	2

These two tables display the top two (left table) and bottom two (right table) countries in each continent ranked by GDP for the year 2023. The ranking is performed separately within each continent,which allows us to directly compare economic disparities.

- The left table lists the highest-GDP countries per continent, with Luxembourg (Europe), Singapore (Asia), and Australia (Oceania) among the wealthiest.

- The right table presents the lowest-GDP countries per continent, including Burundi (Africa), Afghanistan (Asia), and Solomon Islands (Oceania) with significantly lower GDP values.

This comparison highlights economic inequality across and within continents, emphasizing the substantial gap between the wealthiest and poorest nations.

Table 2, ranks the continents based on their overall GDP.

Table 2, continents ranked based on their overall GDP

Continent	Europe	Asia	North America	Africa	South America	Oceania
GDP per Capita (International Dollars)	2,011,913	1,427,721	876,603	350,019	249,050	186,104

According to table 2, the European population has the highest overall GDP per capita with 2,011,913 international dollars; afterwards, Asia holds the second rank with 1,427,721 international dollars, followed by North America, Africa, South America. According to this table, the population of the countries in Oceania has the lowest GDP per capita with 186,104 international dollars per capita.

According to table 1, the top two countries with highest life expectancy are Monaco and San Marino and the two countries with lowest life expectancies are Nigeria and Chad. However, since Monaco is not listed in the GDP dataset, Hong Kong is used instead which ranked 3rd in highest life expectancy and is present in the GDP dataset.

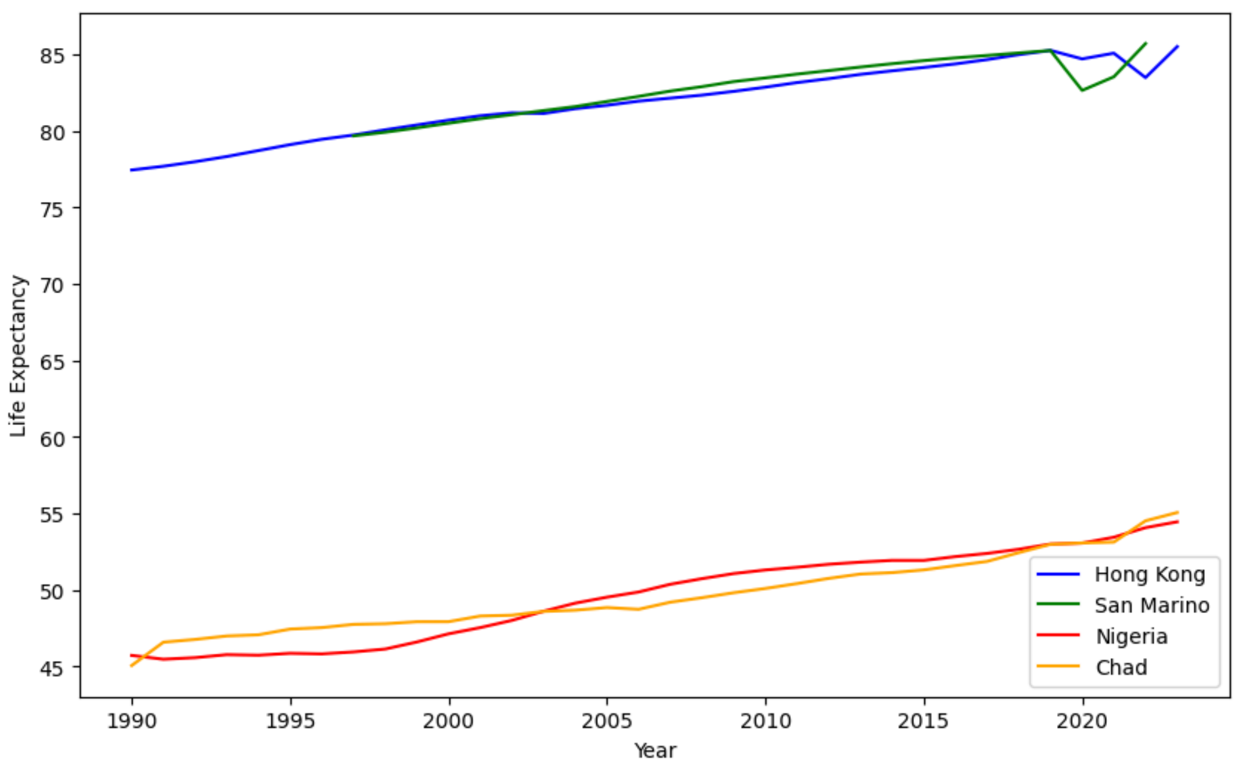


Figure 1, life expectancy of 4 countries over the years

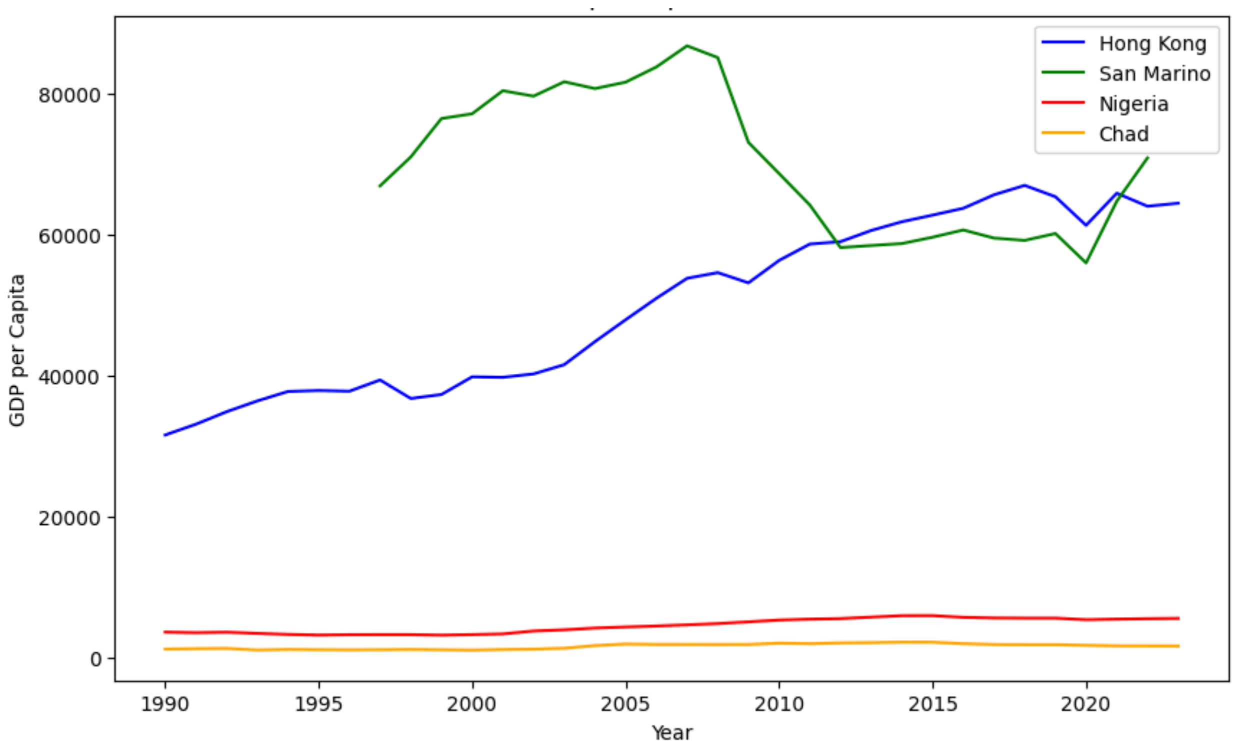


Figure 2, GDP per capita of 4 countries over years

According to figures 1 and 2, due to the fluctuations over the years (specially for San Marino), it is difficult to interpret any accurate result for San Marino, however, it can be seen that for the poor countries of Nigeria and Chad, the increase in life expectancy over the years is independent from their GDP since their GDP has not been growing as much as the life expectancy of the population.

This is a good example of the fact that life expectancy is affected by various factors other than GDP per capita.

In the next step, the correlation between the GDP and life expectancy over the years will be investigated. Table 3 lists the correlation between the life expectancy and GDP per capita of the four aforementioned countries.

Table 3, correlation between the life expectancy and GDP per capita of Hong Kong, San Morino, Nigeria, and Chad

Country	Hong Kong	San Morino	Nigeria	Chad
Correlation	0.963527	-0.599111	0.947889	0.645820

As can be seen from the correlation table above for Hong Kong and Nigeria, there is a strong correlation between the life expectancy of the population and GDP per capita. However, for

Chad the correlation is not strong, meaning that there might be more important factors affecting life expectancy than the GDP per capita. In the case of the San Marino there is a negative correlation, meaning that data are not consistent and we will have to look into more factors to see the overall effect on life expectancy.

Figure 3, illustrates the correlation between GDP and Life Expectancy over the Years.

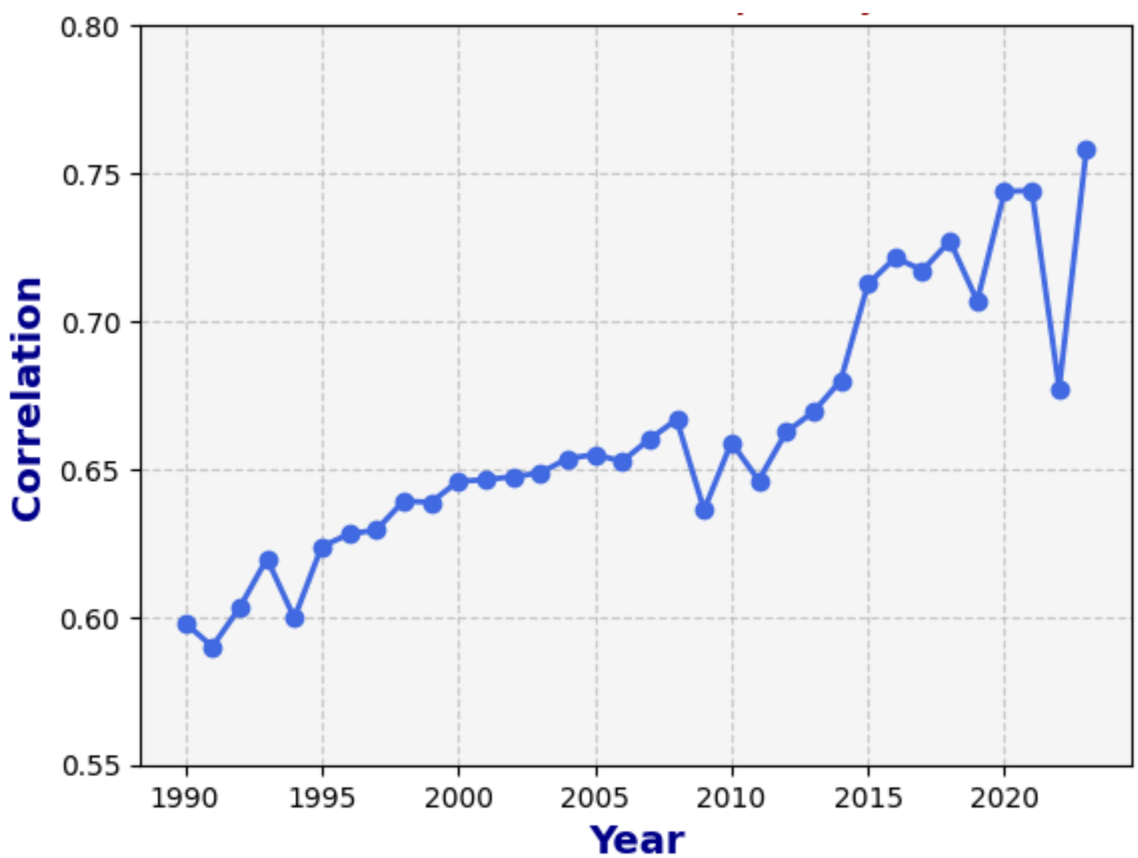


Figure 3, Correlation between GDP and Life Expectancy over the Years

This is an interesting result, showing the correlation between the life expectancy of a country's population and GDP per capita.

