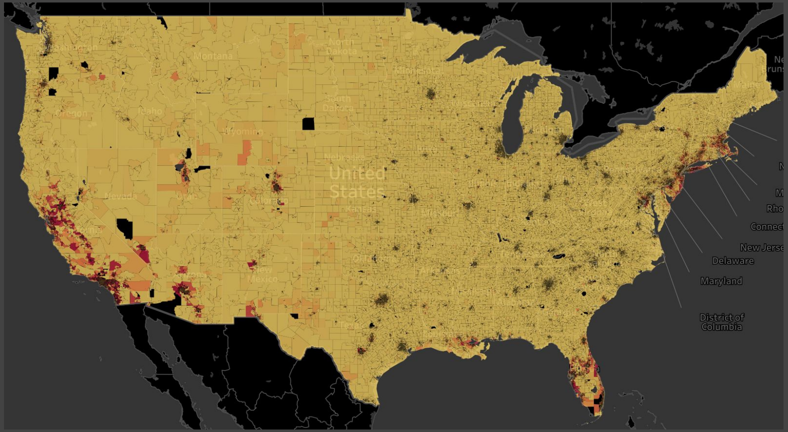
Predicting Solar Panel Adoption



MIDS Capstone - Final Presentation

Team: Laura Williams, Noah Levy, Gabriel Hudson

Mitigating Climate Change

Distributed solar panel electricity generation

- Reduces carbon emissions and reduces impacts of climate change
- Creates more diversity and resilience in electricity production
- Reduces demand on electrical grid but is more complex to integrate
- Requires understanding of complex customer adoption trends

Distributed Residential Solar Panel Adoption

- Identify geographic areas with high predictive factors but low saturation of residential solar panels
- Analyze factors contributing to residential solar panel adoption

