Exploring Patterns in Clinical and Non-Clinical Listeria monocytogenes Isolates

Alexis Gonzalez

**Authors**

* Alexis Gonzalez

**Author affiliations**

1. College of Public Health, University of Georgia, Athens, GA, USA.

# 1. Summary/Abstract

Listeria monocytogenes is major food born pathogen. Though infections are rare, L.monocytogenes has the highest hospitalization rate and death rate of the common food borne illnesses. This pathogen can proliferate in a number of hosts and climates. Understanding temporal and source relationships may allow the public to be better informed on how to protect themselves. This study used whole-genome sequencing data from the NCBI Pathogen Detection database (2010–2024) to explore trends in food, clinical, and environmental isolates of *L. monocytogenes*. We conducted exploratory data analysis to assess the frequency and geographic distribution of isolates. Poisson and logistic regression models were applied to evaluate the effect of seasonality on clinical isolate occurrence. 33,264 isolates were analyzed, with a majority of them being environmental. Geographic analysis concluded that New York state contributed a large percentage of food isolates, and many illness-causing foods originated in Sweden. Statistical modeling showed a consistent seasonal trends. Poisson and negative binomial regression models were used to evaluate the predictive value of seasonality and isolate source type on clinical cases. While clinical isolates exhibited seasonal trends with higher counts in warmer months, the predictive power of the models was weak (R² ≈ 0.17), and performance was similar across Poisson and negative binomial models. These findings suggest that although temporal patterns exist, seasonality alone is insufficient for accurately predicting clinical cases. Enhanced data collection and standardization may improve future outbreak prediction efforts.

# 2. 1. Introduction

Listeria monocytogenes is a ubiquitous, gram-positive, intracellular pathogen renowned for its adaptability and resilience[3]. This bacterium can be found in a wide range of environments such as soil, water, animals, and vegetation. Due to its ability to proliferate in food products and low temperatures, it is considered a major food-borne illness. The ingestion of Listeria-infected food may lead to listeriosis, a gastrointestinal illness that causes vomiting, diarrhea, and even death. It is of even greater concern in pregnant individuals, as its forceful invasion tactics allow the bacterium to cross the placental barrier, infecting the fetus; this can lead to miscarriages and stillbirths[6]. While immunocompetent individuals may only experience mild gastroenteritis, disseminated infection in the immunocompromised has a lethality rate of ~30%.  Listeriosis causes 1,600 infections and 260 deaths each year with a 95% hospitalization rate[1]. The United States Department of Agriculture lists 6 foods that are at high risks for L.monocytogenes contamination; unpasteurized dairy products, soft cheeses, raw fruits and vegetables (particularly sprouts, enoki mushrooms, and melons), Ready-To-Eat deli meats and hot dogs, refrigerated pate, and refrigerated smoked seafood. Listeria contamination typically happens post-lethality which exposes a ready-to-eat product to a processing environment after a form of lethality treatment has taken place. Post lethality processing environments can be deli slicers, submersion in a brine solution, meat encasing, bagging or packaging. The USDA recommends all establishments to take part in additional post lethality treatment such as antimicrobial coatings, pasteurization, or chemical or physical sanitation of equipment[3]. Sanitation tends to be the more cost effective option for many establishments but, due to incorrect sanitation practices and L.monocytogenes robust bio-film formation capabilities sanitation does not always prevent contamination. Outbreaks have been seen in a range of products in recent years. From 2023 to 2025 there have been over 10 multi-state listeria outbreaks[2]. The most recent being an outbreak beginning February 22, 2025 in frozen supplement shakes. These shakes were distributed among institutional facilities such as long term care facilities and hospitals. This outbreak has resulted in 38 cases, 37 hospitalizations, and 12 deaths so far as this investigation is ongoing as of April 2025[5].

This study aims to investigate trends in listeria isolates found in clinical and non-clinical samples in efforts to predict cases. This study will investigate time of year and number of non-clinical isolates as a predictor for clinical isolates , non-clinical isolates being isolates that originate in food products and environmental samples. The data source being used to conduct this study is the National Center for Biotechnology Information’s Pathogen Detection browser. Pathogen Detection is a culmination of pathogen genomic sequences from federal, state, private, and foreign laboratories. It is utilized constantly by public health officials,epidemiologists, and scientists to identify chains of transmission. The new isolates are analyzed in real time to aid in that identification by developing clusters. Listeria monocytogenes is a nationally notifiable disease so, many laboratories opt to share their data nationally through the Pathogen Detection Browser. By utilizing the most up to date data and identifying trends, this study hopes to provide information to form new guidelines on protecting the public from Listeriosis infection.

# 3. 2. Methods

This study utilized publicly available data through NCBI Pathogen Detection. Before downloading the data from the Isolate Browser, the data was filtered to only include Listeria monocytogenes isolates found in or originating in the United States of America. This resulted in 33,264 L.monocytogenes isolates. Once loaded into R the data set was minimized further by selecting isolates that were entered between January 1, 2010 and December 31, 2024. Several variables were removed from the dataset as they did not contribute to the analysis. These were isolate identification and clustering data such as Serovar, Biosample number, Assembly, Min-same and Min-diff SNPs, computed types, and a variable called …18 ,presumably referring to 18s primers, that contained no data.

Two additional columns were created for exploratory analysis to separate clinical and non clinical isolates from the “Isolation type variable”. Entries for the “Clinical” column were coded as “Yes” if the entry for “Isolation type” was clinical. All non-clinical isolates were coded as “environmental/other” so the “Isolation Source” column was utilized to create the “Food” column. All isolates that came from various meat and dairy, and produce food sources were coded as “Yes”. A final source column called “isolation\_type” was created by assiging “food” to food->“Yes” and “clinical” to clinical -> “Yes”, other entries were filled in as “environmental”. Additional columns created were “days\_since\_start” which converted all dates in the selected time period into days since the first date in the data set. A column for year, day, month, and season were also created for statistical analysis.

