

Introduction to the dataRetrieval package

Laura De Cicco¹ and Robert Hirsch¹

¹*United States Geological Survey*

January 18, 2013

Contents

1	Introduction to dataRetrieval	1
1.1	What is dataRetrieval?	2
2	Getting Started	3
2.1	Installing dataRetrieval from downloaded binary:	3
2.2	Installing dataRetrieval from gitHub:	3
2.3	A Simple Web Retrieval Example	4
3	Function Details	7
3.1	Daily Value Retrievals	7
3.2	Water Quality Retrievals	7
3.3	Site Information Retrievals	7

1 Introduction to dataRetrieval

The dataRetrieval package was created to simplify the process of getting hydrologic data in the R environment. It has been specifically designed to work seamlessly with the EGRET package: Exploration and Graphics for RivEr Trends (EGRET).

1.1 What is dataRetrieval?

The dataRetrieval package was created to simplify hydrologic data retrieval. The options are web or user-produced files.

Here is some text. Here is some more.

2 Getting Started

This section describes the options for downloading and installing the dataRetrieval package.

2.1 Installing dataRetrieval from downloaded binary:

The dataRetrieval package is available for download at <https://github.com/USGS-CIDA/WRTDS/downloads>. If the package's tar.gz file is saved in R's working directory, then the following command will fully install the package:

```
> install.packages("dataRetrieval_1.2.0.tar.gz",  
+                  repos=NULL, type="source")
```

If the downloaded file is stored in an alternative location, include the path in the install command. A Windows example looks like this (notice the direction of the slashes, they are in the opposite direction that Windows normally creates paths):

```
> install.packages(  
+   "C:/RPackages/Statistics/dataRetrieval_1.2.0.tar.gz",  
+   repos=NULL, type="source")
```

A Mac example looks like this:

```
> install.packages(  
+   "/Users/userA/RPackages/Statistic/dataRetrieval_1.2.0.tar.gz",  
+   repos=NULL, type="source")
```

It is a good idea to re-start the R environment after installing the package, especially if installing an updated version. Some users have found it necessary to delete the previous version's package folder before installing newer version of dataRetrieval. If you are experiencing issues after updating a package, trying deleting the package folder - the default location for Windows is something like this: C:/Users/userA/Documents/R/win-library/2.15/dataRetrieval the default for a Mac: /Users/userA/Library/R/2.15/library/dataRetrieval Then, re-install the package using the directions above. Moving to CRAN should solve this problem.

2.2 Installing dataRetrieval from gitHub:

Alternatively, R-developers can install the latest version of dataRetrieval directly from gitHub using the devtools package. devtools is available on CRAN. Simply type the following commands into R to install the latest version of dataRetrieval available on gitHub. Rtools (for Windows) and latex tools are required.

```
> library(devtools)
> install_github("dataRetrieval", "USGS-CIDA")
```

To then open the library, simply type:

```
> library(dataRetrieval)
```

2.3 A Simple Web Retrieval Example

In this example, we use 3 dataRetrieval functions to get daily streamflow data and inorganic nitrogen sample results, and site information for a USGS gaging station with the ID 06934500. The station is Missouri River at Hermann, MO (which is discovered in the INFO dataset).

```
> Daily <- getDVDData("06934500", "00060", "1970-10-01", "2011-09-30")
```

There are 14975 data points, and 14975 days.

There are 0 zero flow days

If there are any zero discharge days, all days had 0 cubic meters per second added to the c

```
> head(Daily)
```

	Date	Q	Julian	Month	Day	DecYear	MonthSeq	Qualifier	i	LogQ
1	1970-10-01	3879.408	44102	10	274	1970.747	1450	A 1	8.263438	
2	1970-10-02	3454.655	44103	10	275	1970.750	1450	A 2	8.147478	
3	1970-10-03	3029.903	44104	10	276	1970.753	1450	A 3	8.016286	
4	1970-10-04	2644.793	44105	10	277	1970.755	1450	A 4	7.880348	
5	1970-10-05	2293.665	44106	10	278	1970.758	1450	A 5	7.737906	
6	1970-10-06	2072.793	44107	10	279	1970.761	1450	A 6	7.636652	
	Q7	Q30								
1	NA	NA								
2	NA	NA								
3	NA	NA								
4	NA	NA								
5	NA	NA								
6	NA	NA								

```
> Sample <- getSampleData("06934500", "00631", "1970-10-01", "2011-09-30")
> head(Sample)
```

