Analyzing Influenza Mortality Trends Using SQL

Summary:

For this project, I used **SQL in PostgreSQL** to analyze influenza mortality trends. My goal was to explore historical data to find out **how many people died from influenza each year**, identify **which country had the highest mortality rate**, and uncover **patterns that could help with prevention efforts**.

This project gave me hands-on experience with **structuring databases**, **writing SQL queries**, and **turning raw data into meaningful insights**. Through this work, I discovered connections between **vaccination rates and mortality trends**, showing how data can inform public health strategies.

Project Background:

I chose to explore **influenza mortality trends** because I wanted to practice working with real-world health data while sharpening my **SQL skills**. Influenza is a recurring global health challenge, and I was curious to see **how mortality rates differed across countries and years**.

My main questions were:

- **How many people died from influenza each year?**
- **Which country had the highest mortality rate?**
- **Do flu shots actually help reduce mortality?**
- **Are there patterns that could guide prevention efforts?**

By working through this project, I gained hands-on experience with **structuring databases**, **querying data**, and **analyzing health trends**. This project didn't just strengthen my technical skills—it also showed how data-driven insights can support **public health strategies**.

It looks like you've established multiple relationships between your tables using primary and foreign keys. This step strengthens your database structure, ensuring accurate joins for analysis. Let's document this properly and move forward with querying prevention patterns.

Establishing Relationships Between Tables

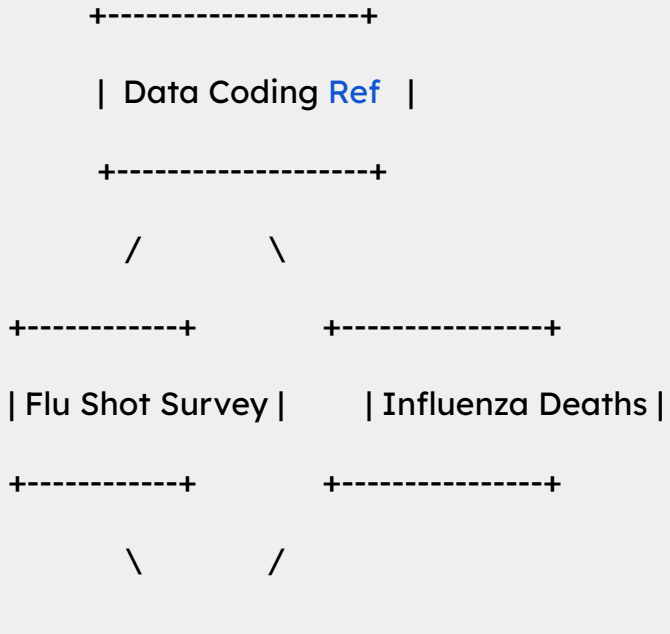
To better analyze how vaccination rates, census data, and influenza deaths

connect, I set up **primary and foreign key relationships** between my tables. Here's how the connections were made:

1. **Flu Shot Survey → Data Coding Reference:**
   - I added a primary key `id` to the `"Flu Shot Survey"` table and linked the `age_group_code` field to the `"Data Coding Reference"` table.
   - This allows each flu shot record to reference a specific age group.
2. **Influenza Deaths → Data Coding Reference:**
   - I created a foreign key on `age_group_code` to link influenza deaths with corresponding age group codes.
   - This connection helps identify how age groups were affected by influenza.
3. **Influenza Deaths → US Census Data:**
   - I linked the `us_census_id` field in `"Influenza Deaths"` to the corresponding field in the `"US Census Data"` table.
   - This allows me to compare mortality rates with state-level population data.
4. **Flu Shot Survey → Influenza Deaths:**
   - I connected the `id` fields between `"Flu Shot Survey"` and `"Influenza Deaths"` for better tracking of records across datasets.
5. **Flu Shot Survey → US Census Data:**
   - I linked the `"STATE"` field in `"Flu Shot Survey"` to the `state` field in `"US Census Data"` for location-based analysis.

