# Part 1: Research Proposal

Authors: Zachary Galante, Mark Herrera, Michael Varner, and Nicholas Lee

1. *What is the research question? Specifically, what is the X concept and what is the Y concept?*

How can we increase a car’s mileage per gallon by making changes to characteristics such as weight, horsepower, acceleration, and engine displacement/cylinders?   
  
*2. What is the data source? What variables will you use to operationalize X and Y?*

This donated dataset from 1993-07-07 includes 398 different models of cars from the years 1970 - 1982. This dataset was taken from the Statlib library, maintained at Carnegie Mellon University and used in the 1983 American Statistical Association Exposition[[1]](#footnote-0). The model will be built to predict the changes that could be made to increase the mpg for a car model. MPG will be modeled by the five continuous variables and three discrete variables. The first model will look at the interaction between weight and horsepower. The second model will add acceleration. Third adding displacement and cylinders since they would be covariates as the more cylinders the car has the more displacement (volume of the gas chamber).

**Dependent Variables**: cylinders, displacement, weight, horsepower, acceleration, and origin

**Independent Variable**: MPG

3. *What is the unit of observation? That is, does each row of the data represent a person, a review, a hotel stay, or something else?*

Each row is a car model for a given year.

**Issues to note:**

* Some car models were used across multiple years. Car models with a longer history have time to improve/tailor their car to perform well in mpg or acceleration.
* There are no classes of cars variable, which could make the comparison between cars challenging since the cars will be tuned for certain performances. Say a sports car would favor acceleration, horsepower and displacement with a lighter weight for increased speed.
* Colinearity between horsepower and acceleration; weight and acceleration
* Omitted variables: Aerodynamics of car, weight of person/cargo, terrain, driver, miles traveled

1. Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science. (<https://archive.ics.uci.edu/ml/datasets/Auto+MPG>) [↑](#footnote-ref-0)