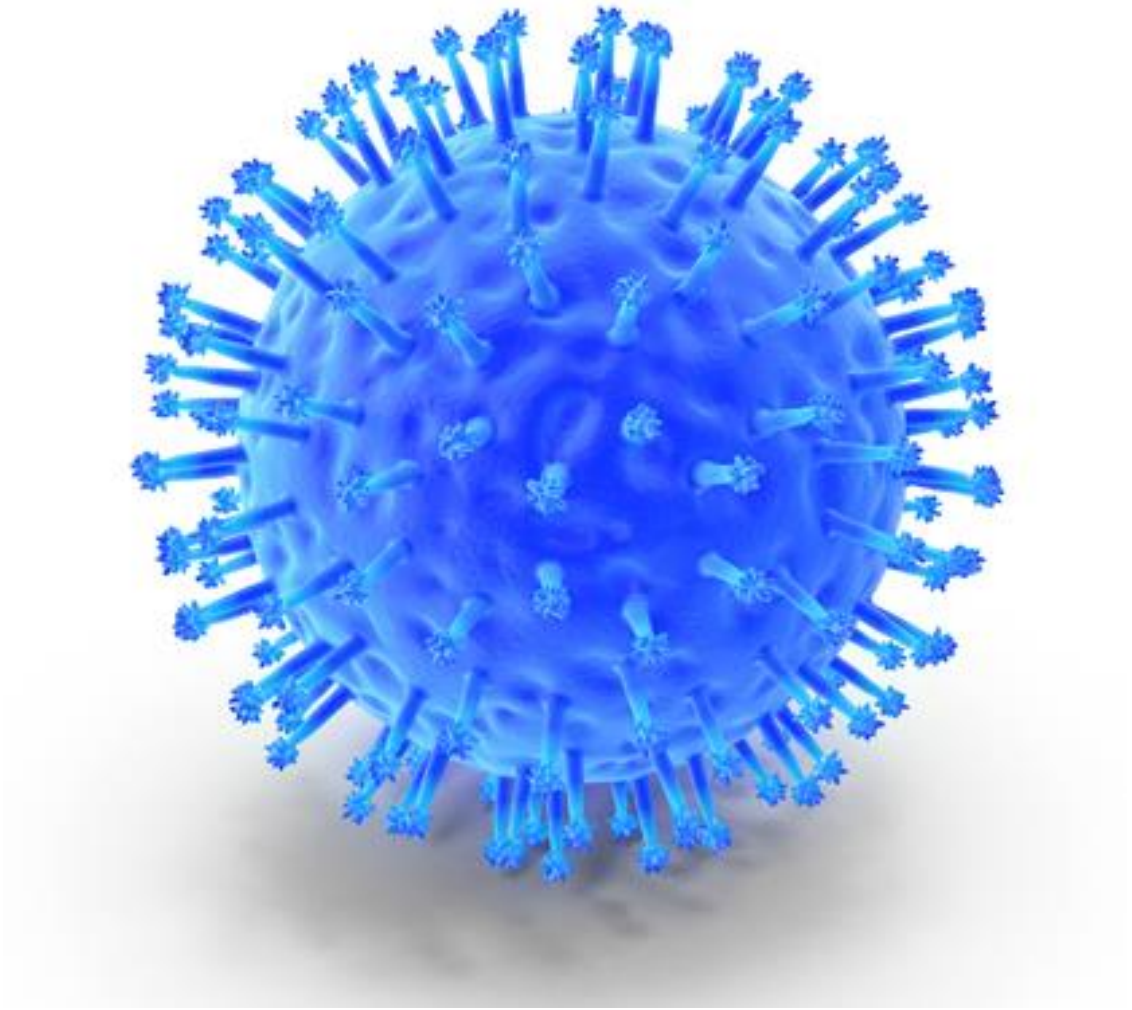
PREPARING FOR FLU SEASON IN THE US

Helping Medical Staffing Agency



INTRODUCTION

PROBLEM

The United States has an influenza season where more people than usual suffer from the flu. Some people, particularly those in vulnerable populations, develop serious complications and end up in the hospital. Hospitals and clinics need additional staff to adequately treat these patients.

