Vehicle Fuel Efficiency & Carbon Dioxide Emissions

A study in factors affecting fuel efficiency, and CO2 emissions as well as building Machine Learning models to predict fuel efficiency

Paul Shoop
Booz | Allen | Hamilton
Shoop_paul@bah.com

Goal

- Create a machine learning model capable of predicting fuel efficiency based on certain variables
- Make the machine learning model accurate enough that it could potentially replace costly and time consuming laboratory tests for fuel efficiency

The Data:

- Open source data available of fueleconomy.gov
- Compilation of vehicle data from 1985-2018
- Over 40,000 entries and 83 variables
- Data contains both traditional and hybrid/electric vehicles*

The data dictionary is here: http://www.fueleconomy.gov/feg/ws/index.shtml#vehicle

First Question:

Have fuel efficiency standards had an impact on CO2 emissions and overall fuel efficiency over time

Second Question:

Which factors, based on available data, most affect fuel efficiency

Third Question:

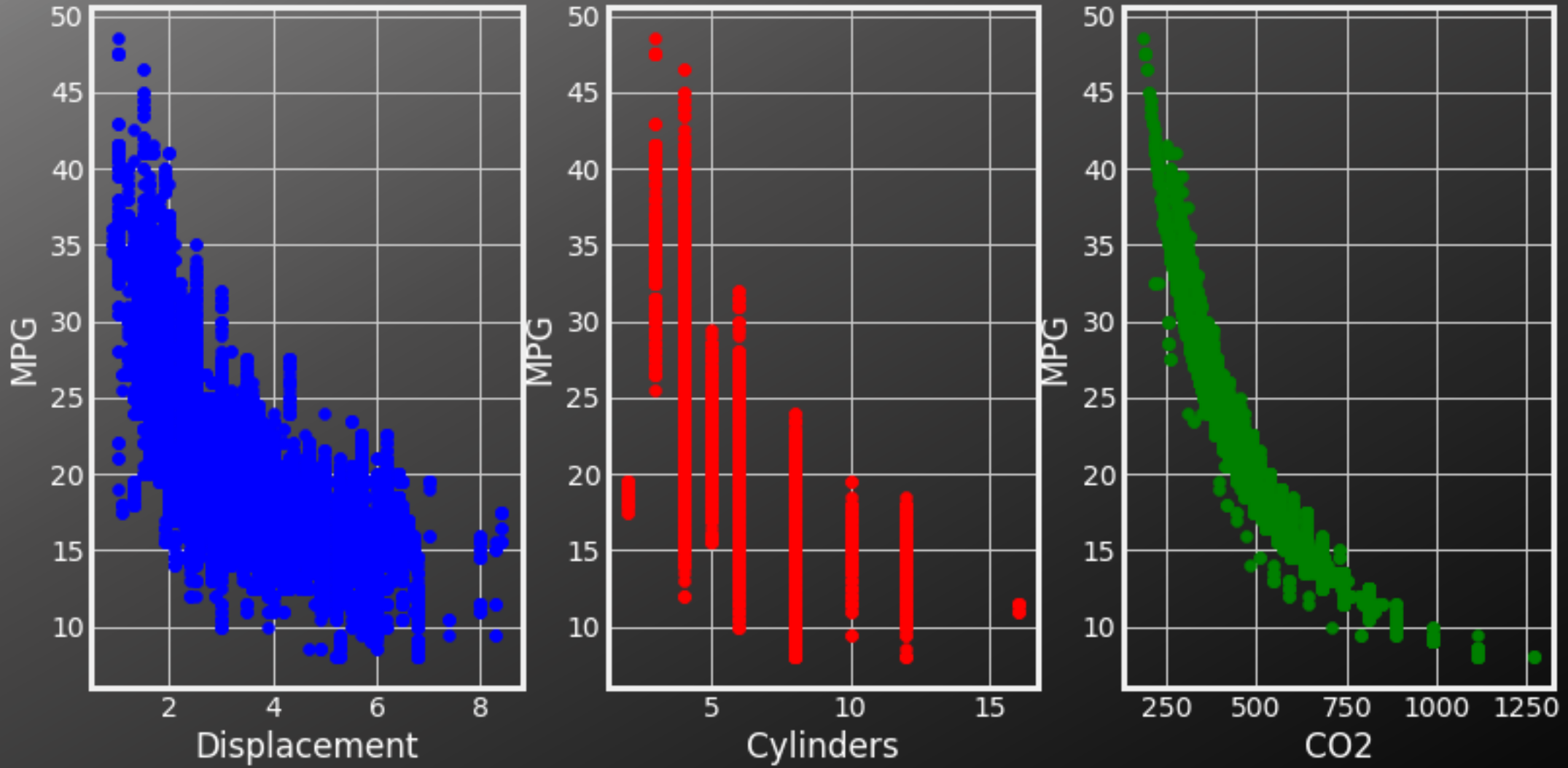
Can a Machine Learning Model be built to predict fuel efficiency

