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   1.5.2
✔ ggplot2   3.5.2     ✔ 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
```

Part 1: Defining Research Question

Chosen Question: Which regions or states show the fastest growth in renewable energy use from 2021 to 2023?

Part 2: Data Preparation and Cleaning

```
# Loading in data files
total_energy_use_2021 <- read_csv("~/Documents/stat133/ev-power-jlpaniangvait/
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.
```

```
total_energy_use_2022 <- read_csv("~/Documents/stat133/ev-power-jlpaniangvait/
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.
```

```
total_energy_use_2023 <- read_csv("~/Documents/stat133/ev-power-jlpaniangvait/
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.
```

```
ev_registrations_by_state <- read_csv("~/Documents/stat133/ev-power-
jlpaniangvait/data/ev-registrations-by-state-2023.csv")
```

```
New names:
Rows: 54 Columns: 2
── Column specification
──────────────────────────────────────────────────────── Delimiter: "," chr
(2): electric vehicle registrations_by_state (2023), ...2
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.
• `` -> `...2`
```

```
renew_use_2021 <- read_csv("~/Documents/stat133/ev-power-jlpaniangvait/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.
```

```
renew_use_2022 <- read_csv("~/Documents/stat133/ev-power-jlpaniangvait/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.
```

```
renew_use_2023 <- read_csv("~/Documents/stat133/ev-power-jlpaniangvait/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.
```

```r
# Standardizing variable names to: Coal, Natural Gas, Petroleum, Nuclear,
Total Renewable Energy
# total_energy_use_2021
total_energy_use_2021$Energy_Source <- sub("Gas.", " Natural Gas",
total_energy_use_2021$Energy_Source)
total_energy_use_2021$Energy_Source <- sub("Petroleum \\(BTU\\)", "Petroleum",
total_energy_use_2021$Energy_Source)
total_energy_use_2021$Energy_Source <- sub("nuclear", "Nuclear",
total_energy_use_2021$Energy_Source)
total_energy_use_2021$Energy_Source <- sub("total_renewable_energy", "Total
Renewable Energy", total_energy_use_2021$Energy_Source)

# total_energy_use_2022
total_energy_use_2022$Energy_Source <- sub("Natural-Gas", "Natural Gas",
total_energy_use_2022$Energy_Source)
total_energy_use_2022$Energy_Source <- sub("petroleum \\(btu\\)", "Petroleum",
total_energy_use_2022$Energy_Source)
total_energy_use_2022$Energy_Source <- sub("Nuclear Energy.", "Nuclear",
total_energy_use_2022$Energy_Source)
total_energy_use_2022$Energy_Source <- sub("total_renewables", "Total
Renewable Energy", total_energy_use_2022$Energy_Source)
total_energy_use_2022$Energy_Source <- sub("coal Consumption", "Coal",
total_energy_use_2022$Energy_Source)

# total_energy_use_2023
total_energy_use_2023$Energy_Source <- sub("NaturalGas", "Natural Gas",
total_energy_use_2023$Energy_Source)
total_energy_use_2023$Energy_Source <- sub("petroleum \\(BTU\\)", "Petroleum",
total_energy_use_2023$Energy_Source)
total_energy_use_2023$Energy_Source <- sub("nuclear-energy +.", "Nuclear",
total_energy_use_2023$Energy_Source)
total_energy_use_2023$Energy_Source <- sub("total renewable-energy", "Total
Renewable Energy", total_energy_use_2023$Energy_Source)
total_energy_use_2023$Energy_Source <- sub("coal_usage", "Coal",
total_energy_use_2023$Energy_Source)

# renew_use_2021
renew_use_2021$Renewable_Use_2021 <-
str_extract(renew_use_2021$Renewable_Use_2021, "[0-9]+")

# renew_use_2022
renew_use_2022$Renewable_Use_2022 <-
str_extract(renew_use_2022$Renewable_Use_2022, "[0-9]+")

# renew_use_2023
renew_use_2023$Renewable_Use_2023 <-
str_extract(renew_use_2023$Renewable_Use_2023, "[0-9]+")
```

Part 3: Joining / Pivoting Datasets for Analysis

