

# R for absolute beginners

## Duncan Golicher



11/19/2008

# Motivation for the course

1. Encourage researchers and students to begin using R
2. Draw on personal experience to flatten the learning curve
3. Provide practical examples of the advantages of R over GUI based statistical packages such as SPSS
4. Provide the basic tools of the R language that allow experimentation and self teaching

# Motivation for the course

1. Encourage researchers and students to begin using R
2. Draw on personal experience to flatten the learning curve
3. Provide practical examples of the advantages of R over GUI based statistical packages such as SPSS
4. Provide the basic tools of the R language that allow experimentation and self teaching

# Motivation for the course

1. Encourage researchers and students to begin using R
2. Draw on personal experience to flatten the learning curve
3. Provide practical examples of the advantages of R over GUI based statistical packages such as SPSS
4. Provide the basic tools of the R language that allow experimentation and self teaching

# Motivation for the course

1. Encourage researchers and students to begin using R
2. Draw on personal experience to flatten the learning curve
3. Provide practical examples of the advantages of R over GUI based statistical packages such as SPSS
4. Provide the basic tools of the R language that allow experimentation and self teaching

# Motivation for the course

1. Encourage researchers and students to begin using R
2. Draw on personal experience to flatten the learning curve
3. Provide practical examples of the advantages of R over GUI based statistical packages such as SPSS
4. Provide the basic tools of the R language that allow experimentation and self teaching

# What can R do for me?

1. Increase your productivity. For example this document includes output from R code automatically.
2. Improve your understanding of statistical procedures. R allows you to test assumptions through simulation.
3. Provide a framework for very advanced data analysis. R can do just about anything.
4. Provide an enjoyable intellectual challenge. R is fun!

# What can R do for me?

1. Increase your productivity. For example this document includes output from R code automatically.
2. Improve your understanding of statistical procedures. R allows you to test assumptions through simulation.
3. Provide a framework for very advanced data analysis. R can do just about anything.
4. Provide an enjoyable intellectual challenge. R is fun!



# What can R do for me?

1. Increase your productivity. For example this document includes output from R code automatically.
2. Improve your understanding of statistical procedures. R allows you to test assumptions through simulation.
3. Provide a framework for very advanced data analysis. R can do just about anything.
4. Provide an enjoyable intellectual challenge. R is fun!

# What can R do for me?

1. Increase your productivity. For example this document includes output from R code automatically.
2. Improve your understanding of statistical procedures. R allows you to test assumptions through simulation.
3. Provide a framework for very advanced data analysis. R can do just about anything.
4. Provide an enjoyable intellectual challenge. R is fun!

# What can R do for me?

1. Increase your productivity. For example this document includes output from R code automatically.
2. Improve your understanding of statistical procedures. R allows you to test assumptions through simulation.
3. Provide a framework for very advanced data analysis. R can do just about anything.
4. Provide an enjoyable intellectual challenge. R is fun!

## What can R do for me?

The text is produced by the code below that looks up quotations in the "fortunes" package.

```
> library(fortunes)
> fortune("pizza")
```

Roger D. Peng: I don't think anyone actually believes that R is designed to make *everyone* happy. For me, R does about 99% of the things I need to do, but sadly, when I need to order a pizza, I still have to pick up the telephone.

Douglas Bates: There are several chains of pizzerias in the U.S. that provide for Internet-based ordering (e.g. [www.papajohnsonline.com](http://www.papajohnsonline.com)) so, with the Internet modules in R, it's only a matter of time before you will have a pizza-ordering function available.

Brian D. Ripley: Indeed, the GraphApp toolkit (used for the RGui interface under R for Windows, but Guido forgot to include it) provides one (for use in Sydney, Australia, we presume as that is where the GraphApp author hails from). Alternatively, a Padovian has no need of ordering pizzas with both home and neighbourhood restaurants ....

```
-- Roger D. Peng, Douglas Bates and Brian D. Ripley
R-help (June 2004)
```

## What can R do for me? R can be fun.

```
> library(sudoku)
> puz <- fetchSudokuUK()
> puz
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	6	1	0	0	4	0	0	0	0
[2,]	4	0	0	0	0	3	0	0	8
[3,]	0	0	0	5	0	0	4	0	0
[4,]	0	0	1	2	9	0	0	0	0
[5,]	8	2	0	0	0	0	0	7	6
[6,]	0	0	0	0	5	8	9	0	0
[7,]	0	0	3	7	0	6	0	0	0
[8,]	1	0	0	4	0	0	0	0	7
[9,]	0	0	0	0	1	0	0	8	9

## What can R do for me? R can cheat!

