NY Bus breakdown and Delays.

Roli Srivastava

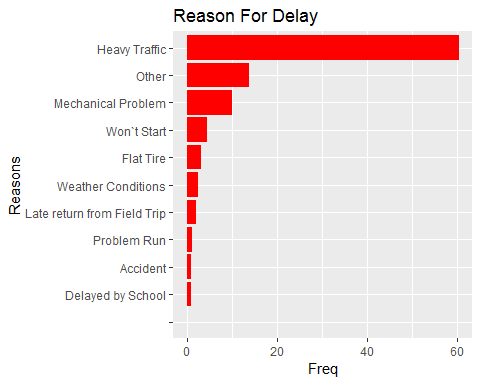
August 19, 2018

load data here

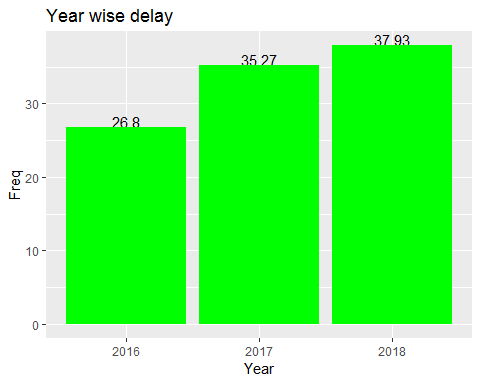
##you can use json link to download data  
##library("rjson")  
##result <- fromJSON(file = "https://data.cityofnewyork.us/resource/fbkk-fqs7.json")  
##busData <- as.data.frame(result)  
## or download from the link below and convert the file to csv  
## kaggle datasets download -d new-york-city/ny-bus-breakdown-and-delays  
  
busData <-read.csv('../Project/ny-bus-breakdown-and-delays/ny-bus-breakdown-and-delays.csv',header= T)

p\_test <- function(confidencelevel)  
{  
 c <- confidencelevel  
 alpha <- 1 -(c/100)  
 z <- 1 - (alpha/2)  
 print(z)  
 pvalue = qnorm(abs(z))  
 print(pvalue)  
}

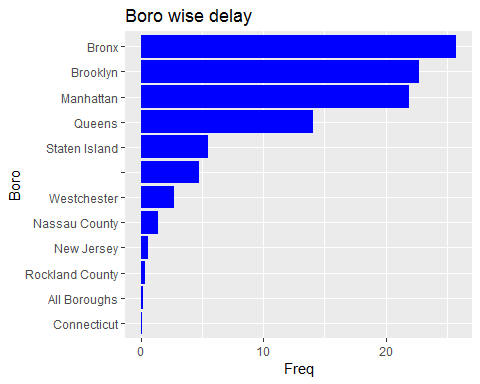
# Histogram to show the reasons for delay  
reason <- data.frame(table(busData$Reason))  
  
reason$Percentage\_Change <- round((reason$Freq/nrow(busData))\*100,digit=2)  
ggplot(reason,aes(x = reorder(reason$Var1,reason$Freq), y = reason$Percentage\_Change))+  
 labs(title = "Reason For Delay" , x= "Reasons",y= "Freq")+  
 geom\_bar(stat="identity" , fill = "red")+  
 coord\_flip()



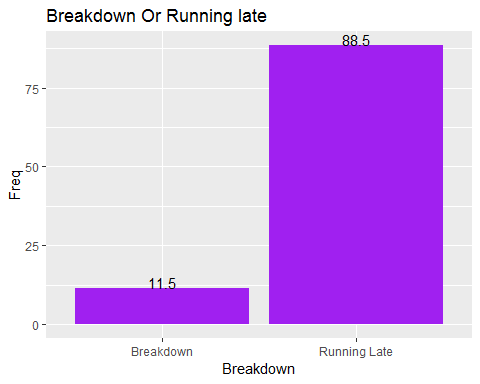
# Histogram to show the most affected year  
Year <- data.frame(table(busData$School\_Year))  
Year$Percentage\_Change <- round((Year$Freq/nrow(busData))\*100,digit=2)  
ggplot(Year,aes(x = reorder(Year$Var1,Year$Freq), y = Year$Percentage\_Change))+  
 labs(title = "Year wise delay" , x= "Year",y= "Freq")+  
 geom\_text(aes(label=Year$Percentage\_Change), vjust=0)+  
 geom\_bar(stat="identity" , fill = "green")



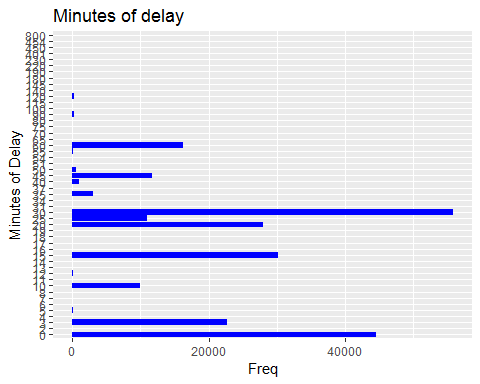
# Histogram to show the most boro  
Boro <- data.frame(table(busData$Boro))  
Boro$Percentage\_Change <- round((Boro$Freq/nrow(busData))\*100,digit=2)  
ggplot(Boro,aes(x = reorder(Boro$Var1,Boro$Freq), y = Boro$Percentage\_Change))+  
 labs(title = "Boro wise delay" , x= "Boro",y= "Freq")+  
 geom\_bar(stat="identity" , fill = "blue")+  
 coord\_flip()



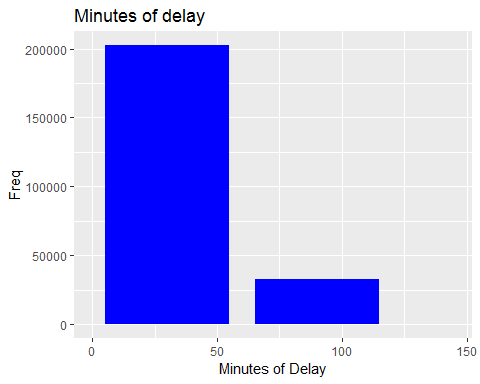
# Histogram to show the breakdown or running late  
BL <- data.frame(table(busData$Breakdown\_or\_Running\_Late))  
BL$Percentage\_Change <- round((BL$Freq/nrow(busData))\*100,digit=2)  
ggplot(BL,aes(x = reorder(BL$Var1,BL$Percentage\_Change), y = BL$Percentage\_Change))+  
 labs(title = "Breakdown Or Running late" , x= "Breakdown",y= "Freq")+  
 geom\_bar(stat="identity" , fill = "purple")+  
geom\_text(aes(label=BL$Percentage\_Change), vjust=0)



