Cambridge Medium (2024), **00**, 1–12

CAMBRIDGE
UNIVERSITY PRESS

RESEARCH

A breakdown of American energy consumers: From the *gluttonous* to the *overladen*

Kyuin Lee

School of Public Policy, Georgetown University, Washington, DC, U.S.A

Abstract

The purpose of this study was to create an overarching typology of American households based on their energy usage patterns, given the lack thereof and its implications for accelerating America's low-carbon transition. Results highlight that the U.S. energy landscape is incredibly multifaceted. According to cluster analysis, American households consist of four types: 1) the *polluting elite*, 2) the *thrifty but energy-inefficient*, 3) the *overburdened*, and 4) the *modest consumer*. Findings suggest that the socioeconomically disadvantaged are either overwhelmed by energy-related expenditures or unknowingly affected by the hidden costs of energy-inefficiency, despite efforts to tighten their belts. In contrast, the *polluting elite*s indulge in gluttony, living in outsized homes given the number of occupants and exerting an outsized impact on the environment. By unveiling the differences across U.S. households, this study points to the sources of wasteful energy consumption as well as the importance of prioritizing the *overburdened* along with the *thrifty but energy-inefficient*, for an equitable transition to clean energy and net-zero.

Keywords: energy sector, residential buildings, energy use intensity, cluster analysis, k-prototype clustering

1. Background

1.1 Greening the U.S. residential energy sector

Whether it is measured in *per capita* or *aggregate* terms, the U.S. has historically been one of the world's largest greenhouse gas emitters (The Global Carbon Atlas 2023). Amid a heightened climate crisis, however, the Biden Administration has put forward a host of ambitious climate goals, vowing to slash emissions across all sectors of the economy and invest in green energy/green infrastructure projects at an unprecedented scale (The White House 2024b).

Particularly critical to the U.S. climate agenda is decarbonization of *residential buildings*, as space heating, cooling, and daily operations of the home contribute to as much as one-fifth of America's energy-related greenhouse gas emissions (U.S. Energy Information Administration 2024b). To attain a green, residential energy sector, the Biden Administration seeks a three-pronged approach: tackling the climate crisis *at its source* by curbing on-site household energy consumption, improving energy efficiency of homes, and reducing household energy bills as a result (The White House 2021b). And cross-cutting these three pillars is the notion of *environmental justice*, as transition to net-zero requires sweeping and fundamental changes to how we power our homes, which results in certain segments of the society to bear a bigger brunt of such potentially costly but inevitable change for a sustainable future (The White House 2021a).

As part of its net-zero policy, the state and local government offers tax credits and rebates for energy efficiency upgrades (The White House 2024a) — such as installation of double-glazed windows, the purchase of all-electric heat pumps [1], and adoption of solar panels. However, in order to ensure the *relevance* of such programs in tandem with *equitable* access to clean, sustainable technology, a good understanding of households' energy consumption behavior and their energy demands is a prerequisite.

1.2 Problem definition

As important as it is to have a thorough understanding of the status quo, the current U.S. energy consumption landscape is either: discussed simply in terms of dichotomy and/or aggregated — that is, simplified down — to the state-level. For instance, households are differentiated along whether they can afford their energy bills or not — the latter which is referred to as *energy insecurity* — and whether these energy-insecure households bear higher utility costs than the other. Ironically, research corroborates that those in straitened conditions (i.e. the energy-insecure) are indeed billed more than their affluent counterparts (U.S. Energy Information Administration 2023a). Alternatively, household energy usage is often conceptualized at the state-level, such as that households in colder states consume more energy than those in warmer states (U.S. Energy Information Administration 2023b). But America is defined by large wealth gaps (Kochhar and Moslimani 2023) and heterogeneity in race and ethnicity (Frey 2020). Moreover, the typical American home (e.g. single-family homes, apartments) varies by location and its urban/rural characteristics (Badger and Ingraham 2015). A unique combination of each of these factors would impact energy consumption in different ways, which a single determinant would fail to account for. Since state-level observations treat as if households in a given state are homogeneous entities, they fail to illustrate differences across households.

In light of these shortcomings, this study sought to provide a bird's-eye perspective yet a fine-grained understanding of Americans' energy consumption patterns, by clustering households along multiple dimensions and without relying on generalizations along state boundaries.

2. Data

2.1 Data collection

The choice of dataset was guided by three criteria: authoritativeness, geographic coverage, and suitability for assessing households' *typical* energy consumption throughout the year. To obtain data specific to American households, I analyzed the Residential Energy Consumption Survey (RECS) for 2020 (U.S. Energy Information Administration 2024a), administered by the U.S. Energy Information Administration — the statistical arm of the U.S. Department of Energy. Spanning 50 U.S. states and the District of Columbia, RECS is the only source of information on the:

1. Characteristics that contribute to Americans' household energy consumption;
2. Total energy usage (*actual*) in a given year;
3. A breakdown of household energy usage, by end-use (*modeled*) [2]; as well as
4. Annual expenditures on energy, in dollars.

Moreover, RECS was particularly well-suited for analyzing America's *typical* household energy usage as it covers only *primary, occupied* housing units — That is, a *usual* place of residence, which excludes vacant, seasonal, and vacation homes.

