Final Project

Stephanie Tolly and David Selemba

2023-12-01

```r
library(readr)
library(dplyr)
library(ggplot2)
library(ggmap)
library(maps)

library(tibble)     # data frame printing
library(dplyr)      # data manipulation

library(data.table)
library(scales)

library(tidyverse)
library(caret)
library(caTools)
library(stargazer)
library(gridExtra)
```

```r
color_blind_friendly_cols <-
  c("#999999", "#E69F00", "#56B4E9", "#009E73",
    "#F0E442", "#0072B2", "#D55E00", "#CC79A7")
```

```r
stations<-read.csv('alt_fuel_stations_2023.csv')
new_york <- read.csv('NY_EV_Registrations.csv')

stations <- stations %>%
  filter(State=="NY") %>%
  select(c('Station.Name','City','State','ZIP','Access.Days.Time','EV.Network','Latitude','Longitude','(
stations<-stations %>%
  mutate(year=substr(Open.Date,1,4))

new_york <- new_york %>%
  select(-Drivetrain.Type,Vehicle.GVWR.Class,DMV.Snapshot.ID,DMV.Snapshot..Date.,Latest.DMV.Snapshot.Fla
new_york <- new_york %>%
  mutate(year=substr(Registration.Date, 5, 9))
new_york$year <- sub("/", "", new_york$year)
```

```r
stations <- stations %>%
  group_by(City)%>%
  mutate(Stations_by_City = sum(row_number()))
```

```r
#Drop two points
stations <- stations %>%
  filter(!(Longitude==-76.04844))

#Unique cities
unique_stations_city_data <- stations %>%
  distinct(City, .keep_all = TRUE) %>%
  dplyr::arrange(-Stations_by_City)

#labels
city_labels <- unique_stations_city_data[1:4,]
```

Introduction

The decision to investigate electric vehicles in New York was one motivated by interest in climate change, policy, and politics. Electric vehicles are being promoted as a way to slow climate change, although little discussion exists around where the electricity to charge these vehicles will come from. Similarly, it is known that new policies are changing what is allowed in and from vehicles; yet many consumers express range anxiety and are not yet ready to fully transition to electric vehicles, claiming the infrastructure is not well-enough established. Lastly, car manufacturing has been tied to politics for years, as it is tied to the manufacturing domestic versus outsourcing discussion. For these reasons, the following analysis has applications to health, government, and the economy. In early 2022, New York State passed a law stating one hundred percent of new passenger cars and trucks offered for sale or lease must be zero-emissions by 2035. This law continued to set objectives for heavy duty vehicles to be zero-emissions by 2045, whenever feasible. Seeking to learn if this seemed possible, our team utilized data regarding electric vehicle registrations over time, as well as electric vehicle charging stations across the state. The first data set utilized contained information regarding when and where the electric vehicle was registered, as well as many interesting details about the specifics of the car. The second data set on electric vehicle charging stations provided important geographical information of charging stations. With these two data sets, it is potentially possible to learn how the presence of charging stations may influence electric vehicle purchasing habits; or rather, if an increase in electric vehicle purchasing is putting pressure on the current charging infrastructure, increasing demand.

Theory and Background

The electric vehicle registration data set is specifically for the state of New York. This data was updated monthly and has been collected since 2018. Having time series data allowed for investigating changes in this market since 2018, as well as comparing electric vehicles to the charging stations over time. Also, this data set provided interesting information regarding the vehicle make, model, and year, allowing us to learn about the vehicle preferences over time. This information proved important to the analysis because we sought to learn how the American market is responding to different makers, knowing many are not domestic. Secondly, the data set on electric vehicle charging stations provides important geographical information, longitude and latitude, and even specifies the station address. This data is also in time-series form and is also updated monthly, which allows for the desired station-to-vehicle comparison over time. With this data, the following analytical questions could be investigated: Is electric vehicle usage truly increasing over time? What relationship exists between electric vehicle usage and charging stations? How have different car manufactures' electric vehicles sold in New York?

Theory

Hypotheses

Within the framework of our project, we are interested in studying if there is a relationship between growth of electric vehicle charging stations in New York state and electric vehicle registration. Null Hypothesis: An increase in the number of electric vehicle charging stations and will have no effect (i.e. show no relationship) on the registration of EV throughout the state of New York. Alternative Hypothesis: Changes in EV charging stations and EV will have a significant effect on New York state EV registration numbers. In addition to this, we will also take on a more focused OLS analysis that aims to examine if charging stations increase impacts all major EV brands in the same fashion. Null Hypothesis: Changes in EV charging station availability throughout New York State will have no relationship with growth of EV brands. Alternative Hypothesis: Changes in EV charging station availability throughout New York State will be correlated with differences in growth of popular EV brands.

