```
PROC IMPORT OUT= WORK.S26
            DATAFILE= "C:\S26.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;
```

```
NOTE: WORK.S26 data set was successfully created.
NOTE: The data set WORK.S26 has 5399 observations and 28 variables.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time           3.09 seconds
      cpu time            0.78 seconds
```

```
data s26_1;
set  s26 ;
rand=ranuni(092765);
     if rand <=.7 then RespHoldout=.;
else if rand  >.7 then do;
   RespHoldout=Resp;
   Resp=.;
end;
run;
```

```
NOTE: There were 5399 observations read from the data set WORK.S26.
NOTE: The data set WORK.S26_1 has 5399 observations and 30 variables.
NOTE: DATA statement used (Total process time):
      real time           0.23 seconds
      cpu time            0.06 seconds
```

```
data s26_2;
set  s26_1;
array orig[11](0, 1, 2,  3,  4,  5,   6,   7,    8,    9, 10);
array new[11] (0,25,75,150,350,750,3000,7500,15000,30000, 30000);
retain orig1-orig11 new1-new11;
do i=1 to dim(orig);
 if PWAPAR=orig[i] then PWAPAR2=new[i];
 if PAANHA=orig[i] then PAANHA2=new[i];
 if PPERSA=orig[i] then PPERSA2=new[i];
end;
drop orig1--orig11 new1--new11 i;
run;
```

```
NOTE: There were 5399 observations read from the data set WORK.S26_1.
NOTE: The data set WORK.S26_2 has 5399 observations and 33 variables.
NOTE: DATA statement used (Total process time):
      real time           0.12 seconds
      cpu time            0.07 seconds
```

```
proc freq data=s26_2;
tables      PPERSA*PPERSA2
            PAANHA*PAANHA2
            PWAPAR*PWAPAR2/list;
run;
```

The SAS System

The FREQ Procedure

PPERSA	PPERSA2	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	0	5355	99.19	5355	99.19
5	750	9	0.17	5364	99.35
6	3000	32	0.59	5396	99.94
7	7500	3	0.06	5399	100.00

PAANHA	PAANHA2	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	0	5192	96.17	5192	96.17
3	150	2	0.04	5194	96.20
4	350	127	2.35	5321	98.56
5	750	28	0.52	5349	99.07
6	3000	48	0.89	5397	99.96
7	7500	2	0.04	5399	100.00

PWAPAR	PWAPAR2	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	0	5322	98.57	5322	98.57
1	25	7	0.13	5329	98.70
2	75	29	0.54	5358	99.24
3	150	21	0.39	5379	99.63
4	350	15	0.28	5394	99.91
5	750	1	0.02	5395	99.93
6	3000	4	0.07	5399	100.00

```
data s26_2;
set  s26_2;
drop PPERSA PAANHA PWAPAR;
run;
```

```
NOTE: There were 5399 observations read from the data set WORK.S26_2.
NOTE: The data set WORK.S26_2 has 5399 observations and 30 variables.
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      cpu time            0.03 seconds
```

```
data s26_3;
set  s26_2;
array orig[11](0,  1, 2, 3, 4, 5, 6, 7, 8,  9, 10);
array new[11] (0,5.5,17,30,43,56,69,82,94,100, 100);
retain orig1-orig11 new1-new11;
do i=1 to dim(orig);
if MSKB1 =orig[i] then MSKB12 =new[i];
if MAUT0 =orig[i] then MAUT02 =new[i];
if MHHUUR =orig[i] then MHHUUR2 =new[i];
if MAUT2 =orig[i] then MAUT22 =new[i];
if MINKGE =orig[i] then MINKGE2 =new[i];
if MFALLE =orig[i] then MFALLE2 =new[i];
if MRELGE =orig[i] then MRELGE2 =new[i];
if MGODRK   =orig[i] then MGODRK2 =new[i];
if MOPLHO  =orig[i] then MOPLHO2 =new[i];
if MFWEKI  =orig[i] then MFWEKI2 =new[i];
if MSKB2   =orig[i] then MSKB22 =new[i];
if MGODPR =orig[i] then MGODPR2 =new[i];
if MSKC  =orig[i] then MSKC2 =new[i];
if MAUT1  =orig[i] then MAUT12 =new[i];
if MSKA  =orig[i] then MSKA2 =new[i];
end;
drop orig1--orig11 new1--new11 i;
run;
```

```
NOTE: There were 5399 observations read from the data set WORK.S26_2.
NOTE: The data set WORK.S26_3 has 5399 observations and 45 variables.
NOTE: DATA statement used (Total process time):
      real time           0.12 seconds
      cpu time            0.12 seconds
```

```
proc freq data=s26_3;
tables
MGODRK*MGODRK2
MGODPR*MGODPR2
MRELGE*MRELGE2
MFALLE*MFALLE2
MFWEKI*MFWEKI2
MOPLHO*MOPLHO2
MSKA*MSKA2
MSKB1*MSKB12
MSKB2*MSKB22
MSKC*MSKC2
MHHUUR*MHHUUR2
MAUT1*MAUT12
MAUT2*MAUT22
MAUT0*MAUT02
MINKGE*MINKGE2
```

```
/list;
run;
```

```
NOTE: There were 5399 observations read from the data set WORK.S26_3.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.91 seconds
      cpu time            0.68 seconds
```

```
data s26_4;
set  s26_3;
drop
MGODRK
MGODPR
MRELGE
MFALLE
MFWEKI
MOPLHO
MSKA
MSKB1
MSKB2
MSKC
MHHUUR
MAUT1
MAUT2
MAUT0
MINKGE;
run;
```

```
NOTE: There were 5399 observations read from the data set WORK.S26_3.
NOTE: The data set WORK.S26_4 has 5399 observations and 30 variables.
NOTE: DATA statement used (Total process time):
      real time           0.08 seconds
      cpu time            0.07 seconds
```

```
%CatToBinWithDrop(s26_4,seqnum,mostyp);
%CatToBinWithDrop(s26_4,seqnum,MOSHOO);

proc means data=s26_4 n nmiss;
run;
```

```
NOTE: There were 5399 observations read from the data set WORK.S26_4.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.48 seconds
      cpu time            0.25 seconds
```

```
data hold00;
set s26_4;
if resp=.;
run;
```

```
NOTE: There were 5399 observations read from the data set WORK.S26_4.
NOTE: The data set WORK.HOLD00 has 1665 observations and 78 variables.
