Predicting the Market Penetration of

Alternative Fuel Stations and Vehicles - United States

Priya Sathish

Abstract

Transportation sector is dependent on Petroleum based fuel that results in higher emission level. The United States consumes approximately 20 million barrels of petroleum per day, about three-fourths of which is used for transportation. Transportation also has a significant economic impact on American businesses and families, accounting for nearly one-sixth of the average household's expenses. Improving efficiency and reducing costs in this sector can thereby make a notable impact on our economy.

Increased economic and energy security aren't the only benefits. Widespread use of alternative fuels and advanced vehicles could reduce the emissions that impact our air quality and public health. Production and distribution of these alternative fuels that can be economically available is a constraint for the increase in the number of alternative fuel driven vehicles.

One known fact is "attaining environmentally beneficial transportation fuel at an affordable price" is the way of the future. This project focuses on the number of alternative fuel stations and the number alternative fuel powered vehicles across the United States, its growth trend and developing a predictive model.

Load Data

Station data and vehicle inventory data has been collected from https://afdc.energy.gov/data_download/ (https://afdc.energy.gov/data_download/). Data wrangling performed and saved in file Alt_Fuel_DW.ipynb from which data from the following cleaned up csv files will be loaded for Exploratory Data Analysis.

Load Shape files

We load the United States shapefile for mapping.
'shape_us_state/tl_2017_us_state.shp'

https://medium.com/@erikgreenj/mapping-us-states-with-geopandas-made-simple-d7b6e66fa20d

We load the geo-json file to map CA.
'ca_california_zip_codes_geo.min.json'

Data Wrangling Explained

Acquired data from https://afdc.energy.gov and loaded csv files as pandas dataframes removing rows with invalid data by setting the parameter error_bad_lines as False. Switching error bad lines turns off the stdout error messages from showing.

Station Data:

Strip leading and trailing white spaces from Zipcode column in the dataframe and selected first 5 characters to be the Zipcode to avoid long codes and converted to numeric datatype. Look for invalid zipcodes and drop them from the dataframe. Clean column names by removing white spaces.

Set proper indices. Drop duplicate rows if any exist.

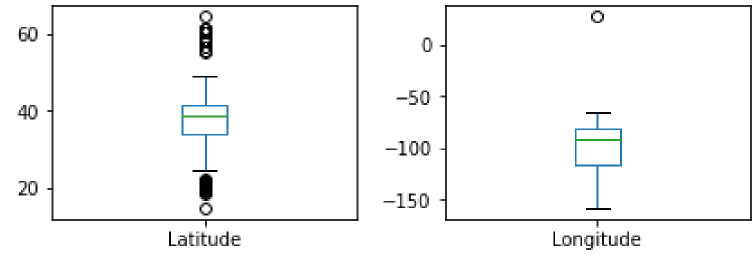
Create a new dataframe by performing a deep copy from the raw data filtering only rows with a valid open date.

We are going to do an anaylsis of the growth of alternative fuel stations across United States, so it is necessary that we have valid open date for the existing stations. Convert the date fields to pandas datetime format.

Create a new column for the Year the station was opened by retrieving the year from the open date.
Convert the OpenYear column to contain numeric values.

Select only required fields from the raw data and save as new dataframe containing the station ID, FuelTypeCode, StationName, StreetAddress, IntersectionDirections, City, State, ZIP, GeocodeStatus, Latitude, Longitude, Country, OpenDate and OpenYear.

Check for outliers for Latitude and Longitude. From the boxplots, we can see that there are few outliers with Latitude and Logitude of 0 also. We can drop those rows which will enable accurate mapping plots.



Save the station data frame to new csv files for future use to perform Exploratory Data Analysis and ML.

Vehicle Data:

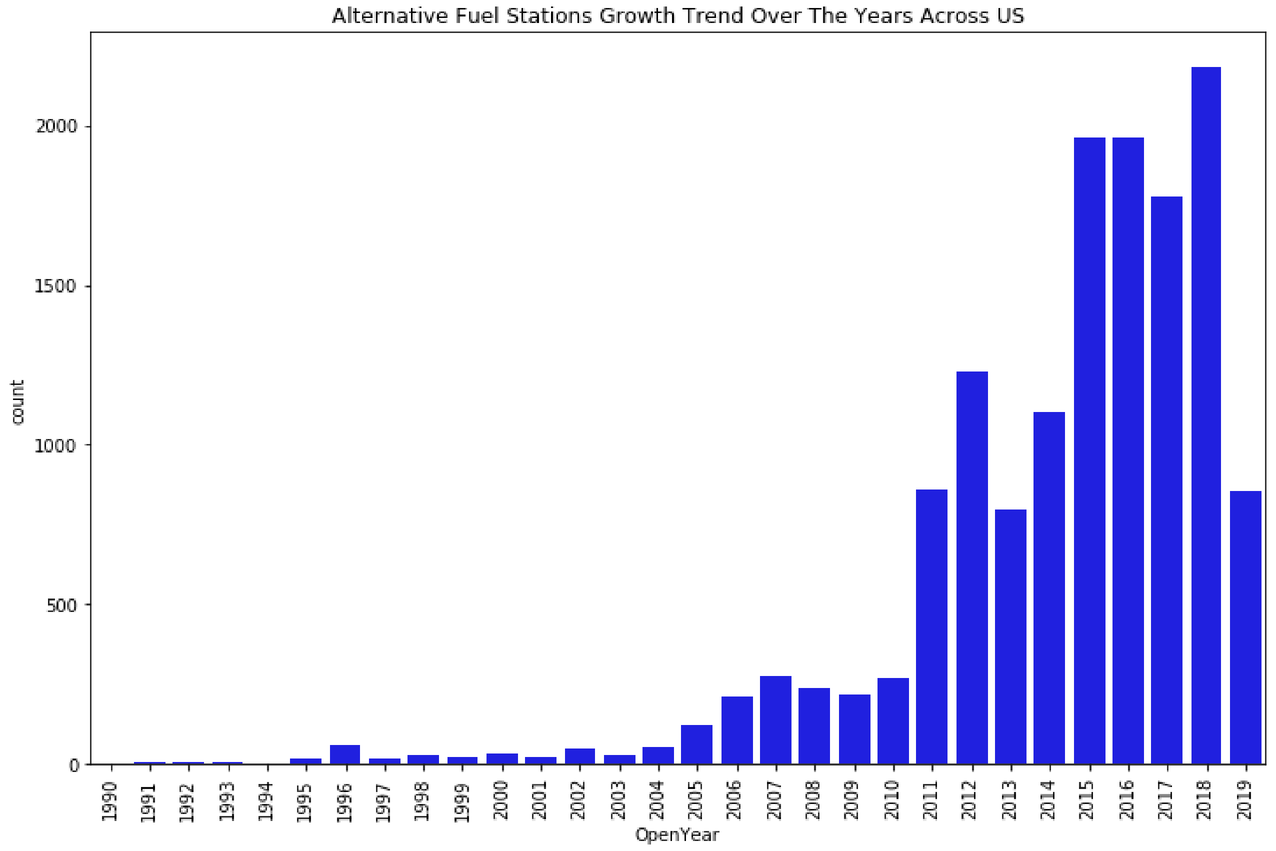
The data frame contains Year as the column header. We want Fuel Type to be the column header.
We transpose the data frame and make first row as the column header to contain Fuel Types as the column names.

PEVs are electric vehicles. Let us keep all electric vehicles under one Fuel Type ELEC.
Add the number of PEVs to the number of ELEC vehicles and group them as one Fuel Type and drop the PEVs column.

Set the index of the data frame to Year.