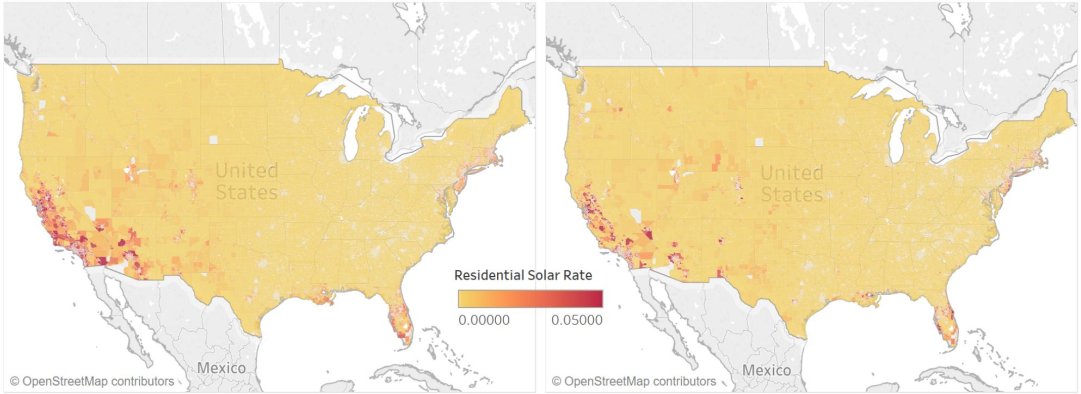
User Interface Demo

http://people.ischool.berkeley.edu/~hudsonga/

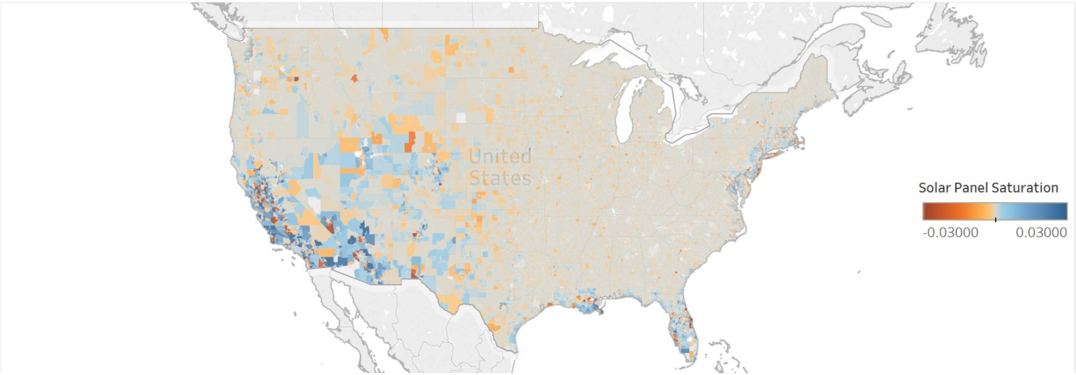
Backup Screenshots

Predicted Solar Adoption Rate

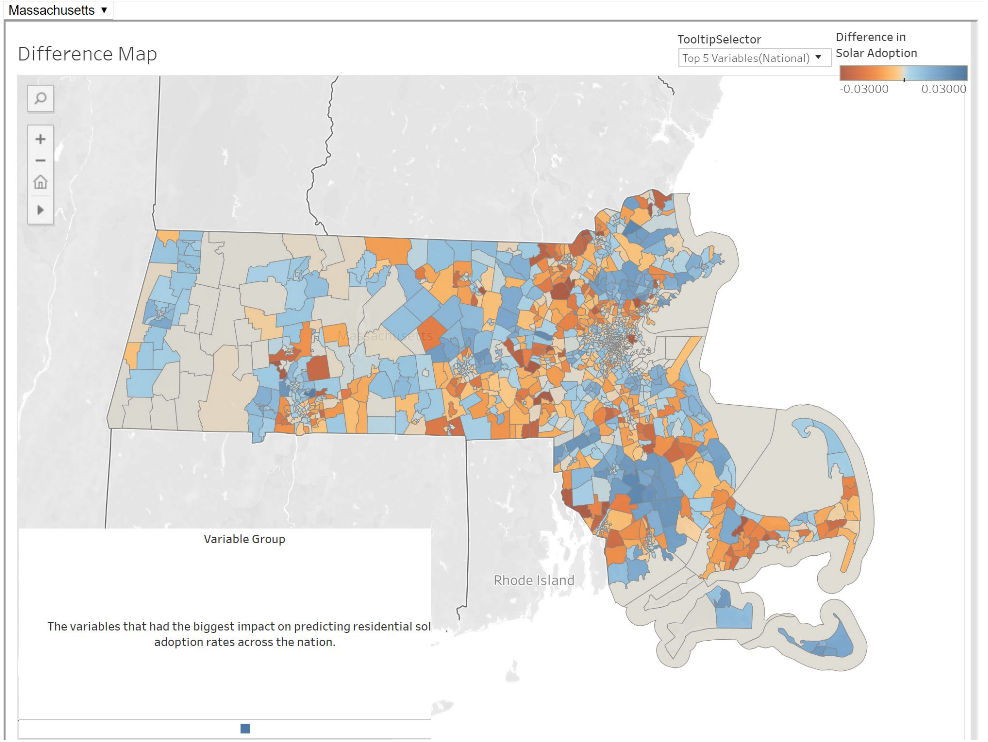
Actual Solar Adoption Rate

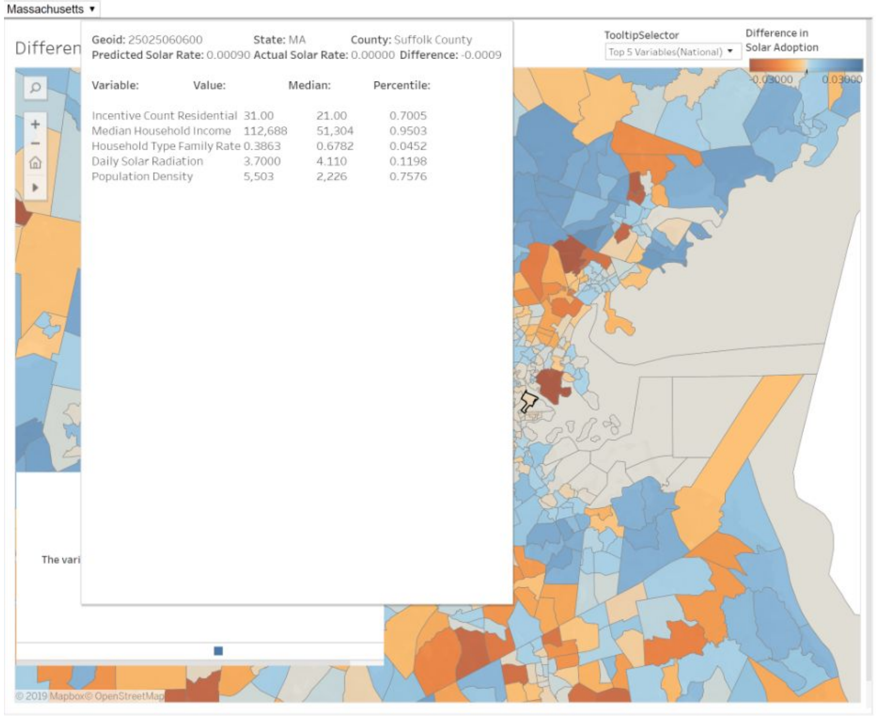


Difference Map

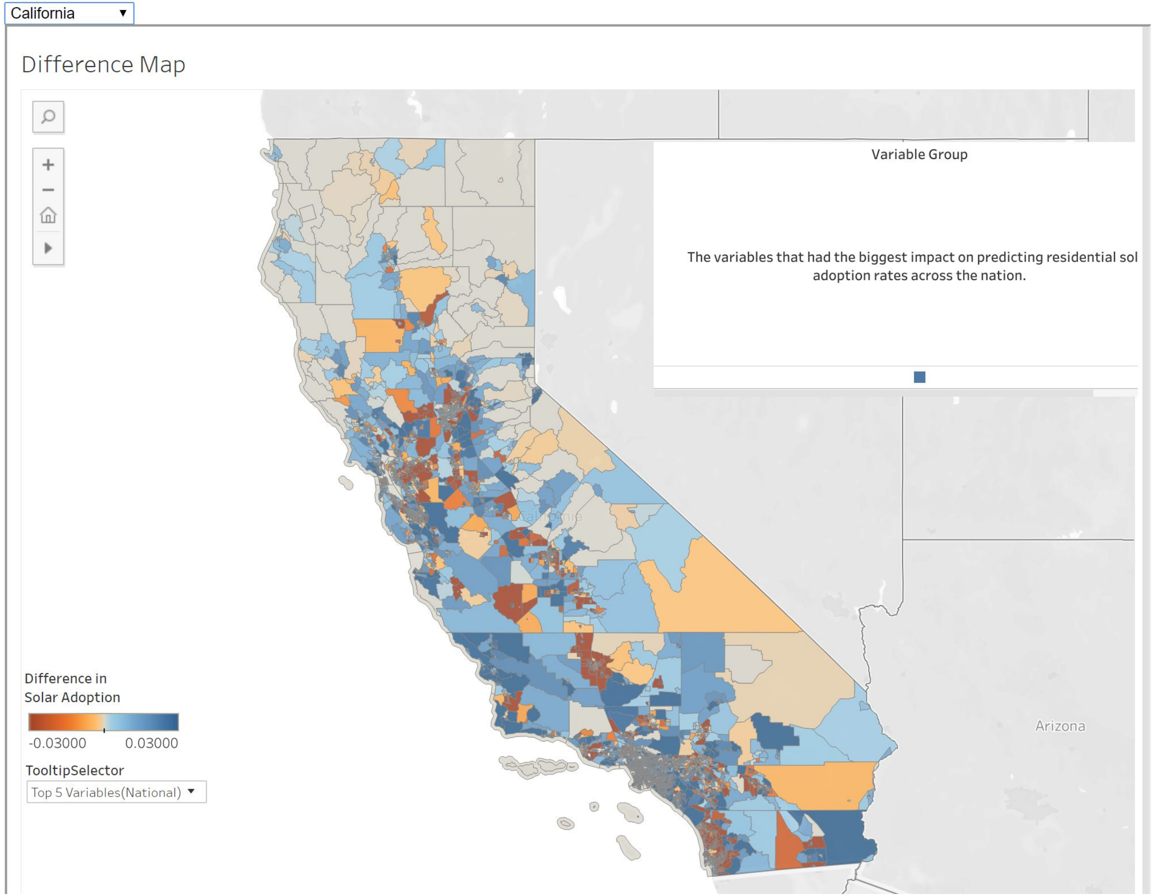


Backup Screenshots

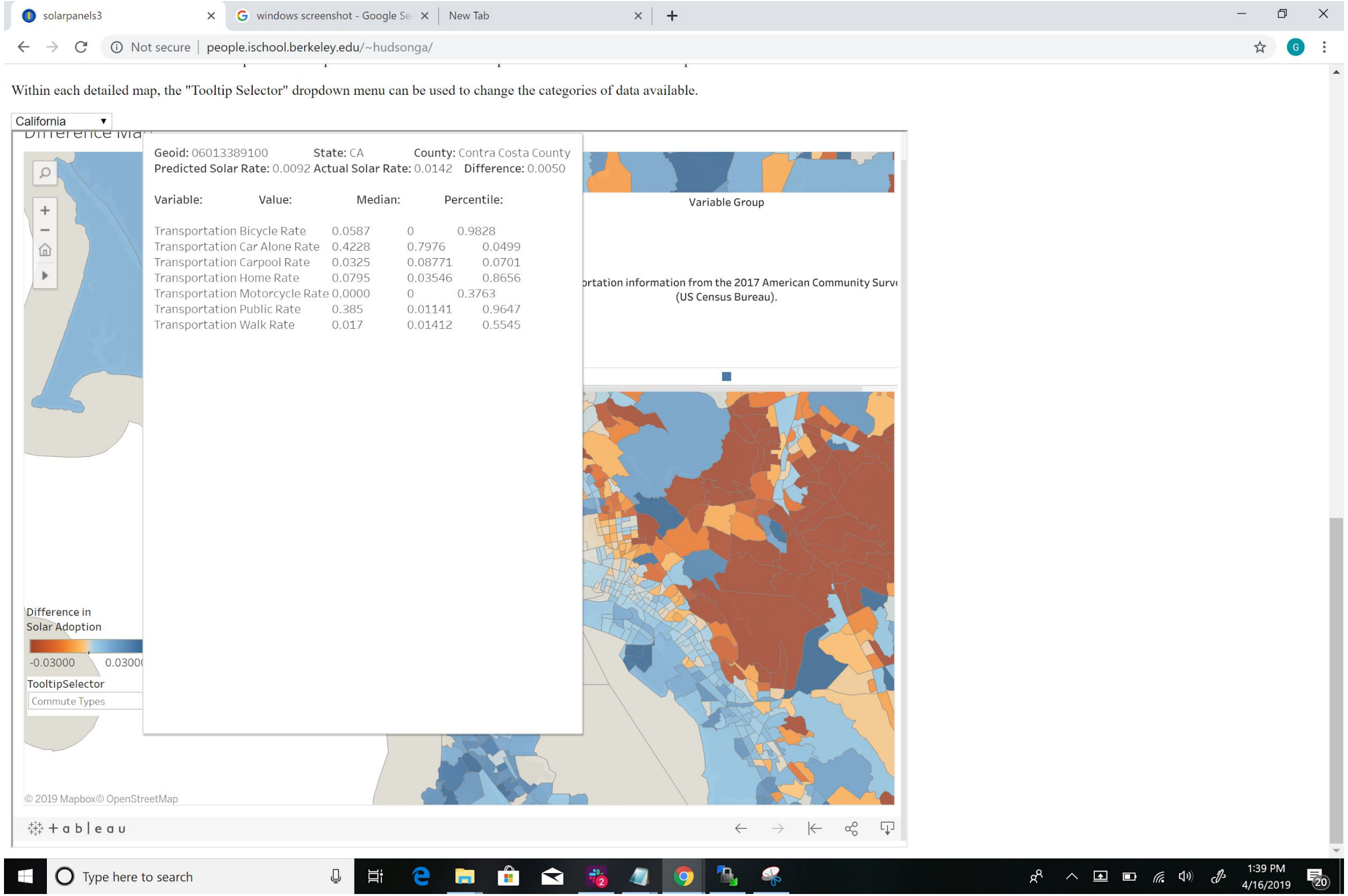




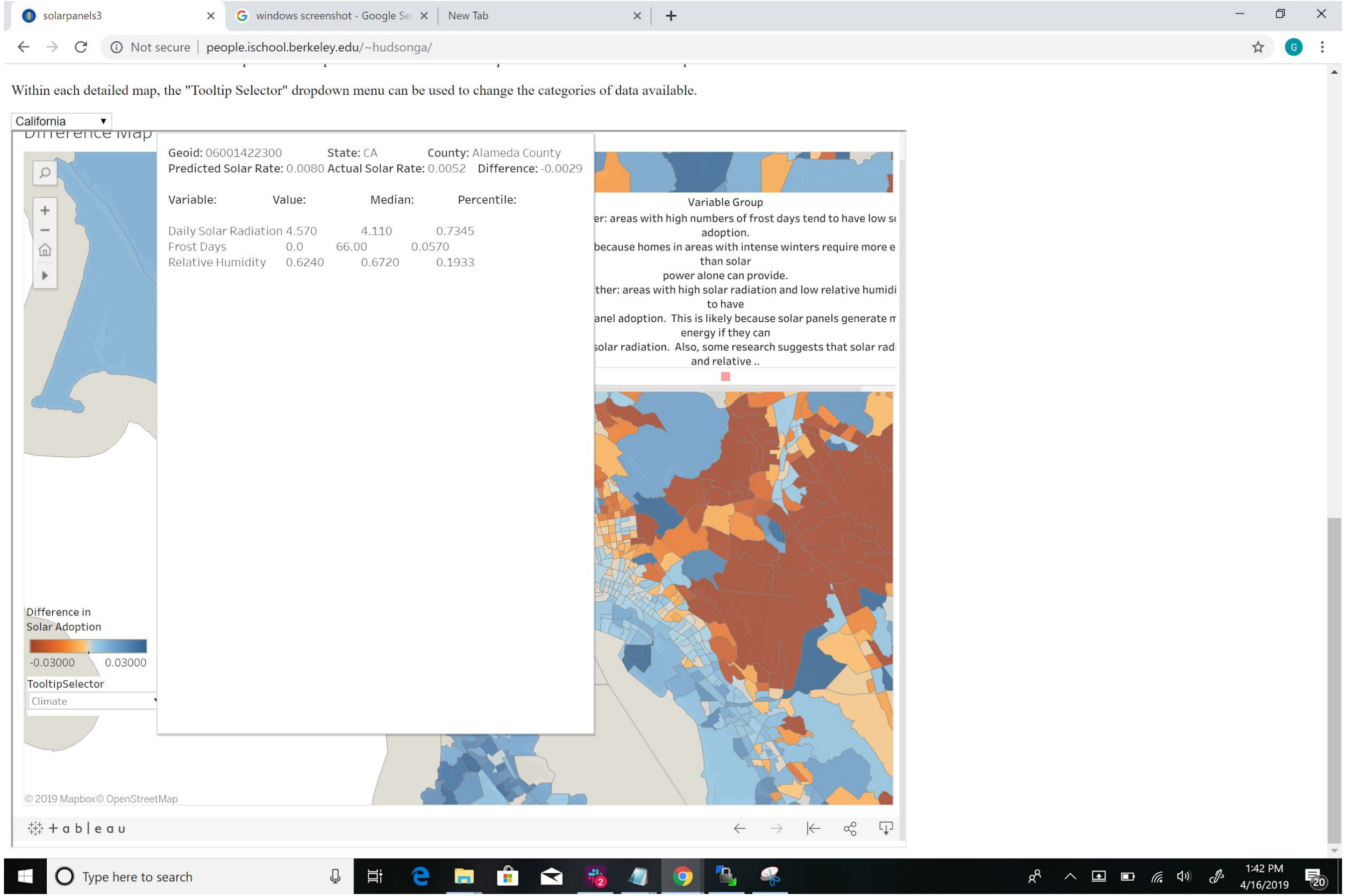
Backup Screenshots



Backup Screenshots



Backup Screenshots



