EV Power - Lab 4 Project Report

Brian Choi

Example Solution 1

Part 0: libraries

```
library(tidyverse)
library(stringr)
library(kableExtra)
library(leaflet)
library(usmap)
library(sf)
```

Part 1: Defining Research Question

Chosen Question: * What is the average growth in renewable energy and non-renewable energy across different states and energy sources from 2023-2021? * Which states led the highest growth in either categories? * What energy sources had the largest change in usage from 2021-2023?

Part 2: Data Preparation and Cleaning

```
## import all data
avg_price = read_csv("data/av-energy-price-2021-2023.csv", show_col_types =
FALSE)
ev_registration = read_csv("data/ev-registrations-by-state-2023.csv",
show_col_types = FALSE)
```

```
New names:
• `` -> `...2`
```

```
renew_use_2021 = read_csv("data/renew-use-2021.csv", show_col_types = FALSE)
renew_use_2022 = read_csv("data/renew-use-2022.csv", show_col_types = FALSE)
renew_use_2023 = read_csv("data/renew-use-2023.csv", show_col_types = FALSE)

total_use_2021 = read_csv("data/total-use-2021.csv", show_col_types = FALSE)
total_use_2022 = read_csv("data/total-use-2022.csv", show_col_types = FALSE)
total_use_2023 = read_csv("data/total-use-2023.csv", show_col_types = FALSE)
```

```
# rename function
energy_source_rename = function(df){
    source = c('Coal','Natural Gas','Petroleum','Nuclear Energy','Total
Renewable Energy')
    df['Energy_Source'] = source
    return(df)
}

# standardize name for all three dfs
total_use_2021 = energy_source_rename(total_use_2021)
total_use_2022 = energy_source_rename(total_use_2022)
total_use_2023 = energy_source_rename(total_use_2023)
```

```
# extract digit function
extract_digit = function(df, column, pattern){
    df = df|>
        mutate(
            ## syntax for dynamically creating a new column
            !!{{column}} := str_extract(df[[{{column}}]], pattern)
        )
    return(df)
}

renew_use_2021 = extract_digit(renew_use_2021, "Renewable_Use_2021" , "\\d+")
renew_use_2022 = extract_digit(renew_use_2022, "Renewable_Use_2022" , "\\d+")
renew_use_2023 = extract_digit(renew_use_2023, "Renewable_Use_2023" , "\\d+")

#notice renew_use_2023 has some state names that are lower case
renew_use_2023$State = toupper(renew_use_2023$State)

#convert the usage into numeric values
renew_use_2021$Renewable_Use_2021 =
as.numeric(renew_use_2021$Renewable_Use_2021)
renew_use_2022$Renewable_Use_2022 =
as.numeric(renew_use_2022$Renewable_Use_2022)
renew_use_2023$Renewable_Use_2023 =
as.numeric(renew_use_2023$Renewable_Use_2023)


## fix registration column name
colnames(ev_registration) = c('State','EV_Count')
ev_registration = slice(ev_registration, c(3:nrow(ev_registration)))
ev_registration = extract_digit(ev_registration, "EV_Count", "\\d+")
```

```
## make column name easier to reference / slice only necessary rows
colnames(avg_price) = "x"
```

```
avg_price = slice(avg_price, c(3:nrow(avg_price)))

## split up by comma
avg_price = avg_price|>
    separate(
        x,
        into = c('State','2021','2022','2023'),
        sep = ','
    )

## extract digits only
avg_price = extract_digit(avg_price, "2021" , "\\d+")
avg_price = extract_digit(avg_price, "2022" , "\\d+")
avg_price = extract_digit(avg_price, "2023" , "\\d+")
```

- In this section, I renamed all the energy sources so that they are standarized across the renewable energy usage tables and the total energy usage tables. From there, I extracted the data values from avg_price and then organized it into a table. I also extracted numerical values where it was necessary.

