

*Introduction to  
R Statistical Software*

**Anthony (Tony) R. Olsen  
USEPA ORD NHEERL  
Western Ecology Division  
Corvallis, OR 97333  
(541) 754-4790  
Olsen.Tony@epa.gov**

# *What is R?*

- A language and environment for statistical computing and graphics
- Based on the “S” system developed by Bell Laboratories
- From R Project for Statistical Computing
- R is available as Free Software
- S-Plus is a commercial implementation of the “S” system



# *The R Environment is an integrated suite of software*

- effective data handling and storage
- calculations on spreadsheets and arrays
- integrated collection of statistical analysis tools
- graphical display on-screen or hardcopy
- a programming language
- add capability using packages (libraries)



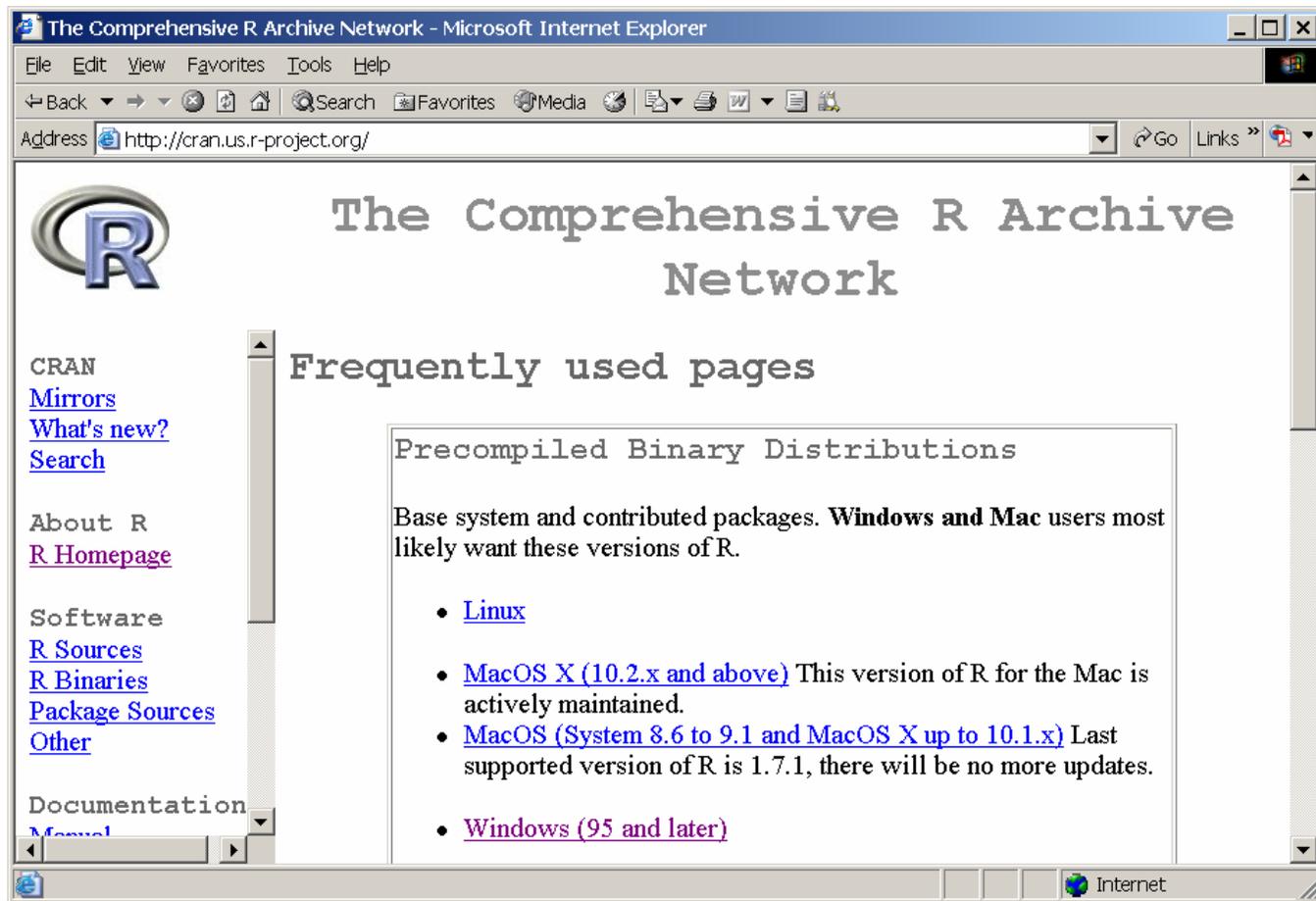
# *Why use R for Aquatic Monitoring Survey Design & Analysis?*

- Needed way to provide survey design and survey analysis algorithms to others (particularly states and tribal nations)
- No existing statistical software has the algorithms required
- Many organizations have limited ability to purchase any recommended commercial statistical software
- R is free, easily installed, and usable with training
- Our approach is to provide libraries that minimize knowledge required about R

