

Factors impacting miles per gallon

a regression analysis

Executive summary

Though there is a difference in fuel consumption between automatic cars, this difference is explained by automatic cars being heavier than manual cars. The impact of the type of transmission for cars of the same weight is negligible for the analysed cars. If one wants to maximize miles per gallon focus should be on decreasing the weight of the car.

Data

I have explored the mtcars data set, which contains information on fuel consumption in miles per gallon (MPG) and 10 aspects of car design and performance for 32 car models. The data set is described in the [R documentation](#). All variables were treated as numerical values except the engine (V/S) and transmission (A/M) types.

Exploratory analysis

I performed an exploratory analysis to investigate the properties of the data and get a first insight into what affects the fuel consumption and the differences between automatic and manual transmission.

At first glance (Figure 1), it seems that transmission type has a clear effect on the fuel consumption, but it is possible that this apparent connection is driven by other variables. From figure 2 we see that fuel consumption for automatic cars (pink dots) is higher than that of manual cars, but we can also see that automatic cars tend to be heavier than manual cars (Figure 3). Taking this into account, automatic cars do not seem to have higher fuel consumption than manual cars of similar weight. The black line shows the result of fitting a linear regression model for MPG depending only on weight. This simple model seems to capture the patterns for both automatic and manual cars reasonably well, but are there other traits of the data that could tell us more about the MPG?

Figures 4-13 explore the relationship between the different variables and MPG. Both number of cylinders and engine type (V/S) (Figures 4-7) look related to MPG when looking at them alone, but when also considering weight, it is not so clear anymore. Figures 8-13 explore the relationship between MPG and the variables displacement, horse power, rear axle ratio, qsec, number of gears and number of carburetors. I have also computed the correlation between each variable and weight. Again, pink dots represent cars with automatic transmission and blue ones those with manual transmission.

Finding a representative model of the data

Based on the strong relationship between weight and MPG identified, I constructed a linear regression model of MPG depending on weight, which gives us a negative relationship (slope = -5) between MPG and weight. The R-squared for this model is 0.75. We can interpret this as 'weight explains 75% of the variance of MPG in the data'.

I then investigated if the model could be improved by adding an interaction term between weight and transmission. Model 2 indicates that the relationship between weight and MPG is different between automatic and manual cars, with a stronger negative relationship for manual cars. The R-squared now increases to 83% of variance explained.

Based on the results from the exploratory analysis I added the variable qsec, the variable with the lowest correlation with weight (model 3). In addition to this, I constructed a model that uses all variables (model 4). The below table displays the R-squared for these 4 models as well as the p value generated by an anova analysis comparing the models.

Model name	Included variables	R-squared (%)	PR(>F)
fit_wt	weight	75	
fit_wt_am	weight * transmissiontype	83	0.0026919
fit_wt_am_qsec	weight * transmissiontype + qsec	90	0.0019822
fit_wt_am_all	weight * transmissiontype + all remaining variables	90	0.9939286

The anova analysis shows no evidence that including all variables would give us a more accurate model than the one we get when including only weight, transmission type and qsec. Since both model number 3 and 4 explain the same amount of variance (90%) there is no reason to believe that adding more variables to the model would improve it and I will select model number 3. The coefficients for this model are:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.723053	5.8990407	1.648243	0.1108925
wt	-2.936531	0.6660253	-4.409038	0.0001489
amManual	14.079428	3.4352512	4.098515	0.0003409
qsec	1.016974	0.2520152	4.035366	0.0004030
wt:amManual	-4.141376	1.1968119	-3.460340	0.0018086

