

SQL code for verifying results in RDF data cube

mja@statgroup.dk

2016-05-16

Derive results in RDF data cube and compare with results in data cube

Setup

Here all libraries are loaded; this should not be necessary.

```
options(width=200) # long lines
library(xlsx)
```

```
## Loading required package: rJava
```

```
## Loading required package: methods
```

```
## Loading required package: xlsxjars
```

```
library(foreign)
library(rrdf)
```

```
## Loading required package: rrdflibs
```

```
library(rrdfqb)
```

```
## Loading required package: RCurl
```

```
## Loading required package: bitops
```

```
##
```

```
## Attaching package: 'RCurl'
```

```
## The following object is masked from 'package:rJava':
```

```
##
```

```
##      clone
```

```
## Loading required package: rrdfancillary
```

```
library(rrdfqbcrnd0)
```

```
## Loading required package: rrdgcdisc
```

```
## Loading required package: devtools
```

```
library(rrdfqbcindex)
library(rrdfqbcindexcheck)
```

Load the data

The SAS dataset is loaded, and factors converted to character.

```
obsFile<- system.file("extdata/sample-xpt", "adsl.xpt", package="rrdfqbcindex")
adsl<-read.xport(obsFile)
ii <- sapply(adsl, is.factor)
adsl[ii] <- lapply(adsl[ii], as.character)
```

The conversion to character can be avoided by using `library(SASxport)`, see (`../rrdfqbcindex/inst/data-raw/create-dm-table-as-csv.Rmd`).

Load the RDF data cube

The RDF data cube is loaded.

```
dataCubeFile<- system.file("extdata/sample-rdf", "DC-DEMO-sample.ttl", package="rrdfqbcindex")
store <- new.rdf() # Initialize
cat("Reading turtle definition from ", dataCubeFile, "\n")
```

```
## Reading turtle definition from /home/ma/R/x86_64-redhat-linux-gnu-library/3.2/rrdfqbcindex/extdata/
```

```
temp<- load.rdf(dataCubeFile, format="TURTLE", appendTo= store)
summarize.rdf(store)
```

```
## [1] "Number of triples: 3088"
```

With a SPARQL query the mean values for the WEIGHTBL variable in RDF data cube the are extracted.

```
rq<- '
prefix crnd-measure: <http://www.example.org/dc/measure#>
prefix code: <http://www.example.org/dc/code/>
prefix crnd-attribute: <http://www.example.org/dc/attribute#>
prefix ds: <http://www.example.org/dc/demo/ds/>
prefix qb: <http://purl.org/linked-data/cube#>
prefix crnd-dimension: <http://www.example.org/dc/dimension#>
select * where {
?obs a qb:Observation ;
      qb:dataSet ds:dataset-DEMO ;
      crnd-dimension:agegr1 code:agegr1-_ALL_ ;
      crnd-dimension:ethnic code:ethnic-_ALL_ ;
      crnd-dimension:factor code:factor-weightbl ;
      crnd-dimension:procedure code:procedure-mean ;
      crnd-dimension:race code:race-_ALL_ ;
      crnd-dimension:sex code:sex-_ALL_ ;
      crnd-dimension:trt01a ?trt01a ;
```

```

    crnd-measure:measure      ?measure .
  }
  ,
knitr::kable(data.frame(sparql.rdf( store, rq)))

```

obs	trt01a	measure
ds:obs100	code:trt01a-Placebo	62.759302326
ds:obs102	code:trt01a-Xanomeline_High_Dose	70.004761905
ds:obs101	code:trt01a-Xanomeline_Low_Dose	67.279518072

The SPARQL query about can be made more generic.

Generate SQL statements

From the RDF data cube loading in the store, the function `GetSQLFromCube` generates the SQL statements for reproducing the data cube. Only the first two select statements are shown.

```

stmtSQL<- GetSQLFromCube( store, srcDsName="adsl" )
cat(paste(unlist(strsplit(stmtSQL$summStatSQL,split="\n"))[1:3],collapse="\n"),"\n")

```

```

## SELECT '_ALL_' as ETHNIC, '_ALL_' as RACE, '_ALL_' as AGEGR1, a.TRT01A, '_ALL_' as SEX, 'count' as p
## UNION
## SELECT '_ALL_' as ETHNIC, '_ALL_' as RACE, '_ALL_' as AGEGR1, a.TRT01A, '_ALL_' as SEX, 'percent' as

