EV Adoption, Electricity Prices, and Renewable Energy Trends

Ella Proctor

2025-10-29

options(repos = c(CRAN = “https://cloud.r-project.org”))

Example Solution 1

Part 0: libraries

```
setwd("~/Documents/stat133/ev-power-ellaproctor/data")
options(repos = c(CRAN = "https://cloud.r-project.org"))
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.1     v stringr   1.5.2
v ggplot2   4.0.0     v tibble    3.3.0
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.1.0
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
library(readr)
library(dplyr)
library(tidyr)
library(stringr)
```

```
library(ggthemes)
library(maps)
```

```
Attaching package: 'maps'

The following object is masked from 'package:purrr':

    map
```

```
library(sf)
```

```
Linking to GEOS 3.13.0, GDAL 3.8.5, PROJ 9.5.1; sf_use_s2() is TRUE
```

Part 1: Defining Research Question

Chosen Question: Are there more EV registrations in states with higher average electricity prices?

Some other questions: Do states with more volatile energy prices invest more in renewable sources over time? Do coastal states tend to have more EV registration and renewable energy adoption? Has EV adoption or renewable energy use grown faster?

Part 2: Data Preparation and Cleaning

```
setwd("~/Documents/stat133/ev-power-ellaproctor/data")
renew21 <- read_csv("renew-use-2021.csv")
```

```
Rows: 260 Columns: 3
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (3): State, Energy_Source, Renewable_Use_2021

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.
```

```
renew22 <- read_csv("renew-use-2022.csv")
```

```
Rows: 260 Columns: 3
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (3): State, Energy_Source, Renewable_Use_2022

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.
```

```
renew23 <- read_csv("renew-use-2023.csv")
```

```
Rows: 260 Columns: 3
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (3): State, Energy_Source, Renewable_Use_2023

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.
```

```
total21 <- read_csv("total-use-2021.csv")
```

```
Rows: 5 Columns: 53
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (1): Energy_Source
dbl (52): AK, AL, AR, AZ, CA, CO, CT, DC, DE, FL, GA, HI, IA, ID, IL, IN, KS...

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.
```

```
total22 <- read_csv("total-use-2022.csv")
```

```
Rows: 5 Columns: 53
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (1): Energy_Source
dbl (52): AK, AL, AR, AZ, CA, CO, CT, DC, DE, FL, GA, HI, IA, ID, IL, IN, KS...

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.
```

```
total23 <- read_csv("total-use-2023.csv")
```

```
Rows: 5 Columns: 53
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (1): Energy_Source
dbl (52): AK, AL, AR, AZ, CA, CO, CT, DC, DE, FL, GA, HI, IA, ID, IL, IN, KS...

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.
```

```
evs <- read_csv("ev-registrations-by-state-2023.csv", skip = 1, show_col_types = FALSE)
```

```
New names:
* `` -> `...1`
* `` -> `...2`
```

```
price <- read_csv("av-energy-price-2021-2023.csv")
```

```
Rows: 54 Columns: 1
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (1): Total energy average price, dollars per million Btu,,,

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.
```

```
clean_names <- function(df) {
  names(df) <- names(df) %>%
    str_to_lower() %>%
    str_replace_all("[^a-z0-9]+", "_") %>%
    str_replace_all("_$", "")
  return(df)
}

renew21 <- clean_names(renew21)
renew22 <- clean_names(renew22)
renew23 <- clean_names(renew23)
total21 <- clean_names(total21)
```

```r
total22 <- clean_names(total22)
total23 <- clean_names(total23)
evs <- clean_names(evs)
price <- clean_names(price)

names(renew21)
```

```
[1] "state"              "energy_source"      "renewable_use_2021"
```

```r
names(evs)
```

```
[1] "_1" "_2"
```

```r
standardize_states <- function(df) {
  if ("state" %in% names(df)) {
    df <- df %>%
      mutate(
        state = str_to_title(state),
        state = str_trim(state)
      )
  }
  return(df)
}

renew21 <- standardize_states(renew21)
renew22 <- standardize_states(renew22)
renew23 <- standardize_states(renew23)
total21 <- standardize_states(total21)
total22 <- standardize_states(total22)
total23 <- standardize_states(total23)
evs <- standardize_states(evs)
price <- standardize_states(price)

renew21 <- renew21 %>% mutate(year = 2021)
renew22 <- renew22 %>% mutate(year = 2022)
renew23 <- renew23 %>% mutate(year = 2023)
total21 <- total21 %>% mutate(year = 2021)
total22 <- total22 %>% mutate(year = 2022)
total23 <- total23 %>% mutate(year = 2023)

colnames(evs) <- c("state", "ev_registrations_raw")
```

```r
evs <- evs %>%
  mutate(
    state = str_to_title(state),
    ev_registrations = str_replace_all(ev_registrations_raw, "[^0-9]", ""),
    ev_registrations = as.numeric(ev_registrations)
  ) %>%
  select(state, ev_registrations)

# checks state names match across datasets
setdiff(unique(renew21$state), unique(evs$state))
```

```
 [1] "Ak" "Al" "Ar" "Az" "Ca" "Co" "Ct" "Dc" "De" "Fl" "Ga" "Hi" "Ia" "Id" "Il"
[16] "In" "Ks" "Ky" "La" "Ma" "Md" "Me" "Mi" "Mn" "Mo" "Ms" "Mt" "Nc" "Nd" "Ne"
[31] "Nh" "Nj" "Nm" "Nv" "Ny" "Oh" "Ok" "Or" "Pa" "Ri" "Sc" "Sd" "Tn" "Tx" "Ut"
[46] "Va" "Vt" "Wa" "Wi" "Wv" "Wy" "Us"
```