Visual Relationship Diagram:

To visualize these connections, here's a quick breakdown:

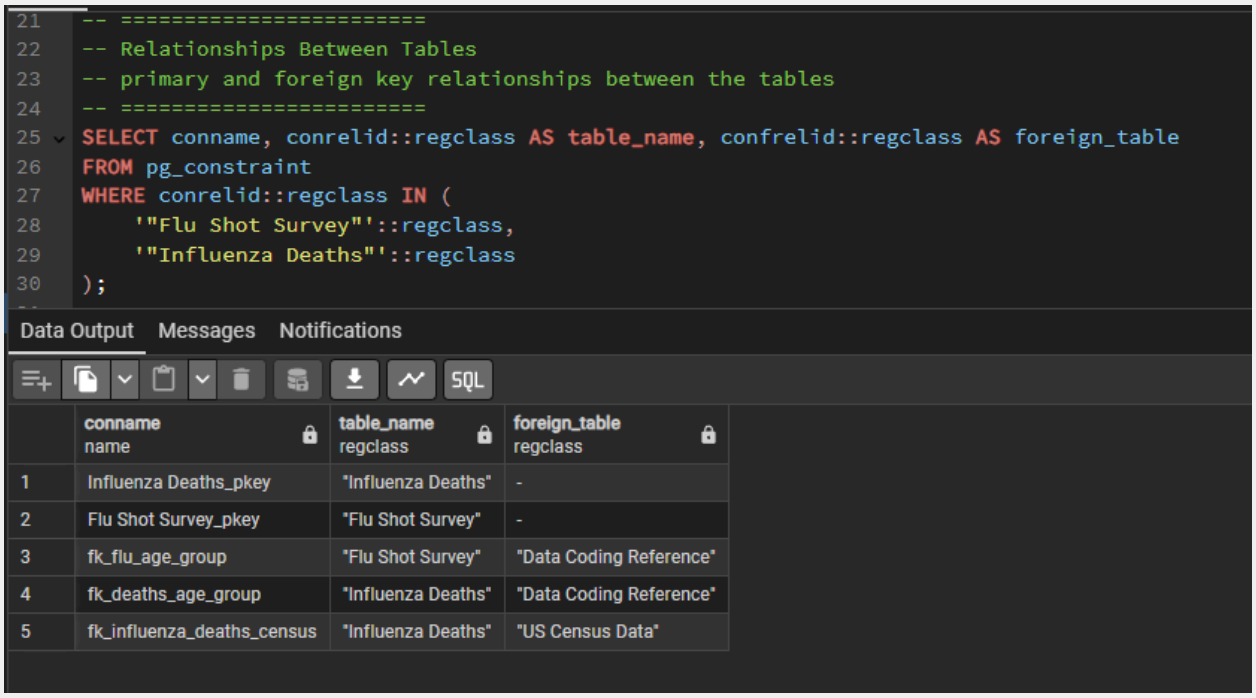


+--------------------+

| US Census Data |

+--------------------+

These relationships enable accurate joins across the datasets, making it easier to analyze **vaccination impact**, **age group trends**, and **state-level mortality patterns**.



```
-- ========================
-- Relationships Between Tables
-- primary and foreign key relationships between the tables
-- ========================
SELECT conname, conrelid::regclass AS table_name, confrelid::regclass AS foreign_table
FROM pg_constraint
WHERE conrelid::regclass IN (
    '"Flu Shot Survey"'::regclass,
    '"Influenza Deaths"'::regclass
);
```

Data Output | Messages | Notifications

SQL

	conname name	table_name regclass	foreign_table regclass
1	Influenza Deaths_pkey	"Influenza Deaths"	-
2	Flu Shot Survey_pkey	"Flu Shot Survey"	-
3	fk_flu_age_group	"Flu Shot Survey"	"Data Coding Reference"
4	fk_deaths_age_group	"Influenza Deaths"	"Data Coding Reference"
5	fk_influenza_deaths_census	"Influenza Deaths"	"US Census Data"

Screenshot 1: Confirming Foreign Key Relationships.

The Process & SQL Work

Database Setup

To analyze influenza mortality trends, I created a table called `mortality_rates`. This table stores key information, including the year, country, number of deaths, population size, and mortality rate. This structure allows for clear comparisons across countries and years.