Here are some of the key observations:

1. Increasing Trend: The correlation has generally increased from the early 1990s to 2023, suggesting a stronger relationship between GDP per capita and life expectancy over time.

2. Fluctuations: Some dips in correlation are visible around certain years, indicating periods where the relationship between GDP and life expectancy weakened temporarily (such as Covid).

3. Higher Correlation in Recent Years: From around 2010 onward, the correlation rises more sharply, implying that in more recent years GDP per capita has played a more consistent role in influencing life expectancy.

Interpretation:

1. Economic Development & Health Improvements: As economies grow, better healthcare, nutrition, and living conditions contribute to increased life expectancy.

2. Variability in Earlier Years: The lower correlation in the 1990s and early 2000s may reflect variations in how different countries allocated economic growth toward healthcare and social welfare.

3. Potential External Factors: Certain fluctuations may be due to global economic crises, pandemics, or geopolitical events affecting health outcomes differently across regions.

Life Expectancy and Violence

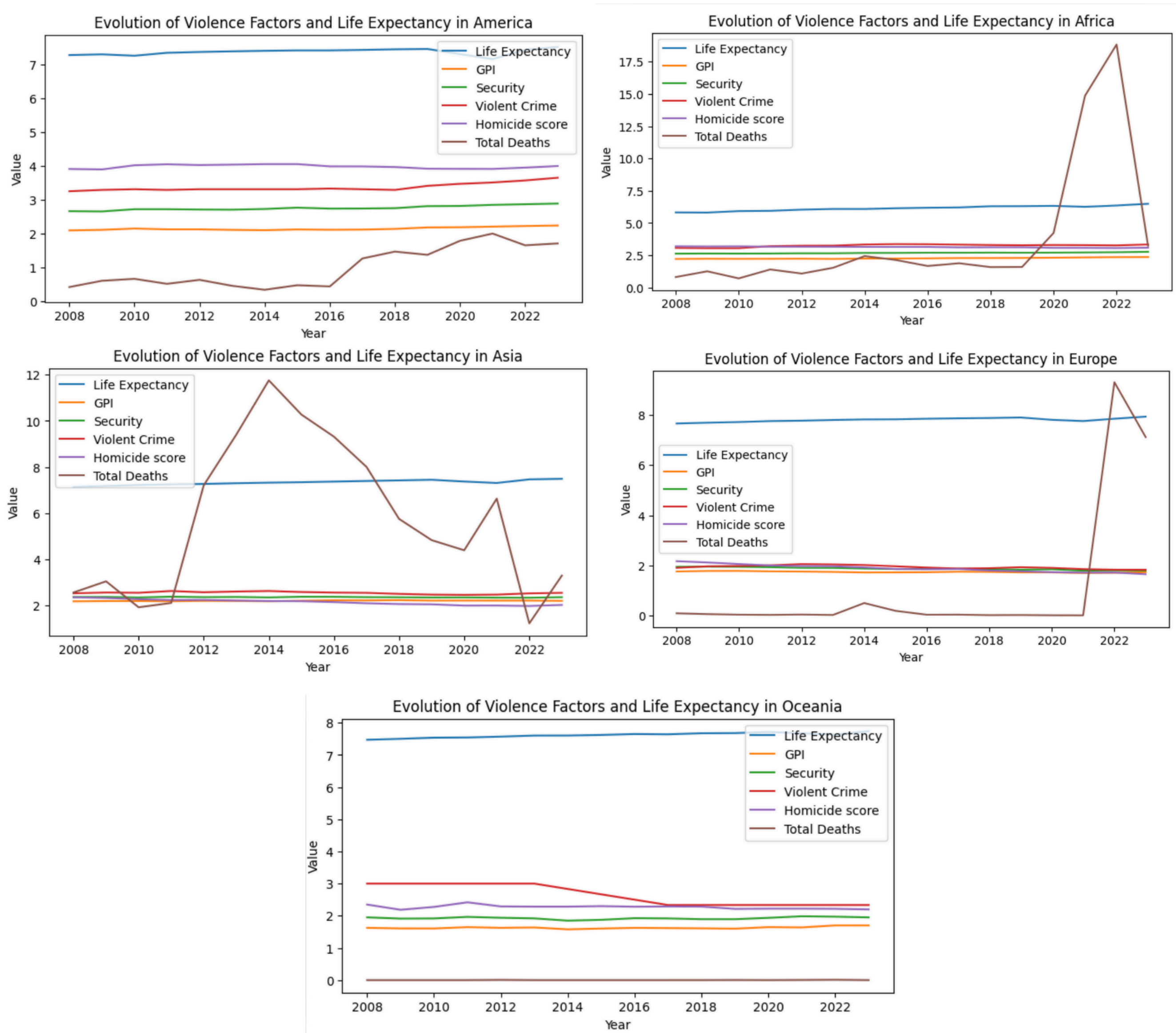
Figure 4. Correlation Between Violence and Life Expectancy Worldwide

For this correlation analysis, all datasets related to the Life Expectancy variable were used. The variables Conflicts, GPI, Homicide Score, Violent Crime Score, and Safety and Security Score represent country-level ratings from 2008 to 2023, where 1 indicates a better condition and 5 a worse condition.

The correlation matrix reveals that Safety and Security Score has the strongest correlation with life expectancy, even surpassing Deaths due to Conflict and Homicide Score, indicating a significant impact on life expectancy. With a negative correlation of -0.6005, the data suggests that as the Safety and Security Score increases (worsens), life expectancy decreases.

The second most correlated factor was Violent Crime Score, with a correlation of -0.5731, exceeding even the Homicide Score. Notably, the Deaths due to Conflict variable showed the weakest correlation with life expectancy, at only -0.0715. This result was surprising, as one would expect that mortality directly affects life expectancy. However, deaths from conflicts over the past decade do not appear to have had a significant impact on overall life expectancy.

Figure 5. Evolution of Violence Factors and life Expectancy in 5 continents



In the graph for the American continent (North America, Central America, and South America), we can observe that it has the highest homicide score among all continents, with an average of 4, compared to the global average of 2, which is nearly double the homicide score of other regions. Regarding total deaths per conflict, there has been a continuous increase since 2016. Meanwhile, the Global Peace Index (GPI) has remained relatively stable since 2008, showing no significant changes. However, one notable trend is that the Violent Crime Score has been rising since 2018. Additionally, there was a decrease in life expectancy between 2020 and 2021, which can be attributed to the impact of COVID-19.

In Africa, total deaths due to conflict quadrupled since 201, yet life expectancy only declined for one year, specifically in 2022, which marked the peak of conflict-related deaths across all continents, with nearly 188,361 recorded deaths. That year, 15 armed conflicts were identified, with major crises occurring in the Western Sahel region, Lake Chad, the Horn of Africa, Central Africa, and northern Mozambique, alongside the COVID-19 pandemic, which likely contributed to the decline in life expectancy. However, since then, life expectancy has followed an upward trend. Meanwhile, other key indicators have remained stable from 2008 to 2023, with Violent Crime Score and Homicide Score consistently showing the highest values.

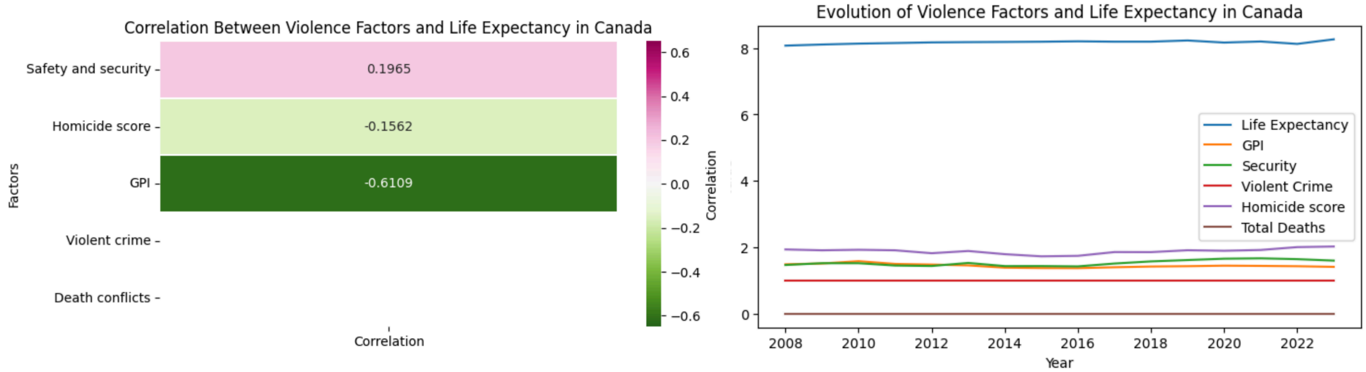
Asia is the continent with the highest number of conflict-related deaths in the last 20 years, averaging 57,325 deaths per year and peaking at over 117,000 deaths in 2014 due to conflicts in Middle Eastern countries. Despite this, life expectancy has not been significantly affected, which seems contradictory, as mortality should directly impact it. A possible explanation is that Asia is home to two of the world's most populous countries—India, with nearly 1.46 billion people, and China, with 1.4 billion inhabitants—which likely exert a strong influence on overall life expectancy, counteracting the effect of conflict-related deaths. Other variables have generally shown a downward trend, except for the Global Peace Index (GPI), which has worsened in recent years compared to 2008. Life expectancy has consistently increased, except for a decline in 2020 and 2021, likely due to the COVID-19 pandemic.

Europe is the only continent where GPI, security score, violent crime score, and homicide score have consistently improved, showing a clear downward trend from 2000 to 2023. This stability makes it unique compared to other regions. However, conflict-related deaths emerged in 2021, whereas before, the number had remained at zero. This increase is directly linked to the ongoing conflict between Russia and Ukraine. Additionally, life expectancy declined in 2020 and 2021, primarily due to the COVID-19 pandemic.

