Project 3 Proposal – CDC Influenza Data

Group 2: Devin Fogerty, Demiko Birden, Corey Zwisler, Irfanullah Ghaffar

Data Visualization Track

We plan to use CDC data for positive Influenza tests in the U.S over the past 5 years to visualize trends in total infections, outbreaks of various strains, and the time of year that these outbreaks occur.

* When do these outbreaks occur, and is it a regular interval / how much does “flu season” vary by year?
* Which strain is the most dominant, and which have increased or decreased in prevalence over time?
* Which states have more confirmed cases relative to their total population?
* How significant of an effect did the coronavirus lockdowns have on influenza rates in 2020 and 2021?

Rough outline:

* Clean CDC data (decide what to do with missing datapoints)
  + Remove data for non-states like Puerto Rico, D.C., and the Virgin Islands
* Create schema for importing data (QuickDBD diagram)
* Import cleaned data into PostgreSQL database
* Extract data to JavaScript
* Create dashboard of visualizations with user inputs, 1 by each group member
  + Heatmap
  + Line chart
  + Stacked columns
  + Cumulative pie chart
* Upload to Git Pages

Dataset link: <https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html>