Executive summary on Motor Trend Data from 1974

-Is Automatic or manual tramission better for MPG?

-Quantification of the MPG difference between automatic and manual transmissions.

This project is designed to answer the above questions. The data used in this project comes from the 1974 Motor Trend Magazine. This edition of the publication contained a dataset with information about fuel consumption and 10 aspects of automobile design for 32 different 1973-74 vehicles. The dataset is reffered to as mtcars can be found at http://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html

The analysis led to the following conclusions and questions.

- Manual transmissions have better MPG than Automatic transmission. However, the ability to predict the MPG of a vehicle based on its transmission type is severly limited and not the most accurate predictive variable.
- Manual transmissions have an approximately 4-7 MPG advantage over automatic transmissions. Again however, the ability to predict a vehicles MPG based on trasmission type is severly limited. More useful predictive variables are weight, quarter mile time and number of carborators.

Data preprocessing

The first step in the analysis of the data is loading it into R from its built in library.

```
data(mtcars)
```

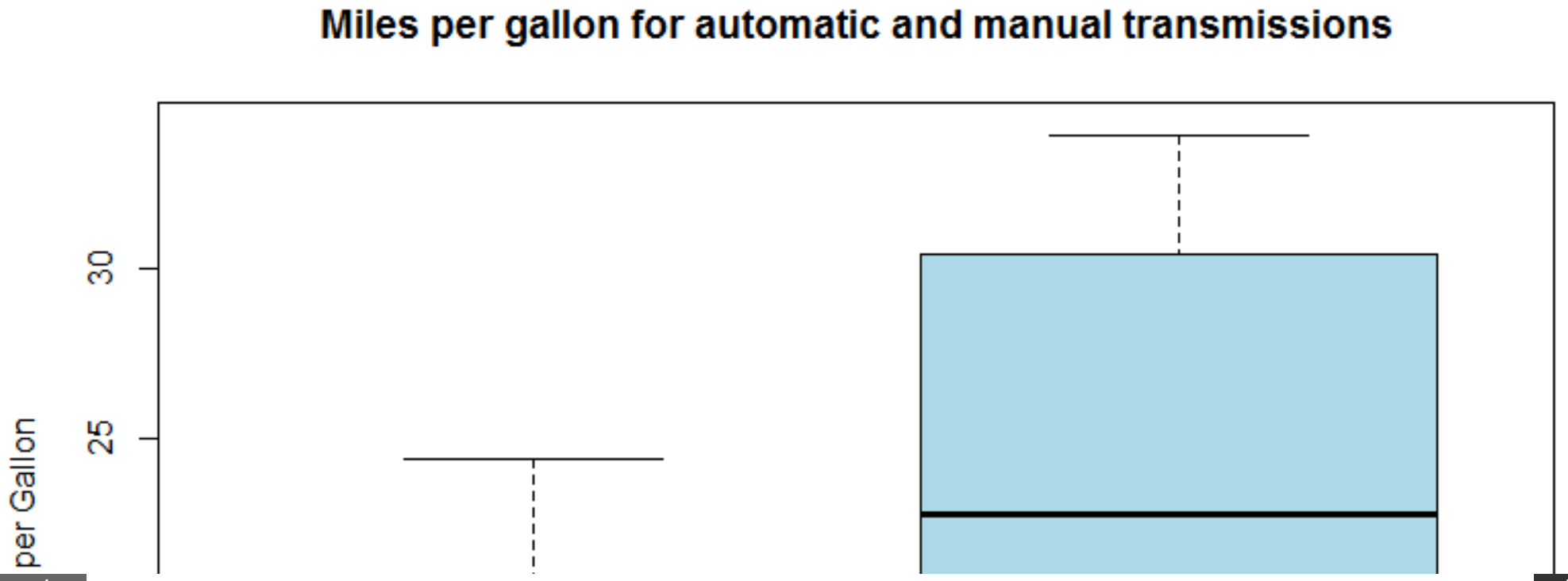
And processing the data so as to allow it to be used by relevant functions found later within the analysis.

