Univariate descriptive statistics of our data

Group 5 MD

2022-09-29

In this document, we provide the basic statistics and plots of our accidents database.

Dimensions of our dataset:

dim(dd)

## [1] 5000 23

n<-dim(dd)[1]  
K<-dim(dd)[2]  
  
K #Number of variables (columns)

## [1] 23

n #Number of instances (rows)

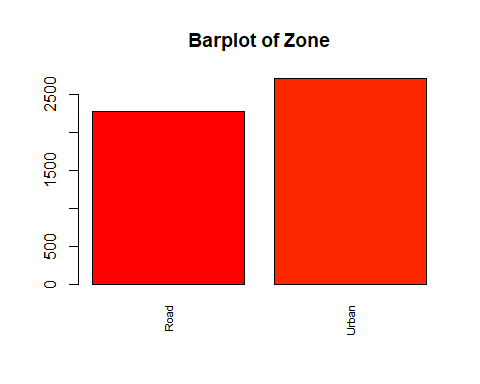
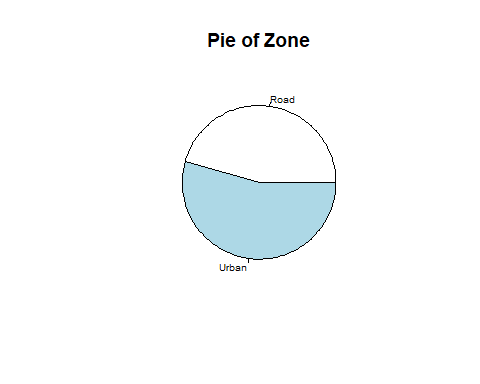
## [1] 5000

Check of 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"

Basic descriptive analysis of our variables:

Variable 1 : Zone 

Number of modalities: 2

Frequency table

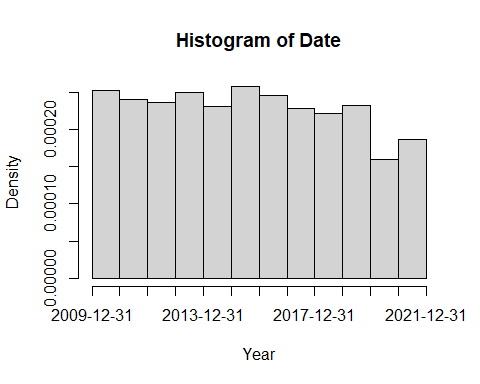
| Var1 | Freq |
| --- | --- |
| Road | 2286 |
| Urban | 2714 |

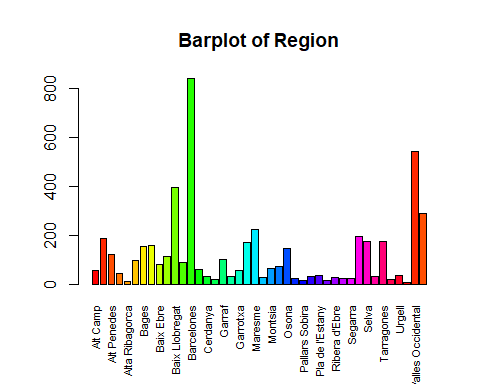
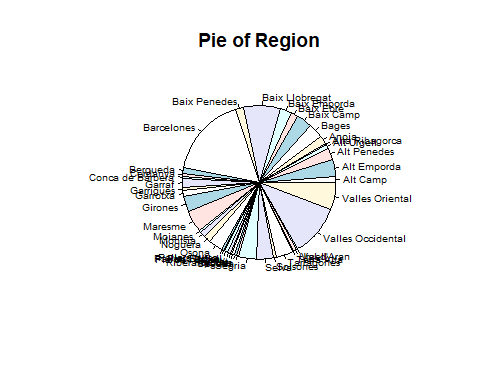
[1] “Relative frequency table (proportions)”

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

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 

Variable 3 : Region 

Number of modalities: 42

Frequency table

| Var1 | Freq |
| --- | --- |
| Alt Camp | 56 |
| Alt Emporda | 187 |
| Alt Penedes | 122 |
| Alt Urgell | 44 |
| Alta Ribagorca | 11 |
| Anoia | 96 |
| Bages | 155 |
| Baix Camp | 159 |
| Baix Ebre | 78 |
| Baix Emporda | 113 |
| Baix Llobregat | 396 |
| Baix Penedes | 88 |
| Barcelones | 840 |
| Bergueda | 59 |
| Cerdanya | 31 |
| Conca de Barbera | 20 |
| Garraf | 102 |
| Garrigues | 30 |
| Garrotxa | 57 |
| Girones | 168 |
| Maresme | 222 |
| Moianes | 25 |
| Montsia | 63 |
| Noguera | 72 |
| Osona | 144 |
| Pallars Jussa | 24 |
| Pallars Sobira | 15 |
| Pla d’Urgell | 31 |
| Pla de l’Estany | 36 |
| Priorat | 13 |
| Ribera d’Ebre | 28 |
| Ripolles | 22 |
| Segarra | 23 |
| Segria | 193 |
| Selva | 176 |
| Solsones | 31 |
| Tarragones | 175 |
| Terra Alta | 18 |
| Urgell | 36 |
| Val d’Aran | 8 |
| Valles Occidental | 543 |
| Valles Oriental | 290 |

