Checkpoint Presentation

LABORATORY FOR Computational Sensing + Robotics

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

Project Mentor: Mathias Unberath

Project Member: Benjamin Killeen (killeen@jhu.edu)

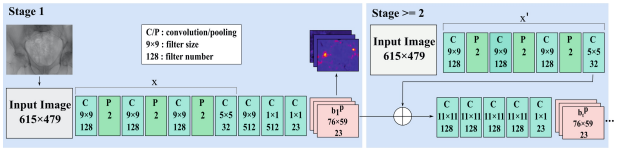
Authors: Benjamin Killeen, Jie Ying Wu, Kinjal Shah, Anna Zapaishchykova, Philipp Nikutta, Aniruddha Tamhane, Shreya Chakraborty, Jinchi Wei, Tiger Gao, Mareike Thies, and Mathias Unberath

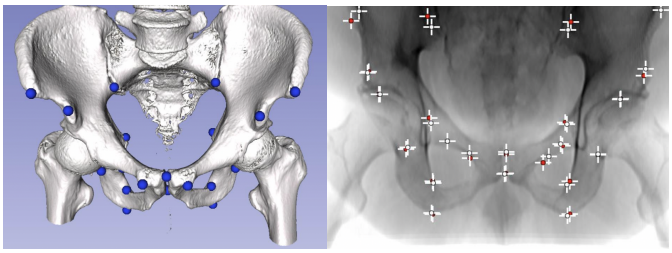
1

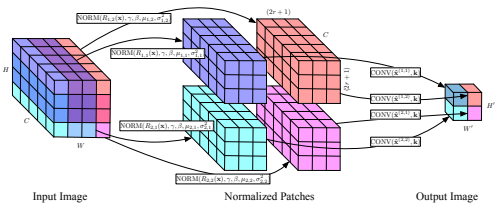
Former Project: Improved Generalization of Pelvis X-ray Landmark Detection

LABORATORY FOR Computational Sensing + Robotics

- Intraoperative registration of hip anatomy from fluoroscopic X-ray.
- Deep-learning based landmark detection.
- Improved generalization leveraging simulated data.







[1] B. Bier et al., "X-ray-transform Invariant Anatomical Landmark Detection for Pelvic Trauma Surgery," arXiv:1803.08608 [cs], Mar. 2018.

2

COVID-19 in the United States

LABORATORY FOR Computational Sensing + Robotics

- Cluster of pneumonia cases reported by China in Wuhan province, December 31, 2019.
- First infection in the United States on January 20, 2020.
- Current number of domestic infections: 378,289

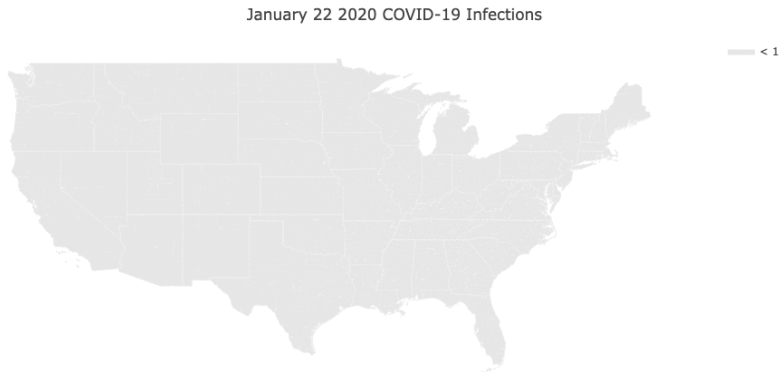


Figure original work.

