week1_caravan_insurance_cx_profiling_r

Renjith Madhavan

1/30/2017

Introduction

Customer profiling for target marketing is an important area for potential application for data mining techniques. In this analysis we derive a typical profile of customers holding Caravan Insurance Policy. Such a profile would help identify customers with a similar profile who do not have caravan policies to both existing nonholders of caravan policies or new customers with similar profiles.

```
library(sqldf)
```

```
## Loading required package: gsubfn
```

```
## Loading required package: proto
```

```
## Could not load tcltk.  Will use slower R code instead.
```

```
## Loading required package: RSQLite
```

```
## Loading required package: DBI
```

```
setwd("/tmp/working/kaggle/rstats/mds564")
dataset1 <- read.csv("ticdata2000.txt", sep = "\t", header = FALSE)
summary(dataset1)
```

```
##        V1              V2              V3             V4
##  Min.   : 1.00   Min.   : 1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:10.00   1st Qu.: 1.000   1st Qu.:2.000   1st Qu.:2.000
##  Median :30.00   Median : 1.000   Median :3.000   Median :3.000
##  Mean   :24.25   Mean   : 1.111   Mean   :2.679   Mean   :2.991
##  3rd Qu.:35.00   3rd Qu.: 1.000   3rd Qu.:3.000   3rd Qu.:3.000
##  Max.   :41.00   Max.   :10.000   Max.   :5.000   Max.   :6.000
##        V5              V6              V7             V8
##  Min.   : 1.000   Min.   :0.0000   Min.   :0.000   Min.   :0.00
##  1st Qu.: 3.000   1st Qu.:0.0000   1st Qu.:4.000   1st Qu.:0.00
##  Median : 7.000   Median :0.0000   Median :5.000   Median :1.00
##  Mean   : 5.774   Mean   :0.6965   Mean   :4.627   Mean   :1.07
##  3rd Qu.: 8.000   3rd Qu.:1.0000   3rd Qu.:6.000   3rd Qu.:2.00
##  Max.   :10.000   Max.   :9.0000   Max.   :9.000   Max.   :5.00
##        V9             V10             V11             V12
##  Min.   :0.000   Min.   :0.000   Min.   :0.0000   Min.   :0.00
##  1st Qu.:2.000   1st Qu.:5.000   1st Qu.:0.0000   1st Qu.:1.00
##  Median :3.000   Median :6.000   Median :1.0000   Median :2.00
##  Mean   :3.259   Mean   :6.183   Mean   :0.8835   Mean   :2.29
##  3rd Qu.:4.000   3rd Qu.:7.000   3rd Qu.:1.0000   3rd Qu.:3.00
##  Max.   :9.000   Max.   :9.000   Max.   :7.0000   Max.   :9.00
##       V13             V14             V15             V16
##  Min.   :0.000   Min.   :0.00   Min.   :0.0   Min.   :0.000
##  1st Qu.:0.000   1st Qu.:2.00   1st Qu.:3.0   1st Qu.:0.000
##  Median :2.000   Median :3.00   Median :4.0   Median :1.000
##  Mean   :1.888   Mean   :3.23   Mean   :4.3   Mean   :1.461
##  3rd Qu.:3.000   3rd Qu.:4.00   3rd Qu.:6.0   3rd Qu.:2.000
```

```
##  Max.   :9.000   Max.   :9.00   Max.   :9.0   Max.   :9.000
##       V17             V18             V19             V20
##  Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:2.000   1st Qu.:3.000   1st Qu.:0.000   1st Qu.:0.000
##  Median :3.000   Median :5.000   Median :2.000   Median :0.000
##  Mean   :3.351   Mean   :4.572   Mean   :1.895   Mean   :0.398
##  3rd Qu.:4.000   3rd Qu.:6.000   3rd Qu.:3.000   3rd Qu.:1.000
##  Max.   :9.000   Max.   :9.000   Max.   :9.000   Max.   :5.000
##       V21             V22             V23             V24
##  Min.   :0.0000   Min.   :0.000   Min.   :0.00   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:2.000   1st Qu.:1.00   1st Qu.:1.000
##  Median :0.0000   Median :3.000   Median :2.00   Median :2.000
##  Mean   :0.5223   Mean   :2.899   Mean   :2.22   Mean   :2.306
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:3.00   3rd Qu.:3.000
##  Max.   :9.0000   Max.   :9.000   Max.   :9.00   Max.   :9.000
##       V25             V26             V27             V28
##  Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:2.000
##  Median :1.000   Median :2.000   Median :2.000   Median :4.000
##  Mean   :1.621   Mean   :1.607   Mean   :2.203   Mean   :3.759
##  3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:3.000   3rd Qu.:5.000
##  Max.   :9.000   Max.   :9.000   Max.   :9.000   Max.   :9.000
##       V29             V30             V31             V32
##  Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :0.00
##  1st Qu.:0.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:5.00
##  Median :1.000   Median :4.000   Median :5.000   Median :6.00
##  Mean   :1.067   Mean   :4.237   Mean   :4.772   Mean   :6.04
##  3rd Qu.:2.000   3rd Qu.:7.000   3rd Qu.:7.000   3rd Qu.:7.00
##  Max.   :9.000   Max.   :9.000   Max.   :9.000   Max.   :9.00
##       V33             V34             V35             V36
##  Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.000   1st Qu.:1.000   1st Qu.:5.000   1st Qu.:1.000
##  Median :1.000   Median :2.000   Median :7.000   Median :2.000
##  Mean   :1.316   Mean   :1.959   Mean   :6.277   Mean   :2.729
##  3rd Qu.:2.000   3rd Qu.:3.000   3rd Qu.:8.000   3rd Qu.:4.000
##  Max.   :7.000   Max.   :9.000   Max.   :9.000   Max.   :9.000
##       V37             V38             V39             V40
##  Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:1.000   1st Qu.:2.000   1st Qu.:1.000   1st Qu.:0.0000
##  Median :2.000   Median :4.000   Median :3.000   Median :0.0000
##  Mean   :2.574   Mean   :3.536   Mean   :2.731   Mean   :0.7961
##  3rd Qu.:4.000   3rd Qu.:5.000   3rd Qu.:4.000   3rd Qu.:1.0000
##  Max.   :9.000   Max.   :9.000   Max.   :9.000   Max.   :9.0000
##       V41             V42             V43             V44
##  Min.   :0.0000   Min.   :0.000   Min.   :1.000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:3.000   1st Qu.:0.0000
##  Median :0.0000   Median :4.000   Median :4.000   Median :0.0000
##  Mean   :0.2027   Mean   :3.784   Mean   :4.236   Mean   :0.7712
##  3rd Qu.:0.0000   3rd Qu.:4.000   3rd Qu.:6.000   3rd Qu.:2.0000
##  Max.   :9.0000   Max.   :9.000   Max.   :8.000   Max.   :3.0000
##       V45             V46             V47             V48
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :5.00   Median :0.00000
```

```
##  Mean   :0.04002   Mean   :0.07162   Mean   :2.97   Mean   :0.04827
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:6.00   3rd Qu.:0.00000
##  Max.   :6.00000   Max.   :4.00000   Max.   :8.00   Max.   :7.00000
##       V49              V50                V51               V52
##  Min.   :0.0000   Min.   :0.000000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.000000   Median :0.00000   Median :0.00000
##  Mean   :0.1754   Mean   :0.009447   Mean   :0.02096   Mean   :0.09258
##  3rd Qu.:0.0000   3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :7.0000   Max.   :9.000000   Max.   :5.00000   Max.   :6.00000
##       V53               V54             V55               V56
##  Min.   :0.00000   Min.   :0.000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.000   Median :0.0000   Median :0.00000
##  Mean   :0.01305   Mean   :0.215   Mean   :0.1948   Mean   :0.01374
##  3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :6.00000   Max.   :6.000   Max.   :9.0000   Max.   :6.00000
##       V57               V58               V59             V60
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.000   Min.   :0.0000000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.0000000
##  Median :0.00000   Median :0.00000   Median :2.000   Median :0.0000000
##  Mean   :0.01529   Mean   :0.02353   Mean   :1.828   Mean   :0.0008588
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:4.000   3rd Qu.:0.0000000
##  Max.   :3.00000   Max.   :7.00000   Max.   :8.000   Max.   :3.0000000
##       V61               V62               V63               V64
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01889   Mean   :0.02525   Mean   :0.01563   Mean   :0.04758
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :6.00000   Max.   :1.00000   Max.   :6.00000   Max.   :5.00000
##       V65              V66               V67               V68
##  Min.   :0.000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.000   Median :0.00000   Median :0.00000   Median :1.0000
##  Mean   :0.403   Mean   :0.01477   Mean   :0.02061   Mean   :0.5622
##  3rd Qu.:1.000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:1.0000
##  Max.   :2.000   Max.   :5.00000   Max.   :1.00000   Max.   :7.0000
##       V69               V70               V71                V72
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.000000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.000000   Median :0.00000
##  Mean   :0.01048   Mean   :0.04105   Mean   :0.002233   Mean   :0.01254
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.00000
##  Max.   :4.00000   Max.   :8.00000   Max.   :3.000000   Max.   :3.00000
##       V73               V74                V75               V76
##  Min.   :0.00000   Min.   :0.000000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.000000   Median :0.00000   Median :0.00000
##  Mean   :0.03367   Mean   :0.006183   Mean   :0.07042   Mean   :0.07661
##  3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :4.00000   Max.   :6.000000   Max.   :2.00000   Max.   :8.00000
##       V77               V78               V79                V80
##  Min.   :0.000000   Min.   :0.000000   Min.   :0.000000   Min.   :0.0000
```

```
##  1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.0000
##  Median :0.000000   Median :0.000000   Median :0.000000   Median :1.0000
##  Mean   :0.005325   Mean   :0.006527   Mean   :0.004638   Mean   :0.5701
##  3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:1.0000
##  Max.   :1.000000   Max.   :1.000000   Max.   :2.000000   Max.   :7.0000
##       V81                V82                V83
##  Min.   :0.0000000   Min.   :0.000000   Min.   :0.00000
##  1st Qu.:0.0000000   1st Qu.:0.000000   1st Qu.:0.00000
##  Median :0.0000000   Median :0.000000   Median :0.00000
##  Mean   :0.0005153   Mean   :0.006012   Mean   :0.03178
##  3rd Qu.:0.0000000   3rd Qu.:0.000000   3rd Qu.:0.00000
##  Max.   :1.0000000   Max.   :2.000000   Max.   :3.00000
##       V84                V85                V86
##  Min.   :0.000000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.000000   Median :0.00000   Median :0.00000
##  Mean   :0.007901   Mean   :0.01426   Mean   :0.05977
##  3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :2.000000   Max.   :2.00000   Max.   :1.00000
```

```
str(dataset1)
```

```
## 'data.frame':    5822 obs. of  86 variables:
##  $ V1 : int  33 37 37 9 40 23 39 33 33 11 ...
##  $ V2 : int  1 1 1 1 1 1 2 1 1 2 ...
##  $ V3 : int  3 2 2 3 4 2 3 2 2 3 ...
##  $ V4 : int  2 2 2 3 2 1 2 3 4 3 ...
##  $ V5 : int  8 8 8 3 10 5 9 8 8 3 ...
##  $ V6 : int  0 1 0 2 1 0 2 0 0 3 ...
##  $ V7 : int  5 4 4 3 4 5 2 7 1 5 ...
##  $ V8 : int  1 1 2 2 1 0 0 0 3 0 ...
##  $ V9 : int  3 4 4 4 4 5 5 2 6 2 ...
##  $ V10: int  7 6 3 5 7 0 7 7 6 7 ...
##  $ V11: int  0 2 2 2 1 6 2 2 0 0 ...
##  $ V12: int  2 2 4 2 2 3 0 0 3 2 ...
##  $ V13: int  1 0 4 2 2 3 0 0 3 2 ...
##  $ V14: int  2 4 4 3 4 5 3 5 3 2 ...
##  $ V15: int  6 5 2 4 4 2 6 4 3 6 ...
##  $ V16: int  1 0 0 3 5 0 0 0 0 0 ...
##  $ V17: int  2 5 5 4 4 5 4 3 1 4 ...
##  $ V18: int  7 4 4 2 0 4 5 6 8 5 ...
##  $ V19: int  1 0 0 4 0 2 0 2 1 2 ...
##  $ V20: int  0 0 0 0 5 0 0 0 1 0 ...
##  $ V21: int  1 0 0 0 4 0 0 0 0 0 ...
##  $ V22: int  2 5 7 3 0 4 4 2 1 3 ...
##  $ V23: int  5 0 0 1 0 2 1 5 8 3 ...
##  $ V24: int  2 4 2 2 0 2 5 2 1 3 ...
##  $ V25: int  1 0 0 3 9 2 0 2 1 1 ...
##  $ V26: int  1 2 5 2 0 2 1 1 1 2 ...
##  $ V27: int  2 3 0 1 0 2 4 2 0 1 ...
##  $ V28: int  6 5 4 4 0 4 5 5 8 4 ...
##  $ V29: int  1 0 0 0 0 2 0 2 1 2 ...
##  $ V30: int  1 2 7 5 4 9 6 0 9 0 ...
##  $ V31: int  8 7 2 4 5 0 3 9 0 9 ...
##  $ V32: int  8 7 7 9 6 5 8 4 5 6 ...
```

```
##  $ V33: int  0 1 0 0 2 3 0 4 2 1 ...
##  $ V34: int  1 2 2 0 1 3 1 2 3 2 ...
##  $ V35: int  8 6 9 7 5 9 9 6 7 6 ...
##  $ V36: int  1 3 0 2 4 0 0 3 2 3 ...
##  $ V37: int  0 2 4 1 0 5 4 2 7 2 ...
##  $ V38: int  4 0 5 5 0 2 3 5 2 3 ...
##  $ V39: int  5 5 0 3 9 3 3 3 1 3 ...
##  $ V40: int  0 2 0 0 0 0 0 0 0 1 ...
##  $ V41: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V42: int  4 5 3 4 6 3 3 3 2 4 ...
##  $ V43: int  3 4 4 4 3 3 5 3 3 7 ...
##  $ V44: int  0 2 2 0 0 0 0 0 0 2 ...
##  $ V45: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V46: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V47: int  6 0 6 6 0 6 6 0 5 0 ...
##  $ V48: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V49: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V50: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V51: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V52: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V53: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V54: int  0 0 0 0 0 0 0 3 0 0 ...
##  $ V55: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V56: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V57: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V58: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V59: int  5 2 2 2 6 0 0 0 0 3 ...
##  $ V60: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V61: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V62: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V63: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V64: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V65: int  0 2 1 0 0 0 0 0 0 1 ...
##  $ V66: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V67: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V68: int  1 0 1 1 0 1 1 0 1 0 ...
##  $ V69: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V70: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V71: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V72: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V73: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V74: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V75: int  0 0 0 0 0 0 0 1 0 0 ...
##  $ V76: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V77: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V78: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V79: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V80: int  1 1 1 1 1 0 0 0 0 1 ...
##  $ V81: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V82: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V83: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V84: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V85: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V86: int  0 0 0 0 0 0 0 0 0 0 ...
```

```
d1 <- cor(dataset1[1:85],dataset1[86], method="spearman")
```

