EV Power - Lab 4 Project Report

Other Formats
PDF

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   1.0.1     ✓ readr     2.1.5
✓ ggplot2   4.0.0     ✓ stringr   1.5.2
✓ lubridate 1.9.4     ✓ tibble    3.3.0
✓ purrr     1.1.0     ✓ 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)
library(ggplot2)
library(sf)
```

```
Linking to GEOS 3.13.1, GDAL 3.11.0, PROJ 9.6.0; sf_use_s2() is TRUE
```

```
library(rnaturalearth)

total_use_21 <- read.csv("data/total-use-2021.csv")
total_use_22 <- read.csv("data/total-use-2022.csv")
total_use_23 <- read.csv("data/total-use-2023.csv")
renew_use_21 <- read.csv("data/renew-use-2021.csv")
renew_use_22 <- read.csv("data/renew-use-2022.csv")
renew_use_23 <- read.csv("data/renew-use-2023.csv")
```

Part 1: Defining Research Question

Chosen Question: Are states with higher renewable energy use also more likely to use nuclear energy?

Part 2: Data Preparation and Cleaning

```
# standardize Energy_Source column for total use tables

total_use_21[1] <- c("Coal", "Natural Gas", "Petroleum", "Nuclear", "Total Renewable Energy")
total_use_22[1] <- c("Coal", "Natural Gas", "Petroleum", "Nuclear", "Total Renewable Energy")
total_use_23[1] <- c("Coal", "Natural Gas", "Petroleum", "Nuclear", "Total Renewable Energy")


# standardize state names for renewable use

renew_use_23$State <- toupper(renew_use_23$State)

# standardize values for renewable use and convert to numeric

renew_use_21 <- renew_use_21 |>
  mutate(Renewable_Use_2021 = str_remove(Renewable_Use_2021, "\\.*$")) |>
  mutate(Renewable_Use_2021 = str_extract_all(Renewable_Use_2021, "\\d+") |>
                    sapply(paste, collapse = "")) |>
  mutate(Renewable_Use_2021 = as.numeric(Renewable_Use_2021))

renew_use_22 <- renew_use_22 |>
  mutate(Renewable_Use_2022 = str_remove(Renewable_Use_2022, "\\.*$")) |>
  mutate(Renewable_Use_2022 = str_extract_all(Renewable_Use_2022, "\\d+") |>
                    sapply(paste, collapse = "")) |>
  mutate(Renewable_Use_2022 = as.numeric(Renewable_Use_2022))

renew_use_23 <- renew_use_23 |>
  mutate(Renewable_Use_2023 = str_remove(Renewable_Use_2023, "\\.*$")) |>
  mutate(Renewable_Use_2023 = str_extract_all(Renewable_Use_2023, "\\d+") |>
                    sapply(paste, collapse = "")) |>
  mutate(Renewable_Use_2023 = as.numeric(Renewable_Use_2023))
```

Part 3: Joining / Pivoting Datasets for Analysis

```
# make table with renewable energy totals for each state in each year

renew_totals <- data.frame(
    State = colnames(total_use_21)[-1],
    totals_21 = as.numeric(slice(total_use_21, 5)[, -1]),
    totals_22 = as.numeric(slice(total_use_22, 5)[, -1]),
    totals_23 = as.numeric(slice(total_use_23, 5)[, -1])
)

# make table that has a row for each state and their nuclear energy usage for each year

nuclear_use <- data.frame(
    State = colnames(total_use_21)[-1],
    totals_21 = as.numeric(slice(total_use_21, 4)[, -1]),
    totals_22 = as.numeric(slice(total_use_22, 4)[, -1]),
    totals_23 = as.numeric(slice(total_use_23, 4)[, -1]))

# add rows indicating total amount of nuclear energy over the three years and whether or not a state uses nuclear

nuclear_use <- nuclear_use |>
    mutate(total_use = rowSums(nuclear_use[c("totals_21", "totals_22", "totals_23")]))|>
    mutate(status = case_when(total_use == 0 ~ FALSE,
                              .default = TRUE))

longer_renew <- renew_totals |>
    pivot_longer(cols = c("totals_21", "totals_22", "totals_23"), names_to = "year", values_to = "usage") |>
    mutate(year = case_when(year == "totals_21" ~ 2021, year == "totals_22" ~ 2022, year == "totals_23" ~ 2023))

longer_nuclear <- pivot_longer(nuclear_use, cols = c("totals_21", "totals_22", "totals_23"), names_to = "year", v

# join renew_totals table with nuclear use (T/F) table

renew_nuclear <- left_join(renew_totals, nuclear_use, by = "State")

final <- mutate(longer_renew, nuclear = longer_nuclear$status)
```

Part 4: Mapping Visualization

```
#nuclear_viz <- final |>
  #ggplot(final, aes(x = year, y = usage, color = nuclear, group = State)) +
  #geom_line(size = 1.1) + scale_x_continuous(breaks = c(2021, 2022, 2023)) +
  #theme_minimal() +
  #scale_y_log10() +
  #labs( title = "Energy Use by State (2021-2023)", x = "Year", y = "Usage (GWh)", color = "Uses Nuclear Energy

#map_1 <- final |>
  # ggplot(aes(x = year, y = usage, color = nuclear, group = State)) +
  # geom_sf(aes(fill = nuclear))

#nuclear_state <- final |>
  #group_by(State) |>
  #summarize(nuclear = any(nuclear)) |>
  #ggplot(final) +
  #geom_sf(aes(fill = nuclear), color = "white") +
  #scale_fill_manual(
    # values = c("TRUE" = "#1b9e77", "FALSE" = "#d95f02"),
    #labels = c("FALSE" = "No Nuclear", "TRUE" = "Has Nuclear")
  #)
```

Part 5: Final Deliverable

Title: Nuclear Energy and Renewable Energy

Overview: Are states with higher renewable energy use also more likely to use nuclear energy?

Data and Methods:

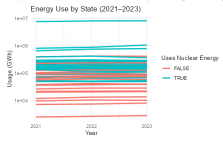
```
head(final)
```

```
# A tibble: 6 × 4
  State  year  usage nuclear
  <chr> <dbl>  <dbl> <lgl>
1 AK     2021   9597 FALSE
2 AK     2022  10410 FALSE
3 AK     2023  10087 FALSE
4 AL     2021 239817 TRUE
5 AL     2022 212035 TRUE
6 AL     2023 222189 TRUE
```

```
head(renew_nuclear)
```

```
  State totals_21.x totals_22.x totals_23.x totals_21.y totals_22.y totals_23.y
1    AK        9597       10410       10087           0           0           0
2    AL      239817      212035      222189      480115      442093      474392
3    AR       89714       90825       87277      141372      149654      154692
4    AZ       99266      101215      108465      329868      333738      329474
5    CA      810020      880995     1065179     1718842     1838814     1859192
6    CO      103955      114917      115061           0           0           0
  total_use status
1         0  FALSE
2   1308480   TRUE
3    447518   TRUE
4    993088   TRUE
5   5488048   TRUE
6         0  FALSE
```

Map Visualization: Include the maps from your analysis



Analysis: Interpret your findings. What patterns do you notice? How does your map help answer the min research question?

I see that there is a correlation between states that have higher total amounts of renewable energy use and states that use nuclear energy.