California 2022 Proposition 30 Analysis - GitData

Weilin Chen, University of California Davis
Hengyuan Liu, University of California Davis
Kathy Mo, University of California Davis
Li Yuan, University of Michigan Ann Arbor

10/24/2022

Contents

I. Introduction and Research Questions

California has been a pioneer in U.S. air quality legislation since the 1940s. In 2006 California passed Assembly Bill 32 (AB 32), which created a comprehensive, multi-year program to reduce greenhouse gas (GHG) emissions in California. In 2050, the emission of greenhouse gas should be cut to 80% below the 1990 level[1]. According to the California Air Resource Board, transportation produced about 41% of greenhouse gas emissions in 2019[2]. Moreover, the California Air Resource Board reported that the wildfire produced 106.7 million metric tons of carbon dioxide in 2020 and 85.1 million metric tons in 2021[3]. The state wants to decrease greenhouse gas emissions by having more electric cars and preventing wildfires. This is what Proposition 30 is going to do.

Proposition 30 is going to increase a 1.75% tax on Californians who earn above $2 million per year to fund programs to increase the amount of zero emission vehicles and prevent wildfires. The tax plans to start in 2023 and will end by 2043 or earlier if we can maintain the expected amount of greenhouse gas emissions for three consecutive calendar years.

80 percent of the tax revenue will be invested in the Zero Emission Vehicle (ZEV) program. Specifically, 45 percent in helping people buy electric vehicles and 35 percent in installing more charging stations[4]. The state government decided to dedicate an additional 10 billion dollars[1] over the next five years to the program in this year's budget.

20 percent of the tax revenue will be spent on Wildfire Response and Prevention Programs. In general, most of this money will be used to hire, train, and retain state firefighters. The rest of the money could be used for other wildfire response and prevention activities[4].

We will figure out whether more electric vehicles can help reduce greenhouse gas emissions. Based on the tax revenue and the additional $10 billion funded by the government, we will find out how many electric vehicles can be used and the amount of greenhouse gas that can be reduced. We will find out how much greenhouse gas can be reduced after funding the Wildfire Response and Prevention Programs. Finally, we will evaluate if our goal can be achieved, which is whether greenhouse gas emissions can be cut to 80% below the 1990 level by 2050.

II. Data Description & Visualization

Before doing the analysis, it is necessary to introduce and visualize the datasets we will use. Generally speaking, we have three major datasets to analyze.

1. EMission FACtor (EMFAC) 2021

The first dataset we use is from the California Air Resource Board called EMission FACtor (EMFAC) 2021 [5], which is the official emissions inventory of on-road mobile sources in California. We only consider light-duty vehicles (passenger cars) because the proposition only helps individuals to buy electric vehicles.

According to the United States Environmental Protection Agency (EPA), greenhouse gas is mainly composed of Carbon dioxide, Methane, and Nitrous oxide[6], so we only consider these three pollutants in our analysis against the fuel types of cars. In order to analyze more accurately, we use the average emission of Carbon dioxide, Methane, and Nitrous oxide per mile per car in a year to be our response variables. We also add up these three pollutants as the average amount of greenhouse gas emitted per mile per car in a year.

i. Data Description

- `Fuel`: The types of fuel each car uses. There are four types of fuel: `Diesel` (37,968 counts), `Electricity` (27,314 counts), `Gasoline` (40,140 counts), and `Plug-in Hybrid` (11,604 counts).

- `CO2_AVG`: The amount of Carbon dioxide, in million metric tons, emitted per mile per car in 2021.

- `CH4_AVG`: The amount of Methane, in million metric tons, emitted per mile per car in 2021.

- `N2O_AVG`: The amount of Nitrous oxide, in million metric tons, emitted per mile per car in 2021.

- `GHG_AVG`: The amount of greenhouse gas (the sum of Carbon dioxide, Methane, and Nitrous oxide), in million metric tons, emitted per mile per car in 2021.

Table 1: Summary Statistics of Fuel Type and Pollutants

Fuel Type	Count	CO2 Total	CH4 Total	N2O Total	GHG Total	CO2 Mean	CH4 Mean	N2O Mean	GHG Mean
Diesel	37968	0.1663633	0.0000013	0.0000262	0.1663908	0.0000044	0	0	0.0000044
Electricity	27314	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0	0	0.0000000
Gasoline	40140	58.4992165	0.0007023	0.0011421	58.5010608	0.0014574	0	0	0.0014574
Plug-in Hybrid	11604	0.6878281	0.0000023	0.0000030	0.6878334	0.0000593	0	0	0.0000593

The table shown above (Table 1) tells us how many cars there are in each type, and how many pollutants they emitted on average and in total. The total greenhouse gas emitted in 2021 is 59.3552849 million metric tons.
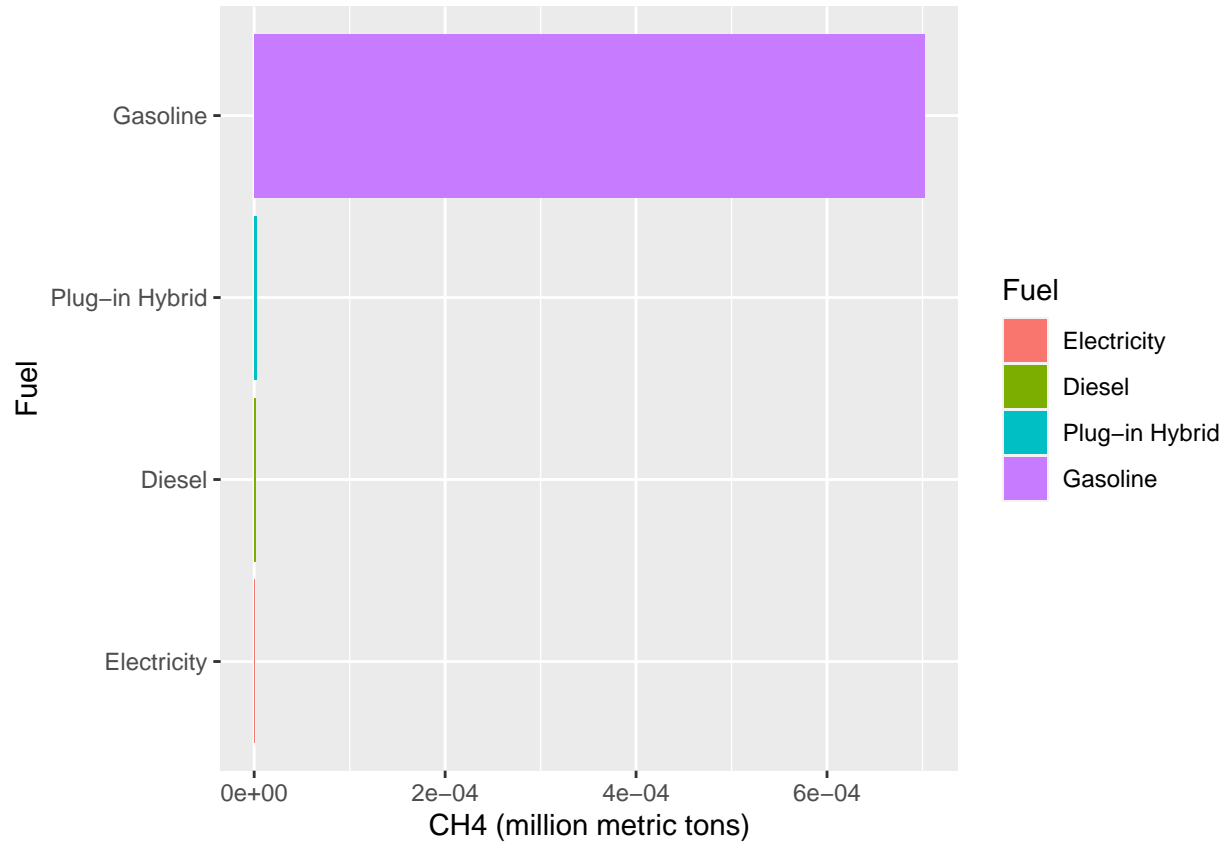
ii. Data Visualization

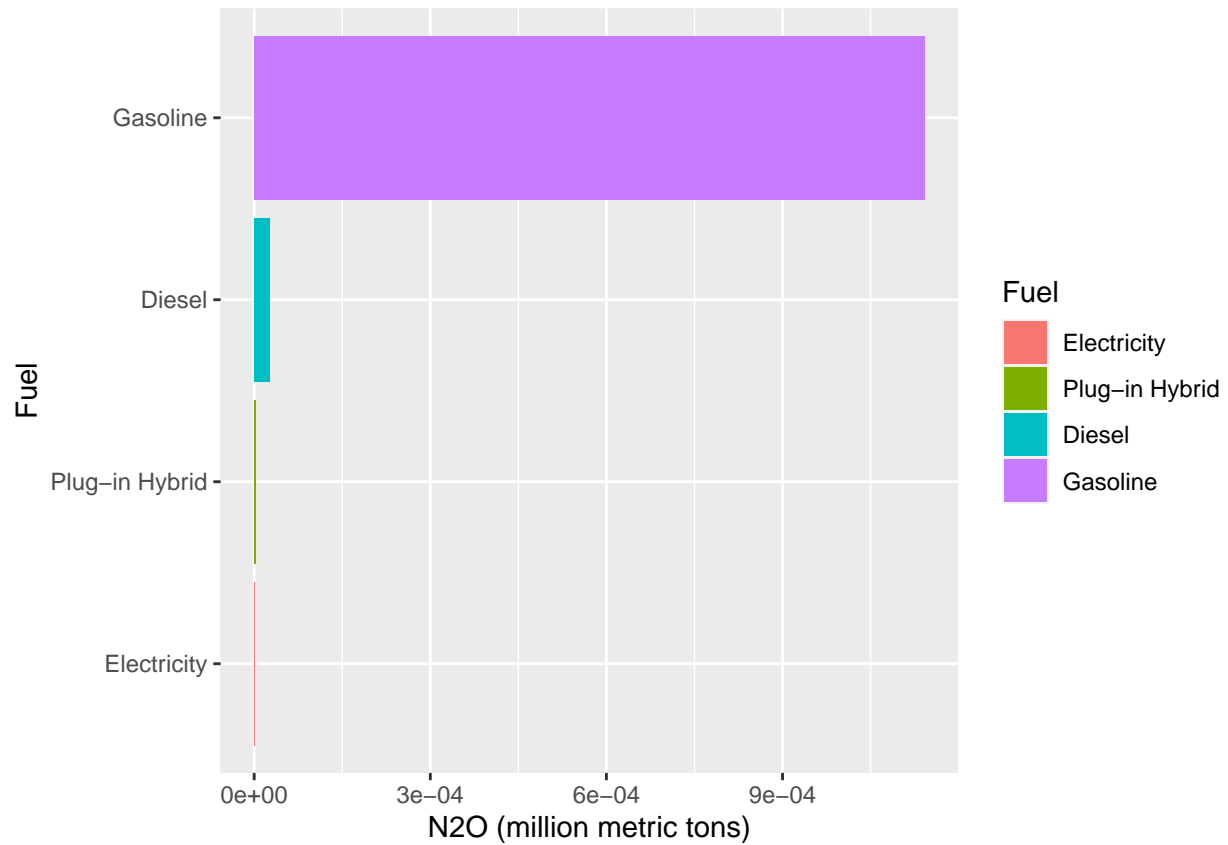Figure 1: Bar Chart of CO2 Emission by Fuel Types
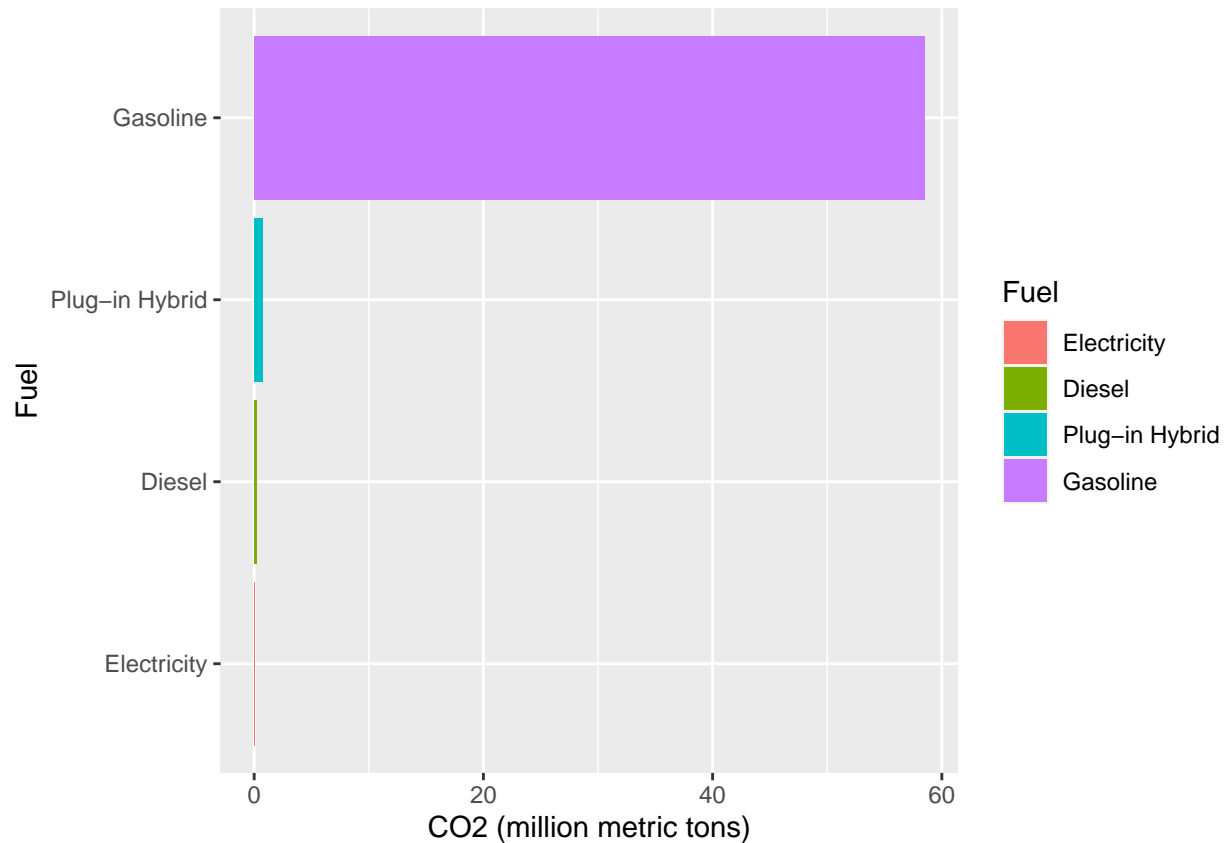
This bar chart shown above (Figure 1) is about the CO2 emissions by passenger cars each year in 2021. The x-axis represents the amount of CO2 generated by each type of car. The y-axis represents the types of fuel each car uses. We can see that the CO2 emission by gasoline cars is the highest. Electric cars emitted zero CO2.

