

EV Power - Lab 4 Project Report

Example Solution 1

Part 0: libraries

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

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v forcats   1.0.0      v readr     2.1.5
v lubridate 1.9.4      v tibble    3.3.0
v purrr     1.1.0
```

```
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag() masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(sf)
```

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

```
library(maps)
```

Attaching package: 'maps'

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

map

Part 1: Defining Research Question

Chosen Question: How does EV registration vary state by state? What is the relationship between average energy price and ev registration? What is the most common type of renewable energy across states in 2023?

Part 2: Data Preparation and Cleaning

```
av_energy_price_2023 <- read.csv("/Users/jayson/Documents/stat133/project-4/ev-power-jayson-a
ev_reg_2023 <- read.csv("/Users/jayson/Documents/stat133/project-4/ev-power-jayson-abarca/da
renewable_2023 <- read.csv("/Users/jayson/Documents/stat133/project-4/ev-power-jayson-abarca,

#Data cleaning for 2023 EV Registration
ev_reg_2023 <- ev_reg_2023 %>%
  rename(State = `electric.vehicle.registrations_by_state..2023.`) %>%
  mutate(
    State = toupper(State),
    State = state.abb[match(State, toupper(state.name))]
  )
```

```

ev_reg_2023 <- ev_reg_2023 %>%
  mutate(
    X = str_remove_all(X, "\\D"),
    X = as.integer(X)
  )

#Data Cleaning for average price
colnames(av_energy_price_2023)[1] <- "energy_price_raw"

av_energy_price_2023 <- av_energy_price_2023 %>%
  separate(
    col = energy_price_raw,
    into = c("State", "price_2021", "price_2022", "price_2023"),
    sep = ",",
    fill = "right"
  )

av_energy_price_2023 <- av_energy_price_2023 %>%
  mutate(
    across(
      .cols = c(price_2021, price_2022, price_2023),
      .fns = ~ .x |>
        str_remove_all("\\$|USD|per MMBtu|est\\.| |about") |>
        str_trim() |>
        str_extract("\\d+\\.?\\d*") |>
        as.numeric()
    )
  )

#Data Cleaning for renewable energy 2023
renewable_2023 <- renewable_2023 %>%
  mutate(State = str_to_upper(State))

renewable_2023 <- renewable_2023 %>%
  mutate(
    Renewable_Use_2023 = str_replace_all(Renewable_Use_2023, "[^0-9]", ""),
    Renewable_Use_2023 = as.integer(Renewable_Use_2023)
  ) %>%
  arrange(State)

renewable_sorted <- renewable_2023 %>%
  group_by(State) %>%

```

```
arrange(desc(Renewable_Use_2023), .by_group = TRUE) %>%
ungroup()
```

Part 3: Joining / Pivoting Datasets for Analysis

```
# Merge EV + price
ev_power <- ev_reg_2023 %>%
  inner_join(av_energy_price_2023, by = "State")

# Get most common renewable source by state
renewable_top <- renewable_sorted %>%
  group_by(State) %>%
  slice_max(Renewable_Use_2023, n = 1) %>%
  ungroup()

# Final master dataset
ev_full <- ev_power %>%
  left_join(renewable_top, by = "State") %>%
  mutate(
    EV_to_Renewable_Ratio = X / Renewable_Use_2023,
    Price_Change_2021_2023 = price_2023 - price_2021
  )
```

Part 4: Mapping Visualization

```
library(ggplot2)
library(dplyr)
library(sf)
library(maps)

# convert built-in state map into sf object
states_sf <- st_as_sf(map("state", plot = FALSE, fill = TRUE))

# `map("state")` uses lowercase full state names
# convert those to state abbreviations
states_sf <- states_sf %>%
  mutate(State = state.abb[match(tolower(ID), tolower(state.name))])
```

```

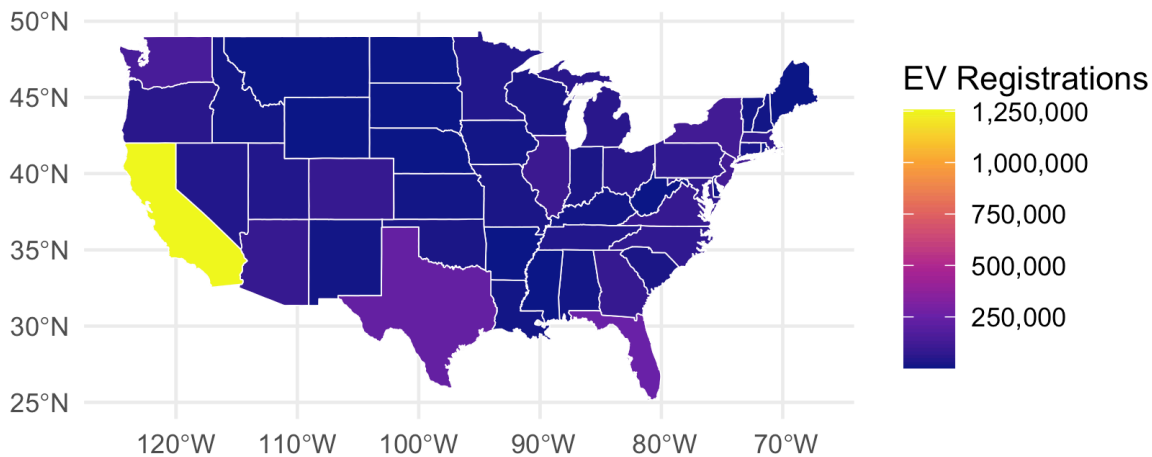
# join with your merged dataset (ev_full must already exist)
states_merged <- states_sf %>%
  left_join(ev_full, by = "State")

# choropleth map of EV registrations
ggplot(states_merged) +
  geom_sf(aes(fill = X), color = "white", size = 0.2) +
  scale_fill_viridis_c(
    option = "plasma",
    labels = scales::comma,
    name = "EV Registrations"
  ) +
  labs(
    title = "EV Registrations by State (2023)",
    subtitle = "EV adoption is highest in coastal and urbanized states",
    caption = "Source: DOE EV dataset (2023)"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 18, face = "bold"),
    legend.position = "right"
  )

```

EV Registrations by State (2023)

EV adoption is highest in coastal and urbanized states



Source: DOE EV dataset (2023)