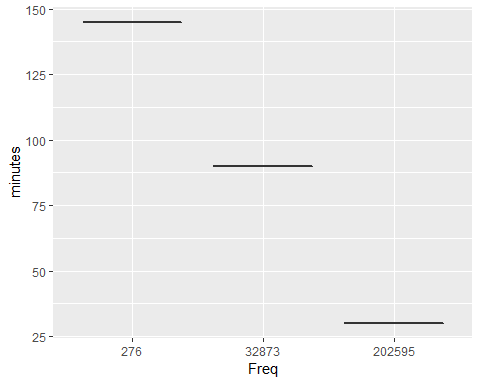
# Histogram to show minutes of delay  
minDelay <- data.frame(table(busData$How\_Long\_Delayed))  
 minDelay <- within(minDelay, {  
 delay <- as.numeric(as.vector.factor(minDelay$Var1))  
})  
ggplot(minDelay,aes(minDelay$Var1,minDelay$Freq ))+  
 labs(title = "Minutes of delay" , x= "Minutes of Delay",y= "Freq")+  
 geom\_bar(stat="identity" , fill = "blue")+  
 coord\_flip()



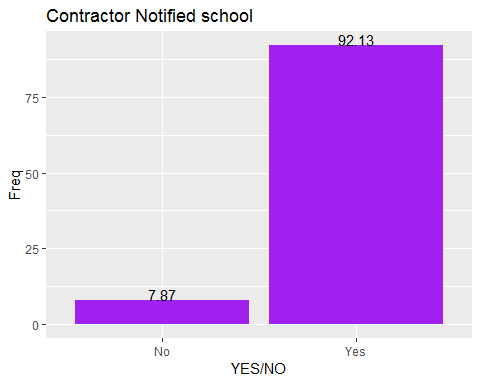
mins30 <- subset(minDelay, minDelay$delay <=30)  
mins90 <- subset(minDelay, minDelay$delay >30 & minDelay$delay <=90)  
mins145 <- subset(minDelay, minDelay$delay >90 & minDelay$delay <=145)  
mins800 <- subset(minDelay, minDelay$delay >145 & minDelay$delay <=800)  
  
minTable <- data.frame(c(sum(mins30$Freq) , sum(mins90$Freq), sum(mins145$Freq)),  
c(30,90,145))  
colnames(minTable)[1] <- "Freq"  
colnames(minTable)[2] <- "minutes"  
  
ggplot(minTable,aes(minTable$minutes,minTable$Freq ))+  
 labs(title = "Minutes of delay" , x= "Minutes of Delay",y= "Freq")+  
 geom\_bar(stat="identity" , fill = "blue")+  
 coord\_cartesian(xlim = c(0,30,90,145))



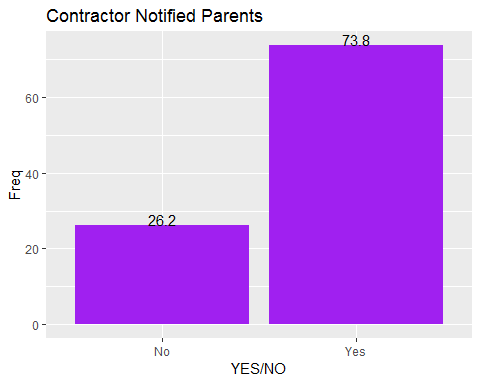
# A really basic boxplot.  
ggplot(minTable, aes(x=as.factor(Freq), y=minutes)) +   
 geom\_boxplot(fill="slateblue", alpha=0.2) +   
 xlab("Freq")



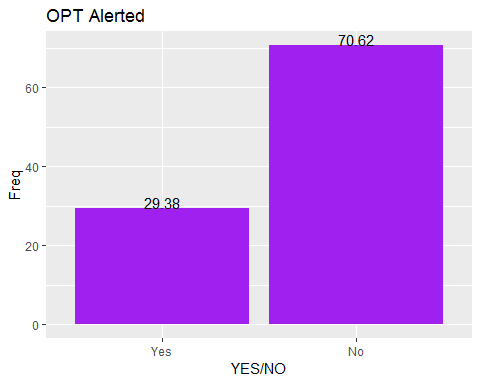
# Histogram to show the contractor notified school  
notifiedSchool <- data.frame(table(busData$Has\_Contractor\_Notified\_Schools))  
notifiedSchool$Percentage\_Change <- round((notifiedSchool$Freq/nrow(busData))\*100,digit=2)  
ggplot(notifiedSchool,aes(x = reorder(notifiedSchool$Var1,notifiedSchool$Freq), y = notifiedSchool$Percentage\_Change))+  
 labs(title = "Contractor Notified school" , x= "YES/NO",y= "Freq")+  
 geom\_bar(stat="identity" , fill = "purple")+  
geom\_text(aes(label=notifiedSchool$Percentage\_Change), vjust=0)



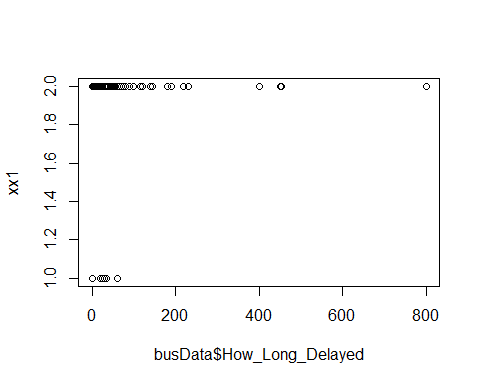
# Histogram to show the contractor notified parents  
notifiedParents <- data.frame(table(busData$Has\_Contractor\_Notified\_Parents))  
notifiedParents$Percentage\_Change <- round((notifiedParents$Freq/nrow(busData))\*100,digit=2)  
ggplot(notifiedParents,aes(x = reorder(notifiedParents$Var1,notifiedParents$Freq),y =notifiedParents$Percentage\_Change))+  
 labs(title = "Contractor Notified Parents" , x= "YES/NO",y= "Freq")+  
 geom\_bar(stat="identity" , fill = "purple")+  
geom\_text(aes(label=notifiedParents$Percentage\_Change), vjust=0)