Figure 2: Bar Chart of CH4 Emission by Fuel Types

This bar chart shown above (Figure 2) is about the CH4 emissions by passenger cars each year in 2021. The x-axis represents the amount of CH4 generated by each type of car. The y-axis represents the types of fuel each car uses. We can see that the CH4 emission by gasoline cars is the highest. Electric cars emitted zero CH4.

Figure 3: Bar Chart of N2O Emission by Fuel Types

This bar chart shown above (Figure 3) is about the N2O emissions by passenger cars each year in 2021. The x-axis represents the amount of N2O generated by each type of car. The y-axis represents the types of fuel each car uses. We can see that the N2O emission by gasoline cars is the highest. Electric cars emitted zero N2O.

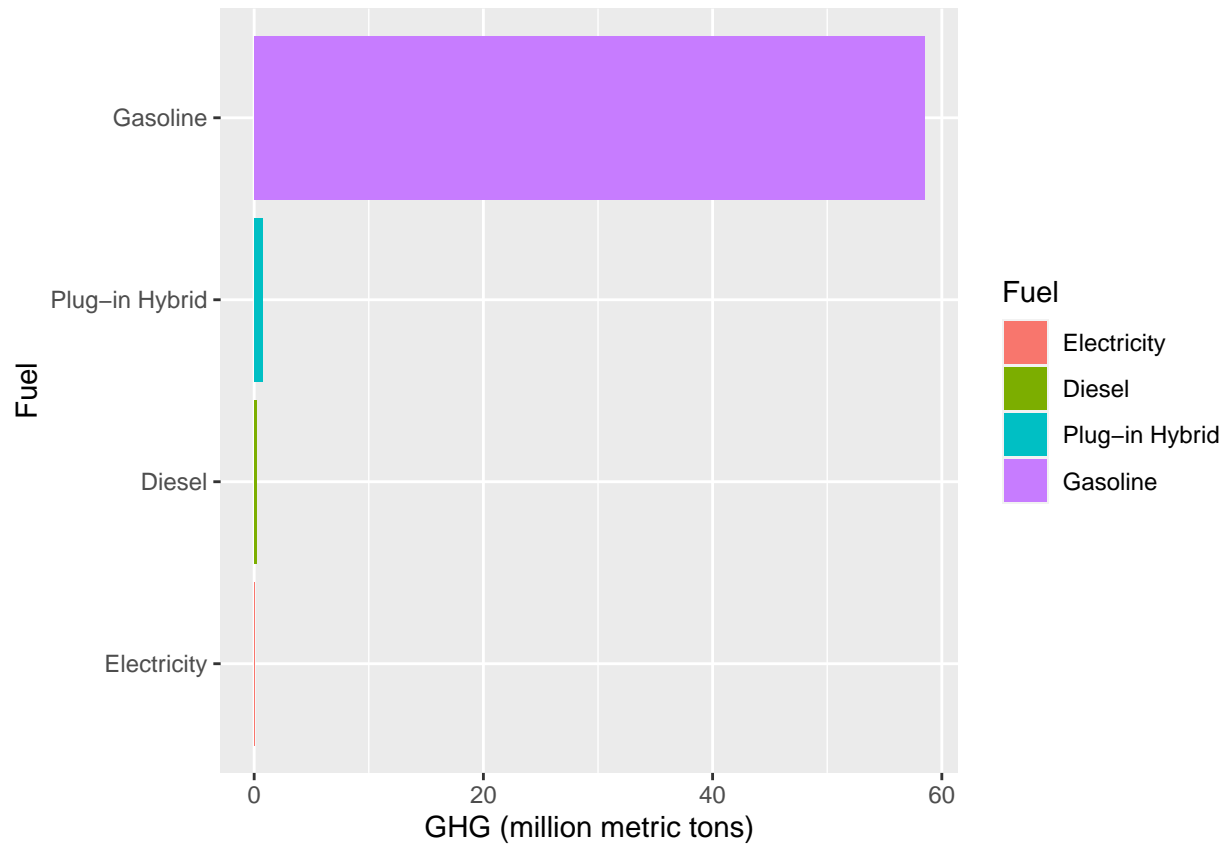Figure 4: Bar Chart of Greenhouse Gas (GHG) Emission by Fuel Types

This bar chart shown above (Figure 4) is about the GHG emissions by passenger cars each year in 2021. The x-axis represents the amount of GHG generated by each type of car. The y-axis represents the types of fuel each car uses. We can see that the GHG emission by gasoline cars is the highest. Electric cars emitted zero GHG.

Figure 5: Box Plot of Average CO2 Emission by Fuel Types
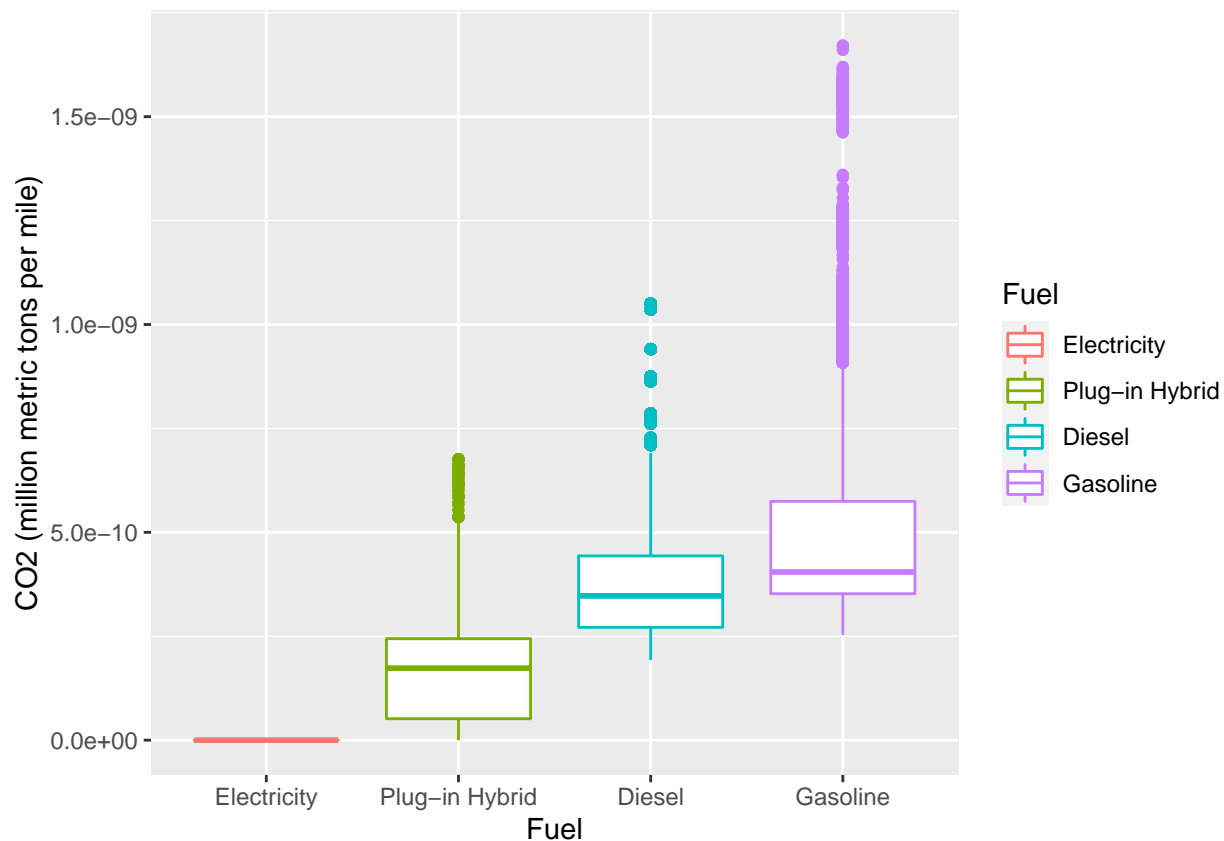
The boxplot shown above (Figure 5) is about the amount of carbon dioxide, in million metric tons, emitted per mile per car in 2021. We observe that gasoline is the main producer of CO2. It has the highest median of carbon dioxide and the most outliers. Diesel is second, and the median between gasoline and diesel is relatively close, especially when compared to the difference between the plug-in hybrid median and diesel median. The larger difference between plug-in hybrid and diesel in outliers, median, and distribution indicates that plug-in hybrid is less similar than diesel and electricity. Meanwhile, electricity does not have a distribution at all, since it does not emit any carbon dioxide. Gasoline and diesel have a right-skewed distribution, while plug-in hybrid has a left-skewed distribution. These differences suggest that gasoline and diesel emit more carbon dioxide than the median.