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      cpu time            0.03 seconds
```

```
data anal00;
set s26_4;
if resp>.;
run;
```

```
NOTE: There were 5399 observations read from the data set WORK.S26_4.
NOTE: The data set WORK.ANAL00 has 3734 observations and 78 variables.
NOTE: DATA statement used (Total process time):
      real time           0.06 seconds
      cpu time            0.03 seconds
```

```
PROC EXPORT DATA= WORK.ANAL00
            OUTFILE= "C:\Users\sailahari\Desktop\IDS 462\Lecture8\ANAL00
.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=YES;
RUN;

PROC EXPORT DATA= WORK.HOLD00
            OUTFILE= "C:\Users\sailahari\Desktop\IDS 462\Lecture8\HOLD00
.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=YES;
RUN;

proc contents data=anal00;
run;
```

R Programming

```
> setwd("C:/Users/sailahari/Desktop/IDS 462/Lecture8")
> train<- read.csv("Anal00.csv")
> str(train)
'data.frame':  3734 obs. of  78 variables:
 $ SeqNum     : int  1 2 4 5 6 7 8 9 10 11 ...
 $ MGEMLE     : int  6 5 4 4 2 6 4 3 6 6 ...
 $ MGEMOM     : int  0 2 1 0 5 4 2 7 2 0 ...
 $ MAANTH     : int  5 5 3 9 3 3 3 1 3 2 ...
 $ AWAPAR     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ APERSA     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ AMOTSC     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ Resp       : int  0 0 0 0 0 0 0 0 0 0 ...
 $ rand       : num  0.0528 0.2797 0.1153 0.4631 0.0133 ...
 $ RespHoldout: logi  NA NA NA NA NA NA ...
 $ PWAPAR2    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PAANHA2    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PPERSA2    : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
$ MSKB12     : num  5.5 17 0 5.5 30 5.5 17 30 17 5.5 ...
$ MAUT02     : num  0 0 0 56 0 0 0 5.5 0 0 ...
$ MHHUUR2    : num  0 5.5 17 5.5 0 17 0 0 30 5.5 ...
$ MAUT22     : num  5.5 0 43 0 17 0 17 5.5 17 0 ...
$ MINKGE2    : num  5.5 0 30 100 17 0 17 5.5 5.5 30 ...
$ MFALLE2    : num  56 43 30 43 56 17 82 5.5 56 43 ...
$ MRELGE2    : num  5.5 0 30 56 0 0 0 0 0 43 ...
$ MGODRK2    : num  69 56 43 0 43 56 56 94 43 0 ...
$ MOPLHO2    : num  17 30 5.5 0 17 43 17 0 5.5 69 ...
$ MFWEKI2    : num  5.5 0 0 43 0 0 0 0 0 0 ...
$ MSKB22     : num  5.5 5.5 17 5.5 0 0 0 30 0 5.5 ...
$ MGODPR2    : num  0 0 0 0 0 0 0 0 0 17 ...
$ MSKC2      : num  17 56 43 43 56 43 30 5.5 43 30 ...
$ MAUT12     : num  17 43 17 0 17 56 17 5.5 30 0 ...
$ MSKA2      : num  17 17 30 17 5.5 17 30 43 30 30 ...
$ mostyp_1   : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_2   : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_3   : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_4   : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_5   : int  1 0 0 0 0 0 1 1 0 0 ...
$ mostyp_6   : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_7   : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_8   : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_9   : int  0 1 0 0 0 0 0 0 0 0 ...
$ mostyp_10  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_11  : int  0 0 0 0 0 1 0 0 0 0 ...
$ mostyp_12  : int  0 0 0 1 0 0 0 0 0 0 ...
$ mostyp_13  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_14  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_15  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_16  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_17  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_18  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_19  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_20  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_21  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_22  : int  0 0 1 0 0 0 0 0 0 0 ...
$ mostyp_23  : int  0 0 0 0 0 0 0 0 0 1 ...
$ mostyp_24  : int  0 0 0 0 0 0 0 0 1 0 ...
$ mostyp_25  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_26  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_28  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_29  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_30  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_31  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_32  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_33  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_34  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_35  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_36  : int  0 0 0 0 1 0 0 0 0 0 ...
$ mostyp_37  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_38  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_39  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_40  : int  0 0 0 0 0 0 0 0 0 0 ...
$ mostyp_41  : int  0 0 0 0 0 0 0 0 0 0 ...
$ MOSHOO_1   : int  0 0 0 0 1 0 0 0 0 0 ...
$ MOSHOO_2   : int  0 0 0 0 0 0 0 0 0 0 ...
$ MOSHOO_3   : int  0 0 0 0 0 0 0 0 0 0 ...
$ MOSHOO_4   : int  1 1 0 0 0 0 1 1 0 0 ...
$ MOSHOO_5   : int  0 0 0 0 0 1 0 0 0 0 ...
$ MOSHOO_6   : int  0 0 0 1 0 0 0 0 0 0 ...
$ MOSHOO_7   : int  0 0 0 0 0 0 0 0 0 0 ...
$ MOSHOO_8   : int  0 0 0 0 0 0 0 0 0 0 ...
$ MOSHOO_9   : int  0 0 1 0 0 0 0 0 1 1 ...
```

```
 $ MOSHOO_10  : int  0 0 0 0 0 0 0 0 0 0 ...
