Correspondence Analysis - DataCrew

Erik Pak

2023-05-23

Import Libraries

```r
# import libraries
library(tidyverse)     # data wrangling
library(corrplot)      # correlation plot
library(psych)         # used for describe
library(ggpubr)        # combine plots


library(ca)            # correspondence analysis
library(polycor)       # hetcor

library(MASS)          # Also contains a qda - Quadratic discriminant analysis function
library(rfUtilities)   # accuracy function

library("FactoMineR")  # Multiple correspondence analysis (MCA)
library("factoextra")  # Multiple correspondence analysis (MCA)
```

Categorical variables mapping

Education:

1 'Below College'

2 'College'

3 'Bachelor'

4 'Master'

5 'Doctor'

Environment Satisfaction:

1 'Low'

2 'Medium'

3 'High'

4 'Very High'

Job Involvement:

1 'Low'

2 'Medium'

3 'High'

4 'Very High'

Job Satisfaction:

1 'Low'

2 'Medium'

3 'High'

4 'Very High'

Relationship Satisfaction:

1 'Low'

2 'Medium'

3 'High'

4 'Very High'

WorkLife Balance:

1 'Bad'

2 'Good'

3 'Better'

4 'Best'

Import HR Data

```r
# import data
hr <- read.csv("HREmployeeAttrition.csv")

# display
head(hr)
```

```
##   Age Attrition     BusinessTravel DailyRate             Department
## 1  41       Yes      Travel_Rarely      1102                  Sales
## 2  49        No Travel_Frequently       279 Research & Development
## 3  37       Yes      Travel_Rarely      1373 Research & Development
## 4  33        No Travel_Frequently      1392 Research & Development
## 5  27        No      Travel_Rarely       591 Research & Development
## 6  32        No Travel_Frequently      1005 Research & Development
##   DistanceFromHome Education EducationField EmployeeCount EmployeeNumber
## 1                1         2  Life Sciences             1              1
## 2                8         1  Life Sciences             1              2
## 3                2         2          Other             1              4
## 4                3         4  Life Sciences             1              5
## 5                2         1        Medical             1              7
## 6                2         2  Life Sciences             1              8
##   EnvironmentSatisfaction Gender HourlyRate JobInvolvement JobLevel
## 1                       2 Female         94              3        2
## 2                       3   Male         61              2        2
## 3                       4   Male         92              2        1
## 4                       4 Female         56              3        1
## 5                       1   Male         40              3        1
## 6                       4   Male         79              3        1
##               JobRole JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate
## 1     Sales Executive               4        Single          5993       19479
## 2   Research Scientist              2       Married          5130       24907
```

```
## 3 Laboratory Technician            3    Single       2090      2396
## 4     Research Scientist           3   Married       2909     23159
## 5 Laboratory Technician            2   Married       3468     16632
## 6 Laboratory Technician            4    Single       3068     11864
##   NumCompaniesWorked Over18 OverTime PercentSalaryHike PerformanceRating
## 1                  8      Y      Yes                11                 3
## 2                  1      Y       No                23                 4
## 3                  6      Y      Yes                15                 3
## 4                  1      Y      Yes                11                 3
## 5                  9      Y       No                12                 3
## 6                  0      Y       No                13                 3
##   RelationshipSatisfaction StandardHours StockOptionLevel TotalWorkingYears
## 1                        1            80                0                 8
## 2                        4            80                1                10
## 3                        2            80                0                 7
## 4                        3            80                0                 8
## 5                        4            80                1                 6
## 6                        3            80                0                 8
##   TrainingTimesLastYear WorkLifeBalance YearsAtCompany YearsInCurrentRole
## 1                     0               1              6                  4
## 2                     3               3             10                  7
## 3                     3               3              0                  0
## 4                     3               3              8                  7
## 5                     3               3              2                  2
## 6                     2               2              7                  7
##   YearsSinceLastPromotion YearsWithCurrManager
## 1                       0                    5
## 2                       1                    7
## 3                       0                    0
## 4                       3                    0
## 5                       2                    2
## 6                       3                    6
```

```
# summary
summary(hr)
```

```
##       Age          Attrition         BusinessTravel        DailyRate
##  Min.   :18.00   Length:1470        Length:1470        Min.   : 102.0
##  1st Qu.:30.00   Class :character   Class :character   1st Qu.: 465.0
##  Median :36.00   Mode  :character   Mode  :character   Median : 802.0
##  Mean   :36.92                                         Mean   : 802.5
##  3rd Qu.:43.00                                         3rd Qu.:1157.0
##  Max.   :60.00                                         Max.   :1499.0
##   Department        DistanceFromHome   Education     EducationField
##  Length:1470        Min.   : 1.000   Min.   :1.000   Length:1470
##  Class :character   1st Qu.: 2.000   1st Qu.:2.000   Class :character
##  Mode  :character   Median : 7.000   Median :3.000   Mode  :character
##                     Mean   : 9.193   Mean   :2.913
##                     3rd Qu.:14.000   3rd Qu.:4.000
##                     Max.   :29.000   Max.   :5.000
##   EmployeeCount EmployeeNumber   EnvironmentSatisfaction   Gender
##  Min.   :1      Min.   :   1.0   Min.   :1.000           Length:1470
##  1st Qu.:1      1st Qu.: 491.2   1st Qu.:2.000           Class :character
##  Median :1      Median :1020.5   Median :3.000           Mode  :character
##  Mean   :1      Mean   :1024.9   Mean   :2.722
```

```
##  3rd Qu.:1      3rd Qu.:1555.8   3rd Qu.:4.000
##  Max.   :1      Max.   :2068.0   Max.   :4.000
##    HourlyRate      JobInvolvement    JobLevel       JobRole
##  Min.   : 30.00   Min.   :1.00   Min.   :1.000   Length:1470
##  1st Qu.: 48.00   1st Qu.:2.00   1st Qu.:1.000   Class :character
##  Median : 66.00   Median :3.00   Median :2.000   Mode  :character
##  Mean   : 65.89   Mean   :2.73   Mean   :2.064
##  3rd Qu.: 83.75   3rd Qu.:3.00   3rd Qu.:3.000
##  Max.   :100.00   Max.   :4.00   Max.   :5.000
##  JobSatisfaction MaritalStatus     MonthlyIncome    MonthlyRate
##  Min.   :1.000   Length:1470      Min.   : 1009   Min.   : 2094
##  1st Qu.:2.000   Class :character  1st Qu.: 2911   1st Qu.: 8047
##  Median :3.000   Mode  :character  Median : 4919   Median :14236
##  Mean   :2.729                     Mean   : 6503   Mean   :14313
##  3rd Qu.:4.000                     3rd Qu.: 8379   3rd Qu.:20462
##  Max.   :4.000                     Max.   :19999   Max.   :26999
##  NumCompaniesWorked   Over18              OverTime          PercentSalaryHike
##  Min.   :0.000      Length:1470       Length:1470        Min.   :11.00
##  1st Qu.:1.000      Class :character  Class :character   1st Qu.:12.00
##  Median :2.000      Mode  :character  Mode  :character   Median :14.00
##  Mean   :2.693                                           Mean   :15.21
##  3rd Qu.:4.000                                           3rd Qu.:18.00
##  Max.   :9.000                                           Max.   :25.00
##  PerformanceRating RelationshipSatisfaction StandardHours StockOptionLevel
##  Min.   :3.000     Min.   :1.000            Min.   :80    Min.   :0.0000
##  1st Qu.:3.000     1st Qu.:2.000            1st Qu.:80    1st Qu.:0.0000
##  Median :3.000     Median :3.000            Median :80    Median :1.0000
##  Mean   :3.154     Mean   :2.712            Mean   :80    Mean   :0.7939
##  3rd Qu.:3.000     3rd Qu.:4.000            3rd Qu.:80    3rd Qu.:1.0000
##  Max.   :4.000     Max.   :4.000            Max.   :80    Max.   :3.0000
##  TotalWorkingYears TrainingTimesLastYear WorkLifeBalance YearsAtCompany
##  Min.   : 0.00     Min.   :0.000         Min.   :1.000   Min.   : 0.000
##  1st Qu.: 6.00     1st Qu.:2.000         1st Qu.:2.000   1st Qu.: 3.000
##  Median :10.00     Median :3.000         Median :3.000   Median : 5.000
##  Mean   :11.28     Mean   :2.799         Mean   :2.761   Mean   : 7.008
##  3rd Qu.:15.00     3rd Qu.:3.000         3rd Qu.:3.000   3rd Qu.: 9.000
##  Max.   :40.00     Max.   :6.000         Max.   :4.000   Max.   :40.000
##  YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager
##  Min.   : 0.000     Min.   : 0.000          Min.   : 0.000
##  1st Qu.: 2.000     1st Qu.: 0.000          1st Qu.: 2.000
##  Median : 3.000     Median : 1.000          Median : 3.000
##  Mean   : 4.229     Mean   : 2.188          Mean   : 4.123
##  3rd Qu.: 7.000     3rd Qu.: 3.000          3rd Qu.: 7.000
##  Max.   :18.000     Max.   :15.000          Max.   :17.000
```

