Renewable Energy Usage and EV Registration Trends

Example Solution 1

Part 0: libraries

```r
library("tidyverse")
```

```
── Attaching core tidyverse packages ─────────────────────── tidyverse 2.0.0
──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.1     ✔ stringr   1.5.2
✔ ggplot2   4.0.0     ✔ tibble    3.3.0
✔ lubridate 1.9.4     ✔ tidyr     1.3.1
✔ purrr     1.1.0
── Conflicts ──────────────────────────────── tidyverse_conflicts()
──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```

```r
library("leaflet")
library("maps")
```

```
Attaching package: 'maps'

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

    map
```

Part 1: Defining Research Question

Chosen Question: Do states with higher renewable energy usage tend to have more electric vehicle (EV) registrations?

Part 2: Data Preparation and Cleaning

```r
setwd("~/Downloads/school/STAT 133/ev-power-jasminehou07/data")
# AV ENERGY DATA
AvgEnergyPrice <- read.csv("av-energy-price-2021-2023.csv")

# convert to tibble
AvgEnergyPrice <- tibble(col = AvgEnergyPrice[[1]])

# delete the first two junk lines
AvgEnergyPrice <- AvgEnergyPrice[-c(1,2), ]

# remove the outer quotes from each line
AvgEnergyPrice <- AvgEnergyPrice %>%
  mutate(col = str_remove_all(col, '^"|"$'))

# split the single column into 4 columns
AvgEnergyPrice <- AvgEnergyPrice %>%
  separate("col", into = c("State", "Avgprice_2021", "Avgprice_2022",
"Avgprice_2023"), sep = ",") %>%
  mutate(across(everything(), str_trim))

# clean numbers, keep only digits and decimal points
AvgEnergyPrice <- AvgEnergyPrice %>%
  mutate(across(starts_with("Avgprice_"),
    ~ .x %>%
      str_replace_all(",", "") %>%
      str_extract("\\d*\\.?\\d+") %>%
      str_replace("^\\.", "0.") %>%
      na_if("") %>%
      as.numeric()
  ))

# change state abbreviations to full names but ignore the last row and DC
AvgEnergyPrice <- AvgEnergyPrice %>%
  mutate(State = case_when(
    row_number() == 8 ~ "District of Columbia", row_number() == n() ~ State,
TRUE ~ state.name[match(State, state.abb)]))

# create csv file
write_csv(AvgEnergyPrice, "AvgEnergyPrice.csv")

# EV REGISTRATION DATA
#skip first 2 lines since they dont have data
EV_Registrations <- read.csv("ev-registrations-by-state-2023.csv", skip = 2)

#convert to tibble
EV_Registrations <- as_tibble(EV_Registrations)

# rename columns
```

```r
colnames(EV_Registrations) <- c("State", "Registrations")

#fix random characters
EV_Registrations <- EV_Registrations %>%
  mutate(across("Registrations",
                ~ str_replace_all(.x, "[^0-9.]", "") %>%
                  as.numeric()))

#create csv file
write_csv(EV_Registrations, "EV_Registrations.csv")

#RENEWABLE USE 2021 DATA
Renewable_Use2021 <- read.csv("renew-use-2021.csv")

# convert to tibble
Renewable_Use2021 <- as_tibble(Renewable_Use2021)

# rename columns
colnames(Renewable_Use2021)[3] <- "Usage"

# clean numbers, keep only digits and decimal points
Renewable_Use2021 <- Renewable_Use2021 %>%
  mutate(across("Usage",
                ~ str_replace_all(.x, "[^0-9.]", "") %>%
                  as.numeric()))

# change state abbreviations to full names but ignore last 5 rows and DC
Renewable_Use2021 <- Renewable_Use2021 %>%
  mutate(State = case_when(
    row_number() %in% 36:40 ~ "District of Columbia", row_number() > n() - 5 ~
State, TRUE ~ state.name[match(State, state.abb)]))

