Pathogen Contamination Issues in the US

For the data visualization thesis project:
Bacteria Invasion: Is Your Favorite Food Contaminated?

Yicen Shi

Master of Science in Data Visualization
Parsons School of Design

Abstract

Food is essential. However, when quality control fails, we can easily put ourselves in danger by consuming pathogen contaminated food. Unfortunately, documented by Food and Drug Administration, there have been over 14,600 recall cases authorized in the United States over the past 7 years, with impacts on civilizations all over the world. What is more appalling is that the number is still growing. Historically, there have been several severe food borne illness outbreaks caused by pathogen contaminations, and the serious consequences were documented by Centers for Disease Control and Prevention.This research project delves deeper into food safety issues and investigates their effects on civilizations. In particular, it closely analyzes food recall cases that are associated with pathogen contaminations in finding patterns of distributions of questionable food sources as well as disposing the "big names" for producing these foods, serving the purpose of informing the public to be conscious of potential problems and be cautious when purchasing foods.

Introduction

Food safety concerns every individual's health. Being the primary audience considered, this project mainly focuses on creating a storyline that effective unfolds different angles of pathogen contaminations through the use of visual and interactive engagements. The entire process, from the ideation to the final implementation, is discussed in great details, uncovering the logistics behind every choices made.

The first component of this research paper is introduction to some basic information towards the food surveillance system in the US and historical regulations in ensuring food safety. It also discusses pathogens that are found in the data which are responsible for the recalls. The second component is thorough walk-through of design thinking and technical process behind the actual making of this project.

The raw data contains not only recalled products produced in the US but also global recalls that were authorized in the US which includes food imported. Additionally, besides pathogen contaminated foods, the raw data contains recall cases that are associated with other food safety issues such as mis-labelling and packaging quality. However, these aspects will not be addressed on. This project is also not focusing on unfolding the timeline of when each recall is being initiated and completed. It serves rather as an educational platform for users to learn and have basic structural grasp of food recalls.

Chapter 1

Due to the fact that this project uses data from the Food and Drug Administration, this chapter introduces some of the basic structural information of the US food system. It further discusses

foodborne illnesses and incorporated definitions from the FDA and CDC to help better understand what defines food safety and provides some general information on food recalls.

Chapter 2

This chapter discusses thoroughly around pathogens, especially the ones listed out by FDA that have caused serious outbreaks in the past. For each type of pathogens, great details are discussed in regards to the biological profile, its effect on civilizations as well as some precautions that can be taken.

Chapter 3

This chapter is where this paper starts the focus on the visualization part of this project. In this chapter, it concentrates on the design thinking and strategies behind every choice made during the making process. Topics include the storyline, layout and color and typographical choices. It guides through the thinking process from ideation to actual implementation, discussing in great detail of how elements are arranged and what tools need to be used at best making the visual communication effective and efficient. For the last part, it concentrates on the discussion of the implementation of tools and the detailed walk-through of transformation towards data in response to the design decisions. Topics include machine learning, data structures and utilizations of programming languages and libraries.

Conclusions (Chapter 4)

This final section summarizes the entire paper and revisits some of the essential messages discussed in previous sections. It also draws some conclusions in regards to general patterns observed from geospatial analysis and some statistical findings from quantitative analysis conducted prior to and after the visualization.

Future Considerations

Safety issues are not resolved if no prevention tips are introduced. This section provides tips that are recommended by organizations such as CDC on actions that can be taken in effective prevention of food consumption induced health risks. Topics include cautions that can be taken during food preparations and some general tips on how to make wiser choices when purchasing food.

Chapter 1

The Background on Food Safety

Food safety concerns have become a major issue for the American public. “Food safety encompasses a wide spectrum of issues—not only the avoidance of foodborne pathogens, chemical toxicants, and physical hazards, but also issues such as nutrition, food quality, labeling, and education.”[1]

For this project, we are only closely examining the effect of food borne pathogens on civilizations. As number of reported foodborne illness cases increases, Americans have become concerned about pathogens such as *Salmonella* and *Escherichia coli* that have potentials of causing serious health issues.

1.1 Foodborne Illnesses

When talking about food safety, it is always important to start with what food is considered “not safe”. This leads to the discussion of consequences after consuming unsafe foods. Foodborne illness is then a topic that is inevitable to be marked on.

---

[1] *Ensuring Safe Food: From Production to Consumption* (Washington, D.C.: National Academy Press, 1998), 17.

"Foodborne illnesses are infections or irritations of the gastrointestinal (GI) tract caused by food or beverages that contain harmful bacteria, parasites, viruses, or chemicals. The GI tract is a series of hollow organs joined in a long, twisting tube from the mouth to the anus. Common symptoms of foodborne illnesses include vomiting, diarrhea, abdominal pain, fever, and chills."[2]

"Foodborne illness is a common, costly—yet preventable—public health problem. CDC estimates that 1 in 6 Americans get sick from contaminated foods or beverages each year, and 3,000 die. The U.S. Department of Agriculture (USDA) estimates that foodborne illnesses cost more than $15.6 billion each year."[3]

"Researchers have identified more than 250 foodborne diseases. Most of them are infections, caused by a variety of bacteria, viruses, and parasites.Harmful toxins and chemicals also can contaminate foods and cause foodborne illness."[4]

1.2 The US Food System

When it comes to major issues like food safety, it always involves structural impact on final productions. As food inspection and quality control fails in any link in the system, small issues can easily lead to major consequences if care or actions are not taken immediately in response to them.

"The US food supply is abundant and affordable and it is acknowledge by many to pose an acceptable level of risk. The food system has evolved from one that provided consumers with minimally processed basic commodities for home meal preparation to today's availability of highly processed products that are ready-to-eat or require minimal preparation. Food preservation processes have changed dramatically from traditional salting, curing, drying, and heating. Today's products are the result of many technological developments such as pasteurization, irradiation, and genetic engineering."[5]

---

[2] "Foodborne Illnesses." National Institute of Diabetes and Digestive and Kidney Diseases. June 01, 2014. Accessed May 14, 2018. https://www.niddk.nih.gov/health-information/digestive-diseases/foodborne-illnesses.

[3] "Food Safety," Centers for Disease Control and Prevention, April 04, 2018, , accessed May 13, 2018, https://www.cdc.gov/foodsafety/cdc-and-food-safety.html.

[4] "Food Safety," Centers for Disease Control and Prevention, February 16, 2018, , accessed May 14, 2018, https://www.cdc.gov/foodsafety/foodborne-germs.html.

