INVESTIGATING THE POTENTIAL IMPACT OF CLIMATE VARIABLES ON THE SPREAD OF SARS-COV-2 (COVID-19) IN THE UNITED STATES

Gus Lipkin, Florida Polytechnic University

A.K.M. Azad Hossain, Assistant Professor, University of Tennessee at Chattanooga

THE UNIVERSITY OF TENNESSEE CHATTANOOGA

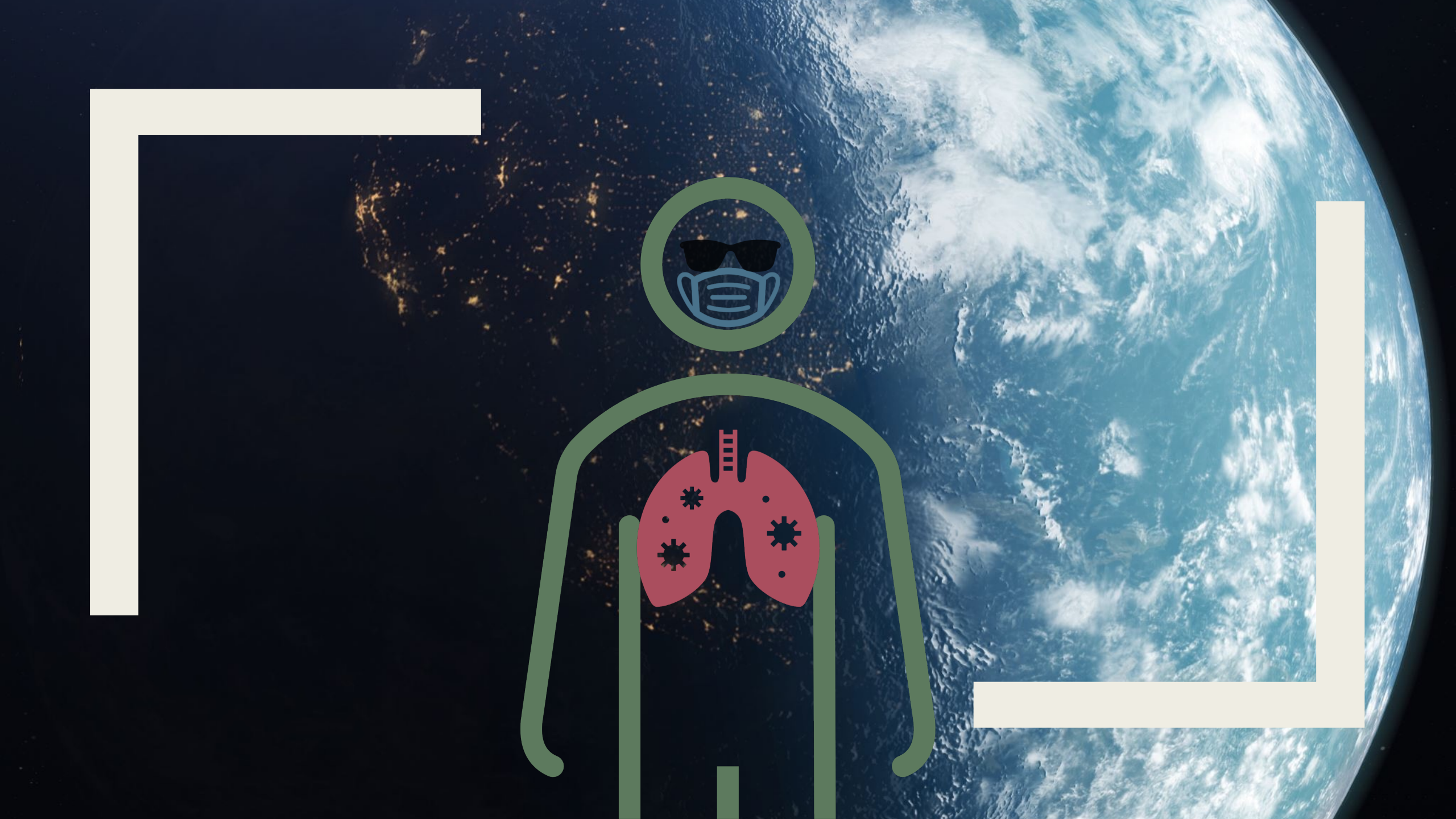
NSF

FLORIDA POLYTECHNIC UNIVERSITY

Outline

- Introduction
- Literature review
- Objective
- Data selection
- Data sources
- Methodology
- Results and Analysis
- Discussion and Conclusions
- Acknowledgements

INTRODUCTION



Since COVID-19 is spread through close contact, do any climate variables influence spread?