```r
# remove EmployeeCount & EmployeeNumber & Over18 & StandardHours
hr <- dplyr::select(hr, -c(EmployeeCount, EmployeeNumber, Over18, StandardHours))


# convert multiple fields into a factor
catCols <- c('Attrition', 'BusinessTravel', 'Department', 'Education', 'EducationField',
        'EnvironmentSatisfaction', 'Gender', 'JobInvolvement', 'JobLevel', 'JobRole',
        'JobSatisfaction', 'MaritalStatus', 'OverTime', 'PerformanceRating',
        'RelationshipSatisfaction', 'StockOptionLevel', 'WorkLifeBalance')
# convert to factors
```

```r
hr[catCols] <- lapply(hr[catCols], factor)

# summary
summary(hr)
```

```
##       Age          Attrition           BusinessTravel    DailyRate
##  Min.   :18.00   No :1233   Non-Travel      : 150   Min.   : 102.0
##  1st Qu.:30.00   Yes: 237   Travel_Frequently: 277   1st Qu.: 465.0
##  Median :36.00              Travel_Rarely    :1043   Median : 802.0
##  Mean   :36.92                                       Mean   : 802.5
##  3rd Qu.:43.00                                       3rd Qu.:1157.0
##  Max.   :60.00                                       Max.   :1499.0
##
##                        Department   DistanceFromHome Education
##  Human Resources         : 63   Min.   : 1.000   1:170
##  Research & Development:961   1st Qu.: 2.000   2:282
##  Sales                 :446   Median : 7.000   3:572
##                               Mean   : 9.193   4:398
##                               3rd Qu.:14.000   5: 48
##                               Max.   :29.000
##
##            EducationField EnvironmentSatisfaction   Gender      HourlyRate
##  Human Resources : 27   1:284                   Female:588   Min.   : 30.00
##  Life Sciences   :606   2:287                   Male  :882   1st Qu.: 48.00
##  Marketing       :159   3:453                                Median : 66.00
##  Medical         :464   4:446                                Mean   : 65.89
##  Other           : 82                                        3rd Qu.: 83.75
##  Technical Degree:132                                        Max.   :100.00
##
##  JobInvolvement JobLevel                      JobRole    JobSatisfaction
##  1: 83          1:543   Sales Executive          :326   1:289
##  2:375          2:534   Research Scientist       :292   2:280
##  3:868          3:218   Laboratory Technician    :259   3:442
##  4:144          4:106   Manufacturing Director   :145   4:459
##                 5: 69   Healthcare Representative:131
##                         Manager                  :102
##                         (Other)                  :215
##   MaritalStatus MonthlyIncome    MonthlyRate     NumCompaniesWorked OverTime
##  Divorced:327   Min.   : 1009   Min.   : 2094   Min.   :0.000      No :1054
##  Married :673   1st Qu.: 2911   1st Qu.: 8047   1st Qu.:1.000      Yes: 416
##  Single  :470   Median : 4919   Median :14236   Median :2.000
##                 Mean   : 6503   Mean   :14313   Mean   :2.693
##                 3rd Qu.: 8379   3rd Qu.:20462   3rd Qu.:4.000
##                 Max.   :19999   Max.   :26999   Max.   :9.000
##
##  PercentSalaryHike PerformanceRating RelationshipSatisfaction StockOptionLevel
##  Min.   :11.00     3:1244            1:276                    0:631
##  1st Qu.:12.00     4: 226            2:303                    1:596
##  Median :14.00                       3:459                    2:158
##  Mean   :15.21                       4:432                    3: 85
##  3rd Qu.:18.00
##  Max.   :25.00
##
##  TotalWorkingYears TrainingTimesLastYear WorkLifeBalance YearsAtCompany
```

```
##  Min.   : 0.00    Min.   :0.000    1: 80     Min.    : 0.000
##  1st Qu.: 6.00    1st Qu.:2.000    2:344     1st Qu.: 3.000
##  Median :10.00    Median :3.000    3:893     Median : 5.000
##  Mean   :11.28    Mean   :2.799    4:153     Mean    : 7.008
##  3rd Qu.:15.00    3rd Qu.:3.000              3rd Qu.: 9.000
##  Max.   :40.00    Max.   :6.000              Max.    :40.000
##
##  YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager
##  Min.   : 0.000     Min.   : 0.000          Min.   : 0.000
##  1st Qu.: 2.000     1st Qu.: 0.000          1st Qu.: 2.000
##  Median : 3.000     Median : 1.000          Median : 3.000
##  Mean   : 4.229     Mean   : 2.188          Mean   : 4.123
##  3rd Qu.: 7.000     3rd Qu.: 3.000          3rd Qu.: 7.000
##  Max.   :18.000     Max.   :15.000          Max.   :17.000
##
```

```r
# select subset of HR data
hrCat <- hr %>% dplyr::select(c(Attrition, BusinessTravel, Department, Education,
                        EducationField, EnvironmentSatisfaction, Gender,
                        JobInvolvement, JobLevel, JobRole, JobSatisfaction,
                        MaritalStatus, OverTime, PerformanceRating,
                        RelationshipSatisfaction, StockOptionLevel,
                        WorkLifeBalance))

# summary
summary(hrCat)
```

```
##  Attrition          BusinessTravel                   Department   Education
##  No :1233   Non-Travel      : 150   Human Resources       : 63   1:170
##  Yes: 237   Travel_Frequently: 277   Research & Development:961   2:282
##             Travel_Rarely   :1043   Sales                 :446   3:572
##                                                                  4:398
##                                                                  5: 48
##
##
##           EducationField EnvironmentSatisfaction    Gender    JobInvolvement
##  Human Resources : 27    1:284                    Female:588   1: 83
##  Life Sciences   :606    2:287                    Male  :882   2:375
##  Marketing       :159    3:453                                 3:868
##  Medical         :464    4:446                                 4:144
##  Other           : 82
##  Technical Degree:132
##
##  JobLevel                        JobRole    JobSatisfaction  MaritalStatus
##  1:543    Sales Executive          :326    1:289            Divorced:327
##  2:534    Research Scientist       :292    2:280            Married :673
##  3:218    Laboratory Technician    :259    3:442            Single  :470
##  4:106    Manufacturing Director   :145    4:459
##  5: 69    Healthcare Representative:131
##           Manager                  :102
##           (Other)                  :215
##  OverTime   PerformanceRating RelationshipSatisfaction StockOptionLevel
##  No :1054   3:1244            1:276                     0:631
##  Yes: 416   4: 226            2:303                     1:596
##                               3:459                     2:158
```
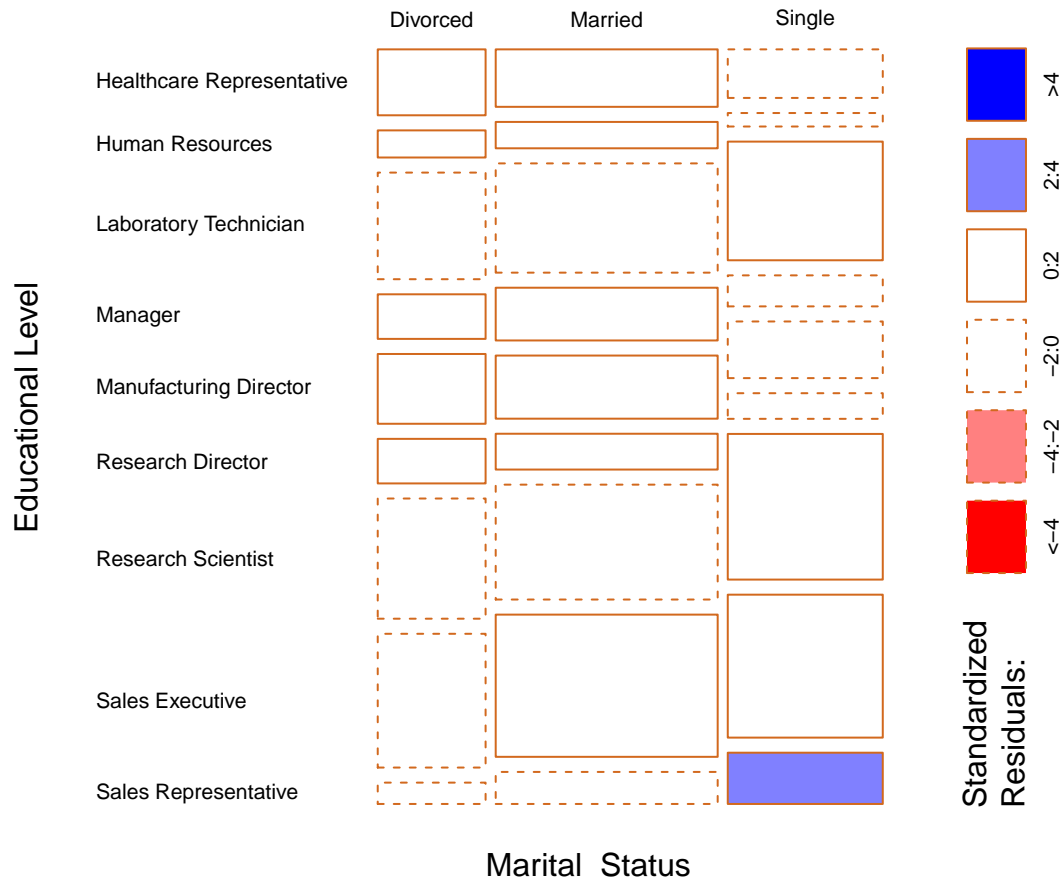
```
##                                      4:432                        3: 85
##
##
##
##   WorkLifeBalance
##   1: 80
##   2:344
##   3:893
##   4:153
##
##
##
```

```r
# number of categories per variable
hrCats = apply(hrCat, 2, function(x) nlevels(as.factor(x)))
# display number of categories per variable
hrCats
```

```
##              Attrition           BusinessTravel               Department
##                      2                        3                        3
##              Education           EducationField    EnvironmentSatisfaction
##                      5                        6                        4
##                 Gender            JobInvolvement                 JobLevel
##                      2                        4                        5
##                JobRole          JobSatisfaction            MaritalStatus
##                      9                        4                        3
##               OverTime        PerformanceRating RelationshipSatisfaction
##                      2                        2                        4
##        StockOptionLevel           WorkLifeBalance
##                      4                        4
```

```r
#######################################################
# mosaic plots
mosaicplot(table(hrCat$MaritalStatus, hrCat$JobRole),
           las = 1, cex.axis = 0.7,
           main = "Marital Status & Job Role\nMosaic Plot",
           xlab = "Marital  Status",
           ylab = "Educational Level",
           border = "chocolate",
           shade = TRUE)
```

