US Utility Rate Disparity Analysis

Amir Zaman
Computer Science
Kean University
Union, United States
zamanam@kean.edu

Ayan Desai
Computer Science
Kean University
Union, United States
desaiay@kean.edu

***Abstract*—Utility policies vary geographically, reflecting differences in utility corporation practices across regions. Our objective is to examine the disparities in housing affordability across different geographic locations by comparing individuals' tax information in the United States to utility rates in 2020. This study aims to shed light on the implications of these discrepancies in essential living expenses and understand how they intersect with socioeconomic factors.**

Methods: We employ data visualization techniques such as scatterplots, histograms and geo-plots of the US by zip code and by county, attribute correlation analysis, and normalization. Histograms illustrate the distribution of individual incomes across income ranges, highlighting the disproportionate burden on lower-income households that spend a larger proportion of their income on utilities. Normalization allows us to uncover uneven distributions, while heat maps visualize higher and lower-priced utility locations in relation to low and high-income housing areas.

Further investigation is necessary to fully comprehend the multitude of factors affecting individuals' financial burdens and the unforeseen influences on utility pricing. Nonetheless, this study represents a significant step towards uncovering inequities inherent in the complexities surrounding essential services.

I. Introduction

Policy that is predatory towards low to mid-income residents is a possible cause of inequity in housing, a concern frequently emphasized in media discussions. Utilities, an integral component of housing, particularly electricity, are the focus of our study. Illuminating trends in utility pricing, particularly profit-driven charges, can offer insights for policymakers to alleviate financial stress among those residing in mid to low-income brackets.

Several related papers have delved into the intricacies of utility pricing and its implications. One study, spanning two decades of policy data from 1980 to 2000, extrapolated that policymakers, upon gaining access to energy information, tend to enact rate decreases rather than increases [1]. Another paper traced the historical roots of electricity back to 1887 and argued that it should be impervious to income levels, while highlighting the significant impact of fuel prices on electric rates. It noted the advantageous position of commercial and industrial complexes in bargaining for regulation, historically overshadowing residential benefits until advancements in technology paved the way for a more equitable distribution of regulatory benefits among consumers [2].

Our objective is to ascertain if there exists a correlation between income levels and residential electricity rates, identifying potential income demographics disproportionately burdened by investor-owned utility rates. With housing affordability in the United States dwindling for middle and lower-income households, we seek to elucidate how utility rates might contribute to the escalating financial strain on housing.

To achieve our goals, we employ correlation analysis on median incomes, residential rates, commercial rates, industrial rates, and the disparity between low and high-income households. Additionally, we utilize visualization techniques to enhance our understanding of the relationship between these variables.

The findings of our study serve as a foundation for further exploration into other utilities and their relationship with census data. They offer insights into the demographic classifications of the United States when analyzing residential data and can inform policymakers in crafting progressive policies aimed at alleviating housing-related financial challenges faced by individuals.

II. Materials

For this study, we utilized data sourced primarily from the US Census and the US Energy Information Administration (EIA) [3]. The US Census data from 2020 provided us with comprehensive information, initially comprising 41,691 records. From this dataset, we extracted relevant variables such as income ranges, median income, zip codes, county, and state. Specifically, we focused on income information by zip code, filtering out non-state zip codes and those lacking population or income data to ensure data quality and minimize noise. The final census dataset consisted of 24,037 records.

In tandem, we obtained utility data from Investor-Owned Utilities (IOUs) [4], starting with 52,178 records. This dataset, compiled from sources within the US Energy Information Administration dataset 861, aggregated IOU rates by zip code. To streamline the analysis, we grouped the rates by zip code, disregarding specific company names and service types, and computed the mean rates served by different providers within each zip code. This approach facilitated subsequent analysis by allowing us to join the datasets by zip code, county, and state.

