Task 4

```
##############################################################################
#  SETTING DIRECTORY AND UPLOADING DATA
##############################################################################
base_url <- '~/GitHub/kul-multivariate-a2/src/'
load(paste0(base_url,"datacar.Rdata"))

##############################################################################
# LIBRARIES NEEDED
##############################################################################
#install.packages("ca")
library(ca)
library(openxlsx)

##############################################################################
# Are X and Y statistically dependent? --> CHI-SQUARE TEST
##############################################################################
chisq.test(datacar)
```

```
##
##  Pearson's Chi-squared test
##
## data:  datacar
## X-squared = 778.2, df = 126, p-value < 2.2e-16
```

```
# Create row profiles table:

row_p = prop.table(data.matrix (datacar), margin=1)

mean = apply(row_p,2,mean)
row_profiles = rbind(row_p,mean)

write.xlsx(row_profiles,file="rows.xlsx")

# Create column profiles table:

col_p = prop.table(data.matrix (datacar), margin=2)

mean = apply(col_p,1,mean)
col_profiles = cbind(col_p,mean)

write.xlsx(col_profiles,file="columns.xlsx")


# P-value < alpha at all significant levels --> reject H0
# where H0: X and Y are not statistically dependent
```

```r
# Consequently, it makes sense to use Correspondence Analysis

##############################################################################
# CORRESPONDENCE ANALYSIS
##############################################################################

ca.out = ca(datacar)
summary(ca.out)
```

```
##
## Principal inertias (eigenvalues):
##
##  dim    value      %   cum%   scree plot
##  1      0.032622  61.7  61.7  ***************
##  2      0.017765  33.6  95.4  ********
##  3      0.000985   1.9  97.2
##  4      0.000727   1.4  98.6
##  5      0.000380   0.7  99.3
##  6      0.000182   0.3  99.7
##  7      0.000118   0.2  99.9
##  8      3.9e-050   0.1 100.0
##  9      2e-05000   0.0 100.0
##         -------- -----
##  Total: 0.052838 100.0
##
##
## Rows:
##      name   mass  qlt  inr    k=1 cor ctr    k=2 cor ctr
## 1  | BMW3 |   98  969   69 | -184 915 102 |  -45  54  11 |
## 2  | Frdx |  108  959  122 |  136 307  61 |  198 652 237 |
## 3  | iJ30 |   91  943   38 |  -43  84   5 | -138 859  97 |
## 4  | JpGC |  111  986  117 |  108 208  39 |  208 778 270 |
## 5  | LES3 |  104  916   57 |  -64 141  13 | -150 775 132 |
## 6  | ChTC |   76  981  210 |  371 942 321 |  -75  39  24 |
## 7  | MC28 |  102  940   33 |  -56 181  10 | -114 759  75 |
## 8  | S900 |  103  541   19 |  -71 523  16 |   13  18   1 |
## 9  | PrsB |  112  979  223 | -296 834 301 |  123 145  96 |
## 