Data and Analyses

Data

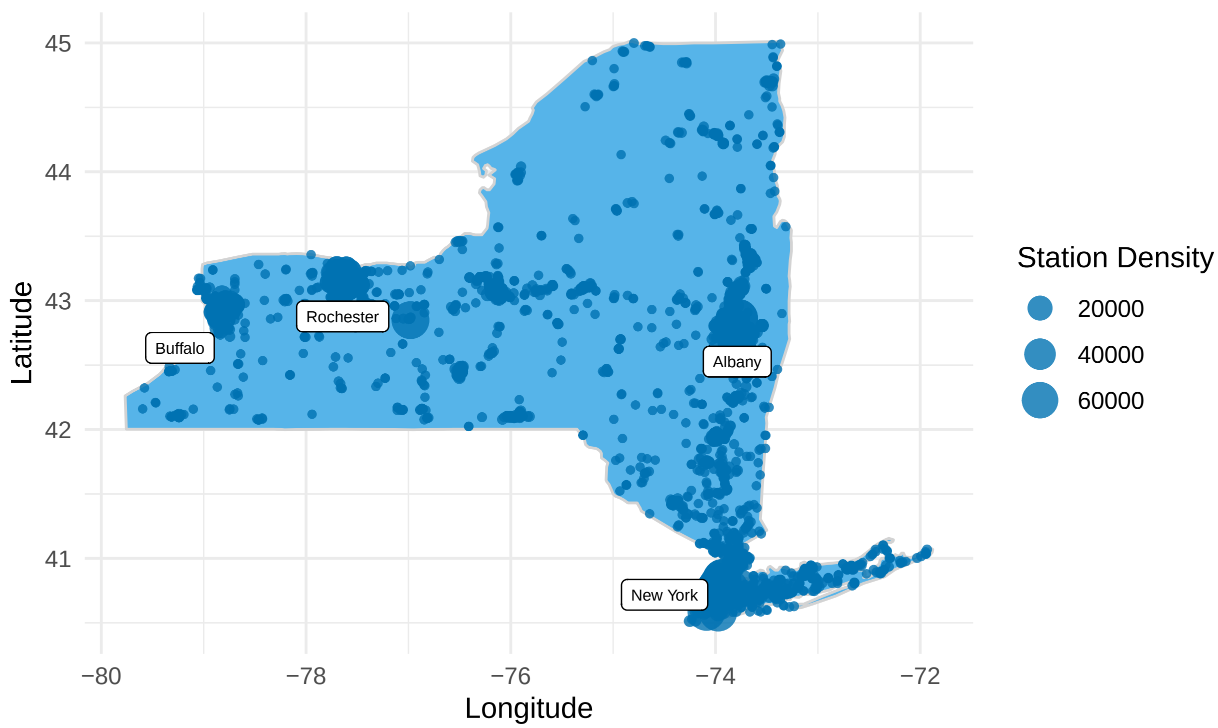
The data we are using for this project focuses on electric vehicle registration and the opening dates of electric charging stations around the state of New York. We obtained the vehicle registration data from the Atlas EV Registration Hub, which aims to provide widely available EV registration data for public policy research. This was time-series data which was interesting to find trends of how the market was shifting. The charging station data was pulled from the public servers of the US Department of Energy. Both data sets have data ranging from 2010 all the way through November 2023, although data collection for 2023 was last completed in August. The station data contains variables such as location, station type, and access, while the registration database contains vehicle characteristics such as make, model, and DMV registration location.

