Do States with Cleaner Energy Have More EV Adoption?

Electric Vehicles and Renewable Energy: A State-Level Analysis

Overview

Main Research Question: *Electric vehicles reduce direct emissions, but does the electricity used to charge them actually come from clean sources?*

Chosen Sub-Question: *Are EV registrations concentrated in states with cleaner energy mixes? Specifically, do states with higher percentages of renewable energy in their total energy use show higher rates of EV adoption?*

This analysis explores the relationship between renewable energy usage and electric vehicle adoption across U.S. states in 2023. Understanding this relationship is crucial because EVs are only as "clean" as the electricity that powers them. If EVs are concentrated in states with coal-heavy energy grids, their environmental benefits are diminished.

Part 0: Load Libraries

```
library(tidyverse)
library(maps)
```

Part 2: Data Preparation and Cleaning

Load and Clean Renewable Energy Data (2021-2023)

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

# Load renewable energy data for all three years
renew_2021 <- read_csv("data/renew-use-2021.csv", show_col_types = FALSE) %>%
  mutate(State = toupper(State),  # Standardize state abbreviations
         Renewable_Use = clean_numeric(Renewable_Use_2021),
         Year = 2021) %>%
  select(State, Energy_Source, Renewable_Use, Year)
```

```r
renew_2022 <- read_csv("data/renew-use-2022.csv", show_col_types = FALSE) %>%
  mutate(State = toupper(State),  # Standardize state abbreviations
         Renewable_Use = clean_numeric(Renewable_Use_2022),
         Year = 2022) %>%
  select(State, Energy_Source, Renewable_Use, Year)

renew_2023 <- read_csv("data/renew-use-2023.csv", show_col_types = FALSE) %>%
  mutate(State = toupper(State),  # Standardize state abbreviations
         Renewable_Use = clean_numeric(Renewable_Use_2023),
         Year = 2023) %>%
  select(State, Energy_Source, Renewable_Use, Year)

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

# Calculate total renewable energy by state and year
renewable_total <- renewable_energy %>%
  group_by(State, Year) %>%
  summarize(Total_Renewable = sum(Renewable_Use, na.rm = TRUE), .groups =
"drop")

head(renewable_total)
```

```
# A tibble: 6 × 3
  State  Year Total_Renewable
  <chr> <dbl>           <dbl>
1 AK     2021            9598
2 AK     2022           10410
3 AK     2023           10088
4 AL     2021          239816
5 AL     2022          232035
6 AL     2023          222189
```

Load and Clean Total Energy Use Data (2021-2023)

```r
# Function to load and pivot total energy data
load_total_energy <- function(filepath, year) {
  read_csv(filepath, show_col_types = FALSE) %>%
    pivot_longer(cols = -Energy_Source,
                 names_to = "State",
                 values_to = "Energy_Use") %>%
    filter(State != "US") %>%  # Remove US total
    mutate(Year = year,
           Energy_Use = as.numeric(Energy_Use))
}
```

```r
# Load all three years
total_2021 <- load_total_energy("data/total-use-2021.csv", 2021)
total_2022 <- load_total_energy("data/total-use-2022.csv", 2022)
total_2023 <- load_total_energy("data/total-use-2023.csv", 2023)

# Combine and calculate total energy by state and year
total_energy <- bind_rows(total_2021, total_2022, total_2023) %>%
  group_by(State, Year) %>%
  summarize(Total_Energy = sum(Energy_Use, na.rm = TRUE), .groups = "drop")

head(total_energy)
```

```
# A tibble: 6 × 3
  State  Year Total_Energy
  <chr> <dbl>        <dbl>
1 AK     2021       684975
2 AK     2022       730276
3 AK     2023       746979
4 AL     2021      2352656
5 AL     2022      2337513
6 AL     2023      2265008
```

Load and Clean Energy Price Data

```r
# Read energy price data - needs special handling due to quoted rows
energy_price_raw <- read_lines("data/av-energy-price-2021-2023.csv")

