Katherine Huerta
DATA 606 SP'20
Deliverable I

Exploring influenza cases in New York State with an interactive dashboard, trend comparison, and statistical analysis.

Project and Purpose.

The aim of this project is to explore seasonal influenza data to glean insights regarding trends, risk factors (such as population, vaccine coverage and effectiveness), and more. Additionally, I will determine the statistical significance of trends/correlations if applicable. Finally, I will create an interactive dashboard geared towards healthcare professionals and policy makers. This dashboard will enable healthcare professionals and researchers to make inferences and/or decisions based on their expert domain. This may include preventative measures, preparedness, and outreach.

Dataset and Data Source.

The main data set I will be exploring is *Influenza Laboratory-Confirmed Cases By County: Beginning 2009-10 Season*.[1] The data set can be acquired from HealthData.gov,[1] or from the original source, the New York State Department of Health (health.data.ny.gov).[2] The dataset contains 62,286 rows and nine columns. The nine columns are: Season (flu season ranging from October through the following May), Region of New York State (region of lab-confirmed cases, such as Central, Western, Capital District, etc.), County (county in New York State, such as Madison County, NY), CDC Week (week number in season), Week Ending Date, Disease (influenza strain - A or B), Count (number of cases), County Centroid (map coordinates), and FIPS (Federal Information Processing Standard - five digit FIPS code that uniquely identifies county and county equivalents in the United States). I plan to use other datasets to supplement my research. For example, I plan to use *State Specific Influenza Vaccination Coverage* (Centers for Disease Control, provided by Kaggle)[3] to analyze New York's vaccine coverage over time, with respect to the number of flu cases and flu strains. I may also use data to determine county populations and overall vaccine effectiveness per year (provided by CDC)[4].

Rationale.

Seasonal flu is deadly and widespread - more deadly and widespread than coronavirus.[5] So far in this 2019 - 2020 flu season, an estimated 10,000 people have died and 180,000 hospitalized in the US.[5] Additionally, seasonal flu and pandemic flu are quite similar (except pandemic flu is much more deadly), thus, seasonal flu data can help us better understand and prepare for the pandemic flu for the next pandemic.

New York State is an ideal state to analyze because it has locations that range from highly populated to rural. New York State also has high influenza-like illness activity.[5] Lastly, plentiful and up-to-date data is available for this State.

References Cited.

1. HealthData.gov. *Influenza Laboratory-Confirmed Cases By County: Beginning 2009-10 Season*. (2020). Retrieved from URL on January 15, 2020.
2. New York State Department of Health. *Influenza Activity, Surveillance, and Reports.* (2020). Retrieved from URL on January 15, 2020.
3. Centers for Disease Control and Prevention. *State Specific Influenza Vaccination Coverage.* (2019). Retrieved from URL on January 25, 2020.
4. Centers for Disease Control and Prevention. *Vaccine Effectiveness Studies.* (2019). Retrieved from URL on January 28, 2020.
5. Centers for Disease Control and Prevention. *Weekly U.S. Influenza Surveillance Report.* (2020). Retrieved from URL on February 1, 2020.

Presentation Link