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.

Hypothesis

If more of the vulnerable population is vaccinated in a state, then fewer people will get influenza, and fewer temporary staff are needed.

Data Overview

Population data by geography, age, and time

Census Data of the population of the U.S. for each state from 2009-2017 divided into age groups

Influenza deaths by geography, age, and time

Influenza deaths of each U.S. state from 2009-2017 divided into age groups

Data Limitations

Population data by geography, age, and time

The Census happens every ten years; I assume estimations for the population between these periods. Also, it is a manual process that is error-prone, but the Census Bureau has a lot of experience doing this kind of work, so errors should be minimal.

Influenza deaths by geography, age, and time

I assume a 1-2 year time lag before the latest data will be published. Death counts of 9 or fewer are suppressed for 54013 of 66096 data points. I imputed these data points with a study from the CDC website where the authors estimate flu death for the 2017-2018 flu season (https://www.cdc.gov/flu/about/burden/2017-2018.htm). Limitations for the imputations are that the study evaluates the whole U.S and not for each state. Second, the age groups are divided in a broader spread and sometimes different age groups altogether. Third, it's an estimation for just one year.

Descriptive Analysis:

a. Mean and standard deviation for core variables.

	Population									
	< 5 years	5-14 years	15-24 years	25-34 years	35-44 years	45-54 years	55-64 years	65-74 years	75-84 years	85+ years
Standard Deviation	452 781	907 593	966 167	953 719	923 112	939 075	765 110	479 173	280 737	121 222
Mean	382 641	779 871	831 411	798 730	786 282	837 168	706 868	440 464	254 032	107 100

	Flu-Deaths									
	< 5 years	5-14 years	15-24 years	25-34 years	35-44 years	45-54 years	55-64 years	65-74 years	75-84 years	85+ years
Standard Deviation	0	0	0	5	11	30	83	100	245	508
Mean	2,16	8	15	15	17	69	100	280	351	520

b. Correlation

The evidence that Diagram 1 gives is that the older the population, the more people die on influenza. For the age groups after 45 years, the correlation lies between 0,73 and 0,94 (1 would be a perfect positive correlation)

