

Predict422-CharityProject Part 1

Artur Mrozowski

April 19, 2017

Exercises 1. Read Data from CSV File (a) Read data from csv file

```
charityData = read.csv(file.choose(),na.strings=c("NA"," "))
```

(c)

```
charityData$DONR = as.factor(charityData$DONR)
charityData$HOME = as.factor(charityData$HOME)
charityData$HINC = as.factor(charityData$HINC)
```

Following variable converted to factors. GENDER and RFA_96 converted due to default setting stringsAsFactors = TRUE

```
str(charityData)
```

```
## 'data.frame': 70871 obs. of 20 variables:
## $ ID : int 1 3 4 6 8 12 14 15 16 17 ...
## $ DONR : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ DAMT : num 0 0 0 0 0 0 0 0 0 0 ...
## $ AGE : int 62 86 66 52 55 49 41 48 58 75 ...
## $ HOME : Factor w/ 2 levels "0","1": 2 NA 2 2 2 NA NA 1 2 2 ...
## $ HINC : Factor w/ 7 levels "1","2","3","4",...: 4 NA 6 4 2 NA NA 5 6 7 ...
## $ GENDER : Factor w/ 6 levels "A","C","F","J",...: 5 3 5 4 3 5 3 3 3 3 ...
## $ MEDAGE : int 45 38 44 32 42 39 40 43 47 49 ...
## $ MEDPPH : int 180 118 225 247 208 266 288 198 236 183 ...
## $ MEDHVAL : int 313 1131 1518 857 440 580 678 1044 871 5913 ...
## $ MEDINC : int 195 323 728 501 236 202 171 562 322 810 ...
## $ MEDEDUC : int 120 160 160 140 120 120 114 120 120 160 ...
## $ NUMPROM : int 34 53 18 43 55 23 14 24 56 32 ...
## $ NUMPRM12: int 14 13 13 14 12 11 8 12 13 12 ...
## $ RAMNTALL: num 53 68 20 84 172 35 15 39 115 44 ...
## $ NGIFTALL: int 6 8 2 7 16 3 1 5 6 5 ...
## $ MAXRAMNT: num 12 15 15 15 15 20 15 10 25 12 ...
## $ LASTGIFT: num 10 15 15 10 11 20 15 7 25 12 ...
## $ TDON : num 16 17 18 16 22 ...
## $ RFA_96 : Factor w/ 69 levels "A1D","A1E","A1F",...: 17 8 43 18 56 2 23 12 4 12 ...
```

2.Data Quality Check

```
dim(charityData) # dimensions of data
```

```
## [1] 70871 20
```

```
names(charityData)    # variable names
```

```
## [1] "ID"      "DONR"    "DAMT"    "AGE"     "HOME"    "HINC"
## [7] "GENDER"  "MEDAGE"  "MEDPPH"  "MEDHVAL" "MEDINC"  "MEDEDUC"
## [13] "NUMPROM" "NUMPRM12" "RAMNTALL" "NGIFTALL" "MAXRAMNT" "LASTGIFT"
## [19] "TDON"    "RFA_96"
```

```
#str(charityData)    # one form of summary of data
# another form of summary
```

(b) We can see that HOME, HINC and GENDER columns contain some missing values.

```
## Check for Missing Values
which(sapply(charityData,anyNA))
```

```
##   HOME   HINC GENDER
##     5     6     7
```

```
# Missing values identified in HINC, GENDER, and RFA_96
# Get counts of missing values for each variable
table(charityData$HOME,useNA="ifany")
```

```
##
##      0      1 <NA>
## 15004 46972 8895
```

```
table(charityData$HINC,useNA="ifany")
```

```
##
##      1      2      3      4      5      6      7 <NA>
##  7084 10616  7189 10983 13454  6770  6657  8118
```

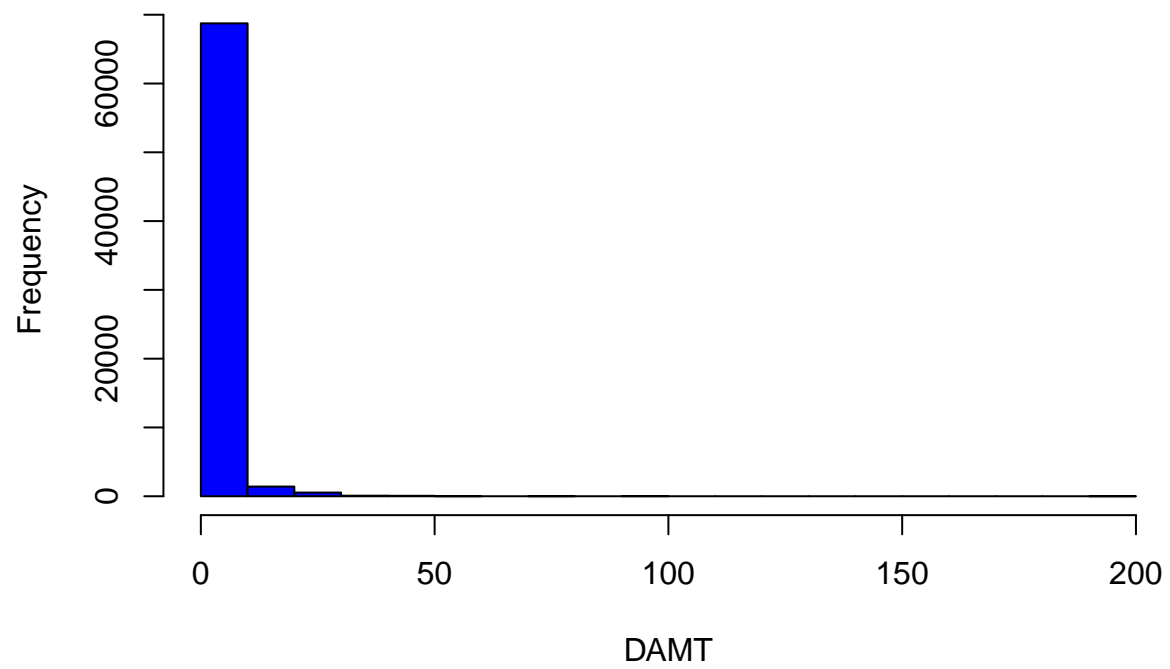
```
table(charityData$GENDER,useNA="ifany")
```

```
##
##      A      C      F      J      M      U <NA>
##      2      1 38183   290 30494   741 1160
```

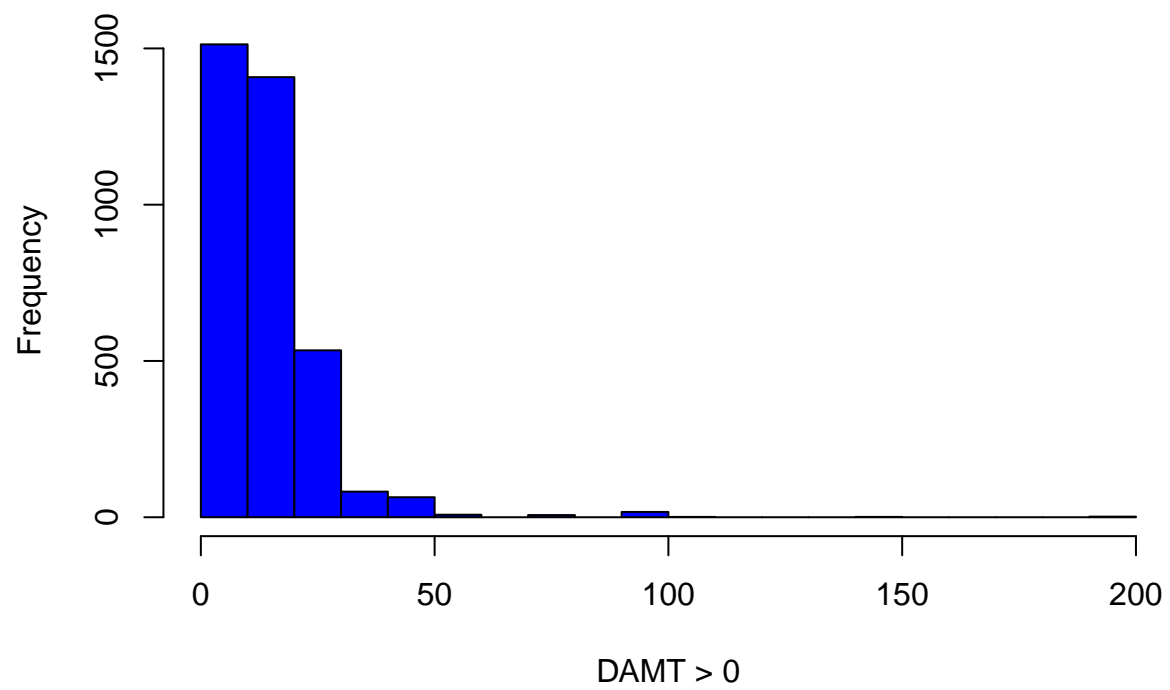
(c) Missing values could be a problem. Replacing the missing values with mean could be one solution.

3(a) Histogram of the response variable DAMT (first with 0s included, second with 0s dropped)

```
hist(charityData$DAMT,col="blue",breaks=20,xlab="DAMT",main="")
```

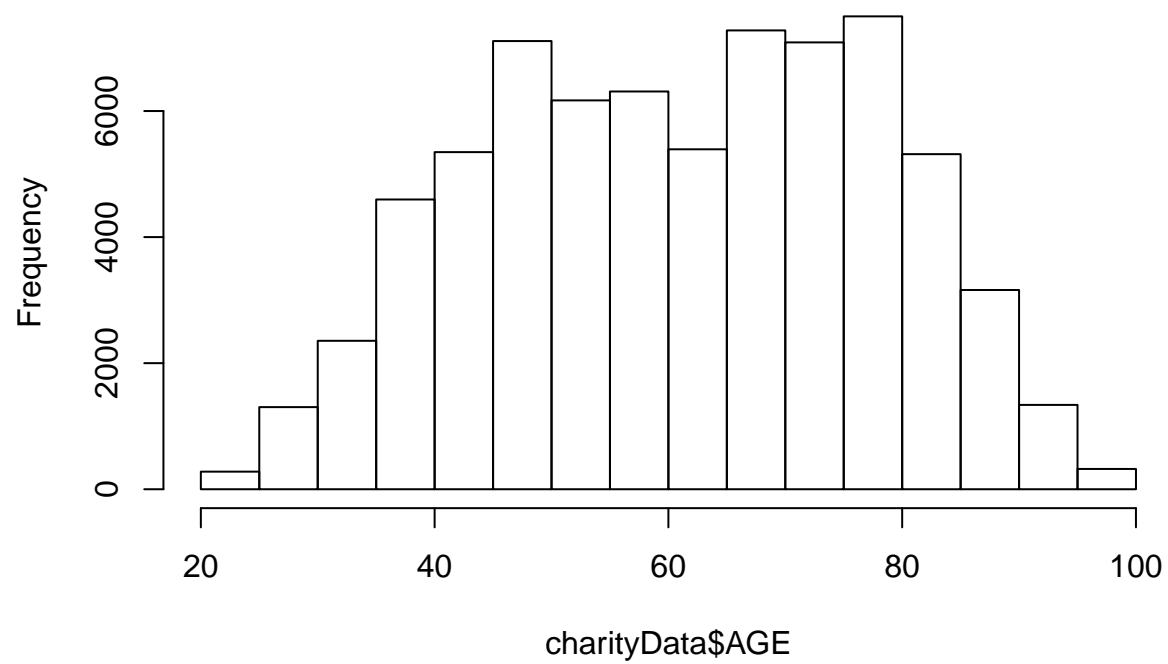


```
hist(charityData$DAMT[charityData$DAMT > 0],col="blue",breaks=20,xlab="DAMT > 0",main="")
```



