IRIDataMergedDataExploration

```
merged.data<-read.csv("cleaned_data_1427_1478_with_panelist_demographics.csv", header = TRUE, na.string
merged.data$COUNT<-1
merged.data<-merged.data[,c(2:58)]
merged.data$PANID<-as.factor(merged.data$PANID)
merged.data$VEND<-as.factor(sprintf("%05d", merged.data$VEND))
merged.data$ITEM<-as.factor(sprintf("%05d", merged.data$ITEM))
merged.data$D<-as.factor(merged.data$D)
merged.data$PR<-as.factor(merged.data$PR)
# < 0.1, 0.1 < x < 0.26, 0.26 < x < 0.35, 0.35 < x < 0.3938, 0.3938 < x < 0.4894, 0.4894 < x < 0.675, 0
merged.data$VOL_EQ<-cut(merged.data$VOL_EQ, breaks = c(-Inf, 0.1, 0.26, 0.35, 0.3938, 0.4894, 0.675, In
merged.data[c(22:57)]<-lapply(merged.data[c(22:57)], factor)
```

Number of panelists:

```
nlevels(merged.data$PANID)
```

```
## [1] 2501
```

Number of transactions:

```
nrow(merged.data)
```

```
## [1] 8975
```

Average purchases per person:

```
nrow(merged.data) / nlevels(merged.data$PANID)
```

```
## [1] 3.588565
```

Lets see how many levels are in each of the factor columns:

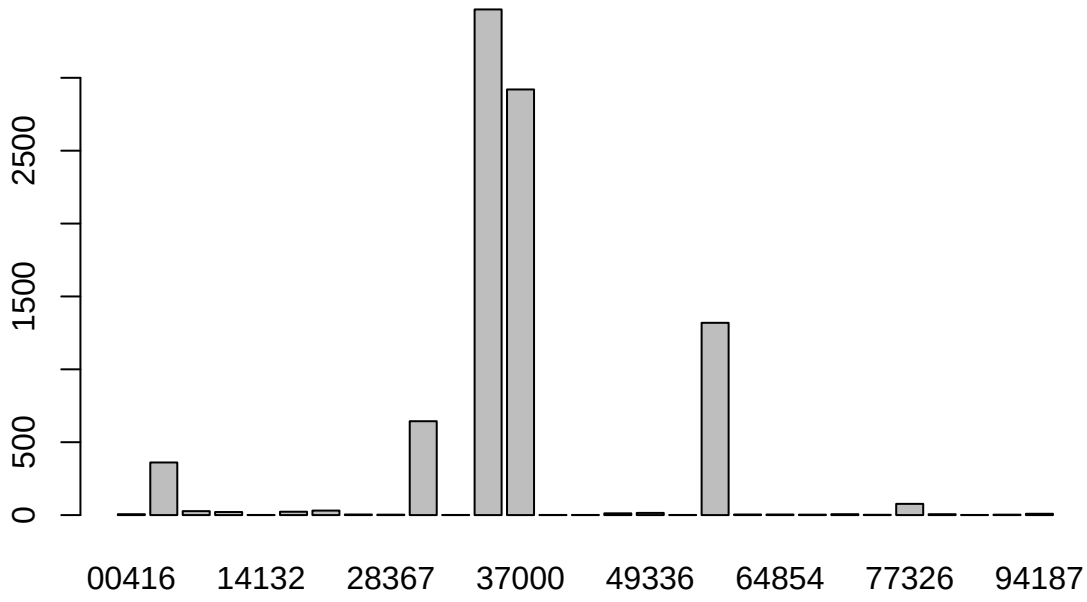
```
for (x in c(1:ncol(merged.data))) {
  if (is.factor(merged.data[,x])) {
    print(paste(colnames(merged.data)[x], "=", nlevels(merged.data[,x])))
  }
}
```

```
## [1] "PANID = 2501"
## [1] "VEND = 29"
## [1] "OUTLET = 2"
## [1] "ITEM = 438"
## [1] "F = 5"
## [1] "D = 3"
## [1] "PR = 2"
## [1] "VOL_EQ = 7"
## [1] "SIZE = 9"
## [1] "FLAVOR.SCENT = 62"
## [1] "FORM = 11"
## [1] "PACKAGE = 17"
## [1] "PRODUCT.TYPE = 4"
## [1] "STORE.LOCATION = 1"
## [1] "ADDITIVES = 25"
## [1] "TYPE.OF.FORMULAT = 46"
## [1] "COLOR = 21"
```

```
## [1] "Panelist.Type = 3"
## [1] "Combined.Pre.Tax.Income.of.HH = 13"
## [1] "Family.Size = 6"
## [1] "HH_RACE = 2"
## [1] "Type.of.Residential.Possession = 3"
## [1] "COUNTY = 2"
## [1] "HH_AGE = 7"
## [1] "HH_EDU = 9"
## [1] "HH_OCC = 12"
## [1] "Age.Group.Applied.to.Male.HH = 8"
## [1] "Education.Level.Reached.by.Male.HH = 9"
## [1] "Occupation.Code.of.Male.HH = 12"
## [1] "Male.Working.Hour.Code = 7"
## [1] "MALE_SMOKE = 2"
## [1] "Age.Group.Applied.to.Female.HH = 8"
## [1] "Education.Level.Reached.by.Female.HH = 9"
## [1] "Occupation.Code.of.Female.HH = 12"
## [1] "Female.Working.Hour.Code = 7"
## [1] "FEM_SMOKE = 2"
## [1] "Number.of.Dogs = 6"
## [1] "Number.of.Cats = 6"
## [1] "Children.Group.Code = 8"
## [1] "Marital.Status = 6"
## [1] "Language = 5"
## [1] "Number.of.TVs.Used.by.HH = 10"
## [1] "Number.of.TVs.Hooked.to.Cable = 10"
## [1] "HISP_FLAG = 2"
## [1] "HISP_CAT = 7"
## [1] "HH.Head.Race..RACE2. = 8"
## [1] "HH.Head.Race..RACE3. = 8"
## [1] "Microwave.Owned.by.HH = 1"
## [1] "ZIPCODE = 28"
## [1] "FIPSCODE = 5"
## [1] "market.based.upon.zipcode = 2"
## [1] "IRI.Geography.Number = 2"
## [1] "EXT_FACT = 1"
```

Should we reduce the number of brands we are dealing with? Now its 29, maybe to 10?

```
barplot(table(merged.data$VEND))
```



According to the bar chart we could reduce it to something like 6 or 7.

Find out which ones they are:

```
vendor.counts<-aggregate(merged.data$COUNT, by=list(Brand<-merged.data$VEND), FUN=sum)
head(vendor.counts[order(-vendor.counts$x),], n = 10)
```

```
##    Group.1    x
## 12   35000 3469
## 13   37000 2920
## 19   53100 1319
## 10   33200  644
## 2    10158  361
## 25   77326   77
## 7    24291   31
## 3    10310   27
## 6    18515   23
## 4    12547   21
```

```
#lets take anything above 30
brands<-vendor.counts[which(vendor.counts$x > 31),1]
brands
```

```
## [1] 10158 33200 35000 37000 53100 77326
## 29 Levels: 00416 10158 10310 12547 14132 18515 24291 27275 28367 ... 94187
```

```
# now remove all non-brand from data
merged.data.reduced<-merged.data[which(merged.data$VEND %in% brands),]
# how to reduce number of levels to match new?
```

```
merged.data.reduced$VEND<-factor(merged.data.reduced$VEND)
levels(merged.data.reduced$VEND)
```

```
## [1] "10158" "33200" "35000" "37000" "53100" "77326"
```

Lets do a barplot of some other interesting ones:

```
barplot(table(merged.data.reduced$PRODUCT.TYPE))
```

We should probably get rid of everything but TOOTHEPASTE, could throw brand purchasing off.

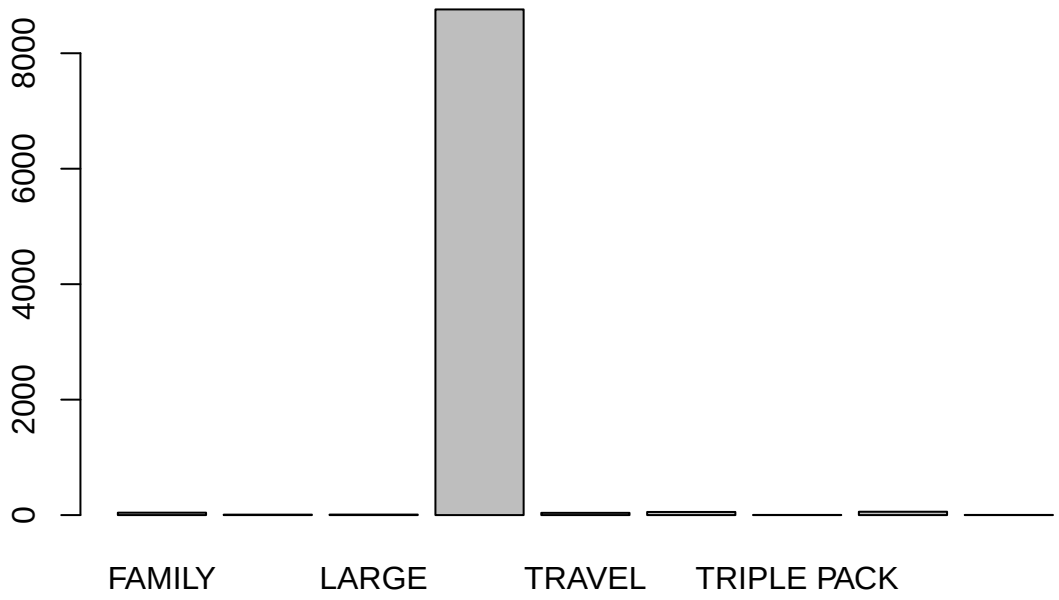
```
merged.data.reduced<-merged.data.reduced[which(merged.data.reduced$PRODUCT.TYPE == "TOOTHPASTE"),]
aggregate(merged.data.reduced$COUNT, by=list(PRODUCT.TYPES=merged.data.reduced$PRODUCT.TYPE), FUN=sum)
```

```
##   PRODUCT.TYPES    x
## 1    TOOTHPASTE 8484
```

```
merged.data.reduced$PRODUCT.TYPE<-factor(merged.data.reduced$PRODUCT.TYPE)
# we should be able to just get rid of this column now
levels(merged.data.reduced$PRODUCT.TYPE)
```

```
## [1] "TOOTHPASTE"
```

```
barplot(table(merged.data$SIZE))
```

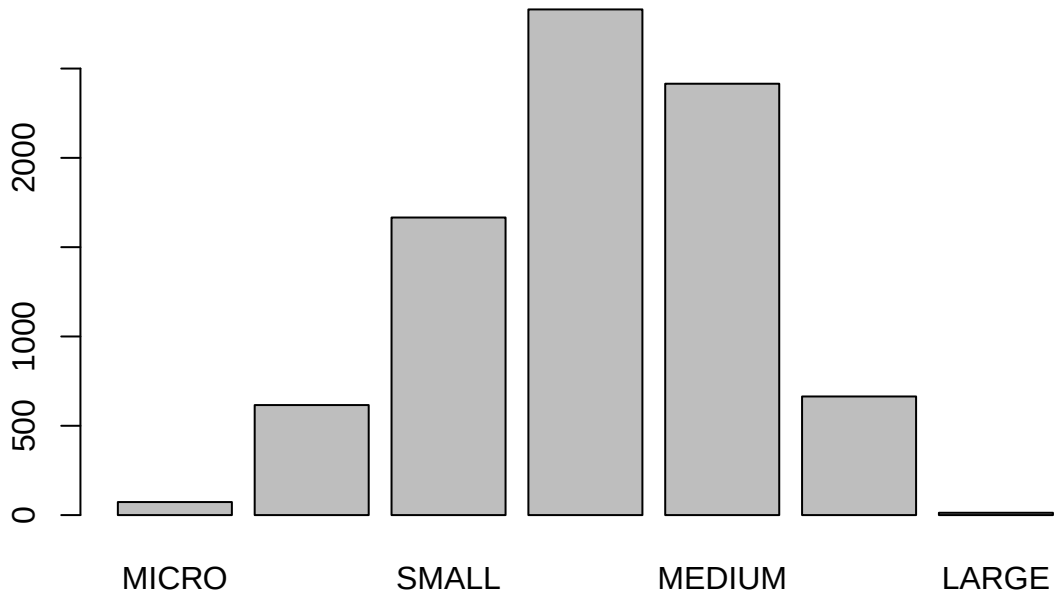


Hmm, maybe only do one "SIZE", then divide that up into "VOL_EQ"s

```
merged.data.reduced<-merged.data.reduced[which(merged.data.reduced$SIZE == "REGULAR"),]
merged.data.reduced$SIZE<-factor(merged.data.reduced$SIZE)
levels(merged.data.reduced$SIZE)
```

```
## [1] "REGULAR"
```

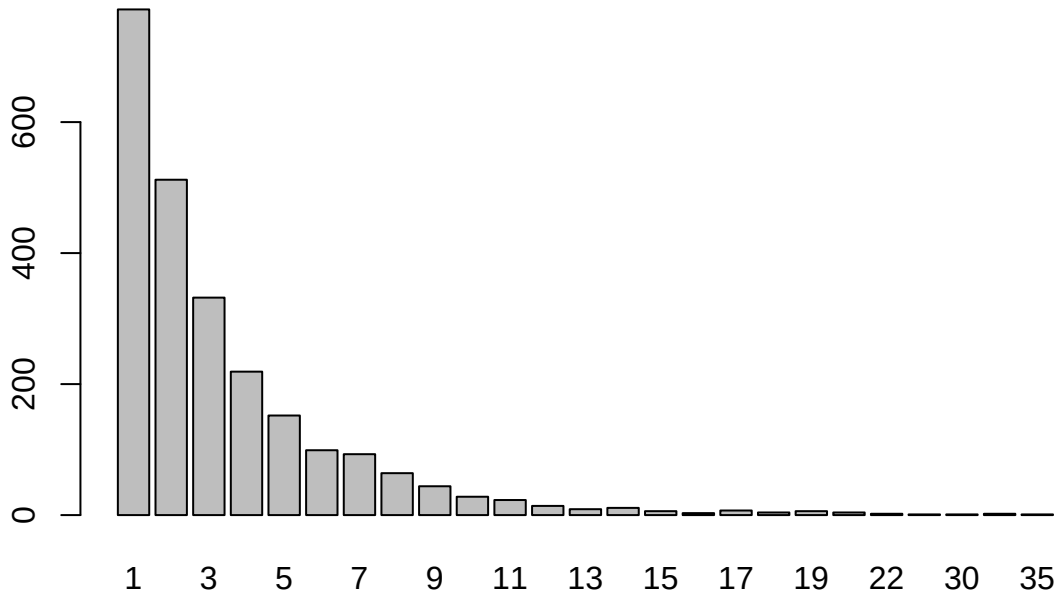
```
# we should be able to just get rid of this column now
barplot(table(merged.data.reduced$VOL_EQ))
```



Maybe get rid of MICRO and LARGE?

Lets look at purchases per person:

```
purchases.per.person<-aggregate(merged.data.reduced$COUNT, by=list(Person=merged.data.reduced$PANID), FU

barplot(table(as.factor(purchases.per.person$x)))
```



Hmm, there are a lot of people that only made 1 purchase. Thats ok I guess.

Lets get an average number of purchases a person made:

```
mean(aggregate(merged.data.reduced$COUNT, by=list(Person=merged.data.reduced$PANID), FUN=sum)$x)
```

```
## [1] 3.436281
```

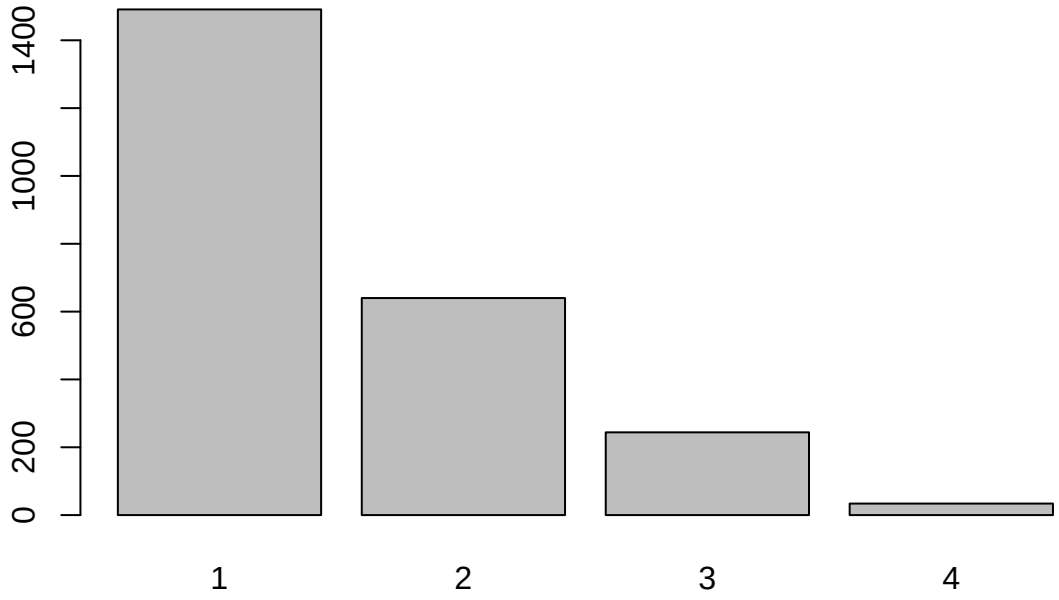
Lets look at for each person how many products in each brand type they purchased:

```
brands.by.person<-aggregate(merged.data.reduced$COUNT, by=list(Person=merged.data.reduced$PANID, Brand=m
brands.by.person<-brands.by.person[order(brands.by.person$Person),]
head(brands.by.person)
```

```
##       Person Brand x
## 170  1100032 33200 1
## 508  1100032 35000 1
## 171  1100057 33200 3
## 2981 1100180 53100 7
## 509  1100214 35000 3
## 2982 1100248 53100 4
```

