Group 08

Activity\_03

###This is our cleaned dataset  
  
cleaned\_data<-Data\_cleaned\_6  
head(cleaned\_data)

## Cement Blast\_Furnace\_Slag Fly\_Ash Water Superplasticizer Coarse\_Aggregate  
## 2 540.0 0.0 0 162 2.5 1055.0  
## 6 266.0 114.0 0 228 0.0 932.0  
## 9 266.0 114.0 0 228 0.0 932.0  
## 11 198.6 132.4 0 192 0.0 978.4  
## 12 198.6 132.4 0 192 0.0 978.4  
## 14 190.0 190.0 0 228 0.0 932.0  
## Fine\_Aggregate Age\_day Concrete\_compressive\_strength  
## 2 676.0 28 61.89  
## 6 670.0 90 47.03  
## 9 670.0 28 45.85  
## 11 825.5 90 38.07  
## 12 825.5 28 28.02  
## 14 670.0 90 42.33

nrow(cleaned\_data)

## [1] 926

#####Install packages#######  
  
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.3.3

##   
## 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

library(tidyr)

## Warning: package 'tidyr' was built under R version 4.3.3

library(MASS)

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

library(Metrics)

## Warning: package 'Metrics' was built under R version 4.3.3

library(ISLR)

## Warning: package 'ISLR' was built under R version 4.3.3

library(leaps)

## Warning: package 'leaps' was built under R version 4.3.3

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.3

library(quantmod)

## Warning: package 'quantmod' was built under R version 4.3.3

## Loading required package: xts

## Warning: package 'xts' was built under R version 4.3.3

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.3.3

##   
## Attaching package: 'zoo'

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

##   
## ######################### Warning from 'xts' package ##########################  
## # #  
## # The dplyr lag() function breaks how base R's lag() function is supposed to #  
## # work, which breaks lag(my\_xts). Calls to lag(my\_xts) that you type or #  
## # source() into this session won't work correctly. #  
## # #  
## # Use stats::lag() to make sure you're not using dplyr::lag(), or you can add #  
## # conflictRules('dplyr', exclude = 'lag') to your .Rprofile to stop #  
## # dplyr from breaking base R's lag() function. #  
## # #  
## # Code in packages is not affected. It's protected by R's namespace mechanism #  
## # Set `options(xts.warn\_dplyr\_breaks\_lag = FALSE)` to suppress this warning. #  
## # #  
## ###############################################################################

##   
## Attaching package: 'xts'

## The following objects are masked from 'package:dplyr':  
##   
## first, last

## Loading required package: TTR

## Warning: package 'TTR' was built under R version 4.3.3

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

library(corrplot)

## Warning: package 'corrplot' was built under R version 4.3.3

## corrplot 0.92 loaded

library(caTools)

## Warning: package 'caTools' was built under R version 4.3.3

library(car)

## Warning: package 'car' was built under R version 4.3.3

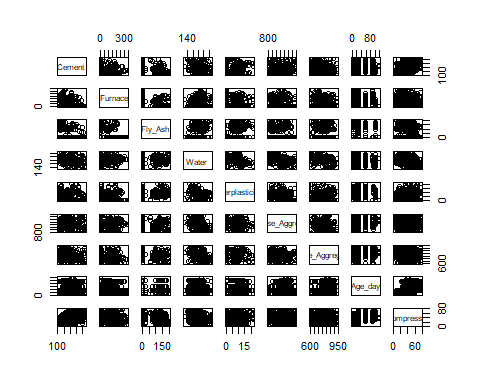
## Loading required package: carData

## Warning: package 'carData' was built under R version 4.3.3

##   
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':  
##   
## recode

###Scatter plots  
pairs(cleaned\_data)



###Correlations  
cor(cleaned\_data)

