MPG performance: manual or automatic transmission?

Guillaume Dreyer

March 20, 2015

Executive summary.

We present quantitative results regarding the MPG difference for manual and automatic transmission cars. Our study reveals that MPG is influenced by different factors in each group.

Main Report.

1. Some exploratory analysis.

As a first look at our data, let's calculate the **mean** of the values shown in the 'mtcars' dataset by aggregating the latter with respect to the (binary) variable 'am' (0: automatic, 1: manual). Also, it is natural to take into account the number of cylinders, so let's aggregate using the variable 'cyl' as well.

```
aggreg_mtcars
```

```
##      mpg cyl    disp      hp  drat    wt   qsec    vs am  gear  carb
## 1 22.900   4 135.867  84.667 3.770 2.935 20.970 1.000  0 3.667 1.667
## 3 19.125   6 204.550 115.250 3.420 3.389 19.215 1.000  0 3.500 2.500
## 5 15.050   8 357.617 194.167 3.121 4.104 17.143 0.000  0 3.000 3.083
## 2 28.075   4  93.612  81.875 4.184 2.042 18.450 0.875  1 4.250 1.500
## 4 20.567   6 155.000 131.667 3.807 2.755 16.327 0.000  1 4.333 4.667
## 6 15.400   8 326.000 299.500 3.880 3.370 14.550 0.000  1 5.000 6.000
```

For a given number of cylinders, the above dataset suggests that MPG performance of cars with manual transmission tends to be higher than that of automatic ones. A series of t tests confirms this obervation for **4 cylinders**. We obtain a **t statistic** equal to 2.8855021 and an **confidence interval** equal to 1.1178919, 9.2321081, which is **conlusive**. However, for **6 and 8 cylinders**, the **confidence intervals** are -1.0327346, 3.9160679 and -1.6395493, 2.3395493, respectively. Since they both contain 0, the t test is **inconclusive in those two cases**.

To conclude our exploritary analysis, we determine the most influencial factor for the variable 'mpg'. By computing the (absolute) values of the **correlation** between 'mpg' and the remaining variables, we conclude that the **weight** 'wt' has the strongest impact, as shown below.

```
cor_mpg
```

```
##    wt   cyl  disp    hp  drat    vs    am  carb  gear  qsec
## 0.868 0.852 0.848 0.776 0.681 0.664 0.600 0.551 0.480 0.419
```

The graph shown in **Fig 1** of the **Appendix** represents the **MPG** 'mpg' as a function of the **weight** 'wt'. We compute the linear regression of 'mpg' against 'wt': the **intercept** is 37.285, and the **slope** is -5.344.

2. Quantifying the MPG difference between automatics and manuals.

t tests in **Part 1** fail to answer whether the MPG difference between manuals and automatics is better for 6 and 8 cylinders. Therefore, We now want to approach the MPG difference differently by **understanding which factors influence the MPG**. Below, we calculate, for each group, the (absolute value of) correlation between 'mpg' and the remaining variables.

```
round(cor_mpg_a,3);round(cor_mpg_m,3);
```

```
##    hp   cyl  disp    wt    vs  qsec  carb  gear  drat
## 0.832 0.796 0.793 0.768 0.736 0.657 0.656 0.540 0.468
```

```
##    wt  disp   cyl  qsec    hp  carb    vs  drat  gear
## 0.909 0.835 0.826 0.802 0.801 0.774 0.725 0.470 0.402
```

For **automatics**, the most influencial factor is the **gross horsepower** 'hp', whereas, for **manuals**, it is the **weight** 'wt'. The graphs **Fig. 2** of the **Appendix** show 'mpg' as a function of 'wt' for manuals, and as a function of 'hp' for automatics, together with a **regression line** of 'mpg' against the other variable. For **manuals**, the diagnostic is very conlusive: we have a nice cluster og points along the regression line, no apparent outlier. For automatics, the cluster is a little more sparsed, though again no obvious outlier. The command 'dfbetas' in **R** enables us to quantify the influence of each point on the regression line, by comparing the difference of coefficients of the line regression, with and without each of the points. We present the values for **automatics** only, for space reason. Also, **Fig. 3** shows the **mpg residuals** in each group.

```
sort(abs(round(df_betas_a[,1],3)),decreasing=TRUE)
```

```
##     5     3     4     6    13    12    14    18     1     8    17    16
## 0.408 0.361 0.344 0.321 0.299 0.260 0.194 0.180 0.179 0.171 0.143 0.125
##    15     7    19     2    10    11     9
## 0.104 0.016 0.015 0.012 0.007 0.004 0.002
```

```
sort(abs(round(df_betas_a[,2],3)), decreasing=TRUE)
```

```
##    13     4    12     5     3    14     6    18     1     8    19    15
## 0.458 0.454 0.444 0.366 0.289 0.270 0.268 0.238 0.140 0.123 0.094 0.086
##     2    17    10    16    11     9     7
## 0.076 0.056 0.054 0.049 0.032 0.017 0.012
```

