EV Power - Lab 4 Project Report

Isabel Aguilar

The change in renewable energy consumption compared to coal consumption across states from 2021 to 2023

Research Question

How has coal consumption changed across states from 2021 to 2023 in comparison to total renewable energy consumption within that same time frame?

Overview

This project investigates how the consumption of renewable energy has changed compared to that of coal across U.S. states between 2021 and 2023.

The goal of this project is to determine whether states are generating a higher proportion of renewable energy over time relative to coal. I chose to compare coal consumption to total renewables because coal historically has been one of the largest contributors to U.S. electricity generation but also one of the most harmful for the environment because of the sheer amount of carbon it uses. In recent years, renewable energy sources have emerged as a way to solve this problem of environmental degradation. Total renewables, including solar, wind, hydropower, biomass, and geothermal power, represent the future of clean energy production and signal progress toward sustainable, low carbon electricity. This is a very important question for understanding whether the electricity that powers electric vehicles truly comes from the cleaner sources of the future or the carbon-intensive ones of the past.

The analysis focuses on the percentage of renewable energy out of the combined total of coal and renewables in each state. By comparing this ratio across years, we can identify trends in the national shift toward sustainable power.

Data Preparation and Cleaning Code

```
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(sf)
```

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

```
renew2021 <- read_csv("~/Desktop/stat133/project4/ev-power-isabelcaguilar/
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.
```

```
renew2022 <- read_csv("~/Desktop/stat133/project4/ev-power-isabelcaguilar/
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.
```

```
renew2023 <- read_csv("~/Desktop/stat133/project4/ev-power-isabelcaguilar/
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.
```

```r
total2021 <- read_csv("~/Desktop/stat133/project4/ev-power-isabelcaguilar/
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.
```

```r
total2022 <- read_csv("~/Desktop/stat133/project4/ev-power-isabelcaguilar/
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.
```

```r
total2023 <- read_csv("~/Desktop/stat133/project4/ev-power-isabelcaguilar/
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.
```

```
cleaner <- function(df) {
  df |>
    rename_with(~ str_replace_all(., " ", "_")) |>
    rename_with(~ str_replace_all(., "-", "_")) |>
    rename_with(~ str_replace_all(., "\\(", "")) |>
    rename_with(~ str_replace_all(., "\\)", "")) |>
    rename_with(tolower)
}

renew2021 <- cleaner(renew2021)
renew2022 <- cleaner(renew2022)
renew2023 <- cleaner(renew2023)

total2021 <- cleaner(total2021)
total2022 <- cleaner(total2022)
total2023 <- cleaner(total2023)


total2021 <- total2021 |>
  mutate(energy_source = str_to_lower(str_trim(energy_source))) |>
  mutate(energy_source = case_when(
    str_detect(energy_source, "coal") ~ "coal",
    str_detect(energy_source, "total_renew") ~ "total_renewables",
    TRUE ~ energy_source
  ))

total2022 <- total2022 |>
  mutate(energy_source = str_to_lower(str_trim(energy_source))) |>
  mutate(energy_source = case_when(
    str_detect(energy_source, "coal") ~ "coal",
    str_detect(energy_source, "total_renew") ~ "total_renewables",
    TRUE ~ energy_source
  ))

total2023 <- total2023 |>
```

```r
  mutate(energy_source = str_to_lower(str_trim(energy_source))) |>
  mutate(energy_source = case_when(
    str_detect(energy_source, "coal") ~ "coal",
    str_detect(energy_source, "total_renew") ~ "total_renewables",
    TRUE ~ energy_source
  ))

colnames(total2021) <- str_to_upper(colnames(total2021))
colnames(total2022) <- str_to_upper(colnames(total2022))
colnames(total2023) <- str_to_upper(colnames(total2023))

print(paste("missing values in total2021:", sum(is.na(total2021))))
```

```
[1] "missing values in total2021: 0"
```

```r
print(paste("missing values in total2022:", sum(is.na(total2022))))
```

```
[1] "missing values in total2022: 0"
```

```r
print(paste("missing values in total2023:", sum(is.na(total2023))))
```

```
[1] "missing values in total2023: 0"
```

```r
tt21 <- total2021 |> mutate(year = 2021)
tt22 <- total2022 |> mutate(year = 2022)
tt23 <- total2023 |> mutate(year = 2023)
```

