

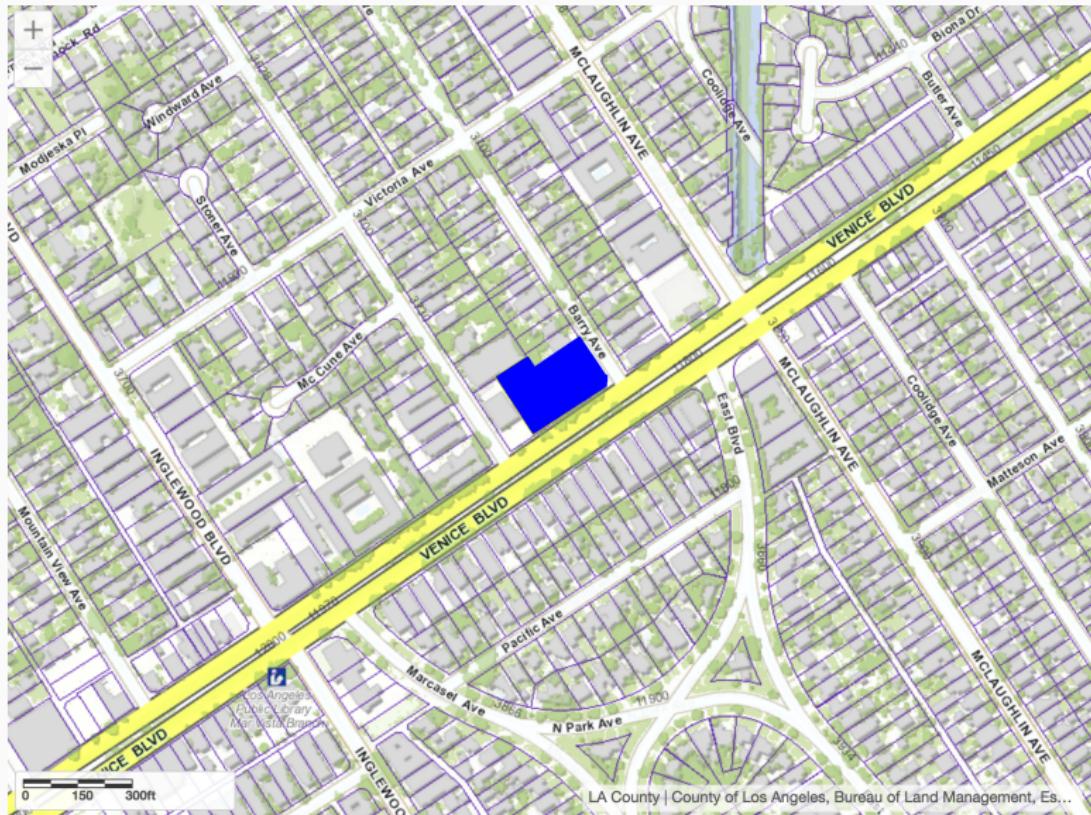
Balancing Data Sufficiency and Privacy

Medha Uppala

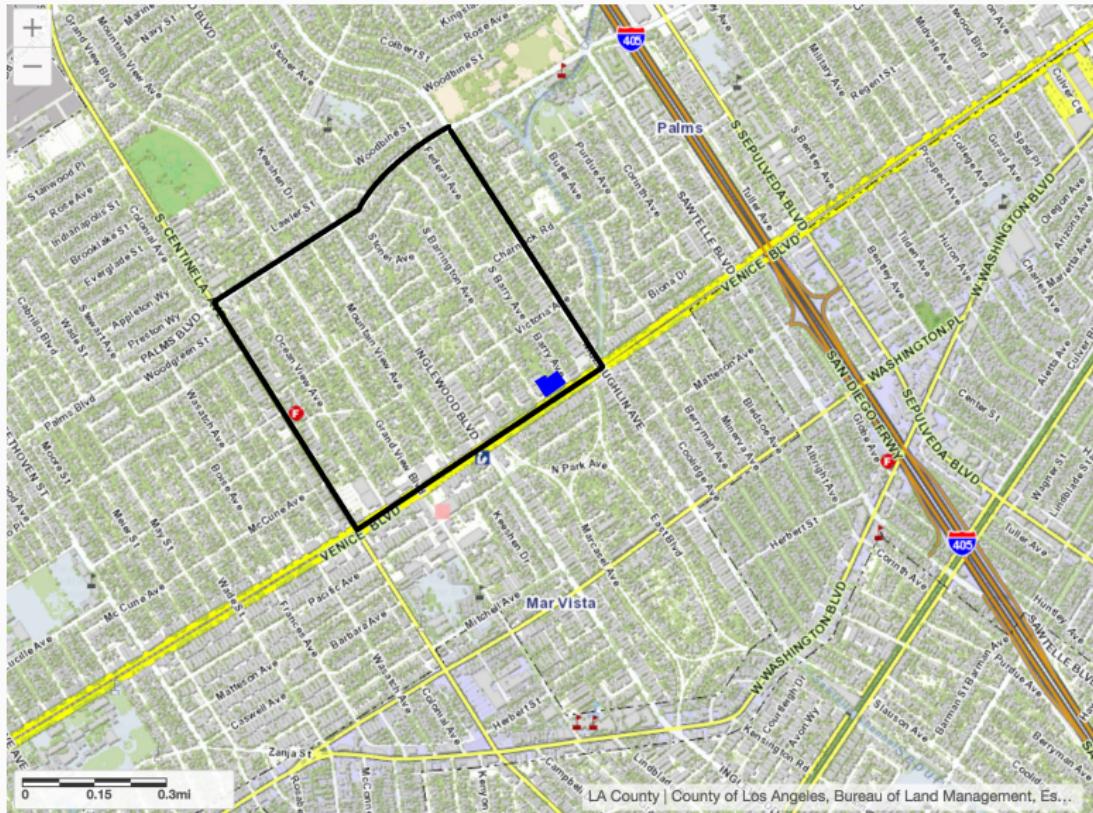
umedha@ucla.edu

February 11, 2019

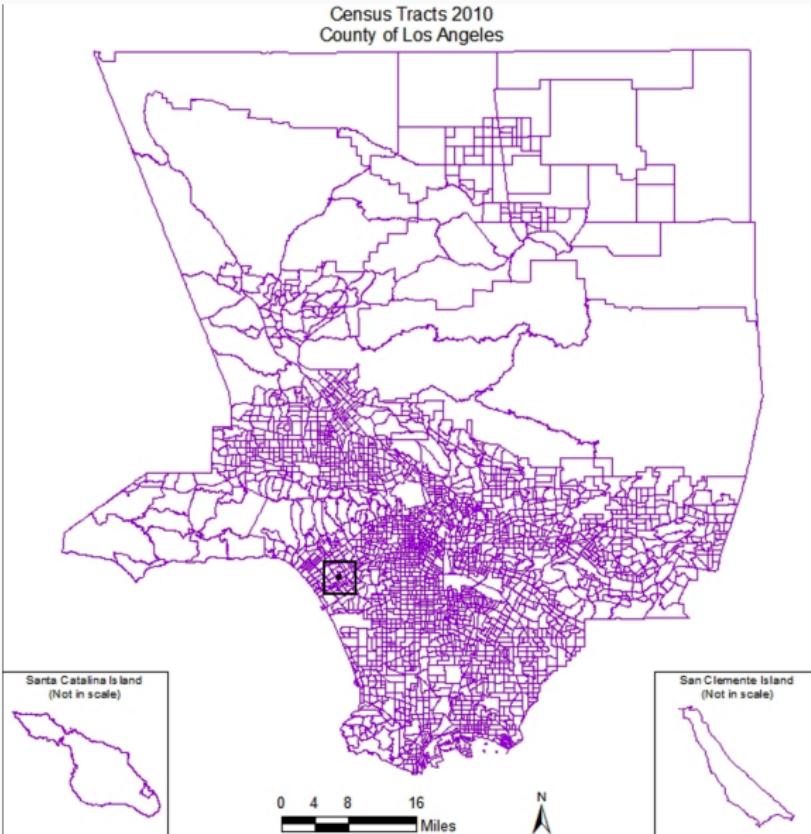
Energy Atlas data resolution



Census tract 2719.02



Census tracts in LA county



Background

- Energy Atlas: Largest disaggregated energy use data in the nation
- High resolution data has not been leveraged to solve policy goals by Utilities.
- Most account information has not been linked to parcels
- Collected data is not used measure energy efficiency savings.

