

Check-in 1

Final Project Update #1

Dan Pacheco - IST 687 | Feb. 16, 2025

For my final project, I want to do an analysis of the electric vehicle charging infrastructure in the United States. I am taking more of a journalistic approach with this than a “business” approach, because journalism is my business.

The purpose of this analysis is to give people who are considering “going electric,” but are holding back, an idea of how their experience may differ from what they are used to with a combustible engine vehicle. For an electric-car driver, the small but growing infrastructure in charging stations in the United States has real-world implications for things like city-to-city trips, cross-country trips and the like. The charging infrastructure is growing differently state by state, so “going electric” can influence where you may be comfortable driving, or not. Finally, the rate at which charging infrastructure is changing also varies state by state.

The data set I will analyze comes from the [Department of Energy’s Alternative Fueling Station Locator](#) site. It has data for all types of alternative fueling stations, including hydrogen, biodiesel and more. But the data I’m interested in is electric. I have already downloaded a subset of electric stations and saved a copy of it in [Google Sheets](#). I could not find a way to import directly into R from the DOE site, but I also have been concerned that data could be deleted – as the Trump administration has been deleting anything related to climate change policy as it seeks to take the country back to a fossil-fuel economy.

While analyzing the data set (see my [investigation notes](#), also available in the second tab at left), I have found several columns that will be relevant for this analysis. They fall into these buckets:

- Location:
 - **City, State, Latitude and Longitude** are most useful. These will allow me to create maps. I also hope that I can use that data to calculate distances between charging spots (need to find out how to do that in R, maybe there is a package for that?)
- Charging speed:
 - **ev_level1_evse_num:** Slowest, mainly for overnight charging at home.
 - **ev_level2_evse_num:** Faster and common in workplaces and public areas
 - **ev_dc_fast_num:** Fastest, ideal for road trips or quick top-ups
- Connector types
 - **ev_connector_types:** Determines if your car can use it. J1772, CCS, CHAdeMO, Tesla
- Pricing:
 - **ev_pricing:** Info on whether the station charges a fee and how much.
- Open Date: I think I could use this to calculate adoption by state and even by city.
- Cost: This column has values that are all over the map, but maybe there is some data that could indicate cost to be in the location (but not to charge). I do notice that “free” is a value.
- Facility Type: Probably the least useful, but could be interesting as it tells you if you are likely to be charging in a parking lot, shopping center, library, museum, and so forth. If charging takes a while, this could tell you what else you could do while waiting for a charge.

Check-in 2

Final Project Update #2

Dan Pacheco - IST 687 | Mar. 4, 2025

I have made quite a lot of progress on my final project in R. Here's an overview of the data I have used so far, what I've done to clean it up, and what I've done to answer some initial questions.

Importing and Cleanup of Primary Data Set

I started out by importing a CSV of 83,000 electric vehicle charging stations which comes from the Department of Energy's Alternative Charging Station Locator web site. I archived a copy of it as a [CSV in Google](#) and exported the CSV, then imported it into R using the jsonlite library.

I then used tidyverse to rename the columns that I want to use, and used subsetting to remove the first row which duplicated column names. I discovered that the data set included states, Canadian provinces, and non-state areas like DC and territories like Puerto Rico. I decided to remove these as I am focusing on U.S. states, though I did keep the Canadian data because I could perhaps do some comparisons later between two countries.

I noted that the data has latitude, longitude and state name columns. These became helpful later.

Two Additional Data Sets

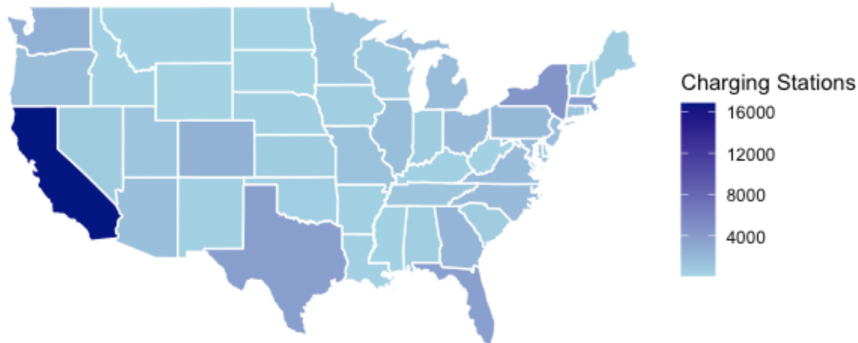
Because I wanted to map things by state, I also loaded these two additional data sets:

- 1) The "states" data from library(maps). I standardized the state name to a field called "region" that is all lower case with the state fully spelled out (e.g. "alabama" vs. "Alabama").
- 2) Another DOE data set that lists the [total number of vehicle registrations](#) by type in each state. I needed this later in order to show things like how many charging stations there are for number of electric cars by state.