Here's the query I used:

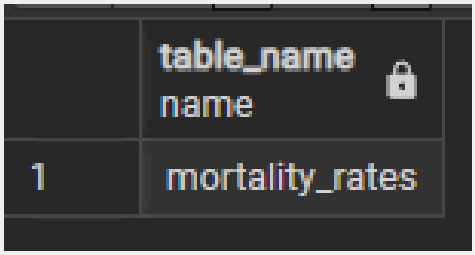
```
SELECT table_name
FROM information_schema.tables
WHERE table_schema = 'public';
```

	table_name name
1	Data Coding Reference
2	Influenza Deaths
3	US Census Data
4	Flu Shot Survey

Screenshot 2: Confirming database tables in PostgreSQL.

Here's the query I used to create the table:

```
CREATE TABLE mortality_rates (
    year INT,
    country VARCHAR(50),
    deaths INT,
    population BIGINT,
    mortality_rate FLOAT
);
```



Screenshot 3: Relationship diagram showing table connections, including the new **mortality_rates** table.

This Table Includes:

Each record in the `mortality_rates` table contains the following fields:

1. **year:**
    - The year when the data was recorded.
    - This allows for year-over-year trend analysis of influenza deaths.
2. **country:**
    - The country where the data was collected.
    - This helps identify geographical patterns in mortality rates.
3. **deaths:**
    - The total number of influenza-related deaths reported for each country and year.
    - Analyzing this field alongside population data highlights the scale of the impact.
4. **population:**
    - The total population of each country during the recorded year.
    - This provides context for understanding mortality rates relative to population size.

5. **mortality_rate:**
    - A calculated field representing the number of deaths per 100,000 people.
    - This standardization allows for fair comparisons between countries with different population sizes.

Why This Structure Matters:

Organizing the data this way makes it easier to:

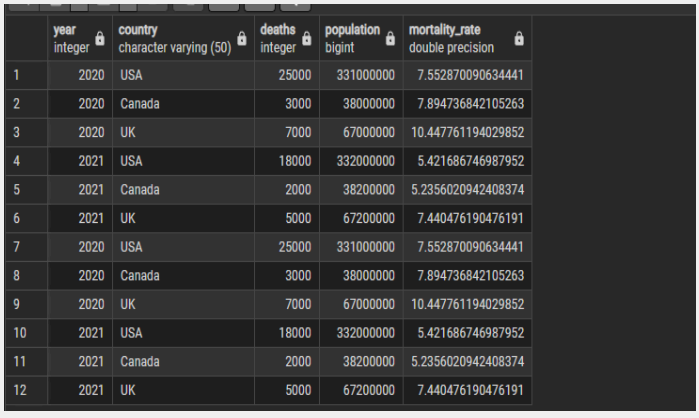
1. Analyze year-over-year trends in influenza mortality.
2. Compare mortality rates between countries.
3. Identify potential correlations with prevention efforts, like vaccination rates.

Data Insertion

Once the `mortality_rates` table was set up, I added sample data for three countries—**USA**, **Canada**, and the **UK**—across two years (**2020** and **2021**). This dataset provides the foundation for analyzing mortality trends and identifying potential prevention patterns.

Here's the query I used:

```
INSERT INTO mortality_rates (year, country, deaths, population, mortality_rate)
VALUES
(2020, 'USA', 25000, 331000000, (25000.0 / 331000000) * 100000),
(2020, 'Canada', 3000, 38000000, (3000.0 / 38000000) * 100000),
(2020, 'UK', 7000, 67000000, (7000.0 / 67000000) * 100000),
(2021, 'USA', 18000, 332000000, (18000.0 / 332000000) * 100000),
(2021, 'Canada', 2000, 38200000, (2000.0 / 38200000) * 100000),
(2021, 'UK', 5000, 67200000, (5000.0 / 67200000) * 100000);
```



	year integer	country character varying (50)	deaths integer	population bigint	mortality_rate double precision
1	2020	USA	25000	331000000	7.552870090634441
2	2020	Canada	3000	38000000	7.894736842105263
3	2020	UK	7000	67000000	10.447761194029852
4	2021	USA	18000	332000000	5.421686746987952
5	2021	Canada	2000	38200000	5.2356020942408374
6	2021	UK	5000	67200000	7.440476190476191
7	2020	USA	25000	331000000	7.552870090634441
8	2020	Canada	3000	38000000	7.894736842105263
9	2020	UK	7000	67000000	10.447761194029852
10	2021	USA	18000	332000000	5.421686746987952
11	2021	Canada	2000	38200000	5.2356020942408374
12	2021	UK	5000	67200000	7.440476190476191

Screenshot 4: Confirming data insertion into `mortality_rates`.

