Preprocessing.Rmd

Christopher Himmel

February 8, 2019

#Introduction

This document is in excerpt from the original Sales\_Prediction, removed and put in its own document for readability. Here we go through each feature and show a bargraph, determining whether there is any information available in the feature to contribute to the model.

#Set up R environment

library(tm)  
library(dplyr)  
library(wordcloud)  
library(SnowballC)  
library(corrplot)  
library(rpart)  
library(rattle)  
library(randomForest)  
library(tidyr)  
library(ggplot2)  
setwd("C:/Users/Christopher/Dropbox/Deep Learning/Projects/Sales-Prediction")

#Upload data pulled out of MMS

Load and view Item Master (INVMST):

Item\_Master = read.csv("data/INVMST - Item Master.csv")  
View(Item\_Master)  
summary(Item\_Master)

## INUMBR ICHECK IDESCR IDSCCD   
## Min. : 1 Min. :0.0000 EVERYDAY CARD: 15339 A:296655   
## 1st Qu.: 458835 1st Qu.:0.0000 GENERAL : 6360 D: 3   
## Median : 583973 Median :0.0000 BDY GEN : 4776 I:198184   
## Mean : 576680 Mean :0.3784 BDY HUM : 4325 N: 1   
## 3rd Qu.: 709587 3rd Qu.:0.0000 BDY CONV : 3070 P: 1286   
## Max. :7777777 Max. :9.0000 ROMANTIC : 2071   
## (Other) :460188   
## ISORT ISTYLN ASNUM   
## CARDS CARL: 68244 Mode:logical Min. : 1   
## CARDS A.G.: 54544 NA's:496129 1st Qu.: 13925   
## CARDS NEUT: 36108 Median : 13925   
## CARDS RPG : 27411 Mean : 52938   
## CARDS TNDR: 25614 3rd Qu.: 18773   
## CARDS TARG: 15667 Max. :957170   
## (Other) :268541   
## IVNDPN IMFGR IMFGNO IDEPT   
## :333989 Min. :0 :165478 Min. :100.0   
## #N/A : 607 1st Qu.:0 1043 : 28407 1st Qu.:100.0   
## CATCHALL : 34 Median :0 1042 : 26704 Median :100.0   
## VENDOR ITE: 22 Mean :0 9011 : 15386 Mean :182.7   
## SCHWAM CATCHALL: 17 3rd Qu.:0 5693 : 8690 3rd Qu.:230.0   
## 9788883707: 13 Max. :0 9029 : 7685 Max. :999.0   
## (Other) :161447 (Other):243779   
## ISDEPT ICLAS ISCLAS BYRNUM   
## Min. : 1.00 Min. : 1.0 Min. : 1.000 Min. : 0.0   
## 1st Qu.: 1.00 1st Qu.: 95.0 1st Qu.: 1.000 1st Qu.: 1.0   
## Median : 2.00 Median :101.0 Median : 1.000 Median : 1.0   
## Mean : 11.96 Mean :115.7 Mean : 9.855 Mean : 2.2   
## 3rd Qu.: 5.00 3rd Qu.:107.0 3rd Qu.: 3.000 3rd Qu.: 3.0   
## Max. :999.00 Max. :999.0 Max. :999.000 Max. :14.0   
## NA's :86735   
## IASPAC IADVYN IBUYCD ISET   
## Min. :0.0e+00 :485192 : 2 Min. :0.000000   
## 1st Qu.:0.0e+00 Y: 10937 C:494704 1st Qu.:0.000000   
## Median :0.0e+00 S: 1423 Median :0.000000   
## Mean :2.6e-05 Mean :0.005116   
## 3rd Qu.:0.0e+00 3rd Qu.:0.000000   
## Max. :1.3e+01 Max. :7.000000   
##   
## IWARNT IPRMPT IPRVNT ITKTTR ITKTTA   
## Min. :1.00 : 569 N:490323 Min. :0.000 Min. :0   
## 1st Qu.:2.00 N:493760 Y: 5806 1st Qu.:0.000 1st Qu.:0   
## Median :2.00 Y: 1800 Median :0.000 Median :0   
## Mean :1.99 Mean :0.261 Mean :0   
## 3rd Qu.:2.00 3rd Qu.:1.000 3rd Qu.:0   
## Max. :2.00 Max. :1.000 Max. :1   
## NA's :133647 NA's :9212 NA's :74606   
## ITKTN ILBLTR ILBLTA IFINLN   
## Min. : 0.0000 Min. :0.000 Min. :0 :214023   
## 1st Qu.: 0.0000 1st Qu.:0.000 1st Qu.:0 CARDS :209098   
## Median : 0.0000 Median :0.000 Median :0 DISC : 12365   
## Mean : 0.2615 Mean :0.249 Mean :0 OTHER : 10656   
## 3rd Qu.: 1.0000 3rd Qu.:0.000 3rd Qu.:0 ACCES : 10360   
## Max. :11.0000 Max. :1.000 Max. :1 GIFT : 9602   
## NA's :8938 NA's :74599 (Other): 30025   
## IPROFL IMODUL ISTYPE ISCOLR   
## :489134 :475906 CM :346610 Min. :0.000000   
## FLAT : 6941 A: 20223 DM :129412 1st Qu.:0.000000   
## GUGUS : 34 IM : 11287 Median :0.000000   
## LEPEN : 11 S1 : 3327 Mean :0.006668   
## MEMKP : 4 NM : 1966 3rd Qu.:0.000000   
## BBEAR : 2 S3 : 987 Max. :7.000000   
## (Other): 3 (Other): 2540   
## ISSIZE IHAZCD MCHNUM ISUBST ICORE   
## :493577 Min. :3 Mode:logical Min. :0 Min. : 0.0   
## C: 2444 1st Qu.:3 NA's:496129 1st Qu.:0 1st Qu.: 0.0   
## P: 5 Median :3 Median :0 Median : 0.0   
## S: 103 Mean :3 Mean :0 Mean : 20.2   
## 3rd Qu.:3 3rd Qu.:0 3rd Qu.: 0.0   
## Max. :3 Max. :0 Max. :607500.0   
## NA's :496128   
## IREPL ISLUM IUMCV2 IBYUM   
## Min. :0 : 12 Min. :0.0000 EA :493798   
## 1st Qu.:0 001: 134 1st Qu.:1.0000 A01 : 634   
## Median :0 500: 1 Median :1.0000 001 : 190   
## Mean :0 A01: 1 Mean :0.9959 012 : 116   
## 3rd Qu.:0 EA :495981 3rd Qu.:1.0000 006 : 114   
## Max. :0 Max. :1.0000 500 : 109   
## (Other): 1168   
## IMINPK ISTDPK IHLDOR IMOQTY   
## Min. : 0.000 Min. : 0.00 : 65917 Min. : 0.0000   
## 1st Qu.: 1.000 1st Qu.: 1.00 N:430187 1st Qu.: 0.0000   
## Median : 1.000 Median : 1.00 Y: 25 Median : 0.0000   
## Mean : 2.613 Mean : 16.28 Mean : 0.2206   
## 3rd Qu.: 1.000 3rd Qu.: 2.00 3rd Qu.: 0.0000   
## Max. :5000.000 Max. :9800.00 Max. :500.0000   
##   
## IMNSTK IMXSTK IMDSTK IDSPLY   
## Min. :0.0e+00 Min. :0.0e+00 Min. : 0 Min. : 0.0000   
## 1st Qu.:0.0e+00 1st Qu.:0.0e+00 1st Qu.: 0 1st Qu.: 0.0000   
## Median :0.0e+00 Median :0.0e+00 Median : 0 Median : 0.0000   
## Mean :4.4e-03 Mean :4.5e+00 Mean : 24 Mean : 0.5193   
## 3rd Qu.:0.0e+00 3rd Qu.:0.0e+00 3rd Qu.: 0 3rd Qu.: 0.0000   
## Max. :5.0e+02 Max. :1.2e+06 Max. :4020040 Max. :500.0000   
##   
## IOMULT IRPLCD IMFPRC IWGHT   
## Min. :0.0e+00 Min. :0.00 Min. :0.0e+00 Min. : 0.0000   
## 1st Qu.:0.0e+00 1st Qu.:0.00 1st Qu.:0.0e+00 1st Qu.: 0.0000   
## Median :0.0e+00 Median :0.00 Median :0.0e+00 Median : 0.0000   
## Mean :4.1e+00 Mean :0.07 Mean :4.7e-05 Mean : 0.0951   
## 3rd Qu.:0.0e+00 3rd Qu.:0.00 3rd Qu.:0.0e+00 3rd Qu.: 0.0000   
## Max. :1.0e+06 Max. :6.00 Max. :7.5e+00 Max. :3001.4000   
## NA's :8430   
## ICUBE IDLGTH IDWDTH   
## Min. : 0.0 Min. : 0.000 Min. : 0.0000   
## 1st Qu.: 0.0 1st Qu.: 0.000 1st Qu.: 0.0000   
## Median : 0.0 Median : 0.000 Median : 0.0000   
## Mean : 189.8 Mean : 1.607 Mean : 0.3503   
## 3rd Qu.: 0.0 3rd Qu.: 0.000 3rd Qu.: 0.0000   
## Max. :112896.0 Max. :9918.700 Max. :3000.0000   
##   
## IDHGHT IMDATE IMCENT IVPLTI   
## Min. : 0.000 Min. : 101 Min. :0.000 Min. : 0.0000   
## 1st Qu.: 0.000 1st Qu.: 81016 1st Qu.:1.000 1st Qu.: 0.0000   
## Median : 0.000 Median :100622 Median :1.000 Median : 0.0000   
## Mean : 1.213 Mean :105550 Mean :0.999 Mean : 0.1722   
## 3rd Qu.: 0.000 3rd Qu.:130517 3rd Qu.:1.000 3rd Qu.: 0.0000   
## Max. :3125.300 Max. :991208 Max. :1.000 Max. :301.0000   
##   
## IVPLHI SHPNUM IDISTM IHLDWO IIGNUM   
## Min. : 0.0000 Min. :0 Min. :1.00 : 6495 Min. :0   
## 1st Qu.: 0.0000 1st Qu.:0 1st Qu.:3.00 N:489633 1st Qu.:0   
## Median : 0.0000 Median :0 Median :3.00 Y: 1 Median :0   
## Mean : 0.0905 Mean :0 Mean :2.89 Mean :0   
## 3rd Qu.: 0.0000 3rd Qu.:0 3rd Qu.:3.00 3rd Qu.:0   
## Max. :801.0000 Max. :0 Max. :4.00 Max. :0   
## NA's :280180   
## ISDIM IVATCD IPLAN IVLRK1 IVLRK2   
## :496126 Mode:logical Min. :0 :481602 :495756   
## N: 1 NA's:496129 1st Qu.:0 A: 5772 B: 1   
## C : 2 Median :0 B: 702 C: 372   
## Mean :0 C: 8053   
## 3rd Qu.:0   
## Max. :0   
##   
## IVLRK3 IVLRK4 IRPLCN IRPLDT   
## Mode:logical Mode:logical Min. :0.000000 Min. : 0   
## NA's:496129 NA's:496129 1st Qu.:0.000000 1st Qu.: 0   
## Median :0.000000 Median : 0   
## Mean :0.004444 Mean : 265   
## 3rd Qu.:0.000000 3rd Qu.: 0   
## Max. :1.000000 Max. :90116   
##   
## ISEASN IDEADC IDEADD   
## :310591 Min. :0.00000 Min. : 0.00   
## XMS : 53671 1st Qu.:0.00000 1st Qu.: 0.00   
## VAL : 29511 Median :0.00000 Median : 0.00   
## MOD : 22969 Mean :0.00126 Mean : 98.66   
## EAS : 19046 3rd Qu.:0.00000 3rd Qu.: 0.00   
## FAD : 17685 Max. :1.00000 Max. :100610.00   
## (Other): 42656   
## INLRTL IHANDL IATRB1 IATRB2   
## Min. : 0.0000 Mode:logical :493576 :496108   
## 1st Qu.: 0.0000 NA's:496129 GF : 709 CF: 19   
## Median : 0.0000 EN : 459 XU: 2   
## Mean : 0.3011 PK : 382   
## 3rd Qu.: 0.0000 BG : 321   
## Max. :3000.0000 KT : 301   
## (Other): 381   
## IATRB3 IATRB4 IATRB5 IPRCCH IPRCZN IPRCST   
## US :309035 :254685 : 4013 B: 79985 N:496129 B: 8255   
## :121991 0:237300 N:485128 N: 5397 N:204584   
## CN : 49960 1: 4142 Y: 6988 P: 68 P:281730   
## GB : 3053 O: 2 R:410676 R: 1560   
## IN : 1498 Y: 3   
## KR : 1459   
## (Other): 9133   
## IASNUM ICORGP ILEAD IHZCOD   
## Min. :0 :406064 Min. : 0.000 Mode:logical   
## 1st Qu.:0 CARDS : 18240 1st Qu.: 0.000 NA's:496129   
## Median :0 SEASN : 13033 Median : 0.000   
## Mean :0 COLLE : 11770 Mean : 1.994   
## 3rd Qu.:0 ASST : 8407 3rd Qu.: 0.000   
## Max. :0 HOMDE : 7637 Max. :800.000   
## (Other): 30978   
## IFRACT IMCRDT IMCRCN   
## Mode:logical Min. : 0 Min. :0.0000   
## NA's:496129 1st Qu.: 80103 1st Qu.:1.0000   
## Median :100511 Median :1.0000   
## Mean : 95556 Mean :0.9157   
## 3rd Qu.:130124 3rd Qu.:1.0000   
## Max. :151208 Max. :1.0000   
##