Since energy consumption is determined by a combination of climatic and physical characteristics of a dwelling along with demographic and behavioral characteristics of its occupants, main variables of interest included:

1. **Climatic characteristics**
    - Heating degree days
    - Cooling degree days [3]

2. **Housing characteristics**
    - Total energy consuming area of the home (ft^2)
    - Draftiness
3. **Household characteristics**
    - Income
    - Energy insecurity [4]
    - Education
    - Race/Ethnicity
    - Household size
4. **Total energy usage, 2020** (*kBtu*)

2.2 Data refinement

The data pipeline began with feature engineering (**Figure** 1). First, I constructed a combination of *adjusted* and *unadjusted* measures of annual household energy consumption:

1. The raw total (*kBtu*);
2. Energy usage adjusted for the *area of the home* ($kBtu/ft^2$); and
3. Energy usage adjusted for the *size of the household* (*kBtu* per member of the household).

Energy consumption adjusted for the *area of the housing unit* is otherwise known as Energy Use Intensity (EUI). It should be noted that the EUI captures a different aspect of energy consumption than the other two measures, as it indicates *energy efficiency* of a home's operations, whereby a lower EUI corresponds to higher energy efficiency.

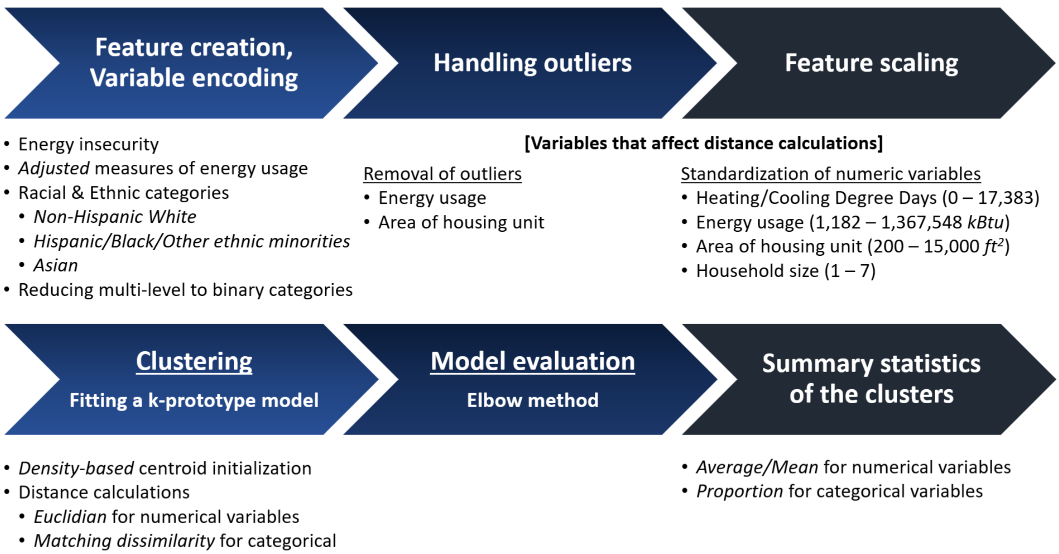

Figure 1. The data pipeline

Since the *area of housing unit* and *household size* were to be incorporated as standalone features of the model, only the *raw aggregate* measure of *energy consumption* was used to cluster the surveyed households. *Adjusted* measures of annual energy consumption were examined only during the subsequent analyses, for a granular understanding of the resultant clusters.

Next, I created a binary indicator for *energy insecurity*, defined as having experienced any of the following:

1. Having to forgo basic necessities in order to afford one's energy bill;
2. Receipt of disconnection notice due to overdue payment; and
3. Keeping the home at unhealthy or unsafe temperatures.

For simplicity, I collapsed the multi-level categorical variables by reclassifying *annual income* as *low*, *medium*, and *high* and *educational attainment* as *no bachelor's degree* and *at least a bachelor's degree*. To distinguish between *Hispanic* Whites from *non-Hispanic* Whites, *race* and *ethnicity* were combined into a single variable [5]. However, due to a predominantly non-Hispanic White sample, it was difficult to make sense of the clusters as (non-Hispanic) Whites accounted for the bulk of every cluster. To alleviate this problem, *all* types of *racial* and *ethnic minority* groups but for Asians were merged into a single category. Note that *Asians* were kept separate from *Hispanic/Black/Other racial minorities* because the former fared far better across multiple socioeconomic indicators and tended to demonstrate different energy consumption patterns.

Subsequently, outliers were removed using the interquartile method, as centroid-based clustering algorithms tend to be sensitive to outliers. However, outliers for *heating* and *cooling degree days* were kept by design, because their removal would result in the exclusion of households from extremely warm or extremely cold climates, compromising the national representativeness of the sample. Given the algorithm's sensitivity to units of measurement, numerical features were also standardized.

Ultimately, 18,496 sampled households were cut down to 17,601 and dimensionality of the dataset was reduced to 10 features.

3. Model evaluation and implementation

Given that the dataset consisted of mixed variable types, I used k-prototype clustering, a centroid-based method which applies *k-modes* on *categorical* data types and *k-means* on *numerical* data.

As for the initialization of centroids, I chose the *Cao* method over that of *Huang*, as the former selects initial centroids based on density (i.e. points that are well-separated from others are more likely to be selected as initial centroids), whereas the latter is simply focused on selecting diverse centroids that represent different patterns in the data. While the *Cao* method tends to be more robust, it can be computationally expensive (Jahangiry 2024). Hence to strike the right balance between computational expense and cluster quality, I opted for smaller number of iterations at *n_init* = 5, instead of the default parameter of *n_init* = 10.