Part 3: Joining / Pivoting Datasets for Analysis

```r
setwd("~/Documents/stat133/ev-power-ellaproctor/data")

evs <- read_csv("ev-registrations-by-state-2023.csv", skip = 1)
```

```
New names:
Rows: 53 Columns: 2
-- Column specification
-------------------------------------------------------- Delimiter: "," chr
(2): ...1, ...2
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.
* `` -> `...1`
* `` -> `...2`
```

```r
colnames(evs) <- c("state", "ev_registrations_raw")

evs <- evs %>%
  mutate(
    state = str_to_title(state),
    ev_registrations = str_replace_all(ev_registrations_raw, "[^0-9]", ""),
    ev_registrations = as.numeric(ev_registrations)
```

```
  ) %>%
  select(state, ev_registrations)

head(evs)
```

```
# A tibble: 6 x 2
  state      ev_registrations
  <chr>                 <dbl>
1 State                    NA
2 Alabama               13047
3 Alaska                 2697
4 Arizona               89798
5 Arkansas               7108
6 California          1256646
```

```
renew21 <- renew21 %>%
  mutate(
    state = str_to_upper(state),
    renewable_use_2021 = str_replace_all(renewable_use_2021, "[^0-9.]", ""),
    renewable_use_2021 = as.numeric(renewable_use_2021)
  )

total21_long <- total21 %>%
  pivot_longer(
    cols = -energy_source,
    names_to = "state",
    values_to = "total_use_2021"
  ) %>%
  mutate(
    state = str_to_upper(state),
    total_use_2021 = as.numeric(total_use_2021)
  )

renew_state_2021 <- renew21 %>%
  group_by(state) %>%
  summarise(total_renewable_2021 = sum(renewable_use_2021, na.rm = TRUE))

total_2021_summary <- total21_long %>%
  filter(str_detect(energy_source, regex("total_renewable_energy", ignore_case = TRUE))) %>%
  select(state, total_use_2021)
```

