What factors influence the price of electricity?

Frances Jay
Hugh Kelly
John Kiersznowski
Carol Love

GT Data Analytics - Project 1
02/02/2023

Overview

Goal:

Compare the relationships between electricity price and several factors: temperature, states, market sectors, time, and consumption.

Another factor that emerged during research was regional differences within the US.

Process:

Get data

Clean/merge data

Create numerical summaries

Create visualizations

Write analysis

Correlations:

Electricity Price + Temperature (John)

Electricity Price + State Sectors (Frances)

Electricity Price + Time (Carol)

Temperature + Consumption* (Hugh)
**We used retail sales of electricity by utilities as a proxy.*

APIs / Datasets:

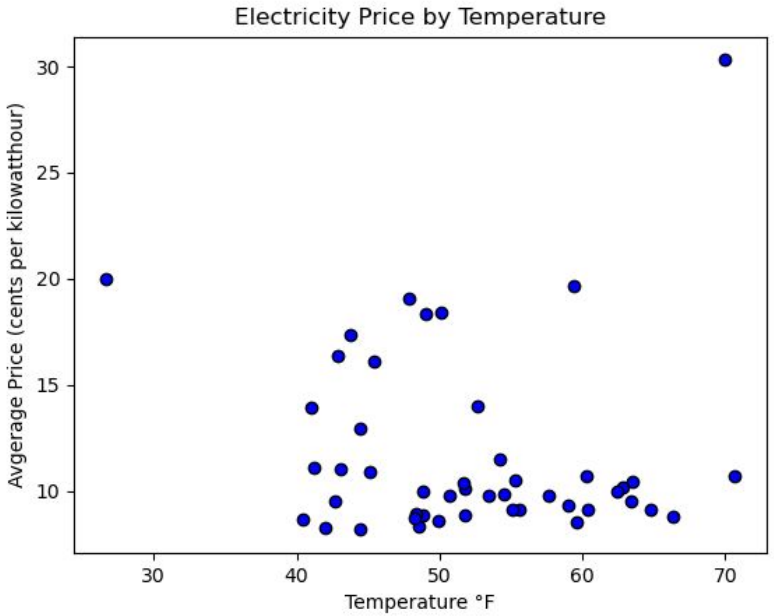
US Energy Information Administration (EIA)

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

Current Results (weather and science facts)