# Remove quotes and parse manually
energy_price <- energy_price_raw[4:length(energy_price_raw)] %>%  # Skip first
3 rows
  str_remove_all('"') %>%  # Remove quotes
  str_split_fixed(",", n = 4) %>%  # Split by comma into 4 columns
  as.data.frame() %>%
  rename(State = V1, Price_2021 = V2, Price_2022 = V3, Price_2023 = V4) %>%
  mutate(State = toupper(State)) %>%  # Standardize state abbreviations
  pivot_longer(cols = starts_with("Price_"),
               names_to = "Year",
               values_to = "Avg_Price") %>%
  mutate(Year = as.numeric(str_extract(Year, "\\d{4}")),
         Avg_Price = clean_numeric(Avg_Price))

head(energy_price)
```

```
# A tibble: 6 × 3
  State  Year Avg_Price
  <chr> <dbl>     <dbl>
```

```
1 AK     2021      20.0
2 AK     2022      27.3
3 AK     2023      23.8
4 AL     2021      17.8
5 AL     2022      23.4
6 AL     2023      21.1
```

Load and Clean EV Registration Data (2023)

```r
# Create state name to abbreviation lookup
state_lookup <- tibble(
  State_Full = state.name,
  State = state.abb
) %>%
  # Add DC manually
  bind_rows(tibble(State_Full = "District of Columbia", State = "DC"))

# Read EV registration data (skip first 2 header rows)
ev_registrations <- read_csv("data/ev-registrations-by-state-2023.csv",
                             skip = 2,
                             show_col_types = FALSE) %>%
  rename(State_Full = 1, EV_Count = 2) %>%
  filter(State_Full != "Total") %>%  # Remove total row
  mutate(EV_Count = clean_numeric(EV_Count)) %>%
  # Join with state lookup to get abbreviations
  left_join(state_lookup, by = "State_Full") %>%
  select(State, EV_Count)

head(ev_registrations)
```

```
# A tibble: 6 × 2
  State EV_Count
  <chr>    <dbl>
1 AL       13047
2 AK        2697
3 AZ       89798
4 AR        7108
5 CA     1256646
6 CO       90083
```

Part 3: Joining / Pivoting Datasets for Analysis

Create Comprehensive Dataset for 2023 Analysis

```r
# Join renewable and total energy for 2023
energy_2023 <- renewable_total %>%
```

```r
  filter(Year == 2023) %>%
  left_join(total_energy %>% filter(Year == 2023),
            by = c("State", "Year")) %>%
  left_join(energy_price %>% filter(Year == 2023),
            by = c("State", "Year"))

# Calculate renewable percentage
energy_2023 <- energy_2023 %>%
  mutate(Renewable_Pct = (Total_Renewable / Total_Energy) * 100)

# Join with EV registration data
analysis_data <- energy_2023 %>%
  left_join(ev_registrations, by = "State") %>%
  mutate(EV_per_Energy = (EV_Count / Total_Energy) * 1000) %>%  # EVs per 1000
BTU
  filter(!is.na(Total_Energy))  # Remove rows with missing energy data

head(analysis_data, 10)
```

```
# A tibble: 10 × 8
   State  Year Total_Renewable Total_Energy Avg_Price Renewable_Pct EV_Count
   <chr> <dbl>           <dbl>        <dbl>     <dbl>         <dbl>    <dbl>
 1 AK     2023           10088       746979      23.8          1.35     2697
 2 AL     2023          222189      2265008      21.1          9.81    13047
 3 AR     2023           87277      1151062      21.8          7.58     7108
 4 AZ     2023          108445      1712667      30.3          6.33    89798
 5 CA     2023         1065179      6429818      35.7         16.6   1256646
 6 CO     2023          115062      1359507      23.8          8.46    90083
 7 CT     2023           48983       789642      32.3          6.20    31557
 8 DC     2023            2796        46323      32.3          6.04     8066
 9 DE     2023            8040       203487      26.7          3.95     8435
10 FL     2023          286307      4237858      28.1          6.76   254878
# ℹ 1 more variable: EV_per_Energy <dbl>
```

Summary Statistics

