Predicting Flood Damages and Property Risks in the US

Team 152 – Allen Yen Chen, Yi-Chun Chen, Kevin Hsu, Parth Patel, David Harold Thompson

Motivation

Problem: Property owners lack reliable, personalized flood risk assessments for their specific properties.

Why It Matters:
- Climate change is increasing flood frequency and severity
- Government shifting flood damage liability to homeowners
- Property owners need accurate damage predictions for protection planning
- Current analyses only provide regional, not property-specific insights

Our Approach

Innovation: Property-specific flood damage prediction model combined with interactive risk visualization

Key Components:
- Statistical Models: Random Forest, Regression, Boosting, K-Nearest Neighbors
- Visualizations: State-level risk mapping, Integration with prediction model results, Geographic risk distribution visualization

Data

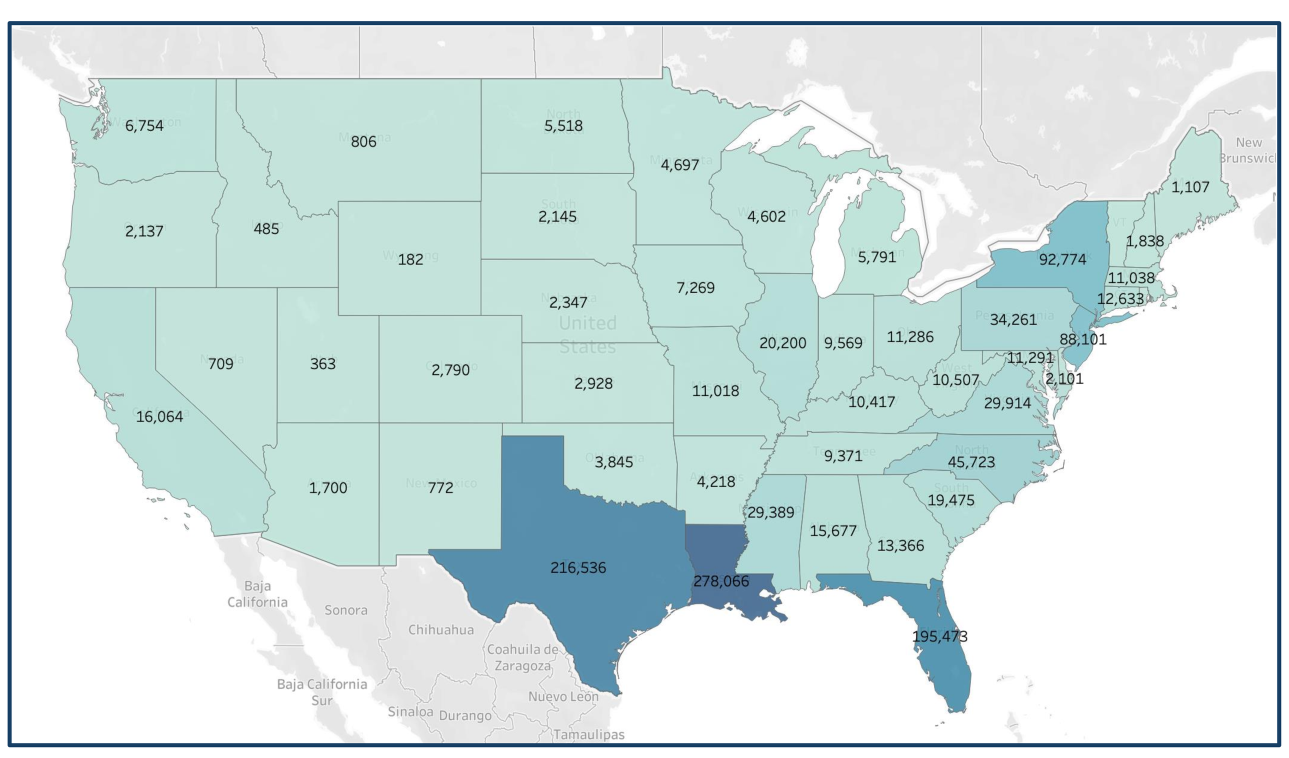
Source: National Flood Insurance Program (NFIP) claims
- Downloaded from FEMA.gov
- 996MB

