Machine Learning Assignment 1

ManojKumar Chalamala

10/28/2019

```
## R Script for Assignment 1 Machine Learning

#Downloaded the dataset from Web
#Source: https://vincentarelbundock.github.io/Rdatasets/datasets.html
#File: worldRecords.csv

#Import the dataset
WorldRecords <- read.csv(file="C://Users/Raahitha/Desktop/worldRecords.csv", header = T)

#View the data
head(WorldRecords)
```

```
##   X Distance roadORtrack        Place      Time      Date
## 1 1     0.10       track       Athens 0.1628333 6/14/2005
## 2 2     0.15       track      Cassino 0.2466667 5/22/1983
## 3 3     0.20       track      Atlanta 0.3220000  8/1/1996
## 4 4     0.30       track     Pretoria 0.5141667 3/24/2000
## 5 5     0.40       track      Seville 0.7196667 8/26/1999
## 6 6     0.50       track Busto Arsizio 1.0013333 5/26/1984
```

```
#Summary of the complete data
summary(WorldRecords)
```

```
##        X             Distance        roadORtrack         Place
##  Min.   : 1.00   Min.   :  0.100   road :16    London      : 3
##  1st Qu.:10.75   1st Qu.:  1.582   track:24    Christchurch: 2
##  Median :20.50   Median : 10.000               La Fleche   : 2
##  Mean   :20.50   Mean   : 32.113               Rieti       : 2
##  3rd Qu.:30.25   3rd Qu.: 30.000               Rome        : 2
##  Max.   :40.00   Max.   :290.221               Tilburg     : 2
##                                                (Other)     :27
##       Time                  Date
##  Min.   :   0.1628   3/22/1981 : 2
##  1st Qu.:   3.6475   3/30/1991 : 2
##  Median :  26.6627   9/4/2005  : 2
##  Mean   : 125.1796   1/15/2006 : 1
##  3rd Qu.:  88.3283   10/16/1982: 1
##  Max.   :1440.0000   10/20/1985: 1
##                      (Other)   :31
```

```
#Descriptive statistics of the Distance travelled
summary(WorldRecords$Distance)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.100   1.582  10.000  32.113  30.000 290.221
```

```
#Import Place column values into separate object
Place1 <- WorldRecords[4]
Place1
```

```
