Marisa Wong
INFO 521

Analyzing Foodborne Outbreaks around the release of FDA's 2017 Food Code

Introduction

Foodborne illnesses are a major cause of morbidity and mortality in the United States and can lead to a detrimental impact on the healthcare system. Every year in the United States, almost 50 million people become ill from contaminated food ingredients or preparation processes, leading to almost 130,000 hospitalizations and over 3,000 deaths[1]. Some common symptoms associated with foodborne illnesses include diarrhea and vomiting that can last up to 7 days. Other less common symptoms include cramping, fever, body aches, and fatigue[2]. Scientists have identified the five most common pathogens related to foodborne illnesses, and they include norovirus, nontyphoidal *Salmonella*, *C. perfringens*, many *Campylobacter* species, and *Staphylococcus aureus*[3]. To intervene and reduce the number of foodborne disease outbreaks in the United States, public health professionals and food safety scientists have developed many regulations and guidance documents to ensure safe food handling practices at retail food establishments.

The U.S. Food and Drug Administration (FDA) is responsible for researching and publishing the national Food Code. This code aims to assist local and state governments in the creation of their own food safety regulations, to serve as a model for local and state governments, and to provide these governments with the most up to date scientific research for regulating the food industry[4]. Many states often adopt the Food Code by short-form (adoption clause) or long-form (by reference) methods. The Food Code is not an enforceable requirement set forth by FDA, so no entities receive reprimands for a lack of adoption. The Food Code is updated every four years, with the most recent update occurring in 2022[4]. In 2017, a new section to the Food Code was added that advises gloves to be worn over any single use bandages or finger cots to prevent these items from contaminating foods[5]. This project intends to assess foodborne disease outbreaks in the United States surrounding the release of FDA's 2017 Food Code by utilizing the National Outbreak Response System (NORS) to understand whether the release of the 2017 Food Code and its new section had any impact on foodborne outbreaks.

Dataset

The National Outbreak Response System (NORS) is an online surveillance site hosted by the U.S. Centers for Disease Control and Prevention (CDC). All reportable outbreaks are detailed on this system by local and state health departments. To ensure that data is communicated in a timely manner, NORS follows a general workflow that consists of the health department being notified of a possible outbreak, investigating the outbreak, and entering the details of the outbreak into NORS. Then, CDC verifies the data entered, summarizes the data, and eventually publishes it to the NORS dashboard[6]. NORS gathers data on waterborne, foodborne, person-to-person, animal contact, environmental contamination, and other enteric disease outbreaks[6]. NORS is a modern step forward for CDC in continuing to strengthen outbreak response systems and share relevant data to the public.

The NORS dashboard is an interactive website that allows users to simplify, find, and gain access to the data that they are interested in. At the top of the site, users can select the specific type of outbreak to be displayed on the dashboard in both quick statistics, a U.S. map, and a few bar and line graphs. The data available is from the years 1971 – 2021. Other specific variables that users can stratify by include custom date ranges, states, etiology, setting, food and ingredients, water exposure, and water type[7]. At the bottom of the site, users can download the data they specified to investigate further or perform additional analyses.

For this project, the dataset was downloaded from NORS on Monday, April 3rd, 2023. The data included in this set represent foodborne outbreaks from 2017 – 2018. The dataset includes 8,276 observations and 21 columns. The columns included in the dataset are detailed in Table 1.

Table 1. NORS Data Dictionary.

Column Name	*Data Type*	*Definition[a]*
Year	Numeric	Year of earliest date of reported illness
Month	Numeric	Month of earliest date of reported illness
State	Character	The state where the exposure occurred. For outbreaks that occurred in multiple states, "Multistate" will be listed.
Primary Mode	Character	Primary mode of transmission
Etiology	Character	Genus and species of identified etiology. Multiple reported etiologies are separated by a semicolon.
Serotype or Genotype	Character	Serotype or genotype of identified etiology. Multiple reported etiologies are separated by a semicolon.
Etiology Status	Character	Indicates whether the identified etiology is laboratory 'Confirmed' or 'Suspected'. Multiple reported etiologies are separated by a semicolon.
Setting	Character	Setting(s) where exposure occurred. The location of food consumed in foodborne outbreaks.
Illnesses	Numeric	Estimated total number of primary cases, including lab-confirmed and probable, based on the outbreak-specific definition.
Hospitalizations	Numeric	Number of primary cases who were hospitalized
Info on Hospitalizations	Numeric	Total number of primary cases for whom information on hospitalization is available
Deaths	Numeric	Number of primary cases who died
Info on Deaths	Numeric	Total number of primary cases for whom information on survival is available
Food Vehicle	Character	For foodborne outbreaks only, the implicated food. Multiple implicated foods are separated by a semicolon.
Food Contaminated Ingredient	Character	For foodborne outbreaks only, indicates the contaminated ingredient. Multiple contaminated ingredients are separated by a semicolon.
IFSAC Category	Character	For foodborne outbreaks only, the IFSAC food category of the contaminated ingredient. If contaminated ingredients fall into multiple IFSAC categories, "Multiple" will be listed.

Water Exposure	Character	For waterborne outbreaks only, the implicated type of water exposure.
Water Type	Character	For waterborne outbreaks only, a description of the venue (for treated and untreated recreational water), water system (for drinking water) or device/structure (for other/unknown; e.g., steam cleaner, cooling tower, ornamental fountain, etc.) that was the vehicle for waterborne exposure to microbial pathogens, chemicals, or toxins.
Animal Type	Character	For waterborne outbreaks only, indicates whether CDC has classified a waterborne disease outbreak report as excluded, cleaned, or closed out.
Animal Type Specify	Character	For animal contact outbreaks only, the type of animal involved. Multiple animal types are separated by a semicolon.
Water Status	Character	For animal contact outbreaks only, the specific animal type. Multiple animal types are separated by a semicolon.

[a]Definitions were provided by downloads from https://wwwn.cdc.gov/norsdashboard/.

Table Creation & Data Management

All four tables created for this project stem from the variables included in the downloaded NORS dataset. Variables were grouped together by relevancy of the data. The four tables that were generated include timeframe, etiology, outcomes, and situation. Tables were constructed using CREATE TABLE statements, and data were imported using INSERT INTO and VALUES statements. Each table contains 8,276 observations representing the number of outbreaks that occurred in 2017 and 2018. Each observation was assigned an identification number, called outbreak_id. To ensure all data associated with a specific outbreak were grouped together, outbreak_id acted as a foreign key in the etiology, outcomes, and situation tables while being a primary key in the timeframe table. Figures 1-6 display the SQL code and editors used to create the database, generate the accompanying tables, and insert the data into those tables. Table 2 documents the variables and their assigned constraints.

```sql
/* Creating a database with a new database schema */
CREATE SCHEMA nors_outbreaks;

CREATE TABLE `nors_outbreaks`.`timeframe`(
    `outbreak_id` INT NOT NULL AUTO_INCREMENT,
    `outbreak_year` YEAR NOT NULL,
    `outbreak_month` INT NOT NULL,
    `state` VARCHAR(50) NOT NULL,
    PRIMARY KEY (`outbreak_id`));

CREATE TABLE `nors_outbreaks`.`etiology` (
  `etiology_id` INT NOT NULL AUTO_INCREMENT,
  `transmission_mode` VARCHAR(1000) NULL,
  `serotype` VARCHAR(1000) NULL,
  `pathogen` VARCHAR(1000) NULL,
  `etiology_status` VARCHAR(1000) NULL,
  `outbreak_id` INT NOT NULL,
  PRIMARY KEY (`etiology_id`));
```

```sql
CREATE TABLE `nors_outbreaks`.`outcomes` (
    `outcomes_id` INT NOT NULL AUTO_INCREMENT,
    `num_illnesses` INT NULL,
    `num_hospitalizations` INT NULL,
    `num_deaths` INT NULL,
    `outbreak_id` INT NOT NULL,
    PRIMARY KEY (`outcomes_id`));

CREATE TABLE `nors_outbreaks`.`situation` (
    `situation_id` INT NOT NULL AUTO_INCREMENT,
    `setting` VARCHAR(1000) NULL,
    `food_vehicle` VARCHAR(1000) NULL,
    `contaminated_ingredient` VARCHAR(1000) NULL,
    `outbreak_id` INT NOT NULL,
    PRIMARY KEY (`situation_id`));
```

Figure 1. SQL code to generate nors_outbreaks schema and timeframe, etiology, outcomes, and situation tables.

