Introducción a la aplicación web Ema



Guzmán López* Asistente en manejo de datos oceanográficos Proyecto FREPLATA URU/09/G31

Diciembre, 2013

Resumen

La aplicación web Ema ("Estaciones de Monitoreo Ambiental") está diseñada para permitirle al usuario visualizar, analizar, y editar de manera interactiva datos oceanográficos. Fue desarrollada en el marco del Proyecto FREPLATA URU/09/G31 dentro del "Programa de Monitoreo y Evaluación y Sistema de Información Integrado y establecido para la toma de decisiones y la Gestión del Río de la Plata y su Frente Marítimo". El objetivo era generar una herramienta que permitiese a los usuarios procesar datos oceanográficos provenientes de las Estaciones de Monitoreo Ambiental del Río de la Plata (Boya oceanográfica, Pilote Norden y Torre Oyarvide). Todo el software y librerías empleadas en esta aplicación son gratuitas, y la mayor parte del software empleado es libre. Para su desarrollo se utilizó fundamentalmente el software R (estadística computacional y gráficos) y la librería Shiny (marco de aplicaciones web para R).

*contacto: <u>guzilop@gmail.com</u>

Contenido

1 - Introducción	3
1.1 - Justificación	3
2 - La aplicación web Ema	3
2.1 - Descripción técnica	3
2.2 - Requerimientos e instalación	4
2.2.1 - Modo local	4
2.2.2 - Modo red	5
2.3 - Utilizando Ema	5
2.3.1 - Descripción de la interfaz general de usuario (GUI)	6
2.3.2 - Cargar datos desde un archivo de texto	6
2.3.3 - Cargar datos desde un archivo de Microsoft Excel	7
2.3.4 – Tabla de datos	8
2.3.5 - Plots	g
2.3.6 - Editar	13
2.3.7 - Resumen de los datos	14
2.3.8 - Convertir conductividad	15
3 - Código fuente	17
3.1 - script server.R	17
3.1 - script ui.R	17
4 - Referencias	18

1 - Introducción

1.1 - Justificación

El proyecto FREPLATA URU/09/G31 a través de su Programa de Monitoreo y un Sistema Binacional Integrado de Información para apoyar la toma de decisiones, busca conformar y consolidar un Sistema de Información Ambiental Costero Marino (SIACM) a nivel nacional para el Río de la Plata y su Frente Marítimo (RPFM). Con esta intención, es que ha evaluado las necesidades de información e intereses por parte de las instituciones competentes en la gestión sobre el RPFM con miras de fomentar las capacidades institucionales y la articulación de estrategias. En tal sentido se contrató a tres asistentes en manejo de datos oceanográficos para que trabajen en el diagnóstico de la información existente y el diseño de bases de datos del Servicio de Oceanografía Hidrografía y Meteorología de la Armada (SOHMA) y de la Dirección Nacional de Recursos Acuáticos (DINARA). Para facilitar el acceso, visualización y análisis de los datos oceanográficos provenientes de las estaciones fijas de monitoreo ambiental del Río de la Plata (Boya oceanográfica, Pilote Norden y Torre Oyarvide) se desarrolló la aplicación web Ema. Buscando que usuarios no expertos puedan acercarse a los datos oceanográficos mediante una interfaz general amigable e interactiva.

2 - La aplicación web Ema

2.1 - Descripción técnica

La aplicación web Ema consta de un directorio con dos scripts: *server.R* y *ui.R*. En el script *server.R* se asignan las entradas para los cálculos a realizarse y se le asignan expresiones reactivas, y en el script *ui.R* se define la interfaz general de usuario de la aplicación. Además, dichos scripts también incluyen información adicional y accesos a

recursos que requiera la aplicación para su funcionamiento. Por más información, en la sección 4 – Código fuente de esta guía se pueden descargar ambos scripts.

2.2 - Requerimientos e instalación

La aplicación web Ema puede ejecutarse localmente o a través de la red con conexión a Internet.

2.2.1 – Modo local

Para correr la aplicación web Ema localmente se requiere instalar la siguiente lista de software y librerías de R:

- R (versión > 2.15.2)
- librería "Shiny, Web Application Framework for R" (versión 0.8.0)
- librería "oce, Analysis of Oceanographic data" (versión 0.9-12)
- librería "gnumeric, Read data from files readable by gnumeric" (versión 0.7-2)
- librería "xts, eXtensible Time Series" (versión 0.9-7)
- librería "googleVis, Interface between R and the Google Chart Tools" (versión 0.4.7)
- Explorador web (ej. Mozilla Firefox, Chromium, Google Chrome, etc).
- Adobe Flash Player plugin.

