Weather Change and
Energy Consumption Relationships

Final Report

- Group 42

Nicholas Wells, Kyle Diep, Vivek Prajapati, Pramod Philip, Vikas Soni

Table of Contents

Choice of Topic, Business Justification, and Problem Statement

With the documented problems of global warming, there are a near infinite number of implications that our society has been, and will continue to deal with, in the coming years. With rising temperatures leading to calamities (e.g., uncontrolled bush fires and electricity blackouts lasting days) in first world countries such as the U.S. and Australia, something that many would have previously considered unheard of, it begs the question of why these developed areas are experiencing such problems in the first place.

The 2021 revenue for the U.S. electricity industry was $424B. Based on historical consumption data, is the level of energy usage associated with that revenue sustainable moving forward, given our existing supply of energy? We are already aware that weather changes are affecting the availability of power generation fuels such as natural gas, which contributed to the greatest single-year increase in electricity prices since 2008.

This shows that weather changes can majorly impact such a huge industry. If we better understand the impact of weather changes on energy consumption, this can help in improving electricity supply logistics, company marketing tactics, and infrastructure planning.

We believe that weather changes are affecting the trends in energy consumption and will verify this using data sources representing multiple locations. Based on the discovery process, we will investigate the related business implications and concerns that can be addressed.

Our hypothesis is that locations with relatively higher than average temperatures also have higher electrical energy consumption and electricity prices compared to locations with relatively lower average temperatures.

If our hypothesis holds true, companies that are focused on heating, ventilation, and air conditioning (HVAC) and electric vehicles (EVs) would benefit by prioritizing their marketing efforts towards areas with higher-than-average temperatures since electricity consumption is higher in those locations. For example, EV companies would need to tailor their marketing campaigns in terms of global warming implications instead of the standard gas vs. electric argument. For EVs, the priority of reducing our reliance on fossil fuels and transitioning towards renewable energy sources would become their main focus towards attracting potential customers.

Relatedly, areas with higher electricity demand would need to have assessments as to the long-term viability of their existing electrical grid system and whether they can support current and future increases without major disruptions.

Understanding of the data and data wrangling

We initially planned to model the data for each team member's home city but were unable to find granular enough data for energy usage. State level information was widely available though. We also found that the weather data is tracked at a daily rate, while energy and GDP were tracked yearly. To standardize the data, we agreed that it would be modelled at the monthly level. We averaged the daily data up to monthly and divided the annual data for each year by 12. After that, we resampled the new data into months and then used linear interpolation to fill in any NA values for the months that didn't initially have values. We also struggled with finding datasets of matching time frames. Due to the varying ranges of time frames from the data sources, we picked the shortest time frame to avoid having too much missing data.

Our temperature data is coming from the University of Dayton's website for everyone in the USA and Kaggle for Australia. We are getting the power utilization data from the EIA.gov website for the USA. Energy.gov is supplying the EV data. Opennem was our source for AUS energy data.

On the next page are two examples of our cleaned data, one from Illinois to represent our domestic region and one for our Australian region; the date range for the Illinois data is from 2001 to 2019 while for Australia it is 1999 to 2019. The screenshots below are a combination of all independent variables we used depending on availability across our two regions.

For example, our four U.S. states (Georgia, Illinois, Missouri, and Texas) and Australia had available data for: average retail electricity price, net electricity generation for renewable and non-renewable resources, average monthly temperature, gross domestic product (GDP), and electrical energy consumption.

For our data, we wanted to make sure that all variables dealing with electricity specifically were scaled from their existing units (e.g. electrical energy consumption was in kilowatt-hr) to gigawatt-hr. This ensured that all of our units of measurement across electrical energy data was being read, interpreted and analyzed in the same fashion. If we had not done this, we ran the risk of our output analyses having unnecessarily large variances due to uneven scaling across variables that were all measuring electricity (even if their respective data origins had different scaling to begin with).

IL data example

```
     date               non_renewables_gwh renewables_gwh       mean_tmp
Min.   :2001-01-01   Min.   :12110       Min.   :   8.75   Min.   :16.01
1st Qu.:2005-09-23   1st Qu.:14444       1st Qu.:  24.65   1st Qu.:34.33
Median :2010-06-16   Median :15382       Median : 338.04   Median :52.30
Mean   :2010-06-16   Mean   :15515       Mean   : 471.33   Mean   :50.80
3rd Qu.:2015-03-08   3rd Qu.:16653       3rd Qu.: 872.62   3rd Qu.:68.16
Max.   :2019-12-01   Max.   :19165       Max.   :1743.88   Max.   :81.17
  all_price        consumption     gdp_monthly_in_mil
Min.   : 5.870   Min.   :12986   Min.   :41796
1st Qu.: 7.452   1st Qu.:14942   1st Qu.:51243
Median : 8.745   Median :15845   Median :56545
Mean   : 8.402   Mean   :15987   Mean   :58046
3rd Qu.: 9.402   3rd Qu.:17027   3rd Qu.:66772
Max.   :10.220   Max.   :19207   Max.   :74207
```

