Martensite Starting Temperature

AUTHOR
Hooman Sabarou & Mounika Chevva-Cross_Validation Group_Fall 2024

```
# Load the dataset
dataset <- read.csv("/Users/hoomansabarou/Downloads/Martensite
head(dataset)
```

```
     C   Mn   Si   Cr   Ni   Mo    V Co   Al W   Cu Nb Ti B     N
Ms
1 0.19 1.25 0.23 0.07 0.66 0.54 0.00  0 0.00 0 0.00  0  0 0 0.000
645.00
2 1.00 0.45 0.00 1.52 3.33 0.00 0.00  0 0.00 0 0.00  0  0 0 0.000
359.00
3 0.26 0.58 0.49 1.65 0.18 0.84 0.38  0 0.00 0 0.07  0  0 0 0.000
653.00
4 0.06 1.96 0.32 0.00 0.00 0.18 0.22  0 0.02 0 0.00  0  0 0 0.005
721.00
5 1.20 1.88 0.00 0.00 0.00 0.00 0.00  0 0.00 0 0.00  0  0 0 0.000
323.00
6 0.38 0.69 0.20 0.95 1.58 0.26 0.00  0 0.00 0 0.00  0  0 0 0.000
589.75
```

```
# Check for missing values
sum(is.na(dataset))
```

```
[1] 0
```

```
# Check the structure of the dataset
str(dataset)
```

```
'data.frame':   1543 obs. of  16 variables:
 $ C : num  0.19 1 0.26 0.06 1.2 0.38 0.43 0.004 0 2 ...
 $ Mn: num  1.25 0.45 0.58 1.96 1.88 0.69 0.83 0.03 0 0.65 ...
 $ Si: num  0.23 0 0.49 0.32 0 0.2 1.55 0.075 0 0.3 ...
 $ Cr: num  0.07 1.52 1.65 0 0 ...
 $ Ni: num  0.66 3.33 0.18 0 0 ...
 $ Mo: num  0.54 0 0.84 0.18 0 0.26 0.4 0 0 0.5 ...
 $ V : num  0 0 0.38 0.22 0 ...
 $ Co: num  0 0 0 0 0 0 0 0 0 0 ...
 $ Al: num  0 0 0 0.02 0 0 0 0 0 0.051 ...
 $ W : num  0 0 0 0 0 0 0 0 0 0 ...
 $ Cu: num  0 0 0.07 0 0 0 0 0 0 0.08 ...
 $ Nb: num  0 0 0 0 0 0 0 0 0 0 ...
 $ Ti: num  0 0 0 0 0 0 0 0 0 0 ...
```

```
$ B : num  0 0 0 0 0 0 0 0 0 0 ...
$ N : num  0 0 0 0.005 0 0 0 0.002 0 0 ...
$ Ms: num  645 359 653 721 323 ...
```

```
# Summary statistics of the dataset
summary(dataset)
```

```
       C                Mn                Si                Cr        
 Min.   :0.0000   Min.   : 0.0000   Min.   :0.0000   Min.   : 0.000  
 1st Qu.:0.1500   1st Qu.: 0.4100   1st Qu.:0.1550   1st Qu.: 0.000  
 Median :0.3200   Median : 0.6500   Median :0.2500   Median : 0.490  
 Mean   :0.3529   Mean   : 0.7624   Mean   :0.3324   Mean   : 1.103  
 3rd Qu.:0.4400   3rd Qu.: 0.9100   3rd Qu.:0.3400   3rd Qu.: 1.130  
 Max.   :2.2500   Max.   :10.2400   Max.   :3.8000   Max.   :16.200  
       Ni               Mo                V                Co        
 Min.   : 0.000   Min.   :0.0000   Min.   :0.0000   Min.   : 0.0000  
 1st Qu.: 0.000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 0.0000  
 Median : 0.180   Median :0.0700   Median :0.0000   Median : 0.0000  
 Mean   : 2.644   Mean   :0.2916   Mean   :0.1052   Mean   : 0.1846  
 3rd Qu.: 1.840   3rd Qu.:0.3550   3rd Qu.:0.0400   3rd Qu.: 0.0000  
 Max.   :31.540   Max.   :8.0000   Max.   :5.0500   Max.   :16.0800  
       Al                W                 Cu                 Nb          
 Min.   :0.00000   Min.   : 0.0000   Min.   :0.00000   Min.   :0.000000  
 1st Qu.:0.00000   1st Qu.: 0.0000   1st Qu.:0.00000   1st Qu.:0.000000  
 Median :0.00000   Median : 0.0000   Median :0.00000   Median :0.000000  
 Mean   :0.02068   Mean   : 0.3723   Mean   :0.04467   Mean   :0.006686  
 3rd Qu.:0.00000   3rd Qu.: 0.0000   3rd Qu.:0.00000   3rd Qu.:0.000000  
 Max.   :3.00679   Max.   :19.2000   Max.   :3.03743   Max.   :1.983240  
       Ti                 B                   N                 Ms        
 Min.   :0.000000   Min.   :0.000e+00   Min.   :0.00000   Min.   :153.2  
 1st Qu.:0.000000   1st Qu.:0.000e+00   1st Qu.:0.00000   1st Qu.:541.0  
 Median :0.000000   Median :0.000e+00   Median :0.00000   Median :599.0  
 Mean   :0.009008   Mean   :9.397e-06   Mean   :0.01454   Mean   :587.6  
 3rd Qu.:0.000000   3rd Qu.:0.000e+00   3rd Qu.:0.00000   3rd Qu.:665.0  
 Max.   :2.524640   Max.   :4.000e-03   Max.   :2.65000   Max.   :788.1  
```

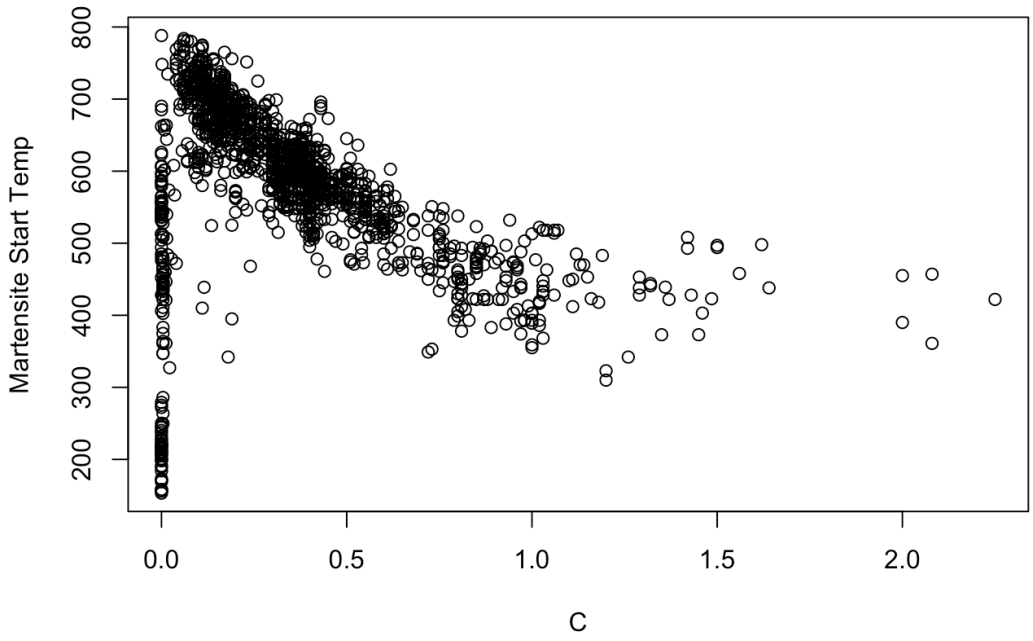
```
# Correlation between variables
cor_matrix <- cor(dataset)
cor_matrix
```

```
            C          Mn           Si           Cr          Ni
Mo
C   1.00000000 -0.05917062  0.076600512  0.133051621 -0.34769552
0.11463635
Mn -0.05917062  1.00000000  0.143637893 -0.040788956 -0.34849206
-0.14600301
Si  0.07660051  0.14363789  1.000000000 -0.003024453 -0.13103635
-0.01174235
```

