Climate Impact on Health in Varying Community Types

PPOL 5203 Final Project

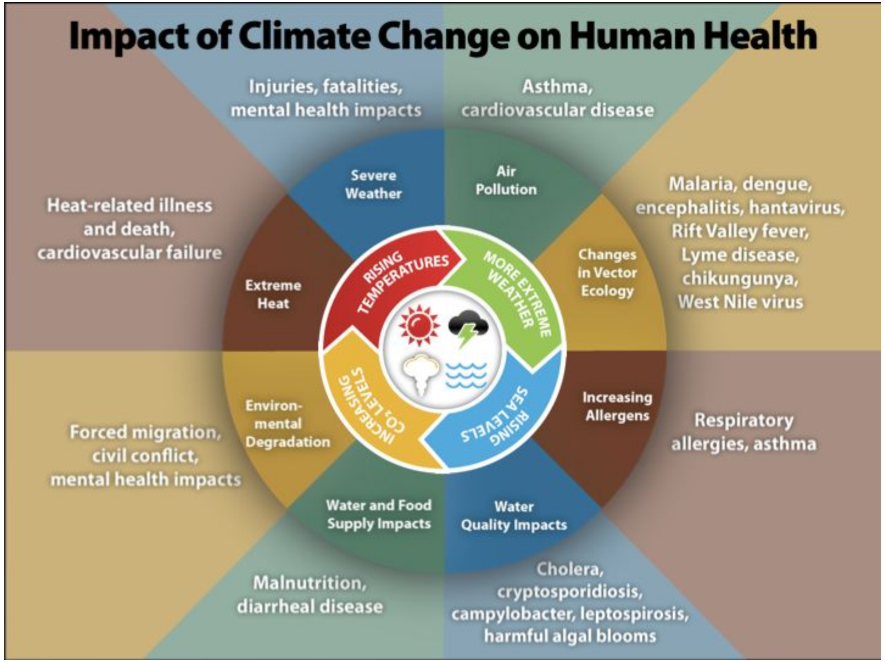
Bridgette Sullivan, Holt Cochran, and Katharyn Loweth

Agenda

1. Motivation & Background
2. Research Questions
3. Data Sources
4. Methods – OLS Modeling and Predictive Machine Learning (Decision Trees)
5. Results using GIS
6. Conclusion/Next Steps

Intensified climate change risks

- Climate change has intensified in recent years, leading to diverse and severe effects in different regions, notably the Southeast.
- The impact of climate change varies, and some communities are better equipped to handle them due to factors like infrastructure and social structures.

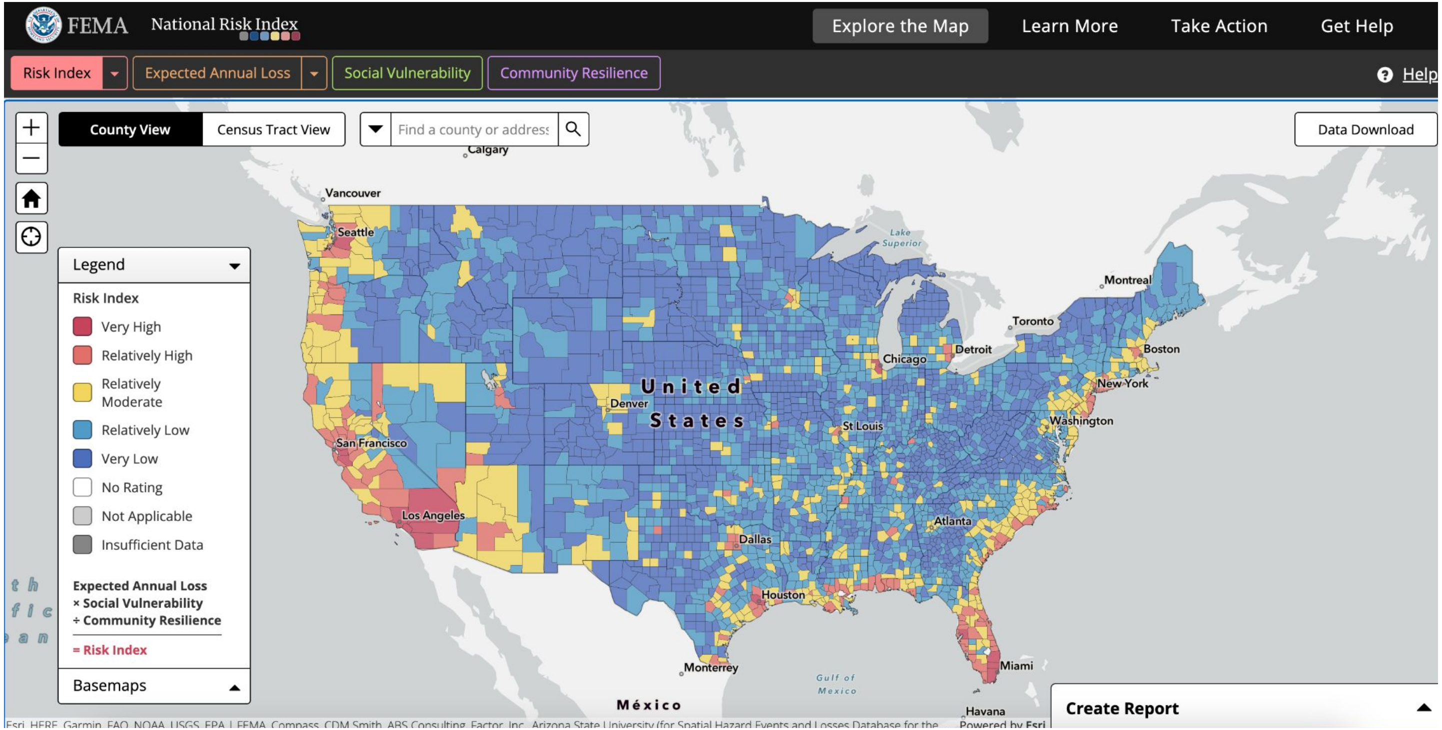


FEMA National Risk Index (NRI)