Map1: Total Charging Stations by State

I used Ggplot to create this map that shows number of charging stations by state. California shows up as a clear leader / outlier and it throws off the contrast for every other state:

Total Charging Stations by State



To get this map I had to do a lot of munging and joining. Specifically:

Count charging stations by state

```
state_counts <- fuelstations_us %>%  
  group_by(StateAbbrev) %>%  
  summarise(StationCount = n()) %>%  
  ungroup()
```

Compute national average

```
national_avg <- mean(state_counts$StationCount, na.rm = TRUE)
```

Calculate percentage difference

```
state_counts <- state_counts %>%  
  mutate(PercentageDiff = ((StationCount - national_avg) / national_avg) * 100)
```

Load built-in state map data

```
states_map <- map_data("state")
```

Convert state abbreviations to full names (needed for mapping)

```
state_abbrev_to_name <- data.frame(  
  StateAbbrev = state.abb,  
  region = tolower(state.name)  
)
```

Merge state_counts with full state names

```
state_data <- state_counts %>%  
  left_join(state_abbrev_to_name, by = "StateAbbrev")
```

Merge charging station data with state map data

```
map_data_final <- left_join(states_map, state_data, by = "region")
```

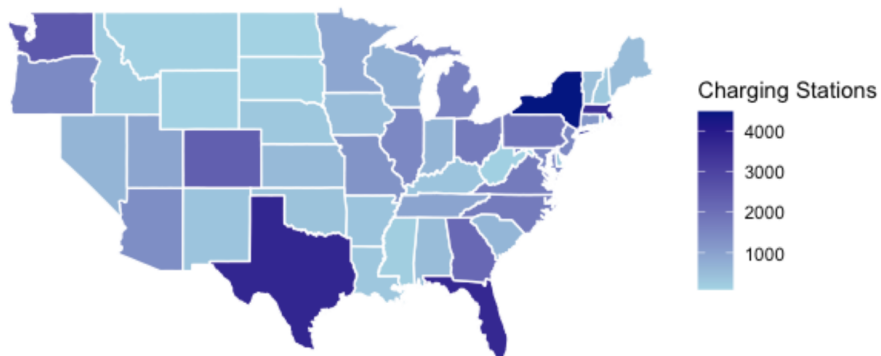
Create the choropleth map with ggplot2

```
ggplot(data = map_data_final, aes(x = long, y = lat, group = group, fill = StationCount)) +  
  geom_polygon(color = "white") +          # Draw state borders in white  
  coord_fixed(1.3) +                      # Fix aspect ratio  
  scale_fill_gradient(low = "lightblue", high = "darkblue",  
    na.value = "grey50", name = "Charging Stations") +  
  theme_void() +                          # Remove background, grid, etc.  
  labs(title = "Total Charging Stations by State")
```

Map 2: Total Charging Stations by State, Excluding California

This is the same map, but with California removed so that we can see how all the other states compare to each other. Texas, New York and Florida and Massachusetts are clear leaders. However, it's worth noting they are also leaders in population so this may not be the best way to compare electric vehicle charging infrastructure.

