Intro to DatABEL

Yurii S. Aulchenko

June 22, 2010

Contents

```
> library(DatABEL)

DatABEL v 0.1-6 (June 21, 2010) loaded
YOU APPEAR TO WORK ON 32-BIT SYSTEM. LARGE FILES ARE NOT SUPPORTED.

> make_random_matrix <- function(range_dim1 = c(2, 10), range_dim2 = c(2,
+     10), range_data = c(-10, 10), type = "double") {
+     dim1 <- round(runif(1, range_dim1[1], range_dim1[2]))
+     dim2 <- round(runif(1, range_dim2[1], range_dim2[2]))
+     data <- runif(dim1 * dim2, range_data[1], range_data[2])
+     data <- as(data, type)
+     data <- matrix(data, nrow = dim1, ncol = dim2)
+     namesCol <- paste("col", c(1:dim2), sep = "_")
+     namesRow <- paste("row", c(1:dim1), sep = "_")
+     dimnames(data) <- list(namesRow, namesCol)
+     return(data)
+ }
> testmatr <- make_random_matrix()
> testmatr

          col_1     col_2
row_1 -5.665633 -1.454002
row_2  7.530301 -5.407012
row_3 -1.263586  2.019335
row_4 -9.207229 -5.476324
row_5  2.611365 -5.945676
row_6 -6.454872  9.370037
row_7 -6.168183  8.442943
row_8  2.748075 -4.762187
row_9 -8.647741 -8.883634

> test_fv <- as(testmatr, "databel")

[1] "./tmp343087"
coersion from 'matrix' to 'databel' of type DOUBLE ; object connected to file ./tmp343087
```

```
> test_fv

uninames$unique.names = TRUE
uninames$unique.rownames = TRUE
uninames$unique.colnames = TRUE
backingfilename = ./tmp343087
cachesizeMb = 64
number of columns (variables) =  2
number of rows (observations) =  9
usedRowIndex: 1  2  3  4  5  ...
usedColIndex: 1  2
Upper-left 2 columns and  5 rows:
          col_1     col_2
row_1 -5.665633 -1.454002
row_2  7.530301 -5.407012
row_3 -1.263586  2.019335
row_4 -9.207229 -5.476324
row_5  2.611365 -5.945676

> as(test_fv, "matrix")

          col_1     col_2
row_1 -5.665633 -1.454002
row_2  7.530301 -5.407012
row_3 -1.263586  2.019335
row_4 -9.207229 -5.476324
row_5  2.611365 -5.945676
row_6 -6.454872  9.370037
row_7 -6.168183  8.442943
row_8  2.748075 -4.762187
row_9 -8.647741 -8.883634

> abs(testmatr - as(test_fv, "matrix")) < 1e-06

      col_1 col_2
row_1  TRUE  TRUE
row_2  TRUE  TRUE
row_3  TRUE  TRUE
row_4  TRUE  TRUE
row_5  TRUE  TRUE
row_6  TRUE  TRUE
row_7  TRUE  TRUE
row_8  TRUE  TRUE
row_9  TRUE  TRUE

> write.table(testmatr, file = "test_matrix_dimnames.dat", row.names = TRUE,
+     col.names = TRUE, quote = FALSE)
> text2filevector(infile = "test_matrix_dimnames.dat", outfile = "test_matrix_dimnames",
+     R_matrix = TRUE)
```

```
Options in effect:
         --infile    = test_matrix_dimnames.dat
         --outfile   = test_matrix_dimnames
         --skiprows  = 1
         --skipcols  = 1
         --cnrow     = ON, using line 1 of 'test_matrix_dimnames.dat'
         --rncol     = ON, using column 1 of 'test_matrix_dimnames.dat'
         --transpose = OFF
         --Rmatrix   = ON
         --nanString = NA
Number of lines in source file is 10
Number of words in source file is 2
skiprows = 1
cnrow = 1
skipcols = 1
rncol = 1
Rmatrix = 1
numWords = 2
Creating file with numRows = 9
Creating file with numColumns = 2
Transposing test_matrix_dimnames_fvtmp => test_matrix_dimnames.
text2fvf finished.
uninames$unique.names = TRUE
uninames$unique.rownames = TRUE
uninames$unique.colnames = TRUE
backingfilename = test_matrix_dimnames
cachesizeMb = 64
number of columns (variables) =  2
number of rows (observations) =  9
usedRowIndex: 1  2  3  4  5  ...
usedColIndex: 1  2
Upper-left 2 columns and  5 rows:
          col_1     col_2
row_1 -5.665633 -1.454002
row_2  7.530301 -5.407012
row_3 -1.263586  2.019335
row_4 -9.207229 -5.476324
row_5  2.611365 -5.945676

> x <- databel("test_matrix_dimnames")
> x

uninames$unique.names = TRUE
uninames$unique.rownames = TRUE
uninames$unique.colnames = TRUE
backingfilename = test_matrix_dimnames
```

```
cachesizeMb = 64
