Statistical computing with R: selecting the right tool for the job—R Commander or something else?

Jason Wilson*

R Commander (Rcmdr) is one of the most popular graphical user interfaces (GUIs) for R. Why is R Commander so popular? What does it do? When should R Commander be used? When should it not be used? What are the alternatives to R Commander? What can be said about using R Commander, or an alternative, to teach introductory statistics courses? The answers to these questions form the sections of the paper. Included is a Box describing many of the alternative R GUIs (RExcel, Deducer, Rattle, and RKWard) and interactive development environments (IDEs—RStudio, REvolution, and RKWard). Advice is given to help the user 'select the right tool for the job'. © 2012 Wiley Periodicals, Inc.

INTRODUCTION

R Commander is a package which gives graphical user interface (GUI) functionality like Minitab or SPSS to R. The purpose of this article is to provide the reader with enough information to make a decision whether or not R Commander is right for them and, if it is not, advice on suitable alternatives. It presumes the reader is familiar with the ascendency of R, the open source professional grade statistical package which has become one of the most widely used statistical software packages in the world.[1] A Google search will turn up lists of reasons for using R. I have three types of readers in mind:

1. Educators (Can I, should I, use R Commander in my teaching?)
2. Practitioners (How can R Commander help me with my R work? And if not R Commander, what are the options?)
3. Graduate students (Is using a GUI, and R Commander in particular, a good approach for learning R?).

Conflict of interest: The author declares no conflict of interest.

*Correspondence to: jason.wilson@biola.edu

Mathematics and Computer Science, Biola University, La Mirada, CA, USA

As R is open source (free), the R Core Team rightly focuses on maintaining and improving the base software. As a result, they do not offer a 'pretty face' GUI with push-button menus which contemporary software users have come to love. This is where R Commander comes in: it is probably the most popular GUI for R.[2] The R package name is 'Rcmdr', which is loaded just like any other R package (although loading all of the dependent packages takes a long time). The package name 'Rcmdr' will therefore be used hereafter (except in headings). One reason for Rcmdr's popularity is that it was one of the first GUIs for R. It has maintained this popularity because it is very intuitive, easy to use, and is actively maintained by its creator, John Fox. If users want to add features of R which are not in Rcmdr, they can write a 'plug-in', of which there are currently 22 available.[3,4] It is perhaps for these reasons that Heiberger and Neuwirth selected Rcmdr as the R GUI of choice for their creation of RExcel, an add-in to Excel which allows the user to run R and Rcmdr, from within Excel in a rather seamless integration (Windows only, see Box 1). In what follows, I will describe (1) the functionality of Rcmdr, (2) when to use it, (3) when not to use it, and (4) tips on using it in the classroom. The first three sections are directed at educators, practitioners, and graduate students while the fourth section primarily targets educators.

ALTERNATIVES TO R COMMANDER
(GRAPHICAL USER INTERFACE = GUI; INTEGRATED DEVELOPMENT ENVIRONMENT = IDE)

RExcel (Windows+Excel)
Add-In to Excel that allows the user to run R from within Excel. It includes all of the Rcmdr GUI menus inside the Excel Window (see Figure 3). The transition between Excel and R is relatively smooth (although not perfect), including the insertion of R output and graphs into the Excel spreadsheet. See the book *R Through Excel*[5] and the online video demonstration at http://rcom.univie.ac.at/

Deducer (Windows, Mac, Linux)
A GUI similar to Rcmdr, this package is probably the closest alternative to Rcmdr. It provides many (although not all) of the same features. For a discussion of the difference between Rcmdr and Deducer, see www.deducer.org/pmwiki/pmwiki.php?n=Main.ComparisonWithThePackageRcmdr
For a video demonstration, see www.deducer.org/pmwiki/pmwiki.php?n=Main.DeducerManual

RStudio (Windows, Mac, Linux)
An IDE (not a GUI) that makes work easier. As an IDE, it allows placement of the console, workspace, plots, programming scripts, etc. all accessible on one screen. A powerful option for teaching with RStudio is that it can be added to a school's server, allowing students access to the software via the web ('on the cloud'). This eliminates the hassle of software loading and adds the flexibility of continuing one project on multiple terminals. Note that while the Rcmdr package works on conventional RStudio, it does not work with RStudio 'on the cloud' (Figure 4) . http://rstudio.org/

