INFECTION *VS* FATALITY OF COVID-19 IN NEW YORK STATE: EFFECTS OF DEMOGRAPHICS AND POOR AIR QUALITY

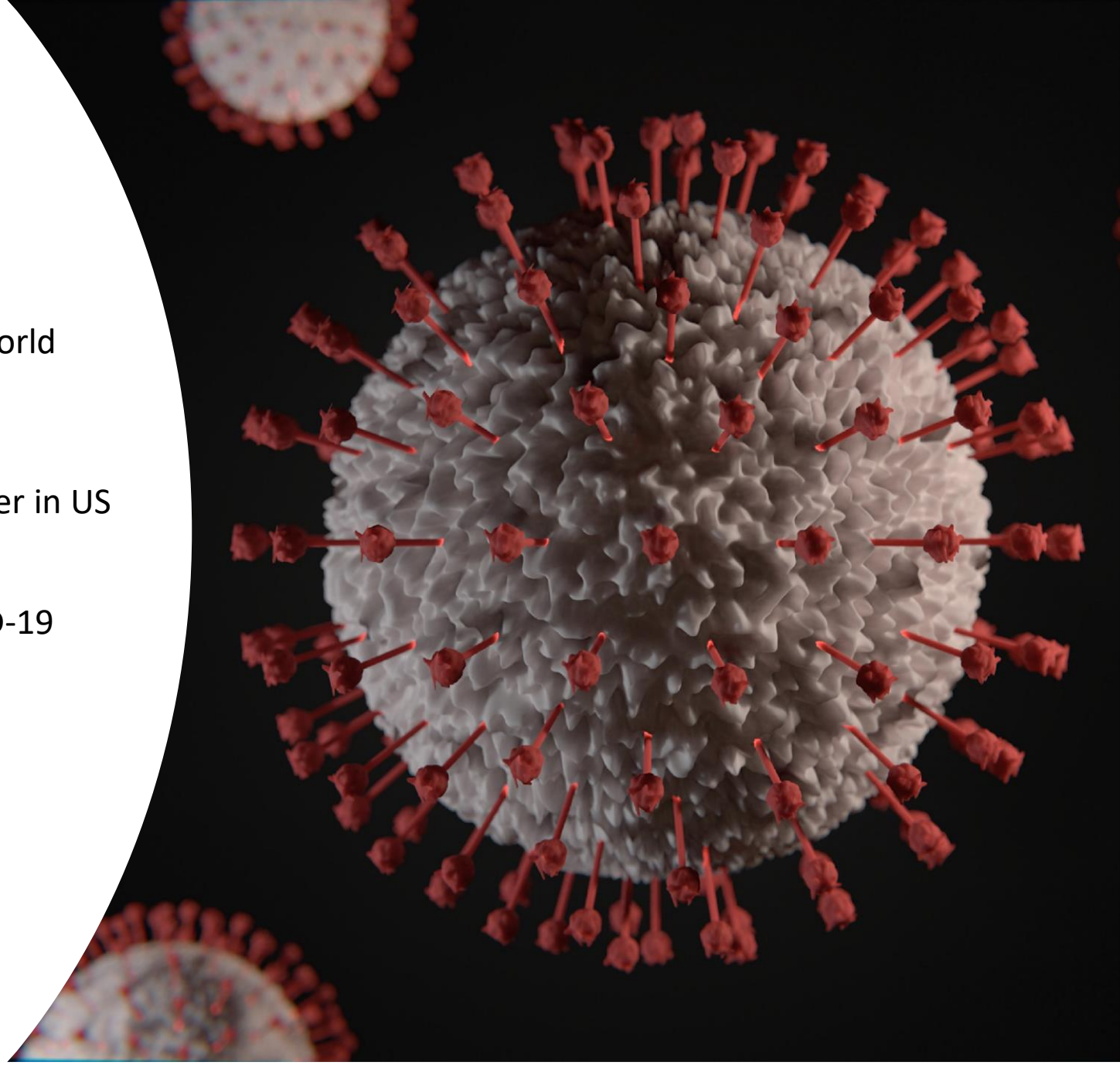
Department of Mathematics, Clarkson University
Presenter: Vijay Kumar
Date: 07/30/2020

