EV Power - Lab 4 Project Report

Example Solution 1

Part 0: libraries

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

# Try common data directories relative to the current working dir
candidates <- c(
  "data",
  "./data",
  "../data",
  "project4/ev-power-JreamIT/data",
  "./project4/ev-power-JreamIT/data"
)

data_dir <- NA_character_
for (p in candidates) {
  if (dir.exists(p) &&
      file.exists(file.path(p, "av-energy-price-2021-2023.csv")) &&
      file.exists(file.path(p, "ev-registrations-by-state-2023.csv"))) {
    data_dir <- p
    break
  }
```

```r
}
if (is.na(data_dir)) stop("Could not locate data directory. Check where the
two CSVs live.")

price_file <- read_csv(file.path(data_dir, "av-energy-price-2021-2023.csv"),
show_col_types = FALSE)
ev_file    <- read_csv(file.path(data_dir, "ev-registrations-by-
state-2023.csv"), show_col_types = FALSE)
```

```
New names:
• `` -> `...2`
```

```r
head(price_file)
```

```
# A tibble: 6 × 1
  `Total energy average price, dollars per million Btu,,,`
  <chr>
1 ,,,
2 State,2021,2022,2023
3 AK,$20.03 per MMBtu,$27.33,$23.84 est.
4 AL,about 17.85 USD,23.37 USD,≈21.11
5 AR,$18.42,$23.84 per MMBtu,$21.76
6 AZ,≈25.07,31.72 USD,about 30.28
```

```r
head(ev_file)
```

```
# A tibble: 6 × 2
  `electric vehicle registrations_by_state (2023)` ...2
  <chr>                                            <chr>
1 <NA>                                             <NA>
2 STATE                                            Count-EVs
3 Alabama                                          #13047
4 Alaska                                           ~2697
5 Arizona                                          89798
6 Arkansas                                         7108 EVs
```

```r
strip_price <- function(x){
  x |>
    str_replace_all("(?i)per\\s*MMBtu|USD|est\\.|about", "") |>
    str_replace_all("[^0-9.\\-]", "") |>
    readr::parse_number()
}
```

```r
strip_count <- function(x){
  suppressWarnings(
    x |>
      str_replace_all("[^0-9.\\-]", "") |>
      readr::parse_number()
  )
}


state_lookup_full  <- c(state.name, "District of Columbia")
state_lookup_abb   <- c(state.abb,  "DC")
to_full <- function(x){
  out <- ifelse(nchar(x) == 2, state_lookup_full[match(toupper(x),
state_lookup_abb)], x)
  out
}
```

Part 1: Defining Research Question

Chosen Question:How have average energy prices changed from 2021 to 2023, and which U.S. states lead in electric vehicle (EV) adoption?

Part 2: Data Preparation and Cleaning

```r
price_clean <- price_file |>
  rename_with(tolower)

has_year_cols <- any(grepl("202[1-3]", names(price_clean)))
if (ncol(price_clean) > 1 && !("state" %in% names(price_clean))) {
  price_clean <- price_clean |> rename(state = 1)
}
if (ncol(price_clean) > 1) {
  price_clean <- price_clean |> mutate(across(-state, strip_price))
}


if (ncol(price_clean) == 1 || !has_year_cols) {
  price_clean <- tibble(line = as.character(price_file[[1]])) |>
    filter(!is.na(line), nchar(trimws(line)) > 0) |>
    mutate(line = str_replace(line, '^"(.*)"$', "\\1")) |>
    separate(line,
             into = c("state","price_2021","price_2022","price_2023"),
             sep = ",", fill = "right", extra = "merge") |>
    mutate(
      state       = str_trim(state),
      price_2021  = strip_price(price_2021),
      price_2022  = strip_price(price_2022),
      price_2023  = strip_price(price_2023)
```

```
    )
}

price_clean <- price_clean |>
  select(state, matches("202[1-3]"))

ev_clean <- ev_file |>
  rename_with(tolower) |>
  select(1, 2) |>
  rename(state = 1, ev_registrations = 2) |>
  mutate(
    state = str_trim(state),
    ev_registrations = strip_count(ev_registrations)
  ) |>
  filter(!is.na(ev_registrations))
```

Part 3: Joining / Pivoting Datasets for Analysis

```
year_cols <- grep("202[1-3]", names(price_clean), value = TRUE)
if (length(year_cols) == 0) stop("No year columns found in price_clean after
cleaning.")

price_long <- price_clean |>
  pivot_longer(cols = dplyr::all_of(year_cols),
               names_to = "year",
               values_to = "price") |>
  mutate(year = readr::parse_integer(stringr::str_replace_all(year, "\\D",
"")))

price_2023 <- price_long |>
  filter(year == 2023) |>
  transmute(
    state_full = ifelse(nchar(state) == 2, to_full(state),
tools::toTitleCase(state)),
    price_2023 = price
  ) |>
  filter(!is.na(price_2023))

ev_ready <- ev_clean |>
  mutate(state_full = ifelse(nchar(state) == 2, to_full(state),
tools::toTitleCase(state))) |>
  select(state_full, ev_registrations)

