Social Science problem: Reduction of the population

Jieun Park 100509696
Advanced Modeling
Universidad Carlos III de Madrid
2023-2024 Academic Year

Introduction

In recent years, the global fertility rate has been declining, leading to an overall decrease in the human population. This phenomenon has prompted considerable attention to understanding the underlying reasons for the reduction in fertility rates, with a focus on various societal factors. In Korean society, for example, factors such as a decrease in savings and investment, an increase in dependency ratios (Hyun-Chool, L. E. E., 2021), and rising housing prices have been identified as contributors to declining fertility rates and population size.

However, I contend that alongside the factors influencing fertility rates, understanding the determinants of mortality rates is equally crucial in addressing global population decline. Mortality levels are influenced by a complex interplay of sociocultural, personal, biological, and medical factors (Tsai, S. P., Lee, E. S., & Hardy, R. J., 1978), underscoring the importance of measuring and investigating these factors. Research by Roth, G. A. et al. highlights a shift in mortality trends, with a decline in deaths from communicable, maternal, neonatal, and nutritional causes, an increase in noncommunicable diseases, and stable rates of injury deaths (Roth, G. A., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., ... & Borschmann, R., 2018).

To assess the influence of factors impacting the global population, I employed unsupervised machine learning methods, namely PCA and K-means clustering. These analyses aim to uncover underlying patterns and associations among various factors affecting mortality rates, providing valuable insights into the dynamics of population decline.

Descriptive Analysis of data

Dataset has 159 rows and 11 columns in total. Variables are country, continent, year, life_value (life expectancy), obesity_value, tobacco_value, doc_den_value (doctor density rate), gob_expenditure (government health expenditure), road_death, birth_by_skilled (birth by skilled health personnel rate), and maternal_death (maternal mortality rate). Year is the same for the whole dataset as 2015. Life expectancy has a minimum value of 47.67 and maximum value of 84.29. Obesity rate has a minimum value of 2.10, maximum value of 53.90 and 2 NAs. Tobacco value has a minimum value of 0.20, maximum value of 39.10 and 21 NAs. Doctor density rate has a minimum value of 0.315, maximum value of 77.586 and 65 NAs. Government health expenditure has a minimum value of 1.330, maximum value of 29.490 and 3 NAs. Road death has a minimum value of 0 and maximum value of 63.37. Birth by skilled health personnels has a minimum value of 39.70, maximum value of 100 and 77 NAs. Finally, the maternal mortality rate has a minimum value of 1.314 and maximum value of 1224.722. Because of the high variance of variable maternal mortality rate, I did the log transformation.

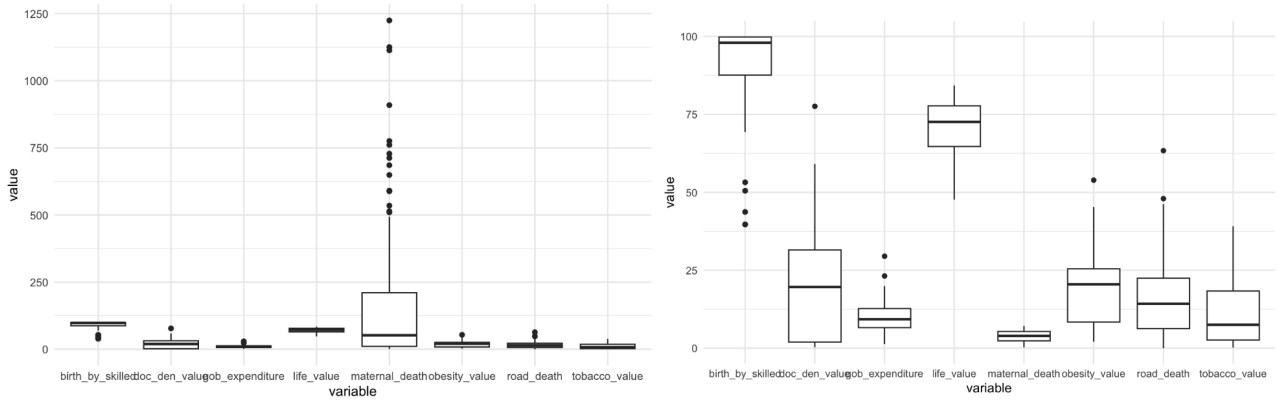

Figure 1. Boxplot before the log-transformation of maternal mortality rate (left) and after (right)

1. Life Expectancy V.S. Obesity

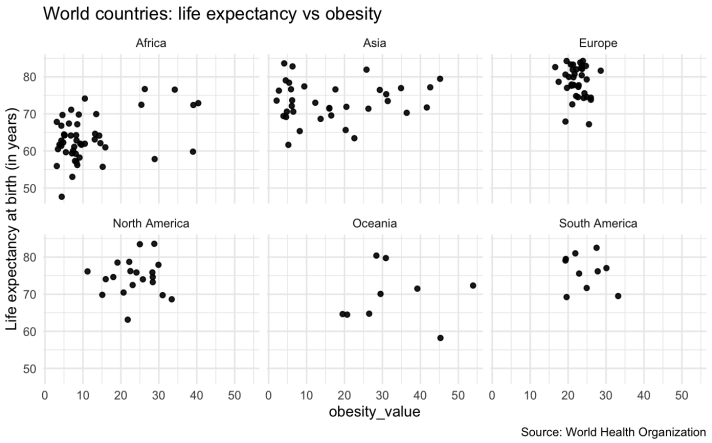

Figure 2. Life Expectancy and Obesity

Countries in Europe, North America, Oceania, and South America generally exhibit neutral obesity rates, while Asia shows a more scattered distribution. Africa demonstrates two extreme tendencies in obesity rates. Despite overall high or neutral life expectancy across continents, certain countries in Africa have notably low life expectancy. When considering the relationship between obesity rate and life expectancy, most continents display neutral or high life expectancy with neutral obesity rates. However, Africa stands out, with most countries showing low obesity rates and varied patterns in life expectancy, including both high and low rates.

