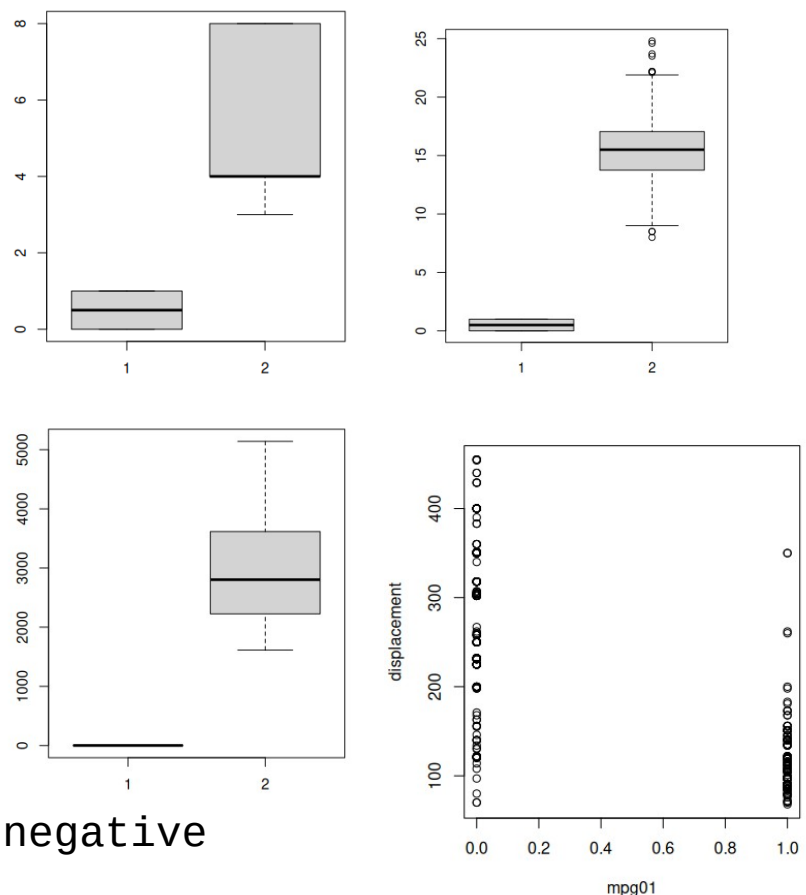


a) I used the median function to pull out the median from the Auto\$mpg column, and then used two lines to create a vector of 0s and 1s based on if the mpg is > median_mpg

```
1 library('ISLR')
2
3 # a)
4 attach(Auto)
5 median_mpg = median(mpg)
6
7 mpg01[mpg >= median_mpg] = 1
8 mpg01[mpg < median_mpg] = 0
9
10 data = data.frame(subset(Auto, select = -mpg), mpg01)
```

b) Here are some plots in order of how I checked some relationships: Box between mpg01 and # of cylinders, Box between mpg01 and acceleration, Box between mpg01 and weight, and scatterplot between mpg01 and displacement. The box plots all show a positive relationship between mpg01 and the corresponding measure, while the plot shows a negative relationship.



c) I created an index column in a set of the randomized data, and then for the training set took the first 300 values, and test set the last 92 values

```
# c)
data_random = data[sample(nrow(data)),]
data_random$index = 1:nrow(data_random)
train = data_random[data_random$index <= 300, ]
test = data_random[data_random$index > 300, ]
```