Exploring Enteric Illness

Outbreaks in the U.S. using Data Visualizations

Summary Report by: Yuwen Li, Kathleen Parsons, Xuan Song, Qiming Zheng

Abstract

The purpose of our research project was to create several data visualization dashboards using the tool Tableau, that would engage general population users in investigating and learning about foodborne and waterborne illness outbreaks in the United States. Our primary goal was to have a single web page containing the dashboards that would allow users to scroll through the different visualizations and interact with them, to view trends and explore the data. Users would then gain insight to answer their pressing questions on reported food and water outbreaks in the U.S. This report discusses our process in-depth, from our initial exploratory needs study to integrated final iterations from our phase three usability study. We discuss what we learned, obstacles we faced and surprising insights from learning a new tool and usability testing. Please view the final website here: Wash-yo-hands website.

Keywords: Data Visualizations, Tableau, Enteric Disease, Foodborne Illness, Outbreaks, United States.

Introduction

Motivated by the international outbreak and spread of the novel coronavirus (COVID - 19) in early 2020, we focused our project on the history of foodborne and waterborne disease outbreaks in the U.S. Please note our data visualizations only contain data from U.S. born outbreaks, meaning a virus like the coronavirus are not accounted for in the dataset because it did not start in the U.S. However, we still believe it's crucial the general public can easily digest large amounts of data because rising incidents of outbreaks. We've learned from author Stephen Few, that visual perception offers the highest bandwidth channel into the brain (*Few, Chapter 1*); and as incidents increase and are becoming of increasing concern, effective displays of information will be important to the general health and well being of the public .

We were interested in the history of these U.S. based outbreaks and past trends. Sparking our curiosity as to where they have historically occurred the most, and in what setting. After our preliminary research in finding a large enough dataset to fulfill project requirements, we settled on the CDC's NORS (National Outbreak Reporting System) dataset. This dataset specifically looks at the spread of enteric disease in the United States. We centered our research and data visualization development around four research questions, including:

- What season are these outbreaks occurring the most?

- Where in the U.S. are they happening the most?
- What setting are they happening in?
- What reported food is the biggest culprit of the spreading illness?

Previous Work

We have explored several sources of related work of national outbreaks, especially those tailored for the general public audience. We've learned a lot from both the strengths and weaknesses of these visualizations and used some of them as inspirations for our own project.

Firstly, we found several static visualizations that are relevant to our topic in the annual report of Surveillance for Foodborne Disease Outbreaks, United States, 2014. See figure 1 shown on the bottom left. While the small multiples of the maps make the geographical distribution of the outbreaks easily comprehensible, it lacks interactivity to allow further exploration on relevant information. We also found the visualization tool HealthMap (figure 2) inspiring for our project because it combines different data sources to present the most up-to-date trends and provide useful information for people who are traveling to specific regions with outbreaks. However, we found this visualization lacks legends or enough context for users to interpret the data.

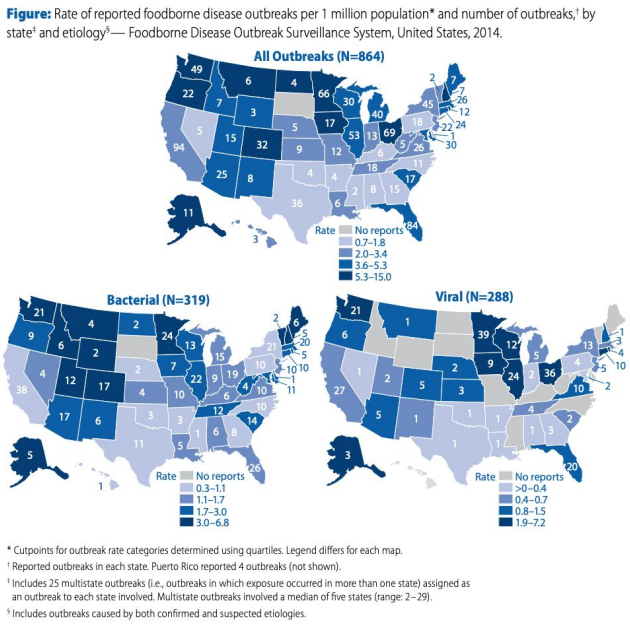


Figure 1. Surveillance for Foodborne Disease Outbreaks, US. (CDC, 2014)

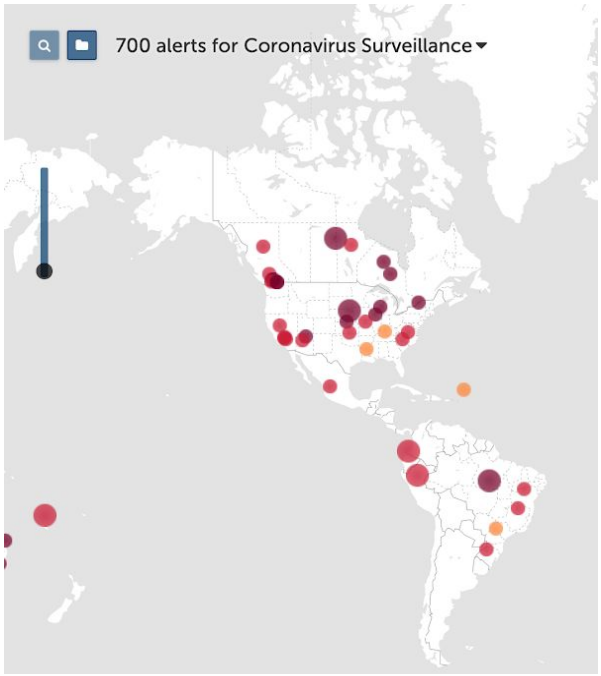


Figure 2. Coronavirus Surveillance Map (HealthMap, 2020)

The most important and relevant previous work we found was the current live dashboard on the CDC website using the NORS dataset to display several graphs (shown in the figure 3 below). There are graphs to show some basic patterns in the data, but we found they are not reflective of the richness of the dataset. For example, we found it difficult to see how food consumption affected the outbreaks. The interactions provided are also very limited. Thus, we found it an opportunity for us to make better use of the data and tell more engaging and insightful stories using visualization techniques such animations, details on demand, brush and linking and so on.

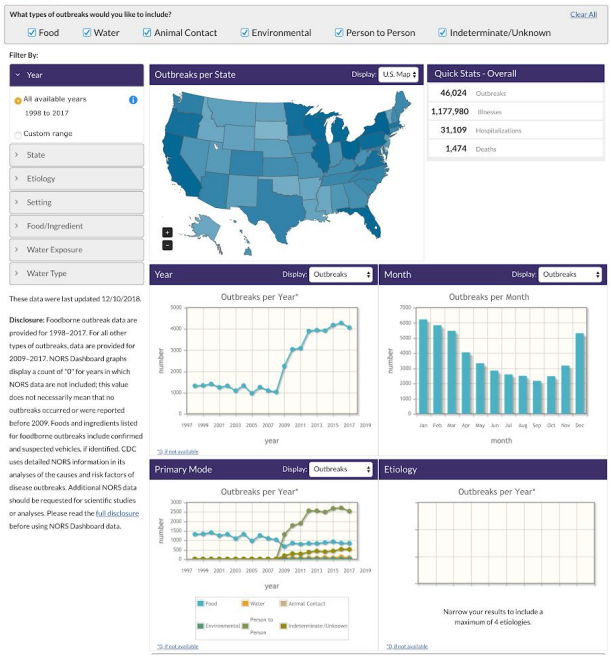


Figure 3. National Outbreak Reporting System (CDC, 2020)

Personas

By talking to potential users of our visualization, we created two personas to help us resolve important questions and guide us in the development of the visualization. We initially thought medical professionals could find this tool useful. Over time as we found that their need is more around real-time monitoring, we realized our dataset which has only historical records is better suited for storytelling. Thus, we decided to design for the general public who are looking for general patterns of the past foodborne and waterborne outbreaks in the U.S. The following shows a typical persona of our target audience.

Vivian Wang is a social worker who has lived in the United States for over five years. She recently read about the Wuhan coronavirus outbreak and realized that she needs to learn more outbreak information, and see how it affects people's life in the U.S. She wants to learn how outbreaks spread over time, and where outbreaks happen the most. Because she is interested in learning about past outbreaks cases, she also desires to learn more in-depth about why certain outbreaks happen through engaging ways.

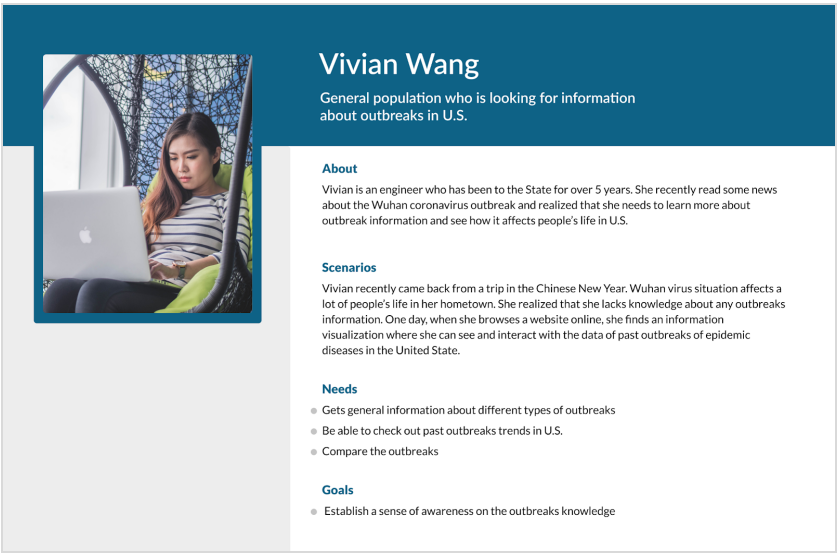


Figure 4. Persona

Dataset

The dataset we used comes from the National Outbreak Reporting System (NORS) provided by the Center for Disease Control and Prevention (CDC).