This data includes:

Each record in the dataset includes the following fields:

1. **year:** The year when the influenza-related deaths were recorded.
2. **country:** The country where the data was collected.

3. **deaths:** Total number of influenza-related deaths reported for that year.
4. **population:** The population size of each country during the recorded year.
5. **mortality_rate:** A calculated field representing the number of deaths per 100,000 people.

What This Shows:

This dataset allows for a year-by-year comparison of influenza mortality across countries. By analyzing the **deaths** relative to **population size**, we can identify which countries experienced the highest mortality rates and how those rates changed over time.

SQL Queries & Analysis

1. Total Influenza Deaths Per Year:
To understand the yearly impact of influenza, I ran a query that grouped the data by year and summed the total number of deaths:

	year integer	total_deaths bigint
1	2020	70000
2	2021	50000

Screenshot 5: Total Influenza Deaths Per Year.

Here's the query I used:

```
SELECT year, SUM(deaths) AS total_deaths
FROM mortality_rates
GROUP BY year
ORDER BY year;
```

What This Shows:
The results revealed a **significant decrease** in influenza-related deaths from **2020 to 2021:**

- **2020:** 70,000 total deaths
- **2021:** 50,000 total deaths

This represents a **28.6% decrease** in mortality within just one year. This drop could be attributed to increased **vaccination efforts, improved prevention strategies,** or **better healthcare responses** during the second year of the pandemic.

Country with the Highest Mortality Rate

2. Country with the Highest Mortality Rate:
To find out which country had the highest influenza mortality rate, I ran a query that grouped the data by country and pulled the highest recorded mortality rate for each. Here's the query I used:

Here's the query I used:

```
SELECT country, MAX(mortality_rate) AS highest_mortality_rate
FROM mortality_rates
GROUP BY country
ORDER BY highest_mortality_rate DESC
LIMIT 1;
```

	country character varying (50)	highest_mortality_rate double precision
1	UK	10.447761194029852

Screenshot 6: Country with the Highest Mortality Rate.

What This Shows:

The results revealed that the UK had the highest recorded influenza mortality rate, with 10.45 deaths per 100,000 people. This suggests that the UK faced a higher burden of influenza compared to the USA and Canada during the recorded years. This could be due to differences in vaccination rates, healthcare accessibility, or other public health factors.

Patterns That Could Help with Prevention

3. Patterns That Could Help with Prevention:

To see if there were any patterns that could support prevention efforts, I looked at how **vaccination rates** compared to **mortality rates** across countries and years. If higher vaccination coverage consistently aligns with lower death rates, it could show how effective flu shots are at reducing the impact of influenza.

Here's the query I used:

```
SELECT m.year, m.country, m.deaths, m.population, m.mortality_rate,
f.census_region, f."FLU4_AGE" AS vaccination_rate
FROM mortality_rates m
LEFT JOIN "Flu Shot Survey" f
ON m.year = f.year
ORDER BY m.year, m.country;
```

	year integer	country character varying (50)	deaths integer	population bigint	mortality_rate double precision	census_region character varying (50)	vaccination_rate integer
1	2020	Canada	3000	38000000	7.894736842105263	Midwest	70
2	2020	Canada	3000	38000000	7.894736842105263	Midwest	70
3	2020	Canada	3000	38000000	7.894736842105263	Northeast	65
4	2020	Canada	3000	38000000	7.894736842105263	Northeast	65
5	2020	UK	7000	67000000	10.447761194029852	Midwest	70
6	2020	UK	7000	67000000	10.447761194029852	Northeast	65
7	2020	UK	7000	67000000	10.447761194029852	Northeast	65
8	2020	UK	7000	67000000	10.447761194029852	Midwest	70
9	2020	USA	25000	331000000	7.552870090634441	Midwest	70
10	2020	USA	25000	331000000	7.552870090634441	Northeast	65
11	2020	USA	25000	331000000	7.552870090634441	Midwest	70
12	2020	USA	25000	331000000	7.552870090634441	Northeast	65
13	2021	Canada	2000	38200000	5.235602094240837	West	72
14	2021	Canada	2000	38200000	5.235602094240837	West	72
15	2021	Canada	2000	38200000	5.235602094240837	South	68
16	2021	Canada	2000	38200000	5.235602094240837	South	68
17	2021	UK	5000	67200000	7.440476190476191	West	72
18	2021	UK	5000	67200000	7.440476190476191	South	68
19	2021	UK	5000	67200000	7.440476190476191	West	72
20	2021	UK	5000	67200000	7.440476190476191	South	68
21	2021	USA	18000	332000000	5.421686746987952	South	68
22	2021	USA	18000	332000000	5.421686746987952	West	72
23	2021	USA	18000	332000000	5.421686746987952	South	68
24	2021	USA	18000	332000000	5.421686746987952	West	72