REvolution R Enterprise (Windows, Linux)
IDE built on R. It is similar to RStudio, but designed specifically for business applications, and therefore is available for a fee (which includes technical support). A single-user license has been available free to educators, and now also to students. Compatible with Rcmdr.[6]
www.revolutionanalytics.com/

Rattle-The R Analytical Tool To Learn Easily (Windows, Mac, Linux)
A GUI developed for R for the specialized purpose of data mining. See the book *Data Mining with Rattle and R.*[7]
http://rattle.togaware.com/

RKWard (Windows, Mac, Linux)
RKWard is both an IDE and a GUI, the only package listed with this attractive feature. It is currently in the Beta version, so the Mac install exists but is purportedly difficult. 'RKWard tries to combine the power of the R-language with the (relative) ease of use of commercial statistics tools. Long-term plans include integration with office suites'.
http://rkward.sourceforge.net/
For a complete list of current R GUI projects see:
 www.sciviews.org/_rgui/

FUNCTIONALITY OF R COMMANDER

In a nutshell, Rcmdr has a Minitab or SPSS type of feel (Figure 1). The R package Rcmdr is available on Windows, Mac, and Linux, with only minor variations due to the operating systems. The full functionality can be found in the accessible documentation.[8,9] There are also online resources available.[10–12] In this section, therefore, I will just give the highlights. As you proceed, please refer to Figure 1. First, the menu buttons are at the top of the Rcmdr interface. Below it is a row of buttons for selecting, viewing, and editing the data and linear models in the workspace. Moving down is the Script Window which contains the R Console commands that Rcmdr generates after a menu selection is made. Any R command may be entered here (followed by clicking 'Submit' at right). This is extremely useful for bridging a user into the R Console. Below the Script Window is the Output Window, which shows the R command and associated output in color coding. At bottom is the Messages window which reports ordinary warnings R would generate along with added notes from using Rcmdr menus. This is the extent of it—everything necessary is contained in this single interface.

Next is the heart of Rcmdr: the menu buttons. I will here highlight them according to the Rcmdr menus, from left to right (top of Figure 1).

File: Allows the user to save the workspace, script window, output window, and exit.

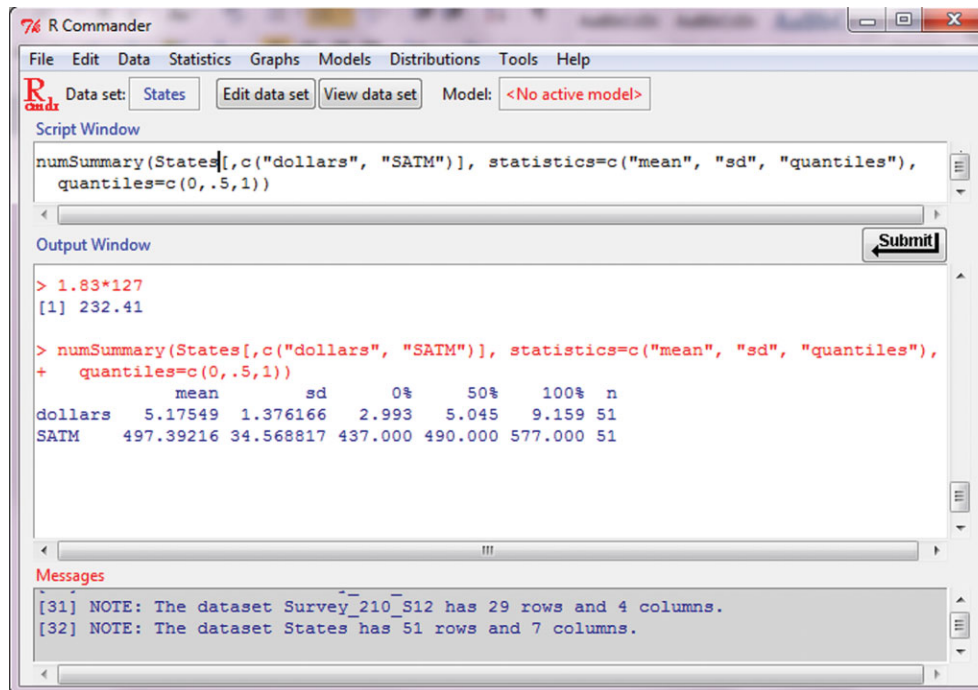*Edit:* Allows common editing features: cut, copy, paste, delete, find, select all, undo, redo.