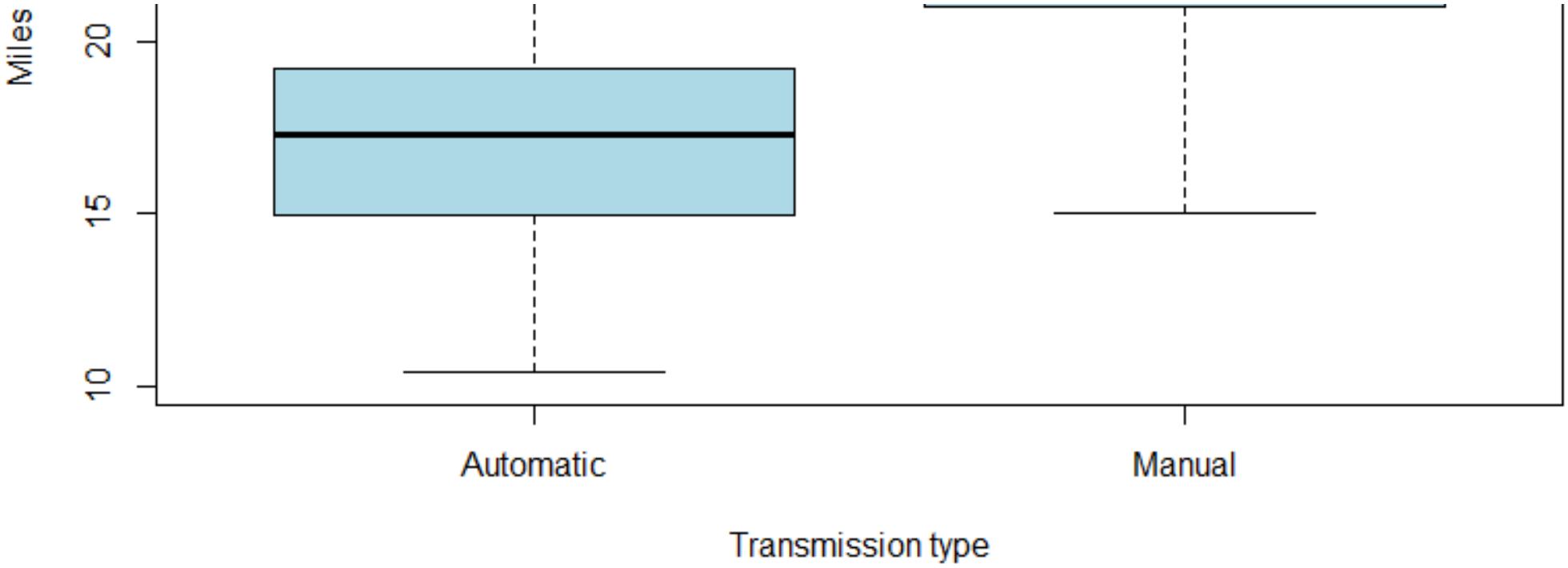
```
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")
```

Exploratory data analysis

An exploratory data analysis of categorical data with real numbered values is best completed using a box plot. As such, one was produced.

```
boxplot(mpg~am, data = mtcars, col = c(" light blue", "light blue"), xlab = "Transmission type", ylab = "Mi
les per Gallon", main = "Miles per gallon for automatic and manual transmissions")
```





The exploratory data analysis using the boxplot shows that the mean value for manual transmissions is higher than the mean value for automatic transmissions. The boxplot also shows that the range of values for automatic and manual transmissions overlaps.

