EV Power - Lab 4 Project Report - Finding which states are trending towards a renewable energy future

Example Solution 1

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

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

```
library(maps)
```

```

Attaching package: 'maps'

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

    map
```

Part 1: Defining Research Question

Some questions I brainstormed include: 1.) Which states increased / decreased their usage of renewable energy between 2021 and 2023? 2.) Is there a correlation between states with a higher average energy price having a greater share of their energy coming from renewable sources? 3.) Which energy source most frequently was the most used in a state, and is there a correlation between which energy source was used most in a state and their EV registration count? 4.) Is there a correlation between EV Registration Count and Average Energy Price?

Chosen Question: Wanted to do 2, but ended up finding no meaningful relationship, so looking at qiestion 1 for the map. Looking to see which states are on the rise in increasing their reusable energy, and which are decreasing and moving towards other options.

Part 2: Data Preparation and Cleaning

```
av_price_raw<-read_csv("data/av-energy-price-2021-2023.csv",col_names=FALSE)|>
separate(X1,into=c("state","2021_av","2022_av","2023_av"),sep=",")
```

```
Rows: 55 Columns: 1
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (1): X1

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.
```

```
Warning: Expected 4 pieces. Additional pieces discarded in 1 rows [1].
```

```
av_price_clean<-av_price_raw|>

mutate(across(.cols=c("2021_av","2022_av","2023_av"),.fns=~as.numeric(str_extract(.x,"[-.0-9]+"))
>
    mutate(state=str_to_lower(state))|>
    slice(-(1:3))|>
    slice(-52)

head(av_price_clean)
```

```
# A tibble: 6 × 4
  state `2021_av` `2022_av` `2023_av`
  <chr>     <dbl>     <dbl>     <dbl>
1 ak         20.0      27.3      23.8
2 al         17.8      23.4      21.1
```

```
3 ar         18.4      23.8      21.8
4 az         25.1      31.7      30.3
5 ca         28.4      37.4      35.7
6 co         20.6      25.8      23.8
```

```
state_name_to_abb_map <- setNames(state.abb, state.name) #Found, used to
streamline changing full state names to their abbreviated form through
matching.

ev_count_raw<-read_csv("data/ev-registrations-by-
state-2023.csv",col_names=FALSE)
```

```
Rows: 55 Columns: 2
── Column specification
────────────────────────────────────────────────────────
Delimiter: ","
chr (2): X1, X2

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_count_clean<-ev_count_raw|>
slice(-(1:3))|>
mutate(state_name = str_to_lower(state_name_to_abb_map[X1])) |>
       mutate(ev_count=as.numeric(str_extract(X2,"[-.0-9]+"))) |>
       select(-X1)|>
       select(-X2)|>
       slice(-52)

head(ev_count_clean)
```

```
# A tibble: 6 × 2
  state_name ev_count
  <chr>         <dbl>
1 al            13047
2 ak             2697
3 az            89798
4 ar             7108
5 ca          1256646
6 co            90083
```

```
total_use_2021_raw <- 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.
```

```
total_use_2021_clean<-total_use_2021_raw|>
    rename_with(str_to_lower)|>
    pivot_longer(cols=-energy_source,names_to="state",values_to =
"total_use_2021")|>
    arrange(state)|>
    filter(state != "us")|>
     mutate(energy_source = case_when(
    str_detect(energy_source, "coal|Coal") ~ "coal",
    str_detect(energy_source, "natural_gas|natural gas|Natural Gas|
NaturalGas") ~ "natural_gas",
    str_detect(energy_source, "petroleum|Petro") ~ "petroleum",
    str_detect(energy_source, "nuclear|Nuclear") ~ "nuclear",
    str_detect(energy_source, "renewable") ~ "renewables"))

head(total_use_2021_clean)
```

```
# A tibble: 6 × 3
  energy_source state total_use_2021
  <chr>         <chr>          <dbl>
1 coal          ak             18694
2 natural_gas   ak            395590
3 petroleum     ak            261094
4 nuclear       ak                 0
5 renewables    ak              9597
6 coal          al            309791
```

```
total_use_2022_raw <- 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.
```

```
total_use_2022_clean<-total_use_2022_raw|>
    rename_with(str_to_lower)|>
    pivot_longer(cols=-energy_source,names_to="state",values_to =
"total_use_2022")|>
    arrange(state)|>
    filter(state != "us")|>
     mutate(energy_source = case_when(
    str_detect(energy_source, "coal| Coal") ~ "coal",
    str_detect(energy_source, "natural_gas|natural gas|Natural Gas|NaturalGas|
Natural-Gas") ~ "natural_gas",
    str_detect(energy_source, "petroleum|Petro") ~ "petroleum",
    str_detect(energy_source, "nuclear|Nuclear") ~ "nuclear",
    str_detect(energy_source, "renewable") ~ "renewables"))

head(total_use_2022_clean)
```

```
# A tibble: 6 × 3
  energy_source state total_use_2022
  <chr>         <chr>          <dbl>
1 coal          ak             18615
2 natural_gas   ak            437916
3 petroleum     ak            263335
4 nuclear       ak                 0
5 renewables    ak             10410
6 coal          al            297654
```

```
total_use_2023_raw <- 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.
```

```
total_use_2023_clean<-total_use_2023_raw|>
    rename_with(str_to_lower)|>
    pivot_longer(cols=-energy_source,names_to="state",values_to =
"total_use_2023")|>
    arrange(state)|>
    filter(state != "us")|>
     mutate(energy_source = case_when(
    str_detect(energy_source, "coal|Coal") ~ "coal",
    str_detect(energy_source, "natural_gas|natural gas|Natural Gas|
NaturalGas") ~ "natural_gas",
    str_detect(energy_source, "petroleum|Petro") ~ "petroleum",
    str_detect(energy_source, "nuclear|Nuclear") ~ "nuclear",
    str_detect(energy_source, "renewable") ~ "renewables"))

head(total_use_2023_clean)
```

```
# A tibble: 6 × 3
  energy_source state total_use_2023
  <chr>         <chr>          <dbl>
1 coal          ak             18414
2 natural_gas   ak            448087
3 petroleum     ak            270391
4 nuclear       ak                 0
5 renewables    ak             10087
6 coal          al            224926
```

```
renew_use_2021_raw <- read_csv("data/renew-use-2021.csv")
```

```
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.
```

```
renew_use_2021_clean <- renew_use_2021_raw |>
  rename_with(str_to_lower)|>
```

```
  mutate(state = str_to_lower(state),
renewable_use_2021 = as.numeric(str_extract(renewable_use_2021, "[-.0-9]+")))|
>
    slice(-(256:260))

head(renew_use_2021_clean)
```

```
# A tibble: 6 × 3
  state energy_source renewable_use_2021
  <chr> <chr>                     <dbl>
1 ak    Biomass                    3153
2 ak    Geothermal                  186
3 ak    Hydropower                 5763
4 ak    Solar Energy                 45
5 ak    Wind Energy                 451
6 al    Biomass                  198543
```

```
renew_use_2022_raw <- read_csv("data/renew-use-2022.csv")
```

```
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.
```

```
renew_use_2022_clean <- renew_use_2022_raw |>
  rename_with(str_to_lower)|>
  mutate(state = str_to_lower(state),
renewable_use_2021 = as.numeric(str_extract(renewable_use_2022, "[-.0-9]+")))|
>
    slice(-(256:260))

head(renew_use_2022_clean)
```

```
# A tibble: 6 × 4
  state energy_source renewable_use_2022 renewable_use_2021
  <chr> <chr>         <chr>                           <dbl>
1 ak    Biomass       ≈3846                            3846
2 ak    Geothermal    $186                              186
3 ak    Hydropower    $5846                            5846
```

```
4 ak    Solar Energy  ~57                                57
5 ak    Wind Energy   $475                              475
6 al    Biomass       193932 USD                     193932
```

```
renew_use_2023_raw <- 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.
```

```
renew_use_2023_clean <- renew_use_2023_raw |>
  rename_with(str_to_lower)|>
  mutate(state = str_to_lower(state),
renewable_use_2023 = as.numeric(str_extract(renewable_use_2023, "[-.0-9]+")))|
>
    slice(-(256:260))

head(renew_use_2023_clean)
```

```
# A tibble: 6 × 3
  state energy_source renewable_use_2023
  <chr> <chr>                      <dbl>
1 ak    Biomass                     3404
2 ak    Geothermal                   186
3 ak    Hydropower                  6051
4 ak    Solar Energy                  67
5 ak    Wind Energy                  380
6 al    Biomass                   189040
```

