#7TotaldescriptivaClean5.Rmd

#install.packages("rmarkdown")

#library("rmarkdown")

Información para la modificación

Para instalar el latex para r poner el comando en la consola: tinytex::install_tinytex()

Algunos de los parametros de cada chunk de codigo para poder deshabilitar algunos outputs. No me sé los valores por defecto :)

include = FALSE prevents code and results from appearing in the finished file. R Markdown still runs the code in the chunk, and the results can be used by other chunks. echo = FALSE prevents code, but not the results from appearing in the finished file. This is a useful way to embed figures. message = FALSE prevents messages that are generated by code from appearing in the finished file. warning = FALSE prevents warnings that are generated by code from appearing in the finished. fig.cap = "..." adds a caption to graphical results.

Ponemos la carpeta en la que estamos y cargamos el dataset elegido

```
setwd("C:/Users/pable/Documents/MD-FIB/")
dd<- read.csv("~/MD-FIB/weathertrain.csv")
```

Imprimos alguna información importante sobre el dataset seleccionado como en este caso el número de filas, el número de columnas y el nombre de las columnas.

```
class(dd)
```

```
## [1] "data.frame"
```

```
dim(dd)
```

```
## [1] 99516    23
```

```
filas<-dim(dd)[1]
columnas<-dim(dd)[2]

filas
```

```
## [1] 99516
```

```
columnas
```

```
## [1] 23
```

```
names(dd)
```

```
##  [1] "row.ID"       "Location"    "MinTemp"      "MaxTemp"
##  [5] "Rainfall"     "Evaporation" "Sunshine"     "WindGustDir"
##  [9] "WindGustSpeed" "WindDir9am"  "WindDir3pm"   "WindSpeed9am"
## [13] "WindSpeed3pm"  "Humidity9am" "Humidity3pm"  "Pressure9am"
## [17] "Pressure3pm"   "Cloud9am"    "Cloud3pm"     "Temp9am"
## [21] "Temp3pm"       "RainToday"   "RainTomorrow"
```

Función dedicada a imprimir los datos en forma de gráficas de cada columna por individual, hay que intentar quitar la primera columna en el caso que la primera columna sea solamente RowX o el índice de la tabla.

```r
descriptiva<-function(X, nom){
  if (!(is.numeric(X) || class(X)=="Date")){
    frecs<-table(as.factor(X), useNA="ifany")
    proportions<-frecs/filas
    #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,breaks="weeks")
    }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)))
    }
  }
}
```

Ejecución de la función por cada columna. La variable c es el rango desde la columna inicial hasta la columna final, es decir si ponemos 1:23 hará la ejecución desde la primera columna hasta la columna 23.
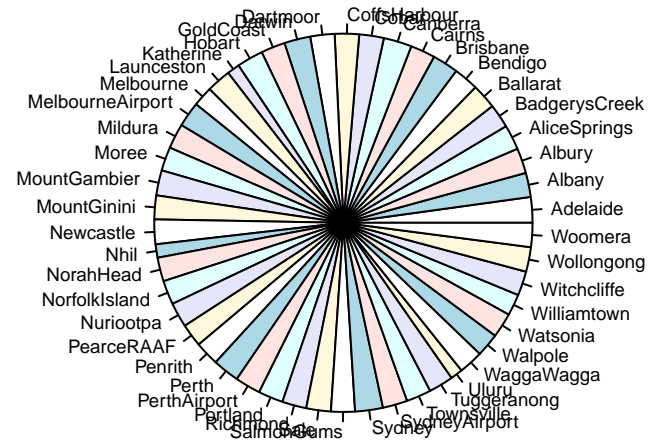
Ponemos el 2 en el inicio del recorrido para skipear la columna de row.
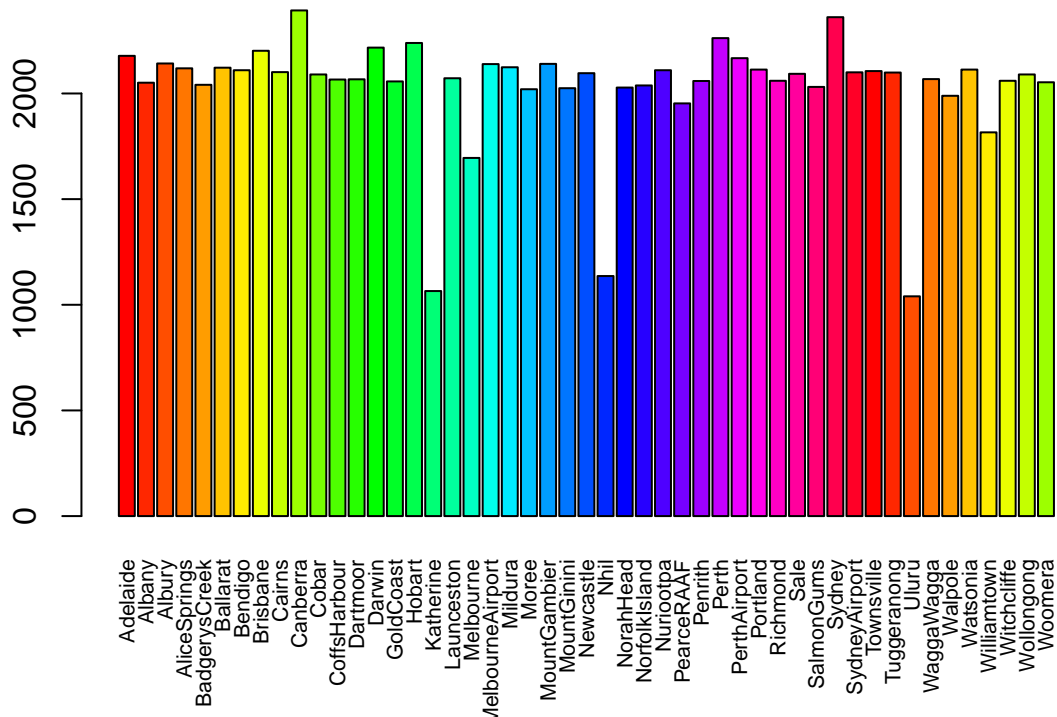
```r
listOfColors<-rainbow(39)

par(ask=TRUE)
actives<-c(2:columnas)
for(k in actives){
  print(paste("variable ", k, ":", names(dd)[k] ))
  descriptiva(dd[,k], names(dd)[k])
}
```

```
## [1] "variable  2 : Location"
```

Pie of Location

Barplot of Location


```
## [1] "Number of modalities:  49"
## [1] "Frequency table"
##
##          Adelaide          Albany          Albury     AliceSprings
##              2178            2051            2142            2119
##     BadgerysCreek         Ballarat         Bendigo        Brisbane
##              2041            2122            2110            2202
##            Cairns         Canberra           Cobar     CoffsHarbour
##              2101            2393            2090            2066
##          Dartmoor           Darwin        GoldCoast          Hobart
##              2067            2217            2057            2239
##          Katherine       Launceston       Melbourne MelbourneAirport
##              1065            2072            1695            2139
##           Mildura            Moree     MountGambier     MountGinini
##              2124            2020            2140            2025
##          Newcastle            Nhil        NorahHead    NorfolkIsland
##              2096            1136            2028            2038
##          Nuriootpa       PearceRAAF         Penrith            Perth
##              2110            1953            2059            2262
##       PerthAirport         Portland        Richmond            Sale
##              2167            2113            2060            2093
##         SalmonGums           Sydney    SydneyAirport      Townsville
##              2031            2361            2100            2106
##       Tuggeranong            Uluru      WaggaWagga          Walpole
##              2099            1040            2068            1989
##           Watsonia      Williamtown      Witchcliffe      Wollongong
```

4

```
##          2113          1816          2060          2090
##       Woomera
##          2053
## [1] "Relative frequency table (proportions)"
##
##         Adelaide         Albany         Albury     AliceSprings
##       0.02188593     0.02060975     0.02152418       0.02129306
##    BadgerysCreek        Ballarat        Bendigo         Brisbane
##       0.02050926     0.02132320     0.02120262       0.02212710
##           Cairns        Canberra          Cobar      CoffsHarbour
##       0.02111218     0.02404638     0.02100165       0.02076048
##         Dartmoor          Darwin       GoldCoast           Hobart
##       0.02077053     0.02227782     0.02067004       0.02249889
##        Katherine       Launceston       Melbourne MelbourneAirport
##       0.01070180     0.02082077     0.01703244       0.02149403
##          Mildura           Moree    MountGambier      MountGinini
##       0.02134330     0.02029824     0.02150408       0.02034849
##        Newcastle            Nhil        NorahHead    NorfolkIsland
##       0.02106194     0.01141525     0.02037863       0.02047912
##        Nuriootpa       PearceRAAF         Penrith            Perth
##       0.02120262     0.01962498     0.02069014       0.02273001
##     PerthAirport         Portland        Richmond             Sale
##       0.02177539     0.02123277     0.02070019       0.02103179
##       SalmonGums           Sydney    SydneyAirport       Townsville
##       0.02040878     0.02372483     0.02110213       0.02116243
##      Tuggeranong            Uluru      WaggaWagga          Walpole
##       0.02109209     0.01045058     0.02078058       0.01998674
##          Watsonia      Williamtown     Witchcliffe       Wollongong
##       0.02123277     0.01824832     0.02070019       0.02100165
##          Woomera
##       0.02062985
## [1] "Frequency table sorted"
##
##         Canberra          Sydney           Perth           Hobart
##             2393            2361            2262             2239
##           Darwin        Brisbane        Adelaide     PerthAirport
##             2217            2202            2178             2167
##           Albury    MountGambier MelbourneAirport         Mildura
##             2142            2140            2139             2124
##         Ballarat    AliceSprings        Portland         Watsonia
##             2122            2119            2113             2113
##          Bendigo       Nuriootpa      Townsville           Cairns
##             2110            2110            2106             2101
##    SydneyAirport     Tuggeranong       Newcastle             Sale
##             2100            2099            2096             2093
##            Cobar      Wollongong      Launceston       WaggaWagga
##             2090            2090            2072             2068
##         Dartmoor    CoffsHarbour        Richmond      Witchcliffe
##             2067            2066            2060             2060
##          Penrith       GoldCoast         Woomera           Albany
##             2059            2057            2053             2051
##    BadgerysCreek   NorfolkIsland      SalmonGums        NorahHead
##             2041            2038            2031             2028
##      MountGinini           Moree         Walpole       PearceRAAF
```

```
##           2025         2020         1989         1953
##     Williamtown    Melbourne         Nhil    Katherine
##           1816         1695         1136         1065
##          Uluru
##           1040
## [1] "Relative frequency table (proportions) sorted"
##
##        Canberra        Sydney         Perth        Hobart
##      0.02404638    0.02372483    0.02273001    0.02249889
##          Darwin      Brisbane      Adelaide   PerthAirport
##      0.02227782    0.02212710    0.02188593    0.02177539
##          Albury  MountGambier MelbourneAirport      Mildura
##      0.02152418    0.02150408    0.02149403    0.02134330
##        Ballarat   AliceSprings      Portland      Watsonia
##      0.02132320    0.02129306    0.02123277    0.02123277
##         Bendigo      Nuriootpa    Townsville        Cairns
##      0.02120262    0.02120262    0.02116243    0.02111218
##    SydneyAirport   Tuggeranong     Newcastle          Sale
##      0.02110213    0.02109209    0.02106194    0.02103179
##           Cobar    Wollongong    Launceston    WaggaWagga
##      0.02100165    0.02100165    0.02082077    0.02078058
##        Dartmoor   CoffsHarbour      Richmond    Witchcliffe
##      0.02077053    0.02076048    0.02070019    0.02070019
##         Penrith      GoldCoast       Woomera        Albany
##      0.02069014    0.02067004    0.02062985    0.02060975
##   BadgerysCreek  NorfolkIsland    SalmonGums      NorahHead
##      0.02050926    0.02047912    0.02040878    0.02037863
##     MountGinini         Moree       Walpole     PearceRAAF
##      0.02034849    0.02029824    0.01998674    0.01962498
##     Williamtown     Melbourne          Nhil     Katherine
##      0.01824832    0.01703244    0.01141525    0.01070180
##           Uluru
##      0.01045058
## [1] "variable  3 : MinTemp"
```
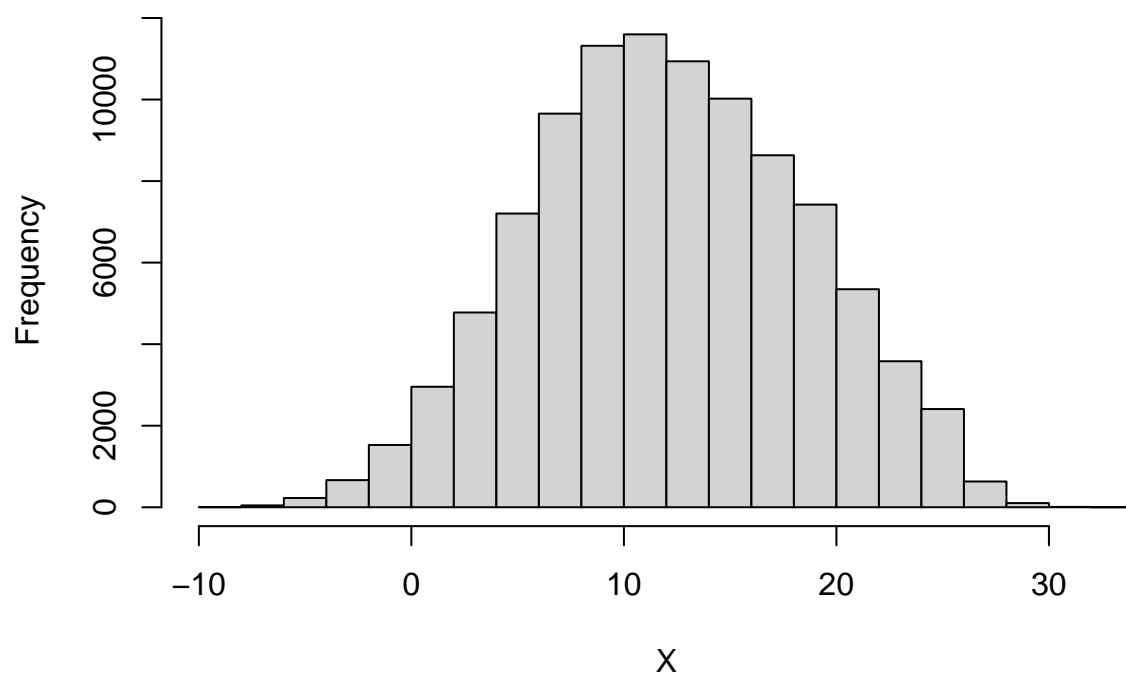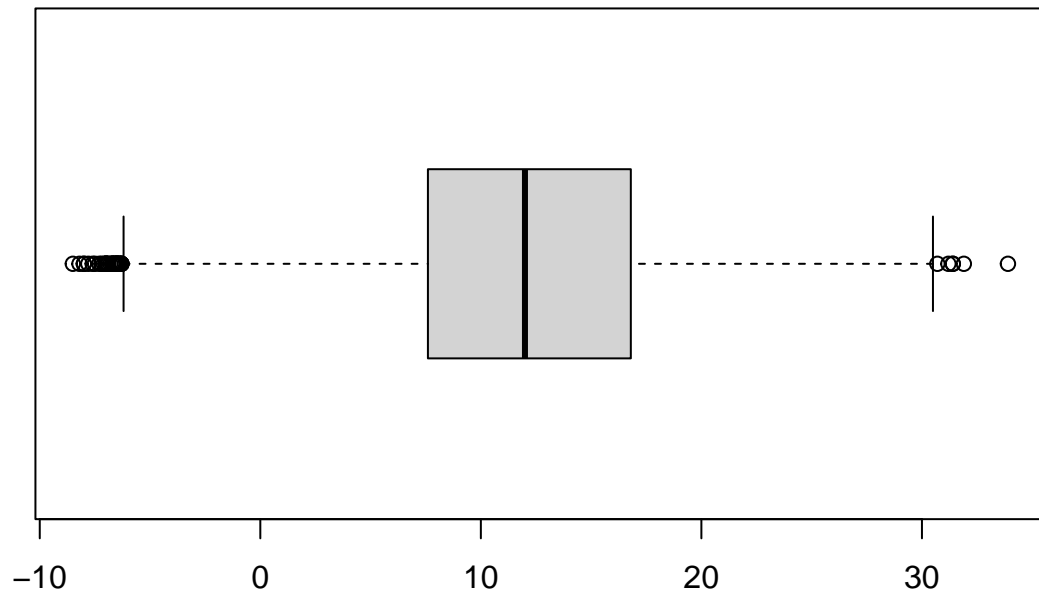
Histogram of MinTemp