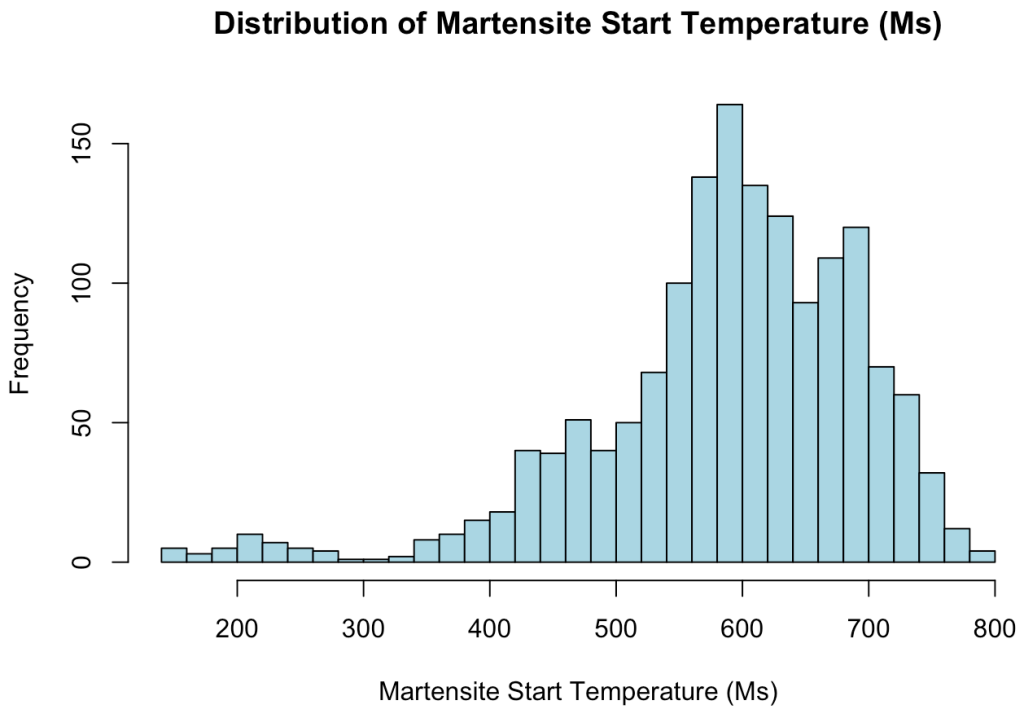
```
Cr  0.13305162 -0.04078896 -0.003024453  1.000000000 -0.14495405
0.20225868
Ni -0.34769552 -0.34849206 -0.131036352 -0.144954052  1.00000000
-0.02131170
Mo  0.11463635 -0.14600301 -0.011742349  0.202258678 -0.02131170
1.00000000
V   0.26802857 -0.14690338  0.020415204  0.233740944  0.01171931
0.40241031
Co  0.06285207 -0.11736792 -0.061284805  0.055137788  0.07811827
0.33894047
Al -0.07543939 -0.05549448 -0.025930426 -0.027279944  0.16949869
-0.03363098
W   0.25293546 -0.13050433 -0.028669274  0.200442198 -0.06329719
0.17030728
Cu -0.05232115 -0.05919812 -0.051157179 -0.046512297  0.12789938
-0.06229066
Nb -0.08143885 -0.06303592 -0.051000046 -0.034834421  0.21724129
-0.02869010
Ti -0.08738355 -0.07660234 -0.049248069 -0.033280530  0.20991758
-0.03131265
B   0.02359553  0.02703234 -0.020884016 -0.022420182 -0.02191465
-0.01437306
N  -0.11664282 -0.01884408 -0.073678342  0.057673828 -0.04105273
-0.04446772
Ms -0.41467171  0.18926516 -0.013710984 -0.112092664 -0.54617556
-0.08461144
             V           Co           Al            W          Cu
Nb
C   0.26802857  0.062852075 -0.075439388  0.252935464 -0.05232115
-0.081438850
Mn -0.14690338 -0.117367920 -0.055494477 -0.130504332 -0.05919812
-0.063035918
Si  0.02041520 -0.061284805 -0.025930426 -0.028669274 -0.05115718
-0.051000046
Cr  0.23374094  0.055137788 -0.027279944  0.200442198 -0.04651230
-0.034834421
Ni  0.01171931  0.078118265  0.169498686 -0.063297195  0.12789938
0.217241285
Mo  0.40241031  0.338940473 -0.033630984  0.170307279 -0.06229066
-0.028690095
V   1.00000000  0.264419567 -0.033061955  0.642312832 -0.04438009
-0.018380708
Co  0.26441957  1.000000000  0.002182604  0.268512394 -0.03393445
-0.006318026
Al -0.03306196  0.002182604  1.000000000 -0.024725669 -0.02149697
-0.008568638
W   0.64231283  0.268512394 -0.024725669  1.000000000 -0.03591799
-0.012666271
Cu -0.04438009 -0.033934454 -0.021496975 -0.035917994  1.00000000
```

```
-0.015557539
Nb -0.01838071 -0.006318026 -0.008568638 -0.012666271 -0.01555754
1.000000000
Ti -0.01800920 -0.010477839 -0.010090645 -0.013142225 -0.01211753
-0.005448801
B  -0.01523008 -0.008181823 -0.008071450 -0.009565916 -0.01292970
-0.003690127
N  -0.02611876 -0.013862435 -0.013261464 -0.018199441 -0.02520760
-0.007211023
Ms -0.20004307 -0.067177142 -0.058711859 -0.170917196 -0.04498824
-0.122781202
             Ti            B            N          Ms
C  -0.087383551  0.023595535 -0.116642823 -0.41467171
Mn -0.076602343  0.027032341 -0.018844080  0.18926516
Si -0.049248069 -0.020884016 -0.073678342 -0.01371098
Cr -0.033280530 -0.022420182  0.057673828 -0.11209266
Ni  0.209917583 -0.021914647 -0.041052728 -0.54617556
Mo -0.031312653 -0.014373055 -0.044467718 -0.08461144
V  -0.018009196 -0.015230081 -0.026118761 -0.20004307
Co -0.010477839 -0.008181823 -0.013862435 -0.06717714
Al -0.010090645 -0.008071450 -0.013261464 -0.05871186
W  -0.013142225 -0.009565916 -0.018199441 -0.17091720
Cu -0.012117528 -0.012929705 -0.025207602 -0.04498824
Nb -0.005448801 -0.003690127 -0.007211023 -0.12278120
Ti  1.000000000 -0.004235017 -0.007578545 -0.08367866
B  -0.004235017  1.000000000 -0.005993731 -0.02009083
N  -0.007578545 -0.005993731  1.000000000 -0.14695161
Ms -0.083678655 -0.020090832 -0.146951608  1.00000000
```

```r
# Scatter plot for one element vs Ms
plot(dataset$C, dataset$Ms, xlab = "C", ylab = "Martensite Star
```




```r
# Plot histogram of Ms (Martensite start temperature)
hist(dataset$Ms,
     main = "Distribution of Martensite Start Temperature (Ms)"
     xlab = "Martensite Start Temperature (Ms)",
     col = "lightblue",
     border = "black",
     breaks = 30)
```



