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
──
✔ forcats   1.0.0     ✔ readr     2.1.5
✔ ggplot2   3.5.1     ✔ stringr   1.5.1
✔ lubridate 1.9.4     ✔ tibble    3.2.1
✔ purrr     1.0.2     ✔ tidyr     1.3.1
```

```
── 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(stringr)
totaluse2021 <- read_csv("data/total-use-2021.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.
```

```
totaluse2022 <- read_csv("data/total-use-2022.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.
```

```
totaluse2023 <- 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.
```

```
avgenergyprice <- 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.
```

```
total_energy_usage_2021 = read_csv("data/total-use-2021.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.
```

Part 1: Defining Research Question

Chosen Question: Comparing total energy usage and renewable energy usage, from 2021-2023, have states with increased renewable energy usage seen a drop in overall average energy spending per year?

Part 2: Data Preparation and Cleaning

```
tidy_2021_energy_use <- totaluse2021 |>
    pivot_longer(cols = 2:length(totaluse2021), names_to = "state", values_to =
"values") |>
    pivot_wider(names_from = "Energy_Source", values_from = "values") |>
    mutate(year = 2021)
# ---- replicate for 2022 ------
tidy_2022_energy_use <- totaluse2022 |>
    pivot_longer(cols = 2:length(totaluse2022), names_to = "state", values_to =
"values") |>
    pivot_wider(names_from = "Energy_Source", values_from = "values") |>
    mutate(year = 2022)
# ---- replicate for 2023 ------
tidy_2023_energy_use <- totaluse2023 |>
    pivot_longer(cols = 2:length(totaluse2023), names_to = "state", values_to =
"values") |>
    pivot_wider(names_from = "Energy_Source", values_from = "values") |>
    mutate(year = 2023)
```

Standardize column names

```
standardized_column_names <- c("State", "Coal Consumption", "Natural Gas",
"Petroleum (BTU)", "Nuclear Energy", "Total Renewable Energy", "Year")
```

```
colnames(tidy_2023_energy_use) = standardized_column_names
colnames(tidy_2022_energy_use) = standardized_column_names
colnames(tidy_2021_energy_use) = standardized_column_names
```

```
## **Part 3: Joining / Pivoting Datasets for Analysis**

::: {.cell}

```{.r .cell-code}
tidy_avg_energy_prices <- avgenergyprice  |>
    separate_wider_delim(
    "Total energy average price, dollars per million Btu,,,",
    delim = ",",
    names = c("State", "2021", "2022", "2023")
    ) |>
    filter(!State %in% c("", "State", NA)) |>
    mutate(
        across('2021':'2023', ~parse_number(.x))
    ) |>
    pivot_longer(cols = c("2021", "2022", "2023"), names_to = "Year", values_to
= "Avg_Energy_Price") |>
    mutate(Year = as.double(Year), Avg_Energy_Price = Avg_Energy_Price/1000000)
```

:::

Joining: Combining 2021-2023 to have panel data.

```
joined_energy_use  <-  bind_rows  (tidy_2021_energy_use,  tidy_2022_energy_use,
tidy_2023_energy_use)

single_table  <-  left_join(joined_energy_use,  tidy_avg_energy_prices,  by  =
c("State", "Year")) |>

    mutate(Total_Non_Renewable_Energy_Use = `Coal Consumption`+ `Natural Gas` +
`Petroleum (BTU)` + `Nuclear Energy`) |>
    mutate(Total_Energy_Use = Total_Non_Renewable_Energy_Use + `Total Renewable
Energy`) |>
                        mutate(Proportion_of_Renewable_Energy      =      `Total
Renewable   Energy`/Total_Energy_Use,   Percentage_Energy_Per_Unit_Of_Price   =
Proportion_of_Renewable_Energy/Avg_Energy_Price) |>
                            select(State,      `Total_Non_Renewable_Energy_Use`,
Year,           Avg_Energy_Price,           `Total_Non_Renewable_Energy_Use`,
Proportion_of_Renewable_Energy, Percentage_Energy_Per_Unit_Of_Price)
```
```

```
## **Part 4: Mapping Visualization**

::: {.cell}

```{.r .cell-code}
library(sf)
```
```

```
Linking to GEOS 3.10.2, GDAL 3.4.1, PROJ 8.2.1; sf_use_s2() is TRUE
```

```
library(dplyr)
library(ggplot2)
install.packages(
  "rnaturalearthhires",
  repos = "https://ropensci.r-universe.dev",
  type = "source")
```

```
Installing package into '/srv/r'
(as 'lib' is unspecified)
```

```
library(rnaturalearthhires)
library(maps)
```

```

Attaching package: 'maps'
```

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

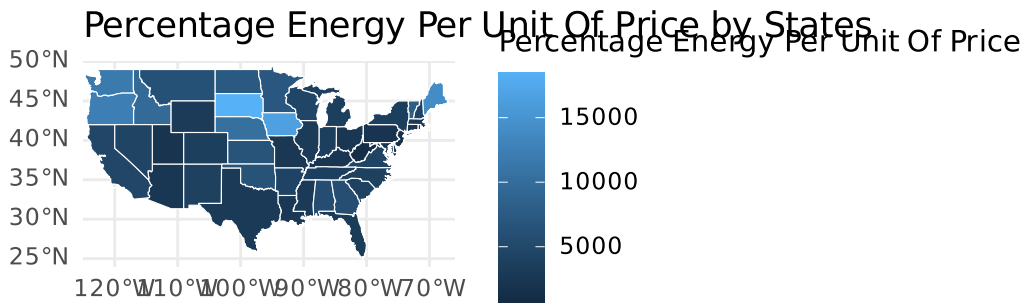
    map
```

```
library(rnaturalearth)

test_2021_table <- single_table |>
    filter(Year == 2021)

us_states <- ne_states(country = "united states of america", returnclass = "sf")
us_joined <- us_states |>
    right_join(test_2021_table, by = join_by("postal"=="State"))
#Plot and Color by Proportion
ggplot(us_joined) +
    geom_sf(aes(fill=Percentage_Energy_Per_Unit_Of_Price), color = "white") +
  scale_fill_continuous(name = "Percentage Energy Per Unit Of Price", na.value
```

```
= "grey90") +
  labs(title = "Percentage Energy Per Unit Of Price by States") +
  coord_sf(xlim = c(-125, -66), ylim = c(24, 50), expand = FALSE) +
  theme_minimal()
```



:::

Analysis: After looking at "Percentage Energy Per Unit of Price by States", the common pattern we see is that the Midwestern states South Dakota and Iowa have the highest renewable energy source usage, with the lowest price spent. The gradient of blue corresponds with a gradient of price values (highest to lowest). This alludes to the conclusion that these states likely have the most renewable energy usage, which is supported by the data in tables totalenergyusage (2021-2023). Further, North and South Dakota are known to have high renewable resource usage–North primarily focusing on wind power and South hydro power.