Several defects of the data collection process compromise the data visualization from showing the full picture of outbreaks in the U.S. First, the CDC collects data of enteric outbreaks, except for some non-enteric illnesses transmitted by food or water (CDC, 2020). Therefore, the visualizations focus only on foodborne and waterborne outbreaks. Second, the dataset includes only outbreaks that originated in the U.S. so non-U.S. origin outbreaks are not under investigation. Thirdly, the information is voluntarily reported to CDC by the public health agency that conducted the investigation on the outbreak. Thus, the outbreaks that are not identified, reported or investigated completely are not included in the dataset.

We narrowed down the dataset by focusing on outbreaks from years 2009 to 2017 and by including only the variables that make sense to the general population users. Eventually, the dataset contains 46,000 rows and 13 variables. Each row in the original excel spreadsheet represents a single outbreak. The nominal and ratio variables we focused on including year, month, state, primary mode (transmission agents), setting, illnesses, hospitalizations, deaths, food vehicle, IFSAS Category, Water type and animal type.

There are some flaws in the dataset. Some variables such as etiology (24.29%), food vehicle, and IFSAC Category (90.37%), water exposure (97.72%), and animal type (99.02%) have high percentages of missing values, not only due incomplete information in the report from health agency, but also due to different types of modes of outbreaks (such as person-to-person and food transmitted) relate to different variables. As for quantitative variables, the empty fields account for 14.21% of the Deaths and 13.98% of the Hospitalizations. These blank fields may compromise some analyses of the data visualization.

Another problem in terms of the quality of the dataset lies in the fields with multiple values. Therefore, some variables have too many distinct values. For example, food vehicle has 3328 distinct values. Some are essentially similar such as ‘beef’, ‘beef, raw’, ‘beef smoked’, ‘beef, stew’. Others are apparently aggregation of a bigger database with smaller granularity such as ‘chips, tortilla; chicken, strips; ground beef, meatballs; stuffed mushroom; quiche; mixed fruit; crudites’. The problem is partly due to the unstructured report survey that takes a variety of values for this field and lack for direct understanding of what food vehicle is at fault. Also, the undetermined nature of the causes of an outbreak makes it hard to put a defined food vehicle but results in a list of potentially contaminated food.

Exploratory Analysis

We noticed that person-to-person illness transmission and food transmission are the two most prominent modes of infection, which contributes to over 98% of all the outbreaks. So we decided to split the dataset into two subsets based on those two primary modes so that we could understand the data better. As included in our previous exploratory analysis reports, we explored both the geospatial and temporal relationships in our data. We also focused on finding patterns around different settings. Some insights we derived that are instrumental for our final prototype are listed below:

- For food transmitted outbreaks, the District of Columbia had the highest numbers of illnesses while Rhode island has the highest death rate among all the states. For person-to-person outbreaks, New York

has the largest number of illnesses while for Alaska, each outbreak tends to affect more people as it has the biggest illnesses per outbreak.

- The food transmitted diseases are generally less severe in the fall season compared to other seasons, while person-to-person transmitted diseases had their lows in the summer.
- We found there were correlations between different quantitative dimensions and the settings where the outbreaks happened. We found that child day care has the largest hospitalization rate but a very low death rate, while in hospitals, the death rate is the highest and the hospitalization is relatively low.
- We also noted some interesting outliers in the above patterns that we would like to explore further.

Also, when exploring, we realized that different variables in the dataset, such as the number of illnesses, hospitalizations and deaths, can all be used to measure the severity of outbreaks. They have different connotations so that one cannot take the place of the other. Therefore, for our final design, we need to consider the different measures to represent the severity of outbreaks comprehensively. After understanding our dataset from all the different perspectives, we wanted to consolidate our insights and form strong narratives of the patterns we believed would be most valuable and interesting to share with our target user group.

Data visualizations

Brainstorming & Sketches

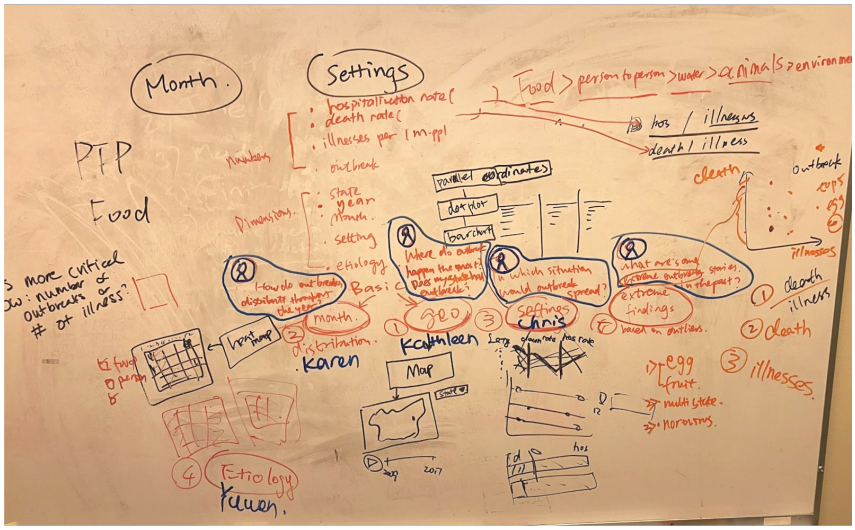


Figure 5. Brainstorming and Whiteboarding

Given the trends, patterns and insights we were able to discover from the data in our exploratory analysis, we came to the agreement that the purpose of our final visualization should be more around guided storytelling instead of free exploration. During our brainstorming and sketching session, we grouped all of our exploratory questions and discussed the best approach of visualization for each. Finally, we decided to present our findings in five aspects: geolocation, time, setting, etiology, and extreme cases.

Implementation

Next, we started building out the dashboards for each category in Tableau so as to find out the best way to encode the variables and showcase the patterns.

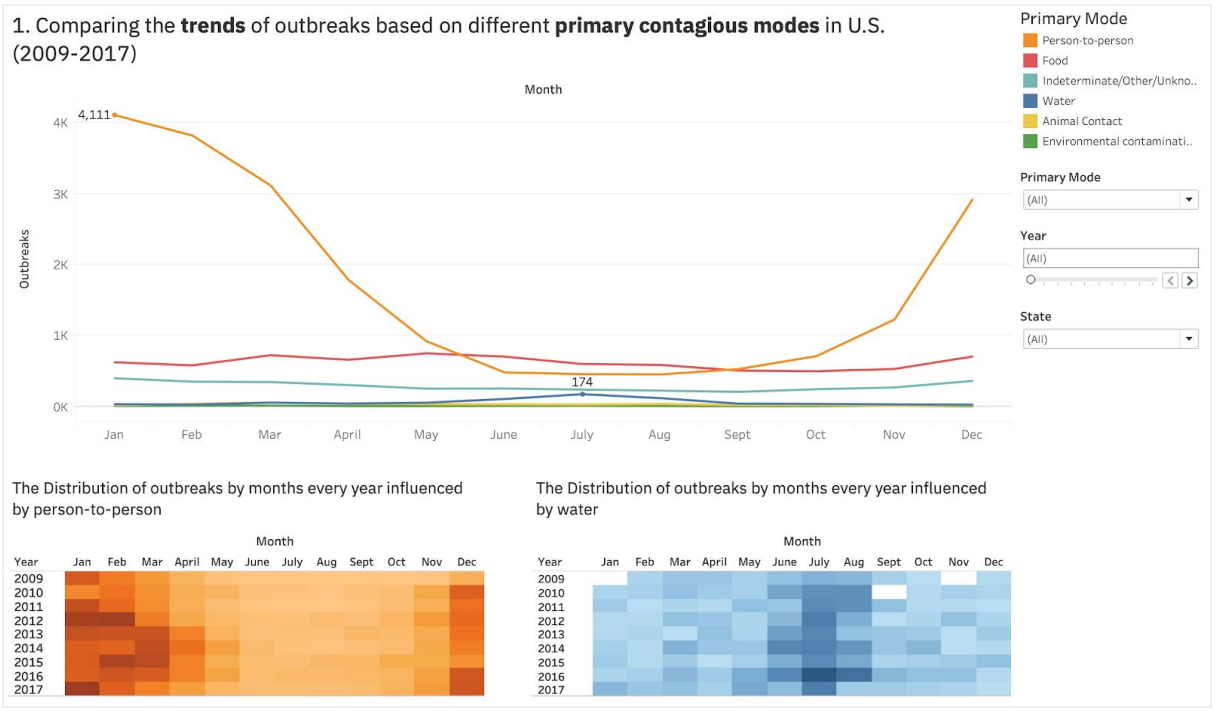


Figure 6. Dashboard for Temporal Patterns

First, to present the time-series relationship of the outbreaks data (Figure 6), we used a line graph with different colors encoding different primary modes. It's made clear that the person-to-person is a dominant transmission mode in almost all seasons and it reaches its peak in the winter season and gradually dies down in the summer. On the contrary, waterborne outbreaks happen the most in the summer. We've added two heatmaps at the bottom of the line graph to compare and contrast the outbreaks distribution of these two modes which showed distinctive patterns.

