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

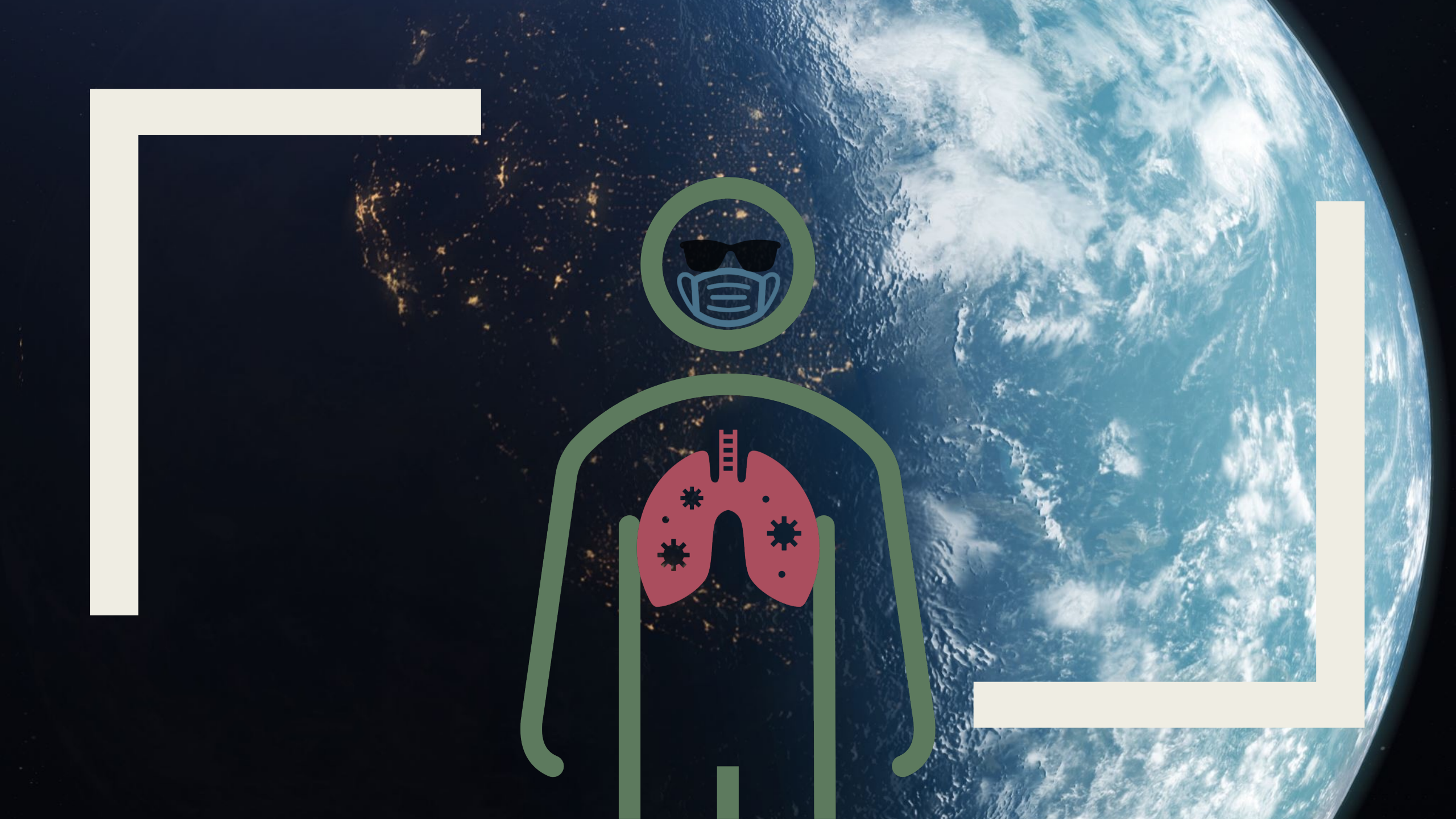
NSF

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

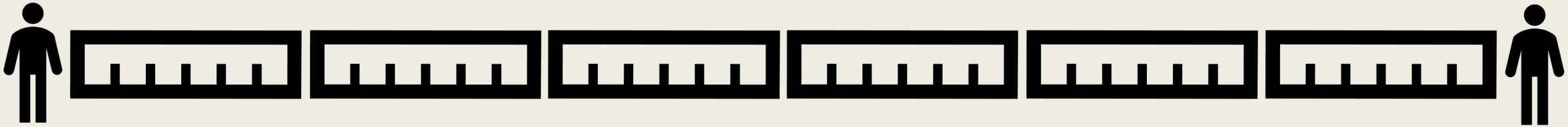
Association of Social Distancing, Population Density, and Temperature With the Instantaneous Reproduction Number of SARS-CoV-2 in Counties Across the United States

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-10 Cases Using GIS and Regressions Incorporating Climate Variables

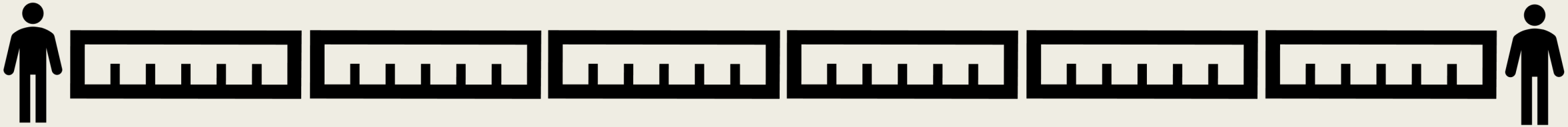
Motivation: Serena Hollis, a UTC REU student from last summer also had Dr Hossain as her advisor

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?