Boxplot of MinTemp


```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   -8.50    7.60   12.00   12.18   16.80   33.90     443
## [1] "sd:  6.39088229056526"
## [1] "vc:  0.524863886685579"
## [1] "variable  4 : MaxTemp"
```
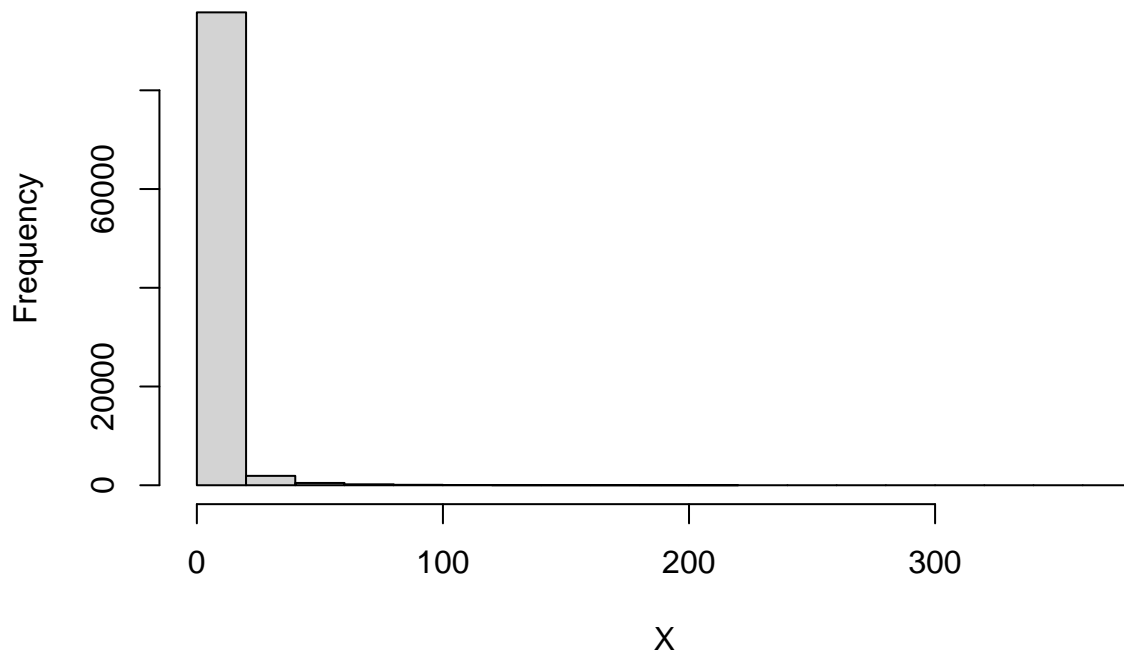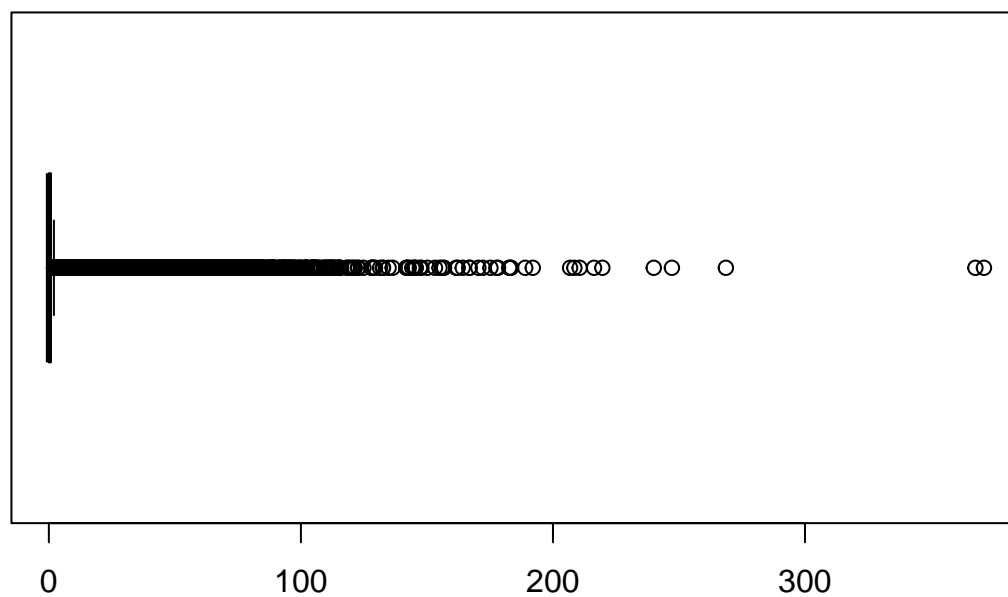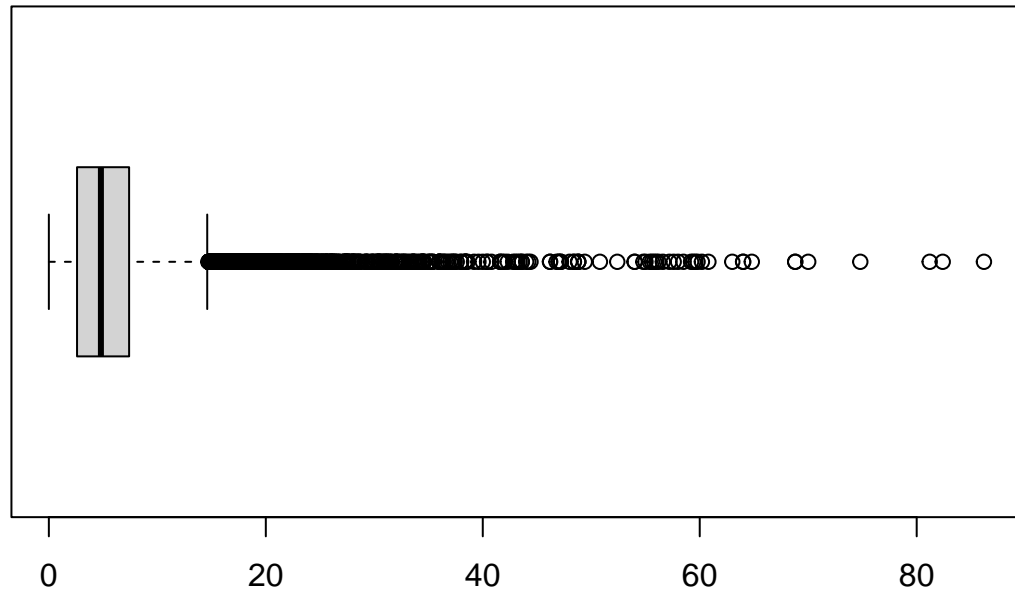
Histogram of MaxTemp
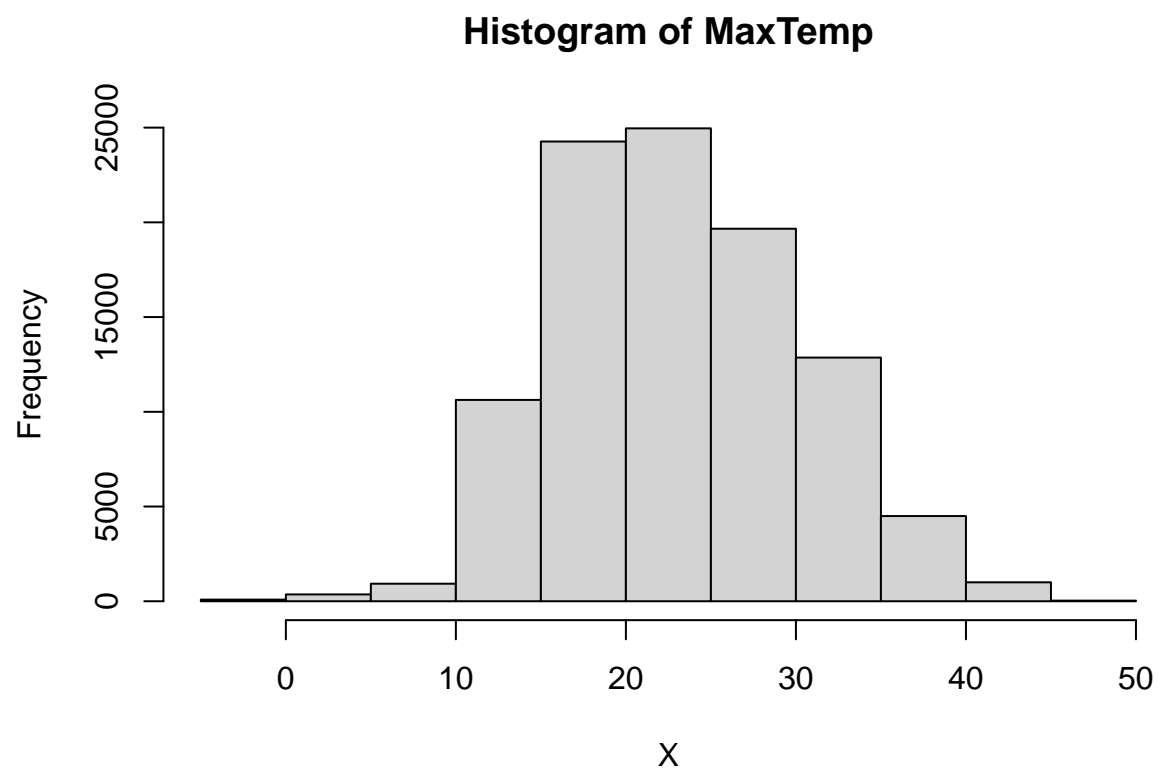
Boxplot of MaxTemp
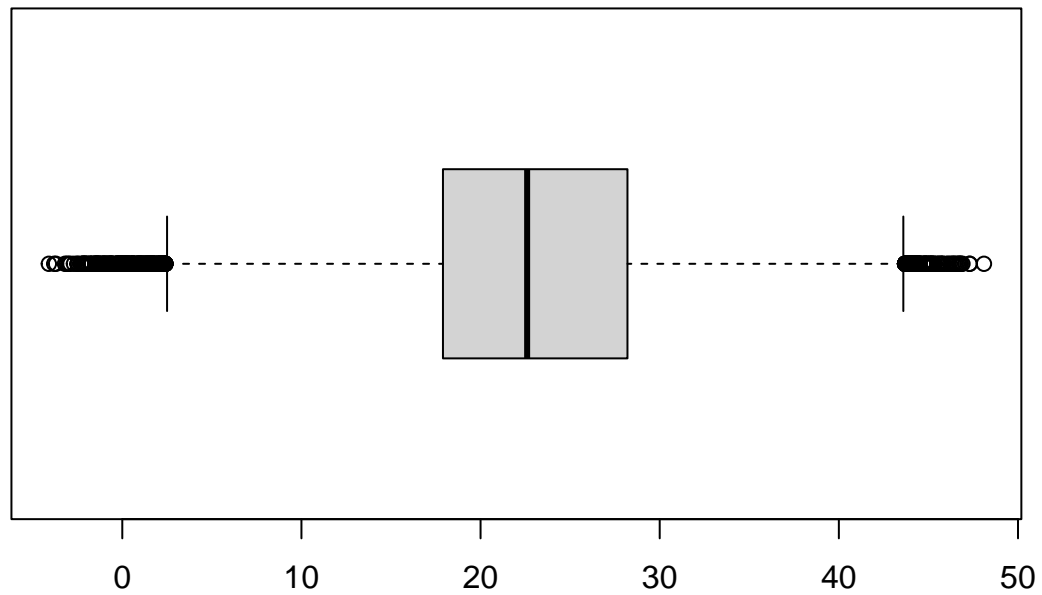

