Project 4

```
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   3.5.2     ✔ tibble    3.3.0
✔ 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(dplyr)
library(stringr)
library(maps)
```

```
Attaching package: 'maps'

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

    map
```

```
library(ggplot2)
```

Part 1:Do states with a higher share of renewable energy also have more electric vehicles registered?

```
# datasets
ev_reg_state23      <- read.csv("data/ev-registrations-by-state-2023.csv")
avg_energy_price2123 <- read.csv("data/av-energy-price-2021-2023.csv")

renew_use_21 <- read.csv("data/renew-use-2021.csv")
renew_use_22 <- read.csv("data/renew-use-2022.csv")
renew_use_23 <- read.csv("data/renew-use-2023.csv")

total_use_21 <- read.csv("data/total-use-2021.csv")
total_use_22 <- read.csv("data/total-use-2022.csv")
total_use_23 <- read.csv("data/total-use-2023.csv")

# EV registrations (clean + numeric + lowercase state)
ev_reg_state23 <- ev_reg_state23 %>%
  rename(State = 1) %>%
  mutate(
```

```
    Count_evs = str_extract(X, "\\d+\\.?\\d*") %>% as.double(),
    State = tolower(State)
  ) %>%
  select(State, Count_evs) %>%
  drop_na()

# states
name_matcher <- tibble(
  full_names = c(state.name, "District of Columbia"),
  abbr = c(state.abb, "DC")
) %>%
  mutate(full_names = str_to_lower(full_names))

#totals
renew_totals <- total_use_23 %>%
  rename(Energy_Source = 1) %>%
  filter(str_detect(str_to_lower(Energy_Source), "total\\s*renewable")) %>%
  pivot_longer(-Energy_Source, names_to = "State", values_to = "total_renewable") %>%
  select(-Energy_Source) %>%
  mutate(State = str_to_upper(State)) %>%
  filter(State != "US")

total_energy <- total_use_23 %>%
  rename(Energy_Source = 1) %>%
  pivot_longer(-Energy_Source, names_to = "State", values_to = "value") %>%
  group_by(State) %>%
  summarise(Total_Use_2023 = sum(value, na.rm = TRUE), .groups = "drop") %>%
  mutate(State = str_to_upper(State)) %>%
  filter(State != "US")

#EV by state w abbs
ev_by_state <- ev_reg_state23 %>%
  mutate(State = str_to_lower(State)) %>%
  left_join(name_matcher, by = c("State" = "full_names")) %>%
  transmute(State = str_to_upper(abbr), Count_evs) %>%
  filter(!is.na(State))

#joins + compute ratio
combined <- renew_totals %>%
  inner_join(total_energy, by = "State") %>%
  mutate(renew_share = 100 * total_renewable / Total_Use_2023) %>%
  inner_join(ev_by_state, by = "State")
```