2. Life Expectancy V.S. Tobacco Usage

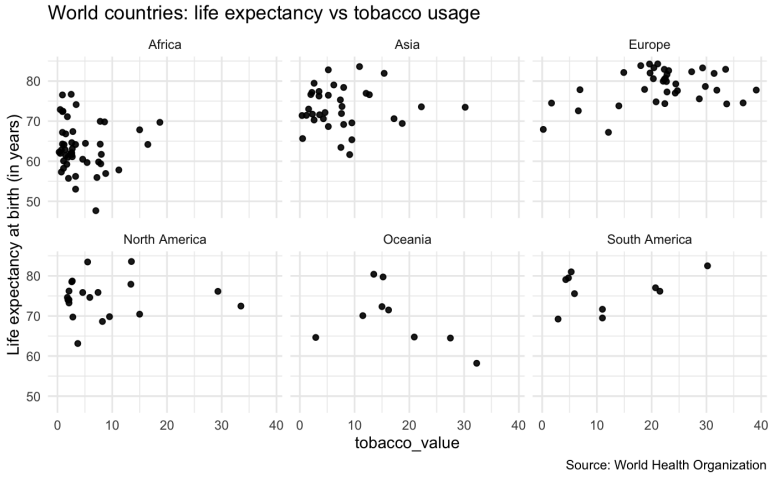

Figure 3. Life Expectancy and Tobacco Usage

In many European countries, tobacco usage is high, while countries in Africa, Asia, and North America generally have lower tobacco usage rates. However, when considering the relationship between tobacco usage and life expectancy, there appears to be no significant impact of tobacco usage on life expectancy rates overall.

3. Life Expectancy V.S. Doctor Density Rate

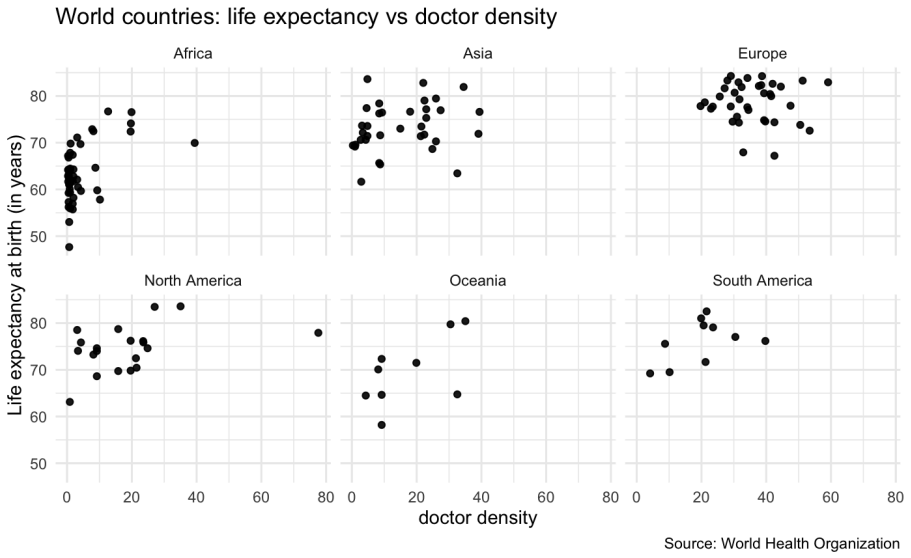

Figure 4. Life Expectancy and Doctor Density Rate

The graph illustrates that, apart from Europe, many countries across different continents, particularly in Africa, Asia, and North America, exhibit low doctor density. Despite the absence of high doctor density, overall life expectancy remains relatively high across countries except for those in Africa. This suggests that the majority of African countries face a shortage of medical professionals, potentially contributing to higher death rates in these regions.

4. Life Expectancy V.S. Government Health Expenditure

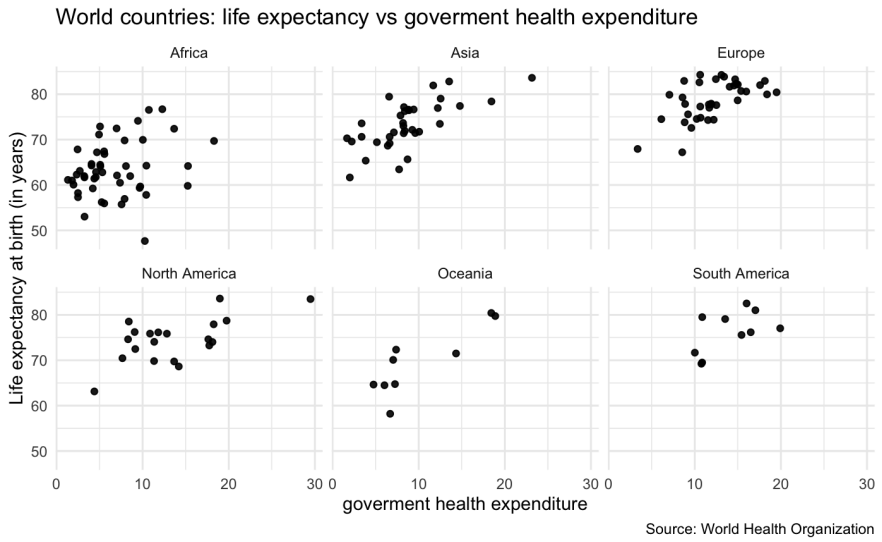

Figure 5. Life Expectancy and Government Health Expenditure

In numerous countries across Africa and Asia, governmental expenditure on healthcare is notably low. Despite the absence of an explicit relationship, there seems to be a positive linear correlation between government health expenditure and life expectancy. This suggests that increased government investment in healthcare is associated with higher life expectancy, highlighting the importance of adequate funding for improving health outcomes.

