친환경

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

library(readxl)  
library(dplyr)

##   
## 다음의 패키지를 부착합니다: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)  
library(lmtest)

## 필요한 패키지를 로딩중입니다: zoo

##   
## 다음의 패키지를 부착합니다: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

options(scipen = 99)

# 

# 14년 01 ~ 20년 12월 까지의 기사량 데이터

news <- read.csv('amCharts.csv')  
head(news)

## date data  
## 1 2014-01-01 176  
## 2 2014-02-01 155  
## 3 2014-03-01 196  
## 4 2014-04-01 156  
## 5 2014-05-01 97  
## 6 2014-06-01 167

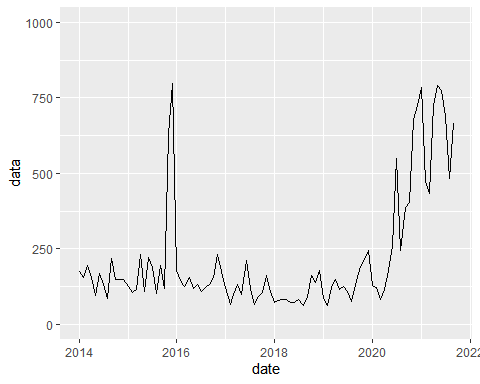
news$date <- as.Date(news$date)  
str(news)

## 'data.frame': 93 obs. of 2 variables:  
## $ date: Date, format: "2014-01-01" "2014-02-01" ...  
## $ data: int 176 155 196 156 97 167 133 86 217 149 ...

# 

# 시계열 그래프로 기사량 동향 확인

ggplot(data = news, aes (x = date, y = data, group =1)) +  
 geom\_line() +  
 ylim(0, 1000)

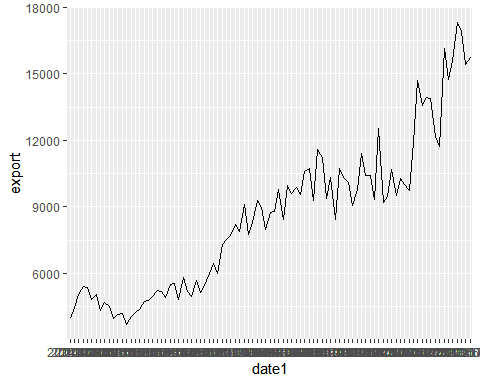


# 14 ~ 21 년 9월 리튬이차전지 수출량 hscode 850760

lit\_export <- read\_excel('리튬 월별 수출입 실적.xlsx')  
lit\_export

## # A tibble: 93 x 8  
## 기간 품목명 품목코드 수출중량 수입중량 수출금액 수입금액 무역수지  
## <chr> <chr> <chr> <dbl> <chr> <chr> <chr> <chr>   
## 1 2014.01 리튬이온 축전지 850760 3982. 577.9 169,350 37,173 132,177   
## 2 2014.02 리튬이온 축전지 850760 4486. 456.6 183,462 33,656 149,806   
## 3 2014.03 리튬이온 축전지 850760 5076. 619.6 208,952 44,355 164,597   
## 4 2014.04 리튬이온 축전지 850760 5414. 517.2 222,229 37,707 184,522   
## 5 2014.05 리튬이온 축전지 850760 5375. 517.0 209,711 39,074 170,637   
## 6 2014.06 리튬이온 축전지 850760 4813. 540.8 191,007 35,859 155,148   
## 7 2014.07 리튬이온 축전지 850760 5039 534.9 197,550 35,947 161,603   
## 8 2014.08 리튬이온 축전지 850760 4314. 457.0 172,434 31,898 140,536   
## 9 2014.09 리튬이온 축전지 850760 4689. 590.2 192,712 43,686 149,026   
## 10 2014.10 리튬이온 축전지 850760 4497. 680.3 190,306 47,829 142,477   
## # ... with 83 more rows

lit\_export <- rename(lit\_export,  
 date1 = '기간',  
 export = '수출중량')  
ggplot(data = lit\_export, aes (x = date1, y = export, group = 1)) +  
 geom\_line()

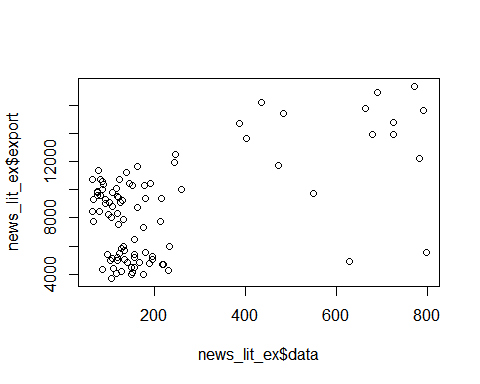


# 기사량 데이터와 리튬이차전지 수출량 의 관계

news\_lit\_ex <- cbind(news,lit\_export)  
news\_lit\_ex <- news\_lit\_ex %>%   
 select(date,data,export)  
head(news\_lit\_ex)

## date data export  
## 1 2014-01-01 176 3982.3  
## 2 2014-02-01 155 4485.5  
## 3 2014-03-01 196 5075.7  
## 4 2014-04-01 156 5413.8  
## 5 2014-05-01 97 5374.7  
## 6 2014-06-01 167 4812.8

# 산점도   
plot(news\_lit\_ex$data, news\_lit\_ex$export)



# 상관관계   
cor(news\_lit\_ex$data, news\_lit\_ex$export)

## [1] 0.5385346

cor.test(news\_lit\_ex$data, news\_lit\_ex$export)

##   
## Pearson's product-moment correlation  
##   
## data: news\_lit\_ex$data and news\_lit\_ex$export  
## t = 6.0969, df = 91, p-value = 0.00000002576  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.3760842 0.6688656  
## sample estimates:  
## cor   
## 0.5385346

# 회귀분석

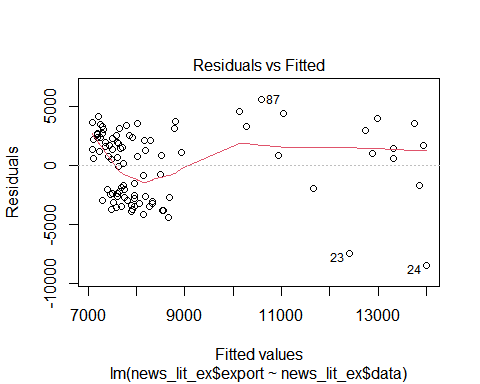
result<- lm(data = news\_lit\_ex, news\_lit\_ex$export ~ news\_lit\_ex$data)  
summary(result)