Figure 6: Box Plot of Average CH4 Emission by Fuel Types
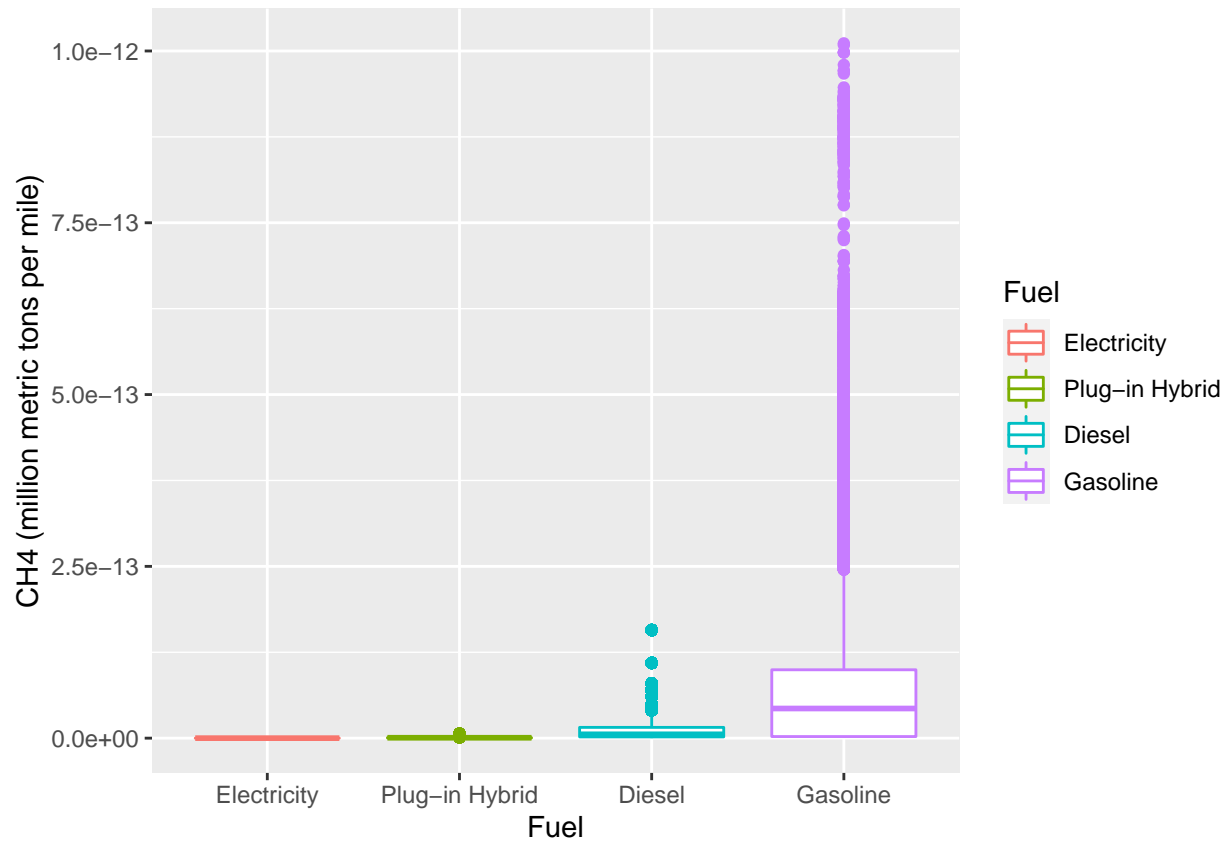
From Figure 6, gasoline produces much more methane emissions than any of the other fuel types. Electricity, as with the other emission types, does not emit any greenhouse gasses. Gasoline has a heavy right tail, a right-skewed distribution, and far more outliers than diesel and plug-in hybrid.

Figure 7: Box Plot of Average N2O Emission by Fuel Types
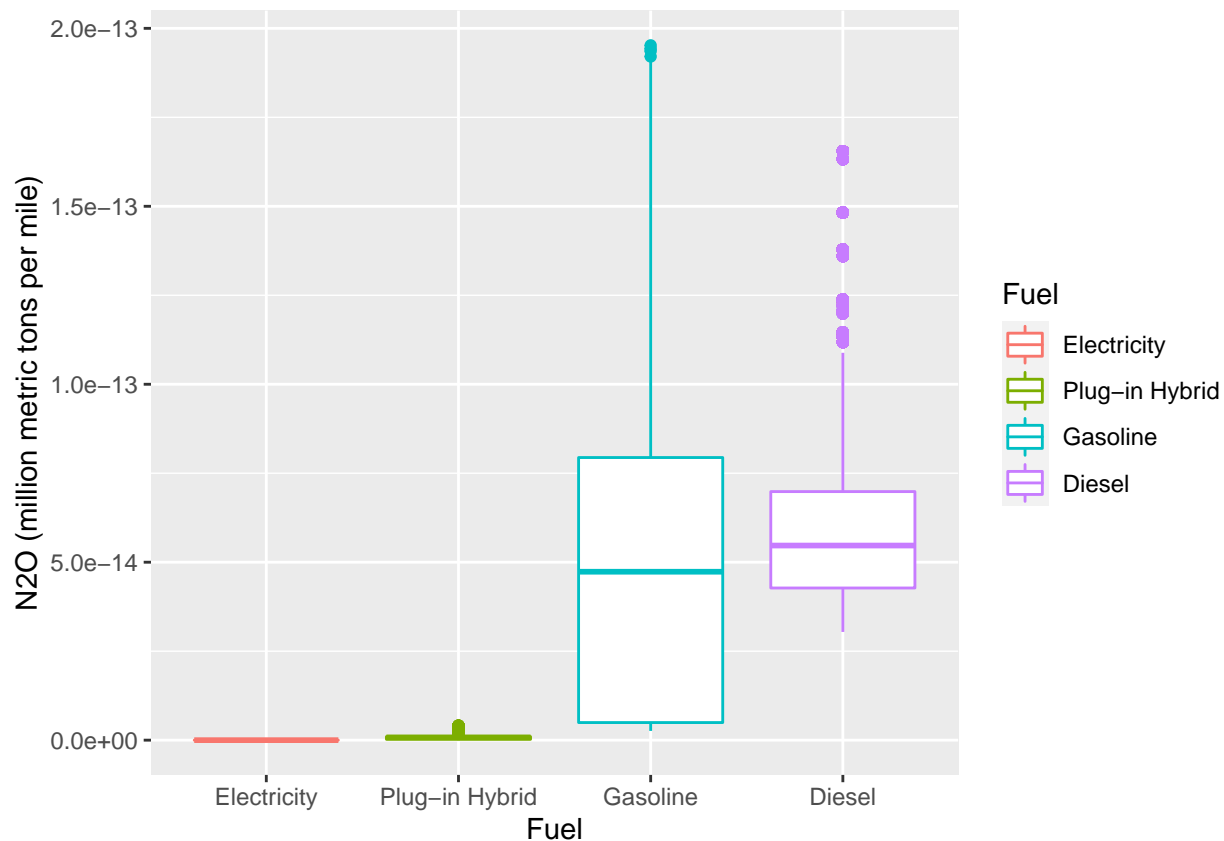
According to the boxplot above (Figure 7), we can see that the median of NO2_AVG from diesel is a little bit higher than Gasoline. Electricity and Plug-in Hybrid produce a negligible amount of N2O. Furthermore, Gasoline has the largest distribution of N2O compared with the other fuel types. Gasoline has fewer outliers than diesel, and both have a heavy right tail. However, gasoline has a left-skewed distribution while diesel has a right-skewed distribution. Plug-in hybrid has few outliers.

Figure 8: Box Plot of Average GHG Emission by Fuel Types
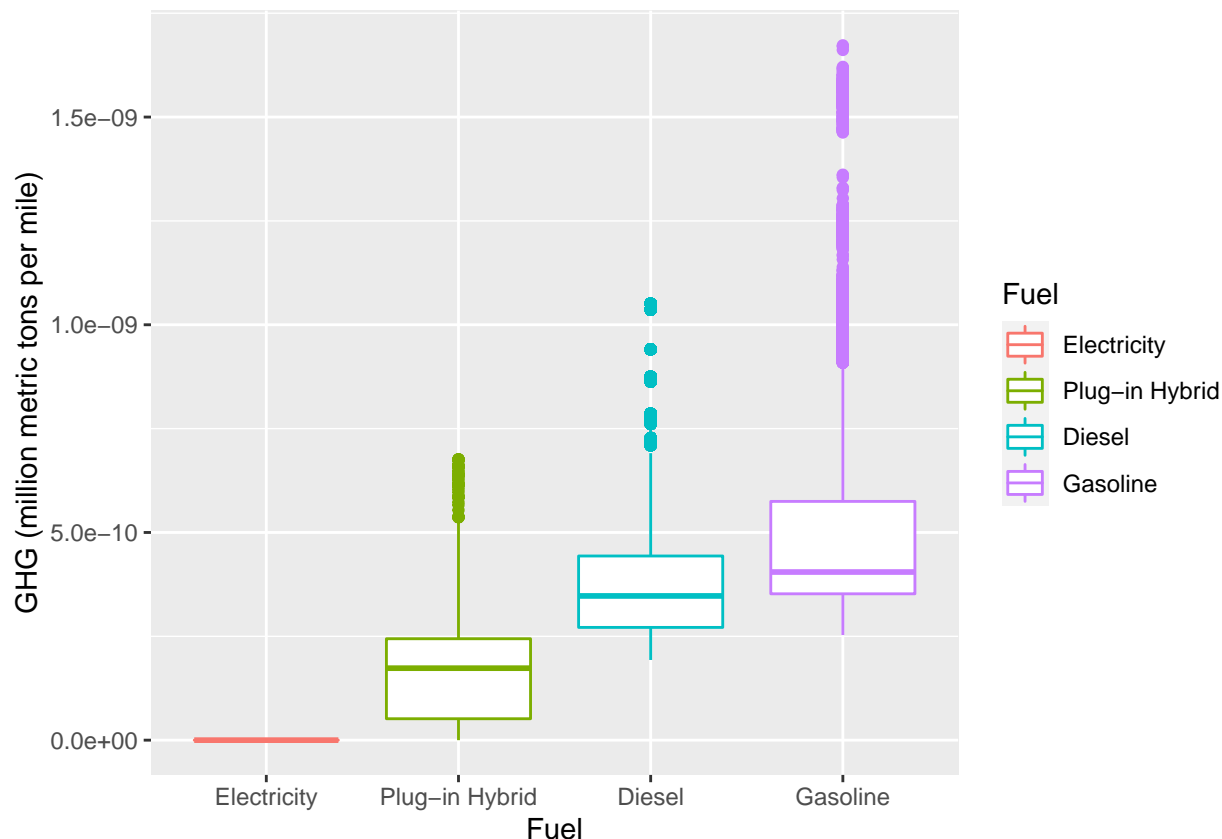
Overall, from Figure 8, gasoline is the largest contributor to greenhouse gas emissions, with diesel second, plug-in hybrid third, and electricity last. Gasoline has the most outliers and also has a right-skewed distribution. Diesel has a heavy right tail, and plug-in hybrid has a left-skewed distribution. Throughout all the plots and this overall plot, electricity has always shown it doesn't emit any greenhouse gasses.

2. Personal Income Tax Data 2020

The second dataset is from the Franchise Tax Board of California[7], which contains the number of taxpayers in California, and the federal adjusted gross income (AGI), which is the figure that the Internal Revenue Service (IRS) uses to determine your income tax liability for the year, for each AGI class. The dataset also divides people into three classes: AGI below 39,999 is lower class, between 40,000 and 149,999 is middle class, and above 150,000 is upper class[8]. The original unit for AGI is thousand dollars. In order to make our analysis more accurate, we use dollars as the unit for AGI. This is the data in 2020, which is the latest we can get.

i. Data Description

- `Adjusted.Gross.Income.Class`: The groups of people based on their adjusted gross income (59 counts).

