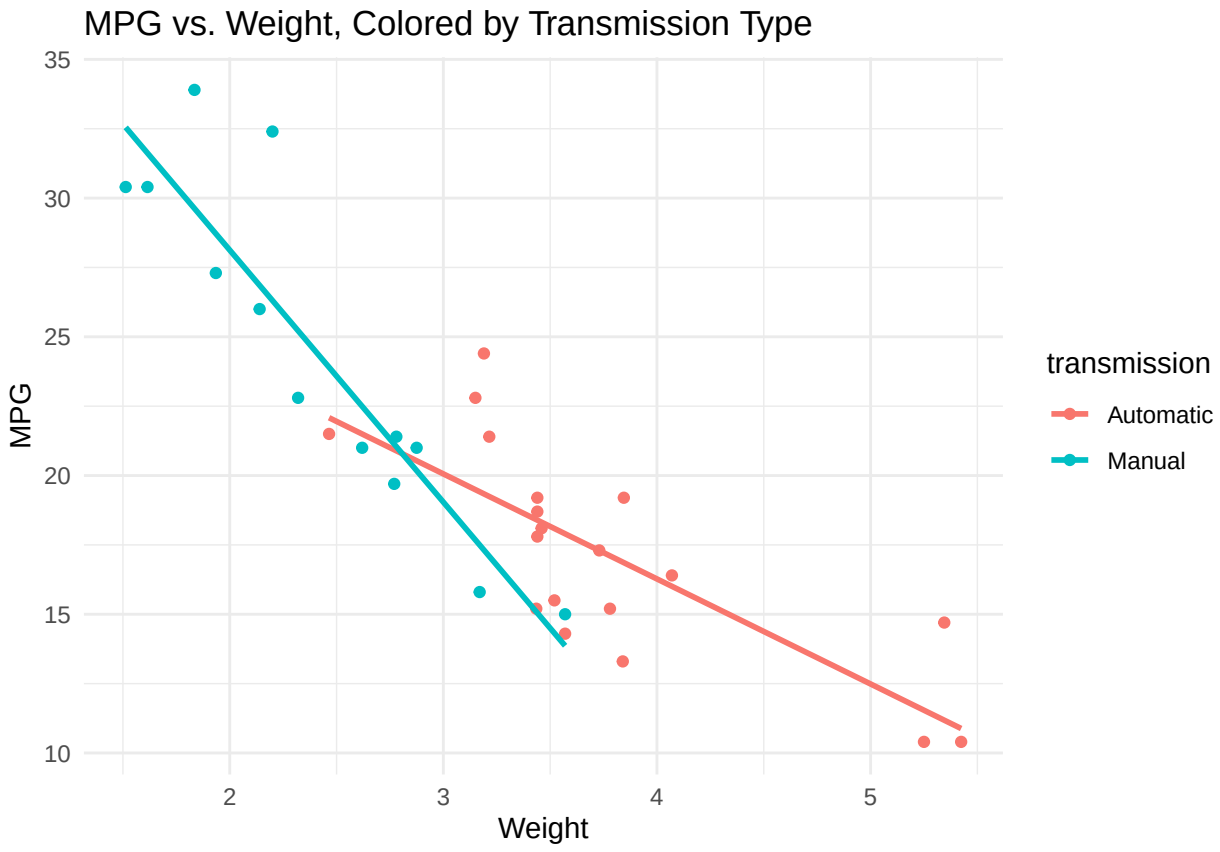
Experimental Design

Johnny Ferrara

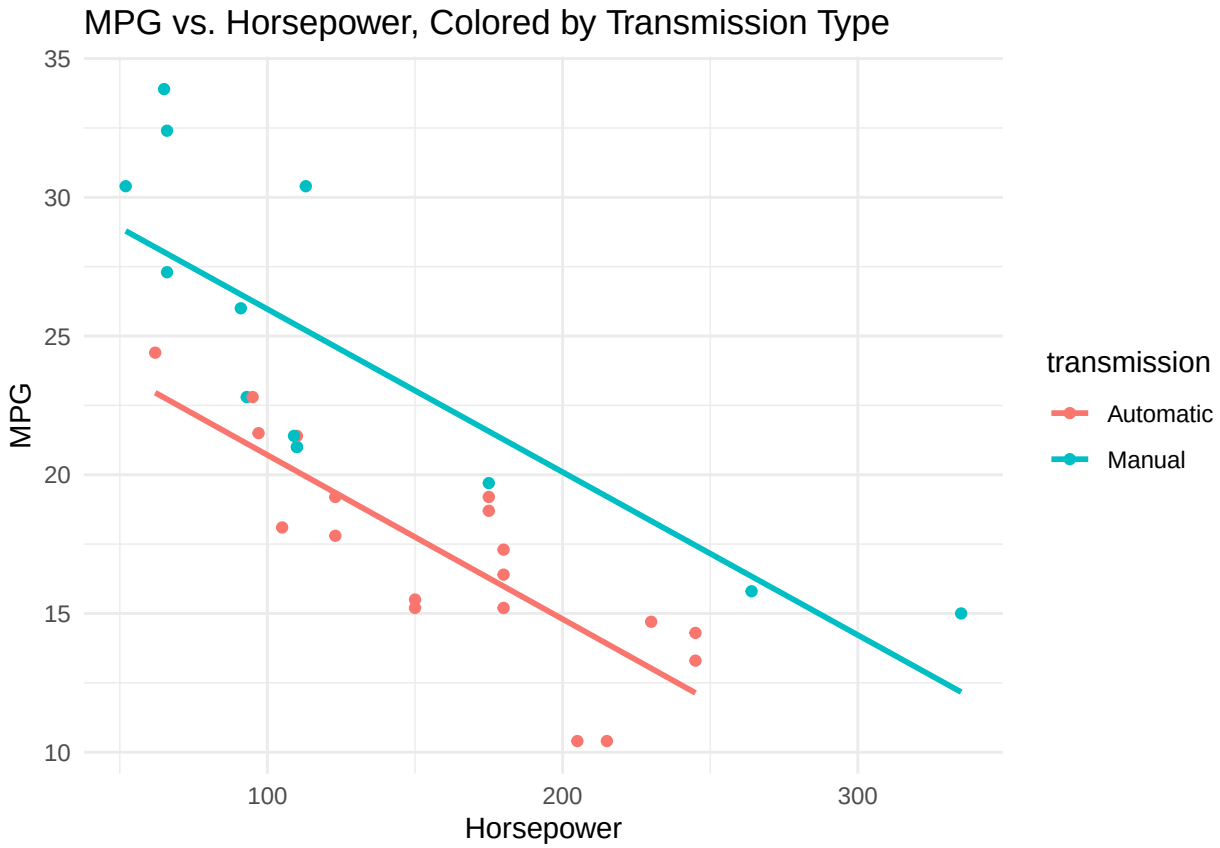
```r
# Load the mtcars dataset and ggplot2 library
data(mtcars)
library(ggplot2)
library(MASS)

# Convert 'am' to a more descriptive factor variable for transmission type
mtcars$transmission <- factor(mtcars$am, levels = c(0, 1), labels = c("Automatic", "Manual"))


# Scatterplot of mpg against weight
ggplot(mtcars, aes(x = wt, y = mpg, color = transmission)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  theme_minimal() +
  labs(title = "MPG vs. Weight, Colored by Transmission Type", x = "Weight", y = "MPG")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



```r
# Scatterplot of mpg against horsepower
ggplot(mtcars, aes(x = hp, y = mpg, color = transmission)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  theme_minimal() +
  labs(title = "MPG vs. Horsepower, Colored by Transmission Type", x = "Horsepower", y = "MPG")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```r
# Complex model with interactions
complex_model <- lm(mpg ~ hp * wt * transmission, data = mtcars)

simple_model <- lm(mpg ~ hp + wt + transmission, data = mtcars)