[5] *Ensuring Safe Food: From Production to Consumption* (Washington, D.C.: National Academy Press, 1998), 18.

Changes in food distributions also have induced changes in the way people make purchases and consume.“ The broad introduction of refrigerated railcars and trucks, freezers, and air transport created a national and now global food system”[6], which has led to the complexity of the US food system. “Americans eat fewer home-prepared meals than ever before, in response to changes in the US workforce and to developments in food processing and food service that offer greater convenience and variety in available foods”(FMI,1998b).

1.3 Food Surveillance

1.3.1 Organizations

The involvement of government agencies plays a significant role for surveillance over food quality in general. The line of duties for ensuring food safety include monitoring, surveillance, inspection, enforcement, outbreak management, research, and education.[7]
From food production to disease control, agencies and organizations of different sizes and levels are assigned with different duties. In total, at federal level, there are at least 12 agencies involved for ensuring food safety, out of which the following are majorly functioning at federal level:

- The Agricultural Marketing Service
- the Animal and Plant Health Inspection Service
- the Economic Research Service
- The Food Safety and Inspection Service
- the Grain Inspection
- Packers and Stockyards Administration of the United States Department of Agriculture
- the Centers for Disease Control and Prevention
- the Food and Drug Administration
- the National Institutes of Health of the Department of Health and Human Services
- the National Marine Fisheries Service of the Department of Commerce
- the Environmental Protection Agency

Data used for this project was adopted mainly from these two agencies that are well-known for food safety surveillance is FDA, CDC and USDA’s Food Safety and Inspection Service. These three government agencies collaborate closely at the federal level to promote food safety, alongside with state and local health departments.

---

[6] *Ensuring Safe Food: From Production to Consumption* (Washington, D.C.: National Academy Press, 1998), 19.

[7] *Ensuring Safe Food: From Production to Consumption* (Washington, D.C.: National Academy Press, 1998), 3.

1.3.1.1 CDC

CDC stands for Centers for Disease Control and Prevention. According to the official website, its role in food safety control is mostly for foodborne infections. Its duties include:

- Building state and local capacity to improve surveillance and investigation of foodborne illnesses through PulseNet, the Integrated Food Safety Centers of Excellence, and other programs.
- Working with local, state, and federal partners to investigate outbreaks, and to implement systems to better detect, stop, and prevent them.
- Using data to evaluate and revise foodborne disease prevention strategies and policies.
- Working with other countries and international agencies to improve surveillance, investigation, and prevention of foodborne infections in the United States and around the world.

1.3.1.2 FDA

FDA stands for Food and Drug Administration. It not only sees over food quality control but also drug regulations. According to official website, its responsibilities lie in the following:

- Protecting the public health by ensuring the safety, efficacy, and security of human and veterinary drugs, biological products, and medical devices; and by ensuring the safety of our nation's food supply, cosmetics, and products that emit radiation.
- Regulating the manufacturing, marketing, and distribution of tobacco products to protect the public health and to reduce tobacco use by minors.
- Advancing the public health by helping to speed innovations that make medical products more effective, safer, and more affordable and by helping the public get the accurate, science-based information they need to use medical products and foods to maintain and improve their health.

1.3.1.3 USDA

USDA’s Food Safety and Inspection Service is another government agency that primarily focuses on food safety control and inspection, especially surveillance over meat, poultry, and processed egg products. Therefore, different from FDA, USDA’s involvement is relatively narrow. It also documents recall cases, but for this project, only recall data from FDA was put to use.

1.4 FDA Recall System

1.4.1 Classifications of Recalls

For this project, in order to bring attention to the serious consequences of contaminations, classifications of recalls were incorporated to enhance the storyline. FDA classifies all recall cases by their potential threat to human health at consumption. There are three types of classifications in total, and each is assigned with a numerical designation, i.e. I, II or III.

Recalls that are labeled as classification I have the potential of causing serious adverse health consequences or deaths if consumer is exposed to these food products. Less serious than Class I products, Classification II recalls are associated with adverse health consequences that are only temporary upon consumption. The probability for serious adverse health consequences is rather remote. Classification III recalls are not likely to cause any sort of serious adverse health consequences upon exposure or consumption.[8]

1.5 Historical Regulations and Laws

It was not until the beginning of the twentieth century that regulations of food safety became a federal responsibility due to the emergence of a national market for food. As early as in 1906, Meat Inspection and the Pure Food and Drugs Act, administrated by USDA, was designed to protect consumers from infections of pathogens. It was the first national laws in the series to initiate concerns for food safety, and compared to nowadays, the implementations of the laws in the early years were quite different. The process involved a lot of on-site inspection by government officials through the uses of smell, sight, and touch.

The administrator of the 1906 Pure Food and Drugs Act, USDA Bureau of Chemistry, was renamed the Food and Drug Administration(FDA) in 1930, and later became a separate entity in the Department of Health and Human Services. It used a sampling method for inspecting food quality, which largely reduced the amount of labor involved for periodic visits to sites. The addition of the Food Additives Amendment to the Food, Drug, and Cometic Act of 1938 gave FDA the authority to approve over chemical additives to foods, while the meat and poultry inspection system still adopts the old inspection method which utilized sight, small and touch for quality testing. Then later in 1985, a new approach called the Hazard Analysis Critical Control Point (HACCP) system was adopted, and it focused on applying science-based controls at each step of the process of inspection. It later was adopted for the inspection of seafood in 1995 and was also implemented for meat and poultry in 1997.[9]

---

[8] FDA Investigations Operations Manual (Arlington, VA: Washington Business Information, 2018), Chapter 7.

[9] *Ensuring Safe Food: From Production to Consumption* (Washington, D.C.: National Academy Press, 1998), 21-23.

Chapter 2

The Pathogens

Foordborne illness are highly to be induced by the consumption of pathogen contaminated food, and such pathogens are biologically distinct from each other in their structure, transmission, target group and effects on human health, some of which are bacteria and some are viruses. Although there are more than five pathogens that can cause adverse health consequences, only the ones that have historically led to large-scale outbreaks are introduced and discussed thoroughly in this paper. The five pathogens are: Salmonella, Escherichia coli (E. coli), Norovirus, Listeria monocytogenes, Campylobacter jejuni (C. jejuni).

2.1 Bacteria

Not all bacteria are harmful to humans. Bacteria are organisms that can lead to infections of the gastrointestinal tract. Some of them are harmful to human health while others are not.