Oceania is the only continent with zero conflict-related deaths in the last 20 years. Additionally, violent crime rates have significantly decreased since 2013, while other security-related scores have remained stable throughout this period. Life expectancy experienced a slight decline during the COVID-19 pandemic—the smallest decrease among all continents. This minimal impact could be attributed to well-implemented public health and policy measures, which helped prevent excessive mortality and, consequently, mitigated the effects on life expectancy.

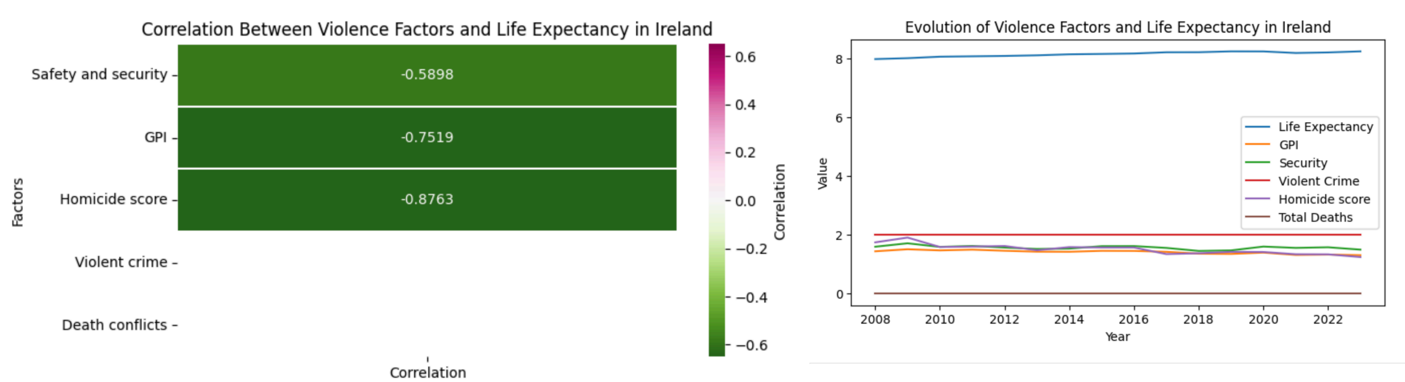
This analysis confirms that life expectancy was more significantly impacted by COVID-19 than by other violence-related variables. Our findings suggest that this is due to the short time frame (1 to 2 years) and the high volume of deaths, which can lower life expectancy even in highly populated regions. In contrast, conflict-related deaths did not have as strong an impact on life expectancy, likely because they are spread out over time rather than concentrated in a short period. Only when a large number of deaths occur in a short timeframe—as seen in Africa in 2022—does life expectancy show a noticeable decline.

Figure 6. Violence Factors and life Expectancy in Canada



In Canada, violence-related variables such as violent crime and deaths from conflict show no correlation with life expectancy, as their levels have been nearly nonexistent or too low to have a significant impact over the past 20 years. However, the most relevant factor is the Global Peace Index (GPI), which has a negative correlation of -0.61—meaning that as the GPI score decreases (indicating improved peace), life expectancy increases. When analyzing the evolution of these factors, we observe that the homicide score is higher than all other variables, suggesting that this issue should be a priority for the government. Overall, life expectancy in Canada has shown an upward trend, with a slight decline during the pandemic.

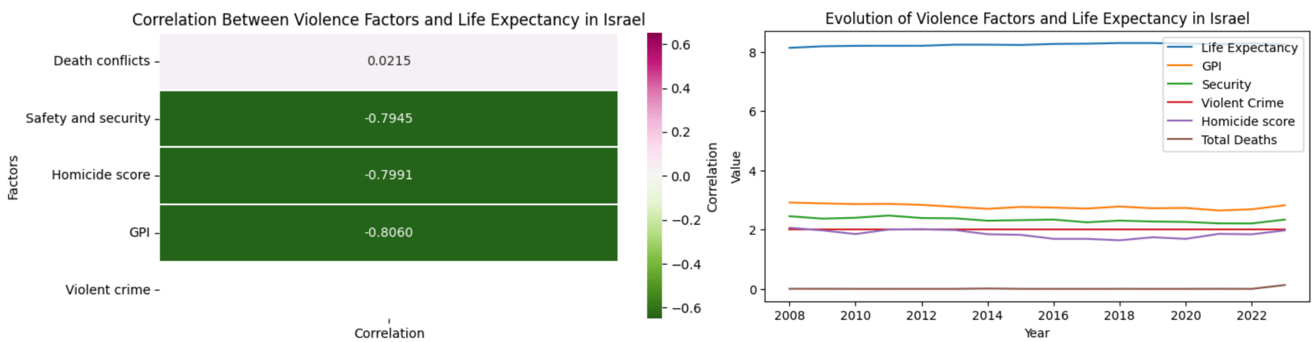
Figure 7. Violence Factors and life Expectancy in Ireland



In Ireland, we observe that violent crime and deaths from conflict do not show a correlation with life expectancy (likely for the same reasons as in Canada). However, the violence-related variable with the strongest correlation is the homicide score, with a negative correlation of

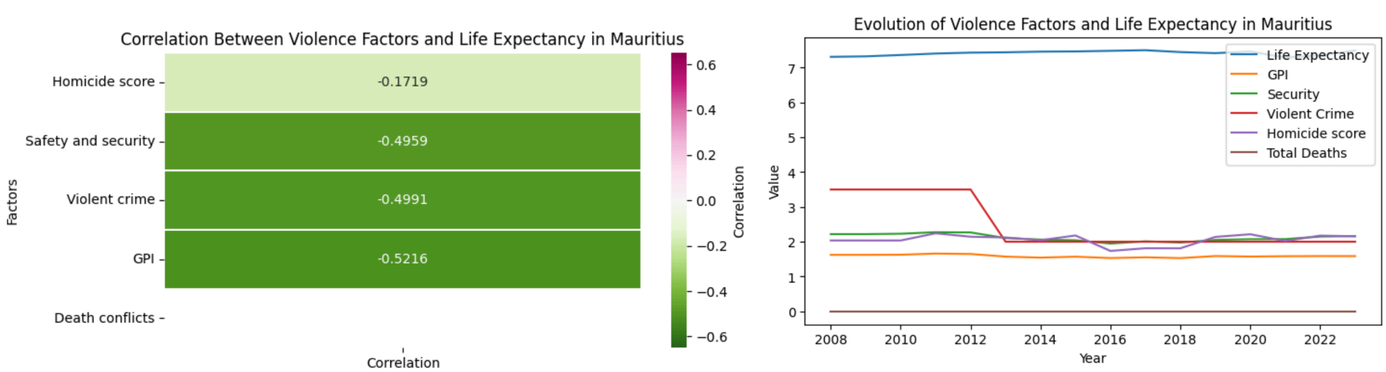
-0.8763, meaning that as the homicide score increases, life expectancy decreases. This highlights the importance of monitoring this rate so the government can take appropriate measures. Life expectancy has shown a slight upward trend, but it briefly decreased during 2020-2022 due to the pandemic. It is important to note that the homicide score has remained consistent over the past 20 years, indicating that the government must pay closer attention to this issue to understand and prevent further occurrences.

Figure 8. Violence Factors and life Expectancy in Israel



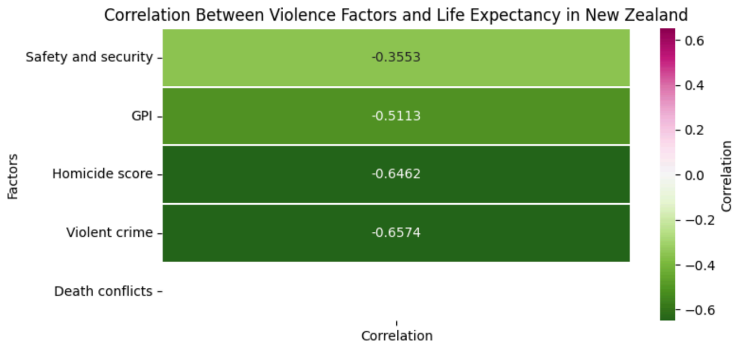
In Israel, the violence-related variable with the highest correlation is the Global Peace Index (GPI), with a negative correlation of -0.81. This means that as the GPI improves, life expectancy increases. However, there is a peculiar phenomenon regarding deaths by conflict, which shows a positive correlation with life expectancy. This means that as the number of deaths increases, life expectancy also rises. This contradictory phenomenon could be explained by the fact that, despite the loss of life due to conflicts, other variables in the country may improve, such as the economy, which could benefit from external investments driven by war-related support, leading to improvements in life expectancy. It is important to note that Israel has the highest GPI score among the countries studied, meaning it is the country with the lowest level of global peace.

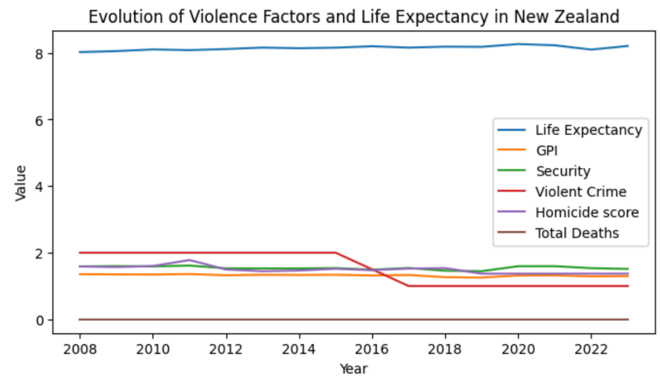
Figure 9. Violence Factors and life Expectancy in Mauritius



In Mauritius, the Global Peace Index (GPI) shows the highest correlation with life expectancy, with a negative correlation of -0.52. From 2008 to 2012, violent crime had the highest score of all the countries evaluated; however, since then, it has shown a declining trend. Life expectancy has generally followed a slight upward trend, though it was also slightly affected by the pandemic.