- FEMA created and published the National Risk Index (NRI) in 2020, assessing the risk of every US county and census tract to various natural disasters. The index is a composite score/rating that considers expected annual economic loss, social vulnerability, and community resilience to 18 natural hazards.

- Few studies have applied FEMA's National Risk Index in the context of public health.

FEMA National Risk Index (NRI)



Research Questions

- How does a county’s climate risk affect its population’s health?
  - Is there a health outcome that is significantly impacted by increased climate risk?
- How does the impact of climate risk vary by urbanicity?
- Can the FEMA NRI Risk Score be used to accurately predict a county’s health outcomes in the Southeast region of the country?

Data Sources

- 2023 FEMA National Risk Score
  - Composite Score consists of the following three variables:
    - Expected Annual Loss (specifically in Building, Population (Fatalities), and Agriculture)
    - Social Vulnerability Score (composite of demographics such as race, income, age, etc.)
    - Community Resilience (factors such as social, economic, infrastructure)

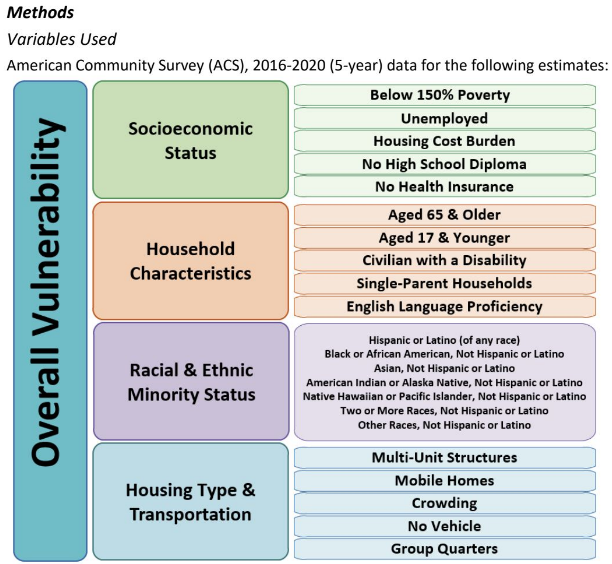
$$\text{Risk} = \frac{\text{Expected Annual Loss} \times \text{Social Vulnerability}}{\text{Community Resilience}}$$

- 2023 County Health Rankings
  - Life Expectancy
  - Age-Adjusted Death Rate
  - Average number of physically unhealthy days reported in past 30 days (age-adjusted).
  - Average number of mentally unhealthy days reported in past 30 days (age-adjusted).

Data Wrangling

- Selected necessary variables when importing data
  - Reviewed the components of the FEMA NRI composite score so that there would not be overlapping information/multicollinearity when we added other variables from the county health dataset.
- Joined tables via common county identifier column (FIPS)
- Renamed columns so that words are joined by "_"

Ex. The Components of the Social Vulnerability Score in the NRI



Data Wrangling: Creating New Variables

- Community Type
  - Used Population to create Community_Type variable based on the CDC's classification of Rural-Urban counties
- Population Density
- Health Outcomes Above or Below National Average
  - Created Binary Variables for the 4 dependent variables to determine whether a county was above or below the national county average

