

## Project Title: **Energy Costs, Temperatures, and Consumption**

### Team Members:

Frances Jay  
Hugh Kelly  
John Kiersznowski  
Carol Love

### Description/Outline:

We want to analyze electricity cost vs temperature and consumption.

### Research Questions to Answer:

- How do energy costs compare across states?
- How have energy costs changed over time (10 years)?
- Maybe include temperature data for comparison...
- Include electricity consumption
- Focus on retail cost of electricity.

### Datasets to be Used:

Energy Information Administration API:

<https://www.eia.gov/opendata>

An example of how to organize an API request

[APIs with Python - Hugh](#)

Open Weather API

<https://openweathermap.org/api>

### Rough Breakdown of Tasks:

- Get data from API
- Clean/merge data
- Create numerical summaries
- Create visualizations

### Next Steps:

- Over the weekend, get data from the API(s)
- Try to answer a question of interest using the data

### Possible Comparisons:

- Electricity Price + Temperature (John)
- Electricity Price + State sectors (Frances)
- Electricity Price + Time (Carol)
- ~~Electricity Price + Consumption~~
- ~~State Electricity Price over time (Pamela)~~
- Temperature + Consumption (Hugh)

Frances: repository and README file

Carol: visualizations

Hugh: talking points

John: presentation

README writeup:

As a group, we wanted to analyze the relationships among temperature, consumption, and the cost of electricity per kilowatt hour in the United States. We looked to the U.S. Energy Information Administration's (EIA) available API to gather data on the price for electricity per month in the past 10 years. Additionally, we used the EIA's API to gather data on the consumption of electricity across the States in the year 2021 (more information [here](#)). Lastly, we used the National Centers for Environmental Information's Climate Data Online to get information on the average temperatures over time in the States in the year 2021 (more information [here](#)).

## Goals

We had three goals in mind for our project: gather and clean our data, graph the relationships as stated above, and prepare an analysis for our presentation to the class. Firstly, we analyzed how the price of electricity has changed over time, and how it has changed for each sector (commercial, industrial, residential, and transportation). Secondly, we wanted to see how temperature impacted the price of electricity across the United States in the year 2021. Finally, we wanted to analyze the relationship between the consumption of electricity and its price in the year 2021. We limited the last two analyses to a single year because the amount of data this would have been far too much for the scope of our project.

## Summary

The first analysis required us to gather data from the EIA on the price of electricity over the past ten years. We requested the data using an API, and then cleaned up the resulting dataset by removing any rows that did not have a value listed in the price column. We then plotted the price of electricity over time and separated the colors out by sector. Over time, the price of electricity has increased for all sectors. There were notably larger increases in 2008 and from 2020 - 2022. Also of interest are the markedly lower price paid by the industrial sector and both the higher price and greater increase in price in the residential sector over time. The boxplots show the variability in price within each year. The green center line of each boxplot tracks the median for the year, and the whiskers show the price variation for that year. The year 2022 shows greater variability than other years, and there is a large increase in the median price from 2020 to 2022.

Next, we analyzed how the average temperature per state impacted the cost of electricity in the year 2021. We merged the pricing data from the EIA, limited to the year 2021, with the temperature data that National Centers for Environmental Information provided through Climate Data Online. Across the 50 states, individually there was no correlation between the average retail price of electricity and the average temperature. When states were categorized by regions in the U.S. the northeast was a significant outlier for average retail price of electricity. While the midwest, southeast, southwest, and all averaged right around 10 cents per kilowatt hour, the northeast stood out with 15.06 center per kilowatt hour.

Lastly, we analyzed the relationship between the consumption per capita and the average temperature across the States in 2021. For this section, our hypothesis was that there would be a significant correlation between average annual temperature by state in 2021 and the amount of electricity consumed per capita.

We gathered data on the consumption of electricity per capita in each state. The source is IEA's State Energy Data System (SEDS), and specifically the data series labeled ESTCP which is electricity consumption in millions of kWh.

Similar to the previous analysis, we merged the resulting data set with the Climate Data Online's temperature data. Additionally, we pulled a list of US states by population in 2021 and divided total consumption in kWh by state population, in order to arrive at a per capita figure. This allows us to better compare the consumption habits of different states.

However, the linear regression model produced a coefficient of determination of 0.0045, which translates to an R score of 0.067. Therefore, the result of the regression test was that there is no evidence for correlation between annual average temperature by state and the quantity of kWh consumed in the United States.

### What We Learned

We learned to use a new API and the corresponding parameters. When we could not find an appropriate API for state weather data, we created a .csv to import data from a website. We cleaned and merged data, created dataframes, and calculated summary statistics to answer questions regarding relationships between the price of electricity and various factors. As we researched the factors affecting the price, we discovered additional factors to consider and included those. We investigated options for data visualization and decided on the most appropriate for highlighting what we learned about each factor.