

Introduction to R

Drew Schmidt

April 2, 2015

Contents

- What is R?
- R Extensions
- R Interfaces
- R Resources

What is R?

What is R?

- Powerful, free, open source.
- Part programming language, part data analysis package.
- An implementation of the S language.
- Written by Ross Ihaka & Robert Gentleman.
- SAS/SPSS have <15 add-ons, R has over 6400!
- Most popular environment for developing new analysis and vis methods.

The SAS/SPSS Way

1. Data input and management
2. Stats and graphics procedures
3. Output management systems
4. Macro language
5. Matrix language

The R Way

All of these concepts are built into the (single) core language.

Comparing R to Other Tools

- R can be harder to “get going” than other analysis tools.
- But once you do...
- Hastily becoming the “lingua franca”.

R Extensions

Extensions

- Community and commercial addons.

- Everything from sophisticated analyses to making your computer beep.
- Written in R, Fortran, C, C++ mostly.
- Performance can sometimes be an issue, but generally accurate.
- Package download trends: <https://dgrtwo.shinyapps.io/cranview/>

The CRAN

- Comprehensive R Archive Network.
- Open source only.
- Over 6400 packages.
- High standard of quality is vigorously enforced.
- Closest thing to a journal for stats software.

```
install.packages("NameOfPackage")
```

GitHub

- Many authors eschewing CRAN for GitHub.
- Faster bug fixes, faster bugs (maybe).

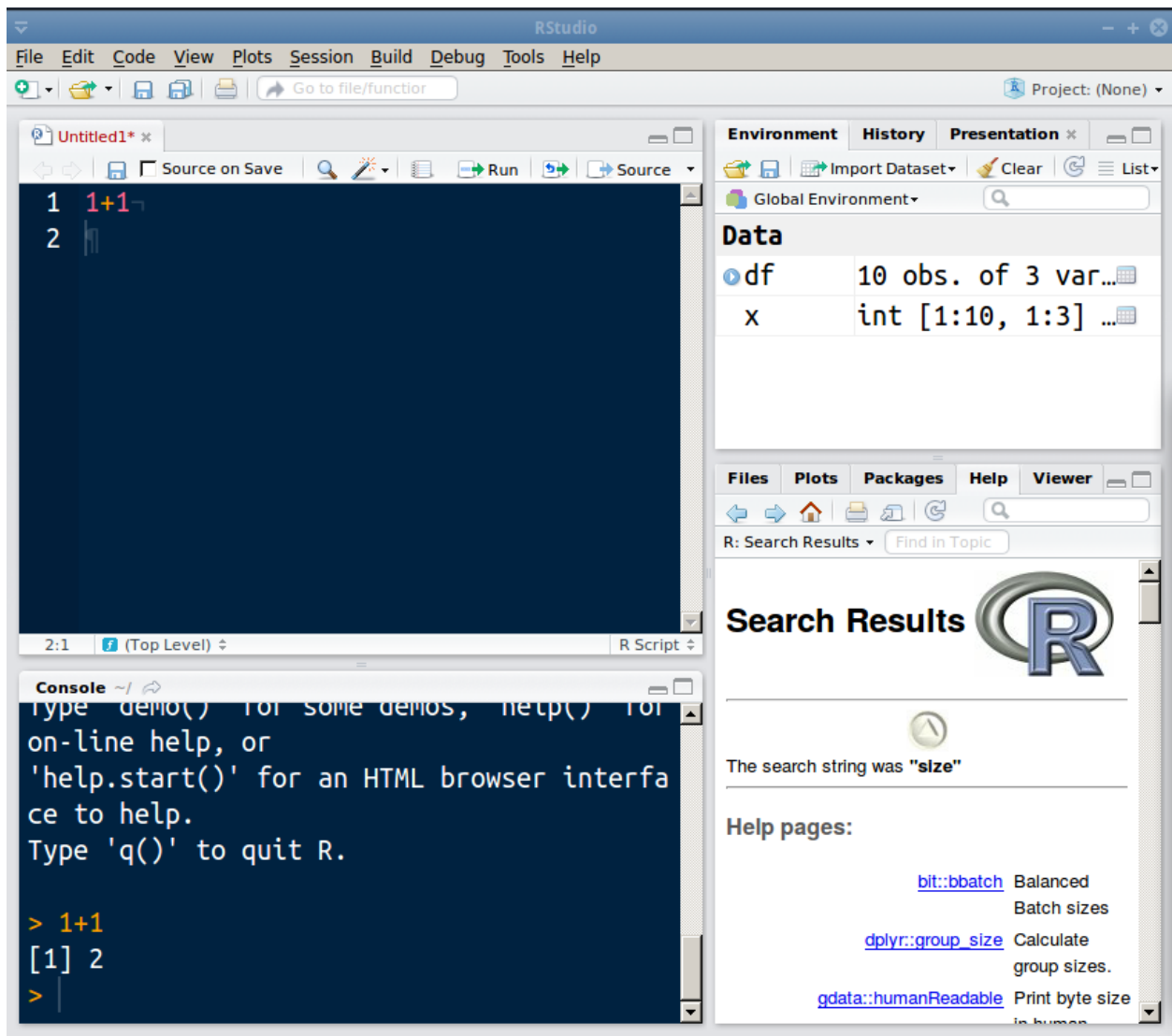
```
install.packages("devtools")
library(devtools)
install_github("GitHubUserName/RepositoryName")
```

