

# PUBLIC HEALTH FORECASTING WITH R: FLU TREND MONITORING FOR U.S. STATEWIDE VACCINE PLANNING

## 1. Background

A U.S. state health department aimed to improve the timing of vaccine stock distribution across counties during seasonal flu outbreaks. Traditional reporting methods were reactive, relying on weekly case submissions that delayed logistical responses by 1–2 weeks.

We were engaged to create a robust statistical forecasting system using R. The objective was to model past flu trends, predict upcoming case surges, and generate early warning outputs to optimize supply-chain decisions and hospital preparedness.

## 2. Objective

- To model county-level flu cases using historical time series data
- To forecast short-term trends and enable preemptive vaccine distribution
- To automate reporting for public health officers via R Markdown and visual dashboards

## 3. Data Used

**Source:** State infectious disease surveillance system (2016–2023)

### **Dataset Details:**

- Weekly reported influenza-like illness (ILI) cases across 24 counties
- Fields: Week, County, ILI\_Cases, Population, Vaccination\_Coverage, Temperature\_Avg

**Data Shape:** ~8,700 records (7 years × 52 weeks × 24 counties)

## 4. Methodology

### **4.1 Data Preprocessing**

- Used dplyr and lubridate for date handling
- Created Cases\_per\_100k as normalized response variable
- Checked for missing values and imputed with time-based rolling average

## 4.2 Time Series Modeling

- Decomposed series using `stl()` for seasonal and trend components
- Modeled using:
  - **ARIMA** for single-county modeling
  - **ETS** (Error-Trend-Seasonal) for comparative accuracy
- Used `auto.arima()` from forecast package to select optimal lags
- Tested accuracy using Mean Absolute Error (MAE) and RMSE

## 4.3 Visualization and Reporting

- Visualized forecasts with `ggplot2` and `plotly`
- Automated summary reports per county using R Markdown

# 5. Statistical Results

County	Model Type	MAE (Last 6 Weeks)	RMSE	Forecast Behavior
Kings County	ARIMA	23.4	31.2	Spike expected mid-February
Lee County	ETS	17.9	28.6	Stable; early peak already passed
Benton County	ARIMA	21.7	30.8	Gradual incline projected

- ARIMA outperformed ETS in 17 out of 24 counties
- Forecasts aligned with previous peak cycles from 2018 and 2020

# 6. Interpretation and Action

- Health department adjusted delivery schedules:
  - Pulled vaccine shipments **one week earlier** in counties with forecasted early spikes
  - Increased awareness messaging in **3 counties** with highest projected growth
- Designed **flu load distribution dashboard** for weekly commissioner briefings
- Identified statistically stable counties for **routine instead of emergency distribution**

## 7. Reporting Output

- **R Markdown Report (PDF + HTML):**
  - 24-county model comparison
  - Time series plots with shaded confidence intervals
  - Policy-ready summaries of risk periods
- **Interactive Plot Dashboard (R Shiny Prototype):**
  - County selector
  - Forecast plots and historical overlays
  - Exportable CSVs for logistics planning
- **Excel Export:**
  - Forecasted weekly values for 12 weeks ahead
  - Risk scoring matrix by county and week

## 8. Public Health Impact

- Enabled **vaccine reallocation within 48 hours** of forecast issuance
- Hospital flu admission rates were **7.3% lower than projected** during peak weeks
- Pilot project scaled to include **COVID + RSV** time series for 2024 rollout
- Forecasting pipeline integrated into department's RStudio Connect server for **monthly automation**