Comparison of Green Energy Growth and Energy Prices

Ryan Eskuri, Benjamin Schmitt, and Stephen Moilanen

Background and Motivation

As climate change makes us question the wisdom of using traditional forms of electricity production, a transition to greener forms of energy has become more and more common. As that transition has progressed, existing infrastructure has been added to the electric system. Critics have argued one major flaw of wind and solar is that you can't turn on the wind with the flick of a switch or turn up the sun with a turn of a knob, which distracts from a utility's responsibility to meet peak demand. This has led some to say that green energy is not a serious replacement for fossil fuels, only an expensive add-on.

Other potential hidden costs of green energy are argued as well. This includes green energy generators indirectly causing more wear and tear in fossil fuel plants (through increased cycling) and necessary changes to the transmission and distribution systems that would not be required otherwise.

With the politics and subsidies surrounding green energy, coupled with the complexity and variations in utility rate designs, identifying the truth in these criticisms may be hard to do at a surface level. Our analysis seeks to make some sense of it and compare how energy costs have changed as the electrical grid and climate have changed.

Goals

1. Determine, through finding correlations or a lack thereof, any relationship between the increase of green energy use and the increase in electricity costs over time.

2. Use weather data to find relationships as described in Goal 1, albeit in similar weather conditions based on temperature, sunshine, etc.

3. Isolate certain states/regions to determine relationships between energy prices and green energy implementation, to hold constant certain regulatory considerations.

Data Sources

Data Sourced from US Energy Information Administration and NOAA

Data from the United State Energy Information Administration

Data from the Energy Information Administration (EIA) was collected in three datasets from the following URL:

https://www.eia.gov/electricity/data.php

The first dataset included monthly data of electricity sales in megawatt-hours, price in cents per kilowatt hour, number of customers, and revenue in dollars, all organized by state for the years 1990 - 2022.

The second dataset includes monthly electricity production organized by production source (i.e., wind, solar, coal, etcetera) and state. Data was only included for the years 2001 to June of 2022. Electric generation values were provided with units of megawatt-hours. The data was organized in a series of sheets within an Excel file, and a Python function was written to extract and combine all the sheets.

The third dataset provided yearly electricity production organized by production source and state from the years 1990 to 2000. Electric generation values were provided with units of megawatt-hours. Monthly data was not available for this period.

Data from the National Oceanic and Atmospheric Administration

Our secondary dataset contains monthly data on temperature in degrees Fahrenheit, precipitation in inches, and the Palmer Z-Index for drought for the continental US. Changes from the mean for each of those parameters was also provided. Each variable for every state's data is available in a separate csv file via the NOAA Nation Centers for Environment Information:

https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/statewide/time-series

While there is data going back to the 1800's, we limited the scope to 1990-2022. When all states are combined, there are 88,704 records for this time period and set of variables.

Data for the entire US was also downloaded. It included an average of the same monthly climate parameters for the 48 contiguous states.

Data Manipulation Methods

Part 1 - Combining and Cleaning Separate Datasets

Preparation of EIA Data

Data from the EIA included three separate sets that were manipulated and reshaped to be merged/concatenated. The first was the energy price data. Prices for the states were grouped together to provide a separate value for the US. Nan values were replaced with '0'.

The second EIA dataset included monthly production data from 2001 to 2022. A function was defined to read in each MS Excel sheet in the file and combine them together. Errors were then corrected, and preparations made for merging/concatenation.

The third EIA dataset was yearly production data from 1990 to 2000. Monthly data was derived from this dataset by taking the monthly average in each year and using that average for the monthly total.

Irrelevant or redundant fields and rows were removed from each of the three EIA datasets, the rows/columns restructured, and a datetime column added so that each could be merged or concatenated together using typical pandas functions.

Preparation of NOAA Data

NOAA has a vast amount of data to choose from. To best match our energy datasets, we opted to pull monthly averages for all states from the past thirty years along with the total US averages. For each state, every metric was stored in a separate csv file. Due to this limitation, we narrowed the metrics chosen to just three: Temperature, Precipitation, and Z-index Drought values.

Each states' three csv files were uploaded in a corresponding state-named folder. We created two main functions to first read and clean each csv into pandas and then merge each dataframe based on state. We had already chosen to focus on the continental US, so Alaska and Hawaii were excluded, but they were also missing drought level measures in the early 1990's. The District of Columbia (DC) had separate energy data, but no weather data; to compensate we made the DC weather data equal to that of Maryland. The other 48 states did not have any missing records. Because of this, we were able to concatenate each state with a dataset of monthly average weather data of the entire United States.