Current Publicly Available State of the Art

Data

- Stanford DeepSolar team neural network: 1.47 million U.S. solar panel systems
- Additional variables, *i.e.*, census data from 2015 and weather data from NASA

DeepSolar Predictive Model

- Random Forest
- R^2 0.722

Table 1. Comparison of the Cross-Validation R^2 Value of Different Solar Deployment Predictive Models

Model	Cross-Validation R^2
LR (quadratic + interaction)	0.181
MARS	0.267
RF regressor	0.412
RF classifier + LR (quadratic + interaction)	0.643
RF classifier + MARS	0.592
SolarForest (RF classifier + RF regressor)	0.722
SolarNN (Feedforward neural network)	0.717

Ten-fold cross-validation is carried out utilizing the census tract data. LR, linear regression; MARS, multivariate adaptive regression splines; RF, random forest. Hierarchical SolarForest proposed in the paper was the best-performing model.

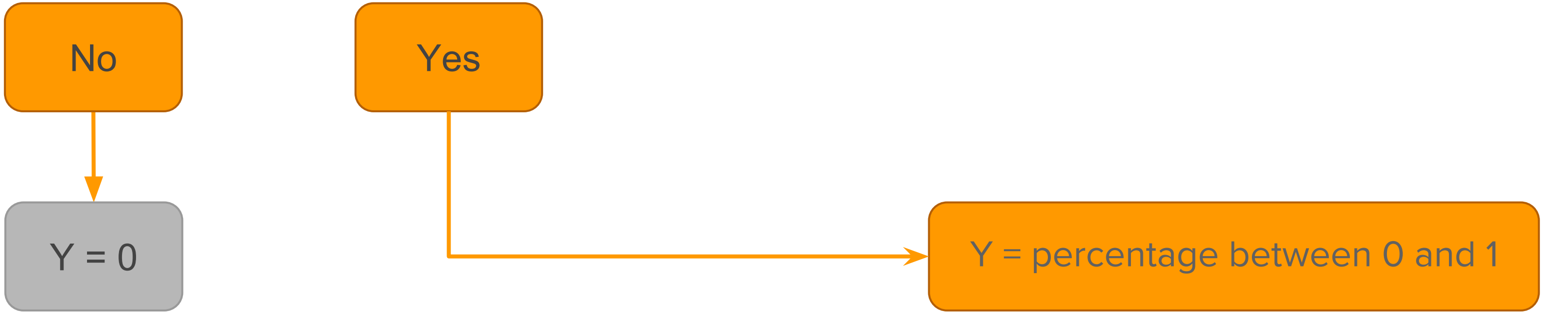
Model Structure

Stage 1: Random Forest Binary Classifier

Outcome variable: Does a census tract have at least one solar panel system?

