Data Analysis of Foodborne Illnesses

Christian Lussier, Nick Tocci, Dillon Thoma, Ben Watto

CS301 - Final Project Presentation

Project Overview

Elimination of Foodborne Illnesses

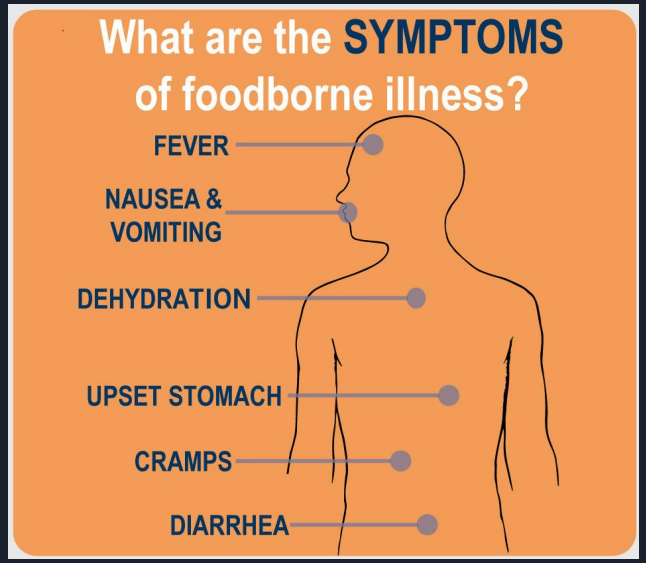
- Analyze trends of foodborne illnesses.
- Correlation between food production & illnesses?
- Have illness numbers increased?
- Used a variety of research methods.
- Policy ideas to eliminate foodborne illnesses.

Project Motivation

- Foodborne illnesses are quite common.
  - Everyone has to eat!
- Background research showed they have widespread effects.
  - Cause around 1300 deaths each year.
  - Around 56000 hospitalizations each year.
- General interest in the effects of foodborne illnesses.
- We all eat in restaurants -- foodborne illnesses originate in restaurants.
- *Scallan says, "estimates of foodborne illnesses can be used to direct food safety policy and interventions."*

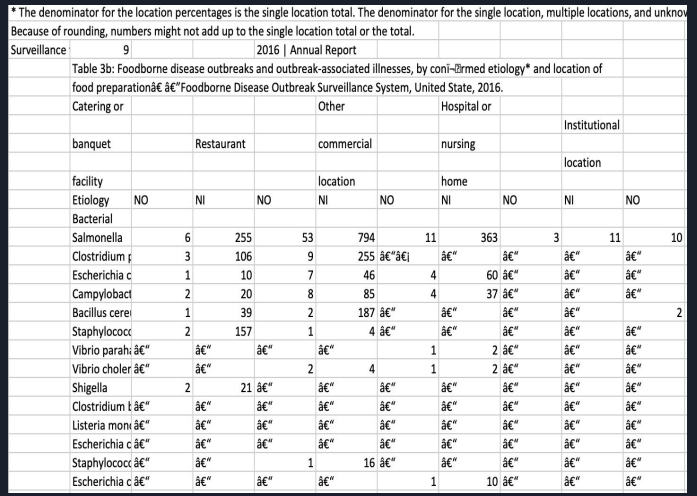
Definition

- **Foodborne Illness =** Illness caused by food contaminated with bacteria, viruses, parasites, or toxins.
- Includes:
  - Norovirus
  - Salmonella
  - Clostridium perfringens
  - Campylobacter
  - Staphylococcus aureus
  - Clostridium botulinum (botulism)
  - Listeria
  - Escherichia coli (E. coli)
  - Vibrio



Challenges

- Finding data.
  - At times it was hard to find data for specific ideas.
  - Sometimes when we found data it was strewn throughout different datasets.
    - Created extra work in combining datasets.
- Actually getting found data into a correct format.
  - PDF to CSV presented challenges.
- Choosing the best ways to analyze data in RStudio.