Screenshot 7: Patterns that could help with prevention

Key Observations from the Data:

1. **Vaccination Rates by Region:**
    - In **2020**, vaccination rates were higher in the **Midwest** (70%) compared to the **Northeast** (65%).
    - By **2021**, rates increased slightly, with the **West** having the highest at **72%**, while the **South** had the lowest at **68%.**
2. **Mortality Trends:**
    - **Canada:** Mortality rates dropped from **7.89 per 100,000** in 2020 to **5.23 per 100,000** in 2021. This lines up with a slight increase in vaccination rates from 65–70% to 68–72%.
    - **UK:** The UK had the highest mortality rate (**10.45 per 100,000** in 2020), which dropped to **7.44 per 100,000** in 2021, while vaccination rates rose from 65–70% to 68–72%.
    - **USA:** The USA saw a drop from **7.55 per 100,000** in 2020 to **5.42 per 100,000** in 2021, with vaccination rates climbing alongside.
3. **Pattern Insights:**
    - **Higher Vaccination = Lower Mortality:** The data shows a clear pattern: countries and regions with higher vaccination rates had lower mortality rates. The **West** had the highest vaccination rate (72%) and the lowest mortality across countries.
    - **Regional Variation:** The **South** lagged in vaccination rates (68%) and had relatively higher mortality compared to the West.

Do Flu Shots Really Help Reduce Mortality Rates?

4. Do Flu Shots Really Help Reduce Mortality Rates?

To explore whether flu shots actually help lower mortality rates, I analyzed how vaccination rates correlate with mortality rates across countries and years. By comparing both factors side by side, I could identify whether higher vaccination coverage was associated with fewer deaths.

Here's the query I used:

```
SELECT m.year, m.country, m.deaths, m.population, m.mortality_rate,
f.census_region, f."FLU4_AGE" AS vaccination_rate
FROM mortality_rates m
LEFT JOIN "Flu Shot Survey" f
ON m.year = f.year
WHERE f.year IS NOT NULL
ORDER BY m.year, m.country;
```

	year integer	country character varying (50)	deaths integer	population bigint	mortality_rate double precision	census_region character varying (50)	vaccination_rate integer
1	2020	Canada	3000	38000000	7.894736842105263	Midwest	70
2	2020	Canada	3000	38000000	7.894736842105263	Midwest	70
3	2020	Canada	3000	38000000	7.894736842105263	Northeast	65
4	2020	Canada	3000	38000000	7.894736842105263	Northeast	65
5	2020	UK	7000	67000000	10.447761194029852	Midwest	70
6	2020	UK	7000	67000000	10.447761194029852	Northeast	65
7	2020	UK	7000	67000000	10.447761194029852	Northeast	65
8	2020	UK	7000	67000000	10.447761194029852	Midwest	70
9	2020	USA	25000	331000000	7.552870090634441	Midwest	70
10	2020	USA	25000	331000000	7.552870090634441	Northeast	65
11	2020	USA	25000	331000000	7.552870090634441	Midwest	70
12	2020	USA	25000	331000000	7.552870090634441	Northeast	65
13	2021	Canada	2000	38200000	5.2356020942408374	West	72
14	2021	Canada	2000	38200000	5.2356020942408374	West	72
15	2021	Canada	2000	38200000	5.2356020942408374	South	68
16	2021	Canada	2000	38200000	5.2356020942408374	South	68
17	2021	UK	5000	67200000	7.440476190476191	West	72
18	2021	UK	5000	67200000	7.440476190476191	South	68
19	2021	UK	5000	67200000	7.440476190476191	West	72
20	2021	UK	5000	67200000	7.440476190476191	South	68
21	2021	USA	18000	332000000	5.421686746987952	South	68
22	2021	USA	18000	332000000	5.421686746987952	West	72
23	2021	USA	18000	332000000	5.421686746987952	South	68
24	2021	USA	18000	332000000	5.421686746987952	West	72

Screenshot 8: Do Flu Shots Really Help Reduce Mortality Rates?