Selecting a regression model

An initial model of mpg against each of the 11 available variables provided can be produced using the following code.

```
lm(mpg ~ factor(am)*.,data=mtcars)
```

The summary statistics of this model are available in the appendix under the section ‘Model 1 summary statistics’. This model produces a model with an R-squared error of .9625 and as such explains 96% of variance between the vehicles. Additional models can and should be produced however, in order to investigate which variables are most signficant.

```
lm(mpg ~ am + wt + qsec + carb, data=mtcars)
```

The summary statistics of this model are available in the appendix under the section ‘Model 2 summary statistics’. As seen in this model by selecting only the distinction between automatic and manual, the weight of the vehicle, 1/4th mile time and the number of carburetors the r-squared error is reduced to .8568, however the number of variables used is significantly reduced.

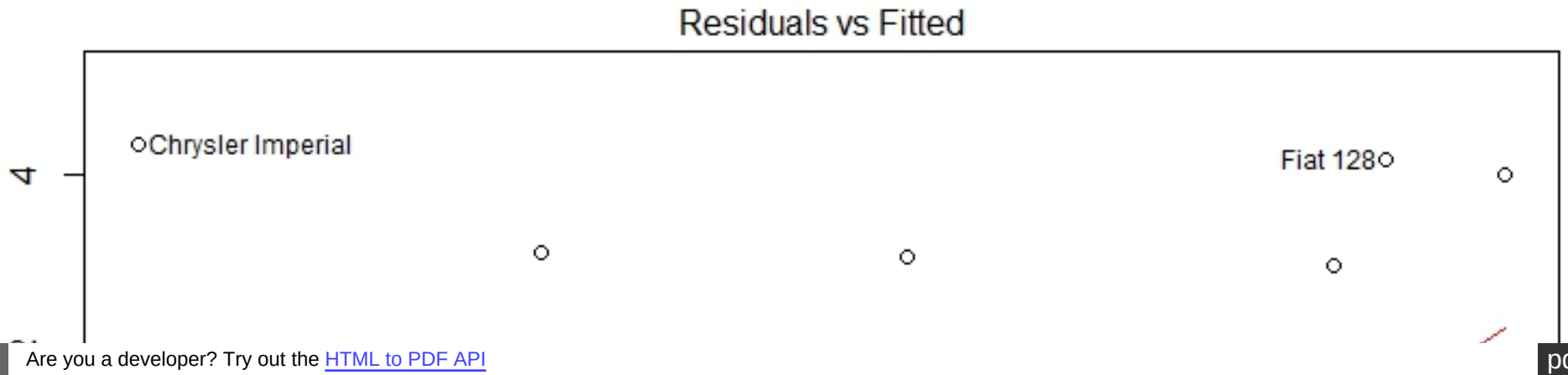
A model only using the distinction between automatic and manual transmissions can also be produced.