> colnames(train)
 [1] "SeqNum"      "MGEMLE"      "MGEMOM"      "MAANTH"      "AWAPAR"
 [6] "APERSA"      "AMOTSC"      "Resp"        "rand"        "RespHoldout"
[11] "PWAPAR2"     "PAANHA2"     "PPERSA2"     "MSKB12"      "MAUT02"
[16] "MHHUUR2"     "MAUT22"      "MINKGE2"     "MFALLE2"     "MRELGE2"
[21] "MGODRK2"     "MOPLHO2"     "MFWEKI2"     "MSKB22"      "MGODPR2"
[26] "MSKC2"       "MAUT12"      "MSKA2"       "mostyp_1"    "mostyp_2"
[31] "mostyp_3"    "mostyp_4"    "mostyp_5"    "mostyp_6"    "mostyp_7"
[36] "mostyp_8"    "mostyp_9"    "mostyp_10"   "mostyp_11"   "mostyp_12"
[41] "mostyp_13"   "mostyp_14"   "mostyp_15"   "mostyp_16"   "mostyp_17"
[46] "mostyp_18"   "mostyp_19"   "mostyp_20"   "mostyp_21"   "mostyp_22"
[51] "mostyp_23"   "mostyp_24"   "mostyp_25"   "mostyp_26"   "mostyp_28"
[56] "mostyp_29"   "mostyp_30"   "mostyp_31"   "mostyp_32"   "mostyp_33"
[61] "mostyp_34"   "mostyp_35"   "mostyp_36"   "mostyp_37"   "mostyp_38"
[66] "mostyp_39"   "mostyp_40"   "mostyp_41"   "MOSHOO_1"    "MOSHOO_2"
[71] "MOSHOO_3"    "MOSHOO_4"    "MOSHOO_5"    "MOSHOO_6"    "MOSHOO_7"
[76] "MOSHOO_8"    "MOSHOO_9"    "MOSHOO_10"
> train<-subset(train, select = -c(rand, RespHoldout))
> colnames(train)
 [1] "SeqNum"    "MGEMLE"    "MGEMOM"    "MAANTH"    "AWAPAR"    "APERSA"
 [7] "AMOTSC"    "Resp"      "PWAPAR2"   "PAANHA2"   "PPERSA2"   "MSKB12"
[13] "MAUT02"    "MHHUUR2"   "MAUT22"    "MINKGE2"   "MFALLE2"   "MRELGE2"
[19] "MGODRK2"   "MOPLHO2"   "MFWEKI2"   "MSKB22"    "MGODPR2"   "MSKC2"
[25] "MAUT12"    "MSKA2"     "mostyp_1"  "mostyp_2"  "mostyp_3"  "mostyp_4"
[31] "mostyp_5"  "mostyp_6"  "mostyp_7"  "mostyp_8"  "mostyp_9"  "mostyp_10"
[37] "mostyp_11" "mostyp_12" "mostyp_13" "mostyp_14" "mostyp_15" "mostyp_16"
[43] "mostyp_17" "mostyp_18" "mostyp_19" "mostyp_20" "mostyp_21" "mostyp_22"
[49] "mostyp_23" "mostyp_24" "mostyp_25" "mostyp_26" "mostyp_28" "mostyp_29"
[55] "mostyp_30" "mostyp_31" "mostyp_32" "mostyp_33" "mostyp_34" "mostyp_35"
[61] "mostyp_36" "mostyp_37" "mostyp_38" "mostyp_39" "mostyp_40" "mostyp_41"
[67] "MOSHOO_1"  "MOSHOO_2"  "MOSHOO_3"  "MOSHOO_4"  "MOSHOO_5"  "MOSHOO_6"
[73] "MOSHOO_7"  "MOSHOO_8"  "MOSHOO_9"  "MOSHOO_10"
> install.packages("randomForest")
Installing package into 'C:/Users/sailahari/Documents/R/win-library/3.1'
(as 'lib' is unspecified)
trying URL 'http://cran.rstudio.com/bin/windows/contrib/3.1/randomForest_4.6-
10.zip'
Content type 'application/zip' length 176424 bytes (172 Kb)
opened URL
downloaded 172 Kb

package 'randomForest' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\sailahari\AppData\Local\Temp\RtmpuYb68I\downloaded_packages
> library(randomForest)
randomForest 4.6-10
Type rfNews() to see new features/changes/bug fixes.
Warning message:
package 'randomForest' was built under R version 3.1.3
> rf<-randomForest(x=train[-c(1,8)],y=train$Resp,ntree=200,importance=TRUE)
Warning message:
In randomForest.default(x = train[-c(1, 8)], y = train$Resp, ntree = 200,  :
  The response has five or fewer unique values.  Are you sure you want to do
regression?
> str(rf)
List of 17
 $ call            : language randomForest(x = train[-c(1, 8)], y = train$Resp
, ntree = 200, importance = TRUE)
 $ type            : chr "regression"
 $ predicted       : Named num [1:3734] 0.01782 0.06389 0.014 0.00278 0.00565
...
  ..- attr(*, "names")= chr [1:3734] "1" "2" "3" "4" ...
```

```
 $ mse            : num [1:200] 0.0891 0.0832 0.08 0.082 0.0787 ...
 $ rsq            : num [1:200] -0.608 -0.5 -0.443 -0.479 -0.42 ...
 $ oob.times      : int [1:3734] 81 60 75 72 80 77 81 72 77 67 ...
 $ importance     : num [1:74, 1:2] 2.92e-03 3.06e-03 2.72e-03 -1.04e-04 -5.9
1e-05 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:74] "MGEMLE" "MGEMOM" "MAANTH" "AWAPAR" ...
  .. ..$ : chr [1:2] "%IncMSE" "IncNodePurity"
 $ importanceSD   : Named num [1:74] 2.65e-04 3.28e-04 3.41e-04 3.13e-05 1.34
e-04 ...
  ..- attr(*, "names")= chr [1:74] "MGEMLE" "MGEMOM" "MAANTH" "AWAPAR" ...
 $ localImportance: NULL
 $ proximity      : NULL
 $ ntree          : num 200
 $ mtry           : num 24
 $ forest         :List of 11
  ..$ ndbigtree    : int [1:200] 527 553 515 525 515 483 513 511 555 543 ...
  ..$ nodestatus   : int [1:575, 1:200] -3 -3 -3 -3 -3 -3 -3 -3 -1 -3 ...
  ..$ leftDaughter : int [1:575, 1:200] 2 4 6 8 10 12 14 16 0 18 ...
  ..$ rightDaughter: int [1:575, 1:200] 3 5 7 9 11 13 15 17 0 19 ...
  ..$ nodepred     : num [1:575, 1:200] 0.0576 0.0555 0.4286 0.0704 0.0206 ..
.
  ..$ bestvar      : int [1:575, 1:200] 6 2 15 24 45 20 28 72 0 43 ...
  ..$ xbestsplit   : num [1:575, 1:200] 0.5 3.5 49.5 62.5 0.5 23.5 0.5 0.5 0
0.5 ...
  ..$ ncat         : Named int [1:74] 1 1 1 1 1 1 1 1 1 1 ...
  .. ..- attr(*, "names")= chr [1:74] "MGEMLE" "MGEMOM" "MAANTH" "AWAPAR" ...