* The denominator for the location percentages is the single location total. The denominator for the single location, multiple locations, and unknow										
Because of rounding, numbers might not add up to the single location total or the total.										
Surveillance	9			2016 | Annual Report						
	Table 3b: Foodborne disease outbreaks and outbreak-associated illnesses, by conî–ïrmed etiology* and location of									
	food preparationâ€ â€"Foodborne Disease Outbreak Surveillance System, United State, 2016.									
	Catering or				Other		Hospital or			
									Institutional	
	banquet		Restaurant		commercial		nursing			
									location	
	facility				location		home			
	Etiology	NO	NI	NO	NI	NO	NI	NO	NI	NO
	Bacterial									
	Salmonella	6	255	53	794	11	363	3	11	10
	Clostridium p	3	106	9	255	â€"â€¡	â€"	â€"	â€"	â€"
	Escherichia c	1	10	7	46	4	60	â€"	â€"	â€"
	Campylobact	2	20	8	85	4	37	â€"	â€"	â€"
	Bacillus cere	1	39	2	187	â€"	â€"	â€"	â€"	2
	Staphylococc	2	157	1	4	â€"	â€"	â€"	â€"	â€"
	Vibrio parah	â€"	â€"	â€"	â€"	1	2	â€"	â€"	â€"
	Vibrio choler	â€"	â€"	2	4	1	2	â€"	â€"	â€"
	Shigella	2	21	â€"	â€"	â€"	â€"	â€"	â€"	â€"
	Clostridium t	â€"	â€"	â€"	â€"	â€"	â€"	â€"	â€"	â€"
	Listeria mon	â€"	â€"	â€"	â€"	â€"	â€"	â€"	â€"	â€"
	Escherichia c	â€"	â€"	â€"	â€"	â€"	â€"	â€"	â€"	â€"
	Staphylococc	â€"	â€"	1	16	â€"	â€"	â€"	â€"	â€"
	Escherichia c	â€"	â€"	â€"	â€"	1	10	â€"	â€"	â€"

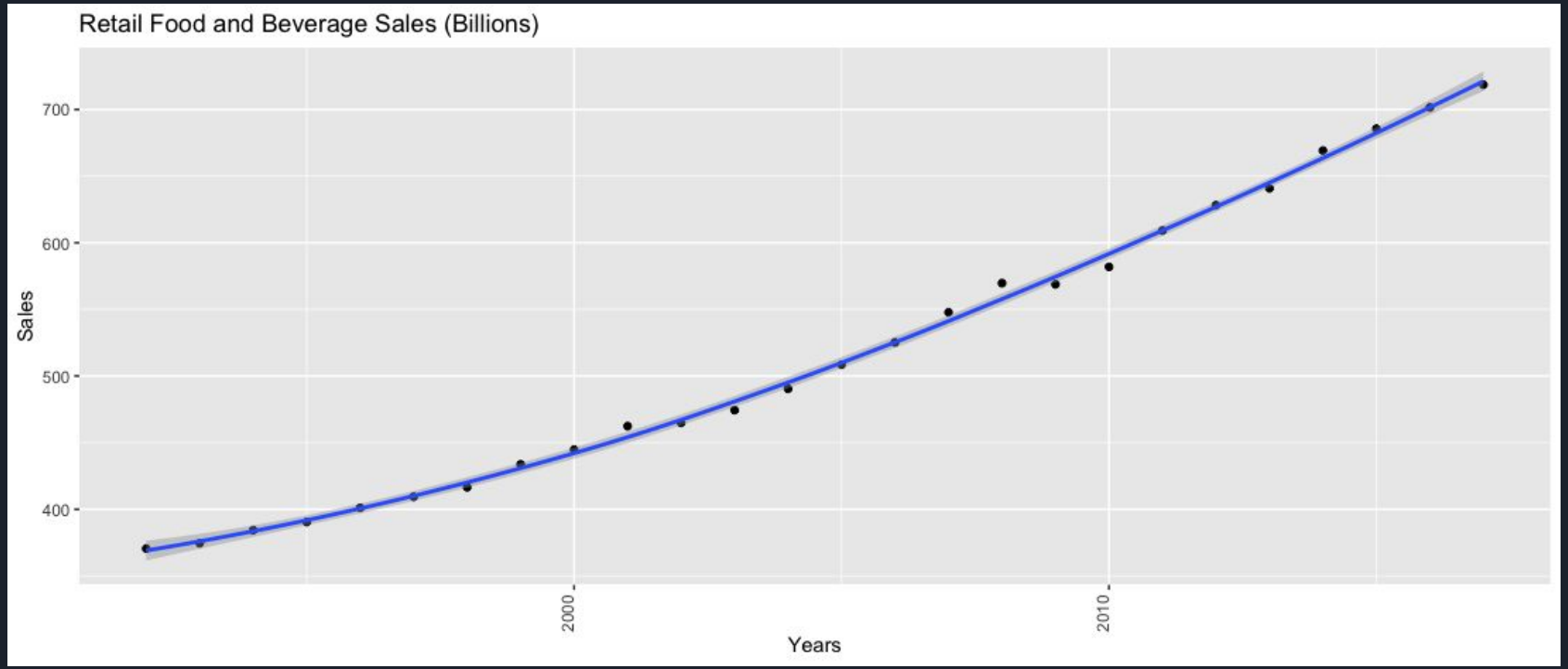
Conversion led to strange symbols (â€“) and non-data text.