[1] “Relative frequency table (proportions)”

Alt Camp Alt Emporda Alt Penedes Alt Urgell   
 0.0112 0.0374 0.0244 0.0088

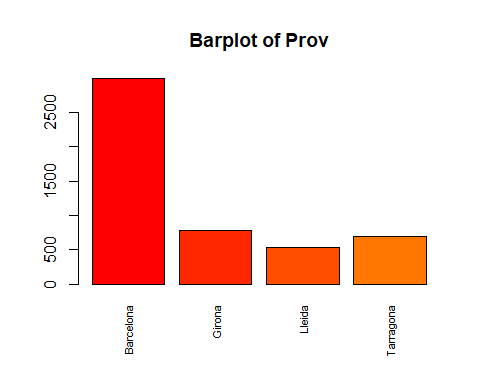
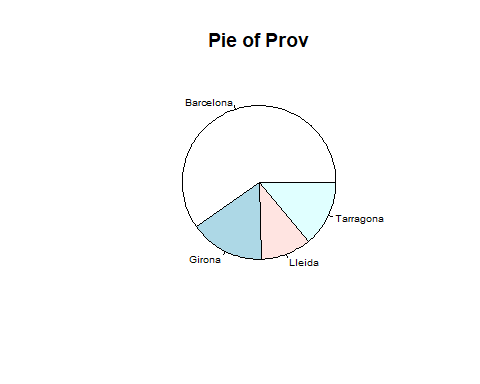
Alta Ribagorca Anoia Bages Baix Camp 0.0022 0.0192 0.0310 0.0318 Baix Ebre Baix Emporda Baix Llobregat Baix Penedes 0.0156 0.0226 0.0792 0.0176 Barcelones Bergueda Cerdanya Conca de Barbera 0.1680 0.0118 0.0062 0.0040 Garraf Garrigues Garrotxa Girones 0.0204 0.0060 0.0114 0.0336 Maresme Moianes Montsia 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

Variable 4 : Prov 

Number of modalities: 4

Frequency table

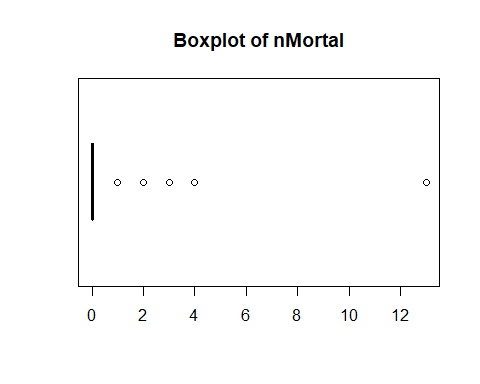
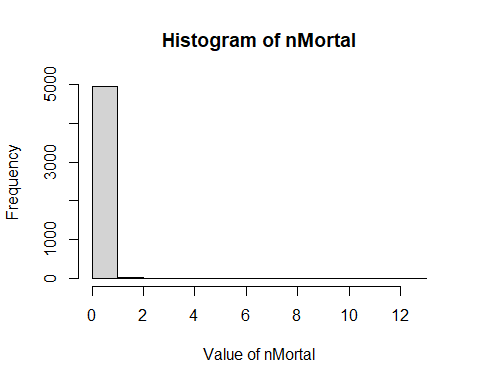
| Var1 | Freq |
| --- | --- |
| Barcelona | 2993 |
| Girona | 783 |
| Lleida | 526 |
| Tarragona | 698 |

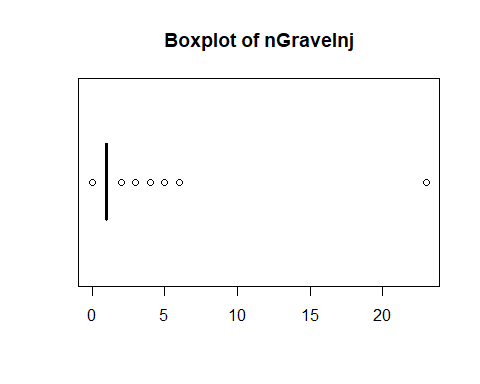
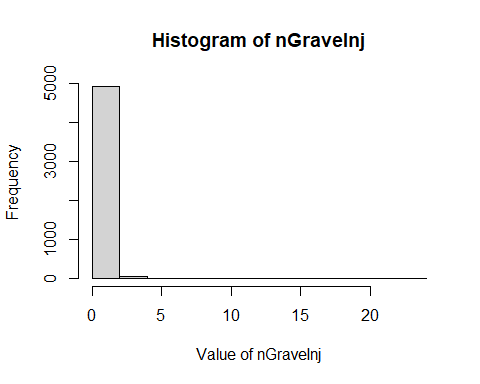
[1] “Relative frequency table (proportions)”