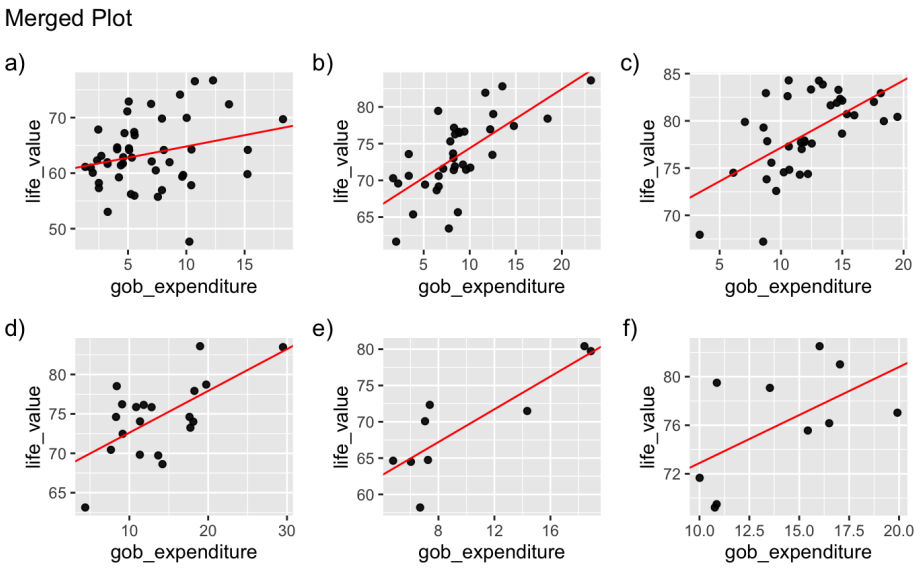

Figure 6. Life Expectancy and Government Health Expenditure with regression analysis

*a) Africa, b) Asia, c) Europe, d) North America, e) Oceania, and f) South America

Figure 5 depicts the linear relationship between government health expenditure and life expectancy across continents. As anticipated, a positive linear relationship is observed, with Africa exhibiting a weaker correlation compared to other continents.

5. Life Expectancy V.S. Road Death

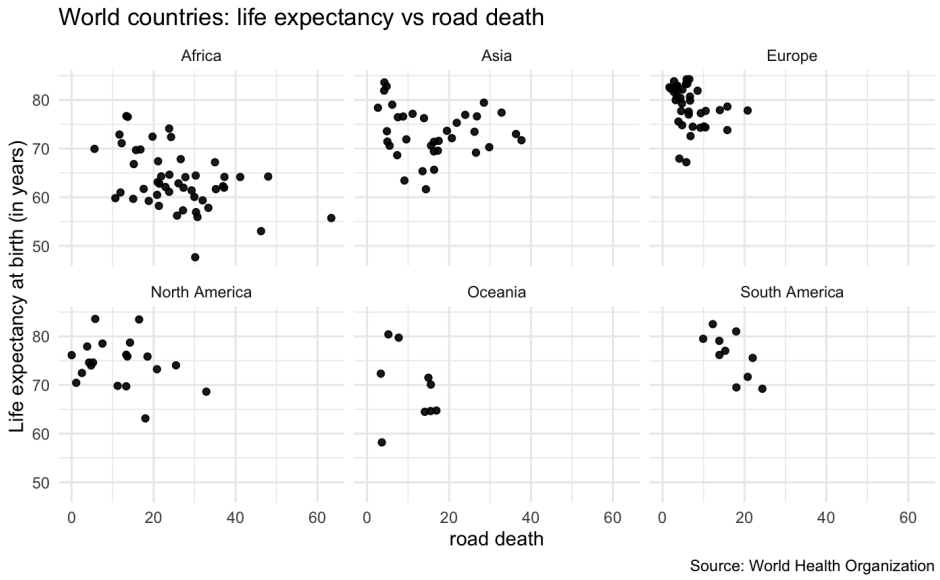

Figure 7. Life Expectancy and Road Death

As anticipated, there appears to be a negative relationship between road deaths and life expectancy, whereby lower road deaths are associated with higher life expectancy. For instance, countries with high road death rates in Africa exhibit notably lower life expectancies. In contrast, Europe, North America, Oceania, and South America generally have low road death rates. However, African and Asian countries demonstrate a more varied distribution, with some exhibiting both low and high road death rates.

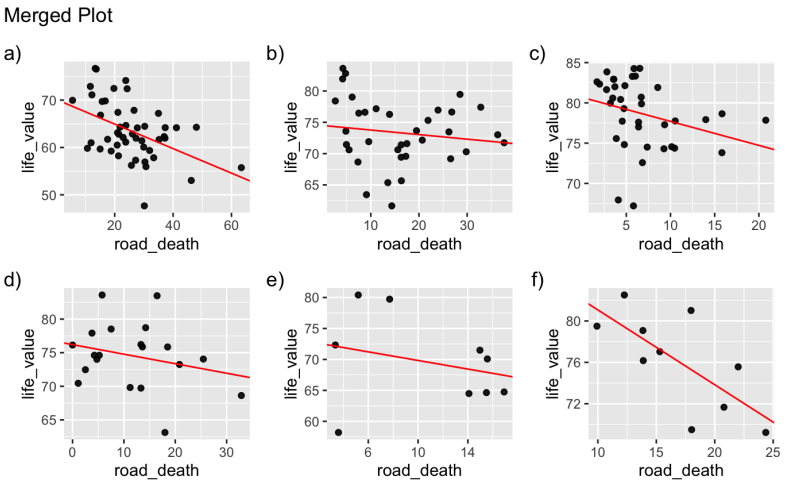

Figure 8. Life Expectancy and Road Death with regression analysis

*a) Africa, b) Asia, c) Europe, d) North America, e) Oceania, and f) South America

Based on Figure 7, Africa and South America exhibit a clear linear relationship between road deaths and life expectancy, while other continents show less pronounced correlations between the two variables.

