

# Report

2023-12-23

## Executive summary

In this report I aim to answer the questions

“Is an automatic or manual transmission better for MPG”

“Quantify the MPG difference between automatic and manual transmissions”

My findings are that the differences between manual/automatic transmission are better explained through other variables.

## Dataset

```
data(mtcars)
head(mtcars)
```

```
##           mpg cyl  disp  hp  drat    wt  qsec vs am gear carb
## Mazda RX4      21.0   6  160 110  3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag  21.0   6  160 110  3.90 2.875 17.02  0  1    4    4
## Datsun 710      22.8   4  108  93  3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive  21.4   6  258 110  3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175  3.15 3.440 17.02  0  0    3    2
## Valiant         18.1   6  225 105  2.76 3.460 20.22  1  0    3    1
```

## Dataset description

*Description sourced from the R documentation*

The 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).

Var	Desc
mpg	Miles/(US) gallon
cyl	Number of cylinders
disp	Displacement (cu.in.)
hp	Gross horsepower
drat	Rear axle ratio
wt	Weight (1000 lbs)
qsec	1/4 mile time
vs	Engine (0 = V-shaped, 1 = straight)
am	Transmission (0 = automatic, 1 = manual)
gear	Number of forward gears

Var	Desc
carb	Number of carburetors

## Initial exploration

vs and am are categorical variables so I will convert the columns from numeric to factor. I will also do the same for gear, cyl, and carb. Though it could be argued that there are numeric variables, in this case I will treat them as categorical.

```
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$gear <- as.factor(mtcars$gear)
mtcars$carb <- as.factor(mtcars$carb)
mtcars$vs <- factor(mtcars$vs, labels=c("V-shaped", "Straight"))
mtcars$am <- factor(mtcars$am, labels=c("Automatic", "Manual"))
```

I will also rename the columns to more readable names

```
names(mtcars) <- c("mpg", "cylinders", "displacement", "horsepower", "rear_axle_ratio", "weight", "quarter_mile_time", "engine", "transmission", "gears", "carburetors")
```

```
str(mtcars)
```

```
## 'data.frame': 32 obs. of 11 variables:
## $ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
## $ cylinders : Factor w/ 3 levels "4","6","8": 2 2 1 2 3 2 3 1 1 2 ...
## $ displacement : num 160 160 108 258 360 ...
## $ horsepower : num 110 110 93 110 175 105 245 62 95 123 ...
## $ rear_axle_ratio : num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
## $ weight : num 2.62 2.88 2.32 3.21 3.44 ...
## $ quarter_mile_time: num 16.5 17 18.6 19.4 17 ...
## $ engine : Factor w/ 2 levels "V-shaped","Straight": 1 1 2 2 1 2 1 2 2 2 ...
## $ transmission : Factor w/ 2 levels "Automatic","Manual": 2 2 2 1 1 1 1 1 1 1 ...
## $ gears : Factor w/ 3 levels "3","4","5": 2 2 2 1 1 1 1 2 2 2 ...
## $ carburetors : Factor w/ 6 levels "1","2","3","4",...: 4 4 1 1 2 1 4 2 2 4 ...
```

```
summary(mtcars)
```

```
##      mpg      cylinders displacement      horsepower      rear_axle_ratio
## Min.   :10.40    4:11      Min.   : 71.1    Min.   : 52.0    Min.   :2.760
## 1st Qu.:15.43    6: 7      1st Qu.:120.8    1st Qu.: 96.5    1st Qu.:3.080
## Median :19.20    8:14      Median :196.3    Median :123.0    Median :3.695
## Mean   :20.09                Mean   :230.7    Mean   :146.7    Mean   :3.597
## 3rd Qu.:22.80                3rd Qu.:326.0    3rd Qu.:180.0    3rd Qu.:3.920
## Max.   :33.90                Max.   :472.0    Max.   :335.0    Max.   :4.930
##      weight      quarter_mile_time      engine      transmission      gears
## Min.   :1.513    Min.   :14.50      V-shaped:18    Automatic:19    3:15
## 1st Qu.:2.581    1st Qu.:16.89      Straight:14    Manual   :13    4:12
## Median :3.325    Median :17.71                5: 5
## Mean   :3.217    Mean   :17.85
## 3rd Qu.:3.610    3rd Qu.:18.90
## Max.   :5.424    Max.   :22.90
```

