Report

# Read data

encounter <- read.csv("/Users/liyi/Desktop/RP/day\_in\_the\_life.csv")

# How many unique subjects in this data set

index <- duplicated(encounter[,1])  
encounter\_subset <- encounter[!index,]  
(nrow(encounter\_subset))

## [1] 1307

# Change age to age range

age\_range\_rep <- encounter\_subset$Age  
i = 1  
while(i<=length(age\_range\_rep)){  
 age = age\_range\_rep[i]  
 if(age>=0&&age<10)  
 {  
 age\_range\_rep[i] = "0-10"  
 }  
 else if(age>=10&&age<20){  
 age\_range\_rep[i] = "10-20"  
 }  
 else if(age>=20&&age<30)  
 {  
 age\_range\_rep[i] = "20-30"  
 }  
 else if(age>=30&&age<40)  
 {  
 age\_range\_rep[i] = "30-40"  
 }  
 else if(age>=40&&age<50)  
 {  
 age\_range\_rep[i] = "40-50"  
 }  
 else if(age>=50&&age<60)  
 {  
 age\_range\_rep[i] = "50-60"  
 }  
 else if(age>=60&&age<70)  
 {  
 age\_range\_rep[i] = "60-70"  
 }  
 else if(age>=70&&age<80)  
 {  
 age\_range\_rep[i] = "70-80"  
 }  
 else if(age>=80&&age<90)  
 {  
 age\_range\_rep[i] = "80-90"  
 }  
 else if(age>=90&&age<100)  
 {  
 age\_range\_rep[i] = "90-100"  
 }  
 else  
 {  
 age\_range\_rep[i] = "no age"  
 }  
 i = i+1  
}

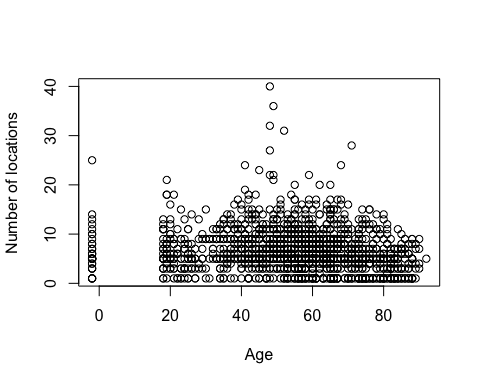
# Relationship between age and num of locations

## Make plot of age and num of locations

num\_of\_locations <- data.frame(num = numeric(0))  
for(subject in encounter\_subset$Subject){  
 #print(subject)  
 num\_of\_locations <- rbind(num\_of\_locations, nrow(encounter[which(encounter$Subject == subject), ]))  
 #nrow(encounter[which(encounter$Subject == subject), ])  
}  
num\_of\_locations$age <- encounter\_subset$Age  
nrow(num\_of\_locations)

## [1] 1307

plot(num\_of\_locations$X1L~num\_of\_locations$age,xlab = "Age",ylab = "Number of locations")

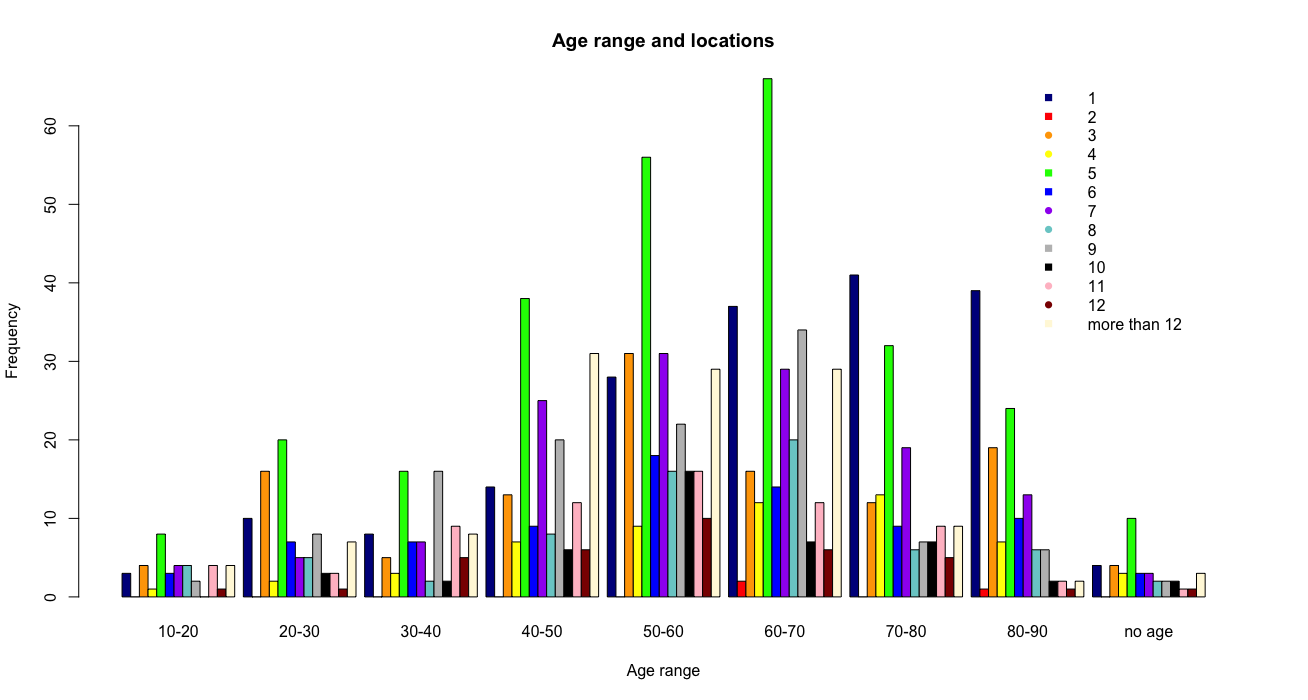
 ## There is no obvious relationship between age and the number of locations, so I rewrite age to age range. ## Rewrite number larger than 12 to “more than 12”

num\_of\_locations$location\_range <- num\_of\_locations$X1L  
num\_of\_locations$location\_range[num\_of\_locations$location\_range > 12] <- 13  
num\_of\_locations$location\_range <- factor(num\_of\_locations$location\_range,  
 levels = c(1,2,3,4,5,6,7,8,9,10,11,12,13),  
 labels = c("1","2","3","4","5","6","7","8","9","10","11","12","more than 12"))  
num\_of\_locations$age\_range <- age\_range\_rep

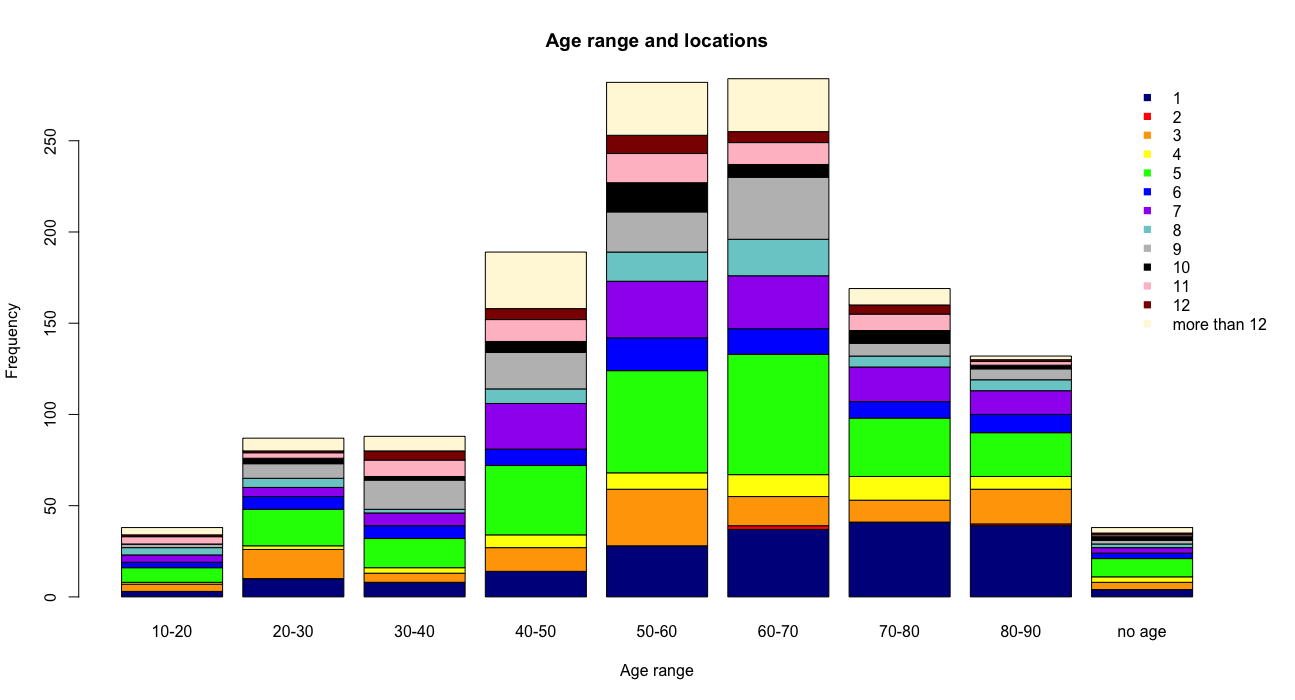
## Make barplot for age range and the number of locations. It is convenient for us to analyze the distribution of different numbers in one range, and the distribution of age range of a number.

## Make barplot

agelocation <- table(num\_of\_locations$location\_range,num\_of\_locations$age\_range)  
lacationage <- table(num\_of\_locations$age\_range,num\_of\_locations$location\_range)  
par(mar=c(5.1, 4.1, 4.1, 8.1), xpd=TRUE)  
barplot(agelocation, main="Age range and locations",  
 xlab="Age range", ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk"),  
 beside=TRUE)  
text.legend=row.names(agelocation)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
legend("topright",pch=c(15,15,16,16),legend=text.legend,col=col2,bty="n",ncol = 1,,inset=c(-0.25,0))

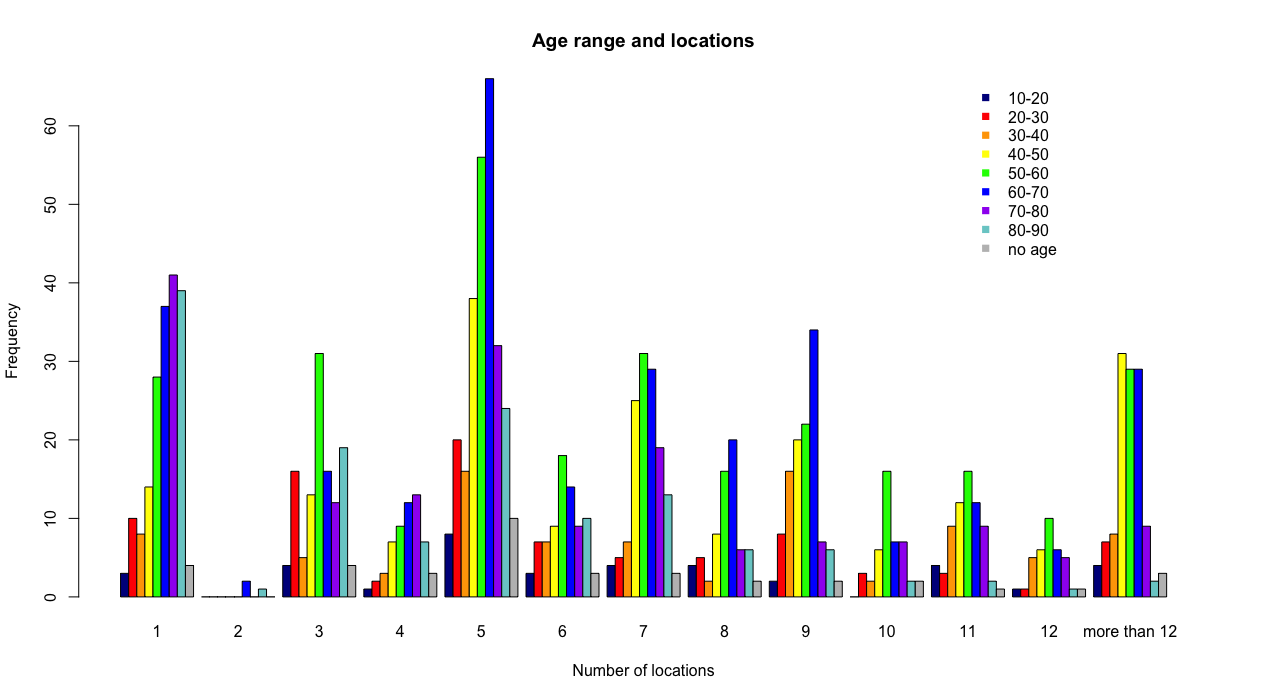


barplot(agelocation, main="Age range and locations",  
 xlab="Age range", ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
 )  
text.legend=row.names(agelocation)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))