# Histogram to show the OPT had been alerted  
optAlerted <- data.frame(table(busData$Have\_You\_Alerted\_OPT))  
optAlerted$Percentage\_Change <- round((optAlerted$Freq/nrow(busData))\*100,digit=2)  
ggplot(optAlerted,aes(x = reorder(optAlerted$Var1,optAlerted$Freq), y = optAlerted$Percentage\_Change))+  
 labs(title = "OPT Alerted" , x= "YES/NO",y= "Freq")+  
 geom\_bar(stat="identity" , fill = "purple")+  
 geom\_text(aes(label=optAlerted$Percentage\_Change), vjust=0)

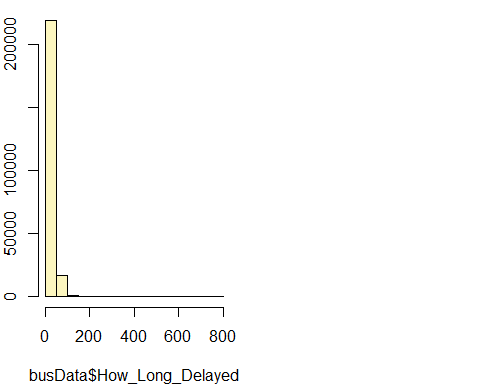


xx1 = as.integer(as.factor(busData$Breakdown\_or\_Running\_Late))  
plot(busData$How\_Long\_Delayed ,xx1)



## 1.Calculate a 95% confidence interval for the average length of delays (minutes) and interpret it in context. Note that since you’re doing inference on a single population parameter, there is no explanatory variable, so you can omit the x variable from the function.  
  
inference(y = busData$How\_Long\_Delayed, est = "mean", type = "ci", null = 0, conflevel = 0.95 ,  
 alternative = "twosided", method = "theoretical")

## Single mean   
## Summary statistics:

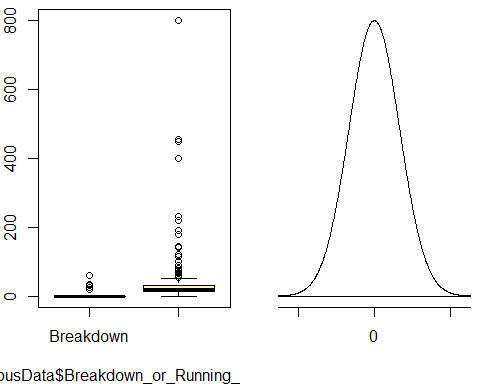


## mean = 20.6923 ; sd = 17.5817 ; n = 235758   
## Standard error = 0.0362   
## 95 % Confidence interval = ( 20.6213 , 20.7633 )

##95 % Confidence interval = ( 20.6214 , 20.7633 )  
  
## I want to know , if delay time is impacted by the run type .  
## Null: H0: mu\_Breakdown - mu\_Running Late = 0   
## ALT : HA: mu\_Breakdown - mu\_Running Late != 0   
inference(y = busData$How\_Long\_Delayed, x = busData$Breakdown\_or\_Running\_Late, est = "mean", type = "ht", null = 0, alternative = "twosided", method = "theoretical")

## Response variable: numerical, Explanatory variable: categorical  
## Difference between two means  
## Summary statistics:  
## n\_Breakdown = 27118, mean\_Breakdown = 0.0116, sd\_Breakdown = 0.6361  
## n\_Running Late = 208640, mean\_Running Late = 23.3803, sd\_Running Late = 16.9241

## Observed difference between means (Breakdown-Running Late) = -23.3687  
##   
## H0: mu\_Breakdown - mu\_Running Late = 0   
## HA: mu\_Breakdown - mu\_Running Late != 0   
## Standard error = 0.037   
## Test statistic: Z = -627.306   
## p-value = 0



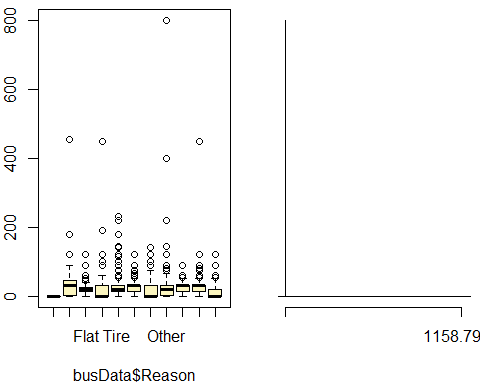
## our p-value 0 < 0.05 so we reject our null hypothesis meaning there is enough evidence in the data to say there is some difference in the average time delayed by buses whose running were considered breakdown and those running late.  
  
  
newData <- subset(busData, Reason == "Mechanical Problem" | Reason == "Heavy Traffic")  
inference(y=busData$How\_Long\_Delayed , x = busData$Reason , est="mean" ,null= 0 , alternative = "greater", type ="ht" , method="theoretical")

## Response variable: numerical, Explanatory variable: categorical

## Warning: Ignoring null value since it's undefined for ANOVA.

## ANOVA  
##   
## Summary statistics:  
## n\_ = 2, mean\_ = 0, sd\_ = 0  
## n\_Accident = 2133, mean\_Accident = 28.9133, sd\_Accident = 27.5675  
## n\_Delayed by School = 2032, mean\_Delayed by School = 19.2269, sd\_Delayed by School = 11.6607  
## n\_Flat Tire = 7296, mean\_Flat Tire = 15.8995, sd\_Flat Tire = 21.4805  
## n\_Heavy Traffic = 142494, mean\_Heavy Traffic = 21.7989, sd\_Heavy Traffic = 15.1623  
## n\_Late return from Field Trip = 5161, mean\_Late return from Field Trip = 26.0949, sd\_Late return from Field Trip = 18.2815  
## n\_Mechanical Problem = 23941, mean\_Mechanical Problem = 13.6925, sd\_Mechanical Problem = 20.8274  
## n\_Other = 32745, mean\_Other = 22.5652, sd\_Other = 19.9155  
## n\_Problem Run = 2892, mean\_Problem Run = 25.7441, sd\_Problem Run = 15.8244  
## n\_Weather Conditions = 6136, mean\_Weather Conditions = 27.0728, sd\_Weather Conditions = 19.2798  
## n\_Won`t Start = 10926, mean\_Won`t Start = 10.3851, sd\_Won`t Start = 17.4851

