ENERGY POVERTY IN TEXAS: EXPLANATORY POWER OF SOCIAL, ECONOMIC, AND HEALTH INDICATORS

Course: Data Management Spring 2019;
Author: Agbim, Chinelo

Energy poverty is a global issue that occurs when a household experiences inadequate energy services for one's dwelling place (Boardman, 1991). Over the past 20-30 years there has been an increasing focus on the issue in developed countries. In Europe, there is understanding amongst researchers that energy poverty is a result of high energy prices, low incomes, inefficient buildings, and individual household practices and needs (Boardman, 2010; Thomson, Bouzarovski, & Snell, 2017). Energy poverty is especially acute in Texas where, like many other states, residential electricity prices have increased over the past two decades and electricity demand has increased due to increasingly severe weather (Yun & Steemers, 2009; DOE, 2016; Wible & King, 2016). These trends often create a compounded burden for vulnerable individuals living in older, less energy efficient housing (Valenzuela, et al., 2014).

The most common way of measuring energy poverty is energy burden; individuals who are living in energy burden spend more than a certain percentage of their income on energy bill. Typically people who spend more than twice the median amount of their income on energy. For example, the 4% is the median amount households spend on energy bills in Texas. Therefore, 8% is the threshold at which a household is considered energy burdened. Roughly 22% of Texas households are energy burdened (Wible & King, 2016). Yet some estimates show that less than 1 million Texans utilized energy assistance in 2016 (MALEWITZ, 2016).

One study estimated that individuals living in rural counties along the Texas-Mexico border may be spending 18.9% of income on energy bill, compared to the 4% state median (Olmedo, 2013). Measuring energy burden at a regional level as well as analyzing nonstructural indicators of energy burden is necessary to inform legislators and policy makers of the causes of the issues. While literature by European scholars have used both qualitative methods to analyze the regional, social, and economic indicators of energy poverty, much of the quantitative literature in the U.S. focused on energy efficiency.

Recent studies have also shown a link between health & food insecurity and energy burden (Tuttle & Beatty, 2017). Since geography is a strong predictor of these disparities (Bouzarovski & Simcock, 2017). Thus, the objective of the study is twofold. First, this study investigates how much of energy burden can be described by socio-demographic, economic, and health indicators at the county level. Secondly, this study seeks to investigate whether being a border county is a significant predictor of energy burden.

The results of this study hold implications for policy makers on what regions to focus energy burden programs. Additionally, this study adds to the sparse, but growing body of literature on what non-structural indicators lead to energy poverty. These results hold implications for identifying eligible customers for energy bill assistance. Findings show that socio-demographic, economic, and health indicators explain nearly half of the county level variance in energy burden. Results also show that border counties are a significant predictor of energy burden.

2. Literature Review

Over the past decade there has been increasing evidence that socio-demographic, geographic, and economic indicators are causes of energy poverty in the United States. Hernandez and Bird interviewed low income individuals in Boston revealing that lack of financial resources, housing instability, health issues, and energy inefficiency compounded into a heightened burden for low income households (Bird & Hernández, 2010). For instance, the authors assert residents who have critical health conditions are more likely to have higher energy bills. Additionally, individuals who cannot afford energy bills are likely to suffer from extreme weather conditions which leads to health problems (Bird & Hernández, 2010).

The authors of that study also posit that individuals living in homes they rented as opposed to owned homes, are less likely to be able to access energy poverty alleviating program like energy efficiency because their landlord must make the decision to make structural upgrade (i.e. principal agent problem) (Bird & Hernández, 2010). However, the authors do not provide any empirical evidence quantifying how many individuals may be facing this issue or its quantifying its impact on energy poverty. Similar to Hernandez and Bird another study posits that energy poverty should be thought of as a network of factors particularly socioeconomic, economic, regional, and structural factors (Bird & Hernández, 2010) (Harrison & Popke, 2011). The study used case study interviews in eastern North Carolina. Furthermore the focus was on heating bills, not energy expenditures overall.

More recently, another study used census blocks in Kansas City to identify socio-demographic and economic predictors of energy inefficiency (Reames, 2016). The results showed that census block level poverty percentage, percentage of individuals over the age of 65, and race/ethnicity as predictors of energy inefficiency (Reames, 2016). However, the study didn't include health indicators or address regional variation in energy burden.

Finally, on a national scale, an ACEEE 2016 report showed that amongst the largest 20 cities in the U.S. the median energy burden varied from 3% to 13% of income from city to city ((Ross & Drehobl, 2016). However, they did not develop any inferential statistics to test the influence of regions on energy burden. Finally in Texas, Olmedo show that in Colonias (informal settlements along the Texas-Mexico border) individuals may be spending 18-19% of income. The author interview ~340 households to create these estimates. However, they were able to focus in four border counties and did not develop any inferential statistics.

This study seeks to fill the gap in energy poverty research by providing an empirical quantitative analysis to investigate the relationship between energy burden (a metric of energy poverty) and social indicators. Specifically, we linear OLS model to assess the power of selected economic, socio-demographic, economic, health, and geographic (i.e. Texas-Mexico border counties) to explain low-moderate income (LMI) energy burden.

2. METHODS

The methods for this study include (1) data cleaning and merging, (2) regression models on a analysis and validation data set that mirror each other, (3) regression robustness checks and (4) GIS visualization. The general work flow of the analysis and validation models is shown in Appendix.

Analysis and Validation Regression Methods

The relationship between energy burden and socio-demographic, health, and economic indicators are tested using a linear regression. Linear model was chosen because all of the variables in the study are continuous, the aim is to understand the relative significance of the variables on energy burden (e.g. compare magnitude of the t-score), and finally linear models have high interoperability (James, Witten, Tibshirani, & Hastie, 2017). Median age of the houses built in each county is included as a control variable to represent relative energy efficiency.

The dependent variable, low to moderate income energy burden, is the estimated median energy expenditure as a percentage of the county low to moderate income[1] (LMI) (National Renewable Energy Labe, 2018). The socio demographic indicators included were percentage of population that is Hispanic[2], the percentage of the population that is African American, the percentage of the population above 65, and the percentage of individuals with some college education. The percentage of Hispanics and African Americans living in a household have been shown as indicators of energy inefficiency and energy burden before (Ross & Drehobl, 2016) (Reames, 2016).

Some studies have postulated that rural regions may be more likely to have customers that face energy poverty due to rural utilities' lack of financial resources to provide energy assistance programs (Bird & Hernández, 2010) (Harrison & Popke, 2011). Thus, percent of the population living in rural area is included as an indicator. The percentage of individuals over the age of 65 (i.e. seniors) at a county level was included as it has been shown that senior spend a disproportionate amount of their income on energy (Bird & Hernández, 2010) (Ross & Drehobl, 2016). Finally, education level has been shown to be strongly correlated with energy burden (Ross & Drehobl, 2016) and a significant predictor of energy inefficiency (Reames, 2016). Thus, the percentage of the population with some years of college education was included.

The economic indicators included were unemployment rate, percent of population living in poverty, percent of households owned, and unemployment rate. Percent living in poverty

[1] Households are considered LMI if income <80% of the median income.
[2] Specifically, the percent of the county that is Hispanic is an estimate the percentage of non-white Hispanic individuals in the county.

has been shown to be a significant indicator of energy inefficiency (Reames, 2016).Ownership has been shown to be a predictors of energy inefficiency (Reames, 2016) as well as a barrier to energy efficiency programs that alleviate energy poverty (Bird & Hernández, 2010) (Ross & Drehobl, 2016).

The health indicators included percent of population uninsured, of adult obesity, of household food insecurity, and of households with low access to a grocery store. Percent of the households uninsured has been shown to be a major indicator of health. Individuals without health insurance are likely to forego preventative care, and more likely to be hospitalized (Kaiser Family Foundation, 2018).

The financial burden of severe illness and severe disabilities can lead to energy poverty (Bird & Hernández, 2010) (Boardman, 2010). Obesity and low access to a grocery store can also be major predictors of severe illness (Food Research and Access Center, 2019). Often one of the qualifiers for energy bill assistance is food stamp eligibility (i.e. SNAP) (Public Utility Commission of Texas, 2017). However, the percent of the population eligible for SNAP was not available at the county level. Instead the percent of households that are food insecurity is used as a proxy for SNAP eligibility. Food rank was used in the validation model instead of food insecurity.

In order to test the level of variation in energy burden that is attributed to region, a dummy variable was included to represent whether or not a county was a border county. Here we use the TX DHS definition of a border county: a county that is with 69 miles of the Texas-Mexico variable (Texas Department of State Health Services, 2017). 32 counties are considered border counties. No energy efficiency data was available on a county level. Thus the median structure age for residential buildings was used as a proxy for how energy efficient households in a given county are. The full list of variables and unique identifiers are listed in the appendix.

Description	Variable Prefix	Data Type
Energy Burden (county energy expenditures as %of LMI)	En_burden	Continuous, nominal
Federal Identification #	fip	Nominal, integer
County Name	cnty	Nominal, integer
Socio Demographic/Economic Indicators		
% pop Hispanic	pcnt_hisp	Continuous, numerical
% pop Black African American	pcnt_blk	Continuous, numerical
% pop in rural community	pcnt_rural	Continuous, numerical
% pop over 65 years old	pcnt_sr	Continuous, numerical
%pop attended some college		
Economic Indicators		
Unemployment Rate	pcnt_unemp	Continuous, numerical
% pop Under Poverty Line	pcnt_pov	Continuous, numerical
% of TX State Median Income	pcnt_income	Continuous, numerical
% of Residential Units Owned	pcnt_own	Continuous, numerical
Food and Health		
% pop with Low Grocery Access	pcnt_lw_access	Continuous, numerical
% pop Obese	pcnt_obese	Continuous, numerical
% pop Food Insecure	food_insec or food_rank	Continuous, numerical
% pop Uninsured	pcnt_uninsured	Continuous, numerical
Control Variable and Dummy		
County is one of 32 Texas-Mexico border counties=1, If not=0	Brdr_cnty	nominal, factor (value 0 or 1)
Median age of the residential structures in the county	House_str_age	Continuous, numerical

Figure 1 Study Variables

In order to validate the results, a dataset with the same independent variables but from different data sources than the analysis set is used. The same was done for the linear model with the dummy variable. The variables were tested for correlation across the different data sources to ensure they were comparable (e.g. comparing Census reported unemployment to the USDA's reported unemployment). All variables had a Pearson's correlation magnitude greater than 0.5. A full list of the data sources can be seen in the next section.

Data Sources:

The data used for this study comes from: United States Department of Agriculture (USDA), County Health Ranking (CHR), United States Census Bureau and SAIPE, Bureau of Labor

Statistics(BLS), United States Housing and Urban Development (HUD). A list of data sources for each variable is listed below (Table 1).

Table 1. Variables and Sources

Analysis Variable and Source	Validation Variable and Source
Socio Demographic Indicators	
Non White Hispanic (% in 2013-2017), US Census 2017	Non White Hispanic (% pop in 2017), County Health Ranking 2019
African Am/Black (% in 2013-2017), US Census 2017	African Am/Black (% pop in 2017), County Health Ranking 2019
Rural (% pop in 2010), County Health Ranking 2019	SAME
65 and over (% pop in 2017), Census Bureau 2017	65 and over (% pop in 2017), County Health Ranking 2019
Some College (% pop with some college), County Health Ranking 2019	SAME
Economic Indicators	
Unemployment Rate (% June 2017), Bureau of statistics	Unemployment Rate (% 2017), County Health Ranking
Poverty (% pop), USDA 2017	Poverty (% pop) SAIPE, US Census Bureau 2017
Tenure Owner (% households 2013-2017%), US Census	Homeownership (% households 2013-2017), County Health Ranking 2019
Food and Health Indicators	
Low Access Grocery Store(% pop in 2015), USDA 2017, “PCT_LACCESS_POP15”	Limited Access to Healthy Food (% pop), County Health Ranking 2019
Adult Obesity (% pop in 2013) ,USDA 2017	Adult Obesity (%pop in 2015), County Health Ranking 2019
Food Insecure (% pop 2016), County Health Ranking 2019	Food Environment index (1-100 percentile scale higher is better, 2016), County Health Ranking 2019
Uninsured Adults (% pop), County Health Ranking 2019	SAME
Control Variable and Dummy	
Border County, Texas Department of Health	SAME
Median Household Structure Age, Housing and Urban Development 2017	SAME

Most of the variables were a percentage on the scale of 0-100. This was done to ease interoperability of coefficients. The only variables that weren’t on a 0-100 % scale were the border county dummy variable and the median age of household in a given county. Food insecurity was measured as the percent of the population that the USDA found to be food insecure. Food-environment index is a variable that ranks overall food health in a county on a percentile (i.e. higher number means better food security).

Research Workflow:

The workflow of the study is as follows: (1) data downloaded and saved, (2) data sets are cleaned merged to create an analysis and validation data set, and (3) linear models were created iteratively. The work flow for each of these stages are shown in the figures below (FIGURE 2-3)

The data and documentation were downloaded from their respective website and saved into folder on a desktop that was linked to a Github repository. This raw data was then loaded into a R script that was used to clean the data. In the script, the datasets that included observations other than Texas were removed. Additionally, irrelevant variables were deleted. The variables that were relevant were renamed (e.g. percentage of Black/ African American in a county renamed “pcnt_blk_datasource”).

Most data sets included a unique county ID number called the “Federal Information Process ID (FIP)” and/or the county name. The FIP was used as the unique identifier to merge that data sets into two new data sets: an analysis data set and a validation data set. All continuous variable were changed to the datatype “numeric”.

The same procedure described above was used to create a the datasets with the border region dummy variable: analysis data with a dummy and validation data set with a dummy. The dummy variable "brdr_cnty" was created using an if statement where is if the name of the county and FIP matched the DHS list of border counties, brdr_cnty=1. Otherwise, the variable brdr_cnty=0.

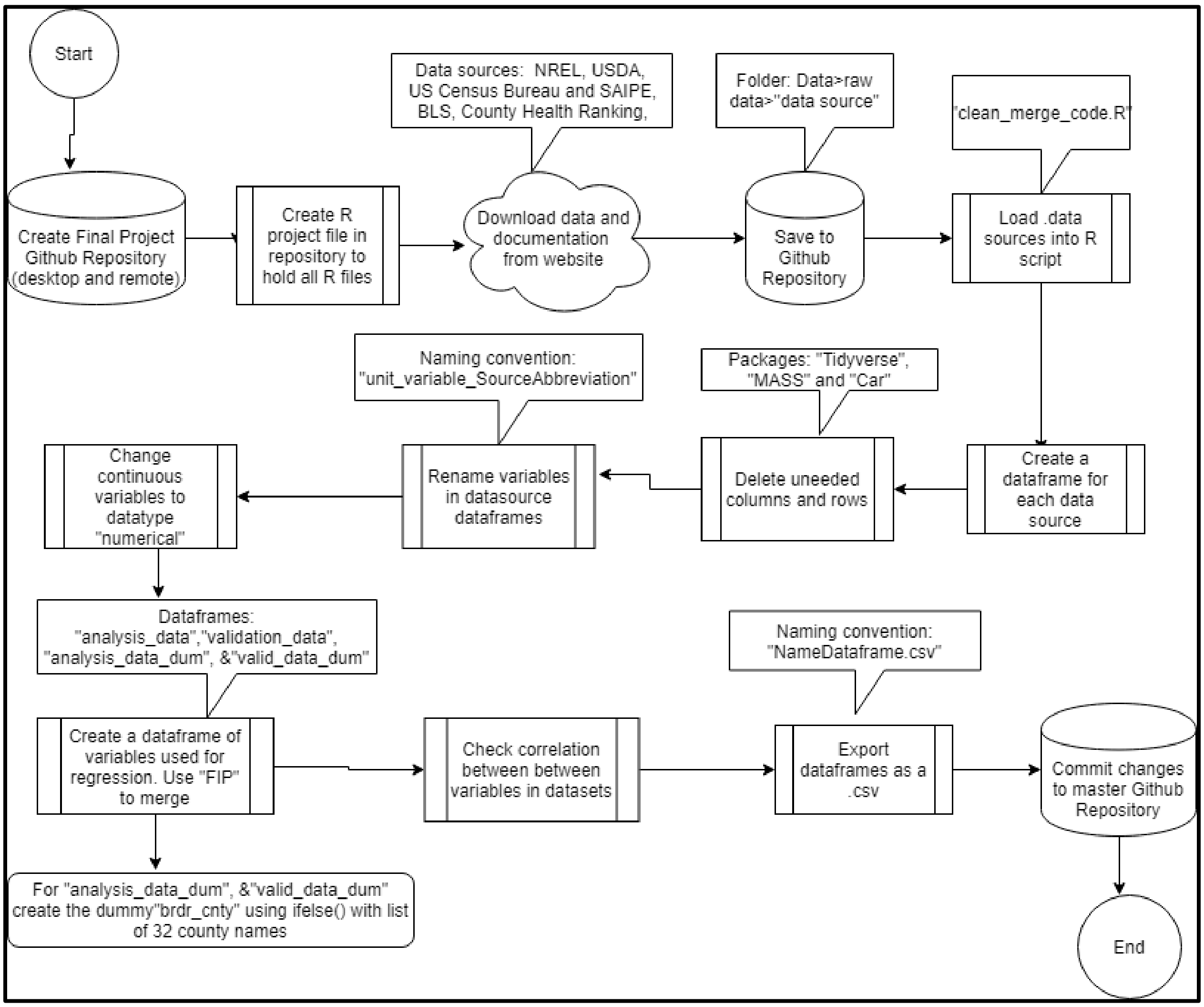


Figure 2 Data Clean and Merge Workflow

The models were run in a script written specifically for linear modelling. After a regression was run a summary table with Adj. R^2 , F statistic, and p-values were developed to assess explanatory power of the model and significance of variables. Robustness check were done to check for typical issues linear models have including: linearity between energy burden and each independent variable, multicollinearity, heteroscedasticity, error dependence were checked. In order to reduce these typical linear model issues, variables were removed in an iterative process.

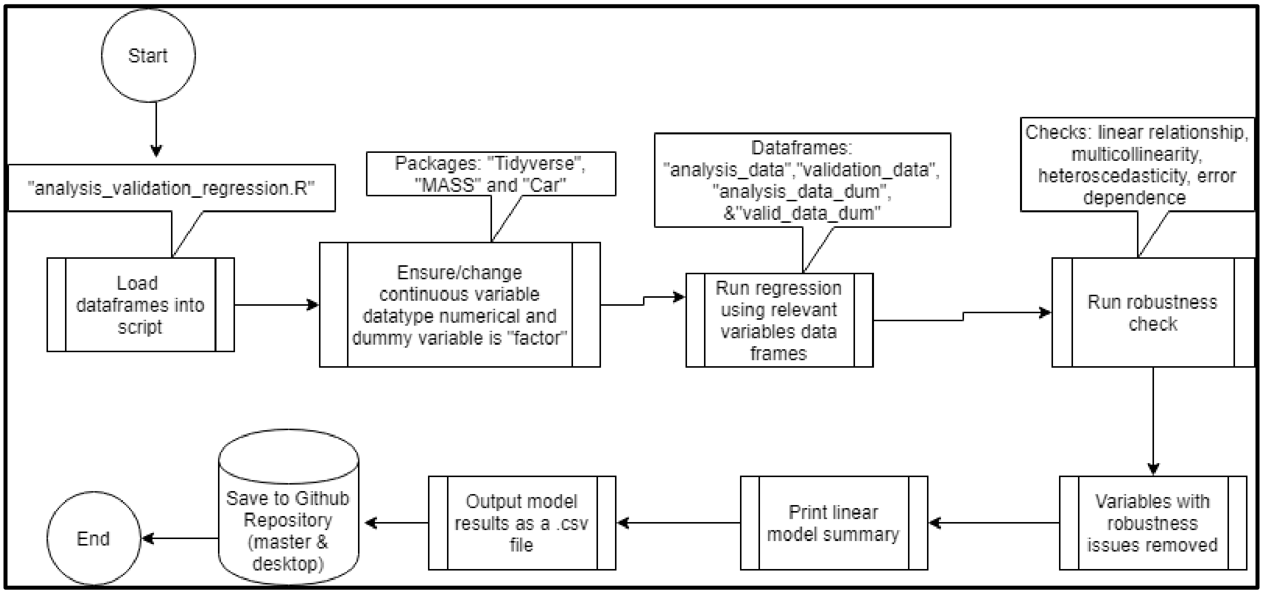


Figure 3 Linear Regression Model Workflow

The map visualizations were done using two datasets that were joined based on FIP. One dataset was a .shp file and the other was a .csv. The datasets were uploaded to ArcGIS and joined in ArcGIS.

Data Management Methods

Data management occurred through: (1) saving the entire study in two locations, (2) folder structuring, (3) file and variable naming conventions. All of the files used for this study where saved in a publicly available Github repository cloud. This repository was linked to the authors desktop and the cloud was updated after any changes were made to any of the files. The folder structure is pictured below (Figure 4). The folders were named after major functionalities: (1) data (raw data and clean data), (2) clean and merge code, (3) regression code, (4) outputs, (5) mapping, (6) writing, literature and presentations.

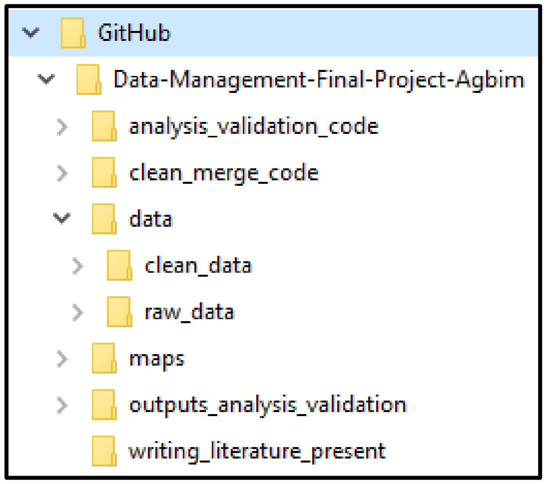


Figure 4 Repository Folder Structure

Within these folders there was a "README" file that described each file in the folder. For folders that had more folders within them, there was another "README" within that folder describing the files. For example, the raw data folder had folders in it that represented each data source (e.g. the raw data from USDA was inside a folder called "USDA"), and each of these folders had their own "README" that described the dataset, listed the URL, and date accessed.

In addition to the raw datasets having their own folders, documentation from the websites were downloaded as well. For instance, the County Health Ranking dataset had documentation explaining the meaning and units for each variable. Some websites had better documentation was better than others.

Variable naming convention was based on the unit, the variable and the data source. For instance, the percentage of the population that is African American measured by CHR was call "pcnt_blk_chr" in order to document where different variables came from as the results were interpreted. The full list of variable names (as they're seen in R script) is included in the Appendix. Additionally, file naming convention was based on functionality and drafts.

3. Results:

The results show that the LMI households spend nearly twice as much on energy bill as non LMI households. The median county level LMI energy burden is 9% whereas the median non LMI is 4% of income spent on energy (Figure 5). Additionally, the map generated in ArcGIS show that energy burden may be concentrated along the border region (Figure 6).

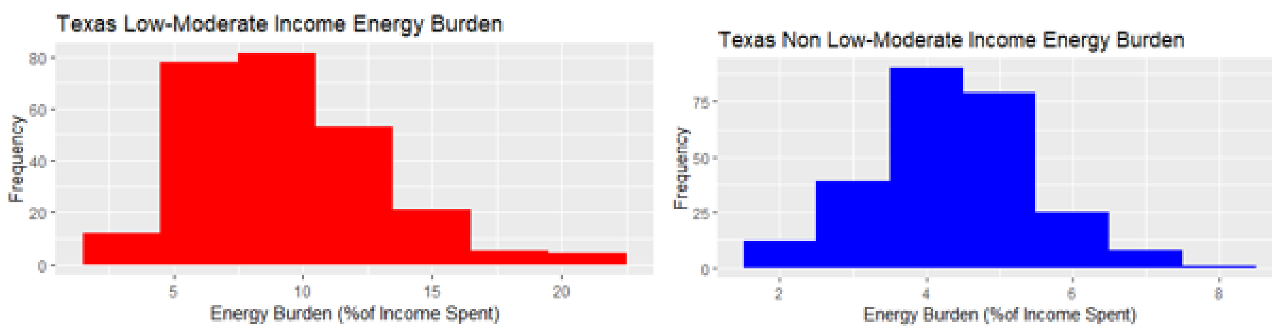


Figure 5 LMI and Non LMI Energy Burden

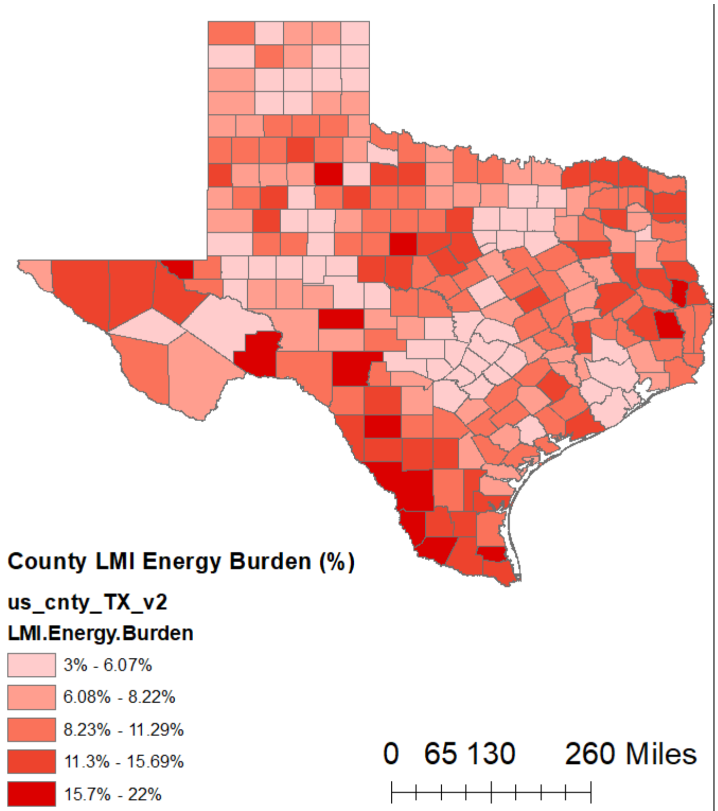


Figure 6 LMI County Level Energy Burden Map

For the all of the regression models the percent Hispanic variable was dropped because it created high multicollinearity with the other variables. Linear regression using the analysis data (without the dummy variable) has an Adjusted R^2=0.454. This Adjusted R^2 means that the model explained ~48% of the variation in energy burden when adjusting for unnecessary variables. The overall analysis model with a dummy had a p value<<0.05 meaning that the model is significant. Percent of county living in poverty and percent of households that own their home, and the food environment ranking were the significant indicators (TABLE 2).

Table 2

Variable	Estimate	Std. Error	t	Pr(>|t|)	Significance
(Intercept)	-10.19	3.44	-2.96	0.00	**
pcnt_pov_usda	0.38	0.05	8.23	0.00	***

pcnt_own_ucb	0.07	0.02	2.73	0.01	**
Summary Statistics					
Multiple R-squared: 0.48	Adjusted R-squared: 0.454	p-value: < 2.2e-16			
Note: ***indicates p<0.001, **indicates p<0.01, *indicates p<0.05,					

Linear regression using the validation data (without the dummy variable) has an Adjusted R^2=0.466. This Adjusted R^2 means that the model explained ~48% of the variation in energy burden when adjusting for unnecessary variables. The overall validation model without a dummy had a p value<<0.05 meaning that the model is significant. Percent of county living in poverty, percent of households that own their home, percent were the only significant indicators (TABLE 3).

Table 3

Variable	Estimate	Std. Error	t	Pr(>|t|)	Significance
pcnt_pov_saipe	0.3874	0.0460	8.4160	0.0000	***
pcnt_own_chr	0.0755	0.0239	3.1520	0.0018	**
pcnt_lw_access_chr	-0.0790	0.0374	-2.1150	0.0354	*
food_rank_chr	-1.2200	0.3797	-3.2130	0.0015	**
Summary Statistics					
Multiple R-squared: 0.4913	Adjusted R-squared: 0.4658	p-value: < 2.2e-16			
Note: ***indicates p<0.001, **indicates p<0.01, *indicates p<0.05,					

Linear regression using the analysis data including the border county has an Adjusted R^2=0.475. This Adjusted R^2 means that the model explained ~47.5% of the variation in energy burden when adjusting for unnecessary variables. Percent of county living in poverty, percent of households that own their home, and the food insecurity, and whether the county is a border county were the significant indicators (TABLE 4). The overall analysis model with a dummy had a p value<<0.05 meaning that the model is significant.

Table 4

Variable	Estimate	Std. Error	t	Pr(>|t|)	Significance
(Intercept)	-10.687	3.379	-3.163	0.002	**
pcnt_pov_usda	0.330	0.048	6.914	0.000	***
pcnt_own_ucb	0.075	0.024	3.087	0.002	**
pcnt_food_insec_chr	0.248	0.076	3.252	0.001	**
brdr_cnty	2.249	0.692	3.252	0.001	**
Summary Statistics					
Multiple R-squared: 0.502	Adjusted R-squared: 0.475	p-value: < 2.2e-16			
Note: ***indicates p<0.001, **indicates p<0.01, *indicates p<0.05,					

The linear regression developed from the validation dataset including the border county dummy variable has an Adjusted R^2=0.48. This Adjusted R^2 means that the model explained ~48% of the variation in energy burden when adjusting for unnecessary variables. Percent of the county that is Black/African American, the percent of county living in poverty, percent of households that own their home, and the food environment ranking, and whether the county is a border county were the significant indicators. The overall validation model with a dummy had a p value<<0.05 meaning that the model is significant.

Table 5

Varable	Estimate	Std. Error	t	Pr(>|t|)	Significance
pcnt_blk_chr	-0.083	0.041	-2.002	0.046	*
pcnt_pov_saipe	0.345	0.048	7.209	0.000	***
pcnt_own_chr	0.081	0.024	3.409	0.001	***
pcnt_lw_access_chr	-0.141	0.043	-3.299	0.001	**
food_rank_chr	-1.759	0.419	-4.193	0.000	***
brdr_cnty	2.041	0.718	2.844	0.005	**
Summary Statistics					
Multiple R-squared: 0.5079	Adjusted R-squared: 0.481	p-value: < 2.2e-16			
Note: ***indicates p<0.001, **indicates p<0.01, *indicates p<0.05,					

4. Discussion and Conclusion:

Poverty, percent Black/African American, and Food Insecurity ……The Adjusted R^2 and significant indicators for results for the model run with the analysis dataset and the validation dataset without a dummy variable are fairly similar. However, the results for the models with dummy variables are different. In the validation dummy variable model, percent of the population that is Black/African American percent of population with low access to grocery stores appear as significant eventhough they weren't in the analysis set. At the time this report is written, the author is uncertain on why these variable are significant.

The results also show that non structural variables (socio-demographic, economic, and health) indicators explain nearly 50% of the county level variation in LMI energy burden. The results also show that border counties are a predictor of energy burden. This means that given that a household is in a border county, this household is more likely to experience energy burden. As energy burden is a metric for energy poverty, I posit that household that are living in border counties are more likely to be living in energy poverty.

The results of this study indicate that there should be a stronger policy program focus on energy affordability for counties near the border. Further studies should be done at a more granular level (i.e. census tract level) to see whether border regions continues to be a strong predictor of energy burden.

Limitations:

Replication Limitations

Overall, this study is highly replicable because the datasets are all publicly available and R and R Studio are publicly available programming software. The only software that is not publicly available is ArcGIS which was used to create the maps. Additionally, the literature that was used in the literature and to develop this study, is not publicly available. As such, validating the assertions made in the introduction and literature review will be difficult for researchers without access to an academic library.

Study Limitations:

The study was done at a county level, which does not have the same statistical power as a study done at the census tract level (i.e. here n=254 compared to census tract level where n>>1000). Additionally, the author attempted to limit variables used in this study to those that were available for more than one source for the data was found. However, county median LMI energy burden, percent of population that attended some college, percent of population that is uninsured, border county, and median housing structure age were only available from one data source. This limitation could lead to validation issues.

Works Cited

Herrero, S. (2017). Energy poverty indicators: A critical review of methods. *Indoor and Built Environment*.

Bird, S., & Hernández, D. (2010). Energy Burden and the Need for Integrated Low-Income Housing and Energy Policy. *Poverty Public Policy*.

Boardman, B. (1991). *Fuel Poverty: From Cold Homes to Affordable Warmth.* London: Belhaven Press.

Boardman, B. (2010). *Fixing Fuel Poverty: Challenges and Solutions.* London: Routledge.

Bouzarovski, S., & Simcock, N. (2017). Spatializing Energy Justice. *Energy Policy*.

DOE, D. (2016). *State of Texas Energy Sector Risk Profile.* Retrieved from https://www.energy.gov/sites/prod/files/2016/09/f33/TX_Energy%20Sector%20Risk%20Profile.pdf

Food Research and Access Center. (2019). *Why Low-Income People and Food Insecure People are Vulnerable to Poor Nutrition*. Retrieved from Food Research and Access Center: Obesity and Health: http://frac.org/obesity-health/low-income-food-insecure-people-vulnerable-poor-nutrition-obesity

Harrison, C., & Popke, J. (2011). “Because You Got to Have Heat”: The Networked Assemblage of Energy Poverty in Eastern North Carolina. *Annals of the Association of American Geographers*.

James, G., Witten, D., Tibshirani, R., & Hastie, T. (2017). *Introduction to Statistical Learning.* Springer.

Kaiser Family Foundation. (2018, December). *Key Facts About the Unninsured Population*. Retrieved from https://www.kff.org/uninsured/fact-sheet/key-facts-about-the-uninsured-population/

MALEWITZ, J. (2016, Saptember 6). *Texas Stops Helping Poor Families Pay Their Electric Bills*. Retrieved from The Texas Tribune: https://www.texastribune.org/2016/09/02/texas-stops-helping-poor-families-pay-their-electr/

National Renewable Energy Labe. (2018). *Energy Affordability and Access in Focus: Metrics and Tools of Relative Energy Vulnerability*. Retrieved from https://www.arcgis.com/apps/MapJournal/index.html?appid=7cbbe0063c1f49b0b167f604f8f40576

Olmedo, C. (2013). *Energy, Housing, Income: Constraints and Opportunities for Affordable Energy Soutions in Texas Colonias.* State Energy Conservation Office.

Public Utility Commission of Texas. (2017). *Assistance Paying Your Bill*. Retrieved from https://www.puc.texas.gov/consumer/lowincome/assistance.aspx

Reames, T. (2016). Targeting energy justice: Exploring spatial, racial/ethnic and socioeconomic disparities in urban residential heating energy efficiency. *Energy Policy*.

Ross, L., & Drehobl, A. (2016). *Lifting the High Energy Burden in America's Largest Cities: How Energy Efficiency Can Improve Low Income and Underserved Communities.* American Council for Energy Efficient Economy.

Schuessler, R. (2014). *Energy Poverty Indicators: Conceptual Review: Part 1 The Ten Percent Rule.* Centre for European Economic Research.

Texas Department of State Health Services. (2017, September). *Map of DSHS Border Area*. Retrieved from https://www.dshs.texas.gov/borderhealth/border_health_map.shtm

Thomson, H., Bouzarovski, S., & Snell, C. (2017). Rethinking the measurement of energy poverty in Europe: A critical analysis of indicators and data. *Indoor and Built Environment*.

Tuttle, C., & Beatty, T. (2017). *The Effects of Energy Price Shocks on Household Food Security in Low-Income Households.* USDA.

Valenzuela, C., Valencia, A., White, S., Jordan, J., Cano, S., Keating, J., . . . Potter, L. (2014). An analysis of monthly household energy consumption among single-family residences in Texas, 2010. *Energy Policy*.

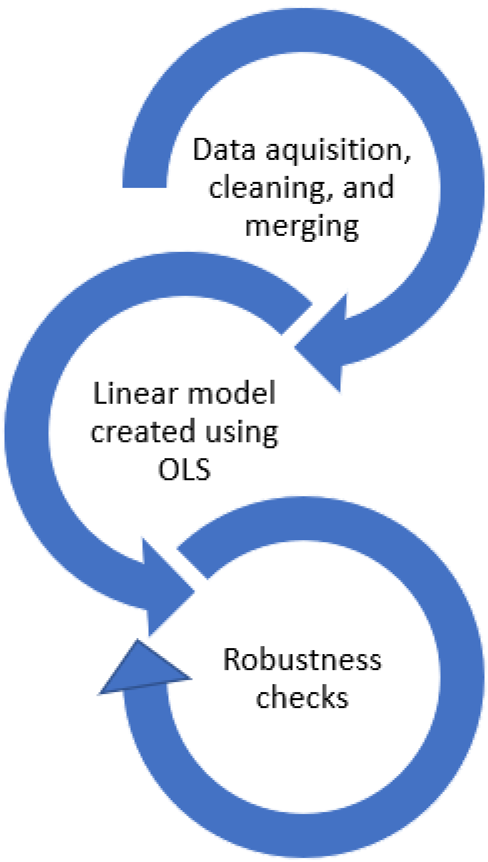
Wible, J., & King, C. (2016). *Household Energy Costs for Texans.* Austin, Texas: University of Texas Energy Institute.

Yun, G., & Steemers, K. (2009). Household Energy Consumption and the Role of Occupants. *Building Research Information*.

APPENDIX

GitHub Repository Link: https://github.com/cnagbim/Data-Management-Final-Project-Agbim

General Workflow



Variable Names

Variable	Analysis Variable	Validation Variable	DataType

En_burden	Lmi_burdenx100	Lmi_burdenx100	Continuous, numerical
fip	fip	fip	Nominal, integer
cnty	cnty	cnty	Nominal, integer
pcnt_hisp	pcnt_hisp_ucb	pcnt_hisp_chr	Continuous, numerical
pcnt_blk	pcnt_blk_ucb	pcnt_blk_chr	Continuous, numerical
pcnt_rural	Pcnt_rural_chr	Pcnt_rural_chr	Continuous, numerical
pcnt_sr	pcnt_sr_ucb	pcnt_sr_chr	Continuous, numerical
pcnt_unemp	pcnt_unemp_bls	pcnt_unemp_chr	Continuous, numerical
pcnt_pov	Pcnt_pov_usda	Pcnt_pov_saipe	Continuous, numerical
pcnt_own	pcnt_own_ucb	pcnt_own_chr	Continuous, numerical
pcnt_lw_access	pcnt_lw_access_usda	pcnt_lw_access_chr	Continuous, numerical
pcnt_obese	Pcnt_obese_usda	Pcnt_obese_chr	Continuous, numerical
food_insec or food_rank	food_insec_chr	Food_rank_chr	Continuous, numerical
pcnt_uninsured	pcnt_uninsured_chr	pcnt_uninsured_chr	Continuous, numerical
Brdr_cnty	Brdr_cnty	Brdr_cnty	Nominal
House_str_age	House_str_age	House_str_age	Continuous, numerical

Source	Acronym
US Census Bureau	ucb
Bureau of Labor Statistics	bls
County Health Ranking	chr
US Department of Agriculture	usda
US Census Small Area Income	saipe