

EV Power — Lab 4 Project Report

Tyler Maldonado

Example Solution 1

Part 0: libraries

```
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(readr)
library(maps)
library(tidyr)
library(ggplot2)
```

Part 1: Defining Research Question

Chosen Question: Do states that use more renewable energy have more electric vehicles registered?

Part 2: Data Preparation and Cleaning

```
#Read in CSV files
renew <- read_csv("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.
```

```
total <- read_csv("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.
```

```
evs <- read_csv("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`
```

```
names(evs) <- c("state", "ev_registrations_2023") # rename columns manually
evs <- evs |>
  mutate(
    state = str_trim(state),
```

```

    state = str_to_title(state)
  )

#Lowercase and replace spaces with underscores
names(renew) <- names(renew) |>
  tolower() |>
  str_replace_all(" ", "_")

names(total) <- names(total) |>
  tolower() |>
  str_replace_all(" ", "_")
names(evs) <- names(evs) |>
  tolower() |>
  str_replace_all(" ", "_")

#Clean up state names consistently (only renew + evs have 'state' columns)
renew$state <- renew$state |>
  str_trim() |>
  str_to_title()

evs$state <- evs$state |>
  str_trim() |>
  str_to_title()

# total is wide (columns are state abbreviations), so pivot it
total_long <- total |>
  # keep only "Total" if that row exists
  filter(str_detect(tolower(energy_source), "total")) |>
  select(-any_of("us")) |> # drop U.S. total
  select(where(is.numeric), everything()) |> # drop non-numeric junk early
  pivot_longer(
    cols = -energy_source,
    names_to = "state_abbr",
    values_to = "total_use_2023"
  ) |>
  mutate(
    state_abbr_up = toupper(state_abbr),
    state = state.name[match(state_abbr_up, state.abb)]
  ) |>
  filter(!is.na(state)) |>
  mutate(state = str_to_title(state)) |>
  select(state, total_use_2023)

```

```

# look at the correct numeric columns
renew_col <- names(renew)[str_detect(names(renew), "2022")]
ev_col <- names(evs)[str_detect(names(evs), "ev")]

#Clean + rename each dataset
renew_clean <- renew |>
  select(state, all_of(renew_col)) |>
  rename(renewable_use_2022 = all_of(renew_col)) |>
  mutate(
    renewable_use_2022 = str_replace_all(renewable_use_2022, "[^0-9.]", ""), # remove anything
    renewable_use_2022 = as.numeric(renewable_use_2022)
  )

evs_clean <- evs |>
  select(state, all_of(ev_col)) |>
  rename(ev_registrations_2023 = all_of(ev_col)) |>
  mutate(
    ev_registrations_2023 = str_replace_all(ev_registrations_2023, "[^0-9.]", ""),
    ev_registrations_2023 = as.numeric(ev_registrations_2023)
  )

```

Part 3: Joining / Pivoting Datasets for Analysis

```

#Join the three datasets
final <- renew_clean |>
  left_join(total_long, by = "state") |>
  left_join(evs_clean, by = "state")

#Add new calculated columns for analysis
final <- final |>
  mutate(
    renewable_share_pct = (renewable_use_2022 / total_use_2023) * 100,
    ev_per_energy = ev_registrations_2023 / total_use_2023
  )

# look at the merged dataset
head(final)

```

A tibble: 6 x 6

```

state renewable_use_2022 total_use_2023 ev_registrations_2023
<chr> <dbl> <dbl> <dbl>
1 Ak 3846 NA NA
2 Ak 186 NA NA
3 Ak 5846 NA NA
4 Ak 57 NA NA
5 Ak 475 NA NA
6 Al 193932 NA NA
# i 2 more variables: renewable_share_pct <dbl>, ev_per_energy <dbl>

```

Part 4: Mapping Visualization

```

# Prepare map data and join
states_map <- map_data("state")

final_map <- final |>
  mutate(region = str_to_lower(state))

map_joined <- states_map |>
  left_join(final_map, by = "region")

#i filled the aesthetics, ai made it look pretty, i promise

# Renewable Share by State
p1 <- ggplot(map_joined, aes(long, lat, group = group, fill = renewable_share_pct)) +
  geom_polygon(color = "white", linewidth = 0.2) +
  coord_fixed(1.3) +
  scale_fill_gradient(name = "Renewable Share (%)",
                      low = "white", high = "darkgreen",
                      na.value = "grey90") +
  labs(
    title = "Renewable Energy Share by State",
  ) +
  theme_void() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    plot.subtitle = element_text(hjust = 0.5)
  )

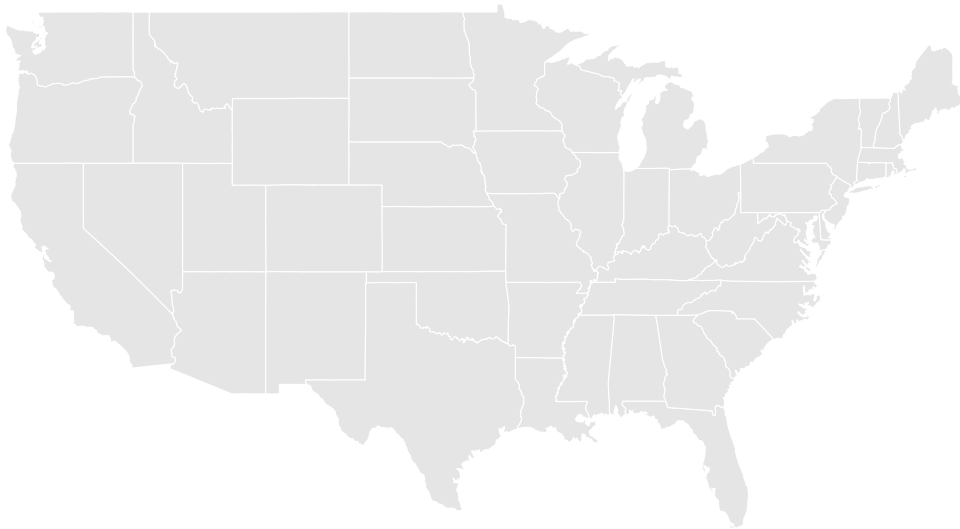
p2 <- ggplot(final |> drop_na(renewable_share_pct, ev_registrations_2023),
  aes(x = renewable_share_pct, y = ev_registrations_2023)) +

```

```
geom_point(alpha = 0.8) +  
geom_smooth(method = "lm", se = FALSE, color = "black") +  
labs(  
  title = "Do States with More Renewable Energy Have More EVs?",  
  x = "Renewable Energy Share (%)",  
  y = "EV Registrations (2023)"  
) +  
theme_minimal()
```

```
print(p1)
```

Renewable Energy Share by State



```
print(p2)
```

Do States with More Renewable Energy Have More EVs?

EV Registrations (2023)

Renewable Energy Share (%)

From the scatterplot and map, a clear regional pattern emerges. States along the West Coast (like California, Oregon, and Washington) and parts of the Northeast show both a higher share of renewable energy use and more EV registrations. Meanwhile, many states across the Midwest and Southeast have lower renewable percentages and fewer EVs. The map shows that states with higher renewable energy use—like California, Oregon, and Washington—also tend to have more EVs. States with lower renewable shares generally have fewer EVs. Overall, there's a clear positive pattern: regions investing in clean energy often have higher EV adoption.