EV Power - Lab 4 Project Report

Part 0: libraries

```
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
```

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

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

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

Part 1: Defining Research Question

Chosen Question: What percent of total energy use in each state comes from clean sources?

Part 2: Data Preparation and Cleaning

```
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.
```

```
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.
```

```
renew_2021 <- renew_2021 |>
    mutate(Energy_Source = str_remove(Energy_Source, "\\s+Energy$"),
        Renewable_Use_2021 = as.numeric(str_extract(Renewable_Use_2021, "\
\d+")),
        State = toupper(State))

renew_2022 <- renew_2022 |>
    mutate(Energy_Source = str_remove(Energy_Source, "\\s+Energy$"),
        Renewable_Use_2022 = as.numeric(str_extract(Renewable_Use_2022, "\
\d+")),
        State = toupper(State))

renew_2023 <- renew_2023 |>
    mutate(Energy_Source = str_remove(Energy_Source, "\\s+Energy$"),
        Renewable_Use_2023 = as.numeric(str_extract(Renewable_Use_2023, "\
\d+")),
        State = toupper(State))

total_2021 <- total_2021 |>
    mutate(Energy_Source = str_to_title(Energy_Source),
        Energy_Source = case_when(
```

```
            str_detect(Energy_Source, "^Coal") ~ "Coal",
            str_detect(Energy_Source, "^Natural") ~ "Natural_Gas",
            str_detect(Energy_Source, "^Petroleum") ~ "Petroleum",
            str_detect(Energy_Source, "^Nuclear") ~ "Nuclear",
            str_detect(Energy_Source, "^Total") ~ "Total_Renew"))

total_2022 <- total_2022 |>
    mutate(Energy_Source = str_to_title(Energy_Source),
        Energy_Source = case_when(
            str_detect(Energy_Source, "^Coal") ~ "Coal",
            str_detect(Energy_Source, "^Natural") ~ "Natural_Gas",
            str_detect(Energy_Source, "^Petroleum") ~ "Petroleum",
            str_detect(Energy_Source, "^Nuclear") ~ "Nuclear",
            str_detect(Energy_Source, "^Total") ~ "Total_Renew"))

total_2023 <- total_2023 |>
    mutate(Energy_Source = str_to_title(Energy_Source),
        Energy_Source = case_when(
            str_detect(Energy_Source, "^Coal") ~ "Coal",
            str_detect(Energy_Source, "^Natural") ~ "Natural_Gas",
            str_detect(Energy_Source, "^Petroleum") ~ "Petroleum",
            str_detect(Energy_Source, "^Nuclear") ~ "Nuclear",
            str_detect(Energy_Source, "^Total") ~ "Total_Renew"))
```