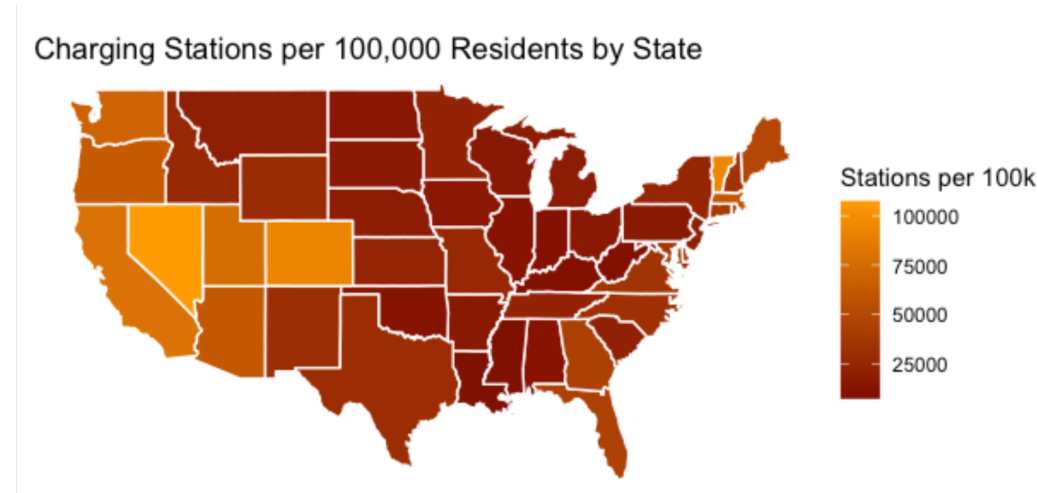
Total Charging Stations by State (Excluding CA)



This was easier than map 1, as I just had to filter out California before drawing the map. Then I used ggplot the same way:

```
map_data_no_ca <- map_data_final %>%  
  filter(region != "california")
```

Map 3: Showing Charging Stations per State Population



This map is really interesting. It shows that the western states (Washington, Oregon, California, Nevada, Utah, Arizona and Colorado) do a significantly better job at providing charging options for the people in their states than most of the other states. In the Northeast, only Vermont and Massachusetts appear to be on par with them. The rest, including my own New York state, are really lagging behind.

Code for this map:

We need to see these in terms of per capita population of each state.

1. Create a new dataframe for state population using state.x77

```
new_state_population <- data.frame(  
  region = tolower(rownames(state.x77)),  
  Population = state.x77["Population"]  
)
```

2. Merge your existing state_data (which has StationCount and region) with the new population data,

creating a new dataframe that includes the per capita metric.

```
new_state_data <- state_data %>%  
  left_join(new_state_population, by = "region") %>%  
  mutate(StationsPer100k = (StationCount / Population) * 100000)
```

3. Create a new map data frame by merging the built-in states map with the new state data

```
new_map_data_final <- left_join(states_map, new_state_data, by = "region")
```

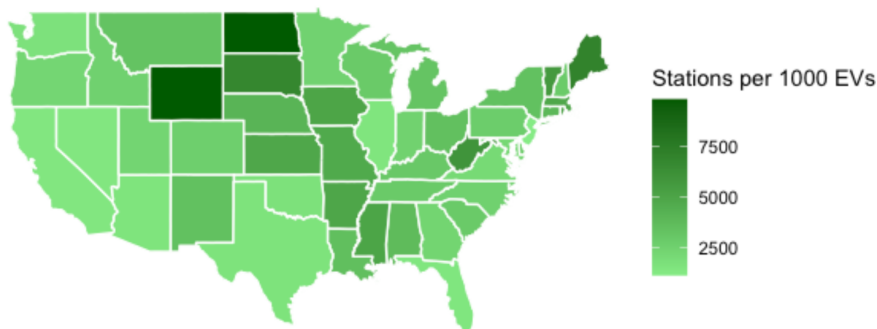
4. Plot the choropleth map using the new map data (charging stations per 100k residents)

```
ggplot(data = new_map_data_final, aes(x = long, y = lat, group = group, fill = StationsPer100k)) +  
  +  
  geom_polygon(color = "white") +  
  coord_fixed(1.3) +  
  scale_fill_gradient(low = "darkred", high = "orange",  
    na.value = "grey50", name = "Stations per 100k") +  
  theme_void() +  
  labs(title = "Charging Stations per 100,000 Residents by State")
```

Map 4: Showing Charging Stations per Number of Electric Cars

However, is chargers per person the best measure? People don't get electric charging, electric cars do. The more I thought about this, the number of chargers per state only makes sense when you consider how many electric cars they need to serve. I chose to compare the number of chargers per electric car in each state. This isn't perfect, as it assumes each state is in its own bubble when in fact people drive between state lines all the time. But it gives us at least a slightly better idea of the electric car charging experience as you move across the country.

