Summary of the data

Janek Idziak

20 grudnia 2015

Please let me know of any suggestions or errors

This is inspired by the fantastic scripts: https://www.kaggle.com/darraghdog/springleaf-marketing-response/explore-springleaf https://www.kaggle.com/thie1e/rossmann-store-sales/exploratory-analysis-rossmann

Read in the data:

```
library(gridExtra)

setwd("C:\\Users\\iWindows\\Desktop\\Kaggle-prudentiak")
source("conf.R")
```

```
## making predictions for  multiclass
```

```
cat("reading the train and test data\n")
```

```
## reading the train and test data
```

```
train <- read_csv("input/train.csv")
test  <- read_csv("input/test.csv")
```

Separate categorical, continous and discrete variables

```
cat.var.names <- c(paste("Product_Info_", c(1:3,5:7), sep=""), paste("Employment_Info_", c(2,3,5), sep=
                   paste("InsuredInfo_", 1:7, sep=""), paste("Insurance_History_", c(1:4,7:9), sep=""),
                   "Family_Hist_1", paste("Medical_History_", c(2:14, 16:23, 25:31, 33:41), sep=""))
cont.var.names <- c("Product_Info_4", "Ins_Age", "Ht", "Wt", "BMI", "Employment_Info_1", "Employment_In
                    "Employment_Info_6", "Insurance_History_5", "Family_Hist_2", "Family_Hist_3", "Fami
                    "Family_Hist_5")
disc.var.names <- c("Medical_History_1", "Medical_History_15", "Medical_History_24", "Medical_History_3
                    paste("Medical_Keyword_", 1:48, sep=""))

train.cat <- train[, cat.var.names]
test.cat <- test[, cat.var.names]

train.cont <- train[, cont.var.names]
test.cont <- test[, cont.var.names]

train.disc <- train[, disc.var.names]
test.disc <- test[, disc.var.names]

train.cat <- as.data.frame(lapply(train.cat, factor))
test.cat <- as.data.frame(lapply(test.cat, factor))
```

Let's take a first look at the data:

```
str(train.cont)
```

```
## Classes 'tbl_df' and 'data.frame':   59381 obs. of  13 variables:
##  $ Product_Info_4     : num  0.0769 0.0769 0.0769 0.4872 0.2308 ...
##  $ Ins_Age            : num  0.6418 0.0597 0.0299 0.1642 0.4179 ...
##  $ Ht                 : num  0.582 0.6 0.745 0.673 0.655 ...
##  $ Wt                 : num  0.149 0.132 0.289 0.205 0.234 ...
##  $ BMI                : num  0.323 0.272 0.429 0.352 0.424 ...
##  $ Employment_Info_1  : num  0.028 0 0.03 0.042 0.027 0.325 0.11 0.12 0.165 0.025 ...
##  $ Employment_Info_4  : num  0 0 0 0 0 0 NA 0 0 0 ...
##  $ Employment_Info_6  : num  NA 0.0018 0.03 0.2 0.05 1 0.8 1 1 0.05 ...
##  $ Insurance_History_5: num  0.000667 0.000133 NA NA NA ...
##  $ Family_Hist_2      : num  NA 0.188 0.304 0.42 0.464 ...
##  $ Family_Hist_3      : num  0.598 NA NA NA NA ...
##  $ Family_Hist_4      : num  NA 0.0845 0.2254 0.3521 0.4085 ...
##  $ Family_Hist_5      : num  0.527 NA NA NA NA ...
```

```
str(train.disc)
```

```
## Classes 'tbl_df' and 'data.frame':   59381 obs. of  52 variables:
##  $ Medical_History_1 : int  4 5 10 0 NA 6 5 6 4 NA ...
##  $ Medical_History_15: int  240 0 NA NA NA NA NA NA NA NA ...
##  $ Medical_History_24: int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Medical_History_32: int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Medical_Keyword_1 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_2 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_3 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_4 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_5 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_6 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_7 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_8 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_9 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_10: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_11: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_12: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_13: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_14: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_15: int  0 0 0 0 0 0 0 0 0 1 ...
##  $ Medical_Keyword_16: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_17: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_18: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_19: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_20: int  0 0 0 0 0 0 0 0 1 0 ...
##  $ Medical_Keyword_21: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_22: int  0 0 0 0 0 1 0 0 0 0 ...
##  $ Medical_Keyword_23: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_24: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_25: int  0 0 0 0 0 0 0 0 0 1 ...
##  $ Medical_Keyword_26: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_27: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_28: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_29: int  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ Medical_Keyword_30: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_31: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_32: int  0 0 0 1 0 0 0 0 0 0 ...
##  $ Medical_Keyword_33: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_34: int  0 0 0 0 0 1 0 0 0 0 ...
##  $ Medical_Keyword_35: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_36: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_37: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_38: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_39: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_40: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_41: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_42: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_43: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_44: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_45: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_46: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_47: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_48: int  0 0 0 0 0 0 0 0 0 0 ...
```

```
str(test.cont)
```

```
## Classes 'tbl_df' and 'data.frame':   19765 obs. of  13 variables:
##  $ Product_Info_4     : num  0.4872 0.0769 0.1447 0.1517 0.0769 ...
##  $ Ins_Age            : num  0.612 0.627 0.582 0.522 0.299 ...
##  $ Ht                 : num  0.782 0.727 0.709 0.655 0.673 ...
##  $ Wt                 : num  0.339 0.312 0.32 0.268 0.247 ...
##  $ BMI                : num  0.472 0.485 0.519 0.487 0.429 ...
##  $ Employment_Info_1  : num  0.15 0 0.143 0.21 0.085 0.075 0.14 0.025 0.035 0.06 ...
##  $ Employment_Info_4  : num  0 0.07 0 0 0 0 0 0 0 0 ...
##  $ Employment_Info_6  : num  0.5 0.2 0.45 1 0.2 0.4 1 0 NA 1 ...
##  $ Insurance_History_5: num  NA 0.001667 NA 0.000667 NA ...
##  $ Family_Hist_2      : num  NA NA 0.667 NA 0.449 ...
##  $ Family_Hist_3      : num  0.627 0.529 NA 0.686 NA ...
##  $ Family_Hist_4      : num  0.761 0.746 0.662 0.676 0.38 ...
##  $ Family_Hist_5      : num  NA NA NA NA NA ...
```

```
str(test.disc)
```

```
## Classes 'tbl_df' and 'data.frame':   19765 obs. of  52 variables:
##  $ Medical_History_1 : int  2 5 3 NA 18 4 21 0 2 NA ...
##  $ Medical_History_15: int  NA 110 240 NA 188 NA 82 NA NA NA ...
##  $ Medical_History_24: int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Medical_History_32: int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Medical_Keyword_1 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_2 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_3 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_4 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_5 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_6 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_7 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_8 : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_9 : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ Medical_Keyword_10: int  0 0 1 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_11: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_12: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_13: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_14: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_15: int  1 0 0 0 0 0 0 0 1 0 ...
##  $ Medical_Keyword_16: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_17: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_18: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_19: int  1 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_20: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_21: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_22: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_23: int  0 0 1 0 0 0 1 0 0 0 ...
##  $ Medical_Keyword_24: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_25: int  1 0 1 0 1 0 0 0 1 0 ...
##  $ Medical_Keyword_26: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_27: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_28: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_29: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_30: int  0 0 0 0 0 0 0 1 0 0 ...
##  $ Medical_Keyword_31: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_32: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_33: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_34: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_35: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_36: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_37: int  0 0 0 1 0 0 0 0 0 0 ...
##  $ Medical_Keyword_38: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_39: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_40: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_41: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_42: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_43: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_44: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_45: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_46: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Medical_Keyword_47: int  0 0 0 1 0 0 0 0 0 0 ...
##  $ Medical_Keyword_48: int  0 0 0 1 0 0 0 0 0 0 ...
```

```
summary(train.cont)
```

```
##  Product_Info_4       Ins_Age             Ht              Wt
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.07692   1st Qu.:0.2388   1st Qu.:0.6545   1st Qu.:0.2259
##  Median :0.23077   Median :0.4030   Median :0.7091   Median :0.2887
##  Mean   :0.32895   Mean   :0.4056   Mean   :0.7073   Mean   :0.2926
##  3rd Qu.:0.48718   3rd Qu.:0.5672   3rd Qu.:0.7636   3rd Qu.:0.3452
##  Max.   :1.00000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
##       BMI         Employment_Info_1 Employment_Info_4 Employment_Info_6
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.000     Min.   :0.000
##  1st Qu.:0.3855   1st Qu.:0.03500   1st Qu.:0.000     1st Qu.:0.060
##  Median :0.4513   Median :0.06000   Median :0.000     Median :0.250
```

```
##  Mean   :0.4695   Mean   :0.07758   Mean   :0.006     Mean   :0.361
##  3rd Qu.:0.5329   3rd Qu.:0.10000   3rd Qu.:0.000     3rd Qu.:0.550
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.000     Max.   :1.000
##                   NA's   :19        NA's   :6779      NA's   :10854
##  Insurance_History_5 Family_Hist_2  Family_Hist_3   Family_Hist_4
##  Min.   :0.000       Min.   :0.000  Min.   :0.00    Min.   :0.000
##  1st Qu.:0.000       1st Qu.:0.362  1st Qu.:0.40    1st Qu.:0.324
##  Median :0.001       Median :0.464  Median :0.52    Median :0.423
##  Mean   :0.002       Mean   :0.475  Mean   :0.50    Mean   :0.445
##  3rd Qu.:0.002       3rd Qu.:0.580  3rd Qu.:0.60    3rd Qu.:0.563
##  Max.   :1.000       Max.   :1.000  Max.   :1.00    Max.   :0.944
##  NA's   :25396       NA's   :28656  NA's   :34241   NA's   :19184
##  Family_Hist_5
##  Min.   :0.00
##  1st Qu.:0.40
##  Median :0.51
##  Mean   :0.48
##  3rd Qu.:0.58
##  Max.   :1.00
##  NA's   :41811
```

```
summary(train.disc)
```

```
##  Medical_History_1 Medical_History_15 Medical_History_24
##  Min.   :  0.000   Min.   :  0.0      Min.   :  0.00
##  1st Qu.:  2.000   1st Qu.: 17.0      1st Qu.:  1.00
##  Median :  4.000   Median :117.0      Median :  8.00
##  Mean   :  7.962   Mean   :123.8      Mean   : 50.64
##  3rd Qu.:  9.000   3rd Qu.:240.0      3rd Qu.: 64.00
##  Max.   :240.000   Max.   :240.0      Max.   :240.00
##  NA's   :8889      NA's   :44596      NA's   :55580
##  Medical_History_32 Medical_Keyword_1 Medical_Keyword_2  Medical_Keyword_3
##  Min.   :  0.00     Min.   :0.000     Min.   :0.000000   Min.   :0.00000
##  1st Qu.:  0.00     1st Qu.:0.000     1st Qu.:0.000000   1st Qu.:0.00000
##  Median :  0.00     Median :0.000     Median :0.000000   Median :0.00000
##  Mean   : 11.97     Mean   :0.042     Mean   :0.008942   Mean   :0.04927
##  3rd Qu.:  2.00     3rd Qu.:0.000     3rd Qu.:0.000000   3rd Qu.:0.00000
##  Max.   :240.00     Max.   :1.000     Max.   :1.000000   Max.   :1.00000
##  NA's   :58274
##  Medical_Keyword_4 Medical_Keyword_5  Medical_Keyword_6 Medical_Keyword_7
##  Min.   :0.00000   Min.   :0.000000   Min.   :0.0000    Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.0000    1st Qu.:0.00000
##  Median :0.00000   Median :0.000000   Median :0.0000    Median :0.00000
##  Mean   :0.01455   Mean   :0.008622   Mean   :0.0126    Mean   :0.01391
##  3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.0000    3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.000000   Max.   :1.0000    Max.   :1.00000
##
##  Medical_Keyword_8 Medical_Keyword_9  Medical_Keyword_10
##  Min.   :0.00000   Min.   :0.000000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.00000
##  Median :0.00000   Median :0.000000   Median :0.00000
##  Mean   :0.01041   Mean   :0.006652   Mean   :0.03646
##  3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.000000   Max.   :1.00000
```

```
##
##  Medical_Keyword_11 Medical_Keyword_12 Medical_Keyword_13
##  Min.   :0.00000    Min.   :0.00       Min.   :0.000000
##  1st Qu.:0.00000    1st Qu.:0.00       1st Qu.:0.000000
##  Median :0.00000    Median :0.00       Median :0.000000
##  Mean   :0.05802    Mean   :0.01       Mean   :0.005961
##  3rd Qu.:0.00000    3rd Qu.:0.00       3rd Qu.:0.000000
##  Max.   :1.00000    Max.   :1.00       Max.   :1.000000
##
##  Medical_Keyword_14 Medical_Keyword_15 Medical_Keyword_16
##  Min.   :0.000000   Min.   :0.0000     Min.   :0.00000
##  1st Qu.:0.000000   1st Qu.:0.0000     1st Qu.:0.00000
##  Median :0.000000   Median :0.0000     Median :0.00000
##  Mean   :0.007848   Mean   :0.1905     Mean   :0.01271
##  3rd Qu.:0.000000   3rd Qu.:0.0000     3rd Qu.:0.00000
##  Max.   :1.000000   Max.   :1.0000     Max.   :1.00000
##
##  Medical_Keyword_17 Medical_Keyword_18 Medical_Keyword_19
##  Min.   :0.000000   Min.   :0.000000   Min.   :0.000000
##  1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000000
##  Median :0.000000   Median :0.000000   Median :0.000000
##  Mean   :0.009161   Mean   :0.007494   Mean   :0.009296
##  3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.000000
##  Max.   :1.000000   Max.   :1.000000   Max.   :1.000000
##
##  Medical_Keyword_20 Medical_Keyword_21 Medical_Keyword_22
##  Min.   :0.000000   Min.   :0.0000     Min.   :0.00000
##  1st Qu.:0.000000   1st Qu.:0.0000     1st Qu.:0.00000
##  Median :0.000000   Median :0.0000     Median :0.00000
##  Mean   :0.008134   Mean   :0.0146     Mean   :0.03717
##  3rd Qu.:0.000000   3rd Qu.:0.0000     3rd Qu.:0.00000
##  Max.   :1.000000   Max.   :1.0000     Max.   :1.00000
##
##  Medical_Keyword_23 Medical_Keyword_24 Medical_Keyword_25
##  Min.   :0.00000    Min.   :0.00000    Min.   :0.00000
##  1st Qu.:0.00000    1st Qu.:0.00000    1st Qu.:0.00000
##  Median :0.00000    Median :0.00000    Median :0.00000
##  Mean   :0.09778    Mean   :0.01889    Mean   :0.08946
##  3rd Qu.:0.00000    3rd Qu.:0.00000    3rd Qu.:0.00000
##  Max.   :1.00000    Max.   :1.00000    Max.   :1.00000
##
##  Medical_Keyword_26 Medical_Keyword_27 Medical_Keyword_28
##  Min.   :0.00000    Min.   :0.00000    Min.   :0.00000
##  1st Qu.:0.00000    1st Qu.:0.00000    1st Qu.:0.00000
##  Median :0.00000    Median :0.00000    Median :0.00000
##  Mean   :0.01344    Mean   :0.01186    Mean   :0.01494
##  3rd Qu.:0.00000    3rd Qu.:0.00000    3rd Qu.:0.00000
##  Max.   :1.00000    Max.   :1.00000    Max.   :1.00000
##
##  Medical_Keyword_29 Medical_Keyword_30 Medical_Keyword_31
##  Min.   :0.00000    Min.   :0.00000    Min.   :0.0000
##  1st Qu.:0.00000    1st Qu.:0.00000    1st Qu.:0.0000
##  Median :0.00000    Median :0.00000    Median :0.0000
##  Mean   :0.01175    Mean   :0.02504    Mean   :0.0109
```

```