FIGURE 1 | The basic R Commander graphical user interface (GUI) Window.

Data: Here are many extremely useful automated data import and modification features that even an intermediate level R user can benefit from. Data can be imported from many formats: tab-delimited, CSV, SPSS, Minitab, STATA, Excel, Access, and dBase. Data can also be selected from R packages, which gives students easy access to data. For an existing dataset, one can perform various manipulations or create new variables relatively easily, including conversion of numeric variables to factors. There are many other features, including creation of a new dataset by direct entry into a spreadsheet-like window (not available on Mac).

Statistics: A great selection is available. All manner of descriptive statistics, all the basic *t*-tests (which include confidence intervals), nonparametric tests, and analysis of variance (ANOVA) are available (only able to be selected if the loaded dataset has variables of the right type). Contingency table analyses can be performed on raw (column) data, and there is a slick feature to enter summarized (two-way table) data for chi-square testing. This is where regression models are created, with the options of linear, generalized linear, multinomial logit, and ordinal. In addition, there are more advanced methods such as correlation, Shapiro-Wilk, Bartlett's, and principal component analyses, among others. The menu items are generally very easy to use, which is a big plus.

Graphs: All of the basic graphs can be created (only able to be selected if the loaded dataset has variables of the right type): histogram, boxplot, QQ, scatterplot (plus Fox's own scatterplot matrix for >2 variables and a 3D scatterplot), line graph, bar graph, and pie chart. There are a few others. An added bonus is that with each graph, many of R's optional arguments are incorporated into straightforward check-box options, allowing the user easy access to R's powerful functionality. In addition to being able to copy/paste the displayed graph as a metafile, bitmap, or PDF, Rcmdr can save graphs in postscript and encapsulated postscript format.

Models: One difference between R and Minitab/SPSS is with linear models. Minitab/SPSS produce voluminous output when the right menu is selected. Instead, R creates a model object. The model object can be saved and then manipulated with subsequent commands to produce the desired output. In Rcmdr then, when a linear model (object) is created, its name appears in the 'Model' box (cf Figure 1 to Figure 2, 'LinearModel.1'). This object can then have, for example, the ANOVA table, or summary statistics, or diagnostic plots, or AIC generated. All of these menus, and quite a few more, are available to explore the models created using the Statistics > Fit Models menu.

Distributions: For each of 13 continuous distributions (normal through Gumbel) and 5 discrete
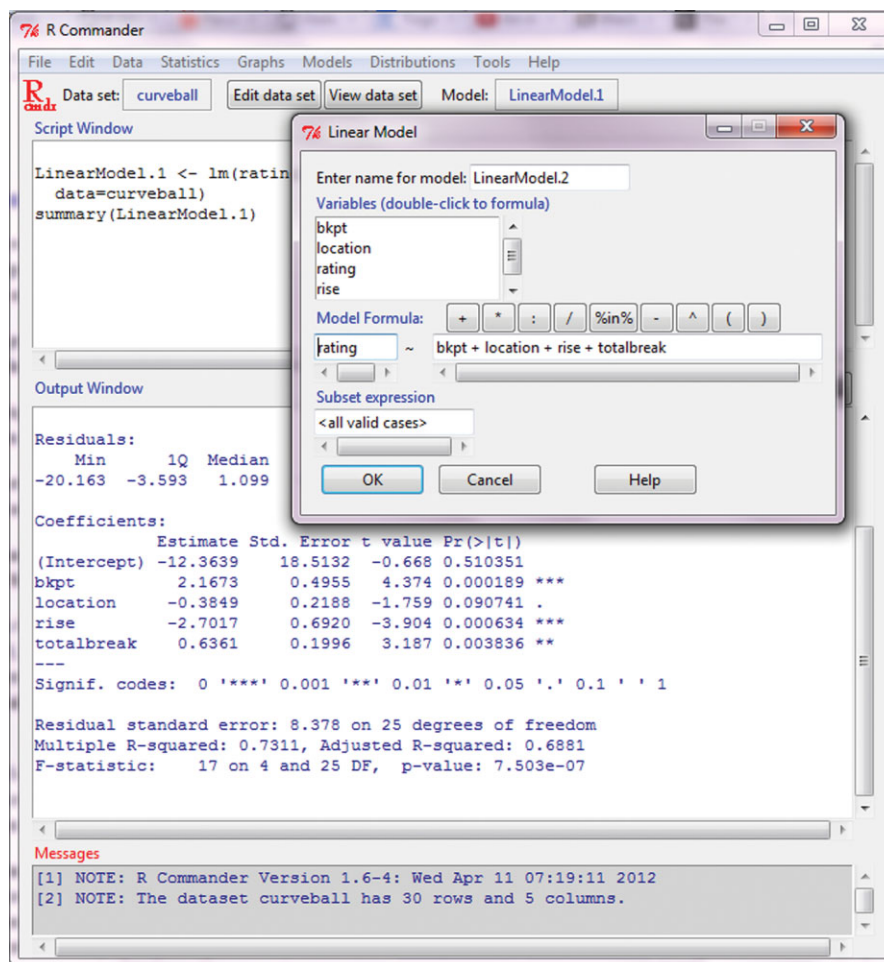
FIGURE 2 | Linear model menu with corresponding output shown. The ease of entry of the numerous model possibilities is featured.

