Mileage and Transmission

Enelen

Executive Summary

This report analyses motor vehicle data to find out the difference in average mileage between vehicles with automatic transmission and those with manual transmission. While a difference in mean mileage between the two was found, the difference was due to other vehicle factors, viz. the number of cylinders, the horsepower and the weight of the vehicle. After adjusting for these two, there was insignificant difference between the mileage of the two transmission systems.

Objective

In this report, we are going to analyse the data collected by *Motor Trend* about various motor vehicles to answer two questions :

1. Automatic or manual transmission, which one gives better mileage?

2. How much is the difference?

Let us perform some initial processing of our data first.

```r
data(mtcars)
df <- mtcars  # copying to make changes to dataset
# the column id of the factor variables in the dataset
fac <- c(2, 8, 9, 10, 11)

# converting factor variables to factor from numeric
for (i in fac) {
    df[, i] <- as.factor(df[, i])
}
df$am <- factor(df$am, labels = c("Automatic", "Manual"))
```

Analysis

A boxplot of the mileage versus transmission (see appendix) seems to indicate a very clear difference between the two transmission modes. Let us just build a very basic regression model between mileage and transmission and see what result we get.

```r
fit <- lm(mpg ~ am, df)
pander(summary(fit)$coef, digits = 3)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	17.1	1.12	15.2	1.13e-15
amManual	7.24	1.76	4.11	0.000285

From this we can infer that am1 (which refers to manual transmission) leads to significantly higher mileage on average (p-value <0.001). Or more specifically, manual transmission on average offers 7.24 more miles per US gallon than automatic transmission. So for automatic transmission, average is (represented by the intercept) is 17.1 miles/gallon, and thus our average mileage for manual transmission becomes 24.39 miles/gallon.

To see the effects of other variables, let us build a more complex model which includes various other variables that are significant. After more exploration of the dataset, and testing various models with different combinations of variables, the model with cylinders, transmission,horsepower and weight was found to be the best, with an adjusted R-squared of ~84% and no discernible pattern or changing variation in the residuals (see appendix for diagnostic plots). Let us look at this model.

