

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    1.0.1    ✓ stringr    1.5.2
✓ ggplot2    4.0.0    ✓ 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
```

```
library(janitor)
```

Attaching package: 'janitor'

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

chisq.test, fisher.test

```
library(readr)
library(stringr)
library(maps)
```

Attaching package: 'maps'

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

```
map
```

```
library(scales)
```

```
Attaching package: 'scales'
```

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

```
discard
```

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

```
col_factor
```

Part 1: Defining Research Question

Chosen Question: Do regional groups (West Coast, East Coast, Midwest, South) differ in renewable energy share and EV registrations in 2023?

Part 2: Data Preparation and Cleaning

```
list.files("data")
```

```
[1] "av-energy-price-2021-2023.csv"      "ev-registrations-by-state-2023.csv"
[3] "renew-use-2021.csv"                "renew-use-2022.csv"
[5] "renew-use-2023.csv"                "total-use-2021.csv"
[7] "total-use-2022.csv"                "total-use-2023.csv"
```

```
#ENERGY: build 2023 totals + renewable share
total23 <- 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.
```

```

#longer format: one row per state_abbreviation, energy_source)
energy_long <- total23 %>%
  pivot_longer(-energy_source, names_to = "state_abb", values_to = "value")
%>% mutate(
  energy_source = str_to_lower(energy_source), is_renew_total =
  str_detect(energy_source, "total\\s*renewable\\s*[- ]?energy"))

#total energy per state
state_totals <- energy_long %>% group_by(state_abb) %>% summarize(total_energy
= sum(value, na.rm = TRUE), .groups = "drop")

#renewable total per state
state_renew <- energy_long %>% filter(is_renew_total) %>% select(state_abb,
renewable_energy = value)

#map abbrev
state_lookup <- tibble(state_name = tolower(state.name), state_abb =
tolower(state.abb)) %>% add_row(state_name = "district of columbia", state_abb
= "dc")

energy23 <- state_totals %>% left_join(state_renew, by = "state_abb") %>%
left_join(state_lookup, by = "state_abb") %>%
  mutate(state = state_name, renew_share = renewable_energy / total_energy)
%>%
  select(state, state_abb, total_energy, renewable_energy, renew_share)

#EV CLEAN
ev23 <- read_csv("data/ev-registrations-by-state-2023.csv") %>% clean_names()
%>%
  rename(state_label = electric_vehicle_registrations_by_state_2023, ev_text =
x2) %>%
  filter(!is.na(state_label), state_label != "STATE") %>%
  mutate(state = str_to_lower(state_label), ev_registrations =
as.numeric(str_replace_all(ev_text, "[^0-9]", ""))) %>%
  select(state, ev_registrations)

```

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`

```
#MERGE AND ADD REGIONS
data23 <- energy23 %>%
  left_join(ev23, by = "state") %>%
  mutate(
    region_group = case_when(
      state %in% c("california","oregon","washington") ~ "West Coast",
      state %in% c("maine","new hampshire","vermont","massachusetts","rhode
island","connecticut",
                  "new york","new
jersey","pennsylvania","delaware","maryland",
                  "district of columbia","virginia","north carolina","south
carolina",
                  "georgia","florida") ~ "East Coast",
      state %in%
c("ohio","michigan","indiana","illinois","wisconsin","minnesota","iowa",
  "missouri","kansas","nebraska","north dakota","south
dakota") ~ "Midwest",
      TRUE ~ "South"
    )
  )

