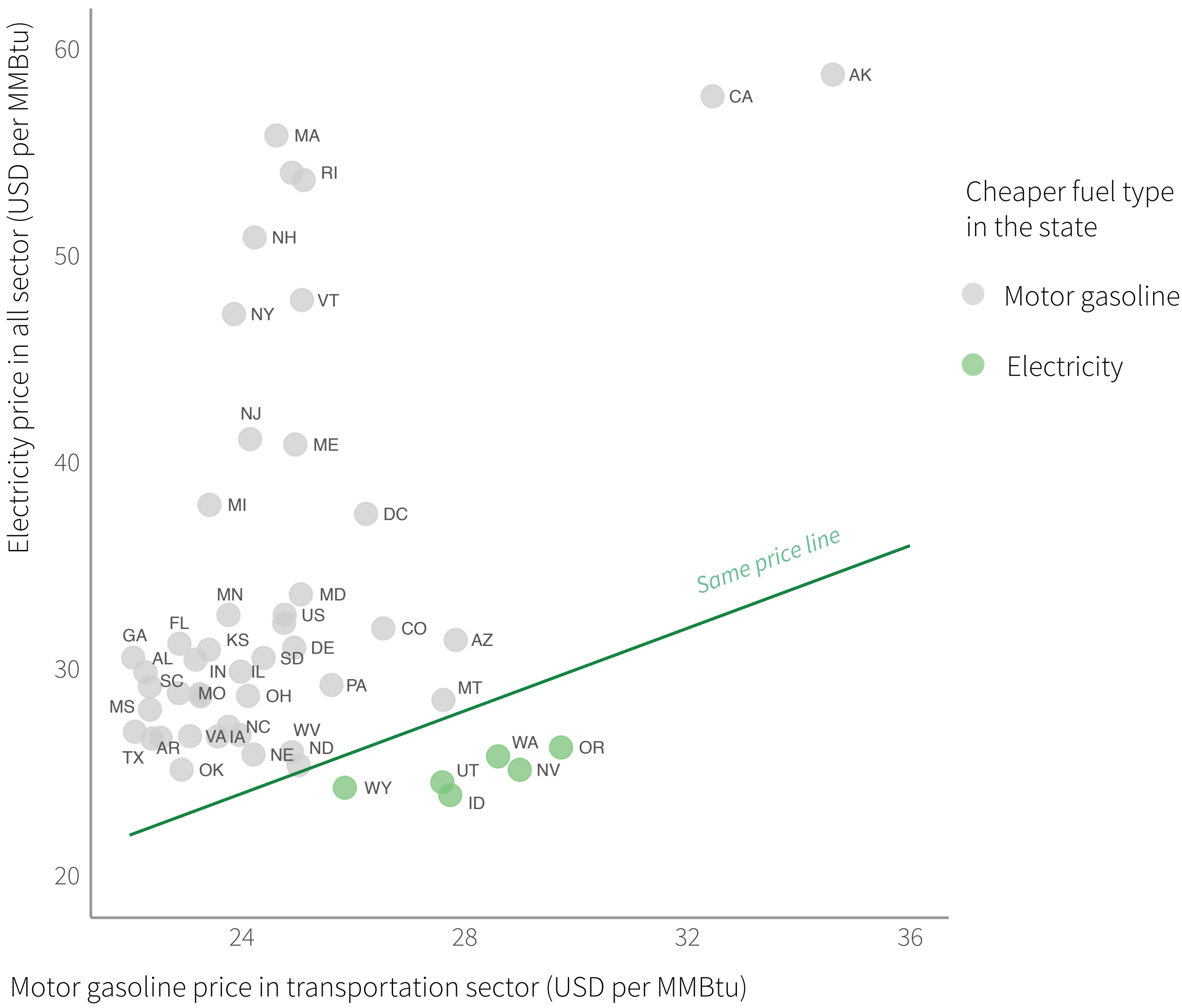
The purpose of my original visualization

First, to practice generating a scatterplot with a similar format to Yau's figure, I am trying to develop my own ability to create this kind of figure.

Second, I created this figure with consideration of my thesis topic. For my thesis, I focused on the factors that influence people's consumption decisions of electric vehicles across different states in the US. Given that the cost of fuel is a key independent variable that affects the overall cost of driving, I chose to represent each state's motor gasoline price and electricity price on the x-axis and the y-axis, respectively, as a measurement of the cost of each fuel. Then, I tried to use each dot to represent a state, and use their locations to position each state's fuel cost for driving both gasoline vehicles and electric vehicles. By doing so, I tried to highlight the potential influence of fuel prices on individuals' consumption decisions across different states.

Six states where electricity is cheaper than motor gasoline

The cost of driving gasoline and electric cars in the United States in 2021



Source: U.S. Energy Information Administration

```r
# Assignment 01 original
# Xiyu Zhang

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

library(tidyverse)
library(sf)

# Original scatterplot ---------------------------------------------------

# Read in the initial dataset

# motor gasoline price for each state, from EIA, in 2021

gasoline_21 <-
  read_csv('data_own/fuel_mg.csv') %>%
  filter(MSN == 'MGACD') %>%
  rename(motor_gasoline = '2021')

# all sector electricity price for each state, from EIA, in 2021

electricity_21 <-
  read_csv('data_own/fuel_es.csv') %>%
  filter(MSN == 'ESTCD') %>%
  rename(electricity = '2021')

# join the two forms into one

price_21 <-
  left_join(gasoline_21,
            electricity_21,
            by = 'State') %>%
  select(State, motor_gasoline, electricity) %>%
  mutate(cheap = ifelse(electricity <= motor_gasoline,
                        T,
                        F))

```

```r

# create an elementary scatterplot

price_21 %>%
  ggplot(
    aes(x = motor_gasoline,
        y = electricity)) +
  geom_point(colour = 'blue')

# It seems there exists an outlier, so descending the electricity price and
# remove the outlier, and create the scatterplot again

price_21 %>%
  arrange(desc(electricity)) %>%
  filter(State != 'HI') %>%
  ggplot(
    aes(x = motor_gasoline,
        y = electricity,
        label = State,
        color = cheap)) +
  geom_point(alpha = 0.7,
             size = 4) +
  geom_segment(
    aes(x = 22,
        y = 22,
        xend = 36,
        yend = 36),
    color = '#238443') +
  geom_text(hjust = -0.7,
            vjust = 0.5,
            size = 2.6,
            check_overlap = T,
            color = '#525252') +
```

```r
            check_overlap = T,
            color = '#525252') +
  scale_y_continuous(limits = c(20, 60)) +
  scale_color_manual(values = c('#cccccc', '#74c476')) +
  labs(title = paste('The cost of driving gasoline and electric cars in the',
                     'United States'),
       subtitle = paste('Motor gasoline price and electricity price in each',
                        'state of US in 2021'),
       caption = 'Source: U.S. Energy Information Administration',
       x = 'Motor gasoline price in transportation sector (USD per MMBtu)',
       y = 'Electricity price in all sector (USD per MMBtu)') +
  theme_bw() +
  theme(
    axis.line = element_line(colour = '#969696'),
    axis.ticks = element_blank(),
    panel.border = element_blank(),
    panel.grid = element_blank())
```

I tried to construct a bubble graph like this assignment's replication practice.