Charging Stations per 1,000 Electric Vehicles by State



This is the most surprising map of all, because it shows that the western states are actually near the bottom of the list in terms of electric vehicle charging infrastructure! If you have an electric car, you are going to have the best luck finding a charging station in the states of Wyoming and Minnesota. Who would have guessed? Your second best experience would be in the midwestern states, a few of the southern ones, and Maine.

Code for this map:

```
# Import registered vehicles by state and fuel tytpe
```

```
carsbystate <- read.csv("~/Library/Mobile Documents/com~apple~CloudDocs/IST 687/My  
Work/Final Project/altfuelstations/2023_light_duty_vehicle_registration_by_fueltype.csv",  
header=FALSE)
```

Rename columns of interest

```
colnames(carsbystate)[1] <- "region"  
colnames(carsbystate)[2] <- "electricvehicle"
```

Remove the first row of carsbystate

```
carsbystate <- carsbystate[-1, ]
```

Make statenames all lower case to match region in state_data

```
carsbystate$region <- tolower(carsbystate$region)
```

Make the new electricvehicles column numeric

```
carsbystate <- carsbystate %>%  
  mutate(electricvehicle = as.numeric(gsub("[^0-9.]", "", electricvehicle)))
```

```
state_data <- state_data %>% mutate(region = as.character(region))
```

new_carsbystate <- new_carsbystate %>%

```
  mutate(electricvehicle = as.numeric(electricvehicle))
```

```
carsbystate %>%  
  mutate(electricvehicle_numeric = as.numeric(electricvehicle)) %>%  
  filter(is.na(electricvehicle_numeric)) %>%  
  select(electricvehicle)
```

```
carsbystate <- carsbystate %>%  
  mutate(electricvehicle = as.numeric(gsub(",", "", electricvehicle)))
```

```
new_carsbystate <- new_carsbystate %>%  
  mutate(electricvehicle = as.numeric(trimws(electricvehicle)))
```

```
new_state_ev_data <- state_data %>%  
  left_join(carsbystate, by = "region") %>%  
  mutate(StationsPer1000EVs = (StationCount / electricvehicle) * 1000)
```

Merge charging station data (state_data) with EV data (new_carsbystate)

```
new_state_ev_data <- state_data %>%  
  left_join(carsbystate, by = "region") %>%  
  mutate(StationsPer1000EVs = (StationCount / electricvehicle) * 100000)
```

Merge the resulting data with the state map data

```
new_map_ev_data <- left_join(states_map, new_state_ev_data, by = "region")
```

Plot the choropleth map

```
ggplot(data = new_map_ev_data, aes(x = long, y = lat, group = group, fill =  
StationsPer1000EVs)) +  
  geom_polygon(color = "white") +  
  coord_fixed(1.3) +  
  scale_fill_gradient(low = "lightgreen", high = "darkgreen",  
                      na.value = "grey50", name = "Stations per 1000 EVs") +  
  theme_void() +  
  labs(title = "Charging Stations per 1,000 Electric Vehicles by State")
```

Next Steps

I have some really great maps, but now I have more questions – in particular about city-to-city travel. The next thing I plan to do is find out how many charging stations you can expect to encounter as you travel in these ways:

- Between major cities and different types of cross-country drives (for example, New York to Washington, D.C.; New York to Los Angeles; New York to Miami; etc.)
- I want to get a sense of what kinds of things you can do while waiting for your car to charge, as this data set also shows things like whether there's a shopping mall nearby.
- I would love to be able to make a Shiny application that lets you pick the cities you want to go between and then show you data about charging stations.

Check-in 3

Final Project Update #3

Dan Pacheco - IST 687 | Due Mar. 21, 2025

Keeping Track of Dataframes

I started to realize that I have so many dataframes in this project, I need to keep better track of them. These are the ones that are used most:

fuelstations / fuelstations_us

The raw altfuelstations data is read from a CSV into the `fuelstations` dataset and then filtered (to remove non-US entries and DC) to create `fuelstations_us`. This is the foundational dataset containing details for each charging station (e.g., FuelTypeCode, StationName, StreetAddress, City, StateAbbrev, Latitude, Longitude, FacilityType, etc.).

state_counts, state_data, and map_data_final

These datasets are created by aggregating `fuelstations_us`:

- `state_counts`: Groups charging stations by state (StateAbbrev) and counts the total per state.
- `state_data`: Merges `state_counts` with state abbreviations (converting them to full state names) for mapping.
- `map_data_final`: Combines the state data with built-in state map data (from the `maps` package) to produce the choropleth maps.