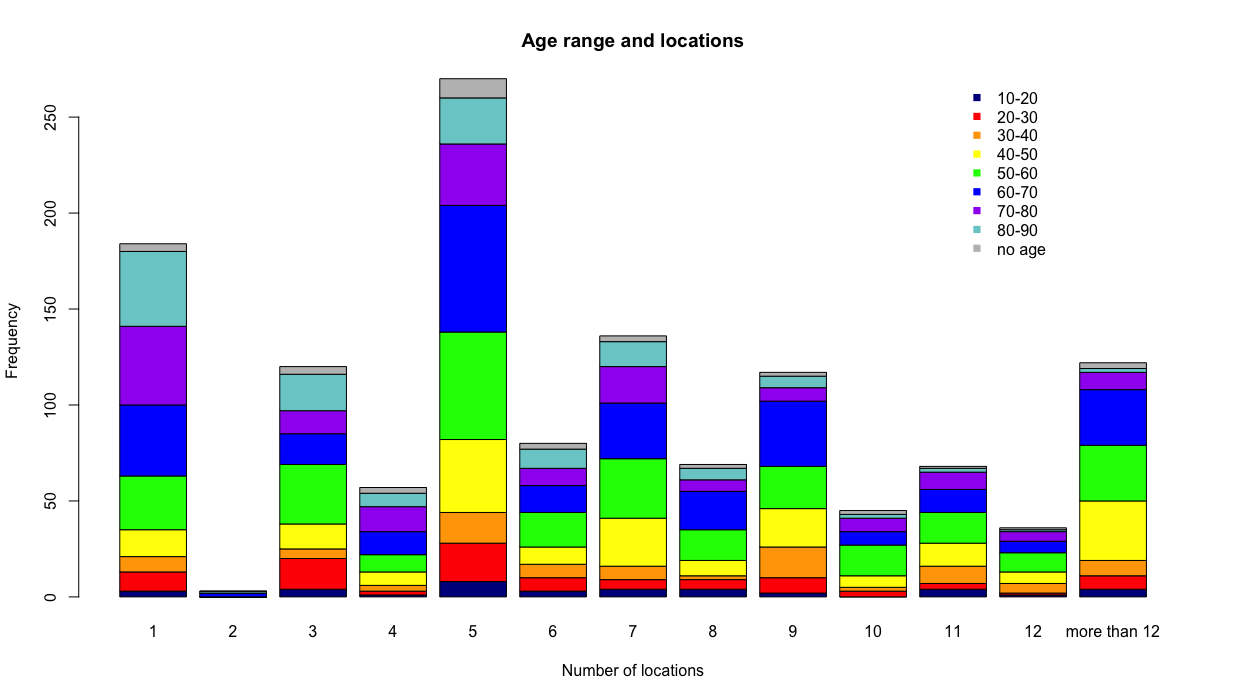


According to information in this plot, the most participants is at age 50-70. In each age range, 5 locations are all have the most frequency.

barplot(lacationage, main="Age range and locations",  
 xlab="Number of locations", ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray"),  
 beside=TRUE)  
text.legend=row.names(lacationage)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



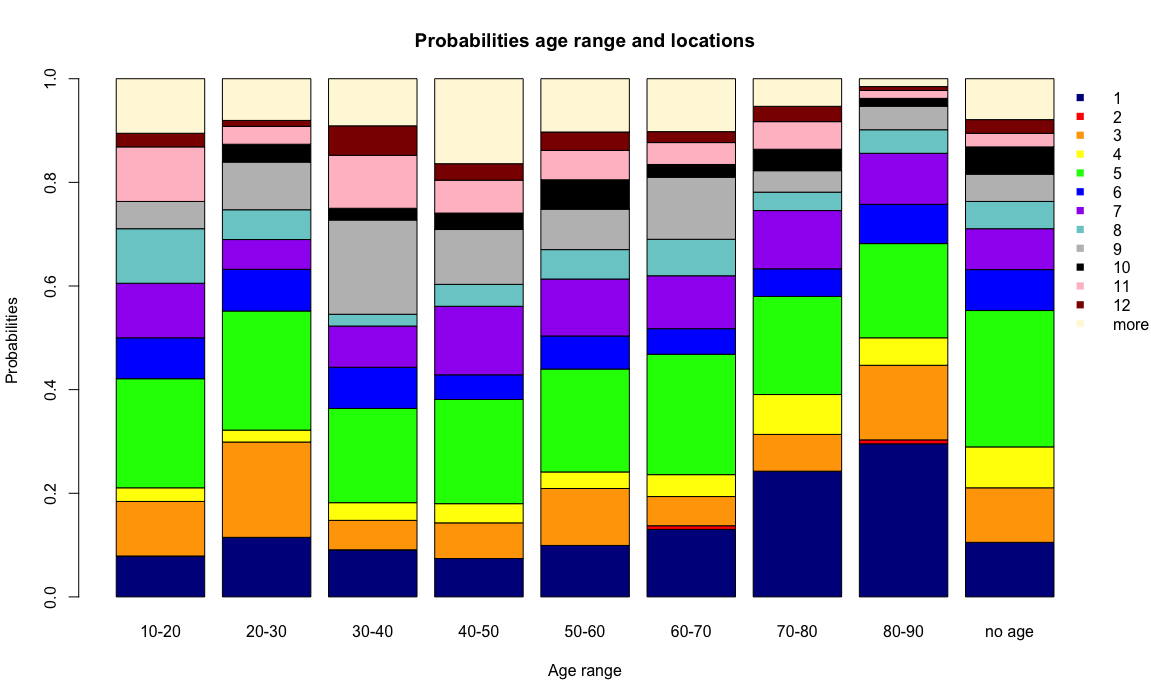
barplot(lacationage, main="Age range and locations",  
 xlab="Number of locations", ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray")  
 )  
text.legend=row.names(lacationage)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



In this plot, two locations has the lowest frequency, it seems abnormal. Actually, it is reasonable, because when one comes out, they go somewhere and come back to home, this is more than three locations, or they stay at home, this is one location, so it is not abnormal.

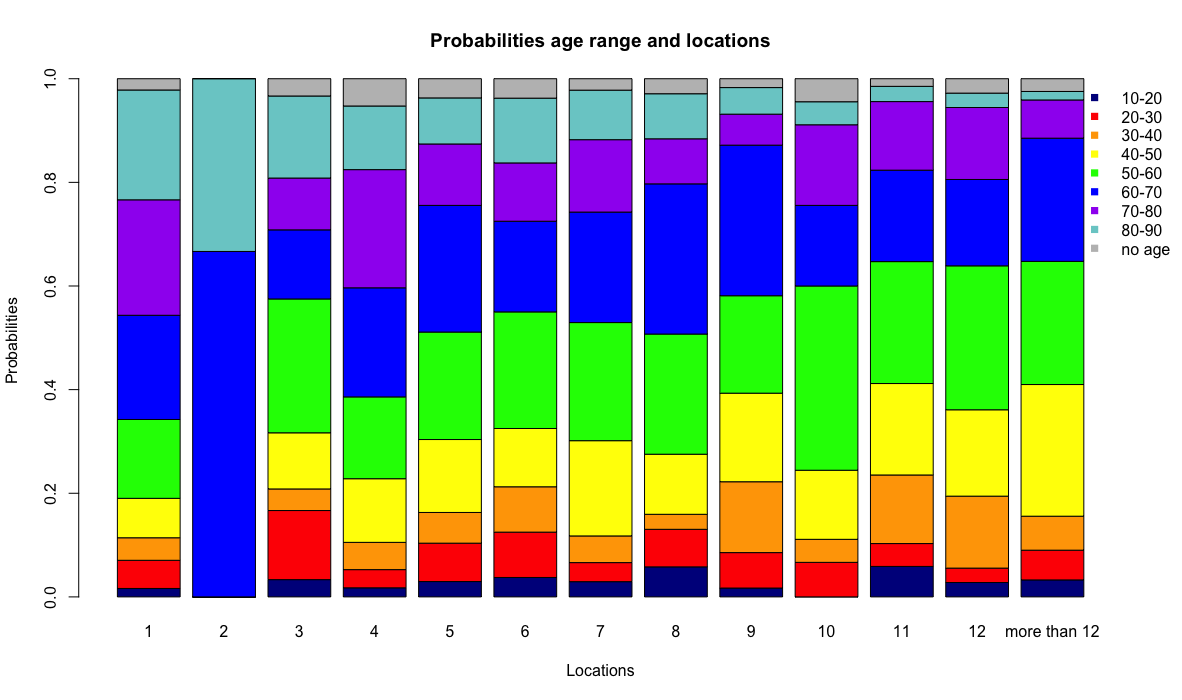
## Make barplot for distribution of probabilities

agelocation\_prop1 <- prop.table(agelocation,2)  
barplot(agelocation\_prop1, main="Probabilities age range and locations",  
 xlab="Age range", ylab = "Probabilities",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk"))  
text.legend=row.names(agelocation\_prop1)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



In this plot, with the increase of age, the frequency of 1 location has a tendency of increase, this is reasonable, because when one is getting older, they may prefer to stay at home rather than going somewhere, the one location can be home. But at age range of 20-30, it is abnormal, the problem may appear at the process of surveying, young people are busy at working, they may be not willing to answer the questions or they treat privacy problem very seriously, so just answer home.

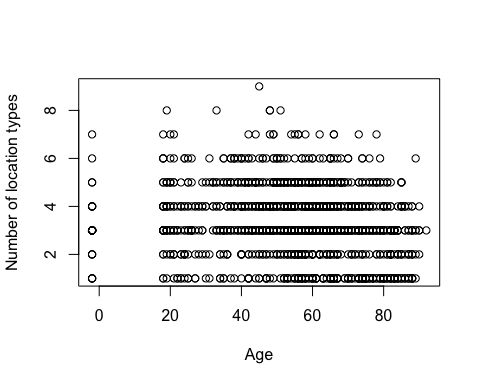
lacationage\_prop1 <- prop.table(lacationage,2)  
barplot(lacationage\_prop1, main="Probabilities age range and locations",  
 xlab="Locations", ylab = "Probabilities",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk"))  
text.legend=row.names(lacationage\_prop1)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



With the increase of the number of locations, the probability of age range 80-90 is smaller and smaller, but with the number of two locations, it is abnormal. But this is probability plot, not frequency, it does not represent people with age range go two locations is more than others.