R Interfaces

R Interfaces

1. RStudio
2. R Commander
3. Deducer
4. Rattle
5. R from Excel
6. Shiny

RStudio



R Commander

The screenshot shows the R Commander application window. The menu bar includes File, Edit, Data, Statistics, Graphs, Models, Distributions, Tools, and Help. The Statistics menu is open, showing options like Summaries, Contingency tables, Means, Proportions, Variances, Nonparametric tests, Dimensional analysis, and Fit models. The Means submenu is also open, listing Single-sample t-test..., Independent samples t-test..., Paired t-test..., One-way ANOVA..., and Multi-way ANOVA... The Data set is 'mydata' and the Model is '<No active model>'. The Script Window shows the commands 'showData(mydata,)' and 'summary(mydata)'. The Output Window displays the summary statistics for the 'mydata' dataset, and the Messages window shows a note about the dataset's dimensions.

R Commander

File Edit Data **Statistics** Graphs Models Distributions Tools Help

Data set: **mydata** Model: **<No active model>**

Statistics

- Summaries
- Contingency tables
- Means**
 - Single-sample t-test...
 - Independent samples t-test...
 - Paired t-test...
 - One-way ANOVA...
 - Multi-way ANOVA...
- Proportions
- Variances
- Nonparametric tests
- Dimensional analysis
- Fit models