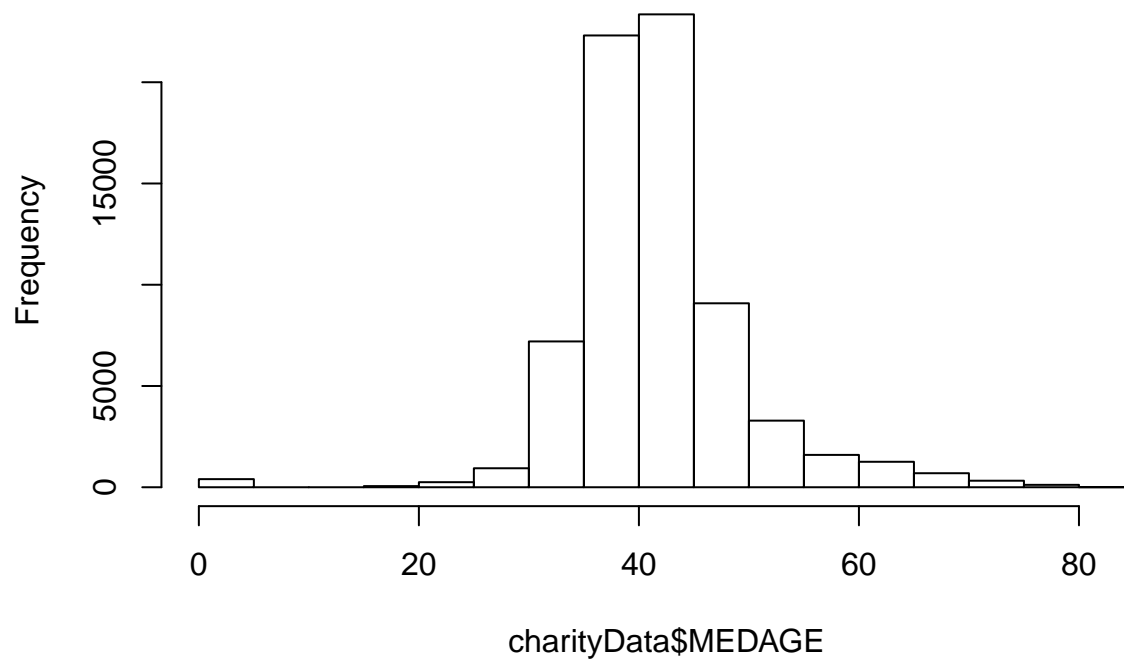
```
hist(charityData$AGE)
```

Histogram of charityData\$AGE



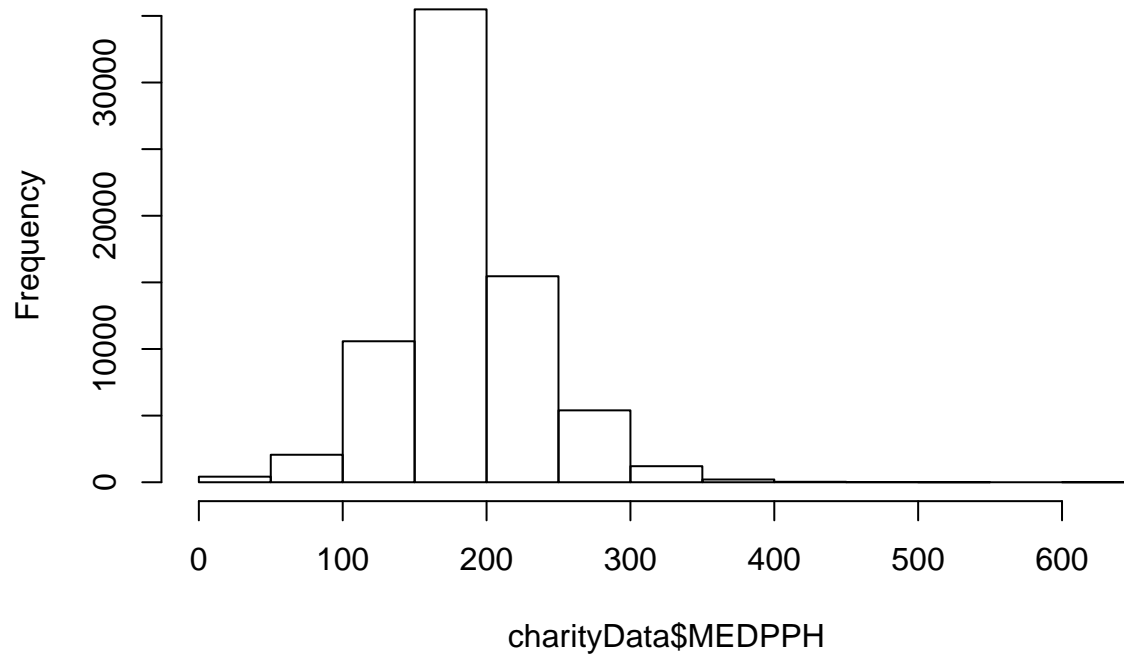
```
hist(charityData$MEDAGE)
```

Histogram of charityData\$MEDAGE



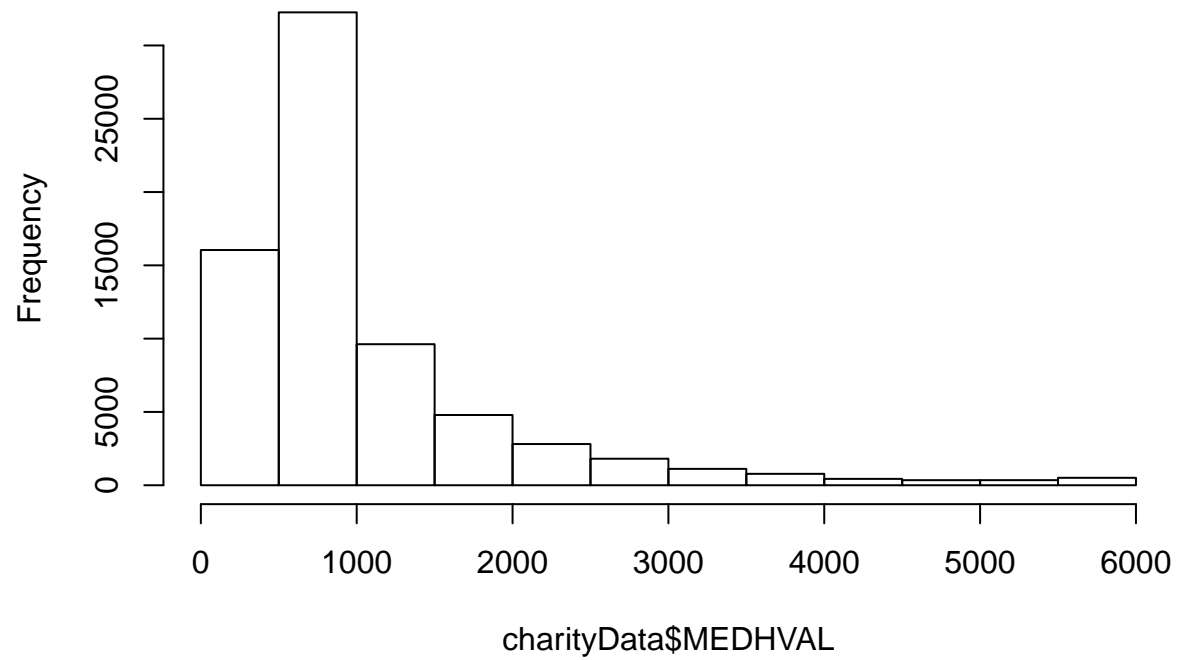
```
hist(charityData$MEDPPH)
```

Histogram of charityData\$MEDPPH



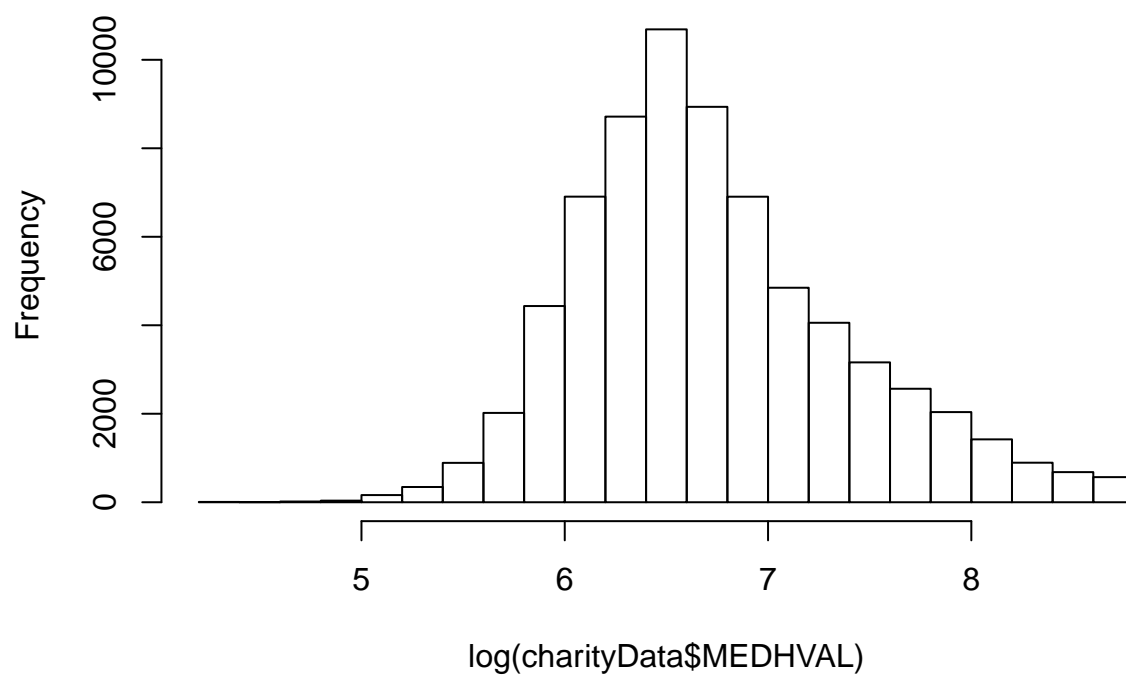
```
hist(charityData$MEDHVAL)
```

Histogram of charityData\$MEDHVAL



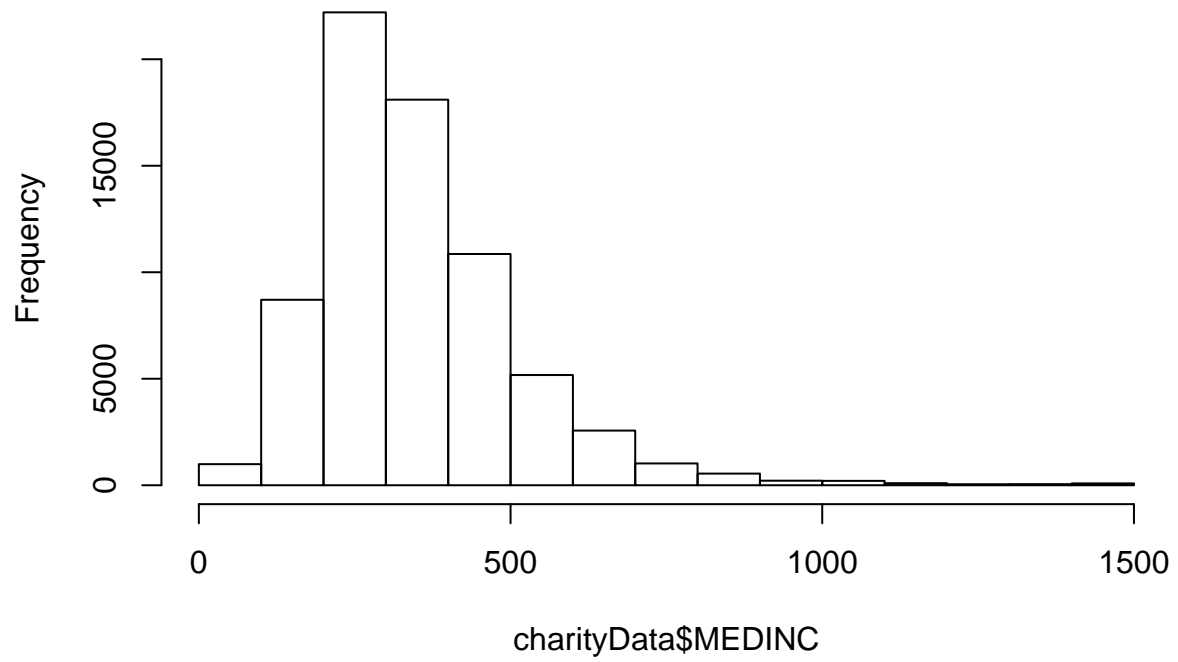
```
hist(log(charityData$MEDHVAL))
```


Histogram of $\log(\text{charityData\$MEDHVAL})$



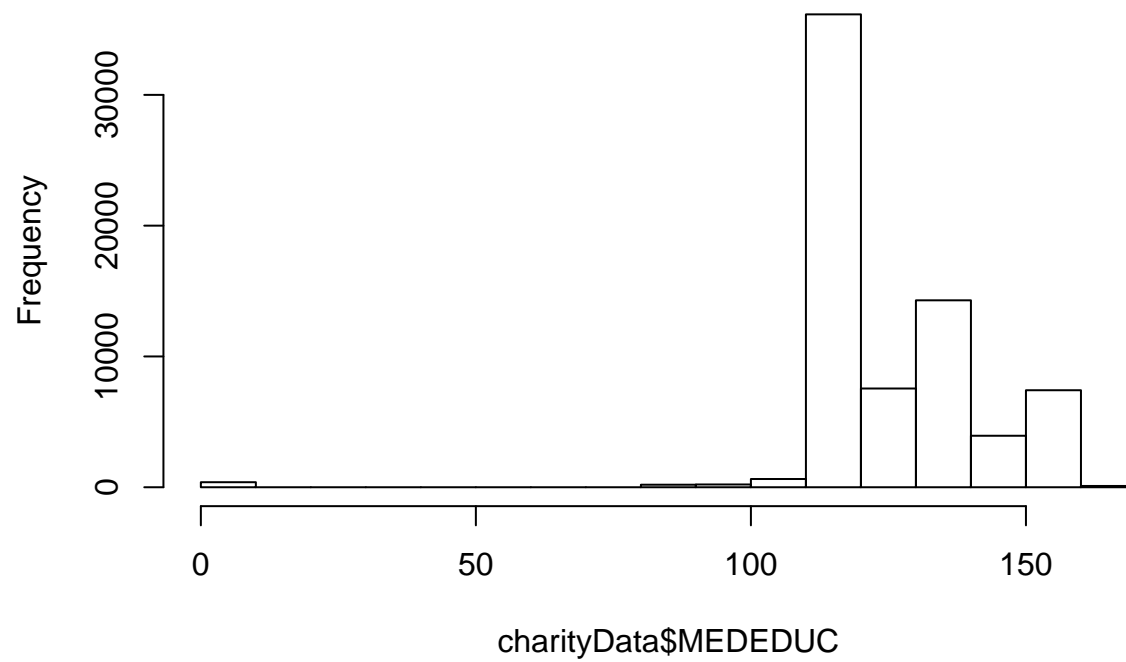
```
hist(charityData$MEDINC)
```

Histogram of charityData\$MEDINC



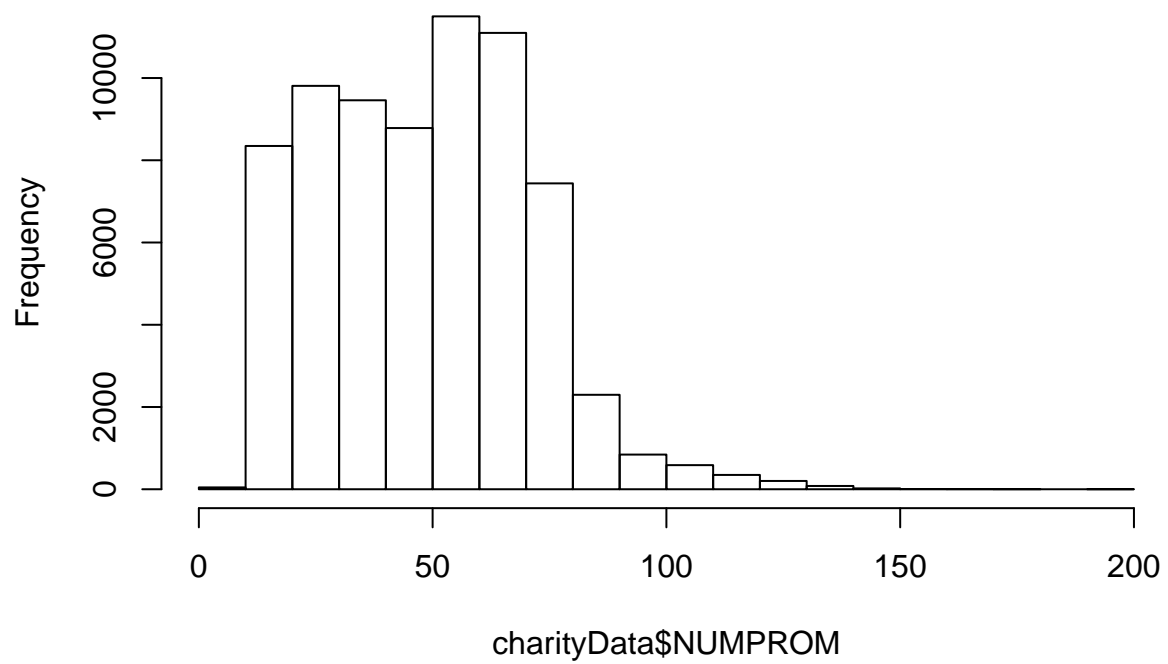
```
hist(charityData$MEDEDUC)
```

Histogram of charityData\$MEDEDUC

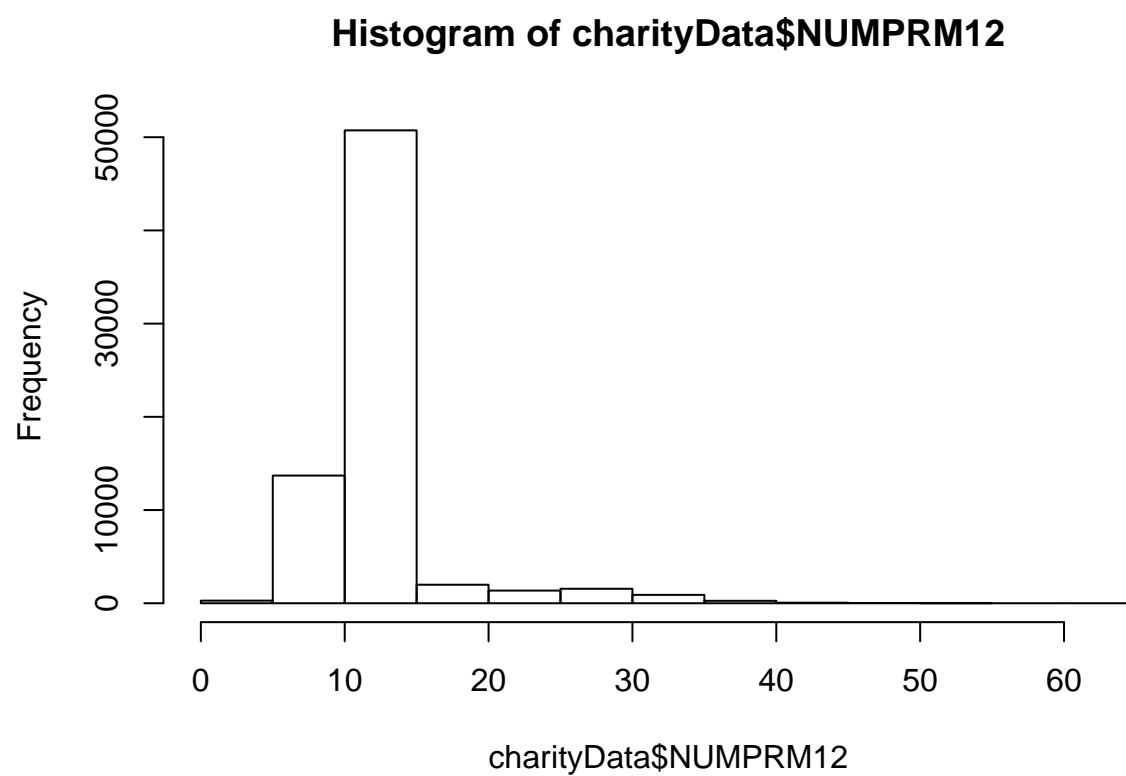


```
hist(charityData$NUMPROM)
```

Histogram of charityData\$NUMFROM

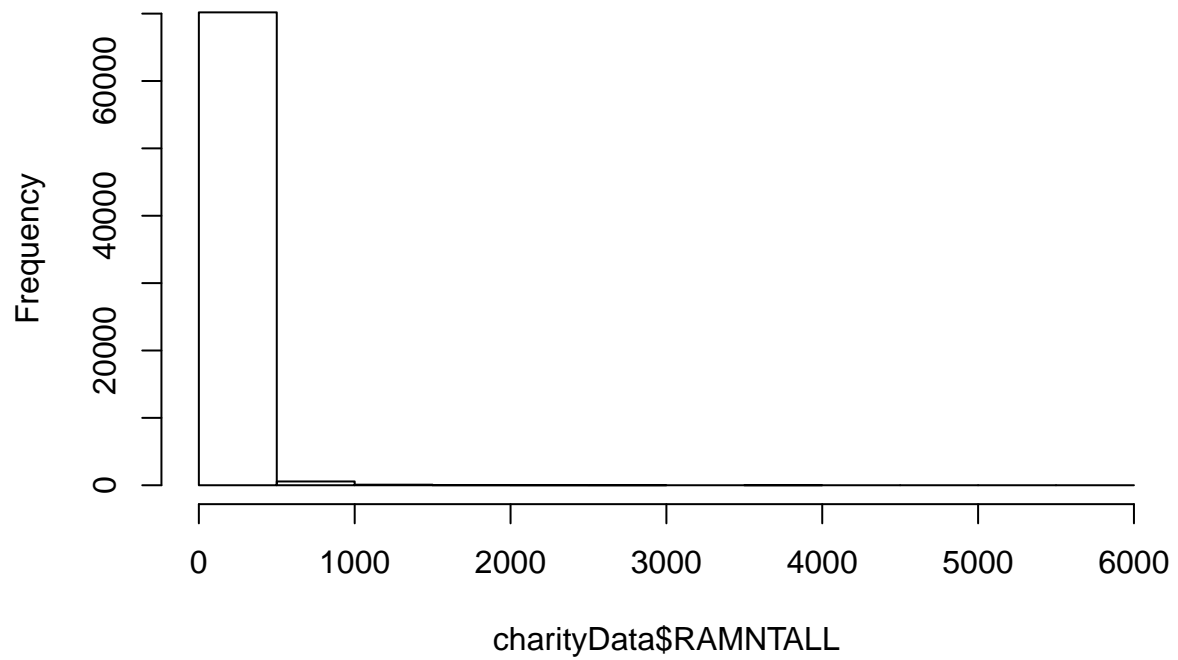


