MarkdownDocument

Group 5 MD

2022-09-25

Here you can write the introduction of your report and include some text that will be transferred to the word file every time you re-run this Markdown

//Escriure INTRO

## [1] "data.frame"

First of all, we obtain the dimensions of our accidents dataset:

dim(dd)

## [1] 5000 23

n<-dim(dd)[1]  
K<-dim(dd)[2]  
  
n

## [1] 5000

K

## [1] 23

Next, we proceed to check the variables’ names:

names(dd)

## [1] "Zone" "Date" "Region" "Prov" "nMortal"   
## [6] "nGraveInj" "nMinorInj" "nInv" "nPedest" "nBikes"   
## [11] "nMotor" "Vel" "Escaped" "Weather" "TrafficInf"  
## [16] "WeatherInf" "LightInf" "VisionInf" "Intersect" "Surface"   
## [21] "DayGroup" "HourGroup" "AccType"

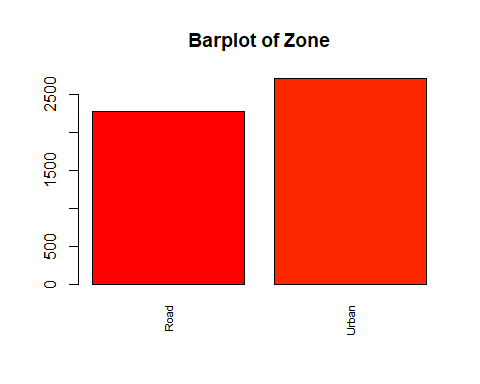
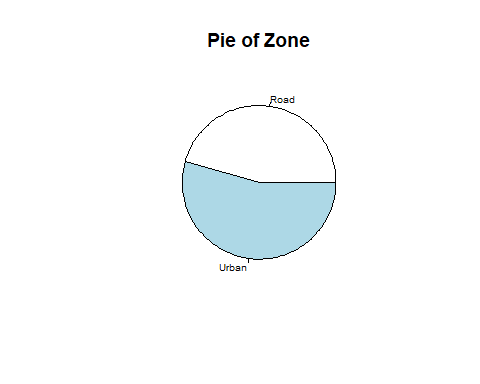
Decide if you need to declare some more factor or date

descriptive<-function(X, nom){  
 if (!(is.numeric(X) || class(X)=="Date")){  
 frecs<-table(as.factor(X), useNA="ifany")  
 proportions<-frecs/n  
 #ojo, decidir si calcular porcentages con o sin missing values  
 pie(frecs, cex=0.6, main=paste("Pie of", nom))  
 barplot(frecs, las=3, cex.names=0.7, main=paste("Barplot of", nom), col=listOfColors)  
 print(paste("Number of modalities: ", length(frecs)))  
 print("Frequency table")  
 print(frecs)  
 print("Relative frequency table (proportions)")  
 print(proportions)  
 print("Frequency table sorted")  
 print(sort(frecs, decreasing=TRUE))  
 print("Relative frequency table (proportions) sorted")  
 print(sort(proportions, decreasing=TRUE))  
 }else{  
 if(class(X)=="Date"){  
 print(summary(X))  
 print(sd(X))  
 #decide breaks: weeks, months, quarters...  
 hist(X, main=paste("Histogram of", nom), breaks="years")  
 }else{  
 hist(X, main=paste("Histogram of", nom))  
 boxplot(X, horizontal=TRUE, main=paste("Boxplot of",nom))  
 print("Extended Summary Statistics")  
 print(summary(X))  
 print(paste("sd: ", sd(X, na.rm=TRUE)))  
 print(paste("vc: ", sd(X, na.rm=TRUE)/mean(X, na.rm=TRUE)))  
 }  
 }  
}  
  
# colDate<-1  
# if (dataset=="platjaDaro")  
# {dd[,colDate]<-as.Date(dd[, colDate], format="%d/%m/%y %h:%m:%s")  
# actives<-c(3:44)  
# }

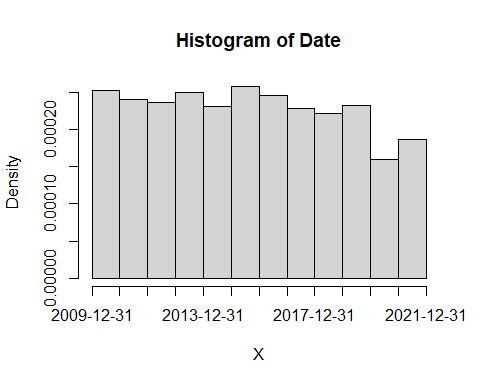
Basic descriptive analysis for numerical variables

(decide the maximum number of colors you can need in a graph based on your metadata file)

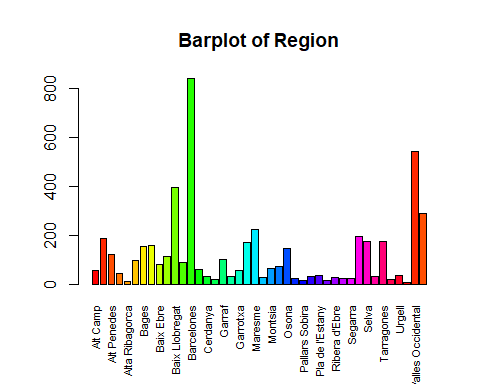
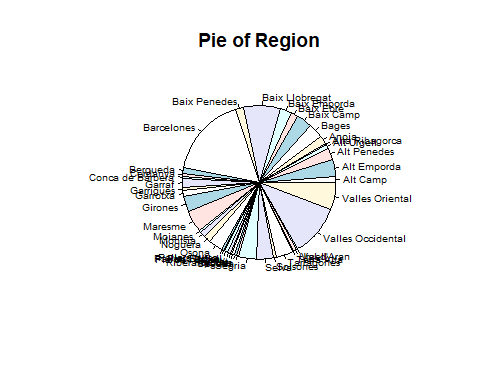
## [1] "variable 1 : Zone"