Poisson and negative binomial modeling were conducted for temporal analysis using predictors such as season and counts of food/environmental isolates to model monthly clinical isolate counts. Overdispersion was assessed using residual deviance and addressed with negative binomial models. Model performance was evaluated using RMSE, MAE, and R-squared values. All plots were generated using ggplot2

# 4. Results

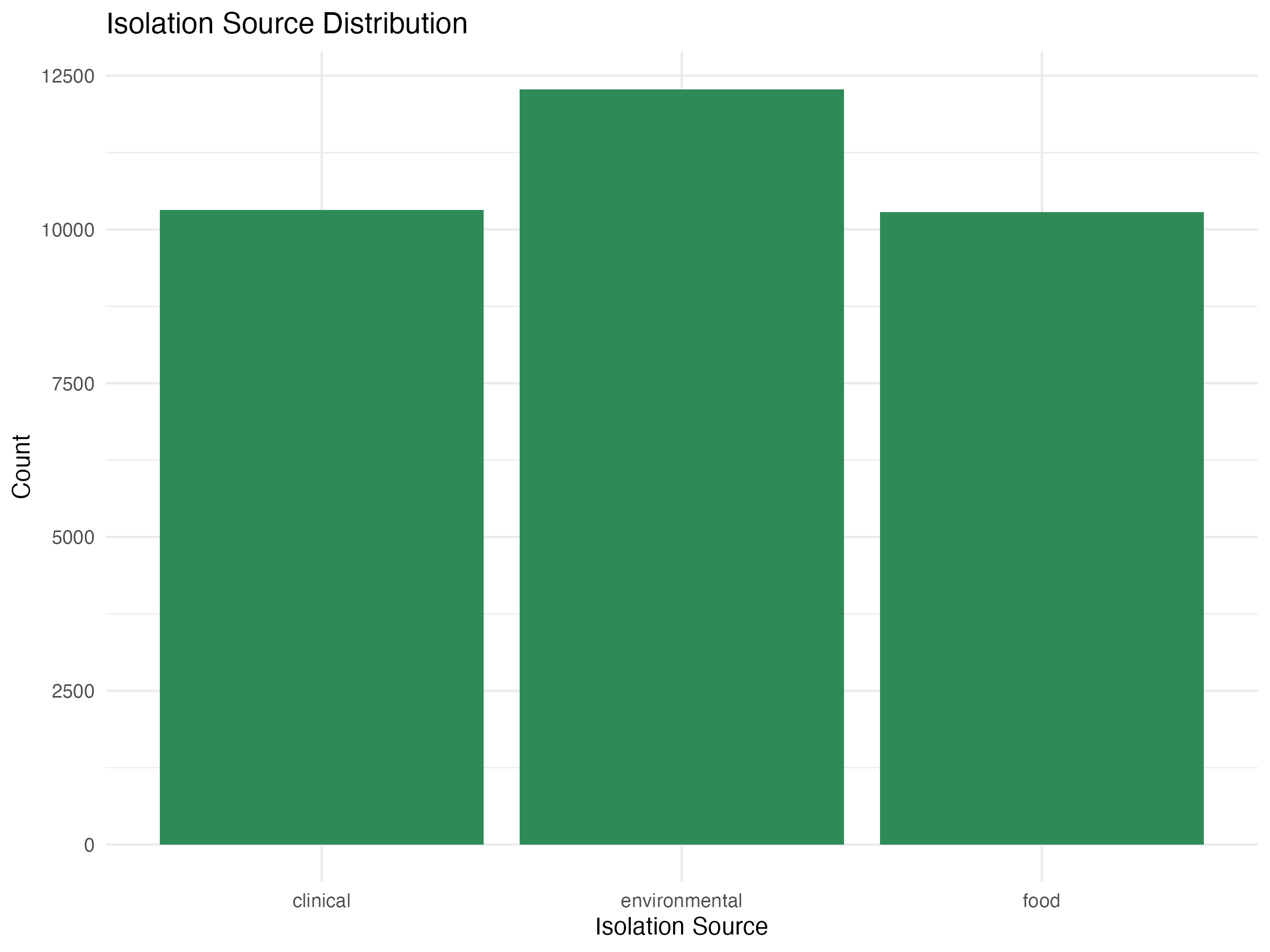
### 4.0.1 Exploratory Data Analysis

Exploratory data analysis started with visualizing the total counts of food, clinical and environmental isolates in the 10 year time period. It is seen in *figure* 1 that food isolates account for the smallest proportion of isolated *Listeria monocytogenes* at 10,284. Clinical isolates fall slightly ahead at 10,319, with a difference of 35 isolates in the 10 year period. Environmental samples dominate the isolate types at 12,777. To get a visualization of where food isolates were coming from a bar graph was made showing the location of origin of each food isolate. Some isolates only logged “USA” as their origin so those were removed from visualization. The majority of food sampled were originate in New York state, followed by Washington, Arizona, Florida, and Texas. The 3rd bar graph *figure 3* shows the origin of clinical cases. This means the origin of the source of infection, likely a food product. An overwhelming percentage of originated in Sweden, followed by Texas and New York. A stacked bar graph was made showing the counts for each isolation source over the 10 year period. A large spike in all sources is seen in early 2020 followed by a large dip.

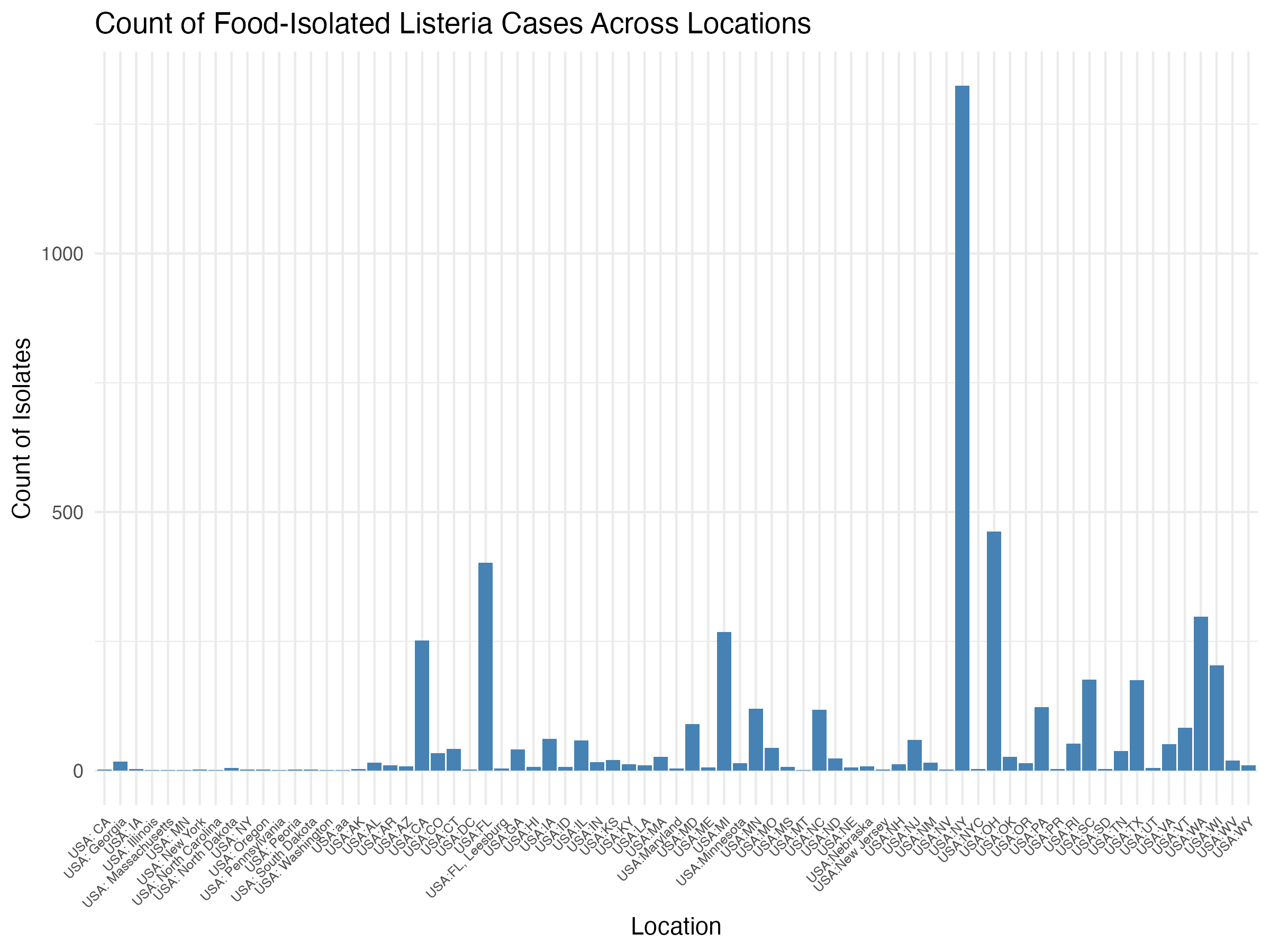
knitr::include\_graphics("../../results/tables/yearly\_isolate\_summary.png")



