

Influenza Season

Project Goals

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 extra patients. The medical staffing agency provides this temporary staff.

In this project I will give advice on when to send staff, and how many, to each state in the upcoming season.

The data that I used

Excel: 471 rows, 38 columns

Time period: 2009 until 2017

Regions: the States of the United States

Other data: per age group: deaths, population and % of deaths

Created: pivot tables, merged datasets and visuals

Data Sources: Influenza deaths, Population data,

The skills that I applied

Excel

Interpret Business Requirements

Data sourcing, Data Profiling Data cleaning,

Data integration, Data transformation

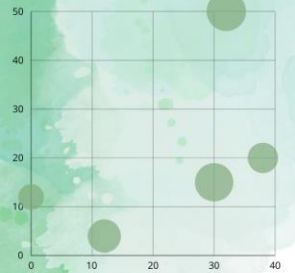
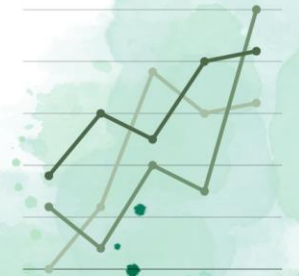
Statistical hypothesis testing

Visual analysis

Forecasting

Storytelling in Tableau

Oral presentation of the Tableau storyboard



The process

For the first part of the project, I had to create a project management plan to manage my time, the meetings with stakeholders and to plan the deadlines for the project deliverables. I had 14 weeks to finish the project. For the structure of the project, I formulated a hypothesis and conducted statistical hypothesis testing to prove or disapprove the hypothesis. After researching the data and creating several clarifying and funneling questions, I came to the following hypothesis:

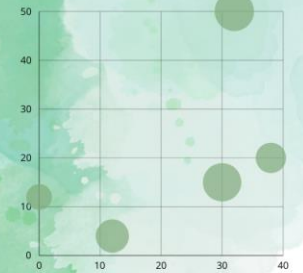
If you are 65 years and older, then the chances of dying from the flu are higher.

As I was working with two datasets in Excel, I had to profile and clean the data of both sets before I merged them together in one big file. With the number of deaths and the population details, I could create the % of deaths of all the different age groups.

After doing thorough analysis on the number of deaths under and from 65 years old I came to the conclusion that the p-value is less than 0.01, lower than the significance level of 0.05 which proves the hypothesis is true.

	% of deaths 0-64 years	% of deaths 65 years and older
Mean	0,000269646	0,001313872
Variance	7,66014E-08	2,75655E-07
P-value	4,61452E-170	

I completed the project by adding my analysis results as a storyboard in Tableau and presenting it in a video recording which I sent to the stakeholders. Some of the visuals of the analysis I have added in the following slides.