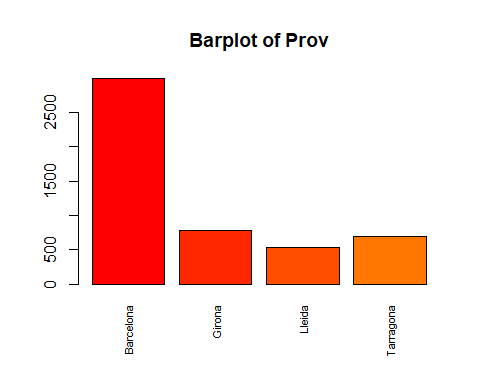
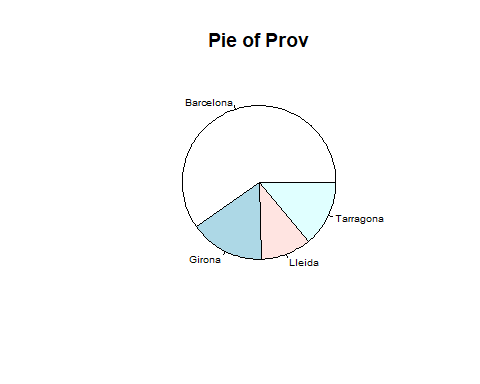
## [1] "Number of modalities: 2"  
## [1] "Frequency table"  
##   
## Road Urban   
## 2286 2714   
## [1] "Relative frequency table (proportions)"  
##   
## Road Urban   
## 0.4572 0.5428   
## [1] "Frequency table sorted"  
##   
## Urban Road   
## 2714 2286   
## [1] "Relative frequency table (proportions) sorted"  
##   
## Urban Road   
## 0.5428 0.4572   
## [1] "variable 2 : Date"  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## "2010-01-01" "2012-10-17" "2015-08-20" "2015-09-18" "2018-07-12" "2021-12-30"   
## [1] 1230.381



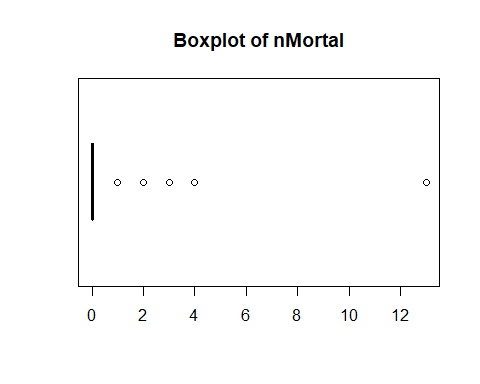
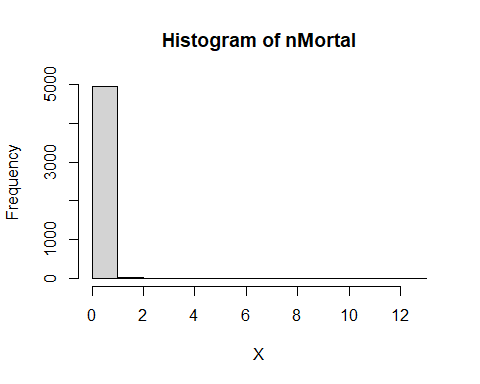
## [1] "variable 3 : Region"