6. Life Expectancy V.S. Birth By Skilled Personnels

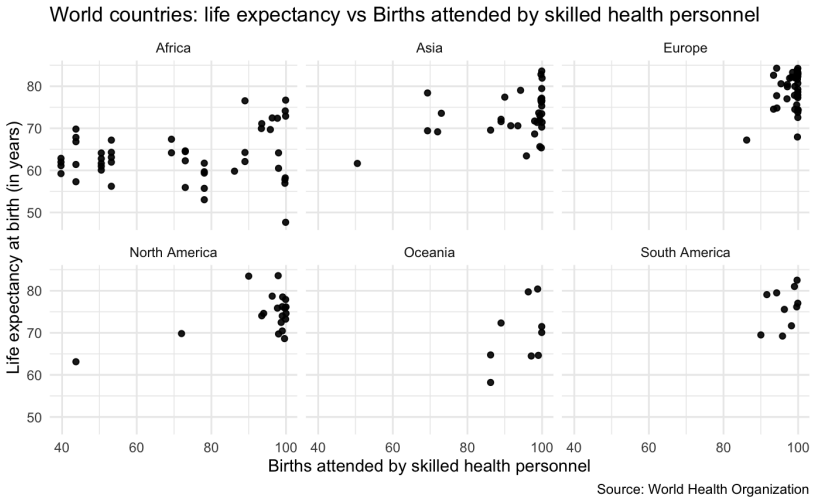

Figure 9. Life Expectancy and Birth by Skilled Personnel

In most continents, countries tend to have relatively high rates of births attended by skilled health personnel. However, many African countries exhibit low rates for this variable, indicating that a significant number of babies are born without skilled health assistance. Despite the low rates of skilled birth attendance in many African countries, the relationship

with life expectancy is not strongly correlated, as overall life expectancy remains relatively low irrespective of the availability of skilled health personnel for childbirth assistance in Africa.

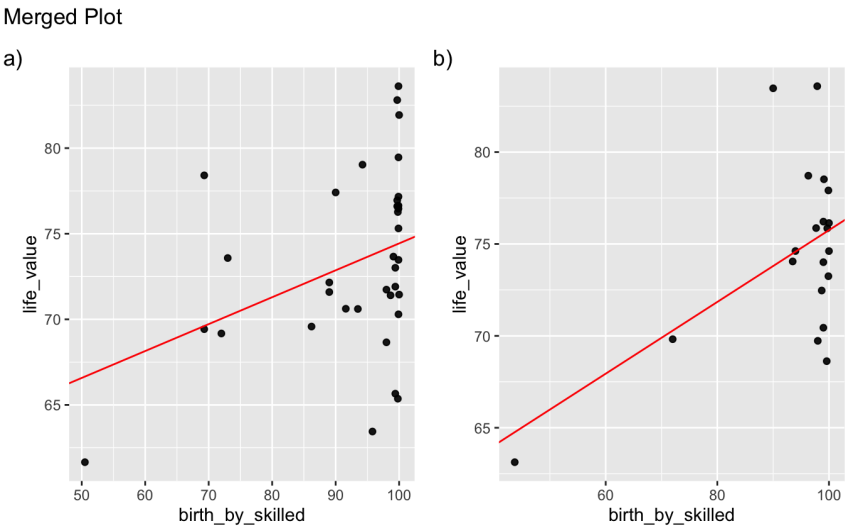

Figure 10. Life Expectancy and Birth by Skilled Health Personnel with regression analysis
*a) Asia and b) North America

Figure 9 illustrates a positive linear relationship between life expectancy and births attended by skilled health personnel in Asia and North America. This indicates that higher utilization of skilled health personnel during childbirth correlates with increased life expectancy in these regions.

7. Life Expectancy V.S. Maternal Mortality

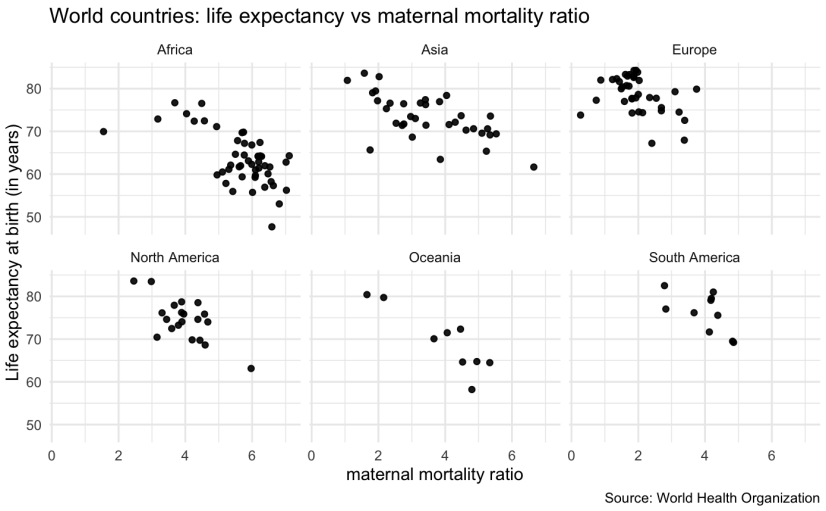

Figure 11. Life Expectancy and Maternal Mortality Ratio

In many African countries, a high maternal mortality rate is prevalent, indicating that a significant number of mothers die during childbirth. The relationship between life expectancy and maternal mortality rate exhibits a pattern wherein higher maternal mortality rates are associated with decreased life expectancy, indicating a negative relationship between the two variables.