https://www.currentresults.com/Weather/US/average-annual-state-temperatures.php

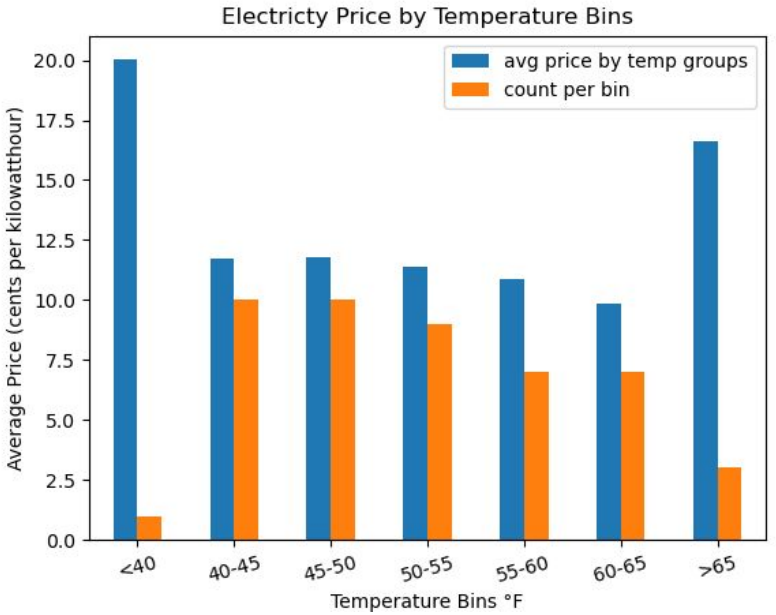
Electricity Price + Temperature



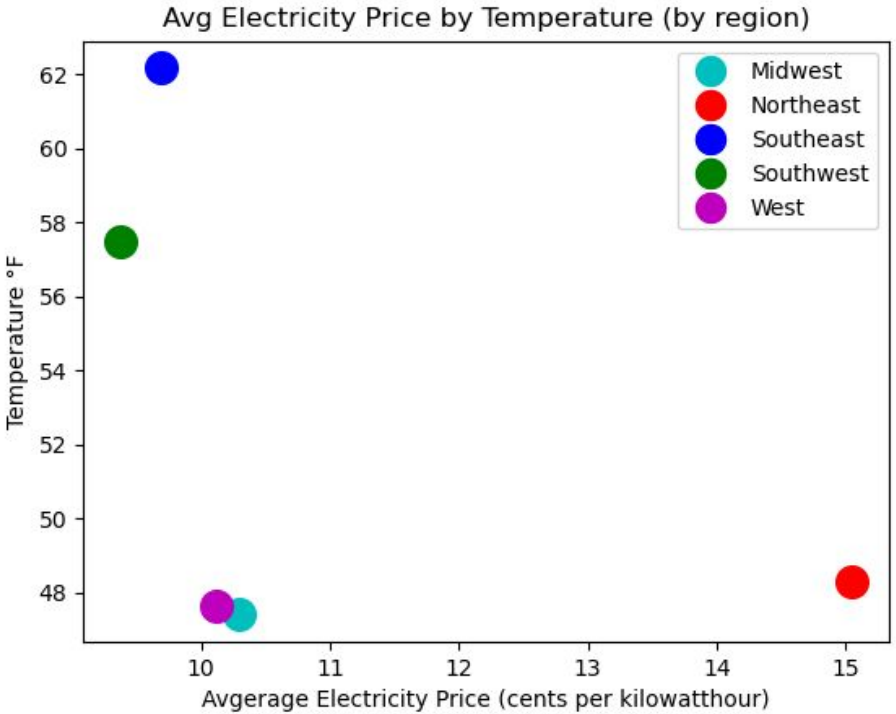
A standard scatter plot shows little to no correlation between electricity price and temperature for 50 states.

- Hawaii and Alaska are notable outliers, intuitively.

Temperature is binned by five degree increments for simpler visual comparison.



Electricity Price + Temperature



When states were categorized by region, the average retail price for states in the northeast was significantly higher than the other fairly consistently priced regions.

Electricity Price + Temperature

```
final_data.head()
```

	stateID	stateDescription	region	population	Avg °F	Avg °C	Rank	period	average-retail-price	average-retail-price-units	temperature bins
0	AL	Alabama	Southeast	5,049,846	62.8	17.1	7.0	2021.0	10.18	cents per kilowatthour	60-65
2	AZ	Arizona	Southwest	7,264,877	60.3	15.7	10.0	2021.0	10.73	cents per kilowatthour	60-65
3	AR	Arkansas	Southeast	3,028,122	60.4	15.8	9.0	2021.0	9.10	cents per kilowatthour	60-65
4	CA	California	West	39,142,991	59.4	15.2	12.0	2021.0	19.65	cents per kilowatthour	55-60
5	CO	Colorado	West	5,811,297	45.1	7.3	39.0	2021.0	10.90	cents per kilowatthour	45-50

Above shows the DataFrame used in this portion of the project.

The right code snippet shows importing, merging, and cleaning regional data as well as the groupby numerical analysis for the summary dataframe.

```python
# Import region data & merge with existing data
region = "data/region_data.csv"
region_df = pd.read_csv(region)
complete_data = pd.merge(region_df, all_data2, how="inner", on=["stateDescription"])

# Clean merged data & drop outliers (Alaska & Hawaii)
renamed_data = complete_data.rename(columns={"stateID_x": "stateID"})
drop_dup_column = renamed_data.drop(columns="stateID_y")
final_data = drop_dup_column.drop([1, 10])
```

```python
# Group data by region and calculate average Avg °F & average-retail-price
region_temp = clean_df.groupby(["region"]).mean()["Avg °F"]
region_price = clean_df.groupby(["region"]).mean()["average-retail-price"]

# Create dictionary & dataframe respectively
region_dict = {
    "average temperature °F": region_temp,
    "average retail price (cents per kilowatthour)": region_price,
}

regions_df = pd.DataFrame(region_dict)
regions_df
```

region	average temperature °F	average retail price (cents per kilowatthour)
Midwest	47.400000	10.291111
Northeast	48.281818	15.050909
Southeast	62.180000	9.685000
Southwest	57.500000	9.365000
West	47.644444	10.121111

