SMU/Bush Institute PostDoc

Bi-weekly progress report

Feb 28th, 2023

Literature review

Methodologies

- Statistical analysis
- Agent-based Modeling (ABM)
- SystDynamics(SD)
  em
- Machine Learning (ML)
- Integrating ML with ABM

Topics

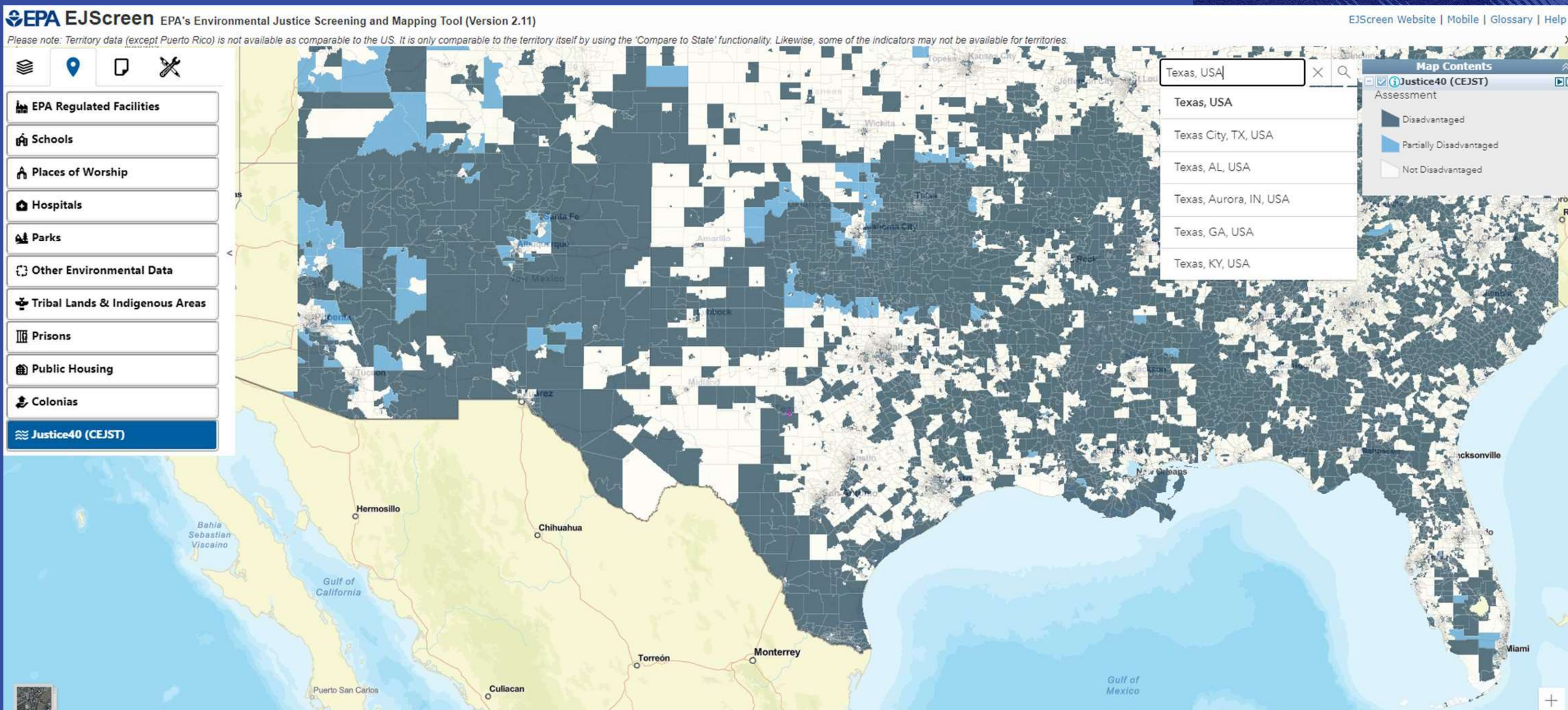
- Equity
- Low-income/disadvantaged communities
- Coupling EV stations with electric grid
- Policies: subsidies, rebates, grants, incentives, tax credits, etc.

Limitations

- Not that much data available at the time
- Research on NYC, LA, Chicago area, no Dallas/Texas
- Difficulty of validating ABM models
- Policy independence between Electricity and EV departments (e.g. independent incentive programs, rebates)

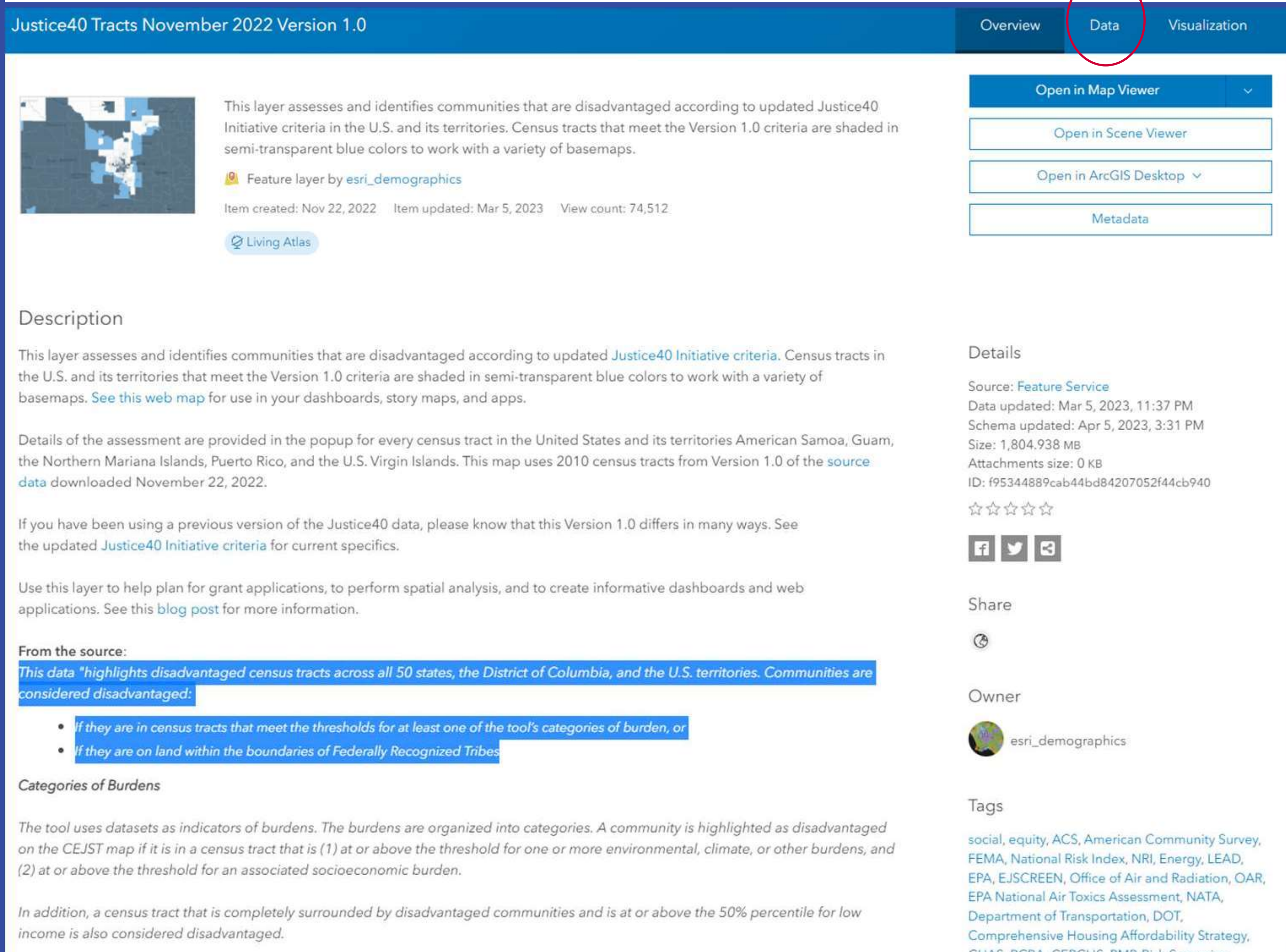
SMU

EPA EJ screen: Justice40 (CEJST)



» https://ejscreen.epa.gov/mapper/

Justice40 Tracts



Justice40 Tracts November 2022 Version 1.0

Overview | Data | Visualization

This layer assesses and identifies communities that are disadvantaged according to updated Justice40 Initiative criteria in the U.S. and its territories. Census tracts that meet the Version 1.0 criteria are shaded in semi-transparent blue colors to work with a variety of basemaps.

Feature layer by esri_demographics

Item created: Nov 22, 2022 Item updated: Mar 5, 2023 View count: 74,512

Living Atlas

Open in Map Viewer

Open in Scene Viewer

Open in ArcGIS Desktop

Metadata

Description

This layer assesses and identifies communities that are disadvantaged according to updated Justice40 Initiative criteria. Census tracts in the U.S. and its territories that meet the Version 1.0 criteria are shaded in semi-transparent blue colors to work with a variety of basemaps. See this web map for use in your dashboards, story maps, and apps.

Details of the assessment are provided in the popup for every census tract in the United States and its territories American Samoa, Guam, the Northern Mariana Islands, Puerto Rico, and the U.S. Virgin Islands. This map uses 2010 census tracts from Version 1.0 of the source data downloaded November 22, 2022.

If you have been using a previous version of the Justice40 data, please know that this Version 1.0 differs in many ways. See the updated Justice40 Initiative criteria for current specifics.

Use this layer to help plan for grant applications, to perform spatial analysis, and to create informative dashboards and web applications. See this blog post for more information.

From the source:

This data "highlights disadvantaged census tracts across all 50 states, the District of Columbia, and the U.S. territories. Communities are considered disadvantaged:

- *If they are in census tracts that meet the thresholds for at least one of the tool's categories of burden, or*
- *If they are on land within the boundaries of Federally Recognized Tribes*

Categories of Burdens

The tool uses datasets as indicators of burdens. The burdens are organized into categories. A community is highlighted as disadvantaged on the CEJST map if it is in a census tract that is (1) at or above the threshold for one or more environmental, climate, or other burdens, and (2) at or above the threshold for an associated socioeconomic burden.

In addition, a census tract that is completely surrounded by disadvantaged communities and is at or above the 50% percentile for low income is also considered disadvantaged.

Details

Source: Feature Service
Data updated: Mar 5, 2023, 11:37 PM
Schema updated: Apr 5, 2023, 3:31 PM
Size: 1,804.938 MB
Attachments size: 0 KB
ID: f95344889cab44bd84207052f44cb940

Share

Owner

esri_demographics

Tags

social, equity, ACS, American Community Survey, FEMA, National Risk Index, NRI, Energy, LEAD, EPA, EJSCREEN, Office of Air and Radiation, OAR, EPA National Air Toxics Assessment, NATA, Department of Transportation, DOT, Comprehensive Housing Affordability Strategy, CHAS, RCRA, CERCLIS, RMP Risk-Screening

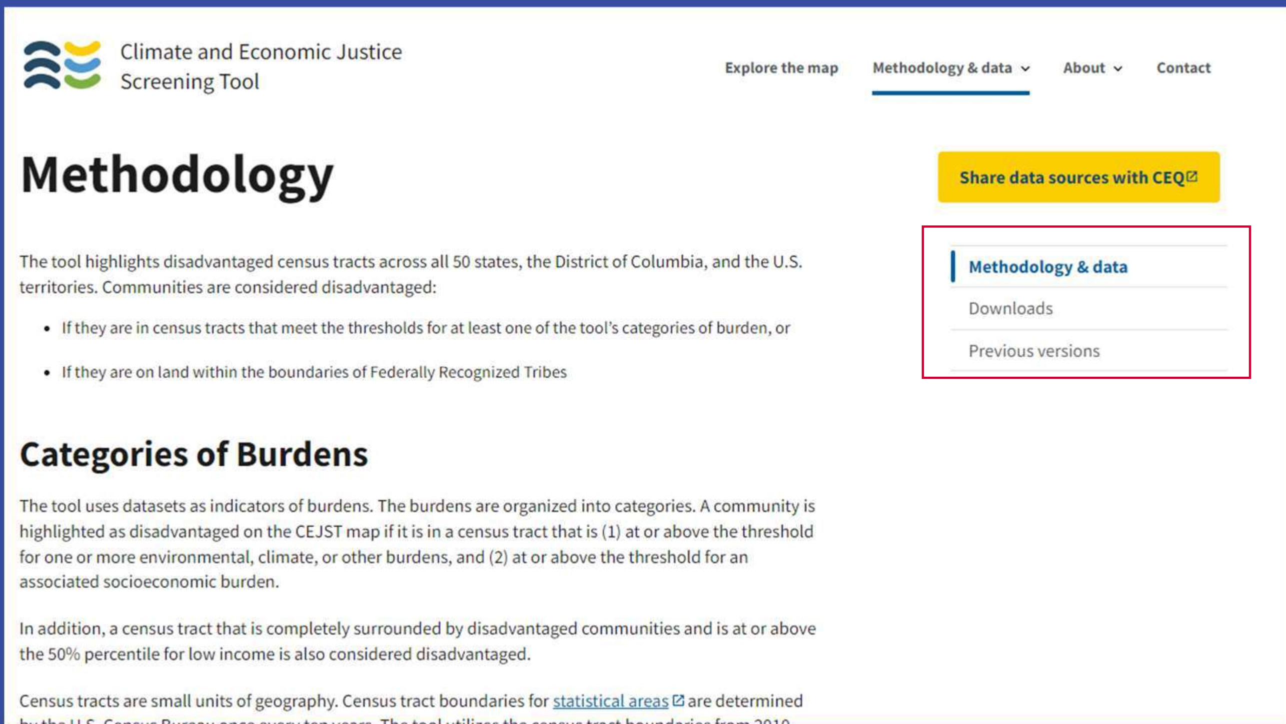
From the source:
This data "highlights disadvantaged census tracts across all 50 states, the District of Columbia, and the U.S. territories.
Communities are considered disadvantaged:
•If they are in census tracts that meet the thresholds for at least one of the tool's categories of burden, or
•If they are on land within the boundaries of Federally Recognized Tribes

Justice40 Initiative criteria

arcgis-blog

SMU

https://www.arcgis.com/home/item.html?id=f95344889cab44bd84207052f44cb940



Climate and Economic Justice Screening Tool

Explore the map | Methodology & data | About | Contact

Methodology

Share data sources with CEQ

- Methodology & data
- Downloads
- Previous versions

The tool highlights disadvantaged census tracts across all 50 states, the District of Columbia, and the U.S. territories. Communities are considered disadvantaged:

- If they are in census tracts that meet the thresholds for at least one of the tool's categories of burden, or
- If they are on land within the boundaries of Federally Recognized Tribes

Categories of Burdens

The tool uses datasets as indicators of burdens. The burdens are organized into categories. A community is highlighted as disadvantaged on the CEJST map if it is in a census tract that is (1) at or above the threshold for one or more environmental, climate, or other burdens, and (2) at or above the threshold for an associated socioeconomic burden.

In addition, a census tract that is completely surrounded by disadvantaged communities and is at or above the 50% percentile for low income is also considered disadvantaged.

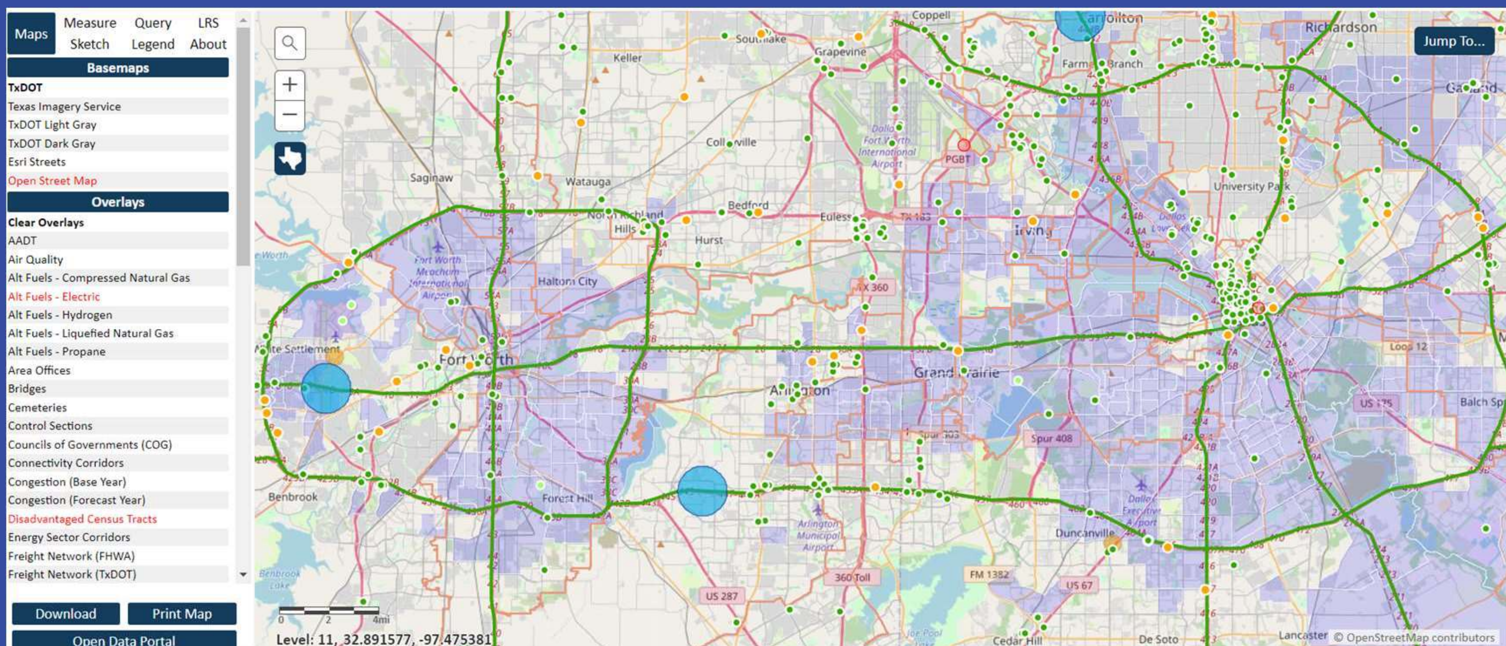
Census tracts are small units of geography. Census tract boundaries for statistical areas are determined by the U.S. Census Bureau once every ten years. The tool utilizes the census tract boundaries from 2010

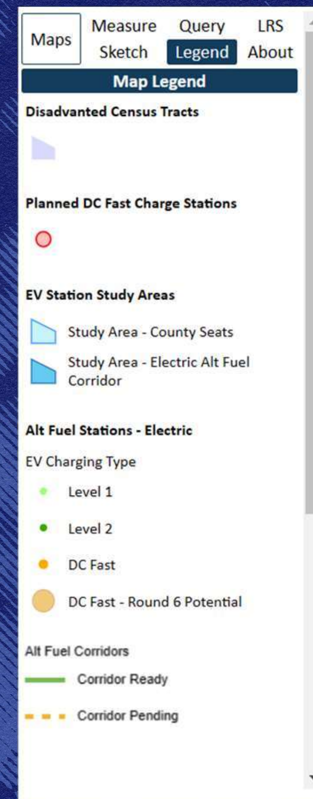
https://screeningtool.geoplatform.gov/en/methodology

SMU

Dataset: o_TXDoT\texas EV stations

EV charging stations and underserved communities mapping





SMU

[Statewide Planning Map (txdot.gov)]

Dataset: 1_DoE\Alternative Fuel Data Center(AFDC)

EV charging stations data:

Fuel Type	Station Na	Street Add	Intersectic	City	State	ZIP	Plus4	Station Ph	Status Cod	Expected D	Groups Wi	Access Da	Cards Acce	BD Blends	NG Fill Typ	NG PSI	EV Level1	EV Level2	EV DC Fast	EV Other I	EV Networ	EV N
ELEC	LADWP - T	11797 Truesdale St		Sun Valley	CA	91352			E		Private	Fleet use only						39	3		Non-Networke	
ELEC	LADWP - V	1394 S Sepulveda Blv		Los Angele	CA	90024			E		Private							4			Non-Networke	

Policies: Laws and incentives

Law Id	State	Title	Text	Enacted D	Amended I	Recent?	Sequence	Type	Agency	Significant	Expired Da	Archived D	Repealed	Topic	Technolog	Incentive (	Regulation	User Categ	Referen
4739	TX	Propane a	The			FALSE	135	Laws and Regulations		2018-09-10 14:12:22 UTC					NG|LPG		OTHER	STATION|	http://w
5309	TX	Clean Vehi	The Texas			FALSE	25	State Incentives		2018-06-18 15:30:15 UTC					AFTMKTC(	GNT		STATION|I	http://w
5312	TX	Natural Ga	The Texas			FALSE	80	Utility/Private Incenti		2015-08-11 15:37:53 UTC					AFTMKTC(	RBATE		MAN|FLEET|IND	

Vehicle info: light/medium/heavy duty

Vehicle ID	Fuel ID	Fuel Confi	Manufactu	Category I	Model	Model Yea	Alternative	Alternative	Alternative	Conventio	Conventio	Conventio	Transmissi	Engine Typ	Engine Size	Engine Cyli	Engine/Mc	Manufactu	Manufactu	Category	Fuel Code	Fuel	F
13140	45	9	377	27	A3	2023				28	38	32	Auto	SI	2.0L	4	2.0L I4	Audi	http://pro	Sedan/Wa	HYBR	Hybrid Ele	H
13139	45	9	377	27	A3 quattro	2023				27	35	30	Auto	SI	2.0L	4	2.0L I4	Audi	http://pro	Sedan/Wa	HYBR	Hybrid Ele	H
13141	45	9	377	27	A4 quattro	2023						26-28	Auto	SI	2.0L	4	2.0L I4	Audi	http://pro	Sedan/Wa	HYBR	Hybrid Ele	H

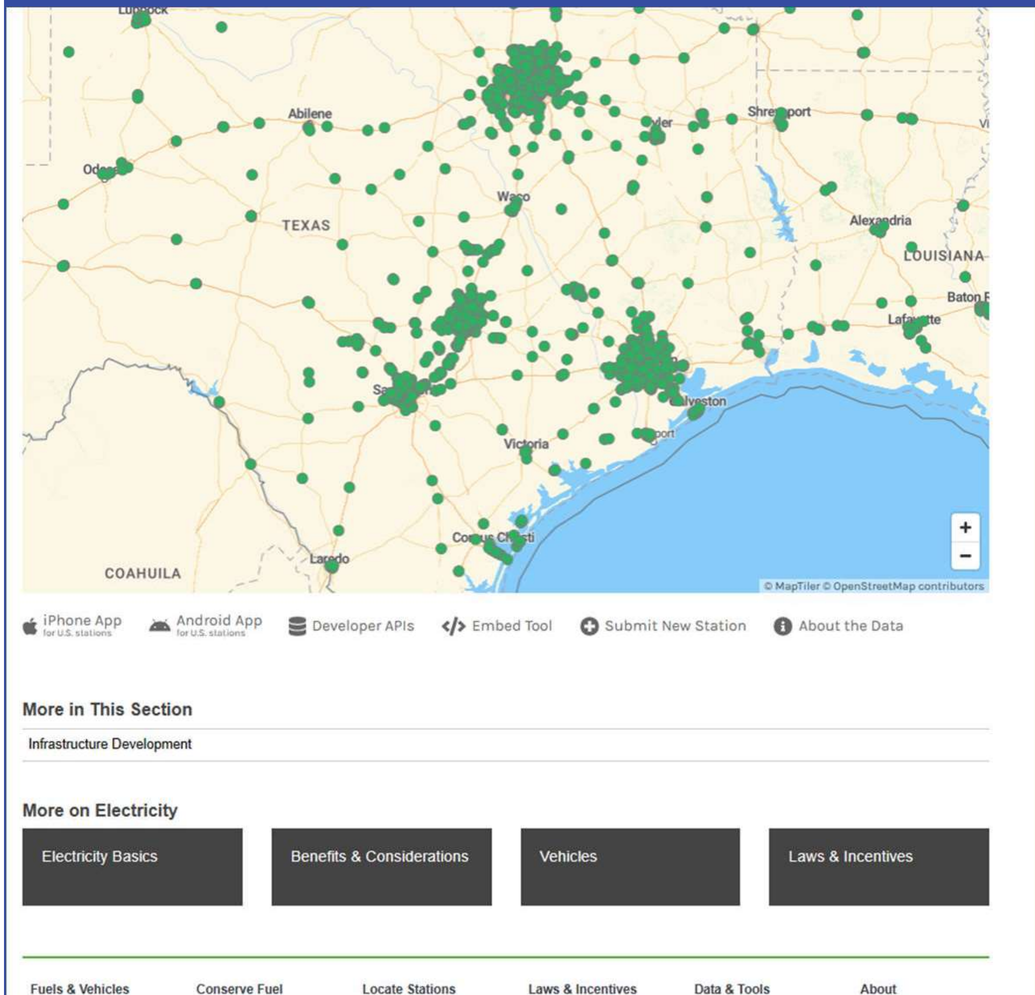
[Map with API access](#)

[Alternative Fuels Data Center: Electric Vehicle Charging Station Locations (energy.gov)](#)

[Alternative Fuels Data Center: Data Downloads (energy.gov)](#)

SMU

Dataset: 1_DoE\Alternative Fuel Data Center(AFDC)

