**INSTRUCTIONS:**

**NORMALIZATION:**

**Min-Max:**

min\_max<- function(x)

{

**#it transforms the data in to minimum value**

return ((x - min(x)) / (max(x) - min(x)))

}

print(minmax)

plot(quakes$depth)

plot(minmax)

**Z-Score:**

**#applying the scale function to normalize the data in to smaller values**

zscore= as.data.frame( scale(trans["depth"] ))

plot(zscore)

**Decimal Scaling:**

normalize.decscale<- function (data)

{

**#decimal scaling to a matrix or dataframe. Decimal scaling transforms the**

**#data into [-1,1] by finding k such that the absolute value of the maximum**

**#value of each attribute divided by 10^k is less than or equal to 1.**

maxvect <- apply(abs(data), 2, max)

kvector <- ceiling(log10(maxvect))

scalefactor <- 10^kvector

scale(data, center = FALSE, scale = scalefactor)

}

print(normalize.decscale(trans))

plot(normalize.decscale(trans))

**#loading the builtin dataset**

trans=head(quakes)

**#applying the lapply function to normalize the data in to smaller values**

minmax = as.data.frame(lapply(trans["depth"], min\_max))

**DISCRETIZATION BY BINNING**

bins<-5

minimumVal<-min(quakes$depth)

maximumVal<-max(quakes$depth)

width=(maximumVal-minimumVal)/bins;

x<-cut(quakes$depth,breaks=seq(minimumVal, maximumVal, width))

x

barplot(quakes$depth)

barplot(table(cut(quakes$depth, breaks=seq(minimumVal, maximumVal, width))))

**DISCRETIZATION BY CLUSTERING**

plot(depth~stations,quakes)

**#Normalization**

z<-quakes[,-c(1,1)]

m<-apply(z,2,mean)

s<-apply(z,2,sd)

z<-scale(z,m,s)

**#Calculate Euclidean distance**

distance<-dist(z)

print(distance,digits = 3)

**#Cluster Dendrogram with complete linkage**

hc.c<-hclust(distance)

plot(hc.c,hang=-1)

**#Cluster Dendrogram with Average linkage**

hc.a<-hclust(distance,method="average")

plot(hc.a,hang=-1)

**DISCRETIZATION BY DECISION TREE**

data<-quakes

str(data)

data$NSPF <- factor(data$stations)

**#Partion data into Training and Validation Datasets**

set.seed(1234)

pd<-sample(2,nrow(data),replace=TRUE,prob=c(0.8,0.2))

train<-data[pd==1,]

validate<-data[pd==2,]

**#Decision Tree with Party**

library(party)

tree<-ctree(NSPF~lat+long+mag,data=train,controls = ctree\_control(mincriterion = 0.99,minsplit = 500))

tree

plot(tree)

**#Decision Tree with rPart**

library(rpart)

tree1<-rpart(NSPF~lat+long+mag,train)

library(rpart.plot)

rpart.plot(tree1,extra=2)

**DISCRETIZATION BY CORRLATION**

cor.test(quakes$depth,quakes$stations)