# create csv file
write_csv(Renewable_Use2021, "Renewable_Use2021.csv")

#RENEWABLE USE 2022 DATA
Renewable_Use2022 <- read.csv("renew-use-2022.csv")

# convert to tibble
Renewable_Use2022 <- as_tibble(Renewable_Use2022)

# rename columns
colnames(Renewable_Use2022)[3] <- "Usage"

# clean numbers, keep only digits and decimal points
Renewable_Use2022 <- Renewable_Use2022 %>%
  mutate(across("Usage",
                ~ str_replace_all(.x, "[^0-9.]", "") %>%
```

```r
                  as.numeric()))

# change state abbreviations to full names but ignore last 5 rows and DC
Renewable_Use2022 <- Renewable_Use2022 %>%
  mutate(State = case_when(
    row_number() %in% 36:40 ~ "District of Columbia", row_number() > n() - 5 ~
State, TRUE ~ state.name[match(State, state.abb)]))

# create csv file
write_csv(Renewable_Use2022, "Renewable_Use2022.csv")

#RENEWABLE USE 2023 DATA
Renewable_Use2023 <- read.csv("renew-use-2023.csv")

# convert to tibble
Renewable_Use2023 <- as_tibble(Renewable_Use2023)

# rename columns
colnames(Renewable_Use2023)[3] <- "Usage"

# clean numbers, keep only digits and decimal points
Renewable_Use2023 <- Renewable_Use2023 %>%
  mutate(across("Usage",
                ~ str_replace_all(.x, "[^0-9.]", "") %>%
                  as.numeric()))

#turn state abbreviations to uppercase
Renewable_Use2023 <- Renewable_Use2023 %>%
  mutate(State = toupper(State))

# change state abbreviations to full names but ignore last 5 rows & DC
Renewable_Use2023 <- Renewable_Use2023 %>%
  mutate(State = case_when(
    row_number() %in% 36:40 ~ "District of Columbia", row_number() > n() - 5 ~
State, TRUE ~ state.name[match(State, state.abb)]))

# create csv file
write_csv(Renewable_Use2023, "Renewable_Use2023.csv")

#TOTAL USE 2021 DATA
Total_Use2021 <- read.csv("total-use-2021.csv")

Total_Use2021 <- as_tibble(Total_Use2021)

#tidy up Energy_Source rows
Total_Use2021$Energy_Source <- Total_Use2021$Energy_Source %>%
  str_replace_all("_", " ") %>%
  str_replace_all("-", " ") %>%
```

```r
  str_replace_all("([a-z])([A-Z])", "\\1 \\2") %>%
  str_remove_all("\\(.*?\\)") %>%
  str_to_lower() %>%
  str_to_title() %>%
  str_replace_all("[^A-Za-z ]", "")

#change column names to state
col_names <- names(Total_Use2021)
col_names[2:(length(col_names) - 1)] <- state.name[match(col_names[2:
(length(col_names) - 1)], state.abb)]
col_names[9] <- "District of Columbia"
names(Total_Use2021) <- col_names

#new csv
write_csv(Total_Use2021, "Total_Use2021.csv")

#TOTAL USE 2022 DATA
Total_Use2022 <- read.csv("total-use-2022.csv")

Total_Use2022 <- as_tibble(Total_Use2022)

#tidy up Energy_Source rows
Total_Use2022$Energy_Source <- Total_Use2022$Energy_Source %>%
  str_replace_all("_", " ") %>%
  str_replace_all("-", " ") %>%
  str_remove_all("\\(.*?\\)") %>%
  str_to_lower() %>%
  str_to_title() %>%
  str_replace_all("[^A-Za-z ]", "")

#change column names to state
col_names <- names(Total_Use2022)
col_names[2:(length(col_names) - 1)] <- state.name[match(col_names[2:
(length(col_names) - 1)], state.abb)]
col_names[9] <- "District of Columbia"
names(Total_Use2022) <- col_names

#new csv
write_csv(Total_Use2022, "Total_Use2022.csv")