Script Window

```
showData(mydata,)  
summary(mydata)
```

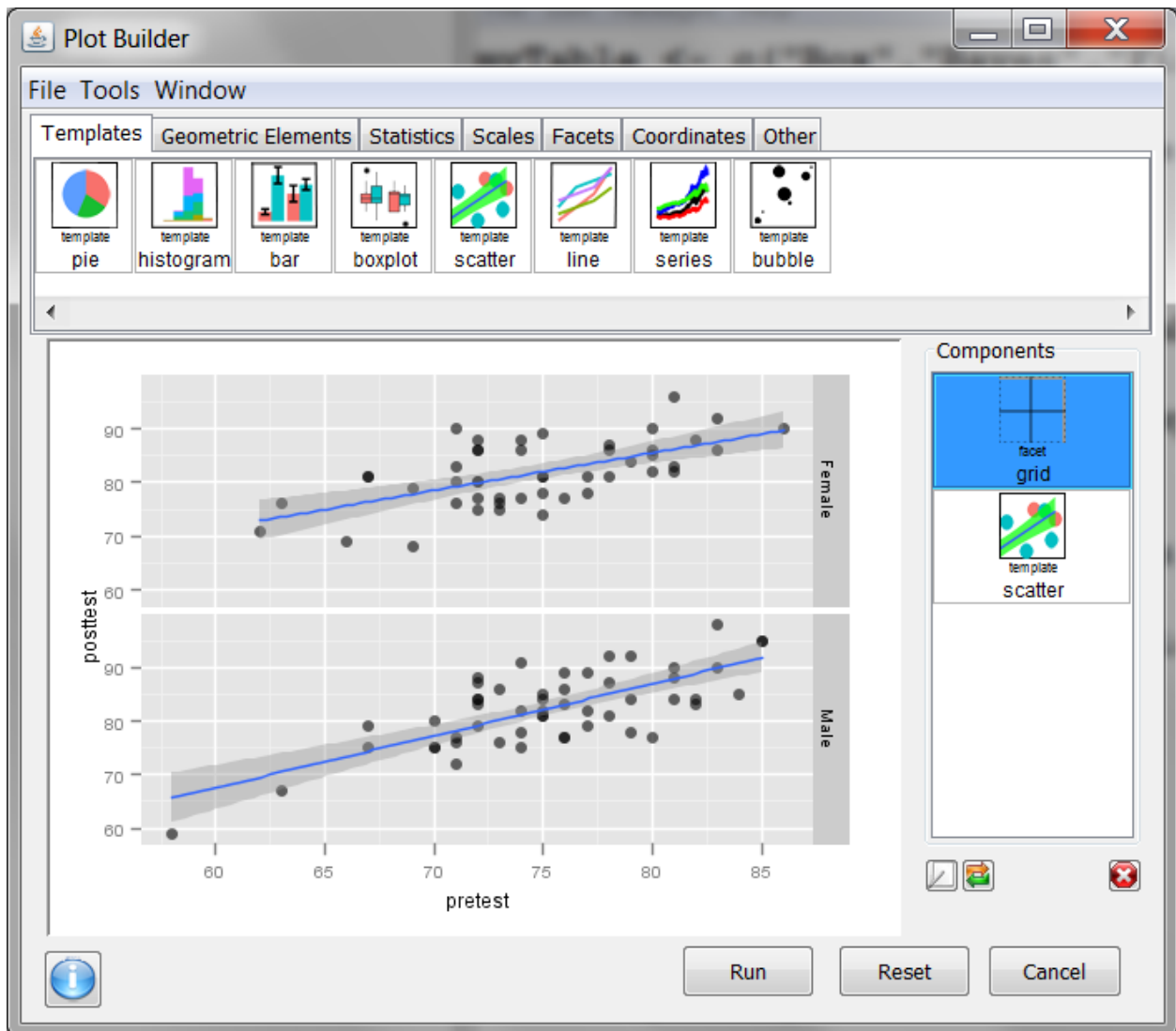
Output Window

```
> summary(mydata)  
  workshop  gender    q1      q2      q3  
Min.   :1.0    f   :3  Min.   :1.00  Min.   :1.00  Min.   :2.000  
1st Qu.:1.0    m   :4  1st Qu.:2.00  1st Qu.:1.00  1st Qu.:4.000  
Median :1.5    NA's:1  Median :3.50  Median :2.50  Median :4.000  
Mean   :1.5                Mean   :3.25  Mean   :2.75  Mean   :4.143  
3rd Qu.:2.0                3rd Qu.:4.25  3rd Qu.:4.25  3rd Qu.:5.000  
Max.   :2.0                Max.   :5.00  Max.   :5.00  Max.   :5.000  
                        NA's   :1.000
```