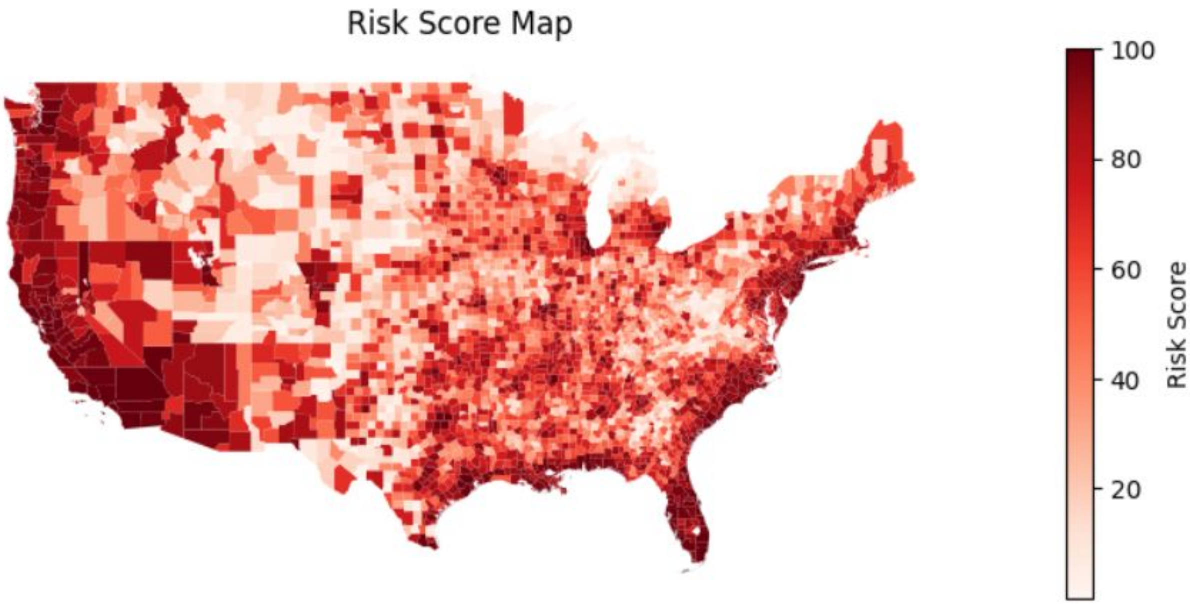
<u>Community Type</u>

- **Rural**
  - **Population < 50,000**
- **Small Urban**
  - **50,000 < Population <250,000**
- **Urban**
  - **Population > 250,000**

GIS - NRI for Continental United States

Initial Observations:

- Coastal areas highest risk
- Higher population density leads to higher risk
- Great plains region appears to have lower risk



Subsetting Data

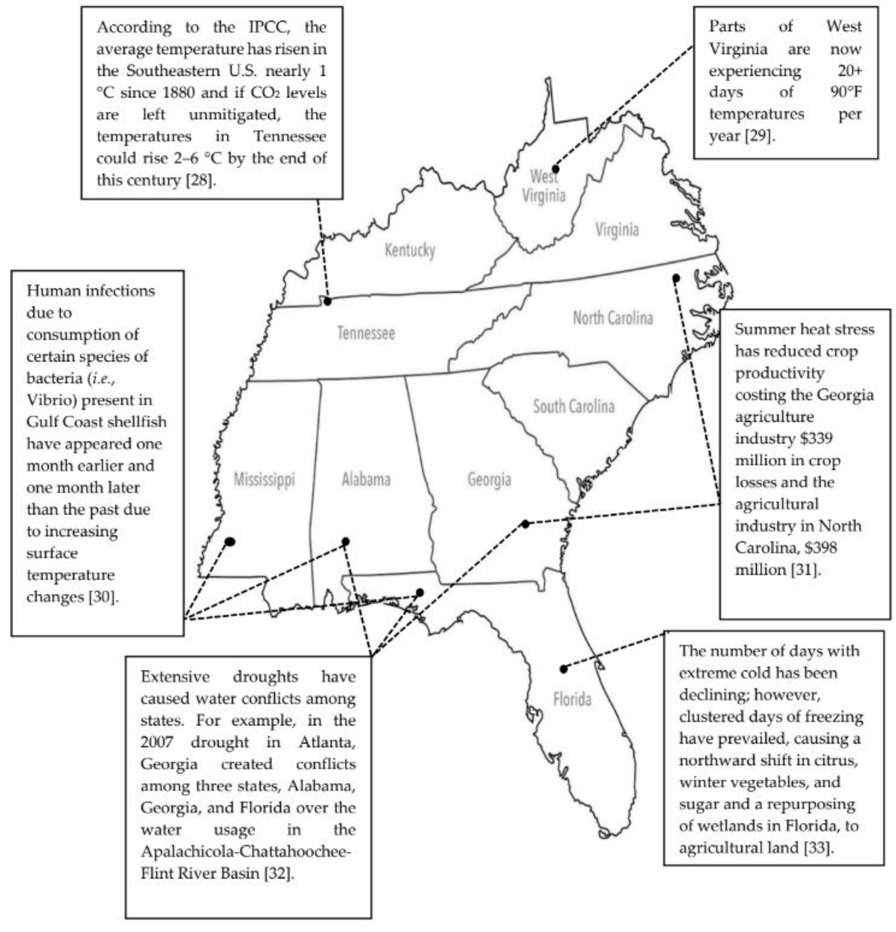
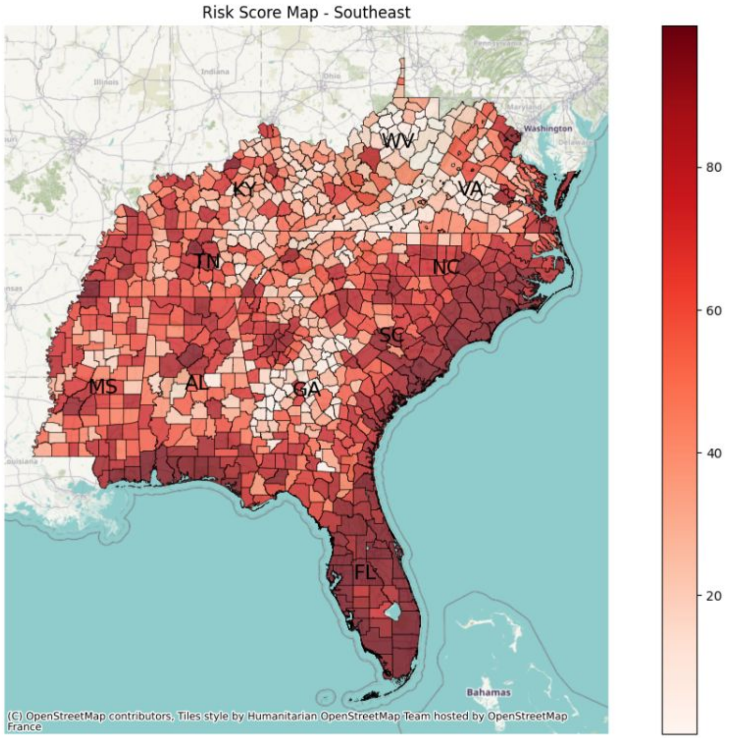