First Question

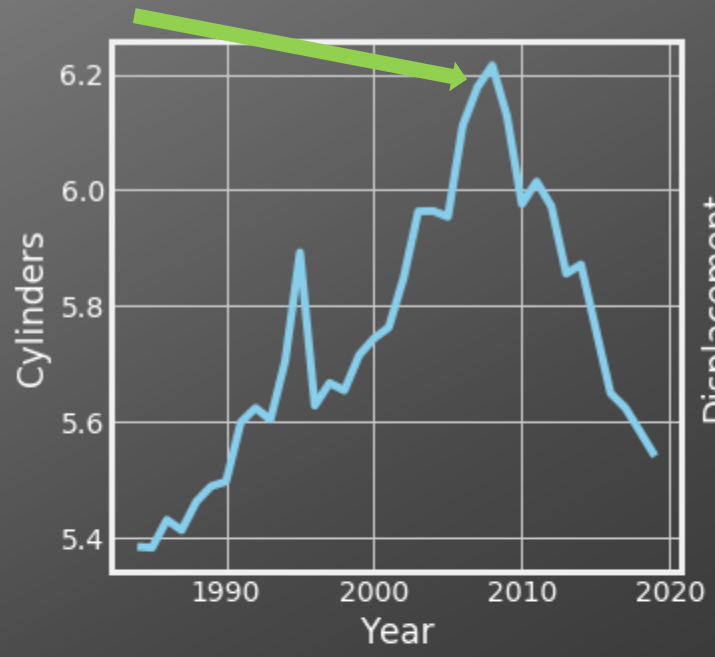
Fuel Efficiency over Time

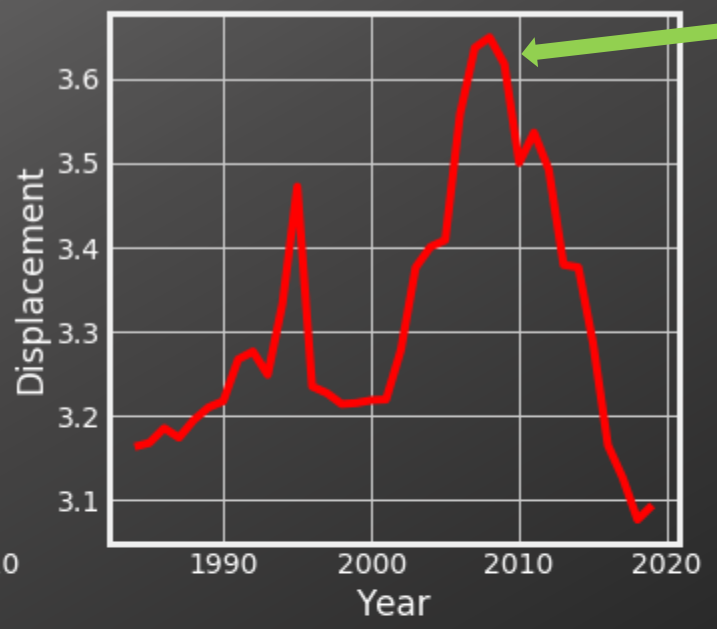
Second Question

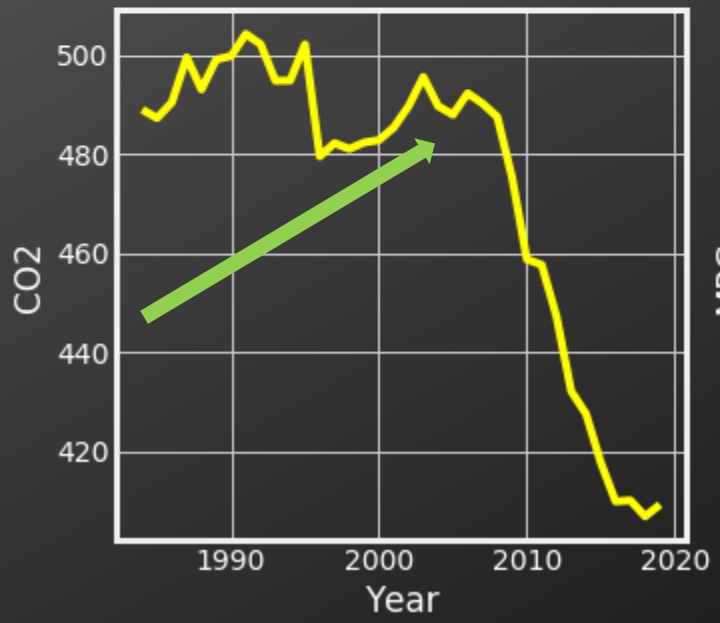
Relationship Between MPG and Displacement, Cylinders, and CO2 Emissions

