

quick view tools for eSets

VJ Carey

October 1, 2012

Contents

1	Introduction	1
2	An alternative to the current show method	1
3	Sample characterization	2

1 Introduction

In teaching a course where a large number of datasets are introduced over a short period of time, the relationship between data content and software infrastructure can be hard to master. This document introduces a number of experimental approaches to getting rapid access to key elements of eSet derivatives.

We will work with the ALL data for demonstration.

```
> library(Biobase)
> library(ALL)
> data(ALL)
> ALL
```

2 An alternative to the current show method

It could be nice to tell the package from which the dataset was loaded.

```
> dataSource = function(dsn) {
+   if (!is(dsn, "character")) dsn = try(deparse(substitute(dsn)))
+   if (inherits(dsn, "try-error")) stop("can't parse dsn arg")
+   dd = data()$results
+   if (is.na(match(dsn, dd[, "Item"]))) return(NULL)
+   paste("package:", dd[ dd[, "Item"] == dsn, "Package" ], sep="")
+ }
```

```
+ }
>
>
```

We use `peek` to get a concise view:

```
> peek(ALL)

ALL [from package:ALL]:
Platform annotation:  hgu95av2
primary assay results are:
Features  Samples
    12625    128
sample attributes are:
first 10 of 21 attributes:

                                labelDescription.truncated.
cod                                Patient ID
diagnosis                        Date of diagnosis
sex                              Gender of the patient
age                              Age of the patient at entry
BT                               does the patient have B-cell or T-cell ALL
remission  Complete remission(CR), refractory(REF) or NA. De
CR                                Original remisson data
date.cr                          Date complete remission if achieved
t(4;11)    did the patient have t(4;11) translocation. Deriv
t(9;22)    did the patient have t(9;22) translocation. Deriv
-----
use varTable to see values/freqs of all sample attributes
-----
```

3 Sample characterization

Getting a handle on sample characterization requires survey of variable names.

```
> varNames(ALL)

[1] "cod"           "diagnosis"      "sex"
[4] "age"           "BT"             "remission"
[7] "CR"           "date.cr"        "t(4;11)"
[10] "t(9;22)"       "cyto.normal"    "citog"
[13] "mol.biol"      "fusion protein" "mdr"
[16] "kinet"         "ccr"            "relapse"
[19] "transplant"    "f.u"            "date last seen"
```

In addition, we need to know values taken. This can be very cumbersome. We have a few parameters on how much detail is provided.

```
> varTable(ALL, max=4)
```

```
$cod
```

```
[1] "10005" "... " "LAL5"
```

```
$diagnosis
```

```
[1] "1/13/1997" "... " "9/4/1997"
```

```
$sex
```

```
[1] "F" "M"
```

```
$age
```

```
[1] "5" "... " "58"
```

In the above, we are only showing 4 attributes. By default all attributes would be shown. Note that the report on range of values is truncated and is character mode. We can show the full range of values using the `full` parameter.

```
> varTable(ALL, full=TRUE, max=4)
```

```
$cod
```

```
[1] "10005" "1003" "1005" "1007" "1010" "11002" "11005"
[8] "12006" "12007" "12008" "12012" "12019" "12026" "14016"
[15] "15001" "15004" "15005" "15006" "16002" "16004" "16007"
[22] "16009" "17003" "18001" "19002" "19005" "19008" "19014"
[29] "19017" "20002" "20005" "2020" "22009" "22010" "22011"
[36] "22013" "24001" "24005" "24006" "24008" "24010" "24011"
[43] "24017" "24018" "24019" "24022" "25003" "25006" "26001"
[50] "26003" "26005" "26008" "26009" "27003" "27004" "28001"
[57] "28003" "28005" "28006" "28007" "28008" "28009" "28019"
[64] "28021" "28023" "28024" "28028" "28031" "28032" "28035"
[71] "28036" "28037" "28042" "28043" "28044" "28047" "30001"
[78] "3002" "31007" "31011" "31015" "33005" "36001" "36002"
[85] "37001" "37013" "4006" "4007" "4008" "4010" "4016"
[92] "4018" "43001" "43004" "43006" "43007" "43012" "43015"
[99] "44001" "48001" "49004" "49006" "56007" "57001" "6002"
[106] "62001" "62002" "62003" "63001" "64001" "64002" "64005"
[113] "65003" "65005" "68001" "68003" "8001" "8011" "8012"
[120] "8018" "8024" "83001" "84004" "9002" "9008" "9017"
[127] "LAL4" "LAL5"
```