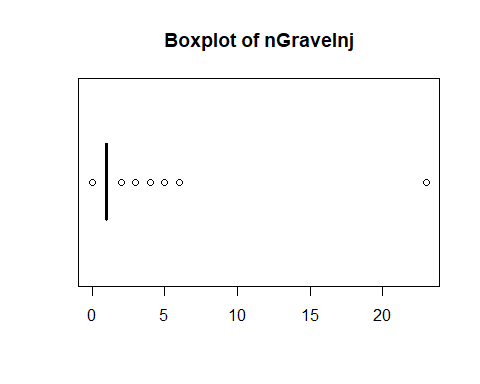
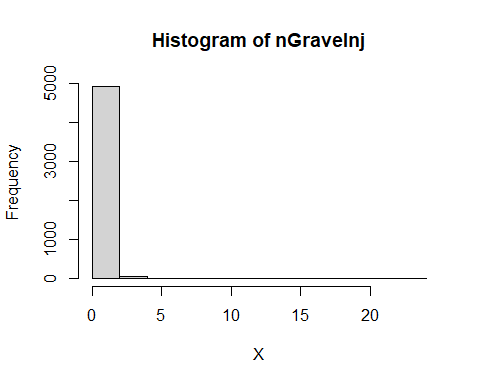
## [1] "Number of modalities: 42"  
## [1] "Frequency table"  
##   
## Alt Camp Alt Emporda Alt Penedes Alt Urgell   
## 56 187 122 44   
## Alta Ribagorca Anoia Bages Baix Camp   
## 11 96 155 159   
## Baix Ebre Baix Emporda Baix Llobregat Baix Penedes   
## 78 113 396 88   
## Barcelones Bergueda Cerdanya Conca de Barbera   
## 840 59 31 20   
## Garraf Garrigues Garrotxa Girones   
## 102 30 57 168   
## Maresme Moianes Montsia Noguera   
## 222 25 63 72   
## Osona Pallars Jussa Pallars Sobira Pla d'Urgell   
## 144 24 15 31   
## Pla de l'Estany Priorat Ribera d'Ebre Ripolles   
## 36 13 28 22   
## Segarra Segria Selva Solsones   
## 23 193 176 31   
## Tarragones Terra Alta Urgell Val d'Aran   
## 175 18 36 8   
## Valles Occidental Valles Oriental   
## 543 290   
## [1] "Relative frequency table (proportions)"  
##   
## Alt Camp Alt Emporda Alt Penedes Alt Urgell   
## 0.0112 0.0374 0.0244 0.0088   
## Alta Ribagorca Anoia Bages Baix Camp   
## 0.0022 0.0192 0.0310 0.0318   
## Baix Ebre Baix Emporda Baix Llobregat Baix Penedes   
## 0.0156 0.0226 0.0792 0.0176   
## Barcelones Bergueda Cerdanya Conca de Barbera   
## 0.1680 0.0118 0.0062 0.0040   
## Garraf Garrigues Garrotxa Girones   
## 0.0204 0.0060 0.0114 0.0336   
## Maresme Moianes Montsia Noguera   
## 0.0444 0.0050 0.0126 0.0144   
## Osona Pallars Jussa Pallars Sobira Pla d'Urgell   
## 0.0288 0.0048 0.0030 0.0062   
## Pla de l'Estany Priorat Ribera d'Ebre Ripolles   
## 0.0072 0.0026 0.0056 0.0044   
## Segarra Segria Selva Solsones   
## 0.0046 0.0386 0.0352 0.0062   
## Tarragones Terra Alta Urgell Val d'Aran   
## 0.0350 0.0036 0.0072 0.0016   
## Valles Occidental Valles Oriental   
## 0.1086 0.0580   
## [1] "Frequency table sorted"  
##   
## Barcelones Valles Occidental Baix Llobregat Valles Oriental   
## 840 543 396 290   
## Maresme Segria Alt Emporda Selva   
## 222 193 187 176   
## Tarragones Girones Baix Camp Bages   
## 175 168 159 155   
## Osona Alt Penedes Baix Emporda Garraf   
## 144 122 113 102   
## Anoia Baix Penedes Baix Ebre Noguera   
## 96 88 78 72   
## Montsia Bergueda Garrotxa Alt Camp   
## 63 59 57 56   
## Alt Urgell Pla de l'Estany Urgell Cerdanya   
## 44 36 36 31   
## Pla d'Urgell Solsones Garrigues Ribera d'Ebre   
## 31 31 30 28   
## Moianes Pallars Jussa Segarra Ripolles   
## 25 24 23 22   
## Conca de Barbera Terra Alta Pallars Sobira Priorat   
## 20 18 15 13   
## Alta Ribagorca Val d'Aran   
## 11 8   
## [1] "Relative frequency table (proportions) sorted"  
##   
## Barcelones Valles Occidental Baix Llobregat Valles Oriental   
## 0.1680 0.1086 0.0792 0.0580   
## Maresme Segria Alt Emporda Selva   
## 0.0444 0.0386 0.0374 0.0352   
## Tarragones Girones Baix Camp Bages   
## 0.0350 0.0336 0.0318 0.0310   
## Osona Alt Penedes Baix Emporda Garraf   
## 0.0288 0.0244 0.0226 0.0204   
## Anoia Baix Penedes Baix Ebre Noguera   
## 0.0192 0.0176 0.0156 0.0144   
## Montsia Bergueda Garrotxa Alt Camp   
## 0.0126 0.0118 0.0114 0.0112   
## Alt Urgell Pla de l'Estany Urgell Cerdanya   
## 0.0088 0.0072 0.0072 0.0062   
## Pla d'Urgell Solsones Garrigues Ribera d'Ebre   
## 0.0062 0.0062 0.0060 0.0056   
## Moianes Pallars Jussa Segarra Ripolles   
## 0.0050 0.0048 0.0046 0.0044   
## Conca de Barbera Terra Alta Pallars Sobira Priorat   
## 0.0040 0.0036 0.0030 0.0026   
## Alta Ribagorca Val d'Aran   
## 0.0022 0.0016   
## [1] "variable 4 : Prov"



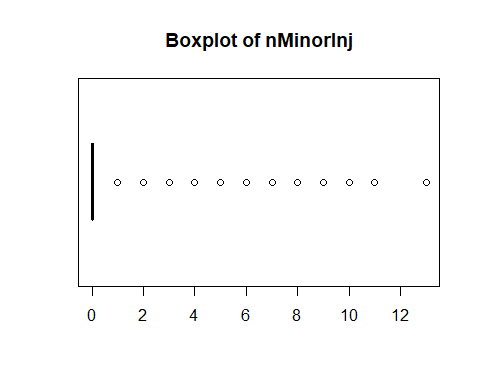
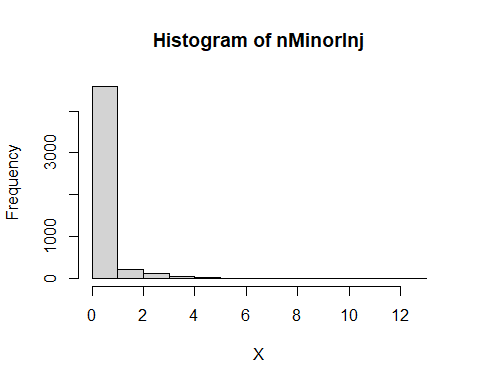
## [1] "Number of modalities: 4"  
## [1] "Frequency table"  
##   
## Barcelona Girona Lleida Tarragona   
## 2993 783 526 698   
## [1] "Relative frequency table (proportions)"  
##   
## Barcelona Girona Lleida Tarragona   
## 0.5986 0.1566 0.1052 0.1396   
## [1] "Frequency table sorted"  
##   
## Barcelona Girona Tarragona Lleida   
## 2993 783 698 526   
## [1] "Relative frequency table (proportions) sorted"  
##   
## Barcelona Girona Tarragona Lleida   
## 0.5986 0.1566 0.1396 0.1052   
## [1] "variable 5 : nMortal"



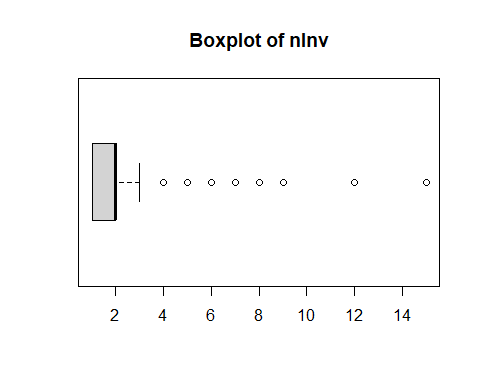
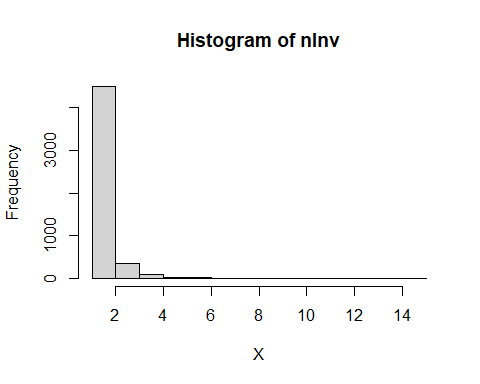
## [1] "Extended Summary Statistics"  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 0.00 0.00 0.14 0.00 13.00   
## [1] "sd: 0.428995113447079"  
## [1] "vc: 3.06425081033628"  
## [1] "variable 6 : nGraveInj"