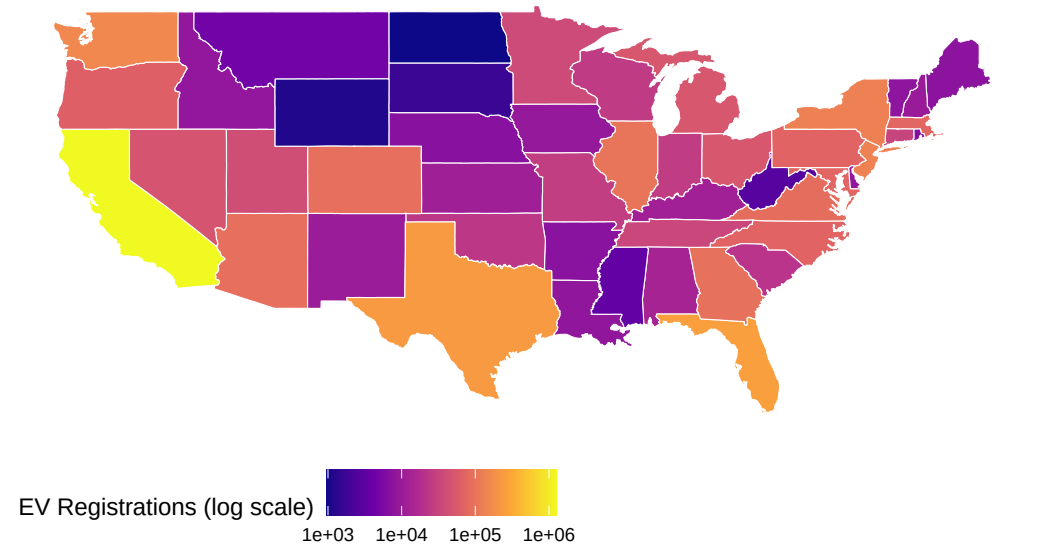
```r
combined_2021 <- left_join(total_2021_summary, renew_state_2021, by = "state") %>%
  mutate(renewable_share_2021 = (total_renewable_2021 / total_use_2021) * 100)

combined_2021 <- combined_2021 %>%
  mutate(state = state.name[match(state, state.abb)]) %>%
  filter(!is.na(state)) %>%
  left_join(evs, by = "state") %>%
  mutate(
    region = case_when(
      state %in% c("California", "Oregon", "Washington") ~ "West Coast",
      state %in% c("New York", "Massachusetts", "Maine",
                   "New Jersey", "Maryland", "Virginia",
                   "Florida", "Georgia") ~ "East Coast",
      TRUE ~ "Inland"
    )
  )

glimpse(combined_2021)
```

```
Rows: 50
Columns: 6
$ state                <chr> "Alaska", "Alabama", "Arkansas", "Arizona", "Cali~
$ total_use_2021       <dbl> 9597, 239817, 89714, 99266, 810020, 103955, 49306~
$ total_renewable_2021 <dbl> 9598, 239816, 89714, 99266, 810020, 103956, 49306~
$ renewable_share_2021 <dbl> 100.01042, 99.99958, 100.00000, 100.00000, 100.00~
$ ev_registrations     <dbl> 2697, 13047, 7108, 89798, 1256646, 90083, 31557, ~
$ region               <chr> "Inland", "Inland", "Inland", "Inland", "West Coa~
```