head(select(data23, state, renew_share, ev_registrations, region_group))
```

```
# A tibble: 6 × 4
  state      renew_share ev_registrations region_group
<chr>      <dbl>          <dbl> <chr>
1 alaska      0.0135            2697 South
2 alabama     0.0981           13047 South
3 arkansas    0.0758            7108 South
4 arizona     0.0633           89798 South
5 california  0.166            1256646 West Coast
6 colorado    0.0846           90083 South
```

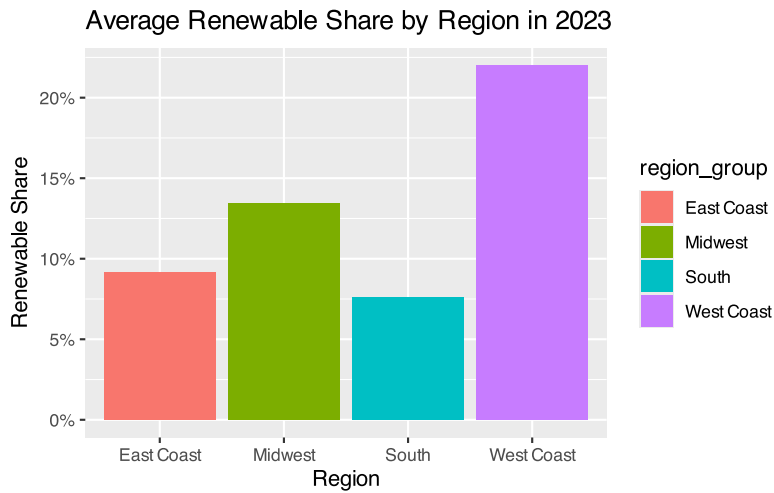
Part 3: Joining / Pivoting Datasets for Analysis

```
# regional summaries
region_summary <- data23 %>% group_by(region_group) %>%
  summarize(mean_renew_share = mean(renew_share, na.rm = TRUE), mean_ev =
mean(ev_registrations, na.rm = TRUE), .groups = "drop")
region_summary
```

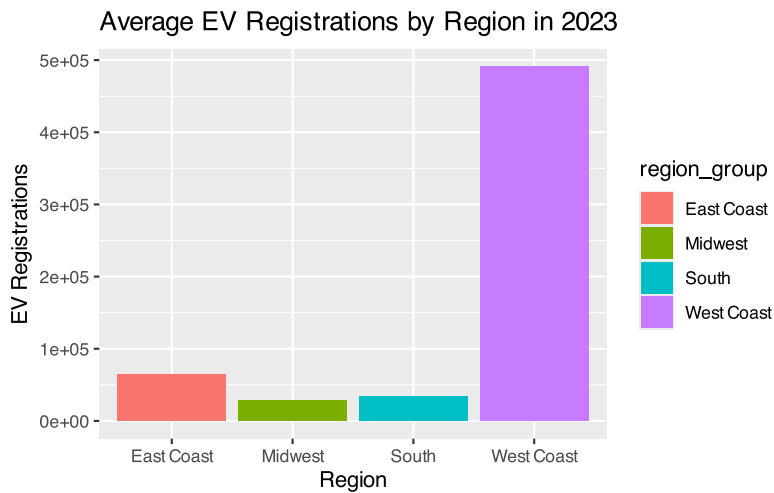
```
# A tibble: 4 × 3
  region_group mean_renew_share mean_ev
<chr>          <dbl>      <dbl>
1 East Coast    0.0916  63811.
2 Midwest       0.135   28755.
```

3	South	0.0762	34341.
4	West Coast	0.220	491036

```
# bar: average renewable share by region
ggplot(region_summary, aes(region_group, mean_renew_share, fill =
region_group)) +
  geom_col() + scale_y_continuous(labels = percent) + labs(title = "Average
Renewable Share by Region in 2023", x = "Region", y = "Renewable Share")
```



```
# bar: average EV registrations by region
ggplot(region_summary, aes(region_group, mean_ev, fill = region_group)) +
  geom_col() + labs(title = "Average EV Registrations by Region in 2023", x =
"Region", y = "EV Registrations")
```



Part 4: Mapping Visualization

```
us_ev <- map_data("state") %>% mutate(state = region)
map_data23 <- us_ev %>%
  left_join(data23, by = "state")

# renewable share colored map
ggplot(map_data23, aes(long, lat, group = group, fill = renew_share)) +
  geom_polygon(color = "white", size = 0.2) + coord_fixed(1.3) +
  scale_fill_continuous(labels = percent) +
  labs(title = "Renewable Share by State (2023)", fill = "Renew.\nShare")
```

Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.

