TTC Bus Delay Analysis using Linear Predictive Model of Delay-Causing Incidents

Load readxl & stringr libraries

library(readxl)  
library(stringr)  
library(ggplot2)  
require(coefplot)

## Loading required package: coefplot

# Create one dataframe for the entire worksheet

df <- data.frame()  
  
for (b in c(14,15,16,17,18)){  
   
  
 x <- sprintf("D:/DataAnalytics/CKME\_136/TTC\_Bus\_Delay/Capstone/Bus\_20%s.xlsx",b)  
  
   
 y <- excel\_sheets(x)  
   
   
 # Create one dataframe for the entire worksheet  
   
 for (a in y){  
   
 df1 <- read\_excel(x,sheet = a)  
   
 df <- rbind(df,df1)  
   
 }  
   
}

# Store Combined file in one CSV file

write.csv(df,"D:/DataAnalytics/CKME\_136/TTC\_Bus\_Delay/Capstone/combined.csv")

# Load the combined file

delay <- read.csv("D:/DataAnalytics/CKME\_136/TTC\_Bus\_Delay/Capstone/combined.csv", header = TRUE, stringsAsFactors = FALSE, sep = ",")  
  
head(delay)

## X Report.Date Route Time Day Location  
## 1 1 2014-01-01 95 1899-12-31 00:23:00 Wednesday York Mills station  
## 2 2 2014-01-01 102 1899-12-31 00:55:00 Wednesday Entire run for route  
## 3 3 2014-01-01 54 1899-12-31 01:28:00 Wednesday lawrence and Warden  
## 4 4 2014-01-01 112 1899-12-31 01:30:00 Wednesday Kipling Station  
## 5 5 2014-01-01 24 1899-12-31 01:37:00 Wednesday VP and Ellesmere  
## 6 6 2014-01-01 129 1899-12-31 01:50:00 Wednesday Scarborough Town  
## Incident Min.Delay Min.Gap Direction Vehicle  
## 1 Mechanical 10 20 E 1734  
## 2 General Delay 33 66 b/w 8110  
## 3 Mechanical 10 20 WB 7478  
## 4 Emergency Services 18 36 N 8084  
## 5 Investigation 10 20 n 7843  
## 6 Mechanical 10 20 N 1755

delay$Incident <- as.factor(delay$Incident)  
  
delay$Day <- as.factor(delay$Day)  
  
class(delay$Incident)

## [1] "factor"

class(delay$Day)

## [1] "factor"

summary(delay$Incident)

## Diversion Emergency Services General Delay   
## 20648 6934 49782   
## Investigation Late Leaving Garage Mechanical   
## 32132 57234 141163   
## Utilized Off Route   
## 60867

# Data Cleaning

# Renaming the columns

names(delay) <- c("SrNo","date","route","time","day","location","incident","delay","gap","direction","vehicle")

# removing “NAs”

delay2 <- delay[which(!is.na(delay$delay)),]  
  
  
ggplot(delay2,aes(y=delay,x=incident))+geom\_boxplot()



# removing outliers

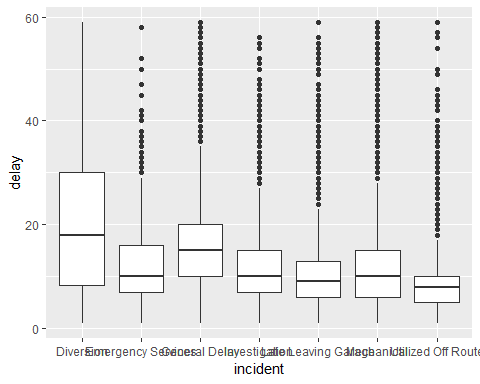
delay3 <- delay2[which(delay2$delay < 60 & delay2$delay > 0 ),]  
  
head(delay3)

## SrNo date route time day location  
## 1 1 2014-01-01 95 1899-12-31 00:23:00 Wednesday York Mills station  
## 2 2 2014-01-01 102 1899-12-31 00:55:00 Wednesday Entire run for route  
## 3 3 2014-01-01 54 1899-12-31 01:28:00 Wednesday lawrence and Warden  
## 4 4 2014-01-01 112 1899-12-31 01:30:00 Wednesday Kipling Station  
## 5 5 2014-01-01 24 1899-12-31 01:37:00 Wednesday VP and Ellesmere  
## 6 6 2014-01-01 129 1899-12-31 01:50:00 Wednesday Scarborough Town  
## incident delay gap direction vehicle  
## 1 Mechanical 10 20 E 1734  
## 2 General Delay 33 66 b/w 8110  
## 3 Mechanical 10 20 WB 7478  
## 4 Emergency Services 18 36 N 8084  
## 5 Investigation 10 20 n 7843  
## 6 Mechanical 10 20 N 1755

class(delay3$date)

