EV Power - Lab 4 Project Report

Sacha Bondeville

EV analysis in the US

Part 0: libraries

```
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
```

```
library(maps)
```

```

Attaching package: 'maps'

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

    map
```

```
library(leaflet)
library(sf)
```

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

```
library(rnaturalearth)
library(RColorBrewer)
```

Part 1: Defining Research Question

Chosen Question: *Which states actually improved their clean energy portfolio in ht elast three years and is there a correlation between their clean energy improvement and the number of evs registration ?*

Part 2: Data Preparation and Cleaning

```
avg_energy_price2123 <- read.csv("/Users/sachabondeville/Documents/stat133/
Projects/Project-4/data/av-energy-price-2021-2023.csv")
ev_reg_state23 <- read.csv("/Users/sachabondeville/Documents/stat133/Projects/
Project-4/data/ev-registrations-by-state-2023.csv")
renew_use_21 <- read.csv("/Users/sachabondeville/Documents/stat133/Projects/
Project-4/data/renew-use-2021.csv")
renew_use_22 <-  read.csv("/Users/sachabondeville/Documents/stat133/Projects/
Project-4/data/renew-use-2022.csv")
renew_use_23 <- read.csv("/Users/sachabondeville/Documents/stat133/Projects/
Project-4/data/renew-use-2023.csv")
total_use_21 <-  read.csv("/Users/sachabondeville/Documents/stat133/Projects/
Project-4/data/total-use-2021.csv")
total_use_22 <-  read.csv("/Users/sachabondeville/Documents/stat133/Projects/
Project-4/data/total-use-2022.csv")
total_use_23 <- read.csv("/Users/sachabondeville/Documents/stat133/Projects/
Project-4/data/total-use-2023.csv")
```

We have to clean data sets using string recognition. For extra challenge I choose to clean everything.

First for the average energy price :

```
avg_energy_price2123 <- avg_energy_price2123|>
    rename(strings = Total.energy.average.price..dollars.per.million.Btu...)

splitted_string <- str_split(avg_energy_price2123$strings, pattern = ",",
simplify = TRUE)

avg_energy_price2123 <-  avg_energy_price2123|>
    mutate(state = splitted_string[,1],
    "2021" = splitted_string[,2],
    "2022" = splitted_string[,3],
    "2023" = splitted_string[,4]
    )

avg_energy_price2123 <- avg_energy_price2123 |>
    mutate(`2021` = as.double(str_extract_all(`2021`, pattern = "\\d+\\.\
\d+")),
```

```
    `2022` = as.double(str_extract_all(`2022`, pattern = "\\d+\\.\\d+")),
    `2023` = as.double(str_extract_all(`2023`, pattern = "\\d+\\.\\d+")))|>
    drop_na() |> # there are no missing values except for first rows.
    select(colnames(avg_energy_price2123)[-1])
```

For the renew uses data set, we define a function that will operate on the three datasets.

```
cleaning_renew <- function(df) {
    name <- colnames(df)[3]
    state <- colnames(df)[1]
    df <- df |>
    mutate(!!name := as.double(str_extract(df[[name]], pattern = "\\d+\\.?\
\d*"))) |>
    mutate(!!state := str_to_upper(df[[state]]))
    return(df)
}

renew_use_21 <- cleaning_renew(renew_use_21)
renew_use_22 <- cleaning_renew(renew_use_22)
renew_use_23 <- cleaning_renew(renew_use_23)
```

We also create a cleaning function for the total datasets we will use later on.

```
cleaning_total <- function(df) {
    df <- df |>
    mutate(Energy_Source = (str_to_upper(str_extract(df$Energy_Source, pattern
= "([Cc]oal|[Gg]as|[Pp]etroleum|[Nn]uclear|[Rr]enewable)"))))
    return(df)
}
```

For the ev_reg_state :

```
ev_reg_state23 <- ev_reg_state23 |>
    mutate(Count_evs = as.double(str_extract(ev_reg_state23$X, pattern = "\\d+
\\.?\\d*")))|>
    rename("State" = electric.vehicle.registrations_by_state..2023.)|>
    mutate(State = tolower(State))|>
    select(State, Count_evs)|>
    drop_na()

#we create a reformatting data frame that will rename each state from the
correct abbreviation !
state_names <- append(state.name, "District of Columbia")
state_abb <- append(state.abb, "DC")
state_names <-  append(state_names, "Total")
```

```
state_abb <- append(state_abb, "US")
name_matcher <- data.frame("full_names" = state_names, "abbr" = state_abb)
```

Part 3: Joining / Pivoting Datasets for Analysis

```
renew_use2123 <- full_join(renew_use_21, renew_use_22, by = join_by("State" ==
"State", "Energy_Source" == "Energy_Source"))

renew_use2123 <- full_join(renew_use2123, renew_use_23, by = join_by("State"
== "State", "Energy_Source" == "Energy_Source"))

total_use_21 <- pivot_longer(total_use_21, cols = colnames(total_use_21)[-1],
names_to = "State", values_to = "total_21")

total_use_22 <-  pivot_longer(total_use_22, cols = colnames(total_use_22)[-1],
names_to = "State", values_to = "total_22")

total_use_23 <- pivot_longer(total_use_23, cols = colnames(total_use_23)[-1],
names_to = "State", values_to = "total_23")

total_use_21 <- cleaning_total(total_use_21)
total_use_22 <- cleaning_total(total_use_22)
total_use_23 <- cleaning_total(total_use_23)

#and joining the total_use :

total_use2123 <-  full_join(total_use_21, total_use_22, by = join_by("State"
== "State", "Energy_Source" == "Energy_Source"))
total_use2123 <- full_join(total_use2123, total_use_23, by = join_by("State"
== "State", "Energy_Source" == "Energy_Source"))
```

The next join that makes a lot of sens to execute : we want to expand in total_use2123 the "renewable" with all of the sub categories displayed in renew_use2123.