LITERATURE REVIEW

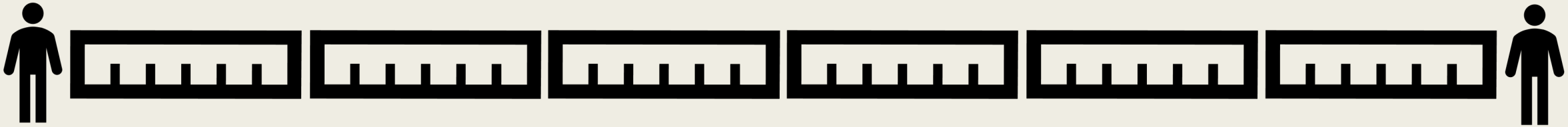
Association of Social Distancing, Population Density, and Temperature With the Instantaneous Reproduction Number of SARS-CoV-2 in Counties Across the United States

Motivation: In July 2020, COVID-19 was relatively new and the researchers thought that maybe the nicer weather would influence spread

Methods: Choose counties based on established criteria about case count, population, and spread. Calculate the R value, how much a virus spreads, from the data. Perform regressions using the EpiEstim and dlnm packages in R.

Results: The R value increased between 11-20 °C (51.8-68 °F). As social distancing increased, the R value decreased. The higher the population density, the more deaths per 100,000 people.

Conclusions: “social distancing, population density, and daily weather may account for variation in the R_t for SARS-COv-2 across the United States.”



Developing Numerical Models to Predict Potential Covid-19 Cases Using GIS and Regressions Incorporating Climate Variables

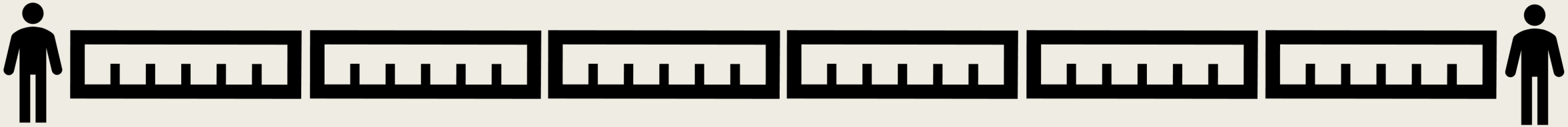
Motivation: Serena Hollis, a UTC REU student from last summer also had Dr Hossain as her advisor

Methods: Linear and non-linear regressions in JMP using soil and air temperature, soil moisture, and population density with data from Johns Hopkins and Copernicus

Results: Population density and air temperature have the greatest effect.

Conclusions: The prediction model was more effective for highly populated areas but cases are likely to rise in the future

Improvements: Smaller scale data, consider new variables like social distancing and mask usage



OBJECTIVES

Can we choose specific counties to capture the diversity in data and allow us to make customized predictions for county profiles?

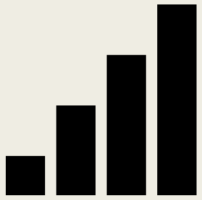
Investigate how the spread of COVID-19 cases relates to climate variables such as air temperature, precipitation, and soil moisture.

COUNTY SELECTION & STRATEGY AND STUDY SITE

COUNTY SELECTION PROCESS

Selection Variables

- Total cases
- Total cases until vaccine introduction (30/11/2020)
- County population
- Google Mobility report total grade
- Population density
- “spread” [total cases / population density]
- Mask usage with never, rarely, sometimes, frequently, and always



Selection Criteria

Goal:

- Find a sample of counties representative of the country at large

Finding the criteria:

