EV Power - Lab 4 Project Report

Example Solution 1

Part 0: libraries

```
library(ggplot2)
library(maps)
library(viridis)
```

```
Loading required package: viridisLite
```

```


Attaching package: 'viridis'
```

```
The following object is masked from 'package:maps':

    unemp
```

```
library(scales)
```

```


Attaching package: 'scales'
```

```
The following object is masked from 'package:viridis':

    viridis_pal
```

```
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
✔ lubridate 1.9.4     ✔ tibble    3.3.0
✔ purrr     1.1.0     ✔ tidyr     1.3.1
```

```
── Conflicts ──────────────────────────────────────── tidyverse_conflicts()
──
✖ readr::col_factor() masks scales::col_factor()
✖ purrr::discard()    masks scales::discard()
✖ dplyr::filter()     masks stats::filter()
✖ dplyr::lag()        masks stats::lag()
✖ purrr::map()        masks maps::map()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```

```
library(tidyr)
library(dplyr)
```

Part 1: Defining Research Question

Chosen Question: Does the state that uses the most renewable energy in 2023 have the most EV registrations in 2023?

Part 2: Data Preparation and Cleaning

```
#Load library and data

ev_2023 <- read.csv("data/ev-registrations-by-state-2023.csv")
renew_2023 <- read.csv("data/renew-use-2023.csv")
renew_2023$Renewable_Use_2023 <- suppressWarnings(as.numeric(gsub("[^0-9.]",
"", renew_2023$Renewable_Use_2023)))
renew_2023$State <- toupper(renew_2023$State)
# Calculate total renewable by state
renewable_total <- renew_2023 |>
  group_by(State) %>%
  summarize(Total_Renewable = sum(Renewable_Use_2023, na.rm = TRUE))
state_lookup <- data.frame(
  Full_Name = toupper(c("Alabama", "Alaska", "Arizona", "Arkansas",
"California",
                        "Colorado", "Connecticut", "Delaware", "District of
Columbia",
                        "Florida", "Georgia", "Hawaii", "Idaho", "Illinois",
"Indiana",
                        "Iowa", "Kansas", "Kentucky", "Louisiana", "Maine",
"Maryland",
                        "Massachusetts", "Michigan", "Minnesota",
"Mississippi", "Missouri",
                        "Montana", "Nebraska", "Nevada", "New Hampshire", "New
Jersey",
                        "New Mexico", "New York", "North Carolina", "North
Dakota", "Ohio",
                        "Oklahoma", "Oregon", "Pennsylvania", "Rhode Island",
```

```r
"South Carolina",
                        "South Dakota", "Tennessee", "Texas", "Utah",
"Vermont", "Virginia",
                        "Washington", "West Virginia", "Wisconsin",
"Wyoming")),
  Abbrev = c("AL", "AK", "AZ", "AR", "CA", "CO", "CT", "DE", "DC", "FL", "GA",
"HI",
             "ID", "IL", "IN", "IA", "KS", "KY", "LA", "ME", "MD", "MA", "MI",
"MN",
             "MS", "MO", "MT", "NE", "NV", "NH", "NJ", "NM", "NY", "NC", "ND",
"OH",
             "OK", "OR", "PA", "RI", "SC", "SD", "TN", "TX", "UT", "VT", "VA",
"WA",
             "WV", "WI", "WY")
)

ev_2023 <- ev_2023 %>%
  rename(State_Full = 1, EV_Count = 2) %>%
  mutate(
    EV_Count = as.numeric(gsub("[^0-9]", "", EV_Count)),
    State_Full = toupper(trimws(State_Full))
  ) %>%
  left_join(state_lookup, by = c("State_Full" = "Full_Name"))|>
  select(State = Abbrev, EV_Count) |>
  filter(!is.na(State), State != "")
```

Part 3: Joining / Pivoting Datasets for Analysis

```r
analysis_data <- left_join(renewable_total, ev_2023, by = "State") %>%
  mutate(EV_per_Renewable = (EV_Count / Total_Renewable) * 100)
cat("Top 5 Renewable Energy States:\n")
```

```
Top 5 Renewable Energy States:
```

```r
top_renewable <- analysis_data |> arrange(desc(Total_Renewable))
clean_data <- top_renewable %>%
  filter(!is.na(Total_Renewable), !is.na(EV_Count))
print(head(clean_data |> select(State, Total_Renewable, EV_Count), 5))
```

```
# A tibble: 5 × 3
  State Total_Renewable EV_Count
  <chr>           <dbl>    <dbl>
1 CA            1065179  1256646
2 TX             791210   230125
3 IA             414801     9031
```

```
4 WA              365955   152101
5 GA              291462    92368
```

```
cat("\nTop 5 EV Registration States:\n")
```

```


Top 5 EV Registration States:
```

```
top_ev <- analysis_data |> arrange(desc(EV_Count))
print(head(top_ev |> select(State, Total_Renewable, EV_Count), 5))
```

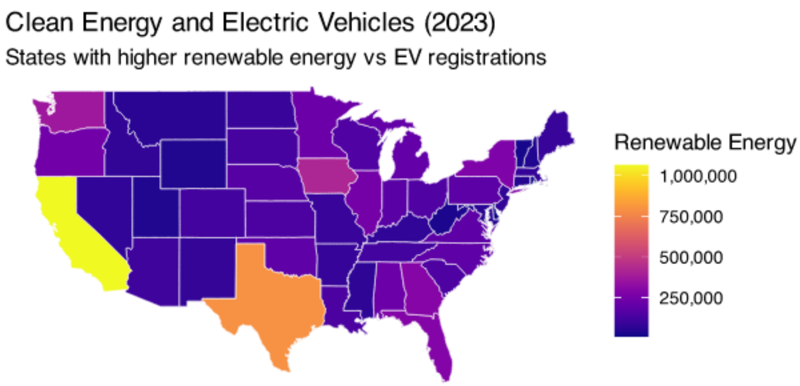
```
# A tibble: 5 × 3
  State Total_Renewable EV_Count
  <chr>           <dbl>    <dbl>
1 CA            1065179  1256646
2 FL             286307   254878
3 TX             791210   230125
4 WA             365955   152101
5 NJ              74409   134753
```

Part 4: Mapping Visualization

```
# US map and join with analysis data
us_map <- map_data("state")
analysis_data$region <- tolower(state.name[match(analysis_data$State,
state.abb)])
map_data <- left_join(us_map, analysis_data, by = "region")

# Plot
ggplot(map_data, aes(long, lat, group = group, fill = Total_Renewable)) +
  geom_polygon(color = "white", size = 0.2) +
  scale_fill_viridis(option = "plasma", name = "Renewable Energy", labels =
comma) +
  labs(title = "Clean Energy and Electric Vehicles (2023)",
       subtitle = "States with higher renewable energy vs EV registrations") +
  coord_fixed(1.3) +
  theme_void() +
  theme(legend.position = "right")
```

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



Part 5: Analysis

Based off my analysis and observations I can say that yes the state that uses the most renewable energy has the most EV registrations in 2023. Through my exploration I found that California had both the most renewable energy at 1065179 and the highest EV registrations at 1256646. I found a pattern that while CA and WA stayed stagnant with both being 1st and 4th in both tables of most renewable energy and EV registrations there were no other patterns for 2,3,and 5th places. Now this may be due to land size as California is huge with Texas following, so by nature more renewable energy will be produced here but that does not mean it is from EV registrations as Texas is third place for that. My map helps pinpoint where the most clean energy comes from.