

An Introduction to the 'R Commander': A Basic-Statistics Graphical User Interface (GUI) for R

Theresa A Scott, MS

Department of Biostatistics
Vanderbilt School of Medicine
theresa.scott@vanderbilt.edu
<http://biostat.mc.vanderbilt.edu/TheresaScott>

Introduction

▷ What is R?

- *Free* interactive programming language and environment.
- Mainly used as a statistical analysis package, but is no way limited to just statistics.
 - Contains facilities for data manipulation, simulation, calculation, and graphical display in general.
- Able to run on Windows, MacOS, and Linux/UNIX platforms.
- **IMPORTANT:** Command- (not menu-) driven.
 - Major hurdle for many new R users.

▷ What is the 'R Commander' ?

- Basic-statistics graphical user interface (GUI) for R.
 - An interface based on menus, buttons, dialog boxes, and information fields instead of (text) commands.

▷ **Purpose of this lecture:** to introduce, describe, and demonstrate the basic use of the R Commander GUI.

▷ **Outline:**

- Motivation for using a GUI
- Design objectives of the R Commander
- Using the R Commander
 - Starting the R Commander
 - Menus, buttons, and sub-windows
 - Data input
 - Making changes to the data set
 - Calculating summary statistics
 - Graphs
 - Odds & ends
- References and resources

Motivation for using a GUI

▷ Why a GUI works for introductory or infrequent use of a software program, like R:

- Does not require you to remember the names and arguments of commands; decreases the chances of syntax/typing errors.

▷ Potential drawback to a GUI:

- Having to drill one's way through successive layers of menus and dialog boxes can be tedious and can make it difficult to reproduce a statistical analysis, perhaps with variations.

▷ Intention is to keep things simple and basic.

- Even though possible, not feasible to provide access to every option and procedure.
 - R based on tens of thousands of *functions* (ie, commands) that are grouped into hundreds of *packages*.

Design objectives of the R Commander

- 1 To provide, through an easy-to-use, cross-platform, extensible GUI, the statistical functionality required for a basic-statistics course.
 - Original target text was David Moore's *The Basic Practice of Statistics, Second Edition* (Freeman, 2000).
 - Current functionality has grown to include support for linear and generalized-linear models, and other more advanced features.
- 2 To make it relatively difficult to do unreasonable things.
 - eg, calculate the mean of a categorical variable.
- 3 To render visible the relationship between choices made in the GUI and the R commands that they generate.
 - Hope is to wean users from the GUI to writing commands.

Starting the R Commander

▷ **NOTE:** This lecture assumes the latest versions of *both* R and the R Commander are installed on your computer – detailed instructions are available at my website.

▷ First, start R.

- If you're using the Windows or Mac versions of R, double-click the R icon on the desktop, or find the R program under the start menu. This will start R in a new *console window* with a *command line subwindow*.
- If you're using the Linux/Unix version of R, launch R by typing 'R' at the shell command prompt (i.e., '\$ R' if we assume that the shell prompt is '\$') of a terminal window. This will cause R to start up as an interactive program in the current terminal window.

Starting the R Commander, *cont'd*

- ▷ Then, start the R Commander within R.
 - Type the following command at the *command line prompt*, which is denoted by the greater than symbol, >, and then press the ENTER key. **NOTE:** R is *case-sensitive*.
`library(Rcmdr)`
 - Recall (from the installation instructions), the R Commander is available via the Rcmdr 'contributed' *package*.
 - Packages are the way R organizes its functionality.
 - The `library` command is the general way we *load* any installed package into memory for its use.
- ▷ If successful, the GUI should be opened in a *new* window (separate from R; floating freely on the desktop) containing several menus, buttons, and sub-windows.

Menus in the R Commander window

- **File** – items for loading and saving script files; for saving output and the R *workspace*; and for exiting.
- **Edit** – items (ie, **Cut**, **Copy**, **Paste**, etc) for editing the contents of the script and output sub-windows.
- **Data** – submenus for reading and manipulating data.
- **Statistics** – submenus for a variety of basic statistical analyses.
- **Graphs** – items for creating simple statistical graphs.
- **Models** – items and submenus for obtaining numerical summaries, confidence intervals, hypothesis tests, diagnostics, and graphs for a statistical model, and for adding diagnostic quantities (eg, residuals) to the data set.
- **Distributions** – submenus for probabilities, quantiles, and graphs of standard statistical distributions.