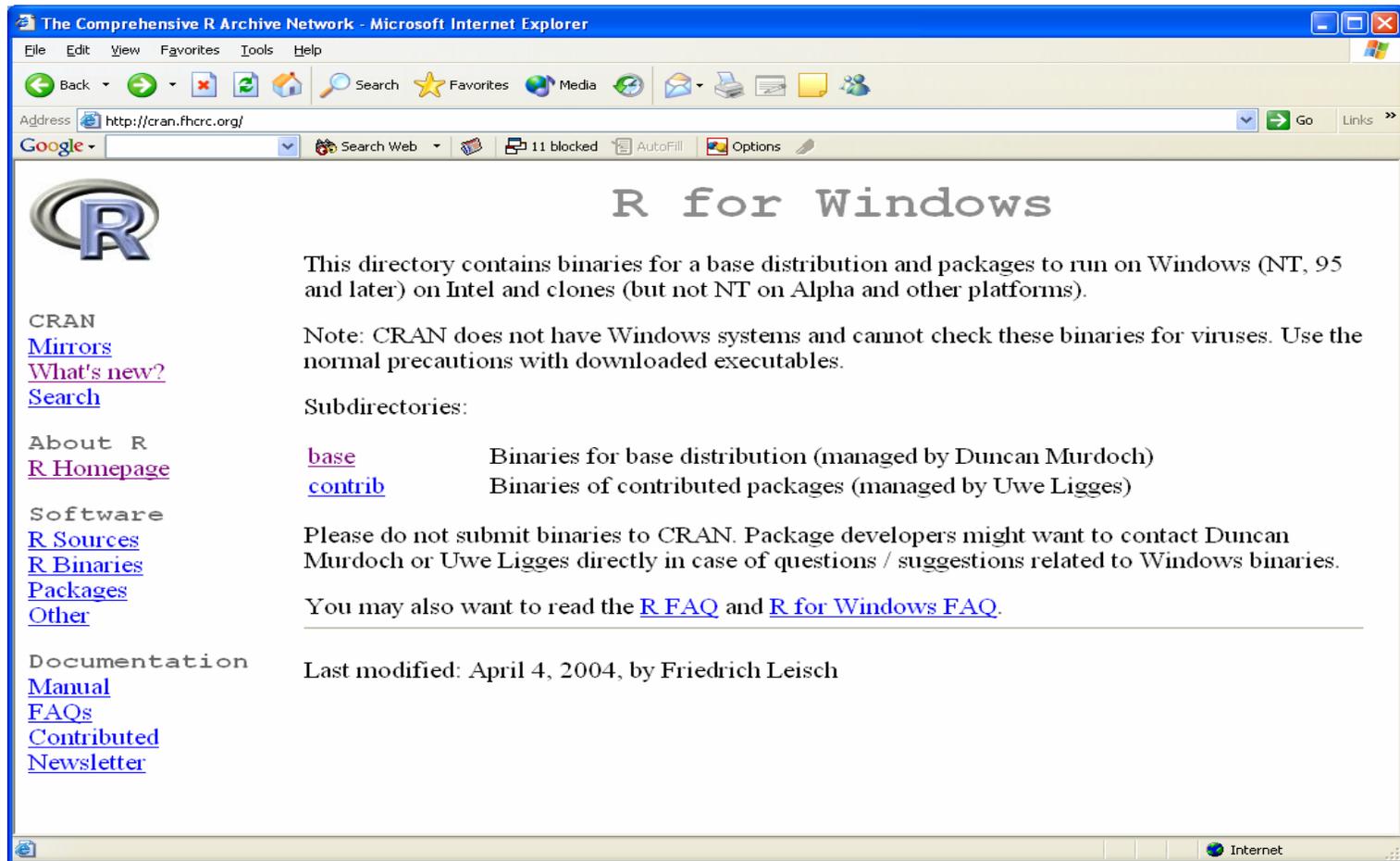


# *Acquiring and Installing R*

*Use web browser to access <http://cran.us.r-project.org/>*



# Click on base



The screenshot shows a Microsoft Internet Explorer browser window titled "The Comprehensive R Archive Network - Microsoft Internet Explorer". The address bar displays "http://cran.fhrc.org/". The page content includes the R logo, the heading "R for Windows", and a description of the directory. The page is organized into several sections: CRAN links (Mirrors, What's new?, Search), About R (R Homepage), Software (R Sources, R Binaries, Packages, Other), and Documentation (Manual, FAQs, Contributed, Newsletter). The page also contains a note about CRAN's virus scanning policy, a list of subdirectories (base and contrib), a warning about submitting binaries, and a reference to FAQs. The page was last modified on April 4, 2004, by Friedrich Leisch.

**R for Windows**

This directory contains binaries for a base distribution and packages to run on Windows (NT, 95 and later) on Intel and clones (but not NT on Alpha and other platforms).

Note: CRAN does not have Windows systems and cannot check these binaries for viruses. Use the normal precautions with downloaded executables.

Subdirectories:

- [base](#) Binaries for base distribution (managed by Duncan Murdoch)
- [contrib](#) Binaries of contributed packages (managed by Uwe Ligges)

Please do not submit binaries to CRAN. Package developers might want to contact Duncan Murdoch or Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Last modified: April 4, 2004, by Friedrich Leisch

**CRAN**  
[Mirrors](#)  
[What's new?](#)  
[Search](#)

**About R**  
[R Homepage](#)

**Software**  
[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Other](#)

**Documentation**  
[Manual](#)  
[FAQs](#)  
[Contributed](#)  
[Newsletter](#)



*Click on **CHANGES** to see changes*  
*Click on **README.R-2.3.0** for instructions*  
*Click on **R-2.3.0-win32.exe** to download*

The Comprehensive R Archive Network - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Refresh Print Mail New Tab

Address <http://cran.fhrc.org/> Go Links

# R-2.3.0 for Windows

This directory contains a binary distribution of R-2.3.0 to run on Windows 95, 98, ME, NT4.0, 2000 and XP on Intel/clone chips.

Patches to this release are incorporated in the [r-patched snapshot build](#).

A build of the development version (which will eventually become the next major release of R) is available in the [r-devel snapshot build](#).

In this directory:

- [README.R-2.3.0](#) Installation and other instructions.
- [CHANGES](#) New features of this Windows version.
- [NEWS](#) New features of all versions.
- [R-2.3.0-win32.exe](#) Setup program (about 27 megabytes). Please download this from a [mirror near you](#). This corresponds to the file named **SetupR.exe** or **rwXXXX.exe** in pre-2.2.0 releases.
- [old](#) The previous release.