  ..$ nrnodes      : int 575
  ..$ ntree        : num 200
  ..$ xlevels      :List of 74
  .. ..$ MGEMLE   : num 0
  .. ..$ MGEMOM   : num 0
  .. ..$ MAANTH   : num 0
  .. ..$ AWAPAR   : num 0
  .. ..$ APERSA   : num 0
  .. ..$ AMOTSC   : num 0
  .. ..$ PWAPAR2  : num 0
  .. ..$ PAANHA2  : num 0
  .. ..$ PPERSA2  : num 0
  .. ..$ MSKB12   : num 0
  .. ..$ MAUT02   : num 0
  .. ..$ MHHUUR2  : num 0
  .. ..$ MAUT22   : num 0
  .. ..$ MINKGE2  : num 0
  .. ..$ MFALLE2  : num 0
  .. ..$ MRELGE2  : num 0
  .. ..$ MGODRK2  : num 0
  .. ..$ MOPLHO2  : num 0
  .. ..$ MFWEKI2  : num 0
  .. ..$ MSKB22   : num 0
  .. ..$ MGODPR2  : num 0
  .. ..$ MSKC2    : num 0
  .. ..$ MAUT12   : num 0
  .. ..$ MSKA2    : num 0
  .. ..$ mostyp_1 : num 0
  .. ..$ mostyp_2 : num 0
  .. ..$ mostyp_3 : num 0
  .. ..$ mostyp_4 : num 0
  .. ..$ mostyp_5 : num 0
  .. ..$ mostyp_6 : num 0
  .. ..$ mostyp_7 : num 0
  .. ..$ mostyp_8 : num 0
  .. ..$ mostyp_9 : num 0
  .. ..$ mostyp_10: num 0
```

```
  .. ..$ mostyp_11: num 0
  .. ..$ mostyp_12: num 0
  .. ..$ mostyp_13: num 0
  .. ..$ mostyp_14: num 0
  .. ..$ mostyp_15: num 0
  .. ..$ mostyp_16: num 0
  .. ..$ mostyp_17: num 0
  .. ..$ mostyp_18: num 0
  .. ..$ mostyp_19: num 0
  .. ..$ mostyp_20: num 0
  .. ..$ mostyp_21: num 0
  .. ..$ mostyp_22: num 0
  .. ..$ mostyp_23: num 0
  .. ..$ mostyp_24: num 0
  .. ..$ mostyp_25: num 0
  .. ..$ mostyp_26: num 0
  .. ..$ mostyp_28: num 0
  .. ..$ mostyp_29: num 0
  .. ..$ mostyp_30: num 0
  .. ..$ mostyp_31: num 0
  .. ..$ mostyp_32: num 0
  .. ..$ mostyp_33: num 0
  .. ..$ mostyp_34: num 0
  .. ..$ mostyp_35: num 0
  .. ..$ mostyp_36: num 0
  .. ..$ mostyp_37: num 0
  .. ..$ mostyp_38: num 0
  .. ..$ mostyp_39: num 0
  .. ..$ mostyp_40: num 0
  .. ..$ mostyp_41: num 0
  .. ..$ MOSHOO_1 : num 0
  .. ..$ MOSHOO_2 : num 0
  .. ..$ MOSHOO_3 : num 0
  .. ..$ MOSHOO_4 : num 0
  .. ..$ MOSHOO_5 : num 0
  .. ..$ MOSHOO_6 : num 0
  .. ..$ MOSHOO_7 : num 0
  .. ..$ MOSHOO_8 : num 0
  .. ..$ MOSHOO_9 : num 0
  .. ..$ MOSHOO_10: num 0
 $ coefs          : NULL
 $ y              : num [1:3734] 0 0 0 0 0 0 0 0 0 0 ...
 $ test           : NULL
 $ inbag          : NULL
 - attr(*, "class")= chr "randomForest"
> test<-read.csv("Hold00.csv")
> test<-subset(test,select=-c(Resp))
> colnames(test)
 [1] "SeqNum"      "MGEMLE"      "MGEMOM"      "MAANTH"      "AWAPAR"
 [6] "APERSA"      "AMOTSC"      "rand"        "RespHoldout" "PWAPAR2"
[11] "PAANHA2"     "PPERSA2"     "MSKB12"      "MAUT02"      "MHHUUR2"
[16] "MAUT22"      "MINKGE2"     "MFALLE2"     "MRELGE2"     "MGODRK2"
[21] "MOPLHO2"     "MFWEKI2"     "MSKB22"      "MGODPR2"     "MSKC2"
[26] "MAUT12"      "MSKA2"       "mostyp_1"    "mostyp_2"    "mostyp_3"
[31] "mostyp_4"    "mostyp_5"    "mostyp_6"    "mostyp_7"    "mostyp_8"
[36] "mostyp_9"    "mostyp_10"   "mostyp_11"   "mostyp_12"   "mostyp_13"
[41] "mostyp_14"   "mostyp_15"   "mostyp_16"   "mostyp_17"   "mostyp_18"
[46] "mostyp_19"   "mostyp_20"   "mostyp_21"   "mostyp_22"   "mostyp_23"
[51] "mostyp_24"   "mostyp_25"   "mostyp_26"   "mostyp_28"   "mostyp_29"
[56] "mostyp_30"   "mostyp_31"   "mostyp_32"   "mostyp_33"   "mostyp_34"
[61] "mostyp_35"   "mostyp_36"   "mostyp_37"   "mostyp_38"   "mostyp_39"
[66] "mostyp_40"   "mostyp_41"   "MOSHOO_1"    "MOSHOO_2"    "MOSHOO_3"
[71] "MOSHOO_4"    "MOSHOO_5"    "MOSHOO_6"    "MOSHOO_7"    "MOSHOO_8"
[76] "MOSHOO_9"    "MOSHOO_10"
```

```
> test$Resp<-test$RespHoldout
> str(test)
'data.frame':  1665 obs. of  78 variables:
 $ SeqNum     : int  3 12 18 21 35 36 37 39 41 42 ...
 $ MGEMLE     : int  2 5 7 6 6 5 3 9 9 4 ...
 $ MGEMOM     : int  4 3 0 3 0 3 0 7 4 2 ...
 $ MAANTH     : int  0 3 3 2 4 3 3 2 0 3 ...
 $ AWAPAR     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ APERSA     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ AMOTSC     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ rand       : num  0.702 0.849 0.833 0.97 0.821 ...
 $ RespHoldout: int  0 0 0 0 0 0 0 0 0 1 ...
 $ PWAPAR2    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PAANHA2    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PPERSA2    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MSKB12     : num  17 17 17 0 5.5 17 0 0 30 0 ...
 $ MAUT02     : num  0 0 17 0 0 17 0 0 5.5 0 ...
 $ MHHUUR2    : num  0 5.5 0 0 0 17 0 0 5.5 17 ...
 $ MAUT22     : num  0 43 0 0 43 17 17 17 17 56 ...
 $ MINKGE2    : num  0 17 17 0 5.5 30 17 0 0 5.5 ...
