

# Clean Energy and Electric Vehicles

## Analyzing the Relationship Between Renewable Energy and EV Adoption

### Example Solution 1

#### Part 0: libraries

```
library(tidyverse)
```

#### Part 1: Defining Research Question

Chosen Question: In 2023, do states that use more renewable energy also have more EVs registered?

This question addresses whether electric vehicle adoption is concentrated in states with cleaner energy grids. If EVs are primarily adopted in states with high renewable energy usage, it suggests that EVs are indeed contributing to emissions reductions. Conversely, if EV adoption is highest in states heavily reliant on fossil fuels, the environmental benefits may be more limited.

#### Part 2: Data Preparation and Cleaning

```
# Part 2: Data Preparation and Cleaning

# Load renewable energy data for 2023
renew_2023 <- read_csv("data/renew-use-2023.csv")

# Load total energy use for 2023
total_2023 <- read_csv("data/total-use-2023.csv")
```

```

# Load EV registrations for 2023 - skip the first 2 rows and set column names
ev_2023 <- read_csv("data/ev-registrations-by-state-2023.csv",
                    skip = 2,
                    col_names = c("state", "ev_count"))

# Clean renewable energy data
renew_2023_clean <- renew_2023 %>%
  mutate(
    renewable_value = parse_number(Renewable_Use_2023),
    state = str_to_upper(str_trim(State))
  ) %>%
  group_by(state) %>%
  summarize(total_renewable = sum(renewable_value, na.rm = TRUE))

# Clean total energy data
total_2023_clean <- total_2023 %>%
  pivot_longer(
    cols = -Energy_Source,
    names_to = "state",
    values_to = "energy_value"
  ) %>%
  mutate(state = str_to_upper(state)) %>%
  group_by(state) %>%
  summarize(total_energy = sum(energy_value, na.rm = TRUE))

# Clean EV registrations data
ev_2023_clean <- ev_2023 %>%
  mutate(
    ev_registrations = parse_number(ev_count),
    state = str_to_upper(str_trim(state))
  ) %>%
  filter(state != "TOTAL") %>%
  select(state, ev_registrations)

# Display samples
head(renew_2023_clean)

```

```

# A tibble: 6 x 2
  state total_renewable
  <chr>          <dbl>
1 AK             10088
2 AL             222189

```

3	AR	87277
4	AZ	108445
5	CA	1065179
6	CO	115062

```
head(total_2023_clean)
```

```
# A tibble: 6 x 2
  state total_energy
  <chr>      <dbl>
1 AK         746979
2 AL        2265008
3 AR        1151062
4 AZ        1712667
5 CA        6429818
6 CO        1359507
```

```
head(ev_2023_clean)
```

```
# A tibble: 6 x 2
  state      ev_registrations
  <chr>          <dbl>
1 STATE              NA
2 ALABAMA          13047
3 ALASKA            2697
4 ARIZONA          89798
5 ARKANSAS           7108
6 CALIFORNIA      1256646
```

### Part 3: Joining / Pivoting Datasets for Analysis

```
# Part 3: Joining and Creating Analysis Variables (WITH DEBUGGING)

# First, let's see what we have
cat("Renewable data rows:", nrow(renew_2023_clean), "\n")
```

```
Renewable data rows: 52
```

```
cat("Total energy data rows:", nrow(total_2023_clean), "\n")
```

Total energy data rows: 52

```
cat("EV data rows:", nrow(ev_2023_clean), "\n")
```

EV data rows: 52

```
# Check a few state names from each dataset
cat("\nSample states from renewable data:\n")
```

Sample states from renewable data:

```
head(renew_2023_clean$state)
```

```
[1] "AK" "AL" "AR" "AZ" "CA" "CO"
```

```
cat("\nSample states from total energy data:\n")
```

Sample states from total energy data:

```
head(total_2023_clean$state)
```

```
[1] "AK" "AL" "AR" "AZ" "CA" "CO"
```

```
cat("\nSample states from EV data:\n")
```