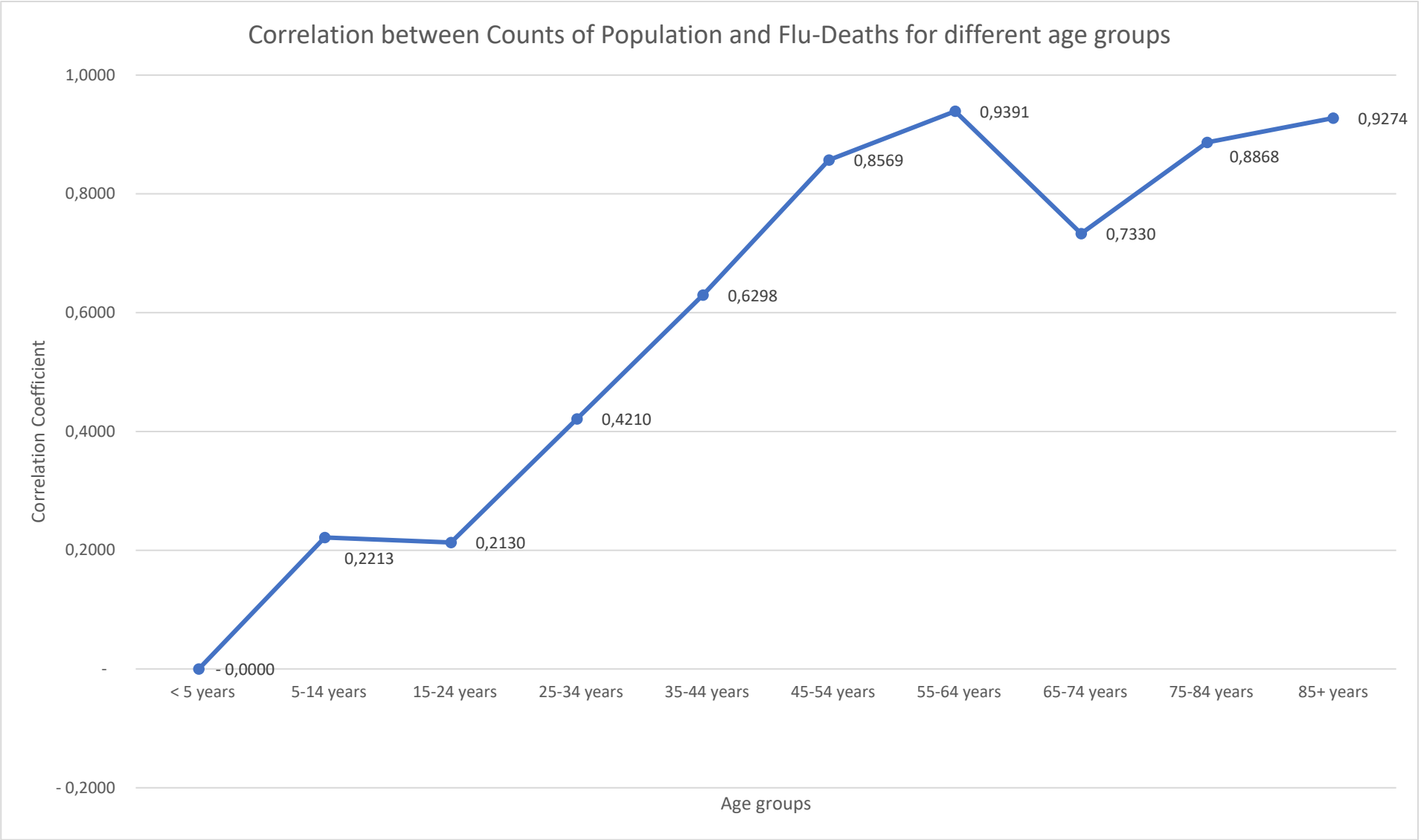

Diagram 1 Correlation Coefficients of the critical variables Population and Flu-Deaths

Results and Insights

c. Statistical hypothesis

H0: The more vulnerable people live in a state, the fewer people will die on influenza, from 2009-2017, overall age groups.
HA: The more vulnerable people live in a state, the more people will die on influenza, from 2009-2017, overall age groups.

p-value = 0,000000000 < 0,05 = α: Because the p-value is smaller than α **I reject H_0** with 95% confidence.

d. and interpretation

One needs to be aware of the limitations of the imputation from point 4. But this hypothesis gives us evidence that the more vulnerable people live in a state, the more people die with influenza.

Remaining analysis and next steps

- For answering my research hypothesis, the critical data for flu vaccination status for the whole population is missing. Therefore, I used the flu mortality rate data as a proxy variable. However, research on the CDC-Website did reveal a study (https://www.cdc.gov/flu/fluvaxview/coverage-1718estimates.htm), where flu vaccination rates are accumulated and estimated via a survey. I will try to implement and test this new data set and hopefully thus get an even deeper understanding of how influenza affects the distribution of our staff.
- I will try to limit further the shortcomings of the CDC-Website estimation data for the influenza mortality rate data.
- Prepare deeper data visualization and develop a narrative with the Tableau storyboard.
- Prepare a video presentation with my key findings

Appendix

t-Test: Two-Sample Assuming Unequal Variances		
	vulnerable Pop	*Flu-Death vulnerable Pop*
Mean	1184237,183	1152,717551
Variance	1731747380468,26	694296,7289
Observations	468	468
Hypothesized Mean Difference	0	
df	467	
t Stat	19,44895724	
P(T<=t) one-tail	0,000000000000	
t Critical one-tail	1,648123038	
P(T<=t) two-tail	0,000000000000	
t Critical two-tail	1,965056762	