# Relationship between age and num of location types ## Plot

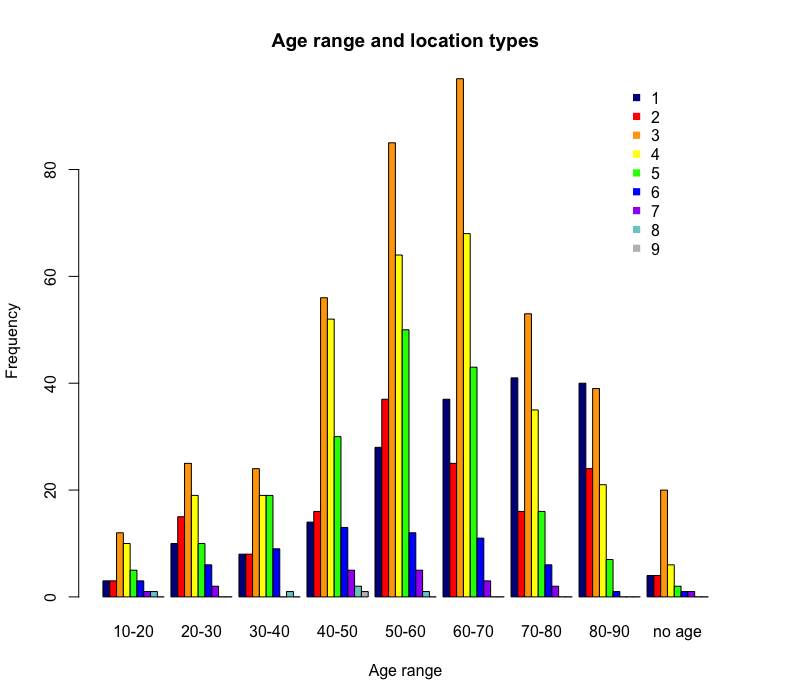
index <- duplicated(encounter[,c(1,6)])  
encounter\_subset\_locationType <- encounter[!index,]  
num\_of\_locationTypes <- data.frame(num = numeric(0))  
for(subject in encounter\_subset$Subject){  
 num\_of\_locationTypes <- rbind(num\_of\_locationTypes, nrow(encounter\_subset\_locationType[which(encounter\_subset\_locationType$Subject == subject), ]))  
}  
num\_of\_locationTypes$age <- encounter\_subset$Age  
plot(num\_of\_locationTypes$X1L~num\_of\_locationTypes$age,xlab = "Age",ylab = "Number of location types")



It is no obvious linear or nonlinear relationship between age and number of location types, so I will count number of location types bases on different age range.

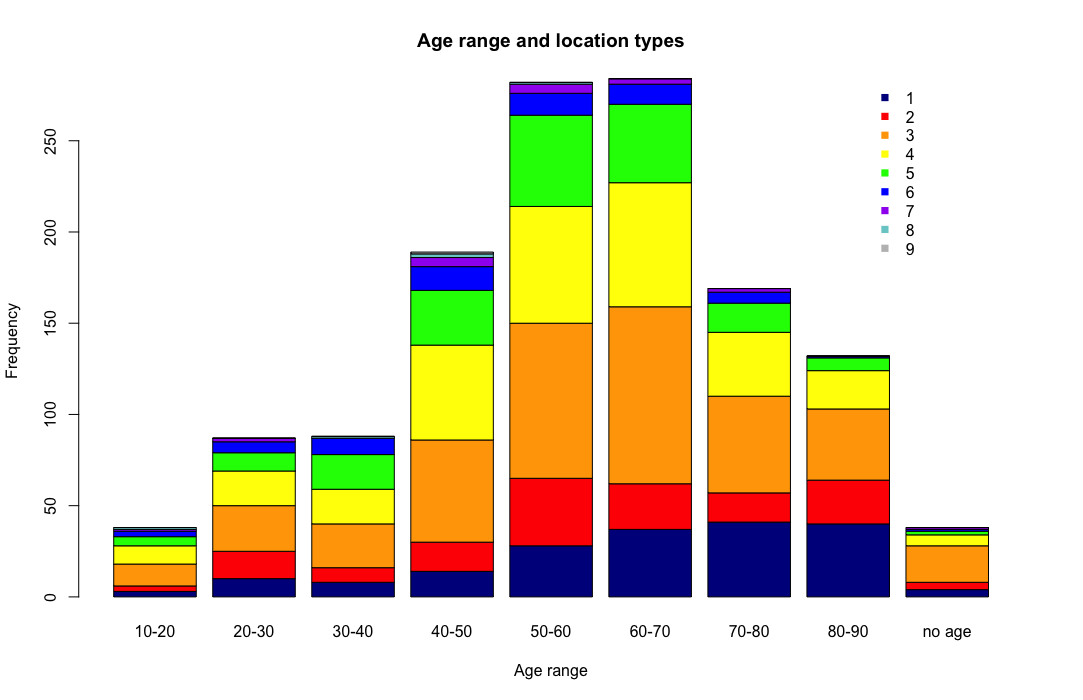
## Make barplot

num\_of\_locationTypes$age\_range <- age\_range\_rep  
agetype <- table(num\_of\_locationTypes$X1L,num\_of\_locationTypes$age\_range)  
typeage <- table(num\_of\_locationTypes$age\_range,num\_of\_locationTypes$X1L)  
barplot(agetype, main="Age range and location types",  
 xlab="Age range", ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray"),  
 beside=TRUE)  
text.legend=row.names(agetype)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



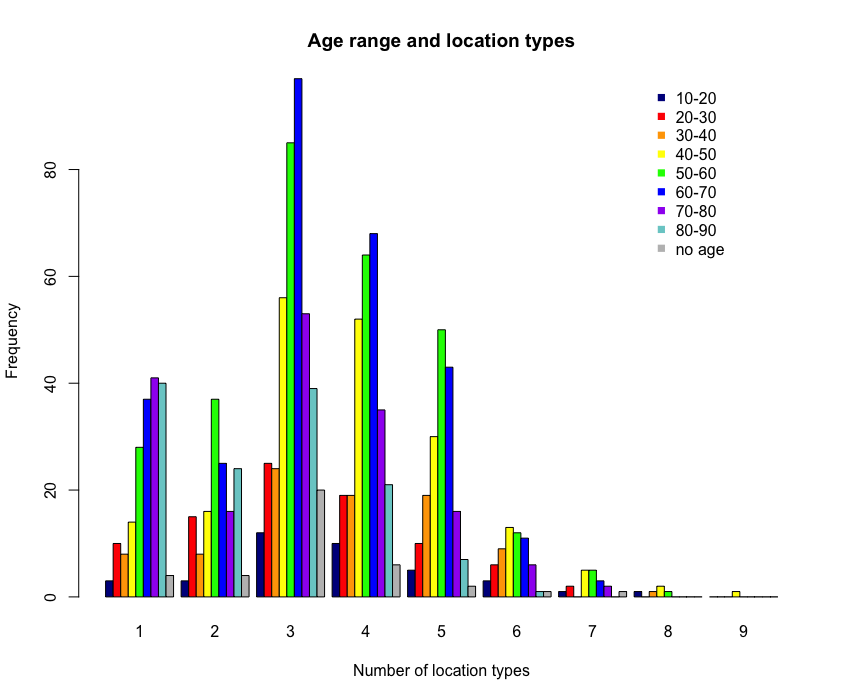
In every age range, three location types always has the most frequency.

barplot(agetype, main="Age range and location types",  
 xlab="Age range", ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray"))  
text.legend=row.names(agetype)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))

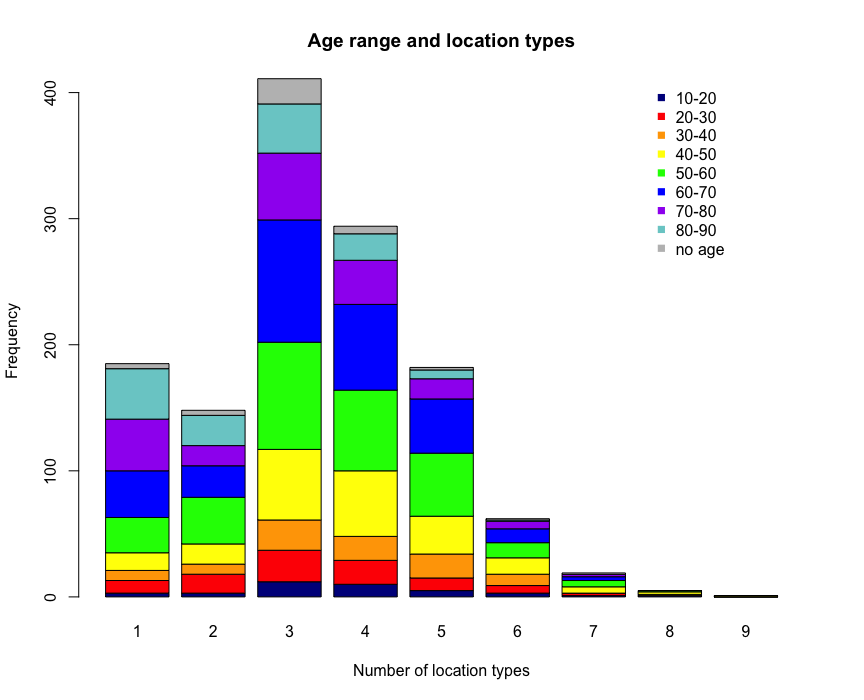


Also, the most participants are at the age range 50-70.

barplot(typeage, main="Age range and location types",  
 xlab="Number of location types", ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray"),  
 beside=TRUE)  
text.legend=row.names(typeage)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))

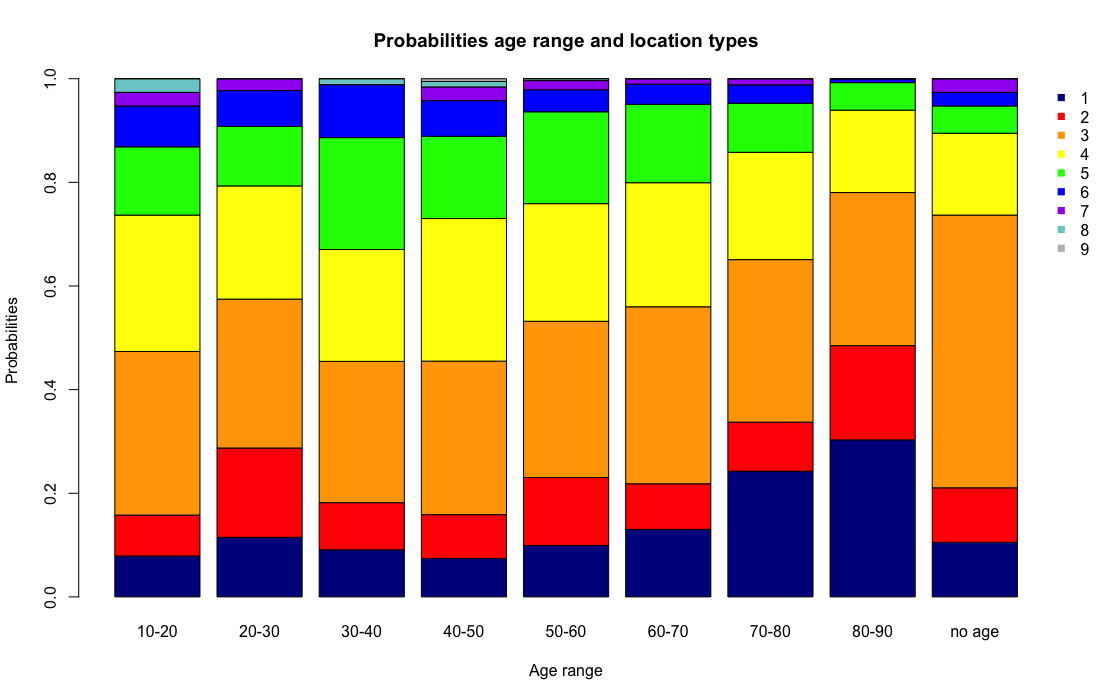


barplot(typeage, main="Age range and location types",  
 xlab="Number of location types", ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray"))  
text.legend=row.names(typeage)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



