Report on Fuel Consumption of automobiles

May 13th, 2015

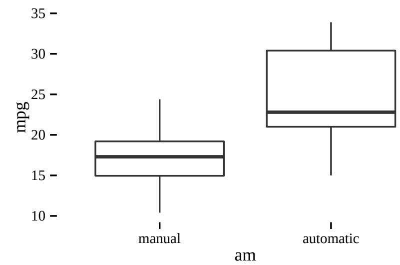

Figure 1: Automatic vs. manual transmission.

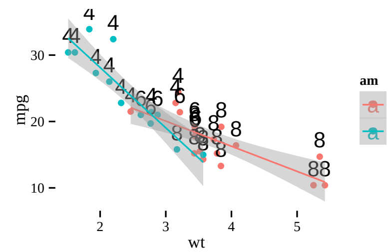

Figure 2: Impact of weight, transmission and number of cylinders on fuel consumption.

Introduction

We investigate data that 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.

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 that we use `regsubsets` function from `leaps` library which select the best model with n predictors, for $n = 1 \dots 8$. This way we can choose from quite large amount of models. Additionally in Appendix we show plots of few models with different sets of predictors.

	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)	*		*	*		*				*	*	*				*

```
ggplot(data = NULL, aes(x = 1:length(reg.summary$rsq),
    y = reg.summary$rsq)) + geom_line() + theme_tufte() +
    labs(x = "n", y = "R2 statistic")
```

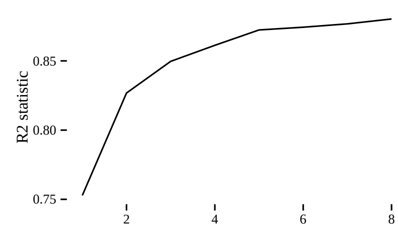

Figure 3: R^2 statistic for the best model with $n = 1, \dots, 8$.

If we take a look at Figure 3 where we plot R^2 statistic for the best model with $n = 1, ..., 8$ predictors, we see that $n = 5$ would be the best choice. Therefore our regression model is the following.

```
fit <- lm(mpg ~ hp + wt + am + I(cyl == 6) + I(vs ==
    1), data = mtcars.dt)
print(xtable(summary(fit)$coefficients), size = "\\small")
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	31.28	3.19	9.81	0.00
hp	-0.03	0.01	-3.25	0.00
wt	-2.37	0.87	-2.73	0.01
am1	2.62	1.30	2.02	0.05
I(cyl == 6)TRUE	-2.21	1.03	-2.14	0.04
I(vs == 1)TRUE	1.88	1.25	1.50	0.14

Is an automatic or manual transmission better for MPG?

```r
ggplot(mtcars.dt, aes(x = am, y = wt)) + geom_boxplot() +
    theme_tufte()
```

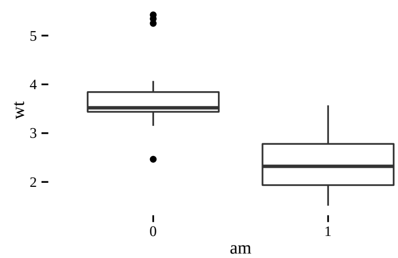

Figure 4: Boxplots of weight

In my opinion the results are not conclusive. This regression model suggest that actually the most significant is characteristics are horsepower and weight. One can clearly see that the more heavy 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.

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.

```r
g1 <- mtcars.dt[am == 0, mpg]
g2 <- mtcars.dt[am == 1, mpg]
t.test(g1, g2, paired = FALSE, var.equal = FALSE)
```

```
Welch Two Sample t-test
```

data: g1 and g2 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

Appendix

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

