USGS
science for a changing world

smwrBase—An R Package for Managing Hydrologic Data, Version 1.1.1

Open-File Report 2015–1202

U.S. Department of the Interior
U.S. Geological Survey

smwrBase—An R Package for Managing Hydrologic Data, Version 1.1.1

By David L. Lorenz

Open-File Report 2015–1202

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2015

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit http://www.usgs.gov or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit http://www.usgs.gov/pubprod/.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:
Lorenz, D.L., 2015, smwrBase—An R package for managing hydrologic data, version 1.1.1: U.S. Geological Survey Open-File Report 2015–1202, 7 p., http://dx.doi.org/10.3133/ofr20151202.

ISSN 2331-1258 (online)

Contents

Abstract....1
Introduction....1
Description of smwrBase....1
Summary....4
Disclaimer....4
Acknowledgments....5
References Cited....5
Appendixes....7
Appendix 1. R Documentation....7
Appendix 2. Regression Strategies Vignette....7
Appendix 3. Summary Statistics Vignette....7

Tables

1. Functions that import or export data....2
2. Functions that apply a transformation to one or more vectors. These functions always produce a single vector that is equal in length to the input....3
3. Functions that manipulate data other than producing a single vector that is equal in length to the input....4

smwrBase—An R Package for Managing Hydrologic Data, Version 1.1.1

By David L. Lorenz

Abstract

This report describes an R package called **smwrBase**, which consists of a collection of functions to import, transform, manipulate, and manage hydrologic data within the R statistical environment. Functions in the package allow users to import surface-water and groundwater data from the U.S. Geological Survey's National Water Information System database and other sources. Additional functions are provided to transform, manipulate, and manage hydrologic data in ways necessary for analyzing the data.

Introduction

The U.S. Geological Survey (USGS) seeks to make hydrologic data highly available and easy to acquire to the many user communities. The R package **smwrBase** was developed to meet those goals for the R user communities that need hydrologic data. This report presents a collection of functions for hydrologic data—streamflow, groundwater elevations, and discrete water-quality concentrations—that can be used to import, transform, manipulate, and manage those data. These functions are provided as an R package called **smwrBase**. The functions in the package have been developed by the USGS to import data from the USGS's National Water Information System (NWIS) database and other sources, and transform, manipulate, and manage hydrologic data in ways necessary for analyzing the data. This R package provides a convenient mechanism for distributing the tools to users of R within the USGS and other communities.

Description of smwrBase

The R package **smwrBase** consists of a collection of functions to import, transform, manipulate, and manage hydrologic data within the R statistical environment. Functions in the package allow users to import surface-water and groundwater data from the USGS's NWIS database and other sources. Additional functions are provided to transform, manipulate, and manage hydrologic data in ways necessary for analyzing the data. Documentation for the functions is provided in appendix 1.

The functions in **smwrBase** are provided as a package in R (http://www.r-project.org/), an open source language and environment for statistical computing and graphics that runs on a variety of operating systems including UNIX®, Linux®, Windows®, and Mac OS®. The R statistical environment can be extended for additional functionality using packages. Additional information on the installation and administration of R and packages that extend R is available in the manual "R Installation and Administration" (R Development Core Team, 2014).

Many of the functions in the **smwrBase** package, version 1.1.1, have been ported from the USGS library for S+ (Lorenz and others, 2011). Other functions have been developed and used within the USGS since the publication of Lorenz and others (2011). All of the functions are in the public domain (see the "Disclaimer" section).

The suggested citation for data from the **smwrBase** package can be acquired by using the *citation* function in R. The call is `citation(package="smwrBase")`.

The import and export functions, and other functions that support or facilitate the import and export of data, are listed in table 1. Data transformation functions are listed in table 2. In general, these data transformation functions accept a vector and other arguments used by the various transformations and return a vector of the transformed values. Random number generating functions for the Pearson type III and log-Pearson type III distributions are also listed in table 2. Functions that manipulate data, such as merging or restructuring datasets, decomposing a vector into components, providing type checking functions, and performing other miscellaneous functions, are listed in table 3.

In addition to these named functions, the **smwrBase** package has support functions for objects of class "timeDay," output from the function *as.timeDay*. These include

Table 1. Functions that import or export data.

