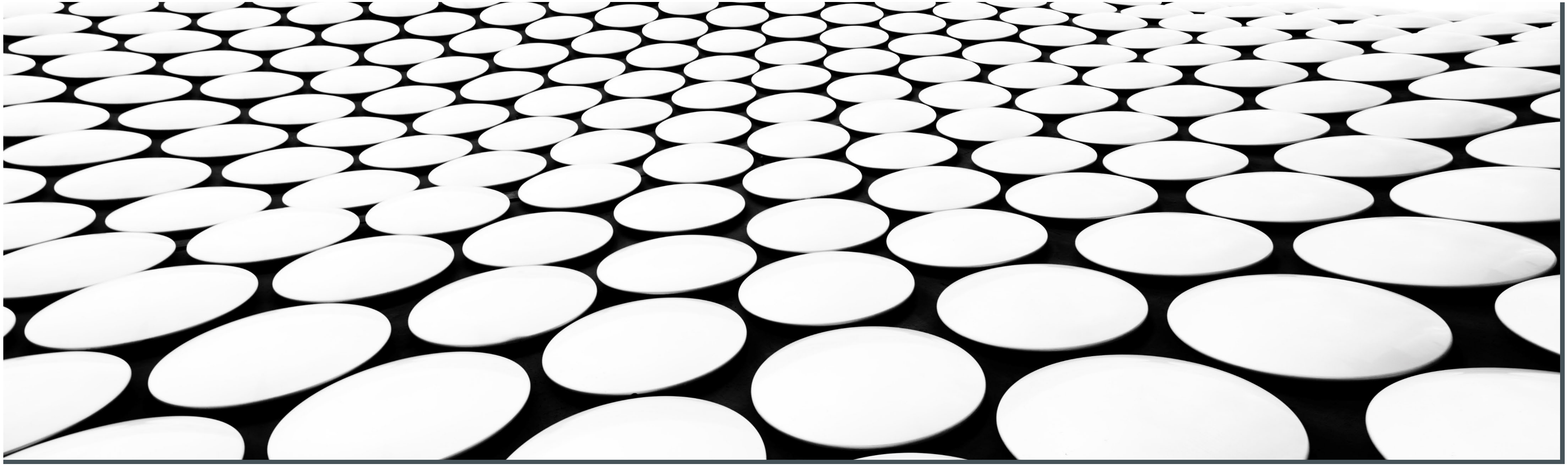
LOCAL DETERMINANTS OF MORBIDITY:

USING MULTI-SCALE SPATIAL MODELING TO EXAMINE ASSOCIATIONS IN US COUNTIES BETWEEN SOCIO-DEMOGRAPHIC INDICATORS, BUILT ENVIRONMENT CHARACTERISTICS AND COVID-19 DEATHS

DAVE COOK- UNIV. OF MD

DR. TAYLOR OSHAN (ADVISOR)- UNIV OF MD

2/26/2022



TODAY: WHAT'S THE REPORTED COST OF COVID-19 IN LIVES LOST?

946,000*

Total Covid 19 Reported Deaths 2/1/20-1/31/22

1,525,000

Actual Estimated Pandemic Deaths 2/1/20-1/31/22

~40%

Under-reported Covid 19 Deaths

*Source: New York Times and Our World in Data

UNDERSTANDING EXCESS DEATHS

- Excess Deaths are defined as the difference between the observed numbers of deaths in a specific time period and the expected numbers of deaths in the same time period
  - ⟶Historical trends identify whether the number of deaths is higher than expected.
  - ⟶Difference + Reported = Actual Death Rate