 $ MFALLE2    : num  43 30 56 69 43 43 82 69 56 82 ...
 $ MRELGE2    : num  0 5.5 17 0 5.5 17 17 0 0 0 ...
 $ MGODRK2    : num  43 5.5 82 69 30 30 0 43 43 17 ...
 $ MOPLHO2    : num  0 43 0 30 30 17 43 43 56 43 ...
 $ MFWEKI2    : num  0 0 0 30 0 0 0 0 5.5 0 ...
 $ MSKB22     : num  17 17 0 0 17 5.5 17 0 5.5 0 ...
 $ MGODPR2    : num  0 0 0 0 0 0 0 0 0 5.5 ...
 $ MSKC2      : num  56 82 5.5 0 82 56 82 30 0 56 ...
 $ MAUT12     : num  17 5.5 56 0 0 30 0 17 17 5.5 ...
 $ MSKA2      : num  17 30 30 30 30 17 17 17 30 30 ...
 $ mostyp_1   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_2   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_3   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_4   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_5   : int  0 0 0 1 0 0 0 0 0 0 ...
 $ mostyp_6   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_7   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_8   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_9   : int  1 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_10  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_11  : int  0 0 0 0 0 0 0 1 0 0 ...
 $ mostyp_12  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_13  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_14  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_15  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_16  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_17  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_18  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_19  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_20  : int  0 0 0 0 0 0 1 0 0 0 ...
 $ mostyp_21  : int  0 0 0 0 0 1 0 0 0 0 ...
 $ mostyp_22  : int  0 1 0 0 0 0 0 0 0 0 ...
 $ mostyp_23  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_24  : int  0 0 0 0 0 0 0 0 0 1 ...
 $ mostyp_25  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_26  : int  0 0 0 0 1 0 0 0 0 0 ...
 $ mostyp_28  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_29  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_30  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_31  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_32  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_33  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_34  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_35  : int  0 0 1 0 0 0 0 0 0 0 ...
```

```
 $ mostyp_36  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_37  : int  0 0 0 0 0 0 0 0 1 0 ...
 $ mostyp_38  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_39  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_40  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_41  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_1   : int  0 0 1 0 0 0 0 0 1 0 ...
 $ MOSHOO_2   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_3   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_4   : int  1 0 0 1 0 0 0 0 0 0 ...
 $ MOSHOO_5   : int  0 0 0 0 0 0 0 1 0 0 ...
 $ MOSHOO_6   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_7   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_8   : int  0 0 0 0 0 1 1 0 0 0 ...
 $ MOSHOO_9   : int  0 1 0 0 1 0 0 0 0 1 ...
 $ MOSHOO_10  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ Resp       : int  0 0 0 0 0 0 0 0 0 1 ...
> test<-subset(test,select=-c(rand,RespHoldout))
> test$rf<-predict(rf,test)
> str(test)
'data.frame':  1665 obs. of  77 variables:
 $ SeqNum   : int  3 12 18 21 35 36 37 39 41 42 ...
 $ MGEMLE   : int  2 5 7 6 6 5 3 9 9 4 ...
 $ MGEMOM   : int  4 3 0 3 0 3 0 7 4 2 ...
 $ MAANTH   : int  0 3 3 2 4 3 3 2 0 3 ...
 $ AWAPAR   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ APERSA   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ AMOTSC   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PWAPAR2  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PAANHA2  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PPERSA2  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MSKB12   : num  17 17 17 0 5.5 17 0 0 30 0 ...
 $ MAUT02   : num  0 0 17 0 0 17 0 0 5.5 0 ...
 $ MHHUUR2  : num  0 5.5 0 0 0 17 0 0 5.5 17 ...
 $ MAUT22   : num  0 43 0 0 43 17 17 17 17 56 ...
 $ MINKGE2  : num  0 17 17 0 5.5 30 17 0 0 5.5 ...
 $ MFALLE2  : num  43 30 56 69 43 43 82 69 56 82 ...
 $ MRELGE2  : num  0 5.5 17 0 5.5 17 17 0 0 0 ...
 $ MGODRK2  : num  43 5.5 82 69 30 30 0 43 43 17 ...
 $ MOPLHO2  : num  0 43 0 30 30 17 43 43 56 43 ...
 $ MFWEKI2  : num  0 0 0 30 0 0 0 0 5.5 0 ...
 $ MSKB22   : num  17 17 0 0 17 5.5 17 0 5.5 0 ...
 $ MGODPR2  : num  0 0 0 0 0 0 0 0 0 5.5 ...
 $ MSKC2    : num  56 82 5.5 0 82 56 82 30 0 56 ...
 $ MAUT12   : num  17 5.5 56 0 0 30 0 17 17 5.5 ...
 $ MSKA2    : num  17 30 30 30 30 17 17 17 30 30 ...
 $ mostyp_1 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_2 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_3 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_4 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_5 : int  0 0 0 1 0 0 0 0 0 0 ...
 $ mostyp_6 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_7 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_8 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_9 : int  1 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_10: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_11: int  0 0 0 0 0 0 0 0 1 0 0 ...
 $ mostyp_12: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_13: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_14: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_15: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_16: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_17: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_18: int  0 0 0 0 0 0 0 0 0 0 ...