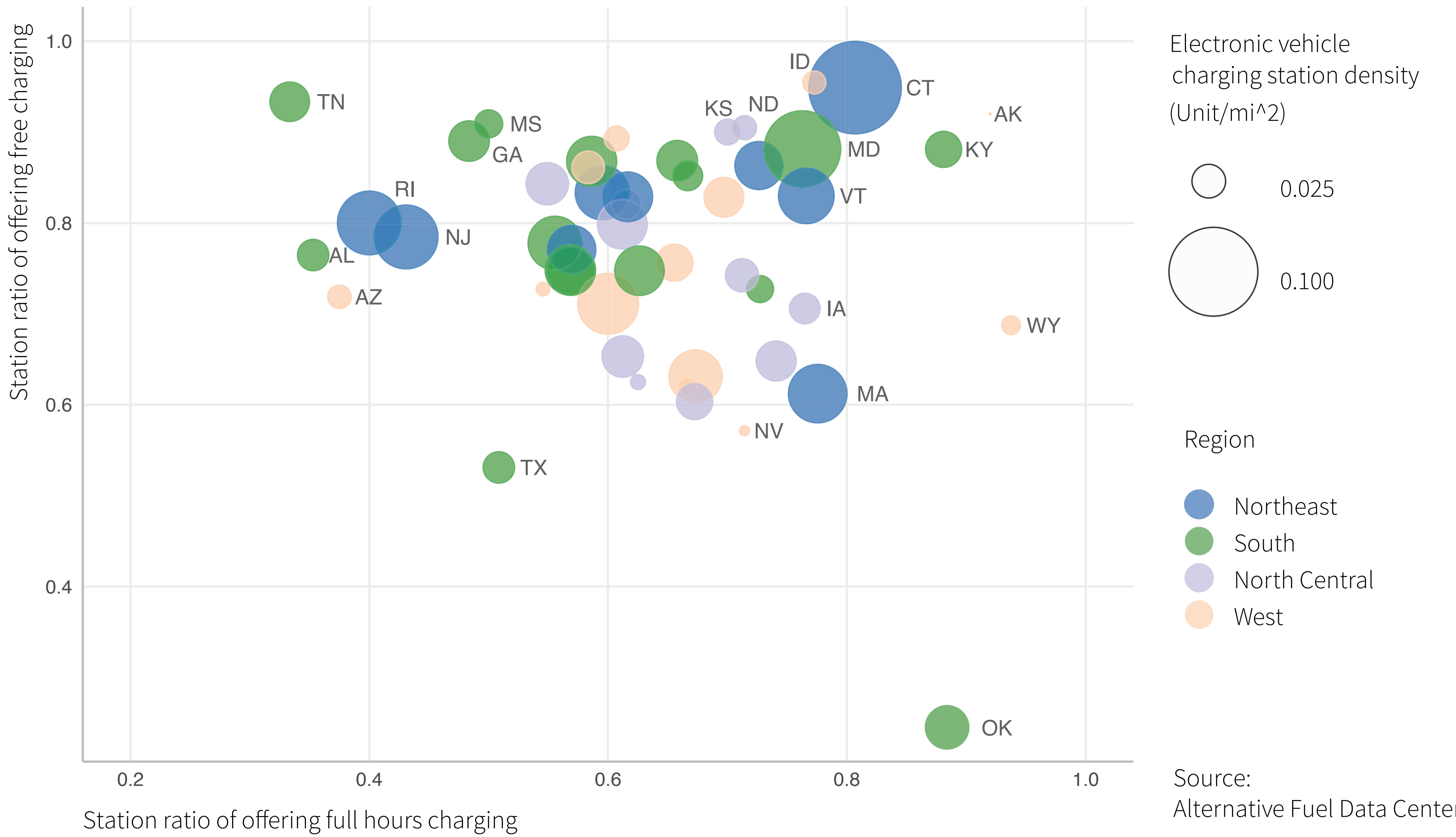
Here, the two dimensions in the x-axis and y-axis show each state's electronic vehicle public charging station's free charging rate and full hours charging rate, respectively. The third dimension shows the electronic vehicle charging station density, which that means on an average area of land in a state, how many publicly accessible electronic vehicles charging stations residents can find.

Overall, the chart tries to present the convenience for residents in different states to charge their vehicles in daily life. As one state approaches the upper and right frontiers of this chart with a large bubble, that indicates charging one's electronic vehicle in this state will be pretty convenient.

My electronic vehicle charging station data comes from afdc.energy.gov. The state area data comes from an R built-in data set named ‘state’.

The Most Convenient and Inconvenient US states to Charge Electric Vehicles in 2023

Connecticut and Maryland Leading While Texas Falling Behind



Note:
The further to the right and up the location of the point representing a state, the larger its area, indicating that it is more convenient for residents to charge their electric vehicles.
The states with more extreme values are marked in the graph. As one could see, Connecticut and Maryland are more convenient to charge because they are located on the upper right side of the graph and have a larger area; Texas, located on the lower left side, is not convenient to charge, as it also has a smaller area.

```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)) %>%
```

```r

  # renames these two Boolean values to full_hours and free_charging

  rename(full_hours = x,
         free_charging = y,
         state.abb = state)

# Gather states information in R build-in data sets

data(state)

# select wanted features and construct a tibble

state_features <-

  # including state abbreviation, state area, and state name

  tibble(state.abb, state.area, state.name, state.region)

# I don't know how to do these concisely so I hard-code to build the wanted
# variables

# 1. construct a variable naming state.amount to describe the full amount of these
# electricity charging station in each US state

temp <-
  elec_station_by_state %>%
  group_by(state.abb) %>%
  summarise(state.amount = n())

# 2. join this variable with the state features tibble construct above

```

```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)

```

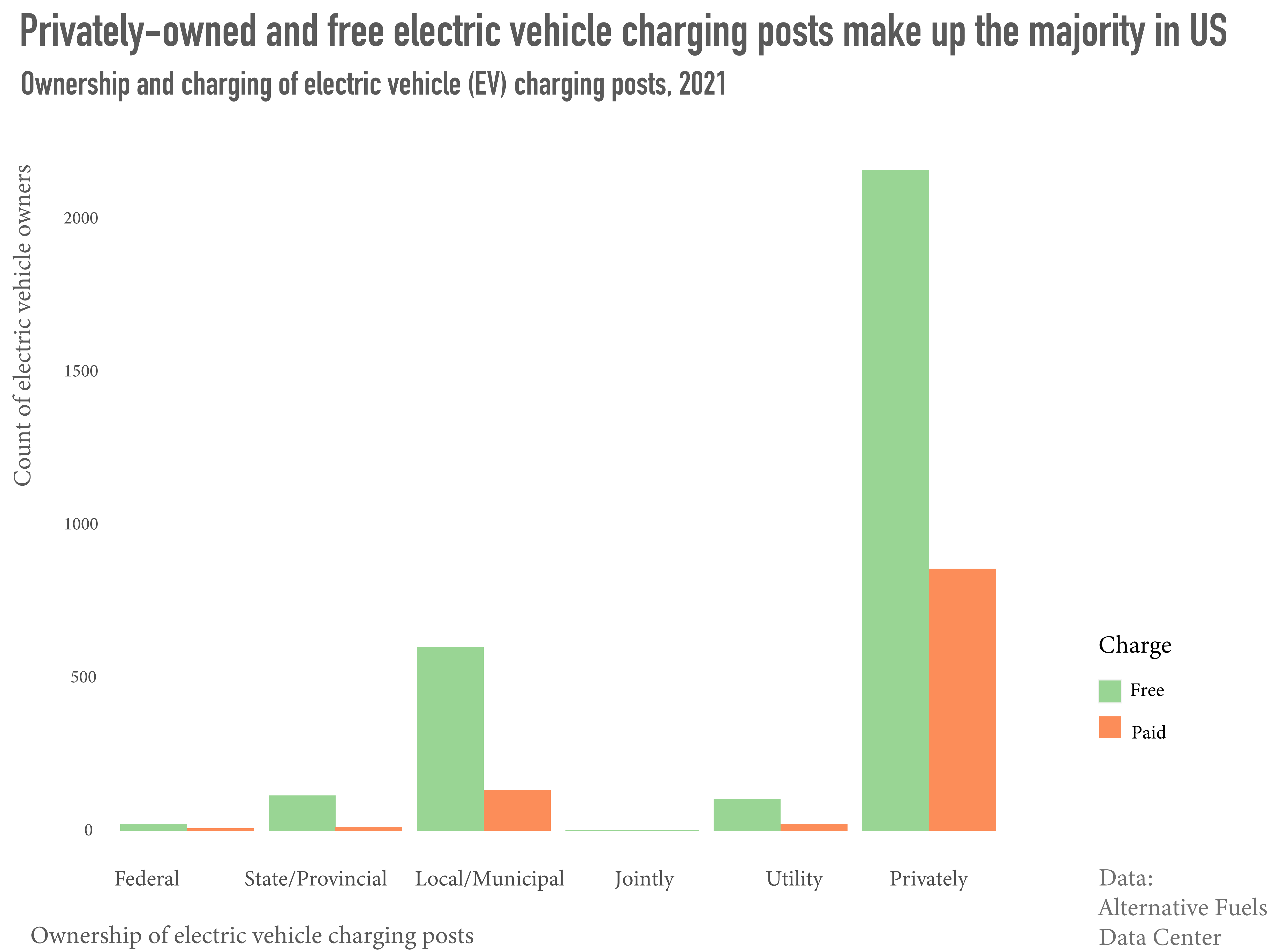
```r
# Data visualization -----------------------------------------------------

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)) +