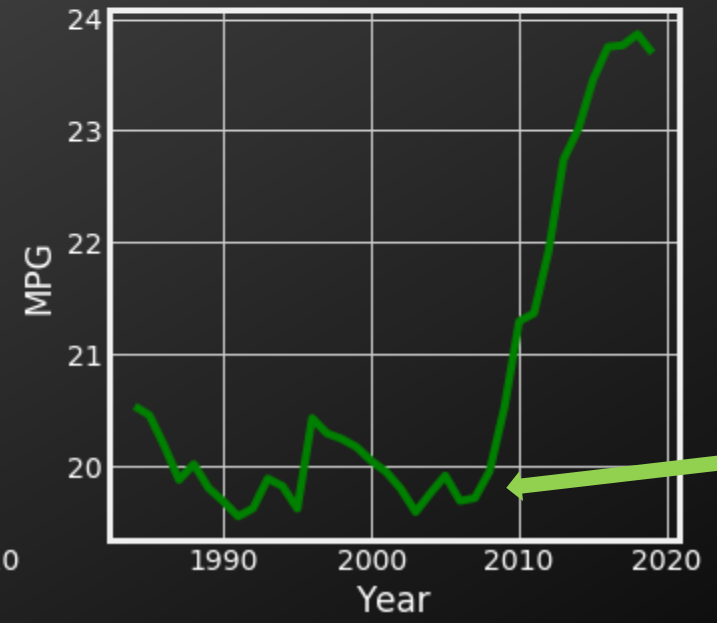


Averages of Major Variables over Time





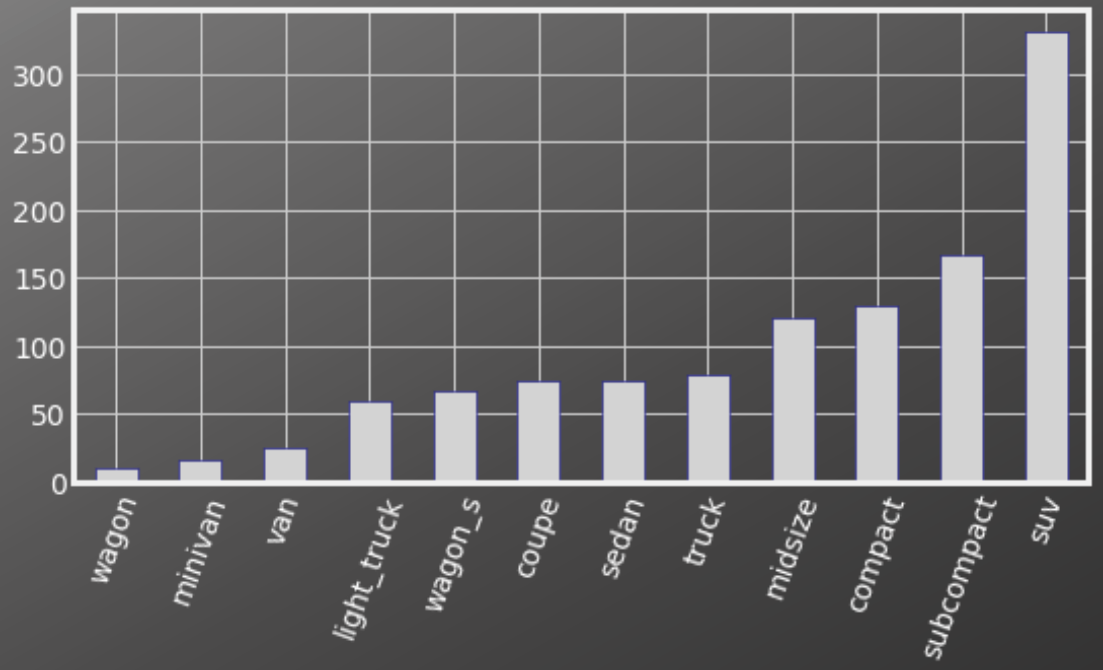




- **Energy Independence and Security Act of 2007**
  - raised the fuel economy standards of America's cars, light