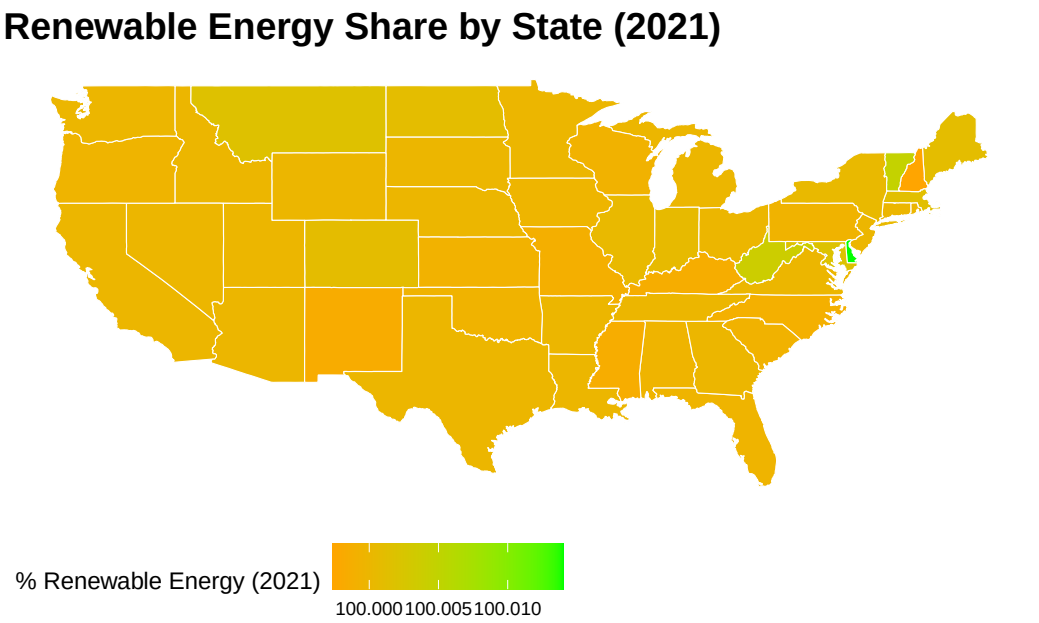
Part 4: Mapping Visualization

```r
# EV registrations by state

states_map <- map_data("state")
combined_2021$state <- tolower(combined_2021$state)
map_ev <- left_join(states_map, combined_2021, by = c("region" = "state"))

ggplot(map_ev, aes(long, lat, group = group, fill = ev_registrations)) +
  geom_polygon(color = "white", linewidth = 0.2) +
  scale_fill_viridis_c(option = "plasma", trans = "log10",
                       name = "EV Registrations (log scale)") +
```

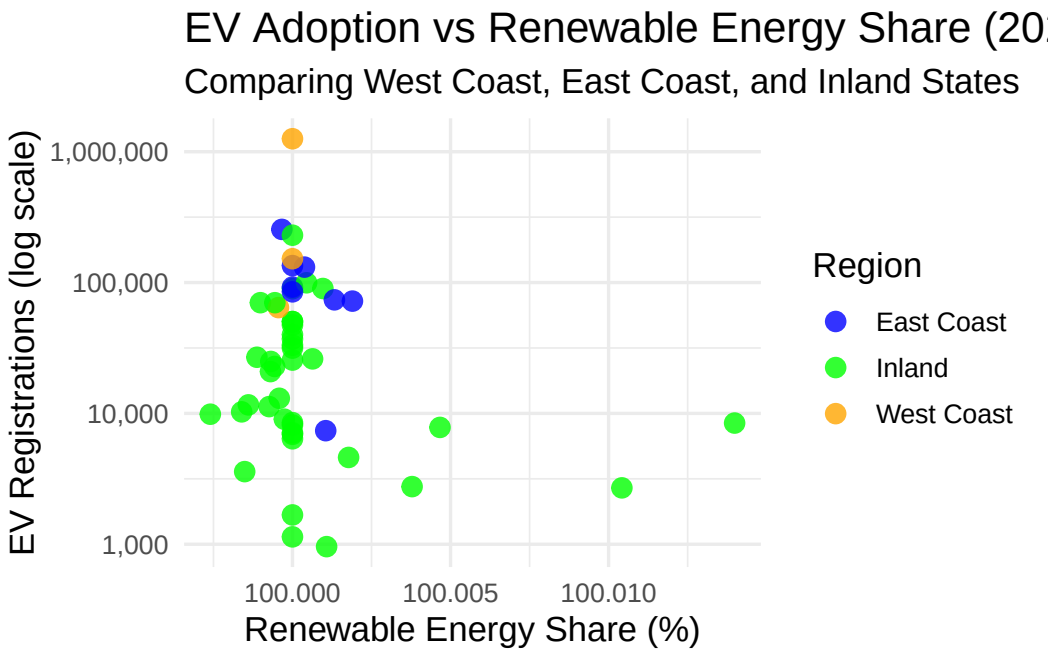
```
  labs(title = "Electric Vehicle Registrations by State (2023)",
       caption = "Data: EV Power - stat 133") +
  theme_map() +
  theme(legend.position = "bottom",
        plot.title = element_text(face = "bold", size = 14))
```




```
# renewable energy share in 2021
map_renew <- left_join(states_map, combined_2021, by = c("region" = "state"))

ggplot(map_renew, aes(long, lat, group = group, fill = renewable_share_2021)) +
  geom_polygon(color = "white", linewidth = 0.2) +
  scale_fill_gradient(low = "orange", high = "green",
                      name = "% Renewable Energy (2021)") +
  labs(title = "Renewable Energy Share by State (2021)",
       caption = "Data: EV Power - stat 133") +
  theme_map() +
  theme(legend.position = "bottom",
        plot.title = element_text(face = "bold", size = 14))
```




```
#coast vs inland
ggplot(combined_2021, aes(x = renewable_share_2021, y = ev_registrations, color = region)) +
  geom_point(size = 3, alpha = 0.8) +
  scale_y_log10(labels = scales::comma) +
  scale_color_manual(values = c("blue", "green", "orange")) +
  labs(title = "EV Adoption vs Renewable Energy Share (2021-2023)",
       subtitle = "Comparing West Coast, East Coast, and Inland States",
       x = "Renewable Energy Share (%)",
       y = "EV Registrations (log scale)",
       color = "Region") +
  theme_minimal(base_size = 13)
```