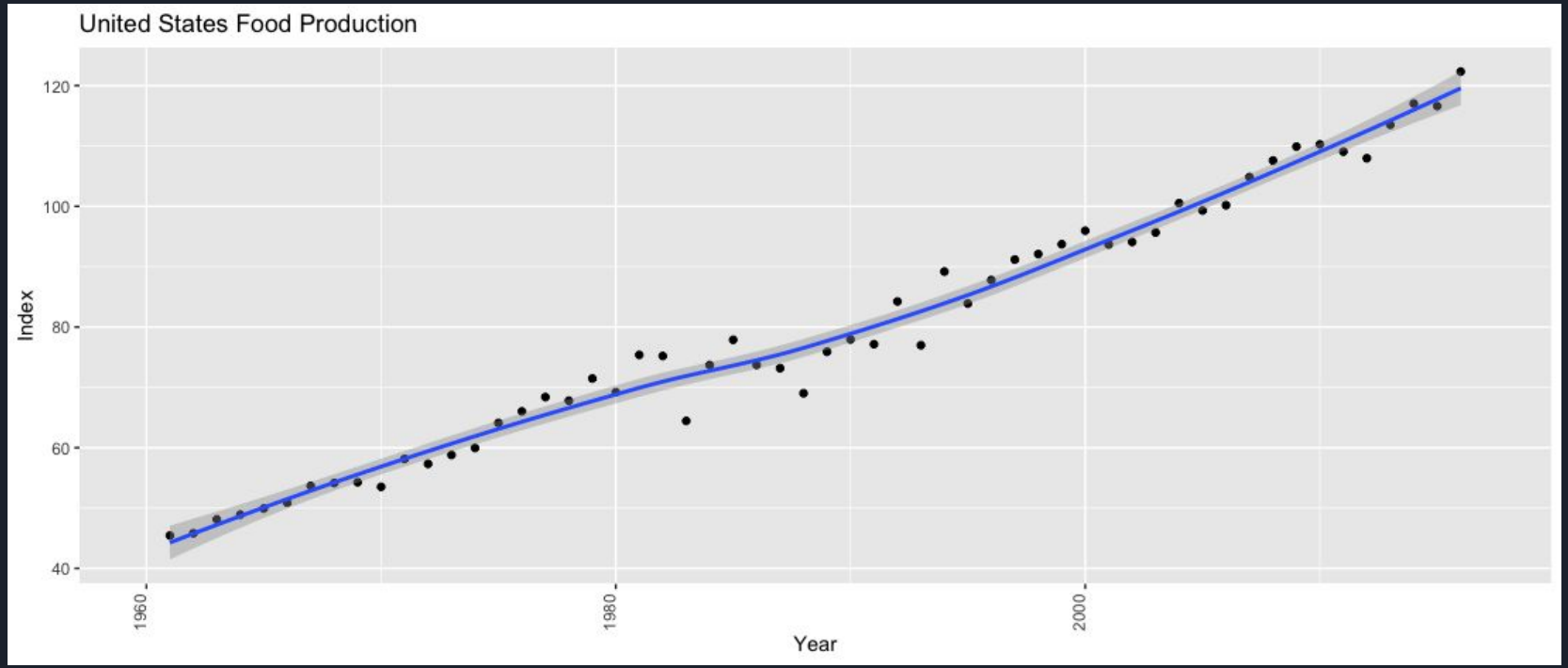
Approaches

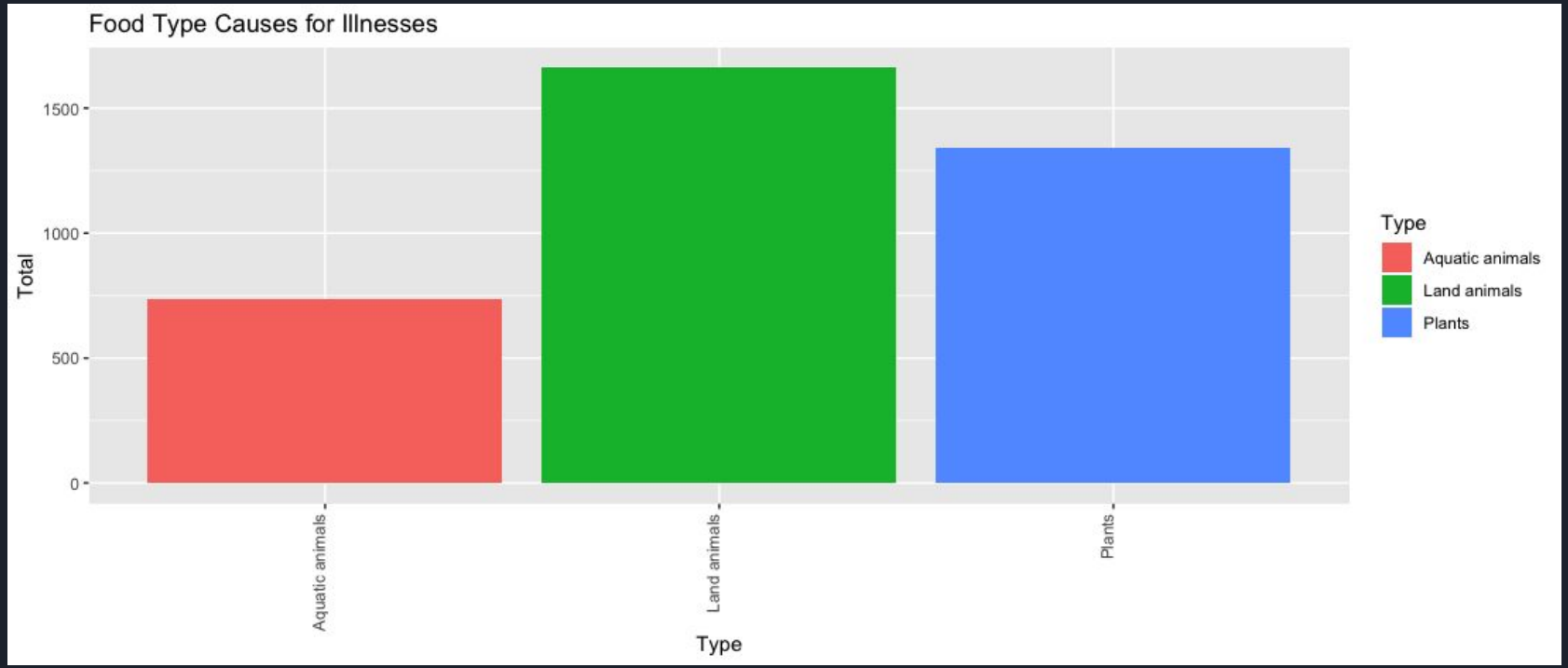
- Compare all information over the years.
- Create a breakdown of the types of illnesses and the locations that they occur in.
- Determine the main cause of contamination.
- Use RStudio to do this!