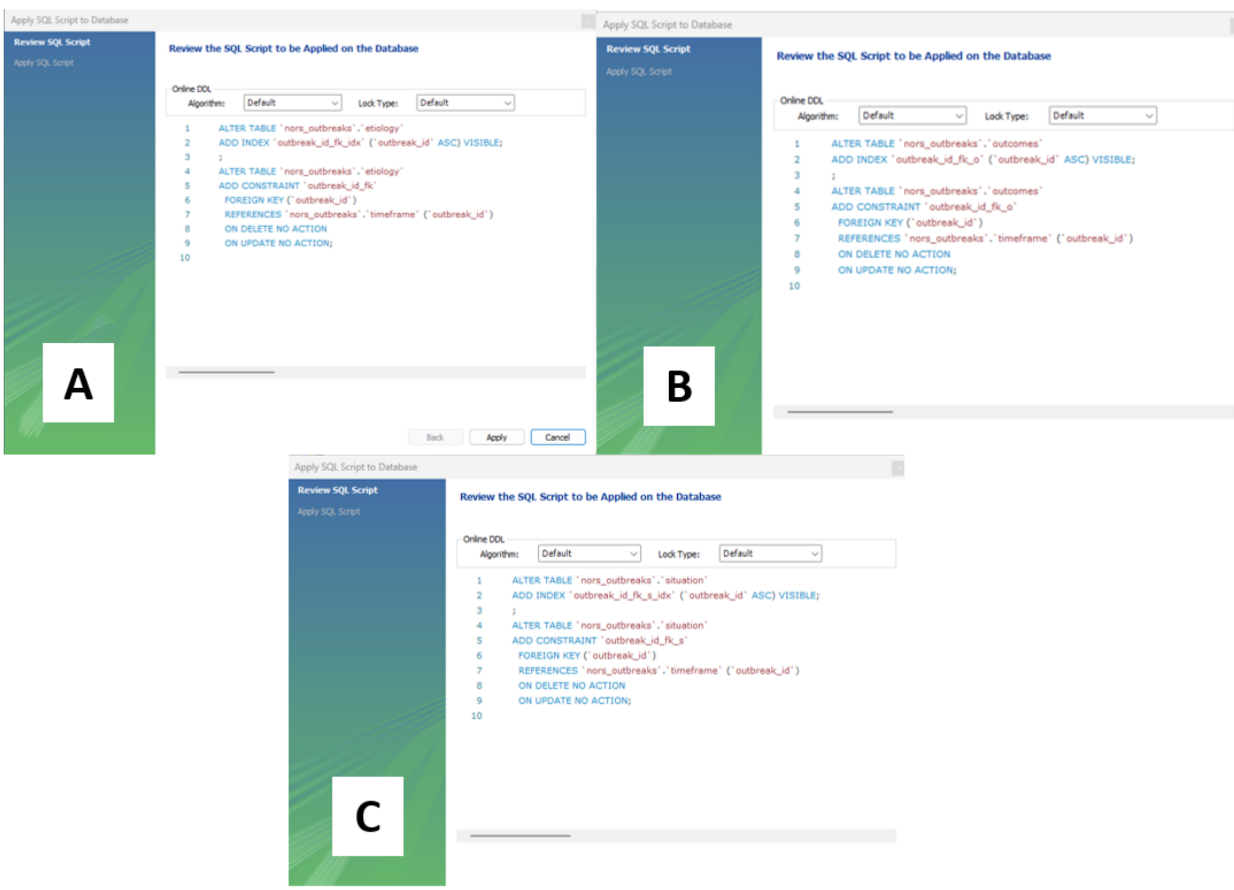

Figure 2. SQL Code that applies Foreign Constraints to child tables in nors_outbreaks schema. Panel A includes foreign constraints for the outbreak_id in the etiology table. Panel B

includes foreign constraints for the outbreak_id in the outcomes table. Panel C includes foreign constraints for the outbreak_id in the situation table.

A

```
INSERT INTO `nors_outbreaks`.`timeframe`
    (`outbreak_year`, `outbreak_month`, `state`)
VALUES
(2017,1,'Virginia'),
(2017,1,'Virginia'),
(2017,1,'Oregon'),
(2017,1,'Minnesota'),
(2017,1,'Ohio'),
```

B

```
INSERT INTO `nors_outbreaks`.`etiology`
    (`outbreak_id`, `transmission_mode`, `pathogen`, `serotype`, `etiology_status`)
VALUES
(1,'Person-to-person','Norovirus','', 'Suspected'),
(2,'Person-to-person','Norovirus Genogroup II','GII.4 Sydney', 'Confirmed'),
(3,'Person-to-person','Norovirus','', 'Suspected'),
(4,'Person-to-person','Norovirus','', 'Suspected'),
(5,'Person-to-person','Rotavirus; Sapovirus','', 'Confirmed; Suspected'),
```

C

```
INSERT INTO `nors_outbreaks`.`outcomes`
    (`outbreak_id`, `num_illnesses`, `num_hospitalizations`, `num_deaths`)
VALUES
(1,11, 0, 0),
(2,19, 0, 0),
(3,39, 0, 0),
(4,32, NULL, 0),
(5,17, 0, 0),
```

D

```
INSERT INTO `nors_outbreaks`.`situation`
    (`outbreak_id`, `setting`, `food_vehicle`, `contaminated_ingredient`)
VALUES
(1,'Hospital', NULL, NULL),
(2,'Other, specify', NULL, NULL),
(3,'Long-term care/nursing home/assisted living facility', NULL, NULL),
(4,'Long-term care/nursing home/assisted living facility', NULL, NULL),
(5,'Long-term care/nursing home/assisted living facility', NULL, NULL),
```

Figure 3. A sample of SQL Insert Into Statements that imported data from NORS into the nors_outbreaks tables. Panel A is a sample from the timeframe table. Panel B is a sample from the etiology table. Panel C is a sample from the outcomes table. Panel D is a sample from the situation table.

Table 2. Relational Database Details of nors_outbreaks Schema.