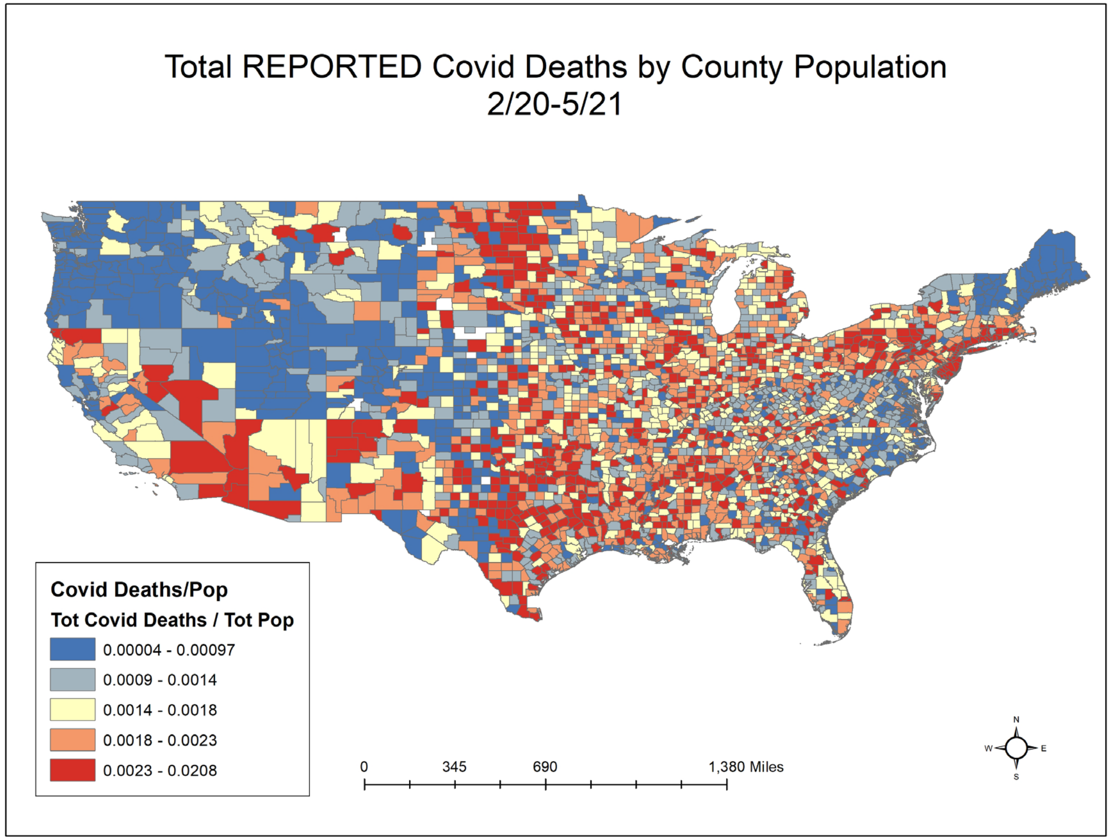
- Excess Deaths can provide information about the burden of mortality and captures some context across various geographic entities

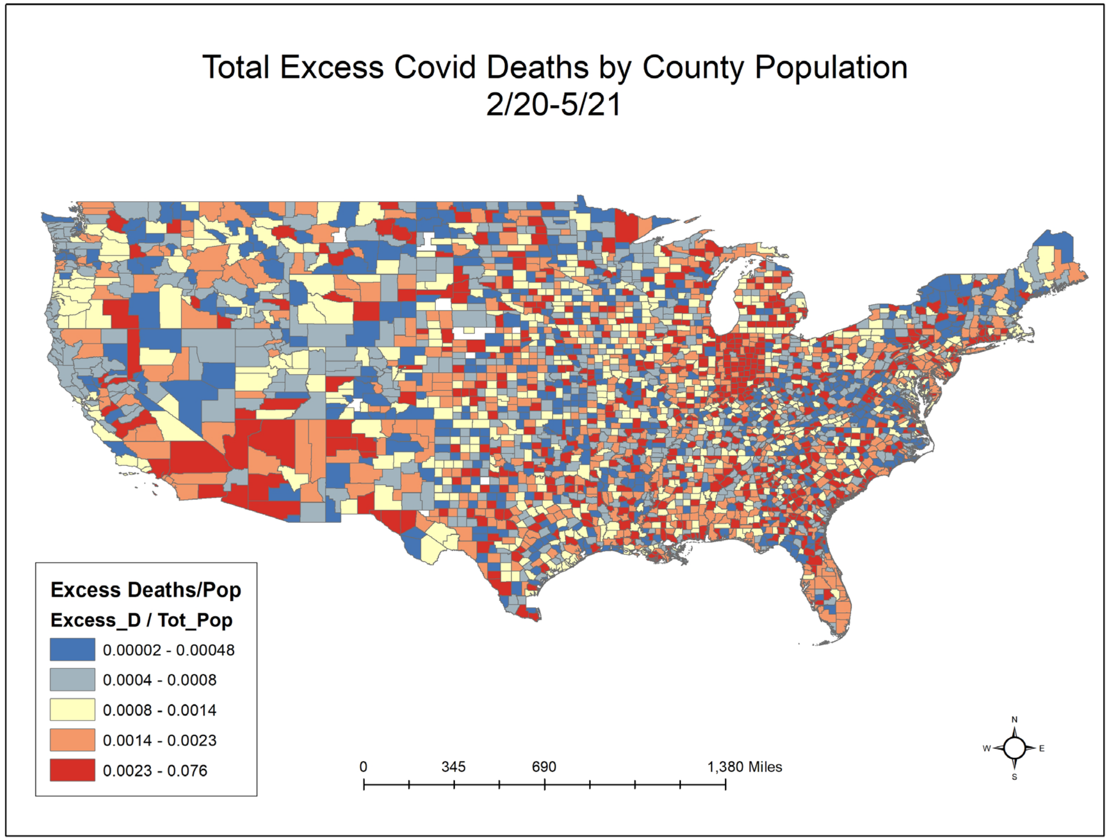
- CDC officially began publicly reporting on Excess Deaths in late 2021
  - ⟶CDC Excess Deaths Dashboard
  - ⟶Prior to that time, Excess Deaths had to be calculated using a variety of sources

