Indicator matrices & the central role they play in CA (correspondence analysis)

Suppose XI be an indicator matrix, first for a single categorical variable.

We might begin from the farms data (MASS), where there are 4 cat. variables.

`>head(farms)` #we shall look initially at `Mois` and `Manure` in farms

```
  Mois Manag Use Manure
1   M1     SF  U2      C4
2   M1     BF  U2      C2
3   M2     SF  U2      C4
4   M2     SF  U2      C4
5   M1     HF  U1      C2
6   M1     HF  U2      C2
attach(farms)
> Mois              Levels: M1 M2 M4 M5
 [1] M1 M1 M2 M2 M1 M1 M1 M5 M4 M2 M1 M4 M5 M5 M5 M5 M2 M1 M5 M5
> Manure           Levels: C0 C1 C2 C3 C4
 [1] C4 C2 C4 C4 C2 C2 C3 C3 C1 C1 C1 C2 C3 C0 C0 C3 C0 C0 C0 C0
```

Now, use the `indicc` function (below) --- first few rows only

`cbind(indicc(Mois),indicc(Manure))` ….TWO INDICATOR MATRICES, Aligned now

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	1	0	0	0	0	0	0	0	1
[2,]	1	0	0	0	0	0	1	0	0
[3,]	0	1	0	0	0	0	0	0	1
[4,]	0	1	0	0	0	0	0	0	1
[5,]	1	0	0	0	0	0	1	0	0
[6,]	1	0	0	0	0	0	1	0	0
[7,]	1	0	0	0	0	0	0	1	0
[8,]	0	0	0	1	0	0	0	1	0
[9,]	0	0	1	0	0	1	0	0	0

… so for the full matrix, we have the sum (of squares) and cross products matrix:

```
X2I=cbind(indicc(Mois),indicc(Manure))
> > t(X2I) %*% X2I
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	7	0	0	0	1	1	3	1	1
[2,]	0	4	0	0	1	1	0	0	2
[3,]	0	0	2	0	0	1	1	0	0
[4,]	0	0	0	7	4	0	0	3	0
[5,]	1	1	0	4	6	0	0	0	0
[6,]	1	1	1	0	0	3	0	0	0
[7,]	3	0	1	0	0	0	4	0	0
[8,]	1	0	0	3	0	0	0	4	0
[9,]	1	2	0	0	0	0	0	0	3