# Compare the simple and complex models
anova(simple_model)
```

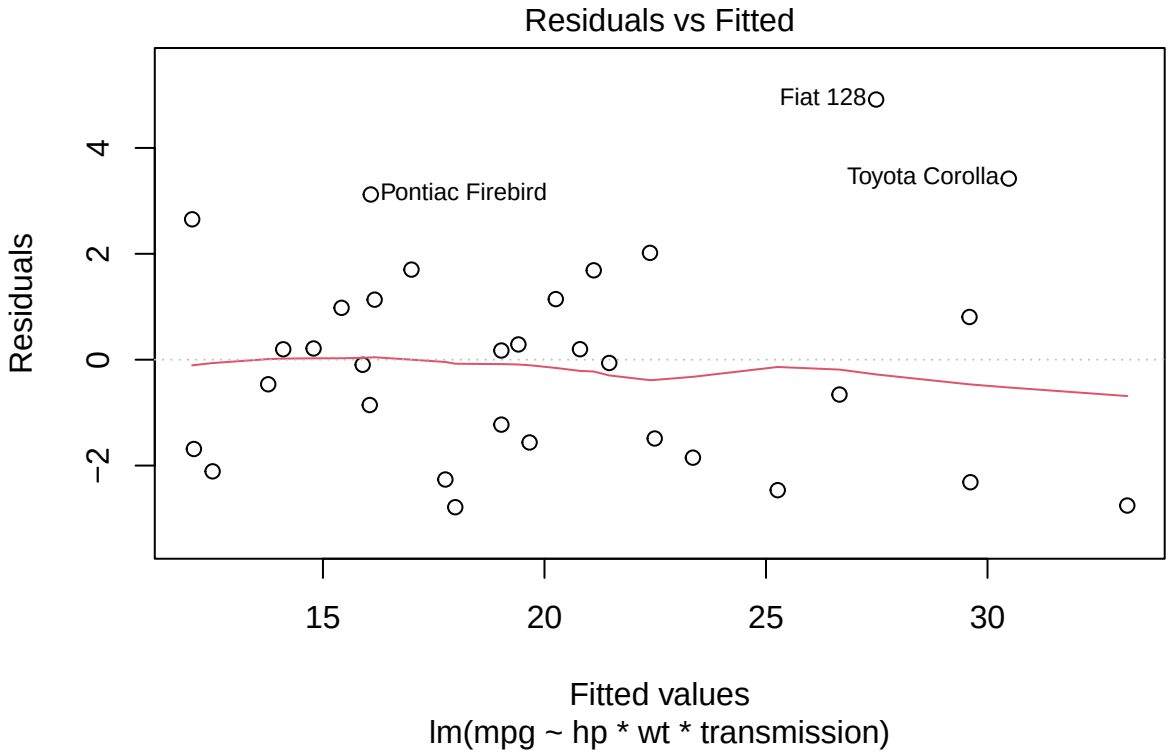
```
## Analysis of Variance Table
##
## Response: mpg
##              Df Sum Sq Mean Sq  F value    Pr(>F)
## hp            1 678.37  678.37 105.3543 5.395e-11 ***
## wt            1 252.63  252.63  39.2340 9.028e-07 ***
## transmission  1  14.76   14.76   2.2918    0.1413
## Residuals    28 180.29    6.44
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