During the data extraction process, we also took into account the need for geographical mapping using tools like Geopandas. Therefore, in addition to median income and income range percentages, we extracted population and

zip	eiaid	utility_name	state	service_type	ownership	comm_rate	ind_rate	res_rate
85321	176	Ajo Improvement Co	AZ	Bundled	Investor Owned	0.087890499	0	0.093887147
36560	195	Alabama Power Co	AL	Bundled	Investor Owned	0.121895087	0.063651538	0.135056714
36513	195	Alabama Power Co	AL	Bundled	Investor Owned	0.121895087	0.063651538	0.135056714
36280	195	Alabama Power Co	AL	Bundled	Investor Owned	0.121895087	0.063651538	0.135056714
35473	195	Alabama Power Co	AL	Bundled	Investor Owned	0.121895087	0.063651538	0.135056714

Fig. 1. Example records from Investor Owned Utilities dataset

state/county information from the census dataset for shape file mapping. Similarly, from the IOU dataset, we focused solely on rates and zip codes. Given that the raw IOU data was organized by zip code and provider, we averaged the rates when grouping providers into zip codes to maintain data coherence, acknowledging that this may lead to minor discrepancies but should preserve overall trends and correlations.

Following data manipulation and refinement, the census dataset was reduced to 24,037 records, while the IOU dataset was trimmed down to 31,364 records. It's important to note that all data manipulation occurred in Python, while the original raw data remained unchanged in the MySQL database, serving as the data warehouse for this study. This systematic approach ensured the integrity of the datasets and facilitated rigorous analysis to explore the correlations between income levels and utility rates at various geographical levels.

III. Methods

A. Methods Used

Our analysis employed various statistical and visualization techniques to explore the relationship between income levels and utility rates. Key methods include:

- Standardized Z-score and original mean, median, standard deviation: Numerically describe the visualizations and statistical measures.
- Correlation analysis: Investigates the direct relationship between median income and residential rates.
- Visualization techniques: Utilized at both Zip and County levels to explore spatial patterns and distributions.
- Histograms: Visualize the distributions of variables.
- Scatterplots: Examine direct relationships between key variables.
- Geographical plots by county: Explore the influence of location on the variables under study.

B. Functions

To facilitate the execution of our study, several Python scripts were developed, each serving a specific purpose:

- experiment.py: Sets up and tests the Anaconda environment and verifies the functionality of map-generating code.
- cleanincome.py: Retrieves data from the warehouse in the form of a Pandas DataFrame, cleans it, and saves it as a CSV file for further analysis.
- cleaniou.py: Performs similar data cleaning procedures for the second dataset pertaining to Investor-Owned Utilities (IOUs).
- hist.py: Generates histograms to visualize distributions of key variables.
- boxplot.py: Creates boxplots to depict the distribution of variables of interest.
- stats.py: Computes statistical measures of interest.
- maps.py: Generates geographical plots to visualize spatial relationships.
- corr.py: Calculates correlations between variables of interest.

The MySQL Workbench was utilized to load the raw data into the data warehouse. Once the initial query was executed, resulting in the extraction and cleaning of data stored in CSV files, all scripts were executed locally to perform subsequent analyses.

C. System and Tools

Our study was conducted using the following system, hardware, software, and tools:

- System: The server operated on Linux v 3.10.0, while the client utilized Windows.
- Hardware: Analysis was performed on an MSI GE 76 Raider laptop.
- Web Server: Apache facilitated web services.
- Database: The data warehouse was a MySQL database server, specifically MariaDB v 5.5.68.
- Languages: Python and MySQL were the primary languages employed.
- API/Libraries: Utilized libraries included Pandas, Geopandas, csv, mysql.connector, Matplotlib, and NumPy.
- Tools: MySQL Workbench for query design and warehouse setup, Anaconda Navigator for Python environment setup, VSCode for code development, and Git for version control and code sharing across multiple machines.

