EV Adoption vs Renewable Energy Mix (2021–2023)

Part 0: libraries

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.2
v ggplot2   4.0.0     v tibble    3.3.0
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.1.0
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to beco
```

```
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(dplyr)
library(ggplot2)
library(stringr)
library(janitor)
```

```
Attaching package: 'janitor'

The following objects are masked from 'package:stats':

    chisq.test, fisher.test
```

```
library(readr)
library(tigris)
```

```
To enable caching of data, set `options(tigris_use_cache = TRUE)`
in your R script or .Rprofile.
```

```
options(tigris_use_cache = TRUE)
library(scales)
```

```
Attaching package: 'scales'

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

    discard

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

    col_factor
```

```
library(dplyr)
library(knitr)
library(htmlwidgets)
library(webshot2)
```

Part 1: Defining Research Question

Chosen Question: In 2023, which states have a high number of EV registrations but a low share of renewable energy in their total energy mix?

Did states with the fastest growth in renewable energy share (from 2021 to 2023) also experience a decrease (or slower increase) in their average energy prices over the same period?

Part 2: Data Preparation and Cleaning

```
re_23 = read.csv('/Users/muhammedbakr/Desktop/Stat-133A/Projects/berkeley-stat133-zhiwei-clas
re_22 = read.csv('/Users/muhammedbakr/Desktop/Stat-133A/Projects/berkeley-stat133-zhiwei-clas
re_21 = read.csv('/Users/muhammedbakr/Desktop/Stat-133A/Projects/berkeley-stat133-zhiwei-clas
total_23 = read.csv('/Users/muhammedbakr/Desktop/Stat-133A/Projects/berkeley-stat133-zhiwei-c
total_22 = read.csv('/Users/muhammedbakr/Desktop/Stat-133A/Projects/berkeley-stat133-zhiwei-c
total_21 = read.csv('/Users/muhammedbakr/Desktop/Stat-133A/Projects/berkeley-stat133-zhiwei-c
ev_reg = read.csv('/Users/muhammedbakr/Desktop/Stat-133A/Projects/berkeley-stat133-zhiwei-cla
avg_e_price <- read_csv(
  "/Users/muhammedbakr/Desktop/Stat-133A/Projects/berkeley-stat133-zhiwei-classroom-ev-power-
  skip = 3,
  quote = "",
  col_names = c("state", "price_2021", "price_2022", "price_2023"),
  trim_ws = TRUE)
```

```
Rows: 52 Columns: 4
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (4): state, price_2021, price_2022, price_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.
```

```
re_23 <- re_23 %>% clean_names()
re_22 <- re_22 %>% clean_names()
re_21 <- re_21 %>% clean_names()

total_23 <- total_23 %>% clean_names()
total_22 <- total_22 %>% clean_names()
total_21 <- total_21 %>% clean_names()

ev_reg <- ev_reg %>% clean_names()
avg_e_price <- avg_e_price %>% clean_names()

avg_e_price <- avg_e_price |>
  mutate(state = str_squish(state),
    across(starts_with("price_"), readr::parse_number))
```

```
re_23 <- re_23 %>% rename(state = 1)
re_22 <- re_22 %>% rename(state = 1)
re_21 <- re_21 %>% rename(state = 1)

total_23 <- total_23 %>% rename(state = 1)
total_22 <- total_22 %>% rename(state = 1)
total_21 <- total_21 %>% rename(state = 1)

ev_reg <- ev_reg %>% rename(state = 1)
avg_e_price <- avg_e_price %>% rename(state = 1)
```

```
STATE_MAP_2_TO_FULL <- c(setNames(state.name, state.abb), "DC" = "District of Columbia")

normalize_state_full <- function(x) {
  x %>%
    str_trim() %>%
    str_remove_all('^"|"$') %>%
    str_replace_all(regex("^D\\.?C\\.?$", ignore_case = TRUE), "DC") %>%
    (\(.) ifelse(str_detect(., "^[A-Za-z]{2}$"),
                 STATE_MAP_2_TO_FULL[toupper(.)],
                 str_to_title(.)))() %>%
    replace_na("Unknown") %>%
    as.character()
}

parse_num <- function(x) readr::parse_number(x)
```

```
avg_e_price <- avg_e_price %>%
  clean_names() %>%
  rename(state = 1,
         price_2021 = 2, price_2022 = 3, price_2023 = 4) %>%
  mutate(
    state = normalize_state_full(state),
    across(starts_with("price_"),
           \(x) if (is.character(x)) readr::parse_number(x) else x))
```