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   -4.10   17.90   22.60   23.22   28.20   48.10     230
## [1] "sd:  7.11507239837278"
## [1] "vc:  0.306439621777097"
## [1] "variable  5 : Rainfall"
```

Histogram of Rainfall

Boxplot of Rainfall


```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   0.000   0.000   2.353   0.800 371.000     979
## [1] "sd:  8.4878657266372"
## [1] "vc:  3.60721636257779"
## [1] "variable  6 : Evaporation"
```

Histogram of Evaporation
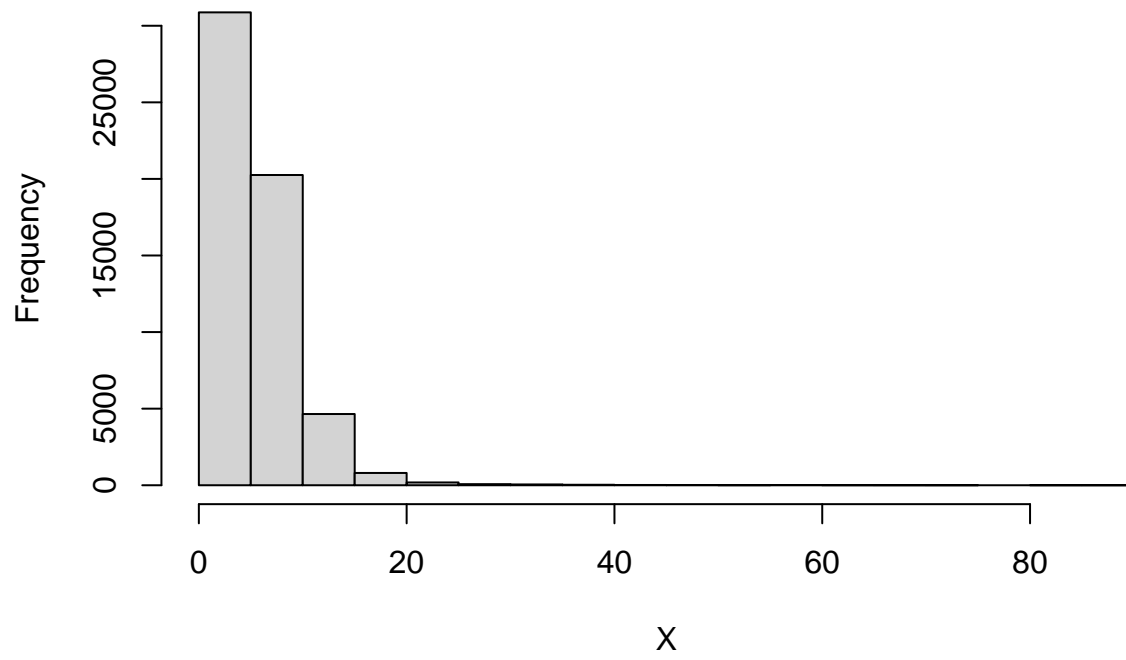
Boxplot of Evaporation


```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00    2.60    4.80    5.46    7.40   86.20   42531
## [1] "sd:  4.16249044462289"
## [1] "vc:  0.762176674283636"
## [1] "variable  7 : Sunshine"
```

Histogram of Sunshine
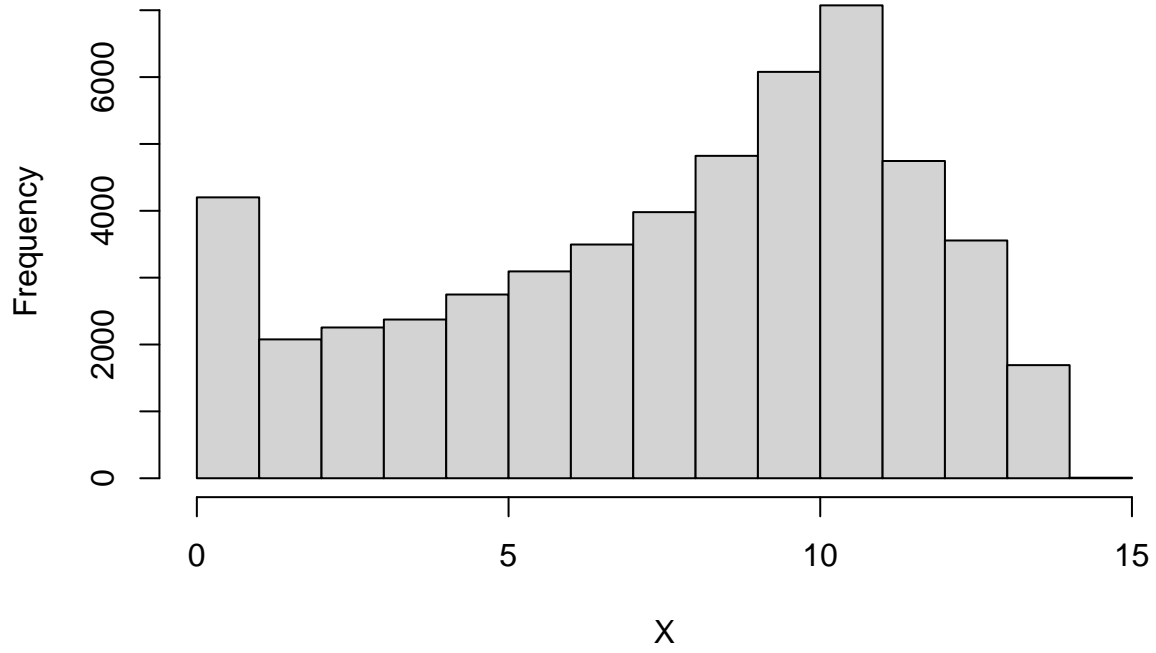
Boxplot of Sunshine
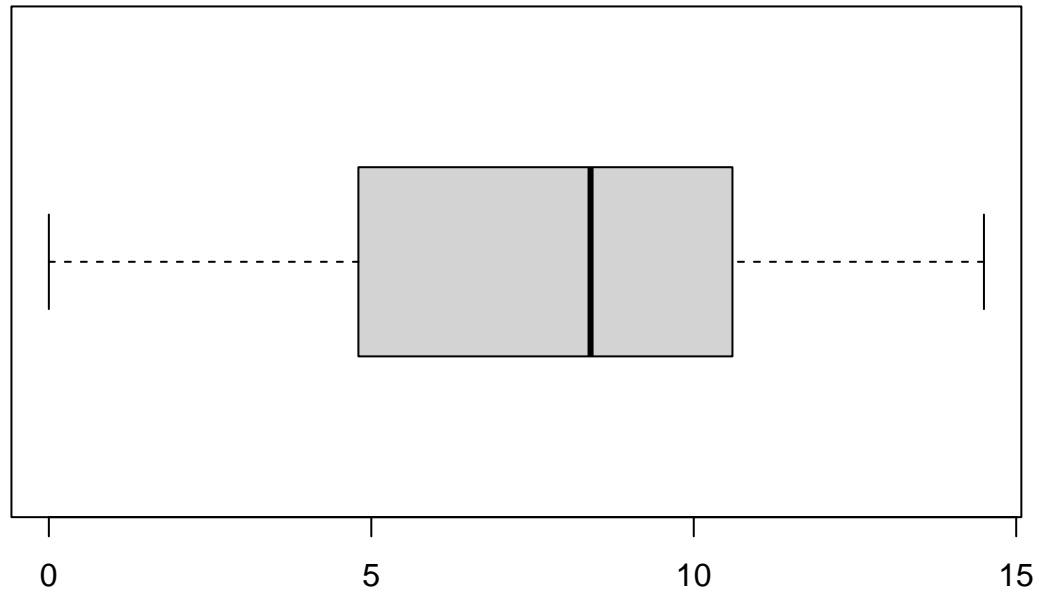

```
## [1] "Extended Summary Statistics"
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##     0.00    4.80    8.40    7.62   10.60   14.50   47317
## [1] "sd:  3.78300764622474"
## [1] "vc:  0.496777777226433"
## [1] "variable  8 : WindGustDir"
```

Pie of WindGustDir
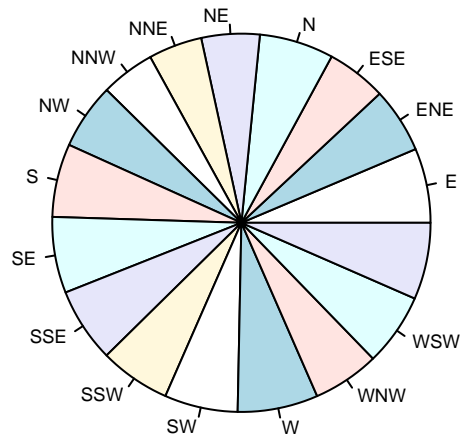
Barplot of WindGustDir
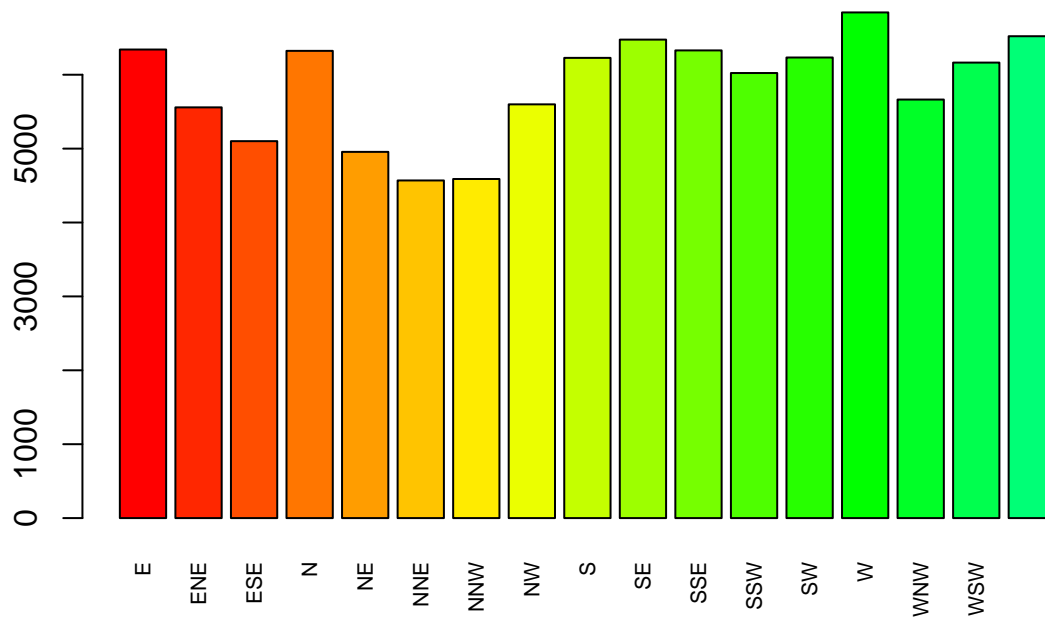

```
## [1] "Number of modalities:  17"
## [1] "Frequency table"
##
##    E  ENE  ESE    N   NE  NNE  NNW   NW    S   SE  SSE  SSW   SW    W  WNW  WSW
## 6341 5558 5101 6323 4956 4569 4589 5599 6228 6475 6329 6023 6233 6843 5664 6164
## <NA>
## 6521
## [1] "Relative frequency table (proportions)"
##
##          E        ENE        ESE          N         NE        NNE        NNW
## 0.06371840 0.05585032 0.05125809 0.06353752 0.04980104 0.04591222 0.04611319
##         NW          S         SE        SSE        SSW         SW          W
## 0.05626231 0.06258290 0.06506491 0.06359781 0.06052293 0.06263314 0.06876281
##        WNW        WSW       <NA>
## 0.05691547 0.06193979 0.06552715
## [1] "Frequency table sorted"
##
##    W <NA>   SE    E  SSE    N   SW    S  WSW  SSW  WNW   NW  ENE  ESE   NE  NNW
## 6843 6521 6475 6341 6329 6323 6233 6228 6164 6023 5664 5599 5558 5101 4956 4589
##  NNE
## 4569
## [1] "Relative frequency table (proportions) sorted"
##
##          W       <NA>         SE          E        SSE          N         SW
## 0.06876281 0.06552715 0.06506491 0.06371840 0.06359781 0.06353752 0.06263314
##          S        WSW        SSW        WNW         NW        ENE        ESE
```