IV. Experiment Results

In this section, we delve into the comprehensive analysis of the experimental results, presenting a meticulous examination of all relevant data, accompanied by insightful plots and rigorous calculations derived from the gathered information. Our exploration not only provides a detailed overview of the experimental outcomes but also offers a deeper understanding of the underlying phenomena observed throughout the study. Through meticulous scrutiny, we aim to elucidate the significance of our findings and their implications in the broader context of the research domain.

V. Discussion

Challenges During Research:

Throughout the course of our research, several challenges were encountered, each requiring thoughtful consideration and strategic approaches to overcome:

1) Geo Plotting Complexity: One significant challenge revolved around generating accurate geo plots. This necessitated further data grouping and the implementation of shape files [5] from the Geopandas documentation. Parsing through numerous fields from tables and determining their significance was paramount in presenting the

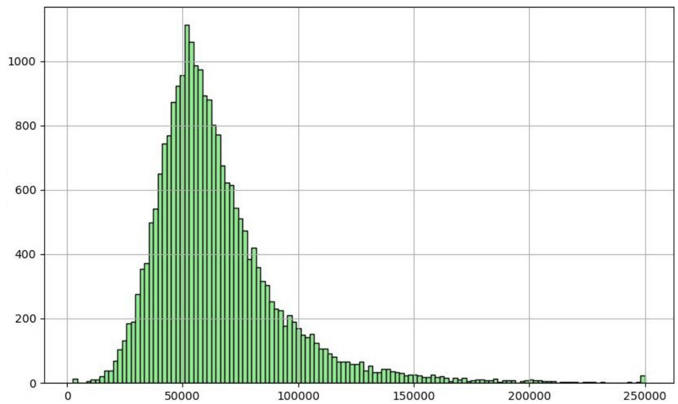

Fig. 2. Distribution of median incomes across the US by zip code.

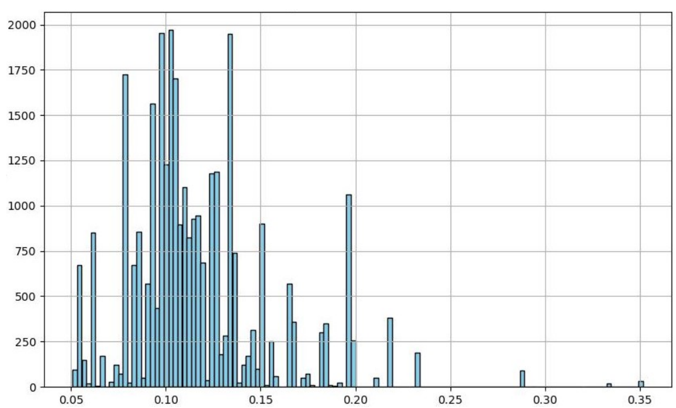

Fig. 3. Distribution of investor owned electricity rates across the US by zip code.

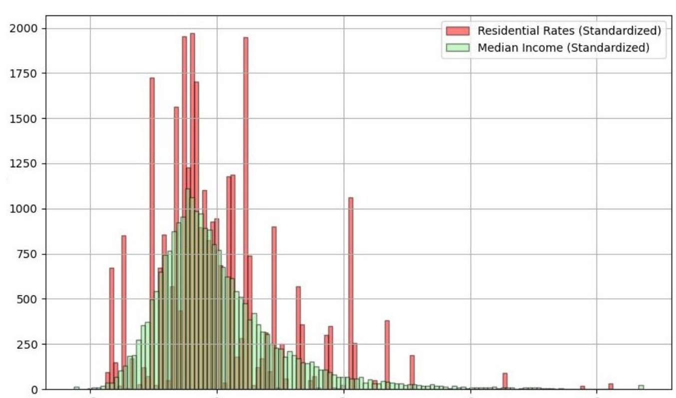

Fig. 4. Normalized distributions of median income (green) layered over residential rates (red).

