

# R for SPSS Users: A User Friendly Introduction Using Rcommander

AAPOR Webinar Series  
April 27, 2016

Trent D. Buskirk, Ph.D.  
Marketing Systems Group  
Instructor

SPSS



R



## Outline for the Short Course

---

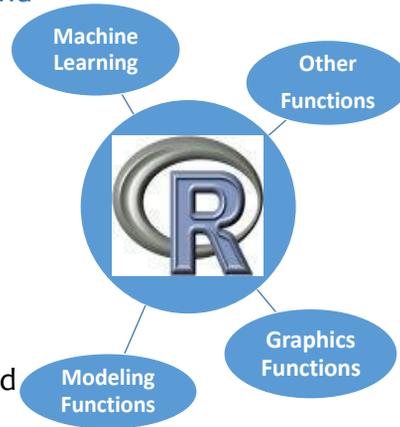
- 🕒 A quick overview of R and R Commander
- 🕒 Loading R Commander into an R Session
- 🕒 Exploring the R Commander Graphical User Interface
- 🕒 Introduction to Basic Data Formatting/Editing using R Commander
- 🕒 Introduction to Descriptive Statistics Using R Commander
  - 📊 Means, proportions
  - 📊 Contingency Tables
- 🕒 Introduction to basic graphics with Rcommander

# What is R?

Besides being the 18<sup>th</sup> letter in the English Alphabet, R is a language and environment for statistical computing and graphics. (i.e. free version of S)

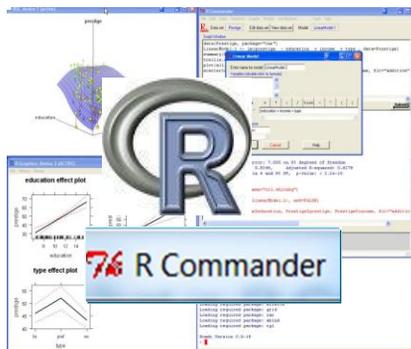
R commands are executed via scripts/syntax in an interactive environment.

R functionality is extended by **packages** that are written for specific statistical/ data management/ machine learning and modeling tasks



# What is RCommander?

R Commander is a bundle of R packages that essentially overlays a graphical user interface (GUI) on top of the R environment.



## Some Key Online Resources

 **R-Seek** ([www.rseek.org](http://www.rseek.org)): Think Google specifically for R.

 Can search for program syntax, packages, R reference cards and more specifically related to R.



 **Quick R** (<http://www.statmethods.net/>): A very top-level synopsis of R commands organized by function:

 Stats, Input, Graphics, etc.

 Contains both package information (i.e. names, descriptions) as well as code snippets.



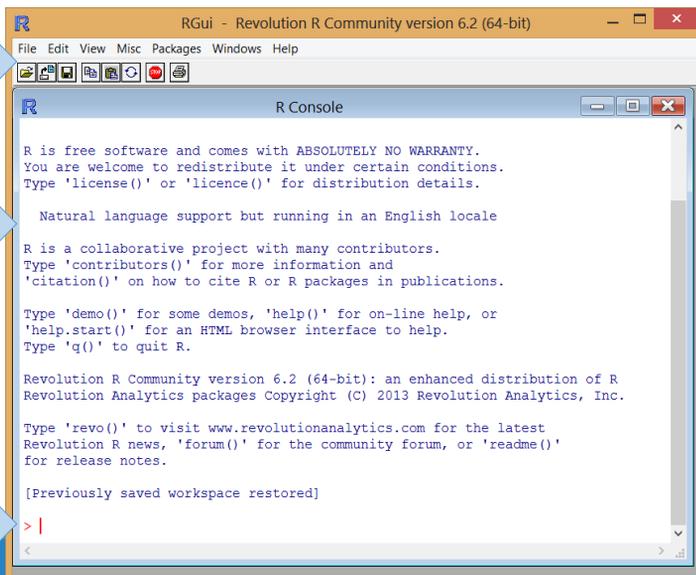
 **R Blogspot** (<http://www.r-bloggers.com/>): Meta Blogging site specifically dedicated to R users.

## Working with R

Master Toolbar  
Menu and  
Shortcut icons

R Console  
Window

R Command  
Prompt



```
RGui - Revolution R Community version 6.2 (64-bit)
File Edit View Misc Packages Windows Help
[Master Toolbar Icons]

R Console
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute 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.

Revolution R Community version 6.2 (64-bit): an enhanced distribution of R
Revolution Analytics packages Copyright (C) 2013 Revolution Analytics, Inc.

Type 'revo()' to visit www.revolutionanalytics.com for the latest
Revolution R news, 'forum()' for the community forum, or 'readme()'
for release notes.

[Previously saved workspace restored]

> |
```

## Steps 1 and 2: Launch R and Select Install Packages



```
> local (pkg <-
+ if (nchar(pkg)
Loading required
Loading required package: car
Loading required package: MASS
Loading required package: nnet

Rcmdr Version 1.8-3

Attaching package: 'Rcmdr'

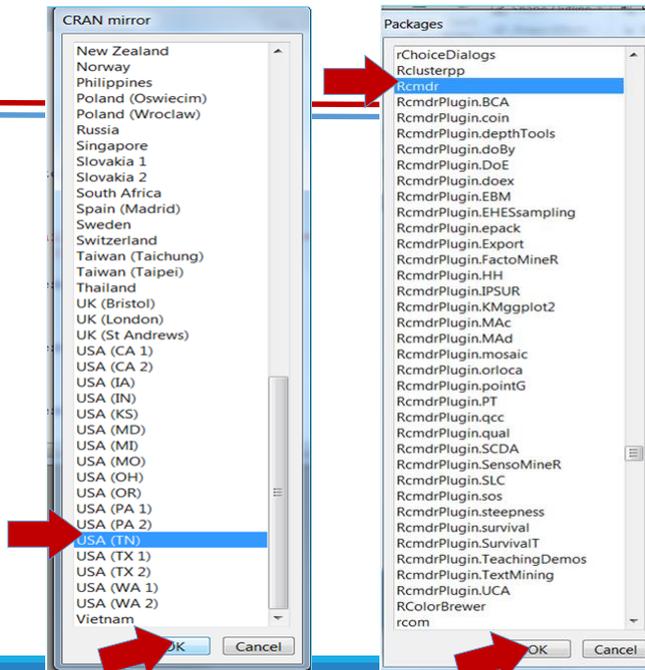
The following object(s) are masked from 'package:tcltk':

  tclvalue

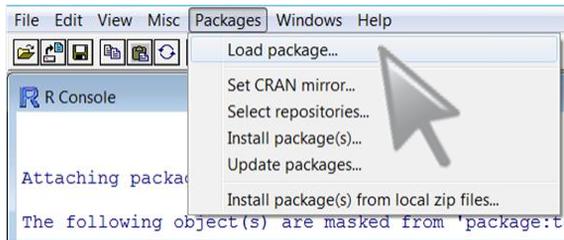
> local (pkg <- select.list(sort(.packages(all.available = TRUE)),graphics=TRUE)
+ if (nchar(pkg)) library(pkg, character.only=TRUE))
> utils::menuInstallPkgs()
--- Please select a CRAN mirror for use in this session ---
Error in contrib.url(repos, type) :
  trying to use CRAN without setting a mirror
> |
```

Steps 3 and 4:  
Select  
Comprehensive R  
Archive Network  
(CRAN) and then  
the Packages you  
need (use ctrl to  
select multiple  
packages)

We want:  
*Rcmdr*

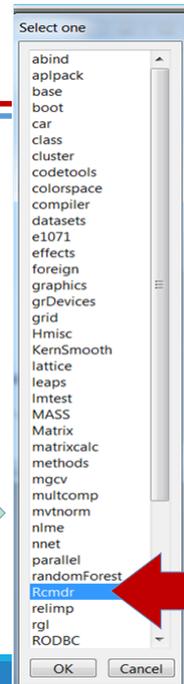


## Step 5: Load/Launch Packages (Rcmdr)



This list contains all of the packages that you have previously installed (or are installed by default in R)

Once you install a package, you do not need to re-install it. You just need to load it (launch) when you start R.