Characteristics:
- 1.2 million claim transactions (residential)
- 73 features per property
- Coverage since 1979
- Data includes property features and actual flood damage amounts

Methods

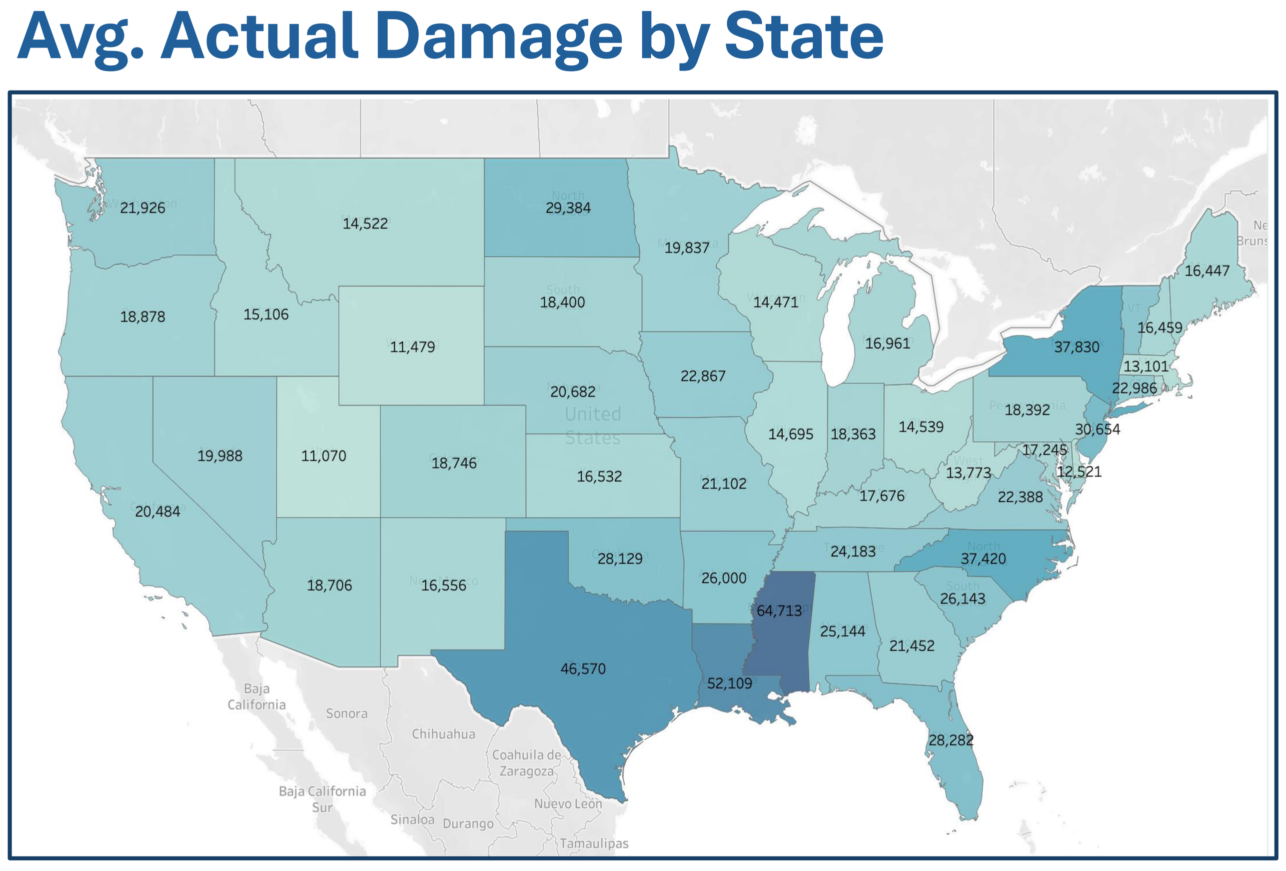
- Employed a variety of modeling approaches to predict potential flood damages on a property, and compared R^2, RMSE, MAE values to determine performance of each model.
- Excluded columns that are irrelevant to the model and performed a correlation analysis.
- Performed forward stepwise regression to decide on numbers of predictors = 36
- Divided data into a training set consisting of records prior to September 2019 and a testing set with records from September 2019 onward.