```
clean_re_year <- function(df, year) {
  nm <- paste0("renewable_use_", year)
  df %>%
    clean_names() %>%
    rename(energy_source = 2, !!nm := 3) %>%
```

```
    mutate(
      state = normalize_state_full(state),
      !!nm := parse_num(.data[[nm]]))
}

re_21 <- clean_re_year(re_21, 2021)
re_22 <- clean_re_year(re_22, 2022)
re_23 <- clean_re_year(re_23, 2023)
```

```
ev_reg <- ev_reg %>%
  clean_names() %>%
  rename(state = 1, ev_count = 2) %>%
  mutate(state = normalize_state_full(state),
    ev_count = parse_num(ev_count))
```

```
clean_total_year <- function(df, year) {
  val_nm <- paste0("total_use_", year)
  df %>%
    clean_names() %>%
    rename(energy_source = 1) %>%
    pivot_longer(
      cols = -energy_source,
      names_to = "state_abbrev",
      values_to = val_nm
    ) %>%
    mutate(
      state = normalize_state_full(toupper(state_abbrev)),
      !!val_nm := as.double(.data[[val_nm]])
    ) %>%
    select(state, energy_source, all_of(val_nm))
}

total_21_long <- clean_total_year(total_21, 2021)
total_22_long <- clean_total_year(total_22, 2022)
total_23_long <- clean_total_year(total_23, 2023)
```

Part 3: Joining / Pivoting Datasets for Analysis

```
re_by_state <- re_21 %>%
  group_by(state) %>%
```

```
  summarise(renewable_use_2021 = sum(renewable_use_2021, na.rm = TRUE)) %>%
  left_join(
    re_22 %>% group_by(state) %>%
      summarise(renewable_use_2022 = sum(renewable_use_2022, na.rm = TRUE)),
    by = "state"
  ) %>%
  left_join(
    re_23 %>% group_by(state) %>%
      summarise(renewable_use_2023 = sum(renewable_use_2023, na.rm = TRUE)),
    by = "state"
  )

state_panel <- re_by_state %>%
  left_join(
    total_21_long %>% group_by(state) %>%
      summarise(total_use_2021 = sum(total_use_2021, na.rm = TRUE)),
    by = "state"
  ) %>%
  left_join(
    total_22_long %>% group_by(state) %>%
      summarise(total_use_2022 = sum(total_use_2022, na.rm = TRUE)),
    by = "state"
  ) %>%
  left_join(
    total_23_long %>% group_by(state) %>%
      summarise(total_use_2023 = sum(total_use_2023, na.rm = TRUE)),
    by = "state"
  ) %>%
  left_join(ev_reg, by = "state") %>%
  left_join(avg_e_price, by = "state")
```

```
state_panel <- state_panel %>%
  mutate(
    renew_share_2021 = renewable_use_2021 / total_use_2021,
    renew_share_2022 = renewable_use_2022 / total_use_2022,
    renew_share_2023 = renewable_use_2023 / total_use_2023
  )
```

```
qs <- state_panel %>%
  summarise(
    ev_q3 = quantile(ev_count, 0.75, na.rm = TRUE),
    rs_q1 = quantile(renew_share_2023, 0.25, na.rm = TRUE)
```

```
)

panel_2023 <- state_panel %>%
  mutate(high_ev = ev_count >= qs$ev_q3,
    low_rs = renew_share_2023 <= qs$rs_q1,
    flag_hiEV_loRS = high_ev & low_rs)

high_ev_low_share_2023 = panel_2023 |>
filter(high_ev == TRUE, low_rs == TRUE)
```

```
kable(head(state_panel, 10), caption = "Joined state panel (excerpt)")
```

Table 1: Joined state panel (excerpt)

state	renewable_2021	renewable_2022	renewable_2023	total_2021	total_2022	total_2023	ev_count	price_2021	price_2022	price_2023	renew_share_2021	renew_share_2022	renew_share_2023
Alabama	239816	232035	222189	2352652	2337513	2265008	13047	17.85	23.37	21.11	0.1019342	0.0992658	0.0980963
Alaska	9598	10410	10088	684975	730276	746979	2697	20.03	27.33	23.84	0.0140122	0.0142549	0.0135051
Arizona	99266	101214	108445	1681257	1651851	1712668	89798	25.07	31.72	30.28	0.0590427	0.0612729	0.0633194
Arkansas	89714	90824	87277	1136025	1178115	1151067	7108	18.42	23.84	21.76	0.0789719	0.0770926	0.0758230
California	810020	880995	1065179	6142256	6244176	6429818	1256646	28.44	37.35	35.72	0.1318767	0.1410907	0.1656624
Colorado	103956	114918	115062	1364155	1411476	1359509	90083	20.64	25.85	23.85	0.0762054	0.0814169	0.0846351
Connecticut	49306	49084	48983	821709	831195	789642	31557	25.85	33.15	32.32	0.0600042	0.0590523	0.0620319
Delaware	7151	7402	8040	208041	210948	203487	8435	21.83	27.74	26.70	0.0343730	0.0350892	0.0395111
District of Columbia	2487	2623	2796	49262	50796	46323	NA	25.67	30.84	32.28	0.0504852	0.0516379	0.0603588
Florida	297290	304605	286307	4145504	4273099	4237858	254878	22.53	29.35	28.12	0.0717138	0.0712843	0.0675594