```
> solveSudoku(puz)
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	6	1	7	8	4	2	3	9	5
[2,]	4	5	2	9	6	3	7	1	8
[3,]	9	3	8	5	7	1	4	6	2
[4,]	5	6	1	2	9	7	8	4	3
[5,]	8	2	9	1	3	4	5	7	6
[6,]	3	7	4	6	5	8	9	2	1
[7,]	2	9	3	7	8	6	1	5	4
[8,]	1	8	5	4	2	9	6	3	7
[9,]	7	4	6	3	1	5	2	8	9

## What can R do for me? Just about anything

```
> fortune("Yoda")
```

```
Evelyn Hall: I would like to know how (if) I can extract  
some of the information from the summary of my nlme.
```

```
Simon Blomberg: This is R. There is no if. Only how.
```

```
-- Evelyn Hall and Simon 'Yoda' Blomberg
```

```
R-help (April 2005)
```

## What can R do for me? More than SAS

```
> fortune(10)
```

Overall, SAS is about 11 years behind R and S-Plus in statistical capabilities (last year it was about 10 years behind) in my estimation.

-- Frank Harrell (SAS User, 1969-1991)

R-help (September 2003)



## What can R do for me? Ensure excellence

```
> fortune(120)
```

```
Rene M. Raupp: Does anybody know any work comparing R with  
other (charged) statistical software (like Minitab, SPSS,  
SAS)? [...] I have to show it's as good as the others.
```

```
Kjetil Brinchmann Halvorsen: Sorry. That will be difficult.  
Couldn't it do to prove it is better?
```

```
-- Rene M. Raupp and Kjetil Brinchmann Halvorsen  
R-help (May 2005)
```

# What is R?

1. R is all things to all people.
2. System for statistical computation and graphics
3. Computer language
4. Interface to system functions
5. Runs commands or scripts
6. So: R is a computer language + a set of procedures that have been implemented to do statistical analysis and more.

# What is R?

1. R is all things to all people.
2. System for statistical computation and graphics
3. Computer language
4. Interface to system functions
5. Runs commands or scripts
6. So: R is a computer language + a set of procedures that have been implemented to do statistical analysis and more.

# What is R?

1. R is all things to all people.
2. System for statistical computation and graphics
3. Computer language
4. Interface to system functions
5. Runs commands or scripts
6. So: R is a computer language + a set of procedures that have been implemented to do statistical analysis and more.

# What is R?

1. R is all things to all people.
2. System for statistical computation and graphics
3. Computer language
4. Interface to system functions
5. Runs commands or scripts
6. So: R is a computer language + a set of procedures that have been implemented to do statistical analysis and more.

# What is R?

1. R is all things to all people.
2. System for statistical computation and graphics
3. Computer language
4. Interface to system functions
5. Runs commands or scripts
6. So: R is a computer language + a set of procedures that have been implemented to do statistical analysis and more.

# What is R?

1. R is all things to all people.
2. System for statistical computation and graphics
3. Computer language
4. Interface to system functions
5. Runs commands or scripts
6. So: R is a computer language + a set of procedures that have been implemented to do statistical analysis and more.

# What is R?

1. R is all things to all people.
2. System for statistical computation and graphics
3. Computer language
4. Interface to system functions
5. Runs commands or scripts
6. So: R is a computer language + a set of procedures that have been implemented to do statistical analysis and more.



# How long will it take me to learn R?

1. How long is a piece of string?
2. Useful things: Ten minutes.
3. Working knowledge, 2- 3 months with some daily practice
4. Dominance: The rest of your life.

# How long will it take me to learn R?

1. How long is a piece of string?
2. Useful things: Ten minutes.
3. Working knowledge, 2- 3 months with some daily practice
4. Dominance: The rest of your life.

# How long will it take me to learn R?

1. How long is a piece of string?
2. Useful things: Ten minutes.
3. Working knowledge, 2- 3 months with some daily practice
4. Dominance: The rest of your life.

# How long will it take me to learn R?

1. How long is a piece of string?
2. Useful things: Ten minutes.
3. Working knowledge, 2- 3 months with some daily practice
4. Dominance: The rest of your life.

# How long will it take me to learn R?

1. How long is a piece of string?
2. Useful things: Ten minutes.
3. Working knowledge, 2- 3 months with some daily practice
4. Dominance: The rest of your life.

