USGS
science for a changing world

smwrData—An R Package of Example Hydrologic Data, Version 1.1.1

Open-File Report 2015–1103

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

smwrData—An R Package of Example Hydrologic Data, Version 1.1.1

By David L. Lorenz

Open-File Report 2015–1103

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/.

Suggested citation:
Lorenz, D.L., 2015, smwrData—An R package of example hydrologic data, version 1.1.1: U.S. Geological Survey Open-File Report 2015–1103, 5 p., http://dx.doi.org/10.3133/ofr20151103.

ISSN 2331-1258 (online)

Contents

Tables

Supplemental Information

Concentrations of chemical constituents in water are given in milligrams per liter (mg/L)

Abbreviations

NASQAN	National Stream Quality Accounting Network
NWISWeb	National Water Information System Web interface
USGS	U.S. Geological Survey

smwrData—An R Package of Example Hydrologic Data, Version 1.1.1

By David L. Lorenz

Abstract

A collection of 24 datasets, including streamflow, well characteristics, groundwater elevations, and discrete water-quality concentrations, is provided to produce a consistent set of example data to demonstrate typical data manipulations or statistical analysis of hydrologic data. These example data are provided in an R package called **smwrData**. The data in the package have been collected by the U.S. Geological Survey or published in its reports, for example Helsel and Hirsch (2002). The R package provides a convenient mechanism for distributing the data to users of R within the U.S. Geological Survey and other users in the R community.

Introduction

Example data are needed to illustrate the functionality of statistical and graphical software. The data in this package are provided so that users can reproduce the results of analyses from Helsel and Hirsch (2002), Hem (1989), and Tesoriero and Voss (1997). The data are selected from a range of disciplines (groundwater, surface water, and water quality) within the field of hydrology to facilitate understanding of any application.

Although these data have been used by the U.S. Geological Survey (USGS), no warranty, expressed or implied, is made by the USGS or the United States Government as to their accuracy 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.

Description of smwrData

The data 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®. R can be extended for additional functionality by using packages that supply datasets or functions not included in the base distribution of R. 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, 2013).

The 24 datasets in the **smwrData** package, version 1.1.1, are derived from several sources within the USGS, including Helsel and Hirsch (2002) and the Web-based version of National Water Information System (NWISWeb; http://waterdata.usgs.gov/usa/nwis/nwis). The datasets include streamflow, well characteristics, groundwater levels, and discrete water-quality concentrations. The datasets are provided to produce a consistent set of example data to demonstrate typical data manipulations or statistical analysis of hydrologic data. All of the datasets in this package are in the public domain.

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

Table 1 is a listing of data extracted from Helsel and Hirsch (2002). Table 2 is a listing of data from other USGS reports; the sources are provided in the documentation for each dataset. Table 3 is a listing of data retrieved from NWISWeb.

The **smwrData** package has detailed help files for each dataset 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, 2013). The documentation for **smwrData** also is included in the appendix of this document.

The **smwrData** package also contains a single function, *dataRd*, that creates templates for datasets in the style used in this package. It can create documentation in either the default format for R, as an Rd file, or a roxygen header (Wickham and others, 2014).

Table 1. Datasets extracted from Helsel and Hirsch (2002).

[Appendixes, tables, and chapters listed are in Helsel and Hirsch (2002). USGS, U.S. Geological Survey]

Dataset name	Description
AppalachianSpecCap	Natural logarithms of specific capacity of wells in four rock types within the Appalachian mountain region of Pennsylvania (appendix C7).
ConecuhFlows	Annual streamflows from 1941 through 1960 for the Conecuh River at Brantley, Alabama, USGS station number 02371500 (appendix C2).
CuyahogaTDS	U.S. Geological Survey National Stream Quality Accounting Network (NASQAN) station: Cuyahoga River at Independence, Ohio (04208000) (appendix C9).
MayFlyNymph	Mayfly nymph counts in small streams above and below an industrial outfall (table 6.1).
MiningIron	Iron concentrations at low flow for small eastern Ohio streams (appendix C6).
PrecipNitrogen	Ammonia plus organic nitrogen concentrations in precipitation (example data in chapter 5).
SaddlePeaks	Annual peak discharges, 1925–1989, for the Saddle River at Lodi, New Jersey, USGS station number 01391500 (appendix C1).
TNLoads	Total nitrogen loads and basin characteristics for urban runoff (appendix C15).
UraniumTDS	Uranium and total dissolved solids (TDS) in groundwaters of differing bicarbonate (HCO3) concentrations (appendix C16).