**OUTPUT:**

**NORMALIZATION:**

**Min-Max:**

depth

1 0.8552632

2 1.0000000

3 0.0000000

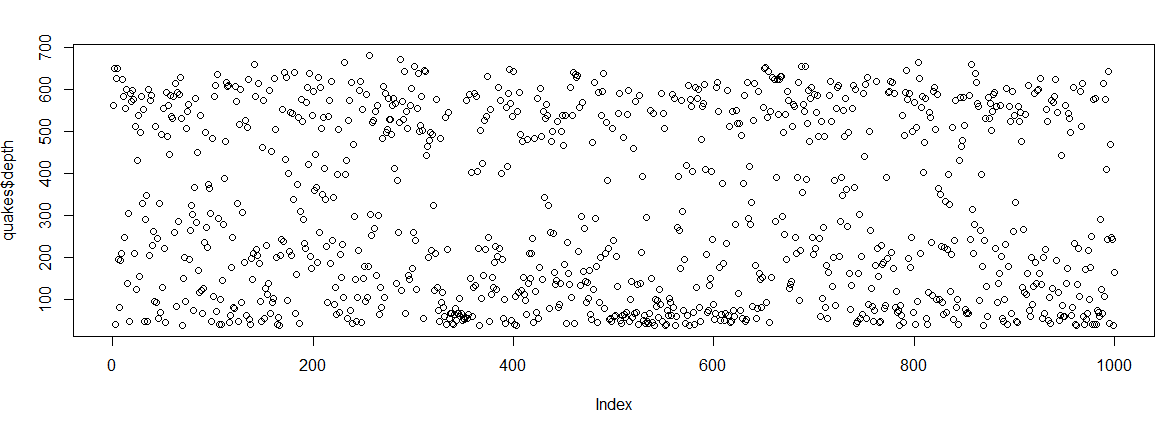
4 0.9605263

5 0.9983553

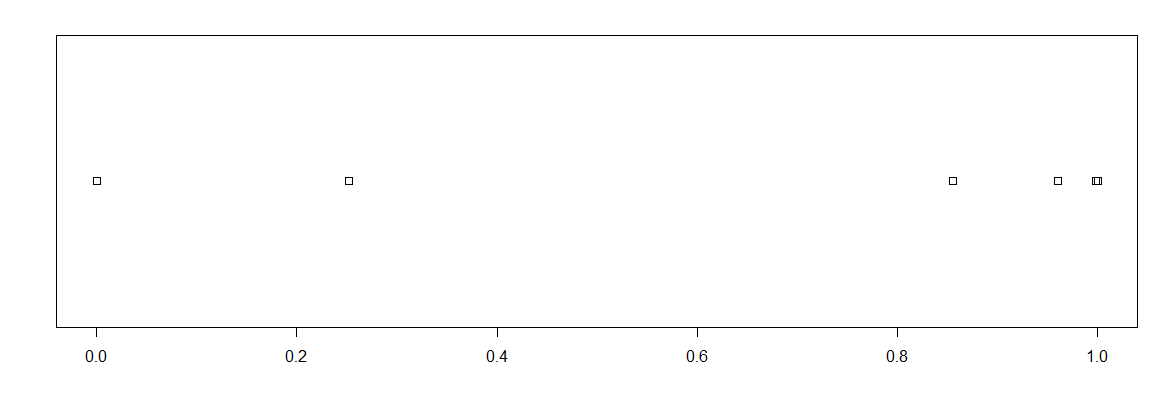
6 0.2516447

**PLOT**

**Before:**

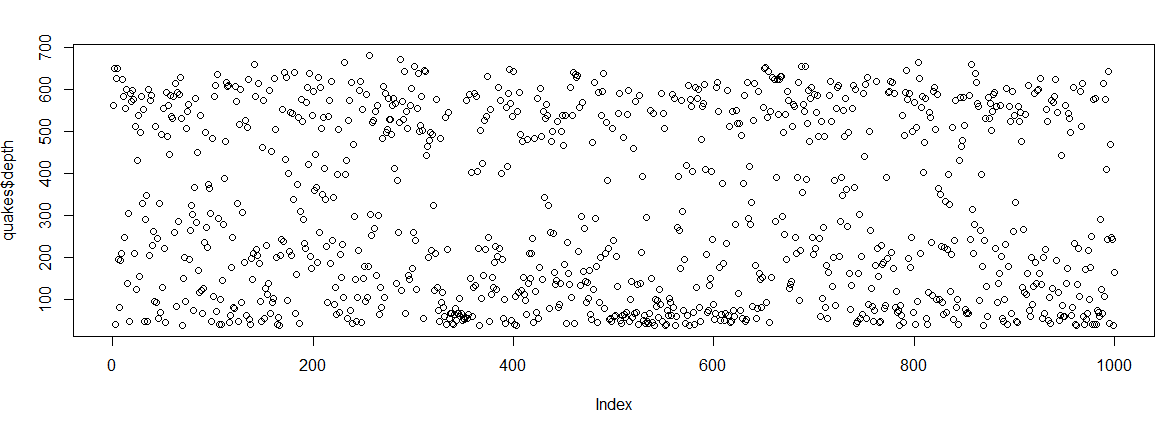


**After:**

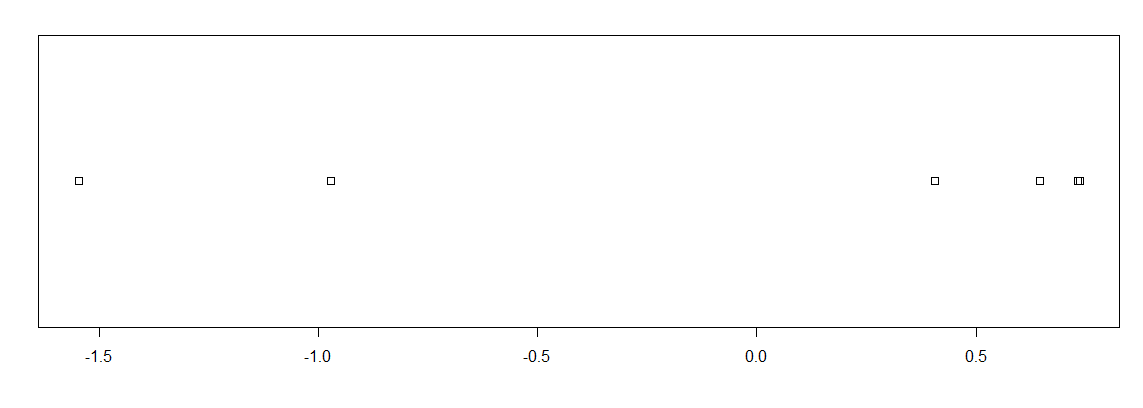
