Age Analysis OBPD

```
##### Simulations #####
SIS <- function(
  obs, #observed table
  n.sim = 1000, # Tables to generate per simulation
  dist = c("unif") #default proposal dist. is uniform but can also take hypergeometric
){
  r.sums <- rowSums(obs); #= c(10,62,13,11,39), #vector of row sums
  c.sums <- colSums(obs); #= c(65,25,45), #vector of column sums
  r.dim <- length(r.sums);#dimensions row
  c.dim <- length(c.sums);# dimensions column

  w <- as(1:n.sim,"mpfr"); # weights for simulated table I
  l.x <- as(1:n.sim,"mpfr"); #proportional dist. to pi(x).

  for(z in 1:n.sim){
    X <- matrix(nrow = r.dim,ncol = c.dim) #Empty Matrix representing a table
    g <- matrix(nrow = (r.dim-1),ncol = (c.dim-1)) #simulated proposal dist.

    #saving the lower and upper values to compute q(T)
    lower <- matrix(nrow = r.dim-1,ncol = c.dim - 1);
    upper <- matrix(nrow = r.dim-1,ncol = c.dim - 1);
    for(j in 1:(c.dim-1)){
      for(i in 1:(r.dim-1)){
        #finding the lower and upper bound for each cell
        lower[i,j]<-max(0,c.sums[j]-sum(X[(1:i),j],na.rm = TRUE)-#total column sum
                          sum(
                            r.sums[(i+1):r.dim], #row sums of non simulated values
                            -sum(X[(i+1):r.dim,1:j],na.rm = TRUE),#row sum of already si
m. values
                            na.rm = TRUE))
        upper[i,j]<- min(
          c.sums[j]- sum(X[1:i,j],na.rm = TRUE),#column sum - values already simulated i
n the col.
          sum(r.sums[i],-sum(X[i,1:j],na.rm = TRUE),na.rm = TRUE))#row sum - values sim.
 in row

        if(dist == c("unif")){#simulate from uniform distribution
          X[i,j] <- round(runif(1,min = lower[i,j],max =upper[i,j]))
          g[i,j] <- 1/(upper[i,j]-lower[i,j] + 1);
        }
        if(dist == c("hyper")){#simulate from hypergeometric distribution
          X[i,j] <- rhyper(1,m = upper[i,j], n = upper[i,j], k = lower[i,j]+upper[i,j]);

          g[i,j] <- dhyper(X[i,j],m = upper[i,j], n = upper[i,j], k =
lower[i,j]+upper[i,j]);
        }
      }
      X[r.dim,j] <- c.sums[j] - sum(X[1:(r.dim-1),j]);
    }

    X[,c.dim]<- (r.sums - apply(X,1,function(x)sum(x,na.rm = TRUE)));
```

```
    l.x[z]<-1/prod(gamma(as(X+1,"mpfr")))
    w[z]<- prod(g)/prod(gamma(as(X+1,"mpfr"))); #weight l.x/g.x
    #g.x is the product of conditionals and l.x is proportional to multinomial


  }
  p.val <- sum((1/prod(factorial(obs))>=l.x)*w)/sum(w); #p-value
  #let h(x) = indicator function. w = weight.
  #use normal importance sampling to get an estimate of p-value

  list(weight = w, #vector with length (n.sim) of weights
       last.weight = g, #example of proposal dist of last table
       upper = upper, #example of the upper bounds of the last table
       lower = lower, #example of the lower bounds of the last table
       X = X, #example of last table
       #n.sim = n.sim, # number of simulated tables
       p.value = p.val #simulated p-value
  )
}


#list of files in dir.
contigency.table.dir <- "/Users/omachowda/Google Drive/StatCom 3/OBPD/Contingency table
s/"
cont.tables.list <- list.files(contigency.table.dir)
#take only .csv files
cont.tables.list <- cont.tables.list[grep(cont.tables.list,pattern = ".csv")]
#get table names
table.name <- gsub(cont.tables.list,pattern = ".csv",replacement = "");
#assign table to name
for(i in 1:length(table.name)){
  tab <- read.csv(cont.tables.list[i]);
  tab <- tab[-c(1,2,nrow(tab)),-c(1,ncol(tab))]; #get rid of row and column sums and fir
st 2 rows
  assign(table.name[i],tab)
}

SIS <- dget("/Users/omachowda/Google Drive/StatCom 3/OBPD/Robin/Sequential Importance Sa
mpling Function.R") #import SIS function
```

```
## Warning: package 'Rmpfr' was built under R version 3.2.5
```

```
## Loading required package: gmp
```

```
## Warning: package 'gmp' was built under R version 3.2.5
```

```
##
## Attaching package: 'gmp'
```

```
## The following objects are masked from 'package:base':
##
##     %*%, apply, crossprod, matrix, tcrossprod
```

```
## C code of R package 'Rmpfr': GMP using 64 bits per limb
```

```
##
## Attaching package: 'Rmpfr'
```

```
## The following objects are masked from 'package:stats':
##
##     dbinom, dnorm, dpois, pnorm
```

```
## The following objects are masked from 'package:base':
##
##     cbind, pmax, pmin, rbind
```

```
#facilities to age output
```