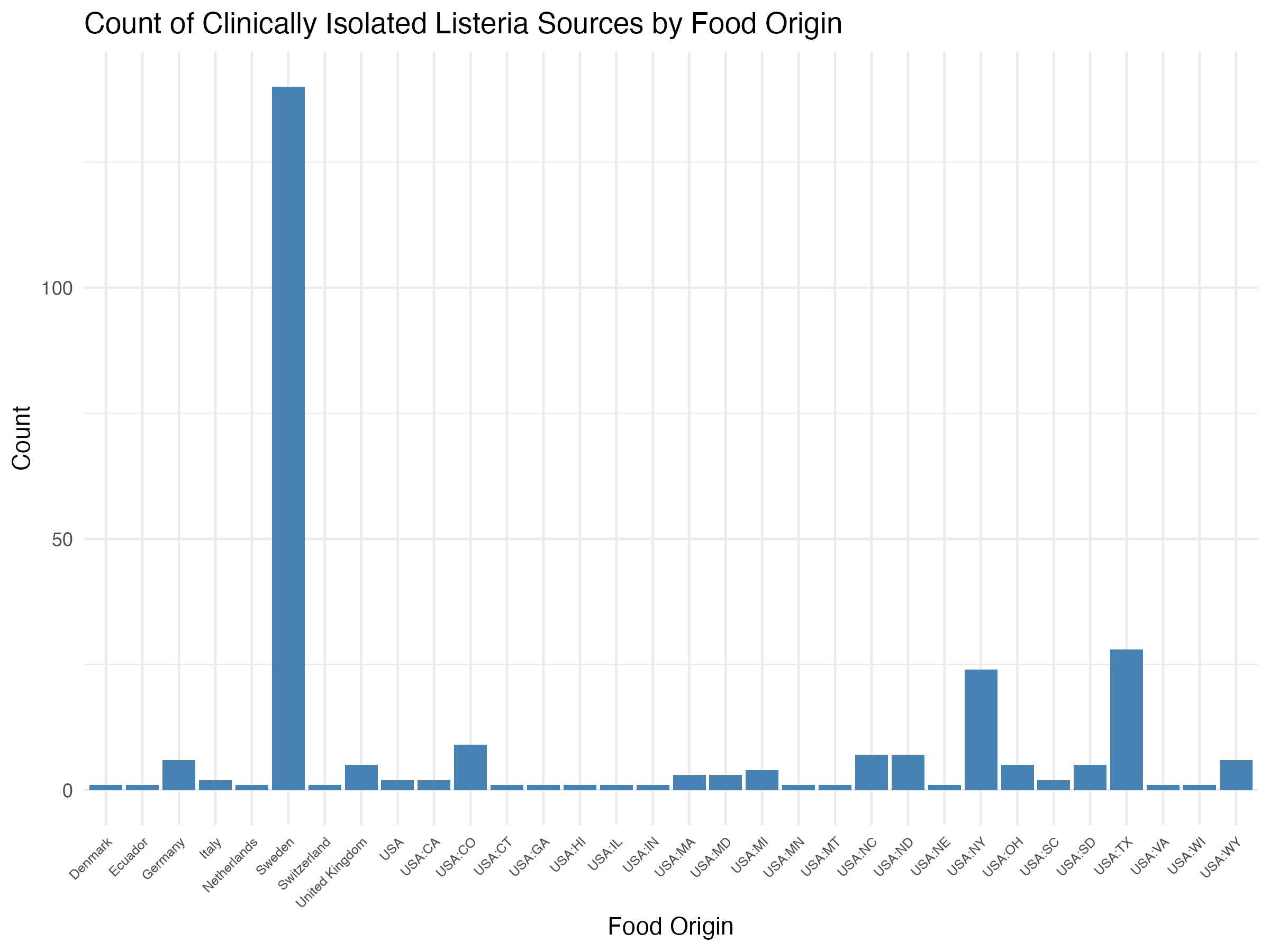
knitr::include\_graphics(here("results/figures", "isolatetypes.png"))



