

# Analyzing Environmental Factors for Foodborne Illness and Recall Prediction

Khusniddin Abdullaev, Alex Alonso, Janessa Davis, Nicholas Filiberti,  
Xiunan Song

Mentors: Dr. Kristen Pogreba Brown & Dr. Erika Austhof

College of Information, University of Arizona

ISTA 498 - Senior Capstone

Professor Michael McKisson

April 28, 2024

## Abstract

In recent years, the dynamic interactions between environmental factors and food safety have attracted much attention in public health. In the face of escalating climate change and the continuing public health threat posed by foodborne illnesses, understanding the complex relationship between weather conditions and disease prevalence is more important than ever. This research delves into the subtle dynamics between environmental factors such as temperature, precipitation, and seasonal variations and the occurrence of foodborne illnesses, with a particular focus on pathogens such as *Salmonella*, *E. Coli*, and *Clostridium*, which exacerbate public health burdens due to their rapid spread in food and water sources.

This research focused on the relationship between weather patterns and foodborne illness in the Midwest region of the United States, focusing on major agricultural states such as Illinois, Iowa, and Wisconsin. By integrating data from the Centers for Disease Control and Prevention (CDC) and the National Weather Administration from 2001 to 2021, statistical analysis was used to uncover patterns and predictors that contribute to foodborne illness outbreaks. We use tools such as machine learning, statistical analysis tools, RStudio, and Tableau, this research focuses particularly on how precipitation and temperature changes affect the frequency and severity of *Salmonella* outbreaks.

Despite advances in detection and surveillance technologies, many challenges remain in fully understanding how specific weather patterns influence the frequency and severity of disease outbreaks. The findings of this research not only look to highlight the role of seasonal and meteorological factors in the dynamics of disease epidemics but also underline the complexity of predicting such events, particularly in terms of the variability and breadth of data. By exploring in depth the relationship between environmental variables, such as temperature fluctuations and precipitation levels and their lagged effects, and pathogen outbreaks, this study provides new insights into how climate information can be utilized to predict the risk of foodborne illness and what the future of this research could look like.

## Introduction and Background

Foodborne illness represents a major global public health challenge, spreading faster than HIV and killing as many people as malaria each year. The U.S. Federal government estimates that there are about 48 million cases of foodborne illness annually, and each year these illnesses result in an estimated 128,000 hospitalizations and 3,000 deaths (FDA, 2022). Foodborne illness impacts global health, with over 200 diseases that have a direct link to eating food contaminated with bacteria, viruses, parasites, or chemical substances. Each year, 600 million people worldwide get sick from foodborne infections leading to 420,000 deaths, according to the World Health Organization (Garcia, 2022).

Despite continued efforts to ensure the safety of the food supply, foodborne illnesses such as *Salmonella*, *E. Coli*, *Clostridium*, etc. continue to pose a significant health risk, especially when affected by climatic conditions such as temperature and precipitation. This research explores the hypothesis that certain weather patterns may significantly increase the risk of these diseases, and the analysis covers the years 2001 to 2021 using a comprehensive dataset that integrates health data from the Centers for Disease Control and Prevention (CDC) with weather and climate data extracted from the National Weather Service and related stations.

Food safety and climate change have continuously been key topics of scientific research and public concern as the agricultural industry continues to expand and daily temperatures continue to rise. With health concerns continuing to peak, scientific investigation has begun to uncover relationships between climate change and related food- and water-borne pathogens. As investigating the role of environmental elements has been a research topic for several agricultural scientists and investigators, notable links and patterns have been identified between environmental conditions and common food-borne illnesses such as *Salmonella*, *Listeria*, and non-cholera *Vibrio* (Lake, 2019). With the prediction of higher temperatures, heat waves, and storm surges, creating connections between climate conditions and outbreak patterns has improved with the introduction of surveillance tools such as the CDC NORs dashboard and quantitative microbial risk assessment (QRMA) (Schijven, 2013).

Preliminary microorganism detection through processes such as the QRMA has allowed for associations such as increased precipitation and rises in *Salmonella* to be discovered

(Schijven, 2013). Thus, research begins to illuminate the seasonality and weather trends between pathogens and environmental circumstances to have greater predictive power in the public health sphere. A greater understanding of the role of environmental and weather factors increases the recognition and priority of food safety in the agricultural production industry (Semenza, 2011). Through both contextual assessments and quantitative analysis, notable patterns have been identified between ecological circumstances and prevalent food- and water-borne pathogens in northern European regions (Semenza, 2011). Therefore, seasonality predictions have become more conceivable with pattern recognition. Weather patterns such as an increase in precipitation or a drought period due to increased temperatures can be telltale indicators for preparedness techniques that should be considered to curb future outbreaks and recalls (Lake, 2019).

Although the conversation about food safety and climate change continues to ramp up within their respective industries and fields, research and experimentation surrounding the connection between the two still offer predictive potential for future public health concerns (Schijven, 2013). Therefore, this research project will provide a new perspective by utilizing the CDC recall and outbreak data to investigate the entire farm-to-table process and those things that can disrupt and contaminate it. While research has been conducted regarding European regions and seasonality, connections here in the United States have not been understood to the same degree. Therefore, this research seeks to acknowledge and discover patterns in the US that can help illuminate outbreak increases seen throughout the years. This data analysis can offer further insight into climate changes witnessed in major US production regions and pathogen spikes to improve public health and food safety and awareness for both agricultural producers and consumers alike.

The goal of this research is to identify certain weather conditions that significantly affect the incidence of foodborne illnesses, emphasizing the potential for using weather data as a predictive tool for public health surveillance and response. By improving our understanding of how climatic factors affect food safety, this research hopes to provide actionable insights for public health policy and response strategies under global climate change, thereby increasing health and safety awareness among agricultural producers and consumers.

## Data and Methodologies

The process began with data collection from two sources: the CDC NORs Dashboard and the National Oceanic and Atmospheric Administration (NOAA) weather database. When selecting the data from the CDC dashboard, a year range of 2001 to 2021 was chosen to help maintain a more updated repertoire of outbreak reports. The CDC provided a variety of information such as recorded bacteria, number of illnesses, and primary mode of spread. After uploading the data and doing an initial investigation, the CDC data was segmented to determine which states had some of the most recorded outbreaks to determine which United States region would become the focus of analysis. After looking through the counts, states within the Midwest provided some of the highest recorded illness counts. Therefore, three states involved in the agricultural industry were chosen—Wisconsin, Illinois, and Iowa. Using these three states, weekly weather data from 2001 to 2021 was pulled from the National Oceanic and Atmospheric Administration weather database containing weekly information on the mean, minimum, and maximum temperatures in Fahrenheit, and measured precipitation, snowfall, and snow depth in inches. The four datasets were then loaded into R for preprocessing and merging.