4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

## A Short Cut for Launching Installed Packages

- ☞ After you have installed a package in R it stays in memory from one launch of R to another.
- ☞ However, generally each time you launch R, you need to re-load packages (such as R commander).
- ☞ You can select Packages → Load package... and select the package name you want to launch (i.e. Rcmdr) ... OR...
- ☞ You can type `library(PackageName)` at the command prompt (`>`) [e.g. `> library(Rcmdr)`]

4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

10

# A Quick Overview of R Commander

- 🕒 R Commander allows users who are familiar with the GUI of SPSS and EXCEL and JMP to access the functionality of R via a window driven graphic user interface.
- 🕒 While R scripts are generated that correspond to commands accessed and selected via a series of drop-down menu selections (similar to paste syntax in SPSS).
- 🕒 The script versions of the commands appear in the Script window and can be copied, edited and rerun by highlighting and submitting (like SPSS syntax window).

# Orientation to the R Commander GUI

1 **Toolbars and Data Sets Index**

2 **Script Pane**  
(commands pasted here or typed) To execute commands entered in this section, highlight them and then click on the submit button

3 **Output Pane**  
Commands pasted in Script Window are executed here. If using R studio, the output window is not included, but is part of the R Studio Console Window

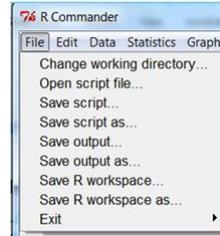
4 **Messages Pane**  
(Warnings/Errors posted here)

The screenshot shows the R Commander window with a menu bar (File, Edit, Data, Statistics, Graphs, Models, Distributions, Tools, Help) and a toolbar with buttons for 'Data set: <No active dataset>', 'Edit data set', 'View data set', and 'Model: <No active model>'. The main area is divided into a 'Script Window' at the top and an 'Output Window' at the bottom. A 'Messages' pane at the very bottom displays a warning: '[2] WARNING: The Windows version of the R Commander works best under RGui with the single-document interface (SDI); see ?Commander.'

# Exploring the R Commander Toolbars

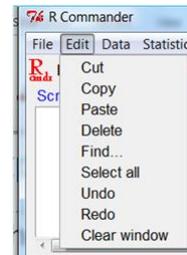
## The File Menu Options For:

- Loading/Saving Script files
- Saving Output and R workspaces
- Exiting



## Edit Menu Options include:

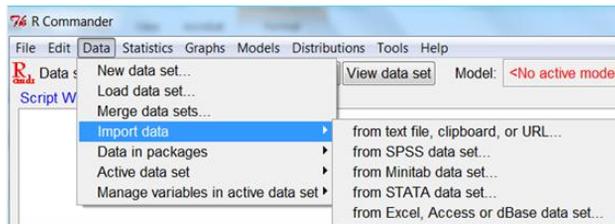
- Copy, cut, paste options for both script and output window contents
- Right click options are also available within each of the panes



# Exploring the R Commander Toolbars, Cont.

## The Data menu contains options for:

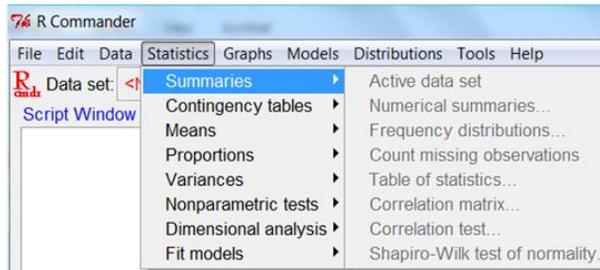
- Creating new R data sets
- Loading data sets that exist on your computer already as well as data that are contained in other R packages
- Importing Data from other Databases/Statistical packages



## Exploring the R Commander Toolbars, Cont.

 The Statistics menu contains options for:

-  Summary Statistics for continuous and categorical data
-  Means, Proportions, and Variance Summaries/Tests
-  Various Nonparametric statistical tests, cluster analysis and linear models



4/26/2016

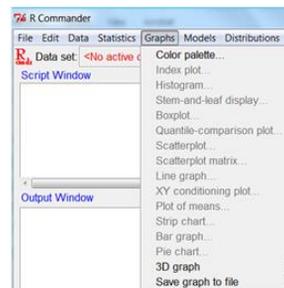
TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

15

## Exploring the R Commander Toolbars, Cont.

 The Graphs menu contains options for:

-  Generating various types of graphs including:
  - ◆ Histograms, Boxplots, QQ plots, 3-D plots, Scatterplots, Bar graphs, Pie Charts.
-  Exporting/Saving graphical output
-  Controlling basic colors of the plots



 The Models menu contains options for:

-  Model summaries, diagnostic statistics, hypothesis testing, residual plots and confidence/prediction intervals



4/26/2016

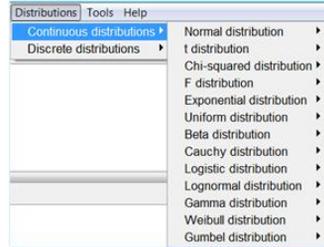
TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

16

## Exploring the R Commander Toolbars

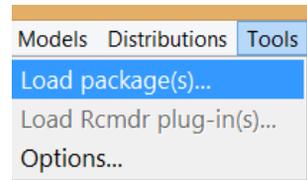
Ⓜ The Distributions Menu Contains Options for:

- 📊 Generating random observations, probabilities and quantiles for various continuous and discrete probability distributions



Ⓜ The Tools Menu Contains Options for:

- 📊 Loading additional packages into R
- 📊 Additional R Commander options (e.g. font and output controls).



## Example 1 – Importing Excel Files Into R

Ⓜ The Excel 2007 data file called R4SPSSData contains data from 15 hypothetical households including information on:

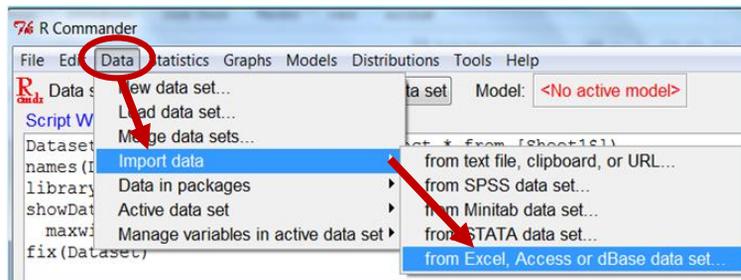
- 📊 household size (Hhsize)
- 📊 rental status (OwnRent)
- 📊 internet connectivity (InternetHH)
- 📊 time in media survey panel (YrsinPanel)
- 📊 total TV time (in hours) from the past week (TotTVLSW)
- 📊 Age of panelist (AgeofPanelist)

Ⓜ For this example, we will demonstrate how to import this Excel file into R for further processing and analysis and call it “Rtest”

## Example 1, Continued

🕒 To Import data from Excel, (and similarly for SPSS, Minitab, Stata or other source) follow the menu sequence:

📊 Data → Import data → from Excel, Access,...



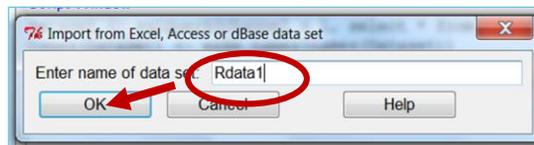
4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

19

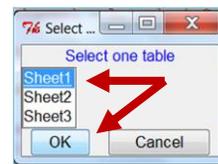
## Example 1, Continued

🕒 Enter the name of the data set – this will be the name you use to reference the data set in R (case matters in R; e.g. SPSS != spss)



🕒 An Open Documents window will appear- in the lower right hand corner be sure to select .XLSX as the extension (default is .XLS)

- Once you have selected the file and clicked the “Open” tab, Select the Excel sheet to import.



4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

20

## Example 1, Continued

If successful, the imported data set name will appear in the tool bar and the View and Edit data set tabs will become active.

The screenshot shows the R Commander interface with the 'Data set' field containing 'Rdata1'. The 'Edit data set' and 'View data set' buttons are active. Below, the 'Data Editor' window displays a table with 15 rows and 7 columns: SpecID, Hhsize, NumTVs, InternetHH, OwnRent, YrsInPanel, and TotTWLSW. The 'Rdata1' window also displays the same data table.

	SpecID	Hhsize	NumTVs	InternetHH	OwnRent	YrsInPanel	TotTWLSW
1	101	1	1	1	1	0.25	12.25
2	102	1	2	1	1	0.5	14.5
3	103	3	1	1	1	1.25	22
4	104	1	1	1	1	1.5	28
5	105	2	1	1	2	1.75	17.25
6	106	4	1	0	2	2.25	40
7	107	2	4	0	2	1.5	75.25
8	108	NA	3	1	2	0.25	10.25
9	109	5	2	0	1	1	47
10	110	4	2	0	1	2	39.75
11	111	6	5	1	1	2.5	26.25
12	112	5	4	1	1	3	21.5
13	113	3	3	0	1	2.75	44.25
14	114	1	16	1	1	3	32.25
15	115	2	4	0	2	4.75	50

## Example 1 – DEMO



Import the dataset R4SPSSData into R via R Commander and name the dataset Rtest.

