R instructions for the 5th seminar

In a data set *DistrictsCR.RData* there are some economic and demographic indicators of all districts in the Czech Republic. By means of the CanCor analyze association between ekonomic (variables 1 - 7) and demographic indicators (variables 8 - 15).

Earning99	average monthly earning in 1999 in crowns
Unempl	measure of registered unemployment in 2000
Unempl2	number of unemployed per 1 job
Unempl3	rate of long term unemployed per all unemployed
BuyingPower	
Progre	indicator of progressivity structure in economic
Enterpr	number of enterprisers per 1000 inhabitants in 1999
Lifex	average life expectancy in 1991-1995
Divorce	number of divorces per 100 marriages in 2000
Abortion	number of abortion per 100 born children
Pop	number of inhabitants living in small villages
Crime	number of crimes per 1000 inhabitants in 2000
Grow	relative growth of population
Migr	average relative migration growth 1991-2000
Pop65	rate of inhabitants older then 65 years in 1999

```
load("DistrictsCR.RData")
```

R Instructions for the problem 1:

•Get familiar with data (correlation matrix,scatterplot matrix,...)

a)

b) As the first case CapitalPraha is an outlier, remove this case from the data set. The new data set entitle "Distr".

```
Distr<-DistrictsCR[-1,]
!library(DescTools )
!library(ellipse)
!library(car)
```

```
x<-cor(Distr,use="pairwise.complete.obs")
pairs(Distr,panel=panel.smooth)
scatterplotMatrix(Distr,smooth=F,diagonal="histogram",col=c(2,1,4))
PlotCorr(x)
plotcorr(x)
```

R Instructions for the problem 2:

•Is CCA appropriate method for our data? (= Is an upper corner of correlation matrix with correlations between left an right set variables significantly different from the unit matrix?)

Visually asses graphs provided by correlation plots from previous task.

R Instructions for the problem 3:

•The matrix `Distr` separate into two matrices U and V where U represents first 7 variables associated with economic indicators and V represents 8 variables associated with demografphic indicators.

```
U<-Distr[,1:7]
V<-Distr[,8:15]
```

R Instructions for the problem 4:
•Create the object in R bearing all essential results of CCA. (Rather then function `cancor` from the package `stats` use function `cc` from a package `CCA`

```
!install.packages("CCA")
Cresults<-cc(U,V)
```

R Instructions for the problem 5:
•a) Find the total redundancy for the left set of variables (resp. right set of variables.) In other words: What proportion of the left set variance of original variables can be explained by "right" canonical variables?
•b) How the canonical variables represent their original variables? (How are Xs represented by Us? How are Ys represented by Vs?)

See R script `redundancy.R`

R Instructions for the problem 6:
•Determine the "reasonable" number of pairs of canonical variables.

a) Find the values of canonical correlation coefficients
`Crho<-Cresults$cor` notice that values in `Crho` are decreasing.

b) Test their significance by χ^2 Bartlett test (with successive roots removed).
This p - values are not available in a packag CCA, for that reason another package CCP has to be installed and loaded.

```
!install.packages("CCP")
p.asym(rho=Crho, N=76, p=7, q=8, tstat = "Wilks")
```

(The first three canonical correlations are significant at the level 5%.)

c) Asses the scree plot of eigenvalues (see chapter 5.3. in a lecture document).

```
Crho^2
plot(Crho^2 ,type="b")
```

d) Find redundancy for the first $k = 3$ pairs of canonical variables.
See R script `redundancy.R`

R Instructions for the problem 7:
•Interpret factor structure, e.g. correlations between particular variables "from the left set" X_i and canonical variable "from the left set" U_j. (Respectively correlations between particular variables "from the right set" Y_i and canonical variable "form the right set" V_j.) Which variables from the left set are well represented by the first canonical variable U_1?

```
Cresults$scores$corr.X.xscores
```

e.g. $R(Earning, U_1) = 0.842$ .The `Earning` is well represented by U_1.

```
Cresults$scores$corr.Y.yscores
```

R Instructions for the problem 8:
•Express cases in a new system of U and V canonical variables.(=Calculate the canonical scores.)

```
Cresults$scores$xscores
Cresults$scores$yscores
```