```
kable(high_ev_low_share_2023, caption = "2023: High EV registrations & Low Renewable Share")
```

Table 2: 2023: High EV registrations & Low Renewable Share

state	renewable_2021	renewable_2022	renewable_2023	total_2021	total_2022	total_2023	ev_count	price_2021	price_2022	price_2023	renew_share_2021	renew_share_2022	renew_share_2023	high_ev	low_rs	flag_hiEV_loRS
Maryland	52713	51256	53711	1011360	989904	975123	72139	23.64	29.98	28.85	0.0521407	0.0517788	0.0550813	TRUE	TRUE	TRUE
New Jersey	70039	73188	74409	1823039	1900068	1879114	134753	21.79	27.11	25.20	0.0384190	0.0385188	0.0395979	TRUE	TRUE	TRUE
Texas	654199	751680	791210	13598122	13708449	14059023	230125	15.38	20.78	17.37	0.0481092	0.0548333	0.0562778	TRUE	TRUE	TRUE

Part 4: Mapping Visualization

```
states_sf <- states(cb = TRUE, year = 2023, class = "sf") %>%
  filter(!STUSPS %in% c("AS","GU","MP","PR","VI")) %>%
  transmute(state = as.character(NAME), geometry)

states_sf$state[states_sf$state == "District of Columbia"] <- "District Of Columbia"

map_sf <- states_sf %>%
  left_join(panel_2023, by = "state")
```

```
pal <- colorNumeric(palette = "Blues", domain = map_sf$renew_share_2023, na.color = "#dddddd"