CRAN

- [Mirrors](#)
- [What's new?](#)
- [Task Views](#)
- [Search](#)

About R

- [R Homepage](#)

Software

- [R Sources](#)
- [R Binaries](#)
- [Packages](#)
- [Other](#)

Documentation

Local intranet



RESEARCH & DEVELOPMENT

*Building a scientific foundation for sound environmental decisions*

# Documentation on Using R

The Comprehensive R Archive Network - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://cran.fhcrc.org/> Go Links

## The R Manuals

*edited by the R Development Core Team.*

Current Version: 2.3.0 (April 2006)

The following manuals for R are downloadable as PDF files:

- **An Introduction to R** is based on the former "Notes on R", gives an introduction to the language and how to use R for doing statistical analysis and graphics. [[browse HTML](#) | [download PDF](#) ]
- A draft of **The R language definition** documents the language *per se*. That is, the objects that it works on, and the details of the expression evaluation process, which are useful to know when programming R functions. [[browse HTML](#) | [download PDF](#) ]
- **Writing R Extensions** covers how to create your own packages, write R help files, and the foreign language (C, C++, Fortran, ...) interfaces. [[browse HTML](#) | [download PDF](#) ]
- **R Data Import/Export** describes the import and export facilities available either in R

Done Local intranet



RESEARCH & DEVELOPMENT

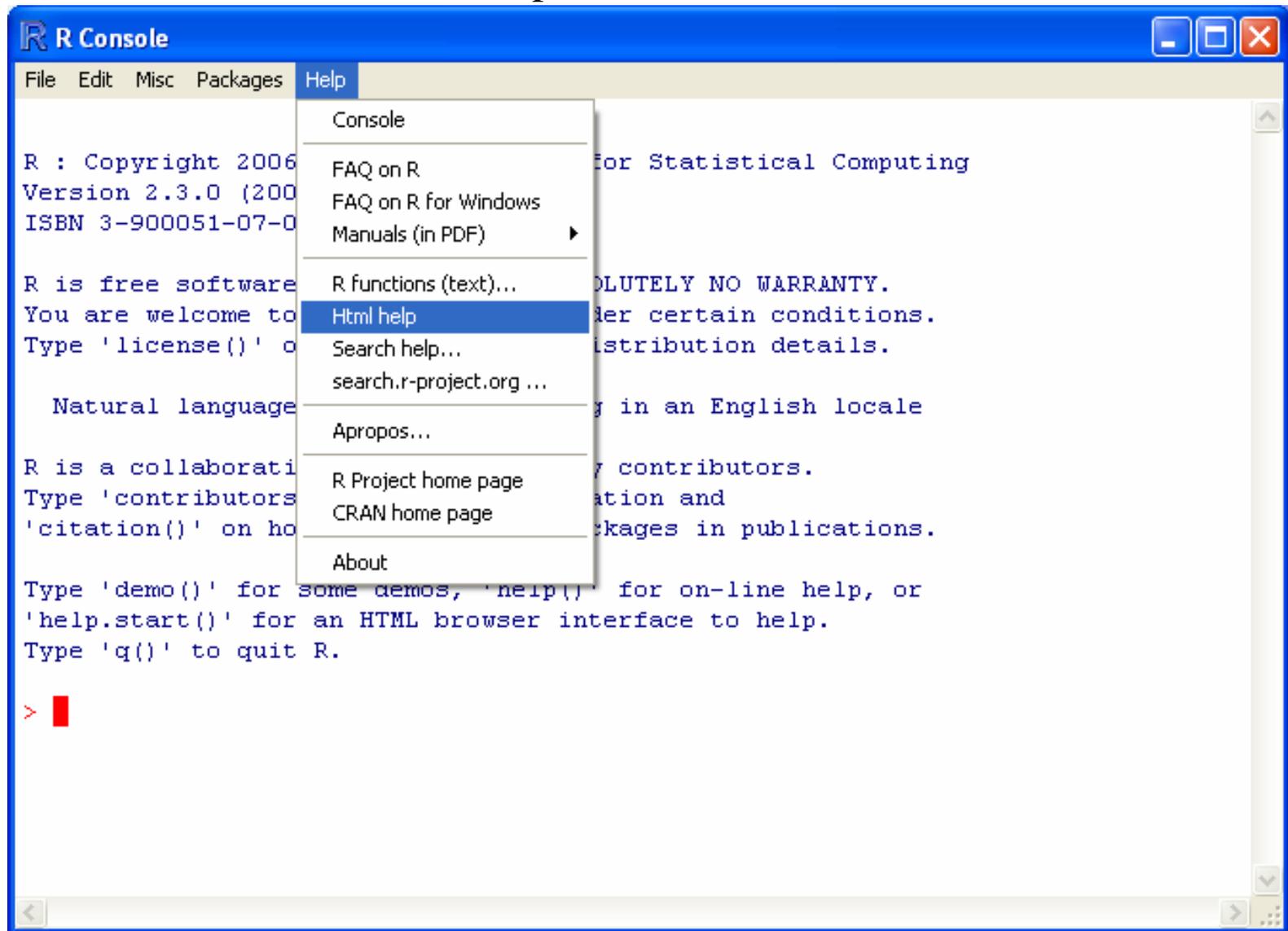
*Building a scientific foundation for sound environmental decisions*

