Title: Global Analysis of Mental Health and Suicide Rates
Team: Upamanyu Mondal, Balakumar Rahul, Geetaharichandana Vemuri
Project Overview

This project conducted an in-depth examination of global mental health infrastructure data and suicide statistics across countries, age groups and genders to uncover insightful patterns, trends and solutions. A structured MySQL database was designed using a normalized entity-relationship diagram for efficient storage and analysis. Rigorous data analysis in Python combined with compelling visualizations revealed several key insights regarding correlates of suicide rates and high-risk demographic groups. The evidence gathered allows us to provide targeted, data-driven recommendations to key stakeholders for future suicide prevention and mental health promotion programs.

My role in this project involved designing the database and data analysis in SQL.

1. Detail your ERD design, including an explanation of each table.

   In the ERD, each table will be represented as an entity, and the relationships between them will be primarily through the Country_ID field, indicating that data in the Facilities, Crude_suicide_rates, Human_Resources, and Age_standardized_suicide_rates tables is all related to the countries defined in the Countries table. This is a typical one-to-many relationship, where one country can have multiple entries in the other tables for different years or categories.

   Countries Table

   Purpose: This table stores information about different countries.
   Fields:
   Country_ID: Unique identifier for each country. Primary Key.
   Country: Name of the country.
   Relationships: Country_ID is a foreign key in other tables, establishing a link to this table.

   Facilities Table

   Purpose: Contains data about various health facilities in each country.
   Fields:
   Country_ID: References Country_ID in the Countries table. Foreign Key.
   Year: The year the data pertains to.
   Mental_Hospitals: Number or capacity of mental hospitals.
   Health_Units: Number or capacity of health units.
   Outpatient_Facilities: Number or capacity of outpatient facilities.
   Day_Treatment: Number or capacity of day treatment facilities.
   Residential_Facilities: Number or capacity of residential facilities.

Primary Key: Composite of (Country_ID, Year).
Relationships: Relates to the Countries table via Country_ID.

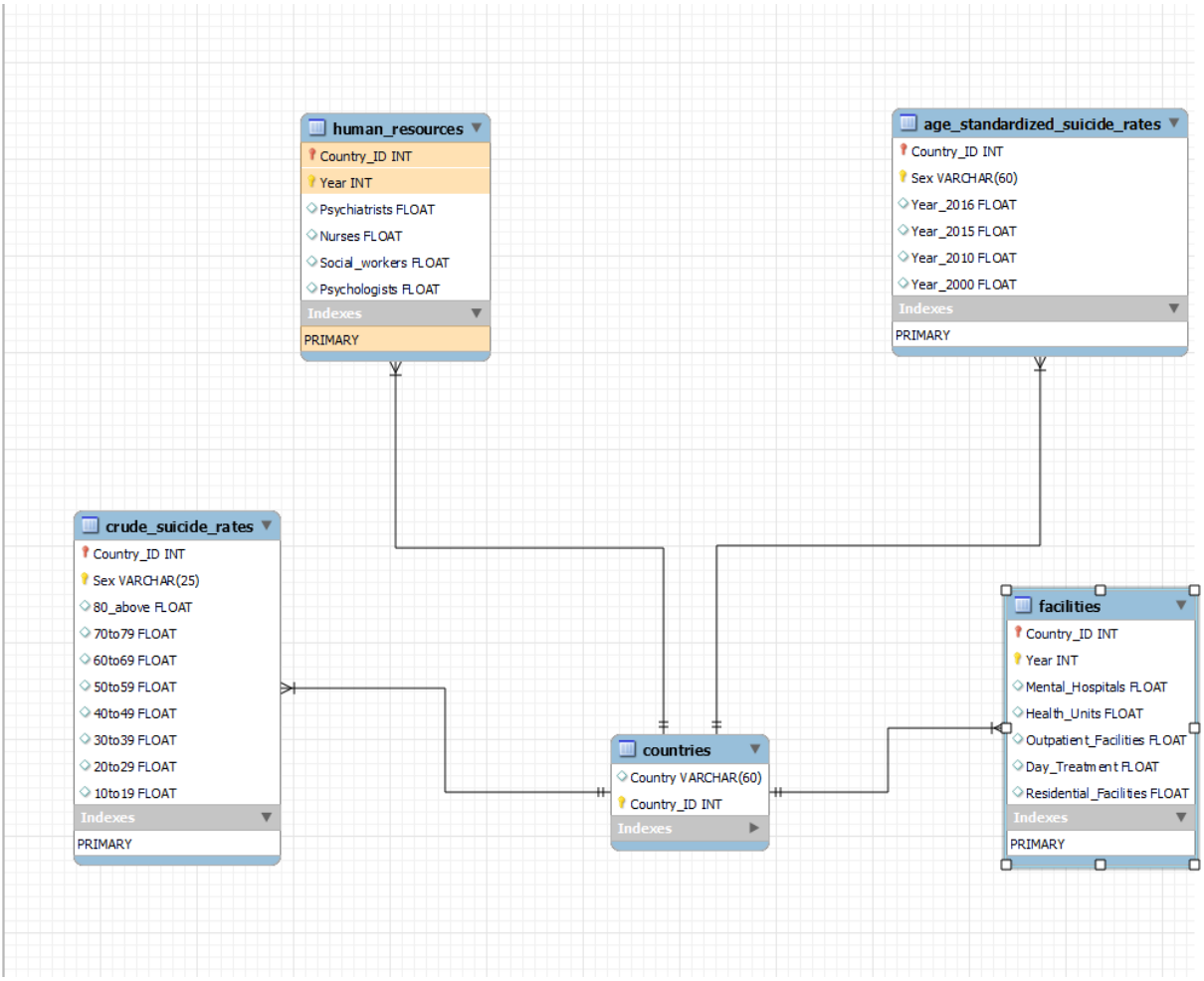
Crude_suicide_rates Table
Purpose: Holds crude suicide rate data for the year 2016 categorized by age groups for each country.
Fields:
Country_ID: References Country_ID in the Countries table. Foreign Key.
Sex: Gender classification.
Age group columns (80_above, 70to79, 60to69, etc.): Suicide rates for different age groups.
Primary Key: Composite of (Country_ID, Sex).
Relationships: Linked to the Countries table through Country_ID.

