**Processing of dataset from 311 service and socio-demographic dataset**

#open dataset with information requested from 311 service of City of Montreal

d <- read.csv(file="C:/Users/Evgeniya/Desktop/DataProcessing/requetes311Big.csv", header = TRUE)

dim(d)

names(d)

#show attributes of dataset (row, fields, class)

attributes(d)

#show first 3 row for each data field

d[1:3,]

#show first 10 row of datafield "information"

d[1:10, "Information"]

# convert single instance of date/time in format year-month-day hour:min:sec

d$Date<-as.POSIXct(d$Date,format="%Y-%m-%dT%H:%M:%S")

#format column "Date" for having only date from dateTime

d$DATE <- format(as.Date(d$Date), "%d/%m/%Y")

d$DATE

#format column "Date" for having only time from dateTime

d$TIME <- format(d$Date, "%H:%M:%S")

d$TIME

d[1:3,]

#show summary per each field

summary(d)

install.packages("Hmisc")

library(Hmisc)

install.packages('plyr')

library(plyr)

install.packages("dplyr")

library(dplyr)

#describe fields 1 and 2

describe(d[,c(1,2)])

#show range for datafield "Date"

range(d$Date)

#create a new fied "hour" in numeric format

d$hour<-as.numeric(format(d$Date, "%H"))

range(d$hour)

#create a new fied "month" in numeric format

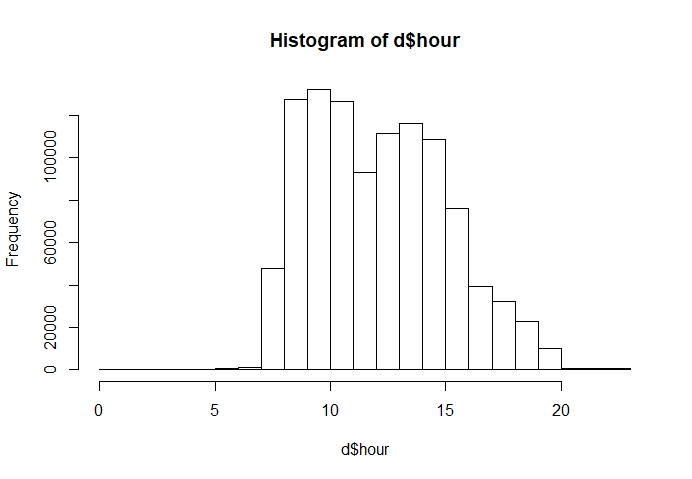
d$month<-as.numeric(format(d$Date, "%m"))

range(d$month)

#create pie chart and histogram based on hours when phone request on information was done

pie(table(d$hour))

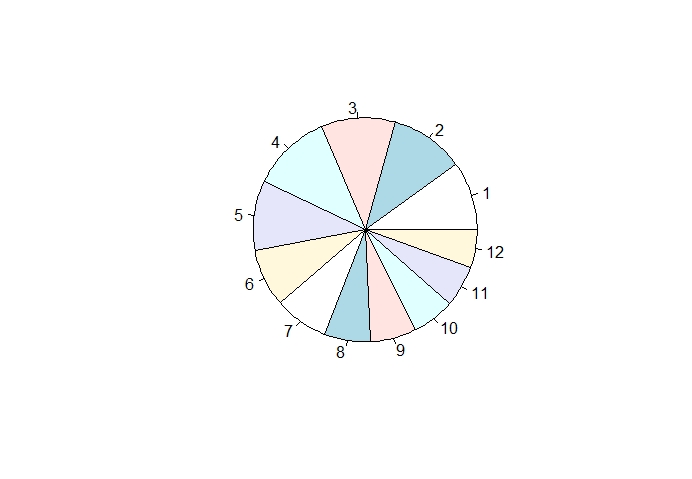
hist(d$hour)

