Descriptive Analysis

GROUP\_03

2024-07-28

#Import the dataset(Copy of Concrete\_Data)  
  
Data<-read.csv("C:\\Users\\User\\OneDrive\\Desktop\\Copy of Concrete\_Data.csv")  
  
colnames(Data)<-c("Cement","Blast\_Furnace\_Slag","Fly\_Ash","Water","Superplasticizer","Coarse\_Aggregate","Fine\_Aggregate","Age\_day","Concrete\_compressive\_strength")  
  
names(Data)

## [1] "Cement" "Blast\_Furnace\_Slag"   
## [3] "Fly\_Ash" "Water"   
## [5] "Superplasticizer" "Coarse\_Aggregate"   
## [7] "Fine\_Aggregate" "Age\_day"   
## [9] "Concrete\_compressive\_strength"

head(Data)

## Cement Blast\_Furnace\_Slag Fly\_Ash Water Superplasticizer Coarse\_Aggregate  
## 1 540.0 0.0 0 162 2.5 1040.0  
## 2 540.0 0.0 0 162 2.5 1055.0  
## 3 332.5 142.5 0 228 0.0 932.0  
## 4 332.5 142.5 0 228 0.0 932.0  
## 5 198.6 132.4 0 192 0.0 978.4  
## 6 266.0 114.0 0 228 0.0 932.0  
## Fine\_Aggregate Age\_day Concrete\_compressive\_strength  
## 1 676.0 28 79.99  
## 2 676.0 28 61.89  
## 3 594.0 270 40.27  
## 4 594.0 365 41.05  
## 5 825.5 360 44.30  
## 6 670.0 90 47.03

###Missing values  
  
print("Total of Missing values")

## [1] "Total of Missing values"

sum(is.na(Data))

## [1] 0

colSums(is.na(Data))

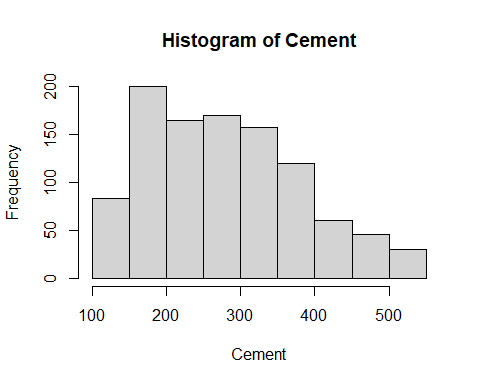
## Cement Blast\_Furnace\_Slag   
## 0 0   
## Fly\_Ash Water   
## 0 0   
## Superplasticizer Coarse\_Aggregate   
## 0 0   
## Fine\_Aggregate Age\_day   
## 0 0   
## Concrete\_compressive\_strength   
## 0

This data set has no missing values.

##########Descriptive Analysis##########  
  
#################Cement############  
  
##Summary  
  
summary(Data$Cement)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 102.0 192.4 272.9 281.2 350.0 540.0

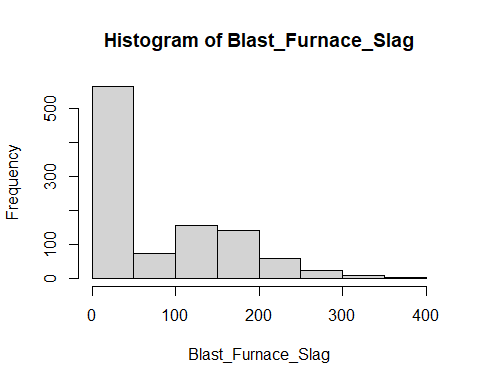
##Histogram  
hist(Data$Cement,xlab = "Cement",main="Histogram of Cement")



#################Blast\_Furnace\_Slag############  
  
##Summary  
  
summary(Data$Blast\_Furnace\_Slag)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0 0.0 22.0 73.9 142.9 359.4