3

Non-pharmaceutical Interventions (NPIs)

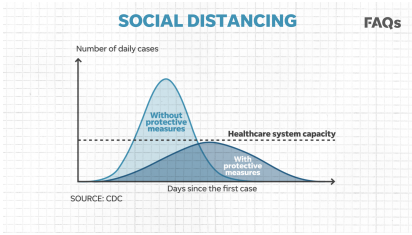
LABORATORY FOR Computational Sensing + Robotics

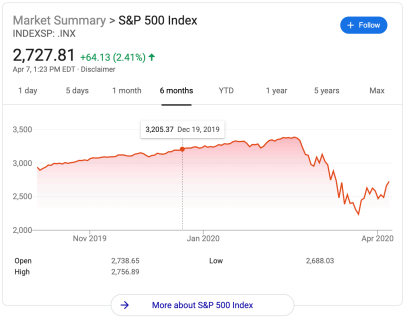
Drastic interventions are necessary to buy time to:

- Provide treatment within our healthcare system's capacity
- Develop effective testing capability
- Establish sophisticated tracing mechanisms
- Discover treatments for the virus

NPIs have adverse side-effects

- Childcare options are limited due to school and childcare closures.
- Closures of bars, restaurants, and other entertainment venues have resulted in layoffs, mainly in the service industry.
- Fears of a major economic recession constrain the job market further.





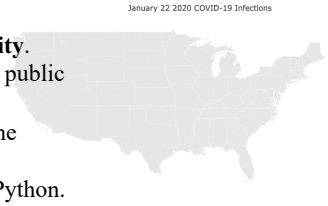
https://www.usatoday.com/story/news/health/2020/03/30/coronavirus-social-distancing-mit-researcher-lydia-bourouiba-27-feet/5091526002/
google.com

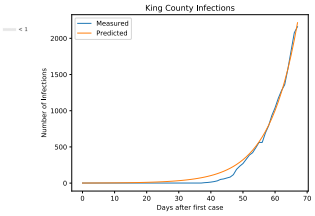
4

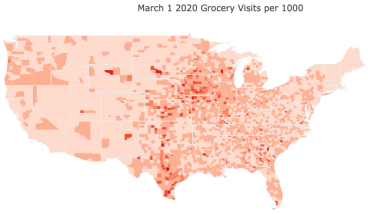
A County-level Dataset for Informing the United States Response

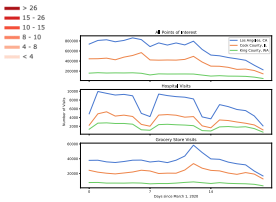
LABORATORY FOR Computational Sensing + Robotics

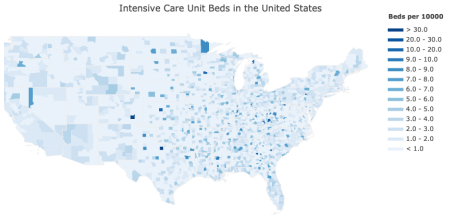
- > 300 county level variables with 90-100% availability, formatted for **machine-readability**.
  - Demographics, socioeconomic, climate, public transit, healthcare capacity.
- Time series for infections, deaths, out-of-home activity, and interventions.
- Visualization and analysis tools provided in Python.









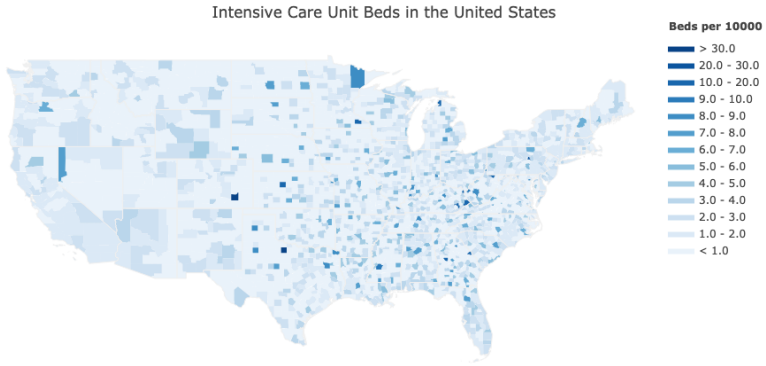


5

Estimating County-level Health Care Capacity

LABORATORY FOR Computational Sensing + Robotics

- ICU beds per-county, estimating detailed healthcare capacity.
- Per-state physicians, separated by specialty.