Malicious bacteria may already be present in foods when they are purchased. Raw foods have high probability of containing bacteria. Specifically, meat, poultry, fish and shellfish, eggs, unpasteurized milk and dairy products, and fresh produce often contain these malicious

bacteria. Moreover, contamination can occur in any stage of production that includes growth, harvesting, slaughter, processing, storage, and shipping.

Foods may also be contaminated with bacteria during food preparation, especially when cleanliness is not paid attention to. If preparers do not thoroughly wash their hands, kitchen utensils, cutting boards, and other kitchen surfaces that come into contact with raw foods, cross-contamination may occur.

Furthermore, bacteria can multiply if conditions are not reaching an ideal state. In particular, if hot food is not kept above 140 degrees or if cold food is not kept below 40 degrees, bacteria multiply quickly. Bacteria multiply more slowly when food is refrigerated, and freezing food can further slow or even stop the spread of bacteria. However, bacteria in refrigerated or frozen foods become active again when food is brought to room temperature. Thoroughly cooking food kills bacteria. [10]

2.1.1 Salmonella

Salmonella is one of the most common bacteria that has led to serious outbreaks in the past. CDC estimates that Salmonella causes approximately around 1.2 million illnesses, 23,000 hospitalizations, and 450 deaths in the United States every year. It is often found in uncooked or undercooked raw meats, poultry, diary products, and seafood. It can also be found in egg shells or inside eggs. The group of people who have a higher likelihood to be severely infected includes the elderly, infants, and those who have weakened immune systems in general.

People who are infected with Salmonella often find themselves develop diarrhea, fever, and abdominal pain in the first three days. The symptoms usually last 4 to 7 days, and in some occasions depending on the serious level of infection, hospitalization is needed. Most people who are infected could fully recover without hospitalization. In rare occasions, for example, when Salmonella is spread from intestines to blood steam, it then can spread to other parts of a body, which can potentially lead to death unless the person is treated immediately with antibiotics.

2.1.2 Escherichia coli (E. coli)

E. coli is another bacteria that is well-known. It has different types of different strains, and only few of which is harmful to human. E. coli O157:H7 is the strain that causes the most severe

[10] "Foodborne Illnesses," National Institute of Diabetes and Digestive and Kidney Diseases, June 01, 2014, , accessed May 17, 2018, https://www.niddk.nih.gov/health-information/digestive-diseases/foodborne-illnesses.

illness.[11] In terms of sources, it is often found in raw or uncooked hamburger, unpasteurized juices and milk. It can also be found in the intestines of human and animals.

After consumption of foods that contain these harmful E coli, people can different physical reactions. Some may only experience bloody diarrhea while others may experience urinary tract infections, respiratory illness and pneumonia. Other symptoms include stomach cramps and vomiting. People who are infected with E coli is likely to start feeling sick in 3 to 4 days after consuming questionable food. However, illness can start anytime in the first 10 days.

2.1.3 Listeria monocytogenes

Similar to E coli and Salmonella, Listeria are also present in raw or uncooked food products, ready-to-eat deli meats and diary products such as soft cheese and unpasteurized milk. According to CDC, approximately 1,600 people get listeriosis each year, and about 260 of which die. The group of people who are likely to get infected include pregnant women and their newborns, the elderly, and those with weakened immune systems.

Like others, Listeria is also hard to be discovered unless severe spread beyond the gut occurs. For pregnant women, they typically experience fever, fatigue and muscle aches. Furthermore, infection of Listeria could lead to life-threatening infections to the newborns, such as miscarriage, stillbirth and premature delivery. For other people, symptoms such as headache, stiff neck, confusion, loss of convulsions can also take place.[12]

2.1.4 Campylobacter Jejuni (C. jejuni)

Campylobacter is another type of bacteria that can cause foodborne illness, and it is often present in raw or uncooked chicken and unpasteurized milk. Drinking contaminated water can also lead to infection. According to CDC, approximately 1.3 million illnesses were caused by Campylobacter infection.

In general, people with weakened immune systems have high likelihood to be infected. Specifically, people who have blood disorders thalassemia and hypogammaglobulinemia, AIDS, or people receiving chemotherapy. In most cases, people who are infected can recover on their

[11] "Foodborne Illnesses," National Institute of Diabetes and Digestive and Kidney Diseases, June 01, 2014, , accessed May 19, 2018, https://www.niddk.nih.gov/health-information/digestive-diseases/foodborne-illnesses.

[12] "Listeria (Listeriosis)," Centers for Disease Control and Prevention, June 29, 2017, , accessed May 19, 2018, https://www.cdc.gov/listeria/symptoms.html.

own, but when infection spreads to bloodstream, it is likely to be life-threatening, which then requires medical treatment. [13]

2.1.5 Clostridium Botulinum (C. botulinum)

Clostridium is also a bacteria that can cause foodborne illness, and it has specific contamination target on canned foods and smoked and salted fish. Different from the other bacteria, C. Botulinum is a rare bacteria, yet upon consumption of contaminated food, it can cause serious illness on human by producing a toxin that attacks nerves.

Infection of C. botulinum can be caused by consuming canned food that is not canned properly, especially for those that are made at home. For manufactured canned food that is often purchased from stores, C. botulinum spores are largely eliminated during the canning process.

Also different from symptoms of other bacteria infections, infection of C.botulinum can cause temporary vision and speech impairment including double vision, drooping eyelids, slurred speech, or physical weaknesses such as dry mouth, muscle weakness, and difficulty swallowing. Symptoms can begin after 6 hours after consumption of contaminated food or as late as 10 days.

2.2 Viruses

Much smaller than bacteria, viruses are tiny capsules that often contain genetic material. Viruses can cause infections that can lead to sickness. Viruses can be transmitted through human interactions and often exit in the stool or vomit of people who are infected. People who are infected with a virus may also contaminate food and drinks that they consume, especially if they do not wash their hands thoroughly after using the bathroom.

2.2.1 Norovirus

Norovirus is one of the common foodborne viruses that can lead to serious illnesses. It is also highly contagious. In total, three are three types of Norovirus, and infection of certain one can’t prevent the person from being infected with the other two. Infections with it can be caused by drinking contaminated water or liquids and accidental contact with touching contaminated surfaces or infected people.

---

[13] "Campylobacter (Campylobacteriosis)," Centers for Disease Control and Prevention, October 02, 2017, , accessed May 19, 2018, https://www.cdc.gov/campylobacter/symptoms.html.