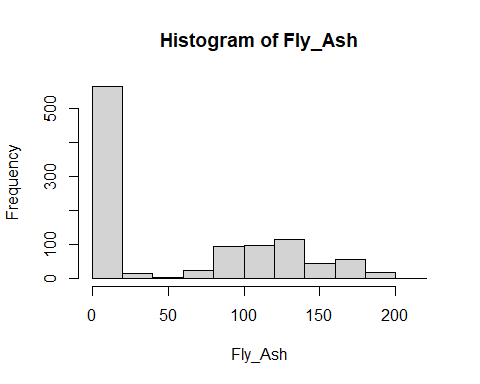
##Histogram  
hist(Data$Blast\_Furnace\_Slag,xlab = "Blast\_Furnace\_Slag",main="Histogram of Blast\_Furnace\_Slag")



#################Fly\_Ash############  
  
##Summary  
  
summary(Data$Fly\_Ash)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 0.00 0.00 54.19 118.30 200.10

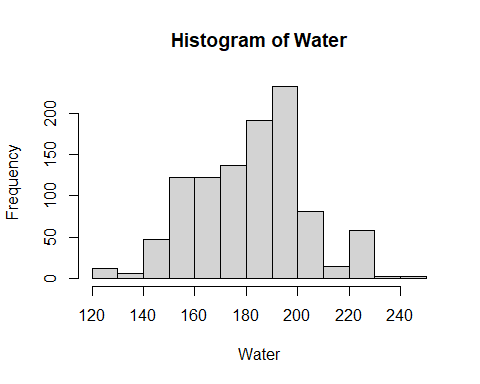
##Histogram  
hist(Data$Fly\_Ash,xlab = "Fly\_Ash",main="Histogram of Fly\_Ash")



#################Water############  
  
##Summary  
  
summary(Data$Water)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 121.8 164.9 185.0 181.6 192.0 247.0

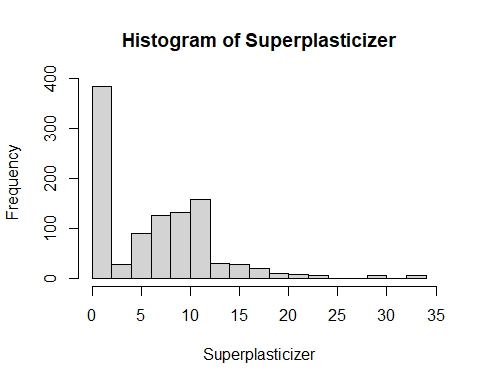
##Histogram  
hist(Data$Water,xlab = "Water",main="Histogram of Water")



###############Superplasticizer############  
  
##Summary  
  
summary(Data$Superplasticizer)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 0.000 6.400 6.205 10.200 32.200

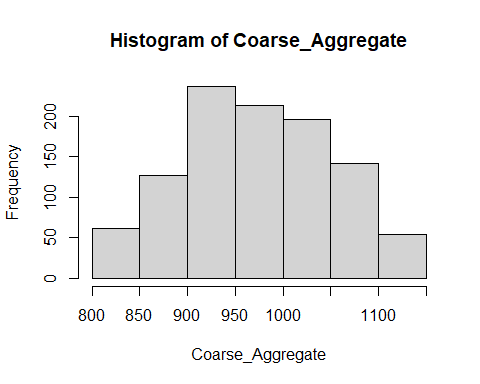
##Histogram  
hist(Data$Superplasticizer,xlab = "Superplasticizer",main="Histogram of Superplasticizer")



###############Coarse\_Aggregate############  
  
##Summary  
  
summary(Data$Coarse\_Aggregate)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 801.0 932.0 968.0 972.9 1029.4 1145.0

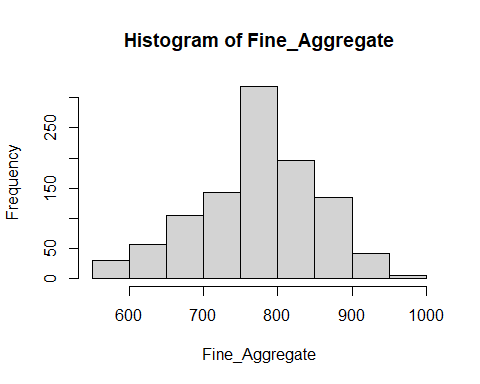
##Histogram  
hist(Data$Coarse\_Aggregate,xlab = "Coarse\_Aggregate",main="Histogram of Coarse\_Aggregate")



###############Fine\_Aggregate############  
  
##Summary  
  
summary(Data$Fine\_Aggregate)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 594.0 731.0 779.5 773.6 824.0 992.6