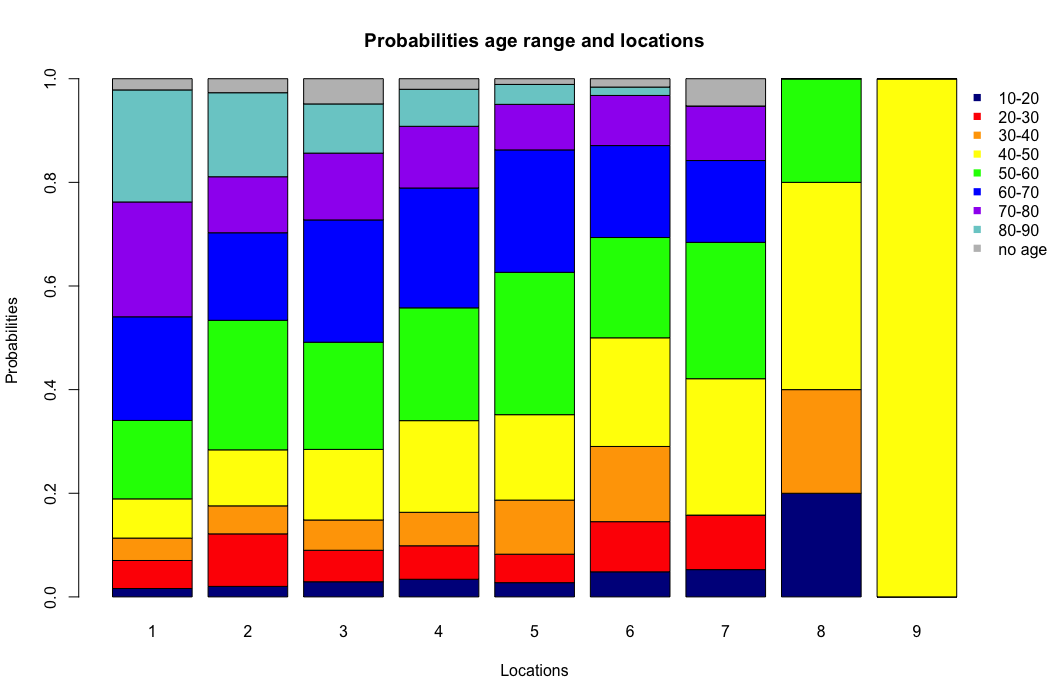
## Make barplot for distribution of probabilities

agetype\_prop1 <- prop.table(agetype,2)  
barplot(agetype\_prop1, main="Probabilities age range and location types",  
 xlab="Age range", ylab = "Probabilities",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray"))  
text.legend=row.names(agetype\_prop1)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



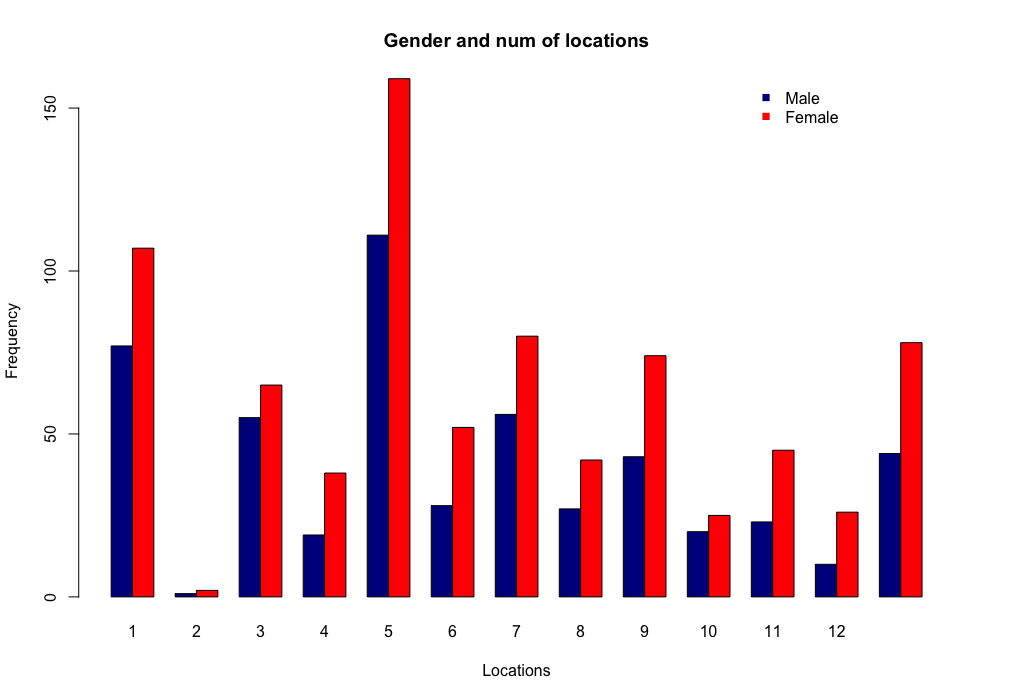
As the number of locations, the abnormal point appears at age range 20-30, its frequency of 1 is not coherent of the tendency of increase, just the same reason, the problem might occur at the process of data collecting.

typeage\_prop <- prop.table(typeage,2)  
barplot(typeage\_prop, main="Probabilities age range and locations",  
 xlab="Locations", ylab = "Probabilities",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk"))  
text.legend=row.names(typeage\_prop)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))

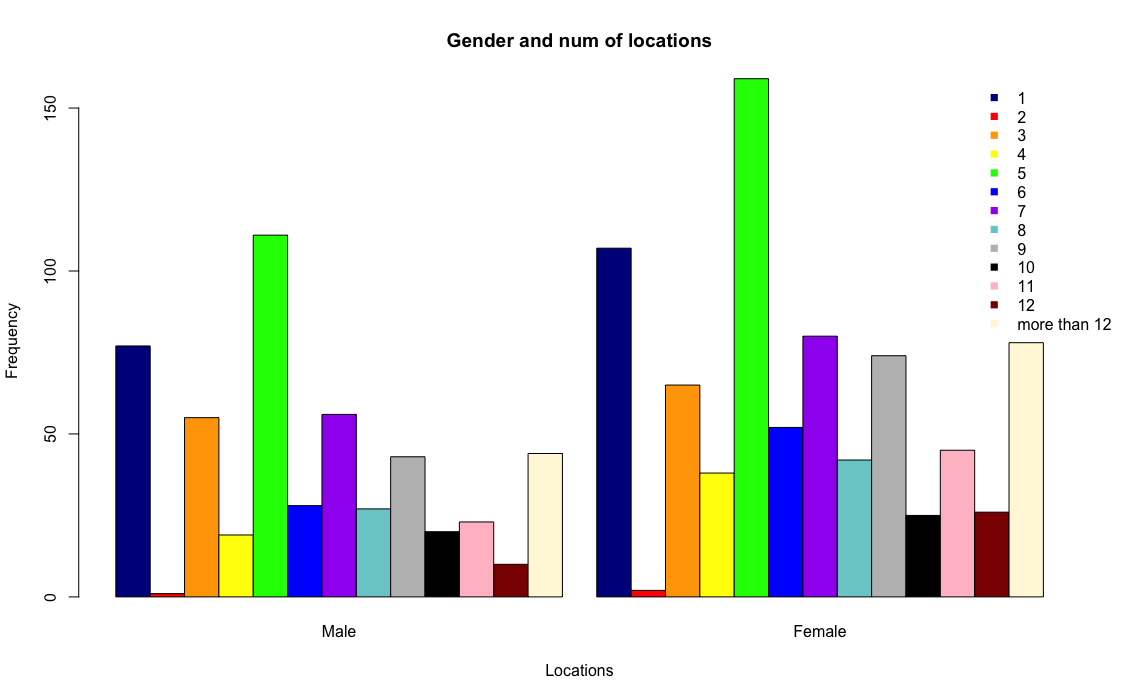


# Relationship between gender and number of locations

num\_of\_locations$gender <- encounter\_subset$Gender  
num\_of\_locations$gender <- factor(num\_of\_locations$gender,  
 levels = c(1,2),  
 labels = c("Male","Female"))  
genderlocation <- table(num\_of\_locations$gender,num\_of\_locations$location\_range)  
locationgender <- table(num\_of\_locations$location\_range,num\_of\_locations$gender)  
barplot(genderlocation,main = "Gender and num of locations",   
 xlab = "Locations", ylab = "Frequency",  
 col = c("darkblue","red"),  
 beside = TRUE)  
text.legend=row.names(genderlocation)  
col2 <- c("darkblue","red")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))

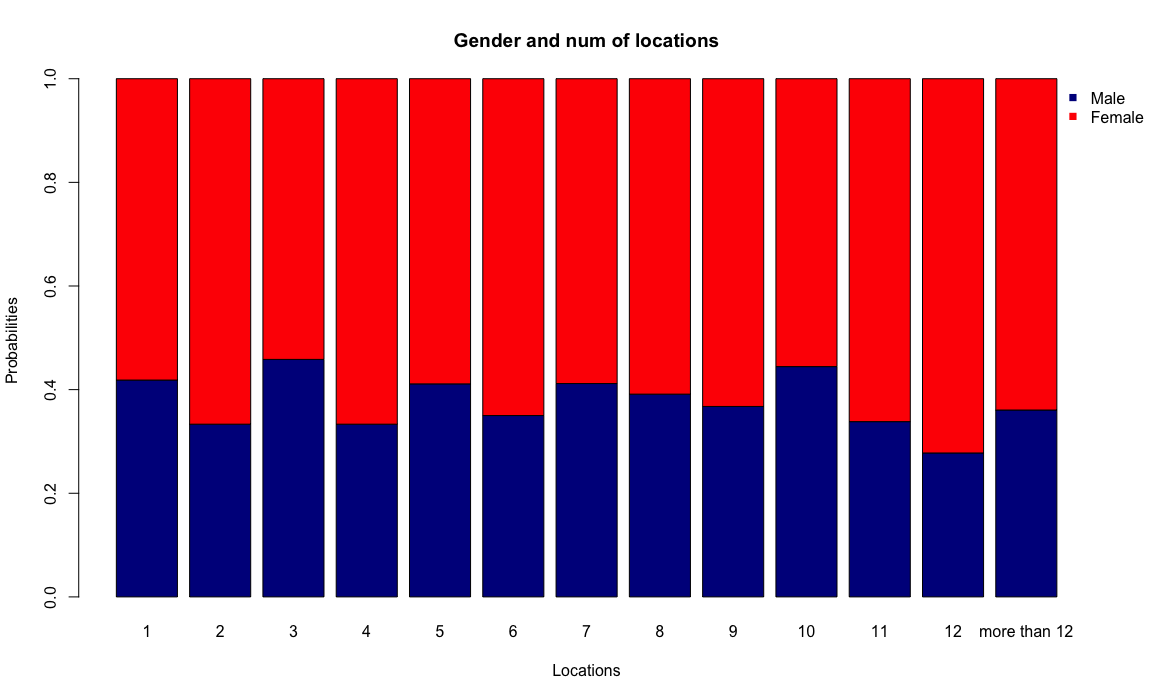


barplot(locationgender,main = "Gender and num of locations",   
 xlab = "Locations", ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk"),  
 beside = TRUE)  
text.legend=row.names(locationgender)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



## Barplot of probabilities

genderlocation\_prop <- prop.table(genderlocation,2)  
barplot(genderlocation\_prop,main = "Gender and num of locations",   
 xlab = "Locations", ylab = "Probabilities",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk"))  
text.legend=row.names(genderlocation\_prop)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