##   
## Call:  
## lm(formula = news\_lit\_ex$export ~ news\_lit\_ex$data, data = news\_lit\_ex)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8474.7 -2640.8 781.8 2364.3 5578.3   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6494.867 456.862 14.216 < 0.0000000000000002 \*\*\*  
## news\_lit\_ex$data 9.403 1.542 6.097 0.0000000258 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2955 on 91 degrees of freedom  
## Multiple R-squared: 0.29, Adjusted R-squared: 0.2822   
## F-statistic: 37.17 on 1 and 91 DF, p-value: 0.00000002576

# 회귀분석 가정 충족

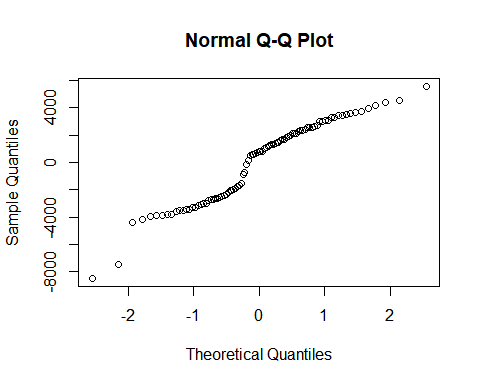
# 등분산성

plot(result, which = 1)



# 정규성 검정

res <- residuals(result)  
qqnorm(res)

 # 독립성 검정

dwtest(result)

##   
## Durbin-Watson test  
##   
## data: result  
## DW = 0.31653, p-value < 0.00000000000000022  
## alternative hypothesis: true autocorrelation is greater than 0

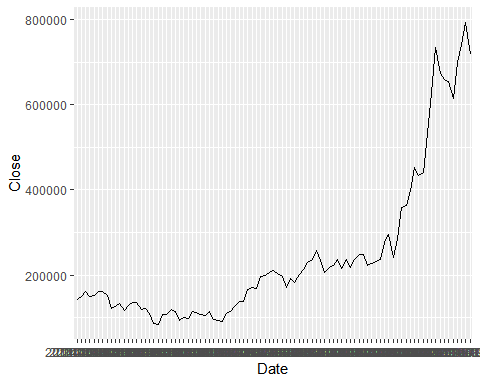
# 리튬이차전지 주요 3사의 주가 데이터 비교

# 삼성 sdi

sam <- read.csv('삼성SDI.KS.csv')  
str(sam)

## 'data.frame': 93 obs. of 7 variables:  
## $ Date : chr "2014-01-01" "2014-02-01" "2014-03-01" "2014-04-01" ...  
## $ Open : num 162000 142500 147500 161500 151000 ...  
## $ High : num 164500 160000 170000 163000 159000 ...  
## $ Low : num 142000 136000 147000 148000 143500 ...  
## $ Close : num 143500 149500 161000 151000 151500 ...  
## $ Adj.Close: num 137521 143271 154292 144709 145188 ...  
## $ Volume : int 5038077 4379955 5741312 5431826 4185254 11076320 8154503 5966175 7062693 9963012 ...

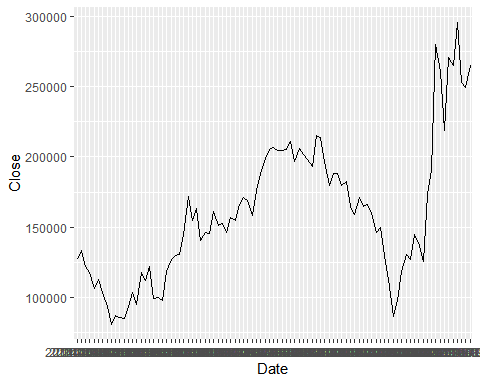
ggplot(data = sam, aes (x = Date, y = Close, group = 1)) +  
 geom\_line()

 # sk 이노베이션

sk <- read.csv('SK이노.KS.csv')  
str(sk)

## 'data.frame': 93 obs. of 7 variables:  
## $ Date : chr "2014-01-01" "2014-02-01" "2014-03-01" "2014-04-01" ...  
## $ Open : num 141500 125500 133000 122000 118000 ...  
## $ High : num 143500 137500 134500 130000 118000 ...  
## $ Low : num 127000 124500 117000 117000 106000 ...  
## $ Close : num 127000 133500 122000 118000 106500 ...  
## $ Adj.Close: num 105600 111004 101442 98116 88554 ...  
## $ Volume : int 6018611 5736103 6321598 6214858 7454252 11386608 10870199 9588768 7874282 11838373 ...

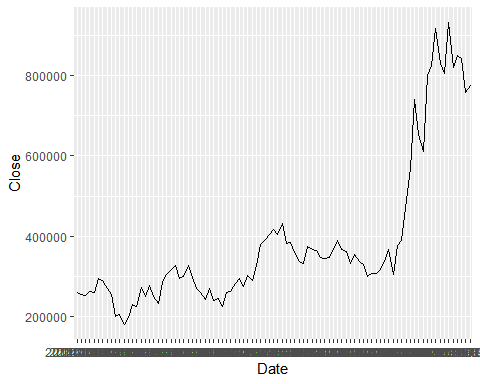
ggplot(data = sk, aes (x = Date, y = Close, group = 1)) +  
 geom\_line()

 # lg 화학

lg <- read.csv('LG화학.KS.csv')  
str(lg)

## 'data.frame': 93 obs. of 7 variables:  
## $ Date : chr "2014-01-01" "2014-02-01" "2014-03-01" "2014-04-01" ...  
## $ Open : num 299500 256000 253500 254000 263000 ...  
## $ High : num 303000 261000 258500 278500 269500 ...  
## $ Low : num 256000 246000 235500 254000 249500 ...  
## $ Close : num 260000 254500 254000 263000 260000 ...  
## $ Adj.Close: num 233905 228957 228507 236604 233905 ...  
## $ Volume : int 4274525 3775592 4576331 5086182 2891374 4207572 4841389 3172903 3407755 10178655 ...

ggplot(data = lg, aes (x = Date, y = Close, group = 1)) +  
 geom\_line()