## How long will it take me to learn R?

```
> fortune("learning curve")
```

The learning curve is steep - but then like many people, I'd like to be able to do sophisticated modelling with deep understanding and no effort :-)

-- Sean O'Riordain (in a thread about the helpfulness of documentation)

R-help (July 2005)

# How much does R cost?

1. Nothing.
2. Free and open source under terms of GPL
3. Can be modified, reused and distributed as seen fit.
4. Comes with “ABSOLUTELY NO WARRANTY”
5. All important code scrutinised by top level academic statisticians.

# How much does R cost?

1. Nothing.
2. Free and open source under terms of GPL
3. Can be modified, reused and distributed as seen fit.
4. Comes with “ABSOLUTELY NO WARRANTY”
5. All important code scrutinised by top level academic statisticians.



# How much does R cost?

1. Nothing.
2. Free and open source under terms of GPL
3. Can be modified, reused and distributed as seen fit.
4. Comes with “ABSOLUTELY NO WARRANTY”
5. All important code scrutinised by top level academic statisticians.

# How much does R cost?

1. Nothing.
2. Free and open source under terms of GPL
3. Can be modified, reused and distributed as seen fit.
4. Comes with “ABSOLUTELY NO WARRANTY”
5. All important code scrutinised by top level academic statisticians.

# How much does R cost?

1. Nothing.
2. Free and open source under terms of GPL
3. Can be modified, reused and distributed as seen fit.
4. Comes with “ABSOLUTELY NO WARRANTY”
5. All important code scrutinised by top level academic statisticians.

# How much does R cost?

1. Nothing.
2. Free and open source under terms of GPL
3. Can be modified, reused and distributed as seen fit.
4. Comes with “ABSOLUTELY NO WARRANTY”
5. All important code scrutinised by top level academic statisticians.

## How much does R cost?

```
> fortune("dodgy software")
```

Mingzhai Sun: When you use it [R], since it is written by so many authors, how do you know that the results are trustable?

Bill Venables: The R engine [...] is pretty well uniformly excellent code but you have to take my word for that.

Actually, you don't. The whole engine is open source so, if you wish, you can check every line of it. If people were out to push dodgy software, this is not the way they'd go about it.

```
-- Mingzhai Sun and Bill Venables  
R-help (January 2004)
```

# How do you install R?

1. Cross platform
2. Windows, Linux, Mac, Solaris
3. Possibly other platforms *“I’d like to see a Nintendo Wii port, just so I can play Super Mario Generalised Linear Modelling by waving the controller around.” (Barry Rowlingson, University of Lancaster)*
4. Windows version, conventional install using setup.  
<http://cran.r-project.org/bin/windows/base/R-2.8.0-win32.exe>
5. Linux (Debian Ubuntu) available from repositories using Aptitude or Synaptic to install r-base, r-base-dev, r-base-core, r-base-html and r-base-latex + extension packages.
6. RKWARD GUI available for Debian Linux
7. Sciviews interface for Windows <http://www.sciviews.org/>
8. SUSE RPMs available.
9. Mac <http://cran.r-project.org/bin/macosx/R-2.8.0.dmg>

# How do you install R?

1. Cross platform
2. Windows, Linux, Mac, Solaris
3. Possibly other platforms *“I’d like to see a Nintendo Wii port, just so I can play Super Mario Generalised Linear Modelling by waving the controller around.” (Barry Rowlingson, University of Lancaster)*
4. Windows version, conventional install using setup.  
<http://cran.r-project.org/bin/windows/base/R-2.8.0-win32.exe>
5. Linux (Debian Ubuntu) available from repositories using Aptitude or Synaptic to install r-base, r-base-dev, r-base-core, r-base-html and r-base-latex + extension packages.
6. RKWARD GUI available for Debian Linux
7. Sciviews interface for Windows <http://www.sciviews.org/>
8. SUSE RPMs available.
9. Mac <http://cran.r-project.org/bin/macosx/R-2.8.0.dmg>

## How do you install R?

1. Cross platform
2. Windows, Linux, Mac, Solaris
3. Possibly other platforms *“I’d like to see a Nintendo Wii port, just so I can play Super Mario Generalised Linear Modelling by waving the controller around.”* (Barry Rowlingson, University of Lancaster)
4. Windows version, conventional install using setup.  
<http://cran.r-project.org/bin/windows/base/R-2.8.0-win32.exe>
5. Linux (Debian Ubuntu) available from repositories using Aptitude or Synaptic to install r-base, r-base-dev, r-base-core, r-base-html and r-base-latex + extension packages.
6. RKWARD GUI available for Debian Linux
7. Sciviews interface for Windows <http://www.sciviews.org/>
8. SUSE RPMs available.
9. Mac <http://cran.r-project.org/bin/macosx/R-2.8.0.dmg>