To conclude, a second correlation analysis between the **mpg residuals** and the remaining variables enables to determine the best regressor for the residuals, thus for 'mpg'. Fig. 4 shows, for each group, the graph of the **mpg residuals** as a function of its **best regressor**, together with the line regression. **In both groups, the clusters are quite sparsed, and as a result, we conclude that a multivariate linear model is not ideal**.

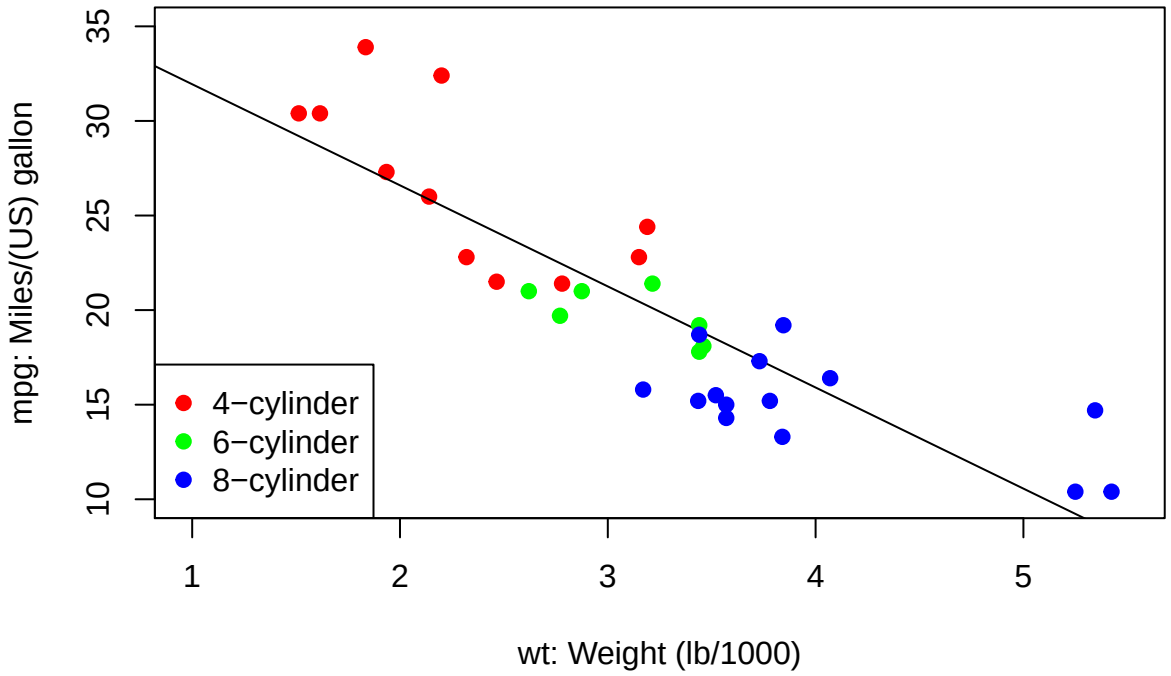
Conclusion.

In the case of **4 cylinders**, **manuals** clearly have a better MPG than **automatics**. However, this difference is unclear for the two entire groups, as the influencing factors differ from one group to another.

Appendix.