Aggregation of Price, Production, and Weather Data

First, the production data from 1990 to 2000 was concatenated with the production data from 2001 to 2022. Then, the energy price data was merged with the production data, using the "Date" column as a basis for merging.

Finally, the weather dataframe was merged with the energy production/sales data (again using the "Date" column) to have all the data in one place. The dataset included monthly values from 1990 to mid-2022. Exploration was then performed to find NaN values which were addressed as described in the notebook.

Data Manipulation Methods

Part 2 - Manipulating and Organizing Data for Visualizations

The aggregated principal dataset from Part 1 of the data manipulation process was used by each of the team members to perform his own separate analysis. Each team member performed his own manipulations of the data as described below.

Comparing Electric Production Growth with Price (pages 5 and 6)

To explore the data in a repetitive way, a function was written in which the user could define parameters to control how the dataset was manipulated. Dates could be truncated, months grouped into years, and certain energy sources left out, among other selections. The function took in the primary dataframe as an argument and output both a dataframe and a plot.

To explore trends in specific states, a second and a third function were defined. The second function calculated several growth metrics and populated them into a new dataframe using groupby functions, merging, and pivoting. Again, manipulation parameters were controlled by the function's input. The third function used the print function to create a report to compare metrics between states.

For report plots, a fourth function was written, similar to the first.

Comparing Green Energy, Price, and Temperature (pages 7 and 8)

To get an easily comparable metric of energy production by source for each State, new variables needed to be added to the existing data. This required filtering the columns of the full dataframe, then pivoting the generation by energy source into columns using the Pandas pivot_table function. Calculated proportion of total electricity generation represented by select energy sources was added as new columns. This produced a dataframe with variables useful for comparing green energy production between States. The Pandas groupby method could now be used to aggregate data by source for visualization. One last element necessary for creating choropleths with the Altair library was the addition of a State ID column corresponding to the State geometry in the geo-json file in Altair's vega datasets. This allowed production of a base map with the state boundaries. Using this new dataframe, a variety of choropleth maps depicting relationships like green energy use by State, temperature by State, and price by State (see Appendix 1 – Section 6) were produced. This new dataframe also facilitated ranking of States by different energy source proportion values.

Comparing Solar Energy, Drought, and Precipitation (page 9)

Comparing some of the main variables from our datasets became straight forward after initial data wrangling. Building a function that filtered states, years, energy production and then plotting alongside remaining weather information made visual EDA quick and simple. Choosing from the many combinations of variables became the more difficult task. We chose matplotlib for flexibility in case a large number of records were displayed, but for visual clarity only four years were used in the final visualizations.

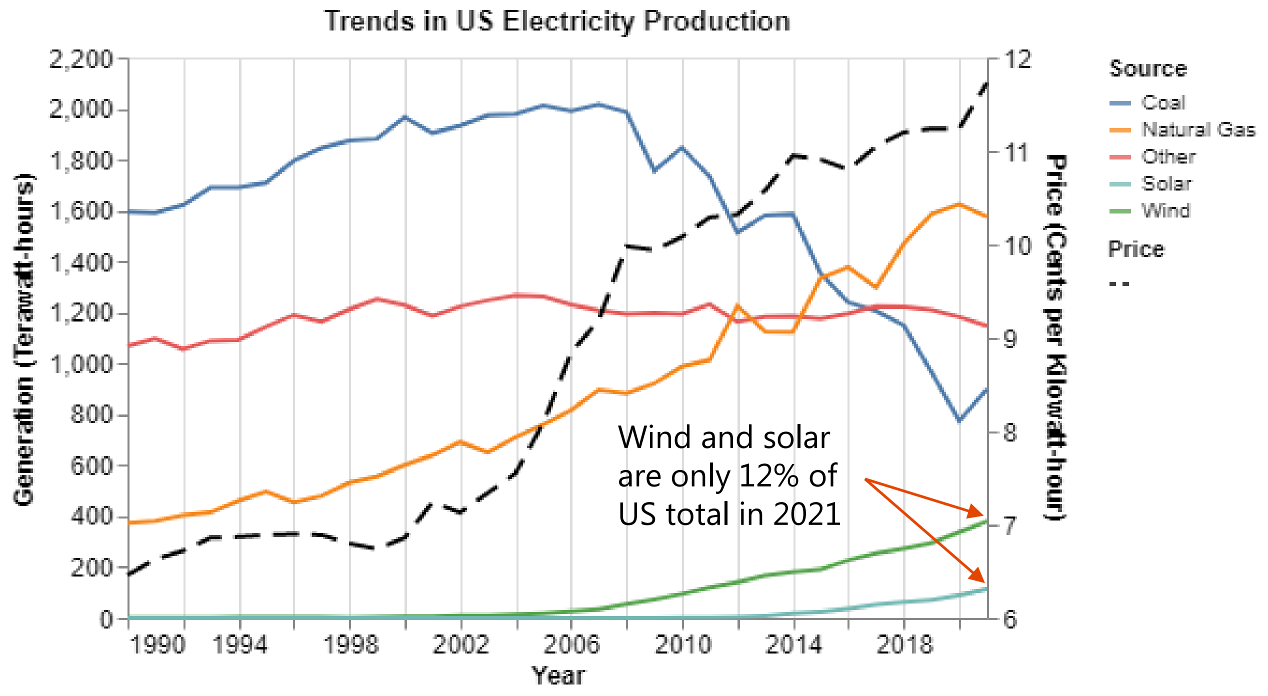
Analysis and Visualization – US Energy Growth from 1990 to 2021