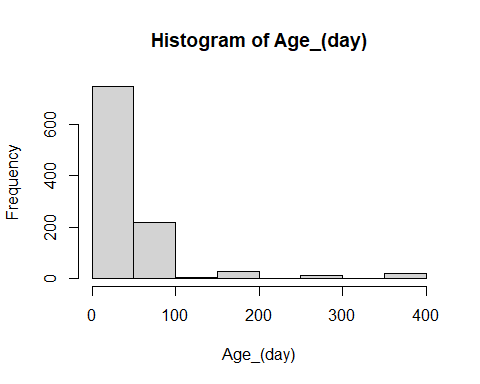
##Histogram  
hist(Data$Fine\_Aggregate,xlab = "Fine\_Aggregate",main="Histogram of Fine\_Aggregate")



###############`Age\_day`############  
  
##Summary  
  
summary(Data$Age\_day)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.00 7.00 28.00 45.66 56.00 365.00

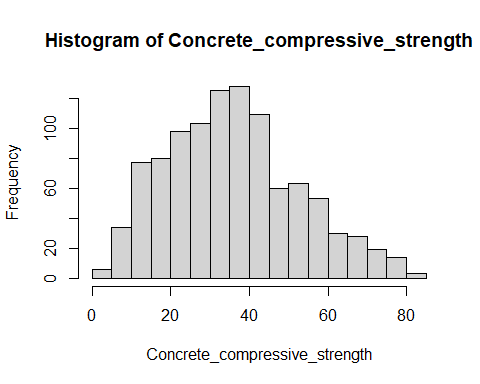
##Histogram  
hist(Data$Age\_day,xlab = "Age\_(day)",main="Histogram of Age\_(day)")



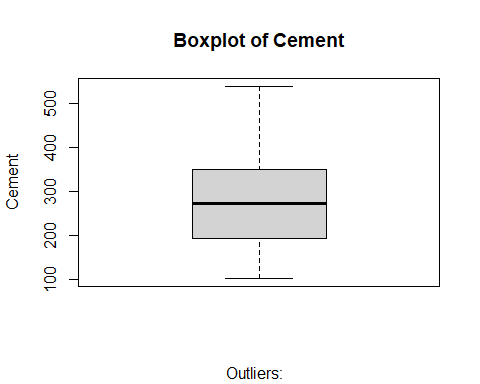
###############Concrete\_compressive\_strength############  
  
##Summary  
  
summary(Data$Concrete\_compressive\_strength)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.33 23.71 34.45 35.82 46.13 82.60

##Histogram  
hist(Data$Concrete\_compressive\_strength,xlab = "Concrete\_compressive\_strength",main="Histogram of Concrete\_compressive\_strength")

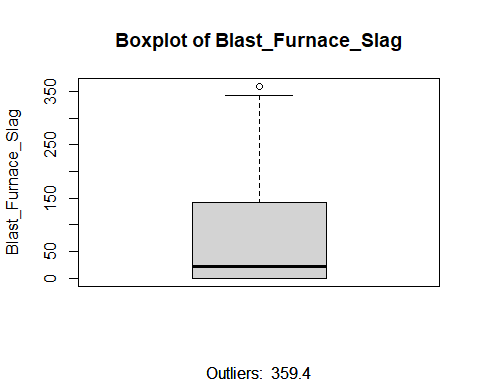


####Cement  
##Boxplot  
boxplot(Data$Cement,ylab="Cement",main="Boxplot of Cement",sub=paste("Outliers: ",boxplot.stats(Data$Cement)$out))



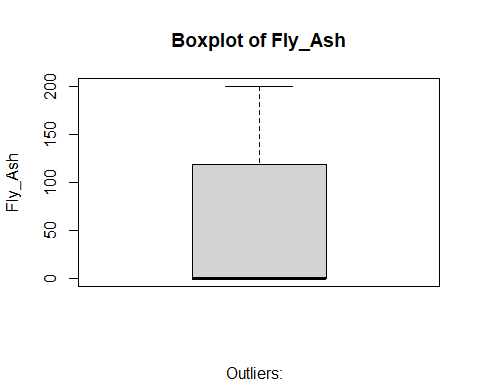
#This boxplot seems that no outliers

#Blast\_Furnace\_Slag  
##Boxplot  
boxplot(Data$Blast\_Furnace\_Slag,ylab="Blast\_Furnace\_Slag",main="Boxplot of Blast\_Furnace\_Slag",sub=paste("Outliers: ",boxplot.stats(Data$Blast\_Furnace\_Slag)$out))



