Title: World Health Organization Mortality Investigation

Team Members: Group 3

Bryan Arment

Jeff Dank

Tony Pearo

Description

We found a database published by the WHO that tracks all cause mortality by various factors extensive classifications for cause, demographics and regions. We hope to mine this data to see what patterns we can discover related to how mortality presents itself. Are certain diseases more prevalent in certain populations? Can having one disease predict another disease? These types of questions have real world impact and we will have to ability to sanity check our results as we go. For example we would expect heart disease to be occurring more often in older people and not children. In addition, we have geographic data available so it would be interesting to see if we can map mortality in this manner to see if interesting, albeit, morbid patterns emerge from the data set.

Prior Work

Lots of work has been done revolving around mortality statistics and other associated health studies. We likely won't be breaking entirely new ground with this project, but it is an interesting and expansive data set that will let us apply what we have been learning in class. A couple of research papers are listed below:

Knowledge and Attitude towards Cervical Cancer: A Case Study of Undergraduate Students in Imo State, Nigeria JOURNAL NAME: **Open Access Library Journal**, Vol.7 No.6, June 23, 2020

Mortality From Ischemic Heart Disease; Analysis of Data From the World Health Organization and Coronary Artery Disease Risk Factors From NCD Risk Factor Collaboration Circ Cardiovasc Qual Outcomes. 2019 Jun; 12(6): e005375.

Datasets

World Health Organization Mortality Database https://www.who.int/healthinfo/statistics/mortality_rawdata/en/
The dataset has been downloaded on Bryan's machine and we will start by working with the US data as a prototype for cleaning and initial processing.

Proposed Work

First, take a look at the entire dataset to get an idea of if anything needs to be cleaned or tossed out. There may be some integration / consolidation that needs to be done, for instance generating our own key-value pairs that correspond to the various types of mortality. Currently, there are several various keys due to where the mortalities were reported from. Load the dataset into python / pandas and explore any correlations we can discover to answer some of our interesting questions. Parse down the data and export the interesting pieces into smaller files. Load the smaller files into a tool like Tableau to be able to visually tell a story of our findings.

List of Tools

Tableau, Excel, Python / Pandas

Evaluation

We will begin with simple analysis of all factors mortality rates across a few countries before trying to ask more difficult questions related to underlying patterns. Try to make a predictions based on factors identified in the project description. For specific mortality factors we can study relevant research on the area as our dataset is widely available and accessed within the medical community.