Sample states from EV data:

```
head(ev_2023_clean$state)
```

```
[1] "STATE"      "ALABAMA"    "ALASKA"     "ARIZONA"    "ARKANSAS"
[6] "CALIFORNIA"
```

```
# Join all datasets
combined_data <- renew_2023_clean %>%
  left_join(total_2023_clean, by = "state") %>%
  left_join(ev_2023_clean, by = "state")

cat("\nRows after joining:", nrow(combined_data), "\n")
```

Rows after joining: 52

```
cat("Rows with EV data:", sum(!is.na(combined_data$ev_registrations)), "\n")
```

Rows with EV data: 0

```
# Now filter
combined_data <- combined_data %>%
  filter(!is.na(ev_registrations),
         !is.na(total_renewable),
         !is.na(total_energy),
         state != "US",
         total_energy > 0) %>%
  mutate(
    renewable_pct = (total_renewable / total_energy) * 100,
    renewable_pct = round(renewable_pct, 2)
  )

# Display the combined dataset
cat("\nFinal rows:", nrow(combined_data), "\n")
```

Final rows: 0

```
head(combined_data, 10)
```

```
# A tibble: 0 x 5
# i 5 variables: state <chr>, total_renewable <dbl>, total_energy <dbl>,
#   ev_registrations <dbl>, renewable_pct <dbl>
```

```
# Summary statistics
summary(combined_data %>% select(renewable_pct, ev_registrations, total_energy))
```

```
renewable_pct ev_registrations total_energy
Min.   : NA   Min.   : NA   Min.   : NA
1st Qu.: NA   1st Qu.: NA   1st Qu.: NA
Median : NA   Median : NA   Median : NA
Mean   :NaN   Mean   :NaN   Mean   :NaN
3rd Qu.: NA   3rd Qu.: NA   3rd Qu.: NA
Max.   : NA   Max.   : NA   Max.   : NA
```

## Part 4: Mapping Visualization

```
# Part 4: Visualization

# Calculate correlation
correlation <- cor(combined_data$renewable_pct,
                   combined_data$ev_registrations)

cat("Correlation:", correlation, "\n")
```

Correlation: NA

```
# Scatter plot: Renewable Percentage vs EV Registrations
combined_data %>%
  mutate(state_label = ifelse(ev_registrations > 80000 | renewable_pct > 30, state, "")) %>%
  ggplot(aes(x = renewable_pct, y = ev_registrations)) +
  geom_point(aes(size = total_energy, color = renewable_pct), alpha = 0.6) +
  geom_smooth(method = "lm", se = TRUE, color = "red", linetype = "dashed") +
  geom_text(aes(label = state_label), vjust = -0.8, size = 3, check_overlap = TRUE) +
  scale_y_log10(labels = scales::comma) +
  scale_color_gradient(low = "blue", high = "green", name = "Renewable %") +
  scale_size_continuous(name = "Total Energy\nUse (BTU)", labels = scales::comma) +
  labs(
    title = "Relationship Between Renewable Energy and EV Adoption (2023)",
    x = "Renewable Energy (% of Total)",
    y = "EV Registrations (log scale)",
    subtitle = paste0("Correlation: ", round(correlation, 3))
  ) +
```

```
theme_minimal() +
theme(
  plot.title = element_text(face = "bold", size = 14),
  legend.position = "right"
)
```

## Relationship Between Renewable Energy and EV Adoption

Correlation: NA

EV Registrations (log scale)

Renewable Energy (% of Total)

The scatterplot compares renewable energy share and EV registrations for 2023. If the pattern shows an upward slope, that suggests states investing in renewable energy also tend to have more EVs, indicating a cleaner transportation-energy ecosystem. If there's no clear relationship, it implies EV adoption may depend more on infrastructure, incentives, or population density than renewable availability.“