**Part 3:Pivots / Joining

```
# Research Question 1: percentage increase or decrease in renewable energy
used between 2021 and 2023.
total_use_2021_2022<-
full_join(total_use_2021_clean,total_use_2022_clean,by=c("state","energy_source"))
total_use_all_renewable<-
full_join(total_use_2021_2022,total_use_2023_clean,by=c("state","energy_source"))|
>
```

```
    mutate(percentage_change=(total_use_2023-total_use_2021)/
total_use_2021*100)|>
    filter(energy_source=="renewables")

head(total_use_all_renewable)
```

```
# A tibble: 6 × 6
  energy_source state total_use_2021 total_use_2022 total_use_2023
  <chr>         <chr>          <dbl>          <dbl>          <dbl>
1 renewables    ak              9597          10410          10087
2 renewables    al            239817         232035         222189
3 renewables    ar             89714          90825          87277
4 renewables    az             99266         101215         108445
5 renewables    ca            810020         880995        1065179
6 renewables    co            103955         114917         115061
# ℹ 1 more variable: percentage_change <dbl>
```

```
#Research Question 2:Is there a correlation between states with a higher
average energy price having a greater share of their energy coming from
renewable sources? (Using 2023 data opposed to averages from across all 3
years)
renewable_share_2023<-total_use_2023_clean|>
    group_by(state)|>
    mutate(total_use_by_state=sum(total_use_2023,na.rm=TRUE))|>
    ungroup()|>
    filter(energy_source=="renewables")|>
    mutate(renewable_share=total_use_2023/total_use_by_state*100)|>
    select(state,renewable_share)

head(renewable_share_2023)
```

```
# A tibble: 6 × 2
  state renewable_share
  <chr>           <dbl>
1 ak               1.35
2 al               9.81
3 ar               7.58
4 az               6.33
5 ca              16.6
6 co               8.46
```

```
question_2_data<-full_join(av_price_clean,renewable_share_2023,by="state")|>
    select("state","renewable_share","2023_av")
```

```r
head(question_2_data)
```

```
# A tibble: 6 × 3
  state renewable_share `2023_av`
  <chr>           <dbl>     <dbl>
1 ak               1.35      23.8
2 al               9.81      21.1
3 ar               7.58      21.8
4 az               6.33      30.3
5 ca              16.6       35.7
6 co               8.46      23.8
```

```r
correlation<-question_2_data|>
    summarize(core_value=cor(renewable_share,`2023_av`))|>
    print()
```

```
# A tibble: 1 × 1
  core_value
       <dbl>
1    0.00633
```

```r
#There is little to no correlation between a states average price of energy
and thee share of their energy production coming from renewable energy.
```

