EV Power - Lab 4 Project Report

Esteban Lopez

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 becor
```

```
library(sf)
```

```
Linking to GEOS 3.13.0, GDAL 3.8.5, PROJ 9.5.1; sf_use_s2() is TRUE
```

```
library(janitor)
```

```
Attaching package: 'janitor'

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

    chisq.test, fisher.test
```

```
library(patchwork)
library(tigris)
```

```
To enable caching of data, set `options(tigris_use_cache = TRUE)`
in your R script or .Rprofile.
```

Part 1: Defining Research Question

Chosen Question: Do states with higher renewable electricity shares also have higher EV adoption and how does price relate?

Part 2: Data Preparation and Cleaning

```
renew_files <- list.files("data", pattern = "renew-use-.*\\.csv$", full.names = TRUE)
total_files <- list.files("data", pattern = "total-use-.*\\.csv$", full.names = TRUE)

price <- read_csv("data/av-energy-price-2021-2023.csv")
```

```
Rows: 54 Columns: 1
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (1): Total energy average price, dollars per million Btu,,,

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.
```

```
ev    <- read_csv("data/ev-registrations-by-state-2023.csv")
```

```
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 <- map_dfr(renew_files, read_csv, .id = "file") |>
  mutate(year = str_extract(file, "202\\d"))
```

```
Rows: 260 Columns: 3
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (3): State, Energy_Source, Renewable_Use_2021

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.
Rows: 260 Columns: 3
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (3): State, Energy_Source, Renewable_Use_2022

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.
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 <- map_dfr(total_files, read_csv, .id = "file") |>
  mutate(year = str_extract(file, "202\\d"))
```

```
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.
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.
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.
```

Part 3: Joining / Pivoting Datasets for Analysis

```r
energy <- left_join(
  renew |> select(1, 2, year),
  total |> select(1, 2, year),
  by = c("file" = "file", "year" = "year"),
  suffix = c("_renew", "_total")) |>
  mutate(pct_renew = 100 * as.numeric(State) / as.numeric(Energy_Source))
```

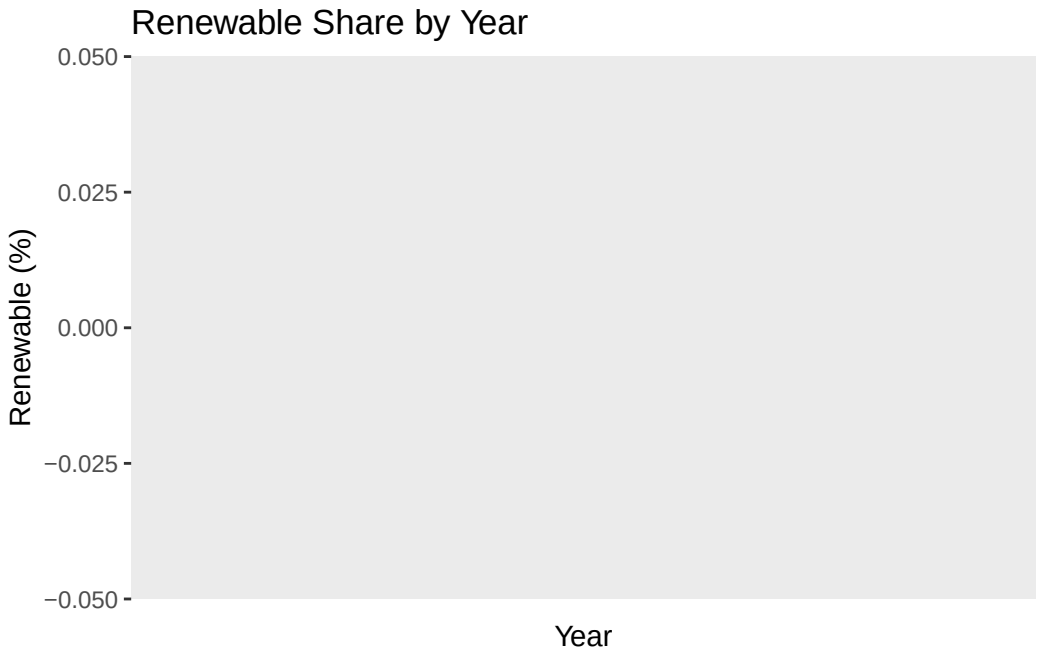
```
Warning in left_join(select(renew, 1, 2, year), select(total, 1, 2, year), : Detected an unex
to-many relationship between `x` and `y`.
i Row 1 of `x` matches multiple rows in `y`.
i Row 1 of `y` matches multiple rows in `x`.
i If a many-to-many relationship is expected, set `relationship =
  "many-to-many"` to silence this warning.

Warning: There were 2 warnings in `mutate()`.
The first warning was:
i In argument: `pct_renew = 100 * as.numeric(State)/as.numeric(Energy_Source)`.
Caused by warning:
! NAs introduced by coercion
i Run `dplyr::last_dplyr_warnings()` to see the 1 remaining warning.
```

Part 4: Mapping Visualization

```r
ggplot(energy, aes(x = year, y = pct_renew)) +
  geom_col(fill = "steelblue") +
  labs(title = "Renewable Share by Year", x = "Year", y = "Renewable (%)")
```

```
Warning: Removed 3900 rows containing missing values or values outside the scale range
(`geom_col()`).
```



Overview: This data is supposed to help me visualize and compare high vs low renewable energy share states so that I can see a direct correlation between energy shares and EV adoption; however, I came across a couple of errors when trying to pivot and join the energy data set. I planned on mutating a new column called pct_renew so that I can visualize the desired data, but I came across various errors when doing so. Despite this, I initially hypothesized that this is not the case since there are other factors to take into account like the price of electricity and state policy.