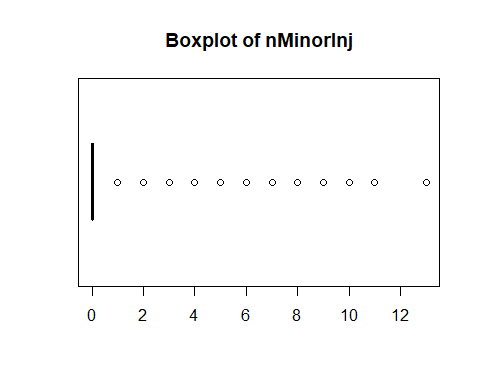
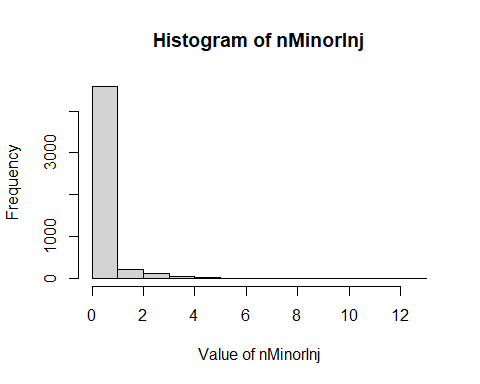
Barcelona Girona Lleida Tarragona 0.5986 0.1566 0.1052 0.1396 [1] “Frequency table sorted”

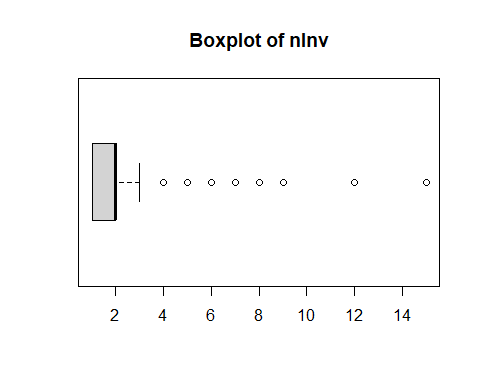
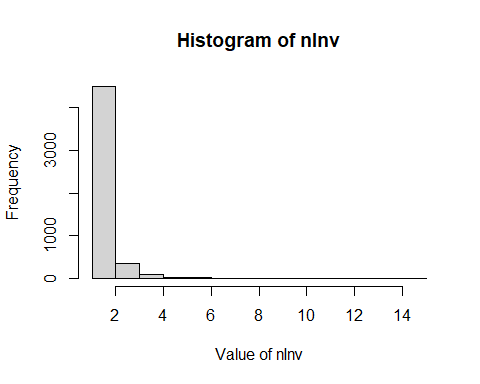
Barcelona Girona Tarragona Lleida 2993 783 698 526 [1] “Relative frequency table (proportions) sorted”

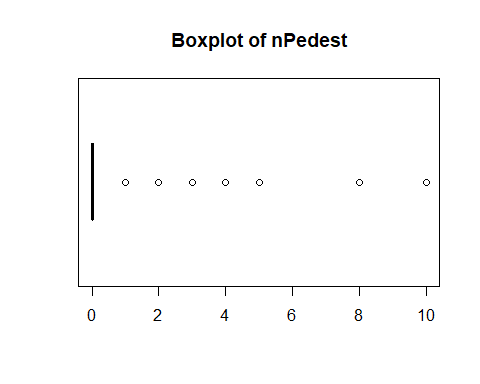
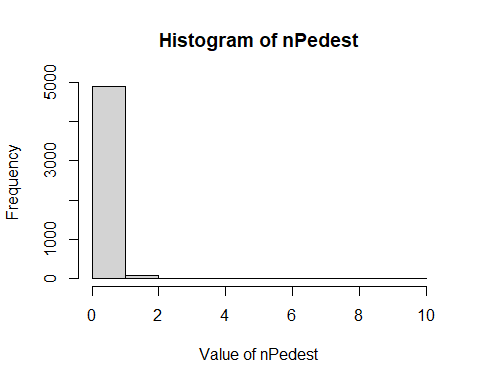
Barcelona Girona Tarragona Lleida 0.5986 0.1566 0.1396 0.1052

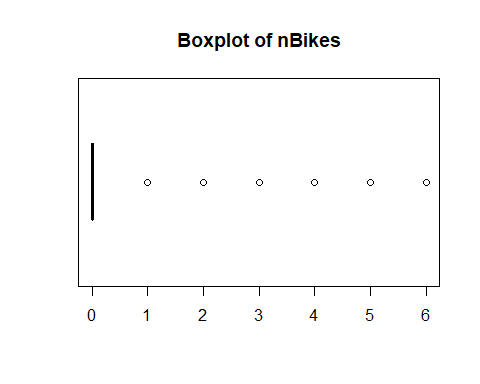
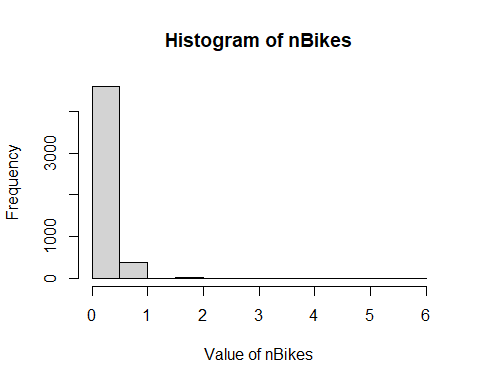
Variable 5 : nMortal [1] “Extended Summary Statistics” Min. 1st Qu. Median Mean 3rd Qu. Max. 0.00 0.00 0.00 0.14 0.00 13.00 [1] “sd: 0.428995113447079” [1] “vc: 3.06425081033628”