A tibble: 6 × 21

Year <dbl>	Month <dbl>	State <chr>	Primary Mode <chr>	Etiology <chr>	Serotype or Genotype <chr>
2001	1	Tennessee	Food	Norovirus Genogroup II	unknown
2001	1	Illinois	Food	Norovirus Genogroup II	other
2001	1	Illinois	Food	NA	NA
2001	1	Illinois	Food	Norovirus Genogroup I	NA
2001	1	Illinois	Food	Salmonella enterica	Berta
2001	1	Illinois	Food	Bacillus cereus; Staphylococcus aureus	NA

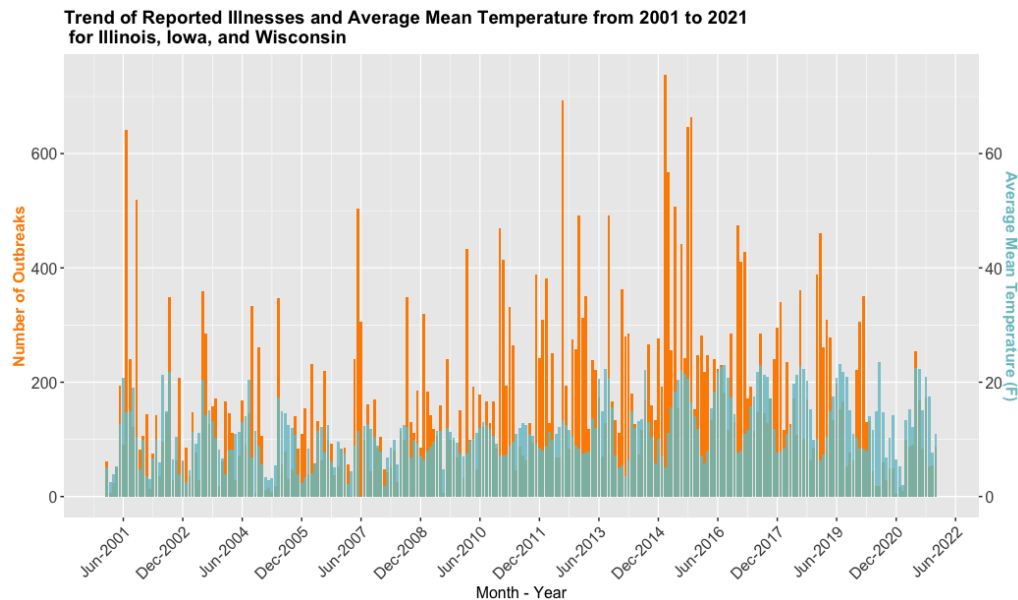
*Table 1 - Example of CDC NORs Dashboard Data*

A tibble: 6 × 7

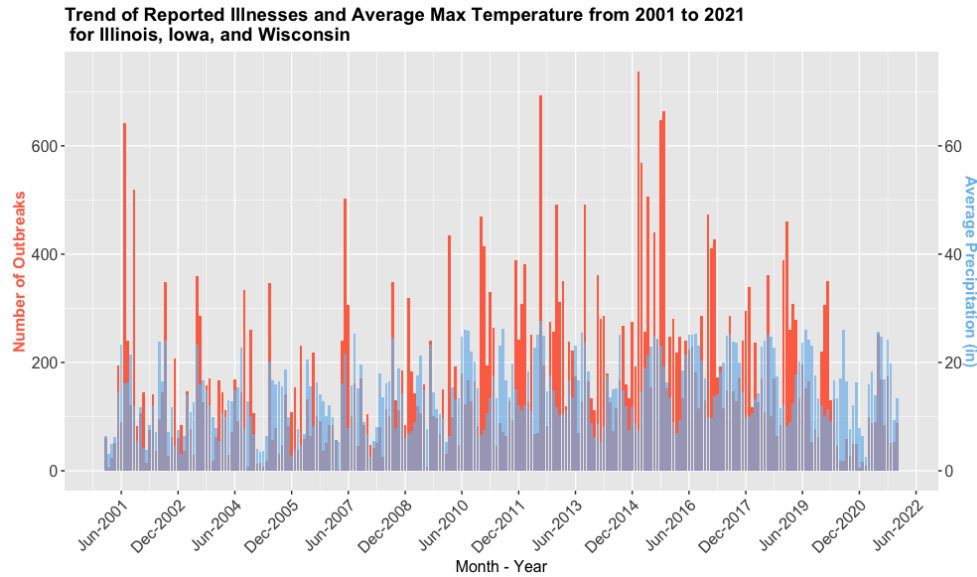
Date <date>	TAVG (Degrees Fahrenheit) <dbl>	TMAX (Degrees Fahrenheit) <dbl>	TMIN (Degrees Fahrenheit) <dbl>	PRCP (Inches) <dbl>
2001-01-03	17.5	28	7	0
2001-01-10	27.5	35	20	0
2001-01-17	24.0	28	20	0
2001-01-24	24.0	29	19	0
2001-01-31	32.5	36	29	0
2001-02-07	29.0	37	21	0

*Table 2 - Example of the NOAA Weather Database*

After a closer investigation of the CDC data, various anomalies were found in the information entries and illness count reports. Therefore, entries were transformed for grouping simplicity and outliers were removed to prevent extenuating values from having influence in future aggregation. The CDC data grouped and cleaned the weather data was merged for all three states on a weekly interval. Due to the CDC outbreaks being reported on a monthly level, the NOAA weather data was averaged at monthly intervals to then be joined with the illness counts in the CDC table. With the outbreak and weather data merged, the information could now be analyzed for initial investigation and machine learning. Example summary plots of the merged data can be found in Figures 1 and 2 below.



*Figure 1 - Trend of Reported Illnesses and Average Mean Temperature after the Data Merge*



*Figure 2 - Trend of Reported Illnesses and Average Maximum Temperature after the Data Merge*

In order to analyze and predict foodborne illnesses and outbreaks, a variety of statistical techniques were applied to the data gathered from the CDC NORs Dashboard and the National Oceanic and Atmospheric Administration weather database. The aim was to identify significant patterns and relationships between environmental factors and the number of recorded illnesses. Linear Regression was the initial model used to examine the relationship between weather patterns and recorded illnesses. This model evaluated the significance of various environmental predictors (e.g., mean temperature, precipitation) in explaining the number of outbreaks. The adjusted R-squared value was used to measure the proportion of variance in the dependent variable that could be predicted from the independent variables, while p-values determined the statistical significance of each variable. Following the fitting of the linear regression model, K-Means Clustering was employed as an unsupervised machine learning method to identify significant groupings and clusters within the monthly weather summaries. This technique aimed to uncover potential patterns or trends in the data that might not be immediately apparent. Additionally, a Time Series Analysis was used to visualize the trends in outbreak counts over time and forecast future outbreak patterns. This method helped to assess any seasonality or long-term trends in the data. As patterns and relationships were still being searched for, further subsetting was done for the top three illnesses observed in the dataset. These subsets helped fuel

the next round of analysis that involved LASSO Regression and a Regression Decision Tree. LASSO Regression was applied to each of the three primary bacteria subsets to identify key weather features that may predict illness. This method penalizes large coefficients to prevent overfitting and multicollinearity, offering a more robust model fit. The Regression Decision Tree was used to further analyze one of the top three datasets based on the subset of predictors identified by LASSO regression. This method created a predictive model to explore potential relationships between weather factors and outbreaks.



