Report on Fuel Consumption of automobiles

May 22, 2015

Introduction

We investigate data extracted from the 1974 *Motor Trend* US magazine. They comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models). We are particularly interested in the following questions:

- Is an automatic or manual transmission better for MPG?
- Quantify the MPG difference between automatic and manual transmissions.

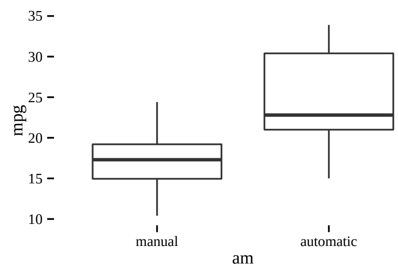

Figure 1: Automatic vs. manual.

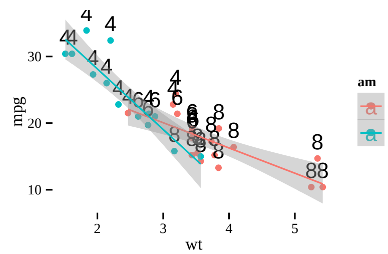

Figure 2: Impact of weight, transmission and number of cylinders.

The boxplot on Figure 1 suggests that manual transmission is better for mpg. However, Figure 2 shows that actually weight or number of gears can be a most important factor. Let's see.

Model Selection

Let us try to identify the subset of the predictors that can be related to the mpg response. For each $n = 1 \ldots 8$ we select the best model with n predictors. And then comparing R^2 statistics we try to select n.

For that we use `regsubsets` function from `leaps` R library. These are the results.

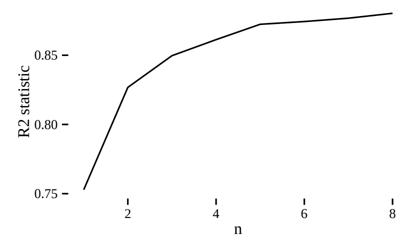

Figure 3: R^2 statistic for the best model for $n = 1, \ldots, 8$.

	cyl6	cyl8	disp	hp	drat	wt	qsec	vs1	am1	gear4	gear5	carb2	carb3	carb4	carb6	carb8
1 (1)						*										
2 (1)				*		*										
3 (1)						*	*		*							
4 (1)	*			*		*			*							
5 (1)	*			*		*		*	*							
6 (1)	*			*		*		*	*		*					
7 (1)		*		*		*	*	*	*		*					
8 (1)	*		*	*		*				*	*	*				*

If we take a look at Figure 3 (where we plot R^2 statistic) we see that $n = 5$ would be the best choice. Therefore our regression model is the following has predictors horsepower (hp), weight (wt), manual transmission (am=1), six cylindes (cyl=6) and V/S (vs=1). And the regression coefficients are given in the following table.

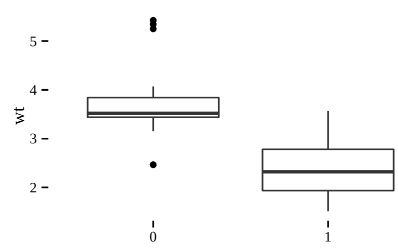

Figure 4: Boxplots of weight

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	31.2824	3.1902	9.8058	0.0000
hp	-0.0339	0.0104	-3.2504	0.0032
wt	-2.3678	0.8686	-2.7262	0.0113
am1	2.6211	1.3000	2.0163	0.0542
I(cyl == 6)TRUE	-2.2052	1.0321	-2.1365	0.0422
I(vs == 1)TRUE	1.8774	1.2481	1.5042	0.1446

Conclusion

Is an automatic or manual transmission better for MPG?

In my opinion the results are not conclusive. This regression model suggest that actually the most significant characteristics are horsepower and weight. One can clearly sees that the heavier cars have automatic transmission (Figure 4). This could be an effect of that economic car (in the sense of price) are smaller and tends to have manual transmission while luxury cars are bigger and its transmission is automatic in general.