- `Number.of.Return_Taxable`: The number of people who are eligible to be taxed (11,332,757 counts).

- `Federal.AGI`: The adjusted gross income, in dollars, of each AGI class.

- `Class`: The general classes (3 counts) people in.

Table 2: Californians Who Earn More than 2 Million in 2020

	AGI Class	People Taxable	Federal AGI
56	2,000,000 to 2,999,999	14911	34229385000
57	3,000,000 to 3,999,999	6280	20578127000
58	4,000,000 to 4,999,999	3479	14510953000
59	5,000,000 and over	10335	150335815000
60	Total:	35005	219654280000

The table shown above gives us all the Californians whose AGI are above 2,000,000 dollars with the number of them and their total AGI, which is 2.1965428×10^{11} in 2020.

ii. Data Visualization

Figure 9: Bar Chart of Number of People Taxable by Classes
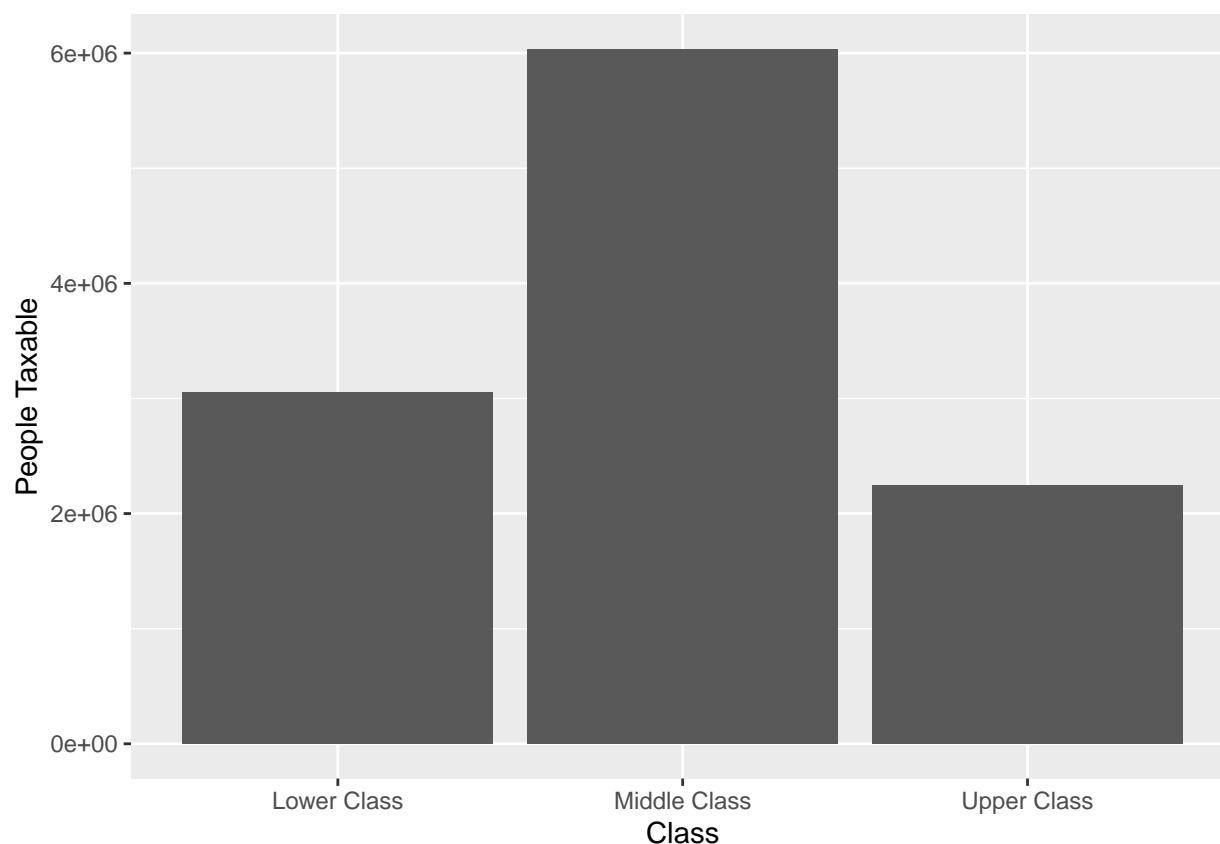
The bar chart shown above (Figure 9) is about the number of people by classes in California. The y-axis represents the number of people. The x-axis represents the three classes. We can see that there are a lot of middle class people in California.

Figure 10: Bar Chart of Federal AGI by Classes

The bar chart shown above (Figure 10) is about the Federal AGI by classes in California. The y-axis represents the total Federal AGI. The x-axis represents the three classes. We can see that the upper class takes up the most wealth.
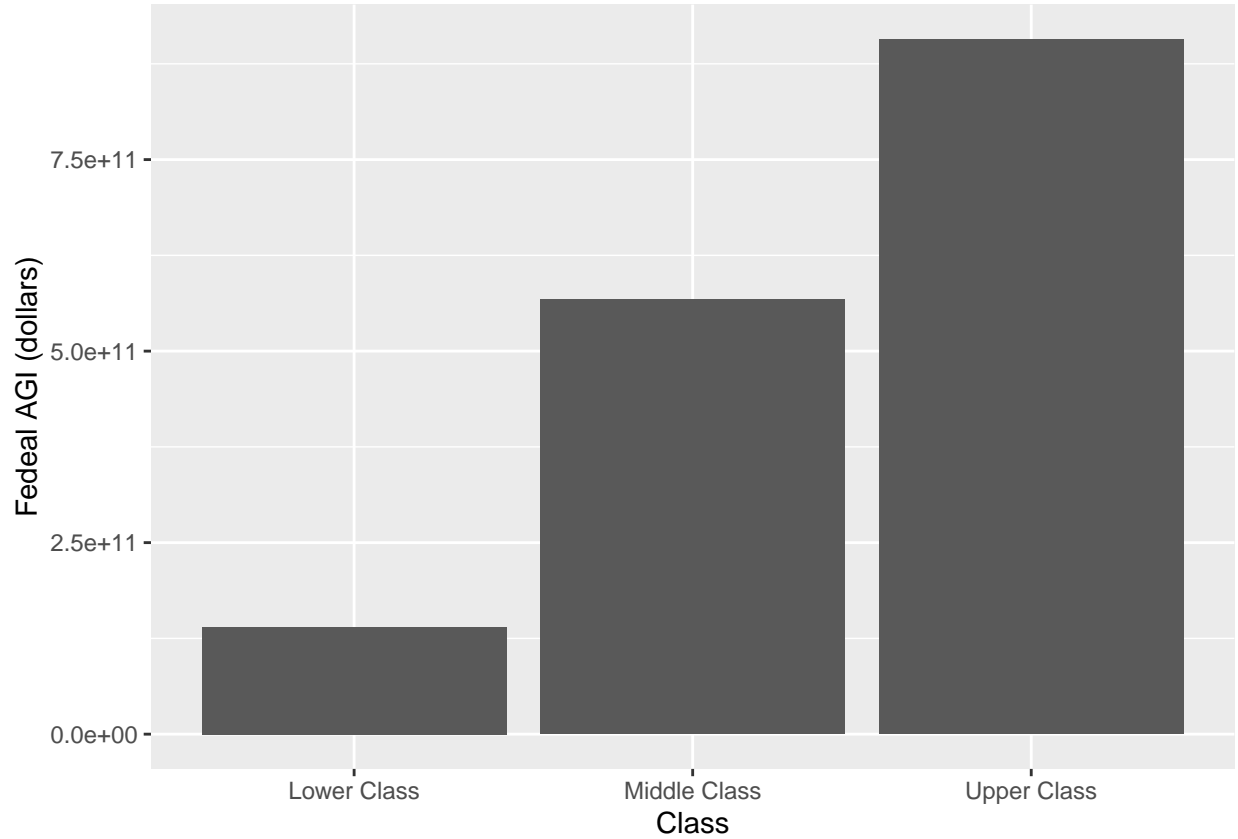
3. Wildfire Response and Prevention Programs fundings 2021

We collect data about two types of funds: Special funds and General fund from the government to the two departments Office of the State Fire Marshall and Department of Forestry and Fire Protection from 2000-2021[9] and the corresponding Carbon dioxide, which is from California Air Resource Board[3].

Special funds are established by a government to collect money that must be used for a specific project. General fund is the primary fund used by a government entity.

Office of the State Fire Marshall (OSFM)'s purpose is to protect life and property through the development and application of fire prevention, engineering, training and education, and enforcement. The Department of Forestry and Fire Protection serves and safeguards the people and protects the property and resources of California.

i. Data Description

- CO_2: The amount of Carbon dioxide produced by wildfires from 2000 to 2021, in million metric tons.

- Fund: The money funded, in dollars, by the government to the Office of the State Fire Marshall and the Department of Forestry and Fire Protection each year from 2000 to 2021.

- Year: The calender years (22 counts) from 2000 to 2021 that wildfire occurred and government funded for the State Fire Marshall and the Department of Forestry and Fire Protection.

ii. Data Visualization

Figure 11: Bar Chart of Fund for Wildfire Protection by Years
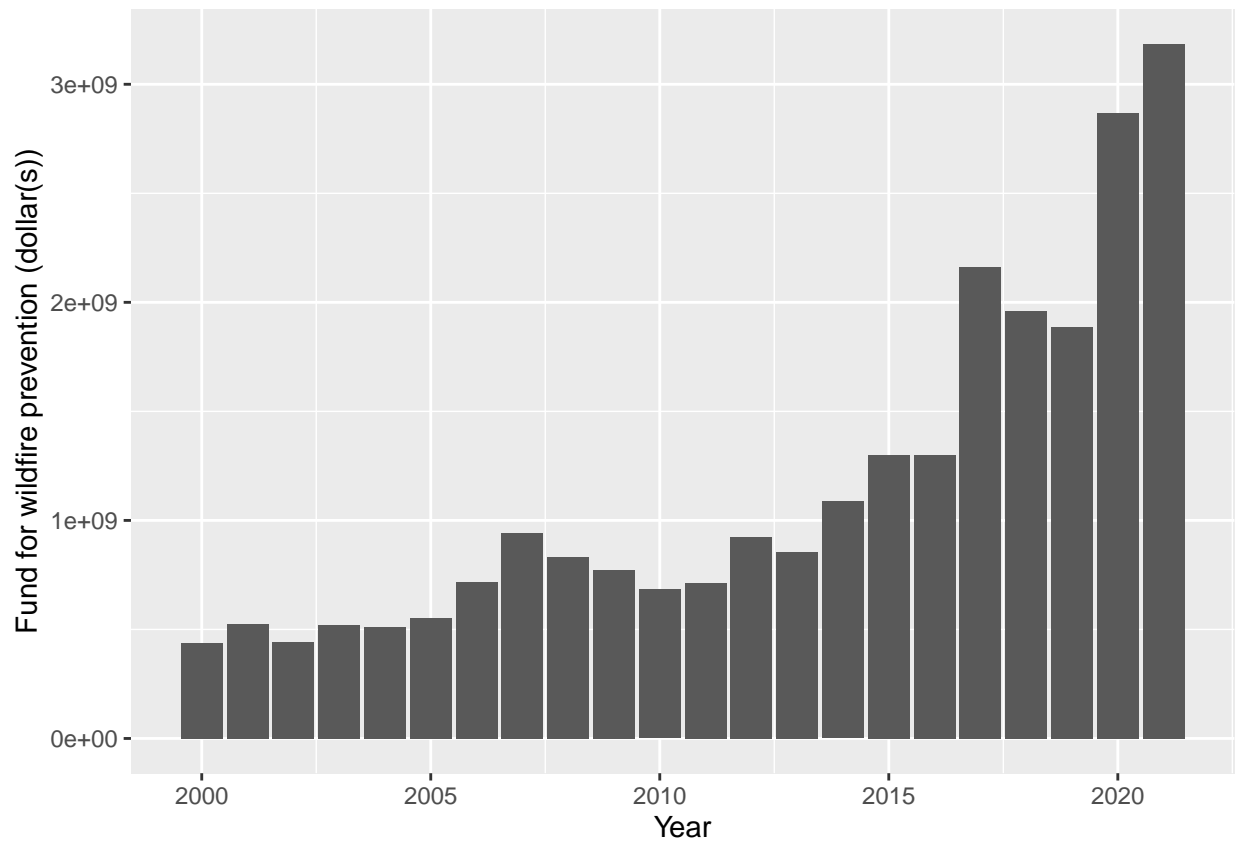
The bar chart shown above (Figure 11) is about the California government providing the Funds for wildfire prevention from 2000 to 2021. The x-axis represents the year of funding. The y-axis represents the amount of funds provided by the state government. We can see that in general, as time goes by, the funding for wildfire prevention increases.