Mentors: Dr. Sumona Mondal, Dr. Suresh Dhaniyala
Dr. Shantanu Sur, Dr. Supraja Gurajala
Collaborators: Chaya Chaipitakporn, Bridget Wangler

AIR LAB
AEROSOL INSTRUMANTATION RESEARCH

Background

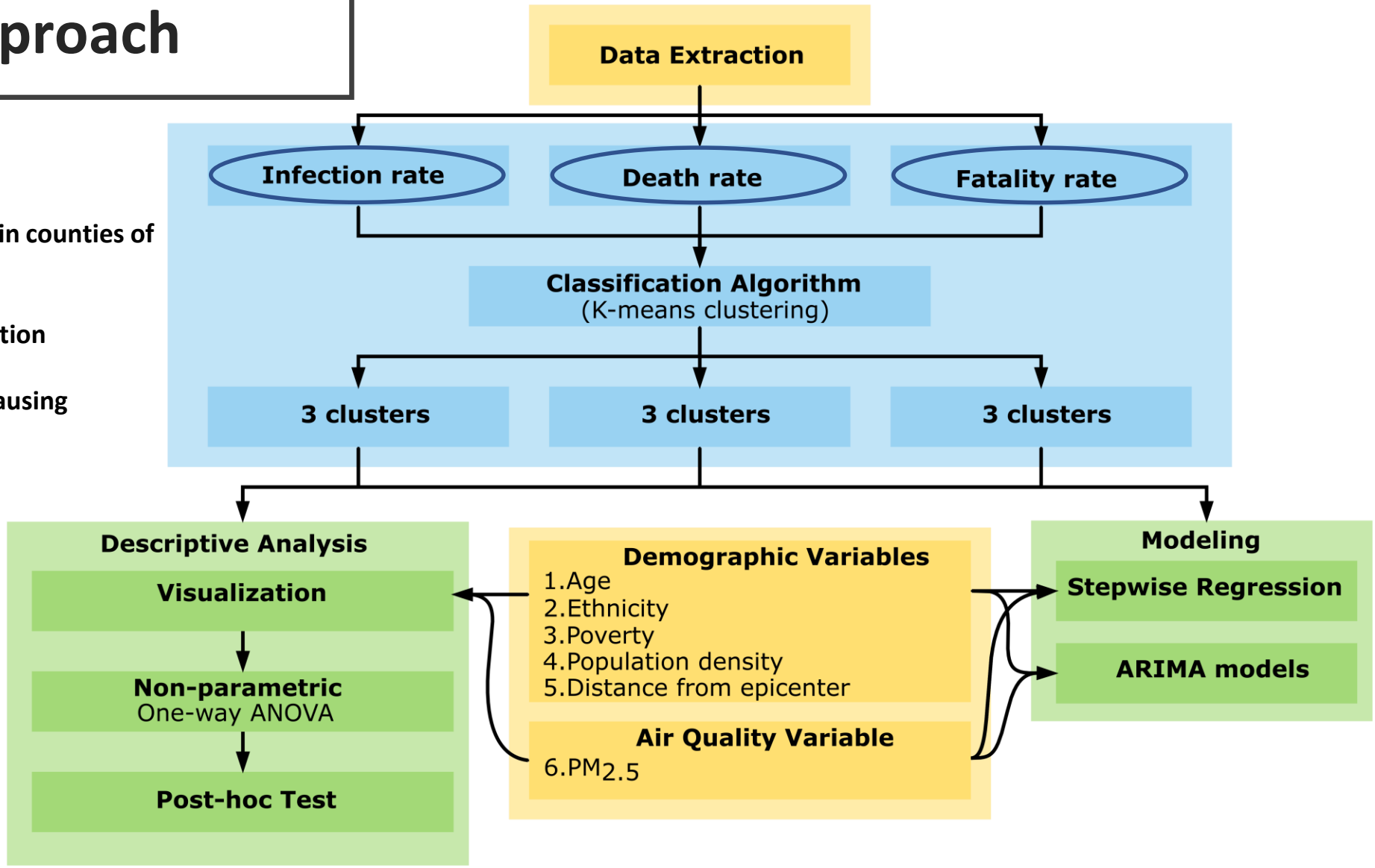
- COVID-19 is ongoing pandemic, infected the whole world
- Caused more than 610K deaths
- New York State was hit badly and become an epicenter in US (March-June,2020)
- Exposure to poor air quality leads to the risk of COVID-19



