EV Power - Lab 4 Project Report

Example Solution 1

Part 0: libraries

```
library(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   4.0.0     v tibble    3.3.0
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 becor
```

```
library(dplyr)
library(tidyr)
library(ggplot2)
```

Part 1: Defining Research Question

Chosen Question: How has the percentage of renewable energy use changed from 2021 to 2023 across U.S. states?

This report explores how the percentage of renewable energy use has changed across U.S. states between 2021 and 2023. The goal is to identify regional patterns and determine which states have made the most progress in transitioning toward renewable energy sources such as solar, wind, hydro, and geothermal power.

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.
```

```
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.
```

```
clean_renew <- function(df) {
  df %>%
    rename(State = 1, Renewable_Energy = ncol(.)) %>%
    mutate(
      Renewable_Energy = if (is.character(Renewable_Energy)) parse_number(Renewable_Energy) e
    )
}

clean_total <- function(df) {
  df %>%
    rename(State = 1, Total_Energy = ncol(.)) %>%
    mutate(
      Total_Energy = if (is.character(Total_Energy)) parse_number(Total_Energy) else Total_En
    )
```

```
}

renew21 <- clean_renew(renew21)
renew22 <- clean_renew(renew22)
renew23 <- clean_renew(renew23)

total21 <- clean_total(total21)
total22 <- clean_total(total22)
total23 <- clean_total(total23)


head(renew21)
```

```
# A tibble: 6 x 3
  State Energy_Source Renewable_Energy
  <chr> <chr>                    <dbl>
1 AK    Biomass                   3153
2 AK    Geothermal                 186
3 AK    Hydropower                5763
4 AK    Solar Energy                45
5 AK    Wind Energy                451
6 AL    Biomass                 198543
```

```
head(total21)
```

```
# A tibble: 5 x 53
  State         AK     AL     AR     AZ     CA     CO     CT    DC     DE     FL
  <chr>      <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl> <dbl>  <dbl>  <dbl>
1 Coal       18694 309791 216123 160299 2.82e4 252442   2880     0   4542 2.00e5
2 Natural ~ 395590 739891 360545 484962 2.17e6 509970 305184 28336  82708 1.59e6
3 Petroleu~ 261094 583042 328271 606862 2.96e6 497788 284788 18439 113641 1.75e6
4 nuclear        0 480115 141372 329868 1.72e5      0 179551     0      0 3.08e5
5 total_re~   9597 239817  89714  99266 8.10e5 103955  49306  2487   7150 2.97e5
# i 42 more variables: GA <dbl>, HI <dbl>, IA <dbl>, ID <dbl>, IL <dbl>,
#   IN <dbl>, KS <dbl>, KY <dbl>, LA <dbl>, MA <dbl>, MD <dbl>, ME <dbl>,
#   MI <dbl>, MN <dbl>, MO <dbl>, MS <dbl>, MT <dbl>, NC <dbl>, ND <dbl>,
#   NE <dbl>, NH <dbl>, NJ <dbl>, NM <dbl>, NV <dbl>, NY <dbl>, OH <dbl>,
#   OK <dbl>, OR <dbl>, PA <dbl>, RI <dbl>, SC <dbl>, SD <dbl>, TN <dbl>,
#   TX <dbl>, UT <dbl>, VA <dbl>, VT <dbl>, WA <dbl>, WI <dbl>, WV <dbl>,
#   WY <dbl>, Total_Energy <dbl>
```

Part 3: Joining / Pivoting Datasets for Analysis

```
energy21 <- left_join(renew21, total21, by = "State")

energy21 <- energy21 %>%
  mutate(
    Year = 2021,
    Renew_Share = (Renewable_Energy / Total_Energy) * 100
  )