Marital Status & Job Role
Mosaic Plot


```r
# correspondence analysis
fitca <-  ca(table(hrCat$MaritalStatus, hrCat$JobRole))
# display the fitca
fitca
```

```
##
##  Principal inertias (eigenvalues):
##            1        2
## Value      0.016462 0.001846
## Percentage 89.92%   10.08%
##
##
##  Rows:
##          Divorced   Married   Single
## Mass     0.222449   0.457823 0.319728
## ChiDist  0.143390   0.078856 0.184532
## Inertia  0.004574   0.002847 0.010887
## Dim. 1  -0.982469 -0.525237 1.435642
## Dim. 2   1.590650 -0.953088 0.258054
##
##
##  Columns:
##          Healthcare Representative Human Resources Laboratory Technician
```

```
## Mass                          0.089116          0.035374                  0.176190
## ChiDist                       0.112382          0.273536                  0.043935
## Inertia                       0.001125          0.002647                  0.000340
## Dim. 1                       -0.818684         -2.114090                  0.342251
## Dim. 2                        0.929853         -0.822125                 -0.032679
##              Manager Manufacturing Director Research Director Research Scientist
## Mass        0.069388               0.098639          0.054422            0.198639
## ChiDist     0.214438               0.076593          0.202027            0.108717
## Inertia     0.003191               0.000579          0.002221            0.002348
## Dim. 1     -1.451081              -0.551643         -1.488517            0.802285
## Dim. 2     -2.476366               0.681343          1.533384            0.814079
##            Sales Executive Sales Representative
## Mass              0.221769             0.056463
## ChiDist           0.026050             0.317917
## Inertia           0.000150             0.005707
## Dim. 1            0.121379             2.431089
## Dim. 2           -0.486022            -1.430606
```

```r
# plot
plot(fitca, main = "Marital Status & Job Role")
```


Marital Status & Job Role

```
####################################################
# mosiac plot
mosaicplot(table(hrCat$JobLevel, hrCat$EducationField),
           las = 1, cex.axis = 0.7,
           main = "Job Level & Education Field",
           xlab = "Job Level",
           ylab = "Education Field",
           border = "chocolate",
           shade = TRUE)
```

Job Level & Education Field
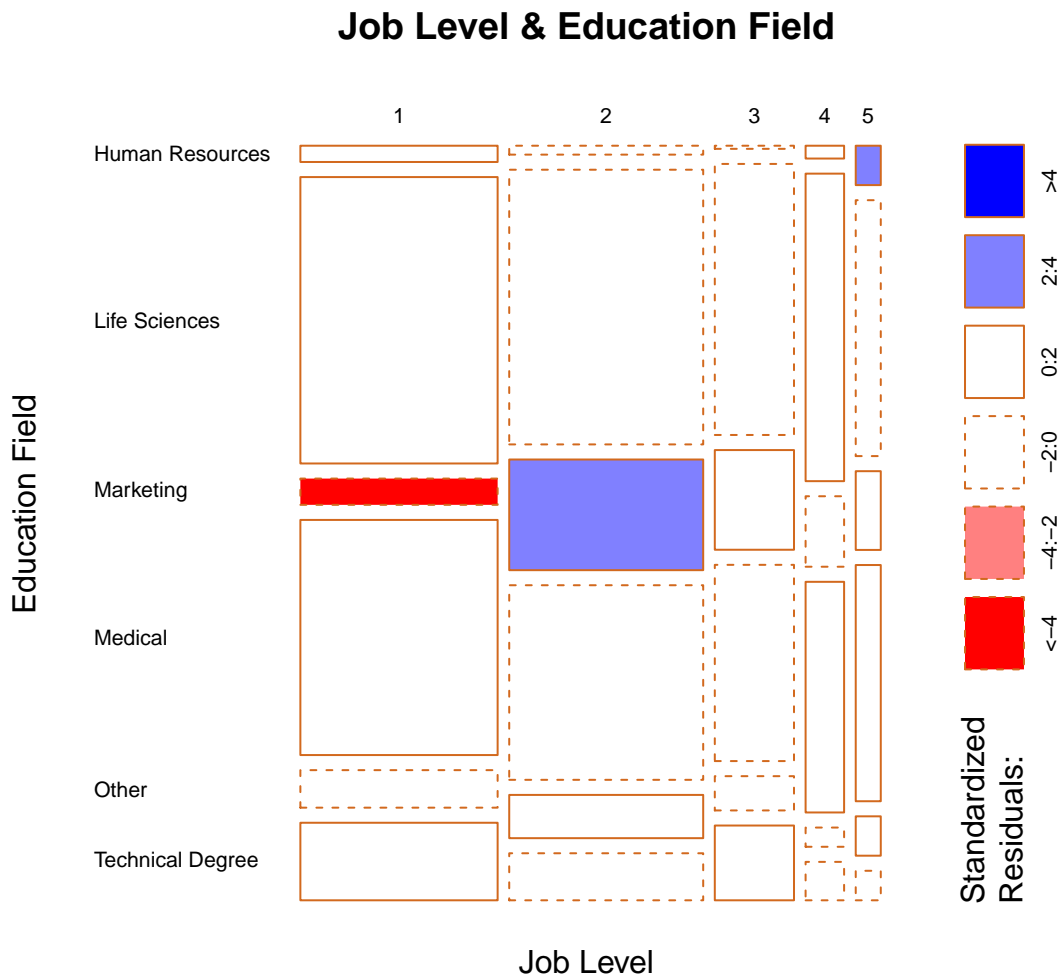

```
# correspondence analysis
fitca <-  ca(table(hrCat$JobLevel, hrCat$EducationField))
# display the fitca
fitca
```

```
##
##  Principal inertias (eigenvalues):
##             1        2        3        4
## Value      0.036199 0.007981 0.002138 0.000655
## Percentage 77.06%   16.99%   4.55%    1.39%
##
##
##  Rows:
##                 1         2         3         4         5
```

```
## Mass      0.369388   0.363265   0.148299 0.072109 0.046939
## ChiDist   0.236943   0.194576   0.178602 0.178369 0.340963
## Inertia   0.020738   0.013753   0.004731 0.002294 0.005457
## Dim. 1     1.239885  -1.013402  -0.630235 0.013013 0.056681
## Dim. 2    -0.205697   0.124101  -1.382895 0.867197 3.695243
##
##
##   Columns:
##          Human Resources Life Sciences Marketing  Medical      Other
## Mass             0.018367        0.412245  0.108163 0.315646   0.055782
## ChiDist          0.604607        0.038998  0.514623 0.092314   0.161929
## Inertia          0.006714        0.000627  0.028646 0.002690   0.001463
## Dim. 1           1.683314        0.106739 -2.700688 0.455048  -0.251642
## Dim. 2           5.525163       -0.053731  0.093326 0.316003   0.029929
##          Technical Degree
## Mass             0.089796
## ChiDist          0.275868
## Inertia          0.006834
## Dim. 1           0.975520
## Dim. 2          -2.125277
```

```r
# plot
plot(fitca, main = "Job Level & Education Field")
```