```
hist(charityData$NUMFROM)
```



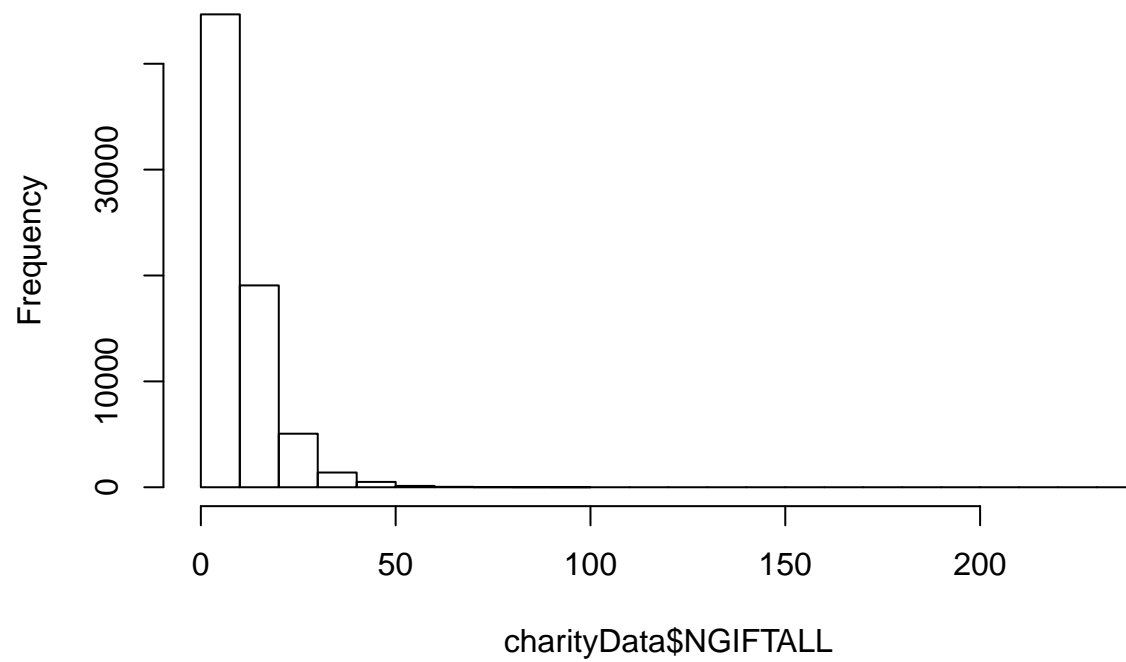
```
hist(charityData$RAMNTALL)
```

Histogram of charityData\$RAMNTALL



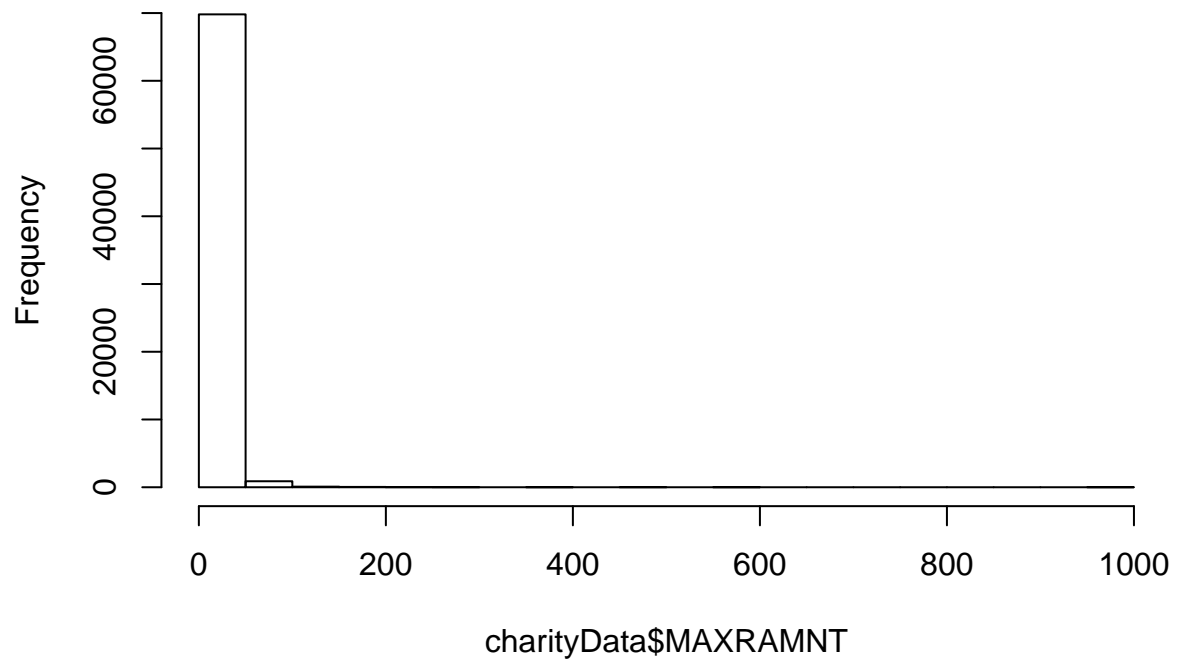
```
hist(charityData$NGIFTALL)
```

Histogram of charityData\$NGIFTALL



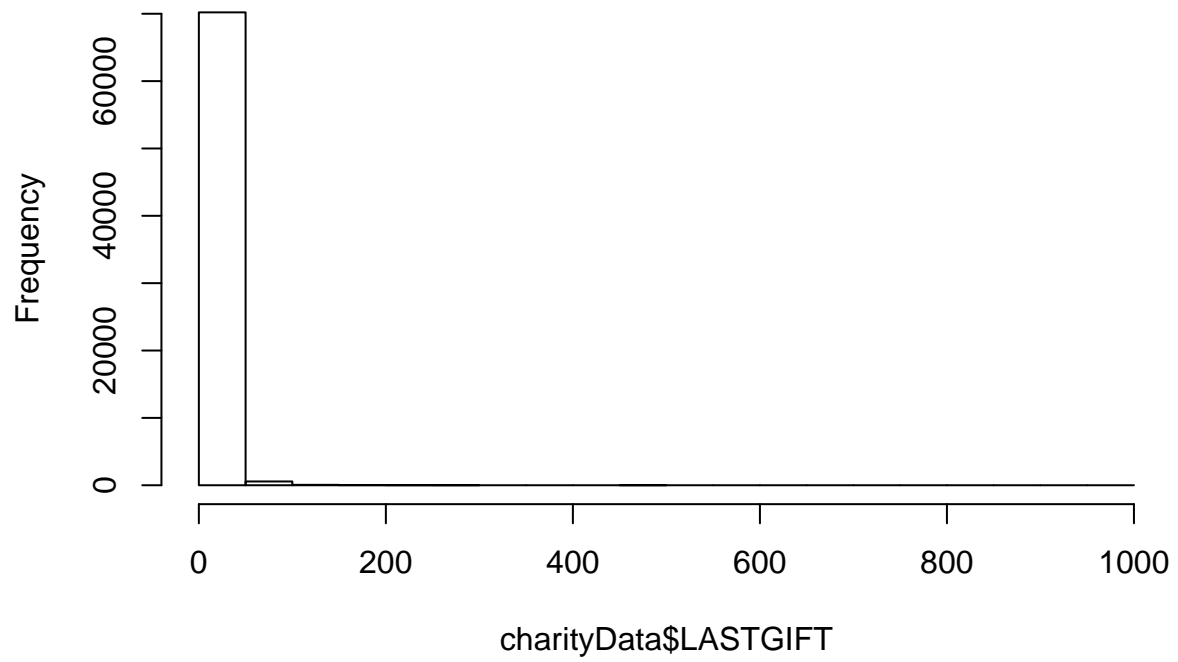
```
hist(charityData$MAXRAMNT)
```

Histogram of charityData\$MAXRAMNT



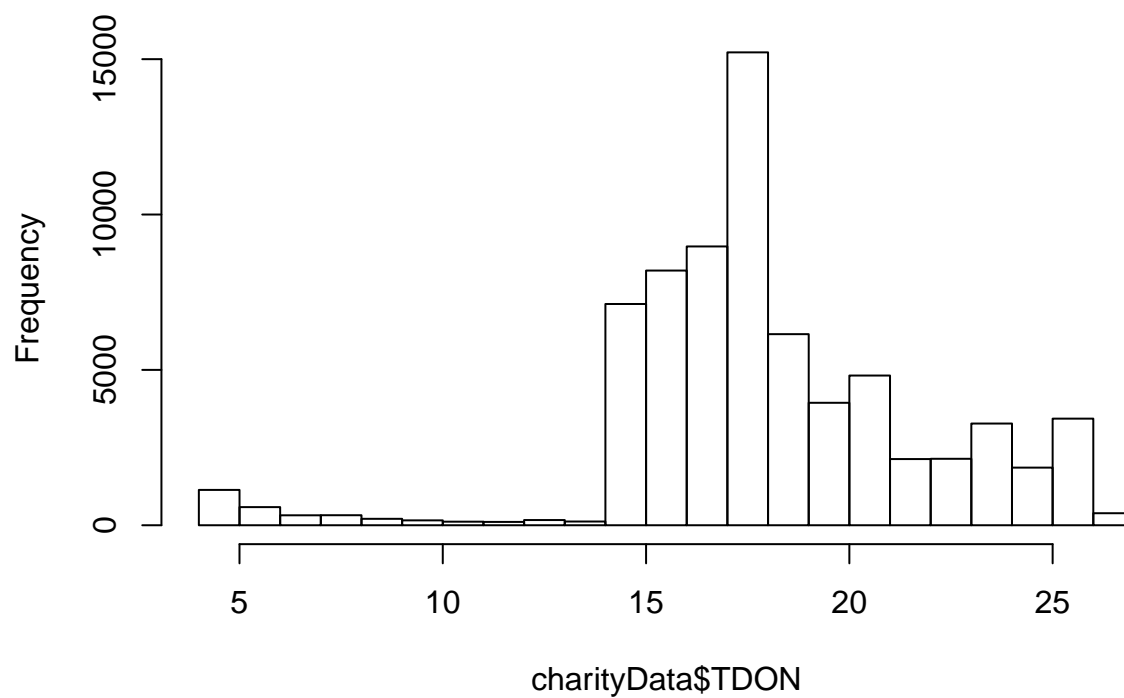
```
hist(charityData$LASTGIFT)
```


Histogram of charityData\$LASTGIFT



```
hist(charityData$TDON)
```

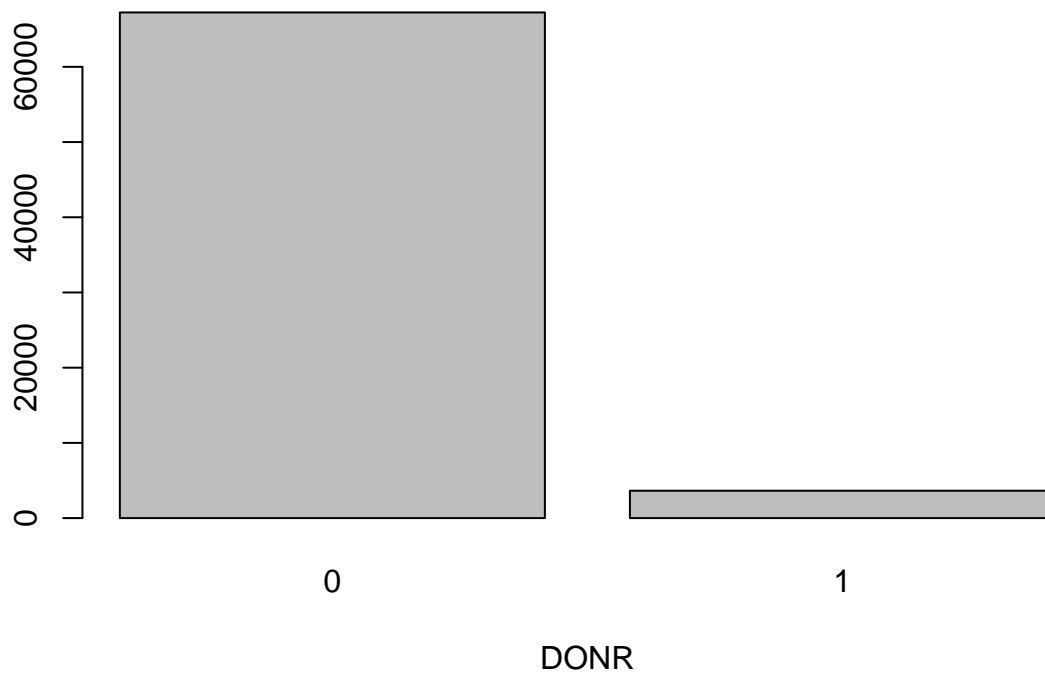
Histogram of charityData\$TDON



```
table(charityData$DONR)
```

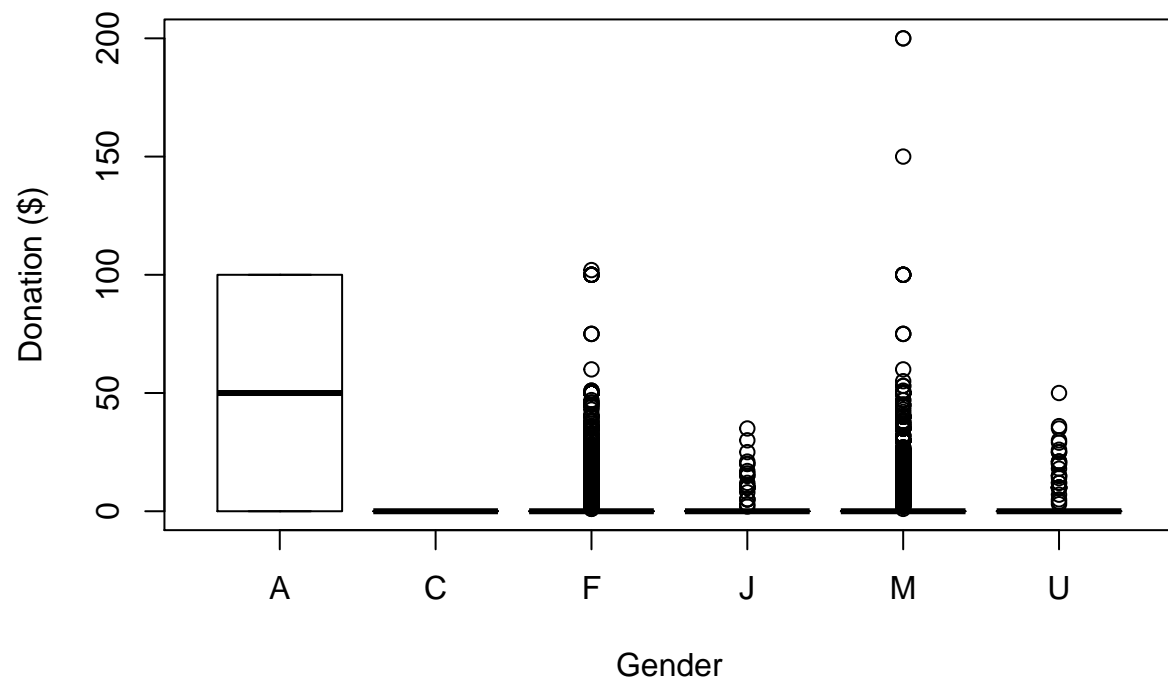
```
##  
##      0      1  
## 67234 3637
```

```
barplot(table(charityData$DONR),xlab="DONR")
```



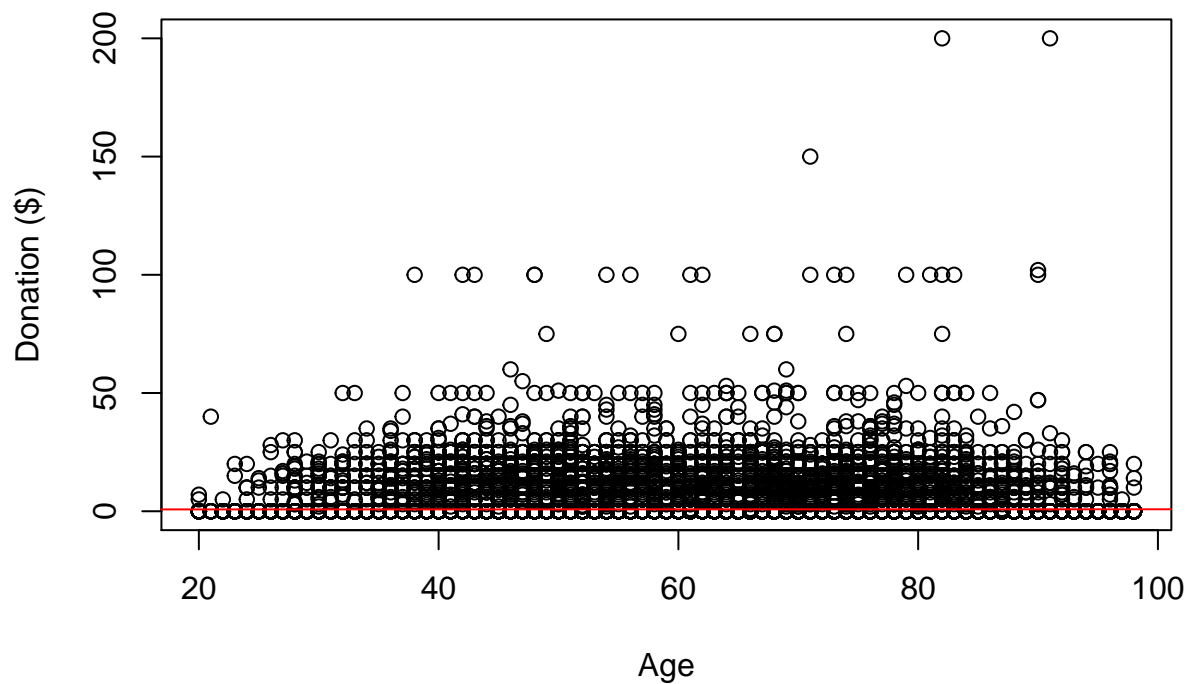
(b) Boxplot of DAMT amount by categories for GENDER

```