#TOTAL USE 2023 DATA
Total_Use2023 <- read.csv("total-use-2023.csv")

Total_Use2023 <- as_tibble(Total_Use2023)

#tidy up Energy_Source rows
Total_Use2023$Energy_Source <- Total_Use2023$Energy_Source %>%
  str_replace_all("_", " ") %>%
```

```
  str_replace_all("-", " ") %>%
  str_replace_all("([a-z])([A-Z])", "\\1 \\2") %>%
  str_remove_all("\\(.*?\\)") %>%
  str_to_lower() %>%
  str_to_title() %>%
  str_replace_all("[^A-Za-z ]", "")

#change column names to state
col_names <- names(Total_Use2023)
col_names[2:(length(col_names) - 1)] <- state.name[match(col_names[2:
(length(col_names) - 1)], state.abb)]
col_names[9] <- "District of Columbia"
names(Total_Use2023) <- col_names

#new csv
write_csv(Total_Use2023, "Total_Use2023.csv")
```

Part 3: Joining / Pivoting Datasets for Analysis

```
#pivot total use 2021
num_cols <- names(Total_Use2021)[sapply(Total_Use2021, is.numeric)]
num_cols <- setdiff(num_cols, "Energy_Source")

Total_Use2021_long <- pivot_longer(Total_Use2021, cols = -Energy_Source,
names_to = "State", values_to = "Value")

Total_Use2021_final <- pivot_wider(Total_Use2021_long, names_from =
Energy_Source, values_from = Value)
write_csv(Total_Use2021_final, "Total_Use2021_final.csv")

#pivot total use 2022
num_cols <- names(Total_Use2022)[sapply(Total_Use2022, is.numeric)]
num_cols <- setdiff(num_cols, "Energy_Source")

Total_Use2022_long <- pivot_longer(Total_Use2022, cols = -Energy_Source,
names_to = "State", values_to = "Value")

Total_Use2022_final <- pivot_wider(Total_Use2022_long, names_from =
Energy_Source, values_from = Value)
write_csv(Total_Use2022_final, "Total_Use2022_final.csv")

#pivot total use 2023
num_cols <- names(Total_Use2023)[sapply(Total_Use2023, is.numeric)]
num_cols <- setdiff(num_cols, "Energy_Source")

Total_Use2023_long <- pivot_longer(Total_Use2023, cols = -Energy_Source,
names_to = "State", values_to = "Value")
```

```r
Total_Use2023_final <- pivot_wider(Total_Use2023_long, names_from =
Energy_Source, values_from = Value)
write_csv(Total_Use2023_final, "Total_Use2023_final.csv")

#pivot renewable use 2021
Renewable_Use2021_final <- Renewable_Use2021 %>%
  pivot_wider(
    names_from = Energy_Source,
    values_from = Usage
  )

write_csv(Renewable_Use2021_final, "Renewable_Use2021_final.csv")

#pivot renewable use 2022
Renewable_Use2022_final <- Renewable_Use2022 %>%
  pivot_wider(
    names_from = Energy_Source,
    values_from = Usage
  )

write_csv(Renewable_Use2022_final, "Renewable_Use2022_final.csv")

#pivot renewable use 2023
Renewable_Use2023_final <- Renewable_Use2023 %>%
  pivot_wider(
    names_from = Energy_Source,
    values_from = Usage
  )

write_csv(Renewable_Use2023_final, "Renewable_Use2023_final.csv")

#joining tables
#renewable
R2021 <- read_csv("Renewable_Use2021_final.csv") %>% rename_with(~paste0(.x,
"_2021"), -State)
```

```
Rows: 52 Columns: 6
── Column specification
────────────────────────────────────────────────────────────
Delimiter: ","
chr (1): State
dbl (5): Biomass, Geothermal, Hydropower, Solar Energy, Wind Energy

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.
```

```r
R2022 <- read_csv("Renewable_Use2022_final.csv") %>% rename_with(~paste0(.x,
"_2022"), -State)
```

