EV Power - Lab 4 Project Report

Example Solution 1

Part 0: libraries

```
# ---- General setup ----
# Always set a CRAN mirror so R can install packages without errors
options(repos = c(CRAN = "https://cloud.r-project.org"))

# Install (if needed) and load packages
pkgs <- c("tidyverse", "sf", "maps", "ggplot2", "viridis")
for(p in pkgs) if(!require(p, character.only = TRUE)) install.packages(p)
```

```
Loading required package: tidyverse
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.1     v stringr   1.5.2
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.1.0
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to beco
Loading required package: sf
```

```
Linking to GEOS 3.13.0, GDAL 3.8.5, PROJ 9.5.1; sf_use_s2() is TRUE
```

```
Loading required package: maps
```

```
Attaching package: 'maps'


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

    map


Loading required package: viridis

Loading required package: viridisLite


Attaching package: 'viridis'


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

    unemp
```

```
lapply(pkgs, library, character.only = TRUE)
```

```
[[1]]
 [1] "viridis"     "viridisLite" "maps"        "sf"          "lubridate"
 [6] "forcats"     "stringr"     "dplyr"       "purrr"       "readr"
[11] "tidyr"       "tibble"      "ggplot2"     "tidyverse"   "stats"
[16] "graphics"    "grDevices"   "utils"       "datasets"    "methods"
[21] "base"

[[2]]
 [1] "viridis"     "viridisLite" "maps"        "sf"          "lubridate"
 [6] "forcats"     "stringr"     "dplyr"       "purrr"       "readr"
[11] "tidyr"       "tibble"      "ggplot2"     "tidyverse"   "stats"
[16] "graphics"    "grDevices"   "utils"       "datasets"    "methods"
[21] "base"

[[3]]
 [1] "viridis"     "viridisLite" "maps"        "sf"          "lubridate"
 [6] "forcats"     "stringr"     "dplyr"       "purrr"       "readr"
[11] "tidyr"       "tibble"      "ggplot2"     "tidyverse"   "stats"
```

```
[16] "graphics"    "grDevices"   "utils"       "datasets"    "methods"
[21] "base"

[[4]]
 [1] "viridis"     "viridisLite" "maps"        "sf"          "lubridate"
 [6] "forcats"     "stringr"     "dplyr"       "purrr"       "readr"
[11] "tidyr"       "tibble"      "ggplot2"     "tidyverse"   "stats"
[16] "graphics"    "grDevices"   "utils"       "datasets"    "methods"
[21] "base"

[[5]]
 [1] "viridis"     "viridisLite" "maps"        "sf"          "lubridate"
 [6] "forcats"     "stringr"     "dplyr"       "purrr"       "readr"
[11] "tidyr"       "tibble"      "ggplot2"     "tidyverse"   "stats"
[16] "graphics"    "grDevices"   "utils"       "datasets"    "methods"
[21] "base"
```

Part 1: Defining Research Question

Chosen Question: Do states with higher shares of renewable energy also have more EV registrations?

Part 2: Data Preparation and Cleaning

```
# Load data
renew_2023 <- 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.
```

```
total_2023 <- 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.
```

```
ev_2023 <- 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(renew_2023)
```

```
[1] "State"              "Energy_Source"      "Renewable_Use_2023"
```

```
names(total_2023)
```

```
 [1] "Energy_Source" "AK"            "AL"            "AR"
 [5] "AZ"            "CA"            "CO"            "CT"
 [9] "DC"            "DE"            "FL"            "GA"
[13] "HI"            "IA"            "ID"            "IL"
[17] "IN"            "KS"            "KY"            "LA"
[21] "MA"            "MD"            "ME"            "MI"
[25] "MN"            "MO"            "MS"            "MT"
[29] "NC"            "ND"            "NE"            "NH"
[33] "NJ"            "NM"            "NV"            "NY"
[37] "OH"            "OK"            "OR"            "PA"
[41] "RI"            "SC"            "SD"            "TN"
[45] "TX"            "UT"            "VA"            "VT"
[49] "WA"            "WI"            "WV"            "WY"
[53] "US"
```

```r
names(ev_2023)
```

```
[1] "electric vehicle registrations_by_state (2023)"
[2] "...2"
```

Part 3: Joining / Pivoting Datasets for Analysis

```r
# ---- Clean and reshape renewable data ----
renew_2023_clean <- renew_2023 %>%
  rename_with(~str_replace_all(.x, "\\s+", "_")) %>%
  rename(State = State, Energy_Source = Energy_Source, Renewable_Use_2023 = Renewable_Use_202
  group_by(State) %>%
  summarise(renewable_energy = sum(as.numeric(Renewable_Use_2023), na.rm = TRUE))
```

```
Warning: There were 5 warnings in `summarise()`.
The first warning was:
i In argument: `renewable_energy = sum(as.numeric(Renewable_Use_2023), na.rm =
  TRUE)`.
i In group 1: `State = "AK"`.
Caused by warning:
! NAs introduced by coercion
i Run `dplyr::last_dplyr_warnings()` to see the 4 remaining warnings.
```

```r
# ---- Clean and reshape total energy data ----
total_2023_long <- total_2023 %>%
  pivot_longer(
    cols = -Energy_Source,
    names_to = "State",
    values_to = "Total_Use_2023"
  ) %>%
  group_by(State) %>%
  summarise(total_energy = sum(as.numeric(Total_Use_2023), na.rm = TRUE))

# ---- Clean EV registrations ----
ev_2023_clean <- ev_2023 %>%
  set_names(c("State", "EV_Registrations")) %>%
  mutate(EV_Registrations = as.numeric(gsub(",", "", EV_Registrations)))
```

```
Warning: There was 1 warning in `mutate()`.
i In argument: `EV_Registrations = as.numeric(gsub(",", "",
  EV_Registrations))`.
Caused by warning:
! NAs introduced by coercion
```

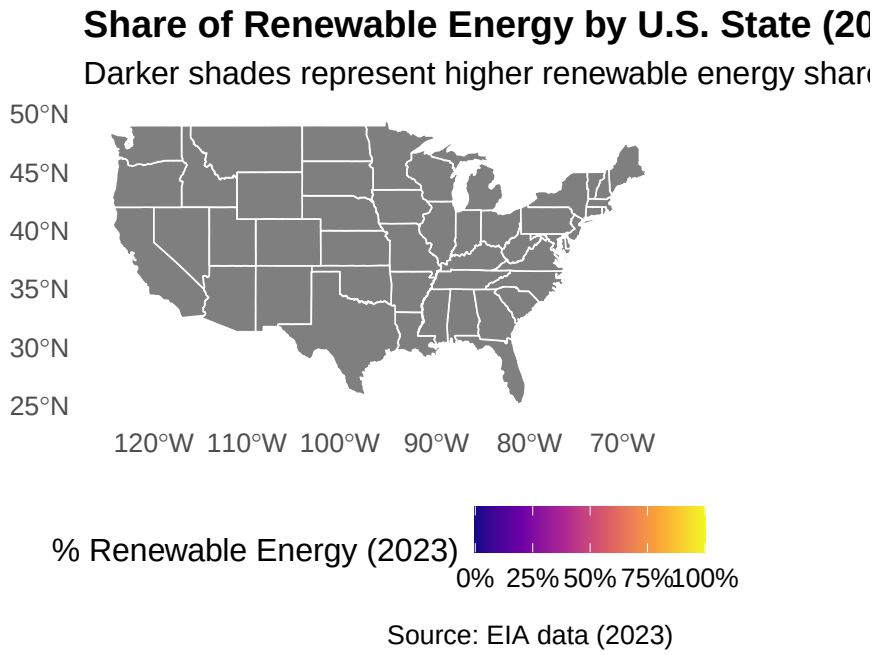
```r
# ---- Merge all datasets ----
energy_merged <- renew_2023_clean %>%
  inner_join(total_2023_long, by = "State") %>%
  inner_join(ev_2023_clean, by = "State") %>%
  mutate(
    pct_renewable = 100 * renewable_energy / total_energy,
    ev_per_million_energy = EV_Registrations / total_energy
  )

head(energy_merged)
```

```
# A tibble: 0 x 6
# i 6 variables: State <chr>, renewable_energy <dbl>, total_energy <dbl>,
#   EV_Registrations <dbl>, pct_renewable <dbl>, ev_per_million_energy <dbl>
```

Part 4: Mapping Visualization

```r
# Load libraries for mapping
usa_states <- st_as_sf(map("state", plot = FALSE, fill = TRUE)) %>%
mutate(state = str_to_title(ID))




map_data <- usa_states %>%
  left_join(energy_merged, by = c("ID" = "State"))



ggplot(map_data) +
geom_sf(aes(fill = pct_renewable), color = "white", linewidth = 0.3) +
scale_fill_viridis_c(
name = "% Renewable Energy (2023)",
option = "C",
```

```
limits = c(0, 100),
labels = scales::label_number(suffix = "%")
) +
labs(
title = "Share of Renewable Energy by U.S. State (2023)",
subtitle = "Darker shades represent higher renewable energy share",
caption = "Source: EIA data (2023)"
) +
theme_minimal(base_size = 12) +
theme(
panel.grid = element_blank(),
plot.title = element_text(face = "bold"),
legend.position = "bottom"
)
```




```
# ---- Scatterplot: Renewable Energy vs EVs ----
#ggplot(energy_merged, aes(x = pct_renewable, y = EV_Registrations)) +
  #geom_point(color = "darkgreen", alpha = 0.7, size = 3) +
  #geom_smooth(method = "lm", se = FALSE, color = "black") +
  #labs(
    #x = "Renewable Energy Share (%)",
    #y = "EV Registrations (2023)",
```

```
  #  title = "Relationship Between Renewable Energy and EV Adoption",
#    caption = "Source: EIA and DMV data (2023)"
#  ) +
#  theme_minimal(base_size = 12)

# Print correlation
#correlation <- cor(energy_merged$pct_renewable, energy_merged$EV_Registrations, use = "compl
#correlation


 # analysis: The map highlights regional differences in renewable energy generation. Western
```