GOAL

The goal is to help a medical staffing agency to plan for the influenza season, a time when additional staffs are in high demand.

ROLE

As a Data Analyst, my job was to examine trends in influenza, to provide some insights about the infection trends across the country, and make suggestions about the staffing needs to the agency

TOOLS

Excel, PowerPoint, Tableau Public

Datasets

CDC Influenza Deaths

Source: Center for Disease Control
Summary: Records of influenza-related deaths by geography (state), time, age and gender

Population Data

Source: US Census Bureau
Summary: population data by geography (state and country), gender and age groups

Influenza-Related Visits

Source: Center for Disease Control
Summary: Count of Influenza related visits by state, year & age groups, count of providers per state by year and week

Techniques & Methodologies

- Sourcing and cleaning of the databases
- Identification of matching data for merging the databases
- Conducting statistical analysis and testing correlation between variables
- Preparing an interim report
- Creating visualization to show the correlation between chosen variables
- Showing results to the stakeholders

SKILLS USED

- Data Cleaning
- Data Integration & Transformation
- Statistical Hypotheses Testing
- Visual Analysis & Forecasting
- Story Telling in Tableau

CHALLENGES

Data Inconsistency

The CDC influenza deaths set has inconsistency

- The reported age groups; start with “under 1 year,” “1-4 years” and then starting the 15-24 years and goes up by 10 years
- Some of the state names are written in abbreviations
- The death counts under 10 are recorded as “Suppressed” values
- The Census Population set lists the population count by state/county, gender and age groups going up by 5 years

Solutions

- I performed extensive data cleaning and transformation on the datasets in this analysis to create an integrated dataset with uniform metrics and variables.
- Since it is impossible to know the exact value of “suppressed” records & we need numeric values to perform statistical analysis, I replaced these values with zero. Considering the suppressed value represents values under 10, the impact on the final hypothesis testing was insignificant.
- I created 2 age groups, one above 65 years old (vulnerable population)and one under 65 years old, which helped create uniform age groups in all sets.

TASK PERFORMED

Defining the Research Questions and Hypothesis

Clarifying Questions:

- Which states are most affected by influenza?
- When is flu season?
- Which age groups have the most vulnerable populations?

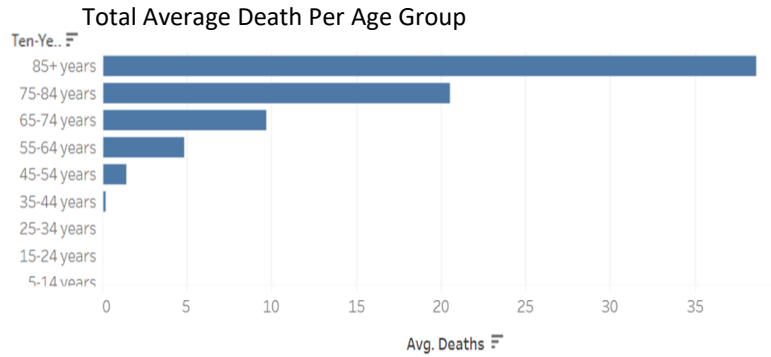
Funnelling Questions:

- Is flu season the same in every state?
- Is flu season the same length every year?
- Is flu season only once a year?

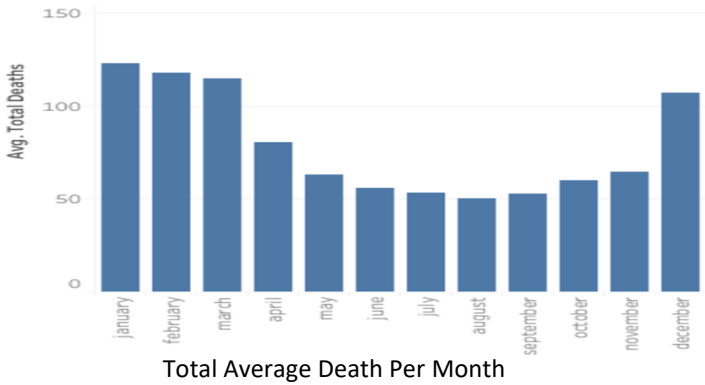
Defining Hypothesis:

If a state has a larger proportion of vulnerable population, then more deaths from flu will occur.

