Team Members

* Amber Lee
* Sai Obijiofor
* Neal Bhatia
* Jared Roussel

Background / Hypothesis

The Chickenpox (Varicella) vaccine is one of the most widely distributed vaccines to children prior to entering grade school in the US. With the rise in popularity of anti-vaccine campaigns, we are interested in investigating if an increase in vaccine exemptions is causing an increase in the number of cases (morbidity) and deaths due to chickenpox among children.

Data Sources

*Data from 2016 - 2019*

* **Vaccine Coverage and Exemptions among Kindergarteners [Data.CDC.gov]** <https://data.cdc.gov/Vaccinations/Vaccination-Coverage-and-Exemptions-among-Kinderga/ijqb-a7ye/data>
* **Nationally Notifiable Infectious Diseases & Conditions - US [Data.CDC.gov]**

<https://wonder.cdc.gov/nndss/static/2016/annual/2016-table2o.html>

Extract, Load and Transform High-Level

Extract

* Use python to read the data from the **Vaccine Coverage and Exemptions among Kindergarteners** data source (CSV format)
* Use BeautifulSoup to scrape the data from the **Nationally Notifiable Infectious Diseases & Conditions - US** data source (HTML table)

Transform

* Use pandas to merge data sources and aggregate based on state level attributes
* Drop unnecessary columns / rows from data
* Create final dataframe for loading into database

Load

* Create ERD schema for database tables
* Create database tables using either PostgreSQL (relational) or MongoDB (non-relational)

Analysis Bonus (if time permits):

*Investigative Questions*

1. Has there been an increase in chickenpox vaccine exemptions during this time?
2. Is there a relationship between the number of chickenpox vaccine exemptions and the number of cases (morbidity) and deaths due to chickenpox among children?
3. What are the types of exemptions recorded during this time? Has a specific exemption been more prevalent?
4. Which states have the highest cases and/or exemptions for chickenpox?