Part 4: Mapping Visualization

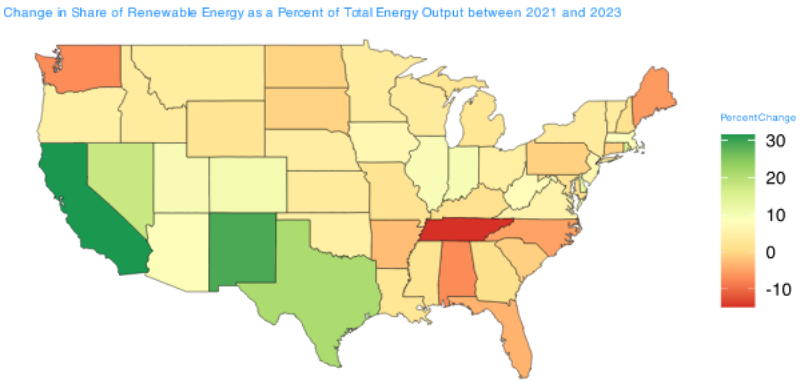
```r
state_map_data<-map("state", fill = TRUE, plot = FALSE)
states_sf_raw <- st_as_sf(state_map_data)
state_name_to_abb_map <- setNames(state.abb, state.name)
state_abb_lowercase<-tolower(state.abb)
state_lookup <- data.frame(ID = tolower(state.name), state =
state_abb_lowercase)
states_sf_clean <- states_sf_raw |>
    left_join(state_lookup, by = "ID") |>
    select(state, geom)
map_data_joined<-states_sf_clean|>
    left_join(total_use_all_renewable,by="state")

head(map_data_joined)
```

```
Simple feature collection with 6 features and 6 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
```

```
Bounding box:  xmin: -124.3834 ymin: 30.24071 xmax: -71.78015 ymax: 42.04937
Geodetic CRS:  +proj=longlat +ellps=clrk66 +no_defs +type=crs
  state energy_source total_use_2021 total_use_2022 total_use_2023
1    al    renewables         239817         232035         222189
2    az    renewables          99266         101215         108445
3    ar    renewables          89714          90825          87277
4    ca    renewables         810020         880995        1065179
5    co    renewables         103955         114917         115061
6    ct    renewables          49306          49084          48981
  percentage_change                           geom
1         -7.350605 MULTIPOLYGON (((-87.46201 3...
2          9.246872 MULTIPOLYGON (((-114.6374 3...
3         -2.716410 MULTIPOLYGON (((-94.05103 3...
4         31.500333 MULTIPOLYGON (((-120.006 42...
5         10.683469 MULTIPOLYGON (((-102.0552 4...
6         -0.659149 MULTIPOLYGON (((-73.49902 4...
```

```
map<-ggplot(map_data_joined)+geom_sf(aes(fill=percentage_change,))+
scale_fill_distiller(palette="RdYlGn",name="Percent Change",direction=1)+
    labs(title="Change in Share of Renewable Energy as a Percent of Total
Energy Output between 2021 and
2023")+theme_void()+theme(plot.title=element_text(size=6.5,color="dodgerblue"),legend.title=eleme
print(map)
```




```
## Analyzing this map, we can see California, New Mexico, Nevada, and Texas
lead the way with increasing their energy output coming from renewable
sources. On the other hand, Tennesee, Washington, Virgina, Maine, and Florida
have increased their usage of non-renewable energy sources since 2021.
```