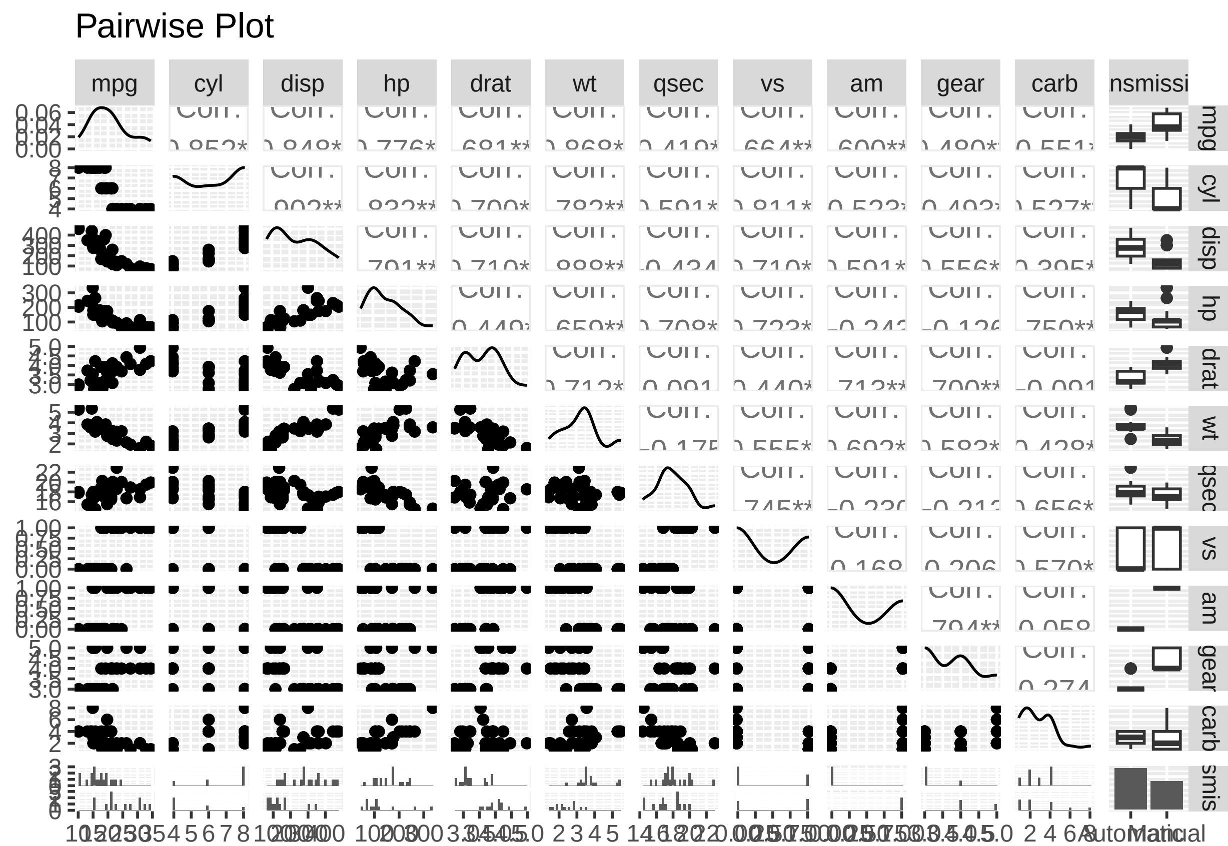
DS Regression Project

Raghda

2024-06-01

Let's load the data and plot the relationship between all variables.



This plot shows a summery of the relationship between all the data

Q1-Is an automatic or manual transmission better for MPG?

Our first model is a simple linear regression of MPG on transmission type.his model indicates a significant difference in MPG between transmission types.

```
# Fit the linear model
model1 <- lm(mpg ~ transmission, data = mtcars)
summary(model1)
```

```
##
## 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
```

```r
head(coef(summary(model1)))
```

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

Interpretation

On average, manual transmission cars have about 7.245 more miles per gallon than automatic transmission car.The Pvalue for transmissionManual is very low (< 0.001), indicating that this difference is statistically significant. R-squared: The model explains about 35.98% of the variance in MPG.So the manual transmissions are better for MPG, and it quantifies the difference as approximately 7.245 MPG higher for manual transmissions.