```r
analysis_data %>%
  summarize(
    Mean_Renewable_Pct = mean(Renewable_Pct, na.rm = TRUE),
    Median_Renewable_Pct = median(Renewable_Pct, na.rm = TRUE),
    Mean_EV_Count = mean(EV_Count, na.rm = TRUE),
    Median_EV_Count = median(EV_Count, na.rm = TRUE),
    Total_EVs = sum(EV_Count, na.rm = TRUE)
  )
```

```
# A tibble: 1 × 5
  Mean_Renewable_Pct Median_Renewable_Pct Mean_EV_Count Median_EV_Count
               <dbl>                <dbl>         <dbl>           <dbl>
1               10.3                 7.96        69715.           25565
# ℹ 1 more variable: Total_EVs <dbl>
```

Time Series Analysis (2021-2023)

```
# Create time series dataset
time_series <- renewable_total %>%
  left_join(total_energy, by = c("State", "Year")) %>%
  mutate(Renewable_Pct = (Total_Renewable / Total_Energy) * 100)

# Calculate change in renewable percentage from 2021 to 2023
renewable_change <- time_series %>%
  filter(Year %in% c(2021, 2023)) %>%
  select(State, Year, Renewable_Pct) %>%
  pivot_wider(names_from = Year,
              values_from = Renewable_Pct,
              names_prefix = "Year_") %>%
  mutate(Renewable_Change = Year_2023 - Year_2021) %>%
  arrange(desc(Renewable_Change))