\$diagnosis

[1]	"1/13/1997"	"1/14/1997"	"1/15/1997"	"1/15/1998"
[5]	"1/15/1999"	"1/16/1997"	"1/17/1998"	"1/29/1997"
[9]	"1/3/1997"	"1/30/1997"	"10/1/1998"	"10/14/1997"
[13]	"10/19/1996"	"10/20/1998"	"10/21/1997"	"10/22/1998"
[17]	"10/23/1998"	"10/30/1997"	"10/4/1996"	"11/1/1998"
[21]	"11/11/1997"	"11/14/1996"	"11/15/1997"	"11/28/1996"
[25]	"12/17/1999"	"12/21/1998"	"12/21/1999"	"12/23/1996"
[29]	"12/23/1998"	"12/27/1996"	"12/3/1998"	"12/30/1998"
[33]	"12/31/1999"	"12/4/1998"	"2/10/1998"	"2/10/2000"
[37]	"2/18/1997"	"2/18/1998"	"2/18/1999"	"2/20/1997"
[41]	"2/21/1997"	"2/26/1998"	"2/29/2000"	"2/3/2000"
[45]	"2/4/1997"	"3/11/1997"	"3/15/2000"	"3/17/2000"
[49]	"3/18/1998"	"3/18/2000"	"3/19/1997"	"3/22/1997"
[53]	"3/23/2000"	"3/24/1999"	"3/24/2000"	"3/27/1997"
[57]	"3/27/1998"	"3/29/2000"	"4/1/1998"	"4/10/1997"
[61]	"4/11/2000"	"4/17/2000"	"4/19/1997"	"4/19/1998"
[65]	"4/23/1997"	"4/29/1998"	"4/7/2000"	"4/8/1997"
[69]	"5/14/1997"	"5/14/1998"	"5/15/1997"	"5/21/1997"
[73]	"5/22/1998"	"5/27/1999"	"5/28/1998"	"5/29/1998"
[77]	"5/4/1999"	"5/4/2000"	"5/9/1997"	"6/1/1998"
[81]	"6/10/1998"	"6/17/1997"	"6/18/1999"	"6/24/1998"
[85]	"6/28/1999"	"6/3/1997"	"7/11/2000"	"7/17/1997"
[89]	"7/20/1999"	"7/20/2000"	"7/22/1997"	"7/30/1997"
[93]	"7/8/1997"	"7/8/1998"	"8/10/1999"	"8/12/1998"
[97]	"8/17/2000"	"8/21/1998"	"8/25/1999"	"8/26/1999"
[101]	"8/27/1999"	"8/28/1997"	"8/5/1997"	"8/6/1999"
[105]	"9/15/1998"	"9/18/1997"	"9/23/1998"	"9/25/1998"
[109]	"9/26/1998"	"9/27/1997"	"9/29/1997"	"9/3/1997"
[113]	"9/30/1997"	"9/30/1998"	"9/4/1997"	

\$sex

[1] "F" "M"

\$age

[1]	"5"	"14"	"15"	"16"	"17"	"18"	"19"	"20"	"21"	"22"	"23"	"24"
[13]	"25"	"26"	"27"	"28"	"29"	"30"	"31"	"32"	"33"	"36"	"37"	"38"
[25]	"39"	"40"	"41"	"43"	"44"	"45"	"46"	"47"	"48"	"49"	"50"	"51"
[37]	"52"	"53"	"54"	"55"	"57"	"58"						