## Cement Blast\_Furnace\_Slag Fly\_Ash  
## Cement 1.00000000 -0.26149938 -0.36726110  
## Blast\_Furnace\_Slag -0.26149938 1.00000000 -0.35142143  
## Fly\_Ash -0.36726110 -0.35142143 1.00000000  
## Water -0.13551917 0.10569333 -0.22696940  
## Superplasticizer 0.05081810 0.05161892 0.44474205  
## Coarse\_Aggregate -0.09251497 -0.28998911 -0.04694758  
## Fine\_Aggregate -0.21606995 -0.29829372 0.03028820  
## Age\_day -0.04225654 -0.03948894 0.06075860  
## Concrete\_compressive\_strength 0.47948503 0.14567612 -0.05453121  
## Water Superplasticizer Coarse\_Aggregate  
## Cement -0.13551917 0.05081810 -0.09251497  
## Blast\_Furnace\_Slag 0.10569333 0.05161892 -0.28998911  
## Fly\_Ash -0.22696940 0.44474205 -0.04694758  
## Water 1.00000000 -0.63365330 -0.19672455  
## Superplasticizer -0.63365330 1.00000000 -0.23831335  
## Coarse\_Aggregate -0.19672455 -0.23831335 1.00000000  
## Fine\_Aggregate -0.29379741 0.07247323 -0.21248995  
## Age\_day -0.03803409 0.05297613 0.02511620  
## Concrete\_compressive\_strength -0.39214102 0.40835045 -0.16832181  
## Fine\_Aggregate Age\_day  
## Cement -0.21606995 -0.04225654  
## Blast\_Furnace\_Slag -0.29829372 -0.03948894  
## Fly\_Ash 0.03028820 0.06075860  
## Water -0.29379741 -0.03803409  
## Superplasticizer 0.07247323 0.05297613  
## Coarse\_Aggregate -0.21248995 0.02511620  
## Fine\_Aggregate 1.00000000 0.06241512  
## Age\_day 0.06241512 1.00000000  
## Concrete\_compressive\_strength -0.16321125 0.52132097  
## Concrete\_compressive\_strength  
## Cement 0.47948503  
## Blast\_Furnace\_Slag 0.14567612  
## Fly\_Ash -0.05453121  
## Water -0.39214102  
## Superplasticizer 0.40835045  
## Coarse\_Aggregate -0.16832181  
## Fine\_Aggregate -0.16321125  
## Age\_day 0.52132097  
## Concrete\_compressive\_strength 1.00000000

###All predictors  
  
lm\_0<-lm(Concrete\_compressive\_strength~.,data = cleaned\_data)  
summary(lm\_0)

##   
## Call:  
## lm(formula = Concrete\_compressive\_strength ~ ., data = cleaned\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -24.2771 -4.8401 -0.3423 5.3339 31.0682   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 50.667660 22.544478 2.247 0.0248 \*   
## Cement 0.101772 0.006980 14.580 < 2e-16 \*\*\*  
## Blast\_Furnace\_Slag 0.075480 0.008508 8.872 < 2e-16 \*\*\*  
## Fly\_Ash 0.047830 0.010420 4.590 5.04e-06 \*\*\*  
## Water -0.249718 0.034826 -7.170 1.54e-12 \*\*\*  
## Superplasticizer 0.214624 0.085112 2.522 0.0118 \*   
## Coarse\_Aggregate -0.010561 0.007881 -1.340 0.1805   
## Fine\_Aggregate -0.010882 0.009129 -1.192 0.2336   
## Age\_day 0.318878 0.009303 34.275 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.979 on 917 degrees of freedom  
## Multiple R-squared: 0.7741, Adjusted R-squared: 0.7722   
## F-statistic: 392.9 on 8 and 917 DF, p-value: < 2.2e-16

**Interpretation**

**The model was fitted using the formula:** Concrete\_compressive\_strength∼Cement+Blast\_Furnace\_Slag+Fly\_Ash+Water+Superplasticizer+Coarse\_Aggregate+Fine\_Aggregate+Age\_day

**Residuals**

The residuals summary indicates the distribution of the differences between the observed and predicted values:

* Minimum Residual: -24.2771
* 1st Quartile: -4.8401
* Median: -0.3423
* 3rd Quartile: 5.3339
* Maximum Residual: 31.0682

**Coefficients**

The coefficients table provides estimates for each predictor along with their standard errors, t-values, and p-values:

1. **Intercept**: 50.667660
   * Interpretation: When all predictors are at zero, the estimated concrete compressive strength is 50.67 MPa.
   * Significance: p-value = 0.0248, indicating significance at the 5% level.
2. **Cement**: 0.101772
   * Interpretation: For each unit increase in cement, the concrete compressive strength increases by approximately 0.1018 MPa.
   * Significance: Highly significant (p-value < 2e-16).
3. **Blast\_Furnace\_Slag**: 0.075480
   * Interpretation: For each unit increase in blast furnace slag, the concrete compressive strength increases by approximately 0.0755 MPa.
   * Significance: Highly significant (p-value < 2e-16).
4. **Fly\_Ash**: 0.047830
   * Interpretation: For each unit increase in fly ash, the concrete compressive strength increases by approximately 0.0478 MPa.
   * Significance: Highly significant (p-value < 5.04e-06).
5. **Water**: -0.249718
   * Interpretation: For each unit increase in water, the concrete compressive strength decreases by approximately 0.2497 MPa.
   * Significance: Highly significant (p-value < 1.54e-12).
6. **Superplasticizer**: 0.214624
   * Interpretation: For each unit increase in superplasticizer, the concrete compressive strength increases by approximately 0.2146 MPa.
   * Significance: Significant (p-value = 0.0118).
7. **Coarse\_Aggregate**: -0.010561
   * Interpretation: For each unit increase in coarse aggregate, the concrete compressive strength decreases by approximately 0.0106 MPa.
   * Significance: Not significant (p-value = 0.1805).
8. **Fine\_Aggregate**: -0.010882
   * Interpretation: For each unit increase in fine aggregate, the concrete compressive strength decreases by approximately 0.0109 MPa.
   * Significance: Not significant (p-value = 0.2336).
9. **Age\_day**: 0.318878
   * Interpretation: For each day increase in the age of the concrete, the compressive strength increases by approximately 0.3189 MPa.
   * Significance: Highly significant (p-value < 2e-16).