trucks, and SUVs to a combined average of at least 35 miles per gallon by 2020—a 10 mpg increase over 2007 levels

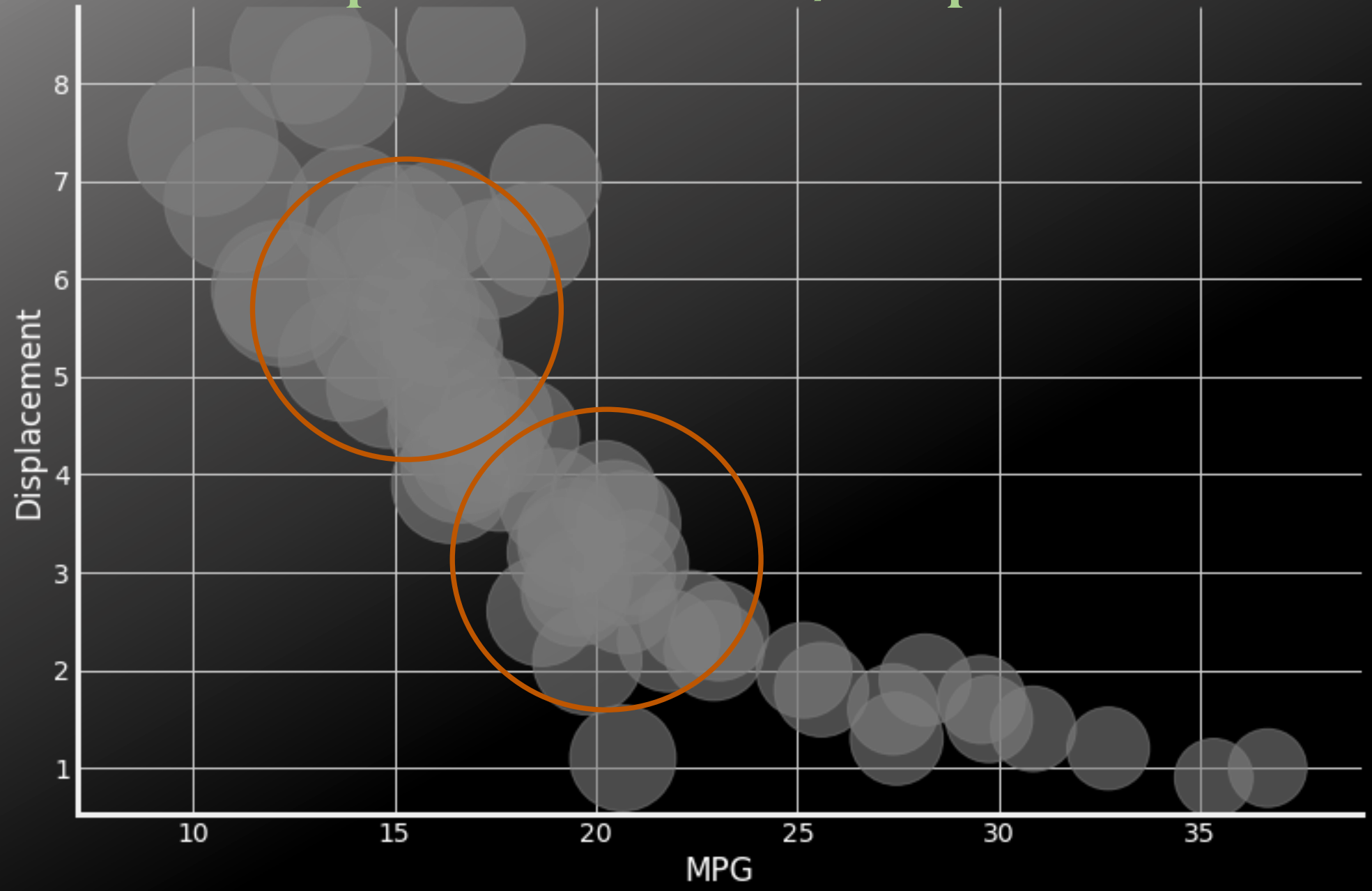
- **Vehicle Sales**
  - The number of SUVs in 2009 in this data set dwarfed all other vehicle types