# *Installing R*

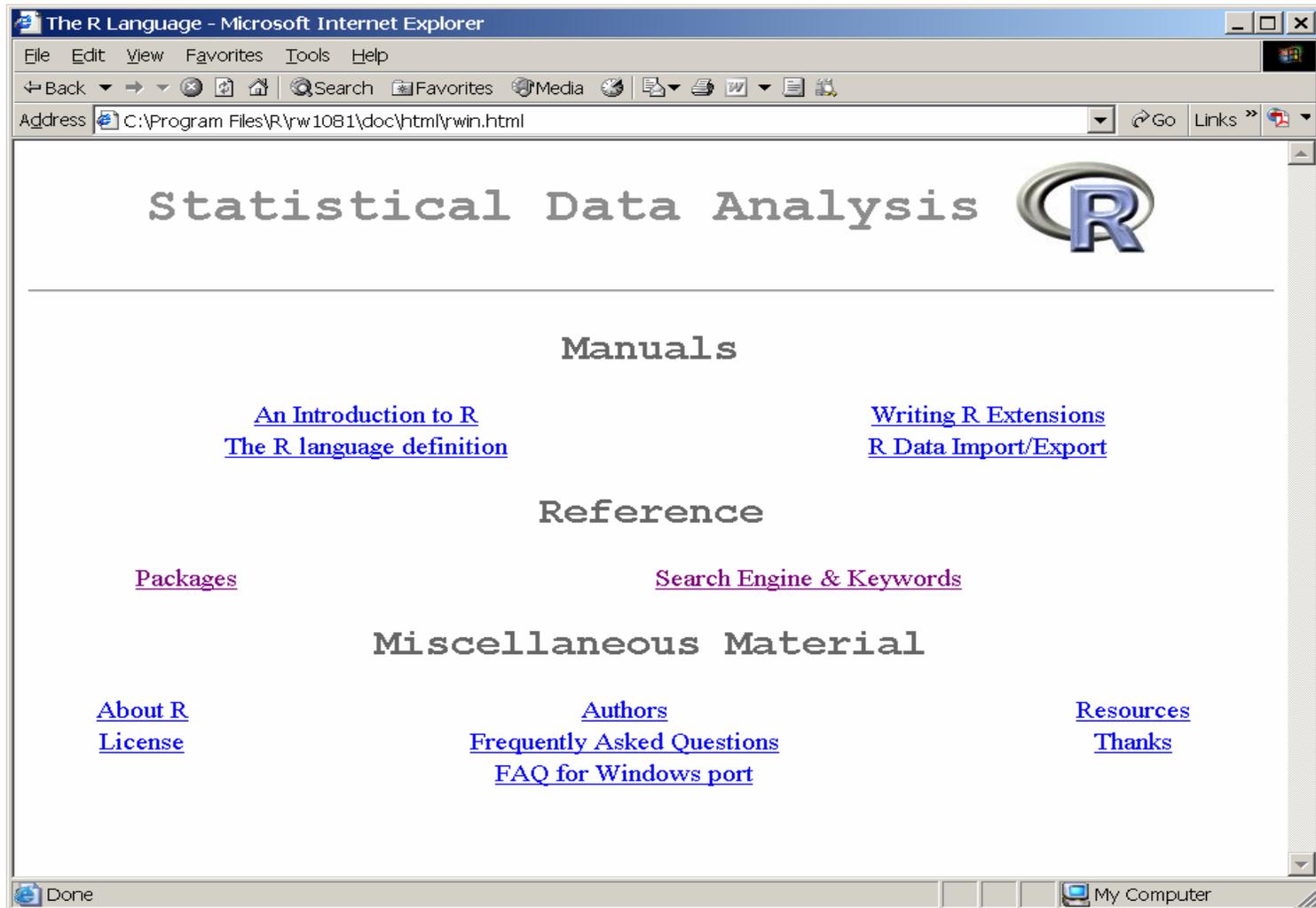
- Save the downloaded file to the desktop
- Uninstall any previous versions of R, if any
  - Not absolutely necessary
- Close all other programs
- Install the downloaded file by double clicking it
  - Standard Windows installer
  - Puts R in your Start Menu
  - Places an R icon on desktop
  - May take 5-10 minutes



- Start R: Double click R desktop icon



# *Accessing Help on commands: base and packages*



R: Package Index - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites Media

Address C:\R\rw2010\doc\html\packages.html Go Links

# Package Index

---

|                                  |  |
|----------------------------------|--|
| <a href="#">base</a>             | The R Base Package   |
| <a href="#">boot</a>             | Bootstrap R (S-Plus) Functions (Canty)                               |
| <a href="#">class</a>            | Functions for Classification   |
| <a href="#">cluster</a>          | Functions for clustering (by Rousseeuw et al.)                       |
| <a href="#">datasets</a>         | The R Datasets Package   |
| <a href="#">foreign</a>          | Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, dBase, ... |
| <a href="#">graphics</a>         | The R Graphics Package   |
| <a href="#">grDevices</a>        | The R Graphics Devices and Support for Colours and Fonts             |
| <a href="#">grid</a>             | The Grid Graphics Package  |
| <a href="#">KernSmooth</a>       | Functions for kernel smoothing for Wand & Jones (1995)               |
| <a href="#">lattice</a>          | Lattice Graphics   |
| <a href="#">MASS</a>             | Main Package of Venables and Ripley's MASS                           |
| <a href="#">methods</a>          | Formal Methods and Classes   |
| <a href="#">mgcv</a>             | GAMs with GCV smoothness estimation and GAMMs by REML/PQL            |
| <a href="#">nlme</a>             | Linear and nonlinear mixed effects models                            |
| <a href="#">nnet</a>             | Feed-forward Neural Networks and Multinomial Log-Linear Models       |
| <a href="#">psurvey.analysis</a> | Probability Survey Analytical Functions                              |
| <a href="#">psurvey.design</a>   | Probability Survey Design Functions                                  |
| <a href="#">RColorBrewer</a>     | ColorBrewer palettes   |
| <a href="#">rpart</a>            | Recursive Partitioning   |
| <a href="#">spatial</a>          | Functions for Kriging and Point Pattern Analysis                     |