plot(charityData$GENDER,charityData$DAMT,xlab="Gender",ylab="Donation ($)")
```



Plot DAMT against a quantitative predictor variable

```
plot(charityData$AGE,charityData$DAMT,xlab="Age",ylab="Donation ($)")
lm_age = lm(DAMT ~ AGE, data=charityData)
abline(lm_age,col="red")
```



Let's see which variable have the greatest potential for the regression.

```
charity_damt<-subset(charityData,DAMT > 0,select=c(DAMT,AGE,MEDAGE,MEDPPH,MEDHVAL,MEDINC,MEDEDUC,NUMPRM12,
,NUMPRM12,RAMNTALL,NGIFTALL,MAXRAMNT,LASTGIFT,TDON))
cor(charity_damt)
```

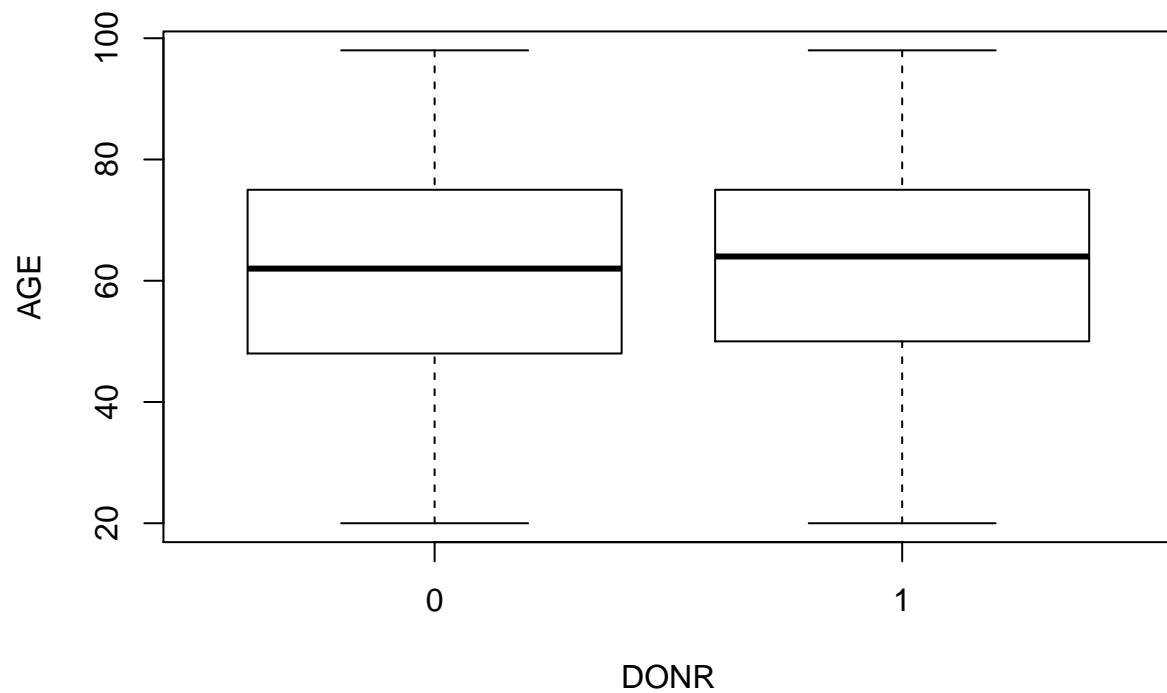
##		DAMT	AGE	MEDAGE	MEDPPH	MEDHVAL
##	DAMT	1.000000000	-0.03859743	-0.00706712	0.006517661	0.116305746
##	AGE	-0.038597430	1.00000000	0.21497461	-0.181142567	0.013425465
##	MEDAGE	-0.007067120	0.21497461	1.00000000	-0.260473255	0.071912722
##	MEDPPH	0.006517661	-0.18114257	-0.26047325	1.00000000	0.048079707
##	MEDHVAL	0.116305746	0.01342547	0.07191272	0.048079707	1.00000000
##	MEDINC	