RBMRB

Kumaran Baskaran

January 12, 2016

RBMRB

RBMRB is an API written in R to import chemical shift data directly from BMRB into R environment as a data frame in R. This will facilitate easy access to BMRB data for statistical analysis and data visualization.

Installation

Requirements

RBMRB requires the package **httr** to import data from BMRB and **reshape2** to format the data into a data frame in R. If you already have these packages installed, then you may proceed with the installation of RBMRB package, otherwise please install these packages first and then proceed further.

Here is the instruction to install those packages. Open your R and use the following command

```
install.packages(c("httr","reshape2"))
```

you may also install packages by navigating through menu in RStudio or R.

Once you have installed the necessary packages, then you can download and install RBMRB. Source package can be downloaded from this link

After downloading use the following command to install RBMRB

```
install.packages("~/Downloads/RBMRB_0.1.tar.gz",repos=NULL,type="source")
```

```
## Installing package into '/home/kumaran/R/x86_64-pc-linux-gnu-library/3.2'
## (as 'lib' is unspecified)
```

Note: provide the right path to the downloaded file.

Usage

RBMRB can be used in a similar way like any other library in R.

```
library(RBMRB)
```

To fetch chemical shift table of a specific entry

```
df<-fetchBMRB('15060')
```

Imported data frame contains following columns

```
colnames(df)
```

```
##  [1] "BMRB_ID"                       "Entry_ID"
##  [3] "Entity_ID"                     "Comp_index_ID"
##  [5] "Comp_ID"                       "Atom_ID"
##  [7] "Atom_type"                     "Chemical_shift"
##  [9] "err"                           "Ambiguity_code"
## [11] "Assigned_chem_shift_list_ID"
```

Its also possible to load multiple entries as a single data frame

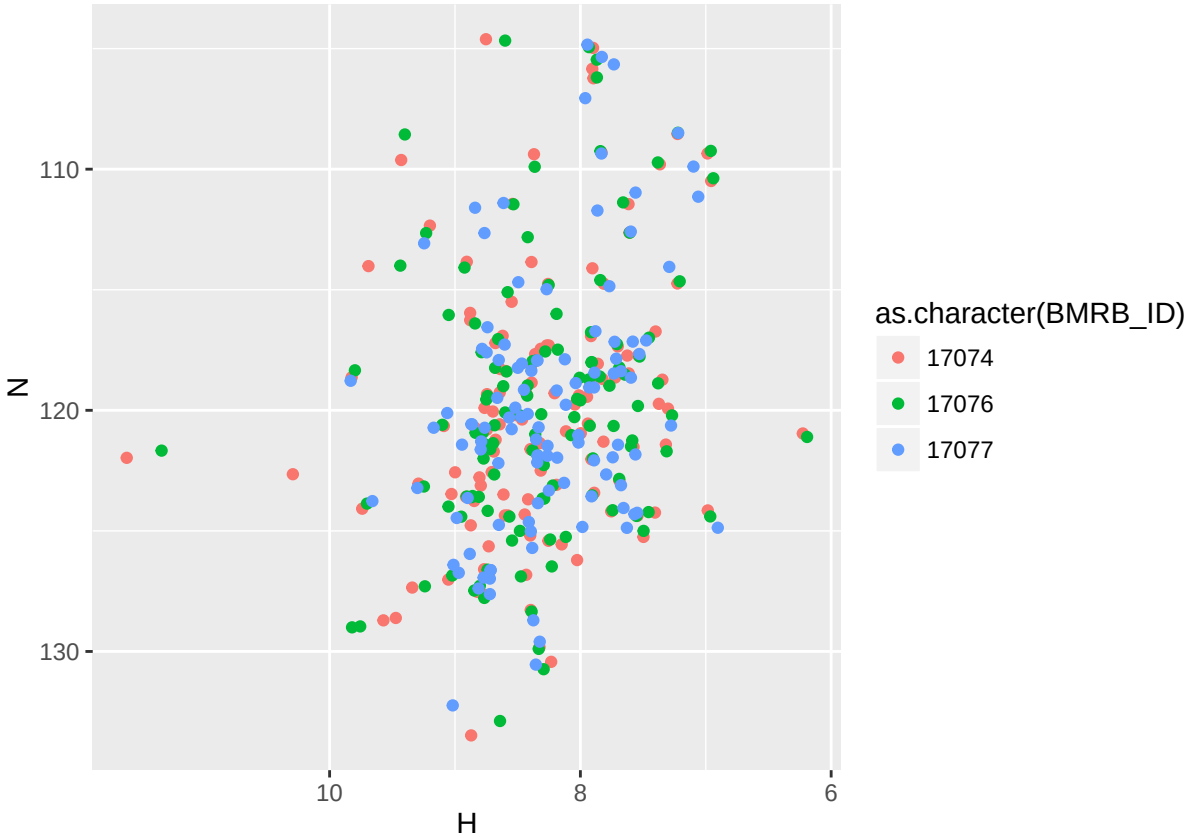
```
df<-fetchBMRB('17074,17076,17077')
```

