Analysis of U.S. Electricity Prices and Renewable Energy Trends (2021–2023)

Overview

This project investigates how U.S. states are transitioning toward cleaner energy and how that relates to electricity costs from 2021 to 2023. The analysis focuses on three questions:

1. How does the share of renewable electricity generation vary across U.S. states in 2023?
2. Which U.S. states experienced the fastest growth in renewable energy use from 2021 to 2023?
3. How do average electricity prices compare across U.S. states from 2021 to 2023?

To answer these questions, multiple raw datasets were loaded and cleaned, including state-level electricity generation by energy source (2021–2023), total electricity use (2023), and average electricity price by state (2021–2023). Columns were renamed for consistency, units and symbols were removed from numeric values, and state names were standardized so they could be joined across files.

After cleaning, the data were summarized by state. For renewable energy, total renewable electricity use in 2023 was divided by total electricity use in 2023 to calculate each state's renewable share. Growth in renewable energy was then measured by comparing total renewable output in 2021 vs. 2023 and computing both absolute and percentage change. For electricity prices, prices from 2021–2023 were reshaped from wide to long format, missing values were replaced with 0 where needed, and a three-year average price was calculated for each state.

Finally, these summarized tables were joined with U.S. state map geometry and visualized as choropleth maps.

Part 0: libraries

```r
library(tidyverse)
```

```
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0
──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.1     ✔ stringr   1.5.2
✔ ggplot2   4.0.0     ✔ tibble    3.3.0
✔ lubridate 1.9.4     ✔ tidyr     1.3.1
✔ purrr     1.1.0
── Conflicts ──────────────────────────────────────────── tidyverse_conflicts()
──
✖ dplyr::filter() masks stats::filter()
```

```
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```

```
library(readr)
library(dplyr)
library(ggplot2)
library(maps   )
```

```


Attaching package: 'maps'

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

    map
```

Part 1: Defining Research Question

Question 1: How does the share of renewable electricity generation vary across U.S. states in 2023?
Question 2: Which U.S. states experienced the fastest growth in renewable energy use from 2021 to 2023?
Question 3: How do average electricity prices across U.S. states compare from 2021 to 2023?

Part 2: Data Preparation and Cleaning

Question 1 Data Cleaning: How does the share of renewable electricity generation vary across U.S. states in 2023?

```
renew_2023 <- read_csv("data/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.
```

```
total_2023 <- read_csv("data/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.
```

```
#renew-use-2023.csv data cleaning
colnames(renew_2023) <- c("state", "energy_source", "renewable_use_2023")

renew_2023 <- renew_2023 |>
    mutate(state = str_to_upper(state))

renew_2023 <- renew_2023 |>
    mutate(renewable_use_2023 = str_remove_all(renewable_use_2023, "kWh|MWh|\
\s"),
           renewable_use_2023 = as.numeric(renewable_use_2023))

head(renew_2023)
```

```
# A tibble: 6 × 3
  state energy_source renewable_use_2023
  <chr> <chr>                      <dbl>
1 AK    Biomass                     3404
2 AK    Geothermal                   186
3 AK    Hydropower                  6051
4 AK    Solar Energy                  67
5 AK    Wind Energy                  380
6 AL    Biomass                   189040
```

```
#total-use-2023.csv data cleaning
total_2023 <- total_2023 |>
    pivot_longer(cols = -Energy_Source, names_to = "state", values_to =
"total_use_2023")

total_2023 <- total_2023 |>
    mutate(state = str_to_upper(state),
           total_use_2023 = as.numeric(total_use_2023),
           energy_source = str_to_lower(Energy_Source)) |>
    select(state, energy_source, total_use_2023)

head(total_2023)
```

```
# A tibble: 6 × 3
  state energy_source total_use_2023
  <chr> <chr>                  <dbl>
1 AK    coal_usage             18414
2 AL    coal_usage            224926
3 AR    coal_usage            180262
4 AZ    coal_usage            137885
5 CA    coal_usage             28746
6 CO    coal_usage            204826
```

Question 2 Data Cleaning: Which U.S. states experienced the fastest growth in renewable energy use from 2021 to 2023?

renew_2023 has already been data cleaned in the previous step.