## Job Level & Education Field
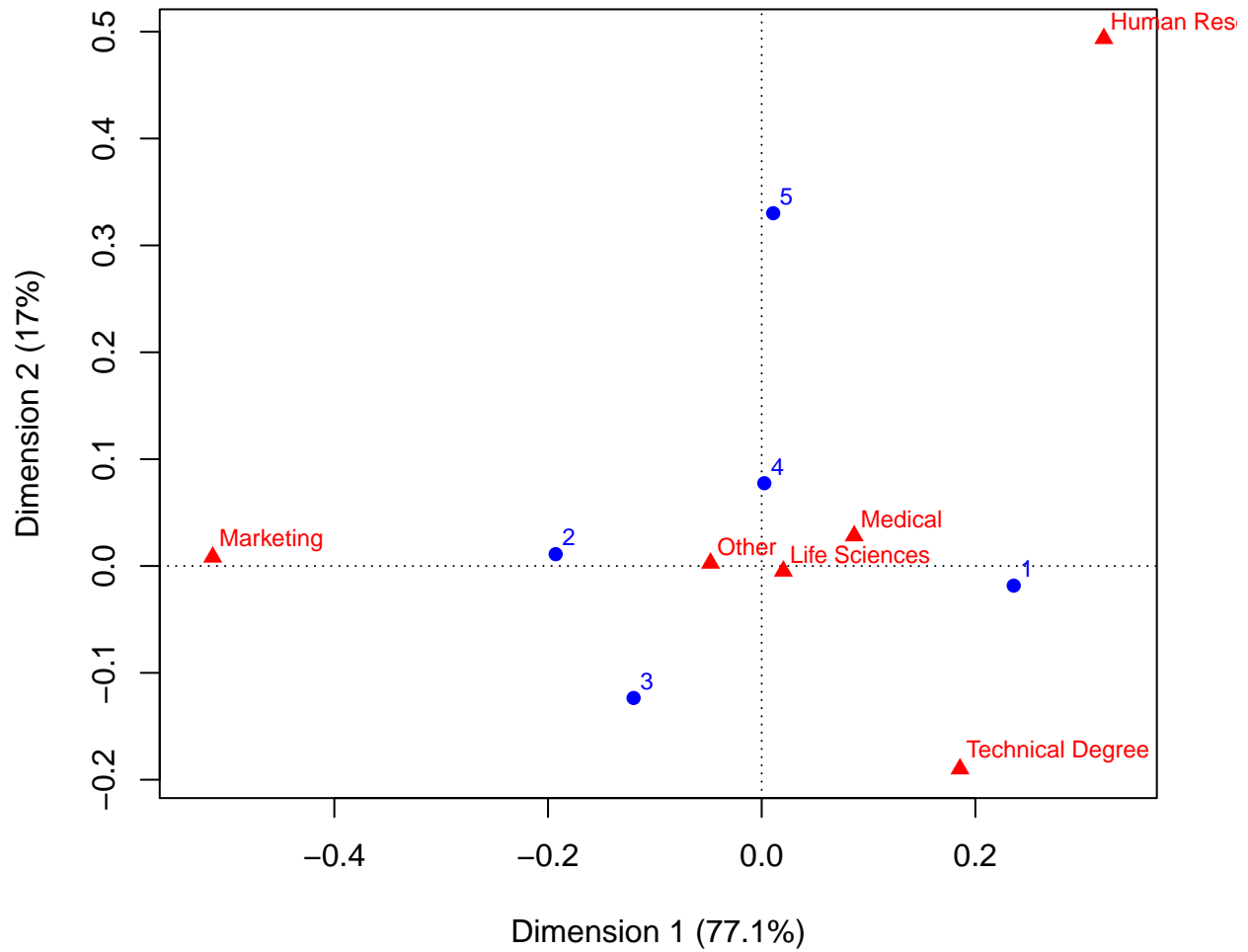

```
# mosaic plot
mosaicplot(table(hrCat$MaritalStatus, hrCat$JobRole),
          las = 1, cex.axis = 0.7,
          main = "Marital Status & Job Role",
          xlab = "Marital  Status",
          ylab = "Educational Level",
          border = "chocolate",
          shade = TRUE)
```
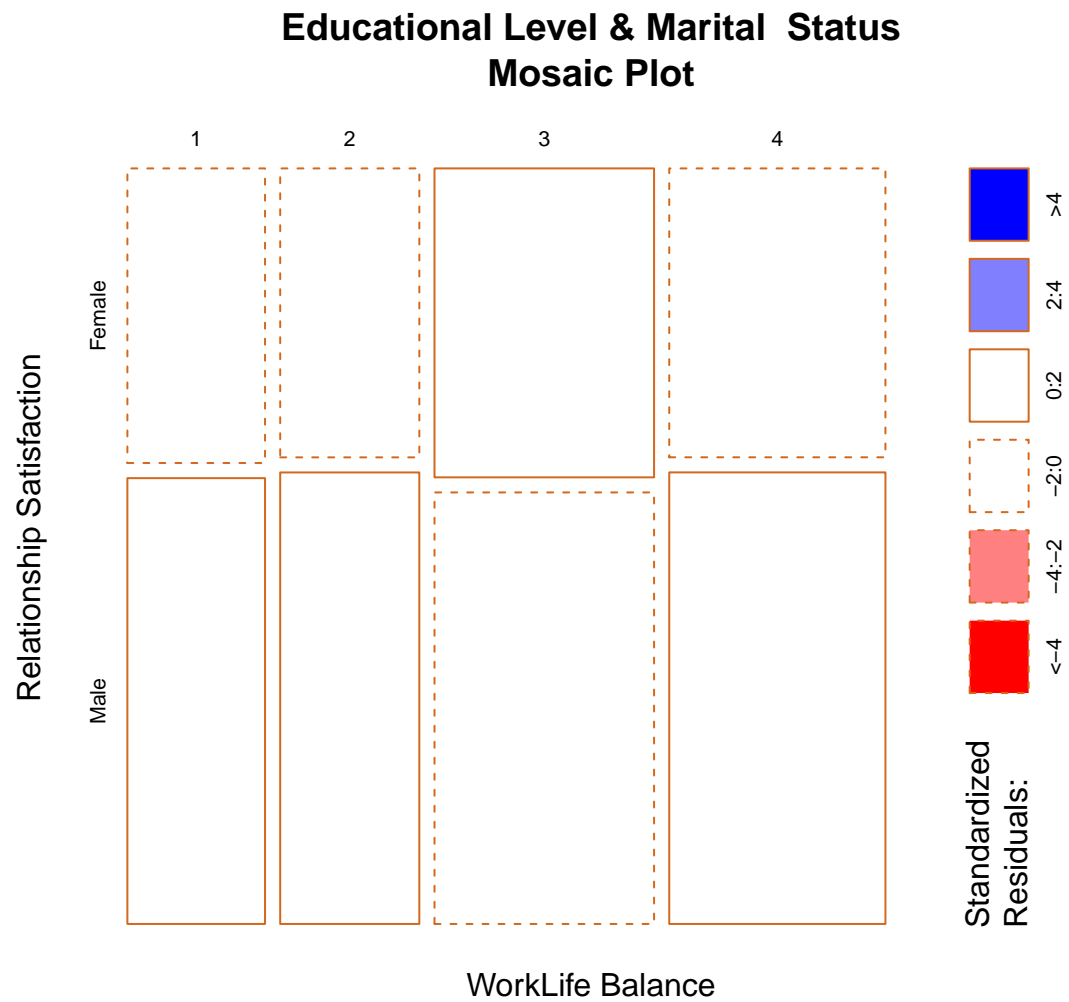
Marital Status & Job Role
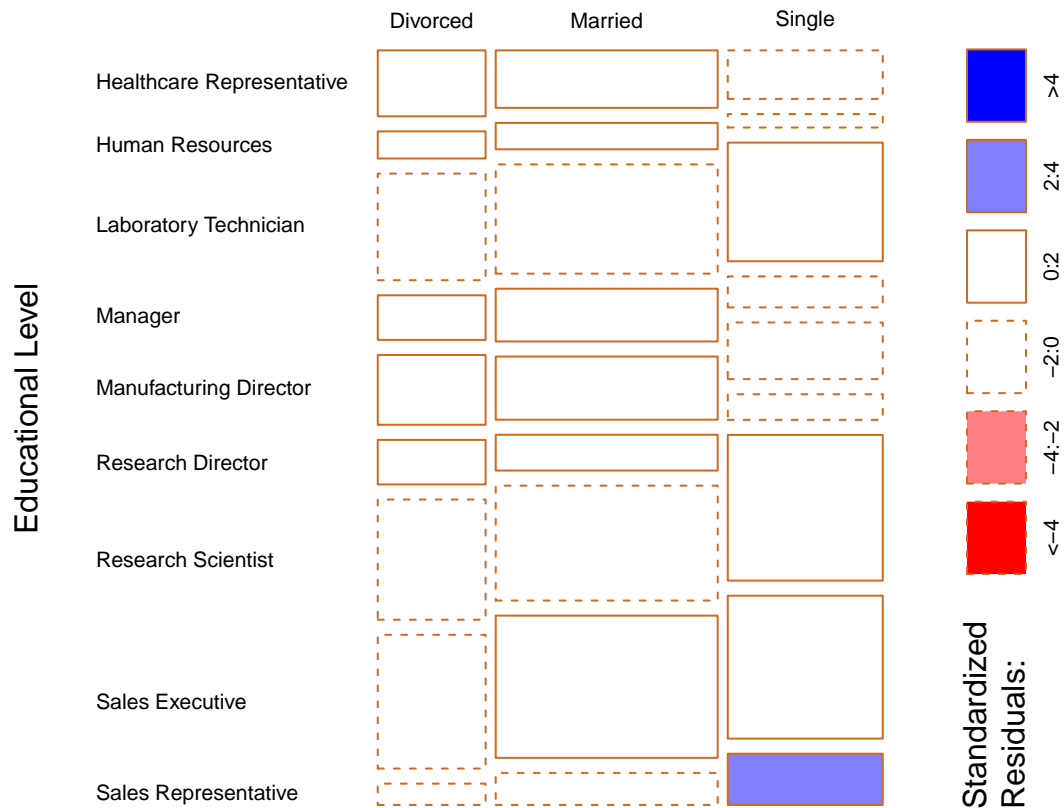

```r
mosaicplot(table(hrCat$WorkLifeBalance, hrCat$Department),
          main = "Educational Level & Marital  Status\nMosaic Plot",
          xlab = "WorkLife Balance",
          ylab = "Relationship Satisfaction",
          border = "chocolate",
          shade = TRUE)
```