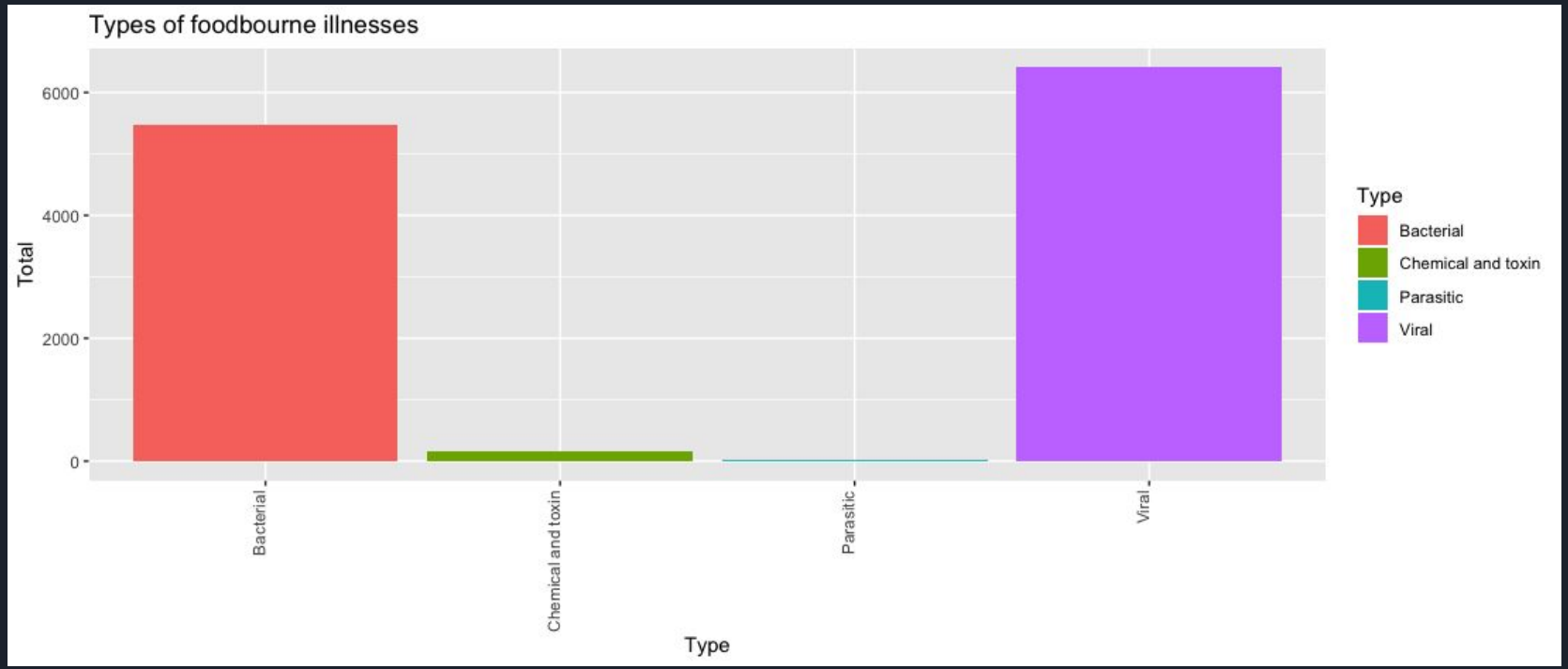
Results

