Regional Differences in Renewable Energy and Electric Vehicle Adoption

Lab 4 Project Report

Overview

Electric vehicles (EVs) are often promoted as a solution to reduce carbon emissions from transportation. However, the environmental benefit of EVs depends critically on the source of electricity used to charge them. This report investigates **regional differences in renewable energy generation and build-out across U.S. states, and examines how these patterns relate to electric vehicle adoption and average energy costs**.

Specifically, this analysis addresses the following research questions:

- What are the regional differences in renewable energy generation across states?
- How has renewable energy build-out changed from 2021 to 2023?
- How do EV registrations relate to renewable energy usage and total energy consumption?
- What is the relationship between average energy prices and renewable energy adoption?

Understanding these patterns helps answer the broader question: **Does the electricity used to charge EVs actually come from clean sources?**

Data and Methods

Data Sources

This analysis uses four datasets from the U.S. Energy Information Administration covering 2021-2023:

1. **Renewable Energy Use by State** - Energy generation from biomass, geothermal, hydropower, solar, and wind sources
2. **Total Energy Use by State** - Total energy consumption across all sources (coal, natural gas, petroleum, nuclear, renewable)
3. **Average Energy Price by State** - Average energy prices in dollars per million BTU
4. **EV Registrations by State (2023)** - Number of registered electric vehicles by state

Data Cleaning and Preparation

```r
# Function to clean numeric values from messy text
clean_numeric <- function(x) {
  x |>
    str_remove_all("about|≈|~|\\$|USD|per MMBtu|est\\.|kWh|EVs|#") |>
```

```
    str_trim() |>
    as.numeric()
}

# 1. Load and clean energy price data
# This file has entire rows in quotes, so we need to read and separate
energy_price_raw <- read.csv("data/av-energy-price-2021-2023.csv", skip = 2,
header = TRUE)
energy_price <- energy_price_raw |>
  separate(col = 1, into = c("state", "price_2021", "price_2022",
"price_2023"), sep = ",") |>
  mutate(
    price_2021 = clean_numeric(price_2021),
    price_2022 = clean_numeric(price_2022),
    price_2023 = clean_numeric(price_2023)
  )

# 2. Load and clean EV registrations data
# Create a mapping of full state names to abbreviations
state_mapping <- c(
  "Alabama" = "AL", "Alaska" = "AK", "Arizona" = "AZ", "Arkansas" = "AR",
  "California" = "CA", "Colorado" = "CO", "Connecticut" = "CT", "Delaware" =
"DE",
  "Florida" = "FL", "Georgia" = "GA", "Hawaii" = "HI", "Idaho" = "ID",
  "Illinois" = "IL", "Indiana" = "IN", "Iowa" = "IA", "Kansas" = "KS",
  "Kentucky" = "KY", "Louisiana" = "LA", "Maine" = "ME", "Maryland" = "MD",
  "Massachusetts" = "MA", "Michigan" = "MI", "Minnesota" = "MN", "Mississippi"
= "MS",
  "Missouri" = "MO", "Montana" = "MT", "Nebraska" = "NE", "Nevada" = "NV",
  "New Hampshire" = "NH", "New Jersey" = "NJ", "New Mexico" = "NM", "New York"
= "NY",
  "North Carolina" = "NC", "North Dakota" = "ND", "Ohio" = "OH", "Oklahoma" =
"OK",
  "Oregon" = "OR", "Pennsylvania" = "PA", "Rhode Island" = "RI", "South
Carolina" = "SC",
  "South Dakota" = "SD", "Tennessee" = "TN", "Texas" = "TX", "Utah" = "UT",
  "Vermont" = "VT", "Virginia" = "VA", "Washington" = "WA", "West Virginia" =
"WV",
  "Wisconsin" = "WI", "Wyoming" = "WY", "District of Columbia" = "DC"
)

ev_registrations <- read.csv("data/ev-registrations-by-state-2023.csv", skip =
2, header = TRUE) |>
  rename(
    state_full = STATE,
    ev_count = Count.EVs
  ) |>
  mutate(
```

```
    ev_count = clean_numeric(ev_count),
    state = state_mapping[state_full],
    year = 2023
  ) |>
  select(state, ev_count, year) |>
  filter(!is.na(state))  # Remove any rows that don't match state names (like
"Total")

# 3. Load and clean renewable energy use data for 2021, 2022, 2023
renew_2021 <- read.csv("data/renew-use-2021.csv") |>
  mutate(
    Renewable_Use_2021 = clean_numeric(Renewable_Use_2021),
    year = 2021
  ) |>
  rename(renewable_use = Renewable_Use_2021)

renew_2022 <- read.csv("data/renew-use-2022.csv") |>
  mutate(
    Renewable_Use_2022 = clean_numeric(Renewable_Use_2022),
    year = 2022
  ) |>
  rename(renewable_use = Renewable_Use_2022)

renew_2023 <- read.csv("data/renew-use-2023.csv") |>
  mutate(
    Renewable_Use_2023 = clean_numeric(Renewable_Use_2023),
    year = 2023
  ) |>
  rename(renewable_use = Renewable_Use_2023)

# Combine renewable energy data
renewable_energy <- bind_rows(renew_2021, renew_2022, renew_2023)

# 4. Load and clean total energy use data for 2021, 2022, 2023
# These files have energy sources as rows and states as columns, so we need to
pivot

total_2021 <- read.csv("data/total-use-2021.csv") |>
  pivot_longer(
    cols = -Energy_Source,
    names_to = "state",
    values_to = "total_use"
  ) |>
  rename(energy_source = Energy_Source) |>
  mutate(
    total_use = as.numeric(total_use),
    year = 2021
  )
```

```r
total_2022 <- read.csv("data/total-use-2022.csv") |>
  pivot_longer(
    cols = -Energy_Source,
    names_to = "state",
    values_to = "total_use"
  ) |>
  rename(energy_source = Energy_Source) |>
  mutate(
    total_use = as.numeric(total_use),
    year = 2022
  )

total_2023 <- read.csv("data/total-use-2023.csv") |>
  pivot_longer(
    cols = -Energy_Source,
    names_to = "state",
    values_to = "total_use"
  ) |>
  rename(energy_source = Energy_Source) |>
  mutate(
    total_use = as.numeric(total_use),
    year = 2023
  )

# Combine total energy use data
total_energy <- bind_rows(total_2021, total_2022, total_2023)

# Display summary of cleaned datasets
cat("Energy Price Data:\n")
```

```
Energy Price Data:
```

```r
print(head(energy_price))
```

```
  state price_2021 price_2022 price_2023
1    AK      20.03      27.33      23.84
2    AL      17.85      23.37      21.11
3    AR      18.42      23.84      21.76
4    AZ      25.07      31.72      30.28
5    CA      28.44      37.35      35.72
6    CO      20.64      25.85      23.85
```

```r
cat("\nEV Registrations Data:\n")
```

```
EV Registrations Data:
```

```r
print(head(ev_registrations))
```

```
  state ev_count year
1    AL    13047 2023
2    AK     2697 2023
3    AZ    89798 2023
4    AR     7108 2023
5    CA  1256646 2023
6    CO    90083 2023
```

```r
cat("\nRenewable Energy Data:\n")
```

```
Renewable Energy Data:
```

```r
print(head(renewable_energy))
```

```
  State Energy_Source renewable_use year
1    AK       Biomass          3153 2021
2    AK    Geothermal            NA 2021
3    AK    Hydropower          5763 2021
4    AK  Solar Energy            45 2021
5    AK   Wind Energy           451 2021
6    AL       Biomass        198543 2021
```

```r
cat("\nTotal Energy Data:\n")
```

```
Total Energy Data:
```

```r
print(head(total_energy))
```

```
# A tibble: 6 × 4
  energy_source state total_use  year
  <chr>         <chr>     <dbl> <dbl>
1 Coal          AK        18694  2021
```

```
2 Coal          AL       309791  2021
3 Coal          AR       216123  2021
4 Coal          AZ       160299  2021
5 Coal          CA        28244  2021
6 Coal          CO       252442  2021
```

Data Integration and Key Variables

The following code joins the datasets and creates key analytical variables:

```
# First, let's aggregate the total energy data by state and year
# (sum across all energy sources to get total energy use per state)
total_energy_by_state <- total_energy |>
  group_by(state, year) |>
  summarise(total_energy_all_sources = sum(total_use, na.rm = TRUE), .groups =
"drop")

# Aggregate renewable energy data by state and year
# (sum across all renewable sources to get total renewable use per state)
renewable_energy_by_state <- renewable_energy |>
  group_by(State, year) |>
  summarise(total_renewable_use = sum(renewable_use, na.rm = TRUE), .groups =
"drop") |>
  rename(state = State)

# Join total energy and renewable energy data
energy_combined <- total_energy_by_state |>
  left_join(renewable_energy_by_state, by = c("state", "year")) |>
  mutate(
    # Calculate percentage of renewable energy out of total energy
    renewable_percentage = (total_renewable_use / total_energy_all_sources) *
100
  )

# Pivot energy price data to long format for easier joining
energy_price_long <- energy_price |>
  pivot_longer(
    cols = c(price_2021, price_2022, price_2023),
    names_to = "year_col",
    values_to = "avg_energy_price"
  ) |>
  mutate(year = as.numeric(str_extract(year_col, "\\d+"))) |>
  select(state, year, avg_energy_price)

# Join energy data with price data
energy_with_price <- energy_combined |>
  left_join(energy_price_long, by = c("state", "year"))
```

```
# For 2023 specifically, join with EV registrations data
# First create a 2023-specific energy dataset
energy_2023 <- energy_with_price |>
  filter(year == 2023)

# Join with EV registrations
energy_ev_2023 <- energy_2023 |>
  left_join(ev_registrations, by = c("state", "year")) |>
  mutate(
    # Calculate ratio of EV registrations to total energy use
    ev_per_energy_unit = ev_count / total_energy_all_sources,
    # Calculate EVs per 1000 energy units for easier interpretation
    ev_per_1000_energy = (ev_count / total_energy_all_sources) * 1000
  )

# Create a dataset showing renewable energy growth from 2021 to 2023
renewable_growth <- energy_with_price |>
  select(state, year, renewable_percentage, total_renewable_use) |>
  pivot_wider(
    names_from = year,
    values_from = c(renewable_percentage, total_renewable_use)
  ) |>
  mutate(
    # Calculate change in renewable percentage from 2021 to 2023
    renewable_pct_change = renewable_percentage_2023 -
renewable_percentage_2021,
    # Calculate percentage growth in absolute renewable use
    renewable_growth_pct = ((total_renewable_use_2023 -
total_renewable_use_2021) /
                             total_renewable_use_2021) * 100
  )

# Display summaries of the joined datasets for the report
```

Key Datasets Summary

Table 1: Combined Energy Data with Prices (2021-2023)

This dataset combines total energy use, renewable energy use, renewable percentage, and average prices by state and year.

state	year	total_energy_all_sources	total_renewable_use	renewable_percentage	avg_energy_price
AK	2021	684975	9412	1.37	20.03
AK	2022	730276	10410	1.43	27.33
AK	2023	746979	10088	1.35	23.84
AL	2021	2352656	239816	10.19	17.85

state	year	total_energy_all_sources	total_renewable_use	renewable_percentage	avg_energy_price
AL	2022	2337513	232035	9.93	23.37
AL	2023	2265008	222189	9.81	21.11
AR	2021	1136025	89714	7.90	18.42
AR	2022	1178115	90824	7.71	23.84
AR	2023	1151062	NA	NA	21.76
AZ	2021	1681257	99266	5.90	25.07

Table 2: Energy and EV Data for 2023

This dataset shows the relationship between EV registrations, renewable energy, and total energy consumption in 2023.

state	renewable_percentage	ev_count	ev_per_1000_energy	avg_energy_price
AK	1.35	2697	3.61	23.84
AL	9.81	13047	5.76	21.11
AR	NA	7108	6.18	21.76
AZ	NA	89798	52.43	30.28
CA	16.57	1256646	195.44	35.72
CO	8.06	90083	66.26	23.85
CT	NA	31557	39.96	32.32
DC	6.04	8066	174.13	32.28
DE	3.95	8435	41.45	26.70
FL	NA	254878	60.14	28.12

Table 3: Renewable Energy Growth (2021-2023)

This dataset tracks changes in renewable energy adoption over the three-year period.

state	renewable_percentage_2021	renewable_percentage_2023	renewable_pct_change	renewable_growth_pct
AK	1.37	1.35	−0.02	7.18
AL	10.19	9.81	−0.38	−7.35
AR	7.90	NA	NA	NA
AZ	5.90	NA	NA	NA
CA	13.19	16.57	3.38	31.50
CO	7.62	8.06	0.44	5.44

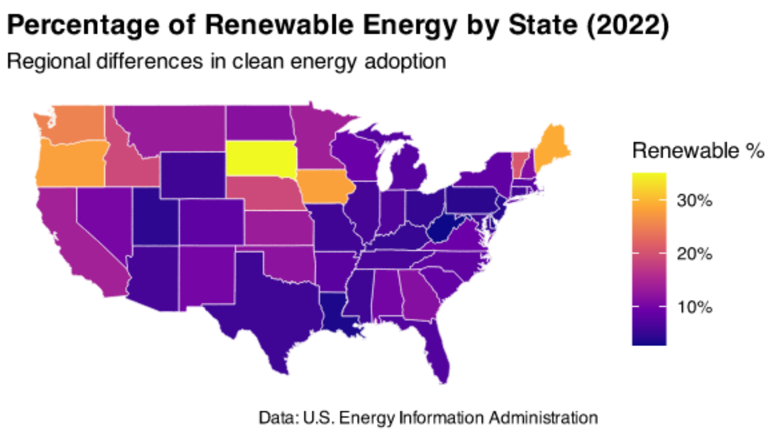
state	renewable_percentage_2021	renewable_percentage_2023	renewable_pct_change	renewable_growth_pct
CT	6.00	NA	NA	NA
DC	5.05	6.04	0.99	12.42
DE	3.44	3.95	0.51	12.43
FL	7.17	NA	NA	NA

Map Visualization

The following maps visualize key findings from the analysis:

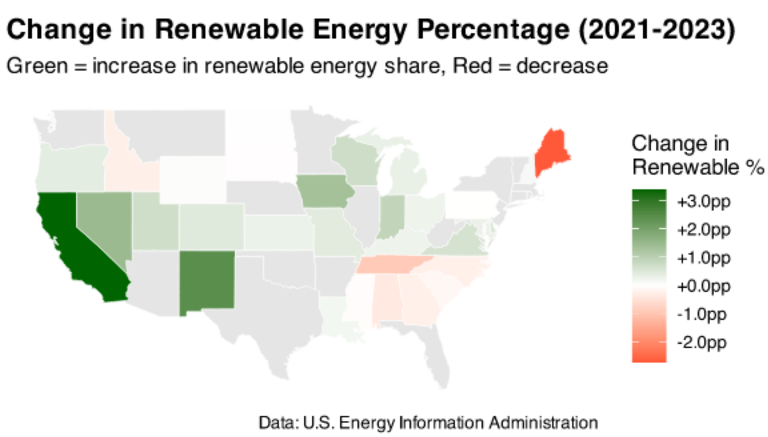
```r
# Load US states spatial data using sf
# Using the built-in US states data from the maps package via sf
us_states_sf <- st_as_sf(maps::map("state", plot = FALSE, fill = TRUE))

# Create mapping between state abbreviations and names
state_lookup <- data.frame(
  state = c(state.abb, "DC"),
  state_name = tolower(c(state.name, "district of columbia"))
)

# Map 1: Renewable Energy Percentage by State (2022)
# Using 2022 data directly from renewable_energy_by_state for complete
coverage
energy_2022 <- energy_with_price |>
  filter(year == 2022)

map_data_renewable_sf <- us_states_sf |>
  mutate(state_name = ID) |>
  left_join(state_lookup, by = "state_name") |>
  left_join(energy_2022, by = "state")

ggplot(data = map_data_renewable_sf) +
  geom_sf(aes(fill = renewable_percentage), color = "white", size = 0.2) +
  scale_fill_viridis_c(
    option = "plasma",
    name = "Renewable %",
    na.value = "grey90",
    labels = function(x) paste0(round(x, 1), "%")
  ) +
  labs(
    title = "Percentage of Renewable Energy by State (2022)",
    subtitle = "Regional differences in clean energy adoption",
    caption = "Data: U.S. Energy Information Administration"
  ) +
  theme_minimal() +
  theme(
```

```
    axis.text = element_blank(),
    axis.title = element_blank(),
    axis.ticks = element_blank(),
    panel.grid = element_blank(),
    plot.title = element_text(face = "bold", size = 14),
    legend.position = "right"
  )
```




```
# Map 2: Renewable Energy Growth (2021-2023)
map_data_growth_sf <- us_states_sf |>
  mutate(state_name = ID) |>
  left_join(state_lookup, by = "state_name") |>
  left_join(renewable_growth, by = "state")

ggplot(data = map_data_growth_sf) +
  geom_sf(aes(fill = renewable_pct_change), color = "white", size = 0.2) +
  scale_fill_gradient2(
    low = "red",
    mid = "white",
    high = "darkgreen",
    midpoint = 0,
    name = "Change in\nRenewable %",
    na.value = "grey90",
    labels = function(x) paste0(sprintf("%+.1f", x), "pp")
  ) +
  labs(
    title = "Change in Renewable Energy Percentage (2021-2023)",
    subtitle = "Green = increase in renewable energy share, Red = decrease",
    caption = "Data: U.S. Energy Information Administration"
  ) +
  theme_minimal() +
```

```r
  theme(
    axis.text = element_blank(),
    axis.title = element_blank(),
    axis.ticks = element_blank(),
    panel.grid = element_blank(),
    plot.title = element_text(face = "bold", size = 14),
    legend.position = "right"
  )
```

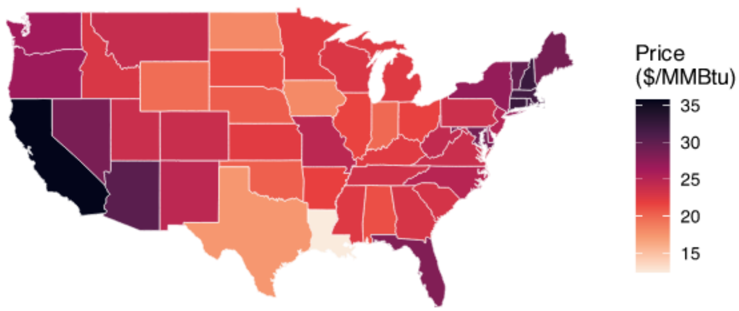



```r
# Map 3: Average Energy Price (2023)
map_data_price_sf <- us_states_sf |>
  mutate(state_name = ID) |>
  left_join(state_lookup, by = "state_name") |>
  left_join(energy_2023, by = "state")

ggplot(data = map_data_price_sf) +
  geom_sf(aes(fill = avg_energy_price), color = "white", size = 0.2) +
  scale_fill_viridis_c(
    option = "rocket",
    name = "Price\n($/MMBtu)",
    na.value = "grey90",
    direction = -1
  ) +
  labs(
    title = "Average Energy Price by State (2023)",
    subtitle = "Relationship between energy costs and renewable adoption",
    caption = "Data: U.S. Energy Information Administration"
  ) +
  theme_minimal() +
  theme(
    axis.text = element_blank(),
```

```
    axis.title = element_blank(),
    axis.ticks = element_blank(),
    panel.grid = element_blank(),
    plot.title = element_text(face = "bold", size = 14),
    legend.position = "right"
  )
```

Average Energy Price by State (2023)

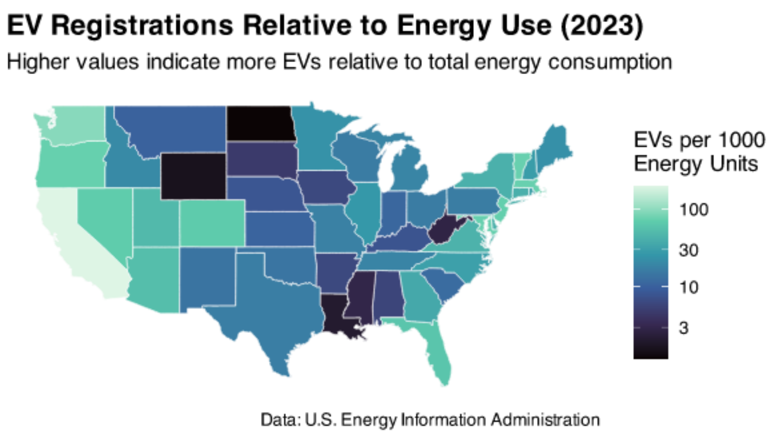
Relationship between energy costs and renewable adoption



Data: U.S. Energy Information Administration

```
# Map 4: EV Registrations per 1000 Energy Units (2023)
map_data_2023_sf <- us_states_sf |>
  mutate(state_name = ID) |>
  left_join(state_lookup, by = "state_name") |>
  left_join(energy_ev_2023, by = "state")

ggplot(data = map_data_2023_sf) +
  geom_sf(aes(fill = ev_per_1000_energy), color = "white", size = 0.2) +
  scale_fill_viridis_c(
    option = "mako",
    name = "EVs per 1000\nEnergy Units",
    na.value = "grey90",
    trans = "log10"
  ) +
  labs(
    title = "EV Registrations Relative to Energy Use (2023)",
    subtitle = "Higher values indicate more EVs relative to total energy
consumption",
    caption = "Data: U.S. Energy Information Administration"
  ) +
  theme_minimal() +
  theme(
    axis.text = element_blank(),
    axis.title = element_blank(),
```

```
    axis.ticks = element_blank(),
    panel.grid = element_blank(),
    plot.title = element_text(face = "bold", size = 14),
    legend.position = "right"
  )
```



Analysis

Key Findings

The maps and data reveal several important patterns about renewable energy adoption and EV registrations across the United States:

Regional Differences in Renewable Energy Generation

Map 1 highlights the percentage of renewable energy by state in 2022. Note that this is different than the 2023 data used for EVs, because it is more complete. States in the Pacific Northwest have high amounts of renewable energy due to hydroelectric. South Dakota and Iowa also have large amounts of wind. Some more liberal states seem to have higher percentages of renewable energy, although once below 20% it is pretty uniform across ideological lines.

Renewable Energy Build-Out Trends (2021-2023)

Map 2 shows the change in renewable energy's share of total electricity generation between 2021 and 2023. States like California, Iowa, New Mexico and Nevada lead the way. The latter two are due to an increase in solar, while Iowa is due to an increase in wind. Most states are close to neutral in the change in share of renewable energy. This is likely due to the fact that while energy consumption grows, fossil fuel energy sources like liquid natural gas are being built at roughly the same rate as new renewable sources.

EV Adoption and Energy Consumption

Map 3 shows the average energy cost per state. It will be helpful to relate this to Map 1 which has renewable energy share. This does not tell a pretty picture, as the data show that renewable energy share is positive correlated with energy cost. There is not necessarily causation, and some states break this pattern like Iowa, Florida, or Maine. This suggests that the relationship may have more to do with local energy markets and infrastructure cost or other environmental regulation than just share of renewable energy.

Energy Prices and Clean Energy

Map 4 shows the EVs per 1000 energy units across different states. It generally seems that states with more renewable energy (California, Washington, Vermont) have higher amounts of EV registrations per energy use. Interestingly, Iowa, a champion of wind, has quite a low EV registration per energy use. There is a noticeable but not significant association between EV registrations per energy use and renewable energy share.

Conclusion

The maps show that the benefit of EVs is dependent on the statewide energy grids. There is however a correlation between EV registrations per energy use and energy generation mix. It shows that the more EVs are used, the more likely it is that these EVs are using renewable energy from the grid. There are however confounding factors and states that break this pattern, which suggest that in some states renewable energy is going other places than EVs and in some states EVs are getting their energy from a fossil fuel dominated grid.

Something we did not consider heavily is how energy is shared between states, because renewable energy may be generated in one state and used in another. We also observe that energy from fossil fuels in the grid is generated much more efficiently than a gasoline engine, and so it is still on the net better to use an EV.