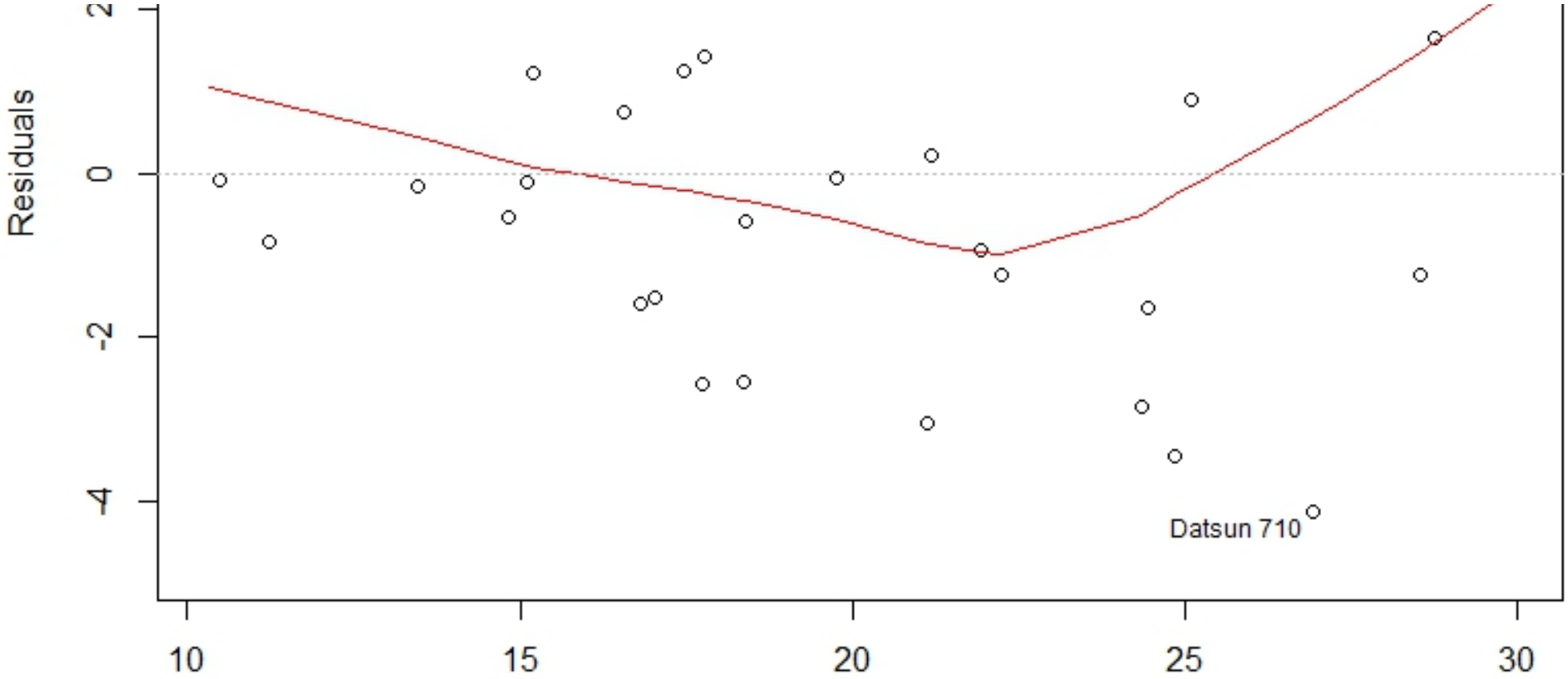
```
lm(mpg ~ am, data=mtcars)
```

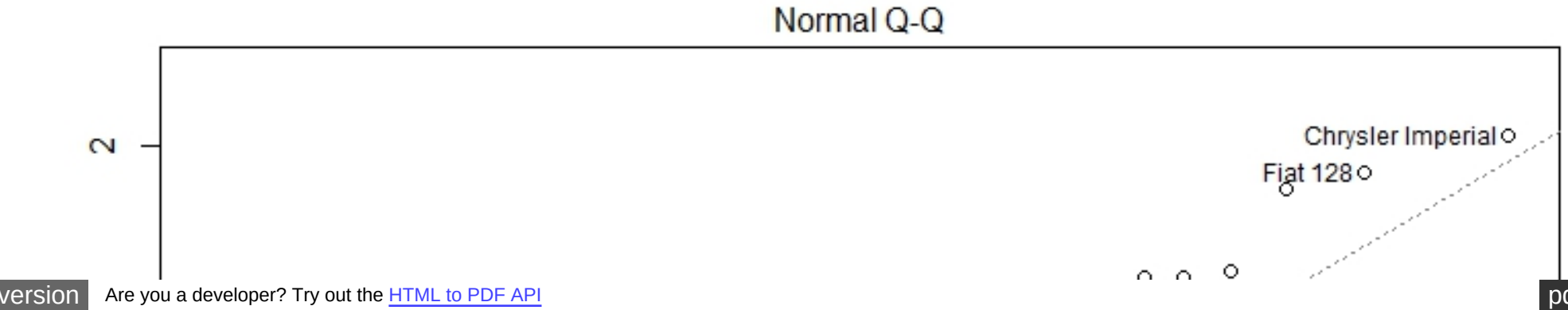
However as the summary in the appendix entiteld ‘Model 3 summary statistics’ bellow shows, this model has a signficantly lower r-squared error of .3598, and as such does not have as much predictive power as do the two other models considered.