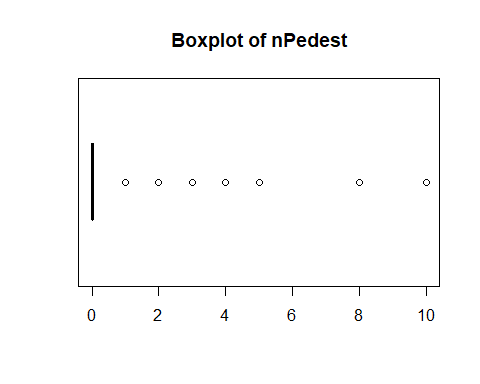
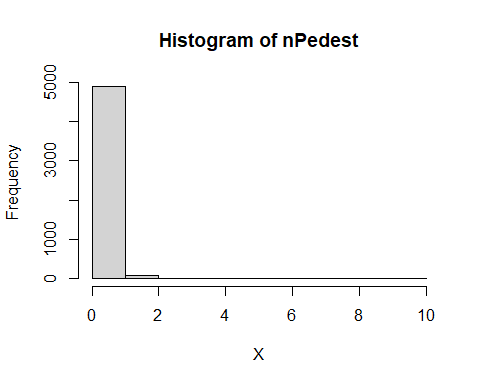
## [1] "Extended Summary Statistics"  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 1.000 1.000 1.003 1.000 23.000   
## [1] "sd: 0.590472837338287"  
## [1] "vc: 0.588472032428031"  
## [1] "variable 7 : nMinorInj"



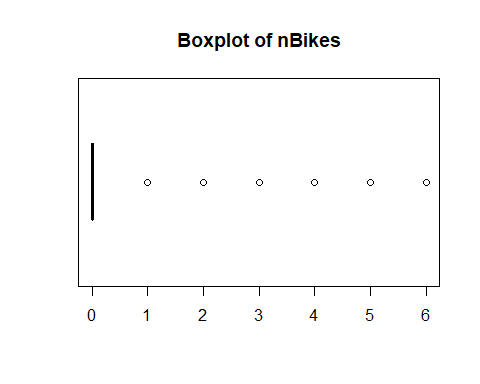
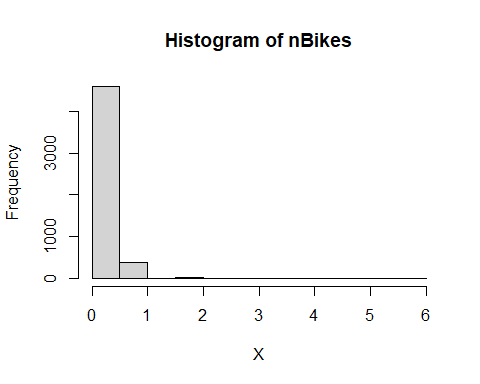
## [1] "Extended Summary Statistics"  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0000 0.0000 0.0000 0.3924 0.0000 13.0000   
## [1] "sd: 0.899212965204216"  
## [1] "vc: 2.29157228645315"  
## [1] "variable 8 : nInv"



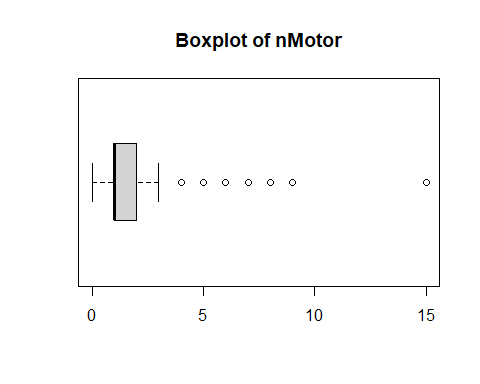
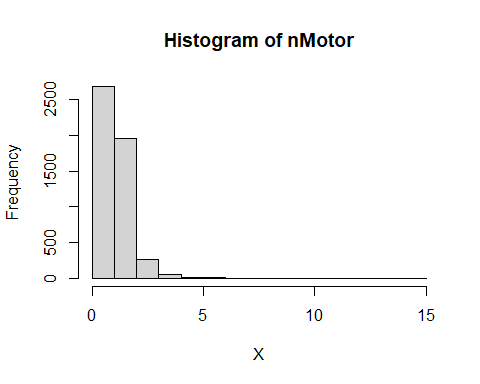
## [1] "Extended Summary Statistics"  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 1.000 2.000 1.881 2.000 15.000   
## [1] "sd: 0.781579051128312"  
## [1] "vc: 0.415600899249342"  
## [1] "variable 9 : nPedest"



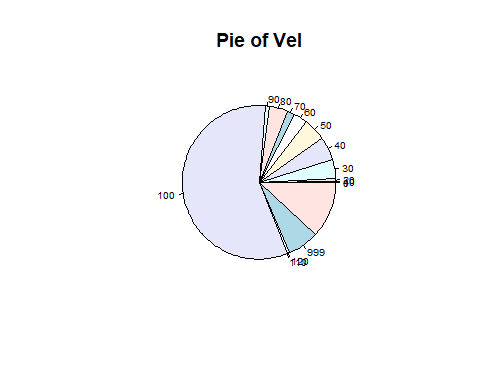
## [1] "Extended Summary Statistics"  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0000 0.0000 0.0000 0.2456 0.0000 10.0000   
## [1] "sd: 0.522814792313121"  
## [1] "vc: 2.12872472440196"  
## [1] "variable 10 : nBikes"



## [1] "Extended Summary Statistics"  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 0.00 0.00 0.09 0.00 6.00   
## [1] "sd: 0.335145437821998"  
## [1] "vc: 3.7238381980222"  
## [1] "variable 11 : nMotor"