distributions (binomial through negative binomial) the following four items can be obtained: quantiles, cumulative probability, random observations, and a distribution plot (user-specified parameters). In addition, the probabilities may be obtained for the discrete distributions.

Tools: As of the time of this writing, users have contributed 22 Rcmdr plug-ins, which are accessed here. Packages may also be loaded here.

Help: In addition to providing general help, one option pulls up the help file on R package datasets, which is good for students exploring. Also, a help button appears on all of the windows which open after menu selection.

WHEN TO USE R COMMANDER

As we have compared R to Minitab and SPSS in functionality, I recommend Rcmdr for those cases similar to when one would use Minitab or

SPSS. There are four particular purposes I believe Rcmdr may be beneficial for: quick entry into R (practitioner, graduate student, and educator), beginning to learn R programming (practitioner and graduate student), occasional use of R (practitioner and graduate student), and teaching introductory statistics (educator). In this section I elaborate on the first three cases. Teaching is covered in a subsequent section. It might be useful for the reader to take a look at Box 1 now to get a feel for the alternatives to Rcmdr. In this way, the reader's interests and the alternative's features may be weighed while reading. My remarks in this section were written to apply to GUIs in general and not merely for Rcmdr in particular. Therefore, for the readers who are 'shopping' for a GUI, I recommend substituting the term *GUI* in place of *Rcmdr* while reading.

Rcmdr provides a quick entry to the power of R because it can bring down the learning curve, relieving the user of the need to learn many commands. The
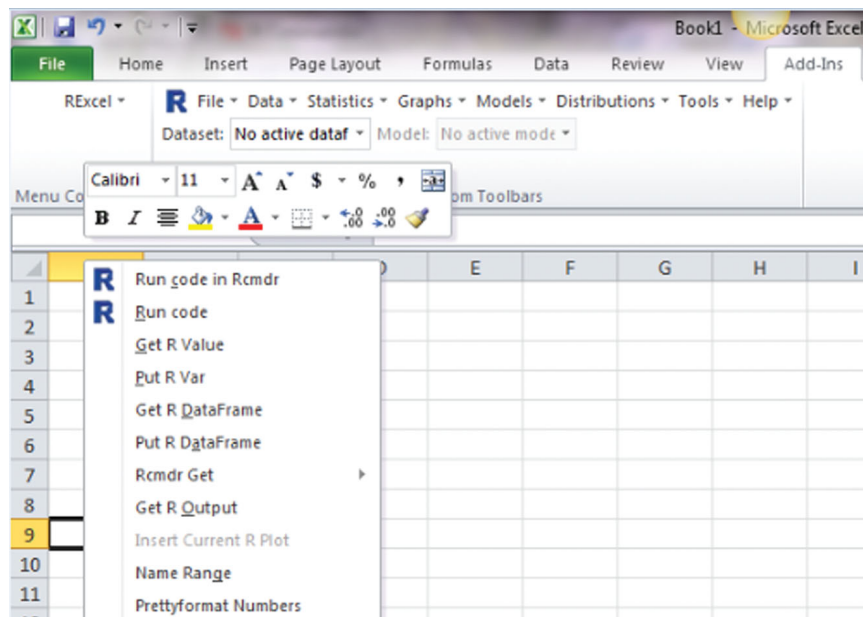
FIGURE 3 | RExcel screenshot highlighting the Rcmdr menu availability (top) as well as the right-click options in Excel to interface with R.

greatest manifestation of this power is in exploratory data analysis—the ability to rapidly compute many statistics and graphs.

