Analysis of findings

Introduction

As part of the project looking into life-expectancy based on health and development indicators per region, the fourth part was to look at the life expectancy data for the American regions in more details. In preparation to the analysis, the original csv dataset was cleaned and reduced to only reflect one decade (2004-2014) and the regions including North America, Central America, South America and the Caribbean.

Main findings

To start off the analysis, the mean life expectancy was calculated and the results sorted by region. The table (Tab 1) shows the results, however, to aid comparison, the highest mean and lowest mean life expectancy with country per region was identified in the following steps. Per region, the highest and lowest life expectancy respectively were as follow:

Tab 2: Highest Mean Life Expectancy per Region

	Country	Region	Life expectancy
14	Jamaica	Caribbean	74.645455
6	Costa Rica	Central America	78.772727
3	Canada	North America	82.463636
4	Chile	South America	79.944444

Tab 3: Lowest Mean Life Expectancy per Region

	Country	Region	Life expectancy
12	Haiti	Caribbean	62.900000
9	El Salvador	Central America	72.145455
3	Canada	North America	82.463636
11	Guyana	South America	65.690909

Along with calculating the mean life expectancy, the actual highest and lowest life expectancy age per region was also identified. The results show that for North America, Central America and South America, the highest life expectancy was recorded in the years 2008 and 2014, while the lowest life expectancy in the same regions was recorded in 2004 and 2005. Only in the Caribbean region, the highest life expectancy was recorded in 2012 and the lowest in 2013 (both in different countries). Though not part of this dataset, a separate search into the potential cause identified Hurricane Sandy to be a strong contender for dishevel in Haiti in 2013 (Ref 1).

To add to the maximum and minimum actual life expectancy, a graph was prepared to visualise the distribution of across the entire dataset. The graph shows the highest life expectancy count for the decade between 2004 and 2014 is between 72 and 76 years.

Fig 1: Histogram of life expectancy distribution across all American regions
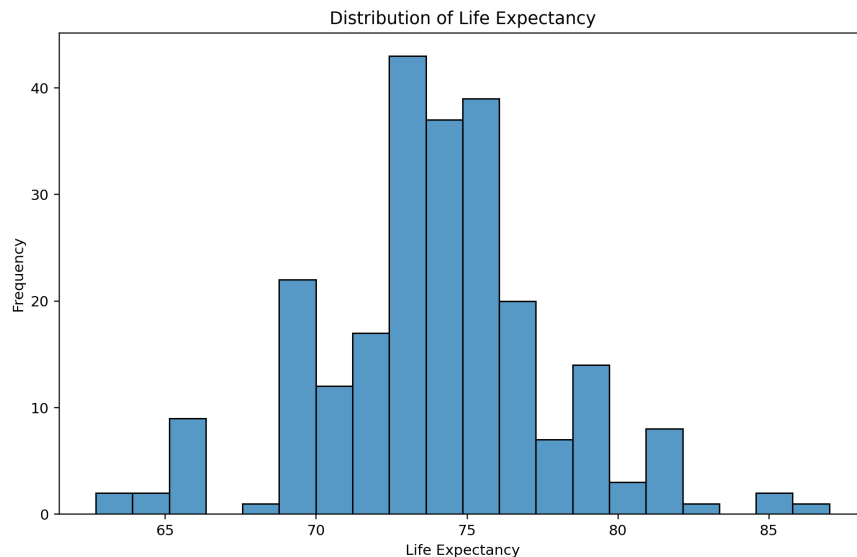
The next step was to have a closer look at the life expectancy distribution and change per country per region. For this purpose, four scatterplots were created as per the below.