Victoria, Australia data example

```
      date            renewables_gwh    renew_gen_gwh     non_renewables_gwh non_renew_gen_gwh total_energy_gwh total_energy_gen_gwh
avg_temp
 Min.   :1999-01-01   Min.   :  74.66   Min.   :  74.66   Min.   :11956      Min.   :3009      Min.   :12171    Min.   :4532
Min.   :35.92
 1st Qu.:2004-03-24   1st Qu.: 543.15   1st Qu.: 543.15   1st Qu.:13760      1st Qu.:4520      1st Qu.:15482    1st Qu.:5613
1st Qu.:58.44
 Median :2009-06-16   Median :1322.09   Median :1322.20   Median :14673      Median :4877      Median :16453    Median :6468
Median :65.47
 Mean   :2009-06-16   Mean   :1552.29   Mean   :1541.14   Mean   :14673      Mean   :4762      Mean   :16226    Mean   :6303
Mean   :64.96
 3rd Qu.:2014-09-08   3rd Qu.:2327.92   3rd Qu.:2289.65   3rd Qu.:15523      3rd Qu.:5174      3rd Qu.:17310    3rd Qu.:6999
3rd Qu.:71.46
 Max.   :2019-12-01   Max.   :4573.37   Max.   :4561.54   Max.   :17870      Max.   :5867      Max.   :19066    Max.   :8185
Max.   :75.71
 electricity_price  gdp_millions
 Min.   : 10.41    Min.   :24369
 1st Qu.: 18.52    1st Qu.:35346
 Median : 25.39    Median :47453
 Mean   : 30.41    Mean   :48684
 3rd Qu.: 36.80    3rd Qu.:59480
 Max.   :127.28    Max.   :79091
```

Literary Review

Per the ijert.org link, they used linear regression, SVM, KNN, and random forest modeling techniques to predict daily electrical consumption behavior. We will investigate whether we can produce similar results to judge our analysis as well as whether we can use the other models that they did with our data.

On the iea.org website, they have in-depth interactive charts comparing the energy consumption versus production of countries around the world and whether they were producing enough and how that is contributing to global pollution levels. This demonstrates a supporting citation for the business need of our analysis.

The University of Malaya has done multiple studies in this area using many types of models. They reference several of the problems they encountered like multicollinearity and using PCA to address the independent variables and we will keep them in mind as they may be applicable to our models.

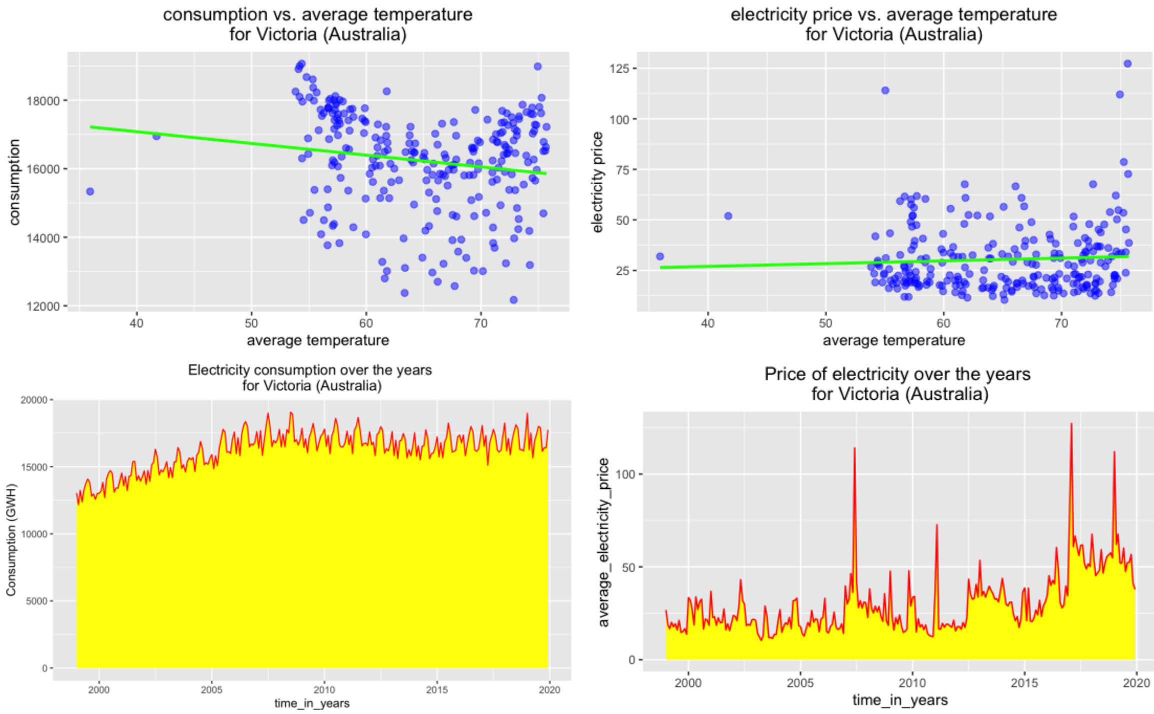
Approach/Methodology

With team members living in multiple hemispheres and biomes, our plan was to model each person's home state, then evaluate our analysis to compare and contrast the anticipated weather pattern changes in each area. The initial modelling approach we used was having electricity consumption and electricity price within each state as our two response variables, while our independent variables were as follows:

- Average monthly temperature
- GDP by month
- Price of electricity
- Consumption of electricity (in gigawatt hours)
- Total electricity generated (in gigawatt hours)
- Electricity generated from renewable resources (in gigawatt hours)
- Electricity generated from non-renewable resources (in gigawatt hours)

Georgia (consumption on left, price on right):

Australia (consumption on left, price on right):



Variable Selection Methods

As we learned in earlier course lectures, we can do a factor-by-factor analysis to ascertain the statistical relevance, or lack thereof, of independent variables with respect to the response variable. We decided to use multiple variable selection methods to establish the significant variables for each location. Generally speaking, we used Akaike Information Criterion (AIC) to determine individual variable significance and adjusted r-squared to identify statistically significant models.