Rcmdr can be a great asset to a user who wants to learn R thoroughly. The reason is that the script window displays and records R commands, making it extremely easy to learn, and to modify, basic commands. One class of useful commands is all of the routine statistical analyses. For example, running a *t*-test on Rcmdr eliminates the need to look up how to use the different arguments in *t.test(),* while providing correct syntax for the user's problem. This is convenient for lesser known analyses such as the Wilcoxon test or factor analysis. If this purpose is selected, the Rcmdr user should make plans to gradually transition themselves from Rcmdr to the R Console by beginning to use the scripts generated in Rcmdr's Script Window.

Rcmdr is also useful for the occasional R user who needs the commands refreshed, or simply to access its functionality. One very useful class of menus is the handy suite for loading data. For example, one menu was written expressly for Rcmdr that allows the user to load data directly from the clipboard (Data > Import Data > from clipboard). Another useful class of menus is the suite of linear model analysis tools.

WHEN NOT TO USE R COMMANDER

While reading this section, the reader be cognizant of the Rcmdr alternatives in Box 1, and their own purpose in exploring GUIs. In general, I would not recommend Rcmdr for purposes outside those given in the previous section. Indeed, the author of Rcmdr himself said in an interview, 'I suppose that the R Commander can ease users into writing commands, in as much as the commands are displayed, but I suspect that most users don't look at them. I think that serious prospective users of R should be encouraged to use the command–line interface along with a script editor of some sort'.[13] I agree. For those who truly want to learn R, I advise caution with Rcmdr. It can be useful, as previously described, in the initial stages of learning R. However, as one advances, he or she should wean themselves off of Rcmdr as a mother weans its child off the breast.

The primary reason for weaning off of Rcmdr is that it inhibits the development of advanced R skills. Statistics is an art. The Rcmdr-artist is confined to a pre-packaged art kit with limited colors and media compared to the R-artist who has free run of the art supply store. There is a different mindset when using Rcmdr versus the R console. Rcmdr rightly frees the user to focus on the big picture and the end results. At the same time, however, it limits the options of the user. Its use can even prevent the development of skills that will lead to advanced R console work, such a modifying a dataset in a way not amenable to Rcmdr's limited options, or performing a bootstrap analysis, for example. Clicking buttons is far easier than typing commands and can lull the user into laziness. For every menu command used, the user has missed an opportunity of reinforcing the learning of an R command.

The example of graphs illustrates the limitations of Rcmdr. While Rcmdr's selection of graphs is nice, it is limited. What if you want to make a double line plot? How about changing the color of the boxes in your multiple boxplot? What if you want to change the title or add a legend? When you create plots with commands, a way of thinking is developed such that the underlying nature of the process is reinforced. In addition, mastery of this skill prepares the user to advanced thinking regarding graphs. For example, on several projects I have needed to create dozens of graphs from successive variables. I used a loop and wrote the graphs into a single long .pdf file. I would not have had the insight to do this if I did most of my work on Rcmdr. Furthermore, the modestly experienced creator of graphs by command can be far quicker than Rcmdr. Although both users have access to the script which created the graph, the Rcmdr user is less equipped to understand or reuse it. That said, these are not faults of Rcmdr, as its purpose was never intended to perform all such tasks. These are merely the buoys marking the transition from the R inlet to the R sea beckoning the user to set sail into deeper waters...

TEACHING WITH R COMMANDER

Ninety-nine percent of my students own a laptop, removing the need for a computer lab—if they can get the software onto their laptop. The fact that R is open source (free) makes this a viable possibility. If someone does use R, should Rcmdr be the R package used for instruction? And just how feasible is this? My remarks on this topic are divided into two subsections which address these questions.