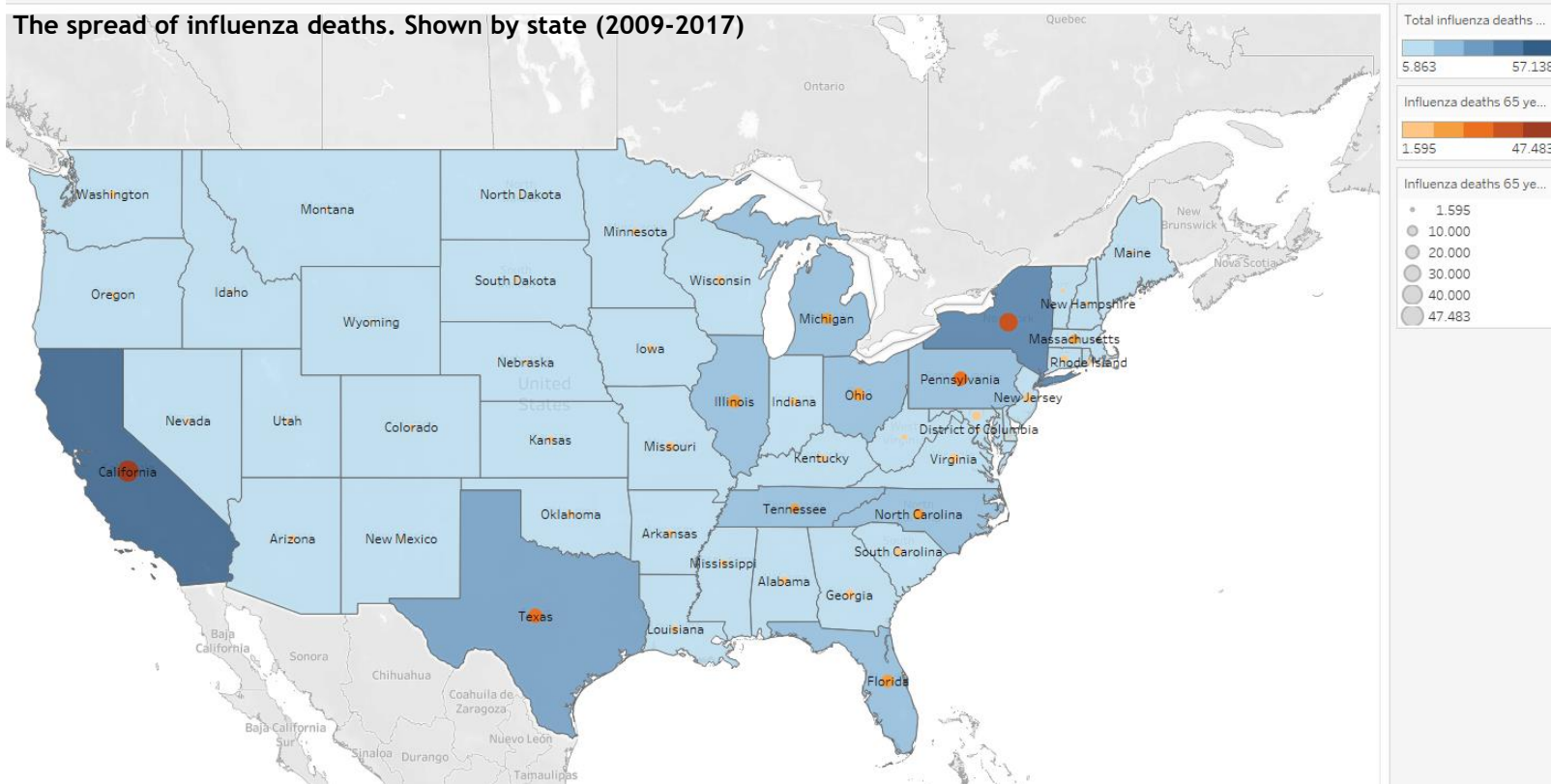
Analysis

The states suffering most
(from highest to lowest) are:

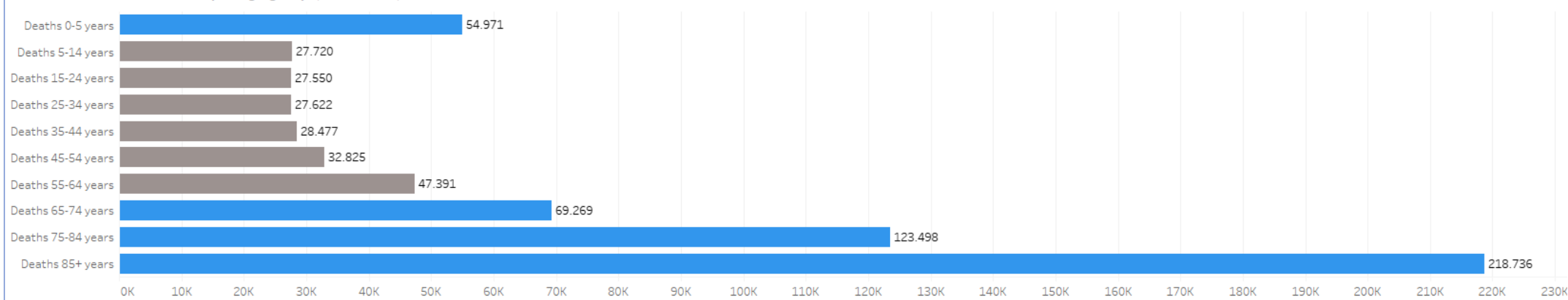
California
New York
Texas
Pennsylvania
Florida

Illinois
Ohio
North Carolina
Michigan
Tennessee

The spread of influenza deaths. Shown by state (2009-2017)



Number of influenza deaths per age group (2009-2017)



