A Deeper Look Into COVID-19 Case and Death Rates in Pennsylvania

Helen Peng, Lily Fransen, and Maximus Liu

The Data

67,335 rows, 17 columns
Date range: 04/01/2020-12/31/2022

County-level & daily
- `cases` and `deaths` columns show COVID stats on that particular date

Includes cumulative measures
- `cases_cume`, `deaths_cume`, `cases_cume_rate`, and `deaths_cume_rate` help display trends over the time period
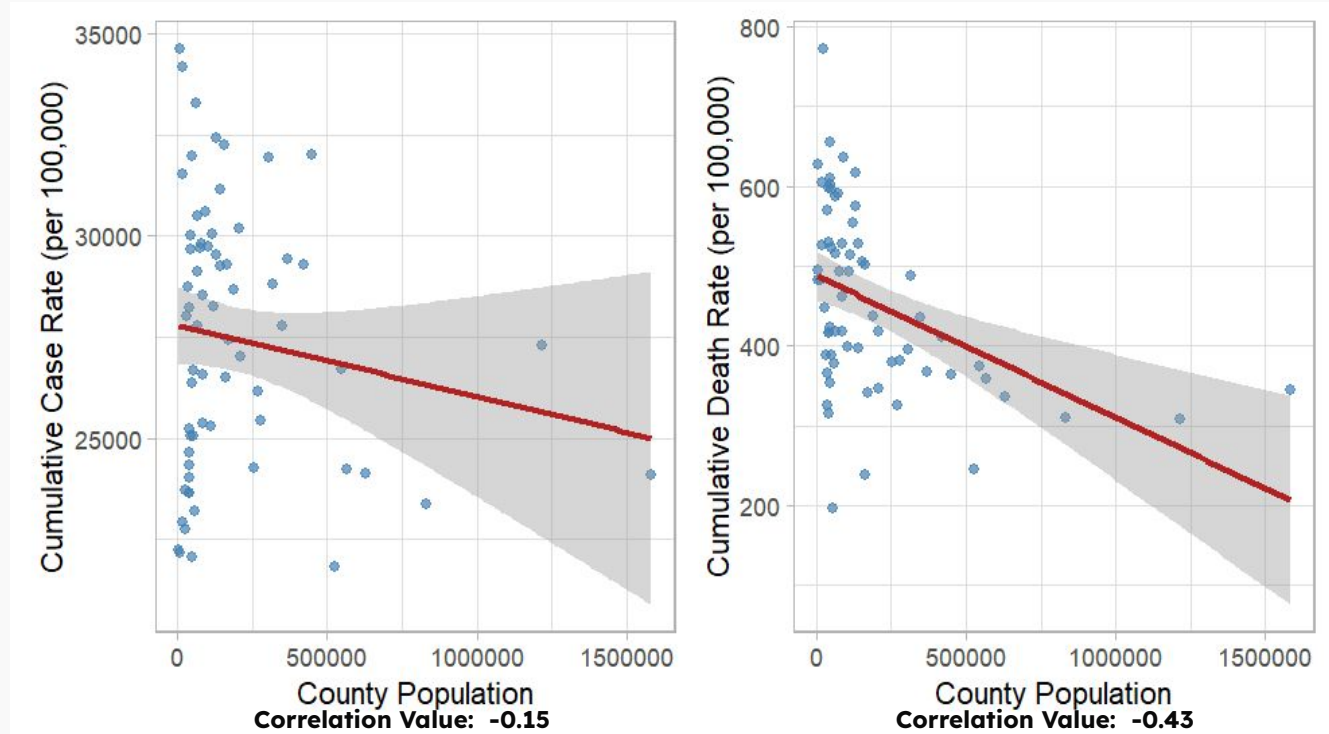
Population
- `population` column allowed us to standardize comparisons between 67 PA counties

01

Is population size associated with higher or lower per-capita COVID-19 case or death rates?

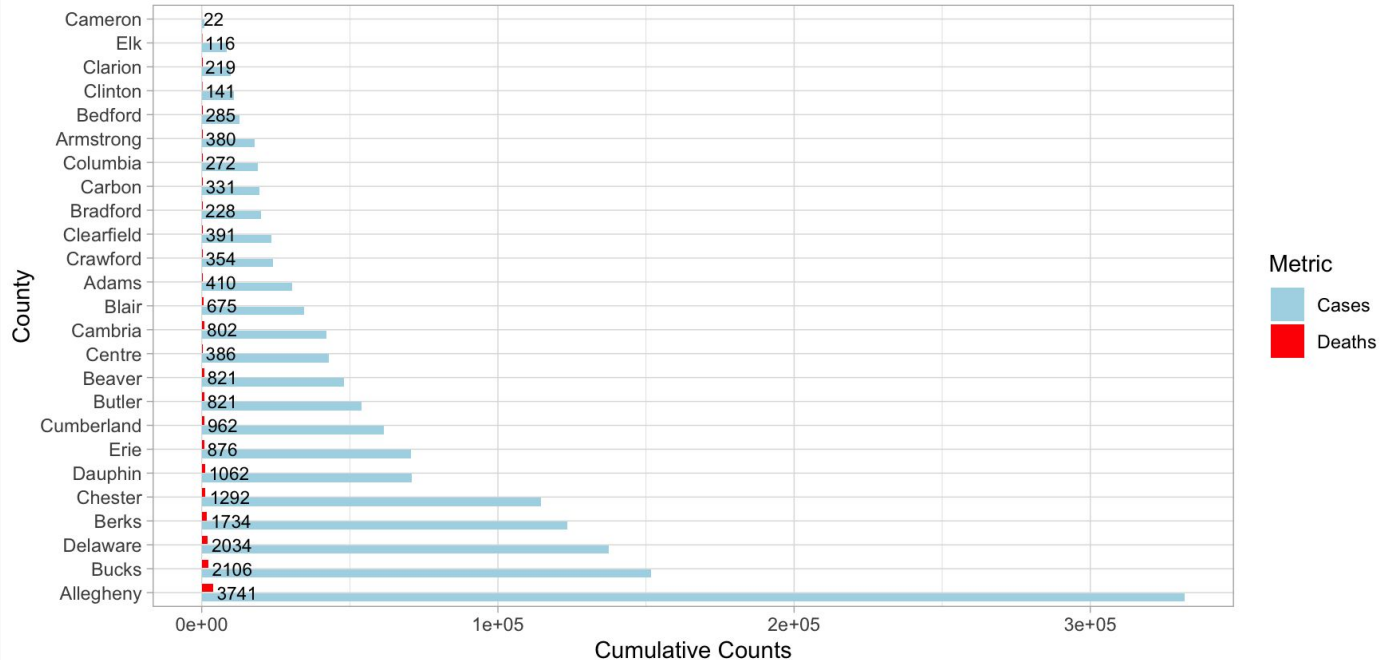# COVID-19 Case and Death Rates Decreases as Population Increases

02

Does a higher number of cases correspond to a higher number of deaths?

Higher Number of Cases Is Associated with Higher Number of Deaths
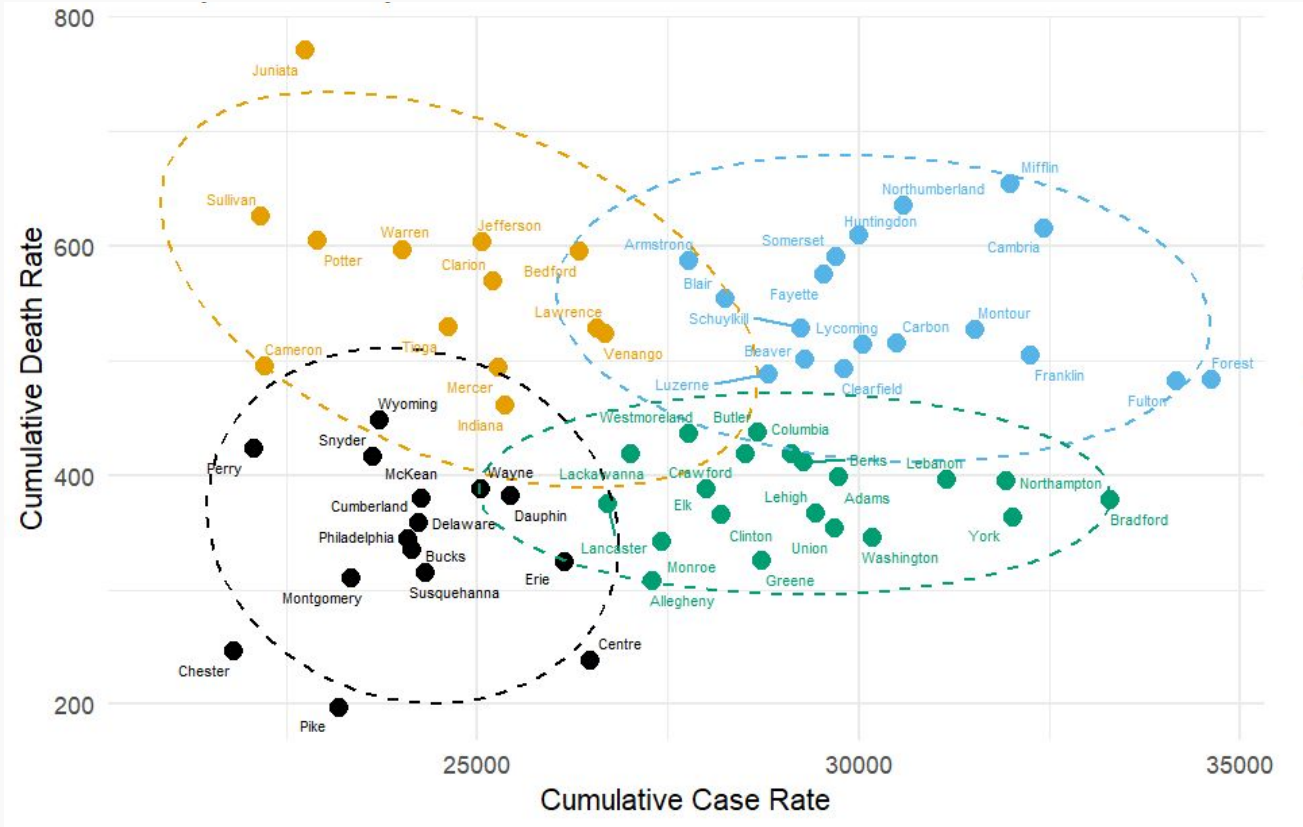
Correlation Value: 0.96

A relationship that may seem obvious in hindsight, but during the early stages of the pandemic, policymakers relied on analyses like this to assess the lethality of COVID-19 in real time.

03

What kind of patterns can clustering reveal about our dataset?

Four Distinct County Clusters Reveal Different COVID-19 Impact Patterns in PA
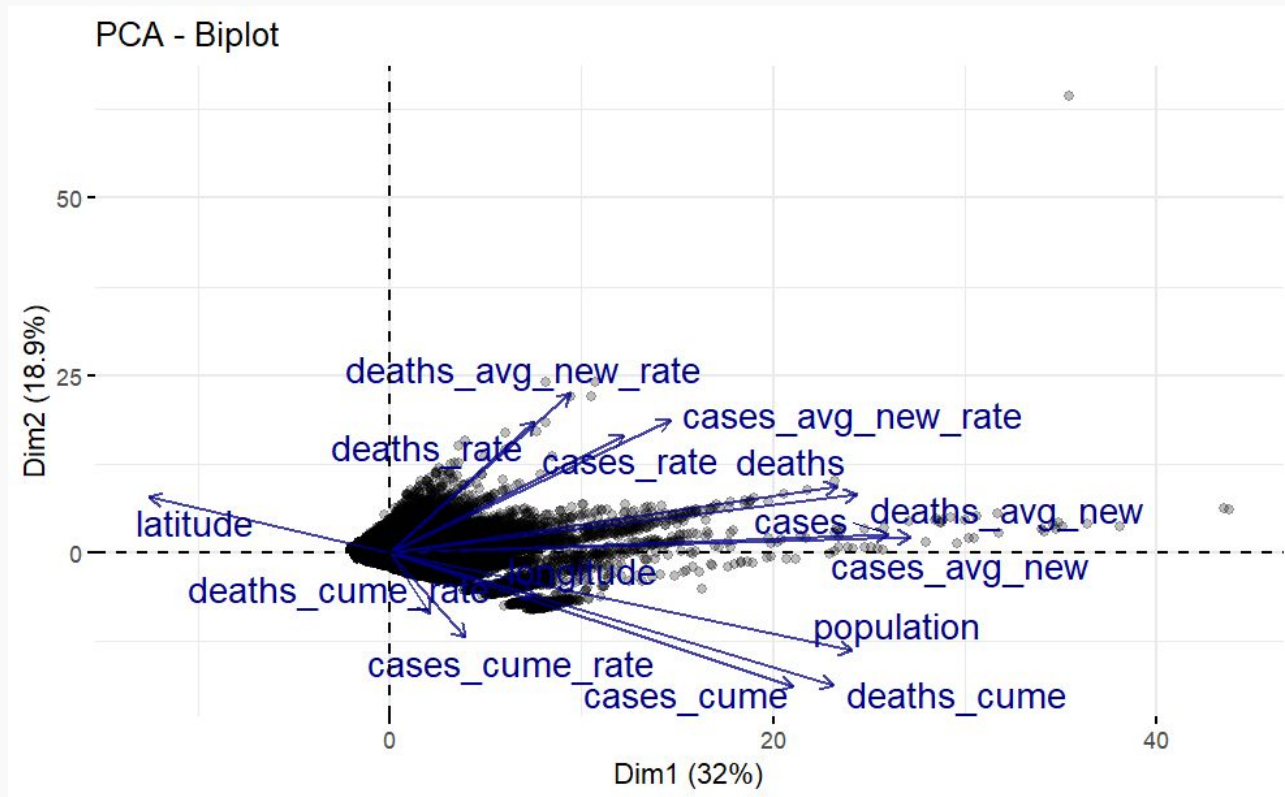
Discussion

Findings:
- Less populated counties often had higher per capita case and death rates
- Population size had a weak correlation with case rates, but a stronger correlation with death rates
- A strong positive correlation was observed between total cases and deaths

Limitations & Next Steps:
- Analysis used only cumulative data as of 12/31/2022
- Real-world decisions require time-specific trends, not just final outcomes
- Future work should apply time series analysis to capture the pandemic's evolution and better inform public health responses

Appendix

Appendix