> library(Hmisc)

> library(lattice)

>

> #START OF DATA IMPORT

>

> CTG=read.csv("C:/Users/jgozal1/Desktop/Data Mining Project/CTG-raw.csv")

> str(CTG)

'data.frame': 2126 obs. of 40 variables:

$ FileName: Factor w/ 352 levels "Aast\_1.txt","Aast\_10.txt",..: 352 69 69 69 69 70 70 75 75 75 ...

$ Date : Factor w/ 48 levels "1/2/1997","1/4/1980",..: 13 31 31 31 31 31 31 18 18 18 ...

$ SegFile : Factor w/ 2126 levels "CTG0001.txt",..: 1 2 3 4 5 6 7 8 9 10 ...

$ b : int 240 5 177 411 533 0 240 62 120 181 ...

$ e : int 357 632 779 1192 1147 953 953 679 779 1192 ...

$ LBE : int 120 132 133 134 132 134 134 122 122 122 ...

$ LB : int 120 132 133 134 132 134 134 122 122 122 ...

$ AC : int 0 4 2 2 4 1 1 0 0 0 ...

$ FM : int 0 0 0 0 0 0 0 0 0 0 ...

$ UC : int 0 4 5 6 5 10 9 0 1 3 ...

$ ASTV : int 73 17 16 16 16 26 29 83 84 86 ...

$ MSTV : num 0.5 2.1 2.1 2.4 2.4 5.9 6.3 0.5 0.5 0.3 ...

$ ALTV : int 43 0 0 0 0 0 0 6 5 6 ...

$ MLTV : num 2.4 10.4 13.4 23 19.9 0 0 15.6 13.6 10.6 ...

$ DL : int 0 2 2 2 0 9 6 0 0 0 ...

$ DS : int 0 0 0 0 0 0 0 0 0 0 ...

$ DP : int 0 0 0 0 0 2 2 0 0 0 ...

$ DR : int 0 0 0 0 0 0 0 0 0 0 ...

$ Width : int 64 130 130 117 117 150 150 68 68 68 ...

$ Min : int 62 68 68 53 53 50 50 62 62 62 ...

$ Max : int 126 198 198 170 170 200 200 130 130 130 ...

$ Nmax : int 2 6 5 11 9 5 6 0 0 1 ...

$ Nzeros : int 0 1 1 0 0 3 3 0 0 0 ...

$ Mode : int 120 141 141 137 137 76 71 122 122 122 ...

$ Mean : int 137 136 135 134 136 107 107 122 122 122 ...

$ Median : int 121 140 138 137 138 107 106 123 123 123 ...

$ Variance: int 73 12 13 13 11 170 215 3 3 1 ...

$ Tendency: int 1 0 0 1 1 0 0 1 1 1 ...

$ A : int 0 0 0 0 0 0 0 0 0 0 ...

$ B : int 0 0 0 0 1 0 0 0 0 0 ...

$ C : int 0 0 0 0 0 0 0 0 0 0 ...

$ D : int 0 0 0 0 0 0 0 0 0 0 ...

$ E : int 0 0 0 0 0 0 0 0 0 0 ...

$ AD : int 0 1 1 1 0 0 0 0 0 0 ...

$ DE : int 0 0 0 0 0 0 0 0 0 0 ...

$ LD : int 0 0 0 0 0 1 1 0 0 0 ...

$ FS : int 1 0 0 0 0 0 0 1 1 1 ...

$ SUSP : int 0 0 0 0 0 0 0 0 0 0 ...

$ CLASS : int 9 6 6 6 2 8 8 9 9 9 ...

$ NSP : int 2 1 1 1 1 3 3 3 3 3 ...