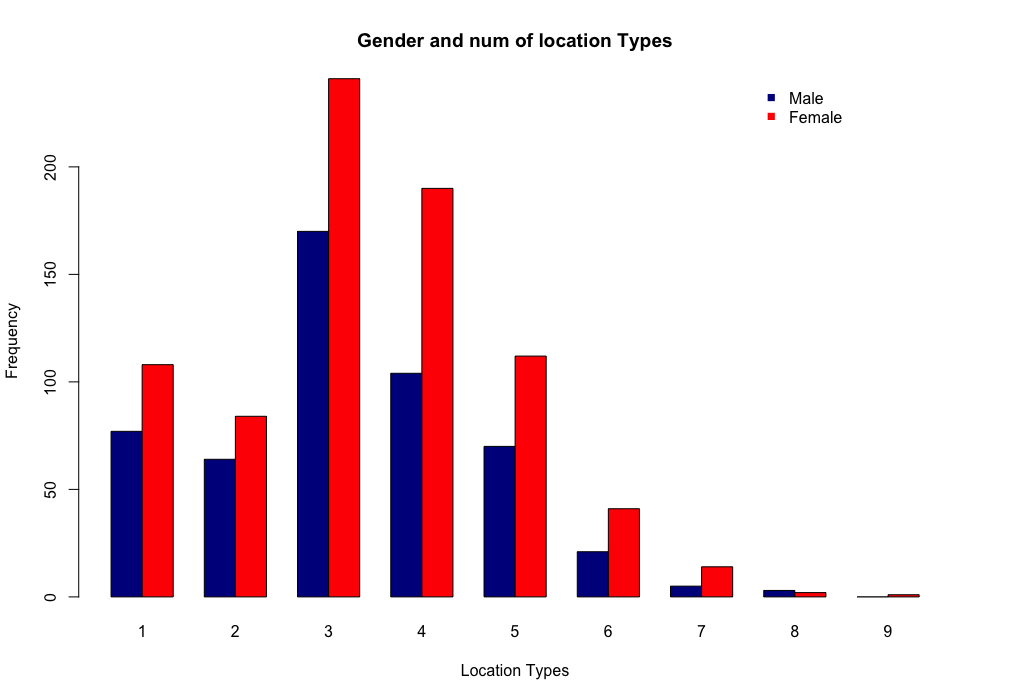
locationgender\_prop <- prop.table(locationgender,2)  
barplot(locationgender\_prop,main = "Gender and num of locations",   
 xlab = "Locations",ylab = "Probabilities",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk"))  
text.legend=row.names(locationgender\_prop)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



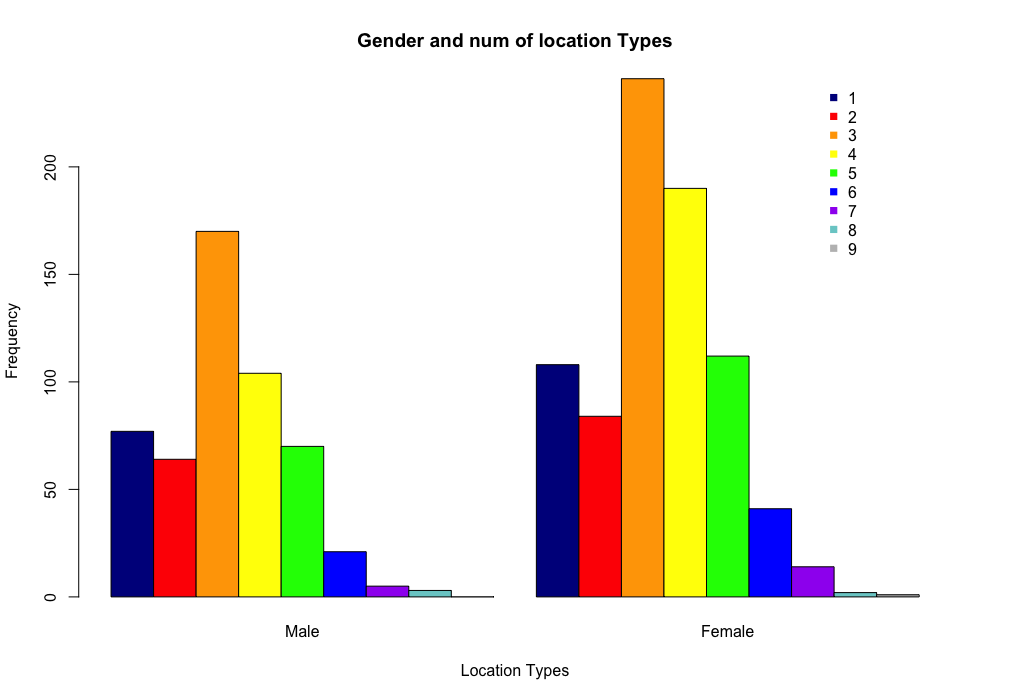
The distribution of different number of locations between male and female are almost the same.

# Relationship between gender and number of location types

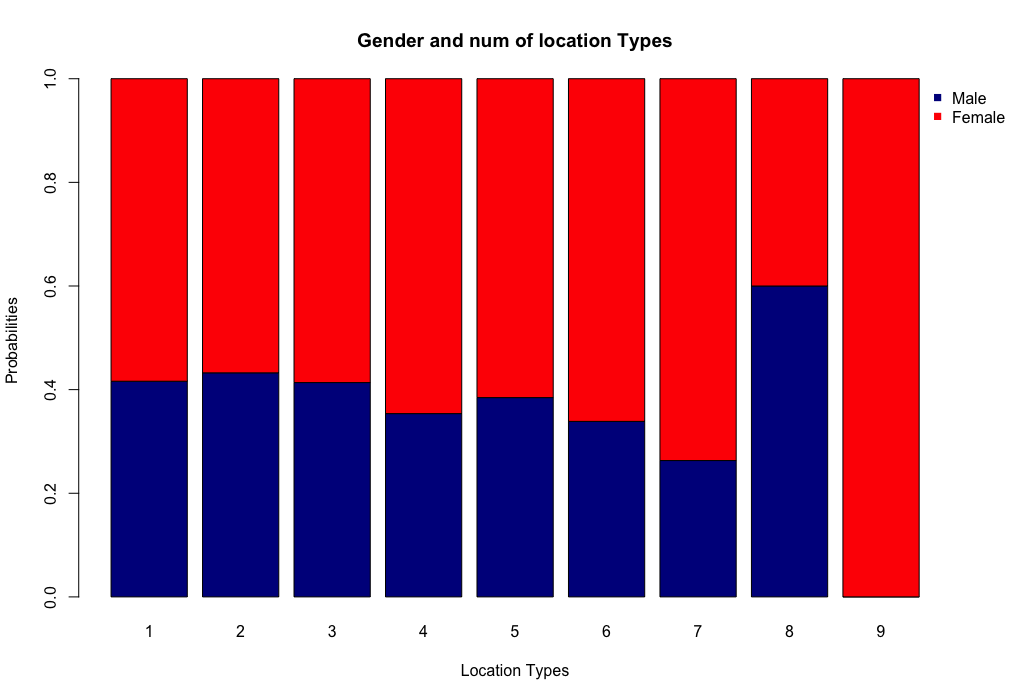
num\_of\_locationTypes$gender <- encounter\_subset$Gender  
num\_of\_locationTypes$gender <- factor(num\_of\_locationTypes$gender,  
 levels = c(1,2),  
 labels = c("Male","Female"))  
gendertype <- table(num\_of\_locationTypes$gender,num\_of\_locationTypes$X1L)  
typegender <- table(num\_of\_locationTypes$X1L,num\_of\_locationTypes$gender)  
barplot(gendertype,main = "Gender and num of location Types",   
 xlab = "Location Types",ylab = "Frequency",  
 col = c("darkblue","red"),  
 beside = TRUE)  
text.legend=row.names(gendertype)  
col2 <- c("darkblue","red")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



barplot(typegender,main = "Gender and num of location Types",   
 xlab = "Location Types",ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray"),  
 beside = TRUE)  
text.legend=row.names(typegender)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))

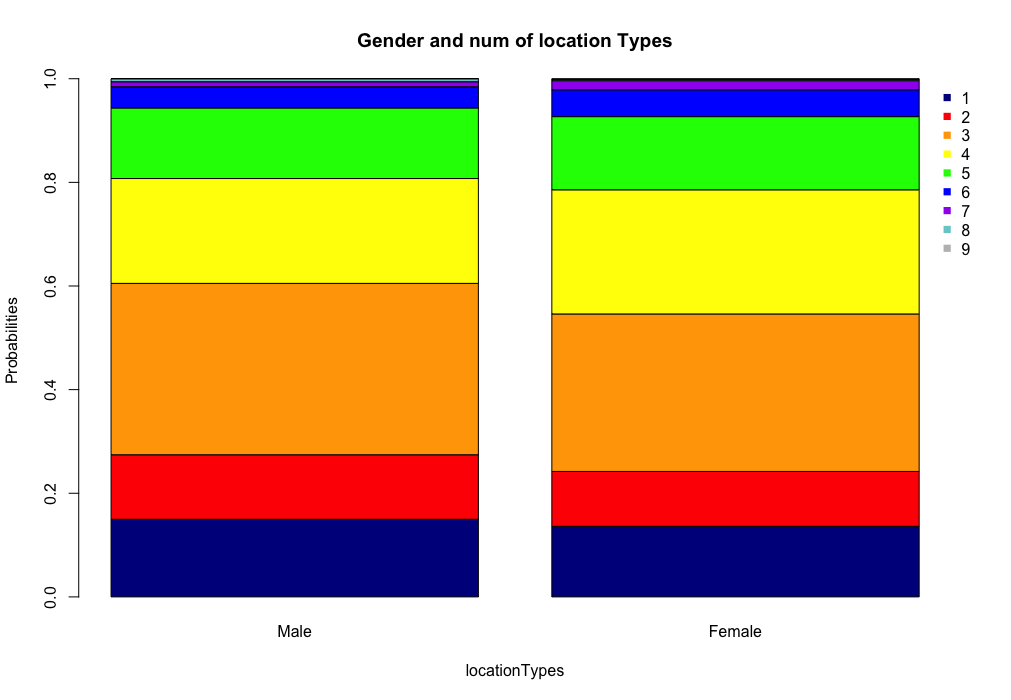
 ## Probabilities

gendertype\_prop <- prop.table(gendertype,2)  
barplot(gendertype\_prop,main = "Gender and num of location Types",   
 xlab = "Location Types",ylab = "Probabilities",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk"))  
text.legend=row.names(gendertype\_prop)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



When locations types is 9, 100%probability it is female, this can be explained as more female need to take care of children and send them to school or somewhere else, which will increase number of location types.

typegender\_prop <- prop.table(typegender,2)  
barplot(typegender\_prop,main = "Gender and num of location Types",   
 xlab = "locationTypes",ylab = "Probabilities",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk"),  
 legend = row.names(typegender\_prop))  
text.legend=row.names(typegender\_prop)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



Also, the distribution are almostly same.

