

EV Power - Lab 4 Project Report

Derek Zhang

Part 0: libraries

```
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)
```

```
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_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_2021 <- read_csv("data/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.
```

```
total_2022 <- read_csv("data/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.
```

```
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.
```

```
ev_2023 <- read_csv("data/ev-registrations-by-state-2023.csv")
```

```
New names:
```

```
Rows: 54 Columns: 2
```

```
— Column specification
```

```
Delimiter: "," chr
(2): electric vehicle registrations_by_state (2023), ...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.
```

```
• `` -> `...2`
```

```
price_2123 <- 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.
```

Part 1: Defining Research Question

Chosen Question: How has each state's share of renewable energy changed from 2021 to 2023?
or Which U.S. regions rely most on renewable energy in 2023?

Part 2: Data Preparation and Cleaning

```
if ("State" %in% names(renew_2021)) renew_2021 <- renew_2021 |> rename(state = State)
if ("Energy_Source" %in% names(renew_2021)) renew_2021 <- renew_2021 |>
  rename(energy_source = Energy_Source)
if ("Renewable_Use_2021" %in% names(renew_2021)) renew_2021 <- renew_2021 |>
  rename(renewable_use_2021 = Renewable_Use_2021)

if ("State" %in% names(renew_2022)) renew_2022 <- renew_2022 |> rename(state = State)
if ("Energy_Source" %in% names(renew_2022)) renew_2022 <- renew_2022 |>
  rename(energy_source = Energy_Source)
if ("Renewable_Use_2022" %in% names(renew_2022)) renew_2022 <- renew_2022 |>
  rename(renewable_use_2022 = Renewable_Use_2022)

if ("State" %in% names(renew_2023)) renew_2023 <- renew_2023 |> rename(state = State)
if ("Energy_Source" %in% names(renew_2023)) renew_2023 <- renew_2023 |>
  rename(energy_source = Energy_Source)
if ("Renewable_Use_2023" %in% names(renew_2023)) renew_2023 <- renew_2023 |>
  rename(renewable_use_2023 = Renewable_Use_2023)

if ("Energy_Source" %in% names(total_2021)) total_2021 <- total_2021 |>
  rename(energy_source = Energy_Source)
if ("Energy_Source" %in% names(total_2022)) total_2022 <- total_2022 |>
  rename(energy_source = Energy_Source)
if ("Energy_Source" %in% names(total_2023)) total_2023 <- total_2023 |>
  rename(energy_source = Energy_Source)

if ("State" %in% names(ev_2023)) ev_2023 <- ev_2023 |> rename(state = State)

if ("State" %in% names(price_2123)) price_2123 <- price_2123 |> rename(state = State)
```

Part 3: Joining / Pivoting Datasets for Analysis

```
#Pivot each year to long and combining

total_2021_long <- total_2021 |>
  pivot_longer(
    cols = -energy_source,
    names_to = "state_abbr",
    values_to = "value"
```

```

) |>
mutate(year = 2021)

total_2022_long <- total_2022 |>
pivot_longer(
  cols = -energy_source,
  names_to = "state_abbr",
  values_to = "value"
) |>
mutate(year = 2022)

total_2023_long <- total_2023 |>
pivot_longer(
  cols = -energy_source,
  names_to = "state_abbr",
  values_to = "value"
) |>
mutate(year = 2023)

# one combined table for 2021–2023
total_all_long <- bind_rows(total_2021_long, total_2022_long, total_2023_long)

# (1) total energy across all sources
total_by_state_year <- total_all_long |>
group_by(state_abbr, year) |>
summarise(total_energy = sum(value, na.rm = TRUE), .groups = "drop")

# (2) renewable-only row(s) per state/year
renew_by_state_year <- total_all_long |>
filter(str_detect(energy_source, "renewable")) |>
group_by(state_abbr, year) |>
summarise(renewable_energy = sum(value, na.rm = TRUE), .groups = "drop")

# (3) join and compute percentage renewable
energy_share <- total_by_state_year |>
left_join(renew_by_state_year, by = c("state_abbr", "year")) |>
mutate(renew_pct = 100 * renewable_energy / total_energy)

energy_share |> arrange(year, desc(renew_pct)) |> head(10)

```

```

# A tibble: 10 × 5
  state_abbr year total_energy renewable_energy renew_pct
  <chr>      <dbl>      <dbl>          <dbl>      <dbl>
1 SD        2021      365263      127382      34.9
2 ME        2021      317953       95141      29.9
3 IA        2021     1446015      389787      27.0
4 OR        2021      849834      225544      26.5

```

5	WA	2021	1616190	394052	24.4
6	VT	2021	107472	21430	19.9
7	ID	2021	400918	74428	18.6
8	NE	2021	874553	158275	18.1
9	MN	2021	1627997	216113	13.3
10	CA	2021	6142252	810020	13.2

Part 4: Mapping Visualization

```
# Load the map data
library(maps)
```

Attaching package: 'maps'

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

map

```
# Step 1: Get map of U.S. states
us_map <- map_data("state")

# The map uses lowercase full state names (e.g., "california"), not "CA".

state_lookup <- tibble(
  state_abbr = state.abb,
  region = tolower(state.name)
)

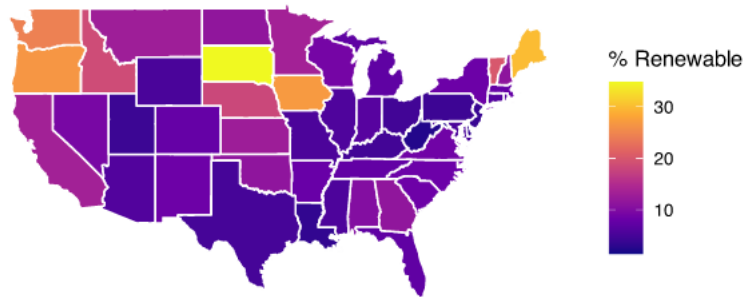
# Step 3: Join energy_share data for 2021 (or 2023)
energy_map <- energy_share |>
  filter(year == 2021) |>
  left_join(state_lookup, by = "state_abbr") |>
  left_join(us_map, by = "region")

# Step 4: Draw the map
ggplot(energy_map, aes(long, lat, group = group, fill = renew_pct)) +
  geom_polygon(color = "white") +
  coord_fixed(1.3) +
  scale_fill_viridis_c(option = "plasma", name = "% Renewable") +
  labs(
    title = "Share of Renewable Energy by State (2021)",
    subtitle = "Percentage of total energy from renewable sources",
    caption = "Source: U.S. Energy Data 2021–2023"
  ) +
```

```
theme_minimal() +
theme(
  axis.text = element_blank(),
  axis.title = element_blank(),
  panel.grid = element_blank()
)
```

Share of Renewable Energy by State (2021)

Percentage of total energy from renewable sources



Source: U.S. Energy Data 2021–2023