- Obtain summary statistics for the selected variables

Selection Method

1. Choose an offset value large enough to capture all data `(max <= mean + offset) && (min >= mean - offset)`
2. Filter data where variables fall inside the given window
3. If the goal number of counties is not found
    1. *If the number of counties found is below the goal, increase the offset value precision (.1 → .01) and reset to the maximum offset*
    2. *If the number of counties found is above the goal, decrease the offset value (100 → 99)*
4. Repeat this process until counties are found that match the Q1, Median, Q3, and Mean summary statistics and are geographically and statistically separated

STUDY SITE

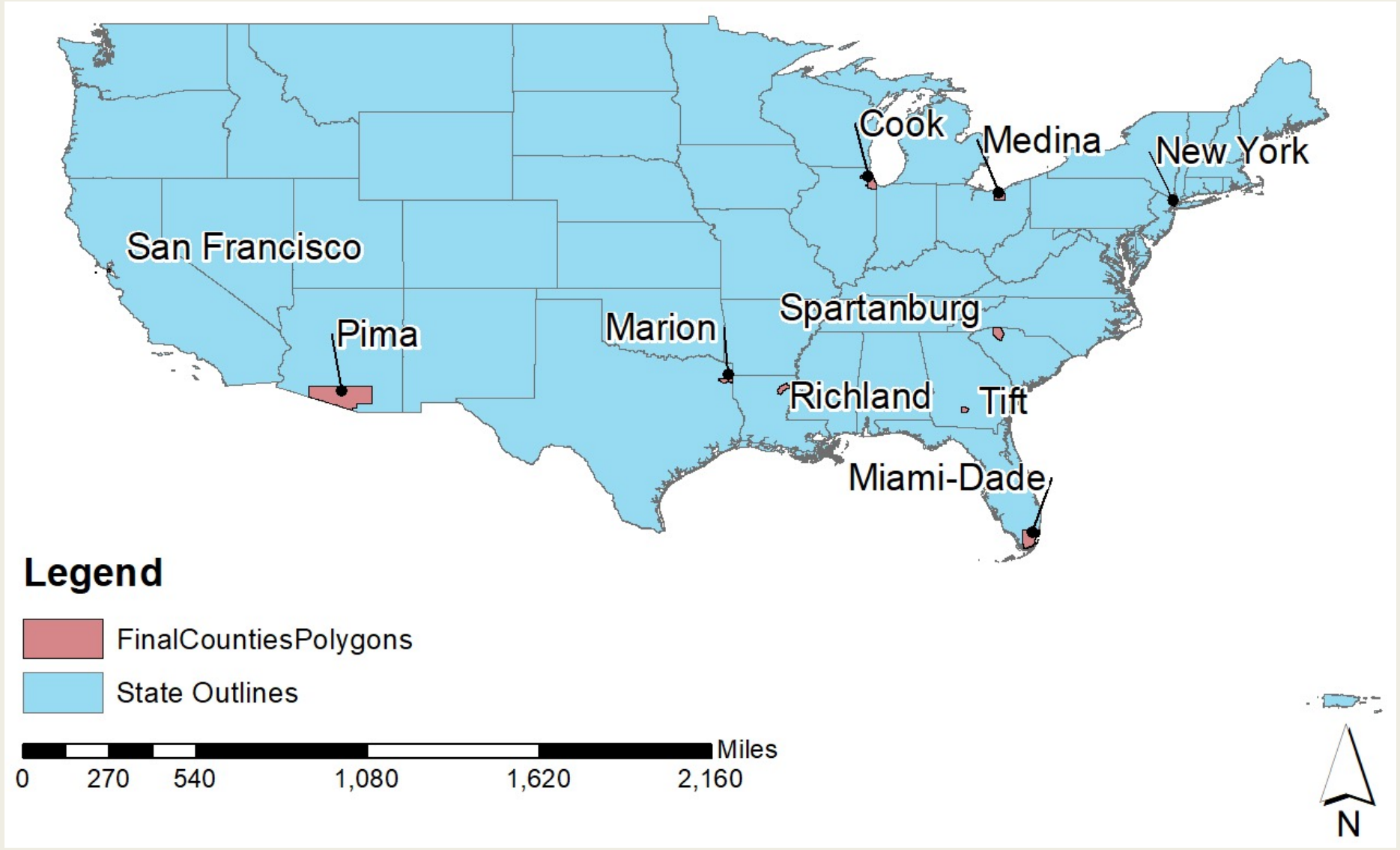
Key County Data

	County	Population	Population Density (per square mile)	Total Cases
1	Pima, Arizona	1,039,073	110	118,938
2	San Francisco, California	883,305	18,790.74	38,367
3	Miami–Dade, Florida	2,761,581	1,000	522,734
4	Tift, Georgia	40,571	155	5,046
5	Cook, Illinois	5,180,493	5,450	559,767
6	Richland, Louisiana	20,192	36	2,597
7	New York, New York	1,628,701	69,467.5	140,314
8	Medina, Ohio	179,146	2,240.87	15,744
9	Spartanburg, South Carolina	313,888	350	42,302
10	Marion, Texas	9,928	28	637