```

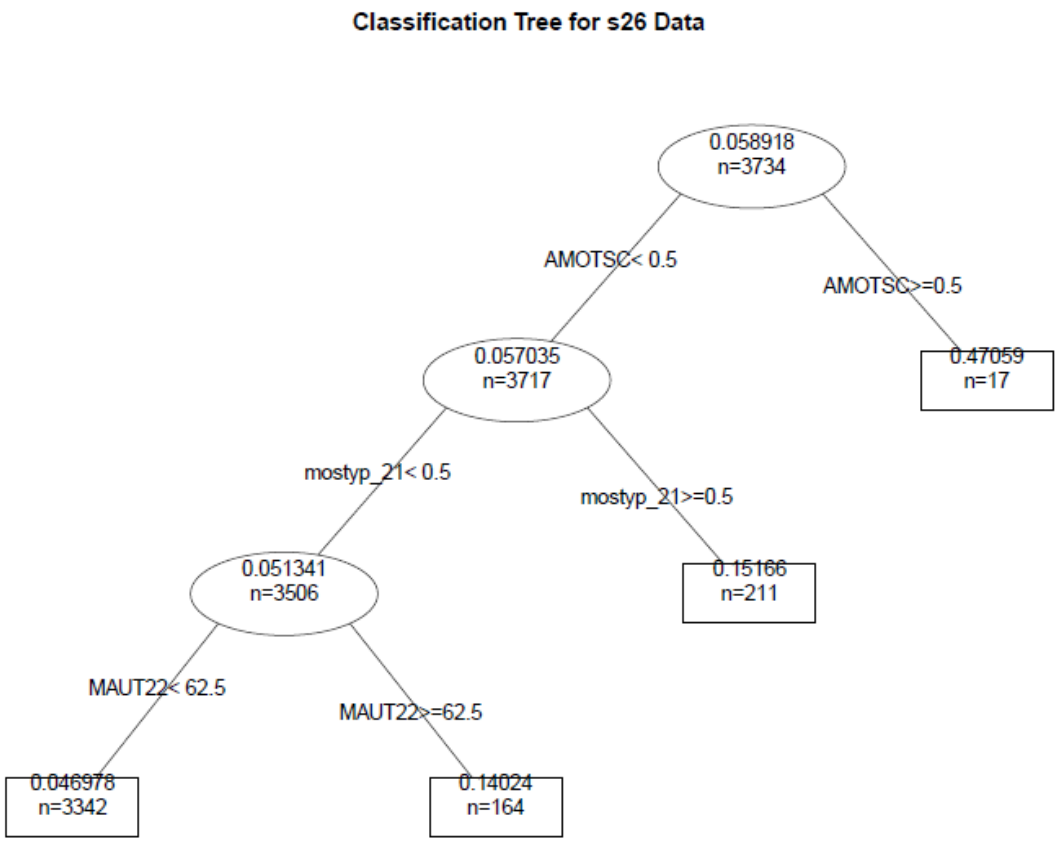
```
 $ mostyp_19: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_20: int  0 0 0 0 0 0 1 0 0 0 ...
 $ mostyp_21: int  0 0 0 0 0 1 0 0 0 0 ...
 $ mostyp_22: int  0 1 0 0 0 0 0 0 0 0 ...
 $ mostyp_23: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_24: int  0 0 0 0 0 0 0 0 0 1 ...
 $ mostyp_25: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_26: int  0 0 0 0 1 0 0 0 0 0 ...
 $ mostyp_28: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_29: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_30: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_31: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_32: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_33: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_34: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_35: int  0 0 1 0 0 0 0 0 0 0 ...
 $ mostyp_36: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_37: int  0 0 0 0 0 0 0 0 1 0 ...
 $ mostyp_38: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_39: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_40: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_41: int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_1 : int  0 0 1 0 0 0 0 0 1 0 ...
 $ MOSHOO_2 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_3 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_4 : int  1 0 0 1 0 0 0 0 0 0 ...
 $ MOSHOO_5 : int  0 0 0 0 0 0 0 1 0 0 ...
 $ MOSHOO_6 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_7 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_8 : int  0 0 0 0 0 1 1 0 0 0 ...
 $ MOSHOO_9 : int  0 1 0 0 1 0 0 0 0 1 ...
 $ MOSHOO_10: int  0 0 0 0 0 0 0 0 0 0 ...
 $ Resp     : int  0 0 0 0 0 0 0 0 0 1 ...
 $ rf       : num  0.002 0.01425 0.17051 0.1932 0.00214 ...
> install.packages("earth")
Installing package into ‘C:/Users/sailahari/Documents/R/win-library/3.1’
(as ‘lib’ is unspecified)
also installing the dependencies ‘plotmo’, ‘plotrix’

trying URL 'http://cran.rstudio.com/bin/windows/contrib/3.1/plotmo_2.2.1.zip'
Content type 'application/zip' length 276725 bytes (270 Kb)
opened URL
downloaded 270 Kb

trying URL 'http://cran.rstudio.com/bin/windows/contrib/3.1/plotrix_3.5-11.zi
p'
Content type 'application/zip' length 650572 bytes (635 Kb)
opened URL
downloaded 635 Kb

trying URL 'http://cran.rstudio.com/bin/windows/contrib/3.1/earth_4.2.0.zip'
Content type 'application/zip' length 1352745 bytes (1.3 Mb)
opened URL
downloaded 1.3 Mb

package ‘plotmo’ successfully unpacked and MD5 sums checked
package ‘plotrix’ successfully unpacked and MD5 sums checked
package ‘earth’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\sailahari\AppData\Local\Temp\RtmpuYb68I\downloaded_packages
```

```
> library("earth")
Loading required package: plotmo
Loading required package: plotrix
Warning messages:
1: package 'earth' was built under R version 3.1.3
2: package 'plotmo' was built under R version 3.1.3
3: package 'plotrix' was built under R version 3.1.3
> mars<-earth(x=train[-c(1,8)],y=train$Resp)
> summary(mars,digits=2,style="pmax")
Call: earth(x=train[-c(1, 8)], y=train$Resp)

train$Resp =
  0.38
  +   0.066 * mostyp_21
  +   0.076 * mostyp_25
  -   0.024 * MOSHOO_9
  -   0.014 * pmax(0,          2 -    MGEMOM)
  -   0.022 * pmax(0,    MGEMOM -         2)
  +   0.023 * pmax(0,    MGEMOM -         4)
  -    0.38 * pmax(0,         1 -    AMOTSC)
  +    0.52 * pmax(0,    AMOTSC -         1)
  + 0.00072 * pmax(0,        43 -    MSKB12)
  -  0.0012 * pmax(0,   MHHUUR2 -       5.5)
  +  0.0019 * pmax(0,    MAUT22 -        56)
  -  0.0055 * pmax(0,   MINKGE2 -        82)
  -   0.001 * pmax(0,        56 -   MRELGE2)
  +  0.0037 * pmax(0,       5.5 -   MFWEKI2)
  +  0.0022 * pmax(0,        43 -   MGODPR2)

Selected 16 of 23 terms, and 12 of 74 predictors
Termination condition: RSq changed by less than 0.001 at 23 terms
Importance: AMOTSC, mostyp_21, MSKB12, MRELGE2, MINKGE2, MGODPR2, MAUT22, MGE
MOM, ...
Number of terms at each degree of interaction: 1 15 (additive model)
GCV 0.054    RSS 197    GRSq 0.031    RSq 0.047
> test$marspred<-predict(mars,test)
> str(test)
'data.frame':  1665 obs. of  78 variables:
 $ SeqNum   : int  3 12 18 21 35 36 37 39 41 42 ...
 $ MGEMLE   : int  2 5 7 6 6 5 3 9 9 4 ...
 $ MGEMOM   : int  4 3 0 3 0 3 0 7 4 2 ...