Done My Computer

*List of packages  
Key ones:*

*base*

*psurvey*



# Spatial Survey Design and Analysis



Documentation for package 'spsurvey' version 1.0

## Help Pages

|                               |   |
|-------------------------------|---|
| <a href="#">adjwgt</a>        | Adjust Initial Survey Design Weights                            |
| <a href="#">albersgeod</a>    | Project Albers Projection in Plane to Spheroid                  |
| <a href="#">cat.analysis</a>  | Categorical Data Analysis for Probability Survey Data           |
| <a href="#">category.est</a>  | Category Proportion and Size Estimates                          |
| <a href="#">cdf.decon</a>     | Cumulative Distribution Function - Deconvolution                |
| <a href="#">cdf.est</a>       | Cumulative Distribution Function - Estimation                   |
| <a href="#">cdf.test</a>      | Cumulative Distribution Function - Inference                    |
| <a href="#">cont.analysis</a> | Continuous Data Analysis for Probability Survey Data            |
| <a href="#">dsgnsum</a>       | Summarize the Sites Selected for a Survey Design                |
| <a href="#">framesum</a>      | Summarize Frame Size for a Survey Design                        |
| <a href="#">grts</a>          | Generalized Random-Tessellation Stratified (GRTS) Survey Design |
| <a href="#">grtsarea</a>      | Select GRTS Sample of an Area Resource                          |
| <a href="#">grtslin</a>       | Select GRTS Sample of a Linear Resource                         |
| <a href="#">grtspts</a>       | Select GRTS Sample of a Finite Resource                         |

adjwgt {spsurvey}

R Documentation

# Adjust Initial Survey Design Weights

## Description

This function adjusts initial survey design weights when implementation results in use of oversample sites or when it is desired to have final weights sum to a known frame size. Adjusted weights are equal to initial weight times the frame size divided by the sum of the initial weights. The adjustment is done separately for each category specified in argument `wcat`.

## Usage

```
adjwgt(sites, wgt, wcat, framesize)
```

## Arguments

- `sites` the logical value for each site, where TRUE = include the site and FALSE = do not include the site.
- `wgt` the initial weight (inverse of the sample inclusion probability) for each site.



# Categorical Data Analysis for Probability Survey Data

## Description

This function organizes input and output for analysis of categorical data generated by a probability survey. Input can be either an object belonging to class `psurvey.analysis` (see the documentation for function `psurvey.analysis`) or through use of the other arguments to this function.

## Usage

```
cat.analysis(sites=NULL, subpop=NULL, design=NULL,  
  data.cat=NULL, N.cluster=NULL, popsize=NULL,  
  stagelsize=NULL, support=NULL, swgt=NULL, swgt1=NULL,  
  unitsize=NULL, vartype="Local", conf=95, psurvey.obj=NULL)
```

## Arguments

- |                     |  |
|---------------------|--|
| <code>sites</code>  | a data frame consisting of two variables: the first variable is site IDs, and the second variable is a logical vector indicating which sites to use in the analysis. If <code>psurvey.obj</code> is not provided, then this argument is required. The default is <code>NULL</code> .   |
| <code>subpop</code> | a data frame describing sets of populations and subpopulations for which estimates will be calculated. The first variable is site IDs. Each subsequent variable identifies a Type of population, where the variable name is used to identify Type. A Type variable identifies each site with one of the subpopulations of that Type. If <code>psurvey.obj</code> is not provided, then this argument is required. The default is <code>NULL</code> . |
| <code>design</code> | a data frame consisting of design variables. If <code>psurvey.obj</code> is not provided, then this argument is required.  |



# *Installing Survey Libraries*

- Download library from ARM web site
  - <http://www.epa.gov/nheerl/arm/>
  - Click on “Download Software”
  - Put zip file on desktop or in folder
- Install library in R using menu





## U.S. Environmental Protection Agency

# Aquatic Resources Monitoring

[Recent Additions](#) | [Contact Us](#) | [Print Version](#) | [Search](#)

[EPA Home](#) > [Office of Research Development \(ORD\)](#) > [National Health Environmental Effects Research Laboratory \(NHEERL\)](#) > [Western Ecology Division \(WED\)](#) > ARM

## Aquatic Resources Monitoring Web Site

Hosted by the  
Monitoring Design and Analysis Team  
USEPA ORD  
National Health and Environmental Effects  
Research Laboratory,  
Western Ecology Division,  
Corvallis, OR



Environmental Monitoring and Assessment Program

This Web site provides information on monitoring of aquatic resources in the US, primarily focused on design and analysis of probability based surveys. Links are provided to other aquatic resources monitoring information available on the internet.



- Use latest version
- Key Files: Users Guide, Changes, zip file for library

**U.S. Environmental Protection Agency**

## Aquatic Resources Monitoring - Software

[Recent Additions](#) | [Contact Us](#) | [Print Version](#) | [Search](#)

[EPA Home](#) > [ORD](#) > [NHEERL](#) > [WED](#) > [ARM](#) > Software

### Software Downloads for Design and Analysis Packages

This information supports GRTS designs, computational options, and analysis software. The software packages are written in S-Plus and R languages (R is a free statistical package). [See Version Notice.](#)

|   |   |
|---|---|
| <a href="#">Users Guide</a> for R psurvey.analysis, version 2.6 Probability Survey Data Analysis Functions (3/31/05)      | Download <a href="#">R Guide 2.6</a> (pdf 16KB)     |
| <a href="#">Users Guide</a> for S-PLUS psurvey.analysis, version 2.6 Probability Survey Data Analysis Functions (3/31/05) | Download <a href="#">S_PLUS Guide 2.6</a> (pdf 7KB) |
| <a href="#">Changes2.6</a> , psurvey.analysis Version 2.6 (3/31/05)   | Download <a href="#">changes2.6</a> (pdf 11 KB)     |