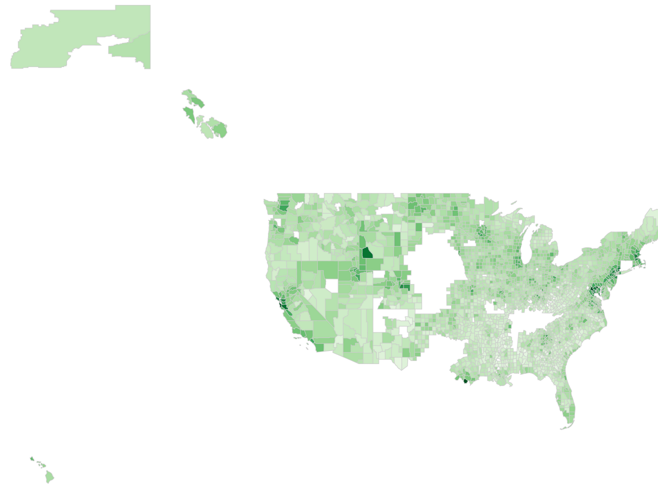

Fig. 5. Distribution of median incomes geographically by county.

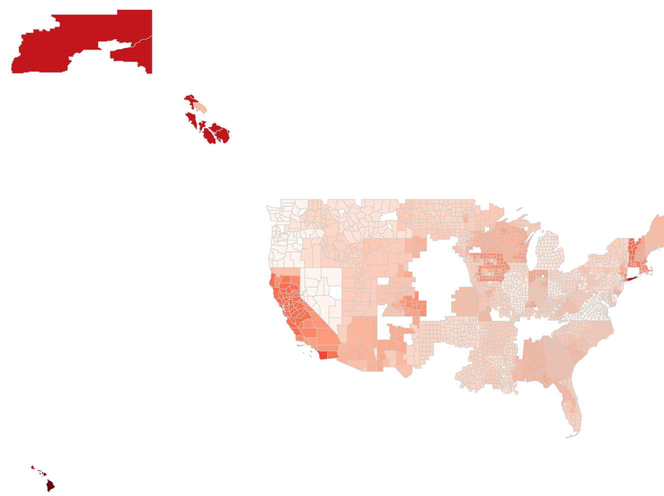

Fig. 6. Distribution of investor owned electric rates geographically by county.

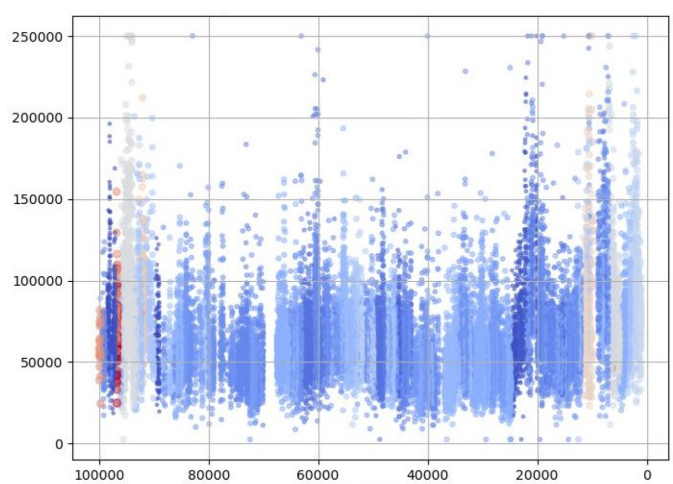

Fig. 7. Zip codes plotted by median incomes (y-axis) and zip code (x-axis) where the redder a point is the higher the electric rates are for that zip code.

State	Abrv	Correlation
Hawaii	HI	0.546632
Alaska	AK	0.438161
Washington	WA	0.430316
Minnesota	MN	0.406674
New York	NY	0.367856

Fig. 8. Table of highest 5 correlations ran at the state level with median income and residential rate. Patterns can be seen relating to the zip codes found in these states and in Fig 7.

information in a meaningful manner to unveil previously undiscovered insights. While visualizing and classifying based on zip code provided a tighter geographical scope, it offered potentially more precise results compared to aggregated county data.

