

Analyzing Foodborne Disease Outbreaks Over Time

A look into illnesses and more from 1998 to 2015

Group 12 Introductions

Trina Shores

Steven Macapagal

Journey Martinez

Yuan Yao

Heather Nagy

Kenneth Porter

Trina Shores

- Major - Graduate Certificate Statistics & Analytics
- Role - group leader
- Responsibility - provide guidance on appropriate research themes and add context to analysis

Steven Macapagal

- Major - Statistics
- Role - editor, facilitator, analysis
- Responsibility - provide graphical analysis, edit slide deck

Journey Martinez

- Major - Statistics
- Role - analysis and computation
- Responsibility - provide statistical analysis, tabulation, and contribute to report

Yuan Yao

- Major - Biology
- Role - statistical analysis; prior literature
- Responsibility - review previous research

Heather Nagy

- Major - Statistics
- Role - Analysis, individual contributor
- Responsibility - Contribute data set insights and preliminary analysis to report

Kenneth Porter

- Major -Statistics
- Role -editor
- Responsibility -Combining team's work into consolidated reports

Data on foodborne illnesses

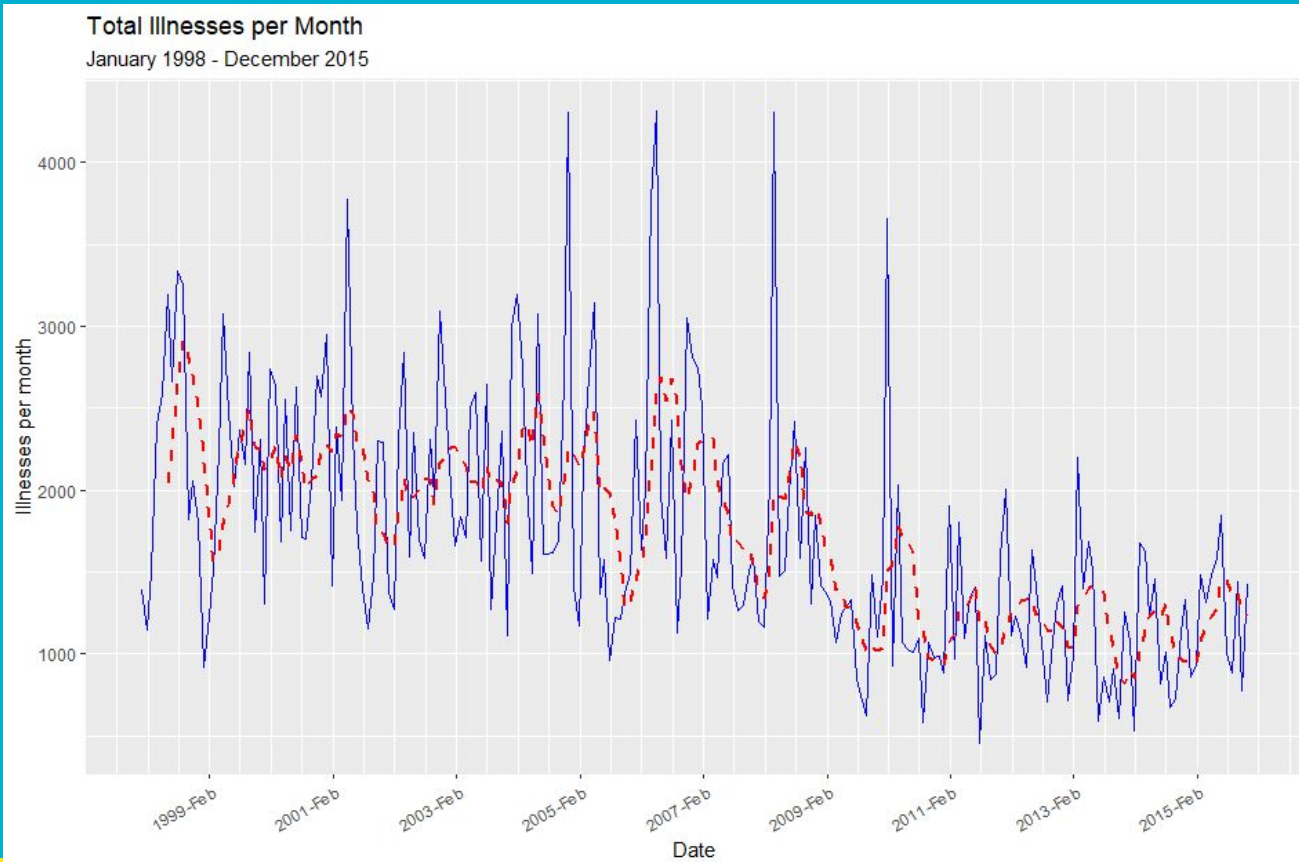
The data set

- Primary measurements are illnesses, hospitalizations, and fatalities for each outbreak
- Other variables include date, US state, location, food or ingredient, type of bacteria or virus
- Data are reported monthly from January 1998 to December 2015
- Over 19,000 different outbreaks recorded
- From a Kaggle data set provided by the Centers for Disease Control and Prevention (<https://www.kaggle.com/datasets/cdc/foodborne-diseases>)

Our goals

- How is the rate of foodborne illness changing over time?
 - Directional trends
 - Changes in variability
 - Cyclical trends
- Is there a seasonal component to this trend?
- Are there relationships to hospitalizations, fatalities, or external factors?

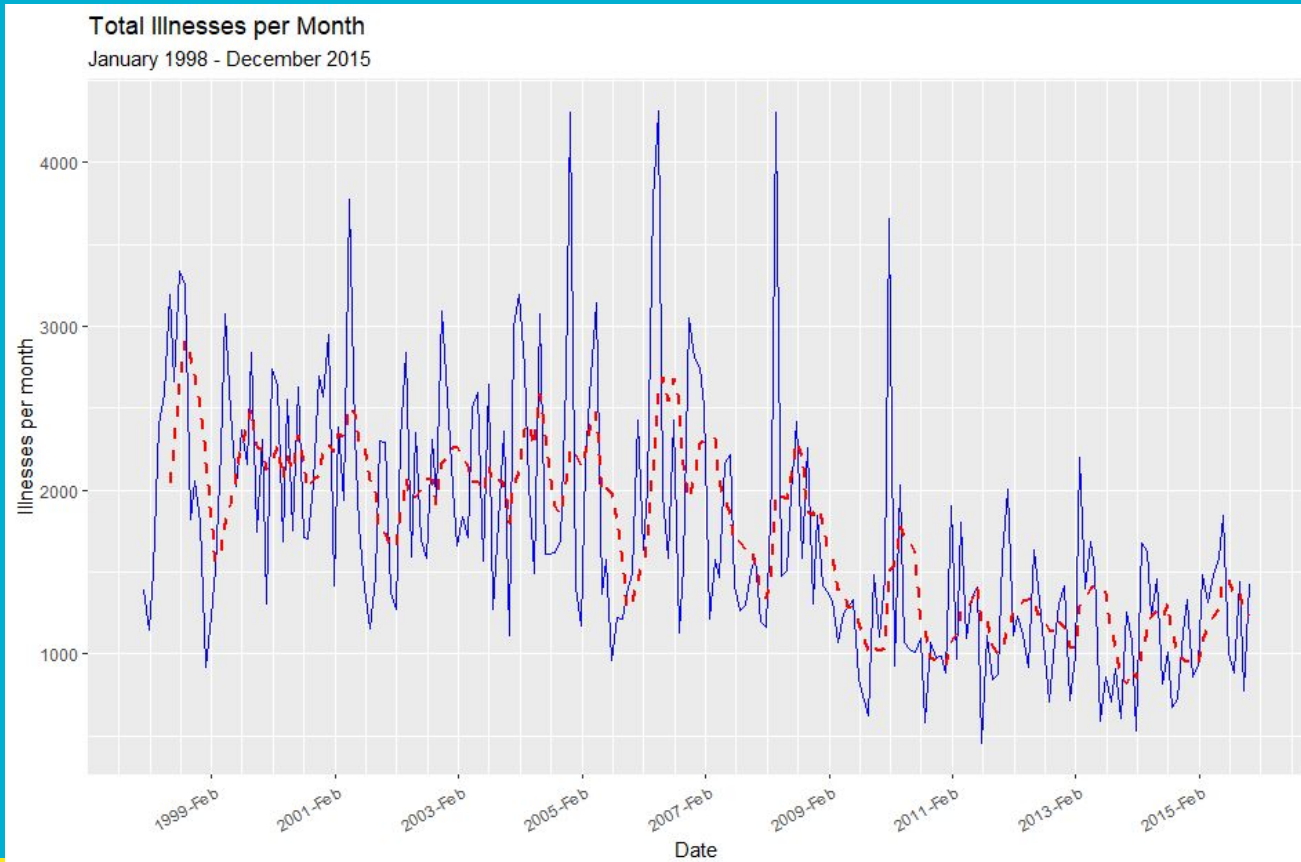
Total illnesses per month appear to decline over time and are less variable over time.