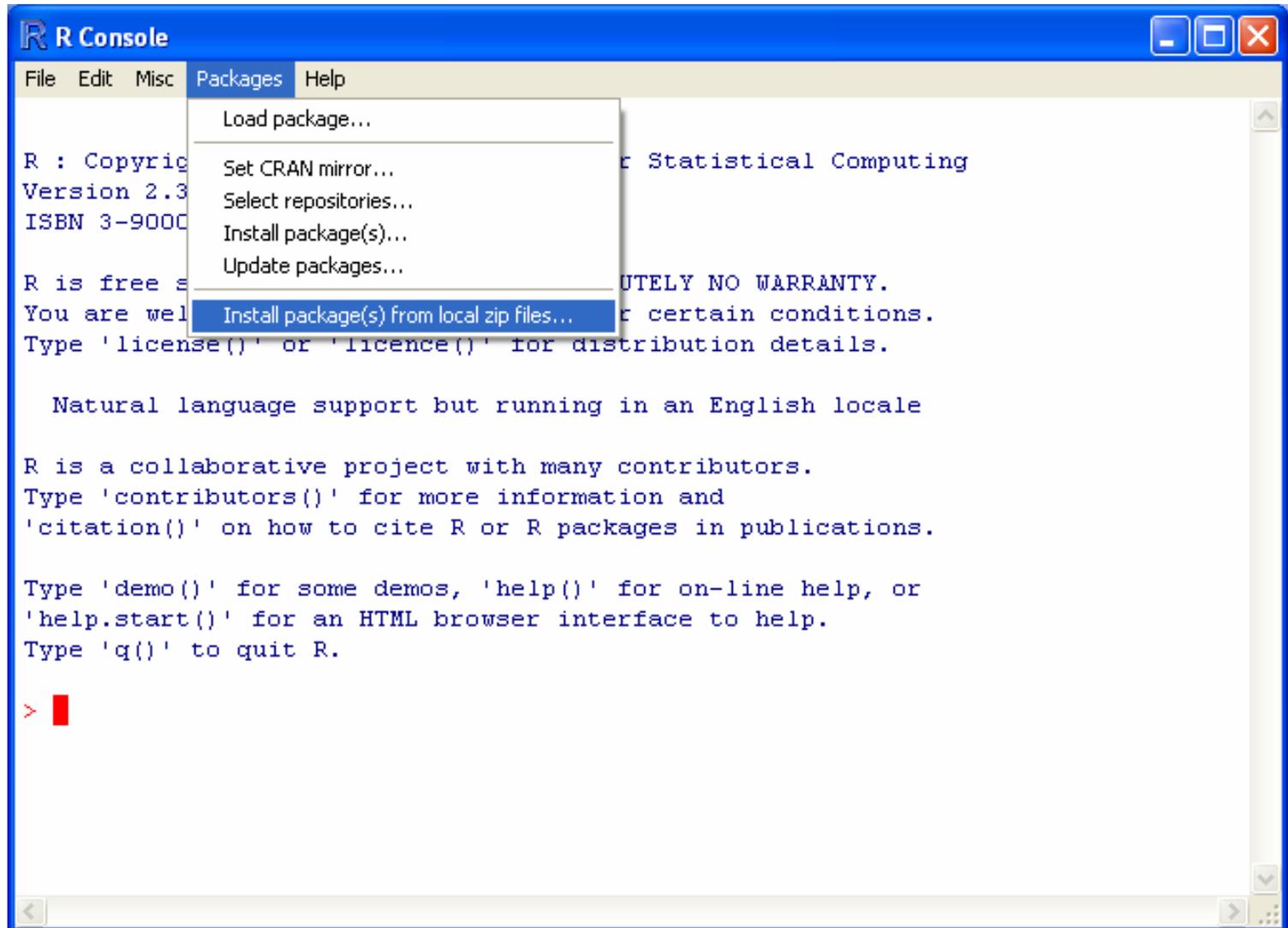
  

|   |
|---|
| R libraries for probability survey population estimation, Version 2.6 (3/31/05)<br><a href="#">psurvey.analysis_2.6.zip</a> (zip 194 KB)  |
| S-Plus libraries for probability survey population estimation, version 2.5 (12/31/04)<br>S-PLUS 6.0: <a href="#">psurvey.analysis_2.6.S-PLUS 6.0.zip</a> (zip 755KB)<br>S-PLUS 6.1: <a href="#">psurvey.analysis_2.6.S-PLUS 6.1.zip</a> (zip 755KB) |