Load and view Chain Level Inventory Balance file (INVCBL):

Inv\_Bal\_Chain = read.csv("data/INVCBL - Chain Level Inventory Balance Data.csv")  
View(Inv\_Bal\_Chain)  
summary(Inv\_Bal\_Chain)

## INUMBR CBHBOY CBHBOP   
## Min. : 0 Min. : -1392.0 Min. :-82678.0   
## 1st Qu.: 431436 1st Qu.: 0.0 1st Qu.: 0.0   
## Median : 546387 Median : 0.0 Median : 0.0   
## Mean : 551447 Mean : 81.1 Mean : 89.6   
## 3rd Qu.: 690742 3rd Qu.: 3.0 3rd Qu.: 5.0   
## Max. :1009699 Max. :1267000.0 Max. :960000.0   
## CBHAND CBTRNS CBGMRI CBWKCR   
## Min. :-91973.0 Min. :0 Min. :0 Min. : -80.0000   
## 1st Qu.: 0.0 1st Qu.:0 1st Qu.:0 1st Qu.: 0.0000   
## Median : 0.0 Median :0 Median :0 Median : 0.0000   
## Mean : 82.6 Mean :0 Mean :0 Mean : 0.9551   
## 3rd Qu.: 5.0 3rd Qu.:0 3rd Qu.:0 3rd Qu.: 0.0000   
## Max. :830000.0 Max. :0 Max. :0 Max. :1511.0000   
## CBWK01 CBWK02 CBWK03   
## Min. : -6.000 Min. : -10.00 Min. : -4.000   
## 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 0.000   
## Median : 0.000 Median : 0.00 Median : 0.000   
## Mean : 3.392 Mean : 2.48 Mean : 1.686   
## 3rd Qu.: 0.000 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :9959.000 Max. :8463.00 Max. :4632.000   
## CBWK04 CBWK05 CBWK06   
## Min. : -4.000 Min. :-249.000 Min. : -10.000   
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.000   
## Median : 0.000 Median : 0.000 Median : 0.000   
## Mean : 1.565 Mean : 1.484 Mean : 1.297   
## 3rd Qu.: 0.000 3rd Qu.: 0.000 3rd Qu.: 0.000   
## Max. :4030.000 Max. :4543.000 Max. :3805.000   
## CBWK07 CBWK08 CBRSUP   
## Min. : -6.000 Min. : -17.000 Min. : -8.000   
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.000   
## Median : 0.000 Median : 0.000 Median : 0.000   
## Mean : 1.142 Mean : 1.186 Mean : 6.568   
## 3rd Qu.: 0.000 3rd Qu.: 0.000 3rd Qu.: 0.000   
## Max. :3290.000 Max. :2475.000 Max. :19725.000   
## CBRSUY CBRSDP CBRSDY CBVRSP   
## Min. : -9.00 Min. : -650.0 Min. : -380.0 Min. :0   
## 1st Qu.: 0.00 1st Qu.: 0.0 1st Qu.: 0.0 1st Qu.:0   
## Median : 0.00 Median : 0.0 Median : 0.0 Median :0   
## Mean : 63.15 Mean : 48.8 Mean : 416.9 Mean :0   
## 3rd Qu.: 1.00 3rd Qu.: 0.0 3rd Qu.: 1.0 3rd Qu.:0   
## Max. :91973.00 Max. :55948.3 Max. :471348.6 Max. :0   
## CBVRSY CBRCSP CBRCSY CBASUP   
## Min. :0 Min. : -260.00 Min. : -72.0 Min. : 0.0000   
## 1st Qu.:0 1st Qu.: 0.00 1st Qu.: 0.0 1st Qu.: 0.0000   
## Median :0 Median : 0.00 Median : 0.0 Median : 0.0000   
## Mean :0 Mean : 13.81 Mean : 101.5 Mean : 0.2592   
## 3rd Qu.:0 3rd Qu.: 0.00 3rd Qu.: 0.0 3rd Qu.: 0.0000   
## Max. :0 Max. :34028.80 Max. :61024.0 Max. :1276.0000   
## CBASUY CBASDP CBASDY CBVASP   
## Min. : -8.000 Min. : -0.400 Min. : -137.20 Min. :0   
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.:0   
## Median : 0.000 Median : 0.000 Median : 0.00 Median :0   
## Mean : 5.275 Mean : 2.456 Mean : 19.89 Mean :0   
## 3rd Qu.: 0.000 3rd Qu.: 0.000 3rd Qu.: 0.00 3rd Qu.:0   
## Max. :5693.000 Max. :10028.720 Max. :32947.34 Max. :0   
## CBVASY CBACSP CBACSY CBRETS   
## Min. :0 Min. : 0.000 Min. : -49.50 Min. : 0.0000   
## 1st Qu.:0 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 0.0000   
## Median :0 Median : 0.000 Median : 0.00 Median : 0.0000   
## Mean :0 Mean : 1.291 Mean : 13.98 Mean : 0.5125   
## 3rd Qu.:0 3rd Qu.: 0.000 3rd Qu.: 0.00 3rd Qu.: 0.0000   
## Max. :0 Max. :6176.600 Max. :20465.52 Max. :1075.0000   
## CBRCUP CBRCUY CBRCDP CBRCDY   
## Min. :-405.000 Min. : -552 Min. : -338.352 Min. : -1171   
## 1st Qu.: 0.000 1st Qu.: 0 1st Qu.: 0.000 1st Qu.: 0   
## Median : 0.000 Median : 0 Median : 0.000 Median : 0   
## Mean : 2.781 Mean : 103 Mean : 4.339 Mean : 158   
## 3rd Qu.: 0.000 3rd Qu.: 0 3rd Qu.: 0.000 3rd Qu.: 0   
## Max. :8800.000 Max. :1656000 Max. :15640.800 Max. :285360   
## CBTIUP CBTIUY CBTOUP   
## Min. : -15.000 Min. : -9.00 Min. :-2.06e+05   
## 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 0.00e+00   
## Median : 0.000 Median : 0.00 Median : 0.00e+00   
## Mean : 0.916 Mean : 23.82 Mean :-3.74e+00   
## 3rd Qu.: 0.000 3rd Qu.: 0.00 3rd Qu.: 0.00e+00   
## Max. :10300.000 Max. :51598.00 Max. : 1.50e+01   
## CBTOUY CBTFDP CBTFDY   
## Min. :-2654000.0 Min. :-24039.600 Min. :-221315.33   
## 1st Qu.: 0.0 1st Qu.: 0.000 1st Qu.: 0.00   
## Median : 0.0 Median : 0.000 Median : 0.00   
## Mean : -54.6 Mean : -0.944 Mean : -1.23   
## 3rd Qu.: 0.0 3rd Qu.: 0.000 3rd Qu.: 0.00   
## Max. : 18.0 Max. : 89.600 Max. : 15216.67   
## CBRVUP CBRVUY CBRVDP CBRVDY   
## Min. :0 Min. :-3264.000 Min. :0 Min. : 0.0   
## 1st Qu.:0 1st Qu.: 0.000 1st Qu.:0 1st Qu.: 0.0   
## Median :0 Median : 0.000 Median :0 Median : 0.0   
## Mean :0 Mean : -0.039 Mean :0 Mean : 0.2   
## 3rd Qu.:0 3rd Qu.: 0.000 3rd Qu.:0 3rd Qu.: 0.0   
## Max. :0 Max. : 0.000 Max. :0 Max. :22470.0   
## CBAJUP CBAJUY CBAJDP   
## Min. :-1.5e+04 Min. :-353800.0 Min. :-1800.0000   
## 1st Qu.: 0.0e+00 1st Qu.: 0.0 1st Qu.: 0.0000   
## Median : 0.0e+00 Median : 0.0 Median : 0.0000   
## Mean :-9.1e-02 Mean : -2.5 Mean : -0.1197   
## 3rd Qu.: 0.0e+00 3rd Qu.: 0.0 3rd Qu.: 0.0000   
## Max. : 7.5e+01 Max. : 3148.0 Max. : 323.4000   
## CBAJDY CBPHUP CBPHUY   
## Min. :-42456.0 Min. :0e+00 Min. :-2450.0000   
## 1st Qu.: 0.0 1st Qu.:0e+00 1st Qu.: 0.0000   
## Median : 0.0 Median :0e+00 Median : 0.0000   
## Mean : -2.8 Mean :5e-05 Mean : 0.2301   
## 3rd Qu.: 0.0 3rd Qu.:0e+00 3rd Qu.: 0.0000   
## Max. : 4295.1 Max. :9e+00 Max. : 2500.0000   
## CBPHDP CBPHDY CBAVGC CBAVOH   
## Min. : 0.0000 Min. :-3.15e+03 Min. : 0.00 Min. :0   
## 1st Qu.: 0.0000 1st Qu.: 0.00e+00 1st Qu.: 0.13 1st Qu.:0   
## Median : 0.0000 Median : 0.00e+00 Median : 0.49 Median :0   
## Mean : 0.0006 Mean : 4.07e-02 Mean : 4.13 Mean :0   
## 3rd Qu.: 0.0000 3rd Qu.: 0.00e+00 3rd Qu.: 2.01 3rd Qu.:0   
## Max. :131.5602 Max. : 6.30e+02 Max. :169274.39 Max. :0   
## CBAVRT CBAVCS CBDEPT CBSDPT CBCLAS   
## Min. :0 Min. :0 Min. : 0 Min. : 0.00 Min. : 0.0   
## 1st Qu.:0 1st Qu.:0 1st Qu.:100 1st Qu.: 1.00 1st Qu.: 95.0   
## Median :0 Median :0 Median :100 Median : 2.00 Median :101.0   
## Mean :0 Mean :0 Mean :197 Mean : 11.91 Mean :113.3   
## 3rd Qu.:0 3rd Qu.:0 3rd Qu.:240 3rd Qu.: 6.00 3rd Qu.:107.0   
## Max. :0 Max. :0 Max. :999 Max. :999.00 Max. :999.0   
## CBSCLS CBVNDR CBSTYL CBPOOQ   
## Min. : 0.00 Min. : 0.00 Mode:logical Min. : 0.0   
## 1st Qu.: 1.00 1st Qu.: 0.00 NA's:338358 1st Qu.: 0.0   
## Median : 1.00 Median : 0.00 Median : 0.0   
## Mean : 11.54 Mean : 45.93 Mean : 20.2   
## 3rd Qu.: 4.00 3rd Qu.: 0.00 3rd Qu.: 0.0   
## Max. :999.00 Max. :61857.00 Max. :1386000.0   
## CBTOOQ CBINTQ CBMASU   
## Min. : 0.000 Min. : 0.000 Min. :-32477.00   
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.00   
## Median : 0.000 Median : 0.000 Median : 0.00   
## Mean : 0.626 Mean : 0.298 Mean : 4.43   
## 3rd Qu.: 0.000 3rd Qu.: 0.000 3rd Qu.: 3.00   
## Max. :5410.000 Max. :5406.000 Max. : 47660.00   
## CBMASD CBMASV CBMASC CBLAVC   
## Min. :-661526069 Min. :0 Min. :-123413 Min. : 0.0000   
## 1st Qu.: 0 1st Qu.:0 1st Qu.: 0 1st Qu.: 0.0000   
## Median : 0 Median :0 Median : 0 Median : 0.0000   
## Mean : -1934 Mean :0 Mean : 26 Mean : 0.2057   
## 3rd Qu.: 10 3rd Qu.:0 3rd Qu.: 2 3rd Qu.: 0.0000   
## Max. : 115916 Max. :0 Max. :5729640 Max. :1443.6333   
## CBARQT CBNSQT CBHLDQ CBTSCN   
## Min. : -75.00 Min. :0.0e+00 Min. :-1500.000 Min. :0.0000   
## 1st Qu.: 0.00 1st Qu.:0.0e+00 1st Qu.: 0.000 1st Qu.:0.0000   
## Median : 0.00 Median :0.0e+00 Median : 0.000 Median :1.0000   
## Mean : 0.77 Mean :9.3e-01 Mean : 0.023 Mean :0.5946   
## 3rd Qu.: 0.00 3rd Qu.:0.0e+00 3rd Qu.: 0.000 3rd Qu.:1.0000   
## Max. :47000.00 Max. :7.0e+04 Max. :10500.000 Max. :1.0000   
## CBTSDT CBTRCN CBTRDT CBCEN2   
## Min. : 0 Min. :0.0000 Min. : 0 Min. :0.0000   
## 1st Qu.: 0 1st Qu.:0.0000 1st Qu.: 0 1st Qu.:1.0000   
## Median : 80725 Median :1.0000 Median : 90227 Median :1.0000   
## Mean : 69544 Mean :0.6216 Mean : 72421 Mean :0.8139   
## 3rd Qu.:130722 3rd Qu.:1.0000 3rd Qu.:130814 3rd Qu.:1.0000   
## Max. :151215 Max. :1.0000 Max. :151215 Max. :1.0000   
## CBALD2 CBFRCN CBFRDT   
## Min. : 0 Min. :0.000 Min. : 0   
## 1st Qu.: 60412 1st Qu.:1.000 1st Qu.: 50315   
## Median : 90727 Median :1.000 Median : 80924   
## Mean : 86913 Mean :0.809 Mean : 83891   
## 3rd Qu.:130318 3rd Qu.:1.000 3rd Qu.:120723   
## Max. :991230 Max. :1.000 Max. :991230