## [1] "character"

#delay3$date <- as.Date(delay3$date, format = "%m/%d/%Y")  
  
ggplot(delay3,aes(y=delay,x=incident))+geom\_boxplot()



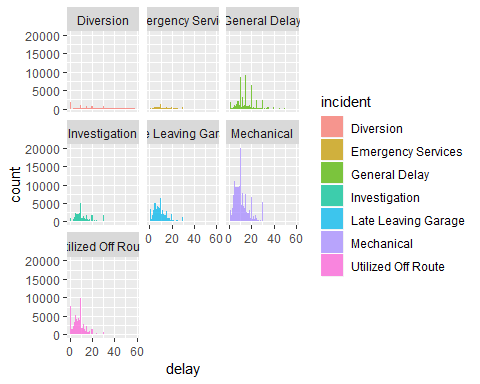
aggregate(delay ~ incident, data = delay3, mean)

## incident delay  
## 1 Diversion 20.563493  
## 2 Emergency Services 12.555166  
## 3 General Delay 15.291940  
## 4 Investigation 11.896197  
## 5 Late Leaving Garage 10.257495  
## 6 Mechanical 11.284636  
## 7 Utilized Off Route 8.399348

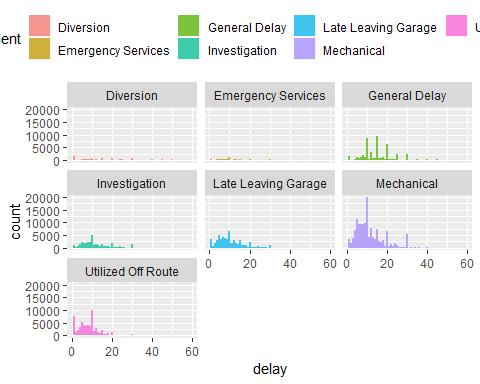
aggregate(delay ~ incident, data = delay3, var)

## incident delay  
## 1 Diversion 244.95540  
## 2 Emergency Services 59.26365  
## 3 General Delay 65.25699  
## 4 Investigation 52.67031  
## 5 Late Leaving Garage 43.21411  
## 6 Mechanical 48.24153  
## 7 Utilized Off Route 31.37129

ggplot(delay3, aes(x= delay,fill=incident))+geom\_histogram(binwidth = 1,alpha=3/4)+facet\_wrap(~incident)



ggplot(delay3, aes(x= delay,fill=incident))+geom\_histogram(binwidth = 1,alpha=3/4)+facet\_wrap(~incident)+theme(legend.position = "top")



# Analysis of Variance “ANOVA” using linear regression coefficient

delay3.lm <- lm(delay ~ incident, data = delay3)  
summary(delay3.lm)

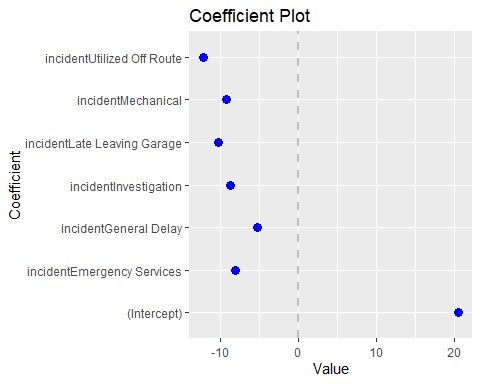
##   
## Call:  
## lm(formula = delay ~ incident, data = delay3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.563 -5.257 -1.285 3.601 50.601   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 20.56349 0.06859 299.80 <2e-16 \*\*\*  
## incidentEmergency Services -8.00833 0.11215 -71.41 <2e-16 \*\*\*  
## incidentGeneral Delay -5.27155 0.07620 -69.18 <2e-16 \*\*\*  
## incidentInvestigation -8.66730 0.07997 -108.38 <2e-16 \*\*\*  
## incidentLate Leaving Garage -10.30600 0.07523 -137.00 <2e-16 \*\*\*  
## incidentMechanical -9.27886 0.07133 -130.07 <2e-16 \*\*\*  
## incidentUtilized Off Route -12.16415 0.07484 -162.54 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.339 on 355906 degrees of freedom  
## Multiple R-squared: 0.1082, Adjusted R-squared: 0.1082   
## F-statistic: 7199 on 6 and 355906 DF, p-value: < 2.2e-16

delay3.lm2 <- lm(delay ~ incident-1, data = delay3)  
summary(delay3.lm2)