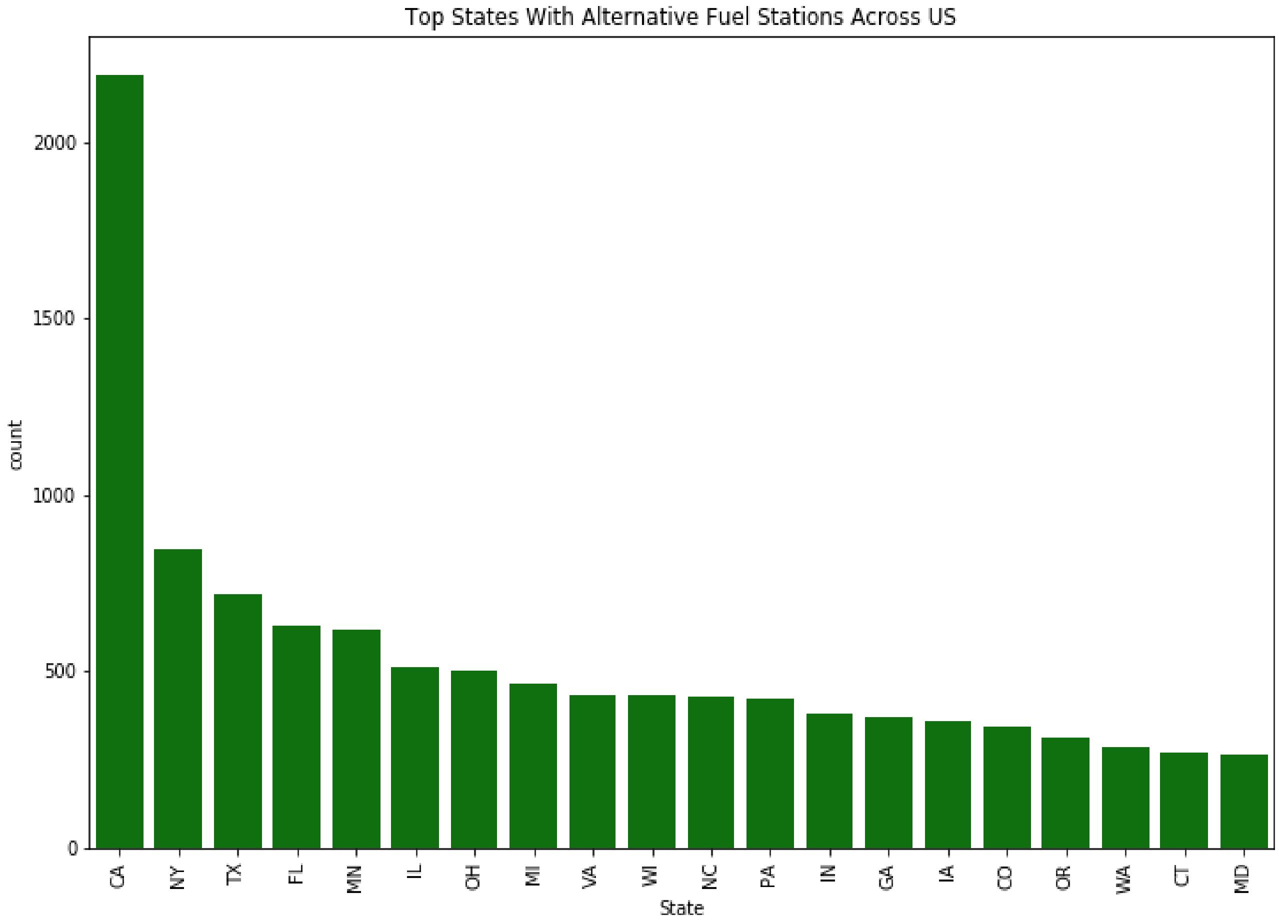
Save the vehicle data frame to new csv file for future use to perform Exploratory Data Analysis and ML.

Perform Exploratory Data Analysis



The trend shows overall growth in the alternative fuel stations year on year, with remarkable growth since 2011. This might be due to increased awareness created by Clean Cities Coalitions in the community. The mission of Clean Cities coalitions is to foster the economic, environmental, and energy security of the United States by working locally to advance affordable, domestic transportation fuels, energy efficient mobility systems, and other fuel-saving technologies and practices. Clean Cities coalitions are comprised of businesses, fuel providers, vehicle fleets, state and local government agencies, and community organizations.

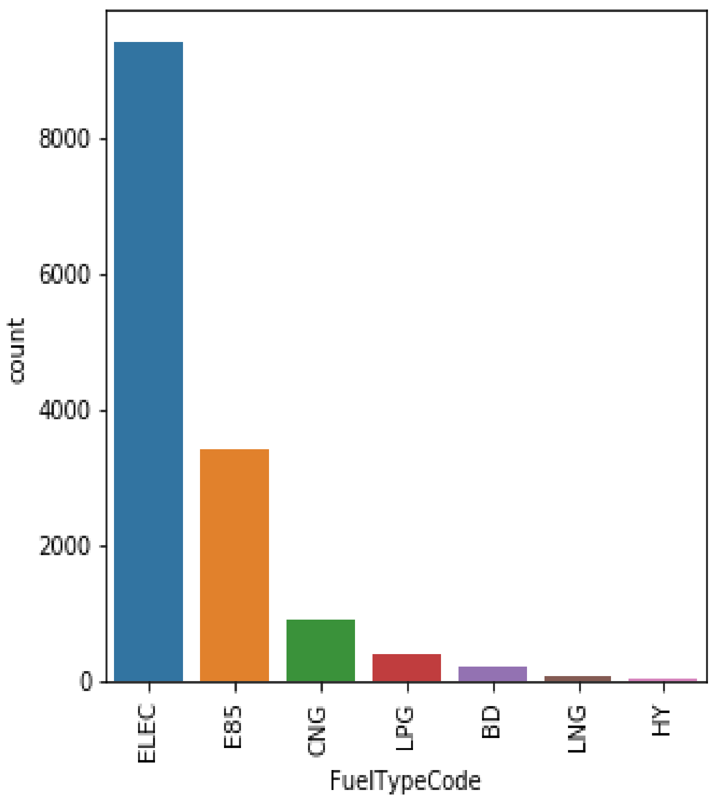
Note: *Number of alternative fuel stations in 2019 is accumulated sum until the month of April. At this rate it is expected to surpass the total number of stations above 2500 for the year ending.*



Synopsis of the above plot shows highest concentration of alternative fuel stations in California followed by New York and Texas states. Upon research State of California has launched a campaign to reduce consumption of petroleum and diesel fuels and shifting towards cleaner fuels. This campaign is known as Low Carbon Fuel Standard (LCFS) http://www.cadelivers.org/low-carbon-fuel-standard/ (http://www.cadelivers.org/low-carbon-fuel-standard/).

From 2011 to 2018 the LCFS has avoided 13.7 billion gallons of petroleum, increased 74% use of clean fuels, invested 2.8 billion in clean fuel production.

The LCFS is working to transform the fuels market from one that relies almost entirely on petroleum-based fuels to a diversified one that uses a variety of alternative fuels. This market-based transition to clean, low carbon fuels is leading to technology innovation that’s helping California meet its long-term climate, clean air, and public health goals. Just as the state requirements are helping shift our electricity mix to more renewables like wind and solar, the LCFS is also delivering the clean fuels we need today and going forward.

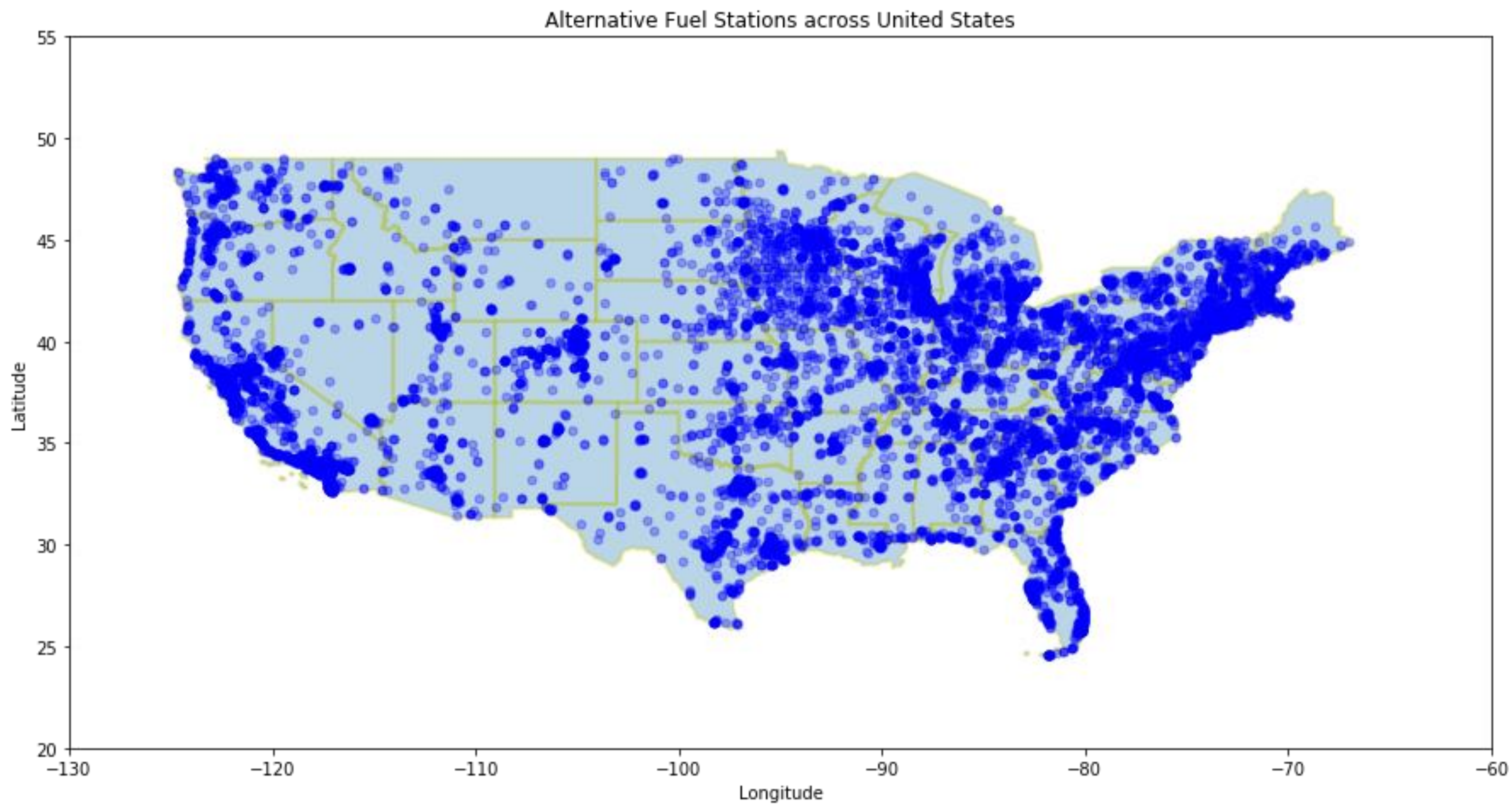


About 90% of the alternative fuel used in transportation sector is comprised of electricity and ethanol85. Among them 65% of the fuel is electric based. The reason for the high number of electric charging stations is attributed to No Tail Pipe Emissions and environmental friendly.