Interpretation of PCA (Principal Component Analysis)

PCA (Principal Component Analysis) is a statistical technique to reduce the dimensionality of the dataset which is widely used in machine learning. From the mathematical perspective, PCA works by projecting the data onto reduced dimension and finding the vector which did not change its previous direction (Eigenvector) and its value (Eigenvalue).
The formula is:

$$A\mathbf{v} = \lambda\mathbf{v}$$

A is a squared matrix, **v** is an eigenvector, and λ is an eigenvalue.

In R, we can conduct the PCA with the code of **pca = prcomp(df_imput_n, scale = TRUE)**.

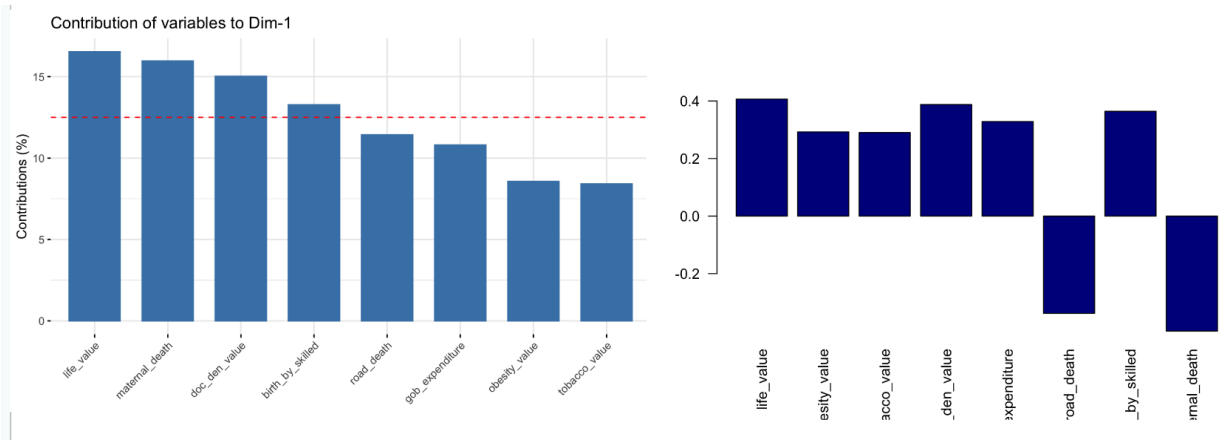

Figure 12. First principal component

After conducting the PCA, we can analyze first principal component by using code of **fviz_contrib(pca, choice = "var", axes = 1)** and **barplot(pca$rotation[,1], las=2, col="darkblue")**. Graph in figure 11 shows the first principal component indicating life

expectancy, maternal mortality rate, and doctor density are three the most important components.

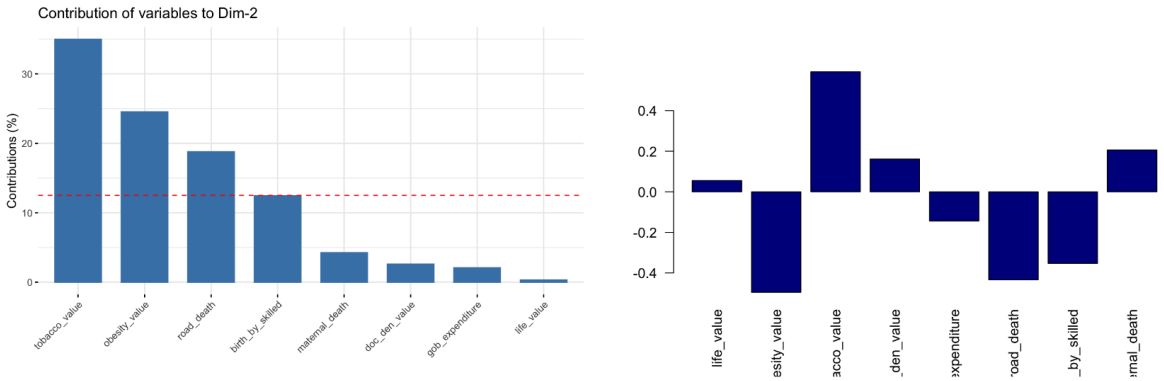

Figure 13. Second principal component

From the figure 12, we can get the second principal component using the same code for the first principal component. Figure 12 indicates that tobacco usage, obesity, and road death rate are three the most important variables since they have the highest absolute loading values among variables.

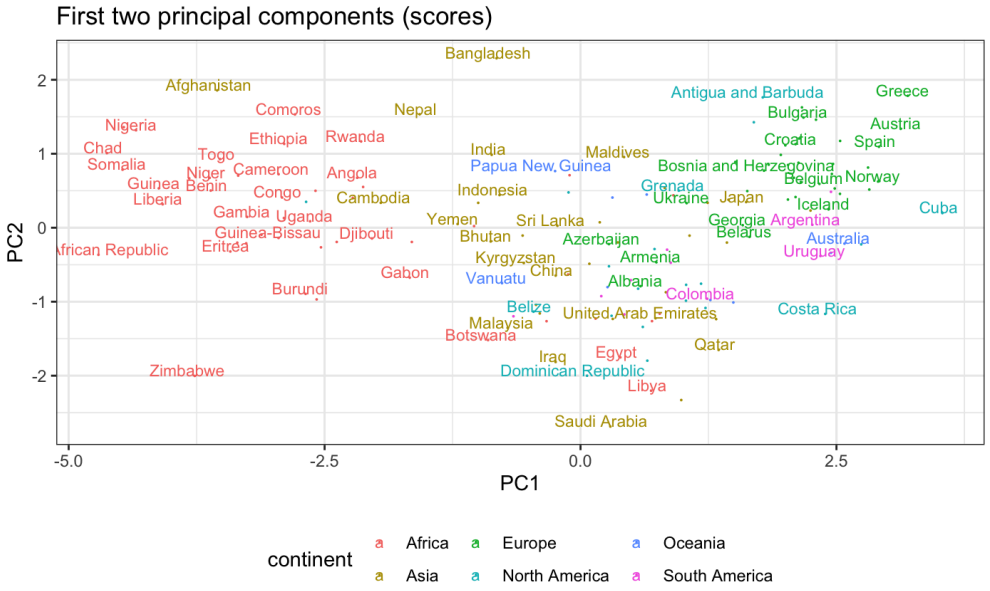

Figure 14. First two principal components

With the result from the first two principal components, we can get figure 13 which shows how countries are located in the 2 dimensional space. We can see that countries in the same