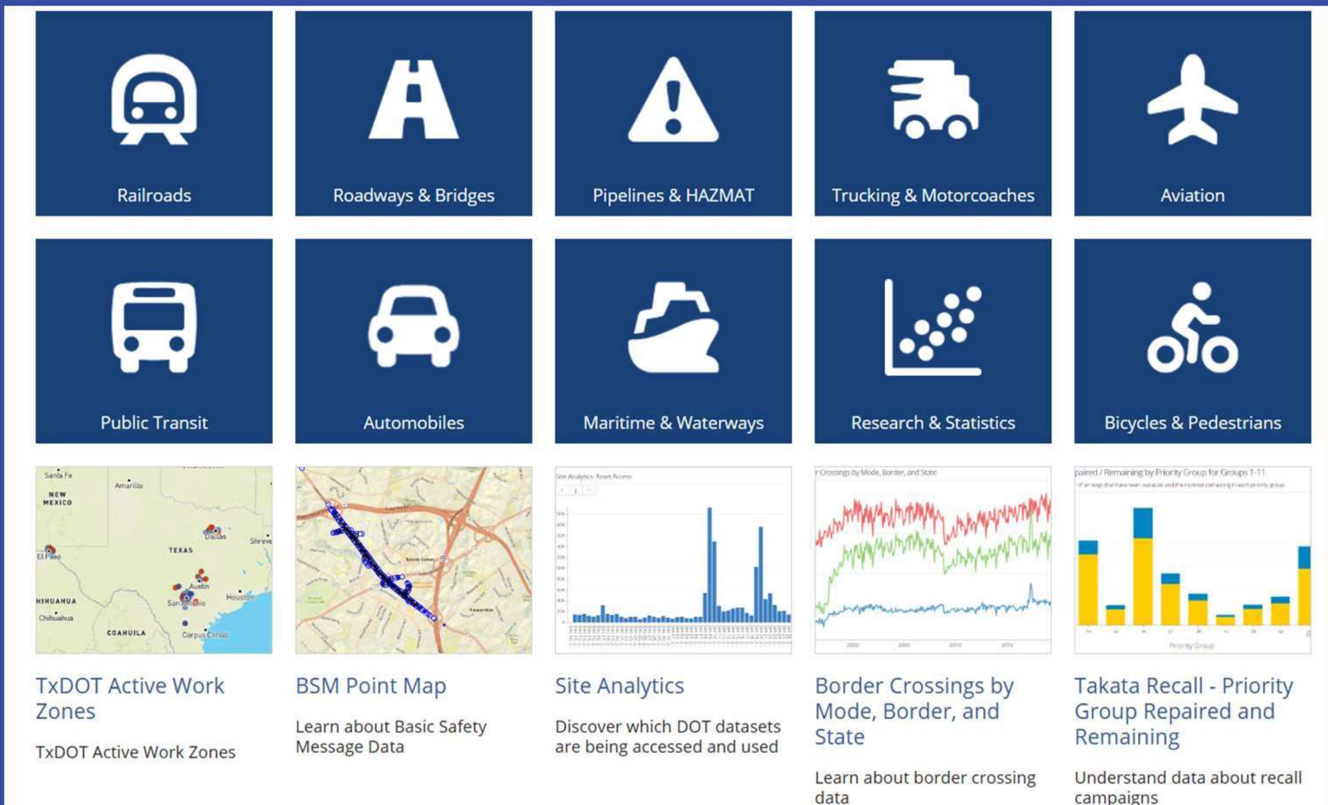


Alternative Fuels Data Center: Tools (energy.gov)

Alternative Fuels Data Center: Data Downloads (energy.gov)

Alternative Fuel Stations API | NREL: Developer Network

Dataset:6_BureauTransportationStatistics



[Department of Transportation - Open Data Portal |](#)

[Video Guides | Department of Transportation - Data Portal](#)

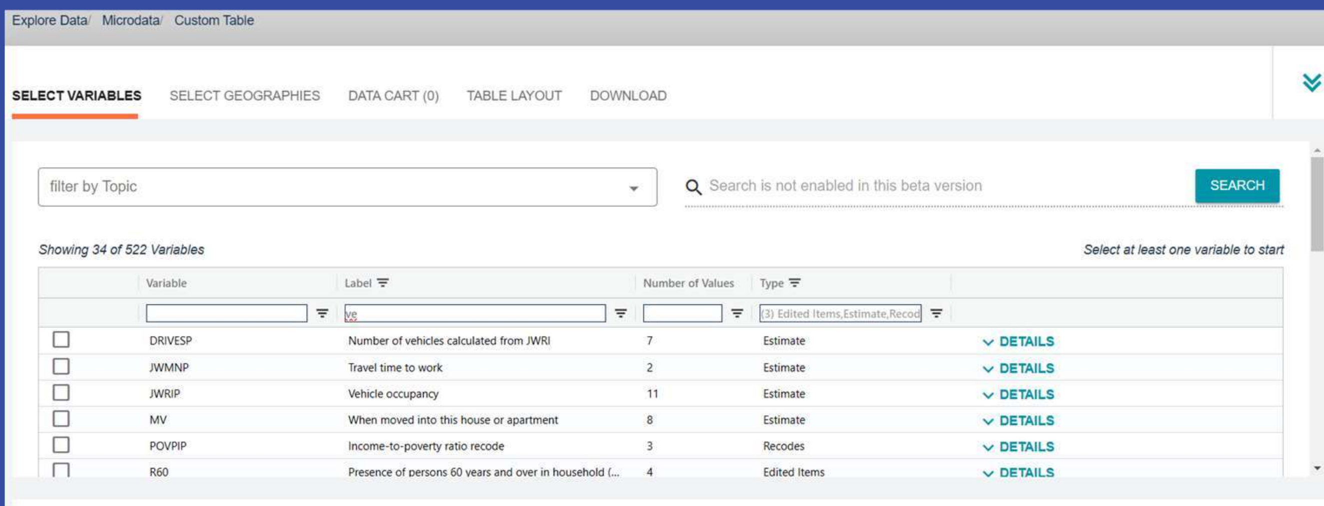
[Socrata Developers | Socrata](#)

SMU

Dataset: 2.U.S. Census Bureau

Public Use Microdata Sample:
This contains lots of **demographic data**, including race, gender, age, household income, employment, commute time, vehicle occupancy, etc.

Custom demographic data tables available but limited to maximum number of variables and its layouts. Need to figure out an approach to retrieve the data programmatically and store it in an organized database for future use.



- American Community Survey (ACS): ACS 1-Year, ACS Migration Flows
- Economic Indicators Time Series
- Decennial Census
- Economic Census
- County Business Patterns and Nonemployer Statistics
- Population Estimates and Projections
- International Trade

Datasets (census.gov)

Available APIs (census.gov)

Census Microdata API

Economic Census

Census Bureau Data

MDAT (census.gov)

Census Academy

Data Science & Visualization Resources (census.gov)

Dataset: 2.U.S. Census Bureau-API

local file: Census Data API User guide

Census Data API: /data

SMU

Dataset:5_US Bureau of Labor Statistics

Inflation & Prices >>

- Consumer Price Index
- Producer Price Indexes
- Import/Export Price Indexes
- Contract Escalation
- Price Index Research

Pay & Benefits >>

- Employment Costs
- Wage Data by Occupation
- Earnings by Demographics
- Earnings by Industry
- County Wages
- Benefits
- Modeled Wage Estimates
- Compensation Research
- Strikes & Lockouts

Occupational Requirements >>

Unemployment >>

- National Unemployment Rate
- State & Local Unemployment Rates
- Unemployment Research

Employment >>

- National Employment
- State & Local Employment
- County Employment
- Worker Characteristics
- American Indian Report
- Employment Projections
- Job Openings & Labor Turnover Survey
- Business Response Survey
- Employment by Occupation
- Work Experience Over Time
- Business Employment Dynamics
- Foreign Direct Investment
- Employment Research

Workplace Injuries >>

Productivity >>

- Labor and Total Factor Productivity
- Productivity Research

Spending & Time Use >>

- Consumer Expenditures
- How Americans Spend Time

International >>

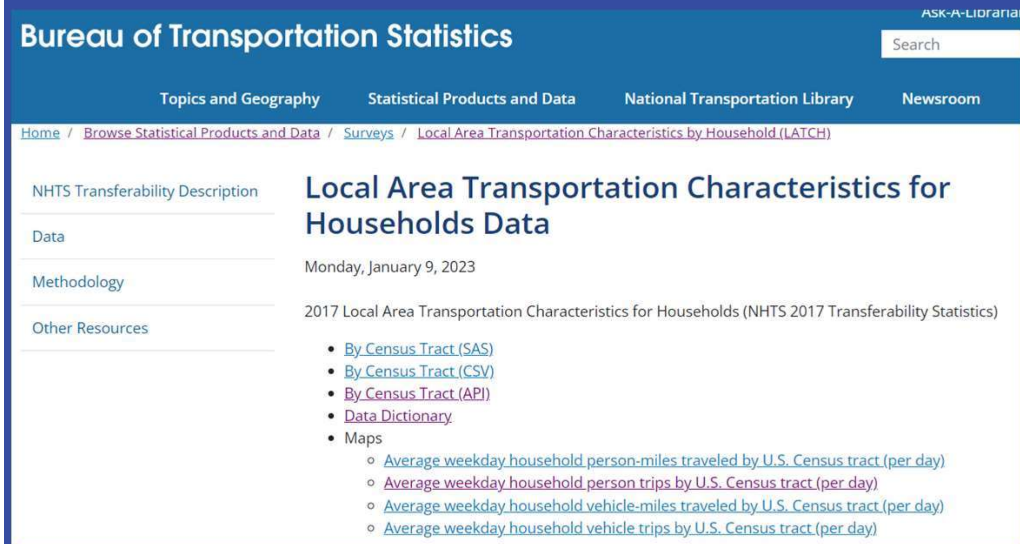
- International Technical Cooperation
- Import/Export Price Indexes

Geographic Information >>

- New England (Boston)
- New York-New Jersey (NY City)
- Mid-Atlantic (Philadelphia)
- Southeast (Atlanta)
- Midwest (Chicago)
- Southwest (Dallas)
- Mountain-Plains (Kansas City)
- West (San Francisco)

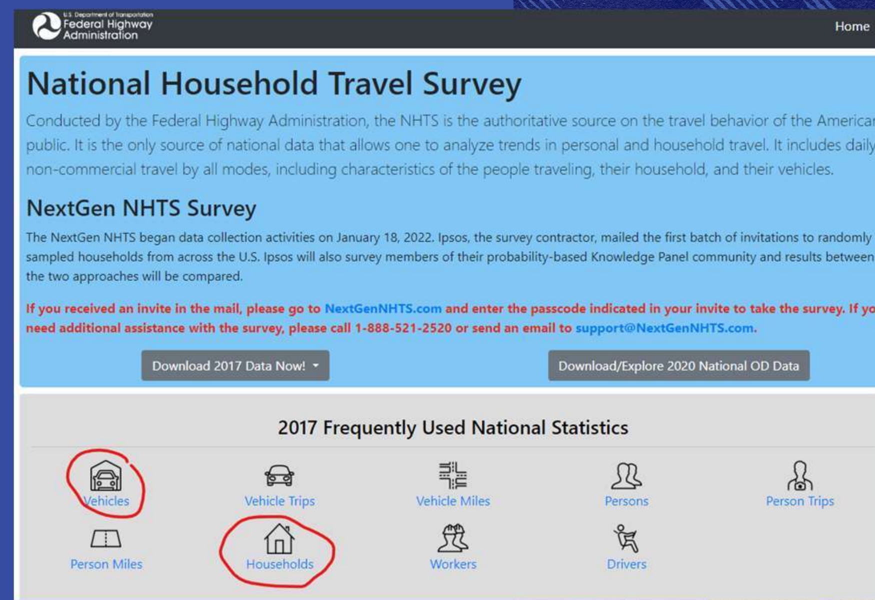
Accessing the Public Data API with R : U.S. Bureau of Labor Statistics (bls.gov)
How many electric vehicle charging stations are there in the US? - USAFacts
How do tax credits for electric cars work? - USAFacts

Dataset:6-1_National Household Travel Survey



https://www.bts.gov/latch/latch-data

https://www.bts.gov/browse-statistical-products-and-data



https://nhts.ornl.gov/

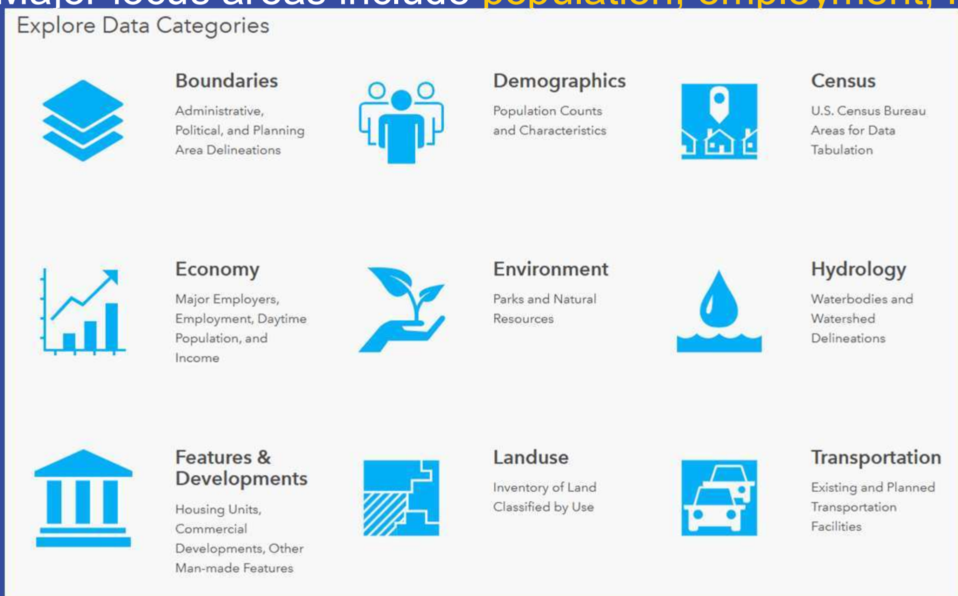
SMU

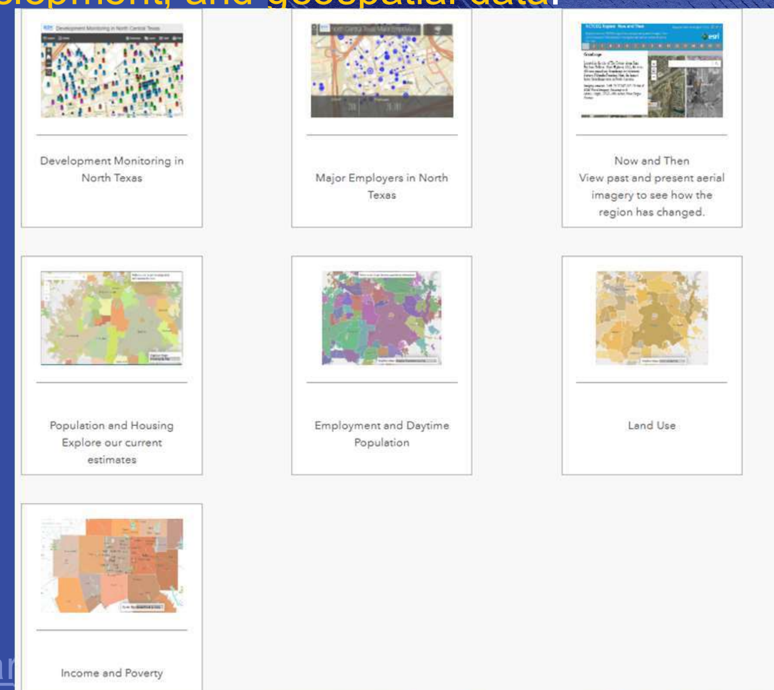
2021 Vehicle Inventory and Use Survey (VIUS)
Available Fall 2023

Dataset: 3_NorthCentralTexasCouncilGov

» The Regional Data Center

» is a service of the Research and Information Services (RIS) department. RIS provides objective information and analysis on the development of the region for use in planning and economic development activities. Major focus areas include population, employment, land use, development, and geospatial data.



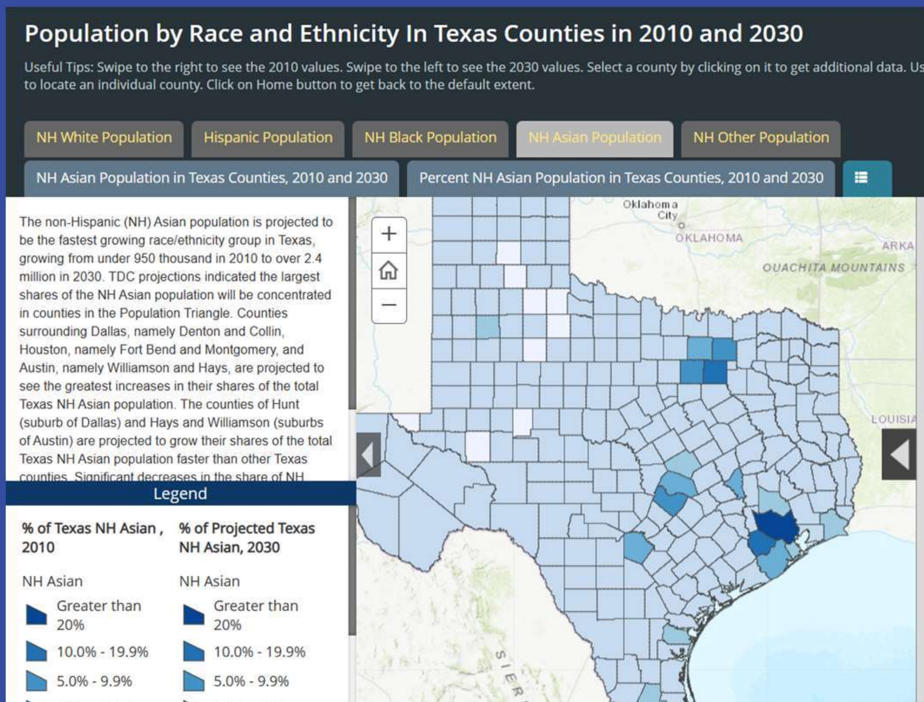


[Regional Data Center - North Central Texas Council of Governments (ar](#)

[Search for '*' | North Central Texas Council of Governments (arcgis.com)](#)

[2045 NCTCOG Demographic Forecast (City) | 2045 NCTCOG Demographic Forecast (City) | North Central Texas Council of Governments (arcgis.com)](#)

Dataset:4_Texas demographic center



[Texas Demographic Center]

[Population by Race and Ethnicity In Texas Counties in 2010 and 2030 (arcgis.com)]

[Population Projections for Texas Counties, 2020-2040 and 2020-2060 (arcgis.com)]

NREL

Developer Network

HOME | DOCUMENTATION | COMMUNITY

Documentation » Electricity

Electricity

Services associated with electricity costs, generation, transmission, delivery, and m

OpenEI Utility Rates

Access complex utility rate structure information (across all sectors) for most U.S. utility companies from the National Utility Rate Database. This information is collected and quality controlled on a continual basis by Illinois State University on behalf of DOE and housed within the OpenEI.org platform.

Utility Rates (GET /api/utility_rates/v3)

This service returns annual average utility rates ($/kWH) for residential, commercial and industrial sectors as well as the local utility name for a specific location. This service does not return complex rate information.

Utility Rates by Census Region (GET /api/census_rate/v3)

Given a location and Census Bureau region level (block, blockgroup, tract), return the Census Bureau ID, utility rate and company information.

Deprecated

Energy Incentives (Version 2 - Deprecated) (/api/energy_incentives/v2/dsire)

Deprecated: This service lists the incentives found in the DSIRE database by location.

Web Services

Available Web Services

- REST - provides a RESTful wrapper around high-value data.
  - Term extraction
    - Documentation
    - includes content recommendation engine
  - Utility Companies
    - Documentation Version 2, Version 3
    - Human-readable source data
    - Also see: the utilities gateway
  - Utility Rates
    - International Utility Rate Database API Documentation
    - U.S. Utility Rate Database API Documentation
    - Also see: the Utilities Gateway
  - Incentives for Renewables & Efficiency
    - Documentation
    - Also see: the incentives & policies gateway
- Ask queries
  - inline queries for Semantic MediaWiki
- SPARQL queries
- NREL's developer network
- Hackathon Resources

Developer Network

HOME | DOCUMENTATION | COMMUNITY

The National Renewable Energy Laboratory's devel
energy and alternative fuel data.

Web Service Documentation ▸

Explore our available Web services.

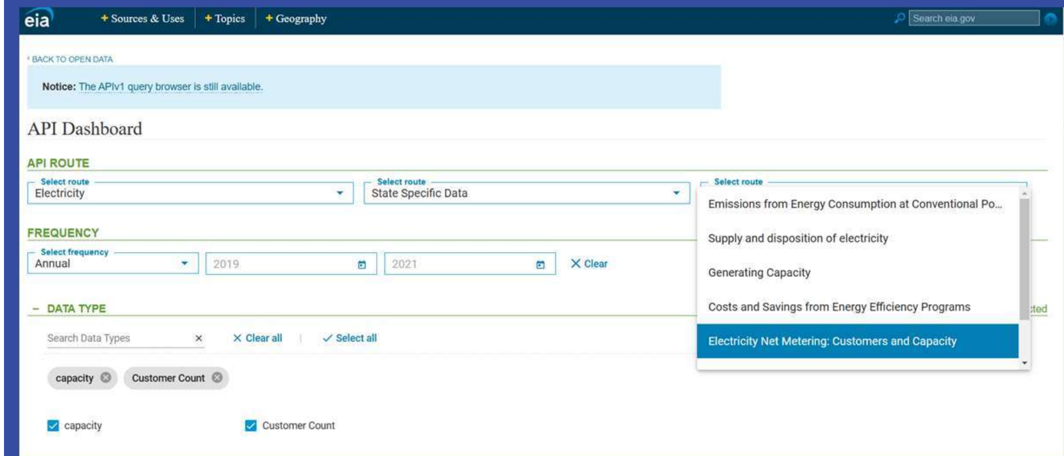
- Buildings
- Electricity
- Energy Optimization
- Partnering
- Solar
- Transportation
- Wave
- Wind

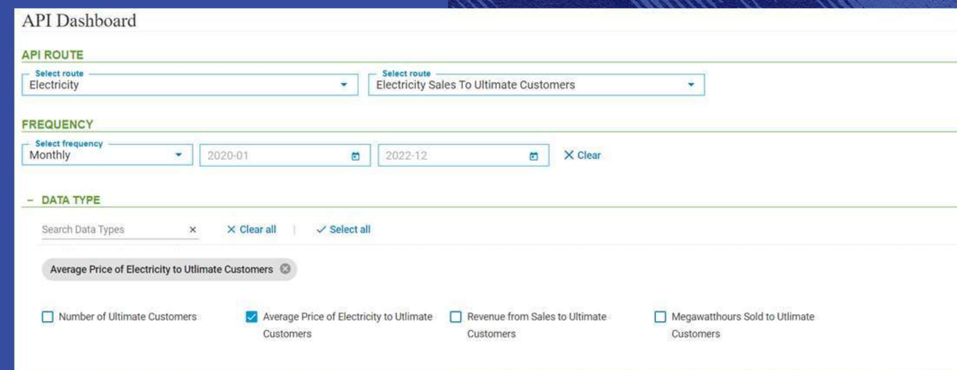
BCL (nrel.gov)

Electricity APIs | NREL: Developer Network

NREL: Developer Network

Dataset:7_USEIA





[API Dashboard - U.S. Energy Information Administration (EIA)]

[API Dashboard - U.S. Energy Information Administration (EIA)]

[Opendata - U.S. Energy Information Administration (EIA)]

SMU

Dataset:8_OEDI

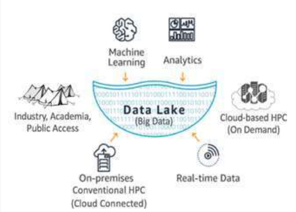
Open Energy Data Initiative

Universal Accessibility

Our open architecture is designed for universal access and dissemination of big data. Data Lakes can be accessed via our cloud partners. There are a few ways to utilize Data Lakes. Jupyter notebooks are a common option for utilizing Data Lakes datasets. However, there are many options for processing Data Lakes.

Multiple Ways to Access

- Jupyter notebooks (example)
- Google Earth Engine
- Direct access to Data Lakes (Requires a cloud account.)
- Data Lake Viewer (Currently only AWS)
- Native cloud command line tools (AWS, Google, Azure)
- Mounting the data as a local read-only drive in a cloud-built computer cluster. Requires same availability zone.



Data stored in the lake is actionable and discoverable, increasing accessibility to large and complex datasets.

Manual Download and AWS CLI

- Download datasets via data-catalog viewer, users may process data in ways they see fit. Alternatively, batch download via Amazon's Command Line Interface.
- **Free** for end users.
- Learn More.

HSDS and Jupyter Notebook

- Highly Scalable Data Service (HSDS) is a REST-based product and solution for reading and writing complex binary data formats within an object-based storage environment, such as the Cloud.
- **Free** for end users.
- Learn More.

S3 Tools and APIs

- Access data with S3 Tools and APIs to analyze datasets.
- **Free** for end users.
- Learn More.

AWS Athena

- Athena is an interactive query service offered by AWS that makes it easy to access data in Amazon S3 using SQL.
- AWS Athena can be used for a small fee, **often less than $1.**
- Learn More.

Google BigQuery

- Google BigQuery is a serverless, highly scalable data warehouse that comes with Google's built-in query engine.
- BigQuery offers two pricing models (On-demand and Flat-rate pricing) for running queries, **often less than $0.10**
- Learn More.

Cloud-based HPC Cluster

-HPC (High Performance Computing) instances are ideal for applications that benefit from high-performance processors, such as large simulations and machine learning workloads.
- Multiple pricing models, **often less than $10.**
- Learn More.

AWS SageMaker Studio

- The SageMaker Studio provides a web-based integrated development environment (IDE) where users can see and interact with all ML workflows on AWS.
- **End user costs are often less than $20.** Use Amazon SageMaker Savings Plan to reduce costs by up to 64%, compared to On-Demand pricing.
- Learn More.

AWS SageMaker Studio Lab

- Amazon SageMaker Studio Lab is a free machine learning (ML) development environment that provides the compute, storage (up to 15GB) and security to learn and experiment with ML.
- **Free** for end users.
- Learn More.