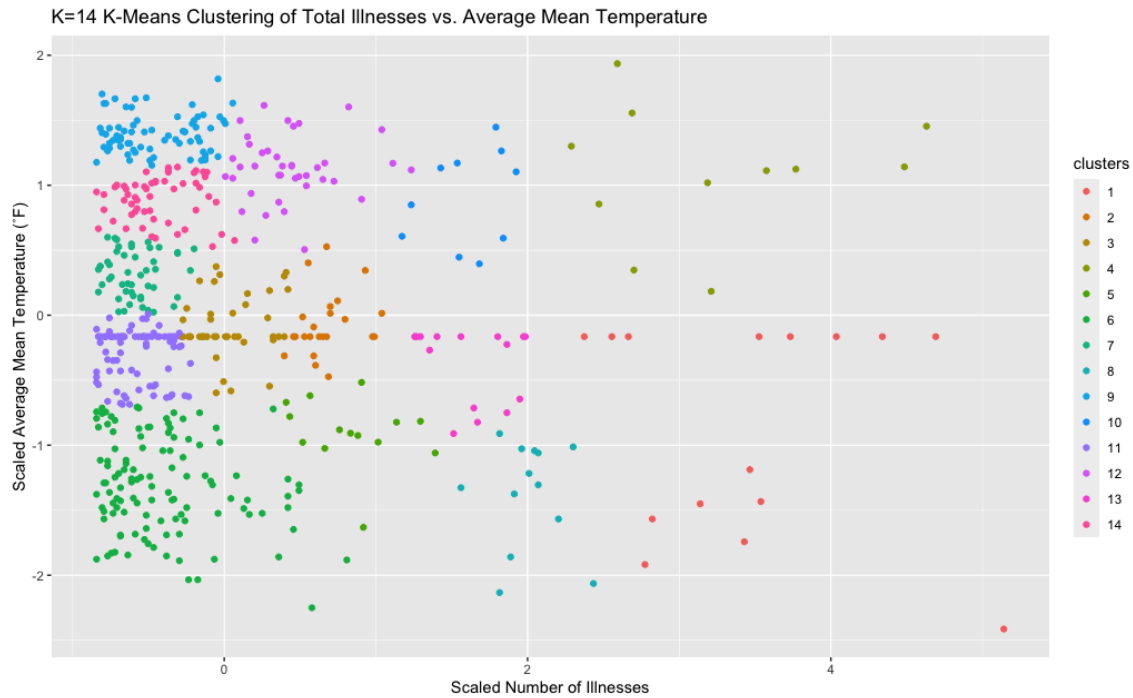
## Analysis and Findings

To begin obtaining an understanding of the relationships between the number of recorded illnesses and monthly weather patterns, a Linear Regression model was fit to discover whether any weather features were significant for predicting the number of outbreaks. The initial regression model only identified two variables significant at a 0.05  $\alpha$ -level out of the nine within the dataset—indicator variable for the state of Iowa and mean temperature (p-value = 0.01231 and 0.00994 respectively). Additionally, the model had an  $R^2$  value of 0.09665 suggesting that the current collection of variables had little impact on describing and predicting the variance within the number of illnesses. To get further information regarding the weather patterns, additional month-to-month lag variables were added for each of the weather predictors. After running another regression model with the additional lag variables, the only significant variable was the indicator for Iowa while  $R^2$  decreased marginally to 0.06553. Therefore, neither model suggested notable weather observations and prompted further analysis.

Linear Regression Fit <i>without</i> Month-to-Month Lag Variables				
Regression Statistics				
R-squared	0.09665			
Adjusted R-squared	0.07998			
Standard Error	80.82			
Observations	756			
Predictors	9			
Residuals				
Min	1st Quartile	Median	3rd Quartile	Max
-150.86	-45.07	-21.79	18.15	399.87
Linear Regression Fit <i>with</i> Month-to-Month Lag Variables				
Regression Statistics				
R-squared	0.06533			
Adjusted R-squared	0.03759			
Standard Error	115.8			
Observations	756			
Predictors	15			
Residuals				
Min	1st Quartile	Median	3rd Quartile	Max
-159.62	-55.99	-28.34	16.85	1253.47

*Table 3 - Linear Regression Summary for Weather Models with and without the Month-to-Month Lag Variables*

With further investigation into the underlying distribution of the data in relation to the number of illnesses, the dispersed scatterplot patterns between the weather predictors offered the opportunity to find significant groupings and clusters within the data pairings. Therefore, an unsupervised machine learning method, K-Means Clustering, was applied to each of the monthly summarized weather values. Using the Elbow Method, the optimal number of clusters was found to minimize the test error of the model. For each of the six pairings, the optimal number of clusters was equal to or greater than 14 as seen in Figures 3 to 8. With Elbow Method k-values applied to each pairing, the calculated groupings can be seen to determine whether significant groups can be found. Because of the large number of clusters, considerable weather patterns could not be identified.



*Figure 3 - K-Means Clustering for Total Number of Illnesses and Average Mean Temperature with  $k = 14$  Clusters*

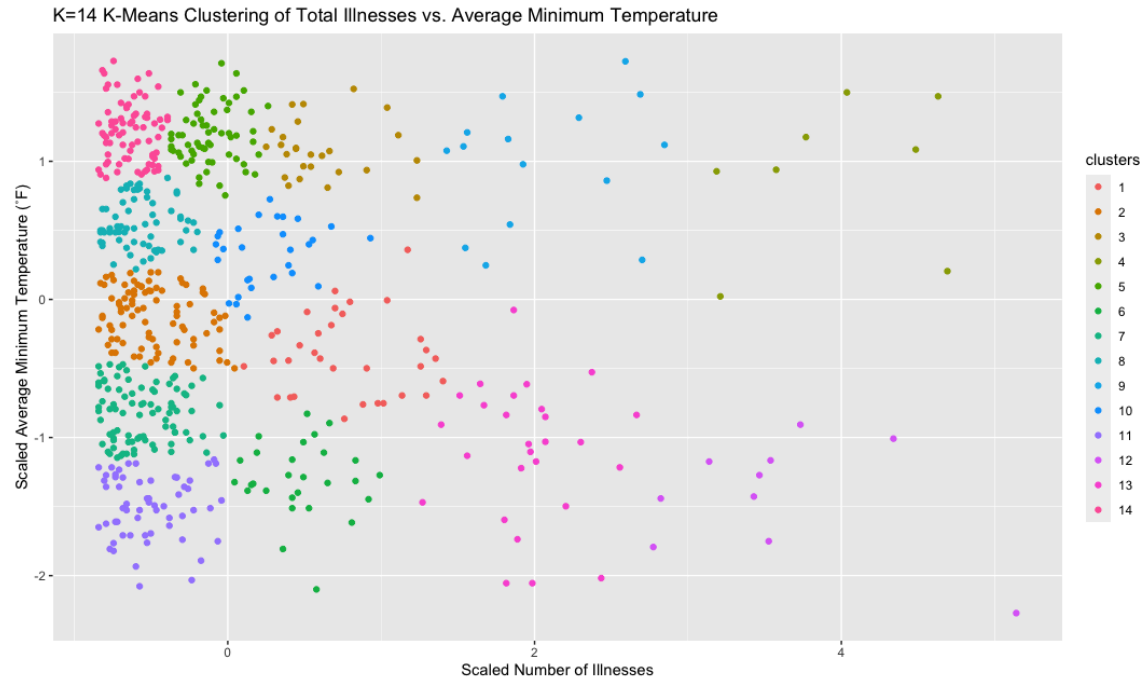


Figure 4 - K-Means Clustering for Total Number of Illnesses and Average Minimum Temperature with  $k = 14$  Clusters

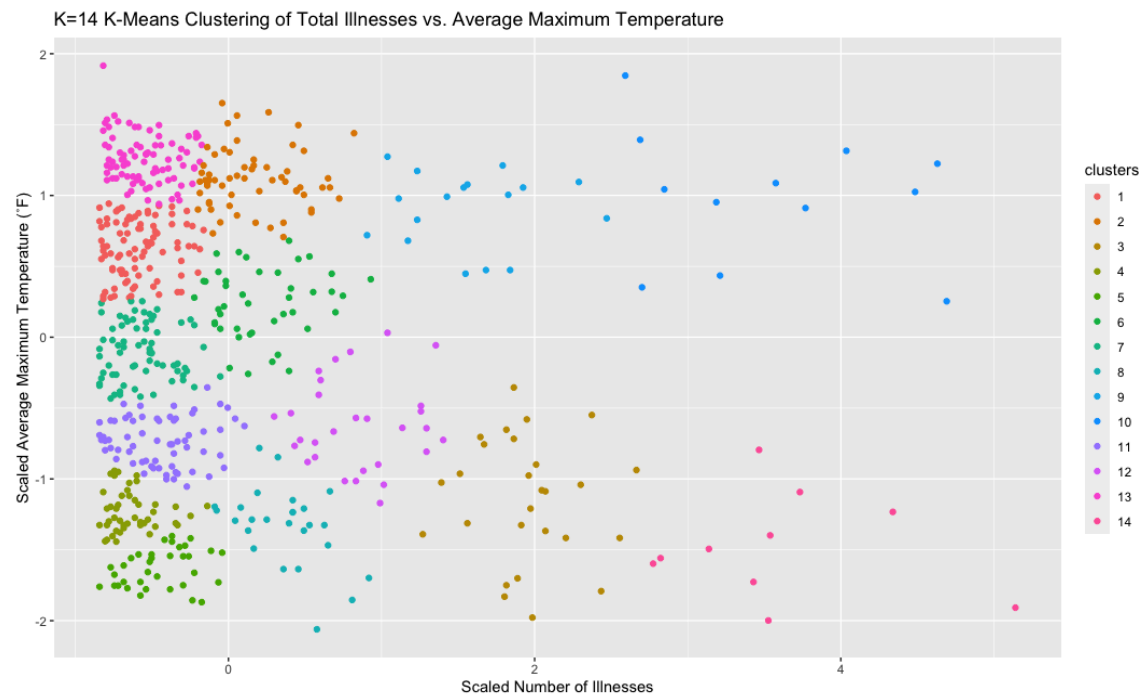


Figure 5 - K-Means Clustering for Total Number of Illnesses and Average Maximum Temperature with  $k = 14$  Clusters