Part 3: Joining / Pivoting Code

```r
# Pivot from wide to long to make the states all go in one column for the 2021
year
tt21_long <- tt21 |>
  pivot_longer(
    cols = -c(ENERGY_SOURCE, year),
    names_to = "state",
    values_to = "energy_use"
  )
# Pivot from wide to long to make the states all go in one column for the 2022
year
tt22_long <- tt22 |>
  pivot_longer(
    cols = -c(ENERGY_SOURCE, year),
```

```r
    names_to = "state",
    values_to = "energy_use"
  )
# Pivot from wide to long to make the states all go in one column for the 2023
year
tt23_long <- tt23 |>
  pivot_longer(
    cols = -c(ENERGY_SOURCE, year),
    names_to = "state",
    values_to = "energy_use"
  )


#The 2021–2023 total energy datasets (tt21, tt22, tt23) were merged to combine
state data at each year
# This way, A direct compatiosn can be made between coal and renewable energy
use across years each of the years and per state
#this allows me to answer my research question of how coal consumption has
changed across states from 2021 to 2023
# in comparison to total renewable energy in the same time frame

energy_long <- bind_rows(tt21_long, tt22_long, tt23_long)

energy_long <- energy_long |>
  mutate(ENERGY_SOURCE = str_to_lower(str_trim(ENERGY_SOURCE))) |>
  mutate(ENERGY_SOURCE = case_when(
    str_detect(ENERGY_SOURCE, "coal") ~ "coal",
    str_detect(ENERGY_SOURCE, "renew") ~ "total_renewables",
    TRUE ~ ENERGY_SOURCE
  ))

# This is filtering only for the energy sources I am comparing (the relevant
ones : coal and total_renewables)
energy_fill <- energy_long |>
  filter(ENERGY_SOURCE %in% c("coal", "total_renewables")) |>
  filter(!is.na(energy_use))

# Pivot wider is done to get one row per state and by year!
energy_wide <- energy_fill |>
  pivot_wider(
    names_from = ENERGY_SOURCE,
    values_from = energy_use
  )

# new variable made to show the percent renewable of the 2 energy sources I am
comapring
energy_wide <- energy_wide |>
  mutate(
    total_energy_twosource = coal + total_renewables,
```

```
    percent_renewable = (total_renewables / total_energy_twosource) * 100
  )

# This is a summary table of the average energy sources that I chose!
energy_summary <- energy_wide |>
  group_by(year) |>
  summarise(
    avg_coal = mean(coal, na.rm = TRUE),
    avg_renew = mean(total_renewables, na.rm = TRUE),
    avg_percent_renewable = mean(percent_renewable, na.rm = TRUE)
  )
```

```
head(tt21_long)
```

```
# A tibble: 6 × 4
  ENERGY_SOURCE  year state energy_use
  <chr>         <dbl> <chr>      <dbl>
1 coal           2021 AK         18694
2 coal           2021 AL        309791
3 coal           2021 AR        216123
4 coal           2021 AZ        160299
5 coal           2021 CA         28244
6 coal           2021 CO        252442
```

```
head(tt22_long)
```

```
# A tibble: 6 × 4
  ENERGY_SOURCE  year state energy_use
  <chr>         <dbl> <chr>      <dbl>
1 coal           2022 AK         18615
2 coal           2022 AL        297654
3 coal           2022 AR        211724
4 coal           2022 AZ        154007
5 coal           2022 CA         30049
6 coal           2022 CO        233256
```

```
head(tt23_long)
```

```
# A tibble: 6 × 4
  ENERGY_SOURCE  year state energy_use
  <chr>         <dbl> <chr>      <dbl>
1 coal           2023 AK         18414
2 coal           2023 AL        224926
```

```
3 coal          2023 AR        180262
4 coal          2023 AZ        137885
5 coal          2023 CA         28746
6 coal          2023 CO        204826
```

Each dataset contains energy sources, which are the rows, and states, which are the columns, that are converted into a longer format using a pivot. This structure makes it easier to compare energy source totals and join over multiple years. Across years, the same energy categories appear (Coal, Natural Gas, Petroleum [BTU], Nuclear, and Total Renewables), allowing a consistent cross-year analysis. After merging, the combined dataset was filtered to include only Coal and Total Renewables, since these two sources are directly relevant to the research question.

energy_long

```
# A tibble: 780 × 4
   ENERGY_SOURCE  year state energy_use
   <chr>         <dbl> <chr>      <dbl>
 1 coal           2021 AK         18694
 2 coal           2021 AL        309791
 3 coal           2021 AR        216123
 4 coal           2021 AZ        160299
 5 coal           2021 CA         28244
 6 coal           2021 CO        252442
 7 coal           2021 CT          2880
 8 coal           2021 DC             0
 9 coal           2021 DE          4542
10 coal           2021 FL        200193
# i 770 more rows
```

This table merges coal and renewable energy data for each state and year. Each row includes numeric values for coal and renewable energy consumption, as well as the calculated percentage of renewable energy out of coal and renewables.

energy_summary

```
# A tibble: 3 × 4
   year avg_coal avg_renew avg_percent_renewable
  <dbl>    <dbl>     <dbl>                 <dbl>
1  2021  405729.   293889.                  53.8
2  2022  380219.   311331.                  56.2
3  2023  314218.   314162.                  60.3
```

This table shows the mean coal usage, mean renewable usage, and average renewable percentage per year. It reveals that renewable energy's share increased from 2021 to 2023, while average coal use decreased, suggesting gradual national progress toward cleaner energy sources.

