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 beco
```

```
library(tidyr)
```

Part 1: Defining Research Question

Chosen Question: Which regions show the fastest growth in renewable energy?

Part 2: Data Preparation and Cleaning

```
data2021 <- read.csv("data/renew-use-2021.csv")
data2022 <- read.csv("data/renew-use-2022.csv")
data2023 <- read.csv("data/renew-use-2023.csv")
```

Part 3: Joining / Pivoting Datasets for Analysis

```
data2021 <- pivot_wider(data2021, names_from = 'Energy_Source', values_from = "Renewable_Use
    mutate(year = '2021')
data2022 <- pivot_wider(data2022, names_from = 'Energy_Source', values_from = "Renewable_Use
    mutate(year = '2022')
data2023 <- pivot_wider(data2023, names_from = 'Energy_Source', values_from = "Renewable_Use
    mutate(year = '2023')
joined_data <- bind_rows(data2021, data2022, data2023)

joined_data <- joined_data |>
    mutate(across(
    c(Biomass, Geothermal, Hydropower, `Solar Energy`, `Wind Energy`, year),
    ~ str_remove_all(.x, "\\D")
  ))

data_long <- joined_data |>
  pivot_longer(
    cols = c(Biomass, Geothermal, Hydropower, `Solar Energy`, `Wind Energy`),
    names_to = "Energy_Source",
    values_to = "Renewable_Use"
  )

energy_growth <- data_long |>
  mutate(Renewable_Use = as.numeric(Renewable_Use)) |>
  group_by(State, Energy_Source, year) |>
  summarise(total_use = sum(Renewable_Use, na.rm = TRUE)) |>
  arrange(State, Energy_Source, year) |>
  group_by(State, Energy_Source) |>
  mutate(growth = c(NA, diff(total_use) / total_use[-length(total_use)] * 100))
```

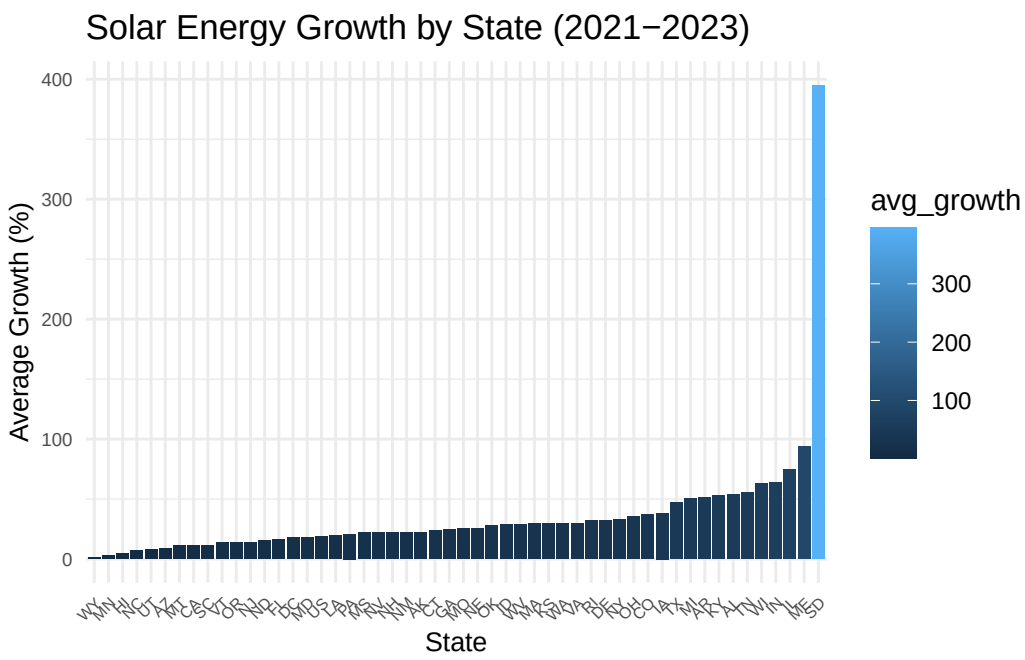
```
`summarise()` has grouped output by 'State', 'Energy_Source'. You can override
using the `.groups` argument.
```

Part 4: Mapping Visualization

```
library(ggplot2)

ggplot(
```

```r
  energy_growth |>
    filter(Energy_Source == "Solar Energy" & !is.na(growth)) |>
    group_by(State) |>
    summarise(avg_growth = mean(growth, na.rm = TRUE)),
  aes(x = reorder(State, avg_growth), y = avg_growth, fill = avg_growth)
) +
  geom_col() +
  labs(
    title = "Solar Energy Growth by State (2021-2023)",
    x = "State",
    y = "Average Growth (%)"
  ) +
  theme_minimal()+
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1, size = 6),  # rotates and sl
    axis.text.y = element_text(size = 7),
    axis.title = element_text(size = 10)
  )
```




```r
#The state with the fastest solar energy growth is South Dakota, followed by Maine and Illino
```