The Rise of Natural Gas, Not Green Energy

The changes in the energy sector regarding production type and price can be broken into decades. The years **1990 to 2000** mainly saw **growth in nuclear, natural gas, and coal** production. Wind and solar growth was negligible as a percentage of the total output. Growth in energy costs for consumers per kWh were negligible (around 6%).

The **2000's** started to see major changes in the energy sector. Growth in coal generation stagnated and started to fall in 2008. Wind electric output grew approximately 1,500%, mostly in the later half of the decade. The highest volume of **growth** by far was from **natural gas (+386,000 GWh)**, followed by **wind (+89,000 GWh)** and **nuclear (+53,000 GWh)**. Price grew 47%.

Between **2010 and 2021** (11 years total) marked the start of the **downfall of coal production** which was cut in half over that time period, dropping about -949,000 GWh. **Wind output accelerated** from the previous decade, growing by +285,000 GWh. **Solar production accelerated** (particularly around 2015), adding **+113,000 GWh** or more than **9,360%** over the previous decade. **Natural gas production growth was again the frontrunner** with +588,000 GWh, **more than twice of wind and more than all green energy growth combined.**



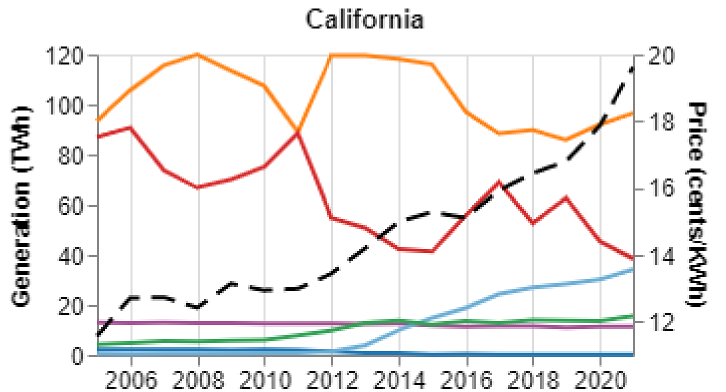
Clearly, the more recent history of the electricity production sector has not necessarily followed a theme of green energy, but rather a theme of pollution reduction. Natural gas, albeit a fossil fuel, burns cleaner than coal and it is relatively economical to covert coal plants to natural gas plants. For this reason, **the rise of natural gas boilers in fossil fuel plants overshadows the rise of green energy**, particularly wind and solar. To gain more insights, it is necessary to take a **closer look at jurisdictions that have produced above average green energy** production.

Analysis and Visualization – Energy and Price Growth in Select States

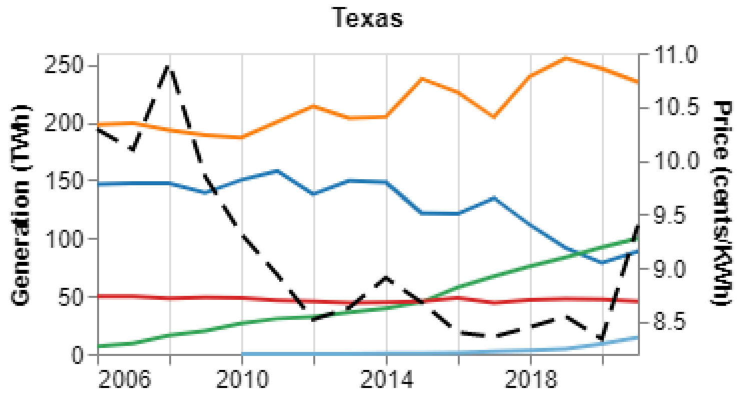
Comparing Green Energy and Energy Costs in Defined Jurisdictions

The four states shown here were selected as leaders in green energy production increases between 2006 and 2021. Price increases are compared to the national average of +32.7% - or +2.9 cents - over the same timeframe. The plots show that three of the four states had price increases **below** the national average. The state that did not beat the national average – California – had the highest price increases in the entire country and was more than double the national average by cents.

