Vignette, QW

Erik.Leppo@tetratech.com and *Jon.Harcum@tetratech.com*

2018-03-26

Purpose

The purpose of this vignette is to cover the basics of exporting and importing water quality data for use in baytrends. In order to have the functionality to deal with censored data, when it exists, the qw object is used in baytrends to store all water quality data. The process of combining two (or more) data frames with qw objects is also described.

The qw object was originally developed by the U.S. Geological Survey (USGS) for use in their `smwrQW` package. For future portability to the CRAN, numerous 'smwrQW' functions along with several supporting functions from USGS' family of smwr packages were incorporated into this package.

Background

Monitoring programs may report laboratory results where the concentration is below the detection limit of the analysis. Bacteriological tests may report very high results as "too numerous to count" (TNTC). Such data typically reported as "<" or ">" some value, x, are referred to as censored data. Censored values are usually associated with limitations of measurement or sample analysis, and are commonly reported as results below or above measurement capacity of the available analytical equipment. Results that are indistinguishable from a blank sample are normally reported as less than the detection limit (DL). The true values of these left-censored observations are considered to lie between zero and the DL. Depending on the laboratory, some results greater than the DL may be identified as less than the quantitation limit (QL) or reported as a single value and given a data qualifier to indicate the value is less than the QL. Typically, results reported as less than the QL indicate that the analyte was detected (i.e., greater than the detection limit), but at a low enough concentration where the precision was deemed too low to reliably report a single value. These interval-censored observations are considered to lie between the DL and QL. Interval-censored observations can also occur when calculating a chemical concentration based on adding or differencing multiple measured concentrations.

Data censoring leads to a loss of information and makes statistical analyses more challenging. Thus beginning in 1999, the Chesapeake Bay Program began to store the uncensored results of chemical measurements. Nevertheless, the Chesapeake Bay Program is interested in performing trend analyses with data back to 1985. Therefore, baytrends was developed explicitly developed to allow for multiple types of censored data.

Export

Use the function *qw.export* to export "qw" formatted data. The sample data set, `dataCensored`, included with `baytrends`, is used for demonstration.

```
library(baytrends)

# load the included data frame, dataCensored, into the global environment
# dataCensored is a sample data set that includes qw formatted data and
# is included with baytrends.
```

```
myDF <- dataCensored

# identify the current working directory as the location to save
# the outputted data set
dir.save <- getwd()

# identify the name of the comma delimited (csv) file for the
# outputted data set
fn.output <- "dataCensored_TEST.csv"

# run function
qw.export(myDF, dir.save, fn.output)
```

The above code chunk will output the raw data to a comma delimited (csv) file named `dataCensored_TEST.csv`. For non-qw variables (such as station, date, and layer in this example), the column names in the csv file are the same as the data frame. For qw-formatted variables, the columns of the data file are in the format of X_Y.

- ○ X = column name
- ○ Y = qw argument of "lo", "hi", and "symbol"

In the context of water quality data, left-censored observations such as "<0.1" would have a lo value of 0 and a hi value of 0.1. Interval-censored observations would have a non-zero lo value. The value of symbol would be "<" and "i", respectively for left- and interval-censored observations. Uncensored observations would have lo and hi values equal to each other.

Selected columns from the first few records of the above example are shown below.

```
df <- read.csv("dataCensored_TEST.csv")
head(df[,c(1:3,16:24)])
```

station	date	layer	tp_lo	tp_hi	tp_symbol	po4f_lo	po4f_hi	po4f_symbol	pp_lo	pp_hi	pp_symbol
CB3.3C	1985-01-23	B	0.028	0.028		0.0000	0.0070	<	0.016	0.028	i
CB3.3C	1985-01-23	S	0.030	0.030		0.0000	0.0070	<	0.018	0.030	i
CB3.3C	1985-02-13	B	0.043	0.043		0.0000	0.0070	<	0.033	0.043	i
CB3.3C	1985-02-13	S	0.021	0.021		0.0000	0.0070	<	0.011	0.021	i
CB3.3C	1985-03-05	B	NA	NA		0.0108	0.0108		NA	NA	
CB3.3C	1985-03-05	S	NA	NA		0.0040	0.0040		NA	NA	

Import

Use the function *qw.import* to import "qw" formatted data. The above exported file, `dataCensored_TEST.csv`, is used for demonstration.

To create a data frame that includes qw objects, the data needs to be in the format that was generated by qw.export. That is, one file with each qw parameter having three columns (x_lo, x_hi, and x_symbol where x is the name of the parameter). The list of parameter codes that will be as column names of the qw-formatted variables are specifed in the argument `qw.names`.

The output of the `qw.import` function must be directed to a variable. Any modification of column classes (e.g., POSIXct, numeric, or integer) will need to be performed by the user. The function str() is useful for examing the structure of the data frame.

```r
library(baytrends)

# Define function arguments
fn.import <- "dataCensored_TEST.csv"
qw.names <- c("secchi", "chla", "do", "tn", "tp", "po4f", "pp", "tdp"
              , "no23f", "nh4f", "tdn", "pn", "tss")

# fun function
dataCensored.test<- qw.import(fn.import, qw.names)

# Check for qw class
str(dataCensored.test)

# convert date field to POSIXct
dataCensored.test[,"date"] <- as.POSIXct(dataCensored.test[,"date"])

# recheck structure (other columns can be converted using
# as.numeric() and as.integer() if desired)
str(dataCensored.test)

# save the data frame for future use
save(dataCensored.test, file="dataCensored.test.rda")
```

Combining QW Data Frames

Use the function *rbindQW* to combine two data frames with "qw" formatted data. The below code chunk demonstrates combining the data from two data frames.

```r
library(baytrends)

newDF <- rbindQW(dataCensored[1:20,], dataCensored[101:120,])
head(newDF)

# Note, that it would have been equivalent and more efficient to
# use the below line of code to extract rows 1-20 and 101-120
newDF <- dataCensored[c(1:20,101:120),]
```

Additional Notes

`qw.import` includes an optional third argument, `rounding`, that controls how the data are rounded when printed. This argument is an integer vector of length 2. The first element is the maximum number of significant figures

and the second element is the maximum number of decimal places to show. The default value used is c(3, 4), but can be changed by modifying this option.