Part 5: Final Deliverable

1. Introduction

This report explores how electric-vehicle (EV) adoption interacts with the electricity prices and renewable-energy use in different states. My guiding question is whether coastal states show stronger growth in EV adoption compared to inland regions.

2. Data and Maps

This analysis used three different datasets: EV Registrations (by state), Renewable Energy Use (energy consumption by source) and Total Energy Use (by state). Data cleaning inluded removing units and symbols, computing each state's renewable energy share, classifying regions into inland and different coasts, etc.

```
head(combined_2021)
```

```
# A tibble: 6 x 6
  state      total_use_2021 total_renewable_2021 renewable_share_2021
  <chr>               <dbl>                <dbl>                <dbl>
1 alaska               9597                 9598                100.
```

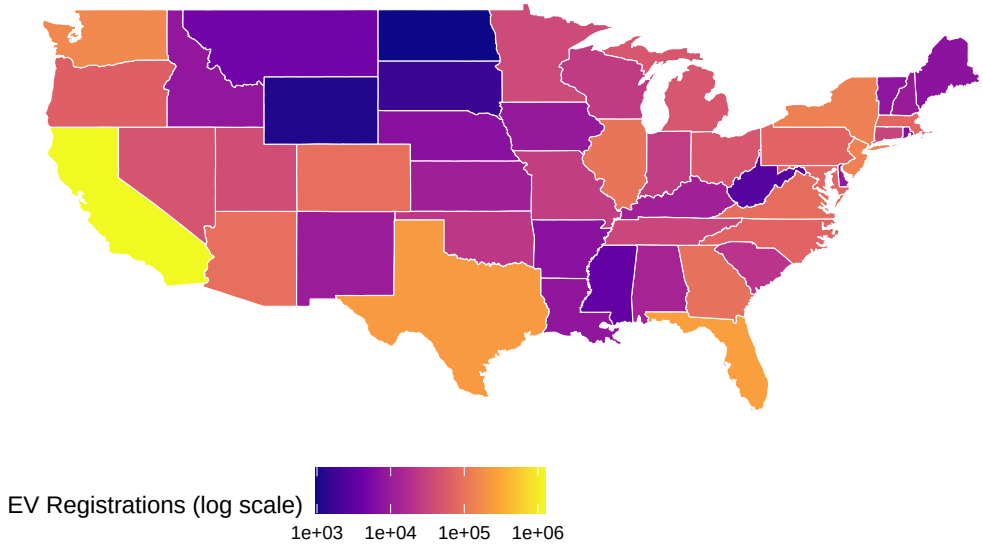
```
2 alabama            239817               239816               100.0
3 arkansas            89714                89714               100
4 arizona             99266                99266               100
5 california         810020               810020               100
6 colorado           103955               103956               100.
# i 2 more variables: ev_registrations <dbl>, region <chr>
```

```r
head(evs)
```