Educational Level & Marital Status
Mosaic Plot
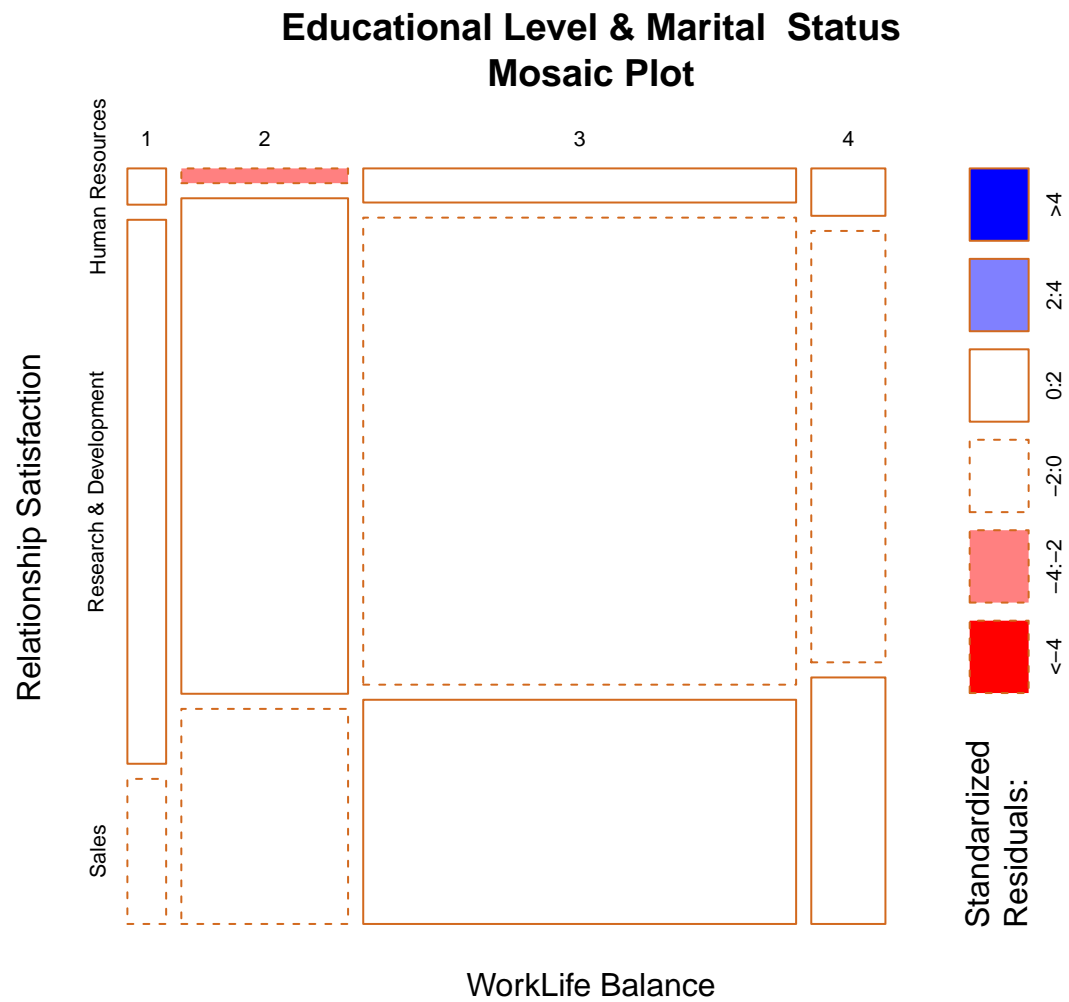

```
mosaicplot(table(hrCat$EnvironmentSatisfaction, hrCat$Gender),
           main = "Educational Level & Marital  Status\nMosaic Plot",
           xlab = "WorkLife Balance",
           ylab = "Relationship Satisfaction",
           border = "chocolate",
           shade = TRUE)
```

Educational Level & Marital Status
Mosaic Plot

Relationship Satisfaction

	1	2	3	4
Female				
Male				

Standardized Residuals:

|	
■ (blue)	>4
■ (light blue)	2:4
□	0:2
⌷ (dashed)	−2:0
▨ (pink)	−4:−2
▨ (red)	<−4

WorkLife Balance

```r
# mosaic plot
mosaicplot(table(hrCat$Attrition, hr$BusinessTravel),
        main = "Business Travel & Attrition\nMosaic Plot",
        xlab = "Attrition",
        ylab = "Business Travel",
        border = "chocolate",
        shade = TRUE)
```

Business Travel & Attrition
Mosaic Plot


```r
# mosaic plot
mosaicplot(table(hrCat$Attrition, hr$BusinessTravel),
          main = "Business Travel & Attrition\nMosaic Plot",
          xlab = "Attrition",
          ylab = "Business Travel",
          border = "chocolate",
          shade = TRUE)


mosaicplot(table(hrCat$Attrition, hr$Department),
          main = "Department & Attrition\nMosaic Plot",
          xlab = "Attrition",
          ylab = "Department",
          border = "chocolate",
          shade = TRUE)
```

Department & Attrition
Mosaic Plot
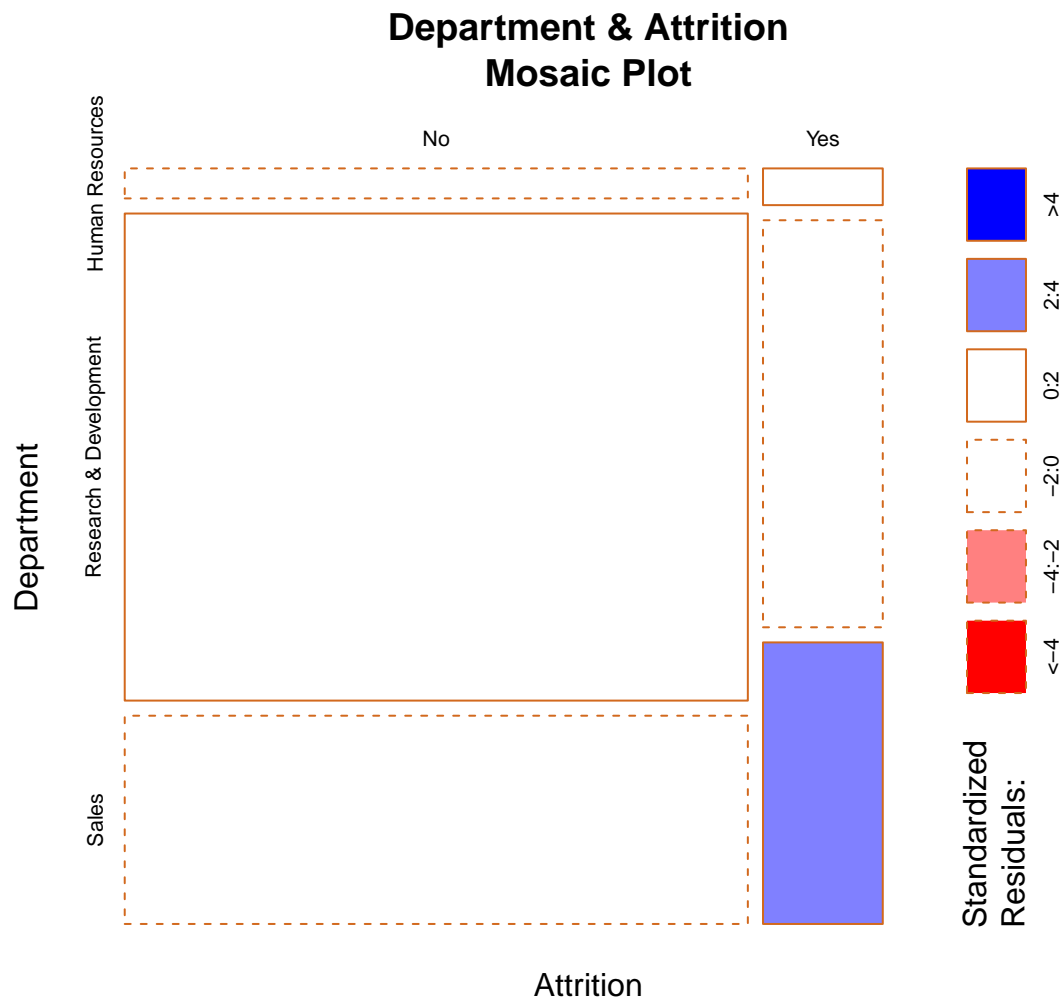

```
mosaicplot(table(hrCat$Attrition, hr$Education),
          main = "Education & Attrition\nMosaic Plot",
          xlab = "Attrition",
          ylab = "Education",
          border = "chocolate",
          shade = TRUE)
```

Education & Attrition
Mosaic Plot
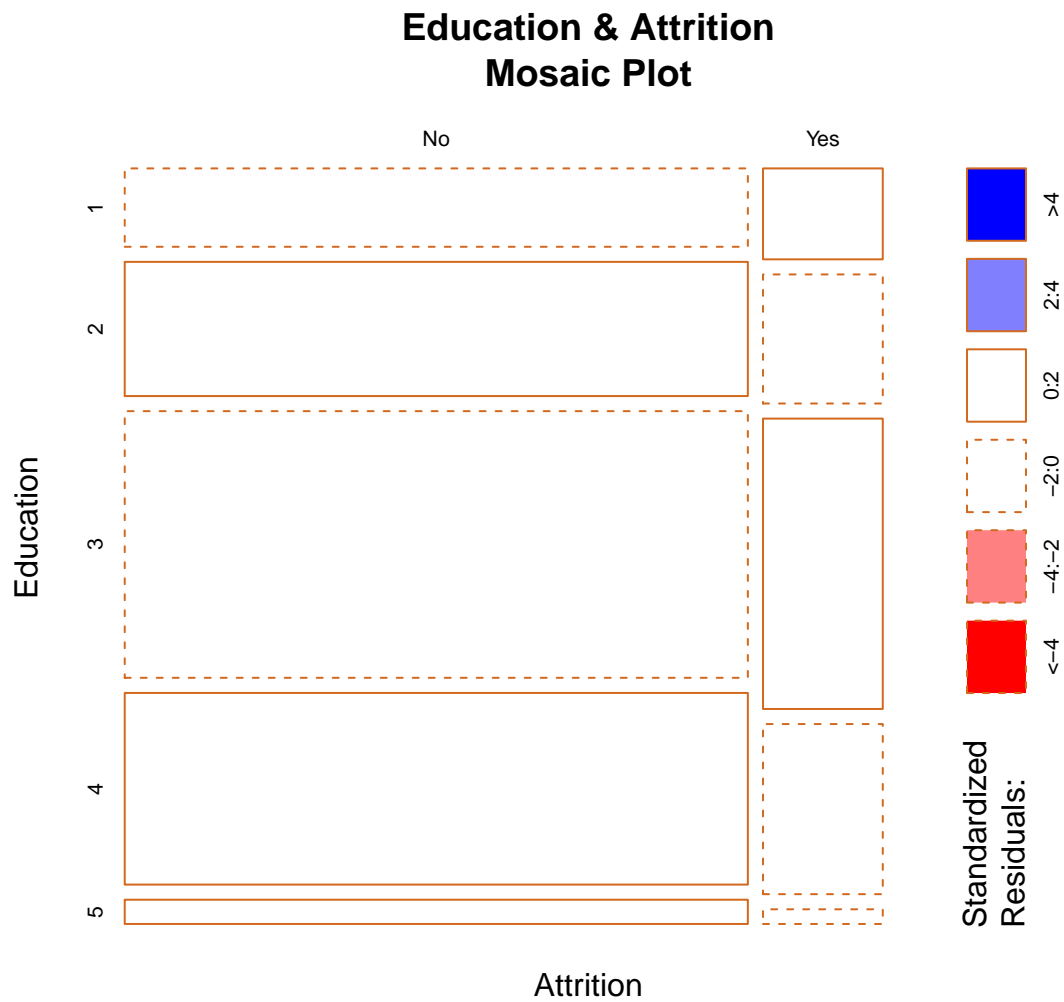

```
mosaicplot(table(hrCat$Attrition, hr$Education, hr$Gender),
           main = "Education & Attrition\nMosaic Plot",
           xlab = "Attrition",
           ylab = "Education",
           border = "chocolate",
           shade = TRUE)
```