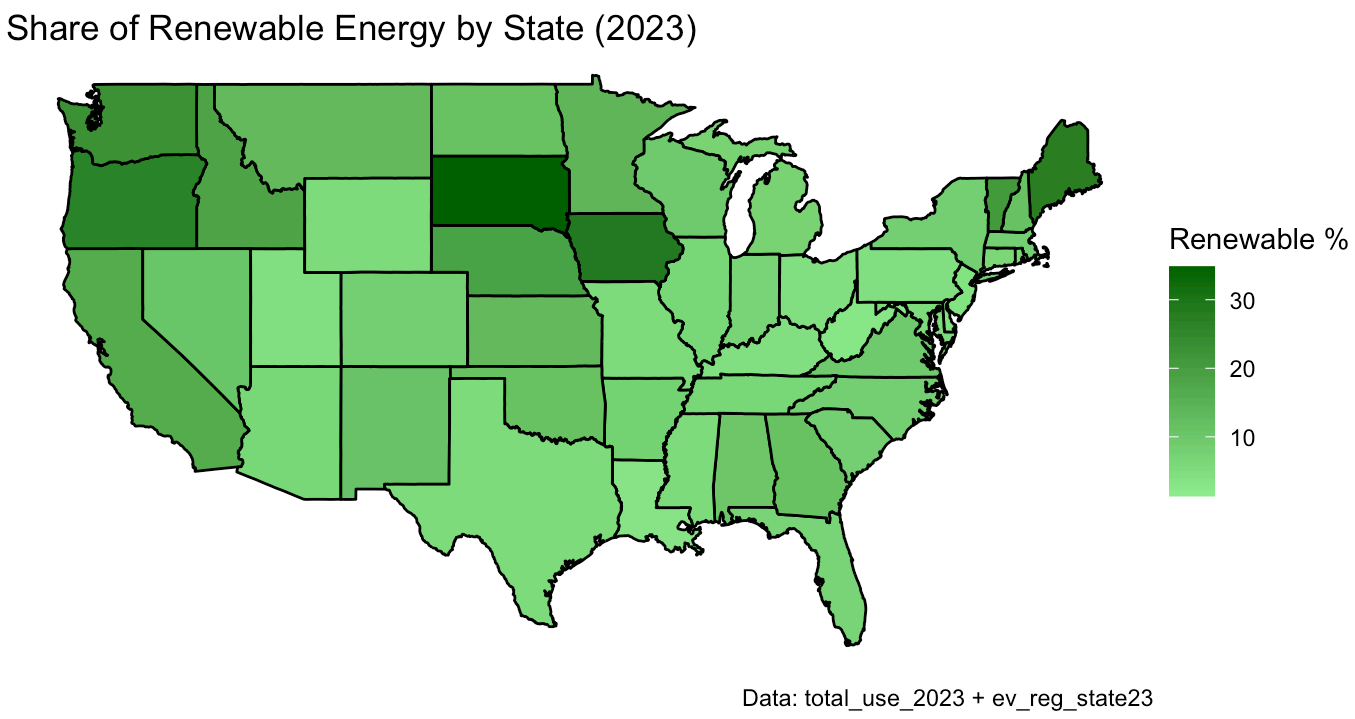
Part 4

```
us_states <- map_data("state")

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

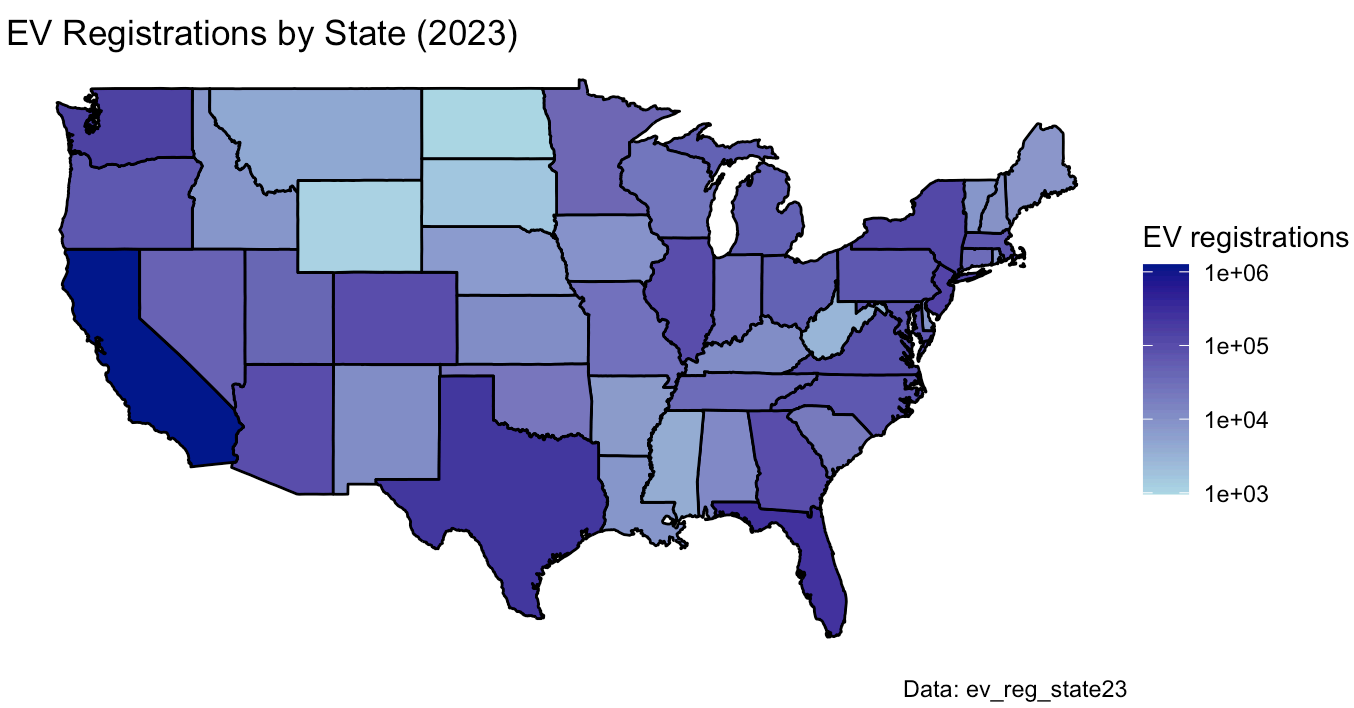
```
map_data_combined <- combined %>%
  left_join(state_ref, by = "State") %>%
  left_join(us_states, by = "region")

ggplot(map_data_combined, aes(long, lat, group = group, fill = renew_share)) +
  geom_polygon(colour = "black") +
  coord_fixed(1.3) +
  scale_fill_gradient(
    low = "lightgreen",
    high = "darkgreen",
    name = "Renewable %"
  ) +
  labs(
    title = "Share of Renewable Energy by State (2023)",
    caption = "Data: total_use_2023 + ev_reg_state23"
  ) +
  theme_void()
```




```
ggplot(map_data_combined, aes(long, lat, group = group, fill = Count_evs)) +
  geom_polygon(colour = "black") +
  coord_fixed(1.3) +
  scale_fill_gradient(
```

```
    low = "lightblue",
    high = "darkblue",
    trans = "log10",
    name = "EV registrations"
  ) +
  labs(
    title = "EV Registrations by State (2023)",
    caption = "Data: ev_reg_state23"
  ) +
  theme_void()
```




```
ggplot(combined, aes(x = renew_share, y = Count_evs, label = State)) +
  geom_point() +
  geom_text(vjust = -0.8, size = 3) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  scale_y_log10() +
  labs(
    title = "EV Count (log scale) vs. Renewable Energy Share by State (2023)",
    x = "Renewable Energy Share (%)",
    y = "Number of EVs (log scale)"
  ) +
  theme_minimal()
```

```
`geom_smooth()` using formula = 'y ~ x'

Warning: The following aesthetics were dropped during statistical transformation: label.
ℹ This can happen when ggplot fails to infer the correct grouping structure in
  the data.
ℹ Did you forget to specify a `group` aesthetic or to convert a numerical
  variable into a factor?
```