Stage 2: Random Forest Regressor

Outcome variable: Density of residential solar panel systems per census tract

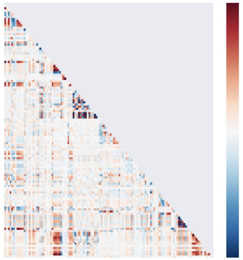


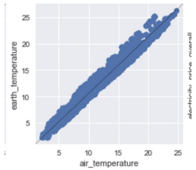
Feature engineering

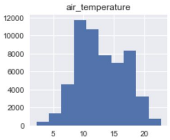
- **Outcome variable**
    - Original: Number of residential solar panel systems per household per census tract
    - Multiplied by rate of owner occupancy to approximate density by residential structure
- **Removed data points** (98.3% of all census tracts retained)
    - Population count under 100
    - Household count under 100
    - Water percentage over 75%
- **Replaced missing values with county medians**

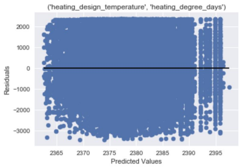
Feature engineering (cont'd)

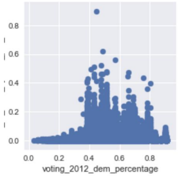
- **Remove redundant variables (reduced to 58 variables)**
    - Unused outcome variables
    - Variables used to calculate other variables
    - Variables not explicitly residential (non-residential electricity prices)
    - Condensed total/land/water area into water percent variable
    - Variables whose coefficients were reduced to zero in a linear regression model with L1 regularization and gradient descent
    - Highly correlated variables (weather and political variables)











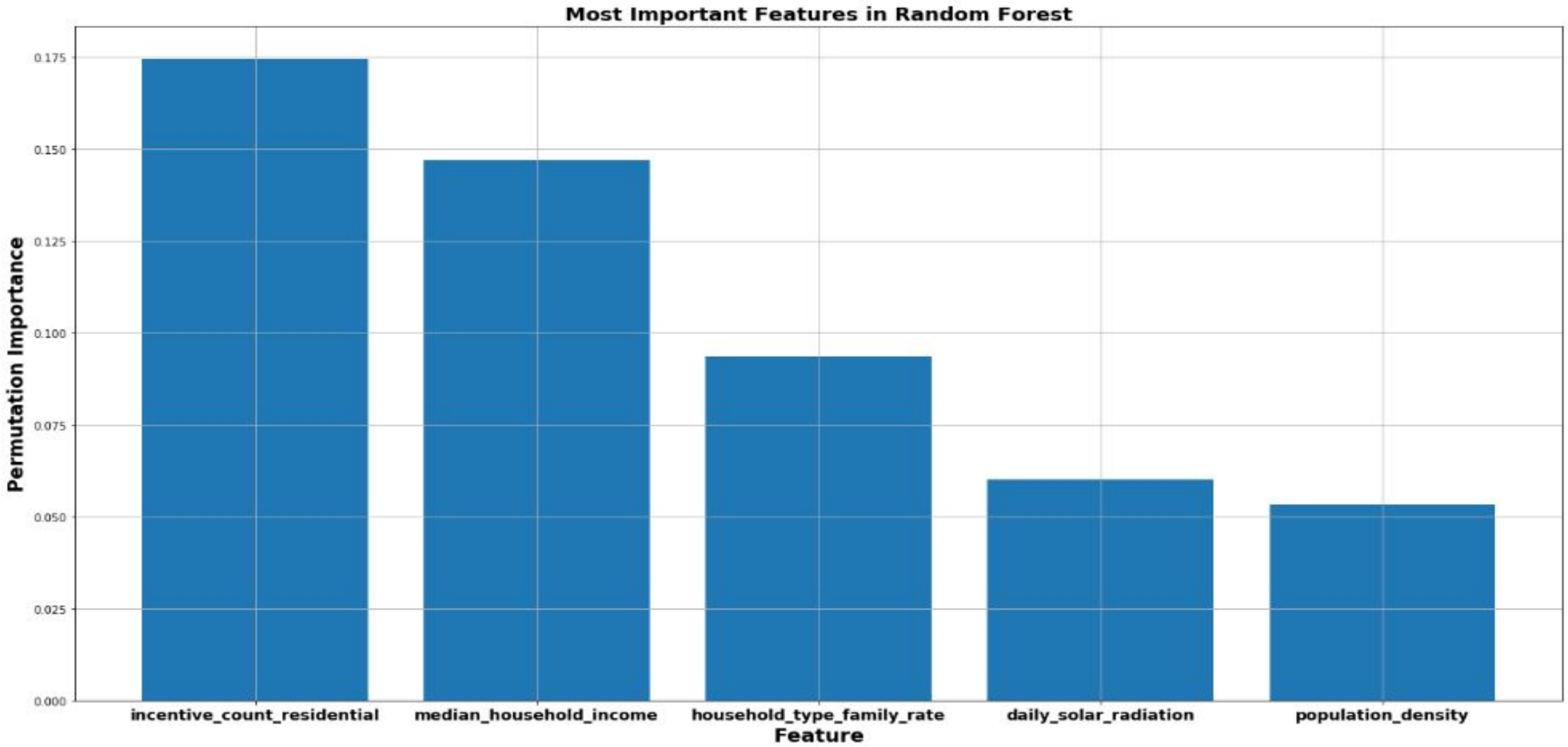
Model performance

	Dependent variable	Stage 1 Classifier	Stage 2 Regressor	Combined R^2 (10-fold CV)
DeepSolar	Residential solar panel systems per household	Not available	Not available	0.722
Our model	Residential solar panel systems per household	0.813	0.714	0.716
Our model - new outcome variable	Residential solar panel systems per owner-occupied housing	0.812	0.744	0.744