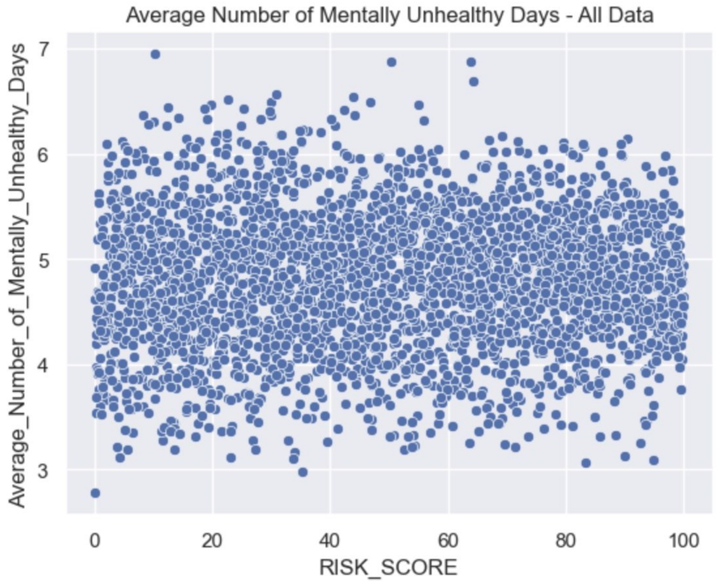
Figure 1. Map of Southeastern U.S. region with corresponding climate change effects identified.

Subsetting Data

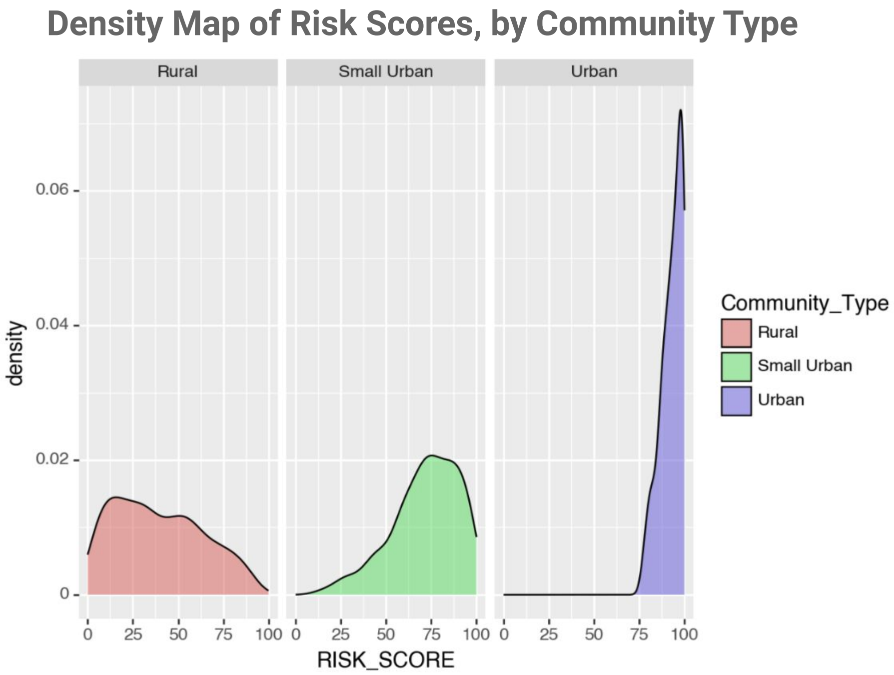
- Why the Southeast?
  - Homogeneous Natural Disasters and Climate Risks
    - Hurricanes, Heat waves, etc.
  - Distribution of Rural, Small Urban, and Urban
  - Highest risk region of the country