# Unique locations # Strip out movement data

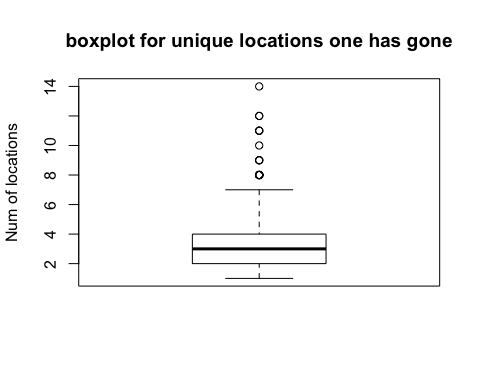
drop\_transportation <- encounter[which(encounter$Location\_Type != "Private Transport (private or rented car, motorbike)"),]  
drop\_transportation\_pri <- drop\_transportation[which(drop\_transportation$Location\_Type != "Public Transport (train, tram, bus or taxi)"),]  
drop\_transportation\_all <- drop\_transportation\_pri[which(drop\_transportation\_pri$Location\_Type != "Car journey (respondent alone in the car)"),]

## Judge unique bases on (lantitude,lontitude)

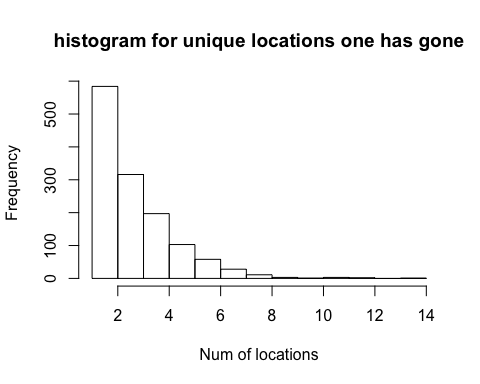
index <- duplicated(drop\_transportation\_all[,c(1,9,10)])  
encounter\_drop\_transportation <- drop\_transportation\_all[!index,]  
  
unique\_location <- data.frame(num = numeric(0))  
for(subject in encounter\_subset$Subject){  
 unique\_location <- rbind(unique\_location, nrow(encounter\_drop\_transportation[which(encounter\_drop\_transportation$Subject == subject), ]))  
}  
unique\_location$subject <- encounter\_subset$Subject  
unique\_location$age <- encounter\_subset$Age  
unique\_location$gender <- encounter\_subset$Gender  
unique\_location$age\_range <- age\_range\_rep

## Boxplot and histogram

boxplot(unique\_location$X1L,main="boxplot for unique locations one has gone",ylab = "Num of locations")

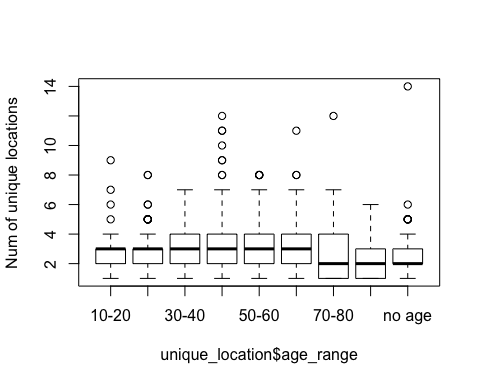


hist(unique\_location$X1L,main="histogram for unique locations one has gone",xlab = "Num of locations")

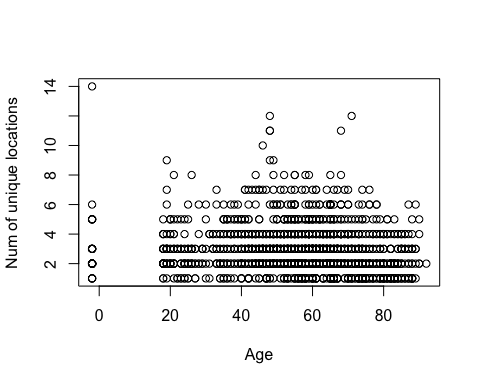


# Relationship between age and unique location

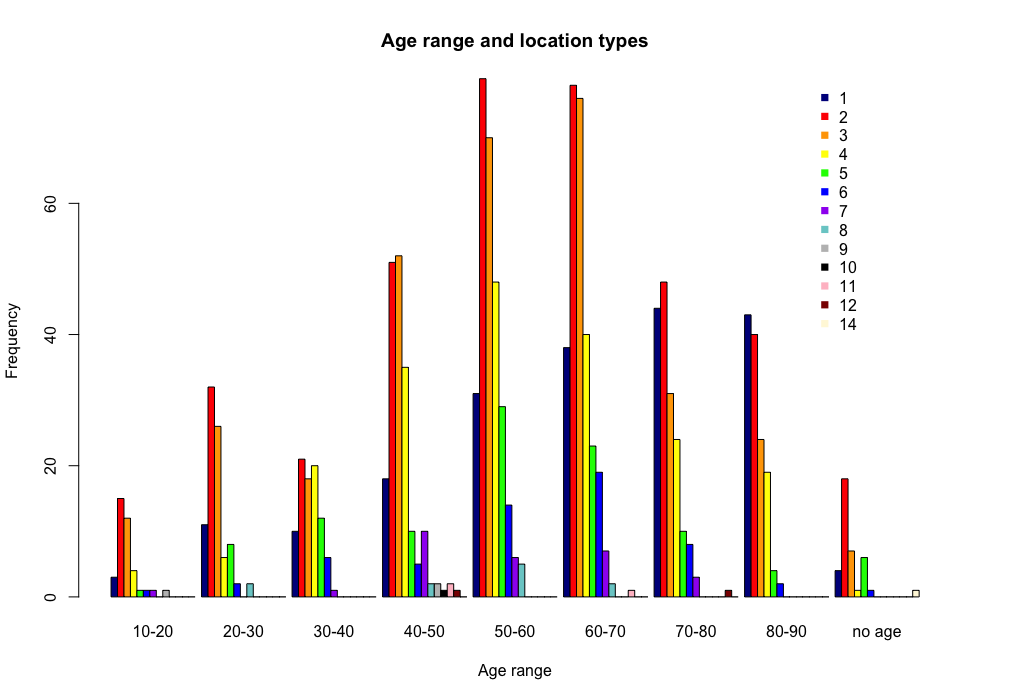
boxplot(unique\_location$X1L~unique\_location$age\_range,ylab = "Num of unique locations")



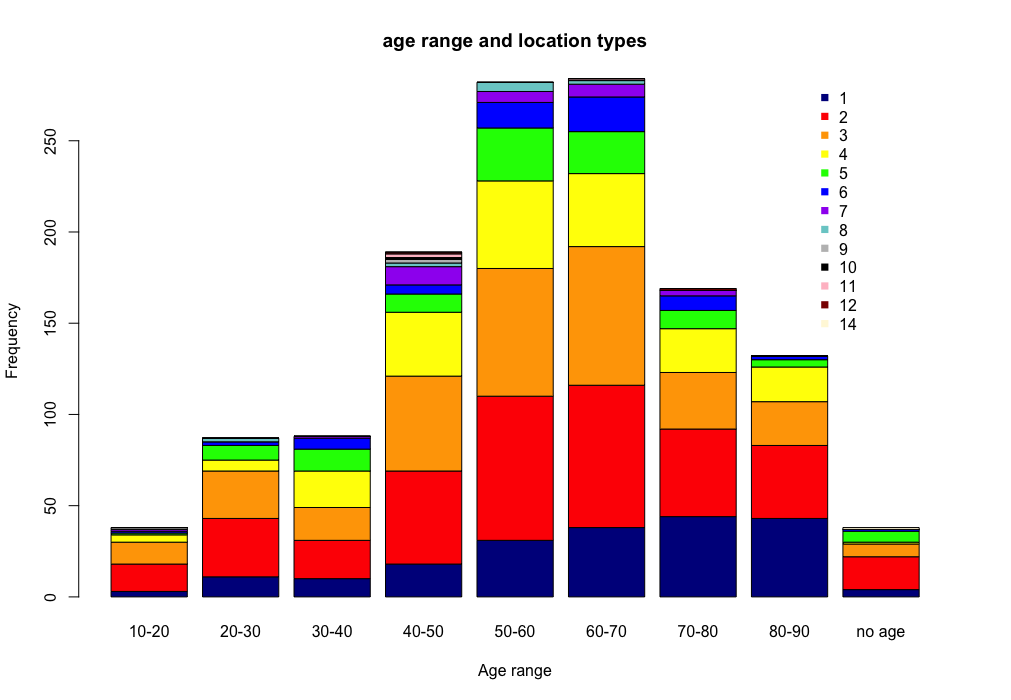
plot(unique\_location$X1L~unique\_location$age,xlab = "Age", ylab = "Num of unique locations")

 ## Barplot of age range and unique location

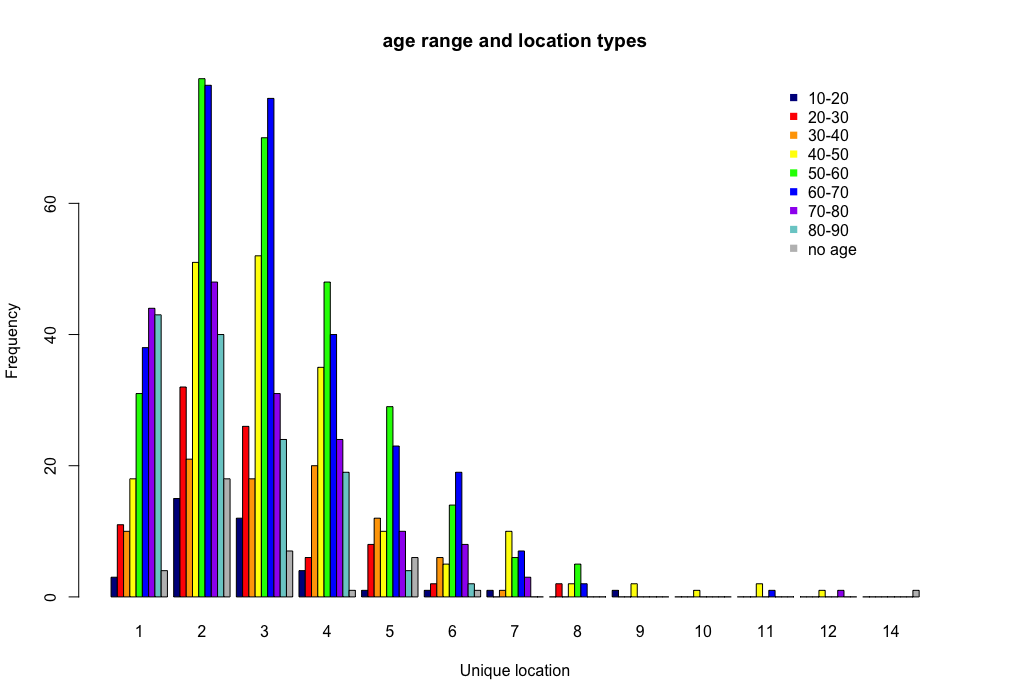
ageunique <- table(unique\_location$X1L,unique\_location$age\_range)  
uniqueage <- table(unique\_location$age\_range,unique\_location$X1L)  
barplot(ageunique, main="Age range and location types",  
 xlab="Age range", ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk"),  
 beside=TRUE)  
text.legend=row.names(ageunique)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



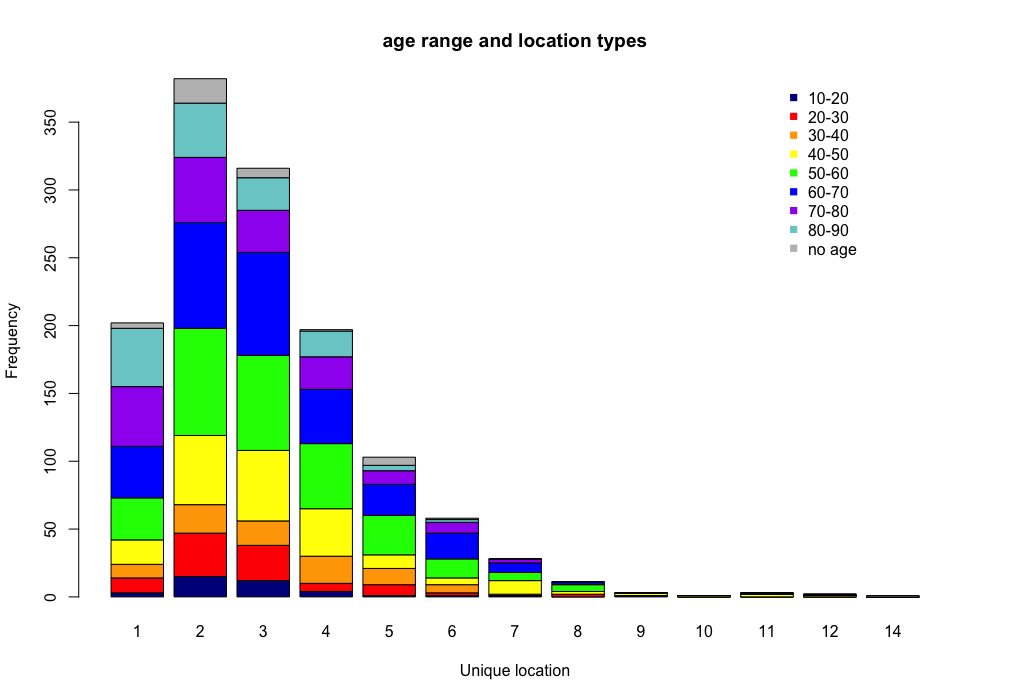
barplot(ageunique, main="age range and location types",  
 xlab="Age range", ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk"))  
text.legend=row.names(ageunique)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