Objectives and Approach

Objectives:

- **Comparing the effect of COVID-19 within counties of New York State**
- **Estimating the different causes of variation**
- **Identifying the important risk factors causing infection, death and fatality rate**



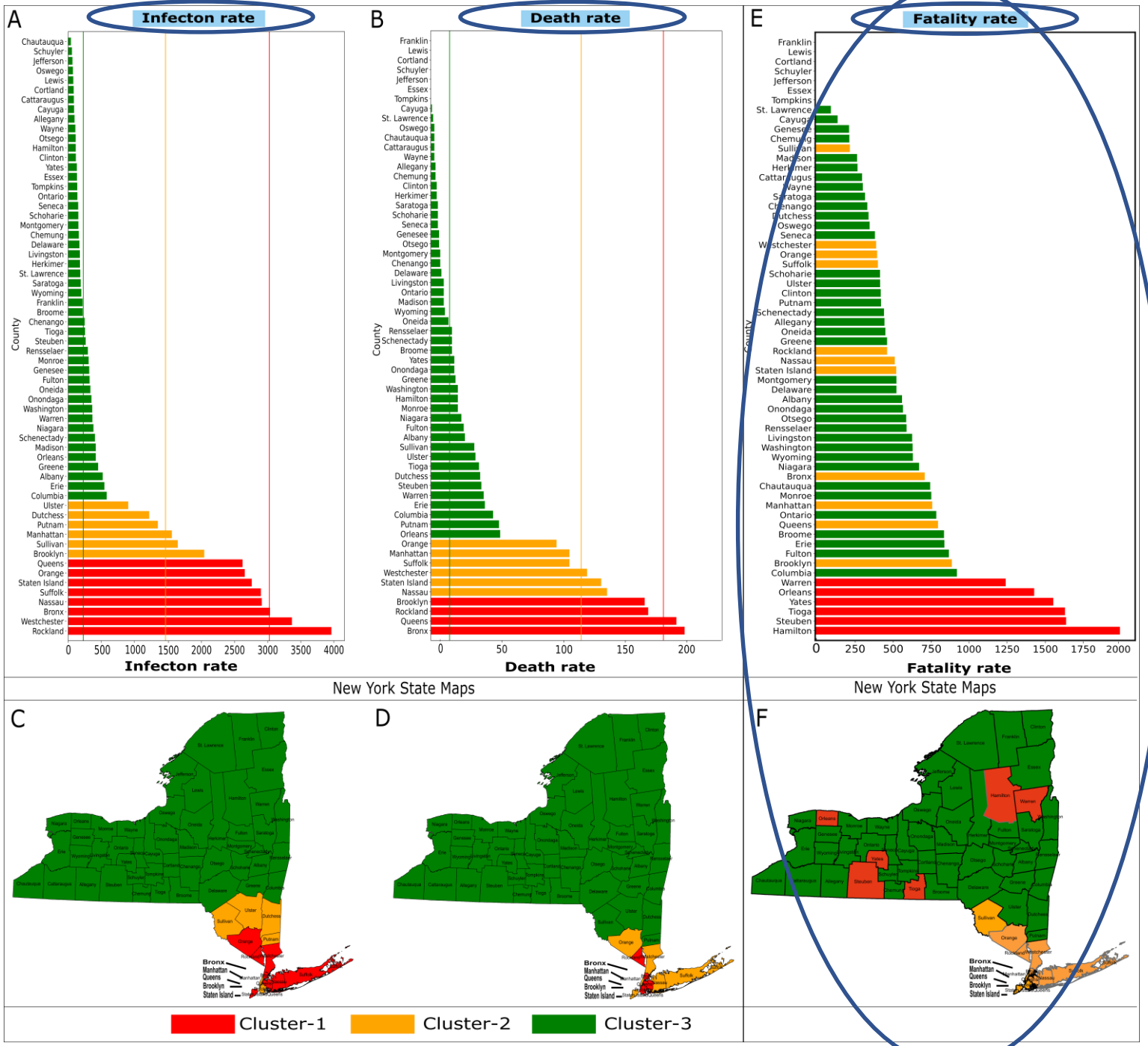
https://www.census.gov/programs-surveys/acs
https://aqs.epa.gov/aqsweb/documents/data_api.html
https://www.syracuse.com/data/