Part 3: Joining / Pivoting Datasets for Analysis

```
# filter clean energy sources for each renew-use dataset (2021-2023)
clean_energy1 <- tibble(Energy_Source = c("Hydropower", "Wind", "Solar"))
renew_2021 <- semi_join(renew_2021, clean_energy1, by = "Energy_Source")

renew_2022 <- semi_join(renew_2022, clean_energy1, by = "Energy_Source")

renew_2023 <- semi_join(renew_2023, clean_energy1, by = "Energy_Source")


# computing total clean energy used by State for each year (for renew-use
datasets)
clean_2021 <- renew_2021 |>
    group_by(State) |>
    summarize(Clean_Energy1 = sum(Renewable_Use_2021, na.rm = TRUE))
head(clean_2021)
```

```
# A tibble: 6 × 2
  State Clean_Energy1
  <chr>         <dbl>
1 AK             6259
```

```
2 AL            41132
3 AR            15967
4 AZ            63634
5 CA           307085
6 CO            66863
```

```
clean_2022 <- renew_2022 |>
    group_by(State) |>
    summarize(Clean_Energy1 = sum(Renewable_Use_2022, na.rm = TRUE))
head(clean_2022)
```

```
# A tibble: 6 × 2
  State Clean_Energy1
  <chr>         <dbl>
1 AK             6378
2 AL            37962
3 AR            15200
4 AZ            64505
5 CA           343055
6 CO            75448
```

```
clean_2023 <- renew_2023 |>
    group_by(State) |>
    summarize(Clean_Energy1 = sum(Renewable_Use_2023, na.rm = TRUE))
head(clean_2023)
```

```
# A tibble: 6 × 2
  State Clean_Energy1
  <chr>         <dbl>
1 AK             6498
2 AL            33008
3 AR            15158
4 AZ            71528
5 CA           411994
6 CO            78790
```

```
total_2021 <- total_2021 |>
    pivot_longer(-Energy_Source, names_to = "State", values_to =
"Energy_Use_2021")
head(total_2021)
```

```
# A tibble: 6 × 3
  Energy_Source State Energy_Use_2021
```

```
  <chr>         <chr>           <dbl>
1 Coal          AK              18694
2 Coal          AL             309791
3 Coal          AR             216123
4 Coal          AZ             160299
5 Coal          CA              28244
6 Coal          CO             252442
```

```
total_2022 <- total_2022 |>
    pivot_longer(-Energy_Source, names_to = "State", values_to =
"Energy_Use_2022")
head(total_2022)
```

```
# A tibble: 6 × 3
  Energy_Source State Energy_Use_2022
  <chr>         <chr>           <dbl>
1 Coal          AK              18615
2 Coal          AL             297654
3 Coal          AR             211724
4 Coal          AZ             154007
5 Coal          CA              30049
6 Coal          CO             233256
```

```
total_2023 <- total_2023 |>
    pivot_longer(-Energy_Source, names_to = "State", values_to =
"Energy_Use_2023")
head(total_2023)
```

```
# A tibble: 6 × 3
  Energy_Source State Energy_Use_2023
  <chr>         <chr>           <dbl>
1 Coal          AK              18414
2 Coal          AL             224926
3 Coal          AR             180262
4 Coal          AZ             137885
5 Coal          CA              28746
6 Coal          CO             204826
```

```
# filter nuclear energy for each total-use dataset then add nuclear energy
consumed to the total clean energy used calculated above

nuclear_2021 <- total_2021 |>
  filter(Energy_Source == "Nuclear") |>
  select(State, Clean_Energy2 = Energy_Use_2021)
```

```
clean_2021 <- clean_2021 |>
  left_join(nuclear_2021, by = "State") |>
  mutate(Clean_Energy = Clean_Energy1 + Clean_Energy2) |>
  select(State, Clean_Energy)
head(clean_2021)
```

```
# A tibble: 6 × 2
  State Clean_Energy
  <chr>        <dbl>
1 AK            6259
2 AL          521247
3 AR          157339
4 AZ          393502
5 CA          478927
6 CO           66863
```

```
nuclear_2022 <- total_2022 |>
  filter(Energy_Source == "Nuclear") |>
  select(State, Clean_Energy2 = Energy_Use_2022)

clean_2022 <- clean_2022 |>
  left_join(nuclear_2022, by = "State") |>
  mutate(Clean_Energy = Clean_Energy1 + Clean_Energy2) |>
  select(State, Clean_Energy)
head(clean_2022)
```

```
# A tibble: 6 × 2
  State Clean_Energy
  <chr>        <dbl>
1 AK            6378
2 AL          480055
3 AR          164854
4 AZ          398243
5 CA          526869
6 CO           75448
```

```
nuclear_2023 <- total_2023 |>
  filter(Energy_Source == "Nuclear") |>
  select(State, Clean_Energy2 = Energy_Use_2023)

clean_2023 <- clean_2023 |>
  left_join(nuclear_2023, by = "State") |>
  mutate(Clean_Energy = Clean_Energy1 + Clean_Energy2) |>
```

```
  select(State, Clean_Energy)
head(clean_2023)
```

```
# A tibble: 6 × 2
  State Clean_Energy
  <chr>        <dbl>
1 AK            6498
2 AL          509400
3 AR          171650
4 AZ          401002
5 CA          597186
6 CO           78790
```

```
# Calculate total energy used by State for each year
Total_2021 <- total_2021 |>
    group_by(State) |>
    summarize(total_energy = sum(Energy_Use_2021, na.rm = TRUE))
head(Total_2021)
```

