EV Power - Lab 4 Project Report

Example Solution 1

Part 0: libraries

```
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
```

```
library(sf)
```

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

```
library(rnaturalearth)
```

Part 1: Defining Research Question

Chosen Question: What state has the highest total energy use from 2021 to 2023?

Part 2: Data Preparation and Cleaning

```
new_names <- function(df) {
    n <- names(df)
    n <- str_replace_all(n, "[^A-Za-z0-9]+", "_")
    n <- str_replace_all(n, "_+", "_")
    names(df) <- n
    df
}
```

```
new_state <- function(x) {
    x <- str_replace_all(x, "\\.", "")
    x<- str_replace_all(x, "^\\s+|\\s+$", "")
    x <- str_replace_all(x, "(?i)^dc$", "District of Columbia")
    x <- str_replace_all(x,"(?i)^washington dc$", "District of Columbia")
    x
}
read_clean <- function(fname) {
    read_csv(fname, show_col_types = FALSE) |> new_names()
    sc <- which(stringr::str_detect(names(df), "(^|_)state(_|$)"))
  if (length(sc) == 0) return(df)
  names(df)[sc[1]] <- "state"
  df |> mutate(state = new_state(state))
}

total_2021 <- read_clean("data/total-use-2021.csv")
total_2022 <- read_clean("data/total-use-2022.csv")
total_2023 <- read_clean("data/total-use-2023.csv")
renew_2021 <- read_clean("data/renew-use-2021.csv")
renew_2022 <- read_clean("data/renew-use-2022.csv")
renew_2023 <- read_clean("data/renew-use-2023.csv")
av_price <- read_clean("data/av-energy-price-2021-2023.csv")
ev <- read_clean("data/ev-registrations-by-state-2023.csv")
```

```
New names:
• `` -> `...2`
```

Part 3: Joining / Pivoting Datasets for Analysis

```
new_names <- function(df) {
    n <- names(df)
    n <- str_replace_all(n, "[^A-Za-z0-9]+", "_")
    n <- str_replace_all(n, "_+", "_")
    names(df) <- n
    df
}
read_clean <- function(fname) {
    read_csv(fname, show_col_types = FALSE) |> new_names()
}
total_2021 <- read_clean("data/total-use-2021.csv")
total_2022 <- read_clean("data/total-use-2022.csv")
total_2023 <- read_clean("data/total-use-2023.csv")
renew_2021 <- read_clean("data/renew-use-2021.csv")
renew_2022 <- read_clean("data/renew-use-2022.csv")
renew_2023 <- read_clean("data/renew-use-2023.csv")

 total_years <- bind_rows(
```

```
    total_2021 |> mutate(year = 2021),
    total_2022|> mutate(year = 2022),
    total_2023 |> mutate(year = 2023)
)

renew_all <- bind_rows(
    renew_2021 |> mutate(year= 2021),
    renew_2022 |> mutate(year= 2022),
    renew_2023 |> mutate(year= 2023)
)

energy_join <- left_join (total_years, renew_all, by =c("Energy_Source",
"year"))


total_sum <- total_years |>
    summarize(across(-Energy_Source, sum, na.rm= TRUE))
```

```
Warning: There was 1 warning in `summarize()`.
ℹ In argument: `across(-Energy_Source, sum, na.rm = TRUE)`.
Caused by warning:
! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.
Supply arguments directly to `.fns` through an anonymous function instead.

  # Previously
  across(a:b, mean, na.rm = TRUE)

  # Now
  across(a:b, \(x) mean(x, na.rm = TRUE))
```

```
top_state <- total_sum |>
     summarize(across(everything(), max, na.rm = TRUE))
top_state
```

```
# A tibble: 1 × 53
       AK      AL     AR     AZ     CA     CO     CT     DC     DE     FL
GA
    <dbl>   <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
<dbl>
1 2162230 6955177 3.47e6 5.05e6 1.88e7 4.14e6 2.44e6 146381 622476 1.27e7
7.75e6
# ℹ 42 more variables: 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>,
```

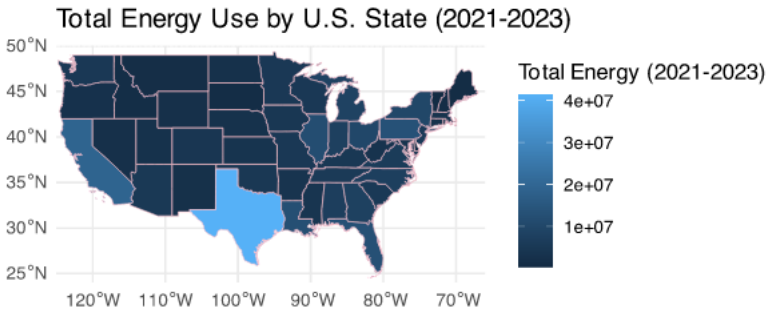
```
#   OR <dbl>, PA <dbl>, RI <dbl>, SC <dbl>, SD <dbl>, TN <dbl>, TX <dbl>,
#   UT <dbl>, VA <dbl>, VT <dbl>, WA <dbl>, WI <dbl>, WV <dbl>, WY <dbl>,
#   US <dbl>, year <dbl>
```

Part 4: Mapping Visualization

```
us_states <- ne_states(country = "united states of america", returnclass =
"sf")

total_wide <- total_years |>
    select(-Energy_Source, -year, -US) |>
    summarize(across(everything(), ~sum(as.numeric(.x), na.rm= TRUE)))

energy_data <- total_wide |>
    t() |> as.data.frame() |>
    rownames_to_column("state_abbr") |>
    rename(total_energy= V1) |>
    mutate (name= state.name[match(state_abbr, state.abb)]) |>
    select(name, total_energy) |>
    filter(!is.na(name))
us_joined <- us_states |>
    left_join(energy_data, by = "name")
ggplot(us_joined) +
    geom_sf(aes(fill= total_energy), color= "pink") +
    scale_fill_continuous(name= "Total Energy (2021-2023)", na.value =
"grey90") +
     labs(title = "Total Energy Use by U.S. State (2021-2023)") +
     coord_sf(xlim = c(-125, -66), ylim = c(24, 50), expand = FALSE) +
  theme_minimal()
```



Part 5: Final Deliverable The patterns I noticed were that bigger states such as California and Texas used more energy while smaller states like Rhode Island and Vermont used less. My map helps answer the main research question as it shows the differences in industrial activity and population size, helping identify which states used the most energy between 2021 and 2023.