Solid blue line represents the original time series

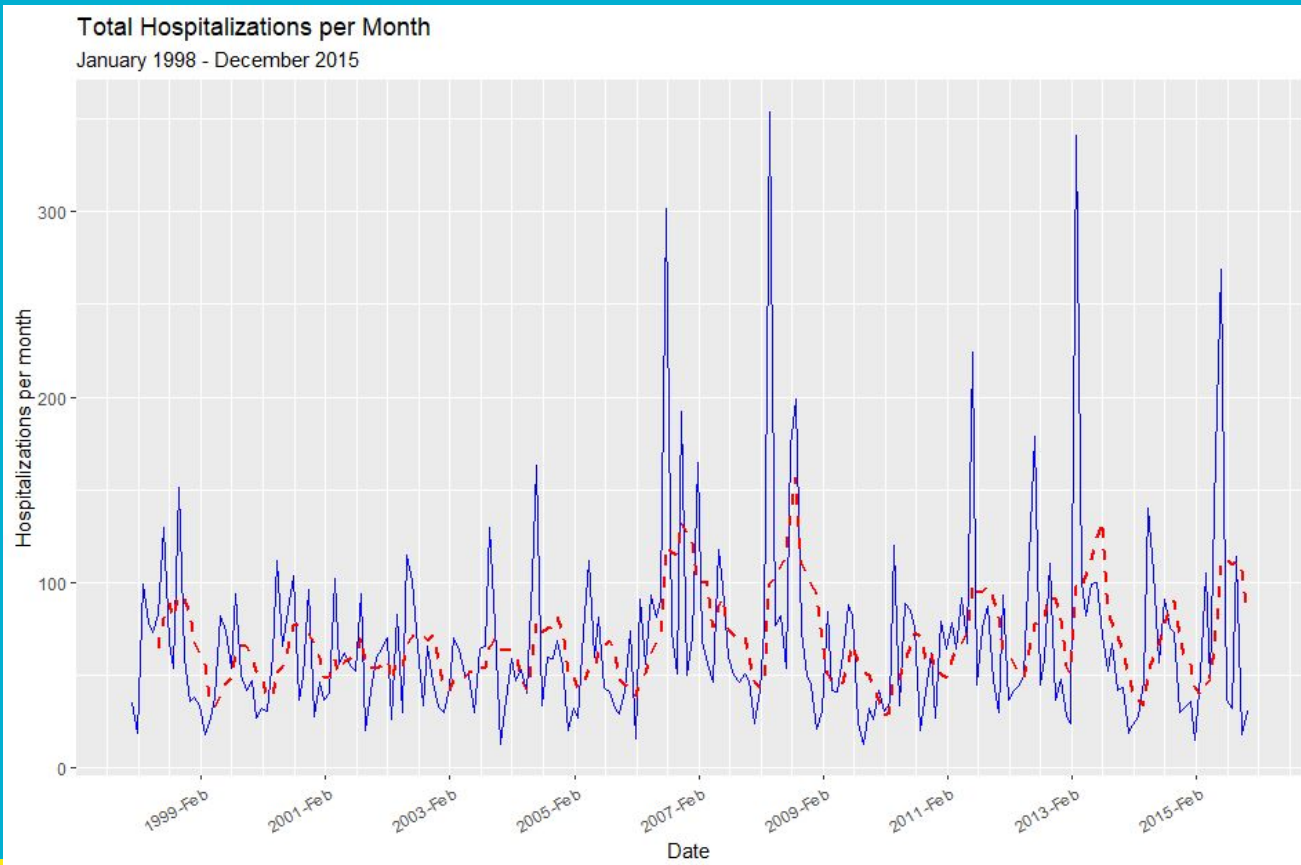
Dotted red line represents a filtered time series over a 6-month period

Total illnesses per month appear to decline over time and are less variable over time.