**Significance Codes**

* \*\*\* denotes highly significant (p < 0.001)
* \*\* denotes significant (0.001 <= p < 0.01)
* \* denotes moderately significant (0.01 <= p < 0.05)
* . denotes marginally significant (0.05 <= p < 0.1)
* No symbol indicates not significant (p >= 0.1)

**Model Fit**

* **Residual Standard Error**: 7.979 on 917 degrees of freedom
* **Multiple R-squared**: 0.7741
* **Adjusted R-squared**: 0.7722
* **F-statistic**: 392.9 on 8 and 917 degrees of freedom
* **Overall Model Significance**: The model is highly significant with a p-value < 2.2e-16, indicating that the predictors jointly have a significant effect on the concrete compressive strength.

The model explains about 77.41% of the variability in concrete compressive strength (R-squared = 0.7741). The significant predictors include cement, blast furnace slag, fly ash, water, superplasticizer, and age of the concrete. Coarse and fine aggregates were not significant predictors in this model. The model's overall F-test indicates it is a good fit for the data.

###Anova test  
  
anova(lm\_0)

## Analysis of Variance Table  
##   
## Response: Concrete\_compressive\_strength  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Cement 1 59427 59427 933.4365 < 2.2e-16 \*\*\*  
## Blast\_Furnace\_Slag 1 20386 20386 320.2074 < 2.2e-16 \*\*\*  
## Fly\_Ash 1 25197 25197 395.7750 < 2.2e-16 \*\*\*  
## Water 1 18760 18760 294.6748 < 2.2e-16 \*\*\*  
## Superplasticizer 1 1378 1378 21.6523 3.749e-06 \*\*\*  
## Coarse\_Aggregate 1 0 0 0.0019 0.9652   
## Fine\_Aggregate 1 162 162 2.5388 0.1114   
## Age\_day 1 74793 74793 1174.7940 < 2.2e-16 \*\*\*  
## Residuals 917 58380 64   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

**Interpretation**

* **Cement**:
  + Degrees of Freedom (Df): 1
  + Sum of Squares (Sum Sq): 59427
  + Mean Squares (Mean Sq): 59427
  + F value: 933.4365
  + p-value: < 2.2e-16
  + Interpretation: The effect of cement on concrete compressive strength is highly significant, contributing significantly to the model.
* **Blast\_Furnace\_Slag**:
  + Df: 1
  + Sum Sq: 20386
  + Mean Sq: 20386
  + F value: 320.2074
  + p-value: < 2.2e-16
  + Interpretation: The effect of blast furnace slag on concrete compressive strength is highly significant.
* **Fly\_Ash**:
  + Df: 1
  + Sum Sq: 25197
  + Mean Sq: 25197
  + F value: 395.7750
  + p-value: < 2.2e-16
  + Interpretation: The effect of fly ash on concrete compressive strength is highly significant.
* **Water**:
  + Df: 1
  + Sum Sq: 18760
  + Mean Sq: 18760
  + F value: 294.6748
  + p-value: < 2.2e-16
  + Interpretation: The effect of water on concrete compressive strength is highly significant.
* **Superplasticizer**:
  + Df: 1
  + Sum Sq: 1378
  + Mean Sq: 1378
  + F value: 21.6523
  + p-value: 3.749e-06
  + Interpretation: The effect of superplasticizer on concrete compressive strength is significant.
* **Coarse\_Aggregate**:
  + Df: 1
  + Sum Sq: 0
  + Mean Sq: 0
  + F value: 0.0019
  + p-value: 0.9652
  + Interpretation: The effect of coarse aggregate on concrete compressive strength is not significant.
* **Fine\_Aggregate**:
  + Df: 1
  + Sum Sq: 162
  + Mean Sq: 162
  + F value: 2.5388
  + p-value: 0.1114
  + Interpretation: The effect of fine aggregate on concrete compressive strength is not significant.
* **Age\_day**:
  + Df: 1
  + Sum Sq: 74793
  + Mean Sq: 74793
  + F value: 1174.7940
  + p-value: < 2.2e-16
  + Interpretation: The effect of age on concrete compressive strength is highly significant.