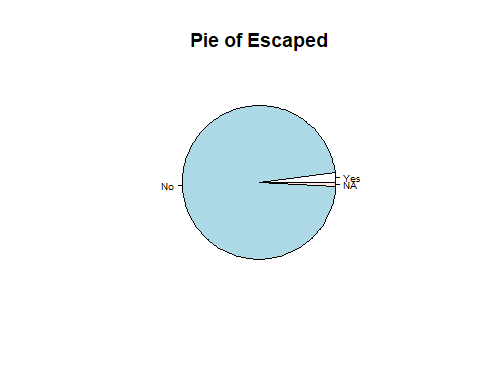
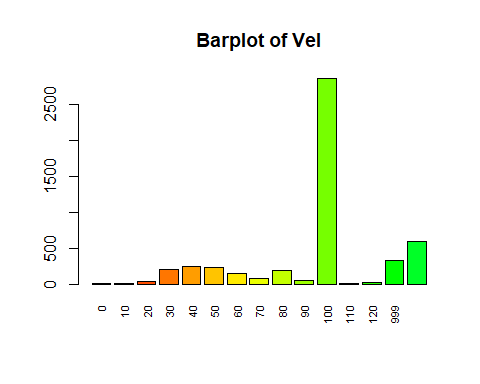


## [1] "Extended Summary Statistics"  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 1.000 1.000 1.521 2.000 15.000   
## [1] "sd: 0.823211744033759"  
## [1] "vc: 0.541088302901117"  
## [1] "variable 12 : Vel"



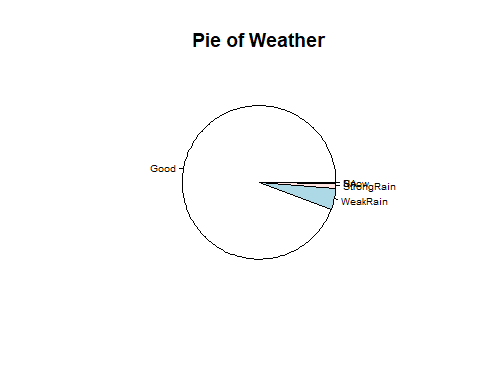
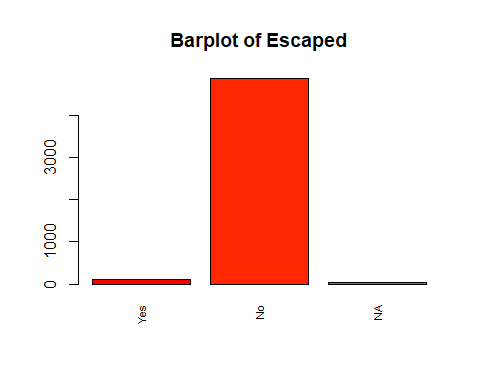
## [1] "Number of modalities: 15"  
## [1] "Frequency table"  
##   
## 0 10 20 30 40 50 60 70 80 90 100 110 120 999 <NA>   
## 2 7 30 204 246 234 150 76 191 49 2856 2 27 330 596   
## [1] "Relative frequency table (proportions)"  
##   
## 0 10 20 30 40 50 60 70 80 90 100   
## 0.0004 0.0014 0.0060 0.0408 0.0492 0.0468 0.0300 0.0152 0.0382 0.0098 0.5712   
## 110 120 999 <NA>   
## 0.0004 0.0054 0.0660 0.1192   
## [1] "Frequency table sorted"  
##   
## 100 <NA> 999 40 50 30 80 60 70 90 20 120 10 0 110   
## 2856 596 330 246 234 204 191 150 76 49 30 27 7 2 2   
## [1] "Relative frequency table (proportions) sorted"  
##   
## 100 <NA> 999 40 50 30 80 60 70 90 20   
## 0.5712 0.1192 0.0660 0.0492 0.0468 0.0408 0.0382 0.0300 0.0152 0.0098 0.0060   
## 120 10 0 110   
## 0.0054 0.0014 0.0004 0.0004   
## [1] "variable 13 : Escaped"

## Warning in is.numeric(X) || class(X) == "Date": 'length(x) = 2 > 1' in coercion  
## to 'logical(1)'



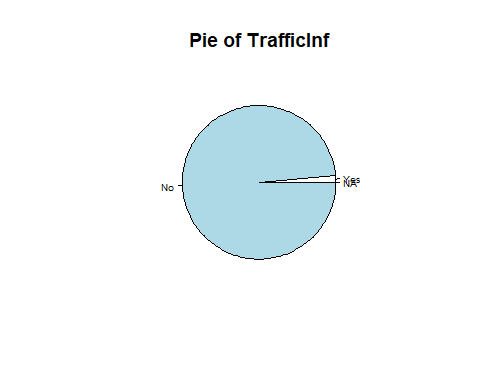
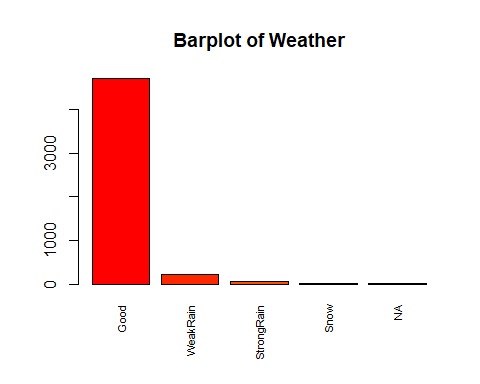
## [1] "Number of modalities: 3"  
## [1] "Frequency table"  
##   
## Yes No NA   
## 98 4861 41   
## [1] "Relative frequency table (proportions)"  
##   
## Yes No NA   
## 0.0196 0.9722 0.0082   
## [1] "Frequency table sorted"  
##   
## No Yes NA   
## 4861 98 41   
## [1] "Relative frequency table (proportions) sorted"  
##   
## No Yes NA   
## 0.9722 0.0196 0.0082   
## [1] "variable 14 : Weather"

