Analysis of Transmission Type on Fuel Consumption

In this report we will analyze the fuel consumption os 32 vehicles contained in the `mtcars` dataset. This data was extracted from the 1974 *Motor Trend US* magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models). Our main questions are to explore the relationship between fuel consumption and car transmission, defined as 'automatic' or 'manual', and quantify this relation (if it exists).

The mains conclusion from the analysis below are the following:

- `Manual` transmission are 1.8 times more efficient than an `Automatic` transmission;
- The transmission type is not the most important factor to analyze fuel consumption.

```
suppressMessages(require(magrittr))
```

Data Processing

The `mtcars` dataset is already presented in a structured way, so our only task here is to load the dataset and define which columns will be treated as qualitative variables in the analysis.

```
data(mtcars)
mtcars$am %<>% factor(levels = c(0,1), labels = c('Automatic', 'Manual'))
mtcars$carb %<>% as.factor; mtcars$cyl %<>% as.factor
mtcars$gear %<>% as.factor; mtcars$vs %<>% as.factor
```

Hypothesis Test

Let's first begin with a simple question: does the type transmission have a visible effect in the mpg value? Assuming the data comes from a normal distribution, we use a `t` test to determine if the two types of transmission are significantly different from each other regarding fuel consumption, or if we should accept accept the null hypothesis H_0 and declare that there's no distinction among them.

```
fit.base <- lm(mpg ~ am, mtcars); t.value <- t.test(mpg ~ am, mtcars)
```

Since the p-value for the `t` test is 0.00137, below the 0.05 limit for a 95% confidence interval, we reject the null hypothesis. We also find that cars with manual transmissions can travel on average 7.245 miles more than cars with automatic transmission with a fuel gallon (figure 2). The R^2 value of 36% for this model indicates that despite the fit the model doesn't does a good job at predicting the fuel consumption.

Model Analysis

Having rejected the null hypothesis, we now examine the remaining variables of the dataset for possible confounders. The pairs plot (figure 1) points us to other variables that seem to have a high correlation with `mpg`. To perform this test we will generate several models with different inputs, starting from a model that contains all the predictors, and measure their relative qualities through the application of the `AIC` criterion.

```
fit.best <- lm(mpg ~ ., mtcars) %>% step
```

The new model contains the variables `cyl`, `hp`, `wt`, plus `am`. The first thing to note is the huge improvement in the R^2 value, now at 87%. Applying the `anova` to the two models presents us a F-statistic of 24.527, indicating that the added variables contribute to the accuracy of the model. The complete details of new model are presented below.

```
summary(fit.best)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## amManual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

Residuals Analysis

In this last section we shall study the residuals of the regression analysis. From the top left plot of figure 3 we can observe that there is no noticeable pattern in the residuals, which indicates a good fit. The Q-Q plot (top right) and the Scale-Location plot we see that the residuals are normally distributed (with a constant variance), although some outliers can be seem. Finally, the leverage plot show the presence of some high leverage points.

Conclusions

After our analysis we can conclude the following:

1. For constant values of `cyl`, `hp`, and `wt`, cars with `Manual` are approximately `1.8` more efficient (on average) than cars with `Automatic` transmission.
2. In the presence of other variables, the `am` type `is not` the main influence in the fuel consumption of a car (p-value of 0.206).
3. `hp` is not a significant factor regarding fuel consumption.
4. Adjusted by `hp`, and `am`, the most significant factors regarding fuel consumption are `cyl` and `wt`.

Appendix

Figure 1