Table	*Column Name*	*Data Type*	*Constraints*[a, b]
timeframe	outbreak_id*	INT	Primary Key, AI, NN
	outbreak_year	YEAR	-
	outbreak_month	INT	-
	state	VARCHAR(50)	-
etiology	etiology_id*	INT	Primary Key, AI, NN

	transmission_mode	VARCHAR(1000)	-
	serotype	VARCHAR(1000)	-
	pathogen	VARCHAR(1000)	-
	etiology_status	VARCHAR(1000)	-
	outbreak_id**	INT	Foreign Key, NN
situation	situation_id*	INT	Primary Key, AI, NN
	setting	VARCHAR(1000)	-
	food_vehicle	VARCHAR(1000)	-
	contaminated_ingredient	VARCHAR(1000)	-
	outbreak_id**	INT	Foreign Key, NN
outcome	outcomes_id*	INT	Primary Key, AI, NN
	num_illnesses	INT	-
	num_hospitalizations	INT	-
	num_deaths	INT	-
	outbreak_id**	INT	Foreign Key, NN

[a]AI: autoincrement
[b]NN: not null

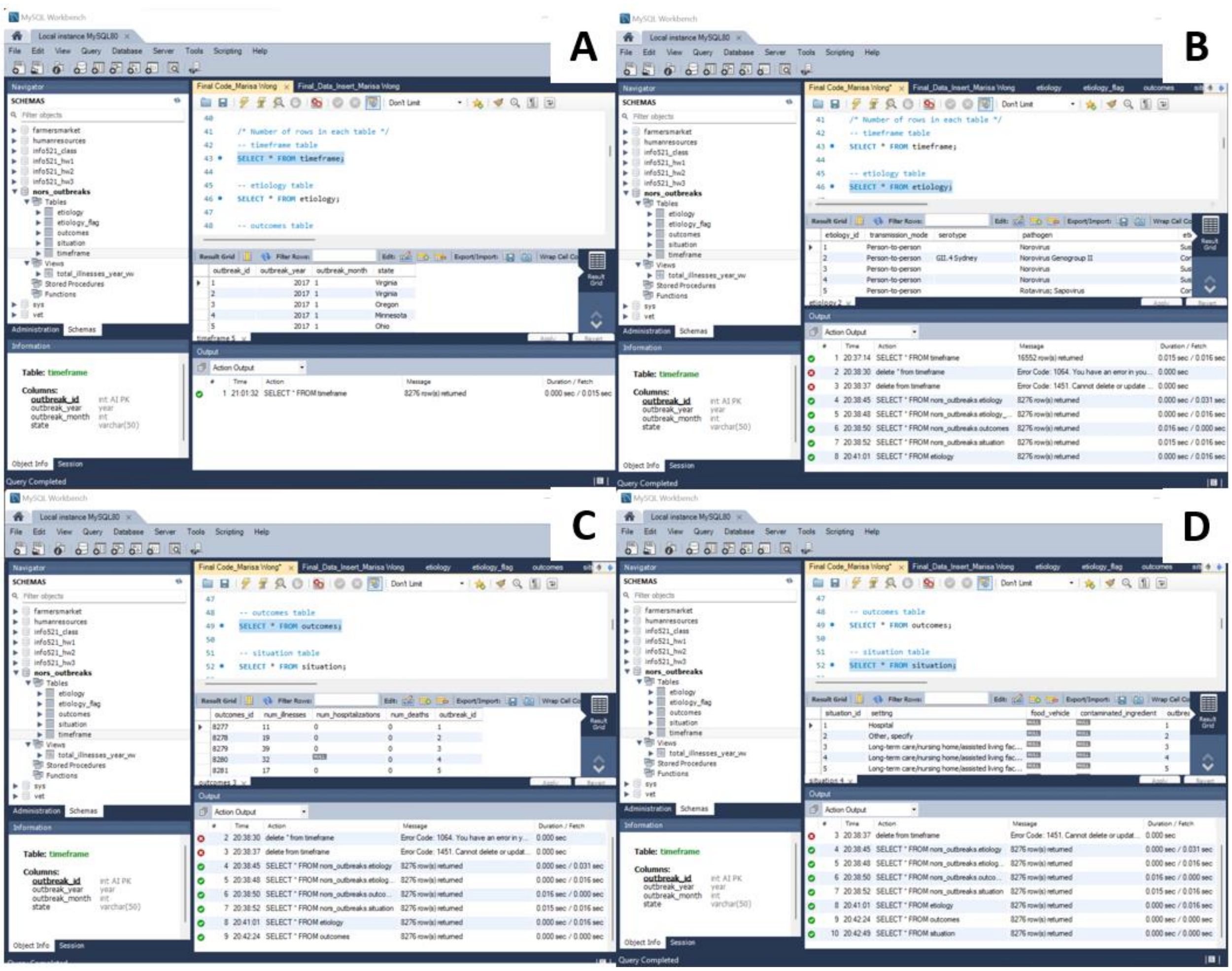

Figure 4. Number of records in each table. Panel A represents the timeframe table. Panel B represents the etiology table. Panel C represents the outcomes table. Panel D represents the situation table.

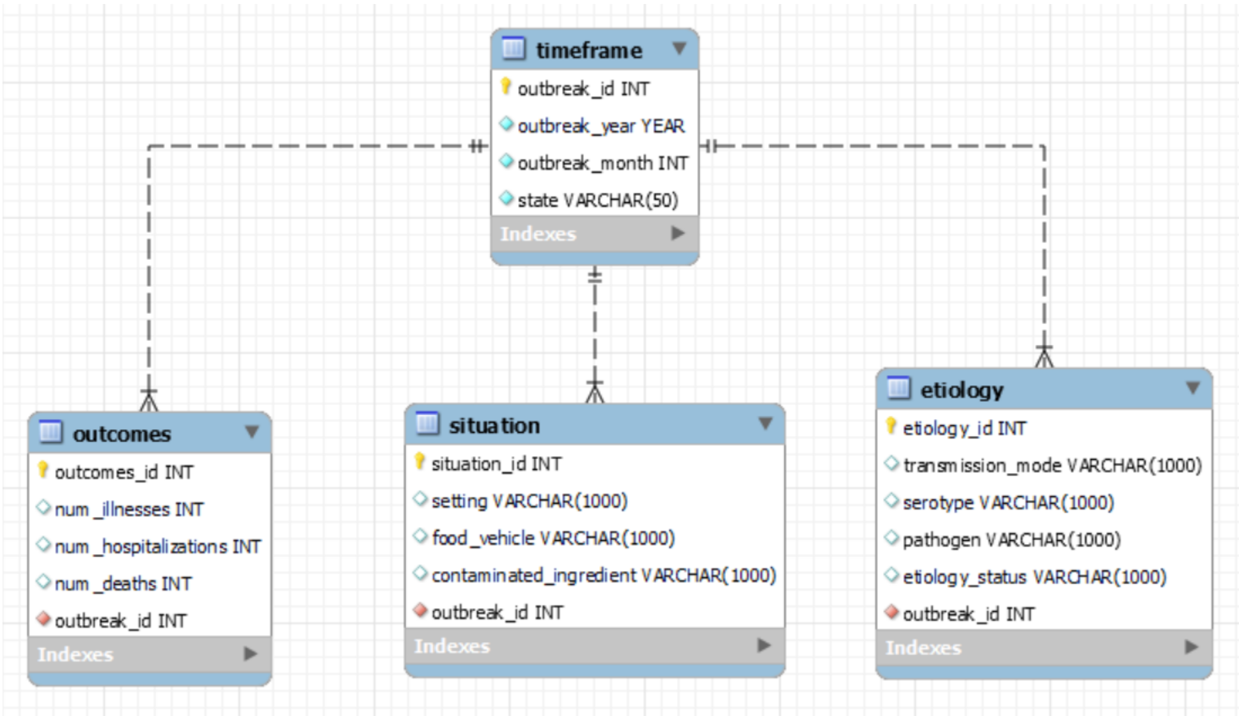

Figure 5. Entity-Relationship Diagram for nors_outbreaks schema. The yellow icon indicates the column is a primary key. The red icon indicates the column is a foreign key.