Subsetting Data



Subsetting Data



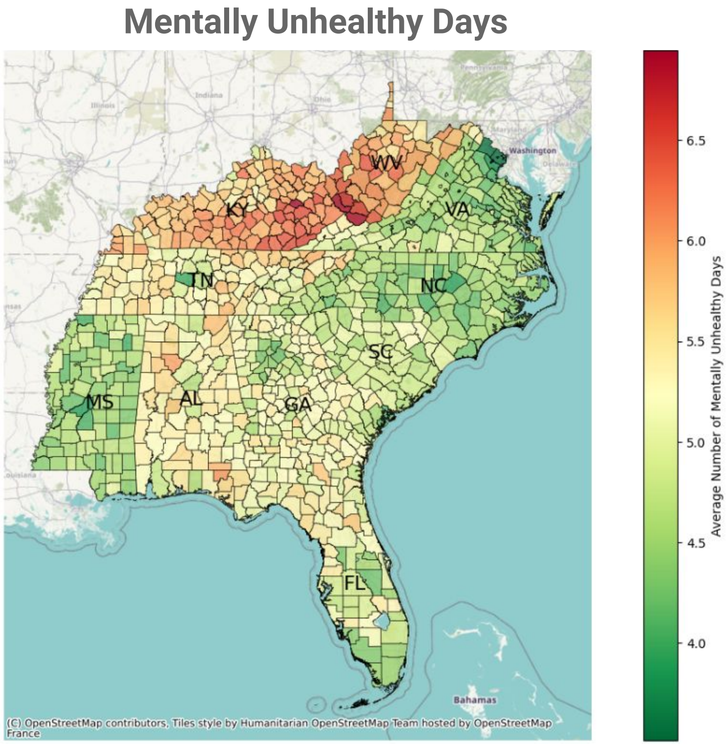
N = 924

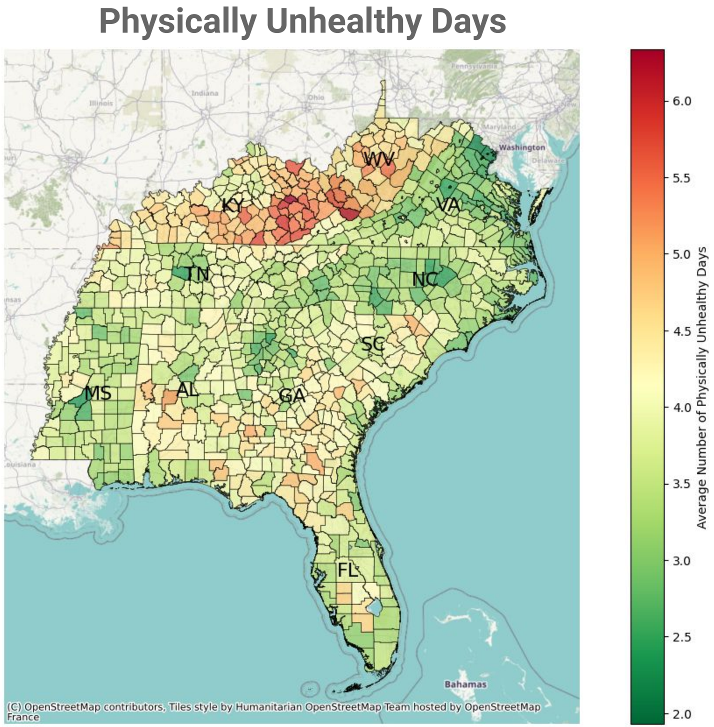
Rural
625

Small
Urban
237

Urban
62

GIS- Mentally v. Physically Unhealthy Days





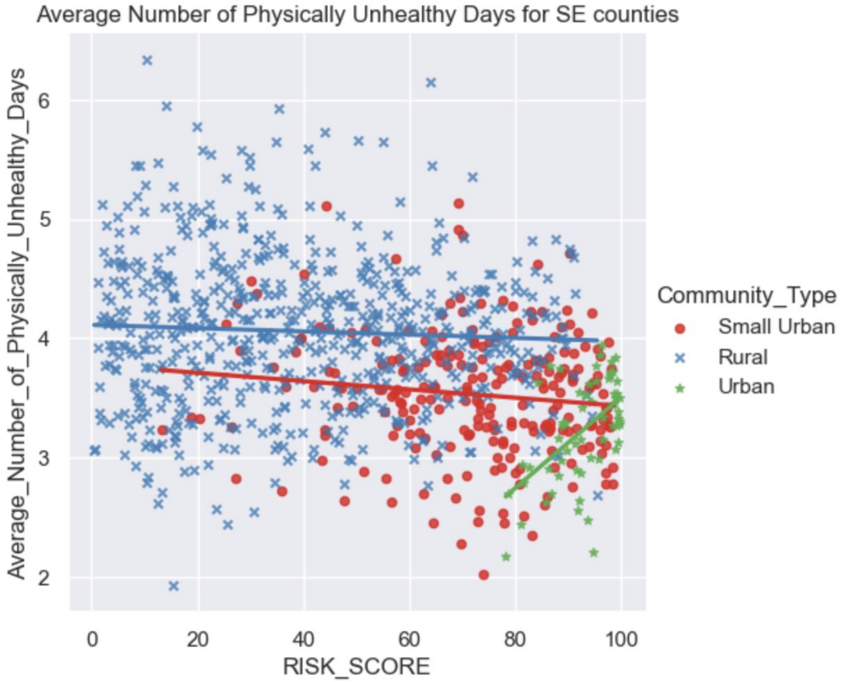
Analysis Methods: OLS and Machine Learning