Key County Data

	County	Population	Population Density (per square mile)	Total Cases
1	Pima, Arizona	1,039,073	110	118,938
2	San Francisco, California	883,305	18,790.74	38,367
3	Miami–Dade, Florida	2,761,581	1,000	522,734
4	Tift, Georgia	40,571	155	5,046
5	Cook, Illinois	5,180,493	5,450	559,767
6	Richland, Louisiana	20,192	36	2,597
7	New York, New York	1,628,701	69,467.5	140,314
8	Medina, Ohio	179,146	2,240.87	15,744
9	Spartanburg, South Carolina	313,888	350	42,302
10	Marion, Texas	9,928	28	637

Map of Selected Counties



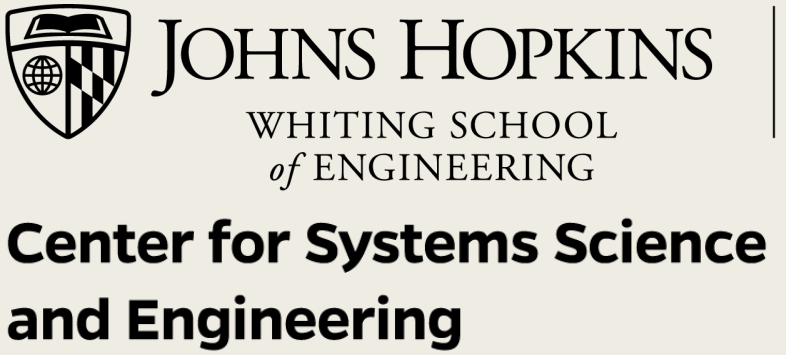
DATA & SOURCES

Climate Data

- Sourced from Copernicus.eu
- Geographical time series data for air temperature, precipitation, and soil moisture every day at 3pm
  - *3pm was chosen because it is generally found to be the hottest time of day*

COVID-19 Data

- Sourced from Center for Systems Science and Engineering (CSSE) at Johns Hopkins University
- Time series data of total case count for each county in the United States



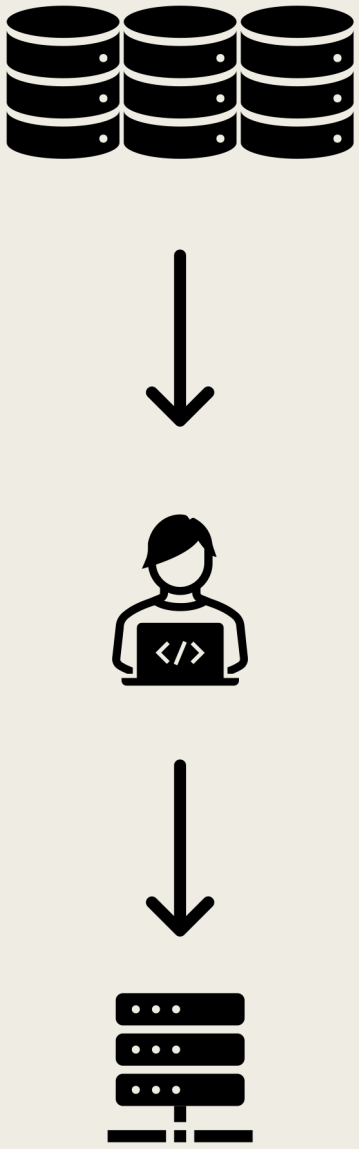
Secondary Data

- New York Times mask survey data from July 2020
- Google Mobility Data is a daily document detailing movement and industry visits (shopping, restaurants, etc)