new_state_population, carsbystate, new_state_ev_data, and new_map_ev_data

These are used for per capita and EV-related visualizations:

- `new_state_population`: A dataset derived from the built-in `state.x77` data, containing state populations.
- `carsbystate`: Contains electric vehicle registration data by state (loaded from a separate CSV).
- `new_state_ev_data`: Created by merging `state_data` with `carsbystate` to compute metrics like stations per 1,000 (or 100,000) electric vehicles.
- `new_map_ev_data`: Merges `new_state_ev_data` with map data for choropleth mapping of EV-related metrics.

facility_summary, state_facility_summary, and state_facility_wide

These summarize the facility type information from `fuelstations_us`:

- `facility_summary`: Aggregates charging stations by FacilityType to show national totals.

- `state_facility_summary`: Provides a breakdown of facility types by state.
- `state_facility_wide`: A pivoted (wide-format) version of `state_facility_summary`, used for interactive Shiny scatterplots comparing different facility type counts across states.

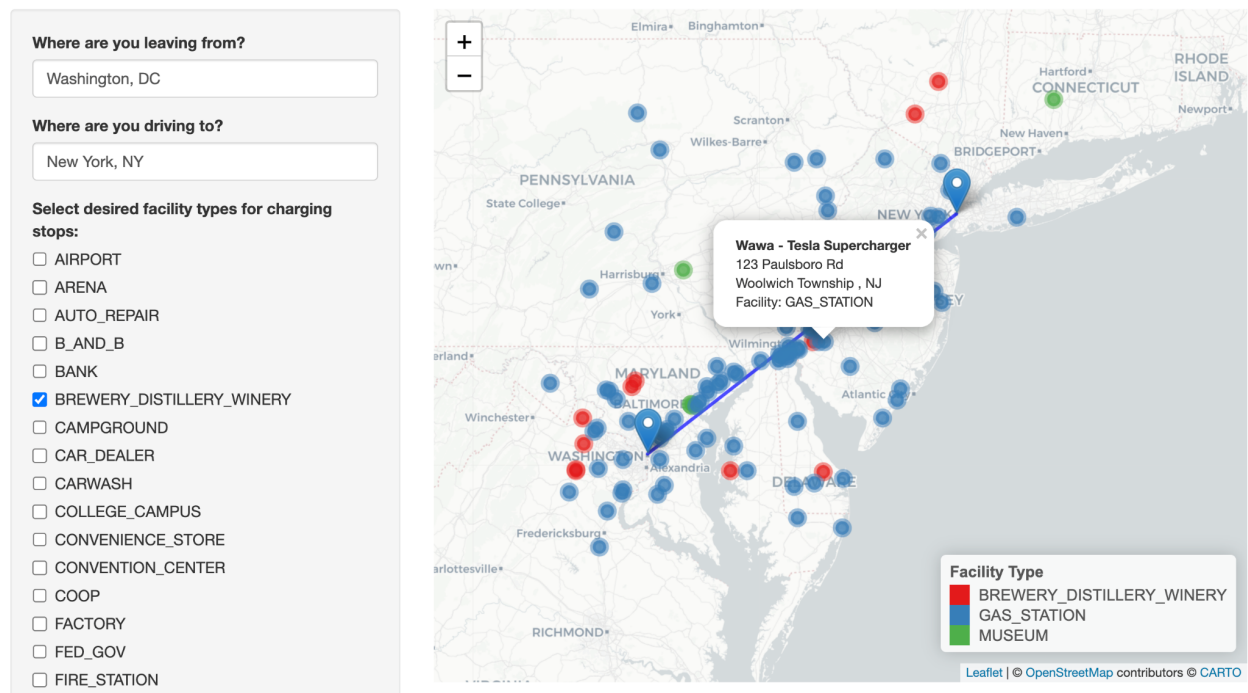
states_map

This is the built-in map dataset (obtained via `map_data("state")`) used to draw state boundaries for the choropleth maps.