Load and view Store Level Inventory Balance file, filtered for Niquea’D stores (INVBAL):

Inv\_Bal\_Store = read.csv("data/INVBAL - Store Level Inventory Balance Data - Niquea'D.csv")  
View(Inv\_Bal\_Store)  
summary(Inv\_Bal\_Store)

## INUMBR ISTORE IBHBOY IBHBOP   
## Min. : 12 Min. :1001 Min. : -9.000 Min. :-652.000   
## 1st Qu.: 698053 1st Qu.:1002 1st Qu.: 0.000 1st Qu.: 0.000   
## Median : 730006 Median :1003 Median : 0.000 Median : 0.000   
## Mean : 727337 Mean :1002 Mean : 2.022 Mean : 2.547   
## 3rd Qu.: 766055 3rd Qu.:1003 3rd Qu.: 2.000 3rd Qu.: 3.000   
## Max. :1009366 Max. :1003 Max. :200.000 Max. : 386.000   
## IBHAND IBTRNS IBGMRI IBWKCR   
## Min. :-767.000 Min. :0 Min. :0 Min. :-2.00000   
## 1st Qu.: 0.000 1st Qu.:0 1st Qu.:0 1st Qu.: 0.00000   
## Median : 0.000 Median :0 Median :0 Median : 0.00000   
## Mean : 2.369 Mean :0 Mean :0 Mean : 0.09702   
## 3rd Qu.: 3.000 3rd Qu.:0 3rd Qu.:0 3rd Qu.: 0.00000   
## Max. : 385.000 Max. :0 Max. :0 Max. :42.00000   
## IBWK01 IBWK02 IBWK03   
## Min. : -1.0000 Min. :-1.00000 Min. :-1.00000   
## 1st Qu.: 0.0000 1st Qu.: 0.00000 1st Qu.: 0.00000   
## Median : 0.0000 Median : 0.00000 Median : 0.00000   
## Mean : 0.1102 Mean : 0.07364 Mean : 0.04908   
## 3rd Qu.: 0.0000 3rd Qu.: 0.00000 3rd Qu.: 0.00000   
## Max. :128.0000 Max. :17.00000 Max. :16.00000   
## IBWK04 IBWK05 IBWK06   
## Min. :-6.00000 Min. :-1.00000 Min. :-6.00000   
## 1st Qu.: 0.00000 1st Qu.: 0.00000 1st Qu.: 0.00000   
## Median : 0.00000 Median : 0.00000 Median : 0.00000   
## Mean : 0.05923 Mean : 0.05473 Mean : 0.04459   
## 3rd Qu.: 0.00000 3rd Qu.: 0.00000 3rd Qu.: 0.00000   
## Max. :42.00000 Max. :16.00000 Max. :25.00000   
## IBWK07 IBWK08 IBNORD IBLEAD   
## Min. :-5.00000 Min. :-1.00000 Min. :0 Min. : 0.000   
## 1st Qu.: 0.00000 1st Qu.: 0.00000 1st Qu.:0 1st Qu.: 0.000   
## Median : 0.00000 Median : 0.00000 Median :0 Median : 0.000   
## Mean : 0.04192 Mean : 0.04819 Mean :0 Mean : 7.702   
## 3rd Qu.: 0.00000 3rd Qu.: 0.00000 3rd Qu.:0 3rd Qu.: 0.000   
## Max. :14.00000 Max. :30.00000 Max. :0 Max. :360.000   
## IBRSUP IBRSUY IBRSDP   
## Min. : -1.0000 Min. : -2.000 Min. :-650.000   
## 1st Qu.: 0.0000 1st Qu.: 0.000 1st Qu.: 0.000   
## Median : 0.0000 Median : 0.000 Median : 0.000   
## Mean : 0.2801 Mean : 2.194 Mean : 3.368   
## 3rd Qu.: 0.0000 3rd Qu.: 2.000 3rd Qu.: 0.000   
## Max. :180.0000 Max. :766.000 Max. :1540.000   
## IBRSDY IBVRSP IBVRSY IBRCSP   
## Min. : -380.00 Min. :0 Min. :0 Min. :-260.000   
## 1st Qu.: 0.00 1st Qu.:0 1st Qu.:0 1st Qu.: 0.000   
## Median : 0.00 Median :0 Median :0 Median : 0.000   
## Mean : 27.49 Mean :0 Mean :0 Mean : 1.332   
## 3rd Qu.: 28.00 3rd Qu.:0 3rd Qu.:0 3rd Qu.: 0.000   
## Max. :18711.10 Max. :0 Max. :0 Max. :1400.000   
## IBRCSY IBASUP IBASUY   
## Min. : -89.000 Min. :0.00000 Min. :-2.00000   
## 1st Qu.: 0.000 1st Qu.:0.00000 1st Qu.: 0.00000   
## Median : 0.000 Median :0.00000 Median : 0.00000   
## Mean : 11.472 Mean :0.00079 Mean : 0.07145   
## 3rd Qu.: 7.976 3rd Qu.:0.00000 3rd Qu.: 0.00000   
## Max. :1400.000 Max. :9.00000 Max. :45.00000   
## IBASDP IBASDY IBVASP IBVASY   
## Min. : 0.00000 Min. :-25.9000 Min. :0 Min. :0   
## 1st Qu.: 0.00000 1st Qu.: 0.0000 1st Qu.:0 1st Qu.:0   
## Median : 0.00000 Median : 0.0000 Median :0 Median :0   
## Mean : 0.00391 Mean : 0.2708 Mean :0 Mean :0   
## 3rd Qu.: 0.00000 3rd Qu.: 0.0000 3rd Qu.:0 3rd Qu.:0   
## Max. :73.63000 Max. : 84.0000 Max. :0 Max. :0   
## IBACSP IBACSY IBRETS   
## Min. : 0.00000 Min. : -7.8000 Min. : 0.00000   
## 1st Qu.: 0.00000 1st Qu.: 0.0000 1st Qu.: 0.00000   
## Median : 0.00000 Median : 0.0000 Median : 0.00000   
## Mean : 0.00171 Mean : 0.1714 Mean : 0.02545   
## 3rd Qu.: 0.00000 3rd Qu.: 0.0000 3rd Qu.: 0.00000   
## Max. :38.25000 Max. :175.0000 Max. :15.00000   
## IBRCUP IBRCUY IBRCDP IBRCDY   
## Min. :-6.00000 Min. : -6.000 Min. :-32.000 Min. : -2.970   
## 1st Qu.: 0.00000 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.000   
## Median : 0.00000 Median : 0.000 Median : 0.000 Median : 0.000   
## Mean : 0.06755 Mean : 2.041 Mean : 0.258 Mean : 9.138   
## 3rd Qu.: 0.00000 3rd Qu.: 0.000 3rd Qu.: 0.000 3rd Qu.: 0.000   
## Max. :24.00000 Max. :400.000 Max. :179.400 Max. :3565.820   
## IBTIUP IBTIUY IBTOUP   
## Min. :-1.00000 Min. : -2.0000 Min. :-37.00000   
## 1st Qu.: 0.00000 1st Qu.: 0.0000 1st Qu.: 0.00000   
## Median : 0.00000 Median : 0.0000 Median : 0.00000   
## Mean : 0.04854 Mean : 0.9286 Mean : -0.01264   
## 3rd Qu.: 0.00000 3rd Qu.: 0.0000 3rd Qu.: 0.00000   
## Max. :37.00000 Max. :204.0000 Max. : 0.00000   
## IBTOUY IBTFDP IBTFDY IBRVUP   
## Min. :-198.0000 Min. :-190.0000 Min. :-1325.016 Min. :0   
## 1st Qu.: 0.0000 1st Qu.: 0.0000 1st Qu.: 0.000 1st Qu.:0   
## Median : 0.0000 Median : 0.0000 Median : 0.000 Median :0   
## Mean : -0.3189 Mean : 0.0914 Mean : 1.846 Mean :0   
## 3rd Qu.: 0.0000 3rd Qu.: 0.0000 3rd Qu.: 0.000 3rd Qu.:0   
## Max. : 0.0000 Max. : 120.0000 Max. : 1458.000 Max. :0   
## IBRVUY IBRVDP IBRVDY IBAJUP   
## Min. :0 Min. :0 Min. :0 Min. :-4.000000   
## 1st Qu.:0 1st Qu.:0 1st Qu.:0 1st Qu.: 0.000000   
## Median :0 Median :0 Median :0 Median : 0.000000   
## Mean :0 Mean :0 Mean :0 Mean :-0.001259   
## 3rd Qu.:0 3rd Qu.:0 3rd Qu.:0 3rd Qu.: 0.000000   
## Max. :0 Max. :0 Max. :0 Max. : 0.000000   
## IBAJUY IBAJDP IBAJDY IBPHUP   
## Min. :-65.00000 Min. :-33.73590 Min. :-354.000 Min. :0   
## 1st Qu.: 0.00000 1st Qu.: 0.00000 1st Qu.: 0.000 1st Qu.:0   
## Median : 0.00000 Median : 0.00000 Median : 0.000 Median :0   
## Mean : -0.04856 Mean : -0.00612 Mean : -0.261 Mean :0   
## 3rd Qu.: 0.00000 3rd Qu.: 0.00000 3rd Qu.: 0.000 3rd Qu.:0   
## Max. : 38.00000 Max. : 0.00000 Max. : 45.900 Max. :0   
## IBPHUY IBPHDP IBPHDY IBCEN1   
## Min. :-14.00000 Min. :0 Min. :-2.04e+03 Min. :0.0000   
## 1st Qu.: 0.00000 1st Qu.:0 1st Qu.: 0.00e+00 1st Qu.:0.0000   
## Median : 0.00000 Median :0 Median : 0.00e+00 Median :1.0000   
## Mean : 0.01071 Mean :0 Mean :-2.40e-03 Mean :0.5615   
## 3rd Qu.: 0.00000 3rd Qu.:0 3rd Qu.: 0.00e+00 3rd Qu.:1.0000   
## Max. : 63.00000 Max. :0 Max. : 3.15e+02 Max. :1.0000   
## IBALD1 IBCEN2 IBALD2 IBCEN3   
## Min. : 0 Min. :0.0000 Min. : 0 Min. :0.0000   
## 1st Qu.: 0 1st Qu.:0.0000 1st Qu.: 0 1st Qu.:0.0000   
## Median :130829 Median :1.0000 Median :130405 Median :0.0000   
## Mean : 79799 Mean :0.5367 Mean : 76299 Mean :0.2674   
## 3rd Qu.:141210 3rd Qu.:1.0000 3rd Qu.:141104 3rd Qu.:1.0000   
## Max. :160611 Max. :1.0000 Max. :151217 Max. :1.0000   
## IBALD3 IBCEN4 IBALD4 IBCEN5   
## Min. : 0 Min. :0.000 Min. : 0 Min. :0.0000   
## 1st Qu.: 0 1st Qu.:1.000 1st Qu.:131123 1st Qu.:1.0000   
## Median : 0 Median :1.000 Median :150107 Median :1.0000   
## Mean : 39110 Mean :0.792 Mean :115909 Mean :0.9711   
## 3rd Qu.:131018 3rd Qu.:1.000 3rd Qu.:150903 3rd Qu.:1.0000   
## Max. :151217 Max. :1.000 Max. :151217 Max. :1.0000   
## IBALD5 IBAVGC IBQLRC IBAVOH   
## Min. : 0 Min. : 0.00 Min. :-54.000 Min. :0   
## 1st Qu.:140420 1st Qu.: 1.20 1st Qu.: 0.000 1st Qu.:0   
## Median :150118 Median : 5.00 Median : 1.000 Median :0   
## Mean :140706 Mean : 17.66 Mean : 3.211 Mean :0   
## 3rd Qu.:150801 3rd Qu.: 16.67 3rd Qu.: 6.000 3rd Qu.:0   
## Max. :160611 Max. :1400.00 Max. :360.000 Max. :0   
## IBAVRT IBAVCS IBOSCR IBOS01 IBOS02 IBOS03 IBOS04   
## Min. :0 Min. :0 :18040 :18394 :18609 :18717 :18642   
## 1st Qu.:0 1st Qu.:0 Y:22466 Y:22112 Y:21897 Y:21789 Y:21864   
## Median :0 Median :0   
## Mean :0 Mean :0   
## 3rd Qu.:0 3rd Qu.:0   
## Max. :0 Max. :0   
## IBOS05 IBOS06 IBOS07 IBOS08 IBDEPT IBSDPT   
## :18722 :18628 :18626 :18901 Min. :100 Min. : 1.00   
## Y:21784 Y:21878 Y:21880 Y:21605 1st Qu.:150 1st Qu.: 3.00   
## Median :300 Median : 12.00   
## Mean :250 Mean : 12.25   
## 3rd Qu.:300 3rd Qu.: 12.00   
## Max. :900 Max. :100.00   
## IBCLAS IBSCLS IBFRCN IBFRDT   
## Min. : 1.00 Min. : 1.0 Min. :0.0000 Min. : 0   
## 1st Qu.: 3.00 1st Qu.: 1.0 1st Qu.:0.0000 1st Qu.: 0   
## Median : 6.00 Median : 26.0 Median :1.0000 Median :130329   
## Mean : 54.53 Mean : 53.1 Mean :0.5367 Mean : 75010   
## 3rd Qu.:103.00 3rd Qu.:103.0 3rd Qu.:1.0000 3rd Qu.:140711   
## Max. :901.00 Max. :140.0 Max. :1.0000 Max. :151217   
## IBVNDR IBSTYL IBPOOQ IBTOOQ   
## Min. :0 Mode:logical Min. : 0.00000 Min. : 0.000000   
## 1st Qu.:0 NA's:40506 1st Qu.: 0.00000 1st Qu.: 0.000000   
## Median :0 Median : 0.00000 Median : 0.000000   
## Mean :0 Mean : 0.01955 Mean : 0.004814   
## 3rd Qu.:0 3rd Qu.: 0.00000 3rd Qu.: 0.000000   
## Max. :0 Max. :12.00000 Max. :12.000000   
## IBINTQ IBMASU IBMASD IBMASV   
## Min. : 0.000000 Min. :-32.00000 Min. :-275.000 Min. :0   
## 1st Qu.: 0.000000 1st Qu.: 0.00000 1st Qu.: 0.000 1st Qu.:0   
## Median : 0.000000 Median : 0.00000 Median : 0.000 Median :0   
## Mean : 0.004567 Mean : 0.01612 Mean : 0.414 Mean :0   
## 3rd Qu.: 0.000000 3rd Qu.: 0.00000 3rd Qu.: 0.000 3rd Qu.:0   
## Max. :12.000000 Max. : 8.00000 Max. :8169.580 Max. :0   
## IBMASC IBLAVC IBARQT IBNSQT   
## Min. :-100.00000 Min. : -75.0001 Min. :0 Min. :0   
## 1st Qu.: 0.00000 1st Qu.: 0.8863 1st Qu.:0 1st Qu.:0   
## Median : 0.00000 Median : 4.9500 Median :0 Median :0   
## Mean : 0.09123 Mean : 17.6149 Mean :0 Mean :0   
## 3rd Qu.: 0.00000 3rd Qu.: 16.8300 3rd Qu.:0 3rd Qu.:0   
## Max. : 186.00000 Max. :1400.0000 Max. :0 Max. :0   
## IBPHND IBHLDQ IBCRCD IBTSCN IBTSDT   
## Min. :0 Min. :0 Mode:logical Min. :0.0000 Min. : 0   
## 1st Qu.:0 1st Qu.:0 NA's:40506 1st Qu.:0.0000 1st Qu.: 0   
## Median :0 Median :0 Median :0.0000 Median : 0   
## Mean :0 Mean :0 Mean :0.1587 Mean : 23031   
## 3rd Qu.:0 3rd Qu.:0 3rd Qu.:0.0000 3rd Qu.: 0   
## Max. :0 Max. :0 Max. :1.0000 Max. :151215   
## IBTRCN IBTRDT   
## Min. :0.0000 Min. : 0   
## 1st Qu.:0.0000 1st Qu.: 0   
## Median :0.0000 Median : 0   
## Mean :0.4383 Mean : 60925   
## 3rd Qu.:1.0000 3rd Qu.:131218   
## Max. :1.0000 Max. :151216