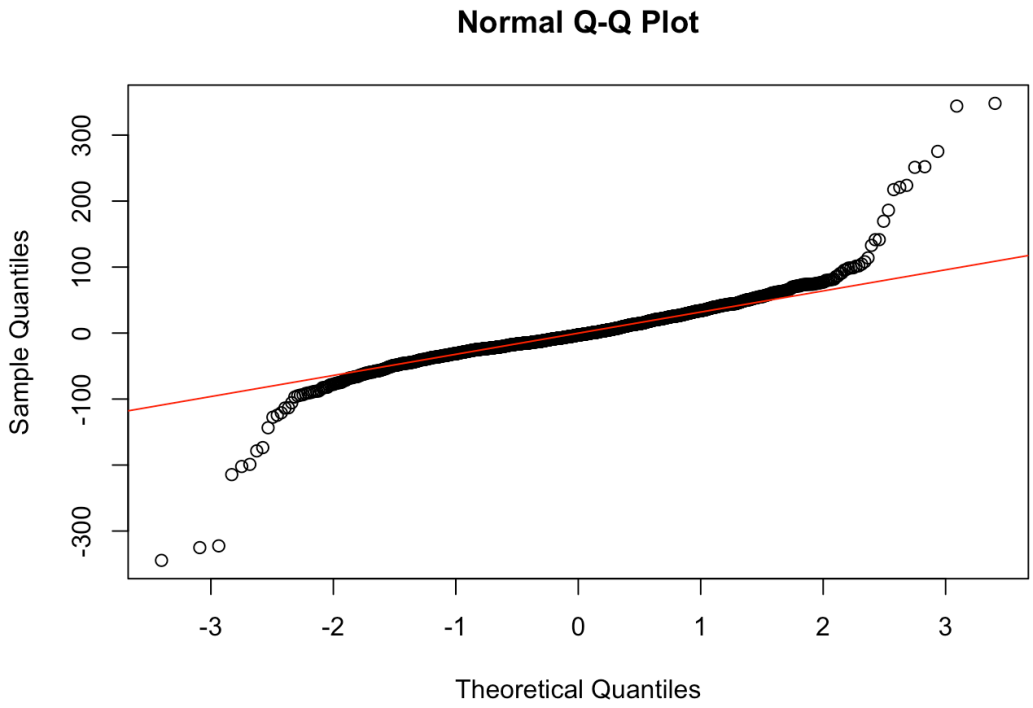

```
# Filter the dataset for Ms between 300°C and 800°C
filtered_dataset <- subset(dataset, Ms >= 300 & Ms <= 800)
```

```
# Build the updated GLM model using filtered data
updated_model <- glm(Ms ~ C + Ni + Mn + Mo + Si + Cr + V + Co,

# Summarize the updated model
summary(updated_model)
```