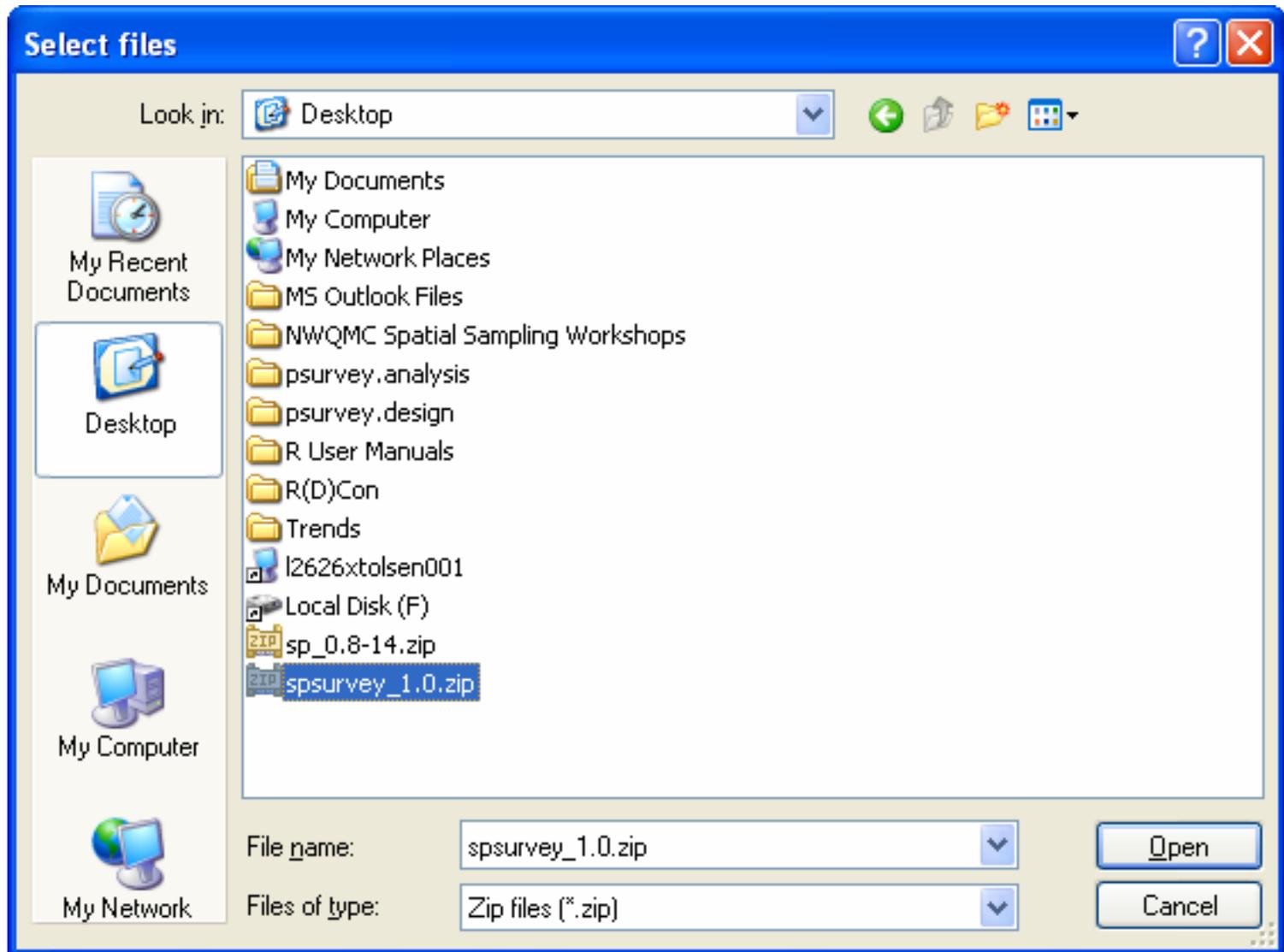
Local intranet



# *Install from local zip file*



Find the saved zip file and click open: That installs the library!



# *Begin Survey Design or Analysis for a Project using R*

- Create a new folder for the survey design or survey analysis task
- Recommend create a sub folder named “Original Data”
  - Place any original data files in this folder
  - Never change these files
- Double click R desktop icon or quick launch icon to start R
- Under R File menu,
  - go to Change Dir
  - Browse to find your project directory
- When done using R, exit and respond YES when asked if want to save workspace.
- When want to start R again to work on this project, go to project folder and open “.RData” file. This will automatically set R to use your project directory



# *Load spsurvey every time start R if doing survey design or survey analysis*

```
R Console
File Edit Misc Packages Help
Load package...
Set CRAN mirror...
Select repositories...
Install package(s)...
Update packages...
Install package(s) from local zip files...

R is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; there is NO WARRANTY.
You are welcome to share it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

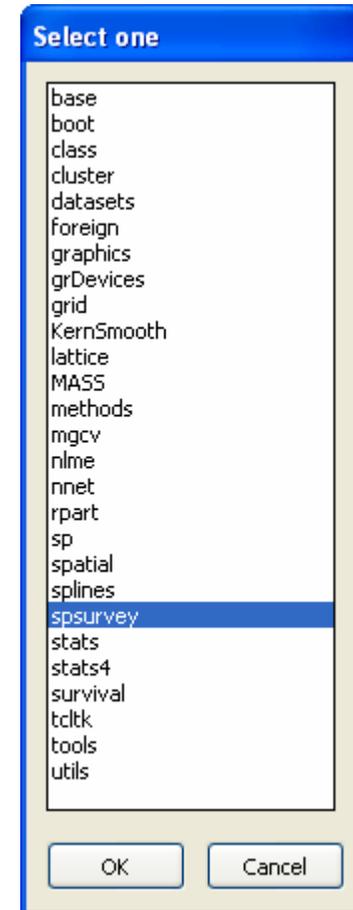
Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

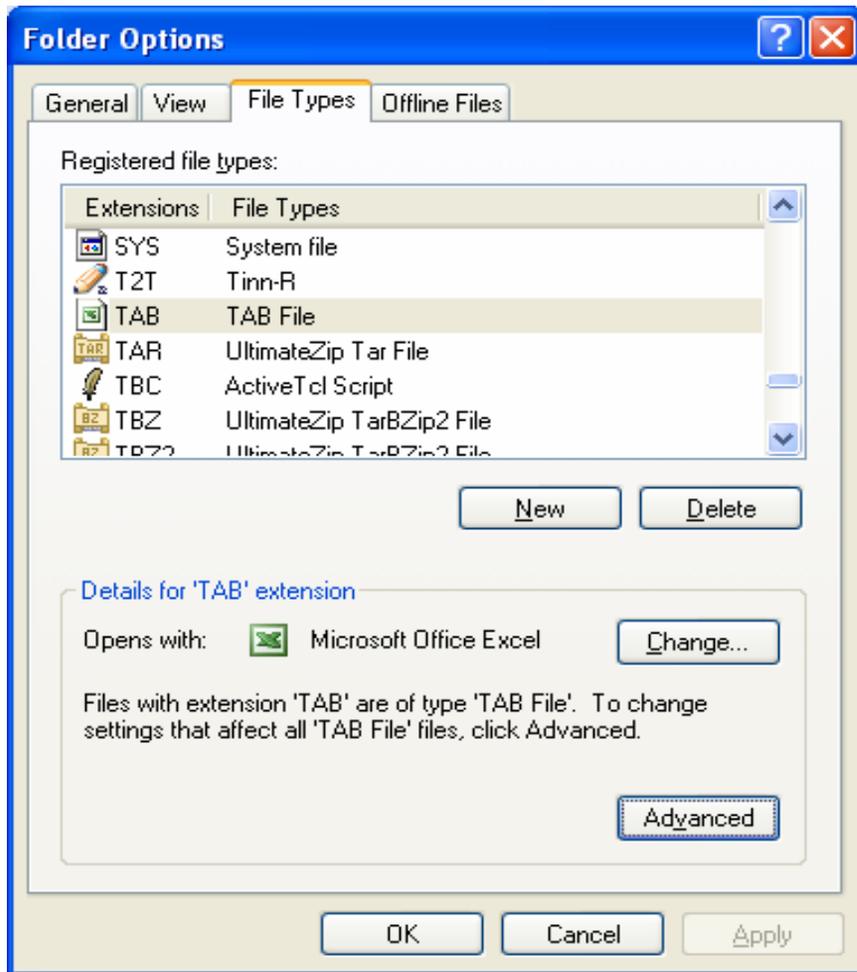
[Previously saved workspace restored]

> local({pkg <- select.list(sort(.packages(all.available = TRUE)))})
+ if(nchar(pkg)) library(pkg, character.only=TRUE)})
> local({pkg <- select.list(sort(.packages(all.available = TRUE)))})
+ if(nchar(pkg)) library(pkg, character.only=TRUE)})
> utils::menuInstallLocal()
Error in install.packages(choose.files("", filters = Filters[c("zip", :
no packages were specified
>
```