An initial review of the data revealed that for the religion parameters those customers who are Roman Catholic are further subdivided into 10 different levels with a range of percentage values (see Table Main Table and Sub Table L3 in Appendix A highlighted in grey). Despite various attempts to get more information about what such percentage value actually represents for a parameter that normally would have a simple binary value—that is, either someone is a Roman Catholic or not—we were unable to come up with a credible explanation. Given this lack of clarity, it was decided to exclude all four religion-related parameters (i.e., MGODRK—Roman Catholic; MGODPR—Protestant; MGODOV—Other religion; MGODGE—No religion) from further analysis presented here. The correlations between customers' religion and holding of Caravan Insurance is also not very significant. Hence, on the basis of this all four religion parameters have been excluded from further analysis presented here.

```
cname <- c('mostype', 'maanthui', 'mgemomv', 'mgemleef', 'moshoofd', 'mgodrk', 'mgodpr', 'mgodov', 'mgo

colnames(dataset1) <- cname
head(dataset1)
```

```
##   mostype maanthui mgemomv mgemleef moshoofd mgodrk mgodpr mgodov mgodge
## 1      33        1       3        2        8      0      5      1      3
## 2      37        1       2        2        8      1      4      1      4
## 3      37        1       2        2        8      0      4      2      4
## 4       9        1       3        3        3      2      3      2      4
## 5      40        1       4        2       10      1      4      1      4
## 6      23        1       2        1        5      0      5      0      5
##   mrelge mrelsa mrelov mfalleen mfgekind mfwekind moplhoog moplmidd
## 1      7      0      2        1        2        6        1        2
## 2      6      2      2        0        4        5        0        5
## 3      3      2      4        4        4        2        0        5
## 4      5      2      2        2        3        4        3        4
## 5      7      1      2        2        4        4        5        4
## 6      0      6      3        3        5        2        0        5
##   mopllaag mberhoog mberzelf mberboer mbermidd mberarbg mberarbo mska
## 1        7        1        0        1        2        5        2    1
## 2        4        0        0        0        5        0        4    0
## 3        4        0        0        0        7        0        2    0
## 4        2        4        0        0        3        1        2    3
## 5        0        0        5        4        0        0        0    9
## 6        4        2        0        0        4        2        2    2
##   mskb1 mskb2 mskc mskd mhhuur mhkoop maut1 maut2 maut0 mzfonds mzpart
## 1     1     2    6    1      1      8     8     0     1       8      1
## 2     2     3    5    0      2      7     7     1     2       6      3
## 3     5     0    4    0      7      2     7     0     2       9      0
## 4     2     1    4    0      5      4     9     0     0       7      2
## 5     0     0    0    0      4      5     6     2     1       5      4
## 6     2     2    4    2      9      0     5     3     3       9      0
##   minkm30 mink3045 mink4575 mink7512 mink123m minkgem mkoopkla pwapart
## 1       0        4        5        0        0       4        3       0
## 2       2        0        5        2        0       5        4       2
## 3       4        5        0        0        0       3        4       2
## 4       1        5        3        0        0       4        4       0
## 5       0        0        9        0        0       6        3       0
## 6       5        2        3        0        0       3        3       0
##   pwabedr pwaland ppersaut pbesaut pmotsco pvraaut paanhang ptractor
```

```
## 1       0       0        6       0       0       0        0        0
## 2       0       0        0       0       0       0        0        0
## 3       0       0        6       0       0       0        0        0
## 4       0       0        6       0       0       0        0        0
## 5       0       0        0       0       0       0        0        0
## 6       0       0        6       0       0       0        0        0
##   pwerkt pbrom pleven ppersong pgezong pwaoreg pbrand pzeilpl pplezier
## 1      0     0      0        0       0       0      5       0        0
## 2      0     0      0        0       0       0      2       0        0
## 3      0     0      0        0       0       0      2       0        0
## 4      0     0      0        0       0       0      2       0        0
## 5      0     0      0        0       0       0      6       0        0
## 6      0     0      0        0       0       0      0       0        0
##   pfiets pinboed pbystand awapart awabedr awaland apersaut abesaut amotsco
## 1      0       0        0       0       0       0        1       0       0
## 2      0       0        0       2       0       0        0       0       0
## 3      0       0        0       1       0       0        1       0       0
## 4      0       0        0       0       0       0        1       0       0
## 5      0       0        0       0       0       0        0       0       0
## 6      0       0        0       0       0       0        1       0       0
##   avraaut aaanhang atractor awerkt abrom aleven apersong agezong awaoreg
## 1       0        0        0      0     0      0        0       0       0
## 2       0        0        0      0     0      0        0       0       0
## 3       0        0        0      0     0      0        0       0       0
## 4       0        0        0      0     0      0        0       0       0
## 5       0        0        0      0     0      0        0       0       0
## 6       0        0        0      0     0      0        0       0       0
##   abrand azeilpl aplezier afiets ainboed abystand caravan
## 1      1       0        0      0       0        0       0
## 2      1       0        0      0       0        0       0
## 3      1       0        0      0       0        0       0
## 4      1       0        0      0       0        0       0
## 5      1       0        0      0       0        0       0
## 6      0       0        0      0       0        0       0
```

```
dataset2 <- dataset1[-c(6,7,8,9)]
dim(dataset2)
```

```
## [1] 5822   82
```

