**Preparing for Influenza Season**

**Tools: Excel (Pivot, VLOOKUP, integration)**

**Tableau Public**

Posted by Graham Field on February 10, 2024 – 10 mins read

**Duration of the project:** 1 week

**Role:** Data Analyst

**Tools:**

* Excel (Pivot, VLOOKUP, integration)
* Tableau Public

**Data:**

* Open Source Data
* 4 Possible Data Sets

**Key Research Questions**

● Provide information to support a staffing plan, detailing what data can help inform the timing and spatial distribution of medical personnel throughout the United States.

● Determine whether influenza occurs seasonally or throughout the entire year. If seasonal, does it start and end at the same time (month) in every state?

● Prioritize states with large vulnerable populations. Consider categorizing each state as low-, medium-, or high-need based on its vulnerable population count.

● Assess data limitations that may prevent you from conducting your desired analyses.

**Introduction**

**Introduction:** The annual influenza season in the United States poses significant challenges for healthcare providers. Vulnerable populations, such as young children and the elderly, are particularly at risk of developing severe complications from the flu, which often leads to increased hospitalizations. Medical staffing agencies play a crucial role in supporting hospitals and clinics by providing temporary staff to manage the surge in patients.

**Objective:** The primary objective of this analysis is to determine the optimal timing and allocation of staff to hospitals across the United States during the influenza season.

**Scope:** The project encompasses all hospitals within the 50 states of the United States, focusing on the upcoming influenza season to ensure adequate staffing and resource distribution. By examining historical data and identifying trends, we aim to enhance preparedness and response strategies for healthcare facilities.

**Step 1: Initial Data Exploration and Cleaning**

After drawing requirements from the stakeholders and the business requirements document I evaluated each data set. There were several considerations with the data namely: its reliability, ownership and ethical concerns. I don't want to present any data that could lead to the identification of individuals, given the scope and scale of this medical data.

The focus of this was to send staff and allocate resources to areas which experience the greatest need. I focussed on areas with the highest mortality rates.

From the 4 posible data sets i selected :

1. US Geographical Census Data
2. US Influenza Deaths

**Step 2: Cleaning and Consistency Checks**

During the exploration of the data i found several inconsistancies namely: the naming of states and areas which were recored but not part of the US such as Puerto Rico.

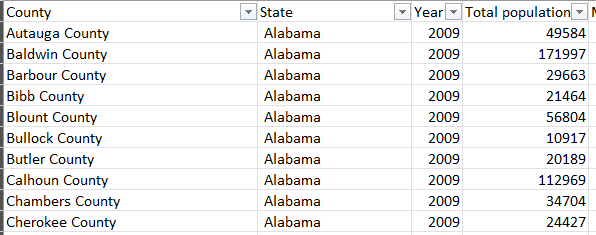
There was some missing data due to privacy requirements where data that was less than 10 was removed. I imputed using both random and averaged methods. But ultimately chose the random seed.

**Step 3: Integration**

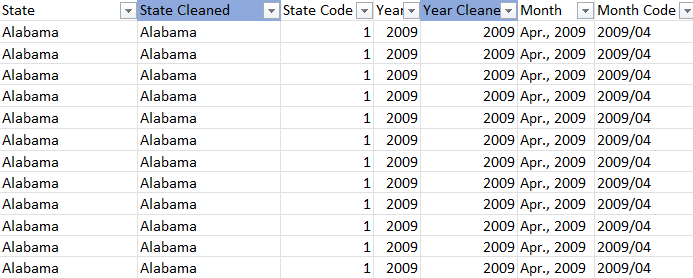
In orderto use the data sets for analysis i had to integrate or merge them. This is due to the data sets often complete and ready for use.

I bean by mapping the data and trying to find commonalities between the data sets. I identified “State” and “Year” as common columns. There was an issue with granularity as The US census data had state and county with year while the Influenza Deaths had State with Year and Month.

Now that I have a common key to work with, I created Pivot tables which summarised the data on both data sets and allowed me to Integrate the two data sets. I used VLOOKUP to integrate the data sets into one.