Electricity Price + State Sectors

```
base_url = "https://api.eia.gov/v2/electricity/retail-sales/data"

X_Params = {
    "api_key": api_key,
    "frequency": "monthly",
    "data[0]": "price",
    "facets":{},
    "start": "2021-01",
    "end": "2021-12",
    "sort[0][column]": "period",
    "sort[0][direction]": "desc",
    "offset": 0,
    "length": 5000
}

response = requests.get(base_url,params=X_Params).json()
```

```
sectorDF = pd.DataFrame(response["response"]["data"])
display(sectorDF.head())
print(f"Number of Rows: {len(sectorDF)}")
```

	period	stateid	stateDescription	sectorid	sectorName	price	price-units
0	2021-12	IN	Indiana	OTH	other	NaN	cents per kilowatthour
1	2021-12	US	U.S. Total	TRA	transportation	10.49	cents per kilowatthour
2	2021-12	US	U.S. Total	RES	residential	13.72	cents per kilowatthour
3	2021-12	US	U.S. Total	OTH	other	NaN	cents per kilowatthour
4	2021-12	US	U.S. Total	IND	industrial	7.06	cents per kilowatthour

```
Number of Rows: 4464
```

```
sectorDF.dropna(subset=["price"],inplace=True,how="all")
```

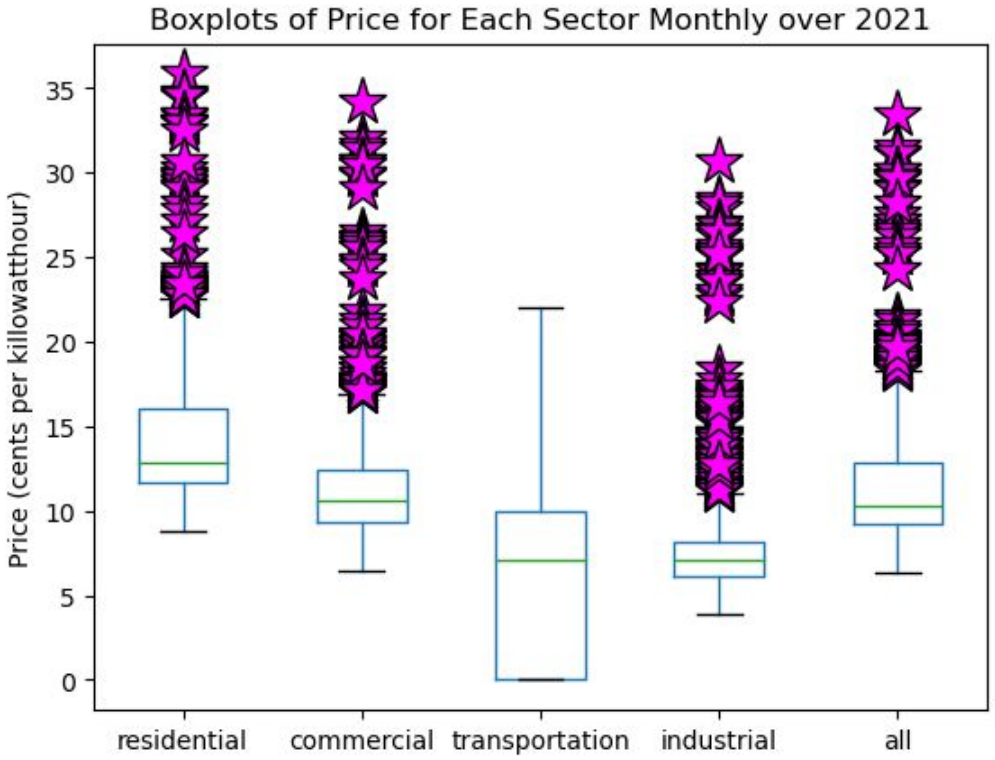
```
len(sectorDF)
```

```
3720
```

I solely used the retail sales data available through the EIA for my analysis. I honed in on 2021 because that was the most recent year with data for every month at the beginning of our project.

