THE UNIVERSITY OF CHICAGO
DATA SCIENCE INSTITUTE

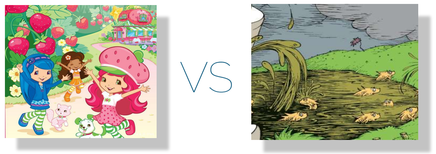
Sociome Data Commons

Investigating Social Determinants of Health

Authored by Elinai Alicia Severiano Perez
Mentored by Sandra Tilmon, MPH MS

MACALESTER

Motivation



Instead of looking only at individual risk behaviors when investigating disease risk, the root causes also need to be evaluated. These are commonly referred to as the *Social Determinants of Health*. Sociome aims to collect traditional SDoH but also expand into the full social context of disease, including social, environmental, behavioral, housing, and economic factors. Considering sociome factors is important as they:

1. Can disrupt social, environmental, or psychological securities needed for best health and/or
2. Can exacerbate or be the main cause of poor health

What is it?

An open source platform with public sociome data to upload, store, and access sociome datasets for researchers' use, especially with clinical data

Project Introduction

Background

Changes in weather conditions can have both rapid and slow effects on human health.

- ➔ High levels of humidity can trigger the narrowing of airways for people with asthma.
- ➔ Wet-bulb temperature resembles how a human body cools down (having a wet cloth on a thermometer resembles sweat cooling the body)
  - ◆ High levels can lead to heat exhaustion and heatstroke

Research Question

1. Can we identify weather anomalies?
   a. Add to the sociome commons for others to use

References

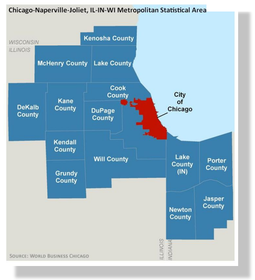
Schinasi LH, Kenyon CC, Hubbard RA, Zhao Y ,et al. Associations between high ambient temperatures and asthma exacerbation among children in Philadelphia, PA: a time series analysis. Occup Environ Med. 2022 May;79(5):326-332. doi: 10.1136/oemed-2021-107823. Epub 2022 Mar 4. PMID: 35246484.

Analyses

Source

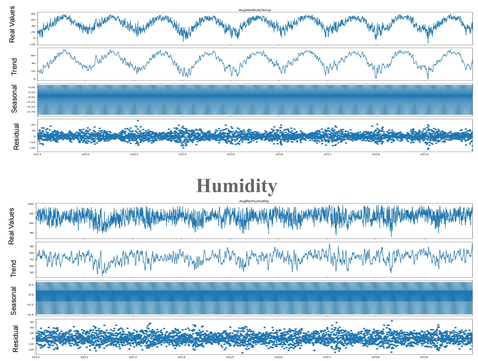
wrangling

Location



- ➔ 2011 to 2019 Chicago area data rather than just Cook County (or Chicago)
  - ◆ Patients at University of Chicago's hospital do not just come from Chicago, some come from Indiana or other nearby counties
  - ◆ Did not consider Wisconsin counties. Removed counties with too much missing data
- ➔ Averaged out all the data rather than looking at each county individually
  - ◆ Weather is relatively the same throughout Chicagoland

Seasonal Decomposition



I used additive decomposition on time series data to break down its components: Trend, seasonality, and noise (residuals). By breaking it down to its elements, it is easier to analyze underlying patterns.

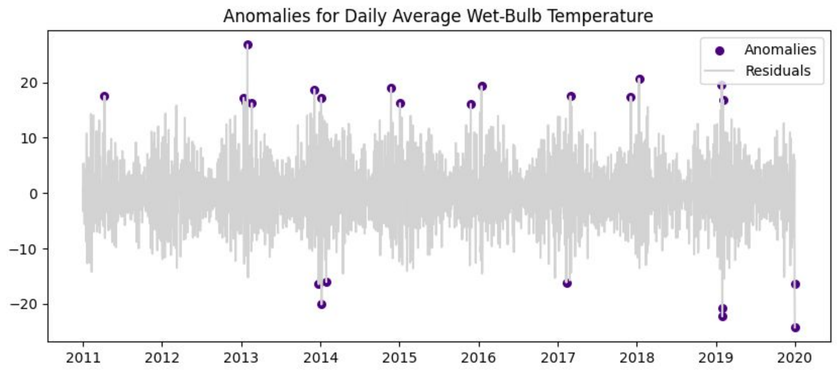
Additive decomposition:

$$y(t) = \text{Level} + \text{Trend} + \text{Seasonality} + \text{Noise}$$

For wet-bulb Temperature and humidity (left), the first block shows the actual values while the second block shows the trend without seasonality. This makes it easier to identify anomalies, as they manifest as outliers in the residuals (seen in the fourth block), standing out as they aren't explained by trend or seasonality.

Anomaly Detection

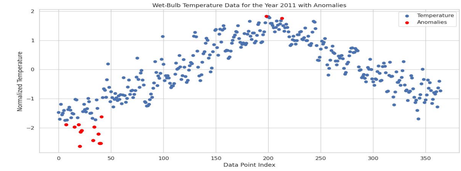
- ❖ An anomaly is defined as three standard deviations away from the mean
- ❖ Used the residuals from the seasonal decomposition
- ❖ Helps identify extreme humidity levels



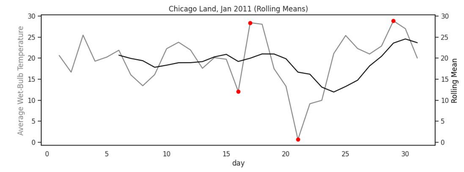
Other methods investigated

Tried to determine anomalies with the following methods:

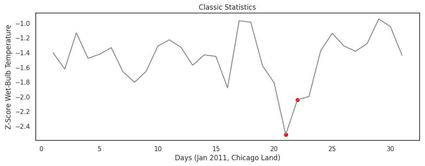
K-nearest Neighbors



Rolling Means



Classic Statistics (Z-Scores)



Didn't end up using these

- ➔ Identified anomalies by month normals rather than by season normals
- ➔ Over-detected temperatures as anomalies

Conclusions

- ❖ Learned different methods of anomaly detection through trial and error

Next Steps

- ➔ Start testing out sociome commons data with clinical data
- ➔ Play with different levels of the z-score anomaly definition
- ➔ Weather anomaly data will be added SDC and included in a prediction model for asthma attacks the day of, the day before, the week before will be tested