The screenshot shows the R Commander interface. The 'Data set' field is empty, and the 'Edit data set' and 'View data set' buttons are disabled. The 'Script Window' is empty, and the 'Output Window' is empty. The 'Messages' window shows the following text: 'RGui with the single-document interface (SDI); see ?Command'.

## Categorical Variables and R Commander

---

- Ⓜ Variable type attribution is extremely critical for maximizing the potential of R commander!
- Ⓜ Specifying nominal categorical variables as factors will allow them to be accessed as grouping/stratification variables, for use in boxplots, to define groups for an ANOVA or to control plotting symbols, for example.
- Ⓜ Binary variables coded with numbers will not automatically be interpreted as a factor (e.g. 1=Male; 2=Female) so you will need to specify such variables as factors and use the numbers or specify the values for the levels.

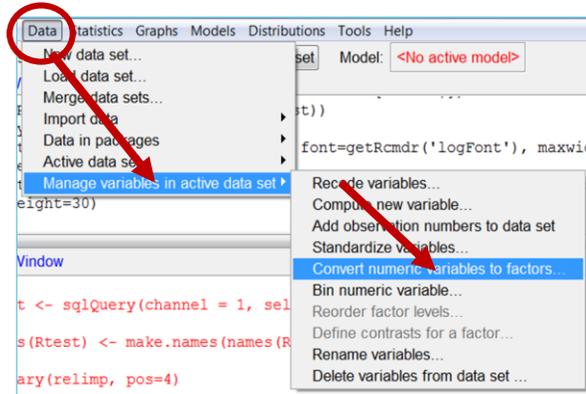
## Example 2

---

- Ⓜ In the R4SPSSData.xlsx file, Internet Connectivity (InternetHH) within a sampled household is coded as a binary variable with 1 implying internet connection and 0, no connection. OwnRent is also a binary variable with 1=Own and 2=Rent.
  - Ⓜ (A) Using R commander, convert the numeric variable InternetHH to a factor called **Internet** with values “YES” and “NO”.
  - Ⓜ (B) Using Rcommander create a 0/1 variable called **Own** that is 1 for Owners and 0 for Renters. (in this case we will use a short cut Own=2-OwnRent).
  - Ⓜ (C) Create a new variable called TotTWQ to represent the quartile in which the case’s TotTWLSW value falls.

## Example 2(A) - Solution

Step 1: In R commander, Follow this menu sequence:  
Data → Manage Variables in Active Data Set →  
Convert Numeric Variables to Factors



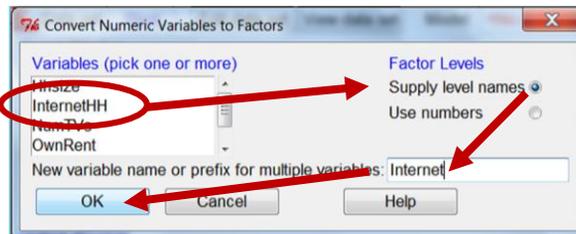
4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

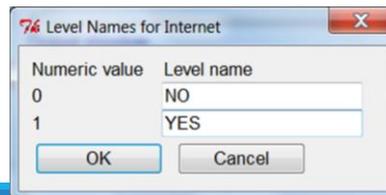
25

## Example 2(A), Solution Continued

Select Variable to convert then specify level names or numbers and the new name, if applicable.



- If levels are selected, then a prompt will appear with as many levels as unique values for the variable.



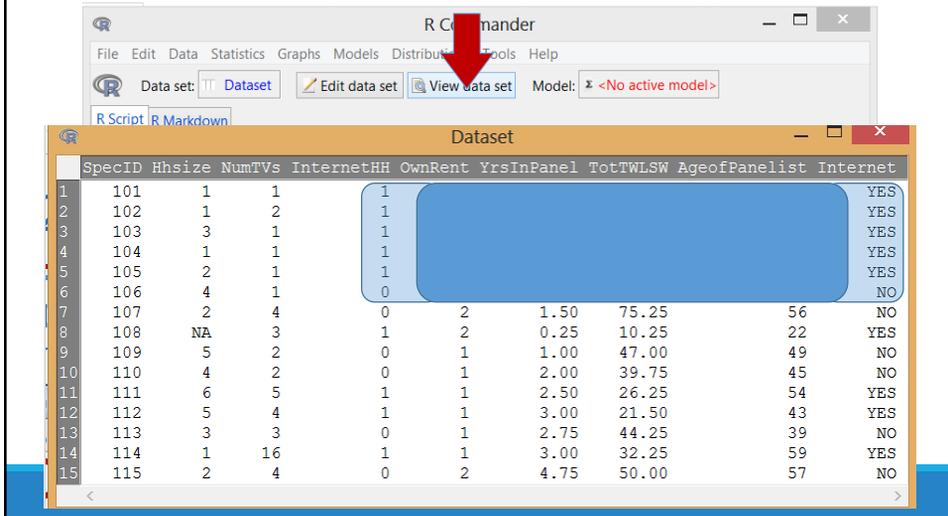
4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

26

## Example 2(A) Solutions Continued

- In the next lecture we will learn how to generate basic tables to check answers. But another way we can do a simple check is simply to view the data –



The screenshot shows the R Commander interface. The 'Dataset' window displays a table with the following data:

	SpecID	Hhsize	NumTVs	InternetHH	OwnRent	YrsInPanel	TotTWLSW	AgeofPanelist	Internet
1	101	1	1	1					YES
2	102	1	2	1					YES
3	103	3	1	1					YES
4	104	1	1	1					YES
5	105	2	1	1					YES
6	106	4	1	0					NO
7	107	2	4	0	2	1.50	75.25	56	NO
8	108	NA	3	1	2	0.25	10.25	22	YES
9	109	5	2	0	1	1.00	47.00	49	NO
10	110	4	2	0	1	2.00	39.75	45	NO
11	111	6	5	1	1	2.50	26.25	54	YES
12	112	5	4	1	1	3.00	21.50	43	YES
13	113	3	3	0	1	2.75	44.25	39	NO
14	114	1	16	1	1	3.00	32.25	59	YES
15	115	2	4	0	2	4.75	50.00	57	NO

## Example 2(B)

- (B) Using Rcommander create a 0/1 variable called **Own** that is 1 for Owners and 0 for Renters.

We want the value of 2 (renter) to be 0

$$2 - 2 = 0$$

We want the value of 1 (owner) to be 1

$$2 - 1 = 1$$

- A simple “equation” to convert a 1/2 variable to a 1/0 variable is:

$$\text{new-1-0-var} = 2 - \text{old-2-1-var}$$

## Example 2(B) Solutions Continued

The screenshot shows the SPSS Data menu with 'Manage variables in active data set' selected. A sub-menu is open, showing 'Compute new variable...' as the selected option. Below, the 'Compute New Variable' dialog box is shown. The 'Current variables' list includes 'OwnRent'. The 'New variable name' is 'Own' and the 'Expression to compute' is '2 - OwnRent'. The 'OK' button is highlighted with a red arrow.

4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

29

## Example 2(B) Solutions Continued

Here's what the data set now looks like after adding the variable "Own"

The screenshot shows the SPSS Dataset window with the following data:

	OwnRent	YrsInPanel	TotTWLSW	AgeofPanelist	Internet	Own	
1	1					1	
2	1					1	
3	1					1	
4	1					1	
5	2					0	
6	2					0	
7	2	1.50	75.25		56	NO	0
8	2	0.25	10.25		22	YES	0
9	1	1.00	47.00		49	NO	1
10	1	2.00	39.75		45	NO	1
11	1	2.50	26.25		54	YES	1
12	1	3.00	21.50		43	YES	1
13	1	2.75	44.25		39	NO	1
14	1	3.00	32.25		59	YES	1
15	2	4.75	50.00		57	NO	0

4/26/2016

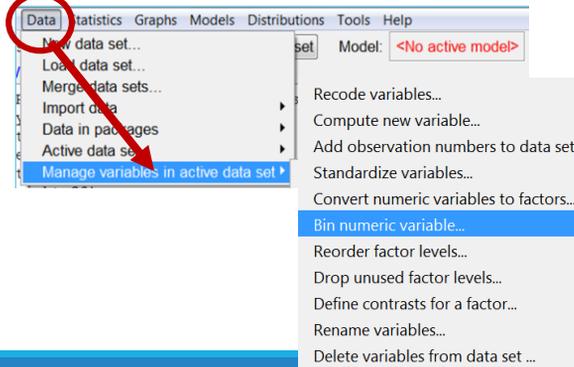
TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

30

## Example 2(C) Solutions

🔗 (C) Create a new variable called TotTWQ to represent the quartile in which the case's TotTWLSW value falls.