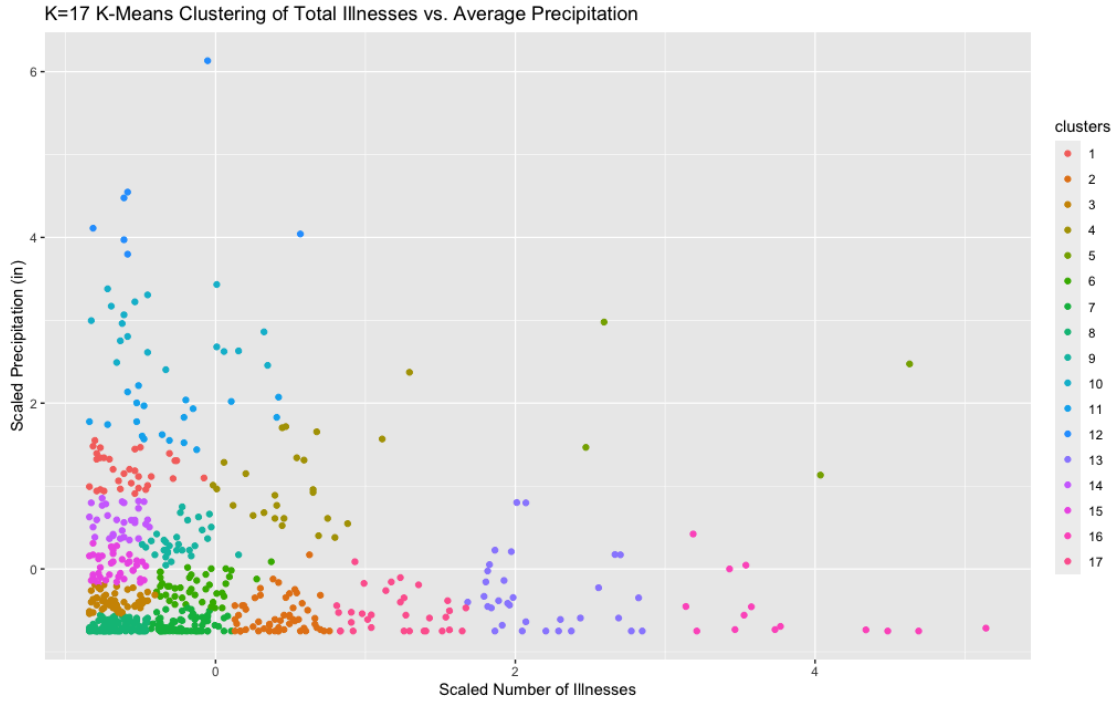


Figure 6 - K-Means Clustering for Total Number of Illnesses and Average Precipitation with  $k = 17$  Clusters

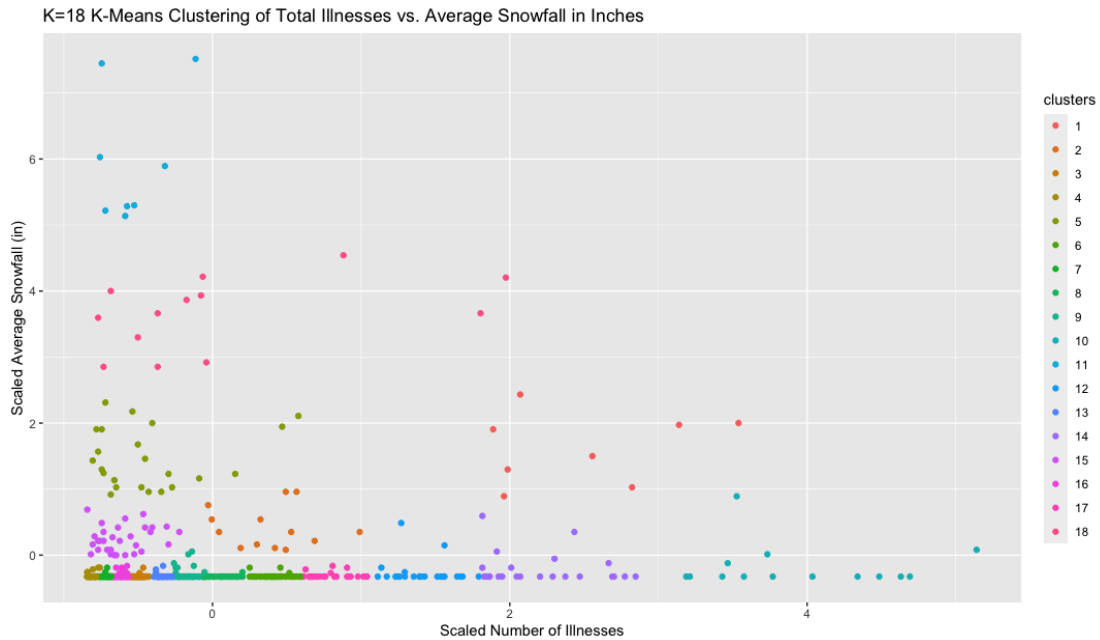


Figure 7 - K-Means Clustering for Total Number of Illnesses and Average Snowfall with  $k = 18$  Clusters

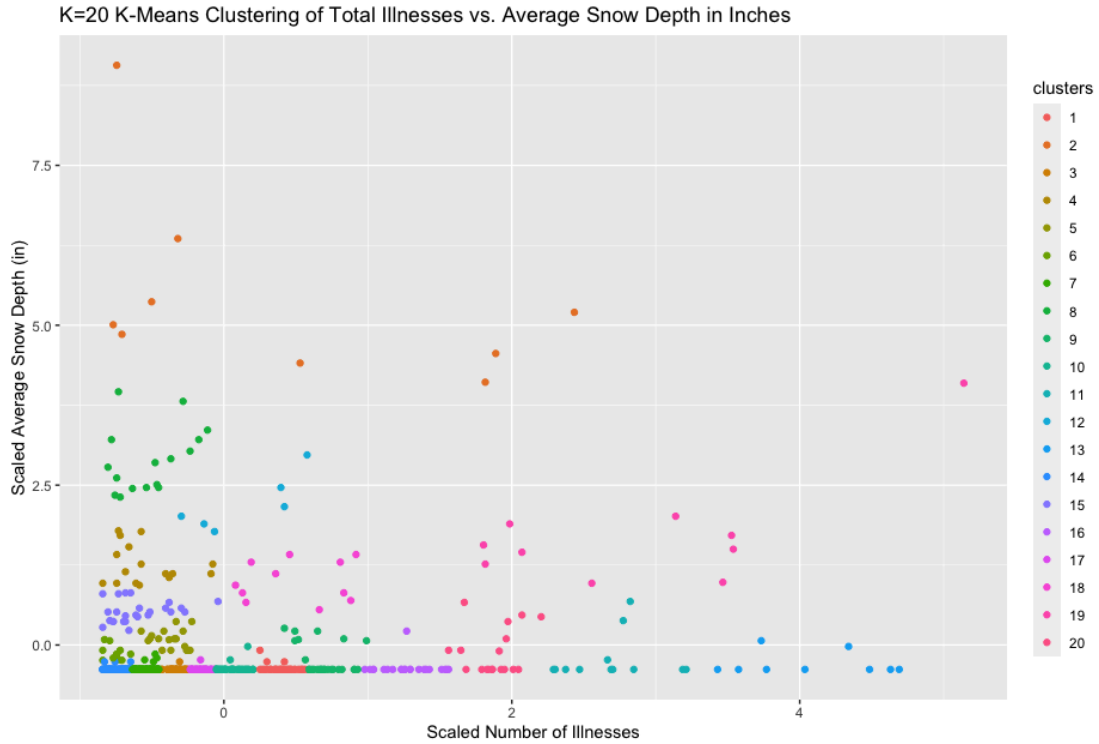


Figure 8 - K-Means Clustering for Total Number of Illnesses and Average Snow Depth with  $k = 20$  Clusters