##                          Place
## 1                        Athens
## 2                       Cassino
## 3                       Atlanta
## 4                      Pretoria
## 5                       Seville
## 6                 Busto Arsizio
## 7                  Santa Monica
## 8                       Cologne
## 9                         Rieti
## 10                         Rome
## 11                         Rome
## 12                       Berlin
## 13                        Rieti
## 14                      Hechtel
## 15                      Hengelo
## 16                     Brussels
## 17                    La Fleche
## 18                    La Fleche
## 19                  Christchurch
## 20                  Christchurch
## 21                    Timperley
## 22                       London
## 23                       London
## 24                       London
## 25                      Carlsbad
## 26                    Groningen
## 27                       Dublin
## 28                  New Orleans
## 29                         Doha
## 30                   Evansville
## 31                      Tilburg
## 32                      Tilburg
## 33                 Phoenix-Tempe
## 34          Alphen aan den Rijn
## 35                     Kumamoto
## 36                      Chicago
## 37       Claremont-Kirstenbosch
## 38                  Lake Saroma
## 39                     New York
## 40                        Basle
```

```
#Descriptive statistics of the Place Column
summary(Place1)
```

```
##            Place
##  London     : 3
##  Christchurch: 2
##  La Fleche  : 2
##  Rieti      : 2
##  Rome       : 2
##  Tilburg    : 2
##  (Other)    :27
```

```
#Descriptive Statistics of the Time Taken
summary(WorldRecords$Time)
```

```
##     Min.   1st Qu.   Median     Mean   3rd Qu.      Max.
##    0.1628    3.6475   26.6628  125.1796   88.3283 1440.0000
```

```
#Descriptive Statistics of roadORtrack
summary(WorldRecords$roadORtrack)
```

```
##  road track
##    16    24
```

```
#Data Transformation
#Transform the Distance column by taking the squareroot of the values
Record1 <- sqrt(WorldRecords$Distance)
Record1
```

```
##  [1]  0.3162278  0.3872983  0.4472136  0.5477226  0.6324555  0.7071068
##  [7]  0.7745967  0.8944272  1.0000000  1.2247449  1.2684636  1.4142136
## [13]  1.7320508  1.7938785  2.2360680  3.1622777  3.8729833  4.4721360
## [19]  5.0000000  5.4772256  7.0710678  8.0224684 10.0000000 12.6846364
## [25]  2.2360680  2.5369273  2.8284271  2.8363709  3.1622777  3.4641016
## [31]  3.8729833  4.0112342  4.4721360  5.0000000  5.4772256  5.6727418
## [37]  7.0710678 10.0000000 12.6846364 17.0358739
```

```
#Transform the Time Column by rasing the power to 3
