# Motor Trend

*Hugo Janssen*

*Monday, July 13, 2015*

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

"Is an automatic or manual transmission better for MPG" "Quantify the MPG difference between automatic and manual transmissions"

Question

Take the mtcars data set and write up an analysis to answer their question using regression models and exploratory data analyses.

Your report must be:

Written as a PDF printout of a compiled (using knitr) R markdown document. Brief. Roughly the equivalent of 2 pages or less for the main text. Supporting figures in an appendix can be included up to 5 total pages including the 2 for the main report. The appendix can only include figures. Include a first paragraph executive summary.

Did the student do some exploratory data analyses? Did the student fit multiple models and detail their strategy for model selection? Did the student answer the questions of interest or detail why the question(s) is (are) not answerable? Did the student do a residual plot and some diagnostics? Did the student quantify the uncertainty in their conclusions and/or perform an inference correctly?

## Executive Summary

Intuitively, cars with larger weights have less miles per gallon. For figures, please refer to Appendix 1: Figures.

## Data Overview

describe fields here

A data frame with 32 observations on 11 variables.

```
xtable(head(mtcars, 4), caption = "The first entries of the mtcars table")
```

% latex table generated in R 3.1.3 by xtable 1.7-4 package % Wed Jul 22 17:26:43 2015

| | mpg | cyl | disp | hp | drat | wt | qsec | vs | am | gear | carb |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Mazda RX4 | 21.00 | 6.00 | 160.00 | 110.00 | 3.90 | 2.62 | 16.46 | 0.00 | 1.00 | 4.00 | 4.00 |
| Mazda RX4 Wag | 21.00 | 6.00 | 160.00 | 110.00 | 3.90 | 2.88 | 17.02 | 0.00 | 1.00 | 4.00 | 4.00 |
| Datsun 710 | 22.80 | 4.00 | 108.00 | 93.00 | 3.85 | 2.32 | 18.61 | 1.00 | 1.00 | 4.00 | 1.00 |
| Hornet 4 Drive | 21.40 | 6.00 | 258.00 | 110.00 | 3.08 | 3.21 | 19.44 | 1.00 | 0.00 | 3.00 | 1.00 |

Table 1: The first entries of the mtcars table

## Exploratory Analysis

See figure 1.

First, fit an univariate model where the MPG is predicted by the transmission type alone. This provides a benchmark model against which other models can be evaluated.

```
fit0 <- lm(mpg ~ am, data = mtcars)
summary(fit0)$coefficients
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am           7.244939   1.764422  4.106127 2.850207e-04
```

This model is .. and with and `R-squared` of explains 0.3597989 of the variance.

```
fit1 <- lm(mpg ~ am + wt, data = mtcars)
summary(fit1)$coefficients
```

```
##                Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
## am          -0.02361522  1.5456453 -0.01527855 9.879146e-01
## wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
```

This model is .. and with and `R-squared` of explains already 0.7528348 of the variance.

When comparing the two models

```
anova(fit0, fit1)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 46.115 1.867e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The residual plot shows..

## Conclusion

You can also embed plots, for example:

## Appendix 1: Figures

### Figure 1. Boxplot of MPG per transmission type