Consequently, without notable clustering or patterns found within the compilation of recorded illnesses, the analysis shifted to drilling down further into the top three outbreak bacteria recorded within the CDC NORs. After aggregating the data, the top three outbreaks within the dataset that have a history of being associated with their environments were *Salmonella*, *Clostridium*, and *E. Coli*. Using these three illnesses, the merged illness and weather data was subsetting for each of the top three bacteria. To begin the search for seasonality patterns for each bacteria strain, a linear regression model was fit on each subset to initial identification of significant weather predictors as found in Table 4. The *Salmonella* model had one significant predictor, Year with a p-value of 0.0211, along with an  $R^2$  value of 0.07553 and a moderate Residual Standard Error (RSE) of 64.21. The *E. Coli* model had zero significant predictors with an  $R^2$  value of 0.07553 and a moderately large RSE of 34.15. The *Clostridium* model also had zero significant predictors with an  $R^2$  value of 0.06811 and a large RSE value of 84.45. The small  $R^2$  values and large RSE and p-values suggest a poor fit of the model on the dataset and

extensive variation occurring between the patterns and combinations of predictors when trying to predict the number of illnesses.

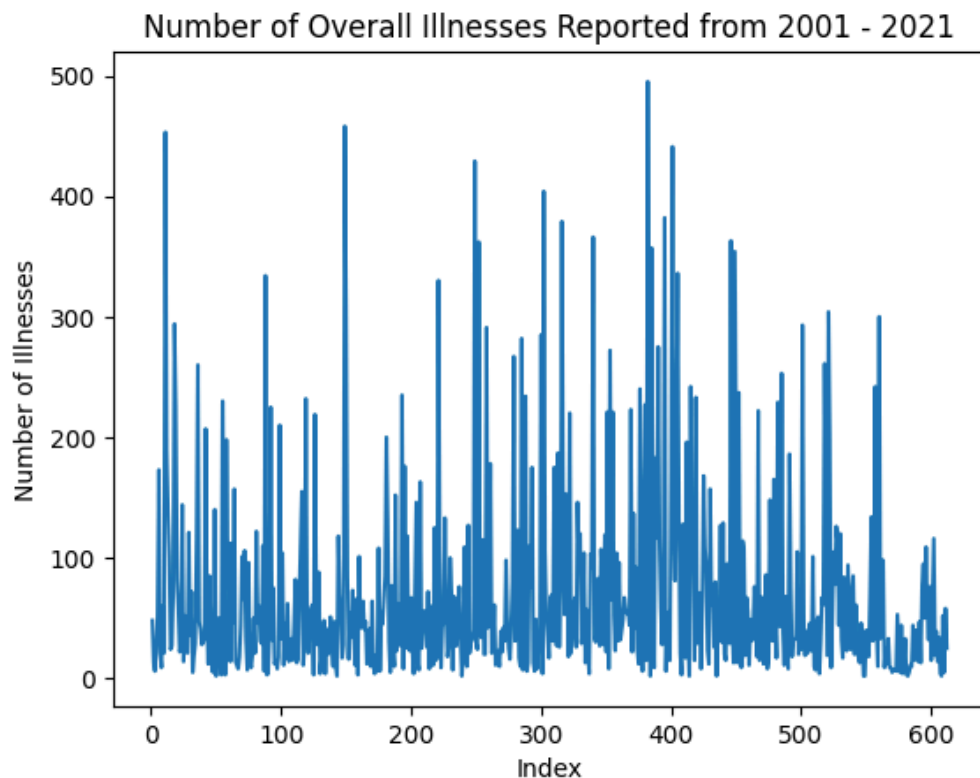
Salmonella Linear Regression Fit				
Regression Statistics				
R-squared	0.07553			
Adjusted R-squared	0.001204			
Standard Error	64.21			
Observations	756			
Predictors	15			
Residuals				
Min	1st Quartile	Median	3rd Quartile	Max
-58.33	-22.99	-9.3	5.36	748.37

E. Coli Linear Regression Fit				
Regression Statistics				
R-squared	0.08756			
Adjusted R-squared	-0.08834			
Standard Error	34.15			
Observations	756			
Predictors	15			
Residuals				
Min	1st Quartile	Median	3rd Quartile	Max
-28.834	-13.449	-5.287	4.297	257.81

Clostridium Linear Regression Fit				
Regression Statistics				
R-squared	0.06811			
Adjusted R-squared	-0.044			
Standard Error	84.45			
Observations	756			
Predictors	15			
Residuals				
Min	1st Quartile	Median	3rd Quartile	Max
-60.7	-29.14	-12.58	12.24	856.52

*Table 4 - Linear Regression Summary for Weather Models for the Top 3 Illnesses found*

A Time Series model was applied to the complete and top three illness subsets to search for any additional insight into the weather and outbreak patterns seen over the 20-year span. The Times Series provided a visual representation of the outbreak counts throughout the years as seen in Figures 9 to 13. Utilizing the weather trends found in the dataset, the forecasting sequence provided ten additional predictions that correspond to January to December of 2022. The following forecasted values for each data group can be found in Table 5. Each forecast takes on exponential shape with the forecast values peaking in January and declining and eventually, plateauing at February, March, or May. The poor forecasting points to high variability and noise within the current set of weather variables that are preventing valuable predictions from being obtained. Although the Time Series forecast provides limited information, it helped guide the research to subset and test for those predictors that are significant for estimation. Therefore, this opened the door for further analysis and variable selection within the larger and three smaller subsets of illness counts.