```r
palette(c('red', 'blue')); par(xpd = TRUE)
pairs(mtcars[, -9], col=mtcars$am, bg=mtcars$am, pch=21, oma=c(4,4,16,4))
legend(x='topright', legend=levels(mtcars$am), horiz=TRUE,
       pt.bg=palette(), pch=21)
```

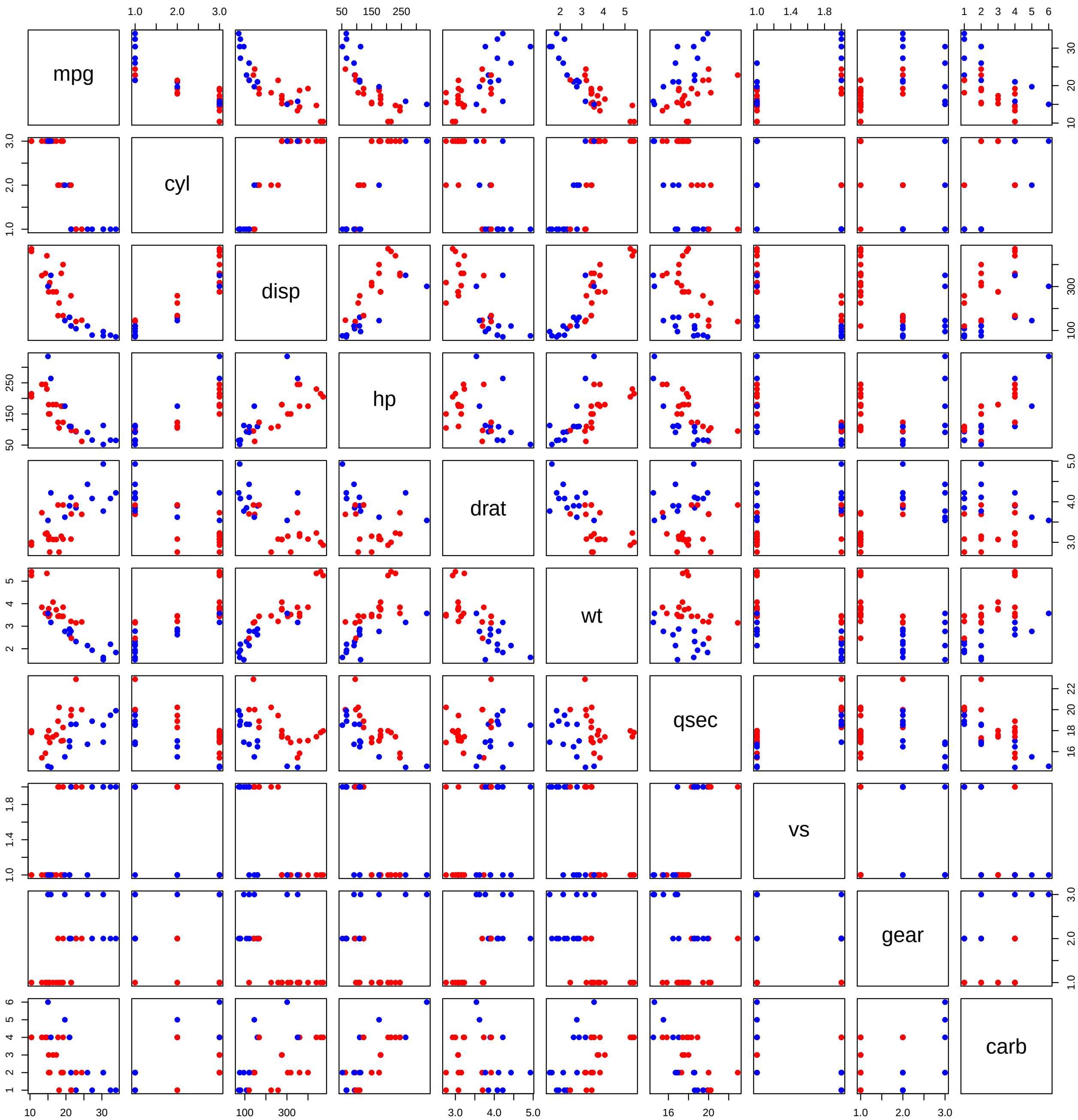
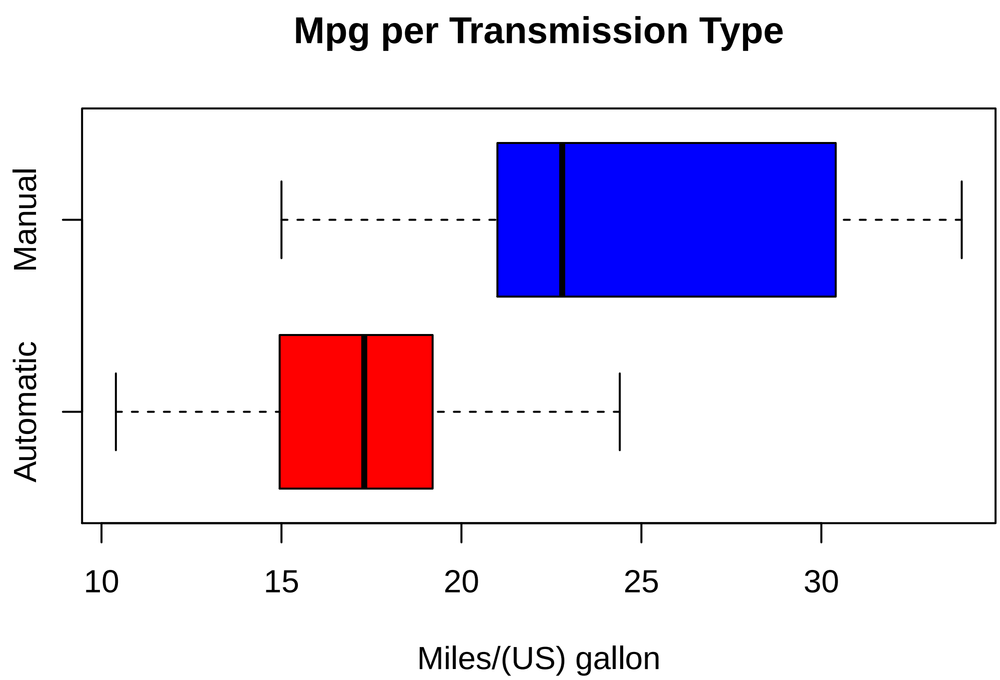


Figure 2