 $ MAANTH   : int  0 3 3 2 4 3 3 2 0 3 ...
 $ AWAPAR   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ APERSA   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ AMOTSC   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PWAPAR2  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PAANHA2  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PPERSA2  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MSKB12   : num  17 17 17 0 5.5 17 0 0 30 0 ...
 $ MAUT02   : num  0 0 17 0 0 17 0 0 5.5 0 ...
 $ MHHUUR2  : num  0 5.5 0 0 0 17 0 0 5.5 17 ...
 $ MAUT22   : num  0 43 0 0 43 17 17 17 17 56 ...
 $ MINKGE2  : num  0 17 17 0 5.5 30 17 0 0 5.5 ...
 $ MFALLE2  : num  43 30 56 69 43 43 82 69 56 82 ...
 $ MRELGE2  : num  0 5.5 17 0 5.5 17 17 0 0 0 ...
 $ MGODRK2  : num  43 5.5 82 69 30 30 0 43 43 17 ...
 $ MOPLHO2  : num  0 43 0 30 30 17 43 43 56 43 ...
 $ MFWEKI2  : num  0 0 0 30 0 0 0 0 5.5 0 ...
 $ MSKB22   : num  17 17 0 0 17 5.5 17 0 5.5 0 ...
 $ MGODPR2  : num  0 0 0 0 0 0 0 0 0 5.5 ...
 $ MSKC2    : num  56 82 5.5 0 82 56 82 30 0 56 ...
 $ MAUT12   : num  17 5.5 56 0 0 30 0 17 17 5.5 ...
 $ MSKA2    : num  17 30 30 30 30 17 17 17 30 30 ...
```

```
 $ mostyp_1 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_2 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_3 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_4 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_5 : int  0 0 0 1 0 0 0 0 0 0 ...
 $ mostyp_6 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_7 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_8 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_9 : int  1 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_10: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_11: int  0 0 0 0 0 0 0 1 0 0 ...
 $ mostyp_12: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_13: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_14: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_15: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_16: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_17: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_18: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_19: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_20: int  0 0 0 0 0 0 1 0 0 0 ...
 $ mostyp_21: int  0 0 0 0 0 1 0 0 0 0 ...
 $ mostyp_22: int  0 1 0 0 0 0 0 0 0 0 ...
 $ mostyp_23: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_24: int  0 0 0 0 0 0 0 0 0 1 ...
 $ mostyp_25: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_26: int  0 0 0 0 1 0 0 0 0 0 ...
 $ mostyp_28: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_29: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_30: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_31: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_32: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_33: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_34: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_35: int  0 0 1 0 0 0 0 0 0 0 ...
 $ mostyp_36: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_37: int  0 0 0 0 0 0 0 0 1 0 ...
 $ mostyp_38: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_39: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_40: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_41: int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_1 : int  0 0 1 0 0 0 0 0 1 0 ...
 $ MOSHOO_2 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_3 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_4 : int  1 0 0 1 0 0 0 0 0 0 ...
 $ MOSHOO_5 : int  0 0 0 0 0 0 0 0 1 0 ...
 $ MOSHOO_6 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_7 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_8 : int  0 0 0 0 0 0 1 1 0 0 ...
 $ MOSHOO_9 : int  0 1 0 0 1 0 0 0 0 1 ...
 $ MOSHOO_10: int  0 0 0 0 0 0 0 0 0 0 ...
 $ Resp     : int  0 0 0 0 0 0 0 0 0 1 ...
 $ rf       : num  0.002 0.01425 0.17051 0.1932 0.00214 ...
 $ marspred : num [1:1665, 1] 0.0349 0.0377 0.067 0.0481 0.0394 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : NULL
  .. ..$ : chr "train$Resp"
> install.packages("rpart")
Installing package into ‘C:/Users/sailahari/Documents/R/win-library/3.1’
(as ‘lib’ is unspecified)
trying URL 'http://cran.rstudio.com/bin/windows/contrib/3.1/rpart_4.1-9.zip'
Content type 'application/zip' length 918753 bytes (897 Kb)
opened URL
downloaded 897 Kb
```

```
package 'rpart' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\sailahari\AppData\Local\Temp\RtmpuYb68I\downloaded_packages
> library(rpart)
Warning message:
package 'rpart' was built under R version 3.1.3
> tree<-rpart(Resp~., data=train[-c(1)], control=rpart.control(cp=.005))
> plot(tree, uniform=TRUE,main="Classification Tree for S26 Data")
> text(tree, use.n=TRUE, all=TRUE, cex=.8)
> test$treepred<-predict(tree,test)
> str(test)
'data.frame':  1665 obs. of  79 variables:
 $ SeqNum   : int  3 12 18 21 35 36 37 39 41 42 ...
 $ MGEMLE   : int  2 5 7 6 6 5 3 9 9 4 ...
 $ MGEMOM   : int  4 3 0 3 0 3 0 7 4 2 ...
 $ MAANTH   : int  0 3 3 2 4 3 3 2 0 3 ...
 $ AWAPAR   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ APERSA   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ AMOTSC   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PWAPAR2  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PAANHA2  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PPERSA2  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MSKB12   : num  17 17 17 0 5.5 17 0 0 30 0 ...
 $ MAUT02   : num  0 0 17 0 0 17 0 0 5.5 0 ...
 $ MHHUUR2  : num  0 5.5 0 0 0 17 0 0 5.5 17 ...
 $ MAUT22   : num  0 43 0 0 43 17 17 17 17 56 ...
 $ MINKGE2  : num  0 17 17 0 5.5 30 17 0 0 5.5 ...
 $ MFALLE2  : num  43 30 56 69 43 43 82 69 56 82 ...
 $ MRELGE2  : num  0 5.5 17 0 5.5 17 17 0 0 0 ...
 $ MGODRK2  : num  43 5.5 82 69 30 30 0 43 43 17 ...
 $ MOPLHO2  : num  0 43 0 30 30 17 43 43 56 43 ...
 $ MFWEKI2  : num  0 0 0 30 0 0 0 0 5.5 0 ...
 $ MSKB22   : num  17 17 0 0 17 5.5 17 0 5.5 0 ...
 $ MGODPR2  : num  0 0 0 0 0 0 0 0 0 5.5 ...
 $ MSKC2    : num  56 82 5.5 0 82 56 82 30 0 56 ...
 $ MAUT12   : num  17 5.5 56 0 0 30 0 17 17 5.5 ...
 $ MSKA2    : num  17 30 30 30 30 17 17 17 30 30 ...