COUNTING THE PROBLEM- DIFFERING VIEWS ON COVID 19 MORBIDITY

617,335	812,576	1,025,750
Total Covid 19 Recorded Deaths 2/1/20-5/1/21	Total Covid 19 Excess Deaths 2/1/20-5/1/21	Total Estimated Pandemic Deaths 2/1/20-5/1/21

CODIFYING THE PROBLEM- THE U.S. GEOGRAPHIC LENS ON COVID 19 MORBIDITY





What can the differences in recorded and estimated Covid 19 morbidity tell us?

How can we estimate and define a better picture?

PROBLEM STATEMENT

- Morbidity measured by Covid 19 counts of death certificates, not differences between expected and actual deaths.
  → **Impact:** Scope/scale is not being properly analyzed or captured geographically
  → **Project Result: Development of an all-cause mortality model and County-level dataset to measure actual versus expected deaths**

- Strong geographic disparities exist in Covid 19's impacts.
  → **Impact:** Scale/spatial context insights have not been but can be developed about the pandemic
  → **Project Result:** Combining socio-demographic indicators and built environment characteristics in models.

- Covid 19 morbidity is a local problem and must account for spatial variance/heterogeneity
  → GWR/MGWR can address the importance of spatial context.
  → **Project Result: GWR/MGWR applied to test hypotheses and separate spatial context**

County Estimates + Identify Predictors + Apply Spatial Models

DATA SOURCES

Layer	Type	Source
Weekly Counts of Deaths by Jurisdiction and Age	Tabular, contains references only to State Level	Centers for Disease Control
Provisional Covid 19 Death Counts	Tabular, contains FIPS code to link to County base data	Centers for Disease Control
Historical Death by County-2012 to 2018 (Wonder)	Tabular, contains FIPS code to link to County base data	Centers for Disease Control
ACS Socio-demographic County-Level estimates	Polygon shapefile	US Census Bureau
Google Street View Built Environment Indicators	Tabular, contains FIPs code to link to County base shapefile	Google Street View, images collected via Google API between 12/15/2017 and 5/14/2018
USDA Economic Research Service	Tabular, contains FIPs code to link to County base shapefile	USDA Economic Research Service
County Health Rankings	Tabular, contains FIPs code to link to County base shapefile	CountyHealthRankings.org
Vaccine Hesitancy Survey	Tabular, contains FIPs code to link to County base shapefile	Dept. of Health and Human Services, Assistant Secretary for Planning and Evaluation
Covid 19 County Cases (Daily Update)	County-level feature layer available through ArcGIS Online	Johns Hopkins University Centers for Civic Impact
US Counties base file	Polygon shapefile	US Census Bureau

ESTIMATING COUNTY LEVEL DEATHS

- Creating an estimate of county level Covid 19 excess deaths for 1/2020 through 5/2021 required a four-step process. The steps included:
  - (1) using county-level historical data downloaded from the CDC *Wonder* online database from 2012-2018 to provide total expected morbidity estimates for 2019 and 2020;
  - (2) forecasting expected deaths using the downloaded historical data applying the R ARIMA method, a commonly used technique for fitting and estimating from time series data;*
  - (3) selecting the monthly forecast values at the upper bound (95%) of the ARIMA forecast as the monthly 2/2020 through 5/2021 expected numbers; and,
  - (4) comparing the 2019 estimated monthly values to known official CDC monthly morbidity values to validate the overall accuracy of the ARIMA model forecast

*Credit: Kimberly Panozzo and Ishfaq Rahman
University of Toldeo, GISAG

VALIDATING THE ACCURACY OF THE MODEL

- The accuracy of the ARIMA model was validated against the 2019 forecast using a county-level seasonality value.
- A seasonality value is designed to represent the normal, historical movement in county level morbidity across months from 2012-2018, then applied to 2019
- The average seasonality value for all years (2012-2018) was used to estimate the expected number of deaths per week per county.
- The assumption is that prior to Covid in 2020, morbidity rates were relatively stable. The county-focused value is calculated by using the monthly morbidity levels per county over the number of all deaths per year in each county to determine the average number of deaths per month by county.
- The 2019 forecast and seasonality values should show similar patterns across months.