compare_2023 <- ev_ready |>
  inner_join(price_2023, by = "state_full")
```
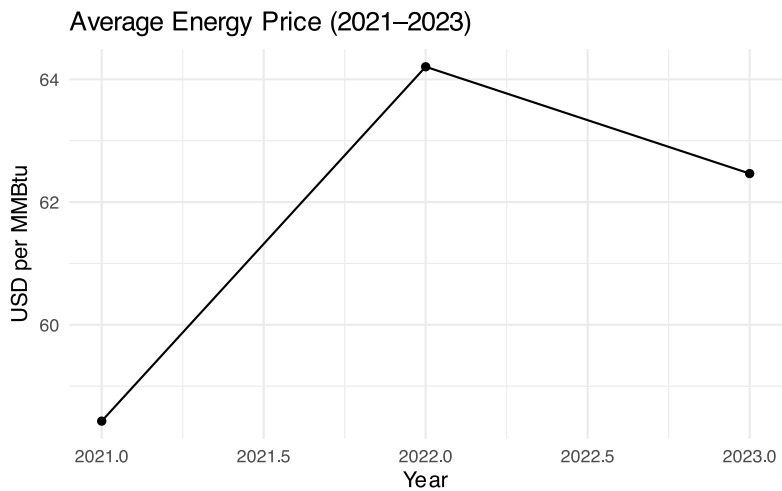
Part 4: Mapping Visualization

```r
avg_price <- price_long |>
  dplyr::group_by(year) |>
  dplyr::summarise(avg_price = mean(price, na.rm = TRUE), .groups = "drop")

p1 <- ggplot(avg_price, aes(year, avg_price)) +
  geom_line() + geom_point() +
  labs(title = "Average Energy Price (2021–2023)",
       x = "Year", y = "USD per MMBtu") +
  theme_minimal()
print(p1)
```


Average Energy Price (2021–2023)

```r
ev_top10 <- ev_ready |>
  arrange(desc(ev_registrations)) |>
  slice_head(n = 10)

p2 <- ggplot(ev_top10, aes(x = reorder(state_full, ev_registrations), y =
ev_registrations)) +
  geom_col() + coord_flip() +
  labs(title = "Top 10 States by EV Registrations (2023)",
       x = "State", y = "EV Registrations") +
  theme_minimal()
print(p2)
```

Top 10 States by EV Registrations (2023)
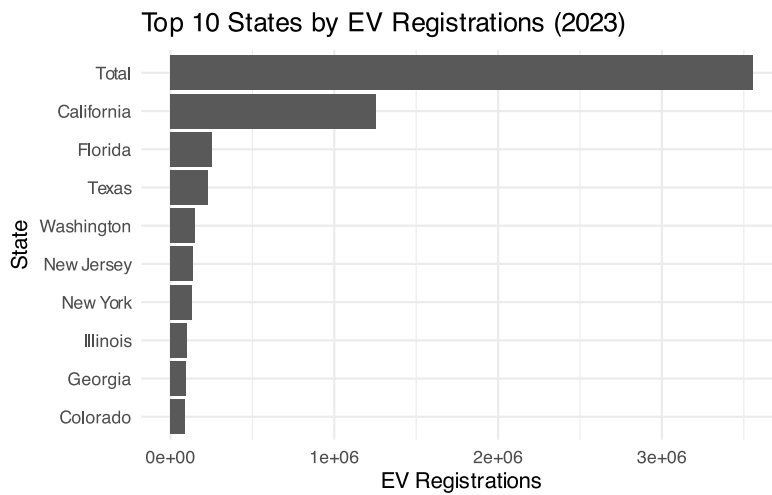

```
p3 <- ggplot(compare_2023, aes(x = price_2023, y = ev_registrations)) +
  geom_point(color = "tomato", size = 2.5) +
  labs(title = "EV Registrations vs Average Energy Price (2023)",
       x = "Average Energy Price (USD per MMBtu)",
       y = "EV Registrations") +
  theme_minimal()
print(p3)
```

EV Registrations vs Average Energy Price (2023)
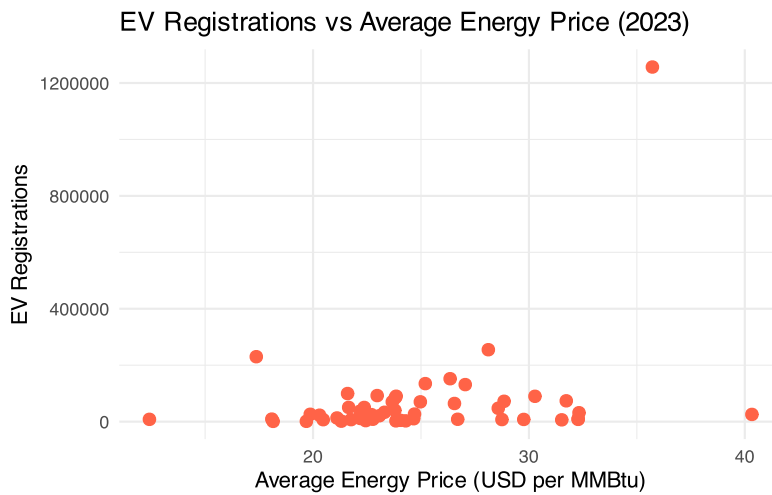

```
us_map <- ggplot2::map_data("state") |>
  mutate(state_full = tools::toTitleCase(region))

map_ev <- us_map |>
  left_join(ev_ready, by = "state_full")

p4 <- ggplot(map_ev, aes(long, lat, group = group, fill = ev_registrations)) +
```

```
  geom_polygon(color = "white", linewidth = 0.2) +
  scale_fill_gradient(low = "white", high = "steelblue", na.value = "grey90")
+
  labs(title = "EV Registrations by State (2023)", fill = "EVs") +
  theme_void()
print(p4)
```

EV Registrations by State (2023)