[OEDI: All Data (openei.org)](https://data.openei.org/submissions/all)

ttps://data.openei.org/submissions/all

End-Use Load Profiles for the U.S. Building Stock

DOI 10.25984/1876417 — Publicly accessible License

The United States is embarking on an ambitious transition to a 100% clean energy economy by 2050, which will require improving the flexibility of electric grids. One way to achieve grid flexibility is to shed or shift demand to align with changing grid needs. To facilitate this, it is critical to understand how and when energy is used. High quality end-use load profiles (EULPs) provide this information, and can help cities, states, and utilities understand the time-sensitive value of energy efficiency, demand response, and distributed energy resources. Publicly available EULPs have traditionally had limited application because of age and incomplete geographic representation. To help fill this gap, the U.S. Department of Energy (DOE) funded a three-year project, End-Use Load Profiles for the U.S. Building Stock, that culminated in this publicly available dataset of calibrated and validated 15-minute resolution load profiles for all major residential and commercial building types and end uses, across all climate regions in the United States. These EULP were created by calibrating the ResStock and ComStock physics-based building stock models using many different measured datasets, as described in the "Technical Report Documenting Methodology" linked in the submission.

- **Aggregate and Individual Building Timeseries End Use Load Profiles** — View Data Lake, 39.68 TB
  Aggregate and individual building timeseries end use load profiles. See README.md file inside the dataset for data hierarchy, organization, etc.
  8,896,627*
- **FAQ for End-Use Load Profiles Dataset** — View
  Answers to frequently asked questions about the End-Use Load Profiles dataset and ComStock/ResStock data viewers.
- **README** — View Data Lake, 39.68 TB
  This file describes the organization of the datasets.
  8,896,470*
- **Technical Report Documenting Methodology** — View

[OEDI: End-Use Load Profiles for the U.S. Building Stock (openei.org)]

City and County Vehicle Inventories

DOI 10.25984/1788088 — Publicly accessible License

This light-duty vehicle inventory dataset provides information on vehicle registrations by vehicle type (car vs. truck), fuel type, and model year showing the changes in adoption trends over time and average fuel economies.

This data is part of a suite of state and local energy profile data available at the "State and Local Energy Profile Data Suite" link below and builds on Cities-LEAP energy modeling, available at the "EERE Cities-LEAP Page" link below. Examples of how to use the data to inform energy planning can be found at the "Example Uses" link below.

- **City and County Vehicle Inventories.xlsb** — Download, 59.71 MB
  The light-duty vehicle inventory dataset. Includes city and county vehicle data. This data includes vehicle type, fuel type, vehicle model year, and vehicle adoption.
  1,238*

[OEDI: City and County Vehicle Inventories (openei.org)]

Sample IEEE123 Bus system for OEDI SI

Publicly accessible License

The aim of this project is to create an easy-to-use platform where various types of analytics can be performed on a wide range of electrical grid datasets. The aim is to establish an open-source library of algorithms that universities, national labs and other developers can contribute to which can be used on both open-source and proprietary grid data to improve the analysis of electrical distribution systems for the grid modeling community. OEDI Systems Integration (SI) is a grid algorithms and data analytics API created to standardize how data is sent between different modules that are run as part of a co-simulation.

This submission includes an IEEE123 bus system with time series load and PV data attached. An example electrical system, named the OEDI SI feeder, is used to test the workflow in the co-simulation. The system used is the IEEE123 test system, which is a well studied test system (see link below to IEEE PES Test Feeder), but some modifications were made to it to add some solar power modules and measurements on the system.

The readme file included in the S3 bucket provides information about the directory structure and how to use the algorithms. The sensors.json file is used to define the measurement locations.

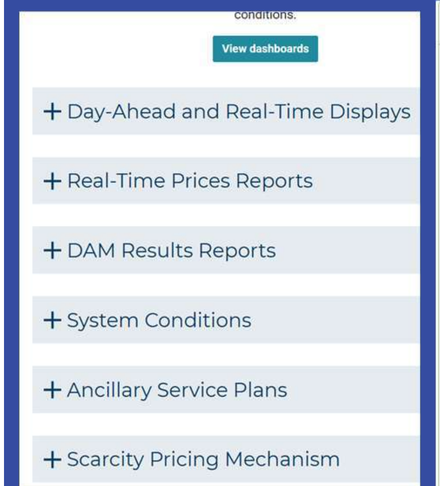
- **Github OEDI SI IEEE123 Bus Test System** — View Repository
  Github link for the location of the IEEE123 bus system for testing the OEDI SI framework.

[OEDI: Sample IEEE123 Bus system for OEDI SI (openei.org)]

SMU

Dataset: 9_ercot

Electric Reliability Council of Texas


conditions.

View dashboards

+ Day-Ahead and Real-Time Displays

+ Real-Time Prices Reports

+ DAM Results Reports

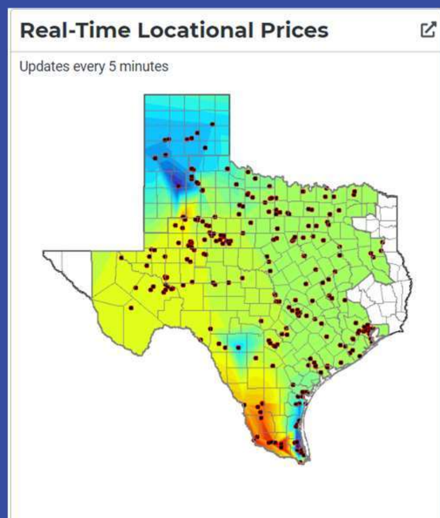
+ System Conditions

+ Ancillary Service Plans

+ Scarcity Pricing Mechanism

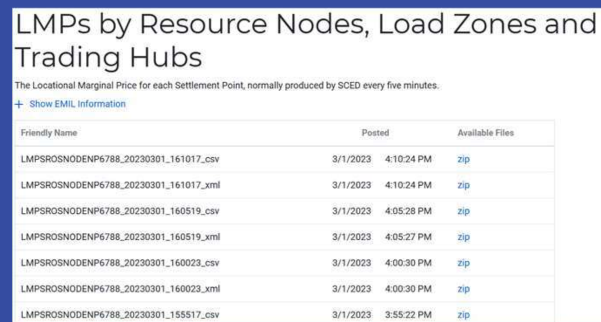
real-time data

[Market Information (ercot.com)]



[Grid and Market Conditions (ercot.com)]

https://www.ercot.com/gridmktinfo/dashboards


LMPs by Resource Nodes, Load Zones and Trading Hubs

The Locational Marginal Price for each Settlement Point, normally produced by SCED every five minutes.

+ Show EMIL Information

Friendly Name	Posted		Available Files
LMPSROSNODENP6788_20230301_161017_csv	3/1/2023	4:10:24 PM	zip
LMPSROSNODENP6788_20230301_161017_xml	3/1/2023	4:10:24 PM	zip
LMPSROSNODENP6788_20230301_160519_csv	3/1/2023	4:05:28 PM	zip
LMPSROSNODENP6788_20230301_160519_xml	3/1/2023	4:05:27 PM	zip
LMPSROSNODENP6788_20230301_160023_csv	3/1/2023	4:00:30 PM	zip
LMPSROSNODENP6788_20230301_160023_xml	3/1/2023	4:00:30 PM	zip
LMPSROSNODENP6788_20230301_155517_csv	3/1/2023	3:55:22 PM	zip


Home > Market Participants > EMIL > Data Product Details

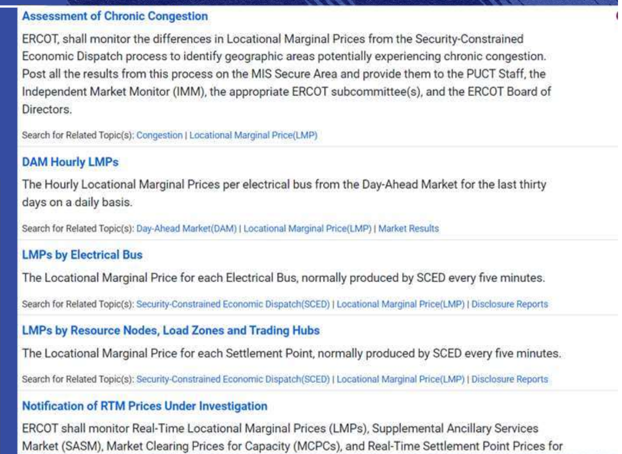
LMPs by Electrical Bus

The Locational Marginal Price for each Electrical Bus, normally produced by SCED every five minutes.

+ Show EMIL Information

Friendly Name	Posted		Available Files
LMPSELECTBUSNP6787_20230301_161017_csv	3/1/2023	4:10:22 PM	zip
LMPSELECTBUSNP6787_20230301_161017_xml	3/1/2023	4:10:22 PM	zip
LMPSELECTBUSNP6787_20230301_160519_csv	3/1/2023	4:05:27 PM	zip
LMPSELECTBUSNP6787_20230301_160519_xml	3/1/2023	4:05:27 PM	zip
LMPSELECTBUSNP6787_20230301_160023_csv	3/1/2023	4:00:31 PM	zip

[Data Product Details (ercot.com)]


Assessment of Chronic Congestion

ERCOT, shall monitor the differences in Locational Marginal Prices from the Security-Constrained Economic Dispatch process to identify geographic areas potentially experiencing chronic congestion. Post all the results from this process on the MIS Secure Area and provide them to the PUCT Staff, the Independent Market Monitor (IMM), the appropriate ERCOT subcommittee(s), and the ERCOT Board of Directors.

Search for Related Topic(s): Congestion | Locational Marginal Price(LMP)

DAM Hourly LMPs

The Hourly Locational Marginal Prices per electrical bus from the Day-Ahead Market for the last thirty days on a daily basis.

Search for Related Topic(s): Day-Ahead Market(DAM) | Locational Marginal Price(LMP) | Market Results

LMPs by Electrical Bus

The Locational Marginal Price for each Electrical Bus, normally produced by SCED every five minutes.

Search for Related Topic(s): Security-Constrained Economic Dispatch(SCED) | Locational Marginal Price(LMP) | Disclosure Reports

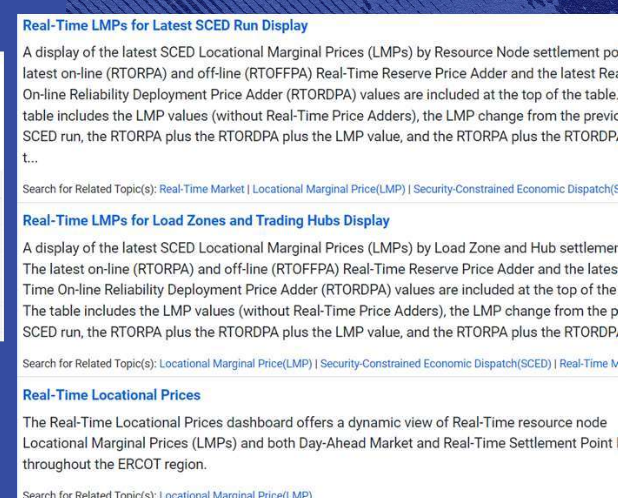
LMPs by Resource Nodes, Load Zones and Trading Hubs

The Locational Marginal Price for each Settlement Point, normally produced by SCED every five minutes.

Search for Related Topic(s): Security-Constrained Economic Dispatch(SCED) | Locational Marginal Price(LMP) | Disclosure Reports

Notification of RTM Prices Under Investigation

ERCOT shall monitor Real-Time Locational Marginal Prices (LMPs), Supplemental Ancillary Services Market (SASM), Market Clearing Prices for Capacity (MCPCs), and Real-Time Settlement Point Prices for


Real-Time LMPs for Latest SCED Run Display

A display of the latest SCED Locational Marginal Prices (LMPs) by Resource Node settlement po latest on-line (RTORPA) and off-line (RTOFFPA) Real-Time Reserve Price Adder and the latest Re On-line Reliability Deployment Price Adder (RTORDPA) values are included at the top of the table table includes the LMP values (without Real-Time Price Adders), the LMP change from the previ SCED run, the RTORPA plus the RTORDPA plus the LMP value, and the RTORPA plus the RTORDP t...

Search for Related Topic(s): Real-Time Market | Locational Marginal Price(LMP) | Security-Constrained Economic Dispatch(S

Real-Time LMPs for Load Zones and Trading Hubs Display

A display of the latest SCED Locational Marginal Prices (LMPs) by Load Zone and Hub settlemer The latest on-line (RTORPA) and off-line (RTOFFPA) Real-Time Reserve Price Adder and the lates Time On-line Reliability Deployment Price Adder (RTORDPA) values are included at the top of the The table includes the LMP values (without Real-Time Price Adders), the LMP change from the p SCED run, the RTORPA plus the RTORDPA plus the LMP value, and the RTORPA plus the RTORDP

Search for Related Topic(s): Locational Marginal Price(LMP) | Security-Constrained Economic Dispatch(SCED) | Real-Time M

Real-Time Locational Prices

The Real-Time Locational Prices dashboard offers a dynamic view of Real-Time resource node Locational Marginal Prices (LMPs) and both Day-Ahead Market and Real-Time Settlement Point throughout the ERCOT region.

Search for Related Topic(s): Locational Marginal Price(LMP)

[EMIL (ercot.com)]

Dataset: 10_policy

Electricity policies:

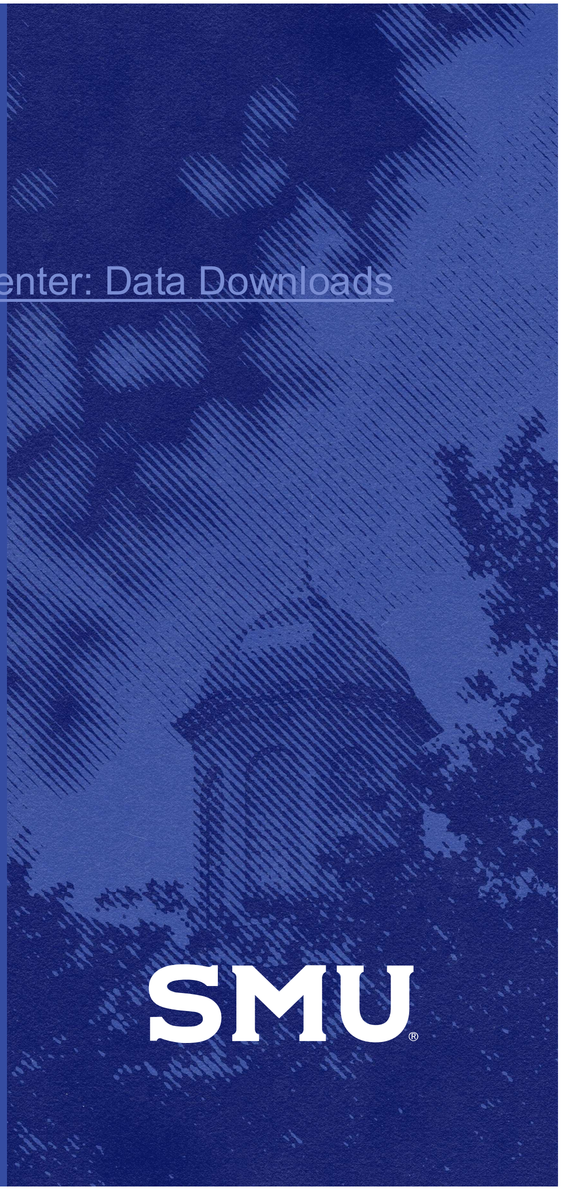
[AEP Texas Energy Incentive Programs | Residential & Commercial (quickelectricity.com)]

EV policies:

[Alternative Fuels Data Center: Data Downloads (energy.gov)]

Energy policies:

[Database of State Incentives for Renewables & Efficiency® - DSIRE (dsireusa.org)]



Prerequisite for Agent based modeling

» Identify key variables and agents:

» key variables and agents to consider:

- Demographic characteristics of residents (e.g. income, age, race, gender)
- Availability and cost of charging infrastructure Statewide Planning Map (txdot.gov)
- Transportation options (e.g. public transit, car ownership)
- Energy usage patterns (e.g. electricity demand) **MDAT (census.gov)**
- Environmental attitudes and awareness (e.g. education)
- Policies and regulations related to EV charging and energy equity

» Find underlying variables and agents:

» Collect as more data as possible

» Conduct statistical/Machine learning modelling to identify latent features/variables

SMU

Ideas

Bigger data

- Collect as much data as possible, since much more data become available

Topics

- Coupling (interaction) effect of electricity and EV policy on equity of EV charging station distribution.
- (Economic Benefits) How advancing EV adoption rate in underrepresented communities can boost local economy growth in the long run. (to come up with more specific question in terms of the first topic)
- Underlying relations among EV charging stations, grid and building energy as well as demographic data.
- More?

New modeling approaches

- Data driven approach
- ML with ABM (later)

Interaction among multiple agents

Metrics/criteria

SMU

Next Steps

Data Collection	• Manual data download • Programmatic data download • Data storage and management
Exploratory Data Analysis (EDA)	• Sanity Check • Statistical analysis (descriptive statistics, inferential statistics, forecasting, A/B testing for different policies) • Further specify the research topic • Find underlying factors using ML • Build predictive and ML models and serve as baseline model for validating ABM
Literature Review	• Conduct paper review while EDA • Propose new ideas or specify topic
Agent Based Modeling	• Refine statistical models and machine learning models as a baseline • Create ABM • Integrate ABM with ML • Validation • Develop real-time dashboard app

Define agents, policies

Validation

SMU

Questions&Discussion

» which geographic area we focus on?

  » Texas Triangle

» Innovation/novelty:

  » new methodology

  » new social/economic/engineering problem

Project Framework

- Cluster Analysis
  - EDA:Quantitative analysis on inequality of EV adoption across different tracts.
    - innovation: combining sociodemographic, economic, political, and techno-infrastructural(EV, Electricity) factor
    - 2022 Inequality and the future of electric mobility in 36 U.S. Cities
  - Cluster analysis to determine factors/attributes affecting equality of EV adoption. Identify different typologies.
    - How inequality is Texas
- Machine Learning
  - Prediction on EV adoption rates across different tracts
    - How factors like electricity policy affect EV purchase decision
    - 2021 Classification of potential electric vehicle purchasers A ML approach
  - Change different policies as input
    - How should we tailor policies to different tracts/typologies, to achieve equity?

- ABM
  - BEAM Core with LBL/City of Dallas
  - C2C program with LBL/DFW
- ML to help agent make decision
  - 2023 A framework proposal for ML-driven ABMs through a case study analysis.pdf

They use micro data, like block, household, need data from city of Dallas and a target community

LO0

Slide 25

LO0 The data is all collected, and I have done EDA for US census. Others are in progress.

Li, Owen, 2023-04-06T21:32:10.687

BEAM activity inputs

Name	Status
activities_location.csv	
blocks (1).csv	
population.csv	
vehicles.csv	
households.csv	
persons.csv	
plans.csv	

blocks.csv

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.

	block_id	y	value_impute	taz_id	square_meters_land	res_rents	rent_impute	residential_unit_capacity	employment_capacity	puma10_id	place_id	x	res_values	TAZ
0	6.00143E+13	37.77027048	1	6.001E+11	0	1475	1	0	0	600105	600562	-122.2338672	777700	576
1	6.00143E+13	37.76946374	0	6.001E+11	79696	1253	0	105	367	600105	600562	-122.2339913	567900	576

households.csv

household_id	gt55	seniors	VEHICL	lcm_cou	hh_child	block_id	gt2	hispanic_	age_of_	race_of_	tenure_m	hh_size	sf_detac	tenure	hh_cars	income	hh_age_	serialno	num_wor	hh_race_
0	1	0	2	6001	no	6E+13	0	no	61	6	own not r	one	yes	1	one	116030	gt35-lt65	2E+12	1	asian
8280	1	0	3	6001	yes	6E+13	1	no	56	1	own not r	three	yes	1	two or mc	215000	gt35-lt65	2E+12	1	white
16534	0	0	1	6001	no	6E+13	0	no	52	1	own not r	one	yes	1	one	110000	gt35-lt65	2E+12	1	white
24796	0	0	2	6001	no	6E+13	1	yes	29	1	rent rece	two	no	2	one	42150	lt35	2E+12	2	white
33076	0	0	1	6001	no	6E+13	0	no	50	2	own not r	one	yes	1	one	78000	gt35-lt65	2E+12	1	black
41326	1	1	2	6001	no	6E+13	1	no	66	2	rent not r	two	yes	2	two or mc	17100	gt65	2E+12	2	black
49597	1	0	1	6001	no	6E+13	0	no	64	2	own not r	one	yes	1	one	9900	gt35-lt65	2E+12	0	black
57872	0	0	3	6001	yes	6E+13	1	no	48	2	rent rece	two	no	2	none	10900	gt35-lt65	2E+12	0	black
66140	1	0	2	6001	no	6E+13	1	no	63	2	own not r	two	yes	1	two or mc	65550	gt35-lt65	2E+12	0	black
74387	0	0	3	6001	yes	6E+13	1	no	35	6	rent rece	four or mc	yes	2	two or mc	100000	gt35-lt65	2E+12	1	asian
82661	0	0	1	6001	no	6E+13	0	no	25	1	rent rece	one	no	2	one	36040	lt35	2E+12	1	white

household_id	hh_incor	recent_m	hh_work	hispanic_	hh_senic	hhsize	hh_type	TAZ	HHT	sample_r	chunk_ic	income_i	income_	median_	hh_value	num_nor	num_driv	num_adu	num_chil	num_you
0	gt100-lt15	0	one	0	no	1	2	1005	1	1	0	116.03	4	12.86	6.6505	0	1	1	0	0
8280	gt150	0	one	0	no	3	4	997	4	1	90	215	4	12.86	22.193	2	2	2	1	0
16534	gt100-lt15	0	one	0	no	1	1	976	1	1	180	110	4	12.86	17.709	0	1	1	0	0
24796	gt30-lt60	1	two or mc	1	no	2	7	925	4	1	270	42.15	2	8.81	4.3446	0	2	2	0	0
33076	gt60-lt10	0	one	0	no	1	1	907	1	1	360	78	3	10.44	13.113	0	1	1	0	0
41326	lt30	0	two or mc	0	yes	2	8	889	4	1	450	17.1	1	6.01	1.8303	0	2	2	0	0
49597	lt30	0	none	0	no	1	2	894	1	1	540	9.9	1	6.01	7.1456	1	1	1	0	0
57872	lt30	1	none	0	no	2	7	888	4	1	630	10.9	1	6.01	3.6756	2	1	1	1	0
66140	gt60-lt10	0	none	0	no	2	4	880	4	1	720	65.55	3	10.44	10.554	2	2	2	0	0
74387	gt100-lt15	1	one	0	no	4	7	1034	4	1	810	100	3	10.44	5.0004	3	2	2	2	2
82661	gt30-lt60	1	one	0	no	1	5	1027	1	1	900	36.04	2	8.81	9.9952	0	1	1	0	0

population.csv

personId	age	isFemale	householdId	householdRank	excludedModes	valueOfTime
57953	6	FALSE	2584691	0		7.25
57954	33	TRUE	2584691	0		7.25
57955	14	TRUE	2584691	0		7.25
57956	10	TRUE	2584691	0		7.25
57957	4	TRUE	2584691	0		7.25
97777	56	FALSE	2548888	0		7.25
161860	61	TRUE	2165040	0		7.25
366699	69	TRUE	1787204	0		7.25
366700	69	FALSE	1787204	0		7.25
574090	30	FALSE	1596570	0		7.25
574092	5	FALSE	1596570	0		7.25
574094	43	FALSE	1596570	0		7.25
574095	43	FALSE	1596570	0		7.25
574096	52	TRUE	1596570	0		7.25
621207	72	TRUE	354085	0		7.25

