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   1.0.0     ✔ stringr   1.5.1
✔ ggplot2   3.5.1     ✔ tibble    3.2.1
✔ lubridate 1.9.4     ✔ tidyr     1.3.1
✔ purrr     1.0.4
── 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(janitor)
```

```

Attaching package: 'janitor'

The following objects are masked from 'package:stats':

    chisq.test, fisher.test
```

```
library(stringr)
```

Part 1: Defining Research Question

Chosen Question: Which 3 states experienced the largest increase in renewable energy usage between 2021 and 2023?

Part 2: Data Preparation and Cleaning

```
energy_2021 <- read_csv("/Users/hibahalam/Desktop/stat133/ev-power-hibahalam/
data/total-use-2021.csv") %>% clean_names()
```

```
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.
```

```
energy_2022 <- read_csv("/Users/hibahalam/Desktop/stat133/ev-power-hibahalam/
data/total-use-2022.csv") %>% clean_names()
```

```
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.
```

```
energy_2023 <- read_csv("/Users/hibahalam/Desktop/stat133/ev-power-hibahalam/
data/total-use-2023.csv") %>% clean_names()
```

```
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.
```

```r
standardize_cols <- function(df) {

  df <- df %>% janitor::clean_names()

  rename_candidates <- list(
    state            = matches("^state$|^state_name$|^region$"),
    year             = matches("^year$"),
    total_energy_use = matches("^total.*energy.*use$|^tot.*energy$"),
    renewable_use    = matches("^renew.*use$|^ren.*energy$"),
    avg_energy_price = matches("^avg.*price$|^average.*price$|
^energy.*price$"),
    ev_registrations = matches("^ev.*registrations?$|^ev_count$")
  )

  for (nm in names(rename_candidates)) {
    cand <- rename_candidates[[nm]]
    if (!nm %in% names(df)) {
      hit <- tryCatch(select(df, !!cand), error = function(e) NULL)
      if (!is.null(hit) && ncol(hit) == 1) {
        df <- df %>% rename(!!nm := !!cand)
      }
    }
  }
  df
}
abbr_to_full <- setNames(datasets::state.name, datasets::state.abb)

standardize_state <- function(x) {
  x %>%
    str_trim() %>%
    str_squish() %>%
    {\(s) {
      s_upper <- str_to_upper(s)
      s2 <- ifelse(s_upper %in% names(abbr_to_full), abbr_to_full[s_upper], s)
      s2
    }}() %>%
    str_replace_all("[,\\.]", "") %>%
    str_to_title() %>%
    recode(
      "Dc"                = "District Of Columbia",
      "Washington Dc"     = "District Of Columbia",
      "Washington D C"    = "District Of Columbia",
      "D C"               = "District Of Columbia",
      "Us Total"          = NA_character_,
      "United States"     = NA_character_
    ) %>%
    (\(s) ifelse(s == "District Of Columbia", "District of Columbia", s))()
}
```

Part 3: Joining / Pivoting Datasets for Analysis

```r
share_from_total_wide <- function(df, year_label) {
  # df has rows = energy_source, columns = states (e.g., ak, al, ...)
  # Make a tidy long table: (energy_source, state, btu)
  long <- df %>%
    pivot_longer(
      cols = -energy_source,
      names_to = "state",
      values_to = "btu"
    ) %>%
    mutate(
      # normalize source names to be safe (remove symbols like †, spaces,
parens)
      energy_source = energy_source %>%
        str_to_lower() %>%
        str_replace_all("\\s+", "_") %>%
        str_replace_all("[^a-z0-9_]", "")
    )

  # Identify rows that represent renewable energy totals
  # (matches things like "total_renewable_energy")
  long <- long %>% mutate(is_renewable = str_detect(energy_source,
"renewable"))

