Causes of Death

Group AKAA

Alaa Almouradi 00030123

Filzah Azeem 00030138

Fatma Khalil 00029999

M Safwan Yasin 00030037

Saleh Alshurafa 00030002
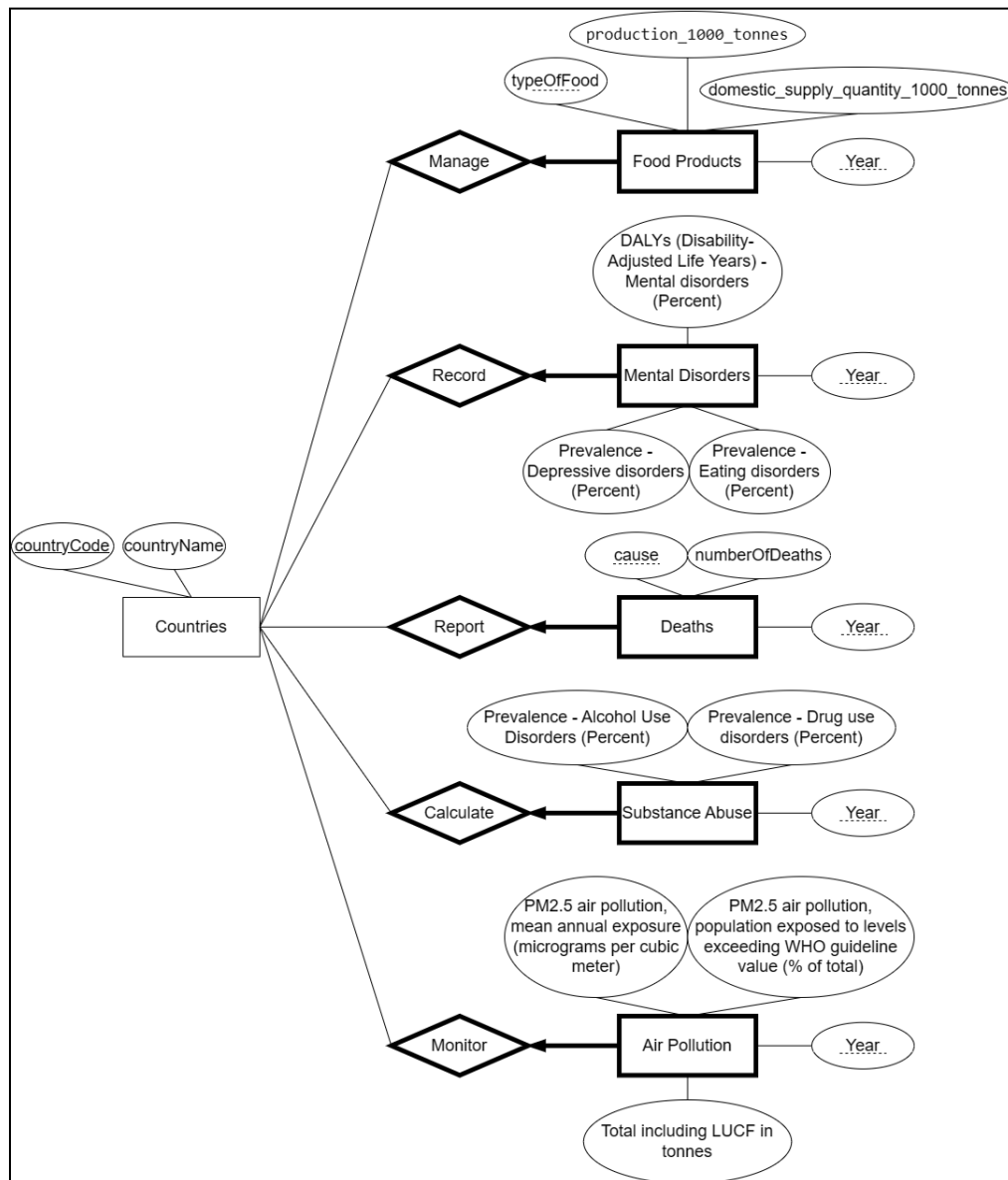
CS306 Spring 22-23

GitHub Repository

Project Description:

In this project our database application will store data on causes of deaths per country throughout time, along with data on other topics that can be related to specific causes of death. With that data, some insight can be discovered about specific causes of death in general or in specific countries. Food data can be related to deaths in famine. Mental disorder data can be related to deaths by direct mental disorders or results of them. Substance Abuse data can be related to violence or other deaths by specific illnesses. Pollution data can be related to deaths by specific illnesses. Our aim is to help countries, responsible authorities, and organizations in identifying the leading causes of deaths around the world so that they can take necessary steps to reduce the numbers of deaths as much as they can.