These data frames can be reformatted using the function **N15HSQC** to simulate HSQC spectrum from the data frame.

```
hsqc<-N15HSQC(df)
```

Now user can use his own plotting function to visualize HSQC spectrum for example

```
library(ggplot2)
ggplot(hsqc)+
  geom_point(aes(H,N,color=as.character(BMRB_ID)))+
  scale_x_reverse()+scale_y_reverse()
```



The entire chemical shift table can be imported using the following command.

```
df_all<-fetchallBMRB()
```

Note: This will take some time, depends on the internet connection speed to download ~500MB csv file from the FTP server. Once it is downloaded, then you have full chemical shift information of all entries.

```
colnames(df_all)
```

```
##  [1] "ID"                          "Assembly_atom_ID"
##  [3] "Entity_assembly_ID"          "Entity_ID"
##  [5] "Comp_index_ID"               "Seq_ID"
##  [7] "Comp_ID"                     "Atom_ID"
##  [9] "Atom_type"                   "Atom_isotope_number"
## [11] "Val"                         "Val_err"
## [13] "Assign_fig_of_merit"         "Ambiguity_code"
## [15] "Occupancy"                   "Resonance_ID"
## [17] "Auth_entity_assembly_ID"     "Auth_asym_ID"
## [19] "Auth_seq_ID"                 "Auth_comp_ID"
## [21] "Auth_atom_ID"                "PDB_record_ID"
## [23] "PDB_model_num"               "PDB_strand_ID"
## [25] "PDB_ins_code"                "PDB_residue_no"
## [27] "PDB_residue_name"            "PDB_atom_name"
## [29] "Original_PDB_strand_ID"      "Original_PDB_residue_no"
## [31] "Original_PDB_residue_name"   "Original_PDB_atom_name"
## [33] "Details"                     "Sf_ID"
## [35] "Entry_ID"                    "Assigned_chem_shift_list_ID"
```

Hint: If you are going to use ***fetchallBMRB*** very often, then it is a good idea to download the all atom chemical shift csv file from this ftp server and specify the file name with correct local path in the function as a parameter

```
df_all=fetchallBMRB('~/Downloads/Atom_chem_shift.csv')
```

Data Visualization

This API is just an interface to fetch the data from BMRB database. This gives the freedom for the user to chose any tool from R to visualize the data. We created some examples in GitHub using this API.

HSQC Spectra

One of the most popular data visualization in BMRB is the simulation of N15-HSQC spectra. Now this simulation can be done on the user computer using this API

You need to have the following tools installed in your R to use this visualization

Required packages:

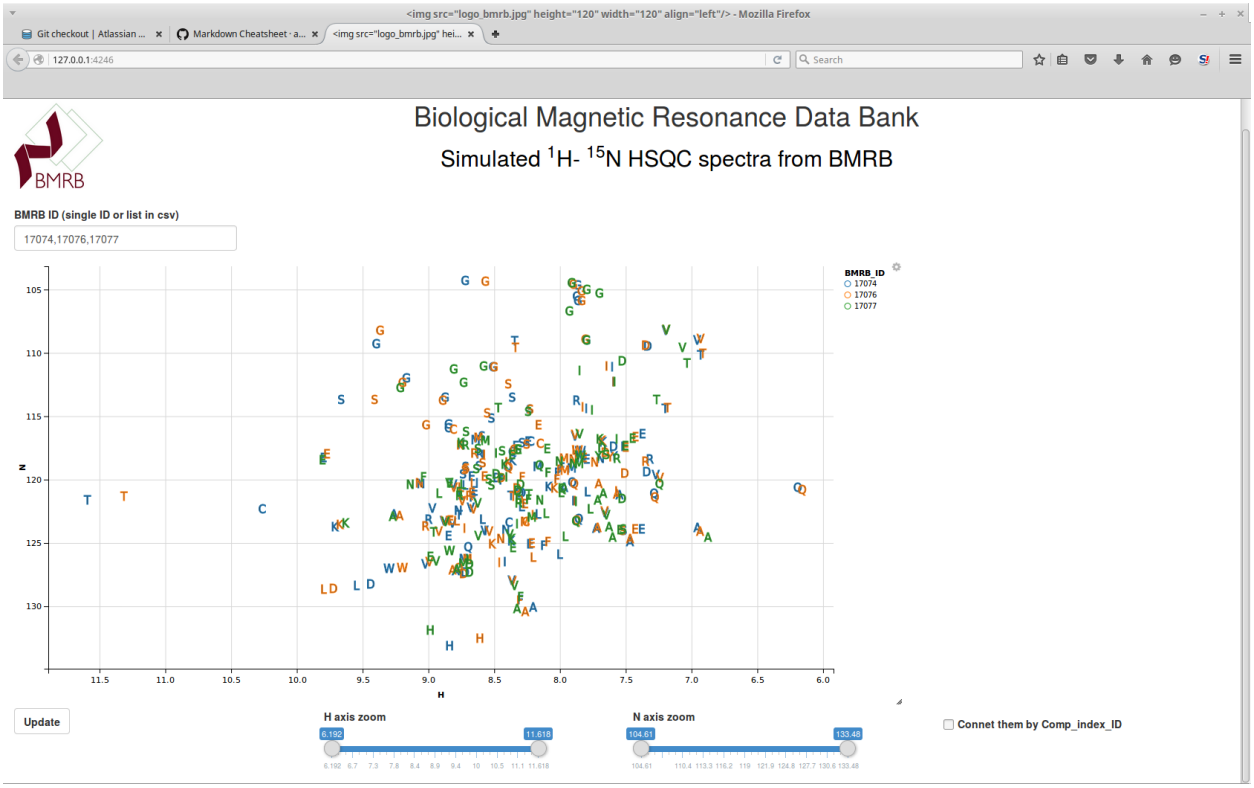
- shiny
- ggplot2
- ggvis
- reshape2

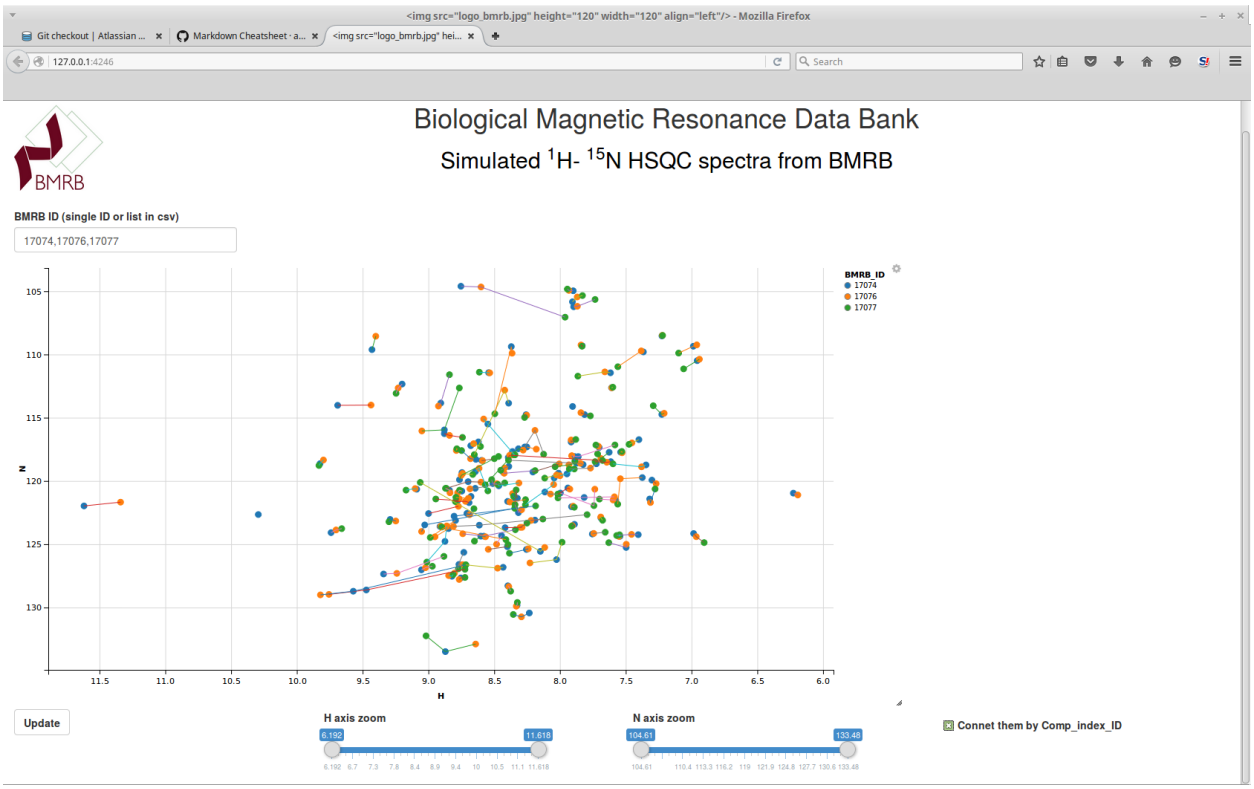
- httr
- plotly you may install these packages using this command

```
install.packages(c('shiny','ggvis','httr','reshape2','ggplot2'))
```