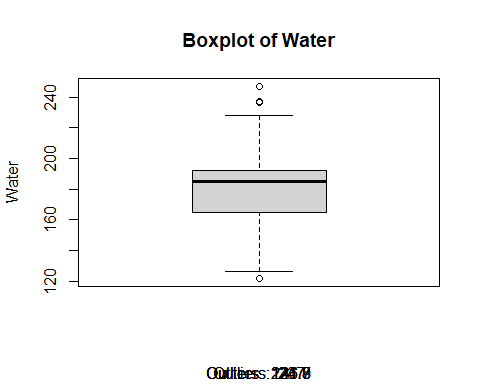
#This box plot seems that one outlier point.  
#Remove the outlier row.  
outlier\_rows\_Blast\_Furnace\_Slag<-boxplot.stats(Data$Blast\_Furnace\_Slag)$out  
  
#Clean Outliers  
Data\_cleaned\_1<-subset(Data,!(Blast\_Furnace\_Slag%in%outlier\_rows\_Blast\_Furnace\_Slag))

##Fly\_Ash  
##Boxplot  
boxplot(Data\_cleaned\_1$Fly\_Ash,ylab="Fly\_Ash",main="Boxplot of Fly\_Ash",sub=paste("Outliers: ",boxplot.stats(Data\_cleaned\_1$Fly\_Ash)$out))



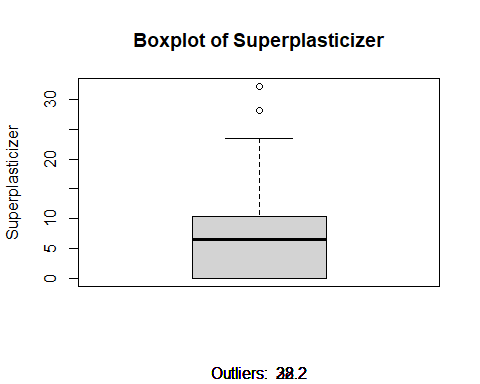
#This box plot seems that no outlier.

#Water  
##Boxplot  
boxplot(Data\_cleaned\_1$Water,ylab="Water",main="Boxplot of Water",sub=paste("Outliers: ",boxplot.stats(Data\_cleaned\_1$Water)$out))



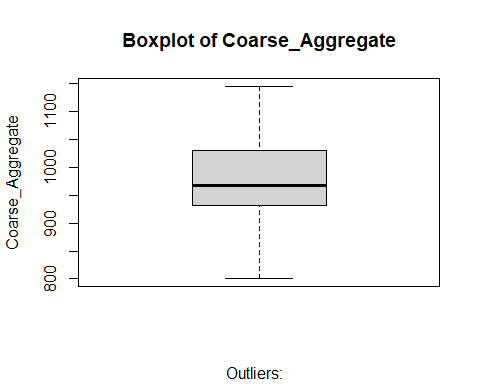
#This box plot seems that three outlier points.  
#Remove the outlier row.  
outlier\_rows\_Water<-boxplot.stats(Data\_cleaned\_1$Water)$out  
  
Data\_cleaned\_2<-subset(Data\_cleaned\_1,!(Water%in%outlier\_rows\_Water))

#Superplasticizer  
##Boxplot  
boxplot(Data\_cleaned\_2$Superplasticizer,ylab="Superplasticizer",main="Boxplot of Superplasticizer",sub=paste("Outliers: ",boxplot.stats(Data\_cleaned\_2$Superplasticizer)$out))



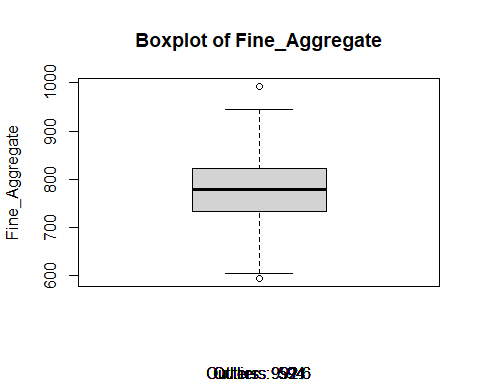
#This box plot seems that two outlier points.  
#Remove the outlier row.  
outlier\_rows\_Superplasticizer<-boxplot.stats(Data\_cleaned\_2$Superplasticizer)$out  
  