```r
ggplot(mtcars.dt, aes(am, mpg)) + geom_boxplot() +
    theme_tufte() + scale_x_discrete(labels = c("manual",
    "automatic"))
```

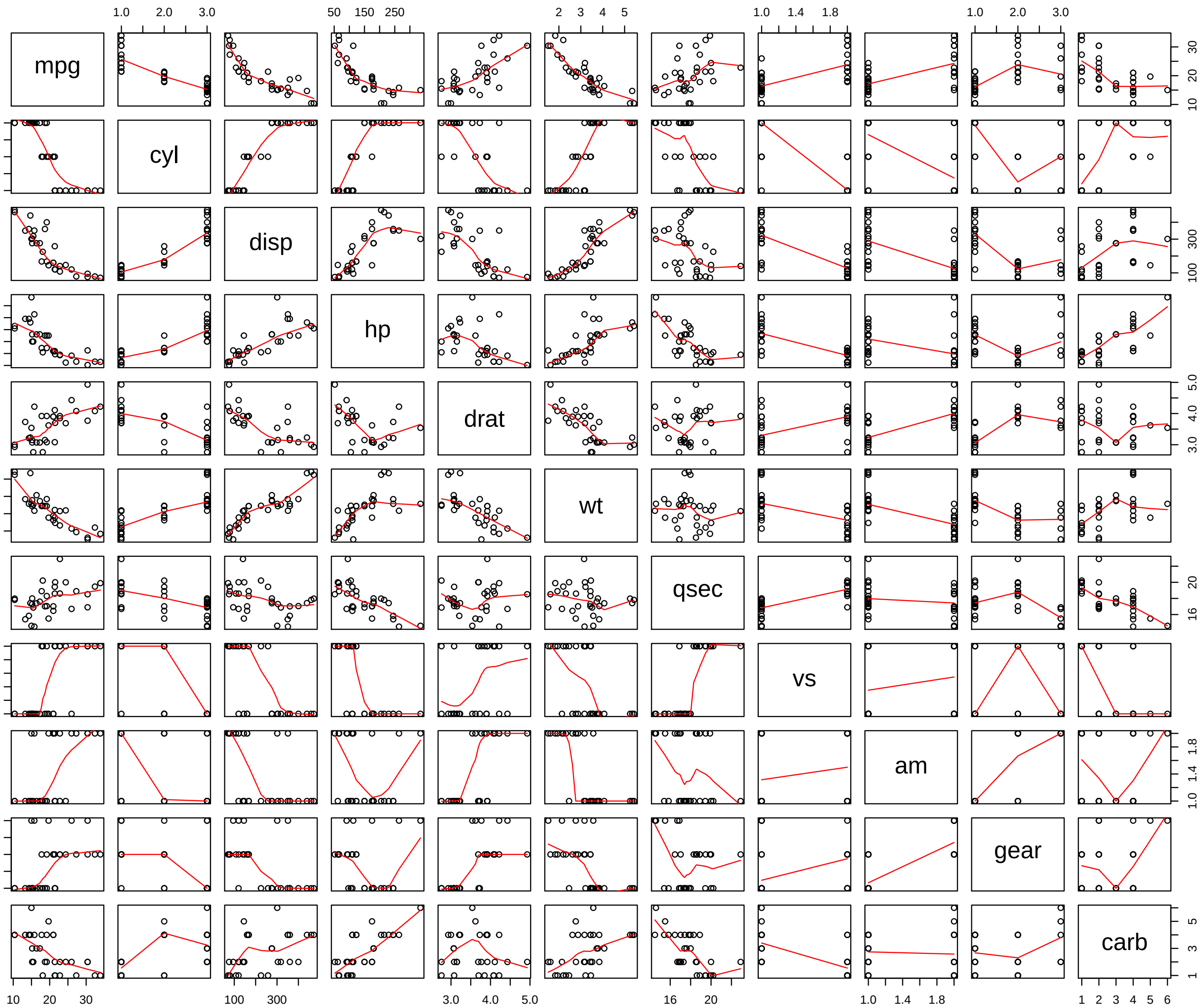

Figure 5: Pairs figure.

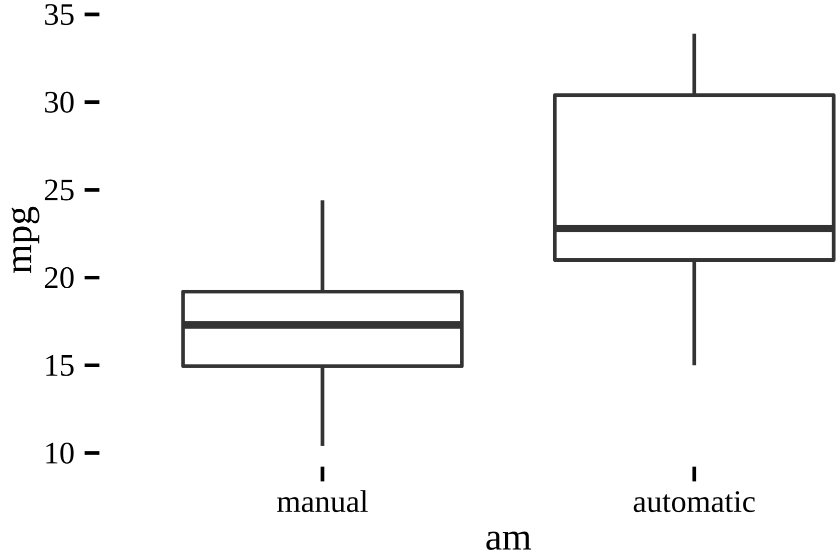

Figure 6: Automatic vs. manual transmission.

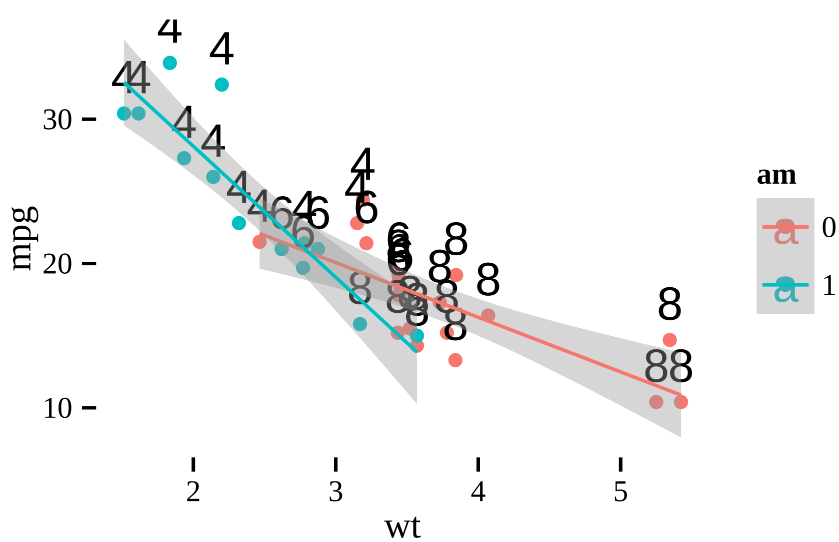

Figure 7: Impact of weight, transmission and number of cylinders on fuel consumption.