The remaining 4 categorical variables are : mostype(41), mgemleef(6), moshoofd(10), pwapart(10)

```
sqldf("select mostype, mgemleef, moshoofd, pwapart from dataset1 limit 10")
```

```
##    mostype mgemleef moshoofd pwapart
## 1       33        2        8       0
## 2       37        2        8       2
## 3       37        2        8       2
## 4        9        3        3       0
## 5       40        2       10       0
## 6       23        1        5       0
## 7       39        2        9       0
## 8       33        3        8       0
## 9       33        4        8       0
## 10      11        3        3       2
```

```
dataset2 <- dataset1
### process categorical variable 1 - mostype
mostype_list <- sort(unique(dataset2$mostype))
for (i in mostype_list){
  dataset2[paste("mostype_", i, sep="")] <- ifelse(dataset2$mostype==i, 1, 0)
}
dataset2$mostype <- NULL
print(length(colnames(dataset2)))
```

```
## [1] 125
```

```
sqldf("select mostype_1, mostype_2, mostype_3 from dataset2 limit 10 ")
```

```
##    mostype_1 mostype_2 mostype_3
## 1          0         0         0
## 2          0         0         0
## 3          0         0         0
## 4          0         0         0
## 5          0         0         0
## 6          0         0         0
## 7          0         0         0
## 8          0         0         0
## 9          0         0         0
## 10         0         0         0
```

```
### process categorical variable 2 - mgemleef
mgemleef_list <- sort(unique(dataset2$mgemleef))
for (i in mgemleef_list){
  dataset2[paste("mgemleef_", i, sep="")] <- ifelse(dataset2$mgemleef==i, 1, 0)
}
dataset2$mgemleef <- NULL
print(length(colnames(dataset2)))
```

```
## [1] 130
```

```
sqldf("select mgemleef_1, mgemleef_2, mgemleef_3 from dataset2 limit 10 ")
```

```
##    mgemleef_1 mgemleef_2 mgemleef_3
## 1           0          1          0
## 2           0          1          0
## 3           0          1          0
## 4           0          0          1
## 5           0          1          0
## 6           1          0          0
## 7           0          1          0
## 8           0          0          1
## 9           0          0          0
## 10          0          0          1
```

```
colnames(dataset2)
```

```
##   [1] "maanthui"   "mgemomv"    "moshoofd"   "mgodrk"     "mgodpr"
##   [6] "mgodov"     "mgodge"     "mrelge"     "mrelsa"     "mrelov"
##  [11] "mfalleen"   "mfgekind"   "mfwekind"   "moplhoog"   "moplmidd"
##  [16] "mopllaag"   "mberhoog"   "mberzelf"   "mberboer"   "mbermidd"
##  [21] "mberarbg"   "mberarbo"   "mska"       "mskb1"      "mskb2"
##  [26] "mskc"       "mskd"       "mhhuur"     "mhkoop"     "maut1"
```

```
##  [31] "maut2"      "maut0"      "mzfonds"    "mzpart"     "minkm30"
##  [36] "mink3045"   "mink4575"   "mink7512"   "mink123m"   "minkgem"
##  [41] "mkoopkla"   "pwapart"    "pwabedr"    "pwaland"    "ppersaut"
##  [46] "pbesaut"    "pmotsco"    "pvraaut"    "paanhang"   "ptractor"
##  [51] "pwerkt"     "pbrom"      "pleven"     "ppersong"   "pgezong"
##  [56] "pwaoreg"    "pbrand"     "pzeilpl"    "pplezier"   "pfiets"
##  [61] "pinboed"    "pbystand"   "awapart"    "awabedr"    "awaland"
##  [66] "apersaut"   "abesaut"    "amotsco"    "avraaut"    "aaanhang"
##  [71] "atractor"   "awerkt"     "abrom"      "aleven"     "apersong"
##  [76] "agezong"    "awaoreg"    "abrand"     "azeilpl"    "aplezier"
##  [81] "afiets"     "ainboed"    "abystand"   "caravan"    "mostype_1"
##  [86] "mostype_2"  "mostype_3"  "mostype_4"  "mostype_5"  "mostype_6"
##  [91] "mostype_7"  "mostype_8"  "mostype_9"  "mostype_10" "mostype_11"
##  [96] "mostype_12" "mostype_13" "mostype_15" "mostype_16" "mostype_17"
## [101] "mostype_18" "mostype_19" "mostype_20" "mostype_21" "mostype_22"
## [106] "mostype_23" "mostype_24" "mostype_25" "mostype_26" "mostype_27"
## [111] "mostype_28" "mostype_29" "mostype_30" "mostype_31" "mostype_32"
## [116] "mostype_33" "mostype_34" "mostype_35" "mostype_36" "mostype_37"
## [121] "mostype_38" "mostype_39" "mostype_40" "mostype_41" "mgemleef_1"
## [126] "mgemleef_2" "mgemleef_3" "mgemleef_4" "mgemleef_5" "mgemleef_6"
```

```
### process categorical variable 3 - moshoofd
moshoofd_list <- sort(unique(dataset2$moshoofd))
for (i in moshoofd_list){
  dataset2[paste("moshoofd_", i, sep="")] <- ifelse(dataset2$moshoofd==i, 1, 0)
}
dataset2$moshoofd <- NULL
print(length(colnames(dataset2)))
```

```
## [1] 139
```

```
sqldf("select moshoofd_1, moshoofd_2, moshoofd_3 from dataset2 limit 10 ")
```

```
##    moshoofd_1 moshoofd_2 moshoofd_3
## 1           0          0          0
## 2           0          0          0
## 3           0          0          0
## 4           0          0          1
## 5           0          0          0
## 6           0          0          0
## 7           0          0          0
## 8           0          0          0
## 9           0          0          0
## 10          0          0          1
```

```
colnames(dataset2)
```

```
##   [1] "maanthui"    "mgemomv"     "mgodrk"      "mgodpr"      "mgodov"
##   [6] "mgodge"      "mrelge"      "mrelsa"      "mrelov"      "mfalleen"
##  [11] "mfgekind"    "mfwekind"    "moplhoog"    "moplmidd"    "mopllaag"
##  [16] "mberhoog"    "mberzelf"    "mberboer"    "mbermidd"    "mberarbg"
##  [21] "mberarbo"    "mska"        "mskb1"       "mskb2"       "mskc"
##  [26] "mskd"        "mhhuur"      "mhkoop"      "maut1"       "maut2"
##  [31] "maut0"       "mzfonds"     "mzpart"      "minkm30"     "mink3045"
##  [36] "mink4575"    "mink7512"    "mink123m"    "minkgem"     "mkoopkla"
##  [41] "pwapart"     "pwabedr"     "pwaland"     "ppersaut"    "pbesaut"
```

```
##  [46] "pmotsco"     "pvraaut"     "paanhang"    "ptractor"    "pwerkt"
##  [51] "pbrom"       "pleven"      "ppersong"    "pgezong"     "pwaoreg"
##  [56] "pbrand"      "pzeilpl"     "pplezier"    "pfiets"      "pinboed"
##  [61] "pbystand"    "awapart"     "awabedr"     "awaland"     "apersaut"
##  [66] "abesaut"     "amotsco"     "avraaut"     "aaanhang"    "atractor"
##  [71] "awerkt"      "abrom"       "aleven"      "apersong"    "agezong"
##  [76] "awaoreg"     "abrand"      "azeilpl"     "aplezier"    "afiets"
##  [81] "ainboed"     "abystand"    "caravan"     "mostype_1"   "mostype_2"
##  [86] "mostype_3"   "mostype_4"   "mostype_5"   "mostype_6"   "mostype_7"
##  [91] "mostype_8"   "mostype_9"   "mostype_10"  "mostype_11"  "mostype_12"
##  [96] "mostype_13"  "mostype_15"  "mostype_16"  "mostype_17"  "mostype_18"
## [101] "mostype_19"  "mostype_20"  "mostype_21"  "mostype_22"  "mostype_23"
## [106] "mostype_24"  "mostype_25"  "mostype_26"  "mostype_27"  "mostype_28"
## [111] "mostype_29"  "mostype_30"  "mostype_31"  "mostype_32"  "mostype_33"
## [116] "mostype_34"  "mostype_35"  "mostype_36"  "mostype_37"  "mostype_38"
## [121] "mostype_39"  "mostype_40"  "mostype_41"  "mgemleef_1"  "mgemleef_2"
## [126] "mgemleef_3"  "mgemleef_4"  "mgemleef_5"  "mgemleef_6"  "moshoofd_1"
## [131] "moshoofd_2"  "moshoofd_3"  "moshoofd_4"  "moshoofd_5"  "moshoofd_6"
## [136] "moshoofd_7"  "moshoofd_8"  "moshoofd_9"  "moshoofd_10"
```

```
### process categorical variable 4 - pwapart
pwapart_list <- sort(unique(dataset2$pwapart))
for (i in pwapart_list){
  dataset2[paste("pwapart_", i, sep="")] <- ifelse(dataset2$pwapart==i, 1, 0)
}
dataset2$pwapart <- NULL
print(length(colnames(dataset2)))
```

```
## [1] 142
```

```
sqldf("select pwapart_1, pwapart_2, pwapart_3 from dataset2 limit 10 ")
```

```
##    pwapart_1 pwapart_2 pwapart_3
## 1          0         0         0
## 2          0         1         0
## 3          0         1         0
## 4          0         0         0
## 5          0         0         0
## 6          0         0         0
## 7          0         0         0
## 8          0         0         0
## 9          0         0         0
## 10         0         1         0
```

```
colnames(dataset2)
```

```
##   [1] "maanthui"    "mgemomv"     "mgodrk"      "mgodpr"      "mgodov"
##   [6] "mgodge"      "mrelge"      "mrelsa"      "mrelov"      "mfalleen"
##  [11] "mfgekind"    "mfwekind"    "moplhoog"    "moplmidd"    "mopllaag"
##  [16] "mberhoog"    "mberzelf"    "mberboer"    "mbermidd"    "mberarbg"
##  [21] "mberarbo"    "mska"        "mskb1"       "mskb2"       "mskc"
##  [26] "mskd"        "mhhuur"      "mhkoop"      "maut1"       "maut2"
##  [31] "maut0"       "mzfonds"     "mzpart"      "minkm30"     "mink3045"
##  [36] "mink4575"    "mink7512"    "mink123m"    "minkgem"     "mkoopkla"
##  [41] "pwabedr"     "pwaland"     "ppersaut"    "pbesaut"     "pmotsco"
##  [46] "pvraaut"     "paanhang"    "ptractor"    "pwerkt"      "pbrom"
```

```
##  [51] "pleven"      "ppersong"    "pgezong"     "pwaoreg"     "pbrand"
##  [56] "pzeilpl"     "pplezier"    "pfiets"      "pinboed"     "pbystand"
##  [61] "awapart"     "awabedr"     "awaland"     "apersaut"    "abesaut"
##  [66] "amotsco"     "avraaut"     "aaanhang"    "atractor"    "awerkt"
##  [71] "abrom"       "aleven"      "apersong"    "agezong"     "awaoreg"
##  [76] "abrand"      "azeilpl"     "aplezier"    "afiets"      "ainboed"
##  [81] "abystand"    "caravan"     "mostype_1"   "mostype_2"   "mostype_3"
##  [86] "mostype_4"   "mostype_5"   "mostype_6"   "mostype_7"   "mostype_8"
##  [91] "mostype_9"   "mostype_10"  "mostype_11"  "mostype_12"  "mostype_13"
##  [96] "mostype_15"  "mostype_16"  "mostype_17"  "mostype_18"  "mostype_19"
## [101] "mostype_20"  "mostype_21"  "mostype_22"  "mostype_23"  "mostype_24"
## [106] "mostype_25"  "mostype_26"  "mostype_27"  "mostype_28"  "mostype_29"
## [111] "mostype_30"  "mostype_31"  "mostype_32"  "mostype_33"  "mostype_34"
## [116] "mostype_35"  "mostype_36"  "mostype_37"  "mostype_38"  "mostype_39"
## [121] "mostype_40"  "mostype_41"  "mgemleef_1"  "mgemleef_2"  "mgemleef_3"
## [126] "mgemleef_4"  "mgemleef_5"  "mgemleef_6"  "moshoofd_1"  "moshoofd_2"
## [131] "moshoofd_3"  "moshoofd_4"  "moshoofd_5"  "moshoofd_6"  "moshoofd_7"
## [136] "moshoofd_8"  "moshoofd_9"  "moshoofd_10" "pwapart_0"   "pwapart_1"
## [141] "pwapart_2"   "pwapart_3"
```

```
# Create multiple column format from MOSTYPE column
# mostype_matrix <- model.matrix(~MOSTYPE, dataset3)
# rm(dataset3)
dataset2$pwapart_4 <- 0
dataset2$pwapart_5 <- 0
dataset2$pwapart_6 <- 0

dim(dataset2)
```

```
## [1] 5822  145
```