number of columns (variables) =  2
number of rows (observations) =  9
usedRowIndex: 1  2  3  4  5  ...
usedColIndex: 1  2
Upper-left 2 columns and  5 rows:
          col_1     col_2
row_1 -5.665633 -1.454002
row_2  7.530301 -5.407012
row_3 -1.263586  2.019335
row_4 -9.207229 -5.476324
row_5  2.611365 -5.945676

> tmp <- as(x, "matrix")
> tmp

          col_1     col_2
row_1 -5.665633 -1.454002
row_2  7.530301 -5.407012
row_3 -1.263586  2.019335
row_4 -9.207229 -5.476324
row_5  2.611365 -5.945676
row_6 -6.454872  9.370037
row_7 -6.168183  8.442943
row_8  2.748075 -4.762187
row_9 -8.647741 -8.883634

> abs(testmatr - tmp) < 1e-06

      col_1 col_2
row_1  TRUE  TRUE
row_2  TRUE  TRUE
row_3  TRUE  TRUE
row_4  TRUE  TRUE
row_5  TRUE  TRUE
row_6  TRUE  TRUE
row_7  TRUE  TRUE
row_8  TRUE  TRUE
row_9  TRUE  TRUE

> text2filevector(infile = "test_matrix_dimnames.dat", outfile = "test_matrix_dimnames_T",
+     R_matrix = TRUE, transpose = TRUE)

Options in effect:
         --infile    = test_matrix_dimnames.dat
         --outfile   = test_matrix_dimnames_T
         --skiprows  = 1
```

```
         --skipcols  = 1
         --cnrow     = ON, using line 1 of 'test_matrix_dimnames.dat'
         --rncol     = ON, using column 1 of 'test_matrix_dimnames.dat'
         --transpose = ON
         --Rmatrix   = ON
         --nanString = NA
Number of lines in source file is 10
Number of words in source file is 2
skiprows = 1
cnrow = 1
skipcols = 1
rncol = 1
Rmatrix = 1
numWords = 2
Creating file with numRows = 9
Creating file with numColumns = 2
text2fvf finished.
uninames$unique.names = TRUE
uninames$unique.rownames = TRUE
uninames$unique.colnames = TRUE
backingfilename = test_matrix_dimnames_T
cachesizeMb = 64
number of columns (variables) =  9
number of rows (observations) =  2
usedRowIndex: 1  2
usedColIndex: 1  2  3  4  5  6  7  8  9
Upper-left 9 columns and  2 rows:
          row_1     row_2     row_3     row_4     row_5     row_6     row_7
col_1 -5.665633  7.530301 -1.263586 -9.207229  2.611365 -6.454872 -6.168183
col_2 -1.454002 -5.407012  2.019335 -5.476324 -5.945676  9.370037  8.442943
          row_8     row_9
col_1  2.748075 -8.647741
col_2 -4.762187 -8.883634

> x <- databel("test_matrix_dimnames_T")
> t(testmatr)

          row_1     row_2     row_3     row_4     row_5     row_6     row_7
col_1 -5.665633  7.530301 -1.263586 -9.207229  2.611365 -6.454872 -6.168183
col_2 -1.454002 -5.407012  2.019335 -5.476324 -5.945676  9.370037  8.442943
          row_8     row_9
col_1  2.748075 -8.647741
col_2 -4.762187 -8.883634

> x

uninames$unique.names = TRUE
uninames$unique.rownames = TRUE
```

```
uninames$unique.colnames = TRUE
backingfilename = test_matrix_dimnames_T
cachesizeMb = 64
number of columns (variables) =  9
number of rows (observations) =  2
usedRowIndex: 1  2
usedColIndex: 1  2  3  4  5  6  7  8  9
Upper-left 9 columns and  2 rows:
          row_1     row_2     row_3     row_4     row_5     row_6     row_7
col_1 -5.665633  7.530301 -1.263586 -9.207229  2.611365 -6.454872 -6.168183
col_2 -1.454002 -5.407012  2.019335 -5.476324 -5.945676  9.370037  8.442943
          row_8     row_9
col_1  2.748075 -8.647741
col_2 -4.762187 -8.883634

> tmp <- as(x, "matrix")
> tmp

          row_1     row_2     row_3     row_4     row_5     row_6     row_7
col_1 -5.665633  7.530301 -1.263586 -9.207229  2.611365 -6.454872 -6.168183
col_2 -1.454002 -5.407012  2.019335 -5.476324 -5.945676  9.370037  8.442943
          row_8     row_9
col_1  2.748075 -8.647741
col_2 -4.762187 -8.883634

> abs(t(testmatr) - tmp) < 1e-06

      row_1 row_2 row_3 row_4 row_5 row_6 row_7 row_8 row_9
col_1  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
col_2  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE

> unlink("*.fv?")
> unlink("test_matrix_*")
```