```
ggplot(mtcars, aes(factor(am), mpg)) +
    geom_boxplot() +
    geom_point(size = 5, aes(colour = factor(am)), alpha = 0.6) +
    labs(title = "MPG per transmission type") +
    xlab("Transmission type (0 = Automatic, 1 = Manual)") +
    ylab("Miles/(US) gallon") +
    theme_minimal(base_family = "")
```

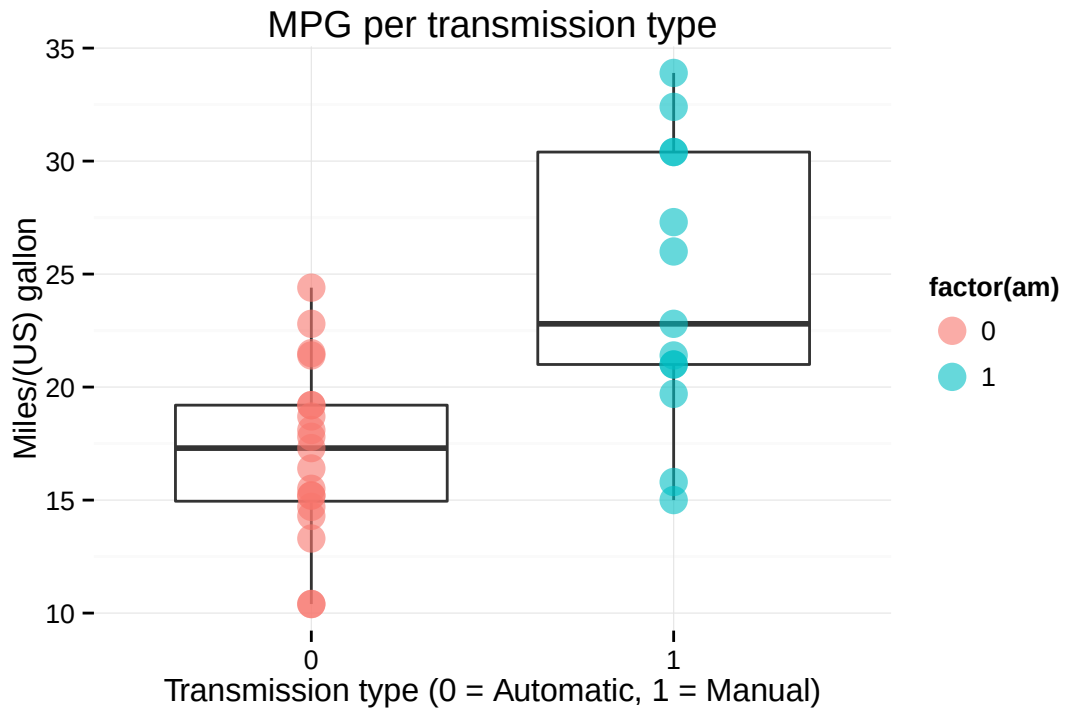



**Figure 2. Regression line**