```
boxplot(mpg ~ am, mtcars, col=palette(), horizontal=TRUE,
        main='Mpg per Transmission Type',
        xlab='Miles/(US) gallon')
```




```
d1 <- density(subset(mtcars, am == 'Automatic')$mpg)
d2 <- density(subset(mtcars, am == 'Manual')$mpg)
plot(d1, col=palette()[1], xlim=range(d1$x, d2$x), ylim=range(d1$y, d2$y),
     main='', xlab='Miles/(US) gallon'); lines(d2, col=palette()[2])
legend('topright', c('Automatic', 'Manual'),
       lty=c(1,1), lwd=c(2.5,2.5),col=palette())
```

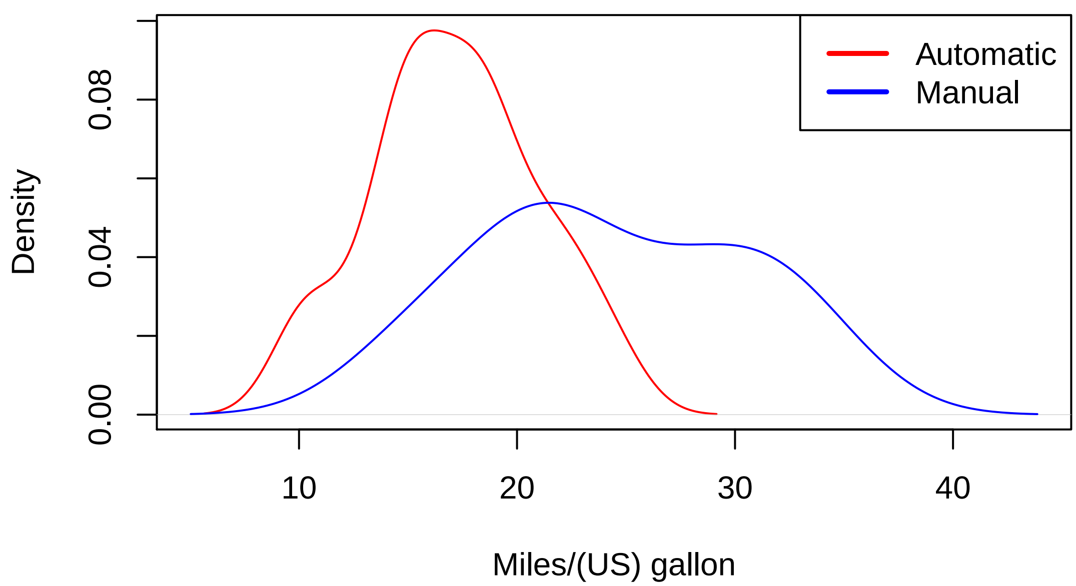


Figure 3

```
par(mfrow=c(2, 2))
plot(fit.best)
```