Preparing for Flu Season: Interim Report

Project Overview

- **Motivation:** 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.
- **Objective:** Determine when to send staff, and how many, to each state.
- **Scope:** The agency covers all hospitals in each of the 50 states of the United States, and the project will plan for the upcoming influenza season.

Research hypothesis

If frontline staff and their families receive flu vaccines, then absenteeism rates during flu season will be lower.

Data Overview

CDC Influenza deaths by state, time, age, and gender

- This data shows flu deaths for each U.S. state categorized in ten-year age groups from 2009 to 2017.
- Deaths are categorized in 10-year age ranges. (5-14 years old, 15-24 years old, 25-34 years old, and so on up to 85 years and older.)

Population data by state

- This data shows population estimates for each state categorized in five-year age groups and sex from 2009 to 2017.
- Age Groups are 5-9 years old, 10-14 years old, 15-19 years old, up to 85+ years old.

Data Limitations

Flu data "Suppressed" values:

- Data representing fewer than ten deaths has been suppressed.
- Missing data could lead to overestimated flu deaths in states with low reported numbers.

Population estimates time lag:

- Estimated population records are from 2009 to 2017, more recent data would be more relevant.
- More recent population estimates could be helpful to see if any state populations increased or decreased significantly.

Descriptive Analysis

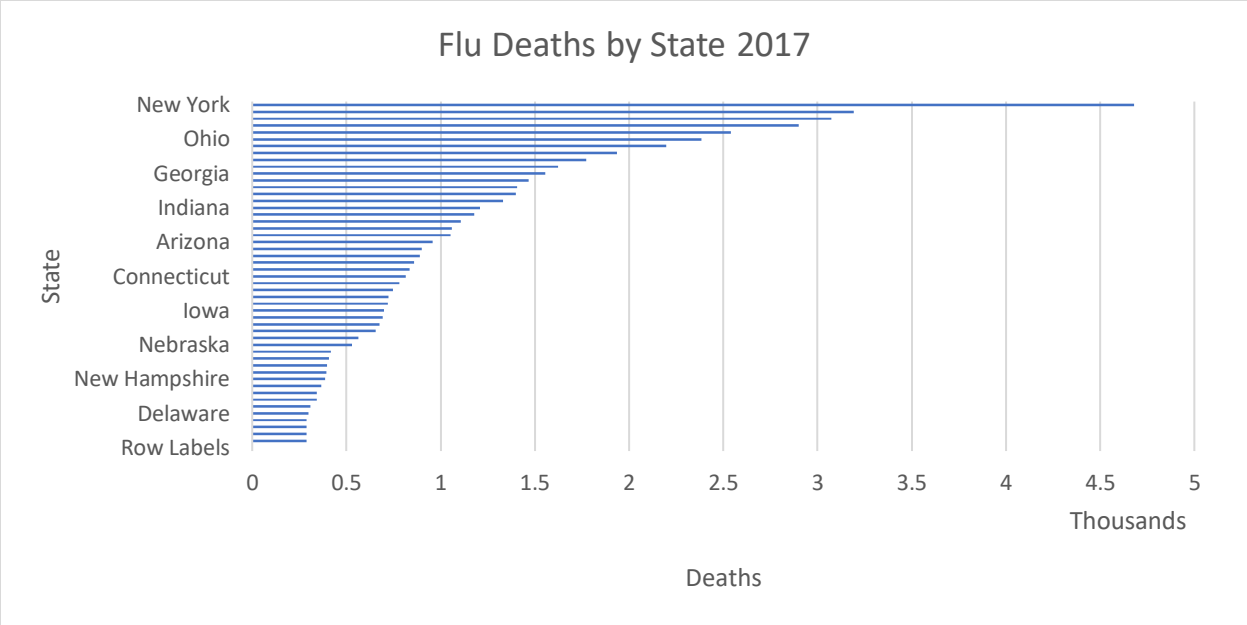
2017 U.S. Flu Death & Population Data

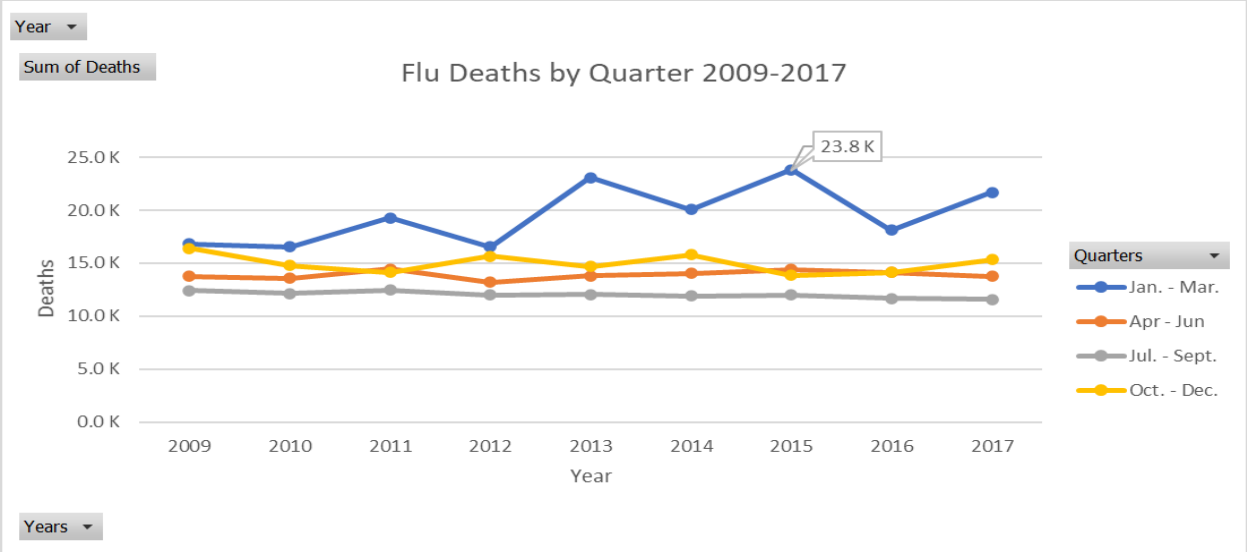
Variables	AVERAGE	STANDARD DEVIATION
Flu Deaths per State	1.2K deaths	1.2K deaths
State Population 2017	6.3M people	7.1M people

Standard deviation is the measure of how spread-out numbers are from the mean/average.