I computed the chart below, Total Average Death per age group, to find out which groups have the most vulnerable population.



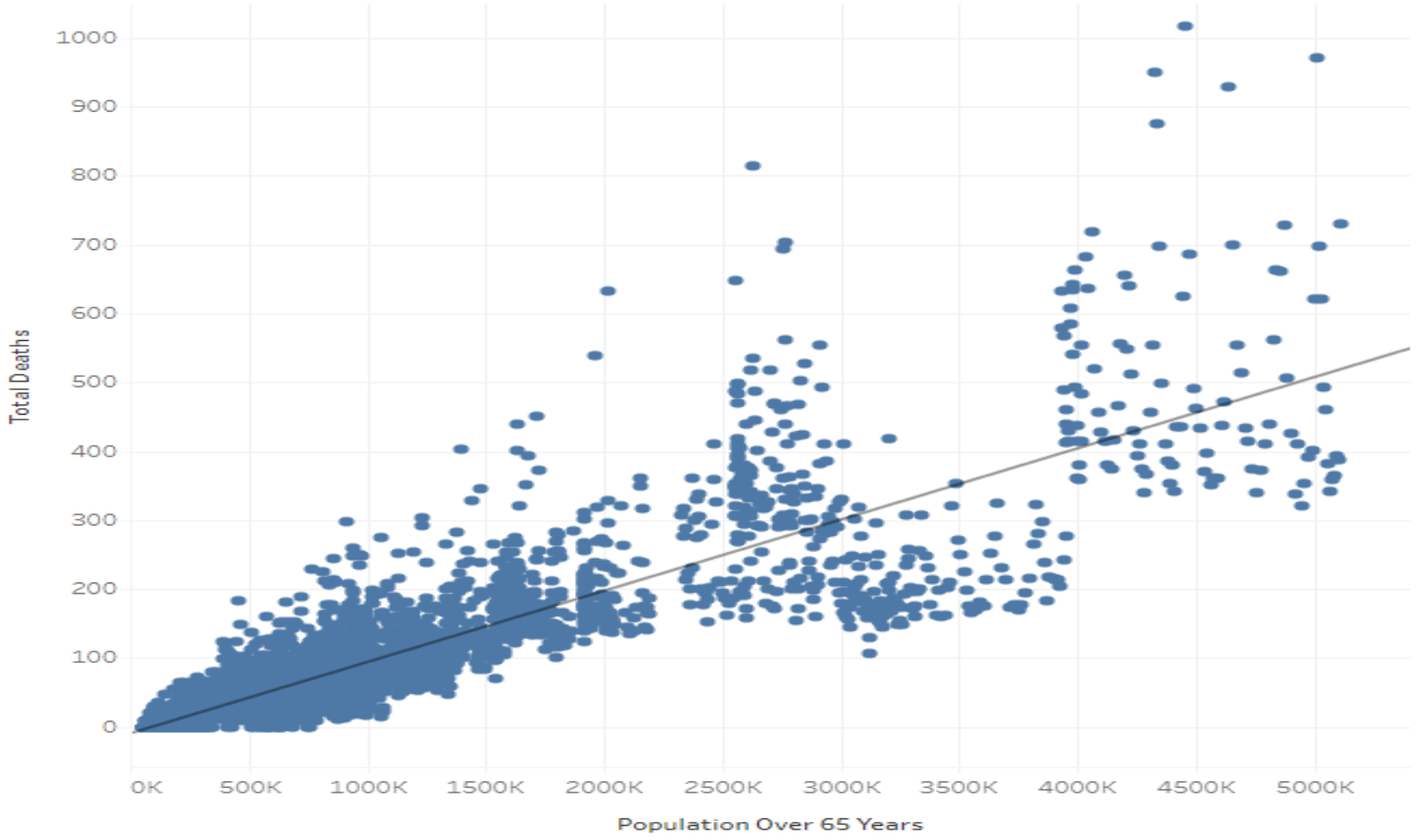
The chart below answers our question, that flu season is not the same every year, it peaked from December to March.