Education & Attrition
Mosaic Plot


```
# correspondence analysis
fitca <-  ca(table(hrCat$MaritalStatus, hrCat$JobRole))

# display the fitca
fitca
```
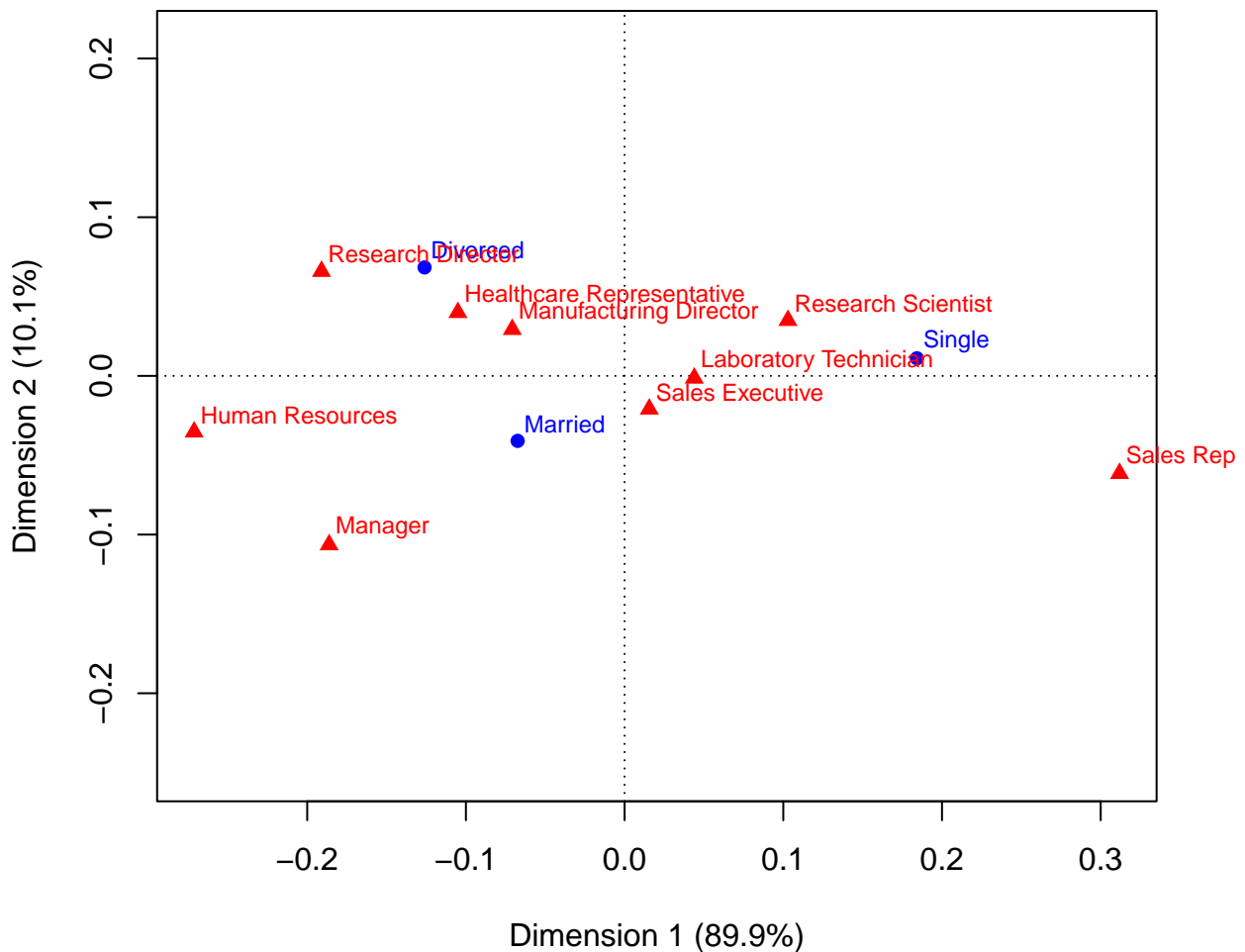
```
##
##  Principal inertias (eigenvalues):
##            1          2
## Value      0.016462 0.001846
## Percentage 89.92%   10.08%
##
##
##  Rows:
##          Divorced   Married   Single
## Mass     0.222449  0.457823 0.319728
## ChiDist  0.143390  0.078856 0.184532
## Inertia  0.004574  0.002847 0.010887
## Dim. 1  -0.982469 -0.525237 1.435642
## Dim. 2   1.590650 -0.953088 0.258054
##
##
##  Columns:
```

```
##         Healthcare Representative Human Resources Laboratory Technician
## Mass                    0.089116         0.035374              0.176190
## ChiDist                 0.112382         0.273536              0.043935
## Inertia                 0.001125         0.002647              0.000340
## Dim. 1                 -0.818684        -2.114090              0.342251
## Dim. 2                  0.929853        -0.822125             -0.032679
##          Manager Manufacturing Director Research Director Research Scientist
## Mass    0.069388              0.098639          0.054422           0.198639
## ChiDist 0.214438              0.076593          0.202027           0.108717
## Inertia 0.003191              0.000579          0.002221           0.002348
## Dim. 1 -1.451081             -0.551643         -1.488517           0.802285
## Dim. 2 -2.476366              0.681343          1.533384           0.814079
##         Sales Executive Sales Representative
## Mass           0.221769             0.056463
## ChiDist        0.026050             0.317917
## Inertia        0.000150             0.005707
## Dim. 1         0.121379             2.431089
## Dim. 2        -0.486022            -1.430606
```

```r
# plot
plot(fitca, main = "Marital  Status & Job Role")
```

Marital Status & Job Role

```r
# correspondence analysis
fitca <-  ca(table(hrCat$MaritalStatus,hrCat$WorkLifeBalance))

# display the fitca
fitca
```

```
##
##  Principal inertias (eigenvalues):
##           1        2
## Value      0.002866 0.000568
## Percentage 83.46%   16.54%
##
##
##  Rows:
##          Divorced   Married    Single
## Mass     0.222449  0.457823  0.319728
## ChiDist  0.099607  0.040006  0.039306
## Inertia  0.002207  0.000733  0.000494
## Dim. 1   1.858347 -0.635489 -0.382968
## Dim. 2  -0.204844 -0.883404  1.407478
##
##
##  Columns:
##                  1         2         3         4
## Mass     0.054422  0.234014  0.607483  0.104082
## ChiDist  0.161791  0.081650  0.021078  0.041490
## Inertia  0.001425  0.001560  0.000270  0.000179
## Dim. 1  -2.826698  1.458444 -0.197598 -0.647804
## Dim. 2  -2.402253 -1.002499  0.765158 -0.955858
```

```r
# plot
plot(fitca, main = "Marital Status & WorkLife Balance")
```