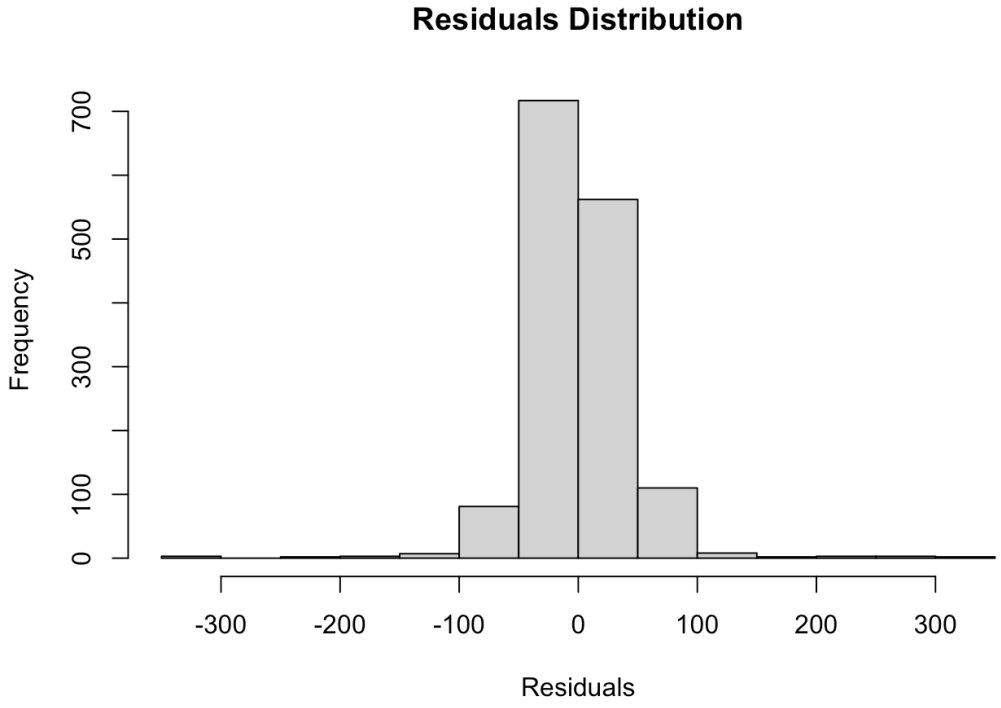
```
Call:
glm(formula = Ms ~ C + Ni + Mn + Mo + Si + Cr + V + Co, family =
gaussian,
    data = filtered_dataset)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  750.6469     3.1633 237.300  < 2e-16 ***
C           -260.1772     4.4241 -58.809  < 2e-16 ***
Ni           -12.4065     0.2859 -43.394  < 2e-16 ***
Mn           -25.5874     1.9657 -13.017  < 2e-16 ***
Mo            -7.7315     2.0623  -3.749 0.000184 ***
Si           -15.6230     2.9154  -5.359 9.69e-08 ***
Cr            -7.1898     0.5425 -13.254  < 2e-16 ***
V              8.3285     3.5131   2.371 0.017881 *
```

```
Co             2.0491     0.9144   2.241 0.025184 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1997.601)

    Null deviance: 12075332  on 1502  degrees of freedom
Residual deviance:  2984416  on 1494  degrees of freedom
AIC: 15699

Number of Fisher Scoring iterations: 2
```

```r
# Check residuals with a QQ plot
qqnorm(residuals(updated_model))
qqline(residuals(updated_model), col = "red")
```




```r
# Histogram of residuals
hist(residuals(updated_model), main = "Residuals Distribution",
```




```
        # Identify outliers
        outlier_residuals <- which(abs(residuals(updated_model)) > 2 *
        dataset[outlier_residuals, ]
```

```
         C   Mn      Si       Cr       Ni      Mo     V   Co    Al   W
Cu Nb
9    0.000 0.00 0.00000  0.53096 30.96549 0.00000 0.000 0.00 0.000 0.0
0.00  0
48   0.280 0.39 0.16000  2.35000  0.06000 0.06000 0.530 0.00 0.000 4.1
0.00  0
97   0.000 0.00 0.00000  0.00000 30.77361 1.36279 0.000 0.00 0.000 0.0
0.00  0
193  0.500 0.35 1.00000  0.11000  0.19000 0.50000 0.000 0.00 0.000 0.0
0.00  0
209  0.300 0.48 2.20000 10.50000  0.12000 1.00000 0.012 0.00 0.000 0.0
0.07  0
220  0.300 1.55 0.20000  0.00000  0.00000 0.28000 0.000 0.00 0.000 0.0
0.00  0
302  0.110 0.50 0.22000  0.00000  0.00000 0.56000 0.000 0.00 0.003 0.0
0.00  0
328  0.160 0.60 0.25000  0.20000  1.50000 0.05000 0.000 0.00 0.000 0.0
0.00  0
331  0.000 0.00 1.92824  0.00000 31.53985 0.00000 0.000 0.00 0.000 0.0
0.00  0
334  0.550 0.75 1.50000  0.70000  0.00000 0.00000 0.000 0.00 0.000 0.0
```

```
0.00  0
365  1.040 0.33 0.26000  1.53000  0.31000 0.01000 0.010 0.00 0.000 0.0
0.20  0
387  0.330 1.12 0.30000  0.11000  0.24000 0.04000 0.000 0.00 0.010 0.0
0.19  0
389  0.190 1.17 0.41000  0.06000  0.00000 0.00000 0.000 0.00 0.000 0.0
0.00  0
403  0.400 1.38 1.50000  0.00000  0.00000 0.80000 0.000 0.00 0.000 0.0
0.00  0
441  0.190 0.46 0.34000  7.83000  0.09000 2.02000 0.010 0.00 0.005 0.0
0.00  0
500  0.260 0.76 0.32000  1.08000  0.72000 1.25000 0.310 0.00 0.000 0.0
0.00  0
555  0.200 1.88 0.00000  0.00000  0.00000 0.00000 0.000 0.00 0.000 0.0
0.00  0
573  0.380 0.82 1.48000  0.72000  0.00000 0.77000 0.000 0.00 0.000 0.0
0.00  0
624  0.440 0.75 0.26000  1.70000  0.17000 0.08000 0.090 0.00 0.000 0.0
0.18  0
629  0.430 0.95 1.38000  1.06000  0.03000 0.10000 0.035 0.00 0.000 0.0
0.05  0
643  0.380 1.45 0.36000  0.00000  0.00000 0.76000 0.000 0.00 0.000 0.0
0.00  0
691  0.100 0.47 0.28000  1.32000  2.34000 0.00000 0.000 0.00 0.000 0.0
0.87  0
701  0.110 0.00 0.00000  0.00000  3.49000 0.00000 0.000 0.00 0.000 0.0
0.00  0
702  0.470 0.82 0.35000  1.20000  0.04000 0.00000 0.110 0.00 0.000 0.0
0.14  0
710  0.150 0.36 0.44000  2.24000  0.09000 0.85000 0.000 0.00 0.097 0.0
0.23  0
821  0.220 0.83 0.24000  0.54000  1.06000 0.51000 0.000 0.00 0.029 0.0
0.30  0
847  0.190 1.00 0.04000  0.62000  0.02000 0.00000 0.000 0.00 0.000 0.0
0.00  0
877  0.360 0.78 0.31000  0.00000  0.73000 0.49000 0.000 0.00 0.000 0.0
0.00  0
882  0.300 0.70 0.20000  0.00000  0.00000 0.00000 0.000 0.00 0.000 0.0
0.00  0
895  0.004 0.03 0.07500  0.00000 29.55000 0.00000 0.000 0.00 0.000 0.0
0.00  0
967  0.430 0.83 1.55000  0.91000  3.02000 0.40000 0.120 0.00 0.000 0.0
0.00  0
1035 0.400 1.47 0.37000  0.00000  0.00000 0.26000 0.000 0.00 0.000 0.0
0.00  0
1069 0.170 0.49 0.29000  0.18000  5.07000 0.24000 0.000 0.00 0.000 0.0
0.10  0
1091 0.330 0.74 0.23000  0.07000  3.47000 0.00000 0.000 0.00 0.000 0.0
0.00  0
```

1112 0.390 0.70 0.20000 1.05000 0.00000 0.00000 0.000 0.00 0.000 0.0
0.00 0
1144 0.120 0.46 0.35000 4.79000 0.20000 0.54000 0.000 0.00 0.000 0.0
0.00 0
1148 0.400 0.80 0.33000 0.00000 0.00000 0.79000 0.000 0.00 0.000 0.0
0.00 0
1166 0.400 0.75 0.27000 0.96000 0.13000 0.07000 0.060 0.00 0.000 0.0
0.20 0
1223 0.540 0.46 0.00000 0.00000 0.00000 0.00000 0.000 0.00 0.000 0.0
0.00 0
1225 0.470 0.40 1.06000 0.00000 0.00000 0.00000 0.000 0.00 0.000 0.0
0.00 0
1226 0.910 0.65 0.23000 0.60000 1.35000 0.00000 0.000 0.00 0.000 0.0
0.03 0
1242 0.870 1.78 0.29000 0.20000 0.15000 0.03000 0.000 0.00 0.000 0.0
0.00 0
1291 0.390 1.67 0.00000 0.00000 0.10000 0.00000 0.000 0.00 0.000 0.0
0.00 0
1302 0.360 1.50 0.20000 0.00000 0.00000 0.00000 0.000 0.00 0.000 0.0
0.00 0
1346 0.430 1.57 0.23000 0.12000 0.20000 0.07000 0.000 0.00 0.000 0.0
0.00 0
1398 0.300 0.40 0.30000 0.86000 3.20000 0.40000 0.000 0.00 0.000 0.0
0.00 0
1436 0.550 0.65 0.20000 0.00000 0.65000 0.00000 0.000 0.00 0.000 0.0
0.00 0
1492 0.760 0.25 0.35000 4.54000 0.00000 5.75000 2.050 0.86 0.000 6.6
0.00 0

