

dbhydroR: An R package to access the DBHYDRO Environmental Database

Joseph Stachelek

July 13, 2015

1 Introduction

This document introduces the `dbhydroR` package and its associated functions. These functions are aimed at improving programmatic workflows that query the DBHYDRO Environmental Database. HTTP requests are facilitated by the `httr` (Wickham 2015) and `RCurl` (Lang 2015) packages.

2 Package installation

The R package `dbhydroR` is distributed via a `.tar.gz` (analagous to `.zip`) package archive file. This package contains the source code for package functions. In RStudio, it can be installed by navigating to **Tools -> Install Packages... -> Install from: -> Package Archive File**. Computers running the Windows operating system can only install binary `.zip` package archive files unless they have additional compiler software installed. The `dbhydroR` binary package can be installed by running the following command from the R console:

```
install.packages(paste("\\\\ad.sfwmd.gov\\DFSRoot\\data\\restoration_sciences",  
  "\\projects\\joe_stachelek\\R\\dbhydroR_0.1-2.zip", sep=""),  
  ,type="win.binary",repos=NULL,dependencies=TRUE)
```

Once installed, the package can be loaded using the following command:

```
> library(dbhydroR)
```

3 Composing database queries

3.1 Water quality data

The workhorse `dbhydroR` functions are `getwq` and `gethydro`. This function takes four required arguments. The user must specify a station ID, a test name, and a date range. Station IDs can be located on the [SFWMD Google Earth portal](#). A list of available test names can be found in the [Appendix](#) to this document. Dates must be specified in YYYY-MM-DD format (e.g. 2015-02-26). The following set of examples retrieve measurements between March 2011 and May 2012. They can be run from the R console by issuing the command:

```
> example(getwq)
```

- One variable at one station

```
> getwq(station_id="FLAB08", date_min="2011-03-01",  
+       date_max="2012-05-01", test_name="CHLOROPHYLLA-SALINE")
```

- One variable at multiple stations

```
> getwq(station_id=c("FLAB08", "FLAB09"), date_min="2011-03-01",  
+       date_max="2012-05-01", test_name="CHLOROPHYLLA-SALINE")
```

- One variable at a wildcard station

```
> getwq(station_id=c("FLAB0%"), date_min="2011-03-01",  
+       date_max="2012-05-01", test_name="CHLOROPHYLLA-SALINE")
```

- Multiple variables at multiple stations

```
> getwq(station_id=c("FLAB08", "FLAB09"), date_min="2011-03-01",  
+       date_max="2012-05-01", test_name=c("CHLOROPHYLLA-SALINE",  
+                                           "SALINITY"))
```

By default, `getwq` returns a "cleaned output". First, the cleaning function `cleanwq` converts the raw output from native DBHYDRO "long" format (each piece of data on its own row) to "wide" format (each site x variable combination in its own column) using the `reshape2` package ([Wickham 2007](#)). Next, the QA blanks/flags and DBHydro database identifiers are stripped. Setting the `raw` flag to `TRUE` will force `getwq` to retain this information. An example query that retains this information is shown below.

```
> getwq(station_id="FLAB08", date_min="2011-03-01",  
+       date_max="2011-05-01", test_name="CHLOROPHYLLA-SALINE",  
+       raw=TRUE)
```

3.2 Hydrologic data

Hydrologic time series data can be retrieved using the `gethydro` function either by specifying a set of known `dbkeys` or by retrieving `dbkeys` on-the-fly with entries in the Station ID (`stationid`), data category (`category`), and parameter (`param`) fields.

Complex `stationid/category/parameter` combinations can easily cause errors or return unexpected results. If `dbkeys` are not known ahead of time (or manually retrieved from the [DBHYDRO Browser](#), it is good practice to plug in your `stationid/category/parameter` combination to the `getdbkey` function to ensure that the listing looks reasonable. The following set of examples can be run from the R console by issuing the command:

```
> example(getdbkey)
```

- Complete specification of the Station ID, Category, and Parameter flags

```
> getdbkey(stationid="JBTS",category="WEATHER",param="WNDS")
```

- Wildcard Station ID with no parameter specification

```
> getdbkey(stationid="C111%",category="SW")
```

```
> getdbkey(stationid="C111%",category="GW")
```

```
> getdbkey(stationid="C111%",category="WQ")
```

- Returning only `dbkeys` with the blind flag

```
> getdbkey(stationid="JBTS",category="WEATHER",param="WNDS", blind=TRUE)
```

In addition to a `dbkey`, users must specify a date range. Dates must be entered in YYYY-MM-DD format (e.g. 2015-02-26). The following set of examples retrieve wind speed measurements between January and February 2013. They can be run from the R console by issuing the command:

```
> example(gethydro)
```