Variable correlation and logistic regression analysis

```
# install.packages("Hmisc")
library(Hmisc)
```

```
## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, round.POSIXt, trunc.POSIXt, units
```

```
describe(dataset2)
```

```
## dataset2
##
##  145  Variables      5822  Observations
## ---------------------------------------------------------------------------
```

```
## maanthui
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        9    0.259    1.111   0.2029
##
## Value          1     2     3     4     5     6     7     8    10
## Frequency   5267   505    39     2     1     1     5     1     1
## Proportion 0.905 0.087 0.007 0.000 0.000 0.000 0.001 0.000 0.000
## ---------------------------------------------------------------------------
## mgemomv
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        5    0.855    2.679   0.8288
##
## Value          1     2     3     4     5
## Frequency    284  2131  2646   693    68
## Proportion 0.049 0.366 0.454 0.119 0.012
## ---------------------------------------------------------------------------
## mgodrk
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.807   0.6965   0.9328        0        0
##      .25      .50      .75      .90      .95
##        0        0        1        2        2
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency   3228  1599   733   152    66    18    13     6     3     4
## Proportion 0.554 0.275 0.126 0.026 0.011 0.003 0.002 0.001 0.001 0.001
## ---------------------------------------------------------------------------
## mgodpr
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.958    4.627    1.869        2        2
##      .25      .50      .75      .90      .95
##        4        5        6        7        7
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency     78   134   396   590  1607  1501   714   564    65   173
## Proportion 0.013 0.023 0.068 0.101 0.276 0.258 0.123 0.097 0.011 0.030
## ---------------------------------------------------------------------------
## mgodov
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        6    0.904     1.07    1.075
##
## Value          0     1     2     3     4     5
## Frequency   2003  2014  1388   257   132    28
## Proportion 0.344 0.346 0.238 0.044 0.023 0.005
## ---------------------------------------------------------------------------
## mgodge
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.961    3.259    1.766        0        1
##      .25      .50      .75      .90      .95
##        2        3        4        5        6
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency    456   230  1055  1453  1334   963   217   101     5     8
## Proportion 0.078 0.040 0.181 0.250 0.229 0.165 0.037 0.017 0.001 0.001
## ---------------------------------------------------------------------------
```

```
## mrelge
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10     0.96    6.183    2.069        2        4
##      .25      .50      .75      .90      .95
##        5        6        7        9        9
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency     64    75   157   246   324   946  1172  1683   361   794
## Proportion 0.011 0.013 0.027 0.042 0.056 0.162 0.201 0.289 0.062 0.136
## ------------------------------------------------------------------------
## mrelsa
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        8    0.877   0.8835   0.9839
##
## Value          0     1     2     3     4     5     6     7
## Frequency   2448  2030  1075   159    78    18    13     1
## Proportion 0.420 0.349 0.185 0.027 0.013 0.003 0.002 0.000
## ------------------------------------------------------------------------
## mrelov
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.954     2.29    1.873        0        0
##      .25      .50      .75      .90      .95
##        1        2        3        4        5
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency   1173   539  1756  1152   648   266   179    64    21    24
## Proportion 0.201 0.093 0.302 0.198 0.111 0.046 0.031 0.011 0.004 0.004
## ------------------------------------------------------------------------
## mfalleen
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.954    1.888    1.942        0        0
##      .25      .50      .75      .90      .95
##        0        2        3        4        5
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency   1757   951  1247   848   519   259   127    67    24    23
## Proportion 0.302 0.163 0.214 0.146 0.089 0.044 0.022 0.012 0.004 0.004
## ------------------------------------------------------------------------
## mfgekind
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10     0.96     3.23    1.777        0        1
##      .25      .50      .75      .90      .95
##        2        3        4        5        6
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency    371   372  1060  1498  1455   606   321    96    14    29
## Proportion 0.064 0.064 0.182 0.257 0.250 0.104 0.055 0.016 0.002 0.005
## ------------------------------------------------------------------------
## mfwekind
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.977      4.3    2.244        1        2
##      .25      .50      .75      .90      .95
##        3        4        6        7        8
##
```

```
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency    153   292   635   973  1137  1106   783   351   206   186
## Proportion 0.026 0.050 0.109 0.167 0.195 0.190 0.134 0.060 0.035 0.032
## ---------------------------------------------------------------------------
## moplhoog
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10     0.93    1.461    1.678        0        0
##      .25      .50      .75      .90      .95
##        0        1        2        4        5
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency   2147  1322  1144   547   326   187    67    51    22     9
## Proportion 0.369 0.227 0.196 0.094 0.056 0.032 0.012 0.009 0.004 0.002
## ---------------------------------------------------------------------------
## moplmidd
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.966    3.351    1.943        0        1
##      .25      .50      .75      .90      .95
##        2        3        4        6        6
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency    423   383   937  1330  1426   738   348   157    37    43
## Proportion 0.073 0.066 0.161 0.228 0.245 0.127 0.060 0.027 0.006 0.007
## ---------------------------------------------------------------------------
## mopllaag
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.984    4.572    2.605        0        2
##      .25      .50      .75      .90      .95
##        3        5        6        7        9
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency    299   243   667   680   851  1009   856   640   254   323
## Proportion 0.051 0.042 0.115 0.117 0.146 0.173 0.147 0.110 0.044 0.055
## ---------------------------------------------------------------------------
## mberhoog
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.957    1.895    1.907        0        0
##      .25      .50      .75      .90      .95
##        0        2        3        4        5
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency   1524  1245  1364   756   397   249   138    92    26    31
## Proportion 0.262 0.214 0.234 0.130 0.068 0.043 0.024 0.016 0.004 0.005
## ---------------------------------------------------------------------------
## mberzelf
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        6    0.623    0.398   0.6223
##
## Value          0     1     2     3     4     5
## Frequency   4171  1202   348    37    12    52
## Proportion 0.716 0.206 0.060 0.006 0.002 0.009
## ---------------------------------------------------------------------------
## mberboer
##        n  missing distinct     Info     Mean      Gmd      .05      .10
```

```
##     5822        0       10    0.627   0.5223   0.8404        0        0
##      .25      .50      .75      .90      .95
##        0        0        1        2        3
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency   4176   854   487   143    77    59    14     3     5     4
## Proportion 0.717 0.147 0.084 0.025 0.013 0.010 0.002 0.001 0.001 0.001
## ------------------------------------------------------------------------
## mbermidd
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.963    2.899    1.996        0        0
##      .25      .50      .75      .90      .95
##        2        3        4        5        6
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency    667   403  1491  1394   953   431   211   178    14    80
## Proportion 0.115 0.069 0.256 0.239 0.164 0.074 0.036 0.031 0.002 0.014
## ------------------------------------------------------------------------
## mberarbg
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.965     2.22    1.905        0        0
##      .25      .50      .75      .90      .95
##        1        2        3        4        5
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency   1167   921  1382  1167   604   310   169    68    24    10
## Proportion 0.200 0.158 0.237 0.200 0.104 0.053 0.029 0.012 0.004 0.002
## ------------------------------------------------------------------------
## mberarbo
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.966    2.306    1.862        0        0
##      .25      .50      .75      .90      .95
##        1        2        3        4        5
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency    968   980  1439  1109   772   331   122    66     9    26
## Proportion 0.166 0.168 0.247 0.190 0.133 0.057 0.021 0.011 0.002 0.004
## ------------------------------------------------------------------------
## mska
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.943    1.621    1.745        0        0
##      .25      .50      .75      .90      .95
##        0        1        2        4        5
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency   1738  1569  1198   685   261   127    96    79    13    56
## Proportion 0.299 0.269 0.206 0.118 0.045 0.022 0.016 0.014 0.002 0.010
## ------------------------------------------------------------------------
## mskb1
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10     0.94    1.607     1.41        0        0
##      .25      .50      .75      .90      .95
##        1        2        2        3        4
##
```

```
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency   1353  1480  1783   775   298    78    25     5     8    17
## Proportion 0.232 0.254 0.306 0.133 0.051 0.013 0.004 0.001 0.001 0.003
## ---------------------------------------------------------------------------
## mskb2
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.958    2.203    1.699        0        0
##      .25      .50      .75      .90      .95
##        1        2        3        4        5
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency    990   861  1676  1175   652   357    96     6     7     2
## Proportion 0.170 0.148 0.288 0.202 0.112 0.061 0.016 0.001 0.001 0.000
## ---------------------------------------------------------------------------
## mskc
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.973    3.759    2.147        0        1
##      .25      .50      .75      .90      .95
##        2        4        5        6        7
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency    364   272   870  1090  1159  1168   487   217    71   124
## Proportion 0.063 0.047 0.149 0.187 0.199 0.201 0.084 0.037 0.012 0.021
## ---------------------------------------------------------------------------
## mskd
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        9    0.887    1.067    1.317
##
## Value          0     1     2     3     4     5     6     7     9
## Frequency   2607  1563   852   441   223   100    22    13     1
## Proportion 0.448 0.268 0.146 0.076 0.038 0.017 0.004 0.002 0.000
## ---------------------------------------------------------------------------
## mhhuur
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.987    4.237    3.537        0        0
##      .25      .50      .75      .90      .95
##        2        4        7        9        9
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency    949   428   717   593   517   519   382   425   532   760
## Proportion 0.163 0.074 0.123 0.102 0.089 0.089 0.066 0.073 0.091 0.131
## ---------------------------------------------------------------------------
## mhkoop
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.987    4.772    3.536        0        0
##      .25      .50      .75      .90      .95
##        2        5        7        9        9
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency    760   530   426   382   499   520   604   724   428   949
## Proportion 0.131 0.091 0.073 0.066 0.086 0.089 0.104 0.124 0.074 0.163
## ---------------------------------------------------------------------------
## maut1
##        n  missing distinct     Info     Mean      Gmd      .05      .10
```

```
##     5822        0       10    0.952     6.04    1.687        3        4
##      .25      .50      .75      .90      .95
##        5        6        7        8        9
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency     19    14    58   231   448  1210  1663  1413   261   505
## Proportion 0.003 0.002 0.010 0.040 0.077 0.208 0.286 0.243 0.045 0.087
## ------------------------------------------------------------------------
## maut2
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        8    0.924    1.316    1.294
##
## Value          0     1     2     3     4     5     6     7
## Frequency   1854  1468  1748   385   301    56     9     1
## Proportion 0.318 0.252 0.300 0.066 0.052 0.010 0.002 0.000
## ------------------------------------------------------------------------
## maut0
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.953    1.959    1.743        0        0
##      .25      .50      .75      .90      .95
##        1        2        3        4        5
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency   1450   776  1625  1066   587   174    89    25    13    17
## Proportion 0.249 0.133 0.279 0.183 0.101 0.030 0.015 0.004 0.002 0.003
## ------------------------------------------------------------------------
## mzfonds
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.969    6.277    2.177        2        4
##      .25      .50      .75      .90      .95
##        5        7        8        9        9
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency     55    15   307   177   357   974   875  1511   699   852
## Proportion 0.009 0.003 0.053 0.030 0.061 0.167 0.150 0.260 0.120 0.146
## ------------------------------------------------------------------------
## mzpart
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.969    2.729    2.183        0        0
##      .25      .50      .75      .90      .95
##        1        2        4        5        7
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency    852   699  1511   849   992   364   178   307    15    55
## Proportion 0.146 0.120 0.260 0.146 0.170 0.063 0.031 0.053 0.003 0.009
## ------------------------------------------------------------------------
## minkm30
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.972    2.574    2.321        0        0
##      .25      .50      .75      .90      .95
##        1        2        4        5        6
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency   1304   630  1094  1079   599   568   293   156    48    51
```

```
## Proportion 0.224 0.108 0.188 0.185 0.103 0.098 0.050 0.027 0.008 0.009
## ---------------------------------------------------------------------------
## mink3045
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.971    3.536    2.082        0        1
##      .25      .50      .75      .90      .95
##        2        4        5        6        7
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency    465   268   919  1147  1356   931   406   205    35    90
## Proportion 0.080 0.046 0.158 0.197 0.233 0.160 0.070 0.035 0.006 0.015
## ---------------------------------------------------------------------------
## mink4575
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.972    2.731    2.112        0        0
##      .25      .50      .75      .90      .95
##        1        3        4        5        6
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency    891   657  1165  1215  1034   498   125    93    53    91
## Proportion 0.153 0.113 0.200 0.209 0.178 0.086 0.021 0.016 0.009 0.016
## ---------------------------------------------------------------------------
## mink7512
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.812   0.7961    1.096        0        0
##      .25      .50      .75      .90      .95
##        0        0        1        2        3
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency   3246  1359   736   246   147    71     8     1     4     4
## Proportion 0.558 0.233 0.126 0.042 0.025 0.012 0.001 0.000 0.001 0.001
## ---------------------------------------------------------------------------
## mink123m
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        8    0.402   0.2027   0.3535
##
## Value          0     1     2     3     4     5     7     9
## Frequency   4900   763    96    36    24     1     1     1
## Proportion 0.842 0.131 0.016 0.006 0.004 0.000 0.000 0.000
## ---------------------------------------------------------------------------
## minkgem
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.928    3.784    1.388        2        2
##      .25      .50      .75      .90      .95
##        3        4        4        5        6
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency     25    49   651  1932  1854   733   355   131    70    22
## Proportion 0.004 0.008 0.112 0.332 0.318 0.126 0.061 0.023 0.012 0.004
## ---------------------------------------------------------------------------
## mkoopkla
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        8    0.971    4.236    2.269
##
```

```
## Value          1     2     3     4     5     6     7     8
## Frequency    587   425  1524   902   583   901   474   426
## Proportion 0.101 0.073 0.262 0.155 0.100 0.155 0.081 0.073
## ---------------------------------------------------------------------------
## pwabedr
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        7    0.042  0.04002  0.07918
##
## Value          0     1     2     3     4     5     6
## Frequency   5740     7    30    23    17     1     4
## Proportion 0.986 0.001 0.005 0.004 0.003 0.000 0.001
## ---------------------------------------------------------------------------
## pwaland
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        4    0.061  0.07162   0.1406
##
## Value          0     2     3     4
## Frequency   5702     3    57    60
## Proportion 0.979 0.001 0.010 0.010
## ---------------------------------------------------------------------------
## ppersaut
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        6    0.819     2.97    2.998
##
## Value          0     4     5     6     7     8
## Frequency   2845     1   613  2319    41     3
## Proportion 0.489 0.000 0.105 0.398 0.007 0.001
## ---------------------------------------------------------------------------
## pbesaut
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        4    0.025  0.04827  0.09578
##
## Value          0     5     6     7
## Frequency   5774    10    35     3
## Proportion 0.992 0.002 0.006 0.001
## ---------------------------------------------------------------------------
## pmotsco
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        6     0.11   0.1754   0.3387
##
## Value          0     3     4     5     6     7
## Frequency   5600     3   136    32    49     2
## Proportion 0.962 0.001 0.023 0.005 0.008 0.000
## ---------------------------------------------------------------------------
## pvraaut
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        4    0.005 0.009447  0.01887
##
## Value          0     4     6     9
## Frequency   5813     1     7     1
## Proportion 0.998 0.000 0.001 0.000
## ---------------------------------------------------------------------------
## paanhang
##        n  missing distinct     Info     Mean      Gmd
```

```
##     5822        0        6    0.033  0.02095  0.04154
##
## Value          0     1     2     3     4     5
## Frequency   5757    19    38     6     1     1
## Proportion 0.989 0.003 0.007 0.001 0.000 0.000
## ---------------------------------------------------------------------------
## ptractor
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        5    0.072  0.09258   0.1812
##
## Value          0     3     4     5     6
## Frequency   5679    79    27    28     9
## Proportion 0.975 0.014 0.005 0.005 0.002
## ---------------------------------------------------------------------------
## pwerkt
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        5    0.011  0.01305  0.02604
##
## Value          0     2     3     4     6
## Frequency   5801     4     6     8     3
## Proportion 0.996 0.001 0.001 0.001 0.001
## ---------------------------------------------------------------------------
## pbrom
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        6     0.19    0.215   0.4035
##
## Value          0     2     3     4     5     6
## Frequency   5426    34   282    63    16     1
## Proportion 0.932 0.006 0.048 0.011 0.003 0.000
## ---------------------------------------------------------------------------
## pleven
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     5822        0       10    0.144   0.1948   0.3737        0        0
##      .25      .50      .75      .90      .95
##        0        0        0        0        1
##
## Value          0     1     2     3     4     5     6     7     8     9
## Frequency   5529     9    28    84    94    35    38     3     1     1
## Proportion 0.950 0.002 0.005 0.014 0.016 0.006 0.007 0.001 0.000 0.000
## ---------------------------------------------------------------------------
## ppersong
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        7    0.016  0.01374  0.02737
##
## Value          0     1     2     3     4     5     6
## Frequency   5791     3    18     4     3     1     2
## Proportion 0.995 0.001 0.003 0.001 0.001 0.000 0.000
## ---------------------------------------------------------------------------
## pgezong
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        3    0.019  0.01529   0.0304
##
## Value          0     2     3
## Frequency   5784    25    13
```

```
## Proportion 0.993 0.004 0.002
## ---------------------------------------------------------------------------
## pwaoreg
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        5    0.012  0.02353  0.04689
##
## Value          0     4     5     6     7
## Frequency   5799     1     1    19     2
## Proportion 0.996 0.000 0.000 0.003 0.000
## ---------------------------------------------------------------------------
## pbrand
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        9     0.89    1.828    2.034
##
## Value          0     1     2     3     4     5     6     7     8
## Frequency   2666   161   535   920  1226   149   155     9     1
## Proportion 0.458 0.028 0.092 0.158 0.211 0.026 0.027 0.002 0.000
## ---------------------------------------------------------------------------
## pzeilpl
##         n   missing  distinct      Info      Mean       Gmd
##      5822         0         3     0.002 0.0008588  0.001717
##
## Value          0     1     3
## Frequency   5819     2     1
## Proportion 0.999 0.000 0.000
## ---------------------------------------------------------------------------
## pplezier
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        7    0.017  0.01889  0.03763
##
## Value          0     1     2     3     4     5     6
## Frequency   5789     5     5     5    13     2     3
## Proportion 0.994 0.001 0.001 0.001 0.002 0.000 0.001
## ---------------------------------------------------------------------------
## pfiets
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.074      147  0.02525  0.04923
##
## ---------------------------------------------------------------------------
## pinboed
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        7    0.023  0.01563  0.03109
##
## Value          0     1     2     3     4     5     6
## Frequency   5777    18    16     6     3     1     1
## Proportion 0.992 0.003 0.003 0.001 0.001 0.000 0.000
## ---------------------------------------------------------------------------
## pbystand
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        5    0.042  0.04758  0.09399
##
## Value          0     2     3     4     5
## Frequency   5740    15    22    44     1
## Proportion 0.986 0.003 0.004 0.008 0.000
```

```
## ---------------------------------------------------------------------------
## awapart
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        3    0.722    0.403   0.4829
##
## Value          0     1     2
## Frequency   3482  2334     6
## Proportion 0.598 0.401 0.001
## ---------------------------------------------------------------------------
## awabedr
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        3    0.042  0.01477  0.02915
##
## Value          0     1     5
## Frequency   5740    81     1
## Proportion 0.986 0.014 0.000
## ---------------------------------------------------------------------------
## awaland
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.061      120  0.02061  0.04038
##
## ---------------------------------------------------------------------------
## apersaut
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        7    0.782   0.5622   0.5974
##
## Value          0     1     2     3     4     6     7
## Frequency   2845  2712   246    12     5     1     1
## Proportion 0.489 0.466 0.042 0.002 0.001 0.000 0.000
## ---------------------------------------------------------------------------
## abesaut
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        5    0.025  0.01048  0.02082
##
## Value          0     1     2     3     4
## Frequency   5774    40     4     3     1
## Proportion 0.992 0.007 0.001 0.001 0.000
## ---------------------------------------------------------------------------
## amotsco
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        4     0.11  0.04105   0.0792
##
## Value          0     1     2     8
## Frequency   5600   211    10     1
## Proportion 0.962 0.036 0.002 0.000
## ---------------------------------------------------------------------------
## avraaut
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        4    0.005 0.002233 0.004461
##
## Value          0     1     2     3
## Frequency   5813     6     2     1
## Proportion 0.998 0.001 0.000 0.000
## ---------------------------------------------------------------------------
```

```
## aaanhang
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        4    0.033  0.01254  0.02483
##
## Value          0     1     2     3
## Frequency   5757    59     4     2
## Proportion 0.989 0.010 0.001 0.000
## ------------------------------------------------------------------------
## atractor
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        5    0.072  0.03367  0.06604
##
## Value          0     1     2     3     4
## Frequency   5679   105    29     3     6
## Proportion 0.975 0.018 0.005 0.001 0.001
## ------------------------------------------------------------------------
## awerkt
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        5    0.011 0.006183  0.01234
##
## Value          0     1     2     3     6
## Frequency   5801    12     6     2     1
## Proportion 0.996 0.002 0.001 0.000 0.000
## ------------------------------------------------------------------------
## abrom
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        3     0.19  0.07042   0.1316
##
## Value          0     1     2
## Frequency   5426   382    14
## Proportion 0.932 0.066 0.002
## ------------------------------------------------------------------------
## aleven
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        6    0.143  0.07661   0.1473
##
## Value          0     1     2     3     4     8
## Frequency   5529   173   100    11     8     1
## Proportion 0.950 0.030 0.017 0.002 0.001 0.000
## ------------------------------------------------------------------------
## apersong
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.016       31 0.005325  0.01059
##
## ------------------------------------------------------------------------
## agezong
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.019       38 0.006527  0.01297
##
## ------------------------------------------------------------------------
## awaoreg
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        3    0.012 0.004638 0.009245
##
```

```
## Value          0     1     2
## Frequency   5799    19     4
## Proportion 0.996 0.003 0.001
## ---------------------------------------------------------------------------
## abrand
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        7    0.765   0.5701   0.5514
##
## Value          0     1     2     3     4     5     7
## Frequency   2666  3017   126     7     3     2     1
## Proportion 0.458 0.518 0.022 0.001 0.001 0.000 0.000
## ---------------------------------------------------------------------------
## azeilpl
##         n   missing  distinct      Info       Sum      Mean       Gmd
##      5822         0         2     0.002         3 0.0005153   0.00103
##
## ---------------------------------------------------------------------------
## aplezier
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        3    0.017 0.006012  0.01196
##
## Value          0     1     2
## Frequency   5789    31     2
## Proportion 0.994 0.005 0.000
## ---------------------------------------------------------------------------
## afiets
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        4    0.074  0.03178  0.06221
##
## Value          0     1     2     3
## Frequency   5675   111    34     2
## Proportion 0.975 0.019 0.006 0.000
## ---------------------------------------------------------------------------
## ainboed
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        3    0.023 0.007901  0.01569
##
## Value          0     1     2
## Frequency   5777    44     1
## Proportion 0.992 0.008 0.000
## ---------------------------------------------------------------------------
## abystand
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        3    0.042  0.01426  0.02812
##
## Value          0     1     2
## Frequency   5740    81     1
## Proportion 0.986 0.014 0.000
## ---------------------------------------------------------------------------
## caravan
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.169      348  0.05977   0.1124
##
## ---------------------------------------------------------------------------
```

```
## mostype_1
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.063      124   0.0213   0.0417
##
## ------------------------------------------------------------------------
## mostype_2
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.042       82  0.01408  0.02778
##
## ------------------------------------------------------------------------
## mostype_3
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.123      249  0.04277  0.08189
##
## ------------------------------------------------------------------------
## mostype_4
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.027       52 0.008932  0.01771
##
## ------------------------------------------------------------------------
## mostype_5
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.023       45 0.007729  0.01534
##
## ------------------------------------------------------------------------
## mostype_6
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2     0.06      119  0.02044  0.04005
##
## ------------------------------------------------------------------------
## mostype_7
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.023       44 0.007558    0.015
##
## ------------------------------------------------------------------------
## mostype_8
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.165      339  0.05823   0.1097
##
## ------------------------------------------------------------------------
## mostype_9
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.136      278  0.04775  0.09096
##
## ------------------------------------------------------------------------
## mostype_10
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.083      165  0.02834  0.05508
##
## ------------------------------------------------------------------------
## mostype_11
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.077      153  0.02628  0.05119
##
```

```
## ------------------------------------------------------------------------
## mostype_12
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.056      111  0.01907  0.03741
##
## ------------------------------------------------------------------------
## mostype_13
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.089      179  0.03075  0.05961
##
## ------------------------------------------------------------------------
## mostype_15
##         n   missing  distinct      Info       Sum      Mean       Gmd
##      5822         0         2     0.003         5 0.0008588  0.001716
##
## ------------------------------------------------------------------------
## mostype_16
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.008       16 0.002748 0.005482
##
## ------------------------------------------------------------------------
## mostype_17
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.005        9 0.001546 0.003087
##
## ------------------------------------------------------------------------
## mostype_18
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2     0.01       19 0.003263 0.006507
##
## ------------------------------------------------------------------------
## mostype_19
##         n   missing  distinct      Info       Sum      Mean       Gmd
##      5822         0         2     0.002         3 0.0005153   0.00103
##
## ------------------------------------------------------------------------
## mostype_20
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.013       25 0.004294 0.008553
##
## ------------------------------------------------------------------------
## mostype_21
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.008       15 0.002576  0.00514
##
## ------------------------------------------------------------------------
## mostype_22
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2     0.05       98  0.01683   0.0331
##
## ------------------------------------------------------------------------
## mostype_23
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.124      251  0.04311  0.08252
```

```
##
## ----------------------------------------------------------------------------
## mostype_24
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2     0.09      180  0.03092  0.05993
##
## ----------------------------------------------------------------------------
## mostype_25
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.042       82  0.01408  0.02778
##
## ----------------------------------------------------------------------------
## mostype_26
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.025       48 0.008245  0.01636
##
## ----------------------------------------------------------------------------
## mostype_27
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.026       50 0.008588  0.01703
##
## ----------------------------------------------------------------------------
## mostype_28
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.013       25 0.004294 0.008553
##
## ----------------------------------------------------------------------------
## mostype_29
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.044       86  0.01477  0.02911
##
## ----------------------------------------------------------------------------
## mostype_30
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2     0.06      118  0.02027  0.03972
##
## ----------------------------------------------------------------------------
## mostype_31
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.102      205  0.03521  0.06795
##
## ----------------------------------------------------------------------------
## mostype_32
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.071      141  0.02422  0.04727
##
## ----------------------------------------------------------------------------
## mostype_33
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.359      810   0.1391   0.2396
##
## ----------------------------------------------------------------------------
## mostype_34
##        n  missing distinct     Info      Sum     Mean      Gmd
```

```
##     5822        0        2    0.091      182  0.03126  0.06058
##
## ----------------------------------------------------------------------------
## mostype_35
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.106      214  0.03676  0.07082
##
## ----------------------------------------------------------------------------
## mostype_36
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.111      225  0.03865  0.07432
##
## ----------------------------------------------------------------------------
## mostype_37
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.066      132  0.02267  0.04432
##
## ----------------------------------------------------------------------------
## mostype_38
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.165      339  0.05823   0.1097
##
## ----------------------------------------------------------------------------
## mostype_39
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.159      328  0.05634   0.1063
##
## ----------------------------------------------------------------------------
## mostype_40
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.036       71   0.0122   0.0241
##
## ----------------------------------------------------------------------------
## mostype_41
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.102      205  0.03521  0.06795
##
## ----------------------------------------------------------------------------
## mgemleef_1
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.038       74  0.01271   0.0251
##
## ----------------------------------------------------------------------------
## mgemleef_2
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.562     1452   0.2494   0.3745
##
## ----------------------------------------------------------------------------
## mgemleef_3
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.749     3000   0.5153   0.4996
##
## ----------------------------------------------------------------------------
## mgemleef_4
```