METHODOLOGY

County Data Aggregation

1. Aggregate all the data possible into one giant data table
2. Perform the county selection process
3. For each county, filter that data and transpose it from wide to long data then group by month
4. Pull the time series climate data for each county from the climate data in ArcMap
5. Merge the climate data and case data in R and write a file for each county



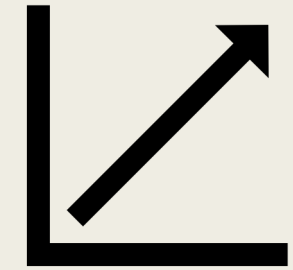
The State of the Data

Example of Data from Pima, Arizona

Month	Year	Total New Cases	Temperature	Soil Moisture	Precipitation
1	2020	0	266.551117	0.21569824	0.00195767
2	2020	0	272.083618	0.33831787	0.00117531
3	2020	202	276.732635	0.39414978	0.00332341
4	2020	1039	283.822433	0.29469299	0.00117324
5	2020	1127	291.20546	0.08180237	0.00049297
6	2020	5636	298.316925	0.02656555	6.4993E-05
7	2020	8163	298.688828	0.02816772	0.0001714
8	2020	5054	298.185226	0.04516602	0.00071669
9	2020	4407	293.498795	0.02424622	5.2507E-07
10	2020	2970	286.173187	0.02497864	3.9767E-07
11	2020	11261	278.698593	0.14277649	0.00105689
12	2020	29663	273.104004	0.19335938	0.00034601

County Data Analysis

1. Perform multiple linear regressions in program of choosing (Excel, R, and STATA) for each county individually
2. Analyze the obtained results