PROPORTIONALITY VALUE

- To capture observed deaths to compare to expected deaths, county level counts are provided via the CDC weekly MMWR (Morbidity and Mortality Weekly Report).
- Two official data collection and publication practices hamper accurate observed Covid-19 numbers for comparison to expected morbidity.
  - CDC dataset currently omits reporting on all counties with less than 10 reported Covid-19 deaths.
  - CDC data also omits deaths which occur in the first 4 weeks of the year.
- The average proportion of deaths that normally fall within each county per month can be substituted for the missing values since county level morbidity rates are a proportion of state level morbidity, by using historical data to identify the proportionality of state deaths per county.
- To address the reporting lag, no data beyond May 2021 was used

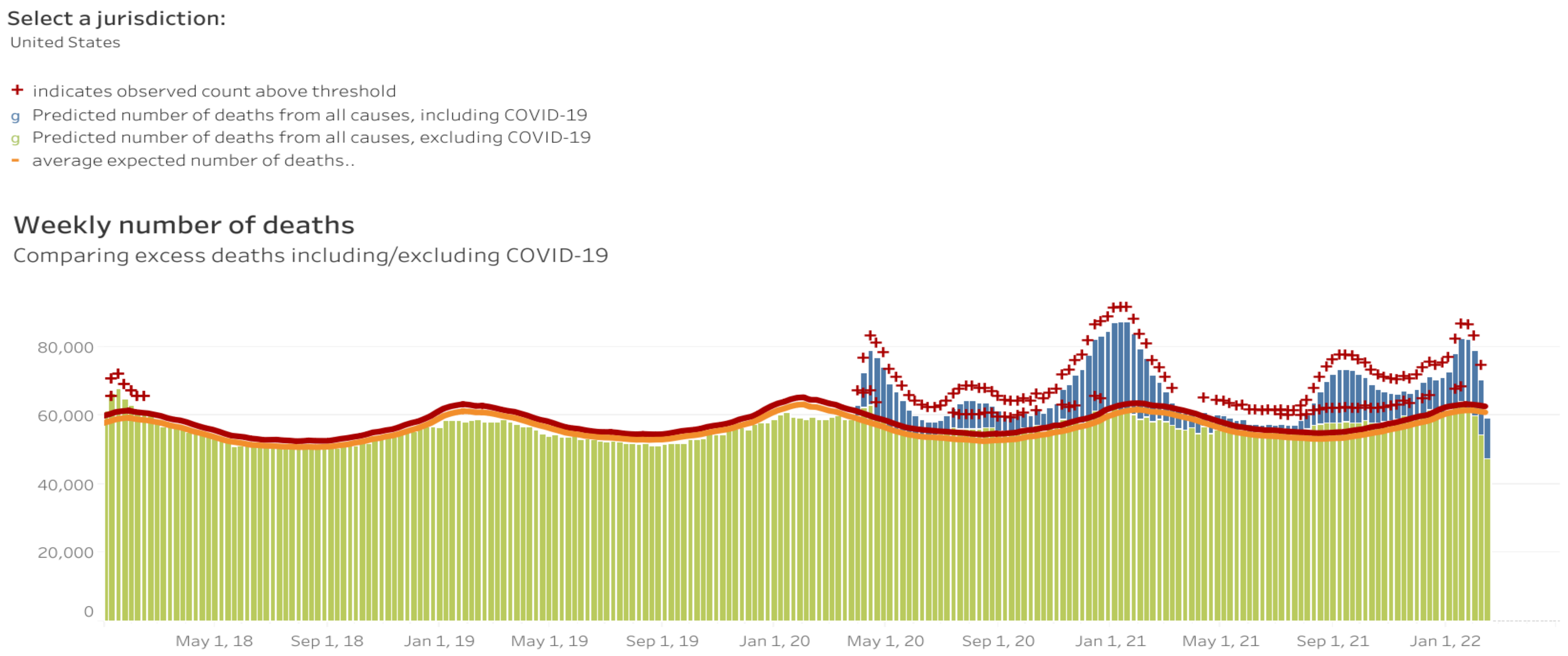
COUNTY ESTIMATES

Table 1- Top 12 Counties for Total Pandemic Deaths and proportion exceeding expected deaths.