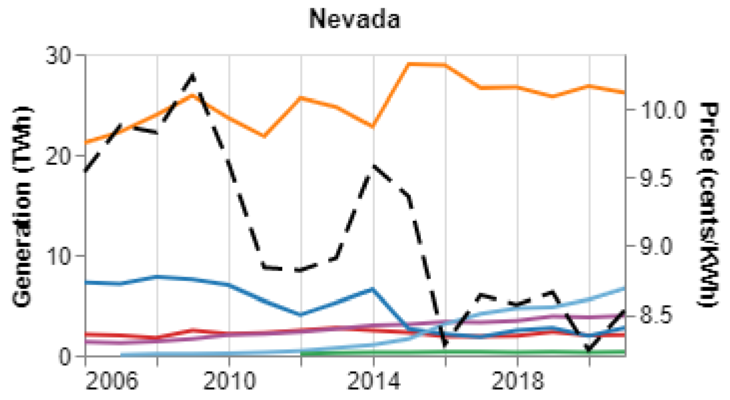
Nevada is the only state that has significant growth in geothermal energy and is ranked second in solar energy production – 15.9% of total production in 2021 came from solar and 9.4% from geothermal. In the same year, California produced 17.9% of it's total energy from solar and 13.7% from wind/geothermal, while Iowa produced 55.3% from wind. Texas was ranked #1 and #2 in wind/solar growth respectively since 2006 (by total MWh). It produced 20.7% of it's total energy from wind but only 2.9% from solar in 2021.



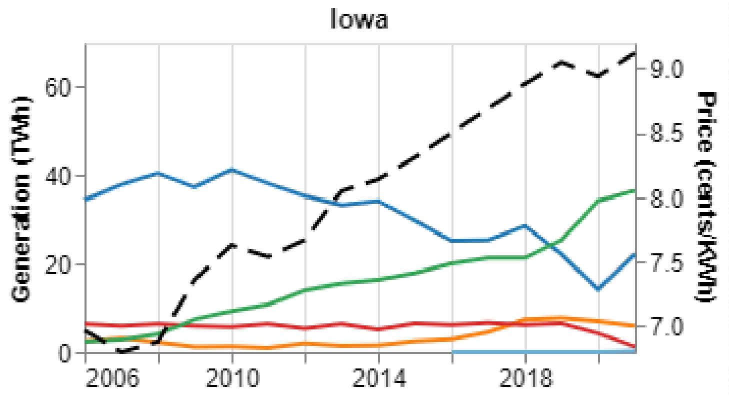
California ranked high in solar growth with some wind growth. Prices increased 6.9 cents (54.3%).



Texas had high wind and some solar growth. Prices decreased 0.9 cents (-8.6%).



Nevada had significant geothermal and solar growth. Prices decreased 1.0 cents (-10.6%).



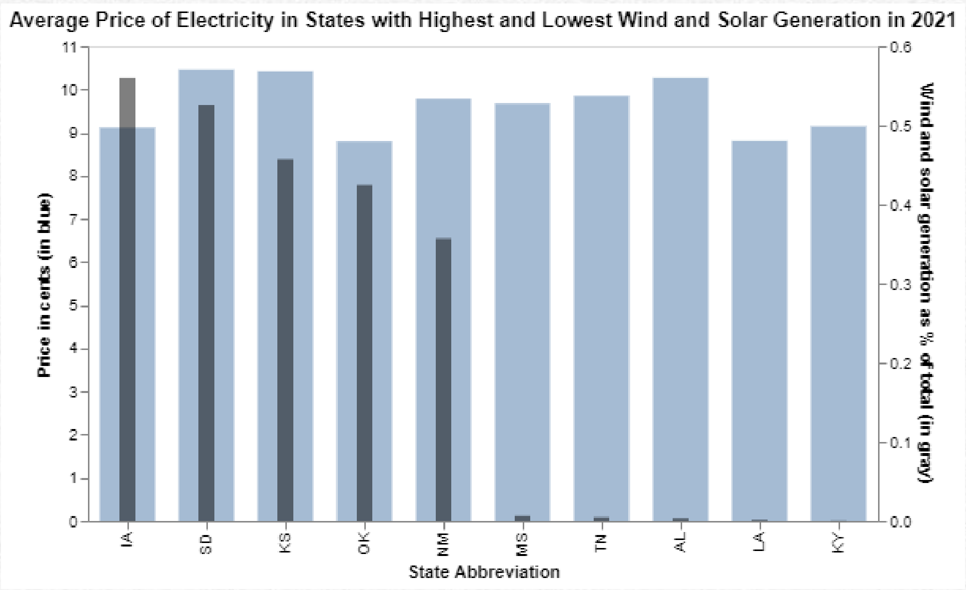
Iowa is currently top-ranked for wind production. Prices increased 2.1 cents (30.7%) since 2006.

