EV Power - Lab 4 Project Report

Example Solution 1

Part 0: Libraries

```r
# Core wrangling & viz
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
```

```r
library(readr)
library(janitor)
```

```


Attaching package: 'janitor'

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

    chisq.test, fisher.test
```

```r
library(scales)
```

```


Attaching package: 'scales'

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

    discard

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

    col_factor
```

```
library(knitr)   # keep knitr for simple tables; DO NOT load kableExtra
```

Part 1: Defining Research Question

Chosen Question: Do states with cleaner electricity mixes (higher renewable share) also have more EV registrations?

Part 2: Data Preparation and Cleaning

```
# ---------- Helpers (robust) ----------
library(dplyr); library(tidyr); library(readr); library(janitor);
library(stringr)

# Normalize to USPS 2-letter abbreviations (handles full state names and DC)
to_abbr <- function(x) {
  full <- c(state.name, "District Of Columbia")
  abbr <- c(state.abb,  "DC")
  x2   <- str_to_title(x)
  out  <- abbr[match(x2, full)]
  keep <- is.na(out) & nchar(x2) == 2
  out[keep] <- toupper(x2[keep])
  toupper(out)
}

# Pivot a “states-as-columns” file to long and force numeric
pivot_states_wide <- function(df, value_name) {
  tmp <- df |>
    clean_names() |>
    pivot_longer(cols = -energy_source, names_to = "abbr", values_to =
"value") |>
    mutate(
      abbr  = toupper(abbr),
      value = readr::parse_number(as.character(value))
    ) |>
    filter(abbr %in% c(state.abb, "DC"))  # drop US/other aggregates
  names(tmp)[names(tmp) == "value"] <- value_name
  tmp
}

# ---------- Renewable use (2023) — already tidy per your preview ----------
# Columns: State | Energy_Source | Renewable_Use_2023  (260 x 3)
```

```r
renew23 <- read_csv("data/renew-use-2023.csv", show_col_types = FALSE) |>
  clean_names() |>
  mutate(
    abbr = to_abbr(state),
    renewable_use_2023 = readr::parse_number(as.character(renewable_use_2023))
  )

# Prefer the “Total” row per state; otherwise sum across sources for each
state
if ("energy_source" %in% names(renew23) && any(renew23$energy_source ==
"Total", na.rm = TRUE)) {
  renew23 <- renew23 |>
    filter(energy_source == "Total") |>
    select(abbr, renewable_use_2023)
} else {
  renew23 <- renew23 |>
    group_by(abbr) |>
    summarize(renewable_use_2023 = sum(renewable_use_2023, na.rm =
TRUE), .groups = "drop") |>
    select(abbr, renewable_use_2023)
}

# ---------- Total energy use (2023) — wide with state abbreviations as
columns ----------
# Columns: Energy_Source | AK | AL | … | WY | US
total23_raw  <- read_csv("data/total-use-2023.csv", show_col_types = FALSE)
total23_long <- pivot_states_wide(total23_raw, value_name =
"total_energy_2023")

# Prefer the “Total” row per state; otherwise sum across sources for each
state
if (any(total23_long$energy_source == "Total", na.rm = TRUE)) {
  total23 <- total23_long |>
    filter(energy_source == "Total") |>
    select(abbr, total_energy_2023)
} else {
  total23 <- total23_long |>
    group_by(abbr) |>
    summarize(total_energy_2023 = sum(total_energy_2023, na.rm =
TRUE), .groups = "drop") |>
    select(abbr, total_energy_2023)
}

# ---------- EV registrations (2023) — messy two columns ----------
# Treat col1 = state, col2 = registrations; parse numbers from text
ev23_raw <- read_csv("data/ev-registrations-by-state-2023.csv", show_col_types
= FALSE)
```

```
New names:
• `` -> `...2`
```

```
ev23 <- ev23_raw |>
  clean_names() |>
  select(1:2) |>
  set_names(c("state_raw", "ev_registrations_raw")) |>
  mutate(
    abbr             = to_abbr(state_raw),
    ev_registrations = readr::parse_number(as.character(ev_registrations_raw))
  ) |>
  filter(!is.na(abbr)) |>
  select(abbr, ev_registrations)
```

```
Warning: There was 1 warning in `mutate()`.
ℹ In argument: `ev_registrations =
  readr::parse_number(as.character(ev_registrations_raw))`.
Caused by warning:
! 1 parsing failure.
row col expected    actual
  2  -- a number Count-EVs
```

```
# ---------- Build 2023 analysis panel (used later by the map) ----------
panel23 <- total23 |>
  left_join(renew23, by = "abbr") |>
  mutate(
    renewable_share = dplyr::if_else(total_energy_2023 > 0,
                                     renewable_use_2023 / total_energy_2023,
                                     NA_real_)
  ) |>
  left_join(ev23, by = "abbr")

# ---------- Quick sanity checks (Typst-safe tables; no HTML) ----------
panel23 |>
  arrange(desc(renewable_share)) |>
  mutate(renewable_share = scales::percent(renewable_share, accuracy = 0.1)) |
>
  slice_head(n = 10) |>
  knitr::kable(format = "pipe", caption = "Top 10 States by Renewable Share
(2023)")
```

abbr	total_energy_2023	renewable_use_2023	renewable_share	ev_registrations
SD	363161	126540	34.8%	1675
IA	1466926	414801	28.3%	9031

abbr	total_energy_2023	renewable_use_2023	renewable_share	ev_registrations
ME	328875	89444	27.2%	7377
OR	876891	236063	26.9%	64361
WA	1624957	365955	22.5%	152101
VT	105445	22209	21.1%	7816
NE	872370	164503	18.9%	6920
ID	421975	77127	18.3%	8501
CA	6429818	1065179	16.6%	1256646
MN	1601319	223864	14.0%	37050