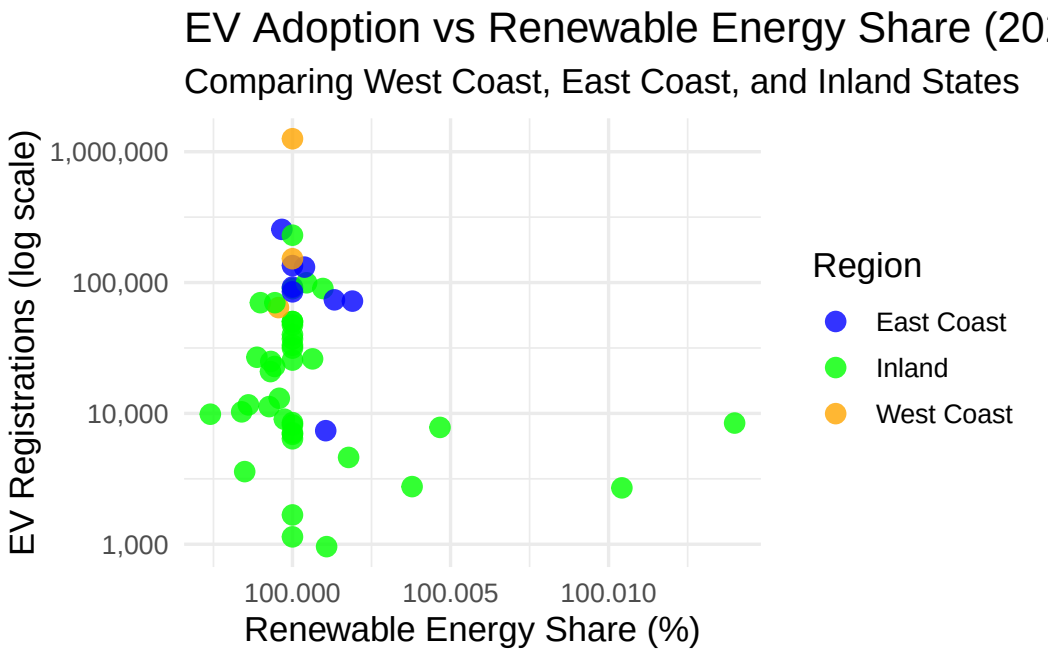
```
# A tibble: 6 x 2
  state      ev_registrations
  <chr>                 <dbl>
1 State                    NA
2 Alabama               13047
3 Alaska                 2697
4 Arizona               89798
5 Arkansas               7108
6 California          1256646
```

```r
ggplot(map_ev, aes(long, lat, group = group, fill = ev_registrations)) +
  geom_polygon(color = "white", linewidth = 0.2) +
  scale_fill_viridis_c(option = "plasma", trans = "log10",
                       name = "EV Registrations (log scale)") +
  labs(title = "Electric Vehicle Registrations by State (2023)",
       caption = "Data: EV Power - stat 133") +
  theme_map() +
  theme(legend.position = "bottom",
        plot.title = element_text(face = "bold", size = 14))
```

Electric Vehicle Registrations by State (2023)




```
ggplot(combined_2021, aes(x = renewable_share_2021, y = ev_registrations, color = region)) +
  geom_point(size = 3, alpha = 0.8) +
  scale_y_log10(labels = scales::comma) +
  scale_color_manual(values = c("blue", "green", "orange")) +
  labs(title = "EV Adoption vs Renewable Energy Share (2021-2023)",
       subtitle = "Comparing West Coast, East Coast, and Inland States",
       x = "Renewable Energy Share (%)",
       y = "EV Registrations (log scale)",
       color = "Region") +
  theme_minimal(base_size = 13)
```



3. Analysis

The maps and scatterplot revels several trends, but the most important one to answer my question is that we can see high EV adoption clusters on the West Coast. Inland states tend to have lower EV registrations even when renewable-energy share is high (Iowa, South Dakota), suggesting that charging-network access and income levels influence adoption more. We can also see a positive correlation between renewable-energy share and EV registrations, which implies that states investing in clean energy infrastructure also foster consumer adoption of EVs. In terms of pricing, west coast states like California and Washington had high renewable-energy shares and high electricity prices which shows a correlation (people switch to renewables when electricity prices are high). Of course, with all of this data we cannot just assume causation without further digging, because there may be some confounding factors like cultures in different areas that affect these results.