R Notebook

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

```
#loading the csv file into r
data <- read.csv("/Users/renaherman/Downloads/auto-mpg(1).csv", header = TRUE)
#converting the horsepower column into numeric values - the ones that cant be converted come up as NA
data$horsepower <- as.numeric(data$horsepower)
```

```
## Warning: NAs introduced by coercion
```

```
#removed all NA values from dataset
data <- na.omit(data)
#splitting the data to only use the first 300 samples
firstHalfdata <- data[1:300, ]
 #creating the linear regression model
simple_model <- lm(mpg ~ weight, data = firstHalfdata)
#getting the summary
summary(simple_model)
```

```
##
## Call:
## lm(formula = mpg ~ weight, data = firstHalfdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.2011 -1.9157 -0.0812  1.7341 15.0246
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 40.5619792  0.6461532   62.77   <2e-16 ***
## weight      -0.0062905  0.0001984  -31.71   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.032 on 298 degrees of freedom
## Multiple R-squared:  0.7714, Adjusted R-squared:  0.7706
## F-statistic:  1005 on 1 and 298 DF,  p-value: < 2.2e-16
```

```
#preforming multiple linear regression
multiple_model <- lm(mpg ~ cylinder + displacement + horsepower + weight + acceleration + `model.year`,
#getting the summary
summary(multiple_model)
```

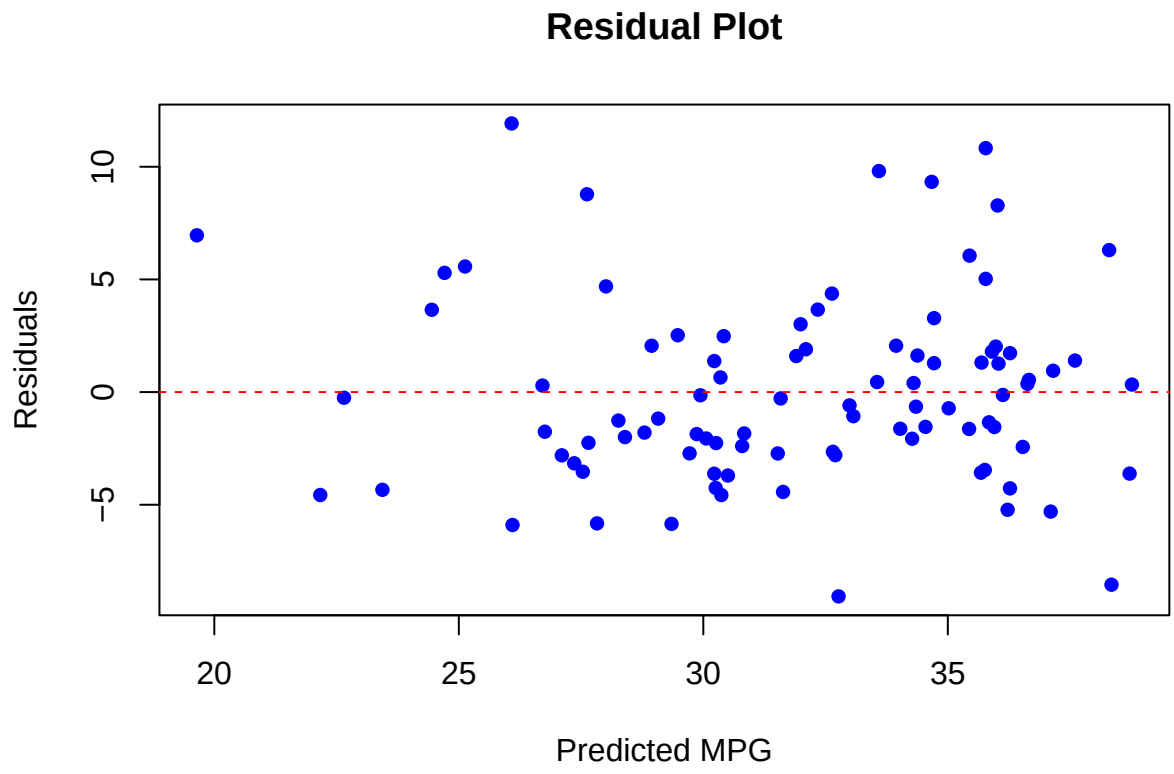
```
##
## Call:
## lm(formula = mpg ~ cylinder + displacement + horsepower + weight +
##     acceleration + model.year, data = firstHalfdata)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.696 -1.728  0.072  1.571 13.703
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.1234027  4.9831096   1.430    0.154
## cylinder     -0.4018046  0.3025860  -1.328    0.185
## displacement  0.0026256  0.0064712   0.406    0.685
## horsepower   -0.0078140  0.0118630  -0.659    0.511
## weight       -0.0054775  0.0005803  -9.440  < 2e-16 ***
## acceleration -0.0417357  0.0951761  -0.439    0.661
## model.year    0.4588389  0.0615861   7.450 1.05e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.745 on 293 degrees of freedom
## Multiple R-squared:  0.8158, Adjusted R-squared:  0.812
## F-statistic: 216.2 on 6 and 293 DF,  p-value: < 2.2e-16
```

```r
#getting the last 98 samples from the dataset
secondHalfdata <- data[301:398,]
#creating linear regression model
model <- lm(mpg ~ weight + model.year, data = secondHalfdata)
#predicting mpg values
predicted_mpg <- predict(model, secondHalfdata)
#displaying the mpg predicted values
predicted_mpg
```