"Millions of older americans live in counties with no ICU beds as pandemic intensifies," Kaiser Health News, Mar. 2020.

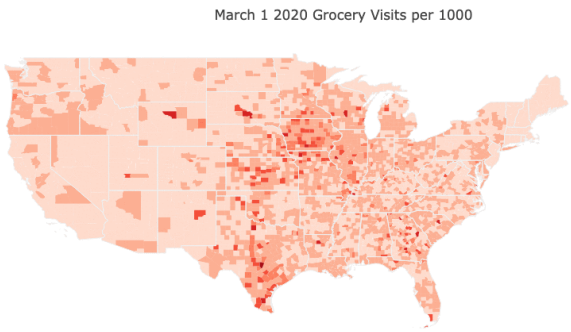
Figure original work.

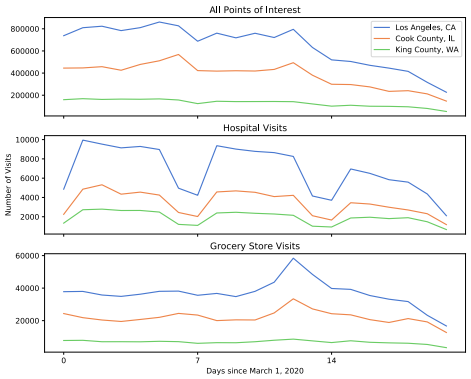
6

Out-of-home Activity in the United States

LABORATORY FOR Computational Sensing + Robotics

- SafeGraph data on out-of-home activity.
- Effects of panic buying and social distancing visible.





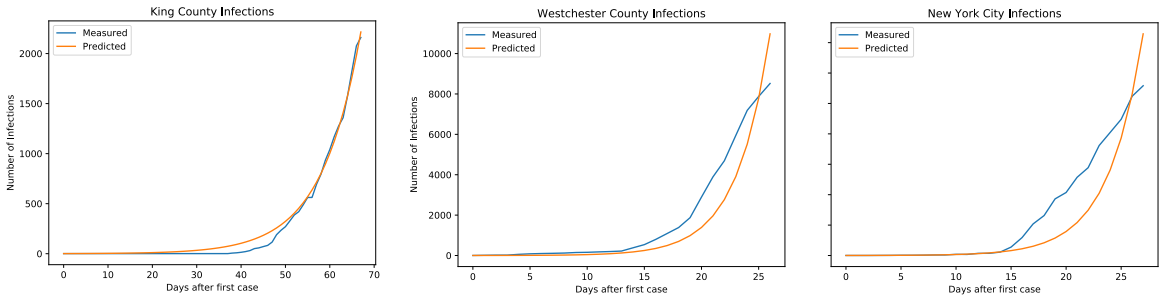
Figures original work.
SafeGraph, "Footprint data," safeGraph, a data company that aggregates anonymized location data from numerous applications in order to provide insights about physical places. To enhance privacy, SafeGraph excludes census block group information if fewer than five devices visited an establishment in a month from a given census block group.

7

Progression of Exponential Growth

LABORATORY FOR Computational Sensing + Robotics

- Exponential models illustrating the growth of the disease.
- Ongoing: advanced epidemiological models with Monte-Carlo methods and agent-based simulation.



