EV Power - Lab 4 Project Report

Example Solution 1

Part 0: libraries

```
library(tidyverse)
```

```
Warning: package 'ggplot2' was built under R version 4.3.3

Warning: package 'purrr' was built under R version 4.3.3

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✔ dplyr     1.1.3     ✔ readr     2.1.4
✔ forcats   1.0.0     ✔ stringr   1.5.1
✔ ggplot2   3.5.2     ✔ tibble    3.2.1
✔ lubridate 1.9.2     ✔ tidyr     1.3.0
✔ purrr     1.0.4
── 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
```

```
library(ggplot2)
library(leaflet)
```

Part 1: Defining Research Question

Chosen Question: Are EV registrations concentrated in states with cleaner (more renewable) energy mixes? I think that states that generate a higher percentage of their electricity from renewable sources will have more electric vehicle registrations, due to the fact that electricity can be more affordable and widely spread.

Part 2: Data Preparation and Cleaning

```
renew21 <- read_csv("data/renew-use-2021.csv")
```

```
Rows: 260 Columns: 3
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (3): State, Energy_Source, Renewable_Use_2021

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.
```

```
renew22 <- 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.
```

```
renew23 <- read_csv("data/renew-use-2023.csv")
```

```
Rows: 260 Columns: 3
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (3): State, Energy_Source, Renewable_Use_2023

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.
```

```
renew21 <- renew21 |> rename(state = State, renewable_use = Rene
renew22 <- renew22 |> rename(state = State, renewable_use = Rene
renew23 <- renew23 |> rename(state = State, renewable_use = Rene

renew21$year <- 2021
renew22$year <- 2022
renew23$year <- 2023
```

```
renew_all <- bind_rows(renew21, renew22, renew23)
renew_all$state <- str_to_title(renew_all$state)

total21 <- read_csv("data/total-use-2021.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.
```

```
total22 <- read_csv("data/total-use-2022.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.
```

```
total23 <- 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.
```

```
total21_long <- total21 |>
  pivot_longer(
    cols = -Energy_Source,
    names_to = "state",
    values_to = "energy_use"
  ) |>
  group_by(state) |>
  summarize(total_energy_use = sum(energy_use, na.rm = TRUE)) |>
  mutate(year = 2021)

total22_long <- total22 |>
  pivot_longer(
    cols = -Energy_Source,
    names_to = "state",
    values_to = "energy_use"
  ) |>
  group_by(state) |>
  summarize(total_energy_use = sum(energy_use, na.rm = TRUE)) |>
  mutate(year = 2022)

total23_long <- total23 |>
  pivot_longer(
    cols = -Energy_Source,
    names_to = "state",
    values_to = "energy_use"
  ) |>
  group_by(state) |>
  summarize(total_energy_use = sum(energy_use, na.rm = TRUE)) |>
  mutate(year = 2023)

total_all <- bind_rows(total21_long, total22_long, total23_long)
total_all$state <- str_to_title(total_all$state)

ev <- read_csv("data/ev-registrations-by-state-2023.csv") |>
  rename(
    state = `electric vehicle registrations_by_state (2023)`,
    ev_registrations = `...2`
  )
```

```
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`
```

```
ev$state <- str_to_title(ev$state)


head(renew_all)
```

```
# A tibble: 6 × 4
  state Energy_Source renewable_use  year
  <chr> <chr>         <chr>         <dbl>
1 Ak    Biomass       ≈3153          2021
2 Ak    Geothermal    186 MMBtu      2021
3 Ak    Hydropower    5763 about     2021
4 Ak    Solar Energy  ~45            2021
5 Ak    Wind Energy   451 USD        2021
6 Al    Biomass       198543 est.    2021
```

```
head(total_all)
```

```
# A tibble: 6 × 3
  state total_energy_use  year
  <chr>            <dbl> <dbl>
1 Ak              684975  2021
2 Al             2352656  2021
3 Ar             1136025  2021
4 Az             1681257  2021
5 Ca             6142252  2021
6 Co             1364155  2021
```

```
head(ev)
```

```
# A tibble: 6 × 2
  state    ev_registrations
  <chr>    <chr>
1 <NA>     <NA>
2 State    Count-EVs
3 Alabama  #13047
4 Alaska   ~2697
5 Arizona  89798
6 Arkansas 7108 EVs
```