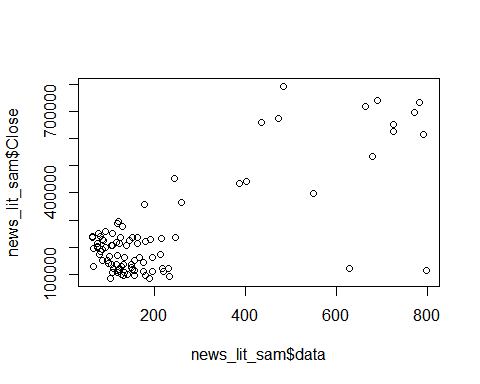
 # 기사 데이터와 주가데이터의 상관관계 # 삼성SDI

news\_lit\_sam <- cbind(news, sam)  
news\_lit\_sam <- news\_lit\_sam %>%   
 select(date, data, Close)  
head(news\_lit\_sam)

## date data Close  
## 1 2014-01-01 176 143500  
## 2 2014-02-01 155 149500  
## 3 2014-03-01 196 161000  
## 4 2014-04-01 156 151000  
## 5 2014-05-01 97 151500  
## 6 2014-06-01 167 162000

# 산점도

plot(news\_lit\_sam$data, news\_lit\_sam$Close)

 # 상관관계

cor(news\_lit\_sam$data, news\_lit\_sam$Close)

## [1] 0.7554396

cor.test(news\_lit\_sam$data, news\_lit\_sam$Close)

##   
## Pearson's product-moment correlation  
##   
## data: news\_lit\_sam$data and news\_lit\_sam$Close  
## t = 10.999, df = 91, p-value < 0.00000000000000022  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.6520794 0.8312304  
## sample estimates:  
## cor   
## 0.7554396

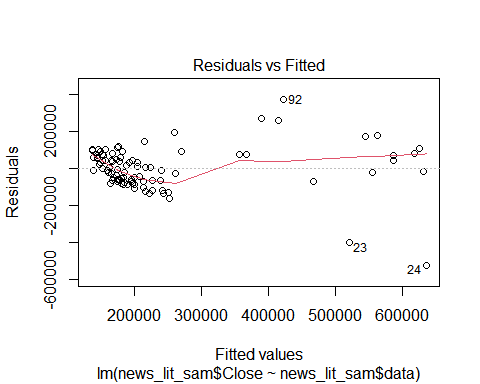
# 회귀분석

result2<- lm(data = news\_lit\_sam, news\_lit\_sam$Close ~ news\_lit\_sam$data)  
summary(result2)

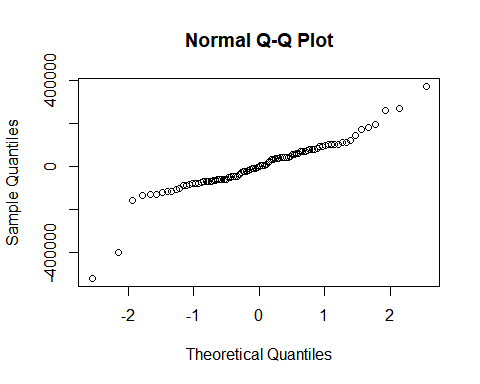
##   
## Call:  
## lm(formula = news\_lit\_sam$Close ~ news\_lit\_sam$data, data = news\_lit\_sam)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -522067 -67057 -2004 69307 370426   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 95216.64 18254.43 5.216 0.00000114 \*\*\*  
## news\_lit\_sam$data 677.76 61.62 10.999 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 118100 on 91 degrees of freedom  
## Multiple R-squared: 0.5707, Adjusted R-squared: 0.566   
## F-statistic: 121 on 1 and 91 DF, p-value: < 0.00000000000000022

# 회귀분석 가정 충족

plot(result2, which = 1)



res2 <- residuals(result2)  
qqnorm(res2)



dwtest(result2)

##   
## Durbin-Watson test  
##   
## data: result2  
## DW = 0.55728, p-value < 0.00000000000000022  
## alternative hypothesis: true autocorrelation is greater than 0

summary(result2)

##   
## Call:  
## lm(formula = news\_lit\_sam$Close ~ news\_lit\_sam$data, data = news\_lit\_sam)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -522067 -67057 -2004 69307 370426   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 95216.64 18254.43 5.216 0.00000114 \*\*\*  
## news\_lit\_sam$data 677.76 61.62 10.999 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 118100 on 91 degrees of freedom  
## Multiple R-squared: 0.5707, Adjusted R-squared: 0.566   
## F-statistic: 121 on 1 and 91 DF, p-value: < 0.00000000000000022

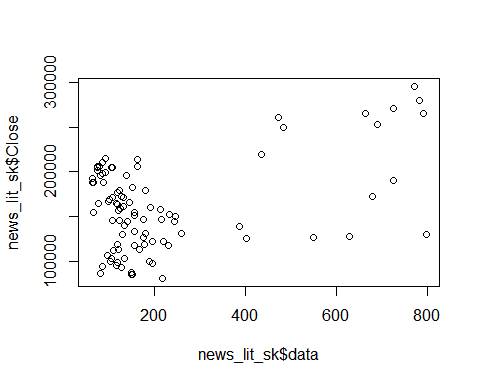
# sk

news\_lit\_sk <- cbind(news, sk)  
news\_lit\_sk <- news\_lit\_sk %>%   
 select(date, data, Close)  
head(news\_lit\_sk)

## date data Close  
## 1 2014-01-01 176 127000  
## 2 2014-02-01 155 133500  
## 3 2014-03-01 196 122000  
## 4 2014-04-01 156 118000  
## 5 2014-05-01 97 106500  
## 6 2014-06-01 167 113000

# 산점도

plot(news\_lit\_sk$data, news\_lit\_sk$Close)

 # 상관관계

cor(news\_lit\_sk$data, news\_lit\_sk$Close)

## [1] 0.4329965

cor.test(news\_lit\_sk$data, news\_lit\_sk$Close)

##   
## Pearson's product-moment correlation  
##   
## data: news\_lit\_sk$data and news\_lit\_sk$Close  
## t = 4.5824, df = 91, p-value = 0.00001458  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.2514690 0.5850963  
## sample estimates:  
## cor   
## 0.4329965