  # Summarize by state: total across all sources, renewable across renewable
rows
  by_state <- long %>%
    group_by(state) %>%
    summarise(
      total_btu = sum(btu, na.rm = TRUE),
      renewable_btu = sum(btu[is_renewable], na.rm = TRUE),
      .groups = "drop"
    ) %>%
    mutate(
      !!paste0("renewable_share_", year_label) := 100 * renewable_btu /
total_btu
    ) %>%
    select(state, starts_with("renewable_share_"))

  # Ensure 2-letter state codes are uppercase for later mapping/joins if
needed
  by_state %>% mutate(state = toupper(state))
}

# Compute renewable share for 2021 and 2023
share_2021 <- share_from_total_wide(energy_2021, "2021")
share_2023 <- share_from_total_wide(energy_2023, "2023")
```

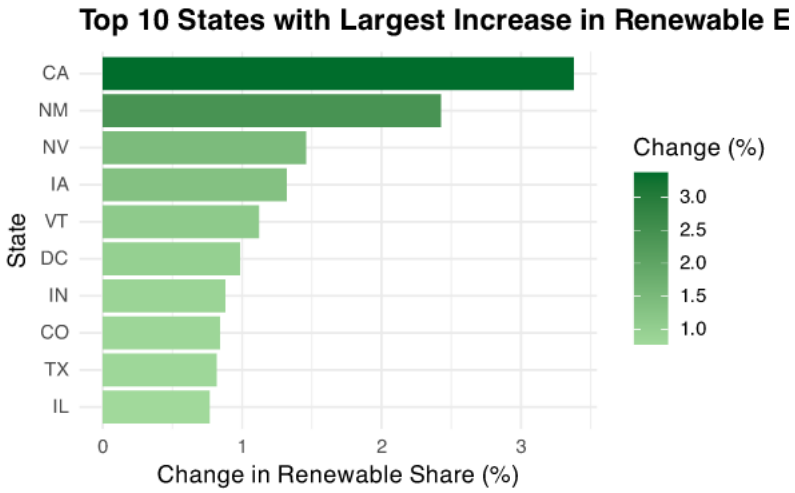
```r
# Join and compute change
renew_change <- share_2021 %>%
  inner_join(share_2023, by = "state") %>%
  mutate(change = renewable_share_2023 - renewable_share_2021) %>%
  arrange(desc(change))

# Top 3 states with largest increase
top3_states <- renew_change %>% slice_max(change, n = 3)
print(top3_states)
```

```
# A tibble: 3 × 4
  state renewable_share_2021 renewable_share_2023 change
  <chr>                <dbl>                <dbl>  <dbl>
1 CA                    13.2                 16.6   3.38
2 NM                     8.36                10.8   2.43
3 NV                     9.21                10.7   1.46
```

Part 4: Mapping Visualization

```r
renew_change %>%
  slice_max(change, n = 10) %>%
  ggplot(aes(x = reorder(state, change), y = change, fill = change)) +
  geom_col() +
  coord_flip() +
  scale_fill_gradient(
    low = "#a1d99b",   #
    high = "#006d2c",
    name = "Change (%)"
  ) +
  labs(
    title = "Top 10 States with Largest Increase in Renewable Energy Share
(2021–2023)",
    x = "State",
    y = "Change in Renewable Share (%)"
  ) +
  theme_minimal(base_size = 12) +
  theme(
    legend.position = "right",
    plot.title = element_text(face = "bold")
  )
```



The graph shows that California, New Mexico, and Nevada experienced the largest increases in renewable energy share between 2021 and 2023, with California leading by a clear margin. These states likely benefited from significant investments in solar and wind energy infrastructure. The steady growth seen in states like Iowa and Vermont suggests continued diversification of energy sources across regions.

Overall, the data indicates that renewable energy generation is expanding most rapidly in the western U.S., highlighting regional disparities in clean energy adoption. This pattern suggests that electric vehicles charged in these states are increasingly powered by renewable sources, strengthening the environmental benefits of EV use.