

6011 Coding Final Project

Andrew Shaw

2025-07-28

US Solar Capacity Growth: What drives adoption?

Andrew Shaw

Georgia Tech, Summer 2025

GitHub: github.com/shaw-a-d/6011_project

Abstract

This project investigates the extent to which state-level policies, political control, and market conditions explain the variation in solar capacity growth across the US from 2014 to 2023. The analysis features Renewable Portfolio Standards (RPS), state partisan control, electricity prices, and CO2 emissions.

Data were collected and compiled from a range of sources. Utility-level solar generator capacity, average retail prices, and CO2 emissions were obtained from the US Energy Information Administration (EIA, 2024; EIA-SEDS, 2025). RPS targets and solar carve-outs were taken from the Lawrence Berkeley National Laboratory 2023 status report (Barbose, 2023). Partisan control data was taken from the National Conference of State Legislatures (NCSL, 2023). These sources were cleaned and merged into state-year panel data to facilitate running several linear regression models to assess the relationships between the variables and the growth of state-level solar generation capacity. The goal is to better understand the drivers of solar adoption and inform future energy policy.

References

- Barbose, G. (2023). *U.S. Renewables Portfolio Standards: 2023 Annual Status Report*. Lawrence Berkeley National Laboratory. Retrieved from <https://emp.lbl.gov>
- National Conference of State Legislatures (NCSL). (2023). *State Partisan Composition: Governors and Legislatures (2014–2023)*. Retrieved from <https://www.ncsl.org/research/about-state-legislatures/partisan-composition>
- U.S. Energy Information Administration (EIA). (2024). *Form EIA-860 Detailed Data — Schedule 3: Generator Data (2023)*. Retrieved from <https://www.eia.gov/electricity/data/eia860/>
- U.S. Energy Information Administration (EIA). (2025). *State Energy Data System (SEDS): 1960–2023 (Complete)*. Release date: June 27, 2025. Retrieved from <https://www.eia.gov/state/seds/>

Background

Over the past decade, particularly starting in the early 2020s, the U.S. saw a marked increase in solar PV capacity alongside expanding Renewable Portfolio Standard (RPS) targets in many states. These trends signaled strong policy momentum and technological readiness for solar growth. However, that trajectory may now be at risk. In recent years, new political dynamics, regulatory rollbacks, and legal challenges in certain states have introduced uncertainty into the policy landscape.

These shifts raise questions about the long-term durability of state-level solar initiatives and whether recent gains can be sustained. This uncertainty is what motivated the current analysis: to explore how past policy and political factors have influenced solar growth.

While the data in this project are new, this type of work, integrating policy, economics, and data analysis, aligns with the broader goal of supporting businesses and organizations that aim to bring solar and other clean energy technologies to market. This project reflects both a learning process and a practical step toward applying economic and policy insight to real-world energy challenges.

Data

Cleaning

The challenges that arose from the various data mentioned in the midterm were accomplished including extracting and summarizing solar capacity data from multiple Excel files, reshaping wide policy data into long format, and standardizing state names and column structures across datasets. Of course, the data presented new challenges along the way

Notably, several states had non-binding RPS targets, which led to missing values in the `rps_target` column. To ensure the regression model could interpret this correctly, NA values for non-binding states were re-keyed with 0s as the NA value doesn't reflect an absence of data, rather a lack of policy. This would have affected the interpretability of models and analysis.

Additionally, because the data spans 50 states over 10 years, the expected panel size was 500 observations. But gaps and inconsistencies across the source datasets led to fewer rows during early cleaning. To address this and maintain the full panel structure, a complete state-year grid was created and all datasets were merged onto it. This ensured that missing values did not drop observations from the final data frame, allowing for consistent modeling and visualization.