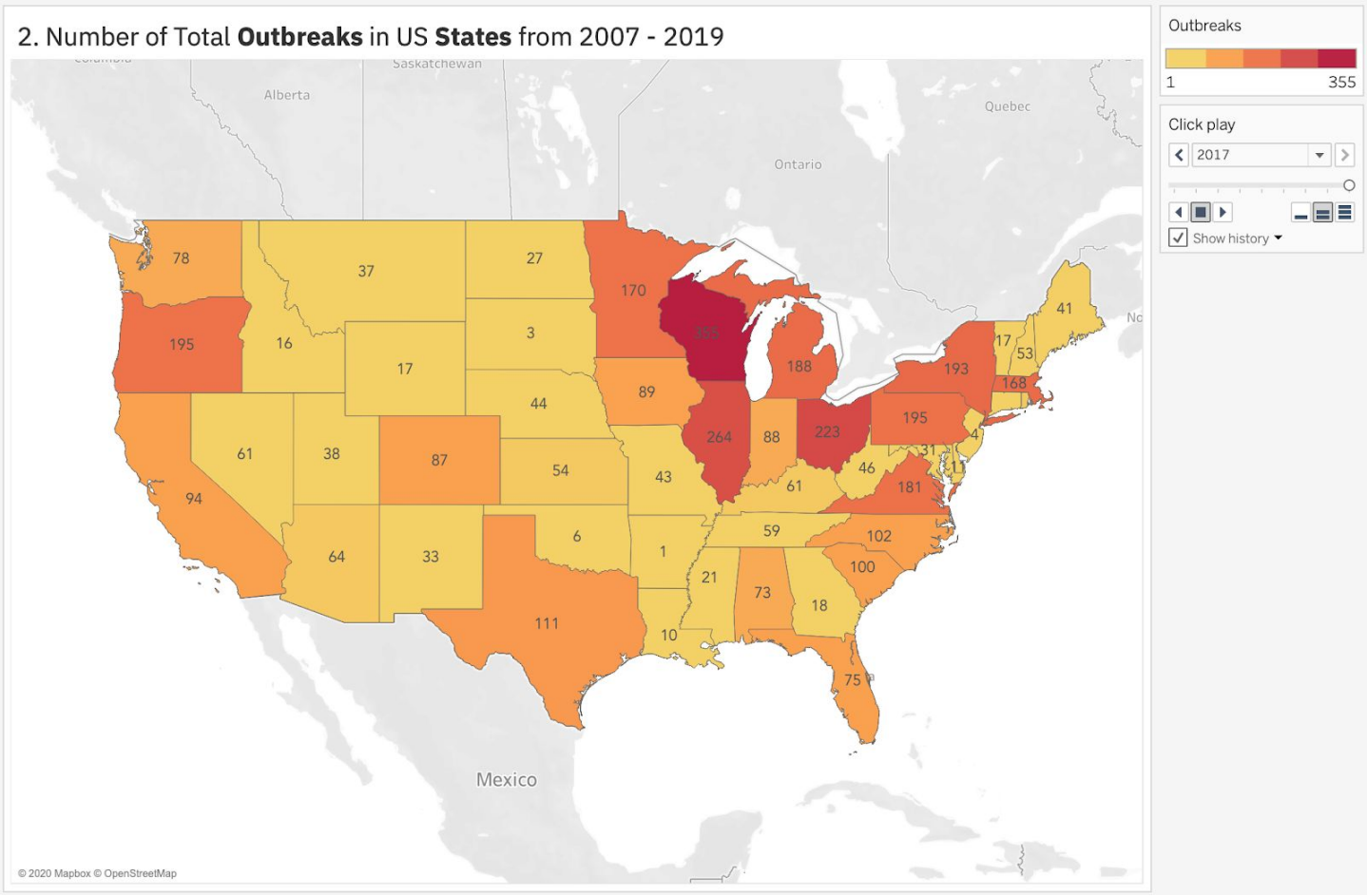


Figure 7. Dashboard for Geolocation Patterns

Second, the geolocation relationship of the outbreaks distribution is mapped out in a geographic visualization (Figure 7). With the color scale, it's easy to tell there were more outbreaks in the northeast states especially the state of Wisconsin in 2017 because the red is darker there. We also used animation to show that such a pattern has been pretty consistent throughout the years from 2007 to 2019. Users can play with the animation control panel on the right to explore further by themselves.

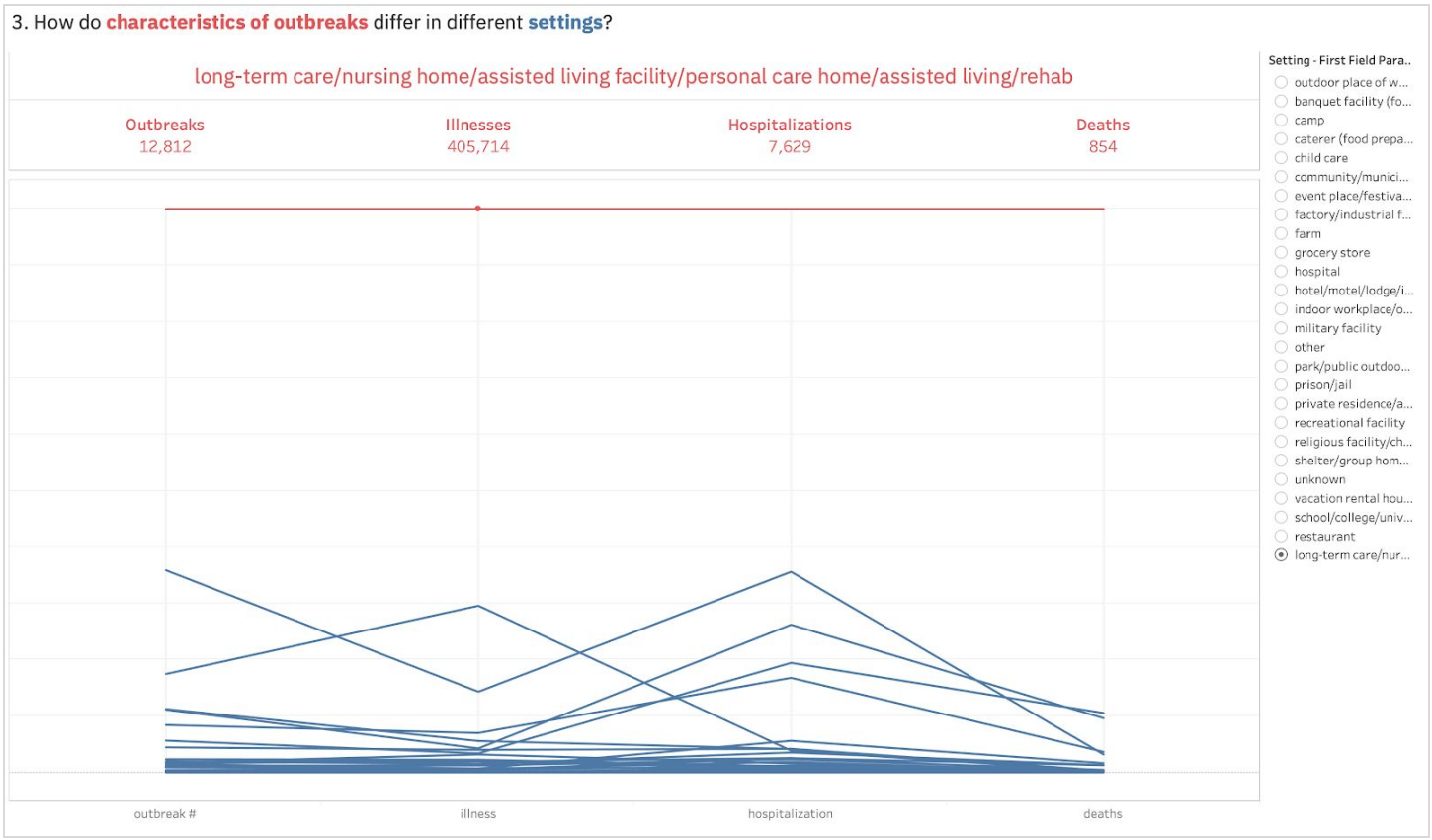


Figure 8. Dashboard for the Profiles of Different Settings

In order to show multiple variables of outbreaks related to different settings, we implemented parallel coordinates in Tableau as shown on Figure 8. Each line in the visualization represents one setting where the outbreaks originated. Users can hover over each line or click the radio button on the right to see the different measures of each setting. The header and different measures automatically change corresponding to the line selected. It's easy to tell that considering all the measures including the number of outbreaks, illnesses, hospitalizations, and deaths, long-term care is the most dangerous setting compared to the rest.