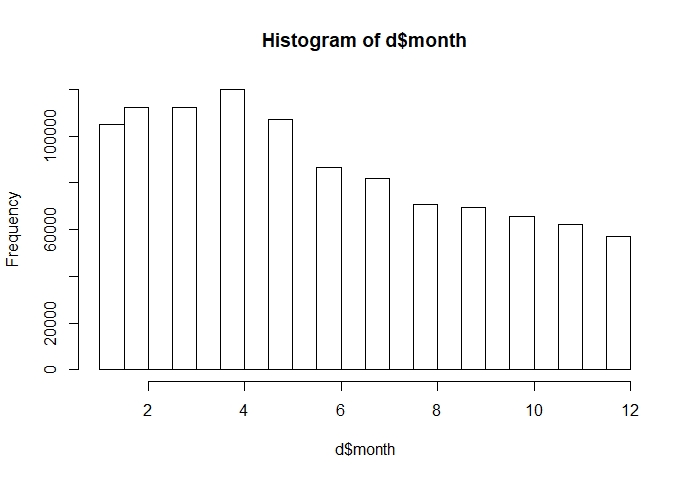


#create pie chart and histogram based on months when phone request on information was done

pie(table(d$month))

hist(d$month)





#create a dataframe with the top-15 requests

factor(d$Information)

w=table(d$Information)

w

t=as.data.frame(w)

t

t<-arrange(t,desc(Freq))

t

tt<-t[1:15,]

tt

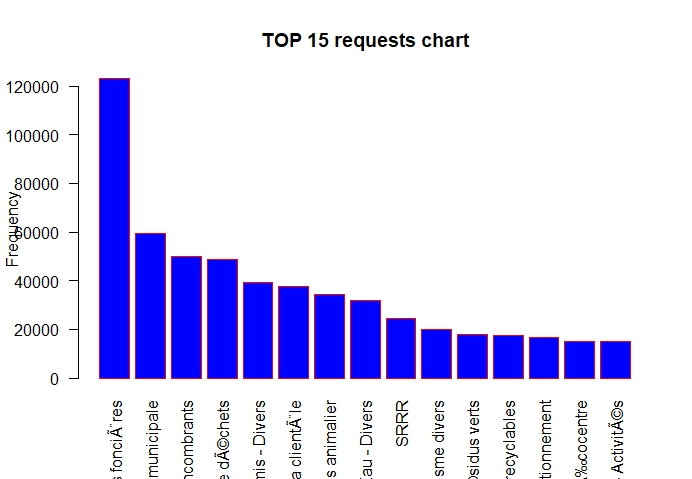
#create barchart (request vs frequency)

inf<-as.vector(tt$Var1)

freq<-as.vector(tt$Freq)

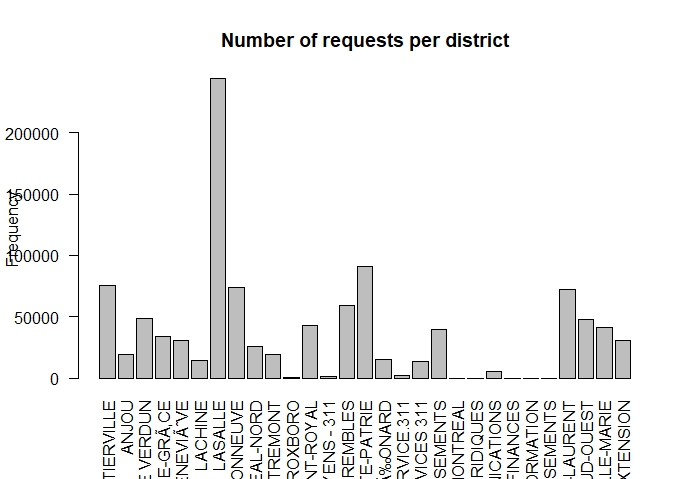
barplot(freq,names.arg=inf,ylab="Frequency", las=2, col="blue",

main="TOP 15 requests chart",border="red")



#build a barplot for datafield "Location"

barplot(table(d$Location),main="Number of requests per district",ylab="Frequency",las=2)



#aggregate columns "Location" by "information"

aggregate(Information~Location, summary, data = d)

#Study requests made in Lasale (as most active area)

dd<-subset(d, Location == "LASALLE")

dd[1:50,]

