Project 4:Clean Energy and Electric Vehicles

AUTHOR
Fiona Li

```
library(readr)
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
library(tidyr)
library(stringr)
library(ggplot2)
library(maps)
```

```
Warning: package 'maps' was built under R version 4.3.3
```

```
av_energy_price_2021_2023 <- read_csv("Downloads/av-energy-pric
```

```
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.
```

```r
ev_registrations_by_state_2023 <- read_csv("Downloads/ev-regist
```

```
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`
```

```r
renew_use_2021 <- read_csv("Downloads/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.
```

```r
renew_use_2022 <- read_csv("Downloads/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.
```

```r
renew_use_2023 <- read_csv("Downloads/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
total_use_2021 <- read_csv("Downloads/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
total_use_2022 <- read_csv("Downloads/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
total_use_2023 <- read_csv("Downloads/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.
```

Part 1: Defining Research Questions

Overview: Introducing Sub Questions 1. Energy Price vs. Renewable Share: Are states with cheaper electricity more or less renewable-powered? 2.EV Adoption vs. Clean Energy: Are electric vehicles more common in states with cleaner (more renewable) energy mixes? 3. Renewable Growth Over Time: Which states have seen the fastest growth in renewable energy use from 2021 to 2023?

Part 2: Data Preparation & Cleaning Cleaning the Renew CSVs

```
clean_renew <- function(df, year_col) {
  df %>%
    rename_with(~ c("state", "energy_source", "value"), everythi
    mutate(year = year_col,
           value = as.numeric(str_replace_all(value, "[^0-9\\.]'
    group_by(state, year) %>%
    summarise(renewable_energy_use = sum(value, na.rm = TRUE), 
}

renew_all <- bind_rows(
  clean_renew(renew_use_2021, 2021),
  clean_renew(renew_use_2022, 2022),
  clean_renew(renew_use_2023, 2023)
)

head(renew_all)
```

```
# A tibble: 6 × 3
  state  year renewable_energy_use
  <chr> <dbl>                <dbl>
```

```
1 AK      2021              9598
2 AL      2021            239816
3 AR      2021             89714
4 AZ      2021             99266
5 CA      2021            810020
6 CO      2021            103956
```

Cleaning the Total CSVs

```
clean_total <- function(df, year_col) {
  df %>%
    pivot_longer(-Energy_Source, names_to = "state", values_to =
    group_by(state) %>%
    summarise(total_energy_use = sum(value, na.rm = TRUE), .grou
    mutate(year = year_col)
}

total_all <- bind_rows(
  clean_total(total_use_2021, 2021),
  clean_total(total_use_2022, 2022),
  clean_total(total_use_2023, 2023)
)

head(total_all)
```

```
# A tibble: 6 × 3
  state total_energy_use  year
  <chr>            <dbl> <dbl>
1 AK              684975  2021
2 AL             2352656  2021
3 AR             1136025  2021
4 AZ             1681257  2021
5 CA             6142252  2021
6 CO             1364155  2021
```

Cleaning the EV Registrations

```
ev_clean <- ev_registrations_by_state_2023 %>%
  rename(
    state = `electric vehicle registrations_by_state (2023)`,
    ev_registrations = `...2`
  ) %>%
  mutate(
    ev_registrations = as.numeric(str_replace_all(ev_registrati
```

```
  )

head(ev_clean)
```

```
# A tibble: 6 × 2
  state    ev_registrations
  <chr>               <dbl>
1 <NA>                   NA
2 STATE                  NA
3 Alabama             13047
4 Alaska               2697
5 Arizona             89798
6 Arkansas             7108
```

Cleaning the Prices

```
price_clean <- read_csv("Downloads/av-energy-price-2021-2023.csv
  separate(X1, into = c("state", "p2021", "p2022", "p2023"), sep
  filter(!str_detect(state, "State|Total|US"), !is.na(p2021)) %>
  mutate(across(c(p2021, p2022, p2023), ~as.numeric(str_extract
  pivot_longer(cols = starts_with("p"), names_to = "year", value
  mutate(year = as.numeric(str_remove(year, "p"))) %>%
  arrange(state, year)
```

```
Rows: 53 Columns: 1
── Column specification ────────────────────────────────────────
Delimiter: ","
chr (1): X1

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(price_clean)
```

```
# A tibble: 6 × 3
  state  year average_price
  <chr> <dbl>         <dbl>
1 AK     2021          20.0
2 AK     2022          27.3
3 AK     2023          23.8
```

```
4 AL     2021          17.8
5 AL     2022          23.4
6 AL     2023          21.1
```

Part 3: Joining / Pivoting Datasets for Analysis Question 1: Are states with cheaper electricity more or less renewable-powered?