[ASCII, American Standard Code for Information Interchange]

Function name	Description
exportCSV	Export a data frame to a comma-separated values file.
exportRDB	Export data to an ASCII relational-database file.
importCSV	Import a data frame from a comma-separated values file.
importRDB	Import a data frame from an ASCII relational-database file.
makeMeta	Create a template meta file for a comma-separated values file.
readList	Import data arranged on lines into a list.
setFileType	Support function to manage file suffixes.

method functions for *addition* (to objects of class "Date" and "POSIXt"), *as.character*, *as.data.frame*, *format*, *is.na*, *length*, *print*, and *show*.

The **smwrBase** package has detailed help files for each function that may be accessed in the same manner as help for other R functions. Help features within R are further described in the manual “An Introduction to R” (Venables and others, 2014).

Table 2. Functions that apply a transformation to one or more vectors. These functions always produce a single vector that is equal in length to the input.

Function or method name	Description
as.timeDay	Convert data to objects of class "timeDay."
baseDay	Compute the "base" day of the year, a reference value that can be used to group days for the computation of summary statistics.
boxCox	Apply a Box-Cox power transformation.
coalesce	Merge a matrix or list of vectors selecting the first non-missing value.
conc2meq	Convert concentration in milligrams per liter to milli-equivalents per liter.
%cn%	Identify character strings that contain the specified pattern.
daysInMonth	The number of days in a month.
dectime	Convert dates and times to decimal time in years.
dectime2Date	Convert decimal time in years to an object of class "Date."
dlpearsonIII	The density of the log-Pearson Type III distribution.
dms2dd	Convert data in degrees, minutes, and seconds to decimal degrees.
dpearsonIII	The density of the Pearson Type III distribution.
eventLen	Compute the length or duration of an event.
eventNum	Compute the number of an event, identified by a TRUE value in a sequence.
eventSeq	Compute the sequence number within an event.
fillMissing	Interpolate missing values in a regular time-series of data.
hyperbolic	Apply a hyperbolic transformation.
hysteresis	Compute a basis for estimating hysteresis effect in some variable related to the argument x.
IboxCox	Apply the inverse Box-Cox power transformation.
Ihyperbolic	Apply the inverse hyperbolic transformation.
index.coalesce	Return the index column number instead of the values for the first non-missing value.
IsCurve	Apply the inverse s-curve transformation.
miss2na	Convert a coded missing value to NA.
movingAve	Compute the moving average in regular time-series data.
movingDiff	Compute the moving difference in regular time-series data.
na2miss	Convert NA to a coded missing value.
pick	Select a value based on the value of a logical, integer, or character reference value.
plpearsonIII	Compute the cumulative probability of the log-Pearson Type III distribution.
ppearsonIII	Compute the cumulative probability of the Pearson Type III distribution.
qlpearsonIII	Compute the quantile of the log-Pearson Type III distribution.
qpearsonIII	Compute the quantile of the Pearson Type III distribution.
recode	Recode distinct values.
rlpearsonIII	Compute the random variates of the log-Pearson Type III distribution.
rpearsonIII	Compute the random variates of the Pearson Type III distribution.
scaleRng	Scale data to a specified range.
sCurve	Apply the s-curve transformation.
seasons	Create seasonal categories from dates.
setTZ	Set the time zone information for dates and times.
shiftData	Shift time-series data forward or backward.
timeDay	Various methods for manipulating time-of-day data, including conversion to and from character, addition, and others.
waterYear	Compute the water year of date data. The water year ends on September 30 of the year.

Table 3. Functions that manipulate data other than producing a single vector that is equal in length to the input..

Function name	Description
anomalies	Break down time-series data into long- and short-term deviations (anomalies) and the high-frequency variation.
conc.meq	Support function for conc2meq—produces a data frame read by conc2meq.
eventSeries	Create regular time-series data from recorded events.
fourier	Compute the Fourier series decomposition from date data.
group2row	Unstack data oriented in columns to rows of data.
isCharLike	Determine whether the data can be treated like character data.
isDateLike	Deterimne whether the data can be treated like date data.
isGroupLike	Determine whether the data can be treated like grouping data.
isNumberLike	Determine whether the data can be treated as numeric data.
mergeNearest	Merge two datasets by the nearest date and time.
mergeQ	Merge flow data with water-quality data.
more	Display the contents of an object by pages.
peaks	Compute the indices of peaks in time-series data.
quadratic	Compute a basis for an orthogonal second-order polynomial.
regularSeries	Put data collected at arbitrary times into a regular time series.
screenData	Screen data for missing values or gaps.
seqCollapse	Collapse a sequence of integers to a compact character string.
sumComposition	Compute the percentages of data within a matrix.
untable	Expand a 2-dimensional table into the raw values.
whichRowCol	Identify the row and column indexes for TRUE values in a logical matrix.