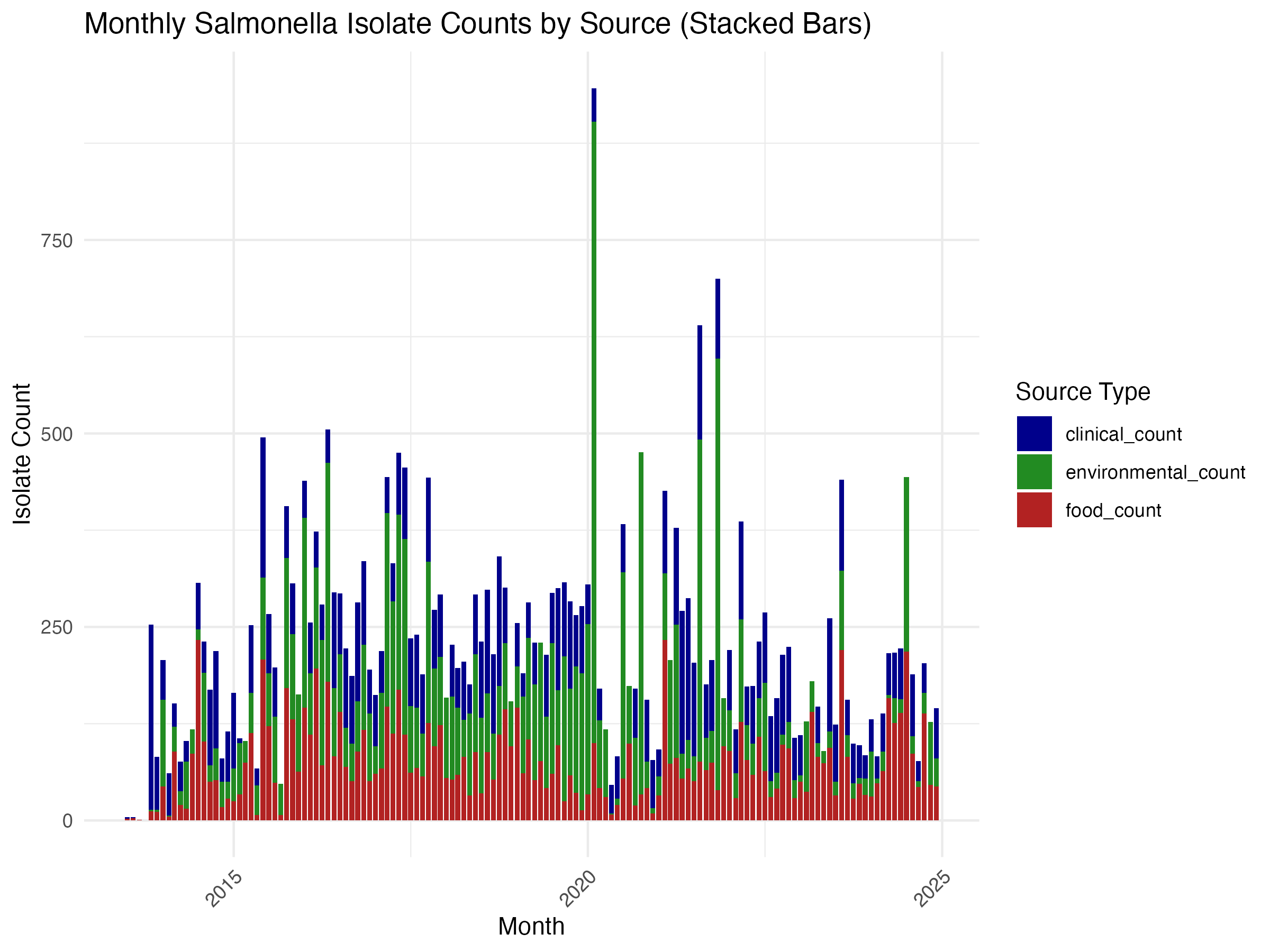
knitr::include\_graphics(here("results/figures","locations.png"))



knitr:: include\_graphics(here("results/figures", "sourcestate.png"))



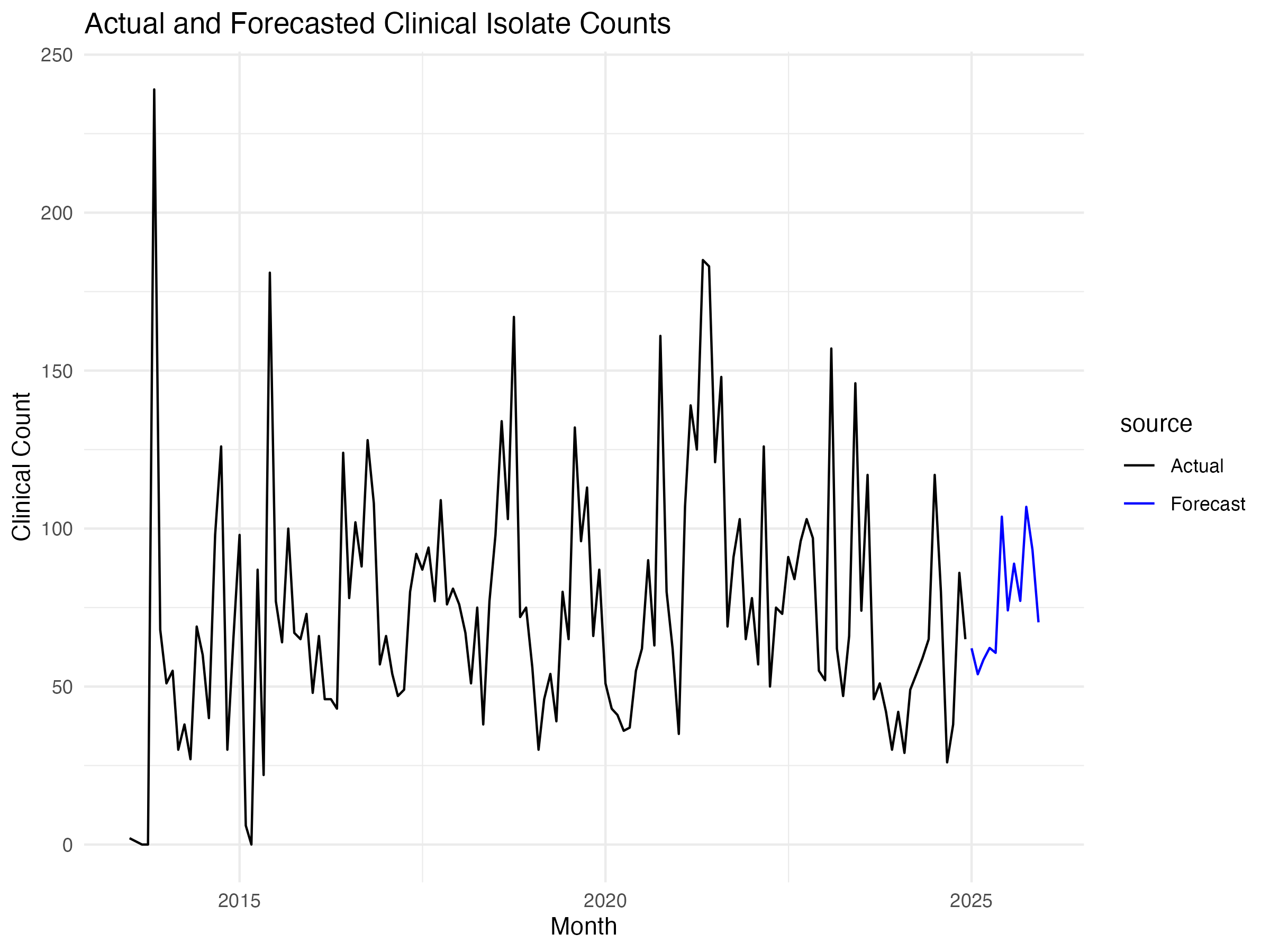
knitr::include\_graphics(here("results/figures", "isolatecounts.png"))



