PBStools: Tools for the efficient execution of onerous tasks

PBStools provides an R interface for algorithms commonly used in fisheries. The scope of this package is by no means comprehensive, having grown from the need to satisfy tasks specific to British Columbia (BC) ocean fisheries. Many of the functions provide a quick way to visualize data, and in some cases perform preliminary analyses. Though oriented to users at the Pacific Biological Station (PBS), these functions may provide a broad utility to users at other locales. The User Guide is organised into sections that loosely classify the functions by theme – (1) Utility, (2) Biology, (3) Fishery, (4) Survey, (5) Spatial, (6) Assessment, (7) Catch Reconstruction, and (8) LaTeX tools. Within each section, the functions are described alphabetically.

PBStools depends heavily on two other R packages: **PBSmapping** and **PBSmodelling**. We use the latter to implement several Graphical User Interfaces (GUIs) that facilitate a few select functions. Most functions, however, are designed for use on the command line or in sourced code. Windows users need to be mindful that the R-statistical language is case sensitive.

Also available in the package directory `./library/PBStools/sql` we provide a number of useful SQL queries for DFO (Department of Fisheries and Oceans, aka Fisheries and Oceans Canada) commercial fisheries databases – `PacHarvest` (trawl, 1996-2007), `PacHarvHL` (hook and line, 1994-2006), `GFCatch` (historical landings, 1954-1995), `GFBioSQL` (biological samples, 1946-2018), and `GFFOS` (integrated fisheries, 2002-2018). To launch SQL queries, **PBStools** relies on the R package **RODBC**. If you have access to the DFO network and have privileges to query these databases, the function `getData` can send the queries to the remote servers and return a data frame called `PBSdat` in the global environment (or any user-specified environment). In the document, we highlight queries for DFO personnel using text with a background shaded 'moccasin'. (Examples are shaded 'aliceblue' and console output is shaded 'honeydew'.) Note that many of these queries might act as useful templates for users outside DFO for similar purposes. Querying databases directly via SQL commands from R usually proves much more efficient than launching Microsoft Access queries from a front-end database shell.

Originally, functions in **PBStools** evolved over time (2007-2012) within the R package **PBSfishery**, along with a convenient GUI tool for interacting with **PBSmapping** and useful datasets (regional boundaries, key codes, example data). In April 2012, we decided to split **PBSfishery** into three separate libraries – **PBStools**, **PBSmapx**, and **PBSdata** – for public distribution (see pbs-software). The three packages experience different rates of change, with **PBStools** undergoing frequent revision, while **PBSdata** and **PBSmapx** can remain unchanged for long periods of time.

Although **PBStools** is not available on CRAN (Comprehensive R Archive Network), the package (Windows binary and source tarball) is built after using CRAN's rigorous `R CMD check --as-cran` routine (using R-devel on a **Windows 7** 64-bit system) and posted to Google Drive. Most of the time, the revision on GitHub can be built (supposedly) in R using `devtools::install_github("pbs-software/pbs-tools")`; however, not every revision has been checked for CRAN worthiness.

In July 2018, the package was updated to include a new function called `linguaFranca`, which translates a limited set of BC stock assessment english words that appear in figures into french. This feature was largely driven by CSAS' mandate to make all scientific research documents accessible to anglophone and francophone readers.

One final note – although influences by Wizard Wickham pervade the Rverse these days, reliance on the convoluted machinations of the overlord will be deprecated in PBS packages wherever possible. While 'ggthis' and 'ggthat' provide code shortcuts, the underlying operations are not transparent. Furthermore, connecting ggblobs using pipes simply makes the mess even more difficult to clean up. Resist assimilation and keep your R code pure.

Another old R curmudgeon,
Rowan Haigh

Fisheries and Oceans Canada Pêches et Océans Canada Canada