Figure 10. Violence Factors and life Expectancy in New Zealand

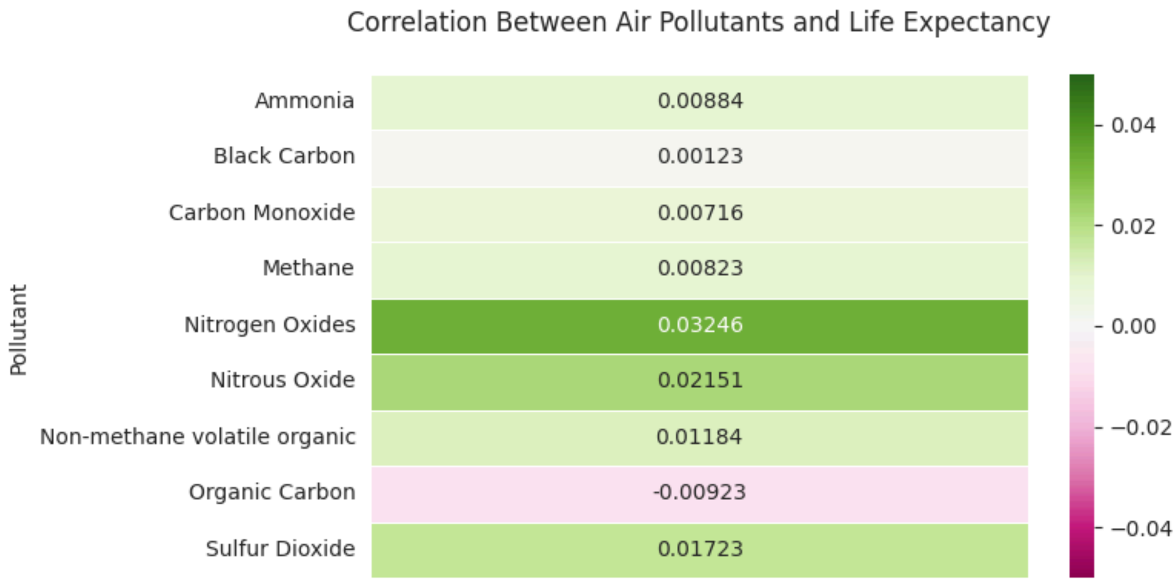




In New Zealand, we observe that violent crime has the greatest impact on life expectancy, with a negative correlation of -0.66, meaning that as the violent crime score increases, life expectancy decreases. When analyzing the evolution of these factors, we can see that while violent crime had the worst score, it is now the least significant factor. Therefore, for policy-making, this may not be a key variable, whereas factors with worse scores, such as security, which is not the best, could be more important. Life expectancy has followed a similar trend to that of other countries, with a noticeable impact during the COVID-19 period.

Life Expectancy and Air Pollutants

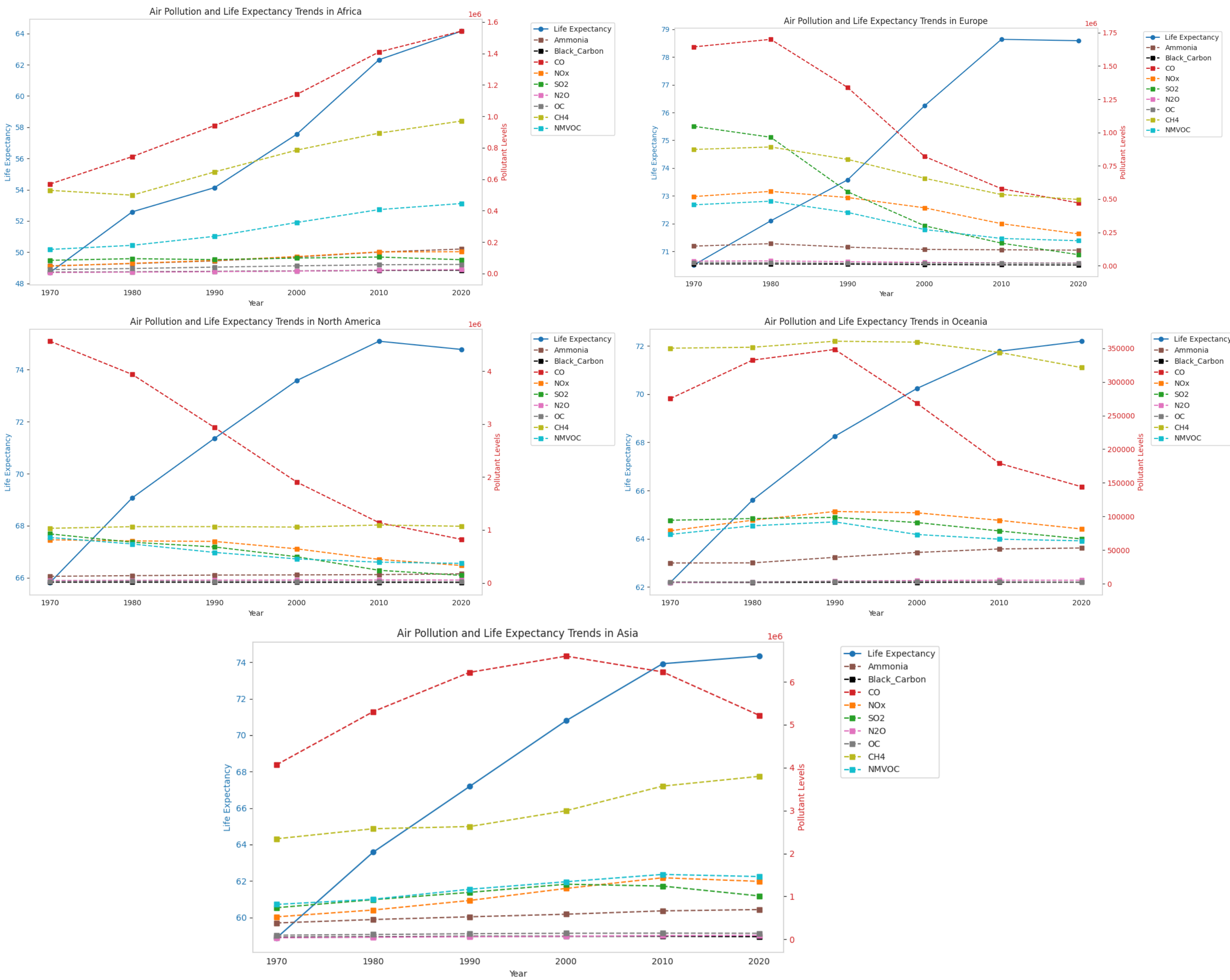
Figure 11. Correlation between Air Pollutants and Life Expectancy Worldwide



The correlation coefficients between air pollutants and life expectancy reveal several important patterns. Nitrogen Oxides show the strongest positive correlation (0.032), suggesting they may have the most noticeable association with reduced life expectancy among the pollutants analyzed. Nitrous Oxide (0.022) and Sulfur Dioxide (0.017) also demonstrate relatively stronger positive correlations. Organic Carbon is the only pollutant with a negative correlation (-0.009), though this weak relationship may not be statistically significant. The weakest correlations appear with Black Carbon (0.001) and Ammonia (0.008), indicating minimal apparent association with life expectancy in this analysis.

While these correlations are generally weak, Nitrogen Oxides emerge as potentially the most dangerous pollutant in this context, showing the strongest association with reduced life expectancy. However, the modest strength of these correlations suggests that other factors beyond these specific pollutants may play more substantial roles in determining life expectancy. The weak relationships could also reflect complex interactions between pollutants, delayed health effects, or the influence of confounding variables not accounted for in this analysis.

Figure 12. Evolution of Air Pollutants Factors and life Expectancy in 5 continents



In **Africa**, the life expectancy trend is from 1970 to 2020, contrasting with the relatively stable trends of various air pollutants. While pollutants like Ammonia, Black Carbon, NO_x, SO_2, N_2O, and OC show minimal fluctuations, CO, CH_4, and NMVOC are notably higher in magnitude, indicating a larger overall contribution to air pollution. Despite this, life expectancy has steadily increased, suggesting that factors beyond these specific pollutants, such as improved healthcare, may be more influential in driving this positive change. It's crucial to acknowledge the higher presence of CO, CH_4, and NMVOC and consider their potential health impacts, even as life expectancy improves, as this graph presents a simplified view of a complex relationship.

The plot for **Asia**, reveals a notable upward trend in life expectancy from 1970 to 2020, with a particularly steep increase after 1990. Conversely, the levels of most air pollutants, including Ammonia, Black Carbon, NO_x, SO_2, N_2O, and OC, show relatively stable and low trends over the same period. However, CO (Carbon Monoxide) exhibits a distinct pattern: it starts at a high level, increases further until around 2010, and then shows a decline, suggesting potential shifts in industrial practices or emissions controls. CH_4 (Methane) is also noteworthy, demonstrating a consistent upward trend, indicating a potential growing source of this pollutant. Despite these fluctuations in CO and the increase in CH_4, life expectancy continues to rise, suggesting that other factors, such as advancements in healthcare and living standards, may be playing a more dominant role in influencing longevity during this period. It's crucial to acknowledge the distinct trends of CO and CH_4 and their potential implications for public health, even as life expectancy improves.

The graph shows a steady increase in life expectancy in **Europe** from 1970 to 2020, with a notable acceleration after 1990. Concurrently, the levels of most air pollutants, including Ammonia, Black Carbon, NO_x, N_2O, and OC, remain relatively stable and low throughout the period. However, CO (Carbon Monoxide) exhibits a significant decline, particularly after 1980, suggesting successful efforts in reducing its emissions. Similarly, CH_4 (Methane) and SO_2 (Sulfur Dioxide) show a noticeable downward trend, indicating improvements in air quality related to these specific pollutants. Despite these reductions, life expectancy continues to rise, suggesting that factors beyond these specific pollutants, such as advancements in healthcare and living standards, may be playing a more dominant role in influencing longevity during this period. The marked decrease in CO, along with the decline in CH_4 and SO_2, points to potential positive impacts of environmental regulations and technological advancements in Europe, even as life expectancy improves.

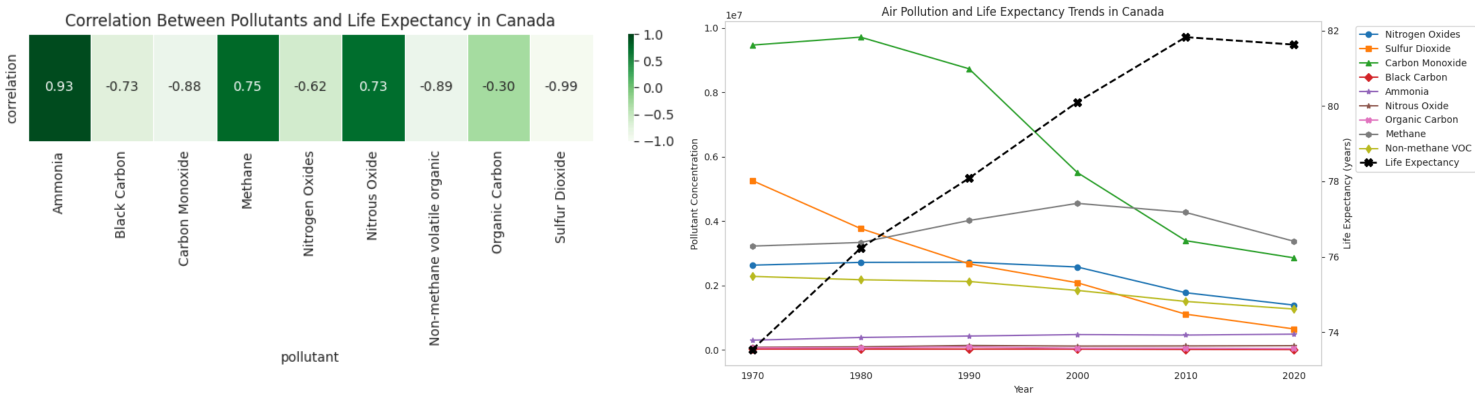
The graph shows a consistent upward trend in life expectancy in **North America** from 1970 to 2020. In contrast, the levels of most air pollutants remain relatively stable and low over the same period. However, CO (Carbon Monoxide) exhibits a notable decrease, particularly after 1980, indicating a significant reduction in its levels. Despite this decrease in CO, life expectancy continues its upward trajectory, suggesting that factors beyond these specific pollutants, such as advancements in healthcare and living standards, may be playing a more dominant role in influencing longevity during this period. The marked decrease in CO highlights the potential positive impacts of environmental regulations and technological advancements in North America, even as life expectancy improves.