Part 3: Joining / Pivoting Datasets for Analysis

```r
# Optional 2021–2023 panel. This forces numerics before any summarize().

renew_year <- function(path, yr) {
  df <- read_csv(path, show_col_types = FALSE) |> clean_names()
  val_col <- paste0("renewable_use_", yr)

  # Harmonize IDs and FORCE numeric in the value column
  df <- df |>
    mutate(
      abbr = to_abbr(df$state),
      !!val_col := readr::parse_number(as.character(.data[[val_col]]))
    )

  if ("energy_source" %in% names(df) && any(df$energy_source == "Total", na.rm
= TRUE)) {
    out <- df |>
      filter(energy_source == "Total") |>
      transmute(abbr, year = yr, renewable = .data[[val_col]])
  } else if (val_col %in% names(df)) {
    out <- df |>
      group_by(abbr) |>
      summarize(renewable = sum(.data[[val_col]], na.rm = TRUE), .groups =
"drop") |>
      mutate(year = yr)
  } else {
    stop("Expected column not found: ", val_col)
  }
  out
}

total_year <- function(path, yr) {
  df <- read_csv(path, show_col_types = FALSE)
  long <- pivot_states_wide(df, value_name = "total_value")  # already
numeric-safe
  if (any(long$energy_source == "Total", na.rm = TRUE)) {
    out <- long |>
      filter(energy_source == "Total") |>
      transmute(abbr, year = yr, total = total_value)
  } else {
    out <- long |>
      group_by(abbr) |>
      summarize(total = sum(total_value, na.rm = TRUE), .groups = "drop") |>
      mutate(year = yr)
  }
  out
}

renew_all <- bind_rows(
  renew_year("data/renew-use-2021.csv", 2021),
  renew_year("data/renew-use-2022.csv", 2022),
  renew_year("data/renew-use-2023.csv", 2023)
)

total_all <- bind_rows(
  total_year("data/total-use-2021.csv", 2021),
  total_year("data/total-use-2022.csv", 2022),
  total_year("data/total-use-2023.csv", 2023)
)

panel_all <- renew_all |>
  left_join(total_all, by = c("abbr","year")) |>
  mutate(renewable_share = dplyr::if_else(total > 0, renewable / total,
NA_real_))

# Typst-safe preview (no kableExtra)
panel_all |>
  arrange(abbr, year) |>
  mutate(renewable_share = scales::percent(renewable_share, 0.1)) |>
  slice_head(n = 12) |>
  knitr::kable(format = "pipe", caption = "Joined 2021–2023 Panel (first
rows)")
```

abbr	renewable	year	total	renewable_share
AK	9598	2021	684975	1.4%
AK	10410	2022	730276	1.4%
AK	10088	2023	746979	1.4%

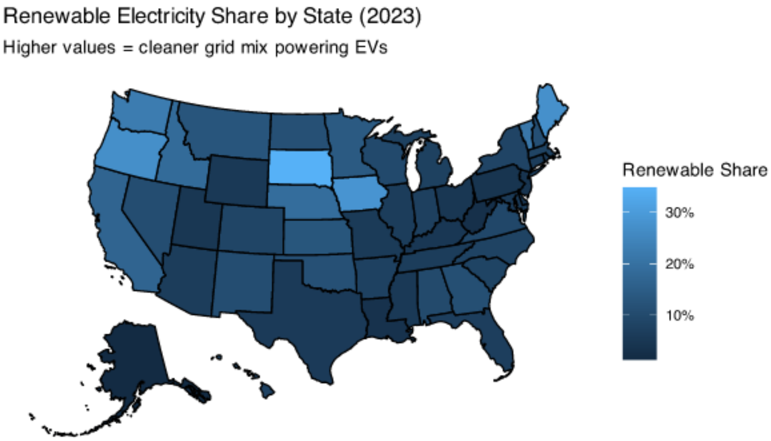
abbr	renewable	year	total	renewable_share
AL	239816	2021	2352656	10.2%
AL	232035	2022	2337513	9.9%
AL	222189	2023	2265008	9.8%
AR	89714	2021	1136025	7.9%
AR	90824	2022	1178115	7.7%
AR	87277	2023	1151062	7.6%
AZ	99266	2021	1681257	5.9%
AZ	101214	2022	1651857	6.1%
AZ	108445	2023	1712667	6.3%

Part 4: Mapping Visualization

```r
# Part 4: Mapping Visualization using usmap
library(usmap)

# Add a full state-name column (required by usmap)
panel23_map <- panel23 |>
  mutate(
    state = state.name[match(abbr, state.abb)],   # convert ABBR → full name
    state = ifelse(abbr == "DC", "District of Columbia", state)
  ) |>
  filter(!is.na(state))   # remove any leftovers where conversion failed

plot_usmap(
  data   = panel23_map,
  values = "renewable_share",
  regions = "states"
) +
  scale_fill_continuous(
    name     = "Renewable Share",
    labels   = percent_format(accuracy = 1),
    na.value = "grey90"
  ) +
  labs(
    title    = "Renewable Electricity Share by State (2023)",
    subtitle = "Higher values = cleaner grid mix powering EVs"
  ) +
  theme(legend.position = "right")
```


Renewable Electricity Share by State (2023)
Higher values = cleaner grid mix powering EVs
Renewable Share
30%
20%
10%