To determine the optimal number of clusters, I compared the total intra-cluster variation (*cost*) across different levels of k, which ranged from 1 to 10. For *numerical* variables, the cost was calculated as the sum of squared distances between the data points and the centroid of their respective clusters (*Euclidean distance*). For *categorical* variables, the cost was measured as the number of mismatches between the data points and the modes of categorical attributes within their respective clusters (*Matching dissimilarity*). Since small-sized clusters result in observations being located closer to their respective centroids, cost constantly decreases as the number of clusters increases, albeit at decreasing rates. Using the elbow method, I identified the point at which the improvement in cost starts to level out, which suggests that increasing the number of clusters beyond this point would provide little improvement in the cost. According to **Figure** 2, the optimal level of k was in between 3 to 5, but I ultimately opted for $k = 4$ based on its interpretability.

4. Results

Upon fitting the k-prototype model at $k = 4$, I generated summary statistics for each cluster, calculating the *mean* for *numerical* features and the *proportion* for *categorical* features. Carbon footprint

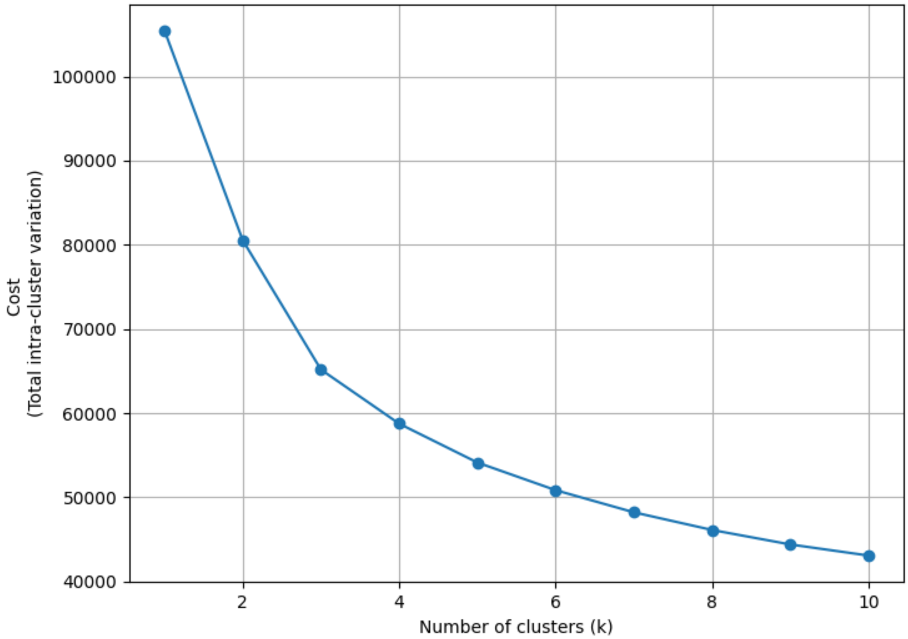

Figure 2. An elbow plot for determining the optimal number of clusters

attributable to the entire workflow – from data cleaning, model evaluation, to cluster analysis – amounted to 56.67g. The analysis revealed four distinct profiles of Americans:

1. **The *polluting elite*** (4,608 households);
2. **The *thrifty* but *energy-inefficient*** (6,720 households);
3. **The *overburdened*** (2,954 households); and
4. **The *modest consumer*** (3,319 households)

All in all, U.S. households differed along 3 dimensions:

1. Demand for energy;
2. Actual energy consumption; and
3. Sociodemographic characteristics

According to **Figure 3**, all but the *modest consumer*s (Cluster 4) were located in cold climates. Accordingly, the *polluting elites* (Cluster 1), the *thrifty but energy inefficient* (Cluster 2), and the *overburdened* (Cluster 3) had higher need for space heating whereas the *modest consumers* had higher demand for air conditioning.

But *actual energy consumption* and *sociodemographic characteristics* are along which I find some interesting distinctions (**Figure 4** – **Figure 11**). Firstly, the *polluting elites* are the energy gluttons, consuming large amounts of energy — whether it is adjusted or unadjusted — and occupying larger-than-average homes, despite the small size of their household. They are affluent, well-educated, and consist of predominantly (non-Hispanic) White households.

Secondly, the *thrifty but energy inefficient* consume the least amount of energy in raw aggregate terms, owing to their 'humble' lifestyle. They tend to live in small-sized homes ideal for two, which is well-suited for the needs of their household. However, a closer examination reveals that they consume large amounts of energy when adjusted for, both in terms of EUI and energy use per capita.