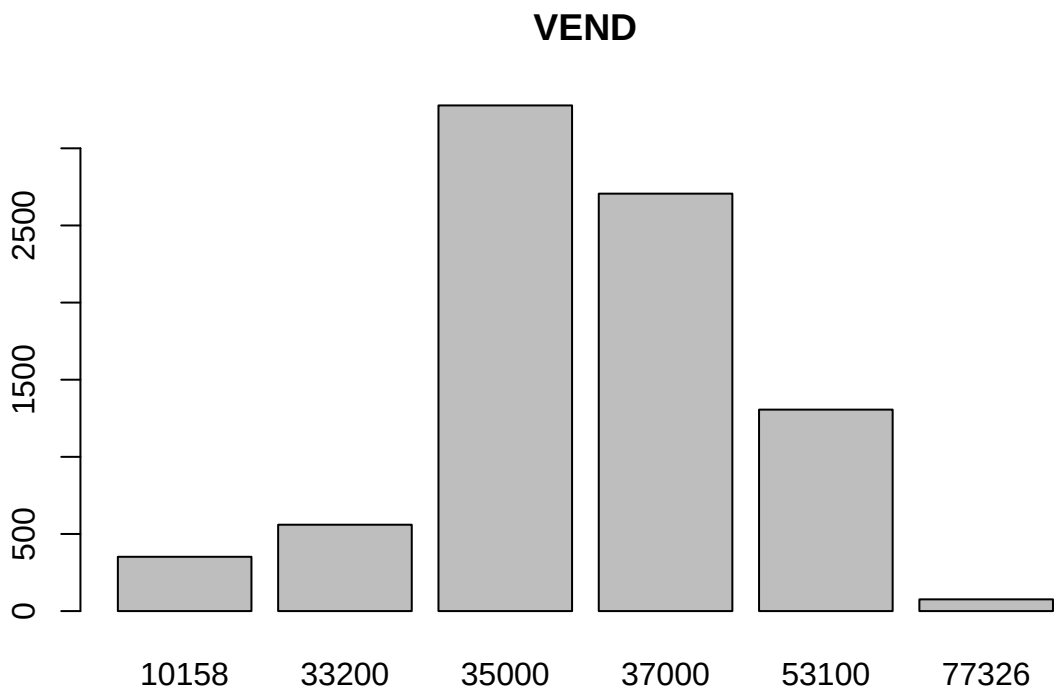
How about number of brands a each person switched between:

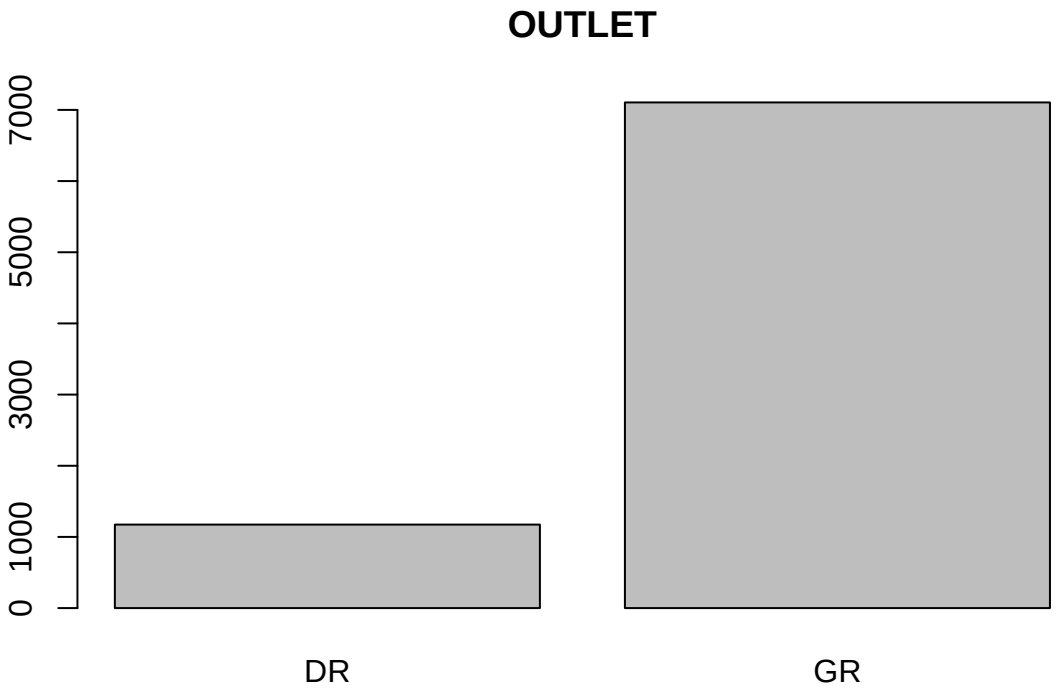
```
brands.by.person$COUNT<-1
brand.per.person<-aggregate(brands.by.person$COUNT, by=list(Person=brands.by.person$Person), FUN=sum)
brand.per.person$x<-as.factor(brand.per.person$x)

barplot(table(brand.per.person$x))
```

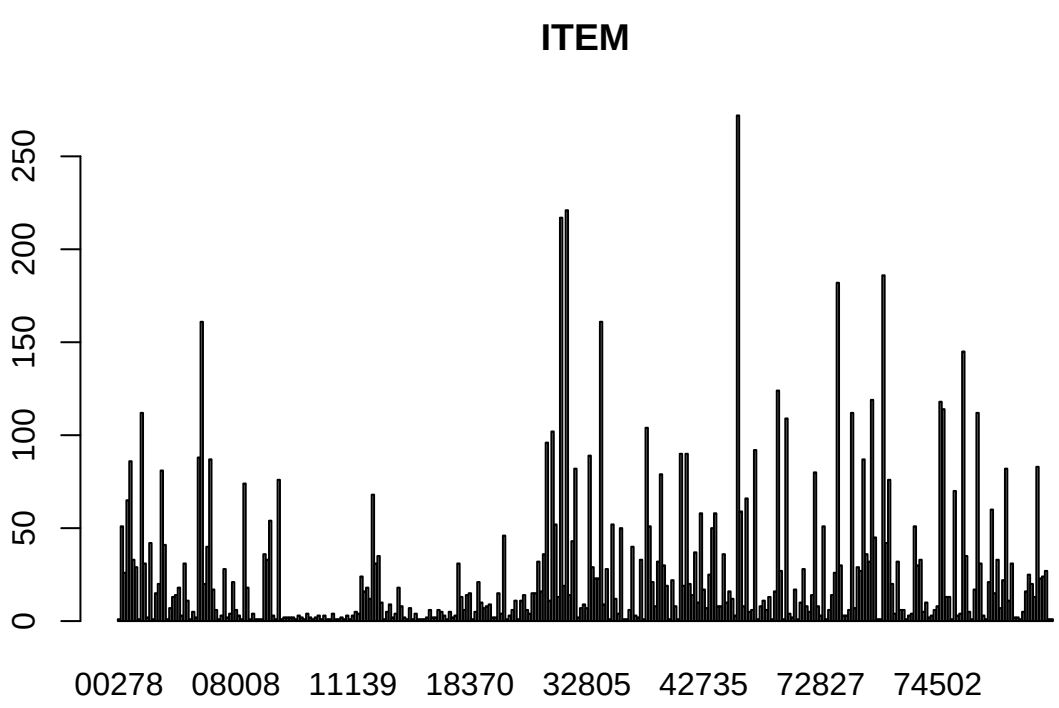


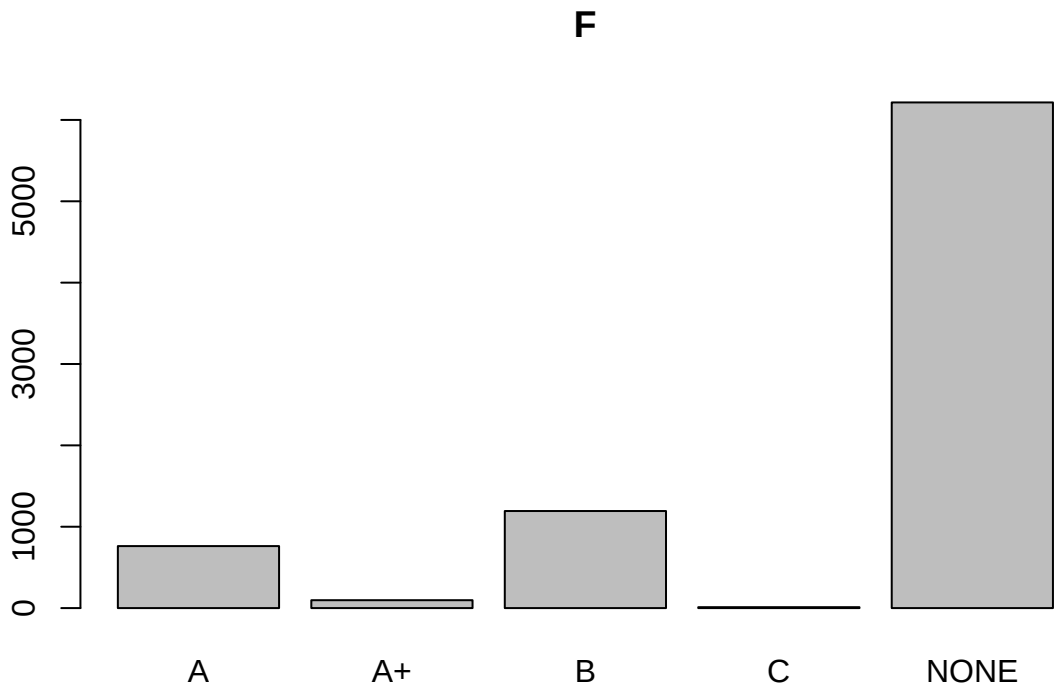
Shows that decent number of people switched between a few brands, but not many above 3.

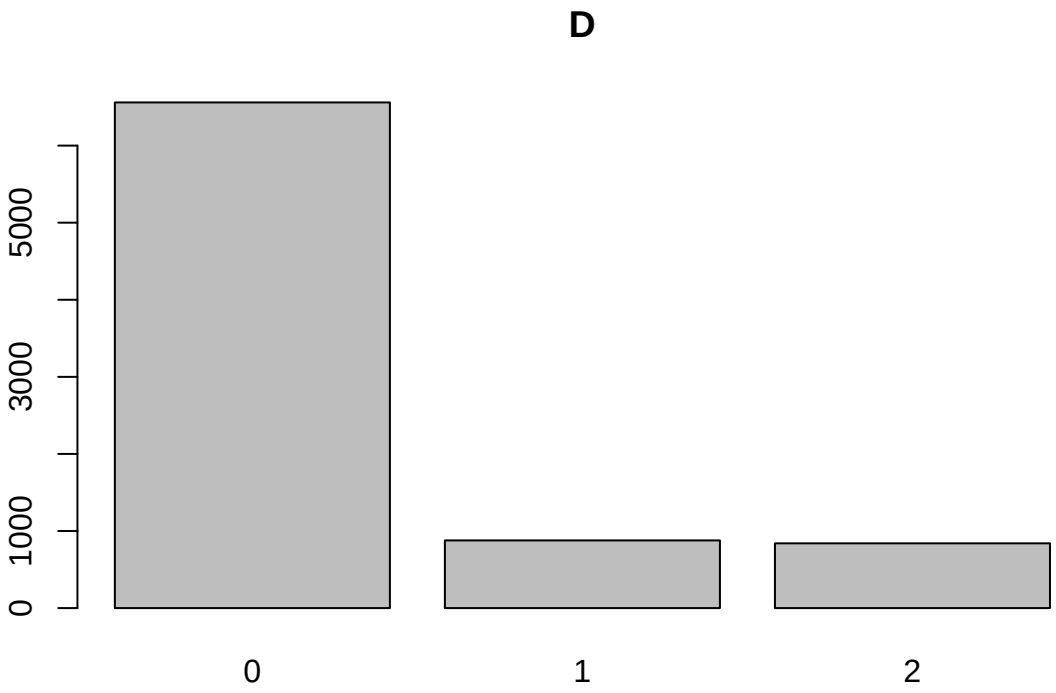
```
for (x in c(2:ncol(merged.data.reduced))) {
  if (is.factor(merged.data.reduced[,x])) {
    merged.data.reduced[,x]<-factor(merged.data.reduced[,x])
    barplot(table(merged.data.reduced[x]), main = colnames(merged.data.reduced)[x])
  }
}
```

