# The QWSR R package

Laura De Cicco[1], Steve Corsi[1], and Austin Baldwin[1]

[1] *United States Geological Survey*

August 6, 2013

## Contents

**1 Introduction to QWSR** **1**

**2 General Workflow** **1**

2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2

**A Getting Started in R** **3**

A.1 New to R? . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

A.2 R User: Installing QWSR . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

## 1 Introduction to QWSR

## 2 General Workflow

```
library(USGSwsQWSR)

Loading required package:  USGSwsQW

Loading required package:  USGSwsBase

Loading required package:  lubridate

Loading required package:  XML
Loading required package:  USGSwsGraphs

Loading required package:  KernSmooth

KernSmooth 2.23 loaded
Copyright M. P. Wand 1997-2009

Loading required package:  akima

Loading required package:  USGSwsStats

Loading required package:  leaps

Loading required package:  boot

Loading required package:  survival

Loading required package:  splines


Attaching package:  ’survival’

The following object is masked from ’package:boot’:

       aml

Loading required package:  mvtnorm

Loading required package:  latticeExtra

Loading required package:  RColorBrewer

Loading required package:  lattice


Attaching package:  ’lattice’

The following object is masked from ’package:boot’:

       melanoma

Loading required package:  MASS


Attaching package:  ’MASS’

The following object is masked from ’package:USGSwsBase’:

       select
```

## 2.1 Introduction

# 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.

```
?plotSteps
```

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

```
plotSteps


function (steps, localDT, transformResponse = "lognormal")
{
    parOriginal <- par(no.readonly = TRUE)
    responseVariable <- steps$response[1]
    nSteps <- nrow(steps)
    logPlot <- ""
    distribution <- transformResponse
    yMin <- min(localDT[, responseVariable]@.Data[, 2], na.rm = TRUE)
    xMin <- yMin
    lmFormula <- "obs ~ pred"
    if ("lognormal" == transformResponse) {
        logPlot <- "xy"
        lmFormula <- "log10(obs) ~ log10(pred)"
    }
    for (i in 2:nSteps) {
        formulaToUse <- substring(steps$scope[i], 3, nchar(steps$scope[i]))
        formulaToUse <- paste(responseVariable, formulaToUse,
            sep = " ~ ")
        cat(formulaToUse, "\n\n")
        modelReturn <- do.call("censReg", list(formulaToUse,
            data = localDT, dist = distribution))
        outlier <- findOutliers(modelReturn, localDT, transformResponse)
        responseValue <- localDT[, responseVariable]@.Data[,
            2]
        df <- data.frame(obs = responseValue, pred = modelReturn$YPRED)
        dfOutliers <- data.frame(obsOut = responseValue[outlier],
            predOut = modelReturn$YPRED[outlier])
        lineFit <- do.call("lm", list(lmFormula, data = df))
        par(tcl = 0.3)
        plot(df$pred, df$obs, ylab = "Observed", xlab = "Predicted",
            main = paste(responseVariable, i, sep = ":"), ylim = c(yMin,
                max(c(df$obs, df$pred))), xlim = c(xMin, max(c(df$obs,
                df$pred))), log = logPlot)
        points(dfOutliers$predOut, dfOutliers$obsOut, col = "red",
            pch = 16)
        if (sum(modelReturn$CENSFLAG) > 0) {
            cenValsX <- modelReturn$YPRED[modelReturn$CENSFLAG]
            cenValsY <- responseValue[modelReturn$CENSFLAG]
            segments(x0 = cenValsX, y0 = cenValsY, x1 = cenValsX,
                y1 = yMin * 0.001)
        }
        abline(lineFit, col = "red")
        abline(0, 1, col = "blue")
        mtext(formulaToUse, side = 3, line = -1, cex = 0.7)
        corStep <- cor(df$obs, df$pred)
        goodness <- paste("slope: ", formatC(lineFit$coefficients[2],
            digits = 4), ", cor: ", formatC(corStep, digits = 4),
            ", rsme: ", formatC(rmse(modelReturn), digits = 4),
            sep = "")
        mtext(goodness, side = 3, line = 0.5, cex = 0.7)
    }
    par(parOriginal)
}
<environment: namespace:USGSwsQWSR>
```

## A.2 R User: Installing QWSR

Before installing QWSR, 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"), repos="http://usgs-r.github.com")
```

It is a good idea to re-start R 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, and the default for a Mac: /Users/userA/Library/R/2.15/li
Then, re-install the package using the directions above. Moving to CRAN should solve this problem.

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)
```