The MPG difference between automatic and manual transmissions

The difference is clearly visible on Figure 1. Let us do however *t-test* to test if the difference in mean is significant.

```
##
##  Welch Two Sample t-test
##
## data:  mtcars[mtcars$am == 0, 1] and mtcars[mtcars$am == 1, 1]
## t = -3.7671, df = 18.332, p-value =
## 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean of x mean of y
##  17.14737  24.39231
```

Hence the difference in mean is significant, however *correlation does not imply causation*, and as we have mentioned above it is likely to be caused by other factors like weight and horsepower. For example, on Figure 5 we presented how mpg depends on weight and on Figure 6 the residual plot for the linear regression. And on the next figures we did the same for horsepower. These relations definitely make sense.

The code of this file is available on my github:
`https://github.com/sbartek/mpg_report`

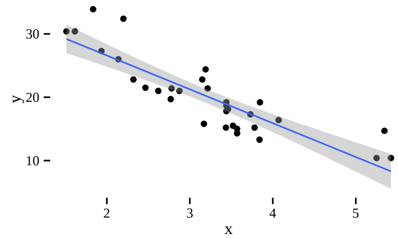

Figure 5: Relation beteewn mpg and weight

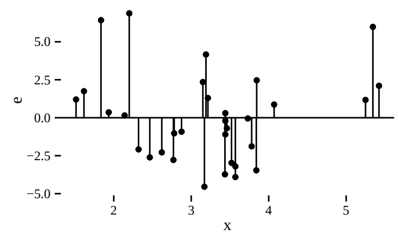

Figure 6: Residual plot for regresion beteewn mpg and weight

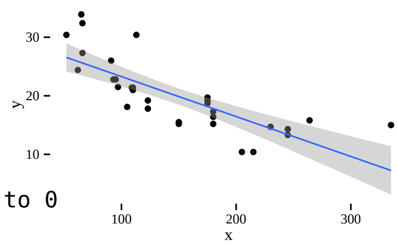

Figure 7: Relation beteewn mpg and horsepower

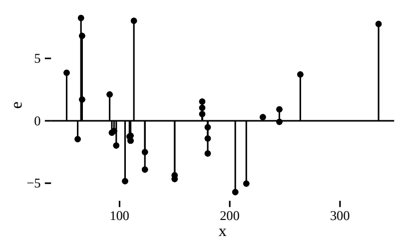

Figure 8: Residual plot for regresion beteewn mpg and horsepower

Appendix

```
pairs(mtcars.dt[, .(mpg, cyl, disp, hp, drat,
    wt, qsec, vs, am, gear, carb)], panel = panel.smooth)
```