0.116530060	-0.09127951	0.03426462	0.333221110	0.716263379
##	MEDEDUC	0.098239525	-0.01074045	0.16332125	0.122422882	0.512371105
##	NUMPRM	-0.059724189	0.23035661	0.02778282	-0.012745549	-0.024119575
##	NUMPRM12	0.054758978	0.08488383	0.01944259	-0.022739982	0.043207453
##	RAMNTALL	0.242751660	0.14025651	0.01156837	-0.021236409	0.025717589
##	NGIFTALL	-0.226396326	0.21498789	0.01523318	-0.020918100	-0.081961843
##	MAXRAMNT	0.412984842	0.01636631	0.00722218	-0.009173518	0.077057583
##	LASTGIFT	0.722788561	-0.03901145	-0.00296728	0.016783856	0.145406985
##	TDON	0.108707853	-0.11618416	-0.03237160	0.040942543	0.003032744
##		MEDINC	MEDEDUC	NUMPRM	NUMPRM12	RAMNTALL
##	DAMT	0.11653006	0.09823952	-0.05972419	0.05475898	0.24275166
##	AGE	-0.09127951	-0.01074045	0.23035661	0.08488383	0.14025651
##	MEDAGE	0.03426462	0.16332125	0.02778282	0.01944259	0.01156837
##	MEDPPH	0.33322111	0.12242288	-0.01274555	-0.02273998	-0.02123641