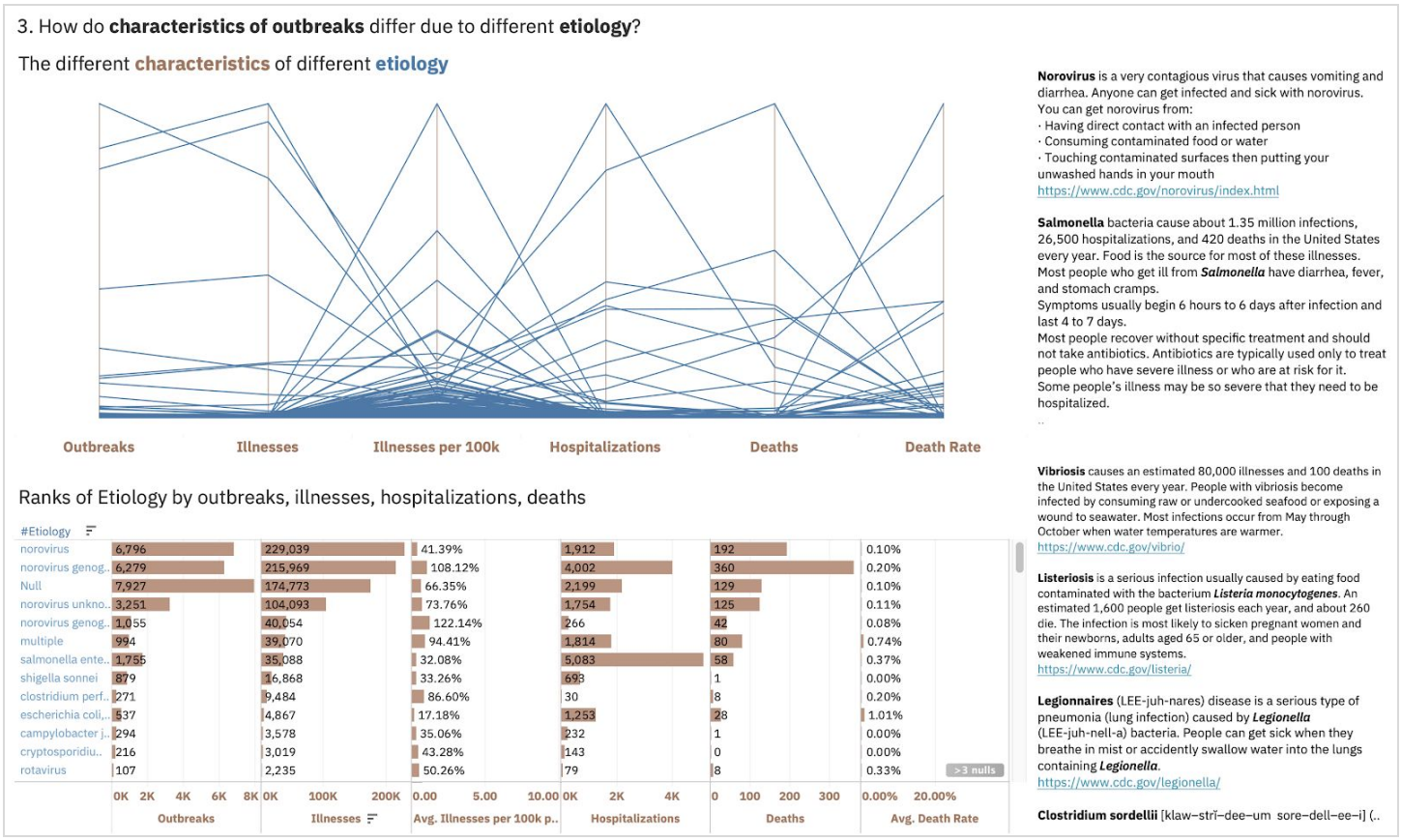


Figure 9. Dashboard for the Profiles of Different Etiology

Similarly, we also used parallel coordinates to compare the profile of different etiology (Figure 9). We used the number outbreaks, illnesses, illnesses per 100k population, hospitalizations, deaths, and death rate to show the characteristics of different etiology. Users can click on each line and the bar chart below will be updated to show the actual numbers in those 6 dimensions. It's easy to compare different etiology in terms of different dimensions. Some etiologies such as the "norovirus" are more widespread than others because they caused more illnesses, while other etiologies such as the vibrio vulnificus are more deadly due to their high death rate. We also included more information to explain different terminologies because they may seem difficult to understand for our target user group who are the general public.

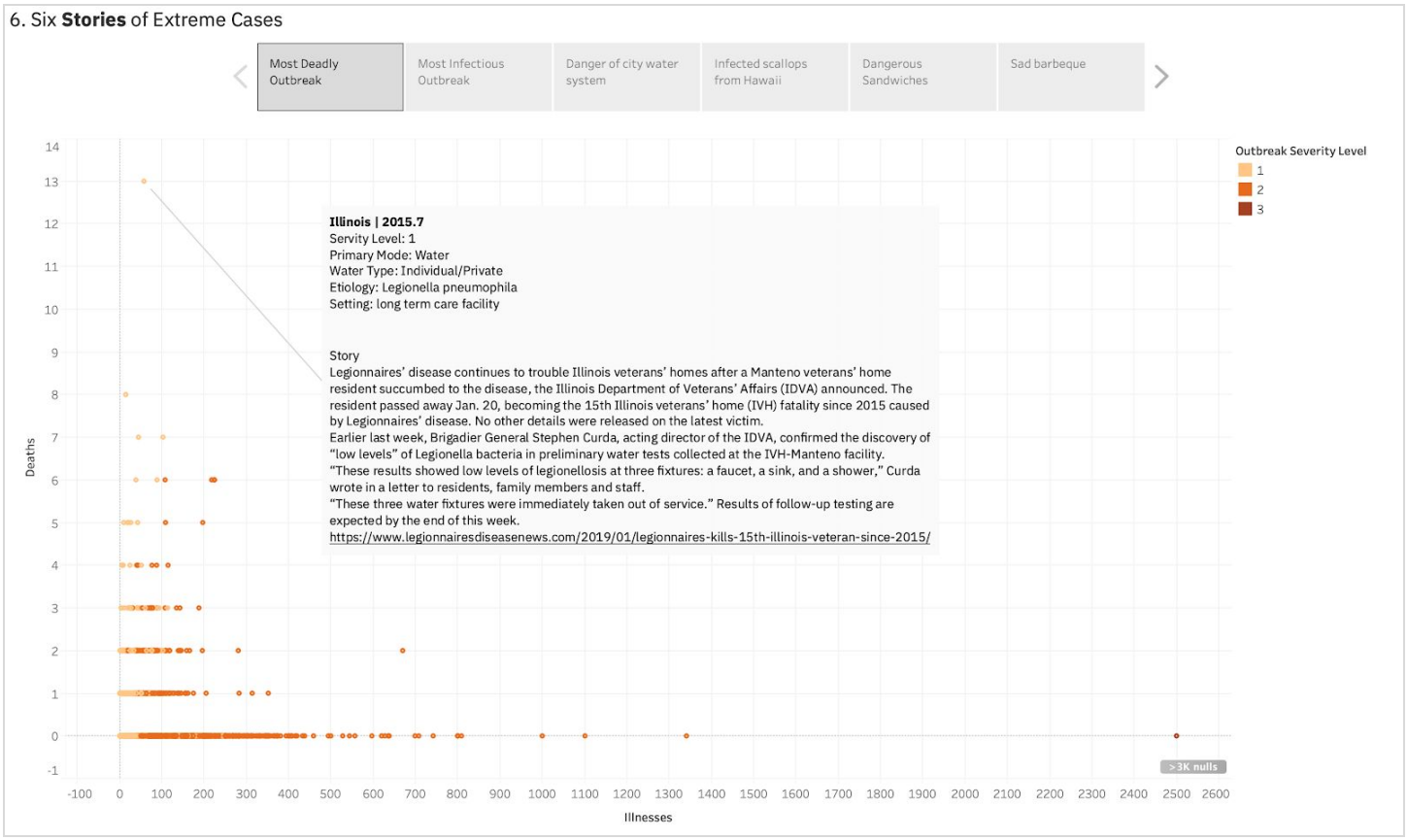


Figure 10. Dashboard for Extreme Cases

Lastly, we found some intriguing extreme cases (Figure 10), searched for news coverage online, and displayed the news stories that could explain those outliers. All outbreaks from 2009 to 2017 are presented in a scatterplot with deaths and illnesses being the two axises. The position of each outbreak encodes those two quantitative values and makes the extreme cases easy to be identified. Some highlighted cases include the most deadly outbreak, the most infectious outbreak, and so on. We hope to use such storytelling data visualization to shed light on those severe outbreaks in the past that may have been forgotten by the public.

Implementation on Interactive Website

Storytelling

In order to tell a better story about the past outbreaks, we explored different ways to showcase data visualization. One is an exploratory dashboard where users are able to explore outbreak data visualizations, while the other one is an interactive web-based data visualization tool that allows the general population to explore and learn insights about past outbreaks data. We chose the latter one because after our initial user testing and competitive analysis, we realized that people are more engaged in learning outbreak data when they know the context. We decided to do a storytelling format where we combine a user-centered design approach with data visualization together. Figure 11 is our first version

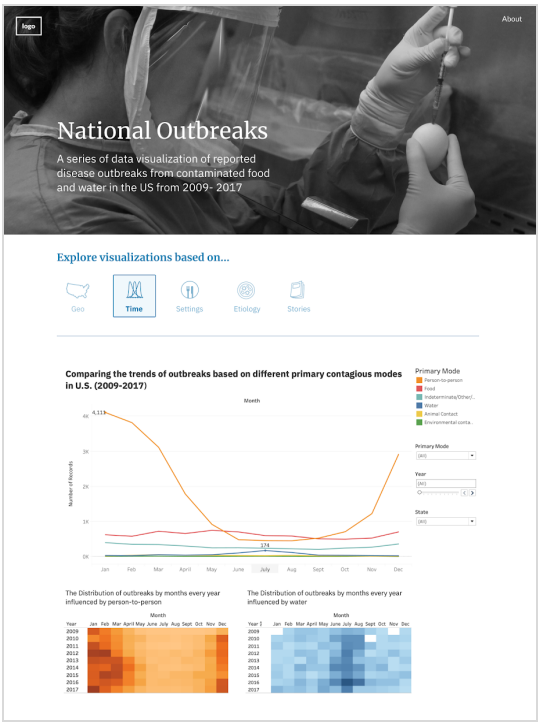


Figure 11. First Version of the Website

Website Iteration

We hosted the website on ZEIT and coded the website in React. Since we have five different dashboards, we initially designed five tabs to allow users to switch between different dashboards. However, there are some technical issues when we integrate tableau dashboards into the website. The loading speed is extremely slow and some graphs were messed up because of the loading method of Tableau's APIs . In addition, we realized the website lacked the definition of terms and description for each dashboard. Thus, in our next iteration, we decided to put different dashboards on one single page to reduce the loading delay, while adding research questions and captions to help viewers understand the context and data visualization better.