Pandemic Deaths (PD)	Pandemic Deaths May Est.	Proportion exceeding expected deaths but not attributed to Covid 19
Los Angeles County, CA	28,354 (2X)	20.72%
Cook County, IL	14,818	31.82%
Maricopa County, AZ	14,721	30.66%
Harris County, TX	12,312	42.50%
Marion County, IN	8,271	69.68%
San Bernardino County, CA	7,973	29.76%
Bexar County, TX	7,768	37.80%
Miami-Dade County, FL	7,416	22.09%
Orange County, CA	7,358	19.75%
New York County, NY	7,007	39.28%
Tarrant County, TX	6,759	39.07%
Wayne County, MI	6,485	38.57%

- Mortality Model reflected trends seen in prior research
  - ⟶ **Magnitude of Covid 19 deaths was less than previous studies, possibly due to vaccines or mitigation**
  - ⟶ **Regional differences in morbidity patterns were in line with prior research**

RECENT CDC ESTIMATES



- Mortality Model reflects broader trends codified by CDC
  - → **Applies 6 years of data, running from 2014 to 2020**

PREDICTOR SELECTION

- Socio-demographic (Census/USDA) and Built Environment Characteristics using Open Sources w/Computer Vision
  ⟶ **Over 50+ County-level predictors**

- Categorized into a common Social Determinants of Health Taxonomy
  ⟶ **"Significant Seven" including Economic Well-Being, Food Insecurity, Social Support, Crime, etc.**

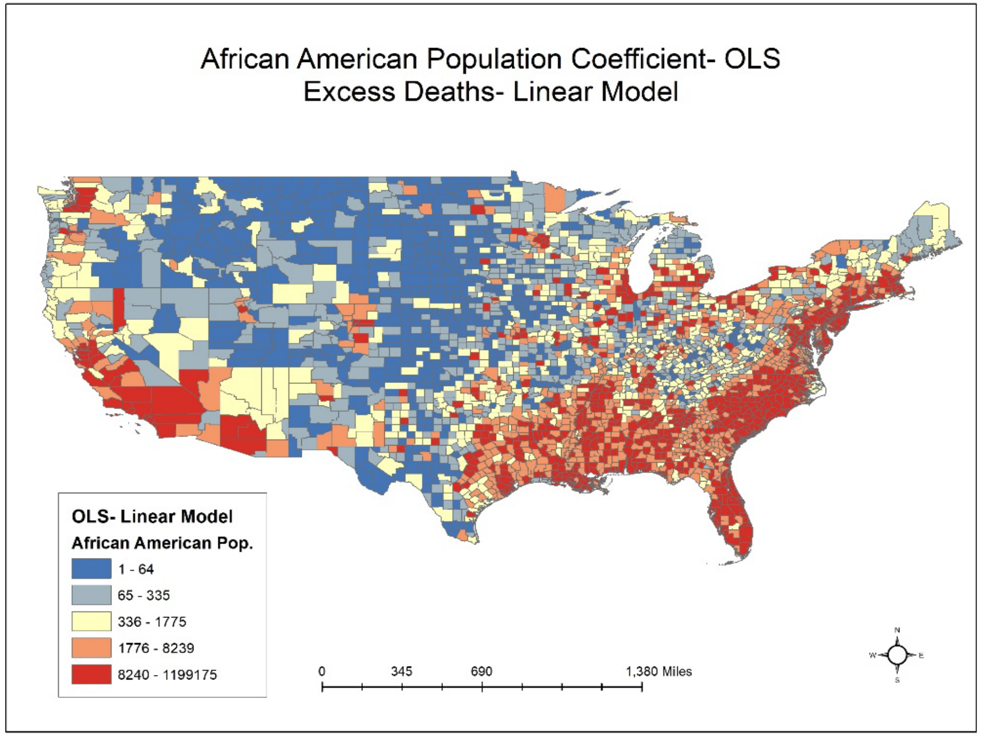
- Hypothesis tests- EDA, Linear/Poisson models, VIF tests, Stepwise Regression techniques
  ⟶ **Predictors were refined into Linear model and Poisson model groups (~6-8 predictors)**

FINAL	Housing	Transportation	Health Literacy	Food Insecurity	Social Support	Crime	Economic Well-Being
Poisson	prop_multi, prop_wires, mhh_inc	lng_com, dr_sing	vac_u, pct_smo	prop_green, pct_obe		prop_dilap, vc_rate	unemployme, deep_pov
Linear	med_ren	lng_com, dr_sing		pct_obe	b_pop	vc_rate	