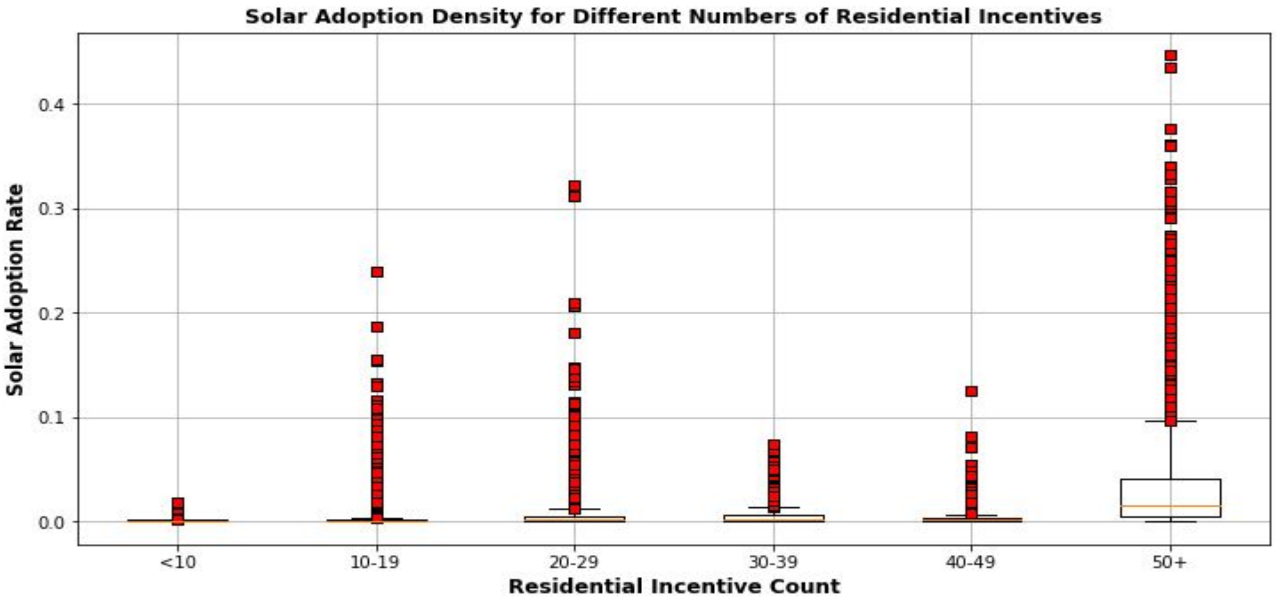
Our model achieves improved R^2 compared to DeepSolar's model using our modified outcome variable with just our regressor instead of their full ensemble. We also used far fewer features for a more parsimonious model.

We tuned hyperparameters separately for both the classifier and the regressor, testing for overfitting by comparing predictions on the training set vs. the test set.

Most Important Features in RF Regressor

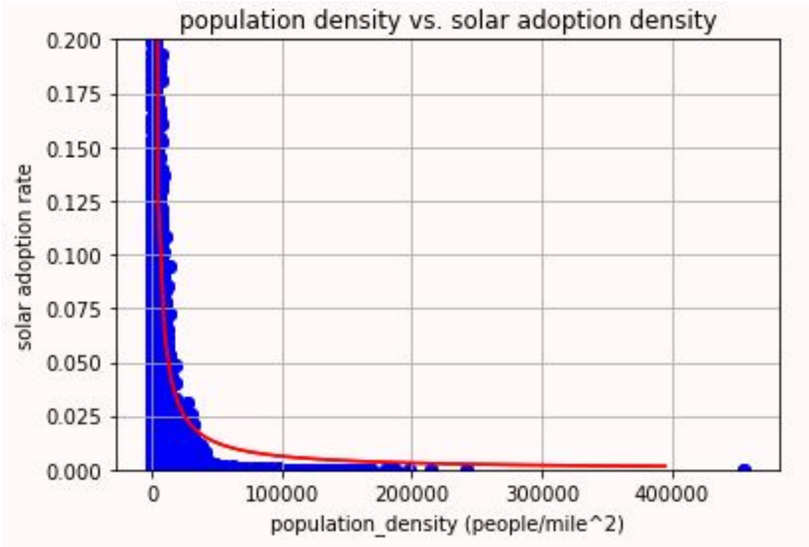


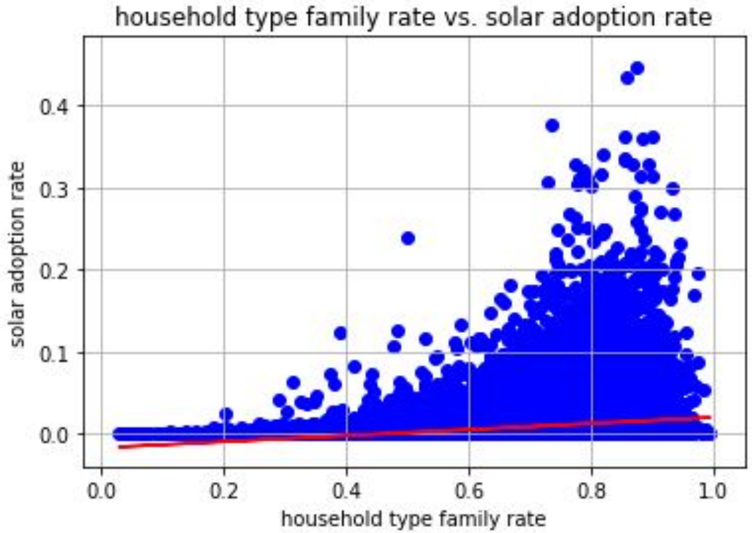
Solar Incentives Drive Solar Adoption



- Census tracts with more local and statewide incentives tend to have higher rates of solar panel adoption
- The effect is uneven; districts with 30-49 incentives seem to have lower adoption rates than districts with 10-29 incentives
- More thorough analysis would be possible with data on specific incentives rather than just counts