Infections of Norovirus could led to inflammation in intestines and stomach, which could result in nausea, stomach pain, and vomiting. Children and older adults are especially prone to get infected by Norovirus. Norovirus can be found in human feces even before the infected person starting to feel sick, and it can still be present for few more days even after the person feeling better. Symptoms include diarrhea, nausea, stomach pain, fever, headache, etc..

The symptoms can be mitigated by drinking lots of water that does not contain alcohol or caffeine due to the fact that infection of Norovirus can lead to severe dehydration after the person throwing up and experiencing diarrhea multiple times a day.[14]

[14] ”Symptoms," Centers for Disease Control and Prevention, April 12, 2012, , accessed May 19, 2018, https://www.cdc.gov/norovirus/about/symptoms.html.

Chapter 3

The Visualization

Given the systematic overview of the current situation within the food safety in the United States, a large picture of food safety is drawn. For the visualization part of this project, a narrative or a storyline is needed to be developed based on the observations of data as well as the background information on the food system as a whole. Therefore, the next section mainly deals with the ideations for this project, including the thinking process and the logistics behind the choices made in color, layout, and typography.

The visualization of this project investigates the severity of pathogen contamination has been historically, looking closely at the food recall cases that were documented by FDA from the year 2011 to 2017 as well as food borne illness cases reported by individuals due to bacteria outbreaks from the year 2011 to 2016. The process of this visualization mainly consists of three parts—data wrangling, tool chaining for the visualization and design solutions for effective communication with viewers, and each part is indispensable in the contribution to the entirety of

the final visualization. Specific details in regards to the methodology, the tools and libraries utilized, and the design thinking are shown in the following sections.

3.1 The Storyline

Every dataset has great potentials for the development of different stories or threads based on focus of analysis. A lot of the times, The storyline for this project begins with the scenario that bacteria is taking over the planet and invading the kingdom of food, and a question is therefore imposed to ask viewers if their favorite food is contaminated. Then, based on personal interest, they enter the project and gain insights. Knowledge is best learned out of interest, and in most cases, passive learning is not effective for comprehension of a concept. Therefore, an invitation is extended to viewers (based on their willingness) to participate in exploring food safety which is a topic that have profound impact on all people.

There are three major perspectives taken to approach the issue of bacteria contamination: the food, the bacteria and the consequences. As an entry point, different types of foods, which are the main characters in the story, are laid out, and number of recalls corresponding to each type is displayed once clicked on. The antagonist in this story is then the six bacteria that are causing chaos. Many of these can lead to serious health issues that are in most cases, detrimental, and some can only cause slight discomfort, which corresponds to the classifications of these recall cases, as explained in detail in the sections above.

3.2 The Design Thinking

3.2.1 The Title: Bacteria Invasion: Is Your Favorite Food Contaminated?

First impression is important in establishing a profile, and triggering emotional responses often would make certain experience memorable. In this case, fear of death or serious health issues plays an essential role in grabbing viewers' attention at the entrance of this visualization. Bacteria is invading the safety of foods. They are then alarmed, and the next question is how to make a personal connection. People always have preferences when it comes to things that are essential like food, and therefore a great way to make connection is to give freedom to allow viewers to choose their favorites, which helps significantly with increasing the level of their involvement.

Questions invites answers. Once the question "is your favorite food contaminated" is asked, and naturally people start searching for answers, which leads to the following sections. The transition of an introduction to actual content is then naturally made.

3.2.2 Icons

The next section is selection of foods. Instead of using only text, icons were used with complimentary text for further clarification. All icons selected for the visualization have round edges and thick lines, which is visually inviting to naked eyes. Each icon is color coded for displaying corresponding information on a chart, and all the colors selected are pastels which have strong contrast against a dark background, making elements stand out.

3.2.3 The Donut Chart

Information itself has its unique hierarchical structures. Even within the same dataset, if approached with different angles, the presentations for the information should vary. The first component for this project was to visualize number of recalls in association with particular food groups. To provide a general sense of the portion of effect in regards to each group, there are two types of relations that are needed to be put to consideration. One is its relation with the total number of recalls, and the other one is its comparative relation with other food groups. A donut chart would be a perfect choice for displaying partitions in relation to the whole, and additionally, because of its circular shape, it disperses content around a center point, which grabs viewers' attention more easily in comparison with other types of graphs. To show the differences among partitions, each section, in other words, each food group, is color coded, and the color used corresponds to the color used for the icon. The display of all partitions with colors help viewers digest information at one glance.

3.2.4 The Qualitative Info

Besides food groups, another layer of information is bacteria. After basic information is shown, viewers have a high tendency of wanting to know more in detail about the bacteria that is affecting the food they are eating. Therefore, in this section, a basic biological profile and its background information are given for each type of food.

3.2.5 The Map

Different from a quantitative analysis in which numbers are often represented by lengths, a map allows viewers to see spatially where qualitative data points are located. Each recall case is marked up on the map with a light color to create drastic contrasts. Naturally, certain areas would be pin-pointed with high density, which can imply that these areas have potential issues such as corruption. Additionally, classifications of recalls were displayed, showing which areas produced most problematic food. Besides a broad view, a drilled down layer of information

containing details such as units being recalled and companies is also available for viewers to explore. They can hover over individual mark-ups and related content will then be displayed on a side banner.

To comply with the dark color scheme, the map was styled using a dark grey background with continents shaded in dark purple and state boundaries colored with teal. The text labelings stand out due to the use of a lighter purple and outlined with a dark blue. The map is centered at the United States since my project is only visualizing food that was produced in the US, and states boundaries were shaded in a light blue, making the contours of each state stand out for a clear view of the spread of markups in relation to states.

3.2.6 Color Scheme

The background for this project is all dark navy, with certain sections highlighted with a lighter shade. All the elements sitting on top, especially the main features, are using much more vibrant colors so that they stand out when viewed from a distance.

The topic of this project itself is scandalous, and a dark color theme would enhance the concept of secrecy and darkness throughout the sections. This gives the stage to the elements sitting on top of the background. Specifically, all features that ties directly with food groups are using light colors that have pastel hues. The choice was made due to the fact that the concept of food should be pleasing to eyes, or in other words, visually stimulating, and the use of pastel colors create a creamy and delicious sensation that is inviting to users. In the part where recall classifications are introduced, it demonstrates the number of recalls that are detrimental which in nature is a serious and alarming concept, and therefore the color red is used in this section to signify danger while all other pastels are toned down to grey.