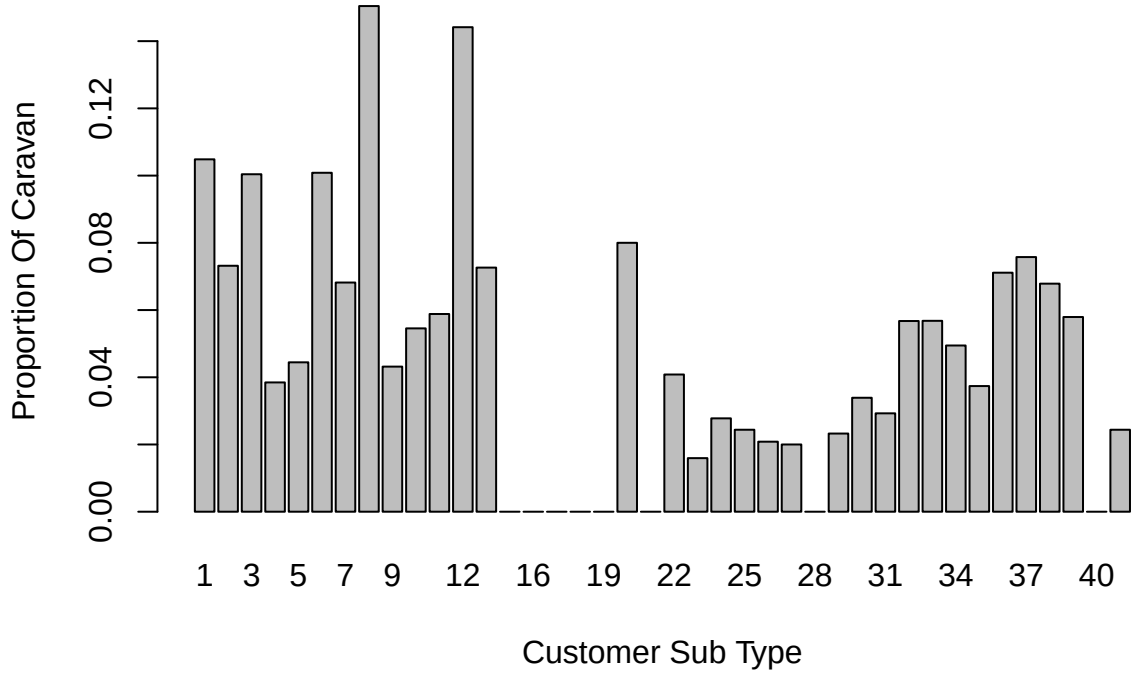
```
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.451     1073   0.1843   0.3007
##
## ------------------------------------------------------------------------
## mgemleef_5
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.096      193  0.03315  0.06411
##
## ------------------------------------------------------------------------
## mgemleef_6
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.015       30 0.005153  0.01025
##
## ------------------------------------------------------------------------
## moshoofd_1
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.257      552  0.09481   0.1717
##
## ------------------------------------------------------------------------
## moshoofd_2
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.236      502  0.08622   0.1576
##
## ------------------------------------------------------------------------
## moshoofd_3
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.387      886   0.1522   0.2581
##
## ------------------------------------------------------------------------
## moshoofd_4
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.027       52 0.008932  0.01771
##
## ------------------------------------------------------------------------
## moshoofd_5
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.265      569  0.09773   0.1764
##
## ------------------------------------------------------------------------
## moshoofd_6
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.102      205  0.03521  0.06795
##
## ------------------------------------------------------------------------
## moshoofd_7
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.257      550  0.09447   0.1711
##
## ------------------------------------------------------------------------
## moshoofd_8
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.589     1563   0.2685   0.3929
##
## ------------------------------------------------------------------------
```

```
## moshoofd_9
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.304      667   0.1146   0.2029
##
## ------------------------------------------------------------------------
## moshoofd_10
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.135      276  0.04741  0.09033
##
## ------------------------------------------------------------------------
## pwapart_0
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.721     3482   0.5981   0.4808
##
## ------------------------------------------------------------------------
## pwapart_1
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2     0.1      201  0.03452  0.06668
##
## ------------------------------------------------------------------------
## pwapart_2
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.696     2128   0.3655   0.4639
##
## ------------------------------------------------------------------------
## pwapart_3
##        n  missing distinct     Info      Sum     Mean      Gmd
##     5822        0        2    0.006       11 0.001889 0.003772
##
## ------------------------------------------------------------------------
## pwapart_4
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        1        0        0        0
##
## Value         0
## Frequency  5822
## Proportion    1
## ------------------------------------------------------------------------
## pwapart_5
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        1        0        0        0
##
## Value         0
## Frequency  5822
## Proportion    1
## ------------------------------------------------------------------------
## pwapart_6
##        n  missing distinct     Info     Mean      Gmd
##     5822        0        1        0        0        0
##
## Value         0
## Frequency  5822
## Proportion    1
## ------------------------------------------------------------------------
```