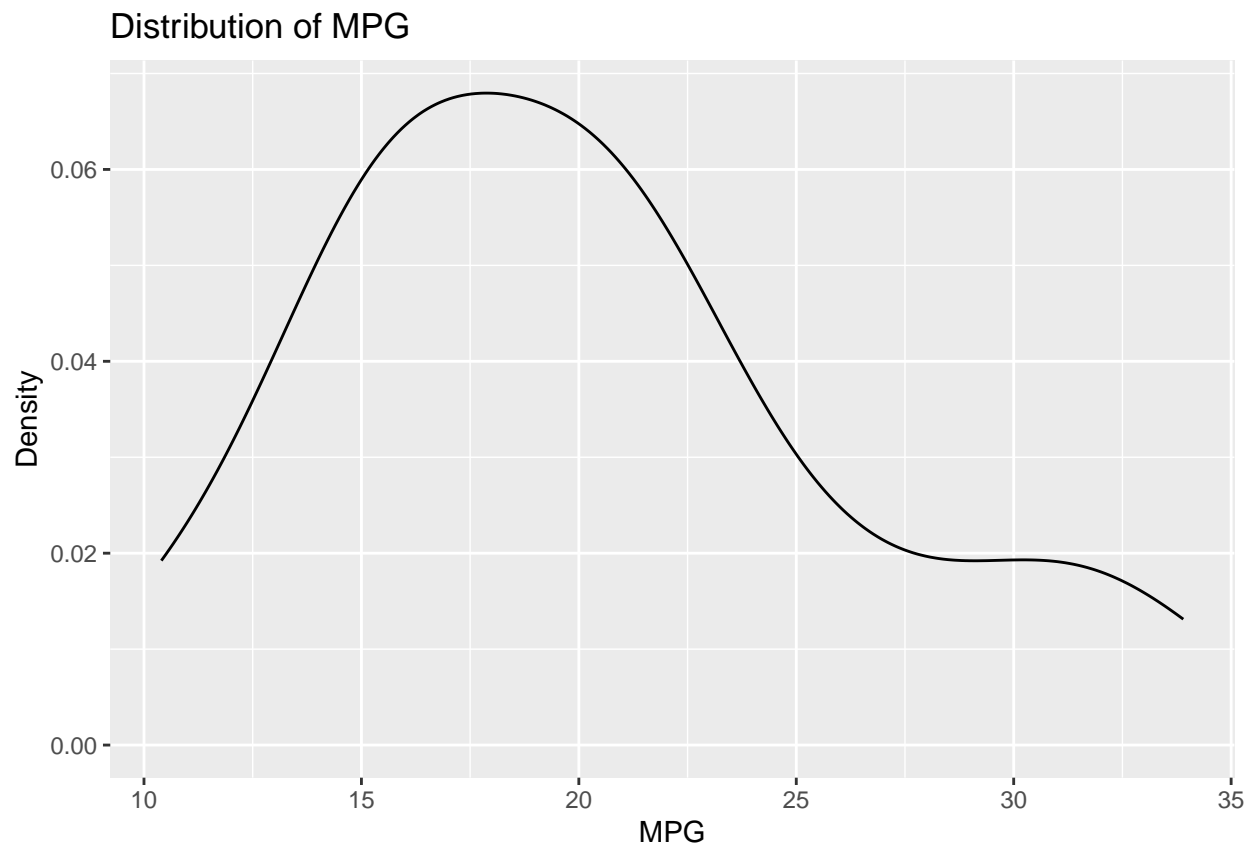
```
## carburetors
## 1: 7
## 2:10
## 3: 3
## 4:10
## 6: 1
## 8: 1
```

No NA values are present, so no imputation needed.