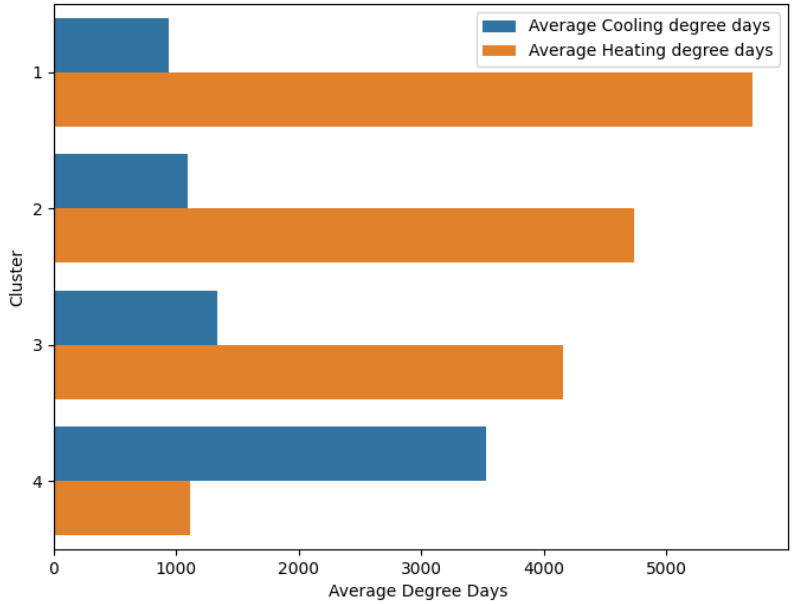

Figure 3. Average demand for energy - for heating (*orange*) and cooling (*blue*)

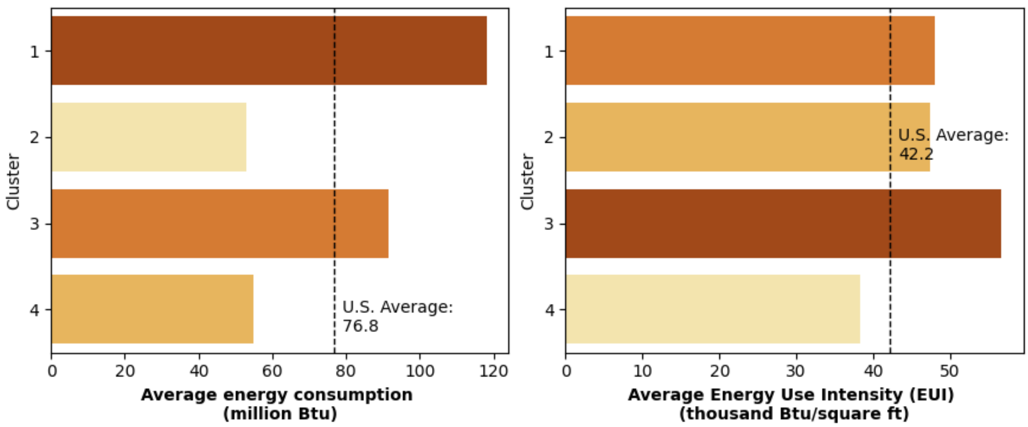

Figure 4. Energy usage, unadjusted (*left*) and adjusted for the size of the home (*right*)

It turns out that they generally live in poorly sealed or poorly insulated homes. Along with the *overburdened*, the *thrifty but energy inefficient* comprise of the socioeconomically disadvantaged. [6]

Thirdly, the *overburdened* consume excessive amounts of energy when adjusted for the size of the home. In support of this finding, a large majority of the *overburdened* live in drafty homes which tends to kick heating systems into an overdrive, jacking up energy costs. Although a sizable proportion of such households grapple with energy insecurity, they consume large amounts of energy even in raw terms. This suggests that energy expenditures can add to their financial constraints. However, they have minimal energy consumption per capita. Contrary to the *polluting elites*, the *overburdened* live in overcrowded homes, have lower educational attainment, and consist of high rates of households of color.

Lastly, the *modest consumers* use relatively low levels of energy, both in aggregate and per capita