head(renewable_change, 10)
```

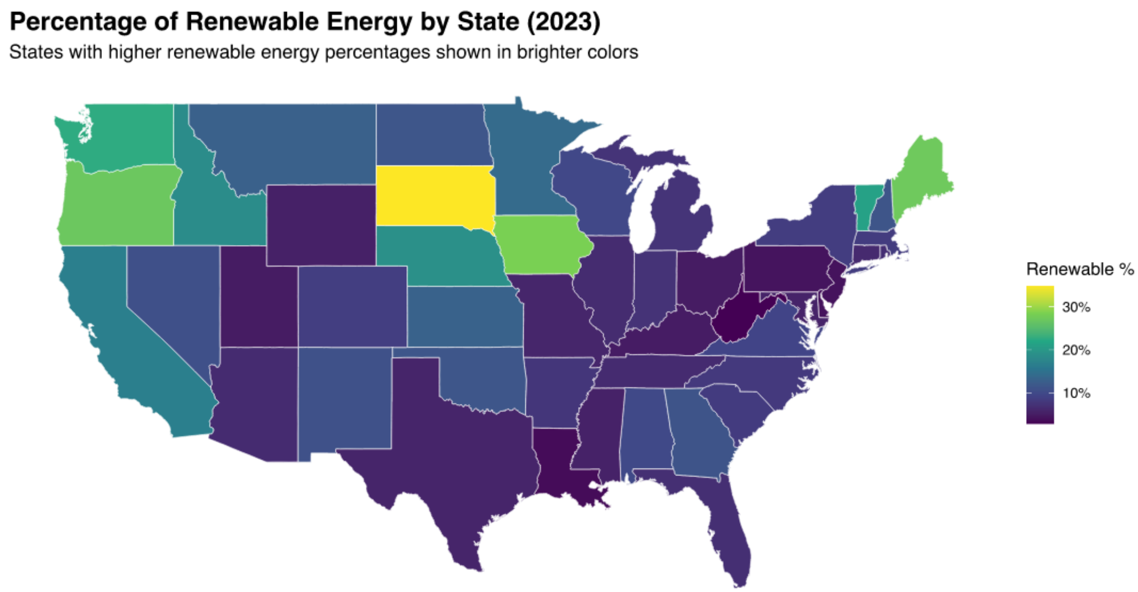
```
# A tibble: 10 × 4
   State Year_2021 Year_2023 Renewable_Change
   <chr>     <dbl>     <dbl>            <dbl>
 1 CA        13.2      16.6             3.38
 2 NM         8.36     10.8             2.43
 3 NV         9.21     10.7             1.46
 4 IA        27.0      28.3             1.32
 5 VT        19.9      21.1             1.12
 6 DC         5.05      6.04            0.987
 7 IN         6.31      7.19            0.880
 8 CO         7.62      8.46            0.843
 9 TX         4.81      5.63            0.817
10 IL         5.63      6.39            0.767
```

Part 4: Mapping Visualization

Map 1: Renewable Energy Percentage by State (2023)

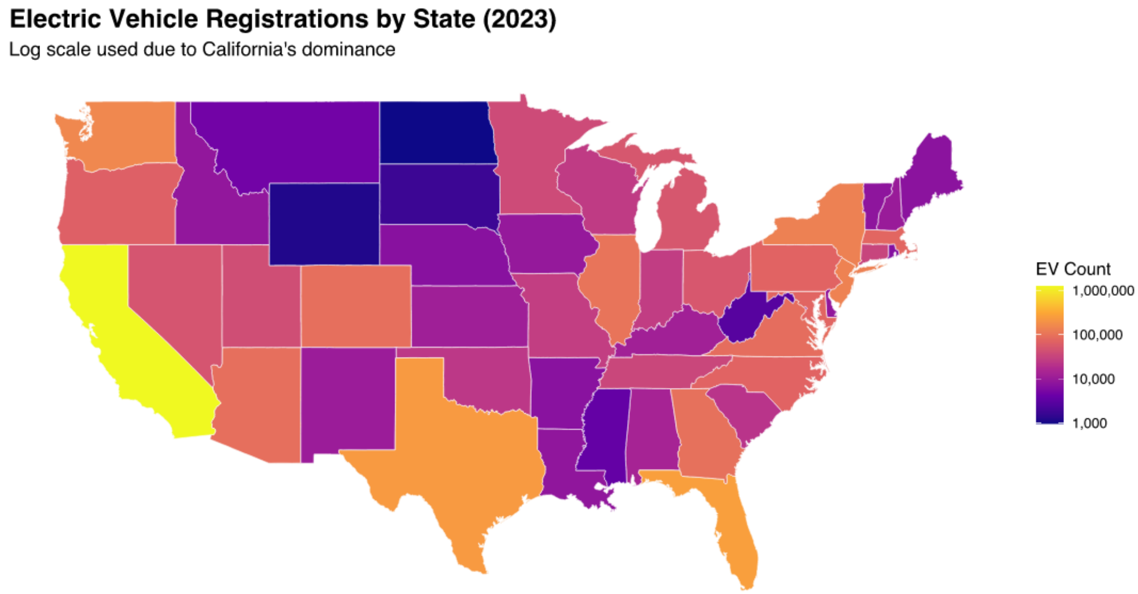
```
# Prepare map data
us_states <- map_data("state") %>%
  rename(state_lower = region)
```

```r
# Create mapping from state abbreviations to full names (lowercase)
state_name_lookup <- tibble(
  State = state.abb,
  state_lower = tolower(state.name)
) %>%
  bind_rows(tibble(State = "DC", state_lower = "district of columbia"))

