Influenza Staffing Plan Preparation Project

Nuno Marques

May 2024

Overview

Client: A medical staffing agency covering all hospitals in the 50 US states wants help to build a staffing plan for next year's influenza season.

Problem: The US has an influenza season where more people than usual suffer from the flu. Vulnerable populations develop serious complications and end up in the hospital. Hospitals and clinics need additional staff to adequately treat these extra patients. It is hard to understand which states and hospitals should be prioritized when designing staff plans.

Objective: Determine the priority states to staff, based on historical data from 2009-2017.

Context: This project was completed in the context of Career Foundry's Data Analytics Program. The data is real, but the scenario and agency are fictitious.

Tools: Data cleaning, wrangling and analysis in Excel. Visualizations in Tableau.

Success Definition and Initial Brainstorming

- Objective: Create an adequate staffing plan for next year's influenza season by identifying critical states.
- Initially, I broke down the solution to this problem in several steps:
  - How do we define what adequate means in regards to staffing plans?
    - Minimal instances of understaffing (staff-to-patient ratio < 0.23) and overstaffing (staff-to-patient ratio > 0.27) across states.
  - What do we define as critical parameters for influenza severity across states?
    - Assumptions: Higher counts of vulnerable populations / Higher total counts of influenza deaths normalized by population / Higher overall influenza death counts / Lower vaccination rates

Staffing Adequacy

- In order to understand past year's critical states, it would be ideal to have data on staffing adequacy. This data was not available.

- Therefore, I advised agency staff to prepare for gathering this data for next year's staffing plan, by both colecting information regarding staff-to-patient ratio, and staff and patient satisfaction surveys regarding staff adequacy.

Data Collection

- Data on the US population by geography, year, age and gender was collected from the U.S. Census Bureau
- Data on US Influenza Deaths by Geography was collected from the the Centers for Disease Control (CDC)
- Data on Influenza-related visits to the hospital was collected from CDC
- Data on lab tests for Influenza in the US was collected from the CDC
- Data on Children Flu vaccination in the US was collected from the CDC.

Data Cleaning

- All data sets were assessed for accuracy, completeness, consistency, uniqueness, timeliness and validity using Excel.
- The datasets regarding Children Vaccination rates, Influenza-related hospital visits and Influenza Positive Lab Results were deemed insufficiently complete to be further analyzed. This is due to anonymization of data leading to suppressed information for several US counties.

Data Cleaning

- While cleaning the Influenza deaths dataset, I created frequency tables for several variables and noticed that a lot of data was missing for all age groups below 45 years. Again, this is a product of anonymization and suppression of data that could identify particular individuals

- No alternative dataset was found that could compliment the missing data, so the analysis proceeded on, focusing on populations older than 45.

Age Group	Count
5-14	1
15-24	1
25-34	21
35-44	81
45-54	510
55-64	1418
65-74	2307
75-84	3435
85+	4309
Grand Total	**12083**