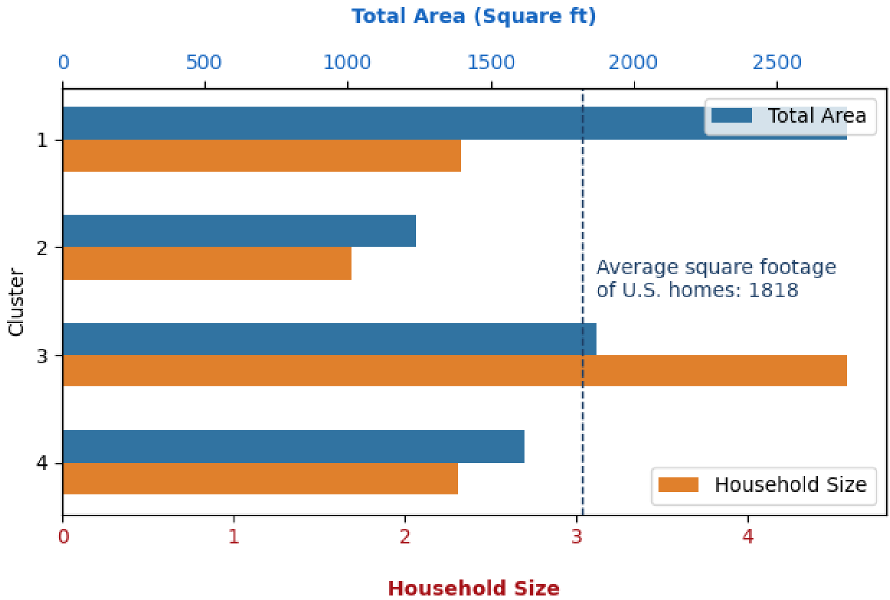

Figure 5. Average *area of housing unit* and average *household size*

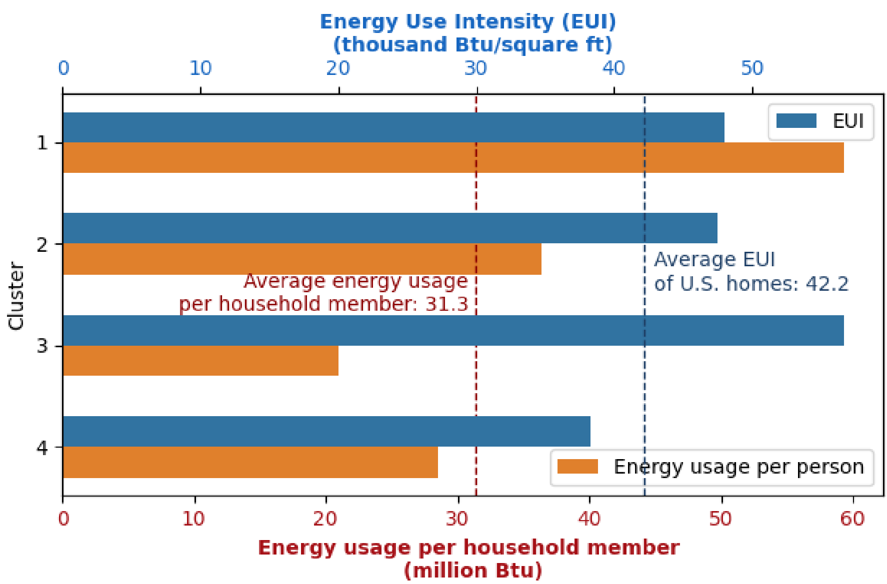

Figure 6. Average energy consumption, adjusted for *area of housing unit* and *household size*