# Windows Topics

- File Associations: Control Panel

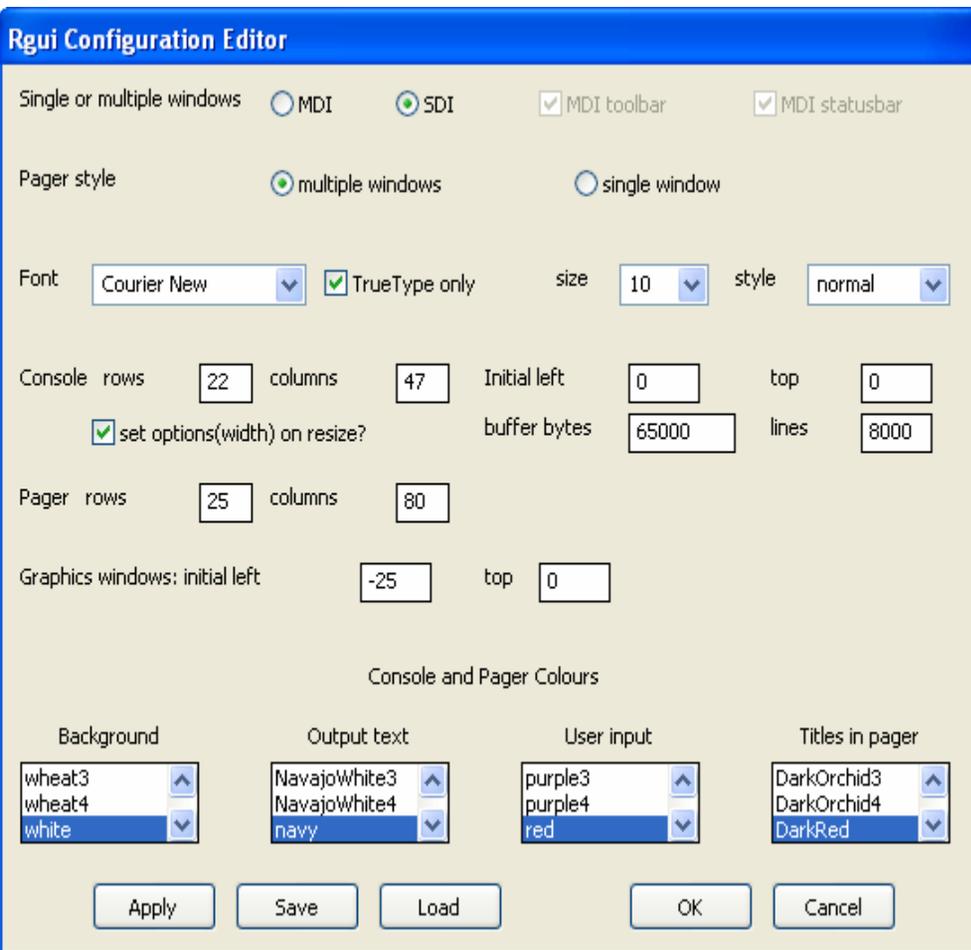


# *Executing Commands in R*

- Three options to execute commands in R
  - Type commands in console window
    - window automatically opens when start R
    - Commands execute when hit “Enter” key
  - Use script window in R
    - Under File menu either create New or Open existing script file
    - Type commands in script window
    - To execute commands, copy and paste into console window
    - Close window and save to project directory
    - Use saved script file when want to continue work on same project
  - Use another text editor program
    - Type commands in editor
    - Copy and Paste into R console window
- Advantages of R script window or text editor approach
  - Can save all your commands in both cases
  - Some text editors recognize R language and structure making it easy to locate errors (e.g. Tinn-R)
  - Redo the analysis if change data with little extra work
  - Can set up an example analysis and use as template for other analyses
  - Aid in QA for analysis process



# Integrating R and Tinn-R



- Install R
- Start R
- Under R Edit Menu select GUI preferences
- Change to SDI from MDI
- Save change in etc folder: located under Program Files; R folder; R 2.30 folder
- Quit R
- Install Tinn-R
- Start Tinn-R



# *Things to be aware of*

- Column names may be modified in R
  - “blanks” will become “.”
  - Unusual characters will become “.”
  - Unusual character at beginning (e.g. %), will change to “X.”
- Blanks in spreadsheet will be turned into missing data
- “NA” in spreadsheet will be interpreted as missing data, except when column is character
- Everything after “#” is ignored. Can be overridden.
- Reading a “csv” file may result in unanticipated results if columns contain “,” in character fields (use “tab” delimited instead)