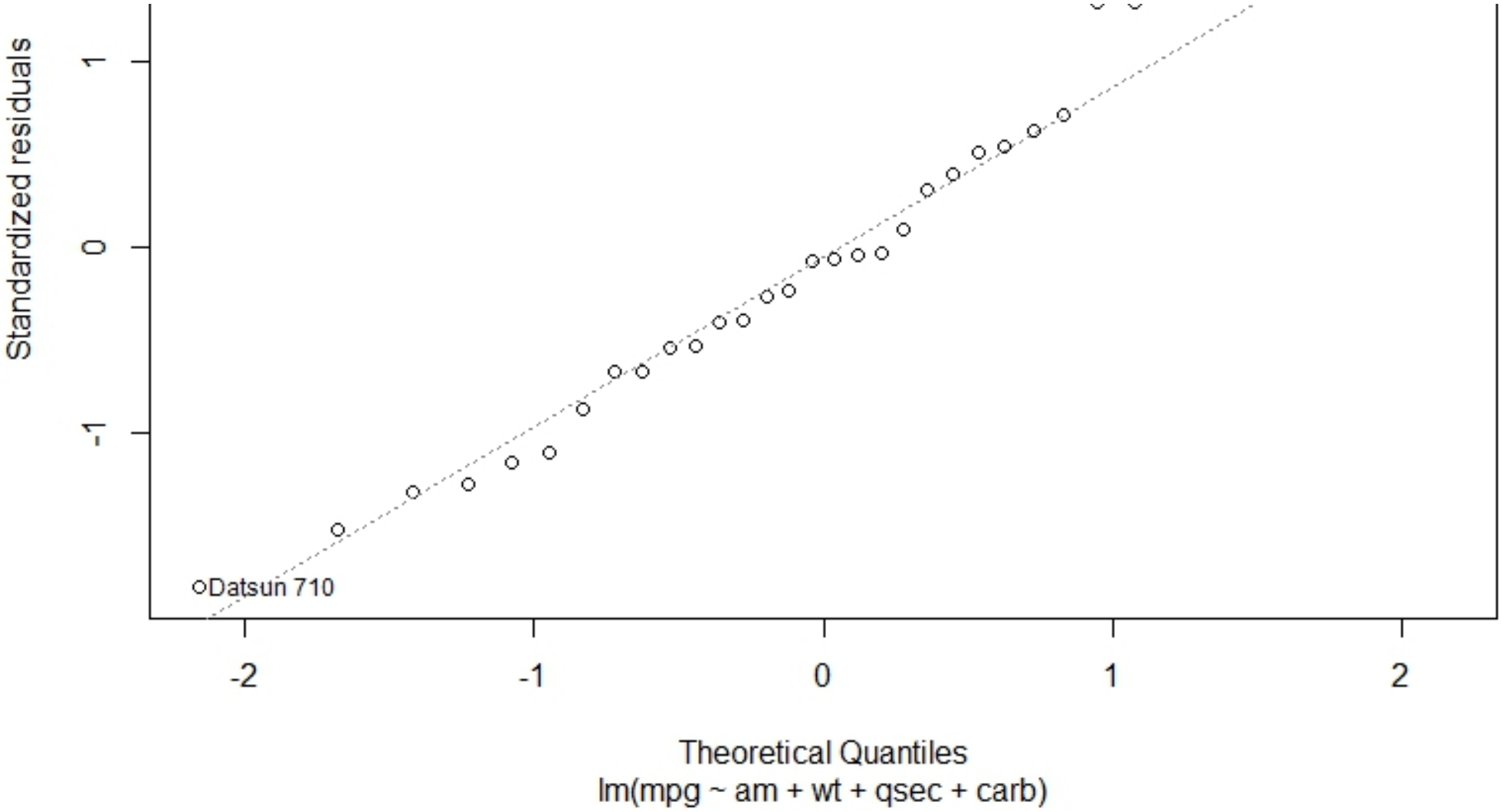
Therefore it is fairly evident that the second model produced is the most useful one.