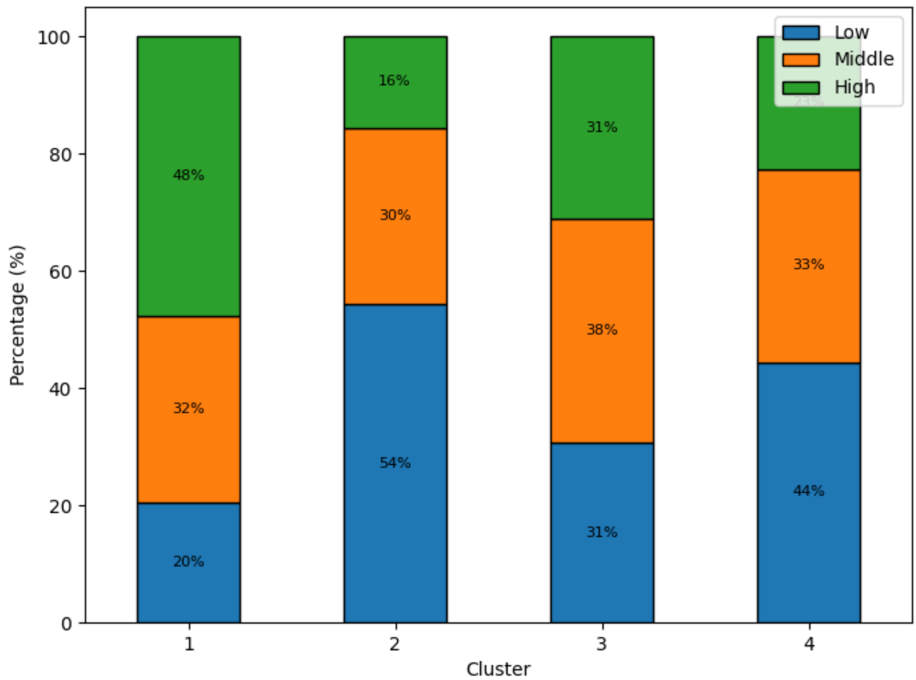

Figure 7. Proportion of households, by *income level*

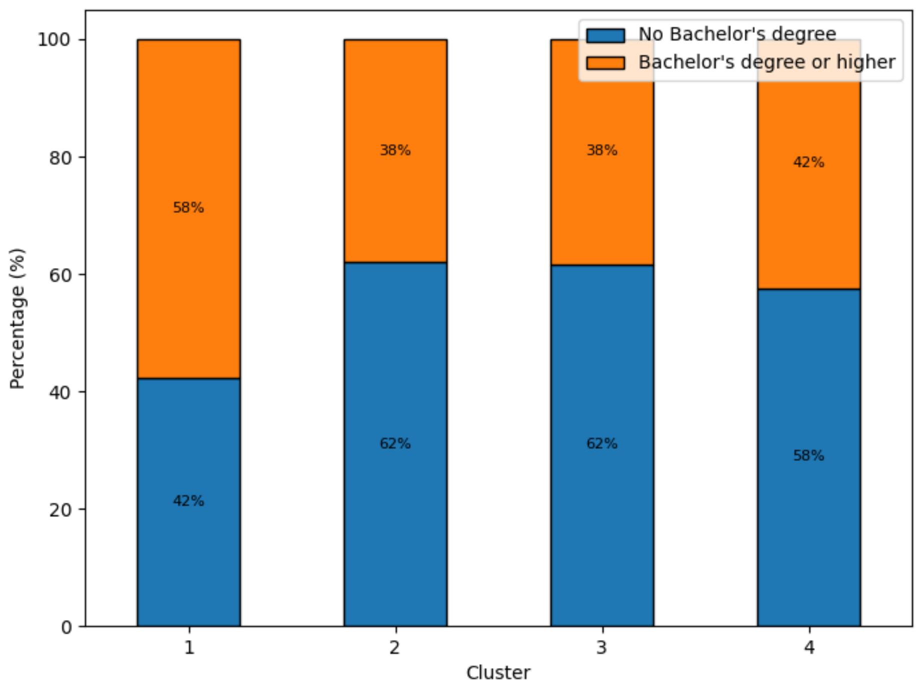

Figure 8. Proportion of households, by *education level*

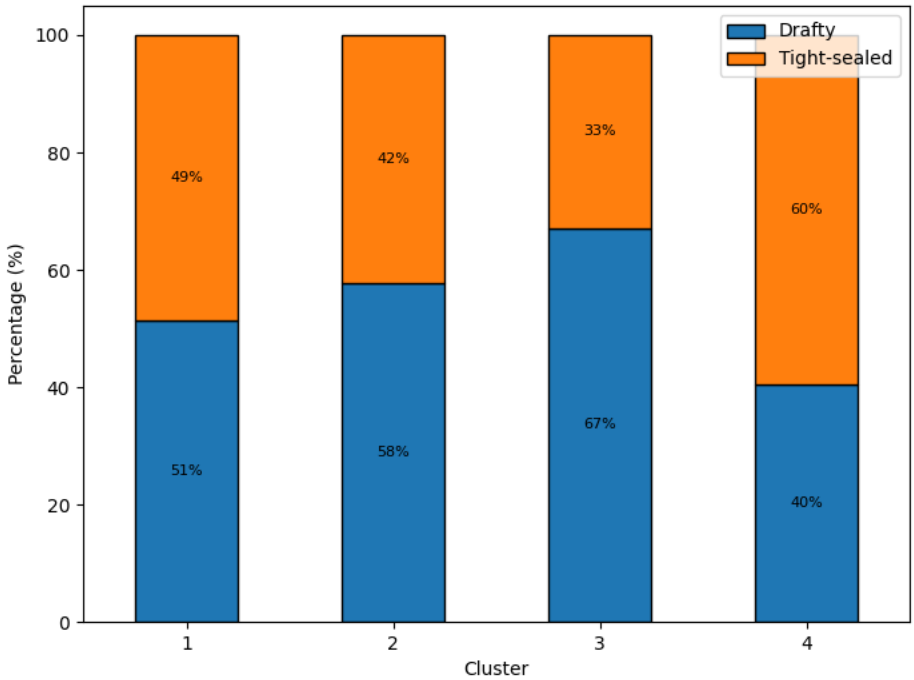

Figure 9. Proportion of households, by *draftiness*

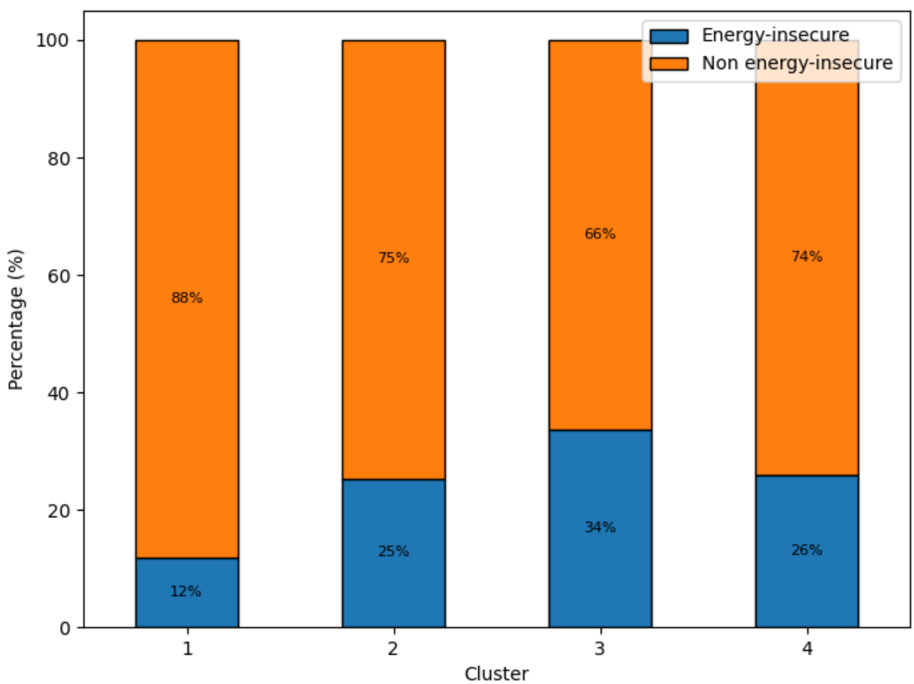

Figure 10. Proportion of households, by *energy insecurity*

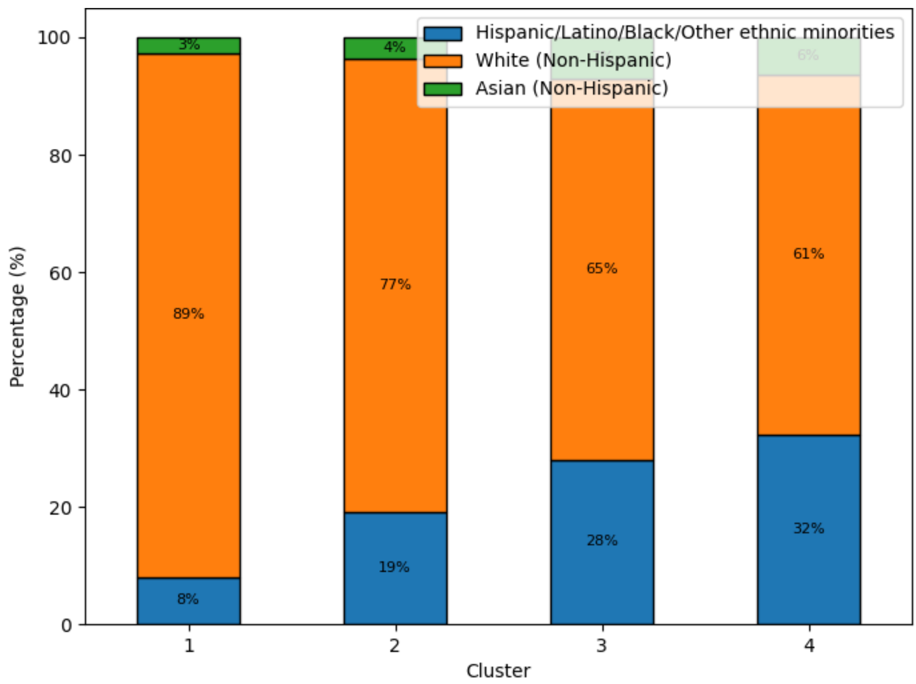

Figure 11. Proportion of households, by *race* and *ethnicity*

terms. The energy performance of their homes is also decent. As their name suggests, the *modest consumers* are moderate in every regard — they live in adequately sized homes given the number of occupants, comprise of mostly low-to-middle income households, and are evenly split across education levels.

In that the *polluting elites* and the *thrifty but energy inefficient* account for the two largest segments of the sampled U.S. households, it raises concerns from both an energy *efficiency* and an energy *equity* perspective.

5. Conclusion

5.1 Policy implications

In sum, Americans' energy consumption patterns and demand for heating and cooling are characterized by diversity. Interestingly, low-income households come in many different forms. Some live in cramped and overcrowded housing to manage their living expenses, but consume disproportionately large amounts of energy given what they can afford. Others cut corners by reining in the amount of energy usage in itself. But in fact, they pay more than they otherwise could have — due to poor energy performance of the home. These results suggest that the government's weatherization and energy efficiency programs should tailor 1) the *overburdened* and 2) the *thrifty but energy-inefficient*. Meanwhile, sacrifice is required on the part of the *polluting elites*, whose downsizing will help cut down America's carbon emissions.

5.2 Limitation and future directions

Although this study presents extensive and fascinating insight into the energy consumption landscape of the U.S., it had several limitations. First, the clustered households were not as clearly distinguishable, particularly along categorical features such as education level. However, by drawing inferences from not only *household level* but also *building unit* characteristics, I was able to identify four profiles of American households that — in qualitative terms — were convincing and tenable.