Electricity Price + State Sectors

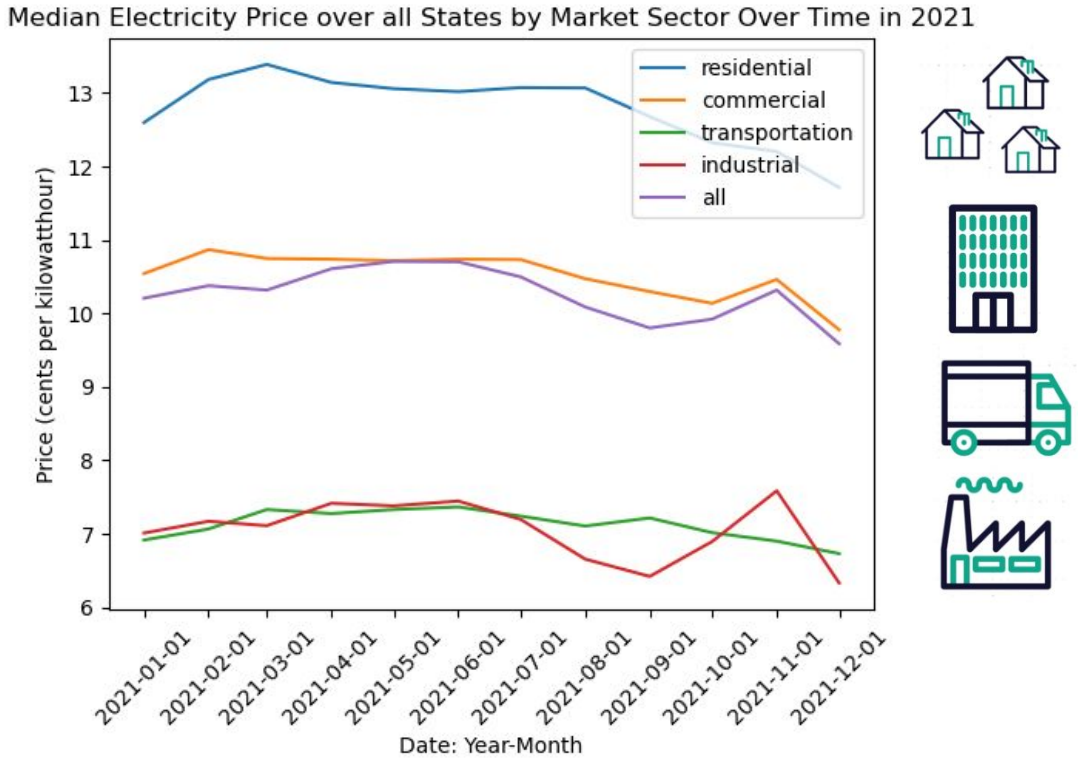
In 2021, the price of electricity across all states varied somewhat by sector. The spread for the residential sector is higher than other sectors, while the transportation sector is lower than the others.



Electricity Price + State Sectors

The rate increases and decreases in the median price of electricity across sectors mirrored each other fairly well in 2021.

Like we saw in the box plots, the residential sector has higher prices than other sectors, and transportation/industrial sectors have lower prices.



Electricity Price + Time

Industrial customers tend to pay less for electricity than other sectors.

Residential customers tend to pay more for electricity than other sectors.

Residential customers also saw a greater increase in price over time.