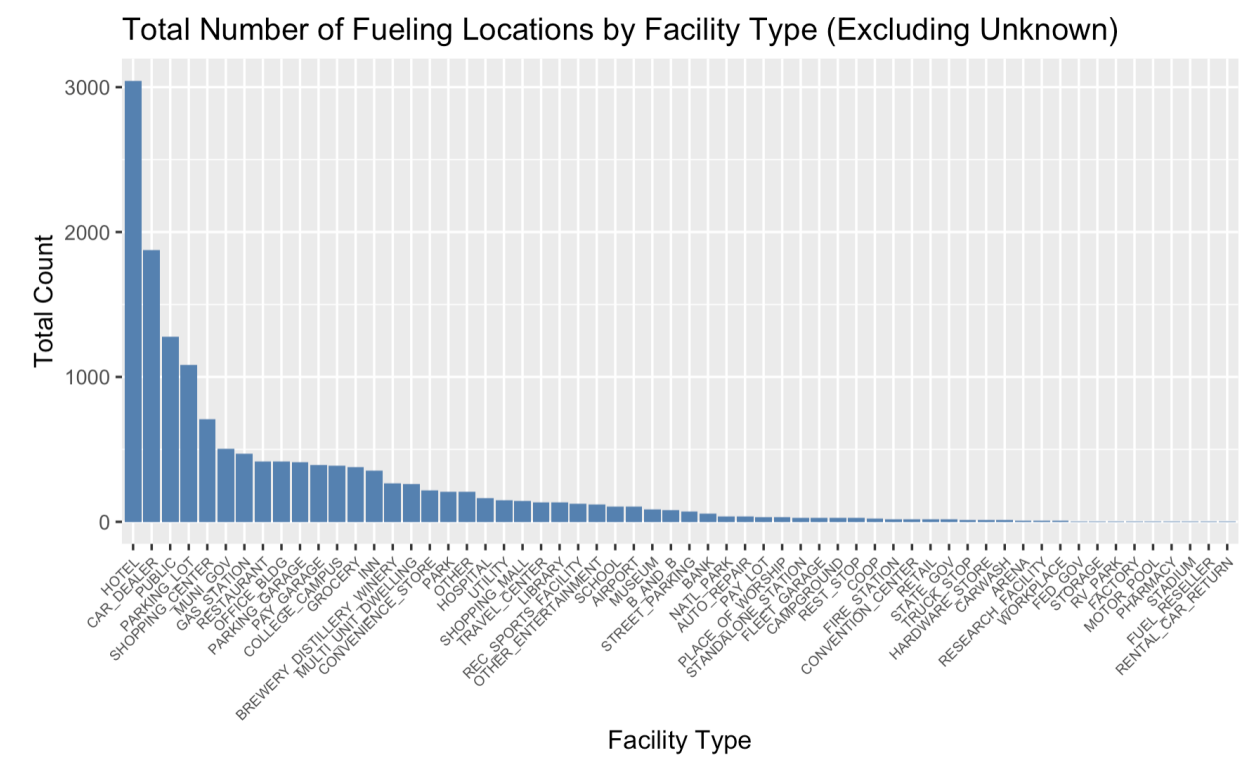
Shiny App Maps

The main addition lately is adding a Shiny App that uses the Leaflet.js library to create an interactive trip planner. The first uses two forms, which ask where you're leaving from and where you are driving to. You check off the kinds of facilities you are interested in charging your car at, and it then uses Leaflet to give you a suggested driving route that allows you to do fun things while you are charging your car on the road trip. When you click on the dots you get info:

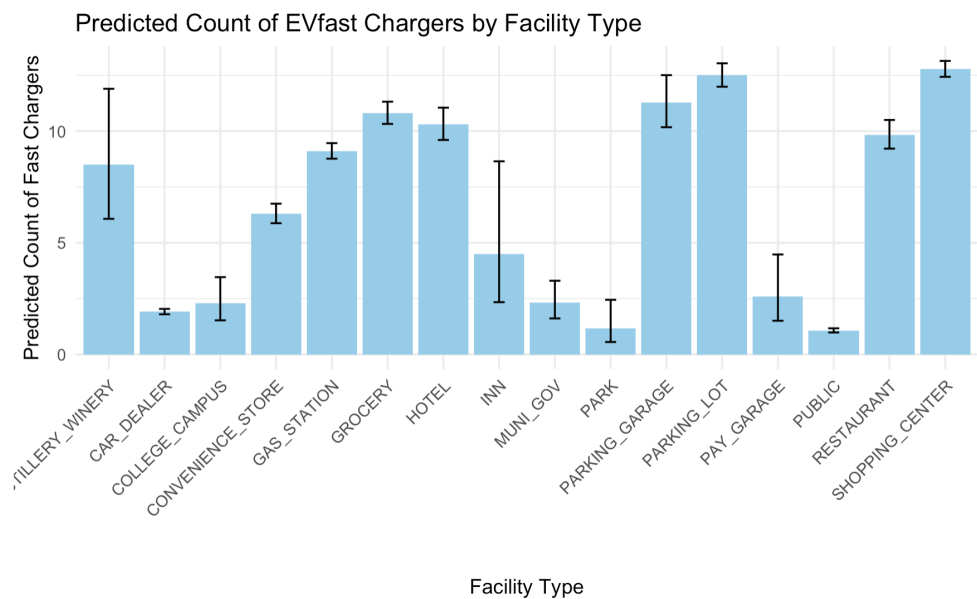
Road Trip Planner for EV Charging & Fun Stops