**Residuals**

* Degrees of Freedom: 917
* Sum of Squares: 58380
* Mean Squares: 64

The ANOVA table confirms that the predictors Cement, Blast Furnace Slag, Fly Ash, Water, Superplasticizer, and Age\_day significantly affect concrete compressive strength, as indicated by their low p-values (< 0.05). The predictors Coarse Aggregate and Fine Aggregate do not have significant effects on the response variable in this model. The high F-values for significant predictors suggest a strong relationship between these predictors and the concrete compressive strength.

########Model Selection###########  
  
######Forward stepwise Selection#########  
fit\_frwd<-regsubsets(Concrete\_compressive\_strength~ .,data = cleaned\_data,nvmax = 8,method = "forward")  
frwd\_summary<-summary(fit\_frwd)  
frwd\_summary

## Subset selection object  
## Call: regsubsets.formula(Concrete\_compressive\_strength ~ ., data = cleaned\_data,   
## nvmax = 8, method = "forward")  
## 8 Variables (and intercept)  
## Forced in Forced out  
## Cement FALSE FALSE  
## Blast\_Furnace\_Slag FALSE FALSE  
## Fly\_Ash FALSE FALSE  
## Water FALSE FALSE  
## Superplasticizer FALSE FALSE  
## Coarse\_Aggregate FALSE FALSE  
## Fine\_Aggregate FALSE FALSE  
## Age\_day FALSE FALSE  
## 1 subsets of each size up to 8  
## Selection Algorithm: forward  
## Cement Blast\_Furnace\_Slag Fly\_Ash Water Superplasticizer  
## 1 ( 1 ) " " " " " " " " " "   
## 2 ( 1 ) "\*" " " " " " " " "   
## 3 ( 1 ) "\*" " " " " " " "\*"   
## 4 ( 1 ) "\*" "\*" " " " " "\*"   
## 5 ( 1 ) "\*" "\*" " " "\*" "\*"   
## 6 ( 1 ) "\*" "\*" "\*" "\*" "\*"   
## 7 ( 1 ) "\*" "\*" "\*" "\*" "\*"   
## 8 ( 1 ) "\*" "\*" "\*" "\*" "\*"   
## Coarse\_Aggregate Fine\_Aggregate Age\_day  
## 1 ( 1 ) " " " " "\*"   
## 2 ( 1 ) " " " " "\*"   
## 3 ( 1 ) " " " " "\*"   
## 4 ( 1 ) " " " " "\*"   
## 5 ( 1 ) " " " " "\*"   
## 6 ( 1 ) " " " " "\*"   
## 7 ( 1 ) "\*" " " "\*"   
## 8 ( 1 ) "\*" "\*" "\*"

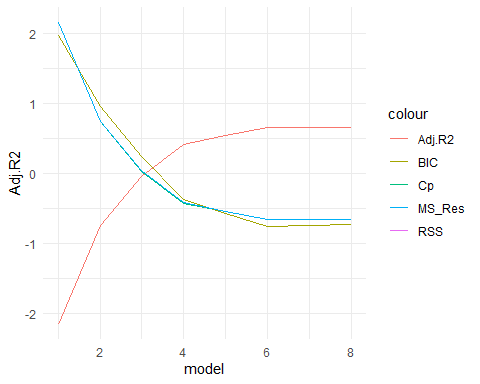
##########Find Adj R^2,Cp,BIC,RSS values  
criterion<-data.frame(model=1:8,  
 Adj.R2=(frwd\_summary$adjr2),  
 Cp=(frwd\_summary$cp),  
 BIC=(frwd\_summary$bic),  
 RSS=(frwd\_summary$rss))  
##Add MS\_Res  
n<-nrow(cleaned\_data)  
criterion$MS\_Res<-criterion$RSS/(n-criterion$model-1)  
  
  
##Standardize  
criterion\_std<-cbind(model=criterion$model,scale(criterion[,-1]))  
criterion\_std<-as.data.frame(criterion\_std)  
  
###Values  
criterion\_std

## model Adj.R2 Cp BIC RSS MS\_Res  
## 1 1 -2.16063154 2.16269269 1.9713194 2.16000747 2.16063154  
## 2 2 -0.74498413 0.74272430 0.9651614 0.74539192 0.74498413  
## 3 3 -0.03653678 0.03358656 0.2414418 0.03754295 0.03653678  
## 4 4 0.42127842 -0.42345558 -0.3789904 -0.41965173 -0.42127842  
## 5 5 0.54734332 -0.54821230 -0.5728855 -0.54646078 -0.54734332  
## 6 6 0.65821831 -0.65754292 -0.7578819 -0.65793194 -0.65821831  
## 7 7 0.65736245 -0.65529905 -0.7411093 -0.65846651 -0.65736245  
## 8 8 0.65794994 -0.65449370 -0.7270554 -0.66043137 -0.65794994