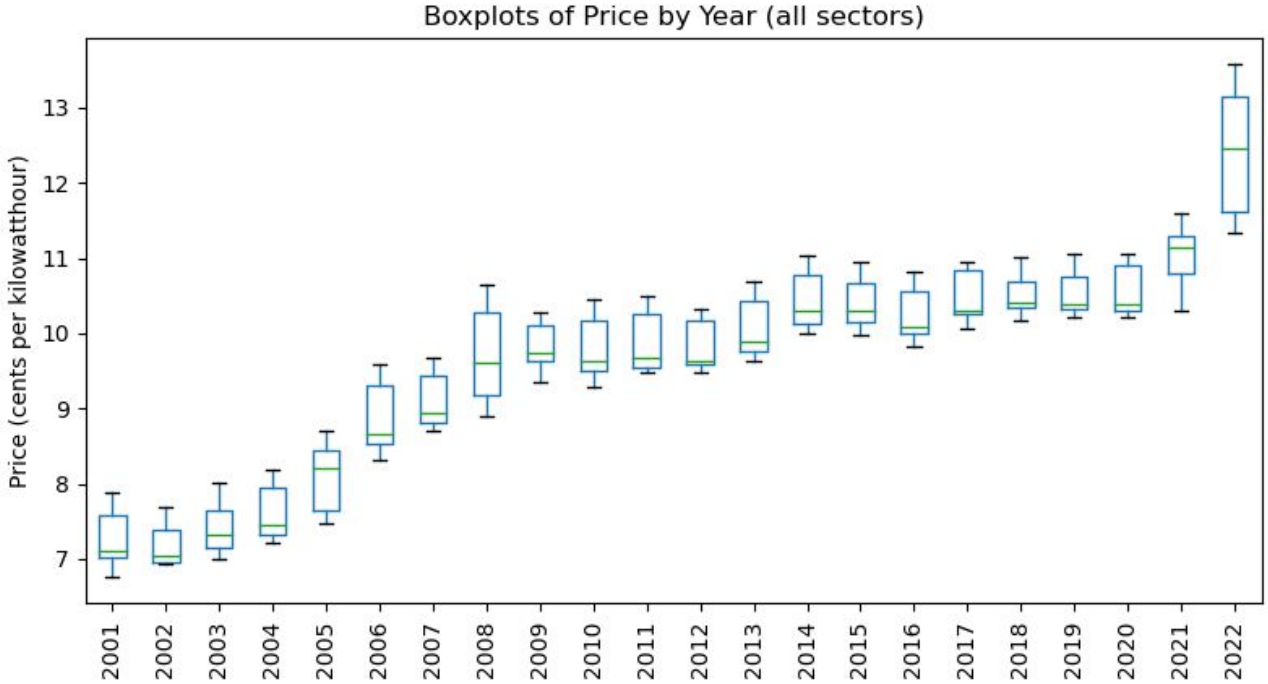
Electricity Price + Time

The boxplots show the median and spread of the prices for each year.

In general, the median price (green line) increases over time.

Most years have similar variability in price, with notably larger spreads in 2008 and 2022.

Also to note: the large increase in median price from 2020 - 2022.



Electricity Price + Time

In [15]:
```
year_summary_stats = time_cost_all.groupby(["year"]).agg(["mean","median","var","std","sem"])["price"]
year_summary_stats
```

Out[15]:

year	mean	median	var	std	sem
2001	7.268333	7.100	0.148342	0.385152	0.111184
2002	7.175000	7.040	0.075373	0.274541	0.079253
2003	7.416667	7.330	0.123970	0.352093	0.101641
2004	7.595000	7.440	0.125100	0.353695	0.102103
2005	8.107500	8.210	0.211893	0.460319	0.132883
2006	8.860833	8.670	0.207081	0.455062	0.131365
2007	9.104167	8.950	0.126863	0.356178	0.102820
2008	9.705000	9.600	0.401100	0.633325	0.182825
2009	9.804167	9.735	0.101808	0.319074	0.092109
2010	9.797500	9.635	0.161711	0.402134	0.116086
2011	9.863333	9.680	0.158297	0.397866	0.114854
2012	9.815833	9.640	0.110263	0.332059	0.095857
2013	10.043333	9.890	0.148188	0.384952	0.111126
2014	10.420833	10.305	0.127990	0.357757	0.103276
2015	10.385000	10.300	0.115391	0.339692	0.098061
2016	10.242500	10.095	0.113384	0.336726	0.097204
2017	10.458333	10.310	0.109070	0.330257	0.095337
2018	10.506667	10.410	0.078861	0.280821	0.081066
2019	10.516667	10.385	0.090842	0.301401	0.087007
2020	10.560833	10.385	0.114227	0.337974	0.097565
2021	11.071667	11.150	0.156397	0.395471	0.114163
2022	12.436000	12.460	0.754071	0.868373	0.274604

We used the API key and parameters from the US Energy Information Administration (EIA) API to get electricity price data, then removed rows that were missing data for the price of electricity.

