



UNIVERSITY OF  
MASSACHUSETTS  
AMHERST

College of Information and  
Computer Sciences 140  
Governors Drive  
Amherst, MA 01003-9264

CENTER FOR DATA SCIENCE

voice: 413.545.1323

fax: 413.545.1249

url: ds.cs.umass.edu

## ***Analyzing the influence of Social Determinants on COVID-19 indicators in Massachusetts***

Anushka Basu, Disha Singh, Ian Birle

### **1. Introduction**

As part of the Data Science for the Common Good(DS4CG) program at UMass Amherst, our team collaborated with the epidemiologists at the Department of Public Health (DPH) in Massachusetts to ideate and implement ways in which data-driven techniques could help their work. DPH is responsible for the health and well-being of MA residents and ensuring that people have equitable access to healthcare.

### **2. Motivation**

With the ongoing COVID-19 pandemic, DPH has been active to not only mitigate the current crisis but also come up with an effective framework for tackling future disease outbreaks like this. One of the most effective solutions to get a complete picture of 'what happened', 'what could have been done better' and 'what should be done next' is to look for patterns in the data that DPH already meticulously collects and publishes on their various COVID-related dashboards. A data science solution could bring to fore patterns that could easily be missed otherwise, for example answer to questions like:

- What are the demographic factors that influence COVID-19 factors like infection rates and vaccine uptake?
- How certain municipalities fare with respect to COVID-19 indicators compared to their peers with similar demographics?
- How are COVID-19 indicators related to other health indicators and comorbidities like heart disease hospitalization, diabetes hospitalization?

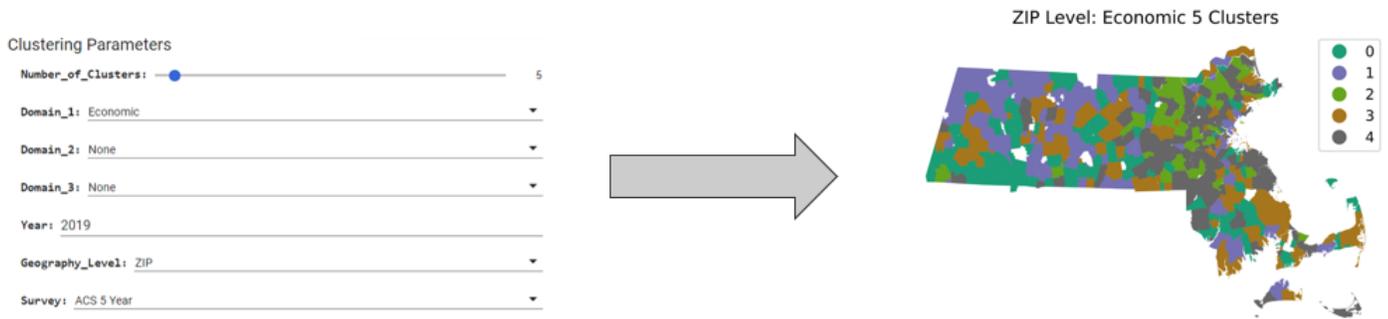
### **3. Implementation**

The problem at hand is expansive and could take various directions. Our team decided to focus specifically on establishing the relation between vaccine uptake numbers, various demographic factors like age, race, and gender, and other indicators of health. As a solution to this problem, we provided DPH with several deliverables:

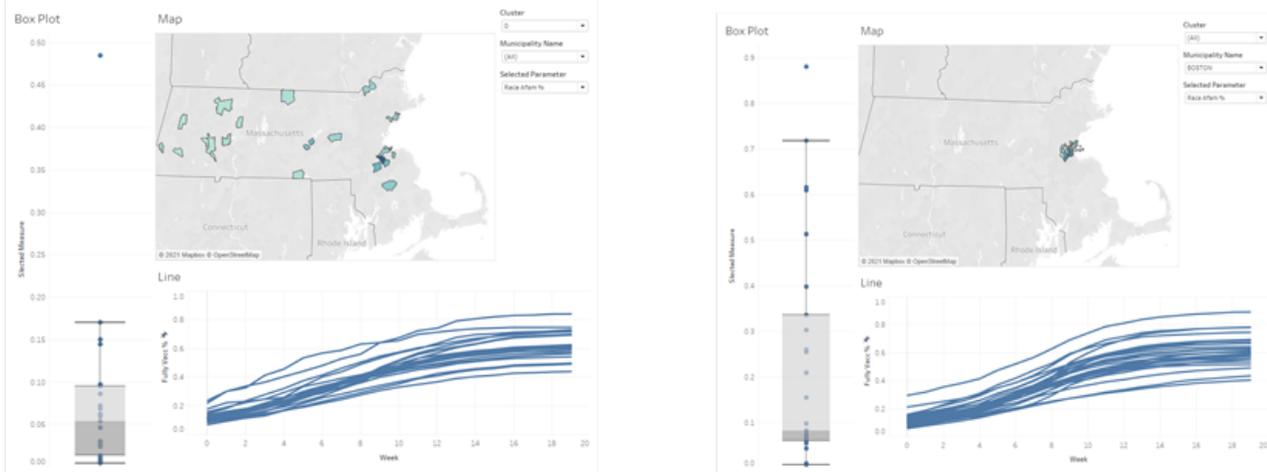
- Tools that automate collection, cleaning and merging of data from the American Community Services (ACS) repository and the COVID-19 vaccine dashboard to minimize DPH's efforts of downloading, processing, and working with data from different sources.
- Clustering tool that groups together ZIP codes in MA into a user-selected number of groups based on several demographic factors like education, employment, housing etc. This is very useful for pointing out patterns and anomalies within a cluster. DPH would take these inferences further and apply their domain knowledge to justify the patterns.

- Tableau Dashboard that allows peers within a cluster to be compared to each other with respect to different indicators.

#### 4. Results



*Clustering tool*



*Tableau Dashboard*

#### 5. Conclusion:

Our work was focussed mainly on vaccine uptake as a COVID-19 feature, this framework could integrate other features like infection rate, mortality rate, etc. DPH could use a similar data-driven approach to visualize patterns for other diseases and health outbreaks.