10 | VV90 |   94  899  111 |  214 731 132 | -103 169  56 |
##
## Columns:
##      name   mass  qlt  inr    k=1 cor ctr    k=2 cor ctr
## 1  | exct |   64  949   58 | -201 849  79 |   69 100  17 |
## 2  | dpnd |   73  876   39 |   78 214  14 | -138 662  78 |
## 3  | lxrs |   69  921   36 |  -97 351  20 | -124 570  60 |
## 4  | otdr |   53  987  139 |  181 240  54 |  320 747 309 |
## 5  | pwrf |   72  781    8 |  -57 556   7 |   36 225   5 |
## 6  | styl |   69  966   48 | -189 960  75 |  -15   6   1 |
## 7  | cmfr |   74  890   29 |   85 349  16 | -105 540  46 |
## 8  | rggd |   51  980  115 |  188 295  55 |  287 685 234 |
## 9  |  fun |   69  941   47 | -175 856  66 |   55  85  12 |
## 10 | safe |   75  919   53 |  132 463  40 | -131 457  72 |
## 11 | prfr |   69  888   49 | -168 744  59 |  -74 144  21 |
## 12 | fmly |   69  986  201 |  375 912 297 | -107  74  44 |
```

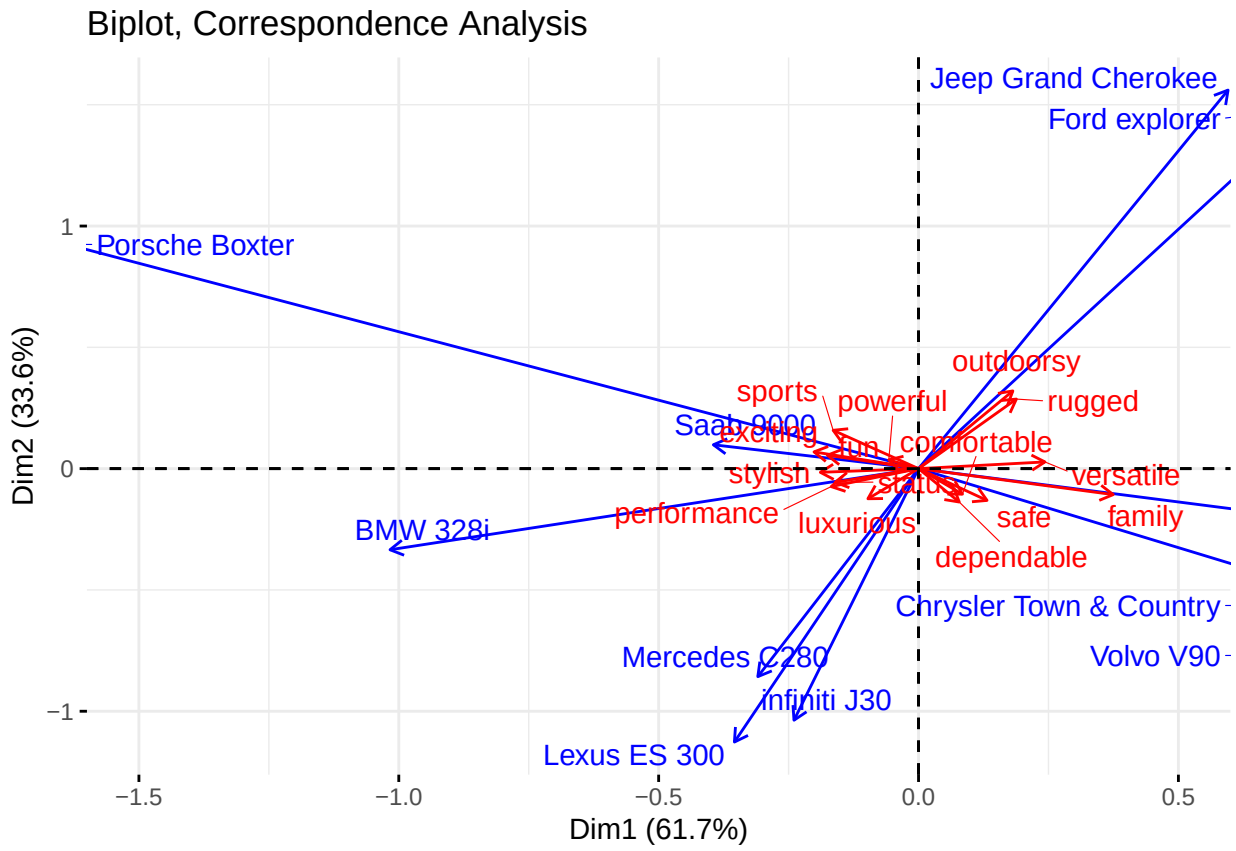
```
## 13 | vrst |   62  968   72 |  243 955 112 |   28  12   3 |
## 14 | sprt |   60  919   65 | -164 477  50 |  158 442  85 |
## 15 | stts |   71  945   41 | -160 842  56 |  -56 104  13 |
```

```
# First two dimensions FINE
# Rows: every car but s900 has qlt>890
# Columns: pwrf (qlt=781)

#install.packages("factoextra")
library(ggplot2)
library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

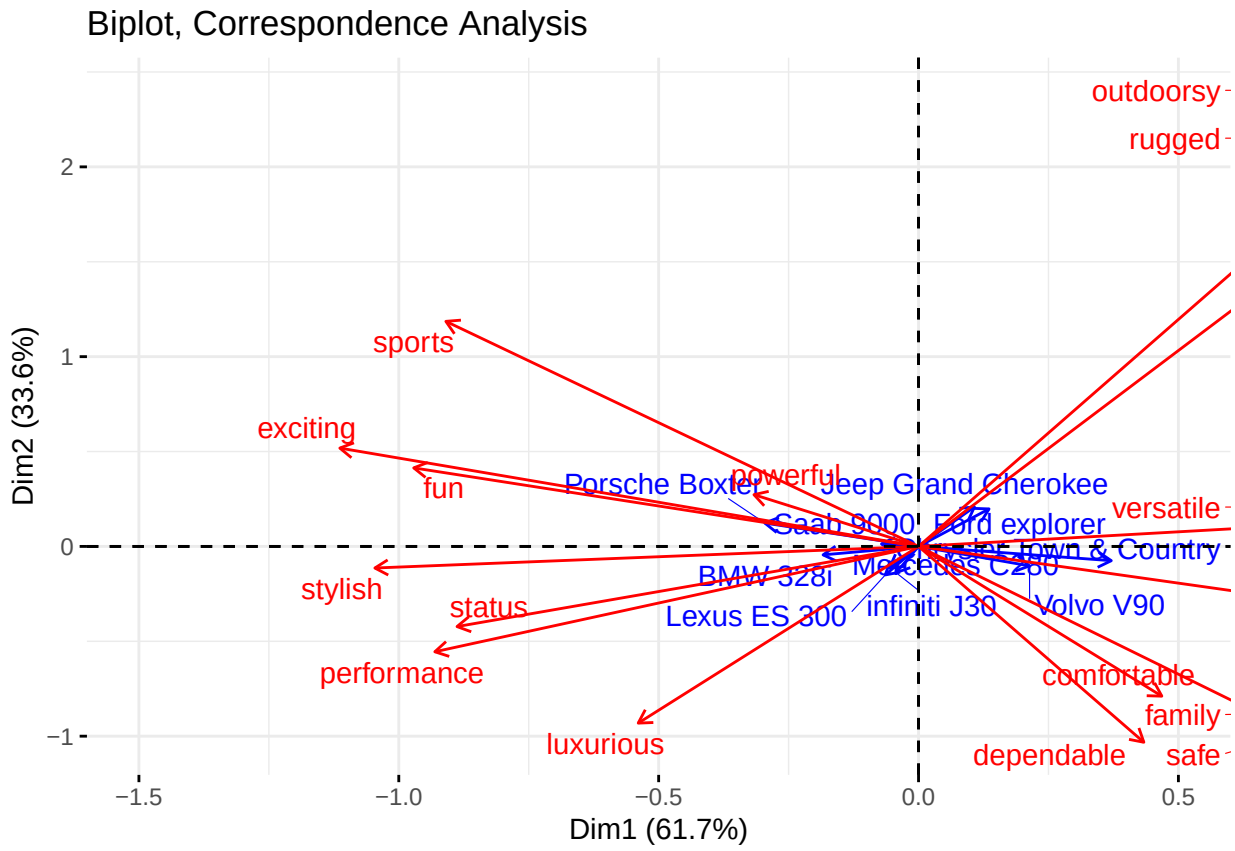
```
library(ggplot2)

