EV Power - Lab 4 Project Report

Andrew Li

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)
library(sf)        # for mapping
```

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

```
library(janitor)   # for clean_names()
```

```

Attaching package: 'janitor'
```

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

    chisq.test, fisher.test
```

```
library(maps)
```

Part 1: Defining Research Question

Chosen Question: Do states with cleaner electricity mixes also have higher EV adoption?

Part 2: Data Preparation and Cleaning

```
# renewable energy by state and year
renew21 <- read_csv("/Users/andrewli/Downloads/ev-power-Andrewli005/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.
```

```
renew22 <- read_csv("/Users/andrewli/Downloads/ev-power-Andrewli005/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.
```

```
renew23 <- read_csv("/Users/andrewli/Downloads/ev-power-Andrewli005/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 energy use by state and year
total21 <- read_csv("/Users/andrewli/Downloads/ev-power-Andrewli005/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.
```

```
total22 <- read_csv("/Users/andrewli/Downloads/ev-power-Andrewli005/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.
```

```
total23 <- read_csv("/Users/andrewli/Downloads/ev-power-Andrewli005/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.
```

```
# average energy price
price_all <- read_csv("/Users/andrewli/Downloads/ev-power-Andrewli005/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.
```

```
# EV registrations
ev_2023 <- read_csv("/Users/andrewli/Downloads/ev-power-Andrewli005/data/ev-
registrations-by-state-2023.csv") |>
  janitor::clean_names() |>
  rename(
    state = electric_vehicle_registrations_by_state_2023,
    ev_registrations = x2
  ) |>
  filter(!is.na(state), state != "STATE") |>   # remove header/junk rows
  mutate(
    ev_registrations = str_replace_all(ev_registrations, "[^0-9]", ""),
    ev_registrations = as.numeric(ev_registrations),
    state = str_to_title(state)                 # make "alabama" -> "Alabama"
  )
```

```
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`
```

```
# Preview cleaned data
head(ev_2023)
```

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

Part 3: Joining / Pivoting Datasets for Analysis

```
# assign an index instead of state names
ev_indexed <- ev_2023 |>
  mutate(index = row_number())

make_tidy_place_table <- function(df, yr = NA) {
  df <- janitor::clean_names(df)

  # try to find any column that looks like a state identifier
  # (full name, abbreviation, region, etc.)
  possible_cols <- intersect(
    c("state_abbr", "state", "state_name", "state_full", "region", "location",
"jurisdiction"),
    names(df)
  )

  # if we found such a column, normalize it to ALL CAPS 2-letter-ish form
  if (length(possible_cols) > 0) {
    keycol <- possible_cols[1]
    df$state_abbr <- toupper(df[[keycol]])
  }

  # add a year column if you give one
  if (!is.na(yr)) {
    df$year <- yr
  }

  df
}
```

```
price_raw <- readr::read_csv("/Users/andrewli/Downloads/ev-power-Andrewli005/
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.
```

```
price_c <- make_tidy_place_table(price_raw)


ev_raw <- readr::read_csv("/Users/andrewli/Downloads/ev-power-Andrewli005/
data/ev-registrations-by-state-2023.csv")
```

```
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`
```

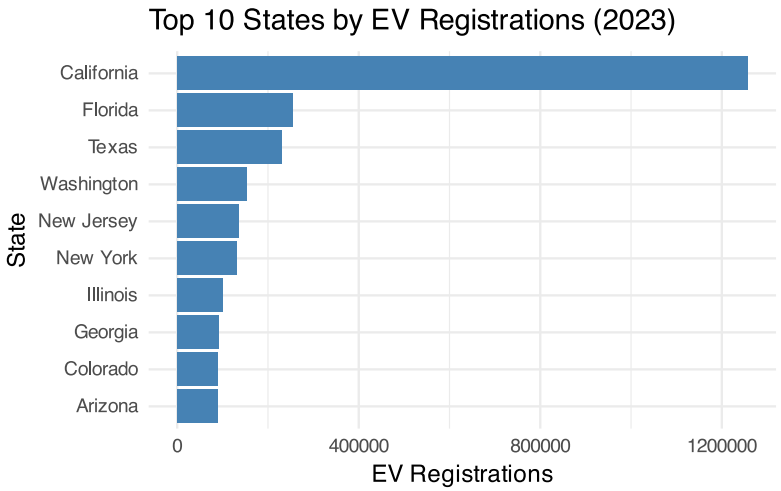
```
ev_c <- make_tidy_place_table(ev_raw, yr = 2023)

ev_clean_sorted <- ev_2023 |>
  # remove totals or non-state rows
  filter(!(state %in% c("Total", "United States", "Usa", "U.s."))) %>%
  # sort by number of EVs
  arrange(ev_registrations) |>
  # preserve sorted order in the factor
  mutate(state = factor(state, levels = state))

ev_top10 <- ev_clean_sorted |>
  top_n(10, ev_registrations)

ggplot(ev_top10, aes(x = state, y = ev_registrations)) +
```

```
  geom_col(fill = "steelblue") +
  coord_flip() +
  theme_minimal(base_size = 12) +
  labs(
    title = "Top 10 States by EV Registrations (2023)",
    x = "State",
    y = "EV Registrations"
  )
```



Part 4: Mapping Visualization

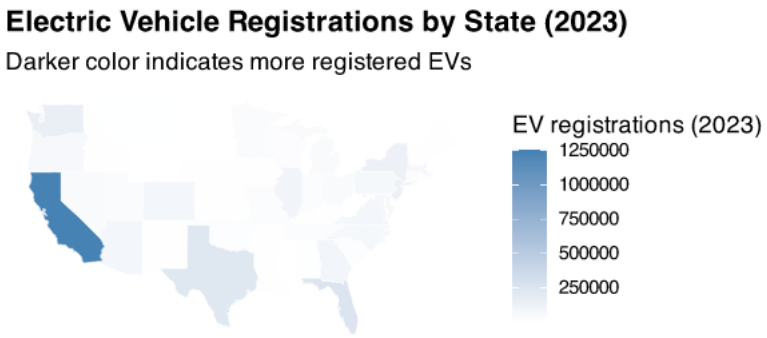
```
# 1. Get US map polygons
us_map <- map_data("state") |>
  mutate(state = str_to_title(region))  # region is lowercase state name

# 2. Join EV data to the map polygons by state name
ev_map_data <- us_map |>
  left_join(ev_2023, by = "state")

# 3. Plot choropleth
ggplot(ev_map_data, aes(long, lat, group = group, fill = ev_registrations)) +
  geom_polygon(color = "white", size = 0.2) +
  coord_fixed(1.3) +
  scale_fill_gradient(
    name = "EV registrations (2023)",
    low = "white",
    high = "steelblue",
    na.value = "grey90"
  ) +
  labs(
    title = "Electric Vehicle Registrations by State (2023)",
```

```
    subtitle = "Darker color indicates more registered EVs",
    x = NULL,
    y = NULL
  ) +
  theme_minimal(base_size = 12) +
  theme(
    axis.text = element_blank(),
    panel.grid = element_blank(),
    legend.position = "right",
    plot.title = element_text(face = "bold")
  )
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
ℹ Please use `linewidth` instead.
```



The graphic depicts the spread of registered electric vehicles (EVs) in different states of the U.S. in 2023. The states were colored according to the total number of EV registrations, with a darker color representing a higher number. By far, California has been the outlier with a lot more EVs than any other state. Besides, Florida and Texas have also a substantial number of EVs, whereas many interior and low, population states have been shaded very lightly. This indicates that the geographical distribution of EVs adoption is not uniform and that it is only a few states with large populations that have the most EV.