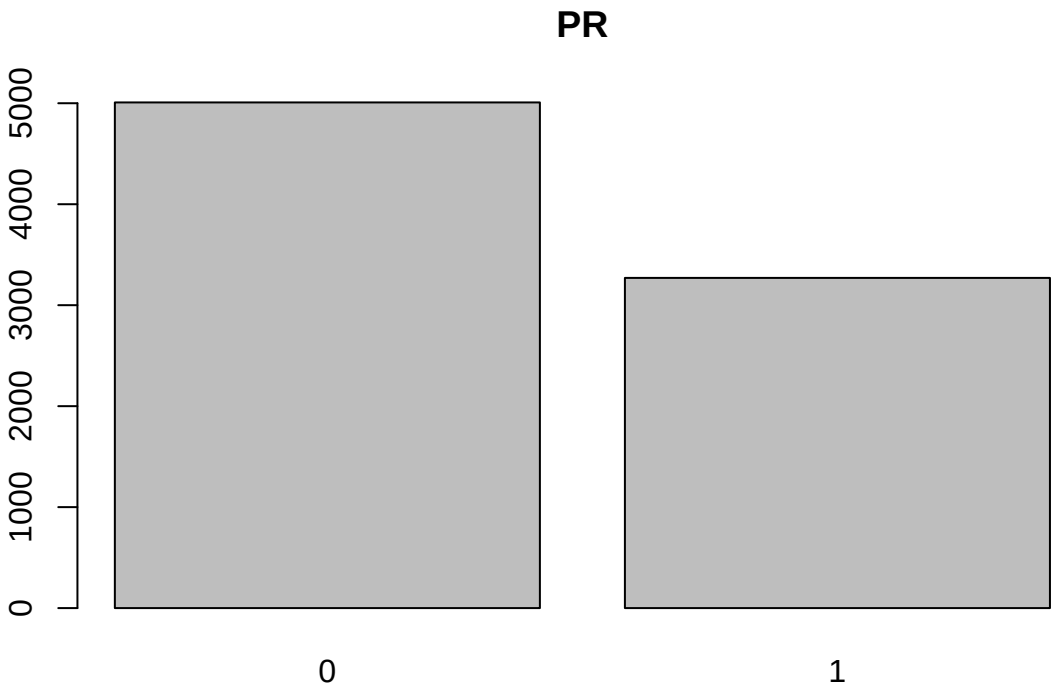

VEND
0
500
1500
2500
10158
33200
35000
37000
53100
77326

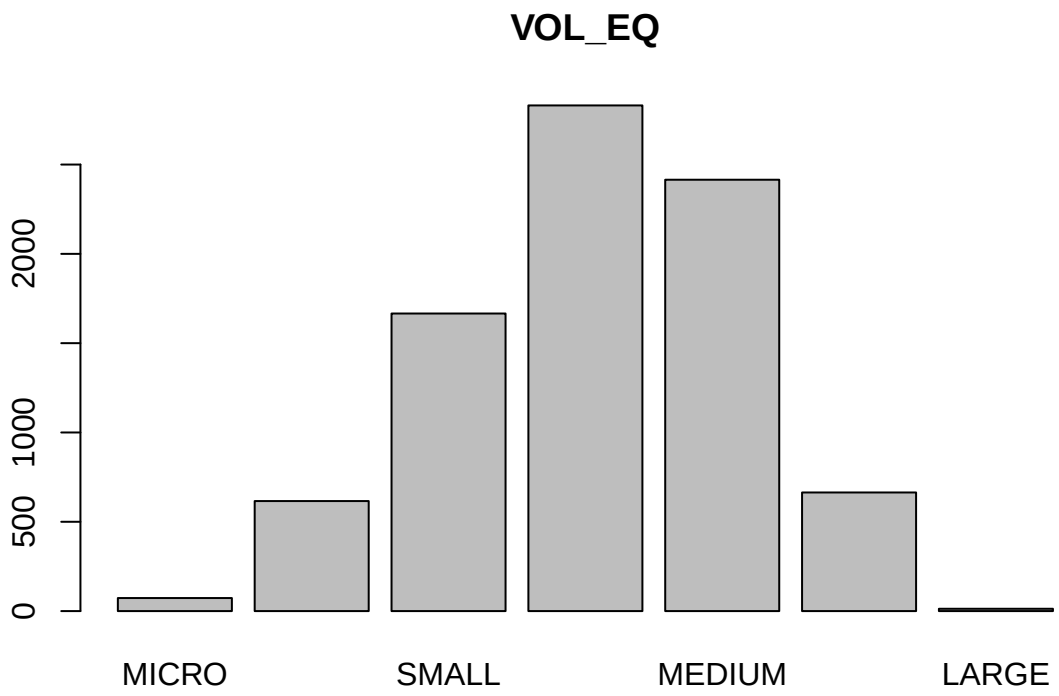

OUTLET
0
1000
3000
5000
7000
DR
GR

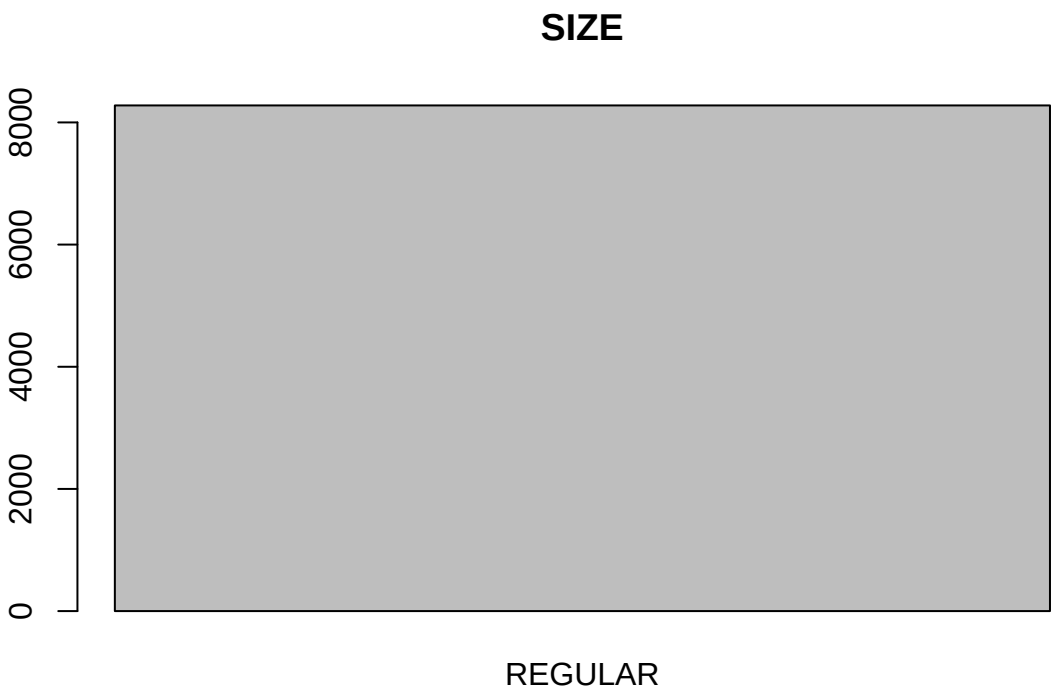

ITEM
0
50
100
150
200
250
00278
08008
11139
18370
32805
42735
72827
74502

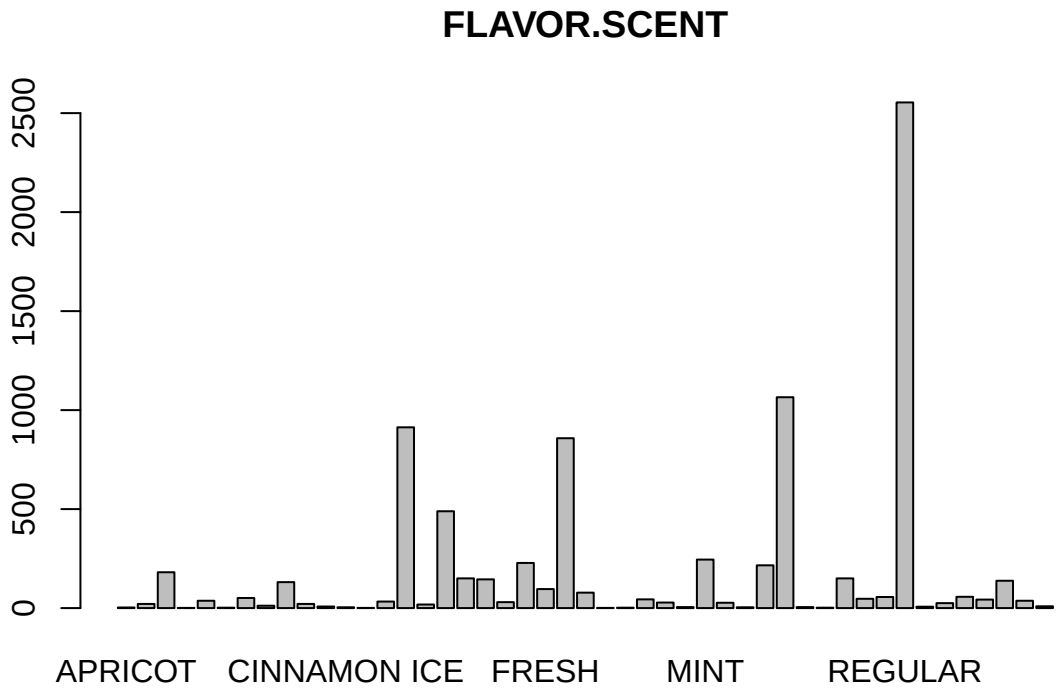

F
0
1000
3000
5000
A
A+
B
C
NONE

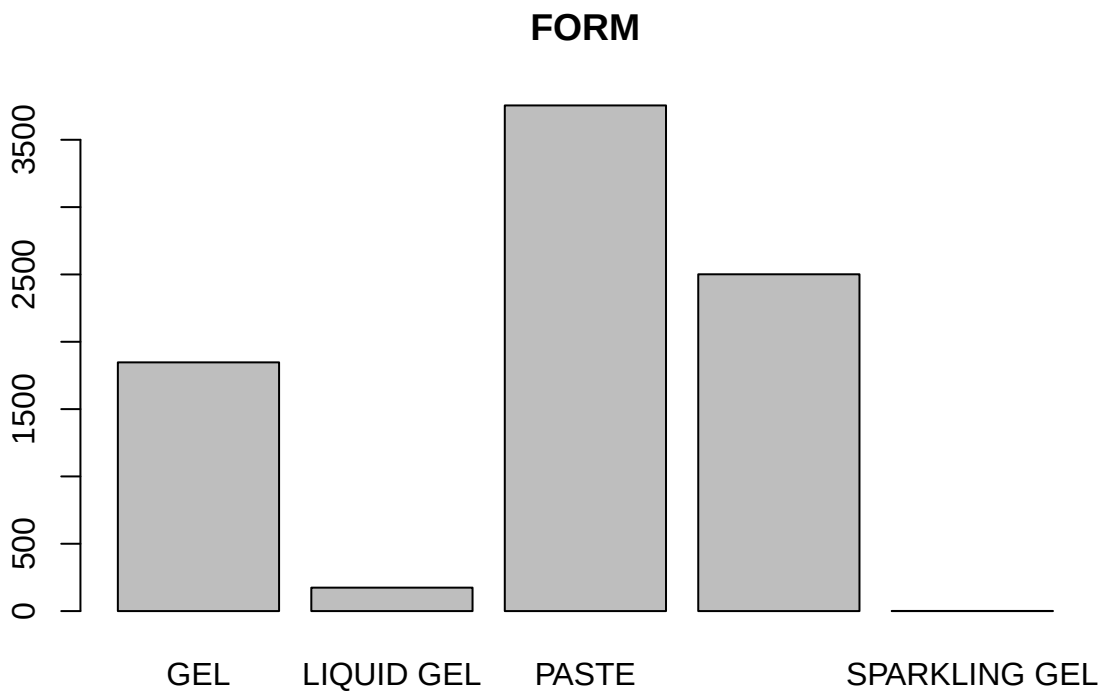

D
0
1000
3000
5000
0
1
2

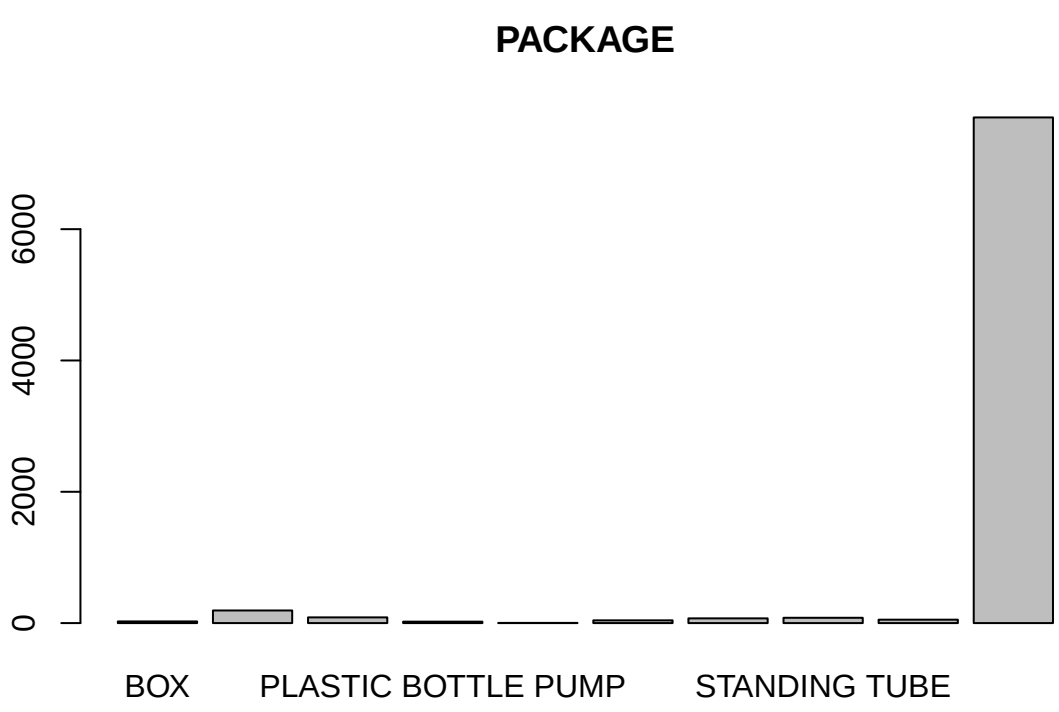

PR
0
1000
2000
3000
4000
5000
0
1

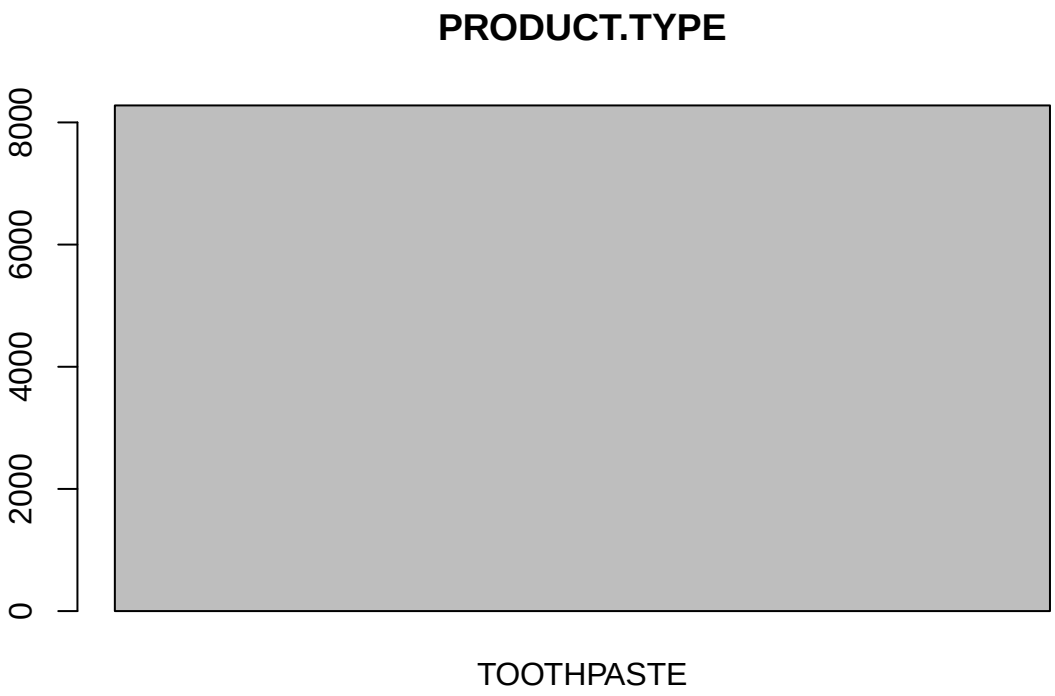

VOL_EQ
0
500
1000
2000
MICRO
SMALL
MEDIUM
LARGE

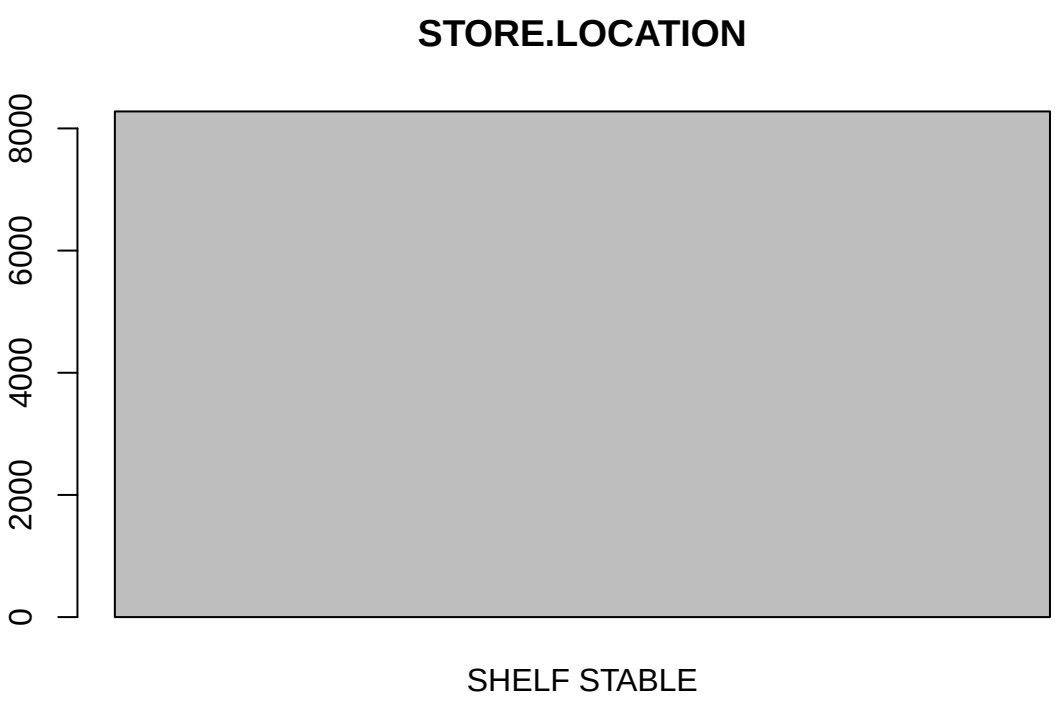

SIZE
0
2000
4000
6000
8000
REGULAR

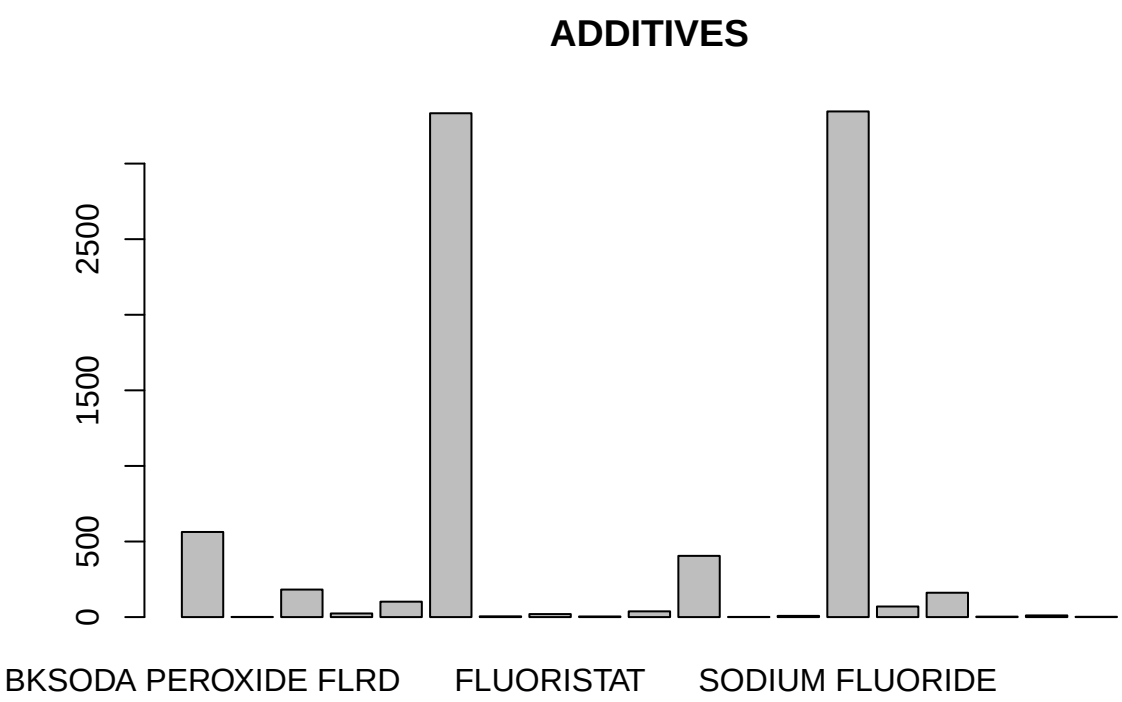

FLAVOR.SCENT
0
500
1000
1500
2000
2500
APRICOT
CINNAMON
ICE
FRESH
MINT
REGULAR

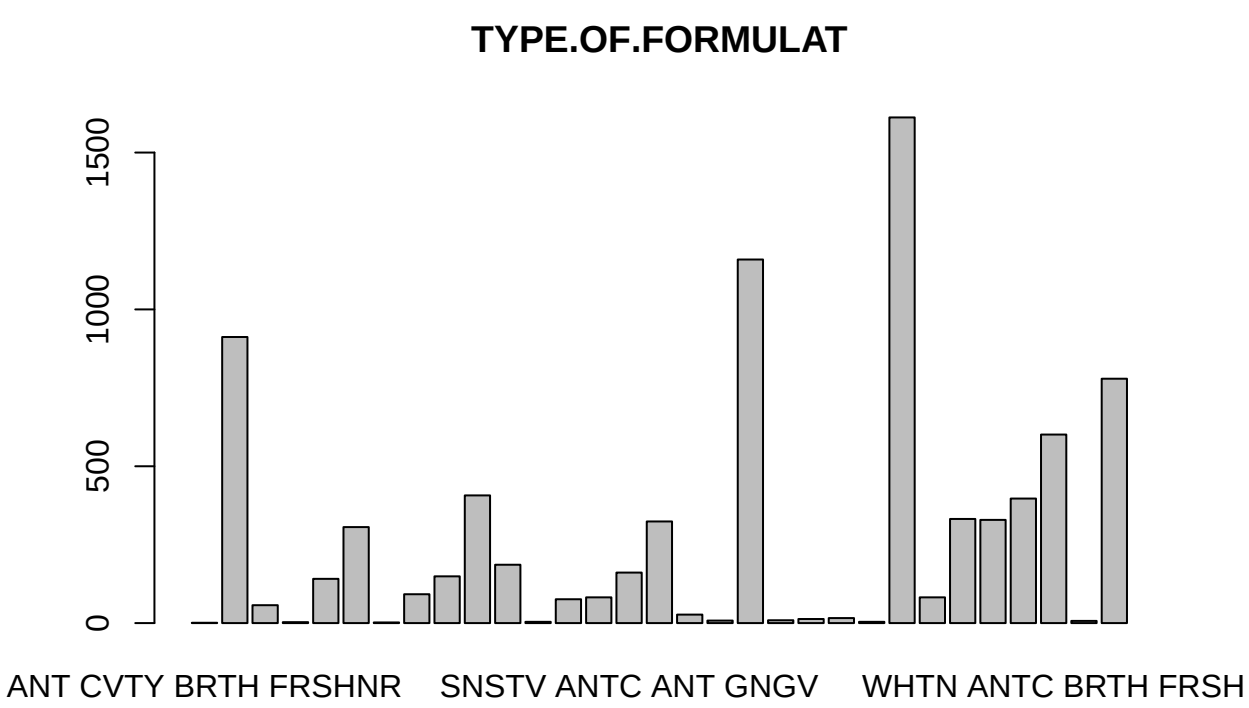

FORM
0
500
1500
2500
3500
GEL