Should I Use R Commander for My Class?

This section consists of my experience in teaching three introductory statistics courses, with remarks pertaining to Rcmdr on each, and a summary at the end. All of the introductory statistics courses at my school are three semester hours with no computer lab section.

Introduction to Probability and Statistics

Some of my students plan on graduate school, but most do not, and will not be active in data analysis in their careers. I have experimented with Excel, R console, and Rcmdr in them and have settled on Rcmdr. We use Rcmdr like a giant 'statistics calculator' obtaining descriptive and inferential statistics very easily. Although we do cover theory, Rcmdr saves class time and allows interpretation of results to be the focus, rather

than the correctness of hand calculations. Educators have posted many resources for using Rcmdr online, including a video introduction,[10] individual video tutorials,[11] and a PDF guide.[12]

Biostatistics

Graduate school plans are mixed, but this group has a stronger math background than my *Intro* group. I tried and rejected SPSS (good software, but the site license made its use for class demonstrations from my laptop clunky, and students were restricted to the campus lab). Next, I cycled between R console and Rcmdr. I expect to settle on the R console with the interactive development environment (IDE) RStudio in Fall 2012 (see Box 1). On the one hand, it is easier to get students to use and appreciate statistical software with Rcmdr. On the other hand, with Rcmdr I have found many reach a point where they have questions and needs beyond it. When this happens, the GUI habits they have learned become a barrier to transitioning to typing commands. When I begin with the R Console, they struggle climbing the steeper learning curve, and it requires more care on my part with instruction and class time. However, most (not all) are still able to 'get it'. Student abilities, our class discussions, and my R demonstrations, are richer at the end.

Business Statistics

Most students have Excel on their laptop, theoretically making Excel a viable choice for statistics instruction on the laptop. This may be desirable for business majors. I encountered two fundamental problems, however. First, there is no longer a single Excel package. The Windows 2000, Windows 2007, Mac 2004, and Mac 2008 versions are all significantly different to students. Classes in which I have attempted 'laptop Excel' have had roughly even distributions of these versions, rendering version-specific classroom demonstration infeasible and creation of supporting materials burdensome. The second problem is Excel's well-known statistical faults. Examples include spotty graph capabilities (e.g., no boxplots), sparse statistical capabilities (e.g., no confidence intervals or many routine hypothesis tests), and uncorrected errors.[14] An adjunct currently teaches our course without software, but when he moves on, my task will be to redesign the course. I will begin by meeting with School of Business faculty to discuss the software. RExcel may be a good option (see Box 1), although it would force Mac users into the computer lab, unless the Open Office version were released in time....

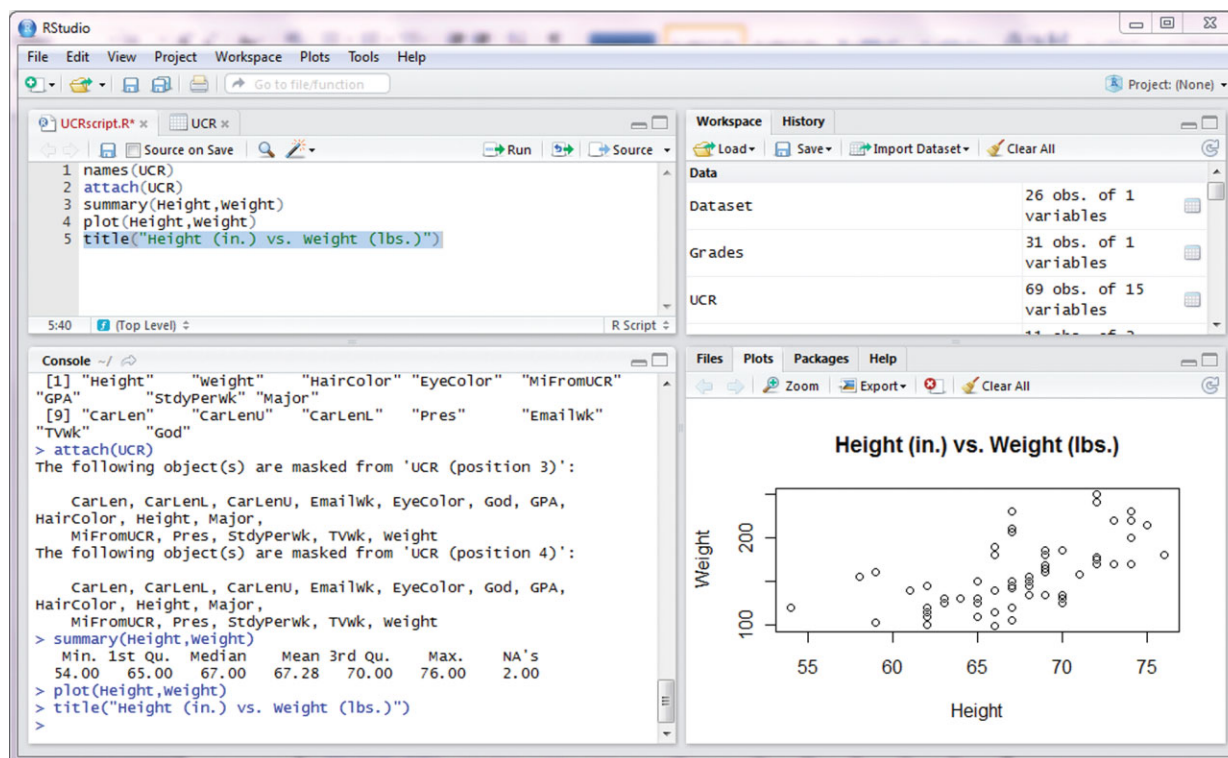In summary, the appropriateness of Rcmdr depends on what the students should know for their