Study Workflow

Infection Group is infection per 100K population
Death Group is death per 100K population
Fatality Group is death per 10k infected population

Clustering Analysis

- Classification of the counties, K-means clustering
- Infection and death rate cluster 1 and 2 are near the epicenter
- But Fatality rate counties cluster 1 is not near epicenter
- Highly infected cluster is the downstate NYS
- Lowly infected cluster are in the upstate and northern counties of NYS



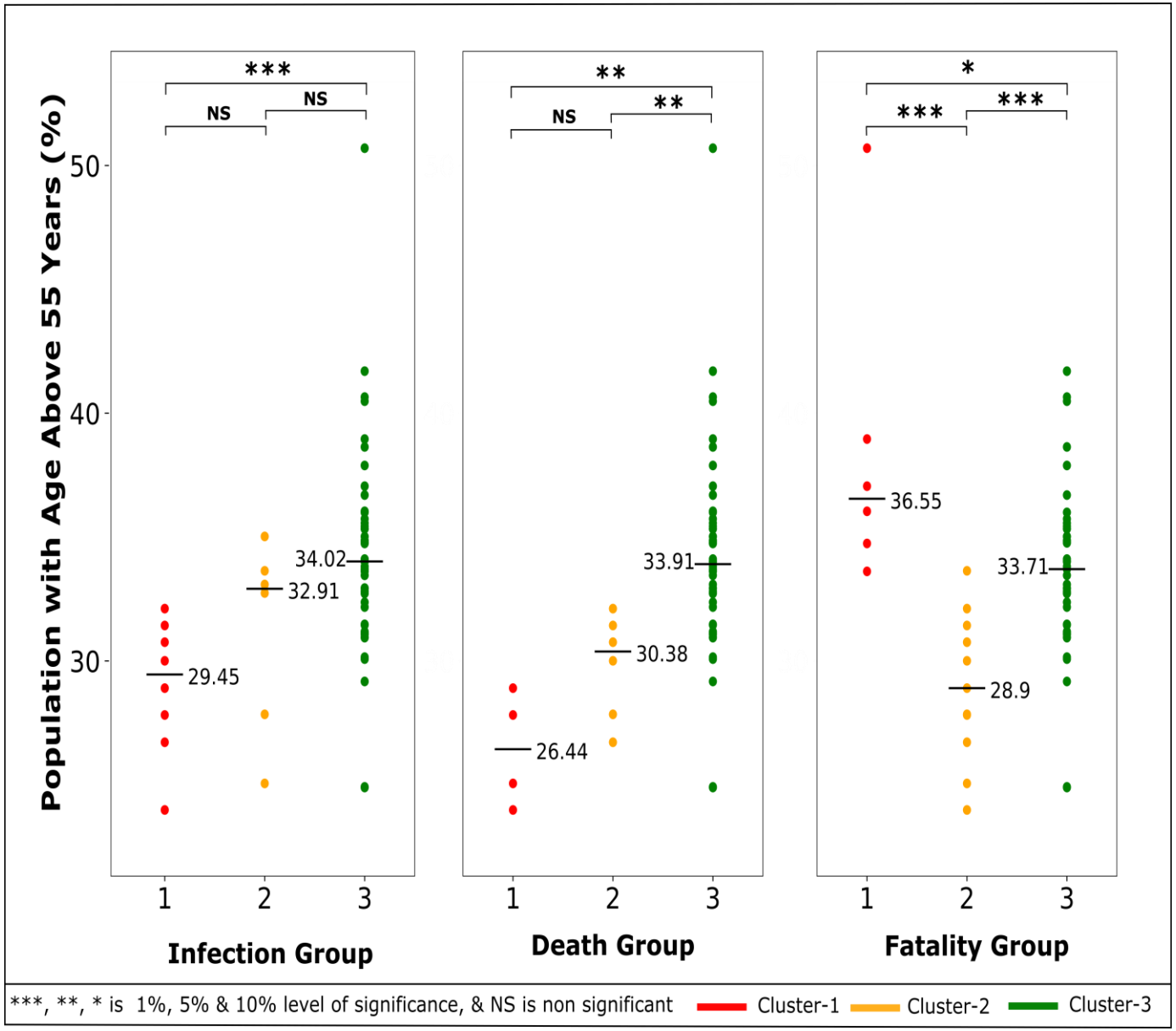
Clusters of counties with infection, death and fatality rates