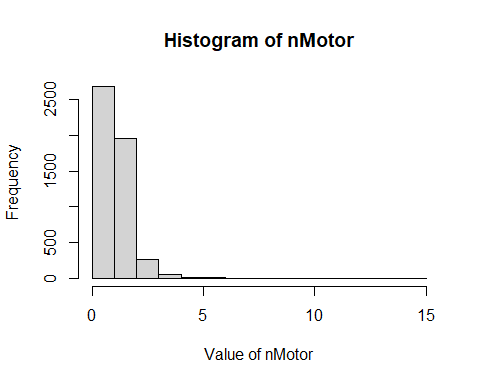
Variable 6 : nGraveInj [1] “Extended Summary Statistics” Min. 1st Qu. Median Mean 3rd Qu. Max. 0.000 1.000 1.000 1.003 1.000 23.000 [1] “sd: 0.590472837338287” [1] “vc: 0.588472032428031”

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”

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”

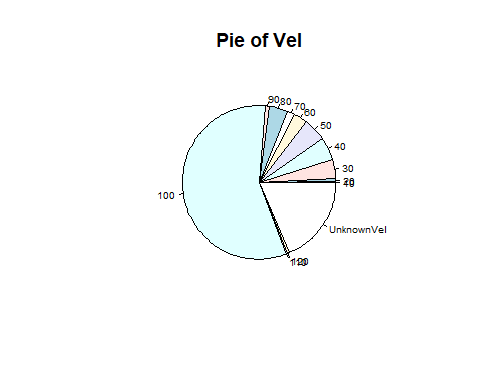
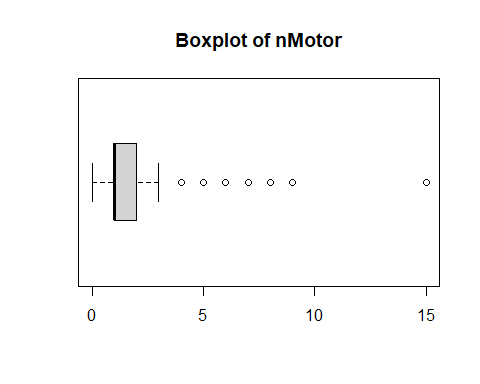
Variable 9 : nPedest [1] “Extended Summary Statistics” Min. 1st Qu. Median Mean 3rd Qu. Max. 0.0000 0.0000 0.0000 0.2456 0.0000 10.0000 [1] “sd: 0.522814792313121” [1] “vc: 2.12872472440196”

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”

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”

Variable 12 : Vel

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



Number of modalities: 13

Frequency table

| Var1 | Freq |
| --- | --- |
| 10 | 7 |
| 20 | 30 |
| 30 | 204 |
| 40 | 246 |
| 50 | 234 |
| 60 | 150 |
| 70 | 76 |
| 80 | 191 |
| 90 | 49 |
| 100 | 2856 |
| 110 | 2 |
| 120 | 27 |
| UnknownVel | 928 |

[1] “Relative frequency table (proportions)”

10 20 30 40 50 60 70   
0.0014 0.0060 0.0408 0.0492 0.0468 0.0300 0.0152   
 80 90 100 110 120 UnknownVel   
0.0382 0.0098 0.5712 0.0004 0.0054 0.1856

[1] “Frequency table sorted”

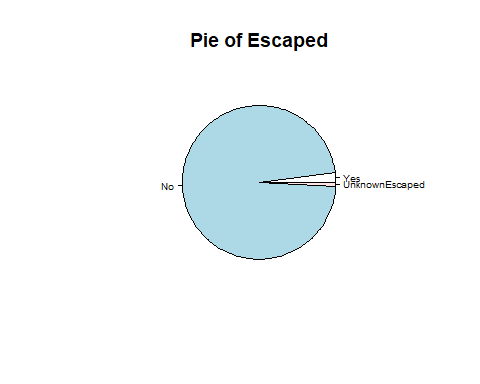
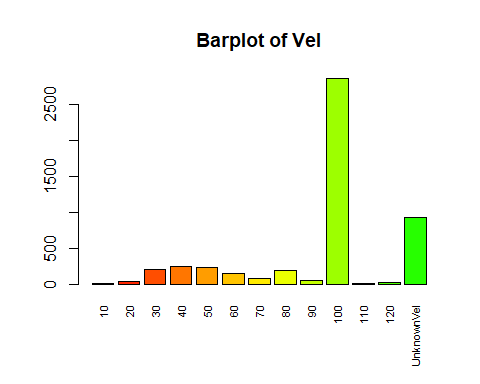
100 UnknownVel 40 50 30 80 60   
 2856 928 246 234 204 191 150   
 70 90 20 120 10 110   
 76 49 30 27 7 2