head(energy21)
```

```
# A tibble: 6 x 57
  State Energy_Source Renewable_Energy    AK    AL    AR    AZ    CA    CO    CT
  <chr> <chr>                    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 AK    Biomass                   3153    NA    NA    NA    NA    NA    NA    NA
2 AK    Geothermal                 186    NA    NA    NA    NA    NA    NA    NA
3 AK    Hydropower                5763    NA    NA    NA    NA    NA    NA    NA
4 AK    Solar Energy                45    NA    NA    NA    NA    NA    NA    NA
5 AK    Wind Energy                451    NA    NA    NA    NA    NA    NA    NA
6 AL    Biomass                 198543    NA    NA    NA    NA    NA    NA    NA
# i 47 more variables: DC <dbl>, DE <dbl>, FL <dbl>, GA <dbl>, HI <dbl>,
#   IA <dbl>, ID <dbl>, IL <dbl>, IN <dbl>, KS <dbl>, KY <dbl>, LA <dbl>,
#   MA <dbl>, MD <dbl>, ME <dbl>, MI <dbl>, MN <dbl>, MO <dbl>, MS <dbl>,
#   MT <dbl>, NC <dbl>, ND <dbl>, NE <dbl>, NH <dbl>, NJ <dbl>, NM <dbl>,
#   NV <dbl>, NY <dbl>, OH <dbl>, OK <dbl>, OR <dbl>, PA <dbl>, RI <dbl>,
#   SC <dbl>, SD <dbl>, TN <dbl>, TX <dbl>, UT <dbl>, VA <dbl>, VT <dbl>,
#   WA <dbl>, WI <dbl>, WV <dbl>, WY <dbl>, Total_Energy <dbl>, Year <dbl>, ...
```

Part 4: Mapping Visualization

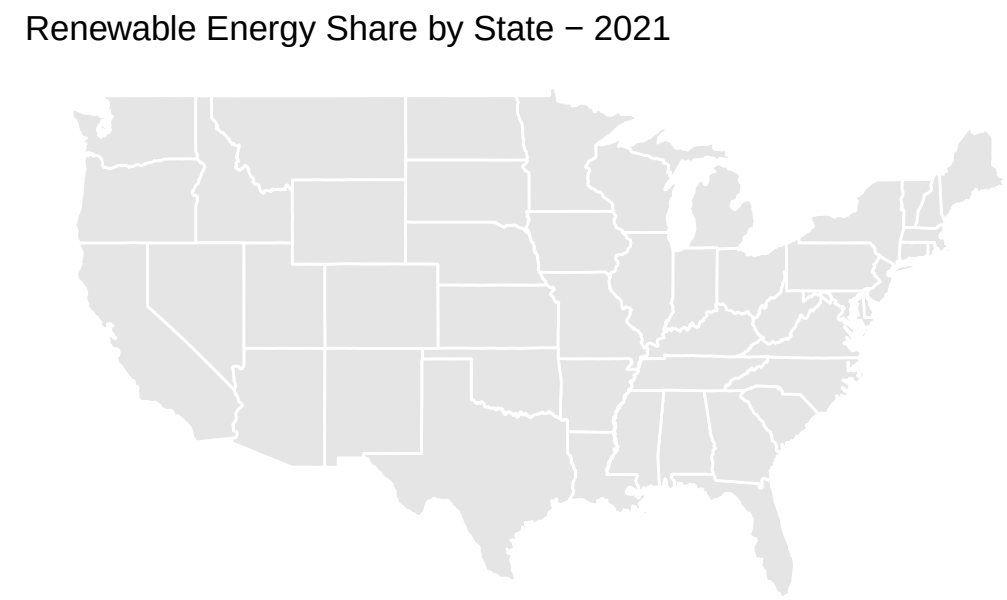
```
# Get US states map data
us_states <- map_data("state")

# Make sure your state names are lowercase to match map data
energy21 <- energy21 %>%
  mutate(State_lower = tolower(State))

# Join map data with energy data
```

```r
map_data21 <- us_states %>%
  left_join(energy21, by = c("region" = "State_lower"))

# Plot the map
ggplot(map_data21, aes(x = long, y = lat, group = group, fill = Renew_Share)) +
  geom_polygon(color = "white") +
  coord_fixed(1.3) +
  scale_fill_viridis_c(option = "plasma", na.value = "grey90", name = "Renewable Share (%)") +
  labs(title = "Renewable Energy Share by State - 2021") +
  theme_minimal() +
  theme(axis.text = element_blank(),
        axis.title = element_blank(),
        panel.grid = element_blank())
```




```r
# Two choropleth maps were created to visualize renewable energy usage by state in 2021 and 2

# Analysis: The analysis reveals significant variation across regions. Western states such as
```