Empowering targeted tenant organizing: geographic forecasting of housing insecurity

Amy Gottsegen, Dr. Jake Williams - Drexel University, College of Computing and Informatics

Introduction

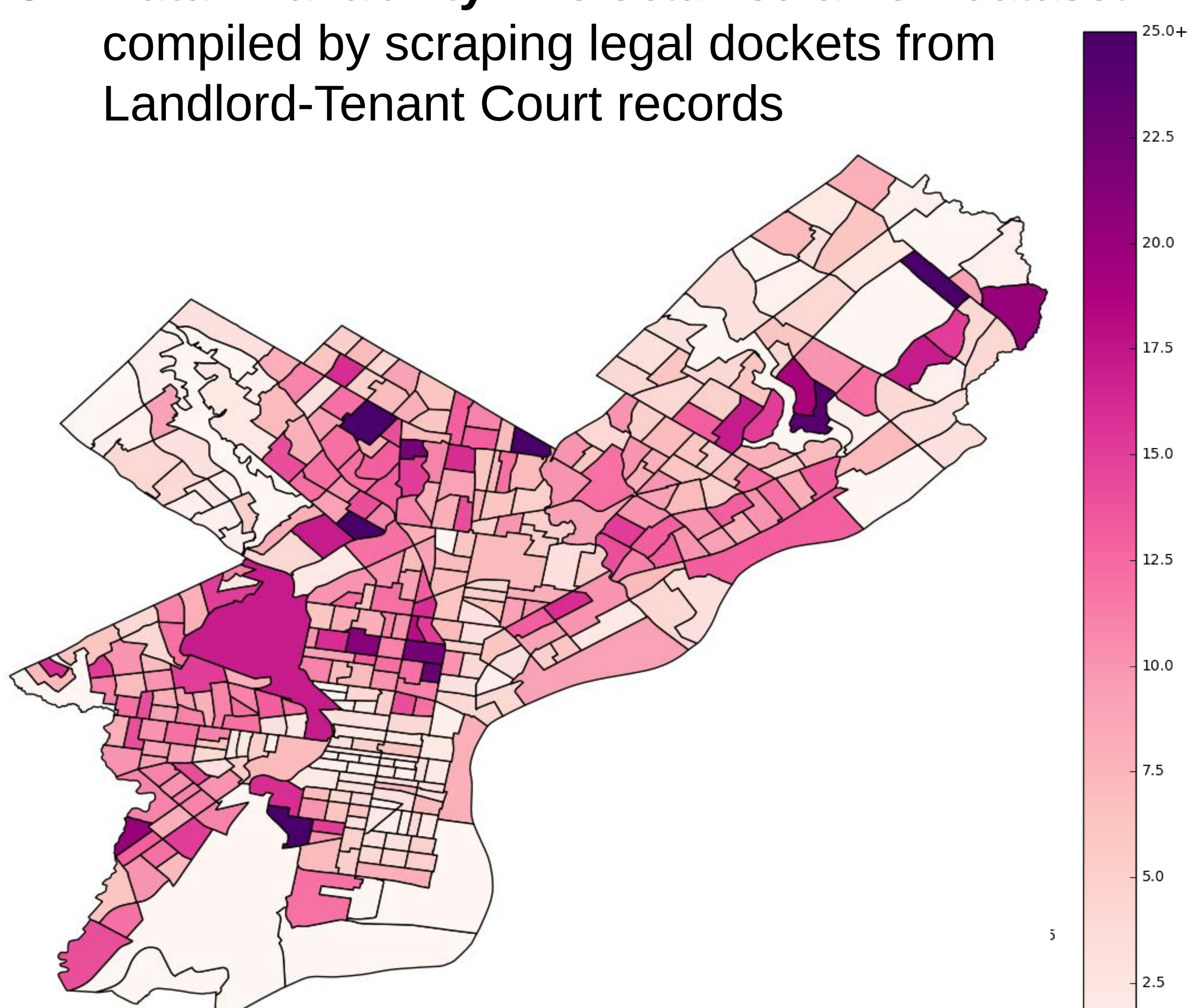
The purpose of this research is to explore how data analysis can support tenant organizing in the mode of the Philadelphia Tenants Union. The PTU seeks to ensure safe, decent, and affordable housing for all using the following model:

PTU organizers lack a methodical approach to step 1, relying mainly on word-of-mouth, which creates a bottleneck in their work.