Data Wrangling

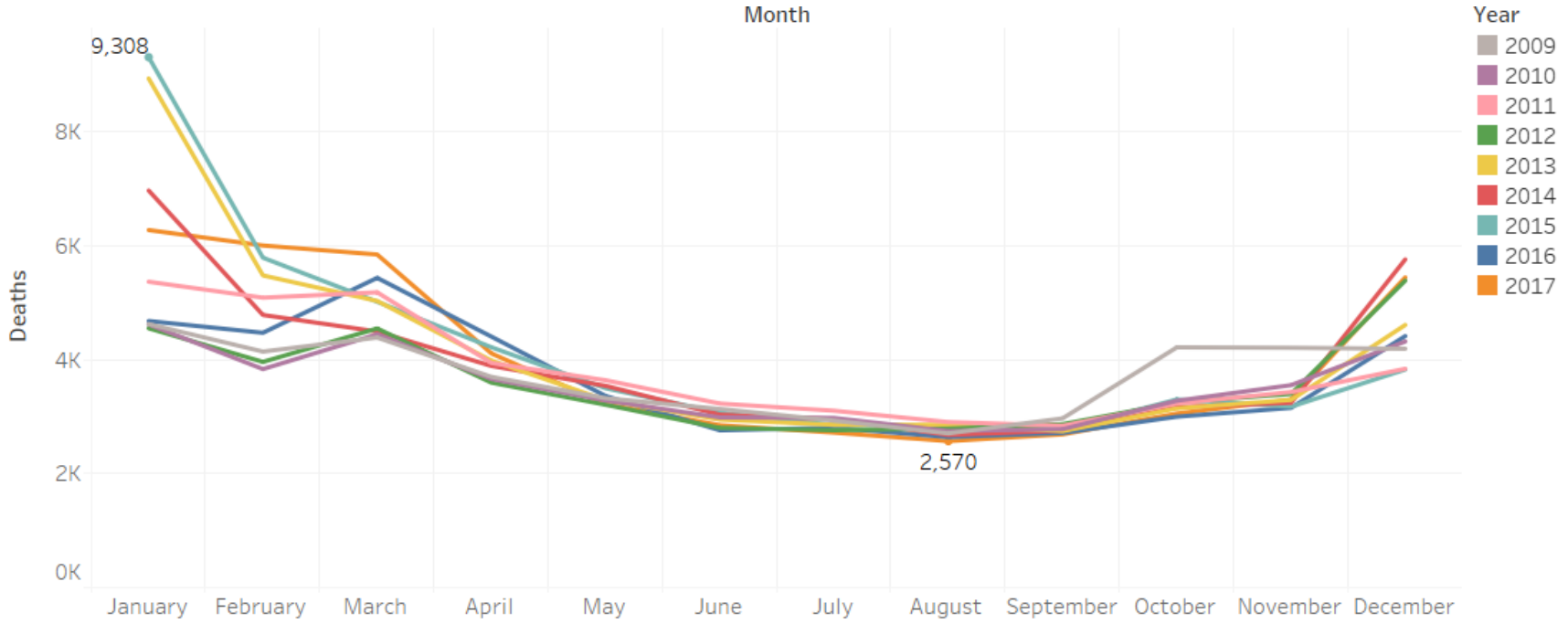
- The common spatial and temporal denominators of US State and Month of the Year(2009-2017) were found for the Influenza Deaths and Population datasets, and the two were grouped by and joined on these denominators.
- The Population was calculated for every state in every month and the Influenza Death Counts were normalized by this population.

Data Analysis – Key Questions

Are there any trends in influenza season's behaviour?

Is there a relation between vulnerable populations and death counts?

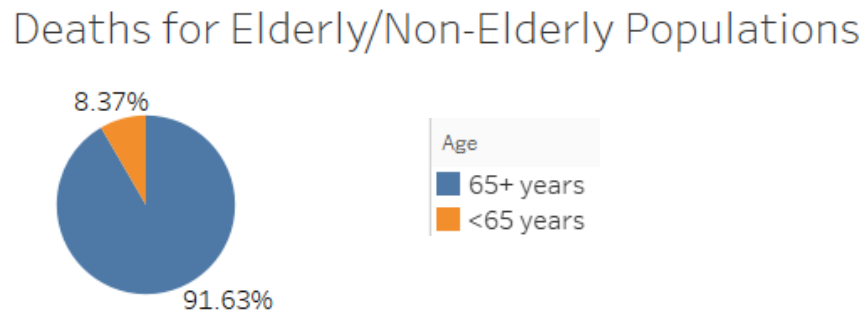
How do we prioritize which states to send staff to? Which are the critical ones?