#Format data set

Create chain level field with combined regular sales and advertised sales:

Inv\_Bal\_Chain\_mut <-  
 Inv\_Bal\_Chain %>%  
 mutate(chn\_sales=CBRSUY+CBASUY)

Create total Niquea’D summarized sales with combined regular sales and advertised sales:

Inv\_Bal\_Store\_sum <-   
 Inv\_Bal\_Store %>%  
 group\_by(INUMBR) %>%  
 summarise(nd\_sales=sum(IBRSUY+IBASUY),  
 IBHAND\_sum=sum(IBHAND),  
 IBWKCR\_sum=sum(IBWKCR),  
 IBWK01\_sum=sum(IBWK01),  
 IBWK02\_sum=sum(IBWK02),  
 IBWK03\_sum=sum(IBWK03),  
 IBWK04\_sum=sum(IBWK04),  
 IBWK05\_sum=sum(IBWK05),  
 IBWK06\_sum=sum(IBWK06),  
 IBWK07\_sum=sum(IBWK07),  
 IBWK08\_sum=sum(IBWK08))  
Inv\_Bal\_Store\_sum <- subset(Inv\_Bal\_Store\_sum, IBHAND\_sum!=0 | nd\_sales!=0)

Combine two new sales numbers into one file by SKU:

Inv\_values <-  
 Inv\_Bal\_Store\_sum %>%  
 left\_join(Inv\_Bal\_Chain\_mut, by="INUMBR") %>%  
 select(INUMBR,nd\_sales,chn\_sales,IBHAND\_sum,IBWKCR\_sum,IBWK01\_sum,IBWK02\_sum,IBWK03\_sum,IBWK04\_sum,  
 IBWK05\_sum,IBWK06\_sum,IBWK07\_sum,IBWK08\_sum)

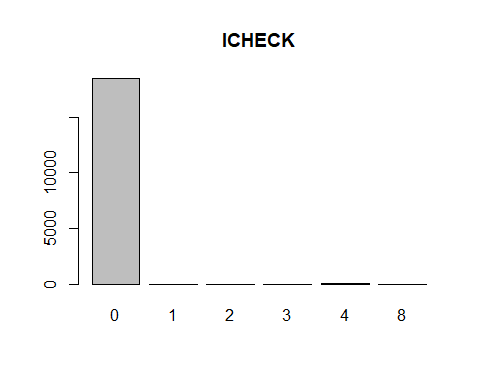
Combine SKU, sales data into one file for analysis:

Inv\_final <-  
 Inv\_values %>%  
 left\_join(Item\_Master, by="INUMBR")

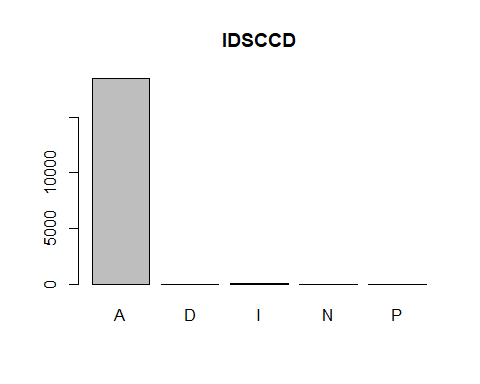
#Test each attribute for significant amount of data

Visualize each column of data:

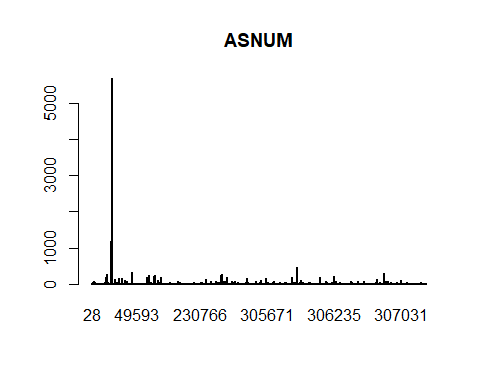
# IDESCR - factors of descriptions  
# ICHECK - all Zeros  
barplot(table(Inv\_final$ICHECK),main="ICHECK")



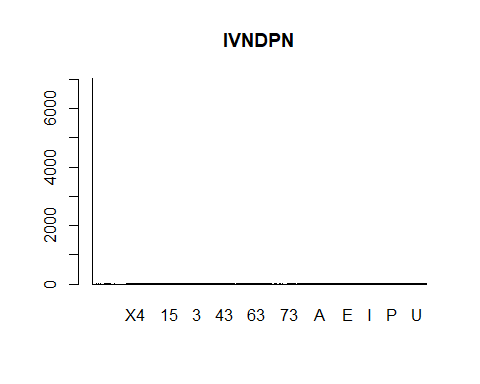
# IDSCCD - minimal values  
barplot(table(Inv\_final$IDSCCD),main="IDSCCD")



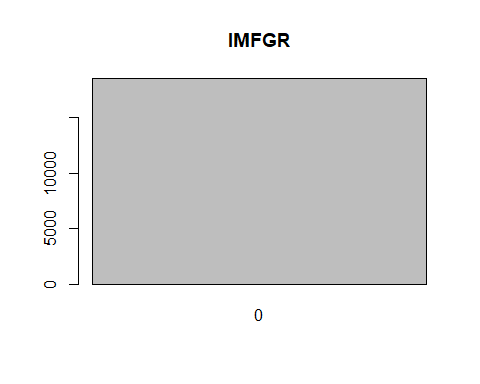
# ISORT - factors of descriptions  
# ISTYLN - all NA's  
# ASNUM - distributed values  
barplot(table(Inv\_final$ASNUM),main="ASNUM")



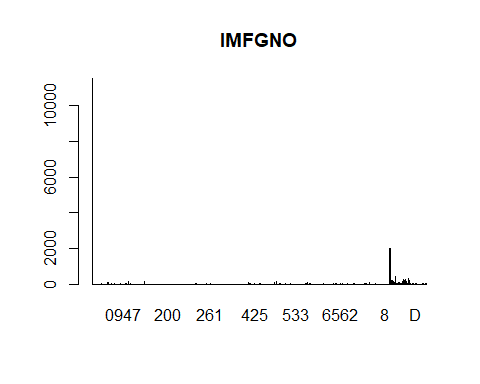
# IVNDPN - empty  
barplot(table(Inv\_final$IVNDPN),main="IVNDPN")



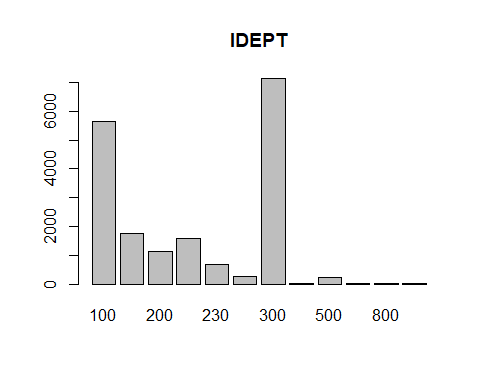
# IMFGR - empty  
barplot(table(Inv\_final$IMFGR),main="IMFGR")



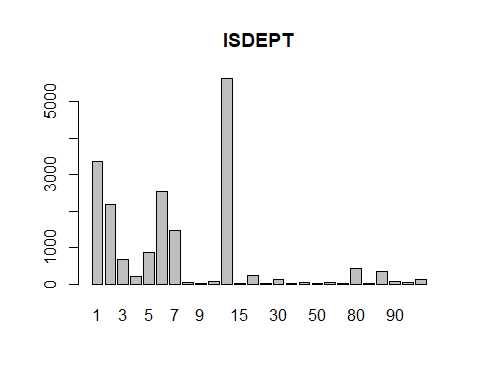
# IMFGNO - some values  
barplot(table(Inv\_final$IMFGNO),main="IMFGNO")



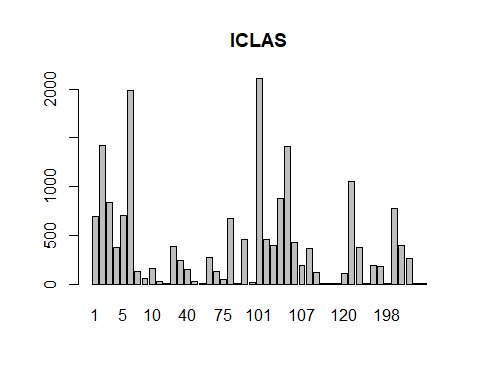
# IDEPT - use  
barplot(table(Inv\_final$IDEPT),main="IDEPT")



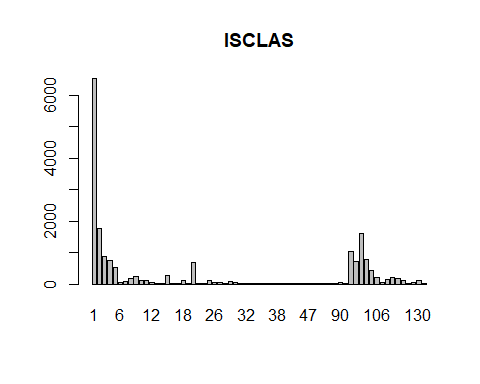
# ISDEPT - use  
barplot(table(Inv\_final$ISDEPT),main="ISDEPT")