Step 5

To see if there is any strong relationship between customer parameters and caravan. The results did not elicit any strong link between Caravan Insurance holding and any other specific variable as can be seen from the plot below. Though the variables in this list are of some interest, the correlation between each pair is unlikely to be statistically significant. Overall, the results also confirmed the relatively low number of Caravan Insurance Policy holders in the database

```
prop1 <- prop.table(table(dataset1$mostype, dataset1$caravan), 1)
barplot(prop1[,2], xlab = "Customer Sub Type", ylab = "Proportion Of Caravan")
```




```
library(caret)
```

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:survival':
##
##     cluster
```

```
colnames(dataset2)
```

```
##   [1] "maanthui"    "mgemomv"     "mgodrk"      "mgodpr"      "mgodov"
##   [6] "mgodge"      "mrelge"      "mrelsa"      "mrelov"      "mfalleen"
##  [11] "mfgekind"    "mfwekind"    "moplhoog"    "moplmidd"    "mopllaag"
##  [16] "mberhoog"    "mberzelf"    "mberboer"    "mbermidd"    "mberarbg"
##  [21] "mberarbo"    "mska"        "mskb1"       "mskb2"       "mskc"
##  [26] "mskd"        "mhhuur"      "mhkoop"      "maut1"       "maut2"
##  [31] "maut0"       "mzfonds"     "mzpart"      "minkm30"     "mink3045"
```

```
##  [36] "mink4575"    "mink7512"    "mink123m"    "minkgem"     "mkoopkla"
##  [41] "pwabedr"     "pwaland"     "ppersaut"    "pbesaut"     "pmotsco"
##  [46] "pvraaut"     "paanhang"    "ptractor"    "pwerkt"      "pbrom"
##  [51] "pleven"      "ppersong"    "pgezong"     "pwaoreg"     "pbrand"
##  [56] "pzeilpl"     "pplezier"    "pfiets"      "pinboed"     "pbystand"
##  [61] "awapart"     "awabedr"     "awaland"     "apersaut"    "abesaut"
##  [66] "amotsco"     "avraaut"     "aaanhang"    "atractor"    "awerkt"
##  [71] "abrom"       "aleven"      "apersong"    "agezong"     "awaoreg"
##  [76] "abrand"      "azeilpl"     "aplezier"    "afiets"      "ainboed"
##  [81] "abystand"    "caravan"     "mostype_1"   "mostype_2"   "mostype_3"
##  [86] "mostype_4"   "mostype_5"   "mostype_6"   "mostype_7"   "mostype_8"
##  [91] "mostype_9"   "mostype_10"  "mostype_11"  "mostype_12"  "mostype_13"
##  [96] "mostype_15"  "mostype_16"  "mostype_17"  "mostype_18"  "mostype_19"
## [101] "mostype_20"  "mostype_21"  "mostype_22"  "mostype_23"  "mostype_24"
## [106] "mostype_25"  "mostype_26"  "mostype_27"  "mostype_28"  "mostype_29"
## [111] "mostype_30"  "mostype_31"  "mostype_32"  "mostype_33"  "mostype_34"
## [116] "mostype_35"  "mostype_36"  "mostype_37"  "mostype_38"  "mostype_39"
## [121] "mostype_40"  "mostype_41"  "mgemleef_1"  "mgemleef_2"  "mgemleef_3"
## [126] "mgemleef_4"  "mgemleef_5"  "mgemleef_6"  "moshoofd_1"  "moshoofd_2"
## [131] "moshoofd_3"  "moshoofd_4"  "moshoofd_5"  "moshoofd_6"  "moshoofd_7"
## [136] "moshoofd_8"  "moshoofd_9"  "moshoofd_10" "pwapart_0"   "pwapart_1"
## [141] "pwapart_2"   "pwapart_3"   "pwapart_4"   "pwapart_5"   "pwapart_6"
```