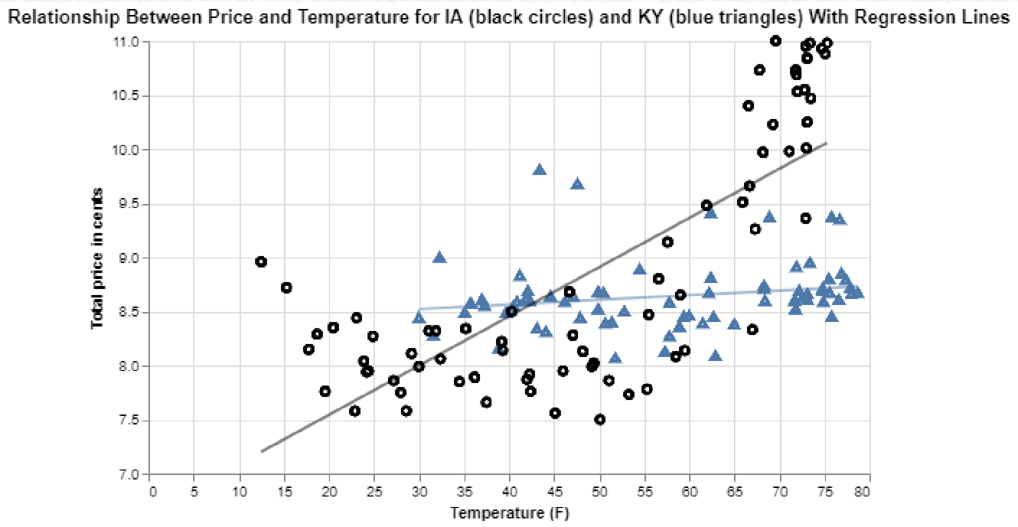
Analysis and Visualization – State Level Relationships

Green Energy, Price, and Temperature

In 2021 the **five States with the highest average wind and solar electricity generation** as percent of total generation, in descending order, **were: Iowa, South Dakota, Kansas, Oklahoma, and New Mexico**. In 2021 the five States with the lowest average wind and solar electricity generation as percent of total generation, in ascending order, were: Kentucky, Louisiana, Alabama, Tennessee, and Mississippi.

As you can see in the bar chart below, **average electricity price is not much different in States that generate a large portion of their electricity from wind and solar as compared to States that do not** get much electricity from wind and solar sources. The blue bars encode electricity price. The gray bars encode wind and solar generation as proportion of total generation.