The graph displays a clear upward trend in life expectancy in **South America** from 1970 to 2020, with a particularly steep increase after 1990. Conversely, the levels of most air pollutants remain relatively stable and low over the same period. However, CH_4 (Methane) exhibits a noticeable upward trend, indicating a potential growing source of this pollutant. CO (Carbon Monoxide) shows a steep increase initially but then experiences a slight decrease towards the later years. Despite these fluctuations in CO and the increase in CH_4, life expectancy continues to rise, suggesting that factors beyond these specific pollutants, such as advancements in healthcare and living standards, may be playing a more dominant role in influencing longevity during this period. The distinct trends of CH_4, with its persistent rise, and CO, with its initial steep increase followed by a slight decrease, highlight the need to consider the specific impacts of these pollutants on public health in South America, even as life expectancy improves.

The graph illustrates a consistent upward trend in life expectancy in **Oceania** from 1970 to 2020. In contrast, the levels of most air pollutants remain relatively stable and low over the same period. However, CH_4 (Methane) is notably high compared to other pollutants, although it shows a slight decrease towards the later years. CO (Carbon Monoxide) exhibits a significant decline, particularly after 1990, indicating a substantial reduction in its levels. Despite these changes in CH_4 and the significant decrease in CO, life expectancy continues its upward trajectory, suggesting that factors beyond these specific pollutants, such as advancements in healthcare and living standards, may be playing a more dominant role in influencing longevity during this period. The marked decrease in CO, coupled with the relatively high but slightly decreasing CH_4 levels, highlights the need to consider the specific impacts of these pollutants on public health in Oceania, even as life expectancy improves.

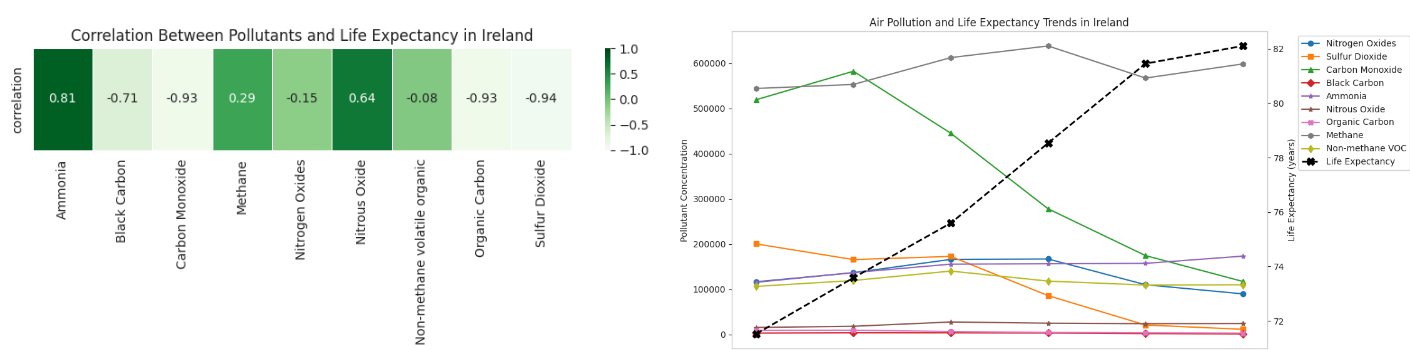
Across all continents, a consistent trend emerges. Life expectancy has shown a significant upward trajectory from 1970 to 2020, despite varying patterns in air pollutant levels. While most pollutants remained relatively stable or exhibited declines, notable exceptions included CO, which showed significant decreases in regions like Europe and North America, and CH4, which displayed increasing trends in Asia and South America. These observations suggest that factors beyond the specific pollutants depicted, such as advancements in healthcare, improved living standards, and potentially effective environmental regulations in certain regions, have played a more dominant role in driving the observed increase in longevity. However, the distinct trends of CO and CH4, along with their potential health implications, warrant continued monitoring and targeted interventions, even as overall life expectancy improves globally.

Figure 13. Air Pollutants factors and Life Expectancy in Canada



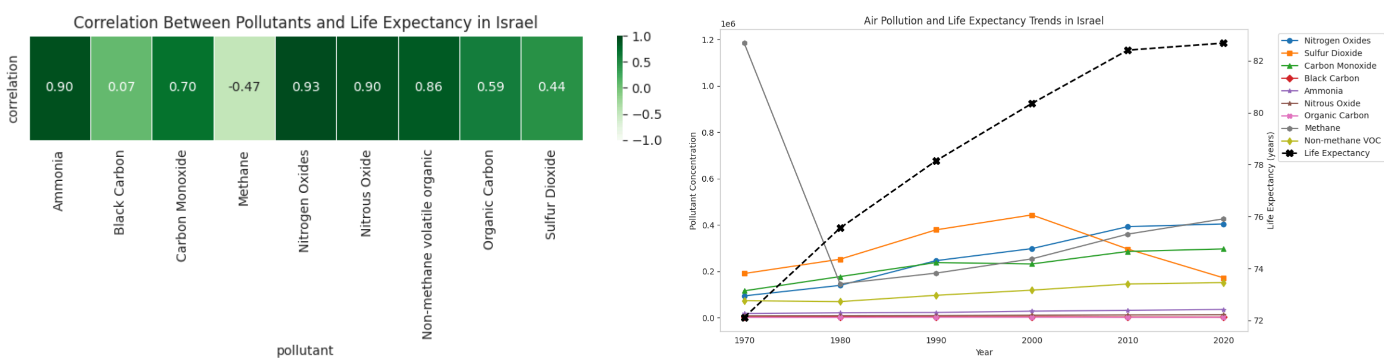
In **Canada**, life expectancy exhibits a consistent upward trend from 1970 to 2020, while the levels of various air pollutants display distinct patterns. Notably, Carbon Monoxide shows a significant decline, paralleled by decreases in Sulfur Dioxide and Nitrogen Oxides, suggesting successful mitigation efforts. Conversely, Methane displays a slight increase until 200 and then decreases, and other pollutants like Black Carbon, Ammonia, and Nitrous Oxide remain relatively stable at low levels. The correlation heatmap reveals strong positive correlations between life expectancy and Ammonia (0.93), Methane (0.75), and Nitrous Oxide (0.73), indicating that increases in these pollutants are associated with increased life expectancy, which is counterintuitive and likely influenced by other confounding factors. Conversely, strong negative correlations are observed for Carbon Monoxide (-0.88), Non-methane volatile organic compounds (-0.89), and Sulfur Dioxide (-0.99), suggesting that decreases in these pollutants are associated with increased life expectancy, aligning with expected trends. The negative correlation with Black Carbon (-0.73) and Nitrogen Oxides (-0.62) also indicate that lower levels of these pollutants are linked to higher life expectancy, while Organic Carbon shows a weak negative correlation (-0.30). These correlations, while statistically significant, likely reflect complex interactions and do not necessarily imply direct causation, necessitating further investigation to understand the underlying mechanisms.

Figure 14. Air Pollutants factors and Life Expectancy in Ireland



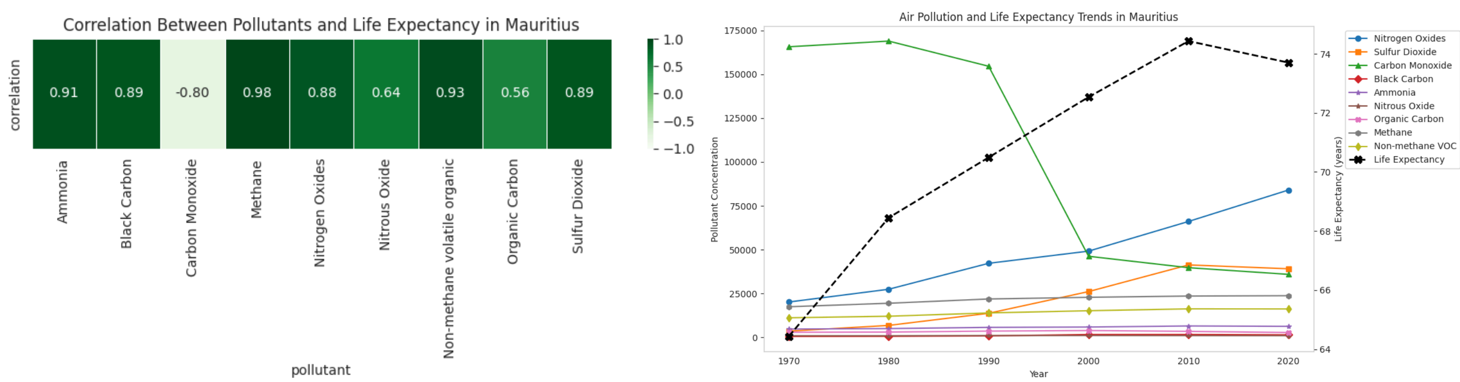
In **Ireland**, life expectancy demonstrates a consistent upward trend from 1970 to 2020, while the patterns of air pollutants are more varied. Carbon Monoxide exhibits a significant decline, particularly after 1980, suggesting effective mitigation strategies. Methane shows a gradual increase until around 2000, followed by a plateau and slight decline. Nitrogen Oxides and Sulfur Dioxide also show decreasing trends, while Ammonia, Black Carbon, Organic Carbon, Nitrous Oxide, and Non-methane VOC remain relatively stable at low levels. The correlation heatmap reveals strong positive correlations between life expectancy and Ammonia (0.81) and Nitrous Oxide (0.64), suggesting that increases in these pollutants are associated with increased life expectancy, which is counterintuitive and likely influenced by other confounding factors. Conversely, strong negative correlations are observed for Carbon Monoxide (-0.93), Non-methane volatile organic compounds (-0.93), and Sulfur Dioxide (-0.94), indicating that decreases in these pollutants are associated with increased life expectancy, aligning with expected trends. Black Carbon (-0.71) also exhibits a strong negative correlation, while Methane (0.29), Nitrogen Oxides (-0.15), and Organic Carbon (-0.08) show weak correlations. These correlations, while statistically significant, likely reflect complex interactions and do not necessarily imply direct causation, necessitating further investigation to understand the underlying mechanisms in Ireland.