	Date	ConcLow	ConcHigh	Uncen	ConcAve	Julian	Month	Day	DecYear	MonthSeq
1	1979-09-26	1.10	1.10	1	1.10	47384	9	269	1979.734	1557

2	1979-10-16	0.42	0.42	1	0.42	47404	10	289	1979.788	1558
3	1979-11-27	2.00	2.00	1	2.00	47446	11	331	1979.903	1559
4	1979-12-18	1.70	1.70	1	1.70	47467	12	352	1979.960	1560
5	1980-01-29	1.30	1.30	1	1.30	47509	1	29	1980.078	1561
6	1980-02-21	1.10	1.10	1	1.10	47532	2	52	1980.141	1562

	SinDY	CosDY
1	-0.9946999	-0.1028210
2	-0.9712570	0.2380333
3	-0.5724040	0.8199718
4	-0.2463613	0.9691781
5	0.4699767	0.8826788
6	0.7733507	0.6339785

```
> INFO <-getMetaData("06934500","00631", interactive=FALSE)
> colnames(INFO)
```

[1] "agency.cd"	"site.no"	"station.nm"
[4] "site.tp.cd"	"lat.va"	"long.va"
[7] "dec.lat.va"	"dec.long.va"	"coord.meth.cd"
[10] "coord.acy.cd"	"coord.datum.cd"	"dec.coord.datum.cd"
[13] "district.cd"	"state.cd"	"county.cd"
[16] "country.cd"	"map.nm"	"map.scale.fc"
[19] "alt.va"	"alt.meth.cd"	"alt.acy.va"
[22] "alt.datum.cd"	"huc.cd"	"basin.cd"
[25] "topo.cd"	"construction.dt"	"inventory.dt"
[28] "drain.area.va"	"contrib.drain.area.va"	"tz.cd"
[31] "local.time.fg"	"reliability.cd"	"project.no"
[34] "queryTime"	"drainSqKm"	"staAbbrev"
[37] "param.nm"	"param.units"	"paramShortName"
[40] "paramNumber"	"constitAbbrev"	

```
> INFO$station.nm
```

```
[1] "Missouri River at Hermann, MO"
```

```
> Sample <- mergeReport()
```

Discharge Record is 14975 days long, which is 41 years
 First day of the discharge record is 1970-10-01 and last day is 2011-09-30
 The water quality record has 437 samples
 The first sample is from 1979-09-26 and the last sample is from 2011-09-29
 Discharge: Minimum, mean and maximum 394 2660 20900
 Concentration: Minimum, mean and maximum 0.02 1.3 4.2
 Percentage of the sample values that are censored is 1.4 %

In the next section, we will go into detail the available functions in `dataRetrieval`, their required input and generated output.

3 Function Details

3.1 Daily Value Retrievals

3.2 Water Quality Retrievals

3.3 Site Information Retrievals

References

- [1] Helsel, D.R. and R. M. Hirsch, 2002. Statistical Methods in Water Resources Techniques of Water Resources Investigations, Book 4, chapter A3. U.S. Geological Survey. 522 pages. <http://pubs.usgs.gov/twri/twri4a3/>
- [2] Hirsch, R. M., Moyer, D. L. and Archfield, S. A. (2010), Weighted Regressions on Time, Discharge, and Season (WRTDS), with an Application to Chesapeake Bay River Inputs. JAWRA Journal of the American Water Resources Association, 46: 857-880. doi: 10.1111/j.1752-1688.2010.00482.x <http://onlinelibrary.wiley.com/doi/10.1111/j.1752-1688.2010.00482.x/full>
- [3] Sprague, L. A., Hirsch, R. M., and Aulenbach, B. T. (2011), Nitrate in the Mississippi River and Its Tributaries, 1980 to 2008: Are We Making Progress? Environmental Science & Technology, 45 (17): 7209-7216. doi: 10.1021/es201221s <http://pubs.acs.org/doi/abs/10.1021/es201221s>