```r
  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())
```

This bar chart still serves the theme of my thesis. The reason I created this bar chart is that I wanted to see how many of the charging posts built by different owners that are open to the public are free and how many of them are paid. I used a light green color for free and a light orange color to show a mild contrast with it for charging.

Ultimately, what this graph tells us is that, first, free charging is the majority of charging posts owned by various entities. Secondly, private charging posts are the type with the most ownership of charging posts.


Privately-owned and free electric vehicle charging posts make up the majority in US
Ownership and charging of electric vehicle (EV) charging posts, 2021
Count of electric vehicle owners
2000
1500
1000
500
0
Federal
State/Provincial
Local/Municipal
Jointly
Utility
Privately
Ownership of electric vehicle charging posts
Charge
Free
Paid
Data:
Alternative Fuels
Data Center

03_1

ShaeChang

2023-03-16

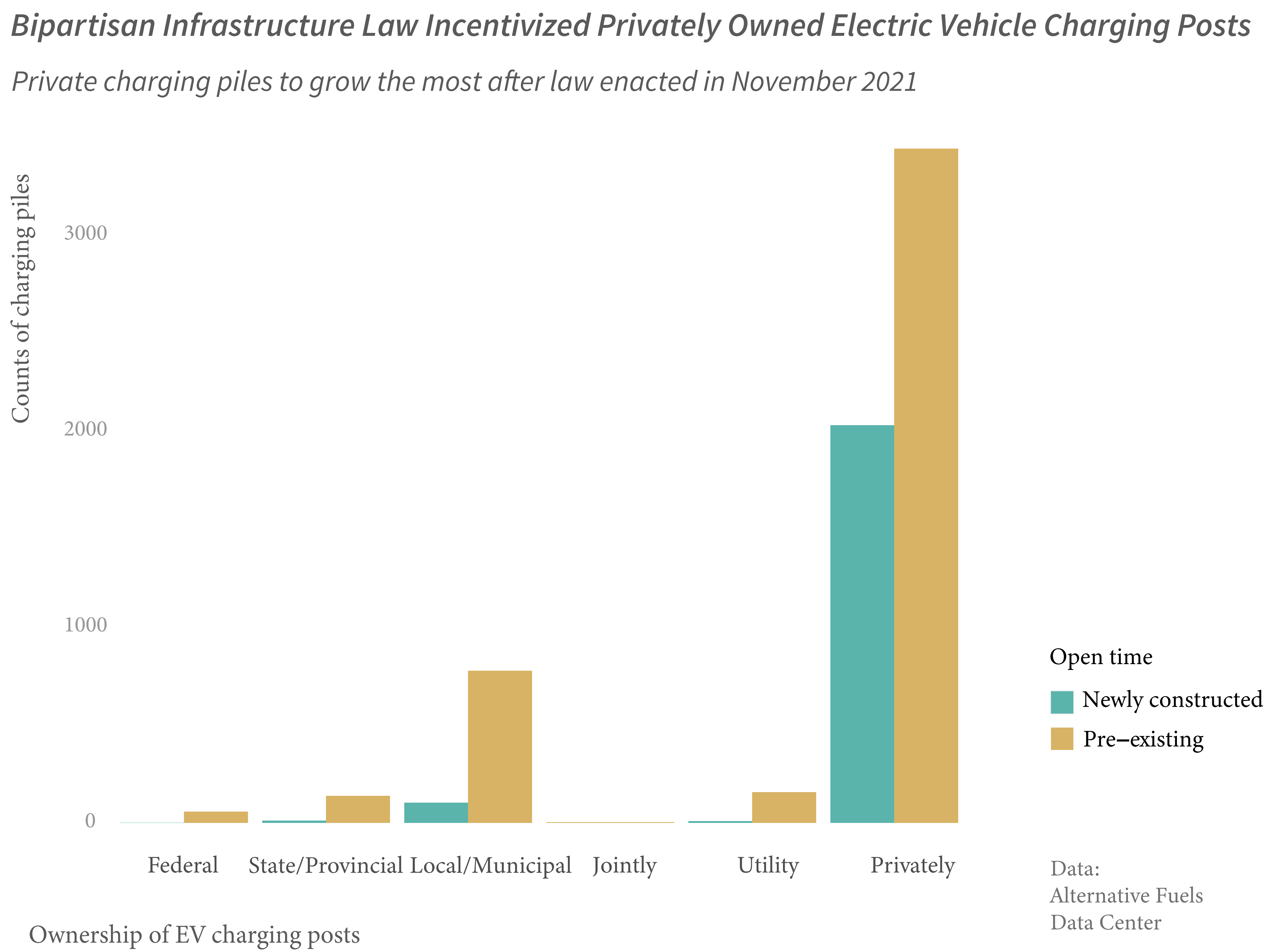
```r
# Load packages ----------------------------------------------------------

library(tidyverse)
library(sf)

# Load the data ----------------------------------------------------------

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()
```

```
## Reading layer `alt_fuel_stations' from data source
##   `/Users/maxzhang/GU/Data_viz/data_own/alt_fuel_stations.geojson'
##   using driver `GeoJSON'
## Simple feature collection with 58698 features and 66 fields
## Geometry type: POINT
## Dimension:     XY
## Bounding box:  xmin: -164.8489 ymin: 0 xmax: 77.64996 ymax: 64.85247
## Geodetic CRS:  WGS 84
```

```r
electricity_station <-
  electricity_station_initial %>%

  # filter for the wanted types

  filter(

    # only include public electricity stations but not private ones

    access_code == 'public',

    # 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))

# Data preperation --------------------------------------------------------

# create the wanted variable

elec_new <-
  electricity_station %>%

  # create a variable measuring this station charge individuals or not

  mutate(Charge =
           if_else(
             str_detect(ev_pricing, 'Free'),
             'Free',
             'Not free')) %>%

  # filter the missing values for the two variables we care

  filter(
    !is.na(Charge),
    !is.na(owner_type_code)) %>%

  # generate new categorical names
```

```r
  mutate(owner_type_new =
           case_when(
             owner_type_code == 'FG' ~ 'Federal',
             owner_type_code == 'J' ~ 'Jointly',
             owner_type_code == 'LG' ~ 'Local/Municipal',
             owner_type_code == 'P' ~ 'Privately',
             owner_type_code == 'SG' ~ 'State/Provincial',
             owner_type_code == 'T' ~ 'Utility'))

  # convert the owner type into a factor with specific levels

elec_new$owner_type_new <-
  factor(elec_new$owner_type_new,
         levels = c('Federal',
                    'State/Provincial',
                    'Local/Municipal',
                    'Jointly',
                    'Utility',
                    'Privately'))

# Data visualization ------------------------------------------------------

p1 <-
  elec_new %>%
  ggplot(mapping =
           aes(x = owner_type_new)) +
  geom_bar(aes(fill = Charge),
           position = 'dodge') +
  scale_x_discrete(drop = FALSE) +

  # to use green to represent free while use a diverging color of orange to
  # represent not free

  scale_fill_manual(values = c('#99d594',
                               '#fc8d59')) +
  labs(title = 'Ownership and charging of electric vehicle charging posts',
       subtitle = paste('Charges for electric vehicle charging stations of',
               'different owners in US'),
       caption = 'Data: afdc.energy.gov',
       x = 'Ownership of EV charging posts',
       y = 'Count (numbers)') +
  theme(
    axis.ticks = element_blank(),
    panel.background = element_blank())
```

This data visualization, like the previous one, serves the topic of the thesis. Its purpose is to show which entity will have the greatest incentive to build electric vehicle charging posts as a result of the Bipartisan Infrastructure Law Incentives issued by the Biden administration in November 2021. Based on the graphs made, it appears that this bill has incentives for Local/Municipal and Private for the construction of charging posts and that the incentives are mainly focused on private entities. This makes sense because under the new bill, private entities can receive tax credits for building EV charging stations. But on the other hand, the Bipartisan Infrastructure Act also allocates significant funds to all levels of government to facilitate the construction of EV charging infrastructure, and the flow and role of these funds should be further examined.


Bipartisan Infrastructure Law Incentivized Privately Owned Electric Vehicle Charging Posts
Private charging piles to grow the most after law enacted in November 2021
Counts of charging piles
3000
2000
1000
0
Federal
State/Provincial
Local/Municipal
Jointly
Utility
Privately
Ownership of EV charging posts
Open time
Newly constructed
Pre–existing
Data:
Alternative Fuels
Data Center

03-2

ShaeChang

2023-03-16

```r
# Load packages ----------------------------------------------------------

library(tidyverse)
library(sf)

# Load the data ----------------------------------------------------------

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()
```

```
## Reading layer `alt_fuel_stations' from data source
##   `/Users/maxzhang/GU/Data_viz/data_own/alt_fuel_stations.geojson'
##   using driver `GeoJSON'
## Simple feature collection with 58698 features and 66 fields
## Geometry type: POINT
## Dimension:     XY
## Bounding box:  xmin: -164.8489 ymin: 0 xmax: 77.64996 ymax: 64.85247
## Geodetic CRS:  WGS 84
```

```
electricity_station <-
  electricity_station_initial %>%

  # filter for the wanted types

  filter(

    # only include public electricity stations but not private ones

    access_code == 'public',

    # 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))

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

p2 <-
  elec_date <-
  electricity_station %>%
  filter(!is.na(open_date),
         !is.na(owner_type_code)) %>%
  mutate(open_time =
           if_else(open_date >= lubridate::ymd('2021-11-16'),
                   'Newly constructed',
                   'Pre-existing')) %>%
  mutate(owner_type_new =
           case_when(
             owner_type_code == 'FG' ~ 'Federal',
             owner_type_code == 'J' ~ 'Jointly',
             owner_type_code == 'LG' ~ 'Local/Municipal',
             owner_type_code == 'P' ~ 'Privately',
             owner_type_code == 'SG' ~ 'State/Provincial',
             owner_type_code == 'T' ~ 'Utility'))

elec_date$owner_type_new <-
  factor(elec_date$owner_type_new,
```

```
       levels = c('Federal',
                  'State/Provincial',
                  'Local/Municipal',
                  'Jointly',
                  'Utility',
                  'Privately'))

# Data visualization------------------------------------------------------

p2 <-
  elec_date %>%
  ggplot(mapping =
           aes(x = owner_type_new)) +
  geom_bar(aes(fill = open_time),
           position = 'dodge') +
  scale_x_discrete(drop = FALSE) +