```
full_dataset <- inner_join(renew_use_2021, renew_use_2022)
```

```
Joining with `by = join_by(State, Energy_Source)`
```

```
full_dataset <- inner_join(full_dataset, renew_use_2023)
```

```
Joining with `by = join_by(State, Energy_Source)`
```

```
full_dataset <- full_dataset |>
    mutate(Renewable_Use_2021 = as.numeric(Renewable_Use_2021),
    Renewable_Use_2022 = as.numeric(Renewable_Use_2023),
    Renewable_Use_2023 = as.numeric(Renewable_Use_2023))

full_dataset[is.na(full_dataset)] <- 0

head(full_dataset)
```

```
# A tibble: 6 × 5
  State Energy_Source Renewable_Use_2021 Renewable_Use_2022 Renewable_Use_2023
  <chr> <chr>                     <dbl>              <dbl>              <dbl>
1 AK    Biomass                    3153               3404               3404
2 AK    Geothermal                  186                186                186
3 AK    Hydropower                 5763               6051               6051
4 AK    Solar Energy                 45                 67                 67
5 AK    Wind Energy                 451                380                380
6 AL    Biomass                  198543             189040             189040
```

```
year_totals <- full_dataset |>
    group_by(State) |>
    summarize(total_renewable_use_2021 = sum(as.numeric(Renewable_Use_2021)),
    total_renewable_use_2023 = sum(as.numeric(Renewable_Use_2023)))

head(year_totals)
```

```
# A tibble: 6 × 3
  State total_renewable_use_2021 total_renewable_use_2023
  <chr>                    <dbl>                    <dbl>
1 AK                        9598                    10088
2 AL                      239816                   222189
3 CA                      810020                  1065179
```

```
4 CO                      103956                   115062
5 DC                        2487                     2796
6 DE                        7151                     8040
```

```
total_growth <- year_totals|>
    mutate(renewable_use_growth_rate = (total_renewable_use_2023 -
total_renewable_use_2021) / total_renewable_use_2021)

head(total_growth)
```

```
# A tibble: 6 × 4
  State total_renewable_use_2021 total_renewable_use_2023
renewable_use_growth…¹
  <chr>                    <dbl>                    <dbl>
<dbl>
1 AK                        9598                    10088
0.0511
2 AL                      239816                   222189
-0.0735
3 CA                      810020                  1065179
0.315
4 CO                      103956                   115062
0.107
5 DC                        2487                     2796
0.124
6 DE                        7151                     8040
0.124
# ℹ abbreviated name: ¹renewable_use_growth_rate
```

Part 4: Mapping Visualization

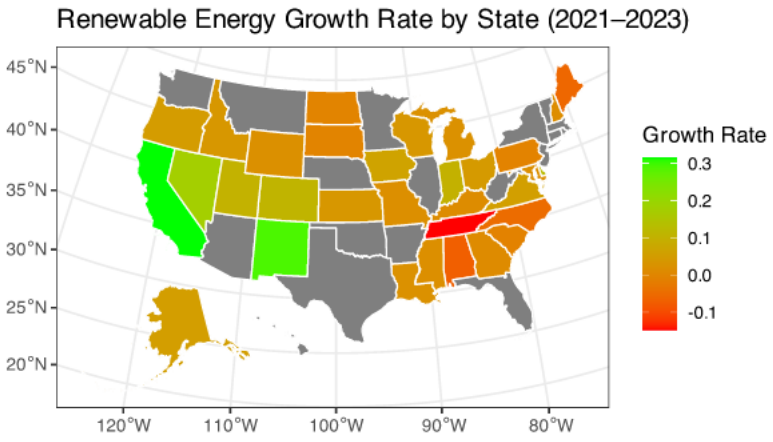
```
library(sf)
```

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

```
library(usmap)

total_growth <- total_growth |> rename(state = State)

plot_usmap(data = total_growth, values = "renewable_use_growth_rate", color =
"white") +
  scale_fill_gradient(low = "red", high = "green", name = "Growth Rate") +
  labs(title = "Renewable Energy Growth Rate by State (2021–2023)") +
  theme_bw()
```



In this map visualization, I only included states that did not have missing data from the years 2021 and 2023. From this map, we can see that California and New Mexico has had the highest growth rate of renewable energy use. On the other hand, there are some states, namely TN, AL, ME, NC, SC, PA, SD, and ND that actually observed a retraction of the amount of renewable energy use from the years 2021 to 2023. From the map, it can also be seen that most of the states with higher growth rate of renewable energy use seems to be located on the West Coast.