continent have similar values of variables. Upper right countries which are usually from Europe have high scores for both PC1 and PC2. For countries in Africa, they tend to have relatively neutral or high scores of PC2 but low scores of PC1.

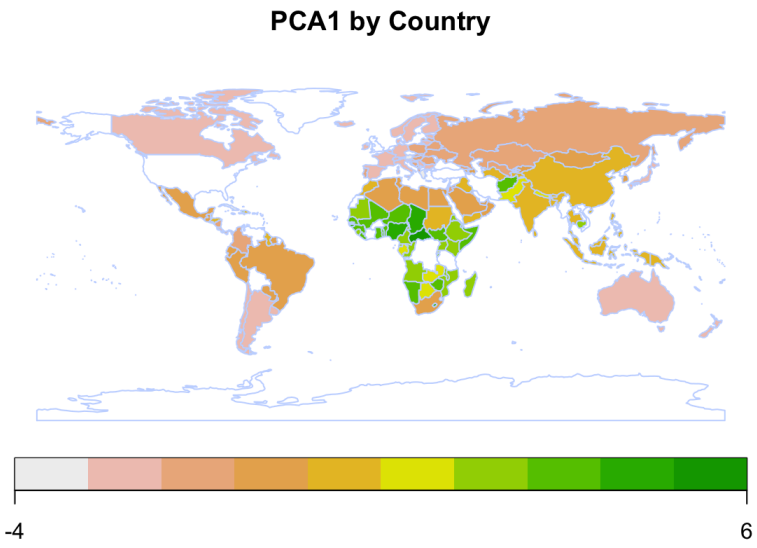

Figure 15. PCA1 by Country

This map visualizes overall PCA values by colors and the countries which have similar color specify that they have similar value of components.

Interpretation of K-means clustering

K-means clustering is an unsupervised machine learning method used for partitioning a dataset into “k” clusters based on the similarity of data points. The algorithm iteratively assigns each data point to the nearest cluster centroid and then recalculates the centroids based on the mean of the data points assigned to each cluster. This process continues until the centroids no longer change significantly, or a specified number of iterations is reached.

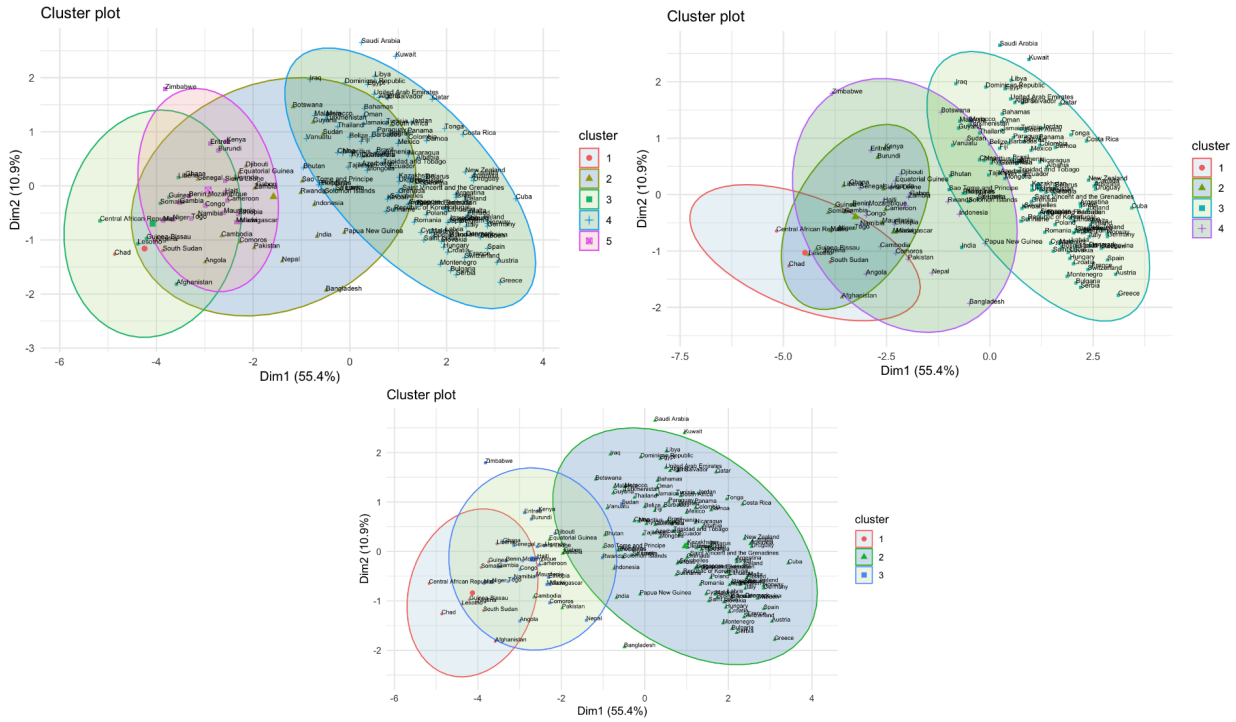

Figure 16. K-means clustering with 5,4 and 3 clusters

K-means clustering graphs visually display how countries are grouped into clusters based on their point locations. With 5 and 4 clusters, there is significant overlap, making it difficult to distinguish clusters. Thus, clustering with 3 clusters appears more suitable for effectively measuring the distinct clusters. In the third graph with 3 clusters, it shows countries in similar continents or have similar characteristics are located adjacently and clustered together. For example, the left cluster contains mainly African countries or less developed countries, the middle cluster contains mainly Asian countries or moderately developed countries, and lastly right cluster contains relatively developed countries from European, North America, some of Asia, and some of North America.