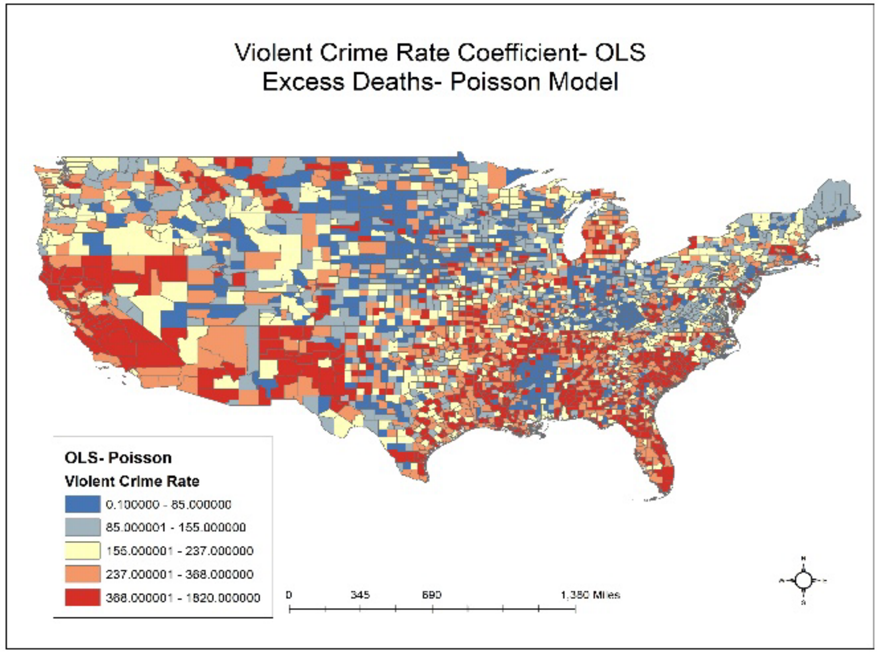
Poisson = MIX (13)

Linear = All Socio-demographic (6)