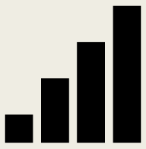
Investigate how the spread of COVID-19 cases relates to climate variables such as air temperature, precipitation, and soil moisture.

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 – 1)]
- Mask usage with never, rarely, sometimes, frequently, and always





Selection Criteria

Goal:

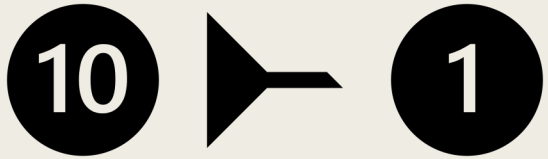
- Find a sample of counties representative of the country at large

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

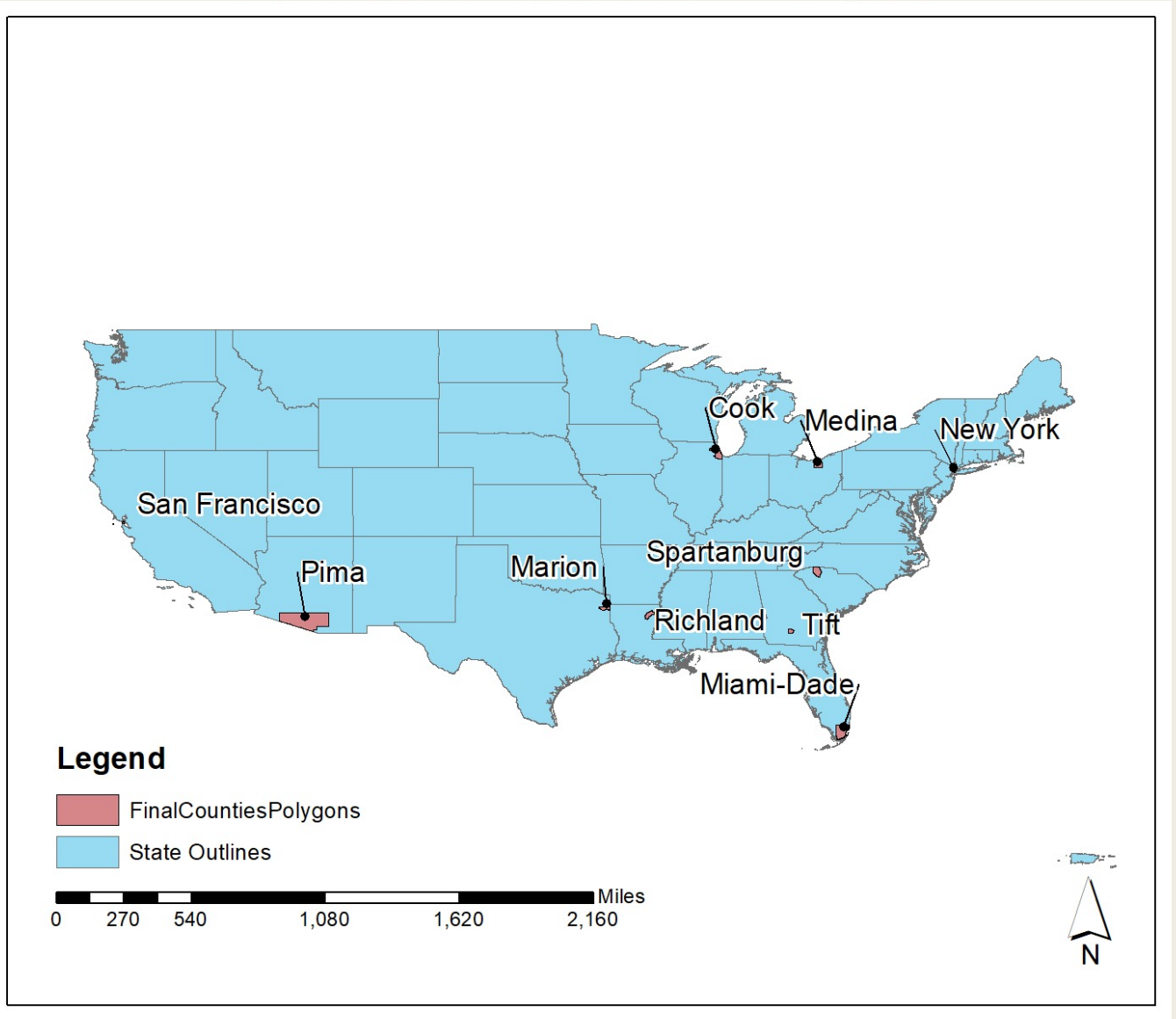


STUDY SITE

Quick Facts

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

Map of Selected Counties



DATA AND 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

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

Google

The New York Times

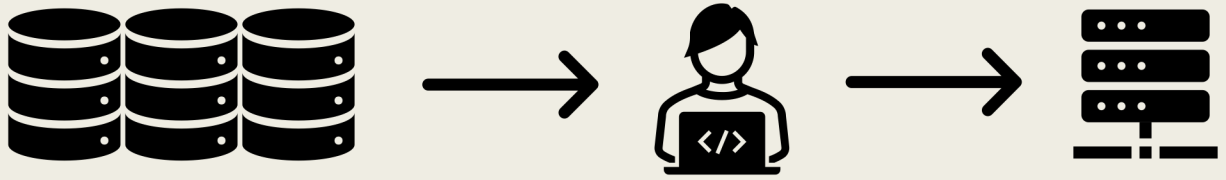
METHODOLOGY

Tying all the data together

Parameters used

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



County Data Analysis

1. Perform linear regressions in program of choosing (Excel, R, and STATA)

RESULTS/ANALYSIS

DISCUSSION & CONCLUSIONS

OBSTACLES

NEXT STEPS

ACKNOWLEDGEMENTS

NSF, UTC