- One variable/station time series

```
> gethydro(dbkey="15081",  
+          date_min="2013-01-01",date_max="2013-02-02")
```

- Multiple variable/station time series

```
> gethydro(dbkey=c("15081","15069"),  
+          date_min="2013-01-01",date_max="2013-02-02")
```

- Retrieiving unknown dbkeys on-the-fly

```
> gethydro(stationid="JBTS",category="WEATHER",
+          param="WNDS",date_min="2013-01-01",date_max="2013-02-02")
```

4 Plotting

Once data has been retrieved with `getwq` or `gethydro` in non-raw form (or raw water quality data that has been cleaned with `cleanwq`), and assigned to an R object, plots can be produced using the `dbhydro_plot` function. This function will create a dedicated panel of figures for each of the available variables. The following set of examples can be run from the R console by issuing the command:

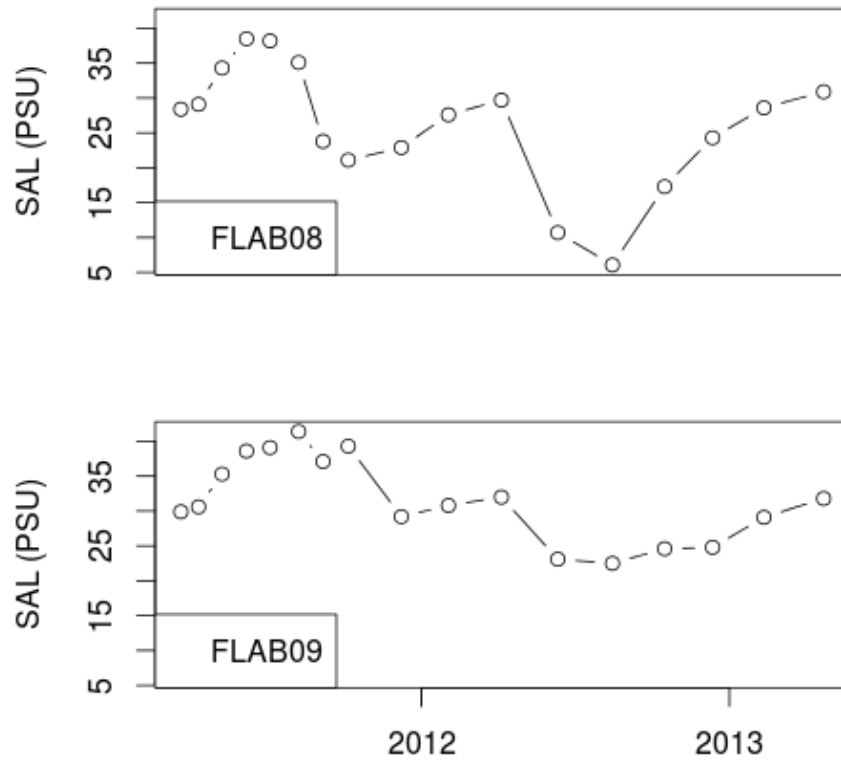
```
> example(dbhydro_plot)
```

- Water quality

```
> dt<-getwq(station_id=c("FLAB08","FLAB09"),
+           date_min="2011-03-01",date_max="2014-05-01",
+           test_name=c("CHLOROPHYLLA-SALINE","SALINITY"))
> dbhydro_plot(dt,label="bottomleft",abb=3)
```

- Hydrologic

```
> dt<-gethydro(dbkey=c("15081","15069"),
+             date_min="2013-01-01",date_max="2013-02-02")
> dbhydro_plot(dt)
```



5 Appendix

5.1 Test names

There are many test names available in DBHYDRO. These are detailed in the following table.

Code
AMMONIA-N
CARBON, TOTAL ORGANIC
CHLOROPHYLL-A(LC)
CHLOROPHYLL-B(LC)
CHLOROPHYLLA-SALINE
DISSOLVED OXYGEN
KJELDAHL NITROGEN,TOTAL
NITRATE+NITRITE-N
NITRITE-N
PHEOPHYTIN-A(LC)

PHOSPHATE,ORTHO AS P PHOSPHATE,TOTAL AS P SALINITY SILICA SP CONDUCTIVITY, FIELD TEMP TOTAL NITROGEN TURBIDITY

5.2 Further reading

See section on URL-based data access in the [DBHYDRO Browser User's Guide](#)

References

Duncan Temple Lang. *RCurl: General network (HTTP/FTP/...) client interface for R*, 2015. URL <http://CRAN.R-project.org/package=RCurl>. R package version 1.95-4.6.

Hadley Wickham. Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12):1–20, 2007. URL <http://www.jstatsoft.org/v21/i12/>.

Hadley Wickham. *httr: Tools for Working with URLs and HTTP*, 2015. URL <http://CRAN.R-project.org/package=httr>. R package version 0.6.1.