  # to use green to represent free while use a diverging color of orange to
  # represent not free

  scale_fill_manual(values = c('#5ab4ac',
                               '#d8b365')) +
  labs(title = paste('Bipartisan Infrastructure Law Incentives for building EV',
                     'charging posts'),
       subtitle = paste('Comparison of the number of charging piles built by',
                        'different entities before and after November 16, 2021'),
       caption = 'Data: afdc.energy.gov',
       x = 'Ownership of EV charging posts',
       y = 'Count (numbers)') +
  theme(
    axis.ticks = element_blank(),
    panel.background = element_blank())
```

01

Statement of Purpose

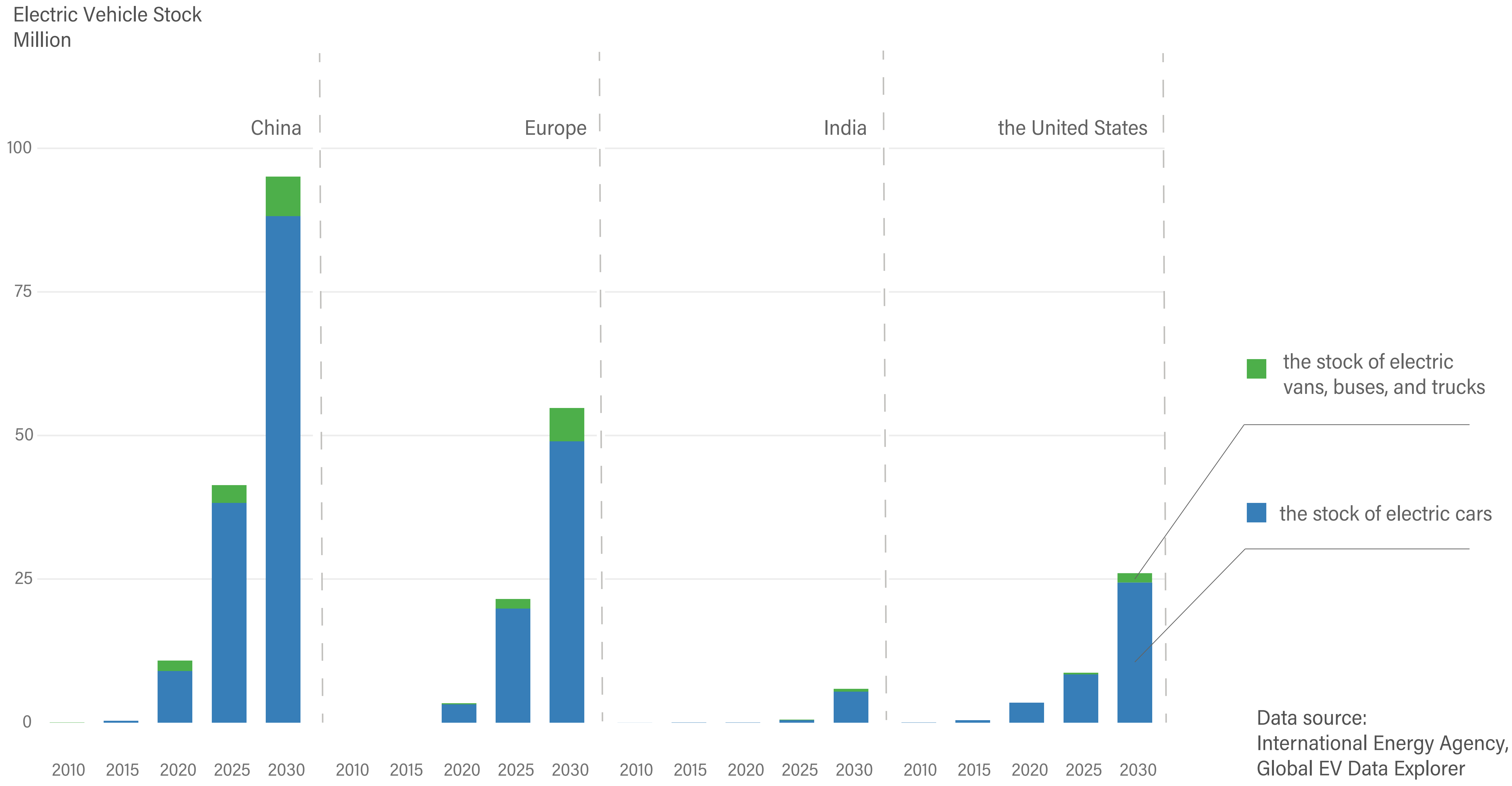
In this diagram, as the title suggests, I want to convey the message that electric cars are dominant in the world among all types of electric vehicles. I plan to include this graph in my thesis. This will give my thesis readers a strong explanation as to why I chose electric cars and no other types of electric vehicles as the subject of my analysis.

By visualizing the International Energy Agency (IEA) dataset, using the agency's records of historical data and forecasts of the number of electric vehicles over the next decade under the IEA Stated Policies Scenario (STEPS), I show the market compositions of the world's four major electric vehicle markets: China, Europe, India, and the United States, between 2010 and 2030.

For clarity, I have combined the categories other than cars: vans, buses, and trucks, because their combined numbers (in green) are far smaller than the share of cars alone.

Cars dominate vehicle electrification in the foreseeable future

Electric vans, buses, and trucks altogether account for little proportion of electric vehicles stock in key markets worldwide, 2010 – 2030



Data source:
International Energy Agency,
Global EV Data Explorer

The world's four major electric vehicle markets (China, Europe, India and the United States) are selected to demonstrate the domination of cars in the wave of vehicle electrification. According to the IEA forecast data, in the foreseeable future to 2030, cars are dominating the electric vehicle market.

01

Xiyu Zhang

2023-04-06

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

library(tidyverse)
library(ggplot2)

# Load data ------------------------------------------------------------------

EV_2010_to_2021 <-
  read_csv('data_own/IEA-EV-data.csv')

# Data wrangling -------------------------------------------------------------

EV_cars_quantity <-
  EV_2010_to_2021 %>%
  filter(region %in%
           c('China', 'Europe', 'USA', 'India')) %>%
  filter(year %in%
           seq(2010, 2030, 5)) %>%
  filter(category != 'Projection-APS') %>%
  filter(parameter %in%
           c('EV stock')) %>%
  mutate(mode_new =
           if_else(mode == 'Cars',
                   'Cars',
                   'Vans, buses & trucks')) %>%
  group_by(region, year, mode_new) %>%
  summarise(value_new =
              sum(value))

  # Set the categorical variables as factors with certain orders

EV_cars_quantity$mode_new <-
  factor(EV_cars_quantity$mode_new,
         levels =
           c('Vans, buses & trucks', 'Cars'))

EV_cars_quantity$region <-
  factor(EV_cars_quantity$region,
         levels =

           # in alphabetical order

           c('China', 'Europe', 'India', 'USA'))
```

```r
# Add percent numbers for reference

EV_cars_percent <-
  EV_cars_quantity %>%
  pivot_wider(names_from = mode_new,
              values_from = value_new) %>%
  mutate(percentage =
           formattable::percent(
             (`Vans, buses & trucks` /
                 (`Vans, buses & trucks` + `Cars`)))) %>%
  select(region, year, percentage)

# Combine the data together

EV_cars_all <-
  EV_cars_quantity %>%
  full_join(EV_cars_percent,
            by = join_by(region, year))

# Data visualization -------------------------------------------------------

p1 <-
  EV_cars_all %>%
  ggplot(mapping =
           aes(x = year,
               y = value_new,
               fill = mode_new)) +

  # only add some of the grid lines

  geom_hline(yintercept = 25000000,
             color = 'gray93') +
  geom_hline(yintercept = 50000000,
             color = 'gray93') +
  geom_hline(yintercept = 75000000,
             color = 'gray93') +
  geom_hline(yintercept = 100000000,
             color = 'gray93') +
  geom_bar(stat = 'identity',
           width = 3.2) +
  facet_wrap(~ region,
             nrow = 1) +
  geom_text(aes(label = percentage),
            size = 2.5,
            check_overlap = T) +
  scale_y_continuous(
    breaks = seq(0, 125000000, 25000000),
    labels = c('0', '25', '50', '75', '100', ''),
    limits = c(0, 125000000)) +
```

```
  scale_fill_manual(values =
                      c('#4daf4a', '#377eb8')) +
  labs(title = paste('Cars dominate vehicle electrification in the past,',
                     'present, and foreseeable future'),
       subtitle = paste('Electric vans, buses and trucks account for a little',
                        'proportion of electric vehicles in key markets worldwide'),
       caption = 'Data source: International Energy Agency',
       y = 'Million') +
  theme_minimal() +
  theme(axis.title.x = element_blank(),
        panel.grid = element_blank(),
        legend.position = 'none')
```