Vehicles by Type in 2009



Vehicles by Type over Time

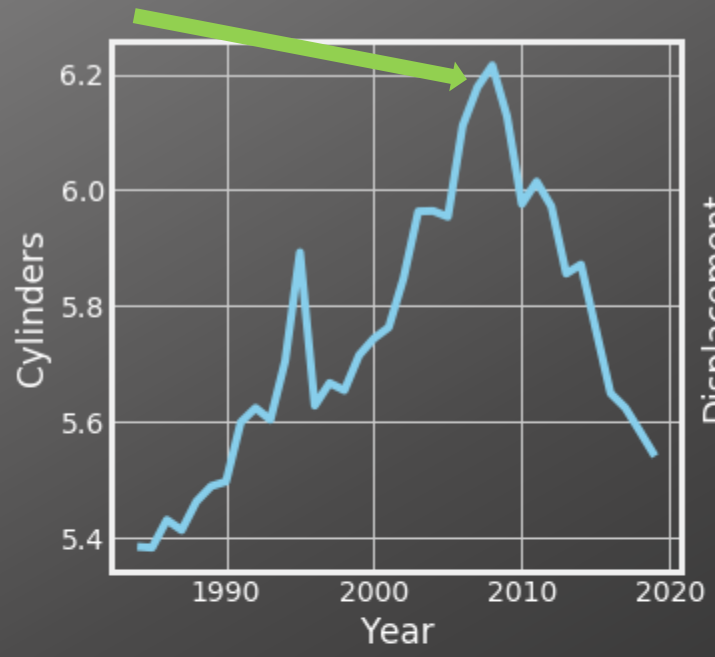
The Relationship Between MPG, Displacement & CO2

