

```
#####
# Project:      Credit Scoring Analysis (petit example)
# Description:  Part 4 - Multiple Correspondence Analysis
# Data:        CleanCreditScoring.csv
# By:          Gaston Sanchez
# url:         www.gastonsanchez.com
#
# Apply MCA on categorical variables, using Status
# as an illustrative projected variable
#####

# remember to change your working directory!!! (don't use mine)
# setwd("/Users/gaston/Documents/Gaston/StatsDataMining")

# load packages FactoMineR and ggplot2
require(FactoMineR)
require(ggplot2)

# read cleaned data set
dd = read.csv("CleanCreditScoring.csv", header=TRUE, stringsAsFactors=TRUE)

# select categorized continuous variables
ddcat = subset(dd, select=c(seniorityR, timeR, ageR, expensesR, incomeR,
    assetsR, debtR, amountR, priceR, finratR, savingsR, Status))

# MCA
mca = MCA(ddcat, quali.sup=12, graph=FALSE)

# barplot of eigenvalues
eigs = mca$eig$eigenvalue
barplot(eigs, border=NA, names.arg=1:length(eigs), las=2,
    cex.names=0.7, main="MCA eigenvalues", cex.main=0.9)

# What is the significant dimension in MCA?
nd = sum(eigs > 1/length(eigs))

# prepare data frame for plotting individuals with ggplot
mca.ind = data.frame(Status=dd$Status, mca$ind$coord)

# plot of individuals on first two dimensions with ggplot
# (this is for exploratory purposes)
ggplot(data=mca.ind, aes(x=Dim.1, y=Dim.2, group=Status)) +
  geom_hline(yintercept=0, colour="gray65") +
  geom_vline(xintercept=0, colour="gray65") +
  geom_point(alpha=0.3, aes(colour=Status)) +
  ggtitle("MCA plot of individuals")

# let's plot the categories of the categorical variables
# as well as the levels of Status (as illustrative variable)
mca.var = data.frame(rbind(mca$var$coord, mca$quali.sup$coord))
# ggplot
ggplot() +
  geom_hline(yintercept=0, colour="gray65") +
  geom_vline(xintercept=0, colour="gray65") +
  geom_text(data=mca.var[1:56,], aes(x=Dim.1, y=Dim.2, label=rownames(mca.var[1:56,])),
    alpha=0.4, size=3.5) +
  geom_text(data=mca.var[57:58,], size=5,
    aes(x=Dim.1, y=Dim.2, label=c("bad", "good"),
    colour=as.factor(c("bad", "good")))) +
  ggtitle("MCA plot of variables")

```