Para ejecutar la aplicación Ema se debe ejecutar la siguiente sentencia en la línea de comandos de R: shiny::runApp(appDir='/directorio/Ema/'). El directorio corresponde a la ubicación donde se encuentra la carpeta "Ema" con los archivos server.R y ui.R.

Nota: el explorador web debe tener habilitado el Adobe Flash Player plugin y permitir JavaScript.

2.2.2 - Modo red

Para correr la aplicación web Ema a través de la red se requiere configurar y mantener un servidor Linux. Además se necesita instalar la siguiente lista de software y librerías de R en el servidor Shiny:

- R (versión > 2.15.2)
- librería "Shiny, Web Application Framework for R" (versión 0.8.0)
- Shiny Server software (http://www.rstudio.com/shiny/server/)
- librería "oce, Analysis of Oceanographic data" (versión 0.9-12)
- librería "gnumeric, Read data from files readable by gnumeric" (versión 0.7-2)
- librería "xts, eXtensible Time Series" (versión 0.9-7)
- librería "googleVis, Interface between R and the Google Chart Tools" (versión 0.4.7).

Utilizando un servidor Linux los usuarios acceden a la aplicación web Ema a través de una página web.

Otra opción es acceder mediante el repositorio GitHub donde se almacena el código fuente ejecutando la siguiente sentencia en la línea de comandos de R: shiny::runGitHub(repo='Ema', username='guzmanlopez').

Nota: como instalar Shiny Server (http://www.rstudio.com/shiny/server/install-opensource) y como configurar Shiny Server (http://rstudio.github.io/shiny-server/latest/).

2.3 - Utilizando Ema

La aplicación web EMA ("Estaciones de Monitoreo Ambiental") está diseñada para permitirle al usuario visualizar, analizar, y editar de manera interactiva datos oceanográficos.

2.3.1 - Descripción de la interfaz general de usuario (GUI)

La aplicación web Ema consta de un panel lateral ubicado a la izquierda de la pantalla donde se especifican los controles de entrada y de un panel principal ubicado a la derecha del panel lateral con una serie de pestañas [Tabla], [Plots], [Editar], [Resumen], [Convertir conductividad] y [Acerca de esta APP] donde se muestran las salidas (Figura 1).

2.3.2 - Cargar datos desde un archivo de texto

Para cargar datos de un archivo de texto, dirigirse a la pestaña [**Tabla**] del panel principal y en **Tipo de archivo** seleccionar **Texto**. Hacer clic en el botón **<Choose File>** y seleccionar el archivo deseado (Figura 1).

En la sección de **Opciones** se deben especificar las características del archivo para que los datos sean cargados correctamente y se vean bien en el panel principal. La opción **Cabecera** se debe seleccionar si el archivo cargado posee columnas con títulos. La opción **Separador de campos** hace referencia a qué símbolo define las celdas, y estos pueden ser *Coma*, *Punto y coma*, o *Tabulador*. La opción **Delimitador de texto** hace referencia a cómo se delimitan los campos de texto y estos pueden ser *Ninguno*, *Comillas dobles*, o *Comillas simples*. Según la configuración regional del archivo seleccionado la opción **Separador decimal** puede ser *Punto* o *Coma*. La opción *Leer desde línea*: define desde que línea de texto se comienza a leer el archivo y es un

valor numérico entero que debe definir el usuario. Por último se debe establecer el **Formato de Fecha y Hora**, cuyas opciones varían en función del orden de aparición de los campos (día, mes y año o mes, día y año), de los delimitadores de campos de la fecha ("/" o "-") y de si el año está por extenso (ej. "2009" o "09").

Las opciones establecidas por defecto en Ema para abrir archivos de texto son: con cabecera, campos separados por comas, sin delimitador de campos de texto, con punto como separador decimal, leer desde la línea cero (no excluir ninguna línea de texto) y formato de fecha y hora como "mes/día/año en extenso y hora".

Las características del archivo a cargarse debe ser como se muestra en la tabla de la Figura 1. Es decir, debe constar de ocho columnas donde la primer columna es la fecha junto con la hora.