# COLPRINCIPAL
fviz_ca_biplot(ca.out, repel=TRUE,
               title = "Biplot, Correspondence Analysis",
               arrows=c(TRUE,TRUE), map="colprincipal",
               labelsize=4, xlim=c(-1.5,0.5))
```

```
## Coordinate system already present. Adding new coordinate system, which will replace the existing one
```



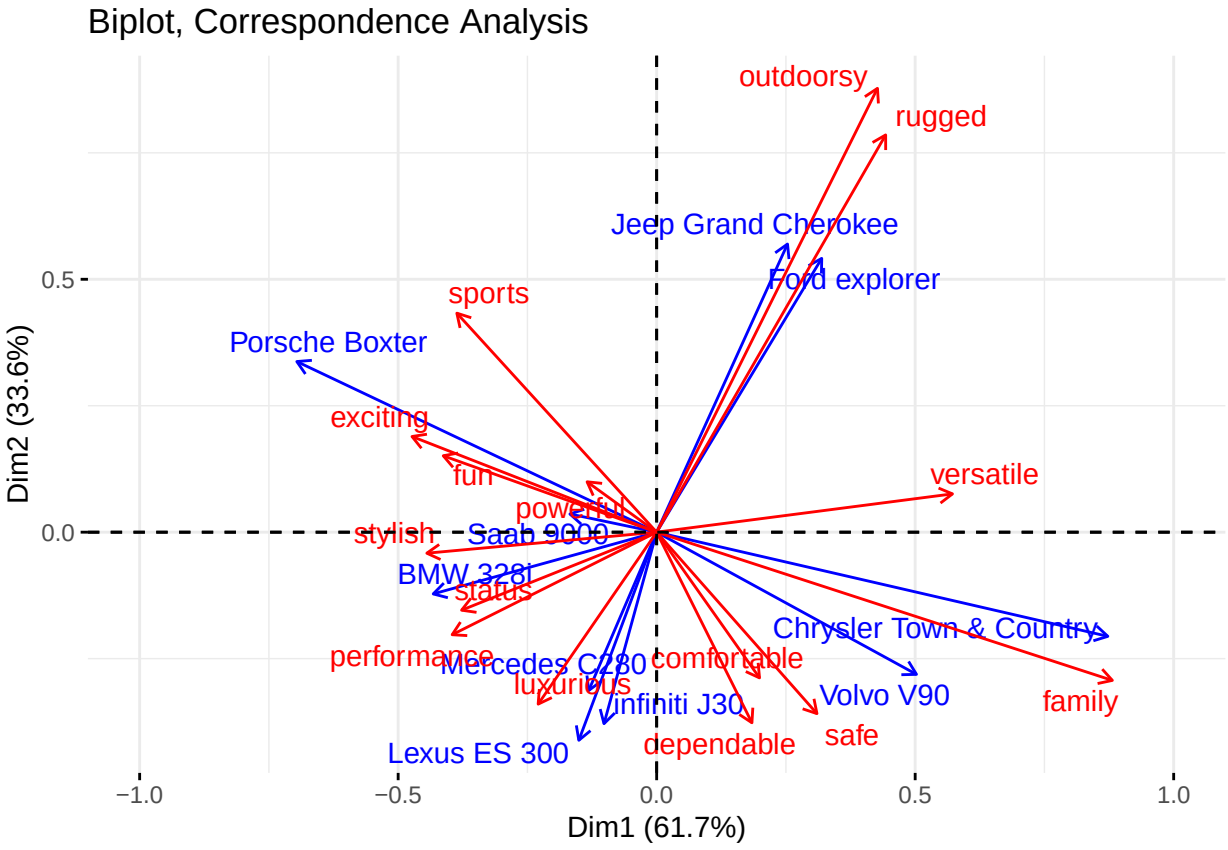
```
#ROWPRINCIPAL
fviz_ca_biplot(ca.out, repel=TRUE,
               title = "Biplot, Correspondence Analysis",
               arrows=c(TRUE,TRUE), map="rowprincipal",
               labelsize=4, xlim=c(-1.5,0.5))
```

```
## Coordinate system already present. Adding new coordinate system, which will replace the existing one
```




```
# SYMBIPLOT
fviz_ca_biplot(ca.out, repel=TRUE,
               title = "Biplot, Correspondence Analysis",
               arrows=c(TRUE,TRUE), map="symbiplot",
               labelsize=4, xlim=c(-1,1))
```

```
## Coordinate system already present. Adding new coordinate system, which will replace the existing one
```


Biplot, Correspondence Analysis
outdoorsy
rugged
Jeep Grand Cherokee
Ford explorer
sports
Porsche Boxter
exciting
fun
versatile
powerful
stylish
Saab 9000
BMW 328i
status
Chrysler Town & Country
performance
Mercedes C280
comfortable
luxurious
Infiniti J30
Volvo V90
family
dependable
safe
Lexus ES 300
0.5
0.0
−1.0
−0.5
0.0
0.5
1.0
Dim2 (33.6%)
Dim1 (61.7%)