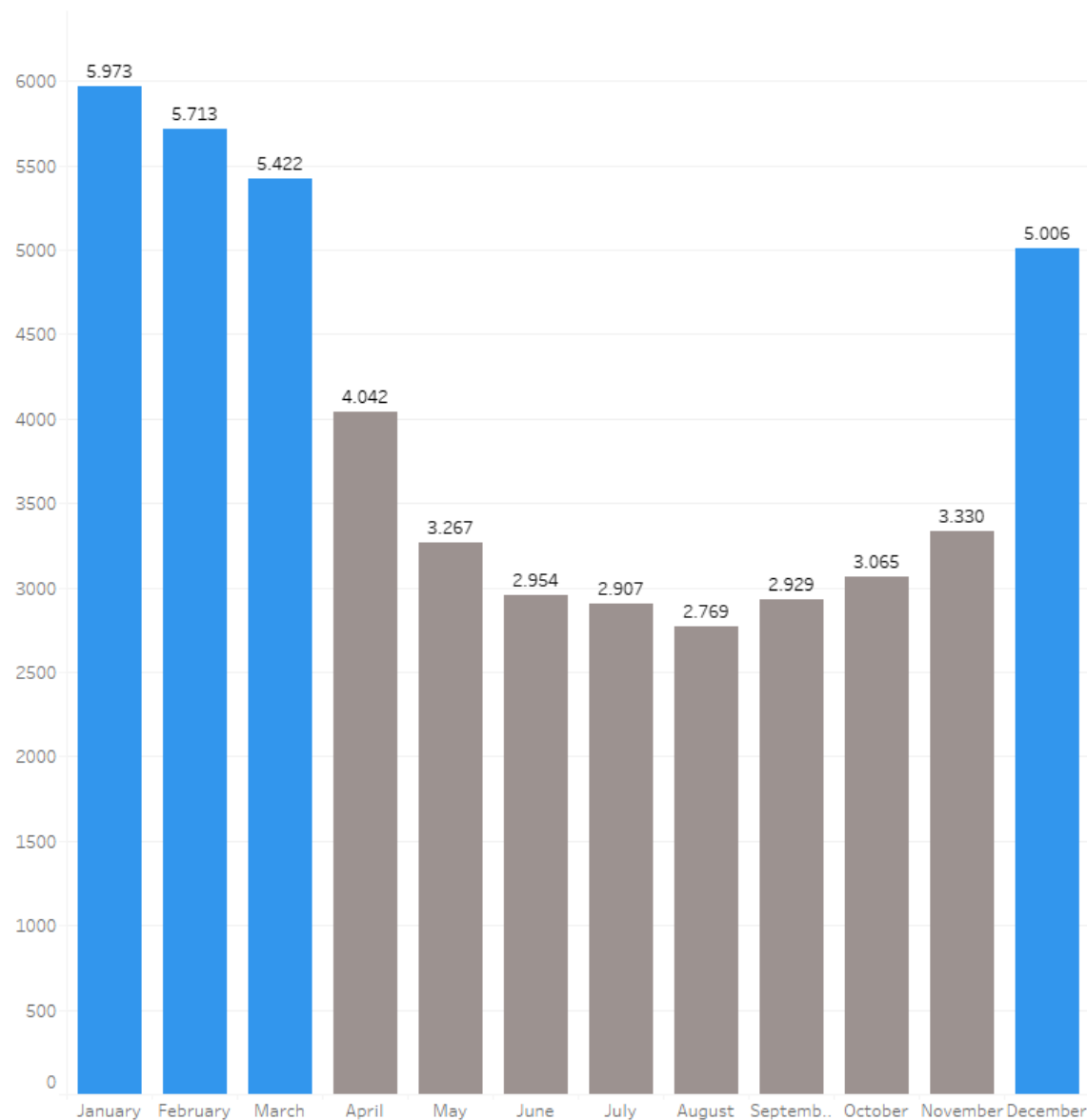
Analysis

The chart shows the influenza deaths of 65 years and older for the year 2017.

The months December until March show the highest numbers in influenza deaths for the vulnerable group of people of 65 years and older.

Important to know is that even in low season influenza deaths keep occurring.

Influenza deaths 65 years and older per month and year



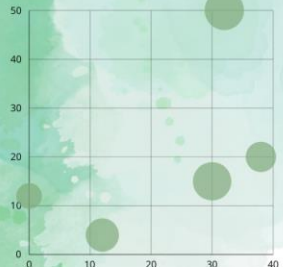
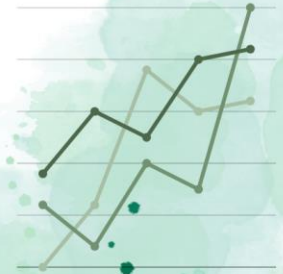
Recommendation

I would advise the Medical Staffing Agency should distribute their additional staff over the 10 states that are most in danger of influenza deaths. California has had already 33% more influenza deaths over the last 9 years than New York, which has the second highest numbers.

The other states are low priority.

Even though the focus should be on the winter months, the other months also show high numbers in influenza deaths. It is wise to still have a capacity of around 75% of additional staff to help out from April until November

I would also recommend to repeat the analysis for the results of 2018, and make modifications in the distribution of additional staff if necessary.



Personal evaluation



Successes

Formulating the hypothesis seemed daunting to me at first, but I came to it quickly and I was able to create a strong analysis around it.

This was my first project working with Tableau. I have always enjoyed working with creative programs such as Canva and Gimp, and working with Tableau was exciting. Some visuals took me longer than others, but overall using this tool goes natural for me.

Challenges and lessons learned

When I merged the two datasets together, there were errors in some values. I found out that I deleted missing values in the wrong way and by redoing a previous task I was able to solve the problem.

At the moment I started with the more statistical parts of the analysis, such as learning about the p-value, variance and standard deviation, it was overwhelming and complicated at first. When I started to look at youtube videos for different examples, it became clearer and I was able to get the correct results and understand them.