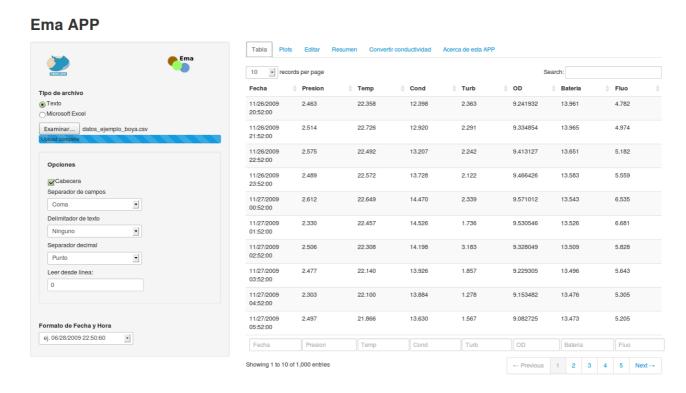


Figura 1 - Captura de pantalla de como cargar datos desde un archivo de texto con Ema. A la izquierda con fondo gris oscuro se ubica el panel lateral donde se configuran las entradas a la aplicación (sección Opciones). A la derecha, donde se visualizan los datos se ubica el panel principal donde se muestran la salidas de la aplicación, en este caso una tabla con los datos cargados.

2.3.3 - Cargar datos desde un archivo de Microsoft Excel

Para cargar datos de un archivo de Microsoft Excel, dirigirse a la pestaña [**Tabla**] del panel principal y en **Tipo de archivo** seleccionar *Microsoft Excel*. Hacer clic en el botón <*Choose File*> y seleccionar el archivo deseado (solo terminación ".xls").

En la sección de **Opciones** se deben especificar las características del archivo para que los datos sean cargados correctamente y se vean bien en el panel principal. La opción **Cabecera** se debe seleccionar si el archivo cargado posee columnas con títulos. En la opción **Nombre de hoja** el usuario debe escribir el nombre de la planilla en el campo de texto vacío. El nombre de la planilla es aquel que aparece en la región inferior cuando el archivo es abierto con un programa que abre hojas de cálculo (ej. Microsoft Excel o LibreOffice Calc). En general, por defecto toma el nombre de "Sheet1" u "Hoja1". El nombre debe ser escrito sin comillas. Por último se debe establecer el **Formato de Fecha y Hora**, cuyas opciones varían en función del orden de aparición de los campos (día, mes y año o mes, día y año), de los delimitadores de campos de la fecha ("/" o "-") y de si el año está por extenso (ej. "2009" o "09").

Las opciones establecidas por defecto en Ema para abrir archivos de Microsoft Excel son: con cabecera y formato de fecha y hora como "mes/día/año en extenso y hora".

Las características del archivo a cargarse debe ser como se muestra en la tabla de la Figura 1. Es decir, debe constar de ocho columnas donde la primer columna es la fecha junto con la hora.

Nota: la opción cargar datos desde un archivo de Microsoft Excel es experimental y puede presentar fallas como no lograr cargar el archivo elegido. Es un complemento ya que siempre es posible convertir el archivo de Microsoft Excel a texto (formato no restrictivo) y ser cargado con la opción de cargar datos desde un archivo de texto.

2.3.4 - Tabla de datos

Una vez que se hayan cargado correctamente los datos y se visualizan los campos como son en el archivo original en el panel principal (Figura 1), Ema permite generar algunos filtros básicos de los mismos. Se puede elegir visualizar entre 10, 25, 50, o 100 registros de datos por página con al opción **records per page**, permite buscar un valor determinado en el campo **Search**, y se pueden ordenar la tabla de datos por columna en orden ascendente o descendente haciendo clic sobre el título de la columna. La búsqueda también se puede hacer por columna introduciendo un valor en los campos vacíos que están al final de cada una. Con los botones **Previous** y **Next**> se pasa a la hoja previa o a la siguiente respectivamente. Con los botones numerados entre **Previous** y **Next**> se accede directamente a la hoja elegida.

Nota: los filtros establecidos en esta sección son meramente informativos y no son aplicados en procesos posteriores ni sobre los datos originales. Ema no modifica ni escribe sobre el archivo cargado.

2.3.5 - Plots

En la pestaña [Plots] es posible acceder a tres gráficos diferentes en la sección Gráfico: Series de tiempo, Relaciones entre variables e Histogramas. Para cada gráfico se puede configurar la Fuente de datos, la Escala temporal y seleccionar la/s variable/s en la sección Elegir variables.