LIQUID GEL
PASTE
SPARKLING GEL

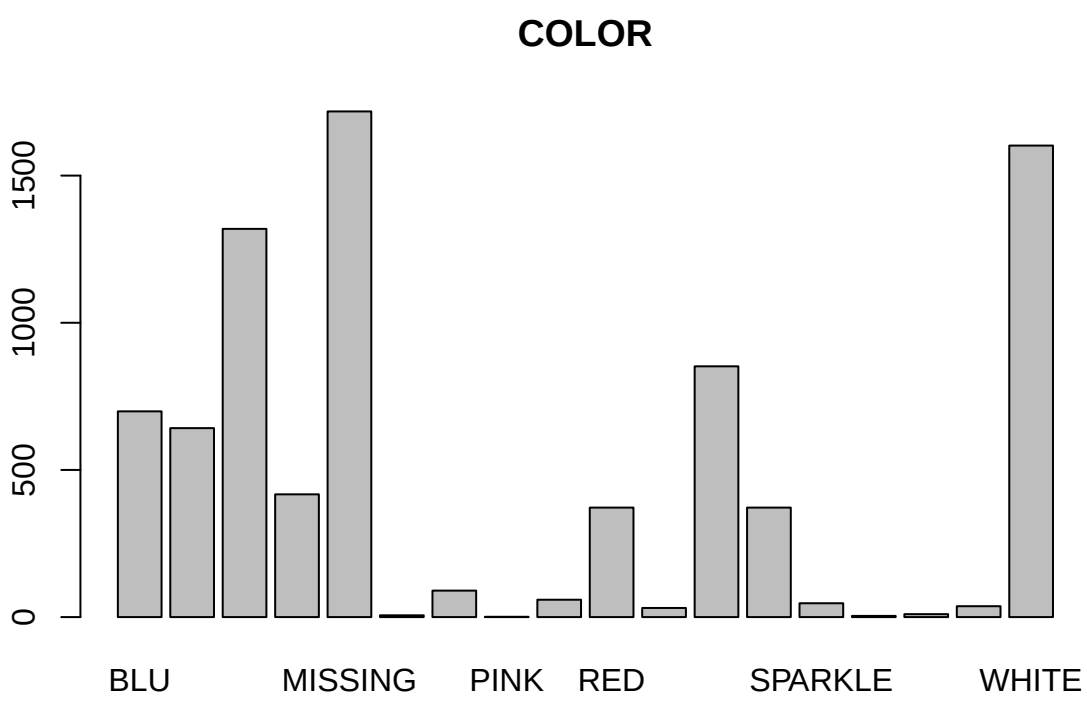

PACKAGE
0
2000
4000
6000
BOX
PLASTIC BOTTLE PUMP
STANDING TUBE

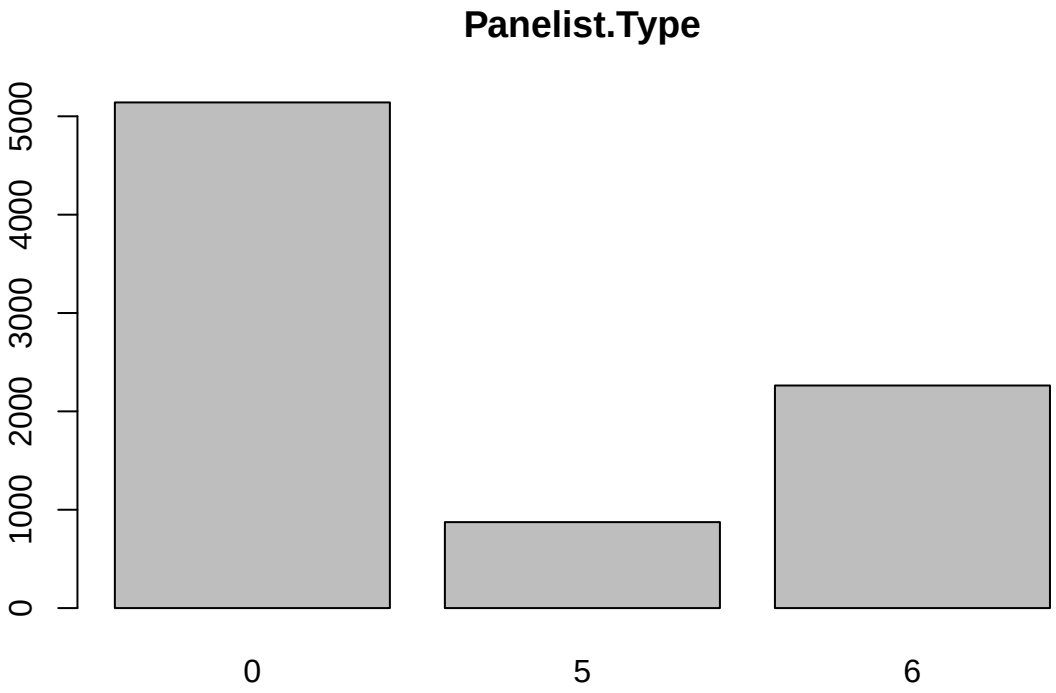

PRODUCT.TYPE
0
2000
4000
6000
8000
TOOTHPASTE

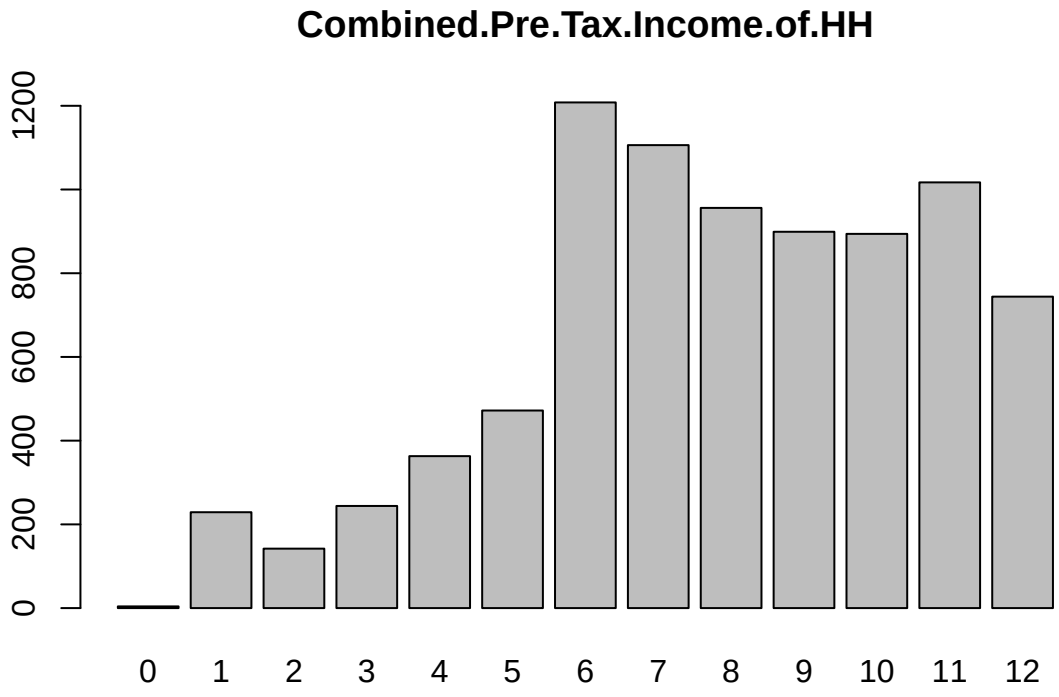

STORE.LOCATION
0
2000
4000
6000
8000
SHELF STABLE

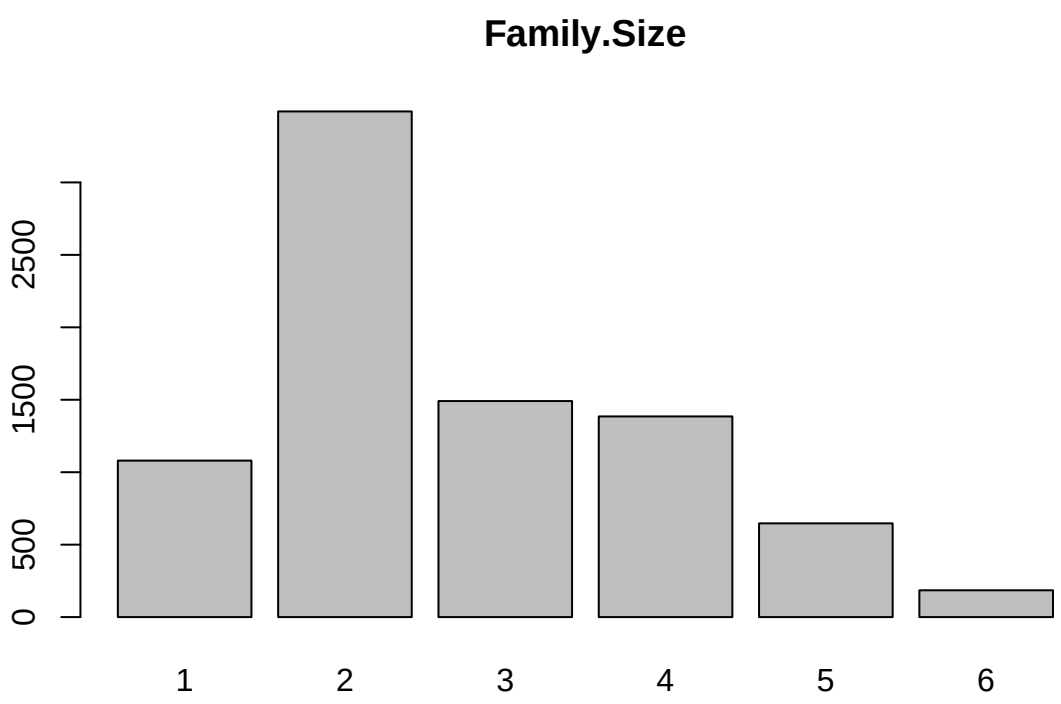

ADDITIVES
0
500
1500
2500
BKSODA
PEROXIDE
FLRD
FLUORISTAT
SODIUM FLUORIDE

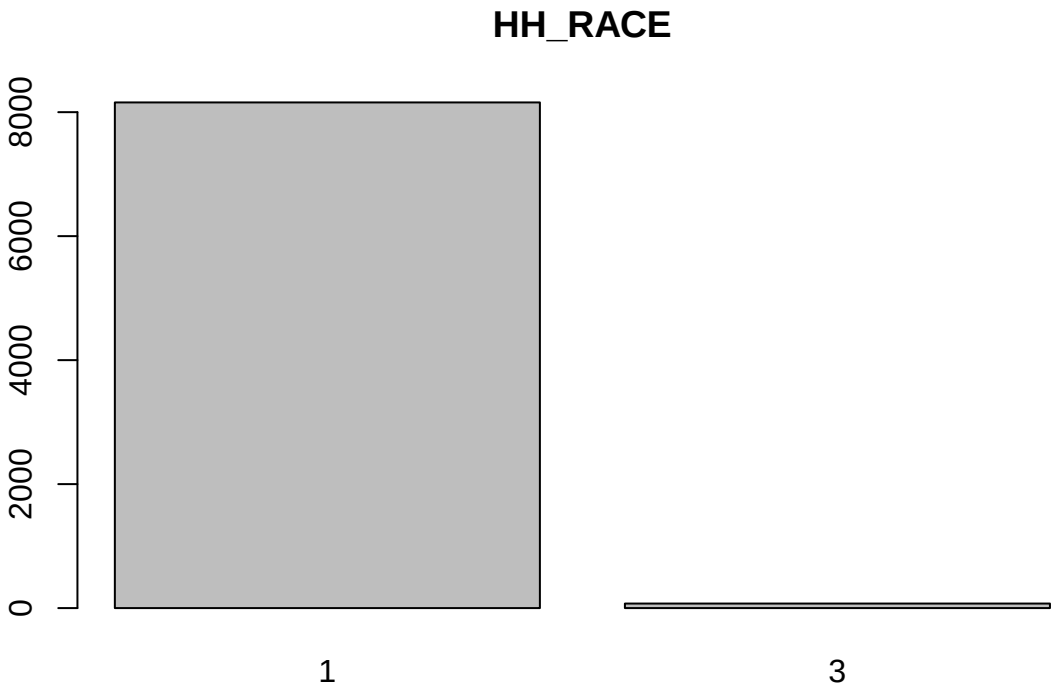

TYPE.OF.FORMULAT
0
500
1000
1500
ANT CVTY BRTH FRSHNR
SNSTV ANTC ANT GNGV
WHTN ANTC BRTH FRSH

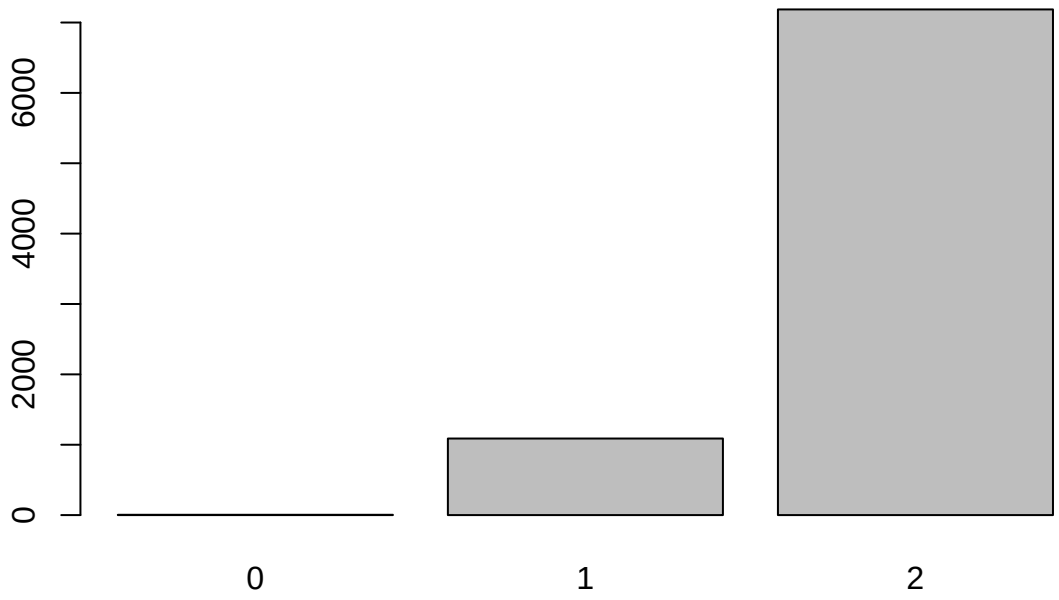

COLOR
0
500
1000
1500
BLU
MISSING
PINK
RED
SPARKLE
WHITE

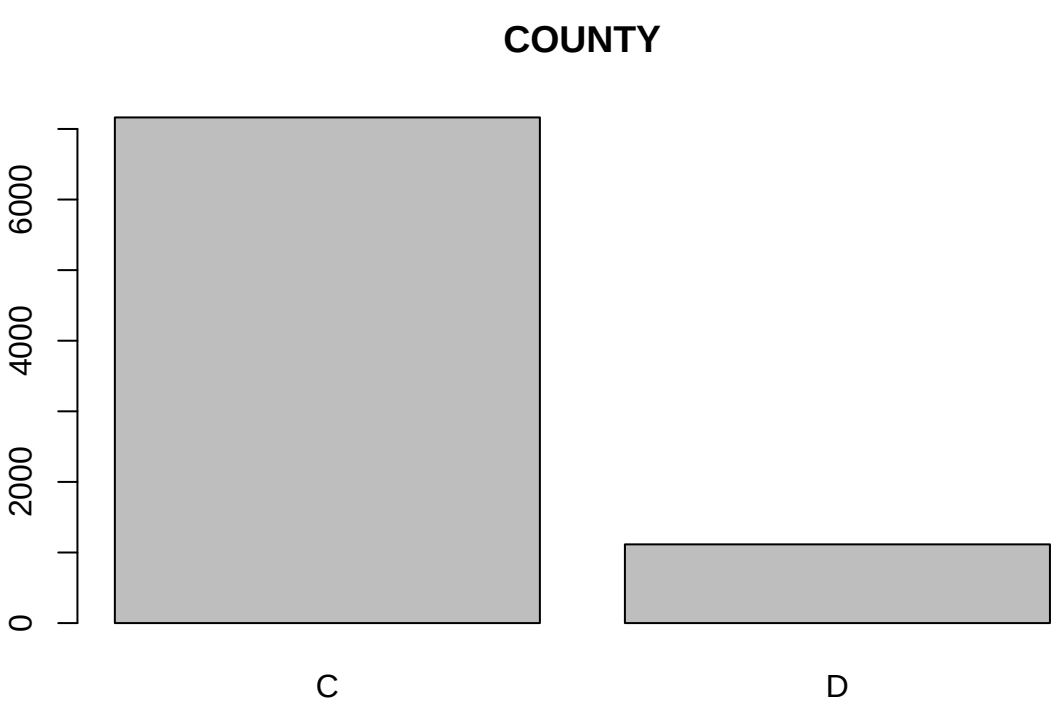

Panelist.Type
0
1000
2000
3000
4000
5000
0
5
6

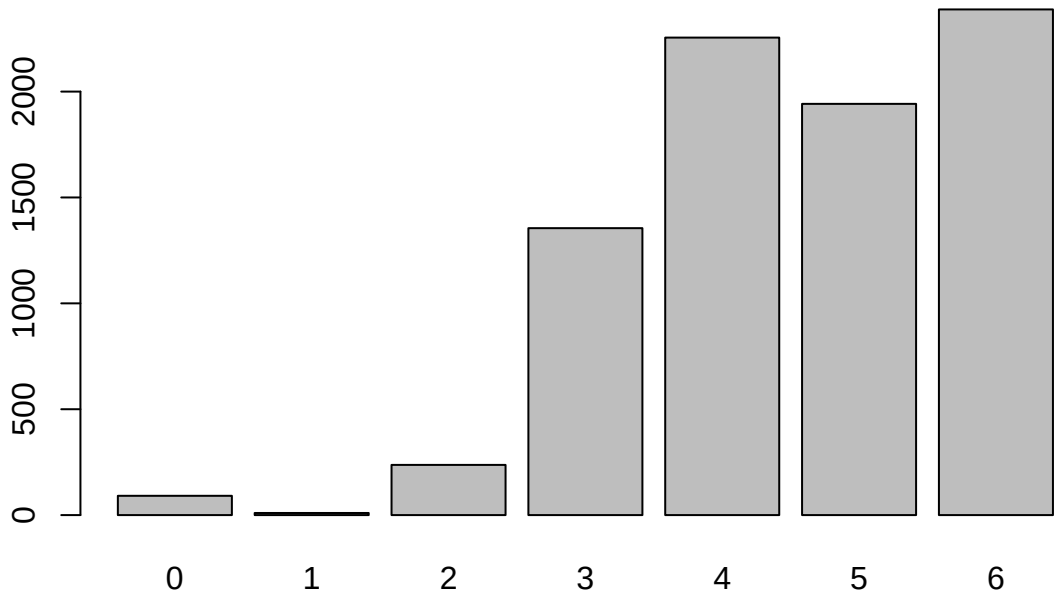

Combined.Pre.Tax.Income.of.HH
0
200
400
600
800
1200
0
1
2
3
4
5
6
7
8
9
10
11
12