Menus in the R Commander window, *cont'd*

- **Tools** – items for loading R packages unrelated to the Rcmdr package (eg, to access data saved in another package), and for setting some options.
- **Help** – items to obtain information on the R Commander, including an introductory manual.

▷ Also...

- Most menu items lead to dialog boxes, which contain Help buttons.
- Menu items are inactive (ie, 'grayed out') if not applicable to the current context.

▷ The complete menu 'tree' for the R Commander (version 1.4-10) is given at the end of these lecture notes.

Buttons of the R Commander window

- ▷ Left-most (flat) button shows the name of the *active* data set.
- Most menus & dialog boxes reference the active data set.
 - Initially, there is no active data set.
 - When a new data set is read, it becomes the active data set.
 - Several data sets may reside in memory, but only one is 'active' at any given time.
 - Can select the desired active data set by pressing this button.
- ▷ The '**Edit data set**' and '**View data set**' buttons allow you to open the R data editor to modify and view the active data set, respectively.
- ▷ Right-most (flat) button shows the name of the active statistical model – similar in functionality to active data set button.

Sub-windows of the R Commander window

▷ **Script Window**

- Displays commands generated by the menu selection and dialog boxes.
- Also provides the ability to edit, enter (ie, type new), and re-execute commands.
 - Pressing the **Submit** button (or **Ctrl-r** for 'run') causes the line containing the cursor to be submitted (or resubmitted) for execution.
 - If several lines are selected (eg, by left-clicking and dragging over them), then pressing the **Submit** button will cause all of them to be executed.
 - Typed commands can extend over more than one line – if they do, lines after the first must be indented with one or spaces or tabs.

Sub-windows of the R Commander window, *cont'd*

▷ **Output Window**

- Displays printed output (in dark blue).
- Includes 'echoing' of the commands displayed in the Script Window (in red).

▷ **Messages Window**

- Displays error messages (in red), warnings (in green), and other information (in dark blue).
- Errors and warning also provide an audible cue by ringing a bell.

▷ **NOTE:** any created graphs will appear in a separate **Graphics Device** window, outside of the main R Commander window.

Data input

- ▷ R Commander provides several ways to get data into R using the **Data** menu:
 - Can enter the data directly – reasonable for a very small data set.
 - Can read a data set that is included in an R package.
 - Can import data from a plain-text (ASCII) file or from another statistical package (Minitab, SPSS, or STATA) – most likely choice.
- ▷ Example: Read-in a comma delimited data file.
 - Select **Data > Import data > from text file...**
 - In the *Read Data From Text File* dialog box, specify characteristics of the data file (eg, field separator).
 - In the *Open* dialog box, navigate to and choose the appropriate file.

Data input, *cont'd*

- ▷ The *Read Data From Text File* dialog box in more detail:
 - Specify a *name* for the data set.
 - Valid names begin with an upper- or lower-case letter and consist entirely of letters, periods, underscores (_), and digits (0-9) – *do not* include any embedded spaces.
 - REMEMBER, R is *case sensitive* (ie, mydata and MyData).
 - Specify whether the first line of the file contain variable names.
 - Specify how missing data are represented in the file.
 - Specify the field separator (eg, commas for a .csv file).
- ▷ Microsoft Excel files can be read into R, but often have issues.
 - Instead, use **Save As** from the **File** drop-menu to save the data set as either a tab-delimited (.txt) or comma-delimited (.csv) file (specified via the **Save as type** drop-box).

- ▷ Double-check that your data set was read-in correctly:
 - The *Messages* sub-window will report any errors.
 - The *Messages* sub-window will also report the number of rows and columns of the data set read-in.
 - Clicking the *View data set* button will bring up the data viewer window.
 - Clicking the *Edit data set* button will bring up the data editor window.
 - Changes are saved upon closing the editor.
 - Good idea to keep a (written) log of all the changes made to the data set.
 - Print the names of the variables of the read-in data set by choosing **Data > Active data set > Variables in the active data set**.