```
>crossd.svd((t(X2I) %*% X2I)[5:9,1:4])    #STUDY THIS a bit
[1] "canonical correlations are: 0.755 0.547 0.233 0.000
[1] "Square roots of singular values for Cont. table analysis are:"
[1] 1.84 1.57 1.02 0.00
[1] "The chi squared statistic for the Cont. table is: 18.57 with d.f.= 12"
[1] "Aim to interpret entries in coefsRt in terms of interdependence structure"
$obsvd
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	7	0	0	0	1	1	3	1	1
[2,]	0	4	0	0	1	1	0	0	2
[3,]	0	0	2	0	0	1	1	0	0
[4,]	0	0	0	7	4	0	0	3	0
[5,]	1	1	0	4	6	0	0	0	0
[6,]	1	1	1	0	0	3	0	0	0
[7,]	3	0	1	0	0	0	4	0	0
[8,]	1	0	0	3	0	0	0	4	0
[9,]	1	2	0	0	0	0	0	0	3

```
$exp
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
 [1,] 2.45  1.4  0.7 2.45  2.1 1.05  1.4  1.4 1.05
 [2,] 1.40  0.8  0.4 1.40  1.2 0.60  0.8  0.8 0.60
 [3,] 0.70  0.4  0.2 0.70  0.6 0.30  0.4  0.4 0.30
 [4,] 2.45  1.4  0.7 2.45  2.1 1.05  1.4  1.4 1.05
 [5,] 2.10  1.2  0.6 2.10  1.8 0.90  1.2  1.2 0.90
 [6,] 1.05  0.6  0.3 1.05  0.9 0.45  0.6  0.6 0.45
 [7,] 1.40  0.8  0.4 1.40  1.2 0.60  0.8  0.8 0.60
 [8,] 1.40  0.8  0.4 1.40  1.2 0.60  0.8  0.8 0.60
 [9,] 1.05  0.6  0.3 1.05  0.9 0.45  0.6  0.6 0.45

$cv    [,1] [,2] [,3]  [,4] [,5]  [,6] [,7] [,8]  [,9]
 [1,]  4.55 -1.4 -0.7 -2.45 -1.1 -0.05  1.6 -0.4 -0.05
 [2,] -1.40  3.2 -0.4 -1.40 -0.2  0.40 -0.8 -0.8  1.40
 [3,] -0.70 -0.4  1.8 -0.70 -0.6  0.70  0.6 -0.4 -0.30
 [4,] -2.45 -1.4 -0.7  4.55  1.9 -1.05 -1.4  1.6 -1.05
 [5,] -1.10 -0.2 -0.6  1.90  4.2 -0.90 -1.2 -1.2 -0.90
 [6,] -0.05  0.4  0.7 -1.05 -0.9  2.55 -0.6 -0.6 -0.45
 [7,]  1.60 -0.8  0.6 -1.40 -1.2 -0.60  3.2 -0.8 -0.60
 [8,] -0.40 -0.8 -0.4  1.60 -1.2 -0.60 -0.8  3.2 -0.60
 [9,] -0.05  1.4 -0.3 -1.05 -0.9 -0.45 -0.6 -0.6  2.55
$chim    [,1]   [,2]   [,3]   [,4]   [,5]    [,6]   [,7]   [,8]    [,9]
 [1,]  2.9069 -1.183 -0.837 -1.565 -0.759 -0.0488  1.352 -0.338 -0.0488
 [2,] -1.1832  3.578 -0.632 -1.183 -0.183  0.5164 -0.894 -0.894  1.8074
 [3,] -0.8367 -0.632  4.025 -0.837 -0.775  1.2780  0.949 -0.632 -0.5477
 [4,] -1.5652 -1.183 -0.837  2.907  1.311 -1.0247 -1.183  1.352 -1.0247
 [5,] -0.7591 -0.183 -0.775  1.311  3.130 -0.9487 -1.095 -1.095 -0.9487
 [6,] -0.0488  0.516  1.278 -1.025 -0.949  3.8013 -0.775 -0.775 -0.6708
 [7,]  1.3522 -0.894  0.949 -1.183 -1.095 -0.7746  3.578 -0.894 -0.7746
 [8,] -0.3381 -0.894 -0.632  1.352 -1.095 -0.7746 -0.894  3.578 -0.7746
 [9,] -0.0488  1.807 -0.548 -1.025 -0.949 -0.6708 -0.775 -0.775  3.8013
```

```
$chm.od [,1]    [,2]    [,3]   [,4]
[1,] -0.7591 -0.183 -0.775  1.31
[2,] -0.0488  0.516  1.278 -1.02
[3,]  1.3522 -0.894  0.949 -1.18
[4,] -0.3381 -0.894 -0.632  1.35
[5,] -0.0488  1.807 -0.548 -1.02
 $coefsRt        #Study patterns in rows, to see relationships, plot if desired.
         [,1]    [,2]
 [1,]  0.8139  0.196
 [2,] -0.2174 -1.402
 [3,]  1.0176  0.106
 [4,] -1.1935  0.808

 [5,] -0.8907  0.276
 [6,]  0.6948 -0.432
 [7,]  1.2114  0.427
 [8,] -0.6470  0.734
 [9,] -0.0868 -1.298
```

```
indicc  <- function(xx) {
#generates indicator matrix w/ 1 entry per row, acc. index in vector xx
if(!is.numeric(xx))xx=as.numeric(xx)
mm <- matrix(0, length(xx), length(unique(xx)))
indx <- ifelse(xx == col(mm), 1, 0)
indx  }
```