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

```
Total_2022 <- total_2022 |>
    group_by(State) |>
    summarize(total_energy = sum(Energy_Use_2022, na.rm = TRUE))
head(Total_2022)
```

```
# A tibble: 6 × 2
  State total_energy
  <chr>        <dbl>
1 AK          730276
2 AL         2337513
3 AR         1178115
4 AZ         1651857
5 CA         6244174
6 CO         1411476
```

```r
Total_2023 <- total_2023 |>
    group_by(State) |>
    summarize(total_energy = sum(Energy_Use_2023, na.rm = TRUE))
head(Total_2021)
```

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

```r
# Calculate percentage of clean energy for each year
percent_clean_2021 <- clean_2021 |>
    left_join(Total_2021, by = "State") |>
    mutate(Clean_Energy_Percent = 100 * Clean_Energy / total_energy)
head(percent_clean_2021)
```

```
# A tibble: 6 × 4
  State Clean_Energy total_energy Clean_Energy_Percent
  <chr>        <dbl>        <dbl>                <dbl>
1 AK            6259       684975                0.914
2 AL          521247      2352656               22.2
3 AR          157339      1136025               13.8
4 AZ          393502      1681257               23.4
5 CA          478927      6142252                7.80
6 CO           66863      1364155                4.90
```

```r
percent_clean_2022 <- clean_2022 |>
    left_join(Total_2022, by = "State") |>
    mutate(Clean_Energy_Percent = 100 * Clean_Energy / total_energy)
head(percent_clean_2022)
```

```
# A tibble: 6 × 4
  State Clean_Energy total_energy Clean_Energy_Percent
  <chr>        <dbl>        <dbl>                <dbl>
1 AK            6378       730276                0.873
2 AL          480055      2337513               20.5
3 AR          164854      1178115               14.0
4 AZ          398243      1651857               24.1
```

```
5 CA          526869      6244174                8.44
6 CO           75448      1411476                5.35
```

```
percent_clean_2023 <- clean_2023 |>
    left_join(Total_2023, by = "State") |>
    mutate(Clean_Energy_Percent = 100 * Clean_Energy / total_energy)
head(percent_clean_2023)
```

```
# A tibble: 6 × 4
  State Clean_Energy total_energy Clean_Energy_Percent
  <chr>        <dbl>        <dbl>                <dbl>
1 AK            6498       746979                0.870
2 AL          509400      2265008               22.5
3 AR          171650      1151062               14.9
4 AZ          401002      1712667               23.4
5 CA          597186      6429818                9.29
6 CO           78790      1359507                5.80
```

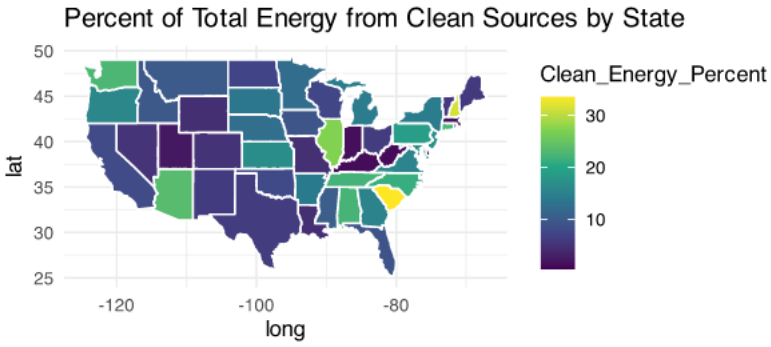
Part 4: Mapping Visualization

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

percent_clean_2021  <- percent_clean_2021 |>
  mutate(region = tolower(state.name[match(State, state.abb)]))

map_data_joined <- states_map |>
  left_join(percent_clean_2021 , by = "region")

ggplot(map_data_joined, aes(long, lat, group = group, fill =
Clean_Energy_Percent)) +
  geom_polygon(color = "white") +
  coord_fixed(1.3) +
  theme_minimal() +
  scale_fill_viridis_c()  +
  labs(title = "Percent of Total Energy from Clean Sources by State")
```



South Carolina has the highest percentage (approx. 33%) of total energy coming from clean sources compared to all other states. Several states have very low percentage of total energy coming from clean sources.