## How do you install R?

1. Cross platform
2. Windows, Linux, Mac, Solaris
3. Possibly other platforms *“I’d like to see a Nintendo Wii port, just so I can play Super Mario Generalised Linear Modelling by waving the controller around.” (Barry Rowlingson, University of Lancaster)*
4. Windows version, conventional install using setup.  
<http://cran.r-project.org/bin/windows/base/R-2.8.0-win32.exe>
5. Linux (Debian Ubuntu) available from repositories using Aptitude or Synaptic to install r-base, r-base-dev, r-base-core, r-base-html and r-base-latex + extension packages.
6. RKWARD GUI available for Debian Linux
7. Sciviews interface for Windows <http://www.sciviews.org/>
8. SUSE RPMs available.
9. Mac <http://cran.r-project.org/bin/macosx/R-2.8.0.dmg>

## How do you install R?

1. Cross platform
2. Windows, Linux, Mac, Solaris
3. Possibly other platforms *“I’d like to see a Nintendo Wii port, just so I can play Super Mario Generalised Linear Modelling by waving the controller around.”* (Barry Rowlingson, University of Lancaster)
4. Windows version, conventional install using setup.  
<http://cran.r-project.org/bin/windows/base/R-2.8.0-win32.exe>
5. Linux (Debian Ubuntu) available from repositories using Aptitude or Synaptic to install r-base, r-base-dev, r-base-core, r-base-html and r-base-latex + extension packages.
6. RKWARD GUI available for Debian Linux
7. Sciviews interface for Windows <http://www.sciviews.org/>
8. SUSE RPMs available.
9. Mac <http://cran.r-project.org/bin/macosx/R-2.8.0.dmg>

## How do you install R?

1. Cross platform
2. Windows, Linux, Mac, Solaris
3. Possibly other platforms *“I’d like to see a Nintendo Wii port, just so I can play Super Mario Generalised Linear Modelling by waving the controller around.”* (Barry Rowlingson, University of Lancaster)
4. Windows version, conventional install using setup.  
<http://cran.r-project.org/bin/windows/base/R-2.8.0-win32.exe>
5. Linux (Debian Ubuntu) available from repositories using Aptitude or Synaptic to install r-base, r-base-dev, r-base-core, r-base-html and r-base-latex + extension packages.
6. RKWARD GUI available for Debian Linux
7. Sciviews interface for Windows <http://www.sciviews.org/>
8. SUSE RPMs available.
9. Mac <http://cran.r-project.org/bin/macosx/R-2.8.0.dmg>

## How do you install R?

1. Cross platform
2. Windows, Linux, Mac, Solaris
3. Possibly other platforms *“I’d like to see a Nintendo Wii port, just so I can play Super Mario Generalised Linear Modelling by waving the controller around.”* (Barry Rowlingson, University of Lancaster)
4. Windows version, conventional install using setup.  
<http://cran.r-project.org/bin/windows/base/R-2.8.0-win32.exe>
5. Linux (Debian Ubuntu) available from repositories using Aptitude or Synaptic to install r-base, r-base-dev, r-base-core, r-base-html and r-base-latex + extension packages.
6. RKWARD GUI available for Debian Linux
7. Sciviews interface for Windows <http://www.sciviews.org/>
8. SUSE RPMs available.
9. Mac <http://cran.r-project.org/bin/macosx/R-2.8.0.dmg>

## How do you install R?

1. Cross platform
2. Windows, Linux, Mac, Solaris
3. Possibly other platforms *“I’d like to see a Nintendo Wii port, just so I can play Super Mario Generalised Linear Modelling by waving the controller around.”* (Barry Rowlingson, University of Lancaster)
4. Windows version, conventional install using setup.  
<http://cran.r-project.org/bin/windows/base/R-2.8.0-win32.exe>
5. Linux (Debian Ubuntu) available from repositories using Aptitude or Synaptic to install r-base, r-base-dev, r-base-core, r-base-html and r-base-latex + extension packages.
6. RKWARD GUI available for Debian Linux
7. Sciviews interface for Windows <http://www.sciviews.org/>
8. SUSE RPMs available.
9. Mac <http://cran.r-project.org/bin/macosx/R-2.8.0.dmg>