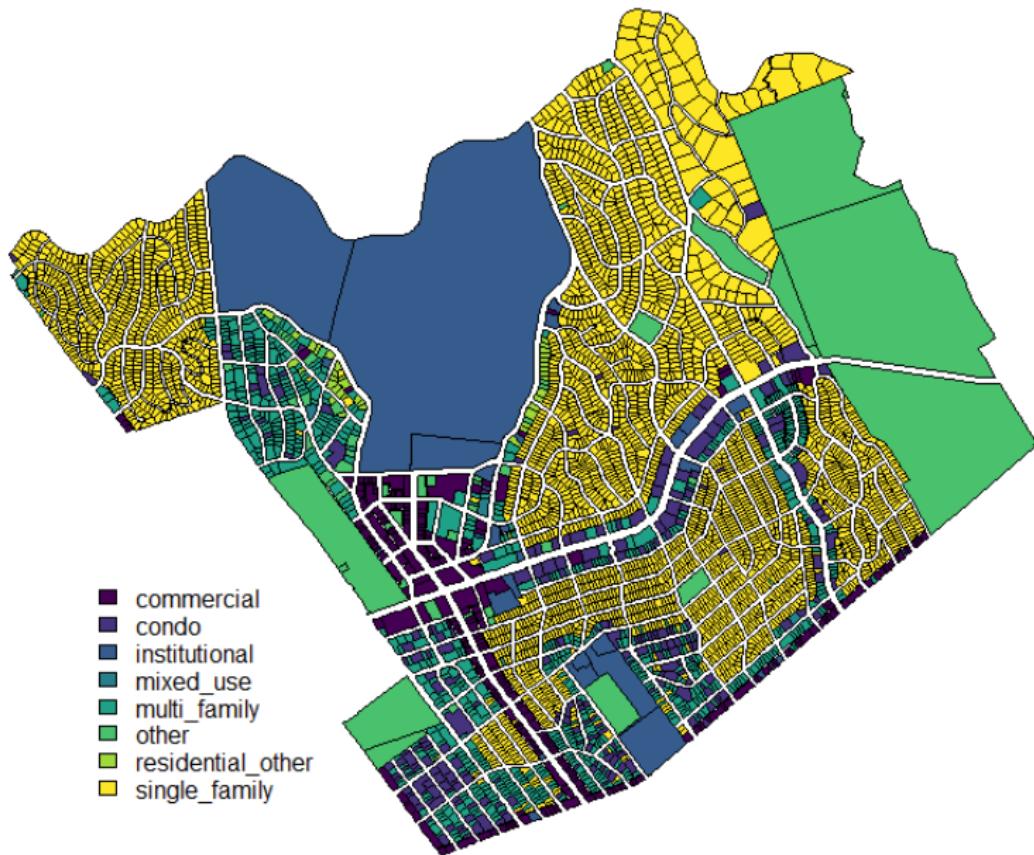
Open Questions

1. Data Ownership and Privacy – ownership could determine onus of securing privacy. Law making lags but privacy preserving methodology does not need to.
2. Data Transparency
3. Data Resolution – can we answer the same questions with less data?

Motivation

- Tackle the main issues by posing specific questions that leverage high resolution data
- California Senate Bill 350:
 1. Increase renewable portfolio standards of utility companies from 33% to 50% by 2030
 2. **Increase California buildings' efficiency by 50% by 2030**
- Aid SB-350's second goal by fitting a quantile regression model to identify *contextual outliers*
- This provides actionable targets for energy efficiency retrofit programs, etc.

Westwood Neighborhood Usetype



Fixed Effects Quantile regression

The conditional quantile function for the monthly energy use of a megaparcel y_{ij} is:

$$Q_{y_{ij}}(\tau|x_{ij}) = \alpha_i + x_{ij}^\top \beta(\tau) + u_{ij}, \quad (1)$$

where

y_{ij} : energy usage of the i th megaparcel in the j th month,

x_{ij} : vectors of covariates on the megaparcels,

α_i : unobservable individual effect introduced by the m megaparcel

u_{ij} : month specific errors

Penalized Fixed Effects Quantile regression

As the number of megaparcel m is much larger than the number of months j , a penalized version of quantile regression is fit and the objective function to be solved is:

$$\min_{(\alpha, \beta)} \sum_{k=1}^q \sum_{j=1}^{n_i} \sum_{i=1}^m w_k \rho_{\tau_k}(y_{ij} - \alpha_i - x_{ij}^\top \beta(\tau)) + \lambda \sum_{i=1}^m |\alpha_i|. \quad (2)$$

where

$\tau_k \in \{\tau_1, \dots, \tau_q\}$: quantiles

$\rho_\tau(u)$: piecewise linear quantile loss function,

w_k : weights that control the relative influence of each of the q quantiles

λ : penalty on the megaparcel fixed effect.

Results

- At the 10% quantile, compared to commercial buildings, Residential Other (single-parcel, 1950s) buildings are the largest energy consumers. Identified as the UCLA Fraternity and Sorority houses on Gayley and Hilgard Avenues.
- Aside a few cases, commercial buildings are consistently the larger consumers at different quantiles. All commercial buildings are single parcel and built over multiple years.
Possibly due to building heating and cooling?
- In Westwood, usetype Other is the largest consumer at the 90% quantile. These megaparcels include the LA Country Club and the UCLA housing on Weyburn.

Statistical approach to Privacy and Resolution

- The range between safeguarding privacy while maintaining data utility is pretty wide.
- Level of necessary data utility is different for researchers vs. general public.
- It is an interesting statistical problem to explore and test the balance of utility and privacy of existing methods.
- Apply differential privacy and synthetic data methods on Energy Atlas, produce masked or synthetic data and re-fit quantile regression models to tests utility of the new data.
- This exercise will help set up guidelines on balancing utility and privacy.
- Metropolitan Water Board Survey: Causal estimation of Efficiency savings accrued.