Total Average Death Per Month

HYPOTHESIS TESTING

The next step was to test the hypothesis. After sourcing the necessary variables, I conducted descriptive analysis and plotted the scatter plot between Total death due to influenza and vulnerable population which is shown below.



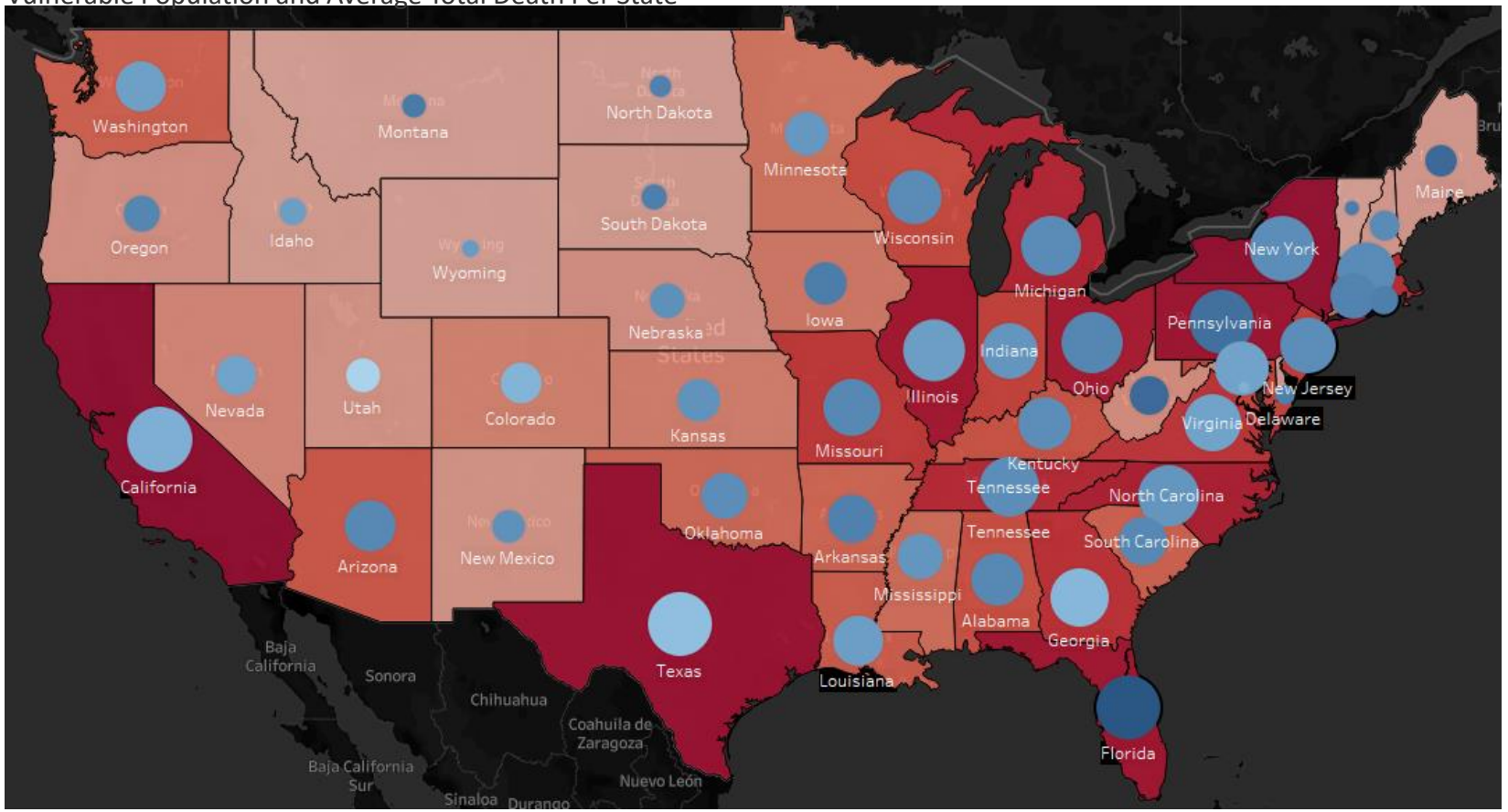
Pearson's correlation coefficient of 0.88, which indicate a strong correlation between population over 65 years and the mortality rate in that state