La **Fuente de datos** puede ser *Original* si queremos ver los datos cargados sin modificación o *Editados* si queremos ver la edición manual e interactiva realizada a los datos originales en la pestaña [Editar].

La **Escala temporal** puede ser *Original* si queremos ver una escala de tiempo idéntica a la de los datos cargados o *Promedios diarios* si queremos ver el promedio por día,

Promedios semanales si queremos ver el promedio por semana, **Promedios mensuales** si queremos ver el promedio por mes, o **Promedios anuales** si queremos ver el promedio por año.

En la sección **Elegir variables** aparecen listadas solamente aquellas variables que contenga el archivo cargado, por ejemplo: *Temperatura, Conductividad, Turbiedad*, etc. Para el plot **Series de tiempo** se pueden seleccionar todas las variables que se deseen, para el plot **Relaciones entre variables** solo se deben seleccionar dos variables, y para el plot de **Histogramas** solo se puede seleccionar una variable.

El plot **Series de tiempo** permite visualizar una o más variables en función del tiempo (Figura 2). Es un gráfico interactivo que cuenta con las siguientes opciones:

- Hacer zoom sobre el gráfico seleccionando una escala temporal prefijada por defecto en un minuto ("1"), cinco minutos ("5"), una hora ("1h"), un día ("1d"), cinco días ("5d"), un mes ("1m"), tres meses ("3m"), seis meses ("6m"), un año ("1y"), y la escala máxima que abarcan los datos cargados ("Max").
- Haciendo clic con el botón izquierdo del ratón sobre el panel principal del gráfico y moviendo el ratón se avanza o retrocede en el tiempo.
- Con la rueda del ratón aumentamos el zoom o lo disminuimos.
- Se puede seleccionar el rango de tiempo que se desee utilizando el panel inferior del gráfico, de manera de estirar o contraer el selector desde sus bordes o desplazándolo.
- Pasando el ratón sobre las gráficas nos arroja el valor que toman las variables.

En la sección **Personalizar** del plot **Series de tiempo** se puede definir si el **Eje "Y":** del gráfico está **Maximizado** (se muestra el rango de valores entre el máximo y el

mínimo para el zoom seleccionado) o es *Fijo* (valor mínimo siempre es cero). También se puede definir el porcentaje de relleno de color (0% a 100%) por debajo de la curva graficada con la opción *Relleno*.

Ema APP



Figura 2 - Captura de pantalla de Series de tiempo para datos de temperatura.

El plot **Relaciones entre variables** permite graficar una variable en función de otra variable (Figura 3). Posicionándonos sobre un punto de la gráfica con el ratón podemos conocer los valores exactos que toma para las dos variables. Para normalizar las escalas de ambas variables que se están relacionando es posible elegir la escala logarítmica seleccionando la opción **Logaritmo**.

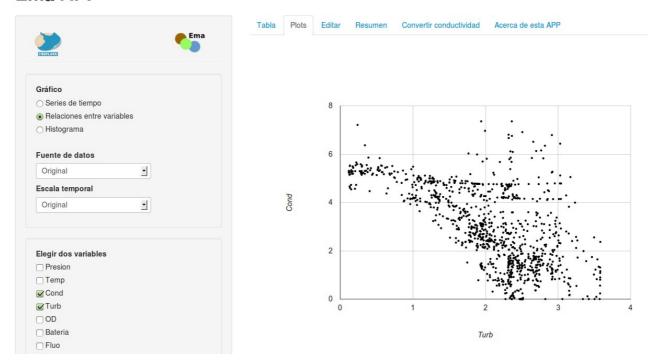


Figura 3 - Captura de pantalla de Relaciones entre variables para datos de Turbiedad vs conductividad en escala logarítmica.

El plot de **Histogramas** nos permite ver la distribución de la variable seleccionada mediante barras que representan las frecuencias de las clases representadas (Figura 4). Las clases dadas por defecto en este histograma están dadas por la fórmula de Sturges (1926).