## Warning in is.numeric(X) || class(X) == "Date": 'length(x) = 2 > 1' in coercion  
## to 'logical(1)'



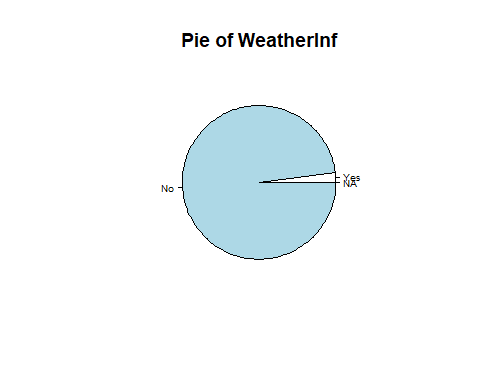
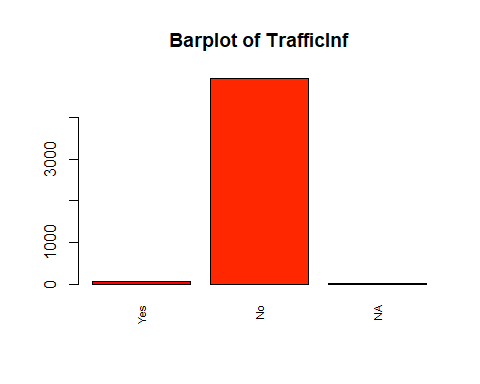
## [1] "Number of modalities: 5"  
## [1] "Frequency table"  
##   
## Good WeakRain StrongRain Snow NA   
## 4717 221 55 6 1   
## [1] "Relative frequency table (proportions)"  
##   
## Good WeakRain StrongRain Snow NA   
## 0.9434 0.0442 0.0110 0.0012 0.0002   
## [1] "Frequency table sorted"  
##   
## Good WeakRain StrongRain Snow NA   
## 4717 221 55 6 1   
## [1] "Relative frequency table (proportions) sorted"  
##   
## Good WeakRain StrongRain Snow NA   
## 0.9434 0.0442 0.0110 0.0012 0.0002   
## [1] "variable 15 : TrafficInf"

## Warning in is.numeric(X) || class(X) == "Date": 'length(x) = 2 > 1' in coercion  
## to 'logical(1)'



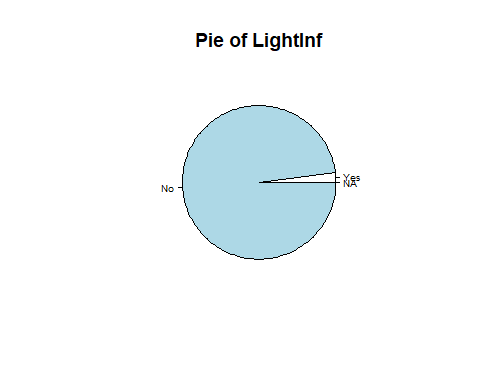
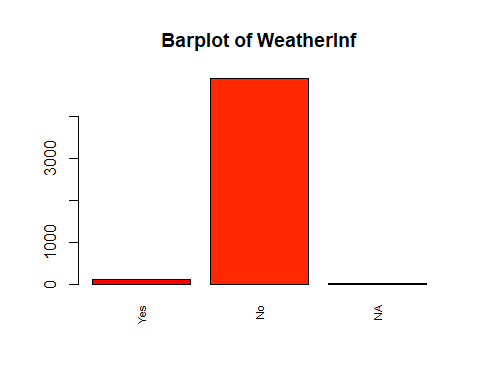
## [1] "Number of modalities: 3"  
## [1] "Frequency table"  
##   
## Yes No NA   
## 69 4930 1   
## [1] "Relative frequency table (proportions)"  
##   
## Yes No NA   
## 0.0138 0.9860 0.0002   
## [1] "Frequency table sorted"  
##   
## No Yes NA   
## 4930 69 1   
## [1] "Relative frequency table (proportions) sorted"  
##   
## No Yes NA   
## 0.9860 0.0138 0.0002   
## [1] "variable 16 : WeatherInf"

## Warning in is.numeric(X) || class(X) == "Date": 'length(x) = 2 > 1' in coercion  
## to 'logical(1)'



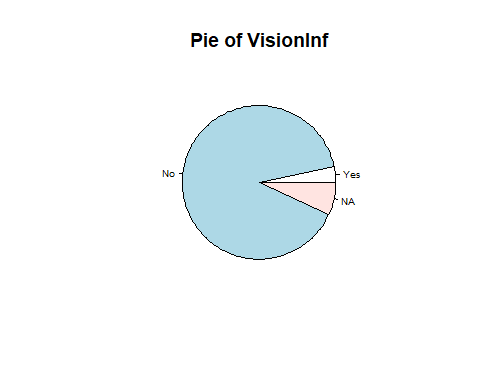
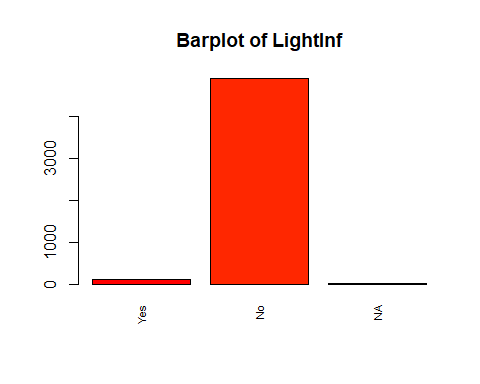
## [1] "Number of modalities: 3"  
## [1] "Frequency table"  
##   
## Yes No NA   
## 106 4893 1   
## [1] "Relative frequency table (proportions)"  
##   
## Yes No NA   
## 0.0212 0.9786 0.0002   
## [1] "Frequency table sorted"  
##   
## No Yes NA   
## 4893 106 1   
## [1] "Relative frequency table (proportions) sorted"  
##   
## No Yes NA   
## 0.9786 0.0212 0.0002   
## [1] "variable 17 : LightInf"

