

EV Power - Lab 4 Project Report

Aimee Pan

Example Solution 1

Part 0: libraries

```
library(readr)
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
library(tidyr)
library(stringr)
library(ggplot2)
library(janitor)
```

Attaching package: 'janitor'

The following objects are masked from 'package:stats':

chisq.test, fisher.test

```
library(usmap)
library(viridis)
```

Loading required package: viridisLite

Part 1: Defining Research Question

Main research question: Do EVs run on clean electricity in the U.S.?

Sub-questions:

How has the share of renewable energy changed from 2021–2023 across states?

Are EV registrations concentrated in states with cleaner energy mixes (2023)?

Do renewable-heavy states have lower average energy prices (2023)?

Part 2: Data Preparation and Cleaning

```
data_dir <- "data"

renew21 <- read_csv(file.path(data_dir, "renew-use-2021.csv")) |>
clean_names()
```

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(file.path(data_dir, "renew-use-2022.csv")) |>
clean_names()
```

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(file.path(data_dir, "renew-use-2023.csv")) |>
clean_names()
```

```
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(file.path(data_dir, "total-use-2021.csv")) |>
clean_names()
```

```
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(file.path(data_dir, "total-use-2022.csv")) |>
clean_names()
```

```
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(file.path(data_dir, "total-use-2023.csv")) |>
clean_names()
```

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.

```
price_raw <- read_csv(file.path(data_dir, "av-energy-price-2021-2023.csv"),  
  show_col_types = FALSE)  
ev23_raw <- read_csv(file.path(data_dir, "ev-registrations-by-  
state-2023.csv"), show_col_types = FALSE)
```

New names:

- `` -> `...2`

```
parse_num <- function(x) readr::parse_number(as.character(x))  
  
to_abb <- function(name) {  
  nm <- str_to_title(name)  
  ix <- match(nm, state.name)  
  ifelse(is.na(ix),  
    ifelse(nm %in% c("District Of Columbia", "Washington, Dc", "Washington  
Dc", "Dc", "D.C."), "DC", NA),  
    state.abb[ix])  
}  
  
add_state_abb <- function(df, state_col = "state") {  
  df |>  
    mutate(  
      !!state_col := str_to_title(.data[[state_col]]),  
      state_abb =  
      toupper(coalesce(to_abb(.data[[state_col]]), .data[[state_col]]))  
    )  
}  
  
pivot_total_sumall <- function(df, year_value) {  
  df |>  
    clean_names() |>  
    pivot_longer(  
      cols = -1,
```

```

    names_to = "state",
    values_to = "value"
  ) |>
  mutate(
    state = stringr::str_to_title(state),
    value = readr::parse_number(as.character(value))
  ) |>
  group_by(state) |>
  summarise(total_use = sum(value, na.rm = TRUE), .groups = "drop") |>
  mutate(year = as.integer(year_value)) |>
  mutate(
    state_abb = toupper(coalesce({
      nm <- stringr::str_to_title(state)
      ix <- match(nm, state.name)
      ifelse(is.na(ix),
        ifelse(nm %in% c("District Of Columbia", "Washington,
Dc", "Washington Dc", "Dc", "D.C."), "DC", NA),
        state.abb[ix])
    }, state))
  ) |>
  select(state_abb, year, total_use)
}

total_all <- bind_rows(
  pivot_total_sumall(total21, 2021),
  pivot_total_sumall(total22, 2022),
  pivot_total_sumall(total23, 2023)
)

pivot_renew_sum <- function(df, year_value) {
  val_col <- names(df)[grepl("renewable_use", names(df))]
  df |>
    mutate(value = parse_num(.data[[val_col]])) |>
    group_by(state) |>
    summarise(renew_use = sum(value, na.rm = TRUE), .groups = "drop") |>
    mutate(year = as.integer(year_value)) |>
    add_state_abb("state") |>
    select(state_abb, year, renew_use)
}

renew_all <- bind_rows(
  pivot_renew_sum(renew21, 2021),
  pivot_renew_sum(renew22, 2022),
  pivot_renew_sum(renew23, 2023)
)

price_tidy <- price_raw |>
  clean_names() |>