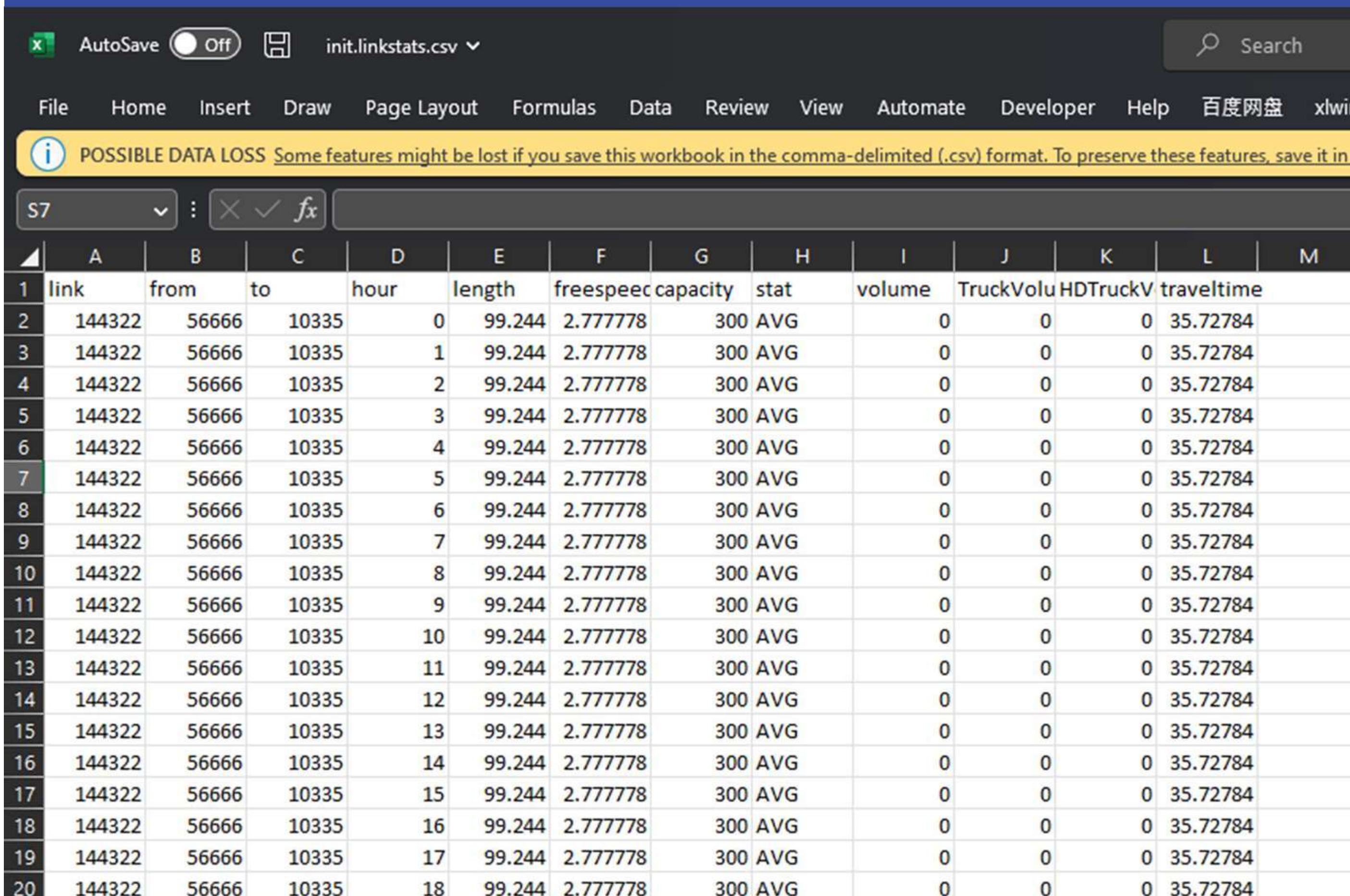
activities_location.csv

person_id	ActivityType	x	y	household_id	TAZ
0	Home	-122.2376959	37.8611524	0	1005
0	work	-122.0669131	37.33911771	0	1005
0	Home	-122.2376959	37.8611524	0	1005
0	shopping	-122.2708216	37.86477182	0	1005
0	Home	-122.2376959	37.8611524	0	1005
1	Home	-121.8064398	37.3503036	1735309	641
1	othmaint	-121.8428209	37.35949054	1735309	641
1	school	-121.8337329	37.37625817	1735309	641
1	Home	-121.8064398	37.3503036	1735309	641
2	Home	-121.8098355	37.34821151	1735309	641
2	work	-121.8064398	37.3503036	1735309	641
2	atwork	-121.6415392	37.12275338	1735309	641
2	Work	-121.6306659	37.11518106	1735309	641
2	othmaint	-121.8387512	37.36208136	1735309	641
2	Home	-121.8098355	37.34821151	1735309	641

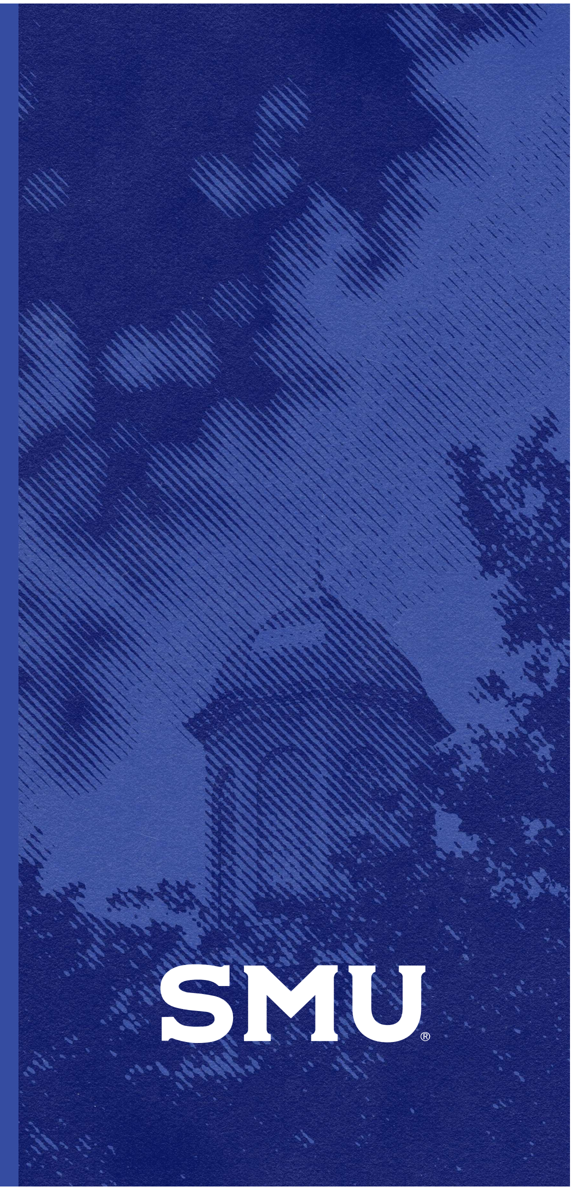
vehicles.csv

vehicleId	vehicleTypeId	stateOfCharge	householdId
0	diesel-L1-10000-to-25000-LowTech-2019		2580575
1	conv-L1-10000-to-25000-LowTech-2019		2580575
2	conv-L1-10000-to-25000-LowTech-2019		2580575
3	BIKE-DEFAULT		2580575
4	conv-L1-10000-to-25000-LowTech-2019		1572860
5	BIKE-DEFAULT		1572860
6	conv-L1-10000-to-25000-LowTech-2019		2732496
7	conv-L1-10000-to-25000-LowTech-2019		2732496
8	conv-L1-10000-to-25000-LowTech-2019		2732496
9	BIKE-DEFAULT		2732496
10	conv-L1-10000-to-25000-LowTech-2019		2529127
11	conv-L1-10000-to-25000-LowTech-2019		2529127

BEAM network inputs



link	from	to	hour	length	freespeec	capacity	stat	volume	TruckVolu	HDTruckV	traveltime
144322	56666	10335	0	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	1	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	2	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	3	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	4	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	5	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	6	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	7	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	8	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	9	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	10	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	11	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	12	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	13	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	14	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	15	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	16	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	17	99.244	2.777778	300	AVG	0	0	0	35.72784
144322	56666	10335	18	99.244	2.777778	300	AVG	0	0	0	35.72784



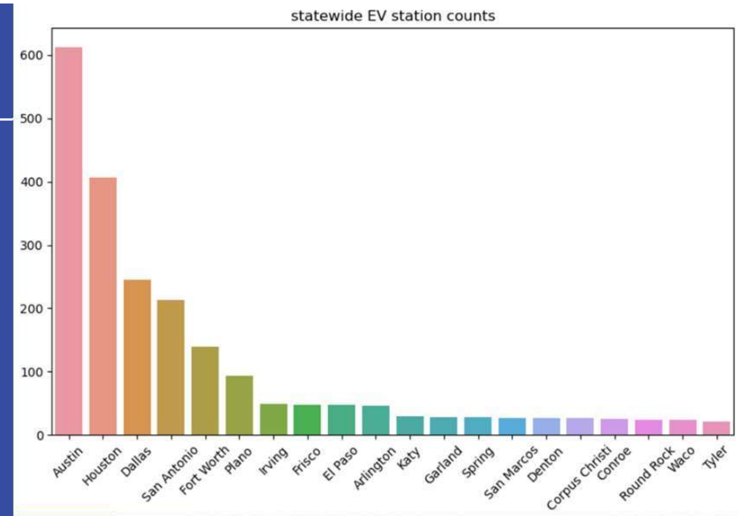
Finalize Data Source

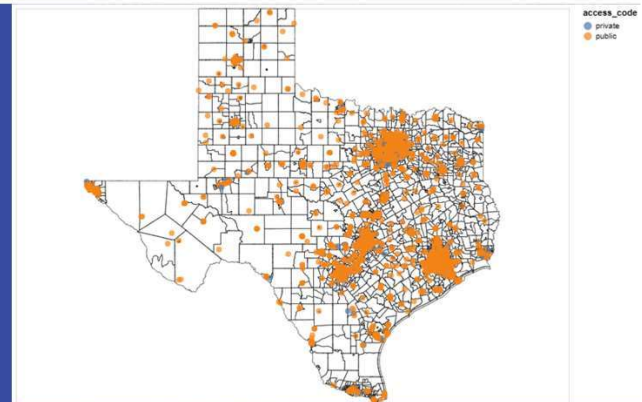
» 2021 ACS social demographic data at tract level:

  » Kmean, hierarchical clustering, DBSCAN (silhouette score: below 0.05)

  » Nomailization (silhouette score: 0.15)

  » Dimension reduction (PCA) (silhouette score: ??)





» EV stations

» EV registration(light/medium/heavy duty)(not associate with location/tract/household)

» EV Law and incentives

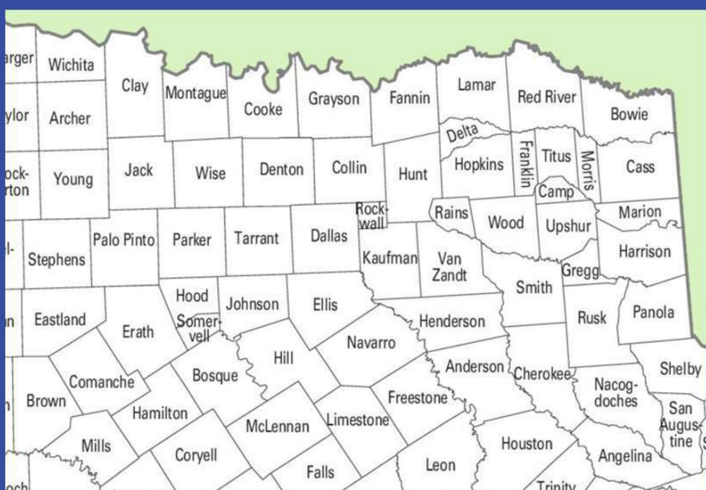
» Electricity incentives

» Utility Rates by tract

» Twitter/facebook/Elsvier =>how people talk about EV policies.

Dallas-Fort Worth

- Narrow down to DFW area(Dallas, Tarrant, Collin, Denton)
  - City of Dallas could provide more micro data if needed
  - Utility Rates by Census Region API has limits
  - DFW (Dallas, Tarrant, Collin, Denton) has 1507 tracts; Texas has 6896.



Questions to be answered

- Study how a certain incentive affects EV adoption in similar typologies at the same duration, using hypothesis testing
- What is the relationship between micro-typology and EV adoption(demand)?
- What is the correlations between EV adoption(demand) and incentives?
- How the same incentive affects different groups of people?
- Which combination of transportation and electricity policy would be most effective?
- Topic from City of Dallas/DFWcleancities
- ABM using BEAM, integrated RL/ML (course?)

5.2. Remaining Research Gaps

There are many areas remaining for further research on the impacts of managed charging on the grid, including:

- Testing different PEV adoption forecasts and different PEV fleet composition (e.g. vehicles with longer range).
- Testing different charging infrastructure scenarios, including the emphasis on fast versus slow charging, and added workplace charging infrastructure.
- Testing more accurate estimation of charging power constraints of the varying available charging infrastructure.
- Using California and/or National Household Travel Survey data to scale PEV charging demand and flexibility in a manner that reflects regional variations in mobility and charging infrastructure.
- Finding correlations between charging demand and mobility profiles (i.e. daily VMT) and including these relationships when scaling demand.
- Simulating the participation of aggregated PEV fleets in other grid services such as regulation and load-following through vehicle-to-grid.
- Testing different renewable renewable generation mixes.
- Testing the impact of competing sources of grid flexibility including increased storage and demand response, varied curtailment assumptions, and higher net export limits.

Finally, there are also many policy changes happening concurrently in California and WECC, which could impact the conclusions of this study. For example, California is already coordinating with neighboring balancing areas through the Energy Imbalance Market, which could alleviate some of the curtailment problems highlighted here [75]. CAISO may also expand to other parts of WECC, and there may be an increase in DR and load management from other end-uses besides PEVs to cope with curtailment. Lastly, there is a push to move residential electric customers in California to opt-out TOU rates in the next few years [83], which may incentivize load shifting during these curtailment periods, without the use of actively managed PEVs.

Next: may 10

Examine how many EV and EV stations typologies have?

» Typology mapping by Tract ID

» EV stations mapping by Location

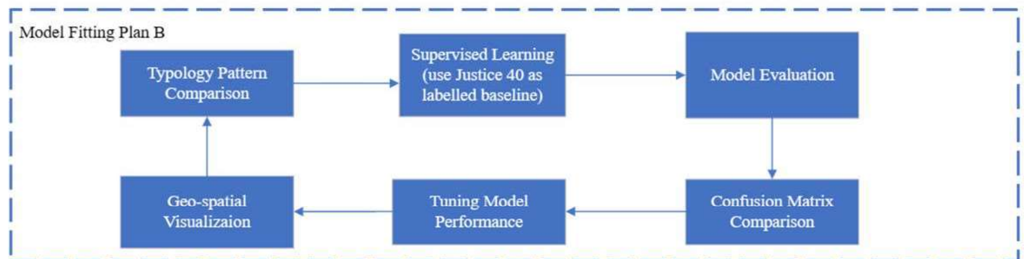
» EV registration by zipcode

Cluster analysis process diagram

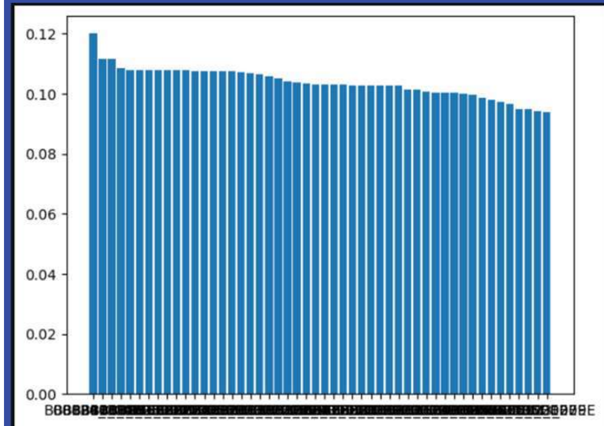
- Data Collection:
  - Dataset: Social Demographic dataset, Electricity dataset, EV registration dataset;
  - Geospatial mapping: Geospatial shapefile, EV charging station location;

- Feature Engineering: the input data are high dimensional; we want to extract the most impactful features/attributes. We plan to use different dimension reduction techniques, to transform and select the features per the importance factor.

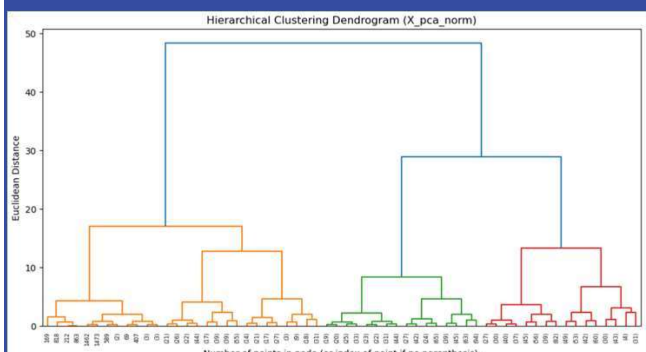
- Modelling fitting:
  - conduct the cluster analysis with consideration of the top important features;
  - Evaluate the clustering results via silhouette analysis;
  - Tuning hyperparameters to improve the results;
  - Visualize the clusters in the maps;
  - Compare different clustering results via patterns;
- Quantifying spatial equity level:
  - comparison between typologies pattern and EVCS placements;
  - develop equity metrics;
  - visualize the variations in equity level

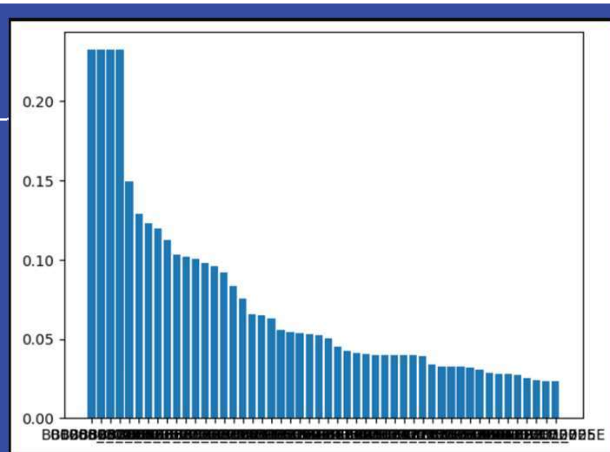


- Develop multi-agent reinforcement learning algorithms to optimize the placement decision making:
  - The Multi-Agent System consists of macro and micro level agents, namely, DFW area agent and tract agents;
  - Design rewards based on the equity metrics;
  - Simulate the decision making process;
  - Discover and understand underlying trending, and make recommendation for policymakers

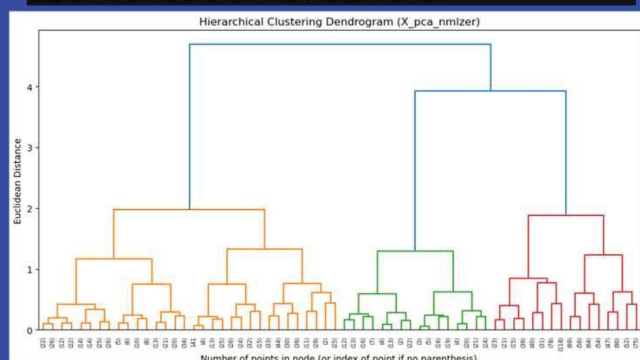


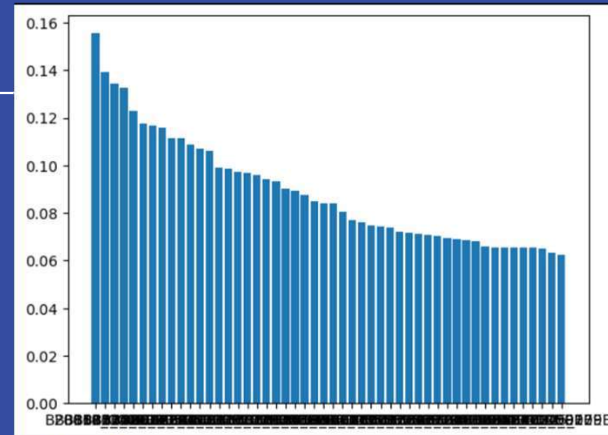
```
B08124 : MEANS OF TRANSPORTATION TO WORK BY OCCUPATION
B08301 : MEANS OF TRANSPORTATION TO WORK
B08006 : SEX OF WORKERS BY MEANS OF TRANSPORTATION TO WORK
B08006 : SEX OF WORKERS BY MEANS OF TRANSPORTATION TO WORK
B11001 : HOUSEHOLD TYPE (INCLUDING LIVING ALONE)
B28010 : COMPUTERS IN HOUSEHOLD
B19001 : HOUSEHOLD INCOME IN THE PAST 12 MONTHS (IN 2021 INFLATION-ADJUSTED DOLLARS)
B28011 : INTERNET SUBSCRIPTIONS IN HOUSEHOLD
B25132 : MONTHLY ELECTRICITY COSTS
B25133 : MONTHLY GAS COSTS
B08122 : MEANS OF TRANSPORTATION TO WORK BY POVERTY STATUS IN THE PAST 12 MONTHS
B08126 : MEANS OF TRANSPORTATION TO WORK BY INDUSTRY
B08124 : MEANS OF TRANSPORTATION TO WORK BY OCCUPATION
B08006 : SEX OF WORKERS BY MEANS OF TRANSPORTATION TO WORK
B08301 : MEANS OF TRANSPORTATION TO WORK
B25132 : MONTHLY ELECTRICITY COSTS
B28010 : COMPUTERS IN HOUSEHOLD
B28011 : INTERNET SUBSCRIPTIONS IN HOUSEHOLD
B08303 : TRAVEL TIME TO WORK
```



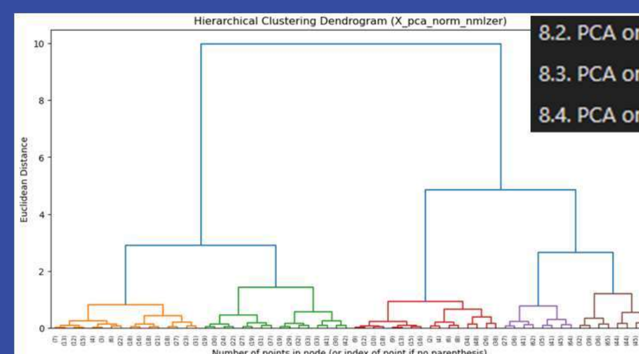


```
B01003 : TOTAL POPULATION
B02001 : RACE
B05003 : SEX BY AGE BY NATIVITY AND CITIZENSHIP STATUS
B05001 : NATIVITY AND CITIZENSHIP STATUS IN THE UNITED STATES
B02001 : RACE
B05001 : NATIVITY AND CITIZENSHIP STATUS IN THE UNITED STATES
B02001 : RACE
B05003 : SEX BY AGE BY NATIVITY AND CITIZENSHIP STATUS
B05003 : SEX BY AGE BY NATIVITY AND CITIZENSHIP STATUS
B05003 : SEX BY AGE BY NATIVITY AND CITIZENSHIP STATUS
B15003 : EDUCATIONAL ATTAINMENT FOR THE POPULATION 25 YEARS AND OVER
B05001 : NATIVITY AND CITIZENSHIP STATUS IN THE UNITED STATES
B05003 : SEX BY AGE BY NATIVITY AND CITIZENSHIP STATUS
B05003 : SEX BY AGE BY NATIVITY AND CITIZENSHIP STATUS
B05003 : SEX BY AGE BY NATIVITY AND CITIZENSHIP STATUS
B02001 : RACE
B02001 : RACE
B08124 : MEANS OF TRANSPORTATION TO WORK BY OCCUPATION
B05003 : SEX BY AGE BY NATIVITY AND CITIZENSHIP STATUS
```





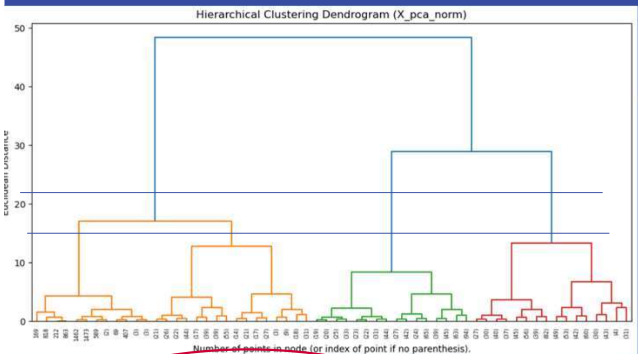
```
B28010 : COMPUTERS IN HOUSEHOLD
B08122 : MEANS OF TRANSPORTATION TO WORK BY POVERTY STATUS IN THE PAST 12 MONTHS
B15003 : EDUCATIONAL ATTAINMENT FOR THE POPULATION 25 YEARS AND OVER
B08124 : MEANS OF TRANSPORTATION TO WORK BY OCCUPATION
B15003 : EDUCATIONAL ATTAINMENT FOR THE POPULATION 25 YEARS AND OVER
B17026 : RATIO OF INCOME TO POVERTY LEVEL OF FAMILIES IN THE PAST 12 MONTHS
B15003 : EDUCATIONAL ATTAINMENT FOR THE POPULATION 25 YEARS AND OVER
B08124 : MEANS OF TRANSPORTATION TO WORK BY OCCUPATION
B08122 : MEANS OF TRANSPORTATION TO WORK BY POVERTY STATUS IN THE PAST 12 MONTHS
B08124 : MEANS OF TRANSPORTATION TO WORK BY OCCUPATION
B08124 : MEANS OF TRANSPORTATION TO WORK BY OCCUPATION
B28011 : INTERNET SUBSCRIPTIONS IN HOUSEHOLD
B02001 : RACE
B15003 : EDUCATIONAL ATTAINMENT FOR THE POPULATION 25 YEARS AND OVER
B19001 : HOUSEHOLD INCOME IN THE PAST 12 MONTHS (IN 2021 INFLATION-ADJUSTED DOLLARS)
B08126 : MEANS OF TRANSPORTATION TO WORK BY INDUSTRY
B28010 : COMPUTERS IN HOUSEHOLD
B11001 : HOUSEHOLD TYPE (INCLUDING LIVING ALONE)
```