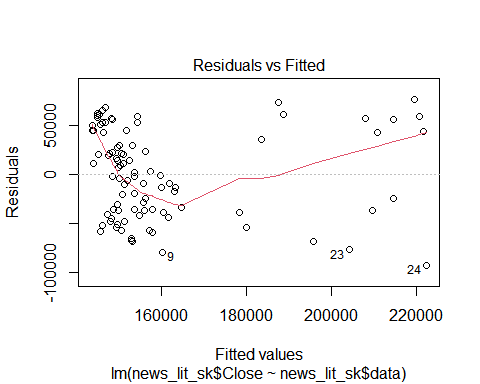
# 회귀분석

result3<- lm(data = news\_lit\_sk, news\_lit\_sk$Close ~ news\_lit\_sk$data)  
summary(result3)

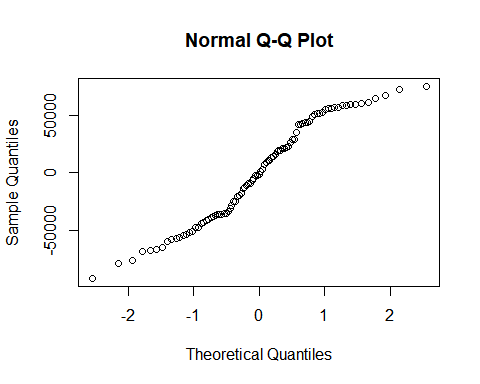
##   
## Call:  
## lm(formula = news\_lit\_sk$Close ~ news\_lit\_sk$data, data = news\_lit\_sk)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -92388 -37354 -1211 43255 75790   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 136890.13 6926.14 19.764 < 0.0000000000000002 \*\*\*  
## news\_lit\_sk$data 107.14 23.38 4.582 0.0000146 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 44790 on 91 degrees of freedom  
## Multiple R-squared: 0.1875, Adjusted R-squared: 0.1786   
## F-statistic: 21 on 1 and 91 DF, p-value: 0.00001458

# 회귀분석 가정 충족

plot(result3, which = 1)



res3 <- residuals(result3)  
qqnorm(res3)



dwtest(result3)

##   
## Durbin-Watson test  
##   
## data: result3  
## DW = 0.19552, p-value < 0.00000000000000022  
## alternative hypothesis: true autocorrelation is greater than 0

summary(result3)

##   
## Call:  
## lm(formula = news\_lit\_sk$Close ~ news\_lit\_sk$data, data = news\_lit\_sk)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -92388 -37354 -1211 43255 75790   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 136890.13 6926.14 19.764 < 0.0000000000000002 \*\*\*  
## news\_lit\_sk$data 107.14 23.38 4.582 0.0000146 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 44790 on 91 degrees of freedom  
## Multiple R-squared: 0.1875, Adjusted R-squared: 0.1786   
## F-statistic: 21 on 1 and 91 DF, p-value: 0.00001458

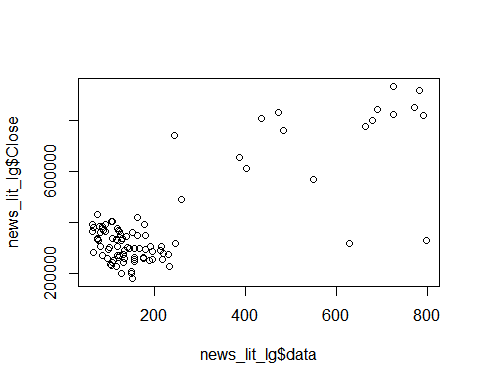
# lg

news\_lit\_lg <- cbind(news, lg)  
news\_lit\_lg <- news\_lit\_lg %>%   
 select(date, data, Close)  
head(news\_lit\_lg)

## date data Close  
## 1 2014-01-01 176 260000  
## 2 2014-02-01 155 254500  
## 3 2014-03-01 196 254000  
## 4 2014-04-01 156 263000  
## 5 2014-05-01 97 260000  
## 6 2014-06-01 167 296000

# 산점도

plot(news\_lit\_lg$data, news\_lit\_lg$Close)

 # 상관관계

cor(news\_lit\_lg$data, news\_lit\_lg$Close)

## [1] 0.7894346

cor.test(news\_lit\_lg$data, news\_lit\_lg$Close)

##   
## Pearson's product-moment correlation  
##   
## data: news\_lit\_lg$data and news\_lit\_lg$Close  
## t = 12.268, df = 91, p-value < 0.00000000000000022  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.6979699 0.8555570  
## sample estimates:  
## cor   
## 0.7894346

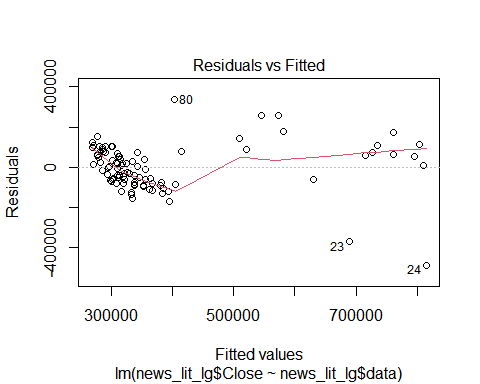
# 회귀분석

result4<- lm(data = news\_lit\_lg, news\_lit\_lg$Close ~ news\_lit\_lg$data)  
summary(result4)

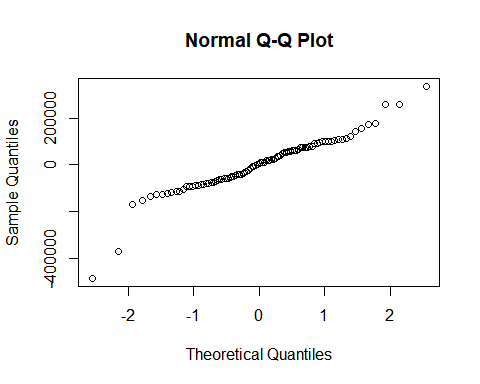
##   
## Call:  
## lm(formula = news\_lit\_lg$Close ~ news\_lit\_lg$data, data = news\_lit\_lg)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -486235 -70005 4961 74395 336159   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 222870.02 17908.65 12.45 <0.0000000000000002 \*\*\*  
## news\_lit\_lg$data 741.69 60.46 12.27 <0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 115800 on 91 degrees of freedom  
## Multiple R-squared: 0.6232, Adjusted R-squared: 0.6191   
## F-statistic: 150.5 on 1 and 91 DF, p-value: < 0.00000000000000022