Figure 12: Bar Chart of CO_2 Generated by Wildfire by Years
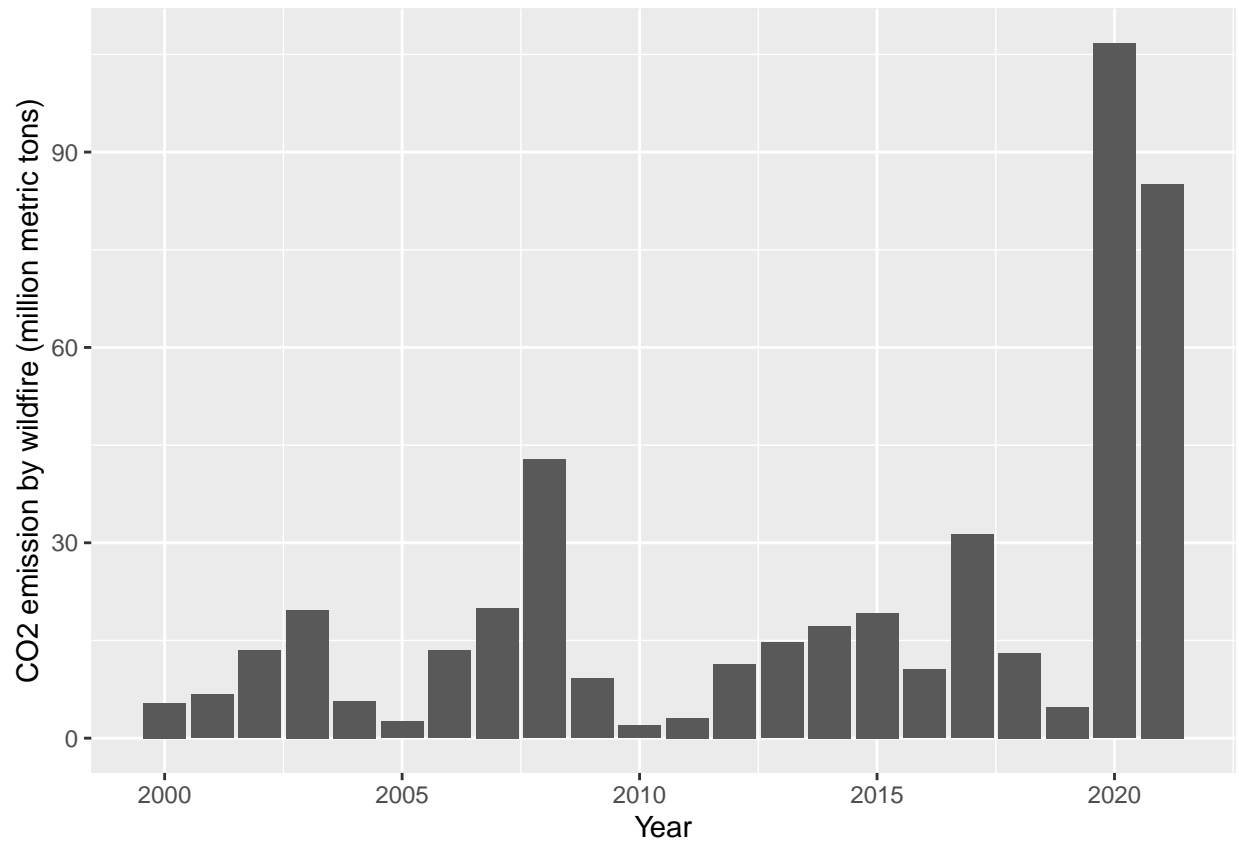
This bar chart shown above (Figure 12) is about the California CO2 emissions by wildfires each year from 2000 to 2021. The x-axis represents the years of wildfires. The y-axis represents the amount of CO2 generated by the wildfires. We can see that as time goes by, the CO2 emissions by wildfires increases.

Figure 13: Box Plot of CO2 Generated by Wildfire
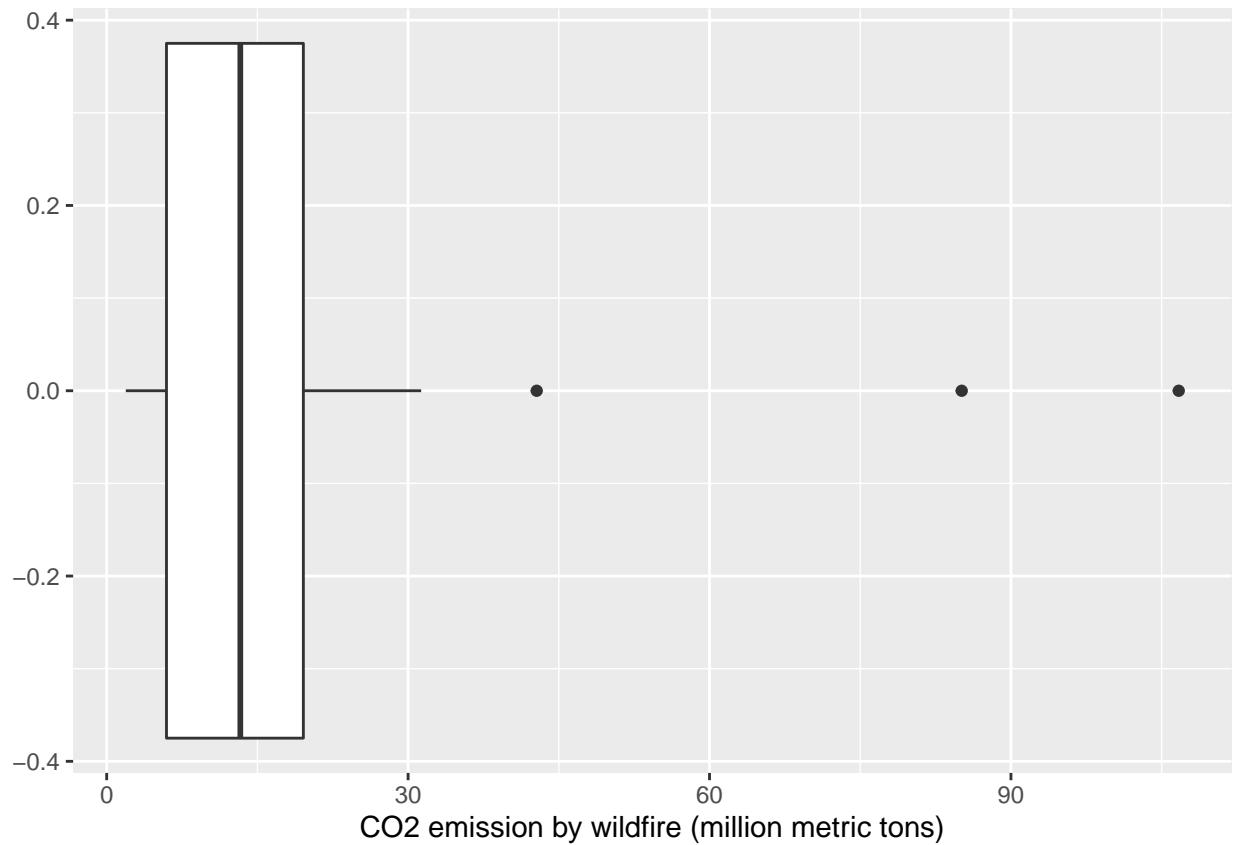
[1] 13.3

The box plot shown above (Figure 13) is about how the CO2 emissions from wildfires from 2000 to 2021 is distributed. The median of CO2 emitted by wildfire from 2000 to 2021 is 13.3 million metric tons. There seems to be three outliers, which are 42.8 million metric tons in 2008, 106.7 million metric tons in 2020, and 85.1 million metric tons in 2021.

Figure 14: Scatter Plot of CO2 Generated by Wildfire and Fund for Wildfire Protection

`geom_smooth()` using formula 'y ~ x'
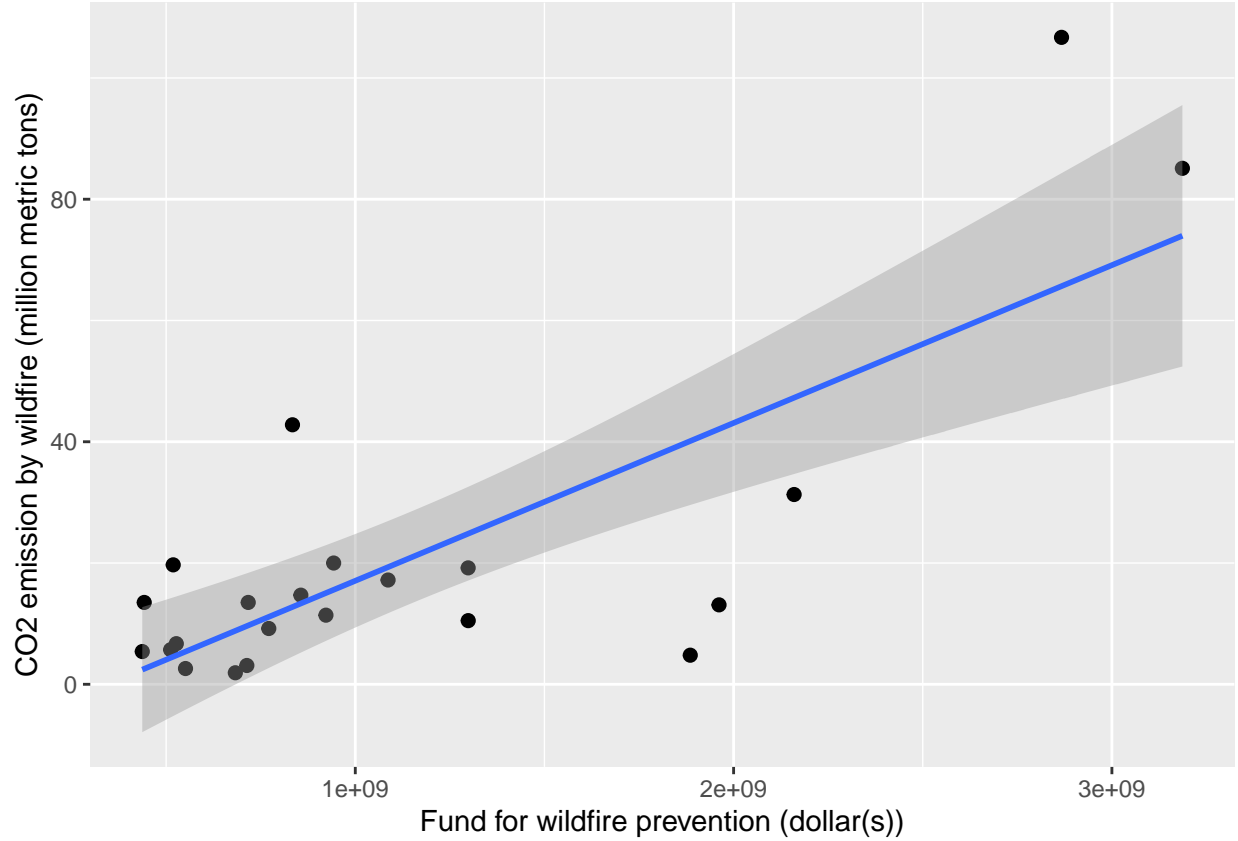
The scatter plot shown above (Figure 14) is about the relationship between California government funded wildfire prevention (x-axis) and CO2 California CO2 emissions from wildfires (y-axis) from 2000 to 2021. We can see that when funding increases CO2 emitted by wildfires also increases. As a result, they might be positively correlated.