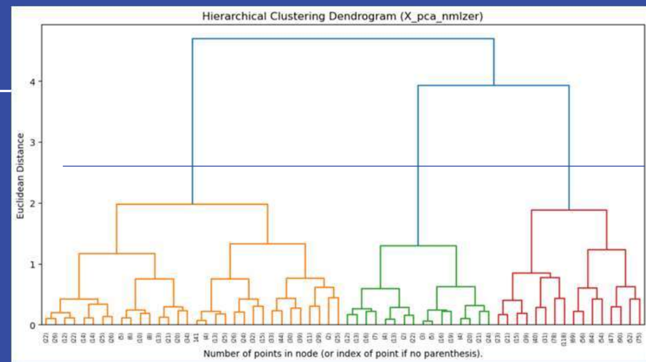


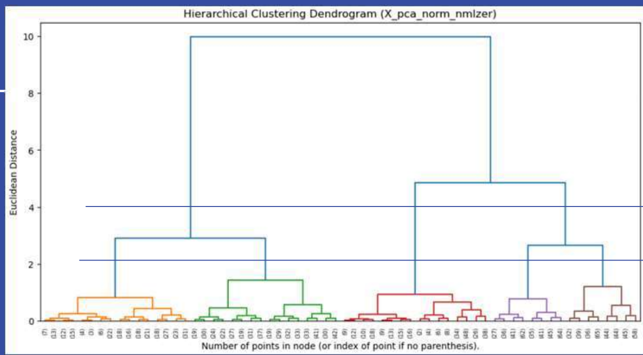
8.2. PCA on X_std

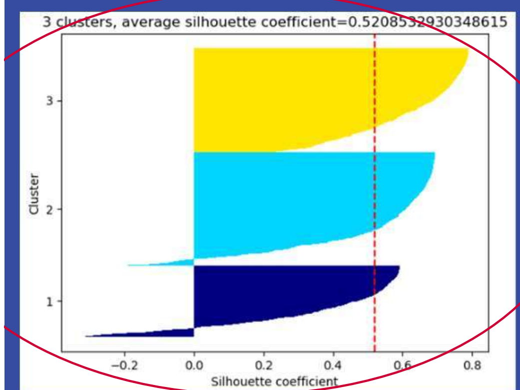
8.3. PCA on X_nmlzer

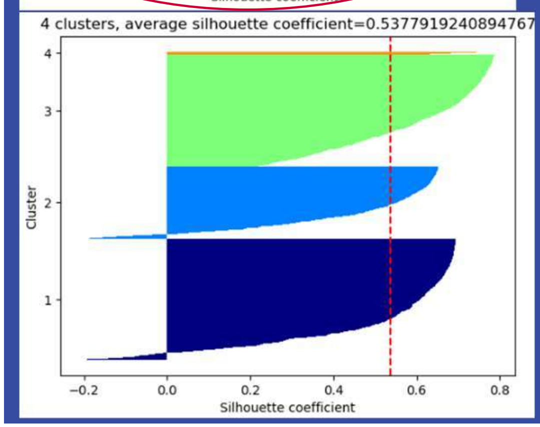
8.4. PCA on X_norm_nm

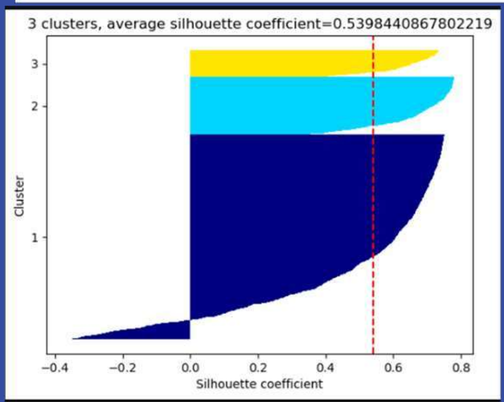


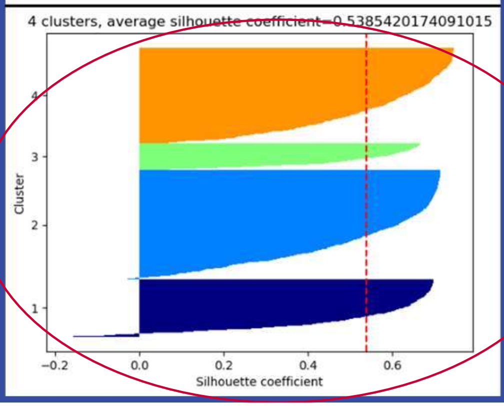


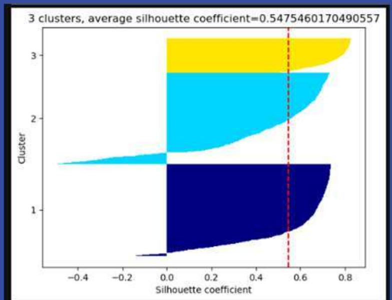


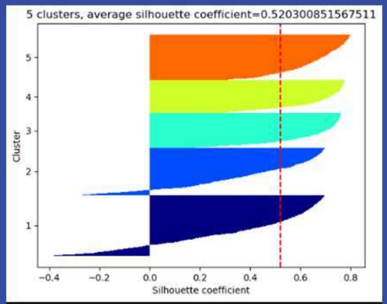


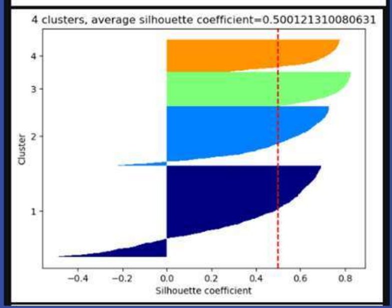


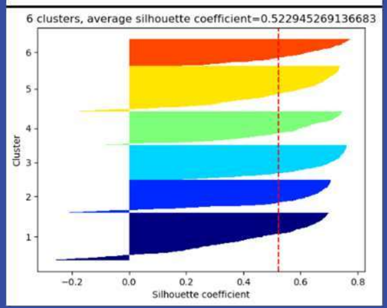




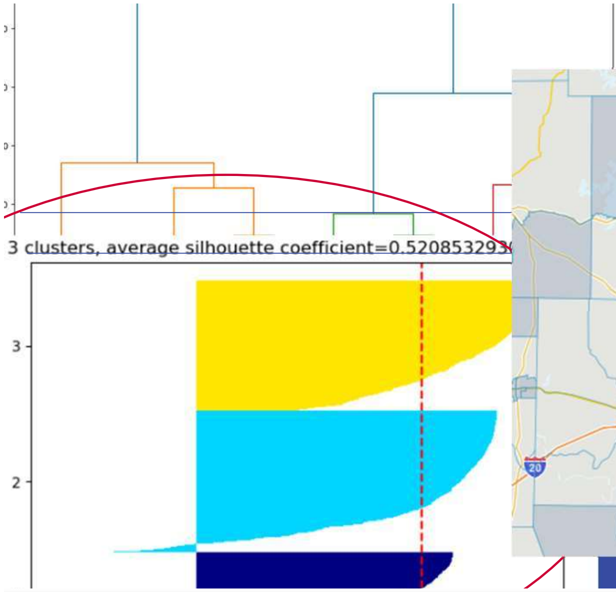


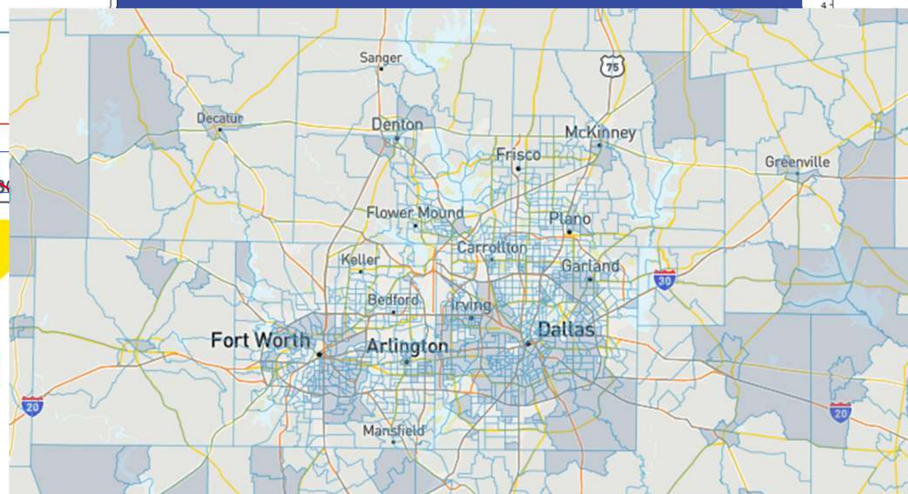


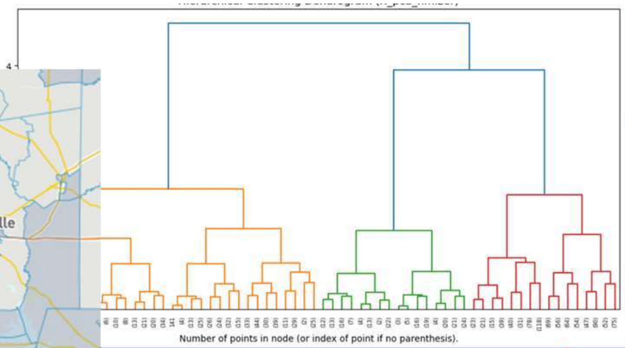




Other papers: less score, less features



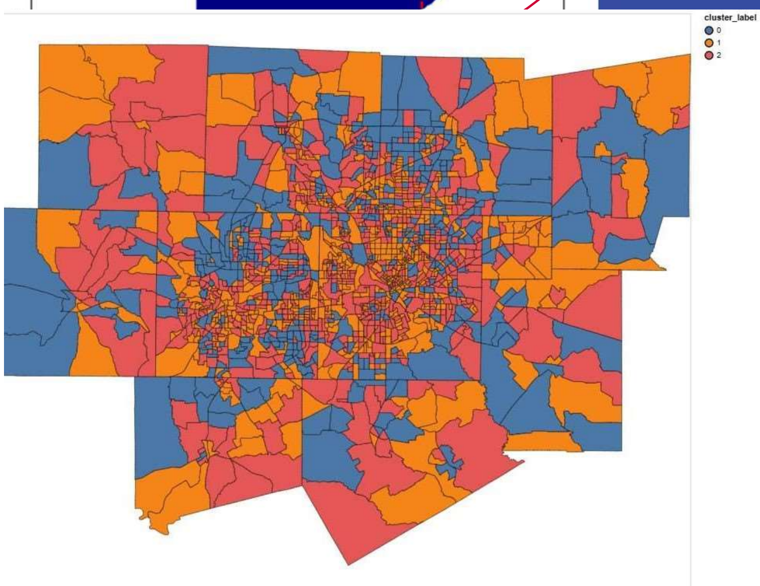


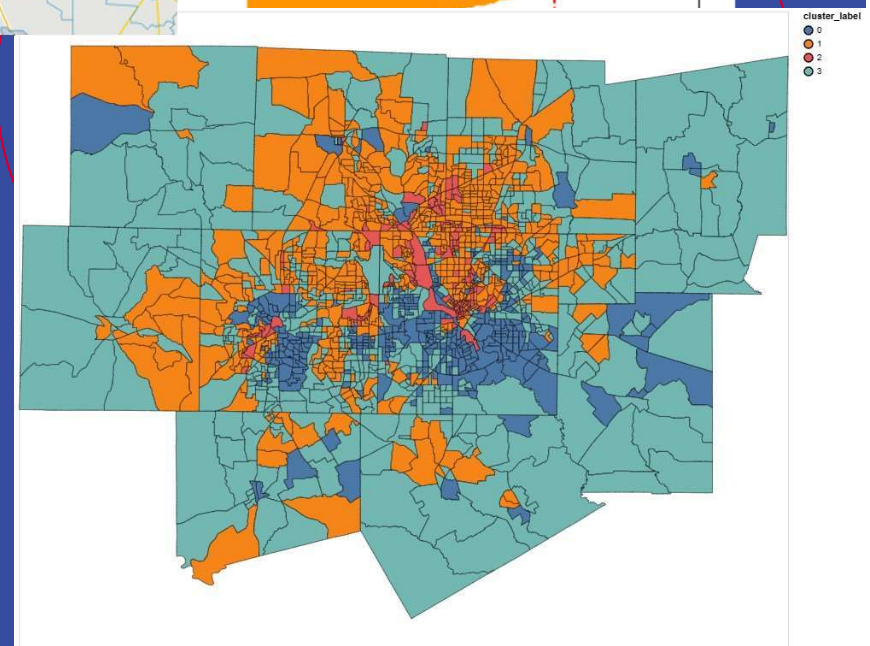


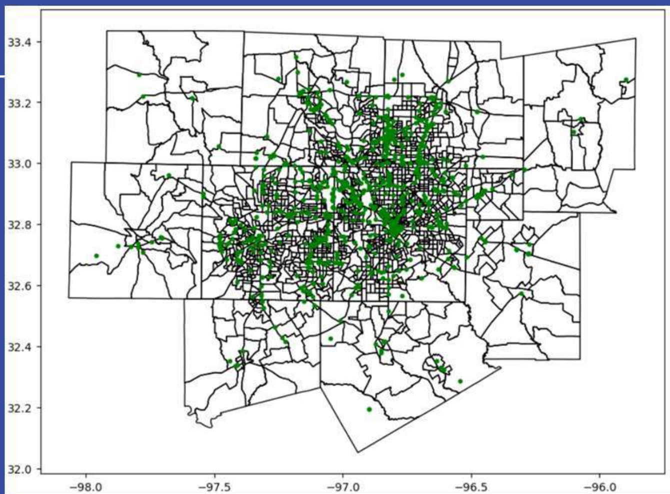
Justice 40

What are the social, economic, demographic, and techno factors that can quantify spatial inequity in EVCS placement?

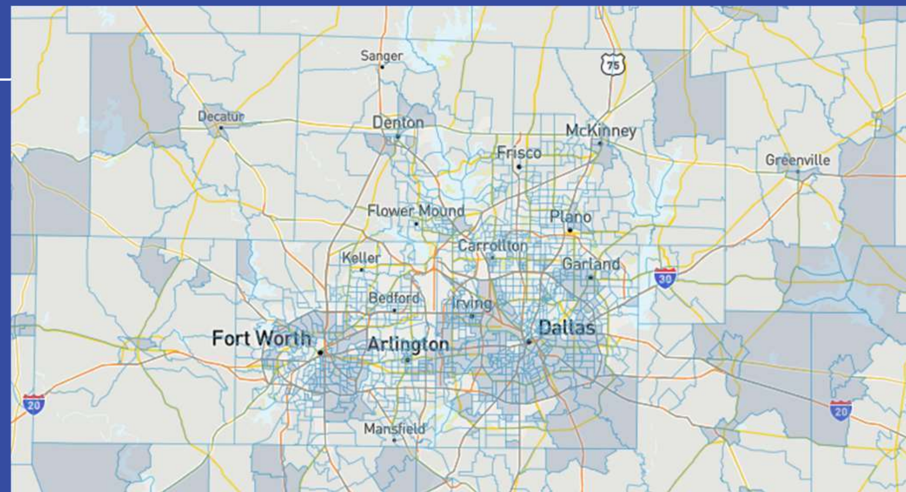
Justice 40 is based on different factors



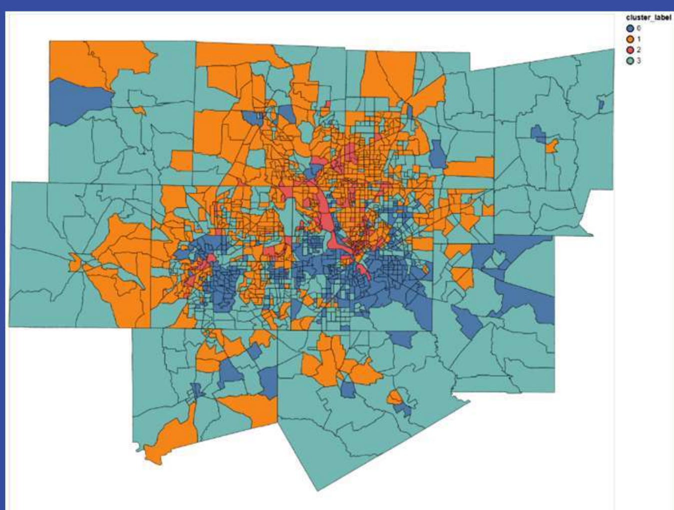


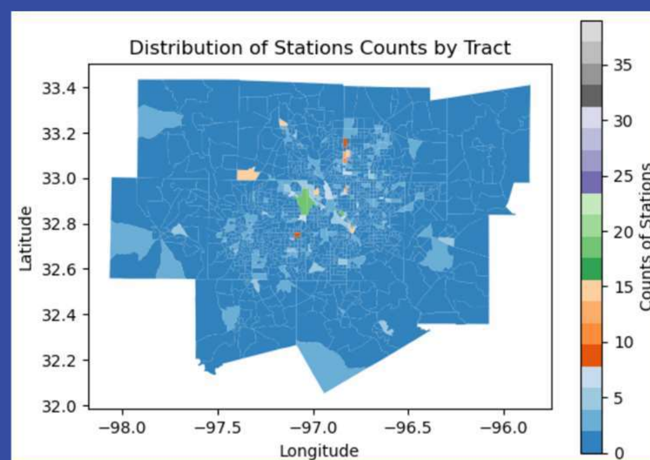


EVCS location



Justice 40 defined disadvantaged tracts





Example Typologies

EVCS counts by Tract

We may use our typology and/or Justice40

Is Justice40 a good fit for solving our problem, we don't know

In the end, we might have potential findings, like if using Justice 40 is good for EVCS placing decision making? Or the advantages of our typology

Feature importance from different algorithms

top_corr_attr(threshold = 0.2, correlation_matrix = merged_X_norm_EVCScnt.corr())

[54]:

	counts	Desc
counts	1.000000	0
B08122_020E	0.256632	[[Estimate, Total:, Walked:, At or above 150 percent of the poverty level]]
B08126_070E	0.249319	69 [Estimate, Total:, Walked:, Professional, scientific, and management, and administrative and waste management services] Name: Label, dtype: object
B08122_017E	0.243381	16 [Estimate, Total:, Walked:] Name: Label, dtype: object
B11001_007E	0.242668	6 [Estimate, Total:, Nonfamily households:] Name: Label, dtype: object
B08124_030E	0.241642	29 [Estimate, Total:, Walked:, Management, business, science, and arts occupations] Name: Label, dtype: object
B25132_005E	0.241469	4 [Estimate, Total, Charged for electricity, 50*to*99] Name: Label, dtype: object
B11001_008E	0.236092	7 [Estimate, Total:, Nonfamily households:, Householder living alone] Name: Label, dtype: object
B08006_032E	0.224437	31 [Estimate, Total:, Male:, Walked] Name: Label, dtype: object
B25132_004E	0.207955	3 [Estimate, Total, Charged for electricity, Less than $50] Name: Label, dtype: object
B08006_015E	0.202705	14 [Estimate, Total:, Walked] Name: Label, dtype: object
B08124_029E	0.202705	28 [Estimate, Total:, Walked:] Name: Label, dtype: object
B08126_061E	0.202705	60 [Estimate, Total:, Walked:] Name: Label, dtype: object
B08301_019E	0.202705	18 [Estimate, Total:, Walked] Name: Label, dtype: object

Correlation:

Correlation measures the statistical relationship between two variables. In the context of feature selection, you can compute the correlation between each feature and the target variable. Features with higher absolute correlation values (positive or negative) are considered more important.

Pros: Simple to compute, provides insight into linear relationships.

Cons: Only captures linear relationships, may miss complex interactions.

SMU

Feature importance from different algorithms

SelectKbest: f_regression

	Top Feature	score	p_values	Desc	Concept
4	B08126_070E	94.019891	1.507031e-21	69 [Estimate, Total:, Walked:, Professional, scientific, and management, and administrative and waste management services] Name: Label, dtype: object	69 MEANS OF TRANSPORTATION TO WORK BY INDUSTRY Name: Concept, dtype: object
3	B08124_030E	93.725772	1.731899e-21	29 [Estimate, Total:, Walked:, Management, business, science, and arts occupations] Name: Label, dtype: object	29 MEANS OF TRANSPORTATION TO WORK BY OCCUPATION Name: Concept, dtype: object
2	B08122_020E	87.892405	2.750332e-20	19 [Estimate, Total:, Walked:, At or above 150 percent of the poverty level] Name: Label, dtype: object	19 MEANS OF TRANSPORTATION TO WORK BY POVERTY STATUS IN THE PAST 12 MONTHS Name: Concept, dtype: object
1	B08122_017E	77.030604	4.905106e-18	16 [Estimate, Total:, Walked:] Name: Label, dtype: object	16 MEANS OF TRANSPORTATION TO WORK BY POVERTY STATUS IN THE PAST 12 MONTHS Name: Concept, dtype: object
6	B11001_007E	73.306123	2.933156e-17	6 [Estimate, Total:, Nonfamily households:] Name: Label, dtype: object	6 HOUSEHOLD TYPE (INCLUDING LIVING ALONE) Name: Concept, dtype: object
7	B11001_008E	70.450944	1.159946e-16	7 [Estimate, Total:, Nonfamily households:, Householder living alone] Name: Label, dtype: object	7 HOUSEHOLD TYPE (INCLUDING LIVING ALONE) Name: Concept, dtype: object
9	B25132_005E	68.930354	2.415743e-16	4 [Estimate, Total, Charged for electricity, 50*to*99] Name: Label, dtype: object	4 MONTHLY ELECTRICITY COSTS Name: Concept, dtype: object
0	B08006_032E	64.549367	2.011133e-15	31 [Estimate, Total:, Male:, Walked] Name: Label, dtype: object	31 SEX OF WORKERS BY MEANS OF TRANSPORTATION TO WORK Name: Concept, dtype: object
8	B25132_004E	58.219604	4.363616e-14	3 [Estimate, Total, Charged for electricity, Less than $50] Name: Label, dtype: object	3 MONTHLY ELECTRICITY COSTS Name: Concept, dtype: object
5	B08301_019E	53.278933	4.882910e-13	18 [Estimate, Total:, Walked] Name: Label, dtype: object	18 MEANS OF TRANSPORTATION TO WORK Name: Concept, dtype: object

SelectKbest: f_classif

	Top Feature	score	p_values	Desc	Conc
4	B08126_070E	94.019891	1.507031e-21	69 [Estimate, Total:, Walked:, Professional, scientific, and management, and administrative and waste management services] Name: Label, dtype: object	69 MEANS OF TRANSPORTATION TO WORK INDUSTRY Name: Concept, dtype: obj
3	B08124_030E	93.725772	1.731899e-21	29 [Estimate, Total:, Walked:, Management, business, science, and arts occupations] Name: Label, dtype: object	29 MEANS OF TRANSPORTATION TO WORK OCCUPATION Name: Concept, dtype: obj
2	B08122_020E	87.892405	2.750332e-20	19 [Estimate, Total:, Walked:, At or above 150 percent of the poverty level] Name: Label, dtype: object	19 MEANS OF TRANSPORTATION TO WORK POVERTY STATUS IN THE PAST 12 MONT Name: Concept, dtype: obj
1	B08122_017E	77.030604	4.905106e-18	16 [Estimate, Total:, Walked:] Name: Label, dtype: object	16 MEANS OF TRANSPORTATION TO WORK POVERTY STATUS IN THE PAST 12 MONT Name: Concept, dtype: obj
6	B11001_007E	73.306123	2.933156e-17	6 [Estimate, Total:, Nonfamily households:] Name: Label, dtype: object	6 HOUSEHOLD TYPE (INCLUDING LIVI ALONE) Name: Concept, dtype: obj
7	B11001_008E	70.450944	1.159946e-16	7 [Estimate, Total:, Nonfamily households:, Householder living alone] Name: Label, dtype: object	7 HOUSEHOLD TYPE (INCLUDING LIVI ALONE) Name: Concept, dtype: obj
9	B25132_005E	68.930354	2.415743e-16	4 [Estimate, Total, Charged for electricity, 50*to*99] Name: Label, dtype: object	4 MONTHLY ELECTRICITY COSTS Nar Concept, dtype: obj
0	B08006_032E	64.549367	2.011133e-15	31 [Estimate, Total:, Male:, Walked] Name: Label, dtype: object	31 SEX OF WORKERS BY MEANS TRANSPORTATION TO WORK Name: Conce dtype: obj
8	B25132_004E	58.219604	4.363616e-14	3 [Estimate, Total, Charged for electricity, Less than $50] Name: Label, dtype: object	3 MONTHLY ELECTRICITY COSTS Nar Concept, dtype: obj
5	B08301_019E	53.278933	4.882910e-13	18 [Estimate, Total:, Walked] Name: Label, dtype: object	18 MEANS OF TRANSPORTATION TO WO Name: Concept, dtype: obj

SelectKBest Score (ANOVA or Chi-Squared):

SelectKBest uses statistical tests (ANOVA F-values or chi-squared) to evaluate the relationship between each feature and the target variable. Features with higher scores are more likely to be informative for the target variable.

Pros: Systematic and principled approach, can handle categorical features.

Cons: May not capture complex non-linear relationships.

XGBoost

[321]:

	Top Feature	Importance	Desc	Concept
0	B05003_020E	0.079453	19 [Estimate, Total:, Female:, 18 years and over:, Native] Name: Label, dtype: object	19 SEX BY AGE BY NATIVITY AND CITIZENSHIP STATUS Name: Concept, dtype: object
1	B05001_002E	0.063960	1 [Estimate, Total:, U.S. citizen, born in the United States] Name: Label, dtype: object	1 NATIVITY AND CITIZENSHIP STATUS IN THE UNITED STATES Name: Concept, dtype: object
2	B08126_070E	0.037749	69 [Estimate, Total:, Walked:, Professional, scientific, and management, and administrative and waste management services] Name: Label, dtype: object	69 MEANS OF TRANSPORTATION TO WORK BY INDUSTRY Name: Concept, dtype: object
3	B08126_097E	0.037280	96 [Estimate, Total:, Worked from home:, Transportation and warehousing, and utilities] Name: Label, dtype: object	96 MEANS OF TRANSPORTATION TO WORK BY INDUSTRY Name: Concept, dtype: object
4	B08124_030E	0.034620	29 [Estimate, Total:, Walked:, Management, business, science, and arts occupations] Name: Label, dtype: object	29 MEANS OF TRANSPORTATION TO WORK BY OCCUPATION Name: Concept, dtype: object
5	B08126_041E	0.028934	40 [Estimate, Total:, Car, truck, or van - carpooled:, Educational services, and health care and social assistance] Name: Label, dtype: object	40 MEANS OF TRANSPORTATION TO WORK BY INDUSTRY Name: Concept, dtype: object
6	B11001_008E	0.024265	7 [Estimate, Total:, Nonfamily households:, Householder living alone] Name: Label, dtype: object	7 HOUSEHOLD TYPE (INCLUDING LIVING ALONE) Name: Concept, dtype: object
7	B08303_009E	0.024017	8 [Estimate, Total:, 35 to 39 minutes] Name: Label, dtype: object	8 TRAVEL TIME TO WORK Name: Concept, dtype: object
8	B05003_007E	0.022446	6 [Estimate, Total:, Male:, Under 18 years:, Foreign born:, Not a U.S. citizen] Name: Label, dtype: object	6 SEX BY AGE BY NATIVITY AND CITIZENSHIP STATUS Name: Concept, dtype: object
9	B28011_003E	0.020557	2 [Estimate, Total:, With an Internet subscription, Dial-up alone] Name: Label, dtype: object	2 INTERNET SUBSCRIPTIONS IN HOUSEHOLD Name: Concept, dtype: object

rf_FI(flag="reg")

			Desc	Concept
0	B08126_041E	0.045506	40 [Estimate, Total:, Car, truck, or van - carpooled:, Educational services, and health care and social assistance] Name: Label, dtype: object	40 MEANS OF TRANSPORTATION TO WORK BY INDUSTRY Name: Concept, dtype: object
1	B08126_070E	0.041445	69 [Estimate, Total:, Walked:, Professional, scientific, and management, and administrative and waste management services] Name: Label, dtype: object	69 MEANS OF TRANSPORTATION TO WORK BY INDUSTRY Name: Concept, dtype: object
2	B28011_003E	0.028419	2 [Estimate, Total:, With an Internet subscription, Dial-up alone] Name: Label, dtype: object	2 INTERNET SUBSCRIPTIONS IN HOUSEHOLD Name: Concept, dtype: object
3	B11001_007E	0.027914	6 [Estimate, Total:, Nonfamily households:] Name: Label, dtype: object	6 HOUSEHOLD TYPE (INCLUDING LIVING ALONE) Name: Concept, dtype: object
4	B11001_008E	0.027084	7 [Estimate, Total:, Nonfamily households:, Householder living alone] Name: Label, dtype: object	7 HOUSEHOLD TYPE (INCLUDING LIVING ALONE) Name: Concept, dtype: object
5	B08126_074E	0.024969	73 [Estimate, Total:, Walked:, Public administration] Name: Label, dtype: object	73 MEANS OF TRANSPORTATION TO WORK BY INDUSTRY Name: Concept, dtype: object
6	B05001_002E	0.022820	1 [Estimate, Total:, U.S. citizen, born in the United States] Name: Label, dtype: object	1 NATIVITY AND CITIZENSHIP STATUS IN THE UNITED STATES Name: Concept, dtype: object
7	B25132_005E	0.021347	4 [Estimate, Total, Charged for electricity, 50to99] Name: Label, dtype: object	4 MONTHLY ELECTRICITY COSTS Name: Concept, dtype: object
8	B08126_026E	0.020744	25 [Estimate, Total:, Car, truck, or van - drove alone:, Educational services, and health care and social assistance] Name: Label, dtype: object	25 MEANS OF TRANSPORTATION TO WORK BY INDUSTRY Name: Concept, dtype: object
9	B08124_030E	0.016790	29 [Estimate, Total:, Walked:, Management, business, science, and arts occupations] Name: Label, dtype: object	29 MEANS OF TRANSPORTATION TO WORK BY OCCUPATION Name: Concept, dtype: object

LightBoost

Pros: Captures non-linear relationships, interactions, and feature dependencies.