## H\_0: All means are equal.  
## H\_A: At least one mean is different.  
## Analysis of Variance Table  
##   
## Response: y  
## Df Sum Sq Mean Sq F value Pr(>F)  
## x 10 3414356 341436 1158.8 < 2.2e-16  
## Residuals 235747 69461880 295   
##   
## Pairwise tests: t tests with pooled SD   
## Accident Delayed by School Flat Tire  
## Accident 0.0173 NA NA NA  
## Delayed by School 0.1134 0 NA NA  
## Flat Tire 0.1903 0 0 NA  
## Heavy Traffic 0.0725 0 0 0  
## Late return from Field Trip 0.0316 0 0 0  
## Mechanical Problem 0.2593 0 0 0  
## Other 0.0630 0 0 0  
## Problem Run 0.0340 0 0 0  
## Weather Conditions 0.0257 0 0 0  
## Won`t Start 0.3923 0 0 0  
## Heavy Traffic Late return from Field Trip  
## Accident NA NA  
## Delayed by School NA NA  
## Flat Tire NA NA  
## Heavy Traffic NA NA  
## Late return from Field Trip 0 NA  
## Mechanical Problem 0 0.0000  
## Other 0 0.0000  
## Problem Run 0 0.3789  
## Weather Conditions 0 0.0026  
## Won`t Start 0 0.0000  
## Mechanical Problem Other Problem Run  
## Accident NA NA NA  
## Delayed by School NA NA NA  
## Flat Tire NA NA NA  
## Heavy Traffic NA NA NA  
## Late return from Field Trip NA NA NA  
## Mechanical Problem NA NA NA  
## Other 0 NA NA  
## Problem Run 0 0 NA  
## Weather Conditions 0 0 6e-04  
## Won`t Start 0 0 0e+00  
## Weather Conditions  
## Accident NA  
## Delayed by School NA  
## Flat Tire NA  
## Heavy Traffic NA  
## Late return from Field Trip NA  
## Mechanical Problem NA  
## Other NA  
## Problem Run NA  
## Weather Conditions NA  
## Won`t Start 0



## High F value 1158.8 , and p-value is less than 0.05 (as suggested by normal scientific standard). Hence we can conclude that for our confidence interval we reject the null hypothesis and there is convincing evidence that there is a significant relationship between delay time and reasons for delay .

.For the data in the NY Delay , how long delayed ,what is the sampling distribution of the mean,the overall mean , and the standard error of the mean?Is a normal distribution an appropriate assumption for the sampling distribution of the mean.

meanTest <- round(mean(busData$How\_Long\_Delayed),digit=2)  
sdTest <- round(sd(busData$How\_Long\_Delayed),digit=2)  
  
n= length(busData$How\_Long\_Delayed)  
x <- p\_test(95)

## [1] 0.975  
## [1] 1.959964

zsc<- round(x, digits = 2)  
cat('zscore for ',x , ' is ' ,zsc)

## zscore for 1.959964 is 1.96

error <- round(zsc \* sdTest / sqrt(n),digit =2)  
lower\_vector <- meanTest - error   
upper\_vector <- meanTest + error  
  
summary <- c("alpha","Standarad deviation","Sample size n","sample average/mean","Confidence Interval","Error","Lower","Upper")  
values <- c("0.05",sdTest,n,meanTest,"95%",error,lower\_vector,upper\_vector)  
kable(data.frame(summary,values))

|  |  |
| --- | --- |
| summary | values |
| alpha | 0.05 |
| Standarad deviation | 17.58 |
| Sample size n | 235758 |
| sample average/mean | 20.69 |
| Confidence Interval | 95% |
| Error | 0.07 |
| Lower | 20.62 |
| Upper | 20.76 |

## Hence the overall mean is 20.69 and the standard error is 0.07 . The assumption of normal distribution is appropriate when the sample is normal distribution population is sufficiently large , and thus the sampling distribution is normal .Also, the normally assumption is valid in most cases .Hence, the assumption of normal distribution is appropriate

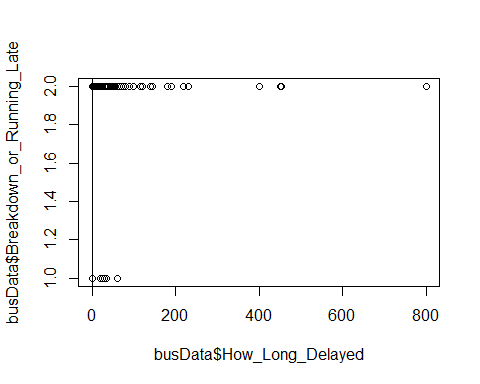
m1 <- lm(How\_Long\_Delayed ~ Breakdown\_or\_Running\_Late, data = busData)  
summary(m1)

##   
## Call:  
## lm(formula = How\_Long\_Delayed ~ Breakdown\_or\_Running\_Late, data = busData)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -23.38 -8.38 -0.01 6.62 776.62   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 0.01162 0.09669 0.12 0.904  
## Breakdown\_or\_Running\_LateRunning Late 23.36868 0.10278 227.36 <2e-16  
##   
## (Intercept)   
## Breakdown\_or\_Running\_LateRunning Late \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 15.92 on 235756 degrees of freedom  
## Multiple R-squared: 0.1798, Adjusted R-squared: 0.1798   
## F-statistic: 5.169e+04 on 1 and 235756 DF, p-value: < 2.2e-16

sum(m1$residuals^2)

## [1] 59770633

plot(busData$How\_Long\_Delayed ,busData$Breakdown\_or\_Running\_Late)  
abline(m1)



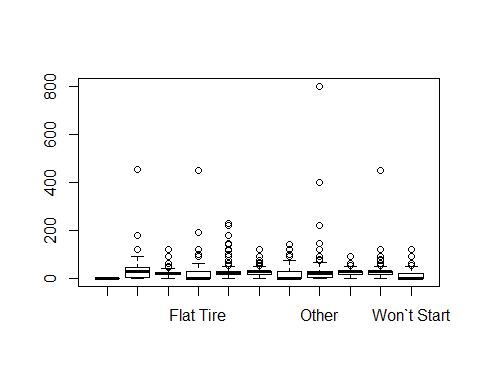
## Y = how long delayed , X = Breakdown\_or\_Running\_Late  
## Y = 0.01162 + 23.36871 \* X  
## So Breakdown is 0 and running late is 1   
## for breakdown line is Y = 0.01162  
## for running late line is Y = 0.01162 + 23.36871 \* 1 = 23.38033  
  
  
m1 <- lm(How\_Long\_Delayed ~ Reason, data = busData)  
summary(m1)