Summary

This report presents a collection of functions for hydrologic data—streamflow, groundwater elevations, and discrete water-quality concentrations—that can be used to import, transform, manipulate, and manage those data. These tools are provided as functions in an R package called **smwrBase**. The tools in the package have been developed by the U.S. Geological Survey to import data from the U.S. Geological Survey's National Water Information System (NWIS) database and other sources, and transform, manipulate, and manage hydrologic data in ways necessary for analyzing the data. This R package provides a convenient mechanism for distributing the tools to users of R within the U.S. Geological Survey and other communities.

Disclaimer

This package was written by U.S. Federal Government employees in the course of their employment and is therefore in the public domain, which means it is not copyrighted and use is unlimited; however, some of the functions depend on other R packages, which, although free and open source, have more restrictive licensing. Those packages are digest and lubridate (GNU [Gnu's Not Unix] GPL [General Public License], memoise [Massachusetts Institute of Technology, MIT], XML [Berkeley Software Distribution, BSD]. The R program itself is released under the free software license GNU GPL, either version 2, June 1991, or version 3, June 2007. Additional information on licensing is available at http://www.r-project.org/Licenses/ and http://www.gnu.org/licenses/license-list.html#SoftwareLicenses.

Although this software package has been used by the U.S. Geological Survey (USGS), no warranty, expressed or implied, is made by the USGS or the U.S. Government as to the accuracy and functioning of the program and related program material nor shall the fact of distribution constitute any such warranty, and no responsibility is assumed by the USGS in connection therewith. This software and related material (functions and documentation) are made available by the USGS to be used in the public interest and the advancement of science. Users may, without any fee or cost, use, copy, modify, or distribute this software, and any derivative works thereof, and its supporting documentation, subject to the USGS Software User's Rights Notice, http://water.usgs.gov/software/help/notice/.

Acknowledgments

The author thanks the time and effort of the U.S. Geological Survey personnel who reviewed this package and contributed suggestions to improve the package. The reviewers were Virginia McGuire, Nebraska Water Science Center; and Mark Brigham, Minnesota Water Science Center.

References Cited

Lorenz, D.L., Ahearn, E.A., Carter, J.M., Cohn, T.A., Danchuk, W.J., Frey, J.W., Helsel, D.R., Lee, K.E., Leeth, D.C., Martin, J.D., McGuire, V.L., Neitzert, K.M., Robertson, D.M., Slack, J.R., Starn, J., Vecchia, A.V., Wilkison, D.H., and Williamson, J.E., 2011, USGS library for S-PLUS for Windows—Release 4.0: U.S. Geological Survey Open-File Report 2011–1130, accessed December 5, 2014, at http://pubs.er.usgs.gov/publication/ofr20111130.

R Development Core Team, 2014, R installation and administration, version 3.1.2 (2014-10-31): accessed February 11, 2015, at http://streaming.stat.iastate.edu/CRAN/doc/manuals/R-admin.pdf.

Venables, W.N., Smith, D.M., and the R Development Core Team, 2014, An introduction to R, version 3.1.2 (2014-10-31): accessed February 11, 2015, at http://cran.r-project.org/doc/manuals/R-intro.pdf.

Appendixes

Appendix 1. R Documentation

The R documentation is a pdf file that can be accessed at http://pubs.usgs.gov/ofr/2015/1202/downloads/appendix1.pdf.

Appendix 2. Regression Strategies Vignette

Vignettes are the established R community method for providing examples of how to use the package. The vignette for regression strategies is a pdf file that can be accessed at http://pubs.usgs.gov/ofr/2015/1202/downloads/appendix2.pdf.

Appendix 3. Summary Statistics Vignette

Vignettes are the established R community method for providing examples of how to use the package. The vignette for summary statistics is a pdf file that can be accessed at http://pubs.usgs.gov/ofr/2015/1202/downloads/appendix3.pdf.

Publishing support provided by:
Rolla Publishing Service Center

For more information concerning this publication, contact:
Director, USGS Minnesota Water Science Center
2280 Woodale Drive
Mounds View, Minnesota 55112
(763) 783–3100

Or visit the Minnesota Water Science Center Web site at:
http://mn.water.usgs.gov/

ISSN 2331-1258 (online)
http://dx.doi.org/10.3133/ofr20151202