

The USGSwsQWSR R package

Laura De Cicco¹, Steve Corsi¹, and Austin Baldwin¹

¹*United States Geological Survey*

January 7, 2014

Contents

1	Introduction to USGSwsQWSR package	1
2	General Workflow	2
3	Workflow Details	3
3.1	Data Retrieval	3
3.2	Data Merging	3
3.3	Data Investigation	3
3.4	Stepwise Regression	3
3.5	Model Analysis	3
A	Getting Started in R	4
A.1	New to R?	4
A.2	R User: Installing QWSR	4

1 Introduction to USGSwsQWSR package

The USGSwsQWSR package was designed to simplify the process of gathering water quality sample data and unit surrogate data, running a stepwise regression using the USGSwsQW censReg function, and analyzing those results. This vignette will first show a general overview workflow (2), then a more detailed description of the workflow with working examples (3).

2 General Workflow

```
library("USGSwsQWSR")

#Sample data included with package:
DTComplete <- StLouisDT
UV <- StLouisUV
QWcodes <- StLouisQWcodes
siteINFO <- StLouisInfo

investigateResponse <- "SuspSed"
transformResponse <- "lognormal"

DT <- DTComplete[c(investigateResponse,
                    getPredictVariables(names(UV)),
                    "decYear", "sinDY", "cosDY", "datetime")]
DT <- na.omit(DT)

predictVariables <- names(DT)[-which(names(DT)
                                     %in% c(investigateResponse, "datetime", "decYear"))]

#Check predictor variables
predictVariableScatterPlots(DT, investigateResponse)

# Create 'kitchen sink' formula:
kitchenSink <- createFullFormula(DT, investigateResponse)

#Run stepwise regression with "kitchen sink" as upper bound:
returnPrelim <- prelimModelDev(DT, investigateResponse, kitchenSink,
                               "BIC", #Other option is "AIC"
                               transformResponse)

steps <- returnPrelim$steps
modelResult <- returnPrelim$modelStuff
modelReturn <- returnPrelim$DT.mod

# Analyze steps found:
plotSteps(steps, DT, transformResponse)
analyzeSteps(steps, investigateResponse, siteINFO)

# Analyze model produced from stepwise regression:
resultPlots(DT, modelReturn, siteINFO)
resultResidPlots(DT, modelReturn, siteINFO)
```

```
# Create prediction plots  
predictionPlot(UV,DT,modelReturn,siteINFO=siteINFO)
```

3 Workflow Details

3.1 Data Retrieval

3.2 Data Merging

3.3 Data Investigation

3.4 Stepwise Regression

3.5 Model Analysis

A Getting Started in R

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

A.1 New to R?

If you are new to R, you will need to first install the latest version of R, which can be found here: <http://www.r-project.org/>.

There are many options for running and editing R code, one nice environment to learn R is RStudio. RStudio can be downloaded here: <http://rstudio.org/>. Once R and RStudio are installed, the dataRetrieval package needs to be installed as described in the next section.

At any time, you can get information about any function in R by typing a question mark before the functions name. This will open a file (in RStudio, in the Help window) that describes the function, the required arguments, and provides working examples.

```
library(USGSwsQWSR)
?plotSteps
```

To see the raw code for a particular code, type the name of the function:

```
plotSteps
```

A.2 R User: Installing QWSR

Before installing USGSwsQWSR, the dependent packages must be first be installed:

```
install.packages(c("XML", "lubridate", "akima",
                  "leaps", "car", "mvtnorm",
                  "relimp", "BSDA", "RODBC"),
                dependencies=TRUE)
install.packages(c("USGSwsBase", "USGSwsData",
                  "USGSwsGraphs", "USGSwsStats",
                  "USGSwsQW", "USGSwsQWSR"), repos="http://usgs-r.github.com")
```

After installing the package, you need to open the library each time you re-start R. This is done with the simple command:

```
library(USGSwsQWSR)
```