III. Analysis and Interpretation

1. Analyze fuel type and greenhouse gas, including CO2, CH4, and N2O

In order to figure out whether there are relationships between fuel types and greenhouse gas, including CO2, CH4, and N2O, we need to use linear regression. Specifically, we will build four linear regression models seaperately on CO2, CH4, N2O, and GHG.

Our linear regression model for CO2 shows that the residuals, which is the difference between the actual and predicted values, is very small. The median residual is zero, which indicates that the residuals are symmetrical and the model is predicting evenly on both ends of our dataset. Our t-value is large for all variables, so the standard error is small compared to our coefficients. We can see that each variable has three asterisks, which indicates that the coefficients are significant and therefore are useful for the model. The small residual standard error tells us that our model is close to the actual values. The multiple R squared is 0.5969, which means that fuel explains 59.69% of the variation within carbon dioxide. Our model rejects the null hypothesis. Therefore, there is a relationship between fuel and carbon dioxide.

If one unit increase in the Electricity, and holding all the other predictors fixed, the `CO2_AVG` will decrease by 3.983×10^{-10} million metric tons per mile on average.

Our linear regression model for CH4 shows that the residuals, which is the difference between the actual

and predicted values, is very small. The median residual is zero, which indicates that the residuals are symmetrical and the model is predicting evenly on both ends of our dataset. Our t-value is relatively small for all variables. We can see that each variable has three asterisks, which indicates that the coefficients are significant and therefore are useful for the model. The small residual standard error indicates that our model is close to the actual values. The multiple R squared is 0.1925, which means that fuel explains 19.25% of the variation within methane. Our model rejects the null hypothesis since the p-value is smaller than any common alpha. Therefore, there is a relationship between fuel and methane.

If one -unit increase in the Electricity, and holding all the other predictors fixed, the `CH4_AVG` will decrease by 1.296×10^{-14} million metric tons per mile on average.

Our linear regression model for N2O shows that the residuals, which is the difference between the actual and predicted values, is very small. The median residual is zero, which indicates that the residuals are symmetrical and the model is predicting evenly on both ends of our dataset. Our absolute t-value is large for all variables, so the standard error is small compared to our coefficients. We can see that each variable has three asterisks, which indicates that the coefficients are significant and therefore are useful for the model. The small residual standard error tells us that our model is close to the actual values. The multiple R squared is 0.4318, which means that fuel explains 43.18% of the variation within carbon dioxide. Our model rejects the null hypothesis. Therefore, there is a relationship between fuel and nitrous oxide.

If one -unit increase in the Electricity, and holding all the other predictors fixed, the `N2O_AVG` will decrease on average by 6.27×10^{-14} million metric tons per mile.

Our linear regression model for GHG shows that the residuals, which is the difference between the actual and predicted values, is very small. The median residual is zero, which indicates that the residuals are symmetrical and the model is predicting evenly on both ends of our dataset. Our t-value is large for all variables, so the standard error is small compared to our coefficients. We can see that each variable has three asterisks, which indicates that the coefficients are significant and therefore are useful for the model. The small residual standard error tells us that our model is close to the actual values. The multiple R squared is 0.5969, which means that fuel explains 59.69% of the variation within carbon dioxide. Our model rejects the null hypothesis. Therefore, there is a relationship between fuel and carbon dioxide.

If one unit increase in the Electricity, and holding all the other predictors fixed, the `GHG_AVG` will decrease by 3.984×10^{-10} million metric tons per mile on average.

2. Analyze funding in wildfire prevention and CO2 emitted by wildfire

In order to figure out the relationship between CO2 emitted by wildfire and funds for wildfire prevention, we will use the nonparametric method called permutation of Kendall's Tau because our sample size is small (22 observations), and there may be some outliers (2008, 2020, and 2021 CO2 emissions by wildfires). The test statistics of Kendall's Tau can be calculated by this equation: $\tau = 2P[(X_i - X_j)(Y_i - Y_j) > 0] - 1$, where X is the funding in years, (ex: in 2019 and 2020, fundings were $X_{2019} = 1885440000$, $X_{2020} = 2866816000$), Y is the CO2 emission in years, (ex: in 2019 and 2020, the CO2 emissions by wildfire were $Y_{2019} = 106.7$ million metric tons and $Y_{2020} = 85.1$ million metric tons). Finally, we get one of our test statistics of Kendall's Tau equals 0.3904565.

By randomly permuting our dataset for 5000 times without replacement, we calculate the test statistics of Kendall Tau for each time. Then we assume that the correlation between CO2 emitted by wildfire (y) and funds for wildfire prevention (y) is smaller or equal to 0 (no correlation or negatively correlated). Otherwise, their correlation is greater than 0 (positively correlated).

After doing the hypothesis test of permutation of Kendall's Tau, we get our p-value equals to 0.005, which is extremely small. As a result, we reject our assumption and confirm that CO2 emitted by wildfire and funds for wildfire prevention are positively correlated, which means as the funding for wildfire prevention increases, the CO2 emitted by wildfire also increases.

## 3. Analyze what we can do with the tax and government funding

After increasing the 1.75% tax for Millionaires, we found there will be 3,843,949,900 dollars that can be applied to our funding for Zero Emission Vehicle (ZEV) program, Wildfire Response and Prevention Programs. Specifically, 1,729,777,455 dollars to help people buy new electric cars, 1,345,382,465 dollars to build charging stations, 768,789,980 for wildfire prevention. Moreover, the CA state already plans to dedicate 10 billion dollars[1] for the next 5 years to zero-emission vehicle programs and Charging stations. By proportion of the funding to the ZEV program, the CA government should put an additional 2,220,382,465 dollars per year to build new charging stations, and an additional 2,854,777,455 dollars per year to help people buy new electric cars from 2023 to 2027.

i. For the number of Charging Stations Prediction

The average building cost of one charging station is about 115,360 dollars [10]. From 2023 to 2027 year we will have 2,220,382,465 dollars in funding per year for our charging stations. Thus, there are about 96,237 new charging stations that could be built in these 5 years in an optimal situation.

From 2028 to the 2043 year, we will have 1,345,382,465 dollars in funding per year for our charging stations. Thus, there are about 186,599 new charging stations that could be built in these 16 years in an optimal situation. The increasing trend is shown as Figure 15 below.

Figure 15: Number of New Electric Cars in Total from 2023 to 2043
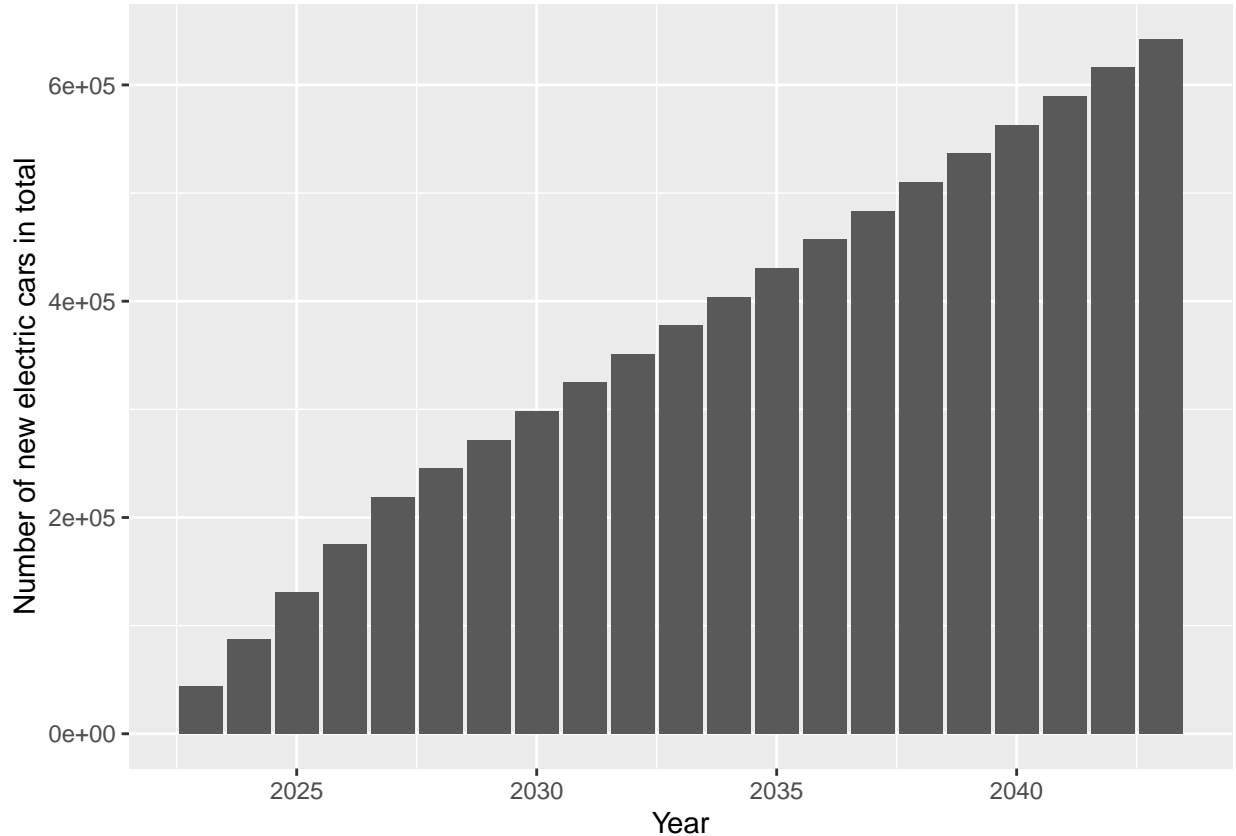
Thus, before 2044, we will be able to build 282,836 new charging stations.

ii. For the number of Electric Car Prediction

The average price of an electric car is about 65,291 dollars[11]. From 2023 to 2027 year we will have 2,854,777,455 dollars in funding per year for our electric car. Thus, there are about 218,619 new electric cars that could be built in these 5 years. And 43,723 new electric cars can be sold in one year in an optimal situation.

From 2028 to the 2043 year, we will have 1,729,777,455 dollars in funding per year for our electric cars. Thus, there are about 423,893 electric cars that could be sold in these 16 years. Thus, before 2044 year, we will be able to have 642,512 new electric cars in an optimal situation. The increasing trend is shown as Figure 16 below.