Human_Resources Table

Purpose: Stores data on medical human resources related to mental health in each country.
Fields:
Country_ID: References Country_ID in the Countries table. Foreign Key.
Year: The year the data pertains to.
Psychiatrists, Nurses, Social_Workers, Psychologists: Counts or ratios of these professionals in the country.
Primary Key: Composite of (Country_ID, Year).
Relationships: Associated with the Countries table via Country_ID.

Age_standardized_suicide_rates Table

Purpose: Contains age-standardized suicide rates for different years in each country.
Fields:
Country_ID: References Country_ID in the Countries table. Foreign Key.
Sex: Gender classification.
Year_2016, Year_2015, Year_2010, Year_2000: Suicide rates for respective years.
Primary Key: Composite of (Country_ID, Sex).
Relationships: Connected to the Countries table through Country_ID.



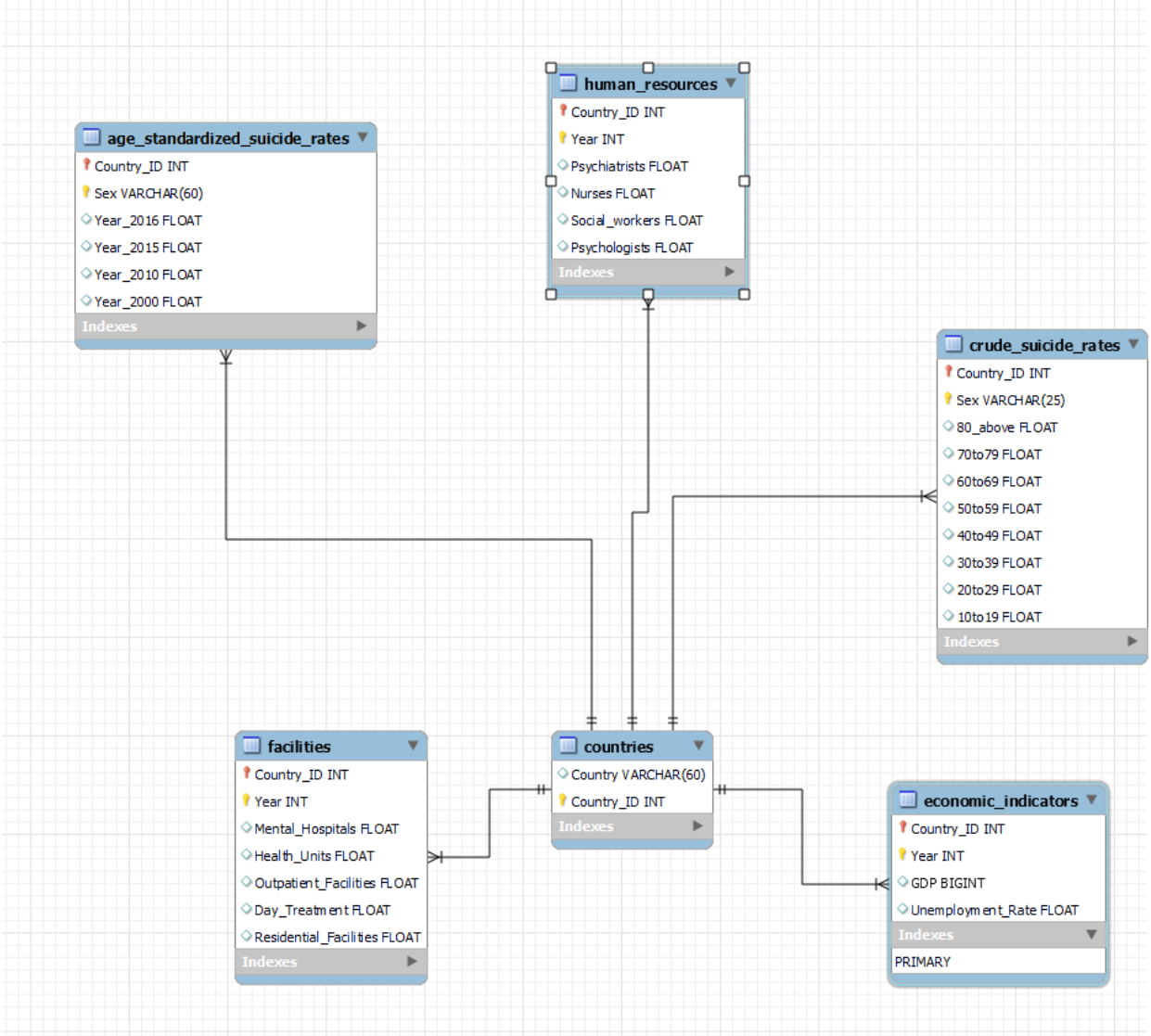
Elective Questions

1) How would the addition of a new table affect your current ERD? Include a revised ERD sketch.

   As we have said during the presentation that economic indicators can be an important part in our analysis. Since we are dealing with suicide rates the governments report of unemployment rate can be an important parameter. Similarly, the GDP of countries can be important as it can give an idea about how the countries are spending in the area of mental health i.e. on the facilities and human resources.

   Though we could not do it for this time, I think our database design is flexible and will allow us to add an economic indicator table containing the GDP and unemployment rate data.

   Updated ERD



2) Describe how your database design allows for the addition of new attributes. How would you implement these changes using SQL?

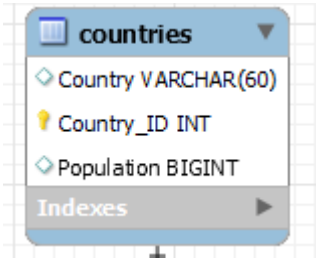
   The flexibility to accommodate new attributes is a key aspect of a well-designed database. The current database design, structured around normalized tables with clearly defined relationships, is conducive to such adaptability.

   Some of the changes I can think of are:

a) Adding New Columns to Existing Tables

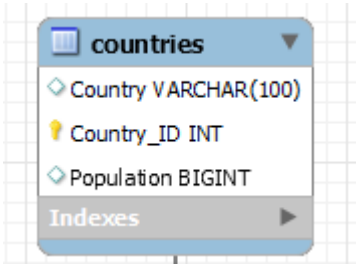
For example, adding population to the countries table can be taken into consideration since all the rates, be it facilities or human resources are for every 100,000 population. We can do it by using the below code.