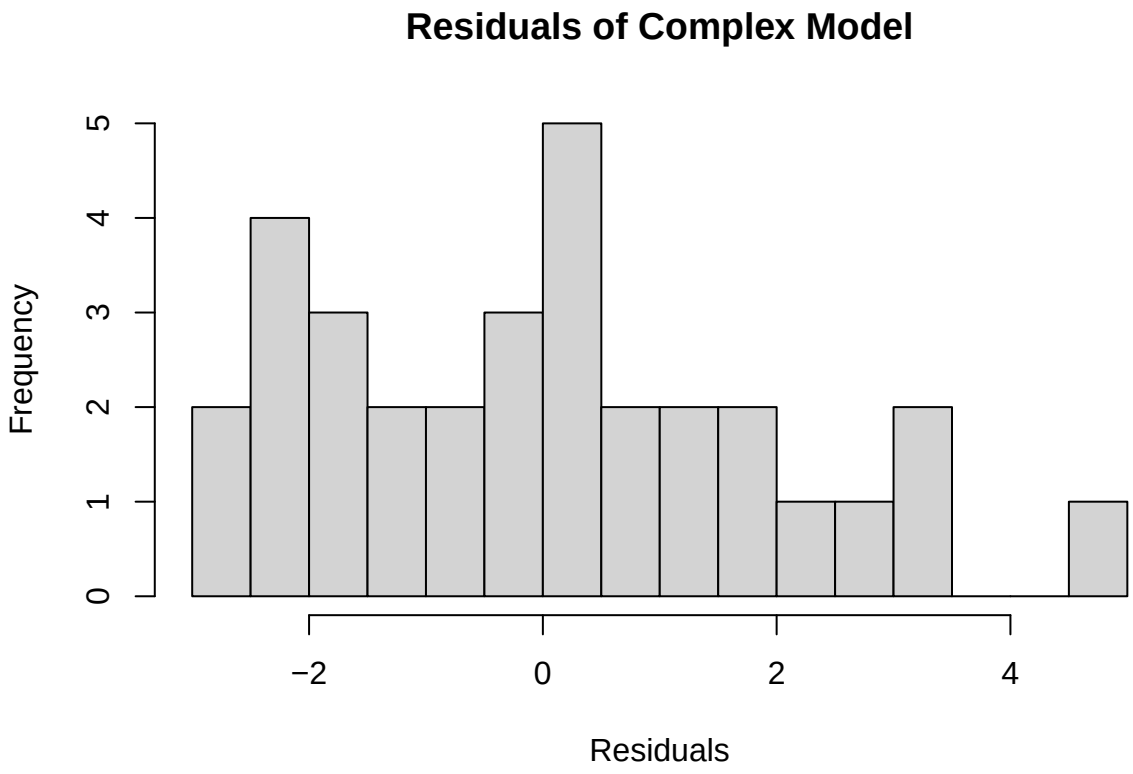
```
anova(complex_model)
```

```
## Analysis of Variance Table
##
## Response: mpg
##                    Df Sum Sq Mean Sq  F value    Pr(>F)
## hp                  1 678.37  678.37 139.0794 1.790e-11 ***
## wt                  1 252.63  252.63  51.7933 1.948e-07 ***
## transmission        1  14.76   14.76   3.0254  0.094782 .
## hp:wt               1  50.57   50.57  10.3682  0.003659 **
## hp:transmission     1   4.76    4.76   0.9754  0.333193
## wt:transmission     1   5.64    5.64   1.1570  0.292791
## hp:wt:transmission  1   2.26    2.26   0.4626  0.502939
## Residuals          24 117.06    4.88
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

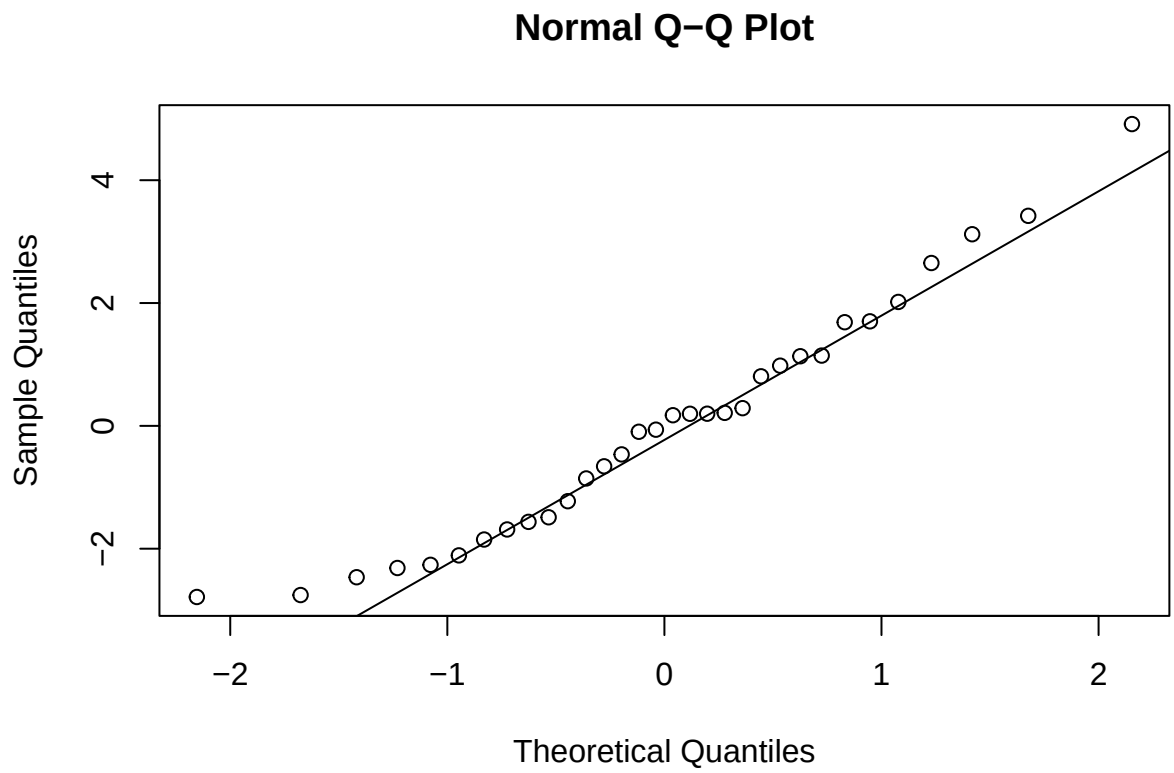
```
# Plotting residuals of the complex model
plot(complex_model, which = 1)
```




```
# Normality check of residuals
hist(residuals(complex_model), breaks = 20, main = "Residuals of Complex Model", xlab = "Residuals")
```