Q2-Quantify the MPG difference between automatic and manual transmissions

Using multiple Models We expand the models to control for all other variables to understand their effect on mgp.

```r
# mpg with all variables
model2<- lm(data = mtcars, mpg~.)
summary(model2)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients: (1 not defined because of singularities)
```

```
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        12.30337   18.71788   0.657   0.5181
## cyl                -0.11144    1.04502  -0.107   0.9161
## disp                0.01334    0.01786   0.747   0.4635
## hp                 -0.02148    0.02177  -0.987   0.3350
## drat                0.78711    1.63537   0.481   0.6353
## wt                 -3.71530    1.89441  -1.961   0.0633 .
## qsec                0.82104    0.73084   1.123   0.2739
## vs                  0.31776    2.10451   0.151   0.8814
## am                  2.52023    2.05665   1.225   0.2340
## gear                0.65541    1.49326   0.439   0.6652
## carb               -0.19942    0.82875  -0.241   0.8122
## transmissionManual       NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

```r
#Quantifying the MPG Difference
head(coef(summary(model2)))
```

```
##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
```

If a p-value is less than 0.05 we can conclude that the corresponding predictor is statistically significant.Larger coefficients suggest stronger associations between predictors and the response variable.Smaller standard errors indicate more precise estimates.

Choosing the best fit

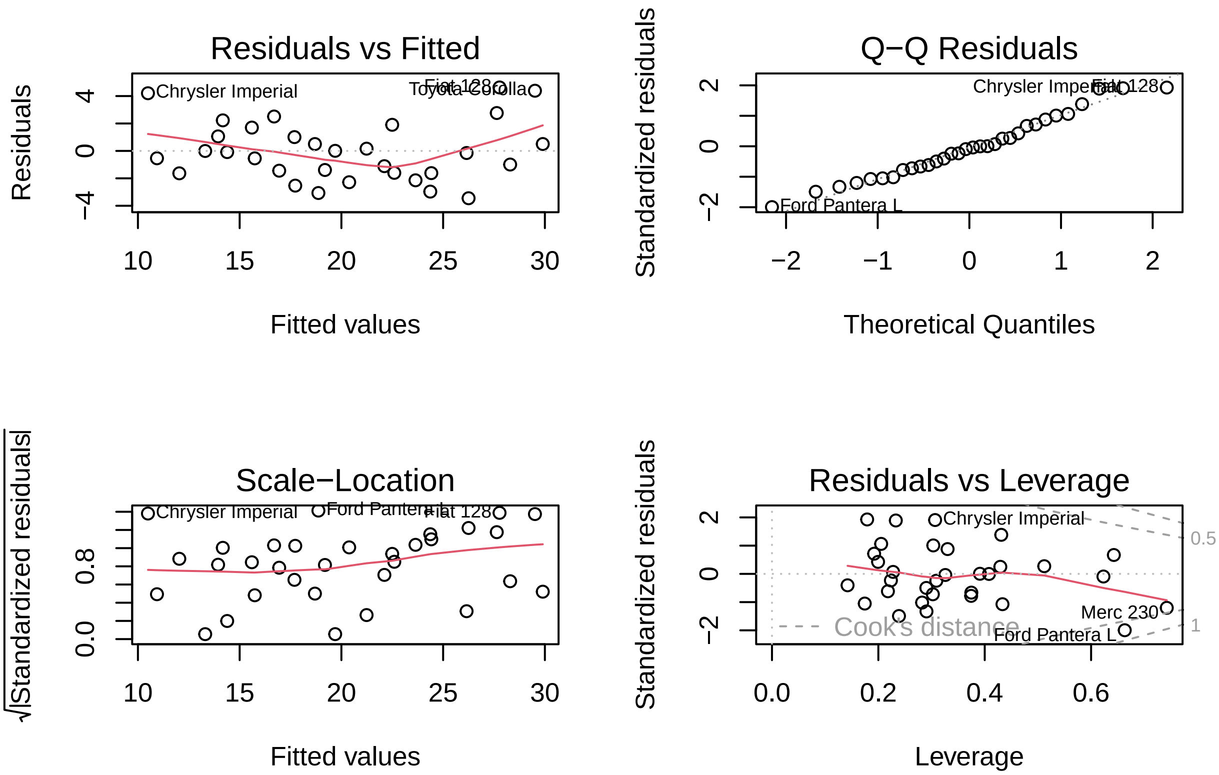
Best selects the best subset of predictors based on the Akaike Information Criterion (AIC), adding or removing predictors one at a time. Best = step(model2, direction = "both")

Interpretation

The stepwise regression procedure selected a final model with three predictors: wt, qsec, and am. This model has the lowest AIC value (61.31) among all the models considered, indicating that it provides the best balance of goodness of fit and model complexity according to the AIC criterion.

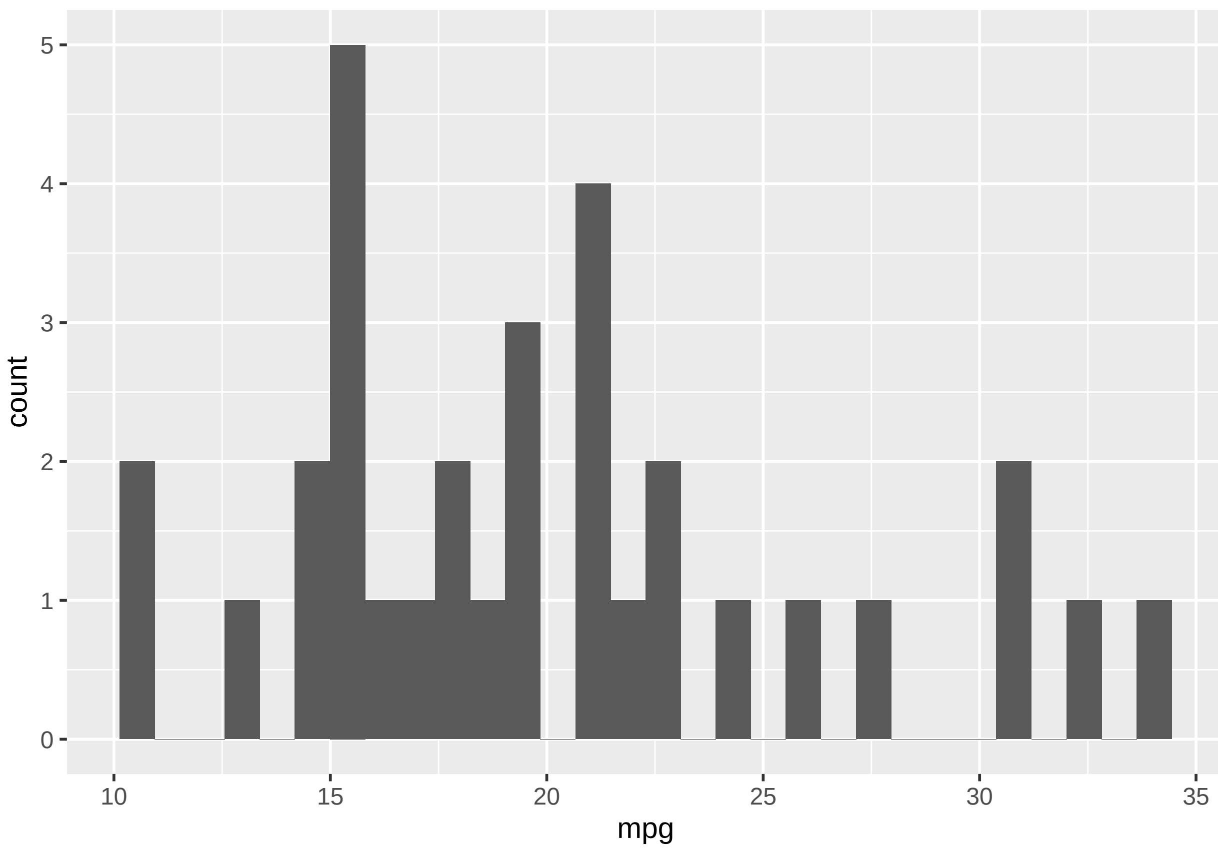
Ploting the Residual

```r
#Residual Plots and Diagnostics
par(mfrow=c(2,2))
plot(model2)
```



The histogram for the mpg data was also plotted and it had a normal distribution.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Conclution

Both models provide insights into the relationship between predictors and the response variable. Model selection depends on the research question, the desired level of complexity, and the importance of interpretability. In this case, while Model 2 explains more variance, Model 1 may be preferred if the focus is solely on the relationship between transmission type and MPG.