Part 3: Joining / Pivoting Datasets for Analysis

```
## pivot the total use data frames
total_use_2021 = total_use_2021|>
    pivot_longer(
        cols = -Energy_Source,
        names_to = 'State',
        values_to = 'Energy_Use'
    )|>
    select("State", everything())

total_use_2022 = total_use_2022|>
    pivot_longer(
        cols = -Energy_Source,
        names_to = 'State',
        values_to = 'Energy_Use'
    )|>
    select("State", everything())

total_use_2023 = total_use_2023|>
    pivot_longer(
        cols = -Energy_Source,
        names_to = 'State',
        values_to = 'Energy_Use'
    )|>
    select("State", everything())
```

```
## join total use data frames and renew use data frames

total_energy_use = total_use_2021 |>
    left_join(
        total_use_2022,
        by = c('State', 'Energy_Source'),
        suffix = c('','_2022')
    ) |>
    left_join(
        total_use_2023,
        by = c('State','Energy_Source'),
        suffix = c('','_2023')
    ) |>
    rename(
        'Energy_Use_2021' = Energy_Use
    )

renew_energy_use = renew_use_2021 |>
    left_join(
        renew_use_2022,
        by = c('State', 'Energy_Source')
    ) |>
    left_join(
        renew_use_2023,
        by = c('State','Energy_Source')
    )
```

```
## create growth tables
## find average growth rate per state per category from 2023-2021
renew_growth = renew_energy_use |>
        mutate(
        Average_Growth = ((Renewable_Use_2023 - Renewable_Use_2021) /
Renewable_Use_2021) * 100
    )|>
    select("State","Energy_Source","Average_Growth")

total_growth = total_energy_use |>
    ## remove the total renewable energy rows for future computatio
    filter(Energy_Source != 'Total Renewable Energy') |>
    mutate(
        Average_Growth = ((Energy_Use_2023 - Energy_Use_2021) /
Energy_Use_2021) * 100
    ) |>
    select("State","Energy_Source", "Average_Growth")


summary_function = function(df, group, summary_name, user_func){
```

```
    f = match.fun(user_func)
    df|>
        group_by({{group}})|>
        summarise(
            !!{{summary_name}} := f(Average_Growth, na.rm = T)
        )
}


total_growth_state = summary_function(total_growth, State, "Total Growth (%)",
"sum")
total_growth_source = summary_function(total_growth, Energy_Source, "Total
Growth (%)", "sum")

renew_growth_state = summary_function(renew_growth, State, "Total Growth (%)",
"sum")
renew_growth_source = summary_function(renew_growth, Energy_Source, "Total
Growth (%)", "sum")
```

- Here I was focused on combining the renewable energy usage and total energy usage from 2021-2023 into usable data frame. First, I pivoted the data frames so that States became a column. Then I combined them based on energy type, making sure to specify the columns by what year they represented. From there, I created a growth data frame that calculated the average growth per State / Energy Source combination. Finally, that allowed me to group by State and Energy Soure to figure out the energy growths per category from 2021-2023.

Part 4: Summary Tables and Mapping Visualization

- Below are the tables that I have created:

Table 1: Growth in Non-Renewable Energy Usage by Source (2021-2023)

```
kable(total_growth_source, caption = "Growth in Non-Renewable Energy Usage by
Source (2021-2023)")
```

Energy_Source	Total Growth (%)
Coal	−1360.34543
Natural Gas	295.29370
Nuclear Energy	73.08843
Petroleum	39.00070

Table 2: Growth in Renewable Energy Usage by Source (2021-2023)

```
kable(renew_growth_source, caption = "Growth in Renewable Energy Usage by
Source (2021-2023)")
```

Energy_Source	Total Growth (%)
Biomass	43.364495
Geothermal	6.556796
Hydropower	−141.648808
Solar Energy	4382.719818
Wind Energy	50.729189

Table 3: Growth in Non-Renewable Energy Usage by State (2021-2023)

```
kable(total_growth_state, caption = "Growth in Non-Renewable Energy Usage by
State (2021-2023)")
```

State	Total Growth (%)
AK	15.3335377
AL	−26.2887335
AR	4.6795744
AZ	−4.5187920
CA	9.9501733
CO	−12.5357034
CT	−117.6770231
DC	−13.6315781
DE	−93.0978061
FL	−23.5758991
GA	5.4385971
HI	−72.8621778
IA	−8.4146466
ID	−47.7823700
IL	−32.6990407
IN	−14.8908290
KS	12.7415467
KY	−10.9032213

State	Total Growth (%)
LA	−69.9289886
MA	0.1450114
MD	−55.1038085
ME	−3.8021453
MI	−36.5963425
MN	−28.7600236
MO	94.8539296
MS	−15.4031362
MT	15.1192574
NC	−26.7964024
ND	5.8780796
NE	−2.2167071
NH	−43.7532779
NJ	−90.6266014
NM	−29.8890042
NV	−16.2644373
NY	−14.2138116
OH	−25.1648022
OK	−36.6206091
OR	−44.2860366
PA	−25.3403566
RI	7.1583864
SC	2.3975643
SD	2.0214597
TN	−8.2457958
TX	−5.1590079
US	−16.2908140
UT	−21.4292774
VA	−32.4901718
VT	−8.5734979

State	Total Growth (%)
WA	39.0529408
WI	−26.7481084
WV	−12.9377381
WY	7.7860674

Table 4: Growth in Renewable Energy Usage by State (2021-2023)

```
kable(renew_growth_state, caption = "Growth in Renewable Energy Usage by State
(2021-2023)")
```