```
fittest <- lm(mpg ~ cyl + am + wt + hp, df)
pander(summary(fittest))
```

	Estimate	Std. Error	t value	Pr(>|t|)
cyl6	-3.03	1.41	-2.15	0.0407
cyl8	-2.16	2.28	-0.947	0.352
amManual	1.81	1.4	1.3	0.206
wt	-2.5	0.886	-2.82	0.00908
hp	-0.0321	0.0137	-2.35	0.0269
(Intercept)	33.7	2.6	12.9	7.73e-13

Table 3: Fitting linear model: mpg ~ cyl + am + wt + hp

Observations	Residual Std. Error	R^2	Adjusted R^2
32	2.41	0.866	0.84

The coefficient of transmission(am1) is now insignificant! It seems the difference in transmission seen earlier, was mostly the result of other variables, most prominently the number of cylinders in the vehicle, its gross horsepower and its weight. The mileage is inversely proportional to all of them (decreases with increase in them).

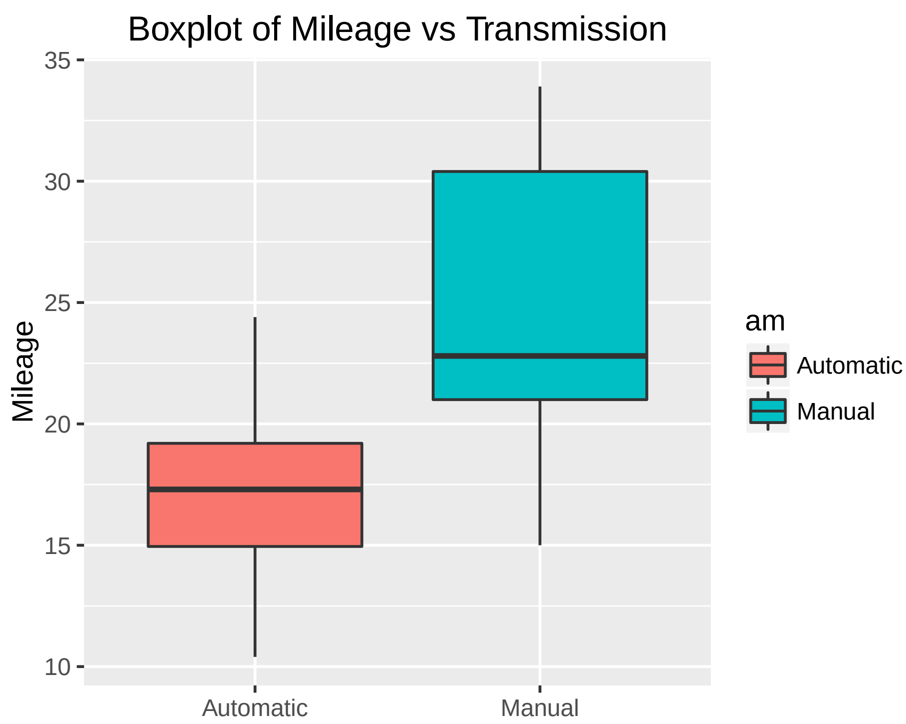
Result

In conclusion, the choice of transmission does not affect mileage significantly. Though it seems to do so at first sight, it is due to the number of cylinders, gross horsepower and weight of the vehicle. Since there is no significant difference between the two choices of transmissions, our second question becomes void.

Appendix

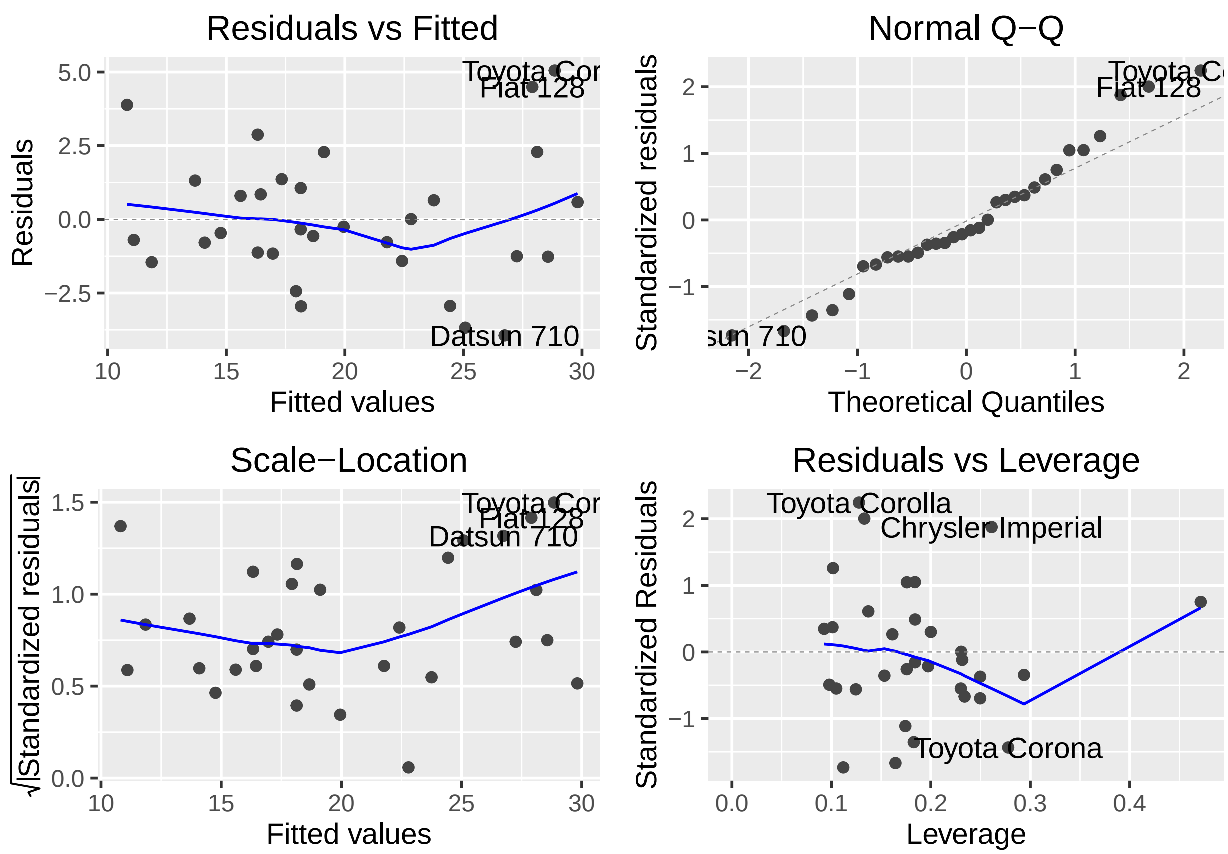
A boxplot of mileage vs transmission.

