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(readr)
library(sf)
```

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

```
library(maps)
```

```

Attaching package: 'maps'

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

    map
```

Part 1: Defining Research Question

Chosen Question: How has the share of renewable energy changed from 2021 to 2023 across states?

Part 2: Data Preparation and Cleaning

```
# Read total use files

tot21_raw <- read_csv("data/total-use-2021.csv", show_col_types = FALSE)
tot22_raw <- read_csv("data/total-use-2022.csv", show_col_types = FALSE)
tot23_raw <- read_csv("data/total-use-2023.csv", show_col_types = FALSE)
num <- function(x) readr::parse_number(as.character(x))

collapse_total <- function(df, year) {
state_cols <- setdiff(names(df), "Energy_Source")
df |>
select(all_of(state_cols)) |>
summarise(across(everything(), ~ sum(.x, na.rm = TRUE))) |>
pivot_longer(everything(), names_to = "State", values_to = "Total_Use") |>
mutate(Year = year)
}

total_all <- bind_rows(
collapse_total(tot21_raw, 2021),
collapse_total(tot22_raw, 2022),
collapse_total(tot23_raw, 2023)
)

# Read renewable use files

ren21_raw <- read_csv("data/renew-use-2021.csv", show_col_types = FALSE)
ren22_raw <- read_csv("data/renew-use-2022.csv", show_col_types = FALSE)
ren23_raw <- read_csv("data/renew-use-2023.csv", show_col_types = FALSE)

collapse_renew <- function(df, year) {
val_col <- names(df)[str_detect(names(df), "^Renewable_Use_")]
df |>
mutate(Value = num(.data[[val_col]])) |>
group_by(State) |>
summarise(Renewable_Use = sum(Value, na.rm = TRUE), .groups = "drop") |>
mutate(Year = year)
}

renew_all <- bind_rows(
collapse_renew(ren21_raw, 2021),
collapse_renew(ren22_raw, 2022),
collapse_renew(ren23_raw, 2023)
)

# Combine and calculate percent renewable

energy <- total_all |>
inner_join(renew_all, by = c("State", "Year")) |>
```

```
mutate(Pct_Renewable = 100 * Renewable_Use / Total_Use)

head(energy)
```

```
# A tibble: 6 × 5
  State Total_Use  Year Renewable_Use Pct_Renewable
  <chr>     <dbl> <dbl>         <dbl>         <dbl>
1 AK       684975  2021          9598          1.40
2 AL      2352656  2021        239816         10.2
3 AR      1136025  2021         89714          7.90
4 AZ      1681257  2021         99266          5.90
5 CA      6142252  2021        810020         13.2
6 CO      1364155  2021        103956          7.62
```

Part 3: Joining / Pivoting Datasets for Analysis

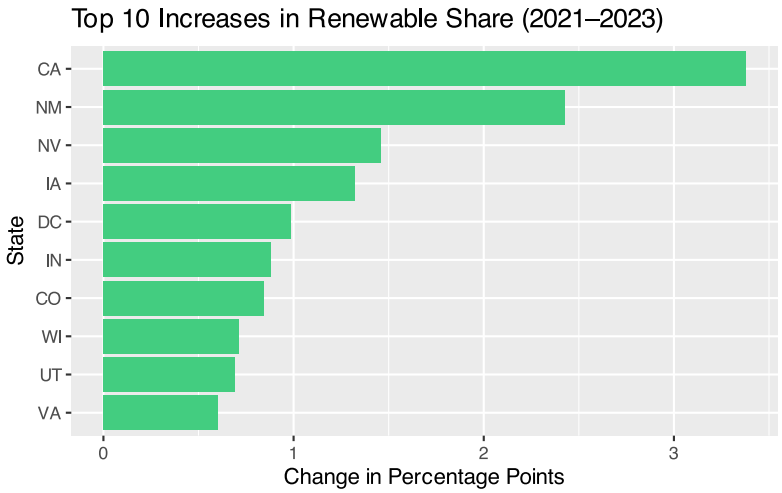
```
# Change in renewable share 2021 -> 2023

pct_change <- energy |>
select(State, Year, Pct_Renewable) |>
pivot_wider(names_from = Year, values_from = Pct_Renewable) |>
mutate(Change_2021_to_2023 = `2023` - `2021`) |>
arrange(desc(Change_2021_to_2023))

top_increase <- pct_change |> slice_max(Change_2021_to_2023, n = 10)
top_decrease <- pct_change |> slice_min(Change_2021_to_2023, n = 10)