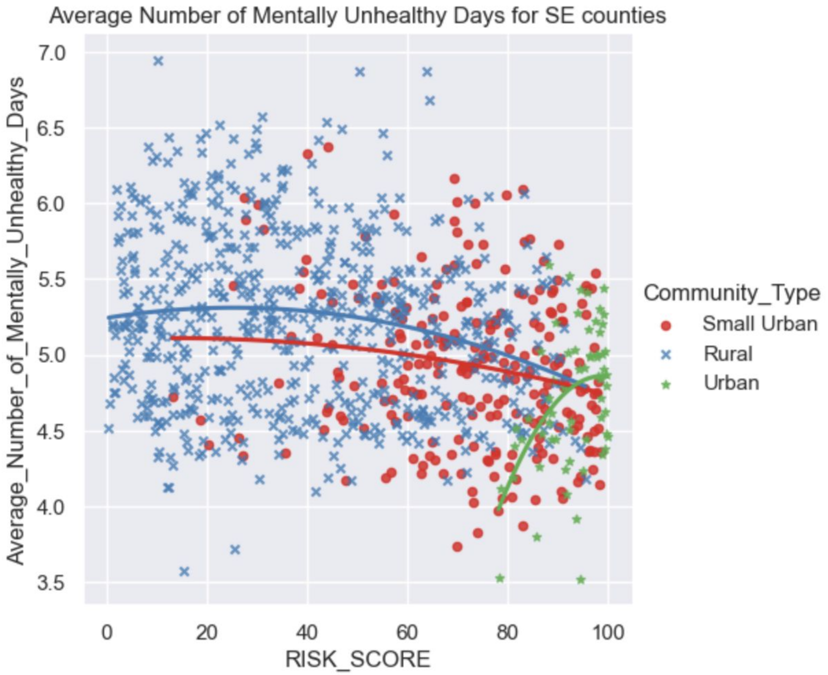
Methods: OLS Models

- Ran various exploratory OLS models
  - Interaction Terms
    - Community Type
  - Risk Score Squared

$$HealthOutcome_i = B_0 + B_1 RiskScore_i + B_2 RiskScore_i^2 + B_3 CommunityType_i$$
$$+ B_4(CommunityType_i * RiskScore_i) + B_5 PopulationDensity_i$$

Example of Scatterplots with Fitted Lines





OLS Models

Findings across the models for the four different dependent variables:

- Risk Score, Risk Score Squared, Community Type, and Population_Density were usually statistically significant
- Risk Score Squared not statistically significant for physical health
- **Signs of the estimated coefficients in our regression models (excluded category Rural):**

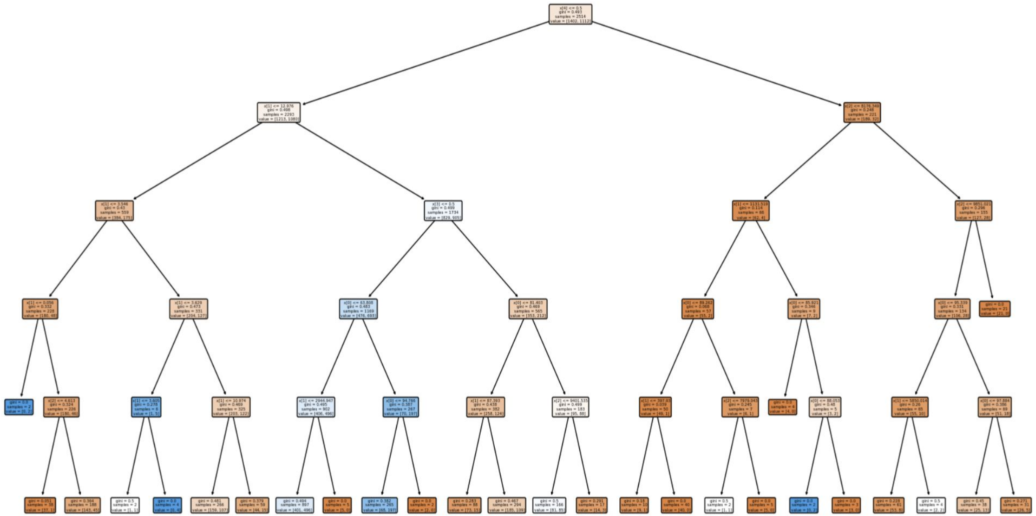
	Life Expectancy	Death Rate	Physical Health	Mental Health
Risk Score	-	+	-	+
Risk Score Sq	+	-	N/A	-
Small Urban	+	-	-	-
Urban	+	-	-	-

OLS Models

- Interaction between Community Type and Risk Score varied but mostly not significant
  - Effect of Interaction between Risk Score and Community_Type Urban (population greater than 250,000) typically stronger than effect of interaction between Risk Score and Community_Type Small Urban (population between 50,000 and 250,000)

Machine Learning – Decision Tree Model

- Predict if Southeast counties are above or below national average in different health metrics
- Independent Variables:
  - Risk Score
  - Risk Score Squared
  - Population Density
  - Community Type
- Dependent Variables
  - Mentally Unhealthy Days
  - Physically Unhealthy Days