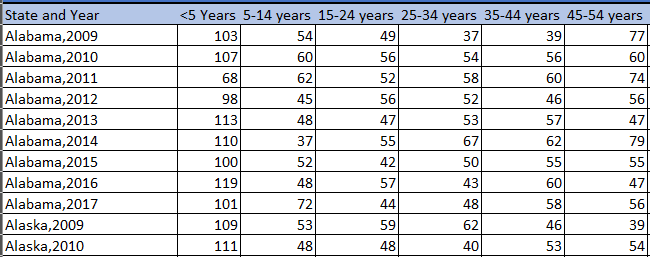
US Census Data



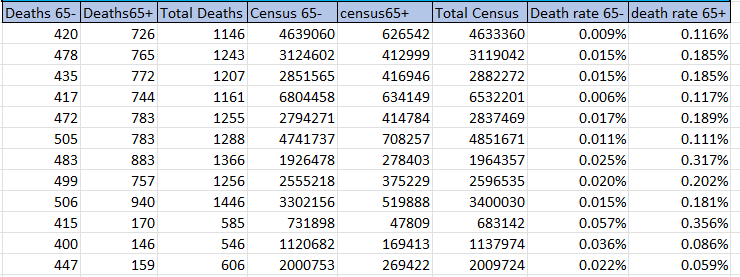
US Influenza Deaths Data

Now that I have the data sets integrated I began to segment them into age groups of 10 years and then created a secondary segment for retired vs non-retired or younger than 65 and older than 65.

A table with numbers and text

Description automatically generated

Now that I had the two tables with matching data I could create a new normalised data table that would give me death rates by dividing the census data to the matching Influenza deaths.

****

**Step 4: Statistical Analysis and Testing**

I began by stating 2 hypothesis.

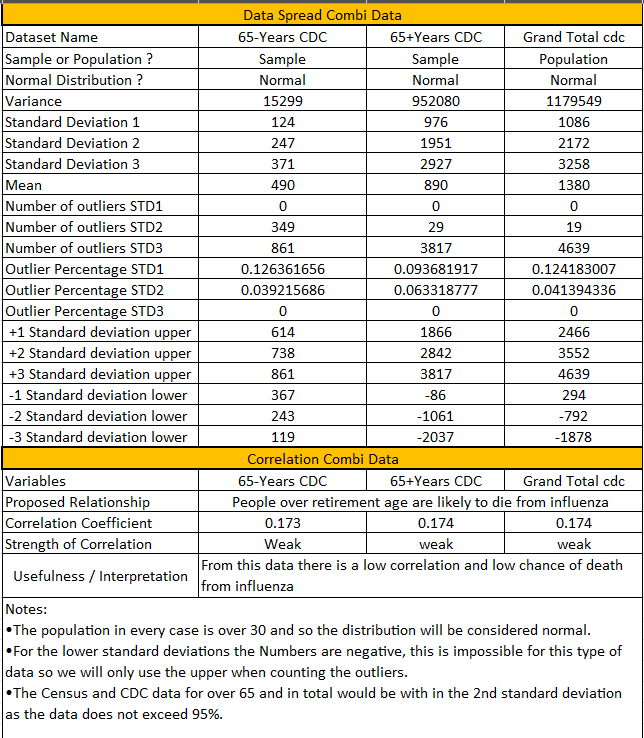
Null Hypothesis (H0): The mean influenza mortality rate for individuals aged 65 and older is equal to the mean influenza mortality rate for individuals under 65.

**H0: µInfluenza mortality 65+ years = µInfluenza mortality <65 years**

Alternative Hypothesis (HA): The mean influenza mortality rate for individuals aged 65 and older is not equal to the mean influenza mortality rate for individuals under 65.

**HA: µInfluenza mortality 65+ years <> µInfluenza mortality <65 years**

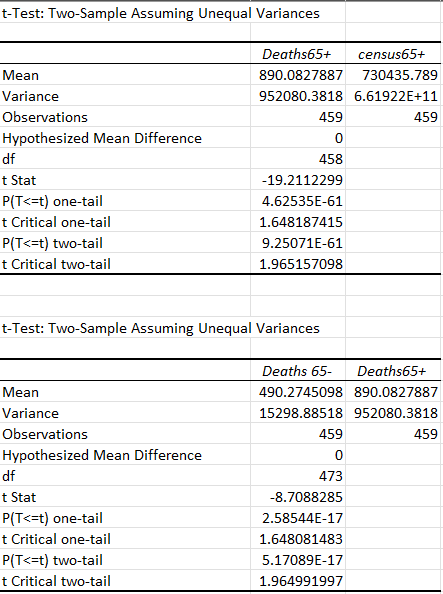
The goal was to test whether the two groups had the same risk of death from influenza the first test I ran was a statistical test to determine: Distribution, Standard Deviation, Number of outliers, and correlation. I wanted to determine if the average mortality rate was higher or lower for the two age groups.