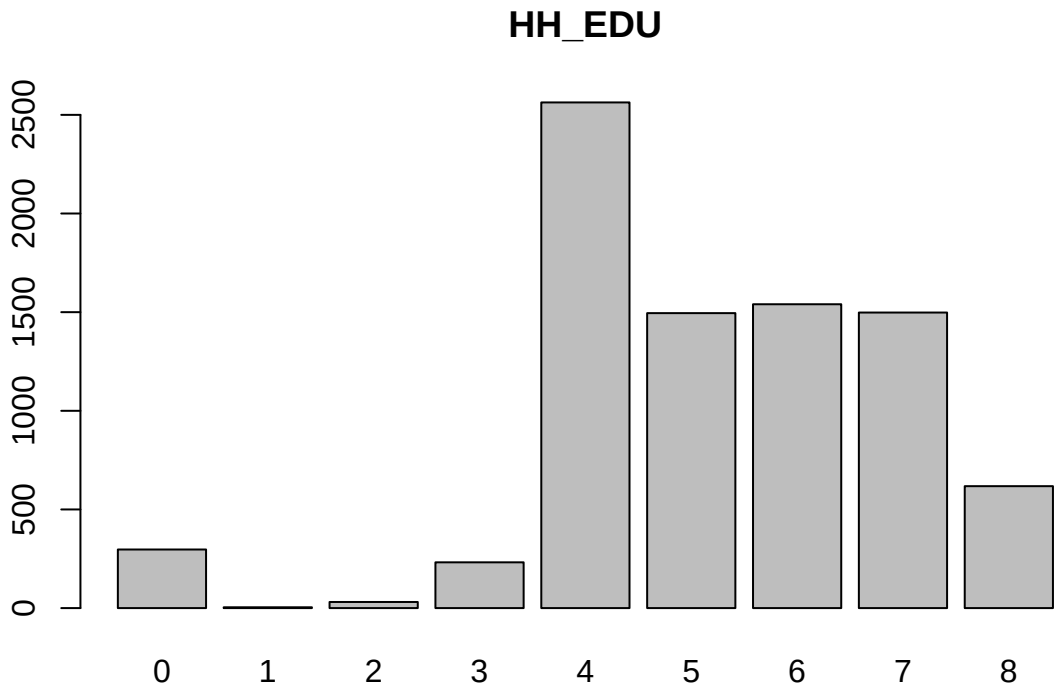

Family.Size
0
500
1500
2500
1
2
3
4
5
6

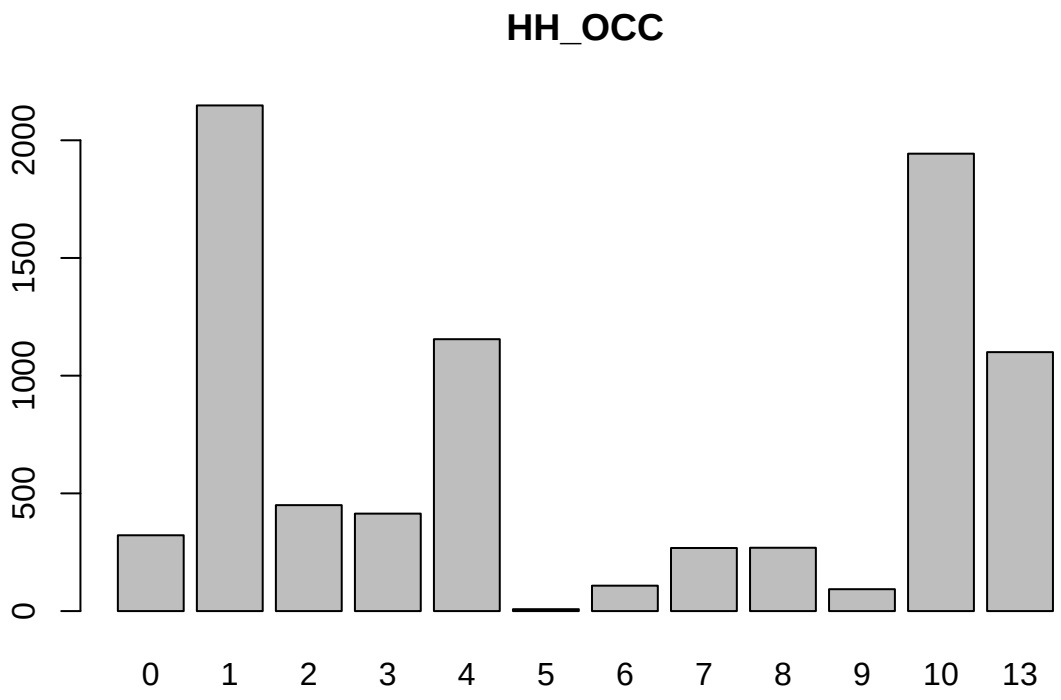

HH_RACE
0
2000
4000
6000
8000
1
3


Type.of.Residential.Possession
0
2000
4000
6000
0
1
2

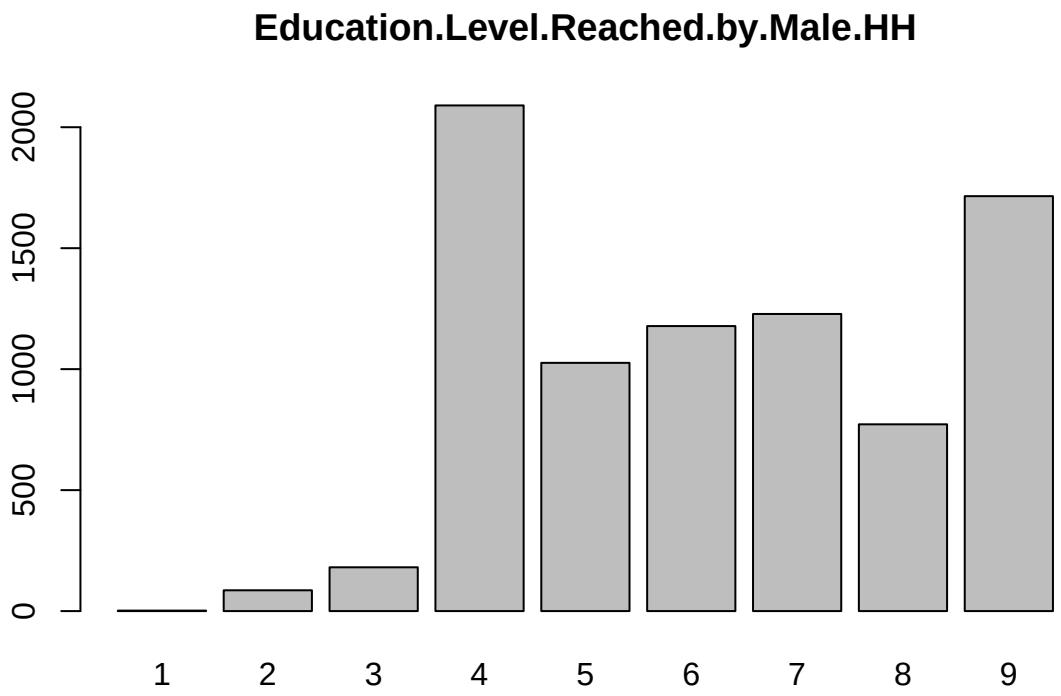

COUNTY
0
2000
4000
6000
C
D

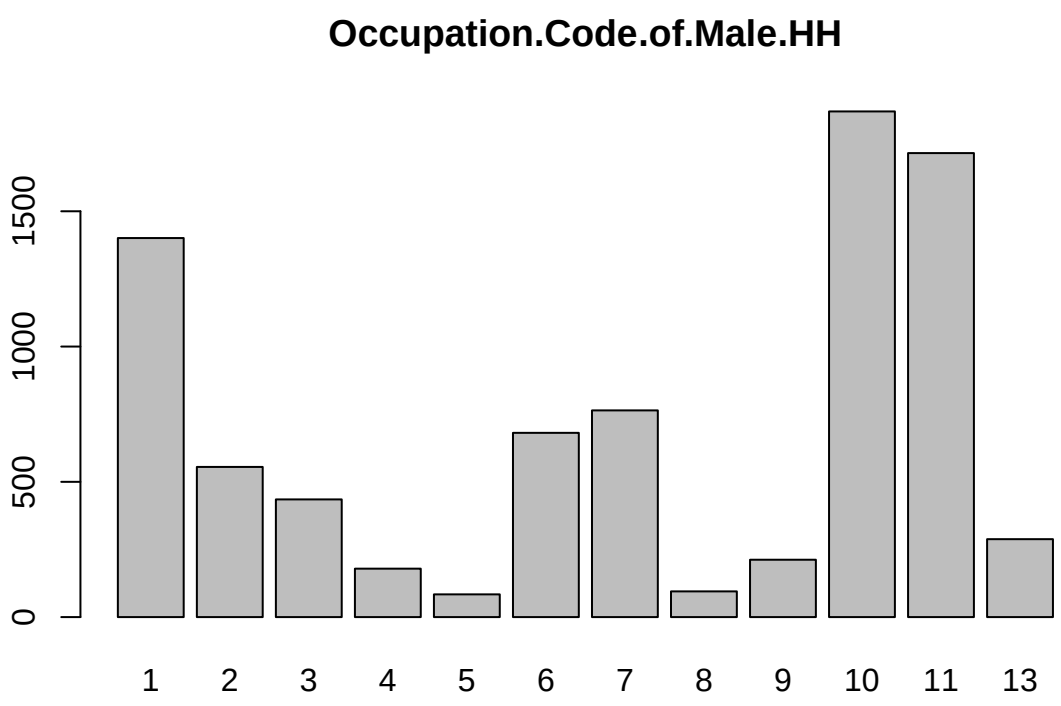

HH_AGE
0
500
1000
1500
2000
0
1
2
3
4
5
6

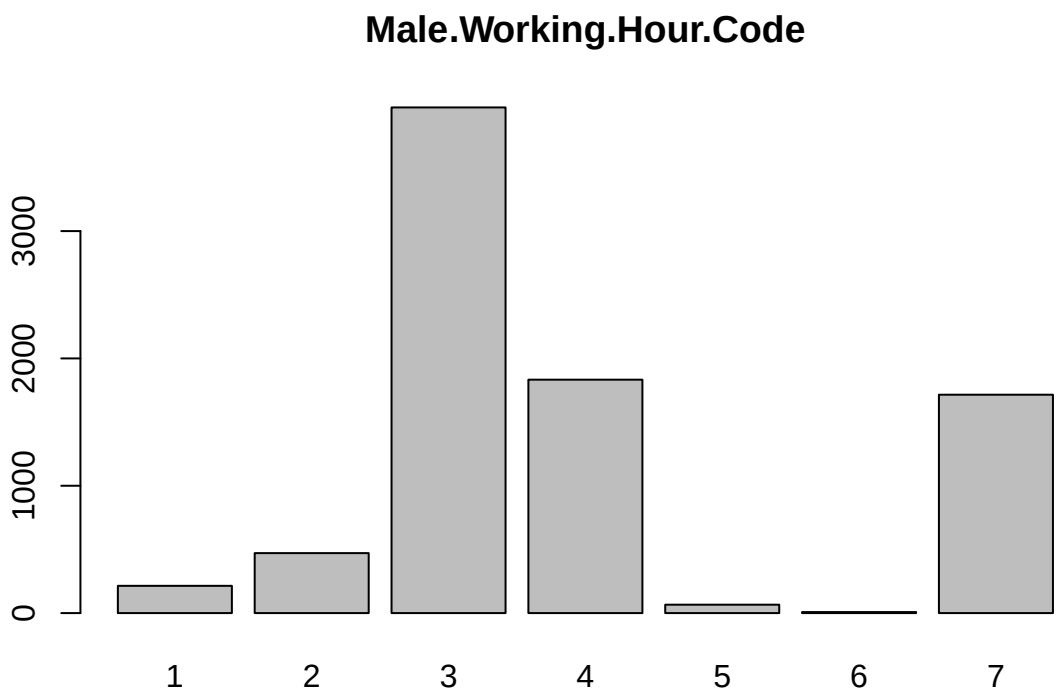

HH_EDU
0
500
1000
1500
2000
2500
0
1
2
3
4
5
6
7
8

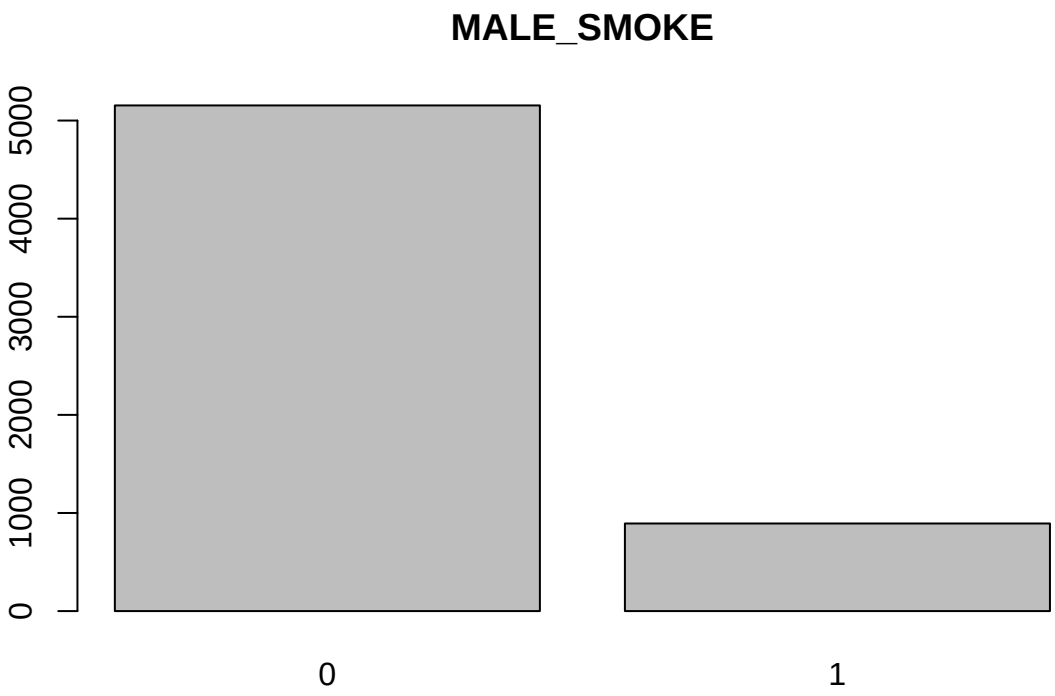

HH_OCC
0
500
1000
1500
2000
0
1
2
3
4
5
6
7
8
9
10
13

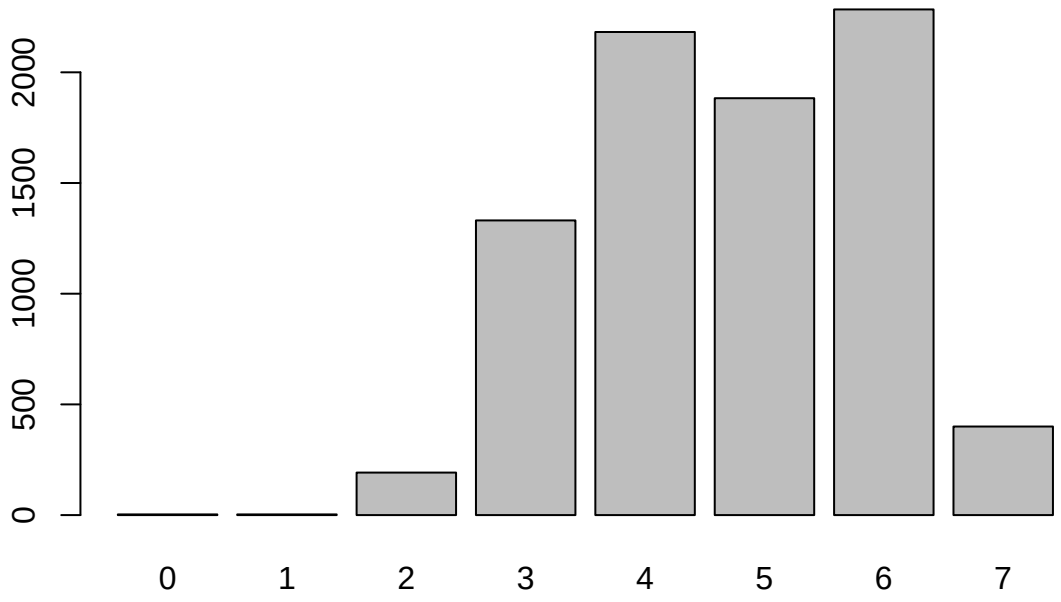

Age.Group.Applied.to.Male.HH
0
500
1000
1500
0
1
2
3
4
5
6
7

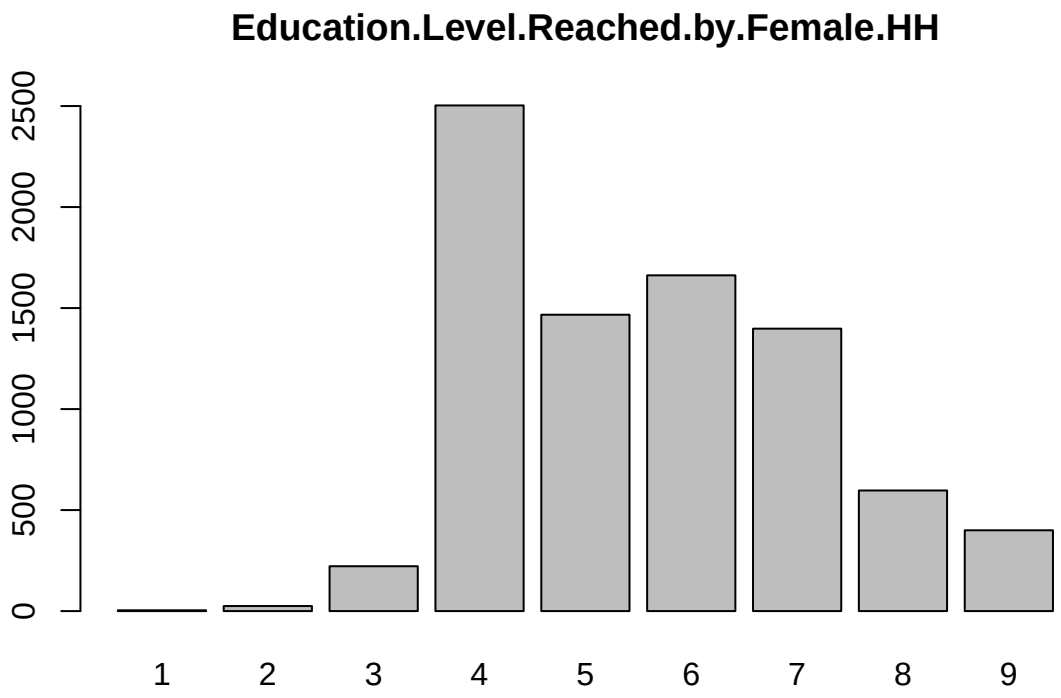

Education.Level.Reached.by.Male.HH
0
500
1000
1500
2000
1
2
3
4
5
6
7
8
9

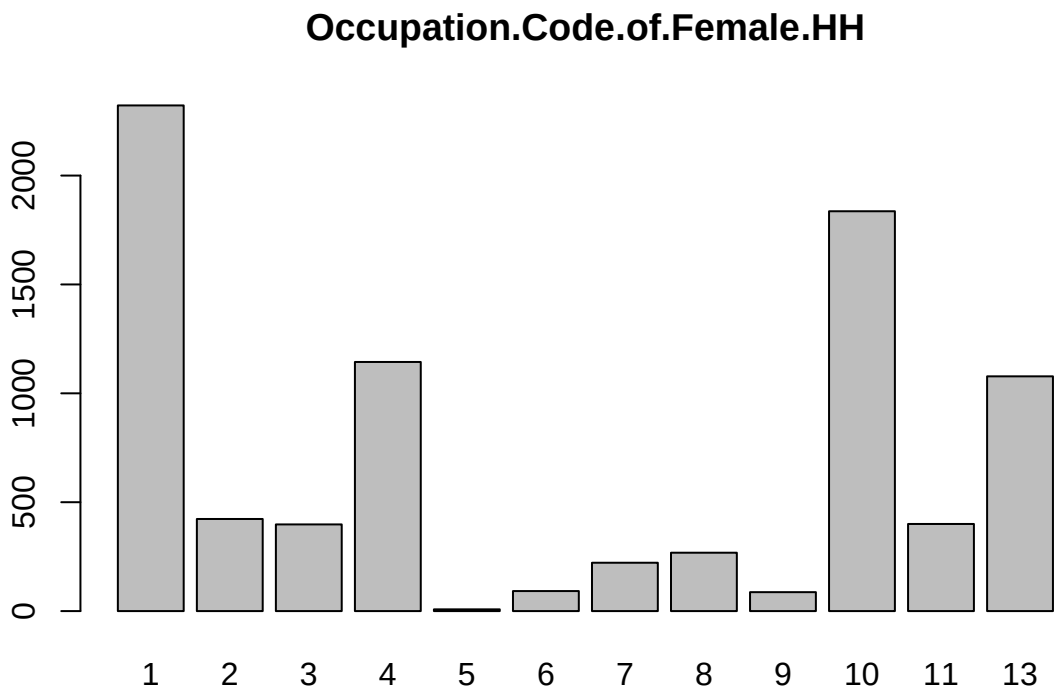

Occupation.Code.of.Male.HH
0
500
1000
1500
1
2
3
4
5
6
7
8
9
10
11
13

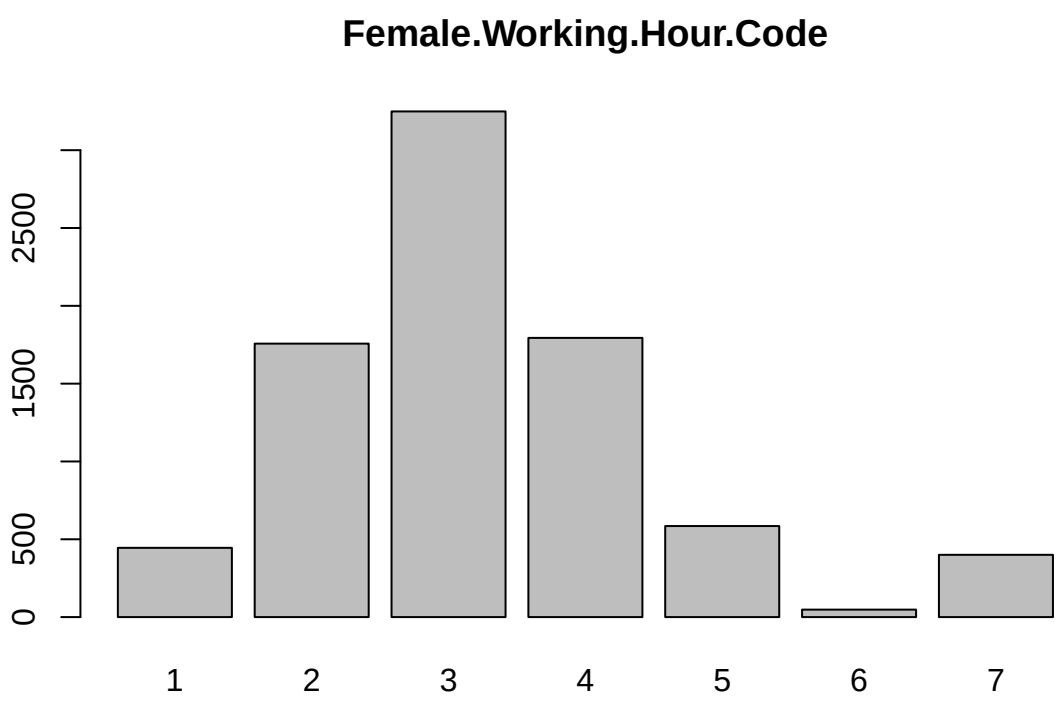

Male.Working.Hour.Code
0
1000
2000
3000
1
2
3
4
5
6
7

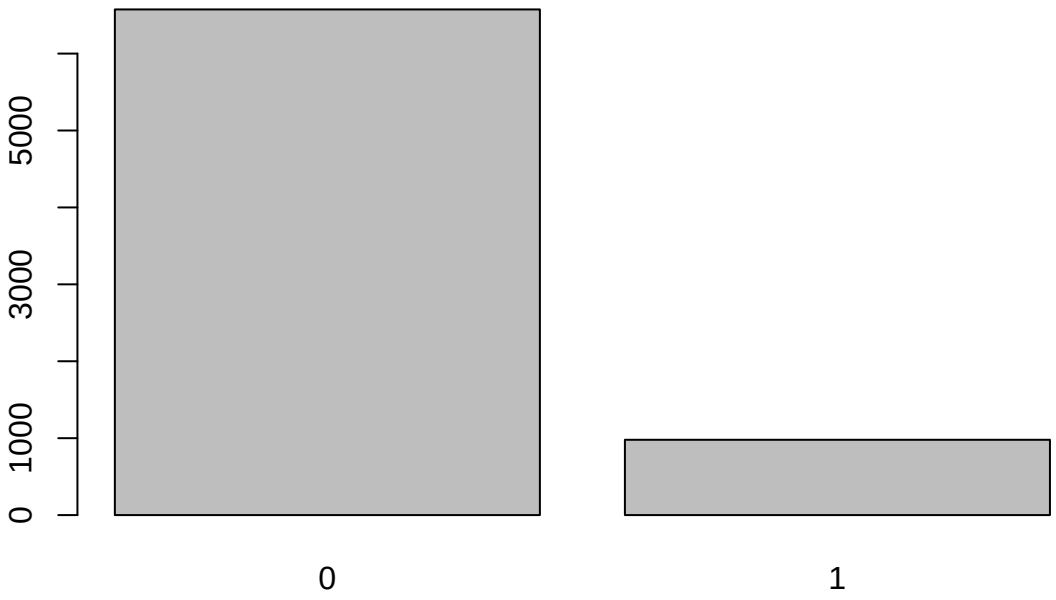

MALE_SMOKE
0
1000
2000
3000
4000
5000
0
1

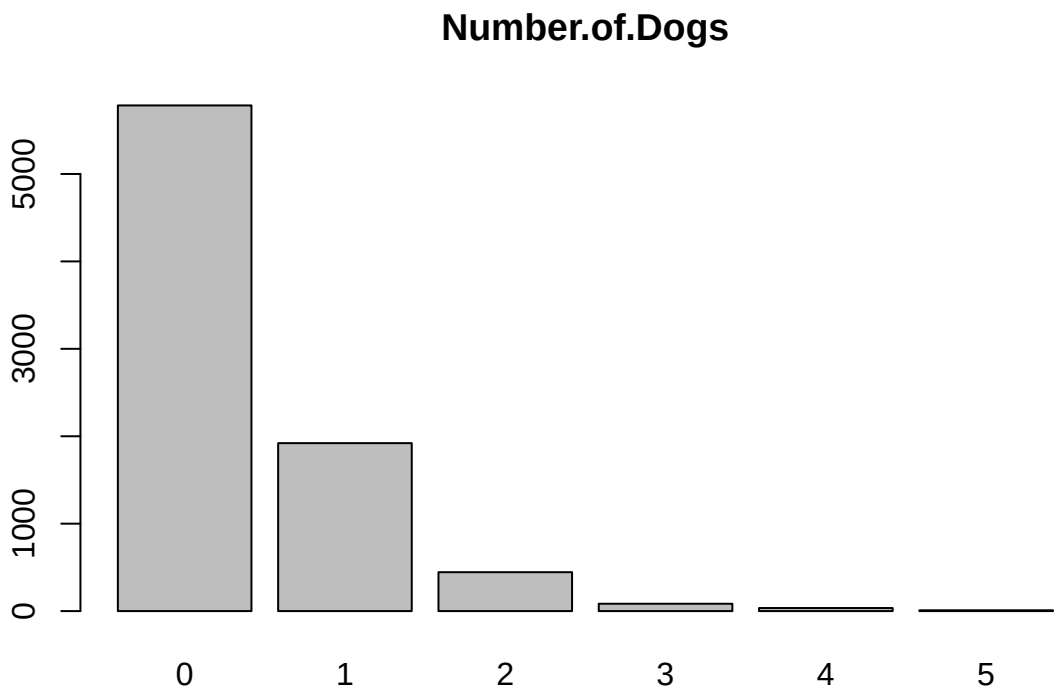

Age.Group.Applied.to.Female.HH
0
500
1000
1500
2000
0
1
2
3
4
5
6
7

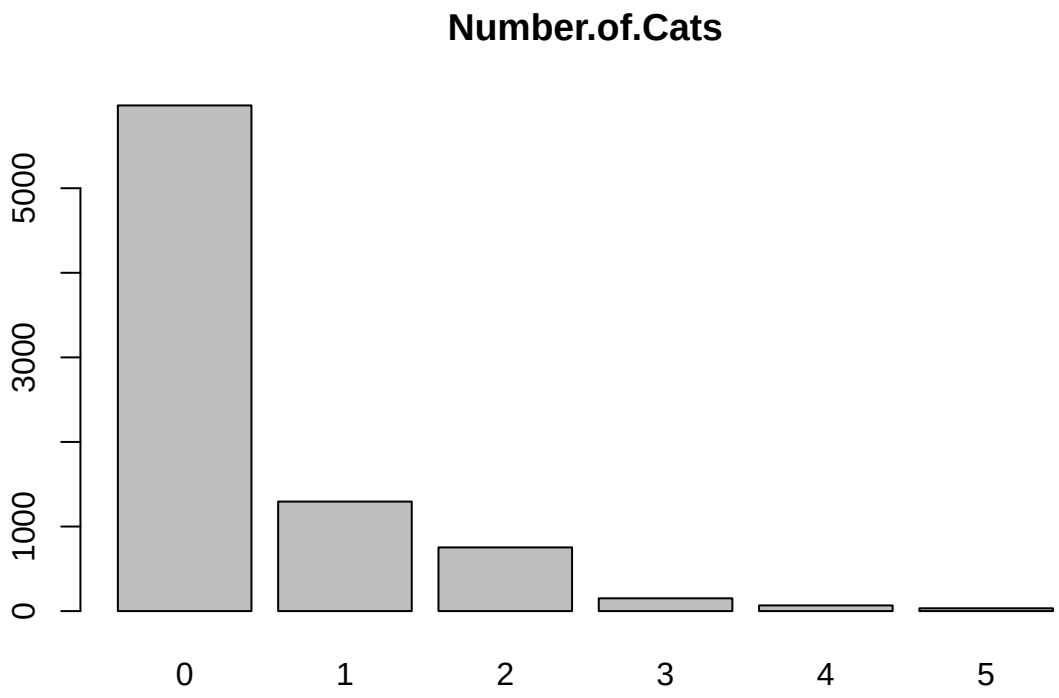

Education.Level.Reached.by.Female.HH
0
500
1000
1500
2000
2500
1
2
3
4
5
6
7
8
9

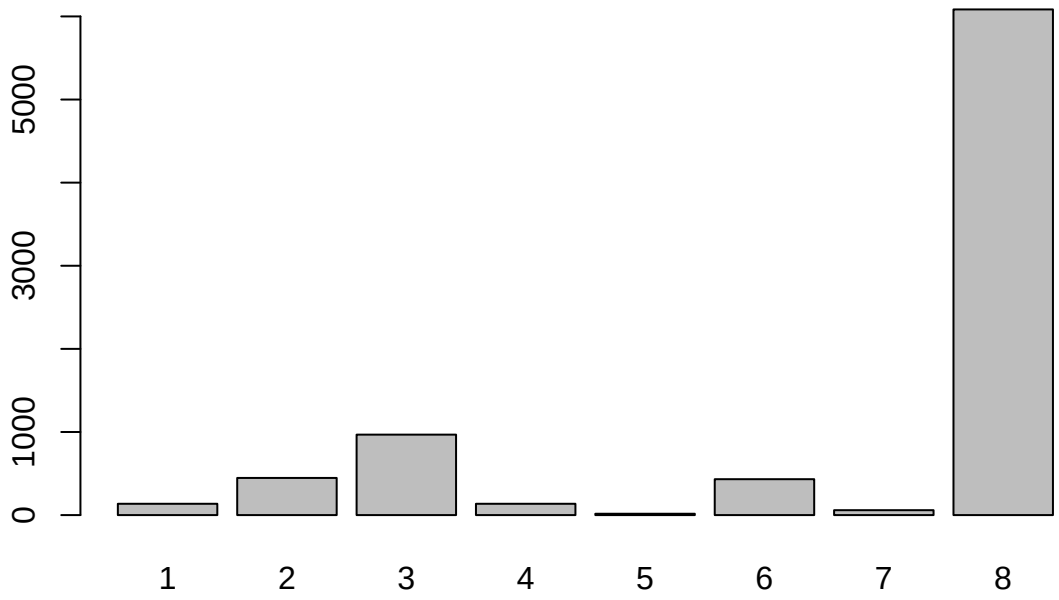

Occupation.Code.of.Female.HH
0
500
1000
1500
2000
1
2
3
4
5
6
7
8
9
10
11
13


Female.Working.Hour.Code
0
500
1500
2500
1
2
3
4
5
6
7

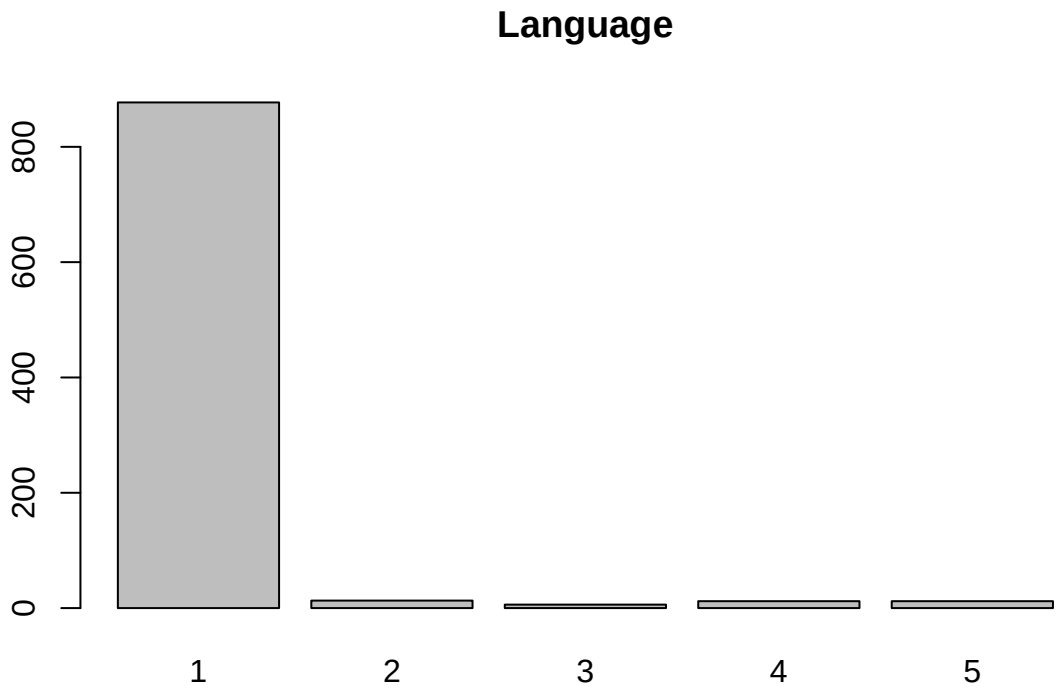

FEM_SMOKE
0
1000
3000
5000
0
1

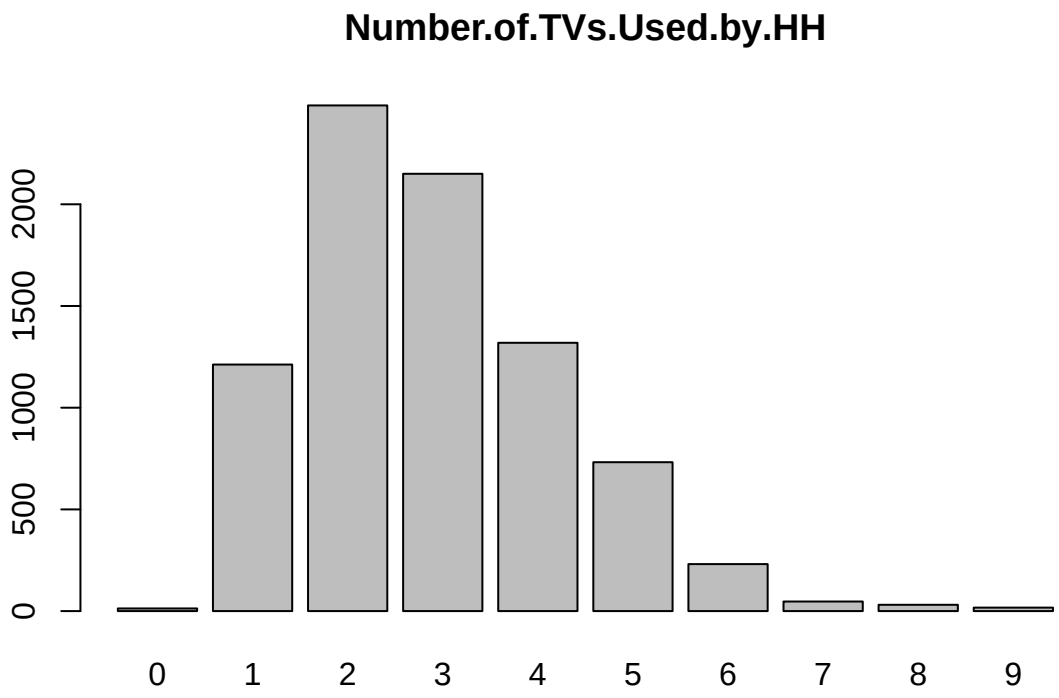

Number.of.Dogs
0
1000
3000
5000
0
1
2
3
4
5

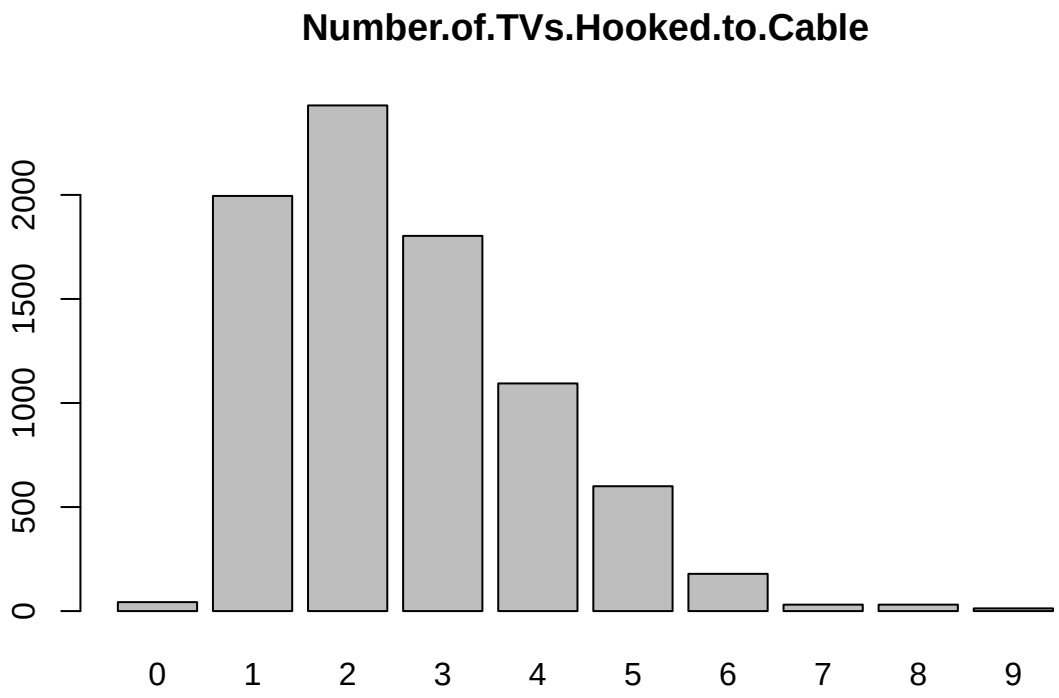

Number.of.Cats
0
1000
3000
5000
0
1
2
3
4
5

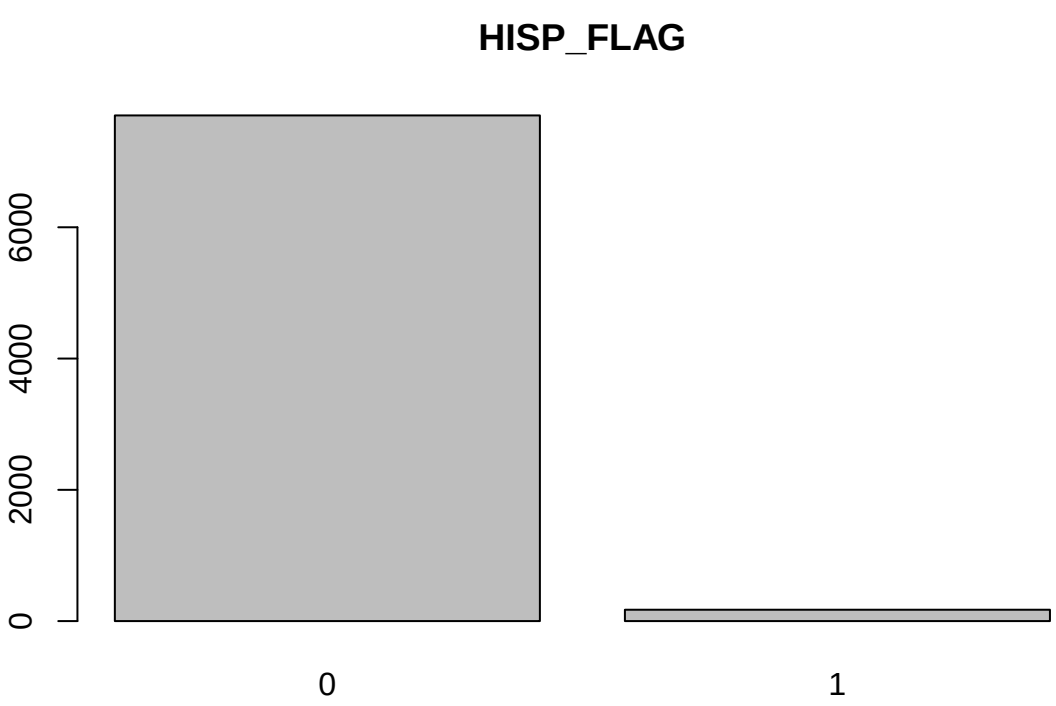

Children.Group.Code
0
1000
3000
5000
1
2
3
4
5
6
7
8

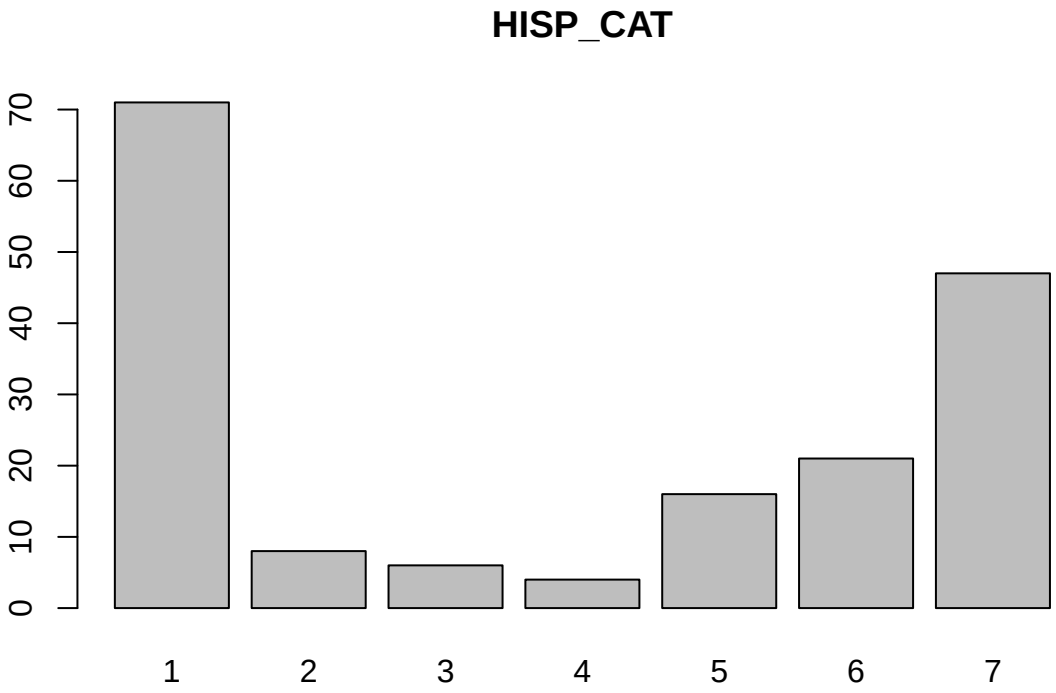

Marital.Status
0
1000
3000
5000
0
1
2
3
4
5

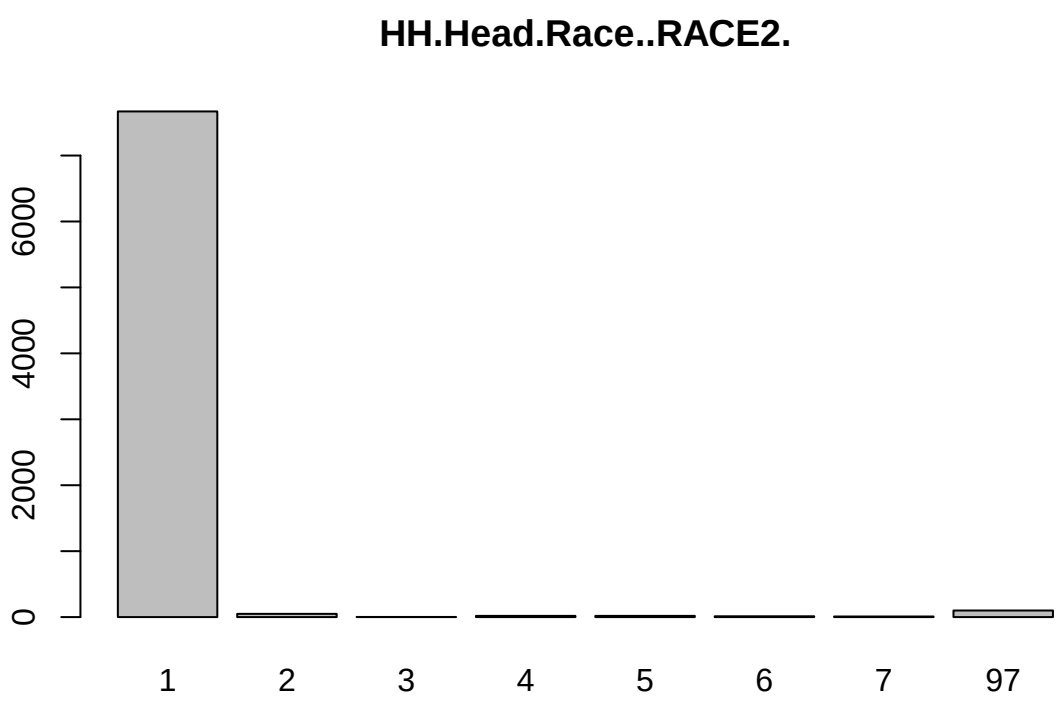

Language
0
200
400
600
800
1
2
3
4
5

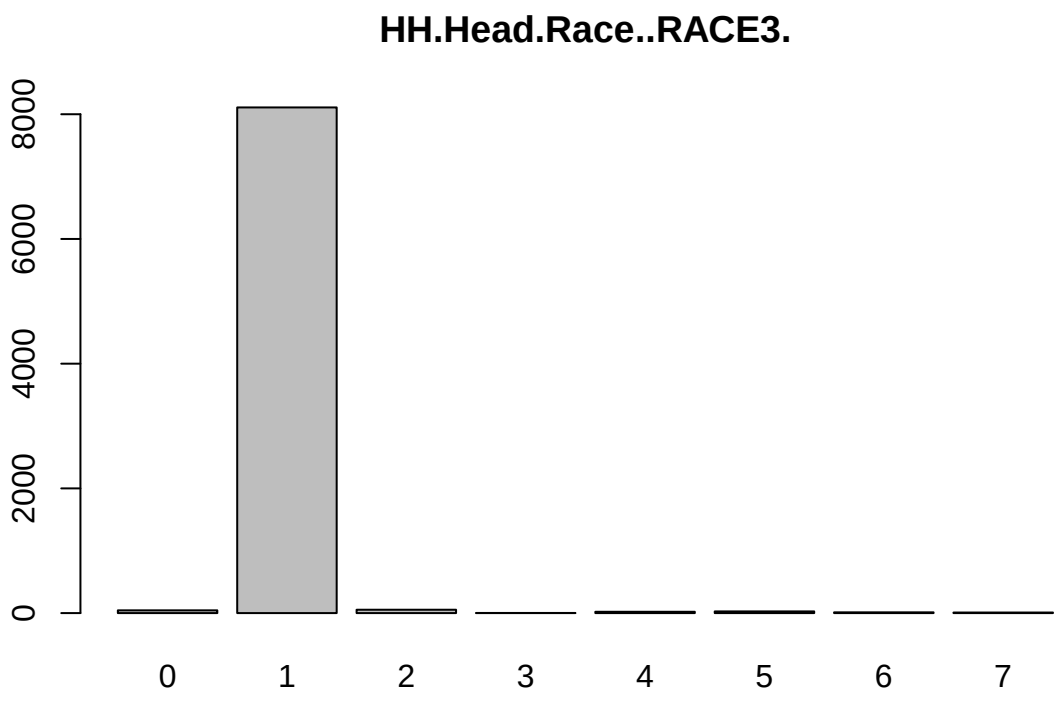

Number.of.TVs.Used.by.HH
0
500
1000
1500
2000
0
1
2
3
4
5
6
7
8
9

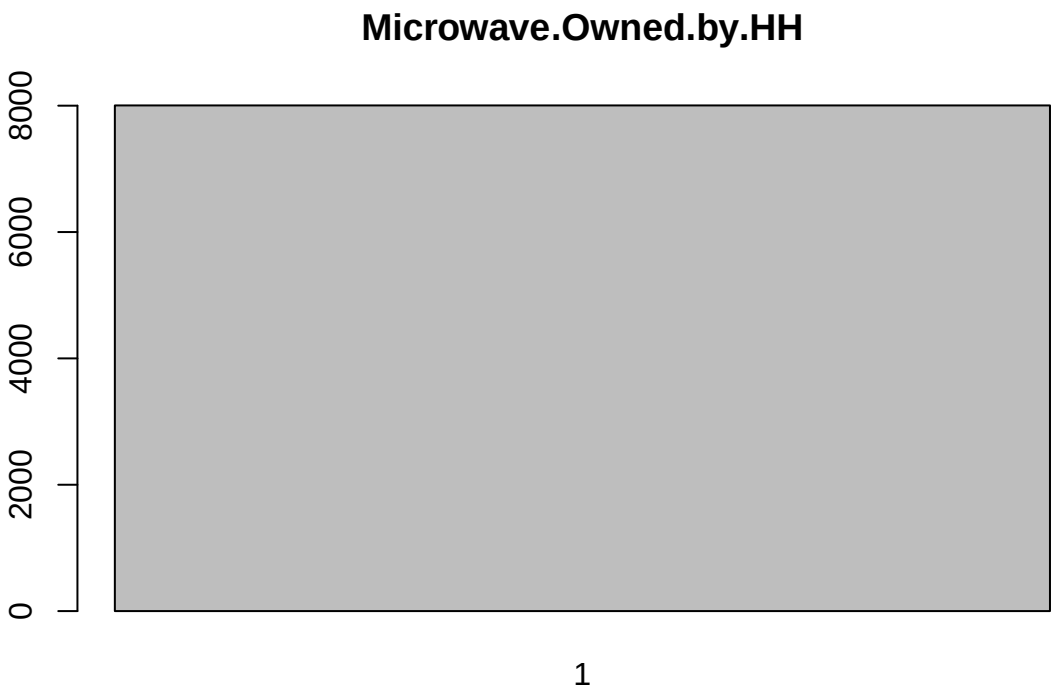

Number.of.TVs.Hooked.to.Cable
0
500
1000
1500
2000
0
1
2
3
4
5
6
7
8
9

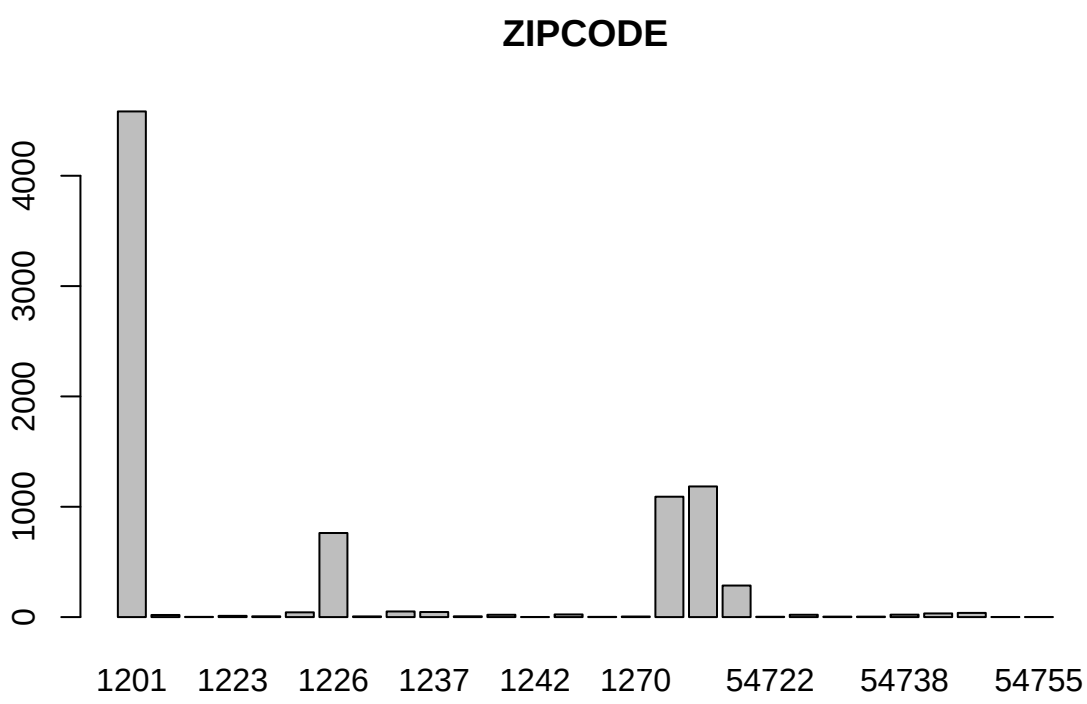

HISP_FLAG
0
2000
4000
6000
0
1

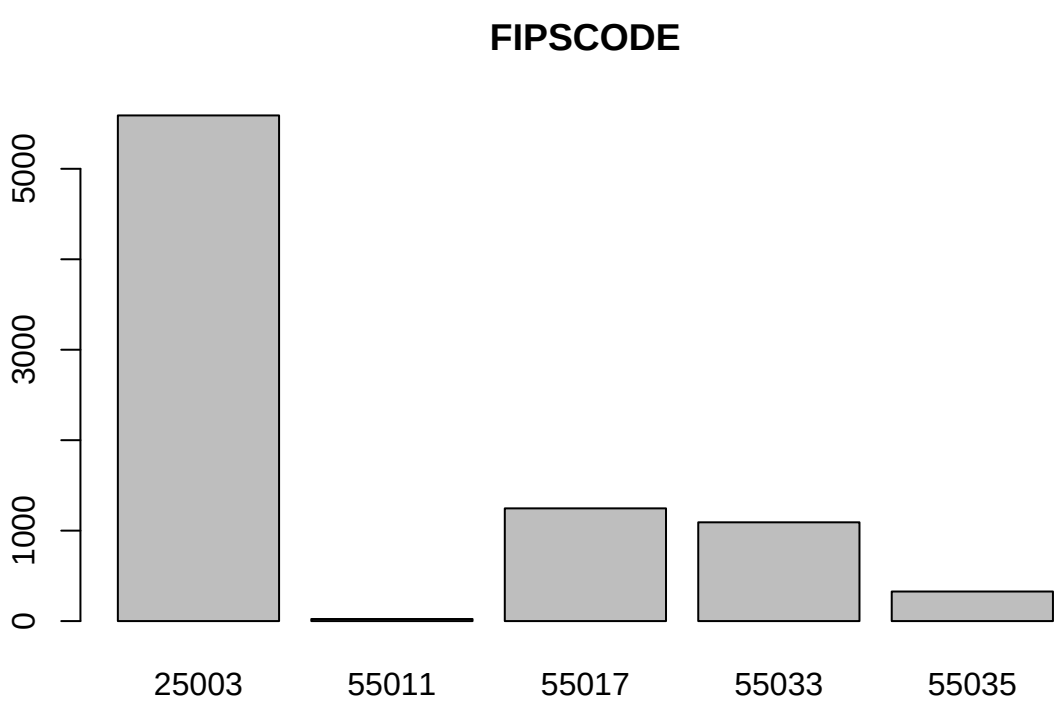

HISP_CAT
0
10
20
30
40
50
60
70
1
2
3
4
5
6
7

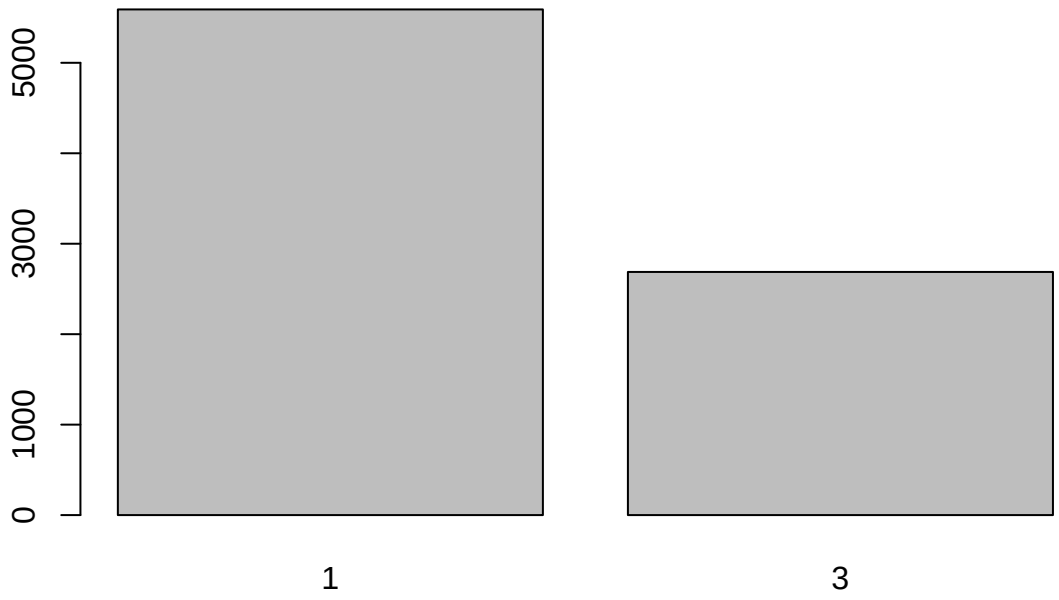

HH.Head.Race..RACE2.
0
2000
4000
6000
1
2
3
4
5
6
7
97

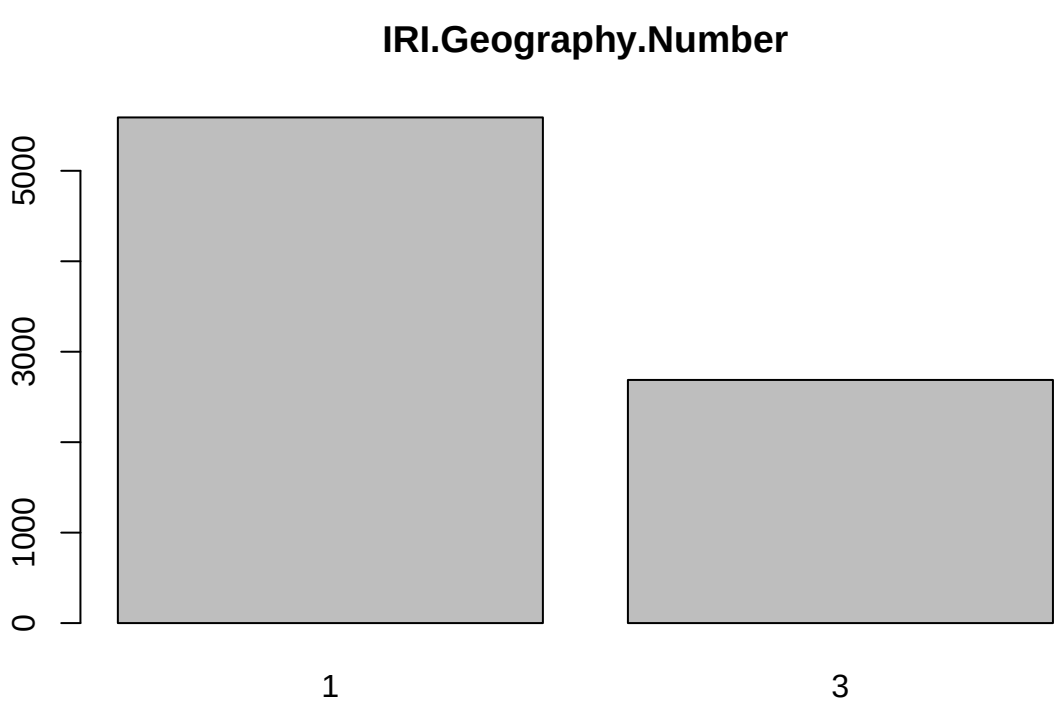

HH.Head.Race..RACE3.
0
2000
4000
6000
8000
0
1
2
3
4
5
6
7


Microwave.Owned.by.HH
0
2000
4000
6000
8000
1


ZIPCODE
0
1000
2000
3000
4000
1201
1223
1226
1237
1242
1270
54722
54738
54755


FIPSCODE
0
1000
3000
5000
25003
55011
55017
55033
55035


market.based.upon.zipcode
0
1000
3000
5000
1
3


IRI.Geography.Number
0
1000
3000
5000
1
3


```
# a lot of "MISSING" values on the COLOR plot, ignore for now
```

```
for (x in c(1:ncol(merged.data.reduced))) {
  if (is.factor(merged.data.reduced[,x])) {
    print(paste(colnames(merged.data.reduced)[x], "=", nlevels(merged.data.reduced[,x])))
  }
}
```

```
## [1] "PANID = 2501"
## [1] "VEND = 6"
## [1] "OUTLET = 2"
## [1] "ITEM = 328"
## [1] "F = 5"
## [1] "D = 3"
## [1] "PR = 2"
## [1] "VOL_EQ = 7"
## [1] "SIZE = 1"
## [1] "FLAVOR.SCENT = 47"
## [1] "FORM = 5"
## [1] "PACKAGE = 10"
## [1] "PRODUCT.TYPE = 1"
## [1] "STORE.LOCATION = 1"
## [1] "ADDITIVES = 19"
## [1] "TYPE.OF.FORMULAT = 31"
## [1] "COLOR = 18"
## [1] "Panelist.Type = 3"
## [1] "Combined.Pre.Tax.Income.of.HH = 13"
```