## 4.1 Poisson Modeling

*Time of year or “season” as a predictor for clinical Listeria monocytogenes isolates*

knitr::include\_graphics(here("results/figures", "seasonalpreds.png"))

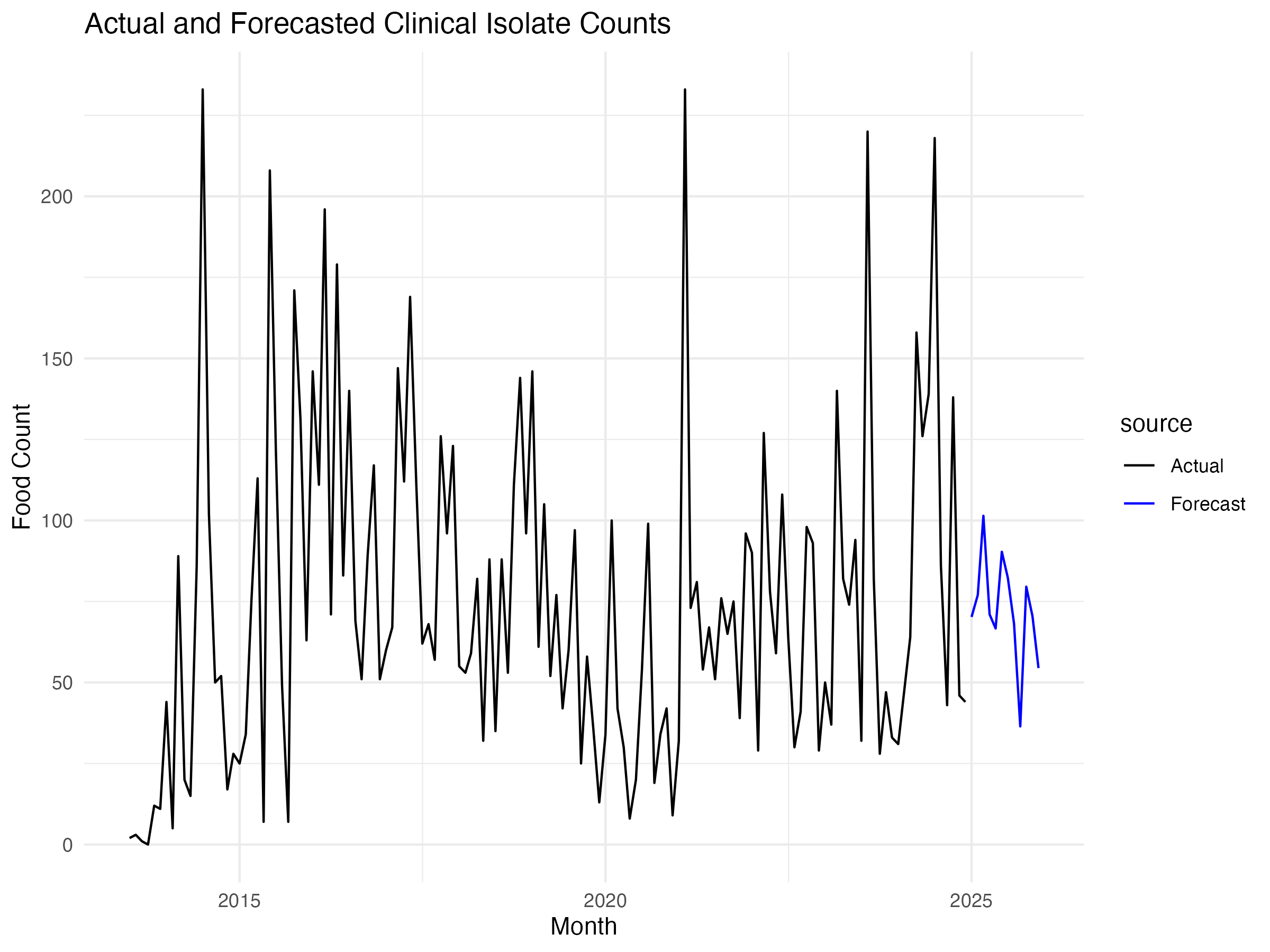


metric\_table <- readRDS(here::here("results", "tables", "metric\_table.rds"))  
knitr::kable(metric\_table, caption = "Model Performance Metrics for Poisson Regression")

Model Performance Metrics for Poisson Regression

| .metric | .estimator | .estimate |
| --- | --- | --- |
| rmse | standard | 37.6469538 |
| rsq | standard | 0.1744102 |
| mae | standard | 26.8444444 |

knitr::include\_graphics(here("results/figures", "pmodel2.png"))