Example of how table is simulated

```
SIS(get(table.name[1]),dist = "hyper")$upper
```

```
##      [,1] [,2] [,3] [,4]
## [1,]   35   16    8    5
## [2,]   50   26   14    7
## [3,]   81   42   21   30
```

```
SIS(get(table.name[1]),dist = "hyper")$lower
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    0    0    0    0
## [2,]    0    0    0    0
## [3,]    0    0    0    0
```

```
SIS(get(table.name[1]),dist = "hyper")$X
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]   17    9    4    2    3
## [2,]   26   11    7    2    4
## [3,]   42   23   10   10   16
## [4,]   39   19   10   24  124
```

```
SIS(get(table.name[1]),dist = "hyper")$p.value
```

```
## 1 'mpfr' number of precision  128   bits
## [1] 0
```

SIS P-values for all questions based on income

```
#reference character items as objects
SIS.results <- lapply(sapply(table.name,get),function(x) SIS(x,dist = "hyper")$p.value)
SIS.results
```

```
## $facilities
## 1 'mpfr' number of precision  128   bits
## [1] 0
##
## $`oversee sports core`
## 1 'mpfr' number of precision  128   bits
## [1] 0
##
## $`park areas`
## 1 'mpfr' number of precision  128   bits
## [1] 0
##
## $`recreation programs`
## 1 'mpfr' number of precision  128   bits
## [1] 0
##
## $`satisfaction with district`
## 1 'mpfr' number of precision  128   bits
## [1] 0
##
## $`satisfaction with facilities`
## 1 'mpfr' number of precision  128   bits
## [1] 0
##
## $`satisfaction with maintenence`
## 1 'mpfr' number of precision  128   bits
## [1] 0
##
## $`satisfaction with outdoor amenities and parks`
## 1 'mpfr' number of precision  128   bits
## [1] 0
##
## $`satisfaction with programs`
## 1 'mpfr' number of precision  128   bits
## [1] 0
##
## $`satisfaction with staff`
## 1 'mpfr' number of precision  128   bits
## [1] 0
##
## $`special events`
## 1 'mpfr' number of precision  128   bits
## [1] 0
##
## $`work with sports core`
## 1 'mpfr' number of precision  128   bits
## [1] 0
```

```
##### M^2 Test ######

corr_calc <- function(df){
  columns = length(df)
  new_df = data.frame()
  v1= c()
  v2= c()
  z = 0
  for (i in 1:4){
    for (j in 1:columns){
      new_df <- rbind(new_df, c(i,j, df[i,j]))
      z = z + df[i,j]
    }
  }

  for (i in 1:(4*columns)){
    v1 = append(v1, c(rep(new_df[i,1],new_df[i,3])))
  }

  for (i in 1:(4*columns)){
    v2 = append(v2, c(rep(new_df[i,2],new_df[i,3])))
  }

  fit = cor(x = v1,y = v2)
  M = (z-1)*(fit)
  p = pchisq(M,12)
  return (p)
}
```

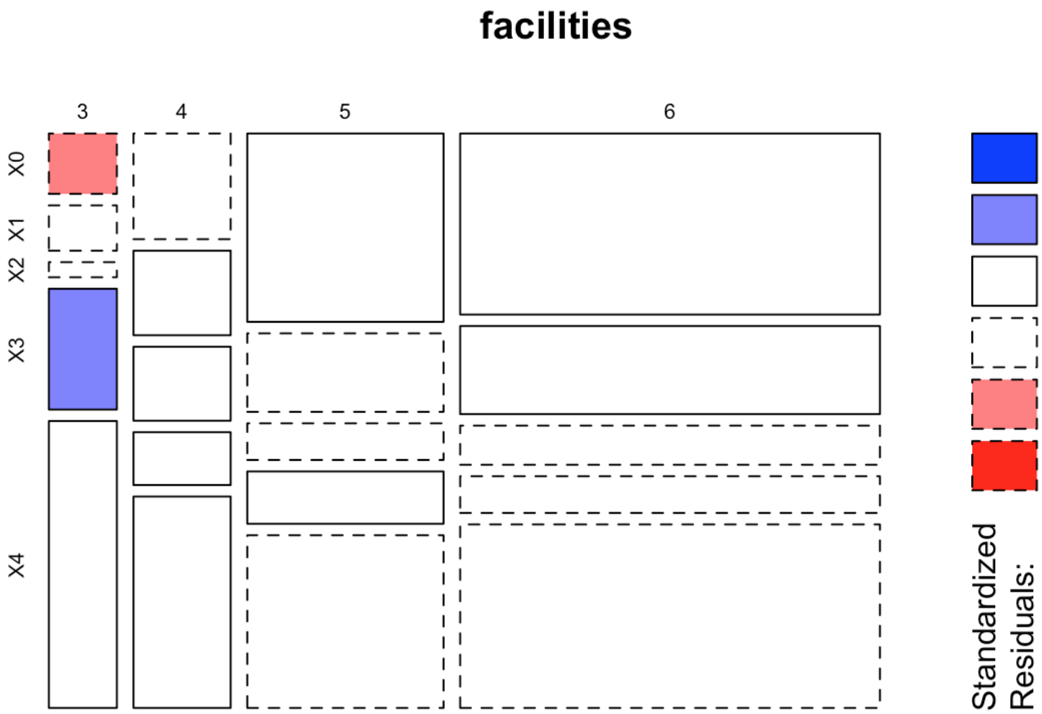
M^2 p-values for all questions based on income