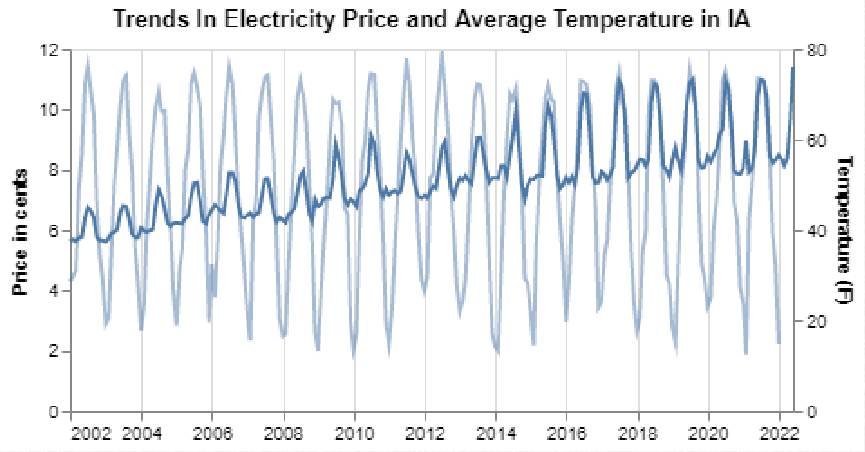


The scatter plot above depicts the relationship between price of electricity and temperature from 2016 to 2021 in Iowa and Kentucky. Monthly average temperature is on the x axis and monthly average price is on the y axis. Iowa data is represented by black circles and Kentucky data is represented by blue triangles. The regression lines are the result of ordinary least squares regression on each State's data. As you can see, the price of electricity in Iowa gets higher as temperature increases. However, in Kentucky the price of electricity does not increase much with increasing temperature. This pricing difference might reflect the respective electric utilities desire to reduce demand. Utilities that rely on wind and solar may not be able to increase generation quickly so they may utilize price increases to reduce demand.

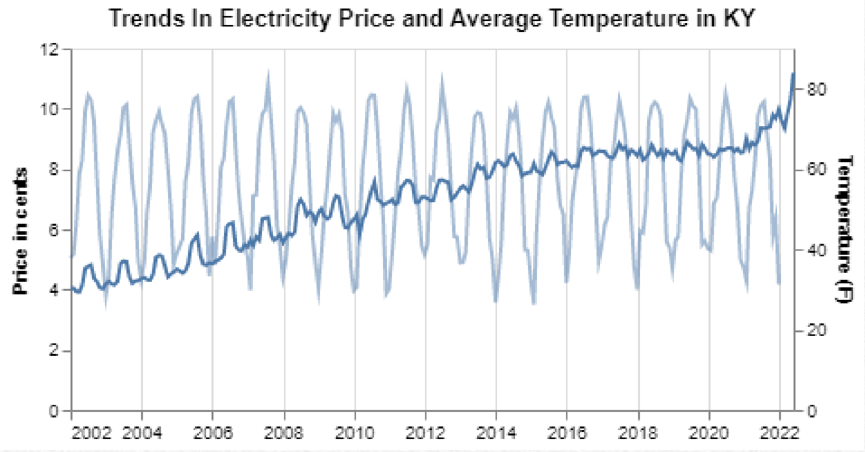
Analysis and Visualization – State Level Relationships

Green Energy, Price, and Temperature

There is a **strong seasonal relationship between temperature and price of electricity** in the United States overall (See Appendix X-Section X) and in individual states. Price of electricity starts to increase in the spring and typically peaks in July or August. This corresponds with average temperature over the same period. There is also **strong seasonal relationship between temperature and total electricity generation** (See Appendix X-Section X). We didn't have demand information in our data, but we can assume that increased generation is a result of increased demand. So, the observed price increases might be the utilities response to increased demand. Changing price in response to demand is a technique that is frequently used by utilities to make up for higher costs and reduce peak demand. Accommodating large fluctuations in demand can be expensive for a utility because they either need to increase production or buy excess power from another source.