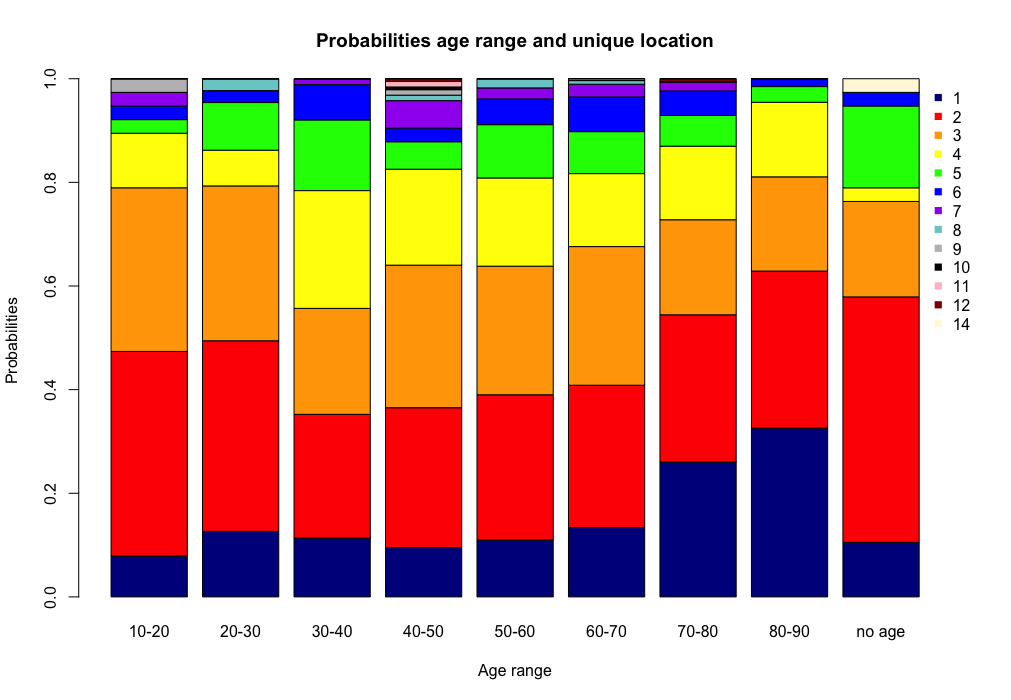
barplot(uniqueage, main="age range and location types",  
 xlab="Unique location", ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray"),  
 beside=TRUE)  
text.legend=row.names(uniqueage)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



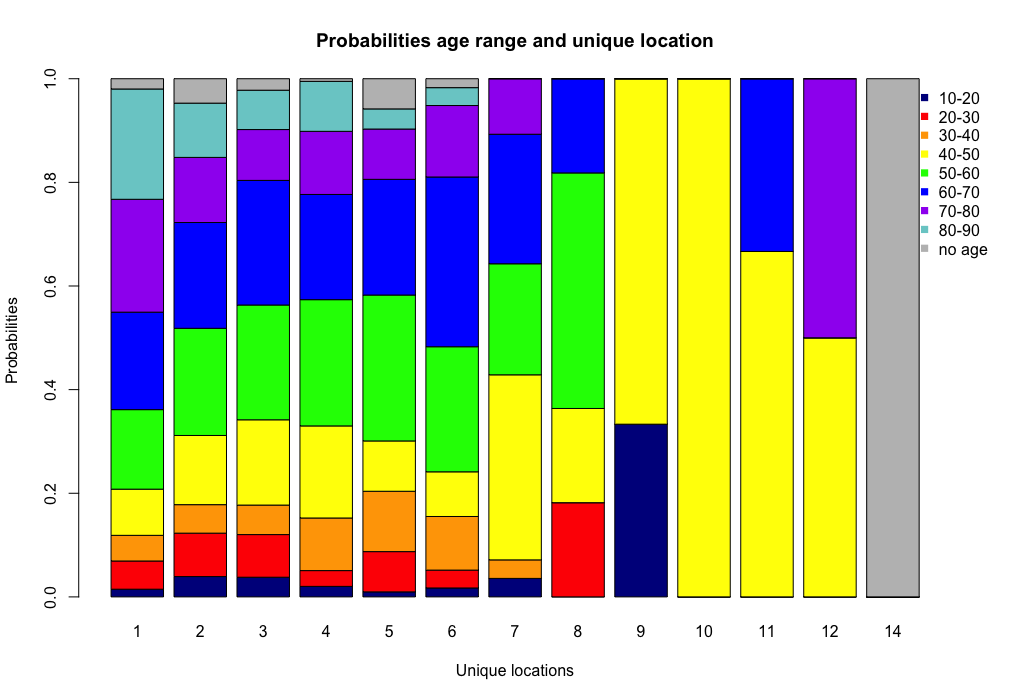
barplot(uniqueage, main="age range and location types",  
 xlab="Unique location", ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray"))  
text.legend=row.names(uniqueage)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))

 ## Probabilities

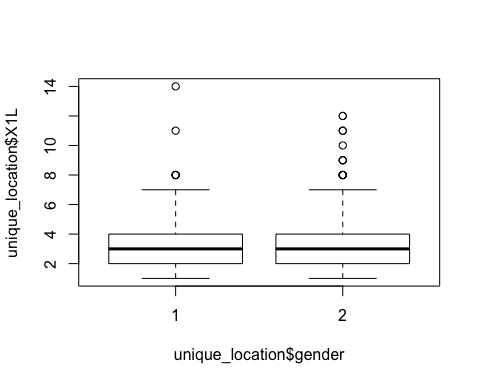
ageunique\_prop1 <- prop.table(ageunique,2)  
barplot(ageunique\_prop1, main="Probabilities age range and unique location",  
 xlab="Age range", ylab = "Probabilities",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk"))  
text.legend=row.names(ageunique\_prop1)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



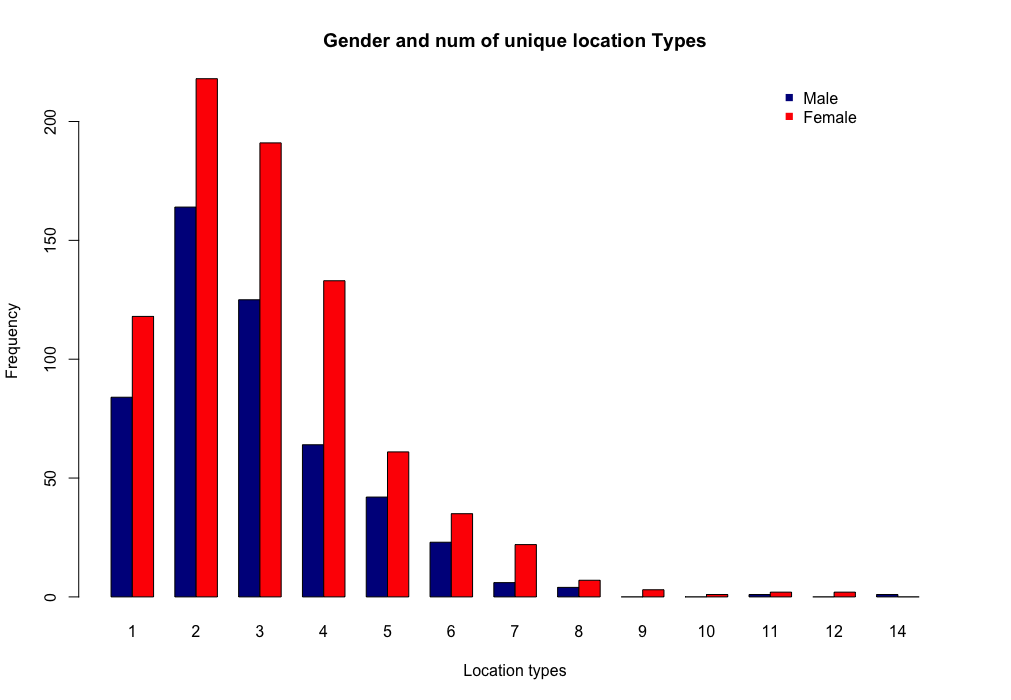
uniqueage\_prop <- prop.table(uniqueage,2)  
barplot(uniqueage\_prop, main="Probabilities age range and unique location",  
 xlab="Unique locations", ylab = "Probabilities",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk"))  
text.legend=row.names(uniqueage\_prop)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))

 # Relationship: gender and num of unique location

boxplot(unique\_location$X1L~unique\_location$gender)

 ## Barplot

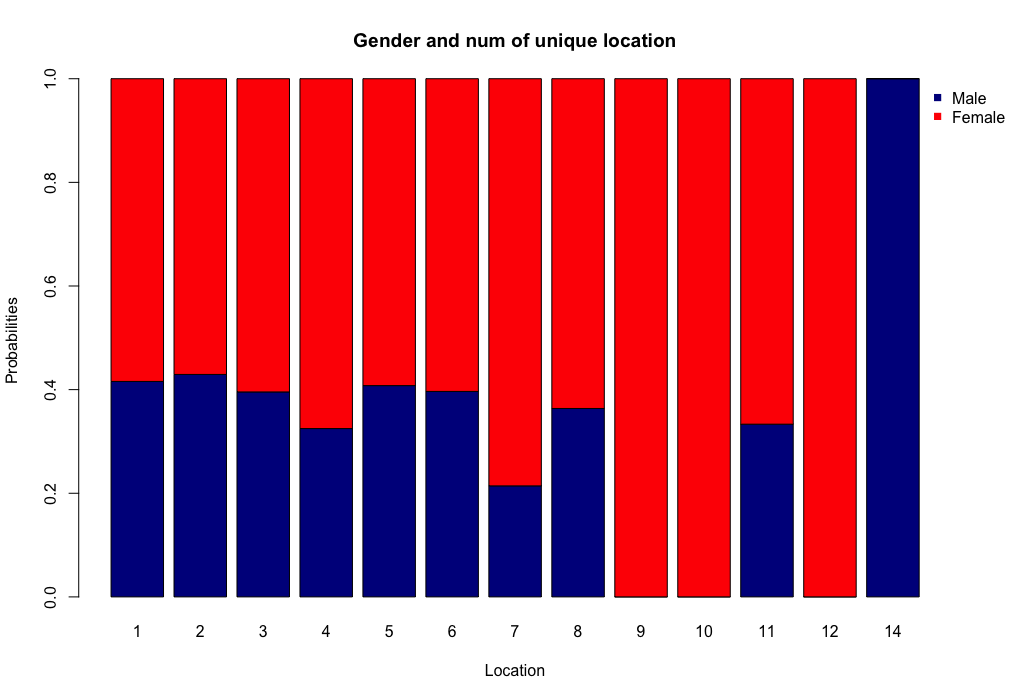
unique\_location$gender <- factor(unique\_location$gender,  
 levels = c(1,2),  
 labels = c("Male","Female"))  
genderunique <- table(unique\_location$gender,unique\_location$X1L)  
uniquegender <- table(unique\_location$X1L,unique\_location$gender)  
barplot(genderunique,main = "Gender and num of unique location Types",   
 xlab = "Location types",ylab = "Frequency",  
 col = c("darkblue","red"),  
 beside = TRUE)  
text.legend=row.names(genderunique)  
col2 <- c("darkblue","red")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



barplot(uniquegender,main = "Gender and num of unique location",   
 xlab = "Location types",ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk"),  
 beside = TRUE)  
text.legend=row.names(uniquegender)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))

 ## Probabilities

genderunique\_prop <- prop.table(genderunique,2)  
barplot(genderunique\_prop,main = "Gender and num of unique location",   
 xlab = "Location", ylab = "Probabilities",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk"))  
text.legend=row.names(genderunique\_prop)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))