RESULTS & ANALYSIS

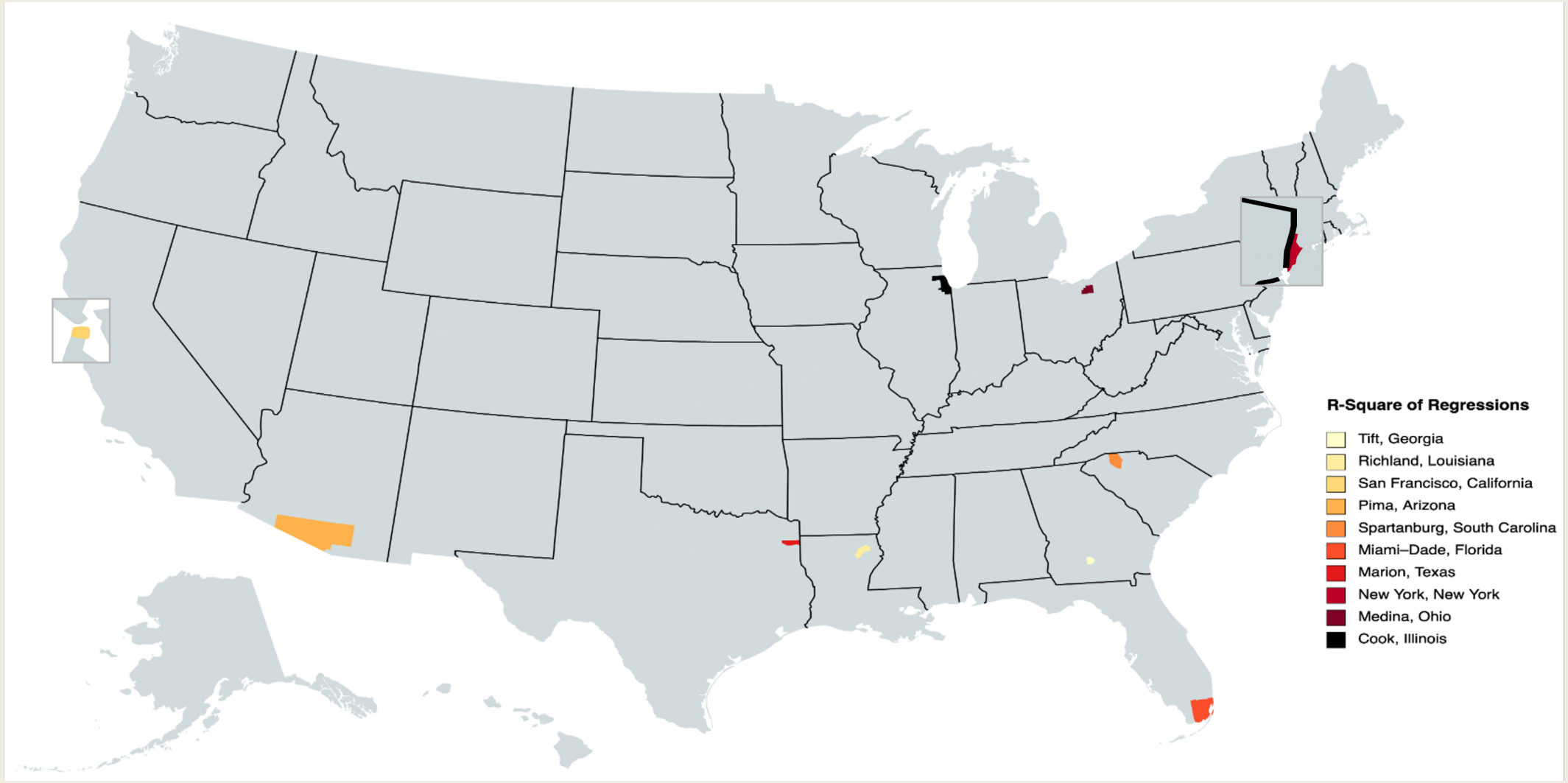
Quick Facts

County	R^2	Temperature P-Value	Soil Moisture P-Value	Precipitation P-Value
Pima, Arizona	0.26790686	0.33439462	0.97643934	0.26287927
San Francisco, California	0.21277399	0.30105978	0.40027999	0.71632608
Miami–Dade, Florida	0.37350872	0.25076117	0.52415012	0.21225603
Tift, Georgia	0.08011189	0.6006376	0.48833484	0.54666331
Cook, Illinois	0.76268884	0.00691822	0.00160986	0.22655311
Richland, Louisiana	0.11374961	0.36307953	0.55731974	0.45032471
New York, New York	0.39831268	0.88800433	0.17927839	0.98352732
Medina, Ohio	0.61505672	0.00826739	0.03470477	0.03327895
Spartanburg, South Carolina	0.30933901	0.10048437	0.19534938	0.2135677
Marion, Texas	0.38347787	0.93600009	0.10936241	0.0895341

Quick Facts

County	R^2	Temperature P-Value	Soil Moisture P-Value	Precipitation P-Value
Pima, Arizona	0.26790686	0.33439462	0.97643934	0.26287927
San Francisco, California	0.21277399	0.30105978	0.40027999	0.71632608
Miami–Dade, Florida	0.37350872	0.25076117	0.52415012	0.21225603
Tift, Georgia	0.08011189	0.6006376	0.48833484	0.54666331
Cook, Illinois	0.76268884	0.00691822	0.00160986	0.22655311
Richland, Louisiana	0.11374961	0.36307953	0.55731974	0.45032471
New York, New York	0.39831268	0.88800433	0.17927839	0.98352732
Medina, Ohio	0.61505672	0.00826739	0.03470477	0.03327895
Spartanburg, South Carolina	0.30933901	0.10048437	0.19534938	0.2135677
Marion, Texas	0.38347787	0.93600009	0.10936241	0.0895341

Map of Selected Counties Colored by R^2 Value



Map of Selected Counties Colored by the Most Impactful Variable by P-Value

CONCLUSIONS

Conclusion 1

- Climate variables such as temperature, precipitation, and soil moisture can be good influential factors of COVID-19 virus spread
  - *The results indicate that this works better for some counties than others*

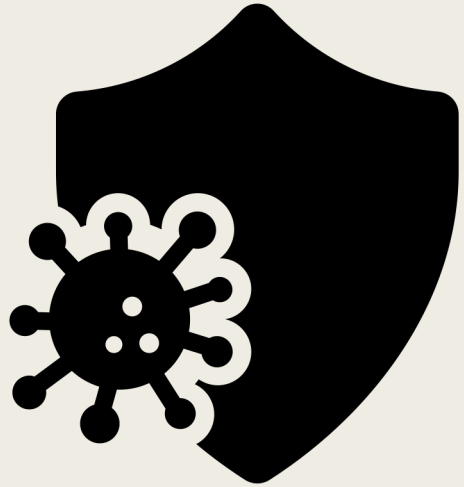
Conclusion 2

- Although the counties affected most by precipitation are in the north and those affected by soil moisture and temperature are in the south, there is not enough data to form a conclusion