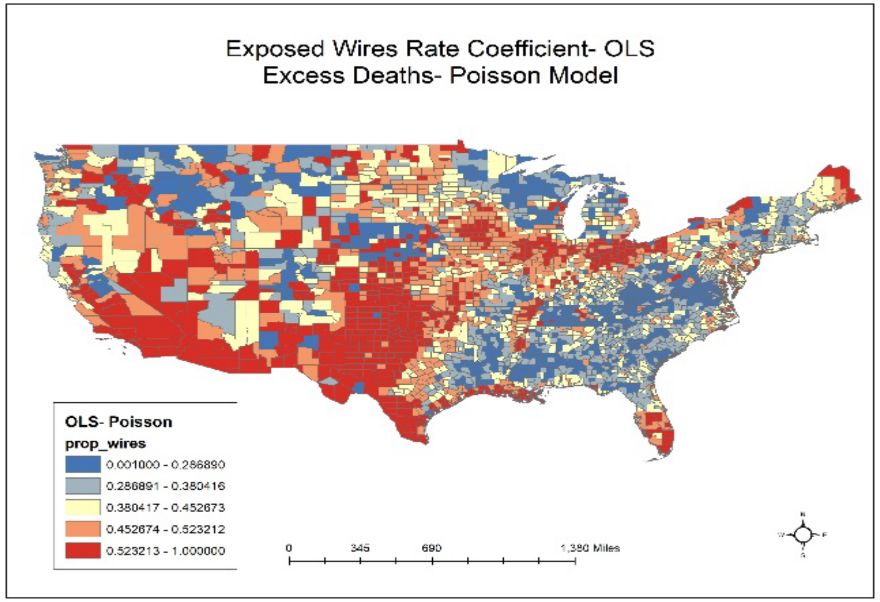
SPATIAL MODELS-OLS



OLS- Linear
R2-(>0.61)
Coeff- 0.001
Positive Autocorrelation

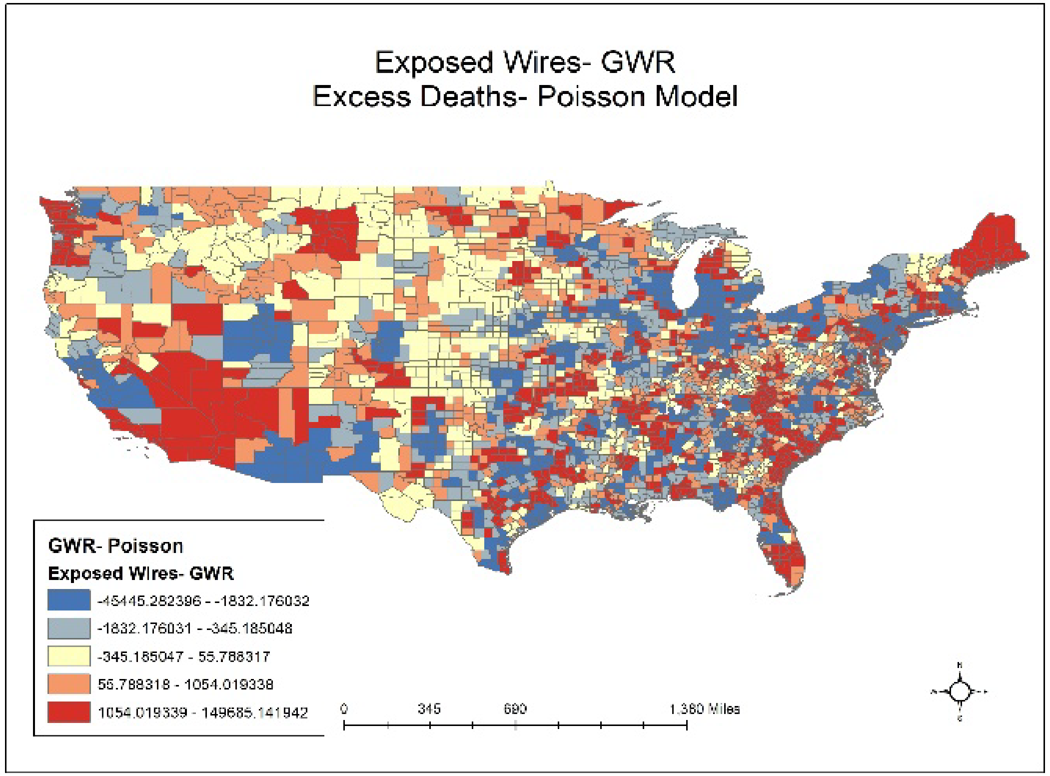


OLS- Poisson
Coeff- 1.39
R2 (>0.26)
Positive
Autocorrelation



OLS- Poisson
Coeff- 478.71
R2 (>0.16)
Positive
Autocorrelation

SPATIAL MODELS-GEOGRAPHICALLY WEIGHTED REGRESSION (POISSON)

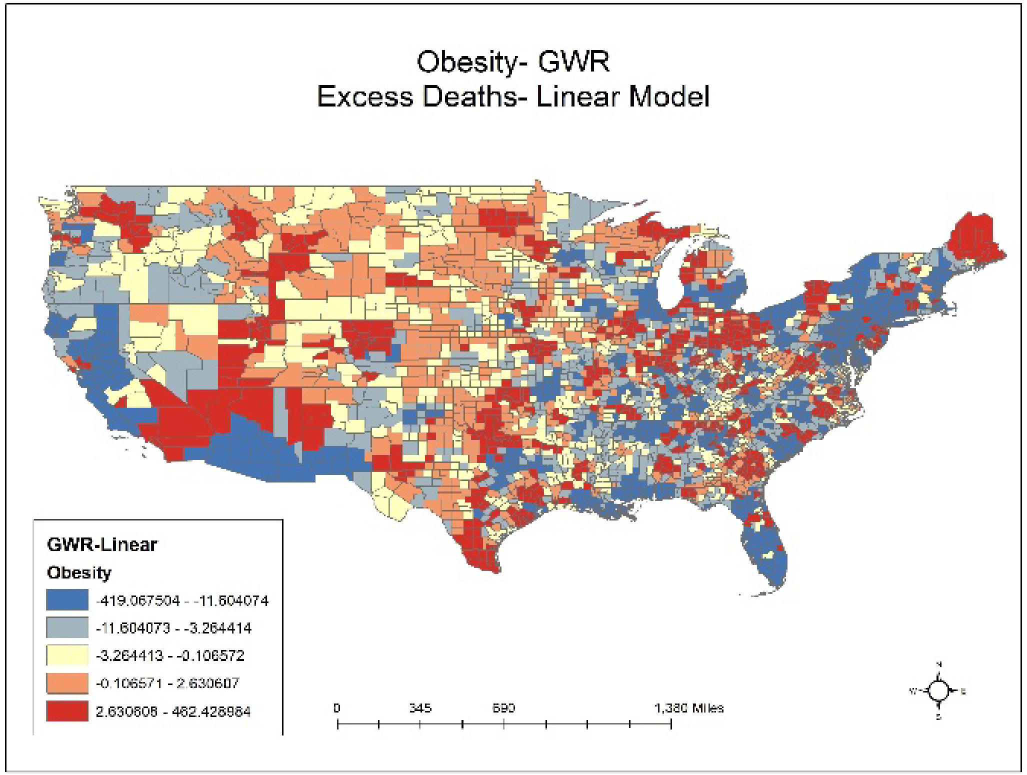


The strongest GWR models were those that applied built environment characteristics

GWR
Poisson- Strong (>0.9)
Minimal Autocorrelation

Test	Housing	Transportati on	Health Literacy	Food Insecurity	Social Support	Crime	Economic Well-Being
Poisson	prop_multi, prop_wires,			prop_green, pct_obe		prop_dilap, vc_rate	unemployme , deep_pov

SPATIAL MODELS-GEOGRAPHICALLY WEIGHTED REGRESSION (LINEAR)



The strongest Linear GWR models were those that applied socio-demographic characteristics