##  3rd Qu.:0.00000    3rd Qu.:0.00000    3rd Qu.:0.0000
##  Max.   :1.00000    Max.   :1.00000    Max.   :1.0000
##
##  Medical_Keyword_32 Medical_Keyword_33 Medical_Keyword_34
##  Min.   :0.00000    Min.   :0.00000    Min.   :0.00000
##  1st Qu.:0.00000    1st Qu.:0.00000    1st Qu.:0.00000
##  Median :0.00000    Median :0.00000    Median :0.00000
##  Mean   :0.02117    Mean   :0.02284    Mean   :0.02065
##  3rd Qu.:0.00000    3rd Qu.:0.00000    3rd Qu.:0.00000
##  Max.   :1.00000    Max.   :1.00000    Max.   :1.00000
##
##  Medical_Keyword_35 Medical_Keyword_36 Medical_Keyword_37
##  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   :0.006938   Mean   :0.01041    Mean   :0.06659
##  3rd Qu.:0.000000   3rd Qu.:0.00000    3rd Qu.:0.00000
##  Max.   :1.000000   Max.   :1.00000    Max.   :1.00000
##
##  Medical_Keyword_38 Medical_Keyword_39 Medical_Keyword_40
##  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   :0.006837   Mean   :0.01366    Mean   :0.05695
##  3rd Qu.:0.000000   3rd Qu.:0.00000    3rd Qu.:0.00000
##  Max.   :1.000000   Max.   :1.00000    Max.   :1.00000
##
##  Medical_Keyword_41 Medical_Keyword_42 Medical_Keyword_43
##  Min.   :0.00000    Min.   :0.00000    Min.   :0.00000
##  1st Qu.:0.00000    1st Qu.:0.00000    1st Qu.:0.00000
##  Median :0.00000    Median :0.00000    Median :0.00000
##  Mean   :0.01005    Mean   :0.04554    Mean   :0.01071
##  3rd Qu.:0.00000    3rd Qu.:0.00000    3rd Qu.:0.00000
##  Max.   :1.00000    Max.   :1.00000    Max.   :1.00000
##
##  Medical_Keyword_44 Medical_Keyword_45 Medical_Keyword_46
##  Min.   :0.000000   Min.   :0.00000    Min.   :0.000000
##  1st Qu.:0.000000   1st Qu.:0.00000    1st Qu.:0.000000
##  Median :0.000000   Median :0.00000    Median :0.000000
##  Mean   :0.007528   Mean   :0.01369    Mean   :0.008488
##  3rd Qu.:0.000000   3rd Qu.:0.00000    3rd Qu.:0.000000
##  Max.   :1.000000   Max.   :1.00000    Max.   :1.000000
##
##  Medical_Keyword_47 Medical_Keyword_48
##  Min.   :0.00000    Min.   :0.0000
##  1st Qu.:0.00000    1st Qu.:0.0000
##  Median :0.00000    Median :0.0000
##  Mean   :0.01991    Mean   :0.0545
##  3rd Qu.:0.00000    3rd Qu.:0.0000
##  Max.   :1.00000    Max.   :1.0000
##
```

```
summary(test.cont)
```

```
##  Product_Info_4       Ins_Age             Ht               Wt
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.3455   Min.   :0.08368
##  1st Qu.:0.07692   1st Qu.:0.2537   1st Qu.:0.6545   1st Qu.:0.22594
##  Median :0.23077   Median :0.4179   Median :0.7091   Median :0.28870
##  Mean   :0.32438   Mean   :0.4149   Mean   :0.7058   Mean   :0.29256
##  3rd Qu.:0.48718   3rd Qu.:0.5821   3rd Qu.:0.7636   3rd Qu.:0.34519
##  Max.   :1.00000   Max.   :0.9701   Max.   :1.0000   Max.   :0.87866
##
##       BMI         Employment_Info_1 Employment_Info_4 Employment_Info_6
##  Min.   :0.0965   Min.   :0.00000   Min.   :0.0000    Min.   :0.000
##  1st Qu.:0.3847   1st Qu.:0.03500   1st Qu.:0.0000    1st Qu.:0.055
##  Median :0.4545   Median :0.06000   Median :0.0000    Median :0.250
##  Mean   :0.4707   Mean   :0.07898   Mean   :0.0065    Mean   :0.369
##  3rd Qu.:0.5345   3rd Qu.:0.10000   3rd Qu.:0.0000    3rd Qu.:0.600
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.0000    Max.   :1.000
##                   NA's   :3         NA's   :2137      NA's   :3787
##  Insurance_History_5 Family_Hist_2   Family_Hist_3   Family_Hist_4
##  Min.   :0.000       Min.   :0.043   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.000       1st Qu.:0.362   1st Qu.:0.412   1st Qu.:0.324
##  Median :0.001       Median :0.464   Median :0.520   Median :0.437
##  Mean   :0.002       Mean   :0.475   Mean   :0.500   Mean   :0.447
##  3rd Qu.:0.002       3rd Qu.:0.580   3rd Qu.:0.608   3rd Qu.:0.563
##  Max.   :0.131       Max.   :0.942   Max.   :0.892   Max.   :1.000
##  NA's   :8105        NA's   :9880    NA's   :11064   NA's   :6677
##  Family_Hist_5
##  Min.   :0.027
##  1st Qu.:0.420
##  Median :0.518
##  Mean   :0.492
##  3rd Qu.:0.589
##  Max.   :0.848
##  NA's   :13624
```

```
summary(test.disc)
```

```
##  Medical_History_1 Medical_History_15 Medical_History_24
##  Min.   :  0.000   Min.   :  0.0      Min.   :  0.00
##  1st Qu.:  2.000   1st Qu.: 19.0      1st Qu.:  1.00
##  Median :  4.000   Median :119.0      Median :  9.00
##  Mean   :  7.827   Mean   :125.7      Mean   : 49.85
##  3rd Qu.:  9.000   3rd Qu.:240.0      3rd Qu.: 60.25
##  Max.   :235.000   Max.   :240.0      Max.   :240.00
##  NA's   :2972      NA's   :14864      NA's   :18585
##  Medical_History_32 Medical_Keyword_1 Medical_Keyword_2 Medical_Keyword_3
##  Min.   :  0.00     Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:  0.00     1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :  0.00     Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   : 10.94     Mean   :0.04341   Mean   :0.00769   Mean   :0.05292
##  3rd Qu.:  1.50     3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :240.00     Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##  NA's   :19414
##  Medical_Keyword_4 Medical_Keyword_5  Medical_Keyword_6 Medical_Keyword_7
##  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   :0.01401   Mean   :0.007943   Mean   :0.01331   Mean   :0.01214
##  3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.000000   Max.   :1.00000   Max.   :1.00000
##
##  Medical_Keyword_8 Medical_Keyword_9  Medical_Keyword_10
##  Min.   :0.00000   Min.   :0.000000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.00000
##  Median :0.00000   Median :0.000000   Median :0.00000
##  Mean   :0.01027   Mean   :0.007589   Mean   :0.03693
##  3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.000000   Max.   :1.00000
##
##  Medical_Keyword_11 Medical_Keyword_12 Medical_Keyword_13
##  Min.   :0.00000    Min.   :0.000000   Min.   :0.000000
##  1st Qu.:0.00000    1st Qu.:0.000000   1st Qu.:0.000000
##  Median :0.00000    Median :0.000000   Median :0.000000
##  Mean   :0.06087    Mean   :0.007741   Mean   :0.007943
##  3rd Qu.:0.00000    3rd Qu.:0.000000   3rd Qu.:0.000000
##  Max.   :1.00000    Max.   :1.000000   Max.   :1.000000
##
##  Medical_Keyword_14 Medical_Keyword_15 Medical_Keyword_16
##  Min.   :0.0000     Min.   :0.0000     Min.   :0.00000
##  1st Qu.:0.0000     1st Qu.:0.0000     1st Qu.:0.00000
##  Median :0.0000     Median :0.0000     Median :0.00000
##  Mean   :0.0085     Mean   :0.1985     Mean   :0.01133
##  3rd Qu.:0.0000     3rd Qu.:0.0000     3rd Qu.:0.00000
##  Max.   :1.0000     Max.   :1.0000     Max.   :1.00000
##
##  Medical_Keyword_17 Medical_Keyword_18 Medical_Keyword_19
##  Min.   :0.00000    Min.   :0.000000   Min.   :0.000000
##  1st Qu.:0.00000    1st Qu.:0.000000   1st Qu.:0.000000
##  Median :0.00000    Median :0.000000   Median :0.000000
##  Mean   :0.01037    Mean   :0.006982   Mean   :0.008601
##  3rd Qu.:0.00000    3rd Qu.:0.000000   3rd Qu.:0.000000
##  Max.   :1.00000    Max.   :1.000000   Max.   :1.000000
##
##  Medical_Keyword_20 Medical_Keyword_21 Medical_Keyword_22
##  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   :0.007539   Mean   :0.01518    Mean   :0.03941
##  3rd Qu.:0.000000   3rd Qu.:0.00000    3rd Qu.:0.00000
##  Max.   :1.000000   Max.   :1.00000    Max.   :1.00000
##
##  Medical_Keyword_23 Medical_Keyword_24 Medical_Keyword_25
##  Min.   :0.0000     Min.   :0.00000    Min.   :0.0000
##  1st Qu.:0.0000     1st Qu.:0.00000    1st Qu.:0.0000
##  Median :0.0000     Median :0.00000    Median :0.0000
##  Mean   :0.1013     Mean   :0.01948    Mean   :0.1049
##  3rd Qu.:0.0000     3rd Qu.:0.00000    3rd Qu.:0.0000
##  Max.   :1.0000     Max.   :1.00000    Max.   :1.0000
##
##  Medical_Keyword_26 Medical_Keyword_27 Medical_Keyword_28
```

```
##  Min.   :0.00000    Min.   :0.00000    Min.   :0.00000
##  1st Qu.:0.00000    1st Qu.:0.00000    1st Qu.:0.00000
##  Median :0.00000    Median :0.00000    Median :0.00000
##  Mean   :0.01305    Mean   :0.01103    Mean   :0.01508
##  3rd Qu.:0.00000    3rd Qu.:0.00000    3rd Qu.:0.00000
##  Max.   :1.00000    Max.   :1.00000    Max.   :1.00000
##
##  Medical_Keyword_29 Medical_Keyword_30 Medical_Keyword_31
##  Min.   :0.00000    Min.   :0.00000    Min.   :0.00000
##  1st Qu.:0.00000    1st Qu.:0.00000    1st Qu.:0.00000
##  Median :0.00000    Median :0.00000    Median :0.00000
##  Mean   :0.01062    Mean   :0.02449    Mean   :0.01113
##  3rd Qu.:0.00000    3rd Qu.:0.00000    3rd Qu.:0.00000
##  Max.   :1.00000    Max.   :1.00000    Max.   :1.00000
##
##  Medical_Keyword_32 Medical_Keyword_33 Medical_Keyword_34
##  Min.   :0.00000    Min.   :0.00000    Min.   :0.00000
##  1st Qu.:0.00000    1st Qu.:0.00000    1st Qu.:0.00000
##  Median :0.00000    Median :0.00000    Median :0.00000
##  Mean   :0.01958    Mean   :0.02423    Mean   :0.01887
##  3rd Qu.:0.00000    3rd Qu.:0.00000    3rd Qu.:0.00000
##  Max.   :1.00000    Max.   :1.00000    Max.   :1.00000
##
##  Medical_Keyword_35 Medical_Keyword_36 Medical_Keyword_37
##  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   :0.006679   Mean   :0.01133    Mean   :0.06592
##  3rd Qu.:0.000000   3rd Qu.:0.00000    3rd Qu.:0.00000
##  Max.   :1.000000   Max.   :1.00000    Max.   :1.00000
##
##  Medical_Keyword_38 Medical_Keyword_39 Medical_Keyword_40
##  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   :0.006982   Mean   :0.01396    Mean   :0.05732
##  3rd Qu.:0.000000   3rd Qu.:0.00000    3rd Qu.:0.00000
##  Max.   :1.000000   Max.   :1.00000    Max.   :1.00000
##
##  Medical_Keyword_41 Medical_Keyword_42 Medical_Keyword_43
##  Min.   :0.00000    Min.   :0.00000    Min.   :0.00000
##  1st Qu.:0.00000    1st Qu.:0.00000    1st Qu.:0.00000
##  Median :0.00000    Median :0.00000    Median :0.00000
##  Mean   :0.01108    Mean   :0.04523    Mean   :0.01007
##  3rd Qu.:0.00000    3rd Qu.:0.00000    3rd Qu.:0.00000
##  Max.   :1.00000    Max.   :1.00000    Max.   :1.00000
##
##  Medical_Keyword_44 Medical_Keyword_45 Medical_Keyword_46
##  Min.   :0.000000   Min.   :0.00000    Min.   :0.000000
##  1st Qu.:0.000000   1st Qu.:0.00000    1st Qu.:0.000000
##  Median :0.000000   Median :0.00000    Median :0.000000
##  Mean   :0.008247   Mean   :0.01356    Mean   :0.008601
##  3rd Qu.:0.000000   3rd Qu.:0.00000    3rd Qu.:0.000000
##  Max.   :1.000000   Max.   :1.00000    Max.   :1.000000
```

```
##
##  Medical_Keyword_47 Medical_Keyword_48
##  Min.   :0.00000    Min.   :0.00000
##  1st Qu.:0.00000    1st Qu.:0.00000
##  Median :0.00000    Median :0.00000
##  Mean   :0.01832    Mean   :0.05631
##  3rd Qu.:0.00000    3rd Qu.:0.00000
##  Max.   :1.00000    Max.   :1.00000
##
```

```
head(train)
```

```
## Source: local data frame [6 x 128]
##
##      Id Product_Info_1 Product_Info_2 Product_Info_3 Product_Info_4
##   (int)          (int)          (chr)          (int)          (dbl)
## 1     2              1             D3             10     0.07692308
## 2     5              1             A1             26     0.07692308
## 3     6              1             E1             26     0.07692308
## 4     7              1             D4             10     0.48717949
## 5     8              1             D2             26     0.23076923
## 6    10              1             D2             26     0.23076923
## Variables not shown: Product_Info_5 (int), Product_Info_6 (int),
##   Product_Info_7 (int), Ins_Age (dbl), Ht (dbl), Wt (dbl), BMI (dbl),
##   Employment_Info_1 (dbl), Employment_Info_2 (int), Employment_Info_3
##   (int), Employment_Info_4 (dbl), Employment_Info_5 (int),
##   Employment_Info_6 (dbl), InsuredInfo_1 (int), InsuredInfo_2 (int),
##   InsuredInfo_3 (int), InsuredInfo_4 (int), InsuredInfo_5 (int),
##   InsuredInfo_6 (int), InsuredInfo_7 (int), Insurance_History_1 (int),
##   Insurance_History_2 (int), Insurance_History_3 (int),
##   Insurance_History_4 (int), Insurance_History_5 (dbl),
##   Insurance_History_7 (int), Insurance_History_8 (int),
##   Insurance_History_9 (int), Family_Hist_1 (int), Family_Hist_2 (dbl),
##   Family_Hist_3 (dbl), Family_Hist_4 (dbl), Family_Hist_5 (dbl),
##   Medical_History_1 (int), Medical_History_2 (int), Medical_History_3
##   (int), Medical_History_4 (int), Medical_History_5 (int),
##   Medical_History_6 (int), Medical_History_7 (int), Medical_History_8
##   (int), Medical_History_9 (int), Medical_History_10 (int),
##   Medical_History_11 (int), Medical_History_12 (int), Medical_History_13
##   (int), Medical_History_14 (int), Medical_History_15 (int),
##   Medical_History_16 (int), Medical_History_17 (int), Medical_History_18
##   (int), Medical_History_19 (int), Medical_History_20 (int),
##   Medical_History_21 (int), Medical_History_22 (int), Medical_History_23
##   (int), Medical_History_24 (int), Medical_History_25 (int),
##   Medical_History_26 (int), Medical_History_27 (int), Medical_History_28
##   (int), Medical_History_29 (int), Medical_History_30 (int),
##   Medical_History_31 (int), Medical_History_32 (int), Medical_History_33
##   (int), Medical_History_34 (int), Medical_History_35 (int),
##   Medical_History_36 (int), Medical_History_37 (int), Medical_History_38
##   (int), Medical_History_39 (int), Medical_History_40 (int),
##   Medical_History_41 (int), Medical_Keyword_1 (int), Medical_Keyword_2
##   (int), Medical_Keyword_3 (int), Medical_Keyword_4 (int),
##   Medical_Keyword_5 (int), Medical_Keyword_6 (int), Medical_Keyword_7
##   (int), Medical_Keyword_8 (int), Medical_Keyword_9 (int),
```

```
##   Medical_Keyword_10 (int), Medical_Keyword_11 (int), Medical_Keyword_12
##   (int), Medical_Keyword_13 (int), Medical_Keyword_14 (int),
##   Medical_Keyword_15 (int), Medical_Keyword_16 (int), Medical_Keyword_17
##   (int), Medical_Keyword_18 (int), Medical_Keyword_19 (int),
##   Medical_Keyword_20 (int), Medical_Keyword_21 (int), Medical_Keyword_22
##   (int), Medical_Keyword_23 (int), Medical_Keyword_24 (int),
##   Medical_Keyword_25 (int), Medical_Keyword_26 (int), Medical_Keyword_27
##   (int), Medical_Keyword_28 (int), Medical_Keyword_29 (int),
##   Medical_Keyword_30 (int), Medical_Keyword_31 (int), Medical_Keyword_32
##   (int), Medical_Keyword_33 (int), Medical_Keyword_34 (int),
##   Medical_Keyword_35 (int), Medical_Keyword_36 (int), Medical_Keyword_37
##   (int), Medical_Keyword_38 (int), Medical_Keyword_39 (int),
##   Medical_Keyword_40 (int), Medical_Keyword_41 (int), Medical_Keyword_42
##   (int), Medical_Keyword_43 (int), Medical_Keyword_44 (int),
##   Medical_Keyword_45 (int), Medical_Keyword_46 (int), Medical_Keyword_47
##   (int), Medical_Keyword_48 (int), Response (int)
```

```
head(test)
```

```
## Source: local data frame [6 x 127]
##
##      Id Product_Info_1 Product_Info_2 Product_Info_3 Product_Info_4
##   (int)          (int)          (chr)          (int)          (dbl)
## 1     1              1             D3             26     0.48717949
## 2     3              1             A2             26     0.07692308
## 3     4              1             D3             26     0.14466667
## 4     9              1             A1             26     0.15170872
## 5    12              1             A1             26     0.07692308
## 6    13              1             D3             26     0.23076923
## Variables not shown: Product_Info_5 (int), Product_Info_6 (int),
##   Product_Info_7 (int), Ins_Age (dbl), Ht (dbl), Wt (dbl), BMI (dbl),
##   Employment_Info_1 (dbl), Employment_Info_2 (int), Employment_Info_3
##   (int), Employment_Info_4 (dbl), Employment_Info_5 (int),
##   Employment_Info_6 (dbl), InsuredInfo_1 (int), InsuredInfo_2 (int),
##   InsuredInfo_3 (int), InsuredInfo_4 (int), InsuredInfo_5 (int),
##   InsuredInfo_6 (int), InsuredInfo_7 (int), Insurance_History_1 (int),
##   Insurance_History_2 (int), Insurance_History_3 (int),
##   Insurance_History_4 (int), Insurance_History_5 (dbl),
##   Insurance_History_7 (int), Insurance_History_8 (int),
##   Insurance_History_9 (int), Family_Hist_1 (int), Family_Hist_2 (dbl),
##   Family_Hist_3 (dbl), Family_Hist_4 (dbl), Family_Hist_5 (dbl),
##   Medical_History_1 (int), Medical_History_2 (int), Medical_History_3
##   (int), Medical_History_4 (int), Medical_History_5 (int),
##   Medical_History_6 (int), Medical_History_7 (int), Medical_History_8
##   (int), Medical_History_9 (int), Medical_History_10 (int),
##   Medical_History_11 (int), Medical_History_12 (int), Medical_History_13
##   (int), Medical_History_14 (int), Medical_History_15 (int),
##   Medical_History_16 (int), Medical_History_17 (int), Medical_History_18
##   (int), Medical_History_19 (int), Medical_History_20 (int),
##   Medical_History_21 (int), Medical_History_22 (int), Medical_History_23
##   (int), Medical_History_24 (int), Medical_History_25 (int),
##   Medical_History_26 (int), Medical_History_27 (int), Medical_History_28
##   (int), Medical_History_29 (int), Medical_History_30 (int),
##   Medical_History_31 (int), Medical_History_32 (int), Medical_History_33
```

```
##   (int), Medical_History_34 (int), Medical_History_35 (int),
##   Medical_History_36 (int), Medical_History_37 (int), Medical_History_38
##   (int), Medical_History_39 (int), Medical_History_40 (int),
##   Medical_History_41 (int), Medical_Keyword_1 (int), Medical_Keyword_2
##   (int), Medical_Keyword_3 (int), Medical_Keyword_4 (int),
##   Medical_Keyword_5 (int), Medical_Keyword_6 (int), Medical_Keyword_7
##   (int), Medical_Keyword_8 (int), Medical_Keyword_9 (int),
##   Medical_Keyword_10 (int), Medical_Keyword_11 (int), Medical_Keyword_12
##   (int), Medical_Keyword_13 (int), Medical_Keyword_14 (int),
##   Medical_Keyword_15 (int), Medical_Keyword_16 (int), Medical_Keyword_17
##   (int), Medical_Keyword_18 (int), Medical_Keyword_19 (int),
##   Medical_Keyword_20 (int), Medical_Keyword_21 (int), Medical_Keyword_22
##   (int), Medical_Keyword_23 (int), Medical_Keyword_24 (int),
##   Medical_Keyword_25 (int), Medical_Keyword_26 (int), Medical_Keyword_27
##   (int), Medical_Keyword_28 (int), Medical_Keyword_29 (int),
##   Medical_Keyword_30 (int), Medical_Keyword_31 (int), Medical_Keyword_32
##   (int), Medical_Keyword_33 (int), Medical_Keyword_34 (int),
##   Medical_Keyword_35 (int), Medical_Keyword_36 (int), Medical_Keyword_37
##   (int), Medical_Keyword_38 (int), Medical_Keyword_39 (int),
##   Medical_Keyword_40 (int), Medical_Keyword_41 (int), Medical_Keyword_42
##   (int), Medical_Keyword_43 (int), Medical_Keyword_44 (int),
##   Medical_Keyword_45 (int), Medical_Keyword_46 (int), Medical_Keyword_47
##   (int), Medical_Keyword_48 (int)
```