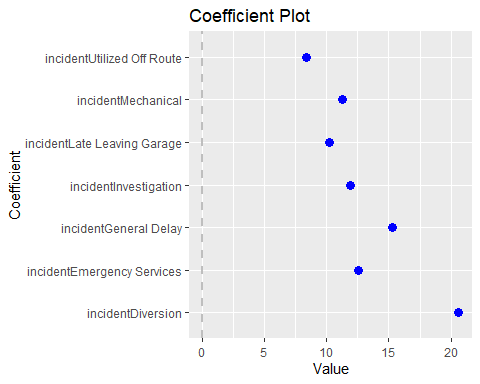
##   
## Call:  
## lm(formula = delay ~ incident - 1, data = delay3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.563 -5.257 -1.285 3.601 50.601   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## incidentDiversion 20.56349 0.06859 299.8 <2e-16 \*\*\*  
## incidentEmergency Services 12.55517 0.08872 141.5 <2e-16 \*\*\*  
## incidentGeneral Delay 15.29194 0.03319 460.7 <2e-16 \*\*\*  
## incidentInvestigation 11.89620 0.04111 289.3 <2e-16 \*\*\*  
## incidentLate Leaving Garage 10.25750 0.03089 332.0 <2e-16 \*\*\*  
## incidentMechanical 11.28464 0.01960 575.9 <2e-16 \*\*\*  
## incidentUtilized Off Route 8.39935 0.02993 280.6 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.339 on 355906 degrees of freedom  
## Multiple R-squared: 0.7225, Adjusted R-squared: 0.7225   
## F-statistic: 1.324e+05 on 7 and 355906 DF, p-value: < 2.2e-16

# Plot linear regression coefficient

coefplot(delay3.lm)



coefplot(delay3.lm2)

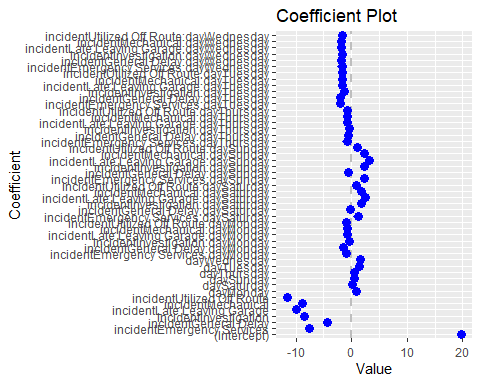


# Consider the interaction with the day of incident

delay3.lm3 <- lm(delay ~ incident \* day, data = delay3)  
summary(delay3.lm3)