## Warning in is.numeric(X) || class(X) == "Date": 'length(x) = 2 > 1' in coercion  
## to 'logical(1)'



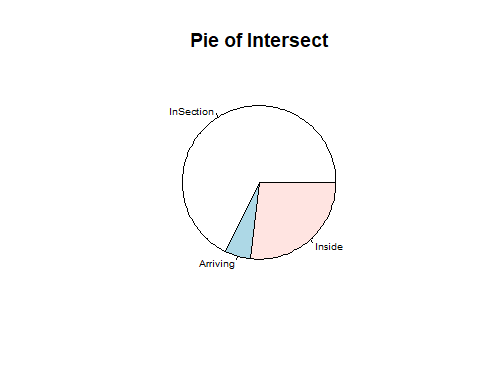
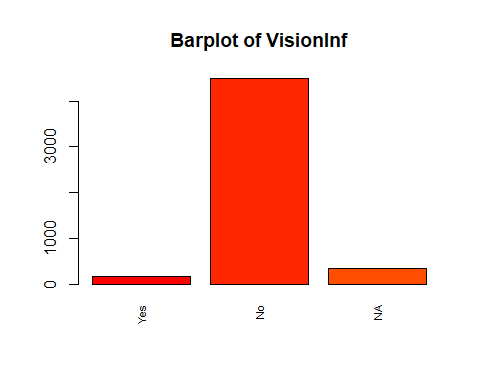
## [1] "Number of modalities: 3"  
## [1] "Frequency table"  
##   
## Yes No NA   
## 103 4896 1   
## [1] "Relative frequency table (proportions)"  
##   
## Yes No NA   
## 0.0206 0.9792 0.0002   
## [1] "Frequency table sorted"  
##   
## No Yes NA   
## 4896 103 1   
## [1] "Relative frequency table (proportions) sorted"  
##   
## No Yes NA   
## 0.9792 0.0206 0.0002   
## [1] "variable 18 : VisionInf"

## Warning in is.numeric(X) || class(X) == "Date": 'length(x) = 2 > 1' in coercion  
## to 'logical(1)'



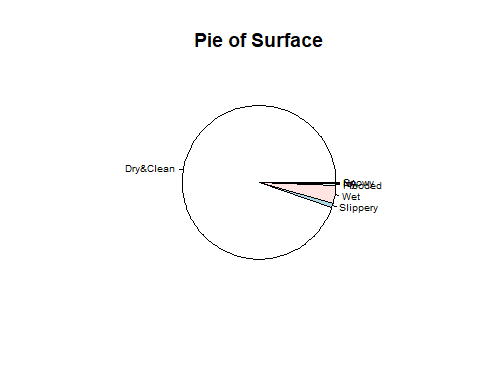
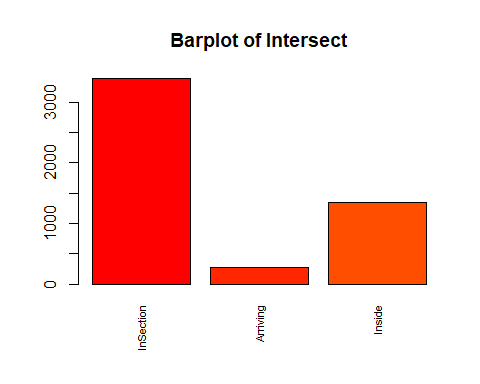
## [1] "Number of modalities: 3"  
## [1] "Frequency table"  
##   
## Yes No NA   
## 163 4494 343   
## [1] "Relative frequency table (proportions)"  
##   
## Yes No NA   
## 0.0326 0.8988 0.0686   
## [1] "Frequency table sorted"  
##   
## No NA Yes   
## 4494 343 163   
## [1] "Relative frequency table (proportions) sorted"  
##   
## No NA Yes   
## 0.8988 0.0686 0.0326   
## [1] "variable 19 : Intersect"

## Warning in is.numeric(X) || class(X) == "Date": 'length(x) = 2 > 1' in coercion  
## to 'logical(1)'



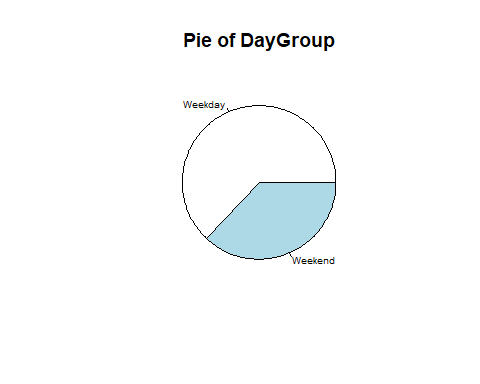
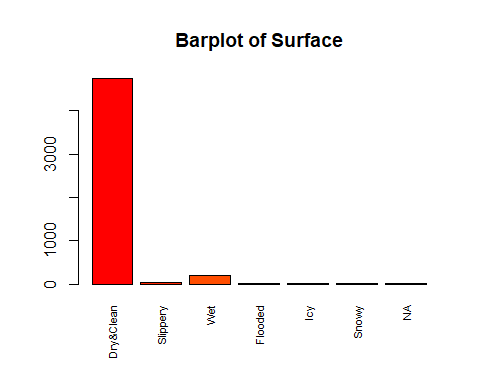
## [1] "Number of modalities: 3"  
## [1] "Frequency table"  
##   
## InSection Arriving Inside   
## 3384 272 1344   
## [1] "Relative frequency table (proportions)"  
##   
## InSection Arriving Inside   
## 0.6768 0.0544 0.2688   
## [1] "Frequency table sorted"  
##   
## InSection Inside Arriving   
## 3384 1344 272   
## [1] "Relative frequency table (proportions) sorted"  
##   
## InSection Inside Arriving   
## 0.6768 0.2688 0.0544   
## [1] "variable 20 : Surface"