Modifying the active data set

- ▷ **Factors:** how R defines *categorical variables*.
 - Raw values can be character, such as 'No' and 'Yes', or numeric, such as 0, 1, and 2.
 - Makes it possible to assign meaningful names to the categories – for example, assign 'Black' to 'B' and 'White' to 'W' of a race variable or 'Censored' to 0, 'Censored due to liver treatment' to 1, and 'Death' to 2 for the a status variable.
 - Also essential for R to be able to distinguish between categorical variables and variables whose values have a direct numerical meaning.
 - Example: the raw numeric values of a categorical variables (eg, 0, 1, and 2) would be interpreted as a continuous variable in a regression model, not a categorical one.

Modifying the active data set, *cont'd*

▷ Factors, *cont'd*

- Non-logical (case sensitive TRUE and FALSE) character columns (whether or not enclosed in quotation marks) are automatically converted to factors when read into R.
- However, must explicitly define numeric variables as factors (if appropriate) using **Data > Manage variables in active data set > Convert numeric variables to factors...**
 - Specify the (character) *label* for each *level* (ie, numeric value).
- Once defined, can 're-order the levels' (ie, modify which group is the *reference* group) using **Data > Manage variables in active data set > Reorder factors levels...**
- NOTE: Several menu items will be 'grayed out' until variables are appropriately defined as factors (also true for specific dialog box choices/specifications).

Modifying the active data set, *cont'd*

▷ Subsetting the active data set

- Use **Data > Active data set > Subset active data set.**
- Can subset the columns *AND/OR* rows.
 - Hold the **Ctrl** key to select more than one variable (ie, column).
 - Subset the rows by specifying a *subset expression*.
 - Reference the *case-sensitive* variable names.
 - == used to express equality; != used to express inequality.
 - & (AND) and | (OR) used to combine multiple subset expressions (ie, logical operators).
 - Use `is.na(varname)` and `!is.na(varname)` to include/exclude, respectively, the missing values of a variable `varname` – NOTE, `varname == NA` will not work.
 - Example: `gender == "Female" & age <= 25`
- Useful to define the subset as a separate data set (ie, give it a different name).

Calculating summary statistics

- ▷ Summarizing every variable in the active data set:
 - Select **Statistics > Summaries > Active data set**.
 - For each numerical variable, reports min & max, 1st & 3rd quartiles, median & mean, and number of missing values.
 - For each categorical variable (defined as a factor), reports number of observations at each level of the factor.
- ▷ Summarizing individual numerical variables.
 - Select **Statistics > Summaries > Numerical summaries**.
 - Only numerical variables are shown to choose from.
 - By default, calculates mean & standard deviation, and quantiles (min, 1st quartile, median, 3rd quartile, & max).
 - Also a *Summarize by groups...* button – allows you to compute summaries within groups defined by the levels of a factor.

Calculating summary statistics, *cont'd*

- ▷ Summarizing individual categorical variables:
 - Select **Statistics > Summaries > Frequency distributions....**
 - Only defined factors are shown to choose from.
 - Reports raw and relative frequencies (ie, proportions) of each level of the factor.
- ▷ Cross-classifying multiple categorical variables.
 - Select **Statistics > Contingency tables > Two-way tables... or Multi-way tables**.
 - Only defined factors are shown to choose from.
 - Have choice of percentages to compute and hypothesis tests to perform.
- ▷ Calculate the number of missing values for each variable.
 - Select **Statistics > Summaries > Count missing observations**.

Graphs

- ▷ Able to generate variety of basic statistical graphs, including
 - Histograms,
 - Single & side-by-side boxplots,
 - Scatterplots & scatterplot matrices,
 - Barplots,
 - Plot of means across groups, and
 - 'XY conditioning plots' (ie, a scatterplot which uses color to distinguish groups defined by a factor variable).
- ▷ Saving a graph to a file.
 - Select **Graphs > Save graph to file > as bitmap... OR as PDF/Postscript/EPS....**
 - Can also right-click on the graph, select 'Copy as meta-file' and past directly into MS Word or PowerPoint.

Graphs, *cont'd*

- ▷ Unfortunately, limited by choices offered in the menu.
 - Instance where impossible to incorporate all possible procedures and options.
 - Using the commands directly, have control over every aspect of a graph (eg, text, margins, color, etc).
 - Can easily modify the appearance of 'pre-defined' graphs, such as a barplot.
 - Can also create 'graphs from scratch'.
 - In turn, choice of possible graphs to create is nearly limitless – much more functionality than other software programs that create graphs (eg, MS Excel, SPSS, STATA, Prism).
 - If interested in learning more, extensive coverage of graphics in R in my 'Introduction to the Fundamentals & Functionality of the R Programming Language' lecture notes – see 'Resources & References'.

