EV Power - Lab 4 Project Report

OVERVIEW

In this report, I investigate the main research question: "Electric vehicles reduce direct emissions, but does the electricity used to charge them actually come from clean sources?

My question for analysis is **Do states with higher renewable usage have lower average electricity prices?**

Data and Methods

Part 0: libraries

```
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(sf)
```

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

```
library(usmap)
library(patchwork)
```

Part 1: Defining Research Question

Chosen Question: Do states with higher renewable usage have lower average electricity prices?

Part 2: Data Preparation and Cleaning

```
renew_2021_raw <- 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_raw <- 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_2023_raw <- 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.
```

```
avg_price_raw <- read_csv("data/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_col_names <- function(col_name) {
  col_name |>
    str_to_lower() |>
    str_replace_all(" ", "_") |>
    str_replace_all("[^a-z0-9_]", "")
}


renew_2021 <- renew_2021_raw |>
  rename_with(clean_col_names) |>
  rename(renewable_use = renewable_use_2021) |>
  mutate(year = 2021)

renew_2022 <- renew_2022_raw |>
  rename_with(clean_col_names) |>
  rename(renewable_use = renewable_use_2022) |>
  mutate(year = 2022)


renew_2023 <- renew_2023_raw |>
  rename_with(clean_col_names) |>
  rename(renewable_use = renewable_use_2023) |>
  mutate(year = 2023)

renew_use_all_years <- bind_rows(
  renew_2021,
  renew_2022,
  renew_2023
)
```

```r
avg_price <- avg_price_raw |>
  rename_with(clean_col_names)

cat("\nRenewable Data (Combined):\n")
```

```
Renewable Data (Combined):
```

```r
glimpse(renew_use_all_years)
```

```
Rows: 780
Columns: 4
$ state         <chr> "AK", "AK", "AK", "AK", "AK", "AL", "AL", "AL", "AL", "A~
$ energy_source <chr> "Biomass", "Geothermal", "Hydropower", "Solar Energy", "~
$ renewable_use <chr> " 3153", "186 MMBtu", "5763 about", "~45", "451 USD", "1~
$ year          <dbl> 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 2021, 20~
```

```r
cat("\nAverage Price Data:\n")
```

```
Average Price Data:
```

```r
glimpse(avg_price)
```

```
Rows: 54
Columns: 1
$ total_energy_average_price_dollars_per_million_btu <chr> ",,,", "State,2021,~
```

Part 3: Joining / Pivoting Datasets for Analysis

```r
renew_summary <- renew_use_all_years |>
  mutate(
    renewable_use_num = str_extract(renewable_use, "\\d+") |> as.numeric()
  ) |>
  group_by(state, year) |>
  summarize(
```

```
    total_renewable_use = sum(renewable_use_num, na.rm = TRUE)
  ) |>
  ungroup()
```

```
`summarise()` has grouped output by 'state'. You can override using the
`.groups` argument.
```

```
cat("Cleaned Renewable Data (Summary):\n")
```

```
Cleaned Renewable Data (Summary):
```

```
glimpse(renew_summary)
```

```
Rows: 156
Columns: 3
$ state               <chr> "AK", "AK", "AK", "AL", "AL", "AL", "AR", "AR", "A~
$ year                <dbl> 2021, 2022, 2023, 2021, 2022, 2023, 2021, 2022, 20~
$ total_renewable_use <dbl> 9598, 10410, 10088, 239816, 232035, 222189, 89714,~
```

```
price_clean <- avg_price |>
  slice(-c(1, 2)) |> # the first two rows are kinda bad so I removed them
  separate(
    col = total_energy_average_price_dollars_per_million_btu,
    into = c("state", "price_2021", "price_2022", "price_2023"),
    sep = ","
  ) |>
  pivot_longer(
    cols = starts_with("price_"),
    names_to = "year_str",
    values_to = "avg_price_str"
  ) |>
  mutate(
    year = str_extract(year_str, "\\d{4}") |> as.numeric(),
    avg_price = str_extract(avg_price_str, "[0-9\\.]+") |> as.numeric()
  ) |>
  select(state, year, avg_price)

cat("\nCleaned Price Data (Tidy):\n")
```