 $ mostyp_1 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_2 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_3 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_4 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_5 : int  0 0 0 1 0 0 0 0 0 0 ...
 $ mostyp_6 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_7 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_8 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_9 : int  1 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_10: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_11: int  0 0 0 0 0 0 0 1 0 0 ...
 $ mostyp_12: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_13: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_14: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_15: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_16: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_17: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_18: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_19: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_20: int  0 0 0 0 0 0 1 0 0 0 ...
 $ mostyp_21: int  0 0 0 0 0 1 0 0 0 0 ...
 $ mostyp_22: int  0 1 0 0 0 0 0 0 0 0 ...
 $ mostyp_23: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_24: int  0 0 0 0 0 0 0 0 0 1 ...
 $ mostyp_25: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_26: int  0 0 0 0 1 0 0 0 0 0 ...
```

```
 $ mostyp_28: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_29: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_30: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_31: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_32: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_33: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_34: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_35: int  0 0 1 0 0 0 0 0 0 0 ...
 $ mostyp_36: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_37: int  0 0 0 0 0 0 0 0 1 0 ...
 $ mostyp_38: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_39: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_40: int  0 0 0 0 0 0 0 0 0 0 ...
 $ mostyp_41: int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_1 : int  0 0 1 0 0 0 0 0 1 0 ...
 $ MOSHOO_2 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_3 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_4 : int  1 0 0 1 0 0 0 0 0 0 ...
 $ MOSHOO_5 : int  0 0 0 0 0 0 0 1 0 0 ...
 $ MOSHOO_6 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_7 : int  0 0 0 0 0 0 0 0 0 0 ...
 $ MOSHOO_8 : int  0 0 0 0 0 1 1 0 0 0 ...
 $ MOSHOO_9 : int  0 1 0 0 1 0 0 0 0 1 ...
 $ MOSHOO_10: int  0 0 0 0 0 0 0 0 0 0 ...
 $ Resp     : int  0 0 0 0 0 0 0 0 0 1 ...
 $ rf       : num  0.002 0.01425 0.17051 0.1932 0.00214 ...
 $ marspred : num [1:1665, 1] 0.0349 0.0377 0.067 0.0481 0.0394 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : NULL
  .. ..$ : chr "train$Resp"
 $ treepred : num  0.047 0.047 0.047 0.047 0.047 ...
> write.csv(test,file="test.csv")
```




```
PROC IMPORT OUT= WORK.DataFromR
            DATAFILE= "C:\Users\sailahari\Desktop\IDS 462\Lecture8\test.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;
```

```
NOTE: WORK.DATAFROMR data set was successfully created.
NOTE: The data set WORK.DATAFROMR has 1665 observations and 80 variables.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time           0.94 seconds
      cpu time            0.81 seconds
```

```
proc freq data= DataFromR;
tables resp;
run;
```

The SAS System

The FREQ Procedure

Resp	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1562	93.81	1562	93.81
1	103	6.19	1665	100.00

```
proc means data=work.DataFromR nmiss mean std cv p1 p10 p25 p50 p75 p90 p99;
var treepred marspred rf;
run;
```

The SAS System

The MEANS Procedure

Variable	N Miss	Mean	Std Dev	Coeff of Variation	1st Pctl	10th Pctl	25th Pctl	50th Pctl	75th Pctl	90th Pctl	99th Pctl
treepred	0	0.0606066	0.0457094	75.4198110	0.0469779	0.0469779	0.0469779	0.0469779	0.0469779	0.1402439	0.1516588
marspred	0	0.0606971	0.0581082	95.7346581	-0.0179995	0.0121440	0.0300416	0.0497850	0.0745602	0.1212540	0.2110994
rf	0	0.0633106	0.1107961	175.0040360	-2.47267E-16	-1.63827E-16	0.0020000	0.0143929	0.0654624	0.1999731	0.5006667

```
data treeanal;
set  DataFromR;
keep treepred resp;
run;
```

```
NOTE: There were 1665 observations read from the data set WORK.DATAFROMR.
NOTE: The data set WORK.TREEANAL has 1665 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.02 seconds
      cpu time            0.03 seconds
```

```
proc sort data=treeanal;
by descending treepred;
run;

data treeanal_1;
set  treeanal;
treecumresp+resp;
treepct=treecumresp/103;
run;
```

```
NOTE: There were 1665 observations read from the data set WORK.TREEANAL.
NOTE: The data set WORK.TREEANAL_1 has 1665 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      cpu time            0.03 seconds
```

```
data Forestanal;
set  DataFromR;
keep rf resp;
```

```
run;
```

```
NOTE: There were 1665 observations read from the data set WORK.DATAFROMR.
NOTE: The data set WORK.FORESTANAL has 1665 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      cpu time            0.03 seconds
```

```
proc sort data=Forestanal;
by descending rf;
run;

data Forestanal_1;
set  Forestanal;
Forestcumresp+resp;
Forestpct=Forestcumresp/103;
run;
```

```
NOTE: There were 1665 observations read from the data set WORK.FORESTANAL.
NOTE: The data set WORK.FORESTANAL_1 has 1665 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.02 seconds
      cpu time            0.01 seconds
```

```
data Marsanal;
set  DataFromR;
keep marspred resp;
run;
```

```
NOTE: There were 1665 observations read from the data set WORK.DATAFROMR.
NOTE: The data set WORK.MARSANAL has 1665 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.13 seconds
      cpu time            0.01 seconds
```

```
proc sort data=Marsanal;
by descending marspred;
run;

data Marsanal_2;
set  marsanal;
Marscumresp+resp;
Marspct=Marscumresp/103;
run;
```

```
NOTE: There were 1665 observations read from the data set WORK.MARSANAL.
NOTE: The data set WORK.MARSANAL_2 has 1665 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.02 seconds
      cpu time            0.03 seconds
```

```
data  compare;
merge ForestAnal_1
```

```
      MarsAnal_2
       TreeAnal_1
       RespAnal;
run;
```

```
NOTE: There were 1665 observations read from the data set WORK.FORESTANAL_1.
NOTE: There were 1665 observations read from the data set WORK.MARSANAL_2.
NOTE: There were 1665 observations read from the data set WORK.TREEANAL_1.
NOTE: There were 1665 observations read from the data set WORK.RESPANAL.
NOTE: The data set WORK.COMPARE has 1665 observations and 16 variables.
NOTE: DATA statement used (Total process time):
      real time           1.15 seconds
      cpu time            0.06 seconds
```