Part 1: Introduction and Research Questions

We live in a globalized world, with increasing mobility and migration around the globe. According to UN migration, there are over 280 million people living in a country different from their place of origin, which amounts to 3.6% of the global population and is threefold the population in 1970[1]. Nations are also increasingly collaborating with one another through diplomatic agreements, free trade, climate consensus etc. as they strive towards Sustainable Development Goals (SDGs). Under such circumstances, it becomes pertinent for both individuals and states to assess health trends and the development of health systems in countries around the world. Countries with better health systems are generally more livable, with higher life expectancy and greater quality of life. These countries serve as home for happier citizens, more attractive destinations for foreigners, and exemplars for less-developed countries.

In this project, we hope to investigate the key factors that determine if a country is suitable for living and develop a general metric for evaluating how good a country is for living. We are also interested in trends in livability of countries over the years, particularly after the world entered the pandemic in December 2019. Therefore, our concrete research question is to construct a model to assess the livability of a country, based on the quality of its health systems and health trends over the years.

This is a substantial research question, because the livability of a country depends on a wide range of factors, from socioeconomic, political, to healthcare, which entails careful statistical evaluation. This project is feasible for three members to complete, because we are narrowing the scope to the public health aspect of a country's livability. Finally, it is a relevant project considering the surge in migration, the importance of global health especially during the Covid-19 pandemic, and the urgency for countries to achieve SDGs given the climate crisis.

Part 2: Data Sources

In this project, the dataset we will use is the World Health Statistics[2], which was provided by Zeus,

[1] UN Migration (2020). World Migration Report 2020. Retrieved from https://worldmigrationreport.iom.int/wmr-2020-interactive/#:~:text=The%20current%20global%20estimate%20is,over%20the%20past%20five%20decades.

[2] Zeus (2020). World Healthy Statistics. Retrieved from https://www.kaggle.com/utkarshxy/who-worldhealth-statistics-2020-complete?select=30-70cancerChdEtc.csv

and the World Happiness Report 2015-2021[3], which was provided by Mathurin Ache. These datasets were collected by Kaggle, which was one of the most famous data websites in the world. For the World Health Statistics, it enumerates information about the national life for various countries, such as Healthy life expectancy, Adolescent birth rate, and so on. We expect to use these variables to predict which country is more livable. Besides, for the World Happiness Report 2015-2021. It contains the happiness index for different countries from 2015-2021. To begin with, we want to use the data to predict the ranked happiness index in 2022. What's more, we can establish a linear regression between the ranked happiness index in 2022 and the variables in World Health Statistics.

In general, I hold the opinion that these datasets are suitable for our project. Firstly, these datasets contain substantial samples, which means the bias error due to insufficient samples will not appear in our project. Then, when we try to build the prediction model based on World Health Statistics, there are tons of variables that can be predictors, and I believe this can make our model more accurate. However, we need to make sure the predictors are independent, otherwise, the multicollinearity will lead to the error for the model.

Part 3: Collaboration Plan

1. Since two people in our group have withdrawn from this class, the project will be completed by three people. Each of us is responsible for the data processing and initial visualization of 13 datasets. We will then work together to explore the relationships between the individual data, select the appropriate parameters, construct the model, and complete testing of the model. In this part, we will decide responsibilities on a weekly basis.
2. We expect each member to spend on this project 3-5 hours per week.
3. We will meet from 9 pm to 10 pm on Sunday each week. We will share some individual results and clarify tasks for the next week during each meeting. As the project progresses, we may extend our meeting time and work together to move the project forward.
4. We will use WeChat and Zoom to communicate between meetings.
5. We will use Box to store our data, use Google Docs to write together, and use Google colab to share our code.

[3] Mathurin Ache (2021). World Happiness Report. Retrieved from https://www.kaggle.com/mathurinache/world-happiness-report-20152021