## MEDHVAL	0.71626338	0.51237111	-0.02411958	0.04320745	0.02571759
## MEDINC	1.00000000	0.64654757	-0.02469300	0.03057805	0.01689872
## MEDEDUC	0.64654757	1.00000000	-0.02984508	0.01427124	0.01217791
## NUMPROM	-0.02469300	-0.02984508	1.00000000	0.52153590	0.66785709
## NUMPRM12	0.03057805	0.01427124	0.52153590	1.00000000	0.40913289
## RAMNTALL	0.01689872	0.01217791	0.66785709	0.40913289	1.00000000
## NGIFTALL	-0.08754909	-0.08206455	0.78496064	0.30351573	0.63483633
## MAXRAMNT	0.06194634	0.04595540	0.11018119	0.17012424	0.47948167
## LASTGIFT	0.13025977	0.10734222	-0.02537302	0.09191341	0.31429272
## TDON	0.02254817	0.00251612	-0.25994471	-0.59919769	-0.20810152
##	NGIFTALL	MAXRAMNT	LASTGIFT	TDON	
## DAMT	-0.22639633	0.412984842	0.72278856	0.108707853	
## AGE	0.21498789	0.016366314	-0.03901145	-0.116184159	
## MEDAGE	0.01523318	0.007222180	-0.00296728	-0.032371603	
## MEDPPH	-0.02091810	-0.009173518	0.01678386	0.040942543	
## MEDHVAL	-0.08196184	0.077057583	0.14540699	0.003032744	
## MEDINC	-0.08754909	0.061946336	0.13025977	0.022548165	
## MEDEDUC	-0.08206455	0.045955398	0.10734222	0.002516120	
## NUMPROM	0.78496064	0.110181186	-0.02537302	-0.259944707	
## NUMPRM12	0.30351573	0.170124239	0.09191341	-0.599197687	
## RAMNTALL	0.63483633	0.479481665	0.31429272	-0.208101522	
## NGIFTALL	1.00000000	0.014362033	-0.21244193	-0.241757439	
## MAXRAMNT	0.01436203	1.000000000	0.53381124	-0.008665962	
## LASTGIFT	-0.21244193	0.533811240	1.00000000	0.100962876	
## TDON	-0.24175744	-0.008665962	0.10096288	1.000000000	

LASTGIFT and MAXRAMNT have the greatest potential for the regression problem.

(c) Boxplot of AGE by DONR status

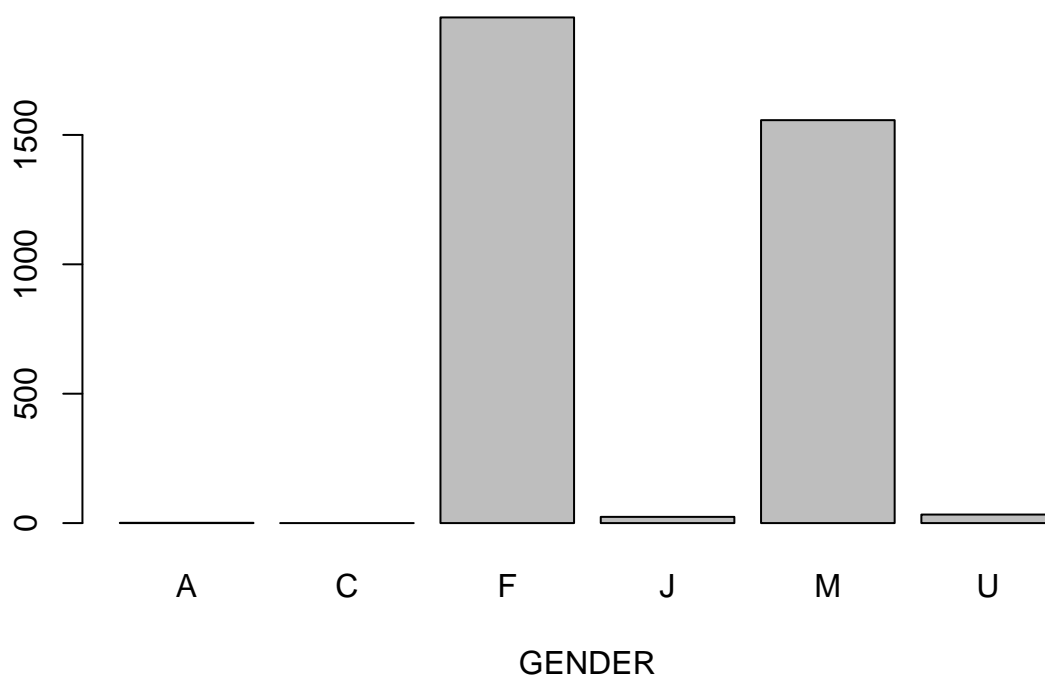
```
plot(charityData$DONR,charityData$AGE,xlab="DONR",ylab="AGE")
```



Wrong" Way. There are more females than males in the dataset as a whole.

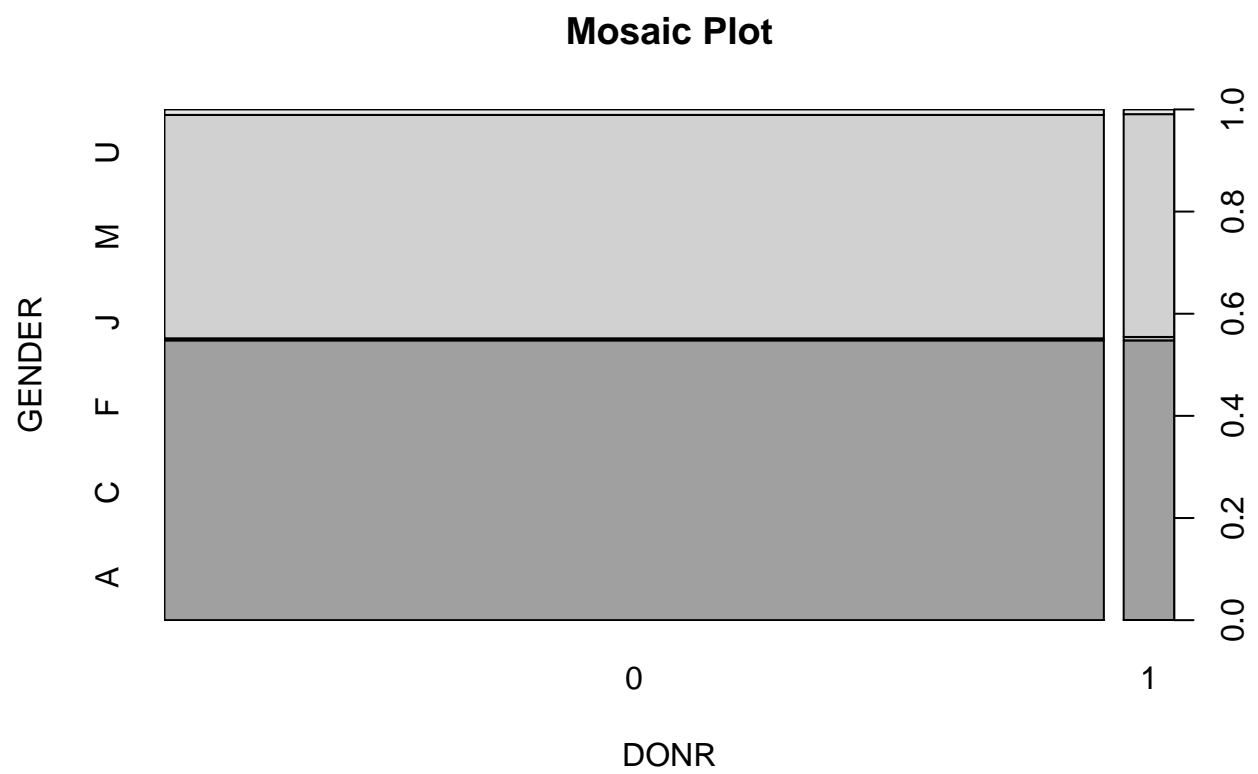
```
barplot(table(charityData$GENDER[charityData$DONR == 1]),xlab="GENDER",main="Barplot of GENDER for DONR
```