Plug-in electric vehicles (PEVs) are capable of drawing electricity from off-board electrical power sources (generally the electricity grid) and storing it in batteries. Electricity for charging vehicles is especially cost effective if drivers are able to take advantage of off-peak residential rates offered by many utilities. In many cities, PEV drivers also have access to public charging stations at libraries, shopping centers, hospitals, and businesses. Charging infrastructure is rapidly expanding, providing drivers with the convenience, range, and confidence to meet more of their transportation needs with PEVs.

The next popular alternative fuel is E85(ethanol) that is predominantly produced by fermenting and distilling starch crops mostly corn. This is 100% renewable fuel and roughly 1 acre of corn can be processed into 330 gallons of combustible ethanol.

Alternative Fuel Stations Distribution in United States

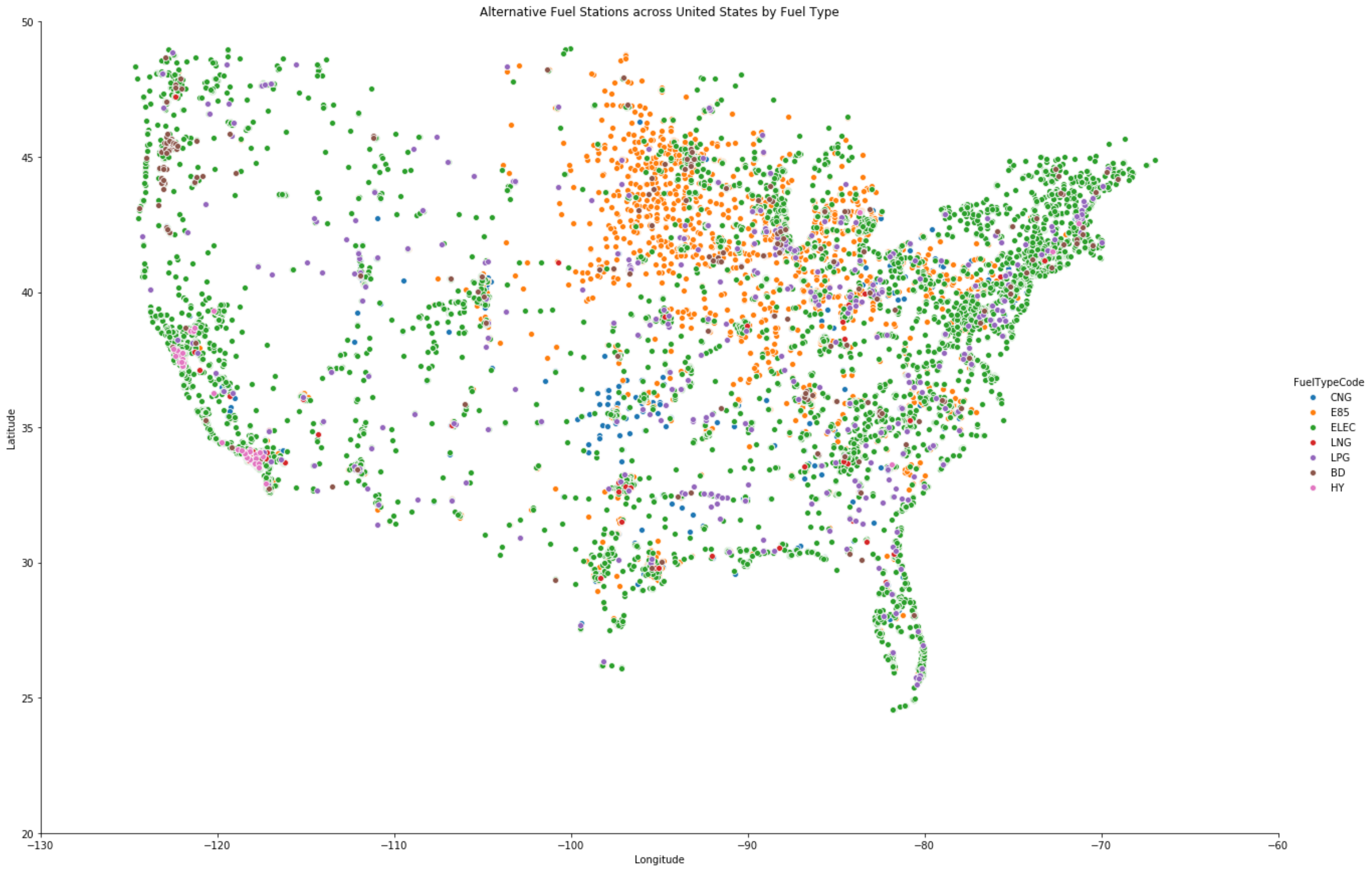


The above geographic plot shows the concentration of alternative fuel stations across United States. It clearly indicates that alternative fuel stations are popular along the coastal line and as well where the population density is high.

Many alternative fuels exist, but few are as bountiful, easily produced and cost effective as traditional fossil fuels. The emissions impact and energy output provided by alternative fuels vary, depending on the fuel source.

Few commonly consumed alternative fuels include Ethanol, Natural Gas, Bio-diesel, electricity (batteries), hydrogen (fuel cells), non-fossil methane, propane and other biomass sources.

Alternative Fuel Stations Distribution by Fuel Type



The above geo plot shows various types of alternative fuel stations across United States. Key points are:

- Electric fuel stations are prominent along the coastline.
- E85 fuel stations are concentrated in North Central states.
- HY stations are concentrated mainly in CA along the coastline.

Individual plots to show the concentration of stations with respective fuel types across the United States

Biodiesel:

Produced from vegetable oil or animal fats, the projected production of biodiesel in the US is nearly 12 billion gallons. North Carolina seems to be leading the way in widespread availability of biodiesel.

Compressed Natural Gas:

If you've seen a municipal bus that's powered by natural gas, it's using compresed methane. With the highest popularity in oil rich nations outside of the US, it's seeing limited use domestically.

E85 (85% Ethanol gas):

An ethanol fuel blend of 85% ethanol, 15% gasoline. Flex fuel vehicles can use E85. Popular in corn growing states where there are local and Federal subsidies.

Electric vehicle charging stations:

Heavy subsidies and the success of Tesla have increased the market viability and variety of all electric vehicles. An increase in pay-as-you-go charging stations as shown in this map make it feasible to travel for longer distances than previously possible.

Hydrogen:

A *hydrogen highway* is a chain of hydrogen-equipped filling stations and other infrastructure along a road or highway. Italy and Germany are collaborating to build a hydrogen highway between Mantua in northern Italy and Munich in southern Germany.

Liquified Natural Gas:

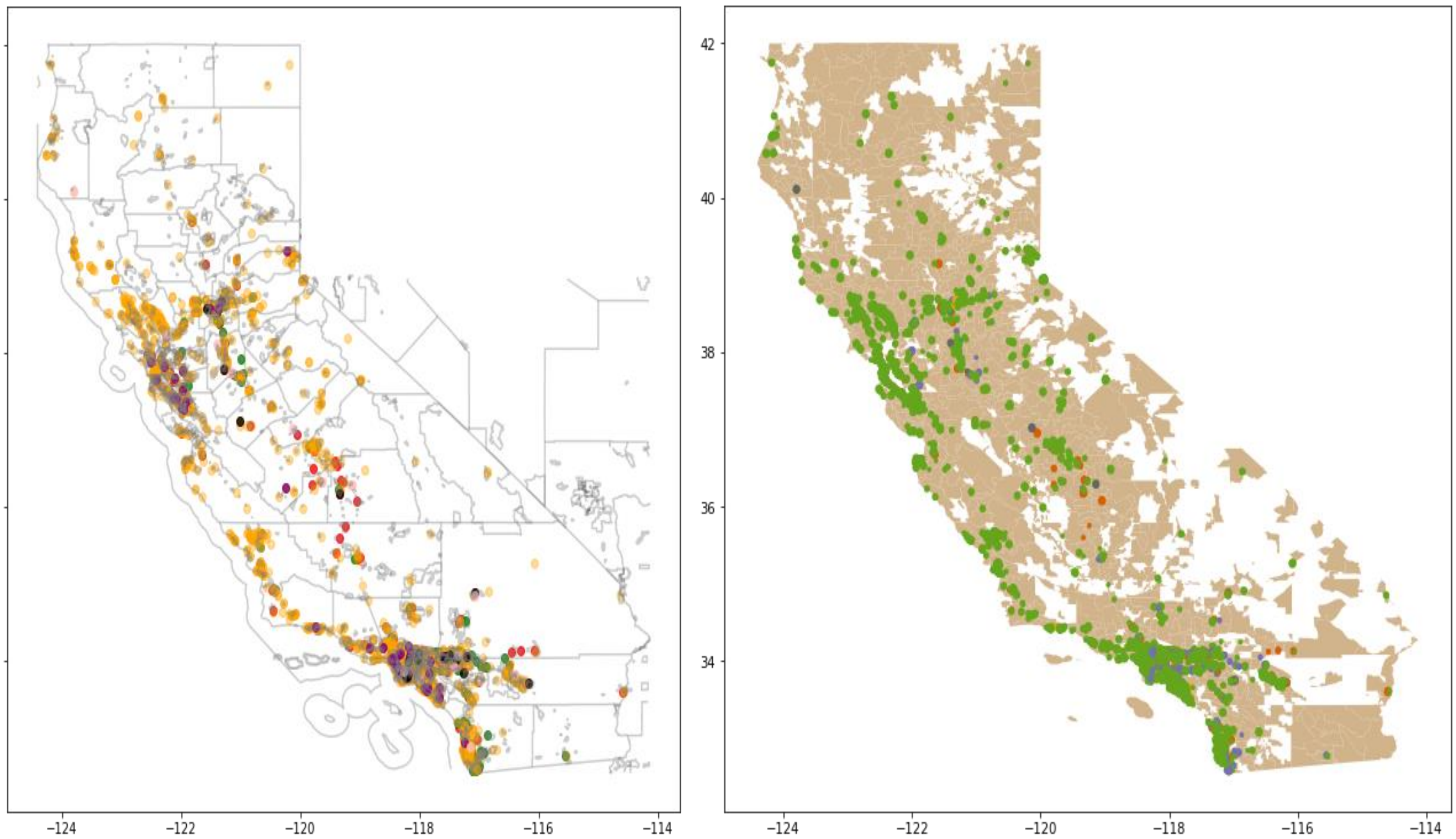
Liquified natural gas is primarily methane that has been converted to liquid for ease of storage or transport.

Propane:

With propane's wide availability and lower maintenance costs, it's an attractive option for light duty industrial vehicles.

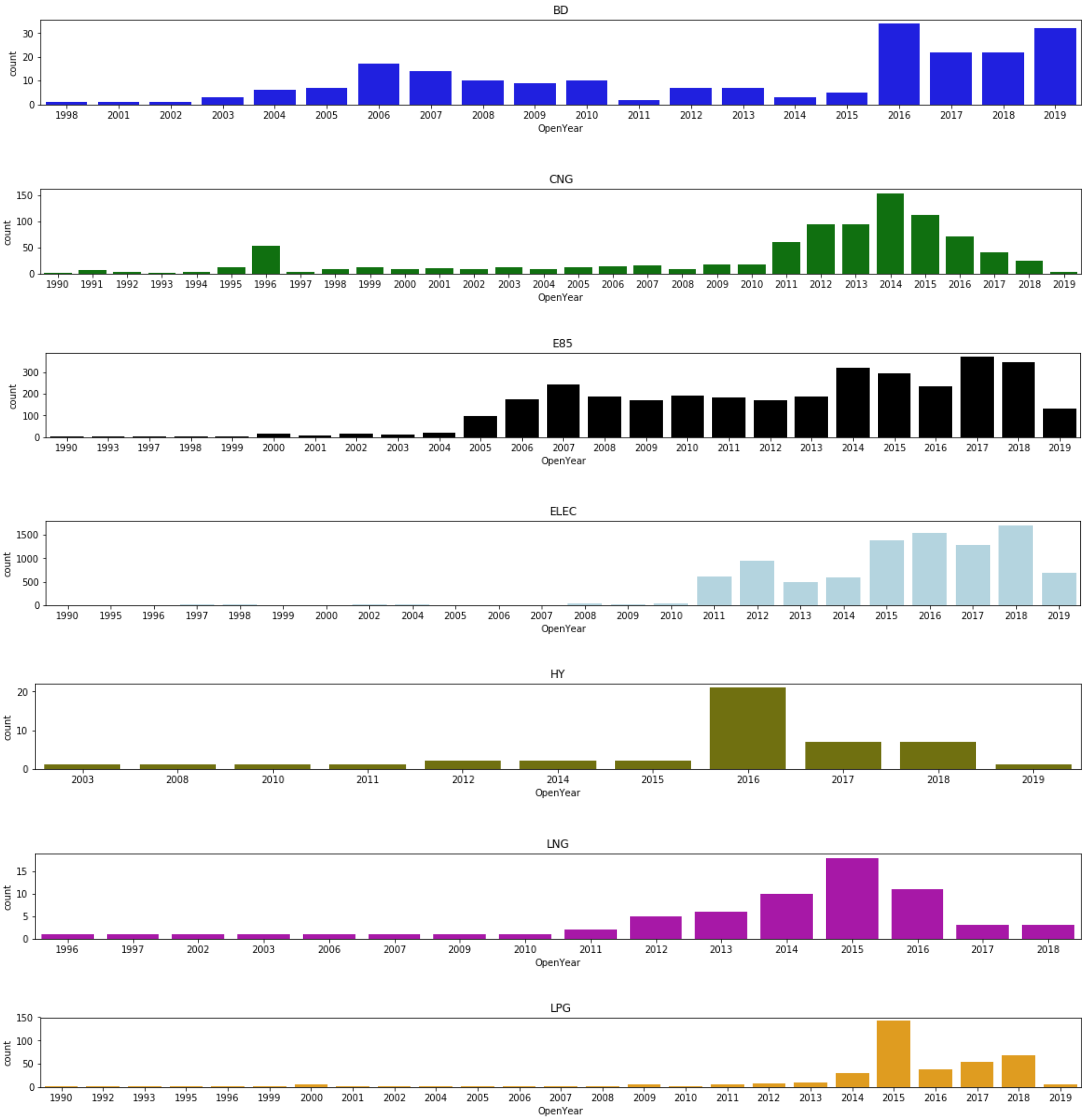
Example

California has the highest number of alternative fuel stations as per the data and as we can see tremendous growth in the electric fuel stations over the past years.



Closer look at the distribution of alternative fuel stations by different fuel types in the state of California