radius_fun <- function(x) sqrt(x) * 0.15

leaflet(map_sf, options = leafletOptions(zoomControl = TRUE)) %>%
  addProviderTiles(providers$CartoDB.Positron) %>%
  addPolygons(
    fillColor = ~pal(renew_share_2023),
    weight = 0.7, color = "#555", opacity = 1,
    fillOpacity = 0.8,
    popup = ~sprintf(
      "<b>%s</b><br/>Renewable share 2023: %s<br/>EV registrations: %s",
      state,
      percent(renew_share_2023, accuracy = 0.1),
      comma(ev_count)
    ),
    group = "Renewable share (2023)"
  ) %>%
  addCircleMarkers(
    data = st_centroid(map_sf),
    radius = ~ifelse(is.na(ev_count), 0, pmax(3, radius_fun(ev_count))),
    stroke = TRUE, weight = 1, color = "#444",
    fillColor = "#2c7fb8", fillOpacity = 0.6,
    label = ~sprintf("%s - EVs: %s", state, comma(ev_count)),
    group = "EV registrations (bubbles)"
  ) %>%
  addPolylines(
    data = subset(map_sf, flag_hiEV_loRS),
    color = "#d7191c", weight = 3, opacity = 0.9,
```

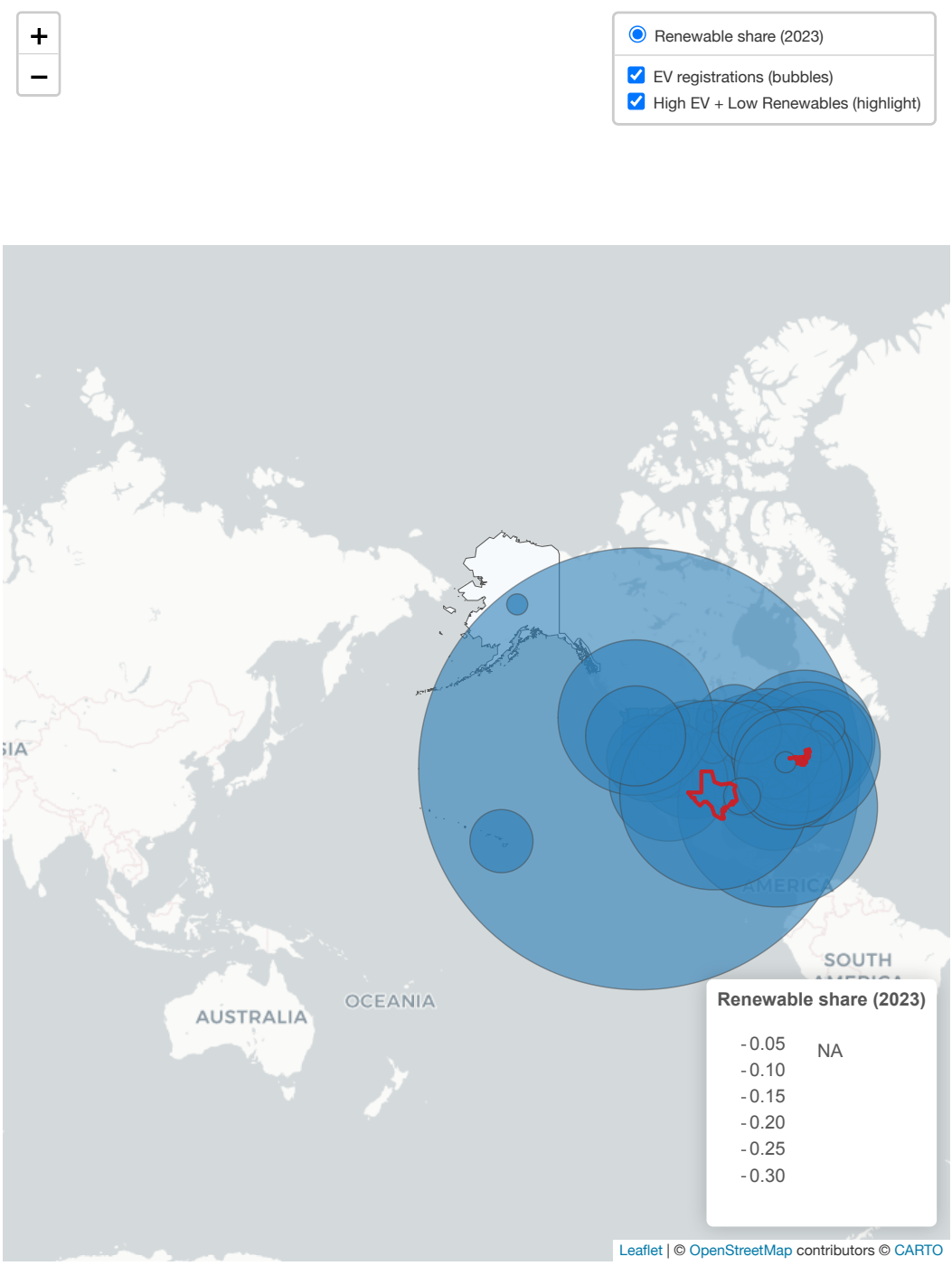
```
    group = "High EV + Low Renewables (highlight)"
  ) %>%
  addLegend("bottomright", pal = pal, values = ~renew_share_2023,
            title = "Renewable share (2023)", labFormat = labelFormat(suffix = ""),
            opacity = 0.8) %>%
  addLayersControl(
    overlayGroups = c("EV registrations (bubbles)", "High EV + Low Renewables (highlight)"),
    baseGroups = c("Renewable share (2023)"),
    options = layersControlOptions(collapsed = FALSE))
```

```
Warning: sf layer has inconsistent datum (+proj=longlat +datum=NAD83 +no_defs).
Need '+proj=longlat +datum=WGS84'
```

```
Warning: st_centroid assumes attributes are constant over geometries
```

```
Warning: sf layer has inconsistent datum (+proj=longlat +datum=NAD83 +no_defs).
Need '+proj=longlat +datum=WGS84'
Warning: sf layer has inconsistent datum (+proj=longlat +datum=NAD83 +no_defs).
Need '+proj=longlat +datum=WGS84'
```

```
file:////private/var/folders/rg/zyndr7495tb3d1pl0sfsmv1r0000gn/T/Rtmp06Ijsw/filea28249dda34b/
```


+
−
Renewable share (2023)
EV registrations (bubbles)
High EV + Low Renewables (highlight)
AUSTRALIA
OCEANIA
AMERICA
SOUTH
Renewable share (2023)
-0.05
-0.10
-0.15
-0.20
-0.25
-0.30
NA
Leaflet | © OpenStreetMap contributors © CARTO