The number of flu deaths has a small correlation with population size, meaning that larger states tend to see more people die as a result of the flu.

Preparing for Flu Season: Interim Report





Results and Insights

Null hypothesis: Flu deaths in states with larger populations are equal to or lower than smaller states.

Alternate hypothesis: States with higher populations will have more deaths from the flu.

California, Texas, New York, Florida, and Illinois had the five largest populations, and the highest flu death totals every year from 2009 to 2017. Flu deaths have been higher during January-March each year.

Remaining Analysis and Next Steps

- Categorize states by low, medium and highly vulnerable population totals.
- Create spatial and temporal visualizations in final deliverable to stakeholders.
- Record video presentation for stakeholders.

Appendix

CDC dataset documentation

Business Requirements Doc

Vulnerable populations: patients likely to develop flu complications requiring additional care. These include adults over 65 years, children under 5 years, and pregnant women, as well as individuals with HIV/AIDs, cancer, heart disease, stroke, diabetes, asthma, and children with neurological disorders.

Hypothesis Development:

Clarifying questions

- Where does flu season typically begin earliest?
- What symptoms/complications require flu patients to be hospitalized?

Funneling questions

- Can some states better prepare for the flu in more ways other than by taking flu shots?
- Do certain states have higher death rates from the flu each year?
- Are there ways to avoid catching the flu from others (besides vaccines)?

Privacy & ethics

- Are there privacy laws related to collecting and storing data for flu patients?
- Do I need to consider privacy laws for data about flu patients that are minors?