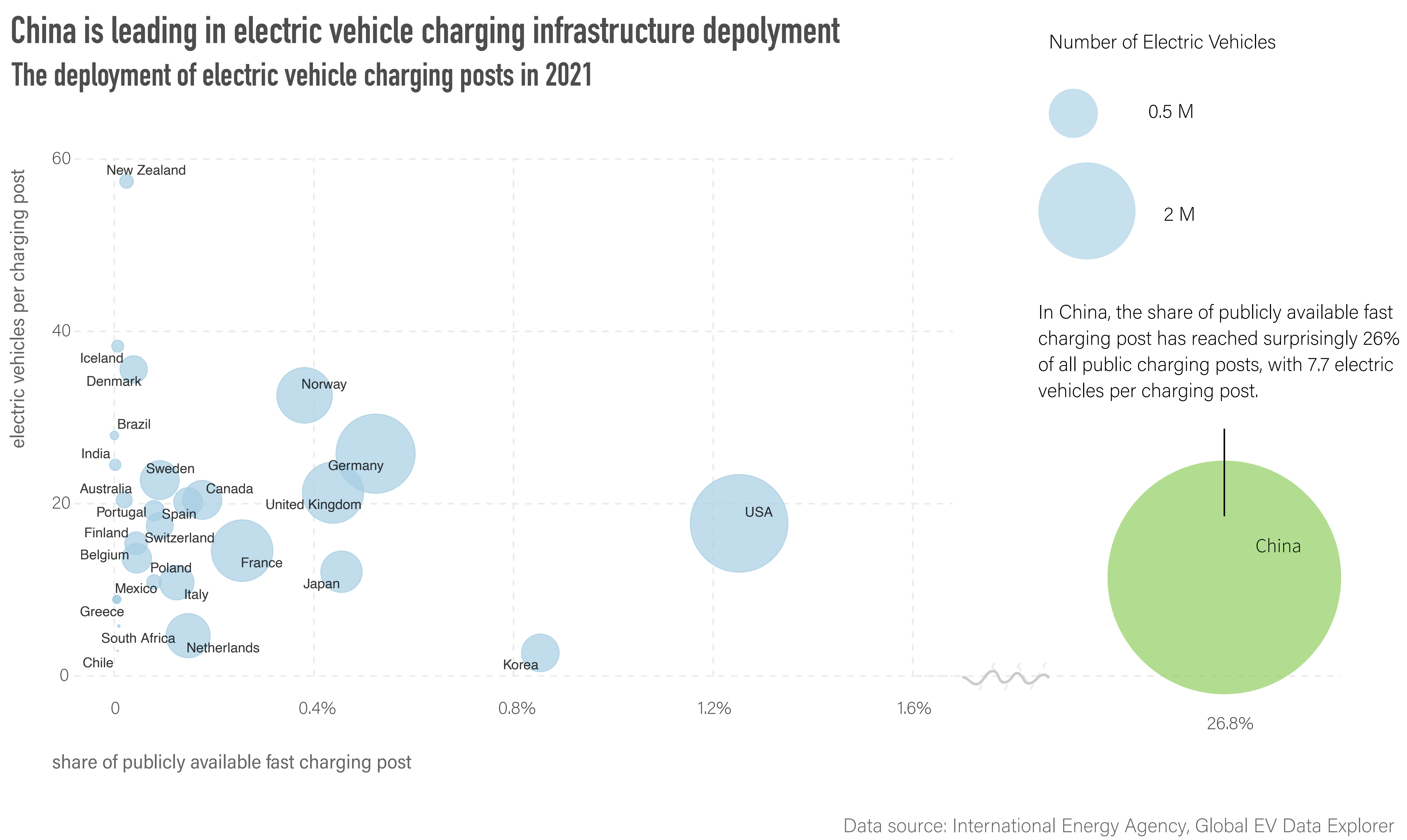
02

Statement of Purpose

The deployment of publicly available chargers for electric vehicles typically lags behind the growth of electric vehicles (IEA, 2022). That is, in today's electrification of transportation, people sometimes buy electric vehicles (EVs) first, and then find out that publicly available charging facilities are insufficient. The deployment of EV charging facilities is the policy issue I want to examine in my paper.

The purpose of this graph is to show the level of EV charging deployment in different countries. The horizontal axis depicts the share of publicly available fast-charging posts, and the publicly available charging posts consist of both "slow" and "fast" charging posts. Fast charging posts can reduce charging time per vehicle and serve more vehicle owners in a given time frame, i.e., provide more charging capacity. The larger this number is, the fewer public charging posts are available for each EV, and the closer to 0, the more charging posts are available for each vehicle. Finally, the size of the bubbles shows the current number of registered electric vehicles in a country and represents the demand for electric vehicle charging in that country.

China (26.8%) is far ahead of all other countries (less than 2%) on the horizontal axis, while also being close to zero on the vertical axis. At the same time, China has the largest bubble, indicating that it has the most electric vehicles. All this reflects the booming development of electric vehicles in China.



China is leading in electric vehicle charging infrastructure depolyment
The deployment of electric vehicle charging posts in 2021
Number of Electric Vehicles
0.5 M
2 M
In China, the share of publicly available fast charging post has reached surprisingly 26% of all public charging posts, with 7.7 electric vehicles per charging post.
electric vehicles per charging post
60
40
20
0
New Zealand
Iceland
Denmark
Norway
Brazil
India
Sweden
Germany
Australia
Canada
United Kingdom
USA
Portugal
Spain
Finland
Switzerland
Belgium
France
Poland
Japan
Mexico
Italy
Greece
South Africa
Netherlands
Chile
Korea
China
0
0.4%
0.8%
1.2%
1.6%
26.8%
share of publicly available fast charging post
Data source: International Energy Agency, Global EV Data Explorer

02

Xiyu Zhang

2023-04-06

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

library(tidyverse)
library(ggplot2)
library(colorspace)

# Load data --------------------------------------------------------------------

EV_2010_to_2021 <-
  read_csv('data_own/IEA-EV-data.csv')

# Data wrangling ---------------------------------------------------------------

# Create a initial data set

EV_initial <-
  EV_2010_to_2021 %>%
  filter(!region %in%

           # exclude the regional data, in order to focus on countries

           c('Europe', 'Rest of the world', 'Other Europe', 'World')) %>%
  filter(parameter %in% c('EV stock', 'EV charging points'),
         year == 2021,
         category == 'Historical')

# Number of electric vehicles per charging point

EV_per_charging <-
  EV_initial %>%
  group_by(region, parameter) %>%
  summarise(value_new =
              sum(value)) %>%
  pivot_wider(
    names_from = parameter,
    values_from = value_new) %>%
  mutate(EV_per_charger =
           `EV stock` / `EV charging points`) %>%
  select(region, EV_per_charger)

# The share of fast chargers in the total number of chargers

Fast_charger_share <-
```

```r
  EV_initial %>%
  filter(parameter == 'EV charging points') %>%
  select(region, powertrain, value) %>%
  pivot_wider(
    names_from = powertrain,
    values_from = value) %>%
  mutate(fast_percentage =
           `Publicly available fast` /
           sum(`Publicly available slow`, `Publicly available fast`)) %>%
  mutate(fast_percentage =
           formattable::percent(fast_percentage)) %>%
  select(region, fast_percentage)

# The total number of electric vehicles

EV_number <-
  EV_initial %>%
  filter(parameter == 'EV stock') %>%
  group_by(region) %>%
  summarise(EV_total = sum(value))

# Combine the created variables into one data frame

EV_analysis <-
  EV_per_charging %>%
  full_join(Fast_charger_share) %>%
  full_join(EV_number) %>%

  # exclude the missing values

  filter(!is.na(EV_per_charger)) %>%

  # to make visualization neater, arrange the order

  arrange(EV_number) %>%

  # we found that the fast charger percentage for China is an outlier, since
  # China's percentage of fast chargers are way higher than other countries
  # Exclude China for now

  filter(region != 'China')

# Data visualization -----------------------------------------------------

# generate a color palette

palette <-
  rainbow_hcl(27)