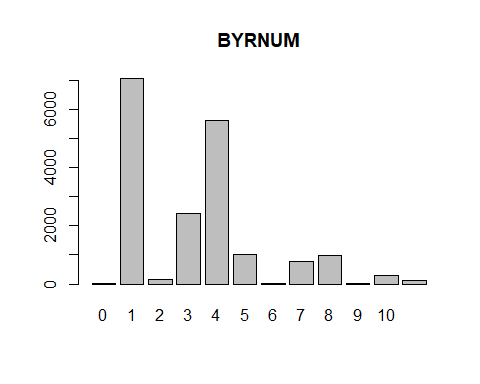
# ICLAS - use  
barplot(table(Inv\_final$ICLAS),main="ICLAS")



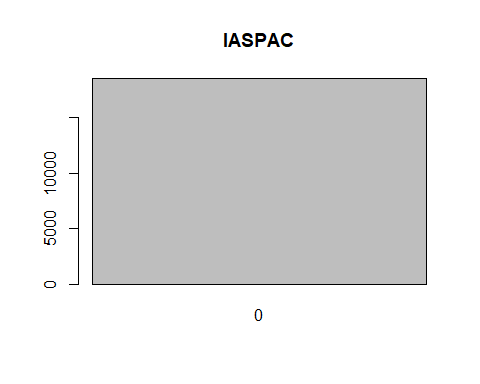
# ISCLAS - use  
barplot(table(Inv\_final$ISCLAS),main="ISCLAS")



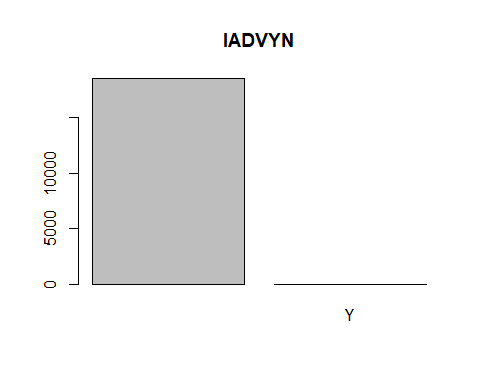
# BYRNUM - convert NA's to 0  
Inv\_final$BYRNUM[is.na(Inv\_final$BYRNUM)]<-0  
barplot(table(Inv\_final$BYRNUM),main="BYRNUM")



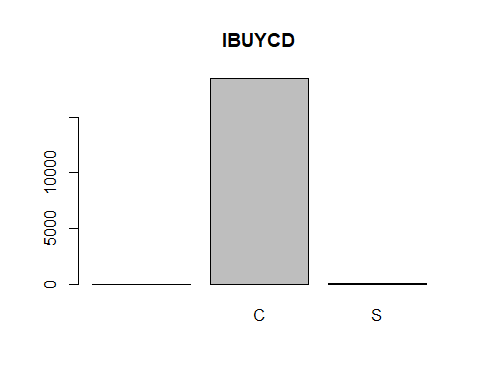
# IASPAC - empty  
barplot(table(Inv\_final$IASPAC),main="IASPAC")



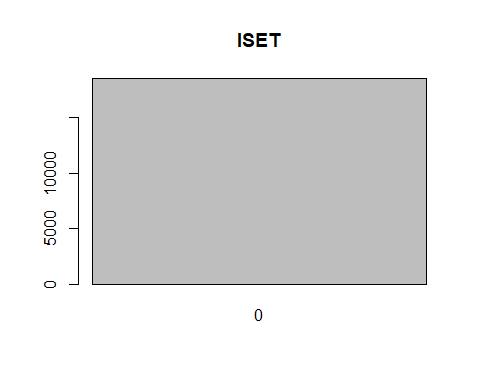
# IADVYN - empty  
barplot(table(Inv\_final$IADVYN),main="IADVYN")



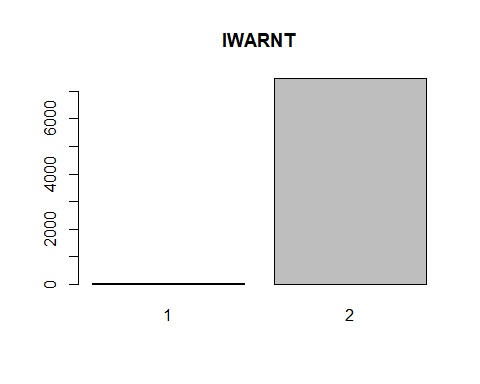
# IBUYCD - mostly empty  
barplot(table(Inv\_final$IBUYCD),main="IBUYCD")



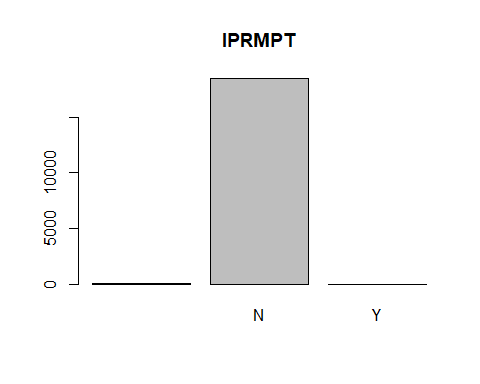
# ISET - empty  
barplot(table(Inv\_final$ISET),main="ISET")



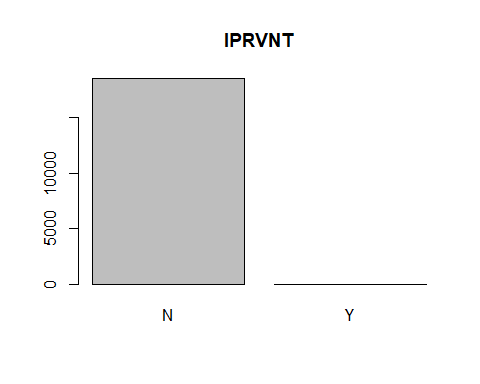
# IWARNT - mostly empty  
barplot(table(Inv\_final$IWARNT),main="IWARNT")



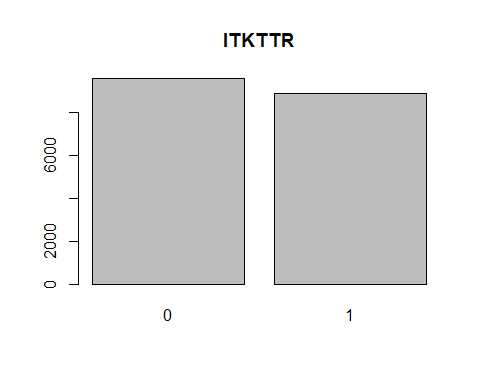
# IPRMPT - mostly empty  
barplot(table(Inv\_final$IPRMPT),main="IPRMPT")



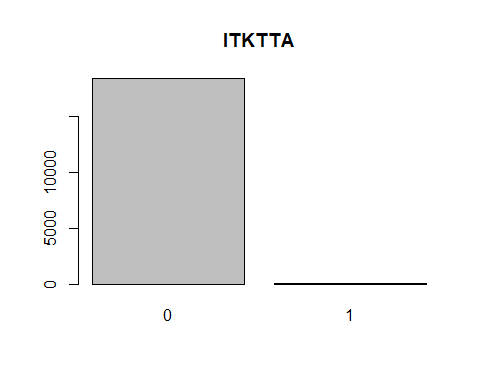
# IPRVNT - mostly empty  
barplot(table(Inv\_final$IPRVNT),main="IPRVNT")



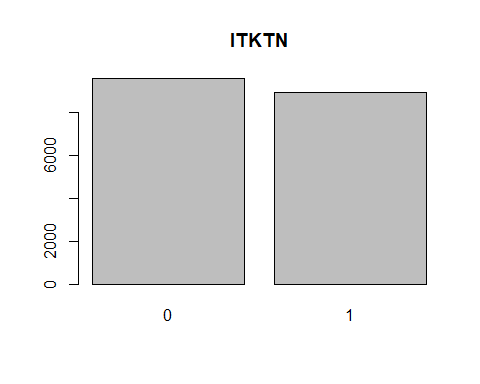
# ITKTTR - convert NA's to 0  
Inv\_final$ITKTTR[is.na(Inv\_final$ITKTTR)]<-0  
barplot(table(Inv\_final$ITKTTR),main="ITKTTR")



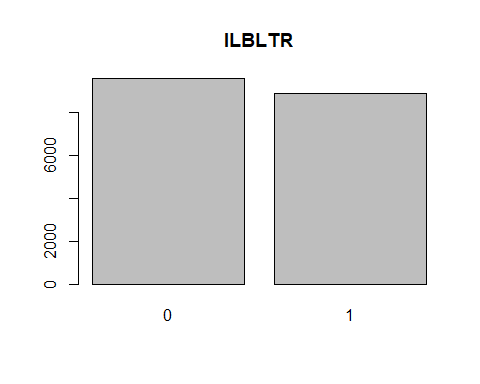
# ITKTTA - mostly empty  
barplot(table(Inv\_final$ITKTTA),main="ITKTTA")



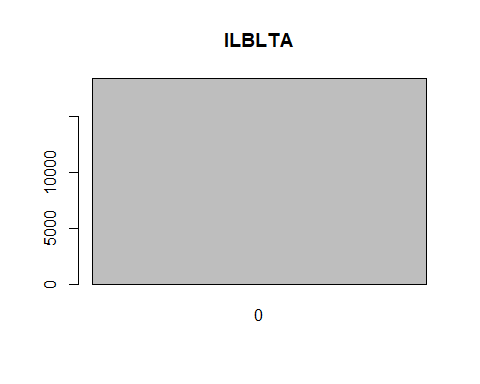
# ITKTN  
barplot(table(Inv\_final$ITKTN),main="ITKTN")



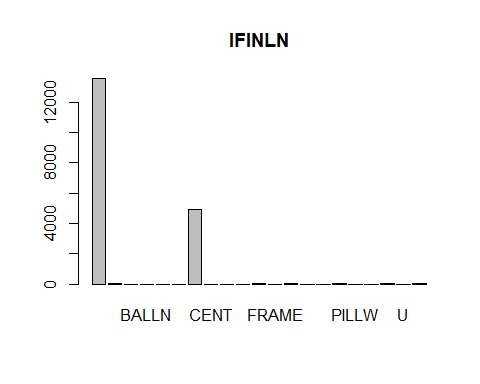
# ILBLTR - convert NA's to 0  
Inv\_final$ILBLTR[is.na(Inv\_final$ILBLTR)]<-0  
barplot(table(Inv\_final$ILBLTR),main="ILBLTR")



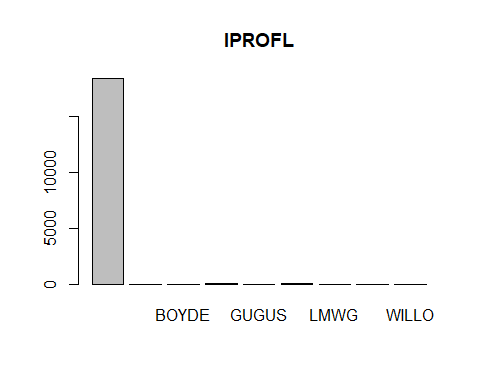
# ILBLTA - empty  
barplot(table(Inv\_final$ILBLTA),main="ILBLTA")



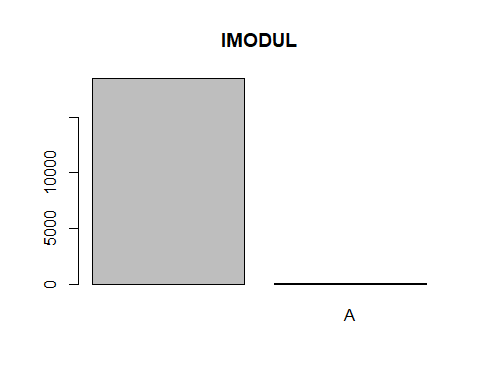
# IFINLN  
barplot(table(Inv\_final$IFINLN),main="IFINLN")



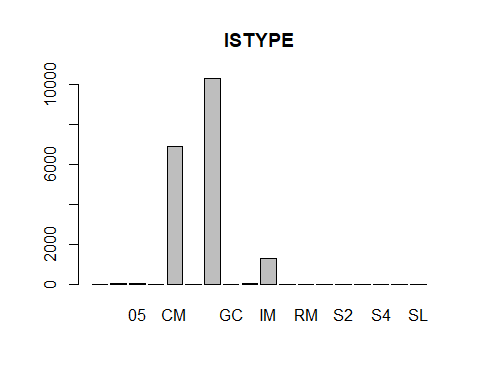
# IPROFL - mostly empty  
barplot(table(Inv\_final$IPROFL),main="IPROFL")



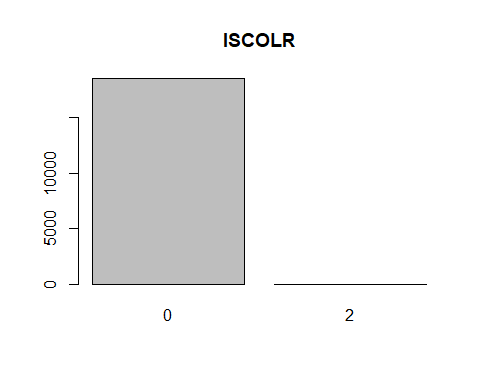
# IMODUL - mostly empty  
barplot(table(Inv\_final$IMODUL),main="IMODUL")