Usability Study Test

Participants

We conducted 2 rounds of usability testing with a total of 7 participants who represent the general public, vary in gender and education levels, but remain in the same age group of 18 - 25. The purpose of the testing was to understand what worked well for potential users and what could be improved upon. We were generally interested in the understandability and interactions of the different visualizations we created.

First round: informal usability testing

The first round of the usability testing was an informal moderated test conducted with only the five dashboards on Tableau. Participants were asked to explore the dashboard without specific tasks and think out loud about their understanding and confusion.

After the first-round testing, we modified the visualization and also built up the website for the next round of testing with higher fidelity..

Second round: task-based usability testing

The second round of the study was a moderated testing of the built website. The research questions we raised include

- Can users understand the variables we encoded in the dashboard?
- Can users complete the tasks and understand the trends
- Do the dashboards support users' expected ways of interactions?
- What information is missing in terms of supporting users' understanding?

The session starts with pre-test interviews concerning the participant's general understanding of outbreak issues and their level of experience interacting with data visualizations. Then, defined tasks for each dashboard are given to the participant. We again used the think-out-loud protocol. The participants were asked to complete both open-ended exploration and defined tasks for each visualization. After each task, we asked follow-up questions concerning their understanding, learning, experience and satisfaction. Finally, a post-test interview is conducted for their overall experience and understanding of the website.

We designed tasks for each of the 5 different dashboards in the same scenario: the recent corornavirus outbreak has sparked your interest in the history of outbreaks in the US. You have come across this website to learn more about outbreaks in U.S..

For the dashboard showing temporal patterns, the task focused on whether participants and quickly grasp the encoded variables, easily notice the seasonal patterns, and easily narrow down their focus by filters and dropdowns if they want to explore in-depth a specific transmission agent. For the one showing the distribution of outbreaks in the U.S., the tasks were designed to check whether participants are able to point out the cluster of outbreaks in the north-east and whether they are able to use the control panel to play the animation. As for the dashboard showing the distributions of illnesses by different settings, the tasks are designed to see whether participants can understand which setting has the highest number of illnesses and death rate and whether the

interaction of demands for details is intuitive. For the dashboard showcasing the profiles of different etiology, tasks are designed to see whether participants can understand how to interact with the complex parallel coordinates and understand which are the most infectious and most deadly etiology. Lastly, we designed tasks for the story dashboard to see whether participants can understand the relationships between the story shown and the scatter plot at the background and whether they are able to use the filters and highlighter to explore the scatter plot. A detailed script for the testing can be found in the appendix.

Positive Findings

From our usability testings, we found that participants prefer seeing trends and insights of each category. In addition, they enjoyed reading the story dashboard because they can see photos of the food and understand the reasons behind some severe outbreaks. 4 out of 7 users said that they enjoy seeing trends and insights which were showing in time dashboards. The majority of participants could finish the tasks and be willing to share this website with their friends.

Problem Space

Generally, in usability testing we found that:

- The dashboards are too complex and demand exploration before users can take away key patterns
- The demands for details are not properly supported by well-designed tooltips, filters and search
- Some interactions (click, hover and filter) with the dashboards are not intuitive enough
- The labels of variables and captions and the title of graphs contain too many professional terms and create barriers for users' quick learning

Technical Limitations

As we were conducting the usability test, our participants ran into several software errors. For example, there was a delay on the map visualization and therefore users had a hard time zooming in and out to view the details. The play button in the map visualization also reacted very slowly. In addition, when switching tabs, it took more than 10 seconds to display the data visualizations and thus resulted in unpleasant user experience. In some testing sections, we also observed that some dashboard didn't show up unless participants reloaded the entire website.

Final Iterations

Based on the findings we got from the usability tests, we went on to more rounds of iterations. We made many adjustments on each of the five visualizations and the overall web design. The changes can be generally summarized into the following 3 major categories and we will highlight some examples for each category.

1. Simplify and streamline the visualizations

1.A. Months & Seasons: kept only the heatmaps with clear patterns

Because our target audience is the general public and many of them found it hard to comprehend some of our visualizations, so we tried very hard to strike a balance between providing enough useful insights and minimizing the cognitive load. For example, we removed the giant heatmap into 2 clear heat maps which show the seasonal pattern of outbreaks spread by person-to-person transmission and water.

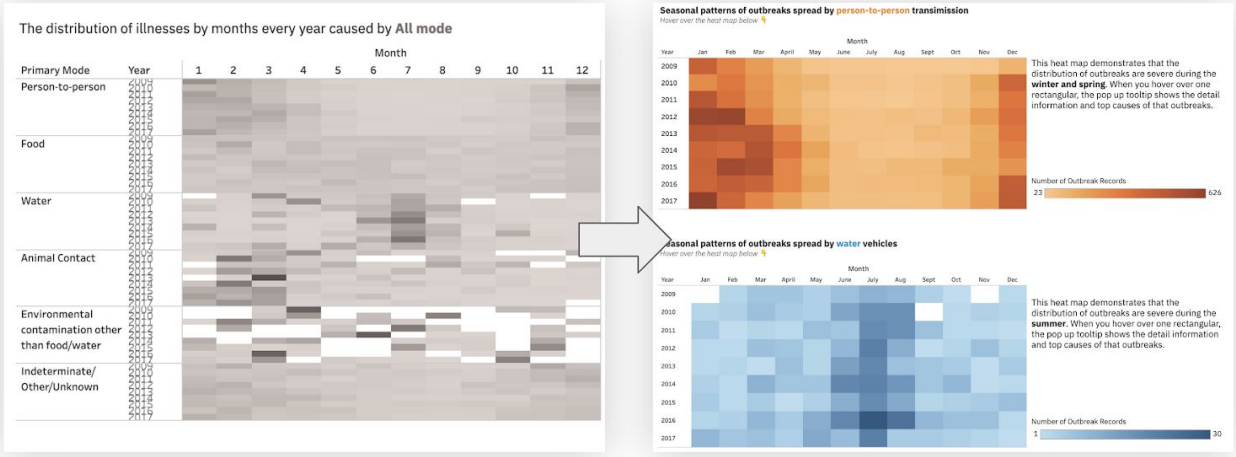


Figure 12. Iteration on the dashboard showing temporal patterns

1.B. States: removed the deaths breakdown of different primary mode

In the geographic map visualization, previously we had the outbreak numbers encoded in different colors shown in each state on the left, and control for animation on the right. There is another bar chart breakdown for the deaths when a state is clicked shown at the bottom. We found that the interaction between the map and the bar chart is distracting and for some states. We removed the bar chart for two reasons: First, it is distracting of the main map interaction. Our goal is to help users to focus more on the pattern of geographic distribution and animation. In addition, users are not interested in clicking the states to view primary mode information which causes cognitive overload.

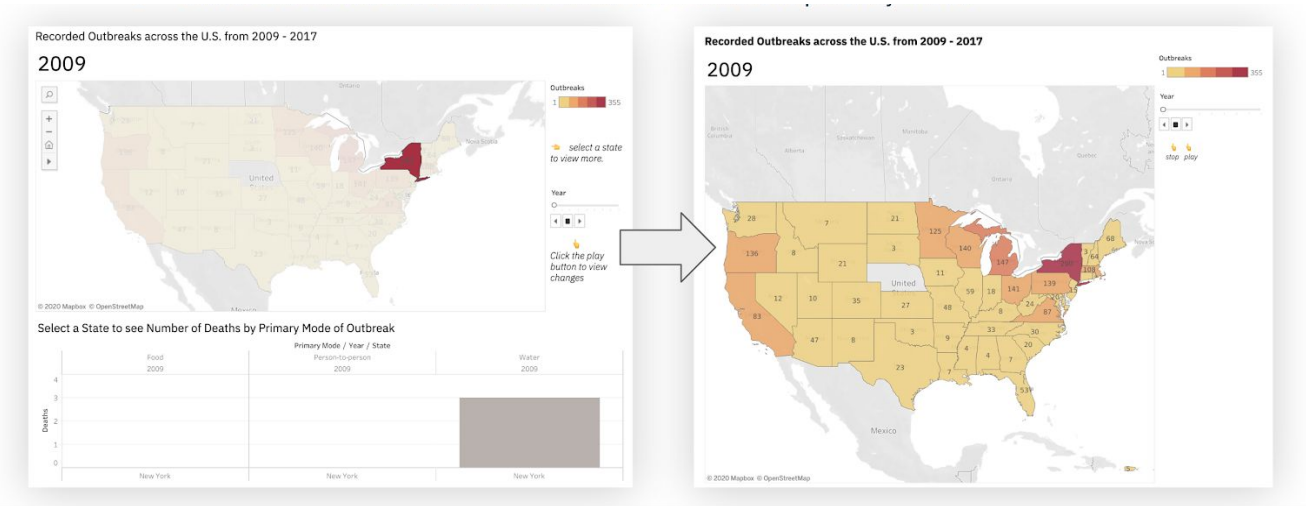


Figure 13. Iteration on the dashboard showing geo-location patterns

1.C. Parallel coordinates of etiology -> scatterplot of food and waterborne illnesses

We also had a debate about whether to keep the parallel coordinate visualization to show the different etiologies.

First, none of our participants had experience of reading a graph like this before. It took them very long to figure out how to interpret the different lines and the values for each coordinate and it even caused some frustrations for some participants we tested. Second, they were not familiar with the term etiology and were more interested in the word cloud especially the food and water types which they are familiar with. Although we put in a lot of effort to implement such interactive parallel coordinates using Tableau, we decided to put our users at the center of design. And reduced the 6 dimensions to the 3 most important ones. And focused on the pattern specifically on the foodborne and waterborne diseases in this scatter plot.