Barplot of GENDER for DONR = 1

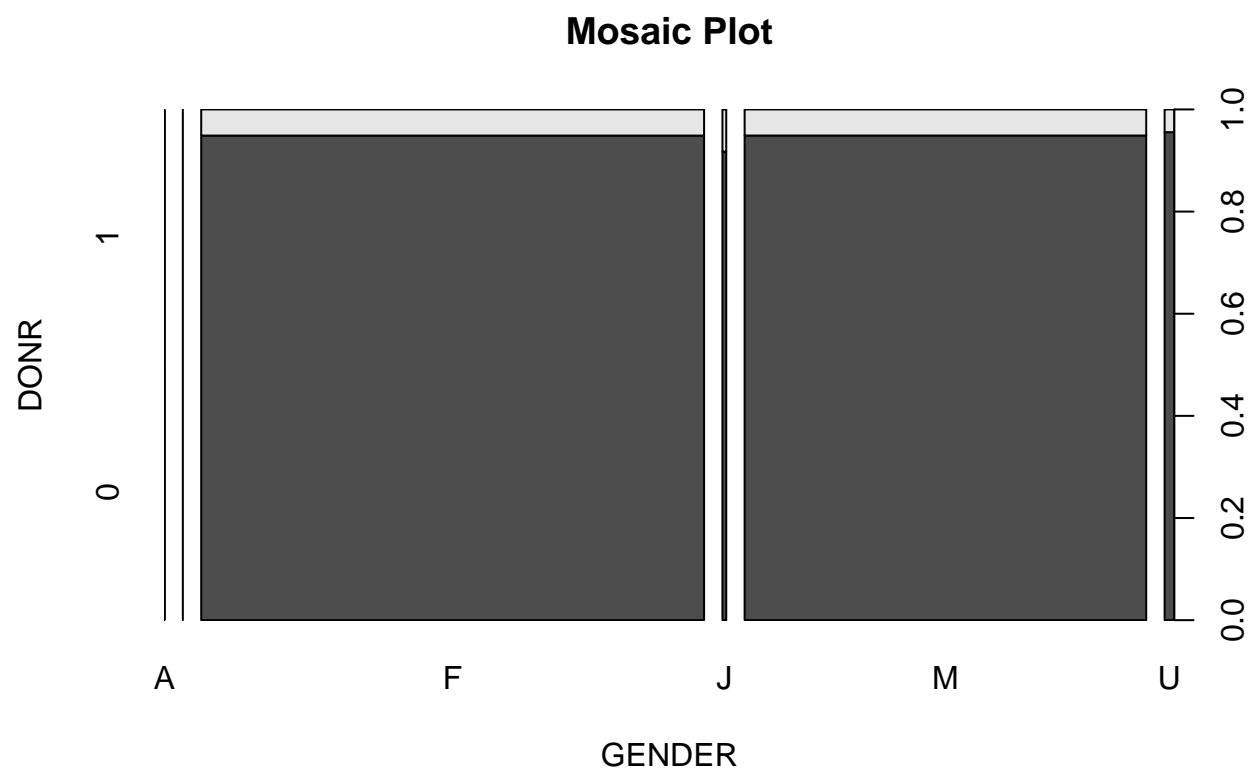


Right Way # A mosaic plot is obtained when we plot one factor variable against another. The # mosaic plot represents the counts as proportions to the whole. A deviation in # overall proportion of females donating compared to males donating is meaningful # whereas the absolute count of females donating compared to males donating was not.

```
plot(charityData$DONR,charityData$GENDER,xlab="DONR",ylab="GENDER",main="Mosaic Plot")
```

```
# Or  
plot(charityData$GENDER,charityData$DONR,xlab="GENDER",ylab="DONR",main="Mosaic Plot")
```



These graphs show that M/F doesn't show any difference in DONR status.

```
charityData = separateRFA(charityData,"RFA_96")
# Check the results
table(charityData$RFA_96,charityData$RFA_96_R)
```

```
##
##      A      F      L      N      S
## A1D    69      0      0      0      0
## A1E  4102      0      0      0      0
## A1F 17247      0      0      0      0
## A1G  7246      0      0      0      0
## A2C     1      0      0      0      0
## A2D    60      0      0      0      0
## A2E  1457      0      0      0      0
## A2F  4839      0      0      0      0
## A2G  2094      0      0      0      0
## A3C     1      0      0      0      0
## A3D   379      0      0      0      0
## A3E  2913      0      0      0      0
## A3F  1524      0      0      0      0
## A3G   641      0      0      0      0
## A4C     3      0      0      0      0
## A4D  1296      0      0      0      0
## A4E  1281      0      0      0      0
## A4F   670      0      0      0      0
```

##	A4G	306	0	0	0	0
##	F1C	0	15	0	0	0
##	F1D	0	46	0	0	0
##	F1E	0	54	0	0	0
##	F1F	0	3935	0	0	0
##	F1G	0	1433	0	0	0
##	L1F	0	0	149	0	0
##	L1G	0	0	54	0	0
##	L2F	0	0	71	0	0
##	L2G	0	0	34	0	0
##	L3E	0	0	19	0	0
##	L3F	0	0	20	0	0
##	L3G	0	0	8	0	0
##	L4E	0	0	11	0	0
##	L4F	0	0	8	0	0
##	L4G	0	0	9	0	0
##	N1C	0	0	0	1	0
##	N1D	0	0	0	23	0
##	N1E	0	0	0	28	0
##	N1F	0	0	0	1026	0
##	N1G	0	0	0	438	0
##	N2C	0	0	0	7	0
##	N2D	0	0	0	38	0
##	N2E	0	0	0	100	0
##	N2F	0	0	0	1013	0
##	N2G	0	0	0	366	0
##	N3D	0	0	0	12	0
##	N3E	0	0	0	549	0
##	N3F	0	0	0	180	0
##	N3G	0	0	0	63	0
##	N4D	0	0	0	22	0
##	N4E	0	0	0	120	0
##	N4F	0	0	0	40	0
##	N4G	0	0	0	15	0
##	S2B	0	0	0	0	2
##	S2C	0	0	0	0	2
##	S2D	0	0	0	0	157
##	S2E	0	0	0	0	2688
##	S2F	0	0	0	0	2055
##	S2G	0	0	0	0	893
##	S3C	0	0	0	0	2
##	S3D	0	0	0	0	1539
##	S3E	0	0	0	0	1757
##	S3F	0	0	0	0	856
##	S3G	0	0	0	0	402
##	S4B	0	0	0	0	2
##	S4C	0	0	0	0	7
##	S4D	0	0	0	0	2027
##	S4E	0	0	0	0	1386
##	S4F	0	0	0	0	741
##	S4G	0	0	0	0	319

```
table(charityData$RFA_96,charityData$RFA_96_F)
```

##				
##		1	2	3
##			4	
##	A1D	69	0	0
##	A1E	4102	0	0
##	A1F	17247	0	0
##	A1G	7246	0	0
##	A2C	0	1	0
##	A2D	0	60	0
##	A2E	0	1457	0
##	A2F	0	4839	0
##	A2G	0	2094	0
##	A3C	0	0	1
##	A3D	0	0	379
##	A3E	0	0	2913
##	A3F	0	0	1524
##	A3G	0	0	641
##	A4C	0	0	0
##	A4D	0	0	0
##	A4E	0	0	0
##	A4F	0	0	0
##	A4G	0	0	0
##	F1C	15	0	0
##	F1D	46	0	0
##	F1E	54	0	0
##	F1F	3935	0	0
##	F1G	1433	0	0
##	L1F	149	0	0
##	L1G	54	0	0
##	L2F	0	71	0
##	L2G	0	34	0
##	L3E	0	0	19
##	L3F	0	0	20
##	L3G	0	0	8
##	L4E	0	0	0
##	L4F	0	0	0
##	L4G	0	0	0
##	N1C	1	0	0
##	N1D	23	0	0
##	N1E	28	0	0
##	N1F	1026	0	0
##	N1G	438	0	0
##	N2C	0	7	0
##	N2D	0	38	0
##	N2E	0	100	0
##	N2F	0	1013	0
##	N2G	0	366	0
##	N3D	0	0	12
##	N3E	0	0	549
##	N3F	0	0	180
##	N3G	0	0	63
##	N4D	0	0	0
##	N4E	0	0	0
##	N4F	0	0	0
##	N4G	0	0	0

```
## S2B 0 2 0 0
## S2C 0 2 0 0
## S2D 0 157 0 0
## S2E 0 2688 0 0
## S2F 0 2055 0 0
## S2G 0 893 0 0
## S3C 0 0 2 0
## S3D 0 0 1539 0
## S3E 0 0 1757 0
## S3F 0 0 856 0
## S3G 0 0 402 0
## S4B 0 0 0 2
## S4C 0 0 0 7
## S4D 0 0 0 2027
## S4E 0 0 0 1386
## S4F 0 0 0 741
## S4G 0 0 0 319
```

```
table(charityData$RFA_96,charityData$RFA_96_A)
```

```
##
##      B      C      D      E      F      G
## A1D 0 0 69 0 0 0
## A1E 0 0 0 4102 0 0
## A1F 0 0 0 0 17247 0
## A1G 0 0 0 0 0 7246
## A2C 0 1 0 0 0 0
## A2D 0 0 60 0 0 0
## A2E 0 0 0 1457 0 0
## A2F 0 0 0 0 4839 0
## A2G 0 0 0 0 0 2094
## A3C 0 1 0 0 0 0
## A3D 0 0 379 0 0 0
## A3E 0 0 0 2913 0 0
## A3F 0 0 0 0 1524 0
## A3G 0 0 0 0 0 641
## A4C 0 3 0 0 0 0
## A4D 0 0 1296 0 0 0
## A4E 0 0 0 1281 0 0
## A4F 0 0 0 0 670 0
## A4G 0 0 0 0 0 306
## F1C 0 15 0 0 0 0
## F1D 0 0 46 0 0 0
## F1E 0 0 0 54 0 0
## F1F 0 0 0 0 3935 0
## F1G 0 0 0 0 0 1433
## L1F 0 0 0 0 149 0
## L1G 0 0 0 0 0 54
## L2F 0 0 0 0 71 0
## L2G 0 0 0 0 0 34
## L3E 0 0 0 19 0 0
## L3F 0 0 0 0 20 0
## L3G 0 0 0 0 0 8
## L4E 0 0 0 11 0 0
```

##	L4F	0	0	0	0	8	0
##	L4G	0	0	0	0	0	9
##	N1C	0	1	0	0	0	0
##	N1D	0	0	23	0	0	0
##	N1E	0	0	0	28	0	0
##	N1F	0	0	0	0	1026	0
##	N1G	0	0	0	0	0	438
##	N2C	0	7	0	0	0	0
##	N2D	0	0	38	0	0	0
##	N2E	0	0	0	100	0	0
##	N2F	0	0	0	0	1013	0
##	N2G	0	0	0	0	0	366
##	N3D	0	0	12	0	0	0
##	N3E	0	0	0	549	0	0
##	N3F	0	0	0	0	180	0
##	N3G	0	0	0	0	0	63
##	N4D	0	0	22	0	0	0
##	N4E	0	0	0	120	0	0
##	N4F	0	0	0	0	40	0
##	N4G	0	0	0	0	0	15
##	S2B	2	0	0	0	0	0
##	S2C	0	2	0	0	0	0
##	S2D	0	0	157	0	0	0
##	S2E	0	0	0	2688	0	0
##	S2F	0	0	0	0	2055	0
##	S2G	0	0	0	0	0	893
##	S3C	0	2	0	0	0	0
##	S3D	0	0	1539	0	0	0
##	S3E	0	0	0	1757	0	0
##	S3F	0	0	0	0	856	0
##	S3G	0	0	0	0	0	402
##	S4B	2	0	0	0	0	0
##	S4C	0	7	0	0	0	0
##	S4D	0	0	2027	0	0	0
##	S4E	0	0	0	1386	0	0
##	S4F	0	0	0	0	741	0
##	S4G	0	0	0	0	0	319

Exercise 4 First we will subset the charity data set and exclude all the categorical variables. Not sure how to handle mix of continuous and categorical variables. We only include DAMT values greater than 0.

```
charity_sub<-subset(charityData,DAMT > 0,select=c(AGE,MEDAGE,MEDPPH,MEDHVAL,MEDINC,MEDEDUC,NUMPROM
,NUMPRM12,RAMNTALL,NGIFTALL,MAXRAMNT,LASTGIFT,TDON))
```

Let's look at mean and variance of the variables. It

```
sapply(charity_sub,mean)
```