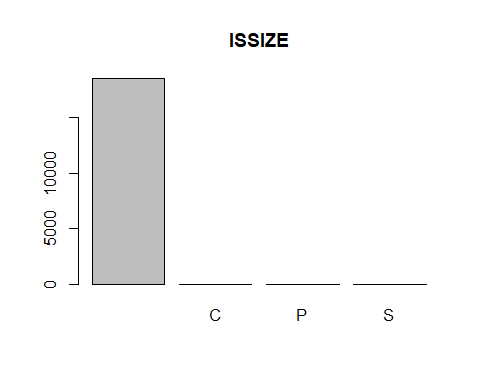
# ISTYPE  
barplot(table(Inv\_final$ISTYPE),main="ISTYPE")



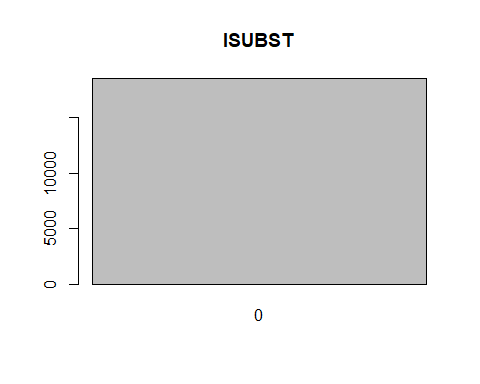
# ISCOLR - mostly empty  
barplot(table(Inv\_final$ISCOLR),main="ISCOLR")



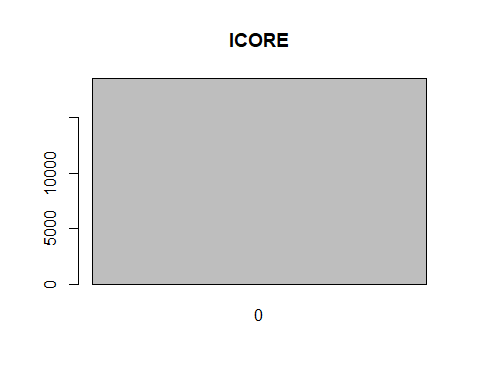
# ISSIZE - mostly empty  
barplot(table(Inv\_final$ISSIZE),main="ISSIZE")



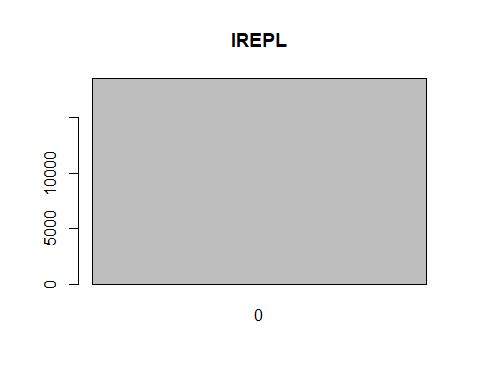
# IHAZCD - all NA's  
# MCHNUM - all NA's  
# ISUBST - empty  
barplot(table(Inv\_final$ISUBST),main="ISUBST")



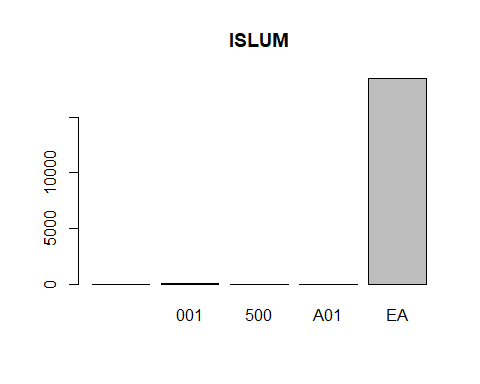
# ICORE - empty  
barplot(table(Inv\_final$ICORE),main="ICORE")



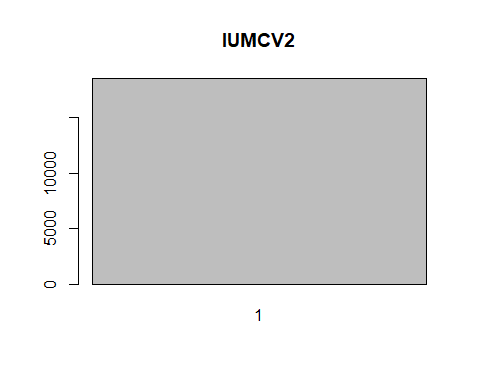
# IREPL - empty  
barplot(table(Inv\_final$IREPL),main="IREPL")



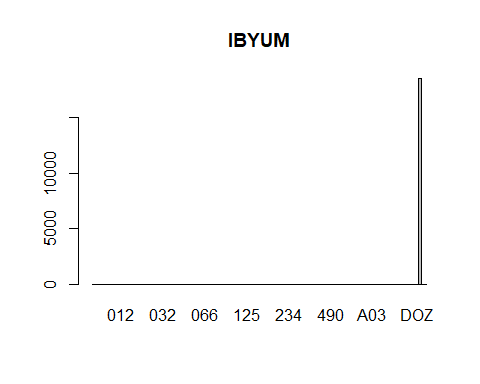
# ISLUM - mostly empty  
barplot(table(Inv\_final$ISLUM),main="ISLUM")



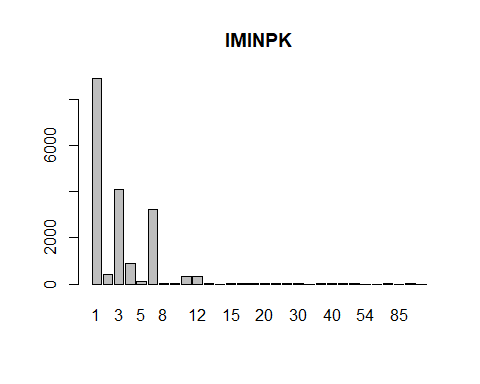
# IUMCV2 - all 1's  
barplot(table(Inv\_final$IUMCV2),main="IUMCV2")



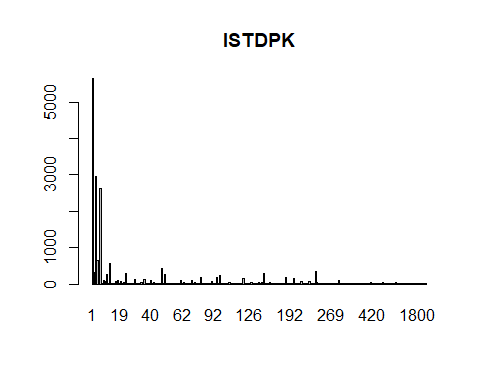
# IBYUM - all EA's  
barplot(table(Inv\_final$IBYUM),main="IBYUM")



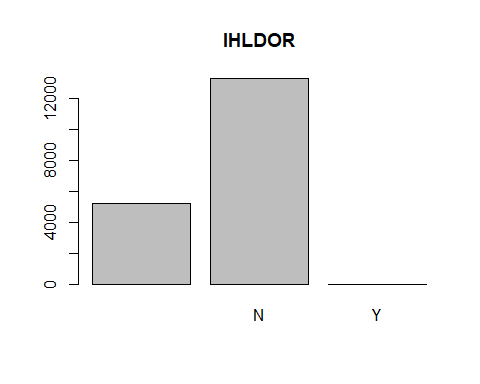
# IMINPK  
barplot(table(Inv\_final$IMINPK),main="IMINPK")



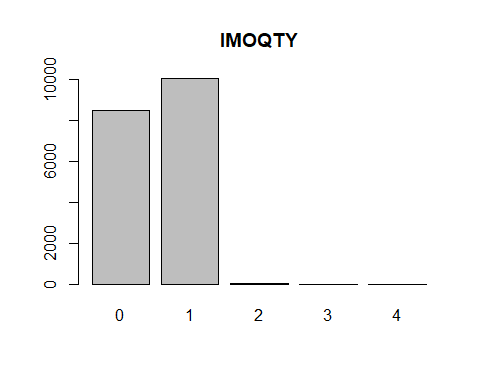
# ISTDPK  
barplot(table(Inv\_final$ISTDPK),main="ISTDPK")



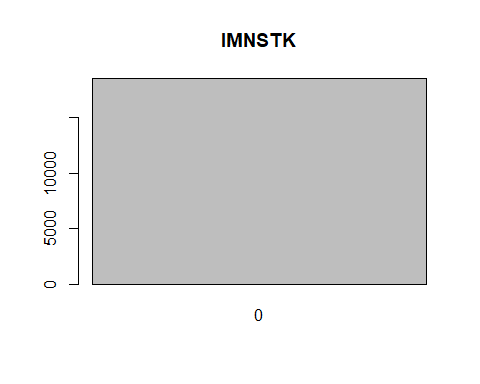
# IHLDOR - mostly empty  
barplot(table(Inv\_final$IHLDOR),main="IHLDOR")



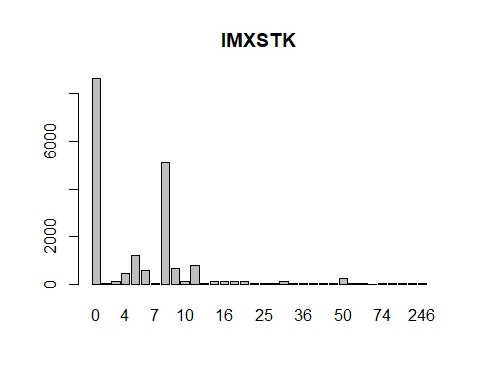
# IMOQTY - mostly empty  
barplot(table(Inv\_final$IMOQTY),main="IMOQTY")



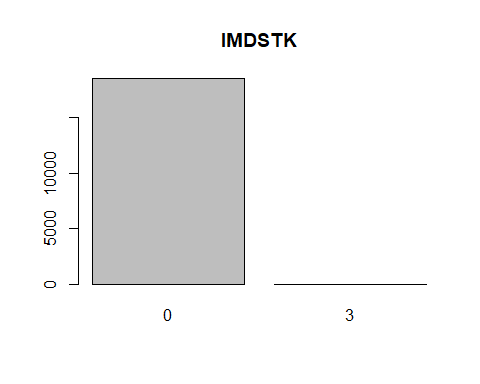
# IMNSTK - empty  
barplot(table(Inv\_final$IMNSTK),main="IMNSTK")



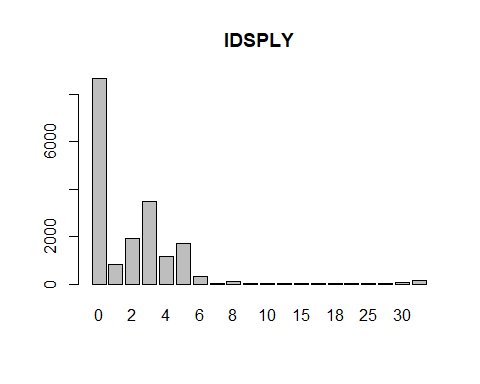
# IMXSTK  
barplot(table(Inv\_final$IMXSTK),main="IMXSTK")



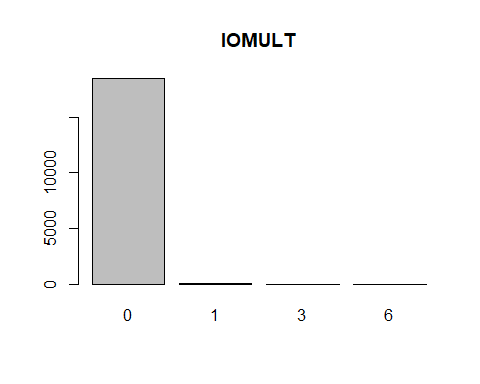
# IMDSTK - empty  
barplot(table(Inv\_final$IMDSTK),main="IMDSTK")



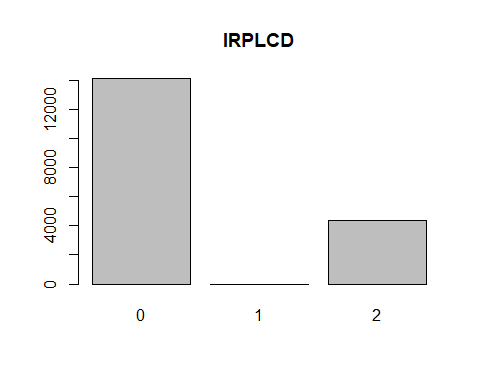
# IDSPLY  
barplot(table(Inv\_final$IDSPLY),main="IDSPLY")



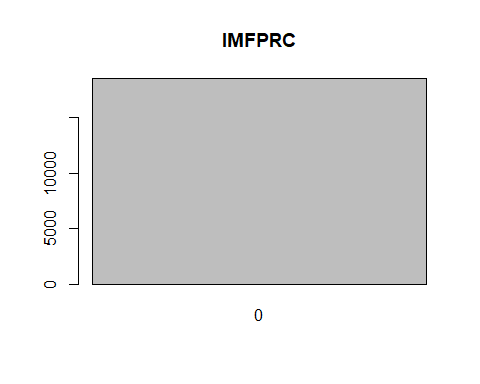
# IOMULT - mostly empty  
barplot(table(Inv\_final$IOMULT),main="IOMULT")