# Prepare analysis data for mapping
map_data_renewable <- analysis_data %>%
  left_join(state_name_lookup, by = "State")

# Join map and data
us_map_renewable <- us_states %>%
  left_join(map_data_renewable, by = "state_lower")

# Create choropleth map
ggplot(us_map_renewable, aes(x = long, y = lat, group = group, fill =
Renewable_Pct)) +
  geom_polygon(color = "white", linewidth = 0.2) +
  coord_fixed(1.3) +
  scale_fill_viridis_c(
    option = "viridis",
    name = "Renewable %",
    na.value = "grey90",
    labels = function(x) paste0(round(x, 1), "%")
  ) +
  theme_minimal() +
  theme(
    legend.position = "right",
    plot.title = element_text(size = 16, face = "bold"),
    plot.subtitle = element_text(size = 12),
    axis.text = element_blank(),
    axis.title = element_blank(),
    panel.grid = element_blank()
  ) +
  labs(
    title = "Percentage of Renewable Energy by State (2023)",
    subtitle = "States with higher renewable energy percentages shown in
brighter colors"
  )
```



Map 2: EV Registrations by State (2023)

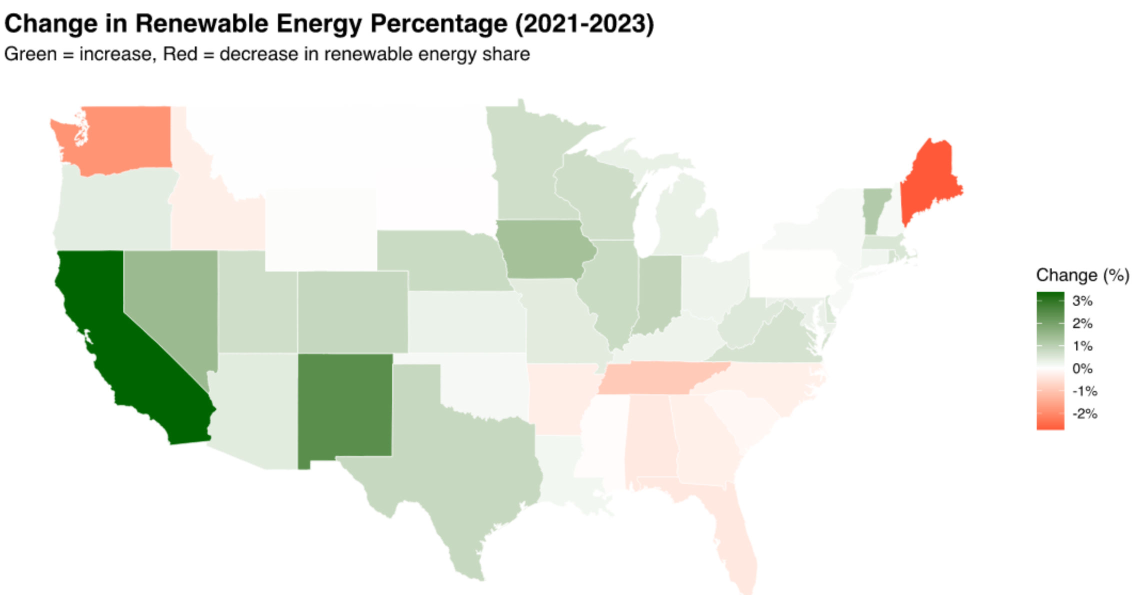
```r
# Create EV registration map
ggplot(us_map_renewable, aes(x = long, y = lat, group = group, fill =
EV_Count)) +
  geom_polygon(color = "white", linewidth = 0.2) +
  coord_fixed(1.3) +
  scale_fill_viridis_c(
    option = "plasma",
    name = "EV Count",
    na.value = "grey90",
    trans = "log10",
    labels = scales::comma
  ) +
  theme_minimal() +
  theme(
    legend.position = "right",
    plot.title = element_text(size = 16, face = "bold"),
    plot.subtitle = element_text(size = 12),
    axis.text = element_blank(),
    axis.title = element_blank(),
    panel.grid = element_blank()
  ) +
  labs(
    title = "Electric Vehicle Registrations by State (2023)",
    subtitle = "Log scale used due to California's dominance"
  )
```



Map 3: Change in Renewable Energy (2021-2023)