Forward Selection

We first used Forward Selection where our models had no independent variables to start with. This method systematically adds new variables, assessing along the way whether the variable being added lowered the resulting AIC at a statistically significant amount. This process repeats until the Forward Selection of an additional variable did not produce a significant reduction in AIC, thus stopping the addition of new variables to our model. We also ran a general summary on our data to assess the adjusted r-squared values to use as a point of comparison with other models.

Backward Elimination

Our next variable selection method, Backward Elimination, began with our model starting with all independent variables. Similar to Forward Selection, Backward Elimination systematically identifies relevant variables and eliminates irrelevant variables on the basis of AIC.

Stepwise Regression

Stepwise Regression combines the Forward Selection and Backward Elimination methods for variable selection. It starts similarly to Forward Selection where our model has no variables, then progressively adds variables one by one. After adding each factor, the model checks to see whether eliminating an existing factor would significantly lower the AIC value as with Backward Elimination. Our model would stop selecting variables when neither adding a new factor nor eliminating an existing one significantly minimizes the AIC any further.

LASSO

Prior to using the LASSO and Elastic Net variable selection methods, we had to separate the predictor variables and response variable by putting them into different matrices. The construction of these two models are contingent on our parameter, alpha, which controls how much L1 and L2 regularization are used.

$$L_{enet}(\hat{\beta}) = \frac{\sum_{i-1}^{n}(y_i - x_i'\hat{\beta})^2}{2n} + \lambda(\frac{1-\alpha}{2}\sum_{j-1}^{m}\hat{\beta}_j^2 + \alpha\sum_{j-1}^{m}|\hat{\beta}_j|),$$

For LASSO regression (where alpha = 1), we plotted the mean squared errors (MSE) against the log lambda to find the optimal lambda value. Based on the optimal lambda value, we calculated the coefficients of the LASSO regression model and created our linear regression model based on the aforementioned significant coefficients/variables.

Elastic Net

For Elastic Net regression, we varied the alpha parameter within the range 0 to 1, exclusive, to avoid alpha = 0 (Ridge regression) or alpha = 1 (LASSO from above). For each alpha, we created an EN model that utilized cross-validation to determine the lambda with the lowest MSE. We then used this lambda value to determine its deviance ratio. We plotted our alphas against the deviance ratio across all of our EN models. The alpha with the largest deviance ratio was chosen as our best alpha value. We constructed our final EN model based on this optimal alpha value and extracted the model’s significant coefficients. These resulting coefficients were then implemented in our final linear regression model utilizing this EN method.

ARIMA

We used AutoRegressive Integrated Moving Average (ARIMA) to forecast the price and consumption of electricity for all 5 of our states. We initially wanted to forecast price, sales, and consumption; however, limitations on availability of data led us to settle on forecasting just price and consumption.

To start, we created a time series object using our dependent variables starting from the year 2001. We plotted our data to get a better idea of any statistical anomalies with respect to the usual parameters. This plot shows any upward trends or seasonality component in our dependent variables. To remove these upward trends, we utilized first-order differencing. We also saw that our series was not stationary with respect to variance; specifically, as the years went on, we saw that the variance was increasing. To get reliable ARIMA forecasting, we needed to make our series stationary for the mean and variance. One of the best ways to make a series stationary is through log transformation, which we applied to our data. While log transforming made our data stationary for the variance, it did not do so for the mean because we are using original data.

After applying log transformation on our differenced original data, our series became stationary for both mean and variance. Since we are using first-order differencing, we expected the integrated part of our ARIMA model to be 1. Our next step involved using autocorrelation factor (ACF) and partial autocorrelation factor (PACF) plots to identify patterns within our transformed data that is now stationary for both mean and variance. In this process, we found out the seasonality component of our data.

Next, we identified the best fit ARIMA model using R's auto arima function. This function uses the minimum AIC and Bayesian Information Criterion (BIC) values to select the best model. With this model, we predicted the dependent variables for the following 12 months, with the assumption that the underlying patterns in the time series would remain the same as predicted. Our final step involved using ACF and PACF plots again to confirm no remaining information was left for extraction in our residuals.

CUSUM

Since our data is of the time series variety, we are also using the CUSUM change detection method to determine whether temperature patterns have changed over time. The CUSUM method is looking to see how our data summary statistics deviate above or below a target value, relative to the standard errors of our summary statistics. Our CUSUM model's parameters were a T-value, which specified the number of standard errors of the summary statistics at which our cumulative sum is out of control, and our C-value, which reflected the amount of shift to detect in the process, measured in the standard errors of our summary statistics. We plan to test 2 CUSUM models with the following parameter values: T-value = 5 and C-value = 1 for the first model and T-value = 5 and C-value = 2.

Results

Forward Selection

For our variable selection method results, we first began using Forward Selection with electricity consumption as our response variable. Our adjusted r-squared values were 0.838 and 0.870 for Illinois and Georgia, respectively. For Illinois, the statistically significant variables were: monthly GDP, electricity price, electricity generated from renewables and electricity generated from non-renewables. For Georgia, the relevant variables were average temperature, electricity generated by non-renewables, and electricity generated by renewables. When using electricity price as our response variable, our adjusted r-squared values were 0.704 and 0.813 for Illinois and Georgia, respectively. For Illinois, the statistically significant variables were: monthly GDP, average temperature, electricity consumption, electricity generated from renewables and electricity generated from non-renewables. For Georgia, the relevant variables were: monthly GDP, average temperature, electricity consumption, and electricity generated from renewables.