Part 4: Mapping Visualization

```r
energy_wide <- energy_wide |>
  mutate(
    percent_renewable_coal = (total_renewables / (coal + total_renewables)) *
100
  )

map_data <- energy_wide |>
  select(state, year, percent_renewable_coal) |>
  mutate(state = tolower(state)) |>
  mutate(state = state.name[match(toupper(state), state.abb)]) |>
  mutate(state = tolower(state)) |>
  filter(!is.na(year))

usa <- st_read(
  "https://raw.githubusercontent.com/PublicaMundi/MappingAPI/master/data/
geojson/us-states.json",
  quiet = TRUE
)

usa <- usa |>
  mutate(state = tolower(name))

map_join <- left_join(usa, map_data, by = "state") |>
  filter(!is.na(year))

ggplot(map_join) +
  geom_sf(aes(fill = percent_renewable_coal), color = "white") +
  facet_wrap(~year) +
  scale_fill_gradient(
    low = "#FAD0C4",
    high = "#6A0DAD",
    name = "% Renewable (of Coal and Renewables)"
  ) +
  labs(
    title = "Renewable Energy Share Out of Coal & Renewables in 2021 - 2023",
    subtitle = "Darker purple shows a higher renewable share relative to
coal") +
  theme_minimal() +
  theme(
    axis.text = element_blank(),
    axis.ticks = element_blank(),
    panel.grid = element_blank(),
```
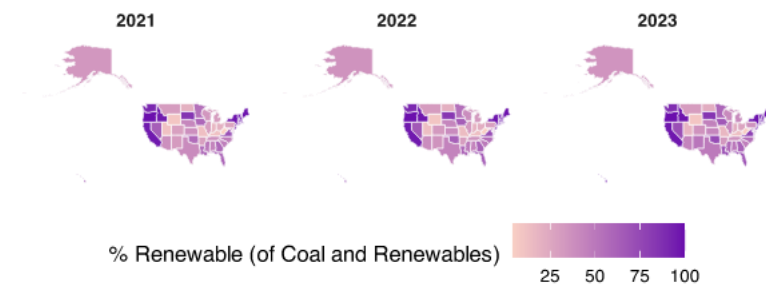
```
    legend.position = "bottom",
    plot.title = element_text(face = "bold", size = 11),
    plot.subtitle = element_text(size = 9),
    strip.text = element_text(face = "bold")
  )
```

Renewable Energy Share Out of Coal & Renewables in 2021 - 2023

Darker purple shows a higher renewable share relative to coal

The map shows each state's percentage of renewable energy out of its coal and renewable total for 2021, 2022, and 2023. The color gradient from light pink to dark purple represents increasing renewable share. In 2021, most states appear in lighter pink shades, indicating that renewables made up a smaller fraction of the coal and renewable mix—often below 40%. The Western states such as California, Oregon, and Washington already show darker purple hues, reflecting their strong hydroelectric and solar capacity. By 2022, those same states maintain high renewable shares, but states like Iowa, Kansas, and Nebraska begin to darken. By 2023, the map shows an overall deepening of color intensity across much of the Western half of the country and parts of the Northeast. This suggests that renewables have had more implementation and usage relative to coal nationwide, though the Southeast and portions of the Midwest remain comparatively light, implying a continued reliance on coal in those specific areas.

Analysis and Final Remarks

The map visualization compares the percentage of renewable energy out of the combined coal and renewable total for each U.S. state from 2021 to 2023. Across the three panels, a visible shift occurs toward darker purple tones, meaning a larger proportion of total energy generation is coming from renewable sources. In the larger context of this project's bigger question, whether electric vehicle power is coming from the clean energy sources of the future or the unsustainable ones of the past this data supports the conclusion that as coal generation declines and renewables expand, the grid supplying EVs becomes less carbon-intensive and actually cleaner. However, this progression is quite slow and difficult to track over the three year jump. Additionally, the progress is not the same everywhere across the map. In fact, the South has shown much slower progression toward cleaner energy sources than the West Coast.