Figure 15. Air Pollutants factor and Life Expectancy in Israel



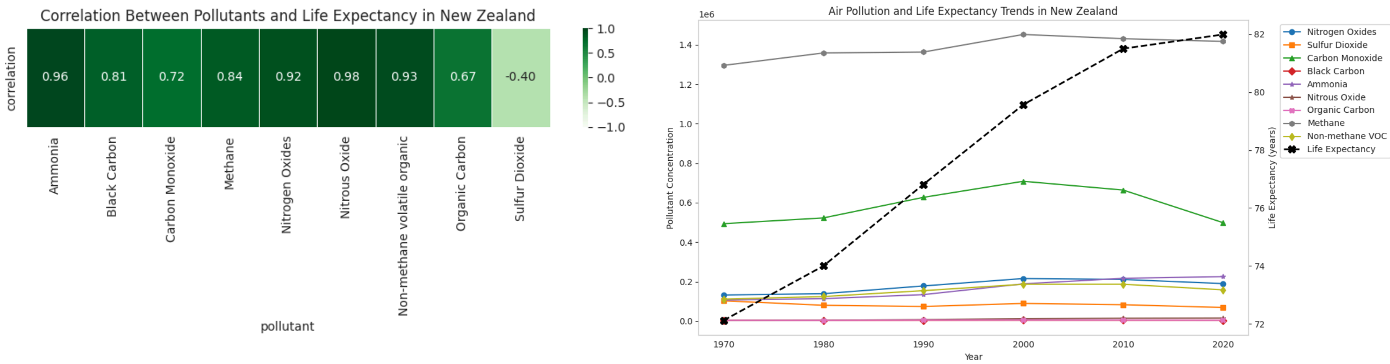
In **Israel**, life expectancy demonstrates a consistent and steep upward trend from 1970 to 2020. Air pollutant patterns, however, are more varied. Methane shows a dramatic spike and a rapid decline, suggesting a significant but temporary emission source early in the period. Non-methane VOC, Carbon Monoxide and Nitrogen Oxides exhibit gradual increasing trends, while Sulfur Dioxide shows a slight overall increase with a fluctuating pattern. Ammonia, Organic Carbon, Black Carbon, Nitrous Oxide remain relatively stable at low levels. The correlation heatmap reveals strong positive correlations between life expectancy and Ammonia (0.90), Carbon Monoxide (0.70), Nitrogen Oxides (0.93), Nitrous Oxide (0.90), Organic Carbon (0.86), and Sulfur Dioxide (0.44), suggesting that increases in these pollutants are associated with increased life expectancy, which is counterintuitive and likely influenced by other confounding factors. Conversely, Methane shows a negative correlation (-0.47), indicating that decreases in this pollutant are associated with increased life expectancy, aligning with expected trends after the initial spike. Black Carbon (0.07 exhibits a weak positive correlation. These correlations, while statistically significant for most pollutants, likely reflect complex interactions and do not necessarily imply direct causation, necessitating further investigation to understand the underlying mechanisms in Israel, especially considering the unusual Methane pattern and the counterintuitive positive correlations with pollutants typically associated with negative health impacts.

Figure 16. Air Pollutants factors and Life Expectancy in Mauritius



In **Mauritius**, life expectancy demonstrates a consistent upward trend from 1970 to 2010, followed by a slight decline towards 2020. Air pollutant patterns are diverse. Carbon Monoxide exhibits a dramatic decline, particularly after 1990, indicating successful mitigation efforts. Nitrogen Oxides and Sulfur Dioxide exhibit increasing trends, especially after 2000. Ammonia, Methane, Black Carbon, Nitrous Oxide, Organic Carbon, and Non-methane VOC remain relatively stable at low levels. The correlation heatmap reveals strong positive correlations between life expectancy and Ammonia (0.91), Black Carbon (0.89), Methane (0.98), Nitrogen Oxides (0.88), Nitrous Oxide (0.64), Non-methane VOC (0.93), Organic Carbon (0.56), and Sulfur Dioxide (0.89), suggesting that increases in these pollutants are associated with increased life expectancy, which is counterintuitive and likely influenced by other confounding factors. Conversely, Carbon Monoxide shows a strong negative correlation (-0.80), indicating that decreases in this pollutant are associated with increased life expectancy, aligning with expected trends. The strong positive correlations for most pollutants, despite the recent decline in life expectancy, necessitate further investigation to understand the complex interactions and potential confounding factors influencing these trends in Mauritius.

Figure 17. Air Pollutants factors and Life Expectancy in New Zealand

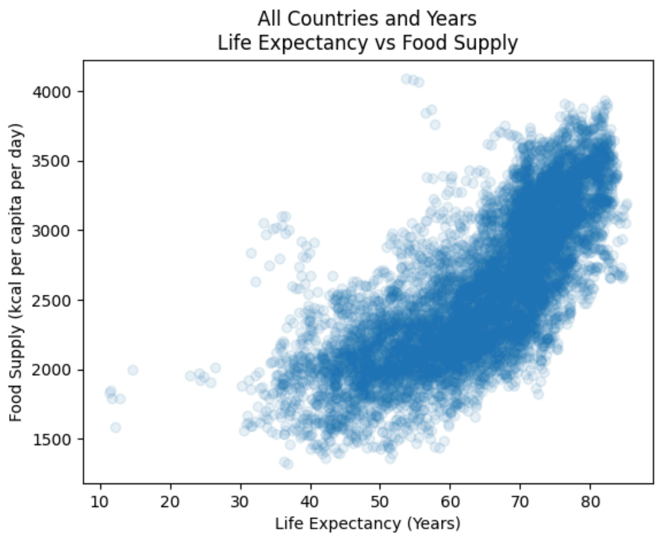


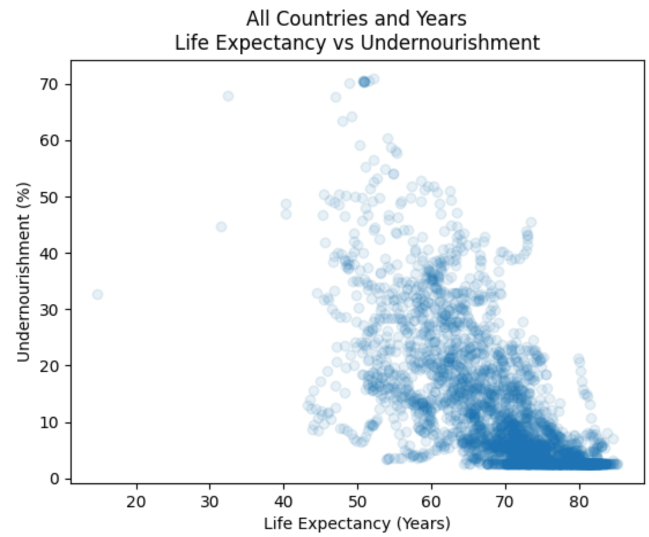
In **New Zealand**, life expectancy exhibits a consistent and strong upward trend from 1970 to 2020. Air pollutant patterns, however, are more varied. Methane has a gradual increase, becoming the dominant pollutant by concentration. Carbon Monoxide initially shows a gradual increase until around 2000, followed by a decline. Nitrogen Oxides and Sulfur Dioxide, Ammonia, Organic Carbon, Black Carbon, Nitrous Oxide, and Non-methane VOC remain relatively stable at low levels. The correlation heatmap reveals strong positive correlations between life expectancy and Ammonia (0.96), Black Carbon (0.81), Carbon Monoxide (0.72), Methane (0.84), Nitrogen Oxides (0.92), Nitrous Oxide (0.98), Non-methane VOC (0.93), and Organic Carbon (0.67), suggesting that increases in these pollutants are associated with increased life expectancy, which is counterintuitive and likely influenced by other confounding factors. Conversely, Sulfur Dioxide shows a negative correlation (-0.40), indicating that decreases in this pollutant are associated with increased life expectancy, aligning with expected trends. The strong positive correlations for most pollutants, particularly the dominant Methane, necessitate further investigation to understand the complex interactions and potential confounding factors influencing these trends in New Zealand.

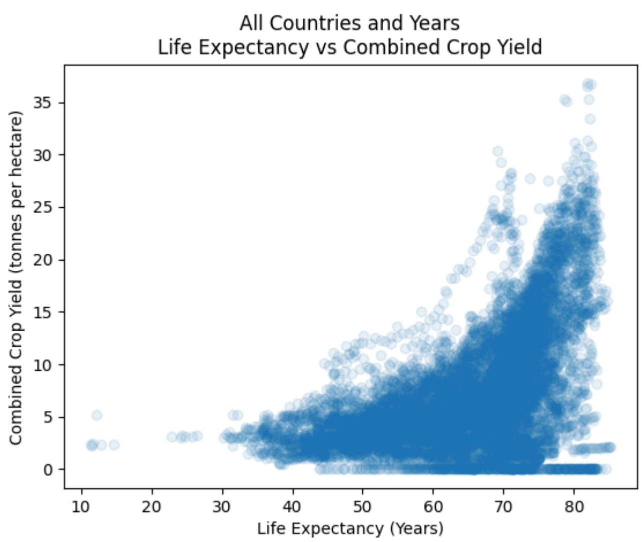
Life Expectancy and Agriculture

When investigating life expectancy and how it relates to agriculture, the datasets of food supply, various crop yields, and undernourishment were chosen to approach the analysis from different angles to see which would be most relevant. All the factors were observed first from a summary and correlation perspective that looked at all countries and years together.

Figure 18. Agriculture metrics for all countries and years with correlation matrix of factors