	Ti	B	N	Ms
9	0.000	0	0.000	218.1500
48	0.000	0	0.000	673.0000
97	0.000	0	0.000	199.1500
193	0.000	0	0.000	573.0000
209	0.000	0	0.000	528.0000
220	0.000	0	0.000	628.0000
302	0.000	0	0.002	775.0000
328	0.000	0	0.000	647.0000
331	0.000	0	0.000	171.1500
334	0.000	0	0.000	551.0000
365	0.000	0	0.000	518.0000
387	0.000	0	0.000	628.0000
389	0.000	0	0.000	700.0000
403	0.000	0	0.000	591.0000
441	0.000	0	0.013	623.0000
500	0.000	0	0.000	634.0000
555	0.000	0	0.000	666.0000
573	0.000	0	0.000	607.0000
624	0.000	0	0.000	573.0000
629	0.000	0	0.000	570.0000

```
643  0.000 0 0.000 601.0000
691  0.046 0 0.000 673.0000
701  0.000 0 0.000 722.0000
702  0.000 0 0.000 571.5000
710  0.010 0 0.000 698.0000
821  0.000 0 0.000 673.0000
847  0.000 0 0.000 705.0000
877  0.000 0 0.000 636.0000
882  0.000 0 0.000 666.0000
895  0.000 0 0.002 249.8167
967  0.000 0 0.000 568.0000
1035 0.000 0 0.000 596.0000
1069 0.000 0 0.000 641.0000
1091 0.000 0 0.000 583.0000
1112 0.000 0 0.000 606.0000
1144 0.000 0 0.000 736.0000
1148 0.000 0 0.000 603.0000
1166 0.035 0 0.010 634.0000
1223 0.000 0 0.000 596.0000
1225 0.000 0 0.000 594.2611
1226 0.000 0 0.000 422.0000
1242 0.000 0 0.000 422.0000
1291 0.000 0 0.000 618.0000
1302 0.000 0 0.000 603.0000
1346 0.000 0 0.000 572.0000
1398 0.000 0 0.000 598.0000
1436 0.000 0 0.000 548.0000
1492 0.000 0 0.000 473.0000
```

```r
# Remove rows where C is 0
filtered_dataset_no_C0 <- subset(filtered_dataset, C != 0)

# Build the updated GLM model using the filtered data
updated_model_no_C0 <- glm(Ms ~ C + Ni + Mn + Mo + Si + Cr + V

# Summarize the updated model
summary(updated_model_no_C0)
```

```
Call:
glm(formula = Ms ~ C + Ni + Mn + Mo + Si + Cr + V + Co, family =
gaussian,
    data = filtered_dataset_no_C0)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  759.1787     3.0472 249.140  < 2e-16 ***
C           -266.4946     4.1713 -63.887  < 2e-16 ***
```

```
Ni           -14.0982     0.3195 -44.121  < 2e-16 ***
Mn           -29.0478     1.8747 -15.494  < 2e-16 ***
Mo            -8.0919     2.0278  -3.991 6.92e-05 ***
Si           -16.6972     2.7500  -6.072 1.61e-09 ***
Cr            -7.5038     0.5069 -14.804  < 2e-16 ***
V              8.3673     3.4113   2.453  0.01429 *
Co             2.3044     0.8627   2.671  0.00764 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1730.735)

    Null deviance: 11796080  on 1456  degrees of freedom
Residual deviance:  2506104  on 1448  degrees of freedom
AIC: 15010

Number of Fisher Scoring iterations: 2
```

```r
# Update the GLM model by removing Co and Mo, and adding an int
updated_model_interaction <- glm(Ms ~ C * Mn + Ni + Si + Cr + V
                                 data = filtered_dataset_no_C0,

# Summarize the updated model
summary(updated_model_interaction)
```

```
Call:
glm(formula = Ms ~ C * Mn + Ni + Si + Cr + V, family = gaussian,
    data = filtered_dataset_no_C0)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  749.7319     3.1524 237.829  < 2e-16 ***
C           -240.0692     5.6937 -42.164  < 2e-16 ***
Mn           -18.6538     2.3738  -7.858 7.54e-15 ***
Ni           -13.7303     0.3167 -43.358  < 2e-16 ***
Si           -13.2855     2.7810  -4.777 1.96e-06 ***
Cr            -8.3034     0.5069 -16.382  < 2e-16 ***
V              1.1567     3.1278   0.370    0.712
C:Mn         -39.6809     6.0769  -6.530 9.09e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1701.745)

    Null deviance: 11796080  on 1456  degrees of freedom
Residual deviance:  2465829  on 1449  degrees of freedom
AIC: 14984
```

```
Number of Fisher Scoring iterations: 2
```

```
        # Update the model: remove V, and add interaction between C and
        updated_model_C_Ni_interaction <- glm(Ms ~ C * Mn + C * Ni + Si
                                              data = filtered_dataset_n
                                              family = gaussian)

        # Summarize the updated model
        summary(updated_model_C_Ni_interaction)
```