Analyses

```r
# Load the map data for the United States
us_map <- map_data("state")

# Extract New York state data
ny_map <- subset(us_map,
                 region %in% c("new york"))

# Plotting
ggplot() +
  geom_polygon(data = ny_map,
               aes(x = long,
                   y = lat,
                   group = group),
               fill = "#56B4E9",
               color = 'lightgrey') +
  geom_point(data=stations,
             aes(x=Longitude,
                 y=Latitude,
                 size=Stations_by_City),
             color= '#0072B2',
```

```r
               alpha=0.8)+
  geom_label(data = city_labels,
             aes(label = City,
                 x=Longitude,
                 y=Latitude),
             nudge_x=c(-0.5,0,0,-0.5),
             nudge_y=c(0,-0.15,-0.2,-0.3),
             size=2)+
  theme_minimal() +
  labs(title="Electric Vehicle Charging Station Densities \nAcross New York State in 2023",
       y="Latitude",
       x="Longitude",
       size="Station Density",
       caption="Only cities with more than 10,000 stations are labeled")+
  theme(plot.caption=element_text(hjust=0))
```

Electric Vehicle Charging Station Densities
Across New York State in 2023



Only cities with more than 10,000 stations are labeled

The above map shows electric vehicle charging station densities for only New York State. Only the cities with more than 10,000 charging stations in 2023 are labeled: Albany, Buffalo, New York City, and Rochester. We chose to include this visual to provide information about the lack of electric vehicle charging stations across the state, as the majority of station density exists in four core areas. We believe this map gives insight into a potential decrease in electric vehicle purchases in New York. According to Forbes, two of the main reasons people choose not to purchase an electric vehicle are what is commonly known as "range anxiety," and a limited charging network. As this map shows, if one does not live in Albany, Buffalo, New York City, or Rochester, his or her charging options decreases dramatically, by roughly 66%. Continuing, those who work in these heavily populated four cities often live outside the city limits. For this group, range anxiety might be the greatest factor. For those who live in areas with fewer charging options, but are expected to complete longer commutes to work than what is typical, may be more averse to purchasing an electric vehicle. Overall, this map indicates that although many electric vehicle charging stations exist across New York State, they may be servicing the wrong populations.

```
registration <- new_york
stations2 = stations%>%
  filter(State == "NY")%>%
  select(Station.Name, City, State, ZIP, Access.Days.Time, EV.Network, Latitude, Longitude, Open.Date, I

registration$Registration.Date <- as.Date(registration$Registration.Date, format = "%m/%d/%Y")
registration$Year <- format(registration$Registration.Date, "%Y")

new_york2 = registration%>%
  group_by(Vehicle.Make, Year)%>%
  summarise(total = n())

stations2$Open.Date <- as.Date(stations2$Open.Date, format = "%Y-%m-%d")
stations2$Year <- format(stations2$Open.Date, "%Y")
```

```
ny_stations = stations2%>%
  group_by(Year)%>%
  summarise(total_stations = n())%>%
  na.omit(Year)
ny_stations$Year = as.numeric(ny_stations$Year)
new_york2$Year = as.numeric(new_york2$Year)
new_york2 = new_york2%>%
  inner_join(ny_stations, by = c("Year" = "Year"))%>%
  filter(Year != 2023)

new_york2$Vehicle.Make = as.factor(new_york2$Vehicle.Make)

makes = split(new_york2, new_york2$Vehicle.Make)

m = list2env(makes, envir = .GlobalEnv)
```

```
chevy_reg = lm(total ~ total_stations + Year, data = CHEVROLET)
tesla_reg = lm(total ~ total_stations + Year, data = TESLA)
ford_reg = lm(total ~ total_stations + Year, data = FORD)
toyota_reg = lm(total ~ total_stations + Year, data = TOYOTA)
```