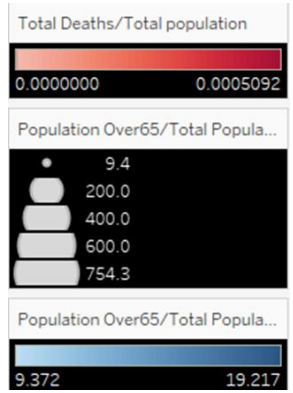
The strong results of the hypothesis testing allowed me to focus the analysis on the older age group.

VISUALIZATION INSIGHT

After finding a strong relationship in the previous step, I continued by creating a map showing the distribution of vulnerable populations per death rate in the U.S.A, using State as the spatial boundary.

Vulnerable Population and Average Total Death Per State





The darker the redness of the state, the higher the death ratio

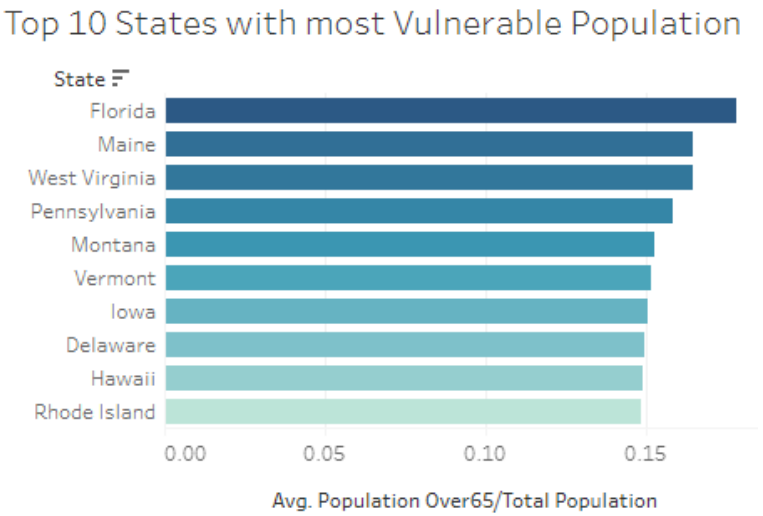
The darker and bigger the blue circle, the higher the ratio of vulnerable population.

The map further confirms our hypothesis, that states with high vulnerable population ratio lead to higher death rates.

CONCLUSIONS & RECOMMENDATIONS

Based on these results, I was able to conclude that if a State has a more vulnerable population it will have more deaths due to influenza (flu). Therefore states will the highest vulnerable ratio should be prioritized planning for staffing.

Furthermore, seasonality should be taken into consideration when planning for staff, more staff should be allocated during the peak season



Limitation

Although the goal of the analysis is to deliver staffing recommendations, there is no record of the number of staff per provider for each state by year. Due to the limitations, I cannot make direct suggestions on the exact number of staff that needs to be allocated to each state, however, I have listed the states in terms of priorities for staffing needs while taking seasonality into consideration. This should provide the staffing agency enough insight to plan for the next flu season

Click to see Tableau Storyboard

Click to see video Presentation

GET IN TOUCH

https://lionel1801.github.io
https://www.linkedin.com/in/lionel1801

Write me an email:

lioneleyong@gmail.com