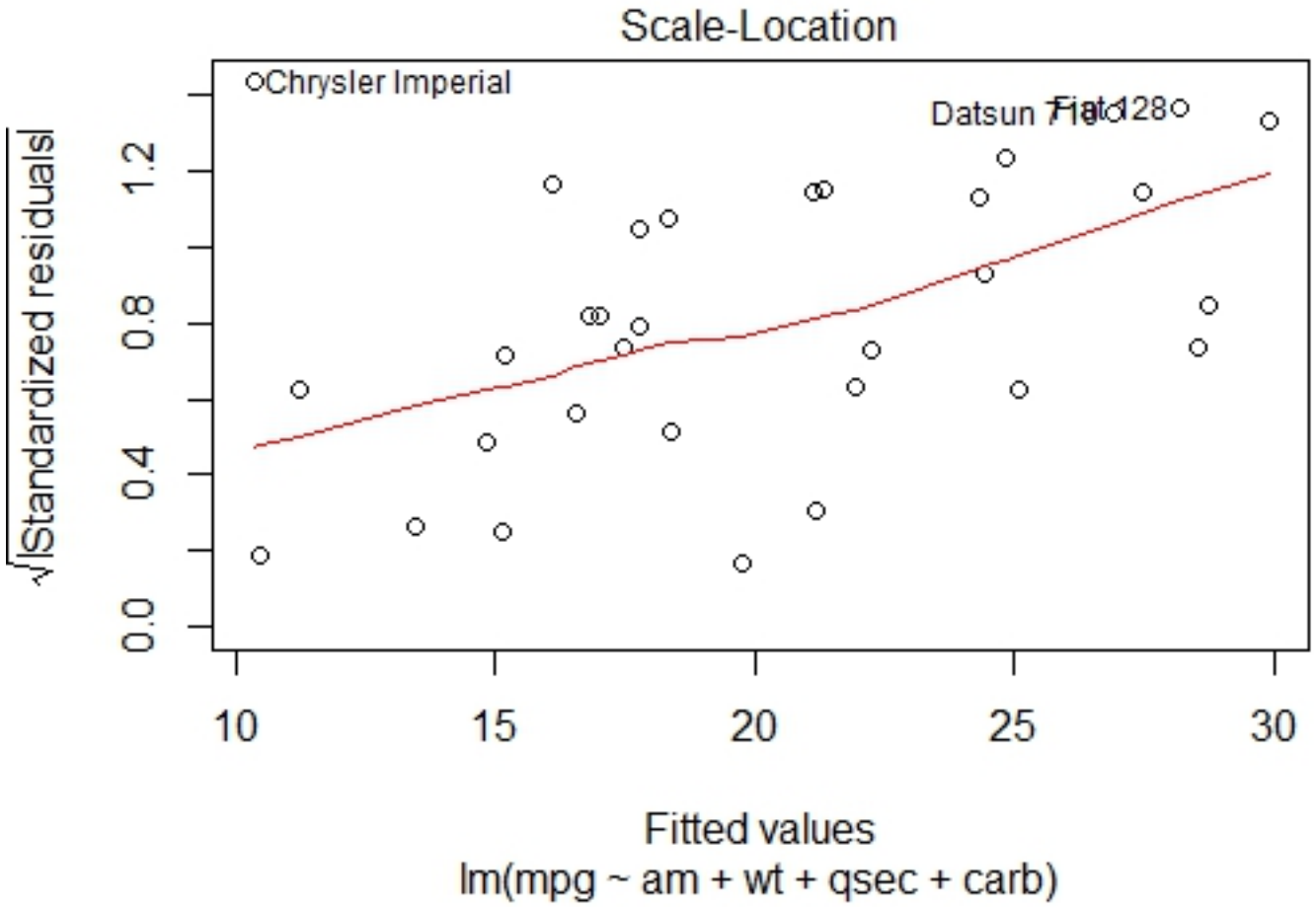
Residual plotting and diagnostic











As seen in the residual plots above, the lines of best fit are sufficiently well fitted upon visual inspection when using a model which includes MPG relative to weight, quarter mile time and number of carborators.

