EV Power - Lab 4 Project Report

Example Solution 1

Part 0: libraries

```
library(readr)
library(dplyr)
```

```


Attaching package: 'dplyr'
```

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

    filter, lag
```

```
The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
library(tidyr)
library(stringr)
library(tibble)
```

Part 1: Defining Research Question

Chosen Question: Which states in 2023 get a higher share of their total energy from renewable sources?

Part 2: Data Preparation and Cleaning

```
## Part 1: helper to standardize column names

clean_names <- function(df) {
  new_names <- names(df) |>
    str_trim() |>
    str_replace_all("[^A-Za-z0-9]+", "_") |>
    tolower() |>
    str_replace_all("^_|_$", "")
  names(df) <- new_names
  df
```

```r
}

renew_2023 <- read_csv("data/renew-use-2023.csv") |> clean_names()
```

```
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
total_2023 <- read_csv("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
ev_raw     <- read_csv("data/ev-registrations-by-state-2023.csv") |>
clean_names()
```

```
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`
```

```r
renew_2023_clean <-
  renew_2023 |>
```

```r
  mutate(
    state = toupper(state),
    renewable_use_2023_num = as.numeric(
      na_if(
        str_replace_all(
          as.character(renewable_use_2023),
          "[^0-9.]",
          ""
        ),
        ""
      )
    )
  )

renew_by_state_2023 <-
  renew_2023_clean |>
  group_by(state) |>
  summarize(
    renewable_total_2023 = sum(renewable_use_2023_num, na.rm = TRUE),
    .groups = "drop"
  )

total_2023_long <-
  total_2023 |>
  pivot_longer(
    cols = -energy_source,
    names_to = "state",
    values_to = "energy_use_2023"
  ) |>
  mutate(
    state = toupper(state)
  )

total_energy_by_state_2023 <-
  total_2023_long |>
  group_by(state) |>
  summarize(
    total_energy_2023 = sum(energy_use_2023, na.rm = TRUE),
    .groups = "drop"
  )


ev_tidy <-
  ev_raw |>
  slice(-(1:2)) |>
  select(1, 2) |>
  rename(
    state_full = 1,
```

```
    ev_reg_raw = 2
  ) |>
  mutate(
    state_full = str_trim(state_full),
    ev_registrations_2023 = as.numeric(
      na_if(
        str_replace_all(
          as.character(ev_reg_raw),
          "[^0-9]",
          ""
        ),
        ""
      )
    )
  )

state_lookup <- tibble(
  state_full = c(

"Alabama","Alaska","Arizona","Arkansas","California","Colorado","Connecticut",
    "Delaware","District of
Columbia","Florida","Georgia","Hawaii","Idaho","Illinois",

"Indiana","Iowa","Kansas","Kentucky","Louisiana","Maine","Maryland","Massachusetts",

"Michigan","Minnesota","Mississippi","Missouri","Montana","Nebraska","Nevada",
    "New Hampshire","New Jersey","New Mexico","New York","North Carolina",
    "North Dakota","Ohio","Oklahoma","Oregon","Pennsylvania","Rhode Island",
    "South Carolina","South
Dakota","Tennessee","Texas","Utah","Vermont","Virginia",
    "Washington","West Virginia","Wisconsin","Wyoming","Total"
  ),
  state = c(
    "AL","AK","AZ","AR","CA","CO","CT",
    "DE","DC","FL","GA","HI","ID","IL",
    "IN","IA","KS","KY","LA","ME","MD","MA",
    "MI","MN","MS","MO","MT","NE","NV",
    "NH","NJ","NM","NY","NC",
    "ND","OH","OK","OR","PA","RI",
    "SC","SD","TN","TX","UT","VT","VA",
    "WA","WV","WI","WY","US"
  )
)

ev_by_state_2023 <-
  ev_tidy |>
  left_join(state_lookup, by = "state_full") |>
  select(state, ev_registrations_2023)
```

```
state_energy_2023 <-
  renew_by_state_2023 |>
  left_join(total_energy_by_state_2023, by = "state") |>
  mutate(
    renewable_share_2023 = renewable_total_2023 / total_energy_2023
  ) |>
  left_join(ev_by_state_2023, by = "state")

state_energy_2023 |>
  arrange(desc(renewable_share_2023)) |>
  head(10)
```

```
# A tibble: 10 × 5
   state renewable_total_2023 total_energy_2023 renewable_share_2023
   <chr>                <dbl>             <dbl>                <dbl>
 1 SD                  126540            363161                0.348
 2 IA                  414801           1466926                0.283
 3 ME                   89444            328875                0.272
 4 OR                  236063            876891                0.269
 5 WA                  365955           1624957                0.225
 6 VT                   22209            105445                0.211
 7 NE                  164503            872370                0.189
 8 ID                   77127            421975                0.183
 9 CA                 1065179           6429818                0.166
10 MN                  223864           1601319                0.140
# ℹ 1 more variable: ev_registrations_2023 <dbl>
```

Part 3: Joining / Pivoting Datasets for Analysis

```
state_energy_2023 <-
  state_energy_2023 |>
  mutate(
    renewable_share_2023 = renewable_total_2023 / total_energy_2023,
    ev_per_total_energy_2023 = ev_registrations_2023 / total_energy_2023
  )

state_energy_2023 |>
  arrange(desc(renewable_share_2023)) |>
  select(state, renewable_share_2023, ev_registrations_2023,
ev_per_total_energy_2023) |>
  head(10)
```