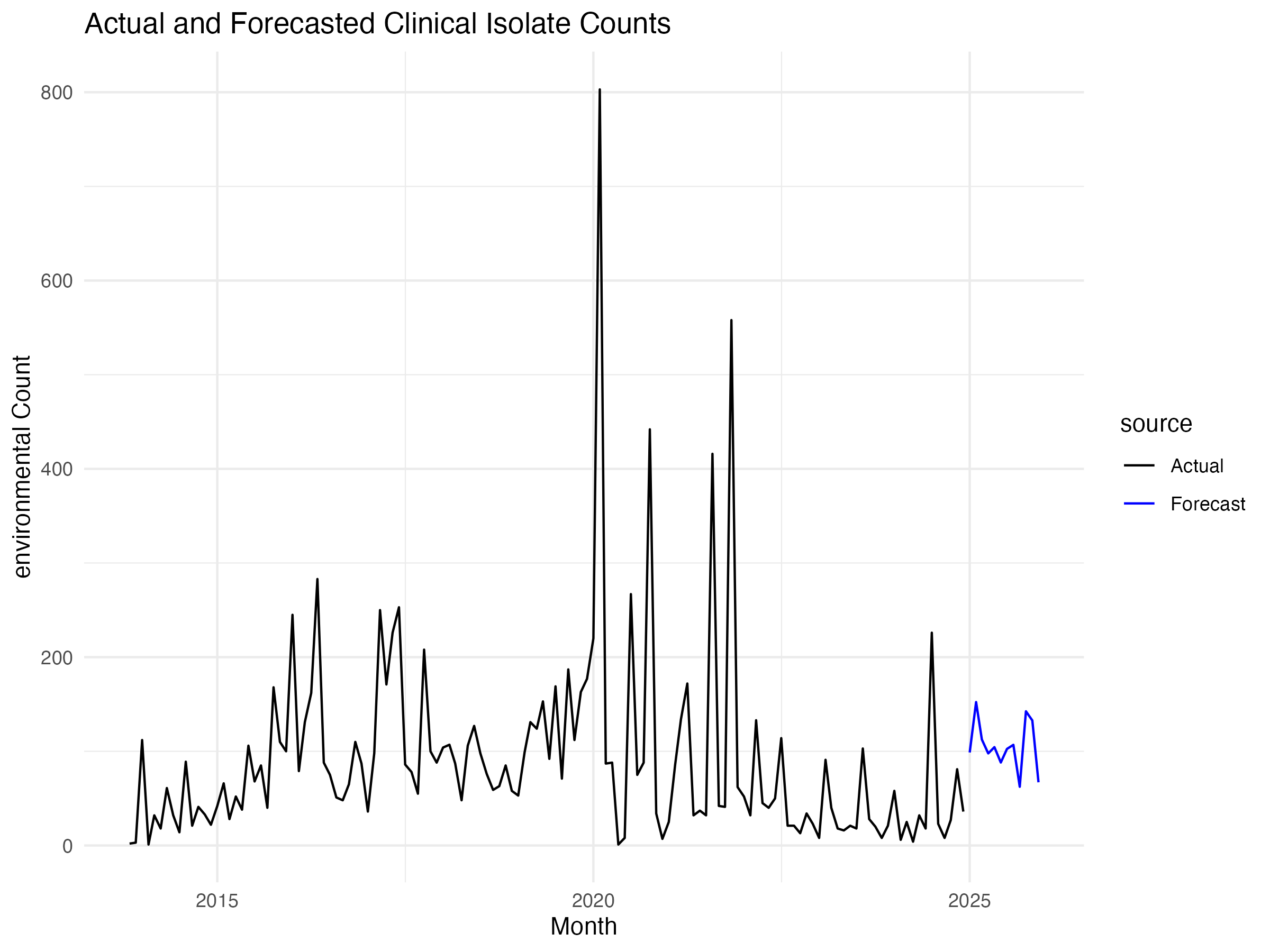


metric\_table2 <- readRDS(here::here("results", "tables", "metric\_table2.rds"))  
knitr::kable(metric\_table2, caption = "Model Performance Metrics for Poisson Regression")

Model Performance Metrics for Poisson Regression

| .metric | .estimator | .estimate |
| --- | --- | --- |
| rmse | standard | 47.4896799 |
| rsq | standard | 0.0989078 |
| mae | standard | 37.3646465 |

knitr::include\_graphics(here("results/figures", "pmodel3.png"))



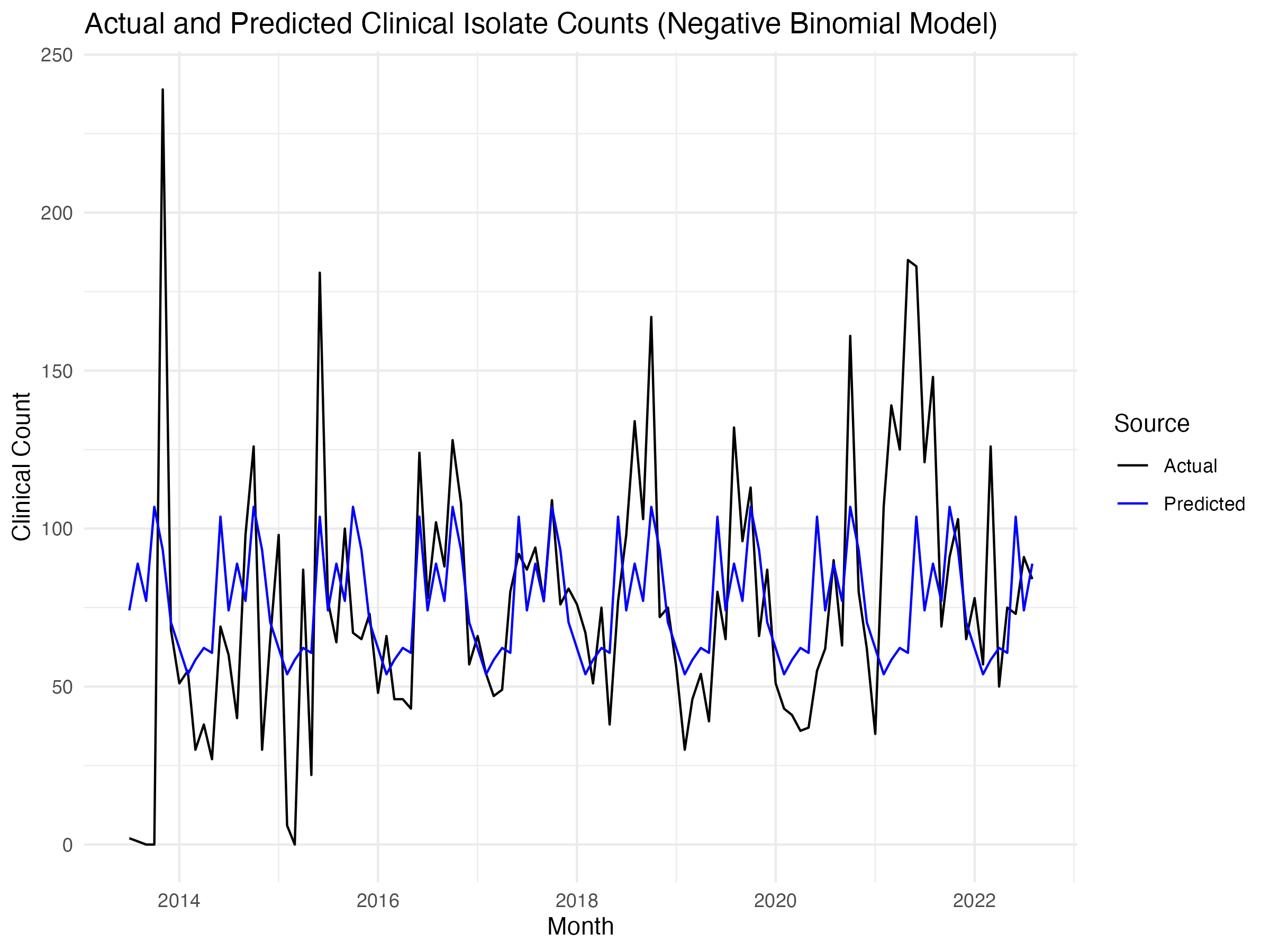
metric\_table3 <- readRDS(here::here("results", "tables", "metric\_table3.rds"))  
knitr::kable(metric\_table3, caption = "Model Performance Metrics for Poisson Regression")

Model Performance Metrics for Poisson Regression

| .metric | .estimator | .estimate |
| --- | --- | --- |
| rmse | standard | 109.362142 |
| rsq | standard | 0.049112 |
| mae | standard | 69.037172 |

## 4.2 Negative Binomial Model

knitr::include\_graphics(here("results/figures", "nmodel\_clinical.png"))