##Interpretation

* The procedure starts with no predictor variable in the model.
* At each step, the algorithm evaluates the impact of adding each predictor variable to the current model. It selects the variable that improves the model the most based on a chosen criterion, typically the adjusted R^2, BIC, Cp, and RSS.
* Each predictor is listed, and the “Force in” and “Force out” columns show whether the variables are included or excluded at each step.
* This process repeated, with one variable being added at each step until all variables are considered or no additional variables improve the model sufficiently.
* The first model includes only “Age\_day” because its adjusted R^2 has the highest value and BIC, Cp, and RSS have the lowest values.
* The second model includes “Age\_day” and “Cement” variables.
* The third model includes “Age\_day”, “Cement” and “Superplasticizer” variables.
* The fourth model includes “Age\_day”, “Cement”, “Superplasticizer”, and “Blast\_Furnace\_Slag” variables.
* The fifth model includes “Age\_day”, “Cement”, “Superplasticizer”, “Blast\_Furnace\_Slag” and “Water” variables.
* The sixth model includes “Age\_day”, “Cement”, “Superplasticizer”, “Blast\_Furnace\_Slag”, “Water” and “Fly\_Ash” variables.
* The seventh model includes “Age\_day”, “Cement”, “Superplasticizer”, “Blast\_Furnace\_Slag”, “Water”, “Fly\_Ash” and “Coarse\_Aggregate” variables.
* The eighth model includes “Age\_day”, “Cement”, “Superplasticizer”, “Blast\_Furnace\_Slag”, “Water”, “Fly\_Ash”, “Coarse\_Aggregate” and “Fine\_Aggregate” variables.
* There are two methods to find the better model
  1. Formal model : Using the values of above criteria, the better model should have maximum value for adjusted R^2 and minimum values for Cp, BIC, RSS and MS\_Res. Therefore according to the above values 6th model is the better model.
  2. Graphical method

###Graphically  
ggplot(criterion\_std,aes(model))+  
 geom\_line(aes(y=Adj.R2,colour="Adj.R2"))+  
 geom\_line(aes(y=Cp,colour="Cp"))+  
 geom\_line(aes(y=BIC,colour="BIC"))+  
 geom\_line(aes(y=RSS,colour="RSS"))+  
 geom\_line(aes(y=MS\_Res,colour="MS\_Res"))+  
 theme\_minimal()



* According to the plot, 6th model has maximum value for adjusted R^2 and minimum values for Cp, BIC, RSS and MS\_Res. Therefore 6th model seems a better model.

##coefficients  
  
coef(fit\_frwd,8)

## (Intercept) Cement Blast\_Furnace\_Slag Fly\_Ash   
## 50.66765975 0.10177234 0.07547985 0.04782991   
## Water Superplasticizer Coarse\_Aggregate Fine\_Aggregate   
## -0.24971813 0.21462388 -0.01056139 -0.01088221   
## Age\_day   
## 0.31887795

better\_frwd<-lm(Concrete\_compressive\_strength~Cement+Blast\_Furnace\_Slag+Fly\_Ash+Water+Superplasticizer+Coarse\_Aggregate+ Fine\_Aggregate +Age\_day,data = cleaned\_data,method = "forward")

## Warning in lm(Concrete\_compressive\_strength ~ Cement + Blast\_Furnace\_Slag + :  
## method = 'forward' is not supported. Using 'qr'

summary(better\_frwd)

##   
## Call:  
## lm(formula = Concrete\_compressive\_strength ~ Cement + Blast\_Furnace\_Slag +   
## Fly\_Ash + Water + Superplasticizer + Coarse\_Aggregate + Fine\_Aggregate +   
## Age\_day, data = cleaned\_data, method = "forward")  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -24.2771 -4.8401 -0.3423 5.3339 31.0682   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 50.667660 22.544478 2.247 0.0248 \*   
## Cement 0.101772 0.006980 14.580 < 2e-16 \*\*\*  
## Blast\_Furnace\_Slag 0.075480 0.008508 8.872 < 2e-16 \*\*\*  
## Fly\_Ash 0.047830 0.010420 4.590 5.04e-06 \*\*\*  
## Water -0.249718 0.034826 -7.170 1.54e-12 \*\*\*  
## Superplasticizer 0.214624 0.085112 2.522 0.0118 \*   
## Coarse\_Aggregate -0.010561 0.007881 -1.340 0.1805   
## Fine\_Aggregate -0.010882 0.009129 -1.192 0.2336   
## Age\_day 0.318878 0.009303 34.275 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.979 on 917 degrees of freedom  
## Multiple R-squared: 0.7741, Adjusted R-squared: 0.7722   
## F-statistic: 392.9 on 8 and 917 DF, p-value: < 2.2e-16

**Interpretation**

* **(Intercept)**: 50.66765975
  + The intercept represents the expected value of the dependent variable (concrete compressive strength) when all predictors are zero. In this context, it means that if all other variables (cement, blast furnace slag, etc.) are zero, the concrete compressive strength would be approximately 50.67 units.