📖 Recall quartiles represent values that evenly split up a distribution (dataset) into four parts – each part containing a quarter (approximately) of the data...



4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

31

## Example 2(C) Solutions, Continued

A screenshot of the 'Bin a Numeric Variable' dialog box in SPSS. The 'Variable to bin (pick one)' list contains 'NumTVs', 'Own', 'OwnRent', 'SpecID', 'TotTWLSW', and 'YrsInPanel'. 'TotTWLSW' is selected. The 'New variable name' field contains 'TotTWQ'. The 'Number of bins' is set to 4. The 'Binning Method' section has 'Equal-count bins' selected. Annotations include: a red arrow pointing to 'TotTWLSW' with the text 'Variable you want to BIN!'; a red arrow pointing to the 'New variable name' field with the text 'How do you want the new variable to be labelled?'; a blue arrow pointing to the 'New variable name' field with the text 'Name of New Variable'; a blue arrow pointing to the 'Number of bins' field with the text 'Number of BINS to USE in creating the new version of the variable'; and a blue arrow pointing to the 'Equal-count bins' radio button with the text 'Choose Equal Count Bins for Quantiles'.

4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

32

## Example 2(C) Solutions, Continued

The screenshot shows the R Commander interface. The 'Dataset' window displays a table with columns: Vs, Internet, HH, OwnRent, YrsInPanel, TotTWL\$W, AgeofPanelist, Internet, Own, and TotTWQ. The console window shows the following R output:

```
counts:
TotTWQ
[10.2,19.4]  (19.4,28]  (28,42.1]  (42.1,75.2]
           4         4         3         4

percentages:
TotTWQ
[10.2,19.4]  (19.4,28]  (28,42.1]  (42.1,75.2]
           26.67      26.67      20.00      26.67
```

4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

33

## Example 2 (C), Solutions Continued

To confirm the conversion, you should now see the new variable Internet appear as an eligible variable for computing frequencies, for example (i.e. can apply the table function to the variable Internet)

Following this menu sequence:  
**Statistics** → **Summaries** → **Frequency Distributions**

The screenshot shows the SPSS Statistics menu with 'Summaries' selected, and the 'Frequency Distributions' dialog box. In the dialog box, the 'Internet' variable is selected in the 'Variables (pick one or more)' list, which is circled in red.

4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

34

## Example 2(C), Solutions Continued

The R Commander Script Window now shows the .Table command sequence to generate the frequencies and distribution (i.e. counts & percentages) of sampled households who do and do not have internet (which are depicted in the output window below)

```
maxheight=30)
Rtest$Internet <- factor(Rtest$InternetHH, labels=c('NO','YES'))
.Table <- table(Rtest$Internet)
.Table # counts for Internet
round(100*.Table/sum(.Table), 2) # percentages for Internet
remove(.Table)

> Rtest$Internet <- factor(Rtest$InternetHH, labels=c('NO','YES'))
> .Table <- table(Rtest$Internet)
> .Table # counts for Internet
NO YES
6 9
> round(100*.Table/sum(.Table), 2) # percentages for Internet
NO YES
40 60
> remove(.Table)
```

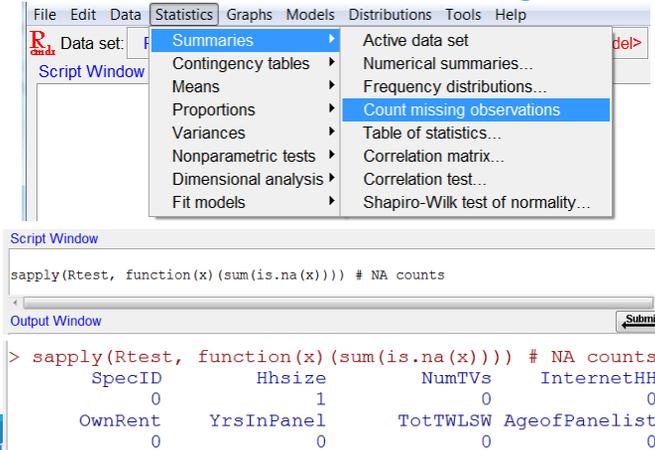
## Example 3

- Part 1: Using R commander, determine how many missing values (coded in R as NA) there are for each of the variables in the Rtest data set.
- Part 2: Using R commander, compute the mean and standard deviation for number of hours of TV viewing for the last week for sampled households.
- Part 3: Compute the average TV viewing for the last week for households that have internet and those that do not.

## Example 3, Part 1 - Solutions

Counts of missing values for every variable in the data set (NAs) can be computed by following the menu :

**Statistics** → **Summaries** → **Count of Missing Values...**



The screenshot shows the R Commander interface. The menu path **Statistics** → **Summaries** → **Count missing observations** is highlighted. Below the menu, the Script Window contains the R code: `sapply(Rtest, function(x) (sum(is.na(x)))) # NA counts`. The Output Window displays the following results:

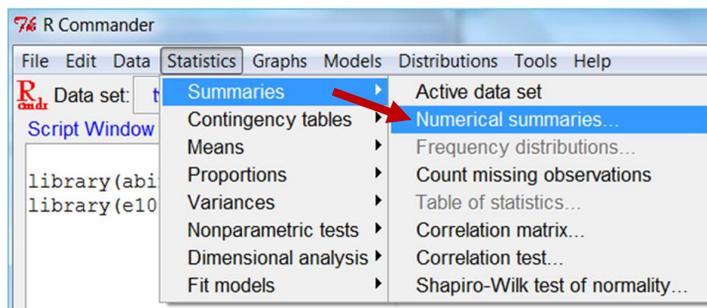
	SpecID	Hhsize	NumTVs	InternetHH
	0	1	0	0
	OwnRent	YrsInPanel	TotTWLSW	AgeofPanelist
	0	0	0	0

## Example 3, Part 2 – Solutions

Part 2: Using R commander, compute the mean and standard deviation for number of hours of TV viewing for the last week for sampled households.

Step 1: Follow the menu path:

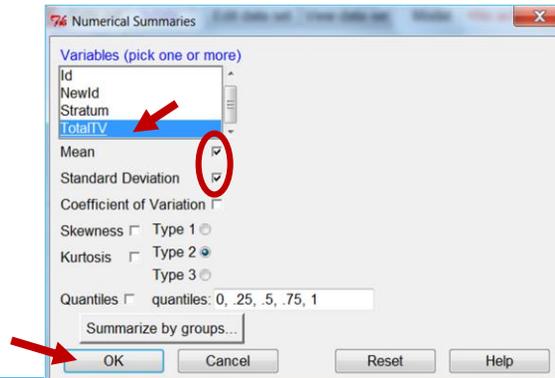
**Statistics** → **Summaries** → **Numerical Summaries**



The screenshot shows the R Commander interface. The menu path **Statistics** → **Summaries** → **Numerical summaries...** is highlighted. A red arrow points to the **Numerical summaries...** option.

## Example 3 – Solutions, Continued

- Step 2: Select the (numeric) variable(s) of interest from the Numerical Summaries Pane; Also select the appropriate options including computation of:
  - the Mean, Standard Deviation and Quartiles.



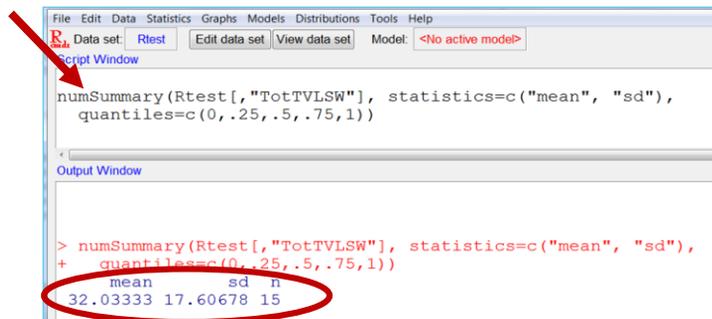
4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

39

## Example 3 – Solutions, Continued

- Notice that the corresponding R command “numSummary” appears with the selected options in the R Commander Script Pane.



- Notice also that the Mean and standard deviation appear in the R Commander Output window.