##   
## Call:  
## lm(formula = How\_Long\_Delayed ~ Reason, data = busData)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -28.91 -12.57 -1.80 8.20 777.43   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.091e-08 1.214e+01 0.000 1.0000   
## ReasonAccident 2.891e+01 1.214e+01 2.381 0.0173 \*  
## ReasonDelayed by School 1.923e+01 1.214e+01 1.583 0.1134   
## ReasonFlat Tire 1.590e+01 1.214e+01 1.310 0.1903   
## ReasonHeavy Traffic 2.180e+01 1.214e+01 1.796 0.0725 .  
## ReasonLate return from Field Trip 2.609e+01 1.214e+01 2.149 0.0316 \*  
## ReasonMechanical Problem 1.369e+01 1.214e+01 1.128 0.2593   
## ReasonOther 2.257e+01 1.214e+01 1.859 0.0630 .  
## ReasonProblem Run 2.574e+01 1.214e+01 2.120 0.0340 \*  
## ReasonWeather Conditions 2.707e+01 1.214e+01 2.230 0.0257 \*  
## ReasonWon`t Start 1.039e+01 1.214e+01 0.856 0.3923   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 17.17 on 235747 degrees of freedom  
## Multiple R-squared: 0.04685, Adjusted R-squared: 0.04681   
## F-statistic: 1159 on 10 and 235747 DF, p-value: < 2.2e-16

sum(m1$residuals^2)

## [1] 69461880

plot(busData$Reason , busData$How\_Long\_Delayed)



delayData <- busData[, c("Reason", "Boro", "How\_Long\_Delayed")]   
head(delayData, 10)

## Reason Boro How\_Long\_Delayed  
## 1 Other Nassau County 25  
## 2 Mechanical Problem Brooklyn 0  
## 3 Other Brooklyn 30  
## 4 Other Brooklyn 20  
## 5 Mechanical Problem Nassau County 30  
## 6 Mechanical Problem Westchester 0  
## 7 Heavy Traffic Manhattan 0  
## 8 Problem Run Brooklyn 90  
## 9 Other Brooklyn 20  
## 10 Heavy Traffic Manhattan 0

delayDataNoOutlier <- filter(delayData, How\_Long\_Delayed <= (quantile(delayData$How\_Long\_Delayed, 0.75) + (1.5 \* IQR(delayData$How\_Long\_Delayed))), How\_Long\_Delayed >= (quantile(delayData$How\_Long\_Delayed, 0.25) - (1.5 \* IQR(delayData$How\_Long\_Delayed))))   
modelNA <- lm(How\_Long\_Delayed ~ Reason + Boro, data = delayDataNoOutlier)   
summary(modelNA)

##   
## Call:  
## lm(formula = How\_Long\_Delayed ~ Reason + Boro, data = delayDataNoOutlier)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -35.077 -12.562 -3.176 9.425 54.425   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.9656 11.4545 0.259 0.795712   
## ReasonAccident 28.4982 11.4588 2.487 0.012883 \*   
## ReasonDelayed by School 19.3950 11.4590 1.693 0.090542 .   
## ReasonFlat Tire 16.7684 11.4549 1.464 0.143233   
## ReasonHeavy Traffic 23.2444 11.4535 2.029 0.042412 \*   
## ReasonLate return from Field Trip 25.9037 11.4556 2.261 0.023746 \*   
## ReasonMechanical Problem 13.8432 11.4538 1.209 0.226813   
## ReasonOther 22.4118 11.4537 1.957 0.050381 .   
## ReasonProblem Run 26.0560 11.4574 2.274 0.022957 \*   
## ReasonWeather Conditions 27.4498 11.4553 2.396 0.016564 \*   
## ReasonWon`t Start 10.9859 11.4543 0.959 0.337505   
## BoroAll Boroughs 1.9062 0.8985 2.122 0.033865 \*   
## BoroBronx -7.1716 0.1673 -42.867 < 2e-16 \*\*\*  
## BoroBrooklyn -1.2342 0.1708 -7.225 5.02e-13 \*\*\*  
## BoroConnecticut -3.7088 1.2554 -2.954 0.003135 \*\*   
## BoroManhattan -6.5189 0.1696 -38.443 < 2e-16 \*\*\*  
## BoroNassau County -1.5089 0.3235 -4.665 3.09e-06 \*\*\*  
## BoroNew Jersey 3.6135 0.4596 7.863 3.78e-15 \*\*\*  
## BoroQueens -2.9656 0.1790 -16.569 < 2e-16 \*\*\*  
## BoroRockland County -0.8981 0.5993 -1.498 0.134025   
## BoroStaten Island -3.0343 0.2102 -14.432 < 2e-16 \*\*\*  
## BoroWestchester 0.8977 0.2551 3.519 0.000434 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 16.2 on 235125 degrees of freedom  
## Multiple R-squared: 0.0779, Adjusted R-squared: 0.07781   
## F-statistic: 945.8 on 21 and 235125 DF, p-value: < 2.2e-16

delayData <- busData[, c("Reason","Breakdown\_or\_Running\_Late","Boro", "Bus\_Company\_Name", "How\_Long\_Delayed" ,"School\_Year")]   
head(delayData, 10)

## Reason Breakdown\_or\_Running\_Late Boro  
## 1 Other Running Late Nassau County  
## 2 Mechanical Problem Breakdown Brooklyn  
## 3 Other Running Late Brooklyn  
## 4 Other Running Late Brooklyn  
## 5 Mechanical Problem Running Late Nassau County  
## 6 Mechanical Problem Breakdown Westchester  
## 7 Heavy Traffic Running Late Manhattan  
## 8 Problem Run Running Late Brooklyn  
## 9 Other Running Late Brooklyn  
## 10 Heavy Traffic Running Late Manhattan  
## Bus\_Company\_Name How\_Long\_Delayed School\_Year  
## 1 BORO TRANSIT, INC. 25 2016  
## 2 RELIANT TRANS, INC. (B2321) 0 2016  
## 3 NEW DAWN TRANSIT, LLC (B2321) 30 2016  
## 4 EMPIRE CHARTER SERVICE IN 20 2016  
## 5 BORO TRANSIT, INC. 30 2016  
## 6 MAR-CAN TRANSPORT CO. INC 0 2016  
## 7 RELIANT TRANS, INC. (B2321) 0 2016  
## 8 CAREFUL BUS 90 2016  
## 9 BORO TRANSIT, INC. 20 2016  
## 10 LEESEL TRANSP CORP (B2192 0 2016

gameshare <- group\_by(delayData,delayData$School\_Year)  
gameshare <- summarize(gameshare,Total\_Delay=sum(How\_Long\_Delayed))  
  
cor(gameshare$`delayData$School\_Year`,gameshare$Total\_Delay)