We therefore need to join by state.

```
renew_use2123 <- renew_use2123|>
rename("total_21"= Renewable_Use_2021,
       "total_22"= Renewable_Use_2022,
       "total_23" = Renewable_Use_2023)|>
mutate(Energy_Source = str_to_upper(Energy_Source))

total_use2123 <- total_use2123 |>
bind_rows(renew_use2123)

#we need to extend the states names
```

```
total_use2123 <- left_join(total_use2123, name_matcher, by = join_by("State"
== "abbr"))

total_use2123 <- total_use2123 |>
mutate(State = tolower(full_names))

head(total_use2123, 5)
```

```
# A tibble: 5 × 6
  Energy_Source State      total_21 total_22 total_23 full_names
  <chr>         <chr>         <dbl>    <dbl>    <dbl> <chr>
1 COAL          alaska        18694    18615    18414 Alaska
2 COAL          alabama      309791   297654   224926 Alabama
3 COAL          arkansas     216123   211724   180262 Arkansas
4 COAL          arizona      160299   154007   137885 Arizona
5 COAL          california    28244    30049    28746 California
```

Now that we have a cool dataset we can conduct an analysis :

I want to summarize for each state the change from non renewable to renewable energy.

For our analysis, we take for each state the absolute change in energy output between 2021 an 2023. And we compute the ratio of the clean energy change over the absolute change.

That allows us : - To display negative clean energy change - To reveal the role clean energy plays relative to the part non clean energy plays.

```
Change_in_renew <- total_use2123 |>
mutate(clean_energy = !Energy_Source %in% c("COAL", "GAS", "PETROLEUM"))|>
#I'm French so I consider Nuclear as a clean energy
mutate(change = total_23 - total_21)|>
filter(Energy_Source != "RENEWABLE")|> #We get rid of the renewable (otherwise
double count !)
group_by(State, clean_energy)|>
mutate(energytypechange = sum(change))|>
ungroup()|>
group_by(State)|>
mutate(propenergytypechange = energytypechange/sum(abs(change)))

# I want to create a categorical variable to make the map more readable.
Change_in_renew <- Change_in_renew|>
group_by(State, clean_energy)|>
summarize(prop_of_clean_in_change = mean(propenergytypechange))|>
filter(clean_energy == TRUE)|>
mutate(prop_of_clean_in_change = case_when(prop_of_clean_in_change < 0 ~
"<0%",
                                           prop_of_clean_in_change > 0 &
```

```
prop_of_clean_in_change < 0.2 ~"<20%",
                                          prop_of_clean_in_change > 0.2 &
prop_of_clean_in_change < 0.4 ~"<40%",
                                          prop_of_clean_in_change > 0.4 &
prop_of_clean_in_change < 0.5 ~"<50%",
                                          prop_of_clean_in_change > 0.6
~">50%"))
```

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

```
Change_in_renew <- Change_in_renew |>
left_join(ev_reg_state23, by = join_by("State" == "State"))

write.csv(Change_in_renew, "/Users/sachabondeville/Documents/stat133/Projects/
Project-4/data/change_in_renew.csv")
```

```
us_states <- ne_states(country = "united states of america", returnclass =
"sf")|>
mutate(name = tolower(name))

us_map_joined <- us_states |>
    left_join(Change_in_renew, by = c("name" = "State"))
```

Part 4: Mapping Visualization

```
us_map_joined$prop_of_clean_in_change <- factor(
  us_map_joined$prop_of_clean_in_change,
  levels = c("<0%", "<20%", "<40%","<50%", ">50%")
)

pal <- colorFactor(
palette = c(
  "#d73027",
  "#fc8d59",
  "#fee08b",
  "#d9ef8b",
  "#1a9850"
),
domain = us_map_joined$prop_of_clean_in_change
)

leaflet(us_map_joined) |>
addTiles() |>
addMarkers(lng = -122.259, lat = 37.872, popup = "UC Berkeley") |>
```

```
addPolygons(
fillColor = ~pal(prop_of_clean_in_change),
color = "white",
weight = 1,
fillOpacity = 0.8,
label = ~paste0(toupper(name),"'s # of EV registrations: ", Count_evs))|>
addLegend(
pal = pal,
values = ~prop_of_clean_in_change,
title = "%change in new clean energy (last 3 years)",
position = "bottomright"
)
```

Part 5: Analysis

The map first shows that a lot of states actually decreased their clean energy supply in their electrical output. These states however can still have a lot of EV registration (like Florida) which can reveal two things : - customers lack information and just believe an electric car is greener (while not necessarily) - customers do not particularly care and just like to have an electric car (for the status, for the speed)

Two states increased considerably the part of clean energy in their portfolio : California and South Carolina. It's a good thing, as California is the first state in terms of EV registrations in 2023.

Let's also remember California's weather allow for an easy use of solar energy. It's harder for other states, but our analysis is pretty generous with the definition of clean energy (we include nuclear) so there is definitely room for improvement.