

Title: Geospatial and Epidemiological Analysis of Influenza Transmission and Social Vulnerability in the United States (2018-2019)

1. Introduction

The 2018-2019 influenza season in the United States presented significant spatial and social variability in transmission patterns. This project seeks to integrate flu surveillance data, household demographic data, and climate information using geospatial tools and statistical modeling to understand the potential interplay between multigenerational household structures and flu positivity rates at the state level.

2. Data Collection

Data sources included:

- State Boundary Shapefile: U.S. state-level boundaries were sourced from the U.S. Census Bureau's Cartographic Boundary Files and projected to WGS 84.
- ILINet CSV: Weekly Influenza-like Illness (ILI) surveillance data from CDC's FLUVIEW portal.
- ICL NREVSS Clinical Labs CSV: State-level lab-confirmed flu results.
- Multigenerational Household Data: ACS 2018 dataset (Table B11017) accessed through data.census.gov.

3. Data Cleaning and Feature Engineering

- ILINet CSV: Removed non-informative columns (zero variance, missing % weighted ILI, and non-numeric age columns). Retained state name, week, and counts.
- ICL NREVSS: Removed 531 rows with 'X' values across all major fields.
- Created a new column Season_Week using the formula: IF(YEAR=2018, WEEK-39, WEEK+13) for consistent flu season indexing.
- Final row counts: ILINet (2809 x 7), Clinical Labs (2278 x 9).

4. Data Integration in ArcGIS Pro 3

- Uploaded cleaned CSVs and shapefiles to ArcGIS Pro 3.
- Conducted joins using Geodatabases to handle one-to-many relationships (e.g., multiple weeks per state).
- Aggregated the following:
 - ILINet: Sum of ILI cases and total patients; Mean of % Unweighted ILI.
 - Clinical Labs: Sum of specimen counts, Type A & B; Mean of positivity rates.

5. Visualization and Mapping

- Flu Activity Maps: Choropleth map using graduated colors for % ILI and lab-confirmed positivity.
- Multigenerational Overlay: Proportional green circles to represent % multigenerational households.
- Hotspot Analysis: Getis-Ord Gi* to identify spatial clusters of high/low flu positivity.
 - Hot spots: Georgia, Mississippi, Alabama (95% confidence).
 - Cold spots: North Dakota, Montana (90-95% confidence).

6. Interpretation and Insights

- High correlation between SUM_ILITOTAL and SUM_TOTAL_PATIENTS (volume indicators).
- Weak/no correlation between % Multigenerational HH and Mean % Positive.
- Some spatial overlap between MGH prevalence and flu activity in Southeast states.

7. Machine Learning Modeling

- Tried Random Forest Regression to predict MEAN_PERCENT_POSITIVE.
- Performance:
 - R^2: -0.46
 - RMSE: 2.46
- Result: Poor model performance; weak correlation among features; ML not pursued further.

8. Conclusion

This study reveals localized clusters of flu activity and highlights potential, though inconclusive, social vulnerability via multigenerational households. While volume-based flu indicators are interrelated, they alone do not predict lab-confirmed positivity effectively. Hotspot analysis provides valuable geographic insights, making this methodology promising for future surveillance and planning.

9. Future Work

- Incorporate additional demographic or environmental variables (e.g., healthcare access, vaccination rates).
- Perform time-series analysis of flu progression.
- Explore local clustering within metropolitan areas.