# create a bubble chart
```

```r
p2 <-
  EV_analysis %>%
  ggplot() +
  geom_point(
    aes(x = EV_per_charger,
        y = fast_percentage,
        size = EV_total,
        color = palette)) +
  scale_size(range = c(.1, 30),
             name = 'Number of Electric Vehicles') +
  coord_flip() +
  geom_text(
    aes(
      x = EV_per_charger,
      y = fast_percentage,
      label = region),
    hjust = 'left',
    size = 3,
    check_overlap = T) +
  scale_x_continuous(
    breaks = seq(0, 60, 20)) +
  scale_y_continuous(
    breaks = seq(0, 0.016, 0.004),
    labels = c('0', '0.4%', '0.8%', '1.2%', '1.6%'),
    limits = c(0, 0.016)) +
  labs(title = paste('China is leading in EV charging infrastructure depolyment,',
                     'follows by the US and Korea'),
       subtitle = 'The deployment of EV charging posts in 2021',
       caption = 'Data source: International Energy Agency',
       x = 'electric vehicles per charging post',
       y = 'share of publicly available fast charging post') +
  theme_minimal() +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_line(linetype = 'dashed'),
        legend.position = 'none')
```

02

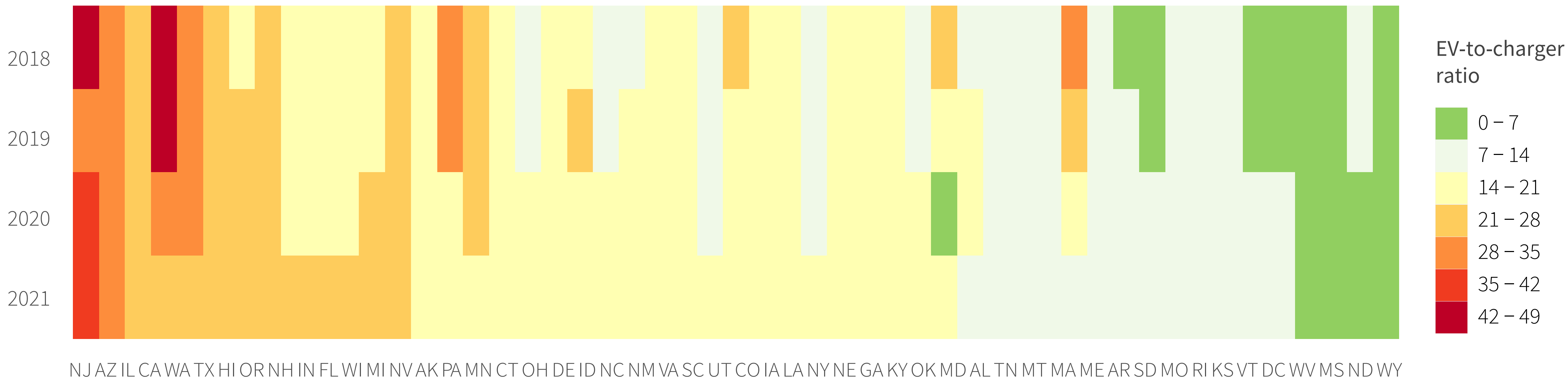
Statement of Purpose

This data set is almost identical to the one generated in the previous section, demonstrating the change in electric vehicle (EV) to charger ratios across U.S. states from 2018 to 2021, with only a few adjustments to fit the graph type better.

However, while a set of choropleth maps are mainly indicating to readers the general trend of the decrease of EV-to-charger ratio across U.S. states, the heat map here focuses on the spectrum of different levels of EV charging infrastructure deployment across U.S. states. On the most left there is New Jersey, with the highest EV-to-charger ratio in 2021, while on the most right there is Wyoming with the lowest ratio. From this aspect, readers could see that New Jersey, Arizona, Illinois, and California are among the states that need EV charging infrastructure the most, while Wyoming, North Dakota, and Mississippi are the states that need the infrastructure the least.

Electric Vehicle (EV) Charging Point Deployment by State in the U.S., 2018 – 2021

From left to right: the state with the least (NJ) to most (WY) adequate deployment of EV charging points



From left to right are the states with the least (NJ) to most (WY) adequate deployment of EV charging points.
From 2018 to 2021, every state in the U.S. was committed to deploying electric vehicle charging infrastructure.
The EV-to-charger ratio decreases in most states.
However, most states still do not meet the Alternative Fuel Infrastructure Directive's (AFID) recommendation to reduce the EV-to-charger ratio to 10 by 2020 (white and green tiles in the chart). A floating standard (ratio 7 - 14 is defined as "meet the standard" with white tiles) rather than fixed standard (rigorously under 10) is used for visualization consideration, since there are only few states meet the threshold of 10.

Data Source:
Alternative Fuels Data Center:
Vehicle Registration Counts by State
Alternative Fueling Station Counts by State

```r
# Load packages ----------------------------------------------------------

library(tidyverse)
library(ggplot2)
library(readxl)

# Load data --------------------------------------------------------------

# Stations

stations <-
  read_csv('assignment5/alt_fuel_stations.csv') %>%

  # ON is a Canada state, exclude non-US states.

  filter(State != 'ON')

# EV registration

LDV_2021 <-
  read_excel('assignment5/2021_LDV_registration.xlsx')

LDV_2020 <-
  read_excel('assignment5/2020_LDV_registration.xlsx')

LDV_2019 <-
  read_excel('assignment5/2019_LDV_registration.xlsx')

LDV_2018 <-
  read_excel('assignment5/2018_LDV_registration.xlsx')

# State name and code

state <-
  read_csv('assignment5/state.csv') %>%
  select(-'abbrev') %>%
  rename('State' = 'state')

# Data wrangling ---------------------------------------------------------

# 1. Repeat the process to generate the 'All_years' data frame as same as
# the first graph

# Electric vehicle supply equipment (EVSE, charging points) in 2021 by state

charger_2021 <-
  stations %>%
  filter(year(`Open Date`) < 2022) %>%
  filter(State != 'ON') %>%
  select(`Station Name`, `EV Level1 EVSE Num`, `EV Level2 EVSE Num`,
         `EV DC Fast Count`, State) %>%
  replace(is.na(.), 0) %>%
```

```r
  mutate(EVSE = `EV Level1 EVSE Num` + `EV Level2 EVSE Num` +
           `EV DC Fast Count`) %>%
  group_by(State) %>%
  summarise(EVSE_state = sum(EVSE))

# the Plug-in Electric Vehicles in 2021 by state

PEV_2021 <-
  LDV_2021 %>%
  filter(State != 'United States') %>%
  mutate(PEV = `Electric (EV)` + `Plug-In Hybrid Electric (PHEV)`) %>%
  select(State, PEV)

# Merge the needed data for 2021

merge_2021 <-
  full_join(state, PEV_2021) %>%
  select(code, PEV) %>%
  rename(State = code) %>%
  full_join(charger_2021) %>%
  mutate(EV_per_charger = PEV / EVSE_state) %>%
  select(State, EV_per_charger) %>%
  mutate(year = 2021)

# Electric vehicle supply equipment (EVSE, charging points) in 2020 by state

charger_2020 <-
  stations %>%
  filter(year(`Open Date`) < 2021) %>%
  filter(State != 'ON') %>%
  select(`Station Name`, `EV Level1 EVSE Num`, `EV Level2 EVSE Num`,
         `EV DC Fast Count`, State) %>%
  replace(is.na(.), 0) %>%
  mutate(EVSE = `EV Level1 EVSE Num` + `EV Level2 EVSE Num` +
           `EV DC Fast Count`) %>%
  group_by(State) %>%
  summarise(EVSE_state = sum(EVSE))

# the Plug-in Electric Vehicles in 2020 by state

PEV_2020 <-
  LDV_2020 %>%
  filter(State != 'United States') %>%
  mutate(PEV = `Electric (EV)` + `Plug-In Hybrid Electric (PHEV)`) %>%
  select(State, PEV)

# Merge the needed data for 2020

merge_2020 <-
  full_join(state, PEV_2020) %>%
  select(code, PEV) %>%
  rename(State = code) %>%
```

```r
  full_join(charger_2020) %>%
  mutate(EV_per_charger = PEV / EVSE_state)%>%
  select(State, EV_per_charger) %>%
  mutate(year = 2020)

# Electric vehicle supply equipment (EVSE, charging points) in 2019 by state

charger_2019 <-
  stations %>%
  filter(year(`Open Date`) < 2020) %>%
  filter(State != 'ON') %>%
  select(`Station Name`, `EV Level1 EVSE Num`, `EV Level2 EVSE Num`,
         `EV DC Fast Count`, State) %>%
  replace(is.na(.), 0) %>%
  mutate(EVSE = `EV Level1 EVSE Num` + `EV Level2 EVSE Num` +
           `EV DC Fast Count`) %>%
  group_by(State) %>%
  summarise(EVSE_state = sum(EVSE))

# the Plug-in Electric Vehicles in 2019 by state

PEV_2019 <-
  LDV_2019 %>%
  filter(State != 'United States') %>%
  mutate(PEV = `Electric (EV)` + `Plug-In Hybrid Electric (PHEV)`) %>%
  select(State, PEV)

# Merge the needed data for 2019

merge_2019 <-
  full_join(state, PEV_2019) %>%
  select(code, PEV) %>%
  rename(State = code) %>%
  full_join(charger_2019) %>%
  mutate(EV_per_charger = PEV / EVSE_state)%>%
  select(State, EV_per_charger) %>%
  mutate(year = 2019)

# Electric vehicle supply equipment (EVSE, charging points) in 2018 by state

charger_2018 <-
  stations %>%
  filter(year(`Open Date`) < 2019) %>%
  filter(State != 'ON') %>%
  select(`Station Name`, `EV Level1 EVSE Num`, `EV Level2 EVSE Num`,
         `EV DC Fast Count`, State) %>%
  replace(is.na(.), 0) %>%
  mutate(EVSE = `EV Level1 EVSE Num` + `EV Level2 EVSE Num` +
           `EV DC Fast Count`) %>%
  group_by(State) %>%
  summarise(EVSE_state = sum(EVSE))
```

```
# the Plug-in Electric Vehicles in 2018 by state

PEV_2018 <-
  LDV_2018 %>%
  filter(State != 'United States') %>%
  mutate(PEV = `Electric (EV)` + `Plug-In Hybrid Electric (PHEV)`) %>%
  select(State, PEV)