```
# A tibble: 10 × 4
   state renewable_share_2023 ev_registrations_2023 ev_per_total_energy_2023
   <chr>                <dbl>                 <dbl>                    <dbl>
```

```
 1 SD                   0.348                 1675                 0.00461
 2 IA                   0.283                 9031                 0.00616
 3 ME                   0.272                 7377                 0.0224
 4 OR                   0.269                64361                 0.0734
 5 WA                   0.225               152101                 0.0936
 6 VT                   0.211                 7816                 0.0741
 7 NE                   0.189                 6920                 0.00793
 8 ID                   0.183                 8501                 0.0201
 9 CA                   0.166              1256646                 0.195
10 MN                   0.140                37050                 0.0231
```

Part 4: Mapping Visualization

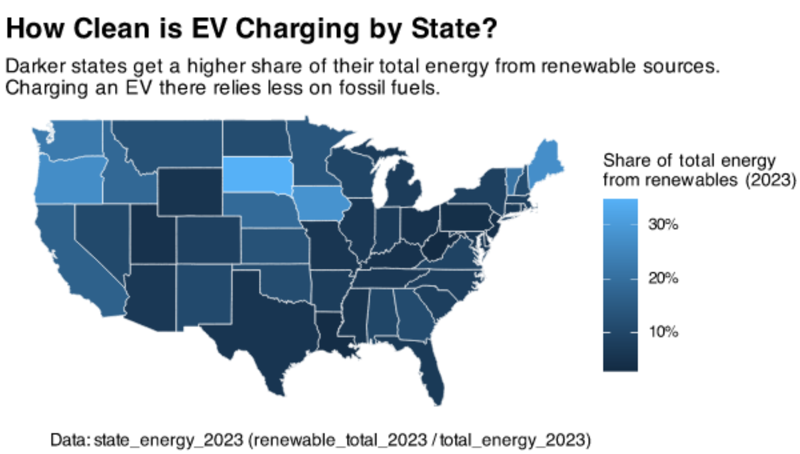
```r
library(maps)
library(ggplot2)
library(scales)
```

```


Attaching package: 'scales'
```

```
The following object is masked from 'package:readr':

    col_factor
```

```r
library(dplyr)
library(tibble)
library(stringr)

us_states_map <- map_data("state")
state_name_lookup <- tibble(
  state = state.abb,
  region = tolower(state.name)
)

plot_data <-
  state_energy_2023 |>
  left_join(state_name_lookup, by = "state") |>
  filter(!is.na(region))

choropleth_data <-
  us_states_map |>
  left_join(plot_data, by = "region")

ggplot(choropleth_data,
       aes(x = long, y = lat, group = group, fill = renewable_share_2023)) +
```

```
  geom_polygon(color = "white", linewidth = 0.2) +
  coord_fixed(1.3) +
  scale_fill_continuous(
    name = "Share of total energy\nfrom renewables (2023)",
    labels = percent_format(accuracy = 1)
  ) +
  labs(
    title = "How Clean is EV Charging by State?",
    subtitle = "Darker states get a higher share of their total energy from renewable sources.\nCharging an EV there relies less on fossil fuels.",
    caption = "Data: state_energy_2023 (renewable_total_2023 / total_energy_2023)"
  ) +
  theme_void() +
  theme(
    legend.position = "right",
    plot.title = element_text(face = "bold", size = 14),
    plot.subtitle = element_text(size = 10),
    legend.title = element_text(size = 9),
    legend.text = element_text(size = 8)
  )
```



Data and Methods

I used three 2023 datasets: renewable energy use by state, total energy use by state, and EV registrations by state. I cleaned the column names, converted messy strings like "3404 kWh" and "#13047" into numbers, and made sure all states used the same two-letter abbreviations. I then summed renewable energy per state and total energy per state, and merged that with EV registrations. From this merged table, I created two new variables. The first is renewable_share_2023, which is renewable energy divided by total energy for each state. This measures how "clean" the electricity is in that state. The second is an EV intensity measure that compares EV registrations to total energy use. I also inspected the joined table (the head() of state_energy_2023) to make sure it looked correct.

Map Visualization

I made a choropleth map of the U.S. where each state is shaded by renewable_share_2023. Darker states get more of their total energy from renewables, which means charging an EV there is more likely to use clean electricity. Lighter states still rely more on fossil fuels to generate electricity. The map includes a title ("How Clean is EV Charging by State?"), a subtitle explaining how to read the colors, and a legend labeled as "Share of total energy from renewables (2023)." This map directly shows where EV charging is closest to truly low-emission.

Analysis

The results are not the same everywhere. States like South Dakota, Iowa, Washington, and Oregon get a very high share of their total energy from renewables. In those states, charging an EV is already relatively clean, because the grid itself is clean. California is especially interesting because it has both a fairly high renewable share and a very large number of EVs on the road, so people there are actually charging lots of EVs on a cleaner grid. In contrast, some states with many EVs, like Texas and Florida, have much lower renewable shares. That means EVs there still improve local air quality (no tailpipe exhaust), but the electricity that charges them still often comes from fossil fuels. There are also states with clean grids but not many EVs yet. Overall, the takeaway is: EVs are not equally "clean" everywhere. Whether charging an EV is truly low-emission depends a lot on which state you're in.