

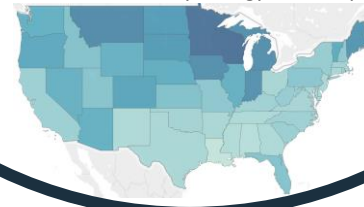
## Introduction & Motivation

- 22% of energy consumed by residential sector.
- Grid modernization and climate change.
- Need for personalized policy recommendations & sustainability solutions.
- Sparse availability of detailed large-scale households and energy data due to consumer privacy, lack of supporting infrastructure, time-consuming long-term data collection.
- Inter-disciplinary approach of data-driven AI, population sciences and energy domain.

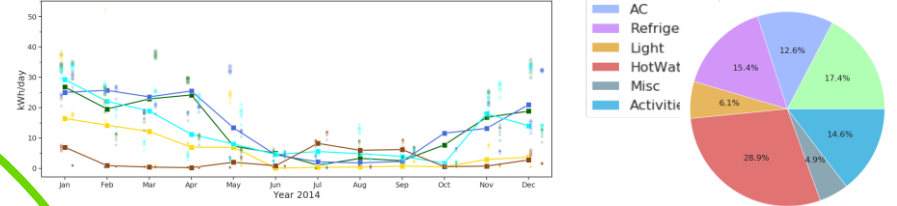
## Research Highlights

- Novel residential energy dataset : Household-level hourly dis-aggregated energy data for over millions of U.S. households.
- Big-data pipelines coupled with microservices-oriented design optimized for resource use.
- Integration of ML, stochastic, and first-principles models and numerous disparate data sources.
- Novel validation metrics for synthetic energy data to address fidelity and diversity.

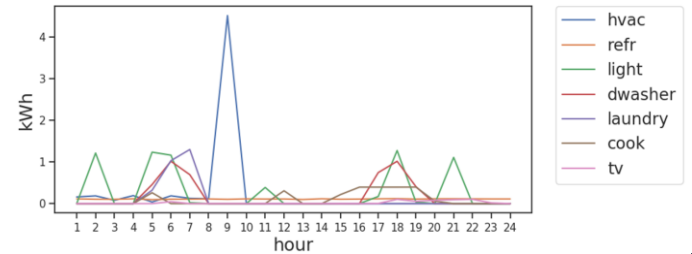
State-level summer day energy consumption



HVAC annual curve for different regions in U.S.



Dis-aggregated energy profile for a household



## Methods

Novel combination of large number of disparate datasets and surveys in one cohesive infrastructure.

Modular and extensible big-data pipelines for data-fusion, modeling, visual analytics, and modeling counterfactuals

Machine learning models for combining synthetic populations, census data, energy surveys, occupancy modeling

Engineering, stochastic, and data-driven behavior & AI models for modeling dis-aggregate energy uses at individual person and household level

## Discussion & Conclusion

- Counterfactual analyses at fine spatio-temporal level to answer important questions in energy sustainability, climate change, and fairness in energy .
- Dis-aggregated energy-use profile available for analyses at hourly intervals at household-level.
- Validation of synthetic data shows good quality and coverage for all types of climate-zones in the U.S.
- Novel data-driven AI framework for energy data generation using big data design principles, data-fusion and machine learning models.