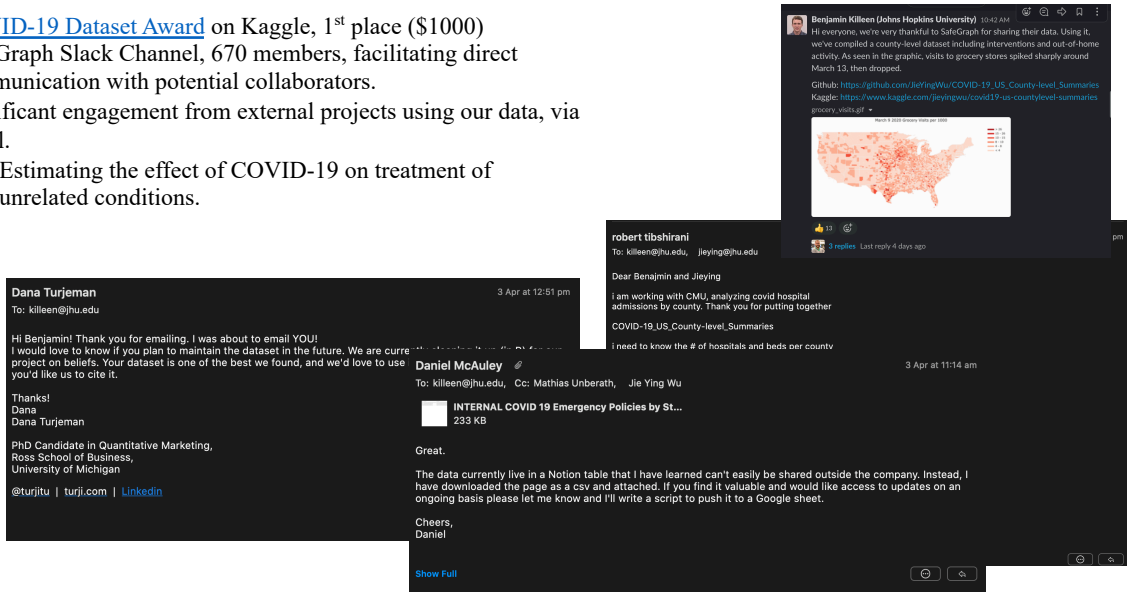
E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," The Lancet Infectious Diseases, vol. 0, no. 0, Feb. 2020.
Figure credit: Jie Ying

8

Visibility and Engagement

LABORATORY FOR Computational Sensing + Robotics

- COVID-19 Dataset Award on Kaggle, 1st place ($1000)
- SafeGraph Slack Channel, 670 members, facilitating direct communication with potential collaborators.
- Significant engagement from external projects using our data, via email.
  - Estimating the effect of COVID-19 on treatment of unrelated conditions.



9

Schedule

LABORATORY FOR Computational Sensing + Robotics

	March				April				May			
	Wk 1	Wk 2	Wk 3	Wk 4	Wk 1	Wk 2	Wk 3	Wk 4	Wk 1	Wk 2	Wk 3	
Brainstorm Contributions												
Gather Data (collaborators)												
Format + Unify Data (myself)												
Create Visualizations (myself, Anna)												
Medium Article (Mathias, myself)												
Arxiv Dataset Paper (myself, others)												
Epidemiological Modeling Medium Article (Jie Ying)												
Brainstorm Further Analysis												
Cluster Counties for better Modelling												
Analyze Roll-back Effects for Interventions as Curve Flattens												
Write Final Report												
In-class Final Presentation												

10

Deliverables

LABORATORY FOR Computational Sensing + Robotics

Minimum

- Dataset: **Structured county-level dataset** including COVID-19 cases, out-of-home activity, and healthcare capacity, available on GitHub and Kaggle.
- Implementation: **Inline-documented formatting tools** using Python, available on GitHub.
- Analysis: **Exponential model** illustrating rapid spread.
- Publication: Medium article describing the dataset in a general overview.

Expected

- Dataset: **Constantly-maintained and up-to-date county-level data** available on GitHub and Kaggle.
- Implementation: **Well-documented example scripts** for using the dataset, including visualizations of county-level time series.
- Analysis: **Detailed Epidemiological Models**
- Publication: Arxiv Dataset Paper providing a detailed description of the dataset.

Maximum

- Dataset: **Constantly-maintained and up-to-date county-level data** available on GitHub and Kaggle.
- Implementation: **Advanced epidemiological modeling**, possibly incorporating political biases.
- Analysis: **Advanced modeling and interactive visualizations** for web page.
- Publication: **Web page highlighting results** and further modeling analyses publications, TBD.