## How do you install R?

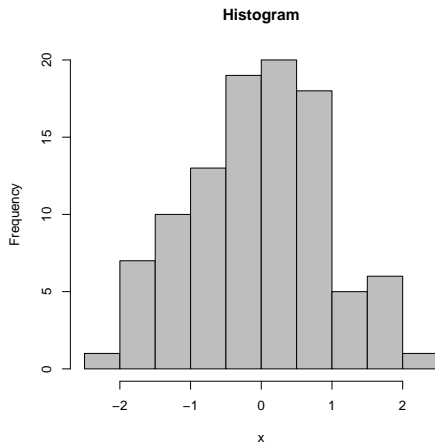
1. Cross platform
2. Windows, Linux, Mac, Solaris
3. Possibly other platforms *“I’d like to see a Nintendo Wii port, just so I can play Super Mario Generalised Linear Modelling by waving the controller around.” (Barry Rowlingson, University of Lancaster)*
4. Windows version, conventional install using setup.  
<http://cran.r-project.org/bin/windows/base/R-2.8.0-win32.exe>
5. Linux (Debian Ubuntu) available from repositories using Aptitude or Synaptic to install r-base, r-base-dev, r-base-core, r-base-html and r-base-latex + extension packages.
6. RKWARD GUI available for Debian Linux
7. Sciviews interface for Windows <http://www.sciviews.org/>
8. SUSE RPMs available.
9. Mac <http://cran.r-project.org/bin/macosx/R-2.8.0.dmg>

# Some simple examples

1. A handful of quick examples of R output
2. All output with one line of code
3. A few other lines to make the data available

# An histogram

```
> x <- rnorm(100)
> hist(x, col = "grey", main = "Histogram")
```





# A t-test

```
> t.test(x)
```

```
One Sample t-test
```

```
data: x
```

```
t = -0.3206, df = 99, p-value = 0.7492
```

```
alternative hypothesis: true mean is not equal to 0
```

```
95 percent confidence interval:
```

```
-0.2236384  0.1614275
```

```
sample estimates:
```

```
mean of x
```

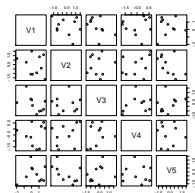
```
-0.03110543
```

# A table

```
> library(xtable)
> x <- rnorm(50)
> dim(x) <- c(10, 5)
> x <- as.data.frame(x)
> xtable(x)
```

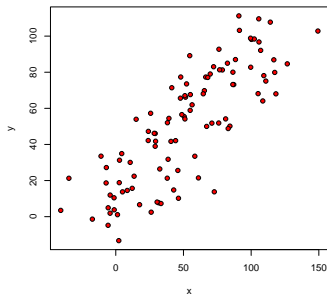
	V1	V2	V3	V4	V5
1	1.05	0.87	-1.02	-1.67	-1.15
2	-0.70	0.25	1.68	-0.62	-0.61
3	0.02	1.50	0.22	0.62	0.51
4	1.30	-0.41	-0.98	0.21	-0.97
5	0.12	-1.10	-0.40	-1.41	0.37
6	1.61	1.11	-0.82	0.55	-1.09
7	-1.97	0.67	-0.09	-0.12	-1.63
8	1.80	-1.67	0.05	0.06	0.66
9	0.68	-0.93	-1.39	-1.74	-0.31
10	-0.10	-1.09	0.31	-0.70	1.94

```
> plot(x)
```



# A regression model

```
> x <- 1:100 + rnorm(100, sd = 20)
> y <- 1:100 + rnorm(100, sd = 10)
> plot(y ~ x, pch = 21, bg = 2)
```

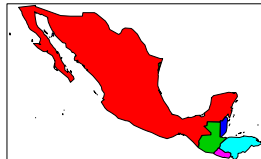


```
> mod <- lm(y ~ x)
> xtable(summary(mod))
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	17.5080	3.0076	5.82	0.0000
x	0.6479	0.0461	14.06	0.0000

# A map

```
> library(maptools)
> library(mapdata)
> map("worldHires")
> box()
```



# Exercises and activities

1. Install the latest version of R on your computer.
2. Investigate the contents of CRAN  
(<http://cran.r-project.org/>)
3. Download all the relevant manuals and courses from the  
“Contributed” section of CRAN.
4. Browse the R graphics gallery  
(<http://addictedtor.free.fr/graphiques/>)