	Top Feature	Importance	Desc	Concept
0	B08126_024E	17	23 [Estimate, Total:, Car, truck, or van - drove alone:, Finance and insurance, and real estate and rental and leasing] Name: Label, dtype: object	23 MEANS OF TRANSPORTATION TO WORK BY INDUSTRY Name: Concept, dtype: object
1	B08303_003E	15	2 [Estimate, Total:, 5 to 9 minutes] Name: Label, dtype: object	2 TRAVEL TIME TO WORK Name: Concept, dtype: object
2	B08124_030E	11	29 [Estimate, Total:, Walked:, Management, business, science, and arts occupations] Name: Label, dtype: object	29 MEANS OF TRANSPORTATION TO WORK BY OCCUPATION Name: Concept, dtype: object
3	B08126_074E	9	73 [Estimate, Total:, Walked:, Public administration] Name: Label, dtype: object	73 MEANS OF TRANSPORTATION TO WORK BY INDUSTRY Name: Concept, dtype: object
4	B05003_004E	9	3 [Estimate, Total:, Male:, Under 18 years:, Native] Name: Label, dtype: object	3 SEX BY AGE BY NATIVITY AND CITIZENSHIP STATUS Name: Concept, dtype: object
5	B15003_025E	8	24 [Estimate, Total:, Doctorate degree] Name: Label, dtype: object	24 EDUCATIONAL ATTAINMENT FOR THE POPULATION 25 YEARS AND OVER Name: Concept, dtype: object
6	B08126_025E	8	24 [Estimate, Total:, Car, truck, or van - drove alone:, Professional, scientific, and management, and administrative and waste management services] Name: Label, dtype: object	24 MEANS OF TRANSPORTATION TO WORK BY INDUSTRY Name: Concept, dtype: object
7	B25132_009E	8	8 [Estimate, Total, Charged for electricity, $250 or more] Name: Label, dtype: object	8 MONTHLY ELECTRICITY COSTS Name: Concept, dtype: object
8	B19001_013E	8	12 [Estimate, Total:, 75,000to99,999] Name: Label, dtype: object	12 HOUSEHOLD INCOME IN THE PAST 12 MONTHS (IN 2021 INFLATION-ADJUSTED DOLLARS) Name: Concept, dtype: object
9	B08126_043E	8	42 [Estimate, Total:, Car, truck, or van - carpooled:, Other services (except public administration)] Name: Label, dtype: object	42 MEANS OF TRANSPORTATION TO WORK BY INDUSTRY Name: Concept, dtype: object
10	B02001_004E	7	3 [Estimate, Total:, American Indian and Alaska Native alone] Name: Label, dtype: object	3 RACE Name: Concept, dtype: object

rf_FI(flag="clf")

	Top Feature	Importance	Desc	Concept
0	B08303_003E	0.011266	2 [Estimate, Total:, 5 to 9 minutes] Name: Label, dtype: object	2 TRAVEL TIME TO WORK Name: Concept, dtype: object
1	B25132_005E	0.007752	4 [Estimate, Total, Charged for electricity, 50to99] Name: Label, dtype: object	4 MONTHLY ELECTRICITY COSTS Name: Concept, dtype: object
2	B25132_009E	0.006329	8 [Estimate, Total, Charged for electricity, $250 or more] Name: Label, dtype: object	8 MONTHLY ELECTRICITY COSTS Name: Concept, dtype: object
3	B11001_007E	0.006308	6 [Estimate, Total:, Nonfamily households:] Name: Label, dtype: object	6 HOUSEHOLD TYPE (INCLUDING LIVING ALONE) Name: Concept, dtype: object
4	B11001_008E	0.005864	7 [Estimate, Total:, Nonfamily households:, Householder living alone] Name: Label, dtype: object	7 HOUSEHOLD TYPE (INCLUDING LIVING ALONE) Name: Concept, dtype: object
5	B11001_009E	0.005853	8 [Estimate, Total:, Nonfamily households:, Householder not living alone] Name: Label, dtype: object	8 HOUSEHOLD TYPE (INCLUDING LIVING ALONE) Name: Concept, dtype: object
6	B08124_030E	0.005624	29 [Estimate, Total:, Walked:, Management, business, science, and arts occupations] Name: Label, dtype: object	29 MEANS OF TRANSPORTATION TO WORK BY OCCUPATION Name: Concept, dtype: object
7	B25132_004E	0.005430	3 [Estimate, Total, Charged for electricity, Less than $50] Name: Label, dtype: object	3 MONTHLY ELECTRICITY COSTS Name: Concept, dtype: object
8	B25133_005E	0.005103	4 [Estimate, Total, Charged for gas, 25to49] Name: Label, dtype: object	4 MONTHLY GAS COSTS Name: Concept, dtype: object
9	B08303_005E	0.005099	4 [Estimate, Total:, 15 to 19 minutes] Name: Label, dtype: object	4 TRAVEL TIME TO WORK Name: Concept, dtype: object

Combine most frequent yet significant features from various algo

	Features	Desc	Concept
0	B11001_007E	6 [Estimate, Total:, Nonfamily households:] Name: Label, dtype: object	6 HOUSEHOLD TYPE (INCLUDING LIVING ALONE) Name: Concept, dtype: object
1	B08124_030E	29 [Estimate, Total:, Walked:, Management, business, science, and arts occupations] Name: Label, dtype: object	29 MEANS OF TRANSPORTATION TO WORK BY OCCUPATION Name: Concept, dtype: object
2	B25132_005E	4 [Estimate, Total, Charged for electricity, 50𝑡𝑜50��99] Name: Label, dtype: object	4 MONTHLY ELECTRICITY COSTS Name: Concept, dtype: object
3	B08126_070E	69 [Estimate, Total:, Walked:, Professional, scientific, and management, and administrative and waste management services] Name: Label, dtype: object	69 MEANS OF TRANSPORTATION TO WORK BY INDUSTRY Name: Concept, dtype: object
4	B11001_008E	7 [Estimate, Total:, Nonfamily households:, Householder living alone] Name: Label, dtype: object	7 HOUSEHOLD TYPE (INCLUDING LIVING ALONE) Name: Concept, dtype: object
5	B08303_003E	2 [Estimate, Total:, 5 to 9 minutes] Name: Label, dtype: object	2 TRAVEL TIME TO WORK Name: Concept, dtype: object
6	B08126_074E	73 [Estimate, Total:, Walked:, Public administration] Name: Label, dtype: object	73 MEANS OF TRANSPORTATION TO WORK BY INDUSTRY Name: Concept, dtype: object
7	B11001_009E	8 [Estimate, Total:, Nonfamily households:, Householder not living alone] Name: Label, dtype: object	8 HOUSEHOLD TYPE (INCLUDING LIVING ALONE) Name: Concept, dtype: object
8	B08126_041E	40 [Estimate, Total:, Car, truck, or van - carpooled:, Educational services, and health care and social assistance] Name: Label, dtype: object	40 MEANS OF TRANSPORTATION TO WORK BY INDUSTRY Name: Concept, dtype: object
9	B08126_069E	68 [Estimate, Total:, Walked:, Finance and insurance, and real estate and rental and leasing] Name: Label, dtype: object	68 MEANS OF TRANSPORTATION TO WORK BY INDUSTRY Name: Concept, dtype: object
10	B25132_009E	8 [Estimate, Total, Charged for electricity, $250 or more] Name: Label, dtype: object	8 MONTHLY ELECTRICITY COSTS Name: Concept, dtype: object
11	B08122_020E	19 [Estimate, Total:, Walked:, At or above 150 percent of the poverty level] Name: Label, dtype: object	19 MEANS OF TRANSPORTATION TO WORK BY POVERTY STATUS IN THE PAST 12 MONTHS Name: Concept, dtype: object
12	B05003_014E	13 [Estimate, Total:, Female:, Under 18 years:] Name: Label, dtype: object	13 SEX BY AGE BY NATIVITY AND CITIZENSHIP STATUS Name: Concept, dtype: object
13	B05001_002E	1 [Estimate, Total:, U.S. citizen, born in the United States] Name: Label, dtype: object	1 NATIVITY AND CITIZENSHIP STATUS IN THE UNITED STATES Name: Concept, dtype: object
14	B28011_003E	2 [Estimate, Total:, With an Internet subscription, Dial-up alone] Name: Label, dtype: object	2 INTERNET SUBSCRIPTIONS IN HOUSEHOLD Name: Concept, dtype: object
15	B25133_002E	1 [Estimate, Total, Not charged, not used, or payment included in other fees] Name: Label, dtype: object	1 MONTHLY GAS COSTS Name: Concept, dtype: object
16	B15003_025E	24 [Estimate, Total:, Doctorate degree] Name: Label, dtype: object	24 EDUCATIONAL ATTAINMENT FOR THE POPULATION 25 YEARS AND OVER Name: Concept, dtype: object
17	B08124_029E	28 [Estimate, Total:, Walked:] Name: Label, dtype: object	28 MEANS OF TRANSPORTATION TO WORK BY OCCUPATION Name: Concept, dtype: object
18	B08122_017E	16 [Estimate, Total:, Walked:] Name: Label, dtype: object	16 MEANS OF TRANSPORTATION TO WORK BY POVERTY STATUS IN THE PAST 12 MONTHS Name: Concept, dtype: object

```
tacked_feat_df["To
11001_007E    5
08124_030E    5
25132_005E    5
08126_070E    4
11001_008E    4
08303_003E    4
08126_074E    3
11001_009E    3
08126_041E    2
08126_069E    2
25132_009E    2
08122_020E    2
05003_014E    2
05001_002E    2
28011_003E    2
25133_002E    2
15003_025E    2
08124_029E    2
08122_017E    2
08301_019E    2
08126_061E    2
08006_015E    2
25132_004E    2
08006_032E    2
08126_043E    1
08006_007E    1
17026_010E    1
08126_024E    1
05003_004E    1
```

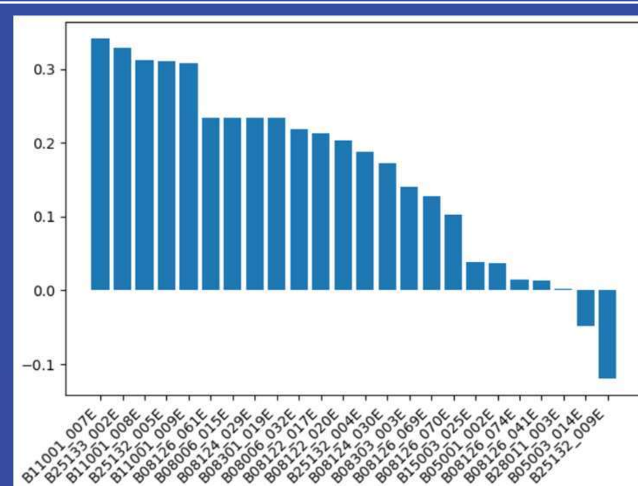
SMU

Feature Extraction using PCA based on Combined Feature Selection

» Keep 1st principal component

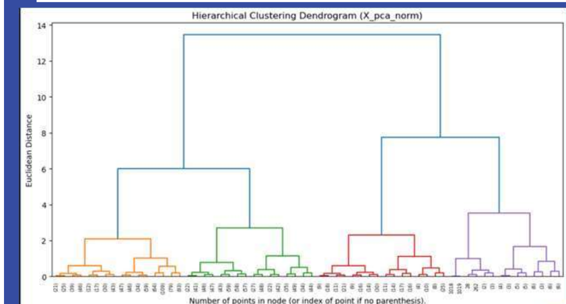
Interesting Findings:

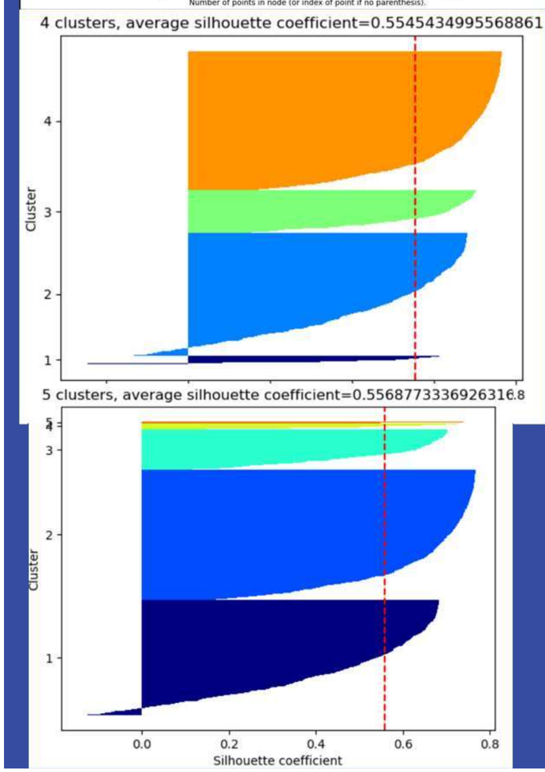
1) Non-family households, walk to work or 5 to 9 mins drive to work, less gas and electricity monthly cost
2) Race does not significantly impact
3) Females under 18 has less accessibility
4) Higher charge for electricity leading to less EVCS

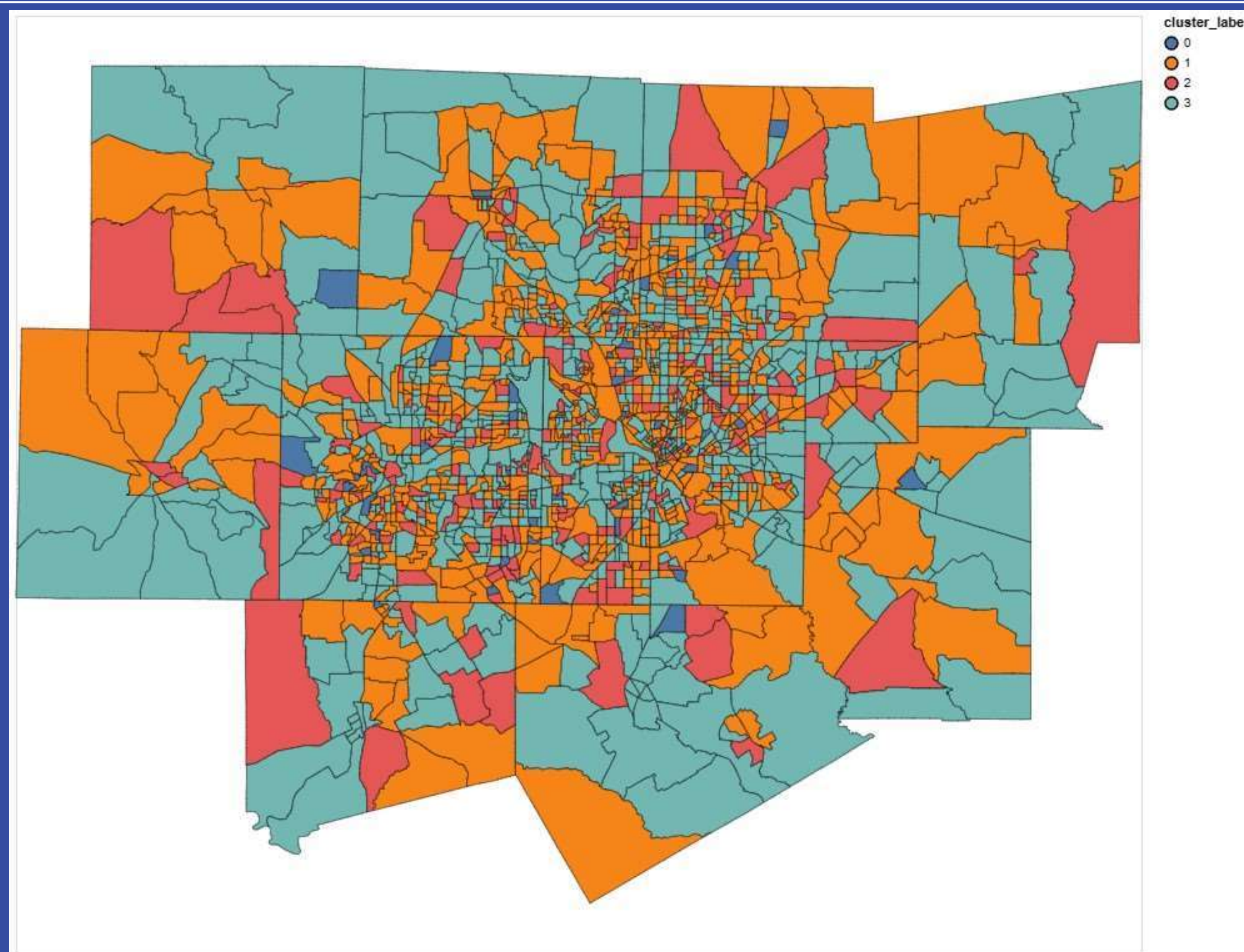


	features	contribution	Desc	Concept
0	B11001_007E	0.341022	6 Estimate!!Total:!!Nonfamily households: Name: Label, dtype: object	6 HOUSEHOLD TYPE (INCLUDING LIVING ALONE) Name: Concept, dtype: object
1	B25133_002E	0.329319	1 Estimate!!Total!!Not charged, not used, or payment included in other fees Name: Label, dtype: object	1 MONTHLY GAS COSTS Name: Concept, dtype: object
2	B11001_008E	0.312600	7 Estimate!!Total:!!Nonfamily households:!!Householder living alone Name: Label, dtype: object	7 HOUSEHOLD TYPE (INCLUDING LIVING ALONE) Name: Concept, dtype: object
3	B25132_005E	0.311248	4 Estimate!!Total!!Charged for electricity!!$50to$99 Name: Label, dtype: object	4 MONTHLY ELECTRICITY COSTS Name: Concept, dtype: object
4	B11001_009E	0.308270	8 Estimate!!Total:!!Nonfamily households:!!Householder not living alone Name: Label, dtype: object	8 HOUSEHOLD TYPE (INCLUDING LIVING ALONE) Name: Concept, dtype: object
5	B08126_061E	0.234539	60 Estimate!!Total:!!Walked: Name: Label, dtype: object	60 MEANS OF TRANSPORTATION TO WORK BY INDUSTRY Name: Concept, dtype: object
6	B08006_015E	0.234539	14 Estimate!!Total:!!Walked Name: Label, dtype: object	14 SEX OF WORKERS BY MEANS OF TRANSPORTATION TO WORK Name: Concept, dtype: object
7	B08124_029E	0.234539	28 Estimate!!Total:!!Walked: Name: Label, dtype: object	28 MEANS OF TRANSPORTATION TO WORK BY OCCUPATION Name: Concept, dtype: object
8	B08301_019E	0.234539	18 Estimate!!Total:!!Walked Name: Label, dtype: object	18 MEANS OF TRANSPORTATION TO WORK Name: Concept, dtype: object
9	B08006_032E	0.217987	31 Estimate!!Total:!!Male:!!Walked Name: Label, dtype: object	31 SEX OF WORKERS BY MEANS OF TRANSPORTATION TO WORK Name: Concept, dtype: object
10	B08122_017E	0.213197	16 Estimate!!Total:!!Walked: Name: Label, dtype: object	16 MEANS OF TRANSPORTATION TO WORK BY POVERTY STATUS IN THE PAST 12 MONTHS Name: Concept, dtype: object
11	B08122_020E	0.203795	19 Estimate!!Total:!!Walked:!!At or above 150 percent of the poverty level Name: Label, dtype: object	19 MEANS OF TRANSPORTATION TO WORK BY POVERTY STATUS IN THE PAST 12 MONTHS Name: Concept, dtype: object
12	B25132_004E	0.188267	3 Estimate!!Total!!Charged for electricity!!Less than $50 Name: Label, dtype: object	3 MONTHLY ELECTRICITY COSTS Name: Concept, dtype: object
13	B08124_030E	0.172386	29 Estimate!!Total:!!Walked:!!Management, business, science, and arts occupations Name: Label, dtype: object	29 MEANS OF TRANSPORTATION TO WORK B OCCUPATION Name: Concept, dtype: obje
14	B08303_003E	0.139835	2 Estimate!!Total:!!5 to 9 minutes Name: Label, dtype: object	2 TRAVEL TIME TO WORK Name: Concept, dtype: obje
15	B08126_069E	0.127180	68 Estimate!!Total:!!Walked:!!Finance and insurance, and real estate and rental and leasing Name: Label, dtype: object	68 MEANS OF TRANSPORTATION TO WORK B INDUSTRY Name: Concept, dtype: obje
16	B08126_070E	0.102291	69 Estimate!!Total:!!Walked:!!Professional, scientific, and management, and administrative and waste management services Name: Label, dtype: object	69 MEANS OF TRANSPORTATION TO WORK B INDUSTRY Name: Concept, dtype: obje
17	B15003_025E	0.037847	24 Estimate!!Total:!!Doctorate degree Name: Label, dtype: object	24 EDUCATIONAL ATTAINMENT FOR THE POPULATIO 25 YEARS AND OVER Name: Concept, dtype: obje
18	B05001_002E	0.036872	1 Estimate!!Total:!!U.S. citizen, born in the United States Name: Label, dtype: object	1 NATIVITY AND CITIZENSHIP STATUS IN THE UNITE STATES Name: Concept, dtype: obje
19	B08126_074E	0.014360	73 Estimate!!Total:!!Walked:!!Public administration Name: Label, dtype: object	73 MEANS OF TRANSPORTATION TO WORK B INDUSTRY Name: Concept, dtype: obje
20	B08126_041E	0.013217	40 Estimate!!Total:!!Car, truck, or van - carpooled:!!Educational services, and health care and social assistance Name: Label, dtype: object	40 MEANS OF TRANSPORTATION TO WORK B INDUSTRY Name: Concept, dtype: obje
21	B28011_003E	0.001944	2 Estimate!!Total:!!With an Internet subscription!!Dial-up alone Name: Label, dtype: object	2 INTERNET SUBSCRIPTIONS IN HOUSEHOLD Name Concept, dtype: obje
22	B05003_014E	-0.048422	13 Estimate!!Total:!!Female:!!Under 18 years: Name: Label, dtype: object	13 SEX BY AGE BY NATIVITY AND CITIZENSHIP STATU Name: Concept, dtype: obje
23	B25132_009E	-0.119232	8 Estimate!!Total!!Charged for electricity!!$250 or more Name: Label, dtype: object	8 MONTHLY ELECTRICITY COSTS Name: Concept, dtype obje

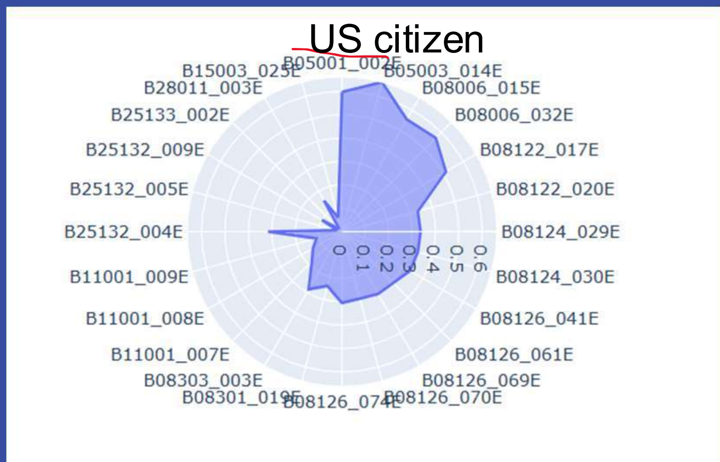
Clustering

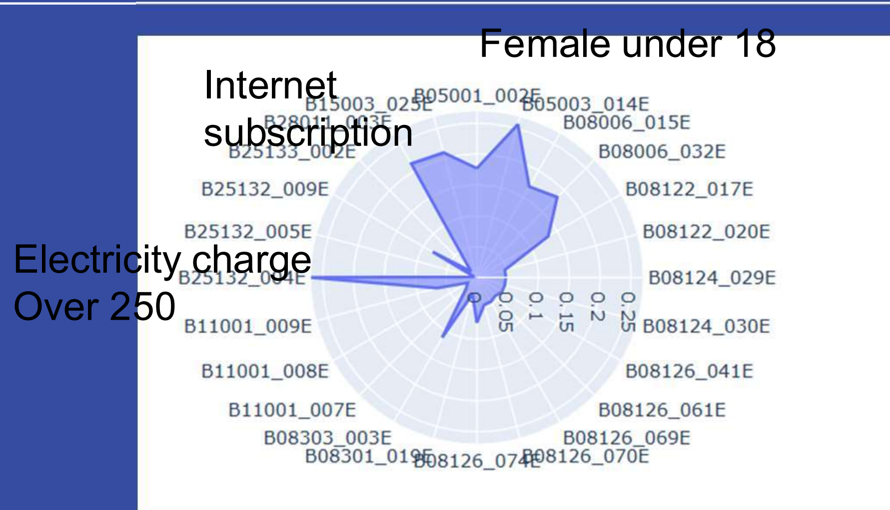

Hierarchical Clustering Dendrogram (X_pca_norm)
Euclidean Distance
Number of points in node (or index of point if no parenthesis).

