Dataset – Overview

Emmanuel Rousseaux

July 20, 2012

Prequisites

The Dataset package provides a class, called Dataset, and some methods allowing to handle more information about a data set in a SPSS-like way: handle variable labels, values labels, and retrieve information about missing values like in a SPSS data set file. In particularity we can switch a specific missing value type from a NA to a level to use it in the analysis. A function allowing to import data from a SPSS file to a Dataset object is provided. It is based on the R language (www.r-project.org).

- MeasureFeature objects should be used instead of numeric vector objects
- CategoricalFeature objects should be used instead of factor objects
- OrdinalFeature objects should be used instead of ordered factor objects
- Dataset objects should be used instead of data.frame objects

How to report a bug

Please provide enough information for us to help you. This typically includes the platform (windows, Unix, Macintosh) that you are using as well as version numbers for R and for the package that seems to be working incorrectly.

Include a small complete example that can be run and demonstrates the problem. In some cases it is also important that you describe what you thought you should get.

Please note:

- bugs in R should be reported to the R community
- missing features are not bugs – they are feature requests.

1 Dataset Design

Dataset relies on the R package system to distribute code and data. Most packages use S4 classes and methods (as described in *Programming with Data* by J. M. Chambers). This adherence to object oriented programming makes it easier to build component software and helps to deal with the complexity of the data. (TO CHANGE, from bioconductor)

1.1 Class Diagram

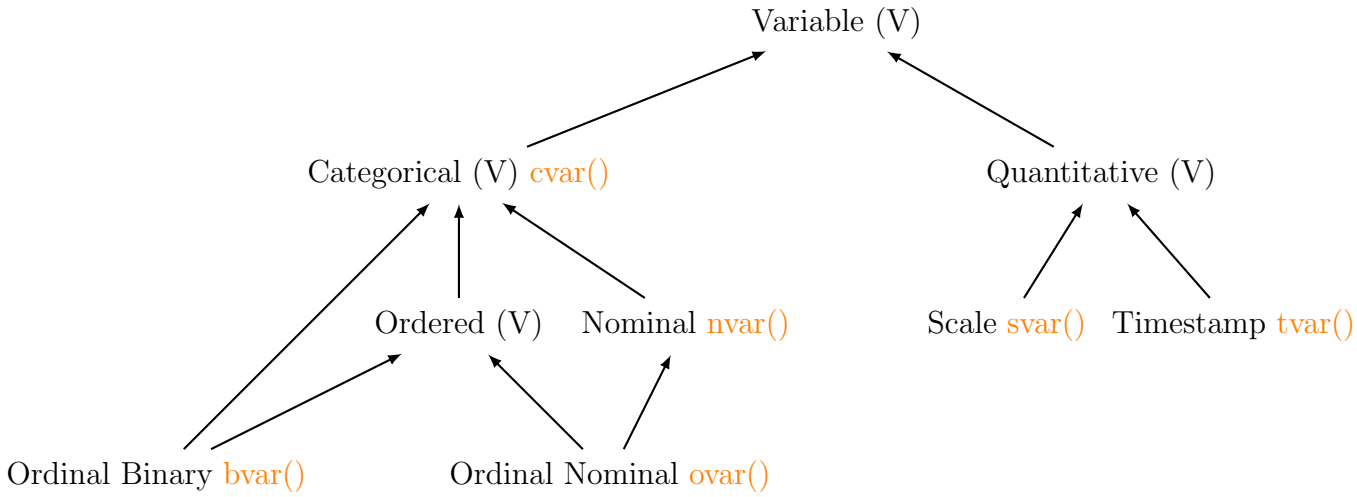

Figure 1: Class diagramme of objects inheriting of the Variable class.

A Variable can't be of type "character", i.e. a character vector with a different value for each individual. We don't expect this kind of variable in survey data. If such a vector is given to build a Variable object, the result will be a CategoricalVariable.

2 Session Information

The version number of R and packages loaded for generating the vignette were:

```
R version 2.15.1 (2012-06-22)
Platform: x86_64-pc-linux-gnu (64-bit)

locale:
[1] C

attached base packages:
```

```
[1] stats     graphics  grDevices utils     datasets  methods   base

loaded via a namespace (and not attached):
[1] tools_2.15.1
```