Conclusion

It is evident that in addition to distinction between automatic and manual transmission, the number of cylinders in the vehicle, the quarter mile time and the weight fo the car are all significant in defining the MPG of the vehicle being investigated. Transmission is useful in defining MPG, however, the aformentioned variables have more predictive power in

relation to the overall MPG than transmission does.

Appendix

- Model 1 summary statistics

```
summary(lm(mpg ~ factor(am)*.,data=mtcars))

lm(formula = mpg ~ factor(am) * ., data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-2.0346 -0.7600  0.1089  0.5484  2.6959

Coefficients: (2 not defined because of singularities)
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)              8.64345   22.37276   0.386   0.7060
factor(am)Manual      -146.55089   66.32350  -2.210   0.0473 *
cyl                     -0.53391    1.17256  -0.455   0.6570
disp                    -0.02025    0.01813  -1.117   0.2859
hp                       0.06223    0.04791   1.299   0.2184
drat                     0.59159    3.13258   0.189   0.8534
wt                       1.95413    2.32068   0.842   0.4162
qsec                    -0.88432    0.78877  -1.121   0.2842
vs                       0.73891    2.61246   0.283   0.7821
amManual                      NA         NA      NA       NA
gear                     8.65416    4.05167   2.136   0.0540 .
carb                    -4.81050    1.97648  -2.434   0.0315 *
```

```
factor(am)Manual:cyl        -0.74737    4.26142  -0.175   0.8637
factor(am)Manual:disp        0.20017    0.15960   1.254   0.2337
factor(am)Manual:hp         -0.22268    0.13808  -1.613   0.1328
factor(am)Manual:drat       -5.54142    5.84742  -0.948   0.3620
factor(am)Manual:wt        -12.49602    5.07276  -2.463   0.0299 *
factor(am)Manual:qsec        8.97928    3.21468   2.793   0.0162 *
factor(am)Manual:vs          0.20419    5.28538   0.039   0.9698
factor(am)Manual:amManual         NA         NA      NA       NA
factor(am)Manual:gear        3.67430    7.25129   0.507   0.6215
factor(am)Manual:carb        9.49905    4.16833   2.279   0.0418 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.877 on 12 degrees of freedom
Multiple R-squared:  0.9625,    Adjusted R-squared:  0.9031
F-statistic:  16.2 on 19 and 12 DF,  p-value: 8.251e-06
```

- Model 2 summary statistics

```
summary(lm(mpg ~ am + wt + qsec + carb, data=mtcars))



Residuals:
    Min      1Q  Median      3Q     Max
-4.1184 -1.5414 -0.1392  1.2917  4.3604

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)  12.8972     7.4725   1.726 0.095784 .
amManual      3.5114     1.4875   2.361 0.025721 *
wt           -3.4343     0.8200  -4.188 0.000269 ***
qsec          1.0191     0.3378   3.017 0.005507 **
carb         -0.4886     0.4212  -1.160 0.256212
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.444 on 27 degrees of freedom
Multiple R-squared:  0.8568,    Adjusted R-squared:  0.8356
F-statistic: 40.39 on 4 and 27 DF,  p-value: 5.064e-11
```

- Model 3 summary statistics

```
summary( lm(mpg ~ am, data=mtcars))

Residuals:
    Min      1Q  Median      3Q     Max
-9.3923 -3.0923 -0.2974  3.2439  9.5077

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   17.147      1.125  15.247 1.13e-15 ***
amManual       7.245      1.764   4.106 0.000285 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 4.902 on 30 degrees of freedom
```

```
Multiple R-squared:  0.3598,    Adjusted R-squared:  0.3385
F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```