```
require(ggplot2)
ggplot(df, aes(x = am, y = mpg, group = am, fill = am)) +
     geom_boxplot() +
     scale_x_discrete(labels = c("Automatic","Manual")) +
     labs(x = "",y = "Mileage",title = "Boxplot of Mileage vs Transmission")
```



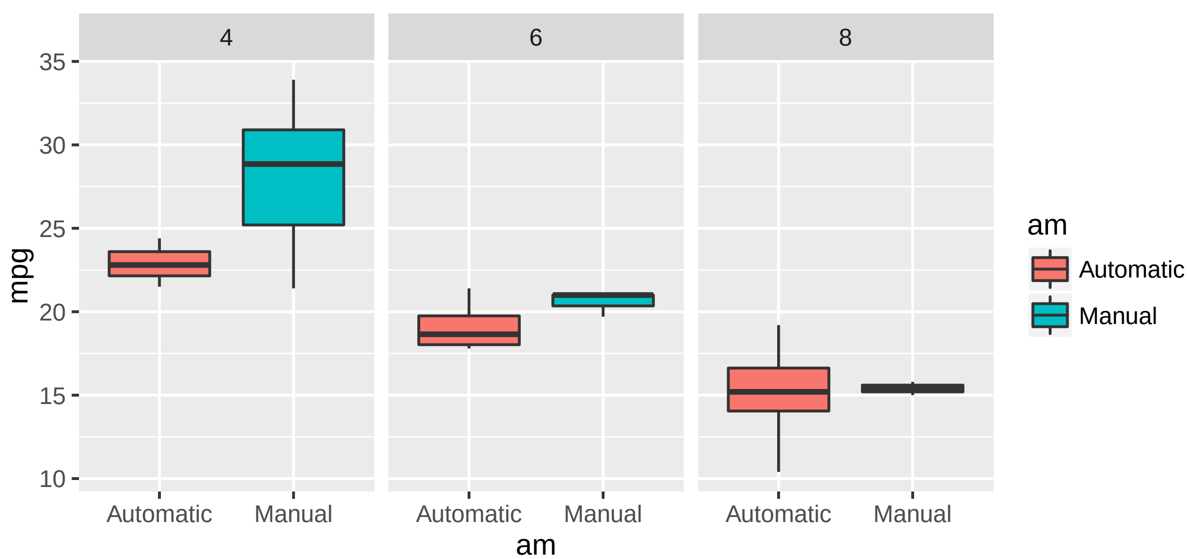
Diagnostic plots of final model