```

Derive the descriptive statistics

To show the full SQL expression use

```
cat(stmtSQL$summStatSQL)
```

```
library(sqldf)
```

```
## Loading required package: gsubfn
```

```
## Loading required package: proto
```

```
## Loading required package: RSQLite
```

```
## Loading required package: DBI
```

```
adsl.summ.stat.res<- sqldf( stmtSQL$summStatSQL)
```

```
## Loading required package: tcltk
```

```
names(adsl.summ.stat.res)<- tolower(gsub("(a|b)\\.","", names(adsl.summ.stat.res)))
knitr::kable(adsl.summ.stat.res)
```

ethnic	race	agegr1	trt01a	sex
HISPANIC OR LATINO	ALL	ALL	Placebo	ALL
HISPANIC OR LATINO	ALL	ALL	Placebo	ALL
HISPANIC OR LATINO	ALL	ALL	Xanomeline High Dose	ALL
HISPANIC OR LATINO	ALL	ALL	Xanomeline High Dose	ALL
HISPANIC OR LATINO	ALL	ALL	Xanomeline Low Dose	ALL
HISPANIC OR LATINO	ALL	ALL	Xanomeline Low Dose	ALL
NOT HISPANIC OR LATINO	ALL	ALL	Placebo	ALL
NOT HISPANIC OR LATINO	ALL	ALL	Placebo	ALL
NOT HISPANIC OR LATINO	ALL	ALL	Xanomeline High Dose	ALL
NOT HISPANIC OR LATINO	ALL	ALL	Xanomeline High Dose	ALL
NOT HISPANIC OR LATINO	ALL	ALL	Xanomeline Low Dose	ALL
NOT HISPANIC OR LATINO	ALL	ALL	Xanomeline Low Dose	ALL
ALL	AMERICAN INDIAN OR ALASKA NATIVE	ALL	Placebo	ALL
ALL	AMERICAN INDIAN OR ALASKA NATIVE	ALL	Xanomeline High Dose	ALL
ALL	AMERICAN INDIAN OR ALASKA NATIVE	ALL	Xanomeline High Dose	ALL
ALL	AMERICAN INDIAN OR ALASKA NATIVE	ALL	Xanomeline Low Dose	ALL
ALL	BLACK OR AFRICAN AMERICAN	ALL	Placebo	ALL
ALL	BLACK OR AFRICAN AMERICAN	ALL	Placebo	ALL
ALL	BLACK OR AFRICAN AMERICAN	ALL	Xanomeline High Dose	ALL
ALL	BLACK OR AFRICAN AMERICAN	ALL	Xanomeline High Dose	ALL
ALL	BLACK OR AFRICAN AMERICAN	ALL	Xanomeline Low Dose	ALL
ALL	BLACK OR AFRICAN AMERICAN	ALL	Xanomeline Low Dose	ALL
ALL	WHITE	ALL	Placebo	ALL
ALL	WHITE	ALL	Placebo	ALL
ALL	WHITE	ALL	Xanomeline High Dose	ALL
ALL	WHITE	ALL	Xanomeline High Dose	ALL
ALL	WHITE	ALL	Xanomeline Low Dose	ALL
ALL	WHITE	ALL	Xanomeline Low Dose	ALL
ALL	ALL	65-80	Placebo	ALL
ALL	ALL	65-80	Placebo	ALL
ALL	ALL	65-80	Xanomeline High Dose	ALL
ALL	ALL	65-80	Xanomeline High Dose	ALL
ALL	ALL	65-80	Xanomeline Low Dose	ALL
ALL	ALL	65-80	Xanomeline Low Dose	ALL
ALL	ALL	<65	Placebo	ALL
ALL	ALL	<65	Placebo	ALL
ALL	ALL	<65	Xanomeline High Dose	ALL
ALL	ALL	<65	Xanomeline High Dose	ALL
ALL	ALL	<65	Xanomeline Low Dose	ALL
ALL	ALL	<65	Xanomeline Low Dose	ALL
ALL	ALL	>80	Placebo	ALL
ALL	ALL	>80	Placebo	ALL
ALL	ALL	>80	Xanomeline High Dose	ALL
ALL	ALL	>80	Xanomeline High Dose	ALL
ALL	ALL	>80	Xanomeline Low Dose	ALL
ALL	ALL	>80	Xanomeline Low Dose	ALL
ALL	ALL	ALL	Placebo	F
ALL	ALL	ALL	Placebo	F
ALL	ALL	ALL	Placebo	M