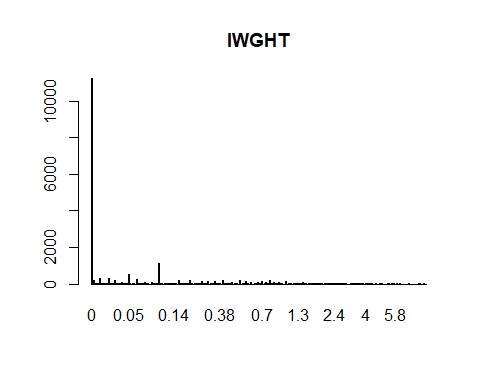
# IRPLCD - convert NA's to 0  
Inv\_final$IRPLCD[is.na(Inv\_final$IRPLCD)]<-0  
barplot(table(Inv\_final$IRPLCD),main="IRPLCD")



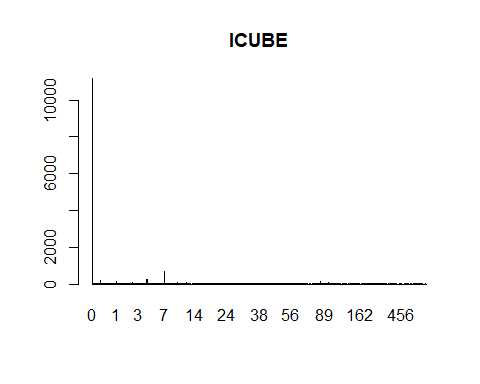
# IMFPRC - empty  
barplot(table(Inv\_final$IMFPRC),main="IMFPRC")



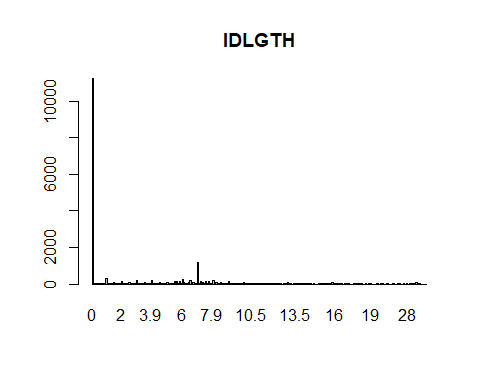
# IWGHT - there is one outlier that throws the significance off  
barplot(table(Inv\_final$IWGHT),main="IWGHT")



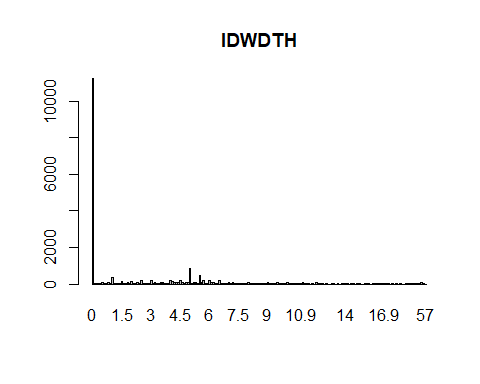
Inv\_final$IWGHT[Inv\_final$IWGHT==3001.4]=0  
# ICUBE - investigate taking the log(ICUBE)  
barplot(table(Inv\_final$ICUBE),main="ICUBE")



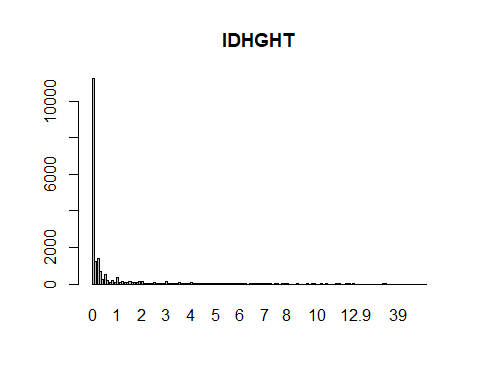
# IDLGTH - investigate taking the log(ICUBE)  
barplot(table(Inv\_final$IDLGTH),main="IDLGTH")



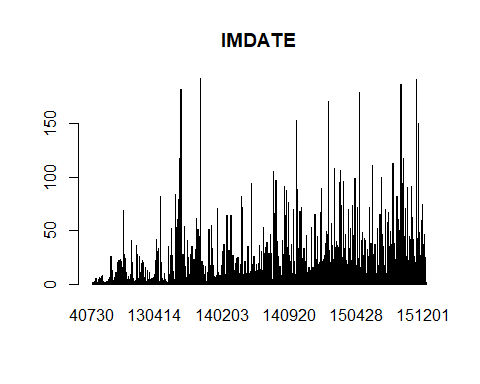
# IDWDTH - investigate taking the log(ICUBE)  
barplot(table(Inv\_final$IDWDTH),main="IDWDTH")



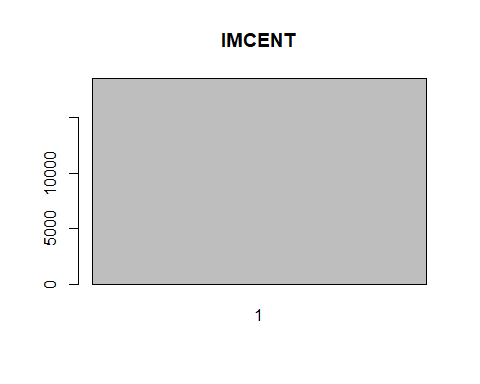
# IDHGHT - investigate taking the log(ICUBE)  
barplot(table(Inv\_final$IDHGHT),main="IDHGHT")



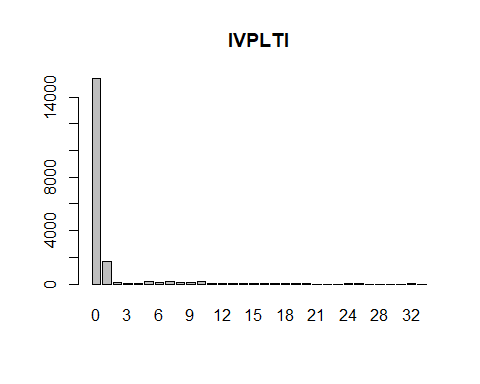
# IMDATE  
barplot(table(Inv\_final$IMDATE),main="IMDATE")



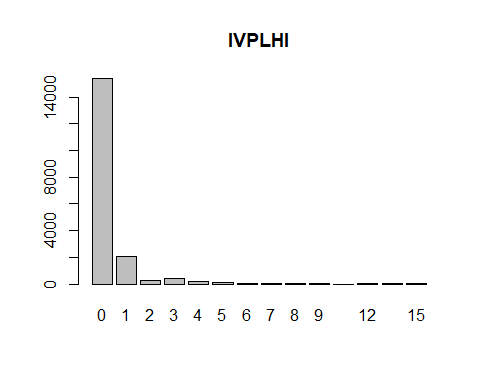
# IMCENT - empty  
barplot(table(Inv\_final$IMCENT),main="IMCENT")



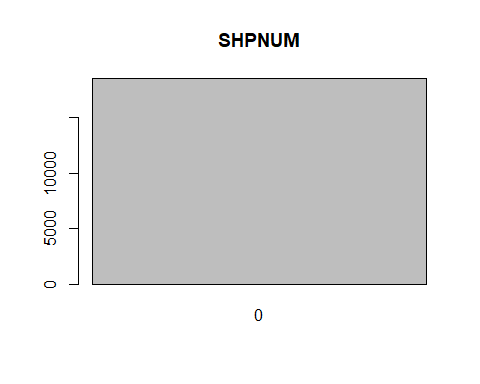
# IVPLTI - not significant  
barplot(table(Inv\_final$IVPLTI),main="IVPLTI")



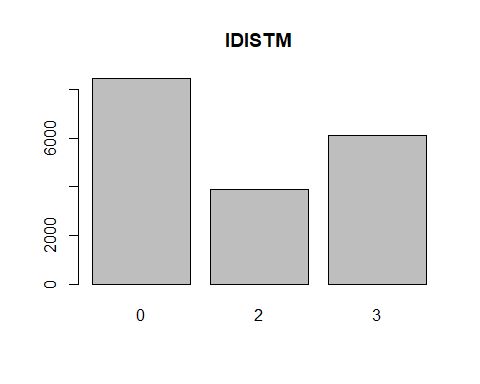
# IVPLHI  
barplot(table(Inv\_final$IVPLHI),main="IVPLHI")



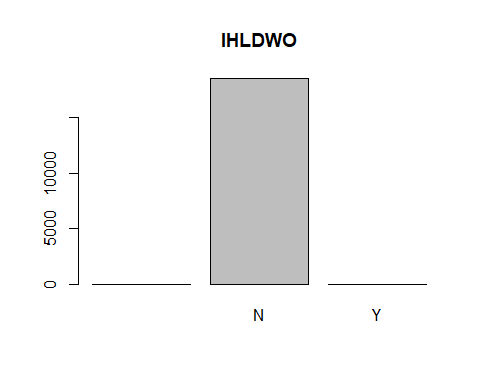
# SHPNUM - empty  
barplot(table(Inv\_final$SHPNUM),main="SHPNUM")



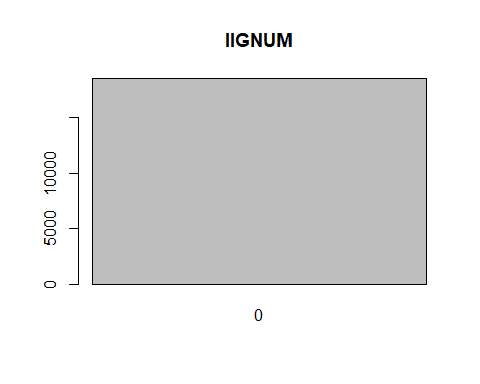
# Lots of NA's in IDISTM, replace with 0  
Inv\_final$IDISTM[is.na(Inv\_final$IDISTM)]<-0  
barplot(table(Inv\_final$IDISTM),main="IDISTM")



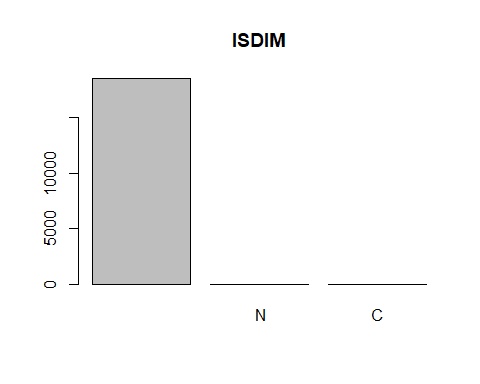
# IHLDWO - empty  
barplot(table(Inv\_final$IHLDWO),main="IHLDWO")



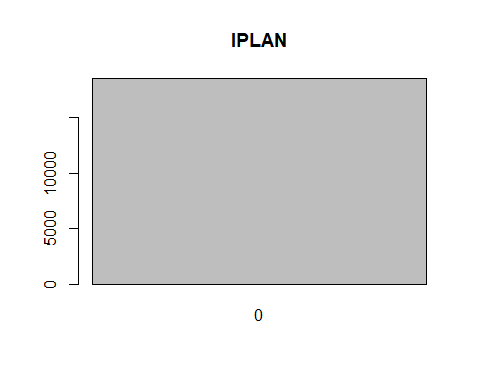
# IIGNUM - empty  
barplot(table(Inv\_final$IIGNUM),main="IIGNUM")



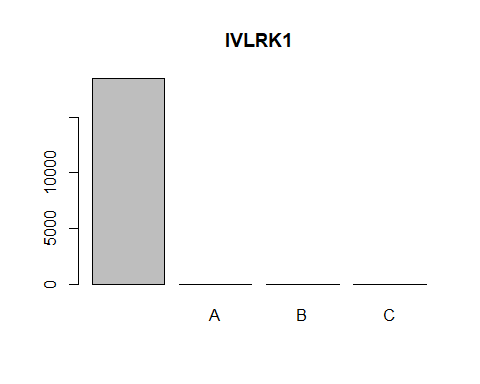
# ISDIM - empty  
barplot(table(Inv\_final$ISDIM),main="ISDIM")



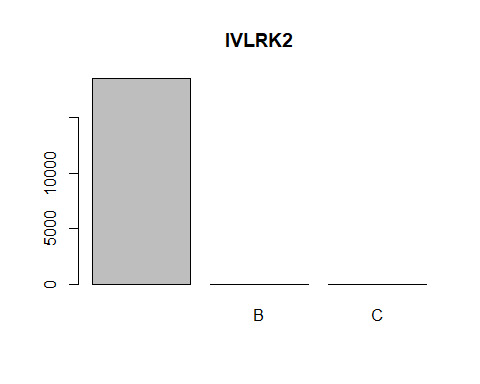
# IVATCD - all NA's  
# IPLAN - empty  
barplot(table(Inv\_final$IPLAN),main="IPLAN")