# Merge the needed data for 2018

merge_2018 <-
  full_join(state, PEV_2018) %>%
  select(code, PEV) %>%
  rename(State = code) %>%
  full_join(charger_2018) %>%
  mutate(EV_per_charger = PEV / EVSE_state)%>%
  select(State, EV_per_charger) %>%
  mutate(year = 2018)

# Merge the data of different years

All_years <-
  rbind(merge_2018,
        merge_2019) %>%
  rbind(merge_2020) %>%
  rbind(merge_2021)

# 2. Minor adjustments

All_years_heat <-
  All_years %>%
  filter(!is.na(EV_per_charger))

# factorize the state vector

level <-
  All_years_heat %>%
  filter(year == 2021) %>%
  arrange(desc(EV_per_charger)) %>%
  pull(State)

All_years_heat$State <-
  factor(All_years_heat$State, levels = level)

# Reset the discrete intervals that suit heat map better

All_years_heat <-
  All_years_heat %>%
  mutate(EV_per_Charger =
           case_when(EV_per_charger <= 7 ~ '0 - 7',
                     EV_per_charger <= 14 ~ '7 - 14',
                     EV_per_charger <= 21 ~ '14 - 21',
```

```
                     EV_per_charger <= 28 ~ '21 - 28',
                     EV_per_charger <= 35 ~ '28 - 35',
                     EV_per_charger <= 42 ~ '35 - 42',
                     .default = '42 - 49'))

All_years_heat$EV_per_Charger <-
  factor(All_years_heat$EV_per_Charger,
         levels = c('0 - 7', '7 - 14', '14 - 21', '21 - 28', '28 - 35',
                    '35 - 42', '42 - 49'))

# Heat map ----------------------------------------------------------------

heat_map <-
  All_years_heat %>%
  ggplot(mapping =
           aes(x = State,
               y = year)) +
  geom_tile(
    aes(fill = EV_per_Charger)) +

  # the mid point is EV-to-charger ratio equals to 10, which is the recommended
  # level of EV charger deployment by the Alternative Fuel Infrastructure
  # Directive (AFID).

  scale_fill_manual(values =
                      c('#91cf60', '#f0f9e8', '#ffffb2',
                        '#fecc5c', '#fd8d3c', '#f03b20', '#bd0026')) +
  labs(title = paste('Electric Vehicle Charging Point Deployment by State in',
                     'the U.S., 2018 - 2021'),
       subtitle = paste('From left to right: the state with the least (NJ) to',
                        'most (WY) adequate deployment of EV charging points'),
       caption = 'Data Source: Alternative Fuels Data Center') +
  theme(axis.title = element_blank(),
        axis.ticks = element_blank(),
        panel.background = element_blank())
```

08 Statement

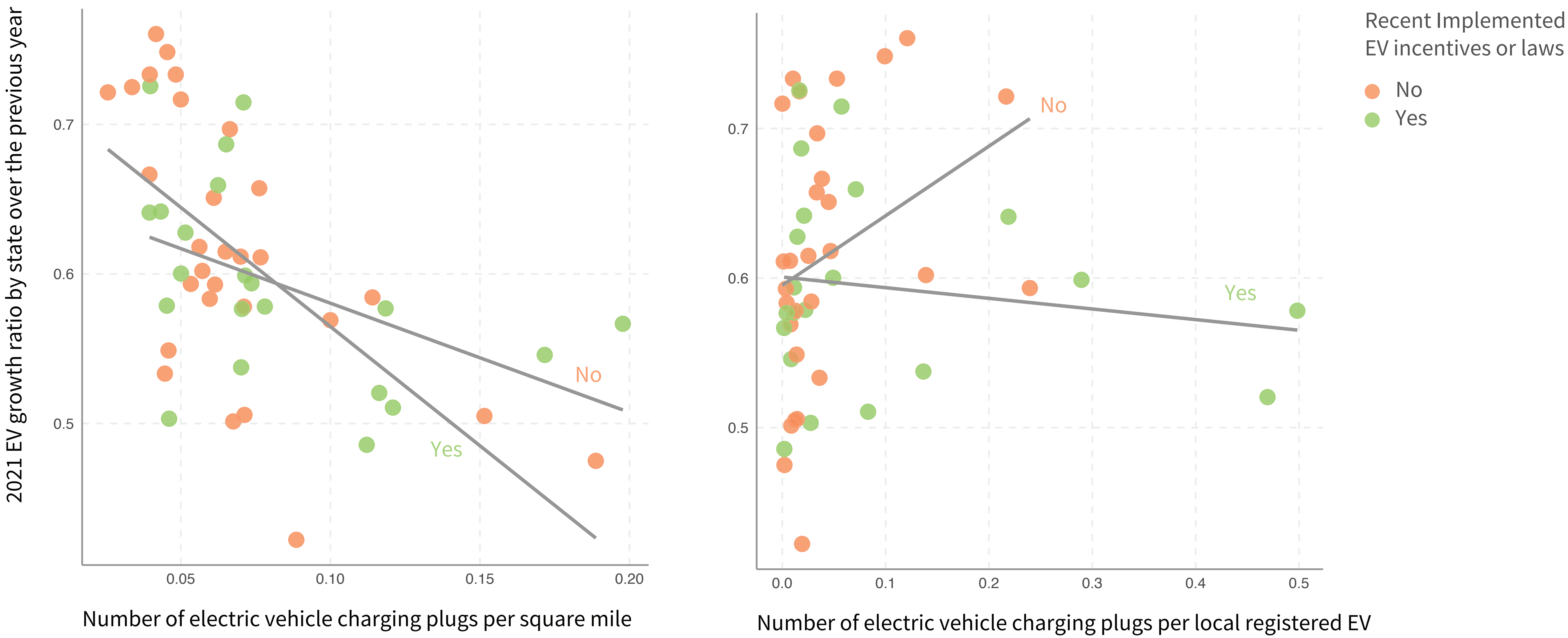
Numerous factors influence a resident's decision to choose an electric vehicle over a conventional one. Merely expanding the number of electric vehicle chargers available may not necessarily stimulate the sale of electric vehicles. The demographics of the local population, the travel characteristics of the local population, and policy scenarios are just as important as the ease of access to EV charging stations in determining EV sales.

This visualization is generated to demonstrate that generally, states without recent policy encouragement benefit from constructing more EV chargers.

The reason behind the inadequate impact of expanding electric vehicle charging stations could be attributed to the fact that states that have already implemented EV incentives have more abundant EV charging posts. Consequently, there are already more charging plugs available for each electric vehicle, resulting in a lower possibility of residents experiencing power shortages. In such circumstances, the development of new EV charging facilities provides little motivation for residents to opt for electric vehicles.

Only in states that have not yet implemented pro-electric vehicle policies, the increase in EV chargers is positively correlated with the growth in the number of EVs

Electric vehicle increase statistics, 2021



Source: Alternative Fuel Data Center: Alternative Fuel Stations, EV registration, Laws and Incentives

```
# Packages ---------------------------------------------------------------

library(sf)
library(tidyverse)
library(readxl)
library(writexl)

# Load Data -------------------------------------------------------------

# 1. stations

stations <-
  read_csv('DataStory/alt_fuel_stations.csv')

# 2. EV stocks

# stocks

EV_registration_by_state <-
  read_excel('DataStory/ev-registration-counts-by-state.xlsx') %>%
  select(State, `Registration Count`) %>%
  filter(State != 'Total')

# registration

LDV_registration_2021 <-
  read_excel('DataStory/2021_LDV_registration.xlsx') %>%
  filter(State != 'United States')

# 3. Housing and other demography data

DMV_housing_temp <-
  read_xlsx('DataStory/DMV_housing.xlsx')

DMV_housing <-
  t(DMV_housing_temp[-1])

colnames(DMV_housing) <- select(DMV_housing_temp, `vehicles available`)

# basic states information

data(state)
```

```
state <-
   cbind(state.name, state.abb, state.area) %>%
   as_tibble() %>%
   rename(State = state.name,
              abb = state.abb,
              area = state.area)

# 4. laws and incentives

EV_law <-
   read_csv('DataStory/EV_law.csv') %>%
   filter(Recent == TRUE,
              State != 'US') %>%
   group_by(State) %>%
   summarise(EV_law = n()) %>%
   rename(state.abb = State)

# PHEV_law is useless since EV_law overlaps with it

PHEV_law <-
   read_csv('DataStory/PHEV_law.csv') %>%
   filter(`Recent?` == TRUE,
              State != 'US') %>%
   group_by(State) %>%
   summarise(PHEV_law = n()) %>%
   rename(state.abb = State)

# states that have incentives or not

state_w_incentive <-
   cbind(state.abb, state.name) %>%
   as_tibble() %>%
   full_join(EV_law) %>%
   full_join(PHEV_law) %>%
   rename(State = state.name,
              abb = state.abb) %>%
   mutate(EV_law_B = if_else(!is.na(EV_law),
                                          T,
                                          F)) %>%
   mutate(PHEV_law_B = if_else(!is.na(PHEV_law),
                                             T,
                                             F))