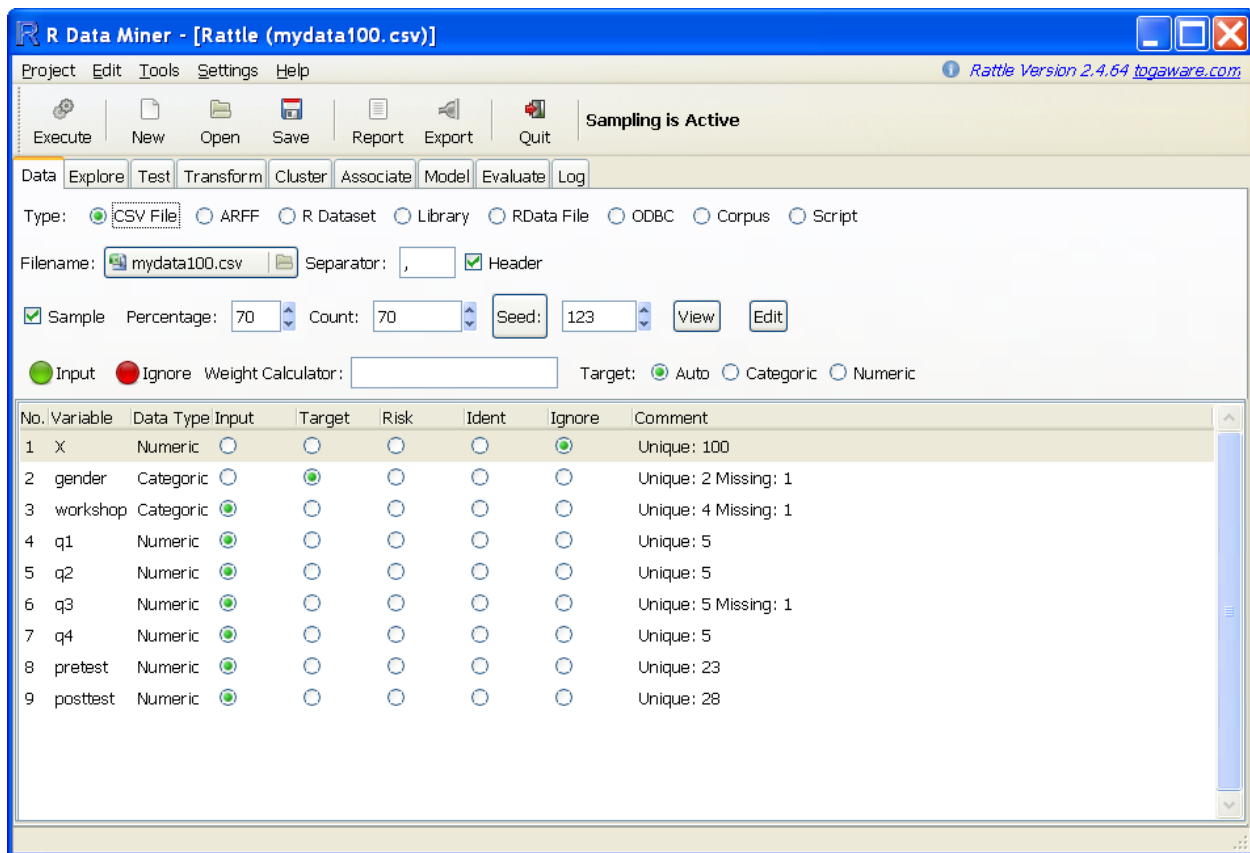
Messages

NOTE: The dataset mydata has 8 rows and 6 columns.

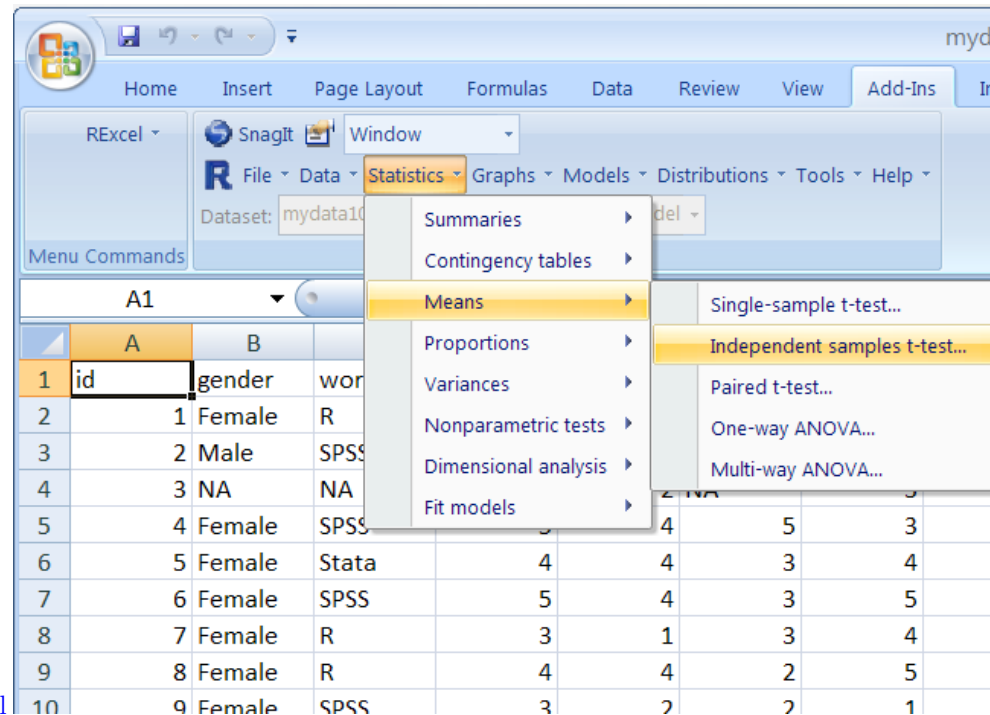
Deducer



Rattle



R from Excel



<http://rcom.univie.ac.at/download.html>

Shiny

- System for creating your own gui's (webapp)
- Code is 100% R (but can use html/css/javascript/...)
- Community quickly getting flooded with webapps

Shiny: Plot Your Data

[Live Source code](#)

Plot Your Data

Hello User!