## Warning in is.numeric(X) || class(X) == "Date": 'length(x) = 2 > 1' in coercion  
## to 'logical(1)'



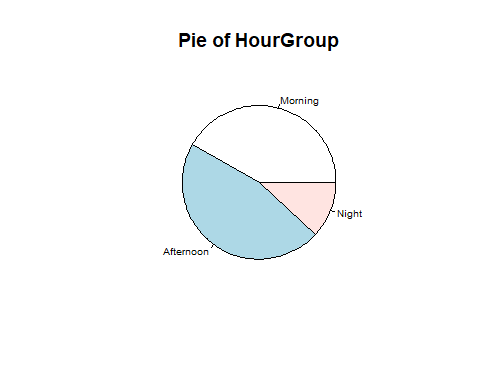
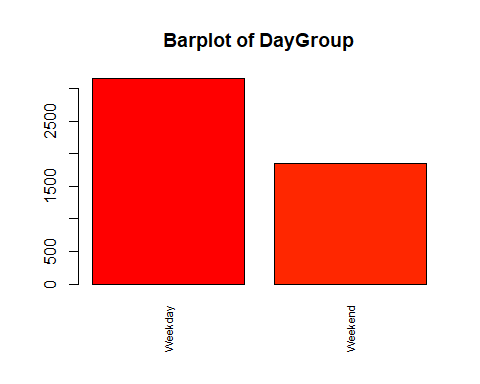
## [1] "Number of modalities: 7"  
## [1] "Frequency table"  
##   
## Dry&Clean Slippery Wet Flooded Icy Snowy NA   
## 4738 42 189 22 4 4 1   
## [1] "Relative frequency table (proportions)"  
##   
## Dry&Clean Slippery Wet Flooded Icy Snowy NA   
## 0.9476 0.0084 0.0378 0.0044 0.0008 0.0008 0.0002   
## [1] "Frequency table sorted"  
##   
## Dry&Clean Wet Slippery Flooded Icy Snowy NA   
## 4738 189 42 22 4 4 1   
## [1] "Relative frequency table (proportions) sorted"  
##   
## Dry&Clean Wet Slippery Flooded Icy Snowy NA   
## 0.9476 0.0378 0.0084 0.0044 0.0008 0.0008 0.0002   
## [1] "variable 21 : DayGroup"

## Warning in is.numeric(X) || class(X) == "Date": 'length(x) = 2 > 1' in coercion  
## to 'logical(1)'



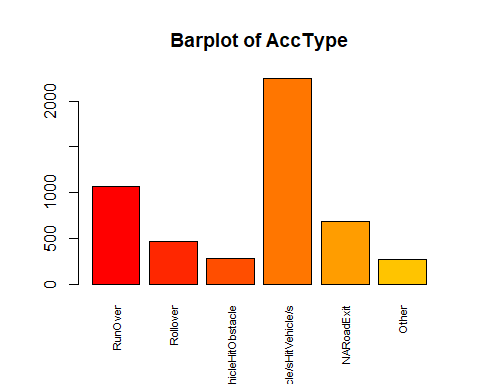
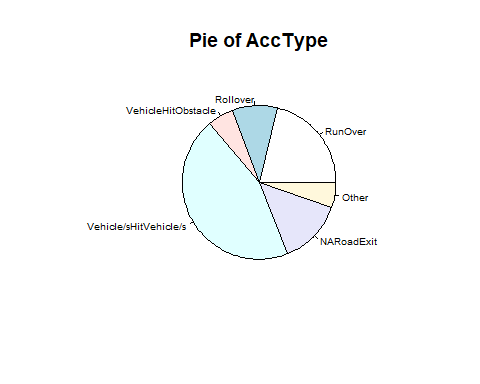
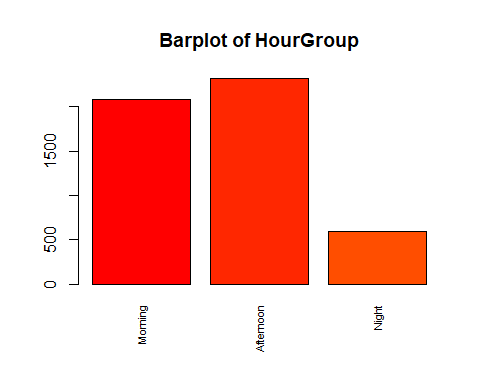
## [1] "Number of modalities: 2"  
## [1] "Frequency table"  
##   
## Weekday Weekend   
## 3152 1848   
## [1] "Relative frequency table (proportions)"  
##   
## Weekday Weekend   
## 0.6304 0.3696   
## [1] "Frequency table sorted"  
##   
## Weekday Weekend   
## 3152 1848   
## [1] "Relative frequency table (proportions) sorted"  
##   
## Weekday Weekend   
## 0.6304 0.3696   
## [1] "variable 22 : HourGroup"

## Warning in is.numeric(X) || class(X) == "Date": 'length(x) = 2 > 1' in coercion  
## to 'logical(1)'



## [1] "Number of modalities: 3"  
## [1] "Frequency table"  
##   
## Morning Afternoon Night   
## 2086 2317 597   
## [1] "Relative frequency table (proportions)"  
##   
## Morning Afternoon Night   
## 0.4172 0.4634 0.1194   
## [1] "Frequency table sorted"  
##   
## Afternoon Morning Night   
## 2317 2086 597   
## [1] "Relative frequency table (proportions) sorted"  
##   
## Afternoon Morning Night   
## 0.4634 0.4172 0.1194   
## [1] "variable 23 : AccType"

## Warning in is.numeric(X) || class(X) == "Date": 'length(x) = 2 > 1' in coercion  
## to 'logical(1)'



## [1] "Number of modalities: 6"  
## [1] "Frequency table"  
##   
## RunOver Rollover VehicleHitObstacle   
## 1062 466 280   
## Vehicle/sHitVehicle/s NARoadExit Other   
## 2246 681 265   
## [1] "Relative frequency table (proportions)"  
##   
## RunOver Rollover VehicleHitObstacle   
## 0.2124 0.0932 0.0560   
## Vehicle/sHitVehicle/s NARoadExit Other   
## 0.4492 0.1362 0.0530   
## [1] "Frequency table sorted"  
##   
## Vehicle/sHitVehicle/s RunOver NARoadExit   
## 2246 1062 681   
## Rollover VehicleHitObstacle Other   
## 466 280 265   
## [1] "Relative frequency table (proportions) sorted"  
##   
## Vehicle/sHitVehicle/s RunOver NARoadExit   
## 0.4492 0.2124 0.1362   
## Rollover VehicleHitObstacle Other   
## 0.0932 0.0560 0.0530