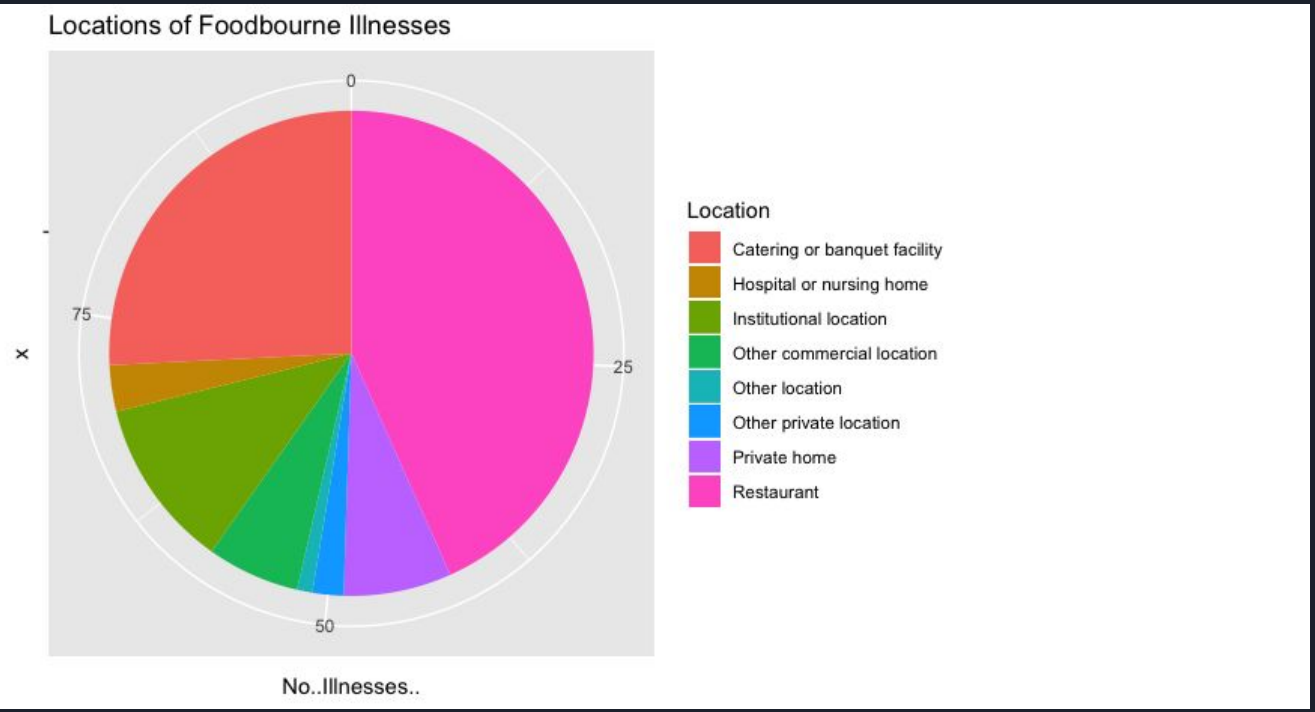
Restaurants are filthy.