EDA



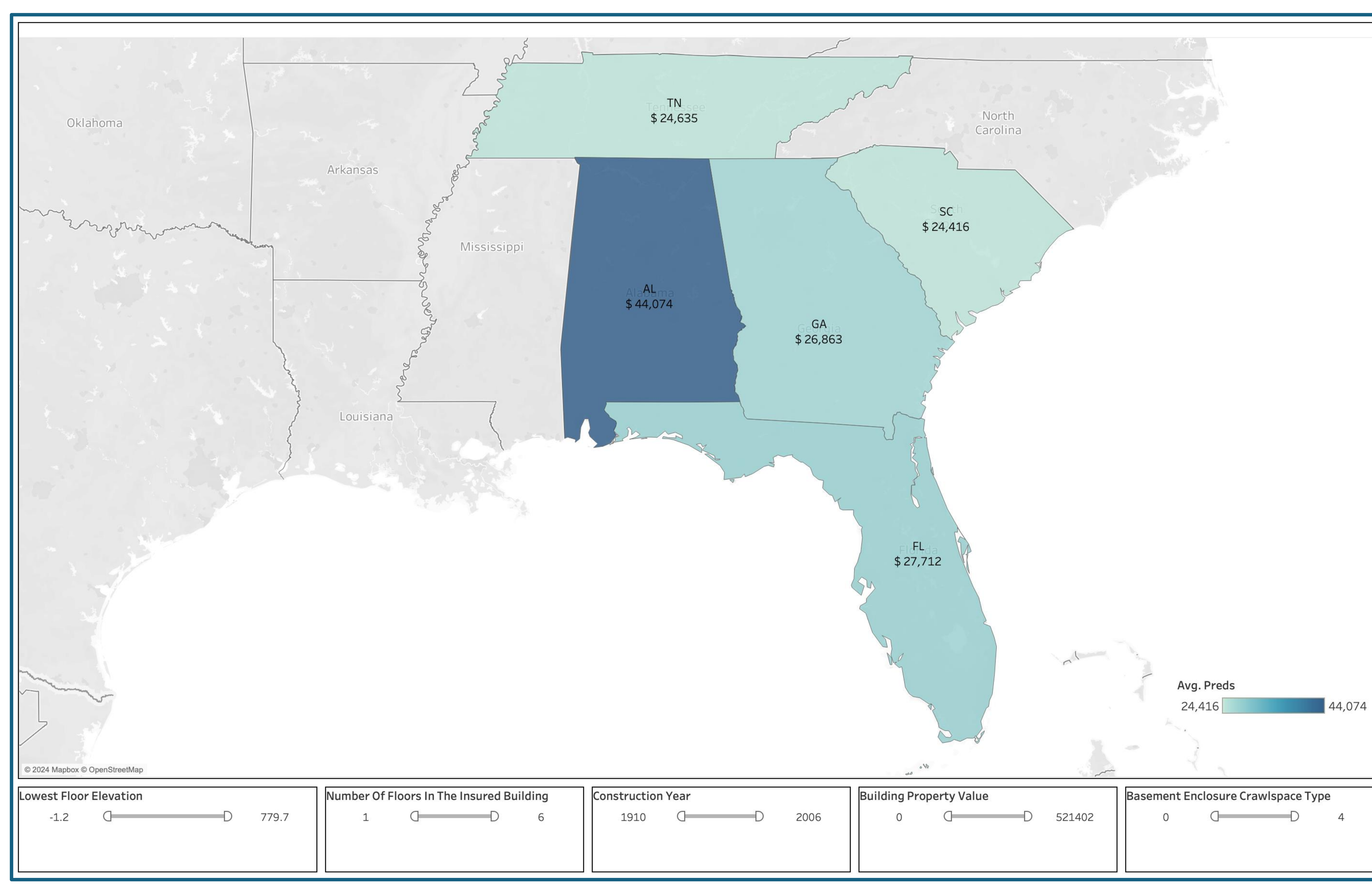
Southern states like Louisiana, Texas, and Florida experience more flood damage claims largely due to their frequent exposure to hurricanes.