State	Total Growth (%)
AK	46.104165
AL	101.295407
AR	64.362496
AZ	31.368129
CA	170.136668
CO	92.157052
CT	7.135217
DC	43.670696
DE	59.911038
FL	10.351794
GA	29.606747
HI	−4.979149
IA	102.886630
ID	53.189262
IL	99.808974
IN	207.073696
KS	17.064276
KY	112.850597
LA	9.128716
MA	30.579052
MD	19.170273

State	Total Growth (%)
ME	280.479097
MI	134.843928
MN	43.119664
MO	11.445482
MS	48.911768
MT	324.553597
NC	−16.684436
ND	38.613264
NE	69.213835
NH	80.738949
NJ	−23.029681
NM	65.488663
NV	5.813501
NY	69.154687
OH	81.011815
OK	45.760389
OR	34.965118
PA	27.887750
RI	129.977380
SC	8.292119
SD	819.855175
TN	57.538843
TX	84.296415
US	53.721495
UT	65.619648
VA	69.175567
VT	52.915192
WA	107.763745
WI	163.635492
WV	90.092103

State	Total Growth (%)
WY	13.679190

```
# grab centeroid per state
state_polys = us_map("states")

# convert the centroid values to an sf object (something called a CRS system)
state_sf = st_as_sf(state_polys)

# compute centroids of each state's polygon
state_centroids <- state_sf |>
  st_centroid() |> # transforms into point from polygons
  st_transform(crs = 4326) |> #transform the CRS values into Long and Lat
  mutate( #grab long and lat
    lon = st_coordinates(geom)[,1],
    lat = st_coordinates(geom)[,2]
  ) |>
  st_drop_geometry() |> # drops the original CRS column
  select(abbr, lon, lat) |>
  rename(State = abbr)

# at this stage, state_centroids is a data frame with state name lon and lat

# Merge with renewable energy growth table
renew_growth_map = left_join(renew_growth_state, state_centroids, by =
"State")

# add color and size column
renew_growth_map = renew_growth_map |>
  mutate(
    color = ifelse(`Total Growth (%)` >= 0, "blue", "red"),
    radius = abs(`Total Growth (%)`) * 500  # 500 seems the most visible
  )

# create leaflet map
leaflet(renew_growth_map) |>
  addProviderTiles("CartoDB.Positron") |> #cartoDB.Positron is just like the
background map we use as base
  addCircles(
    lng = ~lon,
    lat = ~lat,
    radius = ~radius,
    color = ~color,
    fillColor = ~color,
    fillOpacity = 0.6,
    popup = ~paste0(
      "<b>", State, "</b><br>",
```

```
      "Growth: ", round(`Total Growth (%)`, 2), "%"
    ) # add pop up for when you hover above the circle
  ) |>
  addLegend(
    position = "bottomright",
    colors = c("blue", "red"),
    labels = c("Positive Growth", "Negative Growth"),
    title = "Renewable Energy Growth (2021–2023)"
  )

# merge with non-renewable energy growth table
total_growth_map = left_join(total_growth_state, state_centroids, by =
"State") |>
     mutate(
        color = ifelse(`Total Growth (%)` >= 0, "blue", "red"),
        radius = abs(`Total Growth (%)`) * 1000  # adjust to 1000 since
different scales
    )

# create leaflet map
leaflet(total_growth_map) |>
  addProviderTiles("CartoDB.Positron") |> #cartoDB.Positron is just like the
background map we use as base
  addCircles(
    lng = ~lon,
    lat = ~lat,
    radius = ~radius,
    color = ~color,
    fillColor = ~color,
    fillOpacity = 0.6,
    popup = ~paste0(
      "<b>", State, "</b><br>",
      "Growth: ", round(`Total Growth (%)`, 2), "%"
    ) # add pop up for when you hover above the circle
  ) |>
  addLegend(
    position = "bottomright",
    colors = c("blue", "red"),
    labels = c("Positive Growth", "Negative Growth"),
    title = "Non-Renewable Energy Growth (2021–2023)"
  )
```

Part 5: Analysis

- Overall, we can see that the nationwide trend has been to increase Renewable Energy Usage while decreasing Non-Renewable Usage.
- Furthermore, when looking at the growths by Sourced, we can see that the greatest decrease in usage was coal at −1360% while the greatest increase in usage was Solar Energy at 4382%.

- The state with the greatest increase in Renewable Energy Usage was South Dakota at 819.86%.
- The state with the greatest increase in Non-Renewable Energy Usage was Montana at 94.85%.
- The state with the greatest decrease in Non-Renewable Energy Usage was Connecticut at −117.68%.
- We can see that the Map helps out a lot visually with our main questions when it comes to state groupings. It's easy to see the states with the largest growth and with the color coding which highlights negative growth, it's easy to tell the trend of growth between Renewable Energy Growth and Non-Renewable Energy Growth