****

**Z-Score:**

**Before:**



**After:**

****

**Decimal Scaling:**

lat long depth mag stations

1 -0.2042 0.18162 0.562 0.48 0.41

2 -0.2062 0.18103 0.650 0.42 0.15

3 -0.2600 0.18410 0.042 0.54 0.43

4 -0.1797 0.18166 0.626 0.41 0.19

5 -0.2042 0.18196 0.649 0.40 0.11

6 -0.1968 0.18431 0.195 0.40 0.12

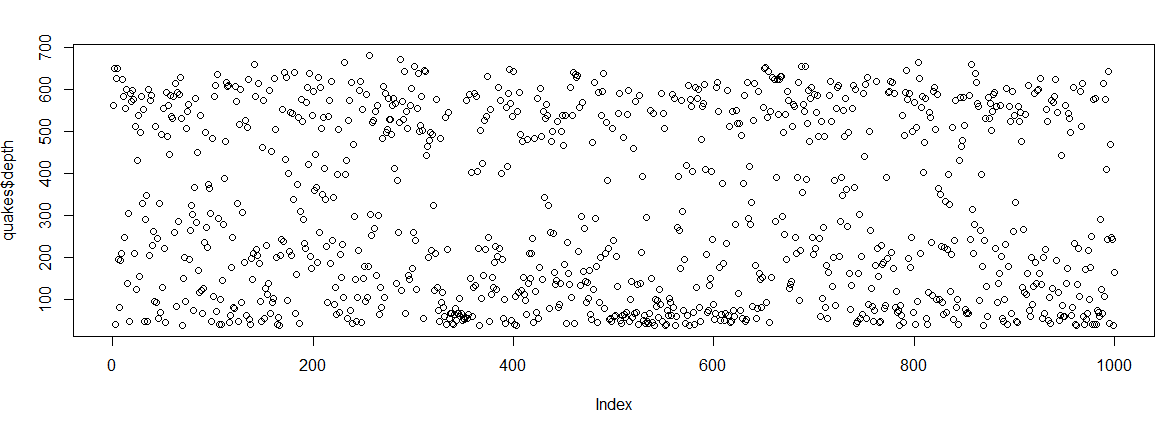
attr(,"scaled:scale")

