Change in US Renewable Energy Share (2021-2023) - Lab 4 Project Report

Emma Mott

Part 0: libraries

```
#install.packages("sf")
library("sf")
```

```
Linking to GEOS 3.13.1, GDAL 3.11.0, PROJ 9.6.0; sf_use_s2() is TRUE
```

```
library("tidyverse")
```

```
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0
──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.0     ✔ 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
```

```
#install.packages("rnaturalearth")
library("rnaturalearth")
library(stringr)
library(dplyr)
library(ggplot2)
#install.packages("maps")
library(maps)
```

```
Attaching package: 'maps'

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

    map
```

Part 1: Defining Research Question

Overview

Chosen Question: How does the share of renewable energy use across states change between 2021-2023?

Data and Methods

Part 2: Data Preparation and Cleaning

```
#Load data for renewables
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.
```

```
# renaming columns
renew_2021 <- renew_2021 |> rename(Renewable_Use = matches("Renewable_Use"))
renew_2022 <- renew_2022 |> rename(Renewable_Use = matches("Renewable_Use"))
renew_2023 <- renew_2023 |> rename(Renewable_Use = matches("Renewable_Use"))

# adding year column
renew_2021 <- renew_2021 |> mutate(Year = 2021)
renew_2022 <- renew_2022 |> mutate(Year = 2022)
renew_2023 <- renew_2023 |> mutate(Year = 2023)

# combining all three years into one data frame
renew_all <- bind_rows(renew_2021, renew_2022, renew_2023)

# extracting only the numeric values from messy Renewable_Use column using
stringr
renew_all <- renew_all |>
  mutate(Renewable_Use_Num = as.numeric(str_extract(Renewable_Use, "\\d+")))

# summarize total renewable use by state and year
renew_summary <- renew_all |>
  group_by(State, Year) |>
  summarise(Total_Renewable_Use = sum(Renewable_Use_Num, na.rm = TRUE)) |>
  mutate(State = str_to_upper(State))
```

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

```
# Load total energy use
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.
```

```
head(renew_summary)
```

```
# A tibble: 6 × 3
# Groups:   State [2]
  State  Year Total_Renewable_Use
```

```
  <chr> <dbl>               <dbl>
1 AK     2021                9598
2 AK     2022               10410
3 AK     2023               10088
4 AL     2021              239816
5 AL     2022              232035
6 AL     2023              222189
```

Part 3: Joining / Pivoting Datasets for Analysis

```
# Pivot longer for 2021–2023
pivot_total <- function(df, year) {
  df |>
    pivot_longer(
      cols = -Energy_Source,
      names_to = "State",
      values_to = "Energy_Use"
    ) |>
    mutate(Year = year)
}

total_2021_long <- pivot_total(total_2021, 2021)
total_2022_long <- pivot_total(total_2022, 2022)
total_2023_long <- pivot_total(total_2023, 2023)

# Combine all years
total_all <- bind_rows(total_2021_long, total_2022_long, total_2023_long)

# Make all state names uppercase
total_all <- total_all |> mutate(State = str_to_upper(State))

# Compute total energy per state
total_all_summary <- total_all |>
  group_by(State, Year) |>
  summarise(Total_Energy_Use = sum(Energy_Use, na.rm = TRUE))
```

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

```
# Compute total renewable energy per state
renewable_all <- total_all |>
  filter(Energy_Source %in% c(
    "total_renewable_energy",
    "total_renewables",
    "total renewable-energy")) |>
```

```
  select(State, Year, Renewable_Energy_Use = Energy_Use)

# join them to compute renewable share properly
energy_joined <- total_all_summary |>
  left_join(renewable_all, by = c("State", "Year")) |>
  mutate(Renewable_Share = Renewable_Energy_Use / Total_Energy_Use)

# Convert state abbreviations to full names
state_fulls <- c(state.name, "District of Columbia")
state_abbs  <- c(state.abb, "DC")
# Build lookup table
state_lookup <- tibble(
  state_abbr = state_abbs,
  state_name = state_fulls
)


#  clean abbreviated states
energy_clean <- energy_joined |>
  left_join(state_lookup, by = c("State" = "state_abbr")) |>
  mutate(State = state_name) |>
  select(-state_name)

#final clean table with all the data!!
head(energy_clean)
```

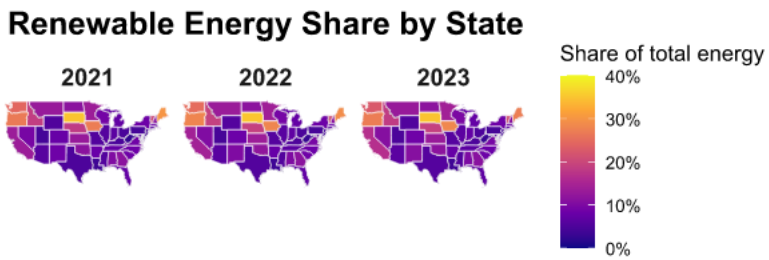
```
# A tibble: 6 × 5
# Groups:   State [2]
  State    Year Total_Energy_Use Renewable_Energy_Use Renewable_Share
  <chr>   <dbl>            <dbl>                <dbl>           <dbl>
1 Alaska   2021           684975                 9597          0.0140
2 Alaska   2022           730276                10410          0.0143
3 Alaska   2023           746979                10087          0.0135
4 Alabama  2021          2352656               239817          0.102
5 Alabama  2022          2337513               232035          0.0993
6 Alabama  2023          2265008               222189          0.0981
```

Part 4: Mapping Visualization

```
# loading US states geometry and excluding territories
us_states <- ne_states(country = "United States of America", returnclass =
"sf") |>
  filter(!name %in% c("Puerto Rico", "Guam", "American Samoa",
                      "Northern Mariana Islands", "United States Virgin
Islands"))
```

```r
# Join the data to the sf object by full state name
us_map <- us_states |>
  left_join(energy_clean, by = c("name" = "State"))

#plotting renewable share over the years
ggplot(us_map) +
  geom_sf(aes(fill = Renewable_Share), color = "white", size = 0.2) +
  facet_wrap(~Year) +
  scale_fill_viridis_c(
  name = "Share of total energy",
  labels = scales::percent_format(accuracy = 1),
  option = "C",
  limits = c(0, 0.4)  # fix the same scale across facets
)+
  coord_sf(
    xlim = c(-125, -66),   # restricts to continental U.S.
    ylim = c(24, 50),      # zooms in to main latitudes
    expand = FALSE
  ) +
  theme_minimal() +
  labs(
    title = "Renewable Energy Share by State",
  ) +
  theme(
    panel.grid.major = element_blank(),
    axis.text = element_blank(),
    axis.ticks = element_blank(),
    strip.text = element_text(size = 12, face = "bold"),
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
    legend.position = "right"
  )
```



Analysis

From the map, we can see that most states do not have a significant change in their renewable energy share of total energy between 2021-2023. However, we do see that there are signifcant differences in renewable energy share across each state.

It seems that the West coast and midwest states have higher renewable shares than the South or Northeast (beside Maine and Vermont, which also have high renewable shares.) Significant West coast states with large shares are Oregon, Washington, and Idaho. Significant midwest states are South Dakota, Nebraska, and Iowa. In fact, South Dakota consistently has the largest renewable energy share of all the states, indicated by the bright yellow color.

Another interesting trend is that California's renewable energy share did grow from 2021-2023, indicated by the lighter purple shade in 2023. It is also interesting that the California renewable energy share is comparatively lower than other states, as California is known to produce a large amount of renewable energy.