## [1] 0.8546084

modelNA <- step(lm(How\_Long\_Delayed ~ Reason+Breakdown\_or\_Running\_Late+Boro+Bus\_Company\_Name, data = delayData),direction="both" )

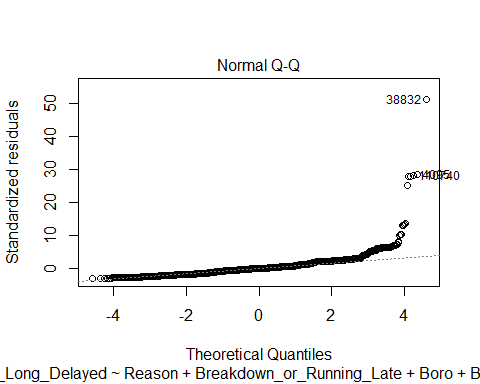
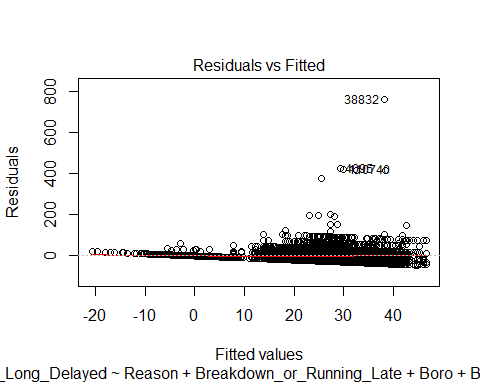
## Start: AIC=1273494  
## How\_Long\_Delayed ~ Reason + Breakdown\_or\_Running\_Late + Boro +   
## Bus\_Company\_Name  
##   
## Df Sum of Sq RSS AIC  
## <none> 52252090 1273494  
## - Boro 11 298188 52550279 1274813  
## - Reason 10 602876 52854966 1276178  
## - Bus\_Company\_Name 56 4663434 56915525 1293536  
## - Breakdown\_or\_Running\_Late 1 9596511 61848602 1313243

summary(modelNA)