Summary Statistics

Here is a table of summary statistics.

```
solar_panel_final <- readr::read_csv("../project_data/clean/solar_panel_final.csv")
solar_panel_final |>
  select(total_mw, rps_target, rps_active, solar_active, avg_retail_price_kwh, mil_metric_tons, state_control)
  summary()
```

```
##      total_mw      rps_target      rps_active      solar_active
## Min.   : 0.00   Min.   :0.00000   Min.   :0.000   Min.   :0.000
## 1st Qu.: 16.65   1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.000
## Median : 120.10   Median :0.06000   Median :1.000   Median :0.000
## Mean   : 805.70   Mean   :0.09866   Mean   :0.576   Mean   :0.224
## 3rd Qu.: 626.80   3rd Qu.:0.15000   3rd Qu.:1.000   3rd Qu.:0.000
## Max.   :18925.00   Max.   :0.63000   Max.   :1.000   Max.   :1.000
## avg_retail_price_kwh mil_metric_tons state_control
## Min.   : 7.130      Min.   : 4.40   Length:500
## 1st Qu.: 9.015      1st Qu.: 16.38   Class :character
## Median : 9.975      Median : 34.05   Mode  :character
## Mean   :11.590      Mean   : 45.40
## 3rd Qu.:12.210      3rd Qu.: 55.00
## Max.   :39.720      Max.   :338.10
```

The summary statistics offer a snapshot of the underlying structure of the panel data. Total solar capacity (`total_mw`) ranges from less than 1 MW to nearly 19,000 MW, with a median of only 120 mW suggesting a heavy skew towards only a few observations. RPS targets (`rps_target`) show a wide range across states, from 0 to 63%, while binary variables such as `rps_active` and `solar_active` confirm that many states had no binding renewable policies in place for part or all of the period.

The variables `rps_active` and `solar_active` were ultimately not used in most regression models, as they were highly collinear with `rps_target` and provided limited additional explanatory value (though they did prove useful for filtering purposes). Instead, the analysis focused on continuous and categorical predictors like RPS target percentage, electricity prices, emissions levels, and state political control.

Models

After running the initial model with all variables procured, the results were lacking. The data collected allowed for multiple ways of combining and allowing relationships to emerge more clearly.

Model 1

Base Policy and Market Controls

As mentioned, the binary predictors `solar_active` and `rps_active` were more noise than signal and were removed due to redundancy with the `rps_target` variable. The baseline for `state_control` was changed to “Rep” to better isolate the effects of Democratic or divided control.

```
model1 <- lm(total_mw ~ rps_target + avg_retail_price_kwh + mil_metric_tons + state_control, data = solar_panel_final)
summary(model1)
```

```
##
## Call:
## lm(formula = total_mw ~ rps_target + avg_retail_price_kwh + mil_metric_tons +
##     state_control, data = solar_panel_final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6393.5  -712.6   -86.0    411.9  12154.1
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1303.994    291.388  -4.475 9.49e-06 ***
## rps_target      2899.056    741.377   3.910 0.000105 ***
## avg_retail_price_kwh    85.362    16.341   5.224 2.59e-07 ***
## mil_metric_tons     24.065     1.353  17.792 < 2e-16 ***
## state_controlDivided  -433.150    201.626  -2.148 0.032177 *
## state_controlRep    -280.788    221.034  -1.270 0.204562
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1584 on 494 degrees of freedom
## Multiple R-squared:  0.4443, Adjusted R-squared:  0.4387
## F-statistic: 78.99 on 5 and 494 DF, p-value: < 2.2e-16
```

Binding RPS targets, electricity prices, and CO₂ emissions are all positively and significantly associated with total solar capacity. Political control is less relevant with the model reach and Adjusted R² of 0.4387.

Model 2

Binding RPS States

```
solar_rps_only <- solar_panel_final |>
  filter(rps_active ==1)
```

```
model_rps_only <- lm(total_mw ~ avg_retail_price_kwh + mil_metric_tons +
                     state_control + rps_target, data = solar_rps_only)

summary(model_rps_only)
```

```
##
## Call:
## lm(formula = total_mw ~ avg_retail_price_kwh + mil_metric_tons +
##     state_control + rps_target, data = solar_rps_only)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5881.8  -905.2  -214.8   548.7 10411.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -3239.871     493.447  -6.566 2.48e-10 ***
## avg_retail_price_kwh    222.708      32.583   6.835 5.06e-11 ***
## mil_metric_tons         25.673       1.784  14.388 < 2e-16 ***
## state_controlDivided  -180.563     252.685  -0.715 0.475463
## state_controlRep      -719.424     325.387  -2.211 0.027840 *
## rps_target          3801.745     1043.179   3.644 0.000319 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1859 on 282 degrees of freedom
## Multiple R-squared:  0.506, Adjusted R-squared:  0.4973
## F-statistic: 57.78 on 5 and 282 DF,  p-value: < 2.2e-16
```

This model sees a small improvement in the Adjusted R^2 . Democratic control also becomes statistically significant in this subset. This could also come from the subset itself in that Democrat-controlled states are more likely to adopt RPS policies to begin with.

Model 3

Fixed year effects

This model treats the year variable as a factor allowing for baseline comparison across years.

```
model_fixed <- lm(total_mw ~ rps_target + avg_retail_price_kwh + mil_metric_tons +
                  state_control + year, data = solar_panel_final)

summary(model_fixed)
```

```
##
## Call:
## lm(formula = total_mw ~ rps_target + avg_retail_price_kwh + mil_metric_tons +
##     state_control + year, data = solar_panel_final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -6024.8 -703.6 -84.4 433.3 11905.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -2.790e+05  5.052e+04  -5.523  5.41e-08 ***
## rps_target      2.045e+03  7.370e+02   2.774  0.00574 **
## avg_retail_price_kwh 7.021e+01  1.612e+01   4.357  1.61e-05 ***
## mil_metric_tons  2.431e+01  1.315e+00  18.486  < 2e-16 ***
## state_controlDivided -3.535e+02  1.965e+02  -1.800  0.07254 .
## state_controlRep    -3.484e+02  2.151e+02  -1.620  0.10593
## year            1.377e+02  2.505e+01   5.497  6.21e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1540 on 493 degrees of freedom
## Multiple R-squared:  0.4764, Adjusted R-squared:  0.47
## F-statistic: 74.75 on 6 and 493 DF, p-value: < 2.2e-16
```

This model sees an Adjusted R^2 of 0.4635 and shows a strong trend between 2021 through 2023 likely reflecting broader federal policies such as several infrastructure bills passed through the US Congress for funding and support for renewable energy projects.

Model Summary

Across all models three predictors stood out: - RPS targets: higher targets more associated with increased adoption - Electricity price: states with higher energy prices were more likely to see growth. - CO₂ emissions: could be a sign of fossil-fuel dependence or large state populations.

Visualization

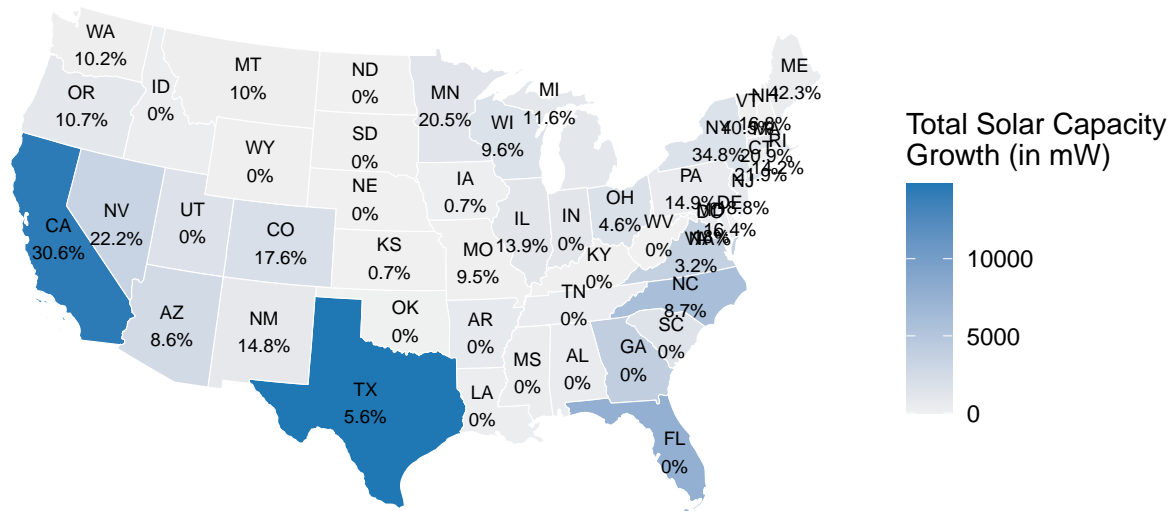
Two visualizations were created to help contextualize the results

US Choropleth Map

The maps displays solar capacity growth by state from 2014 to 2023 with each state also labeled by its average RPS target. California and Texas exhibit the higher solar capacity but have disparate RPS targets. This underscore the complexities of factors driving adoption.

```
states_merge_proj <- readRDS("../project_data/clean/states_merge_proj.rds")

ggplot(states_merge_proj) +
  geom_sf(aes(fill = mw_growth), color = "white") +
  geom_sf_text(aes(label = paste0(STUSPS, "\n", round(mean_rps*100, 1), "%")), size = 2.5,
               color = "black") +
  scale_fill_gradient(
    name = "Total Solar Capacity\nGrowth (in mW)",
    low = "#f0f0f0",
    high = "#1f78b4"
  ) +
  theme_minimal() +
  theme(
    axis.title = element_blank(),
    axis.text = element_blank(),
    axis.ticks = element_blank(),
    panel.grid = element_blank()
  )
```



```
labs(title = "Solar Growth vs. Avg RPS Target (2014-2023)")
```

```
## $title
## [1] "Solar Growth vs. Avg RPS Target (2014-2023)"
##
## attr(,"class")
## [1] "labels"
```

Scatterplot (Solar Growth v. Electricity Price)

The scatterplot shows a relatively flat relationship between electricity price and solar growth. However, the models indicate that electricity price being statistically significant. This suggests that electricity price on its own is not a strong standalone driver but in combination with other variables, it plays a more meaningful role in shaping solar adoption patterns.

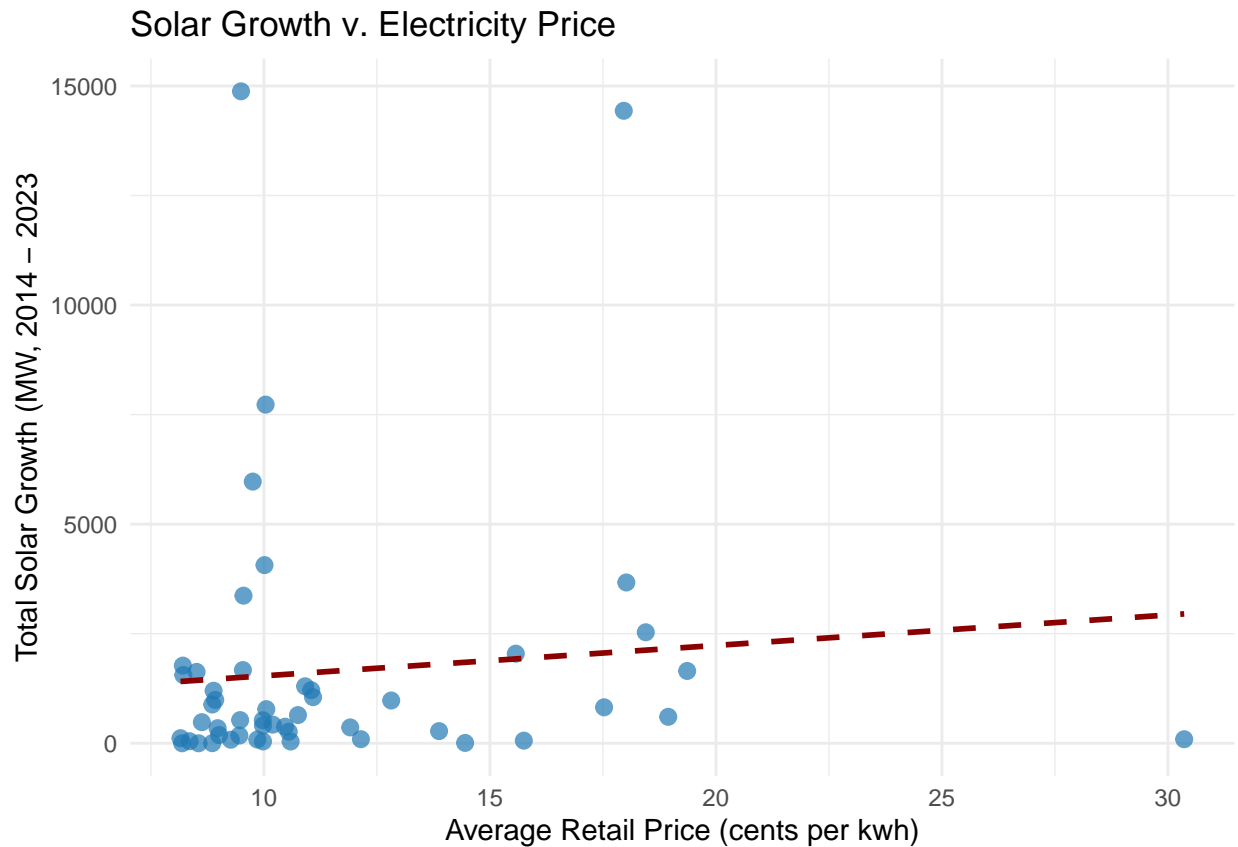
```
solar_growth <- readr::read_csv("../project_data/clean/solar_growth.csv")
```

```
## Rows: 50 Columns: 6
## -- Column specification -----
## Delimiter: ","
## chr (1): state
## dbl (5): mw_2014, mw_2023, mw_growth, mean_rps, avg_retail_price_kwh
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```



```
ggplot(solar_growth, aes(x = avg_retail_price_kwh, y = mw_growth)) +
  geom_point(color = "#1f78b4", size = 2.5, alpha = 0.7) +
  geom_smooth(method = "lm", se = FALSE, color = "darkred", linetype = "dashed") +
  labs(title = "Solar Growth v. Electricity Price",
       x = "Average Retail Price (cents per kwh)",
       y = "Total Solar Growth (MW, 2014 - 2023)" +
  theme_minimal()
```

'geom_smooth()' using formula = 'y ~ x'



Next Steps & Limitations

There are several areas where this project could be extended or improved with more time and exposure to R and other programs. While the choropleth map provided a useful geographic summary of solar growth, the process of working with shapefiles proved significantly more complex than expected. With more time and study visual maps could be more robust improving clarity and storytelling. There are also other potentially really valuable datasets that couldn't be included due to their complexity. The DSIRE database providing granular data on state-level incentives and regulations could add important nuance but the format provided (a large XML file) was too complex to parse. Similarly, solar irradiance data is abundant but learning geospatial data would require a deeper understanding and more time to learn those tools. That said for the bulk of this project, the libraries and packages encountered throughout the duration of the course met the needs of this project. Most of the data used required heavy use of `dyplyr` and `tidyr` to manipulate into panel data. The few packages that were explored were in an attempt to refine basic US maps into something more visually striking. However, time and data constraints prevented more shapely maps like Dorling cartograms or tilemaps. This project was a start and offered a valuable window into what kinds of data is available and what data is lacking. Future iterations of this project can include the harder-to-find data to help provide greater perspective on what drives positive solar and renewable energy policy outcomes.