lat long depth mag stations

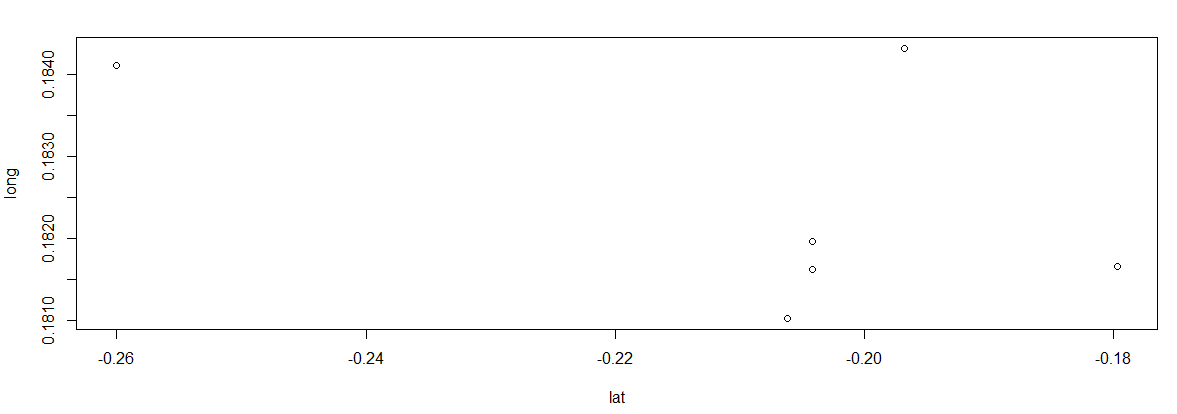
100 1000 1000 10 100

scale(data, center = FALSE, scale = scalefactor)