Trend of Fuel Stations by Fuel Type – 1990's onward

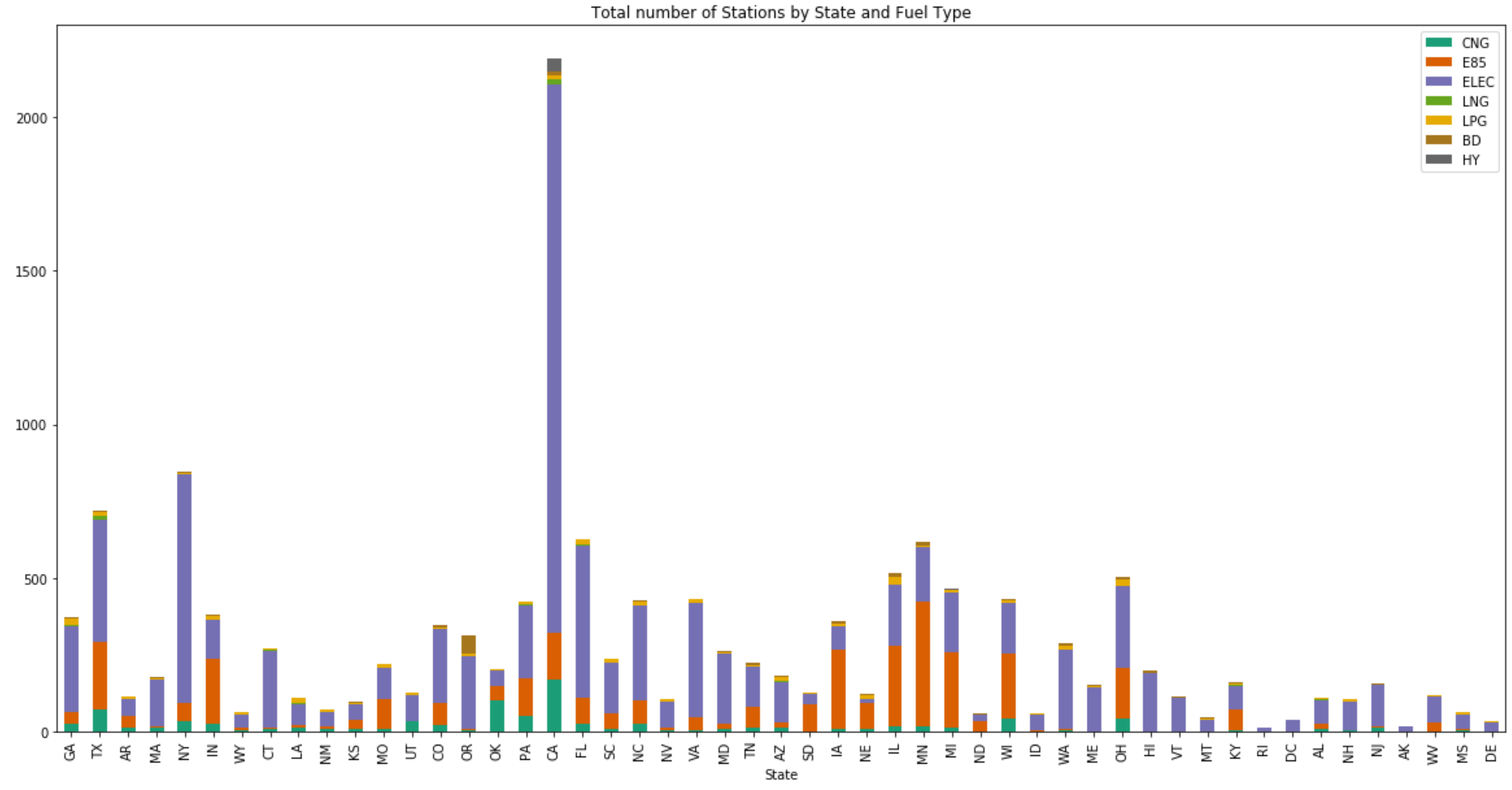


Why there is decline in CNG and LNG in transportation industry?

With growing popularity of EV's combined with steady improvement in battery technology, the pervasiveness of the grid, and services to enable things like high-speed charging makes the EV's market inevitable.

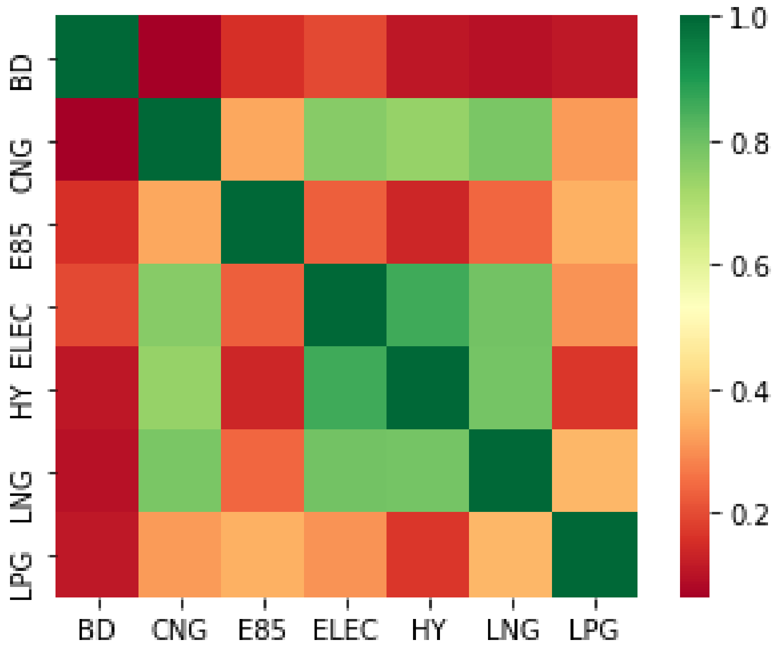
Secondly, the high cost of filling stations has also kept a lid on CNG / LNG. The CNG or LNG driven vehicles get worse mileage than regular gas or EV's.

Nevertheless, CNG is increasingly touted as the fuel of the future for the U.S., in part because of declining prices and large reserves and growing demand.



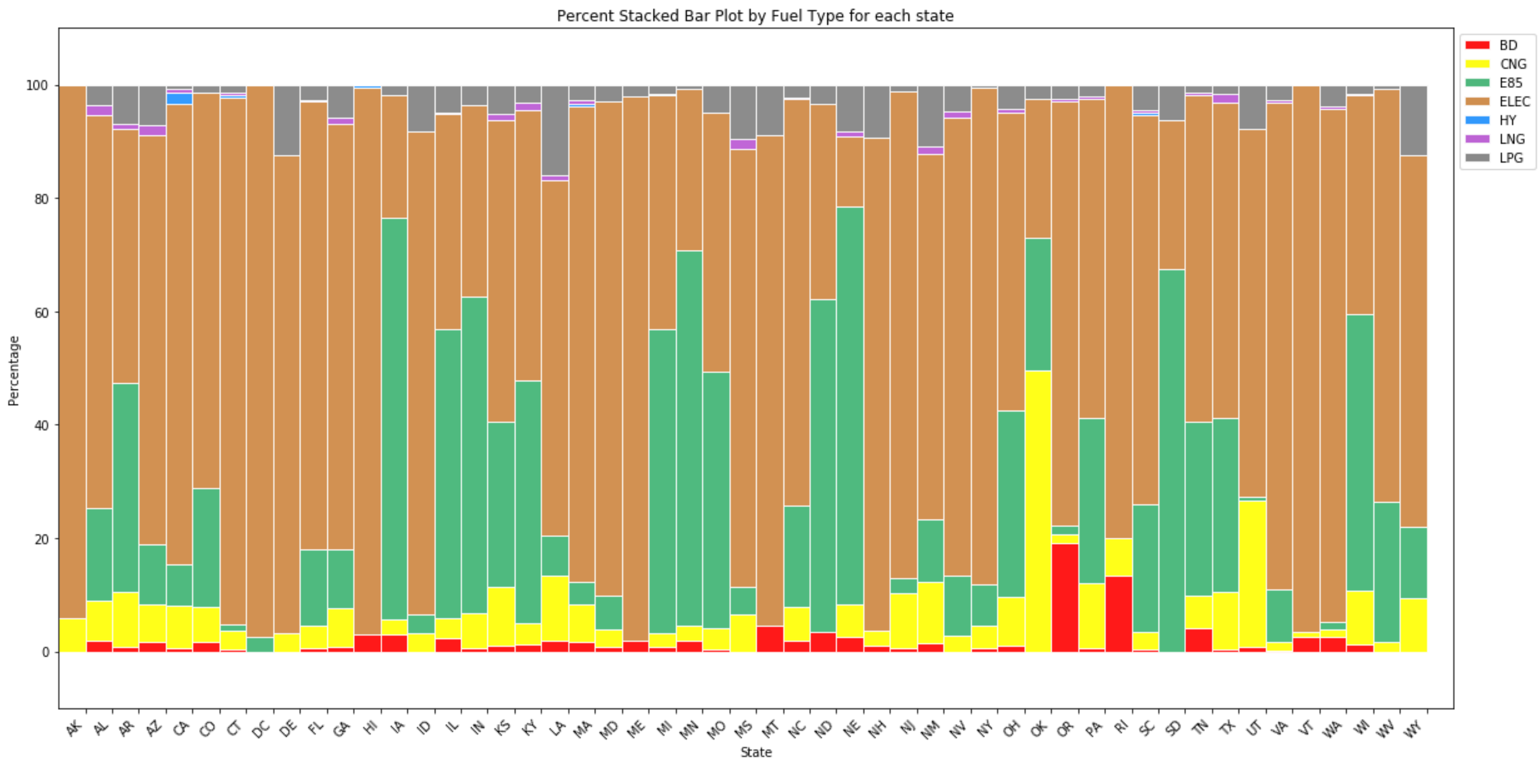
The stacked bar plot shows the distribution of the different alternative fuel stations in each state. CA has the most Electric, CNG, LNG and Hydrogen fuel stations. MN has the most E85 fuel stations. IL has the most LPG fuel stations. OR has the most BD fuel stations.

Correlation Heat Map



The heat map shows correlation among various alternative fuel stations by type. It shows high correlation amongst CNG, HY and ELEC. There is no correlation amongst BD, E85 and LPG.