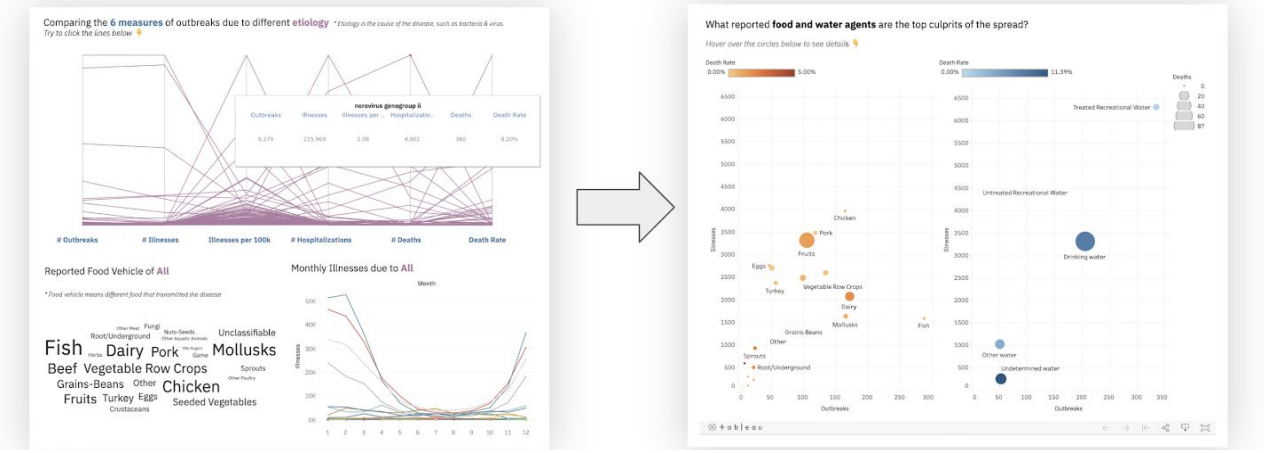


Figure 14. Iteration on the dashboard showing profiles of etiology

2. Enhanced interactions to facilitate exploration (details on demand)

2.A. Settings: show the right-hand graph of detailed breakdown only after click

After we streamlined our data visualization and made the patterns and trends more clear, we also want to facilitate exploration by optimizing our interaction. We used the principle of details on demand when designing such interactions. Participants mentioned that they were confused about the relationship between the two tree graphs. We revised the left tree map where we showed illnesses occurred in different settings and primary mode breakdown in each setting. When users click one settings block, the setting name stays on the top to give users better context.

Figure 15. Iteration on the dashboard showing profiles of settings

2.B. Months & Seasons: treemap changed to bar chart for better clarity

5 out of 7 participants raised concerns that the colors and details in the tooltips are confusing. Some blocks did not include any context and blocked the main data visualization as well. Thus, we changed to bar charts to show the settings breakdown in each specific month.

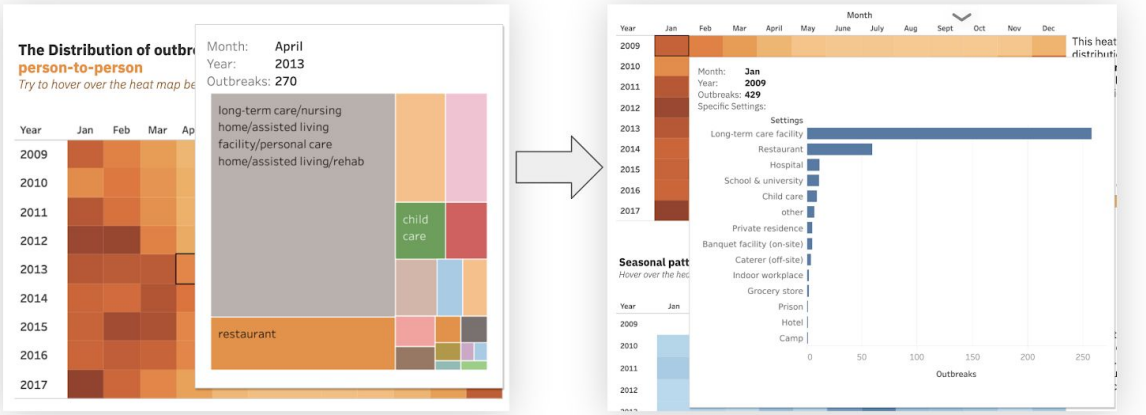


Figure 16. Iteration on the tooltips showing breakdowns of settings and food vehicles of the temporal dashboard

2.C. Word cloud shown on hover

Initially, we have word cloud visualization in the etiology dashboard, however, we changed the dashboard to the scatter plots. We decided to add word clouds into the tooltips as a secondary information to help viewers understand different food types and water types that occurred in outbreaks.

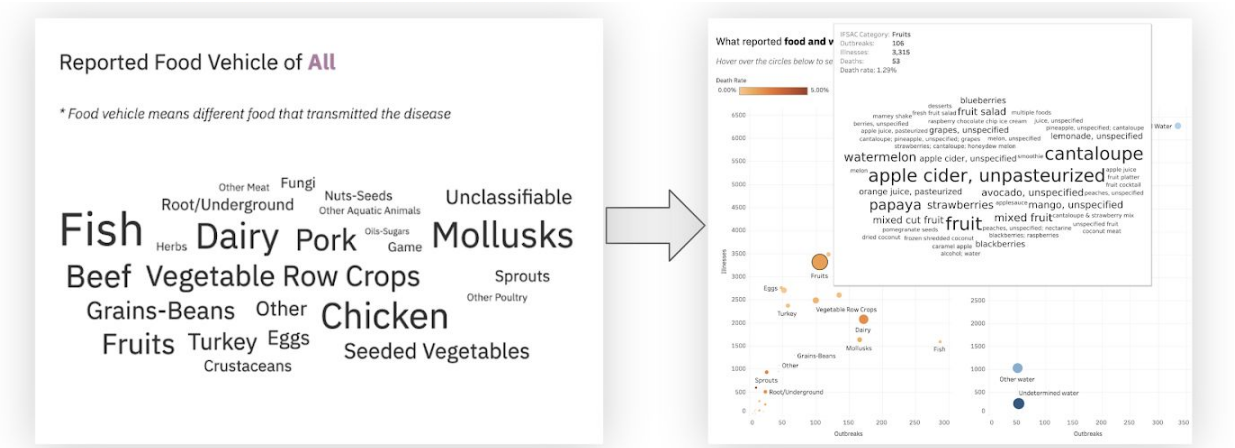


Figure 17. Iteration on the breakdowns of types when hovering

3. Better storytelling with better web contents, labels, annotations, and legends

Last but not least, we fixed a lot of usability issues by providing better contexts, comments and annotations.

We refined the web layout and introductory paragraphs before each visualization. In addition, we add more visible and better explained legends. In the extreme cases dashboards, with pictures and caption, users can get the gist of the stories much faster.

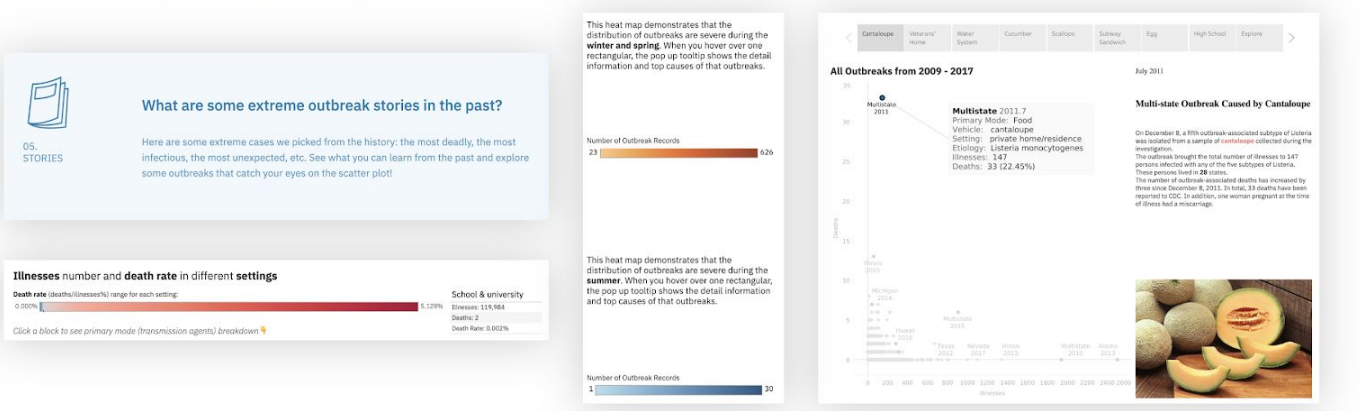
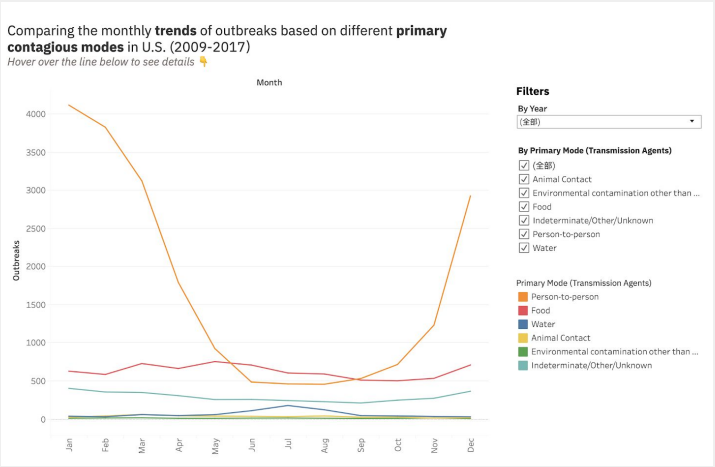


