Tural Novruzov 28632

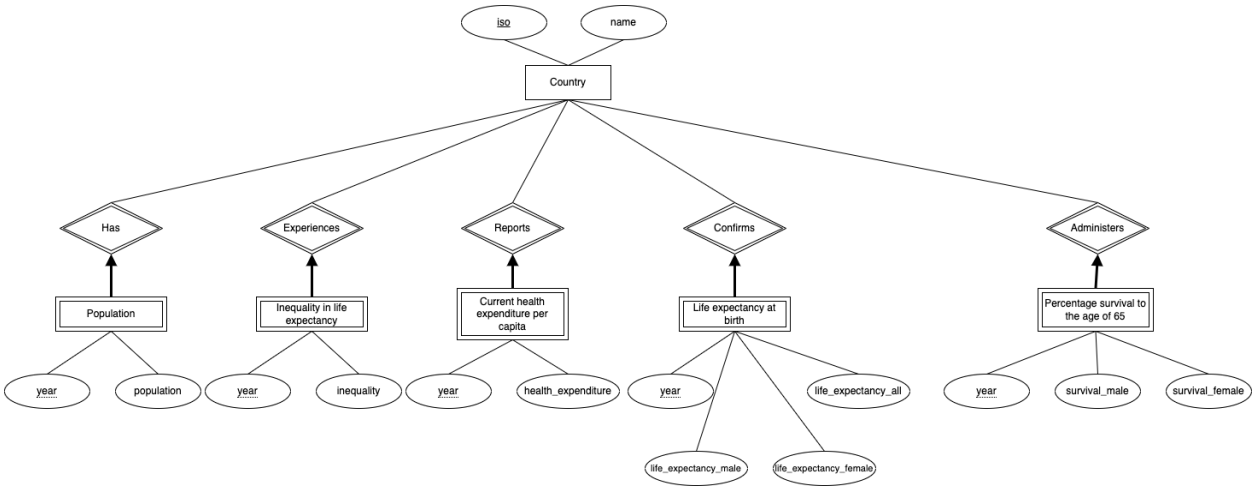
# Project Step 2

## Introduction

My project is titled "Life Expectancy and Inequality" where I will develop a database application to analyze inequality in life expectancy between countries and genders. A secondary objective of the application is to explore the relationship between life expectancy and population in different countries. The GitHub repository for the project: https://github.com/turalnovruzov/cs306-project. **I will have every step in different branches for easier browsing and grading experience. Please refer to the branch "step2" to see my project's state when I finished Step 2.**

In folder "Step 2" you can see the log file as "logs.log" and the SQL file as "create_tables.sql".

## Changes to the ER Diagram




In the second step of the project, modifications were made to the ER diagram. Initially, I attempted to combine the life expectancy CSV files into a single entity named "Life expectancy at birth". However, during the process of importing the data into MySQL, I observed that the number of data points in the "life-expectancy-of-women-vs-life-expectancy-of-women.csv" file is fewer than in the "life-expectancy.csv" file. Consequently, I decided to store these files in separate entities, namely "Life expectancy all" and "Life expectancy gender".

# Changes to the Data

Upon attempting to import data from the "population.csv" file into MySQL, I discovered that empty rows were missed during the cleaning process in step 1. Therefore, all rows with an empty "Code" column in the "population.csv" file were deleted to ensure data accuracy.

# SQL

The "Country" entity was transformed into the "country" table. Additionally, the other tables were created to represent the corresponding weak entities (you can tell which table corresponds to which entity by their names). The weak entity tables have a primary key consisting of ("year","iso") with "iso" being a foreign key referencing the "country" table's "iso" column with “ON DELETE CASCADE” applied to maintain referential integrity.

I have removed the database name from the queries, for an easier reading experience. If you want to see the full queries that I used refer to “Step 2/create_tables.sql”.

SQL queries:

```
CREATE TABLE `country` (
    `iso` CHAR(8) NOT NULL,
    `name` VARCHAR(50) NOT NULL,
    PRIMARY KEY (`iso`));

CREATE TABLE `population` (
     `year` INT NOT NULL,
     `population` INT NOT NULL,
     `iso` CHAR(8) NOT NULL,
     PRIMARY KEY (`year`, `iso`),
     FOREIGN KEY (`iso`)
     REFERENCES `country` (`iso`)
     ON DELETE CASCADE);

CREATE TABLE `inequality_in_life` (
     `year` INT NOT NULL,
     `inequality` FLOAT NOT NULL,
     `iso` CHAR(8) NOT NULL,
     PRIMARY KEY (`year`, `iso`),
     FOREIGN KEY (`iso`)
     REFERENCES `country` (`iso`)
     ON DELETE CASCADE);

CREATE TABLE `health_expenditure` (
     `year` INT NOT NULL,
```

```
      `expenditure` FLOAT NOT NULL,
      `iso` CHAR(8) NOT NULL,
      PRIMARY KEY (`year`, `iso`),
      FOREIGN KEY (`iso`)
      REFERENCES `country` (`iso`)
      ON DELETE CASCADE);

CREATE TABLE `life_expectancy_all` (
      `year` INT NOT NULL,
      `le` FLOAT NOT NULL,
      `iso` CHAR(8) NOT NULL,
      PRIMARY KEY (`year`, `iso`),
      FOREIGN KEY (`iso`)
      REFERENCES `country` (`iso`)
      ON DELETE CASCADE);

CREATE TABLE `life_expectancy_gender` (
      `year` INT NOT NULL,
      `le_male` FLOAT NOT NULL,
      `le_female` FLOAT NOT NULL,
      `iso` CHAR(8) NOT NULL,
      PRIMARY KEY (`year`, `iso`),
      FOREIGN KEY (`iso`)
      REFERENCES `country` (`iso`)
      ON DELETE CASCADE);

CREATE TABLE `survival_to_65` (
      `year` INT NOT NULL,
      `survival_male` FLOAT NOT NULL,
      `survival_female` FLOAT NOT NULL,
      `iso` CHAR(8) NOT NULL,
      PRIMARY KEY (`year`, `iso`),
      FOREIGN KEY (`iso`)
      REFERENCES `country` (`iso`)
      ON DELETE CASCADE);
```