For Texas (adjusted r-squared 0.923), the statistically significant variables when consumption was the response variable were: monthly GDP, electricity generated from non-renewables, and electricity generated from renewables. When electricity price was the response variable with an adjusted r-squared of 0.731, the relevant variables were: electricity consumption, monthly GDP, and electricity generated from renewables. For Missouri (adjusted r-squared 0.787) with consumption as response, the relevant variables were: electricity price, monthly GDP, and electricity generated from non-renewables. When electricity price was the response variable with an adjusted r-squared of 0.884 the statistically significant variables were: average temperature, monthly GDP, and electricity consumption.

For Australia (adjusted r-squared 0.766), the statistically significant variables when electricity consumption was the response variable were: average temperature, electricity price, monthly GDP, electricity generated from non-renewables, and electricity generated from renewables. The adjusted r-squared when electricity price was the response variable was 0.407, and the statistically significant variables were: electricity consumption, monthly GDP, electricity generated from non-renewables, and electricity generated from renewables.

Backward Elimination

With Backward Elimination, when electricity consumption was the response variable for Illinois (adjusted r-squared 0.838), the statistically significant variables were: average temperature, monthly GDP, electricity price, electricity generated from renewables and electricity generated from non-renewables. For the price response variable in Illinois(adjusted r-squared of 0.7096), the significant variables were temperature, monthly GDP, price, renewables, and non-renewables. For Georgia consumption (adjusted r-squared 0.873), the statistically significant variables were: average temperature, monthly GDP, electricity price, electricity generated by non-renewables, and electricity generated by renewables. For Georgia (adjusted r-squared

0.938), the relevant variables when price was the response variable were: electricity consumption, electricity generated by renewables, average temperature, and monthly GDP.

For Texas (adjusted r-squared 0.923), the statistically significant variables when consumption was the response variable were: monthly GDP, electricity generated from non-renewables, and electricity generated from renewables. When electricity price was the response variable with an adjusted r-squared of 0.731, the relevant variables were: electricity consumption, monthly GDP, and electricity generated from renewables. For Missouri (adjusted r-squared 0.787) with consumption as response, the relevant variables were: electricity price, monthly GDP, and electricity generated from non-renewables. When electricity price was the response variable with an adjusted r-squared of 0.884, the relevant variables were: average temperature, monthly GDP, and electricity consumption.

For Australia (adjusted r-squared 0.766), the statistically significant variables when electricity consumption was the response variable were: average temperature, electricity price, monthly GDP, electricity generated from non-renewables, and electricity generated from renewables. The adjusted r-squared when electricity price was the response variable was 0.407, and the statistically significant variables were: electricity consumption, monthly GDP, electricity generated from non-renewables, and electricity generated from renewables.

Stepwise Regression

For Stepwise Regression, when electricity consumption was the response variable for Illinois (adjusted r-squared 0.838), the statistically significant variables were: average temperature, monthly GDP, electricity price, electricity generated from renewables and electricity generated from non-renewables. For Georgia (adjusted r-squared 0.873), the statistically significant variables were: average temperature, monthly GDP, electricity price, electricity generated by non-renewables, and electricity generated by renewables. For Illinois (adjusted r-squared 0.767), the statistically significant variables with electricity price as the response variable were: electricity generated by non-renewables, average temperature, and monthly GDP. For Georgia (adjusted r-squared 0.938), the relevant variables when price was the response variable were: electricity consumption, electricity generated by renewables, average temperature, and monthly GDP.

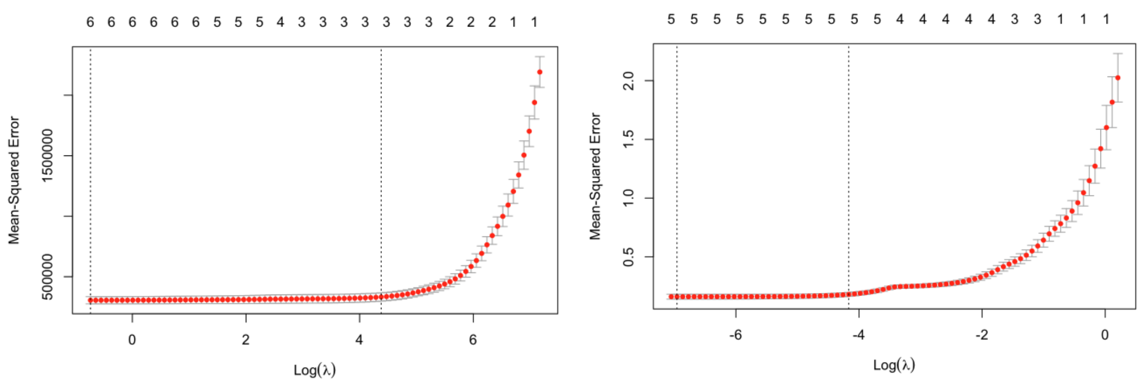
For Texas (adjusted r-squared 0.911), the statistically significant variables when consumption was the response variable were: electricity generated from non-renewables and electricity generated from renewables. When electricity price was the response variable with an adjusted r-squared of 0.731, the relevant variables were: electricity consumption, monthly GDP, and electricity generated from renewables. For Missouri (adjusted r-squared 0.787) with consumption as response, the relevant variables were: electricity price, monthly GDP, and electricity generated from non-renewables. When electricity price was the response variable with an adjusted r-squared of 0.872: average temperature and monthly GDP.

For Australia (adjusted r-squared 0.766), the statistically significant variables when electricity consumption was the response variable were: average temperature, electricity price, monthly GDP, electricity generated from non-renewables, and electricity generated from renewables. The adjusted r-squared when electricity price was the response variable was 0.428, and the statistically significant variables were: electricity consumption, monthly GDP, and electricity generated from non-renewables.