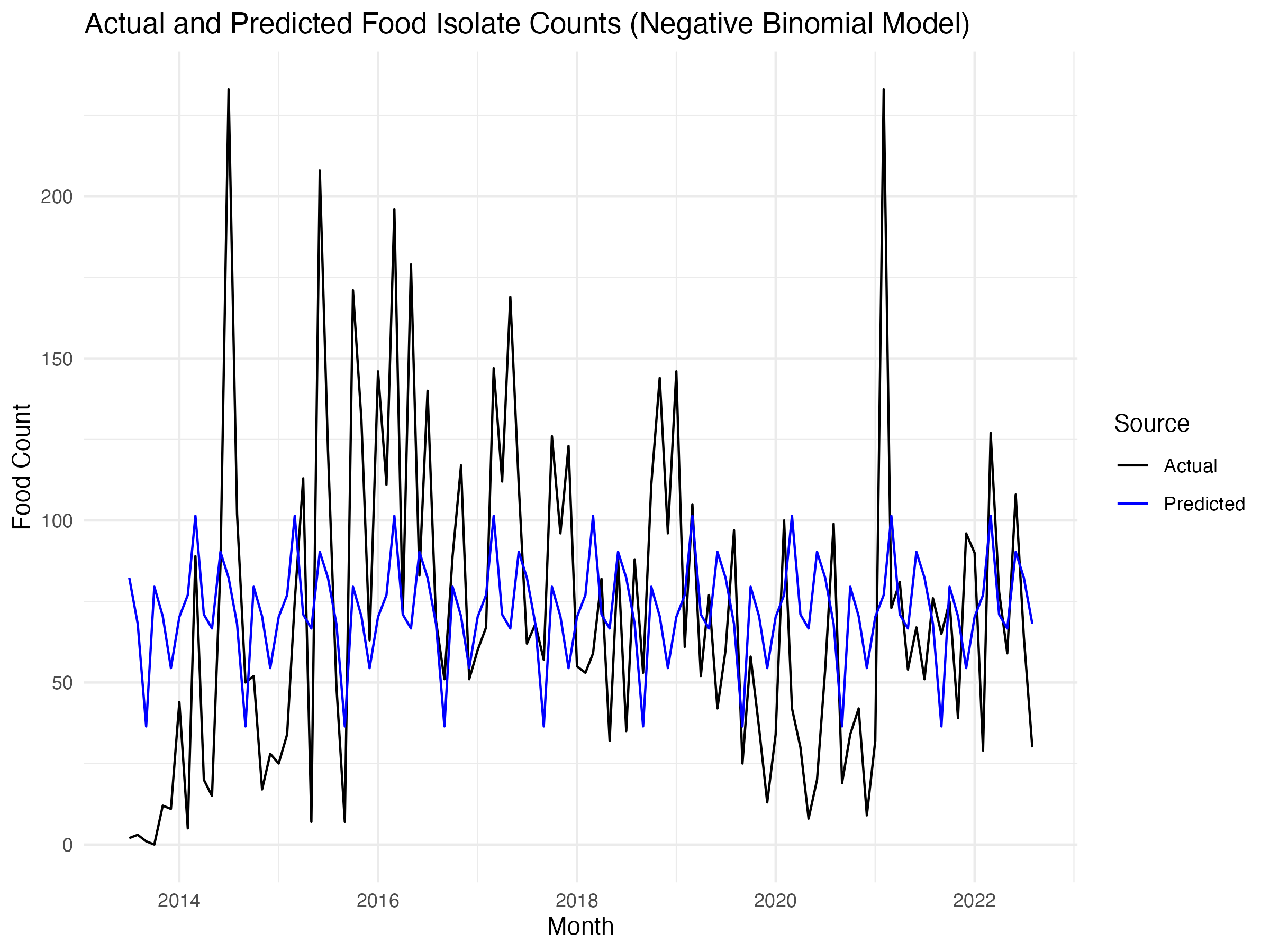


metric\_table4 <- readRDS(here::here("results", "tables", "metric\_table4.rds"))  
knitr::kable(metric\_table4, caption = "Model Performance Metrics for Poisson Regression")

Model Performance Metrics for Poisson Regression

| .metric | .estimator | .estimate |
| --- | --- | --- |
| rmse | standard | 37.6469538 |
| rsq | standard | 0.1744102 |
| mae | standard | 26.8444445 |

knitr::include\_graphics(here("results/figures", "nmodel\_food.png"))