```r
model <- glm(caravan ~ ., data=dataset2)
summary(model)
```

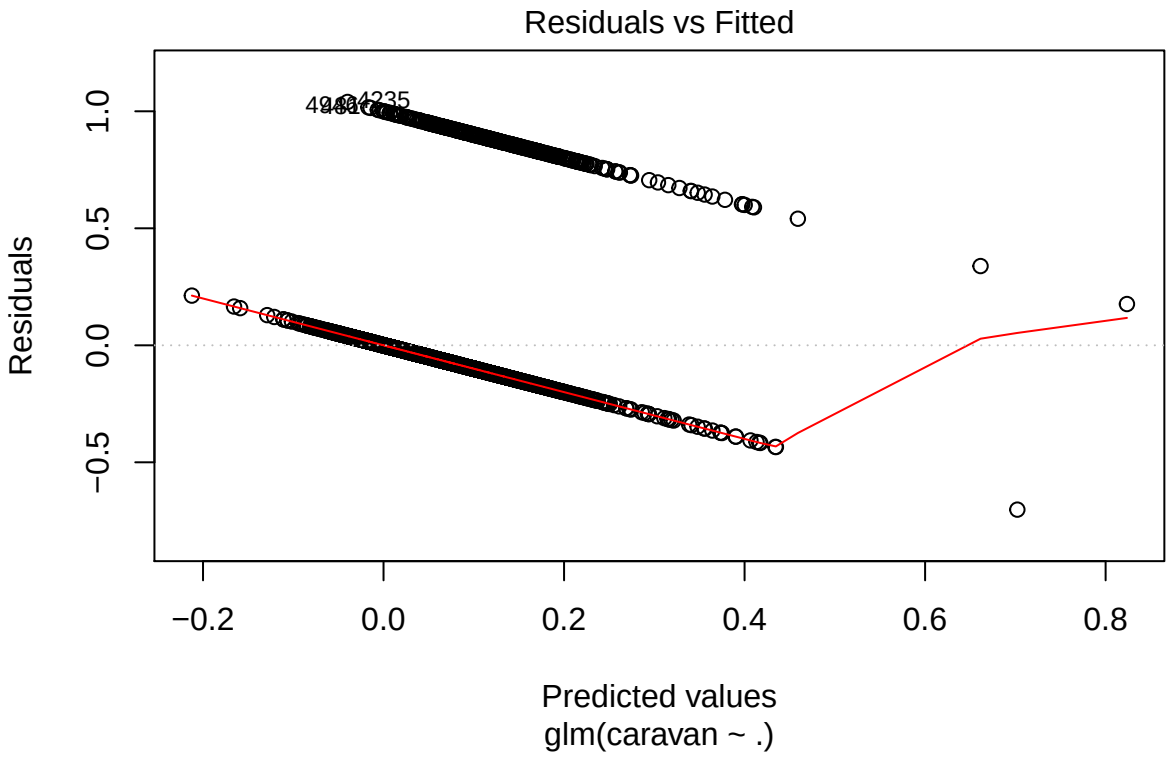
```
##
## Call:
## glm(formula = caravan ~ ., data = dataset2)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.70220  -0.08723  -0.04465  -0.00430   1.03988
##
## Coefficients: (16 not defined because of singularities)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.361e+00  4.818e-01   2.825  0.00475 **
## maanthui    -3.134e-03  8.837e-03  -0.355  0.72289
## mgemomv     -3.812e-04  7.368e-03  -0.052  0.95874
## mgodrk      -4.735e-03  5.751e-03  -0.823  0.41041
## mgodpr       1.098e-03  6.203e-03   0.177  0.85949
## mgodov       1.786e-03  5.563e-03   0.321  0.74826
## mgodge      -2.299e-03  5.926e-03  -0.388  0.69803
## mrelge       1.001e-02  7.715e-03   1.297  0.19464
## mrelsa       5.113e-03  7.339e-03   0.697  0.48597
## mrelov       7.115e-03  7.784e-03   0.914  0.36067
## mfalleen    -1.188e-03  6.706e-03  -0.177  0.85937
## mfgekind    -3.205e-03  6.932e-03  -0.462  0.64386
## mfwekind    -1.714e-03  7.218e-03  -0.237  0.81235
## moplhoog     3.088e-03  6.995e-03   0.441  0.65894
## moplmidd    -4.516e-03  7.317e-03  -0.617  0.53714
## mopllaag    -1.220e-02  7.449e-03  -1.638  0.10145
## mberhoog     2.093e-03  4.682e-03   0.447  0.65489
## mberzelf    -8.655e-04  5.551e-03  -0.156  0.87612
## mberboer    -3.245e-03  5.504e-03  -0.590  0.55554
```

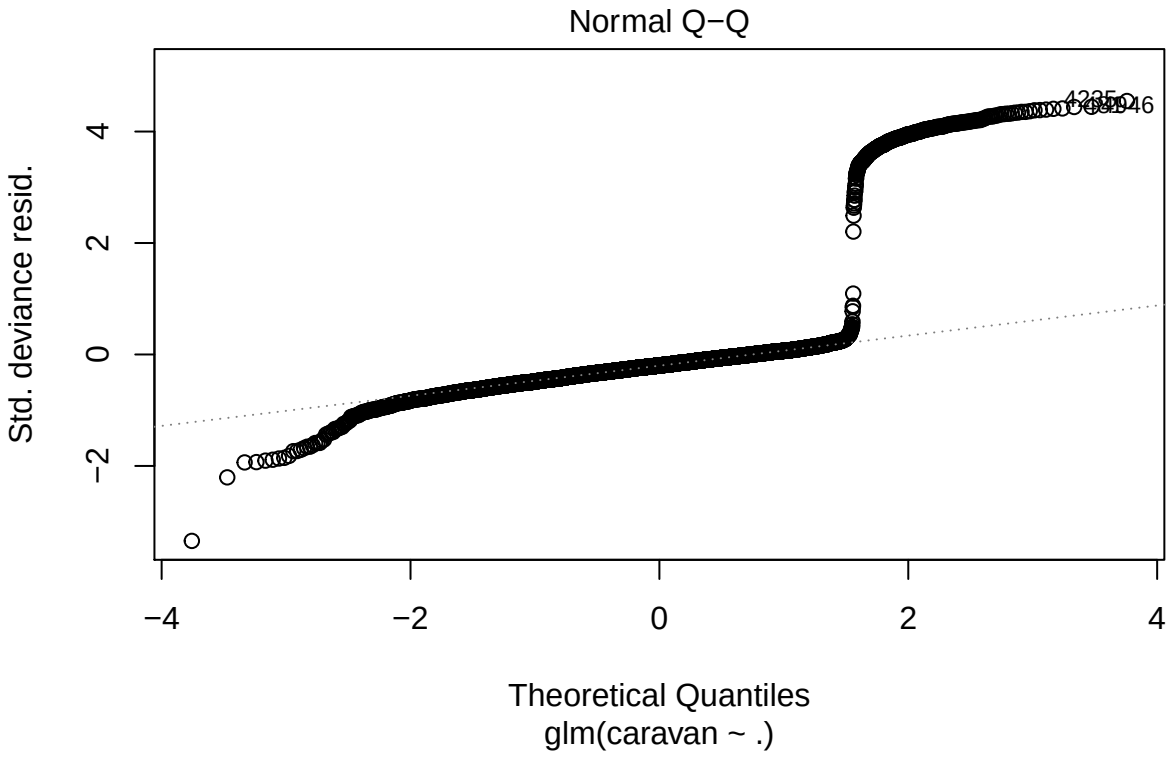
```
## mbermidd     3.106e-03  4.624e-03   0.672  0.50181
## mberarbg    -9.447e-04  4.602e-03  -0.205  0.83735
## mberarbo     1.963e-03  4.575e-03   0.429  0.66786
## mska        -2.206e-03  5.325e-03  -0.414  0.67870
## mskb1       -3.440e-03  5.144e-03  -0.669  0.50367
## mskb2       -5.497e-04  4.600e-03  -0.120  0.90488
## mskc         2.829e-03  5.033e-03   0.562  0.57405
## mskd        -1.734e-03  4.801e-03  -0.361  0.71794
## mhhuur      -4.687e-02  3.811e-02  -1.230  0.21883
## mhkoop      -4.521e-02  3.809e-02  -1.187  0.23532
## maut1        8.876e-03  7.916e-03   1.121  0.26219
## maut2        8.120e-03  7.091e-03   1.145  0.25219
## maut0        5.261e-03  7.569e-03   0.695  0.48702
## mzfonds     -5.935e-02  4.484e-02  -1.323  0.18573
## mzpart      -6.284e-02  4.477e-02  -1.404  0.16046
## minkm30      5.884e-03  5.239e-03   1.123  0.26147
## mink3045     6.166e-03  5.022e-03   1.228  0.21958
## mink4575     5.089e-03  5.105e-03   0.997  0.31886
## mink7512     5.662e-03  5.322e-03   1.064  0.28743
## mink123m    -1.159e-02  7.044e-03  -1.645  0.09995 .
## minkgem      4.557e-03  4.634e-03   0.984  0.32537
## mkoopkla     6.441e-03  1.633e-02   0.394  0.69333
## pwabedr     -5.881e-03  2.056e-02  -0.286  0.77481
## pwaland     -2.036e-02  3.907e-02  -0.521  0.60232
## ppersaut     1.037e-02  2.646e-03   3.920 8.94e-05 ***
## pbesaut      2.220e-03  1.490e-02   0.149  0.88159
## pmotsco     -8.150e-03  7.987e-03  -1.020  0.30761
## pvraaut     -2.641e-02  4.173e-02  -0.633  0.52685
## paanhang     5.915e-02  5.627e-02   1.051  0.29317
## ptractor     1.177e-02  1.426e-02   0.825  0.40926
## pwerkt      -5.804e-03  3.719e-02  -0.156  0.87599
## pbrom        6.319e-03  1.532e-02   0.412  0.68009
## pleven      -1.757e-02  6.529e-03  -2.690  0.00716 **
## ppersong     1.224e-02  3.365e-02   0.364  0.71596
## pgezong      1.953e-01  7.954e-02   2.455  0.01411 *
## pwaoreg      5.519e-02  2.594e-02   2.127  0.03344 *
## pbrand       1.162e-02  3.632e-03   3.200  0.00138 **
## pzeilpl     -1.860e-01  1.442e-01  -1.290  0.19699
## pplezier    -2.659e-02  2.700e-02  -0.985  0.32479
## pfiets      -5.803e-03  5.517e-02  -0.105  0.91623
## pinboed     -4.263e-02  3.090e-02  -1.380  0.16770
## pbystand    -2.114e-02  2.898e-02  -0.729  0.46577
## awapart     -2.830e-01  1.212e-01  -2.335  0.01957 *
## awabedr      1.057e-02  5.308e-02   0.199  0.84217
## awaland      2.136e-02  1.376e-01   0.155  0.87664
## apersaut    -6.862e-04  1.280e-02  -0.054  0.95724
## abesaut     -2.434e-02  6.543e-02  -0.372  0.70994
## amotsco      2.796e-02  3.130e-02   0.893  0.37181
## avraaut      6.375e-02  1.595e-01   0.400  0.68939
## aaanhang    -7.861e-02  9.522e-02  -0.826  0.40906
## atractor    -3.692e-02  3.544e-02  -1.042  0.29758
## awerkt       8.535e-04  7.318e-02   0.012  0.99069
## abrom       -2.133e-02  4.692e-02  -0.455  0.64938
## aleven       4.125e-02  1.549e-02   2.662  0.00778 **
```

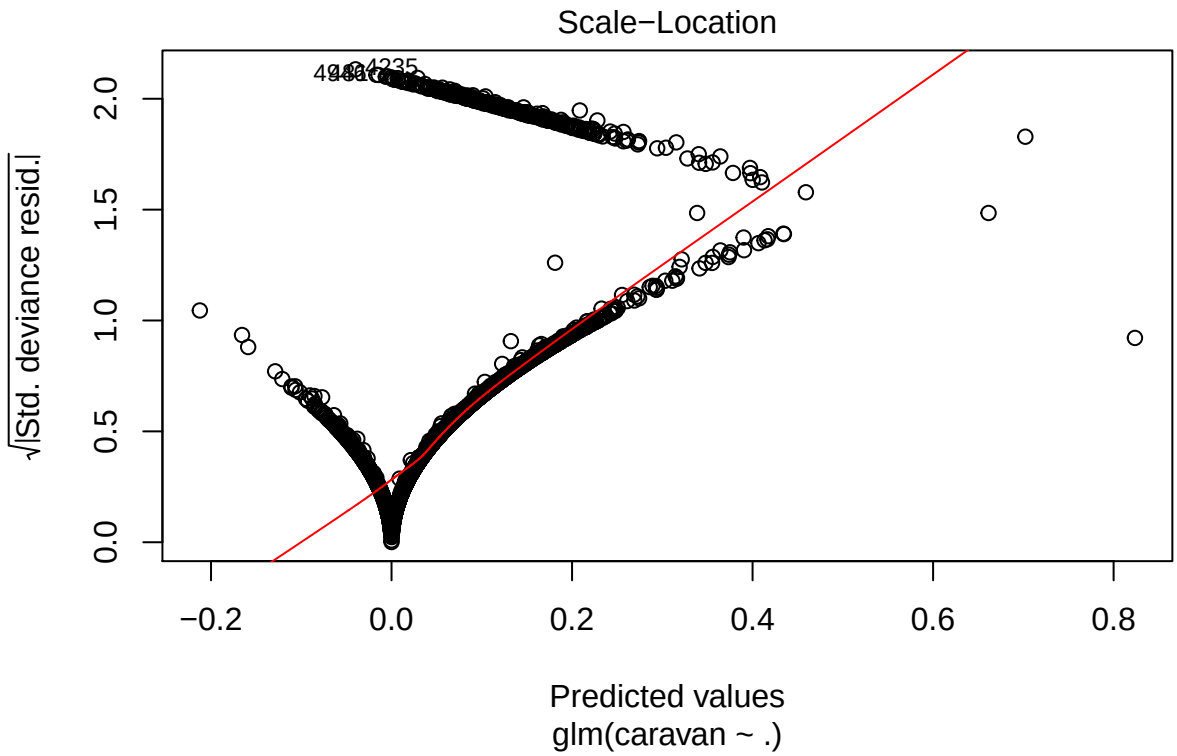
```
## apersong    -4.887e-02  9.572e-02  -0.511  0.60967
## agezong     -4.107e-01  1.903e-01  -2.159  0.03092 *
## awaoreg     -1.958e-01  1.250e-01  -1.567  0.11728
## abrand      -1.765e-02  1.170e-02  -1.509  0.13129
## azeilpl      4.976e-01  2.821e-01   1.764  0.07778 .
## aplezier     3.545e-01  8.877e-02   3.994 6.59e-05 ***
## afiets       3.618e-02  4.100e-02   0.882  0.37764
## ainboed      1.004e-01  7.015e-02   1.431  0.15252
## abystand     1.392e-01  9.867e-02   1.411  0.15843
## mostype_1   -5.408e-03  7.117e-02  -0.076  0.93944
## mostype_2    2.775e-03  4.702e-02   0.059  0.95295
## mostype_3    3.025e-02  4.027e-02   0.751  0.45263
## mostype_4   -4.912e-02  4.984e-02  -0.985  0.32447
## mostype_5    6.204e-03  4.363e-02   0.142  0.88691
## mostype_6    1.367e-02  7.052e-02   0.194  0.84633
## mostype_7   -4.028e-02  5.307e-02  -0.759  0.44793
## mostype_8    4.727e-02  5.426e-02   0.871  0.38370
## mostype_9    6.491e-03  2.482e-02   0.262  0.79369
## mostype_10  -4.458e-02  6.862e-02  -0.650  0.51590
## mostype_11  -5.520e-03  4.233e-02  -0.130  0.89625
## mostype_12   6.079e-02  5.671e-02   1.072  0.28377
## mostype_13  -9.475e-03  4.266e-02  -0.222  0.82423
## mostype_15   3.032e-02  1.198e-01   0.253  0.80017
## mostype_16   6.446e-04  8.450e-02   0.008  0.99391
## mostype_17  -9.461e-02  8.288e-02  -1.142  0.25369
## mostype_18  -8.442e-03  6.750e-02  -0.125  0.90047
## mostype_19  -4.738e-02  1.359e-01  -0.349  0.72732
## mostype_20   5.862e-02  6.085e-02   0.963  0.33542
## mostype_21   1.781e-02  7.935e-02   0.224  0.82239
## mostype_22   1.890e-02  4.551e-02   0.415  0.67795
## mostype_23   5.263e-04  3.232e-02   0.016  0.98701
## mostype_24   2.633e-02  4.380e-02   0.601  0.54768
## mostype_25   5.791e-03  5.970e-02   0.097  0.92272
## mostype_26   3.591e-02  6.613e-02   0.543  0.58716
## mostype_27   2.187e-02  6.456e-02   0.339  0.73483
## mostype_28  -9.030e-03  7.684e-02  -0.118  0.90645
## mostype_29   9.285e-03  3.691e-02   0.252  0.80140
## mostype_30   3.183e-02  4.629e-02   0.688  0.49178
## mostype_31   3.809e-02  5.659e-02   0.673  0.50096
## mostype_32   4.113e-02  5.711e-02   0.720  0.47153
## mostype_33   2.892e-02  2.659e-02   1.087  0.27692
## mostype_34   1.338e-03  4.153e-02   0.032  0.97429
## mostype_35   8.166e-05  2.918e-02   0.003  0.99777
## mostype_36   3.930e-02  3.027e-02   1.298  0.19421
## mostype_37   4.328e-02  2.783e-02   1.555  0.11996
## mostype_38   3.694e-02  2.354e-02   1.569  0.11669
## mostype_39   1.361e-02  2.775e-02   0.491  0.62378
## mostype_40   6.016e-04  3.822e-02   0.016  0.98744
## mostype_41          NA         NA      NA       NA
## mgemleef_1  -7.824e-02  5.707e-02  -1.371  0.17043
## mgemleef_2  -6.885e-02  4.960e-02  -1.388  0.16520
## mgemleef_3  -6.313e-02  4.891e-02  -1.291  0.19689
## mgemleef_4  -5.115e-02  4.788e-02  -1.068  0.28547
## mgemleef_5  -3.501e-02  4.817e-02  -0.727  0.46736
```

```
## mgemleef_6          NA         NA      NA       NA
## moshoofd_1          NA         NA      NA       NA
## moshoofd_2          NA         NA      NA       NA
## moshoofd_3          NA         NA      NA       NA
## moshoofd_4          NA         NA      NA       NA
## moshoofd_5          NA         NA      NA       NA
## moshoofd_6          NA         NA      NA       NA
## moshoofd_7          NA         NA      NA       NA
## moshoofd_8          NA         NA      NA       NA
## moshoofd_9          NA         NA      NA       NA
## moshoofd_10         NA         NA      NA       NA
## pwapart_0   -5.227e-01  1.913e-01  -2.732  0.00631 **
## pwapart_1   -2.413e-01  9.211e-02  -2.620  0.00882 **
## pwapart_2   -2.193e-01  9.065e-02  -2.420  0.01556 *
## pwapart_3           NA         NA      NA       NA
## pwapart_4           NA         NA      NA       NA
## pwapart_5           NA         NA      NA       NA
## pwapart_6           NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.05282748)
##
##     Null deviance: 327.20  on 5821  degrees of freedom
## Residual deviance: 300.75  on 5693  degrees of freedom
## AIC: -469.22
##
## Number of Fisher Scoring iterations: 2
```