ER Model:

production_1000_tonnes

typeOfFood

domestic_supply_quantity_1000_tonnes

Manage — Food Products — Year

DALYs (Disability-Adjusted Life Years) - Mental disorders (Percent)

Record — Mental Disorders — Year

Prevalence - Depressive disorders (Percent)   Prevalence - Eating disorders (Percent)

cause   numberOfDeaths

countryCode   countryName

Countries   Report — Deaths — Year

Prevalence - Alcohol Use Disorders (Percent)   Prevalence - Drug use disorders (Percent)

Calculate — Substance Abuse — Year

PM2.5 air pollution, mean annual exposure (micrograms per cubic meter)   PM2.5 air pollution, population exposed to levels exceeding WHO guideline value (% of total)

Monitor — Air Pollution — Year

Total including LUCF in tonnes

ER Model Explanation:

Countries entity set is a strong entity set and has its primary key as countryCode. The rest of the sets are weak entity sets with complete participation and one to many (one Country can have many entries from each set) constraints in the identifying relationship sets while the countries set has partial participation. The weak entities have their keys as years along with the

primary key of the related owner entity. The food and death sets have one more attribute in the key as they can have multiple entries in the same year per country.

Our CSV Files:

Please note that you can access all of our CSV files by taking a look at our [GitHub](#) Repository.

1. **CountriesTable.csv:**

   Standard ISO3 country paired with its code. Minimal operations on the set were required.

2. **Alaa Almouradi Table STEP1(Food).csv**

   The set had Multiple unrelated columns that had to be removed. No duplicates were present. The organization of the table was undesired as some columns had limited values that would have been better as columns of their own. That undesired organization was fixed by creating columns with the desired names and the related data was transferred.

3. **M Safwan Yasin Table STEP1(Mental).csv**

   My data was sourced from two tables on the website. One of them contained information about DALYs (disability-adjusted life year) which represents the sum of years of life lost due to mental disorders. The second dataset was about the prevalence of mental disorders in each country. Firstly, I linked both datasets together and combined them into a single dataset. Then, I picked two of the most prominent mental disorders - eating disorders and depressive disorders - from the dataset and removed the others.

Furthermore, there was a significant number of countries that had missing country codes. I added them to make sure that there weren't any empty records in the dataset.

4. **Fatma Khalil Table STEP1(Death).csv**

    After downloading the data obtained from the provided website, I correctly sorted the years in chronological order because some countries had the years placed in a random order. Then I removed irrelevant data. Lastly, I fixed the appearance of the data in order to fit the attributes in the ER diagram.

5. **Filzah Azeem Table STEP1(Substance).csv**

    Initially, the dataset contained several irrelevant entries: it had the year spanning all the way from 10000 BC to 2021 whereas the data was only recorded for years 1990 to 2019. After I detected no duplicates, I removed all the unnecessary rows and adjusted the names of the columns.

6. **Saleh Alshurafa Table STEP1(Air Pollution).csv:**

    For my CSV file, I used the data from three tables found in ourworldindata.org. One table contained data about greenhouse gasses emissions, another about exposure to air pollution, and the final about the percentage of population who are exposed to pollution exceeding WHO standards. The exposure to pollution and the percentage of population tables had the same countries and years so no changes were necessary as they could easily be combined by copying and pasting columns. However, the greenhouse gasses emissions table had more years, had some countries that were not in the other 2 tables, and lacked some countries from the other 2 tables. Therefore, the rows that had countries not mentioned in all 3 tables were removed from all tables and likewise for the years that are not covered in all 3 tables. Then, I was finally able to copy and paste the

necessary data. Finally, I used the Remove Duplicates button to remove all the duplicate rows to clean the data.

Conclusion:

Data Was Collected and organized in order to have insight into causes of death and point out relations for the concerned authorities to take action.