Secondly, current research was focused mainly on the structural features that affect energy consumption, such as socioeconomic status, building unit characteristics, and geographic conditions. From a behavioral standpoint, however, structural elements necessitate strong, fundamental levers for change that takes time for it to build momentum. Behavioral interventions and nudges, on the other hand, offer cost-effective measures for cutting down on wasteful energy consumption. Hence, future research should look into energy-saving behaviors that are amenable to change — such as the use of energy-efficient appliances and the use of smart thermostats — and typify households based on these characteristics. Identifying the laggards in the 'pro-climate' space and potential reasons for their behavior would help devise strategic interventions and set America on a promising path towards net-zero.

Notes

1 Heat pumps — which function as a dual heater and air conditioner — are an eco-friendly alternative to their traditional fossil-fuel-reliant counterparts as they run solely on electricity and operate by means of *transferring*, rather than *generating* heat (Kurzius 2022).

2 Major end-uses of energy include space heating, air conditioning, water heating, lighting, refrigeration, cooking, television, clothes washers, and dryers etc. (U.S. Energy Information Administration 2023c).

3 Cooling/heating degree days is a measure of 1) how extreme the outdoor temperature was for a given day and 2) for how long it was at that temperature:

$$\Sigma \left(\left| \textit{Average Outdoor Temperature} - 65^{\circ}F \right| \times \textit{Days} \right)$$

As the baseline ($65^{\circ}F$) refers to the temperature at which it is neither cold nor hot, degree days is a proxy of the amount of energy needed to heat or cool the building to regulate temperatures to comfortable levels. Higher the heating (cooling) degree days, colder (hotter) the average temperature and thus greater the overall need for heating (cooling) (U.S. Energy Information Administration 2023d).

4 Energy insecurity refers to the inability of a household to meet adequate energy needs. It has three dimensions: 1) financial hardship associated with the cost of energy relative to household income (economic), 2) deficiencies in the physical infrastructure of the home which result in thermal discomfort, harmful indoor exposure, and/or increased energy costs (physical), and 3) the use of adaptive strategies to counteract the impacts of economic and physical hardships that are associated with energy insecurity (coping) (Hernández 2016).