```
renew_2021 <- read_csv("data/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.
```

```
renew_2022 <- read_csv("data/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.
```

```
#renew-use-2021.csv data cleaning
renew_2021 <- renew_2021[, 1:3]
colnames(renew_2021) <- c("state", "energy_source", "renewable_use_2021")

renew_2021 <- renew_2021 |>
```

```
    mutate(renewable_use_2021 = str_remove_all(renewable_use_2021,"kWh|MWh|
MMBtu|USD|\\$|~|est\\.|about|per|\\s"),
           renewable_use_2021 = as.numeric(renewable_use_2021)) |>
           filter(!is.na(renewable_use_2021))
```

```
Warning: There was 1 warning in `mutate()`.
i In argument: `renewable_use_2021 = as.numeric(renewable_use_2021)`.
Caused by warning:
! NAs introduced by coercion
```

```
head(renew_2021)
```

```
# A tibble: 6 × 3
  state energy_source renewable_use_2021
  <chr> <chr>                      <dbl>
1 AK    Geothermal                   186
2 AK    Hydropower                  5763
3 AK    Solar Energy                  45
4 AK    Wind Energy                  451
5 AL    Biomass                   198543
6 AL    Geothermal                   141
```

```
#renew-use-2022.csv data cleaning
renew_2022 <- renew_2022[, 1:3]
colnames(renew_2022) <- c("state", "energy_source", "renewable_use_2022")

renew_2022 <- renew_2022 |>
  mutate(renewable_use_2022 = str_remove_all(renewable_use_2022,"kWh|MWh|
MMBtu|USD|\\$|~|est\\.|about|per|\\s"),
         renewable_use_2022 = as.numeric(renewable_use_2022)) |>
  filter(!is.na(renewable_use_2022))
```

```
Warning: There was 1 warning in `mutate()`.
i In argument: `renewable_use_2022 = as.numeric(renewable_use_2022)`.
Caused by warning:
! NAs introduced by coercion
```

```
head(renew_2022)
```

```
# A tibble: 6 × 3
  state energy_source renewable_use_2022
  <chr> <chr>                      <dbl>
```

```
1 AK    Geothermal                    186
2 AK    Hydropower                   5846
3 AK    Solar Energy                   57
4 AK    Wind Energy                   475
5 AL    Biomass                    193932
6 AL    Geothermal                    141
```

<u>Question 3 Data Cleaning:</u> How do average electricity prices across U.S. states compare from 2021 t0 2023?

```r
avg_price2021_2023 <- read_csv("data/av-energy-price-2021-2023.csv", col_names
= FALSE)
```

```
Rows: 55 Columns: 1
── Column specification
────────────────────────────────────────────────────────
Delimiter: ","
chr (1): X1

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.
```

```r
#av-energy-price-2021-2023.csv data clean

price_clean <- avg_price2021_2023 |>
    separate(X1, into = c("state", "price_2021", "price_2022", "price_2023"),
sep = ",") |>
    mutate(price_2021 = str_remove_all(price_2021, "USD|per|MMBtu|about|est\
\.|\\$|\\s"),
           price_2022 = str_remove_all(price_2022, "USD|per|MMBtu|about|est\
\.|\\$|\\s"),
           price_2023 = str_remove_all(price_2023, "USD|per|MMBtu|about|est\
\.|\\$|\\s"),
           price_2021 = as.numeric(price_2021),
           price_2022 = as.numeric(price_2022),
           price_2023 = as.numeric(price_2023))
```

```
Warning: Expected 4 pieces. Additional pieces discarded in 1 rows [1].
```

```
Warning: There were 3 warnings in `mutate()`.
The first warning was:
ℹ In argument: `price_2021 = as.numeric(price_2021)`.
Caused by warning:
```

```
! NAs introduced by coercion
i Run `dplyr::last_dplyr_warnings()` to see the 2 remaining warnings.
```

```r
head(price_clean)
```

```
# A tibble: 6 × 4
  state                        price_2021 price_2022 price_2023
  <chr>                             <dbl>      <dbl>      <dbl>
1 "Total energy average price"       NA         NA         NA  
2 ""                                 NA         NA         NA  
3 "State"                          2021       2022       2023  
4 "AK"                               20.0       27.3       23.8
5 "AL"                               17.8       23.4       NA  
6 "AR"                               18.4       23.8       21.8
```

Part 3: Joining / Pivoting Datasets for Analysis

Question 1 Joining/Pivoting: How does the share of renewable electricity generation vary across U.S. states in 2023?
Step 1) Summarize renewable electricity use by state
Step 2) Summarize total electricity use by state
Step 3) Join the two summaries to calculate renewable share (2023)