The next test I ran was the two-tail test because I was interested in confirming the results from the previous test and confirming or rejecting the null hypothesis with certainty. I decide to include both Equal and unequal variance tests for the following reasons.

The first test shows the difference between the deaths in the 65+ age group and the overall census data for the 65+ population, highlighting the impact of influenza within this age group relative to its population size.

The second test directly compares the mortality rates between the two age groups, showing the disparity in influenza mortality between younger and older individuals.



**Analysis of the t-Tests**

**Test 1: Deaths65+ vs. Census65+**

* **t Stat**: -19.2112299
* **P(T<=t) two-tail**: 9.25071E-61
* **t Critical two-tail**: 1.965157098

This test shows a highly significant difference between the mortality data for the 65+ age group and the census data for the same age group. However, this is not directly related to the null hypothesis about comparing mortality rates between age groups.

**Test 2: Deaths 65- vs. Deaths65+**

* **t Stat**: -8.7088285
* **P(T<=t) two-tail**: 5.17089E-17
* **t Critical two-tail**: 1.964991997

This test directly compares the mean influenza mortality rates between the two age groups:

* **t Stat**: -8.7088285
* **t Critical two-tail**: 1.964991997
* **P(T<=t) two-tail**: 5.17089E-17

**Conclusion**

The p-value for the second test (5.17089E-17) is extremely low, much lower than the typical significance level (α) of 0.05. This indicates that the difference in mean mortality rates between individuals aged 65 and older and those under 65 is statistically significant.

**Decision:**

* Since the p-value is much less than 0.05, we reject the null hypothesis (H0).

**Summary:**

Based on the results of the t-test comparing the mortality rates between the two age groups, we have sufficient evidence to reject the null hypothesis. This indicates that there is a statistically significant difference in the mean influenza mortality rates between individuals aged 65 and older and those under 65.

**Step 5: Visual Analysis**

**A graph with a line and a pie chart

Description automatically generatedDefining Influenza and Who Is at Risk?**

**What is Influenza?** Influenza, commonly known as the flu, is a contagious respiratory illness caused by influenza viruses. It can cause mild to severe illness and at times can lead to death. The flu spreads through respiratory droplets when people with the flu cough, sneeze, or talk.

**Who is at Risk?** Influenza significantly impacts certain age groups more than others. The most vulnerable age groups include:

* **Children younger than 5 years old**, with a mortality rate of 3.2%.
* **People over the age of 65 years**, with an average mortality rate of 4.6%. Within this group:
  + 65-74 years: 1.4% mortality rate
  + 75-84 years: 3.7% mortality rate
  + 85+ years: 8.7% mortality rate

**Vulnerable Groups:** The chart shows the death rates across various age groups, highlighting the higher risk among the very young and the elderly. These populations should be the focus of targeted interventions to reduce mortality rates during the influenza season.

A graph with numbers and a number of percentages

Description automatically generated with medium confidence

**Where and Who is at Risk?**

**Geographical Analysis:** The geographical distribution of influenza deaths shows significant variations. The top 15 states with the highest number of deaths include California, New York, Texas, Pennsylvania, and Florida, which are also among the most densely populated states.

**Regional Deaths Geographically:** Mapping the deaths geographically allows us to identify high-risk areas. This information is crucial for directing resources and planning healthcare strategies.

**Total Deaths by Age Group:** Analyzing the total deaths by age group reveals that older adults, especially those above 65, are at the highest risk. This demographic data is essential for preparing healthcare systems to meet the needs of these vulnerable populations.