Take a look at the data with categorical features:

```
str(train.cat)
```

```
## 'data.frame':    59381 obs. of  61 variables:
##  $ Product_Info_1     : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Product_Info_2     : Factor w/ 19 levels "A1","A2","A3",..: 17 1 19 18 16 16 8 16 17 19 ...
##  $ Product_Info_3     : Factor w/ 34 levels "1","2","3","4",..: 9 23 23 9 23 23 9 23 23 19 ...
##  $ Product_Info_5     : Factor w/ 2 levels "2","3": 1 1 1 1 1 2 1 1 1 1 ...
##  $ Product_Info_6     : Factor w/ 2 levels "1","3": 1 2 2 2 2 1 2 2 2 2 ...
##  $ Product_Info_7     : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Employment_Info_2  : Factor w/ 36 levels "1","2","3","4",..: 11 1 8 8 8 14 1 11 8 1 ...
##  $ Employment_Info_3  : Factor w/ 2 levels "1","3": 1 2 1 1 1 1 2 1 1 2 ...
##  $ Employment_Info_5  : Factor w/ 2 levels "2","3": 2 1 1 2 1 1 2 1 1 2 ...
##  $ InsuredInfo_1      : Factor w/ 3 levels "1","2","3": 1 1 1 2 1 1 1 1 1 2 ...
##  $ InsuredInfo_2      : Factor w/ 2 levels "2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ InsuredInfo_3      : Factor w/ 11 levels "1","2","3","4",..: 6 6 8 8 6 8 3 6 3 3 ...
##  $ InsuredInfo_4      : Factor w/ 2 levels "2","3": 2 2 2 2 2 2 2 2 1 2 ...
##  $ InsuredInfo_5      : Factor w/ 2 levels "1","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ InsuredInfo_6      : Factor w/ 2 levels "1","2": 2 2 1 2 2 1 2 1 1 2 ...
##  $ InsuredInfo_7      : Factor w/ 2 levels "1","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Insurance_History_1: Factor w/ 2 levels "1","2": 1 2 2 2 2 2 1 1 1 2 ...
##  $ Insurance_History_2: Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Insurance_History_3: Factor w/ 3 levels "1","2","3": 3 3 1 1 1 3 3 3 3 3 ...
##  $ Insurance_History_4: Factor w/ 3 levels "1","2","3": 1 1 3 3 3 2 2 1 2 1 ...
##  $ Insurance_History_7: Factor w/ 3 levels "1","2","3": 1 1 3 3 3 1 1 1 1 1 ...
##  $ Insurance_History_8: Factor w/ 3 levels "1","2","3": 1 3 2 2 2 3 1 1 1 3 ...
##  $ Insurance_History_9: Factor w/ 3 levels "1","2","3": 2 2 3 3 3 2 2 2 2 2 ...
##  $ Family_Hist_1      : Factor w/ 3 levels "1","2","3": 2 2 3 3 2 2 3 2 3 3 ...
```

```
##  $ Medical_History_2  : Factor w/ 579 levels "1","2","3","5",..: 103 369 3 313 149 437 534 132 14 149
##  $ Medical_History_3  : Factor w/ 3 levels "1","2","3": 2 2 2 2 2 2 3 2 2 2 ...
##  $ Medical_History_4  : Factor w/ 2 levels "1","2": 1 1 2 2 2 2 2 2 2 2 ...
##  $ Medical_History_5  : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Medical_History_6  : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
##  $ Medical_History_7  : Factor w/ 3 levels "1","2","3": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Medical_History_8  : Factor w/ 3 levels "1","2","3": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Medical_History_9  : Factor w/ 3 levels "1","2","3": 1 1 2 2 2 2 1 1 1 2 ...
##  $ Medical_History_10 : Factor w/ 103 levels "0","1","2","3",..: NA NA NA NA NA NA NA NA NA NA ...
##  $ Medical_History_11 : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
##  $ Medical_History_12 : Factor w/ 3 levels "1","2","3": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Medical_History_13 : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
##  $ Medical_History_14 : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
##  $ Medical_History_16 : Factor w/ 3 levels "1","2","3": 3 1 1 1 1 1 1 1 1 3 ...
##  $ Medical_History_17 : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
##  $ Medical_History_18 : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 2 1 1 1 1 ...
##  $ Medical_History_19 : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Medical_History_20 : Factor w/ 3 levels "1","2","3": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Medical_History_21 : Factor w/ 3 levels "1","2","3": 1 1 1 2 1 2 1 1 1 1 ...
##  $ Medical_History_22 : Factor w/ 2 levels "1","2": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Medical_History_23 : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 1 ...
##  $ Medical_History_25 : Factor w/ 3 levels "1","2","3": 1 1 2 1 2 1 1 1 1 1 ...
##  $ Medical_History_26 : Factor w/ 3 levels "1","2","3": 3 3 2 3 2 3 3 3 3 3 ...
##  $ Medical_History_27 : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
##  $ Medical_History_28 : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Medical_History_29 : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 1 3 1 3 ...
##  $ Medical_History_30 : Factor w/ 3 levels "1","2","3": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Medical_History_31 : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
##  $ Medical_History_33 : Factor w/ 2 levels "1","3": 1 2 2 2 2 2 2 2 2 2 ...
##  $ Medical_History_34 : Factor w/ 3 levels "1","2","3": 3 1 3 3 3 1 3 3 3 3 ...
##  $ Medical_History_35 : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Medical_History_36 : Factor w/ 3 levels "1","2","3": 2 2 3 2 3 2 2 2 2 2 ...
##  $ Medical_History_37 : Factor w/ 3 levels "1","2","3": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Medical_History_38 : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Medical_History_39 : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
##  $ Medical_History_40 : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
##  $ Medical_History_41 : Factor w/ 3 levels "1","2","3": 3 1 1 1 1 3 3 1 3 1 ...
```

```
str(test.cat)
```

```
## 'data.frame':    19765 obs. of  61 variables:
##  $ Product_Info_1     : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Product_Info_2     : Factor w/ 19 levels "A1","A2","A3",..: 17 2 17 1 1 17 3 18 17 3 ...
##  $ Product_Info_3     : Factor w/ 26 levels "2","4","6","7",..: 17 17 17 17 17 17 17 17 17 17 ...
##  $ Product_Info_5     : Factor w/ 2 levels "2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Product_Info_6     : Factor w/ 2 levels "1","3": 2 2 2 1 2 2 2 2 2 2 ...
##  $ Product_Info_7     : Factor w/ 2 levels "1","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Employment_Info_2  : Factor w/ 28 levels "1","2","3","4",..: 3 1 9 9 9 9 9 9 9 9 ...
##  $ Employment_Info_3  : Factor w/ 2 levels "1","3": 1 2 1 1 1 1 1 1 1 1 ...
##  $ Employment_Info_5  : Factor w/ 2 levels "2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ InsuredInfo_1      : Factor w/ 3 levels "1","2","3": 2 1 1 2 1 1 2 1 2 1 ...
##  $ InsuredInfo_2      : Factor w/ 2 levels "2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ InsuredInfo_3      : Factor w/ 9 levels "1","2","3","4",..: 9 6 3 3 6 6 3 2 6 6 ...
##  $ InsuredInfo_4      : Factor w/ 2 levels "2","3": 2 2 2 2 2 2 2 2 2 2 ...
```

```
##  $ InsuredInfo_5      : Factor w/ 2 levels "1","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ InsuredInfo_6      : Factor w/ 2 levels "1","2": 1 1 1 1 2 1 1 2 1 1 ...
##  $ InsuredInfo_7      : Factor w/ 2 levels "1","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Insurance_History_1: Factor w/ 2 levels "1","2": 2 1 2 1 2 2 2 2 2 2 ...
##  $ Insurance_History_2: Factor w/ 2 levels "1","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Insurance_History_3: Factor w/ 2 levels "1","3": 1 2 1 2 1 1 2 1 1 1 ...
##  $ Insurance_History_4: Factor w/ 3 levels "1","2","3": 3 1 3 1 3 3 1 3 3 3 ...
##  $ Insurance_History_7: Factor w/ 3 levels "1","2","3": 3 1 3 2 3 3 1 3 3 3 ...
##  $ Insurance_History_8: Factor w/ 3 levels "1","2","3": 2 1 2 1 2 2 3 2 2 2 ...
##  $ Insurance_History_9: Factor w/ 3 levels "1","2","3": 3 2 3 2 3 3 2 3 3 3 ...
##  $ Family_Hist_1      : Factor w/ 3 levels "1","2","3": 3 2 3 2 2 3 2 2 3 3 ...
##  $ Medical_History_2  : Factor w/ 426 levels "1","2","3","4",..: 13 178 98 116 130 229 84 329 84 116
##  $ Medical_History_3  : Factor w/ 2 levels "2","3": 1 2 1 2 2 1 1 1 1 2 ...
##  $ Medical_History_4  : Factor w/ 2 levels "1","2": 2 1 1 2 1 2 1 2 2 2 ...
##  $ Medical_History_5  : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Medical_History_6  : Factor w/ 2 levels "1","3": 2 2 2 1 2 2 2 2 2 2 ...
##  $ Medical_History_7  : Factor w/ 3 levels "1","2","3": 1 2 2 2 2 2 2 2 2 2 ...
##  $ Medical_History_8  : Factor w/ 3 levels "1","2","3": 2 2 2 3 2 2 2 2 2 2 ...
##  $ Medical_History_9  : Factor w/ 2 levels "1","2": 2 1 2 2 2 2 1 2 2 2 ...
##  $ Medical_History_10 : Factor w/ 56 levels "0","1","2","3",..: NA NA NA NA NA NA NA NA NA NA ...
##  $ Medical_History_11 : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
##  $ Medical_History_12 : Factor w/ 2 levels "2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Medical_History_13 : Factor w/ 3 levels "1","2","3": 1 3 3 3 3 3 3 3 3 3 ...
##  $ Medical_History_14 : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
##  $ Medical_History_16 : Factor w/ 2 levels "1","3": 1 2 1 1 1 1 1 1 1 1 ...
##  $ Medical_History_17 : Factor w/ 2 levels "2","3": 1 2 2 2 2 2 1 2 2 2 ...
##  $ Medical_History_18 : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Medical_History_19 : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 2 1 1 ...
##  $ Medical_History_20 : Factor w/ 3 levels "1","2","3": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Medical_History_21 : Factor w/ 3 levels "1","2","3": 1 1 1 2 1 1 1 1 1 1 ...
##  $ Medical_History_22 : Factor w/ 2 levels "1","2": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Medical_History_23 : Factor w/ 2 levels "1","3": 1 2 2 2 1 2 2 2 1 2 ...
##  $ Medical_History_25 : Factor w/ 3 levels "1","2","3": 2 2 2 1 1 2 1 1 1 1 ...
##  $ Medical_History_26 : Factor w/ 3 levels "1","2","3": 2 2 2 3 3 2 3 3 3 3 ...
##  $ Medical_History_27 : Factor w/ 3 levels "1","2","3": 1 3 3 3 3 3 3 3 3 3 ...
##  $ Medical_History_28 : Factor w/ 3 levels "1","2","3": 1 1 1 2 1 1 1 1 1 1 ...
##  $ Medical_History_29 : Factor w/ 3 levels "1","2","3": 3 3 1 3 1 3 3 3 3 3 ...
##  $ Medical_History_30 : Factor w/ 3 levels "1","2","3": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Medical_History_31 : Factor w/ 2 levels "1","3": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Medical_History_33 : Factor w/ 3 levels "1","2","3": 3 3 1 3 3 3 1 3 3 3 ...
##  $ Medical_History_34 : Factor w/ 3 levels "1","2","3": 3 3 3 1 3 3 3 3 3 1 ...
##  $ Medical_History_35 : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Medical_History_36 : Factor w/ 3 levels "1","2","3": 3 3 3 2 2 3 2 2 2 2 ...
##  $ Medical_History_37 : Factor w/ 2 levels "1","2": 2 2 2 2 2 2 2 2 2 2 ...
##  $ Medical_History_38 : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Medical_History_39 : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
##  $ Medical_History_40 : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
##  $ Medical_History_41 : Factor w/ 2 levels "1","3": 2 1 2 2 2 1 1 2 1 1 ...
```

```
summary(train.cat)
```

```
##  Product_Info_1 Product_Info_2  Product_Info_3  Product_Info_5
##  1:57816        D3     :14321   26     :50892   2:58968
##  2: 1565        D4     :10812   10     : 6008   3:  413
```

```
##                A8     : 6835   29     : 1334
##                D1     : 6554   31     :  411
##                D2     : 6286   15     :  251
##                E1     : 2647   37     :  139
##                (Other):11926   (Other):  346
##  Product_Info_6 Product_Info_7 Employment_Info_2 Employment_Info_3
##  1: 9691        1:58086        9      :34573     1:50447
##  3:49690        2:    2        1      : 8934     3: 8934
##                 3: 1293        12     : 7714
##                                14     : 4819
##                                3      : 1724
##                                15     :  480
##                                (Other): 1137
##  Employment_Info_5 InsuredInfo_1 InsuredInfo_2 InsuredInfo_3
##  2:50892           1:47223       2:58940       8      :18736
##  3: 8489           2:11886       3:  441       3      :16522
##                    3:  272                     6      :12820
##                                                11     : 4141
##                                                2      : 3828
##                                                4      : 1058
##                                                (Other): 2276
##  InsuredInfo_4 InsuredInfo_5 InsuredInfo_6 InsuredInfo_7
##  2: 6908       1:58574       1:35083       1:58237
##  3:52473       3:  807       2:24298       3: 1144
##
##
##
##
##
##  Insurance_History_1 Insurance_History_2 Insurance_History_3
##  1:16175             1:57724             1:25326
##  2:43206             2:    1             2:    1
##                      3: 1656             3:34054
##
##
##
##
##  Insurance_History_4 Insurance_History_7 Insurance_History_8
##  1:27832             1:31201             1:15561
##  2: 6169             2: 2799             2:25380
##  3:25380             3:25381             3:18440
##
##
##
##
##  Insurance_History_9 Family_Hist_1 Medical_History_2 Medical_History_3
##  1:  480             1:  538       112    :11027     1:    4
##  2:33519             2:17556       162    : 8343     2:53306
##  3:25382             3:41287       491    : 5797     3: 6071
##                                    335    : 3032
##                                    161    : 2942
##                                    16     : 2920
##                                    (Other):25320
##  Medical_History_4 Medical_History_5 Medical_History_6 Medical_History_7
```

```
##  1:20494           1:58946           1: 3268           1:  522
##  2:38887           2:  433           2:    2           2:57608
##                    3:    2           3:56111           3: 1251
##
##
##
##
##  Medical_History_8 Medical_History_9 Medical_History_10 Medical_History_11
##  1: 1269           1:13665           240    :  270      1:   88
##  2:54225           2:45712           0      :   75      2:  190
##  3: 3887           3:    4           1      :   22      3:59103
##                                      2      :   13
##                                      5      :   10
##                                      (Other):  167
##                                      NA's   :58824
##  Medical_History_12 Medical_History_13 Medical_History_14
##  1:    1            1: 6883            1:  256
##  2:56018            2:    2            2: 1356
##  3: 3362            3:52496            3:57769
##
##
##
##
##  Medical_History_16 Medical_History_17 Medical_History_18
##  1:49656            1:    1            1:56212
##  2:    1            2: 1304            2: 3159
##  3: 9724            3:58076            3:   10
##
##
##
##
##  Medical_History_19 Medical_History_20 Medical_History_21
##  1:57340            1:  887            1:52913
##  2: 2036            2:58493            2: 6464
##  3:    5            3:    1            3:    4
##
##
##
##
##  Medical_History_22 Medical_History_23 Medical_History_25
##  1: 1090            1:14010            1:48040
##  2:58291            2:    1            2:11105
##                     3:45370            3:  236
##
##
##
##
##  Medical_History_26 Medical_History_27 Medical_History_28
##  1:    3            1:  584            1:55393
##  2:11337            2:    7            2: 3985
##  3:48041            3:58790            3:    3
##
##
##
```

```
##
##  Medical_History_29 Medical_History_30 Medical_History_31
##  1:13576            1:    4            1:  437
##  2:    3            2:56952            2:    1
##  3:45802            3: 2425            3:58943
##
##
##
##
##  Medical_History_33 Medical_History_34 Medical_History_35
##  1: 5801            1: 9230            1:59319
##  3:53580            2:    3            2:    2
##                     3:50148            3:   60
##
##
##
##
##  Medical_History_36 Medical_History_37 Medical_History_38
##  1:  683            1: 3660            1:59093
##  2:47358            2:55719            2:  288
##  3:11340            3:    2
##
##
##
##
##  Medical_History_39 Medical_History_40 Medical_History_41
##  1: 5025            1:  961            1:40347
##  2:    2            2:    2            2:    1
##  3:54354            3:58418            3:19033
##
##
##
##
```

```
summary(test.cat)
```

```
##  Product_Info_1 Product_Info_2 Product_Info_3  Product_Info_5
##  1:19271        D3     :4432   26     :16819   2:19636
##  2:  494        D4     :3259   10     : 2115   3:  129
##                 A8     :2305   29     :  423
##                 D2     :2058   31     :  156
##                 D1     :2057   15     :   80
##                 A2     :1098   37     :   43
##                 (Other):4556   (Other):  129
##  Product_Info_6 Product_Info_7 Employment_Info_2 Employment_Info_3
##  1: 3402        1:19336        9      :11734     1:16548
##  3:16363        3:  429        1      : 3217     3: 3217
##                                12     : 2365
##                                14     : 1443
##                                3      :  551
##                                15     :  128
##                                (Other):  327
##  Employment_Info_5 InsuredInfo_1 InsuredInfo_2 InsuredInfo_3
##  2:16819           1:15822       2:19619       8      :9233
```

```
##  3: 2946           2: 3852       3:  146       3      :5182
##                    3:   91                     2      :3426
##                                                11     : 803
##                                                10     : 641
##                                                1      : 298
##                                                (Other): 182
##  InsuredInfo_4 InsuredInfo_5 InsuredInfo_6 InsuredInfo_7
##  2: 2122       1:19468       1:11445       1:19437
##  3:17643       3:  297       2: 8320       3:  328
##
##
##
##
##
##  Insurance_History_1 Insurance_History_2 Insurance_History_3
##  1: 5805             1:19277             1: 8070
##  2:13960             3:  488             3:11695
##
##
##
##
##
##  Insurance_History_4 Insurance_History_7 Insurance_History_8
##  1:9736              1:10237             1:5469
##  2:1940              2: 1437             2:8087
##  3:8089              3: 8091             3:6209
##
##
##
##
##  Insurance_History_9 Family_Hist_1 Medical_History_2 Medical_History_3
##  1:  212             1:  176       112    :3822      2:16853
##  2:11463             2: 5820       162    :2739      3: 2912
##  3: 8090             3:13769       491    :1878
##                                    161    :1004
##                                    335    : 985
##                                    16     : 984
##                                    (Other):8353
##  Medical_History_4 Medical_History_5 Medical_History_6 Medical_History_7
##  1: 6745           1:19617           1: 1115           1:  168
##  2:13020           2:  148           3:18650           2:19180
##                                                        3:  417
##
##
##
##
##  Medical_History_8 Medical_History_9 Medical_History_10 Medical_History_11
##  1:  421           1: 4431           240    :   93      1:   26
##  2:18051           2:15334           0      :   22      2:   63
##  3: 1293                             7      :    6      3:19676
##                                      176    :    6
##                                      1      :    4
##                                      (Other):   70
##                                      NA's   :19564
```

```
##  Medical_History_12 Medical_History_13 Medical_History_14
##  2:18704            1: 2305            1:   73
##  3: 1061            2:    2            2:  447
##                     3:17458            3:19245
##
##
##
##
##  Medical_History_16 Medical_History_17 Medical_History_18
##  1:16566            2:  440            1:18711
##  3: 3199            3:19325            2: 1050
##                                        3:    4
##
##
##
##
##  Medical_History_19 Medical_History_20 Medical_History_21
##  1:19087            1:  283            1:17640
##  2:  677            2:19481            2: 2124
##  3:    1            3:    1            3:    1
##
##
##
##
##  Medical_History_22 Medical_History_23 Medical_History_25
##  1:  308            1: 5086            1:15954
##  2:19457            3:14679            2: 3739
##                                        3:   72
##
##
##
##
##  Medical_History_26 Medical_History_27 Medical_History_28
##  1:    3            1:  186            1:18414
##  2: 3809            2:    2            2: 1348
##  3:15953            3:19577            3:    3
##
##
##
##
##  Medical_History_29 Medical_History_30 Medical_History_31
##  1: 4370            1:    2            1:  136
##  2:    1            2:18924            3:19629
##  3:15394            3:  839
##
##
##
##
##  Medical_History_33 Medical_History_34 Medical_History_35
##  1: 2001            1: 2930            1:19741
##  2:    1            2:    2            2:    1
##  3:17763            3:16833            3:   23
##
##
```

```
##
##
##  Medical_History_36 Medical_History_37 Medical_History_38
##  1:  228            1: 1284            1:19664
##  2:15724            2:18481            2:  100
##  3: 3813                               3:    1
##
##
##
##
##  Medical_History_39 Medical_History_40 Medical_History_41
##  1: 1718            1:  302            1:13316
##  2:    2            2:    2            3: 6449
##  3:18045            3:19461
##
##
##
##
```