```r
renew_2023 <- renew_2023 |> mutate(state = toupper(state))
total_2023 <- total_2023 |> mutate(state = toupper(state))

renew_state_2023 <- renew_2023 |>
  group_by(state) |>
  summarise(total_renewable_2023 = sum(renewable_use_2023, na.rm = TRUE))

total_state_2023 <- total_2023 |>
  group_by(state) |>
  summarise(total_energy_2023 = sum(total_use_2023, na.rm = TRUE))

energy_joined_2023 <- left_join(renew_state_2023, total_state_2023, by =
"state") |>
  mutate(renewable_share_2023 = total_renewable_2023 / total_energy_2023 *
100)

energy_joined_2023
```

```
# A tibble: 52 × 4
   state total_renewable_2023 total_energy_2023 renewable_share_2023
   <chr>                <dbl>             <dbl>                <dbl>
 1 AK                   10088            746979                 1.35
```

```
 2 AL                   222189           2265008                9.81
 3 AR                    87277           1151062                7.58
 4 AZ                   108445           1712667                6.33
 5 CA                  1065179           6429818               16.6
 6 CO                   115062           1359507                8.46
 7 CT                    48983            789642                6.20
 8 DC                     2796             46323                6.04
 9 DE                     8040            203487                3.95
10 FL                   286307           4237858                6.76
# ℹ 42 more rows
```

Question 2 Joining/Pivoting: Which U.S. states experienced the fastest growth in renewable energy use from 2021 to 2023?
Step 1) Summarize each year's data
Step 2) Join all three years together
Step 3) Calculate growth
Step 4) Identify top states with fastest growth

```
renew_2021_state <- renew_2021 |>
    group_by(state) |>
    summarise(total_renew_2021 = sum(renewable_use_2021, na.rm = TRUE))

renew_2022_state <- renew_2022 |>
    group_by(state) |>
    summarise(total_renew_2022 = sum(renewable_use_2022, na.rm = TRUE))

renew_2023_state <- renew_2023 |>
    group_by(state) |>
    summarise(total_renew_2023 = sum(renewable_use_2023, na.rm = TRUE))
```

```
renew_all <- renew_2021_state |>
    left_join(renew_2022_state, by = "state") |>
    left_join(renew_2023_state, by = "state")

renew_all
```

```
# A tibble: 52 × 4
  state total_renew_2021 total_renew_2022 total_renew_2023
  <chr>            <dbl>            <dbl>            <dbl>
1 AK                6445             6564            10088
2 AL              239816           232035           222189
3 AR               89714            90824            87277
4 AZ               99266           101214           108445
5 CA              759940           820793          1065179
6 CO              103956            57217           115062
```

```
 7 CT                 49262            7112            48983
 8 DC                  2487            2623             2796
 9 DE                  7151            7402             8040
10 FL                 75405           85594           286307
# ℹ 42 more rows
```

```
renew_all <- renew_all |>
    mutate(growth_21_23 = total_renew_2023 - total_renew_2021,
    pct_growth_21_23 = (growth_21_23 / total_renew_2021) * 100)

renew_all
```

```
# A tibble: 52 × 6
   state total_renew_2021 total_renew_2022 total_renew_2023 growth_21_23
   <chr>            <dbl>            <dbl>            <dbl>        <dbl>
 1 AK                6445             6564            10088         3643
 2 AL              239816           232035           222189       -17627
 3 AR               89714            90824            87277        -2437
 4 AZ               99266           101214           108445         9179
 5 CA              759940           820793          1065179       305239
 6 CO              103956            57217           115062        11106
 7 CT               49262             7112            48983         -279
 8 DC                2487             2623             2796          309
 9 DE                7151             7402             8040          889
10 FL               75405            85594           286307       210902
# ℹ 42 more rows
# ℹ 1 more variable: pct_growth_21_23 <dbl>
```

```
top_growth <- renew_all |>
  arrange(desc(pct_growth_21_23))

head(top_growth)
```

```
# A tibble: 6 × 6
  state total_renew_2021 total_renew_2022 total_renew_2023 growth_21_23
  <chr>            <dbl>            <dbl>            <dbl>        <dbl>
1 MS                2478            66614            67304        64826
2 MA               19750            80701            81560        61810
3 FL               75405            85594           286307       210902
4 AK                6445             6564            10088         3643
5 CA              759940           820793          1065179       305239
6 NM               62209            59153            80278        18069
# ℹ 1 more variable: pct_growth_21_23 <dbl>
```

<u>Question 3 Joining/Pivoting:</u> How do average electricity prices compare across U.S. states from 2021 to 2023?

1. Replace NAs values with 0
2. Pivot to longer format
3. Calculate average electricity price per state across 2021-2023