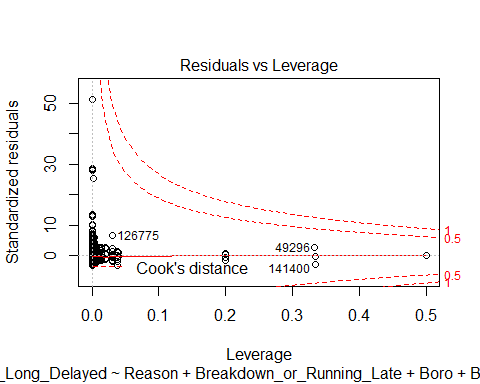
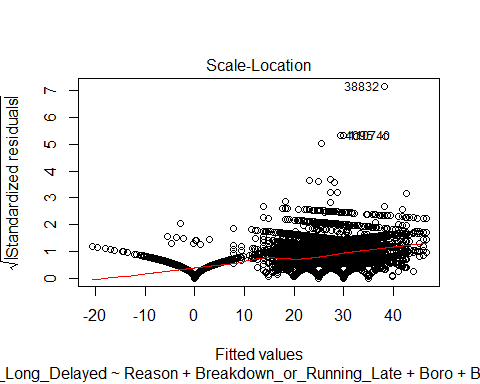
##   
## Call:  
## lm(formula = How\_Long\_Delayed ~ Reason + Breakdown\_or\_Running\_Late +   
## Boro + Bus\_Company\_Name, data = delayData)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -44.82 -8.90 -0.42 6.94 761.76   
##   
## Coefficients:  
## Estimate Std. Error  
## (Intercept) -47.5411 18.2398  
## ReasonAccident 27.8889 10.5350  
## ReasonDelayed by School 17.9681 10.5352  
## ReasonFlat Tire 24.4952 10.5317  
## ReasonHeavy Traffic 19.0803 10.5301  
## ReasonLate return from Field Trip 22.6115 10.5320  
## ReasonMechanical Problem 23.6528 10.5308  
## ReasonOther 20.0997 10.5304  
## ReasonProblem Run 21.5700 10.5338  
## ReasonWeather Conditions 23.5737 10.5317  
## ReasonWon`t Start 22.8801 10.5313  
## Breakdown\_or\_Running\_LateRunning Late 28.4469 0.1367  
## BoroAll Boroughs 0.5165 0.8485  
## BoroBronx -3.9183 0.1897  
## BoroBrooklyn -2.9161 0.1955  
## BoroConnecticut -4.8618 1.1738  
## BoroManhattan -2.1844 0.1935  
## BoroNassau County -1.7211 0.3214  
## BoroNew Jersey -1.4186 0.4374  
## BoroQueens -1.5586 0.1918  
## BoroRockland County 1.0600 0.5776  
## BoroStaten Island 1.2579 0.2392  
## BoroWestchester -2.4025 0.3977  
## Bus\_Company\_Name1992 21.4855 15.0421  
## Bus\_Company\_NameACME BUS CORP. (B2321) 20.9990 14.8961  
## Bus\_Company\_NameADDIES 42.7573 15.1745  
## Bus\_Company\_NameALINA SERVICES CORP. 26.1023 14.9048  
## Bus\_Company\_NameALL AMERICAN SCHOOL BUS C 22.5749 14.8948  
## Bus\_Company\_NameALL COUNTY BUS LLC (B2321 21.2513 14.9191  
## Bus\_Company\_NameALLIED TRANSIT CORP. 21.2927 14.8928  
## Bus\_Company\_NameANOTHER RIDE INC. 20.0829 14.9071  
## Bus\_Company\_NameB & F SKILLED INC.(B2192) 23.7577 14.8933  
## Bus\_Company\_NameBOBBY`S BUS CO. INC. 28.1727 14.8949  
## Bus\_Company\_NameBORO TRANSIT, INC. 25.9748 14.8914  
## Bus\_Company\_NameCAREFUL BUS 30.0531 14.8942  
## Bus\_Company\_NameCHILDREN`S TRANS INC. (B2321) 32.0483 14.9186  
## Bus\_Company\_NameCONSOLIDATED BUS TRANS. I 26.5567 14.8929  
## Bus\_Company\_NameDON THOMAS BUSES, INC. (B 21.6199 14.8920  
## Bus\_Company\_NameEMPIRE CHARTER SERVICE IN 22.0696 14.8929  
## Bus\_Company\_NameFIRST STEPS 29.5551 14.9129  
## Bus\_Company\_NameFIRST STEPS TRANSP INC. (B2192) 31.9123 14.8924  
## Bus\_Company\_NameFORTUNA BUS COMPANY 27.9299 21.0584  
## Bus\_Company\_NameG.V.C. LTD. (B2192) 23.8634 14.8913  
## Bus\_Company\_NameGRANDPA`S BUS CO., INC. 27.6183 14.8935  
## Bus\_Company\_Namegvc 25.1243 16.3121  
## Bus\_Company\_NameHAPPY CHILD TRANS LLC (B2 39.4232 14.8925  
## Bus\_Company\_NameHOYT TRANSPORTATION CORP. 24.4905 14.8920  
## Bus\_Company\_NameI & Y TRANSIT CORP 25.2951 14.9079  
## Bus\_Company\_NameIC BUS INC. (PRE-K) 21.8309 15.1155  
## Bus\_Company\_NameJOFAZ TRANSPORTATION INC. 27.5329 14.8928  
## Bus\_Company\_NameL & M BUS CORP. (B2192) 26.8141 14.8918  
## Bus\_Company\_NameLEESEL TRANSP CORP (B2192 15.9357 14.8914  
## Bus\_Company\_NameLITTLE RICHIE BUS SERVICE 27.0028 14.8913  
## Bus\_Company\_NameLOGAN BUS COMPANY INC. 26.6435 14.8917  
## Bus\_Company\_NameLORINDA ENT. LTD. 28.7422 14.8923  
## Bus\_Company\_NameLORISSA BUS SERVICE INC. 28.4206 14.8981  
## Bus\_Company\_NameMAR-CAN TRANSPORT CO. INC 29.8118 14.8966  
## Bus\_Company\_NameMJT BUS 29.7495 14.9072  
## Bus\_Company\_NameMONTAUK STUDENT TRANS, IN 23.2274 14.9044  
## Bus\_Company\_NameMUTUAL BUS CORP. (B2321) 23.1448 14.9056  
## Bus\_Company\_NameMV TRANSPORTATION, INC. 33.5759 14.9028  
## Bus\_Company\_NameNEW DAWN TRANSIT, LLC (B2321) 30.2854 14.8915  
## Bus\_Company\_NamePENNY TRANSPORTATION 33.3598 14.9218  
## Bus\_Company\_NamePHILLIP BUS CORP (B2192) 24.8953 14.8951  
## Bus\_Company\_NamePIONEER TRANSPORTATION CORP 20.6528 14.8913  
## Bus\_Company\_NamePRIDE TRANSPORTATION (SCH 28.0350 14.8925  
## Bus\_Company\_NameQUALITY TRANSPORTATION CO 24.0185 14.8929  
## Bus\_Company\_NameR & C TRANSIT, INC. (B2321) 26.2907 17.1956  
## Bus\_Company\_NameRELIANT TRANS, INC. (B2321) 31.0950 14.8910  
## Bus\_Company\_NameSAFE COACH INC. (B2321) 17.8128 14.9096  
## Bus\_Company\_NameSELBY TRANS CORP. (B2192) 11.6299 14.8965  
## Bus\_Company\_NameSMART PICK 33.5991 14.9134  
## Bus\_Company\_NameSNT BUS INC 21.2424 14.8920  
## Bus\_Company\_NameTHIRD AVENUE TRANSIT 21.0162 14.9912  
## Bus\_Company\_NameTHOMAS BUSES INC (B2192) 32.6741 14.8954  
## Bus\_Company\_NameTWENTY FIRST AV TRANSP (B 27.1175 14.8943  
## Bus\_Company\_NameVAN TRANS LLC (B2192) 27.6117 14.8920  
## Bus\_Company\_NameVINNY`S BUS SERVICES (B23 17.1645 14.9759  
## Bus\_Company\_NameY & M TRANSIT CORP (B2192 26.2097 14.8936  
## t value Pr(>|t|)   
## (Intercept) -2.606 0.00915 \*\*   
## ReasonAccident 2.647 0.00812 \*\*   
## ReasonDelayed by School 1.706 0.08810 .   
## ReasonFlat Tire 2.326 0.02003 \*   
## ReasonHeavy Traffic 1.812 0.06999 .   
## ReasonLate return from Field Trip 2.147 0.03180 \*   
## ReasonMechanical Problem 2.246 0.02470 \*   
## ReasonOther 1.909 0.05630 .   