#### **Predictors:**

1. **Cement**: 0.10177234

* For each unit increase in the amount of cement, the concrete compressive strength increases by approximately 0.10 units, holding all other variables constant. This coefficient is statistically significant (p < 2e-16).

1. **Blast\_Furnace\_Slag**: 0.07549785

* For each unit increase in the amount of blast furnace slag, the concrete compressive strength increases by approximately 0.08 units, holding all other variables constant. This coefficient is statistically significant (p < 2e-16).

1. **Fly\_Ash**: 0.04782991

* For each unit increase in the amount of fly ash, the concrete compressive strength increases by approximately 0.05 units, holding all other variables constant. This coefficient is statistically significant (p < 2e-16).

1. **Water**: -0.24971813

* For each unit increase in the amount of water, the concrete compressive strength decreases by approximately 0.25 units, holding all other variables constant. This coefficient is statistically significant (p = 1.54e-12).

1. **Superplasticizer**: 0.21462388

* For each unit increase in the amount of superplasticizer, the concrete compressive strength increases by approximately 0.21 units, holding all other variables constant. This coefficient is statistically significant (p < 2e-16).

1. **Coarse\_Aggregate**: -0.01056319

* For each unit increase in the amount of coarse aggregate, the concrete compressive strength decreases by approximately 0.01 units, holding all other variables constant. This coefficient is not statistically significant (p = 0.1805).

1. **Fine\_Aggregate**: -0.01088822

* For each unit increase in the amount of fine aggregate, the concrete compressive strength decreases by approximately 0.01 units, holding all other variables constant. This coefficient is not statistically significant (p = 0.1868).

1. **Age\_day**: 0.31887795

* For each unit increase in the age of the concrete, the compressive strength increases by approximately 0.32 units, holding all other variables constant. This coefficient is statistically significant (p < 2e-16).

### Model Summary:

* **Residual Standard Error**: 7.979 on 917 degrees of freedom

This represents the average distance that the observed values fall from the regression line.

* **Multiple R-squared**: 0.7741

This indicates that approximately 77.41% of the variability in the concrete compressive strength can be explained by the model.

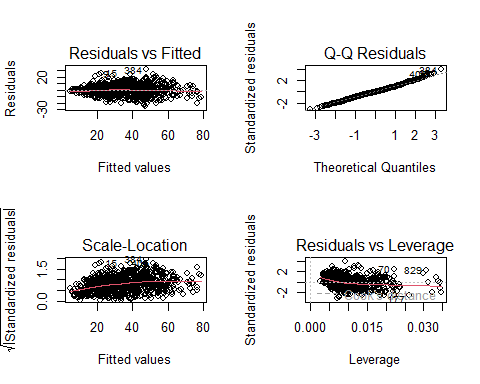
* **Adjusted R-squared**: 0.7722

This value adjusts the R-squared for the number of predictors in the model, indicating that about 77.22% of the variability is explained when adjusting for the number of predictors.

* **F-statistic**: 392.9 on 8 and 917 df, p-value: < 2.2e-16

This tests the overall significance of the model. A p-value less than 0.05 indicates that the model is statistically significant, and at least one of the predictors is significantly related to the dependent variable.

#####Model Diagnostics#####  
  
par(mfrow=c(2,2))  
plot(better\_frwd)



**Interpretation**

Residuals vs Fitted

* Purpose: To check the linearity and homoscedasticity assumptions.
* Ideally, the residuals should be randomly scattered around the horizontal line at 0 without any clear pattern.
* There is a noticeable pattern, suggesting possible issues with non-linearity or heteroscedasticity (non-constant variance of errors).

Normal Q-Q

* Purpose: To check if residuals are normally distributed.
* The points should fall approximately along the reference line.
* The plot shows that most points fall along the line, but there are some deviations at the tails. This indicates that the residuals are roughly normally distributed but with some potential outliers.

Scale-Location (or Spread-Location)

* Purpose: To check homoscedasticity (equal variance of residuals).
* The plot should show a horizontal line with points scattered around it.
* In the plot, there is a slight pattern where the spread of residuals increases with fitted values, suggesting heteroscedasticity.