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

```
# 1. charger number by state

charger_by_state <-
  stations %>%

  # the open date is ealier than 2022, so this is the existence of chargers in
  # 2021

  filter(year(`Open Date`) < 2022) %>%

  # 'ON' is in Canada

  filter(State != 'ON') %>%
  select(`Station Name`, `EV Level1 EVSE Num`, `EV Level2 EVSE Num`,
         `EV DC Fast Count`, State) %>%

  # 'NA' indicates that there is no this type of EVSE in the station, so there
  # is 0 this type of EVSE in the station

  replace(is.na(.), 0) %>%

  # the total number of charging points in a charging station is the sum of
  # 3 types of EVSEs

  mutate(EVSE = `EV Level1 EVSE Num` + `EV Level2 EVSE Num` +
           `EV DC Fast Count`) %>%

  # calculate the total number of EVSE in each state in 2021

  group_by(State) %>%
  summarise(EVSE_state = sum(EVSE))

# 2. increase of EV

# EV count

LDV_2021 <-
  LDV_registration_2021 %>%
  transmute(
    State = State,
    EV = `Electric (EV)`,
    PHEV = `Plug-In Hybrid Electric (PHEV)`) %>%
  mutate(count = EV + PHEV)
```

```
# percentage increase

EV_percentage <-
  LDV_2021 %>%
  full_join(EV_registration_by_state,
            by = join_by(State)) %>%
  mutate(percent = `Registration Count` / count)

# Increase & charger count ---------------------------------------------

# An analysis based on 2021 data

increase_and_charger <-
  charger_by_state %>%
  rename(abb = State) %>%
  full_join(state,
            by = join_by(abb)) %>%
  full_join(EV_percentage,
            by = join_by(State)) %>%

  # combine the DC values into one row, then discard the previous rows
  # containing NA values

  rbind(c('DC', 798, 'District of Columbia', 68, 3700, 2500, 6200, 3700,
          0.596774193548387)) %>%
  na.omit()

# calculate the charger density in each state

increase_and_charger$EVSE_state <- as.numeric(increase_and_charger$EVSE_state)
increase_and_charger$area <- as.numeric(increase_and_charger$area)
increase_and_charger$percent <- as.numeric(increase_and_charger$percent)
increase_and_charger$count <- as.numeric(increase_and_charger$count)

# scatter plot with a fit line

increase_and_charger <-
  increase_and_charger %>%
  mutate(charger_density = EVSE_state / area)

# 1. charger density by area

p1 <-
```

```
  increase_and_charger %>%

  # exclude the outlier for a better visualization

  filter(abb != 'DC') %>%
  ggplot(aes(x = charger_density,
             y = percent)) +
  geom_point(color = 'gray',
             size = 4) +
  stat_smooth(method = 'lm',
              formula = y ~ x,
              se = F,
              color = '#4dac26') +
  theme_bw() +
  theme(axis.line = element_line(colour = '#969696'),
        axis.ticks = element_blank(),
        panel.grid.minor = element_blank(),
        panel.grid.major = element_line(linetype = 'dashed'),
        panel.border = element_blank())

# 2. charger density by EV count

increase_and_charger <-
  increase_and_charger %>%
  mutate(charger_density = EVSE_state / count)

p2 <-
  increase_and_charger %>%

  # exclude the outlier for a better visualization

  filter(abb != 'DC') %>%
  ggplot(aes(x = charger_density,
             y = percent)) +
  geom_point(color = 'gray',
             size = 4) +
  stat_smooth(method = 'lm',
              formula = y ~ x,
              se = F,
              color = '#4dac26') +
  theme_bw() +
  theme(axis.line = element_line(colour = '#969696'),
        axis.ticks = element_blank(),
        panel.grid.minor = element_blank(),
```

```
        panel.grid.major = element_line(linetype = 'dashed'),
        panel.border = element_blank())

# Scatter plot with 2 fit lines ------------------------------------------

# consider incentives for EVs and PHEVs

plot2_data <-
  increase_and_charger %>%
  left_join(state_w_incentive)

p3 <-
  plot2_data %>%
  filter(abb != 'DC') %>%
  ggplot() +
  geom_point(aes(x = charger_density,
                 y = percent,
                 color = EV_law_B),
             size = 4,
             alpha = 0.8) +
  scale_color_manual(values = c('#fc8d59', '#91cf60')) +
  geom_smooth(aes(x = charger_density,
                  y = percent,
                  fill = EV_law_B),
              method = 'lm',
              formula = y ~ x,
              se = F,
              color = '#969696') +
  theme_bw() +
  theme(axis.line = element_line(colour = '#969696'),
        axis.ticks = element_blank(),
        panel.grid.minor = element_blank(),
        panel.grid.major = element_line(linetype = 'dashed'),
        panel.border = element_blank())

# p4

p4 <-
  plot2_data %>%
  filter(abb != 'DC') %>%
  ggplot() +
  geom_point(aes(x = charger_density,
                 y = percent,
                 color = EV_law_B),
```

```
                      size = 4,
                      alpha = 0.8) +
    scale_color_manual(values = c('#fc8d59', '#91cf60')) +
    geom_smooth(aes(x = charger_density,
                              y = percent,
                              fill = EV_law_B),
                        method = 'lm',
                        formula = y ~ x,
                        se = F,
                        color = '#969696') +
    theme_bw() +
    theme(axis.line = element_line(colour = '#969696'),
              axis.ticks = element_blank(),
              panel.grid.minor = element_blank(),
              panel.grid.major = element_line(linetype = 'dashed'),
              panel.border = element_blank())

# Map -------------------------------------------------------------------

write_xlsx(plot2_data, 'DataStory/EV_map.xlsx')
```

02

Statement of Purpose

The electric vehicle market is segmented into four distinct categories: cars, buses, vans, and trucks. China experienced a substantial surge in the stock of electric vehicles from 2019 to 2021, with cars being the primary contributor to this growth. The purpose of this visualization is to illustrate the dominance of cars in this expansion.

Note

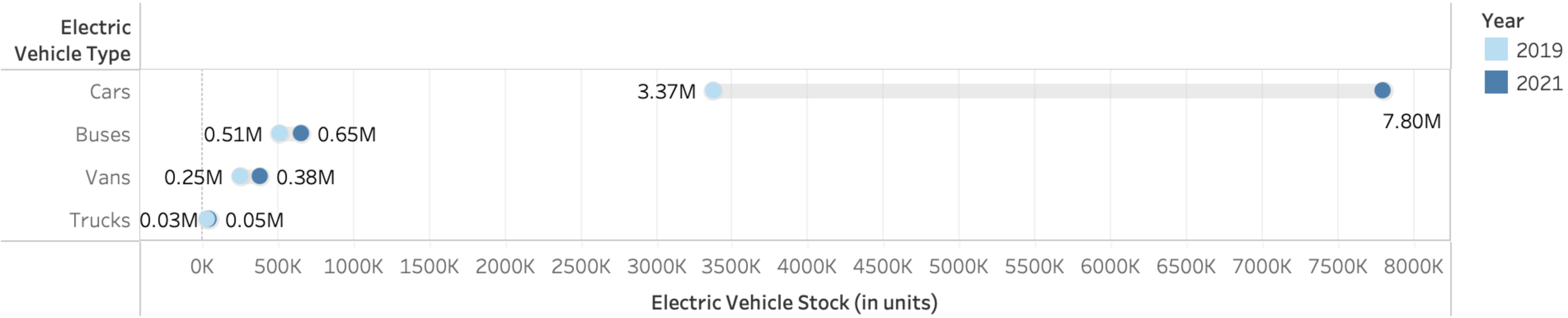
1. Data comes from International Energy Agency (IEA). The tidy format of data imported in Tableau has been processed in R, with the R code below.

```r
library(readr)
library(tidyverse)
library(writexl)

# data wrangling

temp_2 <-
  read_csv('data_own/IEA-EV-data.csv') %>%
  filter(region %in% c('China')) %>%
  filter(year %in% c(2019, 2021)) %>%
  filter(parameter == 'EV stock',
         category == 'Historical') %>%
  select(mode, year, value)

# export the data, and use it in Tableau

write_xlsx(temp_2, 'data_own/assigment6_chinaEV_19to21.xlsx')

```

2. My Tableau software appears to be unable to export a PDF file, which makes it impossible for me to refine the generated image using Adobe Illustrator. When I attempted to select the "Print" option, I received a message stating that no printer is installed. Thus, I am unable to select the option to print to PDF.

The category of "cars" has shown the most rapid growth in the stock of electric vehicles in China

A comparison of China's electric vehicle inventory across categories in 2019 and 2022



The electric vehicle market is segmented into four distinct categories: cars, buses, vans, and trucks. China witnessed a considerable upsurge in the stock of electric vehicles from 2019 to 2021, with cars accounting for the majority of this growth.

Data Source: International Energy Agency, Global EV Outlook 2022

01

Statement of Purpose

Since 1995, the initial recorded year, there has been a surge in the construction of electric vehicle charging stations in every state of the United States, with California boasting the highest number. This map showcases a visualization of the distribution of electric vehicle (EV) direct current (DC) fast charging stations built in each state over time, which represents one of the most advanced and fastest charging ports available.