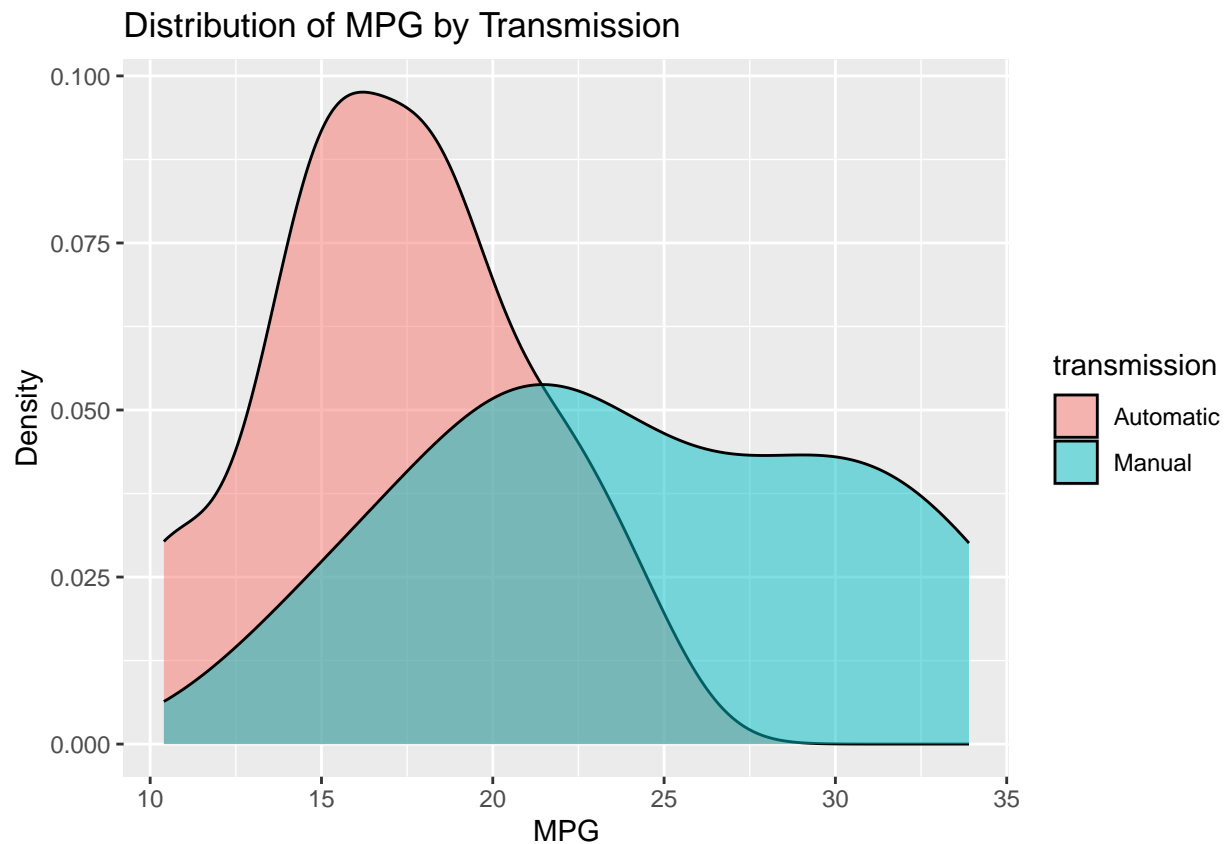
```
sapply(mtcars, function(x) sum(is.na(x)))
```

```
##          mpg          cylinders      displacement      horsepower
##           0              0              0              0
## rear_axle_ratio      weight quarter_mile_time      engine
##           0              0              0              0
##      transmission      gears      carburetors
##           0              0              0
```

```
ggplot(mtcars, aes(x = mpg)) +
  geom_density() +
  labs(title = "Distribution of MPG", x = "MPG", y = "Density")
```



```
ggplot(mtcars, aes(x = mpg, fill = transmission)) +
  geom_density(alpha = 0.5) +
  labs(title = "Distribution of MPG by Transmission", x = "MPG", y = "Density")
```



It appears that automatic cars in general have a small range of MPG values, centred at a lower MPG than that of manual transmission cars, though manual transmission also have a larger range of values. We can verify this using a T-test, using a confidence level of 95% and a null hypothesis that the mean MPG of automatic and manual transmission cars are the same.

```
t.test(mpg ~ transmission, data = mtcars, conf.level = 0.95, alternative = "less")
```

```
##
## Welch Two Sample t-test
##
## data:  mpg by transmission
## t = -3.7671, df = 18.332, p-value = 0.0006868
## alternative hypothesis: true difference in means between group Automatic and group Manual is less than 0
## 95 percent confidence interval:
##      -Inf -3.913256
## sample estimates:
## mean in group Automatic      mean in group Manual
##           17.14737              24.39231
```

We have a p-value of 0.0006 thus we can reject the null hypothesis that the mean MPG of automatic and manual transmission cars are the same, and accept the alternative hypothesis that the mean MPG of automatic transmission cars is less than that of manual transmission cars.

However this may be explaining relationships between other variables and mpg, assuming they are linked to transmission.