11

Dependencies

LABORATORY FOR Computational Sensing + Robotics

Dependency	Solution	Alternative	Status
Demographic, Socioeconomic Data	United States Census Bureau	X	Solved
Climate Data	National Oceanic and Atmosphere Administration	X	Solved
Economic Indicators	United States Department of Agriculture	X	Solved
Healthcare Capacity	Kaiser Family Foundation	X	Solved
Out-of-home Activity	SafeGraph	X	Solved
Public Transit Scores	Center for Neighborhood Technology	X	Solved
COVID-19 Infections COVID-19 Related Deaths Time-series	JHU CSSE COVID-19 Dashboard	New York Times COVID-19 Cases Dataset	Solved
Compute Resources	Personal Workstation or Laptop	Contact Mathias Unberath	Solved
County-level Political Indicators	Election/polling Data	Contact Mark Dredze	In progress

12

Management Plan

- Regular Meetings via Zoom:
  - Mondays 11AM, Wednesdays 4PM, and as needed.
- Communication:
  - Slack channel, one-on-one Zoom meetings.
- Data management:
  - GitHub repository
  - Kaggle dataset
  - Some data subject to licensing restrictions, esp. SafeGraph out-of-home activity.
- Software freely available on GitHub, version control enforced by project leaders.

13

My Responsibilities

- Organized and coalesced raw data into a machine-readable format.
  - Imputed missing values.
  - Coalesced and standardized time-series data
  - Standardized date-of-implementation data for interventions.
  - Coalesced and standardized disparate data from ~10 sources on the county level.
- Created visualizations for arxiv preprint and sharing with academic community.
- First-authored arxiv preprint.
- Ongoing work: clustering counties based on gathered data to allow for more accurate models.

14

Reading List

[1] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," The Lancet Infectious Diseases, vol. 0, no. 0, Feb. 2020.

[2] SafeGraph, "Footprint data," safeGraph, a data company that aggregates anonymized location data from numerous applications in order to provide insights about physical places. To enhance privacy, SafeGraph excludes census block group information if fewer than five devices visited an establishment in a month from a given census block group.

[3] "Millions of older americans live in counties with no ICU beds as pandemic intensifies," Kaiser Health News, Mar. 2020.

[4] M. Vazquez, N. Valencia, J. Acosta, and K. Liptak, "Trump says he wants the country 'opened up and just raring to go by Easter,' despite health experts' warnings," https://www.cnn.com/2020/03/24/politics/trump-easter-economycoronavirus/index.html.

[5] V. Wang and S.-L. Wee, "China to ease coronavirus lockdown on hubei 2 months after imposing it," The New York Times, Mar. 2020.

[63] The New York Times, "We're Sharing Coronavirus Case Data for Every U.S. County," The New York Times, Mar. 2020.

[64] A. Madrigal, J. Hammerbacher, E. Kissane, and COVID Tracking Project Team, "The covid tracking project." [Online]. Available: https://covidtracking.com/

[65] E. Chen, K. Lerman, and E. Ferrara, "COVID-19: The First Public Coronavirus Twitter Dataset," arXiv:2003.07372 [cs, q-bio], Mar. 2020.

[66] [Online]. Available: http://www.socialmediaforpublichealth.org/covid-19/

[67] S. Zhang, M. Diao, W. Yu, L. Pei, Z. Lin, and D. Chen, "Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A datadriven analysis," International Journal of Infectious Diseases, vol. 93, pp. 201–204, Apr. 2020.

[68] K. C. Santosh, "AI-Driven Tools for Coronavirus Outbreak: Need of Active Learning and Cross-Population Train/Test Models on Multitudinal/Multimodal Data," Journal of Medical Systems, 2020.

[69] S. J. Fong, G. Li, N. Dey, R. G. Crespo, and E. Herrera-Viedma, "Composite Monte Carlo Decision Making under High Uncertainty of Novel Coronavirus Epidemic Using Hybridized Deep Learning and Fuzzy Rule Induction," ArXiv, 2020.

[70] S. Fong, G. Li, N. Dey, R. G. Crespo, and E. Herrera-Viedma, "Finding an Accurate Early Forecasting Model from Small Dataset: A Case of 2019-nCoV Novel Coronavirus Outbreak," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 6, no. 1, p. 132, 2020.

15