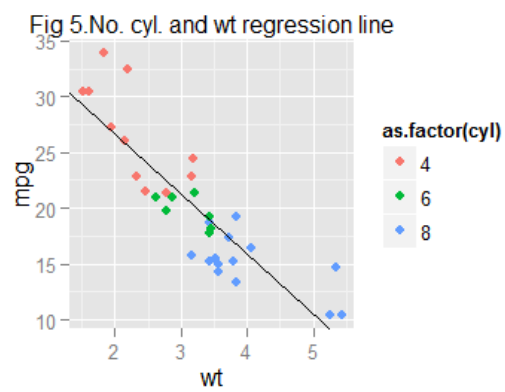
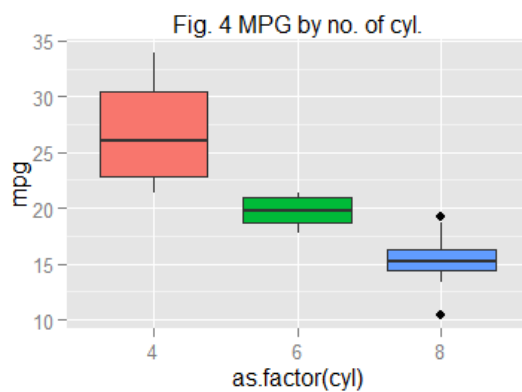
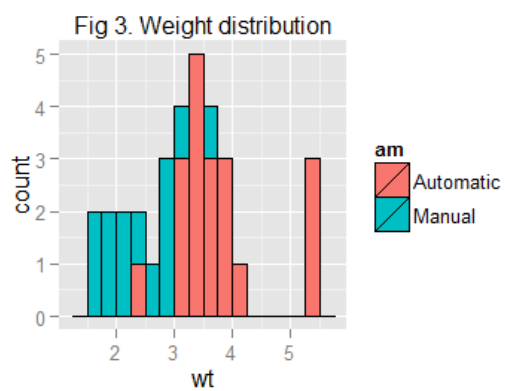
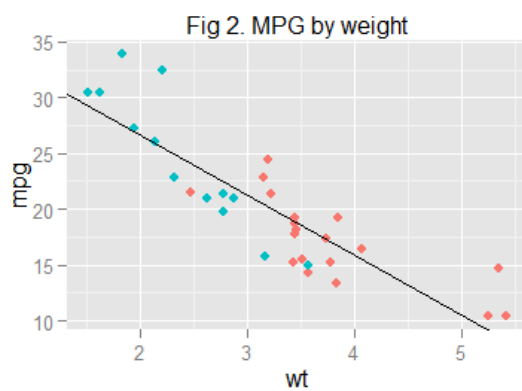
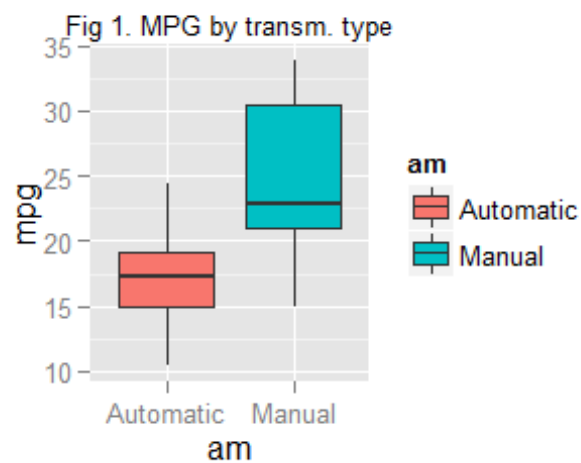
Showing that a greater value of qsec (acceleration) will increase fuel consumption. As in model 2, we see that the connection between MPG and transmission type depends on the weight of the car. For automatic cars, MPG decreases with -2.94 for every 1000 lb increase in weight, and for manual cars MPG decreases with -7.08 for every 1000 lb increase (when holding qsec constant).

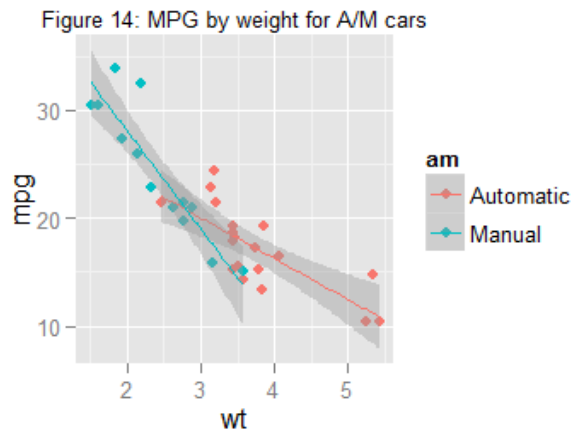
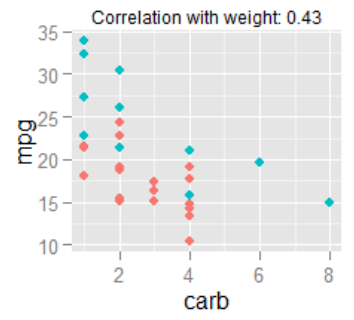
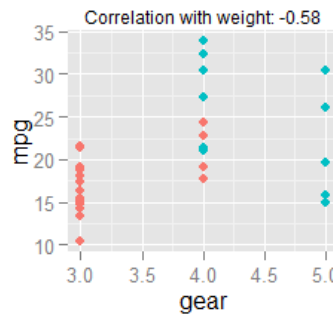
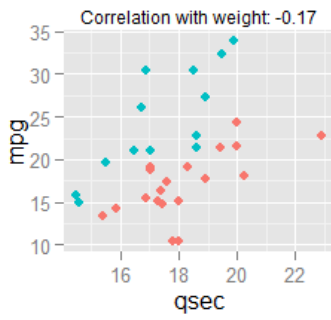
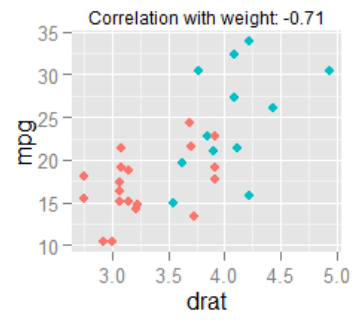
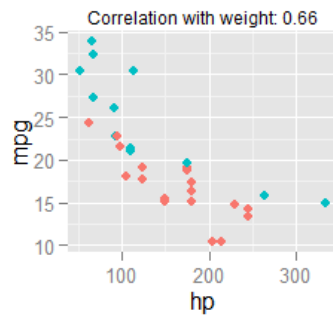
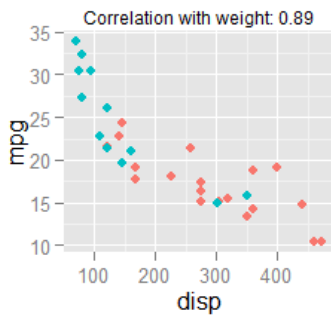
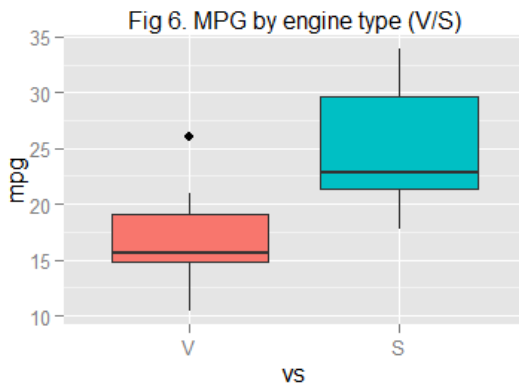
Difference in MPG between automatic and manual transmission cars