Additionally, given the topic is approached from three distinct perspectives—bacteria, food and production sources, key words that in relation to each perspective is paired with different colors. For example, for the title “Bacteria Invasion: Is Your Favorite Food Contaminated?”, “Bacteria Invasion” is shaded with a teal color whereas the next question that is leading to the main composition is filled with red. The logic was carried throughout the project. In the section where pathogens are introduced with intricate details, the title is filled with teal. The points marked up on the map are initially colored with pink, and when they are hovered over, pink turns teal, meaning the food produced at that particular location has been contaminated with some pathogen which is represented by teal.

3.2.7 Typography

There are mainly two type faces chosen for this project. One is Righteous Cursive, and the other one is PT Sans. Righteous is a font that is not strictly serif or sans serif. Its roundness is balanced with the accompany of slanted edges, adding some playfulness to the picture. This is

a great choice for the title “Bacteria Invasion” in that the topic has certain seriousness yet approached with a relatively informal or playful angle. The font PT Sans also has certain roundness to it, yet it is generally thinner and taller than Righteous, which makes it more readable for smaller size and low resolution screens. It is then more appropriate for the body text where legibility is the most essential.

3.2.8 Layout

The layout for this visualization follows a vertical placement of components. The title “Bacteria Invasion: Is Your Favorite Food Contaminated?” is placed in the center when user enter the project. By clicking on a down arrow, they are redirected to the next component where selections of food are laid out. The donut chart is placed under the food categories and becomes the focus of the page with few navigational and explanatory features listed out on the two sides. By clicking on the text on the left hand side, users will be able to see an extra layer of information, and when they click on the two texts on the right hand sides, they will be directed to either the pathogen profile page or the map that displays all the recall cases. The position of these two components will stay the same, when being clicked on, they are layered on top of the other two components that will be hidden. Therefore, generally speaking, there are only two positional changes during the entire exploration of this project.

3.2.9 Interactivity

Interactivity is one of the major functions in enhancing the storyline through prioritizing certain elements and shifting focus of attention. The first layer of interactivity is for navigation through the project, allowing one section led to the next by clicking on certain tabs or texts. The second layer is for shifting focus, which is also the more important part in delivering messages. In the first section of this visualization, interactivity is carried out by allowing users clicking on food categories to retrieve information based on their choice, which largely helps trigger interests and establish emotional connections to the project itself. When clicking on the consequences of food recalls, users shift their focus from food and approach the project from a different perspective. Just like the food selection, in the map section, all recall cases are marked up. They are then given the choice of hovering over any one of these cases to find out more information. After all, the benefit of incorporating interactivity is that it gives users the freedom to explore, which helps their learning of the data, making the entire experience more memorable.

3.3 Methodology

To fully visualize the concepts, raw data gathered need to be re-structured and wrangled in order to meet the structural requirement that otherwise would not need. Appropriate libraries and tools are also selected based on the design. Hence, this section mainly discusses the back-end process of realizing the ideas talked about in the previous section. Specifically, it focuses on the data treatment, tool selections and actual implementation of all aspects.

3.3.1 Data Gathering

For this project, three major data sets are used respectively in regards to three analogies: geospatial, food recalls, bacteria outbreaks. The first being a GeoJSON that contains geospatial features that are needed for mapping and markups for locations. The second is the food recall enforcement data pool that was discovered on FDA's open data channel. There have been over 14,666 cases in total, recorded so far from 2011 to the most recent. In each case, details such as classification, product description, location, distribution pattern and units of recall are documented. The last dataset that is supplementary to the narrative is the bacteria outbreaks that caused food borne illnesses documented by CDC from 2011 to 2016.

3.3.2 Data Types

The recall data set is meta data which contains, by the time of its discovery for this project, 14,666 objects which represent 14,666 recall cases in both the United States and foreign countries such as Australia and Canada. The GeoJSON contains all polygons and paths that are needed for mapping and markups. Additionally, since only US food recalled cases are used The outbreak dataset is a csv file which contains various columns of information such as year, type of virus, location and symptoms.

3.3.3 Data Wrangling

There are two main components of the wrangling. The first task is to identify clusters of food groups within the product descriptions which are all texts that have no clear patterns or entered to the database following certain format. Method of unsupervised machine learning is utilized in this case, specifically, a learning model of K-means in finding clusters in a large sample like this one. First of all, a separate column called food group is created for assigning food category associated with each recall case. The next step is to identify key words in clusters created using K-means and then assign the key word to specific recall case under food group. At last, a new dataset is generated with a new feature and then exported for further analysis.

The next task is to filter the meta data to extract only the cases needed for this particular analysis. Since the original dataset contains not only the food sources in the US but also from foreign soils, cases that do have the feature country as “United States” are filtered out. Additionally, due to the fact that the main interest for this project is bacteria contaminated food, only cases that are associated with bacteria contamination are needed. Therefore, key words search for the major six infamous bacteria are implemented.

In regards to geospatial analysis, two major transformations towards the dataset are implemented. One is data join of GeoJSON with the other two datasets so that corresponding information such as recalling firm and units can be extracted efficiently and displayed properly. Another is to convert text based location entries to latitude and longitude. In realizing that, google geocoding API is utilized. For instance, the very first instance in the dataset has a street address “11455 Moorage Way”, a city name “La Conner” and a state name “WA”. The first step in geocoding is to connect the three parts, separated by commas, for a complete address. The end result is “11455 Moorage Way, La Conner, WA”. Next is to convert the complete address into an url that can be used for requesting transformation from Google using an APIkey generated for this particular project, and an object containing the latitude and longitude is then appended to each recall case.

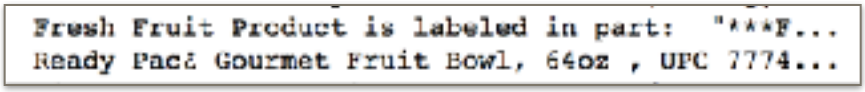
3.3.4 Machine Learning

Machine learning is a process where a large dataset is fed in to a computer program, and based on principles of statistics, machines learn the dataset and develop pattern recognitions and outputs. The reason for incorporating machine learning for this project mainly come from the design thinking. For the purpose of triggering emotional attachments and connections with viewers, the entering to this visualization begins with selections of their personal interests. Specifically, options of twelve types of foods are given so that viewers can select one or more foods they favor to initiate the exploration of information. Such arrangement requires that the dataset provides a classification for each recall case based on the product. However, given the fact that the raw dataset contains only rather random text descriptions for the products being recalled with no noticeable patterns or consistencies, manual labeling of food category is then extremely cumbersome for a large dataset like this one. Hence, in order to realize the design strategy, machine learning is utilized for the sole purpose of reducing the amount of extensive labor that is otherwise needed.