```
ggplot(mtcars, aes(wt, mpg)) +
    geom_point(size = 5, aes(colour = factor(am)), alpha = 0.6) +
    stat_smooth(method = "lm", se = FALSE) +
    ggtitle("MPG per weight") +
    xlab("Weight (lb/1000)") +
    ylab("Miles/(US) gallon") +
    theme_minimal(base_family = "")
```

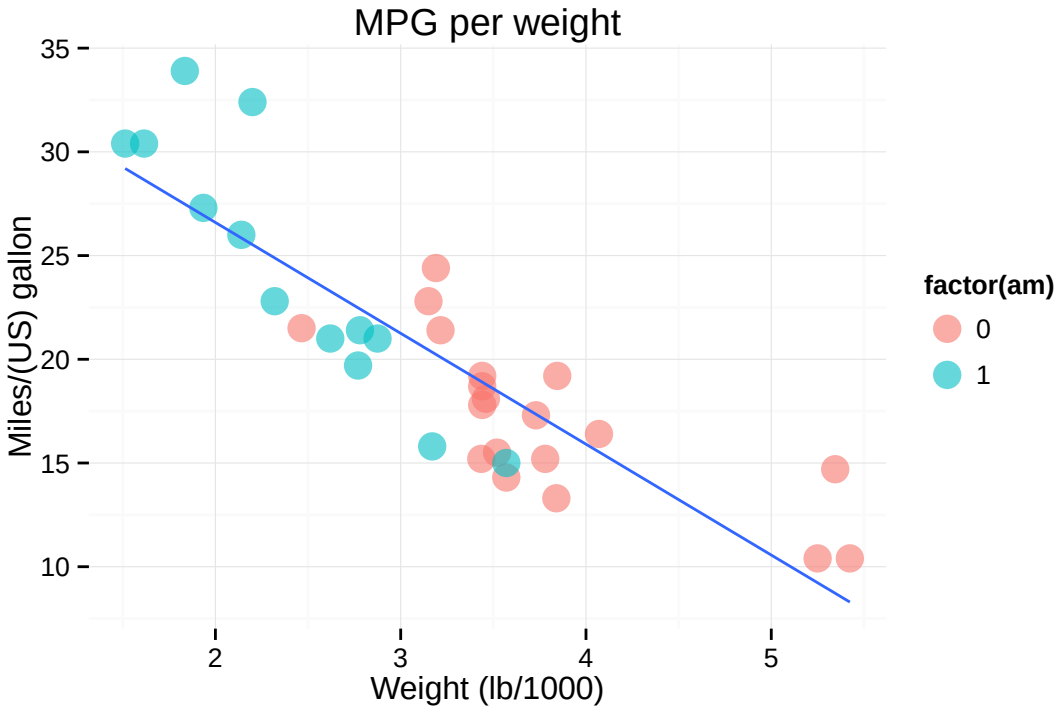
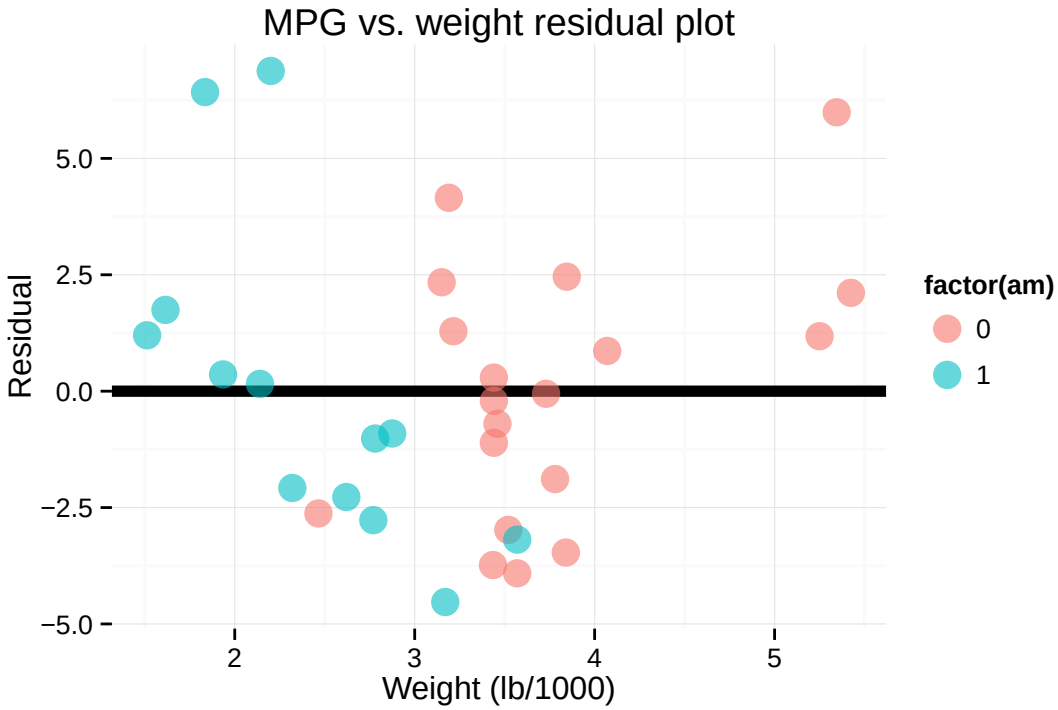



**Figure 3. Residual plot**

```
# residual plot = variation around the regression line
ggplot(mtcars, aes(wt, resid(fit1))) +
    geom_hline(yintercept = 0, size = 2) +
    geom_point(size = 5, aes(colour = factor(am)), alpha = 0.6) +
    ggtitle("MPG vs. weight residual plot") +
    xlab("Weight (lb/1000)") +
    ylab("Residual") +
    theme_minimal(base_family = "", base_size = 12)
```


MPG vs. weight residual plot
Residual
5.0
2.5
0.0
−2.5
−5.0
2
3
4
5
Weight (lb/1000)
factor(am)
0
1