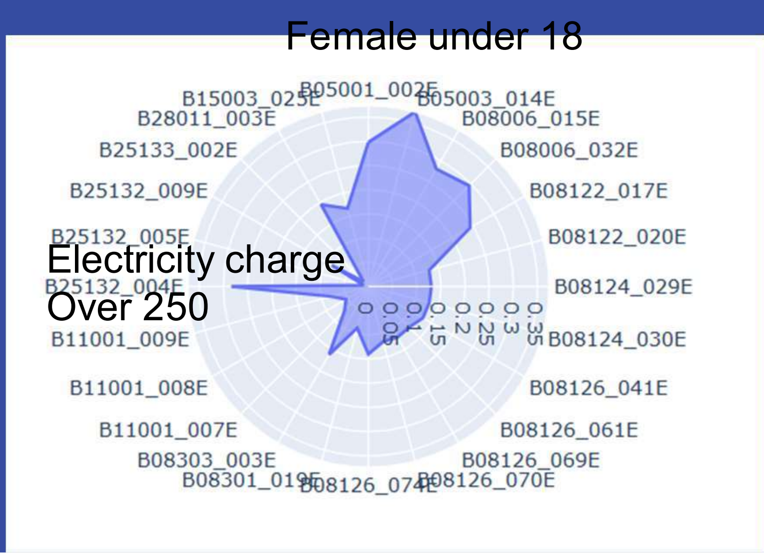

4 clusters, average silhouette coefficient=0.5545434995568861
Cluster
5 clusters, average silhouette coefficient=0.5568773336926316
Cluster
Silhouette coefficient

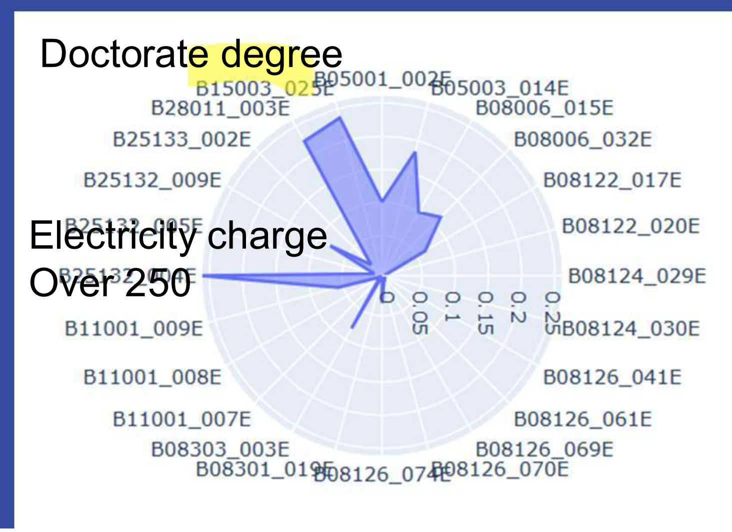

cluster_label
0
1
2
3

Characteristics of Typologies





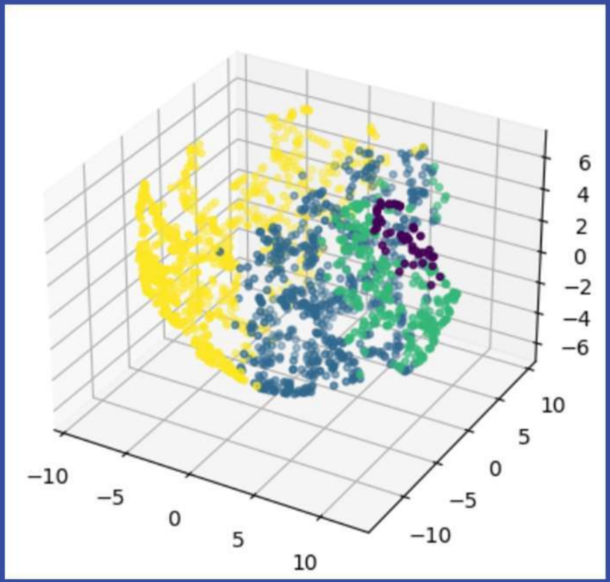


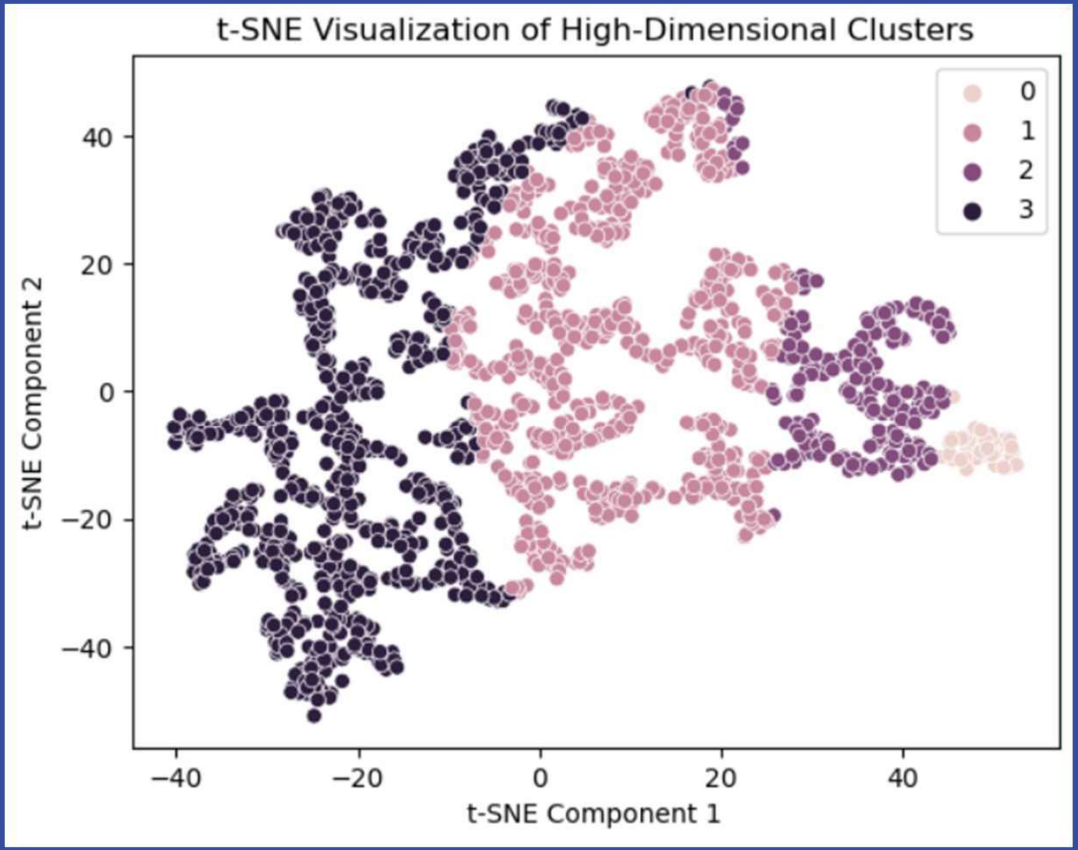


0	B11001_007E	0.341022	6 Es
1	B25133_002E	0.329319	1 Estimate!
2	B11001_008E	0.312600	7 Estimate!!Total:
3	B25132_005E	0.311248	4 Estimate!!
4	B11001_009E	0.308270	8 Estimate!!Tot
5	B08126_061E	0.234539	
6	B08006_015E	0.234539	
7	B08124_029E	0.234539	
8	B08301_019E	0.234539	
9	B08006_032E	0.217987	
10	B08122_017E	0.213197	
11	B08122_020E	0.203795	19 Estimate!!Tot
12	B25132_004E	0.188267	3 Estimate!
13	B08124_030E	0.172386	
14	B08303_003E	0.139835	
15	B08126_069E	0.127180	68 Estim
16	B08126_070E	0.102291	69 Es admi
17	B15003_025E	0.037847	
18	B05001_002E	0.036872	1 Estin
19	B08126_074E	0.014360	73 Estin
20	B08126_041E	0.013217	40 Est
21	B28011_003E	0.001944	2 Estin
22	B05003_014E	-0.048422	1
23	B25132_009E	-0.119232	8 Esti

Visualization of High-Dimensional Clusters

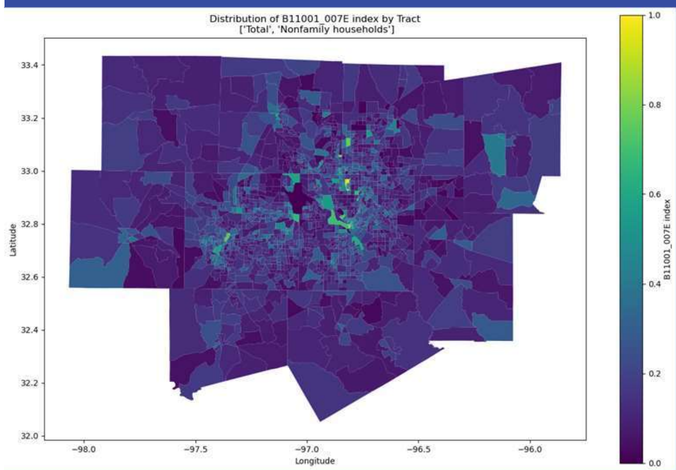
t-SNE (t-Distributed Stochastic Neighbor Embedding) is a machine learning algorithm used for data visualization and dimensionality reduction. It is particularly useful for visualizing high-dimensional data by transforming it into a low-dimensional space, typically two or three dimensions, while preserving the structure of the original data as much as possible.

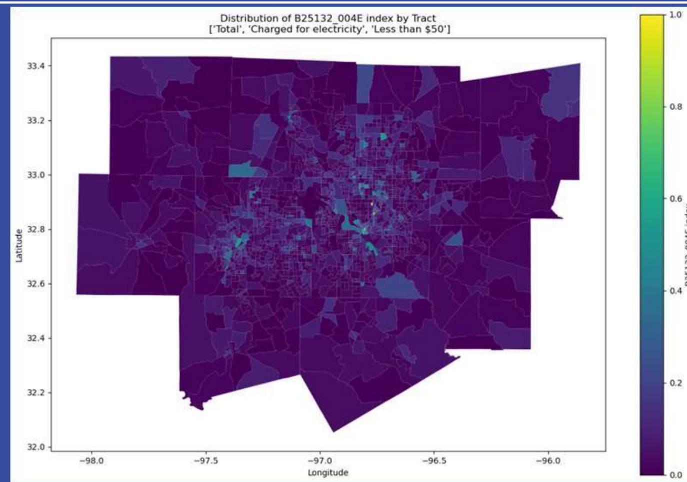


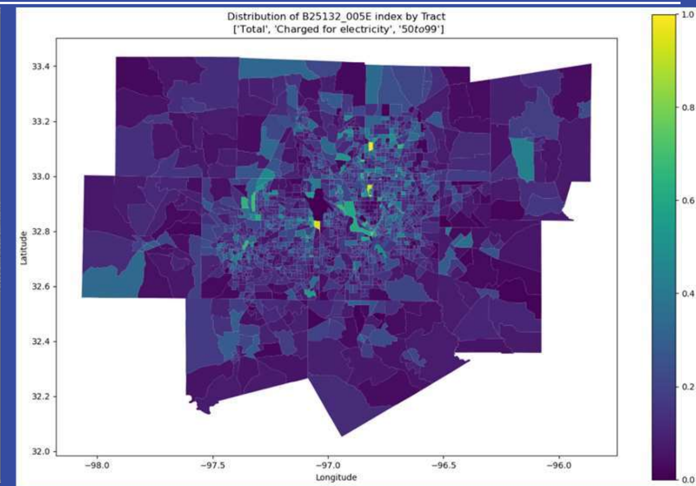


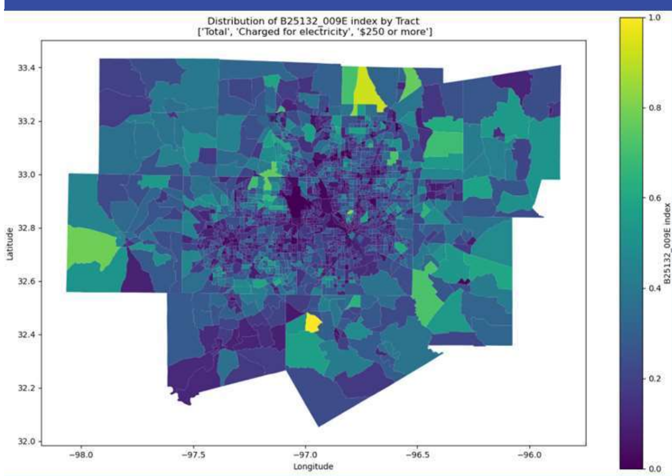
SMU

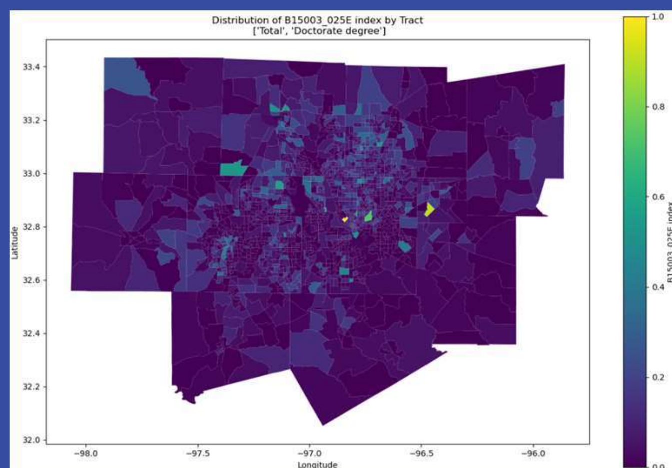
Spatial distribution of socioeconomic composition 1

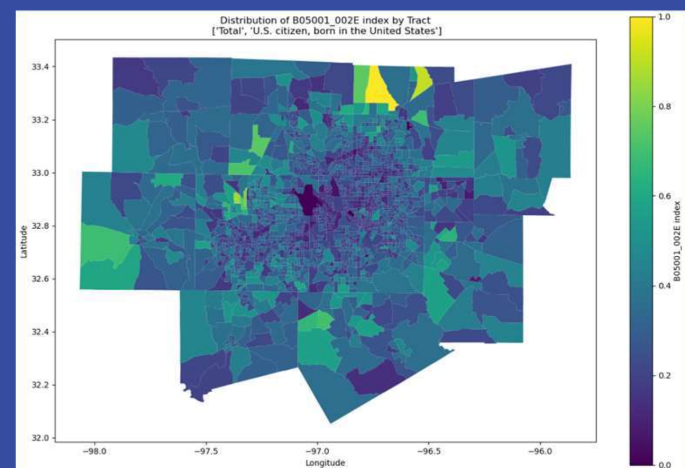


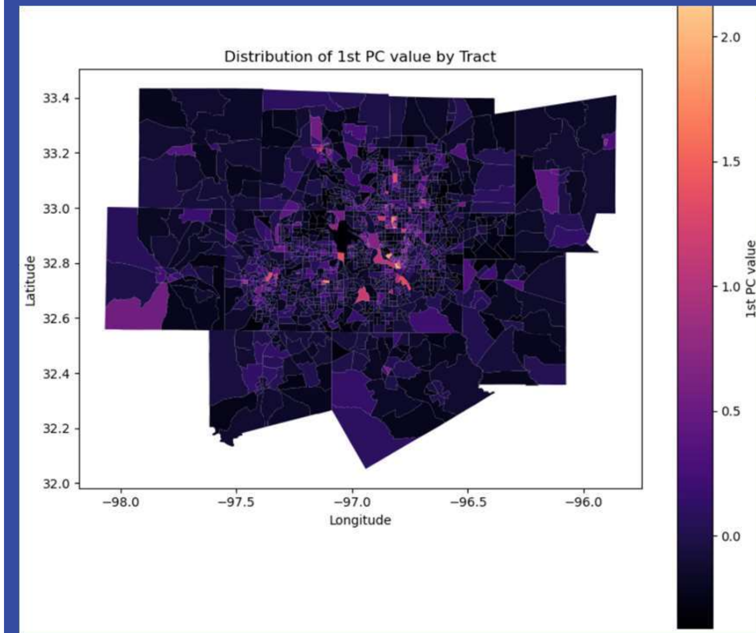


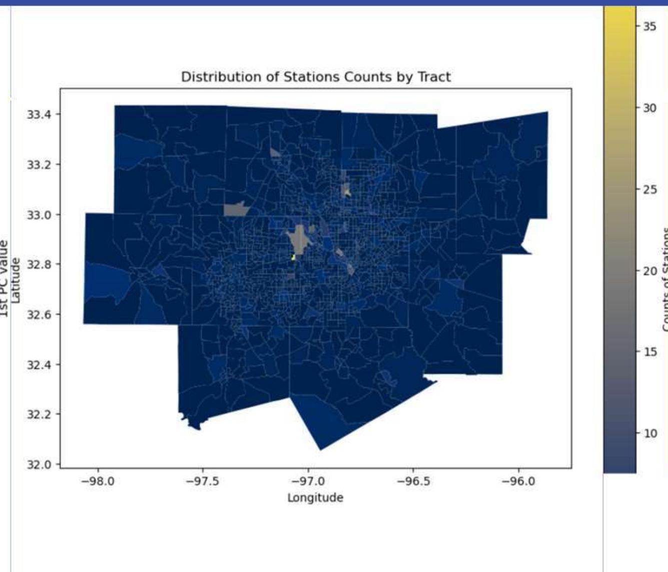


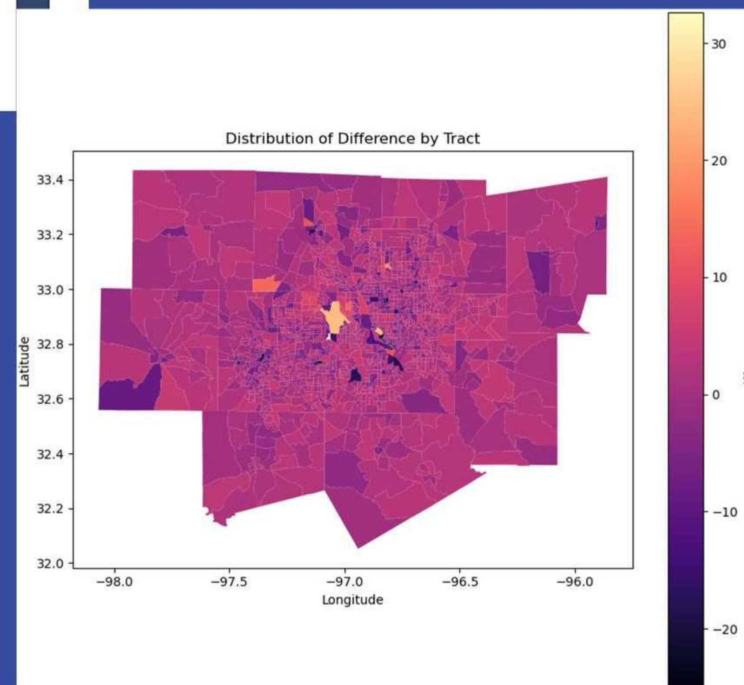












Quantifying Equity Level:

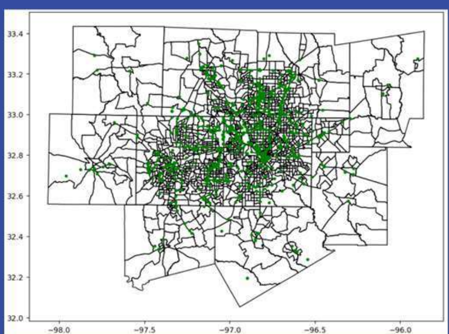
EVCS Equity Index = EVCS_index(counts) – alpha*PCA_PC1(var_1, var_2,….var_k)

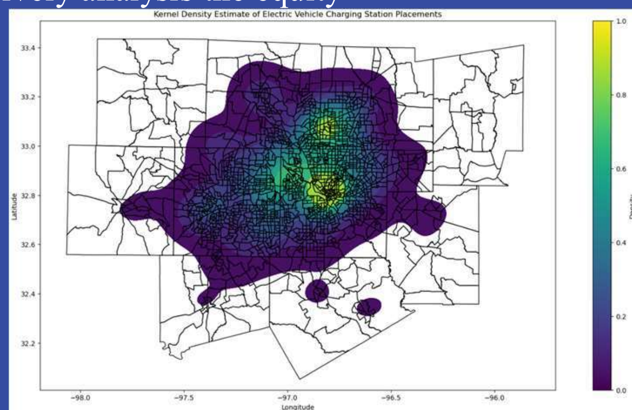
Note:
EVCS_index = func(counts), active term: policymakers to take actions
PCA_PC1 = func(normalized features var_1,var_2, … var_k), passive term: changes as actions taken.
Alpha: an adjustable coefficient based on equity issues cared by local authorities
Normalized features are the most frequent and highest importance features selected from previous 5 algorithms

Quantifying spatial equity level:

How to Quantify:
- Comparison between typologies pattern and EVCS placements;
- Develop **equity metrics**;
- Visualize the variations in equity level
- Develop RL-ABM to quantitatively analysis the equity

- Quantifying spatial equity levels for EVCS placements
- How EVCS placement densities varied spatially with existing socioeconomic inequalities.

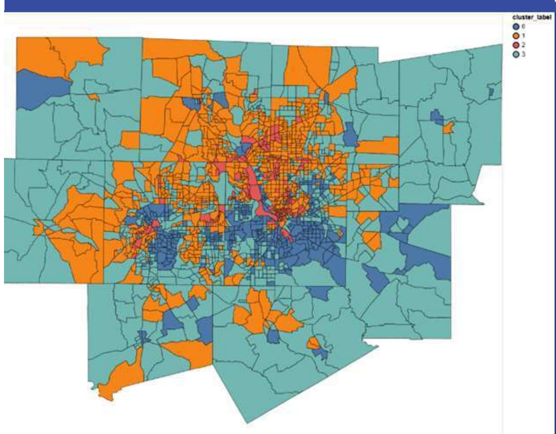


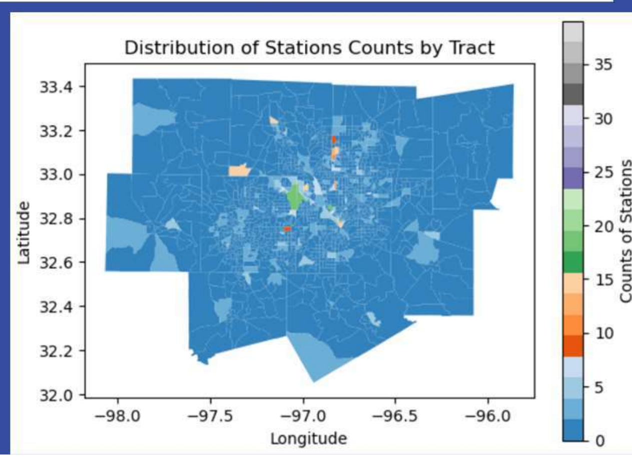


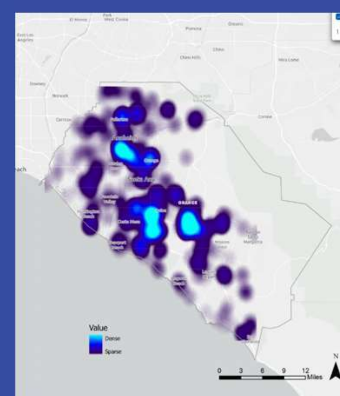
Examining the equity
1. To propose a typology. Its pattern matches current EVCS quantities.
2. Understand the difference btw. Justice 40 and our typology
3. Quantifying spatial equity levels for EVCS placements

Table 4a
List of variables used to create equity indicators for Orange County.