In terms of detailed steps taken, there are mainly three parts. The very first, other than loading the raw data into the program, is observation and exploration of the original product descriptions to obtain a general sense of how the data is structured, and then the creation of another column of food group for later assignment is made. The next is the actual analysis, and the method used is called k-means clustering. It aims to partition n observations into k clusters, and within each cluster, these observations share certain common characteristics or features. In its application for this project, the number of observation is then the total number of data points used for the analysis,, and k is ideally the number of food groups.

In the real practice, a start with k=10 was explored at, followed with expansions of cluster size. At the end, the decision of k=50 was made, which is a lot more than the actual number of

clusters needed. That is due to the fact that machine automatically categorizes two product entries that belong to the same food group to different clusters because of the large difference in format. For example, as shown below, two descriptions for two fruit products are supposed to be clustered



to the group Fruit. However, due to the inconsistency of their format, these two are separated into two different clusters. If other types of product follow the same pattern of entry as one of these shown above, there is a high likelihood that they are group together with one of these two. Therefore, in reality, a lot more clusters are generated to ensure there are not too many misclassified products. To double ensure accurate classifications, manual checks in each clusters were implemented to ensure no product is misclassified. Upon inspection on each cluster, all the products within that cluster are assigned to a food group, and for those that are misclassified, they are re-assigned to the correct one.

3.3.5 Data Structure

There are two versions of the dataset used for this project. For the first component where a food groups are displayed, as mentioned in the above section, assignment of a food group to each recall case was implemented, and one version containing food groups was used for that particular component. The other version includes geocoding of recalling firm locations, which was utilized for the map part where markups of locations require collections of latitude and longitude.

The very first step in the process of making this project, besides filtering through the dataset, was altering the data structure to be compatible with the visual design.The discussion must start with each component separately.

In the first section, data being used gets updated every time when a food group is selected. Hence, first of all, generation of a new dataset based on selection must be made, and then the dataset is inserted as an input to a function to fill out the arc.

This requires the change of the structure. Originally, the data points are entered based on cases. However, in order to make the creation of arcs responsive to selection of food groups, the original data had to be re-organized by food groups. For example, cases that are recalling meats are then grouped together in an array. For each food group, the filling of color is based on the number of recall cases, which is the length of an array for a food group , compared to the total number of cases, and so an array would be created from the inserted dataset containing the percentages. Since two arcs are created, one showing the number of cases for each food group and the other one showing the potential consequences for the recalls, two arrays in which data is extracted from the original dataset are needed for filling the arcs.

Each arc starts with an angle and ends with an angle. To create arcs that are generated by numbers from data, conversions are necessary to be made. Therefore the last step for drawing the arcs is to convert the numbers in each array to start angles and end angles. Furthermore, since the order of arcs depends on the order of selection, the whole arrays are updated every time when some food group is selected, and thereby the start angles and end angles change accordingly due to the change of numbers, hence the length of arcs.

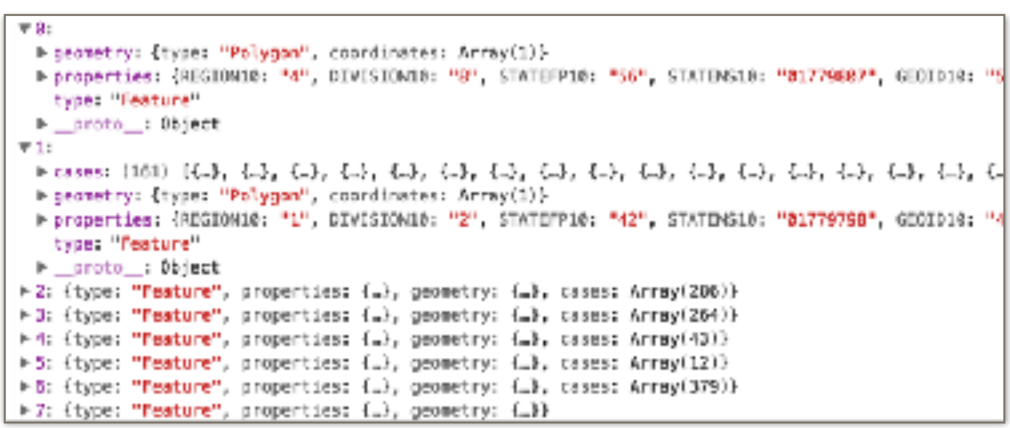
Besides the shape, a color is assigned to each arc. This requires an array of colors inserted to a function for assignments, and the array gets updated according to the change of selection. For the first layer of arcs, arcs are filled in response to the color of the icon for a particular food group. For instance, when “Vegetables” is selected, the corresponding arc is then filled by green since the icon for “Vegetables” is green. Therefore, in order to connect the correct color with an arc, a color is assigned to the original dataset for each case within that particular food category. In other words, each case is assigned with a color, and for the cases within the same category, colors are the same. A new array is then generated containing the colors in subject to arcs.

Additionally, to provide a clear view and a quantitative sense of the amount each arc represents, text labels are displayed at the center of the donut chart. This also requires an array containing all the actual number of cases rather than a percentage for each selection of food group.

Therefore, for arcs in response to food selections, a total number of three arrays were created, and the same transformation method applies to the arcs that relates to classifications of recalls. The only difference is that, since there are three types of classes within each type of food, each array contains triple times the items needed for just a food group.

All of the arrays are wrapped in a an array called “dataForUse”, meaning that it is a new dataset where all data points are either extracted and then re-organized from the original or converted based on the requirements of filling an arc.

The other section that requires re-structuring of the data is the map. Different from a donut chart, as mentioned in the previous sections, a map requires a GeoJSON which contains the tile set needed for constructing the base map. Additionally, since markups of locations are utilized for the visualization, the dataset containing latitudes and longitudes and the GeoJSON were merged into one dataset. As features in GeoJSON are all organized by states, the first step taken was to match the states in the dataset and append all the cases in a particular state to the features of that state in GeoJSON. If no cases were authorized in a state, then the object for that state will not contain a column of cases.



