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
```

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

```
Warning: package 'sf' was built under R version 4.5.2
```

```
Linking to GEOS 3.13.1, GDAL 3.11.4, PROJ 9.7.0; sf_use_s2() is TRUE
```

```
library(maps)
```

```
Warning: package 'maps' was built under R version 4.5.2
```

```

Attaching package: 'maps'

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

```
    map
```

Part 1: Defining Research Question

Chosen Question: Do states with more EV registrations tend to have cheaper electricity?

Part 2: Data Preparation and Cleaning

```
# fixing messy data from csv: av-energy

# read in the messy CSV (skip extra rows if needed)
energy_cost_raw <- read_csv("data/av-energy-price-2021-2023.csv", skip = 2)
```

```
Rows: 52 Columns: 1
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (1): State,2021,2022,2023

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.
```

```
energy_cost_raw
```

```
# A tibble: 52 × 1
   `State,2021,2022,2023`
   <chr>
 1 AK,$20.03 per MMBtu,$27.33,$23.84 est.
 2 AL,about 17.85 USD,23.37 USD,≈21.11
 3 AR,$18.42,$23.84 per MMBtu,$21.76
 4 AZ,≈25.07,31.72 USD,about 30.28
 5 CA,$28.44,$37.35,$35.72 per MMBtu
 6 CO,20.64 USD,≈25.85,23.85
 7 CT,about $25.85,$33.15,$32.32 est.
 8 DC,≈25.67,$30.84,about 32.28 USD
 9 DE,$21.83,$27.74 per MMBtu,$26.70
10 FL,≈22.53,29.35 USD,$28.12
# ℹ 42 more rows
```

```
# Split the one column into four depending on where there is a comma
energy_cost_raw <- separate(
    energy_cost_raw,
```

```
    col = 1,
    into = c("State", "2021", "2022", "2023"),
    sep = ","
)

# preview data
head(energy_cost_raw)
```

```
# A tibble: 6 × 4
  State `2021`           `2022`           `2023`
  <chr> <chr>            <chr>            <chr>
1 AK    $20.03 per MMBtu $27.33           $23.84 est.
2 AL    about 17.85 USD  23.37 USD        ≈21.11
3 AR    $18.42           $23.84 per MMBtu $21.76
4 AZ    ≈25.07           31.72 USD        about 30.28
5 CA    $28.44           $37.35           $35.72 per MMBtu
6 CO    20.64 USD        ≈25.85           23.85
```

```
# clean the table using regex
energy_cost <- energy_cost_raw |>
    mutate(across(
        c(`2021`, `2022`, `2023`),
        ~ as.numeric(str_extract(.x, "\\d+\\.?\\d*"))
    ))
head(energy_cost)
```

```
# A tibble: 6 × 4
  State `2021` `2022` `2023`
  <chr>  <dbl>  <dbl>  <dbl>
1 AK      20.0   27.3   23.8
2 AL      17.8   23.4   21.1
3 AR      18.4   23.8   21.8
4 AZ      25.1   31.7   30.3
5 CA      28.4   37.4   35.7
6 CO      20.6   25.8   23.8
```

```
# fixing messy data from csv: messy-ev-registrations
ev_reg <- read_csv("data/ev-registrations-by-state-2023.csv", skip = 2)
```

```
Rows: 52 Columns: 2
── Column specification ────────────────────────────────────────
Delimiter: ","
chr (2): STATE, Count-EVs
```

```
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.
```

```
ev_reg
```

```
# A tibble: 52 × 2
   STATE                `Count-EVs`
   <chr>                <chr>
 1 Alabama              #13047
 2 Alaska               ~2697
 3 Arizona              89798
 4 Arkansas             7108 EVs
 5 California           1256646
 6 Colorado             90083
 7 Connecticut          EVs 31557
 8 Delaware             8435
 9 District of Columbia 8066
10 Florida              254878
# ℹ 42 more rows
```

```
# cleaning ev_reg
names(ev_reg) <- names(ev_reg) |>
    str_trim() |>
    str_to_lower() |>
    str_replace_all("[-_]", " ") |>
    str_replace_all("\\s+", "_") |>
    str_remove_all("\\(.*\\)") # remove units like (2023)

head(ev_reg)
```

```
# A tibble: 6 × 2
  state      count_evs
  <chr>      <chr>
1 Alabama    #13047
2 Alaska     ~2697
3 Arizona    89798
4 Arkansas   7108 EVs
5 California 1256646
6 Colorado   90083
```

```r
ev_reg <- ev_reg |>
    mutate(count_evs = str_extract(count_evs, "\\d+") %>% as.numeric())
head(ev_reg)
```

```
# A tibble: 6 × 2
  state      count_evs
  <chr>          <dbl>
1 Alabama        13047
2 Alaska          2697
3 Arizona        89798
4 Arkansas        7108
5 California   1256646
6 Colorado       90083
```

```r