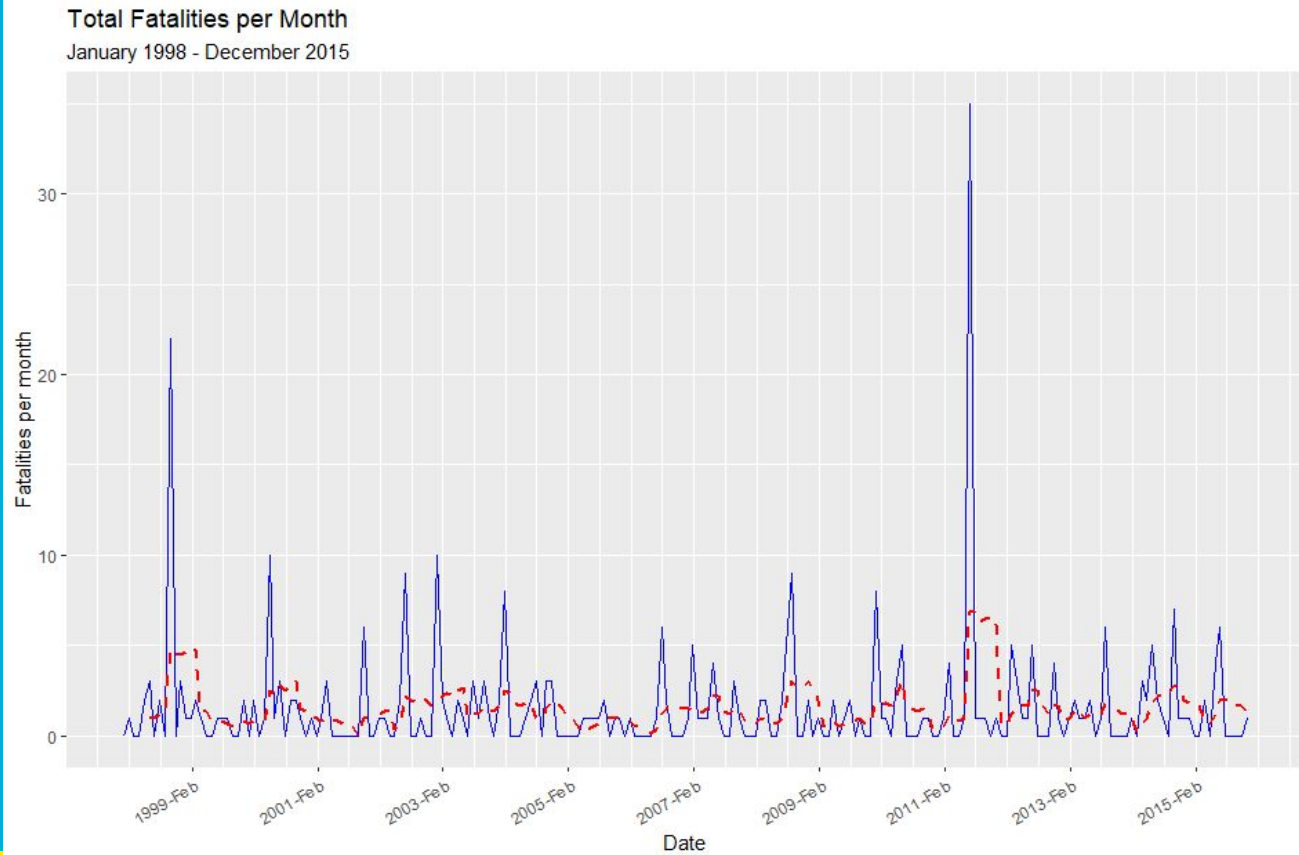
Variability is much greater from 1998 to 2008 and decreases from 2009 to 2015. Illnesses also seem to be cyclical.