Residuals vs Leverage

* Purpose: To identify influential data points.
* Points that stand out with high leverage and high residuals can be influential.
* This plot indicates a few points that might be influential.

##########Multicollinearity#############  
  
vif\_values<-vif(better\_frwd)  
vif\_values

## Cement Blast\_Furnace\_Slag Fly\_Ash Water   
## 7.287089 7.839676 6.553248 5.643127   
## Superplasticizer Coarse\_Aggregate Fine\_Aggregate Age\_day   
## 2.983278 5.501360 6.270621 1.015195

* **VIF < 1**: Indicates no correlation among predictors.
* **1 < VIF < 5**: Indicates moderate correlation but is usually acceptable.
* **VIF > 5**: Indicates high correlation, suggesting potential multicollinearity issues.
* **VIF > 10**: Indicates very high correlation and multicollinearity problem that needs to be addressed.

**Interpretation**

 **Cement**: 7.287089

* High correlation with other predictors.

 **Blast\_Furnace\_Slag**: 7.839676

* High correlation with other predictors.

 **Fly\_Ash**: 6.553248

* High correlation with other predictors.

 **Water**: 5.643127

* High correlation with other predictors.

 **Superplasticizer**: 2.983278

* Moderate correlation, generally acceptable.

 **Coarse\_Aggregate**: 5.501360

* High correlation with other predictors.

 **Fine\_Aggregate**: 6.270621

* High correlation with other predictors.

 **Age\_days**: 1.015195

* Very low correlation, indicating no multicollinearity issues.

###Better model remove the insignificant variable Fine aggregate  
better\_model\_1<-lm(Concrete\_compressive\_strength~Cement+Blast\_Furnace\_Slag+Fly\_Ash+Water+Superplasticizer+Coarse\_Aggregate +Age\_day,data = cleaned\_data)  
summary(better\_model\_1)

##   
## Call:  
## lm(formula = Concrete\_compressive\_strength ~ Cement + Blast\_Furnace\_Slag +   
## Fly\_Ash + Water + Superplasticizer + Coarse\_Aggregate + Age\_day,   
## data = cleaned\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -24.4906 -4.8389 -0.2105 5.1729 31.3343   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 25.342860 7.544678 3.359 0.000814 \*\*\*  
## Cement 0.108906 0.003594 30.306 < 2e-16 \*\*\*  
## Blast\_Furnace\_Slag 0.084208 0.004333 19.433 < 2e-16 \*\*\*  
## Fly\_Ash 0.057468 0.006573 8.743 < 2e-16 \*\*\*  
## Water -0.217373 0.021836 -9.955 < 2e-16 \*\*\*  
## Superplasticizer 0.234803 0.083430 2.814 0.004992 \*\*   
## Coarse\_Aggregate -0.002606 0.004192 -0.622 0.534349   
## Age\_day 0.317978 0.009275 34.284 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.981 on 918 degrees of freedom  
## Multiple R-squared: 0.7738, Adjusted R-squared: 0.7721   
## F-statistic: 448.6 on 7 and 918 DF, p-value: < 2.2e-16

#####Interpretation Remove the only Fine Aggregate but now Coarse aggregate is insignificant in the model.(6)

###Better model remove the insignificant variable Coarse aggregate  
better\_model\_2<-lm(Concrete\_compressive\_strength~Cement+Blast\_Furnace\_Slag+Fly\_Ash+Water+Superplasticizer+Fine\_Aggregate+Age\_day,data = cleaned\_data)  
summary(better\_model\_2)

##   
## Call:  
## lm(formula = Concrete\_compressive\_strength ~ Cement + Blast\_Furnace\_Slag +   
## Fly\_Ash + Water + Superplasticizer + Fine\_Aggregate + Age\_day,   
## data = cleaned\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -24.5235 -4.8902 -0.2229 5.1699 31.1556   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 21.9593341 7.0284656 3.124 0.00184 \*\*   
## Cement 0.1092968 0.0041492 26.342 < 2e-16 \*\*\*  
## Blast\_Furnace\_Slag 0.0847909 0.0049120 17.262 < 2e-16 \*\*\*  
## Fly\_Ash 0.0577973 0.0073004 7.917 7e-15 \*\*\*  
## Water -0.2120130 0.0205341 -10.325 < 2e-16 \*\*\*  
## Superplasticizer 0.2532179 0.0801261 3.160 0.00163 \*\*   
## Fine\_Aggregate -0.0005214 0.0048571 -0.107 0.91454   
## Age\_day 0.3179101 0.0092794 34.260 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.982 on 918 degrees of freedom  
## Multiple R-squared: 0.7737, Adjusted R-squared: 0.772   
## F-statistic: 448.4 on 7 and 918 DF, p-value: < 2.2e-16