# choosing the year 2023 because only have one year of data for EV
registrations

# state names are not equal, so must first change state names so they can be
joined
# adding DC to list of state names
state_fulls <- c(datasets::state.name, "District of Columbia")
state_fulls
```

```
 [1] "Alabama"              "Alaska"               "Arizona"
 [4] "Arkansas"             "California"           "Colorado"
 [7] "Connecticut"          "Delaware"             "Florida"
[10] "Georgia"              "Hawaii"               "Idaho"
[13] "Illinois"             "Indiana"              "Iowa"
[16] "Kansas"               "Kentucky"             "Louisiana"
[19] "Maine"                "Maryland"             "Massachusetts"
[22] "Michigan"             "Minnesota"            "Mississippi"
[25] "Missouri"             "Montana"              "Nebraska"
[28] "Nevada"               "New Hampshire"        "New Jersey"
[31] "New Mexico"           "New York"             "North Carolina"
[34] "North Dakota"         "Ohio"                 "Oklahoma"
[37] "Oregon"               "Pennsylvania"         "Rhode Island"
[40] "South Carolina"       "South Dakota"         "Tennessee"
[43] "Texas"                "Utah"                 "Vermont"
[46] "Virginia"             "Washington"           "West Virginia"
[49] "Wisconsin"            "Wyoming"              "District of Columbia"
```

```r
state_abbs <- c(datasets::state.abb, "DC")
state_abbs
```

```
 [1] "AL" "AK" "AZ" "AR" "CA" "CO" "CT" "DE" "FL" "GA" "HI" "ID" "IL" "IN"
"IA"
[16] "KS" "KY" "LA" "ME" "MD" "MA" "MI" "MN" "MS" "MO" "MT" "NE" "NV" "NH"
"NJ"
[31] "NM" "NY" "NC" "ND" "OH" "OK" "OR" "PA" "RI" "SC" "SD" "TN" "TX" "UT"
"VT"
[46] "VA" "WA" "WV" "WI" "WY" "DC"
```

```
state_lookup <- tibble(state_fulls, state_abbs) # maybe
state_lookup # maybe
```

```
# A tibble: 51 × 2
   state_fulls state_abbs
   <chr>       <chr>
 1 Alabama     AL
 2 Alaska      AK
 3 Arizona     AZ
 4 Arkansas    AR
 5 California  CA
 6 Colorado    CO
 7 Connecticut CT
 8 Delaware    DE
 9 Florida     FL
10 Georgia     GA
# ℹ 41 more rows
```

```
# matching state abbreviations to full names and replacing full names with
abbrvs
matching <- match(energy_cost$State, state_abbs)
matching
```

```
 [1]  2  1  4  3  5  6  7 51  8  9 10 11 15 12 13 14 16 17 18 21 20 19 22 23
25
[26] 24 26 33 34 27 29 30 31 28 32 35 36 37 38 39 40 41 42 43 44 46 45 47 49
48
[51] 50 NA
```

```
energy_cost$State <- ifelse(is.na(matching), ev_reg$State,
state_fulls[matching])
```

```
Error in ans[ypos] <- rep(yes, length.out = len)[ypos]: replacement has length
zero
```

```r
head(energy_cost)
```

```
# A tibble: 6 × 4
  State `2021` `2022` `2023`
  <chr>  <dbl>  <dbl>  <dbl>
1 AK      20.0   27.3   23.8
2 AL      17.8   23.4   21.1
3 AR      18.4   23.8   21.8
4 AZ      25.1   31.7   30.3
5 CA      28.4   37.4   35.7
6 CO      20.6   25.8   23.8
```

```r
# subsetting only 2023 year of energy costs
energy_cost_2023 <- energy_cost[,c(1,4)]

# renaming columns
colnames(energy_cost_2023) <- c("state", "cost")
energy_cost_2023
```

```
# A tibble: 52 × 2
   state  cost
   <chr> <dbl>
 1 AK     23.8
 2 AL     21.1
 3 AR     21.8
 4 AZ     30.3
 5 CA     35.7
 6 CO     23.8
 7 CT     32.3
 8 DC     32.3
 9 DE     26.7
10 FL     28.1
# ℹ 42 more rows
```

```r
colnames(ev_reg) <- (c("state", "count_of_regs"))

ev_reg
```

```
# A tibble: 52 × 2
   state               count_of_regs
   <chr>                       <dbl>
 1 Alabama                     13047
 2 Alaska                       2697
 3 Arizona                     89798
```

```
 4 Arkansas                      7108
 5 California                 1256646
 6 Colorado                     90083
 7 Connecticut                  31557
 8 Delaware                      8435
 9 District of Columbia          8066
10 Florida                     254878
# ℹ 42 more rows
```

```
energy_cost_2023
```

```
# A tibble: 52 × 2
   state  cost
   <chr> <dbl>
 1 AK     23.8
 2 AL     21.1
 3 AR     21.8
 4 AZ     30.3
 5 CA     35.7
 6 CO     23.8
 7 CT     32.3
 8 DC     32.3
 9 DE     26.7
10 FL     28.1
# ℹ 42 more rows
```

Part 3: Joining / Pivoting Datasets for Analysis

```
# Now that the tables have the same state names, join them:

ev_reg_to_energy_cost <- full_join(ev_reg, energy_cost_2023, by = "state")

#remove total column at bottom
ev_reg_to_energy_cost <- ev_reg_to_energy_cost[1:51,]
print(ev_reg_to_energy_cost, n=52)
```

```
# A tibble: 51 × 3
   state                count_of_regs  cost
   <chr>                        <dbl> <dbl>
 1 Alabama                      13047    NA
 2 Alaska                        2697    NA
 3 Arizona                      89798    NA
 4 Arkansas                      7108    NA
 5 California                 1256646    NA
```

```
 6 Colorado                   90083    NA
 7 Connecticut                31557    NA
 8 Delaware                    8435    NA
 9 District of Columbia        8066    NA
10 Florida                   254878    NA
11 Georgia                    92368    NA
12 Hawaii                     25565    NA
13 Idaho                       8501    NA
14 Illinois                   99573    NA
15 Indiana                    26101    NA
16 Iowa                        9031    NA
17 Kansas                     11271    NA
18 Kentucky                   11617    NA
19 Louisiana                   8150    NA
20 Maine                       7377    NA
21 Maryland                   72139    NA
22 Massachusetts              73768    NA
23 Michigan                   50284    NA
24 Minnesota                  37050    NA
25 Mississippi                 3590    NA
26 Missouri                   26861    NA
27 Montana                     4608    NA
28 Nebraska                    6920    NA
29 Nevada                     47361    NA
30 New Hampshire               9861    NA
31 New Jersey                134753    NA
32 New Mexico                 10276    NA
33 New York                  131250    NA
34 North Carolina             70164    NA
35 North Dakota                 959    NA
36 Ohio                       50393    NA
37 Oklahoma                   22843    NA
38 Oregon                     64361    NA
39 Pennsylvania               70154    NA
40 Rhode Island                6396    NA
41 South Carolina             20873    NA
42 South Dakota                1675    NA
43 Tennessee                  33221    NA
44 Texas                     230125    NA
45 Utah                       39998    NA
46 Vermont                     7816    NA
47 Virginia                   84936    NA
48 Washington                152101    NA
49 West Virginia               2758    NA
50 Wisconsin                  24943    NA
51 Wyoming                     1139    NA
```

Part 4: Mapping Visualization

```
length(energy_cost_2023$cost)
```

```
[1] 52
```

```
energy_cost_2023$cost[52]
```

```
[1] 23.59
```

```
us_average_cost <- mean(ev_reg_to_energy_cost$cost)
us_average_cost
```

```
[1] NA
```

```
ev_reg_to_energy_cost <- ev_reg_to_energy_cost |>
    mutate(energy_cost_vs_average = (cost / us_average_cost))
```