Total hospitalizations per month seem to remain constant, but their variability is increasing over time.



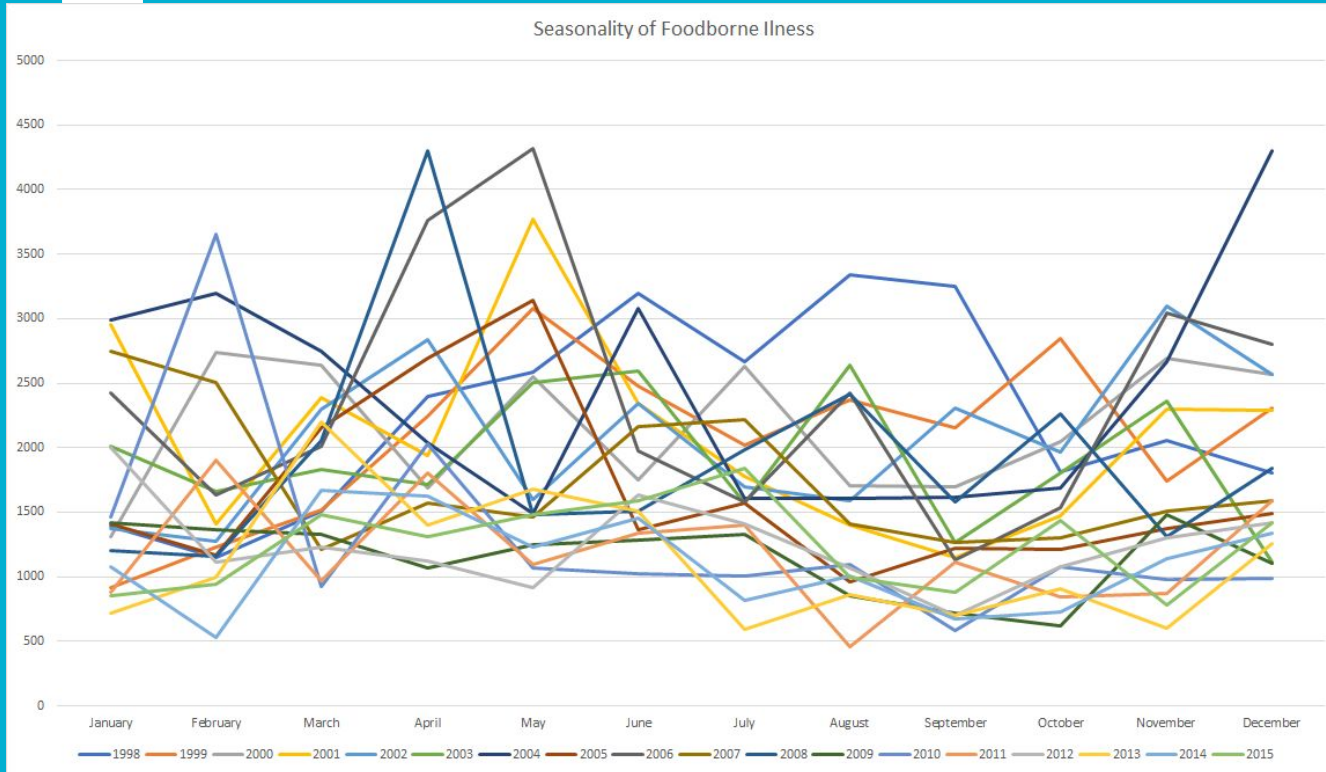
The variability starts increasing from 2006 to 2015. Hospitalizations also show some cyclic nature.

Total fatalities per month remain constant over time with occasional peaks in early 1999 and 2011.



The filtered time series does not seem to show any clear trend over time.

Seasonality of Total Illness per year



There appears to be no visually clear seasonality pattern but the highest counts appear to be between February and May while the lowest are between July and November

Future areas of research

- Looking at autocorrelation over illnesses month-to-month
- Calculating cross-correlations between time series to see if there is a relationship between illnesses and hospitalizations, illnesses and fatalities, etc.
- Build an appropriate model of illnesses over time

Questions, comments, suggestions?

**Thank you for your
engagement and feedback!**