Figure 18. Improvement on messaging

Evaluation of Final Visualization

Months & Seasons



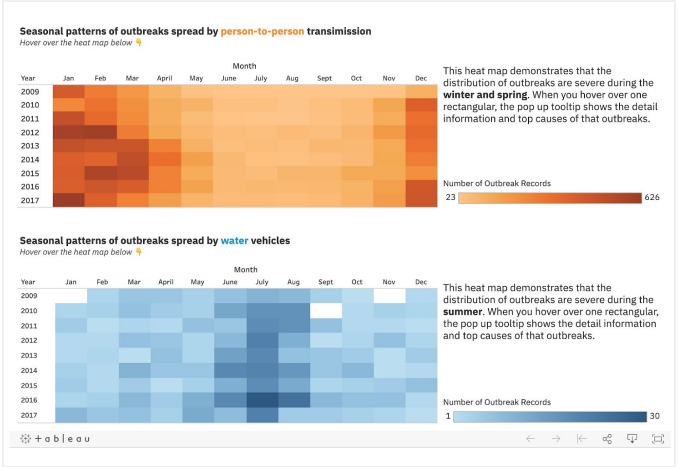


Figure 19. Trends of different primary contagious modes (Left)
Figure 20. Patterns of person-to-person & water contagious modes (Right)

The goal of these data visualizations is to demonstrate the trends and distribution of outbreaks across a year based on different primary contagious modes. The first dashboard (Figure 19) uses line charts to show overview trends of outbreaks. X axis represents the month and Y axis represents outbreak numbers. The color of different line charts encodes different primary contagious modes. Users can either hover over the line graphs to view more details or filter the information by year or by primary modes to see the changes.

The second dashboard (Figure 20) demonstrates the distribution of outbreaks spread by person-to-person transmission and water vehicles. The orange heatmap shows that the distribution of outbreaks spread by person-to-person transmission is severe during winter and spring; while the blue heatmap shows that the distribution of outbreaks spread by water is severe during summer. We use the rectangular heatmap to encode the time and density of color to encode the number of outbreak records. Detail on demand technique is being used here when users are interested in learning more details. They can hover over rectangular to see a popup tooltips showing the top causes of that outbreak (specific settings and types of water vehicles).

States

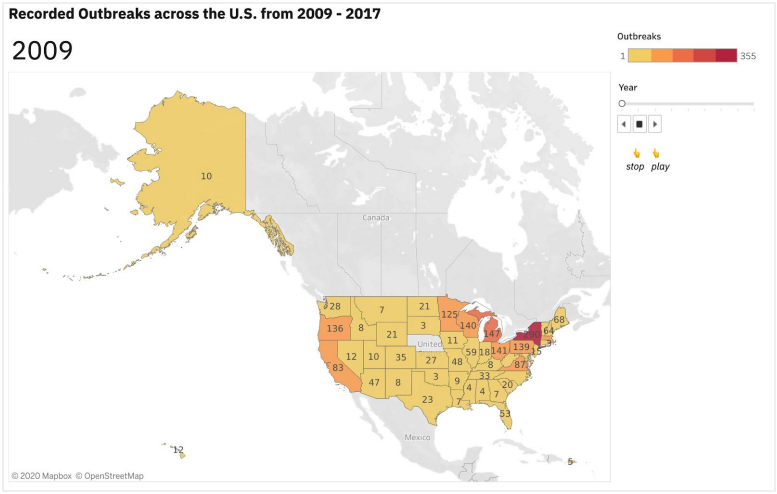


Figure 21. Recorded Outbreaks across the U.S. from 2009 - 2017

The second dashboard (Figure 21) means to showcase the pattern of geographic distribution. Changing red color hue values are used in encoding the number of outbreaks. Users can click to play the animation of how outbreaks distribute across the U.S. in different years and stop to hover for details. The changing number of years on top left indicates the running animation and provides users with necessary context. The encoding of geolocation with a map is intuitive for the general public since they can easily relate it to their knowledge. The different colors of the state can easily direct users' attention to the cluster area with the reddest color: north-east.

Settings

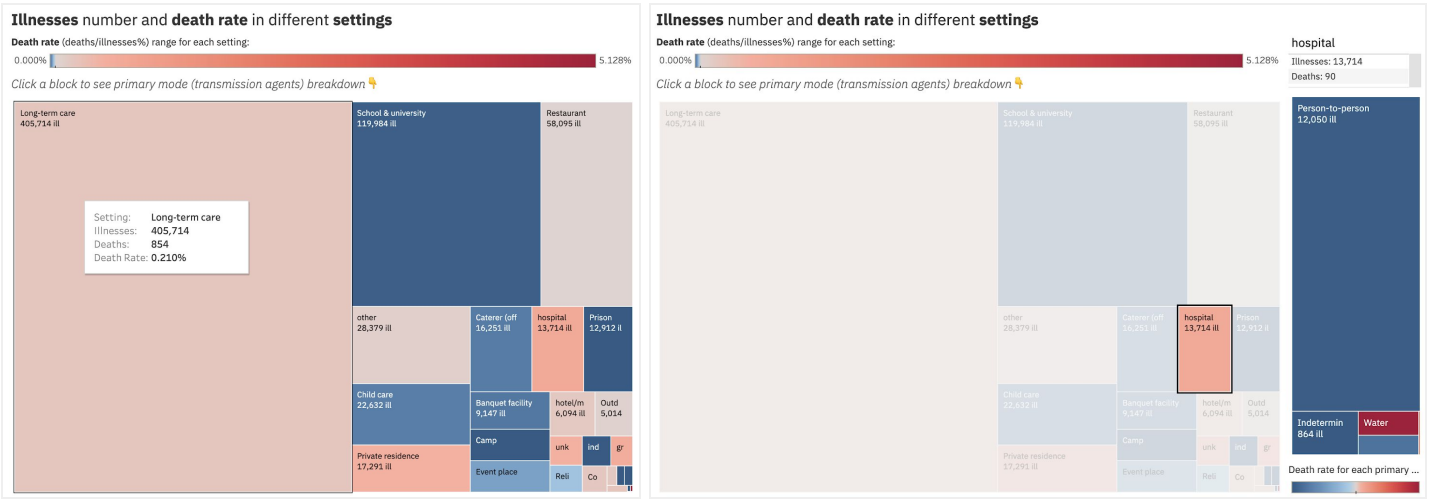


Figure 22. Initial state of dashboard for illlnesses and death rates in different settings (Left)
Figure 23. Interaction of dashboard for illlnesses and death rates in different settings after clicking (Right)

In this section, we sought to show the pattern of outbreaks originated in different settings, which refer to the locations or environments where the patients were exposed to such as restaurants, nursing homes, etc.

The primary treemap shown in figure 22 uses the size of the rectangle to encode the number of illnesses and colors to encode the death rate, that is, the percentage of deaths out of illnesses in each setting. Because the death rates range widely, we set its median value as the middle point of the diverging blue-red legend to make the differences more visible. With the color legend and clear annotations, it's easy for our users to understand that

the larger the block the more illnesses there are, and the redder the color, the higher the death rate. For example, they were able to identify the long-term care is the setting with the most illnesses fairly quickly.

After they click on any block, a secondary tree map showing the breakdown on the primary mode of the chosen setting will be dynamically displayed on the right. (Figure 23) The secondary tree map is hidden until the click and it uses the same rules of data encoding as the primary treemap to reduce the user's cognitive load. Such interaction gives the user an overview first and provides further detailed information on demand.

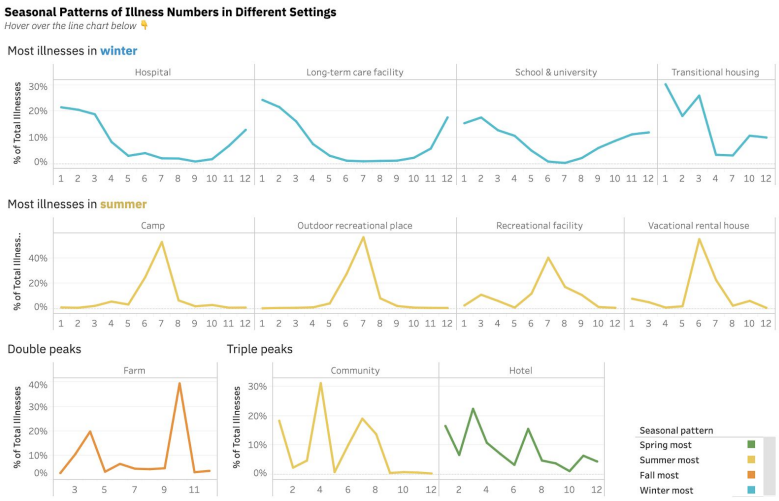


Figure 24. Seasonal patterns of illnesses number of different settings

After exploration on illnesses in different settings, we guided the user to focus on some interesting trends of illness number in different settings. Multiple smalls are used to show the different patterns (Figure 24). The monthly illness number is normalized by calculating the percentage of year total and is encoded by position. Therefore, users can easily identify the trends of distribution. Colors are used to highlight the different seasonal patterns: peak in the winter, peak in the summer, double peaks, and triple peaks.

Our participants could get the gist of the patterns at a glance, and compare and contrast different small multiples of graphs very conveniently given the common coordinate system.