```
ALTER TABLE Countries
ADD Population BIGINT;
```



b) Modifying Existing Columns

If the nature of an existing attribute changes (like changing the data type or size), we can use the ALTER TABLE command. For example, we can extend the size of the Country field in the "Countries" table:



3) Evaluate your database's normalization. To which normal form is it normalized? If not normalized, propose improvements.

Countries Table

In the "Countries" table, Country_ID is the primary key and the only determinant. The only other field i.e. Country depend on Country_ID. There are no non-trivial functional dependencies of attributes on anything other than this primary key. Therefore, the "Countries" table meets the BCNF criteria.

Facilities and Human Resources Tables

In both Facilities and Human Resources table, the composite key (Country_ID, Year) is the determinant for all other fields, which are data points about various health facilities and mental health professionals respectively. There are no non-trivial functional dependencies of attributes on anything other than this composite key. Therefore, this table appears to meet the BCNF criteria.

Crude suicide rates and Age-standardized suicide rates Tables

In both the Crude suicide rates and Age-standardized suicide rates tables, the composite key (Country_ID, Sex) is the determinant for all other fields, which are the different age-specific suicide rates in 2016 and the age standardized suicide rate for the year 2016, 2015, 2010 and 2000 respectively. There are no non-trivial functional dependencies of attributes on anything other than this composite key. Therefore, this table appears to meet the criteria for BCNF.

4) Discuss the selection of primary keys in your design, including the use of composite keys (if any).

The selection of primary keys in a database design is crucial for ensuring data integrity, optimizing performance, and facilitating easy retrieval of information.

Countries Table

This is a single-column primary key. The Country_ID is chosen because it uniquely identifies each country. This is a straightforward choice as it ensures each country can be distinctly identified without ambiguity.