Contrary to the suggested separation of the variables, it seems reasonable to use the variables Medical_History_2 and Medical_History_10 as continuous.

What are the dimensions of the datasets?

```
cat("Train data has", nrow(train), "rows and", ncol(train), "columns! \n")
```

```
## Train data has 59381 rows and 128 columns!
```

```
cat("Test data has", nrow(test), "rows and", ncol(test), "columns! \n")
```

```
## Test data has 19765 rows and 127 columns!
```

In the above structure commands we saw missing data, how much is it?

```
sum(is.na(train)) / (nrow(train) * ncol(train))
```

```
## [1] 0.05171885
```

```
sum(is.na(test)) / (nrow(test) * ncol(test))
```

```
## [1] 0.05205894
```

```
apply(train, 2, function(x) { sum(is.na(x)) })
```

```
##                  Id      Product_Info_1      Product_Info_2
##                   0                   0                   0
##      Product_Info_3      Product_Info_4      Product_Info_5
##                   0                   0                   0
##      Product_Info_6      Product_Info_7             Ins_Age
##                   0                   0                   0
##                  Ht                  Wt                 BMI
```

```
##                   0                   0                   0
##   Employment_Info_1   Employment_Info_2   Employment_Info_3
##                  19                   0                   0
##   Employment_Info_4   Employment_Info_5   Employment_Info_6
##                6779                   0               10854
##       InsuredInfo_1       InsuredInfo_2       InsuredInfo_3
##                   0                   0                   0
##       InsuredInfo_4       InsuredInfo_5       InsuredInfo_6
##                   0                   0                   0
##       InsuredInfo_7 Insurance_History_1 Insurance_History_2
##                   0                   0                   0
## Insurance_History_3 Insurance_History_4 Insurance_History_5
##                   0                   0               25396
## Insurance_History_7 Insurance_History_8 Insurance_History_9
##                   0                   0                   0
##       Family_Hist_1       Family_Hist_2       Family_Hist_3
##                   0               28656               34241
##       Family_Hist_4       Family_Hist_5   Medical_History_1
##               19184               41811                8889
##   Medical_History_2   Medical_History_3   Medical_History_4
##                   0                   0                   0
##   Medical_History_5   Medical_History_6   Medical_History_7
##                   0                   0                   0
##   Medical_History_8   Medical_History_9  Medical_History_10
##                   0                   0               58824
##  Medical_History_11  Medical_History_12  Medical_History_13
##                   0                   0                   0
##  Medical_History_14  Medical_History_15  Medical_History_16
##                   0               44596                   0
##  Medical_History_17  Medical_History_18  Medical_History_19
##                   0                   0                   0
##  Medical_History_20  Medical_History_21  Medical_History_22
##                   0                   0                   0
##  Medical_History_23  Medical_History_24  Medical_History_25
##                   0               55580                   0
##  Medical_History_26  Medical_History_27  Medical_History_28
##                   0                   0                   0
##  Medical_History_29  Medical_History_30  Medical_History_31
##                   0                   0                   0
##  Medical_History_32  Medical_History_33  Medical_History_34
##               58274                   0                   0
##  Medical_History_35  Medical_History_36  Medical_History_37
##                   0                   0                   0
##  Medical_History_38  Medical_History_39  Medical_History_40
##                   0                   0                   0
##  Medical_History_41   Medical_Keyword_1   Medical_Keyword_2
##                   0                   0                   0
##   Medical_Keyword_3   Medical_Keyword_4   Medical_Keyword_5
##                   0                   0                   0
##   Medical_Keyword_6   Medical_Keyword_7   Medical_Keyword_8
##                   0                   0                   0
##   Medical_Keyword_9  Medical_Keyword_10  Medical_Keyword_11
##                   0                   0                   0
##  Medical_Keyword_12  Medical_Keyword_13  Medical_Keyword_14
```

```
##                   0                   0                   0
##  Medical_Keyword_15  Medical_Keyword_16  Medical_Keyword_17
##                   0                   0                   0
##  Medical_Keyword_18  Medical_Keyword_19  Medical_Keyword_20
##                   0                   0                   0
##  Medical_Keyword_21  Medical_Keyword_22  Medical_Keyword_23
##                   0                   0                   0
##  Medical_Keyword_24  Medical_Keyword_25  Medical_Keyword_26
##                   0                   0                   0
##  Medical_Keyword_27  Medical_Keyword_28  Medical_Keyword_29
##                   0                   0                   0
##  Medical_Keyword_30  Medical_Keyword_31  Medical_Keyword_32
##                   0                   0                   0
##  Medical_Keyword_33  Medical_Keyword_34  Medical_Keyword_35
##                   0                   0                   0
##  Medical_Keyword_36  Medical_Keyword_37  Medical_Keyword_38
##                   0                   0                   0
##  Medical_Keyword_39  Medical_Keyword_40  Medical_Keyword_41
##                   0                   0                   0
##  Medical_Keyword_42  Medical_Keyword_43  Medical_Keyword_44
##                   0                   0                   0
##  Medical_Keyword_45  Medical_Keyword_46  Medical_Keyword_47
##                   0                   0                   0
##  Medical_Keyword_48            Response
##                   0                   0
```

```
apply(test, 2, function(x) { sum(is.na(x)) })
```

```
##                  Id      Product_Info_1      Product_Info_2
##                   0                   0                   0
##      Product_Info_3      Product_Info_4      Product_Info_5
##                   0                   0                   0
##      Product_Info_6      Product_Info_7             Ins_Age
##                   0                   0                   0
##                  Ht                  Wt                 BMI
##                   0                   0                   0
##   Employment_Info_1   Employment_Info_2   Employment_Info_3
##                   3                   0                   0
##   Employment_Info_4   Employment_Info_5   Employment_Info_6
##                2137                   0                3787
##       InsuredInfo_1       InsuredInfo_2       InsuredInfo_3
##                   0                   0                   0
##       InsuredInfo_4       InsuredInfo_5       InsuredInfo_6
##                   0                   0                   0
##       InsuredInfo_7 Insurance_History_1 Insurance_History_2
##                   0                   0                   0
## Insurance_History_3 Insurance_History_4 Insurance_History_5
##                   0                   0                8105
## Insurance_History_7 Insurance_History_8 Insurance_History_9
##                   0                   0                   0
##       Family_Hist_1       Family_Hist_2       Family_Hist_3
##                   0                9880               11064
##       Family_Hist_4       Family_Hist_5   Medical_History_1
##                6677               13624                2972
```

```
##   Medical_History_2   Medical_History_3   Medical_History_4
##                   0                   0                   0
##   Medical_History_5   Medical_History_6   Medical_History_7
##                   0                   0                   0
##   Medical_History_8   Medical_History_9  Medical_History_10
##                   0                   0               19564
##  Medical_History_11  Medical_History_12  Medical_History_13
##                   0                   0                   0
##  Medical_History_14  Medical_History_15  Medical_History_16
##                   0               14864                   0
##  Medical_History_17  Medical_History_18  Medical_History_19
##                   0                   0                   0
##  Medical_History_20  Medical_History_21  Medical_History_22
##                   0                   0                   0
##  Medical_History_23  Medical_History_24  Medical_History_25
##                   0               18585                   0
##  Medical_History_26  Medical_History_27  Medical_History_28
##                   0                   0                   0
##  Medical_History_29  Medical_History_30  Medical_History_31
##                   0                   0                   0
##  Medical_History_32  Medical_History_33  Medical_History_34
##               19414                   0                   0
##  Medical_History_35  Medical_History_36  Medical_History_37
##                   0                   0                   0
##  Medical_History_38  Medical_History_39  Medical_History_40
##                   0                   0                   0
##  Medical_History_41   Medical_Keyword_1   Medical_Keyword_2
##                   0                   0                   0
##   Medical_Keyword_3   Medical_Keyword_4   Medical_Keyword_5
##                   0                   0                   0
##   Medical_Keyword_6   Medical_Keyword_7   Medical_Keyword_8
##                   0                   0                   0
##   Medical_Keyword_9  Medical_Keyword_10  Medical_Keyword_11
##                   0                   0                   0
##  Medical_Keyword_12  Medical_Keyword_13  Medical_Keyword_14
##                   0                   0                   0
##  Medical_Keyword_15  Medical_Keyword_16  Medical_Keyword_17
##                   0                   0                   0
##  Medical_Keyword_18  Medical_Keyword_19  Medical_Keyword_20
##                   0                   0                   0
##  Medical_Keyword_21  Medical_Keyword_22  Medical_Keyword_23
##                   0                   0                   0
##  Medical_Keyword_24  Medical_Keyword_25  Medical_Keyword_26
##                   0                   0                   0
##  Medical_Keyword_27  Medical_Keyword_28  Medical_Keyword_29
##                   0                   0                   0
##  Medical_Keyword_30  Medical_Keyword_31  Medical_Keyword_32
##                   0                   0                   0
##  Medical_Keyword_33  Medical_Keyword_34  Medical_Keyword_35
##                   0                   0                   0
##  Medical_Keyword_36  Medical_Keyword_37  Medical_Keyword_38
##                   0                   0                   0
##  Medical_Keyword_39  Medical_Keyword_40  Medical_Keyword_41
##                   0                   0                   0
```

```
##  Medical_Keyword_42  Medical_Keyword_43  Medical_Keyword_44
##                   0                   0                   0
##  Medical_Keyword_45  Medical_Keyword_46  Medical_Keyword_47
##                   0                   0                   0
##  Medical_Keyword_48
##                   0
```

Can we see any different missing data structure depending on the response?

```
train.na.per.response <- sapply(sort(unique(train$Response)), function(x) { apply(train[train$Response =
train.na.per.response
```

```
##                      [,1] [,2] [,3] [,4] [,5]  [,6] [,7]  [,8]
## Id                      0    0    0    0    0     0    0     0
## Product_Info_1          0    0    0    0    0     0    0     0
## Product_Info_2          0    0    0    0    0     0    0     0
## Product_Info_3          0    0    0    0    0     0    0     0
## Product_Info_4          0    0    0    0    0     0    0     0
## Product_Info_5          0    0    0    0    0     0    0     0
## Product_Info_6          0    0    0    0    0     0    0     0
## Product_Info_7          0    0    0    0    0     0    0     0
## Ins_Age                 0    0    0    0    0     0    0     0
## Ht                      0    0    0    0    0     0    0     0
## Wt                      0    0    0    0    0     0    0     0
## BMI                     0    0    0    0    0     0    0     0
## Employment_Info_1      16    0    0    0    0     3    0     0
## Employment_Info_2       0    0    0    0    0     0    0     0
## Employment_Info_3       0    0    0    0    0     0    0     0
## Employment_Info_4     789  679   85  148  560  1752  883  1883
## Employment_Info_5       0    0    0    0    0     0    0     0
## Employment_Info_6    1171 1194  220  333 1221  1688 1365  3662
## InsuredInfo_1           0    0    0    0    0     0    0     0
## InsuredInfo_2           0    0    0    0    0     0    0     0
## InsuredInfo_3           0    0    0    0    0     0    0     0
## InsuredInfo_4           0    0    0    0    0     0    0     0
## InsuredInfo_5           0    0    0    0    0     0    0     0
## InsuredInfo_6           0    0    0    0    0     0    0     0
## InsuredInfo_7           0    0    0    0    0     0    0     0
## Insurance_History_1     0    0    0    0    0     0    0     0
## Insurance_History_2     0    0    0    0    0     0    0     0
## Insurance_History_3     0    0    0    0    0     0    0     0
## Insurance_History_4     0    0    0    0    0     0    0     0
## Insurance_History_5  2848 2710  565  725 2372  4369 2901  8906
## Insurance_History_7     0    0    0    0    0     0    0     0
## Insurance_History_8     0    0    0    0    0     0    0     0
## Insurance_History_9     0    0    0    0    0     0    0     0
## Family_Hist_1           0    0    0    0    0     0    0     0
## Family_Hist_2        4049 3891  479  551 2840  5628 4354  6864
## Family_Hist_3        2709 3101  611  971 3060  6170 4073 13546
## Family_Hist_4        3099 2744  303  355 1929  3679 2994  4081
## Family_Hist_5        3373 4017  745 1119 3701  7835 5221 15800
## Medical_History_1     694  694  231  311  796  1744  998  3421
## Medical_History_2       0    0    0    0    0     0    0     0
```

```
## Medical_History_3      0    0    0    0    0     0    0     0
## Medical_History_4      0    0    0    0    0     0    0     0
## Medical_History_5      0    0    0    0    0     0    0     0
## Medical_History_6      0    0    0    0    0     0    0     0
## Medical_History_7      0    0    0    0    0     0    0     0
## Medical_History_8      0    0    0    0    0     0    0     0
## Medical_History_9      0    0    0    0    0     0    0     0
## Medical_History_10  6106 6468 1003 1419 5379 11042 7980 19427
## Medical_History_11     0    0    0    0    0     0    0     0
## Medical_History_12     0    0    0    0    0     0    0     0
## Medical_History_13     0    0    0    0    0     0    0     0
## Medical_History_14     0    0    0    0    0     0    0     0
## Medical_History_15  3981 4518  111  154 4241  8147 6428 17016
## Medical_History_16     0    0    0    0    0     0    0     0
## Medical_History_17     0    0    0    0    0     0    0     0
## Medical_History_18     0    0    0    0    0     0    0     0
## Medical_History_19     0    0    0    0    0     0    0     0
## Medical_History_20     0    0    0    0    0     0    0     0
## Medical_History_21     0    0    0    0    0     0    0     0
## Medical_History_22     0    0    0    0    0     0    0     0
## Medical_History_23     0    0    0    0    0     0    0     0
## Medical_History_24  5755 6018  928 1327 5078 10047 7580 18847
## Medical_History_25     0    0    0    0    0     0    0     0
## Medical_History_26     0    0    0    0    0     0    0     0
## Medical_History_27     0    0    0    0    0     0    0     0
## Medical_History_28     0    0    0    0    0     0    0     0
## Medical_History_29     0    0    0    0    0     0    0     0
## Medical_History_30     0    0    0    0    0     0    0     0
## Medical_History_31     0    0    0    0    0     0    0     0
## Medical_History_32  6059 6413  957 1352 5338 10690 7999 19466
## Medical_History_33     0    0    0    0    0     0    0     0
## Medical_History_34     0    0    0    0    0     0    0     0
## Medical_History_35     0    0    0    0    0     0    0     0
## Medical_History_36     0    0    0    0    0     0    0     0
## Medical_History_37     0    0    0    0    0     0    0     0
## Medical_History_38     0    0    0    0    0     0    0     0
## Medical_History_39     0    0    0    0    0     0    0     0
## Medical_History_40     0    0    0    0    0     0    0     0
## Medical_History_41     0    0    0    0    0     0    0     0
## Medical_Keyword_1      0    0    0    0    0     0    0     0
## Medical_Keyword_2      0    0    0    0    0     0    0     0
## Medical_Keyword_3      0    0    0    0    0     0    0     0
## Medical_Keyword_4      0    0    0    0    0     0    0     0
## Medical_Keyword_5      0    0    0    0    0     0    0     0
## Medical_Keyword_6      0    0    0    0    0     0    0     0
## Medical_Keyword_7      0    0    0    0    0     0    0     0
## Medical_Keyword_8      0    0    0    0    0     0    0     0
## Medical_Keyword_9      0    0    0    0    0     0    0     0
## Medical_Keyword_10     0    0    0    0    0     0    0     0
## Medical_Keyword_11     0    0    0    0    0     0    0     0
## Medical_Keyword_12     0    0    0    0    0     0    0     0
## Medical_Keyword_13     0    0    0    0    0     0    0     0
## Medical_Keyword_14     0    0    0    0    0     0    0     0
## Medical_Keyword_15     0    0    0    0    0     0    0     0
```

```
## Medical_Keyword_16     0    0    0    0    0     0    0     0
## Medical_Keyword_17     0    0    0    0    0     0    0     0
## Medical_Keyword_18     0    0    0    0    0     0    0     0
## Medical_Keyword_19     0    0    0    0    0     0    0     0
## Medical_Keyword_20     0    0    0    0    0     0    0     0
## Medical_Keyword_21     0    0    0    0    0     0    0     0
## Medical_Keyword_22     0    0    0    0    0     0    0     0
## Medical_Keyword_23     0    0    0    0    0     0    0     0
## Medical_Keyword_24     0    0    0    0    0     0    0     0
## Medical_Keyword_25     0    0    0    0    0     0    0     0
## Medical_Keyword_26     0    0    0    0    0     0    0     0
## Medical_Keyword_27     0    0    0    0    0     0    0     0
## Medical_Keyword_28     0    0    0    0    0     0    0     0
## Medical_Keyword_29     0    0    0    0    0     0    0     0
## Medical_Keyword_30     0    0    0    0    0     0    0     0
## Medical_Keyword_31     0    0    0    0    0     0    0     0
## Medical_Keyword_32     0    0    0    0    0     0    0     0
## Medical_Keyword_33     0    0    0    0    0     0    0     0
## Medical_Keyword_34     0    0    0    0    0     0    0     0
## Medical_Keyword_35     0    0    0    0    0     0    0     0
## Medical_Keyword_36     0    0    0    0    0     0    0     0
## Medical_Keyword_37     0    0    0    0    0     0    0     0
## Medical_Keyword_38     0    0    0    0    0     0    0     0
## Medical_Keyword_39     0    0    0    0    0     0    0     0
## Medical_Keyword_40     0    0    0    0    0     0    0     0
## Medical_Keyword_41     0    0    0    0    0     0    0     0
## Medical_Keyword_42     0    0    0    0    0     0    0     0
## Medical_Keyword_43     0    0    0    0    0     0    0     0
## Medical_Keyword_44     0    0    0    0    0     0    0     0
## Medical_Keyword_45     0    0    0    0    0     0    0     0
## Medical_Keyword_46     0    0    0    0    0     0    0     0
## Medical_Keyword_47     0    0    0    0    0     0    0     0
## Medical_Keyword_48     0    0    0    0    0     0    0     0
## Response               0    0    0    0    0     0    0     0
```

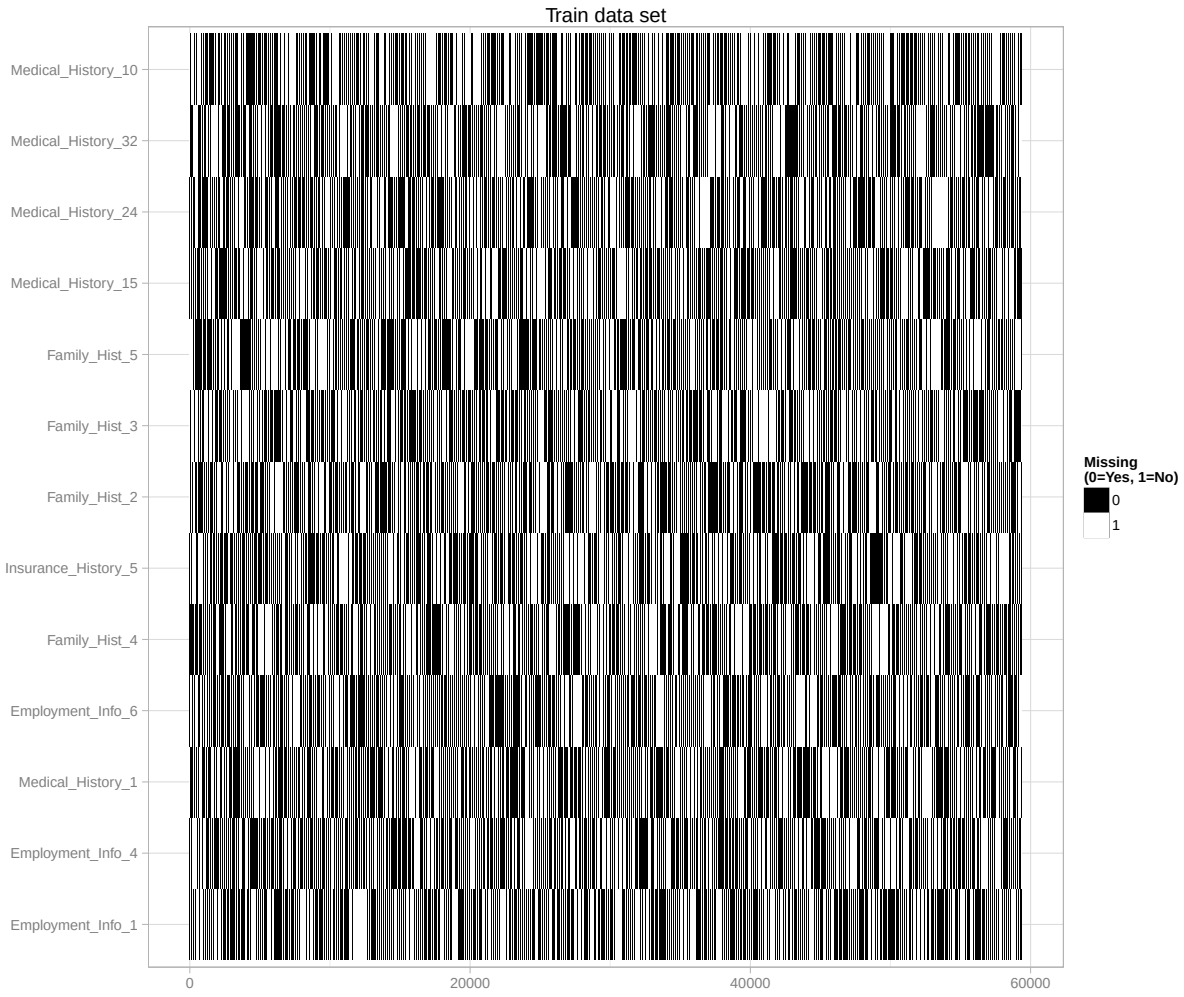
```
round(colSums(train.na.per.response) / sum(train.na.per.response), digits=4)
```

```
## [1] 0.1034 0.1080 0.0159 0.0223 0.0929 0.1852 0.1343 0.3381
```