[1] “Relative frequency table (proportions) sorted”

100 UnknownVel 40 50 30 80 60   
0.5712 0.1856 0.0492 0.0468 0.0408 0.0382 0.0300   
 70 90 20 120 10 110   
0.0152 0.0098 0.0060 0.0054 0.0014 0.0004

Variable 13 : Escaped

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



Number of modalities: 3

Frequency table

| Var1 | Freq |
| --- | --- |
| Yes | 98 |
| No | 4861 |
| UnknownEscaped | 41 |

[1] “Relative frequency table (proportions)”

Yes No UnknownEscaped   
 0.0196 0.9722 0.0082

[1] “Frequency table sorted”

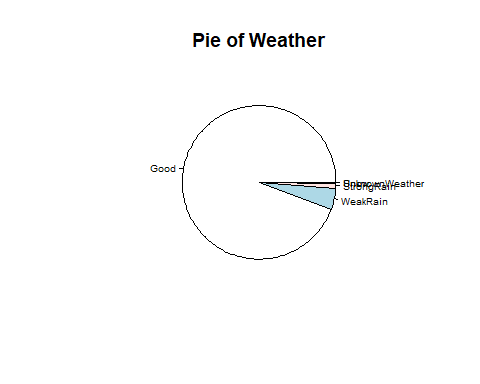
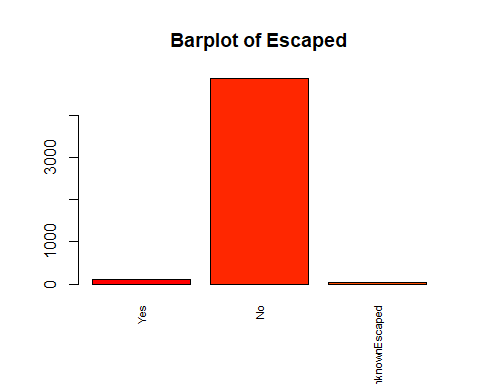
No Yes UnknownEscaped   
 4861 98 41

[1] “Relative frequency table (proportions) sorted”

No Yes UnknownEscaped   
 0.9722 0.0196 0.0082

Variable 14 : Weather

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



Number of modalities: 5

Frequency table

| Var1 | Freq |
| --- | --- |
| Good | 4717 |
| WeakRain | 221 |
| StrongRain | 55 |
| Snow | 6 |
| UnknownWeather | 1 |

[1] “Relative frequency table (proportions)”

Good WeakRain StrongRain Snow UnknownWeather   
 0.9434 0.0442 0.0110 0.0012 0.0002

[1] “Frequency table sorted”

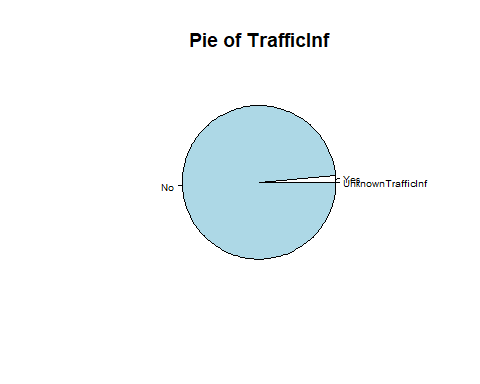
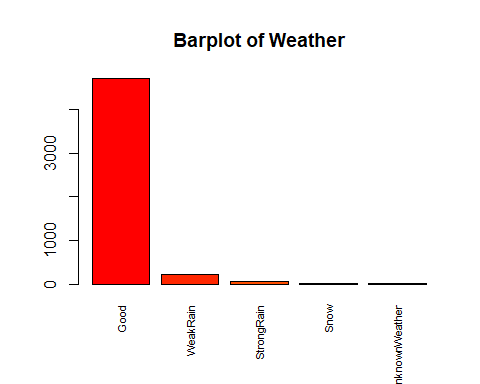
Good WeakRain StrongRain Snow UnknownWeather   
 4717 221 55 6 1

[1] “Relative frequency table (proportions) sorted”

Good WeakRain StrongRain Snow UnknownWeather   
 0.9434 0.0442 0.0110 0.0012 0.0002

Variable 15 : TrafficInf

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



Number of modalities: 3

Frequency table

| Var1 | Freq |
| --- | --- |
| Yes | 69 |
| No | 4930 |
| UnknownTrafficInf | 1 |

[1] “Relative frequency table (proportions)”

Yes No UnknownTrafficInf   
 0.0138 0.9860 0.0002

[1] “Frequency table sorted”