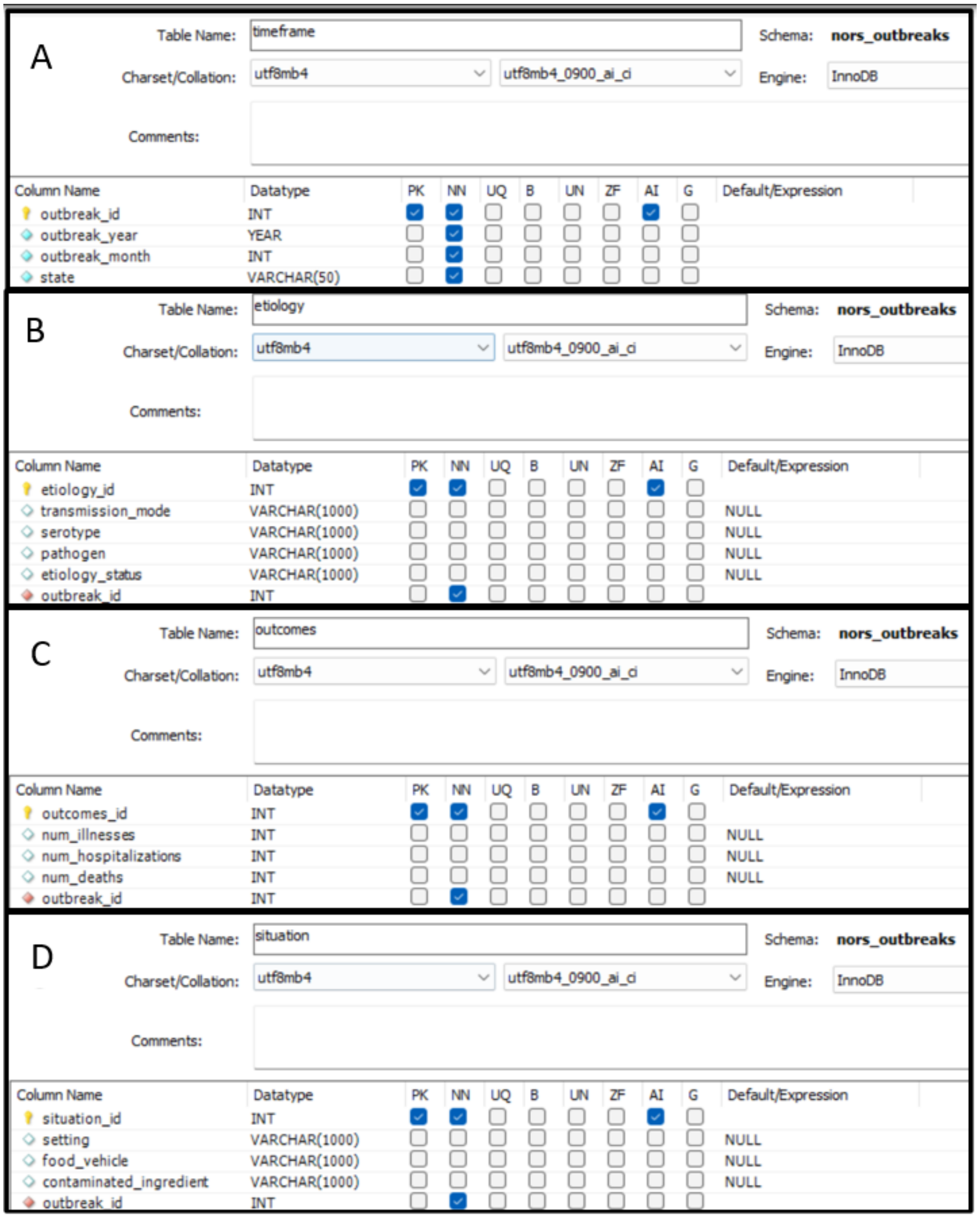

Figure 6. Table Settings and Data Constraints for the nors_outbreak schema. Panel A represents the details for the timeframe table. Panel B represents the details for the etiology table. Panel C represents the details for the outcomes table. Panel D represents the details for the situation table.

<u>Analytical Questions and SQL Query Execution</u>

To determine if the 2017 Food Code affected the number of foodborne illness outbreaks in the U.S. from 2017 to 2018, I developed five analytical questions to assess the change in foodborne outbreaks. The five analytical questions, their answers, and accompanying screenshots are listed below:

1. Question: How many total illnesses were reported to NORS in 2017 compared to 2018, grouped by state and outbreak year? Include outbreak year and total number of illnesses. Report the answer as a view.
    a. Answer: In 2017, there were 104,183 total illnesses. In 2018, there were 95,354 total illnesses. Figures 7 and 8 display the query execution and results.

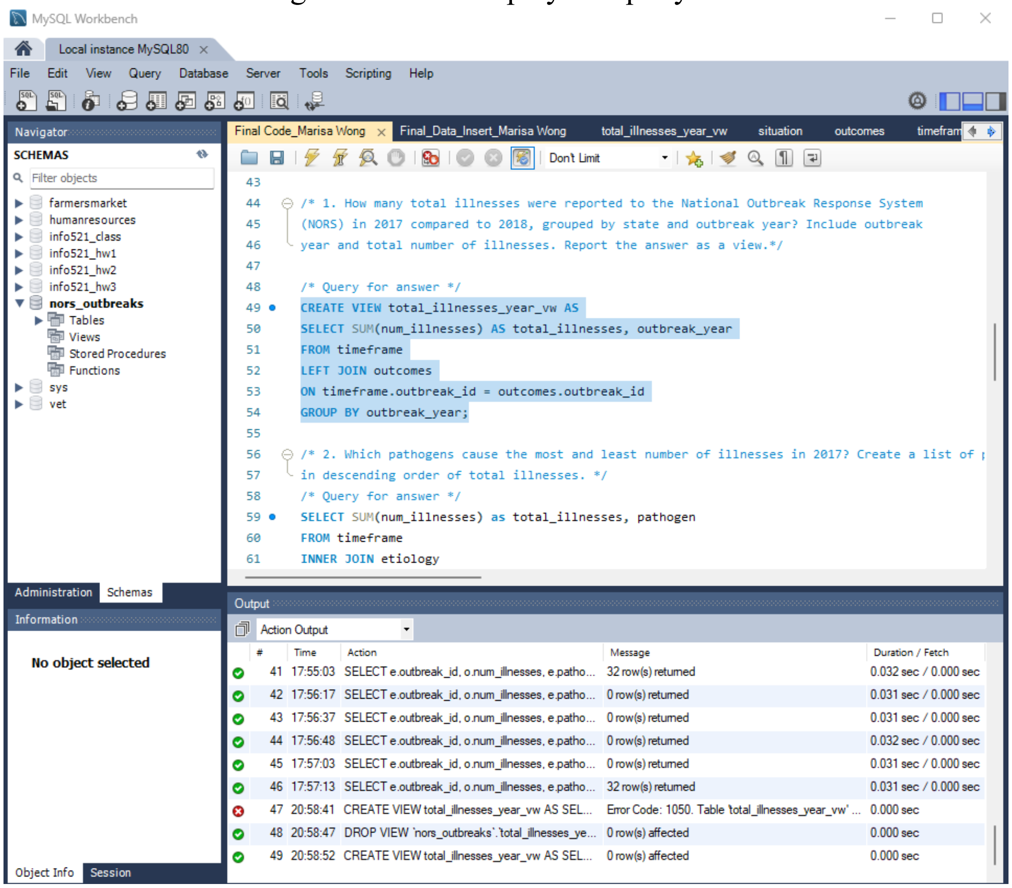

Figure 7. SQL Query for Question 1. The highlighted text represents the query run in this screenshot.