Fig 2: Change in life expectancy between 2004 and 2014 in the Caribbean
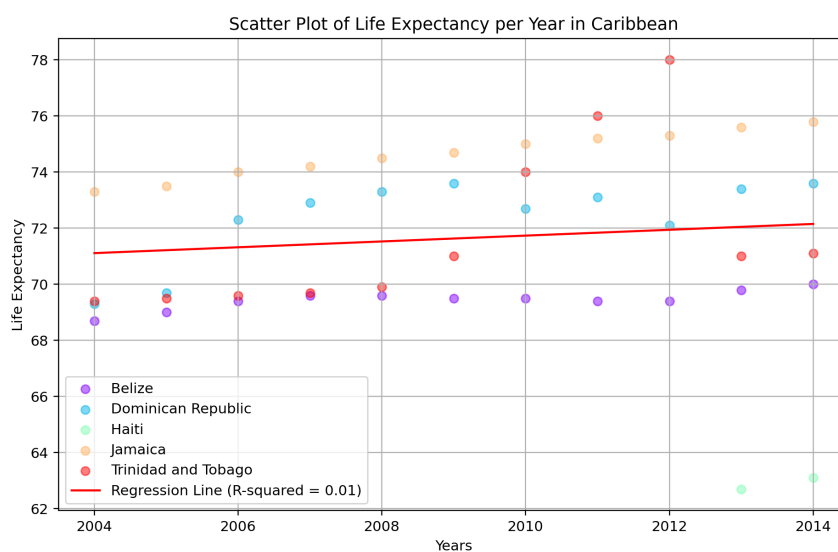
In the Caribbean region, in general, all countries show a rise in life expectancy. Trinidad and Tobago has a rather sharp rise between 2010 and 2012 after which the age falls again in 2013 and 2014.

Fig 3: Change in life expectancy between 2004 and 2014 in Central America

Scatter Plot of Life Expectancy per Year in Central America
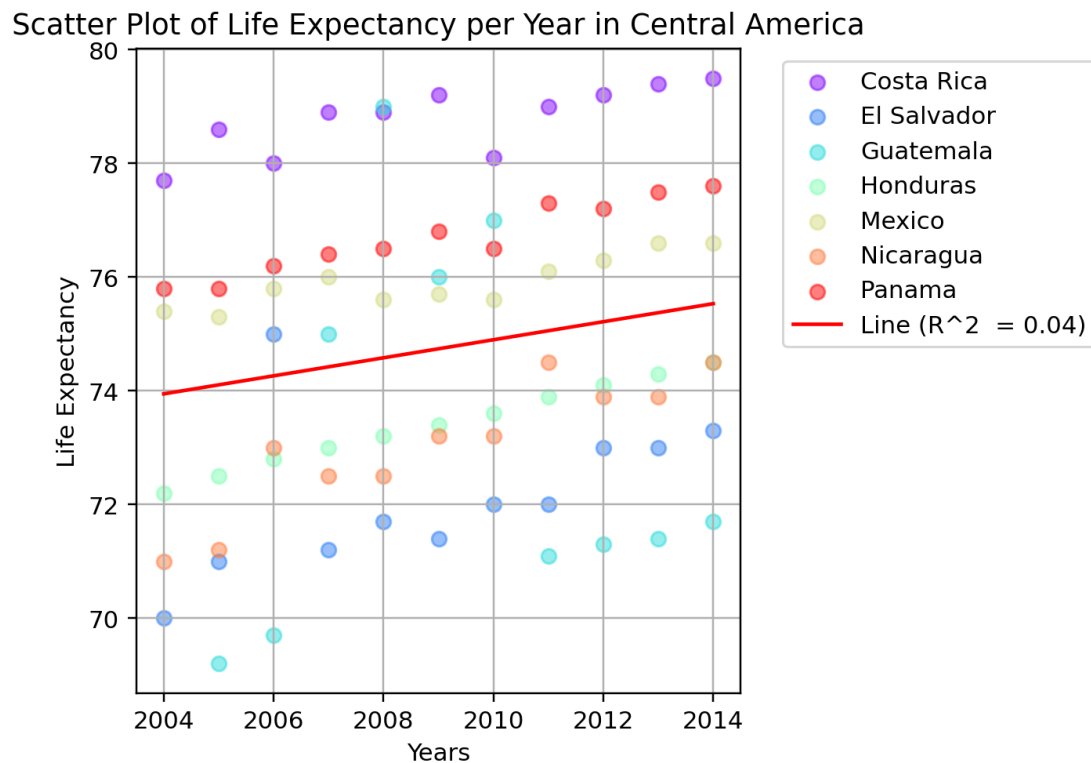
In Central America, all countries show an overall increase in life expectancy. The data indicates to a rather large spike in life expectancy for El Salvador and Guatemala in 2008 and 2006 respectively.