```
plot(model)
```

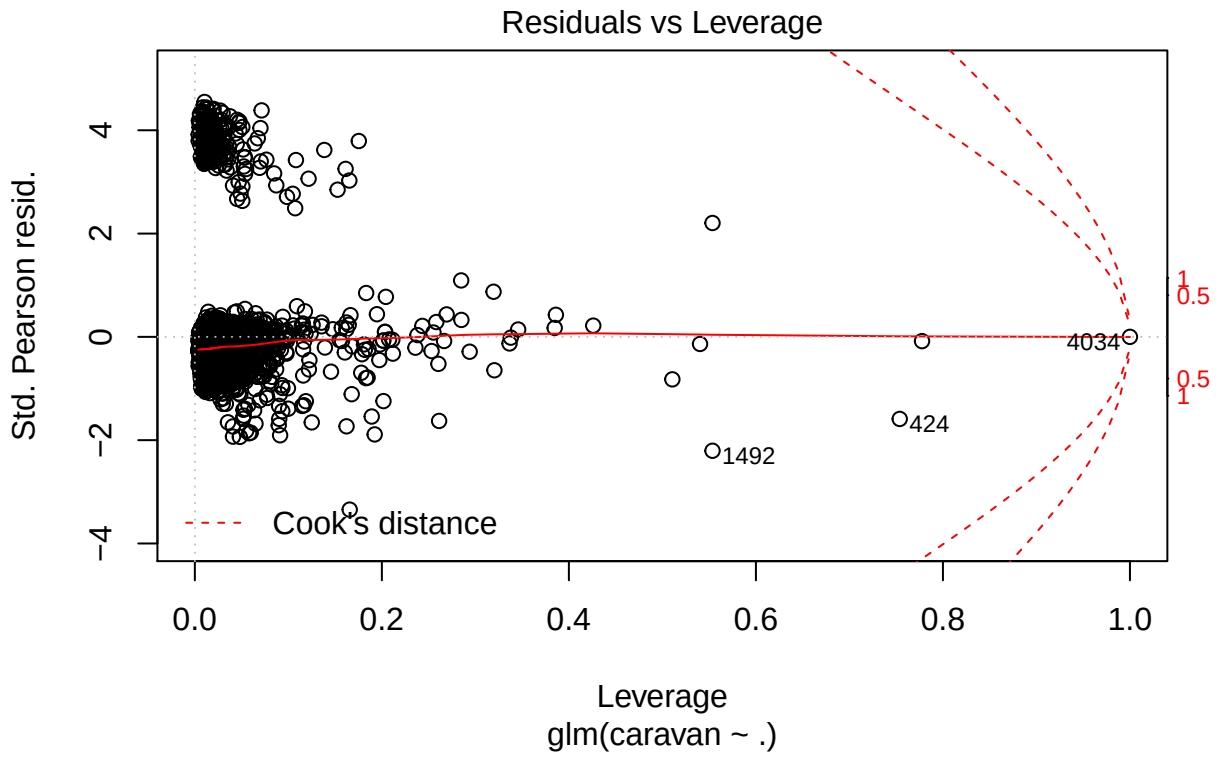

Residuals vs Fitted
Residuals
1.0
0.5
0.0
−0.5
4235
−0.2
0.0
0.2
0.4
0.6
0.8
Predicted values
glm(caravan ~ .)


Normal Q−Q
Std. deviance resid.
4
2
0
−2
4235
−4
−2
0
2
4
Theoretical Quantiles
glm(caravan ~ .)




```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```




```
model$coefficients
```

```
##   (Intercept)      maanthui       mgemomv        mgodrk        mgodpr
##  1.360726e+00 -3.133673e-03 -3.812391e-04 -4.734625e-03  1.098047e-03
##        mgodov        mgodge        mrelge        mrelsa        mrelov
##  1.785592e-03 -2.299402e-03  1.000687e-02  5.113485e-03  7.115483e-03
##      mfalleen      mfgekind      mfwekind      moplhoog      moplmidd
## -1.188102e-03 -3.204841e-03 -1.713660e-03  3.087734e-03 -4.515633e-03
##      mopllaag      mberhoog      mberzelf      mberboer      mbermidd
## -1.220314e-02  2.092835e-03 -8.654511e-04 -3.244541e-03  3.105891e-03
##      mberarbg      mberarbo          mska         mskb1         mskb2
## -9.447260e-04  1.963002e-03 -2.206132e-03 -3.440465e-03 -5.496738e-04
##          mskc          mskd        mhhuur        mhkoop         maut1
##  2.829103e-03 -1.734320e-03 -4.687020e-02 -4.520666e-02  8.876222e-03
##         maut2         maut0       mzfonds        mzpart       minkm30
##  8.120479e-03  5.261283e-03 -5.934722e-02 -6.284309e-02  5.883628e-03
##      mink3045      mink4575      mink7512      mink123m       minkgem
##  6.166085e-03  5.089119e-03  5.662056e-03 -1.158982e-02  4.557415e-03
##      mkoopkla       pwabedr       pwaland       ppersaut       pbesaut
##  6.440536e-03 -5.880913e-03 -2.035961e-02  1.037354e-02  2.219908e-03
##       pmotsco       pvraaut      paanhang      ptractor        pwerkt
## -8.149741e-03 -2.640645e-02  5.915277e-02  1.176973e-02 -5.803787e-03
##         pbrom        pleven      ppersong       pgezong       pwaoreg
##  6.319352e-03 -1.756567e-02  1.224282e-02  1.952957e-01  5.518698e-02
##        pbrand       pzeilpl      pplezier        pfiets       pinboed
##  1.162160e-02 -1.860415e-01 -2.659233e-02 -5.803456e-03 -4.263051e-02
```

```
##      pbystand       awapart       awabedr       awaland      apersaut
## -2.113775e-02 -2.829556e-01  1.056936e-02  2.135885e-02 -6.862120e-04
##       abesaut       amotsco       avraaut      aaanhang      atractor
## -2.433623e-02  2.795608e-02  6.375103e-02 -7.861365e-02 -3.692191e-02
##        awerkt         abrom        aleven      apersong       agezong
##  8.535186e-04 -2.133090e-02  4.124875e-02 -4.887132e-02 -4.107008e-01
##       awaoreg        abrand       azeilpl      aplezier        afiets
## -1.958012e-01 -1.765428e-02  4.975611e-01  3.545203e-01  3.617609e-02
##       ainboed      abystand     mostype_1     mostype_2     mostype_3
##  1.003798e-01  1.391827e-01 -5.407614e-03  2.774758e-03  3.024864e-02
##     mostype_4     mostype_5     mostype_6     mostype_7     mostype_8
## -4.911537e-02  6.204413e-03  1.366813e-02 -4.027754e-02  4.727229e-02
##     mostype_9    mostype_10    mostype_11    mostype_12    mostype_13
##  6.491022e-03 -4.458480e-02 -5.519969e-03  6.079466e-02 -9.474555e-03
##    mostype_15    mostype_16    mostype_17    mostype_18    mostype_19
##  3.032265e-02  6.446004e-04 -9.461386e-02 -8.442045e-03 -4.738392e-02
##    mostype_20    mostype_21    mostype_22    mostype_23    mostype_24
##  5.861920e-02  1.781359e-02  1.889936e-02  5.262732e-04  2.633434e-02
##    mostype_25    mostype_26    mostype_27    mostype_28    mostype_29
##  5.791125e-03  3.591071e-02  2.186740e-02 -9.030165e-03  9.284569e-03
##    mostype_30    mostype_31    mostype_32    mostype_33    mostype_34
##  3.182611e-02  3.808659e-02  4.112511e-02  2.891828e-02  1.338294e-03
##    mostype_35    mostype_36    mostype_37    mostype_38    mostype_39
##  8.165713e-05  3.930008e-02  4.328123e-02  3.694275e-02  1.361307e-02
##    mostype_40    mostype_41    mgemleef_1    mgemleef_2    mgemleef_3
##  6.015787e-04            NA -7.823926e-02 -6.884963e-02 -6.312941e-02
##    mgemleef_4    mgemleef_5    mgemleef_6    moshoofd_1    moshoofd_2
## -5.115068e-02 -3.501129e-02            NA            NA            NA
##    moshoofd_3    moshoofd_4    moshoofd_5    moshoofd_6    moshoofd_7
##            NA            NA            NA            NA            NA
##    moshoofd_8    moshoofd_9   moshoofd_10     pwapart_0     pwapart_1
##            NA            NA            NA -5.226660e-01 -2.413099e-01
##     pwapart_2     pwapart_3     pwapart_4     pwapart_5     pwapart_6
## -2.193484e-01            NA            NA            NA            NA
```

```
lr <- varImp(model, scale = FALSE)
lr
```

```
##               Overall
## maanthui   0.354613257
## mgemomv    0.051741104
## mgodrk     0.823229960
## mgodpr     0.177030067
## mgodov     0.320956241
## mgodge     0.387999160
## mrelge     1.297126396
## mrelsa     0.696779939
## mrelov     0.914169924
## mfalleen   0.177179600
## mfgekind   0.462329201
## mfwekind   0.237410327
## moplhoog   0.441394130
## moplmidd   0.617180708
## mopllaag   1.638121631
## mberhoog   0.446998098
```

```
## mberzelf   0.155899525
## mberboer   0.589511954
## mbermidd   0.671683416
## mberarbg   0.205295865
## mberarbo   0.429109297
## mska       0.414265560
## mskb1      0.668767730
## mskb2      0.119504018
## mskc       0.562124030
## mskd       0.361236648
## mhhuur     1.229787881
## mhkoop     1.186895076
## maut1      1.121338980
## maut2      1.145173544
## maut0      0.695097085
## mzfonds    1.323478144
## mzpart     1.403719265
## minkm30    1.123043548
## mink3045   1.227785729
## mink4575   0.996885530
## mink7512   1.063886828
## mink123m   1.645374330
## minkgem    0.983575013
## mkoopkla   0.394364535
## pwabedr    0.286099136
## pwaland    0.521096915
## ppersaut   3.920490357
## pbesaut    0.148964804
## pmotsco    1.020335631
## pvraaut    0.632864444
## paanhang   1.051286072
## ptractor   0.825254805
## pwerkt     0.156065279
## pbrom      0.412356618
## pleven     2.690203500
## ppersong   0.363880130
## pgezong    2.455181956
## pwaoreg    2.127289793
## pbrand     3.199766998
## pzeilpl    1.290319469
## pplezier   0.984757557
## pfiets     0.105194565
## pinboed    1.379822659
## pbystand   0.729423996
## awapart    2.335130121
## awabedr    0.199125365
## awaland    0.155236156
## apersaut   0.053618059
## abesaut    0.371958347
## amotsco    0.893149280
## avraaut    0.399698618
## aaanhang   0.825605359
## atractor   1.041728206
## awerkt     0.011663188
```

```
## abrom      0.454645976
## aleven     2.662315531
## apersong   0.510575389
## agezong    2.158660665
## awaoreg    1.566532613
## abrand     1.509265721
## azeilpl    1.764018474
## aplezier   3.993630808
## afiets     0.882318666
## ainboed    1.430885944
## abystand   1.410541736
## mostype_1  0.075977854
## mostype_2  0.059009197
## mostype_3  0.751086976
## mostype_4  0.985394935
## mostype_5  0.142221114
## mostype_6  0.193810557
## mostype_7  0.758921055
## mostype_8  0.871168949
## mostype_9  0.261537902
## mostype_10 0.649717297
## mostype_11 0.130410396
## mostype_12 1.072000425
## mostype_13 0.222111669
## mostype_15 0.253135067
## mostype_16 0.007628823
## mostype_17 1.141550835
## mostype_18 0.125067309
## mostype_19 0.348709957
## mostype_20 0.963327889
## mostype_21 0.224483911
## mostype_22 0.415278133
## mostype_23 0.016284988
## mostype_24 0.601275018
## mostype_25 0.097008513
## mostype_26 0.542991806
## mostype_27 0.338726371
## mostype_28 0.117525182
## mostype_29 0.251546189
## mostype_30 0.687530686
## mostype_31 0.673017688
## mostype_32 0.720047922
## mostype_33 1.087377714
## mostype_34 0.032225786
## mostype_35 0.002798309
## mostype_36 1.298377327
## mostype_37 1.555164090
## mostype_38 1.569045213
## mostype_39 0.490533362
## mostype_40 0.015740701
## mgemleef_1 1.371003541
## mgemleef_2 1.387961216
## mgemleef_3 1.290625938
## mgemleef_4 1.068204978
```

```
## mgemleef_5 0.726828303
## pwapart_0  2.732479973
## pwapart_1  2.619851096
## pwapart_2  2.419829696
```

```
v1 <- rownames(lr)
v2 <- lr$Overall
v3 <- cbind("field"=v1, "value"=v2)
v4 <- as.data.frame(v3)
v4$new <- round(as.numeric(v4$value, -2))
head(v4)
```

```
##      field              value new
## 1 maanthui  0.354613257018002  37
## 2  mgemomv 0.0517411036677462   7
## 3   mgodrk  0.823229959672397  73
## 4   mgodpr  0.177030066774465  22
## 5   mgodov  0.320956241332331  34
## 6   mgodge  0.387999159702177  41
```