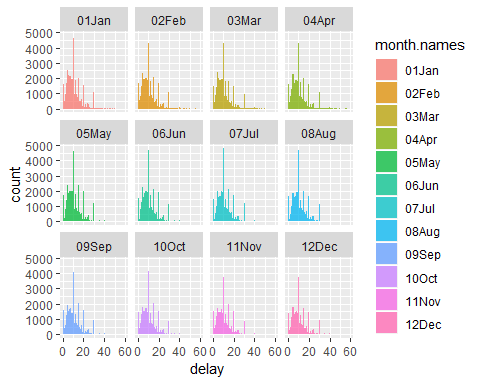
##   
## Call:  
## lm(formula = delay ~ incident \* day, data = delay3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -20.467 -4.865 -0.996 3.305 50.906   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) 19.8425 0.1695 117.040  
## incidentEmergency Services -7.6040 0.2751 -27.644  
## incidentGeneral Delay -4.2905 0.1872 -22.914  
## incidentInvestigation -8.5357 0.1981 -43.094  
## incidentLate Leaving Garage -9.9292 0.1848 -53.735  
## incidentMechanical -8.8461 0.1764 -50.153  
## incidentUtilized Off Route -11.5646 0.1846 -62.659  
## dayMonday 0.7819 0.2456 3.184  
## daySaturday 0.1010 0.2564 0.394  
## daySunday 0.5249 0.2622 2.002  
## dayThursday 0.5455 0.2442 2.234  
## dayTuesday 1.3691 0.2463 5.558  
## dayWednesday 1.6248 0.2422 6.707  
## incidentEmergency Services:dayMonday -0.8692 0.4014 -2.165  
## incidentGeneral Delay:dayMonday -1.4766 0.2746 -5.377  
## incidentInvestigation:dayMonday -0.4122 0.2872 -1.435  
## incidentLate Leaving Garage:dayMonday -0.7286 0.2664 -2.735  
## incidentMechanical:dayMonday -0.8294 0.2550 -3.253  
## incidentUtilized Off Route:dayMonday -0.8762 0.2683 -3.266  
## incidentEmergency Services:daySaturday 1.2149 0.4144 2.932  
## incidentGeneral Delay:daySaturday -0.1929 0.2912 -0.662  
## incidentInvestigation:daySaturday 1.7527 0.3001 5.840  
## incidentLate Leaving Garage:daySaturday 2.4918 0.2896 8.604  
## incidentMechanical:daySaturday 1.7678 0.2681 6.593  
## incidentUtilized Off Route:daySaturday 0.9040 0.2829 3.195  
## incidentEmergency Services:daySunday 2.3765 0.4342 5.473  
## incidentGeneral Delay:daySunday -0.5187 0.3077 -1.686  
## incidentInvestigation:daySunday 2.3114 0.3119 7.412  
## incidentLate Leaving Garage:daySunday 3.1969 0.2977 10.740  
## incidentMechanical:daySunday 2.2346 0.2747 8.134  
## incidentUtilized Off Route:daySunday 1.0599 0.2905 3.649  
## incidentEmergency Services:dayThursday -0.7988 0.3947 -2.024  
## incidentGeneral Delay:dayThursday -0.6247 0.2672 -2.338  
## incidentInvestigation:dayThursday -0.3433 0.2833 -1.212  
## incidentLate Leaving Garage:dayThursday -0.6697 0.2658 -2.519  
## incidentMechanical:dayThursday -0.7591 0.2537 -2.993  
## incidentUtilized Off Route:dayThursday -0.7296 0.2654 -2.749  
## incidentEmergency Services:dayTuesday -2.0094 0.3988 -5.039  
## incidentGeneral Delay:dayTuesday -2.0136 0.2702 -7.452  
## incidentInvestigation:dayTuesday -1.2942 0.2857 -4.529  
## incidentLate Leaving Garage:dayTuesday -1.5877 0.2673 -5.940  
## incidentMechanical:dayTuesday -1.6868 0.2556 -6.599  
## incidentUtilized Off Route:dayTuesday -1.6507 0.2664 -6.195  
## incidentEmergency Services:dayWednesday -1.7384 0.3994 -4.352  
## incidentGeneral Delay:dayWednesday -1.8137 0.2663 -6.810  
## incidentInvestigation:dayWednesday -1.5897 0.2819 -5.639  
## incidentLate Leaving Garage:dayWednesday -1.8690 0.2642 -7.075  
## incidentMechanical:dayWednesday -1.8913 0.2517 -7.513  
## incidentUtilized Off Route:dayWednesday -1.7444 0.2629 -6.634  
## Pr(>|t|)   
## (Intercept) < 2e-16 \*\*\*  
## incidentEmergency Services < 2e-16 \*\*\*  
## incidentGeneral Delay < 2e-16 \*\*\*  
## incidentInvestigation < 2e-16 \*\*\*  
## incidentLate Leaving Garage < 2e-16 \*\*\*  
## incidentMechanical < 2e-16 \*\*\*  
## incidentUtilized Off Route < 2e-16 \*\*\*  
## dayMonday 0.001451 \*\*   
## daySaturday 0.693618   
## daySunday 0.045266 \*   
## dayThursday 0.025510 \*   
## dayTuesday 2.73e-08 \*\*\*  
## dayWednesday 1.99e-11 \*\*\*  
## incidentEmergency Services:dayMonday 0.030372 \*   
## incidentGeneral Delay:dayMonday 7.57e-08 \*\*\*  
## incidentInvestigation:dayMonday 0.151297   
## incidentLate Leaving Garage:dayMonday 0.006243 \*\*   
## incidentMechanical:dayMonday 0.001142 \*\*   
## incidentUtilized Off Route:dayMonday 0.001093 \*\*   
## incidentEmergency Services:daySaturday 0.003368 \*\*   
## incidentGeneral Delay:daySaturday 0.507771   
## incidentInvestigation:daySaturday 5.22e-09 \*\*\*  
## incidentLate Leaving Garage:daySaturday < 2e-16 \*\*\*  
## incidentMechanical:daySaturday 4.32e-11 \*\*\*  
## incidentUtilized Off Route:daySaturday 0.001397 \*\*   
## incidentEmergency Services:daySunday 4.42e-08 \*\*\*  
## incidentGeneral Delay:daySunday 0.091846 .   
## incidentInvestigation:daySunday 1.25e-13 \*\*\*  
## incidentLate Leaving Garage:daySunday < 2e-16 \*\*\*  
## incidentMechanical:daySunday 4.18e-16 \*\*\*  
## incidentUtilized Off Route:daySunday 0.000264 \*\*\*  
## incidentEmergency Services:dayThursday 0.042985 \*   
## incidentGeneral Delay:dayThursday 0.019375 \*   
## incidentInvestigation:dayThursday 0.225544   
## incidentLate Leaving Garage:dayThursday 0.011755 \*   
## incidentMechanical:dayThursday 0.002766 \*\*   
## incidentUtilized Off Route:dayThursday 0.005971 \*\*   
## incidentEmergency Services:dayTuesday 4.69e-07 \*\*\*  
## incidentGeneral Delay:dayTuesday 9.22e-14 \*\*\*  
## incidentInvestigation:dayTuesday 5.92e-06 \*\*\*  
## incidentLate Leaving Garage:dayTuesday 2.85e-09 \*\*\*  
## incidentMechanical:dayTuesday 4.13e-11 \*\*\*  
## incidentUtilized Off Route:dayTuesday 5.82e-10 \*\*\*  
## incidentEmergency Services:dayWednesday 1.35e-05 \*\*\*  
## incidentGeneral Delay:dayWednesday 9.80e-12 \*\*\*  
## incidentInvestigation:dayWednesday 1.71e-08 \*\*\*  
## incidentLate Leaving Garage:dayWednesday 1.50e-12 \*\*\*  
## incidentMechanical:dayWednesday 5.79e-14 \*\*\*  
## incidentUtilized Off Route:dayWednesday 3.27e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.288 on 355864 degrees of freedom  
## Multiple R-squared: 0.1208, Adjusted R-squared: 0.1207   
## F-statistic: 1018 on 48 and 355864 DF, p-value: < 2.2e-16