Defining Housing Insecurity

We use evictions as a heuristic for housing insecurity for three reasons:

1. **PTU Experience**: Threat of eviction is always one of tenants' deepest concerns
2. **Support in Research**: Sociologist Matthew Desmond's 2016 book *Evicted* found eviction is a primary contributor to housing insecurity, and to poverty more generally
3. **Data Availability:** We obtained a new dataset compiled by scraping legal dockets from Landlord-Tenant Court records



Eviction rate by census tract in 2016. Data source: ACS 2016 5 year estimates

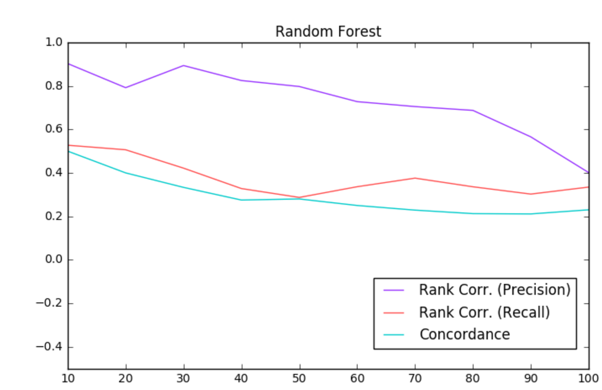
Identifying Features

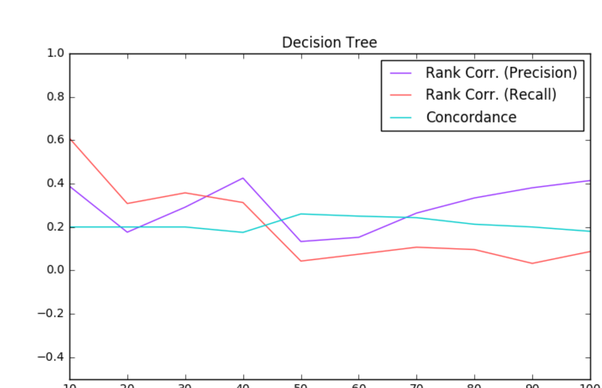
1. **Location**
   Expressed as coordinates
2. **Floor Area Ratio = Building Area / Lot Area**
   Higher values indicate dense urban construction, such as apartment buildings
3. **Assessed Value**
   Normalized over the interior square footage
4. **Owned Units**
   If more than one owned unit at an address, indicates condominiums
5. **Permits**
   PTU organizers often speak of facility improvements immediately preceding evictions

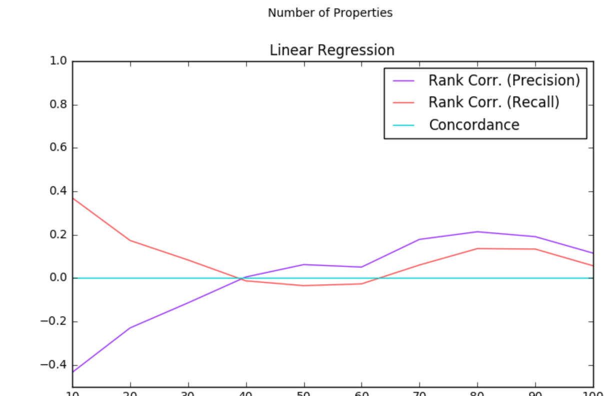
Model Selection

Because canvassing new buildings is a large resource investment, we evaluate models based on their performance on the top *n* properties rated highest-risk. We ultimately chose *n* = 30, about double the number of properties they could canvass in a year, allowing some suggestions could be discarded for limited access via public transit, etc. Furthermore, because the PTU is concerned only with risk *relative* to other properties, we use the Spearman ranked correlation rather than a linear correlation. We thus developed three evaluation metrics:

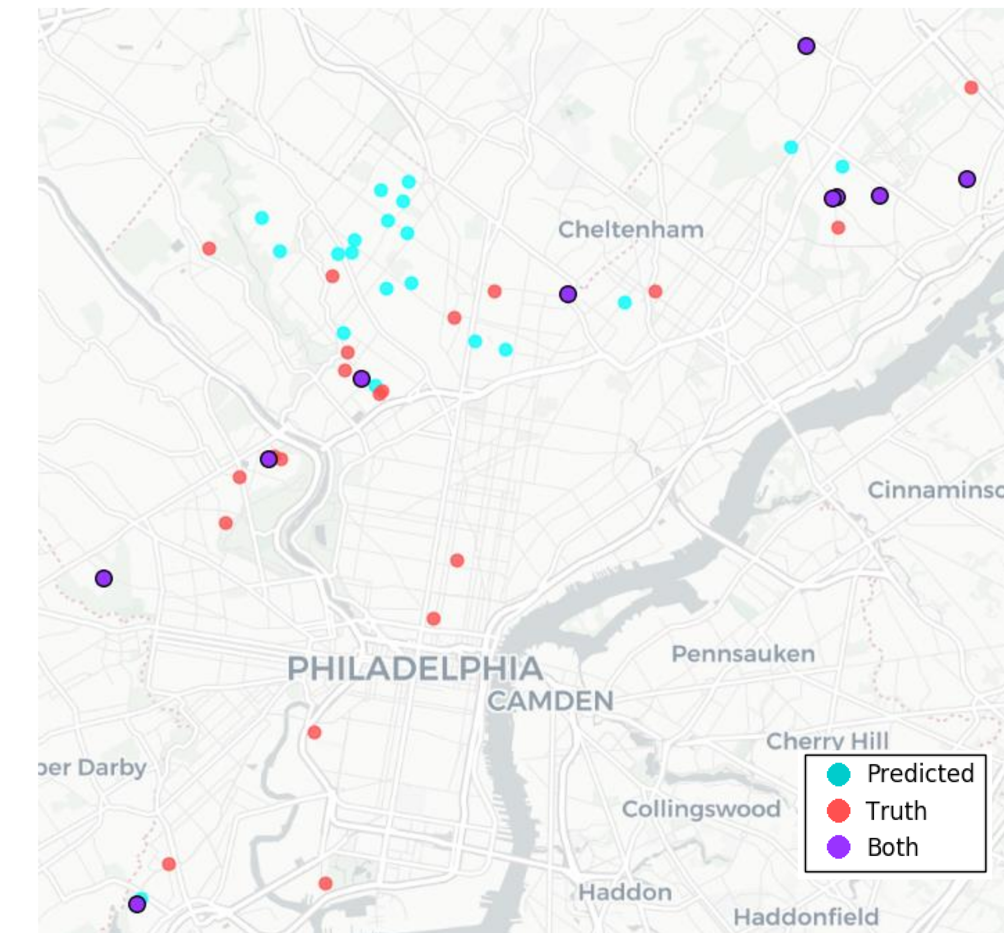
1. **Top *n* Ranked Correlation (ρ_P)**
   The ranked correlation when sorted by predicted evictions. This is a metric of precision.
2. **Top *n* Ranked Correlation (ρ_R)**
   The ranked correlation when values are sorted by true evictions. This is a metric of recall.
3. **Top *n* Concordance (α)**
   Portion of top *n* predictions that are top *n* truths







Output



The random forest model had the best performance of the three, identifying the 5 properties with most evictions

ρ_P	ρ_R	α
0.89	0.42	0.33
0.00	0.02	-

Future Work

- **Constructing a Housing Insecurity Index:** a metric which, in addition to evictions, includes other indicators of overall economic precarity may provide a more robust tool
- **Improved Model:** implementing a model such as a Mixed Effects Random Forest may improve performance by treating year as a fixed effect
- **Further Feature Engineering:** Several features, such as code violations and demographic data, will be assessed as possible features
- **Finer Timescale:** we are building a continuously updating dataset of related social media data, which may be able to provide predications with better temporal resolution

Acknowledgements

Thanks to Lauren Parker at the Philadelphia Mayor's Office of Community Empowerment and Randy Sargent at the CREATE Lab for their collaboration on this project, to Jonathan Pyle at Philadelphia Legal Aid for compiling the evictions data, and to the Steinbright Corporate Partners Fund for funding this work.