Demographic factors

The demographic factors

1. Population Density (people per square miles)
2. Population with age >55 years (%)
3. Population of African American ethnicity (%)
4. Population of Hispanic American ethnicity (%)
5. Population below the poverty line(%)
6. Distance from the epicenter-Manhattan (miles)

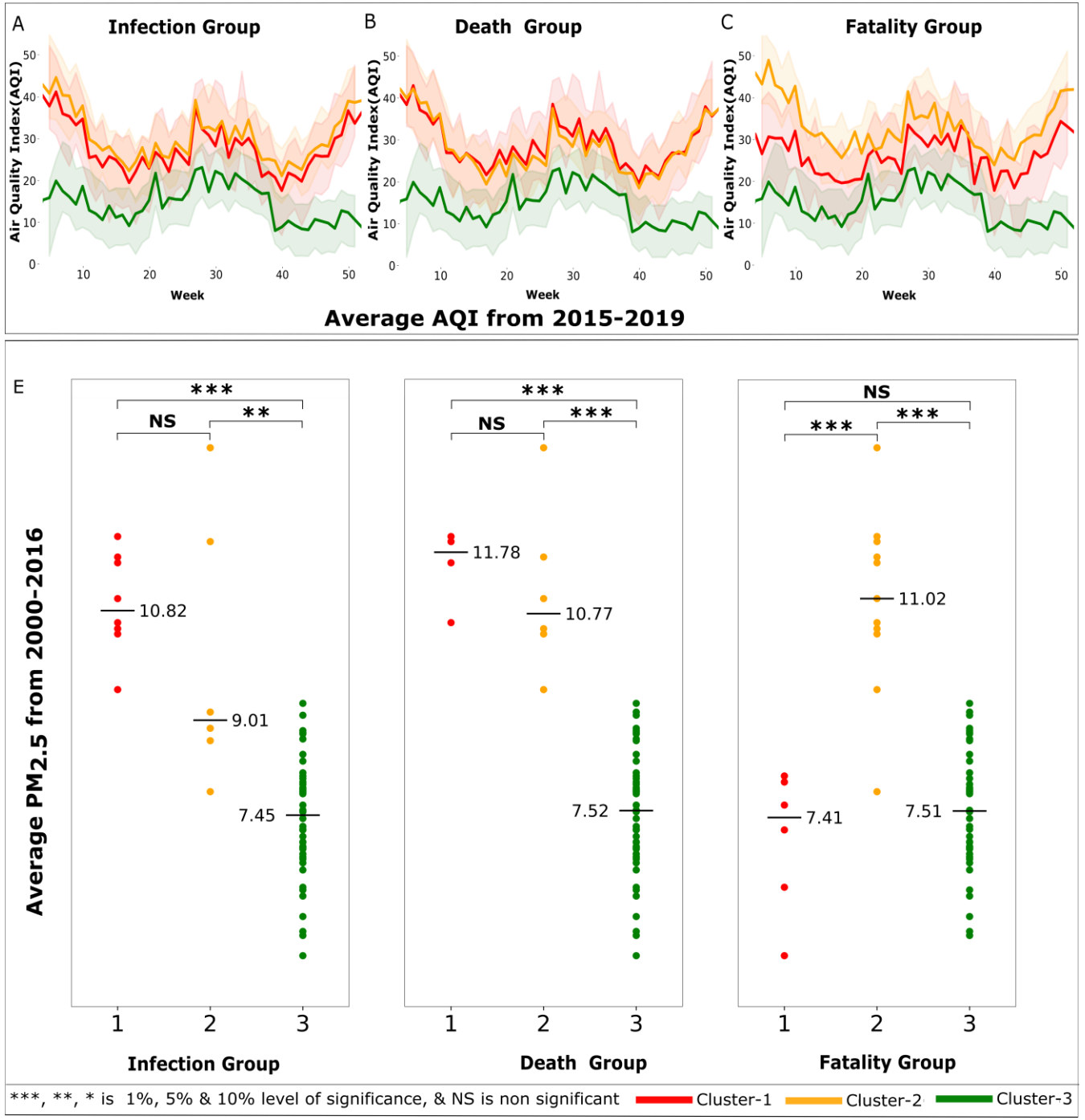
- Horizontal lines represent median with the actual values, while P-values were from non parametric Mann-Whitney-U tests after Bonferroni corrections
- Age> 55 years is taking lead in 1st cluster of Fatality group
- The Kruskal Wallis test demonstrated that all of our demographic parameters were significant at the 95% confidence level except poverty



Dot plots of Demographic variables in three groups with three clusters

ARIMA models

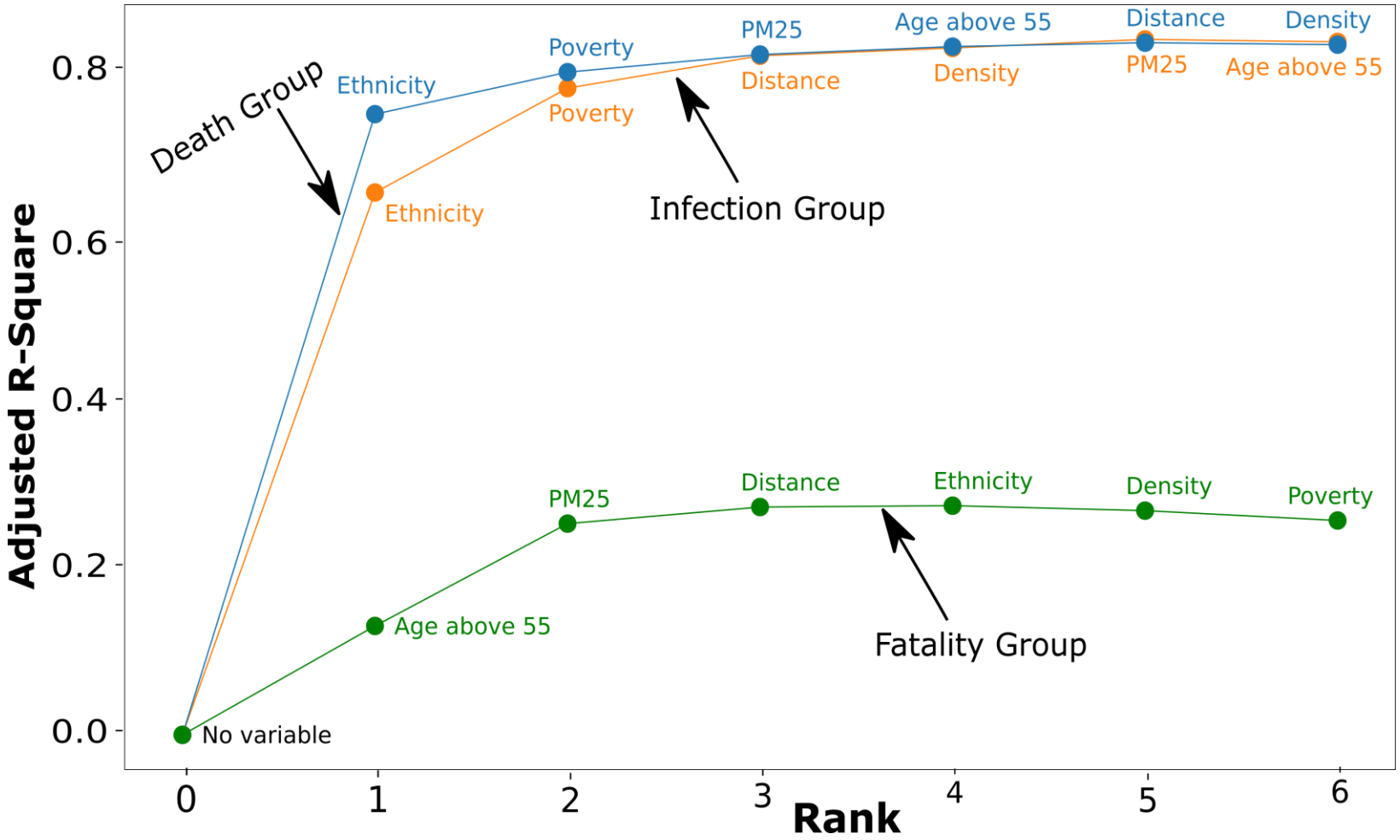
- Line plots are observed AQI values
- 95% Confidence band AQI values using Autoregressive Integrated Moving Average models
- In dot plots horizontal lines represent median with the actual values
- P-values were from non parametric Mann-Whitney U tests
- Cluster 3 is significantly different from 1 & 2 in both plots
- ARIMA plot shows cluster 3 has consistently lower AQI values



Dot plot and ARIMA time series plots for the comparisons of Air Quality variables

Stepwise Regression and significant factors

- Important factors are ranked by forward selection algorithm
- Infection and death groups have ethnicity as the most important predictor
- Fatality group was different than the other two groups
- Older age became the most prominent factor
- Bad air quality is 2nd or 3rd most important factor



Significant factors and relative contribution of different factors in three groups

Conclusion

- Infection and death rates are high in counties located near the New York City
- In Fatality, several other counties take up the topmost positions even having a low infection rate
- Regression model shows **ethnicity (African-American and Hispanic)** and **poverty** are major risk factors for infection and death rate
- Fatality has a strong association with **age and $PM_{2.5}$**
- Our results show distinct contributions by various risk factors to the COVID-19

Acknowledgment

Mentors:

- Prof. Sumona Mondal, Department of Mathematics, Clarkson University
- Prof. Suresh Dhaniyala, Department of Mechanical and Aeronautical Engineering, Clarkson University
- Prof. Supraja Gurajala, Department of Computer Science, SUNY Potsdam
- Prof. Shantanu Sur, Department of Biology, Clarkson University

Colleagues:

- Chaya Chaipitakporn , David D. Reh School of Business, Clarkson University
- Bridget Wangler, Department of Mathematics, Clarkson University
- Morgan Busch, Department of Computer Science, Clarkson University

Personal Contact

- Email: vikumar@clarkson.edu

References

[1] Guo, Y.-R.et al.The origin, transmission and clinical therapies on coronavirus disease 2019 (covid-19) outbreak–an update on the status.Mil. Med. Res.7, 1–10 (2020)

[2] Surveillances, V. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (covid-19)—china, 2020.China CDC Wkly.2, 113–122 (2020)

[3] Richardson, S.et al.Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with covid-19 in the new york city area.Jama(2020)

[4] Dudley, J. P. & Lee, N. T. Disparities in age-specific morbidity and mortality from sars-cov-2 in china and the republic of korea.Clin. Infect. Dis.(2020).

[5] Yang, Z.et al.Acute effects of air pollution on the incidence of hand, foot, and mouth disease in wuhan, china.Atmospheric Environ.225, 117358 (2020)

[6] https://towardsdatascience.com/covid-19-infection-in-italy-mathematical-models-and-predictions-7784b4d7dd8d

Thank You!