```
price_clean <- price_clean |>
    mutate(price_2021 = ifelse(is.na(price_2021), 0, price_2021),
           price_2022 = ifelse(is.na(price_2022), 0, price_2022),
           price_2023 = ifelse(is.na(price_2023), 0, price_2023))
price_clean
```

```
# A tibble: 55 × 4
   state                        price_2021 price_2022 price_2023
   <chr>                             <dbl>      <dbl>      <dbl>
 1 "Total energy average price"        0          0          0
 2 ""                                  0          0          0
 3 "State"                          2021       2022       2023
 4 "AK"                               20.0       27.3       23.8
 5 "AL"                               17.8       23.4        0
 6 "AR"                               18.4       23.8       21.8
 7 "AZ"                                0         31.7       30.3
 8 "CA"                               28.4       37.4       35.7
 9 "CO"                               20.6        0         23.8
10 "CT"                               25.8       33.2       32.3
# ℹ 45 more rows
```

```
price_long <- price_clean |>
    pivot_longer(cols = starts_with("price_"),
                 names_to = "year",
                 values_to = "price") |>
    mutate(year = str_remove(year, "price_"),
           price = as.numeric(price))

price_long
```

```
# A tibble: 165 × 3
   state                        year  price
   <chr>                        <chr> <dbl>
 1 "Total energy average price" 2021      0
 2 "Total energy average price" 2022      0
 3 "Total energy average price" 2023      0
 4 ""                           2021      0
 5 ""                           2022      0
 6 ""                           2023      0
```

```
 7 "State"                        2021  2021
 8 "State"                        2022  2022
 9 "State"                        2023  2023
10 "AK"                           2021    20.0
# ℹ 155 more rows
```

```
price_avg_state <- price_long |>
    group_by(state) |>
    summarize(avg_price = mean(price, na.rm = TRUE))

price_avg_state <- price_avg_state |>
    mutate(state = tolower(state.name[match(state, state.abb)]))

price_avg_state
```

```
# A tibble: 55 × 2
   state       avg_price
   <chr>           <dbl>
 1 <NA>              0
 2 alaska           23.7
 3 alabama          13.7
 4 arkansas         21.3
 5 arizona          20.7
 6 california       33.8
 7 colorado         14.8
 8 connecticut      30.4
 9 <NA>             21.0
10 delaware         25.4
# ℹ 45 more rows
```

Part 4: Mapping Visualization

Question 1: How does the share of renewable electricity generation vary across U.S. states in 2023?

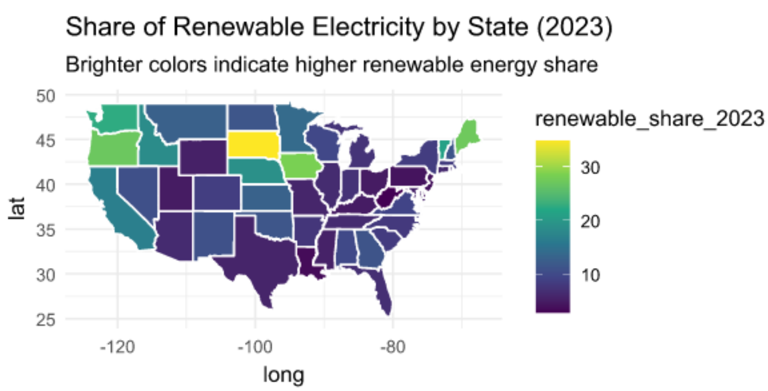
```
us_map <- map_data("state")

state_data <- energy_joined_2023 |>
    mutate(state = ifelse(!is.na(match(toupper(state), state.abb)),
                    tolower(state.name[match(toupper(state), state.abb)]),
                    tolower(state)))


us_map_joined <- us_map |>
    left_join(state_data, by = c("region" = "state"))