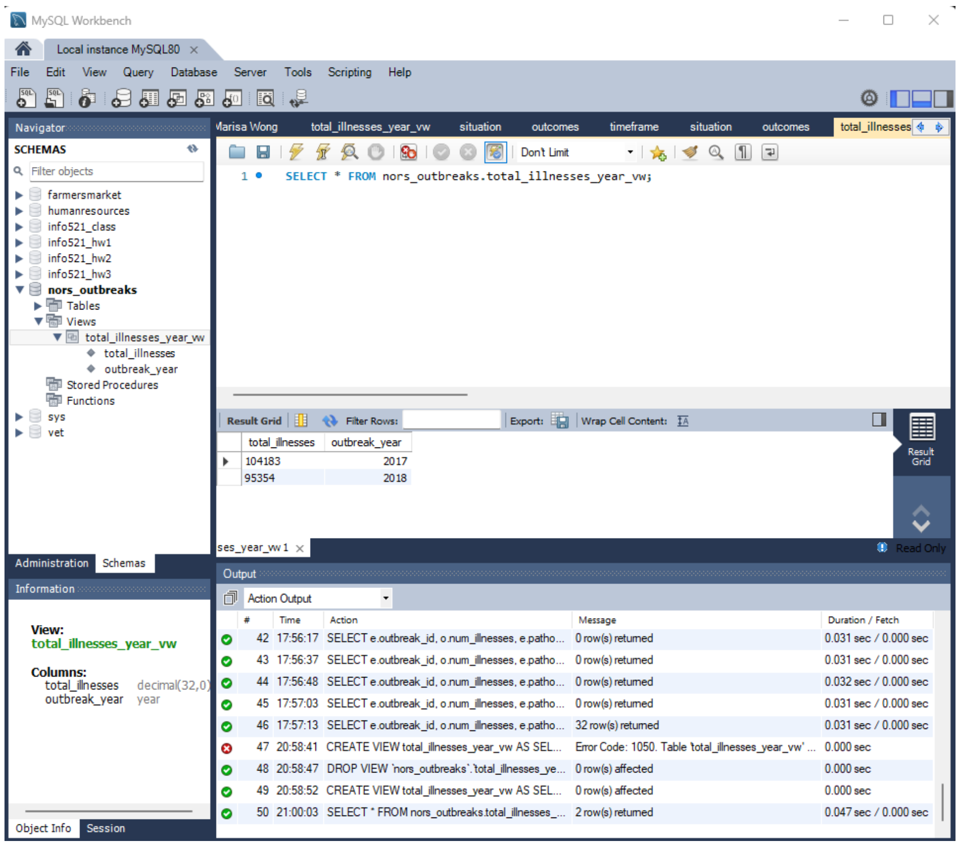

Figure 8. Output for Question 1. A view produced with two rows and two columns.

2. Question: Which pathogens cause the most and least number of illnesses in 2017? Create a list of pathogens and total illnesses in descending order of total illnesses. Conduct the same analysis for 2018.
    a. Answer: In 2017, Norovirus Genogroup II causes the highest number of illnesses with 23,118 illnesses attributed to that pathogen. In 2017, Salmonella other, Vibrio parahaemolyticus (Vibrio unknown), Legionella pneumophila (Legionella anisa), Legionella pneumophila (Legionella pneumophila; Legionella other), Vibrio vulnificus, Vibrio parahaemolyticus (Other – Bacterium), Unknown – Virus, and Other – Virus were all tied with the least number of illnesses coming in at 2 each. See INFO521_Final_MarisaWong_Q2_2017_List.csv for the list. In 2018, Norovirus unknown caused 21,149 illnesses, which was the highest for that year in this dataset. In 2018, Mycotoxins, Campylobacter coli (Campylobacter unknown), Chlorine (Chlorine Gas), Paralytic shellfish poison, and Campylobacter coli were all tied with causing the least number of illnesses coming in at 2 each. See INFO521_Final_MarisaWong_Q2_2018_List.csv for the list.

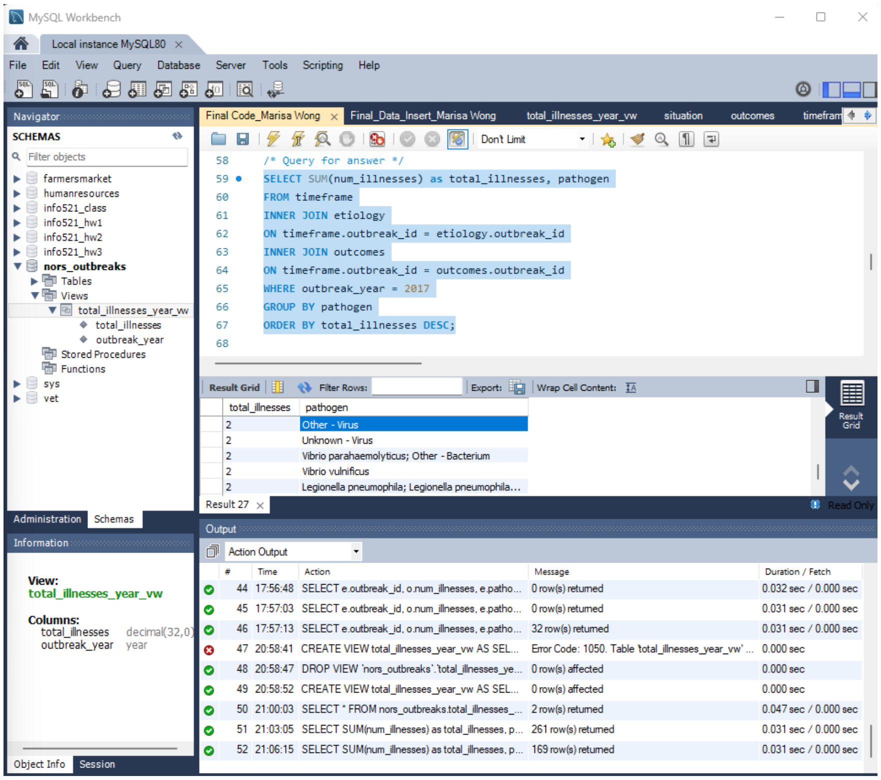

Figure 9. SQL Query for 2017 in Question 2. The highlighted text represents the query run in this screenshot.

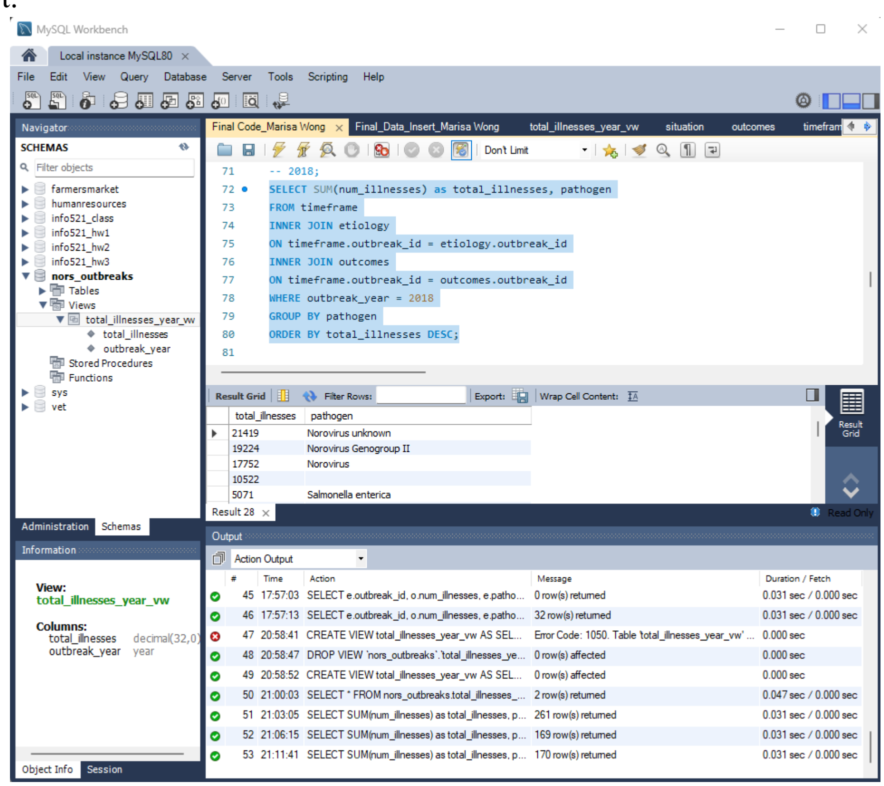

Figure 10. SQL Query for 2018 in Question 2. The highlighted text represents the query run in this screenshot.

