EV Power - Lab 4 Project Report

Example Solution 1

Part 0: libraries

```
#|eval = FALSE
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(ggplot2)
library(dplyr)
library(readr)
library(stringr)
energy_price = read_csv("data/av-energy-price-2021-2023.csv")
```

```
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.
```

```
total_use_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.
```

Part 1: Defining Research Question

Chosen Question: Do states with higher renewable usage have lower average electricity prices in 2021?

Part 2: Data Preparation and Cleaning

```
#|eval = FALSE
#First step is making the energy_price into a workable dataset instead of
whatever it originally is. (A hot mess)
energy_price = energy_price[-1,]
energy_price
```

```
# A tibble: 53 × 1
   `Total energy average price, dollars per million Btu,,,`
   <chr>
 1 State,2021,2022,2023
 2 AK,$20.03 per MMBtu,$27.33,$23.84 est.
 3 AL,about 17.85 USD,23.37 USD,≈21.11
 4 AR,$18.42,$23.84 per MMBtu,$21.76
 5 AZ,≈25.07,31.72 USD,about 30.28
 6 CA,$28.44,$37.35,$35.72 per MMBtu
 7 CO,20.64 USD,≈25.85,23.85
 8 CT,about $25.85,$33.15,$32.32 est.
 9 DC,≈25.67,$30.84,about 32.28 USD
10 DE,$21.83,$27.74 per MMBtu,$26.70
# ℹ 43 more rows
```

```
new_energy_price = energy_price|>
    separate(col = colnames(energy_price),
    into = c("State", "2021_Average_Price", "2022_Average_Price",
"2023_Average_Price"),
    sep = ",") |>
    mutate(`2021_Average_Price` = str_remove_all(`2021_Average_Price`,
```

```
"[^0-9.]"))|>
    mutate(`2022_Average_Price` = str_remove_all(`2022_Average_Price`,
"[^0-9.]"))|>
    mutate(`2023_Average_Price` = str_remove_all(`2023_Average_Price`, "[^0-9.
{1}]"))
new_energy_price
```

```
# A tibble: 53 × 4
   State `2021_Average_Price` `2022_Average_Price` `2023_Average_Price`
   <chr> <chr>                <chr>                <chr>
 1 State 2021                 2022                 2023
 2 AK    20.03                27.33                23.84.
 3 AL    17.85                23.37                21.11
 4 AR    18.42                23.84                21.76
 5 AZ    25.07                31.72                30.28
 6 CA    28.44                37.35                35.72
 7 CO    20.64                25.85                23.85
 8 CT    25.85                33.15                32.32.
 9 DC    25.67                30.84                32.28
10 DE    21.83                27.74                26.70
# ℹ 43 more rows
```

```
#Going to need some data set with the proportion of energy coming from
renewable energy use from the total_use_2021 dataset

#Bro why are there 53 columns.
#DC, LA shouldn't be in there those are cities

total_use_2021
```

```
# A tibble: 5 × 53
  Energy_Source        AK     AL     AR     AZ     CA     CO     CT    DC
DE
  <chr>             <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl> <dbl>
<dbl>
1 Coal              18694 309791 216123 160299 2.82e4 252442   2880     0
4542
2 Natural Gas†     395590 739891 360545 484962 2.17e6 509970 305184 28336
82708
3 Petroleum (BTU)  261094 583042 328271 606862 2.96e6 497788 284788 18439
113641
4 nuclear               0 480115 141372 329868 1.72e5      0 179551     0
0
5 total_renewable…   9597 239817  89714  99266 8.10e5 103955  49306  2487
7150
```

```
# ℹ 43 more variables: FL <dbl>, GA <dbl>, HI <dbl>, IA <dbl>, ID <dbl>,
#   IL <dbl>, IN <dbl>, KS <dbl>, KY <dbl>, LA <dbl>, MA <dbl>, MD <dbl>,
#   ME <dbl>, MI <dbl>, MN <dbl>, MO <dbl>, MS <dbl>, MT <dbl>, NC <dbl>,
#   ND <dbl>, NE <dbl>, NH <dbl>, NJ <dbl>, NM <dbl>, NV <dbl>, NY <dbl>,
#   OH <dbl>, OK <dbl>, OR <dbl>, PA <dbl>, RI <dbl>, SC <dbl>, SD <dbl>,
#   TN <dbl>, TX <dbl>, UT <dbl>, VA <dbl>, VT <dbl>, WA <dbl>, WI <dbl>,
#   WV <dbl>, WY <dbl>, US <dbl>
```

```
total_use_2021_states = total_use_2021[, c(-9, -20)]

total_use_pivot = pivot_longer(total_use_2021_states, cols = c(2:51), names_to
= "State", values_to = "Usage")
total_use_pivot
```

```
# A tibble: 250 × 3
   Energy_Source State  Usage
   <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
 7 Coal          CT      2880
 8 Coal          DE      4542
 9 Coal          FL    200193
10 Coal          GA    203870
# ℹ 240 more rows
```

```
final_pivot = pivot_wider(total_use_pivot, names_from = Energy_Source,
values_from = Usage)
proportions_data = final_pivot|>
    group_by(State)|>
    mutate(proportion = total_renewable_energy/(Coal + `Natural Gas†` +
`Petroleum (BTU)`))|>
    select(State, proportion)
```

Part 3: Joining / Pivoting Datasets for Analysis

```
new_energy_price
```

```
# A tibble: 53 × 4
   State `2021_Average_Price` `2022_Average_Price` `2023_Average_Price`
   <chr> <chr>                <chr>                <chr>
```

```
 1 State 2021                2022                2023
 2 AK    20.03               27.33               23.84.
 3 AL    17.85               23.37               21.11
 4 AR    18.42               23.84               21.76
 5 AZ    25.07               31.72               30.28
 6 CA    28.44               37.35               35.72
 7 CO    20.64               25.85               23.85
 8 CT    25.85               33.15               32.32.
 9 DC    25.67               30.84               32.28
10 DE    21.83               27.74               26.70
# ℹ 43 more rows
```

```
proportions_data
```

```
# A tibble: 50 × 2
# Groups:   State [50]
   State proportion
   <chr>      <dbl>
 1 AK        0.0142
 2 AL        0.147
 3 AR        0.0991
 4 AZ        0.0793
 5 CA        0.157
 6 CO        0.0825
 7 CT        0.0832
 8 DE        0.0356
 9 FL        0.0840
10 GA        0.152
# ℹ 40 more rows
```

```
joined_dataset = left_join(x = new_energy_price, y = proportions_data)
```

```
Joining with `by = join_by(State)`
```

```
necessary_columns = select(joined_dataset, c(1, 2, 5))
necessary_columns
```

```
# A tibble: 53 × 3
   State `2021_Average_Price` proportion
   <chr> <chr>                     <dbl>
 1 State 2021                     NA
 2 AK    20.03                     0.0142
 3 AL    17.85                     0.147
```

```
 4 AR    18.42                    0.0991
 5 AZ    25.07                    0.0793
 6 CA    28.44                    0.157
 7 CO    20.64                    0.0825
 8 CT    25.85                    0.0832
 9 DC    25.67                   NA
10 DE    21.83                    0.0356
# ℹ 43 more rows
```

Part 4: Mapping Visualization

```
ggplot()
```