Food and Water Agents

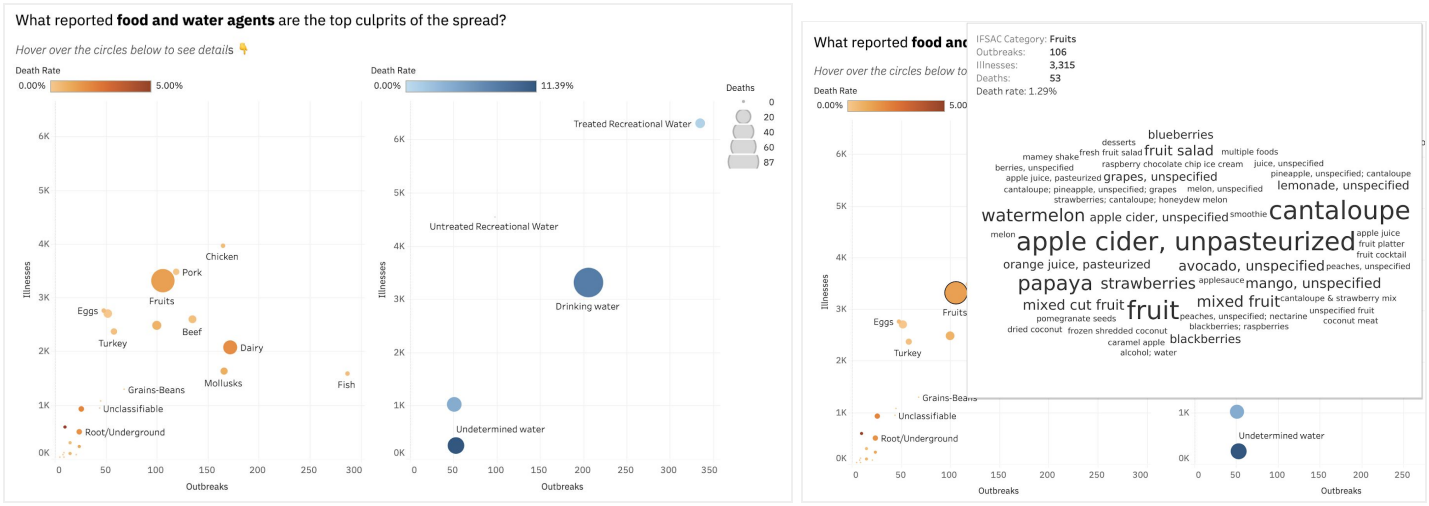


Figure 25. Profiles of Food and water agents

The next section focuses on showing which food and water agents are the top culprits of the spread of illnesses. We placed two scatterplots side by side as shown below. The positions of the circles represent the numbers of outbreaks and illnesses. The size of the circles represent the number of deaths and the color again, encodes the death rate.

As intended, the attention of our participants was drawn to the larger circles first and they found out that fruits and drinking water are the two major agents that caused the most deaths under each primary mode. If they are interested to find out more details about either a food or water category, they can hover over the circle and see the word cloud highlighting the most dangerous food or water types in the chosen category. (Figure 25)

The data encoding is very efficient and didn't overload our users even with four dimensions of quantitative variables being presented at the same time. The word cloud on hover also follows the principle of details-on-demand and provided our general user with a more direct way of making sense of the data.

Stories

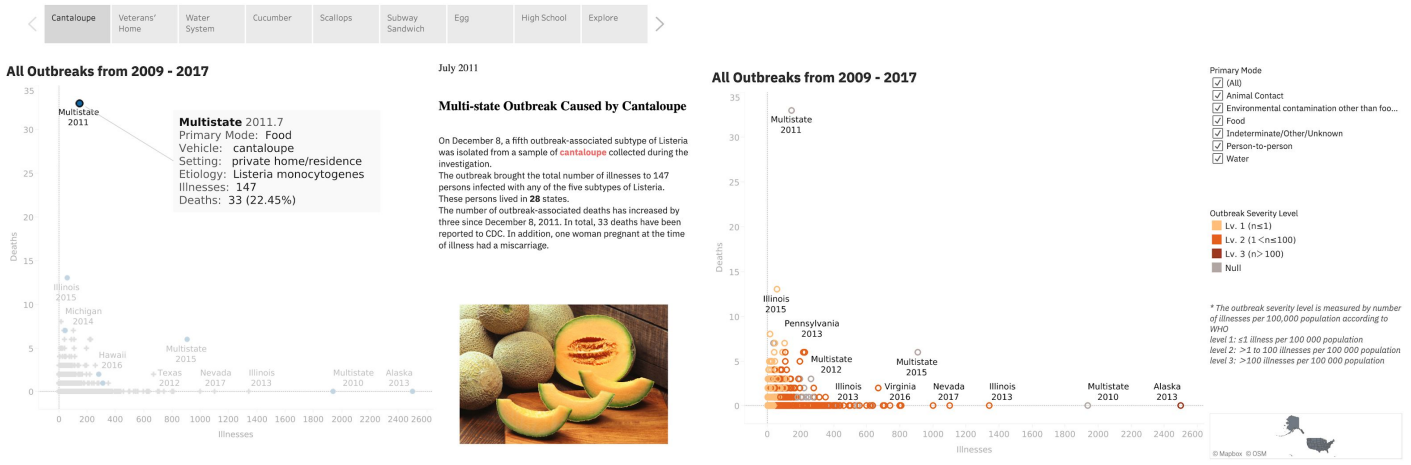


Figure 26. Stories of extreme cases (Left)
Figure 27. Exploratory dashboard of outbreaks (Right)

The last part intends to showcase news coverage of some extreme cases in the history to arouse users' interest and also alert them to food and water borne outbreaks. As shown in figure 26, the background of each story is the same scatter plots which encode illnesses and deaths with positions on the x and y axis. Color and shapes are used to highlight the 8 stories on the scatter plot. The encoding here is effective in providing the context for the stories. In terms of interaction, originally, we wanted to switch among stories by clicking the highlighted dots. However, later we faced some technical limitations of Tableau and failed to do so. As an alternative, top tabs are used to switch among different stories. To compensate for the less intuitive interaction, the information on the tabs is used as chapter titles that attract attention and direct clicking actions..

The last tab (Figure 27) on this storyboard is an exploratory section that allows users to explore more outbreaks in the history. Again, positions are used to encode two ratio variables: deaths and illnesses. Colors represent three levels of severity (World Health Organization, 2008: 61). Users can hover over the outbreak to demand for details of the outbreak including transmission mode, vehicle, settings and causes. Also, they can use filters and highlighters to extract a type or a group of outbreaks. The data encoding with mainly position is effective in terms of helping users quickly locate extreme cases and compare combining two ratio variables in their exploration. The filters and highlighters are an effective way to help users to locate cases of their interest.

Future Work

If we had more time and resources we would employ a UX Content Strategist for consistent and concise messaging for the website. Even though we explained certain terms, it might still take some time for the general public to learn. We would spend more time revising stories and content to make the overall experience more engaging.

We would also like to share this website with other public healthcare domains and social media. We are truly passionate about making this resource available for everyone who is interested in learning more past outbreak information. We want to bring more awareness to the public and potentially make an open data resources platform where we can keep adding data visualization when we can gather more datasets.

Finally, since we were constrained by time, we chose to use Tableau to create our visualization. As a result, we had to work around several limitations; for example, the visualization was quite slow when it was uploaded onto our interactive website. In the future, we would like to use D3 for its efficiency and for including additional features in order to create a more immersive and interactive visualization.

Acknowledgment

We'd like to thank Professor Cecelia Aragon and Teaching Assistant Andrea Figueroa for their guidance and feedback throughout the ten week research project.

References

1. Center for Disease Control (CDC). (2020). NORS Dashboard. Retrieved from https://www.cdc.gov/nors/data/dashboard/faq-using-dashboard.html

2. Center for Disease Control (CDC). (2014). *Surveillance for Foodborne Disease Outbreaks, US.* Retrieved from cdc.gov/foodsafety/pdfs/foodborne-outbreaks-annual-report-2014-508.pdf

3. HealthMap. (2020). Coronavirus Surveillance Map. Retrieved from: https://www.healthmap.org/en/

4. Few, Stephen. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press.

5. World Health Organization. (2008). *Foodborne disease outbreaks: guidelines for investigation and control*. World Health Organization.

Appendix

Appendix 1: Usability Testing Script

Time Visualization:

- [General understanding] How would you go about exploring this data visualization, can you tell me what is being shown in the graphs?
- [Trends] Probing question: What are the trends or patterns have you noticed?

- [Interaction] Now you want to know more about the trend of animal-borne infectious disease, what would you do?
- [What's missing] What other related information do you want to learn that is not supported by the graph?

Geo Visualization:

- [General understanding]Explore this data visualization, can you tell me what is being shown in the graphs?
- [Trends] What are the trends or patterns have you noticed?
- [Expectation] What do you think the buttons on the control panel are trying to do? What are your expectations?
- [Learnings] Now, you can have a try and tell me what you have learned and whether it's different or similar to your expectations?
- [What's missing] What other related information do you want to learn that is not supported by the graph?

Setting Visualization:

- [General understanding] Can you tell me what information the rectangle size and color represent in the two tree maps?
- [Trends] Can you tell me which setting has the highest number of illnesses and the highest death rate? How about the primary mode?
- [Expectation] What are your expectations about clicking the rectangles? How about the interaction between the two tree maps?
- [Learnings] Now, you can have a try and tell me what you have learned and whether it's different or similar to your expectations?
- [What's missing] What other related information do you want to learn that is not supported by the graph?

Etiology Visualization:

- [General understanding] Explore this data visualization, can you tell me what is being shown in the graphs?
- [Interaction] Now, you want to find out the most severe cause of disease, what would you do?
- [Correlations among graphs] (After you click the line), what do you think is being shown here?
- [Learnings] Can you tell me what you have learned?
  (if they don't understand) ask why; what they expect to see?

Stories Visualization:

- Explore Story 1, can you tell me what information is being shown on this page?
  - Follow-up: What do you think the scatter plot is trying to show here?
  - Follow-up: Does this help you understand the story?
- [Interaction/understanding] Please find out a food-borne outbreak with the second severity level that has the most deaths.