Table 2. Datasets extracted from other U.S. Geological Survey reports.

Dataset name	Description
MiscGW	Groundwater-quality data from miscellaneous wells in the United States (Hem, 1989).
PugetNitrate	Selected groundwater nitrate and ancillary data collected near Puget Sound, Washington (Tesoriero and Voss, 1997).

Table 3. Datasets retrieved from the U.S. Geological Survey NWISWeb database.

[USGS, U.S. Geological Survey]

Dataset name	Description
ChoptankFlow	Selected daily flow data for Choptank River near Greensboro, Maryland, USGS station number 01491000.
ChoptankNH3	Selected ammonia concentration data for Choptank River near Greensboro, Maryland, USGS station number 01491000.
GlacialRidge	Daily groundwater data for water year 2008 for selected wells in the Glacial Ridge National Wildlife Refuge in northwestern Minnesota.
EasternIowaNO3	Selected nitrite plus nitrate concentration data from wells in the National Water-Quality Assessment (NAWQA) Program in the eastern Iowa study unit survey.
IonBalance	Selected ion balance data for County Ditch 65 near Maple Bay, Minnesota (SW2), USGS station number 05079250.
KlamathTP	Total phosphorus concentrations and streamflow data for the Klamath River near Klamath, California, USGS station number 11530500.
MenomineeMajorIons	Concentrations of selected major ions in the Menominee River near McAllister, Wisconsin, USGS station number 04067500.
Q05078470	Daily mean flow for Judicial Ditch 64 near Mentor, Minnesota (SW4), USGS station number 05078470 for calendar year 2003.
Q05078770	Daily mean flow for Judicial Ditch 66 near Marcoux Corners, Minnesota (SW6), USGS station number 05078770 for calendar year 2003.
Qall	Daily mean flow for selected USGS streamgages for calendar year 2003.
QW05078470	Selected water-quality data for Judicial Ditch 64 near Mentor, Minnesota (SW4), USGS station number 05078470 for calendar year 2003.
QWall	Whole-water or total phosphorus data for selected USGS streamgages for calendar year 2003.
QWstacked	Selected water-quality data for Judicial Ditch 64 near Mentor, Minnesota (SW4), USGS station number 05078470 for calendar year 2003.

Summary

A collection of 24 hydrologic datasets, including streamflow, well characteristics, groundwater elevations, and discrete water-quality concentrations is provided to produce a consistent set of example data to demonstrate typical data manipulations or statistical analysis. These example data are provided in an R package called **smwrData**. The data in the package have been collected by the USGS or published in its reports. The package provides a convenient mechanism for distributing the data to users of R within the USGS, and other users in the R community.

Although these data have been used by the USGS, no warranty, expressed or implied, is made by the USGS or the United States Government as to their accuracy 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.

Acknowledgments

The author thanks the time and effort of the USGS personnel who reviewed this report and contributed suggestions to improve the package. The reviewers were Sarah Elliott, Minnesota Water Science Center; and Naomi Nakagaki, California Water Science Center.

References Cited

Helsel, D.R. and R. M. Hirsch, 2002, Statistical methods in water resources: U.S. Geological Survey Techniques of Water Resources Investigations, book 4, chapter A3, 522 pages. (Also available at http://pubs.usgs.gov/twri/twri4a3/).

Hem J.D., 1989, Study and interpretation of the chemical characteristics of natural water: U.S. Geological Survey Water-Supply Paper 2254, 263 p.

R Development Core Team, 2013, R installation and administration, version 3.0.2, 2013-09-25, accessed November 14, 2013, at http://streaming.stat.iastate.edu/CRAN/doc/manuals/R-admin.pdf.

Tesoriero, A.J., and Voss, F.D., 1997, Predicting the probability of elevated nitrate concentrations in the Puget Sound Basin–Implications for aquifer susceptibility and vulnerability: Groundwater, v. 35, no. 6, p. 1029–1039.

Venables, W.N., Smith, D.M., and the R Development Core Team, 2013, An introduction to R, version 3.0.2, 2013-09-25, accessed November 14, 2013, at http://cran.r-project.org/doc/manuals/R-intro.pdf.

Wickham H., Danenberg, P, and Eugster, M., 2014, roxygen2: In-source documentation for R, R package version 4.0.2. at http://CRAN.R-project.org/package=roxygen2.

Appendix

The documentation pdf file can be accessed at http://pubs.usgs.gov/ofr/2015/1103/downloads/appendix.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/ofr20151103