I was able to generate a histogram showing the most common types of facilities at stations. The list of the top 16 was used in a later statistical analysis.

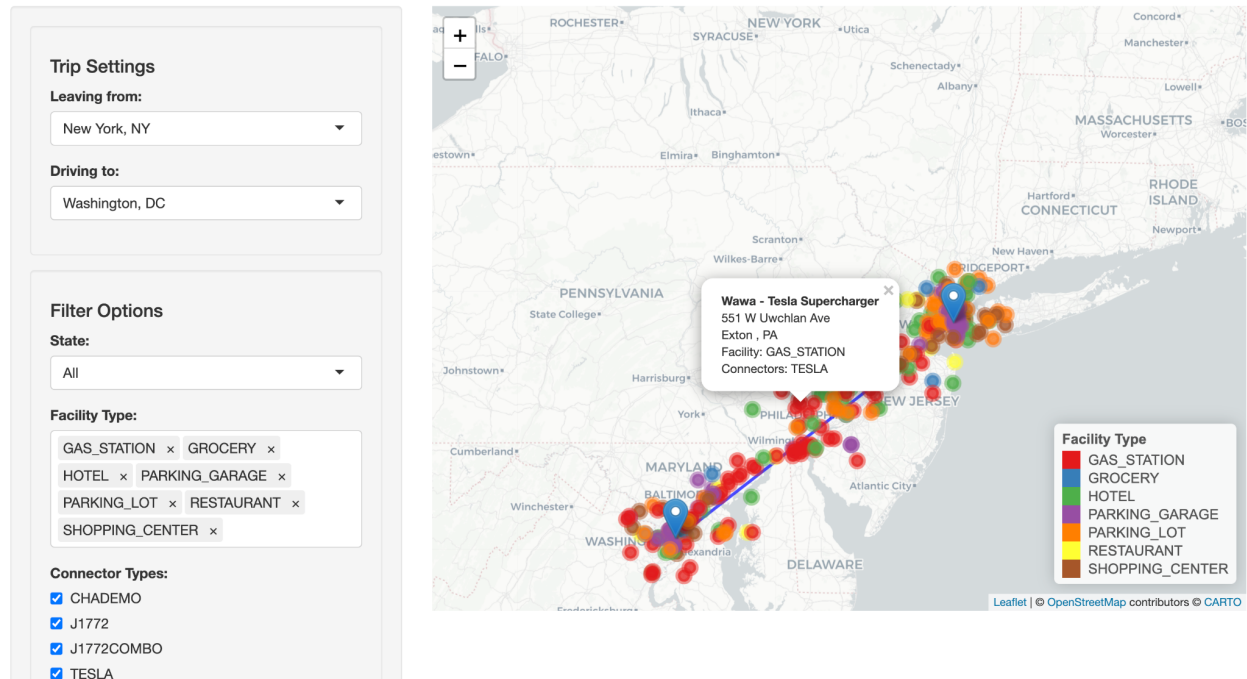


I did some regressions on the top 16 most common facility types, and found that some types of facilities are likely to have higher numbers of fast chargers. That generated this:



This information allows you to then use the road trip planner to maximize the locations you can stop at for the least amount of time, like this. I'm thinking of maybe adding a single checkbox to "Show only the locations with the fastest chargers," and will then filter out everything except the locations in those top 16.