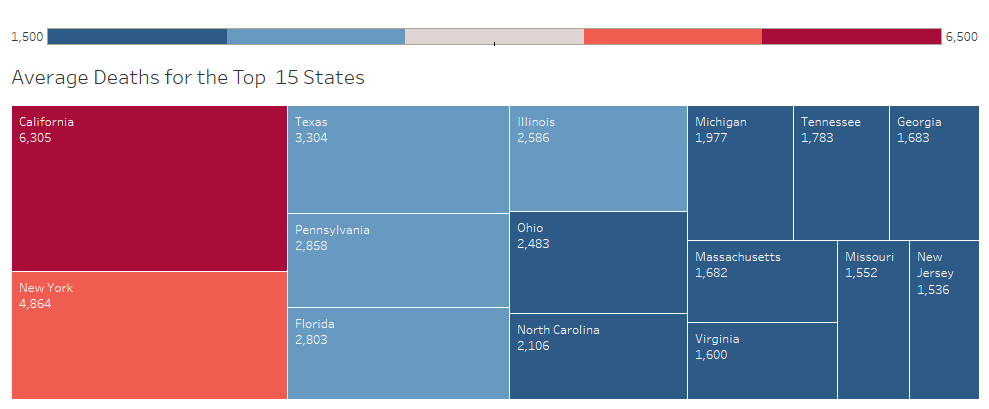
A screenshot of a computer screen

Description automatically generated

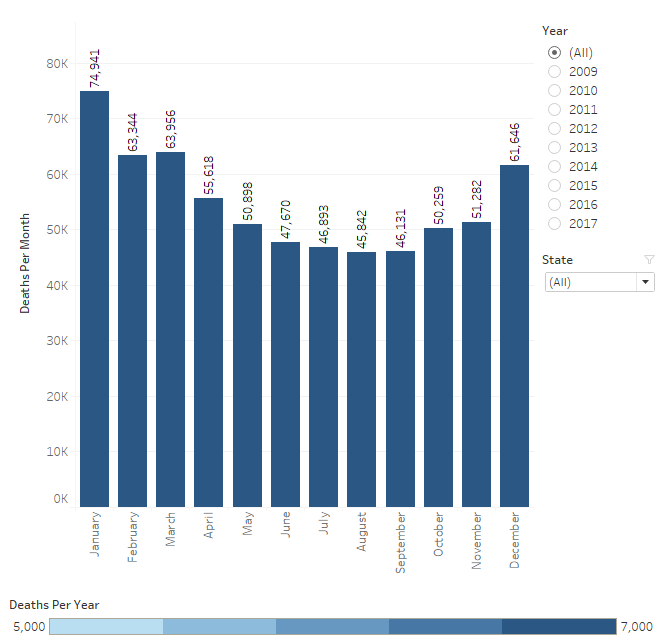
**When is Influenza Season?**

**Influenza Seasonality:** Influenza season typically starts in December, peaks in January, and declines by March. These months coincide with colder weather, which contributes to the spread of the virus.

**State-specific Analysis:** California, New York, and Texas have the highest number of deaths, correlating with their large populations. This trend underscores the need for targeted resource allocation during peak months.

****

**Monthly Distribution of Deaths:** Understanding the monthly distribution of deaths can help in planning the deployment of medical staff and resources. Ensuring that healthcare facilities are adequately staffed during peak months can save lives and improve patient outcomes.



**Conclusion and Recommendations**

**Conclusion:** The analysis shows that certain age groups and geographical regions require more resources during the influenza season. Specifically, children under 5 and adults over 65 years need prioritized care due to higher mortality rates.

**Recommendations:**

1. **Prioritize Care for Vulnerable Age Groups:** Ensure that resources are directed towards children under 5 and adults over 65.
2. **Prepare for Seasonal Peaks:** Staffing and resources should be heightened from December to March.
3. **Allocate Resources to High-Risk States:** States like California, New York, Texas, Pennsylvania, and Florida should receive the majority of resources. The next tier of states should receive moderate resources.
4. **Enhance Public Health Campaigns:** Increase awareness and vaccination efforts, particularly among high-risk groups and in high-risk states.

**Further Analysis:** Future research should include additional data on vaccination rates, living standards, nutrition, and substance use to better understand the spread and survival rates of influenza. Collaboration with public health agencies can provide a more comprehensive approach to managing influenza seasons.

**Ongoing Monitoring and Data Limitations**

**Ongoing Monitoring:**

1. Continue to track death rates over time to identify trends and emerging patterns.
2. Include comparisons between vaccinated and unvaccinated populations to assess the effectiveness of vaccination campaigns.
3. Assess how influenza deaths compare to other leading causes of death to allocate healthcare resources effectively.

**Data Limitations:**

1. Limited information on general health, fitness, diet, and hereditary factors can affect the analysis. More detailed health data could improve predictive models.
2. High-population density states with low death rates should be studied to identify effective health practices that could be implemented elsewhere.
3. Data quality and completeness are crucial; ensuring accurate reporting from healthcare facilities will enhance the reliability of future analyses.

**Assessment Recommendations:**

* Regularly update and review data to reflect current trends.
* Engage with local health departments to validate and supplement data.
* Implement technological solutions to improve data collection and analysis.

By addressing these limitations and continuing to monitor relevant data, we can improve our understanding of influenza patterns and enhance our preparedness for future seasons.