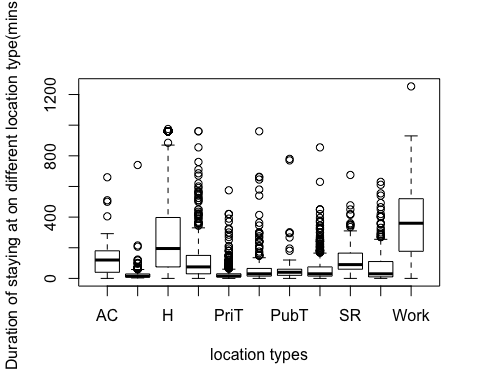


uniquegender\_prop <- prop.table(uniquegender,2)  
barplot(uniquegender\_prop,main = "Gender and num of unique location",   
 xlab = "Location", ylab = "Probabilities",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk"))  
text.legend=row.names(uniquegender\_prop)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink","darkred","cornsilk")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))

 # The relationship between duration and different location types

encounter\_duration <- encounter[which(encounter$Duration!="NULL"), ]  
encounter\_type\_duration <- encounter[which(encounter$Location\_Type != "NULL"),]  
  
encounter\_type\_duration$Location\_Type <- factor(encounter\_type\_duration$Location\_Type,  
 levels = c("Home","Private Transport (private or rented car, motorbike)",  
 "Retail and hospitality (bars, cafes, shops, hair dressing, etc.) ",  
 "Car journey (respondent alone in the car)",  
 "Sport and recreation","Other ","Work",  
 "Public spaces (parks, streets, stations, airports etc.)",  
 "Study (Uni, school, kindergarden, childcare centre etc)",  
 "Arts and culture (cinema, library, gigs, museum, theatre etc.) ",  
 "Public Transport (train, tram, bus or taxi)"),  
 labels = c("H","PriT",  
 "RT",  
 "CJ",  
 "SR","Other","Work",  
 "PS",  
 "Study",  
 "AC",  
 "PubT"))  
Location\_Type <- data.frame(sapply(encounter\_type\_duration$Location\_Type, function(x) { if(is.factor(x)) {  
 as.character(x)  
} else {  
 x  
}  
}))  
  
Duration <- data.frame(sapply(encounter\_type\_duration$Duration, function(x) { if(is.factor(x)) {  
 as.numeric(as.character(x))  
} else {  
 x  
}  
}))

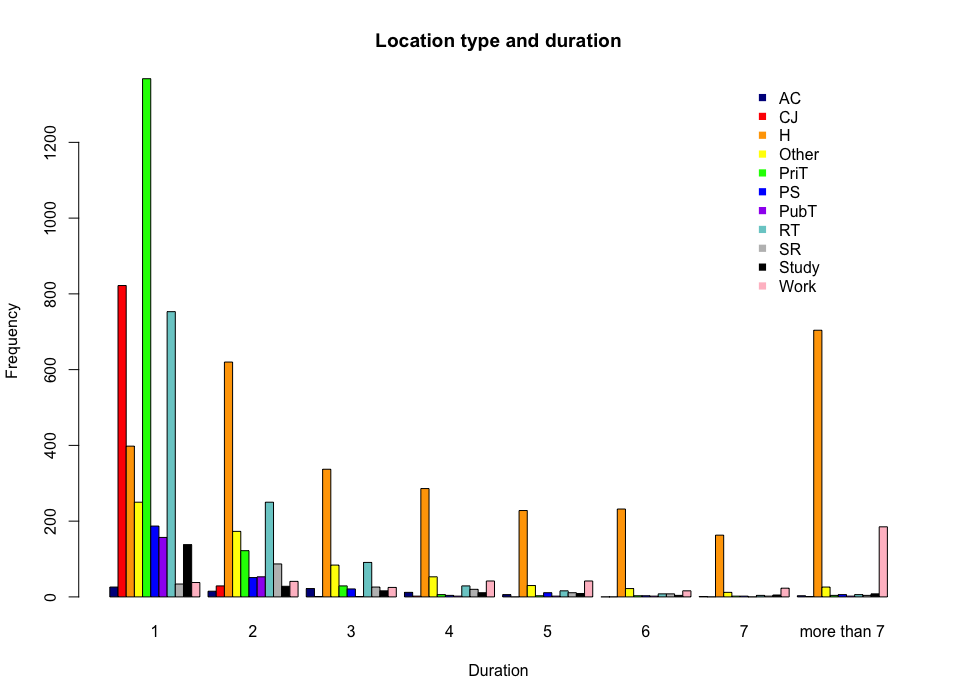
encounter\_type\_duration$Location\_Type <- Location\_Type$sapply.encounter\_type\_duration.Location\_Type..function.x...  
encounter\_type\_duration$Duration <- Duration$sapply.encounter\_type\_duration.Duration..function.x...  
boxplot(encounter\_type\_duration$Duration ~ encounter\_type\_duration$Location\_Type, xlab = "location types",ylab = "Duration of staying at on different location type(mins)")



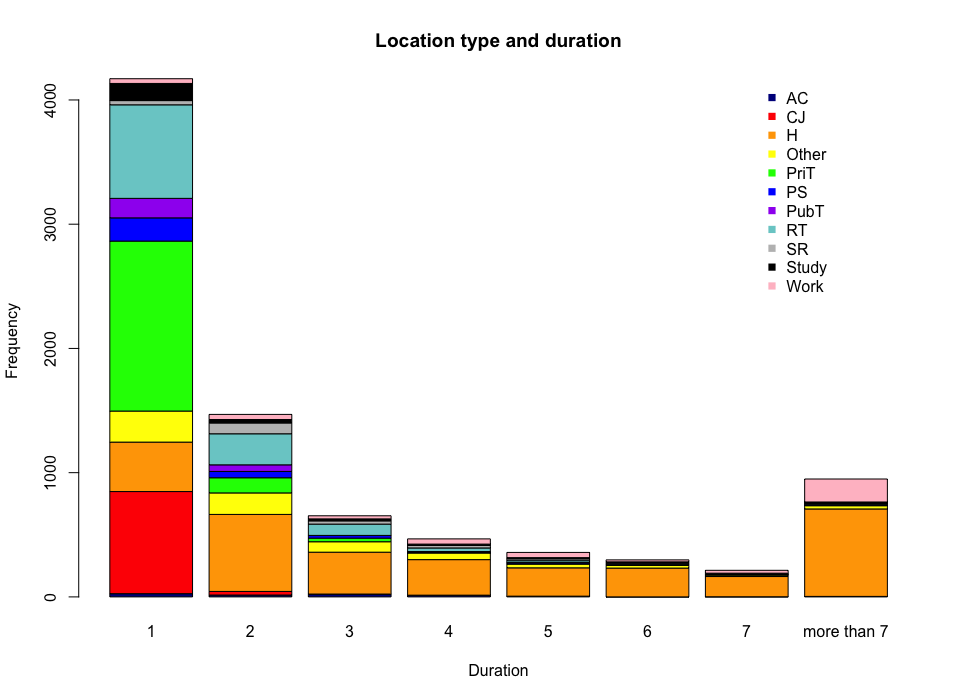
encounter\_type\_duration$Duration[which(encounter\_type\_duration$Duration >= 0 & encounter\_type\_duration$Duration < 60)] <- 1  
encounter\_type\_duration$Duration[which(encounter\_type\_duration$Duration >= 60 & encounter\_type\_duration$Duration < 120)] <- 2  
encounter\_type\_duration$Duration[which(encounter\_type\_duration$Duration >= 120 & encounter\_type\_duration$Duration < 180)] <- 3  
encounter\_type\_duration$Duration[which(encounter\_type\_duration$Duration >= 180 & encounter\_type\_duration$Duration < 240)] <- 4  
encounter\_type\_duration$Duration[which(encounter\_type\_duration$Duration >= 240 & encounter\_type\_duration$Duration < 300)] <- 5  
encounter\_type\_duration$Duration[which(encounter\_type\_duration$Duration >= 300 & encounter\_type\_duration$Duration < 360)] <- 6  
encounter\_type\_duration$Duration[which(encounter\_type\_duration$Duration >= 360 & encounter\_type\_duration$Duration < 420)] <- 7  
encounter\_type\_duration$Duration[which(encounter\_type\_duration$Duration >= 420)] <- "more than 7"

## Barplot

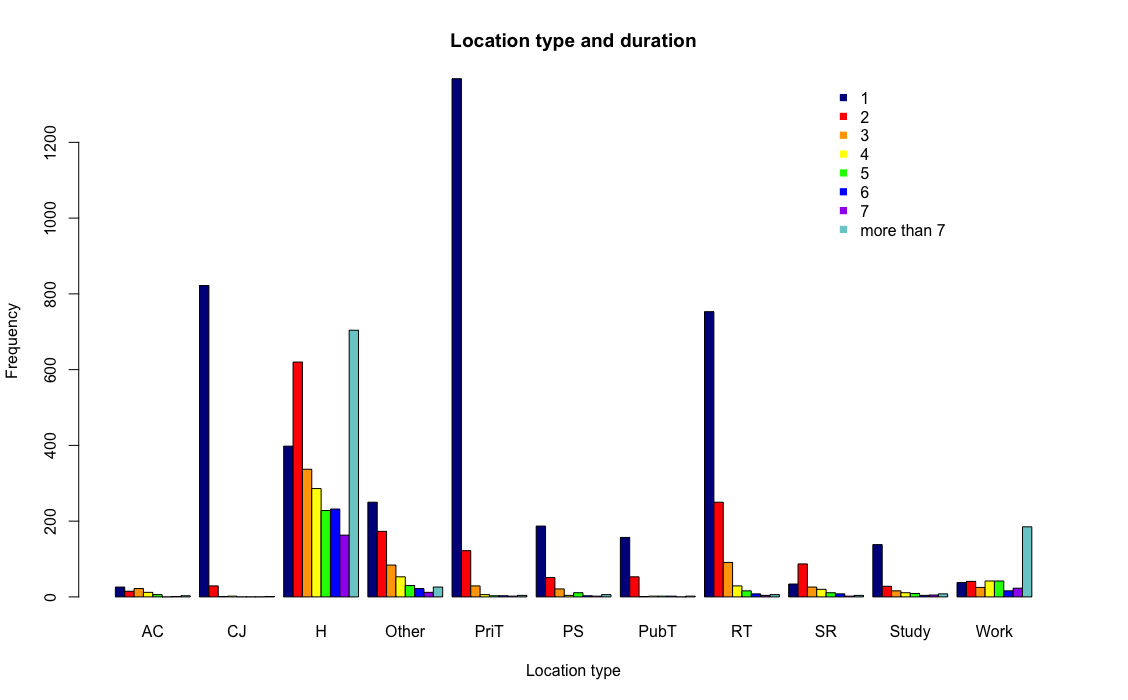
type\_duration <- table(encounter\_type\_duration$Location\_Type,encounter\_type\_duration$Duration)  
duration\_type <- table(encounter\_type\_duration$Duration,encounter\_type\_duration$Location\_Type)  
barplot(type\_duration,main = "Location type and duration",   
 xlab = "Duration",ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink"),  
 beside = TRUE)  
text.legend=row.names(type\_duration)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



barplot(type\_duration,main = "Location type and duration",   
 xlab = "Duration",ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink"))  
text.legend=row.names(type\_duration)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3","gray","black","pink")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



barplot(duration\_type,main = "Location type and duration",   
 xlab = "Location type",ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3"),  
 beside = TRUE)  
text.legend=row.names(duration\_type)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))



barplot(duration\_type,main = "Location type and duration",   
 xlab = "Location type",ylab = "Frequency",  
 col=c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3"),  
 legend = row.names(duration\_type))  
text.legend=row.names(duration\_type)  
col2 <- c("darkblue","red","orange","yellow","green","blue","purple","darkslategray3")  
legend("topright",pch=c(15),legend=text.legend,col=col2,bty="n",ncol = 1,inset=c(-0.25,0))