*Figure 9 - Time Series Model of All Reported Illnesses*

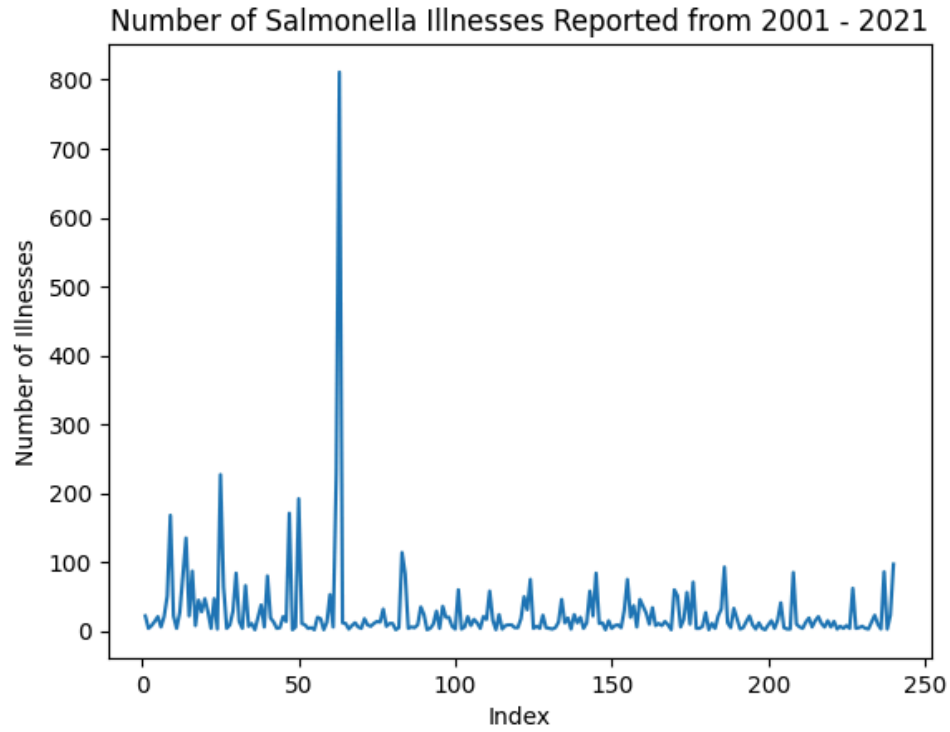


Figure 10 - Time Series Model of All Reported Salmonella Illnesses

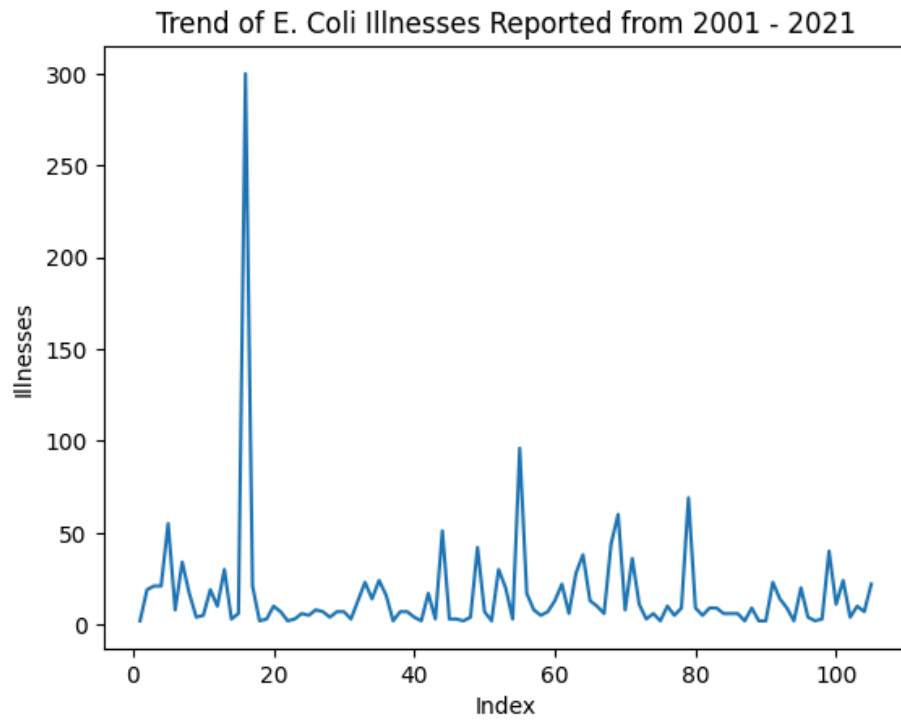


Figure 11 - Time Series Model of All Reported E. Coli Illnesses



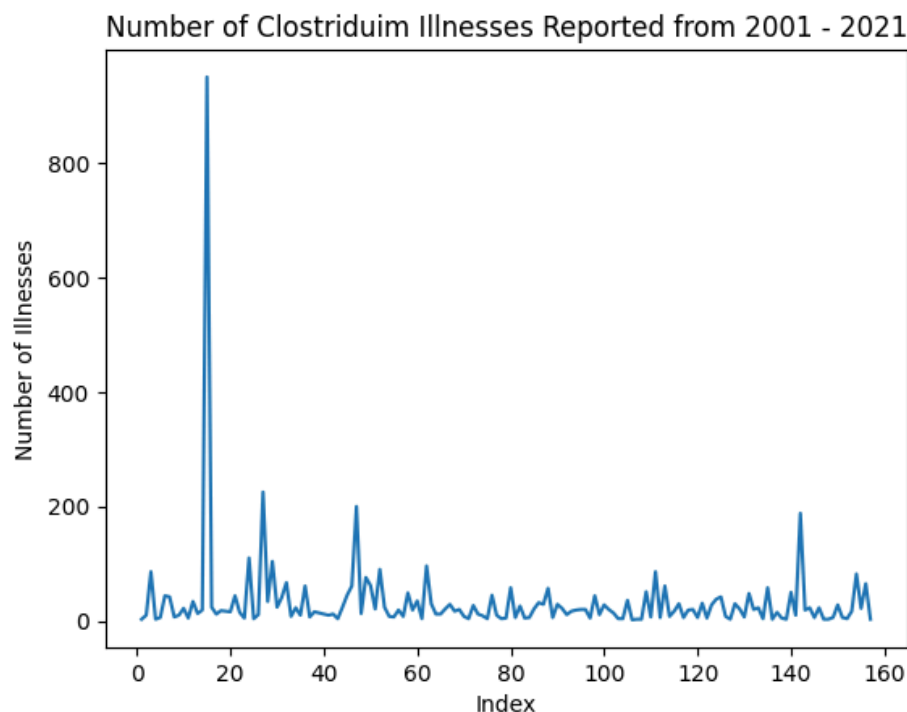


Figure 13 - Time Series Model of All Reported Clostridium Illnesses

Forecasted Illnesses 2022				
	Full Set of Illnesses	Salmonella	E. Coli	Clostridium
January	43.500655	33.296804	16.583069	26.863816
February	42.712219	22.690855	16.568294	25.916282
March	42.745819	20.92507	16.568294	25.953905
April	42.744388	20.631084	16.568294	25.952411
May	42.744449	20.582139	16.568294	25.95247
June	42.744446	20.57399	16.568294	25.952468
July	42.744446	20.572633	16.568294	25.952468
August	42.744446	20.572407	16.568294	25.952468
September	42.744446	20.57237	16.568294	25.952468
October	42.744446	20.572363	16.568294	25.952468
November	42.744446	20.572362	16.568294	25.952468
December	42.744446	20.572362	16.568294	25.952468

Table 5 - Linear Regression Summary for Weather Models for the Top 3 Illnesses found

Due to a great deal of noise occurring within the linear regression model and minimal insights through the time series, a more restrictive and computationally costly method was applied to determine the weather features that are notable for illness prediction using conservative mathematical parameters. Therefore, LASSO Regression was applied to each of the bacteria subsets to identify which variables have coefficient values that reduce to zero while managing the possibilities for overfitting and multicollinearity that might occur between predictors. Using the cross-validation and an 80-20 training and test split, a model that minimizes test error was found for *Salmonella*, *Clostridium*, and *E. Coli*. Out of the three LASSO fittings, the only model that found significant weather features was *Salmonella*. The following variables and coefficient estimates were found for the trend of *Salmonella* illnesses reported in the last 20 years.