# 회귀분석 가정 충족

plot(result4, which = 1)



res4 <- residuals(result4)  
qqnorm(res4)



dwtest(result4)

##   
## Durbin-Watson test  
##   
## data: result4  
## DW = 0.63992, p-value = 0.000000000000005855  
## alternative hypothesis: true autocorrelation is greater than 0

summary(result4)

##   
## Call:  
## lm(formula = news\_lit\_lg$Close ~ news\_lit\_lg$data, data = news\_lit\_lg)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -486235 -70005 4961 74395 336159   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 222870.02 17908.65 12.45 <0.0000000000000002 \*\*\*  
## news\_lit\_lg$data 741.69 60.46 12.27 <0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 115800 on 91 degrees of freedom  
## Multiple R-squared: 0.6232, Adjusted R-squared: 0.6191   
## F-statistic: 150.5 on 1 and 91 DF, p-value: < 0.00000000000000022

# 리튬의 원자재 가격 비교

lit\_price <- read\_excel('리튬 가격.xls')  
head(lit\_price)

## # A tibble: 6 x 5  
## 기준일 기준가격 등락가 등락비율 재고량  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 2014-01 36 0 0 0  
## 2 2014-02 36 0 0 0  
## 3 2014-03 36 0 0 0  
## 4 2014-04 36 0 0 0  
## 5 2014-05 36 0 0 0  
## 6 2014-06 35.4 -0.6 -1.67 0

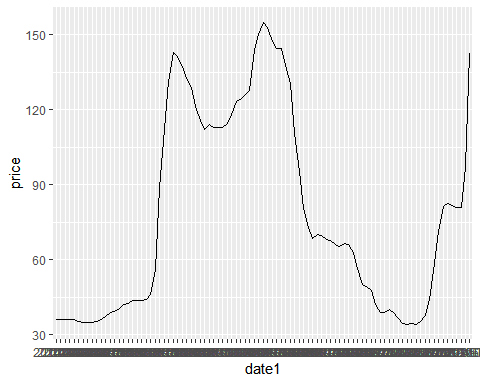
str(lit\_price)

## tibble [93 x 5] (S3: tbl\_df/tbl/data.frame)  
## $ 기준일 : chr [1:93] "2014-01" "2014-02" "2014-03" "2014-04" ...  
## $ 기준가격: num [1:93] 36 36 36 36 36 35.4 35 35 35 35.4 ...  
## $ 등락가 : num [1:93] 0 0 0 0 0 -0.6 -0.4 0 0 0.4 ...  
## $ 등락비율: num [1:93] 0 0 0 0 0 -1.67 -1.13 0 0 1.14 ...  
## $ 재고량 : num [1:93] 0 0 0 0 0 0 0 0 0 0 ...

lit\_price <- rename(lit\_price,  
 price = '기준가격',  
 date1 = '기준일')

# 시계열 그래프

ggplot(data = lit\_price, aes (x = date1, y = price ,group =1)) +  
 geom\_line()

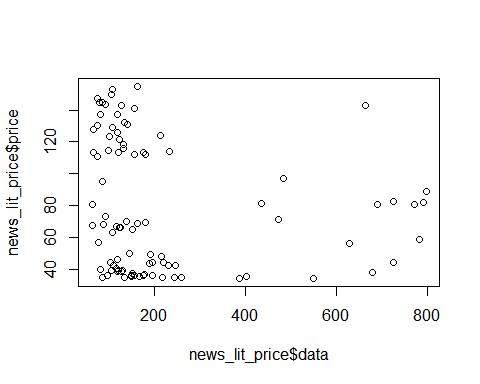
 # 기사량 데이터와 리튬 원자재가의 관계

news\_lit\_price <- cbind(news,lit\_price)  
news\_lit\_price <- news\_lit\_price %>%   
 select(date, data, price)  
head(news\_lit\_price)

## date data price  
## 1 2014-01-01 176 36.0  
## 2 2014-02-01 155 36.0  
## 3 2014-03-01 196 36.0  
## 4 2014-04-01 156 36.0  
## 5 2014-05-01 97 36.0  
## 6 2014-06-01 167 35.4

# 산점도

plot(news\_lit\_price$data, news\_lit\_price$price)

 # 상관관계

cor(news\_lit\_price$data, news\_lit\_price$price)

## [1] -0.1452915

cor.test(news\_lit\_price$data, news\_lit\_price$price)

##   
## Pearson's product-moment correlation  
##   
## data: news\_lit\_price$data and news\_lit\_price$price  
## t = -1.4009, df = 91, p-value = 0.1647  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.33896729 0.06019852  
## sample estimates:  
## cor   
## -0.1452915

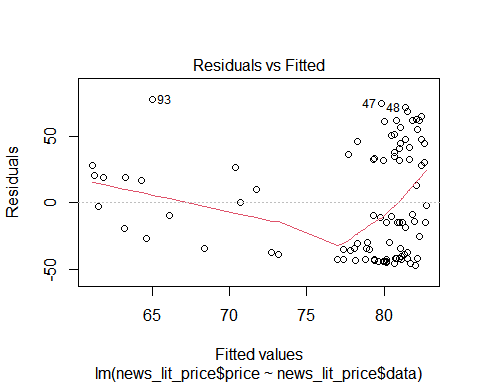
# 회귀분석

result5<- lm(data = news\_lit\_price, news\_lit\_price$price ~ news\_lit\_price$data)  
summary(result5)

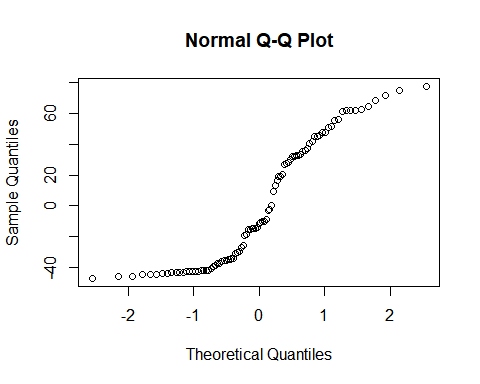
##   
## Call:  
## lm(formula = news\_lit\_price$price ~ news\_lit\_price$data, data = news\_lit\_price)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -47.02 -38.81 -11.40 35.12 77.61   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 84.54766 6.21374 13.607 <0.0000000000000002 \*\*\*  
## news\_lit\_price$data -0.02938 0.02098 -1.401 0.165   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 40.19 on 91 degrees of freedom  
## Multiple R-squared: 0.02111, Adjusted R-squared: 0.01035   
## F-statistic: 1.962 on 1 and 91 DF, p-value: 0.1647