Figure 16: Number of New Charging Stations in Total from 2023 to 2043
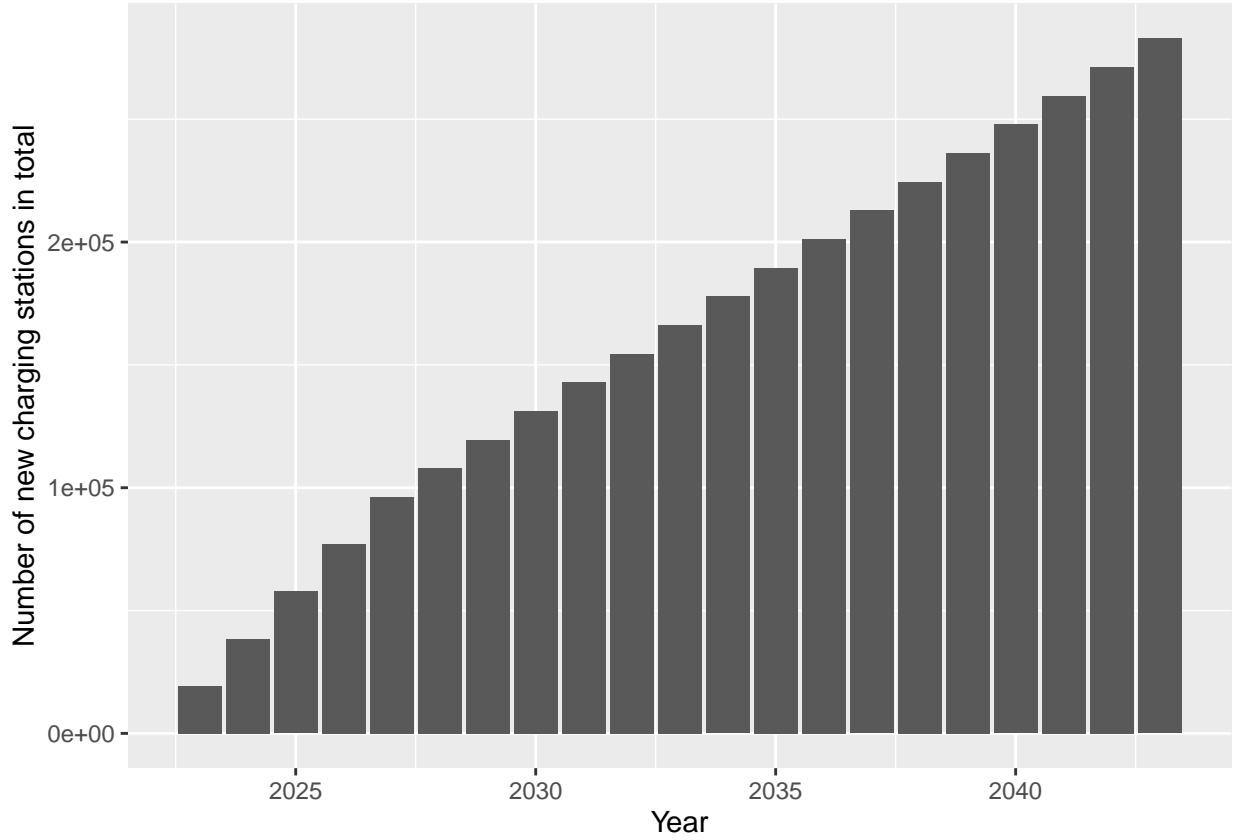
Assume the total number of passenger cars is fixed (117,026 counts) in California. Therefore, in 2026, all passenger cars will be using electricity with zero greenhouse gas emission.

iii. Maximum amount of greenhouse gas we can reduce

In California. Assume other greenhouse gas emission sources stay the same emission amounts. Based on our data, the total greenhouse gas emitted by vehicles is 59.35528 million metric tons in 2021. Moreover, the CO_2, which is a major component of greenhouse gas, generated by wildfires in 2021 was 85.1 million metric tons. According to the California Air Resource Board the total greenhouse gas 1990 was 426.855 million metric tons[12]. Our goal is to cut greenhouse gas emissions to 80% below the level in 1990[1], which is 85.371 million metric tons by calculation. By the estimation from Rhodium Group, the greenhouse gas emission in 2021 in California should be about 377.431 million metric tons (6.2% increase from 2020 level)[13]. By

subtracting the amount produced by cars and wildfire in 2021, we still have 232.97572 million metric tons remaining, which is still a lot compared to our goal in 2050, which is 85.371 million metric tons state-wise.

IV. Conclusion and Future Discussion

From our linear regression models, we find out that there are relationships between fuel types and all greenhouse gasses. Gasoline is the largest contributor to greenhouse gas emissions in all these three pollutants (CO_2, CH_4, and N_2O). On the other hand, electricity did not produce any emissions of greenhouse gasses. Diesel is always the second largest contributor while plug-in hybrids remain third.

By doing the permutation of Kendall's Tau, we find there is a positive correlation between the CO_2 emitted by wildfire and funds for wildfire prevention in California. Based on our analysis and research, we consider the Wildfire Response and Prevention Programs might not be effectively reducing greenhouse gasses in the past. The methods of management of the land and prevention of wildfires were not on the right track. This leads to more money spent but more greenhouse gasses by wildfire as we see from our analysis. To improve this, according to the World Economic Forum, we should focus more on "forest thinning, prescribed burning, Indigenous cultural burning, and managed wildfires.[14]"

Through our analysis, we might not be able to achieve the goal, which is to cut greenhouse gas emissions to 80% below the level in 1990, which is 85.371 million metric tons by 2050, only by performing proposition 30, because we still have 232.97572 million metric tons left even if we do not have wildfire and fossil fuel burning passenger cars. Instead, we should also control other sources of greenhouse gas such as trucks, cargo ships, industrial pollution, air conditioners, public transportation, planes, and so on.

Reference

[1] California Proposition 30: Income Tax for Electric Vehicles https://calmatters.org/california-voter-guide-2022/propositions/prop-30-income-tax-electric-cars/

[2] Current California GHG Emission Inventory Data https://ww2.arb.ca.gov/ghg-inventory-data

[3] Wildfire Emission Estimates for 2021 https://ww2.arb.ca.gov/sites/default/files/classic/cc/inventory/Wildfire%20Emission%20Estimates%202000-2021.pdf

[4] Proposition 30: Provides Funding for Programs to Reduce Air Pollution and Prevent Wildfires by Increasing Tax on Personal Income Over $2 Million. Initiative Statute https://lao.ca.gov/handouts/resources/2022/Proposition-30-081022.pdf

[5] EMFAC https://arb.ca.gov/emfac/emissions-inventory/26ef0d9f7c3739b95b469bcce274ff3b6bef927e

[6] Sources if Greenhouse Gas Emissions https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions

[7] PIT Annual Report https://data.ftb.ca.gov/California-Personal-Income-Tax/PIT-Annual-Report-2020/s2q7-rtsh

[8] Are you in the American middle class? Find out with our income calculator https://www.pewresearch.org/fact-tank/2020/07/23/are-you-in-the-american-middle-class/

[9] Historical Data | Legislative Analysis Office https://lao.ca.gov/policyareas/state-budget/historical-data

[10] CALeVIP DC Fast Chargers, Average Rebate, Unit Cost, and Total Project Cost per Charger https://www.energy.ca.gov/programs-and-topics/programs/clean-transportation-program/california-electric-vehicle/calevip-dc

[11] The Average Price of an Electric Car (Updated Monthly) https://joinyaa.com/guides/average-price-of-an-electric-car/

[12] Greenhouse Gas Inventory Data Explorer https://cfpub.epa.gov/ghgdata/inventoryexplorer/

[13] Preliminary US Greenhouse Gas Emissions Estimates for 2021 https://rhg.com/research/preliminary-us-emissions-2021/#:~:text=Based%20on%20preliminary%20data%20for,remained%205%25%20below%202019%20levels.

[14] California's wildfires are causing more and more damage https://www.weforum.org/agenda/2022/07/californian-wildfires-danger-environment/