Record2 <- (WorldRecords$Time)^3
Record2
```

```
##  [1] 4.317476e-03 1.500830e-02 3.338625e-02 1.359289e-01 3.727298e-01
##  [6] 1.004005e+00 1.786979e+00 4.785514e+00 1.063832e+01 4.047137e+01
## [11] 5.143043e+01 1.069351e+02 3.961747e+02 5.075649e+02 2.011111e+03
## [16] 1.817520e+04 7.712744e+04 1.844791e+05 4.040751e+05 7.124410e+05
## [21] 4.750104e+06 1.194356e+07 5.079002e+07 3.257317e+08 2.197000e+03
## [26] 5.058921e+03 1.069647e+04 1.076946e+04 1.975599e+04 3.765152e+04
## [31] 7.078669e+04 8.752838e+04 1.737411e+05 3.673181e+05 6.814720e+05
## [36] 8.673409e+05 4.381424e+06 5.212502e+07 3.528187e+08 2.985984e+09
```

```r
#Transform the Distance column by taking the base 10 logarithm of Distance
Record3 <- log10(WorldRecords$Distance)
Record3
```

```
##  [1] -1.00000000 -0.82390874 -0.69897000 -0.52287875 -0.39794001
##  [6] -0.30103000 -0.22184875 -0.09691001  0.00000000  0.17609126
## [11]  0.20655604  0.30103000  0.47712125  0.50758604  0.69897000
## [16]  1.00000000  1.17609126  1.30103000  1.39794001  1.47712125
## [21]  1.69897000  1.80861604  2.00000000  2.20655604  0.69897000
## [26]  0.80861604  0.90308999  0.90552605  1.00000000  1.07918125
## [31]  1.17609126  1.20655604  1.30103000  1.39794001  1.47712125
## [36]  1.50758604  1.69897000  2.00000000  2.20655604  2.46272883
```

```r
#Conversion from one data type to another
is.numeric(WorldRecords$Distance)
```

```
## [1] TRUE
```
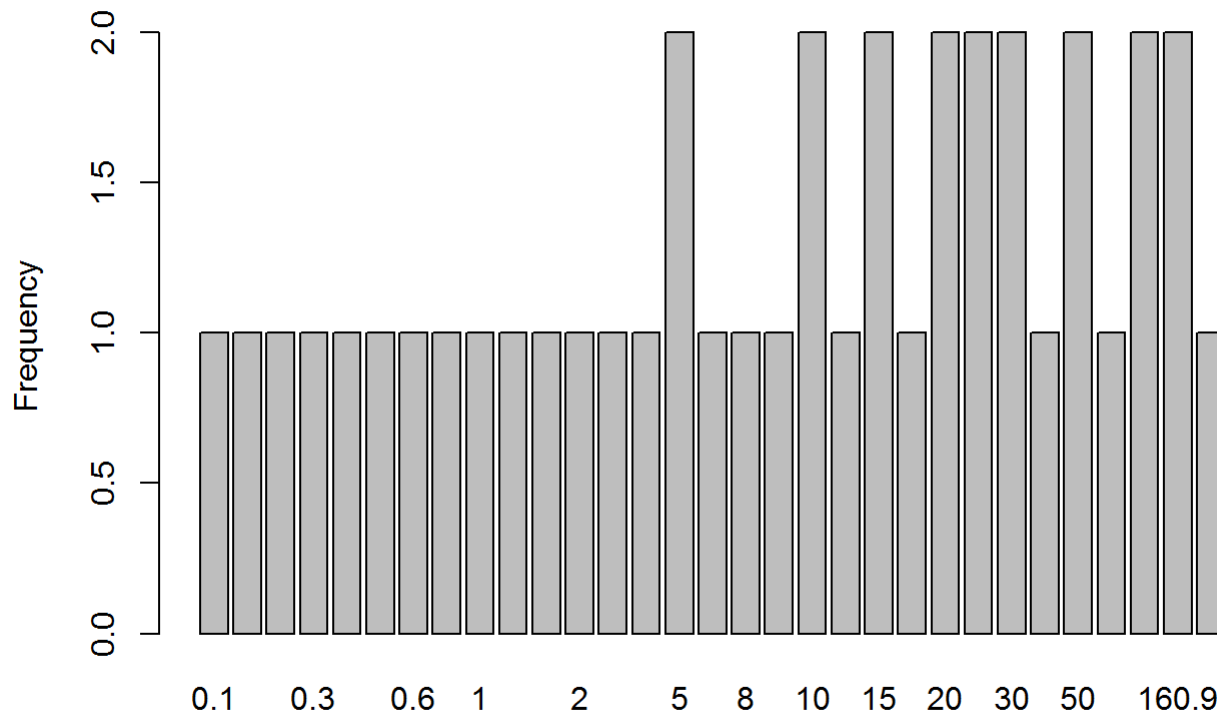
```r
DisChar <- as.character(WorldRecords$Distance)
is.character(DisChar)
```

```
## [1] TRUE
```
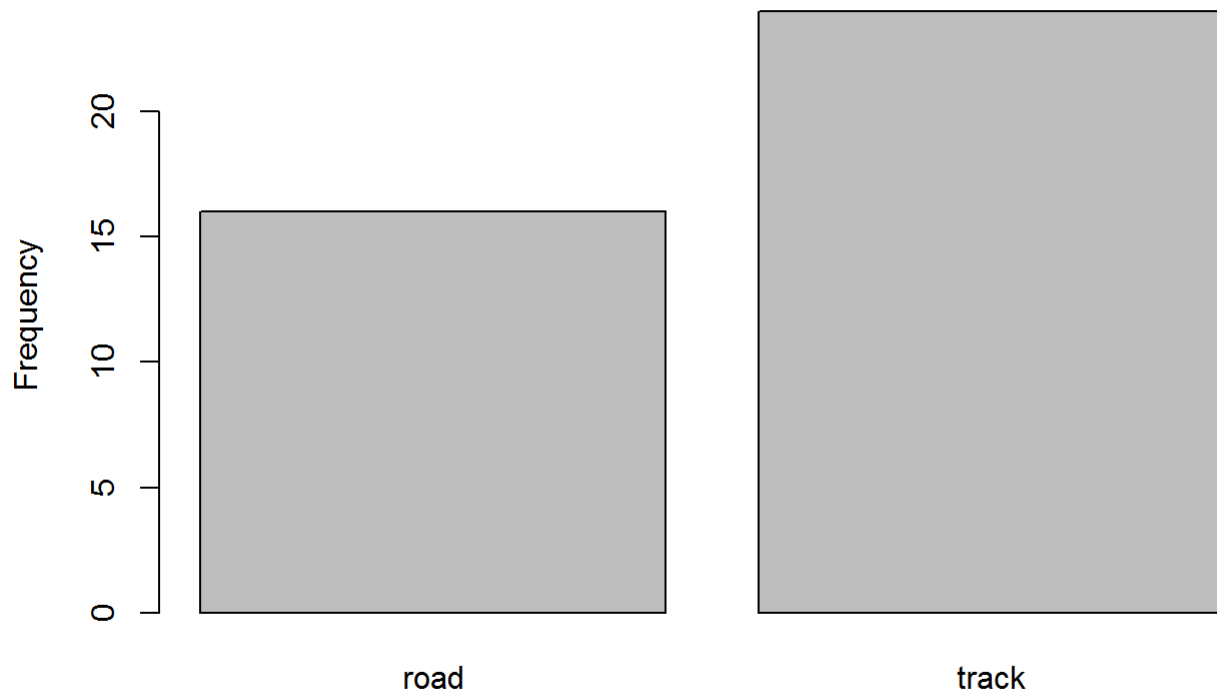
```r
#Plot the Data for Distance Travelled
barplot(table(WorldRecords$Distance), ylab='Frequency', main='Distribution of Distance Travelle
d')
```

Distribution of Distance Travelled


```
#Plot the data for RoadORTrack
barplot(table(WorldRecords$roadORtrack), ylab='Frequency', main='Barplot of roadORtrack')
```

Barplot of roadORtrack


```
#Scatterplot the data for Distance travelled by Time taken
plot(WorldRecords$Distance, WorldRecords$Time, xlab='Distance Travelled', ylab='Time Taken', mai
n='Distance travelled by time', xlim=c(0,60), ylim=c(0,100), pch=10, col='blue')
points(WorldRecords$Distance[WorldRecords$Place=='Christchurch'], WorldRecords$Time[WorldRecords
$Place=='Christchurch'], pch=20, col='red')
```

Distance travelled by time