```r
qqnorm(residuals(complex_model))
qqline(residuals(complex_model))
```




```r
# Summary of the complex model
summary(complex_model)
```

```
##
## Call:
## lm(formula = mpg ~ hp * wt * transmission, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7864 -1.5944  0.0556  1.1364  4.9140
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              40.32715   13.00819   3.100  0.00489 **
## hp                       -0.08875    0.06469  -1.372  0.18276
## wt                       -4.79683    4.00169  -1.199  0.24235
## transmissionManual       12.83705   14.22240   0.903  0.37571
## hp:wt                     0.01446    0.01915   0.755  0.45770
## hp:transmissionManual    -0.03257    0.08894  -0.366  0.71739
## wt:transmissionManual    -5.36196    4.59745  -1.166  0.25496
## hp:wt:transmissionManual  0.01776    0.02612   0.680  0.50294
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.209 on 24 degrees of freedom
## Multiple R-squared:  0.896,  Adjusted R-squared:  0.8657
## F-statistic: 29.55 on 7 and 24 DF,  p-value: 2.605e-10
```

```r
# Backward stepwise regression on the complex model
backward_stepwise <- step(complex_model, direction = "both")
```

```
## Start:  AIC=57.5
## mpg ~ hp * wt * transmission
##
##                       Df Sum of Sq    RSS    AIC
## - hp:wt:transmission  1    2.2562 119.32 56.114
## <none>                             117.06 57.503
##
## Step:  AIC=56.11
## mpg ~ hp + wt + transmission + hp:wt + hp:transmission + wt:transmission
##
##                       Df Sum of Sq    RSS    AIC
## - wt:transmission     1    5.6432 124.96 55.593
## <none>                             119.32 56.114
## - hp:transmission     1    9.9990 129.32 56.689
## + hp:wt:transmission  1    2.2562 117.06 57.503
## - hp:wt               1   16.5816 135.90 58.278
##
## Step:  AIC=55.59
## mpg ~ hp + wt + transmission + hp:wt + hp:transmission
##
##                   Df Sum of Sq    RSS    AIC
## - hp:transmission  1     4.758 129.72 54.788
## <none>                         124.96 55.593
## + wt:transmission  1     5.643 119.32 56.114
## - hp:wt            1    51.457 176.42 64.628
##
## Step:  AIC=54.79
```

```
## mpg ~ hp + wt + transmission + hp:wt
##
##                   Df Sum of Sq    RSS    AIC
## - transmission     1     0.042 129.76 52.799
## <none>                         129.72 54.788
## + hp:transmission  1     4.758 124.96 55.593
## + wt:transmission  1     0.402 129.32 56.689
## - hp:wt            1    50.572 180.29 63.323
##
## Step:  AIC=52.8
## mpg ~ hp + wt + hp:wt
##
##                Df Sum of Sq    RSS    AIC
## <none>                      129.76 52.799
## + transmission  1     0.042 129.72 54.788
## - hp:wt         1    65.286 195.05 63.840
```

```
summary(backward_stepwise)
```

```
##
## Call:
## lm(formula = mpg ~ hp + wt + hp:wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0632 -1.6491 -0.7362  1.4211  4.5513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 49.80842    3.60516  13.816 5.01e-14 ***
## hp          -0.12010    0.02470  -4.863 4.04e-05 ***
## wt          -8.21662    1.26971  -6.471 5.20e-07 ***
## hp:wt        0.02785    0.00742   3.753 0.000811 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.153 on 28 degrees of freedom
## Multiple R-squared:  0.8848, Adjusted R-squared:  0.8724
## F-statistic: 71.66 on 3 and 28 DF,  p-value: 2.981e-13
```

```
both_stepwise <- step(simple_model, direction = "both")
```

```
## Start:  AIC=63.32
## mpg ~ hp + wt + transmission
##
##                Df Sum of Sq    RSS    AIC
## <none>                      180.29 63.323
## - transmission  1    14.757 195.05 63.840
## - wt            1    65.148 245.44 71.194
## - hp            1    98.029 278.32 75.217
```

```
summary(both_stepwise)
```

```
##
## Call:
## lm(formula = mpg ~ hp + wt + transmission, data = mtcars)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4221 -1.7924 -0.3788  1.2249  5.5317
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        34.002875   2.642659  12.867 2.82e-13 ***
## hp                 -0.037479   0.009605  -3.902 0.000546 ***
## wt                 -2.878575   0.904971  -3.181 0.003574 **
## transmissionManual  2.083710   1.376420   1.514 0.141268
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.538 on 28 degrees of freedom
## Multiple R-squared:  0.8399, Adjusted R-squared:  0.8227
## F-statistic: 48.96 on 3 and 28 DF,  p-value: 2.908e-11
```