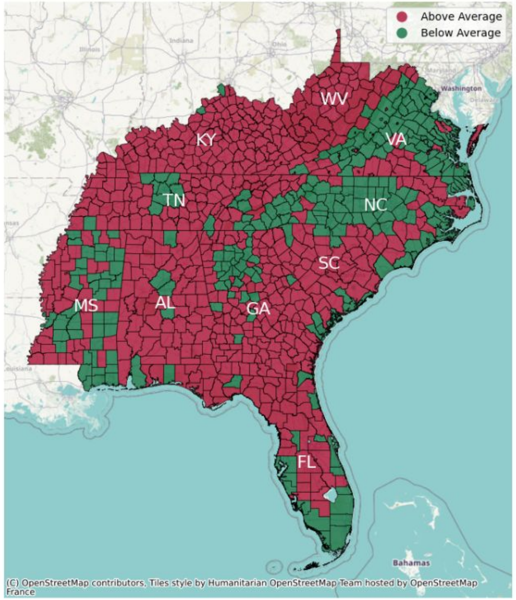


Results of Machine Learning

GIS Results - Machine Learning Predictions for Physically Unhealthy Days

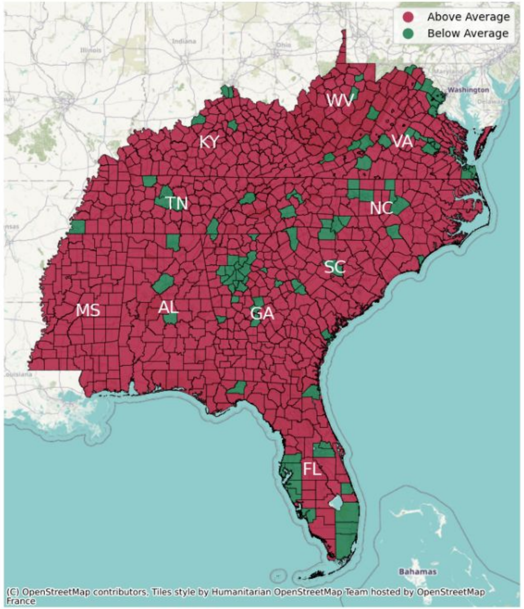
Actual

Average Number of Physically Unhealthy Days Above National Average Map - Southeast



Predicted

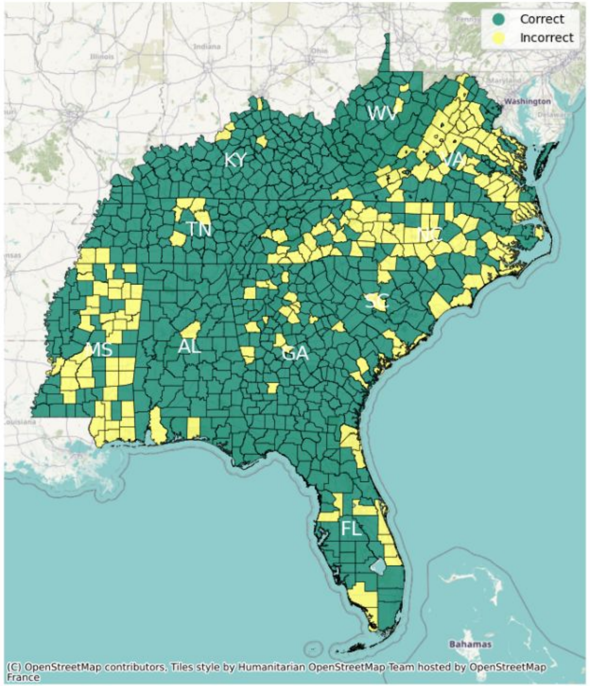
Average Number of Physically Unhealthy Days Prediction Map - Southeast



GIS Results - Machine Learning Accuracy for Physically Unhealthy Days

- Model predicts outcomes for 7/10 states fairly accurately
  - Distinguishes between rural, small urban and urban counties
- Does not predict well for three states: Virginia, North Carolina, and Mississippi
  - Very clear differences along these state lines

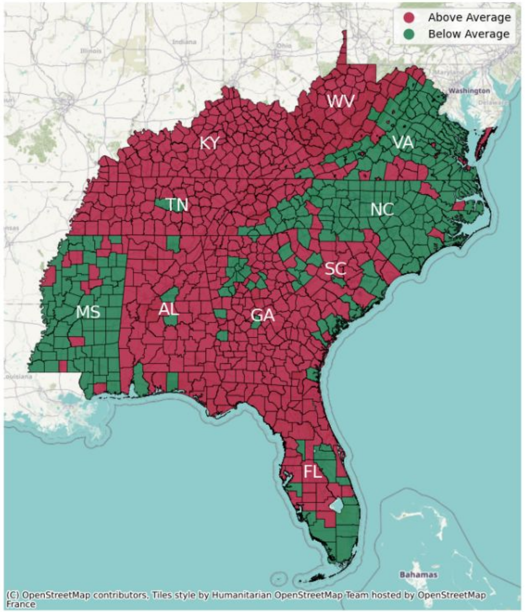
Average Number of Physically Unhealthy Days Accuracy Map - Southeast