Appendix: R Script

```r
knitr::opts_chunk$set(echo = TRUE)
rm(list=ls())
library(knitr)
library(ggplot2)
library(dplyr)
library(GGally)
library(lmtest)
library(car)
library(onewaytests)
library(ggiraph)
library(ggiraphExtra)
library(plyr)
library(ggpubr)
library(forcats)
set.seed(100)
# Load the 1st dataset
pollution = read.csv("https://caelection2022.gitdata.ga/dataset/EMFAC2021-EI-202x.csv",
                     header = TRUE)
# Pollution per mile per car
pollutionAVG = pollution %>%
            mutate(CO2_AVG = CO2_RUNEX / Total.VMT / 1000000,
                   CH4_AVG = CH4_RUNEX / Total.VMT / 1000000,
                   N2O_AVG = N2O_RUNEX / Total.VMT / 1000000,
                   GHG_AVG = (CO2_RUNEX + CH4_RUNEX + N2O_RUNEX) / Total.VMT / 1000000,
                   CO2_MILL = CO2_RUNEX / 1000000,
                   CH4_MILL = CH4_RUNEX / 1000000,
                   N2O_MILL = N2O_RUNEX / 1000000,
                   GHG_MILL = (CO2_RUNEX + CH4_RUNEX + N2O_RUNEX) / 1000000)
# Number of passenger cars
num_car = nrow(pollutionAVG)
num_car
# Number of cars by fuel types
table(pollutionAVG$Fuel)
# Total CO2
total_co2 = sum(pollutionAVG$CO2_RUNEX / 1000000)
total_co2
# Total CH4
total_ch4 = sum(pollutionAVG$CH4_RUNEX / 1000000)
total_ch4
# Total N2O
total_n2o = sum(pollutionAVG$N2O_RUNEX / 1000000)
total_n2o
# Total greenhouse gas
total_ghg = sum(pollutionAVG$CO2_RUNEX / 1000000) +
           sum(pollutionAVG$CH4_RUNEX / 1000000) +
           sum(pollutionAVG$N2O_RUNEX / 1000000)
total_ghg
# Summary table by fuel types
co2_total_fuel = as.data.frame(aggregate(pollutionAVG$CO2_MILL,
                     by=list(Category=pollutionAVG$Fuel), FUN=sum))
ch4_total_fuel = as.data.frame(aggregate(pollutionAVG$CH4_MILL,
```

```r
                        by=list(Category=pollutionAVG$Fuel), FUN=sum))
n2o_total_fuel = as.data.frame(aggregate(pollutionAVG$N2O_MILL,
                        by=list(Category=pollutionAVG$Fuel), FUN=sum))
ghg_total_fuel = as.data.frame(aggregate(pollutionAVG$GHG_MILL,
                        by=list(Category=pollutionAVG$Fuel), FUN=sum))
co2_avg_fuel = as.data.frame(aggregate(pollutionAVG$CO2_MILL,
                        by=list(Category=pollutionAVG$Fuel), FUN=mean))
ch4_avg_fuel = as.data.frame(aggregate(pollutionAVG$CH4_MILL,
                        by=list(Category=pollutionAVG$Fuel), FUN=mean))
n2o_avg_fuel = as.data.frame(aggregate(pollutionAVG$N2O_MILL,
                        by=list(Category=pollutionAVG$Fuel), FUN=mean))
ghg_avg_fuel = as.data.frame(aggregate(pollutionAVG$GHG_MILL,
                        by=list(Category=pollutionAVG$Fuel), FUN=mean))
fuel_count = as.data.frame(table(pollutionAVG$Fuel))
fuel_count$CO2_TOL = co2_total_fuel$x
fuel_count$CH4_TOL = ch4_total_fuel$x
fuel_count$N2O_TOL = n2o_total_fuel$x
fuel_count$GHG_TOL = ghg_total_fuel$x
fuel_count$CO2_AVG = co2_avg_fuel$x
fuel_count$CH4_AVG = ch4_avg_fuel$x
fuel_count$N2O_AVG = n2o_avg_fuel$x
fuel_count$GHG_AVG = ghg_avg_fuel$x
colnames(fuel_count) = c('Fuel Type','Count','CO2 Total','CH4 Total','N2O Total',
                        'GHG Total','CO2 Mean','CH4 Mean','N2O Mean','GHG Mean')
kable(fuel_count)
# CO2 total
pollutionAVG %>%
  mutate(Fuel = fct_reorder(Fuel, CO2_MILL)) %>%
  ggplot(aes(y=CO2_MILL, x=Fuel, fill=Fuel)) +
  geom_bar(stat='identity') + labs(y = "CO2 (million metric tons)") + coord_flip()
# CH4 total
pollutionAVG %>%
  mutate(Fuel = fct_reorder(Fuel, CH4_MILL)) %>%
  ggplot(aes(y=CH4_MILL, x=Fuel, fill=Fuel)) +
  geom_bar(stat='identity') + labs(y = "CH4 (million metric tons)") + coord_flip()
# GHG total
pollutionAVG %>%
  mutate(Fuel = fct_reorder(Fuel, N2O_MILL)) %>%
  ggplot(aes(y=N2O_MILL, x=Fuel, fill=Fuel)) +
  geom_bar(stat='identity') + labs(y = "N2O (million metric tons)") + coord_flip()
# GHG total
pollutionAVG %>%
  mutate(Fuel = fct_reorder(Fuel, GHG_MILL)) %>%
  ggplot(aes(y=GHG_MILL, x=Fuel, fill=Fuel)) +
  geom_bar(stat='identity') + labs(y = "GHG (million metric tons)") + coord_flip()
pollutionAVG %>%
  mutate(Fuel = fct_reorder(Fuel, CO2_AVG)) %>%
  ggplot(aes(x=Fuel, y=CO2_AVG, color=Fuel)) +
  geom_boxplot() + labs(y = "CO2 (million metric tons per mile)")
pollutionAVG %>%
  mutate(Fuel = fct_reorder(Fuel, CH4_AVG)) %>%
  ggplot(aes(x=Fuel, y=CH4_AVG, color=Fuel)) +
  geom_boxplot() + labs(y = "CH4 (million metric tons per mile)")
```

```r
pollutionAVG %>%
  mutate(Fuel = fct_reorder(Fuel, N2O_AVG)) %>%
  ggplot(aes(x=Fuel, y=N2O_AVG, color=Fuel)) +
  geom_boxplot() + labs(y = "N2O (million metric tons per mile)")
pollutionAVG %>%
  mutate(Fuel = fct_reorder(Fuel, GHG_AVG)) %>%
  ggplot(aes(x=Fuel, y=GHG_AVG, color=Fuel)) +
  geom_boxplot() + labs(y = "GHG (million metric tons per mile)")
# Load the 2nd dataset
income = read.csv("https://caelection2022.gitdata.ga/dataset/2020_Combined_AR_PIT_B4A.csv",
                  header = TRUE)
AGI_class = income$Adjusted.Gross.Income.Class[-c(1, 2, 62, 63)]
People_Taxable = as.numeric(gsub(",","",income$X[-c(1, 2, 62, 63)]))
Federal_AGI = as.numeric(gsub(",","",income$Federal.AGI..Thousands.[-c(1, 2, 62, 63)])) * 1000
Class = c(rep("Lower Class", 42), rep("Middle Class", 7), rep("Upper Class", 10))
income_clean = data.frame(AGI_class, People_Taxable, Federal_AGI, Class)
income_clean[is.na(income_clean)] = 0
# Total people taxable 2020
sum(income_clean$People_Taxable)
# Total AGI 2020
sum(income_clean$Federal_AGI)
# Calculate the tax revenue generated by the 1.75% tax
people_200m = income_clean$Federal_AGI[c(56, 57, 58, 59)]
people_200m
AGI_200m = sum(people_200m)
AGI_200m
tax_revenue = AGI_200m * 0.0175
tax_revenue
colnames(income_clean) = c('AGI Class', 'People Taxable', 'Federal AGI', 'Class')
income_clean_copy = income_clean
income_clean_copy[60,] = c('Total:', 35005,
                           AGI_200m, '')
kable(income_clean_copy[56:60,-4])
class_data1 = as.data.frame(aggregate(income_clean$`People Taxable`,
          by=list(Class=income_clean$Class), FUN=sum))
ggplot(class_data1, aes(y = x, x = Class)) +
  geom_bar(stat='identity') + labs(y = "People Taxable", x="Class")
class_data2 = as.data.frame(aggregate(income_clean$`Federal AGI`,
          by=list(Class=income_clean$Class), FUN=sum))
ggplot(class_data2, aes(y = x, x = Class)) +
  geom_bar(stat='identity') + labs(y = "Fedeal AGI (dollars)", x="Class")
budget = read.csv("https://caelection2022.gitdata.ga/dataset/Historical_Expenditures.csv",
                  header = TRUE)
budget_clean = budget %>%
              filter(Department %in% c("Office of the State Fire Marshall",
                                       "Department of Forestry and Fire Protection"))
budget_clean = budget_clean %>%
              filter(Fund %in% c("General Fund", "Special Funds")) %>%
              arrange(desc(Amount))
budget_clean = budget_clean[1:160,] %>% arrange(desc(Fiscal.Year))
budget_clean = budget_clean[-c(1:5, 86:160),]
budget_clean$Amount = gsub(",","",budget_clean$Amount)
budget_clean$Amount = as.numeric(gsub("\\$","",budget_clean$Amount)) * 1000
```

```r
budget_clean = as.data.frame(aggregate(budget_clean$Amount,
                             by=list(Category=budget_clean$Fiscal.Year), FUN=sum))
Fire_CO2 = c(5.4, 6.7, 13.5, 19.7, 5.7, 2.6, 13.5, 20.0, 42.8, 9.2, 1.9, 3.1,
             11.4, 14.7, 17.2, 19.2, 10.5, 31.3, 13.1, 4.8, 106.7, 85.1)
budget_clean$Fire_CO2 = Fire_CO2
budget_clean = budget_clean[,-1]
budget_clean$Year = rep(2000:2021)
colnames(budget_clean) = c('Fund','CO2','Year')
# Fund by year
ggplot(budget_clean, aes(y = Fund, x = Year)) +
  geom_bar(stat='identity') + labs(y = "Fund for wildfire prevention (dollar(s))")
# CO2 by year
ggplot(budget_clean, aes(y = CO2, x = Year)) +
  geom_bar(stat='identity') + labs(y = "CO2 emission by wildfire (million metric tons)")
# CO2 distribution
ggplot(budget_clean, aes(CO2)) +
  geom_boxplot() + labs(x = "CO2 emission by wildfire (million metric tons)")
median(budget_clean$CO2)
# CO2 and fund
ggplot(budget_clean, aes(y = CO2, x = Fund)) +
  geom_point(size=2) + labs(x = "Fund for wildfire prevention (dollar(s))",
                            y = "CO2 emission by wildfire (million metric tons)") +
  geom_smooth(method = lm, se = TRUE)
# Linear regression for CO2
co2.model.avg = lm(CO2_AVG ~ as.character(Fuel), data = pollutionAVG)
summary(co2.model.avg)
# Linear regression for CH4
ch4.model.avg = lm(CH4_AVG ~ as.character(Fuel), data = pollutionAVG)
summary(ch4.model.avg)
# Linear regression for N2O
n2o.model.avg = lm(N2O_AVG ~ as.character(Fuel), data = pollutionAVG)
summary(n2o.model.avg)
# Linear regression for GHG
ghg.model.avg = lm(GHG_AVG ~ as.character(Fuel), data = pollutionAVG)
summary(ghg.model.avg)
rp = cor(budget_clean$Fund, budget_clean$CO2)
rs = cor(budget_clean$Fund, budget_clean$CO2,method = "spearman")
rt = cor(budget_clean$Fund, budget_clean$CO2,method = "kendall")
all.correlations = c(rp,rs,rt)
all.correlations
names(all.correlations) = c("Pearson","Spearman","Kendall")
all.correlations
R = 5000
n = nrow(budget_clean)
all.perm.cor = sapply(1:R,function(i){
  the.data = budget_clean
  budget_clean$Fund = sample(budget_clean$Fund,n,replace = FALSE)
  rp = cor(budget_clean$Fund, budget_clean$CO2)
  rs = cor(budget_clean$Fund, budget_clean$CO2,method = "spearman")
  rt = cor(budget_clean$Fund, budget_clean$CO2,method = "kendall")
  all.correlations = c(rp,rs,rt)
  names(all.correlations) = c("Pearson","Spearman","Kendall")
  return(all.correlations)
```

```r
})
all.perm.pvalues = sapply(1:nrow(all.perm.cor),function(i){
  lower = mean(all.perm.cor[i,] < all.correlations[i])
  upper = mean(all.perm.cor[i,] > all.correlations[i])
  two = mean(abs(all.perm.cor[i,]) > abs(all.correlations[i]))
  results = c(lower,upper,two)
  names(results) = c("lower tail","upper tail","two sided")
  return(results)
})
colnames(all.perm.pvalues) = names(all.correlations)
all.perm.pvalues
# Amount of money used in each section
buy_car = tax_revenue * 0.45
charging_station = tax_revenue * 0.35
fire_control = tax_revenue * 0.2
# Only for car and charging station in the next 5 years
addition = 10000000000
addition_car_year = addition / 5 * 0.5625
addition_charging_year = addition / 5 * 0.4375
total_car_2027_year = buy_car + addition_car_year
total_charging_2027_year = charging_station + addition_charging_year
#2023-2027 charging stations and electric (5years) per year funding
total_car_2027_year = buy_car + addition_car_year
total_charging_2027_year = charging_station + addition_charging_year
#2028-2043 (16 years) charging stations and electric and fire control per year funding
buy_car = tax_revenue * 0.45
charging_station = tax_revenue * 0.35
fire_control = tax_revenue * 0.2
year_21 = c(2023:2043)
electric_car = c(43723, 43723*2, 43723*3, 43723*4, 43723*5,
                 218615+26493, 218615+26493*2, 218615+26493*3,
                 218615+26493*4, 218615+26493*5, 218615+26493*6,
                 218615+26493*7, 218615+26493*8, 218615+26493*9,
                 218615+26493*10, 218615+26493*11, 218615+26493*12,
                 218615+26493*13, 218615+26493*14, 218615+26493*15,
                 218615+26493*16)
charging_number = c(19247, 19247*2, 19247*3, 19247*4, 19247*5,
                 96235+11662, 96235+11662*2, 96235+11662*3,
                 96235+11662*4, 96235+11662*5, 96235+11662*6,
                 96235+11662*7, 96235+11662*8, 96235+11662*9,
                 96235+11662*10, 96235+11662*11, 96235+11662*12,
                 96235+11662*13, 96235+11662*14, 96235+11662*15,
                 96235+11662*16)
prediction = as.data.frame(cbind(year_21, electric_car, charging_number))
#2023-2027 charging stations (5years) per year # numbers
ave_charging = 115360
number_charging_5year = total_charging_2027_year/115360
number_charging_5year
Total_5year = 5*number_charging_5year
floor(Total_5year)
#2028-2043 (16 years) #number per year
ave_charging = 115360
number_charging_16year = charging_station/115360
```

```r
number_charging_16year
Total_16year = 16*number_charging_16year
floor(Total_16year)
TOTAL_Plan_Cha =floor(Total_16year) + floor(Total_5year)
TOTAL_Plan_Cha
ggplot(prediction, aes(y = electric_car, x = year_21)) +
    geom_bar(stat='identity') + labs(y = "Number of new electric cars in total", x = "Year")
#2023-2027 EC (5years) per year # numbers
ave_car = 65291
number_car_5year = total_car_2027_year/65291
number_car_5year
Total_5_caryear = 5*number_car_5year
floor(Total_5_caryear)
#2028-2043 (16 years) #number per year
ave_car = 65291
number_car_16year = buy_car/65291
number_car_16year
Total_16_caryear = 16*number_car_16year
floor(Total_16_caryear)
TOTAL_Plan_car=floor(Total_16_caryear) + floor(Total_5_caryear)
TOTAL_Plan_car
89712/number_car_5year
ggplot(prediction, aes(y = charging_number, x = year_21)) +
    geom_bar(stat='identity') + labs(y = "Number of new charging stations in total", x = "Year")
```