According to our model, the dependence of MPG on transmission type interacts with the weight of the cars. Due to this, it is difficult to determine the effect of only transmission. The 14.1 offset for manual transmission compared to automatic transmission only applies to cars with weight = 0, and so does not tell us much. For cars with weights within the span of the analysed cars the MPG values overlap each other. Figure 14 illustrates this by plotting the dependence of MPG on weight for automatic and manual cars.

Diagnostics of model

There is no obvious pattern in the residuals, so the model does not seem to be biased. There is no single data point that has an extreme influence on the model. All in all, the diagnostics indicate that this is a reliable model.





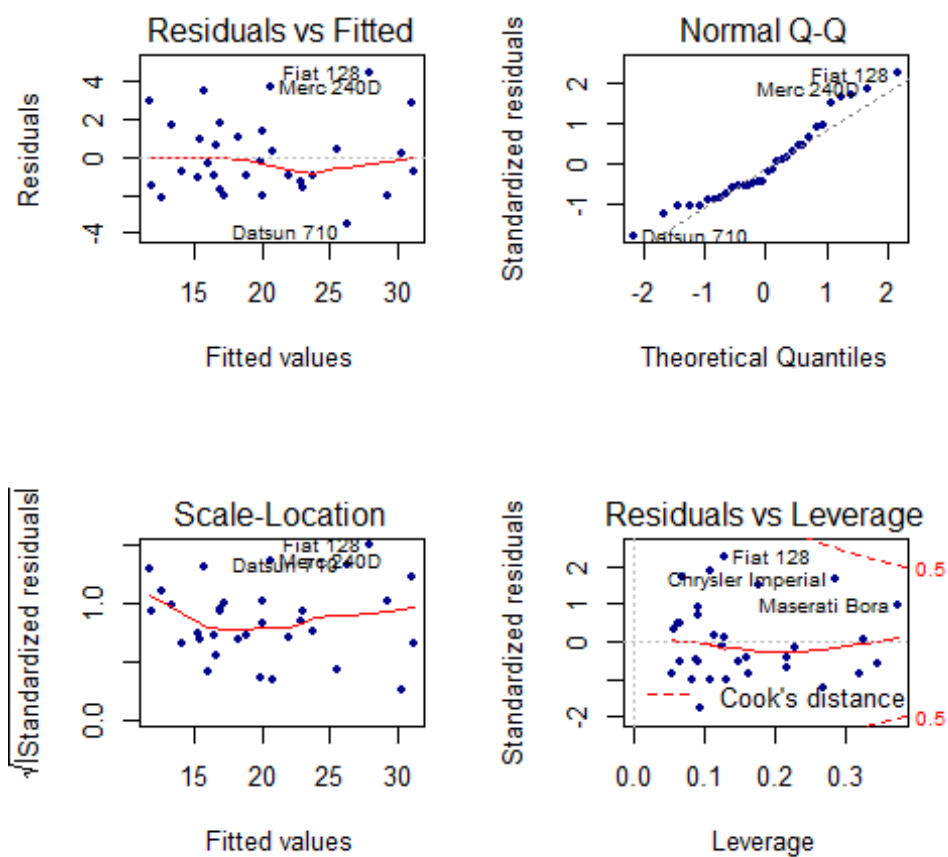


Figure 15: Diagnostic plots