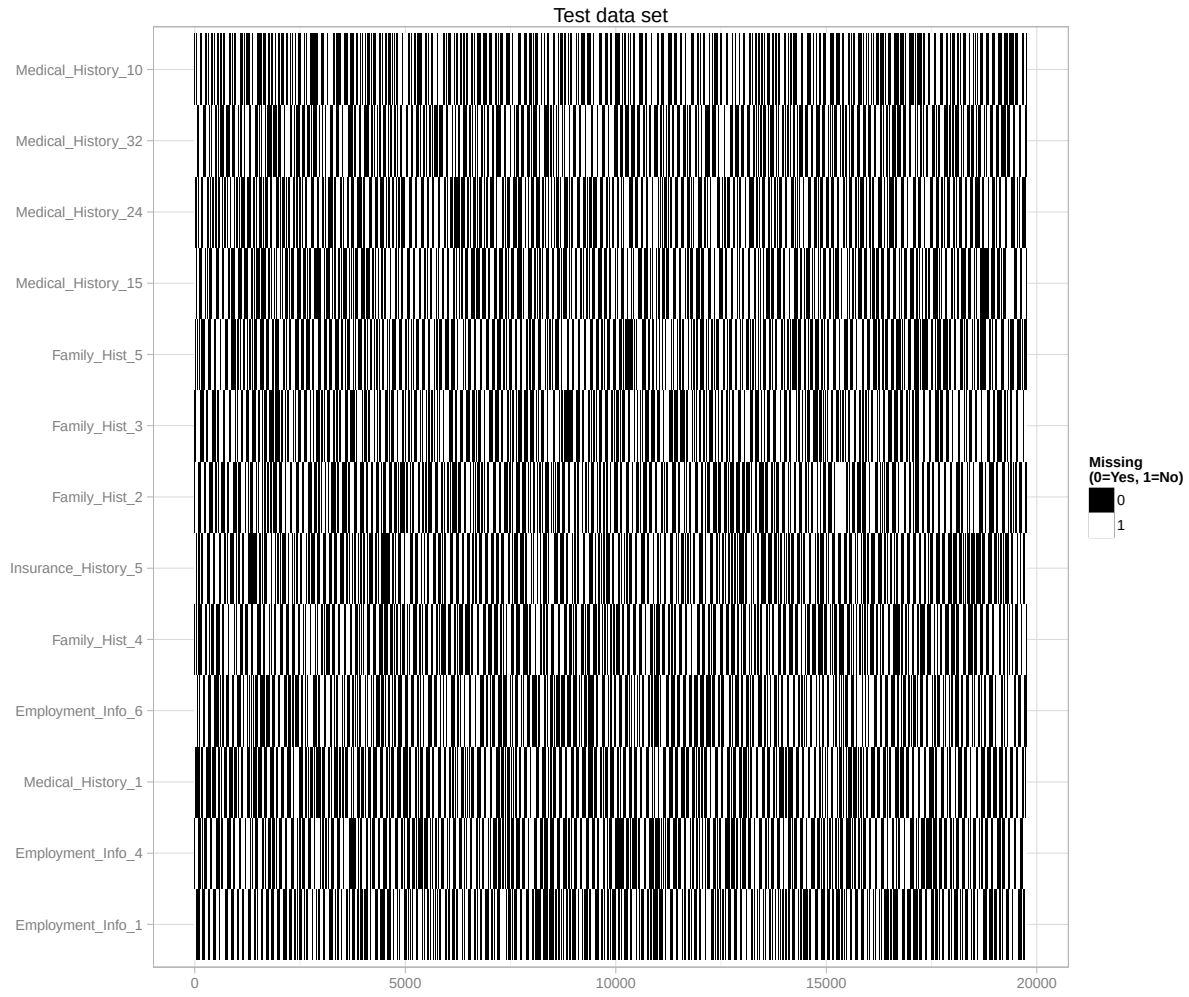
Data with response equal 8 has the most and response equal 3 the least missing data.

Plot the missingness structure (only for those features with missing data) where the feature with the most missing data is on top and with the least missing data on bottom.

```
plotMissingnessStr <- function(data.in, title=NULL) {
  r <- as.data.frame(ifelse(is.na(data.in), 0, 1))
  r <- r[,order(colMeans(is.na(data.in)))]
  dat.tmp <- expand.grid(list(x=1:nrow(r), y=colnames(r)))
  dat.tmp$r <- as.vector(t(r))

  ggplot(dat.tmp) + geom_tile(aes(x=x, y=y, fill=factor(r))) + scale_fill_manual(values=c("black", "whi
}

plotMissingnessStr(data.in=train[,apply(train, 2, function(x) { sum(is.na(x)) > 0 })], title="Train data
```




```
plotMissingnessStr(data.in=test[,apply(test, 2, function(x) { sum(is.na(x)) > 0 })], title="Test data s
```



Are there any duplicate rows?

```
cat("Train data set - Number of duplicated rows:", nrow(train) - nrow(unique(train)), "\n")
```

```
## Train data set - Number of duplicated rows: 0
```

```
cat("Test data set - Number of duplicated rows:", nrow(test) - nrow(unique(test)), "\n")
```

```
## Test data set - Number of duplicated rows: 0
```

Are there any constant columns?

```
train.const <- sapply(train, function(x) { length(unique(x)) == 1 })
test.const <- sapply(test, function(x) { length(unique(x)) == 1 })
cat("Train data set - Number of constant columns:", sum(train.const), "\n")
```

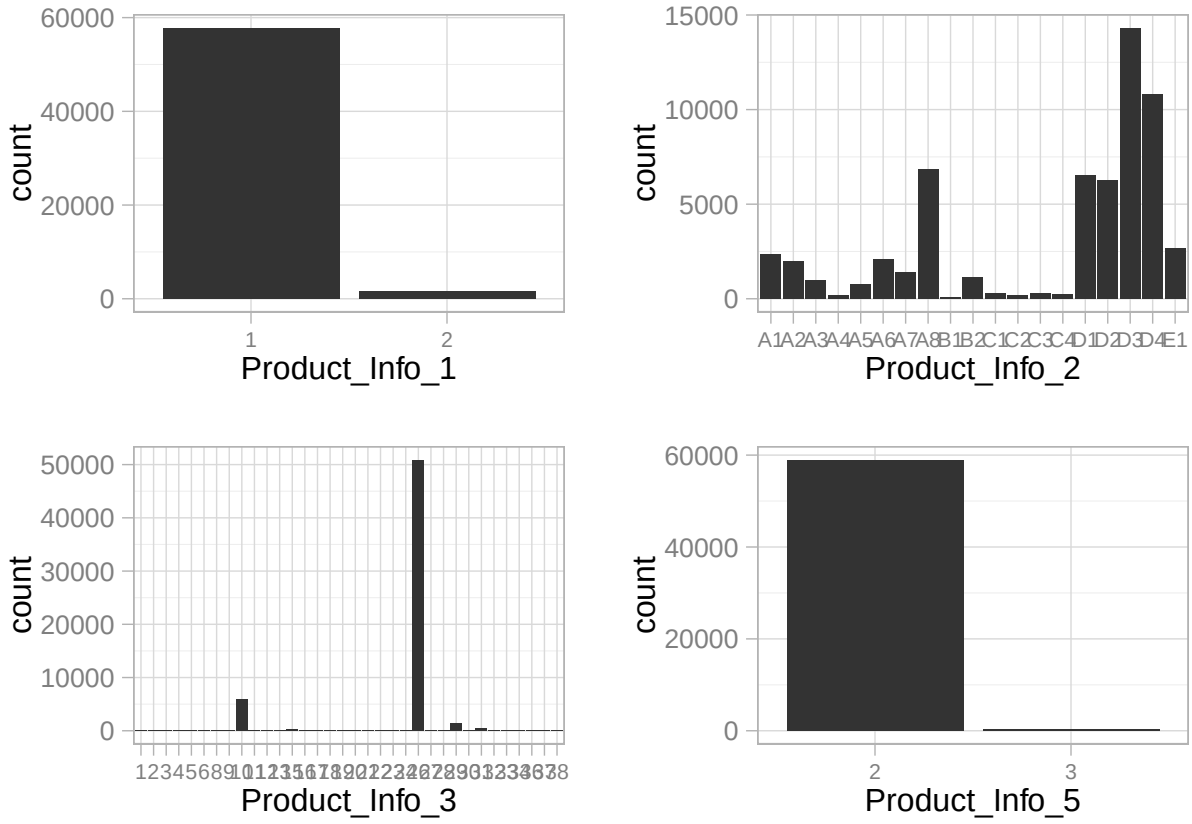
```
## Train data set - Number of constant columns: 0
```

```
cat("Test data set - Number of constant columns:", sum(test.const), "\n")
```

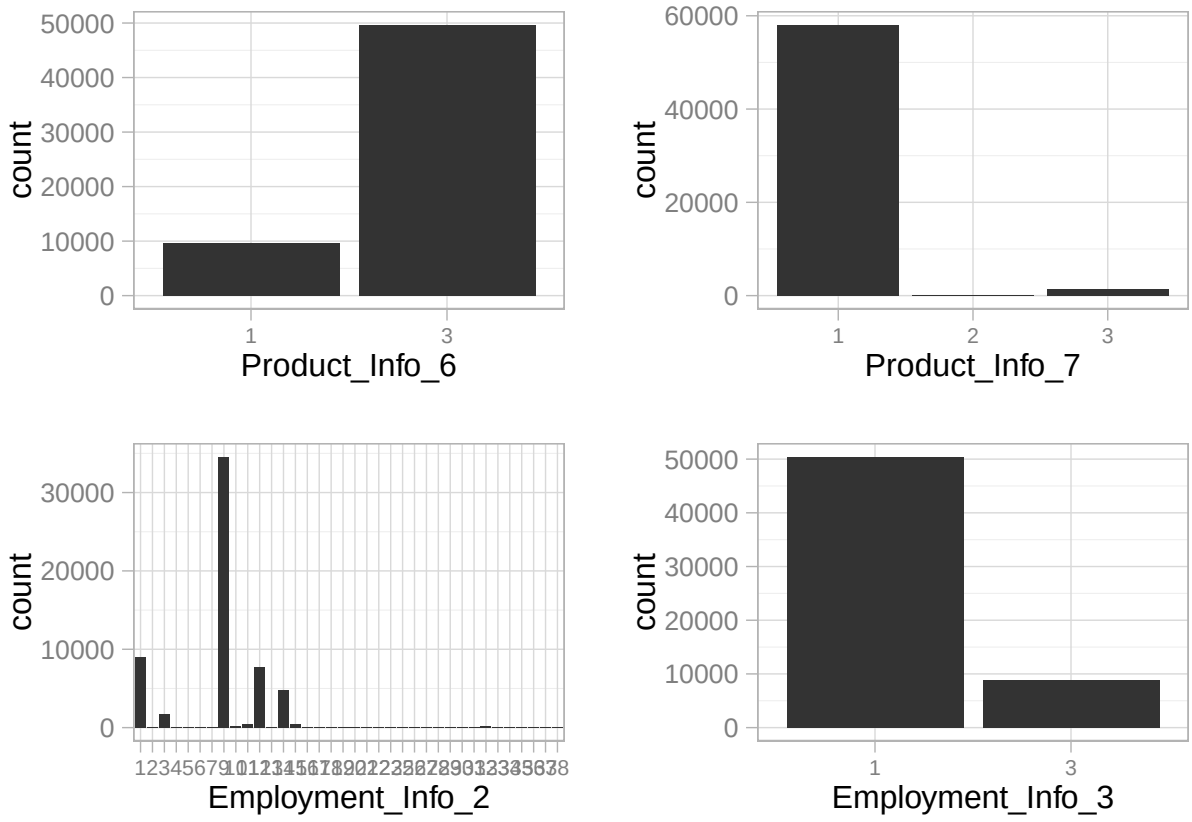
```
## Test data set - Number of constant columns: 0
```

Plot histograms of categorical variables

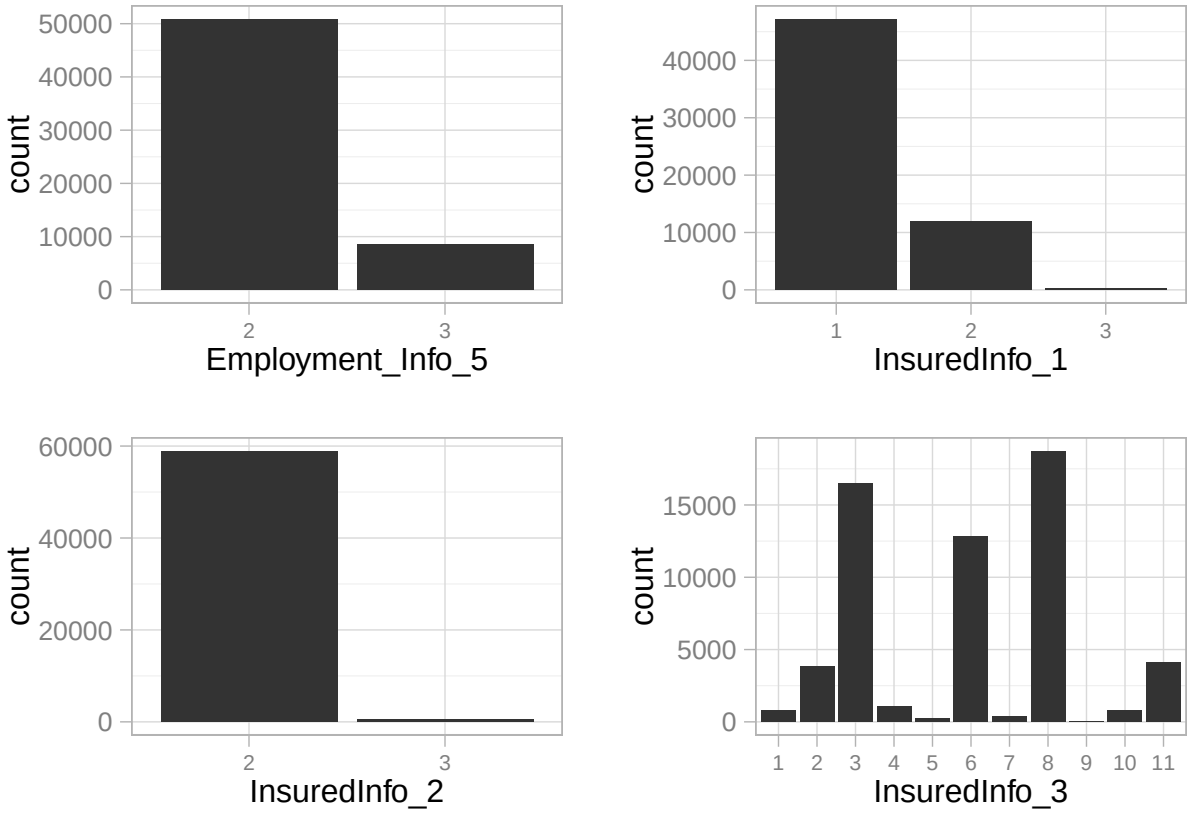
```
plotHist <- function(data.in, i) {
  data <- data.frame(x=data.in[,i])
  p <- ggplot(data=data, aes(x=factor(x))) + geom_histogram() + xlab(colnames(data.in)[i]) + theme_ligh
    theme(axis.text.x=element_text(size=8))
  return (p)
}

doPlots <- function(data.in, fun, ii, ncol=3) {
  pp <- list()
  for (i in ii) {
    p <- fun(data.in=data.in, i=i)
    pp <- c(pp, list(p))
  }
  do.call("grid.arrange", c(pp, ncol=ncol))
}

doPlots(data.in=train.cat, fun=plotHist, ii=1:4, ncol=2)
```



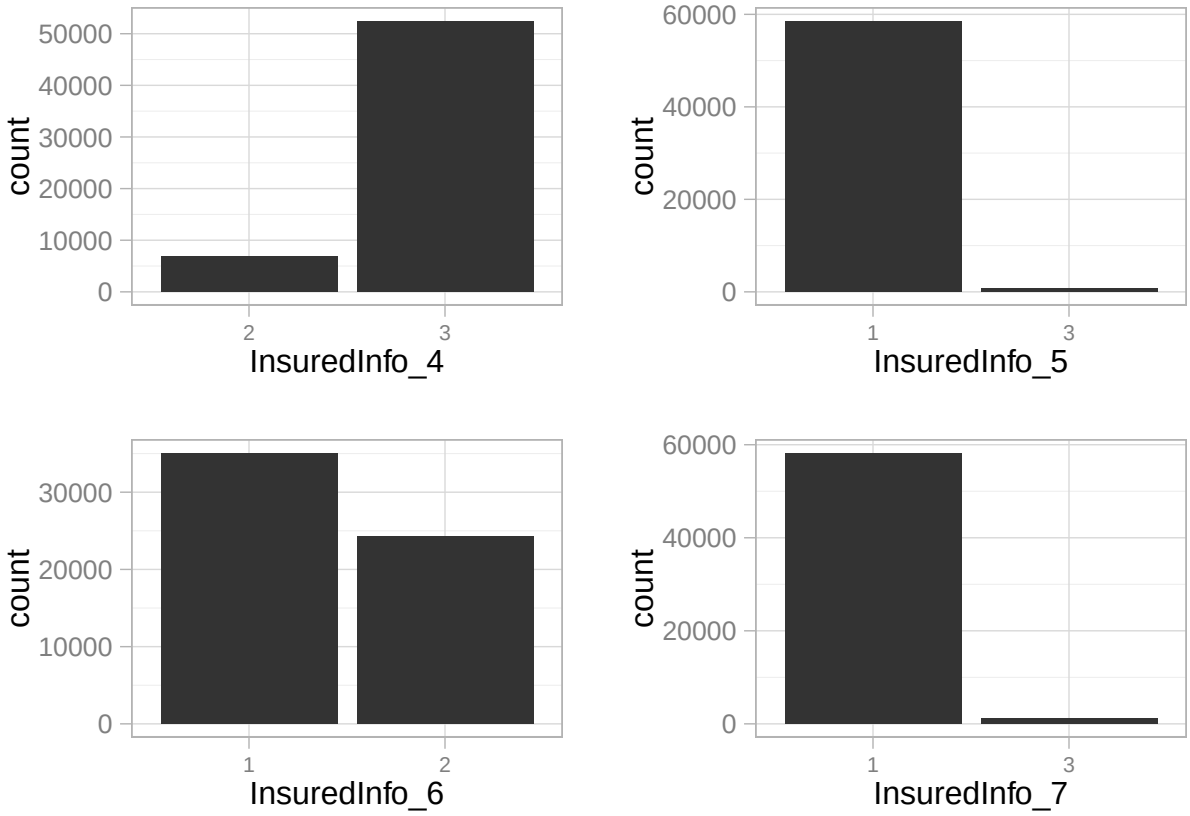
```
doPlots(data.in=train.cat, fun=plotHist, ii=5:8, ncol=2)
```



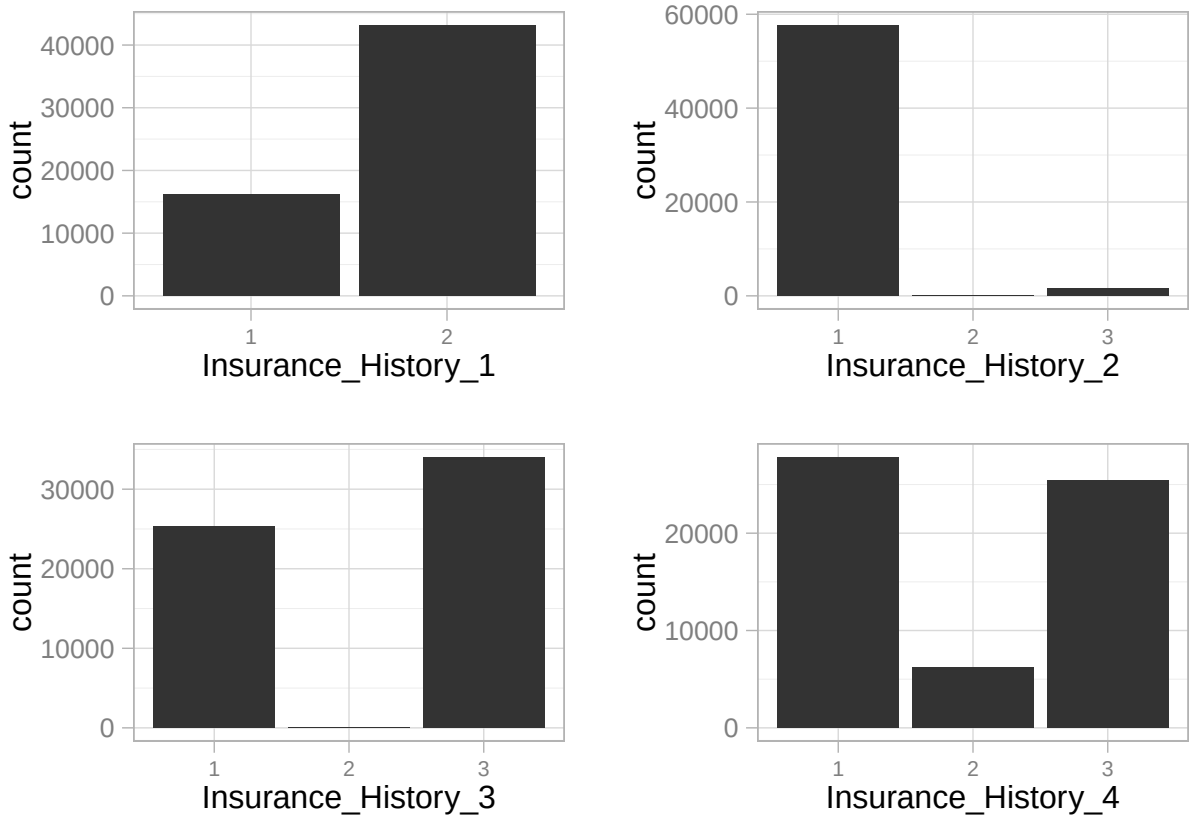

```
doPlots(data.in=train.cat, fun=plotHist, ii=9:12, ncol=2)
```