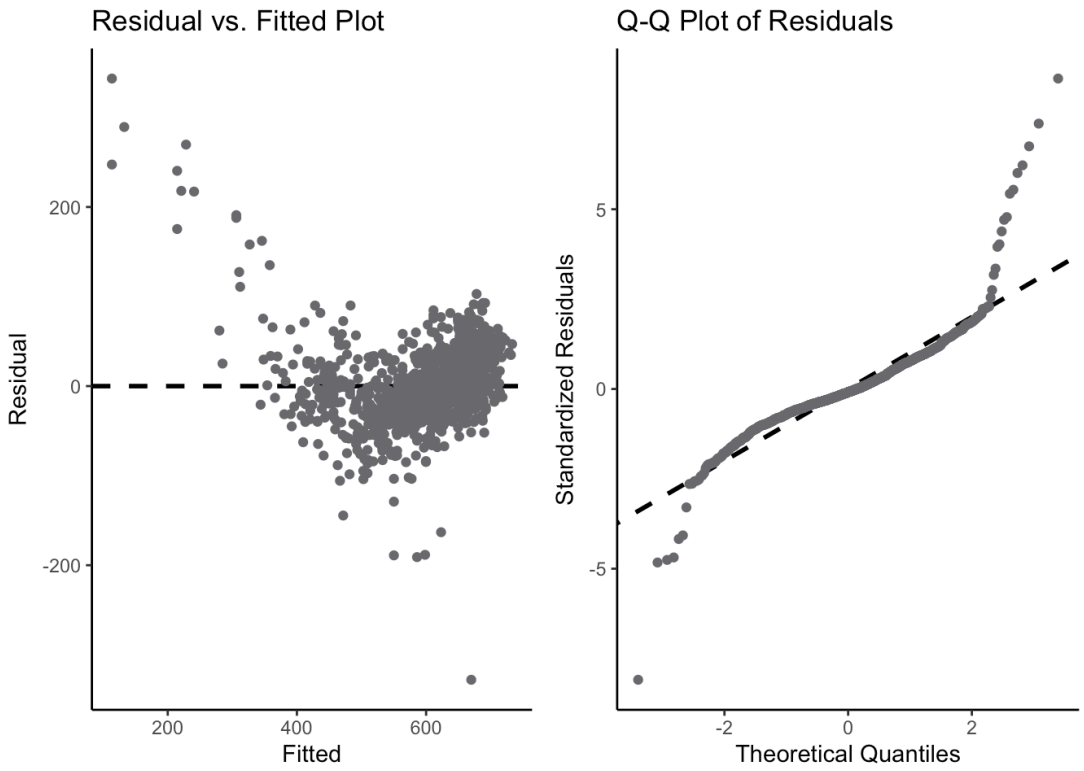
```
Call:
glm(formula = Ms ~ C * Mn + C * Ni + Si + Cr, family = gaussian,
    data = filtered_dataset_no_C0)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  751.5794     3.0913 243.129  < 2e-16 ***
C           -231.2782     5.4214 -42.660  < 2e-16 ***
Mn           -19.0683     2.3316  -8.178 6.23e-16 ***
Ni           -13.0691     0.3241 -40.326  < 2e-16 ***
Si           -15.1211     2.7398  -5.519 4.03e-08 ***
Cr            -8.5661     0.4913 -17.436  < 2e-16 ***
C:Mn         -41.8580     5.8686  -7.133 1.55e-12 ***
C:Ni         -13.8302     1.9610  -7.053 2.71e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1645.423)

    Null deviance: 11796080  on 1456  degrees of freedom
Residual deviance:  2384218  on 1449  degrees of freedom
AIC: 14935

Number of Fisher Scoring iterations: 2
```

```
        library(classpackage)
```

```
Loading required package: ggplot2

Loading required package: ggpubr

Loading required package: purrr

Loading required package: tidyr

Loading required package: dplyr
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

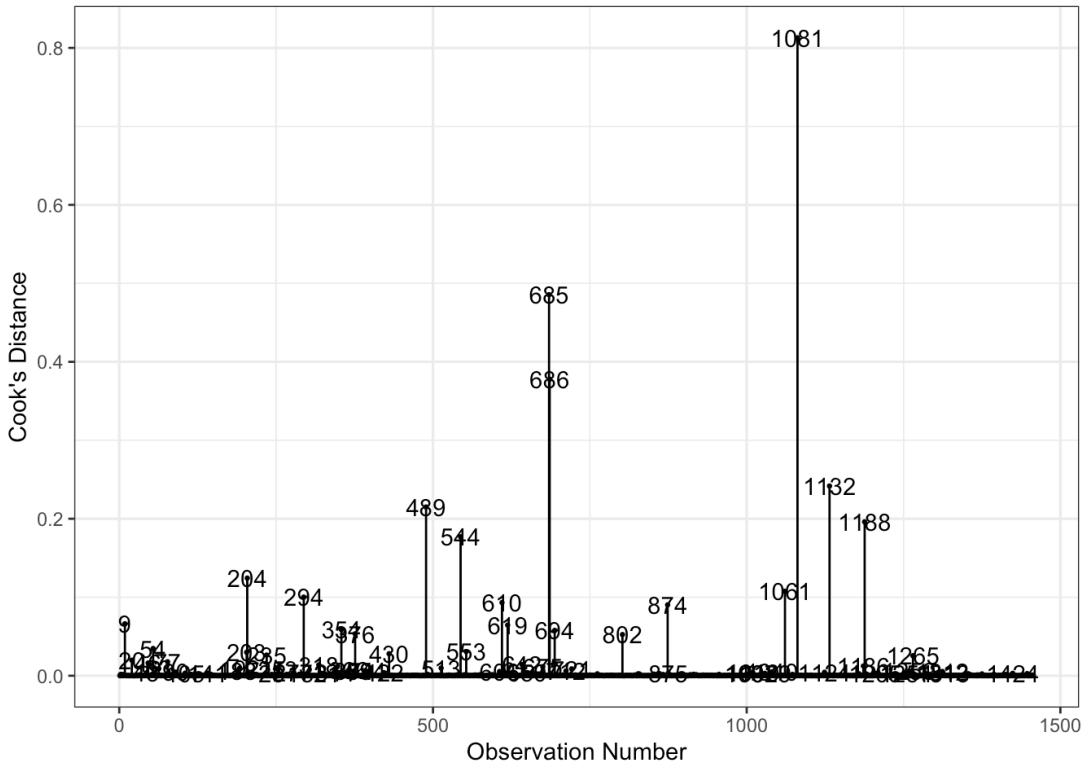
```
# Perform ANOVA on the updated model
anova_check(updated_model_C_Ni_interaction)
```




```
library(glmtoolbox)
(adjR2(updated_model_C_Ni_interaction))
```

```
[1] 0.7969
```

```
cooks(updated_model_C_Ni_interaction)
```




```
# Investigate the influential points
influential_points <- c(1081, 685, 686)
filtered_dataset_no_C0[influential_points, ]
```

```
      C    Mn   Si    Cr   Ni   Mo    V Co Al W   Cu Nb Ti B     N
Ms
1145 0.022 10.24 0.00  8.19 0.00 0.00 0.00  0  0 0 0.00  0  0 0 0.206
327.15
718  2.250  0.00 0.00 11.50 0.00 0.80 0.20  0  0 0 0.00  0  0 0 0.000
422.00
719  2.080  0.39 0.28 11.48 0.31 0.02 0.04  0  0 0 0.15  0  0 0 0.000
457.00
```

```
# Remove the influential points
filtered_dataset_no_influential <- filtered_dataset_no_C0[-c(10

# Rebuild the model after removing the influential points
updated_model_no_influential <- glm(Ms ~ C * Mn + C * Ni + Si +
                                    data = filtered_dataset_no_
                                    family = gaussian)

# Summarize the updated model
summary(updated_model_no_influential)
```

```
Call:
glm(formula = Ms ~ C * Mn + C * Ni + Si + Cr, family = gaussian,
    data = filtered_dataset_no_influential)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  753.7295     3.2150 234.445  < 2e-16 ***
C           -242.2463     5.5389 -43.735  < 2e-16 ***
Mn           -16.9666     2.6015  -6.522 9.58e-11 ***
Ni           -13.2149     0.3142 -42.053  < 2e-16 ***
Si           -15.0063     2.5997  -5.772 9.55e-09 ***
Cr            -9.1535     0.4711 -19.429  < 2e-16 ***
C:Mn         -40.5155     5.9696  -6.787 1.67e-11 ***
C:Ni         -12.4910     1.8647  -6.699 3.01e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1481.353)

    Null deviance: 11670630  on 1453  degrees of freedom
Residual deviance:  2142037  on 1446  degrees of freedom
AIC: 14751

Number of Fisher Scoring iterations: 2
```

```
# Investigate the observations with moderate Cook's Distance
moderate_influential_points <- c(1132, 1188, 1061)
filtered_dataset_no_C0[moderate_influential_points, ]
```