LASSO

For LASSO, when electricity consumption was the response variable for Illinois (adjusted r-squared 0.838), the statistically significant variables were: average temperature, electricity price, monthly GDP, electricity generated by non-renewables and electricity generated by renewables. For Georgia (adjusted r-squared 0.873), the statistically significant variables were: average temperature, monthly GDP, electricity price, electricity generated by non-renewables, and electricity generated by renewables. For Illinois (adjusted r-squared 0.767), the statistically significant variables with electricity price as the response variable were: electricity generated by non-renewables, electricity generated by renewables, average temperature, and monthly GDP. For Georgia (adjusted r-squared 0.937), the relevant variables when price was the response variable were: electricity consumption, electricity generated by non-renewables, electricity generated by renewables, average temperature, and monthly GDP.

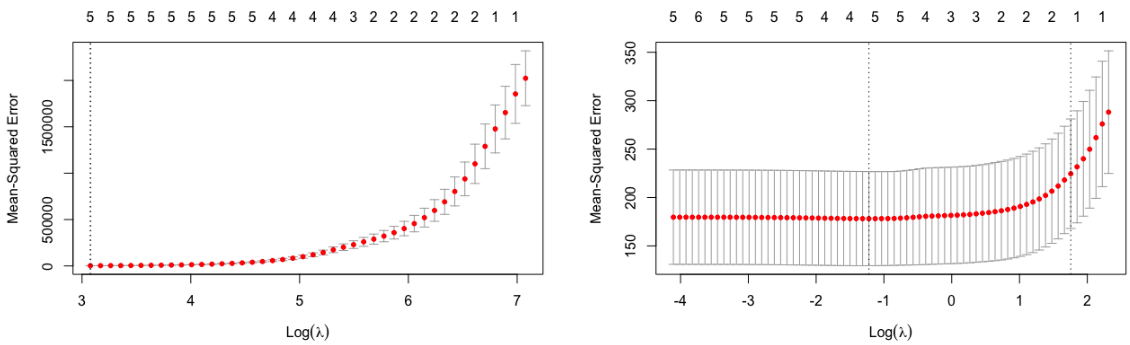
Georgia (consumption on left, price on right):



For Texas (adjusted r-squared 0.923), the statistically significant variables when consumption was the response variable were: monthly GDP, electricity generated from non-renewables and electricity generated from renewables. When electricity price was the response variable with an adjusted r-squared of 0.326, the relevant variables were: average temperature, monthly GDP, and electricity generated from renewables. For Missouri (adjusted r-squared 0.787) with consumption as response, the relevant variables were: electricity price, monthly GDP, and electricity generated from non-renewables. When electricity price was the response variable with an adjusted r-squared of 0.884: average temperature, monthly GDP, and electricity consumption.

For Australia (adjusted r-squared 0.743), the statistically significant variables when electricity consumption was the response variable were: total electricity generated, monthly GDP and electricity generated from renewables. The adjusted r-squared when electricity price was the response variable was 0.409, and the statistically significant variables were: electricity consumption, monthly GDP, electricity generated from non-renewables.

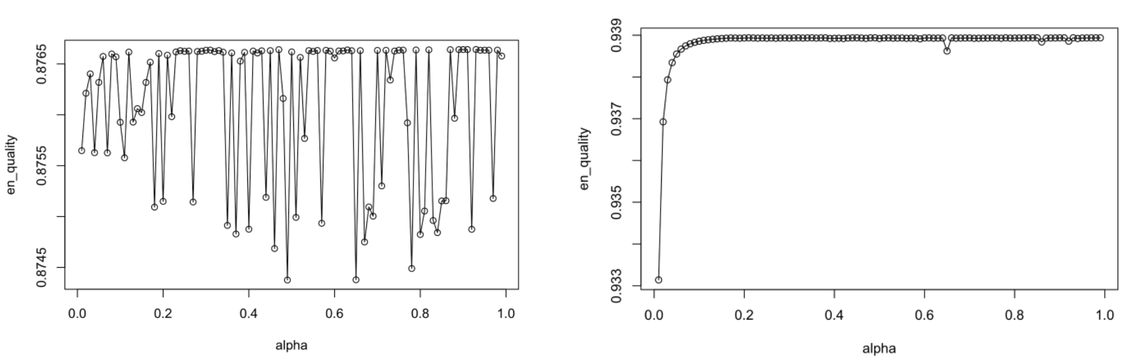
Australia (consumption on left, price on right):



Elastic Net

For EN, when electricity consumption was the response variable for Illinois (adjusted r-squared 0.838), the statistically significant variables were: average temperature, electricity price, monthly GDP, electricity generated by non-renewables and electricity generated by renewables. For Georgia (adjusted r-squared 0.873), the statistically significant variables were: average temperature, monthly GDP, electricity price, electricity generated by non-renewables, and electricity generated by renewables. For Illinois (adjusted r-squared 0.767), the statistically significant variables with electricity price as the response variable were: electricity generated by non-renewables, electricity generated by renewables, average temperature, and monthly GDP. For Georgia (adjusted r-squared 0.937), the relevant variables when price was the response variable were: electricity generated by non-renewables, electricity generated by renewables, average temperature, electricity consumption, and monthly GDP.

