

Qeco: Una aplicación de arquitectura abierta para el desarrollo de aplicaciones de R con interfaz gráfica

**Julio A. Di Rienzo¹, Laura Pla³, Mauricio J. Di Rienzo¹,
Sergio Vílchez², Fernando Casanoves²**

¹ Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba, Argentina;
² Centro Agronómico Tropical de Investigación y Enseñanza - CATIE. Turrialba. Costa Rica; ³ Universidad Nacional Experimental Francisco de Miranda. Venezuela.

Resumen

Qeco surge para dar respuesta a los requerimientos de la comunidad de ecólogos para implementar métodos estadísticos y cuantitativos modernos basados en R en el marco de una aplicación con una interfaz gráfica. La principal característica que hace a Qeco diferente es que une lo mejor de las aplicaciones basadas en menús con el poder de R en una aplicación de arquitectura abierta diseñada para crecer de acuerdo a las necesidades y el conocimiento de los usuarios. Esta arquitectura permite crear aplicaciones personalizadas que pueden aprovechar el esfuerzo de múltiples desarrolladores, favoreciendo el nacimiento de proyectos colaborativos.

Introducción

Qeco nace de una iniciativa de un equipo multidisciplinario internacional con vasta y diversa experiencia en el área de la estadística, ciencias biológicas y desarrollo de software: InfoStat[1], FDiversity [2], Runner [3], fgStatistics [4]. Qeco es un software orientado al análisis de datos ecológicos implementando métodos estadísticos y de ecología cuantitativa. Se caracteriza por presentar una interfaz de uso sencillo, intuitivo y transparente lo que hace una gran diferencia en la experiencia de analizar datos utilizando R. La principal característica que hace único a Qeco, es la combinación de su interfaz guiada por menú con el poder que ofrece R y su capacidad de adaptarse y personalizarse según las necesidades de los usuarios y sus habilidades para programar en R. Los ecólogos son demandantes de nuevos e integrados métodos cuantitativos. R es una plataforma de desarrollo de software muy utilizada en la comunidad de ecólogos, quienes generan scripts y paquetes ad hoc. La gran dificultad de estos esfuerzos de desarrollo es que no se encuentran integrados, son difíciles de ejecutar y no se encuentran disponibles en un repositorio común. Esto implica que muchos de los esfuerzos en el desarrollo de nuevos métodos queden aislados lo que se traduce en una pérdida de productividad. El propósito de Qeco es dar un contenedor estandarizado de aplicaciones desarrolladas en R que serán el núcleo de una aplicación estadística poderosa, compleja y fácil de usar.

Keywords: Qeco, ecología, script R, colaboración, estadística, bioinformática.

Elementos del Trabajo y Metodología

Qeco fue programado íntegramente en Delphi [5] para el desarrollo de la interfaz gráfica y R [6] para la implementación y desarrollo de métodos estadísticos. Funciona bajo la plataforma Windows en sus versiones XP o posteriores con arquitectura 32Bits y 64 Bits (En sistemas de 64 bits puede ser instalado pero se ejecuta bajo 32 bits). Para establecer comunicación entre la aplicación y R se utiliza Statconn DCOM [7].

Resultados

Qeco utiliza una interfaz gráfica de ventanas guiadas por menús. La aplicación presenta menús para el manejo de datos, análisis estadístico, gráficos, ayuda y menús basados en funciones personalizadas.

Datos: El manejo de datos permite, leer y escribir desde/hacia diferentes formatos estándar (*.xls, *.txt, *.r, entre otros). Las herramientas disponibles para el manejo de datos permiten editar, ordenar, categorizar, manipular columnas y filas, seleccionar casos con criterios definidos, unir tablas, aplicar formulas, etc.

El análisis exploratorio de datos resulta muy simple gracias a las flexibles e intuitivas herramientas estadísticas y graficas que Qeco tiene en su configuración básica. Estas herramientas permiten al usuario interactuar en un entorno de trabajo simple y poderoso para el análisis exploratorio.

Análisis cuantitativo: Para el análisis ecológico, Qeco dispone de métodos estándar y avanzados incluidos como parte del paquete de distribución del software. Las principales rutinas se encuentran organizadas bajo los menús: Statistics, Population, Communiuty, Ecosystem y Others. Cada rutina al ejecutarse muestra una ventana para seleccionar qué variables del archivo de datos utilizar y opciones para aplicar al análisis.

Intérprete de R: Una característica importante de Qeco es la integración con R. El usuario puede invocar un intérprete de R sin necesidad de abandonar el entorno de la aplicación. Así, el usuario no solo puede compartir datos entre el ambiente de Qeco y R, si no también realizar análisis complementarios utilizando la amplia variedad de paquetes desarrollados por la comunidad de usuarios R. Además dispone de una interfaz amigable para leer data frames a partir de paquetes R, permitiendo así, acceder a conjuntos de datos de prueba referenciados en la literatura de R. El usuario puede invocar o copiar scripts R de su colección o de recursos publicados, e incluso ejecutarlos y modificarlos para satisfacer demandas específicas.

Menús personalizados: Además de los menús pre-configurados de Qeco, esta plataforma tiene la capacidad de agregar contribuciones (“add-ons”) desarrolladas por los usuarios. Esto se reflejará en nuevos menús que el usuario incorpora de manera personalizada desde un repositorio común.