<b>LASSO Regression Model - Coefficients and MSE</b>				
	Full Set of Illnesses	<i>Salmonella</i>	<i>E. Coli</i>	<i>Clostridium</i>
Intercept	95.6288	3041.874869	16.57143	32.84713
Year	-	-1.5028434	-	-
Month	-0.3953514	-1.900154	-	-
State	-	-	-	-
Avg Mean Temperature	-	-	-	-
Avg Max Temperature	-0.2982835	-	-	-
Avg Min Temperature	-	-	-	-
Avg Precipitation	-	-19.8192519	-	-
Avg Snowfall	-	8.0240629	-	-
Avg Snow Depth	1.323132	-1.7726657	-	-
Avg Lag in Mean Temperature	0.0002979134	-	-	-
Avg Lag in Max Temperature	-	0.3893929	-	-
Avg Lag in Min Temperature	-	-	-	-
Avg Lag in Precipitation	-37.28636	-13.1709709	-	-
Avg Lag in Snowfall	-6.645793	-	-	-
Avg Lag in Snow Depth	-	-0.4236402	-	-
<b>Mean Square Error</b>	8116.67	1199.772	300.9638	626.7394

Table 6 - LASSO Regression Coefficient Estimates and Mean Square Error

The following coefficients in Table 6 point to possible patterns that have occurred over time as outbreak surveillance has improved. The negative coefficient associated with Year and Month predictors suggests a decrease in the number of recorded illnesses over time as surveillance has improved. Additionally, a possible seasonality trend throughout the years demonstrates a decrease in *Salmonella* illnesses recorded throughout the year. The monthly and month-to-month precipitation and snow depth variables also discovered a decline in the number of outbreaks. However, the snow and month-to-month maximum temperature found a positive relationship with the recorded *Salmonella* illnesses from 2001 to 2021. The final model produced a Mean Squared test error of 1199.72; thus, suggesting a highly variable and loose fit of the data. Therefore, the LASSO regression resulted in a poor model for illness prediction, but the method provided the opportunity for subset selection of the key predictors in the *Salmonella* dataset. This reduced model helped form the matrix of predictors that would be utilized to fit a Regression Decision Tree model.

Utilizing the subset of predictors found by the LASSO regression model, a Regression Decision Tree was fit to the *Salmonella* dataset to create a beginning framework for a predictive model for outbreaks. The tree R package was used to test and train the model; an initial fit was applied using the LASSO subset that resulted in a 7 terminal node tree with training residual mean deviance of 971.4 and a mean squared test error of 115.4061 as seen in Figures 14 and 15 and Table 7. The large values of these two error values suggest a moderately weak fit to the dataset and a great deal of unexplained variance occurring within the dataset; therefore, with the hopes of improving the fit of the regression tree, a cost complexity pruning was applied to the subset. After the pruning and 10-fold cross-validation were applied, the algorithm found that the original model provided the best model to minimize the variance between the predictors. Based on the currently available CDC data, the following decision tree provides a baseline model for predicting the number of *Salmonella* outbreaks based on previous Midwest weather trends. Thus, initial observations can be used to catapult future research. A notable observation is the right-hand side of the tree for outbreaks that were in 2004 and beyond; following the month split, the diagram continues to divert into precipitation and month-to-month maximum temperature splits. The increased mean prediction of outbreaks points to a 20-year pattern where drier seasons tend to see a higher number of reported *Salmonella* illnesses. Additionally, the split at the month-to-month max temperature for lags less than 63.375° F predicts higher outbreak

counts; hence, suggesting months and seasons with smaller lags and similar temperature patterns predict more reported illnesses. Although the decision tree model did not provide the best fit and fairly wide error margins, it provides valuable insights into *Salmonella* outbreak patterns that warrant deeper investigation.

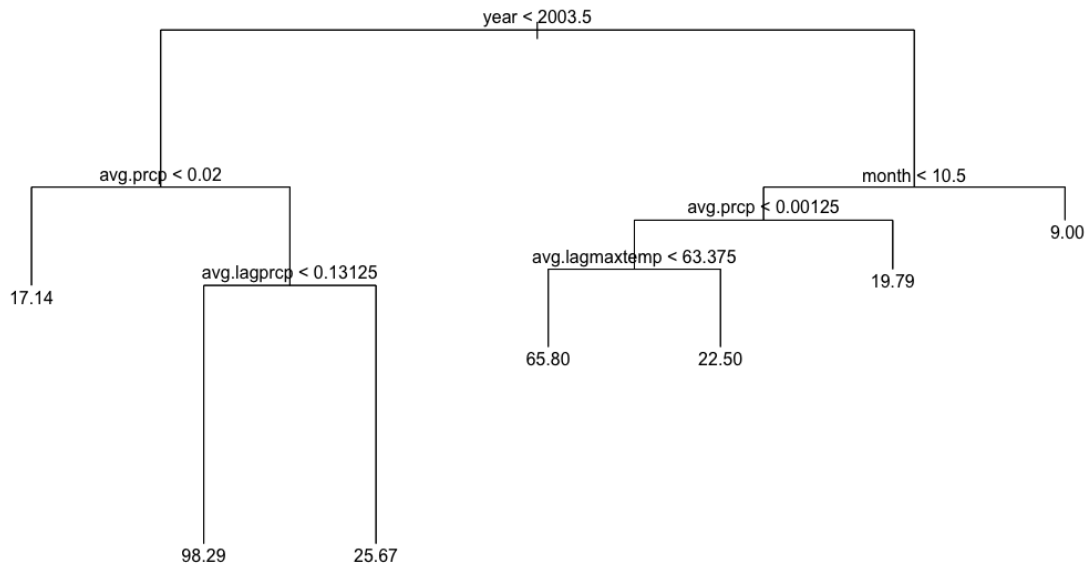


Figure 14 - Original Regression Decision Tree for Salmonella Outbreak Prediction

Regression Decision Tree Summary					
Regression Statistics					
Residual Mean Deviance	971.4				
Mean Squared Error	115.4061				
Numer of Terminal Nodes	7				
Residuals					
Min	1st Quartile	Median	Mean	3rd Quartile	Max
-84.29	-14.79	-7.786	0	3.124	172.2