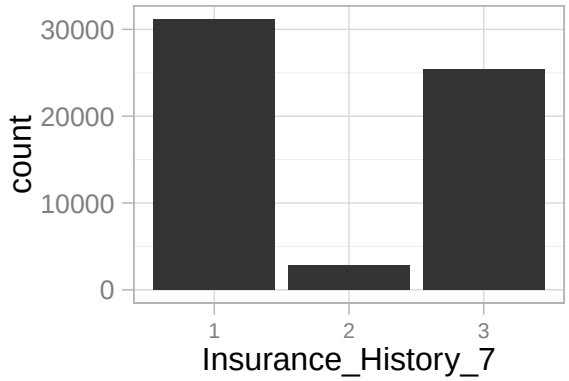

```
doPlots(data.in=train.cat, fun=plotHist, ii=13:16, ncol=2)
```

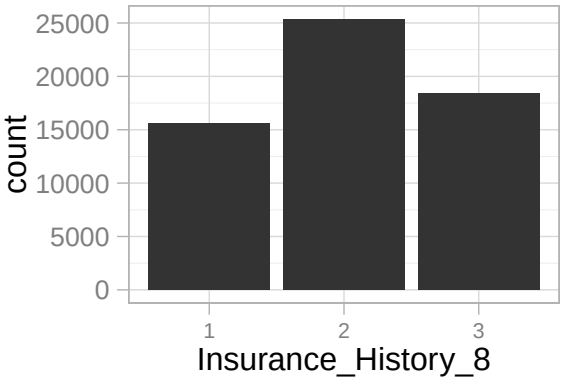


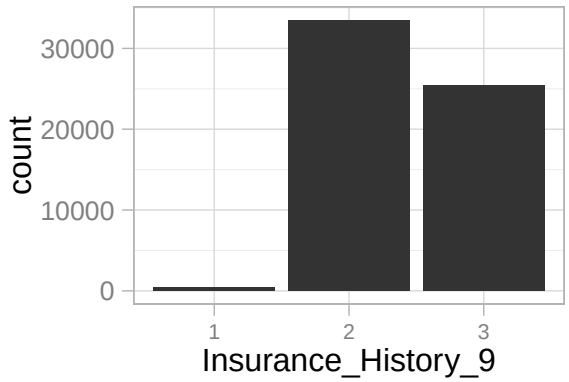

```
doPlots(data.in=train.cat, fun=plotHist, ii=17:20, ncol=2)
```

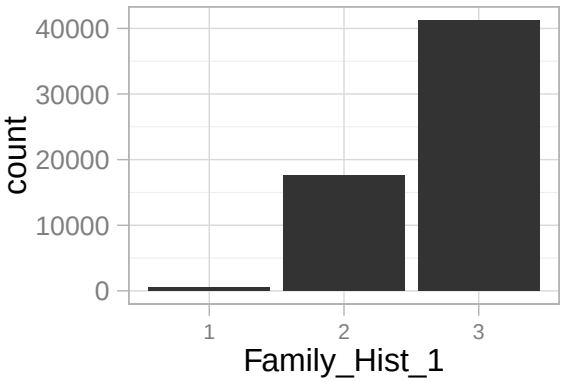



```
doPlots(data.in=train.cat, fun=plotHist, ii=21:24, ncol=2)
```

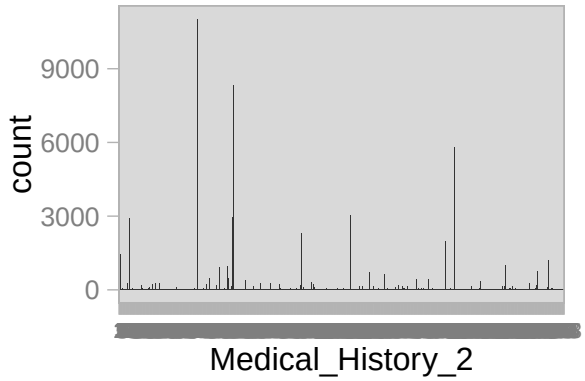


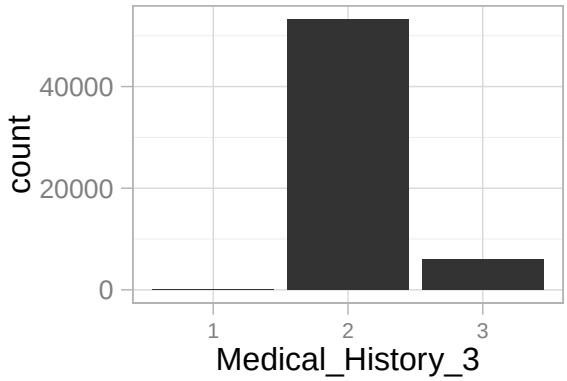


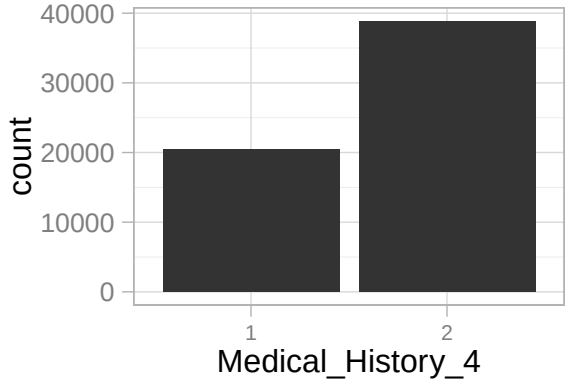


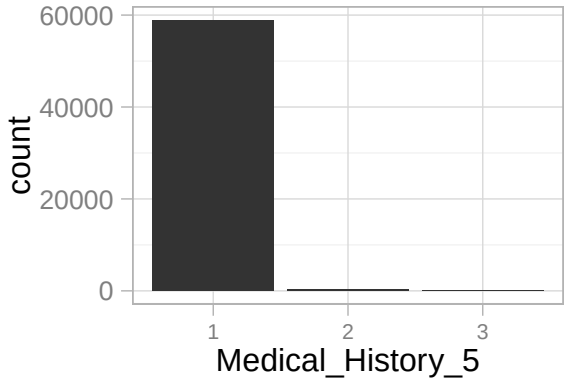



```
doPlots(data.in=train.cat, fun=plotHist, ii=25:28, ncol=2)
```

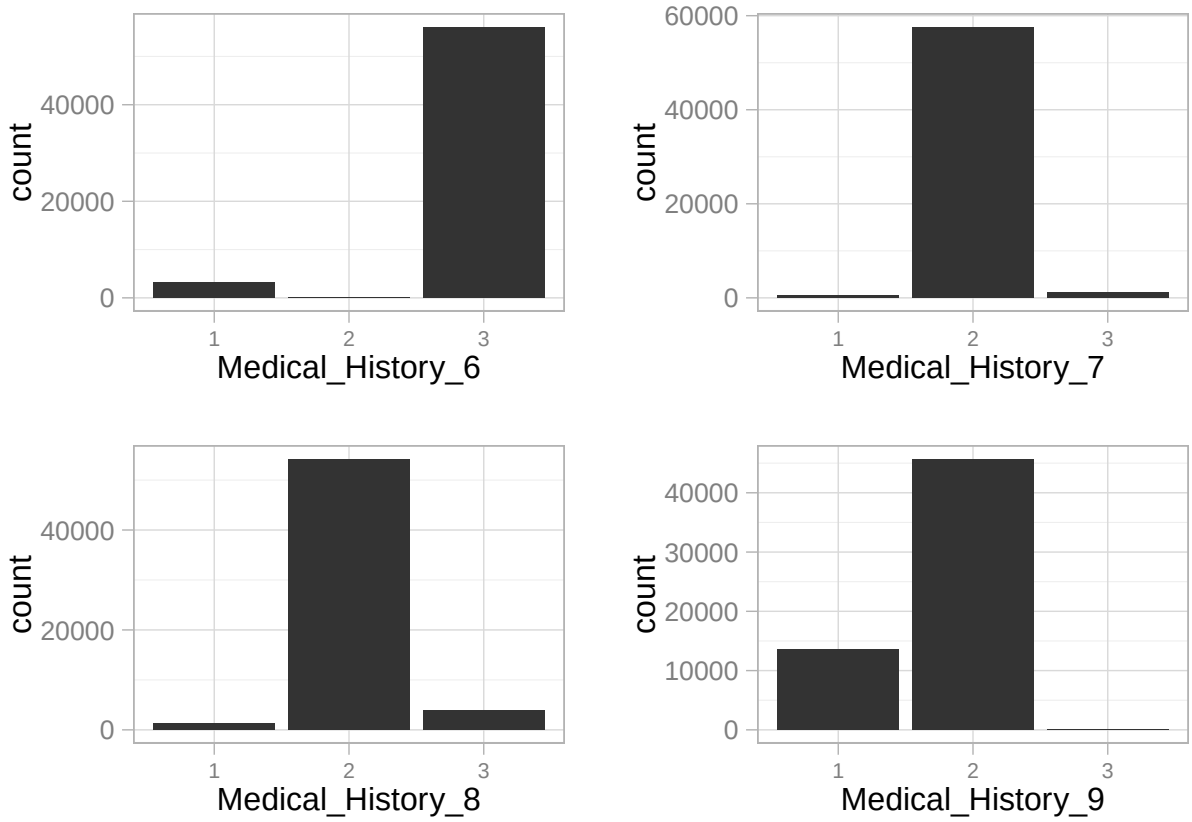




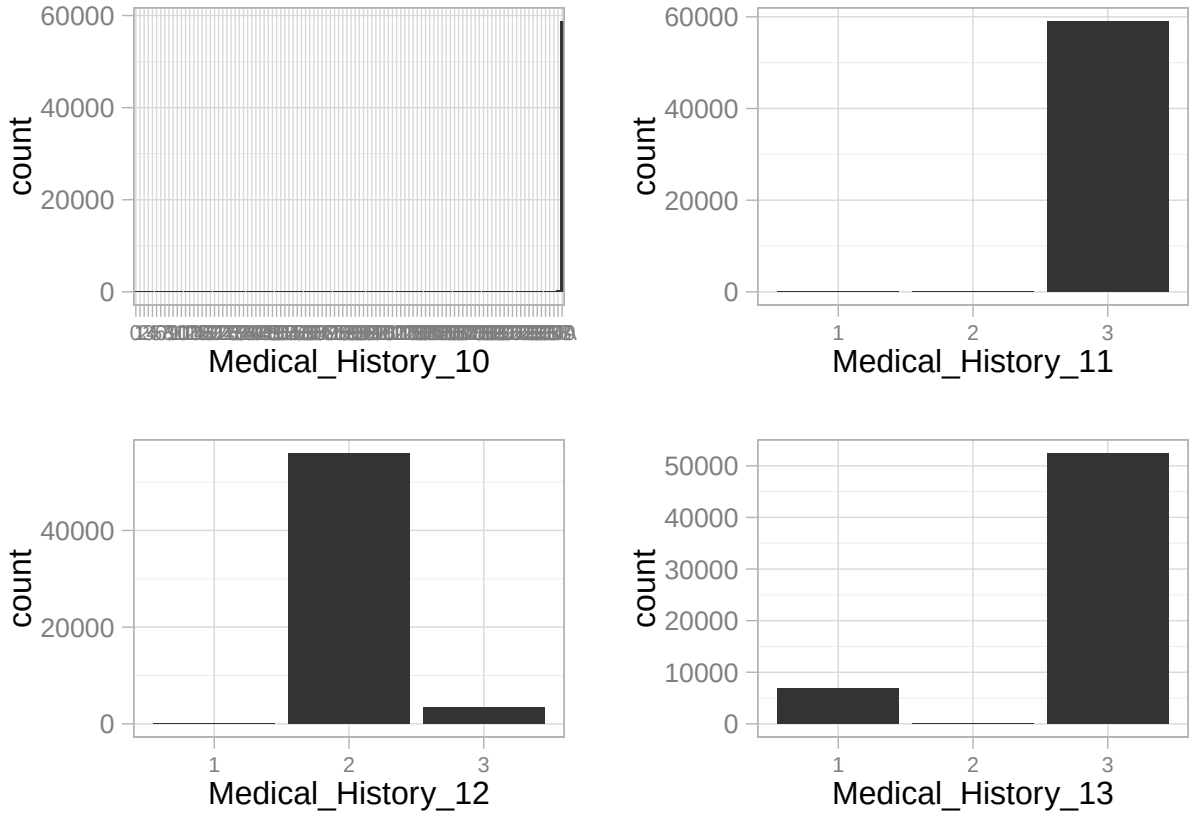




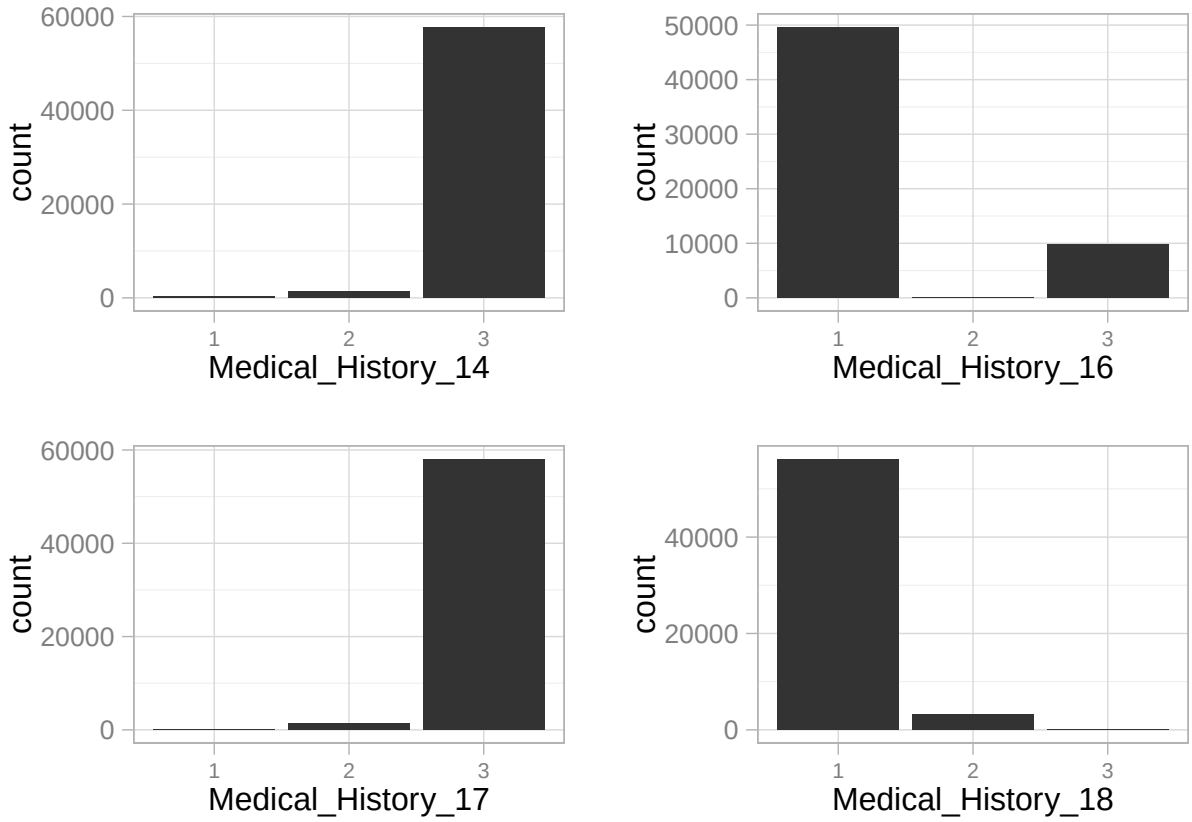

```
doPlots(data.in=train.cat, fun=plotHist, ii=29:32, ncol=2)
```



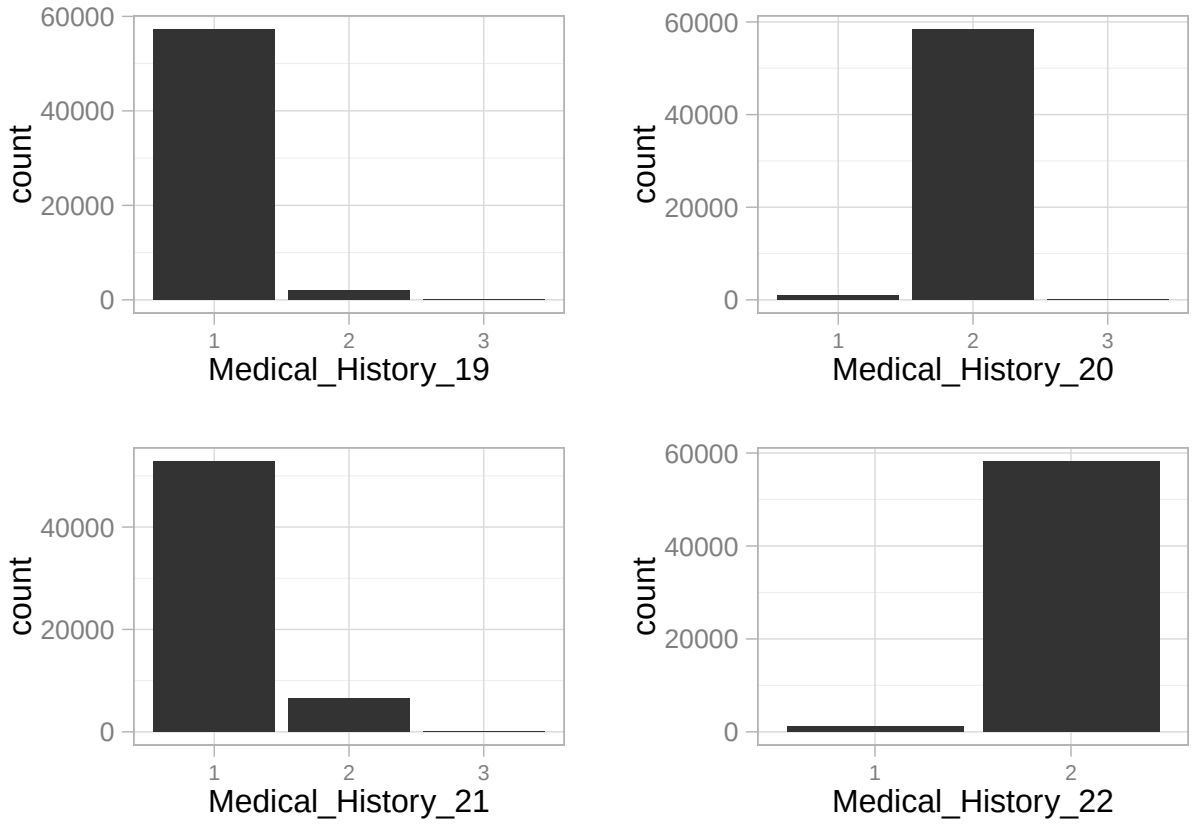

```
doPlots(data.in=train.cat, fun=plotHist, ii=33:36, ncol=2)
```