3. Question: What transmission mode led to the most outbreaks in 2017? What about in 2018?
    a. Answer: In 2017, person-to-person transmission led to 2,551 outbreaks, which was the most for 2017. See INFO521_Final_MarisaWong_Q3_2017_List.csv for a complete list. In 2018, person-to-person transmission led to 2,397 outbreaks, which was the most for 2018. See INFO521_Final_MarisaWong_Q3_2018_List.csv for a complete list.

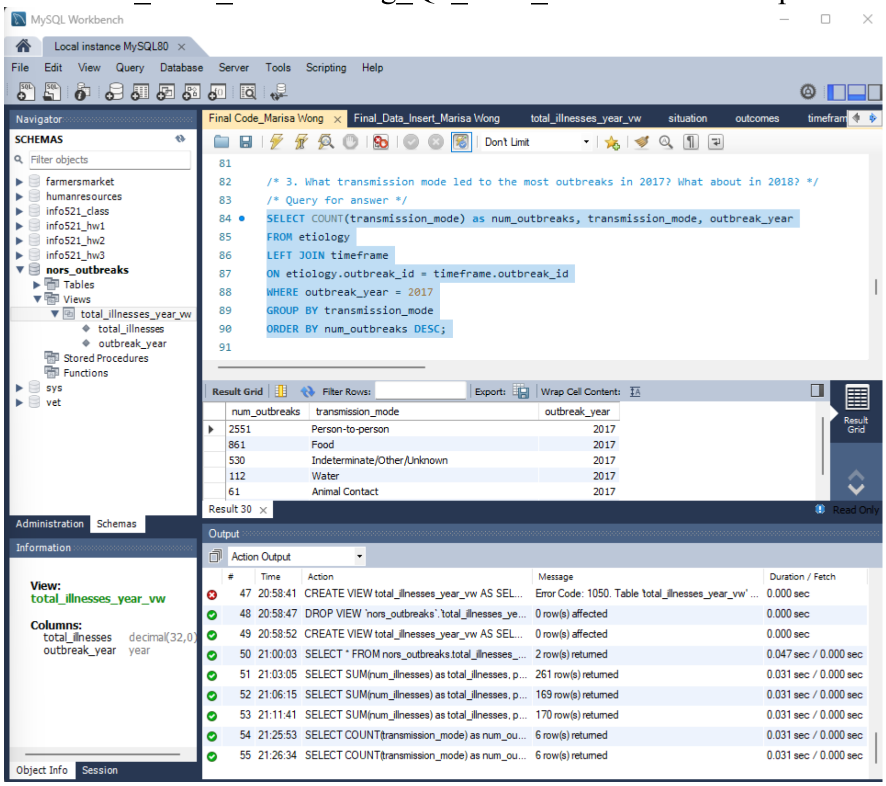

Figure 11. SQL Query for 2017 in Question 3. The highlighted text represents the query run in this screenshot.

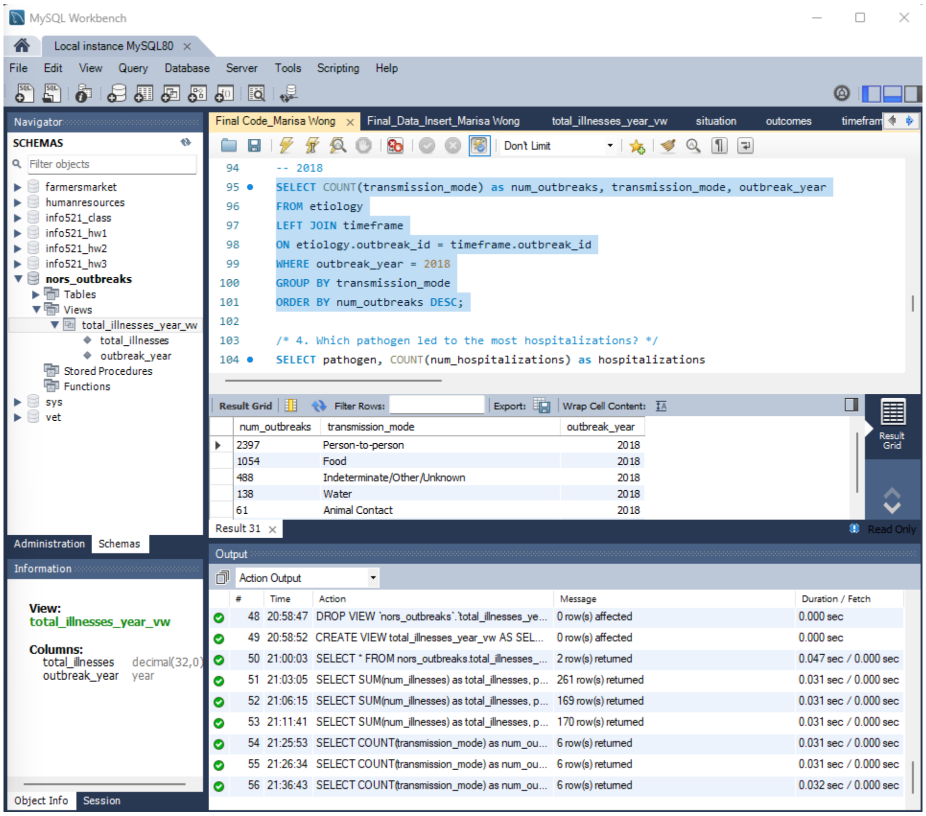

Figure 12. SQL Query for 2018 in Question 3. The highlighted text represents the query run in this screenshot.

4. Question: Create a flag for all norovirus related outbreaks from 2017-2018. How many norovirus outbreaks occurred in California that led to more than 10 illnesses?
    a. Answer: See INFO521_Final_MarisaWong_Q4_table for the etiology_flag table. 32 norovirus outbreaks occurred in California during 2017-2018 that led to more than 10 illnesses. See INFO521_Final_MarisaWong_Q4_List for the entire list of outbreaks that met that criterion.

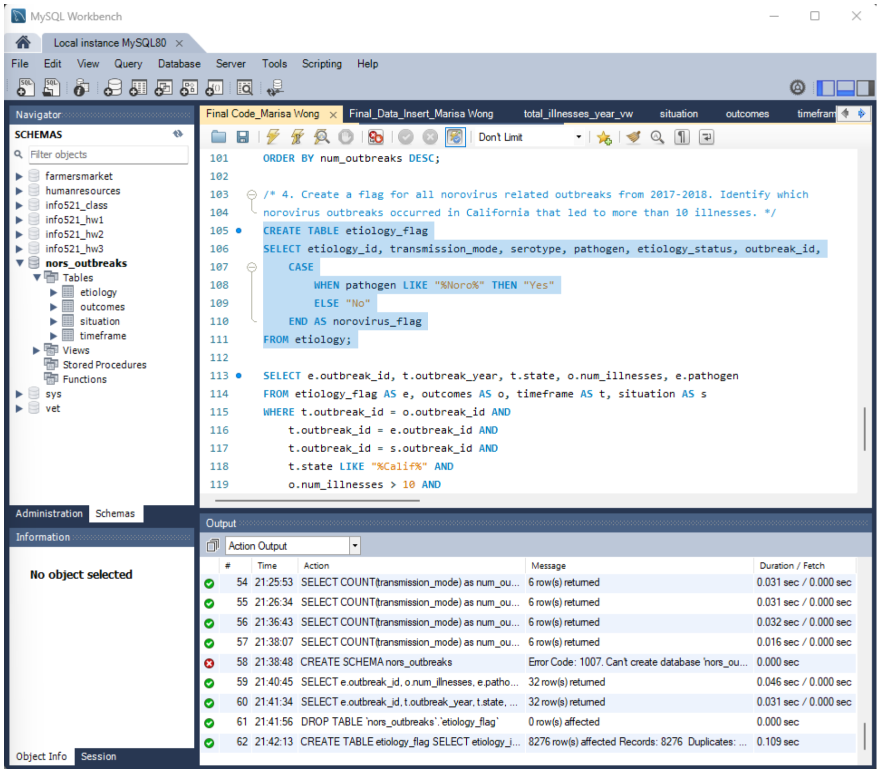

Figure 13. SQL Query for Creating a Flag in Question 4. The highlighted text represents the query run in this screenshot.