I joined the renewable energy, total energy use, and average energy price datasets by state and year. This allowed me to calculate the renewable energy share as: renewable_share = renewable_energy_use / total_energy_use This variable represents the percentage of each state's total energy use that comes from renewable sources. I used this combined table to compare average_price against renewable_share across years.

```
energy_joined <- renew_all %>%
  left_join(total_all, by = c("state", "year")) %>%
  left_join(price_clean, by = c("state", "year")) %>%
  mutate(renewable_share = renewable_energy_use / total_energy_u

energy_joined
```

```
# A tibble: 156 × 6
   state  year renewable_energy_use total_energy_use
average_price
   <chr> <dbl>                <dbl>            <dbl>
<dbl>
 1 AK     2021                 9598           684975
20.0
 2 AL     2021               239816          2352656
17.8
 3 AR     2021                89714          1136025
18.4
 4 AZ     2021                99266          1681257
25.1
 5 CA     2021               810020          6142252
28.4
 6 CO     2021               103956          1364155
20.6
 7 CT     2021                49306           821709
25.8
 8 DC     2021                 2487            49262
25.7
 9 DE     2021                 7151           208041
```

```
21.8
10 FL     2021                297290          4145505
22.5
# ℹ 146 more rows
# ℹ 1 more variable: renewable_share <dbl>
```

Question 2: Are electric vehicles more common in states with cleaner energy mixes?

For this question, I merged the 2023 renewable and total energy tables with the EV registrations dataset using the state column. This allowed me to compute the EV density relative to total energy use or renewable share.

The new variable is: ev_ratio = ev_registrations / total_energy_use

```
ev_combined <- renew_all %>%
  filter(year == 2023) %>%
  left_join(total_all %>% filter(year == 2023), by = "state") %>
  left_join(ev_clean, by = "state") %>%
  mutate(ev_ratio = ev_registrations / total_energy_use)

ev_combined
```

```
# A tibble: 52 × 7
   state year.x renewable_energy_use total_energy_use year.y
ev_registrations
   <chr>  <dbl>                <dbl>            <dbl>  <dbl>
<dbl>
 1 AK      2023                10088           746979   2023
NA
 2 AL      2023               222189          2265008   2023
NA
 3 Ar      2023                87277               NA     NA
NA
 4 CA      2023              1065179          6429818   2023
NA
 5 CO      2023               115062          1359507   2023
NA
 6 DC      2023                 2796            46323   2023
NA
 7 DE      2023                 8040           203487   2023
NA
 8 GA      2023               291462          2627553   2023
NA
```

```
 9 IA      2023               414801          1466926   2023
NA
10 ID      2023                77127           421975   2023
NA
# ℹ 42 more rows
# ℹ 1 more variable: ev_ratio <dbl>
```

Question 3: Which states have seen the fastest growth in renewable energy use from 2021 to 2023?

To measure growth, I combined all three years of renewable energy data into a single long table (renew_all) and grouped by state. Then I computed each state's percentage change in renewable energy use from 2021 to 2023.

The new variable is: renewable_growth = (renewable_use_2023 - renewable_use_2021) / renewable_use_2021

```
renew_growth <- renew_all %>%
  pivot_wider(names_from = year, values_from = renewable_energy_
  mutate(renewable_growth = (`2023` - `2021`) / `2021`)

renew_growth
```

```
# A tibble: 70 × 5
   state `2021` `2022`  `2023` renewable_growth
   <chr>  <dbl>  <dbl>   <dbl>            <dbl>
 1 AK      9598  10410   10088           0.0511
 2 AL    239816 232035  222189          -0.0735
 3 AR     89714  90824      NA          NA
 4 AZ     99266 101214      NA          NA
 5 CA    810020 880995 1065179           0.315
 6 CO    103956 114918  115062           0.107
 7 CT     49306  49084      NA          NA
 8 DC      2487   2623    2796           0.124
 9 DE      7151   7402    8040           0.124
10 FL    297290 304605      NA          NA
# ℹ 60 more rows
```