4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

40

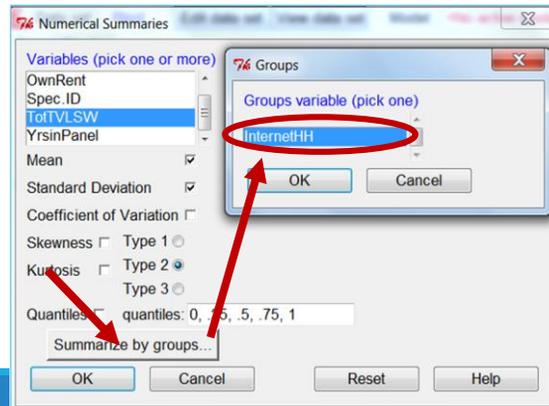
## Example 3, Part 3 – Solutions

Part 3: Compute the average TV viewing for the last week for households that have internet and those that do not.

- Follow the same path to obtain the Numerical Summaries Pane and select the numerical variable (TotTVLSW) and Summary Statistics Required. Then click on the “Summarize by Groups Tab” and then select the grouping variable...



**Caution:** If the grouping variable you need does not appear in the list, check to see that that variable was declared a factor as described in Example 2.



4/26/2016

## Example 3, Part 2 – Solutions, Continued

- Notice that the script window has an additional numSummary call with a “groups” option added

```
File Edit Data Statistics Graphs Models Distributions Tools Help
Data set: Rtest Edit data set View data set Model: <No active model>
Script Window
numSummary(Rtest[, "TotTVLSW"], groups=Rtest$InternetHH,
statistics=c("mean", "sd"), quantiles=c(0,.25,.5,.75,1))
Output Window
> numSummary(Rtest[, "TotTVLSW"], statistics=c("mean", "sd"),
+ quantiles=c(0,.25,.5,.75,1))
  mean  sd  n
32.03333 17.60678 15

> numSummary(Rtest[, "TotTVLSW"], groups=Rtest$InternetHH,
+ statistics=c("mean", "sd"), quantiles=c(0,.25,.5,.75,1))
  mean  sd data:n
NO 49.37500 13.284154 6
YES 20.47222 7.502546 9
```

4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

42

## Example 4 – Contingency Tables!

- Do a higher proportion of renting HH have internet compared to owning HHs?
- Is there a significant association between Household Ownership Status and Internet Connectivity?

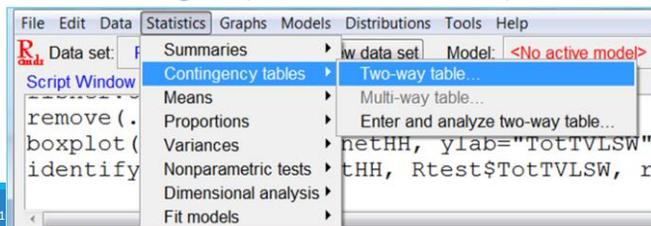
4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

43

## Example 4 - Solutions

- These questions can be answered by computing a 2-by-2 contingency table using OwnRent and InternetHH.
- We will also compute a Fisher's Exact test of association (or Chi squared test of association) between these two variables.
- In R Commander, follow the menu path:  
Statistics → Contingency tables → Two way tables...



4/26/2016

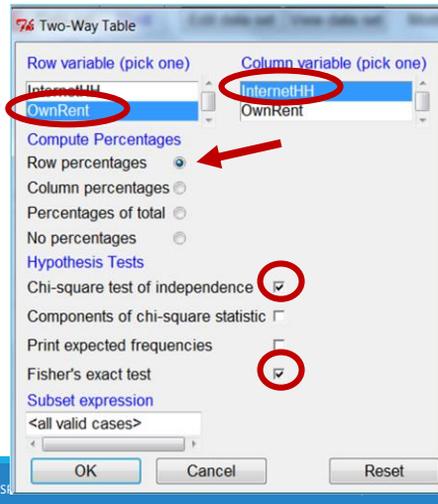
44

## Example 4 – Solutions, Continued

Step 2: Select the row and column variables as well as any percentages or statistical tests that are desired.

Step 3: Here we want to know an attribute of Owners/Renters, so I have selected this variable as the row and InternetHH as the column;

Consistent with this logic, I have also requested row percentages. Chi squared and Fischer's Exact tests have also been checked



4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR S

## Example 4, Solutions Continued

Rcommander  
Script Window  
Contents

Output Window  
Contents

```
.Table <- xtabs(~OwnRent+InternetHH,
data=Rtest)
.Table
rowPercents(.Table) # Row Percentages
.Test <- chisq.test(.Table,
correct=FALSE)
.Test
remove(.Test)
fisher.test(.Table)
remove(.Table)
```

```
> .Table <- xtabs(~OwnRent+InternetHH, data=Rtest)
> .Table
```

	InternetHH	
OwnRent	NO	YES
Own	3	7
Rent	3	2

```
> rowPercents(.Table) # Row Percentages
```

	InternetHH		
OwnRent	NO	YES	Total Count
Own	30	70	100
Rent	60	40	100

X-squared = 1.25, df = 1, p-value = 0.2636

```
> fisher.test(.Table)
```

```
Fisher's Exact Test for Count Data
data: .Table
p-value = 0.3287
alternative hypothesis: true odds ratio is not
equal to 1
95 percent confidence interval:
0.01692169 4.29385832
sample estimates:
odds ratio
0.3129331
```

SPSS USERS AAPOR WEBINAR

46

# Graphics in Rcommander

- R in general is known for its extensive graphical capabilities and language (ggplot) that provide an unprecedented level of customization (and coding).
- Rcommander has a moderate number of supported graphical procedures that are built in. These graphs can be customized with edits you can provide to the general syntax that is generated by Rcommander.
  - The color palettes can be customized for many of the plots.
  - Generally, plotting symbols, lines types and other plot features have predefined defaults built in.

# Graphics in Rcommander, Continued

col= #

Reference for Colors in R (i.e. color names/numbers and RGB's)  
<http://research.stowers.institute.org/efg/R/Color/Chart/ColorChart.pdf>

# Graphics in Rcmdr: Histograms

Rcmdr can generate histograms for any continuous variable in your dataset. Some customization is available through the Rcmdr window interface and other customization is easily available through revising the generated syntax.

Both frequency and relative frequency histograms are available.

The screenshot shows the Rcmdr interface. The 'Graphs' menu is open, and 'Histogram...' is selected. The 'Histogram' dialog box is open, showing the 'Options' sub-dialog. The 'Options' sub-dialog has two sections: 'Plot Options' and 'Plot Labels'. In the 'Plot Options' section, 'Percentages' is selected. In the 'Plot Labels' section, the x-axis label is 'Household Size'. Red circles highlight the 'Graphs' menu, the 'Histogram' dialog box, and the 'Options' sub-dialog. A red arrow points from the 'Options' sub-dialog back to the main 'Histogram' dialog box.

4/26/2016

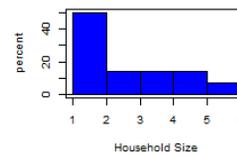
TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

49

# Histograms, Continued

R Script R Markdown

```
with(Rtest, Hist(Hhsize, scale="percent", breaks="Sturges", col="darkgray", xlab="Household Size"))  
with(Rtest, Hist(Hhsize, scale="percent", breaks="Sturges", col=3, xlab="Household Size"))  
with(Rtest, Hist(Hhsize, scale="percent", breaks="Sturges", col=4, xlab="Household Size"))
```



4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

50

# Bar graphs with Rcommander

The screenshot illustrates the steps to create a stacked bar graph in R Commander. The 'Graphs' menu is open, and the 'Bar graph...' option is selected. The 'Bar Graph' dialog box is shown with 'Style of Group Bars' set to 'Divided (stacked)' and 'Position of Legend' set to 'Left'. The 'Data Options' dialog box is also open, showing 'InternetHH' and 'OwnRent' as variables. The 'Groups' dialog box is open, showing 'OwnRent' as the 'Groups variable'. The resulting bar graph shows two bars: 'No Internet' and 'Internet'. The 'No Internet' bar has a frequency of approximately 6, and the 'Internet' bar has a frequency of approximately 10. The bars are stacked with 'Owner' (pink) at the bottom and 'Renter' (teal) on top.

# Scatterplots in Rcommander

The screenshot illustrates the steps to create a scatterplot in R Commander. The 'Graphs' menu is open, and the 'Scatterplot...' option is selected. The 'Scatterplot' dialog box is shown with 'Plot Options' set to 'Marginal boxplots' and 'Least-squares line'. The 'Data Options' dialog box is also open, showing 'count' and 'education' as variables. The 'Span for smooth' slider is set to 50. A blue callout box points to the 'Span for smooth' slider with the text: "Plots can add Least squares regression lines or a smoothed, loess line. If you choose smooth line, you can adjust the 'bandwidth' in the span slider."