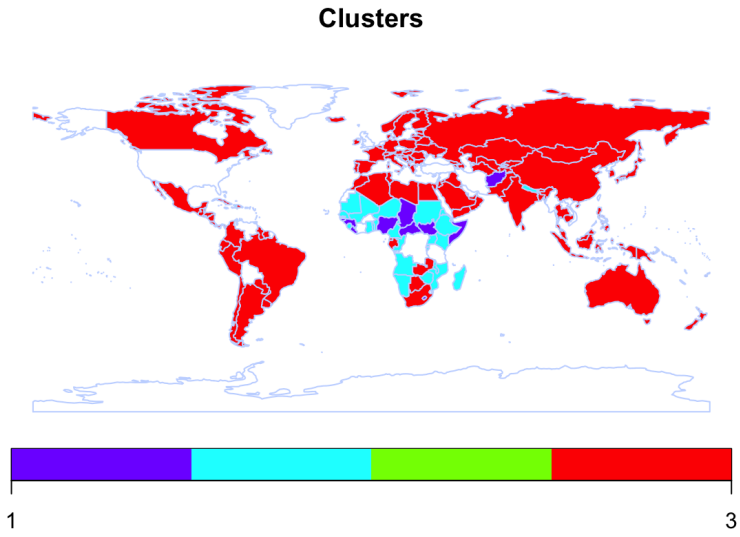

Figure 17. K-means clustering with 3 clusters

Figure 1.6 displays a map visualizing the clustering result with 3 clusters. It illustrates that countries in Africa and other continents typically exhibit distinct features.

Conclusion

Based on the objective of addressing population size decline by investigating the most dangerous components contributing to death rates worldwide, the PCA and K-means clustering analyses provided valuable insights into understanding these factors and clustering countries based on their socio-economic characteristics.

The PCA revealed that life expectancy, maternal mortality rate, and tobacco usage rate are the most important variables contributing to death rates. Countries with higher life expectancy tend to have lower impacts from other variables, suggesting the significance of life expectancy in mitigating death-related risks. Additionally, socio-economically similar countries exhibit similar values for these components, indicating shared patterns among countries with comparable socio-economic statuses.

The K-means clustering analysis identified three clusters as appropriate for the analysis, effectively grouping countries based on their socio-economic characteristics. The varying number of countries within each cluster suggests that the majority share similar features,

highlighting commonalities in death-related factors among countries with comparable socio-economic profiles.

These analyses offer valuable insights into understanding the reasons for death worldwide and how countries can be categorized based on these factors. By identifying factors that contribute to reducing death rates and observing clusters with similar characteristics, tailored strategies can be devised to address death-related issues specific to each cluster of countries. This approach enables the development of targeted interventions aimed at mitigating death-related risks and ultimately addressing population size decline more effectively.