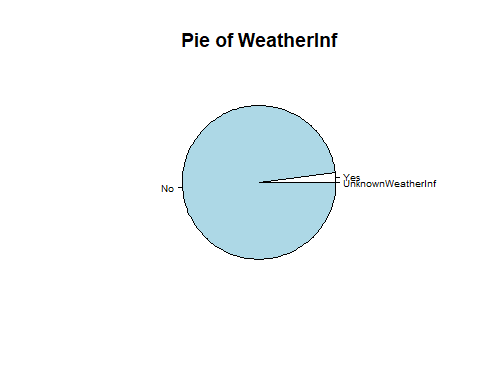
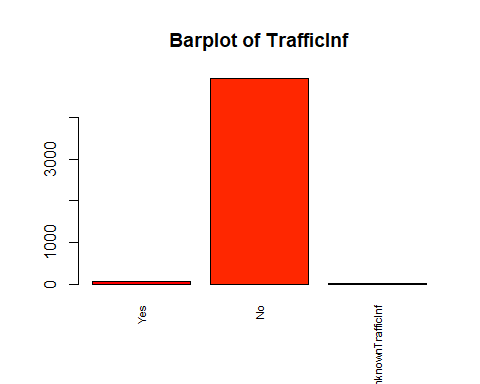
No Yes UnknownTrafficInf   
 4930 69 1

[1] “Relative frequency table (proportions) sorted”

No Yes UnknownTrafficInf   
 0.9860 0.0138 0.0002

Variable 16 : WeatherInf

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



Number of modalities: 3

Frequency table

| Var1 | Freq |
| --- | --- |
| Yes | 106 |
| No | 4893 |
| UnknownWeatherInf | 1 |

[1] “Relative frequency table (proportions)”

Yes No UnknownWeatherInf   
 0.0212 0.9786 0.0002

[1] “Frequency table sorted”

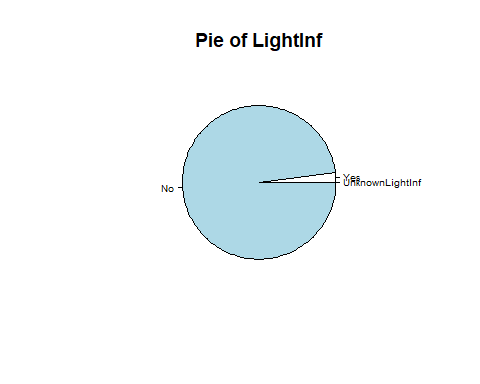
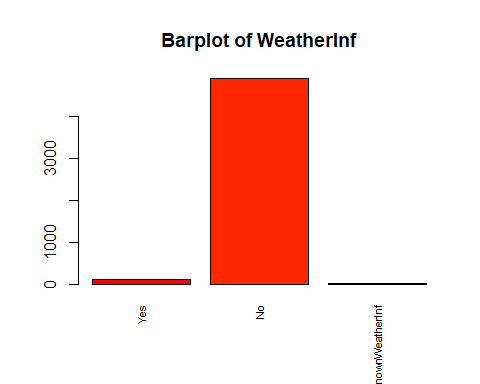
No Yes UnknownWeatherInf   
 4893 106 1

[1] “Relative frequency table (proportions) sorted”

No Yes UnknownWeatherInf   
 0.9786 0.0212 0.0002

Variable 17 : LightInf

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



Number of modalities: 3

Frequency table

| Var1 | Freq |
| --- | --- |
| Yes | 103 |
| No | 4896 |
| UnknownLightInf | 1 |

[1] “Relative frequency table (proportions)”

Yes No UnknownLightInf   
 0.0206 0.9792 0.0002

[1] “Frequency table sorted”

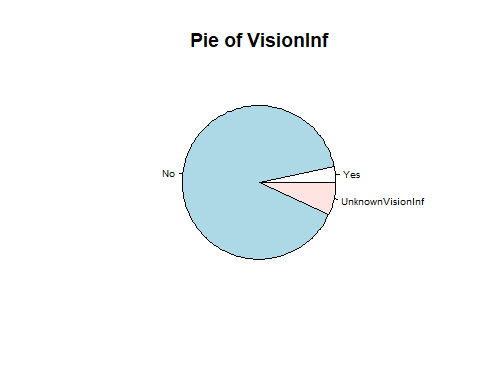
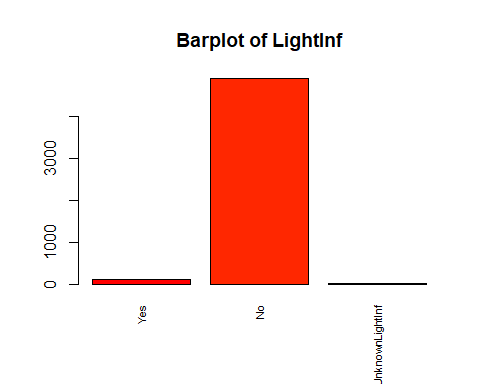
No Yes UnknownLightInf   
 4896 103 1

[1] “Relative frequency table (proportions) sorted”

No Yes UnknownLightInf   
 0.9792 0.0206 0.0002

Variable 18 : VisionInf

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



Number of modalities: 3

Frequency table

| Var1 | Freq |
| --- | --- |
| Yes | 163 |
| No | 4494 |
| UnknownVisionInf | 343 |