Marital Status & WorkLife Balance
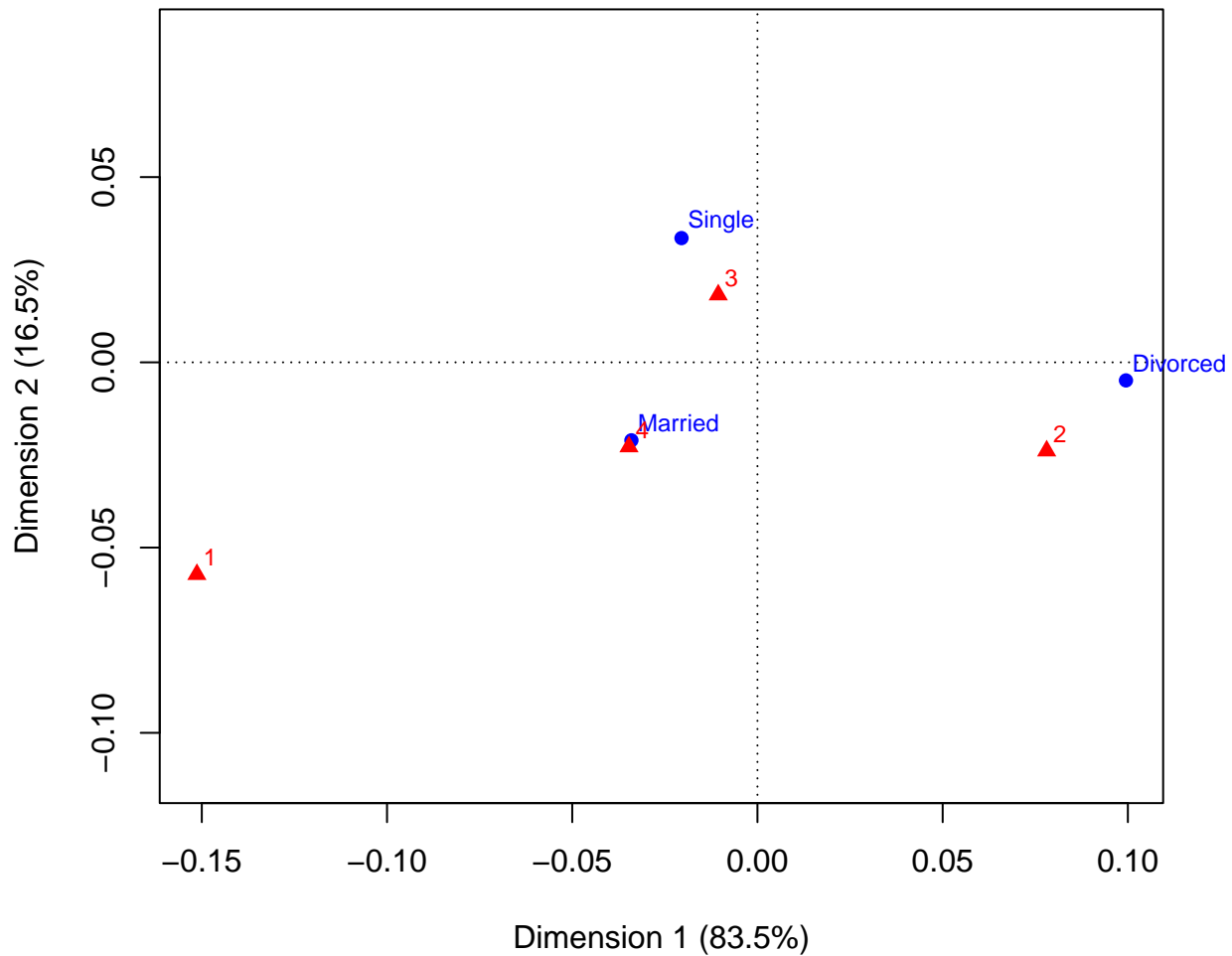

```
#######
# MCA #
#######
# select these columns
newHR <- hr[, c("Attrition", "BusinessTravel")]

# number of categories per variable
cats = apply(newHR, 2, function(x) nlevels(as.factor(x)))

cats

##      Attrition BusinessTravel
##              2              3
# apply MCA
mHR = MCA(newHR, graph = FALSE)

# list of results
mHR

## **Results of the Multiple Correspondence Analysis (MCA)**
## The analysis was performed on 1470 individuals, described by 2 variables
```

```
## *The results are available in the following objects:
##
##    name                description
## 1  "$eig"              "eigenvalues"
## 2  "$var"              "results for the variables"
## 3  "$var$coord"        "coord. of the categories"
## 4  "$var$cos2"         "cos2 for the categories"
## 5  "$var$contrib"      "contributions of the categories"
## 6  "$var$v.test"       "v-test for the categories"
## 7  "$ind"              "results for the individuals"
## 8  "$ind$coord"        "coord. for the individuals"
## 9  "$ind$cos2"         "cos2 for the individuals"
## 10 "$ind$contrib"      "contributions of the individuals"
## 11 "$call"             "intermediate results"
## 12 "$call$marge.col"   "weights of columns"
## 13 "$call$marge.li"    "weights of rows"
```

```r
# table of eigenvalues
mHR$eig
```

```
##       eigenvalue percentage of variance cumulative percentage of variance
## dim 1    0.56413               37.60867                          37.60867
## dim 2    0.50000               33.33333                          70.94200
## dim 3    0.43587               29.05800                         100.00000
```

```r
# data frame with variable coordinates
mHR_vars_df = data.frame(mHR$var$coord, Variable = rep(names(cats), cats))

# data frame with observation coordinates
mHR_obs_df = data.frame(mHR$ind$coord)

# plot of variable categories
ggplot(data=mHR_vars_df,
       aes(x = Dim.1, y = Dim.2,label = rownames(mHR_vars_df))) +
 geom_hline(yintercept = 0, colour = "gray70") +
 geom_vline(xintercept = 0, colour = "gray70") +
 geom_text(aes(colour=Variable)) +
 ggtitle("MCA plot of variables using R package FactoMineR")
```

MCA plot of variables using R package FactoMineR
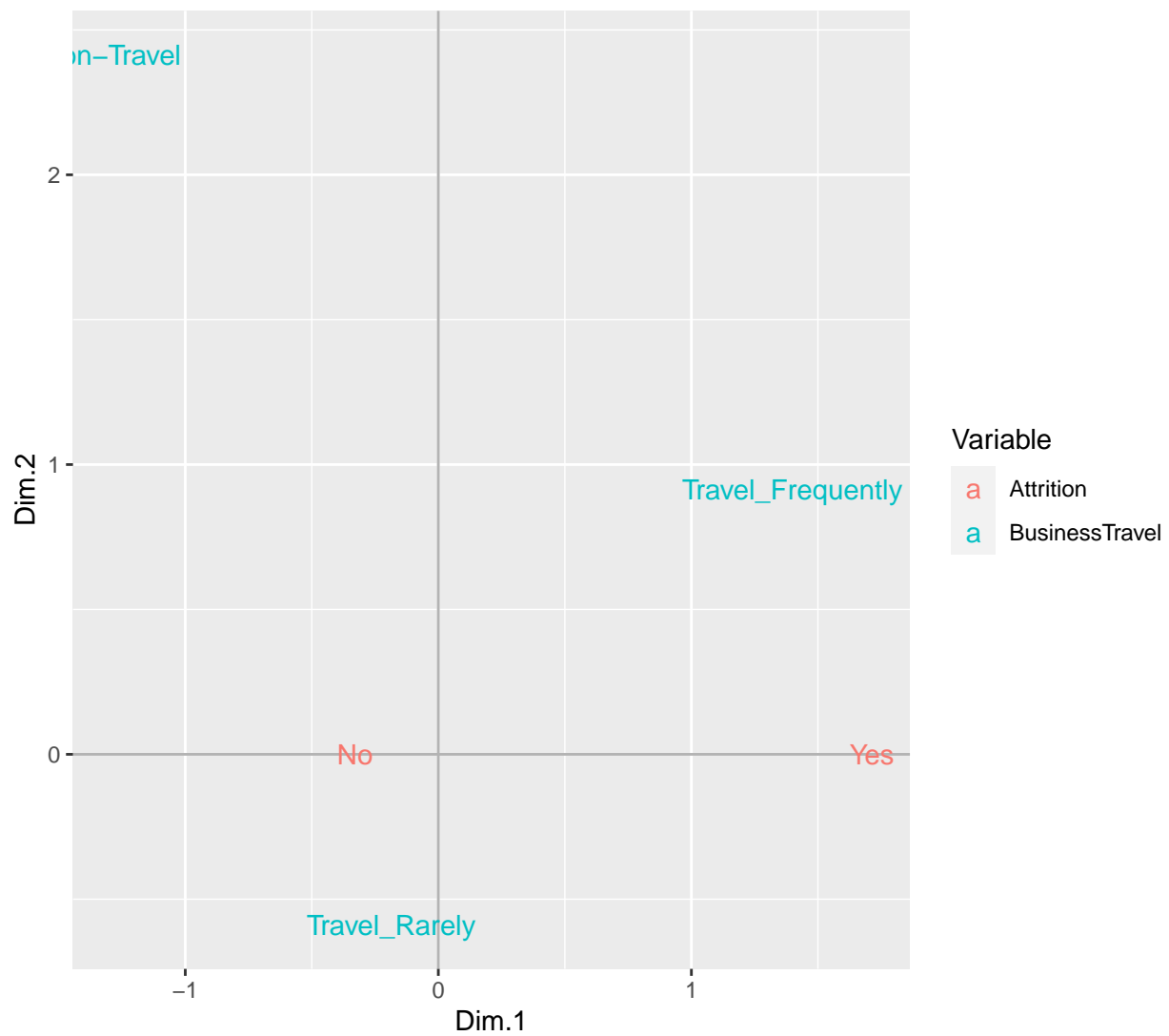

```r
table(hr$Attrition, hr$BusinessTravel)
```
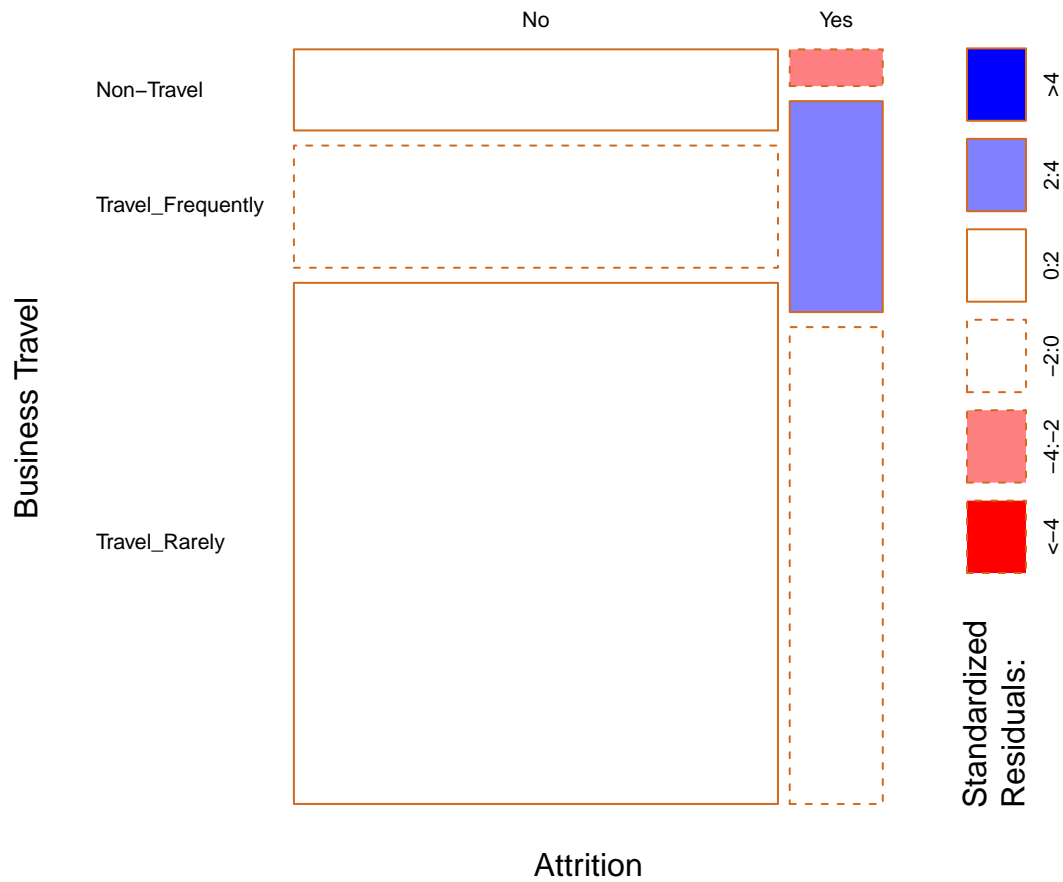
```
##
##         Non-Travel Travel_Frequently Travel_Rarely
##   No           138               208           887
##   Yes           12                69           156
```

```r
mosaicplot(table(hrCat$Attrition, hr$BusinessTravel),
           las = 1, cex.axis = 0.7,
           main = "Business Travel & Attrition\nMosaic Plot",
           xlab = "Attrition",
           ylab = "Business Travel",
           border = "chocolate",
           shade = TRUE)
```