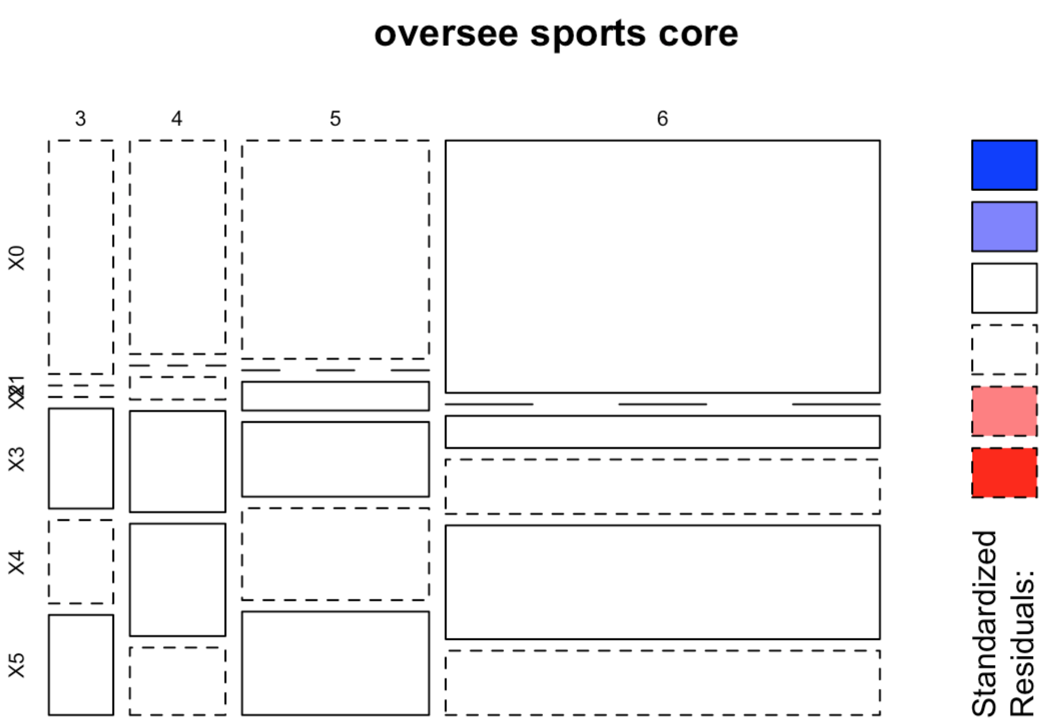
```
resulted = lapply(sapply(table.name,get),FUN= corr_calc)
resulted
```