coefplot(delay3.lm3)

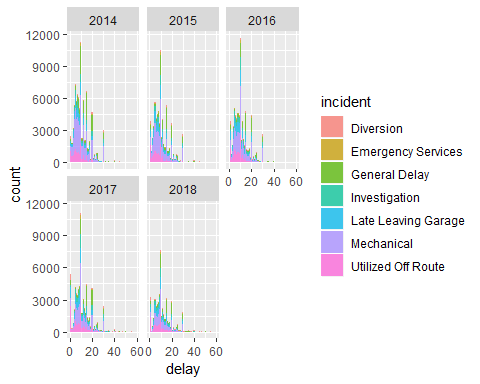


# Studying monthly distribution of the delay

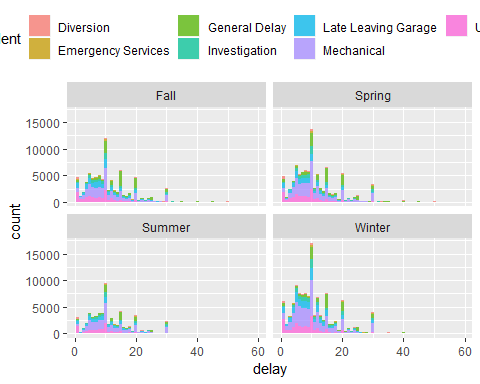
month.col <- str\_sub(delay3$date,start = 6, end = 7)  
  
month.name <- function(x){  
   
 switch(x,  
 "01"="01Jan",  
 "02"="02Feb",  
 "03"="03Mar",  
 "04"="04Apr",  
 "05"="05May",  
 "06"="06Jun",  
 "07"="07Jul",  
 "08"="08Aug",  
 "09"="09Sep",  
 "10"="10Oct",  
 "11"="11Nov",  
 "12"="12Dec"  
 )  
   
}  
  
month.names <- sapply(month.col, month.name)  
  
delay4 <- data.frame(delay3,month.names)  
  
ggplot(delay4, aes(x= delay,fill=month.names))+geom\_histogram(binwidth =1,alpha=3/4)+facet\_wrap(~month.names)