We used .agg to create a dataframe of summary statistics for the price of electricity by year (shown) and by month (not shown).

https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.agg.html

In [3]:
```
# Set the base url

elec_url = "https://api.eia.gov/v2/electricity/retail-sales/data/"

# Define the search parameters

parameters = {"api_key": elec_key,
              "frequency": "monthly",
               "data[0]": "price",
               "facets[sectorid][]": ["ALL", "COM", "IND", "RES", "TRA", "OTH"],
               "facets[stateid][]": "US",
               "start": "2001-01",
               "end": "2022-10",
               "sort[0][column]": "period",
               "sort[0][direction]": "asc",
               "length": "5000",}

# Get the json data from the API

time_cost = requests.get(elec_url, params = parameters).json()

# Create a DataFrame from the json data

time_cost = pd.DataFrame(time_cost["response"]["data"])
time_cost.head()
```

Out[3]:

	period	stateid	stateDescription	sectorid	sectorName	price	price-units
0	2001-01	US	U.S. Total	ALL	all sectors	6.75	cents per kilowatthour
1	2001-01	US	U.S. Total	TRA	transportation	NaN	cents per kilowatthour
2	2001-01	US	U.S. Total	RES	residential	7.73	cents per kilowatthour
3	2001-01	US	U.S. Total	OTH	other	6.48	cents per kilowatthour
4	2001-01	US	U.S. Total	IND	industrial	4.73	cents per kilowatthour

In [4]:
```
# Remove rows that are missing prices

time_cost.dropna(subset=["price"], how = "all", inplace = True)
time_cost.head()
```

Out[4]:

	period	stateid	stateDescription	sectorid	sectorName	price	price-units
0	2001-01	US	U.S. Total	ALL	all sectors	6.75	cents per kilowatthour
2	2001-01	US	U.S. Total	RES	residential	7.73	cents per kilowatthour
3	2001-01	US	U.S. Total	OTH	other	6.48	cents per kilowatthour
4	2001-01	US	U.S. Total	IND	industrial	4.73	cents per kilowatthour
5	2001-01	US	U.S. Total	COM	commercial	7.25	cents per kilowatthour

Temperature v. Electricity Consumption per Capita by US State - 2021

United States overall total electricity consumption 2021

$$\frac{\text{total kWh consumption}}{\text{total population}} = \frac{\sim 3{,}500{,}000 \text{ million kWh}}{\sim 332 \text{ million people}}$$

$$\sim 11{,}500 \ \text{kWh/capita}$$

source: EIA SEDS ESTCP kWh units

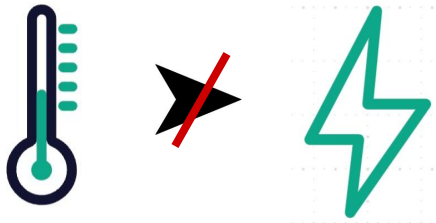
Temperature v. Electricity Consumption per Capita by US State - 2021

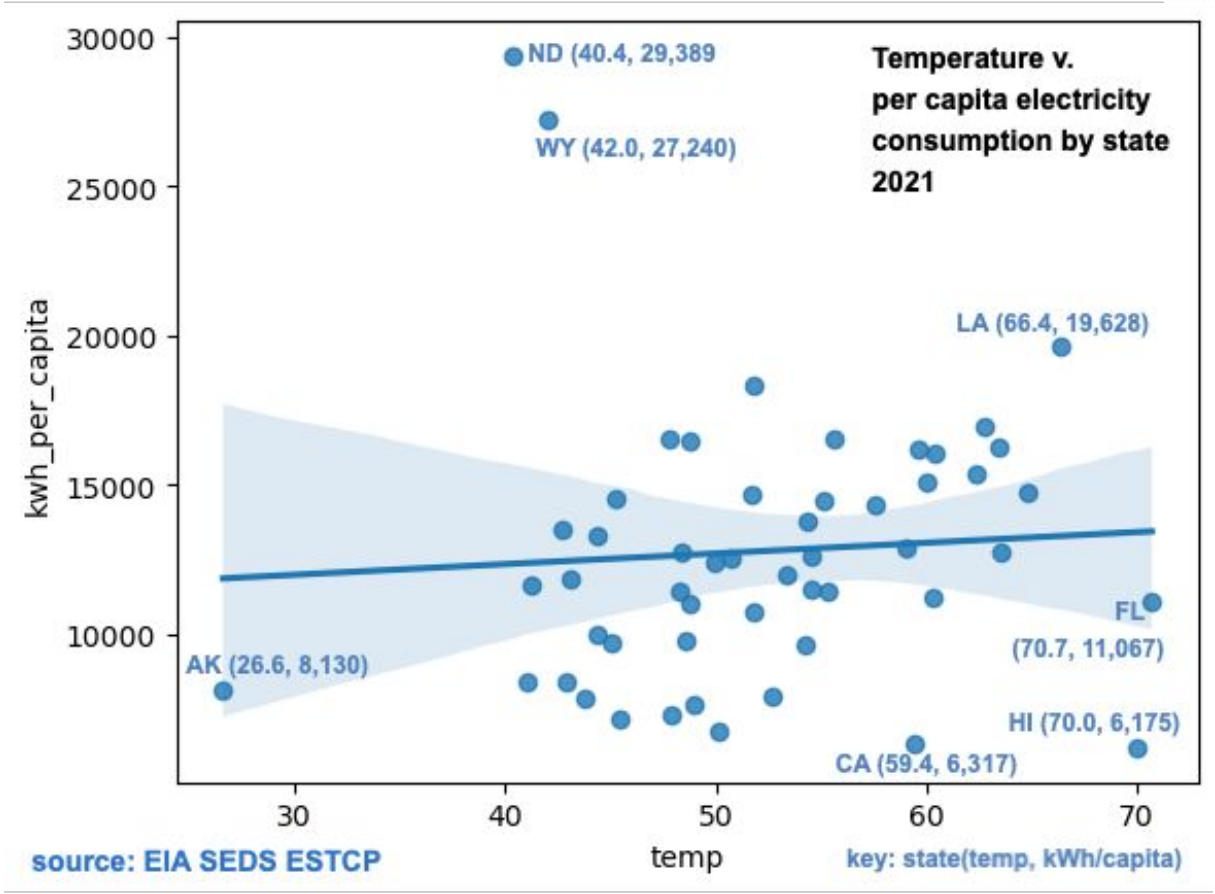
Linear regression
independent variable: temperature
dependent variable: consumption