```r
# Prepare change data for mapping
map_data_change <- renewable_change %>%
  left_join(state_name_lookup, by = "State")

us_map_change <- us_states %>%
  left_join(map_data_change, by = "state_lower")

# Create change map
ggplot(us_map_change, aes(x = long, y = lat, group = group, fill =
Renewable_Change)) +
  geom_polygon(color = "white", linewidth = 0.2) +
  coord_fixed(1.3) +
  scale_fill_gradient2(
    low = "red",
    mid = "white",
    high = "darkgreen",
    midpoint = 0,
    name = "Change (%)",
    na.value = "grey90",
    labels = function(x) paste0(round(x, 1), "%")
  ) +
  theme_minimal() +
  theme(
    legend.position = "right",
    plot.title = element_text(size = 16, face = "bold"),
```

```
    plot.subtitle = element_text(size = 12),
    axis.text = element_blank(),
    axis.title = element_blank(),
    panel.grid = element_blank()
  ) +
  labs(
    title = "Change in Renewable Energy Percentage (2021-2023)",
    subtitle = "Green = increase, Red = decrease in renewable energy share"
  )
```

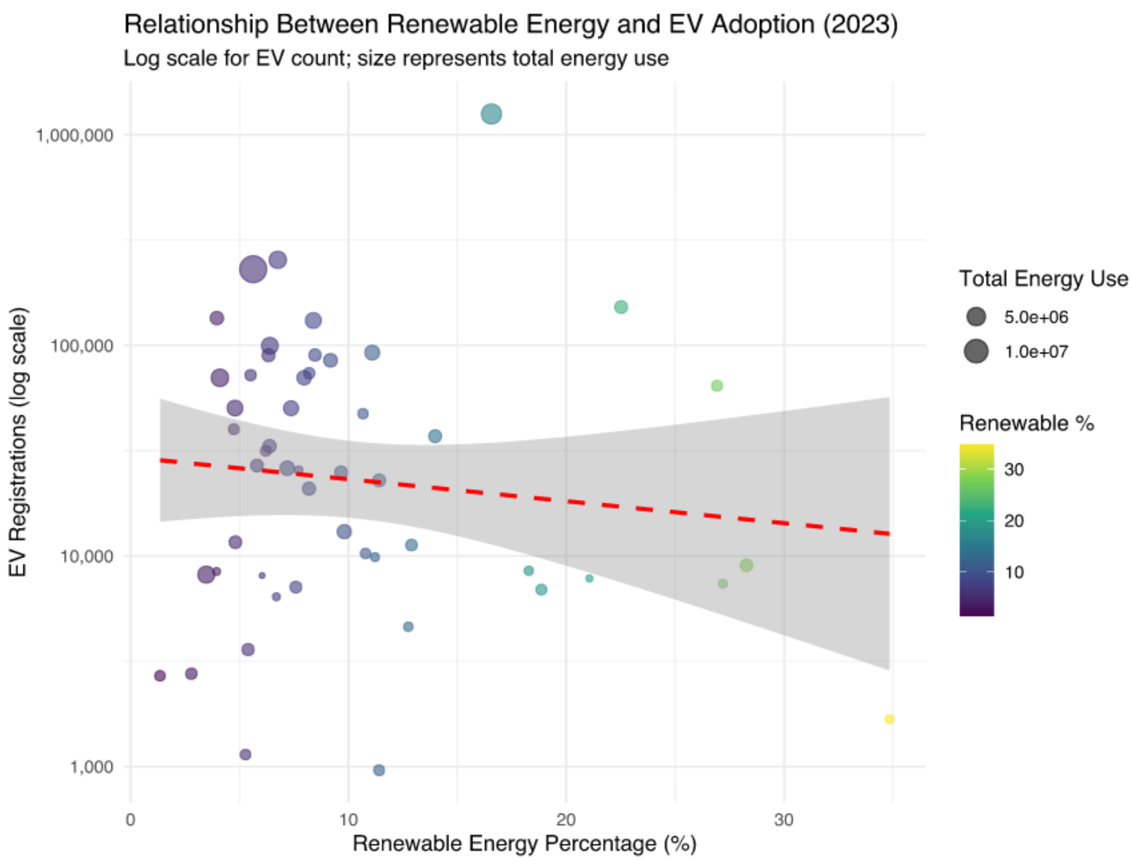


Scatter Plot: Renewable Energy vs EV Adoption