Although *income level*s and *energy insecurity* are correlated with each other, both measures of socioeconomic status were incorporated into the clustering algorithm because no single measure could draw an accurate picture of the hardships associated with meeting energy costs. For instance, households in the lowest income bracket did not necessarily struggle with paying their energy bills, if they had retired and had sufficient savings. Also, when climatic conditions were mild as such that it did not require much heating or air conditioning throughout the year, low-income households were subject to less of these financial constraints. On the other hand, if the algorithm were to cluster observations solely based on *energy insecurity*, it would fail to identify households that are on the borderline of socioeconomic vulnerability due to the binary nature of the category. Hence, the two measures of socioeconomic status had to be examined in tandem.

5 Hispanic or Latino origin (i.e. a person of Cuban, Mexican, Puerto Rican, South/Central American, or other Spanish culture or origin) relates to *ethnicity*, not *race*. Note that Hispanic/Latinos can be of *any race*, such as White, Black, or mixed (U.S. Census Bureau 2022).

6 Due to class imbalance, non-energy insecure households and non-Hispanic White samples comprise of the lion's share of all the clusters. Hence the focus of the analysis is to identify clusters with larger percentage of energy-insecure households and households of color, in *relative* terms.

References

Badger, Emily, and Christopher Ingraham. 2015. The most popular type of home in every major American city, charted. *Washington Post* (September). Accessed May 13, 2024. https://www.washingtonpost.com/news/wonk/wp/2015/09/21/the-most-popular-type-of-home-in-every-major-american-city-charted/.

Frey, William. 2020. The nation is diversifying even faster than predicted, according to new census data. *The Brookings Institution* (July). Accessed May 13, 2024. https://www.brookings.edu/articles/new-census-data-shows-the-nation-is-diversifying-even-faster-than-predicted/.

Hernández, Diana. 2016. Understanding 'energy insecurity' and why it matters to health. *Social Science & Medicine* 167 (October): 1–10. Accessed April 2, 2024. https://doi.org/10.1016/j.socscimed.2016.08.029. https://www.sciencedirect.com/science/article/pii/S0277953616304658.

Jahangiry, Pedram. 2024. *Clustering: K-mode and K-prototypes,* February. Accessed April 1, 2024. https://github.com/PJalgotrader/Machine_Learning-USU/blob/main/Lectures%20and%20codes/Module%2012-%20Clustering/python/Kmode_kprototypes.ipynb.

Kochhar, Rakesh, and Mohamad Moslimani. 2023. *Wealth surged in the pandemic, but debt endures for poorer black and Hispanic families.* Technical report. Pew Research Center, December. https://www.pewresearch.org/wp-content/uploads/sites/20/2023/12/RE_2023.12.04_Race-Wealth_Report.pdf.

Kurzius, Rachel. 2022. What is a heat pump, and should I get one? *Washington Post* (December). Accessed March 26, 2024. https://www.washingtonpost.com/home/2022/12/13/what-is-a-heat-pump/.

The Global Carbon Atlas. 2023. *Carbon emissions.* Accessed May 1, 2024. https://globalcarbonatlas.org/emissions/carbon-emissions/.

The White House. 2021a. *Fact sheet: President Biden takes executive actions to tackle the climate crisis at home and abroad, create jobs, and restore scientific integrity across federal government,* January. Accessed May 6, 2024. https://www.whitehouse.gov/briefing-room/statements-releases/2021/01/27/fact-sheet-president-biden-takes-executive-actions-to-tackle-the-climate-crisis-at-home-and-abroad-create-jobs-and-restore-scientific-integrity-across-federal-government/.

———. 2021b. *President Biden's historic climate agenda,* January. Accessed May 2, 2024. https://www.whitehouse.gov/climate/.

———. 2024a. *Fact sheet: Lowering home energy costs through president Biden's Inflation Reduction Act,* February. Accessed May 2, 2024. https://www.whitehouse.gov/briefing-room/statements-releases/2024/02/12/fact-sheet-lowering-home-energy-costs-through-president-bidens-inflation-reduction-act/.

———. 2024b. *Fact sheet: President Biden marks Earth Day 2024 with historic climate action,* April. Accessed May 10, 2024. https://www.whitehouse.gov/briefing-room/statements-releases/2024/04/22/fact-sheet-president-biden-marks-earth-day-2024-with-historic-climate-action/.

U.S. Census Bureau. 2022. *About the Hispanic population and its origin,* April. Accessed February 27, 2024. https://www.census.gov/topics/population/hispanic-origin/about.html.

U.S. Energy Information Administration. 2023a. *U.S. energy insecure households were billed more for energy than other households,* May. Accessed April 5, 2024. https://www.eia.gov/todayinenergy/detail.php?id=56640.

———. 2023b. *U.S. households in warmer states consume less site energy than households in colder states,* May. Accessed April 5, 2024. https://www.eia.gov/todayinenergy/detail.php?id=56380.

———. 2023c. *Use of energy in homes,* December. Accessed February 24, 2024. https://www.eia.gov/energyexplained/use-of-energy/homes.php.

———. 2023d. *What is a degree day?,* August. Accessed April 1, 2024. https://www.eia.gov/energyexplained/units-and-calculators/degree-days.php.

———. 2024a. *2020 Residential Energy Consumption Survey (RECS) data,* January. Accessed March 9, 2024. https://www.eia.gov/consumption/residential/data/2020/index.php?view=microdata.

———. 2024b. *Monthly energy review.* Technical report. June. https://www.eia.gov/totalenergy/data/monthly/pdf/mer.pdf.