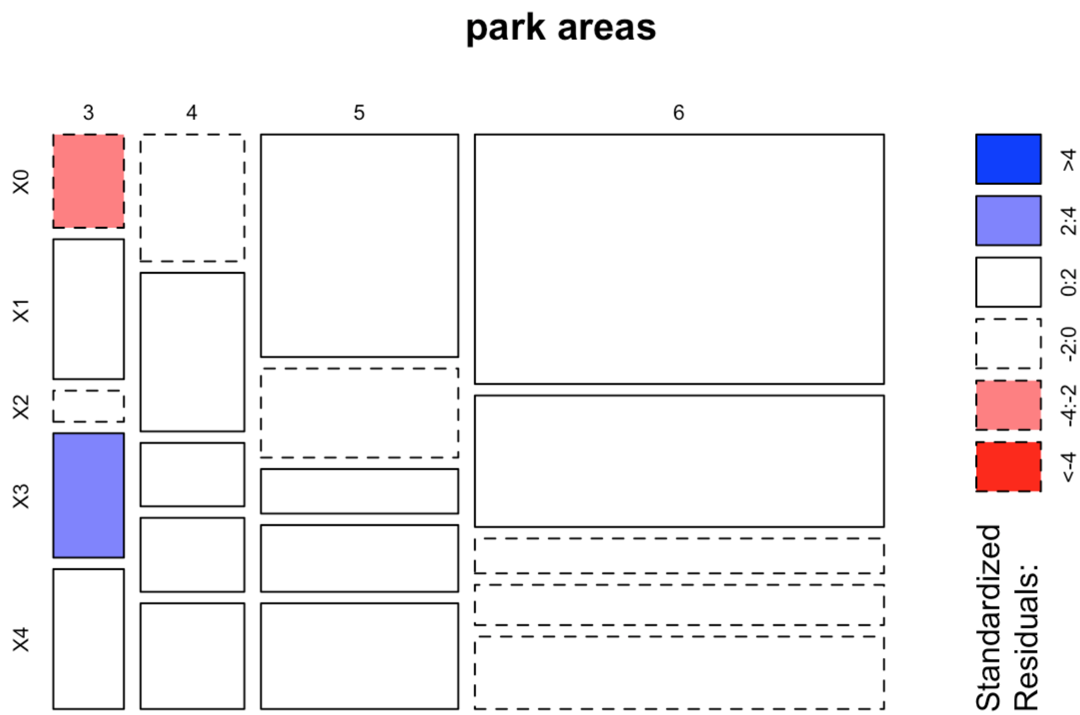
```
## $facilities
## [1] 0
##
## $`oversee sports core`
## [1] 0
##
## $`park areas`
## [1] 0
##
## $`recreation programs`
## [1] 0
##
## $`satisfaction with district`
## [1] 0
##
## $`satisfaction with facilities`
## [1] 0
##
## $`satisfaction with maintenence`
## [1] 0
##
## $`satisfaction with outdoor amenities and parks`
## [1] 0
##
## $`satisfaction with programs`
## [1] 0
##
## $`satisfaction with staff`
## [1] 0
##
## $`special events`
## [1] 0
##
## $`work with sports core`
## [1] 0
```

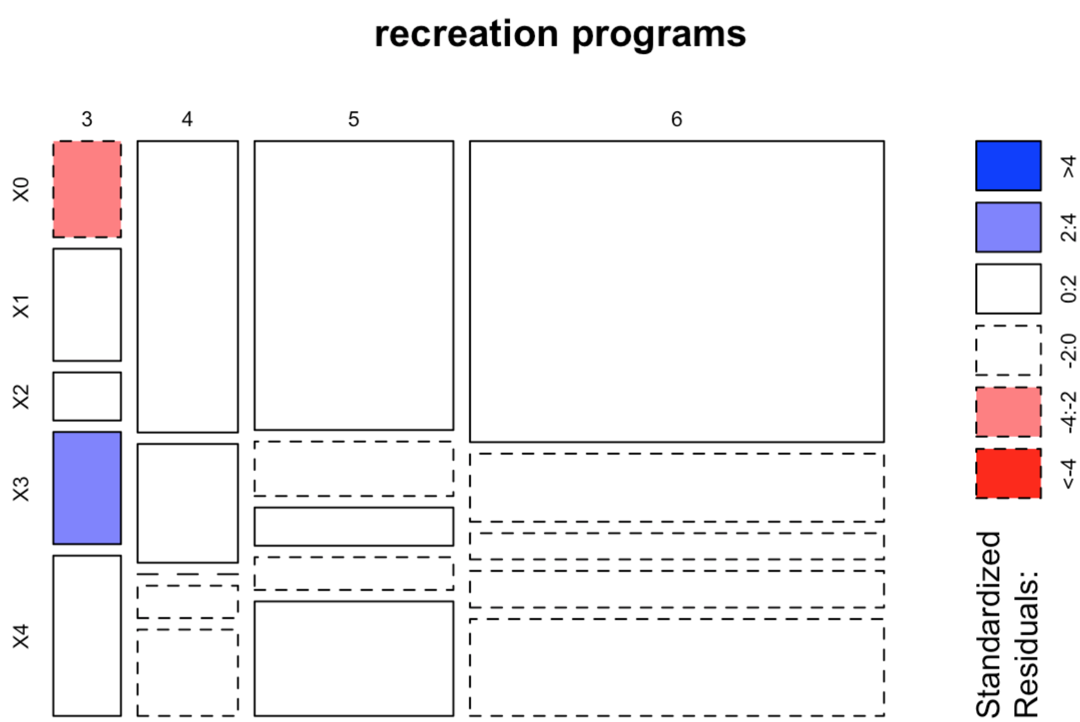
Mosaic Plots for questions based on income