```
# Create scatter plot with state labels for top states
top_ev_states <- analysis_data %>%
  arrange(desc(EV_Count)) %>%
  head(10)

ggplot(analysis_data, aes(x = Renewable_Pct, y = EV_Count)) +
  geom_point(aes(size = Total_Energy, color = Renewable_Pct), alpha = 0.6) +
  geom_smooth(method = "lm", se = TRUE, color = "red", linetype = "dashed") +
  scale_y_log10(labels = scales::comma) +
  scale_color_viridis_c(option = "viridis", name = "Renewable %") +
  scale_size_continuous(name = "Total Energy Use") +
  theme_minimal() +
  labs(
    title = "Relationship Between Renewable Energy and EV Adoption (2023)",
```

```r
    subtitle = "Log scale for EV count; size represents total energy use",
    x = "Renewable Energy Percentage (%)",
    y = "EV Registrations (log scale)"
  ) +
  theme(legend.position = "right")
```

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



Analysis

Key Findings

1. Geographic Patterns in Renewable Energy

The first map reveals significant geographic variation in renewable energy usage. Western states, particularly those in the Pacific Northwest and Mountain West, show the highest percentages of renewable energy. This is largely due to abundant hydroelectric power (Washington, Oregon, Idaho) and wind resources (Montana, Wyoming). In contrast, many Midwestern and Southern states remain heavily dependent on fossil fuels, showing lower renewable percentages.

2. EV Adoption Concentration

The second map demonstrates that EV adoption is extremely concentrated. California dominates with over 1.2 million EVs (requiring a log scale for visualization), followed by Florida, Texas, and Washington. This concentration reflects a combination of factors including state incentives, charging infrastructure, and population density rather than purely energy mix.

3. Limited Correlation Between Clean Energy and EV Adoption

The scatter plot reveals a **weak positive relationship** between renewable energy percentage and EV adoption. While some high-renewable states like Washington show high EV numbers, the relationship is not strong. California, despite having a moderate renewable percentage (~8%), has by far the most EVs due to aggressive policy support and population size.

4. Progress in Renewable Energy (2021-2023)

The change map shows that many states made progress in increasing their renewable energy share between 2021 and 2023, particularly in the Midwest and South. This suggests the energy grid is gradually becoming cleaner, which bodes well for the environmental impact of EVs.

Implications for the Main Research Question

Does the electricity used to charge EVs actually come from clean sources?

The answer is **mixed and geography-dependent**:

- **Best case**: EV owners in Washington, Oregon, Idaho, and Montana are charging with 20-40% renewable electricity, making their vehicles significantly cleaner than gas cars.
- **Moderate case**: States like California, despite moderate renewable percentages, have strong EV adoption. These EVs are cleaner than gas cars but still rely substantially on natural gas.
- **Challenging case**: Many Southern and Midwestern states with coal-heavy grids have relatively few EVs, but those that exist are being charged with mostly fossil fuel electricity, reducing their environmental benefits.

Recommendations

1. **Policy Alignment**: States should coordinate EV incentives with renewable energy investments to maximize environmental benefits.
2. **Grid Improvement**: As renewable energy infrastructure improves (as shown in the 2021-2023 change map), the existing EV fleet automatically becomes cleaner.
3. **Regional Strategies**: EV adoption strategies should account for regional energy mixes—states with cleaner grids should receive priority for EV infrastructure investments to maximize emission reductions.

The maps clearly show that the environmental benefits of EV adoption are not uniform across the country and depend critically on the source of electricity in each state.