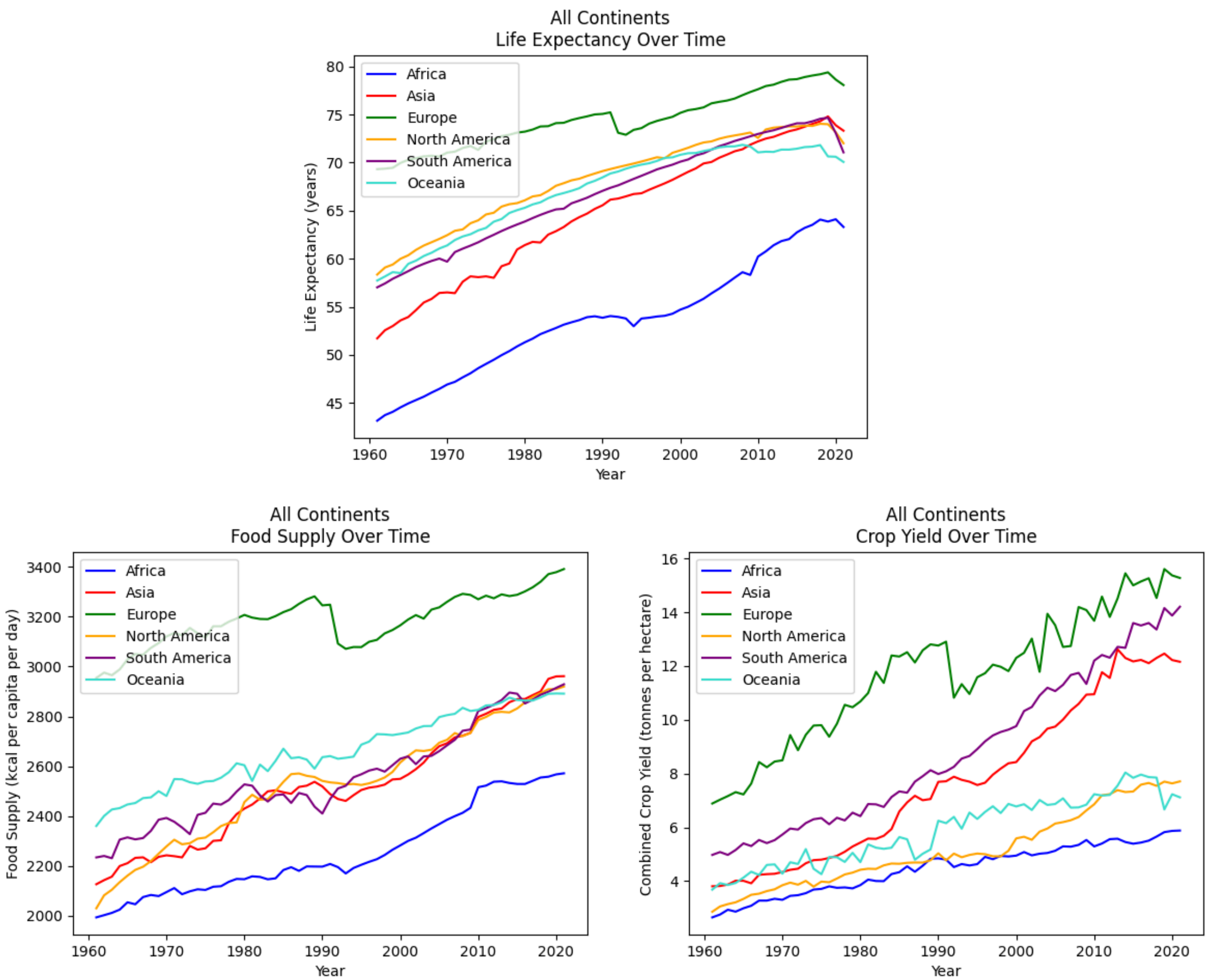
	LifeExpectancy	FoodSupply	CornYield	OatYield	RiceYield	WheatYield	CombYield
LifeExpectancy	1.000000	0.759688	0.442896	0.464684	0.140786	0.413304	0.523829
FoodSupply	0.759688	1.000000	0.488140	0.584440	0.166874	0.548232	0.626822
CornYield	0.442896	0.488140	1.000000	0.354137	0.210920	0.414372	0.826568
OatYield	0.464684	0.584440	0.354137	1.000000	-0.009503	0.788513	0.658622
RiceYield	0.140786	0.166874	0.210920	-0.009503	1.000000	0.048867	0.487573
WheatYield	0.413304	0.548232	0.414372	0.788513	0.048867	1.000000	0.732161
CombYield	0.523829	0.626822	0.826568	0.658622	0.487573	0.732161	1.000000

	LifeExpectancy	Undernourishment
LifeExpectancy	1.000000	-0.697082
Undernourishment	-0.697082	1.000000

These figures show that of the three observed agriculture metrics, food supply had by far the strongest correlation with life expectancy. Undernourishment was also a highly correlated factor, but the dataset had fewer countries and years to analyze, so food supply was determined to be the main metric moving forward with the analysis.

To analyze the effect of agriculture on life expectancy from a different perspective, the same metrics were observed over time while being split per continent. This view shows us how the metrics might have evolved over time and how the relative age of countries in each continent could impact how important agriculture is in determining life expectancy.

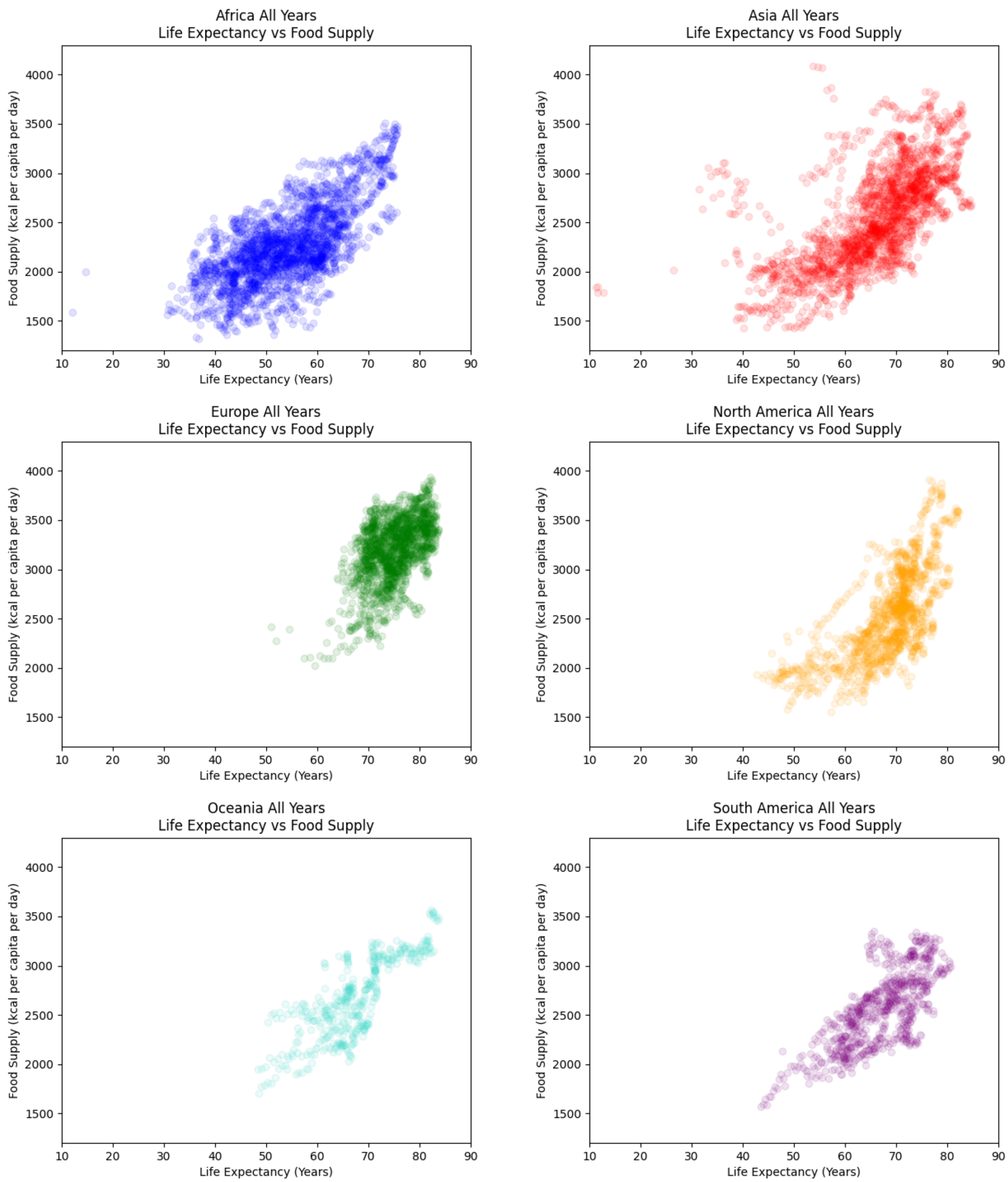
Figure 19. Life expectancy, food supply, and crop yield by continent over time



These figures show a positive trend over time for all three metrics, with occasional simultaneous decreases for large events such as the fall of the Soviet Union in the 1990's for Europe or the global pandemic in 2020 across the globe. These graphs more clearly show why food supply was most correlated with life expectancy over crop yield because there is a clear split in the graph for crop yield between Asia & South America, and North America & Oceania despite all four continents having similar life expectancies and food supplies. Europe and Africa remain consistently at the top and bottom respectively of all three metrics.

To further investigate the difference in continents for the most relevant factor of food supply, we will return to a more in-depth look at the first style of chart where life expectancy is compared to food supply across all countries and years, but this time we'll observe the unique differences in each continent and how they might have contributed to the initial plot.

Figure 20. Life expectancy vs food supply across all years and countries, split by continent



Africa seems to have the largest variability of all continents. There are a few countries in Africa that have food supply and life expectancies similar to other continents, but the vast majority of countries in Africa seem to be low on both metrics with a wide spread of how much they impact each other.

Asia has the most visible outliers of all continents, but overall has a large and noticeable curve with many countries being low on both food supply and life expectancy, but then a shift towards just as many countries that have an increased life expectancy without increasing food supply as much and also countries with both high food supply and life expectancy.

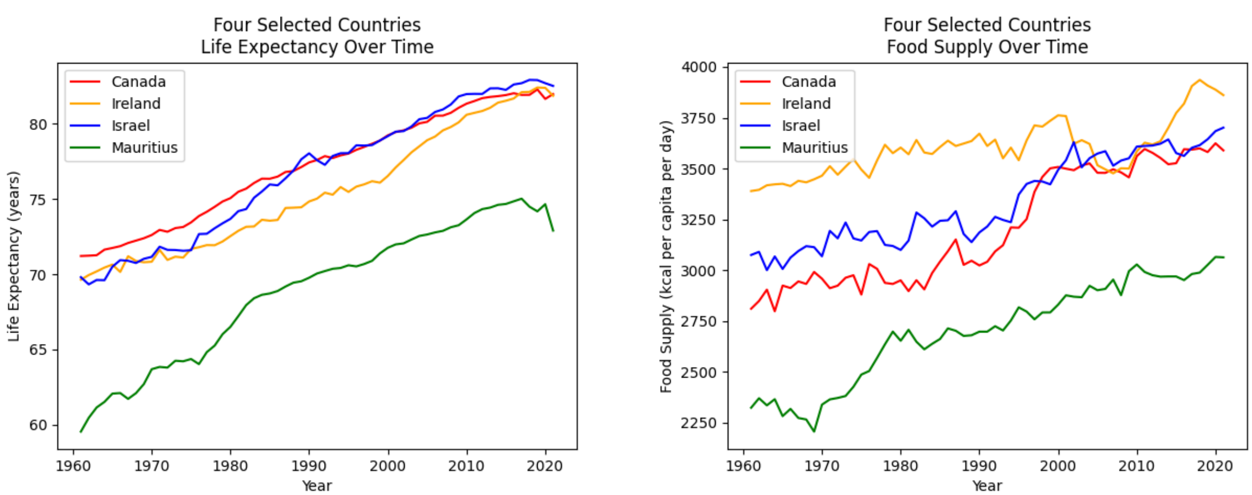
Europe is very densely focussed on both high food supply and high life expectancy. This continent has the least variation in both metrics and overall seems to be very well-established, with all points being at the top end of the trend curve seen in other continents.

North America has a trend very similar to Asia. It captures all points of the growth curve, but with a thinner overall trend, showing how there's fewer countries and years to observe in this continent when compared to Asia.

Oceania and **South America** can be analyzed together since they both show very similar trends and both have the fewest data points of all the continents. They seem to capture the midpoint of the growth curve, with very few countries at the high end of food supply, but also very few countries at the low end of life expectancy.

Finally, to try and compare agriculture and life expectancy to the other major metrics, we selected four countries (Canada, Ireland, Israel, and Mauritius) to investigate for any specific trends. We see that Ireland is the most inconsistent among these selected countries, with higher food supply than the others, but consistent or lower life expectancy across all years.