```
## [1] "Family.Size = 6"
## [1] "HH_RACE = 2"
## [1] "Type.of.Residential.Possession = 3"
## [1] "COUNTY = 2"
## [1] "HH_AGE = 7"
## [1] "HH_EDU = 9"
## [1] "HH_OCC = 12"
## [1] "Age.Group.Applied.to.Male.HH = 8"
## [1] "Education.Level.Reached.by.Male.HH = 9"
## [1] "Occupation.Code.of.Male.HH = 12"
## [1] "Male.Working.Hour.Code = 7"
## [1] "MALE_SMOKE = 2"
## [1] "Age.Group.Applied.to.Female.HH = 8"
## [1] "Education.Level.Reached.by.Female.HH = 9"
## [1] "Occupation.Code.of.Female.HH = 12"
## [1] "Female.Working.Hour.Code = 7"
## [1] "FEM_SMOKE = 2"
## [1] "Number.of.Dogs = 6"
## [1] "Number.of.Cats = 6"
## [1] "Children.Group.Code = 8"
## [1] "Marital.Status = 6"
## [1] "Language = 5"
## [1] "Number.of.TVs.Used.by.HH = 10"
## [1] "Number.of.TVs.Hooked.to.Cable = 10"
## [1] "HISP_FLAG = 2"
## [1] "HISP_CAT = 7"
## [1] "HH.Head.Race..RACE2. = 8"
## [1] "HH.Head.Race..RACE3. = 8"
## [1] "Microwave.Owned.by.HH = 1"
## [1] "ZIPCODE = 28"
## [1] "FIPSCODE = 5"
## [1] "market.based.upon.zipcode = 2"
## [1] "IRI.Geography.Number = 2"
## [1] "EXT_FACT = 1"
```

We can get rid of all the columns that have a only 1 factor: “Microwave.Owned.by.HH = 1” “EXT_FACT = 1” “STORE.LOCATION = 1” “PRODUCT.TYPE = 1” “SIZE = 1”

```
cnames<-colnames(merged.data.reduced)
cols.remove<-c(which(cnames == "Microwave.Owned.by.HH"),
  which(cnames == "EXT_FACT"),
  which(cnames == "STORE.LOCATION"),
  which(cnames == "PRODUCT.TYPE"),
  which(cnames == "SIZE"),
  which(cnames == "COUNT"),
  which(cnames == "ITEM"),
  which(cnames == "VENDORITEMCOUNT"))

merged.data.reduced<-merged.data.reduced[,-cols.remove]

for (x in c(1:ncol(merged.data.reduced))) {
  if (is.factor(merged.data.reduced[,x])) {
    print(paste(colnames(merged.data.reduced)[x], "=", nlevels(merged.data.reduced[,x])))
  }
}
```

```
## [1] "PANID = 2501"
## [1] "VEND = 6"
## [1] "OUTLET = 2"
## [1] "F = 5"
## [1] "D = 3"
## [1] "PR = 2"
## [1] "VOL_EQ = 7"
## [1] "FLAVOR.SCENT = 47"
## [1] "FORM = 5"
## [1] "PACKAGE = 10"
## [1] "ADDITIVES = 19"
## [1] "TYPE.OF.FORMULAT = 31"
## [1] "COLOR = 18"
## [1] "Panelist.Type = 3"
## [1] "Combined.Pre.Tax.Income.of.HH = 13"
## [1] "Family.Size = 6"
## [1] "HH_RACE = 2"
## [1] "Type.of.Residential.Possession = 3"
## [1] "COUNTY = 2"
## [1] "HH_AGE = 7"
## [1] "HH_EDU = 9"
## [1] "HH_OCC = 12"
## [1] "Age.Group.Applied.to.Male.HH = 8"
## [1] "Education.Level.Reached.by.Male.HH = 9"
## [1] "Occupation.Code.of.Male.HH = 12"
## [1] "Male.Working.Hour.Code = 7"
## [1] "MALE_SMOKE = 2"
## [1] "Age.Group.Applied.to.Female.HH = 8"
## [1] "Education.Level.Reached.by.Female.HH = 9"
## [1] "Occupation.Code.of.Female.HH = 12"
## [1] "Female.Working.Hour.Code = 7"
## [1] "FEM_SMOKE = 2"
## [1] "Number.of.Dogs = 6"
## [1] "Number.of.Cats = 6"
## [1] "Children.Group.Code = 8"
## [1] "Marital.Status = 6"
## [1] "Language = 5"
## [1] "Number.of.TVs.Used.by.HH = 10"
## [1] "Number.of.TVs.Hooked.to.Cable = 10"
## [1] "HISP_FLAG = 2"
## [1] "HISP_CAT = 7"
## [1] "HH.Head.Race..RACE2. = 8"
## [1] "HH.Head.Race..RACE3. = 8"
## [1] "ZIPCODE = 28"
## [1] "FIPSCODE = 5"
## [1] "market.based.upon.zipcode = 2"
## [1] "IRI.Geography.Number = 2"
```

OK, now we have a pretty decent dataset, lets try a model:

```
library("mlogit")
```

```
## Warning: package 'mlogit' was built under R version 3.3.3
```