The markup of locations requires transforms of the dataset, specifically, extraction and separation of information. Since the primary information needed for markups on a map is collection of coordinates, an array containing just latitudes and longitudes was created. Additionally, positional switch of latitude and longitude, which complies with Mapbox a library for mapping. Besides the display of the spread of locations, users have the tendency to want to know about details in regards to each markup. Therefore, another layer of information needed for the mapping is collection of details such as recalling firm and units being recalled. By looping through cases within each state, a new array is generated for each case. In each array, latitude and longitude will be inserted to a column while other detail information is appended to another column called “info”. When users hovering over markups, corresponding information is then extracted from “info” and displayed on a sidebar next to the map.

3.4 Tools and Libraries

For projects that are constructed digitally like this one, tools mainly consist of two parts: programming languages and libraries. In terms of interface design and visual arrangements, HTML and CSS are the main platforms used for front-end development, while for the back-end processing of data transformations and manipulations, Javascript and Python were used. Additionally, Javascript also served the purpose of connecting the front-end and back-end.

3.4.1 HTML and CSS

HTML provides the basic structural development for any web based products. It dissects a webpage and divides it into various components. For this visualization, as mentioned in the previous sections, there are four components—the title, the donut chart, the information of pathogens and the map. Naturally, four divs were created, and within each four of the divs, subsequent divs were also created for different layout arrangements. For example, for the donut chart, a selection of food icons takes up a whole div and the actual donut chart is placed in another div under icons. Also in the map section, the map itself is wrapped in one div including markups and a base map, and all additional information in relation to each markup is displayed on a sidebar next to the map, which is another div.

The styling of the visualization was realized in two ways, and all elements that are structured using HTML are directly styled using CSS. The styling ranges from structural placement such as positions and visibility to visual enhancements such as colors and sizes. For the positioning of divs, all primary components are absolute, meaning that each of them is placed exactly at certain position and is only subject to the title. For sizes and colors, the choices made followed the design strategy which is discussed in the previous sections. Instead of externalizing the

CSS styling sheet, it was imbedded within the HTML file and inserted in the head section where links and CDNs are also noted.

3.4.2 JavaScript

JavaScript is utilized to load the datasets and create functions that allow data to be processed and then projected to the web. In other words, it is both a developmental tool and a middle ground which connects HTML and additional manipulations through other libraries such as D3.

In regards to the manipulations of the raw data, the process mainly took place at the back end, and so such processing were not inserted to HTML. The transformed versions were then loaded to the script inserted in the body section of HTML. In the script, within each of the three major components, dataset is loaded and then inserted to a function which visualizes the data. In each of the functions, the first part is where a complete dataset is transformed. For instance, in the donut chart section, data points were extracted from the original and then arrays were created. The following part involves the integration of libraries such as D3 and Mapbox GL which will be discussed in greater detail in later sections.

3.4.3 Python

Since most of the data wrangling and re-structuring was processed using JavaScript, python was used only for machine learning. Python generally follows different semantic rules from Javascript. However, the order of steps taken is similar. For example, at the beginning, the dataset is loaded into the system, and at the end data is exported to a folder where other files that relates to the project reside. The difference takes place in the body section in that Javascript is used dealing with filtering and re-structuring of a dataset whereas Python is great for other types of data wrangling such as machine learning. Also, due to the fact that many of packages used, for instance, scikit learn, for machine learning is primarily designed for Python, it is then ideal to use Python instead of Javascript even though both outputs data that maintains the same format.

3.4.4 D3

D3 is a library designed for displaying infographics based on data, and the full name is Data-Driven Document. D3 library includes various types of geometric representations of data. For this project, a donut chart is chosen in that it is appropriate for displaying proportional relationships among data ,which is discussed in detail in previous sections. In terms of actual implementation of the library, it has to follow the logistics and semantic rules from the D3 API.

D3 also has instructions for mapping and markups, which is used for this project.

3.4.5 Mapbox GL

Mapbox GL is another Javascript Library that is developed specially for geospatial analysis. It provides services that range from map styling, geocoding to location markups. For this visualization, the base map is created and styled using the Mapbox API. The geocoding was accomplished using Google Geocode API which was mentioned earlier. The styling choices were also explained in the previous section.

Conclusions

Conclusions from Analysis

With dataset that is as large as the one from FDA, there are many potential insights that can be retrieved, and many of which could be contradictory with intuitive reading of the data. For this project, two types of analysis were performed. One is quantitative analysis which is mostly performed back-end, and the other one is geospatial analysis which is realized on a map. These two different types of analysis offered two different perspectives. A quantitive analysis helps understand the size of effects whereas a geospatial analysis unfolds the distribution of data points.

4.1 Quantitative Analysis

From filtering through the raw data, we gain much better understanding of the distribution of all recall cases. While most of the unsafe food were produced in the US, there are roughly 100 cases for which food was imported from a foreign source. Even for the productions that were initiated in the US, many of which were distributed to foreign continents.

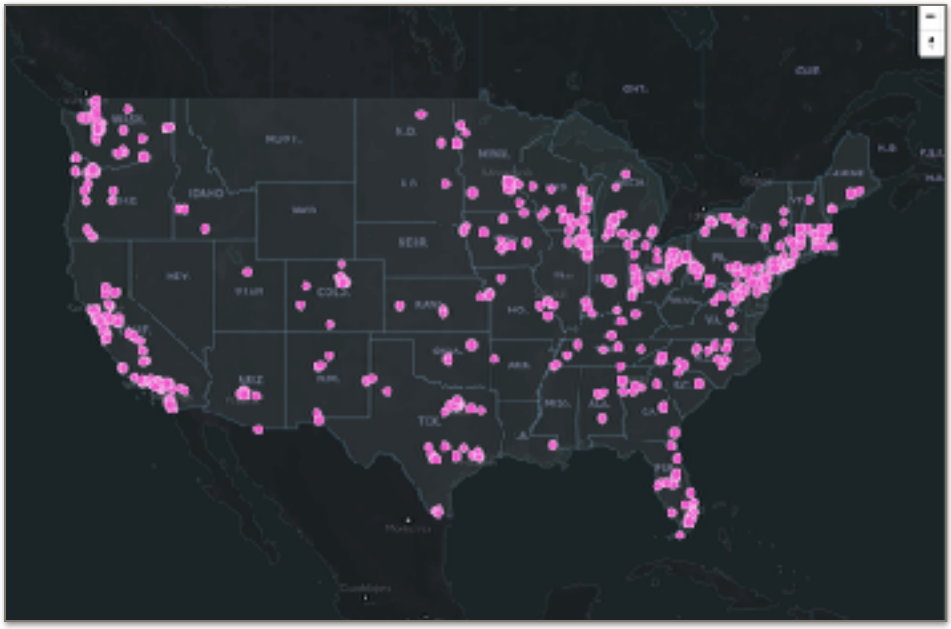
For parts that are visualized for this project, especially in association with the donut chart, number of recalls for each of the 12 food groups is displayed and compare with others by being encapsulated in a full circle. In summary, Ice Creams (828 recalls) and Condiments (732 recalls)

are the two groups that have been recalled the most, collectively over 1500 recalls in the past 7 years. Being the least type recalled, flour was only recalled 13 times in total in comparison with others. Vegetables (603 recalls) and Salads (571 recalls) are the next two largest group. However, since salads contain a large amount of vegetables such as lettuce and corn, the emergence of two groups is also a large food category. Given the background knowledge that pathogens are easily transmitted through diary products, some categories such as Ice Creams and Cheese that contain large amount of diary ingredients are likely to be recalled due to the contamination of such ingredients.