Part 3: Joining / Pivoting Datasets for Analysis

```
state_lookup <- tibble(
  state = str_to_title(state.abb),
  full_name = state.name
)

energy_joined <- renew_all |>
  left_join(total_all, by = c("state", "year")) |>
  mutate(
    renewable_use = as.numeric(gsub("[^0-9.]", "", renewable_use
    total_energy_use = as.numeric(gsub("[^0-9.]", "", total_ener
    pct_renew = renewable_use / total_energy_use * 100
  ) |>
  left_join(state_lookup, by = "state") |>
  mutate(state = full_name) |>
  select(-full_name)

head(energy_joined)
```

```
# A tibble: 6 × 6
  state   Energy_Source renewable_use  year total_energy_use
pct_renew
  <chr>   <chr>                 <dbl> <dbl>            <dbl>
<dbl>
1 Alaska  Biomass                3153  2021           684975
0.460
2 Alaska  Geothermal              186  2021           684975
0.0272
3 Alaska  Hydropower             5763  2021           684975
0.841
4 Alaska  Solar Energy             45  2021           684975
0.00657
5 Alaska  Wind Energy             451  2021           684975
0.0658
6 Alabama Biomass              198543  2021          2352656
8.44
```

```
ev_energy_2023 <- energy_joined |>
  filter(year == 2023) |>
  left_join(ev, by = "state") |>
  mutate(
    ev_registrations = as.numeric(gsub("[^0-9.]", "", ev_regist
    ev_per_energy = ev_registrations / total_energy_use
  )
```

```
head(ev_energy_2023)
```

```
# A tibble: 6 × 8
  state   Energy_Source renewable_use  year total_energy_use
pct_renew
  <chr>   <chr>                 <dbl> <dbl>            <dbl>
<dbl>
1 Alaska  Biomass                3404  2023           746979
0.456
2 Alaska  Geothermal              186  2023           746979
0.0249
3 Alaska  Hydropower             6051  2023           746979
0.810
4 Alaska  Solar Energy             67  2023           746979
0.00897
5 Alaska  Wind Energy             380  2023           746979
0.0509
6 Alabama Biomass              189040  2023          2265008
8.35
# ℹ 2 more variables: ev_registrations <dbl>, ev_per_energy
<dbl>
```

```
# Check
sum(!is.na(ev_energy_2023$ev_registrations))
```

```
[1] 250
```

```
# Summary statistics
summary(ev_energy_2023$pct_renew)
```

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
 0.00000  0.06978  0.45138  2.05981  2.92614 21.43650
```

```
summary(ev_energy_2023$ev_registrations)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
    959    8150   25833   70948   72139 1256646      10
```

```
cor(ev_energy_2023$pct_renew, ev_energy_2023$ev_registrations,
```

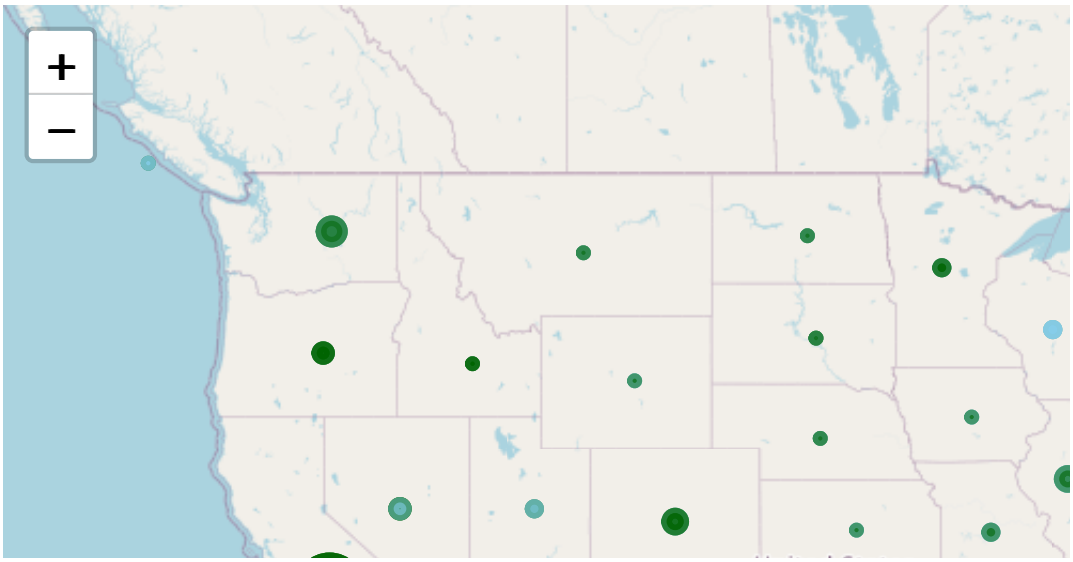
```
[1] 0.02980727
```

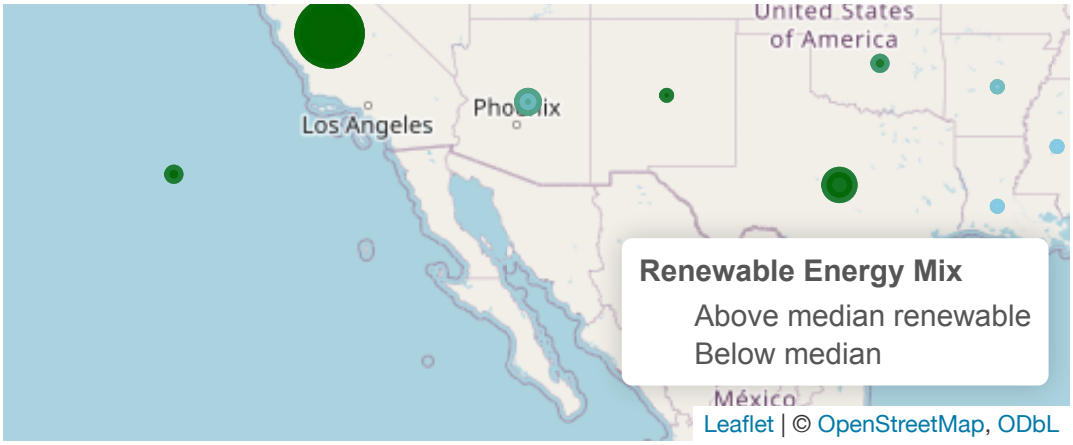
Part 4: Mapping Visualization

```r
state_coords <- tibble(
  state = state.name,
  lat = state.center$y,
  lng = state.center$x
)


map_data <- ev_energy_2023 |>
  filter(!is.na(pct_renew), !is.na(ev_registrations)) |>
  left_join(state_coords, by = "state")


leaflet(map_data) |>
  addTiles() |>
  addCircleMarkers(
    lng = ~lng,
    lat = ~lat,
    radius = ~sqrt(ev_registrations) / 80,
    color = ~ifelse(pct_renew > median(pct_renew, na.rm = TRUE),
                    "darkgreen", "skyblue"),
    fillOpacity = 0.6,
    popup = ~paste0(
      "<b>", state, "</b><br>",
      "EV Registrations: ", ev_registrations, "<br>",
      "Renewable Share: ", round(pct_renew, 1), "%"
    )
  ) |>
  addLegend(
    position = "bottomright",
    colors = c("darkgreen", "skyblue"),
    labels = c("Above median renewable", "Below median"),
    title = "Renewable Energy Mix"
  )
```





I developed this interactive map using the leaflet library which visualizes EV registrations and renewable energy share across U.S. states for 2023. Each circle's size represents the number of EVs in that specific state, while its color indicates whether the state's renewable share is above or below the median. Although some high-renewable states also have many EVs (such as CA & WA), EV adoption doesn't seem to be best indicated by renewable energy, which might point us to the fact that the quantity of EV registrations might depend on additional factors such as income levels, and infrastructure.