# EV Power - Lab 4 Project Report

## Project 4: EV Power - Colin Choe

### Part 0: libraries

```
library(knitr)
knitr::opts_chunk$set(dev = "ragg_png", dpi = 200, fig.width = 7, fig.height =
4.3, fig.align = "center")
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
```

```
library(readr)
library(dplyr)
library(ggplot2)
library(stringr)
```

### Part 1: Defining Research Question

Chosen Question: **Which states increased their renewable energy use most from 2022 to 2023?**

### Part 2: Data Preparation and Cleaning

```
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.
```

```
clean_numbers <- function(x) {
  x |>
    str_replace_all("[^0-9.]", "") |>  # keep only digits and periods
    as.numeric()
}

names(renew_2022) <- names(renew_2022) |>
  tolower() |>
  str_replace_all("[- ]", "_")

names(renew_2023) <- names(renew_2023) |>
  tolower() |>
  str_replace_all("[- ]", "_")


renew_2022 <- renew_2022 %>%
     rename(state = state, source = energy_source, renewable_use =
renewable_use_2022) %>%
  mutate(
    state = str_to_upper(state),
    source = str_to_title(source),
    renewable_use = clean_numbers(renewable_use)
  )

renew_2023 <- renew_2023 %>%
     rename(state = state, source = energy_source, renewable_use =
renewable_use_2023) %>%
  mutate(
```

```
    state = str_to_upper(state),
    source = str_to_title(source),
    renewable_use = clean_numbers(renewable_use)
  )
```

## Part 3: Joining / Pivoting Datasets for Analysis

```
renew_2022_state <- renew_2022 %>%
  group_by(state) %>%
  summarise(total_renew_2022 = sum(renewable_use, na.rm = TRUE))

head(renew_2022_state, 10)
```

```
# A tibble: 10 × 2
   state total_renew_2022
   <chr>            <dbl>
 1 AK               10410
 2 AL              232035
 3 AR               90824
 4 AZ              101214
 5 CA              880995
 6 CO              114918
 7 CT               49084
 8 DC                2623
 9 DE                7402
10 FL              304605
```

```
renew_2023_state <- renew_2023 %>%
  group_by(state) %>%
  summarise(total_renew_2023 = sum(renewable_use, na.rm = TRUE))

head(renew_2023_state, 10)
```

```
# A tibble: 10 × 2
   state total_renew_2023
   <chr>            <dbl>
 1 AK               10088
 2 AL              222189
 3 AR               87277
 4 AZ              108445
 5 CA             1065179
 6 CO              115062
 7 CT               48983
 8 DC                2796
```

```
 9 DE                8040
10 FL              286307
```

```
renew_compare <- renew_2022_state %>%
  inner_join(renew_2023_state, by = "state")

renew_compare <- renew_compare %>%
  mutate(
    absolute_change = total_renew_2023 - total_renew_2022,
    percent_change = ((total_renew_2023 - total_renew_2022) / total_renew_2022)
* 100
  )

renew_compare <- renew_compare %>%
  arrange(desc(absolute_change))

head(renew_compare, 10)
```

```
# A tibble: 10 × 5
   state total_renew_2022 total_renew_2023 absolute_change percent_change
   <chr>            <dbl>            <dbl>           <dbl>          <dbl>
 1 CA              880995          1065179          184184         20.9
 2 US             8107355          8187317           79962          0.986
 3 TX              751680           791210           39530          5.26
 4 AZ              101214           108445            7231          7.14
 5 NY              269883           272968            3085          1.14
 6 NM               77285            80278            2993          3.87
 7 MD               51256            53711            2455          4.79
 8 UT               37369            39675            2306          6.17
 9 NV               72735            74879            2144          2.95
10 IN              170987           172891            1904          1.11
```

## Part 4: Mapping Visualization

```
library(sf)
```

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

```
library(scales)
```

```

Attaching package: 'scales'
```

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

    discard
```

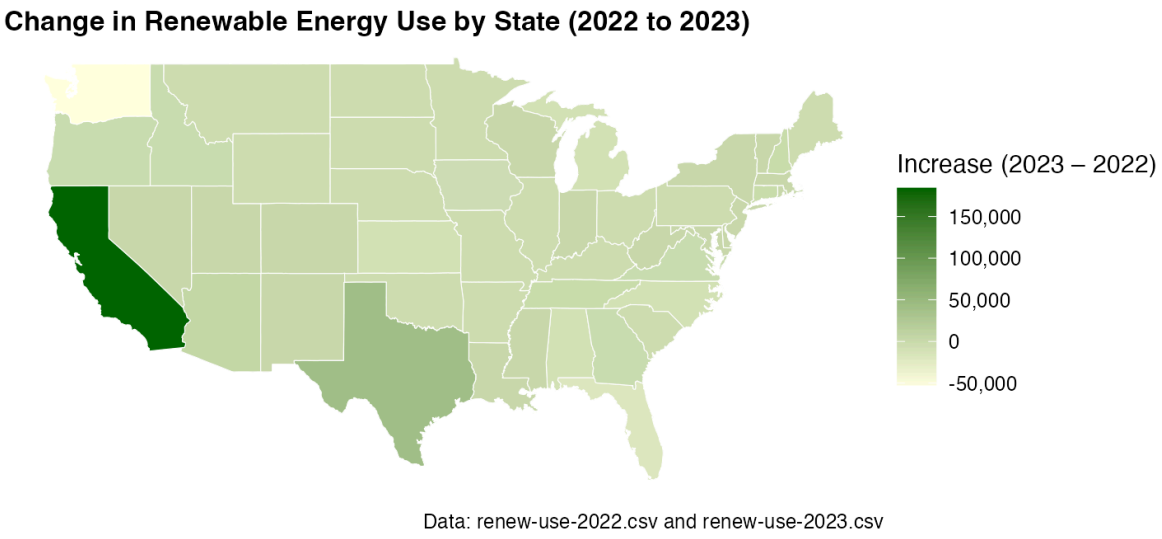
```
The following object is masked from 'package:readr':

    col_factor
```

```
library(maps)
```

```

Attaching package: 'maps'
```

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

    map
```

```
states_sf <- st_as_sf(map("state", plot = FALSE, fill = TRUE))

renew_compare_clean <- renew_compare %>%
  mutate(state_name = tolower(state.name[match(toupper(state), state.abb)]))

map_data <- states_sf %>%
  left_join(renew_compare_clean, by = c("ID" = "state_name"))

ggplot(map_data) +
  geom_sf(aes(
    fill = absolute_change),
    color = "white",
    linewidth = 0.2) +
  scale_fill_gradient(
    low = "lightyellow",
    high = "darkgreen",
    name = "Increase (2023 – 2022)",
    labels = comma) +
  labs(
    title = "Change in Renewable Energy Use by State (2022 to 2023)",
    caption = "Data: renew-use-2022.csv and renew-use-2023.csv"
  ) +
  theme_void() +
  theme(
    legend.position = "right",
    plot.title = element_text(size = 12, face = "bold")
  )
```




## Part 5: Analysis

Analysis: The map shows that renewable energy use increased the most in states like California, which clearly stands out from the rest of the country. Most other states saw smaller or steady changes, especially in the Midwest and South. Thi suggests that while renewable energy is growing overall, most of the progress is driven by a few key states with stronger clean energy policies. It helps visualize where that growth is happening and how uneven it still is across the U.S.