>

> #END OF DATA IMPORT

> describe(CTG)

CTG

40 Variables 2126 Observations

----------------------------------------------------------------------------------------------------------------------

FileName

n missing unique

2126 0 352

lowest : Aast\_1.txt Aast\_10.txt Aast\_11.txt Aast\_12.txt Aast\_13.txt

highest: Tffm\_2.txt Tffm\_3.txt Tffm\_4.txt Tffm\_5.txt Variab10.txt

----------------------------------------------------------------------------------------------------------------------

Date

n missing unique

2126 0 48

lowest : 1/2/1997 1/4/1980 1/8/1995 10/1/1997 10/14/1996

highest: 9/18/1996 9/20/1996 9/23/1996 9/25/1996 9/26/1996

----------------------------------------------------------------------------------------------------------------------

SegFile

n missing unique

2126 0 2126

lowest : CTG0001.txt CTG0002.txt CTG0003.txt CTG0004.txt CTG0005.txt

highest: CTG2124.txt CTG2125.txt CTG2126.txt CTG2127.txt CTG2128.txt

----------------------------------------------------------------------------------------------------------------------

b

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 979 1 878.4 0 0 55 538 1521 2304 2654

lowest : 0 1 2 3 4, highest: 3188 3194 3199 3274 3296

----------------------------------------------------------------------------------------------------------------------

e

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 1064 1 1703 592 697 1009 1241 2435 3241 3538

lowest : 287 301 307 321 346, highest: 3591 3593 3595 3597 3599

----------------------------------------------------------------------------------------------------------------------

LBE

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 48 1 133.3 119 121 126 133 140 146 149

lowest : 106 110 112 114 115, highest: 156 157 158 159 160

----------------------------------------------------------------------------------------------------------------------

LB

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 48 1 133.3 119 121 126 133 140 146 149

lowest : 106 110 112 114 115, highest: 156 157 158 159 160

----------------------------------------------------------------------------------------------------------------------

AC

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 22 0.92 2.722 0 0 0 1 4 8 10

lowest : 0 1 2 3 4, highest: 17 18 19 21 26

----------------------------------------------------------------------------------------------------------------------

FM

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 96 0.76 7.241 0 0 0 0 2 9 22

lowest : 0 1 2 3 4, highest: 325 443 489 557 564

----------------------------------------------------------------------------------------------------------------------

UC

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 19 0.99 3.66 0 0 1 3 5 7 9

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 23

Frequency 332 238 236 294 272 235 199 114 82 57 29 16 9 6 3 1 1 1 1

% 16 11 11 14 13 11 9 5 4 3 1 1 0 0 0 0 0 0 0

----------------------------------------------------------------------------------------------------------------------

ASTV

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 75 1 46.99 21 23 32 49 61 68 75

lowest : 12 13 14 15 16, highest: 82 83 84 86 87

----------------------------------------------------------------------------------------------------------------------

MSTV

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 57 1 1.333 0.3 0.4 0.7 1.2 1.7 2.5 3.0

lowest : 0.2 0.3 0.4 0.5 0.6, highest: 5.9 6.0 6.3 6.9 7.0

----------------------------------------------------------------------------------------------------------------------

ALTV

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 87 0.8 9.847 0 0 0 0 11 38 56

lowest : 0 1 2 3 4, highest: 85 86 88 90 91

----------------------------------------------------------------------------------------------------------------------

MLTV

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 249 1 8.188 0.00 1.30 4.60 7.40 10.80 15.00 18.48

lowest : 0.0 0.1 0.2 0.3 0.4, highest: 35.7 36.9 40.8 41.8 50.7

----------------------------------------------------------------------------------------------------------------------

DL

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 15 0.8 1.57 0 0 0 0 3 5 7

0 1 2 3 4 5 6 7 8 9 10 11 12 14 16

Frequency 1231 232 127 112 123 108 67 45 28 25 7 12 6 2 1

% 58 11 6 5 6 5 3 2 1 1 0 1 0 0 0

----------------------------------------------------------------------------------------------------------------------

DS

n missing unique Info Sum Mean

2126 0 2 0.01 7 0.003293

----------------------------------------------------------------------------------------------------------------------

DP

n missing unique Info Mean

2126 0 5 0.23 0.1261

0 1 2 3 4

Frequency 1948 109 49 19 1

% 92 5 2 1 0

----------------------------------------------------------------------------------------------------------------------

DR

n missing unique Info Mean

2126 0 1 0 0

----------------------------------------------------------------------------------------------------------------------

Width

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 154 1 70.45 16.0 22.0 37.0 67.5 100.0 126.0 138.0

lowest : 3 5 6 7 8, highest: 161 162 163 176 180

----------------------------------------------------------------------------------------------------------------------