```
## 0.06258290 0.06193979 0.06052293 0.05691547 0.05626231 0.05585032 0.05125809
##         NE        NNW        NNE
## 0.04980104 0.04611319 0.04591222
## [1] "variable  9 : WindGustSpeed"
```

Histogram of WindGustSpeed
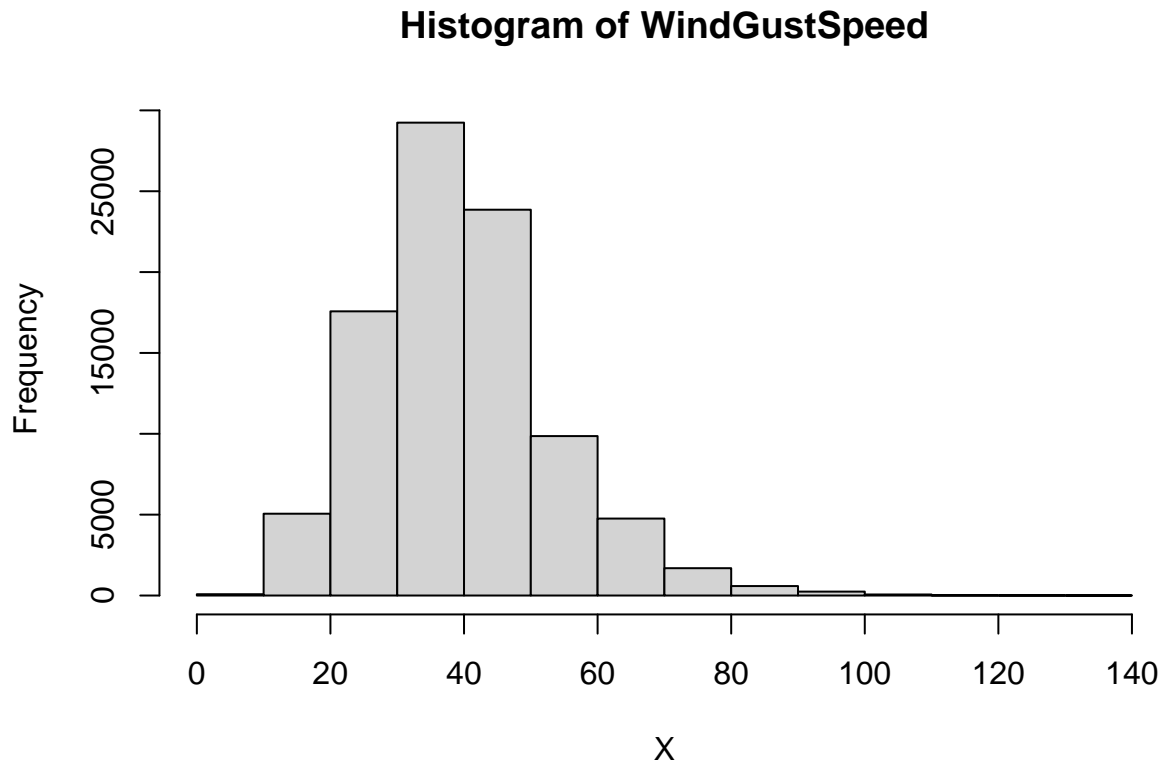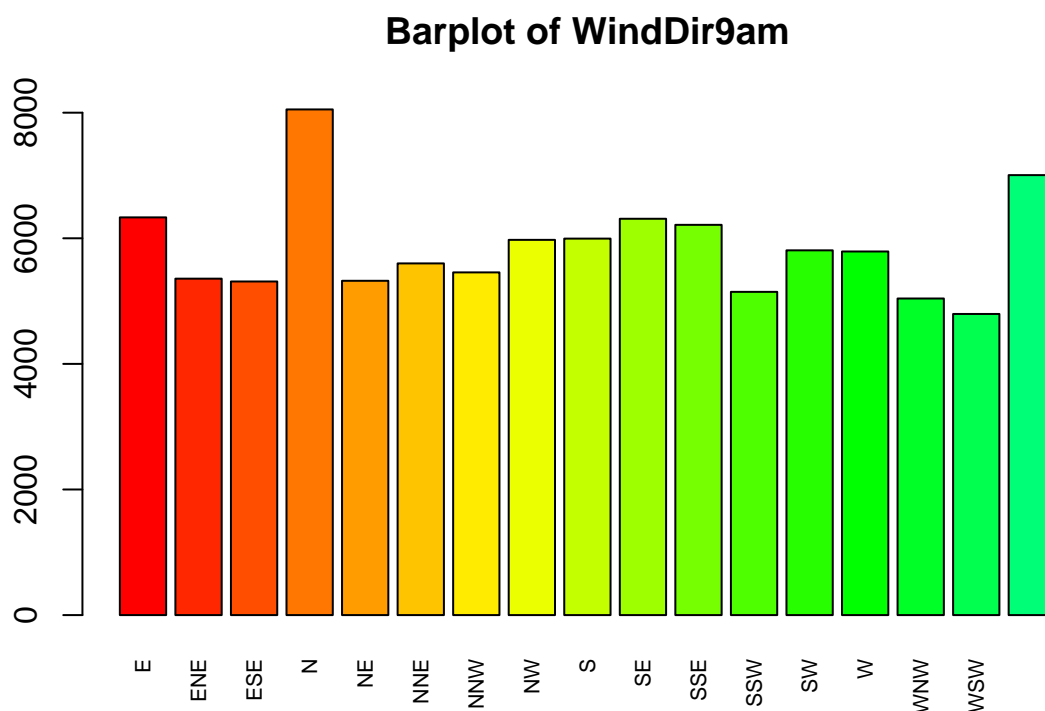
Boxplot of WindGustSpeed
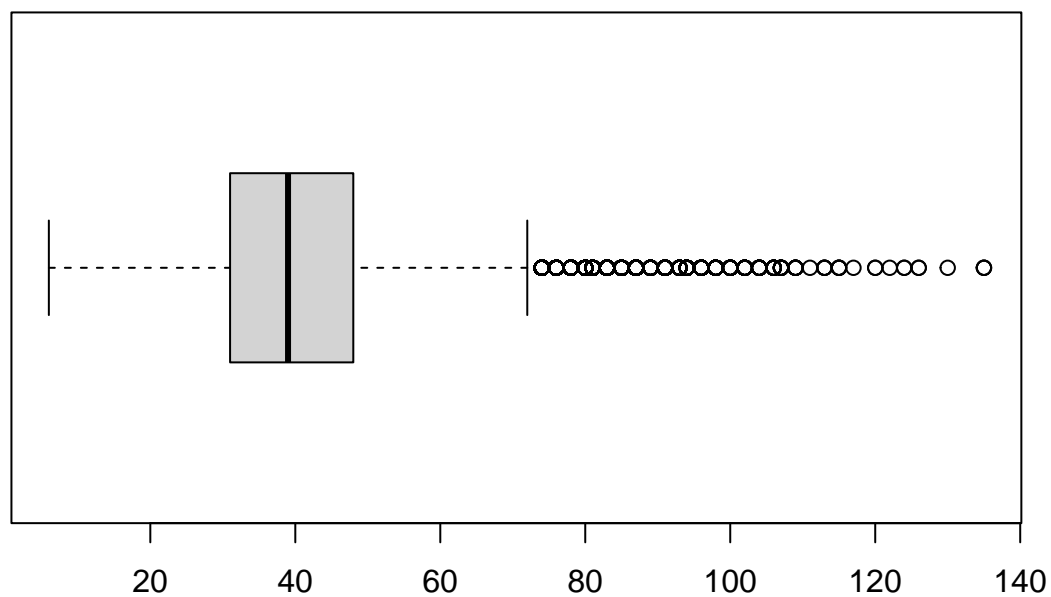

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    6.00   31.00   39.00   39.98   48.00  135.00    6480
## [1] "sd:  13.5815235085385"
## [1] "vc:  0.3397337241797"
## [1] "variable  10 : WindDir9am"
```

Pie of WindDir9am

Barplot of WindDir9am
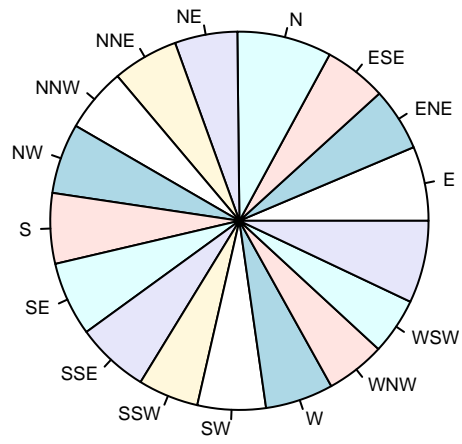

```
## [1] "Number of modalities:  17"
## [1] "Frequency table"
##
##    E  ENE  ESE    N   NE  NNE  NNW   NW    S   SE  SSE  SSW   SW    W  WNW  WSW
## 6333 5357 5312 8052 5323 5600 5457 5975 5995 6311 6214 5147 5808 5790 5041 4795
## <NA>
## 7006
## [1] "Relative frequency table (proportions)"
##
##          E        ENE        ESE          N         NE        NNE        NNW
## 0.06363801 0.05383054 0.05337835 0.08091161 0.05348889 0.05627236 0.05483540
##         NW          S         SE        SSE        SSW         SW          W
## 0.06004060 0.06024157 0.06341694 0.06244222 0.05172033 0.05836247 0.05818160
##        WNW        WSW       <NA>
## 0.05065517 0.04818321 0.07040074
## [1] "Frequency table sorted"
##
##    N <NA>    E   SE  SSE    S   NW   SW    W  NNE  NNW  ENE   NE  ESE  SSW  WNW
## 8052 7006 6333 6311 6214 5995 5975 5808 5790 5600 5457 5357 5323 5312 5147 5041
##  WSW
## 4795
## [1] "Relative frequency table (proportions) sorted"
##
##          N       <NA>          E         SE        SSE          S         NW
## 0.08091161 0.07040074 0.06363801 0.06341694 0.06244222 0.06024157 0.06004060
##         SW          W        NNE        NNW        ENE         NE        ESE
```

```
## 0.05836247 0.05818160 0.05627236 0.05483540 0.05383054 0.05348889 0.05337835
##        SSW        WNW        WSW
## 0.05172033 0.05065517 0.04818321
## [1] "variable  11 : WindDir3pm"
```

Pie of WindDir3pm
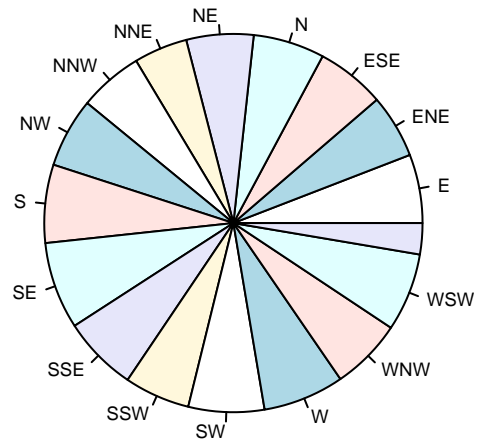
## Barplot of WindDir3pm
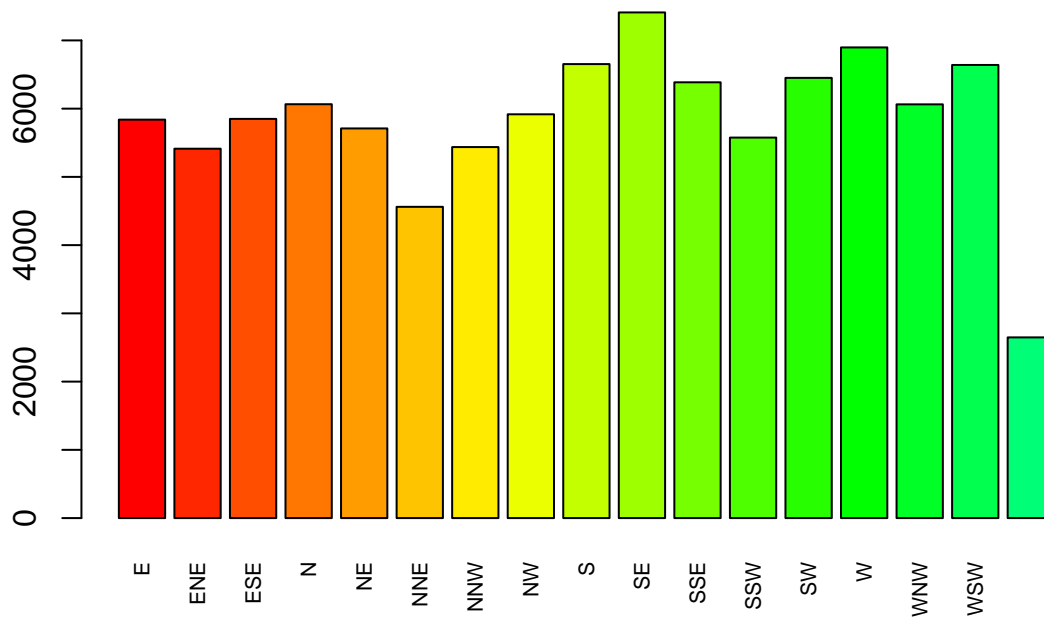

```
## [1] "Number of modalities:  17"
## [1] "Frequency table"
##
##    E  ENE  ESE    N   NE  NNE  NNW   NW    S   SE  SSE  SSW   SW    W  WNW  WSW
## 5839 5413 5850 6065 5710 4561 5437 5917 6653 7410 6386 5576 6451 6897 6063 6640
## <NA>
## 2648
## [1] "Relative frequency table (proportions)"
##
##          E        ENE        ESE          N         NE        NNE        NNW
## 0.05867398 0.05439326 0.05878452 0.06094497 0.05737771 0.04583183 0.05463443
##         NW          S         SE        SSE        SSW         SW          W
## 0.05945778 0.06685357 0.07446039 0.06417059 0.05603119 0.06482375 0.06930544
##        WNW        WSW       <NA>
## 0.06092488 0.06672294 0.02660879
## [1] "Frequency table sorted"
##
##   SE    W    S  WSW   SW  SSE    N  WNW   NW  ESE    E   NE  SSW  NNW  ENE  NNE
## 7410 6897 6653 6640 6451 6386 6065 6063 5917 5850 5839 5710 5576 5437 5413 4561
## <NA>
## 2648
## [1] "Relative frequency table (proportions) sorted"
##
##         SE          W          S        WSW         SW        SSE          N
## 0.07446039 0.06930544 0.06685357 0.06672294 0.06482375 0.06417059 0.06094497
##        WNW         NW        ESE          E         NE        SSW        NNW
```

```
## 0.06092488 0.05945778 0.05878452 0.05867398 0.05737771 0.05603119 0.05463443
##         ENE        NNE       <NA>
## 0.05439326 0.04583183 0.02660879
## [1] "variable  12 : WindSpeed9am"
```