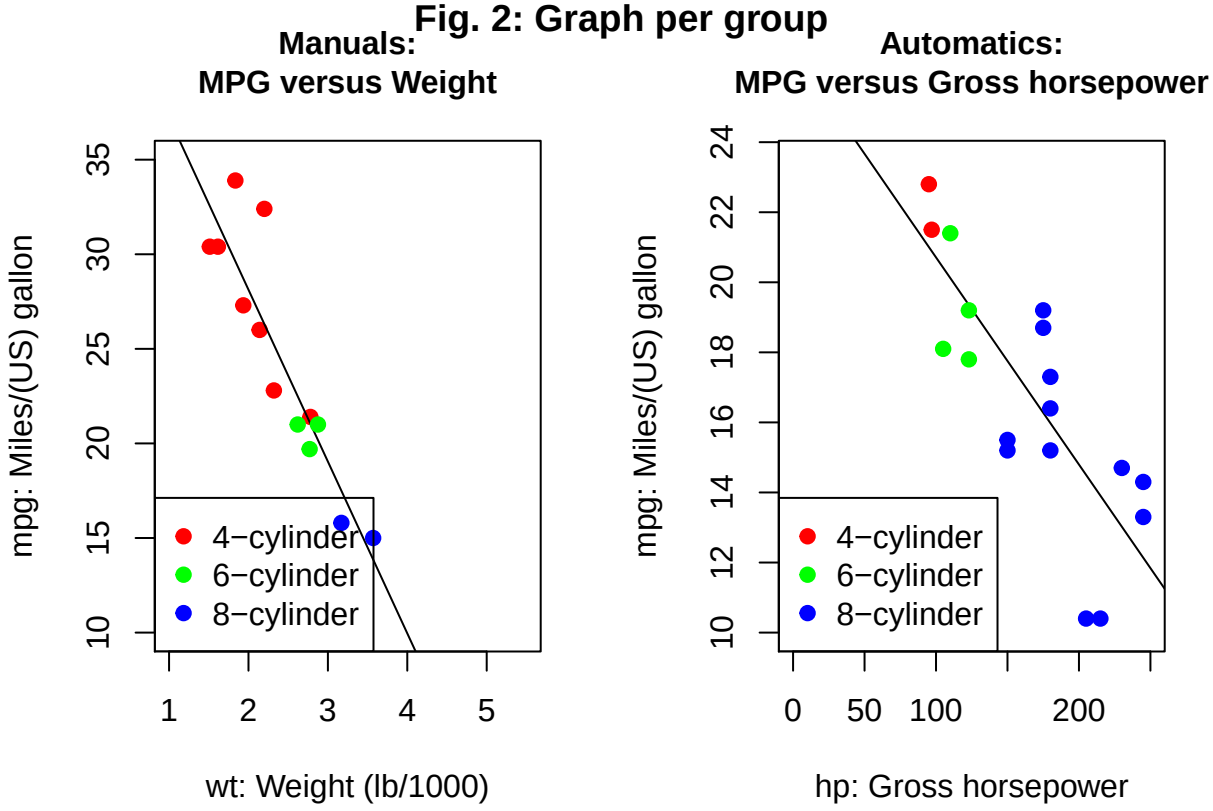
```
graph1()
```

Fig. 1: MPG versus Weight



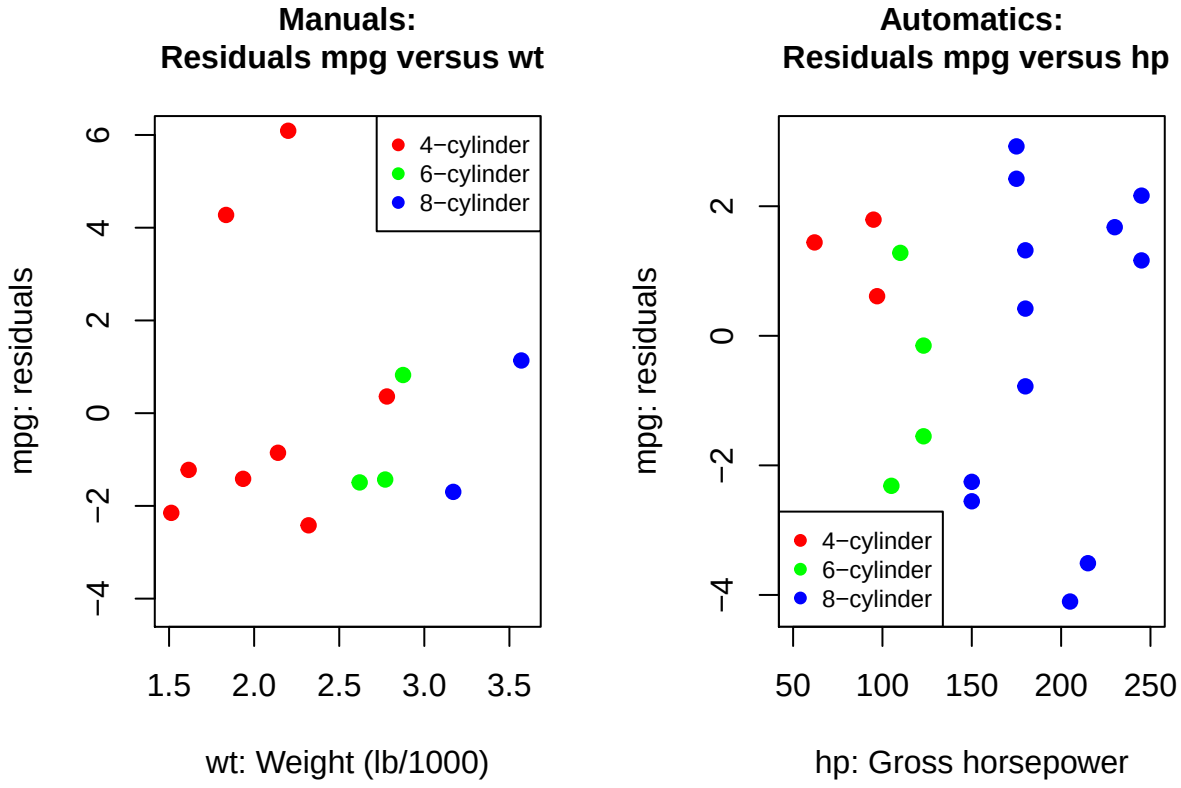

```
graph2()
```

Fig. 2: Graph per group