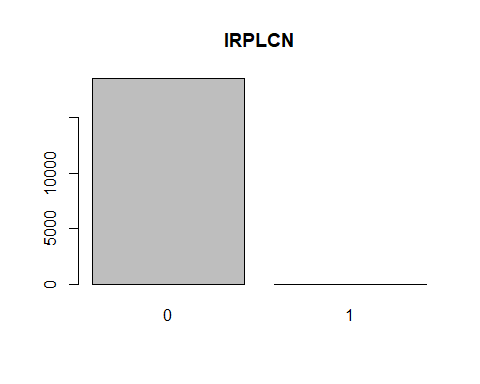
# IVLRK1 - empty  
barplot(table(Inv\_final$IVLRK1),main="IVLRK1")



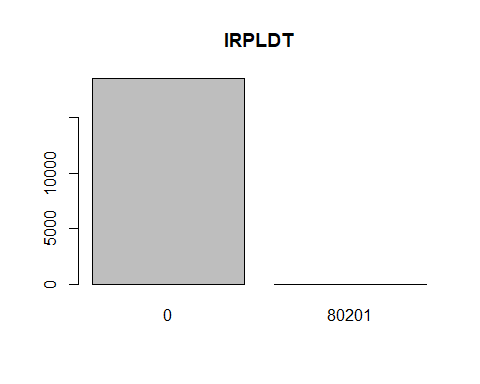
# IVLRK2 - empty  
barplot(table(Inv\_final$IVLRK2),main="IVLRK2")



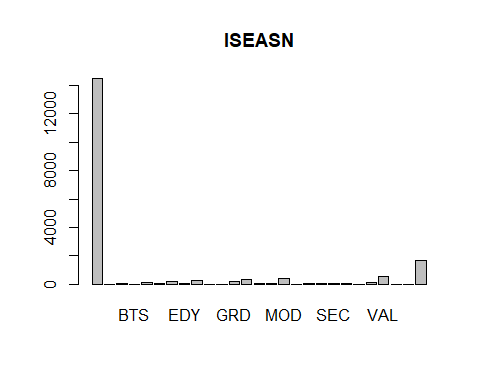
# IVLRK3 - all NA's  
# IVLRK4 - all NA's  
# IRPLCN - empty  
barplot(table(Inv\_final$IRPLCN),main="IRPLCN")



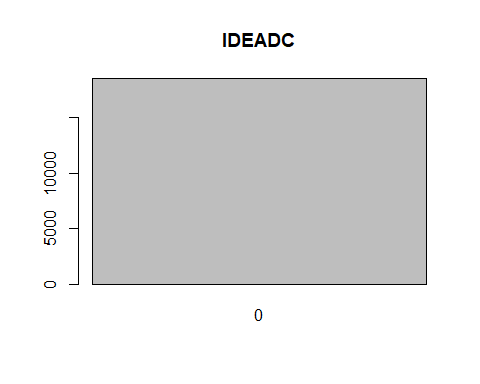
# IRPLDT - empty  
barplot(table(Inv\_final$IRPLDT),main="IRPLDT")



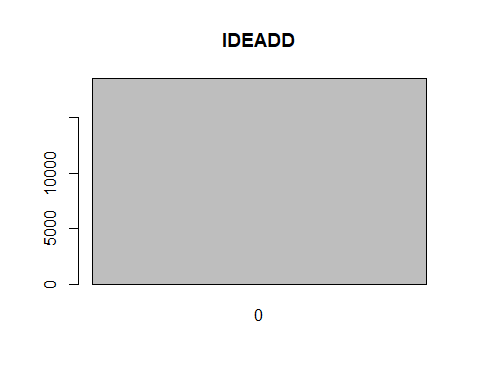
# ISEASN  
barplot(table(Inv\_final$ISEASN),main="ISEASN")



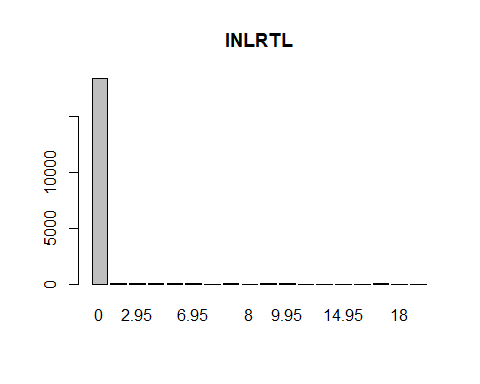
# IDEADC  
barplot(table(Inv\_final$IDEADC),main="IDEADC")



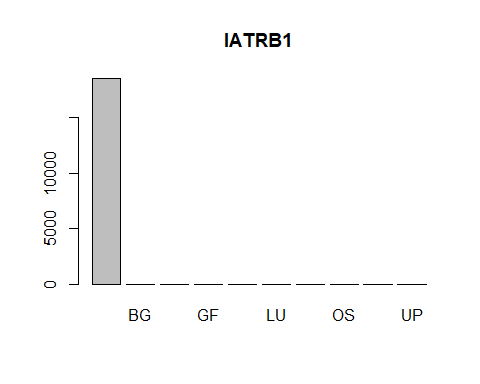
# IDEADD - empty  
barplot(table(Inv\_final$IDEADD),main="IDEADD")



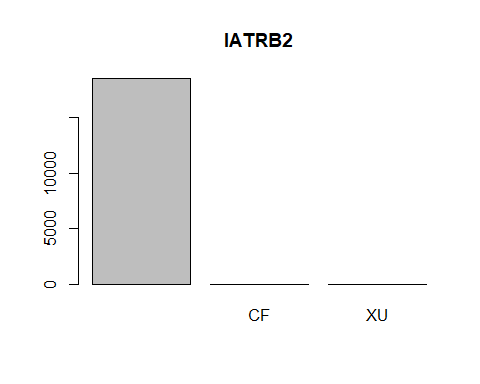
# INLRTL  
barplot(table(Inv\_final$INLRTL),main="INLRTL")



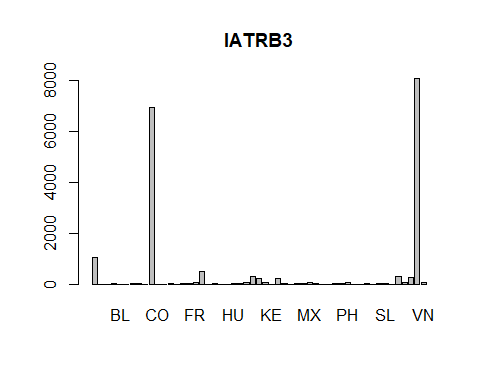
# IHANDL is all NA's  
# IATRB1 - empty  
barplot(table(Inv\_final$IATRB1),main="IATRB1")



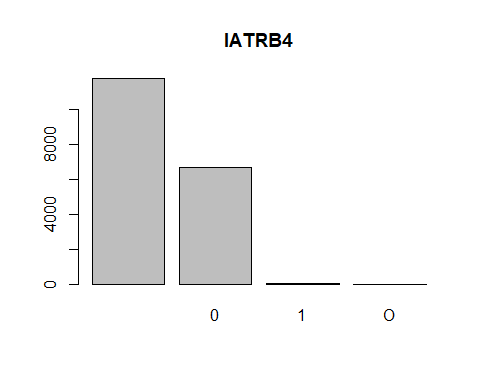
# IATRB2 - empty  
barplot(table(Inv\_final$IATRB2),main="IATRB2")



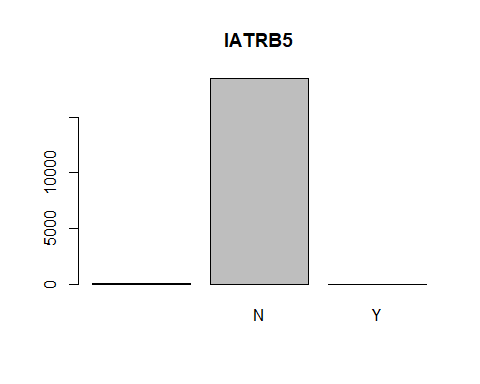
# IATRB3  
barplot(table(Inv\_final$IATRB3),main="IATRB3")



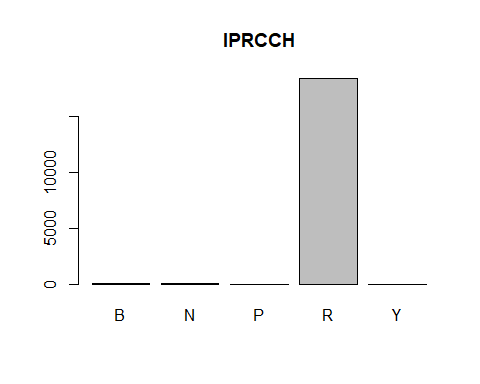
# IATRB4 - empty  
barplot(table(Inv\_final$IATRB4),main="IATRB4")



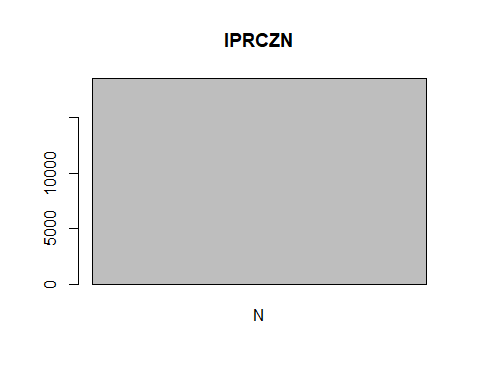
# IATRB5 - all N's  
barplot(table(Inv\_final$IATRB5),main="IATRB5")



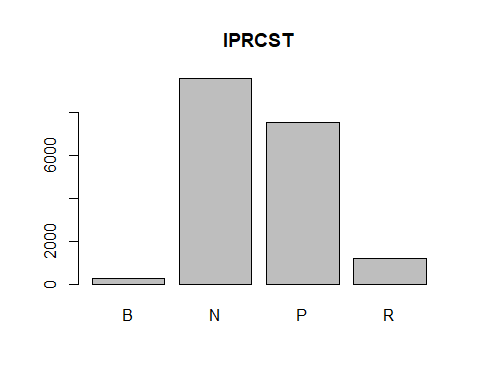
# IPRCCH  
barplot(table(Inv\_final$IPRCCH),main="IPRCCH")



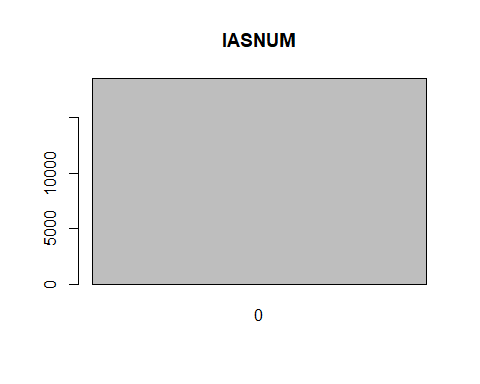
# IPRCZN - all N's  
barplot(table(Inv\_final$IPRCZN),main="IPRCZN")



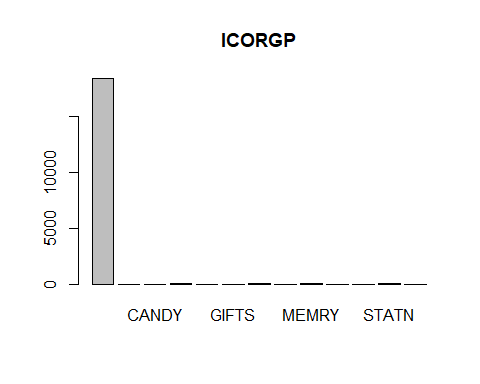
# IPRCST  
barplot(table(Inv\_final$IPRCST),main="IPRCST")



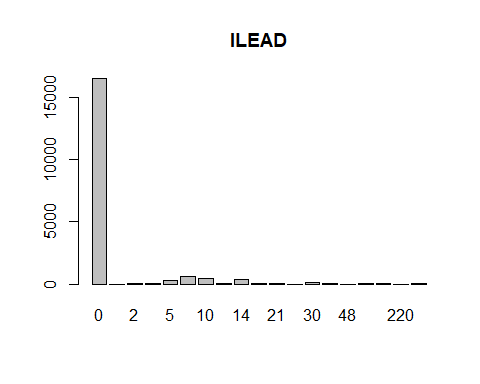
# IASNUM - empty  
barplot(table(Inv\_final$IASNUM),main="IASNUM")



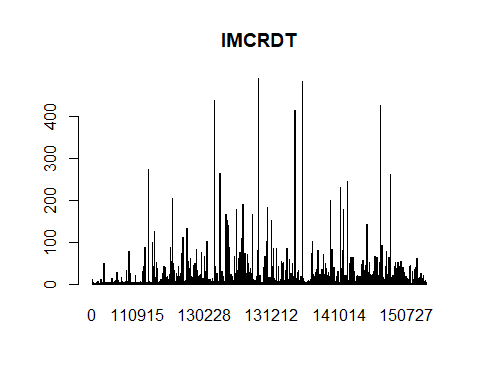
# ICORGP  
barplot(table(Inv\_final$ICORGP),main="ICORGP")



# ILEAD  
barplot(table(Inv\_final$ILEAD),main="ILEAD")



# IHZCOD - all NA's  
# IFRACT - all NA's  
# IMCRDT - distributed  
barplot(table(Inv\_final$IMCRDT),main="IMCRDT")



# IMCRCN - all 1's  
barplot(table(Inv\_final$IMCRCN),main="IMCRCN")

