

NATIONAL UNIVERSITY OF SINGAPORE  
SCHOOL OF COMPUTING



Project Proposal  
Team 12

---

# CS5346 INFORMATION VISUALISATION

---

Student Number	Student Name
A0154690B	DEBJYOTI GHOSH
A0195628N	MORATH BASTIAN DOMINIK
A0154660H	MUHAMMAD HUZAIFAH BIN MD SHAHRIN
A0191374B	SMITHA SHESHADRI

# Many more happy returns! - Life expectancy, what affects it and how?

## Project Overview

One of the major indicators of the health condition of a population is its life expectancy. Mankind has been trying to stretch its lifespan and live *long and happy* lives from time immemorial. Generation after generation try and experiment with techniques to extend the human lifespan. Even in literature and pop culture, fantastical creatures with no death or with inhumanely long life spans are revered (Vampire and Thor come to mind!). While immortality continues to be confined to Sci-Fi, leading a long and healthy life is definitely within one's reach.

In order to understand the factors that contribute to one's lifespan, one needs to look for patterns and indicators of correlation in a population in addition to analyzing the human body biologically. This project attempts to introduce readers to several potential factors that may impact longevity. Our target audience is diverse and through our visualizations we aim to cater to this audience with different interpretations of the data: the class cohort and the general population (general awareness), specific target groups such as the policymakers (interaction effect between indicators of health and life expectancy) and the industry (effect of food and fitness consumerism on life expectancy).

We have decided to look into four measures: Body Mass Index(BMI), time spent on fitness, meat consumption patterns and healthcare. The reason we chose these measures is that we hypothesized that the major contributing factors to life expectancy would be the genetic factor (nature), the food and exercise factor (nurture) and the received medical support (external). We carefully analyzed these four contributors and chose the measures based on accurate measurability,

uniqueness, and availability of integrated datasets. Via these indicators, we will study global trends with a country-level granularity.

## Datasets

We will be using data from various sources and have CSV files of the following:

**Better:** <https://ourworldindata.org/life-expectancy>  
**Life expectancy**, which comes from the [United Nations Population Division](#). The data contains the life expectancy at birth, i.e. the number of years a newborn infant would live if prevailing patterns of mortality at the time of birth were to stay the same throughout life, from 1960 to 2016 for 237 countries and regions.

**BMI** comes from the [NCD Risk Factor Collaboration](#). We have data for 200 countries from 1975 to 2016, both for women and men.

**Meat production and consumption** comes from the [Food and Agriculture Organization of the United Nations](#). We have data from 1961 to 2014 for 234 countries and regions.

**Healthcare** comes from the [World Health Organization](#). We have data about the healthcare expenditure as a share of a country's GDP, the expenditure per capita and also the so-called 'out-of-pocket' expenditure, which refers to direct outlays made by households, including gratuities and in-kind payments, to healthcare providers. The data is available for 230 countries and regions.

**Fitness** comes from the [World Health Organization](#) and includes the prevalence of insufficient physical activity among adults in 161 countries.

**Smoking:** <https://ourworldindata.org/smoking>



## **TEAM 12**

### **Health related**

- What is the problem being addressed?
- What is the data source and format ? Where is the data coming from and what are its characteristics?
- Who would be interested in understanding this data better?
- What would these people want to know about the data?

Idea: Use the BMI, Fitness routines/sports access, meat consumption and healthcare data (e.g. the healthcare expenditure of GDP) to visualize and also predict/correlate with things such as the life expectancy

Data:

- **Life expectancy** (for each country, at birth, 1960-2015)  
<https://data.worldbank.org/indicator/SP.DYN.LE00.IN>
- **BMI**  
<https://www.who.int/nutrition/databases/en/>
- **Fitness/sports access**
- **Meat consumption + production**  
<https://ourworldindata.org/meat-and-seafood-production-consumption>
- **Healthcare** (Share Expenditure as share of GDP + Out-of-Pocket expenditure)  
<https://ourworldindata.org/health-meta>

## **Migration**

**Migration data portal** - very extensive website with map visualizations that pulls from plenty of data sources

<https://migrationdataportal.org/data?t=null>

some raw data available here:

<https://migrationdataportal.org/themes/iom-data-overview>

thematic overviews (e.g. type of migration):

<https://migrationdataportal.org/themes>

## **Migration policy institute**

<https://www.migrationpolicy.org/programs/migration-data-hub>

## **UN data**

<https://www.un.org/en/development/desa/population/migration/data/estimates2/estimates17.asp>

**UC San Diego** – has good links to datasets / studies

<https://ucsd.libguides.com/c.php?g=90870&p=4488628>

including e.g. World bank data on remittances

## **US Migration Flow Data released by the US Govt.**

<https://catalog.data.gov/dataset/migration-flow-data>

## **Can look at different aspects and themes:**

- What are the major driving forces of migration currently (compared to the past)
- Where are the major flows of people, why are they occurring, how are the people moving etc.
- Singapore case study? - how this impacts policy

e.g.

- Push vs pull (forced, labour, different factors e.g war, environment)
- Female vs male

- Adults vs children
- Where are the key migration nodes
- Migration patterns over time
- Stats around condition of migrants (migrants/population, integration, living condition, health, economic condition etc.)