# 회귀분석 가정 충족

plot(result5, which = 1)



res5 <- residuals(result5)  
qqnorm(res5)



dwtest(result5)

##   
## Durbin-Watson test  
##   
## data: result5  
## DW = 0.054231, p-value < 0.00000000000000022  
## alternative hypothesis: true autocorrelation is greater than 0

summary(result5)

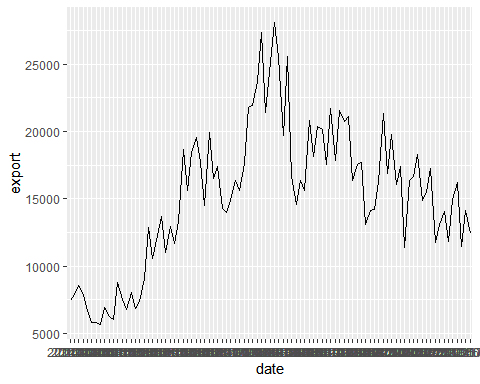
##   
## Call:  
## lm(formula = news\_lit\_price$price ~ news\_lit\_price$data, data = news\_lit\_price)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -47.02 -38.81 -11.40 35.12 77.61   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 84.54766 6.21374 13.607 <0.0000000000000002 \*\*\*  
## news\_lit\_price$data -0.02938 0.02098 -1.401 0.165   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 40.19 on 91 degrees of freedom  
## Multiple R-squared: 0.02111, Adjusted R-squared: 0.01035   
## F-statistic: 1.962 on 1 and 91 DF, p-value: 0.1647

# 14 ~ 21 년 9월 태양광 모듈 + 전지 수출량 hscode

sun\_export <- read\_excel('태양광 수출량.xlsx')  
sun\_export

## # A tibble: 93 x 4  
## 기간 품목명 품목코드 수출중량  
## <chr> <chr> <chr> <dbl>  
## 1 2014.01 감광성 반도체 디바이스(광전지는 모듈에 조립되었거나~ 854140 7446   
## 2 2014.02 감광성 반도체 디바이스(광전지는 모듈에 조립되었거나~ 854140 7866.  
## 3 2014.03 감광성 반도체 디바이스(광전지는 모듈에 조립되었거나~ 854140 8569.  
## 4 2014.04 감광성 반도체 디바이스(광전지는 모듈에 조립되었거나~ 854140 7904.  
## 5 2014.05 감광성 반도체 디바이스(광전지는 모듈에 조립되었거나~ 854140 6678.  
## 6 2014.06 감광성 반도체 디바이스(광전지는 모듈에 조립되었거나~ 854140 5822   
## 7 2014.07 감광성 반도체 디바이스(광전지는 모듈에 조립되었거나~ 854140 5795.  
## 8 2014.08 감광성 반도체 디바이스(광전지는 모듈에 조립되었거나~ 854140 5660.  
## 9 2014.09 감광성 반도체 디바이스(광전지는 모듈에 조립되었거나~ 854140 6882.  
## 10 2014.10 감광성 반도체 디바이스(광전지는 모듈에 조립되었거나~ 854140 6226.  
## # ... with 83 more rows

sun\_export <- rename(sun\_export,  
 date = '기간',  
 export = '수출중량')  
ggplot(data = sun\_export, aes (x = date, y = export, group = 1)) +  
 geom\_line()

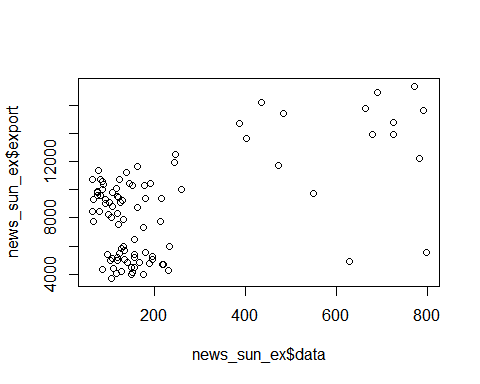
 # 기사량 데이터와 태양전지 수출량 의 관계

news\_sun\_ex <- cbind(news,sun\_export)  
news\_sun\_ex <- news\_lit\_ex %>%   
 select(date, data, export)  
head(news\_sun\_ex)

## date data export  
## 1 2014-01-01 176 3982.3  
## 2 2014-02-01 155 4485.5  
## 3 2014-03-01 196 5075.7  
## 4 2014-04-01 156 5413.8  
## 5 2014-05-01 97 5374.7  
## 6 2014-06-01 167 4812.8

# 산점도

plot(news\_sun\_ex$data, news\_sun\_ex$export)

 # 상관관계

cor(news\_sun\_ex$data, news\_sun\_ex$export)

## [1] 0.5385346

cor.test(news\_sun\_ex$data, news\_sun\_ex$export)

##   
## Pearson's product-moment correlation  
##   
## data: news\_sun\_ex$data and news\_sun\_ex$export  
## t = 6.0969, df = 91, p-value = 0.00000002576  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.3760842 0.6688656  
## sample estimates:  
## cor   
## 0.5385346

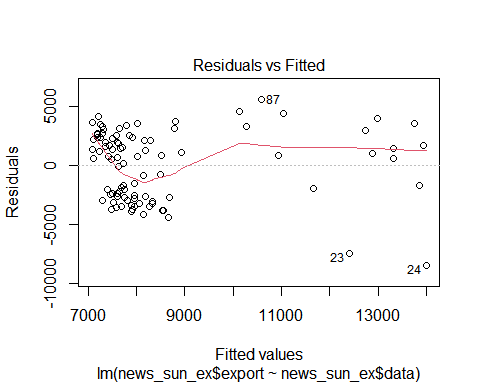
# 회귀분석

result6<- lm(data = news\_sun\_ex, news\_sun\_ex$export ~ news\_sun\_ex$data)  
summary(result6)