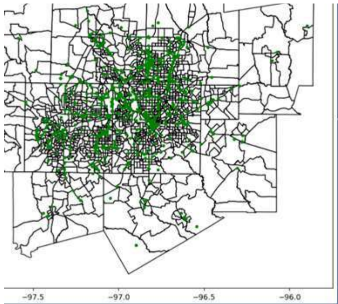
Variable	Description	Relevance
Vehicles Available	Average number of vehicles available in each household	Mobility indicator (Hanke et al., 2014, Luo et al., 2015, Javid & Nejat, 2017)
Population Density	Spatial distribution of population, calculated by total population/Land area	Demographic indicator (Plötz et al., 2014, Chakraborty, 2019)
Poverty	Percentage of population that lives under the poverty line	Socioeconomic indicator (Li, 2017, Hsu & Fingerman, 2021, Javid & Nejat, 2017)
Education	Percentage of population that holds a bachelor's degree or higher	Socioeconomic indicator (Hanke et al., 2014, Nayum et al., 2016, Javid & Nejat, 2017)
Income	Average median household income	Socioeconomic indicator (Hanke et al., 2014, Nayum et al., 2016)
Household Size	Average number of persons living in a household	Demographic indicator (Plötz et al., 2014, Chakraborty, 2019)
Housing Affordability	Percentage of population that spend more than 30% of monthly income on housing	Socioeconomic indicator (Westin et al., 2018, Hanke et al., 2014, Nayum et al., 2016)
Average Commute Time	Average one-way travel time to work	Mobility indicator (Luo et al., 2015, Javid & Nejat, 2017)
Age Between 25 to 45	Percentage of population that are between age 25 to 45	Demographic indicator (Westin, 2018, Chakraborty, 2019)
Employment Rate	Percentage of employment for population over age of 16	Socioeconomic indicator (Li, 2017, Hsu & Fingerman, 2021, Javid & Nejat, 2017)
Distance to EVCS	Distance between centroid of Census block group and the closest EVCS	Mobility indicator (Luo et al., 2015, Javid & Nejat, 2017)
PM 2.5 Level	Annual mean concentration particulate matter level	Environmental indicator ((Nayum, Klöckner and Mehmetoglu, 2016), Nordlund et al., 2016; Zeise & Blumenfeld, 2022)
Traffic Impact	Sum of traffic volumes (vehicle-kilometers per hour) divided by total road length (kilometers) within 150 meters of the census tract	Mobility and environmental indicator (Luo et al., 2015, Nordlund et al., 2016, Javid & Nejat, 2017)

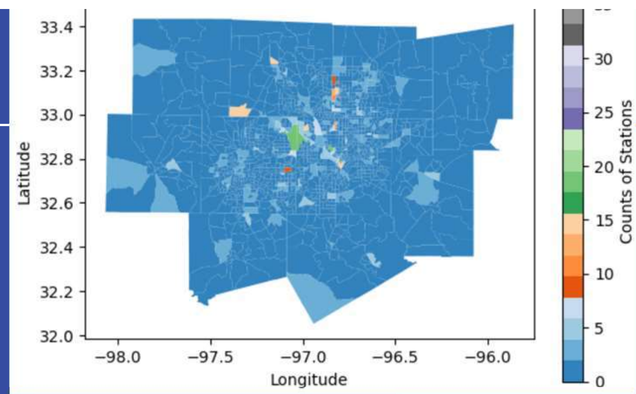


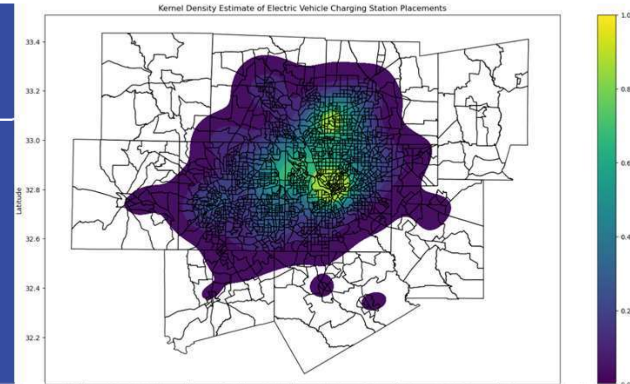


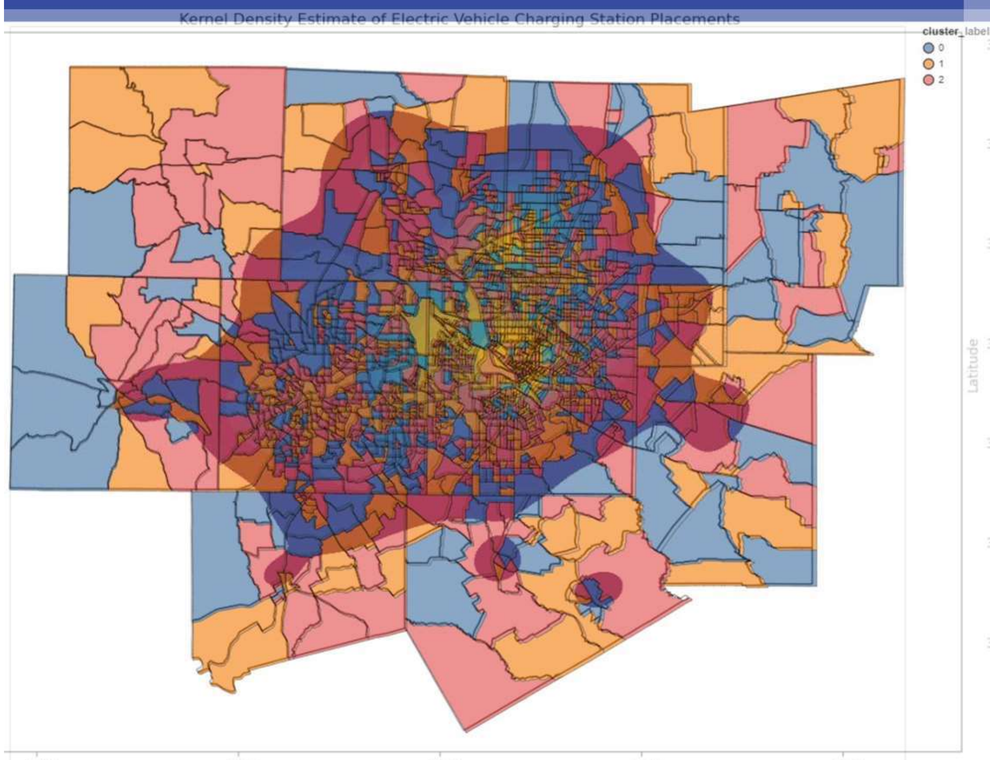


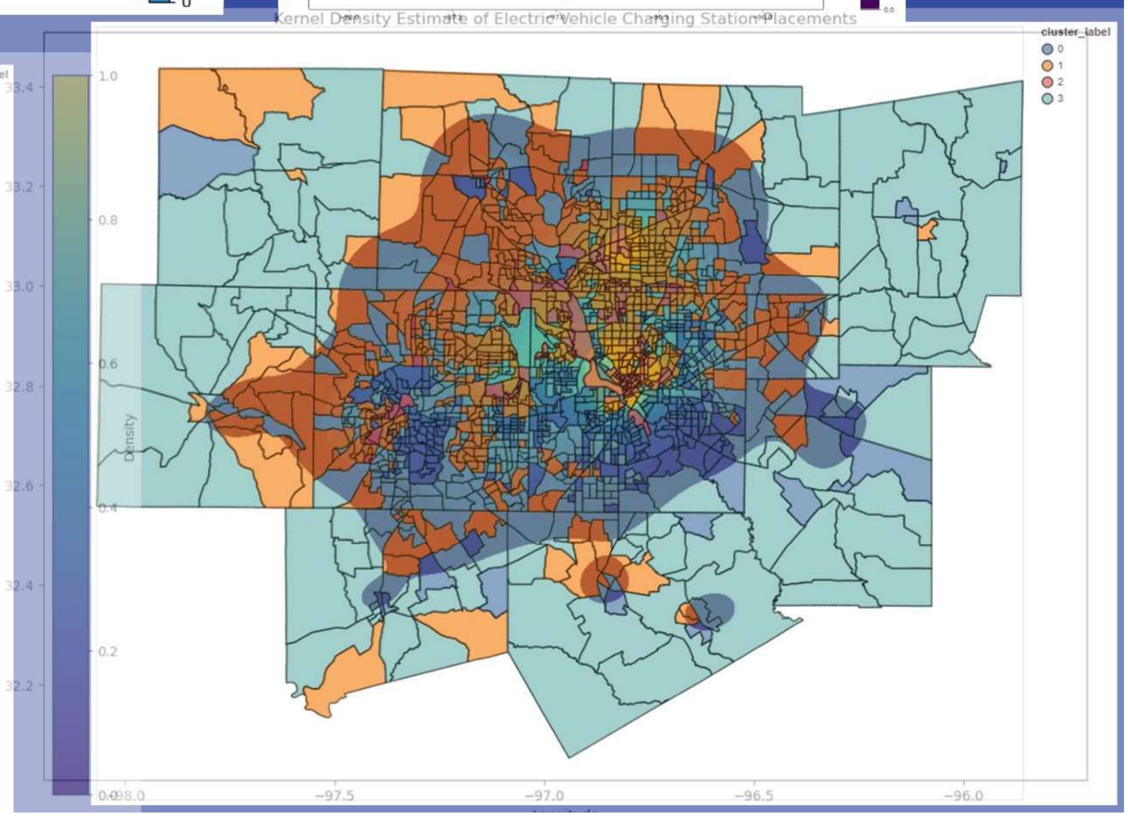
SMU

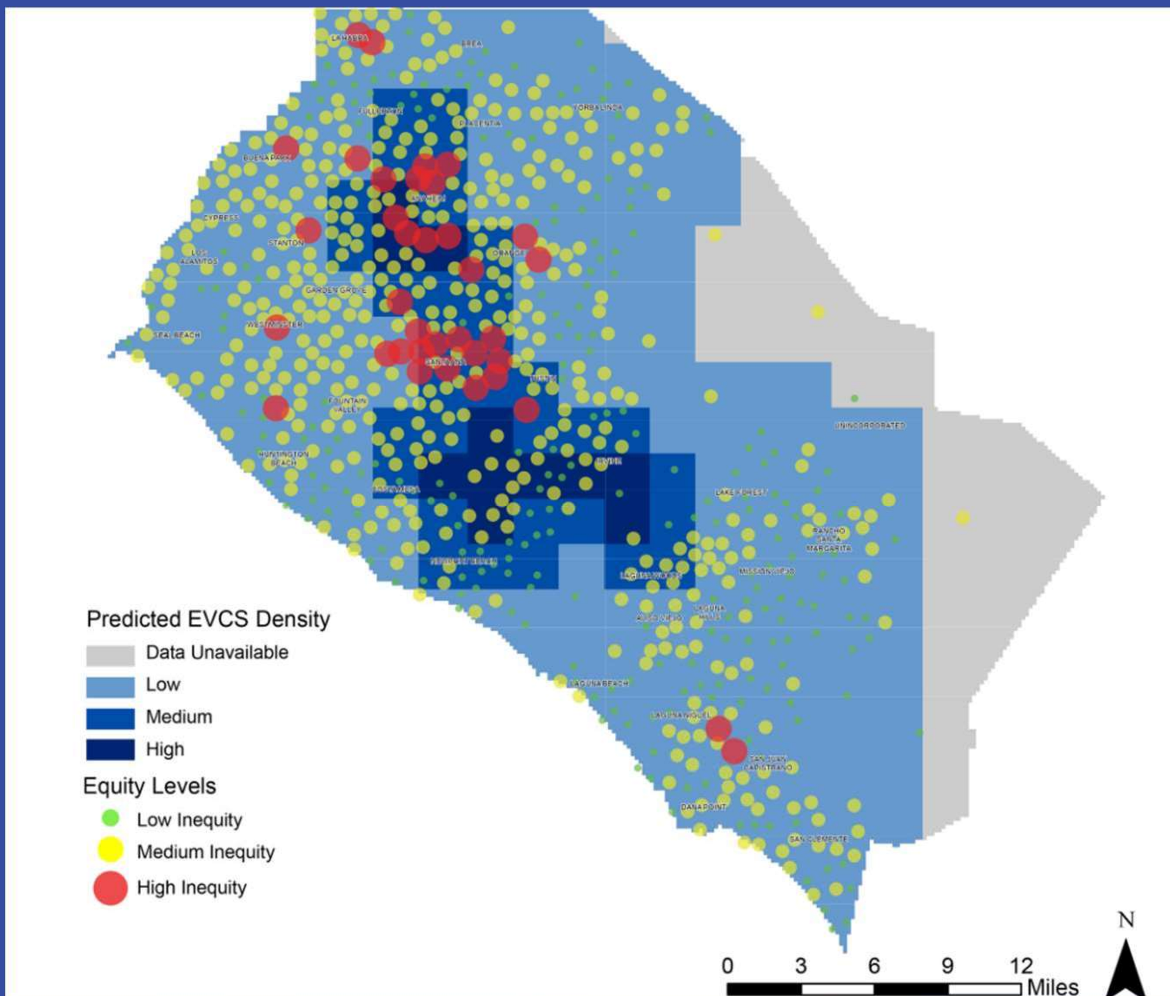

−97.5
−97.0
−96.5
−96.0


Latitude
Longitude
Counts of Stations
33.4
33.2
33.0
32.8
32.6
32.4
32.2
32.0
−98.0
−97.5
−97.0
−96.5
−96.0
30
25
20
15
10
5
0


Kernel Density Estimate of Electric Vehicle Charging Station Placements
Latitude
Density


Kernel Density Estimate of Electric Vehicle Charging Station Placements
cluster_label
0
1
2
−98.0
−97.5
−97.0
−96.5
−96.0


Kernel Density Estimate of Electric Vehicle Charging Station Placements
cluster_label
0
1
2
3
Latitude
Density
1.0
0.8
0.6
0.4
0.2
33.4
33.2
33.0
32.8
32.6
32.4
32.2
−98.0
−97.5
−97.0
−96.5
−96.0

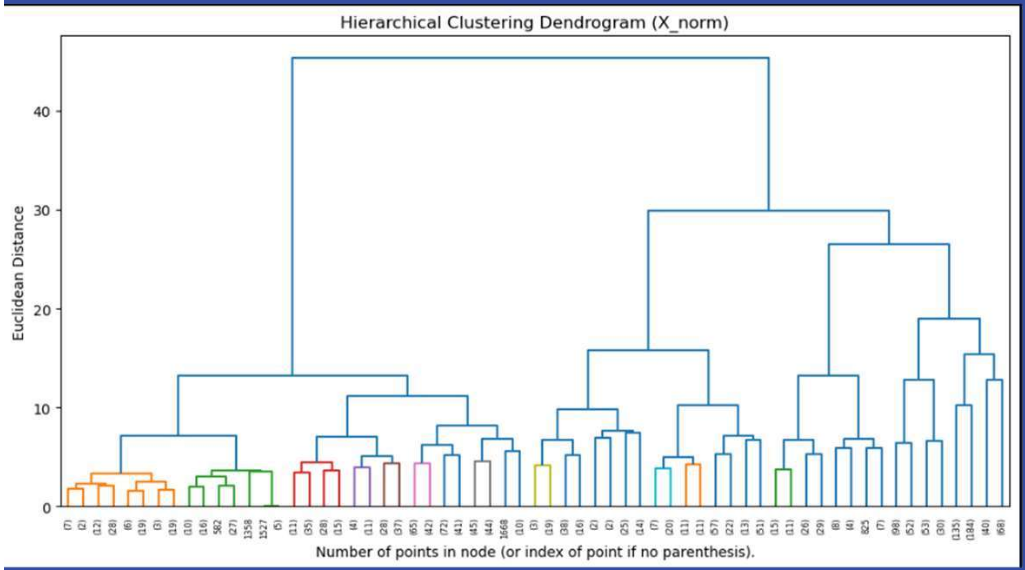

Predicted EVCS Density
Data Unavailable
Low
Medium
High
Equity Levels
Low Inequity
Medium Inequity
High Inequity
0 3 6 9 12 Miles
N

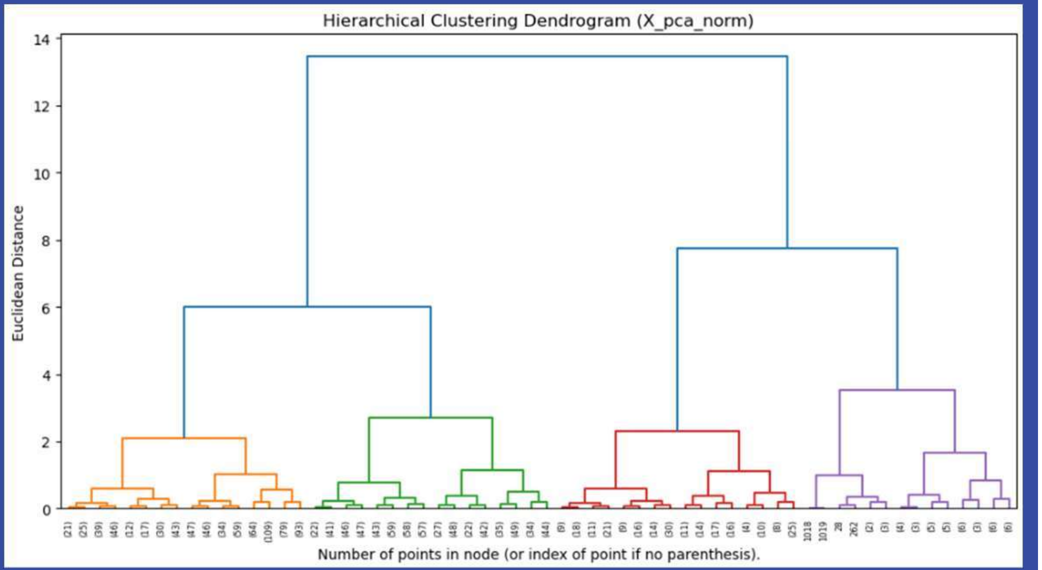
Comparison: dendrogram

» **NREL paper: Inequality and the future of electric mobility in 36 U.S. Cities: An innovative methodology and comparative assessment**

» **Proposed approach**

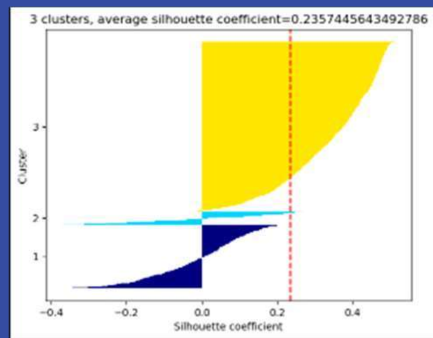
- feature extraction
- dimension reduction (PCA)

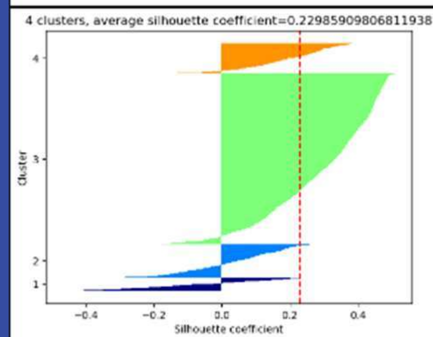


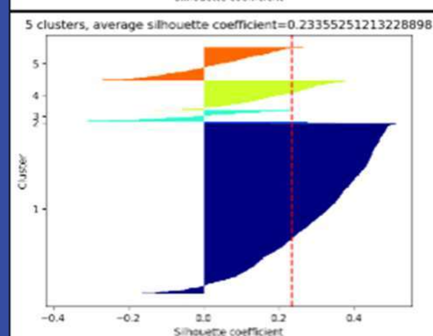


Comparison: Silhouette analysis

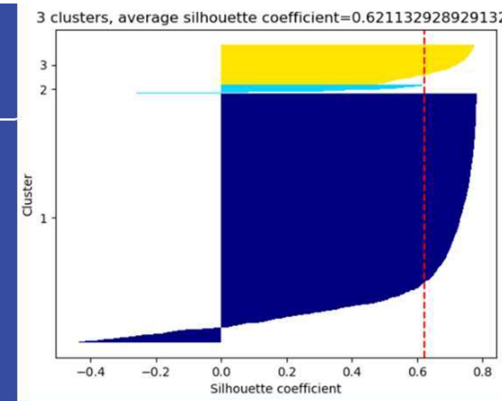
» **Inequality and the future of electric mobility in 36 U.S. Cities: An innovative methodology and comparative assessment**

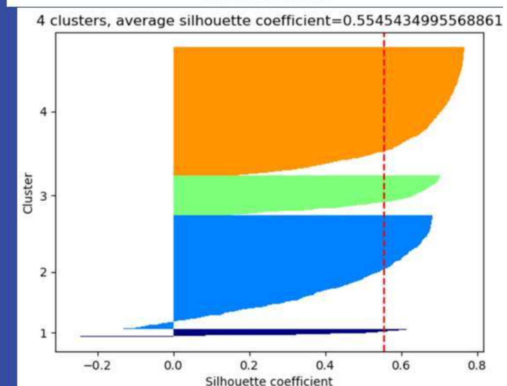


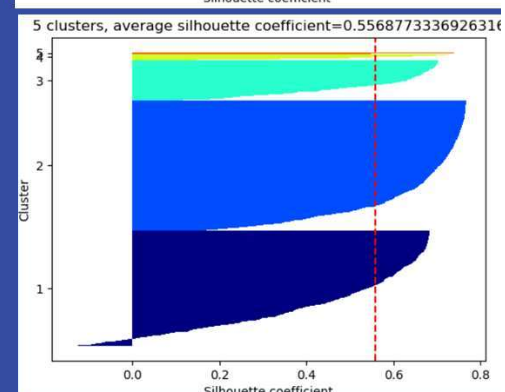




» **Proposed approach**

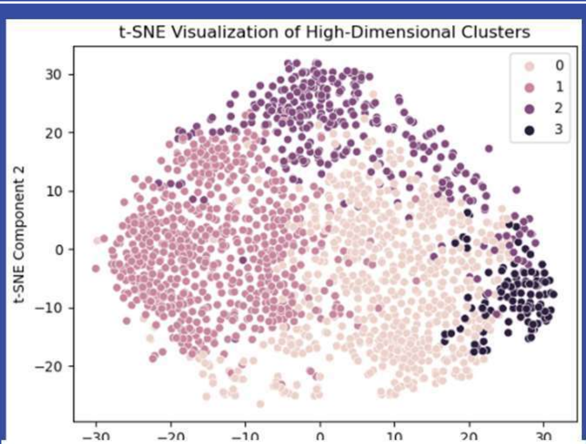


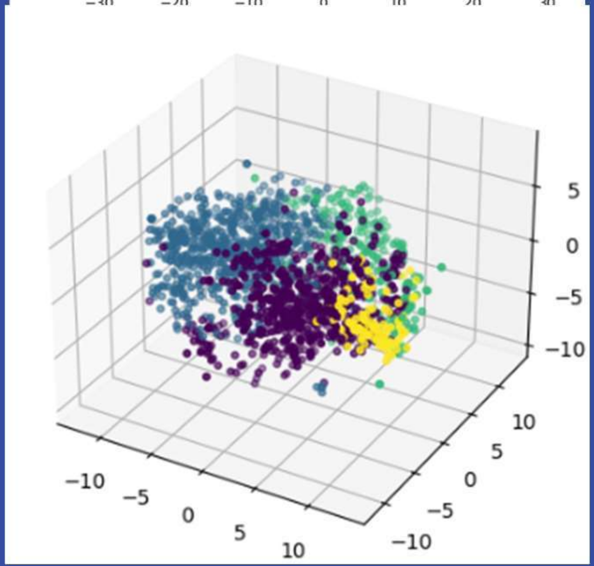




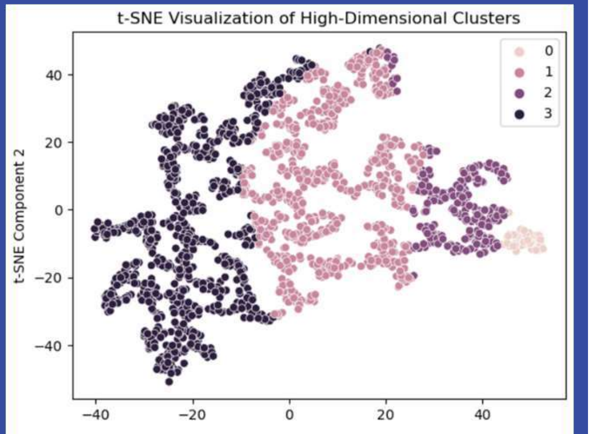
Comparison: cluster visualization

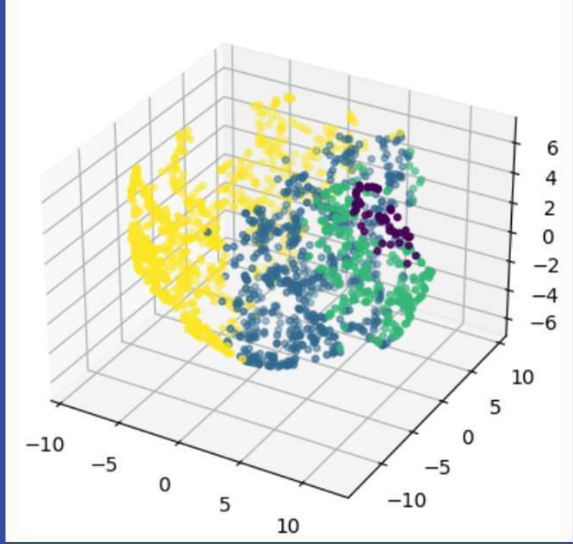
» **Inequality and the future of electric mobility in 36 U.S. Cities: An innovative methodology and comparative assessment**





» **Proposed approach**





SMU

Agent-Based Modeling vs Reinforcement Learning

Table 1. Comparison of Agent Based Modeling and Reinforcement Learning

Theory-driven Agent-Based Modeling	vs	Data-driven Reinforcement Learning
Agents' behavior generate a system behaviour		Focus on one agents' decision
Each step and parameter are interpretable, based on theories and laws		Uninterpretable, each parameter has no physical significance
Use factors that already known to influence social/physical/natural phenomenon		Capable of finding latent factors
Iteration always required (computationally expensive)		Iteration only required in the process of training (Faster in prediction)
Less data needed		Need significant large amount of data
Sensitive to missing data or attributes		Not sensitive to missing data or attributes
Model based		Model free
Deterministic and robust policies		Stochastic policy (Exploration and exploitation tradeoff)

Hybrid modeling for Multi-agent system

Paradigm of agent based modeling for

Reinforcement Learning

SMU

Single-RL-ABM framework