**Before:**



**After:**

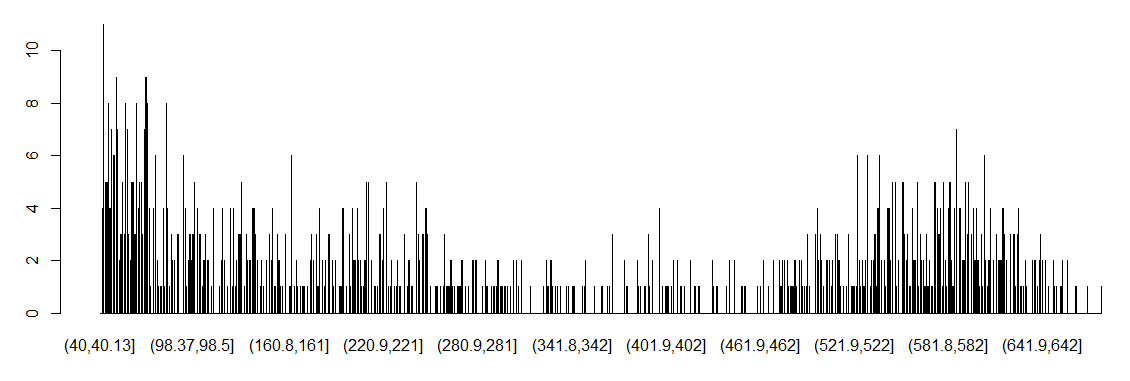


**DISCRETIZATION BY BINNING**

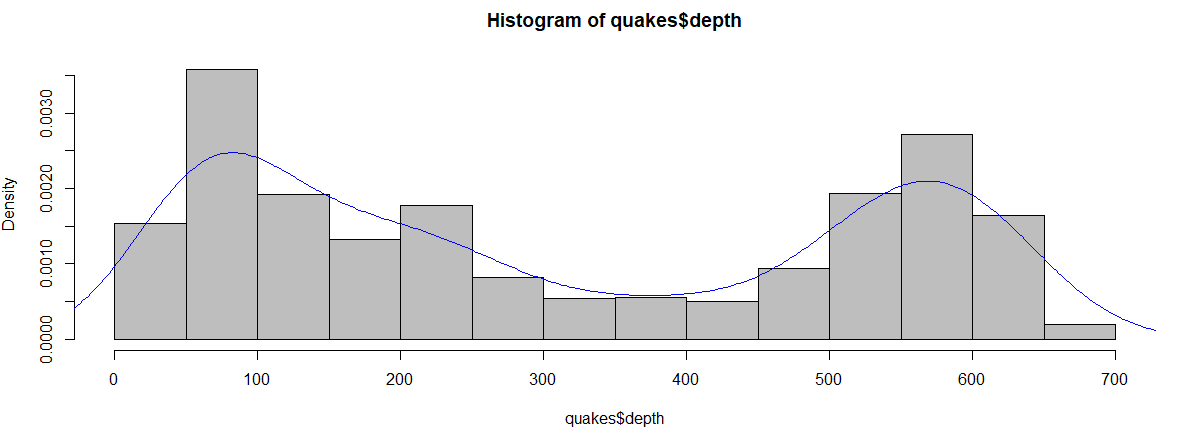
**Before:**

****

**After:**

****

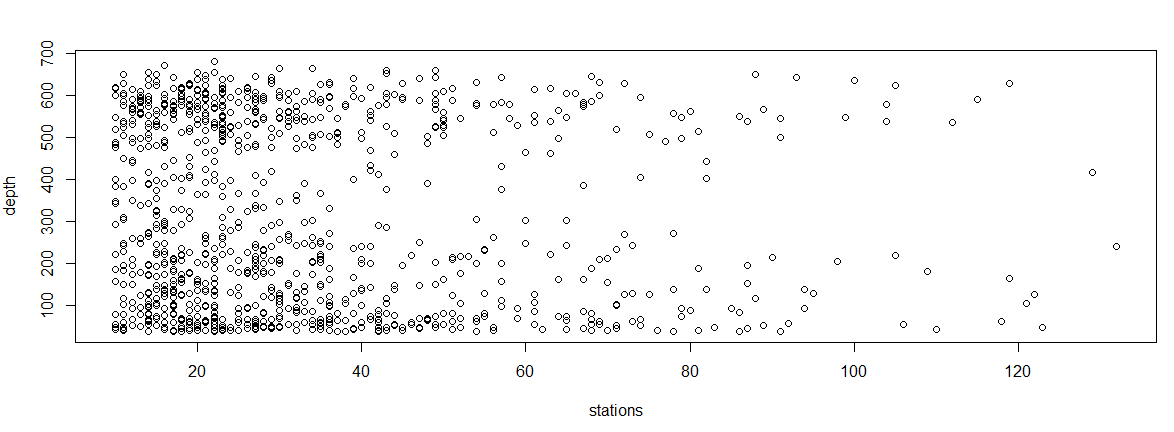
**DISCRETIZATION BY HISTOGRAM ANALYSIS**