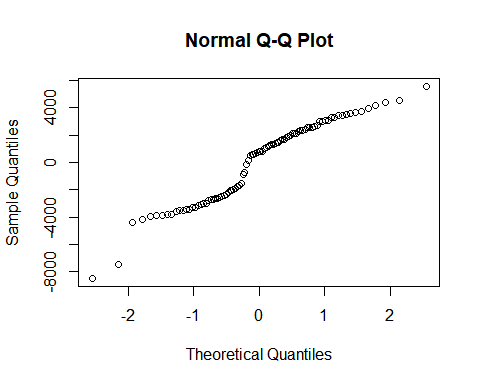
##   
## Call:  
## lm(formula = news\_sun\_ex$export ~ news\_sun\_ex$data, data = news\_sun\_ex)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8474.7 -2640.8 781.8 2364.3 5578.3   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6494.867 456.862 14.216 < 0.0000000000000002 \*\*\*  
## news\_sun\_ex$data 9.403 1.542 6.097 0.0000000258 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2955 on 91 degrees of freedom  
## Multiple R-squared: 0.29, Adjusted R-squared: 0.2822   
## F-statistic: 37.17 on 1 and 91 DF, p-value: 0.00000002576

# 회귀분석 가정 충족

plot(result6, which = 1)



res6 <- residuals(result6)  
qqnorm(res6)



dwtest(result6)

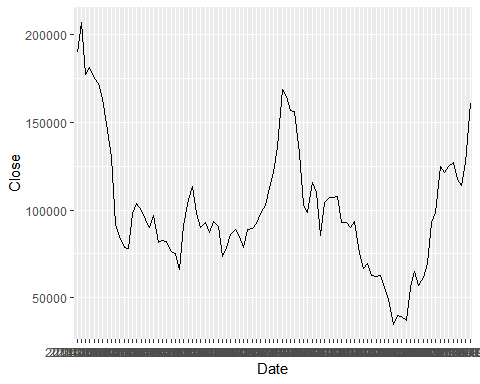
##   
## Durbin-Watson test  
##   
## data: result6  
## DW = 0.31653, p-value < 0.00000000000000022  
## alternative hypothesis: true autocorrelation is greater than 0

# 태양광전지 주요 2사의 주가 데이터 비교

oci <- read.csv('OCI.KS.csv')  
str(oci)

## 'data.frame': 93 obs. of 7 variables:  
## $ Date : chr "2014-01-01" "2014-02-01" "2014-03-01" "2014-04-01" ...  
## $ Open : num 191000 186500 205000 178500 181000 ...  
## $ High : num 207500 214500 215000 186500 183500 ...  
## $ Low : num 184500 181000 175000 171500 164500 ...  
## $ Close : num 190000 207000 177500 181000 175000 ...  
## $ Adj.Close: num 184170 200649 172054 175446 169630 ...  
## $ Volume : int 4413858 3234930 5055190 3590555 2606618 2948899 3575072 3199815 3644207 10785915 ...

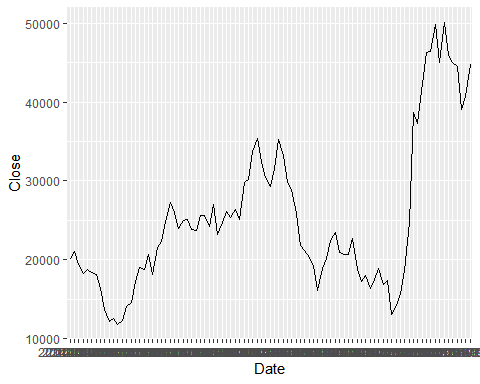
ggplot(data = oci, aes (x = Date, y = Close, group = 1)) +  
 geom\_line()



hanwha <- read.csv('한화솔루션.KS.csv')  
str(hanwha)

## 'data.frame': 93 obs. of 7 variables:  
## $ Date : chr "2014-01-01" "2014-02-01" "2014-03-01" "2014-04-01" ...  
## $ Open : num 21550 19650 20850 19500 18250 ...  
## $ High : num 22450 21600 21550 19950 19500 ...  
## $ Low : num 19700 18950 18550 17300 17350 ...  
## $ Close : num 20000 21000 19350 18250 18800 ...  
## $ Adj.Close: num 18793 19733 18183 17149 17666 ...  
## $ Volume : int 17148806 15880265 25105642 40025896 44092467 23657346 22655301 18011258 17205429 33170997 ...

ggplot(data = hanwha, aes (x = Date, y = Close, group = 1)) +  
 geom\_line()

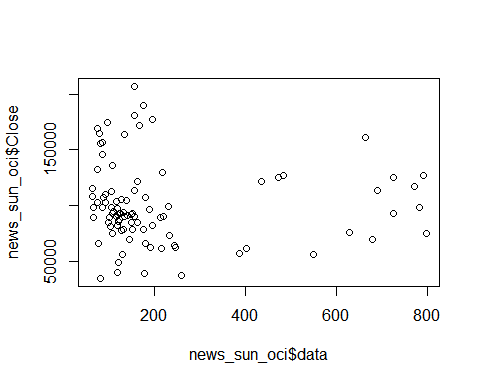
 # 기사 데이터와 주가데이터의 상관관계 # OCI

news\_sun\_oci <- cbind(news, oci)  
news\_sun\_oci <- news\_sun\_oci %>%   
 select(date, data, Close)  
head(news\_sun\_oci)

## date data Close  
## 1 2014-01-01 176 190000  
## 2 2014-02-01 155 207000  
## 3 2014-03-01 196 177500  
## 4 2014-04-01 156 181000  
## 5 2014-05-01 97 175000  
## 6 2014-06-01 167 172000

# 산점도

plot(news\_sun\_oci$data, news\_sun\_oci$Close)

 # 상관관계

cor(news\_sun\_oci$data, news\_sun\_oci$Close)

## [1] -0.02381614

cor.test(news\_sun\_oci$data, news\_sun\_oci$Close)

##   
## Pearson's product-moment correlation  
##   
## data: news\_sun\_oci$data and news\_sun\_oci$Close  
## t = -0.22726, df = 91, p-value = 0.8207  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.2264259 0.1807691  
## sample estimates:  
## cor   
## -0.02381614