```
ggplot(mtcars.dt, aes(wt, mpg, colour = am)) +
    geom_point() + theme_tufte() + geom_text(aes(label = cyl,
    colour = NULL), vjust = -0.6) + geom_smooth(method = "lm")
```

```
ggplot(mtcars.dt, aes(hp, mpg, colour = am)) +
    geom_point() + geom_text(aes(label = cyl,
    colour = NULL), vjust = -0.6) + theme_tufte() +
    geom_smooth(method = "lm")
```

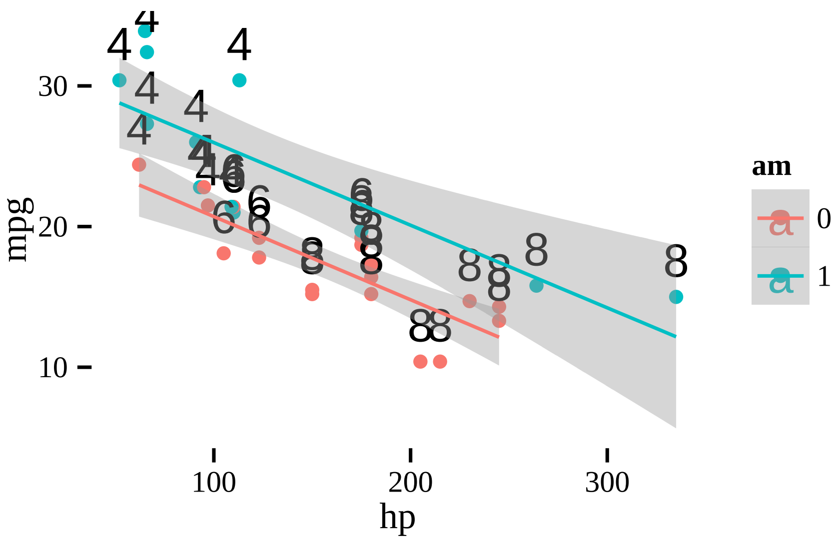

Figure 8: Sepal length vs. petal length, colored by species

```
ggplot(mtcars.dt, aes(hp, wt, colour = am)) +
    geom_point() + geom_text(aes(label = cyl,
    colour = NULL), vjust = -0.6) + theme_tufte() +
    geom_smooth(method = "lm")
```

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