Avg. Actual Damage by State



Actual damage amounts across the states vary. Mississippi's average is the highest at around $64,713.

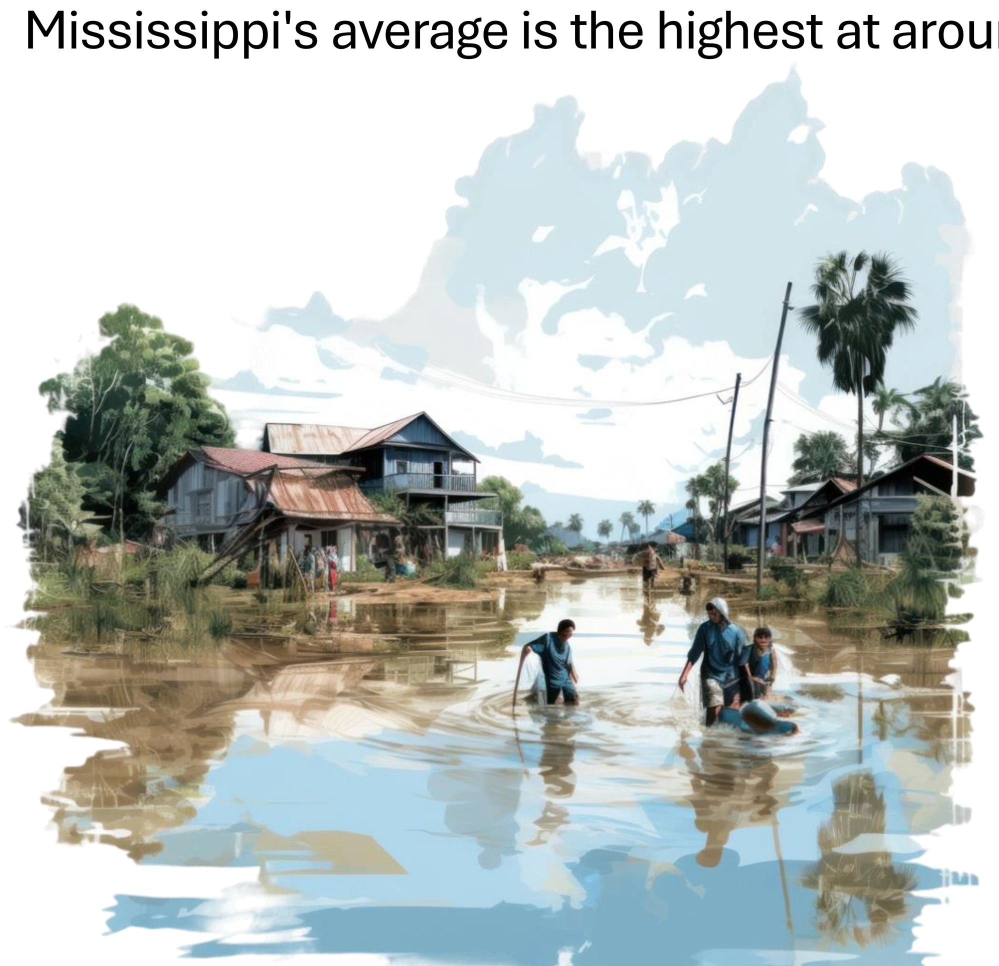
Average Predicted Damage by State (Example)



Results

Model	R^2	RMSE	MAE
Boosting	0.32	42,571	13,316
Random Forest	0.30	39,293	15,826
Linear Regression	-67.01	414,503	50,810
KNN (k=10)	-189.94	694,555	37,637
KNN (k=15)	-236.39	774,427	44,515
KNN (k=5)	-278.25	839,939	36,485

Most important features (Random Forest & Boosting)
Building Property Value
Longitude
Year of Loss
Latitude
Building Replacement Cost
Construction Year



Key Takeaways

- More recent data tends to provide better predictions for the damage amount.
- k-NN performed poorly because it calculates its closest neighbors are all from the training set, which is pre-09/2019 data, which could be outdated.
- Boosting and random forest have similar results, but random forest is far more resource demanding
- Southern states should implement a more effective flood management system.

Next Steps

- Explore the performance of the model using just a subset of variables
- Adjust damage amount for inflation before model training and testing