2) Scope Constriction: As we delved deeper into the dataset, it became evident that the scope of the issue was constrained by the size of the dataset. Expanding the analysis to include more column correlations could potentially reveal additional insights. However, this expansion poses its own set of challenges, including data management and computational complexity.
3) Interpretation Challenges: An additional hurdle arose during the interpretation of scatterplots, particularly when comparing population on one axis against population percentage. The variability in readings obtained from these plots underscored the need for robust analytical methodologies to ensure accurate interpretation and reliable conclusions.

Future Work:

Looking ahead, several avenues for future research and exploration present themselves, each promising further insights into the dynamics of utility rates and their impact on different demographic groups:

1) Commercial and Industrial Rates Analysis: Extending our analysis to encompass commercial and industrial utility rates holds significant promise. Enterprises often wield greater negotiating power compared to individual consumers, potentially resulting in disparities in kilowatt-hour rates. Exploring these trends and their implications could shed light on inequities in utility pricing across different sectors.
2) Temporal Analysis: Examining utility rate trends over different time periods, beyond the scope of the current census data, could offer valuable insights into the evolution of policy and its impact on utility pricing. However, this endeavor presents logistical challenges, given the lengthy process of aggregating utility information and the infrequency of US Census data collection, which occurs every ten years.
3) Further Geographical Exploration: Expanding our geographical analysis beyond the confines of zip codes to include other spatial units, such as census tracts or blocks, could provide a more granular understanding of regional disparities in utility rates. This approach would require careful consideration of data aggregation and visualization techniques to ensure meaningful interpretation of results.

In addressing these areas of future work, we aim to deepen our understanding of the factors influencing utility pricing and advance efforts to promote equitable access to essential services for all demographic groups.

VI. Conclusion

In summary, our analysis suggests that there is limited evidence of a direct correlation between residential rates and household income. The observed similarities in certain trends can largely be attributed to external factors, particularly geographic location. States situated far from power generation sources or characterized by high levels of urban development, such as Long Island, New York, exhibit higher residential rates and consequently demonstrate a stronger correlation. However, it is essential to recognize that this issue is multifaceted, influenced by various factors including fuel costs, political climates, and other socioeconomic dynamics not explicitly explored in this study. Thus, while our findings contribute valuable insights to the discourse on energy, housing, and financial stability across the United States, they represent only a preliminary step in the broader exploration of this complex interplay. Further research incorporating additional variables and a more comprehensive analysis framework is necessary to fully elucidate the nuanced relationships at play.

References

[1] Adam R. Fremeth, Guy L. F. Holburn, Information Asymmetries and Regulatory Decision Costs: An Analysis of U.S. Electric Utility Rate Changes 1980–2000, *The Journal of Law, Economics, and Organization*, Volume 28, Issue 1, April 2012, Pages 127–162, https://doi.org/10.1093/jleo/ewp042https://doi.org/10.1093/jleo/ewp042

[2] Jackson, R. (1969). Regulation and Electric Utility Rate Levels. *Land Economics*, *45*(3), 372–376. https://doi.org/10.2307/3145025

[3] "Department of Energy - U.S. Electric Utility Companies and Rates: Look-up by Zipcode (2020)." *Catalog*, Publisher National Renewable Energy Laboratory (NREL), 11 June 2023, catalog.data.gov/dataset/u-s-electric-utility-companies-and-rates-look-up-by-zipcode-2020.

[4] US Energy Administration. (2019, August 15). Investor-owned utilities served 72% of U.S. electricity customers in 2017 - U.S. Energy Information Administration (EIA). https://www.eia.gov/todayinenergy/detail.php?id=40913

[5] Bureau, US Census. "Cartographic Boundary Files." *Census.Gov*, Census Bureau, 1 Apr. 2024, www.census.gov/geographies/mapping-files/time-series/geo/cartographic-boundary.2020.html#list-tab-1883739534.