In [83]:

```
library(ggplot2)
```

1. Loading Data

In [84]:

```
supercon_data <- read.csv('dataset/train.csv')
```

In [85]:

```
head(supercon_data,5)
```

number_of_elements	mean_atomic_mass	wtd_mean_atomic_mass	gmean_atomic_mass	wtd_gm
4	88.94447	57.86269	66.36159	
5	92.72921	58.51842	73.13279	
4	88.94447	57.88524	66.36159	
4	88.94447	57.87397	66.36159	
4	88.94447	57.84014	66.36159	

In [86]:

```
dim(supercon_data)
```

21263 82

In [87]:

```
sum(is.na(supercon_data))
```

0

In [88]:

```
table(supercon_data$number_of_elements)
```

```
   1    2    3    4    5    6    7    8    9 
 285 3280 3895 4496 5792 2666  774   61   14 
```

In [89]:

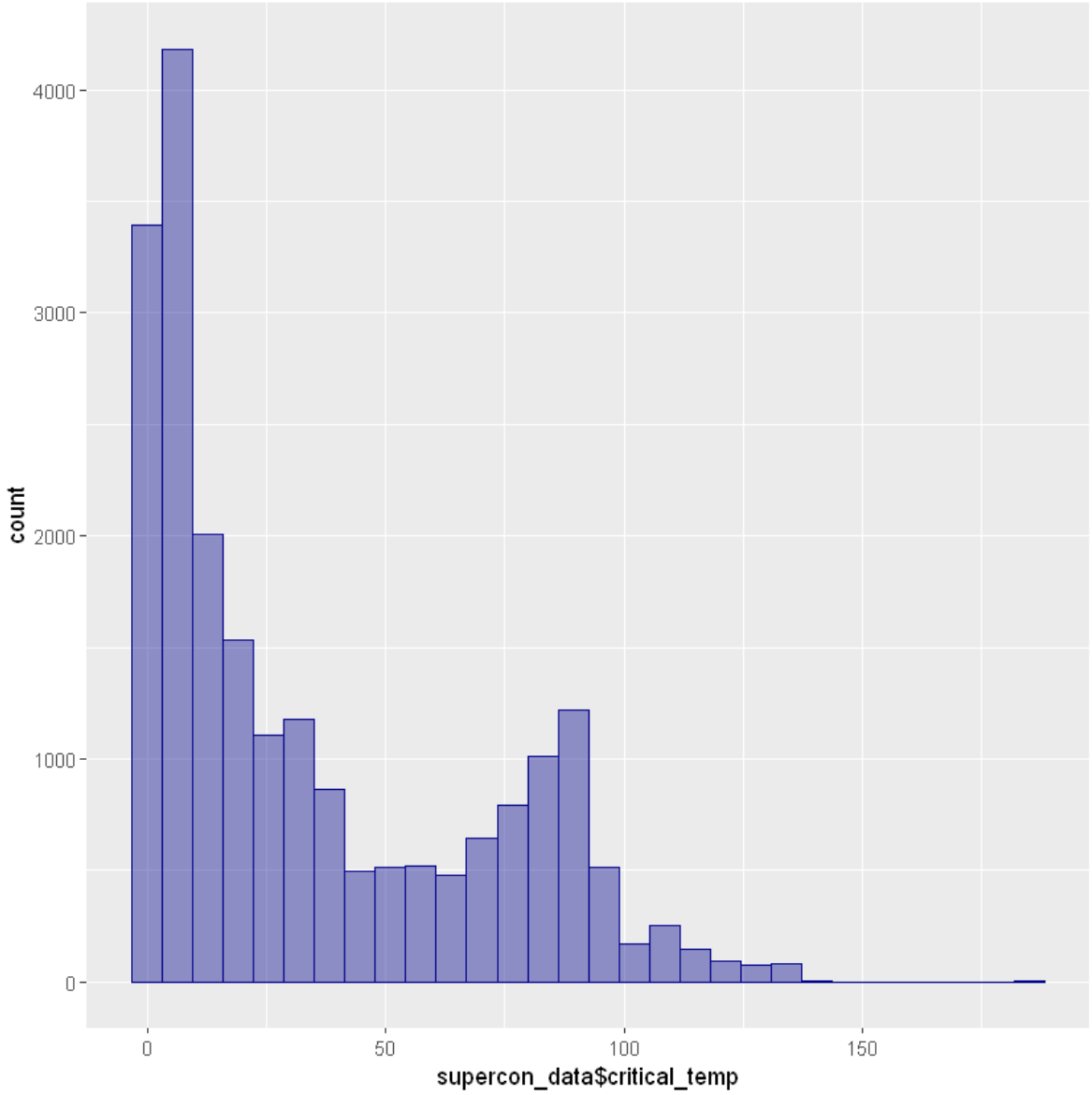
```
table(supercon_data$range_Valence)
```

```
   0    1    2    3    4    5    6 
1398 7387 5131 4335 2321  686    5 
```

In [91]:

```
ggplot(data=supercon_data, aes(supercon_data$critical_temp)) +
  geom_histogram(col="darkblue",
                 fill="darkblue",
                 alpha=.4)
```

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



In [92]:

```
colNames <- names(supercon_data)
```

In [42]:

```
for (i in colNames){
    ggplot(supercon_data, aes_string(x=i,y='critical_temp'))+geom_point()
    ggsave(paste("plots/",i,".jpg",sep=""))
}
```

...

2. Data Cleaning

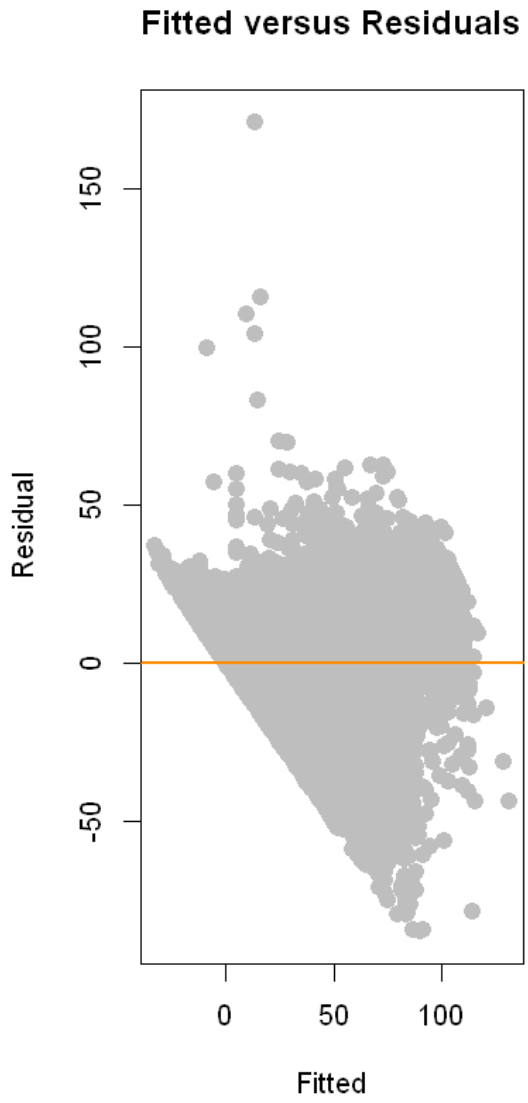
2.1.1. Leverages, Outliers and Influence

In [93]:

```
lev_fit <- lm(critical_temp ~ ., data = supercon_data)
# Use function "hatvalues" for the leverages from the lm object
leverages = hatvalues(lev_fit)
max_i <- which.max(leverages)
```

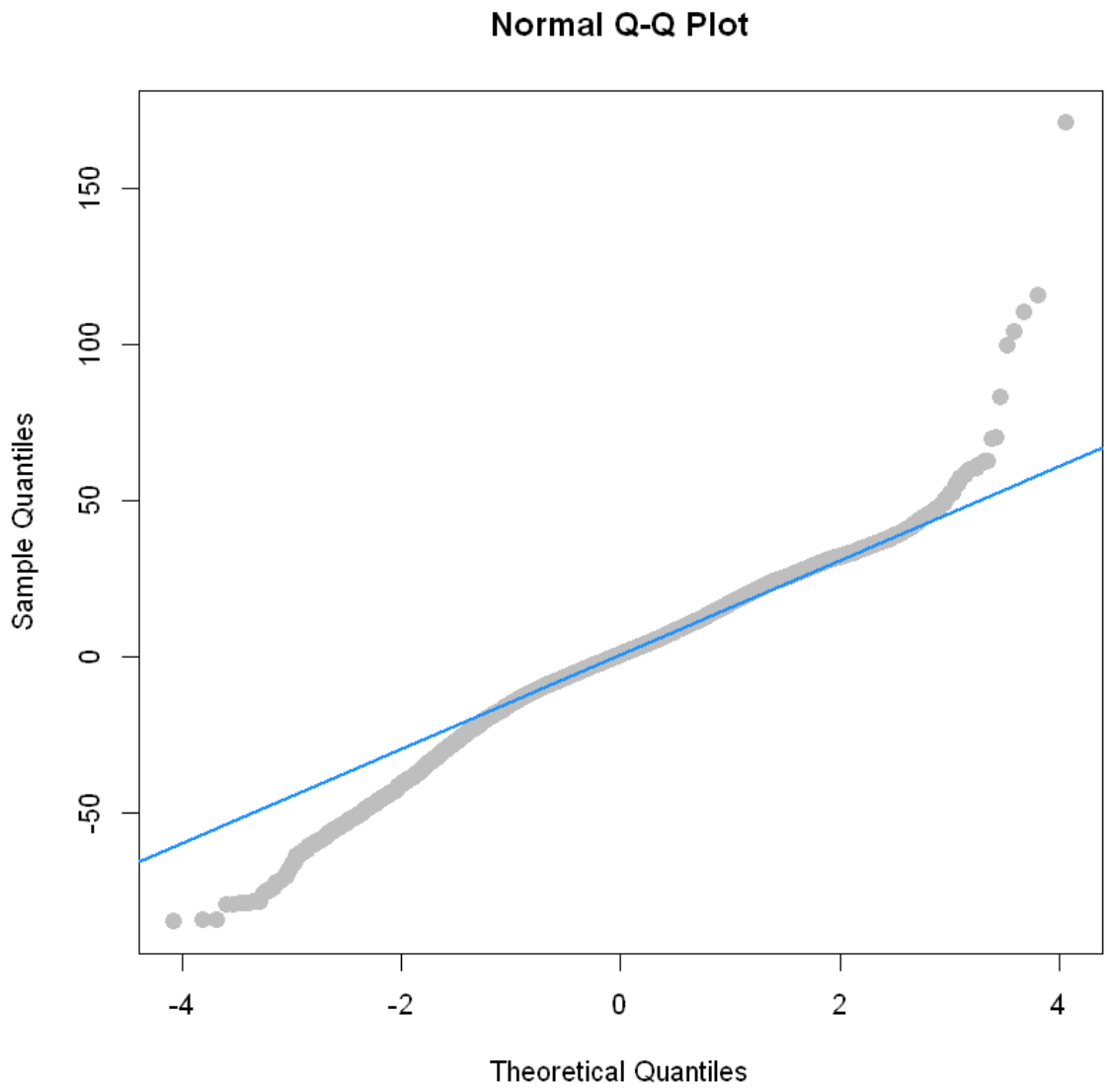
In [94]:

```
par(mfrow=c(1,2))
plot(fitted(lev_fit), resid(lev_fit), col = "grey", pch = 20,
     xlab = "Fitted", ylab = "Residual",cex=2,
     main = "Fitted versus Residuals")
abline(h = 0, col = "darkorange", lwd = 2)
```