Percentage Distribution of Alternative Fuels in Each State

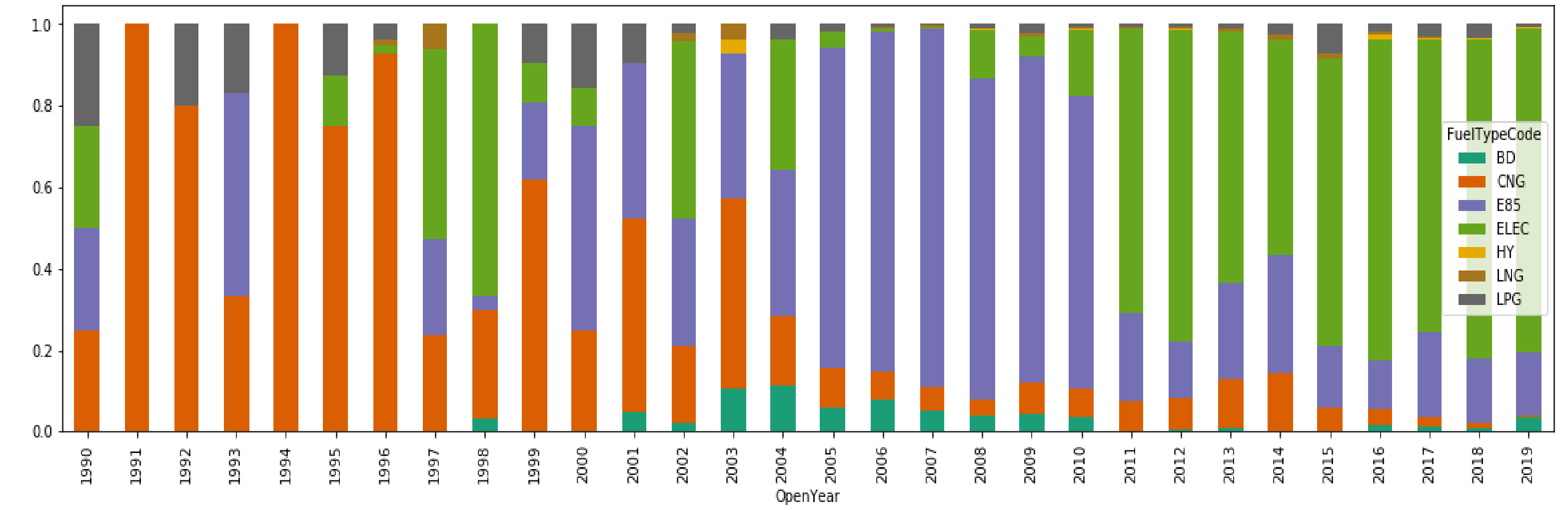


In general, electric fuel stations are predominantly higher in proportion in most of the states except for those where ethanol type fuel is produced more which is supported by the corn fields. This can be seen in states like NE, IA, MN, SD and WI.

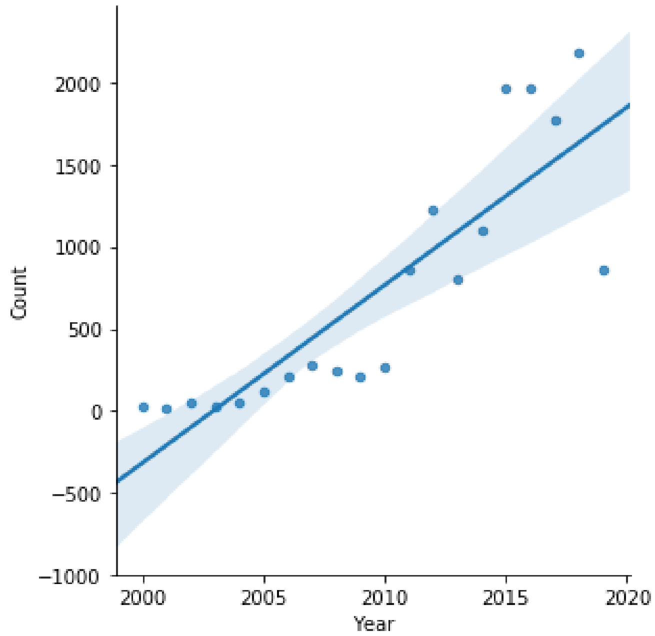
State regulations and incentives are spurring CNG station investment and vehicle adoption. It is no accident that states like Oklahoma and Utah have the highest concentrations of CNG stations and vehicles. Oklahoma provides waivers of state tax liability to in-state suppliers of CNG station components. Utah requires that CNG be supplied by regulated utilities (effectively capping prices).

CA has adopted HY based fuel to meet its long term climate, clean air and public health goals. HY fuel stations are clustered in major metropolitan cities in CA that are also tech savvy. CA has about 40 HY fuel stations for about 6500 fuel cell cars as of May 2019. There is now 1 retail station per 163 vehicles. Though it is in growing trend it is expected to not reach the states set goal of 100 HY fueling stations by 2020.

Percentage Distribution of Alternative Fuels by Year

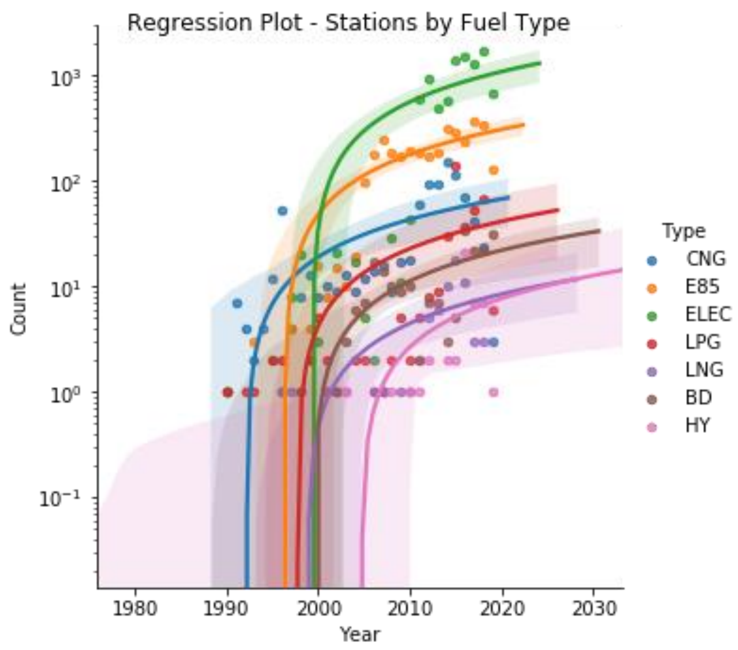


In 90's the majority of alternative fuels stations opened were Biodiesel based which has shifted towards electric and E85 type alternative fuels mainly due to advancement in technology and affordable cost. In the past decade it is distictly focused on growth towards electric based stations. This combined with higher number of electric vehicles justify the growth of number of stations.

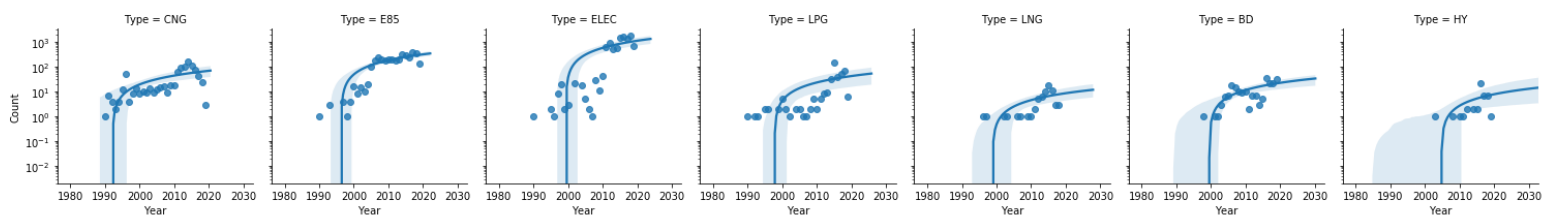


Over the last 15 years the total number of alternative fuel stations is in a steeper growth that is at a rate of 100 to 125 fuel stations year over year.

Regression plot with log scale on y axis enables the separation between various fuel type stations data points for better visualization.



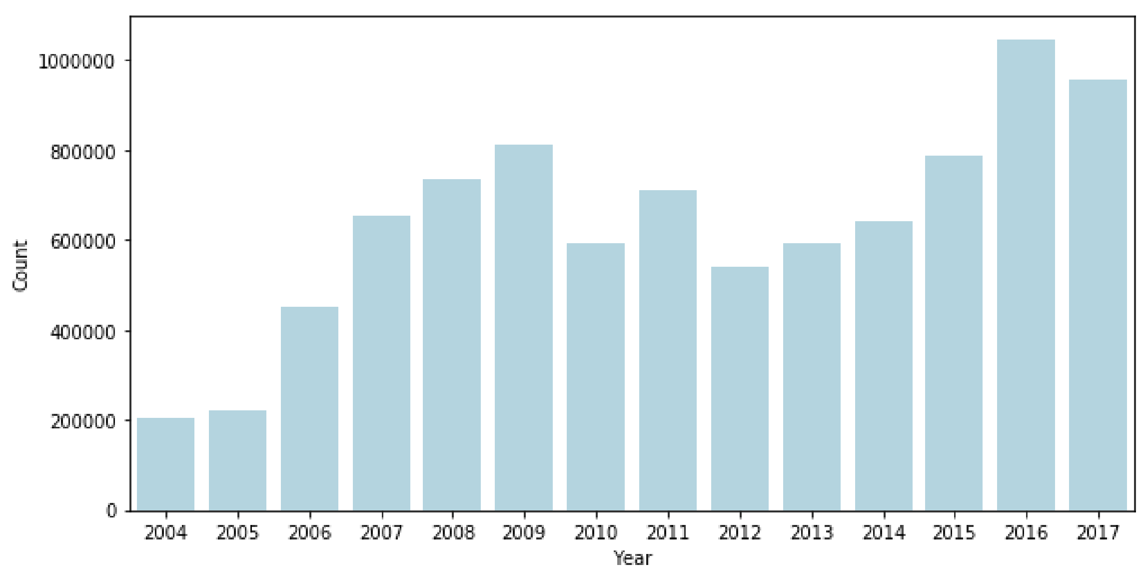
sns lm plot to show the regression by fuel type