Odds & Ends

- ▷ Saving syntax: (via the *File* menu)
 - Useful to edit this file (ie, in WordPad), removing anything unnecessary or incorrect.
 - During the next R session, can 'get back to where you left off' by opening a saved script (also via the *File* menu) and submitting the syntax.
- ▷ Saving output: (via the *File* menu)
 - If opened in MS Word, use a mono-spaced ('typewriter') font, such as *Courier New*; otherwise output will not line up neatly.
- ▷ When exiting from R, will be asked whether to save the 'workspace image' (ie, the data R keeps in memory).
 - Best to answer 'No' – can be easily confused by objects carried over from one session to another; can replicate anything previously created as long as script is saved.

References and Resources

- ▷ References for the R Commander:
 - John Fox's article in the December 2007 issue of the R News.
 - Also discusses how to extend the functionality of the R Commander by using plug-in packages.
- ▷ Resources for learning R in general:
 - My website
 - 'An Introduction to the Fundamentals & Functionality of the R Programming Language' (two PDF files that amount to a 2-day short-course).
 - The *R Clinic*
 - Every Thursday from 2:00 - 3:00 PM in MCN D-2221.

Current menu “tree” of the R Commander (version 1.4-10)

<p>File</p> <ul style="list-style-type: none"> Change working directory... Open script file... Save script... Save script as... Save output... Save output as... Save R workspace... Save R workspace as... Exit <ul style="list-style-type: none"> From Commander From Commander and R <p>Edit</p> <ul style="list-style-type: none"> Cut Copy Paste Delete Find... Select all Undo Redo Clear Window <p>Data</p> <ul style="list-style-type: none"> New data set... Load data set... Import data <ul style="list-style-type: none"> from text file, clipboard, or URL... from SPSS data set... from Minitab data set... from STATA data set... from Excel, Access, or dbase data set... Data in packages <ul style="list-style-type: none"> List data sets in packages Read data set from an attached package... Active data set <ul style="list-style-type: none"> Select active data set... Refresh active data set Help on active data set (if applicable) Variables in active data set Set case names... Subset active data set Remove row(s) from active data set... Stack variables in active data set... Remove cases w/ missing data... Save active data set... Export active data set... Manage variables in active data set <ul style="list-style-type: none"> Recode variables... Compute new variable... Add observation numbers to data set Standardize variables... Convert numeric variables to factors... Bin numeric variable... Reorder factor levels... Define contrasts for a factor... Rename variables... Delete variables from data set... 	<p>Statistics</p> <ul style="list-style-type: none"> Summaries <ul style="list-style-type: none"> Active data set Numerical summaries... Frequency distributions... Count missing observations Table of statistics Correlation matrix... Correlation test... Shapiro-Wilk test of normality... Contingency tables <ul style="list-style-type: none"> Two-way table... Multi-way table... Enter and analyze two-way table... Means <ul style="list-style-type: none"> Single-sample t-test... Independent samples t-test... Paired t-test... One-way ANOVA... Multi-way ANOVA... Proportions <ul style="list-style-type: none"> Single-sample proportion test... Two-sample proportions test... Variances <ul style="list-style-type: none"> Two-variances F-test... Bartlett’s test... Levene’s test... Nonparametric tests <ul style="list-style-type: none"> Two-sample Wilcoxon test... Paired-samples Wilcoxon test... Kruskal-Wallis test... Friedman rank-sum test... Dimensional analysis <ul style="list-style-type: none"> Scale reliability... Principal-components analysis... Factor analysis... Cluster analysis <ul style="list-style-type: none"> k-means cluster analysis... Hierarchical cluster analysis... Summarize hierarchical clustering... Add hierarchical clustering to data set... Fit models <ul style="list-style-type: none"> Linear regression... Linear model... Generalized linear model... Multinomial logit model... Ordinal regression model... <p>Graphs</p> <ul style="list-style-type: none"> Color palette... Index plot... Histogram... Stem-and-leaf display... Boxplot... Quantile-comparison plot... Scatterplot... Scatterplot matrix... Line graph...
---	---

Current menu “tree” of the R Commander (version 1.4-10)