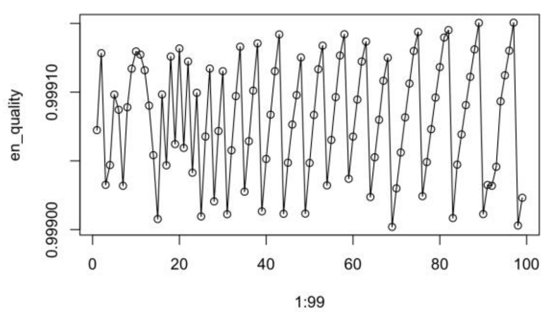
Georgia (consumption on left, price on right):

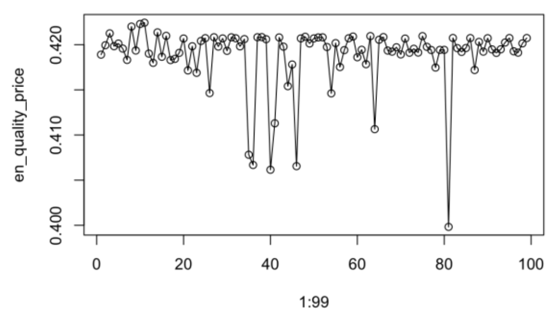


For Texas (adjusted r-squared 0.923), the statistically significant variables when consumption was the response variable were: monthly GDP, electricity generated from non-renewables and electricity generated from renewables. When electricity price was the response variable with an adjusted r-squared of 0.326, the relevant variables were: average temperature, monthly GDP, and electricity generated from renewables. For Missouri (adjusted r-squared 0.782) with consumption as response, the relevant variables were: monthly GDP and electricity generated from non-renewables. When electricity price was the response variable with an adjusted r-squared of 0.872: average temperature and monthly GDP.

For Australia (adjusted r-squared 0.743), the statistically significant variables when electricity consumption was the response variable were: total electricity generated, monthly GDP and electricity generated from renewables. The adjusted r-squared when electricity price was the response variable was 0.387, and the statistically significant variables were: monthly GDP and electricity generated from non-renewables.

Australia (consumption on left, price on right):

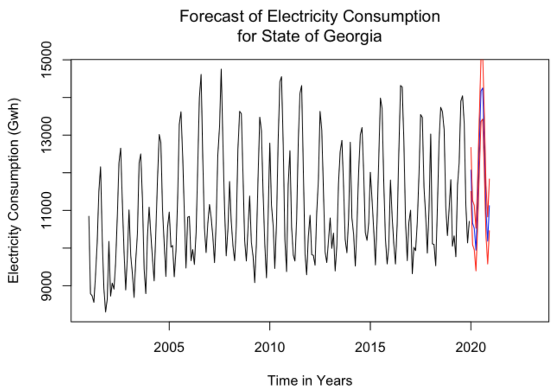


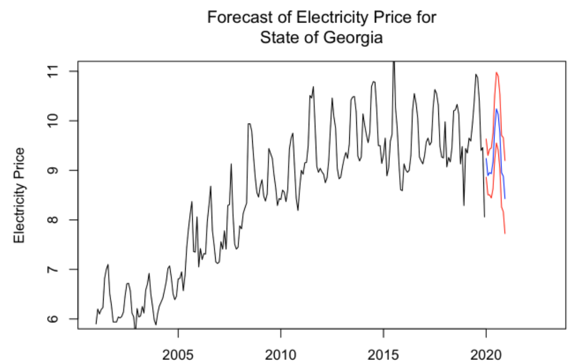


ARIMA

For ARIMA, the models are chosen by the minimum AIC and BIC values. With consumption as the response variable, the final model for Texas was (0,0,1)(0,1,1)[12]. This model has a seasonal integrated component equal to 1 (representing differencing of order-1) and MA value of order-1 with lag of 12. Similarly, with price as the response variable, the final model was (2,1,2)(0,1,1)[12]. For Missouri with consumption as the response variable, the final model was (3,0,0)(0,1,1)[12]. Similarly, with price as the response variable, the final model was (1,1,2)(0,1,1)[12].

Georgia (consumption on left, price on right):

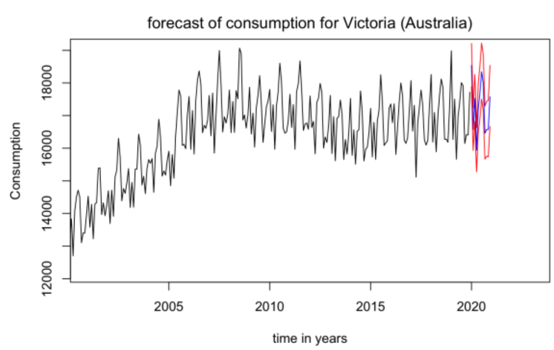


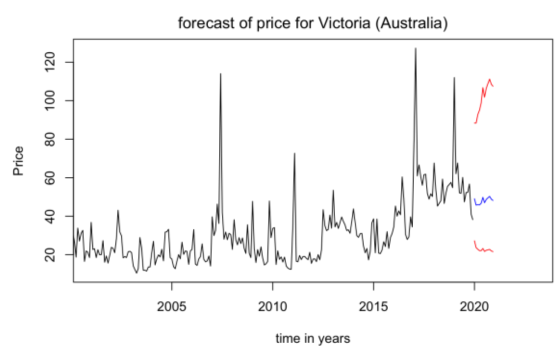


With consumption as the response variable, the final model for Illinois was (1,1,1)(2,1,0)[12]. This model has a seasonal integrated component equal to 1 (representing differencing of order-1) and MA value of order-1 with lag of 12. Similarly, with price as the response variable, the final model was (1,0,1)(0,1,1)[12]. For Georgia with consumption as the response variable, the final model was (1,0,2)(0,1,1)[12]. Similarly, with price as the response variable, the final model was (2,1,0)(0,1,1)[12].

For Australia with consumption as the response variable, the final model was (0,1,2)(0,1,1)[12]. This model has a seasonal integrated component equal to 1 (representing differencing of order-1) and MA value of order-1 with lag of 12. Similarly when price was the response, the final model was (1,1,2)(2,0,0)[12].