```
Rows: 52 Columns: 6
── Column specification ─────────────────────────────────────────
Delimiter: ","
chr (1): State
dbl (5): Biomass, Geothermal, Hydropower, Solar Energy, Wind Energy

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.
```

```r
R2023 <- read_csv("Renewable_Use2023_final.csv") %>% rename_with(~paste0(.x,
"_2023"), -State)
```

```
Rows: 52 Columns: 6
── Column specification ─────────────────────────────────────────
Delimiter: ","
chr (1): State
dbl (5): Biomass, Geothermal, Hydropower, Solar Energy, Wind Energy

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.
```

```r
# join all by state
Renewable_All <- R2021 %>%
  left_join(R2022, by = "State") %>%
  left_join(R2023, by = "State")

write_csv(Renewable_All, "Renewable_All.csv")

#renewable
T2021 <- read_csv("Total_Use2021_final.csv") %>% rename_with(~paste0(.x,
"_2021"), -State)
```

```
Rows: 52 Columns: 6
── Column specification ─────────────────────────────────────────
Delimiter: ","
chr (1): State
```

```
dbl (5): Coal, Natural Gas, Petroleum, Nuclear, Total Renewable Energy

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.
```

```r
T2022 <- read_csv("Total_Use2022_final.csv") %>% rename_with(~paste0(.x,
"_2022"), -State)
```

```
Rows: 52 Columns: 6
── Column specification
──────────────────────────────────────────────────────────
Delimiter: ","
chr (1): State
dbl (5): Coal Consumption, Natural Gas, Petroleum, Nuclear Energy, Total
Ren...

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.
```

```r
T2023 <- read_csv("Total_Use2023_final.csv") %>% rename_with(~paste0(.x,
"_2023"), -State)
```

```
Rows: 52 Columns: 6
── Column specification
──────────────────────────────────────────────────────────
Delimiter: ","
chr (1): State
dbl (5): Coal Usage, Natural Gas, Petroleum, Nuclear Energy, Total
Renewable...

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.
```

```r
# join all by state
Total_Use_All <- T2021 %>%
  left_join(T2022, by = "State") %>%
  left_join(T2023, by = "State")

write_csv(Total_Use_All, "Total_Use_All.csv")

# find totals for each renewable energy
```

```r
Renewable_All$renew_2021 <- with(Renewable_All,
  Biomass_2021 + Geothermal_2021 + Hydropower_2021 + `Solar Energy_2021` +
`Wind Energy_2021`
)
Renewable_All$renew_2022 <- with(Renewable_All,
  Biomass_2022 + Geothermal_2022 + Hydropower_2022 + `Solar Energy_2022` +
`Wind Energy_2022`
)
Renewable_All$renew_2023 <- with(Renewable_All,
  Biomass_2023 + Geothermal_2023 + Hydropower_2023 + `Solar Energy_2023` +
`Wind Energy_2023`
)

# row num the US total is in
us_r <- nrow(Renewable_All)

#find proportions for renewable energy usage
Renewable_All$prop_2021 <- Renewable_All$renew_2021 /
Renewable_All$renew_2021[us_r]
Renewable_All$prop_2022 <- Renewable_All$renew_2022 /
Renewable_All$renew_2022[us_r]
Renewable_All$prop_2023 <- Renewable_All$renew_2023 /
Renewable_All$renew_2023[us_r]

# row num the US total is in
us_ev <- nrow(EV_Registrations)

#find props for EV registrations
EV_Registrations$prop <- EV_Registrations$Registrations /
EV_Registrations$Registrations[us_ev]

#remove the last row of US
Renewable_All <- slice(Renewable_All, -nrow(Renewable_All))
EV_Registrations <- slice(EV_Registrations, -nrow(EV_Registrations))

#merge renewable and EV data
comparison <- Renewable_All %>%
  select(State, prop_2021, prop_2022, prop_2023) %>%
  left_join(EV_Registrations %>% select(State, prop), by = "State")
```

Part 4: Mapping Visualization