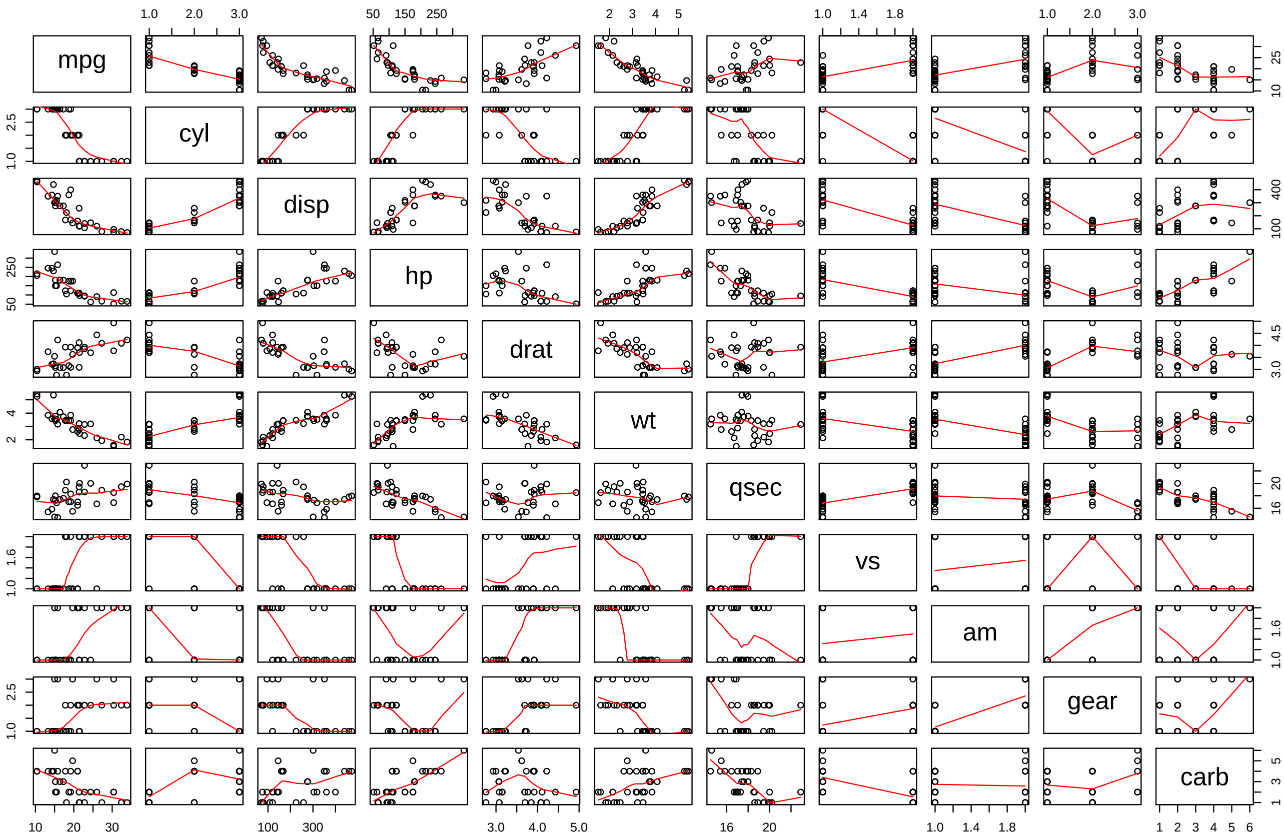

Figure 9: Pairs figure. Here one sees how strongly the variables are corealted with each other.

```
g1 <- ggplot(mtcars.dt, aes(wt, hp, colour = am)) +
    geom_point() + geom_text(aes(label = cyl,
    colour = NULL), vjust = -0.6, size = 1.5) +
    theme.bartek + geom_smooth(method = "lm")
g2 <- ggplot(mtcars.dt[!(cyl == 8 & am == 1)],
    aes(wt, hp, colour = am)) + geom_point() +
    geom_text(aes(label = cyl, colour = NULL),
        vjust = -0.6, size = 1.5) + theme.bartek +
    geom_smooth(method = "lm")
grid.arrange(g1, g2, ncol = 2)
```

```
ggplot(mtcars.dt, aes(disp, mpg, colour = am)) +
    geom_point() + theme.bartek + geom_smooth(method = "lm")
```