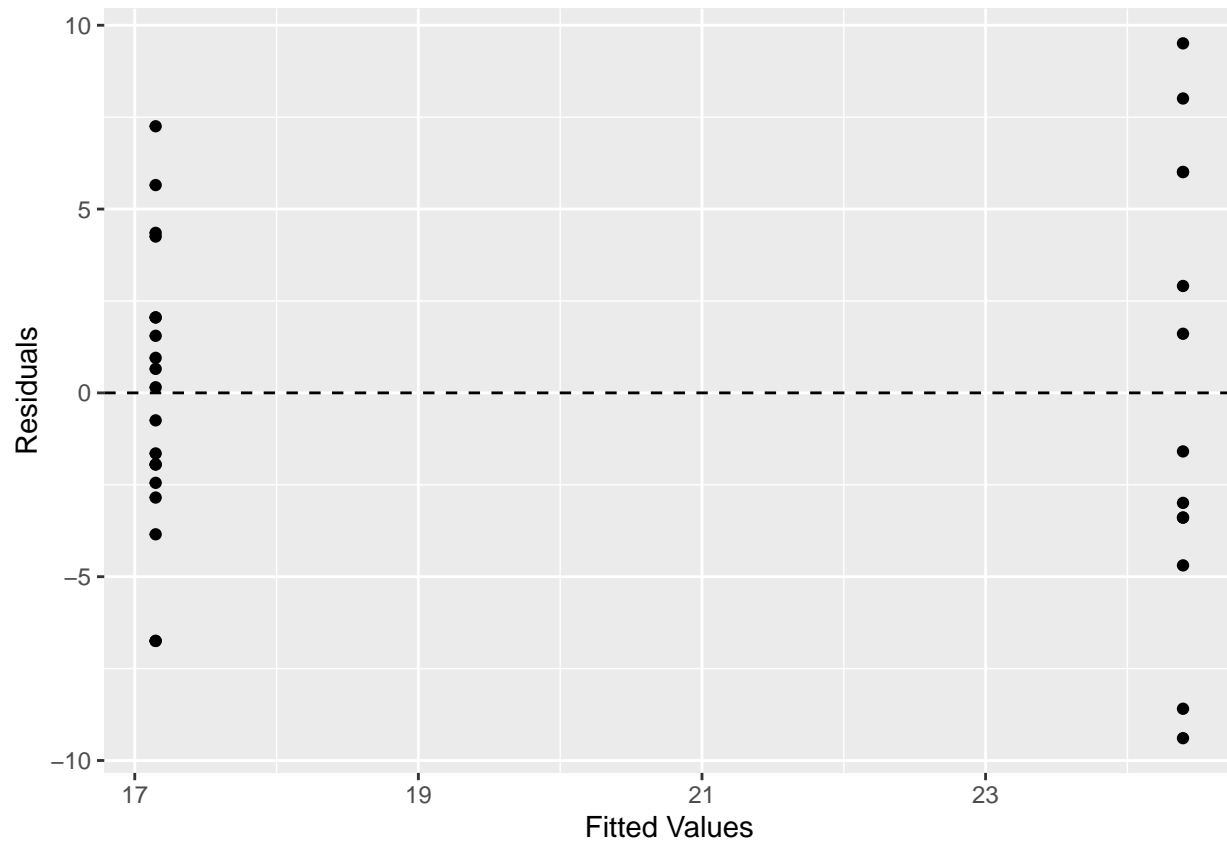
## Regression

```
model <- lm(mpg ~ transmission, data = mtcars)
summary(model)

##
## Call:
## lm(formula = mpg ~ transmission, 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 ***
## transmissionManual    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
```

By fitting a linear model from just the transmission to mpg it seems that moving from a transmission of 0 (automatic) to 1 (manual) increases the mpg by 7.245. However this is not a very good model as the R-squared value is only 0.3598, meaning that only 35.98% of the variance in mpg is explained by the transmission.

```
ggplot(model, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  xlab("Fitted Values") +
  ylab("Residuals")
```



```
model <- lm(mpg ~ ., data=mtcars)
summary(model)
```