Flu Data Profile

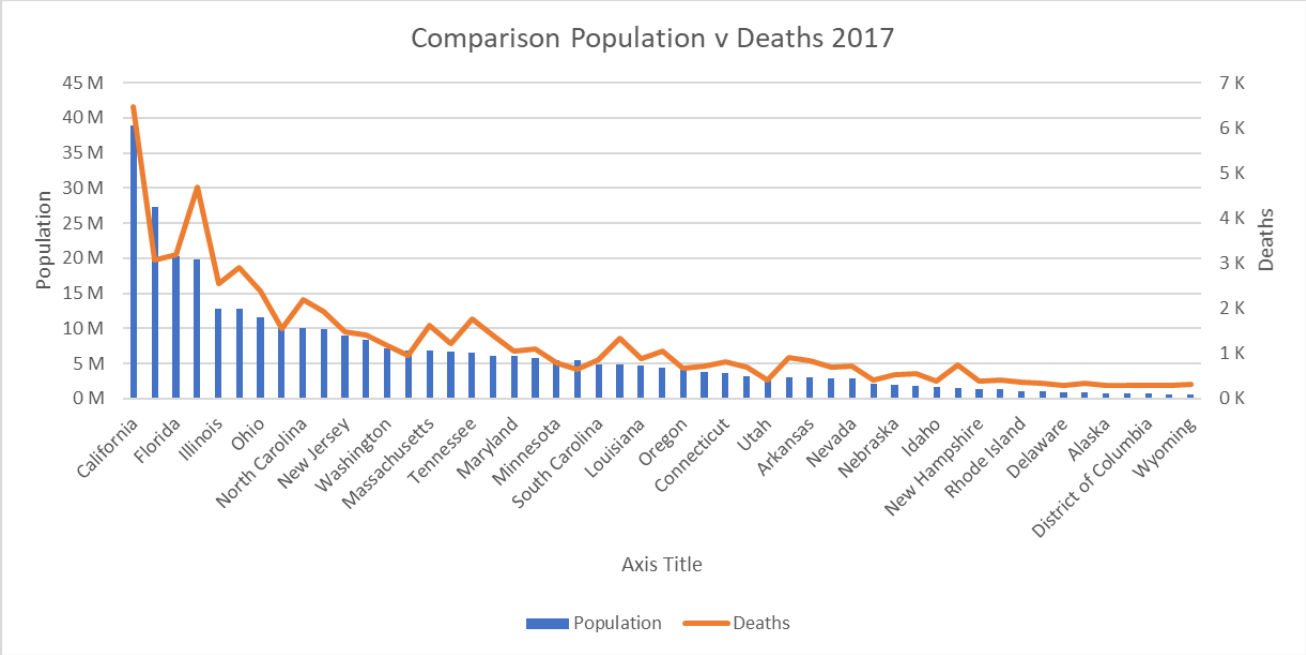
Variables	**Data Type**	**Time Component**	**Structured or Unstructured**	**Qualitative/ binary, nominal or ordinal**	**Quantitative/ discrete or continuous**
State	String	time-invariant	Structured	Qualitative/ nominal	no
State Code	numeric	time-invariant	Structured	Qualitative/ ordinal	no
Year	numeric	time variant	Structured	Qualitative/ ordinal	no
Month	string	time variant	Structured	Qualitative/ ordinal	no
Month Code	string	time variant	Structured	Qualitative/ ordinal	no
Ten-Year Age Groups	string	time variant	Structured	Qualitative/ ordinal	no
Ten-Year Age Groups Code	numeric	time variant	Structured	Qualitative/ ordinal	no
Deaths	numeric	time variant	Structured	no	Quantitative/ discrete

Preparing for Flu Season: Interim Report

Census Population Data Profile

Variables	**Data Type**	**Time Constraint**	**Structured or Unstructured**	**Qualitative / binary, nominal, or ordinal**	**Quantitative/ discrete or continuous**
State	string	time-invariant	Structured	Qualitative/ nominal	x
Year	numeric	time variant	Structured	Qualitative/ ordinal	x
Total population	numeric	time variant	Structured	x	Quantitative/ discrete
Male Total population	numeric	time variant	Structured	x	Quantitative/ discrete
Female Total population	numeric	time variant	Structured	x	Quantitative/ discrete
5-year age groups (0 to 85+)	numeric	time variant	Structured	x	Quantitative/ discrete

Preparing for Flu Season: Interim Report



Results and Insights:

Population total directly correlates with the number of flu deaths reported. It makes sense to gauge staffing needs based on estimated state populations. Younger age groups are harder to predict due to the lack of reported data pertaining to flu death totals.

2020 census results will be released by the Census Bureau later this year.

Reference publications about vulnerable population (65+ years and older):

Census Bureau - The Older Population in Rural America: 2012-2016

Census Bureau - The Population 65 Years and Older in the United States