

## Where did this car come from?

I am analyzing many factors of a car in order to predict where the car was made. These factors include the miles per gallon, number of cylinders, horsepower, acceleration, and weight. The places of origin can be USA, Europe, or Japan.

My entire life I have always wondered if there was any difference between American cars and Japanese cars. I have always heard rumors that American cars are more “powerful” and Japanese cars are more “reliable.” Predicting whether there was any correlation between the actual properties of a car and where it came from would show if there was any significant difference between cars from different countries.

This dataset was taken from the StatLib library which is maintained at Carnegie Mellon University. Previously it has been used to predict the miles per gallon of a car by Ross Quinlan (1993).

### Attributes

- |                  |                             |
|------------------|-----------------------------|
| 1. MPG:          | continuous                  |
| 2. Cylinders:    | multi-valued discrete {0-8} |
| 3. Horsepower:   | continuous                  |
| 4. Weight:       | continuous                  |
| 5. Acceleration: | continuous                  |

### Classification

Origin {1,2,3} : 1 = USA; 2 = Europe; 3 = Japan;

I am not using the attributes “displacement” and “model year” because I did not want my predictions based on either of them. I did not want my algorithm to be based on the model year because it was not a “physical property” of a car and I wanted to focus on “physical properties.” I did not use “displacement” because it doesn’t add anything different from cylinders and horsepower. Engine **displacement** is the volume of an engine's cylinders, a general indicator of its size and power. Since the measurement is very general and not specific, I did not believe it was useful.

There are 240 data points in this dataset.

This dataset was acquired from: <http://cs.nyu.edu/courses/fall00/G22.3033-001/weka/weka-3-0-2/data/auto-mpg.arff>

I made no changes to the algorithm. The algorithm worked perfectly as soon as I formatted my dataset correctly.

Class Distribution:

Class	Value	Number of Instances
USA	1	159
Europe	2	42
Japan	3	39

### Statistical Analysis:

Attribute	Mean	Standard Deviation
MPG	23.006249999999998	7.776845964720908
Cylinders	5.5625	1.732126296283314
Horsepower	107.11666666666666	39.7881804390441
Weight	3018.5458333333333	865.189193752876
Acceleration	15.336249999999996	2.6445448904914186

In order to test the importance of each of the attributes, I ran the algorithm twenty-five times. First, I ran it five times without the MPG attribute, then five times without the cylinders attribute (and all of the other attributes), etc. Then, I looked at the change in accuracy of the algorithm. After doing this I concluded that weight impacted the accuracy the most, so it was the most important. The next most impactful attribute was the weight attribute. The other attributes did not have as dramatic of an impact on the accuracy, so they were equally important.