<ul style="list-style-type: none"> XY conditioning plot... Plot of means... Bar graph... Pie chart... 3D graph <ul style="list-style-type: none"> 3D scatterplot... Identify observations with mouse Save graph to file Save graph to file <ul style="list-style-type: none"> as bitmap... as PDF/Postscript/EPS... 3D RGL graph... Models <ul style="list-style-type: none"> Select active model Summarize model Add observation statistics to data Confidence intervals Akaike Information Criterion (AIC) Bayesian Information Criterion (BIC) Hypothesis tests <ul style="list-style-type: none"> ANOVA table... Compare two models... Linear hypothesis... Numerical diagnostics <ul style="list-style-type: none"> Variance-inflation factors Breusch-Pagan test for heteroscedasticity Durbin-Watson test for autocorrelation RESET test for nonlinearity Bonferroni outlier test Graphs <ul style="list-style-type: none"> Basic diagnostic plots Residual quantile-comparison plot Component+residual plots Added-variable plots Influence plot Effect plots Distributions <ul style="list-style-type: none"> Continuous distributions <ul style="list-style-type: none"> Normal distribution <ul style="list-style-type: none"> Normal quantiles... Normal probabilities... Plot Normal distribution... Sample from Normal distribution... t distribution <ul style="list-style-type: none"> t quantiles... t probabilities... Plot t distribution... Sample from t distribution... Chi-squared distribution <ul style="list-style-type: none"> Chi-squared quantiles... Chi-squared probabilities... Plot Chi-squared distribution... Sample from Chi-squared distribution... F distribution <ul style="list-style-type: none"> F quantiles... 	<ul style="list-style-type: none"> F probabilities... Plot F distribution... Sample from F distribution... Exponential distribution <ul style="list-style-type: none"> Exponential quantiles... Exponential probabilities... Plot Exponential distribution... Sample from Exponential distribution... Uniform distribution <ul style="list-style-type: none"> Uniform quantiles... Uniform probabilities... Plot Uniform distribution... Sample from Uniform distribution... Beta distribution <ul style="list-style-type: none"> Beta quantiles... Beta probabilities... Plot Beta distribution... Sample from Beta distribution... Cauchy distribution <ul style="list-style-type: none"> Cauchy quantiles... Cauchy probabilities... Plot Cauchy distribution... Sample from Cauchy distribution... Logistic distribution <ul style="list-style-type: none"> Logistic quantiles... Logistic probabilities... Plot Logistic distribution... Sample from Logistic distribution... Lognormal distribution <ul style="list-style-type: none"> Lognormal quantiles... Lognormal probabilities... Plot Lognormal distribution... Sample from Lognormal distribution... Gamma distribution <ul style="list-style-type: none"> Gamma quantiles... Gamma probabilities... Plot Gamma distribution... Sample from Gamma distribution... Weibull distribution <ul style="list-style-type: none"> Weibull quantiles... Weibull probabilities... Plot Weibull distribution... Sample from Weibull distribution... Gumbel distribution <ul style="list-style-type: none"> Gumbel quantiles... Gumbel probabilities... Plot Gumbel distribution... Sample from Gumbel distribution... Discrete distributions <ul style="list-style-type: none"> Binomial distribution <ul style="list-style-type: none"> Binomial quantiles... Binomial tail probabilities...
--	--

Current menu “tree” of the R Commander (version 1.4-10)

- Binomial probabilities...
- Plot Binomial distribution...
- Sample from Binomial distribution...
- Poisson distribution
 - Poisson quantiles...
 - Poisson tail probabilities...
 - Poisson probabilities...
 - Plot Poisson distribution...
 - Sample from Poisson distribution...
- Geometric distribution
 - Geometric quantiles...
 - Geometric tail probabilities...
 - Geometric probabilities...
 - Plot Geometric distribution...
 - Sample from Geometric distribution...
- Hypergeometric distribution
 - Hypergeometric quantiles...
 - Hypergeometric tail probabilities...
 - Hypergeometric probabilities...
 - Plot Hypergeometric distribution...
 - Sample from Hypergeometric distribution...
- Negative binomial distribution
 - Negative binomial quantiles...
 - Negative binomial tail probabilities...
 - Negative binomial probabilities...
 - Plot Negative binomial distribution...
 - Sample from Negative binomial distribution...

Tools

- Load package(s)...
- Load Rcmdr plug-in(s)...
- Options...

Help

- Commander help
- Introduction to the R Commander
- Help on active data set (if applicable)
- About Rcmdr