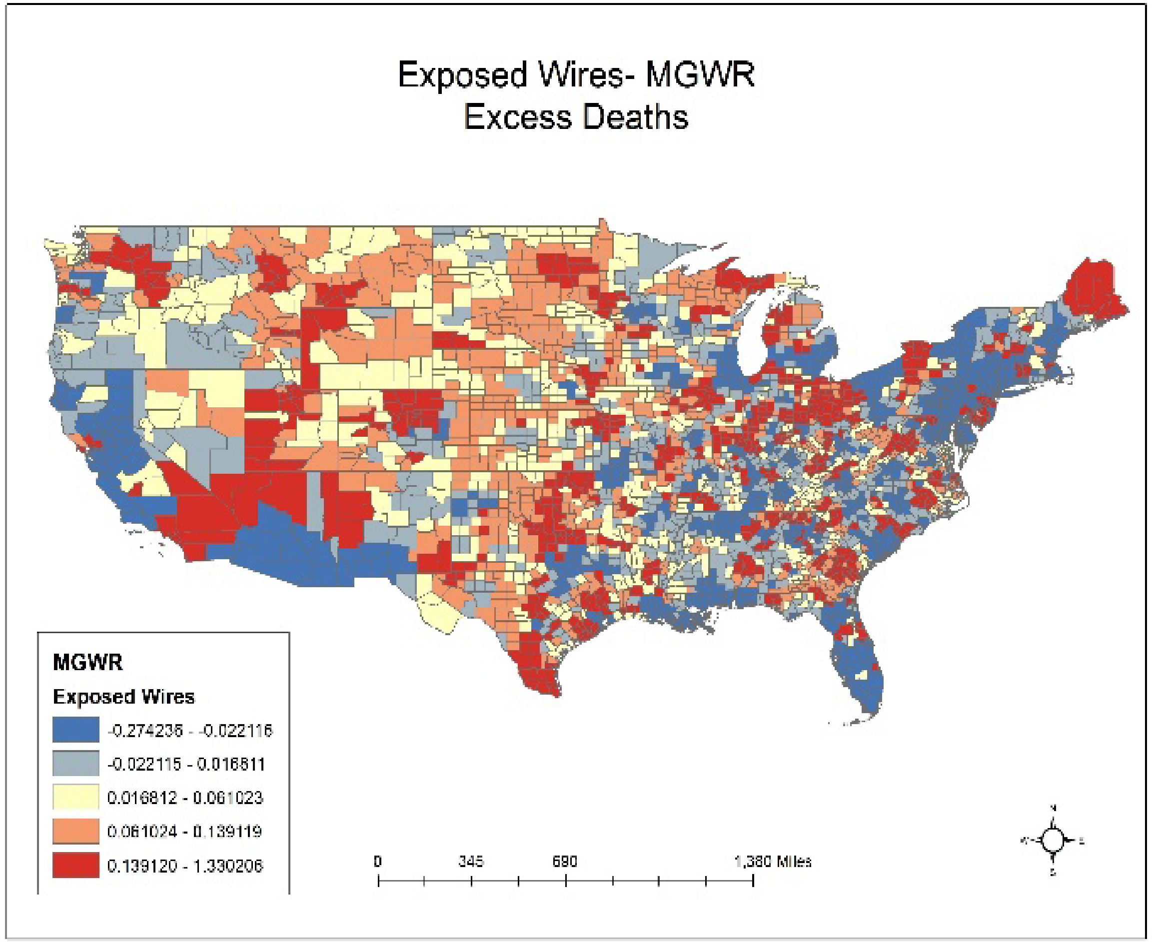
GWR
Linear- Strong (>0.9)
Minimal Autocorrelation

Test	Housing	Transportation	Health Literacy	Food Insecurity	Social Support	Crime	Economic Well-Being
Linear	med_ren	lng_com, dr_sing		pct_obe	b_pop	vc_rate	

MULTI-SCALE GEOSPATIALLY WEIGHTED REGRESSION

Variable	Bandwidth	Confidence Interval
prop_wires	155.000	(149.0, 174.0)
dr_sing	130.000	(108.0, 139.0)
mhh_inc	64.000	(61.0, 73.0)
vc_rate	44.000	(44.0, 44.0)
deep_pov	31.000	(2411.0, 3141.0)

Lower Bandwidths = Local Phenomena
Minimal Autocorrelation
Mixed Model- Fair (0.6)



FUTURE RECOMMENDATIONS

- Greater focus on use of Poisson models and tools for Pandemic analyses
  - Data lends itself to counts and rates

- Built environment matters, and presents an opportunity for more robust predictors
  - Strongest models were a mix, not a single type

- Variable selection was time consuming but important
  - Less focus on the R-squared for this research, more on developing a set of predictors

- Multi-scale modeling can help inform all aspects of the Emergency Management lifecycle, and target resources effectively

- Machine learning could enhance future applications of MGWR

MANY THANKS TO DR. TAYLOR OSHAN AND DR. NGUYEN FROM UNIVERSITY OF MARYLAND FOR THEIR SUPPORT AND DATA!