```
        C   Mn   Si    Cr   Ni   Mo    V Co Al    W   Cu Nb Ti B N  Ms
1199 1.36 1.84 1.14  0.15 1.81 1.41 0.00  0  0 0.00 0.00  0  0 0 0 439
1259 2.08 0.39 0.28 11.48 0.31 0.02 0.04  0  0 0.00 0.15  0  0 0 0 361
1122 1.56 0.37 0.20 12.46 0.26 0.54 0.65  0  0 0.28 0.10  0  0 0 0 458
```

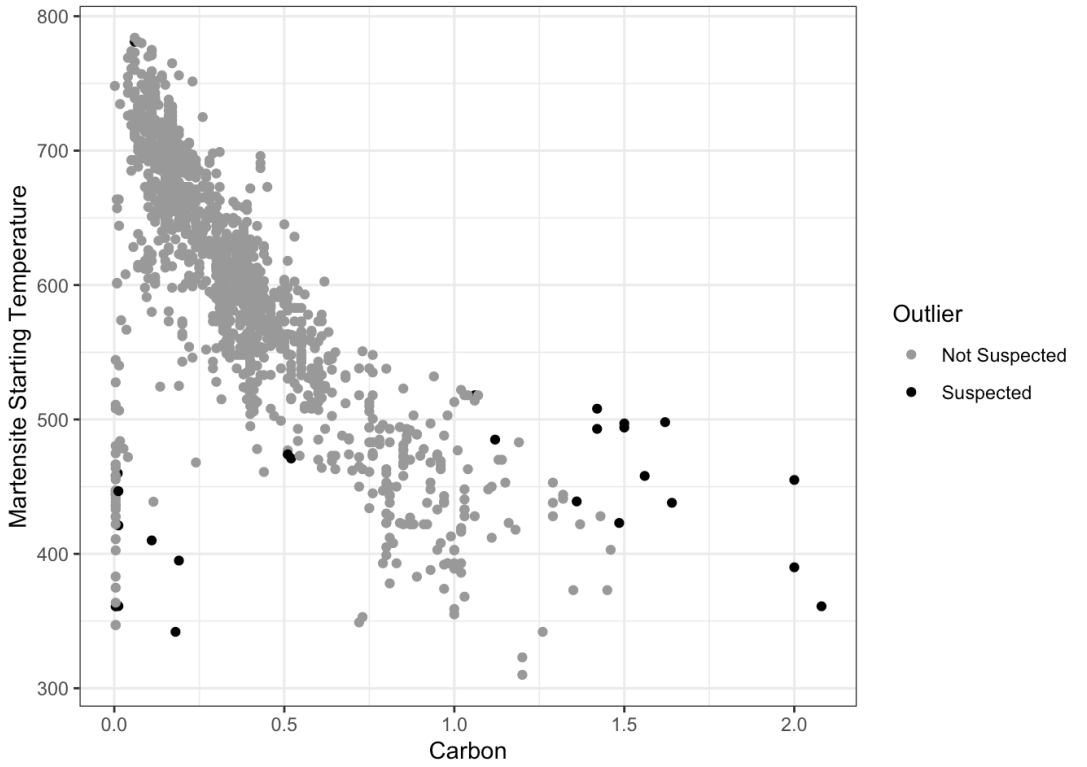
```
# Apply the residuals check to the correct dataset (filtered_da
filtered_dataset_no_influential <- filtered_dataset_no_influent
  mutate(outlier = if_else(abs(rstandard(updated_model_no_influ
                           "Suspected", "Not Suspected"))
```

```
filtered_dataset_no_influential %>% count(outlier)
```

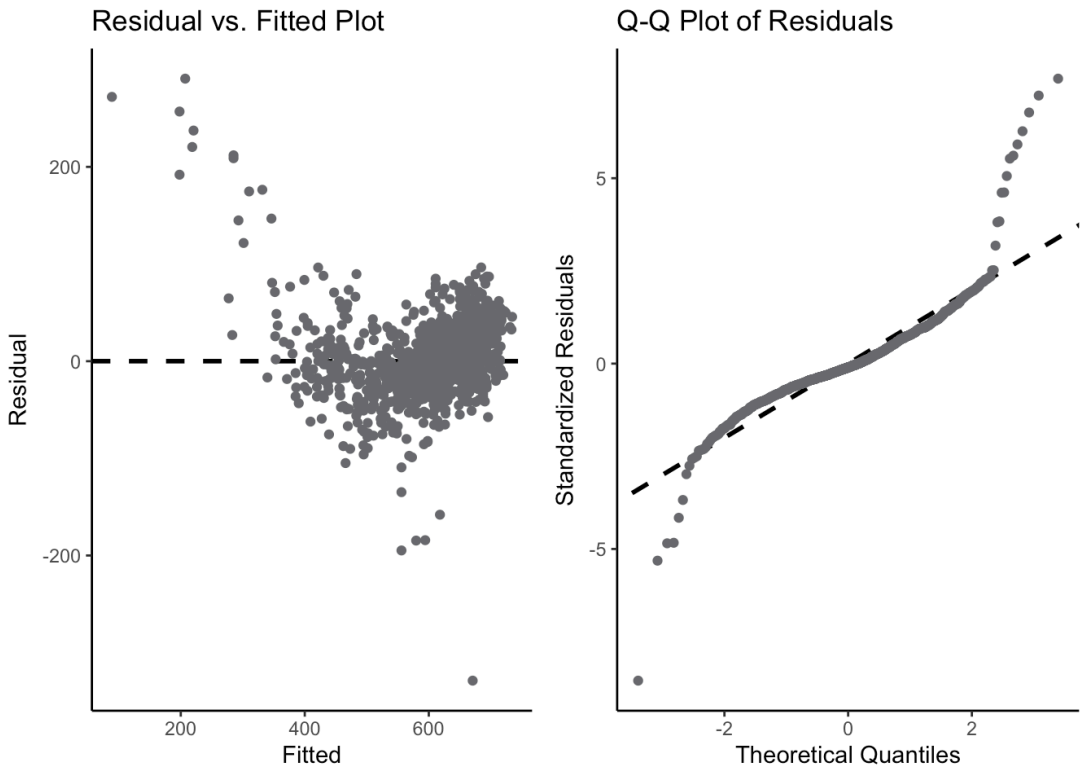
```
        outlier    n
1 Not Suspected 1429
2     Suspected   25
```

```
library(ggplot2)
filtered_dataset_no_influential %>% ggplot(aes(x = C, y = Ms, c
```

```
        geom_point() +
        scale_color_manual(values = c("#999999", "#000000")) +
        labs(x = "Carbon", y = "Martensite Starting Temperature", col
        theme_bw()
```