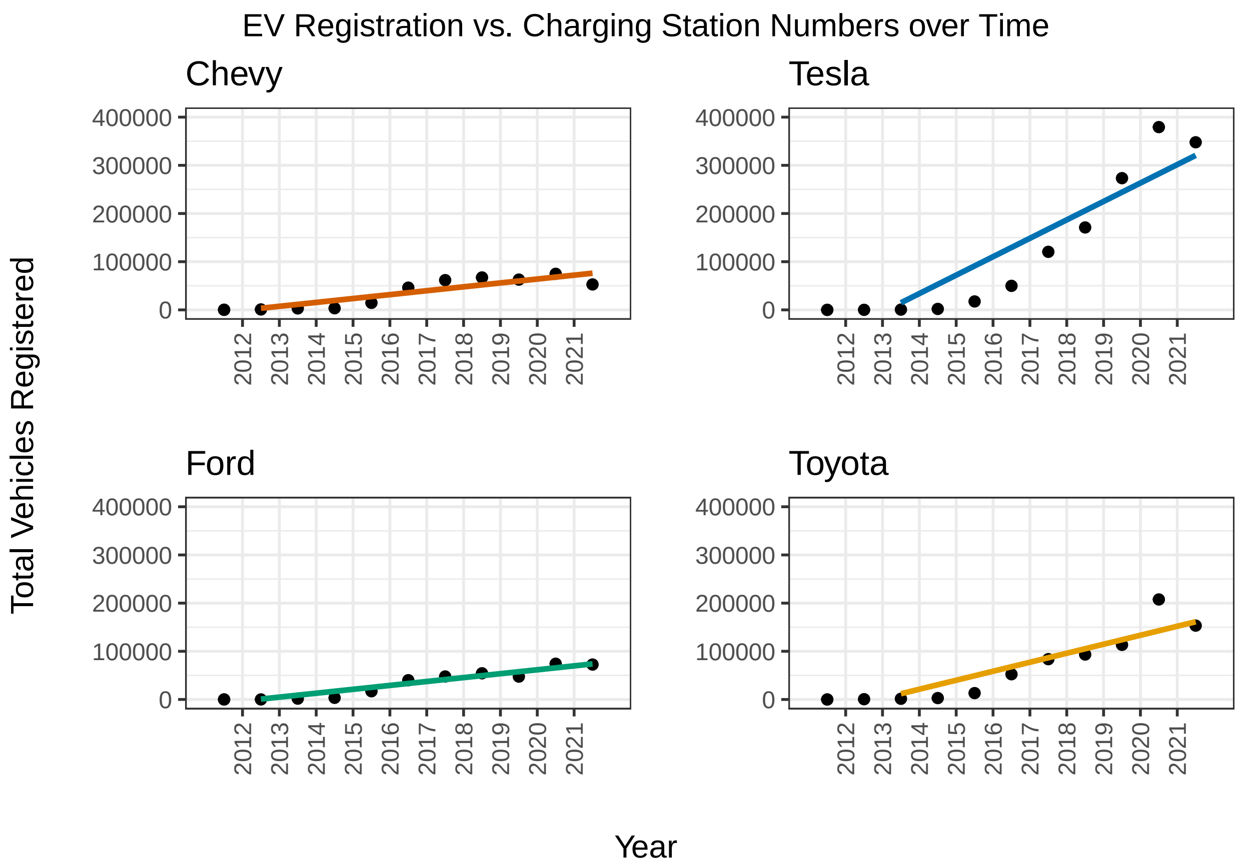
```
makes = c("Chevy", "Tesla", "Ford", "Toyota")
make_list = list(chevy_reg, tesla_reg, ford_reg, toyota_reg)
stargazer(make_list, type = "text", title = "EV Registration Regression by Car Make", column.labels = ma
```

```
##
## EV Registration Regression by Car Make
## ===================================================================================================
##                                                        Dependent variable:
##                              ----------------------------------------------------------------------
##                                                               total
##                                   Chevy             Tesla              Ford              Toyota
##                                    (1)               (2)                (3)                (4)
## ---------------------------------------------------------------------------------------------------
```

```
## total_stations                      -19.020         373.476***          18.455         148.684***
##                                     (28.892)         (61.245)          (18.641)         (42.504)
##
## Year                              9,169.678***      8,167.582       6,354.525***       6,559.714
##                                   (2,488.153)      (5,274.387)      (1,605.349)       (3,660.410)
##
## Constant                       -18,453,400.000*** -16,453,046.000  -12,788,792.000*** -13,205,977.000
##                                 (5,010,815.000)   (10,621,926.000)  (3,232,963.000)   (7,371,588.000)
##
## ----------------------------------------------------------------------------------------------------
## Observations                           12               12               12               12
## R2                                    0.823            0.964            0.918            0.927
## Adjusted R2                           0.784            0.956            0.900            0.910
## Residual Std. Error (df = 9)       14,352.970       30,425.420        9,260.492        21,115.160
## F Statistic (df = 2; 9)            20.982***        120.602***        50.525***         56.792***
## ====================================================================================================
## Note:                                                                  *p<0.1; **p<0.05; ***p<0.01
```