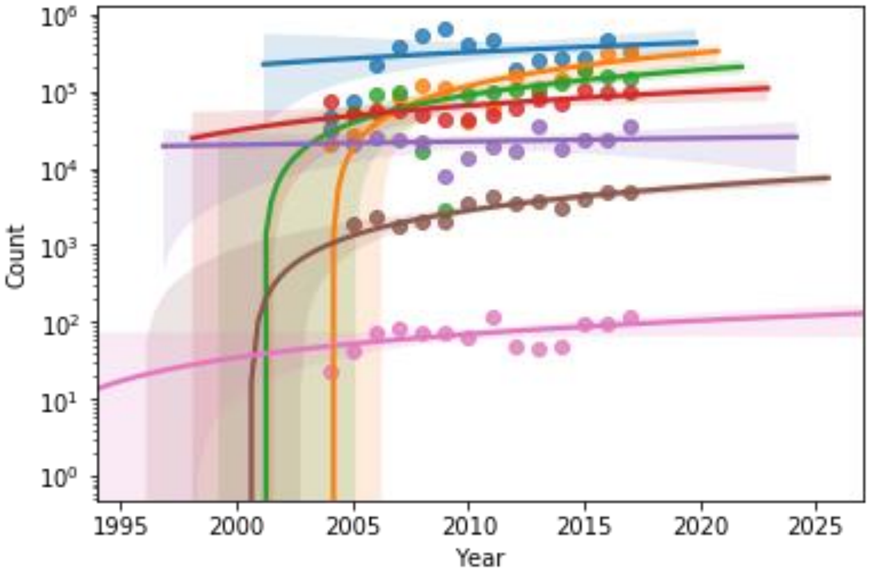


Individual regression plots of alternative fuel stations for each fuel type.

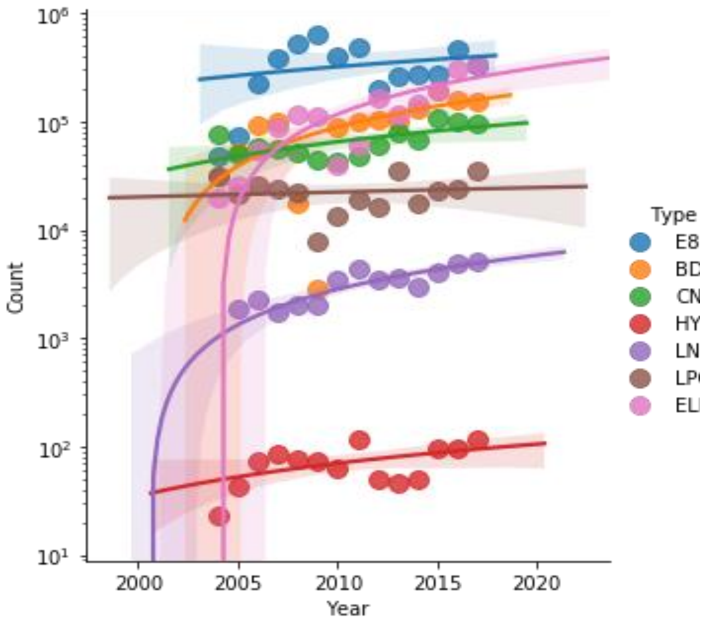
Vehicle Data

Alternative fuel equipped vehicles trend over the years

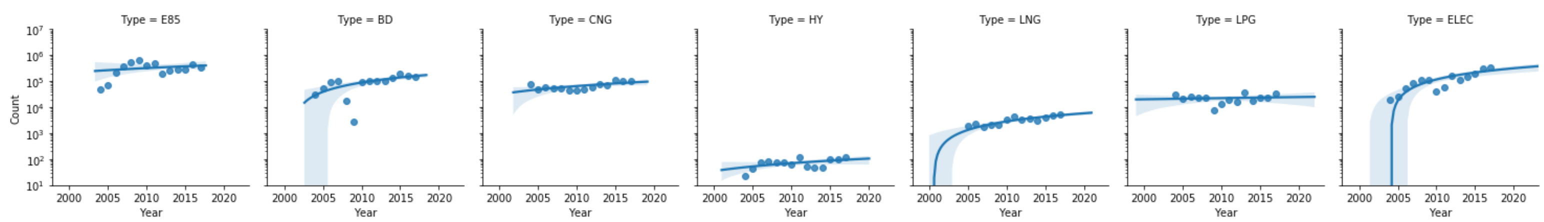




Regression plot with log scale for vehicles equipped with alternative fuels shows an increase in growth trend which coincides with the growth trend of number of alternative fuel stations.



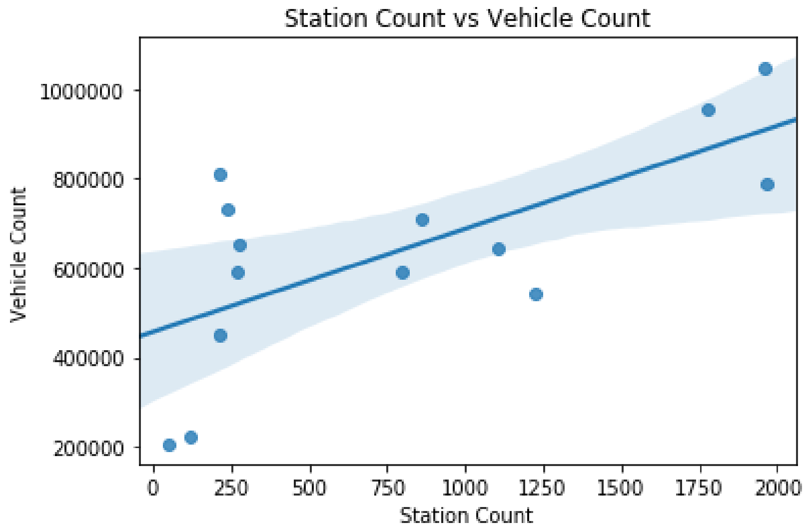
sns lmplot for the same vehicle data.

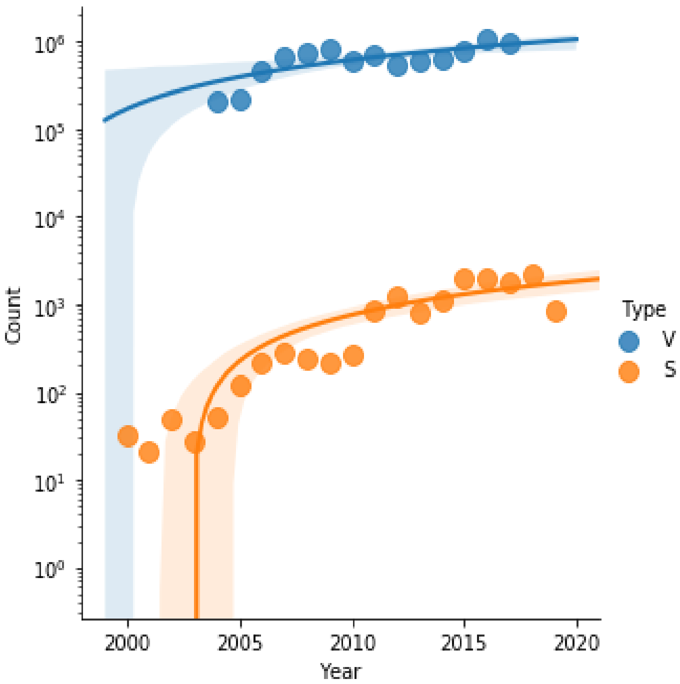


Individual regression plots for vehicles equipped with alternative fuels shows growth trend except for CNG and LNG that flattens which is again supported by stagnation in stations trend.

Let’s look at the correlation between number of alternative fuel stations and number of alternative fuel vehicles across US using the scipy stats pearsonr function that returns the pearsonr coefficient and the pvalue.
Selecting a signifance level of 0.05, and so if the pvalue is less than 0.05, we can find out if the correlation coefficienct returned is significant.