Table 7 - Regression Decision Tree Summary Statistics

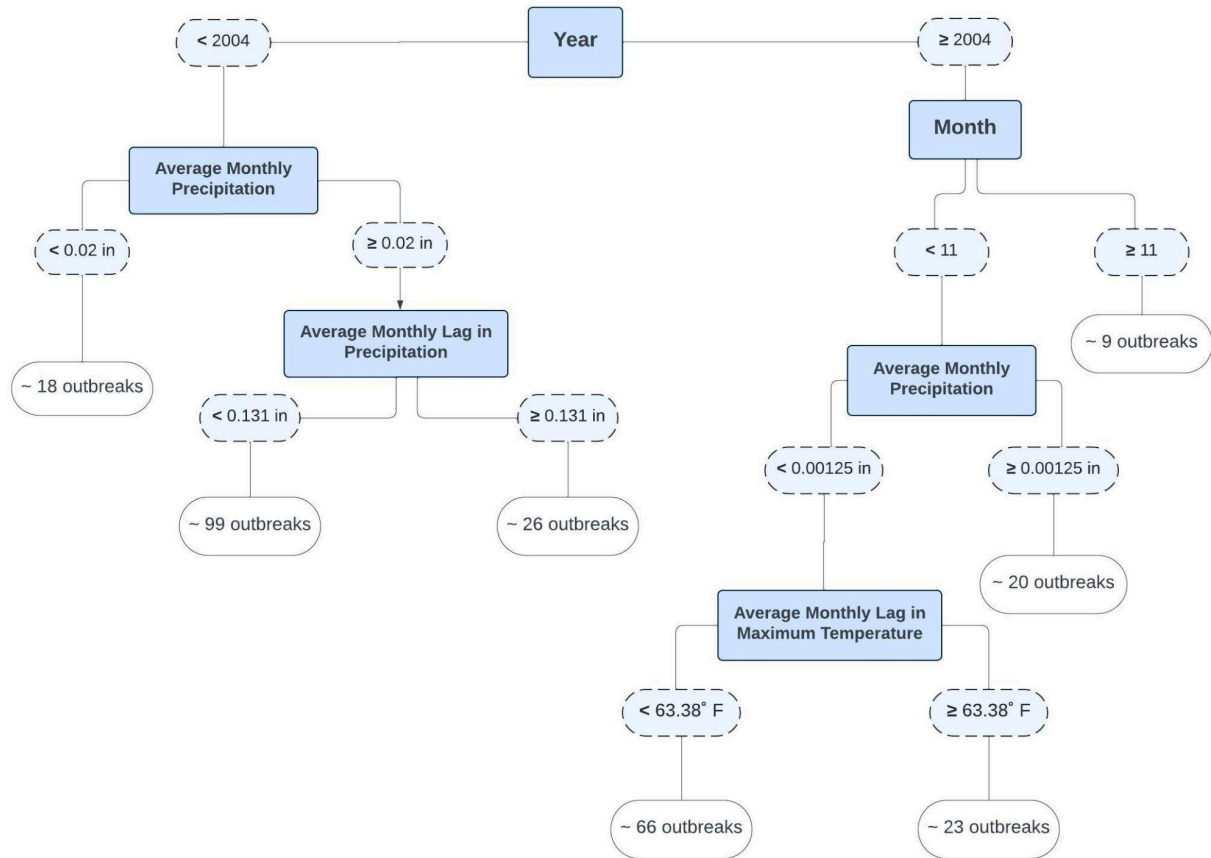


Figure 15 - Final Regression Decision Tree for Salmonella Outbreak Prediction

## Discussion and Limitations

After completing a series of supervised and unsupervised machine learning methods, foundational insights into foodborne illness patterns were gained. Nevertheless, several limitations were identified that may impact future analysis and exploration within this area. The decision tree model, while foundational, provided initial insights into potential relationships between climate change and Salmonella cases. For instance, the trend of more predicted illnesses during months with reduced precipitation points to a notable seasonality trend across the Midwest. The pattern suggests further investigation into the potential correlation between drier summer months and increased Salmonella occurrences, given that the optimal growth range temperature for the bacteria is around 98.6°F. While these findings may initially appear minimal in comparison to the broad scope of the dataset, they offer a crucial starting point for more targeted future research and data modeling. Additionally, the observed trends may guide the development of strategic measures for mitigating foodborne illness during peak risk periods.

However, to expand upon these insights, the data limitations encountered during analysis must be addressed. Despite advances in surveillance, significant challenges remain in understanding the complex relationship between weather conditions and foodborne illness outbreaks. One key issue is the presence of missing data, which can introduce biases and compromise the reliability of statistical analyses. Although methods exist to handle missing data, they often require assumptions that may impact the accuracy of outcomes. Another limitation arises from variability in outbreak reporting practices across different regions, leading to data inconsistencies that complicate trend analysis. Extensive time and effort were required to preprocess the data and resolve reporting discrepancies. Additionally, the CDC data's monthly frequency constrained the depth of analysis, necessitating monthly averaging to align it with weather data. This simplification limited the detection of trends at a more granular, weekly level, potentially overlooking significant anomalies. Human input and reporting introduce the possibility of errors and biases, further emphasizing the need for improved data quality and analytical precision. Addressing these challenges is vital for enhancing the robustness and depth of future analyses.

Beyond the data and analytical aspects, this research emphasizes the importance of improving outbreak surveillance and data collection practices. Access to more detailed,

high-resolution recall data could enable more precise data merging and analysis. Analyzing recall and weather data on a daily or weekly basis would yield more accurate insights into potential patterns. These findings underscore the need for enhanced surveillance systems and data collection methods to provide comprehensive and high-quality data for in-depth analysis. Improved data quality and surveillance practices could lead to a better understanding of the dynamic interactions between weather patterns and foodborne illness outbreaks, ultimately informing more effective interventions. The public health sector stands to gain significantly from these insights. With access to more comprehensive data and advanced analytical techniques, public health officials can proactively identify and respond to emerging foodborne illness outbreaks. By developing more accurate predictive models, agencies can allocate resources more efficiently and implement targeted prevention strategies.

Furthermore, integrating weather data with outbreak surveillance could offer new perspectives on potential risk factors for foodborne illnesses. This integrated approach can enhance early detection and improve the speed and effectiveness of public health responses. Advancing the precision and granularity of data collection could lead to significant improvements in outbreak tracking and investigation. Public health agencies may benefit from partnerships with meteorological services to access more detailed climate data for analysis. Additionally, collaborations with other sectors, such as agriculture and food production, may provide valuable insights into the environmental factors contributing to foodborne illness outbreaks. Overall, this research opens the door for transformative improvements in public health surveillance and response. By leveraging advanced data collection and analytical techniques, public health authorities can better protect communities from foodborne illnesses and enhance food safety measures.

## Tableau Dashboard

To round out the analysis, an interactive Tableau dashboard was created to summarize the research findings. This dashboard serves as a powerful tool for presenting our research in an engaging and accessible manner. By providing a visually immersive and user-friendly experience, the dashboard allows stakeholders to explore our findings and gain a deeper

understanding of the complex relationships between weather patterns and foodborne illness outbreaks. The dashboard offers a narrative-driven journey through our research, with carefully crafted visualizations and accompanying descriptions that guide users through key insights and trends. Stakeholders can interact with the data by adjusting filters and parameters, allowing them to examine specific aspects of the research and tailor their experience to their interests.

Visualizations such as scatterplots and time series plots enable users to quickly grasp the patterns and correlations uncovered in our analysis. These graphical representations provide a compelling story that can resonate with a diverse audience, from public health officials to researchers and policymakers. Through the dashboard, users can engage with our research on multiple levels, fostering a greater appreciation of the nuances and implications of our findings. This dynamic tool not only communicates our research effectively but also opens the door for further discussion and collaboration among stakeholders.



## Sources

- Center for Food Safety and Applied Nutrition. “What You Need to Know about Foodborne Illnesses.” *U.S. Food and Drug Administration*, FDA, 17 Feb. 2022, [www.fda.gov/food/consumers/what-you-need-know-about-foodborne-illnesses](http://www.fda.gov/food/consumers/what-you-need-know-about-foodborne-illnesses).
- Garcia, Posted by Jesus. “World Food Safety Day Spotlights Global Health Threat.” *USDA*, 7 June 2022, [www.usda.gov/media/blog/2022/06/07/world-food-safety-day-spotlights-global-health-threat#:~:text=Foodborne%20illness%20impacts%20global%20health,to%20the%20World%20Health%20Organization](http://www.usda.gov/media/blog/2022/06/07/world-food-safety-day-spotlights-global-health-threat#:~:text=Foodborne%20illness%20impacts%20global%20health,to%20the%20World%20Health%20Organization).
- Semenza, J. C., Herbst, S., Rechenburg, A., Suk, J. E., Höser, C., Schreiber, C., & Kistemann, T. (2012). Climate Change Impact Assessment of Food- and Waterborne Diseases. *Critical reviews in environmental science and technology*, 42(8), 857–890. <https://doi.org/10.1080/10643389.2010.534706>
- Schijven, J., Bouwknegt, M., de Roda Husman, A.M., Rutjes, S., Sudre, B., Suk, J.E. and Semenza, J.C. (2013), A Decision Support Tool to Compare Waterborne and Foodborne Infection and/or Illness Risks Associated with Climate Change. *Risk Analysis*, 33: 2154-2167. <https://doi.org/10.1111/risa.12077>
- European Centre for Disease Prevention and Control (Ed.). (2016, June). ECDC quantitative microbial risk assessment (QMRA) tool. *Climate Adapt*. Retrieved March 10, 2024, from <https://climate-adapt.eea.europa.eu/en/metadatas/tools/ecdc-quantitative-microbial-risk-assessment-qmra-tool>
- Lake, I. R., & Barker, G. C. (2018). Climate Change, Foodborne Pathogens and Illness in Higher-Income Countries. *Current environmental health reports*, 5(1), 187–196. <https://doi.org/10.1007/s40572-018-0189-9>
- Centers for Disease Control and Prevention. (n.d.). CDC National Outbreak Reporting System (NORS) Dashboard. <https://wwwn.cdc.gov/norsdashboard/>
- National Oceanic and Atmospheric Administration. (n.d.). Climate at a Glance. <https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/national>