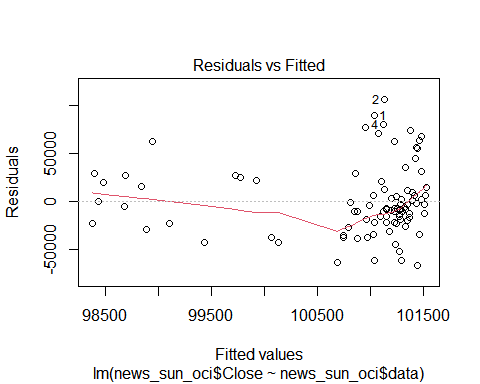
# 회귀분석

result7<- lm(data = news\_sun\_oci, news\_sun\_oci$Close ~ news\_sun\_oci$data)  
summary(result7)

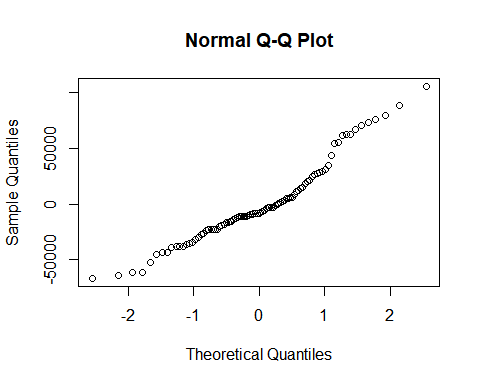
##   
## Call:  
## lm(formula = news\_sun\_oci$Close ~ news\_sun\_oci$data, data = news\_sun\_oci)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -66643 -22336 -7954 15166 105874   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 101789.867 5584.845 18.226 <0.0000000000000002 \*\*\*  
## news\_sun\_oci$data -4.284 18.853 -0.227 0.821   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 36120 on 91 degrees of freedom  
## Multiple R-squared: 0.0005672, Adjusted R-squared: -0.01042   
## F-statistic: 0.05165 on 1 and 91 DF, p-value: 0.8207

# 회귀분석 가정 충족

plot(result7, which = 1)



res7 <- residuals(result7)  
qqnorm(res7)



dwtest(result7)

##   
## Durbin-Watson test  
##   
## data: result7  
## DW = 0.12485, p-value < 0.00000000000000022  
## alternative hypothesis: true autocorrelation is greater than 0

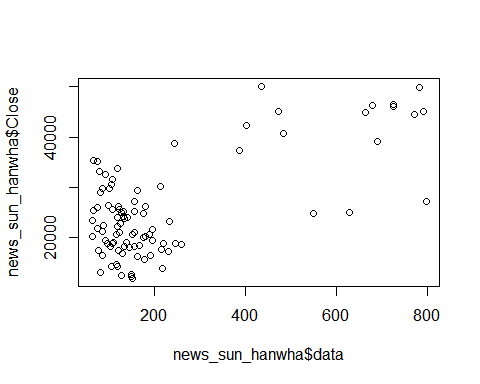
# 한화솔루션

news\_sun\_hanwha <- cbind(news, hanwha)  
news\_sun\_hanwha <- news\_sun\_hanwha %>%   
 select(date, data, Close)  
head(news\_sun\_hanwha)

## date data Close  
## 1 2014-01-01 176 20000  
## 2 2014-02-01 155 21000  
## 3 2014-03-01 196 19350  
## 4 2014-04-01 156 18250  
## 5 2014-05-01 97 18800  
## 6 2014-06-01 167 18350

# 산점도

plot(news\_sun\_hanwha$data, news\_sun\_hanwha$Close)

 # 상관관계

cor(news\_sun\_hanwha$data, news\_sun\_hanwha$Close)

## [1] 0.671575

cor.test(news\_sun\_hanwha$data, news\_sun\_hanwha$Close)

##   
## Pearson's product-moment correlation  
##   
## data: news\_sun\_hanwha$data and news\_sun\_hanwha$Close  
## t = 8.6464, df = 91, p-value = 0.0000000000001733  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.5420177 0.7699500  
## sample estimates:  
## cor   
## 0.671575

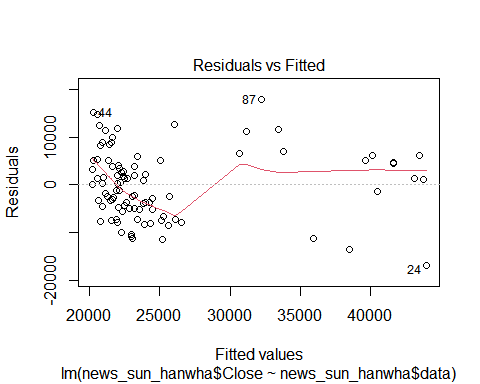
# 회귀분석

result8<- lm(data = news\_sun\_hanwha, news\_sun\_hanwha$Close ~ news\_sun\_hanwha$data)  
summary(result8)

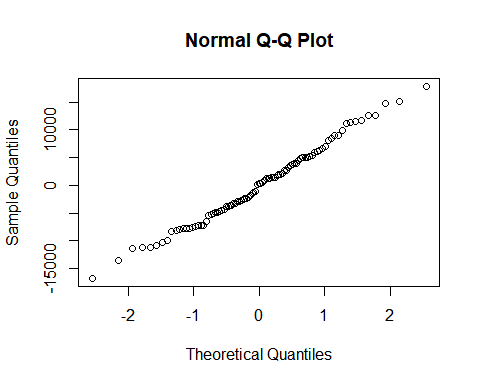
##   
## Call:  
## lm(formula = news\_sun\_hanwha$Close ~ news\_sun\_hanwha$data, data = news\_sun\_hanwha)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -16800.1 -4953.9 244.7 5042.5 17858.5   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 18150.579 1109.796 16.355 < 0.0000000000000002 \*\*\*  
## news\_sun\_hanwha$data 32.393 3.746 8.646 0.000000000000173 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7177 on 91 degrees of freedom  
## Multiple R-squared: 0.451, Adjusted R-squared: 0.445   
## F-statistic: 74.76 on 1 and 91 DF, p-value: 0.0000000000001733

# 회귀분석 가정 충족

plot(result8, which = 1)



res8 <- residuals(result8)  
qqnorm(res8)



dwtest(result8)

##   
## Durbin-Watson test  
##   
## data: result8  
## DW = 0.45965, p-value < 0.00000000000000022  
## alternative hypothesis: true autocorrelation is greater than 0

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.