[1] “Relative frequency table (proportions)”

Yes No UnknownVisionInf   
 0.0326 0.8988 0.0686

[1] “Frequency table sorted”

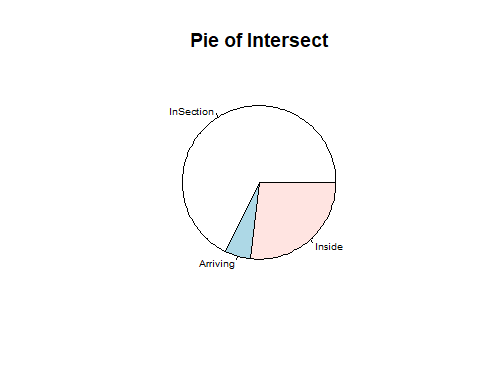
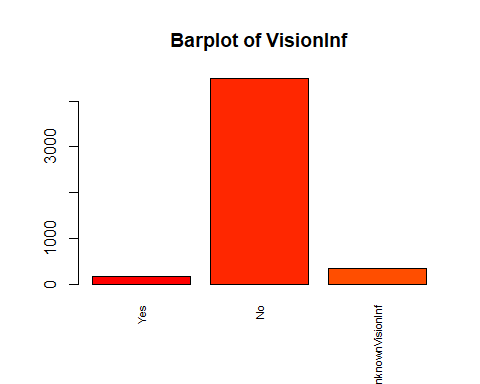
No UnknownVisionInf Yes   
 4494 343 163

[1] “Relative frequency table (proportions) sorted”

No UnknownVisionInf Yes   
 0.8988 0.0686 0.0326

Variable 19 : Intersect

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



Number of modalities: 3

Frequency table

| Var1 | Freq |
| --- | --- |
| InSection | 3384 |
| Arriving | 272 |
| Inside | 1344 |

[1] “Relative frequency table (proportions)”

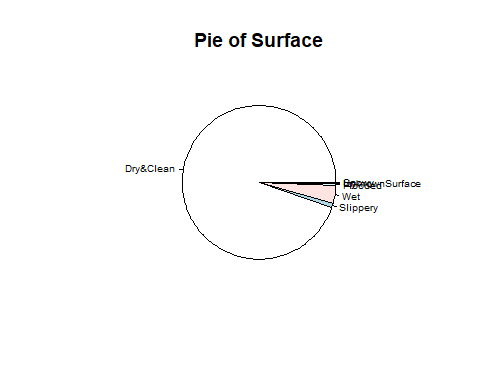
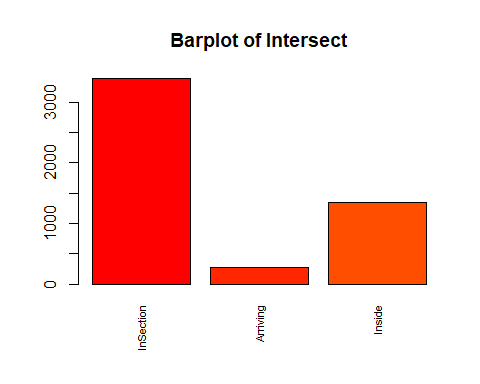
InSection Arriving Inside 0.6768 0.0544 0.2688 [1] “Frequency table sorted”

InSection Inside Arriving 3384 1344 272 [1] “Relative frequency table (proportions) sorted”

InSection Inside Arriving 0.6768 0.2688 0.0544

Variable 20 : Surface

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



Number of modalities: 7

Frequency table

| Var1 | Freq |
| --- | --- |
| Dry&Clean | 4738 |
| Slippery | 42 |
| Wet | 189 |
| Flooded | 22 |
| Icy | 4 |
| Snowy | 4 |
| UnknownSurface | 1 |

[1] “Relative frequency table (proportions)”

Dry&Clean Slippery Wet Flooded Icy   
 0.9476 0.0084 0.0378 0.0044 0.0008   
 Snowy UnknownSurface   
 0.0008 0.0002

[1] “Frequency table sorted”

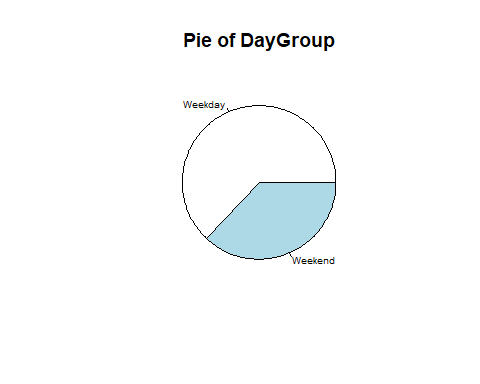
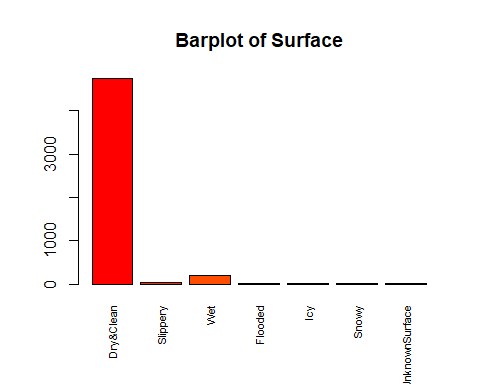
Dry&Clean Wet Slippery Flooded Icy   
 4738 189 42 22 4   
 Snowy UnknownSurface   
 4 1