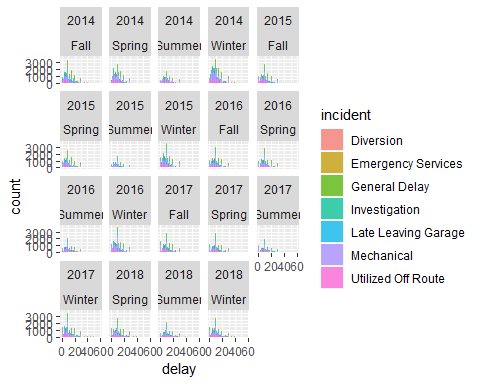
yearly <- str\_sub(delay4$date,start = 1, end = 4)  
  
delay5 <- data.frame(delay4,yearly)  
  
ggplot(delay5, aes(x= delay,fill=incident))+geom\_histogram(binwidth = 1,alpha=3/4)+facet\_wrap(~yearly)



season <- function(x){  
   
 switch(x,  
 "01"="Winter",  
 "02"="Winter",  
 "03"="Winter",  
 "04"="Spring",  
 "05"="Spring",  
 "06"="Spring",  
 "07"="Summer",  
 "08"="Summer",  
 "09"="Fall",  
 "10"="Fall",  
 "11"="Fall",  
 "12"="Winter"  
 )  
   
}  
season.names <- sapply(month.col, season)  
  
delay6 <- data.frame(delay5,season.names)  
  
ggplot(delay6, aes(x= delay,fill=incident))+geom\_histogram(binwidth = 1,alpha=3/4)+facet\_wrap(~season.names)+theme(legend.position = "top")



ggplot(delay6, aes(x= delay,fill=incident))+geom\_histogram(binwidth = 1,alpha=3/4)+facet\_wrap(~yearly+season.names)



# Analyze Incidents per location

delay6$location <- as.factor(delay6$location)  
  
#levels(delay6$location)  
  
location2 <- aggregate(incident ~ location, delay6 ,length)  
location3 <- location2[order(-location2$incident),]  
summary(location3)

## location incident   
## !00 Flemington Route : 1 Min. : 1.000   
## !00 Flemmington Route : 1 1st Qu.: 1.000   
## !16 Morningside Route : 1 Median : 1.000   
## !20 McLevin Avenue : 1 Mean : 4.549   
## !44 DownTown . Don Valley Express: 1 3rd Qu.: 2.000   
## !44 Downtown Route : 1 Max. :4985.000   
## (Other) :78116

quantile(location3$incident, probs = .995)

## 99.5%   
## 103

length(location3$incident[location3$incident > 1000])

## [1] 30

# Create a prediction model for the delay through linear regression

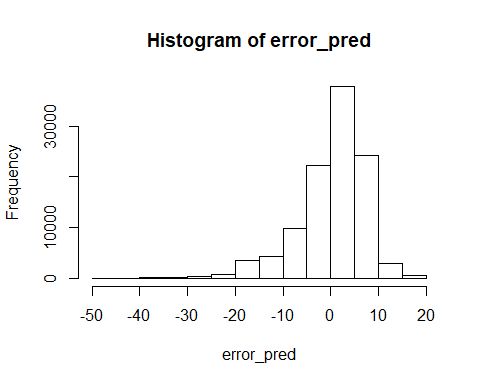
index <- sample(length(delay6$route), .7 \*length(delay6$route))  
  
train <- delay6[index,]  
test <- delay6[-index,]  
  
  
delay\_single <- lm(delay ~ incident-1, data = train)  
  
summary(delay\_single)

##   
## Call:  
## lm(formula = delay ~ incident - 1, data = train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.582 -5.254 -1.284 3.604 50.604   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## incidentDiversion 20.58177 0.08165 252.1 <2e-16 \*\*\*  
## incidentEmergency Services 12.56581 0.10679 117.7 <2e-16 \*\*\*  
## incidentGeneral Delay 15.26502 0.03963 385.2 <2e-16 \*\*\*  
## incidentInvestigation 11.94176 0.04920 242.7 <2e-16 \*\*\*  
## incidentLate Leaving Garage 10.25395 0.03693 277.7 <2e-16 \*\*\*  
## incidentMechanical 11.28447 0.02343 481.7 <2e-16 \*\*\*  
## incidentUtilized Off Route 8.39620 0.03581 234.5 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.342 on 249132 degrees of freedom  
## Multiple R-squared: 0.7225, Adjusted R-squared: 0.7225   
## F-statistic: 9.267e+04 on 7 and 249132 DF, p-value: < 2.2e-16