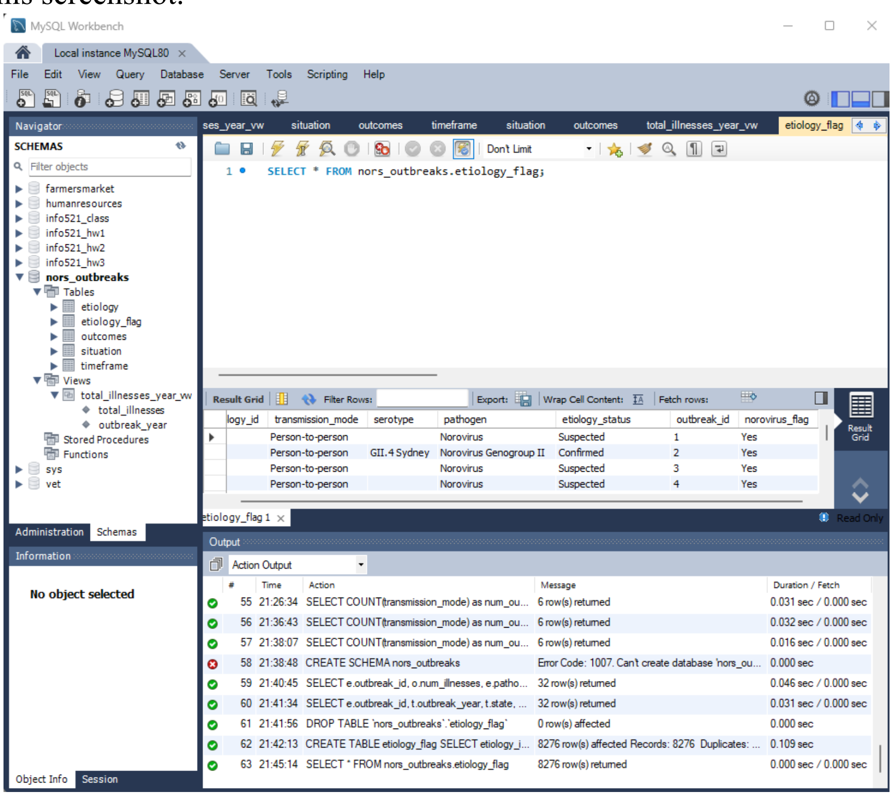

Figure 14. New Table Created called etiology_flag that contains the new norovirus_flag column.

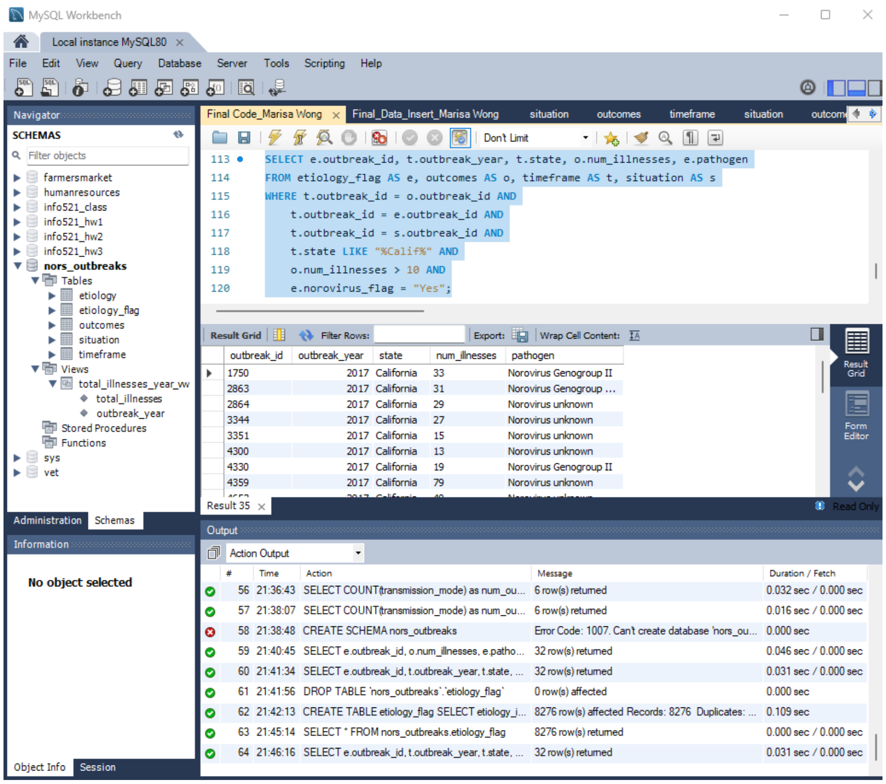

Figure 15. SQL Query for Question 4. The highlighted text represents the query run in this screenshot.

5. Question: How many unique food vehicles were associated with more hospitalizations than outbreak_id 982 in 2017? What about 2018? Create both years as a view.
    a. Answer: In 2017, 12 distinct food vehicles were associated with more hospitalizations than that of outbreak #982. In 2018, 19 distinct food vehicles were associated with more hospitalization than that of outbreak #982.

Marisa Wong
INFO 521

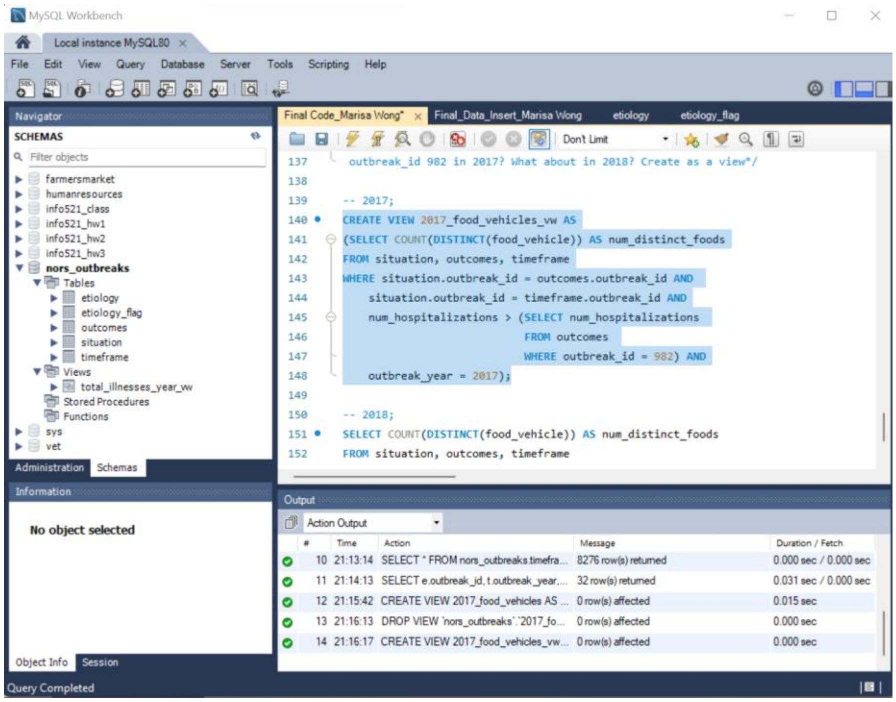

Figure 16. SQL Query for 2017 in Question 5. The highlighted text represents the query run in this screenshot.