```
anova_check(updated_model_no_influential)
```




```
#Check Multicollinearity
car::vif(updated_model_no_influential)
```

```
there are higher-order terms (interactions) in this model
consider setting type = 'predictor'; see ?vif
```

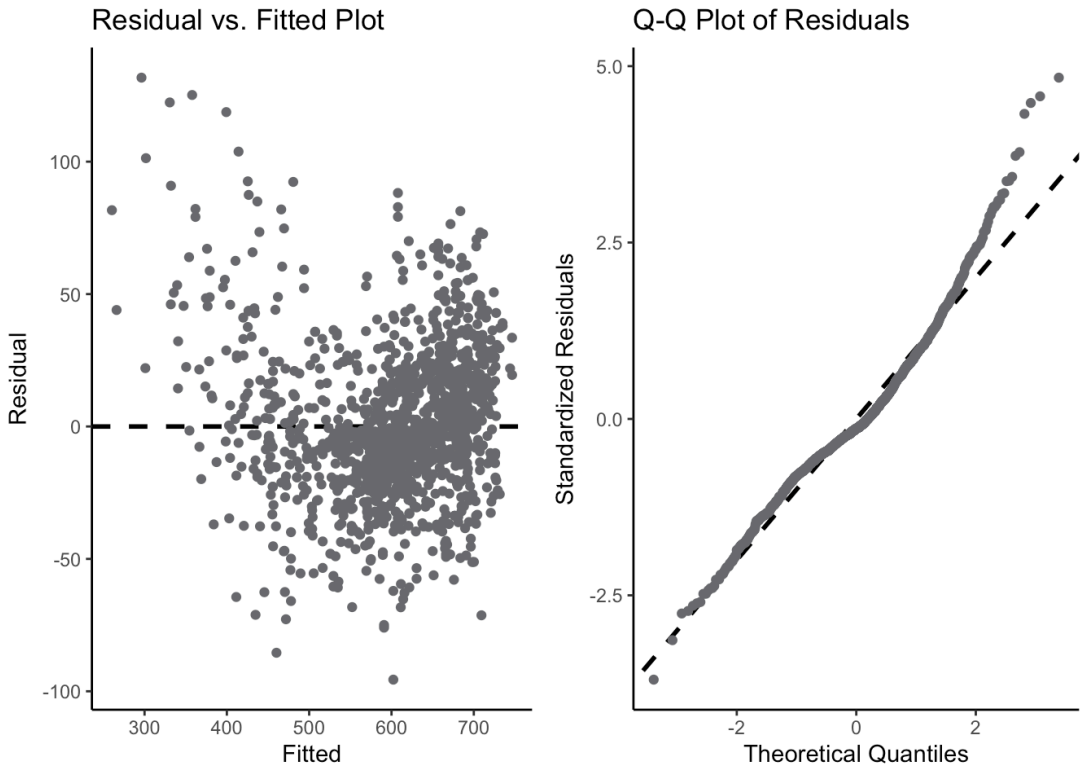
```
       C       Mn       Ni       Si       Cr     C:Mn     C:Ni 
2.392626 2.309298 1.417059 1.067519 1.073750 2.869351 1.194094 
```

```
#Removing identified outliers
filtered_dataset_no_influential_no_outlier <- filtered_dataset_

updated_model_no_influential_2 <- glm(Ms ~ C * Mn + C * Ni + Si
                                      data = filtered_dataset_no_
                                      family = gaussian)

summary(updated_model_no_influential_2)
```

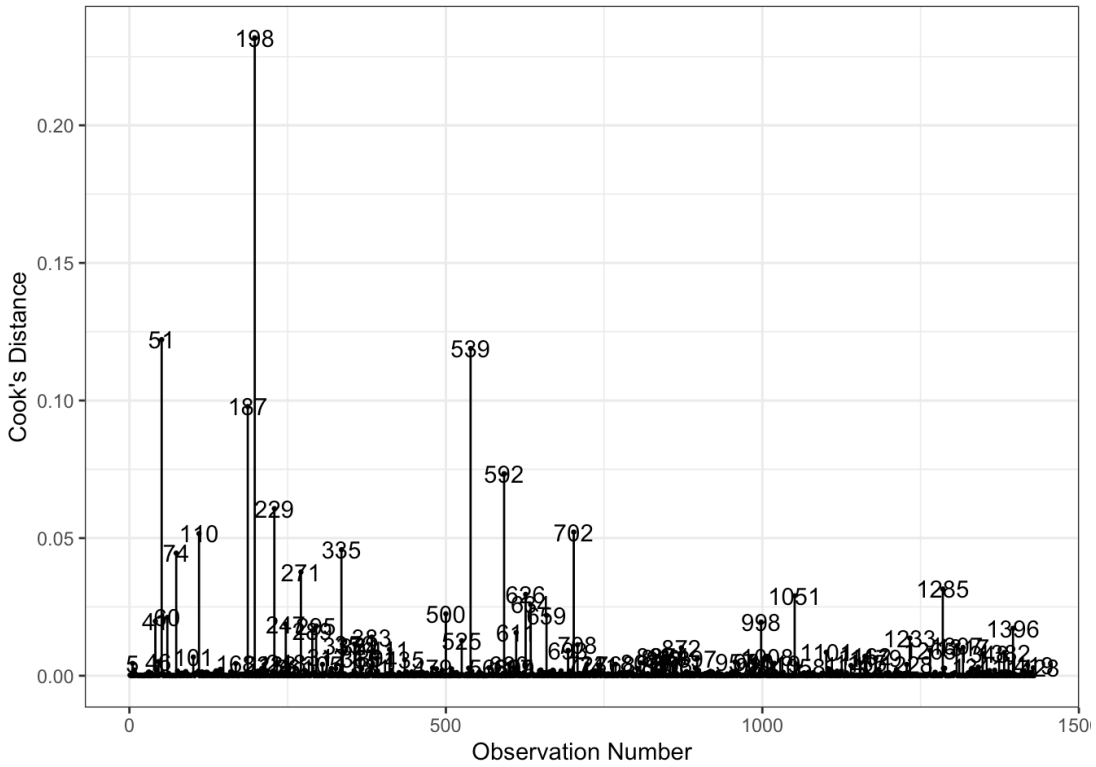
```
Call:
glm(formula = Ms ~ C * Mn + C * Ni + Si + Cr, family = gaussian,
    data = filtered_dataset_no_influential_no_outlier)

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

```
(Intercept)  769.4121     2.6216 293.489  < 2e-16 ***
C           -286.7145     4.5296 -63.299  < 2e-16 ***
Mn           -16.4197     2.1971  -7.473 1.36e-13 ***
Ni           -14.0369     0.2361 -59.454  < 2e-16 ***
Si           -13.8905     1.8663  -7.443 1.70e-13 ***
Cr           -10.1289     0.3825 -26.478  < 2e-16 ***
C:Mn         -41.4509     4.7323  -8.759  < 2e-16 ***
C:Ni          -8.3589     1.3578  -6.156 9.68e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 760.2129)

    Null deviance: 10969684  on 1428  degrees of freedom
Residual deviance:  1080263  on 1421  degrees of freedom
AIC: 13545

Number of Fisher Scoring iterations: 2
```

```
anova_check(updated_model_no_influential_2)
```




```
cooks(updated_model_no_influential_2)
```



Ms	C	Mn	Ni	Si	Cr
Min. :310.0	Min. :0.0016	Min. :0.0000	Min. : 0.000	Min. :0.0000	Min. : 0.000
1st Qu.:553.5	1st Qu.:0.1600	1st Qu.:0.4700	1st Qu.: 0.000	1st Qu.:0.2000	1st Qu.: 0.000
Median :605.0	Median :0.3300	Median :0.6900	Median : 0.150	Median :0.2600	Median : 0.520
Mean :601.8	Mean :0.3617	Mean :0.7917	Mean : 1.558	Mean :0.3475	Mean : 1.043
3rd Qu.:670.0	3rd Qu.:0.4400	3rd Qu.:0.9700	3rd Qu.: 1.580	3rd Qu.:0.3400	3rd Qu.: 1.150
Max. :784.0	Max. :1.4600	Max. :4.9500	Max. :27.200	Max. :3.8000	Max. :16.200

```
write.csv(summary(filtered_dataset_no_influential_no_outlier[c(
```