Para implementar contribuciones-R es necesario seguir una serie de requisitos para que Qeco pueda interpretarlas y ubicarlas en la estructura de la aplicación. Todas las contribuciones-R deben estar al menos en tres carpetas en una jerarquía (una dentro de la otra). La primera es la carpeta donde residen todas las contribuciones del usuario. Esa carpeta es declarada por el usuario la primera vez que quiere agregar una contribución.

El nombre de la segunda carpeta define el menú donde la contribución aparecerá listada. El nombre de la tercera carpeta define el nombre del ítem del menú correspondiente a la contribución. El script de R que contiene el código a ejecutar se encuentra en esta última carpeta. Si el usuario desea crear un ítem llamado `MyFirstItem` dentro del menú `Statistics` (`Statistics` es un menú estándar dentro de la aplicación), deberá crear una carpeta llamada “`Statistics`” dentro de carpeta que declaró como repositorio de las contribuciones y dentro de “`Statistics`” crear la carpeta llamada `MyFirstItem`.

Para finalizar la instalación de la contribución, la carpeta más interna, en este caso `MyFirstItem`, deberá contener un script de R que debe ser nombrado: `MyMainRStrip.R`. Si este archivo no se encuentra presente, la aplicación no agregará el nuevo ítem del menú. Con esta estructura de carpetas, la aplicación agregará, cuando arranca, el nuevo ítem llamado `MyFirstItem` al final del menú “`Statistics`”.

MyMainRStrip.R: Este script es el núcleo de la contribución-R. Es un script de R que contiene la interfaz para interactuar con Qeco y el código R para procesar los datos. Las instrucciones para la interfaz deberán incluirse dentro de *MyMainRStrip.R* como comentarios de R. El script contiene tres secciones: `HEADER`, `OPTIONS` y `CODIGO`

La sección `HEADER` es obligatoria y se encuentra delimitada por las directivas `#HEADER.BEGIN` y `#HEADER.END`. Las declaraciones en esta sección van a definir el comportamiento de una ventana de dialogo llamada “selección de variables” que se muestra cuando se ejecuta el script desde el menú de Qeco. El dialogo emergente permite seleccionar que parte del archivo de datos disponible será utilizado por el script de R. Para ello Qeco crea objetos R (data-frames) que contienen los datos que el script debe procesar.

Dentro del `HEADER` también se encuentra la declaración de la base de datos de prueba de procedimiento y la fuente de la ayuda. Esta puede estar en formato pdf o estar declarada como una URL. También en esta sección se puede incluir la información sobre autoría del desarrollo, referencias bibliográficas, fecha de actualización y otra información que aparece en el encabezamiento de los resultados.

La sección `OPTIONS` es opcional. Describe el contenido de una ventana de diálogo que permite al desarrollador de la contribución solicitar al usuario información adicional y opciones para la ejecución del script. En esta ventana el desarrollador puede incluir un dispositivo para elección de opciones no excluyentes (`CHECKLIST`), un dispositivo para opciones alternativas (`RADIOGROUP`), un dispositivo para que el usuario ingrese datos alfanuméricos en un campo de edición (`LISTEDIT`) y finalmente un dispositivo para que el usuario pueda ingresar texto (`MEMO`). Además, se encuentran otras opciones relacionadas a la selección de archivos de entrada y salida e instalación de librerías requeridas por el script entre otras.

Finalmente el `CODIGO` es una sección de instrucciones ejecutables por R y está integrada principalmente por una función: *TheRMainFunction()*. Esta contiene el código principal de la contribución R y se escribe utilizando los objetos R que surgen de las secciones `HEADER` y `OPTIONS`.

Discusión

R es una plataforma de desarrollo potente que tiene una creciente aceptación en el ámbito académico. Sin embargo su interfaz es muy pobre y demanda un importante nivel de experiencia en programación para poder hacer un uso productivo del mismo. Los usuarios de R pueden sentirse frustrados por las dificultades que encuentra para ejecutar scripts simples. La disponibilidad de una plataforma de desarrollo de aplicaciones basadas en R pero con una interfaz gráfica resuelve el dilema de aquellos que desean una interfaz simplificada pero que no quieren renunciar al poder de cálculo y a la amplia base de librerías disponibles en R. Qeco inaugura un modelo de desarrollo de aplicaciones que redundará en una mayor productividad para la comunidad de ecólogos.

Agradecimientos

Este trabajo es soportado parcialmente por los proyectos: Conservation International. PICT 32905.

Referencias

- [1] Di Rienzo J.A., Casanoves F., Balzarini M.G., Gonzalez L., Tablada M., Robledo C.W. (2011) InfoStat versión 2011. Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina. URL <http://www.infostat.com.ar>.
- [2] Casanoves F., Pla L., Di Rienzo J.A., Díaz S. (2010). FDiversity: a software package for the integrated analysis of functional diversity. *Methods in Ecology & Evolution* doi: 10.1111/j.2041-210X.2010.00082.x
- [3] Di Rienzo J.A. (2011). fgStatistics. Statistical software for the analysis of experiments of functional genomics. URL <http://sites.google.com/site/fgStatistics/>
- [4] Di Rienzo J.A., FCA, Universidad Nacional de Córdoba, Argentina. URL <http://sites.google.com/site/runner/>
- [5] <http://www.codegear.com/products/delphi/win32>
- [6] R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>
- [7] Thomas Baier and Erich Neuwirth (2007) Excel::COM::R. *Computational Statistics*, 22, Number 1/April. Physica Verlag.