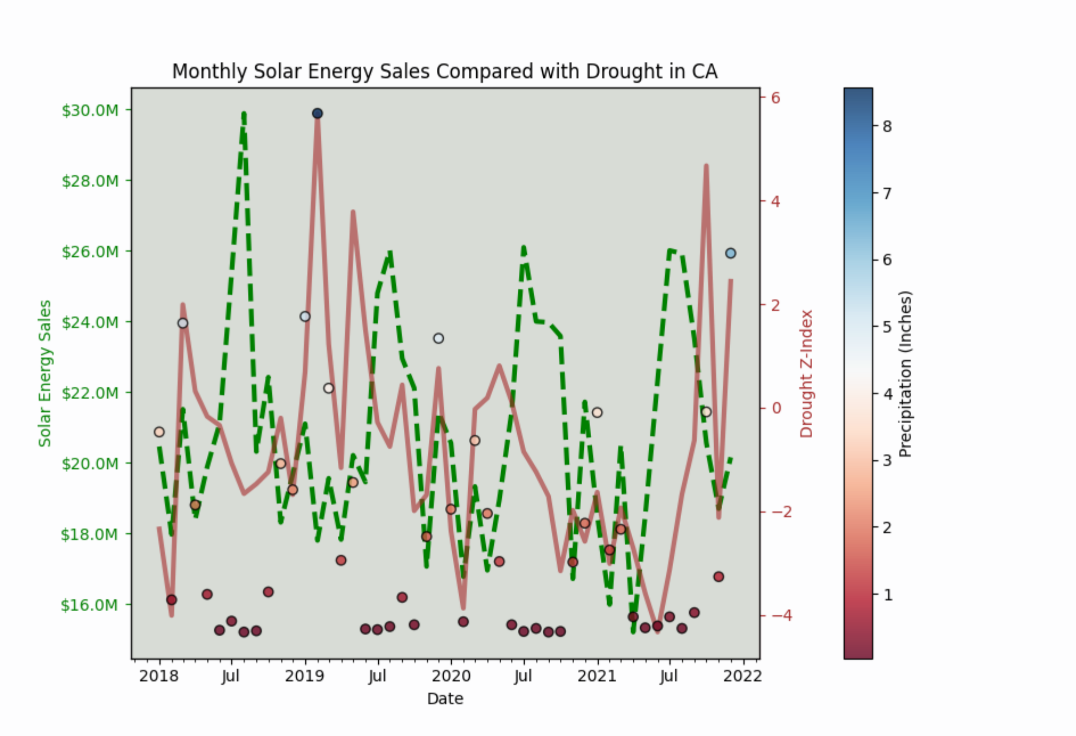


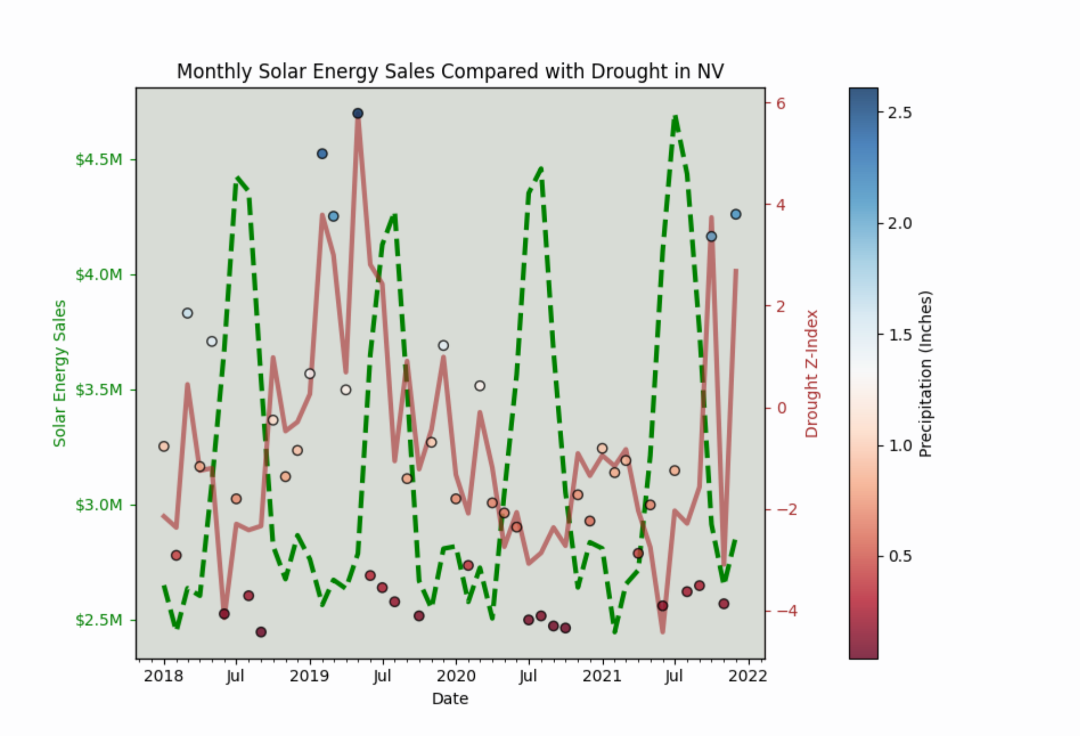
We observed **different relationships between price and temperature in different states**. In some states, like Iowa, price is highly correlated with temperature in the spring and summer. This trend is more pronounced after 2014. This can be seen in the top right line chart, where the lines for temperature and price overlap in spring and summer months. In other states, like Kentucky, price is not as strongly correlated with temperature. This can be seen in line chart on the bottom right. It looks like temperature became less corelated with price in Kentucky around 2014.



This difference in the price temperature relationship may be related to green energy use. We picked Iowa and Kentucky for comparison because, in 2021, Iowa had the highest average wind and solar electricity generation as percent of total generation and Kentucky had the lowest. Utilities that rely on wind and solar may not be able to increase generation quickly so they may utilize price increases more frequently than utilities the use other energy sources.