```
require(ggfortify)  ## extends ggplot for other types of objects.
autoplot(fittest)
```

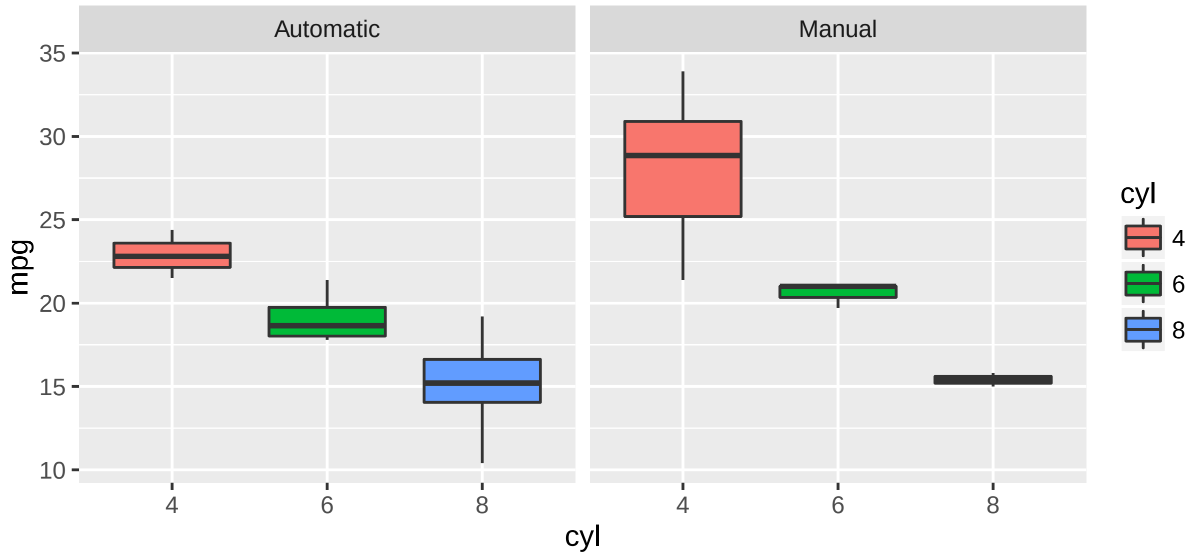


Some other exporatory analysis

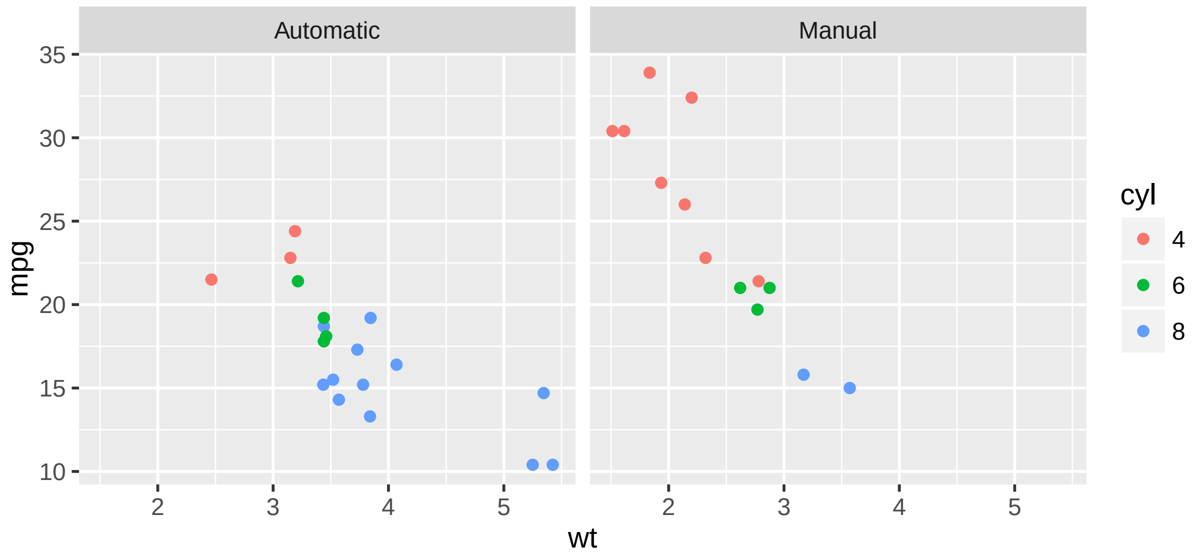
```r
ggplot(df, aes(am, mpg, fill = am)) + geom_boxplot() + facet_wrap(facets = ~cyl)
```



```r
ggplot(df, aes(cyl, mpg, fill = cyl)) + geom_boxplot() + facet_wrap(facets = ~am)
```




```r
ggplot(df, aes(wt, mpg, col = cyl)) + geom_point() + facet_wrap(facets = ~am)
```




```r
# wt and cyl seem to be much greater predictor of mpg from these
```