Coefficient of determination (R-squared):
0.0045

R score (correlation coefficient)
$\sqrt{R^2} = R = 0.067$

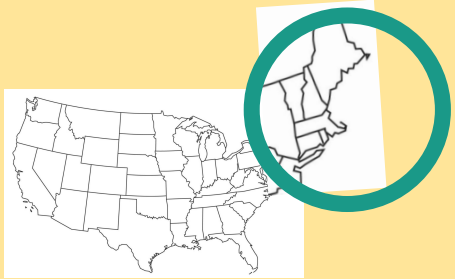
No discernable relationship between temperature and electricity consumption by state





Takeaways

John:



Electricity costs more in the Northeast

Frances:

Residential sector price > Transportation/Industrial

Similar rate of change across all sectors.

Carol:

Greatest price increase from 2020 - 2022.

Hugh:

2021 consumption per capita

~11,500 kWh/year

~1,000 kWh/month

Questions, Comments, Concerns?

Temperature v. Electricity Consumption per Capita by US State - 2021

How outliers help to understand limitations of the dataset & weaknesses in the hypothesis that per capita electricity consumption is correlated with temperature

State	Avg. 2021 temp	kWh per capita	possible reason	corresponding limitation
CA	59.4	6,317	high cost($0.239/kWh) geographic temp variation higher proliferation of DERs	price varies among states population not evenly distributed according to temperature variation a higher proportion of kWh consumed not accounted for in utility-based retail sales data

Temperature v. Electricity Consumption per Capita by US State - 2021

State	Avg. 2021 temp	kWh per capita	possible reason	corresponding limitation
HI	70.0	6,175	high cost($0.30/kWh) less temp variation higher wind speeds	price varies among states less work for HVAC less work for HVAC
FL	70.7	11,067	lowish cost($0.125/kWh) more seasonal variation of temp high humidity	price varies among states more work in summer & winter for HVAC all else equal, more work for HVAC
LA	66.4	19,628	low cost($0.117/kWh) more seasonal variation of temp high humidity	price varies among states more work in summer & winter for HVAC all else equal, more work for HVAC

Temperature v. Electricity Consumption per Capita by US State - 2021

State	Avg. 2021 temp	kWh per capita	possible reason	corresponding limitation
WY	40.4	27,240	very low cost($0.109) more seasonal temp variation	price varies among states more work in summer & winter for HVAC
ND	42.0	29,398	very low cost($0.109) more seasonal temp variation	price varies among states more work in summer & winter for HVAC
AK	26.6	8,130	high cost($.226) less electricity used in building energy	price varies among state more thermal energy sources for building energy