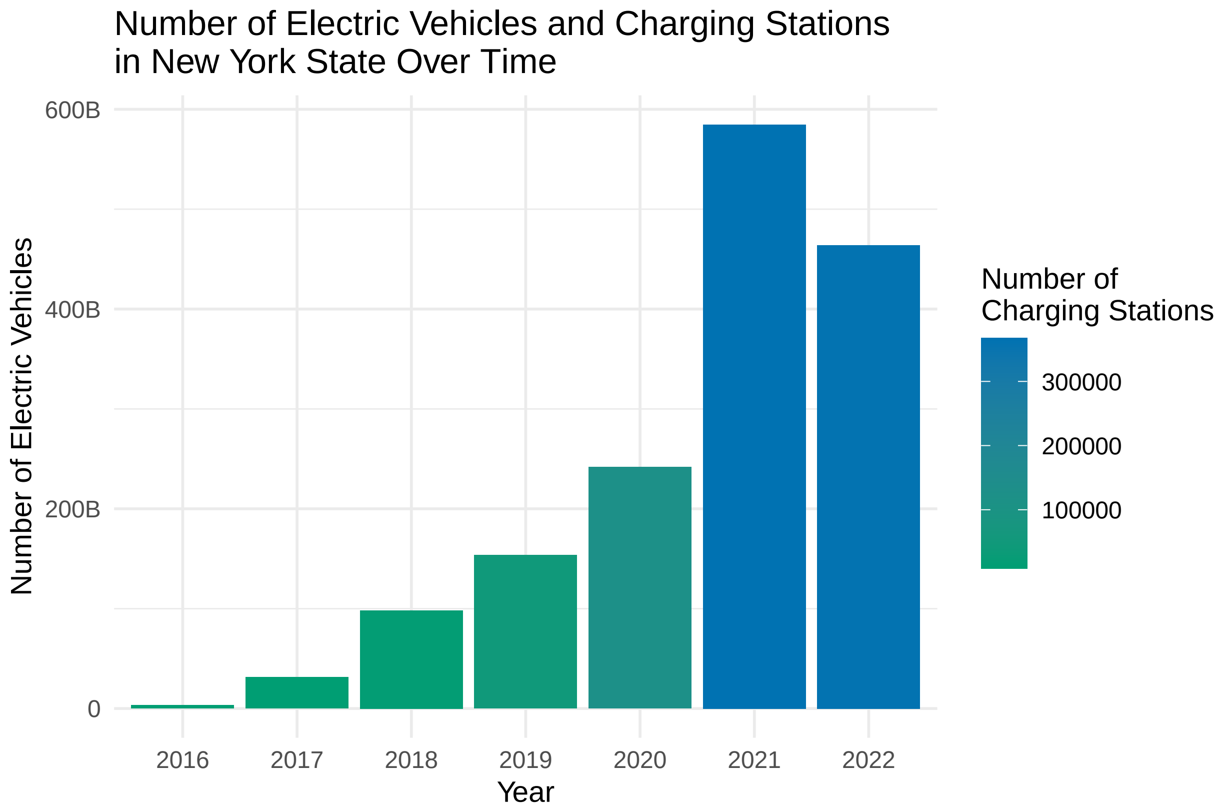
```r
options(scipen = 999)
grid.arrange(
  ggplot(data = CHEVROLET,
         aes(x = Year,
             y = total))+
    geom_point()+
    geom_smooth(method = "lm", se = F, color = "#D55E00")+
    theme_bw()+labs(x = "", y = "", title = "Chevy")+
    scale_x_binned()+
    ylim(0,400000)+
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)),

ggplot(data = TESLA,
       aes(x = Year,
           y = total))+
  geom_point()+
  geom_smooth(method = "lm", se = F, color = "#0072B2")+
  theme_bw()+labs(x = "", y = "", title = "Tesla")+
  scale_x_binned()+
    ylim(0,400000)+
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)),

ggplot(data = FORD,
       aes(x = Year,
           y = total))+
  geom_point()+
  geom_smooth(method = "lm", se = F, color = "#009E73")+
  theme_bw()+
  labs(x = "", y = "", title = "Ford")+
  scale_x_binned()+
  ylim(0,400000)+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)),

ggplot(data = TOYOTA,
       aes(x = Year,
           y = total))+
```

```
  geom_point()+
  geom_smooth(method = "lm", se = F, color = "#E69F00")+
  theme_bw()+
  labs(x = "", y = "", title = "Toyota")+
  scale_x_binned()+
  ylim(0,400000)+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)),
nrow = 2, ncol = 2, top = "EV Registration vs. Charging Station Numbers over Time", left = "Total Vehicl
)
```