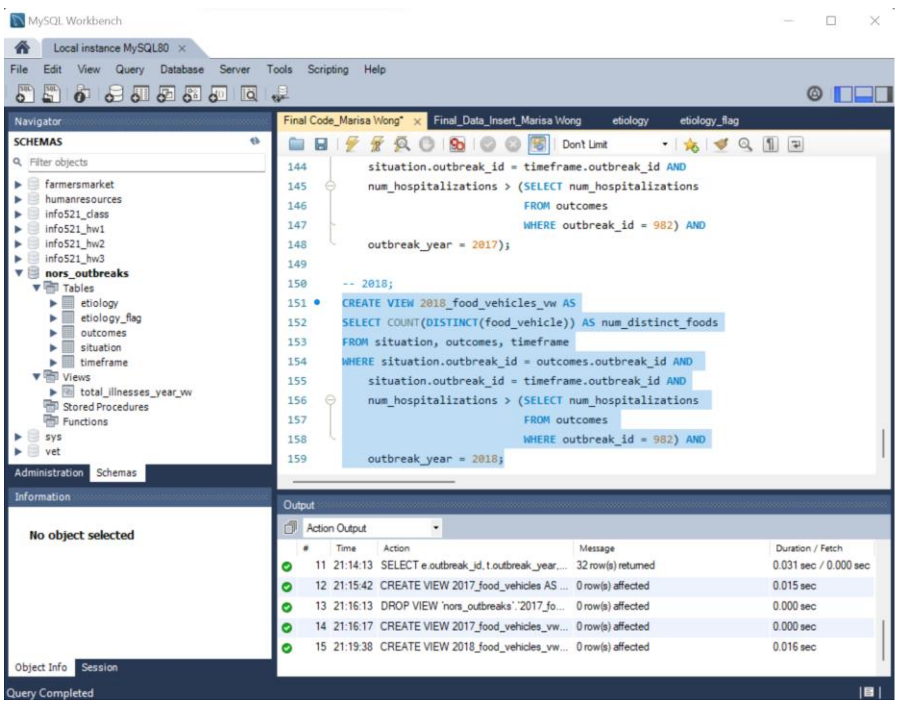

Figure 17. SQL Query for 2018 in Question 5. The highlighted text represents the query run in this screenshot.

<u>Summary of Results</u>

After performing the queries to my analytical questions that compared outcomes from foodborne disease outbreaks in 2017 to 2018, it seems that outbreaks in 2018 were slightly less severe than that of outbreaks in 2017. In 2017, 8,829 more illnesses were reported and associated with foodborne outbreaks compared to 2018. In both years, norovirus was the pathogen that caused the most illnesses; however, the specific genotype of norovirus differed in both years. In 2017, norovirus genogroup II was the pathogen with the highest number of illnesses. In 2018, norovirus unknown caused the highest number of illnesses. In both years, person-to-person transmission served as the mode that led to the most illnesses. Of the 32 outbreaks that occurred in California in 2017-2018 and led to more than 10 illnesses, 13 outbreaks occurred in 2018, and 19 outbreaks occurred in 2017. Lastly, there was a difference in the number of unique food vehicles between both years. In 2017, only 12 distinct food types led to outbreaks. On the other hand, 2018 identified 19 distinct food types that led to outbreaks.

Discussion

FDA Food Code is a guiding framework for local and state health jurisdictions to follow when implementing their own food safety policies. The food code is updated on a four-year basis to ensure the utmost scientific evidence backs the recommended food safety policies. The 2017 version of the Food Code served as a significant step in advancing food safety as it incorporated a completely new section advising for glove use over any type of finger bandage. On the other hand, NORS is a U.S. based surveillance system where local and state health departments submit information regarding foodborne, waterborne, and animal outbreaks to the CDC. CDC compiles the data acquired on NORS and makes it available to the public. Both systems work together to both inform and prevent further foodborne disease outbreaks.

My project used NORS to determine if the 2017 Food Code made any difference on the number of foodborne illness outbreaks in the U.S. The results from my analytical questions show a general decrease in number and severity of outbreaks in 2018, compared to 2017. In addition, the number of overall illnesses attributed to foodborne diseases decreased. According to FDA, norovirus is the pathogen that causes the most foodborne illnesses and outbreaks in the US[8]. My project confirmed this, as norovirus had the highest number of illnesses in both 2017 and 2018. However, the number of norovirus illnesses dropped in 2018 compared to 2017. Person-to-person transmission led to the greatest number of outbreaks in both years but led to more outbreaks in 2017. My only finding that seems to diverge from the trend of less severe and less frequent illnesses in 2018 was the fact that more distinct food types led to outbreaks during that year. Overall, 2018 seemed like a year of improvement in terms of foodborne illness outbreaks.

While my project identified a decreasing trend in frequency and intensity of foodborne illnesses in 2018 compared to 2017, true causality cannot be determined. Regulations take time to be interpreted, incorporated, and implemented in many jurisdictions, so one year from the release of FDA's Food Code might not give jurisdictions enough time to fully digest the new recommendations that led to this decrease in foodborne outbreaks. A wider gap in years may have been more appropriate to truly test the impact of the 2017 Food Code. Additionally, the category of person-to-person transmission is quite broad, and we cannot determine if the new section in the 2017 Food Code directly prevented any of the person-to-person outbreaks. Lastly, no statistical tests were conducted as part of this project, so I am unable to determine if the differences in both years are statistically significant. To continue this project in the future, I would like to gather more data on regulatory implementation timelines, access more detailed data surrounding transmission modes, and perform statistical testing on the quantitative measures in NORS. Protecting the public from foodborne illnesses remains a high priority to food safety experts. Using data to inform these practitioners of the impact of food safety interventions provides evidence for greater food protections for a healthier future.

Marisa Wong
INFO 521

References

1. Burden of Foodborne Illness: Overview | Estimates of Foodborne Illness | CDC. Published November 15, 2018. Accessed April 19, 2023. https://www.cdc.gov/foodborneburden/estimates-overview.html

2. Causes and Symptoms of Foodborne Illness - MN Dept. of Health. Accessed April 19, 2023. https://www.health.state.mn.us/diseases/foodborne/basics.html

3. Burden of Foodborne Illness: Findings | Estimates of Foodborne Illness | CDC. Published November 19, 2018. Accessed April 19, 2023. https://www.cdc.gov/foodborneburden/2011-foodborne-estimates.html

4. FDA Food Code. FDA. Published February 8, 2023. Accessed April 19, 2023. https://www.fda.gov/food/retail-food-protection/fda-food-code

5. New 2017 Food Code Section on Bandages, Finger Cots, or Finger Stalls. Accessed April 22, 2023. https://www.fda.gov/media/111463/download

6. About National Outbreak Reporting System (NORS) | CDC. Published March 13, 2023. Accessed April 19, 2023. https://www.cdc.gov/nors/about.html

7. National Outbreak Reporting System (NORS) Dashboard | CDC. Accessed April 20, 2023. https://wwwn.cdc.gov/norsdashboard/

8. FDA. Most Common Foodborne Illnesses. Accessed April 22, 2023. https://www.fda.gov/files/food/published/Most-Common-Foodborne-Illnesses-%28PDF%29.pdf