#####Interpretation Remove the only Coarse Aggregate but now Fine aggregate is insignificant in the model.(7)

###Better model remove the insignificant variable Fine aggregate and Coarse Aggregate  
better\_model\_final<-lm(Concrete\_compressive\_strength~Cement+Blast\_Furnace\_Slag+Fly\_Ash+Water+Superplasticizer+Age\_day,data = cleaned\_data)  
summary(better\_model\_final)

##   
## Call:  
## lm(formula = Concrete\_compressive\_strength ~ Cement + Blast\_Furnace\_Slag +   
## Fly\_Ash + Water + Superplasticizer + Age\_day, data = cleaned\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -24.5282 -4.9021 -0.2091 5.1402 31.1895   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 21.332558 3.910394 5.455 6.29e-08 \*\*\*  
## Cement 0.109545 0.003442 31.823 < 2e-16 \*\*\*  
## Blast\_Furnace\_Slag 0.085081 0.004098 20.763 < 2e-16 \*\*\*  
## Fly\_Ash 0.058158 0.006477 8.979 < 2e-16 \*\*\*  
## Water -0.211345 0.019558 -10.806 < 2e-16 \*\*\*  
## Superplasticizer 0.251711 0.078845 3.192 0.00146 \*\*   
## Age\_day 0.317881 0.009270 34.290 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.978 on 919 degrees of freedom  
## Multiple R-squared: 0.7737, Adjusted R-squared: 0.7722   
## F-statistic: 523.7 on 6 and 919 DF, p-value: < 2.2e-16

###Interpretation Now all are the significant.(8)

\*Residual Standard Error (RSE)

This indicates that the average distance of the observed values from the predicted values is about 7.978 units of Concrete Compressive Strength. A lower RSE suggests a better fit of the model to the data.

\*R-Squared

Approximately 77.37% of the variability in Concrete Compressive Strength can be explained by the model that includes cement, Blast furnace slag, fly ash, water, superplasticizer, age day.

\*Adjusted R-Squared

An adjusted R-Squared of 77.22% indicates that the model still explains a substantial amount of variance in Concrete Compressive Strength consider the number of predictors.

H0: The regression model is not significant.

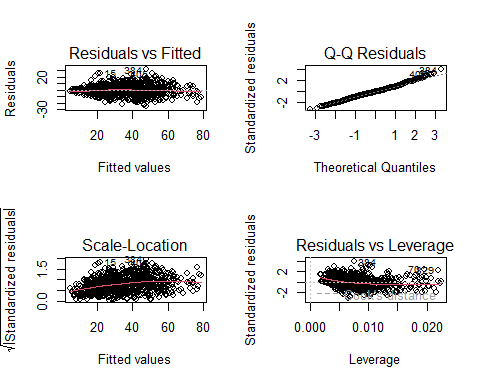
H1: The regression model is significant.

F0.05, 6, 919 = 2.108

F-Statistic = 523.7 > 2.108

We have sufficient evidence to reject the null hypothesis. Concluding that at least one of the predictors in the model significantly contributes to explaining the variability the variability in Concrete Compressive Strength.

########Residual plots for better model######  
  
par(mfrow=c(2,2))  
plot(better\_model\_final)



**Interpretation**

** Residuals vs. Fitted Values (Top Left)**

* Purpose: To check for non-linearity, unequal error variances (heteroscedasticity), and outliers.
* Interpretation: The residuals appear to be randomly scattered around the horizontal axis (y=0), suggesting that the linearity assumption is reasonable. However, the spread of residuals increases slightly as the fitted values increase, indicating some heteroscedasticity. This means that the variance of the residuals may not be constant across all levels of the fitted values.

** Normal Q-Q Plot (Top Right)**

* Purpose: To check if the residuals are normally distributed.
* Interpretation: The points on the Q-Q plot fall approximately along the reference line, suggesting that the residuals are roughly normally distributed. Some deviations at the tails might be observed, but they do not appear to be severe.

** Scale-Location Plot (Bottom Left)**

* Purpose: To check for homoscedasticity (constant variance of residuals).
* Interpretation: The plot shows the square root of the standardized residuals against the fitted values. The residuals should be spread equally along the range of predictors for homoscedasticity. The slightly increasing trend suggests some degree of heteroscedasticity, meaning that the variance of the residuals might be increasing with the fitted values.

** Residuals vs. Leverage Plot (Bottom Right)**

* Purpose: To identify influential data points that might unduly influence the model.
* Interpretation: This plot helps to find points with high leverage (influential points). The presence of points outside the red dashed lines (Cook’s distance) indicates influential observations. In this plot, there do not appear to be any points that are significantly outside these lines, suggesting no highly influential data points.