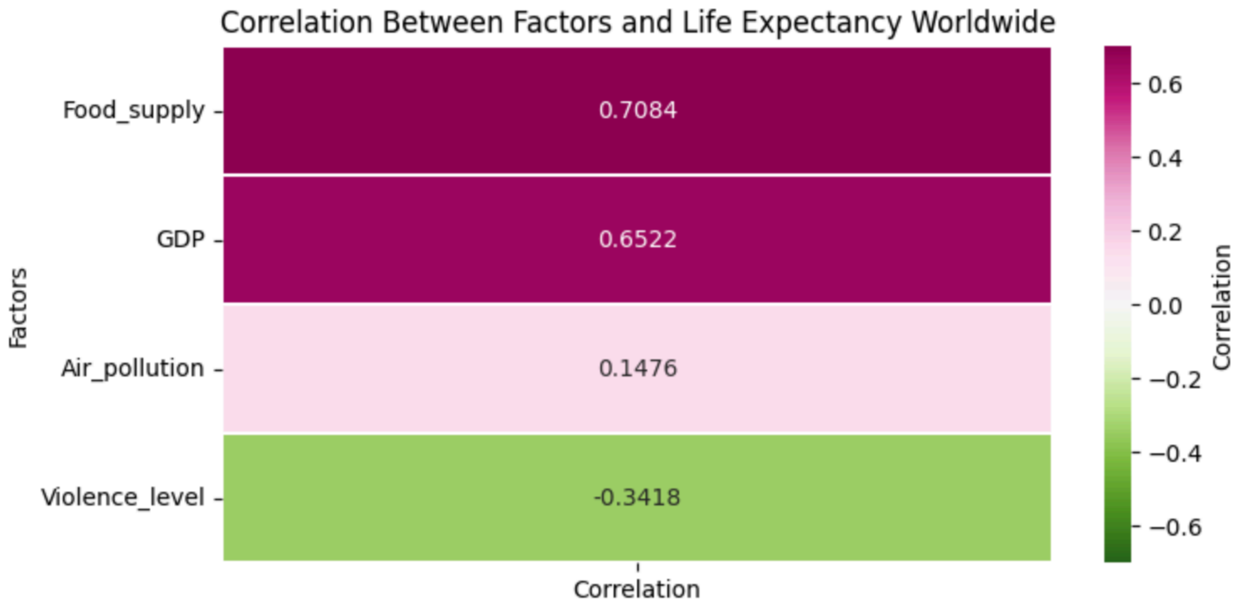
Figure 21. Comparison of cour selected countries in life expectancy and food supply over time



The World

Based on the results for each factor, we identified the variable with the greatest impact on life expectancy within each category. In the Economy, the most influential variable is GDP; in Violence, it is the Safety and Security Score; in Air Pollution, it is Nitrogen Oxides; and in Food/Agriculture, it is Food Supply.

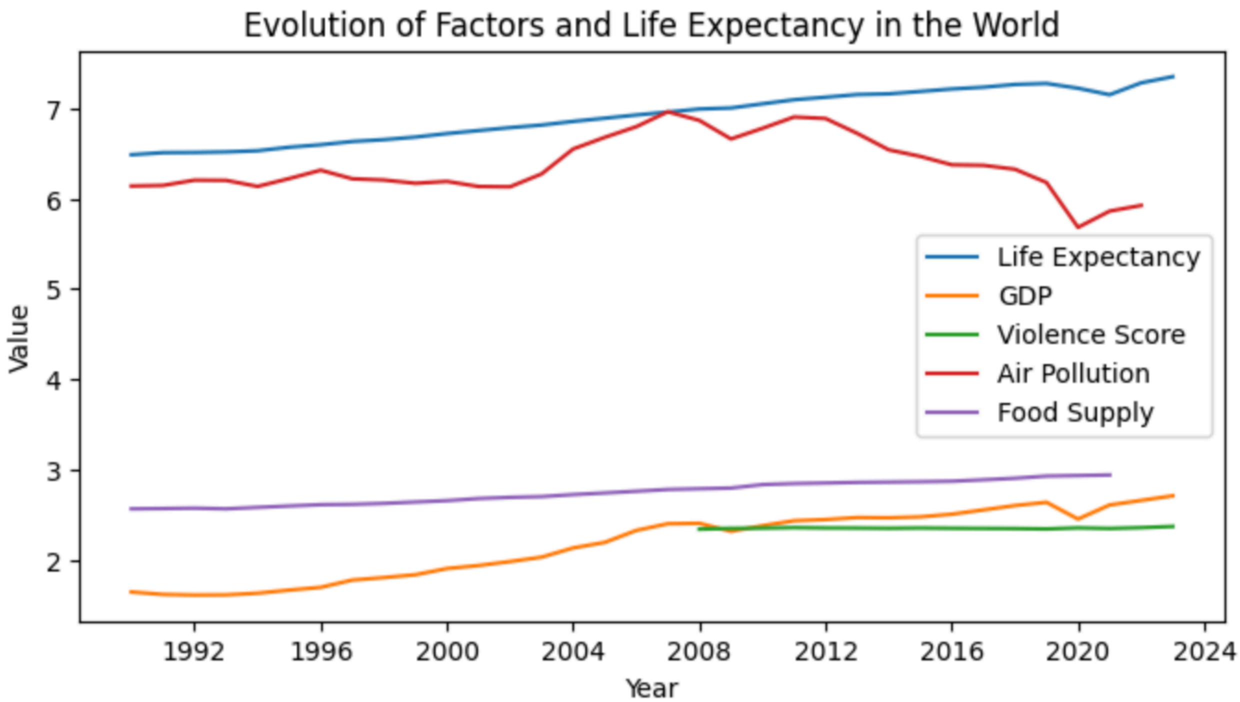
Figure 22. Global Correlation Between Key Factors and Life Expectancy



The factor with the highest correlation to life expectancy is Agriculture (Food Supply), with a positive correlation of 0.71, meaning that as food supply increases, life expectancy also increases. The second strongest correlation is with Economic Factors (GDP), showing a positive correlation of 0.65, indicating that higher GDP is associated with increased life expectancy.

On the other hand, Violence Levels exhibit a negative correlation of -0.342, meaning that as violence decreases, life expectancy tends to rise. Lastly, Air Pollution shows a weak positive correlation of 0.1476; however, this could be due to a dataset-wide increase in both air pollution and life expectancy, potentially introducing bias into the correlation. Further analysis is necessary to better understand this relationship.

Figure 23. Global Evolution Between Key Factors and Life Expectancy



For a more comprehensive analysis, the variables were adjusted to be displayed on a single graph: Life Expectancy was divided by 10, GDP by 10,000, Air Pollution by 100,000, and Food Supply by 1,000.

In this graph, we can observe that air pollution has been declining since 2008, which reflects the political, economic, and social shifts driven by growing awareness of the dangers of these gases. This has led to a global reduction in nitrogen oxide emissions.

GDP has increased significantly, particularly since the 2000s, maintaining an upward trend. However, there is a noticeable dip in 2009, likely due to the global economic instability between 2008 and 2010.

Another major achievement for humanity is Agriculture (Food Supply), which has shown a steady upward trend since 1992. As previously noted, this factor has the highest correlation with life expectancy, reinforcing the idea that ensuring proper nutrition—one of the most fundamental human needs—is essential for increasing lifespan.

Finally, we can also observe a declining trend in violence levels over time.

Conclusion

Economy Summary

This analysis shows that GDP plays an important role in life expectancy. Countries with higher GDP tend to have better healthcare, living conditions, and access to essential services, leading to longer lifespans. On the other hand, lower-GDP countries often face challenges like poor healthcare and higher disease rates, resulting in shorter life expectancy. While GDP is not the only factor, it is a key indicator of a country's ability to support the well-being of its people. This highlights the need for policies that promote both economic growth and fair access to resources to improve life expectancy worldwide.

Violence Summary

The factor with the greatest impact on violence was safety and security across most continents. However, in countries with the highest life expectancy per continent, the most significant variable was the Global Peace Index (GPI). In these countries, the GPI was generally much better than the average, making it the key factor affecting life expectancy. It is also important to note that, beyond just examining the correlation, it is essential to consider the trends of these variables in recent years so governments can take necessary actions. This is particularly relevant for the countries with the worst scores in recent years, as seen in three out of five countries for violent crime.

Air Pollution Summary

Across the analysis of air pollutants and life expectancy, trends varied by continent and country, revealing complex relationships. While life expectancy generally increased, patterns in pollutant levels differed significantly. In some regions, like Europe and North America, reductions in pollutants such as Carbon Monoxide (CO) and Sulfur Dioxide (SO_2) coincided with rising life expectancy, suggesting positive impacts from environmental regulations.However, in other areas, certain pollutants showed counterintuitive positive correlations with life expectancy, as seen in Israel and Mauritius. New Zealand presented a unique case with Methane becoming the dominant pollutant. Therefore, effective strategies to improve public health require careful consideration of regional and pollutant-specific trends.

Agriculture Summary

Overall, the analysis of agriculture and life expectancy showed that of the factors investigated, food supply was the most significant. We saw that the main trend across all continents together was a reverse ‘L’ curve where life expectancy increased gradually with food supply, with more pronounced effects occurring in the middle range of countries. We saw a similar curve with crop yield, but with a larger variability due to the limited selection of crops that were selected to be analyzed. Undernourishment also showed a distinct trend where life expectancy increased as undernourishment decreased, but the dataset had far fewer countries and years available for analysis, so we didn't analyze beyond the surface-level trend. When looking further into food supply, we saw that each continent captured a unique portion of the overall trend, with Europe having a small, condensed trend at the high end of the plot, and Africa having a large, varied trend at the lower end of the plot.

Global Summary

On a global scale, we observed that the factor with the strongest correlation was agriculture (food supply), which confirms that basic human needs, such as food, have a significant impact on life expectancy. The second most influential factor was the economy, highlighting the importance of a country's economic stability in improving life expectancy. In the evolution of the air pollution factor, we can see a significant decrease in nitrogen oxide emissions, a change driven by political, economic, and social trends addressing the dangers of polluting gases and efforts to reduce their emissions.

Conclusion

By investigating the effect of various elements on the life expectancy of a population, this study has provided a data-driven approach on understating the trends of life expectancy in various countries and continents. We observed a variety of large-scale metrics and analyzed them using a variety of smaller-scale datasets. We were able to analyze all these datasets through our use of SQL servers and SQL queries to extract valuable insights from databases with numerous rows. We learned how to process big data and produce insights beyond what a surface-level observation of the data would be able to accomplish.

References

- Our World in Data. (n.d.). Our World in Data. Retrieved March 2, 2025, from https://ourworldindata.org/
- World Health Organization. (n.d.). World Health Organization. Retrieved March 2, 2025, from https://www.who.int/
- World Bank. (n.d.). World Bank Open Data. Retrieved March 2, 2025, from https://data.worldbank.org/
- United Nations Development Programme. (n.d.). United Nations Development Programme. Retrieved March 2, 2025, from https://www.undp.org/
- Institute for Economics & Peace. Global Peace Index March 2, 2024. https://www.visionofhumanity.org/maps/#/