Road Trip Planner for EV Charging & Fun Stops



Investigation Notes

Alternative Fuel Charging Station Data Set Research / Notes

Project Updates So Far: [Update 1](#) | [Update 2](#) | [Update 3](#)

3/3 update: New supplementary data set: vehicle registrations by state:

<https://afdc.energy.gov/vehicle-registration>

I can use this to get things like the ratio of EVs to charging stations, state by state.

I archived this here:

<https://docs.google.com/spreadsheets/d/1CSWyIU9mSks9T1YzPVbBiVEqL0BCimecbjom-n4tygY/edit?usp=sharing>

My ongoing rambling thoughts and investigation notes

One of the biggest hindrances to people adopting electric cars is knowing that they will be able to charge them on longer trips. At least, that's my own personal hindrance! So I think I will focus on the prevalence of stations between different common transportation corridors.

In addition to knowing you can charge, you want to know that you can charge your car in under an hour at each stop. The type of charging station makes a big difference.

Here is what I'm learning so far about stations. My source so far is simply talking with ChatGPT Plus and I will verify what it tells me through more primary sources as I go along.

Analysis of fields

This is a list of every single field and what it means:

https://afdc.energy.gov/data_download/alt_fuel_stations_format

Here are the fields that focus most on charging type. I've highlighted fields that I think deliver some interesting insights for drivers.

- **ev_level1_evse_num:** The number of Level 1 EV charging ports at the station.
- **ev_level2_evse_num:** The number of Level 2 EV charging ports at the station.
- **ev_dc_fast_num:** The number of DC fast-charging ports at the station.
- **ev_other_evse:** Other types of EVSE (Electric Vehicle Supply Equipment) ports available, such as inductive or Avcon conductive ports.
- **ev_connector_types:** The types of connectors available at the station (e.g., J1772, CCS, CHAdeMO, Tesla).

- **ev_network**: The charging network that operates the station (e.g., ChargePoint, Tesla, Electrify America).
- **ev_network_ids**: The unique IDs assigned to the station by the charging network.
- **ev_network_web**: The website of the charging network.
- **ev_pricing**: Information on whether the station charges a fee and how much.
- **ev_renewable_source**: The type of renewable energy (if any) used to power the charging station (e.g., solar, wind).
- **ev_workplace_charging**: Whether the station is primarily used for employee charging at a workplace.

Backgrounding

I don't know as much as I thought I did about electric car charging, so here's what I'm learning as I go along. Here's what ChatGPT says this about the difference between Level 1, Level 2, and DC Fast charging:

Comparison of Charging Levels

Charging Level	Voltage	Power Output	Charging Speed	Typical Use Case
Level 1	120V AC	1-1.9 kW	~3-5 miles of range per hour	Home charging (standard household outlet)
Level 2	208V-240V AC	3-19 kW	~12-60 miles of range per hour	Homes, workplaces, public stations
DC Fast Charging	400V-900V DC	25-350 kW	~60-200+ miles of range in 15-45 minutes	Highways, commercial locations, Tesla Superchargers

Key Differences

1. **Charging Speed:**
 - **Level 1** is the slowest, mainly for overnight charging at home.
 - **Level 2** is much faster and common in workplaces and public areas.
 - **DC Fast Charging** is the fastest, ideal for road trips or quick top-ups.

2. Power Delivery:

- **Level 1 and Level 2** chargers supply AC power, requiring the vehicle's **onboard charger** to convert it to DC.
- **DC Fast Chargers** bypass the onboard charger, delivering high-voltage DC power directly to the battery, which speeds up charging.

3. Connector Types:

- **Level 1 & 2:** J1772 (North America) or Type 2 (Europe) [Note: knowing that there is a connector type in particular for a European car if you have one seems pretty important!]
- **DC Fast Charging:** CCS (Combo), CHAdeMO (older standard), Tesla Supercharger (Tesla vehicles)

4. Installation & Cost:

- **Level 1:** Cheapest, no special equipment needed.
 - **Level 2:** Moderate cost (\$500-\$2,000 for installation).
 - **DC Fast Charging:** Expensive to install (\$10,000-\$100,000+), mainly used in commercial areas.
-

When to Use Each?

- **Level 1** → Best for overnight home charging.
 - **Level 2** → Great for daily use at home or work.
 - **DC Fast Charging** → Ideal for road trips and quick refueling.
-

Analysis in R