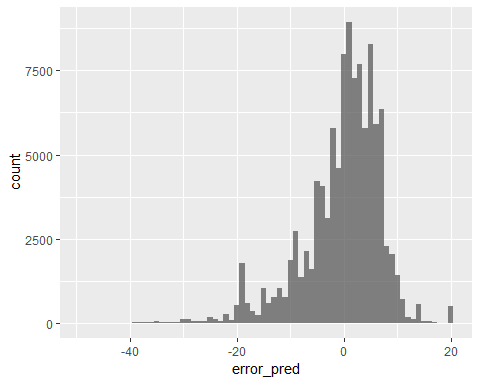
delay\_pred <- predict(delay\_single, newdata = test, interval = "prediction")  
  
pred <- data.frame(delay\_pred)  
  
  
error\_pred <- pred$fit - test$delay  
  
error <- data.frame(error\_pred)  
head(error)

## error\_pred  
## 1 -17.734977  
## 2 -4.715525  
## 3 -4.715525  
## 4 -13.058242  
## 5 -13.058242  
## 6 -17.434194

hist(error\_pred)



ggplot(error,aes(x=error\_pred))+geom\_histogram(binwidth = 1,alpha=3/4)



summary(error\_pred)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -48.74605 -3.60380 1.28447 0.00005 5.25395 19.58177

mse\_single <- sqrt(sum((error\_pred)^2)/nrow(test))  
  
mse\_single

## [1] 7.334493

relative\_error <- 1- ((test$delay - abs(error\_pred))/test$delay)  
#relative\_error  
relative\_error\_less25 <- table(relative\_error < .25)["TRUE"]/nrow(test)  
relative\_error\_less25\_pct <- paste("relative\_error\_less 25% is ",100\*relative\_error\_less25)  
relative\_error\_less25\_pct

## [1] "relative\_error\_less 25% is 34.4700020604267"

# Hypothesis: delays due to “Mechanical” incidents has no significant difference from delays caused by “All incidents”

Mechanical\_mean <- mean(delay6$delay[delay6$incident == "Mechanical"])  
  
Mechanical\_mean

## [1] 11.28464

Mechanical\_sd <- sd(delay6$delay[delay6$incident == "Mechanical"])  
  
Mechanical\_var <- var(delay6$delay[delay6$incident == "Mechanical"])  
  
  
  
Population\_mean <- mean(delay6$delay)  
  
Population\_mean

## [1] 11.56248

Population\_sd <- sd(delay6$delay)  
  
Population\_sd

## [1] 7.772019

Population\_var <- var(delay6$delay)  
  
Population\_var

## [1] 60.40428

n <- length(delay6$delay[delay6$incident == "Mechanical"])  
  
n

## [1] 140281

z <- (Mechanical\_mean - Population\_mean)/(Population\_sd/sqrt(n))  
  
z

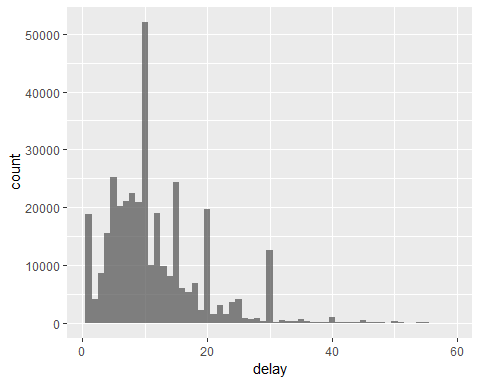
## [1] -13.38958

z1 <- (Mechanical\_mean - Population\_mean)/(Mechanical\_sd/sqrt(n))  
  
p\_value <- pnorm(z)  
  
p\_value

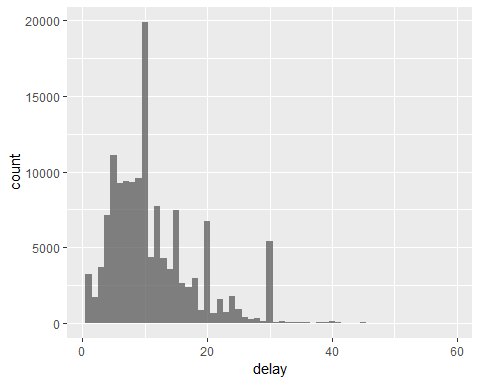
## [1] 3.478466e-41

# side by side histograms of whole delays and delays due to Mechanical reasons

ggplot(delay6, aes(x=delay))+geom\_histogram(binwidth = 1,alpha=3/4)



delay6\_mechanical <- delay6[delay6$incident == "Mechanical",]  
  
ggplot(delay6\_mechanical, aes(x=delay))+geom\_histogram(binwidth = 1,alpha=3/4)