****

**DISCRETIZATION BY CLUSTERING**

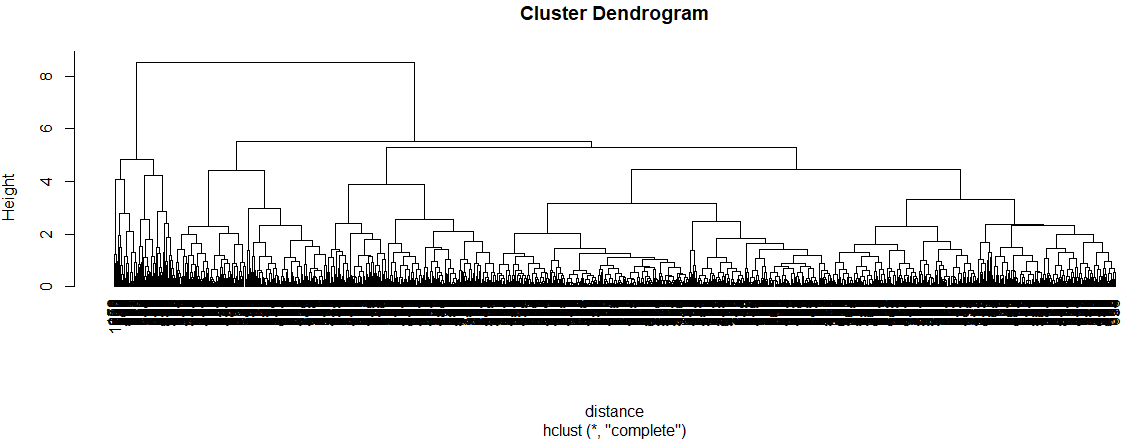
**PLOT**

**Before:**

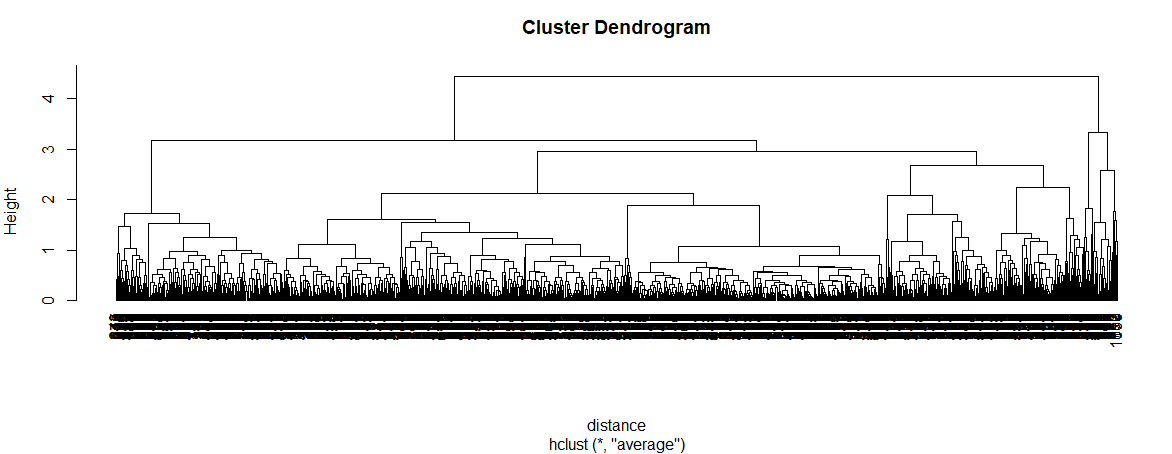
****

**After:**

**Complete Clustering:**

****

**Average Clustering:**

****

**DISCRETIZATION BY DECISION TREE**

**Tree:**

Conditional inference tree with 7 terminal nodes

Response: NSPF

Inputs: lat, long, mag

Number of observations: 792

1) mag <= 5.6; criterion = 1, statistic = 611.093

2) mag <= 4.2; criterion = 1, statistic = 562.84

3)\* weights = 149

2) mag > 4.2

4) mag <= 5.2; criterion = 1, statistic = 455.569

5) mag <= 5.1; criterion = 1, statistic = 352.735

6) mag <= 5; criterion = 1, statistic = 298.307