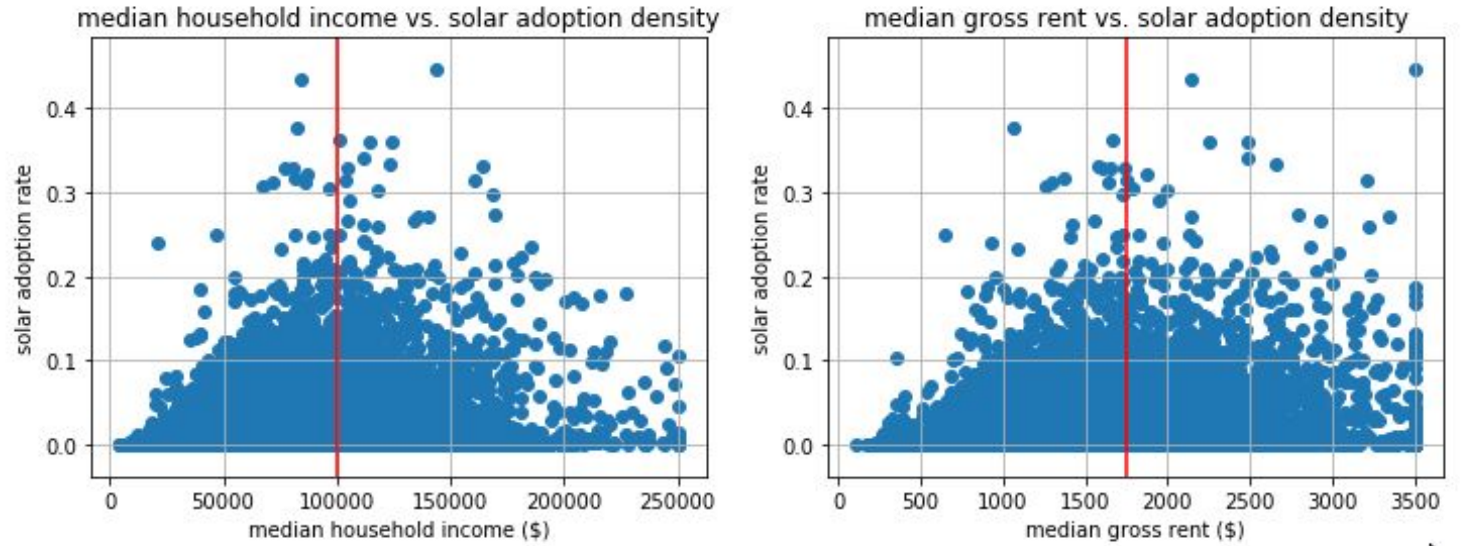
Urban Areas Have Lower Solar Adoption Rates





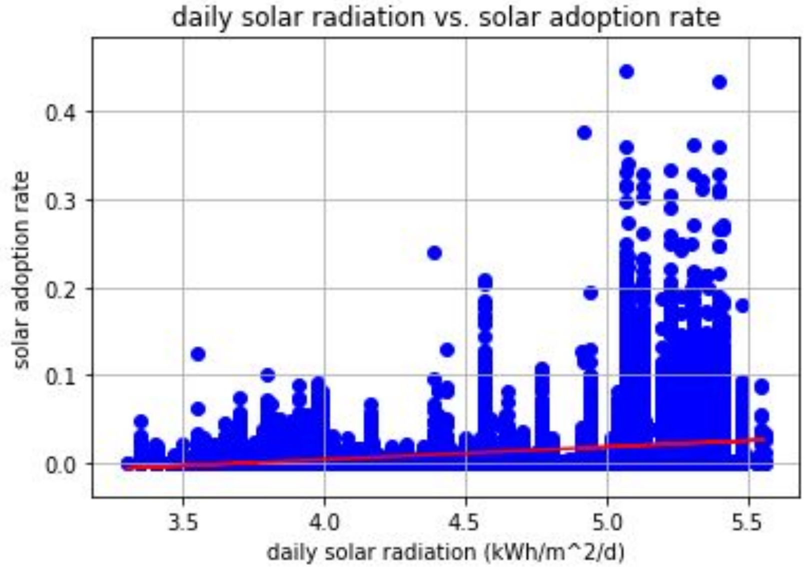
- Urban areas tend to lack available rooftop space
- Buildings in these areas tend to absorb solar radiation on their facades rather than their roofs

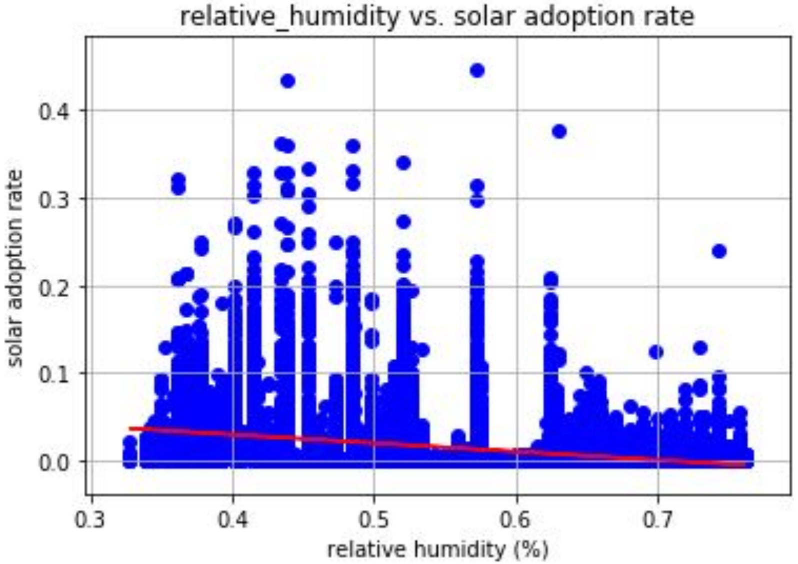
Solar Adoption Increases with Prosperity, up to a Point