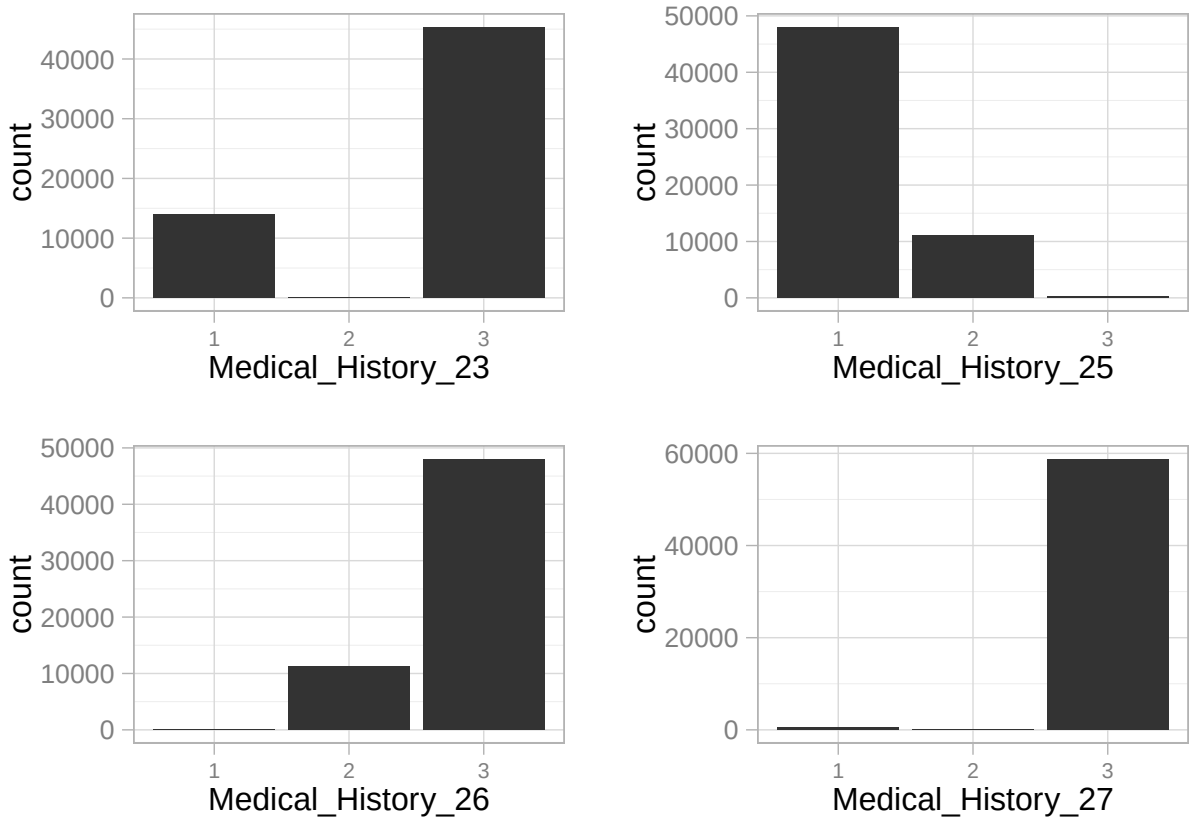

```
doPlots(data.in=train.cat, fun=plotHist, ii=37:40, ncol=2)
```



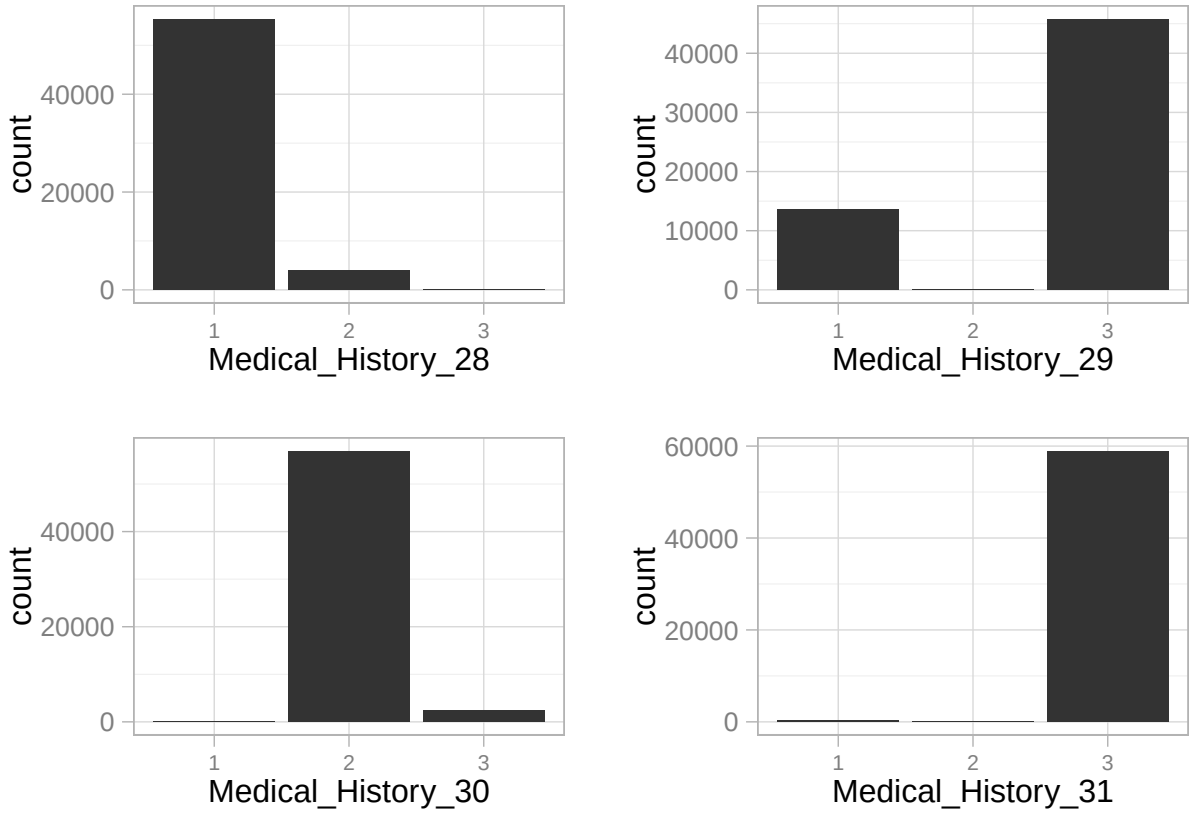

```
doPlots(data.in=train.cat, fun=plotHist, ii=41:44, ncol=2)
```



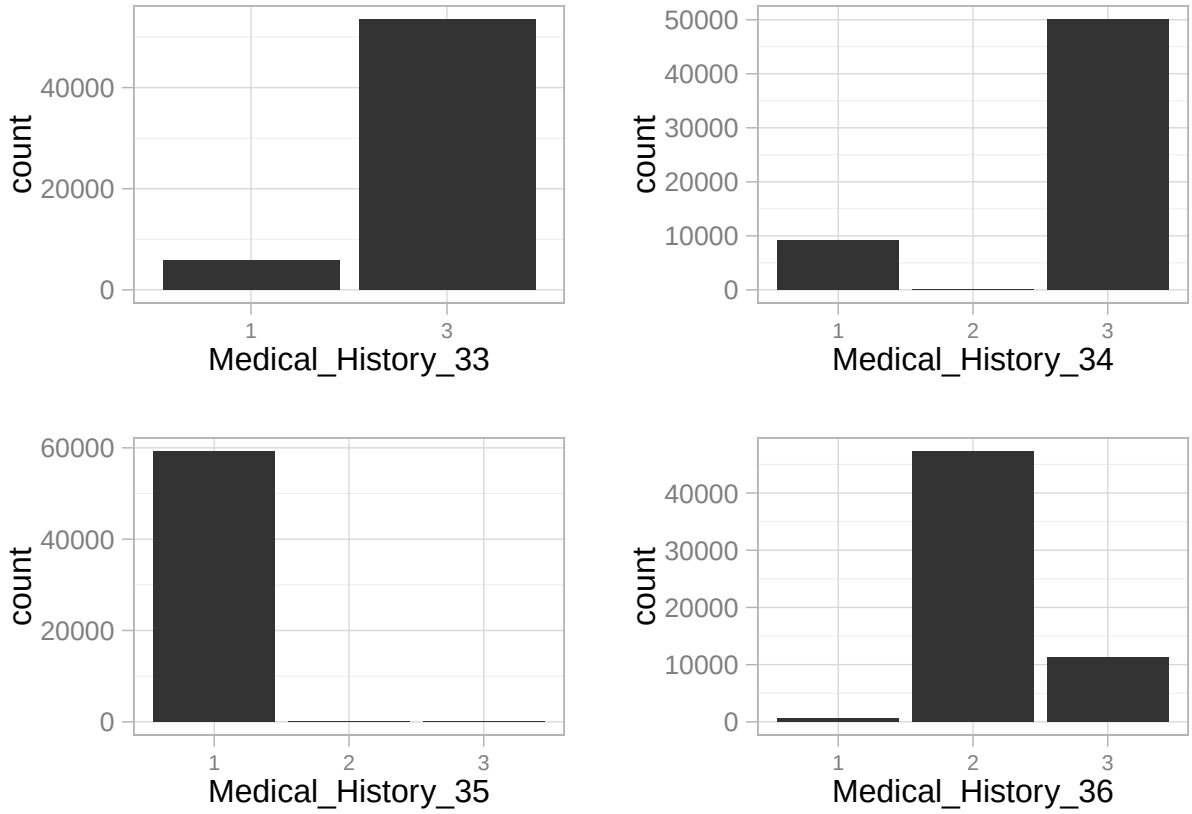

```
doPlots(data.in=train.cat, fun=plotHist, ii=45:48, ncol=2)
```



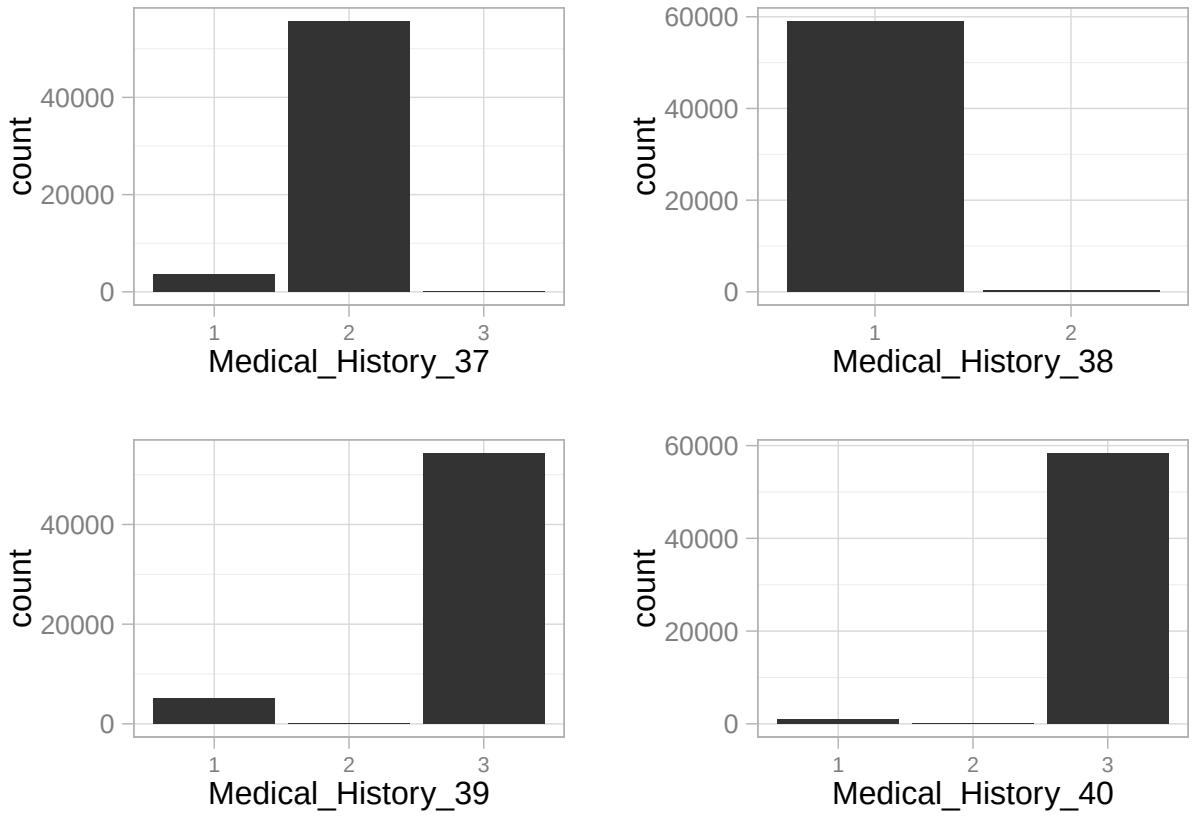

```
doPlots(data.in=train.cat, fun=plotHist, ii=49:52, ncol=2)
```



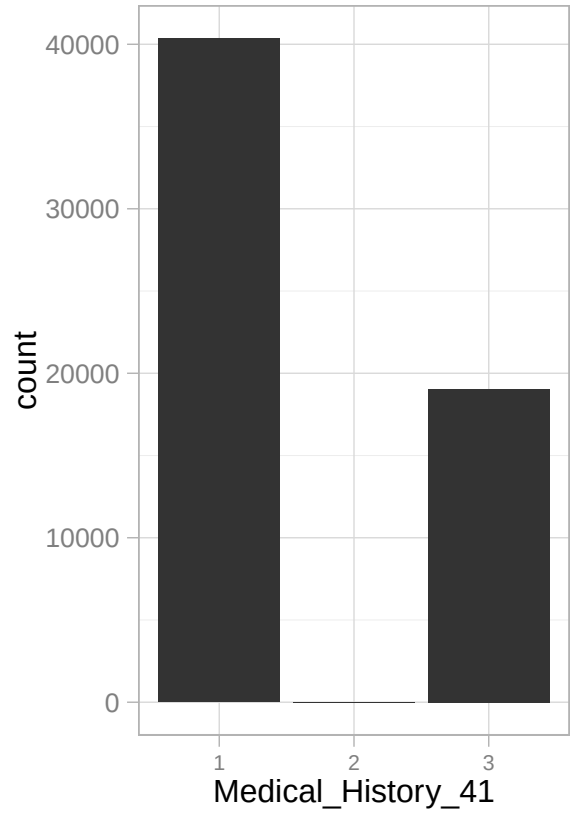

```
doPlots(data.in=train.cat, fun=plotHist, ii=53:56, ncol=2)
```




```
doPlots(data.in=train.cat, fun=plotHist, ii=57:60, ncol=2)
```

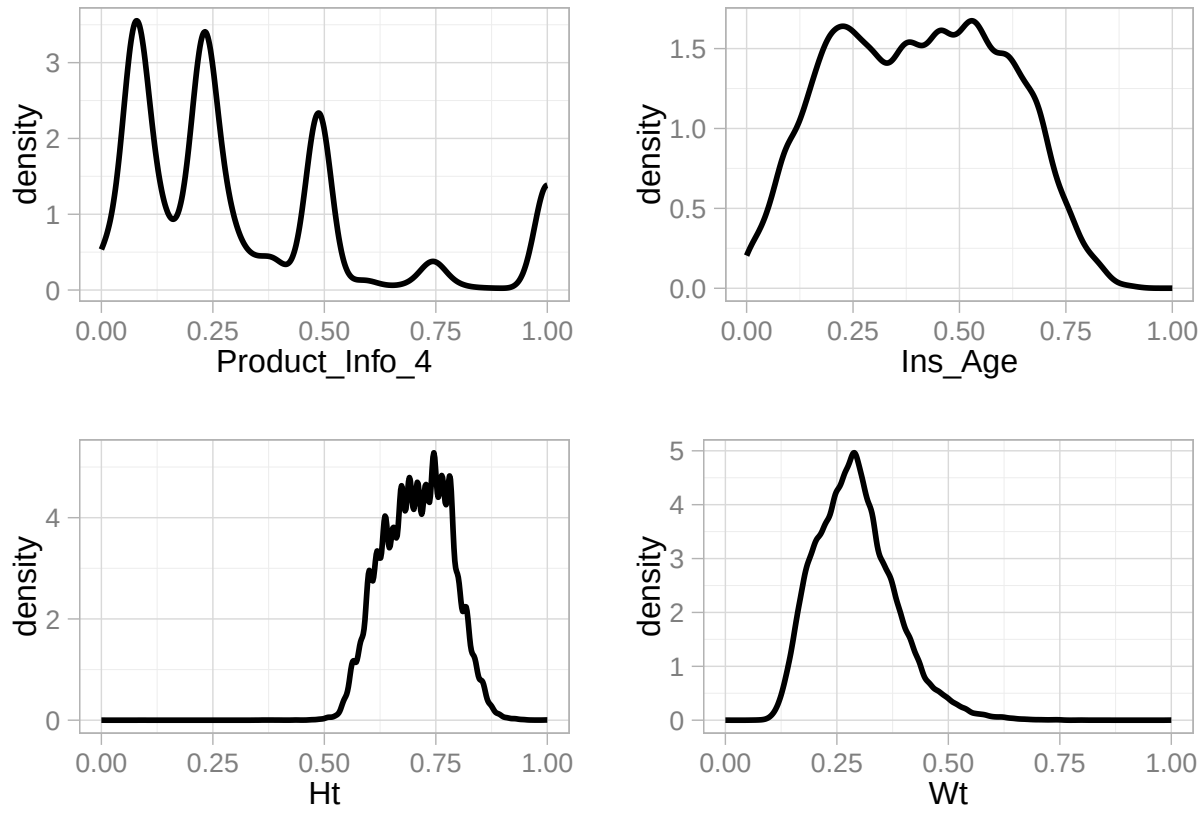



```
doPlots(data.in=train.cat, fun=plotHist, ii=61, ncol=2)
```