Appendices

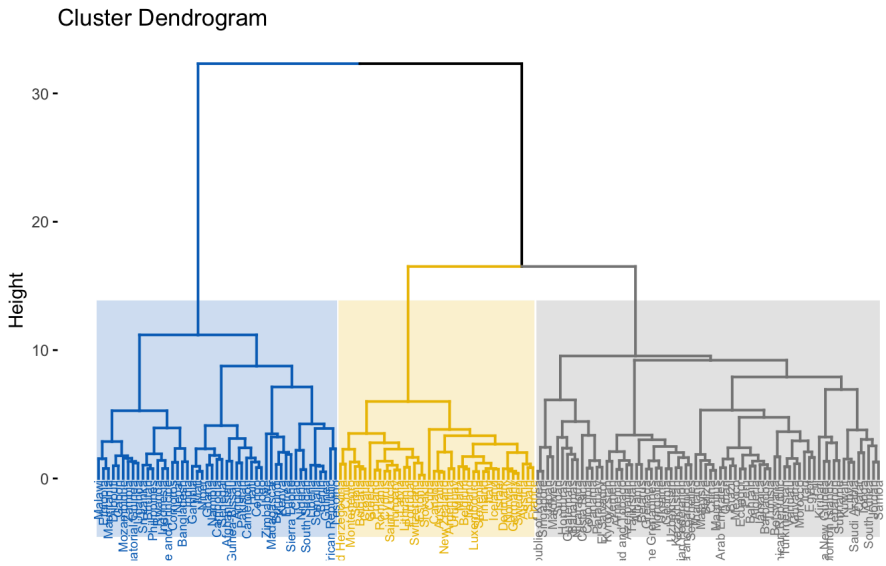

Figure 18. Cluster Dendrogram with 3 clusters

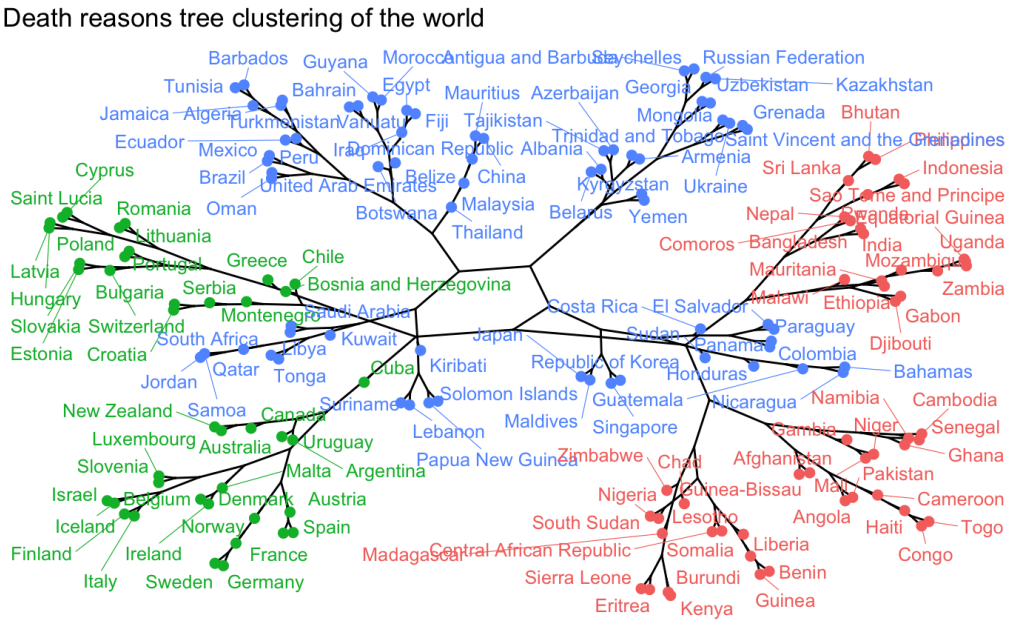

Figure 19. Death reasons tree clustering of the world with 3 clusters

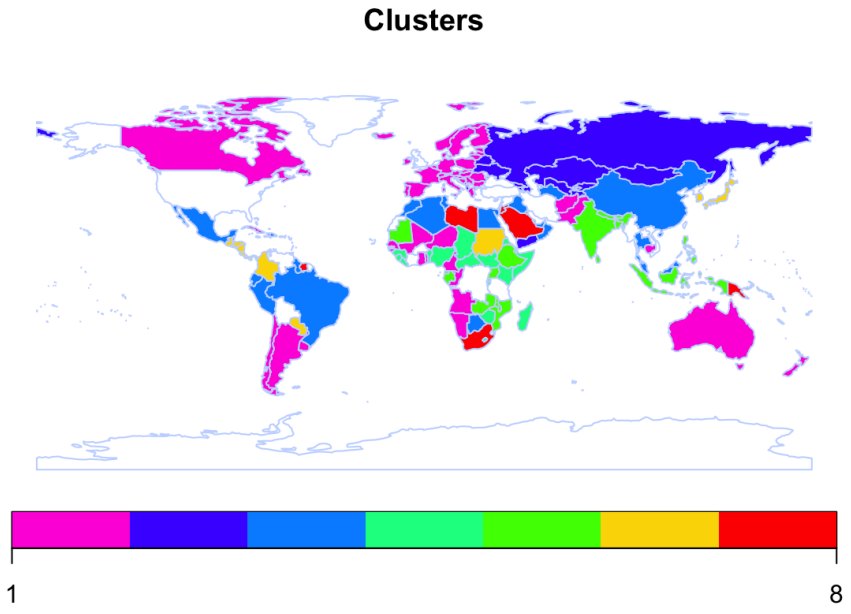

Figure 20. Map with 8 clustered countries

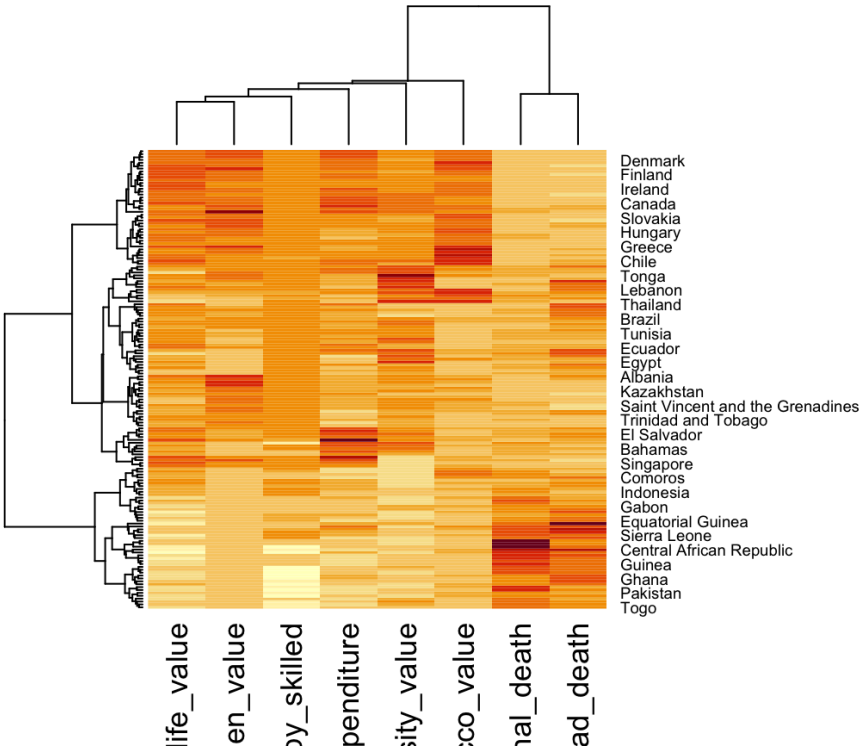

Figure 21. Heatmap

Reference

Tsai, S. P., Lee, E. S., & Hardy, R. J. (1978). The effect of a reduction in leading causes of death: potential gains in life expectancy. *American journal of public health*, *68*(10), 966-971.

Roth, G. A., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., ... & Borschmann, R. (2018). Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, *392*(10159), 1736-1788.

Hyun-Chool, L. E. E. (2021). Population aging and Korean society. *Korea Journal*, *61*(2), 5-20.