FIGURE 4 | RStudio screenshot illustrating functionality. Upper left is a script window, with a tab to show the attached dataset. Upper right is the list of variables in the workspace, with a tab to show the history. Lower right is the latest plot, with tabs to show the files, manage packages, and help menus. Lower left is the conventional R Console.

careers. I maintain that Rcmdr is suitable for students who will not be analyzing data, but that a command line version of R is more suitable for students who will need to conduct data analysis, perhaps in a graduate program or on the job. Of course, classes are not entirely homogenous in this regard and wisdom should be used to guide the choice.

What Will It Take to Conduct a Class using Rcmdr on Student Laptops?

The three main issues I encounter when working with students and Rcmdr on laptops are: (1) the initial installation of Rcmdr, (2) the loading of the dependent Rcmdr packages, and (3) finding Rcmdr help. First, regarding installation, installing the R software is simple and generally presents no problems. Most Windows users are able to get Rcmdr to install, although I encounter a few quirky issues particular to specific hardware/software configurations that are easy to troubleshoot individually. For the Macs, some versions have issues with broken or absent installations of X11, and others have issues with Mac OS' tcltk or R's tcltk package. This semester, I successfully troubleshot all seven of the significant installation issues (out

of 95 students), but this required extra out-of-class time.

Second, Rcmdr depends upon a large number of dependent packages. I ask students to attempt installation on their own before class, but perhaps only 1/3 succeed. Guiding a classroom of 30 students on wifi to download and install R works fine in my school's classrooms. Guiding them to install Rcmdr conventionally does not, because of the high download time required by all of the dependent packages. To avoid this, I have them call the following two commands:

```
>install.packages(''Rcmdr'', depen
dencies=FALSE)
>install.packages(''car'', depen
dencies=FALSE)
```

This installs Rcmdr plus the one necessary dependent package, freeing the class to engage in a hands on demonstration immediately. I ask them to finish installing the other packages on their own. A few do not, which leaves them short later in the class (e.g., some summary statistics and chi-square options do not appear in the Rcmdr menus).

Lastly, 'help' in the R help files is for the R functions and therefore not usually useful to Rcmdr

users. They need help remembering which commands to use, and how to use them in Rcmdr, which are not available in Rcmdr itself. To supplement, I have created a document listing the desired function and the series of menus needed to achieve it, as covered in class. In addition, I am building a set of video tutorials for the main Rcmdr features we use.[11]