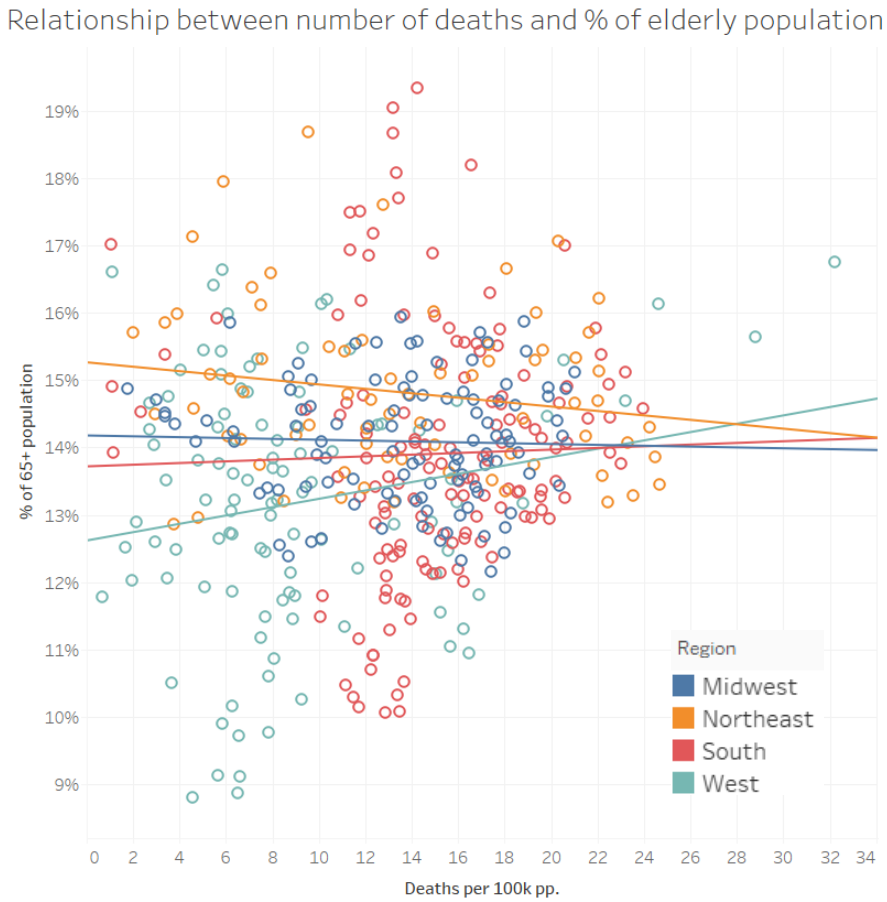
- The analysis of data from 2009 to 2017 reveals that the U.S. Influenza season predominantly occurs from November to March, with a notable surge in flu-related deaths starting in November which then tapers off post-March

Trends in Influenza season

Relation Between Vulnerable Population and Death Counts



Elderly Population, those aged 65 and over, is particularly susceptible, accounting for 91.63% of the recorded deaths, indicating a distinct and consistent risk to this demographic across the years observed.