Histogram of WindSpeed9am
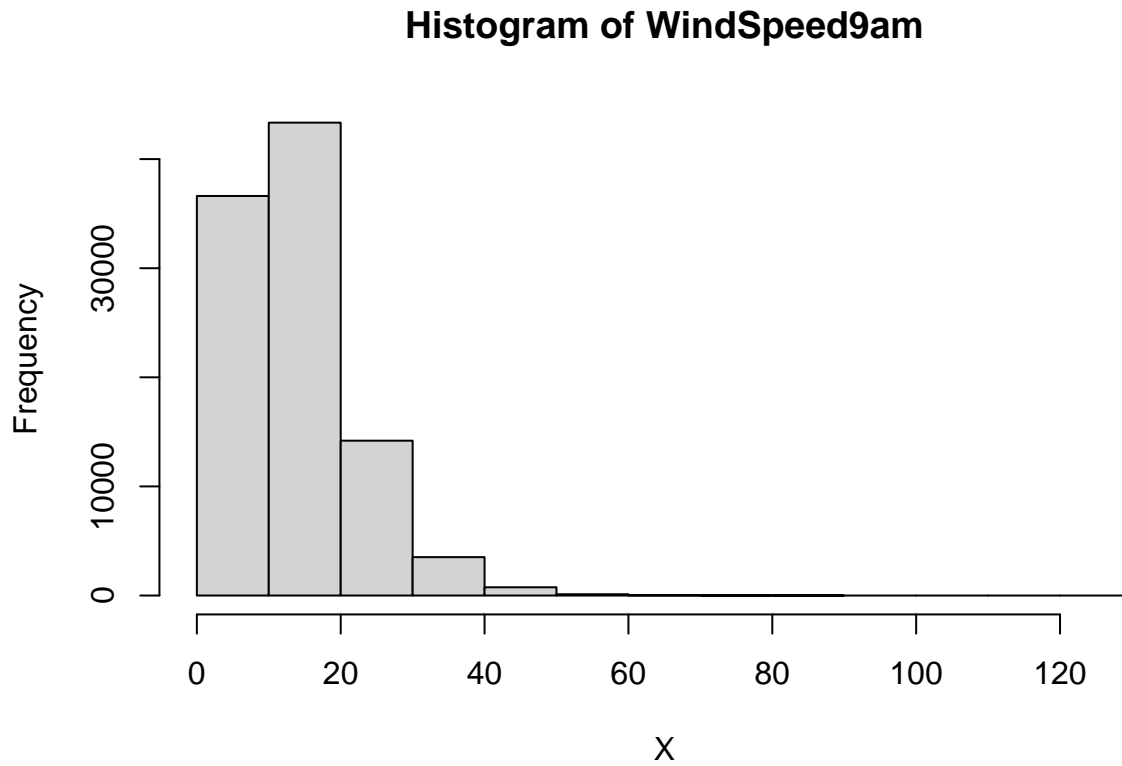
Boxplot of WindSpeed9am
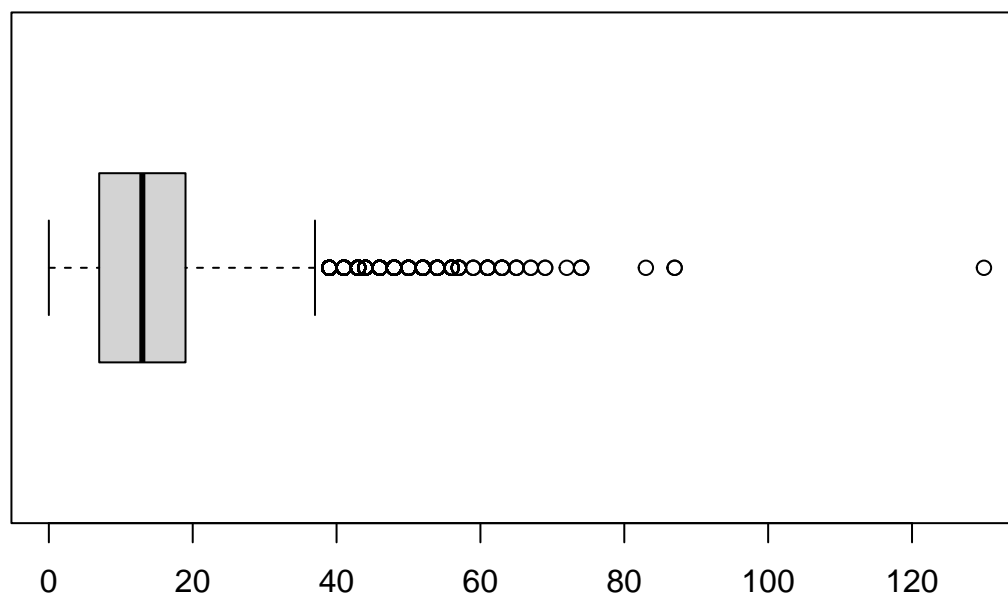

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##       0       7      13      14      19     130     935
## [1] "sd:  8.90232250230192"
## [1] "vc:  0.635660022221613"
## [1] "variable  13 : WindSpeed3pm"
```
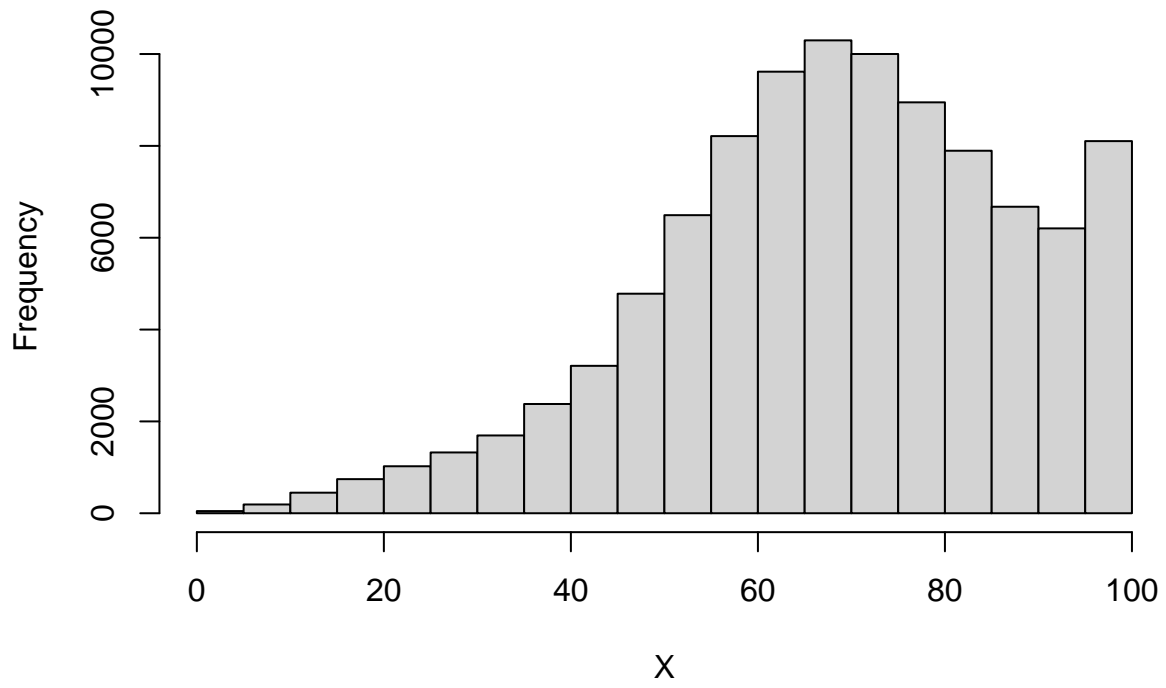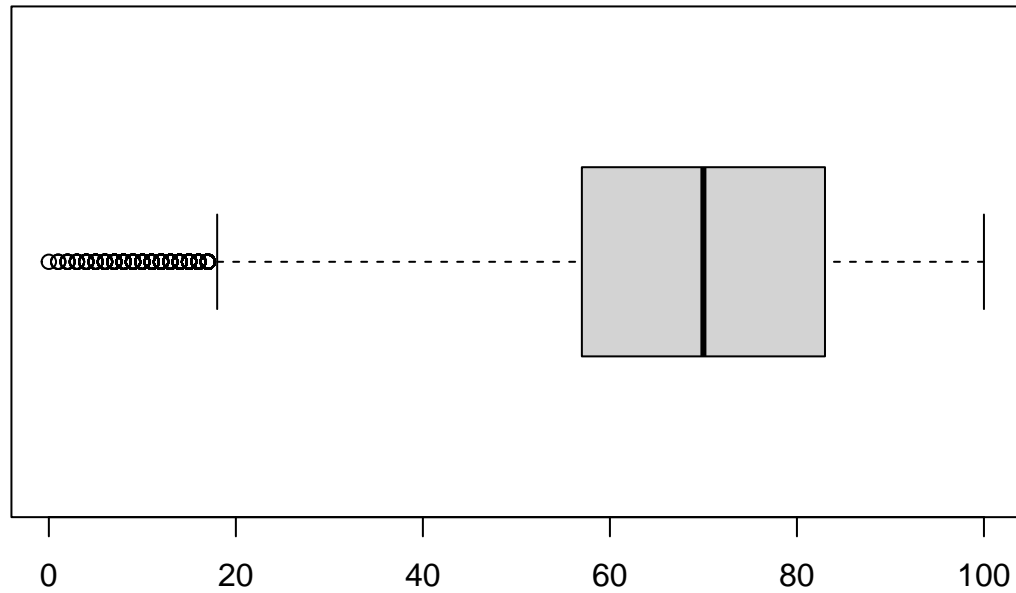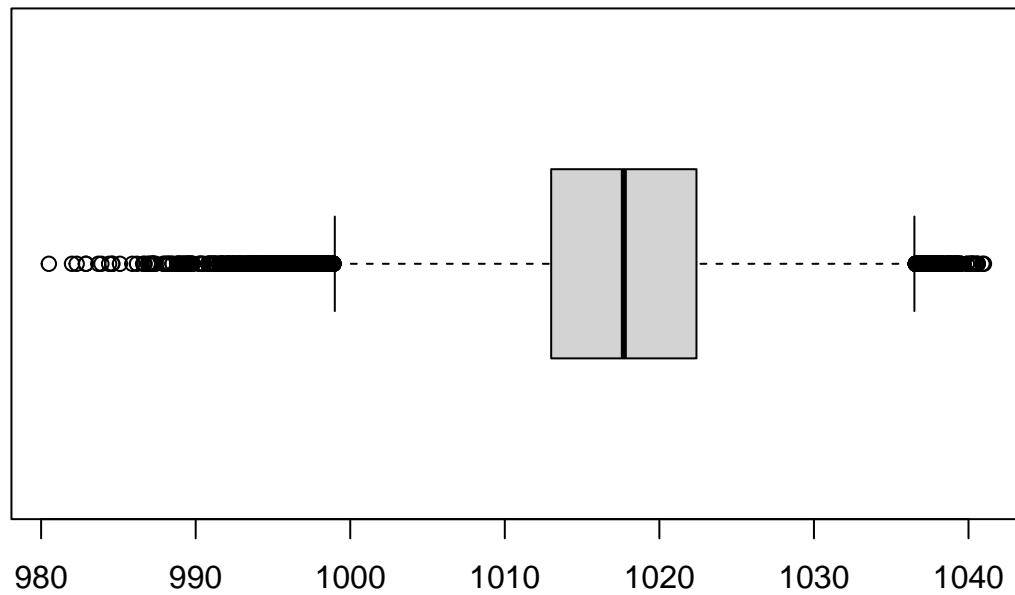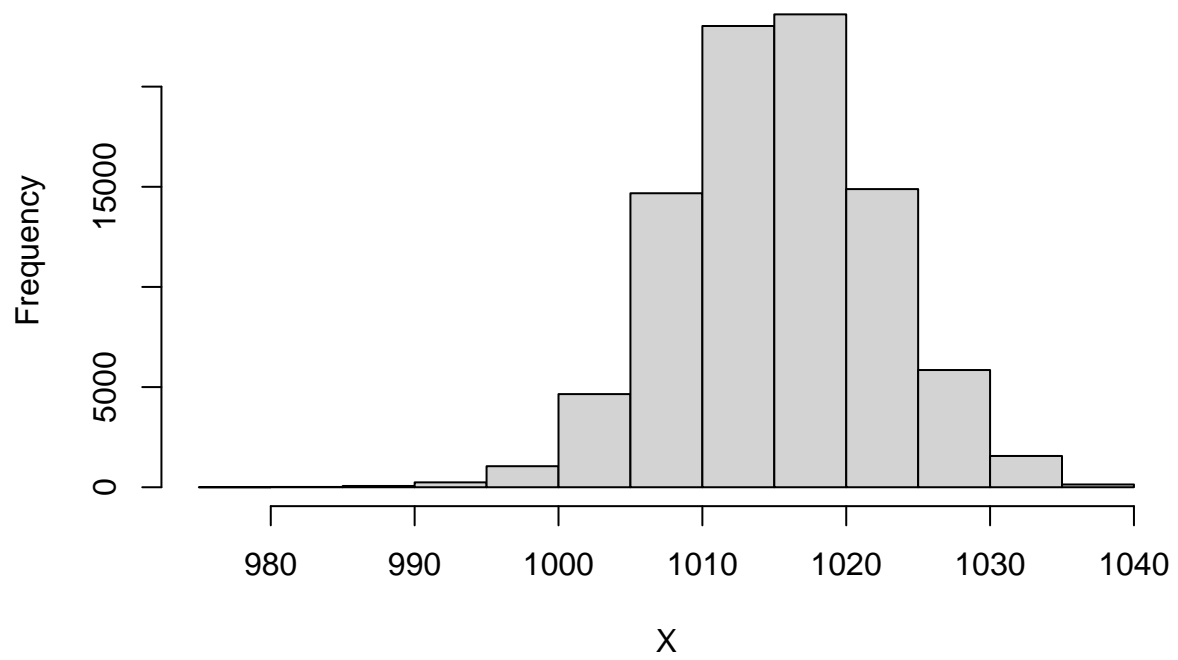
Histogram of WindSpeed3pm
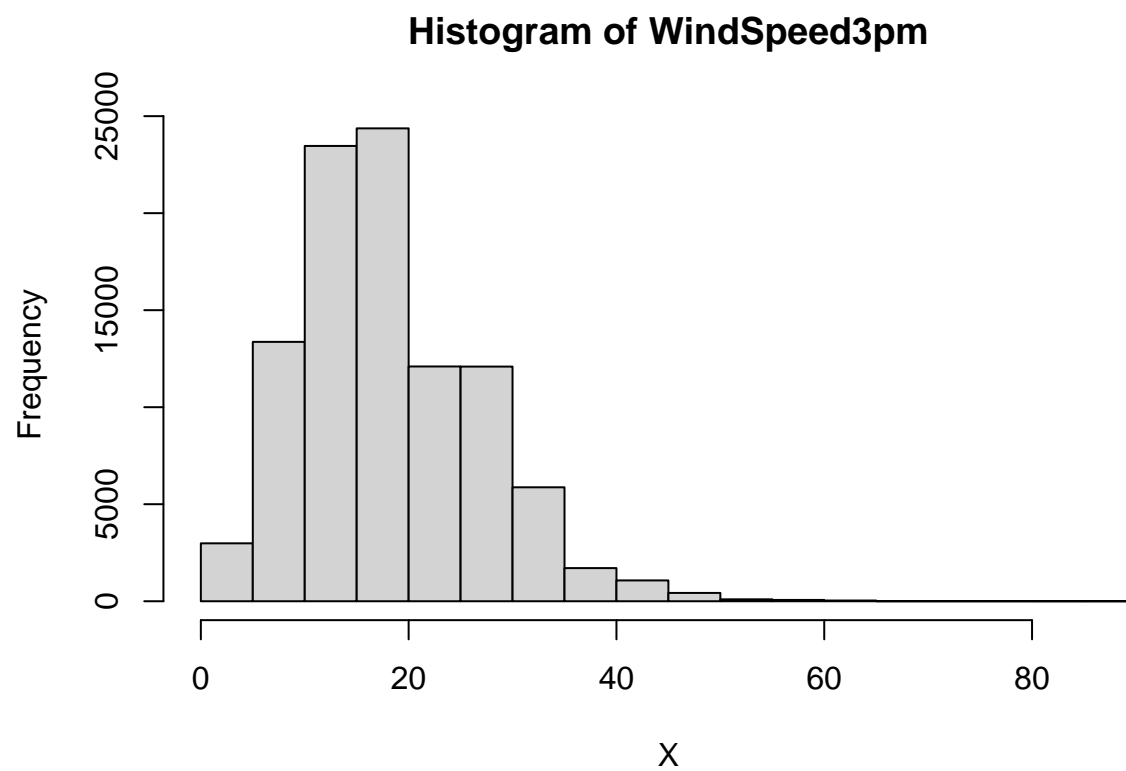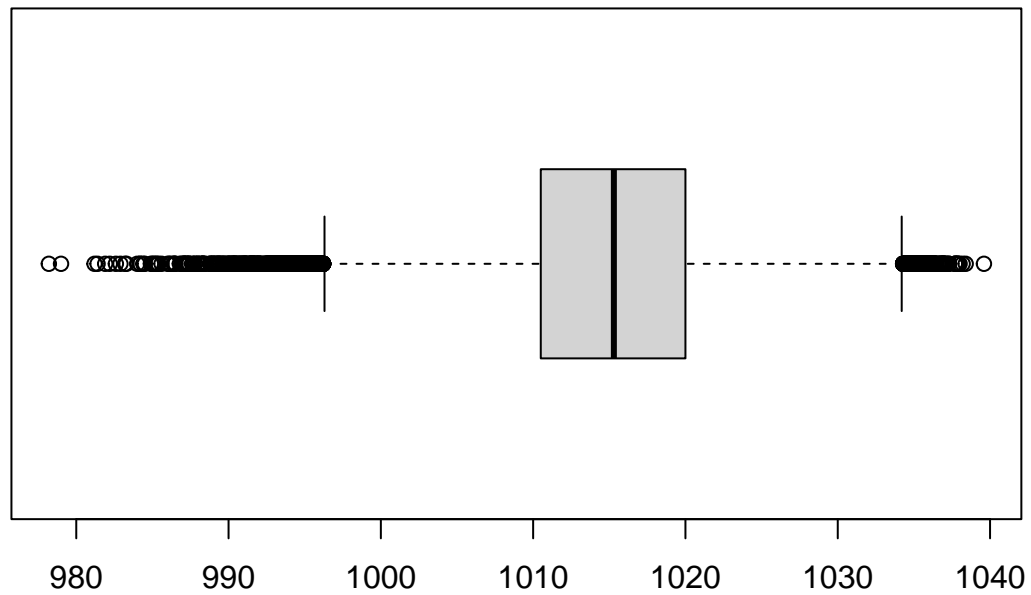
Boxplot of WindSpeed3pm
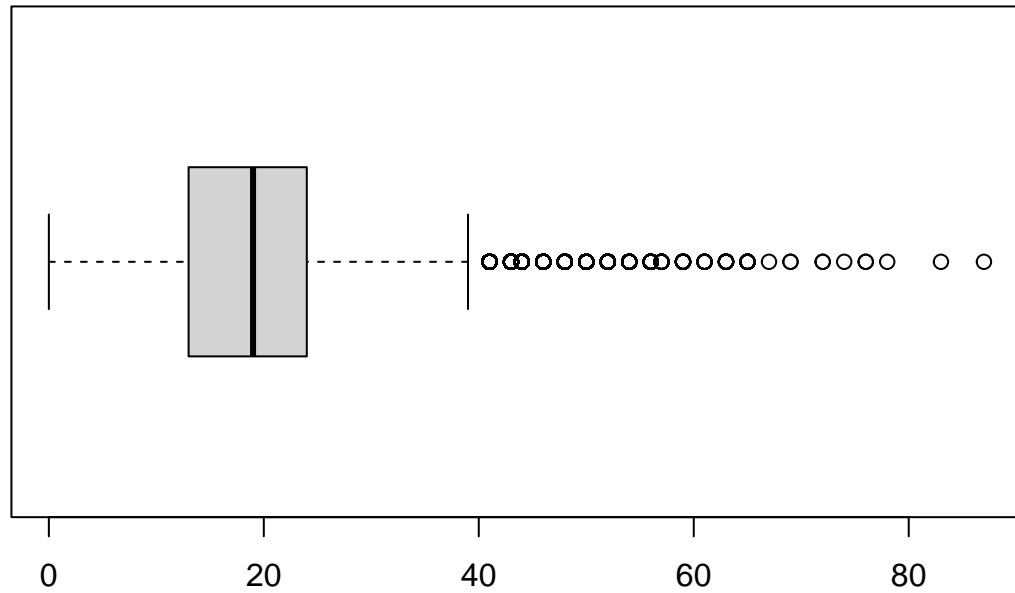

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##    0.00   13.00   19.00   18.65   24.00   87.00     1835
## [1] "sd:  8.80182720672024"
## [1] "vc:  0.471936091296523"
## [1] "variable  14 : Humidity9am"
```