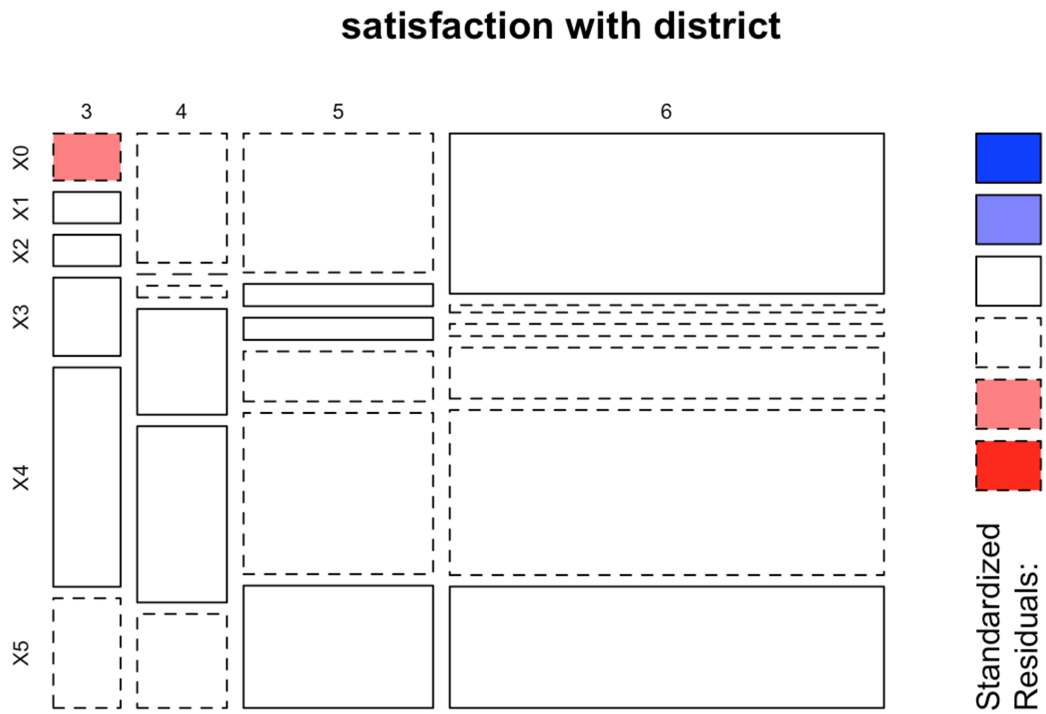
```
counter=1;
par(mfrow=c(1,1))
for( i in 1:12){
  mosaicplot(x = get(table.name[i]), shade = TRUE,color = TRUE, main= table.name[i])
}
```

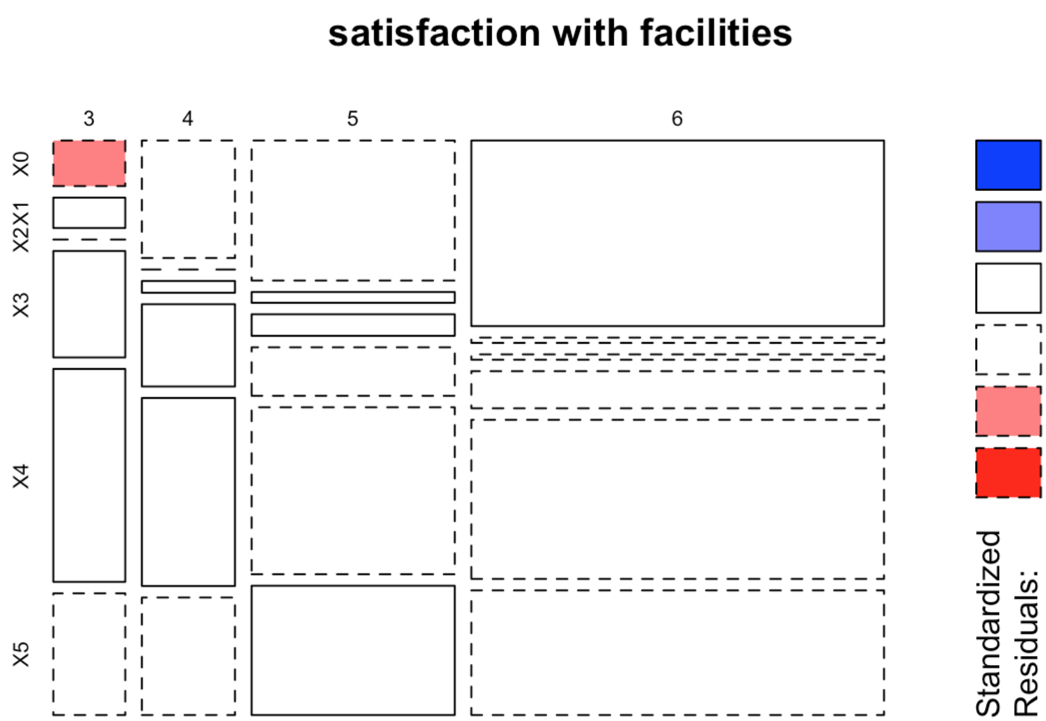

facilities
3
4
5
6
X0
X1
X2
X3
X4
>4
2:4
0:2
-2:0
-4:-2
<-4
Standardized
Residuals:


oversee sports core
3
4
5
6
X0
X1
X2
X3
X4
X5
>4
2:4
0:2
-2:0
-4:-2
<-4
Standardized
Residuals:

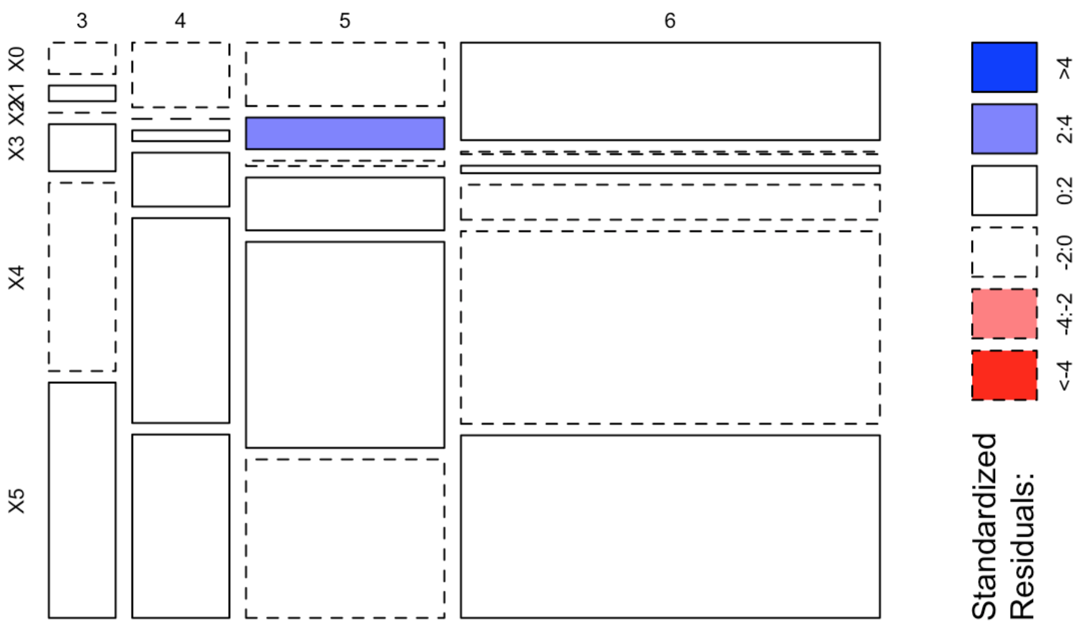

park areas
3
4
5
6
X0
X1
X2
X3
X4
>4
2:4
0:2
-2:0
-4:-2
<-4
Standardized
Residuals:


recreation programs
3
4
5
6
X0
X1
X2
X3
X4
>4
2:4
0:2
-2:0
-4:-2
<-4
Standardized
Residuals:

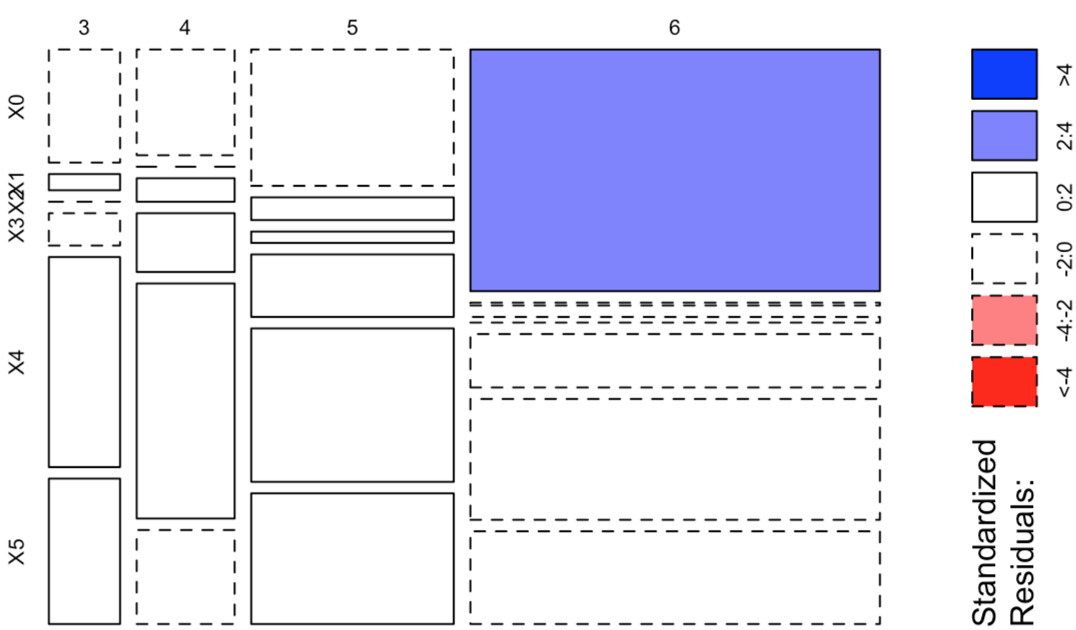

satisfaction with district
3
4
5
6
X0
X1
X2
X3
X4
X5
>4
2:4
0:2
-2:0
-4:-2
<-4
Standardized
Residuals:

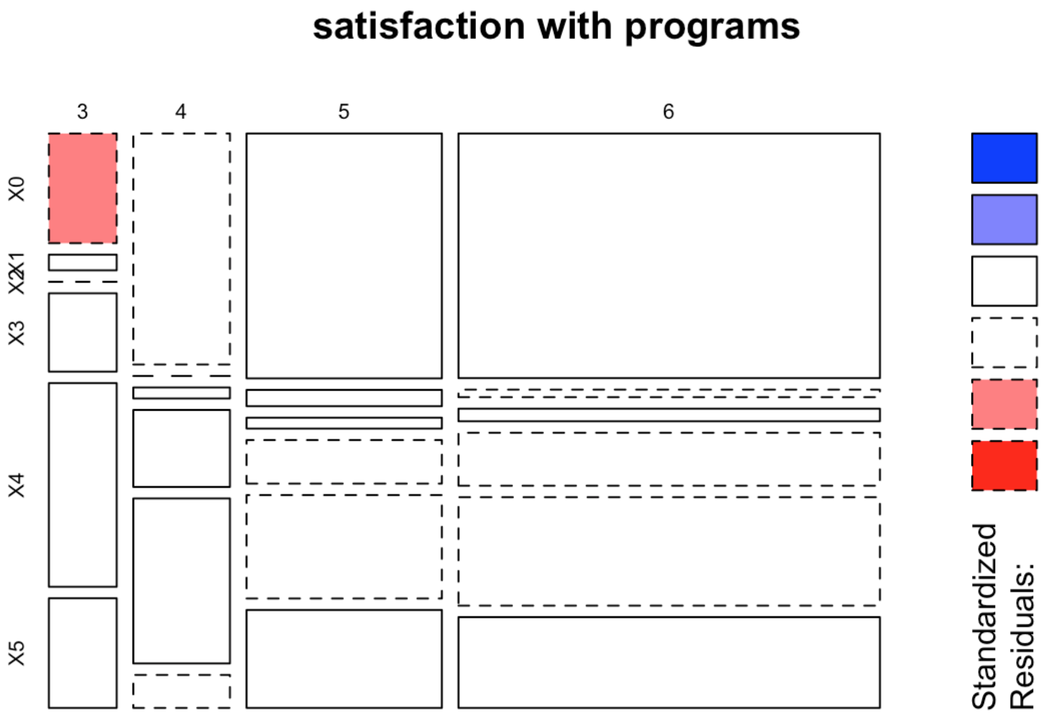

satisfaction with facilities
3
4
5
6
X0
X2X1
X3
X4
X5
>4
2:4
0:2
-2:0
-4:-2
<-4
Standardized
Residuals:

satisfaction with maintenence

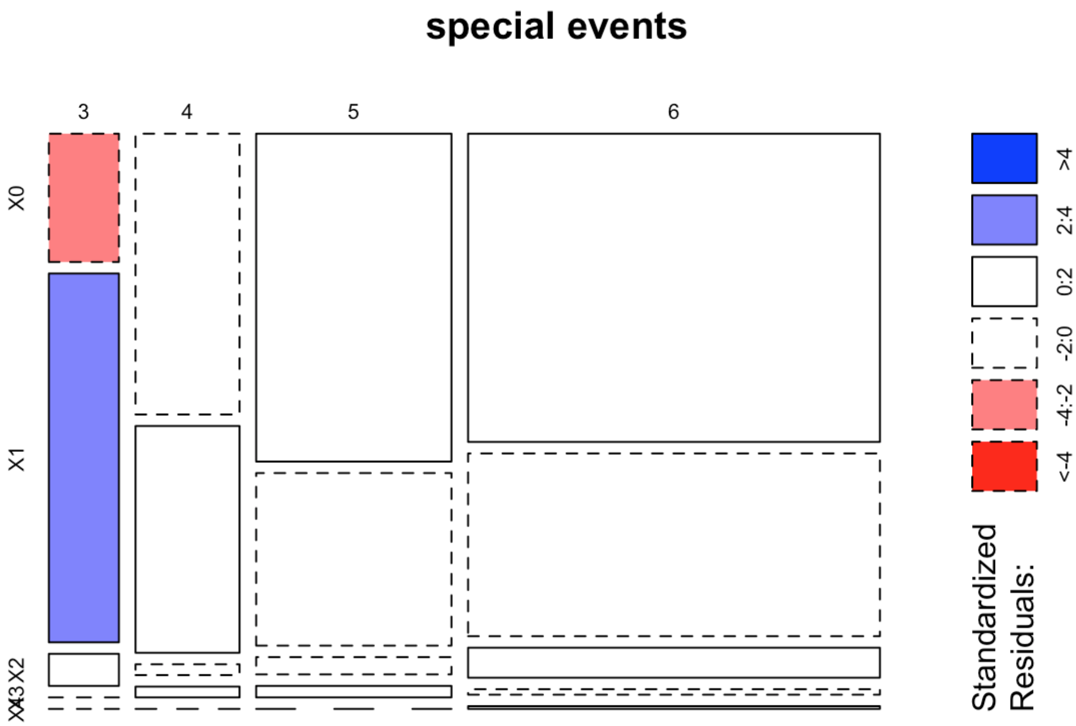


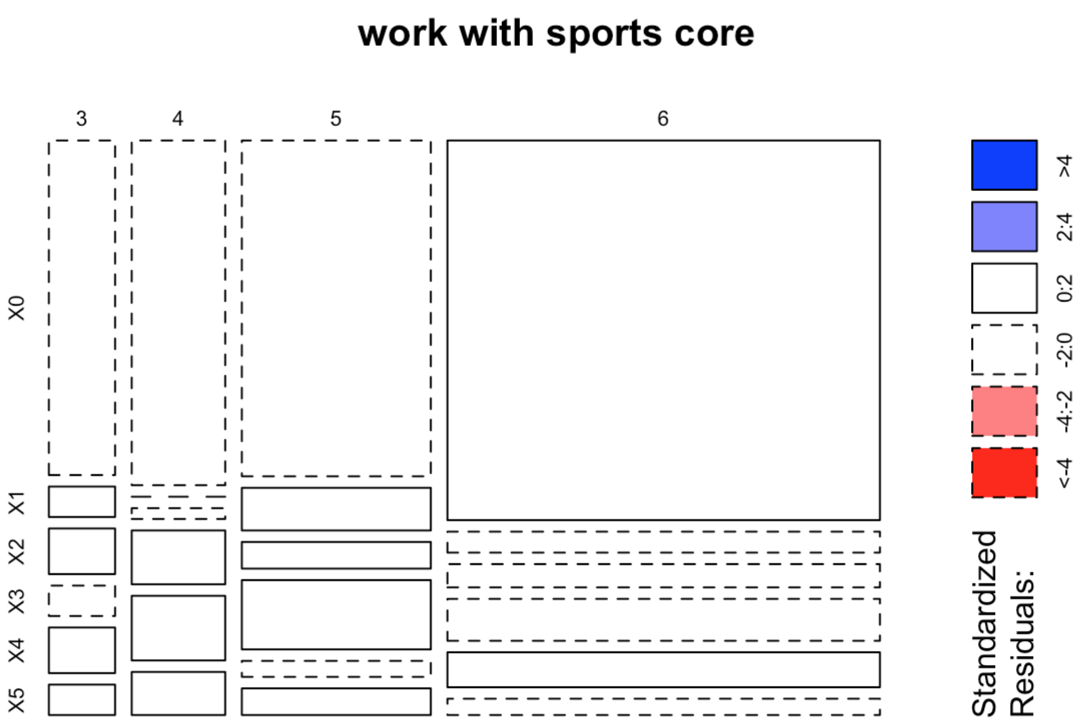
satisfaction with outdoor amenities and parks




satisfaction with programs
3
4
5
6
X0
X2X1
X3
X4
X5
>4
2:4
0:2
-2:0
-4:-2
<-4
Standardized
Residuals:


satisfaction with staff
3
4
5
6
X0
X2X1
X3
X4
X5
>4
2:4
0:2
-2:0
-4:-2
<-4
Standardized
Residuals:


special events
3
4
5
6
X0
X1
X2
X3
X4
>4
2:4
0:2
-2:0
-4:-2
<-4
Standardized
Residuals:


work with sports core
3
4
5
6
X0
X1
X2
X3
X4
X5
>4
2:4
0:2
-2:0
-4:-2
<-4
Standardized
Residuals: