Narrative of Study on Global Life Expectancy - Group 6 (L2)

1) Introduction and Background

Since ancient times, humans have tried various ways to prolong their lives. With the continuous development of society, people began to use big data to find out the relationship between life expectancy and various factors. Numerous studies were conducted in the past on factors affecting life expectancy, but the effects of immunization and the human development index had not been considered. Also, some of them only covered datasets of one year for all countries, which gave us motivation to study a wider range of factors for a longer period. The dataset we found is more comprehensive, as it includes immunization factors, mortality factors, economic factors, social factors, and other health-related factors on life expectancy over 15 years. Performing visual analytics on such a dataset gives insights to effectively increase life expectancy of its population.

2) Objectives and Goals

Generally, from a national perspective, our goal is to find the factors that may have associations with people's life expectancy. It will help the government easier to detect which area further research should focus on and what kind of corresponding assistance or policy should be implemented. Furthermore, from the personal perspective, we want to exploit which type of lifestyle may have an association with life expectancy and the insightful conclusion will warn people to reduce their bad habits. Visualization based on correlation analysis and other statistical methods would help us to get more intuitive and convincing conclusions. Specifically, we raise the following questions.

- How does life expectancy evolve over time/within different regions?
- Do citizens from developed countries live longer than those from underdeveloped countries?
- Do residents from densely populated countries characterized by lower life expectancy?
- Do adults' and children's mortality rates greatly associated with the country's average citizen life span?
- To what extent do different kinds of immunization correlate with life expectancy?
- How is lifestyle (alcohol assumption, smoking habits) correlated with an individual's life span?
- Is education level associated with life expectancy?

3) Description of Datasets

We found the dataset on Kaggle (https://www.kaggle.com/kumarajarshi/life-expectancy-who), which is the world's largest data science community with powerful tools and resources. The original data was collected and posted by GHO (Global Health Observatory), which is a portal of WHO (World Health Organization), tracking and monitoring data related to global health situations. Nevertheless, we chose to adopt the dataset from the source mentioned above, but re-organized by Kumar Rajarshi, who integrated multiple variables into one dataset and excluded a few less known countries for missing values. As a result, there are 2938 observations in the dataset, spanning from 2000 to 2015 over 193 countries. Also, there are 22 variables in the dataset. 2 out of 22 are Categorical variables, which are about country and the associated development status. Thus, a choropleth/filled map is able to be utilized for the visualization. The remaining variables are the numerical variables, which are either discrete or continuous. The numerical variables (excluding life expectancy) can be classified into economical, mortality, immunization related or social attributes. For example, there is a variable of ***Total expenditure*** indicating the percentage of the amount the government spends on health over the total expenditure, which is a continuous variable and belongs to the economical attribute. However, we will only choose the most important variables of interest to perform our visual analytics.