Histogram of Humidity9am

Boxplot of Humidity9am


```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   57.00   70.00   68.87   83.00  100.00    1233
## [1] "sd:  19.0749513463571"
## [1] "vc:  0.276984975043417"
## [1] "variable  15 : Humidity3pm"
```

Histogram of Humidity3pm
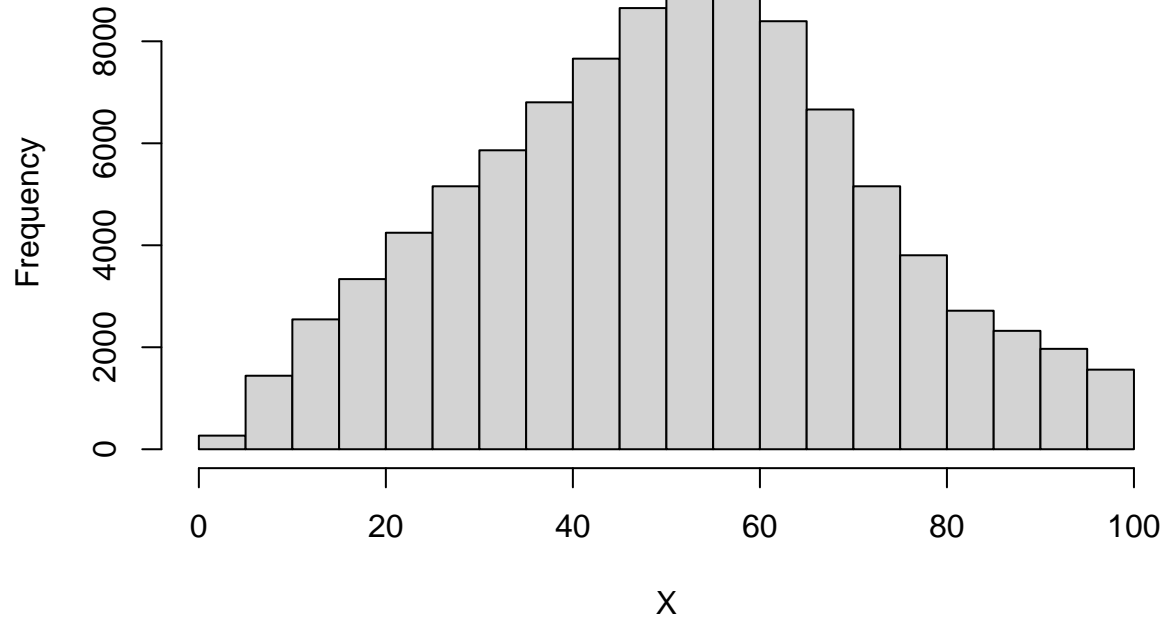
Boxplot of Humidity3pm
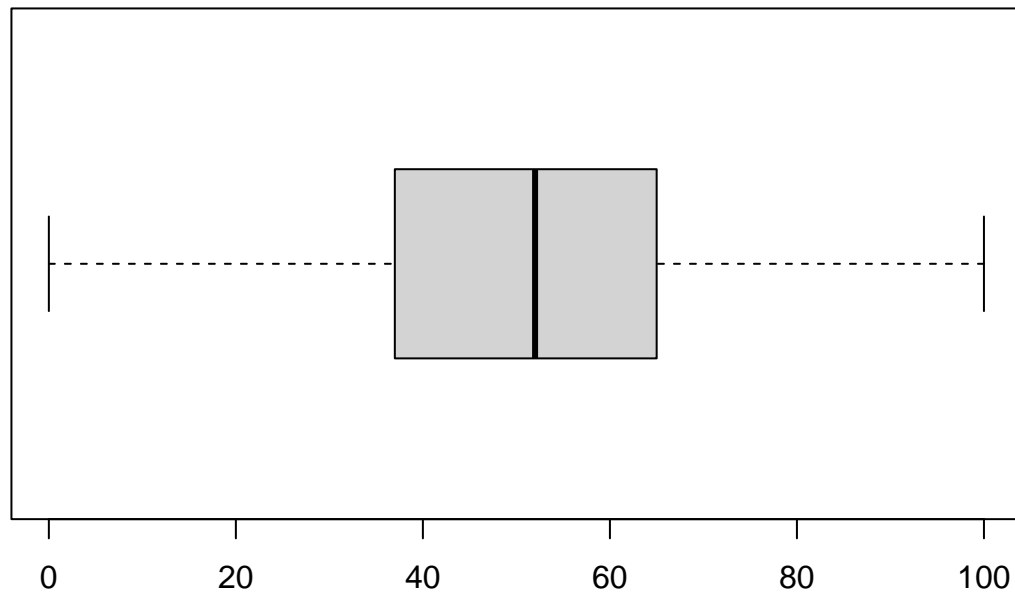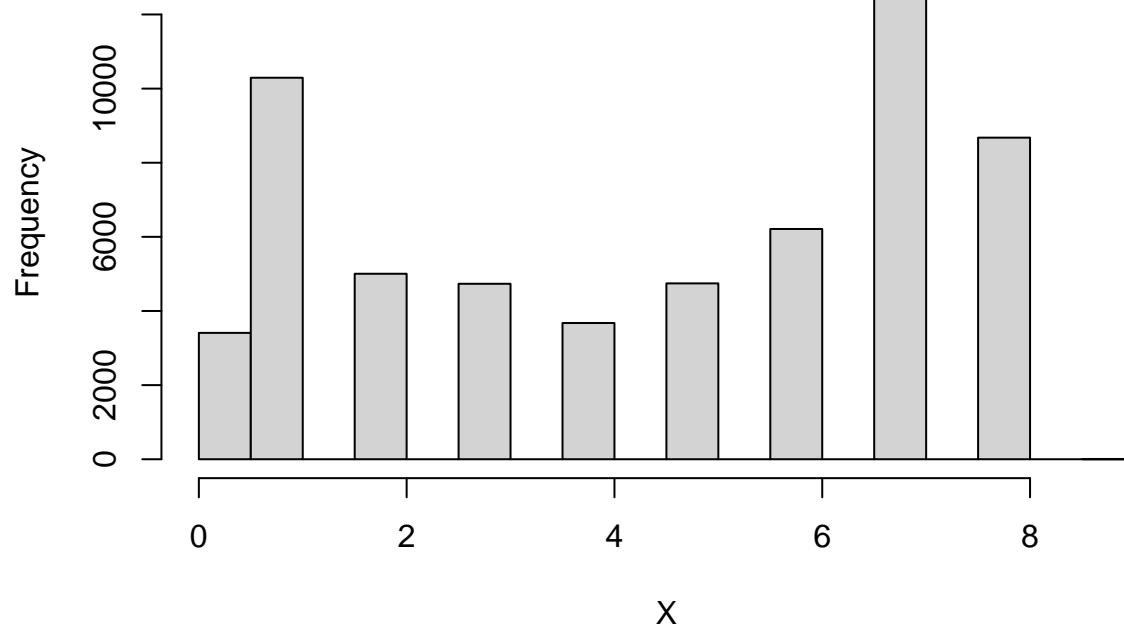


```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   37.00   52.00   51.43   65.00  100.00    2506
## [1] "sd:  20.7776161952411"
## [1] "vc:  0.403972095866944"
## [1] "variable  16 : Pressure9am"
```

