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(readr)
library(stringr)
library(dplyr)
library(tidyr)
library(ggplot2)
library(maps)
```

```
Attaching package: 'maps'

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

    map
```

Part 1: Defining Research Question

Chosen Question: - How has each state's share of total renewable energy changed from 2021 to 2023.

```
renew21 <- read_csv("data/renew-use-2021.csv") |>
    mutate(year = 2021)
```

```
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") |>
    mutate(year = 2022)
```

```
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") |>
    mutate(year = 2023)
```

```
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") |>
    mutate(year = 2021)
```

```
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") |>
    mutate(year = 2022)
```

```
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") |>
    mutate(year = 2023)
```

```
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.
```

Part 2: Data Preparation and Cleaning

```r
clean_num <- function(x) {
  x|>
    str_replace_all("[^0-9.]", "") |>
    as.numeric()
}
```

Part 3: Joining / Pivoting Datasets for Analysis

```r
renew_all <- bind_rows(renew21, renew22, renew23) |>
    mutate(State = str_trim(toupper(State)),
    Renewable_Use = coalesce(clean_num(Renewable_Use_2021), clean_num(Renewable_Use_2022), cl
        group_by(State, year) |>
        summarise(Renewable_Use = sum(Renewable_Use, na.rm = TRUE), .groups = "drop")

total_all <- bind_rows(total21, total22, total23) |>
    pivot_longer(cols = -c(Energy_Source, year),
    names_to = "State",
    values_to = "total_energy_use") |>
        mutate(State = str_trim(toupper(State)),
        total_energy_use = as.numeric(str_replace_all(total_energy_use, "[^0-9.]", ""))) |>
            filter(Energy_Source == "total_renewable_energy") |>
            select(State, year, total_energy_use)

price <- read_csv("data/av-energy-price-2021-2023.csv", skip = 2)
```

```
Rows: 52 Columns: 1
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (1): State,2021,2022,2023

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.
```

```r
price <- price |>
    rename_with(~ str_replace_all(., "[^A-Za-z0-9]", "")) |>
    mutate(across(matches("20[0-9]{2}"), ~ as.numeric(str_replace_all(., "[^0-9.]", ""))))
```

```
Warning: There was 1 warning in `mutate()`.
```

```
i In argument: `across(...)`.
Caused by warning:
! NAs introduced by coercion
```

```
ev <- read_csv("data/ev-registrations-by-state-2023.csv", show_col_types = FALSE) |>
  rename_with(~ str_replace_all(., "[^A-Za-z0-9]", "")) |>
  rename(State = matches("(?i)state"))
```

```
New names:
* `` -> `...2`
```

```
energy <- renew_all |>
  left_join(total_all, by = c("State", "year")) |>
  mutate(renew_share = Renewable_Use / total_energy_use)

renew_change <- energy |>
    pivot_wider(names_from = year, values_from = renew_share) |>
    mutate(change_21_23 = `2023` - `2021`)
```

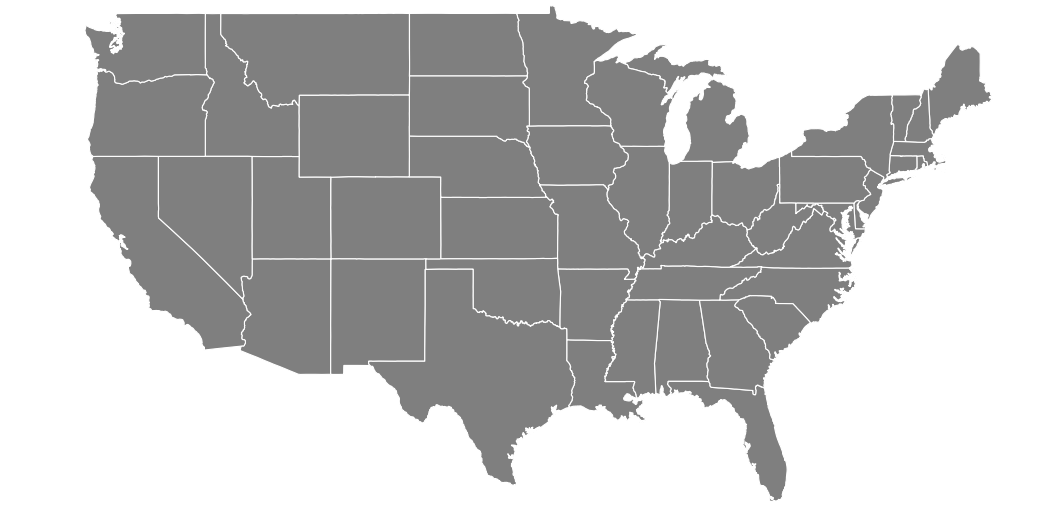
Part 4: Mapping Visualization

```
state_names <- tibble(State = state.abb, region = tolower(state.name))

map_data_join <- renew_change |>
    left_join(state_names, by = "State") |>left_join(ev, by = "State")

us_map <- map_data("state")
map_df <- left_join(us_map, map_data_join, by = "region")
```

```
Warning in left_join(us_map, map_data_join, by = "region"): Detected an unexpected many-to-ma
i Row 1 of `x` matches multiple rows in `y`.
i Row 4 of `y` matches multiple rows in `x`.
i If a many-to-many relationship is expected, set `relationship =
  "many-to-many"` to silence this warning.
```

```
ggplot(map_df, aes(long, lat, group = group, fill = change_21_23)) +
  geom_polygon(color = "white", linewidth = 0.2) +
  scale_fill_gradient2(
```

```
  low = "red", mid = "white", high = "green", midpoint = 0,
  name = "Change in Renewable Share (2021-2023)"
) +
coord_fixed(1.3) +
theme_void() +
labs(
  title = "Change in Renewable Energy Share by State (2021-2023)",
  subtitle = "Green = increase, Red = decrease",
  caption = "Source: Data Files Stat 133)"
)
```



Analysis

This map was supposed to show how the share of renewable energy changed between 2021 and 2023 across the US. I wanted the red states to show a decrease and the green states to show an increase. Maybe ambitiously with darkers shades showing more growth. I couldn't figure out why the map didn't end up working. I had to look up some new ways to use code and maybe I implemented it wrong because I couldn't understand it and maybe I iit was outside of teh scope for this project and I ended up just using it wrong. I thought this project was super interesting though! I might go back and try to go step by step through my mistakes. and try to find the issue.