```

```

rename(line = 1) |>
filter(!is.na(line), !str_detect(line, "^\\s*$")) |>
filter(!str_detect(line, regex("^"?state\\s*,?\\s*2021\\s*,?\\s*2022\\s*,?\\s*2023", ignore_case = TRUE))) |>
separate(col = line, into = c("state_raw", "p2021", "p2022", "p2023"),
  sep = ",", fill = "right", extra = "merge") |>
mutate(
  state = str_to_title(str_trim(state_raw)),
  p2021 = parse_num(p2021),
  p2022 = parse_num(p2022),
  p2023 = parse_num(p2023)
) |>
select(state, p2021, p2022, p2023) |>
pivot_longer(cols = starts_with("p"), names_to = "year", values_to =
"avg_price") |>
mutate(year = as.integer(str_extract(year, "\\d{4}"))) |>
add_state_abb("state") |>
select(state_abb, year, avg_price) |>
arrange(year, state_abb)

ev23_tidy <- ev23_raw |>
clean_names() |>
rename(state = 1, ev_text = 2) |>
filter(!is.na(state)) |>
slice(-(1:2)) |>
mutate(
  state = str_to_title(state),
  ev_registrations = parse_num(ev_text)
) |>
add_state_abb("state") |>
select(state_abb, ev_registrations) |>
arrange(state_abb)

```

I cleaned seven CSVs. total-use-YYYY: states are columns, last row stores totals. renew-use-YYYY: I parsed numeric strings and summed across renewable sources to get (state_abb, year, renew_use). av-energy-price-2021-2023: a single text column. I split by commas, parsed numbers, and produced (state_abb, year, avg_price). ev-registrations-by-state-2023: I dropped the first two header lines, parsed numbers to get (state_abb, ev_registrations). I standardized two-letter state codes for consistent joins and mapping.

Part 3: Joining / Pivoting Datasets for Analysis

```

energy_all <- renew_all |>
inner_join(total_all, by = c("state_abb", "year")) |>
mutate(
  renew_share = if_else(total_use > 0, 100 * renew_use / total_use,
NA_real_)

```

```

)

energy_2023 <- energy_all |>
  filter(year == 2023) |>
  left_join(ev23_tidy, by = "state_abb") |>
  left_join(price_tidy |>
    filter(year == 2023) |>
    select(state_abb, avg_price),
    by = "state_abb") |>
  transmute(
    state_abb, year, renew_use, total_use, renew_share,
    ev_registrations, avg_price,
    ev_per_energy = dplyr::if_else(total_use > 0, ev_registrations /
total_use, NA_real_)
  )

names(energy_all); names(energy_2023)

```

```
[1] "state_abb" "year" "renew_use" "total_use" "renew_share"
```

```

[1] "state_abb" "year" "renew_use"
"total_use"
[5] "renew_share" "ev_registrations" "avg_price"
"ev_per_energy"

```

```
summary(energy_2023$renew_share); summary(energy_2023$ev_registrations)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.351	5.758	8.076	10.299	11.407	34.844

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
959	8087	25833	70848	71645	1256646	2

For 2023 I merged EV registrations and prices to obtain energy_2023 used in RQ2 and RQ3.

Part 4: Mapping Visualization

```

map_key <- usmap::statepop |> select(fips, abbr) |> rename(state_abb = abbr)
map_2023 <- map_key |> left_join(energy_2023, by = "state_abb")