Histogram of Pressure9am
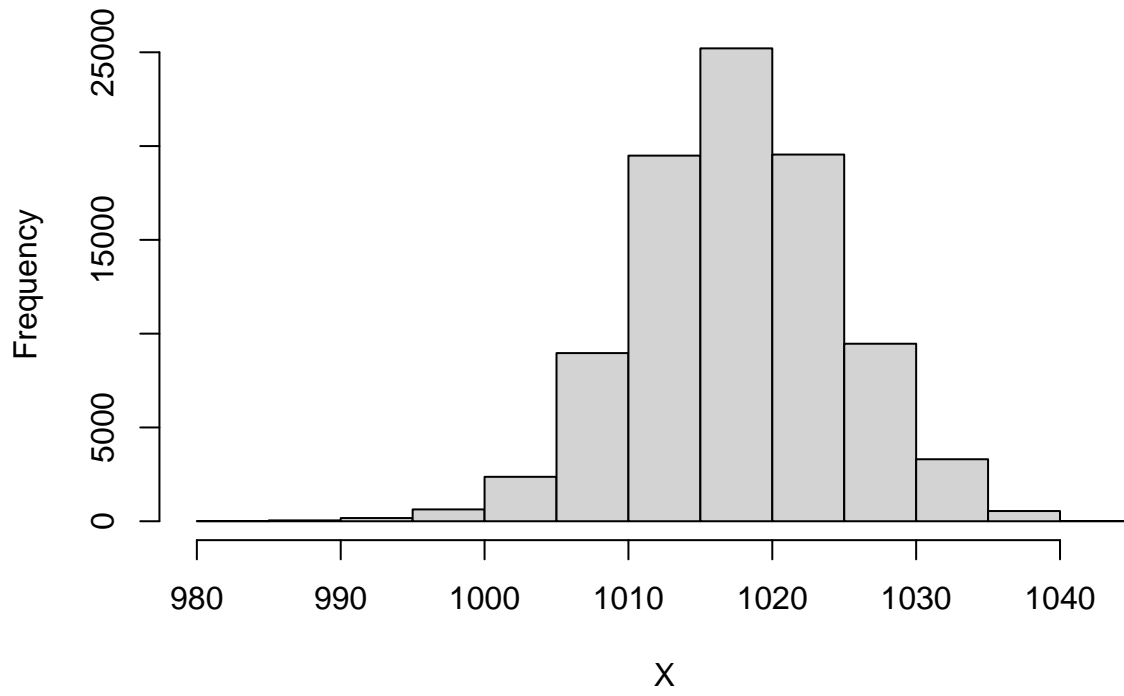
Boxplot of Pressure9am


```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   980.5  1013.0  1017.7  1017.7  1022.4  1041.0    9748
## [1] "sd:  7.11016624760789"
## [1] "vc:  0.00698661056762593"
## [1] "variable  17 : Pressure3pm"
```

Histogram of Pressure3pm

Boxplot of Pressure3pm


```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   978.2  1010.5  1015.3  1015.3  1020.0  1039.6    9736
## [1] "sd:  7.04518877022009"
## [1] "vc:  0.00693911602553753"
## [1] "variable  18 : Cloud9am"
```

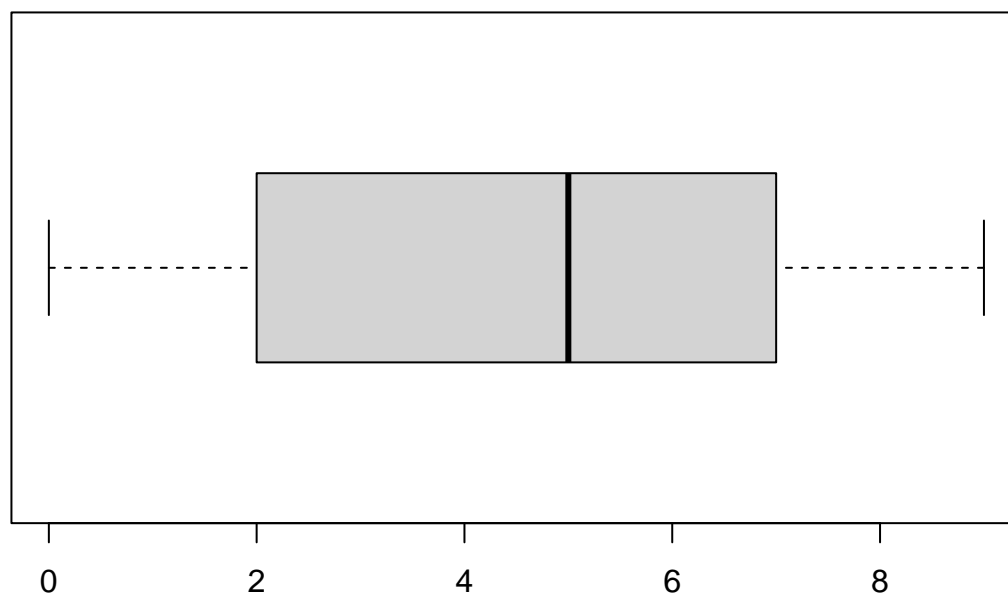# Histogram of Cloud9am
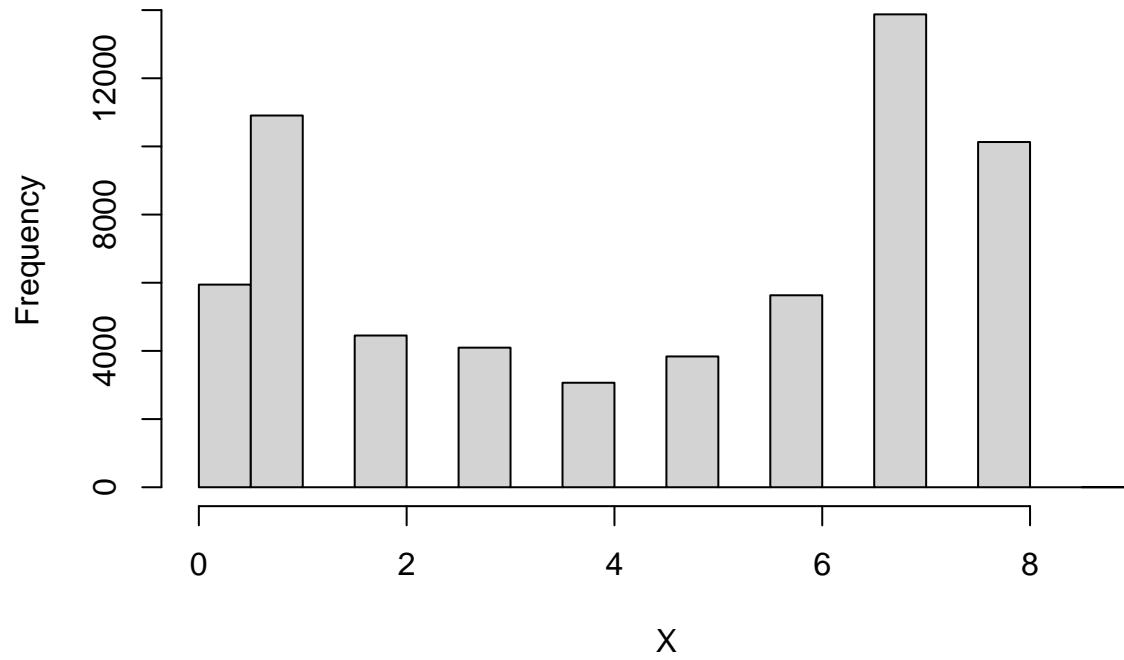
Boxplot of Cloud9am
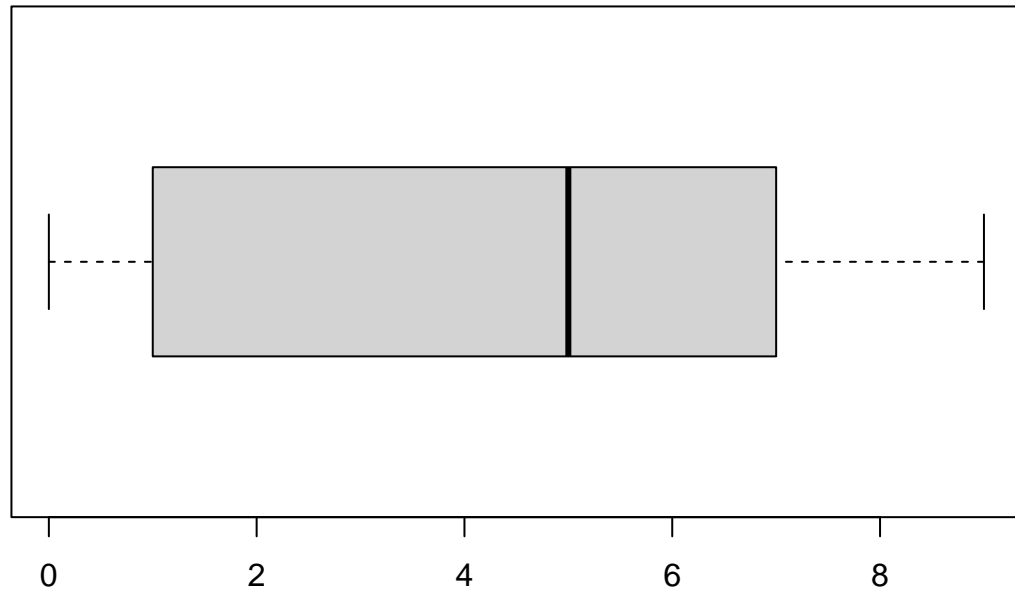

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00    1.00    5.00    4.45    7.00    9.00   37572
## [1] "sd:  2.88658005089183"
## [1] "vc:  0.648963490459861"
## [1] "variable  19 : Cloud3pm"
```
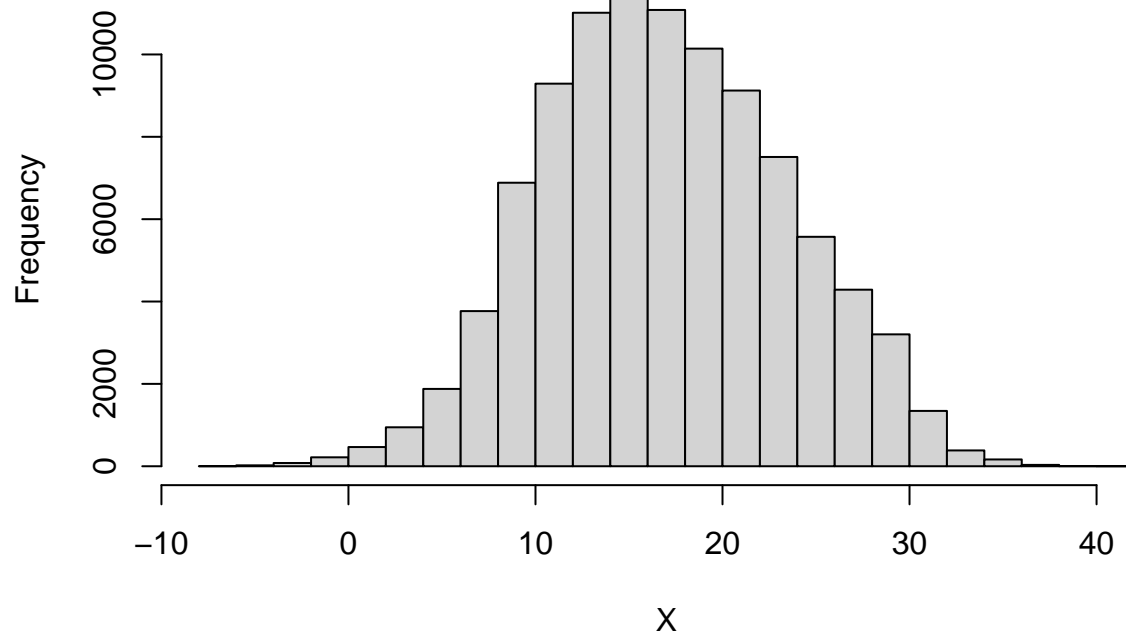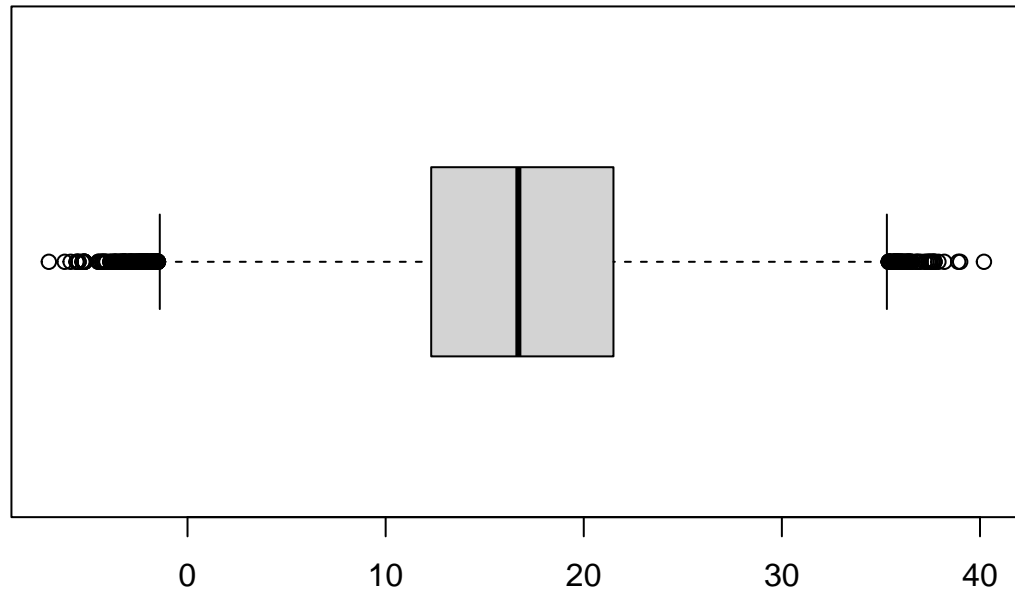
Histogram of Cloud3pm

Boxplot of Cloud3pm