```
## Loading required package: Formula
```

```
## Loading required package: maxLik
## Warning: package 'maxLik' was built under R version 3.3.3
## Loading required package: miscTools
## Warning: package 'miscTools' was built under R version 3.3.3
##
## Please cite the 'maxLik' package as:
## Henningsen, Arne and Toomet, Ott (2011). maxLik: A package for maximum likelihood estimation in R. C
##
## If you have questions, suggestions, or comments regarding the 'maxLik' package, please use a forum o
## https://r-forge.r-project.org/projects/maxlik/
```

```
head(merged.data.reduced)
```

```
##     PANID  VEND UNITS OUTLET DOLLARS    F D PR       VOL_EQ FLAVOR.SCENT
## 3 1100032 33200     2     GR    1.00    B 2  1 MEDIUM_SMALL     ORIGINAL
## 4 1100032 35000     1     GR    1.49    A 1  1       MEDIUM      REGULAR
## 5 1100057 33200     1     GR    3.99 NONE 0  0        SMALL   FRESH MINT
## 6 1100057 33200     2     GR    7.98 NONE 0  0        SMALL   FRESH MINT
## 7 1100057 33200     2     GR    7.98 NONE 0  0 MEDIUM_SMALL   FRESH MINT
## 8 1100180 53100     5     GR    4.80    B 2  1       MEDIUM      REGULAR
##            FORM     PACKAGE            ADDITIVES     TYPE.OF.FORMULAT
## 3         PASTE TUBE IN BOX      SODIUM FLUORIDE ANTI CAVITY FLUORIDE
## 4         PASTE TUBE IN BOX      SODIUM FLUORIDE          ANTI CAVITY
## 5         PASTE TUBE IN BOX BKSODA PEROXIDE FLRD WHTNN TRT CTR AN CVT
## 6         PASTE TUBE IN BOX BKSODA PEROXIDE FLRD WHTNN TRT CTR AN CVT
## 7         PASTE TUBE IN BOX BKSODA PEROXIDE FLRD WHTNN TRT CTR AN CVT
## 8 PASTE AND GEL TUBE IN BOX      SODIUM FLUORIDE    TRIPLE PROTECTION
##              COLOR Panelist.Type Combined.Pre.Tax.Income.of.HH Family.Size
## 3          MISSING             0                             6           2
## 4            WHITE             0                             6           2
## 5          MISSING             0                            10           2
## 6          MISSING             0                            10           2
## 7          MISSING             0                            10           2
## 8 RED WHITE & BLUE             6                            11           2
##   HH_RACE Type.of.Residential.Possession COUNTY HH_AGE HH_EDU HH_OCC
## 3       1                              2      C      5      7      6
## 4       1                              2      C      5      7      6
## 5       1                              2      C      6      6      4
## 6       1                              2      C      6      6      4
## 7       1                              2      C      6      6      4
## 8       1                              2      C      5      5      1
##   Age.Group.Applied.to.Male.HH Education.Level.Reached.by.Male.HH
## 3                            7                                   9
## 4                            7                                   9
## 5                            6                                   4
## 6                            6                                   4
## 7                            6                                   4
## 8                            5                                   6
##   Occupation.Code.of.Male.HH Male.Working.Hour.Code MALE_SMOKE
## 3                         11                      7       <NA>
## 4                         11                      7       <NA>
## 5                         10                      4          0
## 6                         10                      4          0
```