In [95]:

```
qqnorm(resid(lev_fit), col = "grey",pch=20,cex=2)
qqline(resid(lev_fit), col = "dodgerblue", lwd = 2)
```



In [96]:

```
lev_fit_cd <- cooks.distance(lev_fit)
```

In [97]:

```
sum(lev_fit_cd > 4/length(lev_fit_cd))
```

792

In [98]:

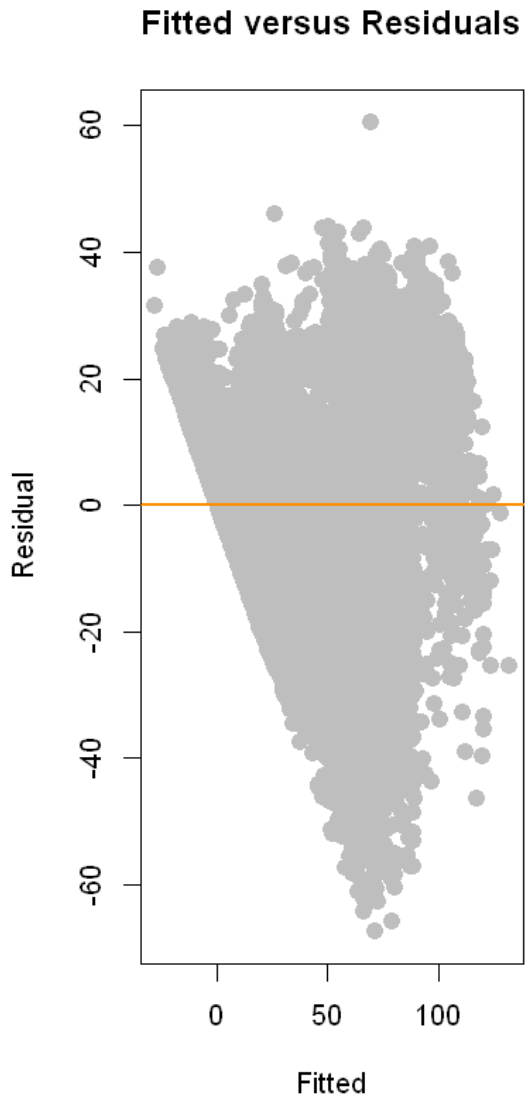
```
inf_i = which(lev_fit_cd > 4/length(lev_fit_cd))
supercon_data_ou_tr = supercon_data[-inf_i,]
```

In [99]:

```
lev_fit_2 = lm(critical_temp ~ ., data = supercon_data_ou_tr)
```

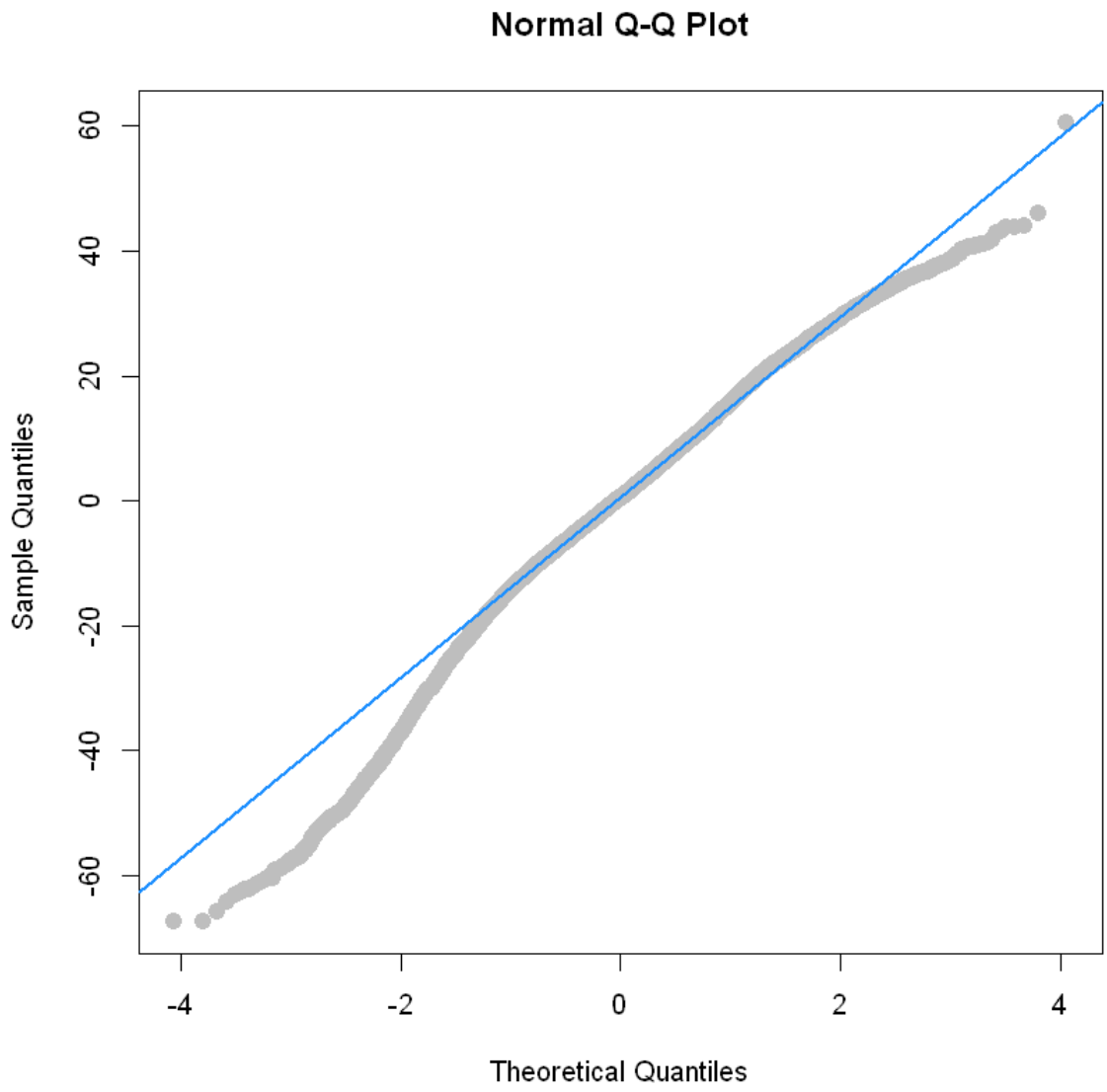
In [100]:

```
par(mfrow=c(1,2))
plot(fitted(lev_fit_2), resid(lev_fit_2), col = "grey", pch = 20,
     xlab = "Fitted", ylab = "Residual",cex=2,
     main = "Fitted versus Residuals")
abline(h = 0, col = "darkorange", lwd = 2)
```



In [101]:

```
qqnorm(resid(lev_fit_2), col = "grey",pch=20,cex=2)
qqline(resid(lev_fit_2), col = "dodgerblue", lwd = 2)
```



Removing High leverage points has improved the residual and normality assumptions

In [102]:

```
supercon_data = supercon_data_ou_tr
```

3. Correlation Analysis

In [103]:

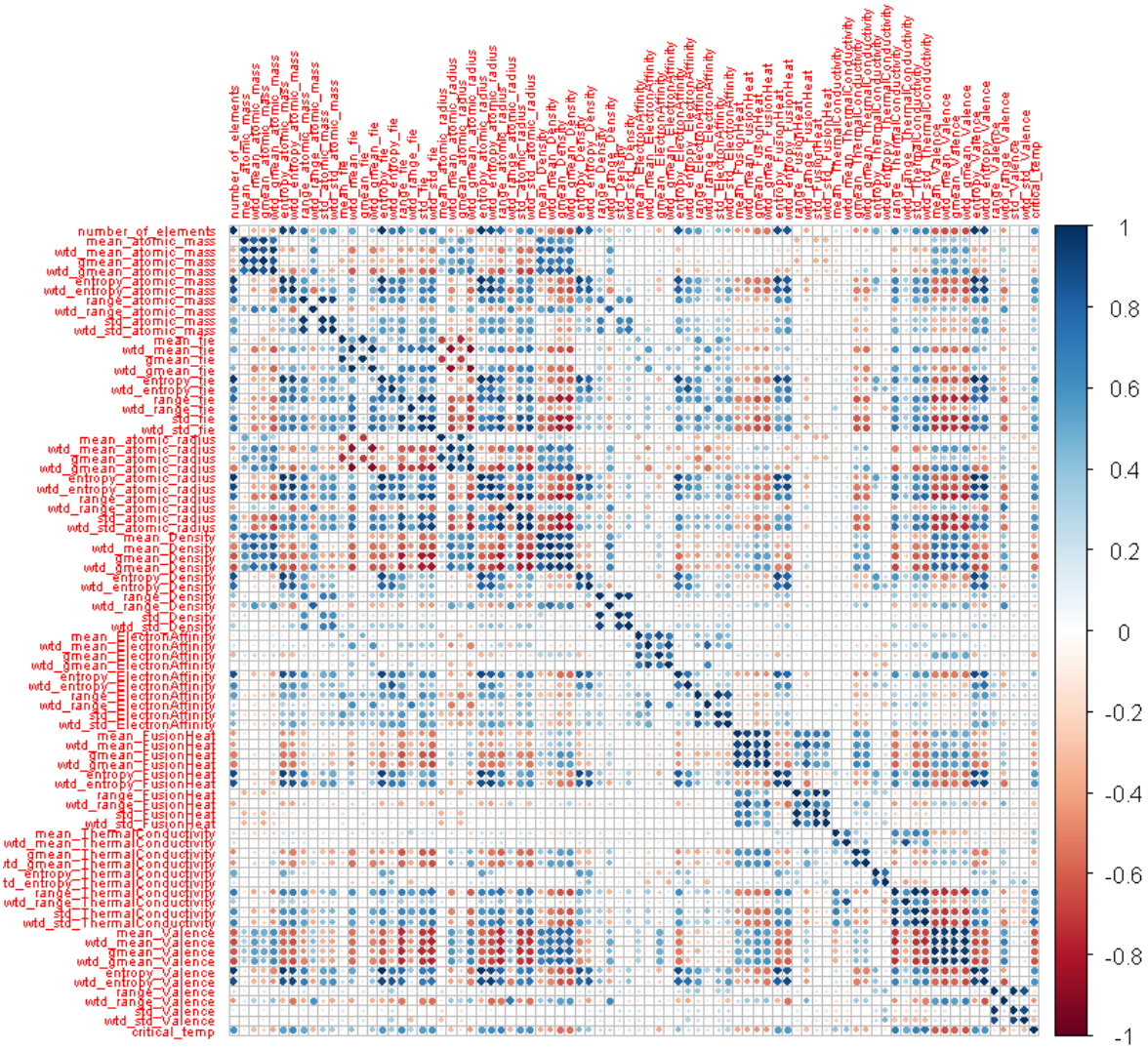
```
library(corrplot)
```

In [104]:

```
library(dplyr)
library(caret)
```

In [105]:

```
# Correlation plot - High Correlation among variables
# It can be seen that high correlation exists!
correlations <- cor(supercon_data)
corrplot(correlations, method="circle",tl.cex=0.5)
```



In [106]:

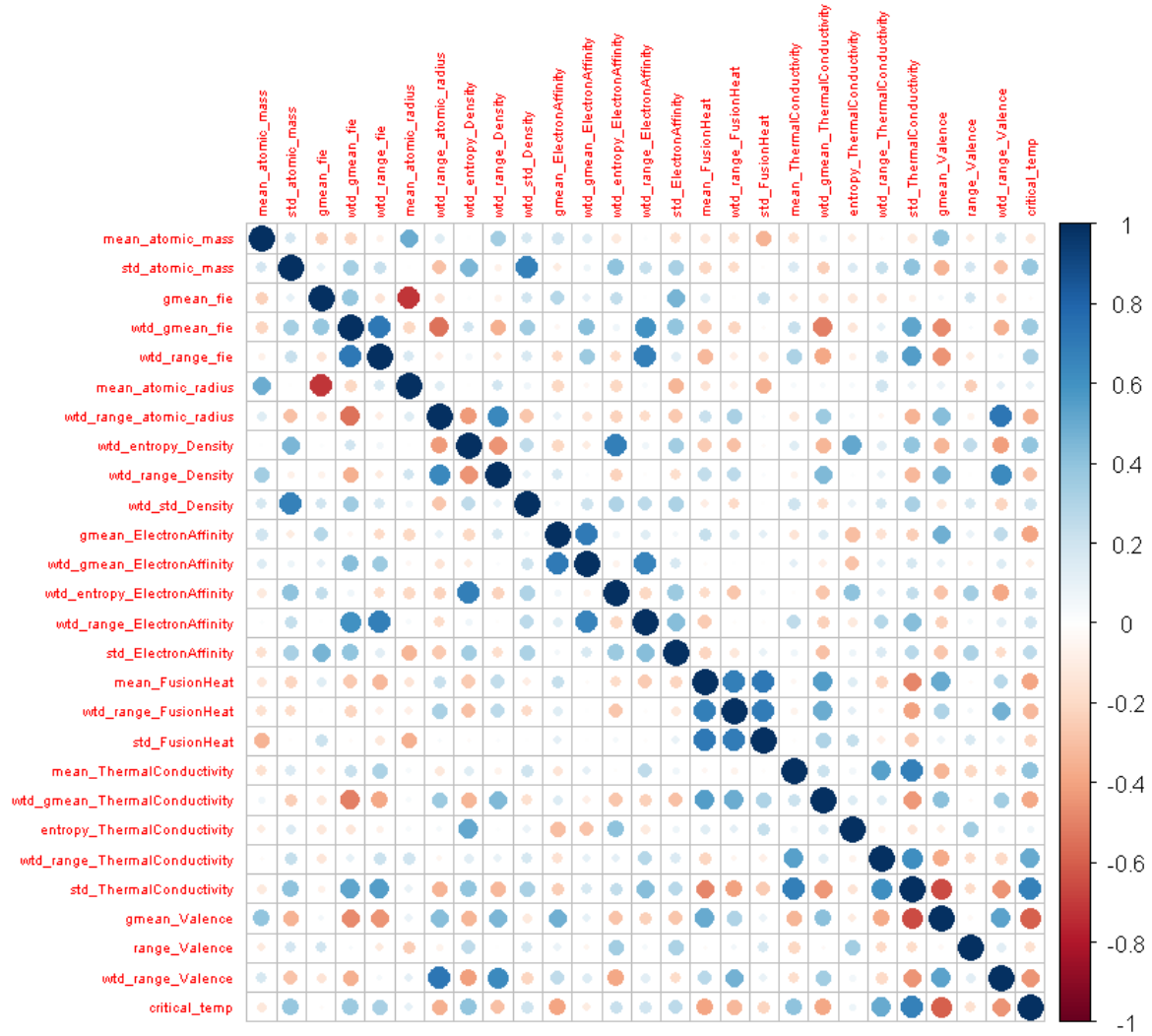
```
cor_matrix = cor(supercon_data)
```

In [123]:

```
drop = findCorrelation(cor_matrix, cutoff = .75) #function that returns a vector of integer
drop = names(supercon_data)[drop]
supercon_data_corr_rem = supercon_data[ , !(names(supercon_data) %in% drop)]
```

In [124]:

```
correlations <- cor(supercon_data_corr_rem)
corrplot(correlations, method="circle",tl.cex=0.5)
```



In [125]:

```
#Getting the highly correlated variable groups for better explainability
row_col_mat = which(cor_matrix>=0.6, arr.ind=TRUE)
rc_df = as.data.frame(row_col_mat)
correlated_var_groups = rc_df %>% group_by(row) %>%
summarize(col = paste(sort(unique(col)),collapse=", "))
```

In [126]:

```
smp_size <- floor(0.70 * nrow(supercon_data_corr_rem))

## set the seed to make your partition reproducible
set.seed(10)
train_ind <- sample(seq_len(nrow(supercon_data_corr_rem)), size = smp_size)

train <- supercon_data_corr_rem[train_ind, ]

test <- supercon_data_corr_rem[-train_ind, ]
```

First Cut Model Creation

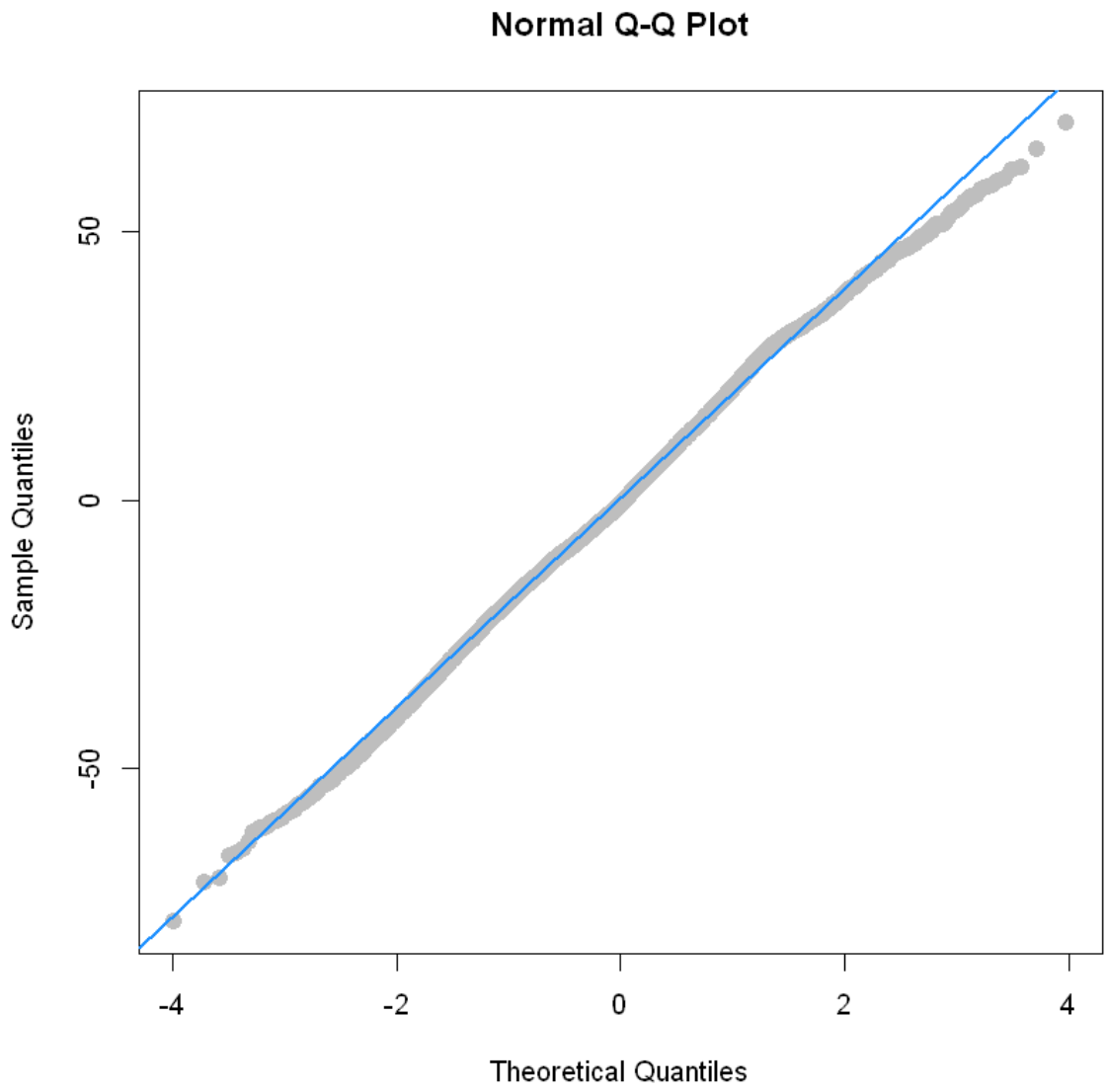
In [243]:

```
model_correlated_rem<-lm(critical_temp~.,data=train)
summary(model_correlated_rem)
```

```
Call:
lm(formula = critical_temp ~ ., data = train)

Residuals:
    Min      1Q  Median      3Q     Max 
-78.441 -12.825  -0.885  13.471  70.165 

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)    
(Intercept)                  -5.346e+01  4.896e+00 -10.918  < 2e-16 ***
mean_atomic_mass             -8.954e-02  9.772e-03  -9.163  < 2e-16 ***
std_atomic_mass               4.140e-01  1.450e-02  28.556  < 2e-16 ***
gmean_fie                     5.396e-02  5.976e-03   9.030  < 2e-16 ***
wtd_gmean_fie                 1.799e-02  5.962e-03   3.018 0.002550 ** 
wtd_range_fie                -9.772e-03  2.525e-03  -3.870 0.000109 ***
mean_atomic_radius            2.714e-01  1.670e-02  16.257  < 2e-16 ***
wtd_range_atomic_radius      -1.591e-01  1.162e-02 -13.689  < 2e-16 ***
wtd_entropy_Density           5.919e+00  1.403e+00   4.220 2.46e-05 ***
wtd_range_Density             8.660e-04  1.448e-04   5.982 2.26e-09 ***
wtd_std_Density              -3.193e-03  1.841e-04 -17.342  < 2e-16 ***
gmean_ElectronAffinity        1.775e-01  1.566e-02  11.336  < 2e-16 ***
wtd_gmean_ElectronAffinity   -2.363e-01  1.746e-02 -13.533  < 2e-16 ***
wtd_entropy_ElectronAffinity -2.712e+01  1.604e+00 -16.904  < 2e-16 ***
wtd_range_ElectronAffinity   -1.178e-01  1.870e-02  -6.297 3.12e-10 ***
std_ElectronAffinity          1.392e-01  1.304e-02  10.678  < 2e-16 ***
mean_FusionHeat               8.484e-02  3.599e-02   2.357 0.018416 *  
wtd_range_FusionHeat         -3.453e-02  3.106e-02  -1.112 0.266350    
std_FusionHeat               -3.004e-01  4.292e-02  -6.998 2.71e-12 ***
mean_ThermalConductivity      1.311e-01  1.416e-02   9.256  < 2e-16 ***
wtd_gmean_ThermalConductivity -3.445e-01  1.165e-02 -29.567  < 2e-16 ***
entropy_ThermalConductivity   2.418e+01  1.191e+00  20.305  < 2e-16 ***
wtd_range_ThermalConductivity 3.079e-01  8.510e-03  36.184  < 2e-16 ***
std_ThermalConductivity       6.780e-02  1.503e-02   4.510 6.54e-06 ***
gmean_Valence                -4.495e+00  3.634e-01 -12.370  < 2e-16 ***
range_Valence                -3.867e+00  1.898e-01 -20.371  < 2e-16 ***
wtd_range_Valence             1.355e+00  4.453e-01   3.043 0.002345 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.83 on 14302 degrees of freedom
Multiple R-squared:  0.6606,    Adjusted R-squared:  0.6599 
F-statistic:  1070 on 26 and 14302 DF,  p-value: < 2.2e-16
```

In [235]:

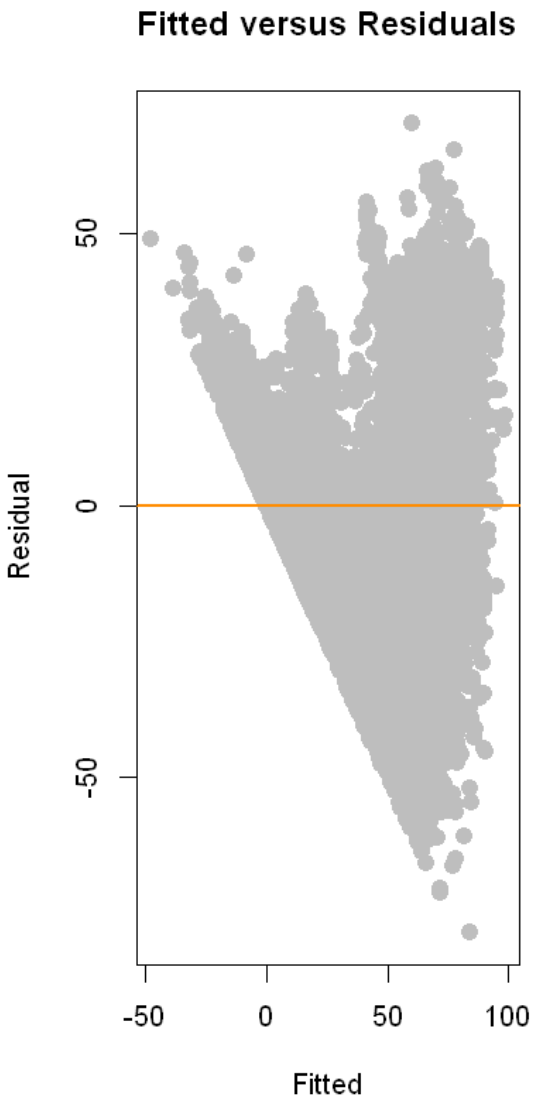
```
qqnorm(resid(model_correlated_rem), col = "grey",pch=20,cex=2)
qqline(resid(model_correlated_rem), col = "dodgerblue", lwd = 2)
```



Normal Q-Q Plot seems to be good

In [236]:

```
par(mfrow=c(1,2))
plot(fitted(model_correlated_rem), resid(model_correlated_rem), col = "grey", pch = 20,
     xlab = "Fitted", ylab = "Residual",cex=2,
     main = "Fitted versus Residuals")
abline(h = 0, col = "darkorange", lwd = 2)
```



The residual plot has equal variance but shows that that y should be transformed as as seen below.

2.1. Train-Test Split

4. Data Transformations

4.1. Taylor Expansion

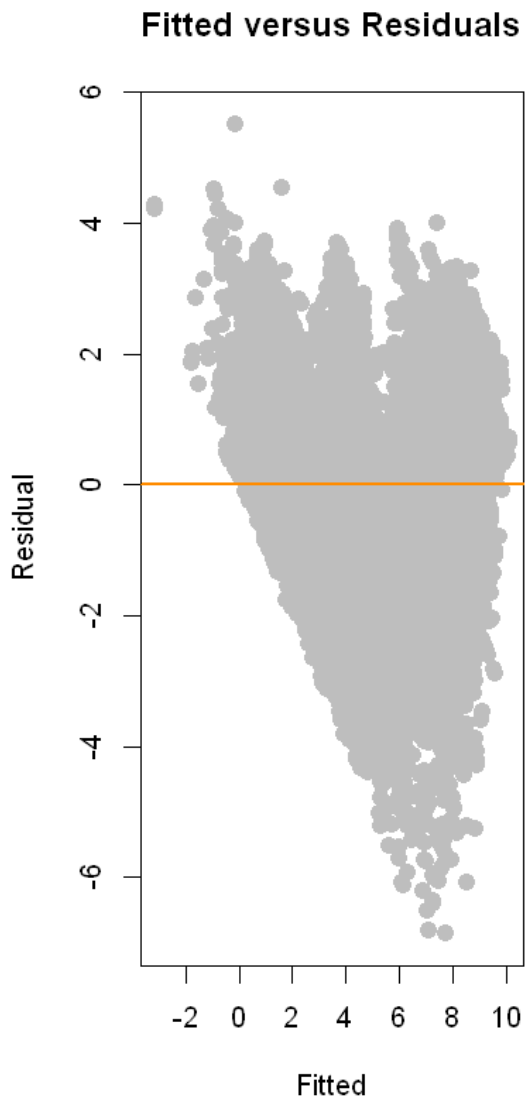
In [168]:

```
model_taylor_tf<-lm(sqrt(critical_temp)~.,data=train)
summary(model_taylor_tf)
```

```
Call:
lm(formula = sqrt(critical_temp) ~ ., data = train)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.8585 -1.0654  0.0565  1.2521  5.5098 

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)    
(Intercept)                  -2.172e+00  4.025e-01  -5.397 6.89e-08 ***
mean_atomic_mass             -1.121e-02  8.032e-04 -13.952  < 2e-16 ***
std_atomic_mass               3.136e-02  1.192e-03  26.317  < 2e-16 ***
gmean_fie                     2.760e-03  4.912e-04   5.618 1.96e-08 ***
wtd_gmean_fie                 3.281e-03  4.901e-04   6.695 2.23e-11 ***
wtd_range_fie                -9.001e-04  2.076e-04  -4.337 1.46e-05 ***
mean_atomic_radius            2.164e-02  1.372e-03  15.769  < 2e-16 ***
wtd_range_atomic_radius      -9.829e-03  9.552e-04 -10.290  < 2e-16 ***
wtd_entropy_Density           5.278e-01  1.153e-01   4.578 4.74e-06 ***
wtd_range_Density             5.464e-05  1.190e-05   4.592 4.44e-06 ***
wtd_std_Density              -2.877e-04  1.514e-05 -19.010  < 2e-16 ***
gmean_ElectronAffinity        1.169e-02  1.287e-03   9.082  < 2e-16 ***
wtd_gmean_ElectronAffinity   -1.879e-02  1.435e-03 -13.092  < 2e-16 ***
wtd_entropy_ElectronAffinity -1.945e+00  1.319e-01 -14.754  < 2e-16 ***
wtd_range_ElectronAffinity   -1.281e-02  1.537e-03  -8.330  < 2e-16 ***
std_ElectronAffinity          1.359e-02  1.072e-03  12.678  < 2e-16 ***
mean_FusionHeat               1.328e-02  2.958e-03   4.489 7.21e-06 ***
wtd_range_FusionHeat          6.068e-03  2.553e-03   2.376   0.0175 *  
std_FusionHeat               -3.046e-02  3.528e-03  -8.635  < 2e-16 ***
mean_ThermalConductivity      5.849e-03  1.164e-03   5.024 5.12e-07 ***
wtd_gmean_ThermalConductivity -2.866e-02  9.578e-04 -29.923  < 2e-16 ***
entropy_ThermalConductivity   2.058e+00  9.789e-02  21.029  < 2e-16 ***
wtd_range_ThermalConductivity 2.316e-02  6.995e-04  33.107  < 2e-16 ***
std_ThermalConductivity       1.215e-02  1.236e-03   9.831  < 2e-16 ***
gmean_Valence                -3.930e-01  2.987e-02 -13.158  < 2e-16 ***
range_Valence                -2.604e-01  1.561e-02 -16.690  < 2e-16 ***
wtd_range_Valence             7.647e-02  3.660e-02   2.089   0.0367 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.63 on 14302 degrees of freedom
Multiple R-squared:  0.7086,    Adjusted R-squared:  0.7081 
F-statistic:  1338 on 26 and 14302 DF,  p-value: < 2.2e-16
```

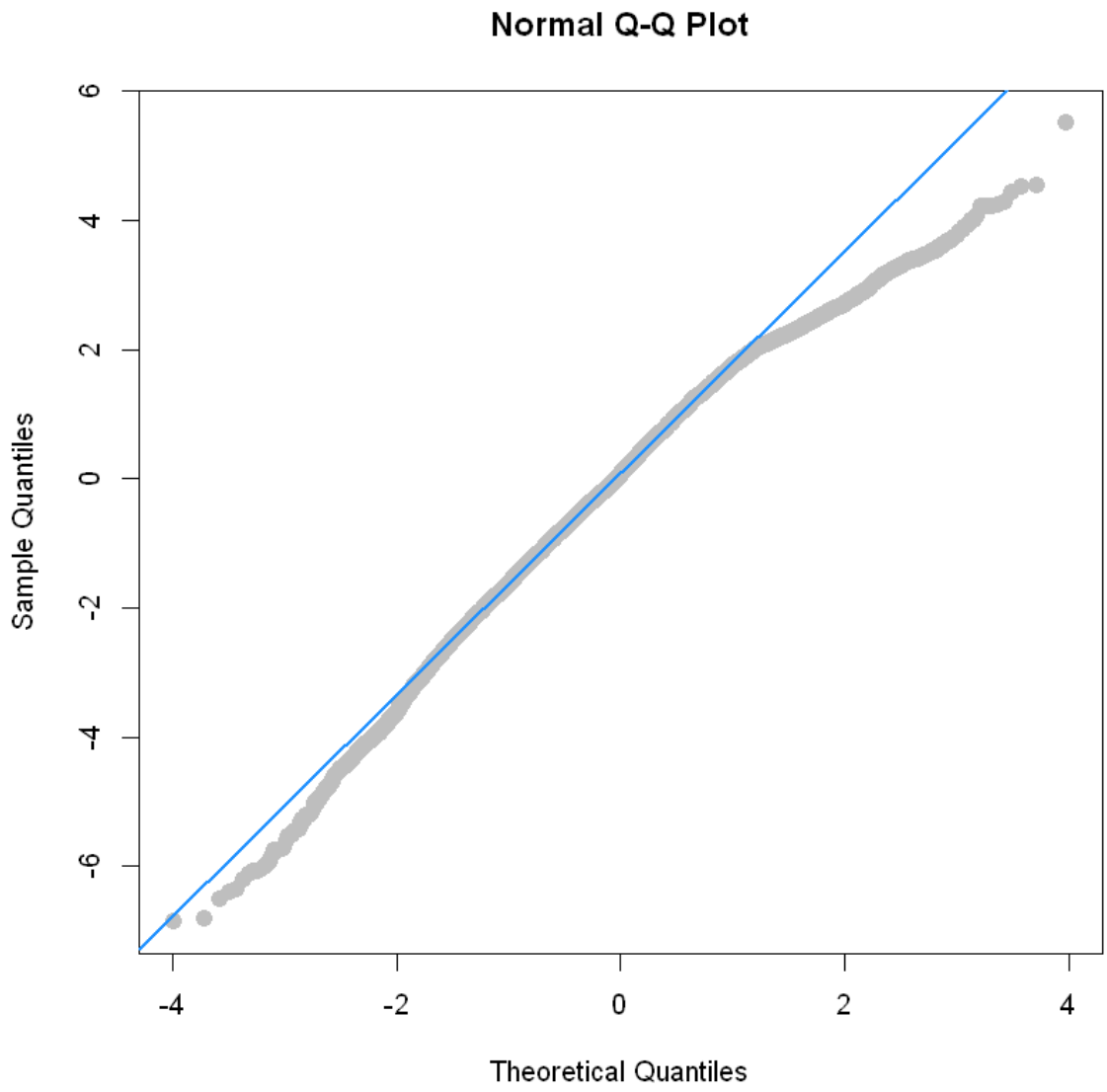
In [169]:

```
par(mfrow=c(1,2))
plot(fitted(model_taylor_tf), resid(model_taylor_tf), col = "grey", pch = 20,
     xlab = "Fitted", ylab = "Residual",cex=2,
     main = "Fitted versus Residuals")
abline(h = 0, col = "darkorange", lwd = 2)
```



In [170]:

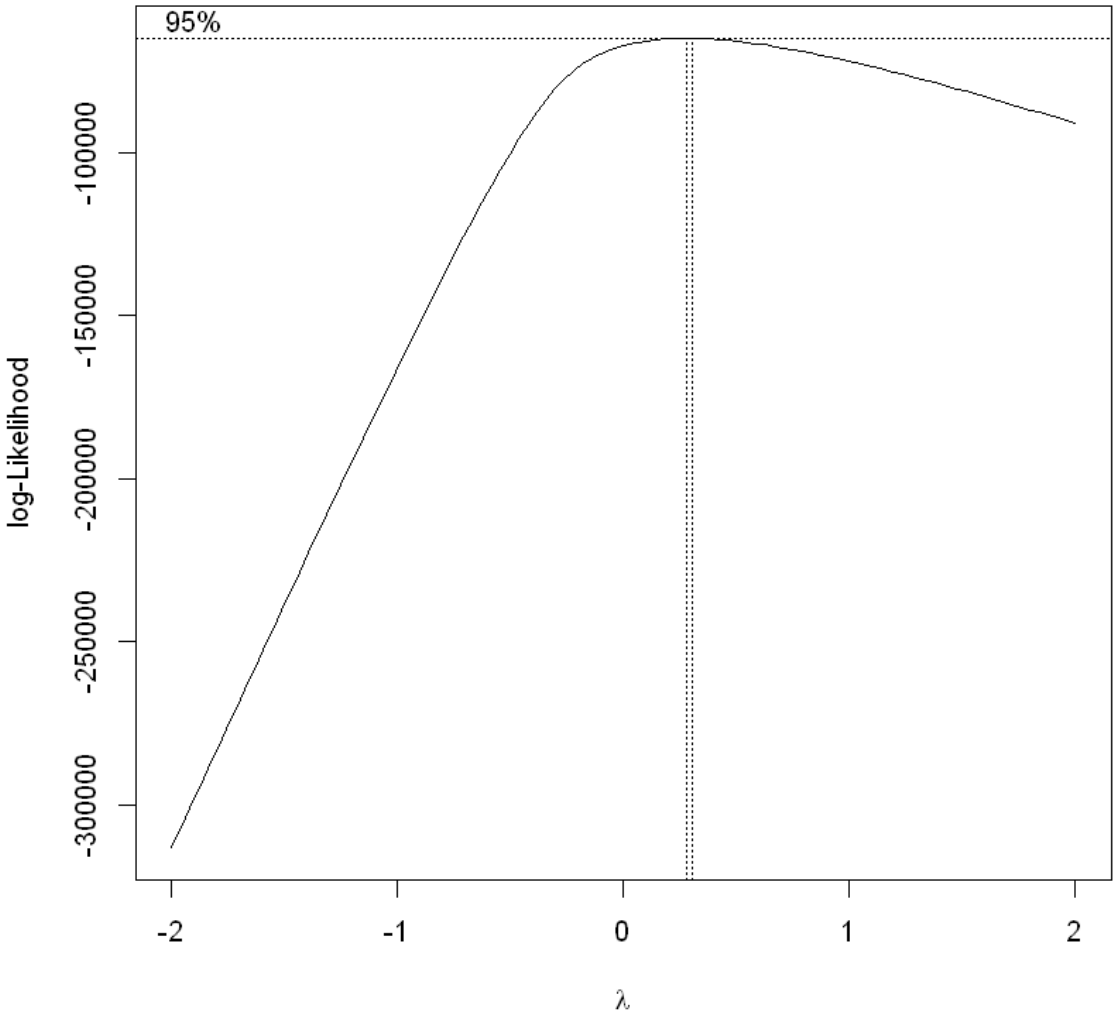
```
qqnorm(resid(model_taylor_tf), col = "grey",pch=20,cex=2)
qqline(resid(model_taylor_tf), col = "dodgerblue", lwd = 2)
```



4.2. Box Cox method

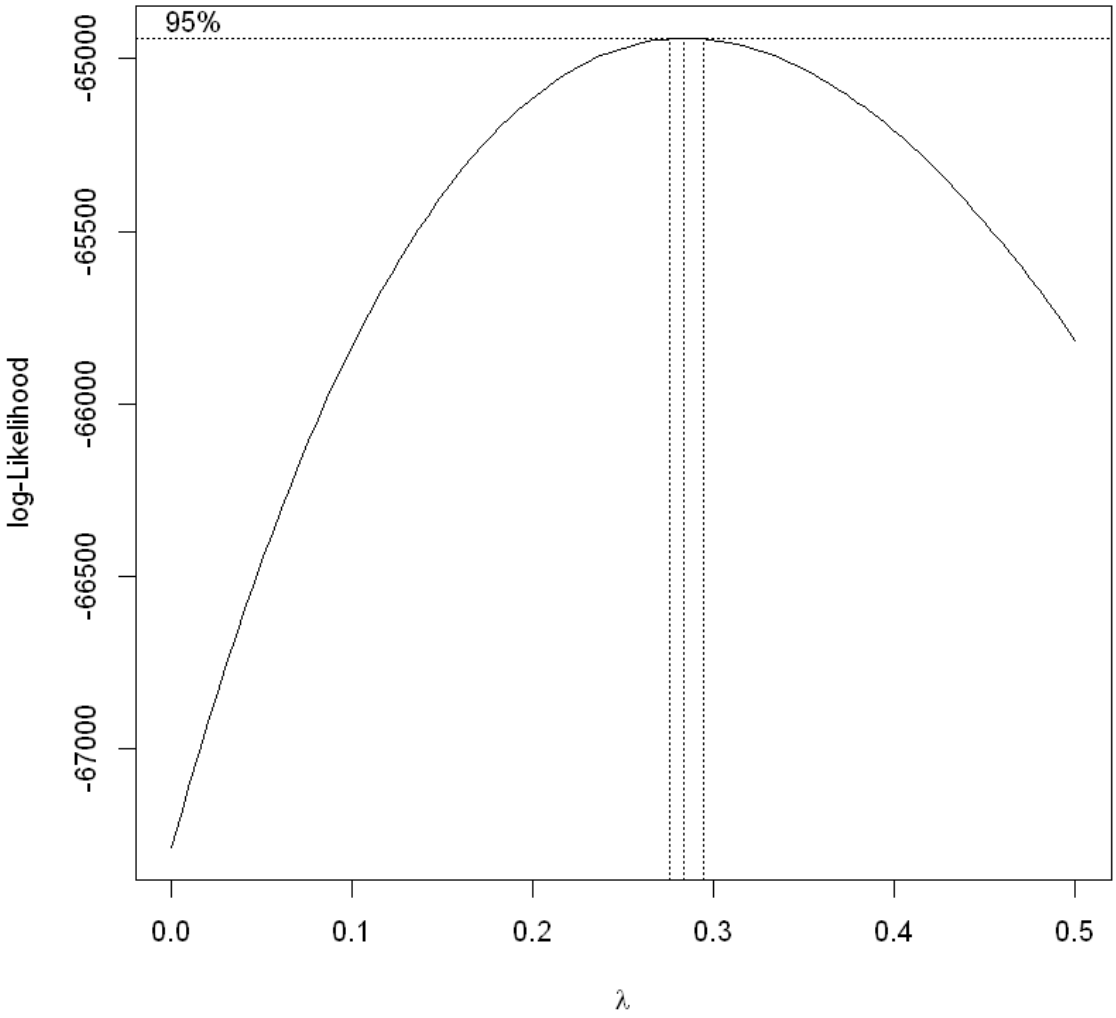
In [244]:

```
boxcox(model_correlated_rem)
```



In [245]:

```
boxcox(model_correlated_rem, lambda = seq(-0, 0.5, by = 0.01))
```



In [247]:

```
lambda = 0.29
model_boxcox_supercon <- lm(((critical_temp^(lambda)-1)/(lambda))~.,data=train)
```

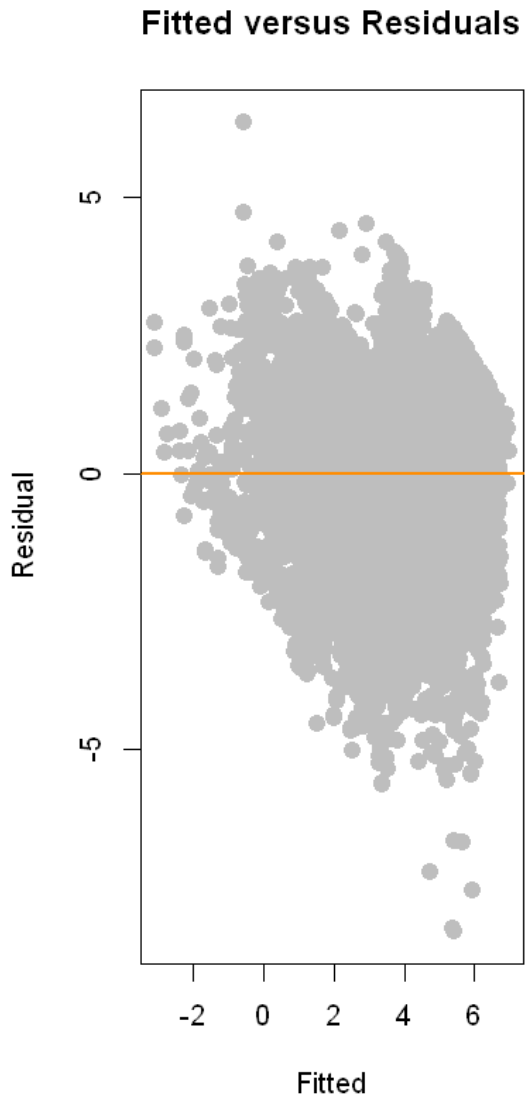
In [248]:

```
summary(model_boxcox_supercon)
```

```
Call:
lm(formula = ((critical_temp^(lambda) - 1)/(lambda)) ~ ., data = train)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.2224 -1.0711  0.1646  1.2750  6.7521 

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)    
(Intercept)                  -1.830e+00  4.255e-01  -4.300 1.72e-05 ***
mean_atomic_mass             -1.338e-02  8.491e-04 -15.755  < 2e-16 ***
std_atomic_mass               3.008e-02  1.260e-03  23.878  < 2e-16 ***
gmean_fie                     1.553e-03  5.192e-04   2.990  0.00279 ** 
wtd_gmean_fie                 4.192e-03  5.181e-04   8.091 6.37e-16 ***
wtd_range_fie                -1.031e-03  2.194e-04  -4.698 2.65e-06 ***
mean_atomic_radius            2.063e-02  1.451e-03  14.217  < 2e-16 ***
wtd_range_atomic_radius      -7.506e-03  1.010e-03  -7.433 1.12e-13 ***
wtd_entropy_Density           4.774e-01  1.219e-01   3.917 9.02e-05 ***
wtd_range_Density             2.571e-05  1.258e-05   2.044  0.04097 *  
wtd_std_Density              -2.859e-04  1.600e-05 -17.868  < 2e-16 ***
gmean_ElectronAffinity        1.013e-02  1.361e-03   7.444 1.03e-13 ***
wtd_gmean_ElectronAffinity   -1.969e-02  1.517e-03 -12.979  < 2e-16 ***
wtd_entropy_ElectronAffinity -1.689e+00  1.394e-01 -12.118  < 2e-16 ***
wtd_range_ElectronAffinity   -1.365e-02  1.625e-03  -8.400  < 2e-16 ***
std_ElectronAffinity          1.455e-02  1.133e-03  12.840  < 2e-16 ***
mean_FusionHeat               1.930e-02  3.127e-03   6.172 6.93e-10 ***
wtd_range_FusionHeat          1.266e-02  2.699e-03   4.690 2.75e-06 ***
std_FusionHeat               -3.370e-02  3.730e-03  -9.036  < 2e-16 ***
mean_ThermalConductivity      2.761e-03  1.231e-03   2.243  0.02489 *  
wtd_gmean_ThermalConductivity -2.964e-02  1.013e-03 -29.270  < 2e-16 ***
entropy_ThermalConductivity   2.079e+00  1.035e-01  20.087  < 2e-16 ***
wtd_range_ThermalConductivity 2.097e-02  7.394e-04  28.366  < 2e-16 ***
std_ThermalConductivity       1.609e-02  1.306e-03  12.319  < 2e-16 ***
gmean_Valence                -4.095e-01  3.158e-02 -12.967  < 2e-16 ***
range_Valence                -2.301e-01  1.650e-02 -13.951  < 2e-16 ***
wtd_range_Valence             8.284e-02  3.869e-02   2.141  0.03231 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.723 on 14302 degrees of freedom
Multiple R-squared:  0.7094,    Adjusted R-squared:  0.7089 
F-statistic:  1343 on 26 and 14302 DF,  p-value: < 2.2e-16
```