```r
states_map <- map_data("state")

# make sure state names match
comparison$region <- tolower(comparison$State)

# merge with map coordinates
```

```r
map_2023 <- left_join(states_map, comparison, by = "region")

# create the map
ev_points <- map_2023 %>%
  group_by(region) %>%
  summarise(long = mean(long), lat = mean(lat), EV = mean(prop, na.rm = TRUE))

# renewable energy is shown by color, ev registration is shown by circle size
ggplot(map_2023, aes(x = long, y = lat, group = group)) +
  geom_polygon(aes(fill = prop_2023), color = "white") +
  geom_point(data = ev_points,
             aes(x = long, y = lat, size = EV),
             color = "red", alpha = 0.6, inherit.aes = FALSE) +
  coord_fixed(1.3) +
  scale_fill_gradient(low = "lightblue", high = "darkgreen", name = "Renewable
Share (2023)") +
  scale_size_continuous(name = "EV Registration Share", range = c(2, 10)) +
  labs(title = "Renewable Energy & EV Registrations by State (2023)",
       subtitle = "Color = renewable energy share; Circle size = EV
registration share",
       caption = "Data: Renewable_All & EV_Registrations (2023)") +
  theme_minimal() +
  theme(axis.text = element_blank(),
        axis.title = element_blank(),
        panel.grid = element_blank())
```

Renewable Energy & EV Registrations by State (2023)
Color = renewable energy share; Circle size = EV registration share

Renewable Share (2023)

0.125
0.100
0.075
0.050
0.025

EV Registration Share

0.1
0.2
0.3

Data: Renewable_All & EV_Registrations (2023)
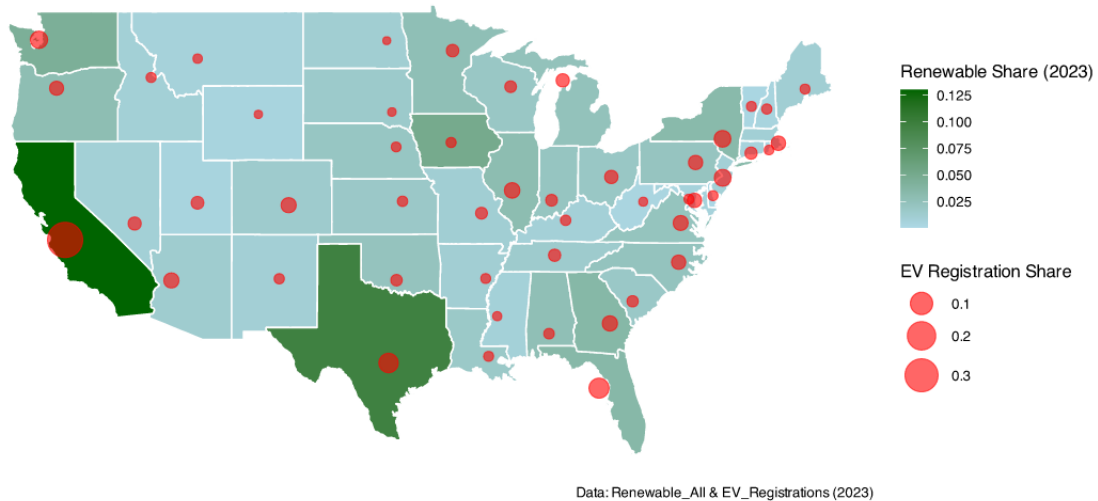
Findings

Based on the combined dataset and map visualization, there is a relationship between renewable energy usage and EV registrations but it's not very strong and varies by region. States on the west coast have both high renewable energy and EV registration. THis shows that states that invest in more renewables may also encourage EV. However, states like Iowa, South Dakota, and Idaho have a high renewable energy proportion but have low EV registrations. On the other hand, states like Florida, Texas, and New York have high EV registrations but lower renewable energy. Overall, there is not a clear one-to-one relationship between renewable energy and EV registrations.