The next visual we decided to include was a regression over time of how different make's EV registration was impacted by the increasing number of charging stations. We decided to take a look at the top four brands in the EV market, which included early adopters and industry leaders in Tesla and Toyota as well as their American counterparts Ford and Chevrolet. This was aimed to see if companies with a greater infrastructure in place were more strongly impacted due to their ability to quickly scale up production. Our results showed an effect that seemed to back this idea, as the coefficient on total_stations for Tesla showed an increase of 373.476 EV registrations for each additional charging station that was built that year. Toyota showed an increase of 148.684 registered EV's for each additional charging station. Both values were statistically significant at a one percent alpha level, furthering the importance of these results. Overall, we feel these regression results support the alternative hypothesis that the increase in electric vehicle charging stations is correlated to growth presented by the different popular EV makers. Meanwhile, for Ford and Chevy, the coefficient was much smaller than that of the competitors, while also holding no statistical significance in the model. However, their time coefficient was very similar to that of the Tesla and Toyota, which would seemingly show that they are growing as a whole over time at a similar rate as these other brands. This signals that the companies with a larger background in the EV market are better set up to grow with the increasing number of stations around the state. This could also be due in part to consumer preferences for the brands with longer track records that consumers feel they will more comfortable with range wise.

```r
# Manipulate data
new_york <- new_york %>%
  mutate(year=substr(Registration.Date, 5, 9))

new_york$year <- sub("/", "", new_york$year)

new_york <- new_york %>%
  group_by(year)%>%
  mutate(vehicles_per_year = sum(row_number()))

stations <- stations %>%
  group_by(year) %>%
  mutate(stations_per_year = sum(row_number()))

stations <- stations %>%
  distinct(year,.keep_all=TRUE) %>%
  filter(year %in% 2016:2022)

ny_1 <- new_york %>%
  distinct(year,.keep_all=TRUE) %>%
  filter(year %in% 2016:2022)

stations_1<- stations %>%
  distinct(year,.keep_all = TRUE)

stacked_plot <- merge(ny_1,stations_1)

# Grouped
options(scipen = 999)
ggplot(stacked_plot, aes(x=year,
                         y=vehicles_per_year,
                         fill=stations_per_year)) +
    geom_bar(position="dodge",
             stat="identity")+
  scale_fill_continuous(low="#009E73",high="#0072B2")+
  theme_minimal()+
  labs(y='Number of Electric Vehicles',
       x="Year",
       fill="Number of \nCharging Stations",
       title="Number of Electric Vehicles and Charging Stations \nin New York State Over Time",
       caption='Data for 2023 was excluded because only observations through August existed.')+
    scale_y_continuous(labels = scales::label_number_si())+
  theme(plot.caption=element_text(hjust=0))
```



Data for 2023 was excluded because only observations through August existed.

Finally, this third visualization shows the relationship between the number of electric vehicles registered in New York compared to the number of electric vehicle charging stations. We chose to display this relationship to further challenge our hypothesis of whether or not electric vehicle charging station numbers will have a zero effect on the number of electric vehicles registered. Overall, the the number electric vehicles has increased from 2016 to 2021, and experienced a sharp fall in 2022; and over the same time period the number of electric vehicle charging stations has continued to increase. This would support, but not yet confirm, our null hypothesis, as 2022 posed a sharply inverse relationship between the two figures. With the support of the map, we know range anxiety still may be a factor contributing to why consumers' purchases of electric vehicles has slowed. This graph, however, points to other factors, potentially including unfamiliarity with the product or financial restraints, again according to Forbes. Perhaps this graph is detailing that the early adopters have been satisfied, and significant changes will need to be made to declare electric vehicles preferred by consumers and the industry standard.

Conclusion

Although we feel confident in this body of analysis, one limitation to consider is that this may only apply to the state of New York and not the nation, nor the electric vehicle market, as a whole. Additionally, we were not able to investigate all variables that could influence the number of electric vehicles registered. For example, this analysis did not cover the impact of vehicle cost on the number of vehicles registered, which we feel could have a significant impact considering electric vehicles are still considerably more expensive than their gas-fueled alternatives. Overall, we feel we presented a strong case for the lack of correlation between electric vehicle charging stations and electric vehicles. Up until 2022, this relationship seemed to be strongly, positively correlated, but recent data has weakened this relationship. Similarly, a natural experiment or a regression controlling for many observables would be required to make a statement about causality. In regards to the Maker regression models, we feel more would need to be accounted for and removed from the error to declare this relationship as a causal one. Overall, building new charging stations (at least in current popular areas) is not enough to increase the use of electric vehicles. For the desired climate, health, and economic effects to be achieved, different incentives must be applied in conjunction, and perhaps more foreign made vehicles be imported, as this is what the market currently prefers.