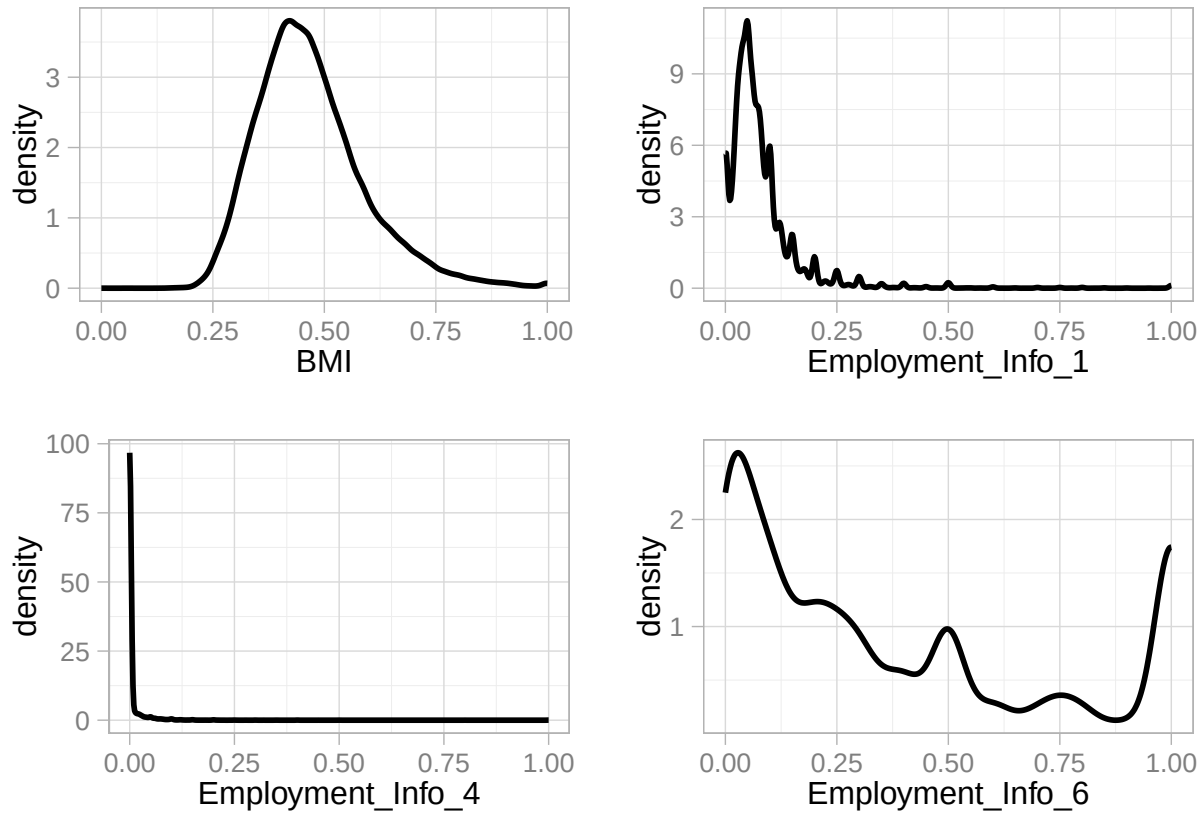


Densities of continous Features

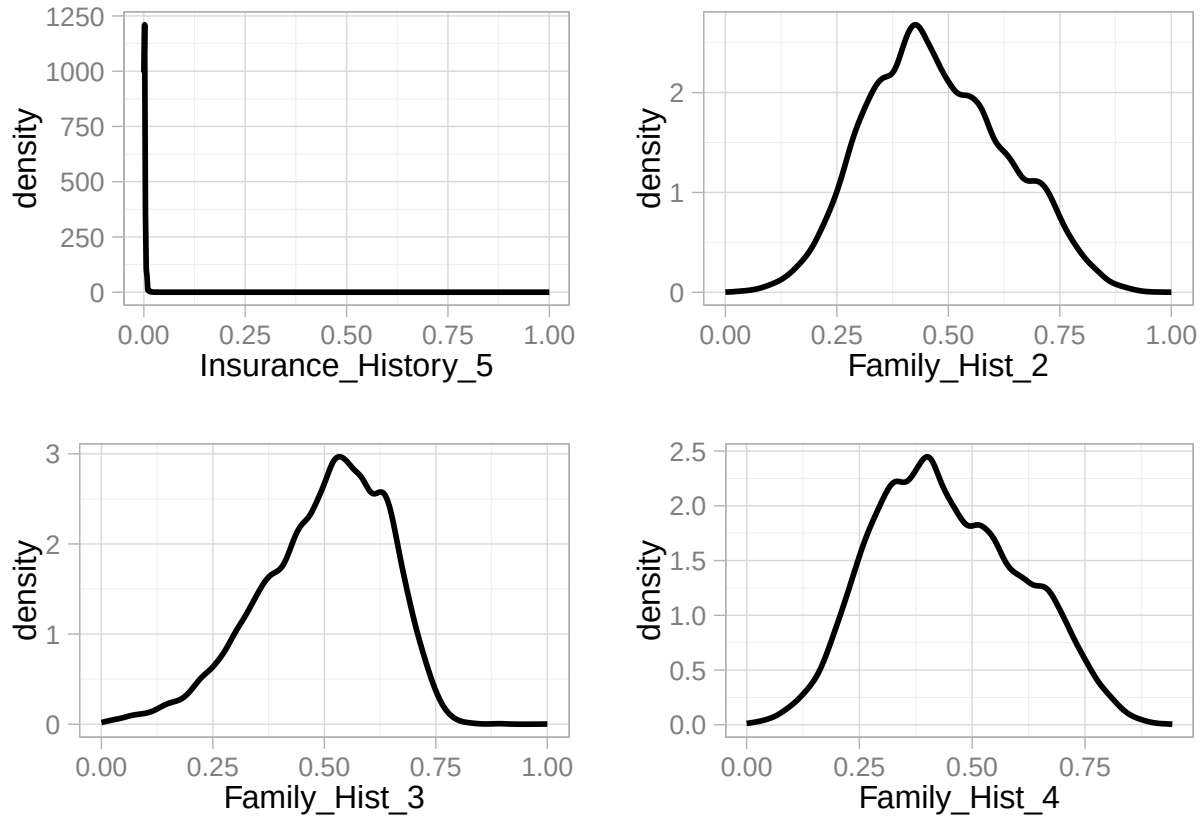
```
train.cont <- data.frame(train.cont, Response=train$Response)
plotDensity <- function(data.in, i) {
  data <- data.frame(x=data.in[,i], Response=data.in$Response)
  p <- ggplot(data) + #geom_density(aes(x=x, colour=factor(Response))) +
    geom_line(aes(x=x), stat="density", size=1, alpha=1.0) +
    xlab(colnames(data.in)[i]) + theme_light()
  return (p)
}

doPlots(data.in=train.cont, fun=plotDensity, ii=1:4, ncol=2)
```



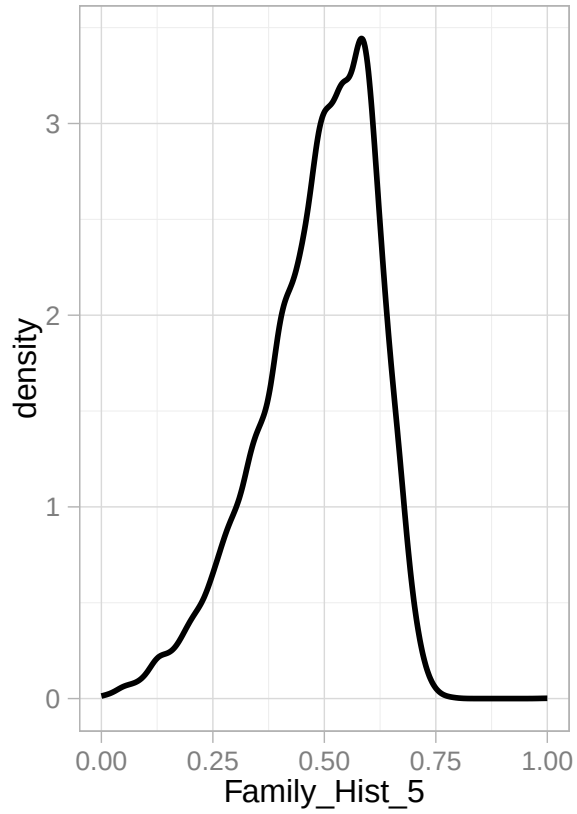

```
doPlots(data.in=train.cont, fun=plotDensity, ii=5:8, ncol=2)
```




```
doPlots(data.in=train.cont, fun=plotDensity, ii=9:12, ncol=2)
```

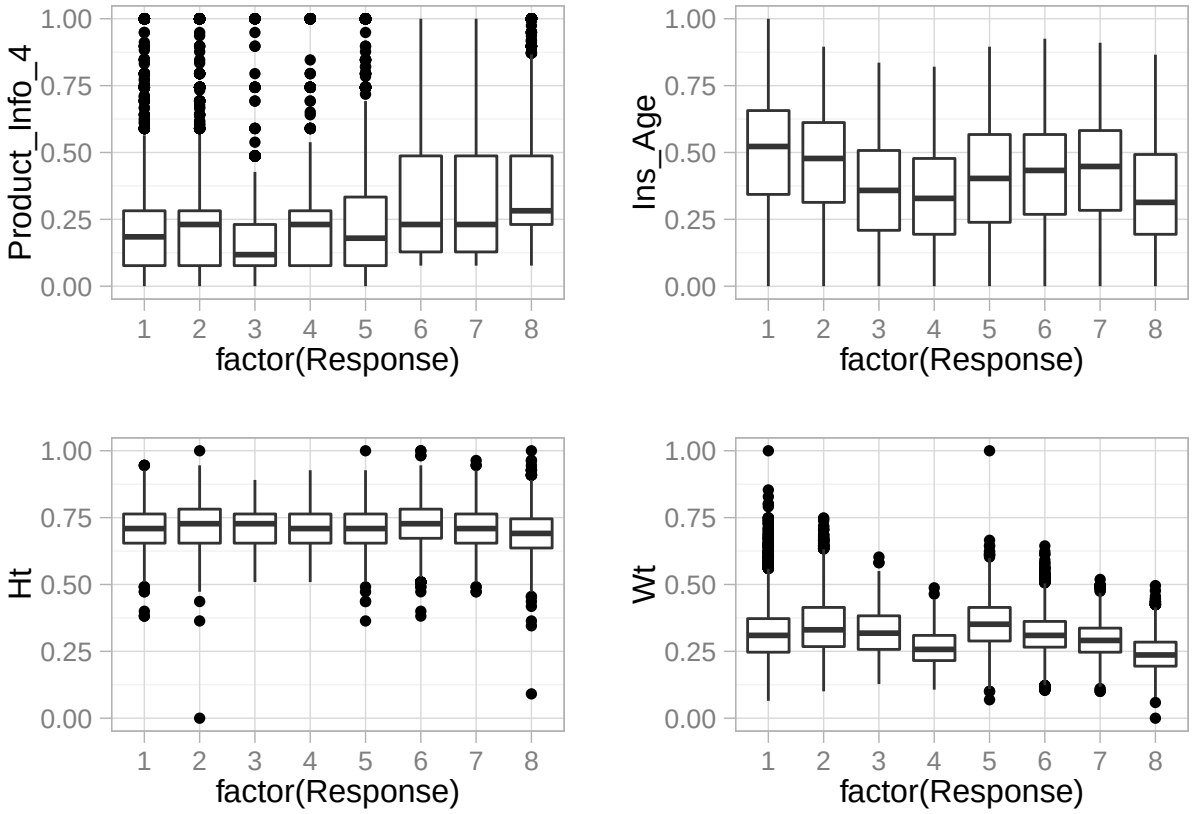



```
doPlots(data.in=train.cont, fun=plotDensity, ii=13, ncol=2)
```

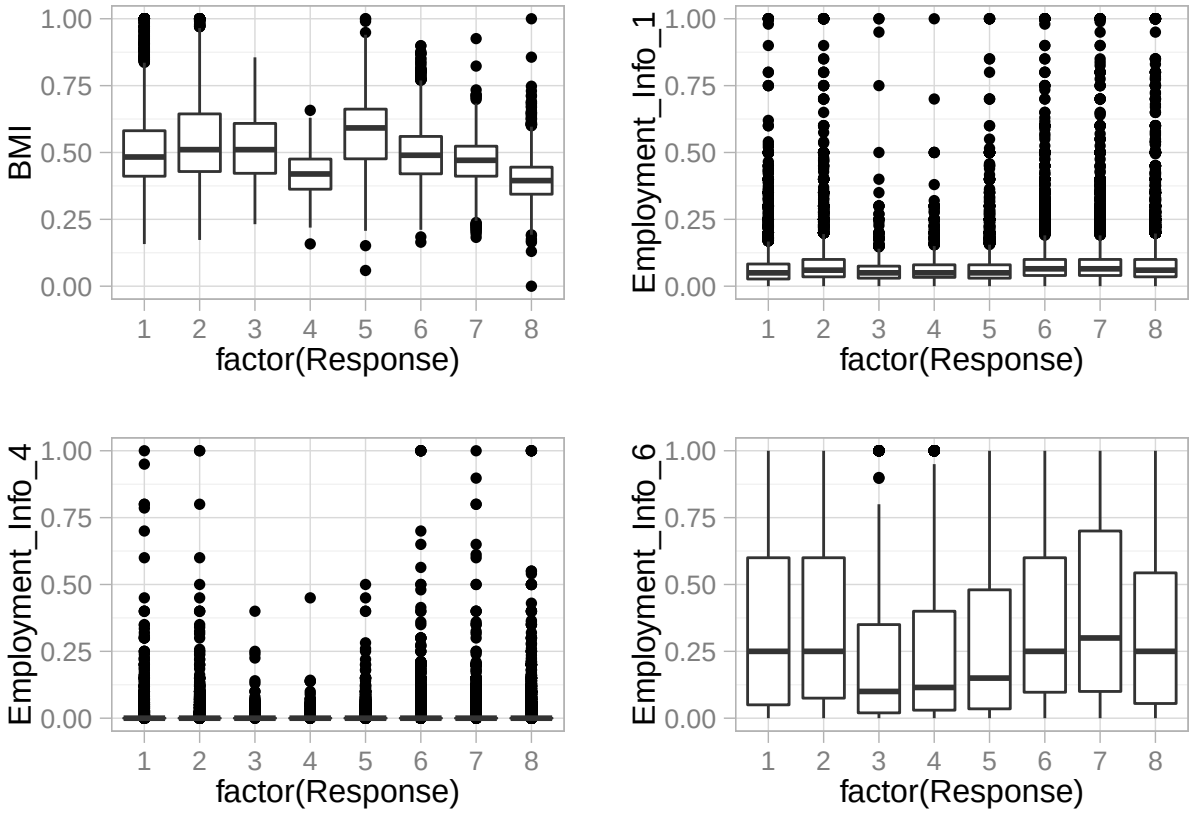


Boxplots of continous Features depending on Response

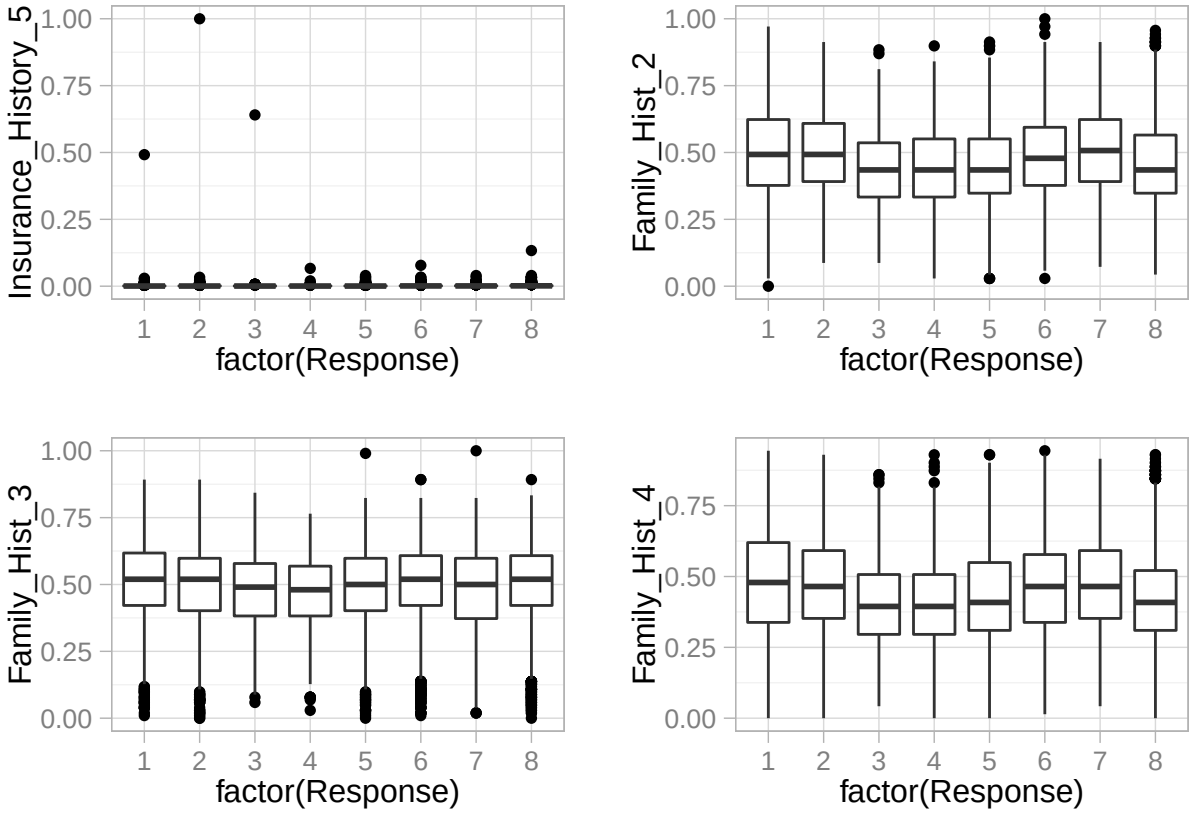
```
plotBox <- function(data.in, i) {
  data <- data.frame(y=data.in[,i], Response=data.in$Response)
  p <- ggplot(data, aes(x=factor(Response), y=y)) + geom_boxplot() + ylab(colnames(data.in)[i]) + theme_
  return (p)
}

doPlots(data.in=train.cont, fun=plotBox, ii=1:4, ncol=2)
```



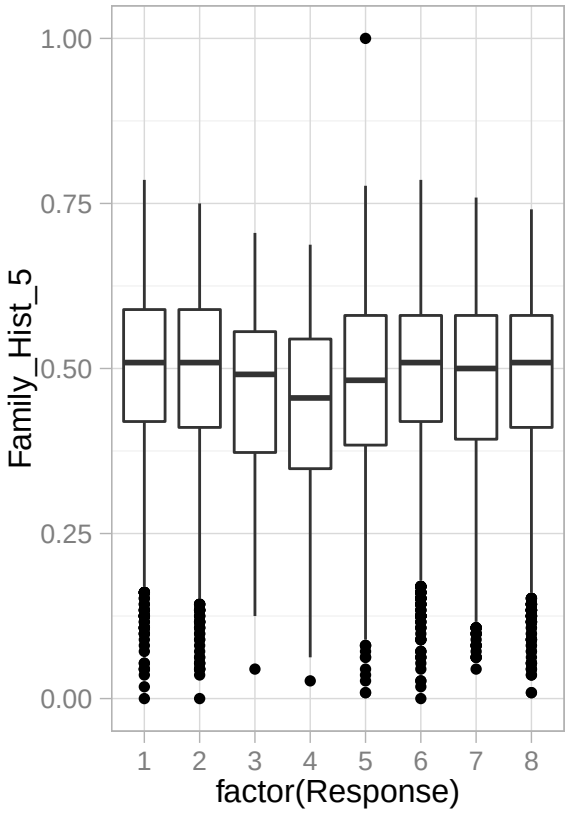

```
doPlots(data.in=train.cont, fun=plotBox, ii=5:8, ncol=2)
```




```
doPlots(data.in=train.cont, fun=plotBox, ii=9:12, ncol=2)
```

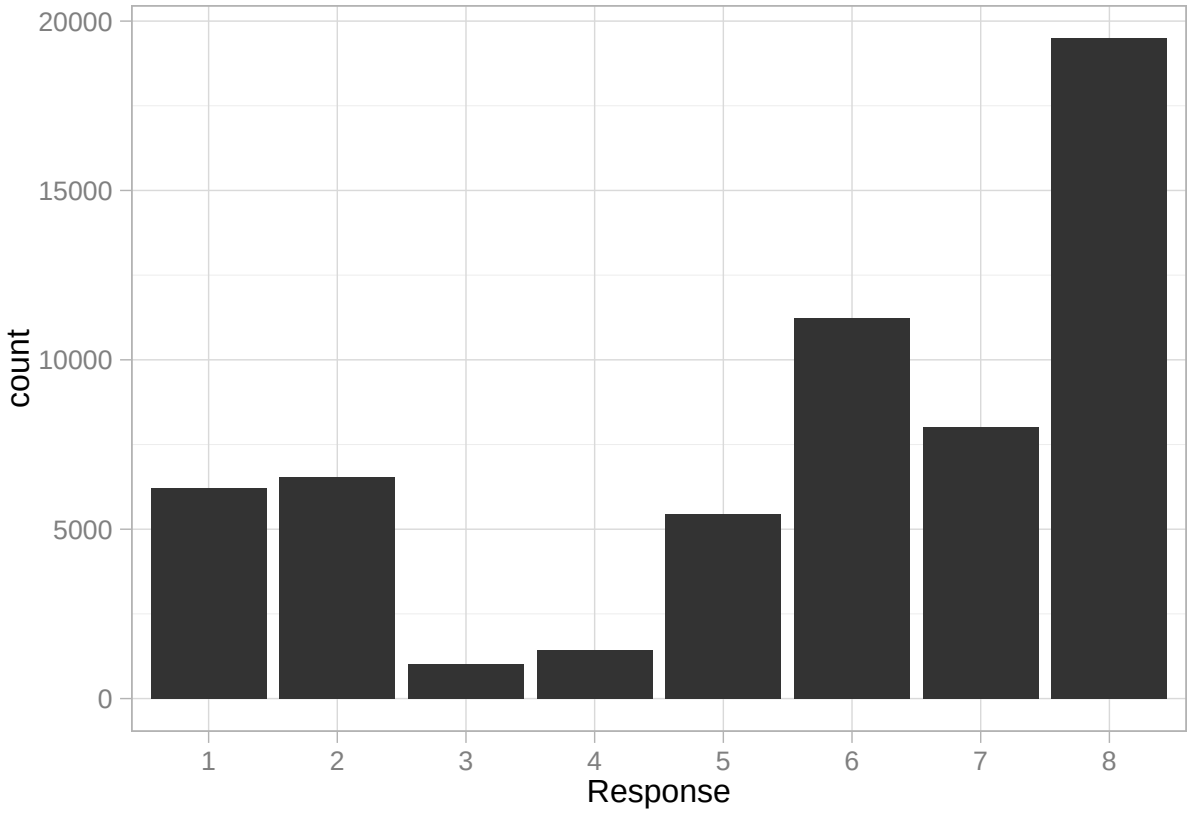



```
doPlots(data.in=train.cont, fun=plotBox, ii=13, ncol=2)
```



Take a look at the response

```
ggplot(train) + geom_histogram(aes(factor(Response))) + xlab("Response") + theme_light()
```


20000
15000
10000
5000
0
count
1
2
3
4
5
6
7
8
Response