## ReasonProblem Run 2.048 0.04059 \*   
## ReasonWeather Conditions 2.238 0.02520 \*   
## ReasonWon`t Start 2.173 0.02981 \*   
## Breakdown\_or\_Running\_LateRunning Late 208.049 < 2e-16 \*\*\*  
## BoroAll Boroughs 0.609 0.54268   
## BoroBronx -20.658 < 2e-16 \*\*\*  
## BoroBrooklyn -14.914 < 2e-16 \*\*\*  
## BoroConnecticut -4.142 3.45e-05 \*\*\*  
## BoroManhattan -11.287 < 2e-16 \*\*\*  
## BoroNassau County -5.355 8.57e-08 \*\*\*  
## BoroNew Jersey -3.244 0.00118 \*\*   
## BoroQueens -8.128 4.38e-16 \*\*\*  
## BoroRockland County 1.835 0.06648 .   
## BoroStaten Island 5.258 1.46e-07 \*\*\*  
## BoroWestchester -6.041 1.53e-09 \*\*\*  
## Bus\_Company\_Name1992 1.428 0.15319   
## Bus\_Company\_NameACME BUS CORP. (B2321) 1.410 0.15863   
## Bus\_Company\_NameADDIES 2.818 0.00484 \*\*   
## Bus\_Company\_NameALINA SERVICES CORP. 1.751 0.07990 .   
## Bus\_Company\_NameALL AMERICAN SCHOOL BUS C 1.516 0.12962   
## Bus\_Company\_NameALL COUNTY BUS LLC (B2321 1.424 0.15432   
## Bus\_Company\_NameALLIED TRANSIT CORP. 1.430 0.15279   
## Bus\_Company\_NameANOTHER RIDE INC. 1.347 0.17792   
## Bus\_Company\_NameB & F SKILLED INC.(B2192) 1.595 0.11067   
## Bus\_Company\_NameBOBBY`S BUS CO. INC. 1.891 0.05857 .   
## Bus\_Company\_NameBORO TRANSIT, INC. 1.744 0.08111 .   
## Bus\_Company\_NameCAREFUL BUS 2.018 0.04362 \*   
## Bus\_Company\_NameCHILDREN`S TRANS INC. (B2321) 2.148 0.03170 \*   
## Bus\_Company\_NameCONSOLIDATED BUS TRANS. I 1.783 0.07456 .   
## Bus\_Company\_NameDON THOMAS BUSES, INC. (B 1.452 0.14656   
## Bus\_Company\_NameEMPIRE CHARTER SERVICE IN 1.482 0.13837   
## Bus\_Company\_NameFIRST STEPS 1.982 0.04750 \*   
## Bus\_Company\_NameFIRST STEPS TRANSP INC. (B2192) 2.143 0.03213 \*   
## Bus\_Company\_NameFORTUNA BUS COMPANY 1.326 0.18474   
## Bus\_Company\_NameG.V.C. LTD. (B2192) 1.603 0.10905   
## Bus\_Company\_NameGRANDPA`S BUS CO., INC. 1.854 0.06368 .   
## Bus\_Company\_Namegvc 1.540 0.12351   
## Bus\_Company\_NameHAPPY CHILD TRANS LLC (B2 2.647 0.00812 \*\*   
## Bus\_Company\_NameHOYT TRANSPORTATION CORP. 1.645 0.10007   
## Bus\_Company\_NameI & Y TRANSIT CORP 1.697 0.08974 .   
## Bus\_Company\_NameIC BUS INC. (PRE-K) 1.444 0.14866   
## Bus\_Company\_NameJOFAZ TRANSPORTATION INC. 1.849 0.06450 .   
## Bus\_Company\_NameL & M BUS CORP. (B2192) 1.801 0.07177 .   
## Bus\_Company\_NameLEESEL TRANSP CORP (B2192 1.070 0.28456   
## Bus\_Company\_NameLITTLE RICHIE BUS SERVICE 1.813 0.06978 .   
## Bus\_Company\_NameLOGAN BUS COMPANY INC. 1.789 0.07359 .   
## Bus\_Company\_NameLORINDA ENT. LTD. 1.930 0.05361 .   
## Bus\_Company\_NameLORISSA BUS SERVICE INC. 1.908 0.05644 .   
## Bus\_Company\_NameMAR-CAN TRANSPORT CO. INC 2.001 0.04537 \*   
## Bus\_Company\_NameMJT BUS 1.996 0.04597 \*   
## Bus\_Company\_NameMONTAUK STUDENT TRANS, IN 1.558 0.11913   
## Bus\_Company\_NameMUTUAL BUS CORP. (B2321) 1.553 0.12048   
## Bus\_Company\_NameMV TRANSPORTATION, INC. 2.253 0.02426 \*   
## Bus\_Company\_NameNEW DAWN TRANSIT, LLC (B2321) 2.034 0.04198 \*   
## Bus\_Company\_NamePENNY TRANSPORTATION 2.236 0.02538 \*   
## Bus\_Company\_NamePHILLIP BUS CORP (B2192) 1.671 0.09465 .   
## Bus\_Company\_NamePIONEER TRANSPORTATION CORP 1.387 0.16547   
## Bus\_Company\_NamePRIDE TRANSPORTATION (SCH 1.882 0.05977 .   
## Bus\_Company\_NameQUALITY TRANSPORTATION CO 1.613 0.10680   
## Bus\_Company\_NameR & C TRANSIT, INC. (B2321) 1.529 0.12628   
## Bus\_Company\_NameRELIANT TRANS, INC. (B2321) 2.088 0.03678 \*   
## Bus\_Company\_NameSAFE COACH INC. (B2321) 1.195 0.23220   
## Bus\_Company\_NameSELBY TRANS CORP. (B2192) 0.781 0.43497   
## Bus\_Company\_NameSMART PICK 2.253 0.02426 \*   
## Bus\_Company\_NameSNT BUS INC 1.426 0.15375   
## Bus\_Company\_NameTHIRD AVENUE TRANSIT 1.402 0.16095   
## Bus\_Company\_NameTHOMAS BUSES INC (B2192) 2.194 0.02827 \*   
## Bus\_Company\_NameTWENTY FIRST AV TRANSP (B 1.821 0.06866 .   
## Bus\_Company\_NameVAN TRANS LLC (B2192) 1.854 0.06372 .   
## Bus\_Company\_NameVINNY`S BUS SERVICES (B23 1.146 0.25174   
## Bus\_Company\_NameY & M TRANSIT CORP (B2192 1.760 0.07844 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 14.89 on 235679 degrees of freedom  
## Multiple R-squared: 0.283, Adjusted R-squared: 0.2828   
## F-statistic: 1193 on 78 and 235679 DF, p-value: < 2.2e-16

modelNA.aov <- aov(How\_Long\_Delayed ~ Reason+Breakdown\_or\_Running\_Late+Boro+Bus\_Company\_Name, data = delayData)  
#Displaying the reidual plots  
plot(modelNA)

## Warning: not plotting observations with leverage one:  
## 47570, 235508



## Warning: not plotting observations with leverage one:  
## 47570, 235508



#Displaying results of ANOVA analysis  
summary(modelNA.aov)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Reason 10 3414356 341436 1540.0 <2e-16 \*\*\*  
## Breakdown\_or\_Running\_Late 1 11067595 11067595 49919.5 <2e-16 \*\*\*  
## Boro 11 1478761 134433 606.3 <2e-16 \*\*\*  
## Bus\_Company\_Name 56 4663434 83276 375.6 <2e-16 \*\*\*  
## Residuals 235679 52252090 222   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1