Cleaned Price Data (Tidy):

```
glimpse(price_clean)
```

```
Rows: 156
Columns: 3
$ state     <chr> "AK", "AK", "AK", "AL", "AL", "AL", "AR", "AR", "AR", "AZ", ~
$ year      <dbl> 2021, 2022, 2023, 2021, 2022, 2023, 2021, 2022, 2023, 2021, ~
$ avg_price <dbl> 20.03, 27.33, 23.84, 17.85, 23.37, 21.11, 18.42, 23.84, 21.7~
```

```
energy_data_joined <- left_join(
  renew_summary,
  price_clean,
  by = c("state", "year")
) |>
  rename(total_reusable_use_MMBTU = total_renewable_use) |>
  mutate(state = str_to_upper(state))

cat("\nJoined Data for everything:\n")
```

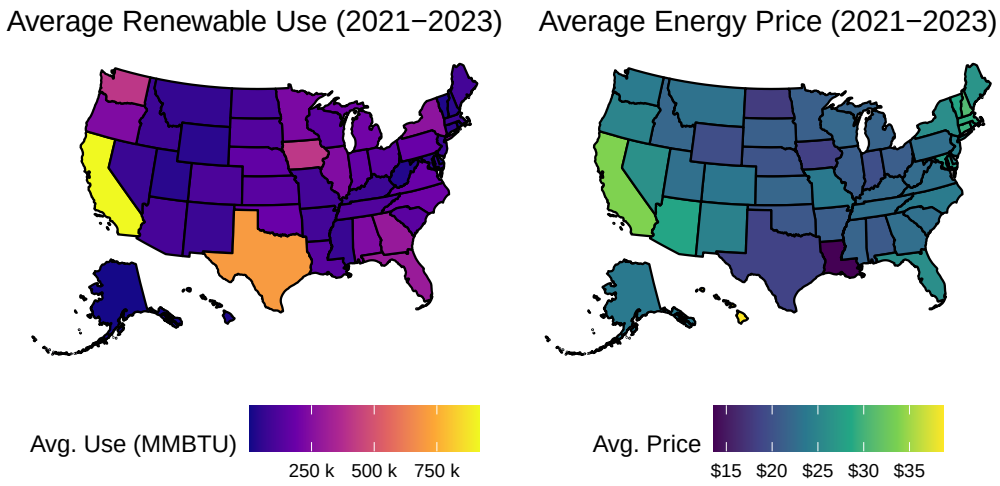
```
Joined Data for everything:
```

```
glimpse(energy_data_joined)
```

```
Rows: 156
Columns: 4
$ state                    <chr> "AK", "AK", "AK", "AL", "AL", "AL", "AR", "AR~
$ year                     <dbl> 2021, 2022, 2023, 2021, 2022, 2023, 2021, 202~
$ total_reusable_use_MMBTU <dbl> 9598, 10410, 10088, 239816, 232035, 222189, 8~
$ avg_price                <dbl> 20.03, 27.33, 23.84, 17.85, 23.37, 21.11, 18.~
```

```
energy_summary_avg <- energy_data_joined |>
  filter(!is.na(total_reusable_use_MMBTU), !is.na(avg_price)) |>
  group_by(state) |>
  summarize(
    avg_renewable_use = mean(total_reusable_use_MMBTU, na.rm = TRUE),
    avg_energy_price = mean(avg_price, na.rm = TRUE)
  ) |>
  ungroup()
```

Part 4: Mapping Visualization



Part 5: Analysis

I noticed that California has very high average renewable energy use, but also a high energy price. Texas on the other hand also has a very high average renewable energy use, but has a much lower average energy price. This is likely due to the fact that the cost of living in Texas is generally lower.

As for the the other states states, there appear to be almost no correlation between a states' energy price and renewable energy usage.

Lastly, the maps show that for the vast majority of the United States (the large dark purple areas), the total renewable energy use is low. This suggests that in most states, the electricity grid is not predominantly clean. So, EVs in those regions are likely being charged by electricity from non-renewable sources.