```r
# Assignment 2 original data viz
# Xiyu Zhang

# Load packages ------------------------------------------------------------

library(sf)
library(tidyverse)
library(ggplot2)

# Data preparation ---------------------------------------------------------

# Read in the initial dataset

electricity_station_initial <-

  # read in the dataset

  st_read('data_own/alt_fuel_stations.geojson') %>%

  # convert an sf object into a pure tibble

  as_tibble()

# Filter for the wanted data

electricity_station <-
  electricity_station_initial %>%

  # filter for the wanted types

  filter(

    # only include public electricity stations but not private ones

    access_code == 'public',
```

```r

    # only include those are currently available but not planned nor
    # temporarily unavailable

    status_code == 'E',

    # only include those in the US

    country == 'US',

    # only include the charging stations open to the public

    restricted_access == FALSE) %>%

  # select the wanted traits of those electricity charging stations

  select(
    c(access_days_time, id, open_date, owner_type_code, state,
    ev_pricing, ev_renewable_source, facility_type))

# Create the wanted variables

# To create a dataset, as for every state, including the free pricing rate,
# 24-hour pricing rate, and the density of the charging stations in different
# states open to the public

# The intention of this data visualization is to visualize the convenience for
# people to charge their private electronic vehicles

# develop wanted variables

elec_station_by_state <-
  electricity_station %>%

```

```r
  # select the wanted features

  select(id, state, access_days_time, ev_pricing) %>%

  # exclude missing values

  filter(!is.na(access_days_time),
         !is.na(ev_pricing)) %>%

  # construct two Boolean values describing whether a station operates for 24
  # hours or not, and whether this station offers free charging, respectively

  transmute(
    id,
    state,
    x =
      if_else(
        str_detect(access_days_time, '24'),
        TRUE,
        FALSE),
    y =
      if_else(
        str_detect(ev_pricing, 'Free'),
        TRUE,
        FALSE)) %>%
```

```
96
97    # renames these two Boolean values to full_hours and free_charging
98
99    rename(full_hours = x,
100          free_charging = y,
101          state.abb = state)
102
103 # Gather states information in R build-in data sets
104
105 data(state)
106
107 # select wanted features and construct a tibble
108
109 state_features <-
110
111   # including state abbreviation, state area, and state name
112
113   tibble(state.abb, state.area, state.name, state.region)
114
115 # I don't know how to do these concisely so I hard-code to build the wanted
116 # variables
117
118 # 1. construct a variable naming state.amount to describe the full amount of these
119 # electricity charging station in each US state
120
121 temp <-
122   elec_station_by_state %>%
123   group_by(state.abb) %>%
124   summarise(state.amount = n())
125
126 # 2. join this variable with the state features tibble construct above
127
```

```r

temp_2 <-
  elec_station_by_state %>%
  left_join(temp) %>%
  left_join(state_features)

# 3. Respectively, calculate free_charging_rate and full_hour_rate for each
# state, referring to among all of the electric vehicle charging station, the
# ratio of free charging stations and the ratio of charging stations operating
# 24 hours everyday

temp_3 <-
  temp_2 %>%
  group_by(state.abb, free_charging, state.amount) %>%
  summarise(free_charging_amount = n(),

            # ignore 'state.amount', only group by the first two variables

            .groups = 'drop_last') %>%

  # calculating the charging stations that offers free charging

  filter(free_charging == TRUE) %>%

  # calculate the wanted variable by divide the free charging station amount by
  # the whole charging station amount in each state

  mutate(free_charging_rate = free_charging_amount / state.amount) %>%

  # select useful variables for future data visualization

  select(state.abb, free_charging_rate)

# temp_4 is basically the same as the previous one, but for full-hours rate
```

```r

# temp_4 is basically the same as the previous one, but for full-hours rate

temp_4 <-
  temp_2 %>%
  group_by(state.abb, full_hours, state.amount) %>%
  summarise(full_hours_amount = n(),
            .groups = 'drop_last') %>%
  filter(full_hours == TRUE) %>%
  mutate(full_hours_rate = full_hours_amount / state.amount) %>%
  select(state.abb, full_hours_rate, state.amount)

# 4. Finally, combine the constructed variables together in one tibble

temp_fin <-
  left_join(state_features,
            temp_3) %>%
  left_join(temp_4) %>%

  # calculate the electronic vehicle charging station density in each state
  # by divide the amount in each state by the stata area

  mutate(station_density = state.amount / state.area)

```

```r
# Data visualization ----------------------------------------------------

temp_fin %>%
  ggplot() +
  geom_point(
    aes(x = full_hours_rate,
        y = free_charging_rate,
        size = sqrt(station_density / pi),
        color = state.region),
    alpha = 0.7) +
  geom_text(
    aes(x = full_hours_rate,
        y = free_charging_rate),
    label = ifelse(
      ((temp_fin$full_hours_rate >= 0.75 |
          temp_fin$free_charging_rate >= 0.9) |
         (temp_fin$full_hours_rate < 0.5 |
            temp_fin$free_charging_rate < 0.6)),
      state.abb,
      ''),
    size = 3.5,
    color = '#636363',
    hjust = 0,
    nudge_x = 0.003) +
  scale_color_manual(values =
                       c('#1f78b4', '#33a02c', '#bebada', '#fdcdac')) +
  scale_size(range = c(.1, 20),
             name = paste('Electronic vehicle\ncharging station density',
                          '\n(Unit/mi^2)')) +
  scale_x_continuous(limits = c(0.2, 1.0)) +
  scale_x_continuous(limits = c(0.2, 1.0)) +
  labs(title = paste('The Most Convenient and Inconvenient US states to',
                     'Charge Electric Vehicles in 2023'),
       subtitle = paste('Connecticut and Maryland Leading While Texas Falling',
                        'Behind'),
       caption = 'Source: Alternative Fuel Data Center',
       x = 'Station ratio of offering full hours charging',
       y = 'Station ratio of offering free charging') +
  theme_bw() +
  theme(
    axis.ticks = element_blank(),
    axis.line = element_line(colour = 'gray'),
    panel.border = element_blank(),
    panel.grid = element_blank())
```