GIS Results - Machine Learning Predictions for Mentally Unhealthy Days

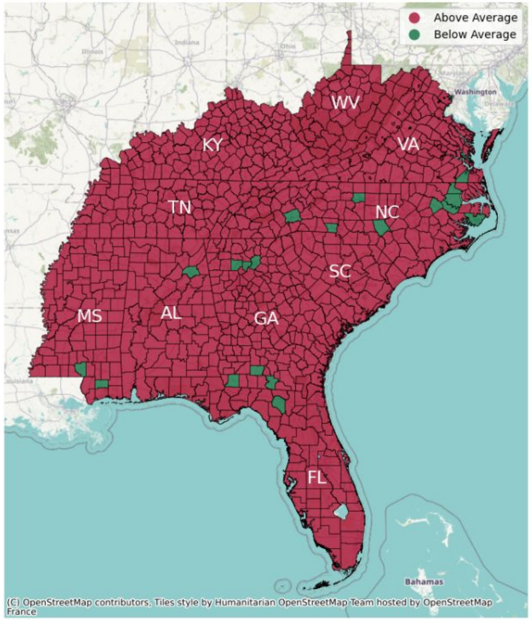
Actual

Average Number of Mentally Unhealthy Days Above National Average Map - Southeast



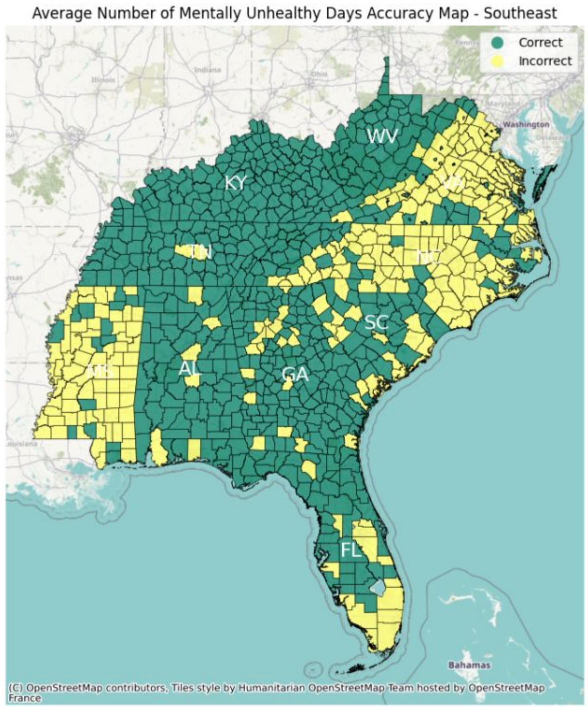
Predicted

Average Number of Mentally Unhealthy Days Prediction Map - Southeast



GIS Results - Machine Learning Accuracy for Mentally Unhealthy Days

- Model predicts fairly well, but not as accurate as physical health model
  - Clear discrepancies in Atlanta and Miami urban and metro (small urban) counties
- Predictions are worse for Virginia, North Carolina, and Mississippi than physical health model
  - Next, we would want to look at specific policies and factors attributed specifically to those states that accounts for this difference.



GIS Results – Machine Learning Accuracy Comparison

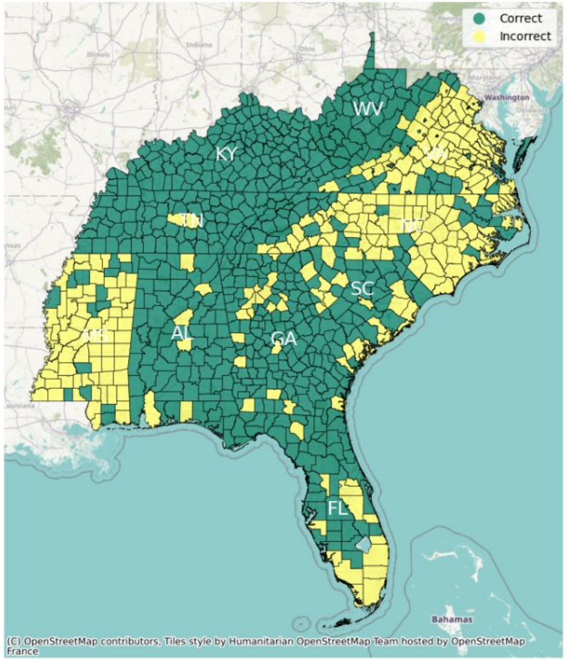
Accuracy

Rural
74%

Small Urban
58%

Urban
53%

Mentally Unhealthy Days



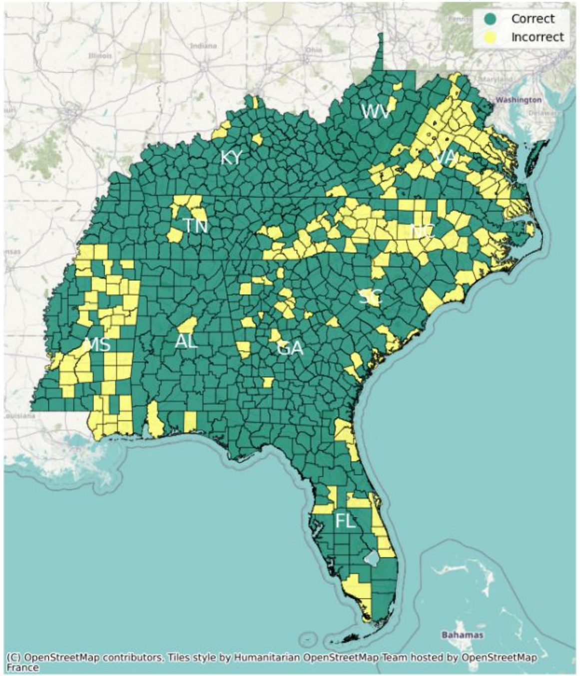
Accuracy

Rural
81%

Small Urban
61%

Urban
77%

Physically Unhealthy Days



Conclusion

Conclusion

- Urban communities tend to have higher climate risk scores than small urban and rural communities.
- Health outcomes in small urban communities are harder to predict than urban and rural in the chosen machine learning model.
- Climate risk score is statistically significant and is able to predict accurately in some scenarios, but overall it has room for improvement.
    - Including additional variables might help improve accuracy

Limitations

- Inherent error in the Physical and Mental Health metrics because it reflects someone's personal response to a survey question
  - People's answers may not be 100% accurate
- The calculation of the composite risk score
  - We briefly looked at the Expected Annual Loss (EAL), Social Vulnerability (SOVI), and Community Resilience (RESL) scores by themselves in relation to the health outcomes that showed clear patterns
    - SOVI score positively correlated with average number of mental health days

Future Considerations

- Why are certain states average mental unhealthy days worse than others and how does climate play into that?
- How does the score play into real life impacts? Are resources allocated based off of this score?
- What would be a better way to calculate the composite risk score? Does the resilience have too much of an impact by just division?
- What impact would adding temperature and elevation have on our models?
- How would comparison across regions affect our predictions and conclusions?