#create a dataframe with the top-15 requests for LASALLE

names(dd)

factor(dd$Information)

a=table(dd$Information)

a

b=as.data.frame(a)

b

b<-arrange(b,desc(Freq))

b

bb<-b[1:15,]

bb

names(bb)

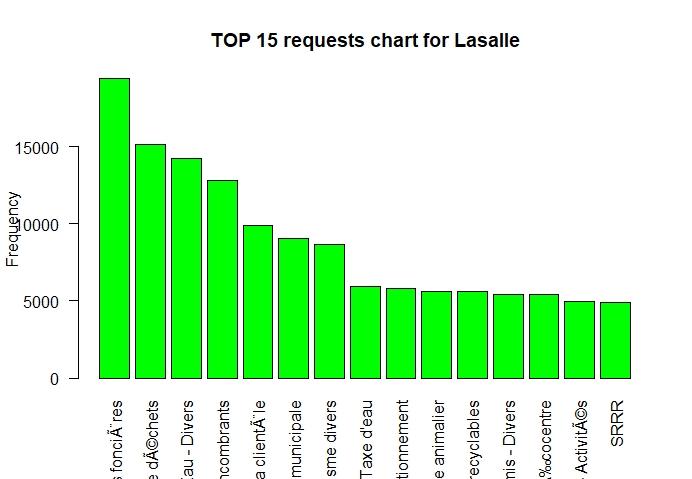
#build a barplot for datafield "Information" for LASALLE

infLas<-as.vector(bb$Var1)

freqLas<-as.vector(bb$Freq)

barplot(freqLas,names.arg=infLas,ylab="Frequency", las=2, col="green",

main="TOP 15 requests chart for Lasalle",border="black")



#create a new data frame with requests for whole montreal and Lasalle

# for visual comparison

cc<-bb

cc$InfoTotal<-tt$Var1

cc$InfoLasalle<-bb$Var1

cc<-cc[,-(1:2)]

names(cc)

cc

install.packages("geo")

library(geo)

library(ggplot2)

install.packages("eeptools")

library(eeptools)

#open dataset with information on socio-demographic statistics of City of Montreal

s <- read.csv("stat.csv", header = TRUE)

dim(s)

str(s)

names(s)

#transform all rows with numbers with comma to numbers without comma (keeping the same data field names)

s$Densité <- as.numeric(gsub(",","",s$Densité))

s$Population <- as.numeric(gsub(",","",s$Population))

s$X0.4.ans <- as.numeric(gsub(",","",s$X0.4.ans))

s$X5.9.ans <- as.numeric(gsub(",","",s$X5.9.ans))

s$Population\_m <- as.numeric(gsub(",","",s$Population\_m))

s$Population\_f <- as.numeric(gsub(",","",s$Population\_f))

s$Anglais <- as.numeric(gsub(",","",s$Anglais))

s$Français <- as.numeric(gsub(",","",s$Français))

s$Anglais.et.français <- as.numeric(gsub(",","",s$Anglais.et.français))

s$Ni.anglais.ni.français <- as.numeric(gsub(",","",s$Ni.anglais.ni.français))

s$Citoyens.canadiens <- as.numeric(gsub(",","",s$Citoyens.canadiens))

s$Pas.citoyens.canadiens <- as.numeric(gsub(",","",s$Pas.citoyens.canadiens))

s$Non.immigrants <- as.numeric(gsub(",","",s$Non.immigrants))

s$Immigrants <- as.numeric(gsub(",","",s$Immigrants))

s$Résidents\_NP <- as.numeric(gsub(",","",s$Résidents\_NP))

s$Certificat\_universite <- as.numeric(gsub(",","",s$Certificat\_universite))

s$Population.active <- as.numeric(gsub(",","",s$Population.active))

s$Population.inactive <- as.numeric(gsub(",","",s$Population.inactive))

#to see population and density distribution in Montreal

s<-arrange(s,desc(Population))

ss<-s[1:15,1:12]

ss

names(ss)

barplot(ss$Population,names.arg=ss$Location, yaxt = "n", las=2, col=rgb(0,1,0,alpha=0.3),

main="TOP 15 districs based on population and density",border="black")

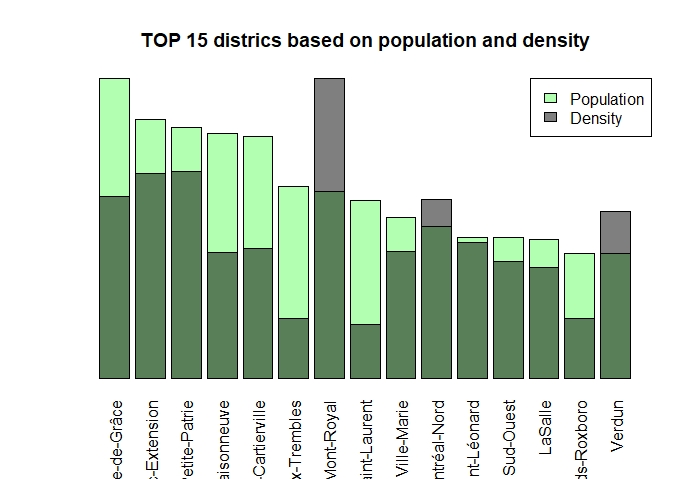
par(new=TRUE)

barplot(ss$Densité,names.arg=ss$Location, axes=F, xaxt = "n", col=rgb(0,0,0,alpha=0.5),border="black")

legend("topright",

legend = c("Population", "Density"),

fill = c(rgb(0,1,0,alpha=0.3), rgb(0,0,0,alpha=0.5)))



ss$Anglais

#Plotting two barplots (Francais and Anglais) on the same graph

barplot(ss$Français,names.arg=ss$Location, ylab="Number of people", ylim=c(0,65000), las=2, col=rgb(1,0,0,alpha=0.3),

main="French and English language in Montreal",border="black")

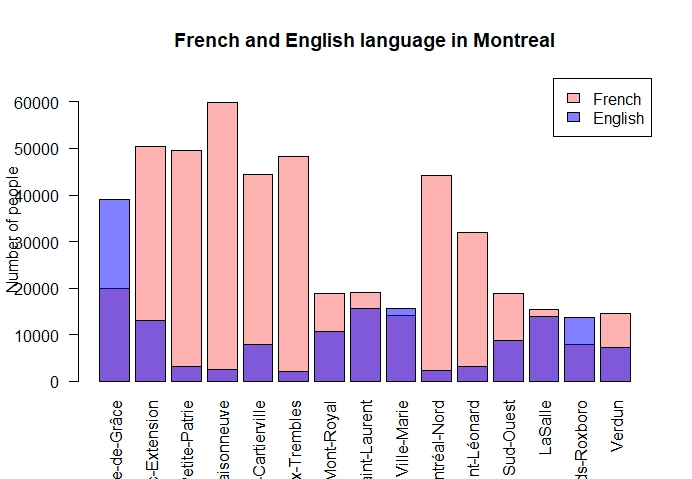
par(new=TRUE)

barplot(ss$Anglais,names.arg=ss$Location, xaxt = "n", yaxt = "n", ylim=c(0,65000), col=rgb(0,0,1,alpha=0.5),border="black")

legend("topright",

legend = c("French", "English"),

fill = c(rgb(1,0,0,alpha=0.3), rgb(0,0,1,alpha=0.5)))



**Conclusions based on data processing:**

-most of requests were done during working hours (8am-3pm) and from January to May, which may be due to problems caused by winter period

-the most frequent issue from all requests concerns property taxes

-Lasalle dominates with numbers of requests

-Among other popular requests were ones about garbage disposal and recycling, water and snow issues

-Most populated district is CDN-NDG, and most dense is Mont-Royal. Lasalle is only #13 in population rating

-Lasalle is one the districts where people speak both English and French.

*>>>Based on data processing one can conclude that Lasalle is very attractive district to live both for Francophones and Anglophones; however, there are still a lot of issues concerning properties, garbage collection, etc. City of Montreal thus way have to pay more attention to this district. One can solve issues in Lasalle by comparing the facilities of this district with more successful districts (like Roxborro, for example)*