Fig 4: Change in life expectancy between 2004 and 2014 in South America

Scatter Plot of Life Expectancy per Year in South America
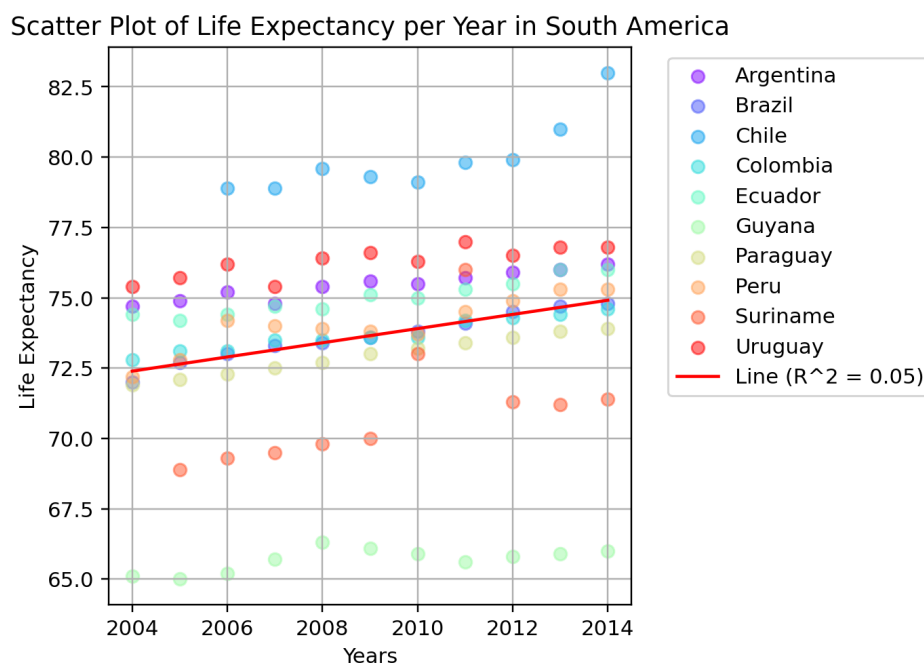
The scatterplot for South America shows similar trends to that of the Central American countries. For majority of countries, life expectancy rose at the same time and was mostly around the yer 72 and 77. The highest disparity was that between Chile and Guyana.

Fig 5: Change in life expectancy between 2004 and 2014 in North America
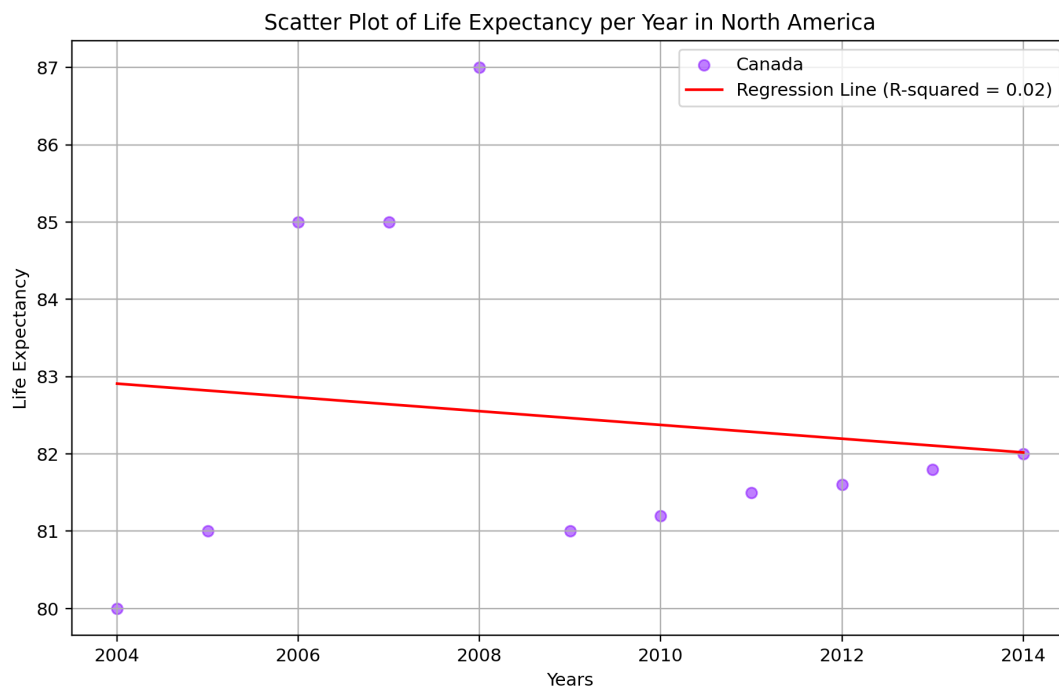
The dataset only had Canada as a country. For uniformities sake, the data was also visualised in a scatterplot. Purely looking at the data points, Canada shows a steady rise of life expectancy between the age of 80 and 82, with three spikes between 2006 and 2008. Due to these outlier, the regression is showing a downward trend.