  
Data\_cleaned\_3<-subset(Data\_cleaned\_2,!(Superplasticizer%in%outlier\_rows\_Superplasticizer))

##Coarse\_Aggregate  
##Boxplot  
boxplot(Data\_cleaned\_3$Coarse\_Aggregate,ylab="Coarse\_Aggregate",main="Boxplot of Coarse\_Aggregate",sub=paste("Outliers: ",boxplot.stats(Data\_cleaned\_3$Coarse\_Aggregate)$out))



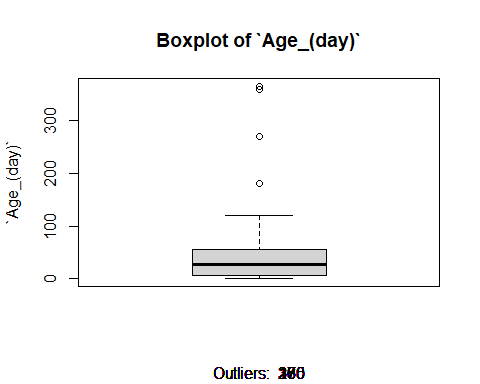
#This box plot seems that no outlier points.

##Fine\_Aggregate  
##Boxplot  
boxplot(Data\_cleaned\_3$Fine\_Aggregate,ylab="Fine\_Aggregate",main="Boxplot of Fine\_Aggregate",sub=paste("Outliers: ",boxplot.stats(Data\_cleaned\_3$Fine\_Aggregate)$out))



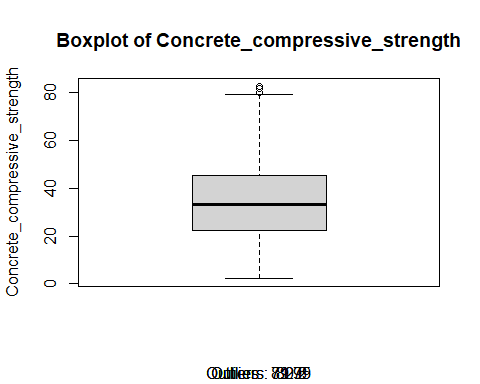
#This box plot seems that one outlier point.  
#Remove the outlier row.  
outlier\_rows\_Fine\_Aggregate<-boxplot.stats(Data\_cleaned\_3$Fine\_Aggregate)$out  
  
  
Data\_cleaned\_4<-subset(Data\_cleaned\_3,!(Fine\_Aggregate%in%outlier\_rows\_Fine\_Aggregate))

#Age\_(day)  
##Boxplot  
boxplot(Data\_cleaned\_4$Age\_day,ylab="`Age\_(day)`",main="Boxplot of `Age\_(day)`",sub=paste("Outliers: ",boxplot.stats(Data\_cleaned\_4$Age\_day)$out))



#This box plot seems that four outlier points.  
#Remove the outlier row.  
outlier\_rows\_Age<-boxplot.stats(Data\_cleaned\_4$Age\_day)$out  
  
Data\_cleaned\_5<-subset(Data\_cleaned\_4,!(Age\_day%in%outlier\_rows\_Age))

#Concrete\_compressive\_strength  
##Boxplot  
boxplot(Data\_cleaned\_5$Concrete\_compressive\_strength,ylab="Concrete\_compressive\_strength",main="Boxplot of Concrete\_compressive\_strength",sub=paste("Outliers: ",boxplot.stats(Data\_cleaned\_5$Concrete\_compressive\_strength)$out))



#This box plot seems that four outlier points.  
#Remove the outlier row.  
outlier\_rows\_Concrete\_compressive\_strength<-boxplot.stats(Data\_cleaned\_5$Concrete\_compressive\_strength)$out  
  
  
  
Data\_cleaned\_6<-subset(Data\_cleaned\_5,!(Concrete\_compressive\_strength%in%outlier\_rows\_Concrete\_compressive\_strength))

###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