- The solar adoption rate actually starts to decrease as median income increases past $100,000.00

Sunny, Dry Areas Have the Highest Adoption Rate





- One potential explanation for the inverse relationship between humidity and solar adoption is that humidity correlates with fewer sunny days

Market and industry feedback

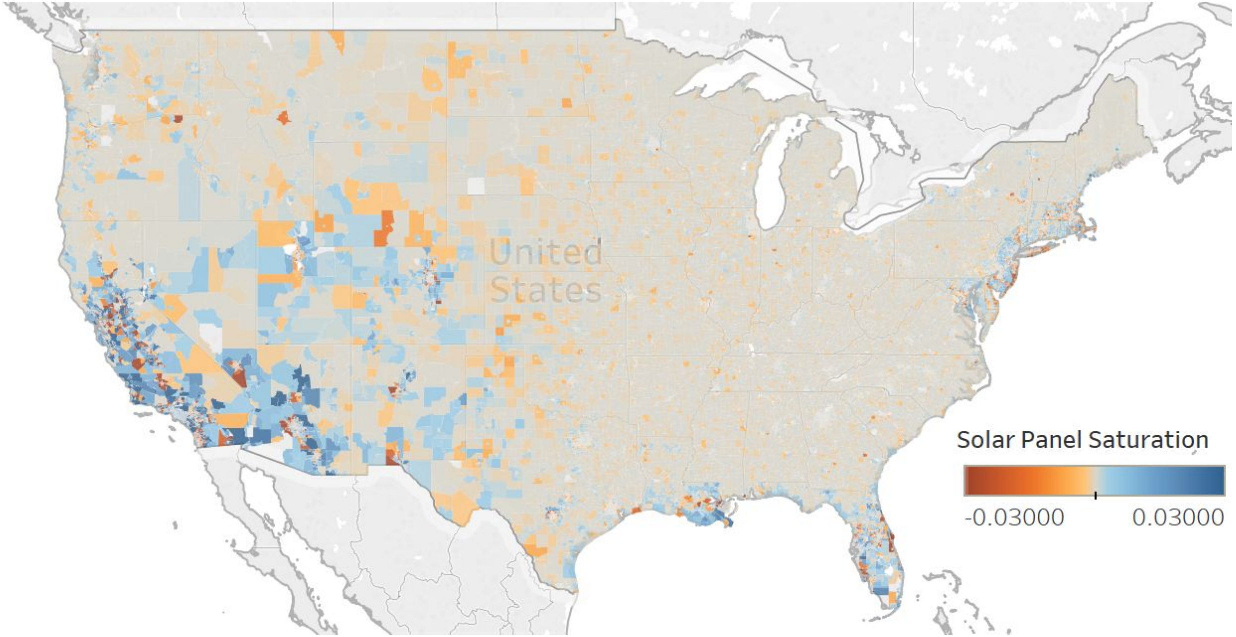
- Residential Solar Market expected to reach ~$23 Billion by 2025
- Solar panel installer one of fastest growing jobs in U.S.
- Industry feedback
  - We connected with NextEra Analytics, Florida Solar Energy Center, NREL, and Oregon Department of Energy
  - This type of prediction is very much needed and is in process at larger business and agencies (i.e., NREL) but is not currently publicly available.
  - Understanding predictive factors is key
  - Policy makers want to understand role of incentives
  - Solar panel businesses want to understand role of business models
  - UI may need to be tailored for different audiences, depending on statistical knowledge

Future opportunities for growth

- Detailed business model and incentive data could be collected or possibly purchased from vendors.
- Partition dataset on buckets of one or two importance feature values (i.e., daily solar radiation or incentives) and explore changes in important features.
- Represent time factors in the dataset - how long have certain incentives or other factors have been in place.
- Collect data on residential structures per census tract to improve outcome variable.
- Removing latitude and longitude from our dataset reduces model performance, which tells us some important data particular to geographic areas is not in the dataset.
- Expand visualization capability with commercial resources.

Code

Github repo: [Predicting Solar Panel Adoption]()



Citations

Abdullahi, Ayegba S, et al. Impacts of Relative Humidity and Mean Air Temperature on Global Solar Radiations of Ikeja, Lagos, Nigeria. International Journal of Scientific and Research Publication, 2 Feb. 2017, www.ijsrp.org/research-paper-0217.php?rp=P626216.

Stettiner, Sam, et al. Urban U.S. Solar Electric Usage and Population Density. Apr. 2017, static1.squarespace.com/static/56b25eb9cf80a1861cb737c8/t/5994aaa6e58c62554404bf62/1502915238902/Urban+U.S.+Per+Capita+PV+Usage+-+Population+Density+4-27-17.docx.

Yu, J. et al (2018). DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States, *Joule 2*, 2605–2617, December 19, 2018, *https://doi.org/10.1016/j.joule.2018.11.021*

Yu, J. et al. The DeepSolar Project. *DeepSolar by Stanford.* 2018. *http://web.stanford.edu/group/deepsolar/home*