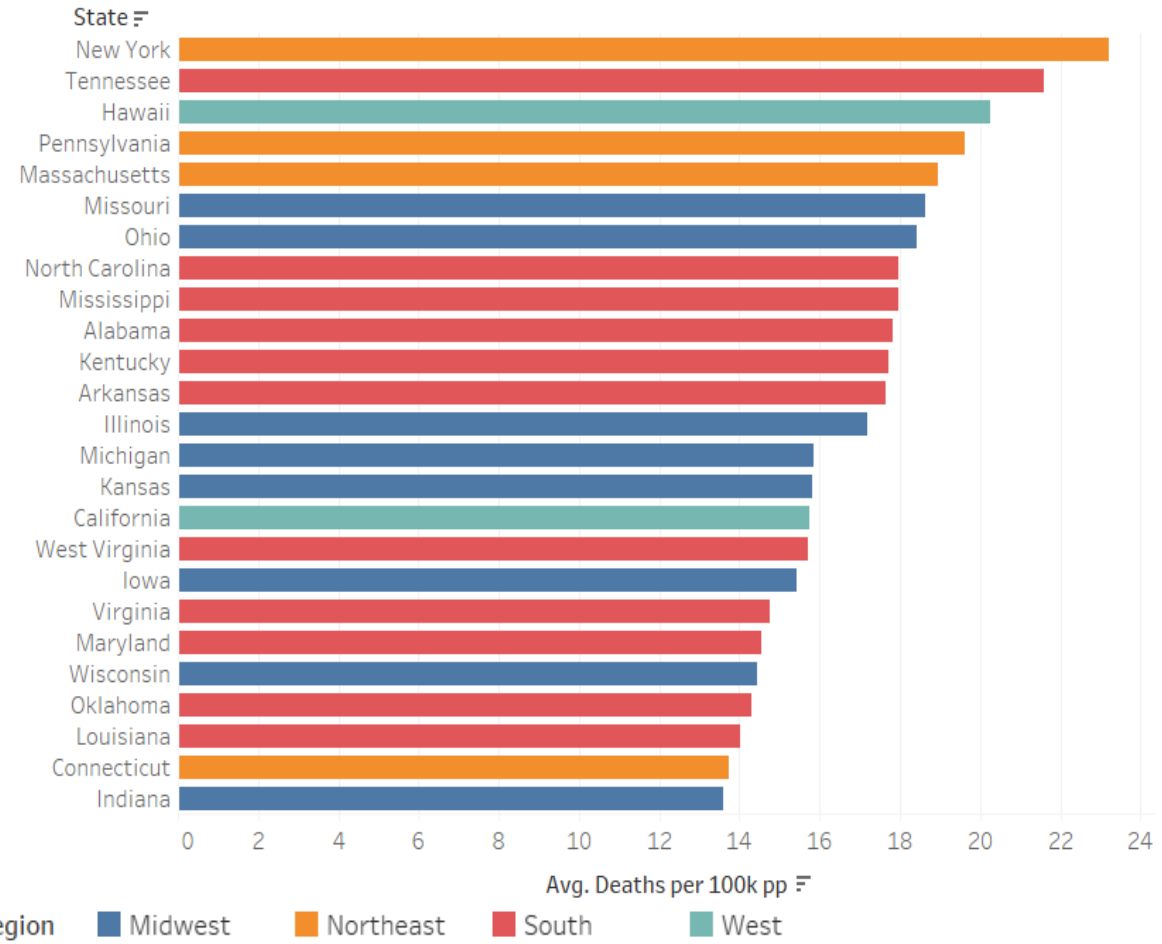


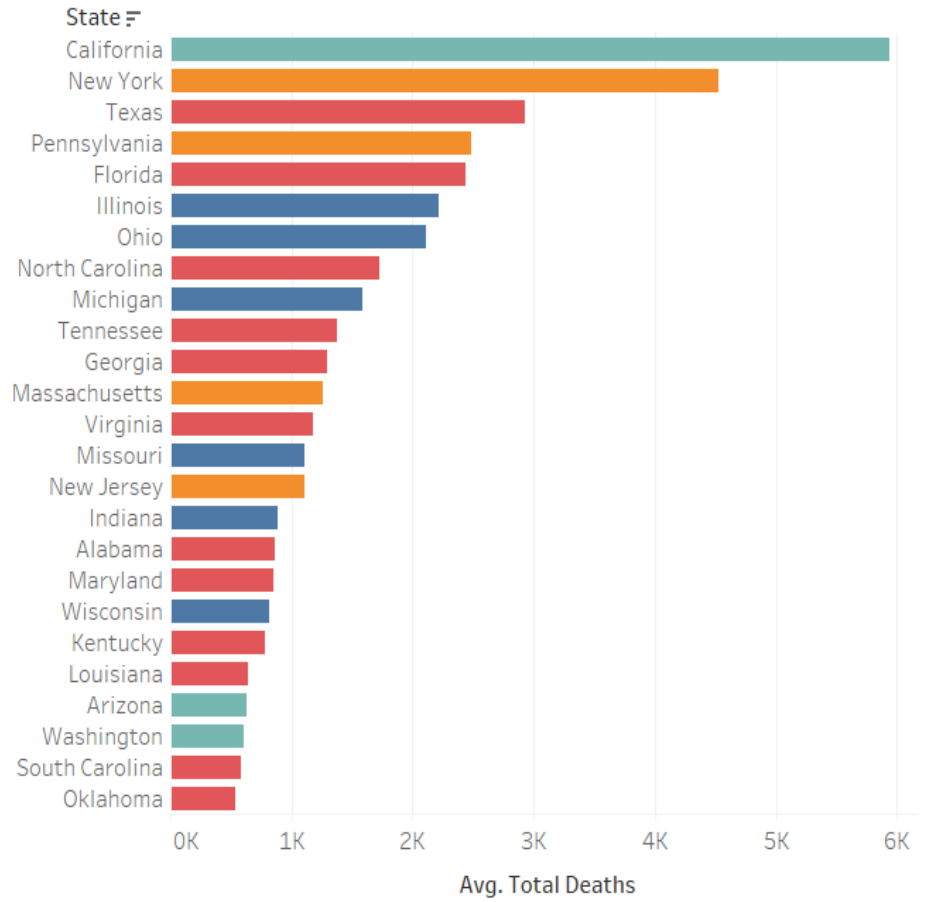
Despite this, **states with more elderly population do not consistently have more deaths relative to their population, as can be seem from the scatterplot to the left**. The trend lines are all different and mostly flat

