EV Power - Lab 4 Project Report

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
```

```
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.
```

Part 1: Defining Research Question

Chosen Question:How has renewable energy use changed from 2021 to 2023 across states? This question compares renewable energy use across U.S. states over three years. It examines which states increased or decreased their use from 2021 to 2023. The goal is to understand overall trends in renewable energy adoption nationwide.

Part 2: Data Preparation and Cleaning

```
renew_2021_clean <- renew_2021|>
    rename_with(~ str_to_lower(.)) |>
    rename_with(~ str_replace_all(., "[[:space:]|\\.]+", "_")) |>
    rename_with(~ str_replace_all(., "[^a-z0-9_]", "")) |>
    mutate(state_abbr = str_to_upper(state),state_abbr =
str_trim(state_abbr))|> # str_trim standardizes the state abbreviation column
by removing any spaces, tabs, or newlines from the start and end of each state
string.
    select(-state)|>
    relocate(state_abbr)
head(renew_2021_clean)
```

```
# A tibble: 6 × 3
  state_abbr energy_source renewable_use_2021
  <chr>      <chr>         <chr>
1 AK         Biomass       ≈3153
2 AK         Geothermal    186 MMBtu
3 AK         Hydropower    5763 about
4 AK         Solar Energy  ~45
```

```
5 AK         Wind Energy   451 USD
6 AL         Biomass       198543 est.
```

```
renew_2022_clean <- renew_2022 |>
    rename_with(~ str_to_lower(.)) |>
    rename_with(~ str_replace_all(., "[[:space:]|\\.]+", "_")) |>
    rename_with(~ str_replace_all(., "[^a-z0-9_]", "")) |>
    pivot_longer(cols = !energy_source,names_to = "state_abbr",values_to =
"consumption" ) |>
    mutate(state_abbr = str_to_upper(state_abbr)) |>
    relocate(state_abbr)
head(renew_2022_clean)
```

```
# A tibble: 6 × 3
  state_abbr          energy_source consumption
  <chr>               <chr>         <chr>
1 STATE               Biomass       AK
2 RENEWABLE_USE_2022 Biomass       ≈3846
3 STATE               Geothermal    AK
4 RENEWABLE_USE_2022 Geothermal    $186
5 STATE               Hydropower    AK
6 RENEWABLE_USE_2022 Hydropower    $5846
```

```
renew_2023_clean <- renew_2023|>
    rename_with(~ str_to_lower(.)) |>
    rename_with(~ str_replace_all(., "[[:space:]|\\.]+", "_")) |>
    rename_with(~ str_replace_all(., "[^a-z0-9_]", "")) |>
    mutate(state_abbr = str_to_upper(state),state_abbr =
str_trim(state_abbr))|>
    select(-state)|>
    relocate(state_abbr)
head(renew_2023_clean)
```

```
# A tibble: 6 × 3
  state_abbr energy_source renewable_use_2023
  <chr>      <chr>         <chr>
1 AK         Biomass       3404 kWh
2 AK         Geothermal    186.0
3 AK         Hydropower    6051
4 AK         Solar Energy  67
5 AK         Wind Energy   380
6 AL         Biomass       189040 kWh
```

Part 3: Joining / Pivoting Datasets for Analysis

```
renew_joined <- renew_2021_clean |>
  rename(value_2021 = renewable_use_2021) |>
  left_join(renew_2022_clean |> rename(value_2022 = consumption) |>
              select(state_abbr, energy_source, value_2022),
            by = c("state_abbr", "energy_source")) |>
  left_join(renew_2023_clean |> rename(value_2023 = renewable_use_2023) |>
              select(state_abbr, energy_source, value_2023),
            by = c("state_abbr", "energy_source"))

head(renew_joined)
```

```
# A tibble: 6 × 5
  state_abbr energy_source value_2021  value_2022 value_2023
  <chr>      <chr>         <chr>       <chr>      <chr>
1 AK         Biomass       ≈3153       <NA>       3404 kWh
2 AK         Geothermal    186 MMBtu   <NA>       186.0
3 AK         Hydropower    5763 about  <NA>       6051
4 AK         Solar Energy  ~45         <NA>       67
5 AK         Wind Energy   451 USD     <NA>       380
6 AL         Biomass       198543 est. <NA>       189040 kWh
```

Part 4: Mapping Visualization

```
library(ggplot2)
library(maps)
```

```

Attaching package: 'maps'
```

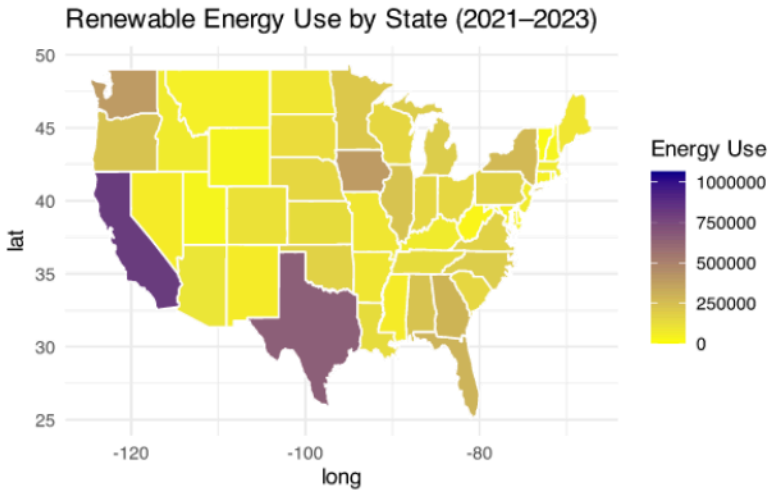
```
The following object is masked from 'package:purrr':

    map
```

```
library(dplyr)
library(tidyr)

us_map <- map_data("state")

renew_sum <- renew_joined |>
  mutate(across(starts_with("value_"), ~ as.numeric(gsub("[^0-9\\.]",
"", .)))) |>
  group_by(state_abbr) |>
```

```
  summarise(value_2021 = sum(value_2021, na.rm = TRUE),value_2022 =
sum(value_2022, na.rm = TRUE),value_2023 = sum(value_2023, na.rm = TRUE)) |>
  mutate(region = tolower(state.name[match(state_abr, state.abb)])) |>
  pivot_longer(cols = starts_with("value_"),names_to = "year",names_prefix =
"value_", values_to = "energy_use")
renew_map <- left_join(us_map, renew_sum, by = "region")
```

```
Warning in left_join(us_map, renew_sum, by = "region"): Detected an unexpected
many-to-many relationship between `x` and `y`.
i Row 1 of `x` matches multiple rows in `y`.
i Row 4 of `y` matches multiple rows in `x`.
i If a many-to-many relationship is expected, set `relationship =
  "many-to-many"` to silence this warning.
```

```
ggplot(renew_map, aes(long, lat, group = group, fill = energy_use)) +
  geom_polygon(color = "white") + # Draws each state as a filled polygon,
using longitude/latitude coordinates
  scale_fill_continuous(low = "yellow", high = "darkblue", name = "Energy
Use") +
  labs(title = "Renewable Energy Use by State (2021–2023)") +
  theme_minimal()
```



Analysis: From 2021 to 2023, most states showed a gradual increase in renewable energy use. Western states such as California and Oregon had higher total renewable consumption, while southern states remained lower overall. The map highlights clear regional patterns, showing where renewable adoption has grown most rapidly across the country.