Facilities and Human Resources Tables

A composite key is used in both these two tables because the combination of Country_ID and Year uniquely identifies each record. Country_ID alone would not suffice, as there could be multiple records for the same country across different years. The composite key ensures each row is unique and specific to a country and year.

Crude suicide rates and Age-standardized suicide rates Tables

In both these two table requires a composite key because the data is specific to both a country and a gender classification. There might be separate records for the same country, differentiated by the Sex column (e.g., male, female). Therefore, both Country_ID and Sex together ensure the uniqueness of each record.

5) How have you implemented referential integrity? Provide a specific example from your project.

Referential integrity is a key concept in relational databases, ensuring that relationships between tables remain consistent. As shown in the first question while creating the tables all the sub tables referenced the countries table using the Country_ID.

For example

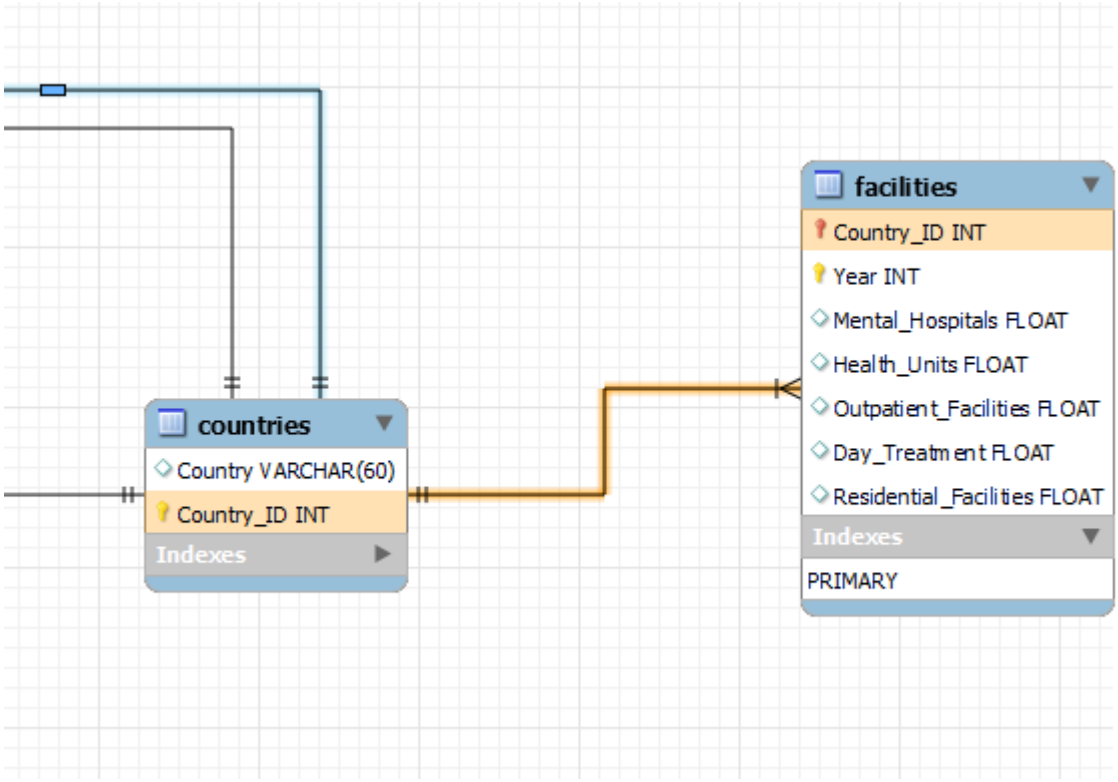
Countries Table

This table includes a column Country_ID, which serves as the primary key. Each Country_ID represents a unique country.

Facilities Table

Apart from the various data about health facilities, the Facilities table also contains Country_ID column. The Country_ID in the "Facilities" table is a foreign key that references the Country_ID in the "Countries" table. This prevents the insertion of a record in the "Facilities" table that references a Country_ID that doesn't exist in the "Countries" table.

```
CREATE TABLE IF NOT EXISTS Facilities (
   Country_ID INT,
   Year INT,
   Mental_Hospitals FLOAT,
   Health_Units FLOAT,
   Outpatient_Facilities FLOAT,
   Day_Treatment FLOAT,
   Residential_Facilities FLOAT,
   FOREIGN KEY (Country_ID) REFERENCES Countries(Country_ID),
   PRIMARY KEY (Country_ID, Year)
);
```


facilities
Country_ID INT
Year INT
Mental_Hospitals FLOAT
Health_Units FLOAT
Outpatient_Facilities FLOAT
Day_Treatment FLOAT
Residential_Facilities FLOAT
Indexes
PRIMARY
countries
Country VARCHAR(60)
Country_ID INT
Indexes