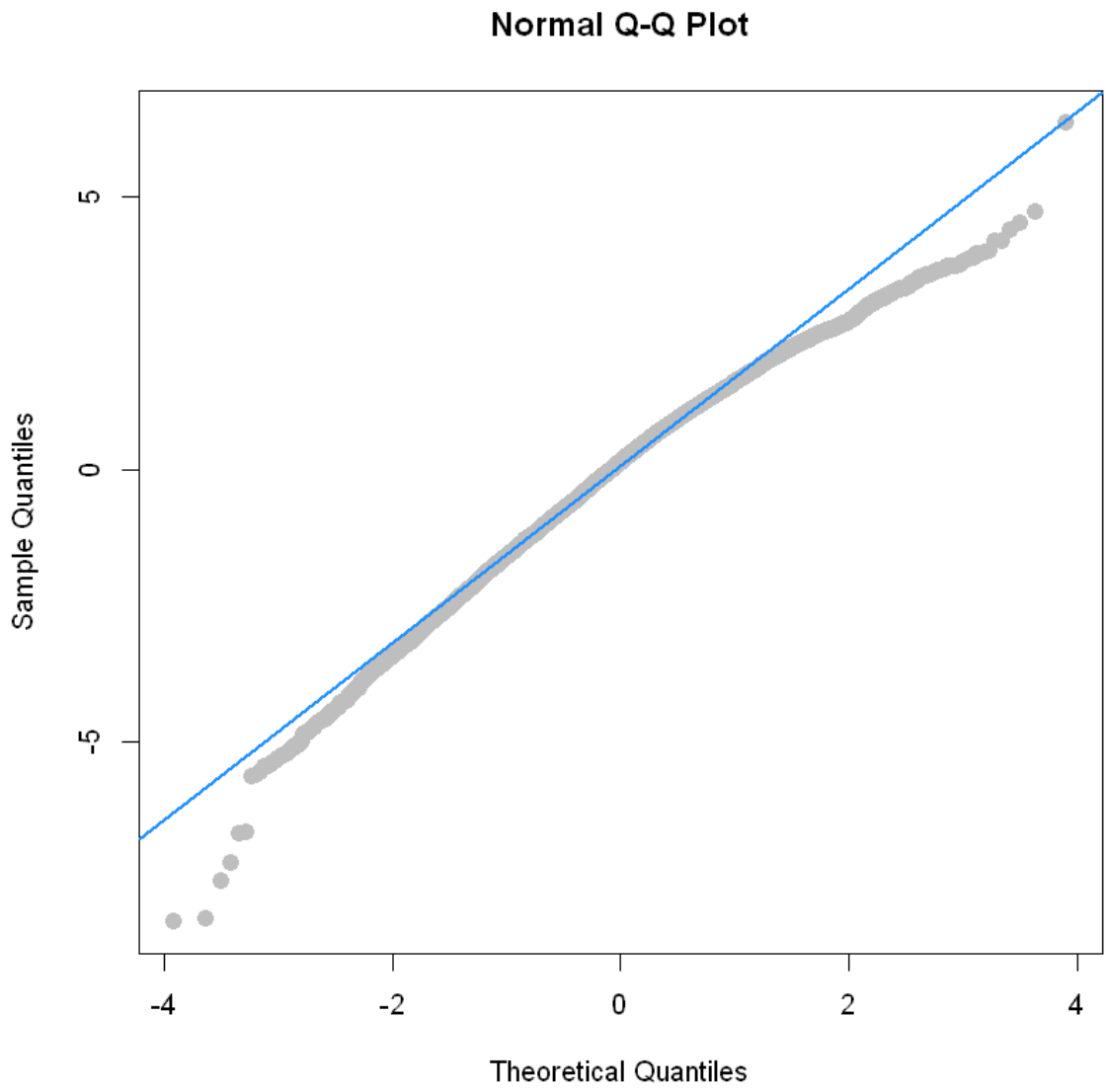
In [231]:

```
par(mfrow=c(1,2))
plot(fitted(model_boxcox_supercon), resid(model_boxcox_supercon), col = "grey", pch = 20,
     xlab = "Fitted", ylab = "Residual",cex=2,
     main = "Fitted versus Residuals")
abline(h = 0, col = "darkorange", lwd = 2)
```



In [232]:

```
qqnorm(resid(model_boxcox_supercon), col = "grey",pch=20,cex=2)
qqline(resid(model_boxcox_supercon), col = "dodgerblue", lwd = 2)
```



4. Regularization

In [249]:

```
library(ISLR)
library(glmnet)
```

```
Loading required package: Matrix
Loading required package: foreach
Loaded glmnet 2.0-16
```

In [251]:

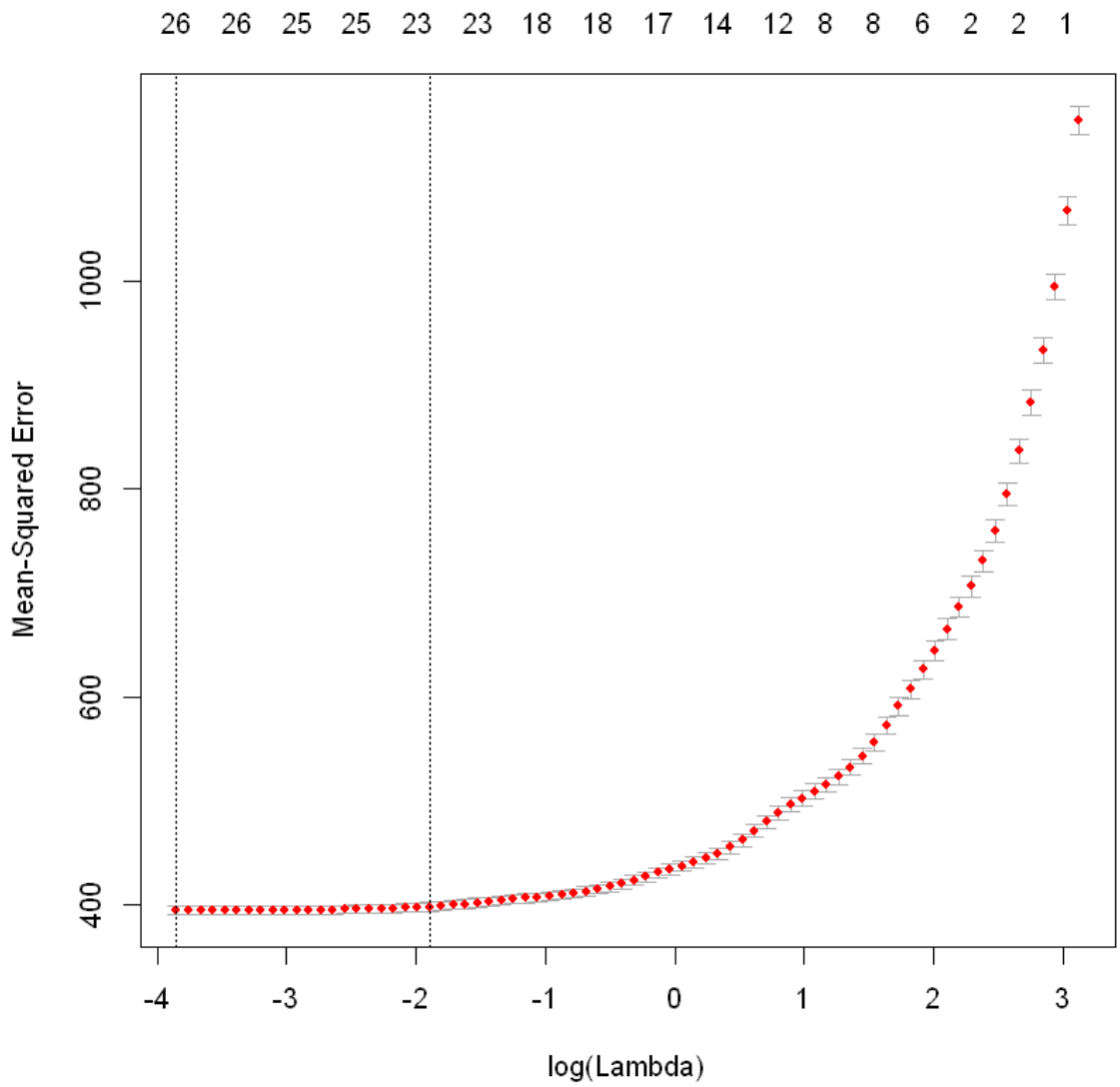
```
X_train = model.matrix(critical_temp~. ,train)[, -1]# the last column (for intercept) is el
y_train = train$critical_temp
X_test = model.matrix(critical_temp~. ,test)[, -1]
y_test = test$critical_temp
```

In [310]:

```
eval_results <- function(true, predicted, df) {
  SSE <- sum((predicted - true)^2)
  SST <- sum((true - mean(true))^2)
  R_square <- 1 - SSE / SST
  RMSE = sqrt(SSE/nrow(df))


  # Model performance metrics
  data.frame(
    RMSE = RMSE,
    Rsquare = R_square,
    adj.r.squared=1- ((1-R_square)*(21263-1)/(21263-27-1))
  )

}
```

In [311]:

```
cv_model_regularization = cv.glmnet(X_train, y_train, alpha = 1)
plot(cv_model_regularization)
```



In [312]:

```
bestlam = cv_model_regularization$lambda.min
bestlam
```

0.0212775519871423

In [313]:

```
model_regularization <- glmnet(X_train, y_train, alpha = 1, lambda = 0.02)


ypr4 = predict(model_regularization, X_test, lambda = 0.02)
eval_results(y_test, ypr4, test)

lasso_coef = coef(model_regularization)
sum(lasso_coef != 0)
```

RMSE	Rsquare	adj.r.squared
19.7324	0.6636291	0.6632014

27

In [ ]:

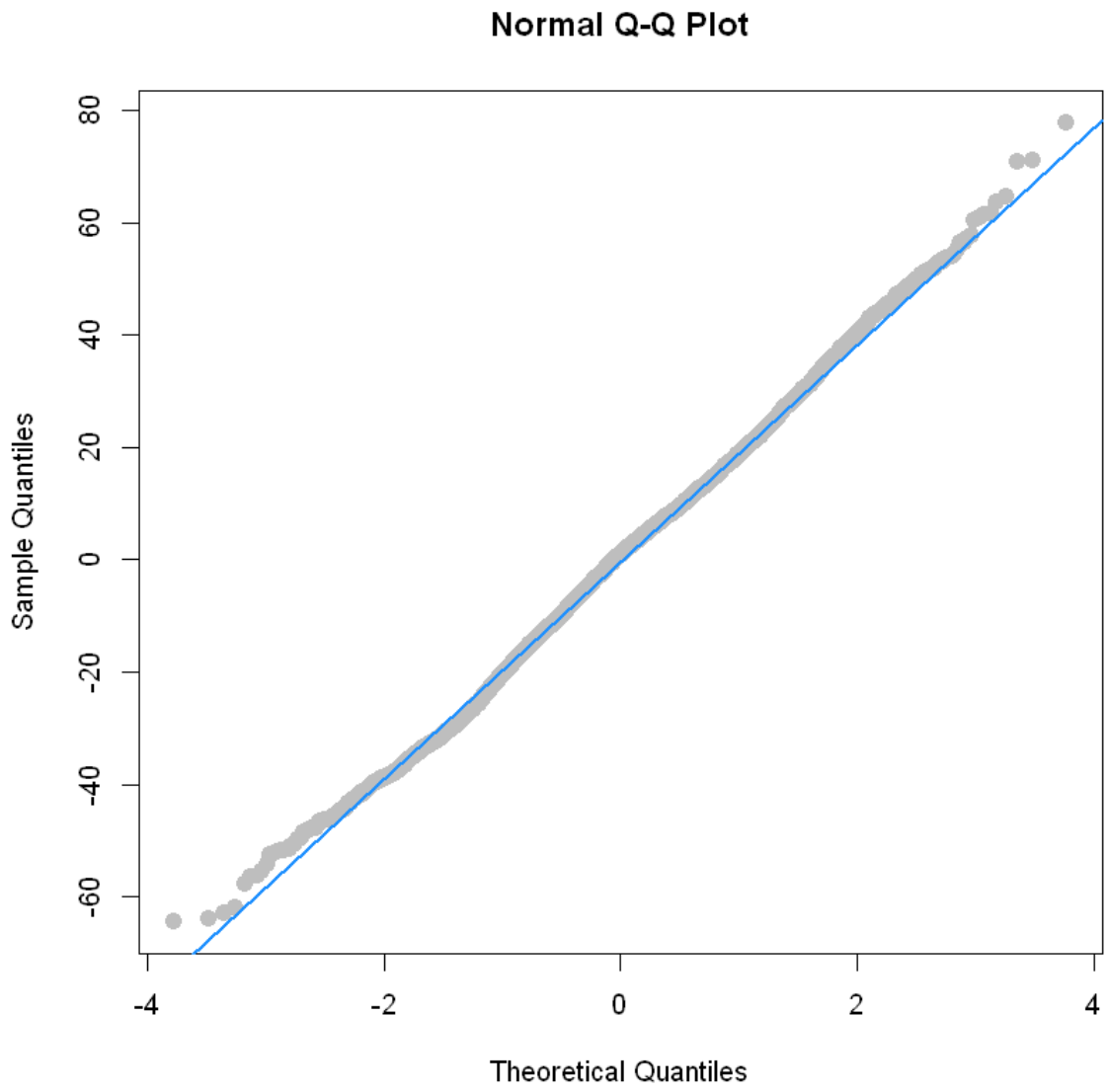
```
adj.r.squared=
```

In [288]:

```
resid_reg = ypr4-y_test
fitted_reg = ypr4
```

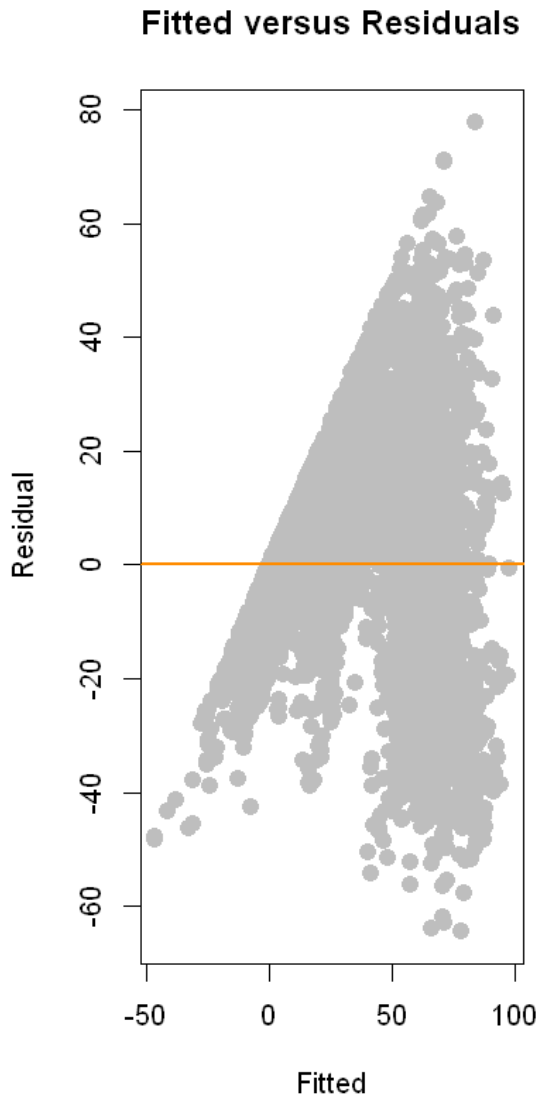
In [289]:

```
qqnorm(resid_reg, col = "grey",pch=20,cex=2)
qqline(resid_reg, col = "dodgerblue", lwd = 2)
```



In [290]:

```
par(mfrow=c(1,2))
plot(fitted_reg, resid_reg, col = "grey", pch = 20,
     xlab = "Fitted", ylab = "Residual",cex=2,
     main = "Fitted versus Residuals")
abline(h = 0, col = "darkorange", lwd = 2)
```



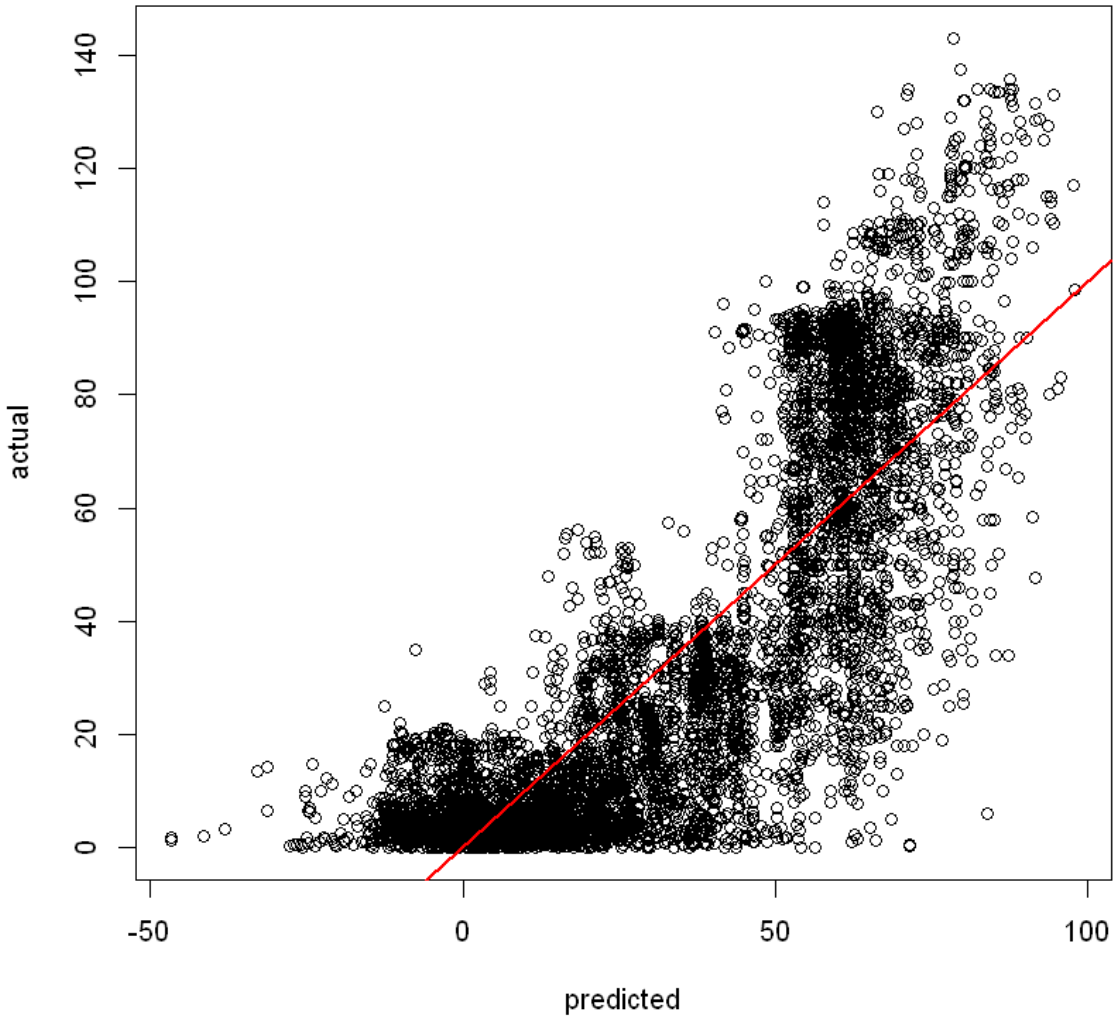
In [296]:

```
coef(model_regularization)
```

```
27 x 1 sparse Matrix of class "dgCMatrix"
                                        s0
(Intercept)                   -5.125209e+01
mean_atomic_mass              -8.475087e-02
std_atomic_mass                4.070248e-01
gmean_fie                      5.437067e-02
wtd_gmean_fie                  1.488815e-02
wtd_range_fie                 -8.325990e-03
mean_atomic_radius             2.637769e-01
wtd_range_atomic_radius       -1.560860e-01
wtd_entropy_Density            5.604194e+00
wtd_range_Density              8.069471e-04
wtd_std_Density               -3.110983e-03
gmean_ElectronAffinity         1.632561e-01
wtd_gmean_ElectronAffinity    -2.313093e-01
wtd_entropy_ElectronAffinity  -2.555549e+01
wtd_range_ElectronAffinity    -1.070722e-01
std_ElectronAffinity           1.335118e-01
mean_FusionHeat                8.074163e-02
wtd_range_FusionHeat          -2.269603e-02
std_FusionHeat                -3.023492e-01
mean_ThermalConductivity       1.372186e-01
wtd_gmean_ThermalConductivity -3.453793e-01
entropy_ThermalConductivity    2.341209e+01
wtd_range_ThermalConductivity  3.075683e-01
std_ThermalConductivity        6.025478e-02
gmean_Valence                 -4.350924e+00
range_Valence                 -3.805605e+00
wtd_range_Valence              1.054035e+00
```

In [316]:

```
plot(ypr4,y_test,xlab="predicted",ylab="actual")
 abline(a=0,b=1,col='red',lwd=2)
```



4.2. AIC/BIC Methods

In [5]:

```
#construct full model
model_full = lm(critical_temp~.,data=supercon_data)
```

In [9]:

```
fit_back_aic = step(model_full, direction = "backward")
fit_back_aic
```

```
Start:  AIC=122021.5
critical_temp ~ number_of_elements + mean_atomic_mass + wtd_mean_atomic_ma
ss +
    gmean_atomic_mass + wtd_gmean_atomic_mass + entropy_atomic_mass +
    wtd_entropy_atomic_mass + range_atomic_mass + wtd_range_atomic_mass +
    std_atomic_mass + wtd_std_atomic_mass + mean_fie + wtd_mean_fie +
    gmean_fie + wtd_gmean_fie + entropy_fie + wtd_entropy_fie +
    range_fie + wtd_range_fie + std_fie + wtd_std_fie + mean_atomic_radius
+
    wtd_mean_atomic_radius + gmean_atomic_radius + wtd_gmean_atomic_radius
+
    entropy_atomic_radius + wtd_entropy_atomic_radius + range_atomic_radiu
s +
    wtd_range_atomic_radius + std_atomic_radius + wtd_std_atomic_radius +
    mean_Density + wtd_mean_Density + gmean_Density + wtd_gmean_Density +
    entropy_Density + wtd_entropy_Density + range_Density + wtd_range_Dens
ity +
    std_Density + wtd_std_Density + mean_ElectronAffinity + wtd_mean_Elect
ronAffinity +
```