```
ggplot(mtcars.dt, aes(wt, disp, colour = am)) +
    geom_point() + theme_tufte() + geom_smooth(method = "lm")
```

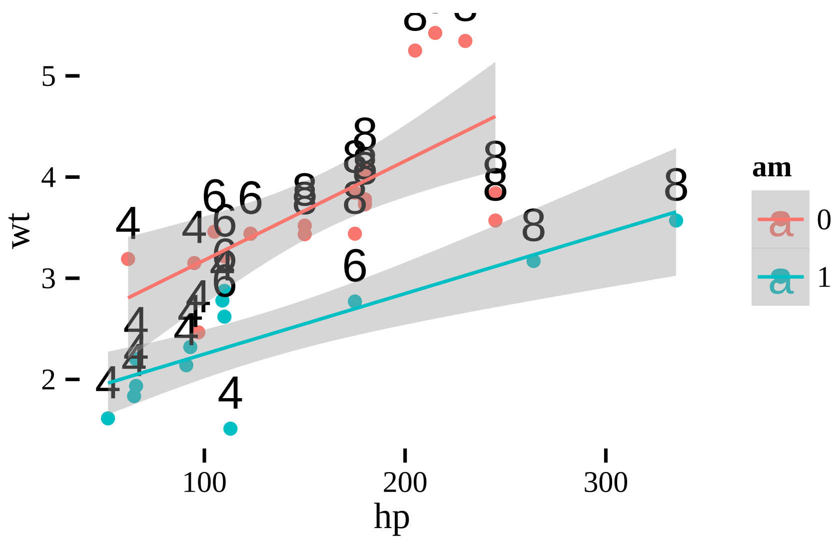

Figure 9: Sepal length vs. petal length, colored by species

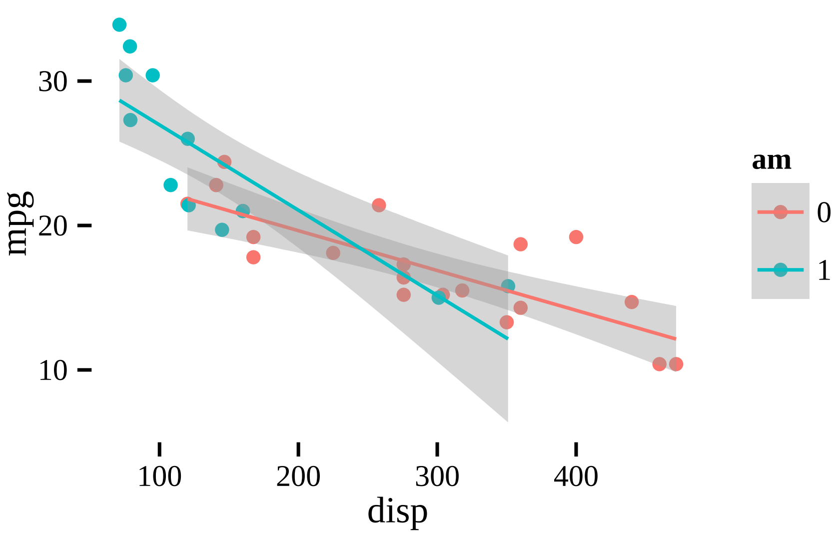

Figure 10: Sepal length vs. petal length, colored by species

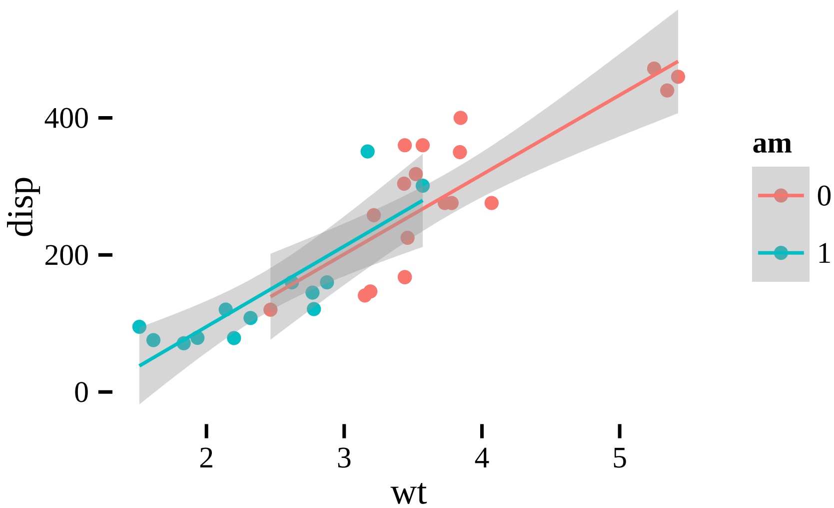

Figure 11: Sepal length vs. petal length, colored by species