Min

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 109 1 93.58 51 54 67 93 120 133 139

lowest : 50 51 52 53 54, highest: 154 155 156 158 159

----------------------------------------------------------------------------------------------------------------------

Max

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 86 1 164 138 143 152 162 174 189 198

lowest : 122 123 125 126 127, highest: 211 213 228 230 238

----------------------------------------------------------------------------------------------------------------------

Nmax

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 18 0.99 4.068 0.25 1.00 2.00 3.00 6.00 8.00 10.00

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 18

Frequency 107 357 331 269 258 210 158 145 106 67 49 28 22 10 5 1 2 1

% 5 17 16 13 12 10 7 7 5 3 2 1 1 0 0 0 0 0

----------------------------------------------------------------------------------------------------------------------

Nzeros

n missing unique Info Mean

2126 0 9 0.55 0.3236

0 1 2 3 4 5 7 8 10

Frequency 1624 366 108 21 2 2 1 1 1

% 76 17 5 1 0 0 0 0 0

----------------------------------------------------------------------------------------------------------------------

Mode

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 88 1 137.5 111.2 120.0 129.0 139.0 148.0 156.0 160.0

lowest : 60 67 69 71 75, highest: 176 179 180 186 187

----------------------------------------------------------------------------------------------------------------------

Mean

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 103 1 134.6 108 115 125 136 145 153 157

lowest : 73 75 76 78 79, highest: 173 175 178 180 182

----------------------------------------------------------------------------------------------------------------------

Median

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 95 1 138.1 113 120 129 139 148 155 159

lowest : 77 78 79 82 86, highest: 177 178 180 183 186

----------------------------------------------------------------------------------------------------------------------

Variance

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 133 1 18.81 0 1 2 7 24 53 76

lowest : 0 1 2 3 4, highest: 241 243 250 254 269

----------------------------------------------------------------------------------------------------------------------

Tendency

n missing unique Info Mean

2126 0 3 0.79 0.3203

-1 (165, 8%), 0 (1115, 52%), 1 (846, 40%)

----------------------------------------------------------------------------------------------------------------------

A

n missing unique Info Sum Mean

2126 0 2 0.44 384 0.1806

----------------------------------------------------------------------------------------------------------------------

B

n missing unique Info Sum Mean

2126 0 2 0.59 579 0.2723

----------------------------------------------------------------------------------------------------------------------

C

n missing unique Info Sum Mean

2126 0 2 0.07 53 0.02493

----------------------------------------------------------------------------------------------------------------------

D

n missing unique Info Sum Mean

2126 0 2 0.11 81 0.0381

----------------------------------------------------------------------------------------------------------------------

E

n missing unique Info Sum Mean

2126 0 2 0.1 72 0.03387

----------------------------------------------------------------------------------------------------------------------

AD

n missing unique Info Sum Mean

2126 0 2 0.4 332 0.1562

----------------------------------------------------------------------------------------------------------------------

DE

n missing unique Info Sum Mean

2126 0 2 0.31 252 0.1185

----------------------------------------------------------------------------------------------------------------------

LD

n missing unique Info Sum Mean

2126 0 2 0.14 107 0.05033

----------------------------------------------------------------------------------------------------------------------

FS

n missing unique Info Sum Mean

2126 0 2 0.09 69 0.03246

----------------------------------------------------------------------------------------------------------------------

SUSP

n missing unique Info Sum Mean

2126 0 2 0.25 197 0.09266

----------------------------------------------------------------------------------------------------------------------

CLASS

n missing unique Info Mean .05 .10 .25 .50 .75 .90 .95

2126 0 10 0.97 4.51 1 1 2 4 7 9 10

1 2 3 4 5 6 7 8 9 10

Frequency 384 579 53 81 72 332 252 107 69 197

% 18 27 2 4 3 16 12 5 3 9

----------------------------------------------------------------------------------------------------------------------

NSP

n missing unique Info Mean

2126 0 3 0.53 1.304

1 (1655, 78%), 2 (295, 14%), 3 (176, 8%)

----------------------------------------------------------------------------------------------------------------------