ggplot(us_map_joined, aes(long, lat, group = group, fill =
```

```r
renewable_share_2023)) +
    geom_polygon(color = "white") +
    coord_fixed(1.3) +
    scale_fill_viridis_c() +
    theme_minimal() +
    labs(title = "Share of Renewable Electricity by State (2023)",
    subtitle = "Brighter colors indicate higher renewable energy share")
```



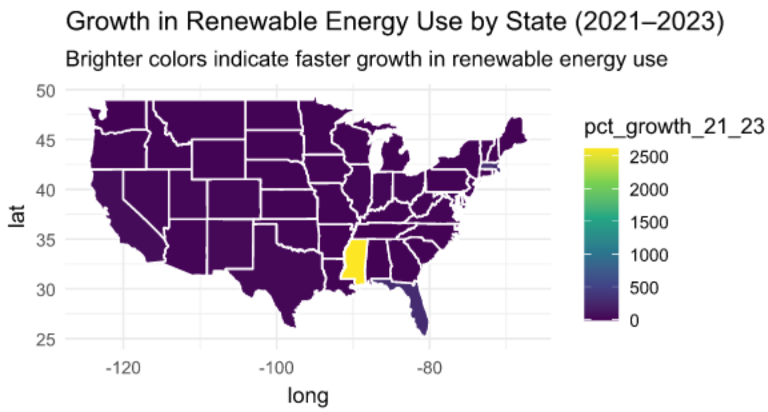
Question 2: Which U.S. states experienced the fastest growth in renewable energy use from 2021 to 2023?

```r
us_map <- map_data("state")

renew_all <- renew_all |>
  mutate(state = ifelse(
      !is.na(match(toupper(state), state.abb)),
      tolower(state.name[match(toupper(state), state.abb)]),
      tolower(state)))

us_map_joined <- us_map |>
    left_join(renew_all, by = c("region" = "state"))

ggplot(us_map_joined, aes(long, lat, group = group, fill = pct_growth_21_23))
+
    geom_polygon(color = "white") +
    coord_fixed(1.3) +
    scale_fill_viridis_c() +
    theme_minimal() +
    labs(title = "Growth in Renewable Energy Use by State (2021–2023)",
         subtitle = "Brighter colors indicate faster growth in renewable
energy use")
```

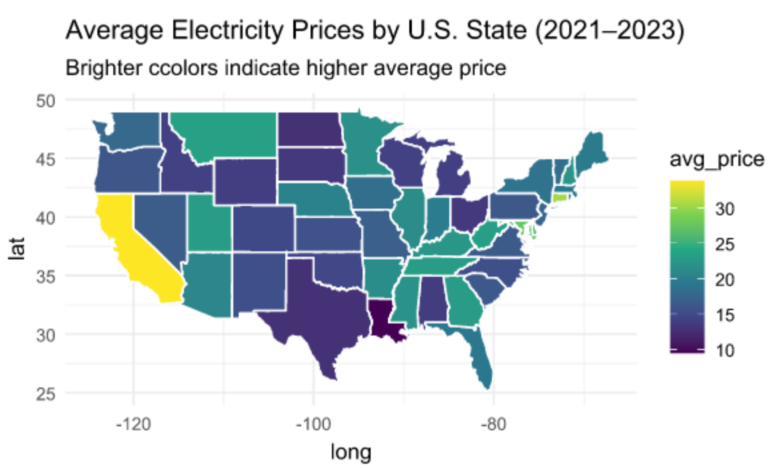


Question 3: How do average electricity prices compare across U.S. states from 2021 to 2023?

```
us_map <- map_data("state")

us_map_joined <- left_join(us_map, price_avg_state, by = c("region" =
"state"))

ggplot(us_map_joined, aes(long, lat, group = group, fill = avg_price)) +
    geom_polygon(color = "white") +
    coord_fixed(1.3) +
    theme_minimal() +
    scale_fill_viridis_c() +
    labs(title = "Average Electricity Prices by U.S. State (2021–2023)",
         subtitle = "Brighter ccolors indicate higher average price")
```



Analysis

The first map shows that in 2023, states with high renewable electricity shares—such as California, Oregon, Washington, and Maine—are better positioned to charge EVs using cleaner energy. In contrast, many southern and midwestern states still rely heavily on fossil fuels, meaning that EVs charged there may indirectly depend on non-renewable sources.

The second map, which tracks renewable energy growth from 2021 to 2023, indicates that most states have seen slow but positive progress toward cleaner electricity generation. Mississippi stands out as an outlier with particularly high growth, suggesting that some traditionally fossil-fuel-dependent regions are beginning to transition. However, the generally modest increases highlight that nationwide progress toward renewable-based electricity remains uneven.

The third map on average electricity prices reveals that states with higher renewable shares, including California and northeastern states, often face higher electricity costs. This may reflect investments in renewable infrastructure and regional market differences. In contrast, central and southern states, where fossil fuels remain dominant, tend to have lower prices but also less clean electricity.

Taken together, these findings suggest that EVs are not equally "clean" across all states. While EVs reduce direct tailpipe emissions everywhere, the carbon intensity of charging depends on each state's energy mix. States with stronger renewable portfolios provide genuinely cleaner charging, while others may still rely heavily on coal or natural gas–based electricity.