Business Travel & Attrition
Mosaic Plot


```
# MCA plot of observations and categories
ggplot(data = mHR_obs_df, aes(x = Dim.1, y = Dim.2)) +
  geom_hline(yintercept = 0, colour = "gray70") +
  geom_vline(xintercept = 0, colour = "gray70") +
  geom_point(colour = "gray50", alpha = 0.7) +
  geom_density2d(colour = "gray80") +
  geom_text(data = mHR_vars_df,
            aes(x = Dim.1, y = Dim.2,
                label = rownames(mHR_vars_df), colour = Variable)) +
  ggtitle("MCA plot of variables using R package FactoMineR") +
  scale_colour_discrete(name = "Variable")
```

MCA plot of variables using R package FactoMineR
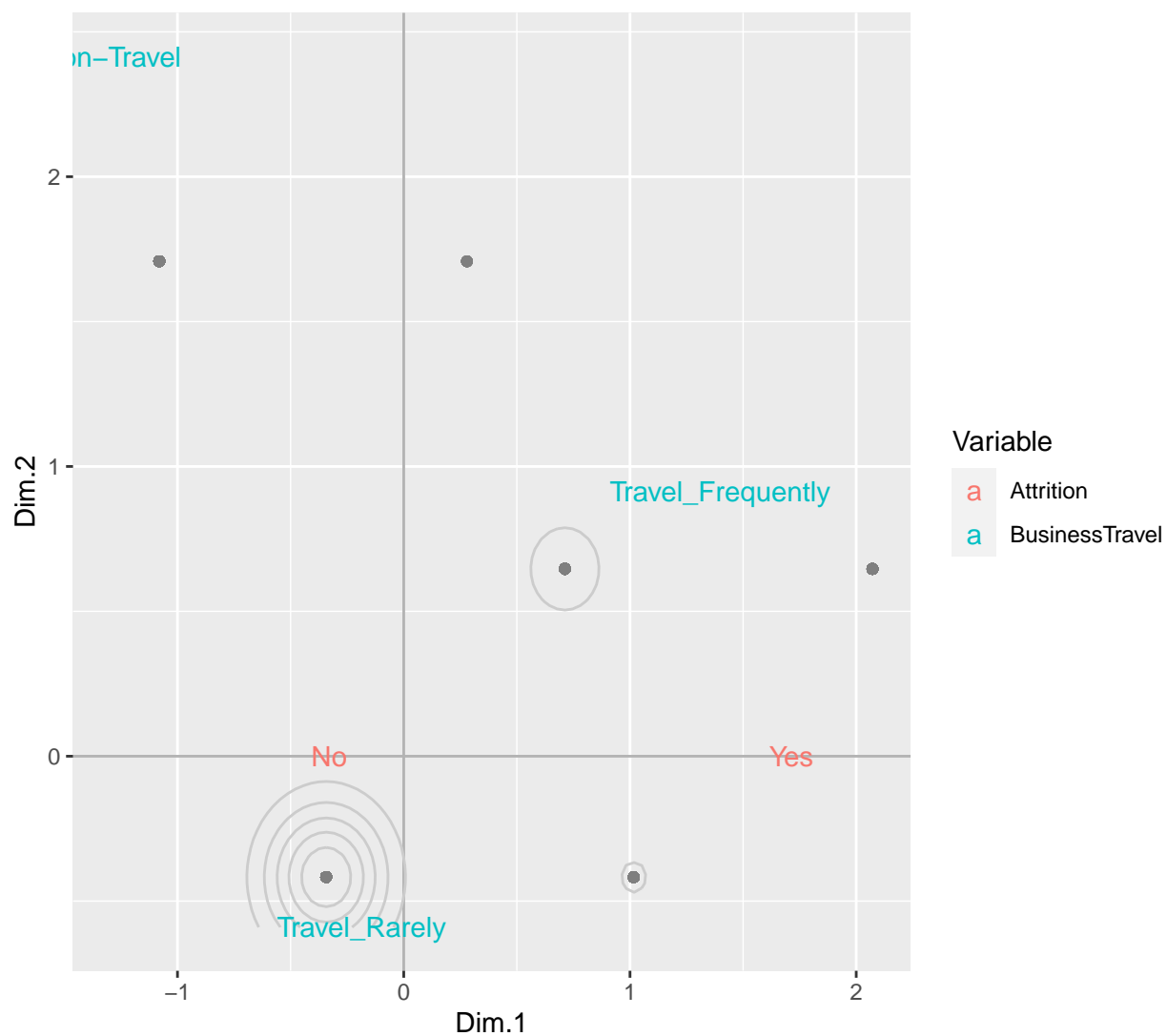

```r
# load data tea
data(tea)



# select these columns
newtea = tea[, c("Tea", "How", "how", "sugar", "where", "always")]

# number of categories per variable
cats = apply(newtea, 2, function(x) nlevels(as.factor(x)))


cats

##     Tea    How    how  sugar  where always
##       3      4      3      2      3      2
# apply MCA
mca1 = MCA(newtea, graph = FALSE)
```

```r
# list of results
mca1
```

```
## **Results of the Multiple Correspondence Analysis (MCA)**
## The analysis was performed on 300 individuals, described by 6 variables
## *The results are available in the following objects:
##
##     name               description
## 1   "$eig"             "eigenvalues"
## 2   "$var"             "results for the variables"
## 3   "$var$coord"       "coord. of the categories"
## 4   "$var$cos2"        "cos2 for the categories"
## 5   "$var$contrib"     "contributions of the categories"
## 6   "$var$v.test"      "v-test for the categories"
## 7   "$ind"             "results for the individuals"
## 8   "$ind$coord"       "coord. for the individuals"
## 9   "$ind$cos2"        "cos2 for the individuals"
## 10  "$ind$contrib"     "contributions of the individuals"
## 11  "$call"            "intermediate results"
## 12  "$call$marge.col"  "weights of columns"
## 13  "$call$marge.li"   "weights of rows"
```

```r
# table of eigenvalues
mca1$eig
```

```
##         eigenvalue percentage of variance cumulative percentage of variance
## dim 1   0.27976178              15.259733                          15.25973
## dim 2   0.25774772              14.058967                          29.31870
## dim 3   0.22013794              12.007524                          41.32622
## dim 4   0.18792961              10.250706                          51.57693
## dim 5   0.16876495               9.205361                          60.78229
## dim 6   0.16368666               8.928363                          69.71065
## dim 7   0.15288834               8.339364                          78.05002
## dim 8   0.13838682               7.548372                          85.59839
## dim 9   0.11569167               6.310455                          91.90885
## dim 10  0.08612637               4.697802                          96.60665
## dim 11  0.06221147               3.393353                         100.00000
```

```r
# data frame with variable coordinates
mca1_vars_df = data.frame(mca1$var$coord, Variable = rep(names(cats), cats))

# data frame with observation coordinates
mca1_obs_df = data.frame(mca1$ind$coord)

# plot of variable categories
ggplot(data=mca1_vars_df,
       aes(x = Dim.1, y = Dim.2, label = rownames(mca1_vars_df))) +
 geom_hline(yintercept = 0, colour = "gray70") +
 geom_vline(xintercept = 0, colour = "gray70") +
 geom_text(aes(colour=Variable)) +
 ggtitle("MCA plot of variables using R package FactoMineR")
```

MCA plot of variables using R package FactoMineR

Variable
- a always
- a how
- a How
- a sugar
- a Tea
- a where