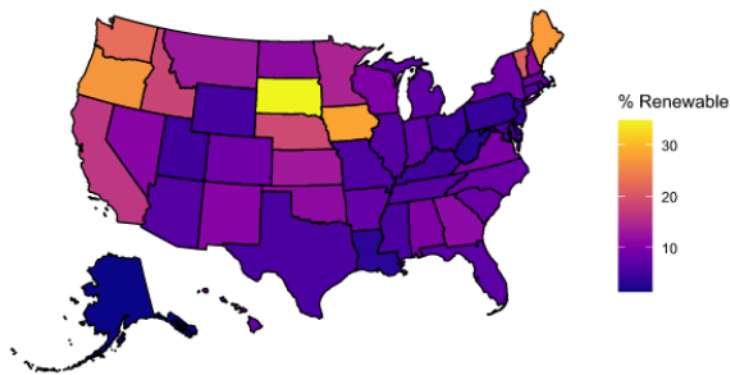
plot_usmap(data = map_2023, values = "renew_share", regions = "states") +
  scale_fill_viridis(option = "C", name = "% Renewable", na.value = "grey90")
+

```

```
labs(
  title = "Renewable Energy Share by State (2023)",
  subtitle = "renew_share = 100 × renew_use / total_use",
  caption = "STAT133 Project 4"
) +
theme(legend.position = "right")
```

Renewable Energy Share by State (2023)

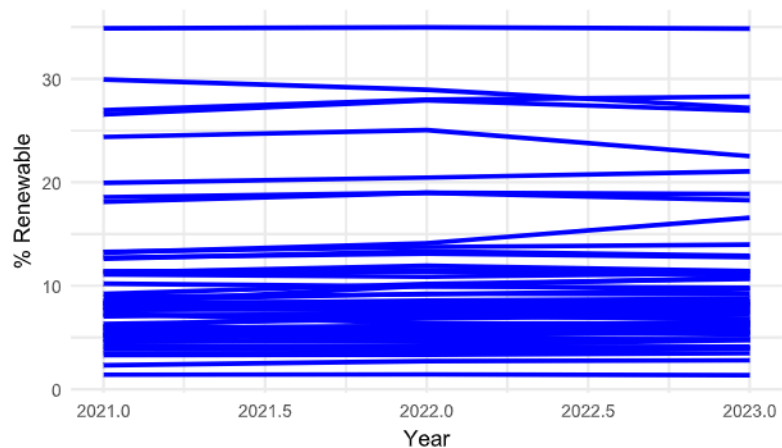
$\text{renew_share} = 100 \times \text{renew_use} / \text{total_use}$



STAT133 Project 4

```
ggplot(energy_all, aes(x = year, y = renew_share, group = state_abb)) +
  geom_line(alpha = 0.25, color = "grey40") +
  stat_summary(fun = mean, geom = "line", linewidth = 1.1, color = "blue") +
  labs(
    title = "Trends in Renewable Energy Share by State (2021–2023)",
    x = "Year", y = "% Renewable"
  ) +
  theme_minimal()
```

Trends in Renewable Energy Share by State (2021–2023)

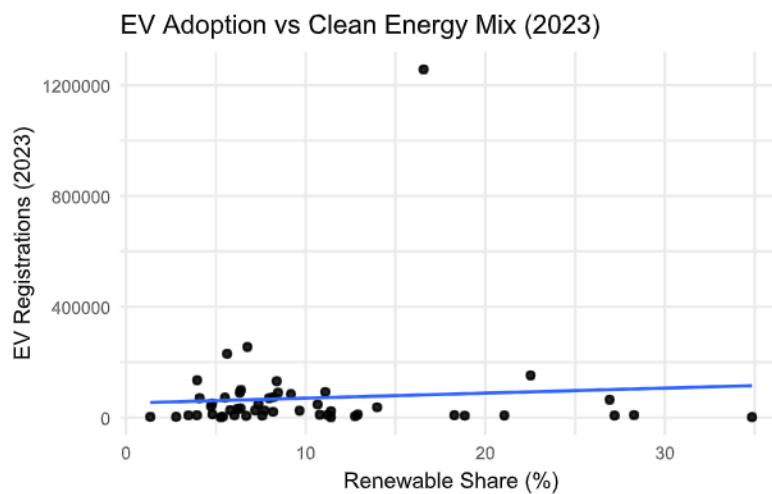