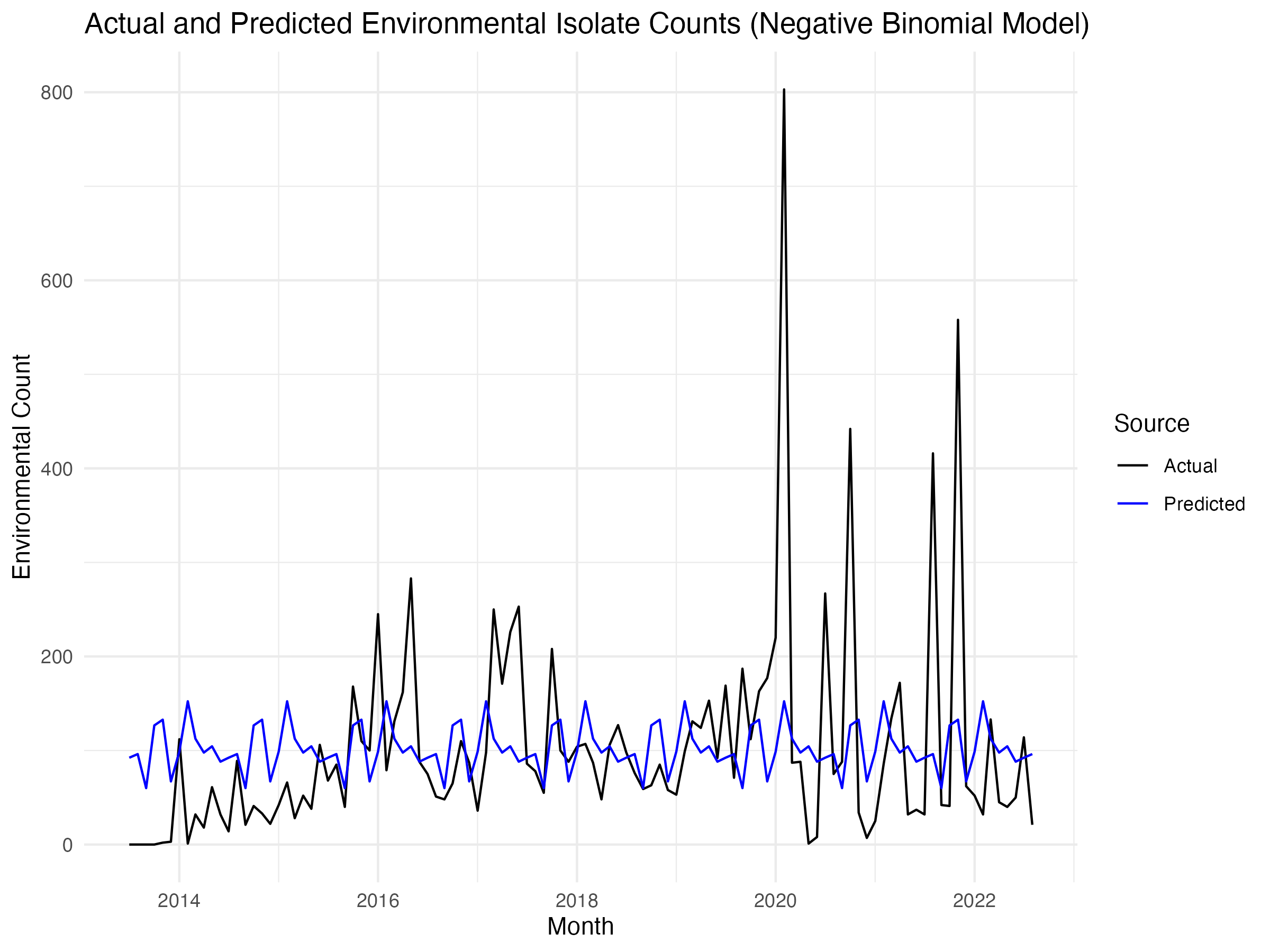


metrics\_table5 <- readRDS(here::here("results", "tables", "metrics\_table5.rds"))  
knitr::kable(metrics\_table5, caption = "Model Performance Metrics for Poisson Regression")

Model Performance Metrics for Poisson Regression

| .metric | .estimator | .estimate |
| --- | --- | --- |
| rmse | standard | 47.4896799 |
| rsq | standard | 0.0989078 |
| mae | standard | 37.3646465 |

knitr::include\_graphics(here("results/figures", "nmodel\_env.png"))



metrics\_table6 <- readRDS(here::here("results", "tables", "metrics\_table6.rds"))  
knitr::kable(metrics\_table6, caption = "Model Performance Metrics for Poisson Regression")

Model Performance Metrics for Poisson Regression

| .metric | .estimator | .estimate |
| --- | --- | --- |
| rmse | standard | 109.362142 |
| rsq | standard | 0.049112 |
| mae | standard | 69.037172 |

# 5. Discussion

## 5.1 Summary and Interpretation

This study examined temporal and source-based trends in Listeria monocytogenes isolates reported in the United States from 2010 to 2024 using publicly available data from the NCBI Pathogen Detection database. Isolates were classified into clinical, environmental, and food-related categories to assess patterns and predictors of clinical infections.

Exploratory analysis showed that environmental samples were the most frequently reported isolate type, which supports previous findings that L. monocytogenes is widespread in processing environments and natural reservoirs. Clinical and food isolates were nearly equal in count, suggesting a potential link between food exposure and clinical outcomes.

Temporal modeling using Poisson regression revealed a seasonal pattern in clinical isolates, with elevated counts in warmer months,especially summer. This pattern may reflect greater bacterial growth in food products stored at warmer temperatures, seasonal shifts in food handling practices, or increased consumption of ready-to-eat items during these months. Though it must be noted that the predictive power was low. The R² value was calculated to be 0.17, meaning that there is an effect from season/time of year but that alone is not sufficient in predicting clinical cases. Food and environmental models had lower R² values. All models all had very high dispersion statistics at 20,34, and 112 for clinical, food, and environmental respectively. Negative binomial models produced lower dispersion statistics at 0.923,0.989, 0.94, indicating the model was better fit to the data. Unfortunately, predictive performance remained the same, reinforcing the conclusion that seasonality alone is not sufficient for predicting clinical, food, and environmental isolates

These results loosely align with prior epidemiological data suggesting that Listeria infections may follow seasonal trends, which has important implications for food safety monitoring and public health messaging. Summer could be a critical time for intensified surveillance and intervention efforts.

## 5.2 Strengths and Limitations

While the data set contains thousands of isolates, the amount of clinical isolates is quite small due to the rarity of infection. This made statistical analysis complicated as some months during the 10 year period experienced 0 cases of listeriosis. There also were not many predictors to work with from the raw data. Isolation source would have been an association to explore but, the isolation sources are entered manually bu researchers and inspectors so there is no standardization in nomenclature of sources. For example, one source can say “hot-dog” while another says “premium hot-dog”. More spatial analysis could have been done if more entries specify the state of origin. Many had just “USA” or NA. This is also not standardized for NCBI and can be entered however the submitter pleases. It may have also been helpful to expand the timeline from 10 years to 15 or 20 years to allow for more clinical isolate data.

## 5.3 Conclusions

This study identified seasonal patterns in clinical Listeria monocytogenes isolates, with potential implications for timing surveillance and prevention strategies. Accurate predictions may allow state and federal governments to take necessary precautions ahead of predicted spikes. While environmental and food isolates remain a significant reservoir for transmission, more refined metadata and consistent data entry practices are essential for improving future predictive models.

Future research should focus on expanding spatial analysis, refining source categorization through natural language processing, and incorporating additional predictors such as food recall data or consumer behavior trends. With enhanced data quality and analytical tools, real-time surveillance systems can be better leveraged to prevent and control outbreaks of listeriosis.

# 6. References

[1] Centers for Disease Control and Prevention. (n.d.). *About listeria infection*. Centers for Disease Control and Prevention. https://www.cdc.gov/listeria/about/index.html

[2] Centers for Disease Control and Prevention. (n.d.-b). *Listeria outbreaks*. Centers for Disease Control and Prevention. https://www.cdc.gov/listeria/outbreaks/index.html

[3] Farber JM, Peterkin PI. Listeria monocytogenes, a food-borne pathogen. Microbiol Rev. 1991 Sep;55(3):476-511. doi: 10.1128/mr.55.3.476-511.1991. Erratum in: Microbiol Rev 1991 Dec;55(4):752. PMID: 1943998; PMCID: PMC372831.

[4] *Food Safety and Inspection Service*. Listeria Monocytogenes | Food Safety and Inspection Service. (n.d.). https://www.fsis.usda.gov/inspection/compliance-guidance/microbial-risk/listeria-monocytogenes

[5] Smith, P. (2025, February 24). *Nutritional shakes recalled after listeria infections killed 12 and hospitalized dozens dating back to 2018*. NBCNews.com. https://www.nbcnews.com/news/us-news/nutritional-shakes-recalled-listeria-infections-killed-11-hospitalized-rcna193411

[6] Vázquez-Boland JA, Krypotou E, Scortti M.2017.Listeria Placental Infection. mBio8:10.1128/mbio.00949-17.<https://doi.org/10.1128/mbio.00949-17>