Select Inputs

How To

Choose file to upload

Browse...

iris.csv

Upload complete

☐ Use Example Data

x variable:

Sepal.Length

Include x range:

4.3

7.9

y variable:

Sepal.Width

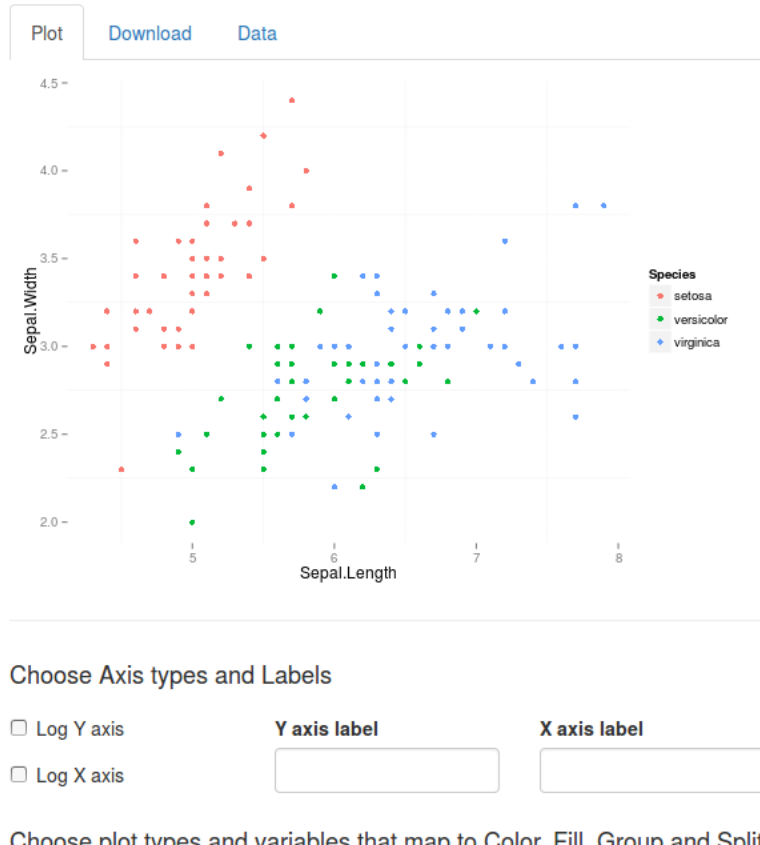
Include y range:

2

4.4

Max number of unique values for filter variables:

250



Shiny: Radiant

[Live Source code](#)

Radiant · Marketing
Data
Sample
Base
Regression
Maps
Factor
Cluster
Conjoint
R
Quit
Help

Datasets:

diamonds

☐ Add/edit data description
☐ Rename data

Load data:

☒ rda
☐ csv
☐ clipboard

☐ examples
☐ state

Browse...
No files selected.

Save data:

☒ rda
☐ csv
☐ clipboard

☐ state

Save

☐ Remove data from memory

Manage
View
Visualize
Pivot
Explore
Transform
Merge

price	carat	clarity	cut	color	depth	table	x	y	z
580	0.32	VS1	Ideal	H	61.00	56.00	4.43	4.45	2.71
650	0.34	SI1	Very Good	G	63.40	57.00	4.45	4.42	2.81
630	0.30	VS2	Very Good	G	63.10	58.00	4.27	4.23	2.68
706	0.35	VVS2	Ideal	H	59.20	56.00	4.60	4.65	2.74
1080	0.40	VS2	Premium	F	62.60	58.00	4.72	4.68	2.94
3082	0.60	VVS1	Ideal	E	62.50	53.70	5.35	5.43	3.38
3328	0.88	SI1	Ideal	I	61.70	56.00	6.14	6.18	3.80
4229	0.93	SI1	Premium	E	61.40	57.00	6.34	6.23	3.86
1895	0.51	VVS2	Very Good	G	63.40	57.00	5.09	5.06	3.22
3546	1.01	SI2	Good	E	63.90	58.00	6.31	6.37	4.05

10 (max) rows shown. See View-tab for details.

Diamond prices

Prices of 3,000 round cut diamonds

Description

A dataset containing the prices and other attributes of a sample of 3000 diamonds. The variables are as follows:

Variables

- price = price in US dollars (\$338–\$18,791)
- carat = weight of the diamond (0.2–3.00)

R Resources

Books

- *The Art of R Programming*, Norm Matloff
- *R for SAS and SPSS Users*, Muenchen
- *R for Stata Users*, Muenchen & Hilbe
- [An Introduction to R](#)
- [The R Inferno](#)

Comparisons to Other Languages

- [aRrgh: a newcomer's \(angry\) guide to R](#)
- [Mathesaurus](#)
- [R programming for those coming from other languages](#)

Online Help

- [r] tag on stackoverflow
- The mailing lists (r-help)
- #rstats on twitter

Questions?

This presentation is available at github.com/wrathematics/2015SFSURworkshop

Exercises are also available there.