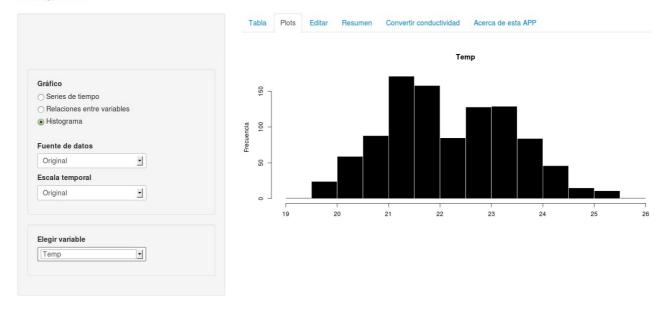


Figura 4 - Captura de pantalla de un Histograma para datos de Temperatura.

2.3.6 - Editar

En la pestaña [**Editar**] es posible procesar los datos cargados de manera interactiva para señalar datos erróneos (Figura 5). En la sección **Elegir variable** definimos la variable a editar. Luego en la sección **Intervalo** definimos la cantidad de datos que queremos observar en el gráfico y las opciones predefinidas son: **25**, **50**, **100**, **250**, **500**, **750**, **1000**, **1500** y **2000**. Con los botones **Atrás** y **Siguiente** podemos retroceder y avanzar el intervalo de datos que hayamos definido anteriormente.

Para seleccionar un dato que consideramos erróneo solo se debe hacer clic sobre el punto de interés e inmediatamente dicho punto será marcado de color rojo ("Errores") en la gráfica.

Todos los datos editados pueden descargarse junto con los originales como un archivo de texto (".csv") apretando el botón **Descargar datos editados**> en el panel lateral.

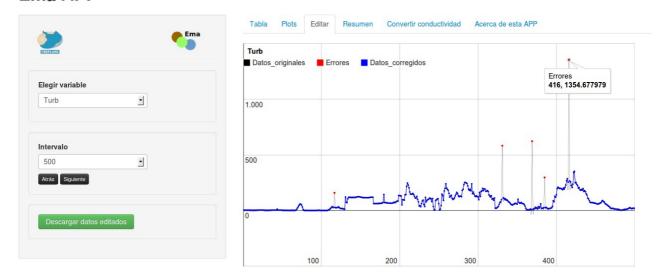


Figura 5 - Captura de pantalla de Editar para datos de Turbiedad. En el gráfico se pueden observar los puntos seleccionados como errores en color rojo, la línea negra como la serie de datos originales y la azul como la serie de datos corregidos (datos originales sin los marcados como errores).

Nota: la aplicación web Ema no elimina ningún dato. Los datos que el usuario seleccione como errores solamente son marcados para que puedan filtrarse.

2.3.7 - Resumen de los datos

En la pestaña [**Resumen**] es posible obtener estadísticas descriptivas básicas ("Promedio", "Desvío estándar", "Mediana", "Valor máximo", "Valor mínimo", "Cantidad de datos") de cada una de las variables (Figura 6). En el panel lateral en la sección **Resumen** se debe definir si la fuente de los datos son los **Datos originales** o los **Datos editados**. Haciendo clic sobre el nombre de la columna podemos ordenar la tabla en orden ascendente o descendente.



Figura 6 - Captura de pantalla de Resumen de los datos originales cargados a la aplicación donde se muestran estadísticas descriptivas básicas de los datos.

2.3.8 - Convertir conductividad

En la pestaña [Convertir conductividad] es posible convertir los datos de conductividad a salinidad utilizando el algoritmo de la UNESCO descrito por Fofonoff and Millard (1983) (Figura 7). Para ello es necesario definir la Fuente (Datos originales o Datos editados). Para cada sección de cada una de las variables (Conductividad, Temperatura y Profundidad) se debe definir si tienen un valor Constante (Si o No); esto quiere decir que si elegimos que "Si" debemos ingresar un único valor que se empleará en el cálculo, mientras que si elegimos que "No" se utilizan los valores cargados en los datos originales. Además debe especificarse correctamente el nombre de la variable para cada sección en Elegir variable. Para la conductividad se debe especificar las unidades en Unidades, y estas pueden ser mS/cm (mili siemens por centímetro) o S/m (siemens por metro). Luego de configurar todos los elementos de entrada en el panel lateral, se muestran los valores de salinidad en unidades prácticas de salinidad en el panel principal. Dicha columna puede ser filtrada como la salida de los datos cargados en la pestaña [Tabla] (ver sección 2.2.4 – Tabla de datos).