In [7]:

```
n = nrow(supercon_data)
fit_back_bic = step(model_full, direction = "backward", k=log(n))
fit_back_bic
```

```
- wtd_mean_ThermalConductivity     1    108061 6662720 123012
- range_ElectronAffinity           1    138826 6693485 123110
- std_ElectronAffinity             1    141384 6696043 123118

Step:  AIC=122664.6
critical_temp ~ number_of_elements + mean_atomic_mass + wtd_mean_atomic_ma
ss +
    gmean_atomic_mass + wtd_gmean_atomic_mass + entropy_atomic_mass +
    wtd_entropy_atomic_mass + range_atomic_mass + wtd_range_atomic_mass +
    std_atomic_mass + wtd_std_atomic_mass + mean_fie + wtd_mean_fie +
    gmean_fie + wtd_gmean_fie + entropy_fie + wtd_entropy_fie +
    range_fie + wtd_range_fie + std_fie + wtd_std_fie + mean_atomic_radius
+
    wtd_mean_atomic_radius + gmean_atomic_radius + wtd_gmean_atomic_radius
+
    entropy_atomic_radius + wtd_entropy_atomic_radius + range_atomic_radiu
s +
    wtd_range_atomic_radius + std_atomic_radius + wtd_std_atomic_radius +
    mean_Density + wtd_mean_Density + gmean_Density + wtd_gmean_Density +
```

In [10]:

```
summary(fit_back_aic)
```

```
Call:
lm(formula = critical_temp ~ number_of_elements + mean_atomic_mass +
    wtd_mean_atomic_mass + gmean_atomic_mass + wtd_gmean_atomic_mass +
    entropy_atomic_mass + range_atomic_mass + std_atomic_mass +
    wtd_std_atomic_mass + mean_fie + wtd_mean_fie + gmean_fie +
    wtd_gmean_fie + entropy_fie + wtd_entropy_fie + range_fie +
    wtd_range_fie + std_fie + mean_atomic_radius + wtd_mean_atomic_radius
+
    wtd_gmean_atomic_radius + entropy_atomic_radius + wtd_entropy_atomic_r
adius +
    range_atomic_radius + wtd_range_atomic_radius + std_atomic_radius +
    wtd_std_atomic_radius + mean_Density + gmean_Density + wtd_gmean_Densi
ty +
    entropy_Density + wtd_entropy_Density + range_Density + std_Density +
    wtd_std_Density + mean_ElectronAffinity + wtd_mean_ElectronAffinity +
    gmean_ElectronAffinity + wtd_gmean_ElectronAffinity + entropy_Electron
Affinity +
    wtd_entropy_ElectronAffinity + range_ElectronAffinity + wtd_range_Elec
```

In [8]:

```
summary(fit_back_bic)
```

```
Call:
lm(formula = critical_temp ~ number_of_elements + mean_atomic_mass +
    wtd_mean_atomic_mass + gmean_atomic_mass + wtd_gmean_atomic_mass +
    entropy_atomic_mass + range_atomic_mass + std_atomic_mass +
    mean_fie + wtd_mean_fie + gmean_fie + wtd_gmean_fie + entropy_fie +
    wtd_entropy_fie + range_fie + wtd_range_fie + std_fie + mean_atomic_radi
us +
    wtd_mean_atomic_radius + wtd_gmean_atomic_radius + entropy_atomic_radius
+
    wtd_entropy_atomic_radius + range_atomic_radius + wtd_range_atomic_radiu
s +
    std_atomic_radius + wtd_std_atomic_radius + mean_Density +
    gmean_Density + wtd_gmean_Density + entropy_Density + wtd_entropy_Densit
y +
    range_Density + std_Density + wtd_std_Density + wtd_mean_ElectronAffinit
y +
    gmean_ElectronAffinity + wtd_gmean_ElectronAffinity + wtd_entropy_Electr
onAffinity +
    range_ElectronAffinity + wtd_range_ElectronAffinity + std_ElectronAffini
ty +
    wtd_std_ElectronAffinity + mean_FusionHeat + wtd_mean_FusionHeat +
    gmean_FusionHeat + wtd_gmean_FusionHeat + entropy_FusionHeat +
    wtd_entropy_FusionHeat + range_FusionHeat + wtd_range_FusionHeat +
    wtd_std_FusionHeat + mean_ThermalConductivity + wtd_mean_ThermalConducti
vity +
    wtd_gmean_ThermalConductivity + entropy_ThermalConductivity +
    range_ThermalConductivity + wtd_range_ThermalConductivity +
    std_ThermalConductivity + gmean_Valence + wtd_gmean_Valence +
    entropy_Valence + wtd_entropy_Valence + range_Valence + wtd_std_Valence,
    data = supercon_data)

Residuals:
    Min      1Q  Median      3Q     Max
-83.947  -9.317   0.574  10.915 170.776

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                 -1.861e+01  4.796e+00  -3.880 0.000105 ***
number_of_elements          -3.560e+00  7.167e-01  -4.967 6.84e-07 ***
mean_atomic_mass             7.717e-01  5.928e-02  13.018  < 2e-16 ***
wtd_mean_atomic_mass        -7.805e-01  5.121e-02 -15.241  < 2e-16 ***
gmean_atomic_mass           -4.370e-01  6.284e-02  -6.954 3.66e-12 ***
wtd_gmean_atomic_mass        5.296e-01  5.476e-02   9.672  < 2e-16 ***
entropy_atomic_mass         -3.365e+01  4.051e+00  -8.307  < 2e-16 ***
range_atomic_mass            2.077e-01  1.613e-02  12.873  < 2e-16 ***
std_atomic_mass             -4.709e-01  4.379e-02 -10.752  < 2e-16 ***
mean_fie                     1.655e-01  4.264e-02   3.882 0.000104 ***
wtd_mean_fie                -2.170e-01  3.470e-02  -6.252 4.13e-10 ***
gmean_fie                   -1.630e-01  4.214e-02  -3.867 0.000111 ***
wtd_gmean_fie                2.400e-01  3.464e-02   6.928 4.40e-12 ***
entropy_fie                 -1.037e+02  1.789e+01  -5.799 6.78e-09 ***
wtd_entropy_fie              4.357e+01  3.976e+00  10.959  < 2e-16 ***
range_fie                    6.865e-02  6.246e-03  10.990  < 2e-16 ***
wtd_range_fie                1.990e-02  3.124e-03   6.372 1.91e-10 ***
std_fie                     -2.104e-01  1.591e-02 -13.221  < 2e-16 ***
mean_atomic_radius          -3.499e-01  2.682e-02 -13.047  < 2e-16 ***
```

```
wtd_mean_atomic_radius         3.055e+00  1.513e-01  20.187  < 2e-16 ***
wtd_gmean_atomic_radius       -2.653e+00  1.432e-01 -18.533  < 2e-16 ***
entropy_atomic_radius          6.245e+01  1.632e+01   3.826 0.000131 ***
wtd_entropy_atomic_radius      5.158e+01  3.633e+00  14.199  < 2e-16 ***
range_atomic_radius            2.073e-01  2.133e-02   9.718  < 2e-16 ***
wtd_range_atomic_radius       -8.302e-02  1.214e-02  -6.841 8.10e-12 ***
std_atomic_radius             -4.630e-01  7.014e-02  -6.600 4.21e-11 ***
wtd_std_atomic_radius         -2.683e-01  6.173e-02  -4.346 1.39e-05 ***
mean_Density                  -4.892e-03  3.459e-04 -14.145  < 2e-16 ***
gmean_Density                  1.204e-03  3.496e-04   3.444 0.000574 ***
wtd_gmean_Density              2.334e-03  2.573e-04   9.070  < 2e-16 ***
entropy_Density                1.299e+01  3.132e+00   4.146 3.39e-05 ***
wtd_entropy_Density           -1.629e+01  2.023e+00  -8.051 8.67e-16 ***
range_Density                 -1.522e-03  2.096e-04  -7.265 3.87e-13 ***
std_Density                    5.628e-03  6.406e-04   8.786  < 2e-16 ***
wtd_std_Density               -1.245e-03  3.816e-04  -3.262 0.001107 **
wtd_mean_ElectronAffinity      4.527e-01  3.355e-02  13.492  < 2e-16 ***
gmean_ElectronAffinity         1.092e-01  1.418e-02   7.703 1.39e-14 ***
wtd_gmean_ElectronAffinity    -5.317e-01  3.265e-02 -16.282  < 2e-16 ***
wtd_entropy_ElectronAffinity  -1.926e+01  1.768e+00 -10.895  < 2e-16 ***
range_ElectronAffinity        -3.642e-01  1.624e-02 -22.427  < 2e-16 ***
wtd_range_ElectronAffinity    -1.178e-01  1.726e-02  -6.824 9.07e-12 ***
std_ElectronAffinity           1.141e+00  4.609e-02  24.755  < 2e-16 ***
wtd_std_ElectronAffinity      -4.943e-01  3.129e-02 -15.795  < 2e-16 ***
mean_FusionHeat                1.413e+00  1.371e-01  10.308  < 2e-16 ***
wtd_mean_FusionHeat           -1.631e+00  1.536e-01 -10.618  < 2e-16 ***
gmean_FusionHeat              -1.318e+00  1.358e-01  -9.707  < 2e-16 ***
wtd_gmean_FusionHeat           1.379e+00  1.518e-01   9.081  < 2e-16 ***
entropy_FusionHeat            -1.870e+01  2.479e+00  -7.540 4.88e-14 ***
wtd_entropy_FusionHeat         2.400e+01  1.715e+00  13.995  < 2e-16 ***
range_FusionHeat              -4.753e-01  3.920e-02 -12.127  < 2e-16 ***
wtd_range_FusionHeat           5.111e-01  5.575e-02   9.168  < 2e-16 ***
wtd_std_FusionHeat             4.594e-01  1.069e-01   4.297 1.74e-05 ***
mean_ThermalConductivity      -1.302e-01  1.508e-02  -8.630  < 2e-16 ***
wtd_mean_ThermalConductivity   5.435e-01  1.935e-02  28.087  < 2e-16 ***
wtd_gmean_ThermalConductivity -3.461e-01  1.243e-02 -27.848  < 2e-16 ***
entropy_ThermalConductivity    1.432e+01  1.421e+00  10.079  < 2e-16 ***
range_ThermalConductivity     -1.006e-01  1.257e-02  -8.008 1.23e-15 ***
wtd_range_ThermalConductivity -2.353e-01  1.368e-02 -17.202  < 2e-16 ***
std_ThermalConductivity        3.443e-01  3.251e-02  10.591  < 2e-16 ***
gmean_Valence                  5.493e+00  7.646e-01   7.184 7.01e-13 ***
wtd_gmean_Valence             -5.579e+00  7.064e-01  -7.897 3.00e-15 ***
entropy_Valence                7.813e+01  9.009e+00   8.672  < 2e-16 ***
wtd_entropy_Valence           -7.540e+01  3.775e+00 -19.975  < 2e-16 ***
range_Valence                  5.796e+00  4.037e-01  14.357  < 2e-16 ***
wtd_std_Valence               -1.939e+01  8.640e-01 -22.444  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.6 on 21198 degrees of freedom
Multiple R-squared:  0.7367,    Adjusted R-squared:  0.7359
F-statistic: 926.8 on 64 and 21198 DF,  p-value: < 2.2e-16
```