The visualization can be directly run from Git Hub

```
library(shiny)
runGitHub('ShinySpectra','uwbmrb')
```

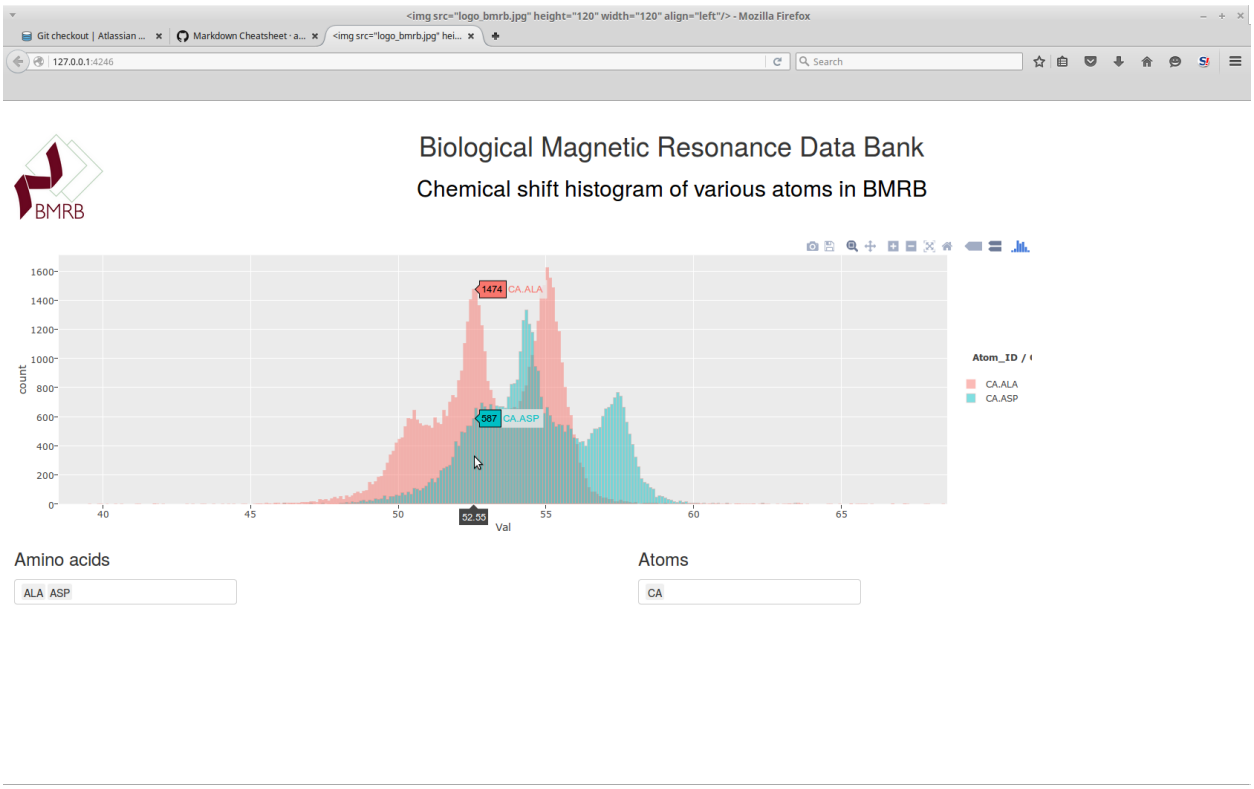




Chemical shift histograms

Next example is the chemical shift histogram of all atoms in the BMRB data base. This can also be run directly from Git Hub

```
library(shiny)
runGitHub('ShinyHistogram2','uwbmrb')
```


 - Mozilla Firefox
Git checkout | Atlassian ...
Markdown Cheatsheet · a...
<img src="logo_bmrb.jpg" hei...
127.0.0.1:4246
Search
BMRB
Biological Magnetic Resonance Data Bank
Chemical shift histogram of various atoms in BMRB
count
1600
1400
1200
1000
800
600
400
200
0
1474
CA.ALA
567
CA.ASP
40
45
50
55
60
65
52.55
Val
Atom_ID / (
CA.ALA
CA.ASP
Amino acids
ALA ASP
Atoms
CA