

EV Power - Lab 4 Project Report

Example Solution 1

Part 0: libraries

```
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(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
v forcats 1.0.0      v readr    2.1.5
```

```
v ggplot2  4.0.0      v stringr  1.5.1
```

```
v lubridate 1.9.4      v tibble   3.3.0
```

```
v purrr     1.1.0      v tidyr    1.3.1
```

```
-- 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 become
```

```
library(sf)
```

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

```
library(ggplot2)
library(stringr)
library(rnaturalearth)
```

Part 1: Defining Research Question

Chosen Question: What is the share of electricity that comes from clean sources by state in 2023?

Part 2: Data Preparation and Cleaning

```
clean <- read_csv("data/renew-use-2023.csv")
```

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.

```
total <- read_csv("data/total-use-2023.csv")
```

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.

```
colnames(clean) <- tolower(colnames(clean))
clean_2023 <- select(clean, -energy_source)

head(clean_2023)
```

```
# A tibble: 6 x 2
  state renewable_use_2023
  <chr> <chr>
1 AK    3404 kWh
2 AK    186.0
3 AK    6051
4 AK     67
5 AK    380
6 AL   189040 kWh
```

Part 3: Joining / Pivoting Datasets for Analysis

```
total_2023 <- total |>
  select(-Energy_Source) |>
  pivot_longer(
    cols = everything(),
    names_to = "state",
    values_to = "total_use_2023"
  )

head(total_2023)
```

```
# A tibble: 6 x 2
  state total_use_2023
  <chr>      <dbl>
1 AK          18414
2 AL        224926
3 AR        180262
4 AZ        137885
5 CA         28746
6 CO        204826
```

```

renew <- clean_2023$renewable_use_2023
clean_2023$renewable_use_2023 <-
  str_replace(renew, "kWh|MWh", "") |>
  as.numeric()

renew_2023 <- clean_2023 |>
  group_by(state) |>
  summarize(renewable_use = sum(renewable_use_2023, na.rm = TRUE))

use_2023 <-
  left_join(renew_2023, total_2023, by = "state")|>
  group_by(state, renewable_use) |>
  summarize(total_use = sum(total_use_2023, na.rm = TRUE))

```

`summarise()` has grouped output by 'state'. You can override using the
 `groups` argument.

```

use_2023 |>
  mutate(renewable_percentage = renewable_use / total_use * 100)

```

```

# A tibble: 52 x 4
# Groups:   state [52]
  state renewable_use total_use renewable_percentage
  <chr>         <dbl>     <dbl>             <dbl>
1 AK             10088     746979             1.35
2 AL            222189    2265008             9.81
3 Ar              87277         0             Inf
4 CA           1065179    6429818            16.6
5 CO            115062    1359507             8.46
6 DC              2796     46323             6.04
7 DE              8040    203487             3.95
8 GA            291462    2627553            11.1
9 IA            414801    1466926            28.3
10 ID             77127     421975            18.3
# i 42 more rows

```

```
head(use_2023)
```

```

# A tibble: 6 x 3
# Groups:   state [6]

```

	state	renewable_use	total_use
	<chr>	<dbl>	<dbl>
1	AK	10088	746979
2	AL	222189	2265008
3	Ar	87277	0
4	CA	1065179	6429818
5	CO	115062	1359507
6	DC	2796	46323

Part 4: Mapping Visualization

```
us_states <- ne_states(country = "united states of america", returnclass = "sf")
us_joined <- us_states |>
  left_join(use_2023, by = join_by(name == state))

# ggplot(us_joined)+
#   geom_sf(aes(fill = renewable_percentage), color = "white")+
#   scale_fill_continuous(name = "Percentage", na.value = "grey90") +
#   labs(title = "The Share of Electricity from Clean Sources by State") +
#   theme_minimal()
```

****Part 5:**

I analyzed the percentage of clean energy in total energy use by state in 2023, but when creating the map, I encountered an issue that I was unable to solve, and the map can't display.