7) mag <= 4.9; criterion = 1, statistic = 232.966

8)\* weights = 483

7) mag > 4.9

9)\* weights = 38

6) mag > 5

10)\* weights = 33

5) mag > 5.1

11)\* weights = 25

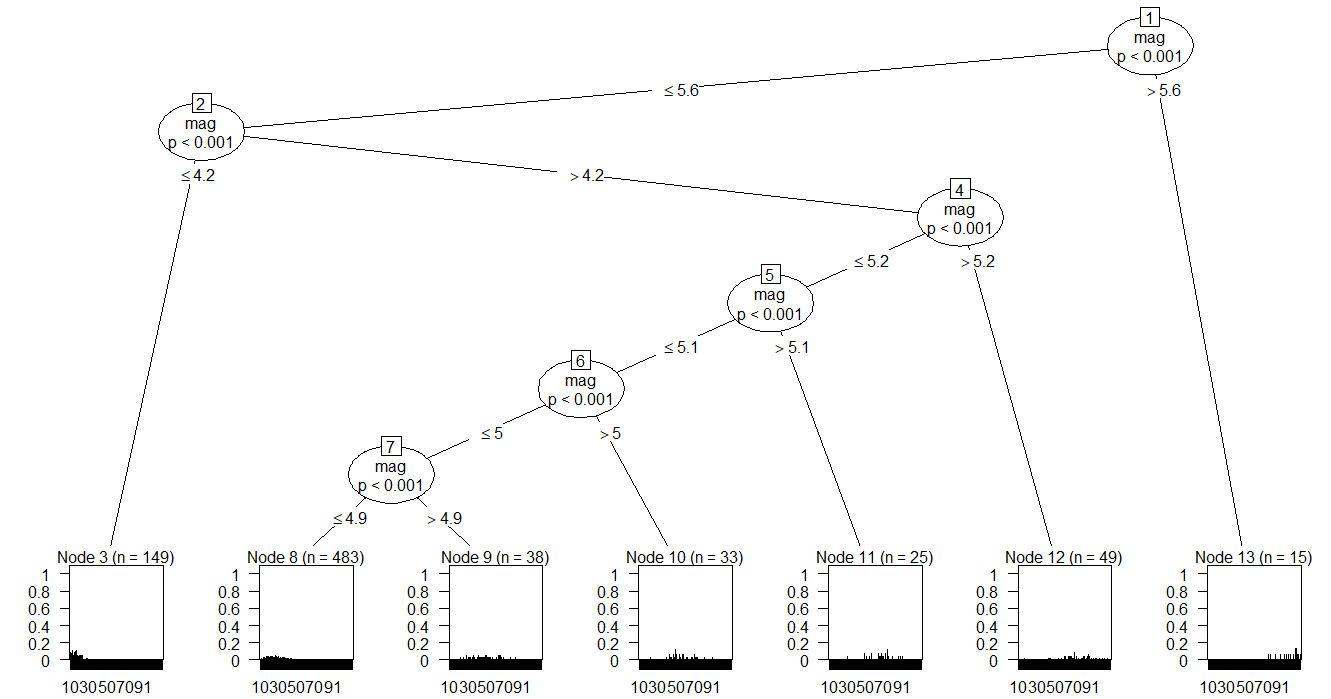
4) mag > 5.2

12)\* weights = 49

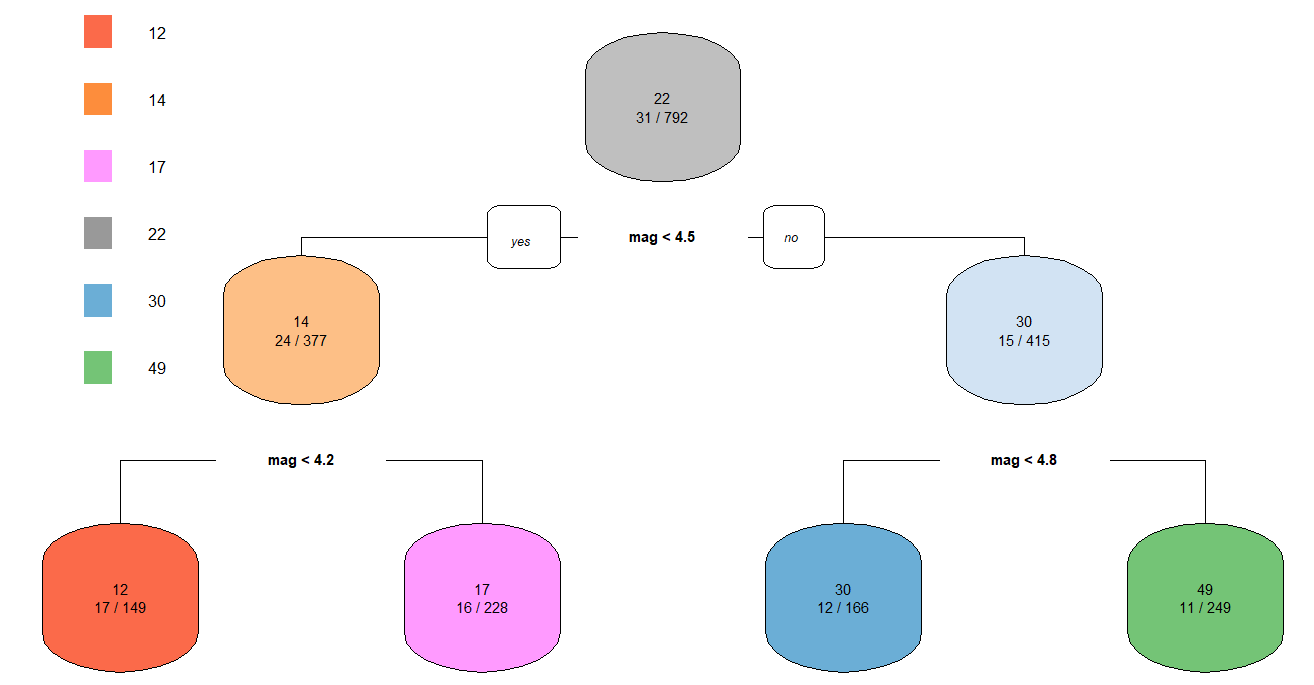
1) mag > 5.6

13)\* weights = 15

**Tree With Party Package:**

****

**Tree With rpart Package:**

****

**DISCRETIZATION BY CORRELATION**

Pearson's product-moment correlation

data: quakes$depth and quakes$stations

t = -2.3287, df = 998, p-value = 0.02007

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.13489341 -0.01157477

sample estimates:

cor

-0.0735151