```
ggplot(energy_2023, aes(x = renew_share, y = ev_registrations)) +
  geom_point(alpha = 0.85) +
  geom_smooth(method = "lm", se = FALSE, linewidth = 0.8) +
  labs(
    x = "Renewable Share (%)",
    y = "EV Registrations (2023)",
    title = "EV Adoption vs Clean Energy Mix (2023)"
  ) +
  theme_minimal()
```

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

Warning: Removed 2 rows containing non-finite outside the scale range
(`stat_smooth()`).

Warning: Removed 2 rows containing missing values or values outside the scale range
(`geom_point()`).

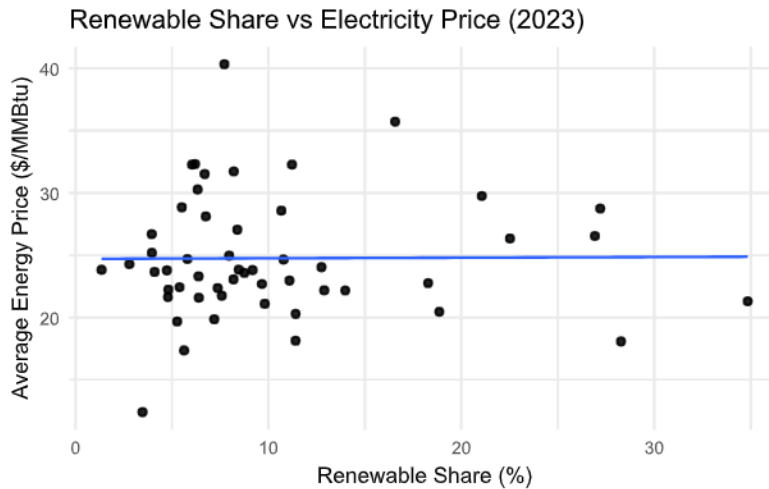


```
energy_price_2023 <- energy_2023 |> filter(!is.na(avg_price))

ggplot(energy_price_2023, aes(x = renew_share, y = avg_price)) +
  geom_point(alpha = 0.85) +
  geom_smooth(method = "lm", se = FALSE, linewidth = 0.7) +
  labs(
    x = "Renewable Share (%)",
    y = "Average Energy Price ($/MMBtu)",
    title = "Renewable Share vs Electricity Price (2023)"
  )
```

```
) +  
theme_minimal()
```

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



I created three main visualizations to answer my questions.

The first map shows how renewable energy is distributed across states in 2023. States in the West and Northeast tend to have higher renewable shares, while many southern states still rely heavily on fossil fuels.

The second plot tracks how each state's renewable share changed from 2021 to 2023. Most states show a slow but steady increase, although the pace is uneven. A few states barely moved, while others, especially those with strong clean-energy policies, grew faster.

The third set of scatterplots connects renewable energy with other factors. One plot compares EV registrations and renewable share: states like California and Washington stand out with both high EV adoption and cleaner electricity. Another plot compares energy prices and renewable share: there's no simple relationship, but it seems that cleaner grids don't automatically mean cheaper electricity.

Part 5: Analysis

When I look at these results, I get a clearer sense of how uneven the clean-energy transition really is. Some states are already running on cleaner grids, while others are still catching up. The gradual rise from 2021 to 2023 suggests steady progress, but not enough to close the gap.

In 2023, EV adoption appears to line up with cleaner electricity in a few key regions—mainly the West Coast and parts of the Northeast. But in places like Texas, EV numbers are high even though the energy mix is still relatively dirty. It shows that EV growth doesn't always wait for the grid to go green.

I also thought renewable energy would make electricity cheaper, but that didn't really show up in the data. Prices seem to depend more on local market structure and energy demand than on renewable share alone.

To sum up, I think EVs are an important step toward cutting emissions, but the real environmental payoff only comes when the electricity itself is clean. The data makes me feel both hopeful and a bit impatient—it's progress, just slower than I'd like.