CSCI 5388: Big Data Analytics

Term Project:

Good Health and Well Being

Human Mortality

Team:
Suhail Arfaath
Bhanusree Penukonda
Sri Hari Krishna Perla
Vishal Kumar

Instructor:
Yalong Wu

Table of Contents:

Introduction:

Overview or Background

Ever wondered why we are born and why we die. What is the factor which affects our mortality?

Is it diseases or is it the aging which reduces our mortality?

In the evolutionary journey of human beings from Ramapithecus > Australopithecus > Homo Erectus > Homo Sapiens > Modern Humans, there isn't any way to gauge what is the major reason for death in humans.

Scientist have been using statistical analysis to estimate the cause of death in the past. Which is a way to get observations and draw conclusions. Now in the modern era after the advent of Bigdata Analytics around 2010, evolved numerous ways to efficiently collect and store data and perform complex computation.

Problem Statement

Human Mortality data plays a huge role in finding the challenging insights within the data and to extract patterns. These patterns could help us in understanding the human evolution also analyze the dataset by using Bigdata technologies. This dataset signifies the mortality rate of different age groups. This data set has data right from 1950 and it's a collection of data from past decades that contains the mortality rate of different people from different countries and from different age groups.

The age groups are bucketed based on different ages, age 1 to age 5 fall in bucket 1 or group in a bin.

This data has been collected from the National Registry from each country

What we are trying to solve

We are trying to solve this problem by identifying the patterns within the data and try to extract the pattern and cause of problems.

Why we are solving this

This may give us valuable insights through which we can overcome the problem and better help the humanity.

How the data is collected

The data available for download from this web site are official national statistics in the sense that they have been transmitted to the World Health Organization by the competent authorities of the countries concerned.

The WHO Mortality Database comprises deaths registered in national vital registration systems, with underlying cause of death as coded by the relevant national authority. Underlying cause of death is defined as “the disease or injury which initiated the train of morbid events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury” in accordance with the rules of the International Classification of Diseases.

The database contains number of deaths by country, year, sex, age group and cause of death as far back from 1950. Data are included only for countries reporting data properly coded according to the International Classification of Diseases (ICD).

How we are going to solve

We're going to cut the data from different sources and then try to preprocess the data and then try to visualize the data and find the patterns within it.

Research Goal and Objectives

We can try to extract the knowledge from this data in terms of what kind of diseases or pandemics or epidemics occur and how the mortality of populations from different countries.
This includes the cause of death for each age group. With this data we can try to analyze what are the major trends in terms of panic or epidemics and what's the reason why there is a high death toll or mortality for the population.

What knowledge patterns we are observing

in terms of knowledge if we try to understand like what's the main reason and what kind of symptoms or diseases occurring at different countries and if we came to know like what the most common factor is that is affecting the mortality rate of each population.

What insights do we gain?

With these insights we can try to better help each country in terms of specific disease aid such as in the case of Africa, if we see we have malaria disease which one is the most prominent, and we have high death toll because of the same.

What we give to society back?

If we were able to address this problem, we could solve the human crisis between different countries in “Third World Countries”.

Are we doing any savings?

In terms of savings, we can try to invent better medicines and efficient vaccines and we can also try to deliver the vaccines and medicines to the target location which may aid in solving the problem

In Terms of monetary $ value

We could try to better address the problem by increasing the supply chain management of delivering the vaccine or medicine to the location where the disease is more prominent especially in terms of covid-nineteen where a lot of people were unable to find the vaccines or the medications due to which a high death toll was recorded.

What is the trend in data?

The trend in data will be dependent on the death rate at that time and the causes of death in that particular year. It gives the top 10 global causes of death.
And, health disaggregated by sex, annual global deaths and DALY’s among women were around 15% lower than for men.

how mortality rate differs based on

A mortality rate is a measurement of how frequently people die within a given population during a certain period. The choice of measuring disease or death depends on whether you want to use morbidity or mortality metrics, which are frequently mathematically equivalent.

Literature Review

Our research journey began with an exploration of literature on conditions for well-being in Health and mortality rate.
Some literature already done on this topic:

Yang Yang

Yang (2008) analyzes the impact of changes in cause of death overtime on age, period, and cohort effects in the United States during the 20th century. Table 8 below provides an overview of her findings by cause of death.

Cause of Death	Age	Period	Cohort
Heart disease	Increases exponentially with age	Modest impact	Large monotonic decline from the earliest to the latest cohort
Stroke	Increases exponentially with age	Modest impact	Large monotonic decline from the earliest to the latest cohort
Lung cancer	Increases rapidly with age from early adulthood to peak near ages 80–85, then levels off	Monotonic increase over time	Increases for cohorts through 1905 and decreases for recent cohorts
Breast cancer	Increases with age, but increases slow around menopause	Modest impact	Steady declines in mortality from breast cancer from the earliest to the latest cohort

Yang discovered that significant declines in mortality that began in the late 1960s persisted far into the late 1990s and were primarily related to cohort effects. Although cohort effects vary depending on the precise cause of death, overall survival rates have significantly increased.

Review of the Literature and Evaluation of Mortality Improvement 58 Rates in the US Population Her data offer further proof of enduring cohort disparities in mortality rates across all investigated causes of death. The predominance of cohort effects in explaining current trends in mortality

decreases is a significant conclusion. Birth cohort effects reflect the processes of uneven cohort accumulation of lifelong exposure to risk variables that include education, diet and nutrition, physical activity, and smoking.
Yang discovered that when birth cohort and age effects are jointly controlled, period effects are typically negligible or low. She discovered a very slight decline in stroke mortality since 1975 and almost no period effect for heart disease mortality. Period effects are probably more pronounced at times of war or other significant events with significant social repercussions (Yang 2008).

Research Methodology Contribution to the field/ Significance and / or Impact of Proposed Research

People are more likely to die young or have poor health outcomes when certain behaviors, exposures, and predispositions are present. The seminars were designed to help participants better understand the risk factors, primarily behavioral risk factors, that are most receptive to preventative and health policy measures. There are various ways to define "early." Although his latter study has centered on fatalities before age 80, Michael McGinnis' earlier work concentrated on deaths before age 75. The Global Burden of Disease report from the World Health Organization compares years of life lost to a reference age of 86, which is the maximum average longevity for a nation with a population of over 5 million people.

The methodology used are

- Maternal mortality ratio (per 100 000 live births)
- Age factor as a mortality rate
- Medical health workers working per capital ratio.
- Infectious, parasitic, and respiratory diseases
- Cancer as a health Issue
- Diseases of the circulatory system
- All other causes of the digestive system.

Form the data available finding the factors/above methodologies we can find the mortality rates how it varies based on the different health conditions.

Research Plan and Outline:

- Firstly, on the data main objective is to find the mortality rate for the health and wellbeing.
- WHO have provided the major factors affecting the mortality rate.
- Each factor percentage and influence for the occurrence of death.
- In details study of few health conditions affecting the cause of death.
- Age factor and the mortality rate as per death.
- Any possibilities or suggestions for the improve of the above factors.

References

- Mortality Committee. 2005. "Projecting Future Mortality: Towards a Proposal for a Stochastic Methodology." CMI Working Paper 15.

- https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section3.html

- Data source and other information:
- https://www.who.int/data/data-collection-tools/who-mortality-database

- Andreev, Kirill F., and James W. Vaupel. 2005. "Patterns of Mortality Improvement Over Age and Time in Developed Countries: Estimation, Presentation and Implications for Mortality Forecasting." Paper presented at the Population Association of America Annual Meeting, Philadelphia. http://paa2005.princeton.edu/abstracts/51061.