NEXT STEPS

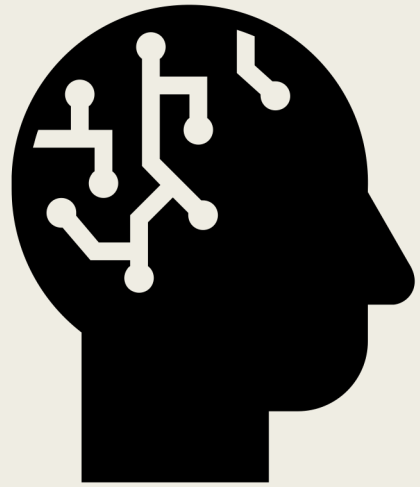
Incorporate Vaccination Data

- As the vaccine rollout continues and the number of vaccinated stabilize, we will be able to better forecast future outbreaks provided the vaccines provide similar levels of protection against future variants



Forecasting and Cross-Validation

- Cross-Validation allows us to forecast on data we already have
- Forecasting future COVID-19 cases helps combat the pandemic



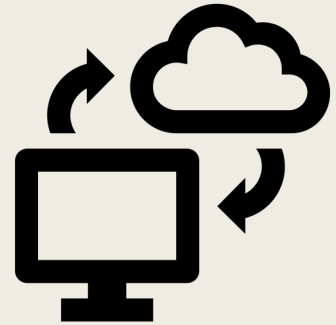
Perform More Advanced Regressions

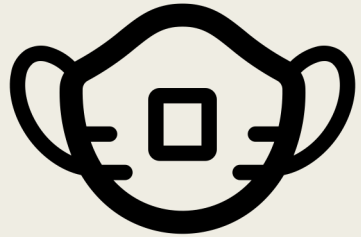
- With geographically weighted regressions, we can see if placement in the country has any effect
- We will also be able to make forecasts for all counties, rather than specific counties

OBSTACLES

Technological and Resource Challenges

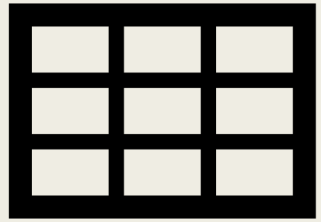
- Took a long time to set up and learn the basics of ArcGIS
- Challenges with file sharing and dataset storage especially with regards to remote work
  - *GitHub provides lots of opportunity, but severely limits file sizes*
- Data on masking and social distancing is not very complete





Knowledge

- I’ve never used ArcGIS before and it has a very steep learning curve
- I’ve never worked with panel data before in any capacity
  - *Time series data for multiple items (i.e. counties)*
- Some file types are harder to work with and not well supported by familiar tools (R, Excel, or Stata)



ACKNOWLEDGEMENTS

National Science Foundation

- Funding came from the National Science Foundation Undergraduate Research Opportunity Grant #1852042

University of Tennessee at Chattanooga

- This REU was organized and supported by the University of Tennessee at Chattanooga
- Special thanks to Dr Hong Qin, Principal Investigator of the iCompBio REU program

Research Mentor and Supporting Staff

- Dr Azad Hossain, without whom this project would not be possible and to who I am forever grateful for being so patient with me and helping me learn
- Nyssa Hunt from the IGTLab at UTC who helped me through many ArcGIS problems

THE UNIVERSITY OF TENNESSEE CHATTANOOGA

INVESTIGATING THE POTENTIAL IMPACT OF CLIMATE VARIABLES ON THE SPREAD OF SARS-COV-2 (COVID-19) IN THE UNITED STATES

Gus Lipkin, Florida Polytechnic University

A.K.M. Azad Hossain, Assistant Professor, University of Tennessee at Chattanooga

THE UNIVERSITY OF TENNESSEE CHATTANOOGA

NSF

FLORIDA POLYTECHNIC UNIVERSITY