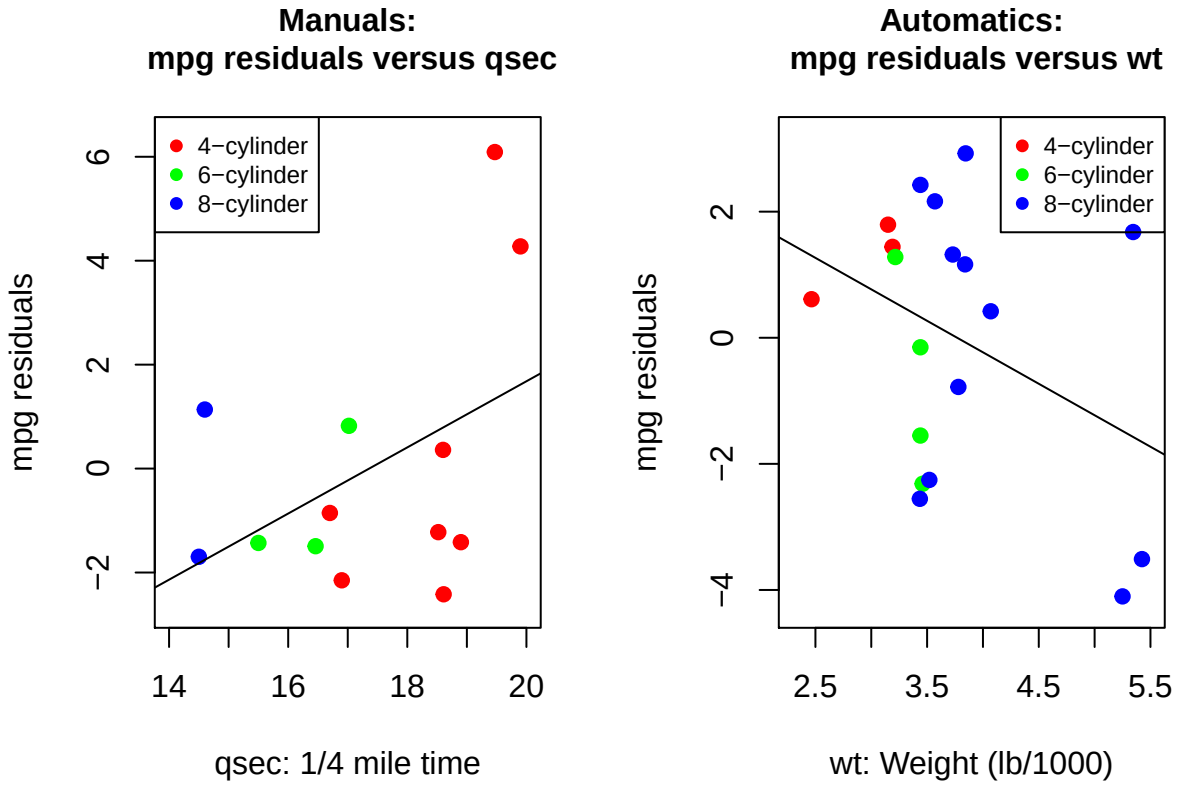

```
graph3()
```

Fig. 3: Residuals per group



```
graph4()
```

Fig. 4: Residuals per group against other variables




```
## [1] 29
```