## Saving Graphs: Method 1 – Inside R Console

In general, Rcmdr produces no graphical output inside its own output window. Instead you can find all graphs within the main R console window.

The screenshot shows the R GUI interface. On the left is the R console window with the following code and output:

```
> library(Rcmdr)
Loading required package: splines
Loading required package: RcmdrMisc
Loading required package: car
Loading required package: sandwich

Rcmdr Version 2.1-7

> load("C:\\Users\\Trent\\Dropbox\\Analysis Fa
XLConnect 0.2-11 by Miral Solutions GmbH [aut]
Martin Stüder [cre],
The Apache Software Foundation [ctb, cph] (A
Codec),
Stephen Colebourne [ctb, cph] (Joda-Time Jav
http://www.miral-solutions.com ,
http://mraisolutions.wordpress.com
starting httpd help server ... done
> |
```

On the right is the R Graphics: Device 2 (ACTIVE) window showing a bar chart. The x-axis is labeled 'Household Size' with values 1 through 6. The y-axis is labeled 'Percent' with values 0 through 50. A red arrow points from the R console window to the bar chart window, with a text box that says "Click on the boarder to make this panel 'Active'".

4/26/2016 TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR 53

## Saving Graphs – Method 2: Via Rcmdr

The screenshot shows the Rcmdr software interface. The 'Graphs' menu is open, and the 'Save graph to file' option is selected. The 'Save Graph as Bitmap' dialog box is open, showing the following settings:

- Graphics File Type:  JPEG
- Units:  inches
- Fixed aspect ratio (height:width):
- Width (inches): 7
- Height (inches): 7
- Resolution (pixels/inch): 71
- Text size (points): 16

Buttons: Help, Reset, OK, Cancel

4/26/2016 TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR 54

## Example 5

- Ⓜ (a) Using the Rtest dataset plot side-by-side boxplots for the ages of panelists by whether or not they live in internet households. Make the boxplots “purple”.
- Ⓜ (b) Using this same dataset, plot a scatterplot of the total television viewed in the last sweeps week by the number of years they panelist has been on the panel.
  - 📊 For this scatterplot be sure the to include marginal boxplots
  - 📊 Make the plotted points sized at 1.8
  - 📊 Add a title to the plot ‘TV viewing by Years in Panel’

4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

55

## Example 3.1(d) Solutions

- Ⓜ Follow this menu sequence/steps:
  - 📊 Graphs → Boxplot...
  - 📊 Then select “AgeofPanlist” in the Variable Window
  - 📊 Then click “Plot by Goups” and select InternetHH, then OK.
  - 📊 In the Rcommander script window, add, **col='purple'** inside the Boxplot syntax. Highlight the entire statement and click submit.

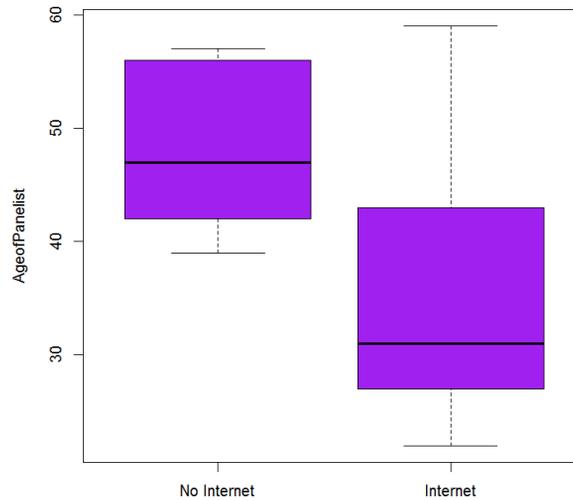
The screenshot shows the R Commander interface. On the left, the 'Boxplot' dialog box is open, with 'AgeofPanelist' selected in the 'Variable (pick one)' list and 'InternetHH' selected in the 'Groups variable (pick one)' list. A red arrow points from the 'AgeofPanelist' variable to the 'Groups' dialog box. In the background, the R script window shows the command: `Boxplot(AgeofPanelist~InternetHH, data=Rtest, id.method="none")`. A red arrow points from the 'Submit' button in the 'Boxplot' dialog to the 'col='purple'' text in the script window, which is highlighted with a pink box.

4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

56

## Example 5(a) Solutions Continued



4/26/2016

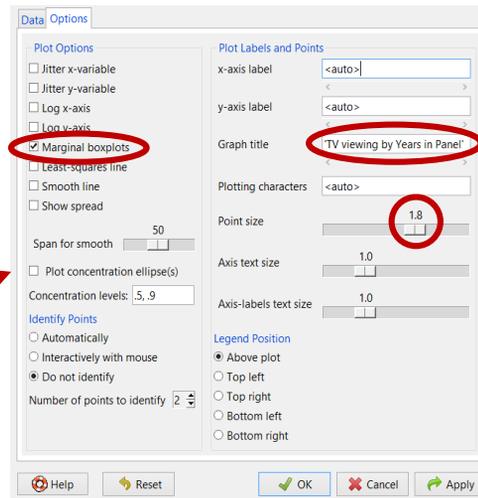
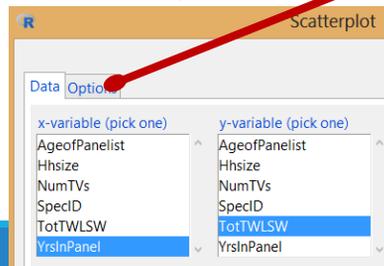
TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

57

## Example 5 (b) Solutions

Follow this menu sequence/steps:

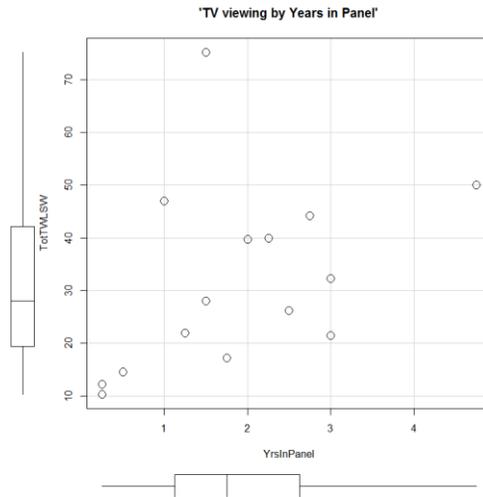
- 1. Graphs → Scatterplot...
- 2. Then select "YrsInPanel" for x
- 3. Select "TotTWSW" for y
- 4. The select options to select marginal boxplots on the left, enter the title 'TV viewing by Years in Panel' and change point size to 1.8 (both on the right).



PH.D. R FOR SPSS USERS AAPOR WEBINAR

58

## Example 5(b) solutions Cont.



4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

59

## Thank You!



[TBuskirk@m-s-g.com](mailto:TBuskirk@m-s-g.com)  
[tdbuskirk@gmail.com](mailto:tdbuskirk@gmail.com)



314-695-1378



[www.m-s-g.com](http://www.m-s-g.com)

4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

60

# Appendix

---

---

- 📄 A0 – Installing R
- 📄 A1 – Take home example for variable conversion and basic frequency tables.
- 📄 A2 – Take home example for frequency distributions
- 📄 A3 – Graphical plotting symbols and lines reference
- 📄 A4 – Other R packages for Descriptive Statistics
- 📄 A5 – Linear Regression using Rcommander

# A0: Installing R

---

---



 <http://cran.r-project.org/bin/windows/base/>

❖ Current version as of June 2014:

❖ <http://cran.r-project.org/bin/windows/base/>

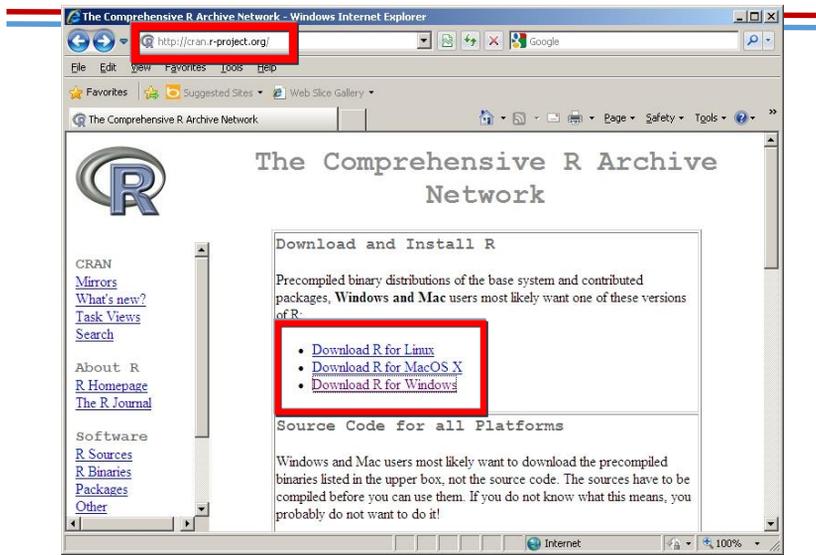


 <http://rstudio.org/download/desktop>

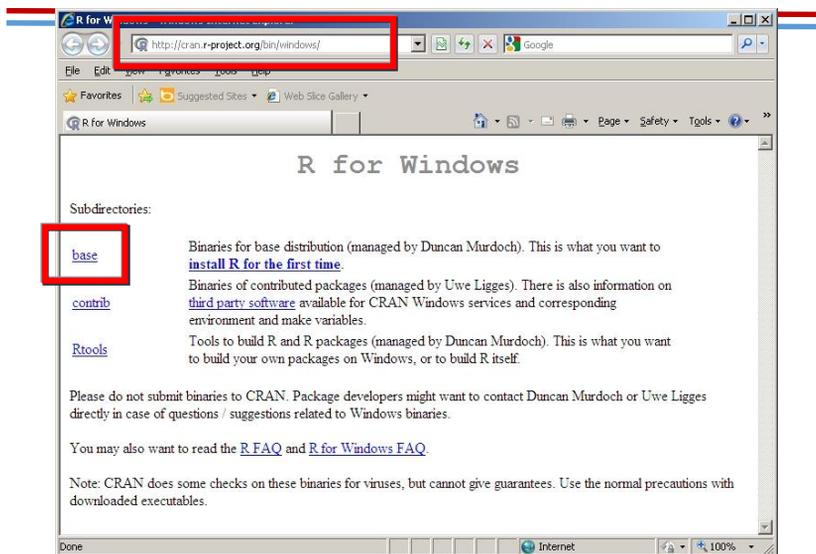
❖ Current version as of June 2012:

❖ <http://download1.rstudio.org/RStudio-0.96.231.exe>

# A0: Download R



# A0: Download R



## A0: Download R <http://cran.r-project.org/bin/windows/base/>

R-3.2.5 for Windows (32/64 bit)

[Download R 3.2.5 for Windows](#) (62 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)

If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the [md5sum](#) of the .exe to the [true fingerprint](#). You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

### Frequently asked questions

- [How do I install R when using Windows Vista?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

## A0: Install R



## A1– Basic Recoding and Frequencies



- Convert the OwnRent variable into a factor with the levels 1 corresponding to OWN and 2 to RENT.
- What proportion of sampled households are renters?

## Appendix A1– Demo Video

Roll over the Graphical Image at the bottom and then press the play triangle in lower left

```
R Commander
File Edit Data Statistics Graphs Models Distributions Tools Help
Data set: Rtest Edit data set View data set Model: <No active model>
Script Window
Output Window Submit
OWN RENT
10 5
> round(100*.Table/sum(.Table), 2) # percentages for OwnRent
OWN RENT
66.67 33.33
> remove(.Table)
> .Workbook <-
+ loadWorkbook("C:/Users/Trent/Documents/AnalysisFactorRtrain
> Rtest <- readWorksheet(.Workbook, "Sheet1")
> remove(.Workbook)
Messages
[10] NOTE: The dataset Rtest has 15 rows and 8 columns.
[11] NOTE: The dataset Rtest has 15 rows and 8 columns.
```

## Appendix 2: Working more with Frequency Distributions



- Are there any outliers in the Number of TVs per household? **Don't cheat** – use an interactive boxplot here.
- What is the 75<sup>th</sup> Percentile (i.e. 3<sup>rd</sup> quartile) for the number of TVs owned?
- Which type of household – owner or renter – has a longer tenure on the panel, on average?

4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

69

## A2– Video Solutions

- Are there any outliers in the Number of TVs per household

```
h many contributors.  
nformation and  
R packages in publications.  
elp()' for on-line help, or  
er interface to help.  
  
.2 (64-bit): an enhanced distribution  
pyright (C) 2013 Revolution Analytics  
utionanalytics.com for the latest  
the community forum, or 'readme()'  
  
ored)  
  
e
```

Roll over the  
Graphical Image  
at the bottom  
and then press  
the play triangle  
in lower left

A screenshot of the R Commander software interface. The window title is "R Commander". The menu bar includes File, Edit, Data, Statistics, Graphs, Models, Distributions, Tools, and Help. Below the menu bar are buttons for "Data set:", "Rtest", "Edit data set", "View data set", and "Model: <No active model>". The "Script Window" contains the following R code:

```
> round(100*.Table/sum(.Table), 2) # percentages for OwnRent  
OWN RENT  
66.67 33.33  
  
> remove(.Table)  
  
> sapply(Rtest, function(x) (sum(is.na(x)))) # NA counts  
SpecID Hssize NumTVs InternetHH  
0 1 0  
OwnRent YrsInPanel TotTWLSW AgeofPanelist  
0 0 0 0  
  
> Boxplot(~ NumTVs, data=Rtest, id.method="y")  
[1] "14"  
  
> Boxplot(~ NumTVs, data=Rtest, id.method="identify")
```

The "Output Window" shows the results of the first command:

```
OWN RENT  
66.67 33.33
```

The "Messages" window at the bottom shows two notes:

```
[11] NOTE: The dataset Rtest has 15 rows and 8 columns.  
[12] NOTE: The dataset Rtest has 15 rows and 8 columns.
```

4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

70

## A2, Solutions Cont.

- What is the 75<sup>th</sup> Percentile (i.e. 3<sup>rd</sup> quartile) for the number of TVs owned?

Roll over the Graphical Image at the bottom and then press the play triangle in lower left

```

R Commander
File Edit Data Statistics Graphs Models Distributions Tools Help
Data set: Rtest Edit data set View data set Model: <No active model>
Script Window
numSummary(Rtest[, "NumTVs"], statistics=c("mean", "sd",
"quantiles"), quantiles=c(.75))
Output Window
> library(abind, pos=4)
> library(e1071, pos=4)
> numSummary(Rtest[, "NumTVs"], statistics=c("mean", "sd",
+ "IQR", "quantiles"), quantiles=c(.75))
  mean      sd IQR  n
3.333333 3.754363  3 15
> numSummary(Rtest[, "NumTVs"], statistics=c("mean", "sd",
+ "quantiles"), quantiles=c(.25,.50,.75))
  mean      sd 25% 50% 75%  n
3.333333 3.754363  1  2  4 15
> numSummary(Rtest[, "NumTVs"], statistics=c("mean", "sd",
+ "quantiles"), quantiles=c(.75))
  mean      sd  n
3.333333 3.754363 15
Messages
[7] NOTE: The dataset Dataset has 15 rows and 8 columns.
[8] NOTE: The dataset Dataset has 15 rows and 8 columns.
[9] NOTE: The dataset Rtest has 15 rows and 8 columns.
    
```

## A2, Cont.

- Which type of household – owner or renter – has a longer tenure on the panel, on average?

Roll over the Graphical Image at the bottom and then press the play triangle in lower left

```

R Commander
File Edit Data Statistics Graphs Models Distributions Tools Help
Data set: Rtest Edit data set View data set Model: <No active model>
Script Window
conf.level=.95, var.equal=FALSE, data=Rtest)
Output Window
  mean      sd 25% 50% 75%  n
3.333333 3.754363  1  2  4 15
> t.test(YrsInPanel~OwnRent, alternative='two.sided',
+ conf.level=.95, var.equal=FALSE, data=Rtest)

Welch Two Sample t-test

data:  YrsInPanel by OwnRent
t = -0.4024, df = 5.587, p-value = 0.7023
alternative hypothesis: true difference in means is not equal to
95 percent confidence interval:
-2.337217  1.687217
sample estimates:
mean in group OWN mean in group RENT
      1.775          2.100
Messages
[7] NOTE: The dataset Dataset has 15 rows and 8 columns.
[8] NOTE: The dataset Dataset has 15 rows and 8 columns.
[9] NOTE: The dataset Rtest has 15 rows and 8 columns.
    
```

## A2, Cont.

- Which type of household – owner or renter – has a longer tenure on the panel, on average?