En la sección **Descargar Salinidad** es posible descargar solo la columna que se

muestra en el panel principal eligiendo **Solo salinidad** o descargar la salinidad junto con los datos cargados originales (no incluye los editados) eligiendo **Salinidad junto con datos cargados**. Se debe apretar el botón **Descargar datos**> y automáticamente podemos descargar la opción elegida como un archivo de texto (".csv").

Ema APP

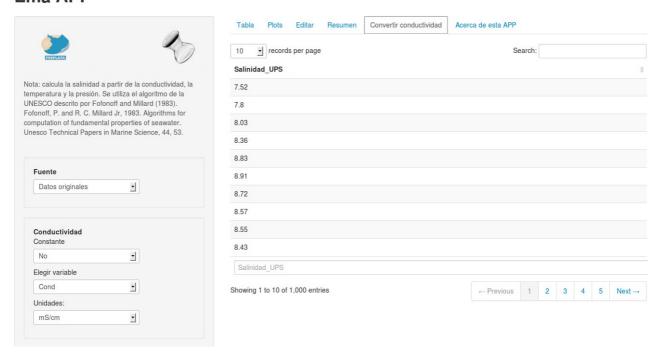


Figura 7 - Captura de pantalla de Convertir conductividad. En el panel lateral se especifica la referencia de la ecuación implementada para el cálculo, la fuente de los datos y las variables utilizadas. En el panel principal se visualizan los valores de salinidad en unidades prácticas de salinidad para los datos cargados.

Nota: la aplicación web Ema sugiere una columna de los datos originales para cada una de las variables, pero no siempre puede coincidir ya que depende de cómo se le haya llamado en los datos originales. Se sugiere chequear estos campos siempre.

2.3.9 - Acerca de esta aplicación web

En la pestaña [Acerca de esta APP] es posible acceder a una descripción general de la aplicación (Figura 8). En la misma se especifica para qué sirve Ema, en el marco del proyecto en que se estuvo desarrollando y el objetivo de su desarrollo. Incluye también las referencias bibliográficas a las distintas librerías de R utilizadas para correr la aplicación y un contacto con el autor del desarrollo de la aplicación para que si a un usuario le surgen dudas pueda consultarlas o pueda reportar "bugs".

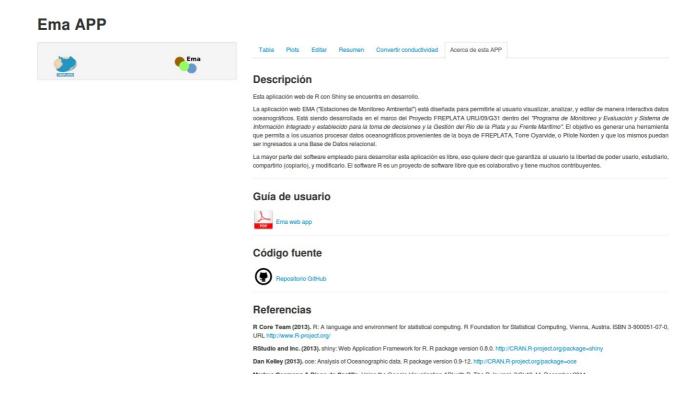


Figura 8 - Captura de pantalla de Acerca de esta APP. En esta parte se describe la aplicación web y se puede acceder mediante hipervínculos a esta guía de usuario, al código fuente y a las citas del software y librerías utilizadas.

3 - Código fuente

3.1 - script server.R

https://github.com/guzmanlopez/Ema/blob/master/server.R

3.1 - script ui.R

https://github.com/guzmanlopez/Ema/blob/master/ui.R

4 - Referencias

Dan Kelley (2013). oce: Analysis of Oceanographic data. R package version 0.9-12. http://CRAN.R-project.org/package=oce

Jeffrey A. Ryan & Joshua M. Ulrich (2013). xts: eXtensible Time Series. R package version 0.9-7. http://r-forge.r-project.org/projects/xts/

Karoly Antal. (2012). gnumeric: Read data from files readable by gnumeric. R package version 0.7-2. http://CRAN.R-project.org/package=gnumeric

Markus Gesmann & Diego de Castillo. Using the Google Visualisation API with R. The R Journal, 3(2):40-44, December 2011.

R Core Team (2013). 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/

RStudio and Inc. (2013). shiny: Web Application Framework for R. R package version 0.8.0. http://CRAN.R-project.org/package=shiny

Sturges, H. A. (1926). The choice of a class interval. Journal of the American Statistical Association 21, 65–66.