As a last task, a simple comparison between developing and developed countries across all regions in the dataset was drawn. The below box-plot shows a wider data distribution for developing countries and a higher mean age for developed countries.

Fig 6: Life expectancy distribution in developing vs developed countries
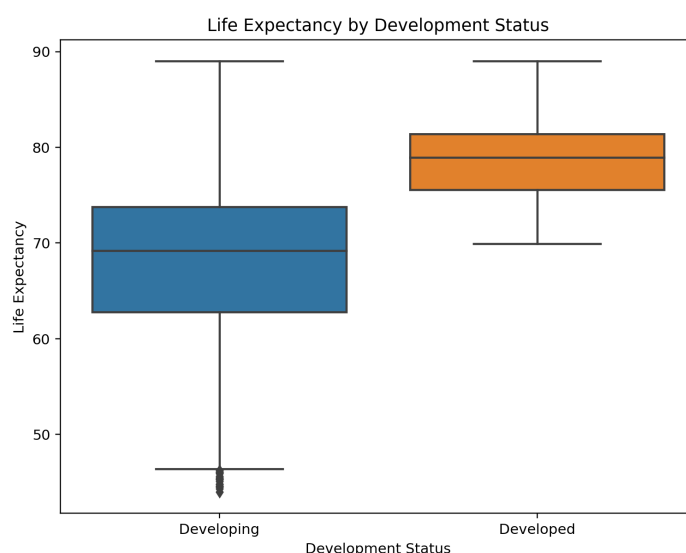
Limitations

During the analysis, a few issues with the dataset were noted. Firstly, for North America, USA as a country was missing. As such, a comparison between just Canada and the other regions is not likely to give a tangible indication of difference or similarity. Further, the status of countries were often unexpected, for example Canada was labeled as a developing country which is incorrect (Ref 2). Additionally, political and environmental factors were missing from the dataset which would have helped in exploring the cause for the sudden drop or rise in life expectancy in some countries over the decade. Lastly, given more time, a further search for country wise data would have aided in cleaning the dataset to make it more accurate.

Conclusion

Over the decade between 2004 and 2014, most countries in the American regions shows a rise in live expectancy. The causation of the trends and potential outliers is outside of the scope of this analysis.

Reference

The tables and figures indicated are to be found in the Jupyter Notebook related to this analysis on the GitHub repository:

Tab 1: 3. Mean life expectancy across all countries per region (Project_1_American Regions.ipynb)
Tab 2: 6. Mean life expectancy - highest per region (Project_1_American Regions.ipynb)
Tab 3: 7. Mean life expectancy - lowest per region (Project_1_American Regions.ipynb)

Ref 1: https://www.bbc.co.uk/news/world-latin-america-20320487
Ref 2: https://www.un.org/en/development/desa/policy/wesp/wesp_current/2014wesp_country_classification.pdf

Fig 1: 8. Actual life expectancy distribution across all countries (Project_1_American Regions.ipynb)
Fig 2: 1. CARIBBEAN (Project_1_American Regions.ipynb)
Fig 3: 2. CENTRAL AMERICA (Project_1_American Regions.ipynb)
Fig 4: 3. SOUTH AMERICA (Project_1_American Regions.ipynb)
Fig 5: 4. NORTH AMERICA (Project_1_American Regions.ipynb)
Fig 6. C. Bonus Analysis (Project_1_American Regions.ipynb)