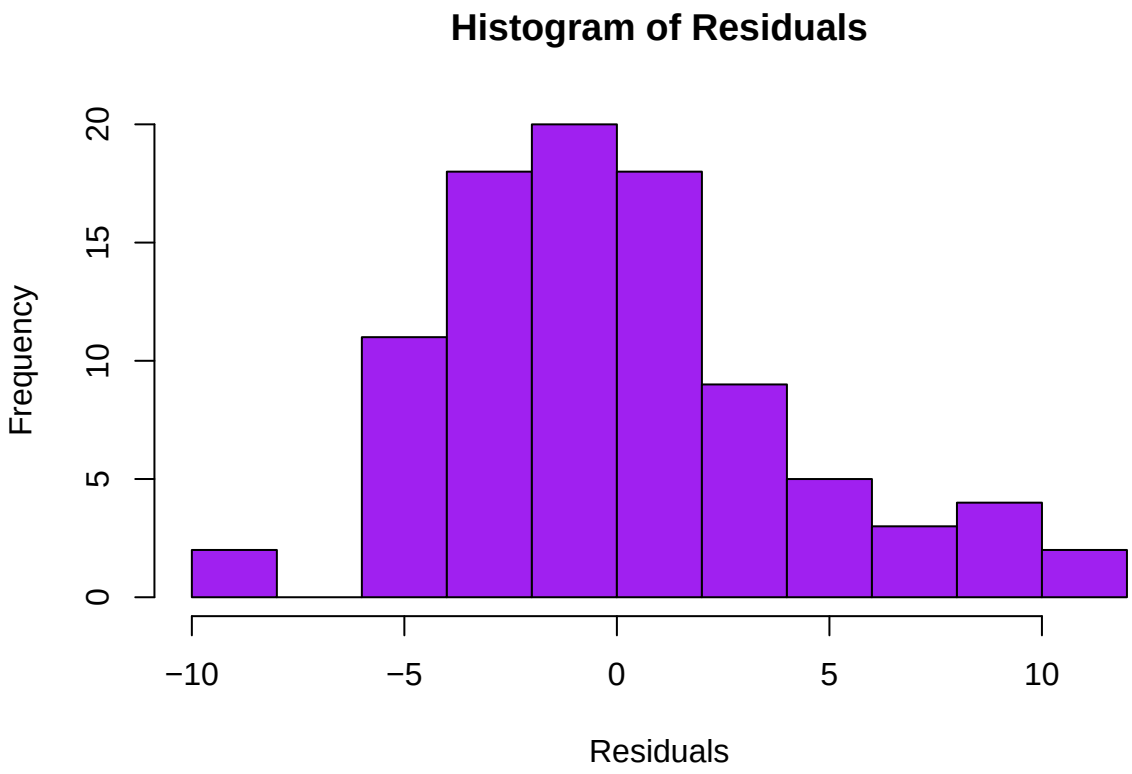
```
##      303      304      305      306      307      308      309      310
## 35.84220 37.10413 36.03634 30.79450 31.52254 30.50329 31.90111 35.44506
##      311      312      313      314      315      316      317      318
## 37.15351 35.67803 36.65845 30.26146 28.39769 27.10665 23.43736 35.01795
##      319      320      321      322      323      324      325      326
## 29.94113 31.58163 32.63000 34.27050 35.77510 29.07719 35.77510 36.01778
##      327      328      329      330      332      333      334      335
## 33.59100 27.62113 24.70899 38.29895 35.43535 38.34748 28.00941 32.76590
##      336      338      339      340      341      342      343      344
## 31.98933 34.02782 31.63101 30.22348 30.36909 29.34984 32.65026 38.76574
##      345      346      347      348      349      350      351      352
## 37.60089 38.71720 35.75653 36.63017 35.90214 36.53310 34.30047 35.95068
##      353      354      356      357      358      359      360      361
## 32.69879 34.54314 34.34900 32.99001 30.41762 30.22348 24.44775 25.12725
##      362      363      364      365      366      367      368      369
## 27.65109 27.35988 22.65193 19.64273 26.09796 22.16658 30.05931 29.71956
##      370      371      372      373      374      376      377      378
## 32.09780 30.35052 30.83588 28.79738 27.53546 36.12625 35.68943 36.22332
##      379      380      381      382      383      384      385      386
## 34.71872 34.71872 34.37897 33.94215 33.55387 36.27186 36.27186 35.98065
```

```
##      387      388      389      390      391      392      393      394
## 26.75889 26.07939 30.25345 27.82667 29.47688 32.34048 26.71035 28.26349
##      395      396      397      398       NA     NA.1     NA.2     NA.3
## 34.67019 33.06851 29.86516 28.94299       NA       NA       NA       NA
##     NA.4     NA.5
##       NA       NA
```

```r
#Calculating the residuals - by seeing the difference between the predicted and actual mpg values
residuals <- secondHalfdata$mpg - predicted_mpg
#plotting the residuals
plot(predicted_mpg, residuals,
     main = "Residual Plot",
     xlab = "Predicted MPG",
     ylab = "Residuals",
     pch = 16, col = "blue")
abline(h = 0, col = "red", lty = 2)
```




```r
#making histogram of residuals
hist(residuals,
     main = "Histogram of Residuals",
     xlab = "Residuals",
     col = "purple",
     border = "black")
```



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.