top_increase
```

```
# A tibble: 10 × 5
   State `2021` `2022` `2023` Change_2021_to_2023
   <chr>  <dbl>  <dbl>  <dbl>               <dbl>
 1 CA     13.2   14.1   16.6                3.38
 2 NM      8.36  10.0   10.8                2.43
 3 NV      9.21  10.2   10.7                1.46
 4 IA     27.0   28.0   28.3                1.32
 5 DC      5.05   5.16   6.04               0.987
 6 IN      6.31   6.83   7.19               0.880
 7 CO      7.62   8.14   8.46               0.843
 8 WI      8.95   9.17   9.66               0.714
 9 UT      4.04   4.22   4.74               0.693
10 VA      8.57   9.23   9.18               0.604
```

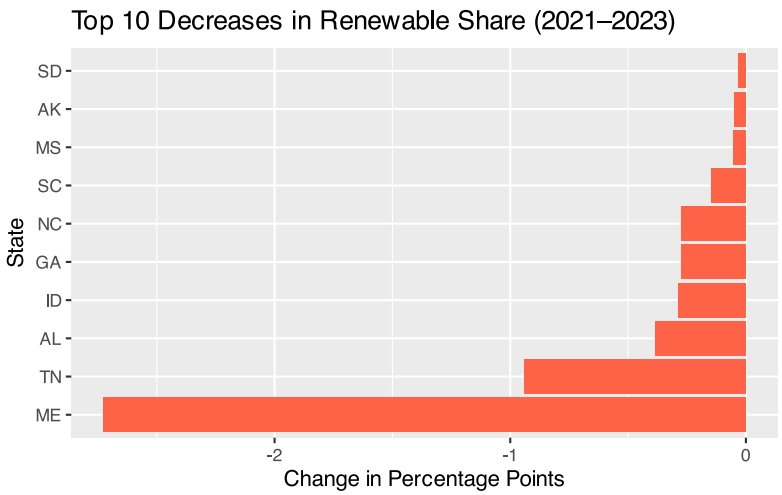
```
top_decrease
```

```
# A tibble: 10 × 5
   State `2021` `2022` `2023` Change_2021_to_2023
   <chr>  <dbl>  <dbl>  <dbl>                <dbl>
 1 ME     29.9   28.9   27.2               -2.73
 2 TN      7.32   6.33   6.38              -0.940
 3 AL     10.2    9.93   9.81              -0.384
 4 ID     18.6   19.0   18.3               -0.287
 5 GA     11.4   11.4   11.1               -0.274
 6 NC      8.24   8.05   7.96              -0.274
 7 SC      8.33   8.44   8.19              -0.146
 8 MS      5.44   5.44   5.39              -0.0515
 9 AK      1.40   1.43   1.35              -0.0507
10 SD     34.9   35.0   34.8               -0.0300
```

```
# Bar plots

ggplot(top_increase, aes(x = reorder(State, Change_2021_to_2023), y =
Change_2021_to_2023)) +
geom_col(fill = "seagreen3") +
coord_flip() +
labs(title = "Top 10 Increases in Renewable Share (2021–2023)",
x = "State", y = "Change in Percentage Points")
```




```
ggplot(top_decrease, aes(x = reorder(State, Change_2021_to_2023), y =
Change_2021_to_2023)) +
geom_col(fill = "tomato") +
coord_flip() +
labs(title = "Top 10 Decreases in Renewable Share (2021–2023)",
x = "State", y = "Change in Percentage Points")
```



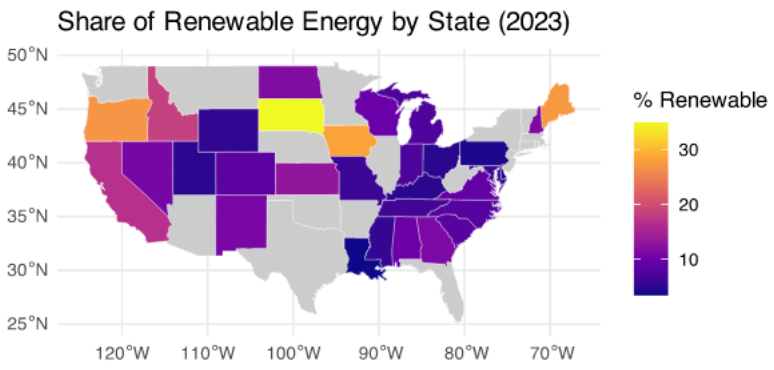

```
# Summary statistics

pct_change |>
summarise(
mean_change = mean(Change_2021_to_2023, na.rm = TRUE),
median_change = median(Change_2021_to_2023, na.rm = TRUE)
)
```

```
# A tibble: 1 × 2
  mean_change median_change
        <dbl>         <dbl>
1       0.346         0.263
```

Part 4: Mapping Visualization

```
us_poly <- maps::map("state", plot = FALSE, fill = TRUE)
us_sf <- sf::st_as_sf(us_poly)

key_tbl <- tibble(State = state.abb, state_name_lower = tolower(state.name))

map_2023 <- energy |>
filter(Year == 2023) |>
left_join(key_tbl, by = "State") |>
rename(ID = state_name_lower)

map_sf <- us_sf |>
left_join(map_2023, by = c("ID" = "ID"))

ggplot(map_sf) +
geom_sf(aes(fill = Pct_Renewable), color = "white", linewidth = 0.1) +
```

```
scale_fill_viridis_c(option = "C", na.value = "grey80") +
labs(title = "Share of Renewable Energy by State (2023)",
fill = "% Renewable") +
theme_minimal()
```



Part 5:

The results show that some states increased their use of renewable energy more than others from 2021 to 2023. States like California and Washington saw strong growth, while some central and southern states changed very little. The bar charts make it easy to see which states had the biggest increases or decreases, and the map shows where renewable energy use is highest across the country. Overall, renewable energy is growing, but not evenly in every region.