Influenza Vaccination Status in the United States of America Between 2014 to 2021

by Joshua Nguyen

Abstract Influenza, commonly known as *flu*, is an infectious upper respiratory illness or disease caused by the *influenza virus.* Luckily, the *influenza vaccine* can be used to help protect against infections caused by the influenza virus. Because, the influezna virus rapidly changes, mutations in the virus can cause new arising strains. Therefore, repeated annual influenza vaccinations ia recommended to offer consistent yearly protection against influenza and the influenza virus.

Project Overview

1. This project explores different statistical metrics to analyze and assay annual influenza vaccination status in the United States of America between 2014 to 2021. Additionally, this project will display the results that is easily read and interpreted by the non-technical population who may not easily understand statistical vernacular.
2. This project will explore the relationship between vaccination status of USA health care personnel in all 50 US states between the years mentioned above.
3. This project aims to provide conclusions that help explain the influenza vaccination trend amongst health care workers in the USA. In particular, this project can help improve influenza vaccination compliance among USA health care personnel.
4. This project utilizes **Microsoft Excel**, **MySQL** and **Tableau**.

Objectives

The overall goal of this project is to build a report that showcases summarized information about influenza vaccination status in a healthcare setting in the USA and understand how influenza vaccination status can increase among US health care personnel.

Questions of Interest

1. How has the population of health care personnel in the United States of America (USA) changed within the past five years?
2. Which USA state has the most health care personnel?
3. What is the trend of the vaccinated and non-vaccinated health care personnel in the USA during the past four to five years?
4. Which state has the second highest health care population and how does this state rank when comparing the total vaccinated health care personnel in the past four to five years?
5. Each state has different total health care personnel population - which state in 2021 had the highest percentage of vaccinated personnel compared to its health care personnel population?

Data Cleaning, Transformation and Preparation

The Dataset

This project utilizes the dataset found at https://catalog.data.gov/dataset/vaccination-coverage-among-health-care-personnel-b0240. It explores the influenza vaccination coverage among Health Care Personnel in the USA. It takes measurements from the National Healthcare Safety Network for hospital-based health care personnel.

Viewing Dataset in SQL

This project utilized MySQL to extract the data into a csv file. First, import the dataset and select applicable fields of interest.

```
SELECT Vaccine,
        Geography,
        CONCAT(20, SUBSTRING_INDEX(season, '-', -1)) AS Season_updated,
        `Personnel Type`,
        `Estimate (%)` AS Estimate,
        `Sample Size`,
FROM vaccination
        WHERE `Geography Type` = 'States'
ORDER BY 2,3;
```

Next, it would be interesting to find the total population size for each row. We can get this value by multiplying the Sample Size of vaccinated personnel by the inverse of its estimate percentage. We will create a new column to add in our table with these values.

```
ALTER TABLE vaccination
        ADD COLUMN tot FLOAT;

UPDATE vaccination
        SET tot = `Sample Size` * 100 / `Estimate (%)`;
```

Next, we will create a view that shows the total number of vaccinated and non-vaccinated health care personnel. Obtaining the total non-vaccinated population is intuitive and trivial. We take the difference between the total population we found above and the total vaccinated population.

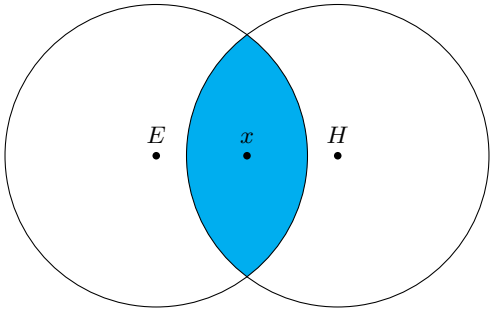
```
SELECT CONCAT(20, SUBSTRING_INDEX(season, '-', -1)) AS Season_updated,
    Geography,
    ROUND(SUM(`tot`)) AS total_vaccinated,
    ROUND(((SUM(`tot`)) * 100 / `Estimate (%)`) - (SUM(`tot`))) AS
        Not_vaccinated,
    ROUND((SUM(`tot`)) * 100 / `Estimate (%)`) AS Tot_Health_population
FROM vaccination
WHERE `Personnel Type` IN ('Employees', 'Adult Students/Trainees and Volunteers',
↪  'Licensed Independent Practitioners')
GROUP BY Season_updated, Geography
ORDER BY Geography, Season_updated;

-- Copy field names and values into a text file and save as a csv file called 'Flu
↪  Vaccination_states.csv'
```

Notice, that there are four types of personnel types in the dataset:

1. All Health Care Personnel
2. Employees
3. Adults Students/Trainees and Volunteers
4. Licensed Independent Practitioners

Additionally, noticed that we filtered out **All Health Care Personnel**. This is a necessary filter because we are looking for the total population size. Not all health care personnel are considered employees, i.e. *Adult Students/Trainees and Volunteers* and *Licensed Independent Practitioners*. Moreover, not all employees are health care personnel such as custodians, data analyst, engineers, ect. Therefore, the total population size, say Y is the sum of Employees and Health Care Personnel minus their overlap as demonstrated below:



where E = Employees, H = All Health Care Personnel and x = Health Care Personnel who are hospital employees. See that

$$H = x + \alpha + \beta \tag{1}$$

where α =Adult Students/Trainees and Volunteers and β =Licensed Independent Practitioners. Therefore,

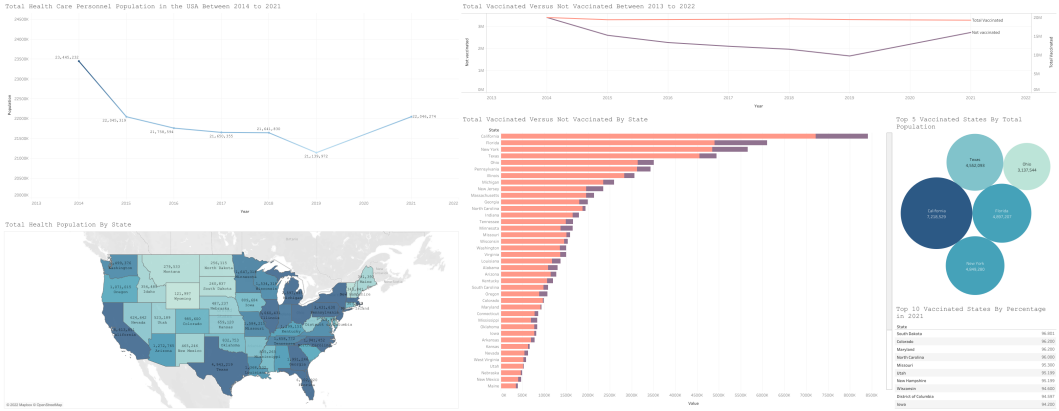
$$Y = E + H - x \tag{2}$$

$$\implies Y = E + (x + \alpha + \beta) - x \tag{3}$$

$$\implies Y = E + \alpha + \beta \tag{4}$$

Visualizations

Importing the dataset into Tableau gives us the following simple one-page dashboard that focuses on visualizing the questions of interests above. Click HERE to access the interactive dashboard.



Conclusion

Based on the analysis above, the population of the health care personnel in the USA has steadily declined between 2014 to 2019. However, the population size started to increase in 2020 and 2021; interestingly, this could have possibly been a result of the COVID-19 pandemic that started in January 2020. The pandemic and viral disease may have caused an increased number of health care personnel to obtain season 2020 and 2021's influenza vaccination.

Overall, California intuitively has the most health care personnel calculated during the past five years. In particular, California possesses 8,413,851 health care personnel during this time span. Florida follows behind California at 6,102,920 personnel followed by New York and then Texas.

Next, within the past five years, the total number of vaccinated health care personnel in the USA has remained relatively stable but did decrease slightly. In 2014, when data collection for this dataset began, the total vaccinated population was 20,035,343 personnel; and, in 2021, the population size decreased to 19,317,740 personnel. In contrast to this population, the total number of non-vaccinated health care personnel decreased within the first five years of collection but increased again during the final year. In particular, in 2014, the total population size of non-vaccinated personnel was 3,409,893

and decreased to a minimum of 1,659,698 in 2019 (maybe due to the pandemic?). It increased again in 2021 when the population of non-vaccinated personnel was 2,728,528. Overall, there was also a negative decline in the total non-vaccinated population.

Recall from above that Florida has the second highest health care personnel population. Again, Florida ranks second when comparing total vaccinated population. Florida has overall 4,897,207 which is behind California's 7,218,529 total population of vaccinated health care personnel.

Finally, in 2021, the state with the highest percentage of vaccinated health care personnel compared to its total health care personnel population is interestingly South Dakota which has a vaccinated percentage of 96.801%. This is closely followed behind by Colorado, Maryland and North Carolina. Interestingly, using the interaction tool in Tableau to filter all states in descending order, we see that California, the state with the most vaccinated health care personnel population, ranks 46 overall in percentage vaccinated in 2021. In fact, the top four states with the highest vaccinated health care personnel population, California (6th), Florida (third), New York (fifth) and Texas (20th), all place in the bottom 20 States.