##	AGE	MEDAGE	MEDPPH	MEDHVAL	MEDINC	MEDEDUC
##	62.39126	42.52543	185.60682	1165.70003	360.15947	129.53148
##	NUMPROM	NUMPRM12	RAMNTALL	NGIFTALL	MAXRAMNT	LASTGIFT
##	51.87930	13.32279	115.18877	11.68985	18.27874	15.17563
##	TDON					
##	17.70993					

```
supply(charity_sub, var)
```

```
##          AGE          MEDAGE          MEDPPH          MEDHVAL          MEDINC
## 2.481018e+02 6.539794e+01 2.354235e+03 1.055142e+06 3.049313e+04
##          MEDEDUC          NUMPROM          NUMPRM12          RAMNTALL          NGIFTALL
## 2.831281e+02 5.138163e+02 2.555749e+01 1.275299e+04 8.500444e+01
##          MAXRAMNT          LASTGIFT          TDON
## 4.474424e+02 1.268361e+02 1.759488e+01
```

We now perform principal component analysis using prcomp function.

```
pr.out<-prcomp(charity_sub,scale=TRUE)
pr.out$sdev
```

```
## [1] 1.7468653 1.5615092 1.2753145 1.1910868 1.0645638 0.8807616 0.8437433
## [8] 0.6911847 0.6813230 0.6053720 0.4460684 0.4425991 0.3752499
```

The center and scale variables correspond to mean and standard deviations of variable prior to PCA.

```
pr.out$center
```

```
##          AGE          MEDAGE          MEDPPH          MEDHVAL          MEDINC          MEDEDUC
## 62.39126 42.52543 185.60682 1165.70003 360.15947 129.53148
##          NUMPROM          NUMPRM12          RAMNTALL          NGIFTALL          MAXRAMNT          LASTGIFT
## 51.87930 13.32279 115.18877 11.68985 18.27874 15.17563
##          TDON
## 17.70993
```

```
pr.out$scale
```

```
##          AGE          MEDAGE          MEDPPH          MEDHVAL          MEDINC          MEDEDUC
## 15.751248 8.086899 48.520463 1027.201029 174.622834 16.826410
##          NUMPROM          NUMPRM12          RAMNTALL          NGIFTALL          MAXRAMNT          LASTGIFT
## 22.667516 5.055442 112.929153 9.219785 21.152834 11.262151
##          TDON
## 4.194625
```

```
pr.out$rotation
```

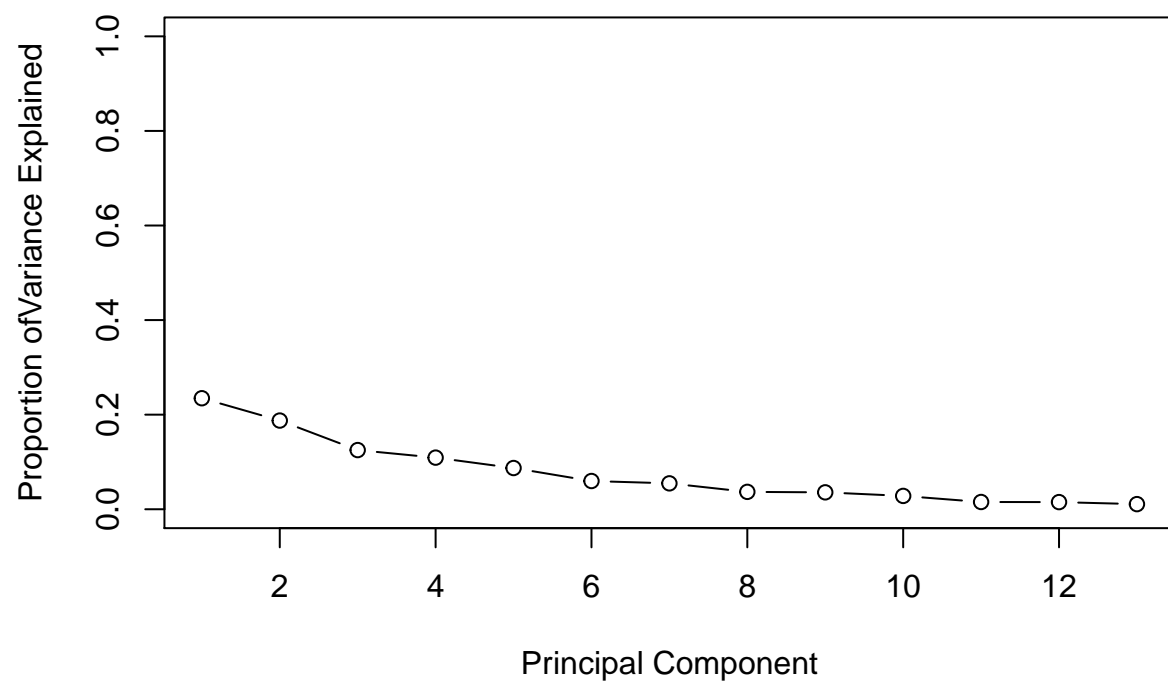
```
##          PC1          PC2          PC3          PC4          PC5
## AGE      0.17981397 0.06195935 0.14354868 -0.43790261 0.172836557
## MEDAGE   0.04359051 -0.05164401 0.13650338 -0.62358498 0.002226247
## MEDPPH   -0.05336775 -0.19045402 0.03538389 0.58273954 0.115523629
## MEDHVAL  -0.01830202 -0.51271462 0.17564772 -0.10455262 0.015440314
## MEDINC   -0.04045772 -0.56720783 0.19880445 0.08969708 0.059428940
## MEDEDUC  -0.03025628 -0.49317982 0.20952796 -0.11114998 0.039765987
## NUMPROM   0.49689818 0.01850947 0.11913819 0.08974645 0.238985165
## NUMPRM12 0.39873651 -0.05916862 0.07282530 0.08201776 -0.514500737
## RAMNTALL 0.47667400 -0.09692337 -0.23060492 0.03790078 0.208804271
## NGIFTALL 0.44893195 0.09744319 0.20099196 0.10929265 0.379334933
```

```

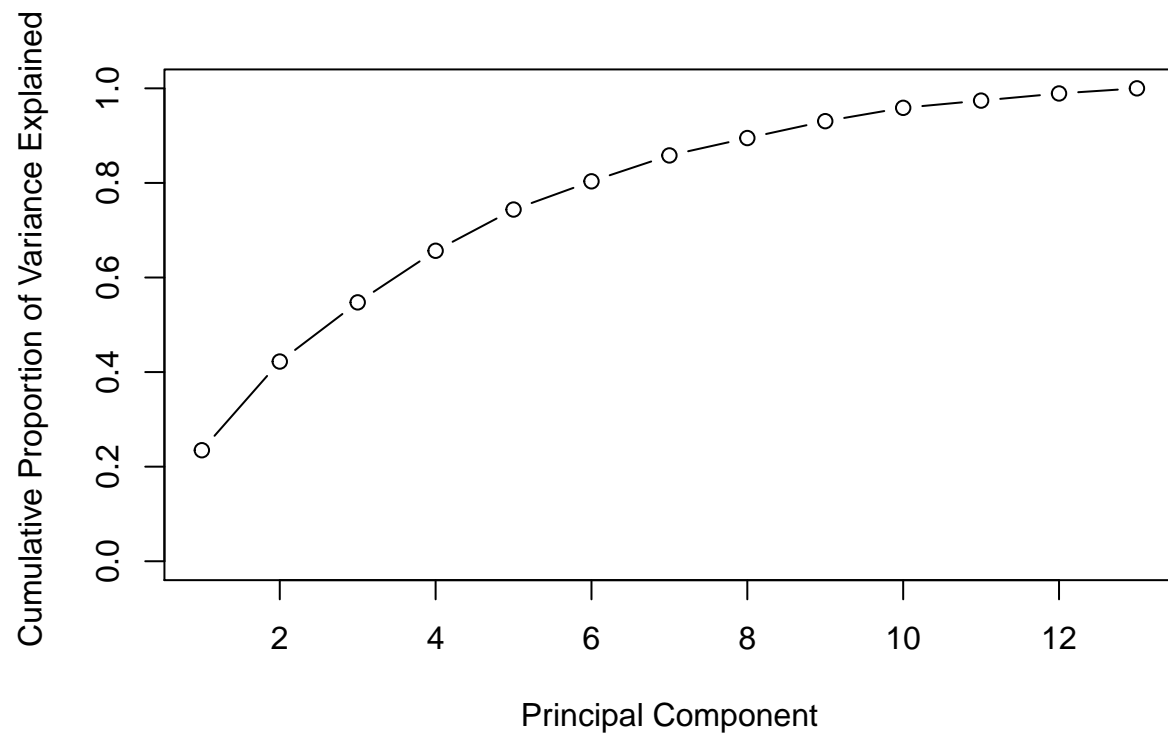
## MAXRAMNT  0.19153112 -0.19875009 -0.57204491 -0.08549752 -0.038346487
## LASTGIFT  0.06441568 -0.25851848 -0.59320900 -0.10335678 -0.097926038
## TDON      -0.29075986 -0.01810023 -0.23027194 -0.04807731  0.656541110
##           PC6          PC7          PC8          PC9          PC10
## AGE       -0.834713823  0.1058180067 -0.01417260  0.05596316 -0.056250405
## MEDAGE    0.251390795 -0.6757429585  0.16258097 -0.19911380  0.033115988
## MEDPPH    -0.400603136 -0.6003448089  0.08590519 -0.11801232  0.019802428
## MEDHVAL   0.028966866  0.3999744560  0.26621732 -0.43806101  0.064052641
## MEDINC    -0.014974483  0.0144252802  0.06532814 -0.13843591 -0.008776728
## MEDEDUC   0.089713537 -0.0762350486 -0.46818726  0.62929252 -0.068664885
## NUMPROM   0.096426396  0.0089423619  0.22980430  0.14304286 -0.185288722
## NUMPRM12  0.008076649 -0.0230706235  0.30198933  0.14263393 -0.551102324
## RAMNTALL  0.114637318  0.0003830476 -0.09472328 -0.01854907  0.232188875
## NGIFTALL  0.146290047  0.0226111334 -0.10389267 -0.07347749  0.224312287
## MAXRAMNT -0.055944872 -0.0552792290 -0.49543516 -0.38339398 -0.345961164
## LASTGIFT -0.098843528 -0.0223449809  0.42750812  0.35466357  0.400833979
## TDON      0.116855254  0.0328931032  0.28223248  0.12999231 -0.513636142
##           PC11          PC12          PC13
## AGE       -0.05991367  0.05515000  0.022758480
## MEDAGE    0.02797040  0.01226849 -0.001455955
## MEDPPH    0.24095252  0.07717345 -0.009310533
## MEDHVAL   0.50463386  0.10828806 -0.026359967
## MEDINC    -0.75722700 -0.16464540  0.033096933
## MEDEDUC   0.24972566  0.04128453  0.013028624
## NUMPROM   0.09566517 -0.48612483 -0.557617229
## NUMPRM12 -0.01997478  0.23172717  0.301358758
## RAMNTALL -0.16598410  0.69854752 -0.270710137
## NGIFTALL  0.06683942 -0.17951971  0.683821455
## MAXRAMNT  0.06345266 -0.25200189  0.044109949
## LASTGIFT  0.03762744 -0.21429823  0.173273045
## TDON      -0.01215741  0.17513972  0.151043619

```

```
biplot(pr.out,scale=0)
```

```
plot(cumsum (pve ), xlab=" Principal Component ", ylab ="Cumulative Proportion of Variance Explained ",
```



Looks like first 6 principal components explain 80% of the variance in