Solar Energy, Drought and Precipitation:
A Tenuous Connection





We were curious if the amount of solar energy sold to customers was affected by other factors than seasonality. Drought levels and precipitation were used as **proxies for general cloud cover** in a state. Note that on the Drought Z-Index a **lower value indicates more severe drought**. California and Nevada were chosen for closer examination as they **produce the largest ratio of solar energy.** They might show trends more readily. The most surprising part of this exploratory analysis is how **solar sales almost double during Summer** compared to Winter.

Leading into the summer of 2019, both Nevada and California saw large amounts of precipitation and low drought which led to **lower Summer sales than 2018**. However, California sales did not change much the following two summers with lower precipitation and higher drought. Winter sales are consistently low across the board with **variation regardless of different precipitation or drought levels**. This makes it difficult to draw any concrete conclusions and there are other factors to consider such as reduced solar production due to large wildfires or our proxy being inaccurate. **Monthly averages may be too broad** of a scale to see trends that could potentially exist, especially in Summer months.

Discussion, Conclusions, and Limitations
Many Correlations, Few Clear Answers

Discussion

The data selection and analysis options for this topic seem infinite. We spent considerable time and effort to produce our analysis and were left with more questions than answers. Much of our time was spent exploring correlations. If we had more time, it would be very interesting to investigate causal relationships between some of the correlated variables.

Summary of Goals

1. Determine relationship between green energy use and electricity costs over time.

2. Use weather data to find relationships as described in Goal 1, albeit in similar weather conditions.

3. Isolate certain states/regions to determine relationships between energy prices and green energy.

Conclusions

Goal 1. In the US, both green energy use and electricity price increased over time, but there was no obvious causal relationship.

Goal 2. Some States that use lots of green energy have higher electricity costs during the hottest months of the year. This relationship does not always hold.

Goal 3. In some States, electricity price decreased while green energy increased. In other States, electricity price increased while green energy increased.

Limitations

The choice to use US States as a geographic boundary for our analysis may have limited our ability to identify important relationships. Using States as geographic units is problematic because most States do not operate on an independent electric grid. Instead, states are constantly importing and exporting electricity with surrounding states. Looking at State level data might obscure important relationships between generation source, price, and weather.

Federal subsidies likely influence the price of electricity. Our analysis only examined price of electricity and did not consider government programs designed to influence electricity price. Both fossil fuel and green energy sources receive federal subsidies.

Another factor that influenced our price analysis was the EIA's method of price calculation. The calculated price included energy charges, demand charges, consumer service charges, environmental surcharges, fuel adjustments, and other miscellaneous charges. This may or may not have influenced our ability to identify relationships between price and other variables of interest.

Data sample frequency may have limited our analysis. The monthly time steps in our data provided for more manageable data analysis and visualization. However, we were not able to explore finer scale phenomena. For example, it would have been interesting to see how local heat waves influenced generation and price. This was not possible with State level monthly average temperature and price data.

Lastly, the free version of Deepnote, which we used for collaboration, proved to have limited RAM. This imposed some restrictions on the maximum size of our dataframes.

Statement of Work

Summary of Team Contributions and Endnotes

Ryan Eskuri

- NOAA data cleaning
- EDA of state weather data
- Energy/drought visualization creation
- Notebook cleanup
- Report creation

Benjamin Schmitt

- EDA of energy and weather data
- Green energy, price, and temperature visualization
- Notebook cleanup
- Report creation

Stephen Moilanen

- EIA data cleaning and data aggregation
- EDA and visualizations of price/production trends
- Comparisons of energy price and generation
- Notebook cleanup
- Report creation

Endnotes

- Appendix 1 contains the notebook, coded in Python
- Due to the length of Appendix 1, some supplementary csv files were created by the notebook and then called in later cells
- For better collaboration, future work could include a detailed task chart assigning specific and ordered tasks to team members

References/Citations

U. S. Energy Information Administration. (2022, April 20). Electricity explained: Factors affecting electricity prices. https://www.eia.gov/energyexplained/electricity/prices-and-factors-affecting-prices.php

U. S. Department of Energy Office of Energy Efficiency & Renewable Energy. (n.d.). Demand Response and Time-Variable Pricing Programs. https://www.energy.gov/eere/femp/demand-response-and-time-variable-pricing-programs

NOAA National Centers for Environmental information, Climate at a Glance: Nationwide and Statewide Time Series, Average Temperature, Average Precipitation, Z-Index Drought. published October 2022, retrieved on October 5th, 2022 from https://www.ndc.noaa.gov/cag/

NOAA Logo
https://commons.wikimedia.org/wiki/File:NOAA_logo.svg

EIA Logo
https://commons.wikimedia.org/wiki/File:Energy_Information_Administration_logo.svg