# study Multi regression cosidering “gap” attribute

# Pearson correlation test between Delay and Gap

cor.test(delay6$delay,delay6$gap, method = "pearson")

##   
## Pearson's product-moment correlation  
##   
## data: delay6$delay and delay6$gap  
## t = 431.09, df = 355760, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.5836084 0.5879254  
## sample estimates:  
## cor   
## 0.585771

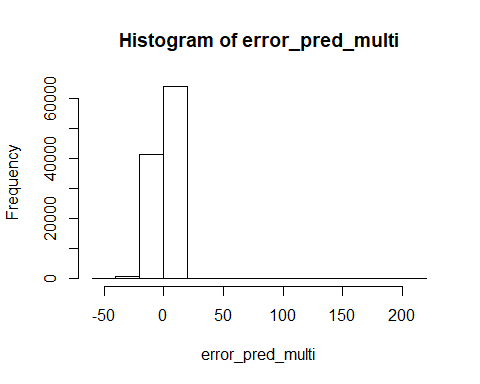
delay\_Multi <- lm(delay ~ incident-1+gap, data = train)  
  
summary(delay\_Multi)

##   
## Call:  
## lm(formula = delay ~ incident - 1 + gap, data = train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -249.892 -3.136 -0.816 2.083 50.755   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## incidentDiversion 1.108e+01 7.238e-02 153.10 <2e-16 \*\*\*  
## incidentEmergency Services 7.324e+00 8.896e-02 82.33 <2e-16 \*\*\*  
## incidentGeneral Delay 1.060e+01 3.520e-02 301.17 <2e-16 \*\*\*  
## incidentInvestigation 6.962e+00 4.284e-02 162.51 <2e-16 \*\*\*  
## incidentLate Leaving Garage 6.178e+00 3.250e-02 190.07 <2e-16 \*\*\*  
## incidentMechanical 6.614e+00 2.346e-02 281.95 <2e-16 \*\*\*  
## incidentUtilized Off Route 4.900e+00 3.106e-02 157.75 <2e-16 \*\*\*  
## gap 2.101e-01 6.048e-04 347.41 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.024 on 249015 degrees of freedom  
## (116 observations deleted due to missingness)  
## Multiple R-squared: 0.8131, Adjusted R-squared: 0.8131   
## F-statistic: 1.354e+05 on 8 and 249015 DF, p-value: < 2.2e-16

delay\_pred <- predict(delay\_Multi, newdata = test, interval = "prediction")  
  
pred\_Multi <- data.frame(delay\_pred)  
  
error\_pred\_multi <- pred\_Multi$fit - test$delay  
  
error\_multi <- data.frame(error\_pred\_multi)  
head(error\_multi)

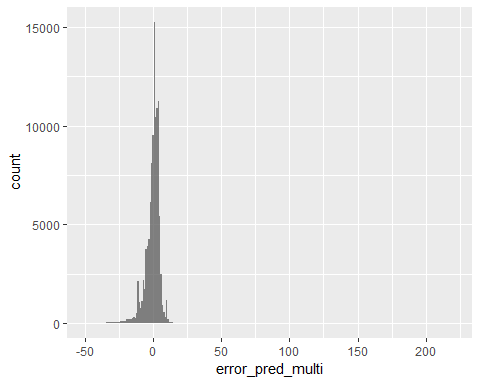
## error\_pred\_multi  
## 1 -8.531554  
## 2 -2.662325  
## 3 -3.082540  
## 4 -7.532565  
## 5 -7.532565  
## 6 -10.069447

hist(error\_pred\_multi)



ggplot(error\_multi,aes(x=error\_pred\_multi))+geom\_histogram(binwidth = 1,alpha=3/4)

## Warning: Removed 38 rows containing non-finite values (stat\_bin).



summary(error\_pred\_multi)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## -48.9987 -2.0572 0.8164 0.0011 3.1355 219.9797 38