```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##    0.00    2.00    5.00    4.52    7.00    9.00   40002
## [1] "sd:  2.71661801454954"
## [1] "vc:  0.601138514145333"
## [1] "variable  20 : Temp9am"
```
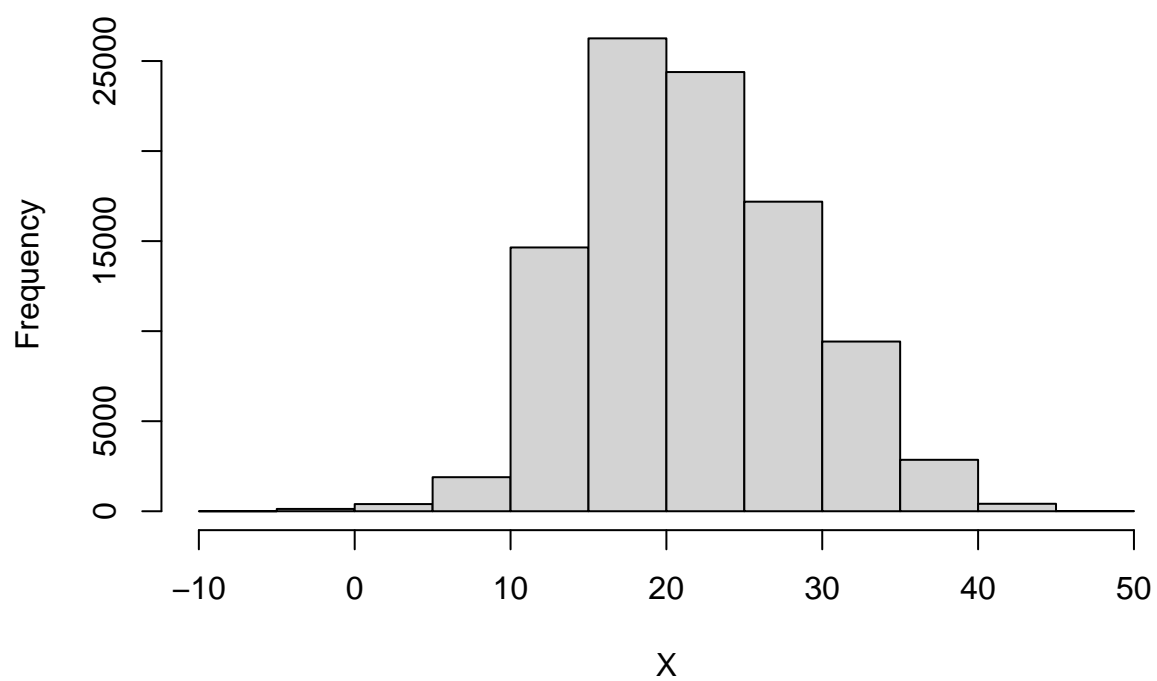
Histogram of Temp9am

Boxplot of Temp9am
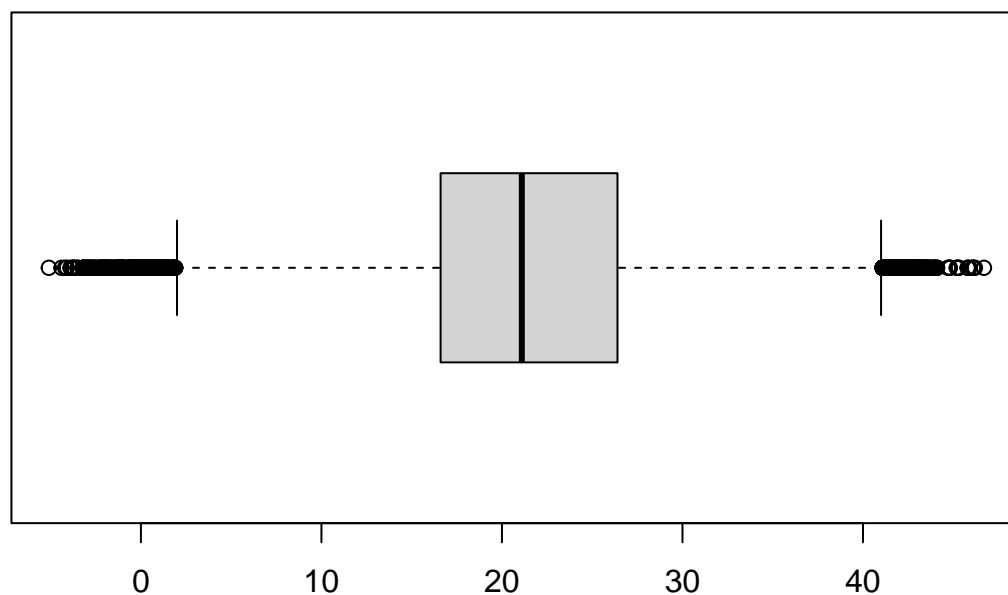


```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   -7.00   12.30   16.70   16.97   21.50   40.20     614
## [1] "sd:  6.48896079385892"
## [1] "vc:  0.38237743647515"
## [1] "variable  21 : Temp3pm"
```

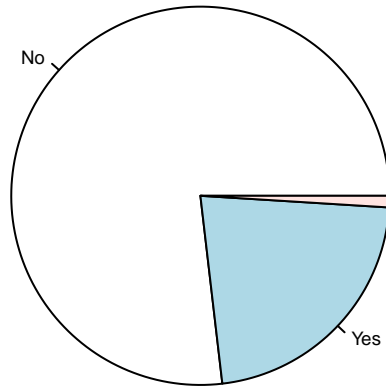# Histogram of Temp3pm

Boxplot of Temp3pm
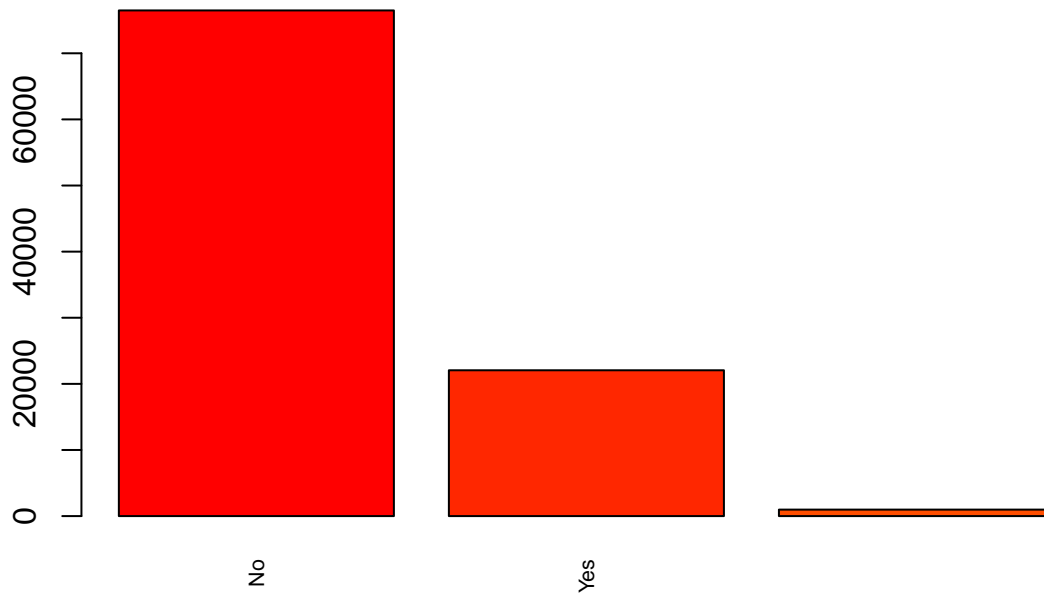


```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   -5.10   16.60   21.10   21.68   26.40   46.70    1904
## [1] "sd:  6.93168147161278"
## [1] "vc:  0.319707238614558"
## [1] "variable  22 : RainToday"
```

Pie of RainToday

Barplot of RainToday


```
## [1] "Number of modalities:  3"
## [1] "Frequency table"
##
##    No   Yes  <NA>
## 76481 22056   979
## [1] "Relative frequency table (proportions)"
##
##         No        Yes       <NA>
## 0.768529684 0.221632702 0.009837614
## [1] "Frequency table sorted"
##
##    No   Yes  <NA>
## 76481 22056   979
## [1] "Relative frequency table (proportions) sorted"
##
##         No        Yes       <NA>
## 0.768529684 0.221632702 0.009837614
## [1] "variable  23 : RainTomorrow"
```

Histogram of RainTomorrow
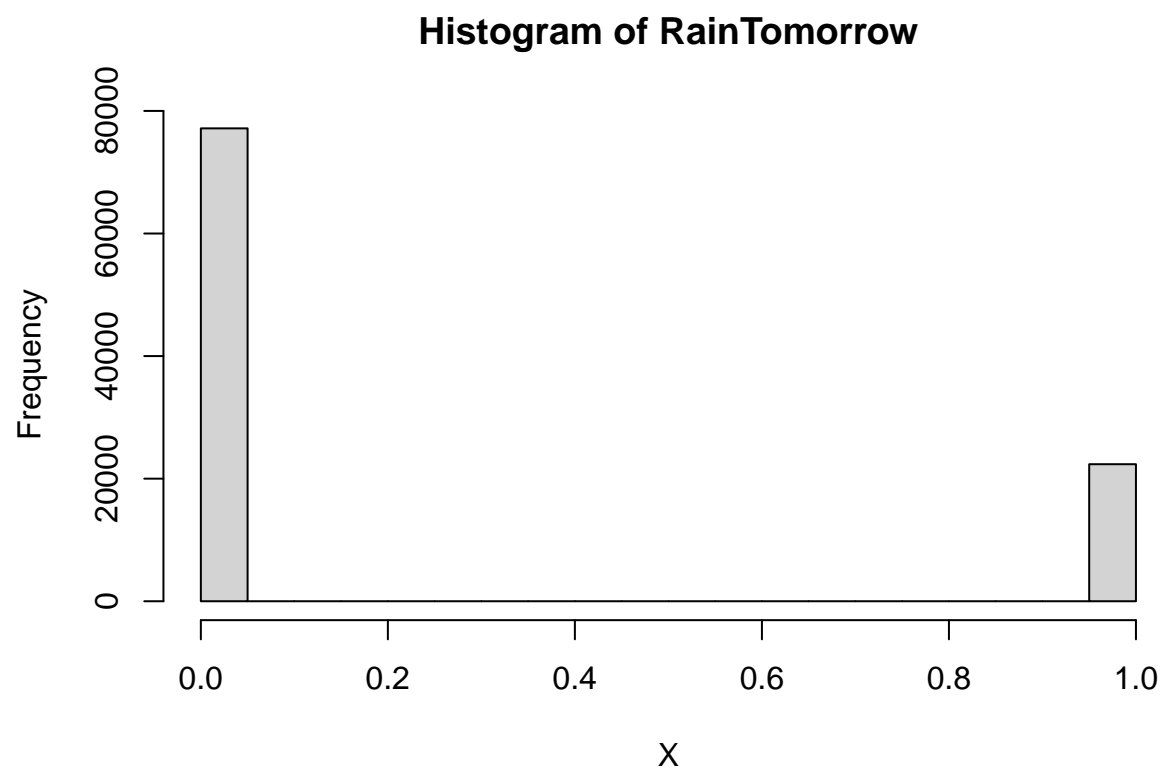
Boxplot of RainTomorrow
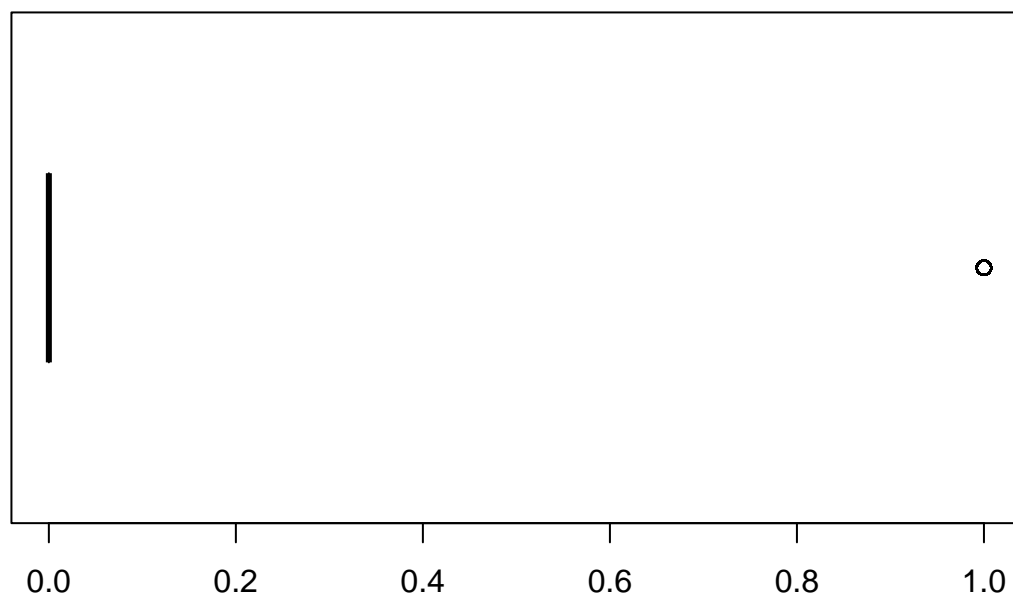

```
## [1] "Extended Summary Statistics"
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.2247  0.0000  1.0000
## [1] "sd:  0.417371821953755"
## [1] "vc:  1.85764901084798"
```

```r
par(ask=FALSE)
```