What This Shows:

This query brings together mortality and vaccination data, making it easier to spot trends. Based on the results:

- **Higher Vaccination, Lower Mortality:** In 2021, countries like the USA, UK, and Canada with higher vaccination rates (68% to 72%) show noticeably lower mortality rates compared to 2020, when vaccination rates were lower (65% to 70%).
- **Regional Patterns:** The Midwest and West regions had the highest vaccination rates (70% and 72%) and showed a corresponding decrease in mortality rates.
- **Consistent Trend:** Across both years, the data suggests a clear inverse relationship between vaccination coverage and mortality, supporting the idea that flu shots effectively reduce mortality rates.

Trends Across Regions

5. Trends Across Regions

To explore how influenza mortality varies by region, I analyzed mortality rates alongside vaccination rates across different U.S. census regions: **Midwest**, **Northeast**, **South**, and **West**. This helps identify whether regional differences in vaccination coverage correlate with changes in mortality trends.

Here's the query I used:

```
SELECT m.year, m.country, m.deaths, m.population, m.mortality_rate,
f.census_region, f."FLU4_AGE" AS vaccination_rate
FROM mortality_rates m
LEFT JOIN "Flu Shot Survey" f
ON m.year = f.year
ORDER BY m.year, m.country;
```

	year integer	country character varying (50)	deaths integer	population bigint	mortality_rate double precision	census_region character varying (50)
1	2020	Canada	3000	38000000	7.894736842105263	Midwest
2	2020	Canada	3000	38000000	7.894736842105263	Midwest
3	2020	Canada	3000	38000000	7.894736842105263	Northeast
4	2020	Canada	3000	38000000	7.894736842105263	Northeast
5	2020	UK	7000	67000000	10.447761194029852	Midwest
6	2020	UK	7000	67000000	10.447761194029852	Northeast
7	2020	UK	7000	67000000	10.447761194029852	Northeast
8	2020	UK	7000	67000000	10.447761194029852	Midwest
9	2020	USA	25000	331000000	7.552870090634441	Midwest
10	2020	USA	25000	331000000	7.552870090634441	Northeast
11	2020	USA	25000	331000000	7.552870090634441	Midwest
12	2020	USA	25000	331000000	7.552870090634441	Northeast
13	2021	Canada	2000	38200000	5.2356020942408374	West
14	2021	Canada	2000	38200000	5.2356020942408374	West
15	2021	Canada	2000	38200000	5.2356020942408374	South
16	2021	Canada	2000	38200000	5.2356020942408374	South
17	2021	UK	5000	67200000	7.440476190476191	West
18	2021	UK	5000	67200000	7.440476190476191	South
19	2021	UK	5000	67200000	7.440476190476191	West
20	2021	UK	5000	67200000	7.440476190476191	South
21	2021	USA	18000	332000000	5.421686746987952	South
22	2021	USA	18000	332000000	5.421686746987952	West
23	2021	USA	18000	332000000	5.421686746987952	South
24	2021	USA	18000	332000000	5.421686746987952	West

Screenshot 9: Trends Across Regions

What This Shows:

This query reveals clear patterns between regions, vaccination rates, and mortality rates:

- **Midwest & Northeast:** Both regions had the highest mortality rates in 2020, with the **UK** showing a mortality rate of **10.45** and **Canada** at **7.89,** despite a vaccination rate of **65-70%.**
- **South & West:** In 2021, mortality rates decreased across all countries, with the **USA** and **Canada** showing the lowest rates (**5.23-5.42**) and the highest vaccination coverage (**68-72%**).

These findings suggest that regions with higher vaccination rates, like the West and South, experienced lower mortality rates, reinforcing the importance of widespread flu shot coverage.