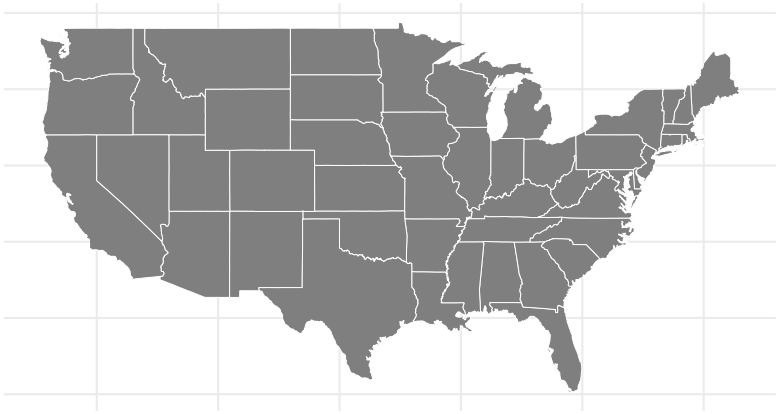
```
# mapping ratio of energy cost

# create map of US states
states_map <- st_as_sf(map("state", plot = FALSE, fill = TRUE))

# Match state names: map names are lowercase
ev_reg_to_energy_cost$state <- tolower(ev_reg_to_energy_cost$state)
# Join map data
ratio_map <- states_map |>
    left_join(ev_reg_to_energy_cost, by = c("ID" = "state"))
```

```
# map of energy cost in states
ggplot(ratio_map) +
    geom_sf(aes(fill = energy_cost_vs_average), color = "white") +
    scale_fill_gradient2(
        low = "green", mid = "white", high = "red",
        midpoint = 1
    ) +
    labs(
        title = "Electricity Price by State Compared to Average",
        subtitle = "Red = above US average, Green = below US average"
    ) +
    theme_minimal() +
```

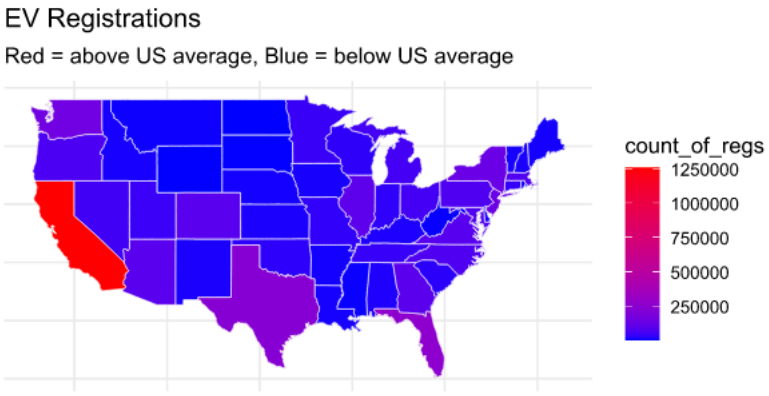
```
    theme(
        axis.text = element_blank(),
        axis.ticks = element_blank()
)
```

Electricity Price by State Compared to Average

Red = above US average, Green = below US average



```
# map of number of EVs
ggplot(ratio_map) +
    geom_sf(aes(fill = count_of_regs), color = "white") +
    scale_fill_gradient(
        low = "blue", high = "red"
    ) +
    labs(
        title = "EV Registrations",
        subtitle = "Red = above US average, Blue = below US average"
    ) +
    theme_minimal() +
    theme(
        axis.text = element_blank(),
        axis.ticks = element_blank()
)
```



Part 5: Analysis

These graphs suggest a weak relationship between EV registrations and the cost of energy compared to the national average. This effect appears negligible since electricity is still cheaper than gasoline. Perhaps a more meaningful comparison would be EV registrations per state compared to that state's cost of gasoline.

However, there does seem to be a geographic relationship to EV registrations. The highest levels of EV registrations is at the states closest to the ocean, as well as states closer to the equator. California, a leader in renewable energy, stands out as leading the EV industry as well.