Part 4: Mapping and Dashboard Visualization

```
energy_joined <- energy_joined %>%
  mutate(state = case_when(
    state %in% state.abb ~ tolower(state.name[match(state, state
```

```
    TRUE ~ tolower(state)
  ))
```

```
energy_joined <- renew_all %>%
  left_join(total_all, by = c("state", "year")) %>%
  left_join(price_clean, by = c("state", "year")) %>%
  mutate(
    renewable_share = renewable_energy_use / total_energy_use
  )
```

```
distinct(energy_joined, state) %>% head(10)
```

```
# A tibble: 10 × 1
   state
   <chr>
 1 AK
 2 AL
 3 AR
 4 AZ
 5 CA
 6 CO
 7 CT
 8 DC
 9 DE
10 FL
```

```
energy_joined <- energy_joined %>%
  mutate(
    state = case_when(
      state %in% state.abb ~ tolower(state.name[match(state, sta
      TRUE ~ tolower(state)
    )
  )
```

```
distinct(energy_joined, state) %>% head(10)
```

```
# A tibble: 10 × 1
   state
   <chr>
 1 alaska
 2 alabama
 3 arkansas
```

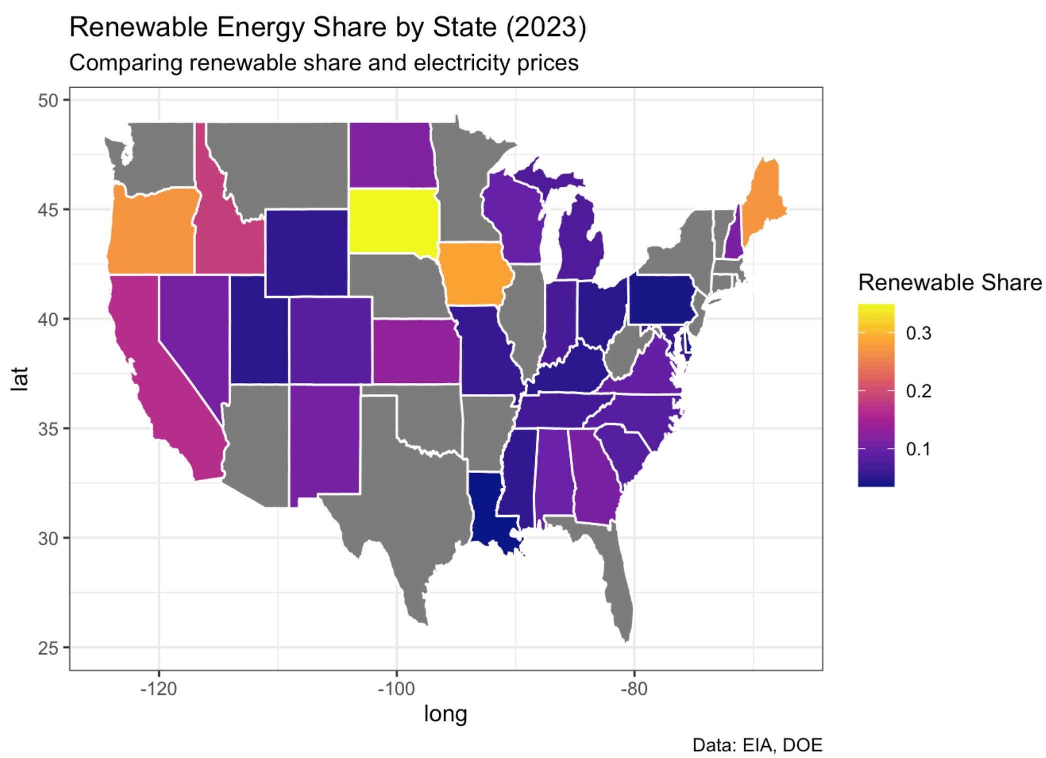
```
 4 arizona
 5 california
 6 colorado
 7 connecticut
 8 dc
 9 delaware
10 florida
```

```
states_map <- map_data("state")

map_data_2023 <- energy_joined %>% filter(year == 2023)

map_joined <- left_join(states_map, map_data_2023, by = c("regio

ggplot(map_joined, aes(long, lat, group = group, fill = renewab
  geom_polygon(color = "white") +
  scale_fill_viridis_c(option = "plasma", name = "Renewable Shar
  labs(
    title = "Renewable Energy Share by State (2023)",
    subtitle = "Comparing renewable share and electricity prices
    caption = "Data: EIA, DOE"
  ) +
  theme_bw()
```



Analysis

The map reveals clear geographic patterns in renewable energy use across the U.S. in 2023. States such as Iowa, Maine, and South Dakota have some of the highest renewable energy shares, while many Southeastern and Mid-Atlantic states remain more fossil-fuel-reliant. This visual pattern suggests that access to wind and hydro resources plays a major role in driving renewable adoption. When compared with average electricity prices, there is no simple one-to-one relationship—some states with high renewable shares still have higher prices, while others benefit from both clean and affordable energy. Overall, the map helps answer the main research question by showing that while EVs reduce direct emissions, the environmental impact of the electricity they use depends heavily on each state’s energy mix—highlighting regional disparities in how “clean” electric power truly is.