In addition to the number of recalls, classifications of recalls towards each food group were also explored at, which was explained in full detail in the above section. For all of the groups, recalls that are potential detrimental occupy a significant portion, and in total, there are 4,438 productions, which is more than 80% of all recalls, that can cause serious adverse health consequences. This shows the serious impact of pathogens on human lives if accidentally consumed.

4.2 Geospatial Analysis

Maps allow users visually see the spread of recall cases over a flat surface.



As shown above, recall cases gather around the West Coast as well as the East Coast extending to the Middle West, contributing to two major clusters. The general pattern is that there are huge amount of recalls gathered along the northeast coast of the US with extension to some of the southern areas especially Florida. Tristate area and California are where there are huge concentrations of markups also. Companies involved include “big names” such as Whole Foods Market and General Mills, Inc.

4.3 In Conclusion

Food safety is so essential that any minor issue within the food production system can lead to potential problems that have impact on all civilizations. With impact broad as they are, it is a must to ensure effective control over food production despite the fact that pathogens contaminations are difficult to surveil. It is appalling to see the number for recalling cases over the past seven years, and the fact that it is still growing requires almost immediate attention to food quality control.

From another perspective, since pathogens can be transmitted through multiple ways, it is important for the general public to be aware of potential issues and take cautions on daily activities. Proportionally speaking, in consideration of the total number of productions each year, the amount being problematic and recalled may not be significant, but it is always important to be aware of the fact that such small things can lead to catastrophic consequences, especially when human lives are involved.

Future Considerations

Actions from all levels need to be taken to detect contaminations in time so that issues can then be tackled effectively. This requires systematic and more structured levels of involvement, including government surveillance both on state and federal level, law enforcement, as well as consumer awareness.

Since this project takes primarily the perspective of the general public, suggestions towards government agencies involvement or law enforcement are not to be discussed in detail. One of the main actions that is urgent to be taken, in consideration of consumers' benefits, is transparency towards the process of production with accurate documentation. Although complete documentations of production are difficult to implement in real situations given the additionally labor that is otherwise not needed, it is still worthwhile to make effort to contribute to a safer food network.

From consumers' perspective, it is important to be aware of these potential food safety issues, especially when children, the old, the pregnant and the weak are the main targeting group of these foodborne germs. At least, basic understanding of these pathogens would be helpful in avoiding unnecessary exposure. It requires a general knowledge towards the transmission, symptoms, and prevention in association with each type of pathogen. Although they share some common features in the ways they are transmitted, there are also fundamental differences in biology. However, in terms of prevention, there are still some basic steps can be taken. According to CDC, the four principles need to be considered for effective prevention of infection of foodborne pathogens:

Clean

- Wash hands 20 seconds throughly with hand soap before, during and after food preparation

- Clean all fresh fruits and vegetables under running water before eating
- Wash all utensils, cookwares, cutting boards and countertops with hot water and soap

Separate

- Separate raw foods such as meats, poultry, eggs, and seafood from read-to-eat food to prevent germs from spreading
- Use separate cutting boards and plates for meats, poultry, and seafood
- Use separate bags for raw foods from others when grocery shopping
- Put raw foods in different sections in the fridge, away from others

Cook

Germs that reside in raw foods can get killed when the internal temperature is high enough

- Use a food thermometer for detecting the internal temperature when food is being cooked
- Constantly check color and texture change when food is being cooked
    - 145°F for whole cuts of beef, pork, veal, and lamb (then allow the meat to rest for 3 minutes before carving or eating)
    - 160°F for ground meats, such as beef and pork
    - 165°F for all poultry, including ground chicken and turkey
    - 165°F for leftovers and casseroles
    - 145°F for fresh ham (raw)
    - 145°F for fin fish or cook until flesh is opaque

Chill

- Keep the temperature in fridge below 40°F for perishable foods (those that are likely to decay, spoil or became unsafe if not kept refrigerated)
- Refrigerate perishable foods within 2 hours. If outdoor temperature is above 90°F, refrigerate within 1 hour
- Thaw frozen food in the fridge, microwave or cold water

Bibliography

"Botulism." Centers for Disease Control and Prevention. May 08, 2017. Accessed May 19, 2018. https://www.cdc.gov/botulism/symptoms.html.

"Campylobacter (Campylobacteriosis)." Centers for Disease Control and Prevention. October 02, 2017. Accessed May 19, 2018. https://www.cdc.gov/campylobacter/symptoms.html.

Ensuring Safe Food: From Production to Consumption. Washington, D.C.: National Academy Press, 1998.

FDA Investigations Operations Manual. Arlington, VA: Washington Business Information, 2018.

"Food Safety." Centers for Disease Control and Prevention. April 04, 2018. Accessed May 13, 2018. https://www.cdc.gov/foodsafety/cdc-and-food-safety.html.

"Food Safety." Centers for Disease Control and Prevention. February 16, 2018. Accessed May 14, 2018. https://www.cdc.gov/foodsafety/foodborne-germs.html.

"Foodborne Illnesses." National Institute of Diabetes and Digestive and Kidney Diseases. June 01, 2014. Accessed May 14, 2018. https://www.niddk.nih.gov/health-information/digestive-diseases/foodborne-illnesses.

FMI. 1998b. *A study of consumer trends toward irradiation*. Food Marketing Institute, Washington, DC.

"Listeria (Listeriosis)." Centers for Disease Control and Prevention. June 29, 2017. Accessed May 19, 2018. https://www.cdc.gov/listeria/symptoms.html.

"Symptoms." Centers for Disease Control and Prevention. April 12, 2012. Accessed May 19, 2018. https://www.cdc.gov/norovirus/about/symptoms.html.