A few parting comments to prospective teachers who would use Rcmdr are in order. First, when attempting the use of Rcmdr on student laptops, I have found it best to work with our IT department to ensure that some campus computers have R with Rcmdr and its dependent packages available. In this way, students without Rcmdr on their computers (possibly due to persistent installation issues) can still do their homework. IT can be a valuable resource and some IT departments even handle student installation issues! Second, if you are an experienced R user, remember that Rcmdr may not be able to do everything you are used to. For example, showing a table of the columns of the steps needed to compute sums of squares is trivial on the R Console, but takes many menu commands in Rcmdr. I recommend 'learning' Rcmdr itself in order to use it effectively in class. Third, some recommend using Rcmdr while running RStudio (Figure 4). It adds good features and the interface is the same on all platforms. This may make it more likely that students will migrate from Rcmdr to RStudio than from Rcmdr to the older R Console interfaces. It should be noted, however, that Rcmdr will not work with RStudio when it is being used 'on the cloud' (see Box 1). Finally, be aware that Rcmdr is not without its issues. For example, it can occasionally become unstable. Also, for many graphs you cannot change the default x and y axis labels.

CONCLUSION

It has been said, 'Guns don't kill, people do'. To this could be added, 'Computers don't do statistics, people do'. Guns and computers are simply tools, and so is statistical software. The trick is selecting the right tool for the job. The purpose of this article has been to inform and advise the reader on wisely selecting the right tool for enhancing their statistical work with R. Rcmdr is very good for what it was designed for: quick entry into R, beginning to learn R programming, occasional use of R, and teaching introductory statistics. There are alternative GUIs to consider for these purposes as well (see Box 1). To the teacher—the laptop approach with Rcmdr can be fruitful, but the benefits should be balanced with the costs of installation issues and sharing class time with the added software. To the student and practitioner—if you plan on learning R as a serious tool, then I would recommend starting right off with an IDE (see Box 1) instead of Rcmdr or other GUIs. If you do plan on going with a GUI, then Rcmdr is a good one to use, as described, particularly because of the Script Window's ease at transitioning to writing R commands. In all—may you have wisdom in your search.

REFERENCES

1. Muenchen RA. The Popularity of Data Analysis Software. 2012. Available at: http://r4stats.com/popularity. (Accessed April 11, 2012).

2. Ohri A. *The Top Statistical Softwares (GUI)*. 2010. Available at: www.decisionstats.com/the-top-statistical-softwares-gui/. (Accessed April 11, 2012).

3. Fox J. Extending the R Commander by "Plug-in" packages. *R News* 2007, 7:46–52. Available at: http://tolstoy.newcastle.edu.au/R/e3/help/att-2806/wrapper.pdf. (Accessed April 13, 2012).

4. Fox J. The R Commander: A Basic-Statistics GUI for R. 2010. Available at: http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/. (Accessed April 11, 2012).

5. Heiberger R, Neuwirth E. 2009. *R Through Excel*. New York, NY: Springer.

6. Revolution Analytics On-Site R Commander Training Course. Available at: http://www.revolutionanalytics.com/products/training/on-site/introduction-to-r-commander.php. (Accessed May 29, 2012).

7. Williams G. 2011. *Data Mining with Rattle and R*. New York, NY: Springer.

8. Fox J. Getting Started With the R Commander. 2006, 2–8. Available at: http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/Getting-Started-with-the-Rcmdr.pdf. (Accessed April 11, 2012).

9. Fox J. Rcmdr: The R Commander: a basic-statistics graphical user interface to R. *J Stat Softw* 2005, 14:1–42.

10. Wagaman A. An Introduction to R with R Commander for Introductory Statistics. 2011. Available at: http://www.amstat.org/education/webinars/index.cfm. (Accessed May 29, 2012).

11. Wilson JR. Commander Demonstrations. 2012. Available at: https://sites.google.com/a/biola.edu/r-commander-demonstrations/home. (Accessed April 11, 2012).

12. Karp NA. R Commander An Introduction. 2010. Available at: http://cran.r-project.org/doc/contrib/Karp-Rcommander-intro.pdf. (Accessed May 29, 2012).

13. Ohri A. Interview of Professor John Fox, Creator of R Commander. 2009. Available at: www.decisionstats.com/interview-professor-john-fox-creator-r-commander/. (Accessed April 11, 2012).

14. McCullough BD, Heiser DA. On the accuracy of statistical procedures in Microsoft Excel 2007. *Comput Stat Data Anal* 2008, 52:4570–4578. Available at: www.pages.drexel.edu/~bdm25/excel2007.pdf. (Accessed April 11, 2012).