[1] “Relative frequency table (proportions) sorted”

Dry&Clean Wet Slippery Flooded Icy   
 0.9476 0.0378 0.0084 0.0044 0.0008   
 Snowy UnknownSurface   
 0.0008 0.0002

Variable 21 : DayGroup

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



Number of modalities: 2

Frequency table

| Var1 | Freq |
| --- | --- |
| Weekday | 3152 |
| Weekend | 1848 |

[1] “Relative frequency table (proportions)”

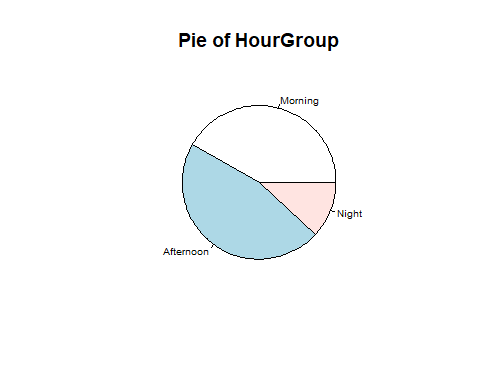
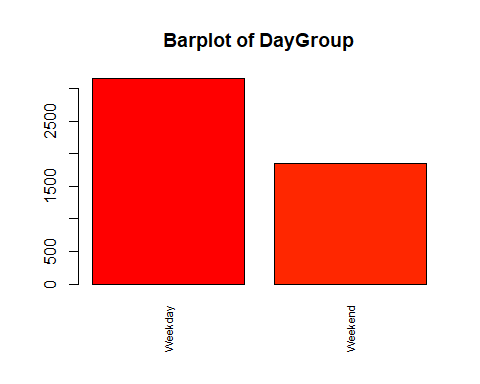
Weekday Weekend 0.6304 0.3696 [1] “Frequency table sorted”

Weekday Weekend 3152 1848 [1] “Relative frequency table (proportions) sorted”

Weekday Weekend 0.6304 0.3696

Variable 22 : HourGroup

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



Number of modalities: 3

Frequency table

| Var1 | Freq |
| --- | --- |
| Morning | 2086 |
| Afternoon | 2317 |
| Night | 597 |

[1] “Relative frequency table (proportions)”

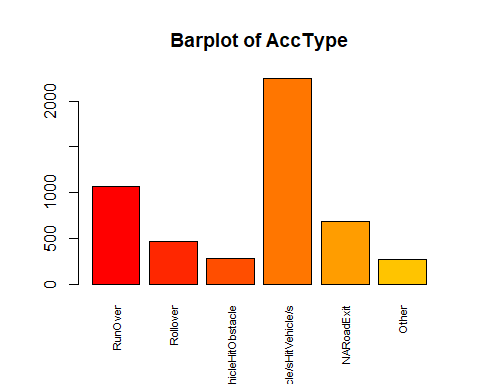
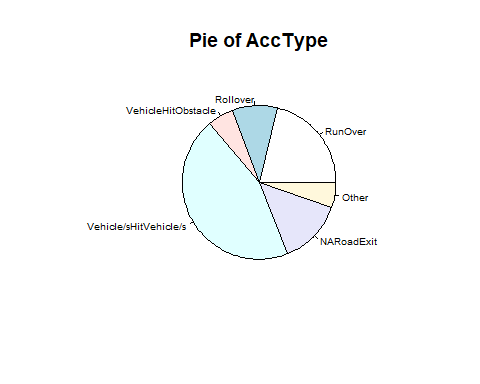
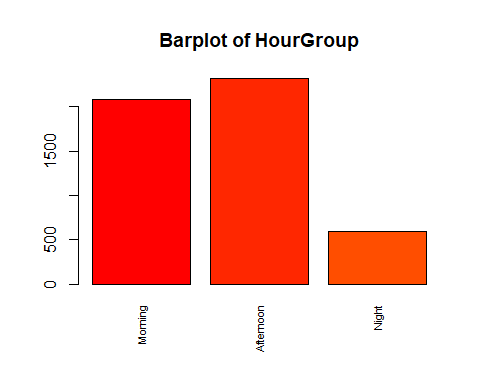
Morning Afternoon Night 0.4172 0.4634 0.1194 [1] “Frequency table sorted”

Afternoon Morning Night 2317 2086 597 [1] “Relative frequency table (proportions) sorted”

Afternoon Morning Night 0.4634 0.4172 0.1194

Variable 23 : AccType

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



Number of modalities: 6

Frequency table

| Var1 | Freq |
| --- | --- |
| RunOver | 1062 |
| Rollover | 466 |
| VehicleHitObstacle | 280 |
| Vehicle/sHitVehicle/s | 2246 |
| NARoadExit | 681 |
| Other | 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