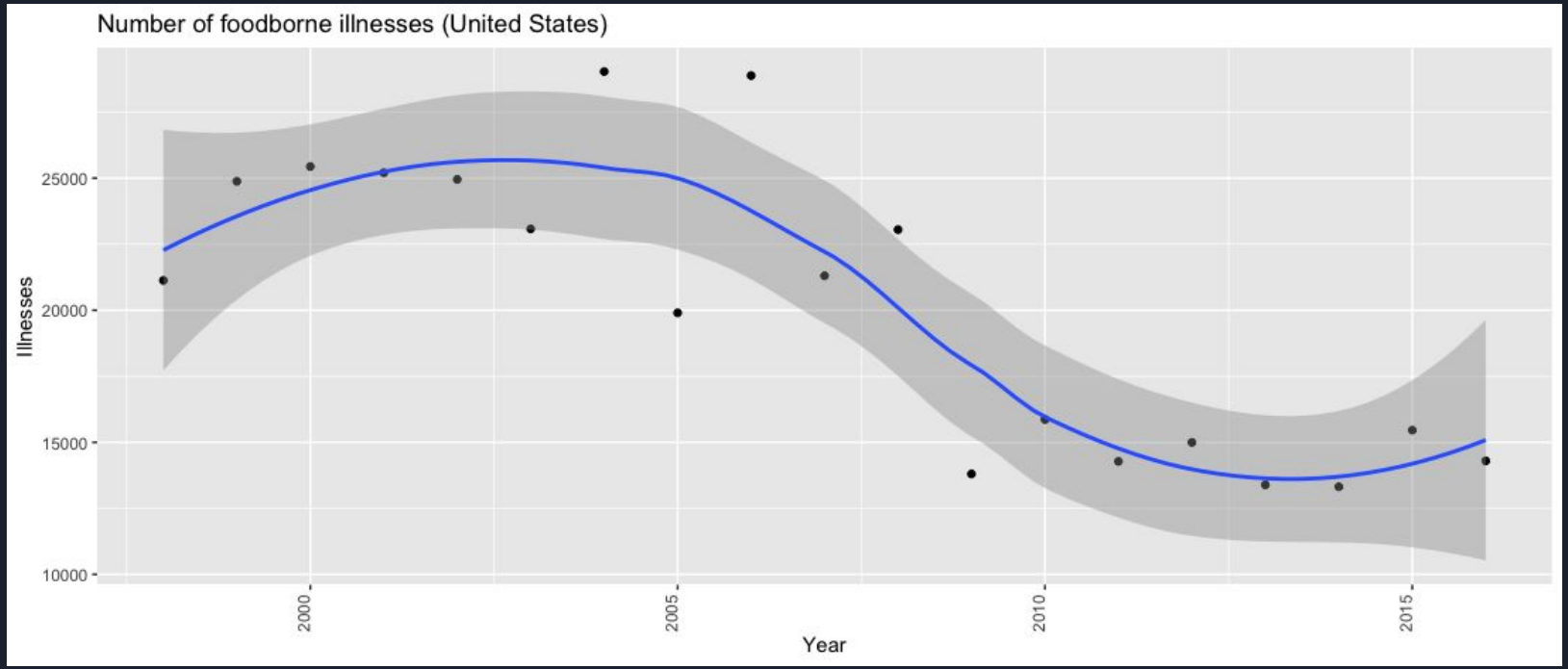

Retail Food and Beverage Sales (Billions)
700
600
500
400
Sales
2000
2010
Years

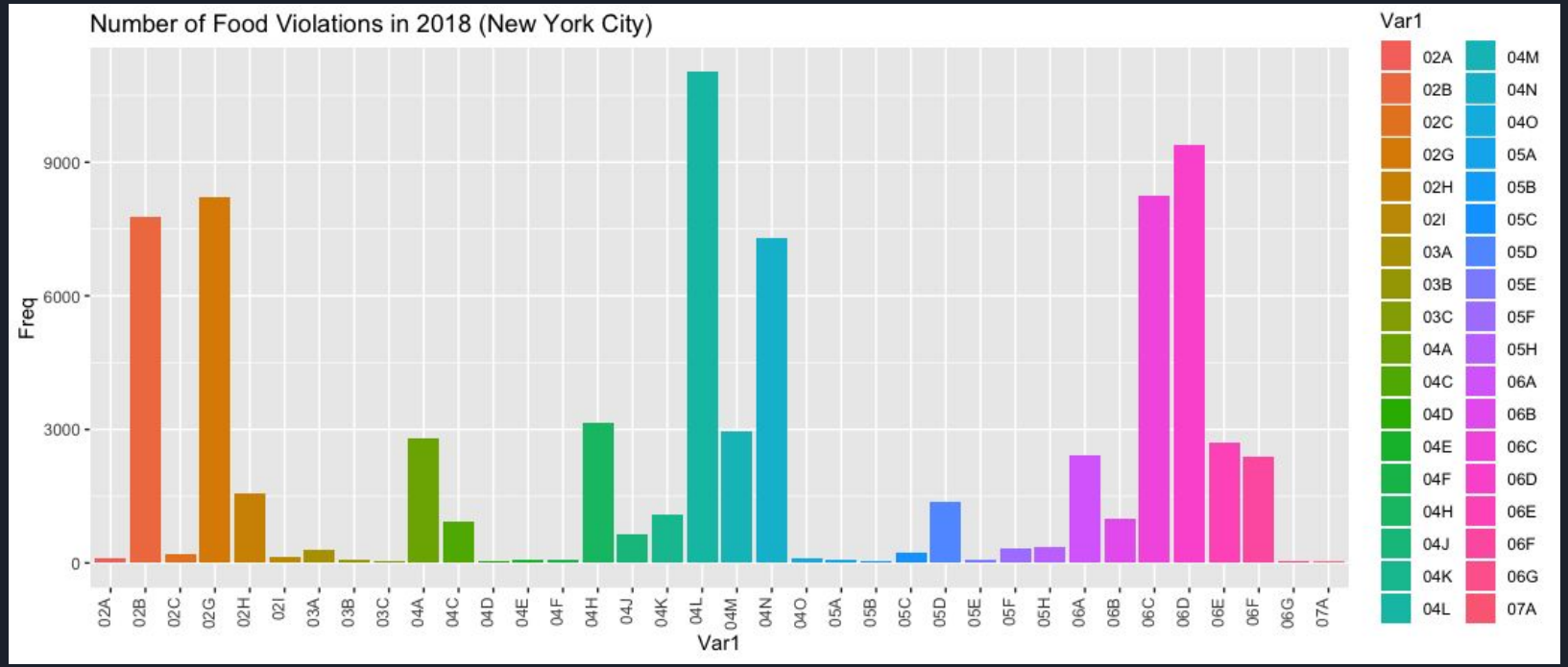

United States Food Production
Index
120
100
80
60
40
1960
1980
2000
Year


Food Type Causes for Illnesses
Total
1500
1000
500
0
Aquatic animals
Land animals
Plants
Type
Type
Aquatic animals
Land animals
Plants


Types of foodbourne illnesses
Total
6000
4000
2000
0
Bacterial
Chemical and toxin
Parasitic
Viral
Type
Type
Bacterial
Chemical and toxin
Parasitic
Viral


Locations of Foodbourne Illnesses
0
25
50
75
x
No..Illnesses..
Location
Catering or banquet facility
Hospital or nursing home
Institutional location
Other commercial location
Other location
Other private location
Private home
Restaurant


Number of foodborne illnesses (United States)
Illnesses
25000
20000
15000
10000
2000
2005
2010
2015
Year


Number of Food Violations in 2018 (New York City)
Freq
9000
6000
3000
0
02A
02B
02C
02G
02H
02I
03A
03B
03C
04A
04C
04D
04E
04F
04H
04J
04K
04L
04M
04N
04O
05A
05B
05C
05D
05E
05F
05H
06A
06B
06C
06D
06E
06F
06G
07A
Var1
Var1
02A
02B
02C
02G
02H
02I
03A
03B
03C
04A
04C
04D
04E
04F
04H
04J
04K
04L
04M
04N
04O
05A
05B
05C
05D
05E
05F
05H
06A
06B
06C
06D
06E
06F
06G
07A

Thank You!
Questions?