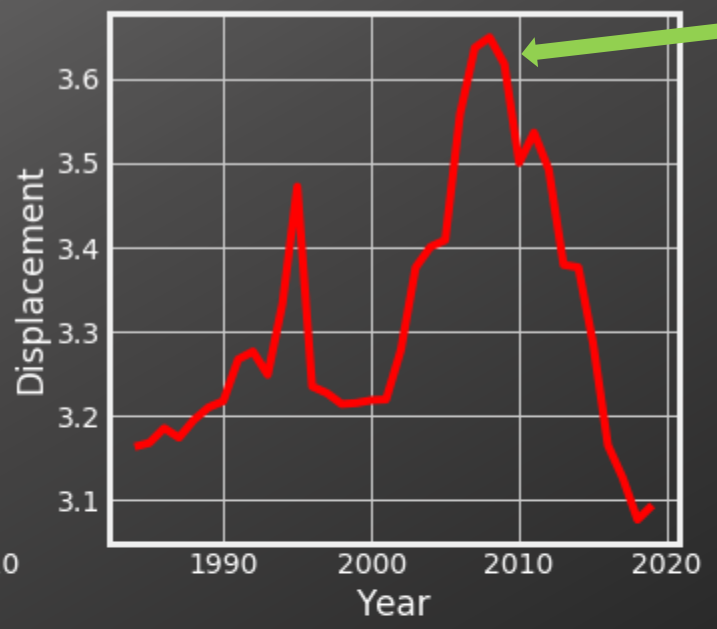


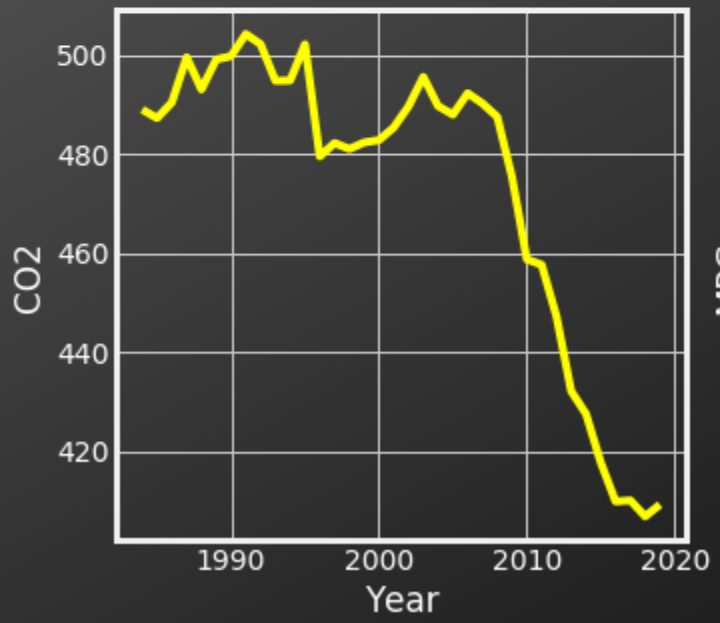
Third Question

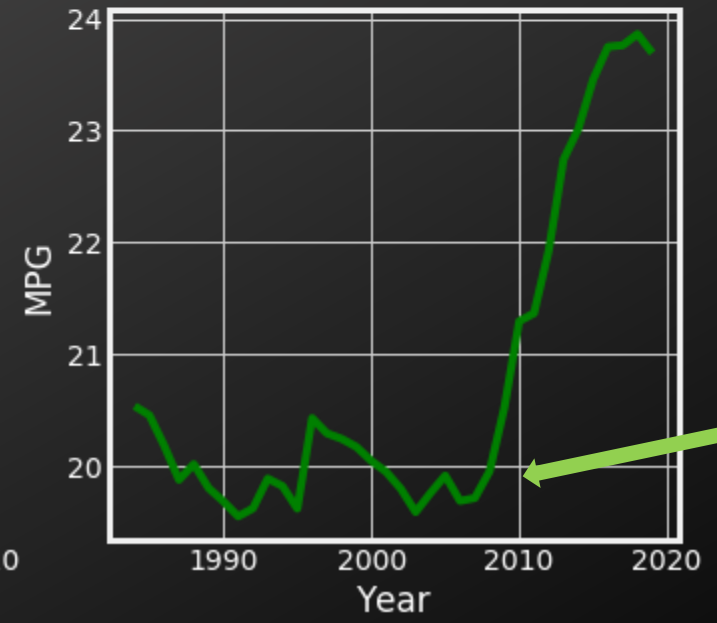
Machine Learning Demonstration

Averages of Major Variables over Time









Considerations:

- This data does not capture all variables that could affect fuel efficiency and CO2 emissions
- Hybrid and Electric Vehicles may skew data

The Way Ahead:

- Build a separate machine learning model specifically for hybrid and electric vehicles
- Build a separate machine learning model specifically for predicting CO2 emissions based on available data
- Deploy a front-end user model on a webpage for educational purposes
- Deploy a machine learning model for making predictions over large sets of new data for EPA, DOE, and other Agencies and NGOs

Questions