Introduction to the `surveydata` package for working with survey data.

Andrie de Vries

December 27, 2012

`surveydata` is a package that makes it easy to work with typical survey data that originated in SPSS or other formats.

A surveydata object consists of:

- A data frame with a row for each respondent and a column for each question. Column names are typically names in the pattern Q1, Q2_1, Q2_2, Q3 - where underscores separate the subquestions when these originated in a grid (array) of questions.

- Question metadata gets stored in the `variable.labels` attribute of the data frame. This typically contains the original questionnaire text for each question.

- Information about the subquestion separator (typically an underscore) is stored in the `patterns` attribute.

Data processing a survey file can be tricky, since the standard methods for dealing with data frames does not conserve the `variable.labels` attribute. The `surveydata` package defines a `surveydata` class and the following methods that knows how to deal with the `variable.labels` attribute:

- `as.surveydata`

- `[.surveydata`

- `[<-.surveydata`

- `$.surveydata`

- `$<-.surveydata`

- `merge.surveydata`

  In addition, `surveydata` defines the following convenient methods for extracting and working with the variable labels:

  - varlabels

  - varlabels<-

1 Example

Load the package.

```
> library(surveydata)
```

Create sample data and a surveydata object:

```
> sdat <- data.frame(
+     id   = 1:4,
+     Q1   = c("Yes", "No", "Yes", "Yes"),
+     Q4_1 = c(1, 2, 1, 2),
+     Q4_2 = c(3, 4, 4, 3),
+     Q4_3 = c(5, 5, 6, 6),
+     Q10 = factor(c("Male", "Female", "Female", "Male")),
+     crossbreak  = c("A", "A", "B", "B"),
+     weight      = c(0.9, 1.1, 0.8, 1.2)
+ )
> varlabels(sdat) <- c(
+     "RespID",
+     "Question 1",
+     "Question 4: red", "Question 4: green", "Question 4: blue",
+     "Question 10",
+     "crossbreak",
+     "weight"
+   )
```

Now create a surveydata object.

```
> sv <- as.surveydata(sdat, renameVarlabels=TRUE)
```

It is easy to extract specific questions:

```
> sv[, "Q1"]
    Q1
1 Yes
2  No
3 Yes
4 Yes

> sv[, "Q4"]
  Q4_1 Q4_2 Q4_3
1    1    3    5
2    2    4    5
3    1    4    6
4    2    3    6
```

The extraction makes use of the underlying metadata, contained in the `varlabels` and `pattern` attributes:

```
> varlabels(sv)
                id                 Q1                Q4_1                Q4_2
          "RespID"       "Question 1"   "Question 4: red" "Question 4: green"
              Q4_3                Q10          crossbreak              weight
 "Question 4: blue"      "Question 10"        "crossbreak"            "weight"
```

```
> pattern(sv)
```

```
$sep
[1] "_"
```

```
$exclude
[1] "other"
```

It is easy to query the surveydata object to find out which questions it contains, as well as which columns store the data for those questions.

```
> questions(sv)
```

```
[1] "id"         "Q1"         "Q4"         "Q10"         "crossbreak"
[6] "weight"
```

```
> which.q(sv, "Q1")
```

```
[1] 2
```

```
> which.q(sv, "Q4")
```

```
[1] 3 4 5
```

The function **qText** gives access to the questionnaire text.

```
> qText(sv, "Q1")
```

```
[1] "Question 1"
```

```
> qText(sv, "Q4")
```

```
[1] "Question 4: red"   "Question 4: green" "Question 4: blue"
```

```
> qTextCommon(sv, "Q4")
```

```
[1] "Question 4"
```

```
> qTextUnique(sv, "Q4")
```

```
[1] "red"   "green" "blue"
```

2 Final thoughts

The last word.