Australia (consumption on left, price on right):

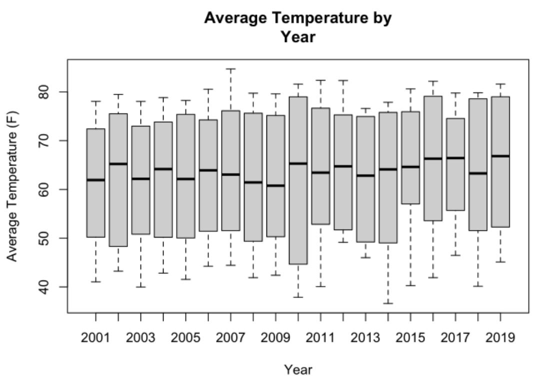


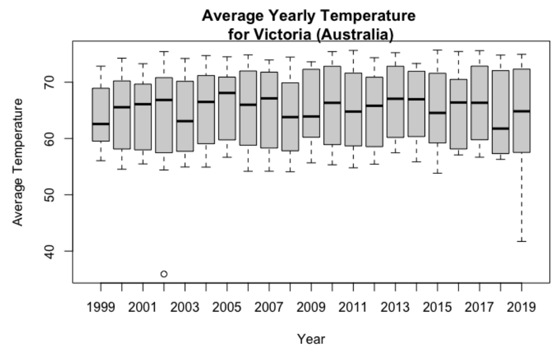


CUSUM

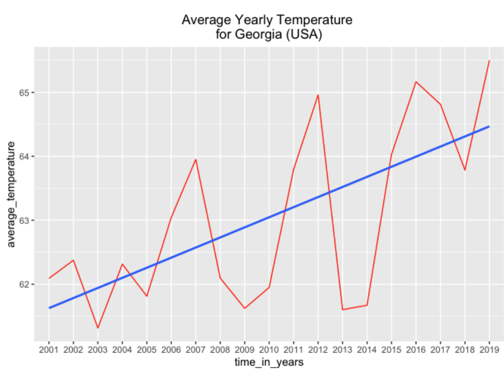
On visual observation of the box plots for all five geographies it was not clear that the mean is consistently going up or down, which would have indicated temperature warming or cooling trends in those respective regions.

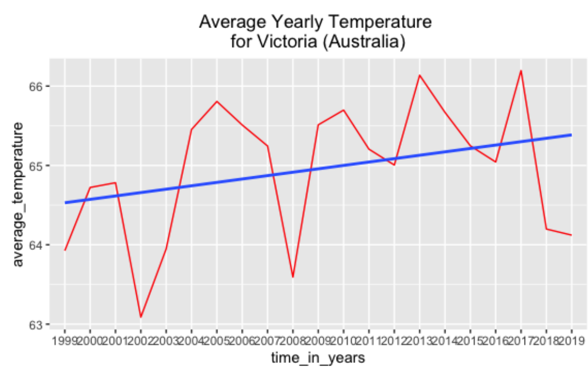
Average Temperature by Year for Georgia (left) and Australia (right):





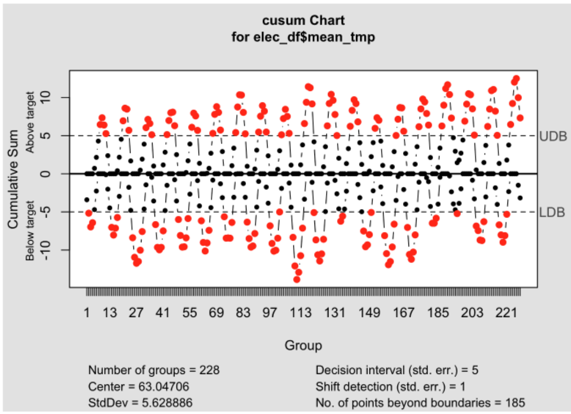
This trend line clearly indicated that the average temperature was trending upwards.

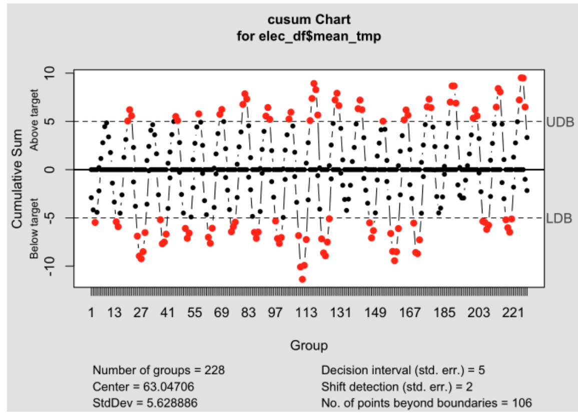




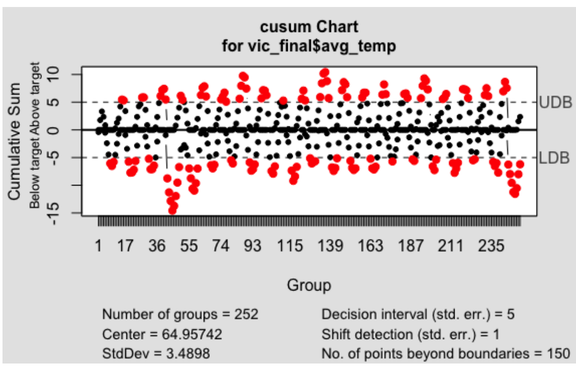
For CUSUM, our two models had the following parameter values: T-value = 5 and C-value = 1 for the first model and T-value = 5 and C-value = 2 for the second model. For both models we ran, it was not clear that the number of observations were consistently exceeding the higher or lower decision boundaries as we progressed throughout the years. This meant we could not definitively conclude that we were seeing higher variations in temperature in more recent years compared to our data's earlier years.

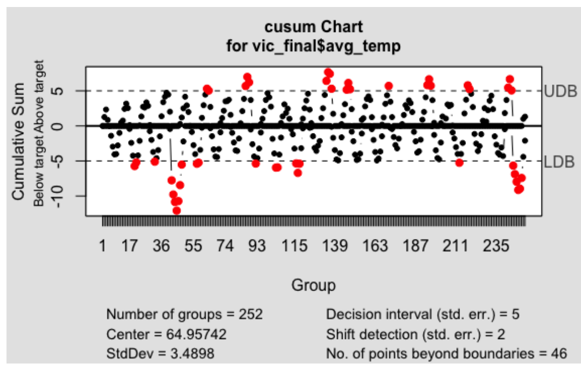
Georgia:





Australia:



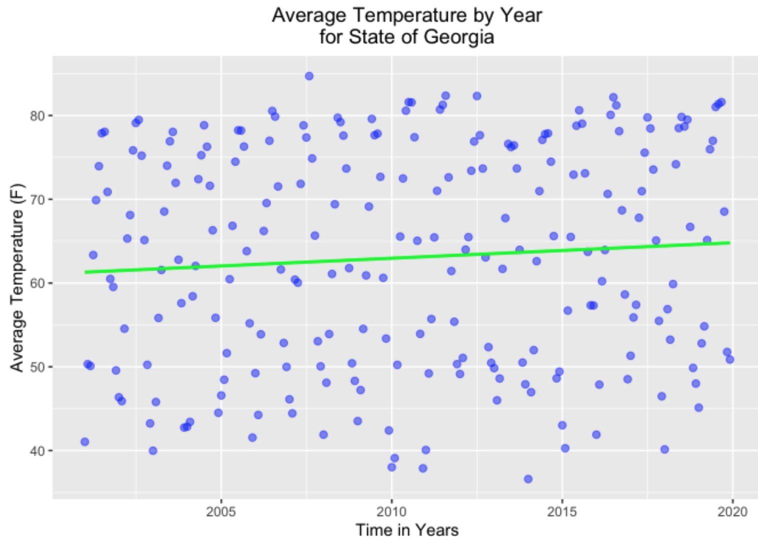


One caveat to our CUSUM models' results is that our data's time frames ranged between 8-15 years rather than across multiple decades / centuries. Our results may have been different if we had more time series data with a larger sample size. These results were consistent across all five of our selected states.

Conclusion

After assessing the results of our models, our initial hypothesis seems to be incorrect when we thought that locations with relatively higher average temperatures would also have higher electrical energy consumption and electricity prices compared to locations with relatively lower average temperatures.

Using a linear regression model, with average temperature as the response variable and time as the independent variable, we can see that while there is a slight positive correlation below for Georgia, overall there is no discernible trend.



From our variable selection methods, we ultimately see that average temperature was not a statistically significant variable for electricity consumption. However, when electricity price was the response variable, average temperature was in fact a statistically significant variable, except for Australia which follows as their average temperature range is much smaller than the USA states used.

From our CUSUM method, we found that there was no conclusive evidence that average temperature was changing over our specific timeframe.

The overall business implications of our results can be broken down into micro and macro levels. For example, the design of consumer goods that are reliant on electricity would have to be able to withstand more extreme temperatures on both ends of the spectrum. Relatedly, electrical appliances that historically have used relatively large amounts of electricity, e.g. refrigerators and HVAC systems, would have to be more efficient in terms of their electricity consumption for long-term sustainability across the globe.

Even though the results of our specific project were not conclusive, our suggestion would be that this level of analysis should still be pursued by private and public entities. Given our limited sample size of data and relatively small timeframe analyzed, more definitive conclusions may be found with a larger time frame.

References:

- https://www.eia.gov/todayinenergy/detail.php?id=51438#:~:text=In%20our%20latest%20Short%2DTerm,%25%20to%2014.26%20cents%2FkWh.
- https://www.statista.com/statistics/190548/revenue-of-the-us-electric-power-industry-since-1970/
- https://www.ijert.org/weather-based-electricity-consumption-prediction-using-big-data-analytics-model
- https://azpdf.org/article/electricity-demand-literature-review-university-malaya.q7wl7m4r
- https://www.iea.org/data-and-statistics/data-tools/weather-for-energy-tracker
- http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arima-modeling-in-r-manufacturing-case-study-example/
- Temperature:
  - https://academic.udayton.edu/kissock/http/Weather/citylistUS.htm
- Texas:
  - https://www.eia.gov/beta/states/states/tx/data/dashboard/electricity
  - https://www.eia.gov/beta/states/states/tx/data/dashboard/renewables
- Georgia:
  - https://www.eia.gov/beta/states/states/ga/data/dashboard/electricity
  - https://www.eia.gov/beta/states/states/ga/data/dashboard/renewables
- Illinois:
  - https://www.eia.gov/beta/states/states/il/data/dashboard/electricity
  - https://www.eia.gov/beta/states/states/il/data/dashboard/renewables
- Missouri:
  - https://www.eia.gov/beta/states/states/mo/data/dashboard/electricity
  - https://www.eia.gov/beta/states/states/mo/data/dashboard/renewables
- Australia:
  - https://opennem.org.au/energy/vic1/?range=7d&interval=30m
  - https://www.abs.gov.au/statistics/economy/national-accounts/australian-national-accounts-national-income-expenditure-and-product/latest-release#data-download
  - https://www.kaggle.com/datasets/mohanpal/victoriaaustralia-electricity-data-2018-2020