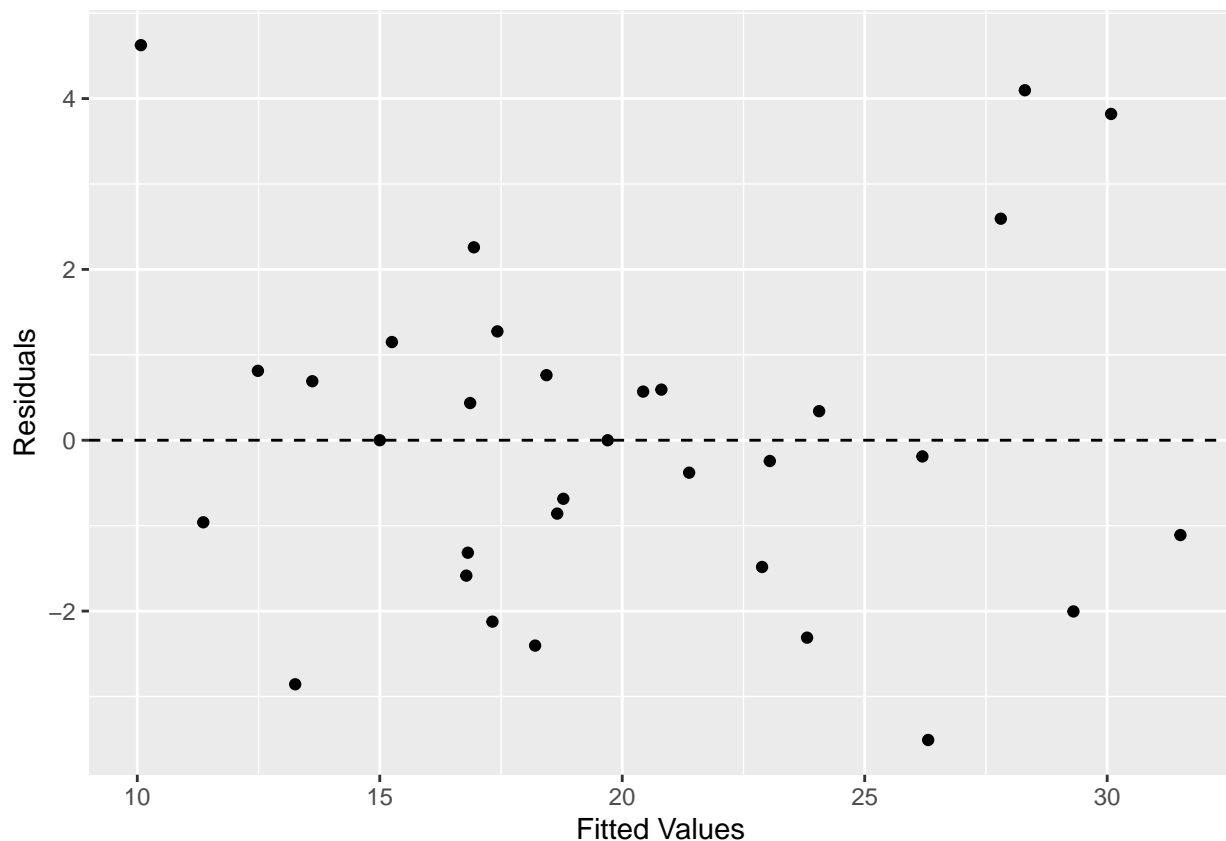
```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.5087 -1.3584 -0.0948  0.7745  4.6251
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    23.87913    20.06582   1.190   0.2525
## cylinders6     -2.64870     3.04089  -0.871   0.3975
## cylinders8     -0.33616     7.15954  -0.047   0.9632
## displacement   0.03555     0.03190   1.114   0.2827
## horsepower    -0.07051     0.03943  -1.788   0.0939 .
## rear_axle_ratio  1.18283     2.48348   0.476   0.6407
## weight        -4.52978     2.53875  -1.784   0.0946 .
## quarter_mile_time 0.36784     0.93540   0.393   0.6997
## engineStraight   1.93085     2.87126   0.672   0.5115
## transmissionManual 1.21212     3.21355   0.377   0.7113
## gears4          1.11435     3.79952   0.293   0.7733
## gears5          2.52840     3.73636   0.677   0.5089
```

```
## carburetors2      -0.97935    2.31797   -0.423    0.6787
## carburetors3       2.99964    4.29355    0.699    0.4955
## carburetors4       1.09142    4.44962    0.245    0.8096
## carburetors6       4.47757    6.38406    0.701    0.4938
## carburetors8       7.25041    8.36057    0.867    0.3995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.833 on 15 degrees of freedom
## Multiple R-squared:  0.8931, Adjusted R-squared:  0.779
## F-statistic:  7.83 on 16 and 15 DF,  p-value: 0.000124
```

When including all the variables we can see that the changing the transmission only increases the mpg by 1.212, suggesting that most of the change in mpg is due to other factors than just manual/automatic. The high p value of 0.7 suggests that we cannot attribute the change in mpg to transmission.

In particular horsepower and weight seem to contribute the most to capturing the variance of mpg. This model has an R-squared of 89%, and it is clear from the residual plot that it is a much better predictor of mpg.

```
ggplot(model, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  xlab("Fitted Values") +
  ylab("Residuals")
```



## Conclusion

In conclusion: manual transmission is better for MPG than automatic, switching to manual will increase MPG by approximately 1.212, but I have shown that this transmission is not a statistically significant contributor to change in mpg when taking other variables into account.