Roll over the Graphical Image at the bottom and then press the play triangle in lower left

The screenshot shows the R Commander interface. The Script Window contains the following code:

```
numSummary(Rtest[, "YrsInPanel"], groups=Rtest$OwnRent,
```

The Output Window displays the results of a Two Sample t-test:

```
+ conf.level=.95, var.equal=TRUE, data=Rtest)

Two Sample t-test

data: YrsInPanel by OwnRent
t = -0.4739, df = 13, p-value = 0.6434
alternative hypothesis: true difference in means is not equal to
95 percent confidence interval:
-1.806604 1.156604
sample estimates:
mean in group OWN mean in group RENT
1.775 2.100

> tapply(Rtest$YrsInPanel, list(OwnRent=Rtest$OwnRent), mean, na.
```

A red box highlights the following table in the Output Window:

OwnRent	OWN	RENT
	1.775	2.100

The Messages window at the bottom shows several notes about the dataset structure.

## A2 Cont.

The screenshot shows the R Commander interface. The Output Window displays the following code and results:

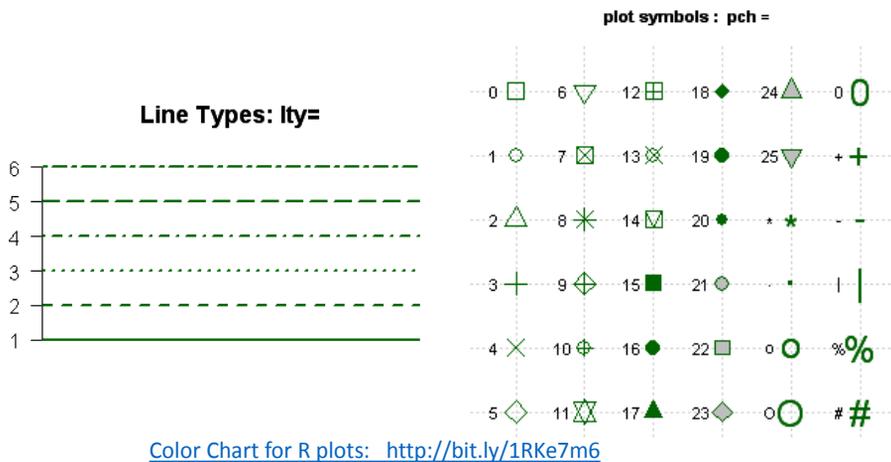
```
> library(abind, pos=4)
> library(e1071, pos=4)

> numSummary(Rtest[, "NumTVs"], statistics=c("mean", "sd", "quantiles"),
+ quantiles=c(0,.25,.5,.75,1))
  mean      sd 0% 25% 50% 75% 100%  n
3.333333 3.754363 1 1 2 4 16 15

> numSummary(Rtest[, "YrsinPanel"], groups=Rtest$OwnRent, statistics=c("mean",
+ "sd", "quantiles"), quantiles=c(0,.25,.5,.75,1))
  mean      sd 0% 25% 50% 75% 100% data:n
Own 1.775 1.023678 0.25 1.0625 1.75 2.6875 3.00 10
Rent 2.100 1.654539 0.25 1.5000 1.75 2.2500 4.75 5
```

A red box highlights the 'mean' column in the second summary table. Another red box highlights the '75%' column in the first summary table. To the right, a boxplot for 'NumTVs' is shown, with a y-axis labeled 'NumTVs' ranging from 0 to 15. The boxplot shows a median around 3, with whiskers extending from approximately 1 to 5, and an outlier at 14.

## A3: Basic plotting symbols in R



4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

75

## A4: Some Other R packages for Descriptive Statistics

Ⓜ There are hundreds of R packages that extend the analytic and graphics capabilities of R.

Ⓜ These packages are like “mini” SPSS modules.

📊 Some popular examples include:

- lessR – streamlined and simplified syntax for General Data Analysis
- Boot – bootstrap statistics
- NLME – nonlinear models
- RandomForest – random forest classification models
- Lattice – scatterplot matrix and other graphical enhancements
- MatchIt – statistical matching models
- Ggplot 2 – enhanced graphics capabilities/forms
- Survey – design and analysis of survey samples
- PASWR – Basic statistics package that contains functions for nonparametric analysis and Exploratory Data Analysis.

4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

76

## A4: Additional Examples of Packages focused on Descriptive Statistics

• The **Hmisc** package:  
>library(Hmisc)  
>describe(mydata)  
**REQUESTS:** n, nmiss, unique, mean, 5, 10, 25, 50, 75, 90, 95th percentiles, 5 lowest and 5 highest scores

The **pastecs** package:  
>library(pastecs)  
>stat.desc(mydata)  
**REQUESTS:** nbr.val, nbr.null, nbr.na, min max, range, sum, median, mean, SE.mean, CI.mean, var, std.dev, coef.var

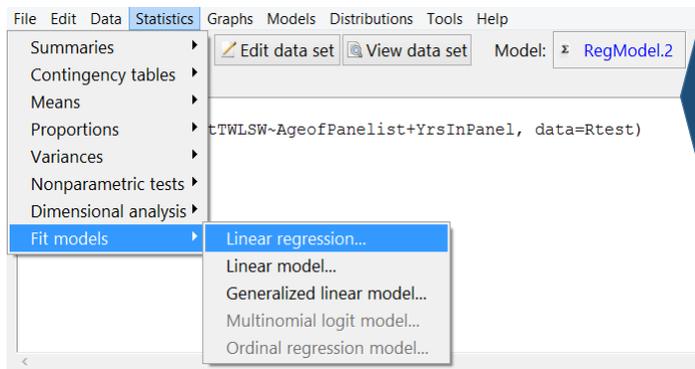
The **psych** package:  
>library(psych)  
>describe(mydata)  
>describe.by(mydata, grouping variable)  
**REQUESTS:** item name, item number, nvalid, mean, sd, median, mad, min, max, skew, kurtosis, se  
(Summary statistics as a whole or by the grouping variable)

4/26/2016

77

## A5: Regression Modeling in Rcommander

Step 1: Select the type of model under the Statistics Menu: **Statistics Menu** → **Fit Model** → **Linear Regression**



**R names models as Objects and stores model results in these objects**

4/26/2016

TRENT D. BUSKIRK, PH.D. R FOR SPSS USERS AAPOR WEBINAR

78

# A5: Linear Regression in Rcommander

Step 2: Select the form of the model – predictors and outcome variables

Linear Regression dialog box showing the selection of response and explanatory variables. The response variable is 'OwnRent' and the explanatory variables are 'SpecID', 'TotTWLSW', and 'YrsInPanel'. A callout box indicates: "For Multiple Linear Regression, select multiple predictors using the 'CTRL' key".

# A5: Linear Regression in Rcommander

Step 3: Summarize model and explore model fit diagnostics

Rcommander interface showing the 'Models' menu and the 'Select Model' dialog box. The 'Models' menu is open, and 'Select Model' is highlighted. The 'Select Model' dialog box shows 'Current Model: RegModel.3' and a list of models including 'RegModel.2', 'RegModel.3', 'resp.lrf', and 'resp.lrf2'.

# A5: Linear Regression in Rcommander

Summarize Function presents overall model fit and coefficient estimates table

Provides choices for model fit and summary plots

The screenshot shows the Rcommander 'Models' menu. The 'Summarize model' option is highlighted, which corresponds to the first callout. The 'Hypothesis tests' and 'Graphs' options are also visible, corresponding to the other callouts. The background shows a regression model named 'RegModel.3' with a coefficient of  $\alpha$ .

# A5: R, Rcommander Regression Output

Script Window

```
RegModel.2 <- lm(TotTWLSW~AgeofPanelist+YrsInPanel, data=Rtest)
summary(RegModel.2)
anova(RegModel.2)
```

Output Window

Call:  
lm(formula = TotTWLSW ~ AgeofPanelist + YrsInPanel, data = Rtest)

Residuals:

Min	1Q	Median	3Q	Max
-20.075	-4.771	-3.231	7.059	23.586

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-12.2067	12.2485	-0.997	0.33864
AgeofPanelist	1.2183	0.3878	3.141	0.00851
YrsInPanel	-2.9021	3.9846	-0.728	0.48038

Residual standard error: 12.85 on 12 degrees of freedom  
Multiple R-squared: 0.5436, Adjusted R-squared: 0.4675  
F-statistic: 7.146 on 2 and 12 DF, p-value: 0.00904

Analysis of Variance Table

Response: TotTWLSW

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
AgeofPanelist	1	2271.58	2271.58	13.7613	0.002983
YrsInPanel	1	87.57	87.57	0.5305	0.480378
Residuals	12	1980.84	165.07		