```
## 7                         10                      4          0
## 8                          6                      3          1
##   Age.Group.Applied.to.Female.HH Education.Level.Reached.by.Female.HH
## 3                              5                                     7
## 4                              5                                     7
## 5                              6                                     6
## 6                              6                                     6
## 7                              6                                     6
## 8                              5                                     5
##   Occupation.Code.of.Female.HH Female.Working.Hour.Code FEM_SMOKE
## 3                            6                        3         0
## 4                            6                        3         0
## 5                            4                        3         0
## 6                            4                        3         0
## 7                            4                        3         0
## 8                            1                        3         0
##   Number.of.Dogs Number.of.Cats Children.Group.Code Marital.Status
## 3              0              1                   3              1
## 4              0              1                   3              1
## 5              1              0                   8              2
## 6              1              0                   8              2
## 7              1              0                   8              2
## 8              1              1                   8              2
##   Language Number.of.TVs.Used.by.HH Number.of.TVs.Hooked.to.Cable
## 3     <NA>                        2                             1
## 4     <NA>                        2                             1
## 5     <NA>                        3                             3
## 6     <NA>                        3                             3
## 7     <NA>                        3                             3
## 8     <NA>                        2                             2
##   HISP_FLAG HISP_CAT HH.Head.Race..RACE2. HH.Head.Race..RACE3. ZIPCODE
## 3         0     <NA>                    1                    1    1201
## 4         0     <NA>                    1                    1    1201
## 5         0     <NA>                    1                    1    1201
## 6         0     <NA>                    1                    1    1201
## 7         0     <NA>                    1                    1    1201
## 8         0     <NA>                    1                    1    1201
##   FIPSCODE market.based.upon.zipcode IRI.Geography.Number
## 3    25003                         1                    1
## 4    25003                         1                    1
## 5    25003                         1                    1
## 6    25003                         1                    1
## 7    25003                         1                    1
## 8    25003                         1                    1
```

```
alevels<-c(levels(merged.data.reduced$VEND))
cvars<-c(colnames(merged.data.reduced)[c(16:49)])
vnames<-c(colnames(merged.data.reduced)[c(3:15)])
# I don't think this is correct yet
transformed.data<-mlogit.data(merged.data.reduced, shape = "wide", varying = 3:15, v.names=vnames, choi
```