Moving forward, the Death counts (Total and Normalized) will be the major factor in deciding which states will be prioritized for staffing needs, instead of vulnerable population counts.

State Priorization

The states with higher avg normalized deaths have **higher mortality per person**. However when creating a staffing plan, we need to also consider **total volume of deaths**, as high volume states will be more likely to **overwhelm health staff and facilities**

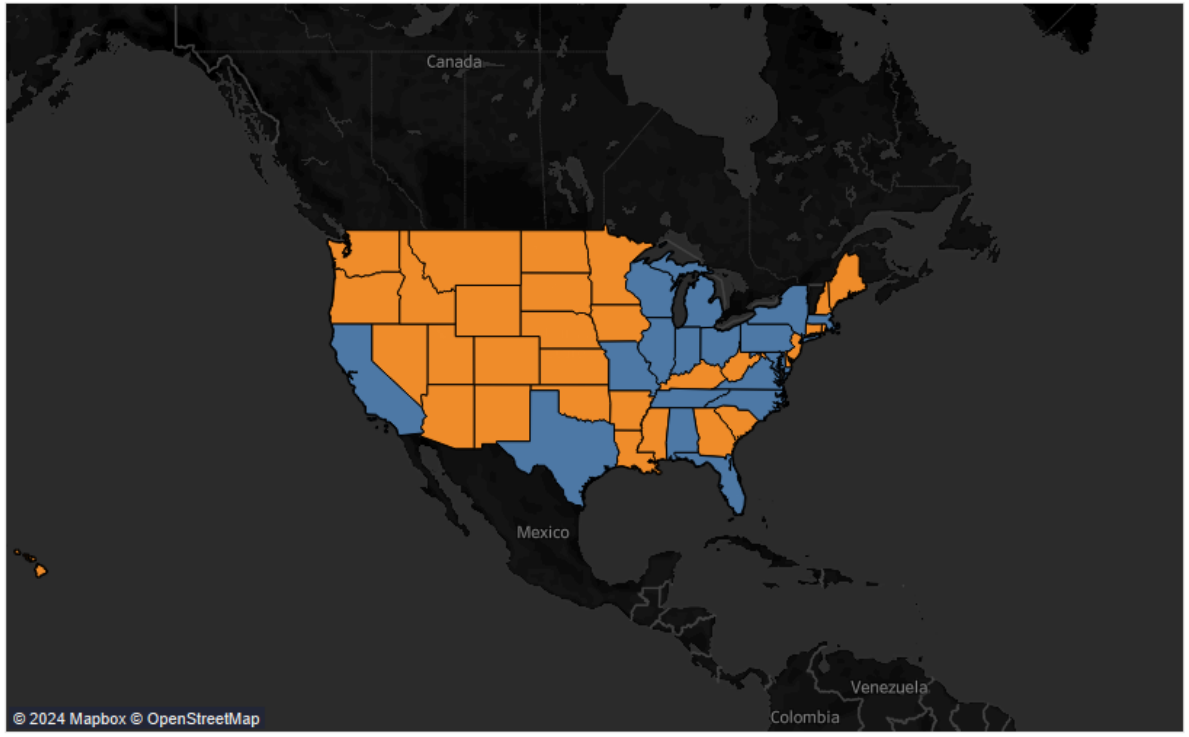




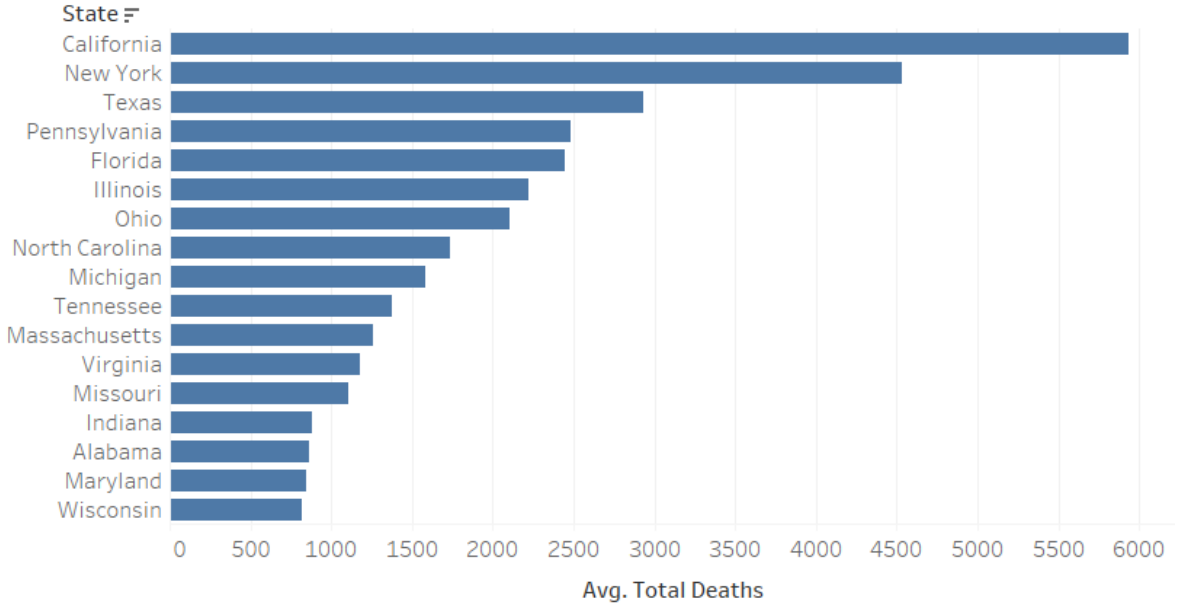
Recommendations

Due to elevated Total and Normalized Death counts, the states shown below will require additional staff during the months of November-March.

US States by Priority Level



Priority .. High Priority Low Priority



Conclusions & Next Steps

Are there any trends in influenza season's behaviour?

Yes, the influenza season lasts from November to March

Is there a relation between vulnerable populations and death counts?

There does not appear to be a significant relation between vulnerable population counts and death counts

How do we prioritize which states to send staff to? Which are the critical ones?

By taking into account both higher normalized and total influenza death counts, we were able to create a list of the most critical states

Next Steps

Seasonality should be assessed on a per state basis, particularly for the critical states

Data for younger age groups needs to be found to reach a conclusion, as this dataset misses a large portion of younger vulnerable population

Procedures for surveying efficacy of staffing plan need to be planed out and executed

Project Retrospective

- Major Challenges:
  - Dealing with missing data
  - Aggregating and displaying information for 51 states in an understandable and statistically meaningfull manner
- Learnings:
  - Sometimes, there is no complementary data source available, so you need to be ready to work with clients/colleages to set up the correct data collection processes
  - Researching standard aggregations of our specific data type(geographical in this case), and applying them as necessary to make data easier to read, but still statistically meaningful

Supporting Material

Data

1. Influenza deaths by geography
Source: CDC
Download Data Set

2. Population data by geography, time, age, and gender
Source: US Census Bureau
Download Data Set

Project Walkthroughs

Full analysis breakdown available on Tableau Public

Video Presentation Walkthrough

Skills/Tools

- Excel
- Designing a data research project
- Data profiling and integrity
- Data transformation & integration
- Statistical hypothesis testing
- Data visualization
- Spatial & textual analysis
- Forecasting
- Storytelling in Tableau

Contact info

Reach out at: