

## Final Presentation



# A County-level Dataset for Informing the United States' Response to COVID-19

**Project Mentor:** Mathias Unberath ([unberath@jhu.edu](mailto:unberath@jhu.edu))

**Project Member:** Benjamin Killeen ([killeen@jhu.edu](mailto:killeen@jhu.edu))

**Collaborators:** Benjamin Killeen, Jie Ying Wu, Kinjal Shah, Anna Zapaishchykova, Philipp Nikutta, Aniruddha Tamhane, Shreya Chakraborty, Jinchi Wei, Tiger Gao, Mareike Thies, and Mathias Unberath

1

## Presentation Outline



- Background on COVID-19
- Dataset for informing the United States Response to COVID-19
- Unsupervised Clustering and Interactive Dashboard (Demonstration)

United States Counties Embedding

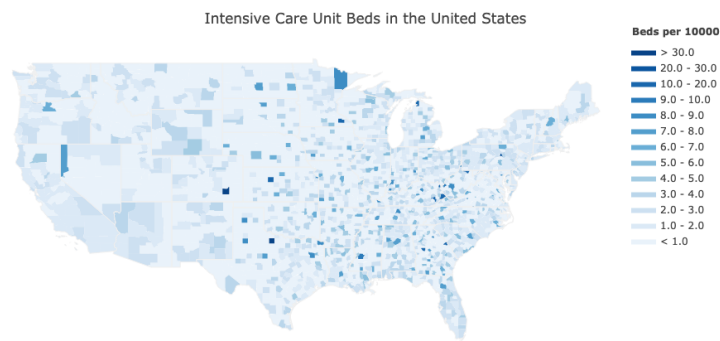


Figure (left) from B. D. Killeen et al., "A County-level Dataset for Informing the United States' Response to COVID-19," arXiv:2004.00756 [physics, q-bio], Apr. 2020, Accessed: May 05, 2020, [Online]. Available: <http://arxiv.org/abs/2004.00756>.  
Figure (right) by Benjamin Killeen.

2

## Background: COVID-19 in the United States



- **December 31:** China reports pneumonia cases in Wuhan province.
- **January 20:** First confirmed case in U.S.
- **January – March:** Containment efforts hindered by testing capability.
- **March:** Non-pharmaceutical interventions, e.g. stay-at-home orders, go into effect.

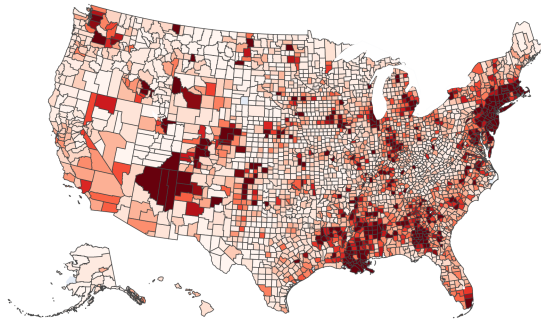
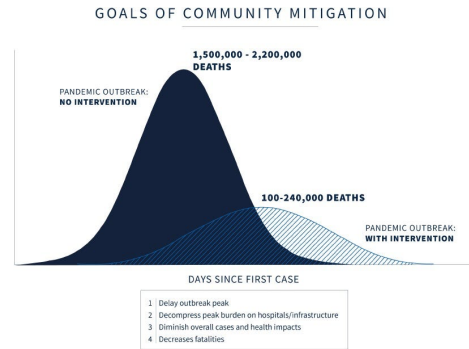


Figure (left) by Benjamin Killen. Case data from JHU CSSE COVID-19 Dashboard.  
Figure (right) from the White House.



3

## Background: Epidemiological Modeling



- **Goal:** predict future cases based on assumptions to *inform policy decisions*.
  - Local transmission rate of the disease.
  - Local availability of diagnostic tests.
- **Methods:** for example,
  - Susceptible, Recovered, and Infected (SRI) Model
- **Problem:** too many unknowns in United States.

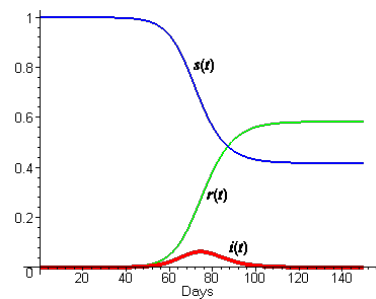


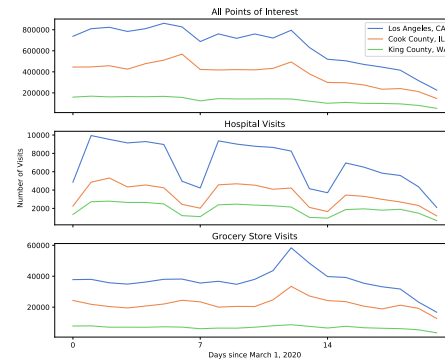
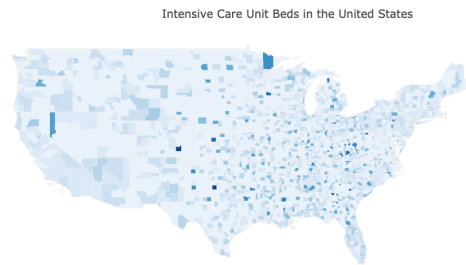
Figure (left) from "The SIR Model for Spread of Disease - The Differential Equation Model | Mathematical Association of America." <https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model> (accessed May 05, 2020).  
Figure (right) by the White House, model by Chris Murray.

4

## A County-level Dataset for Informing the United States' Response to COVID-19



- **Goal:** augment epidemiological models with **machine-readable** data on the county-level.
  - **Over 300 variables** detailing demographic, economic, and health-care capacity data.
  - **Intervention dates** for stay-at-home orders, school closures, etc.
  - **Out-of-home** activity data from SafeGraph, Google.
- Exhaustive efforts from *multiple collaborators* allowed dataset publication on **March 25**.
- **My contributions:**
  - Organized and formatted disparate data sources.
  - Created visualizations and documentation.
  - First-authored publication, published on arxiv.org, **April 1**.



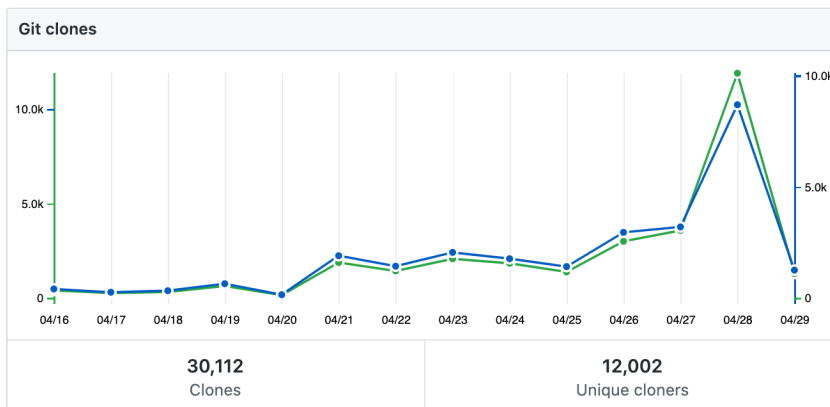
B. D. Killeen et al., "A County-level Dataset for Informing the United States' Response to COVID-19," arXiv:2004.00756 [physics, q-bio], Apr. 2020, Accessed: May 05, 2020. [Online]. Available: <http://arxiv.org/abs/2004.00756>.

5

## Dataset Significance: Visibility and Engagement



- [COVID-19 Dataset Award](#) on Kaggle, 1<sup>st</sup> place (\$1000)
- More than **500 downloads** on Kaggle
- More than **31,000 downloads** on GitHub (~12,000 unique cloners)
- **Arxiv paper:** 3 citations and *multiple ongoing communications* with researchers.

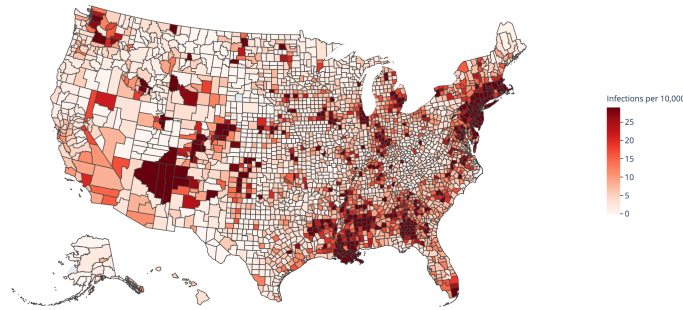


6

### Application: County-level Clustering



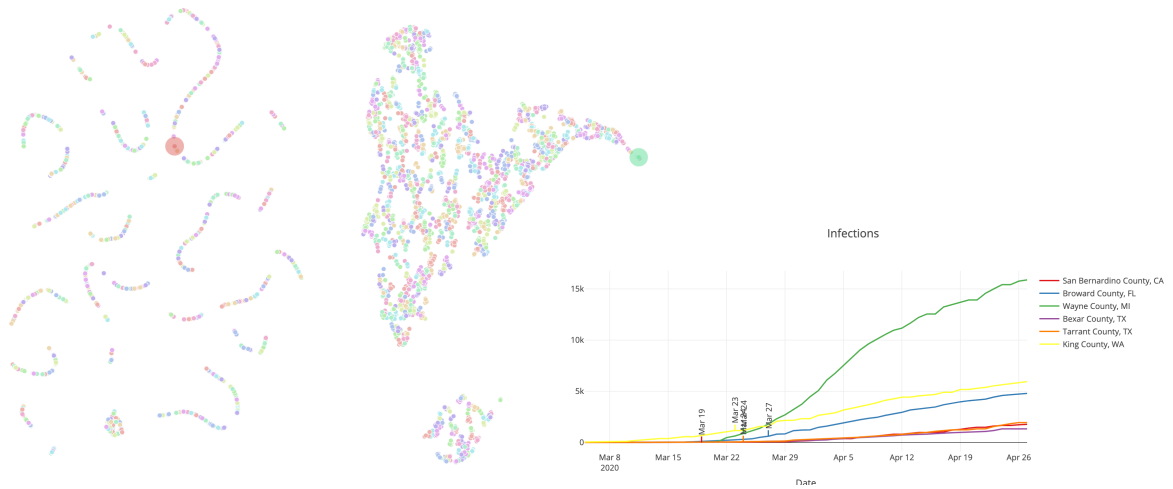
- **Goal:** inform response for counties with less spread or testing capability.
- **Method:**
  - **Clustering** based on curated variables: growth rates, age demographics, education, ICU capacity.
  - **UMap embedding (right):** reduces multi-dimensional data to preserve high-dim relationships.
- **Visualization:** interactive dashboard to verify meaningful clusters.
- **My Contribution:** all dashboard and clustering efforts, shown here, by project member.



Figures by Benjamin Killeen.

7


### Application: Dashboard Visualization of Clustering



Figures by Benjamin Killeen.

8

## Deliverables and Future Work




---

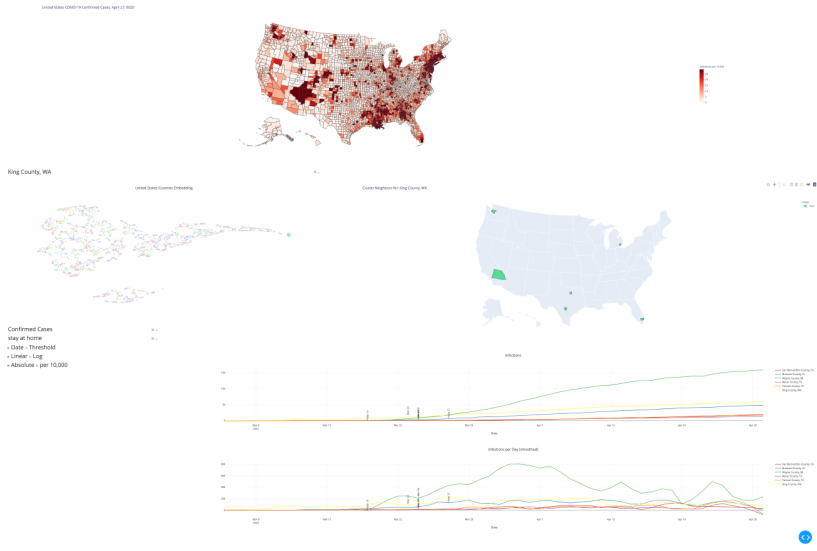
Minimum	Dataset	• <b>Structured county-level dataset</b> including COVID-19 cases, out-of-home activity, and healthcare capacity, available on GitHub and Kaggle. <span style="float: right; color: green;">✓</span>
	Implementation	• <b>Inline-documented formatting tools</b> using Python, available on GitHub. <span style="float: right; color: green;">✓</span>
	Analysis	• <b>Exponential model</b> illustrating rapid spread. <span style="float: right; color: green;">✓</span>
	Publication	• <a href="#">Medium article</a> describing the dataset in a general overview. <span style="float: right; color: green;">✓</span>
<hr/>		
Expected	Dataset	• <b>Constantly-maintained and up-to-date county-level data</b> available on GitHub and Kaggle. <span style="float: right; color: green;">✓</span>
	Implementation	• <b>Well-documented example scripts</b> for using the dataset, including visualizations of county-level time series. <span style="float: right; color: green;">✓</span>
	Analysis	• <b>Detailed Epidemiological Models</b> <span style="float: right; color: blue;">↻</span>
	Publication	• <a href="#">Arxiv Dataset Paper</a> providing a detailed description of the dataset. <span style="float: right; color: green;">✓</span>
<hr/>		
Maximum	Dataset	• <b>Constantly-maintained and up-to-date county-level data</b> available on GitHub and Kaggle. <span style="float: right; color: green;">✓</span>
	Implementation	• <b>Advanced epidemiological modeling</b> , based on county clustering. <span style="float: right; color: blue;">↻</span>
	Analysis	• <b>Advanced modeling and interactive visualizations</b> for web page. <span style="float: right; color: green;">✓</span>
	Publication	• <b>Publicly available web page highlighting results</b> and further modeling analyses publications, next two weeks. <span style="float: right; color: blue;">↻</span>

9

## Clustering Dashboard: Demonstration



---



Dashboard by Benjamin Killeen.

10