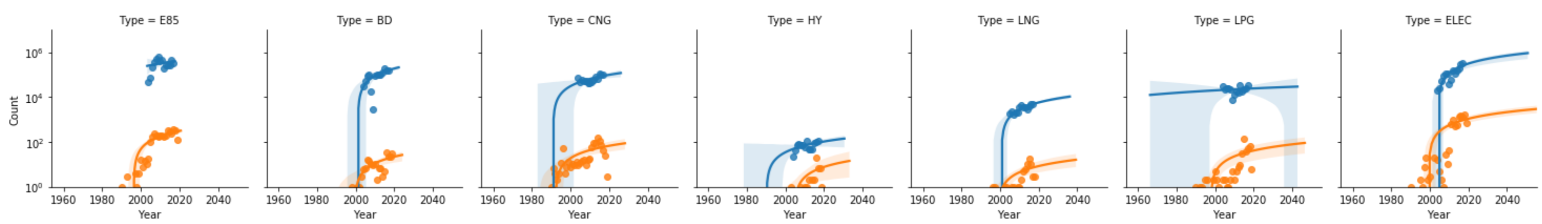
Calculate the pearsonr coefficient between the stations and vehicles using scipy stats:
scipy.stats.pearsonr(result_corr_df.Count_x, result_corr_df.Count_y) = (0.6822392170959737, 0.007186427141207331)
With pearsonr coefficient of about 68% and p value of 0.007 which is less than the significance level of 0.05 we can conclude that the correlation is significant.

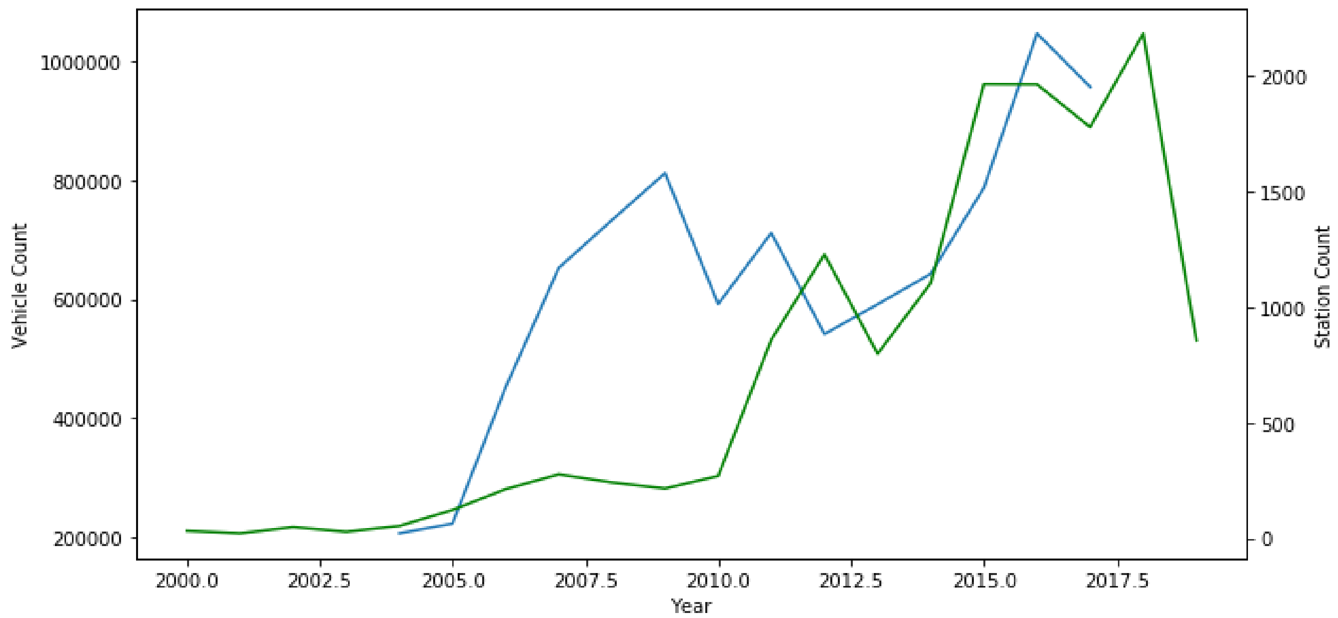




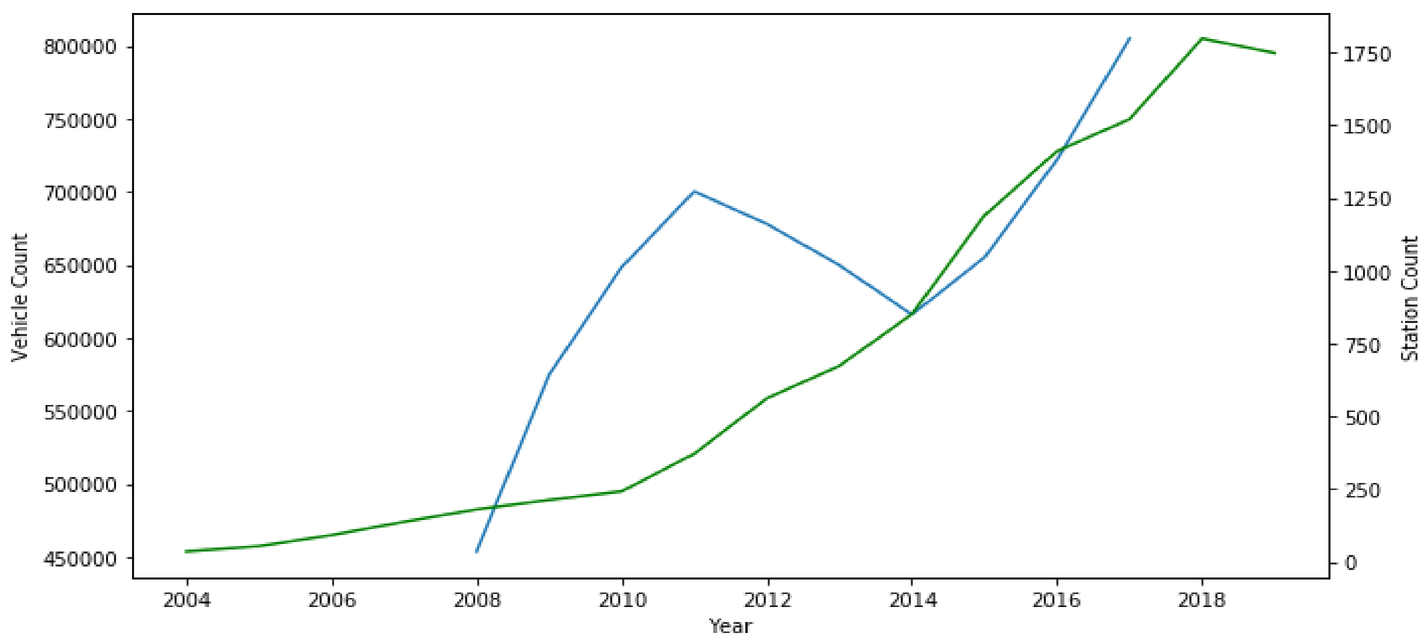
We can see the trend between the number of alternative fuel stations and vehicles equipped with alternative fuels being parallel to each other from the year 2005 onwards showing strong correlation.

Individual regression plots to see the trend between the number of stations and number of vehicles for each fuel type. Strong correlation seems to be the case with alternative fuel types like BD, CNG, LNG and ELEC whereas with HY, the number of stations seem to progress at a higher rate than the number of HY fuel cell based vehicles.

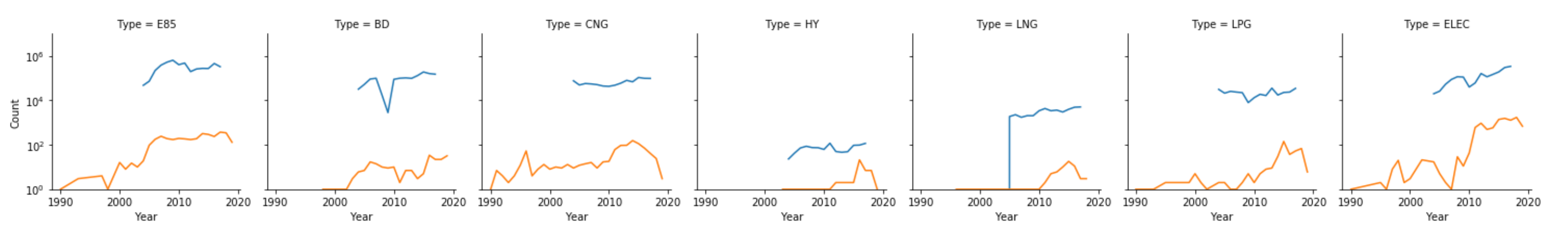




We can clearly see that from 2005 onwards the alternative fuels station count and alternative fuels vehicle count trends similar and follow each other showing that some correlation exists between them.



This plot shows the rolling mean of the previous plot with lines being smoothed out. Overall, the number of alternative fuel stations and alternative fuel vehicles seems to be on the rise in the past 15 years.



These individual plots clearly show increase in E85, BD and Electric vehicles and stations while CNG, LNG, LPG and HY show decline in the number of stations.

Future Work

- Build a model to predict number of alternative fuel stations and number of alternative fuel equipped vehicles for any given future year - either by fuel type or by total.
- Build a model to predict best location for the future alternative fuel station.
- Assessing individual alternative fuel cost trend's effect on the rate of change in the number of alternative fuel stations and vehicles equipped with that particular fuel type.