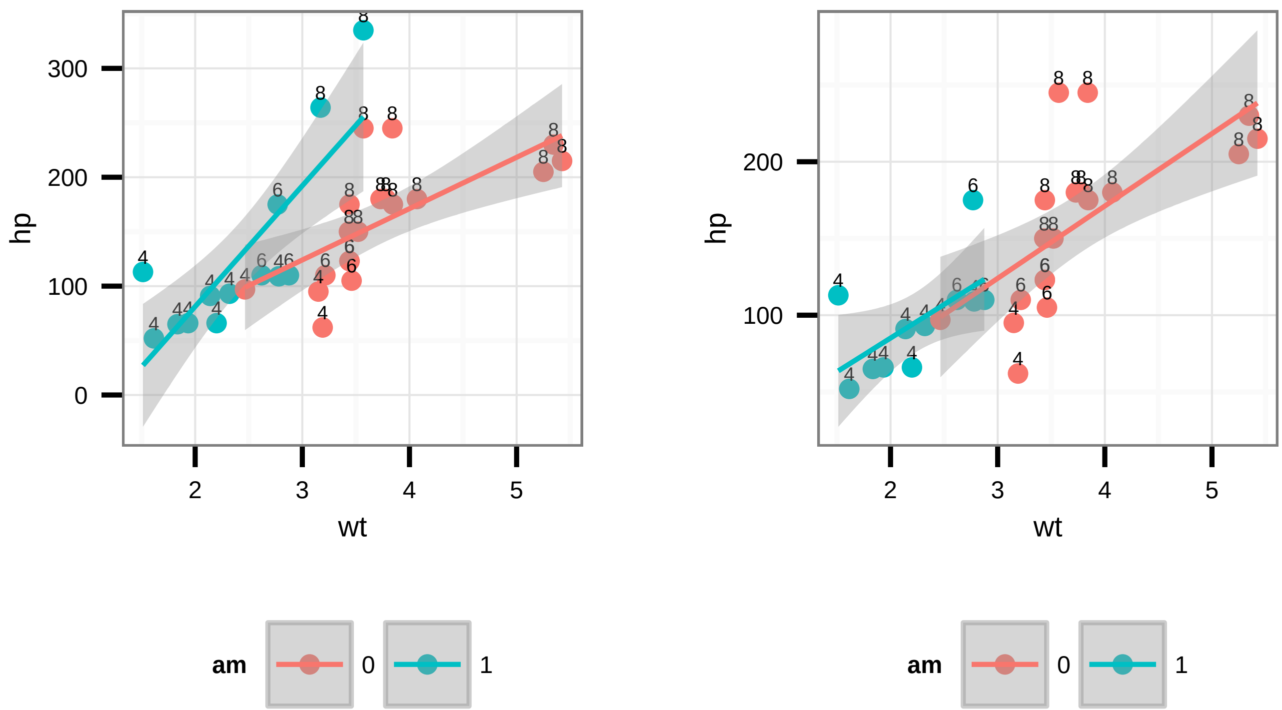

Figure 10: Horsepower vs. weight. If we remove two cars that have manual transmission but 8 cylinders (probably sport cars) there is almost no interaction between weight and type of transmission.

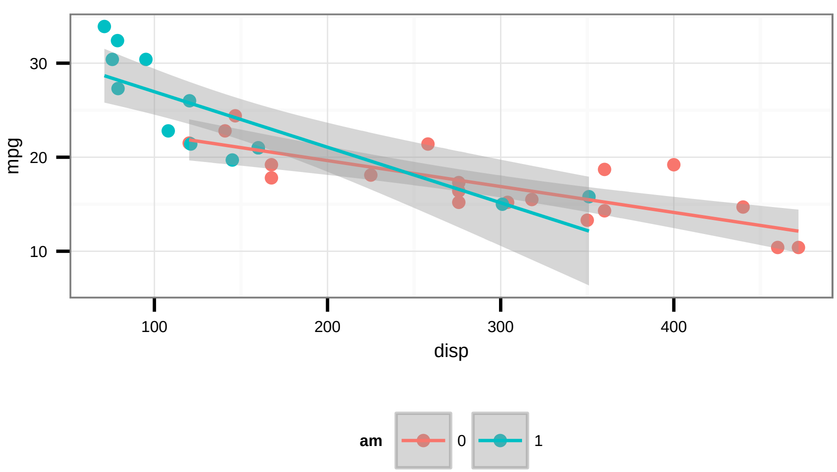

Figure 11: Displacement vs. mpg. Almost no interaccion with transmition type.