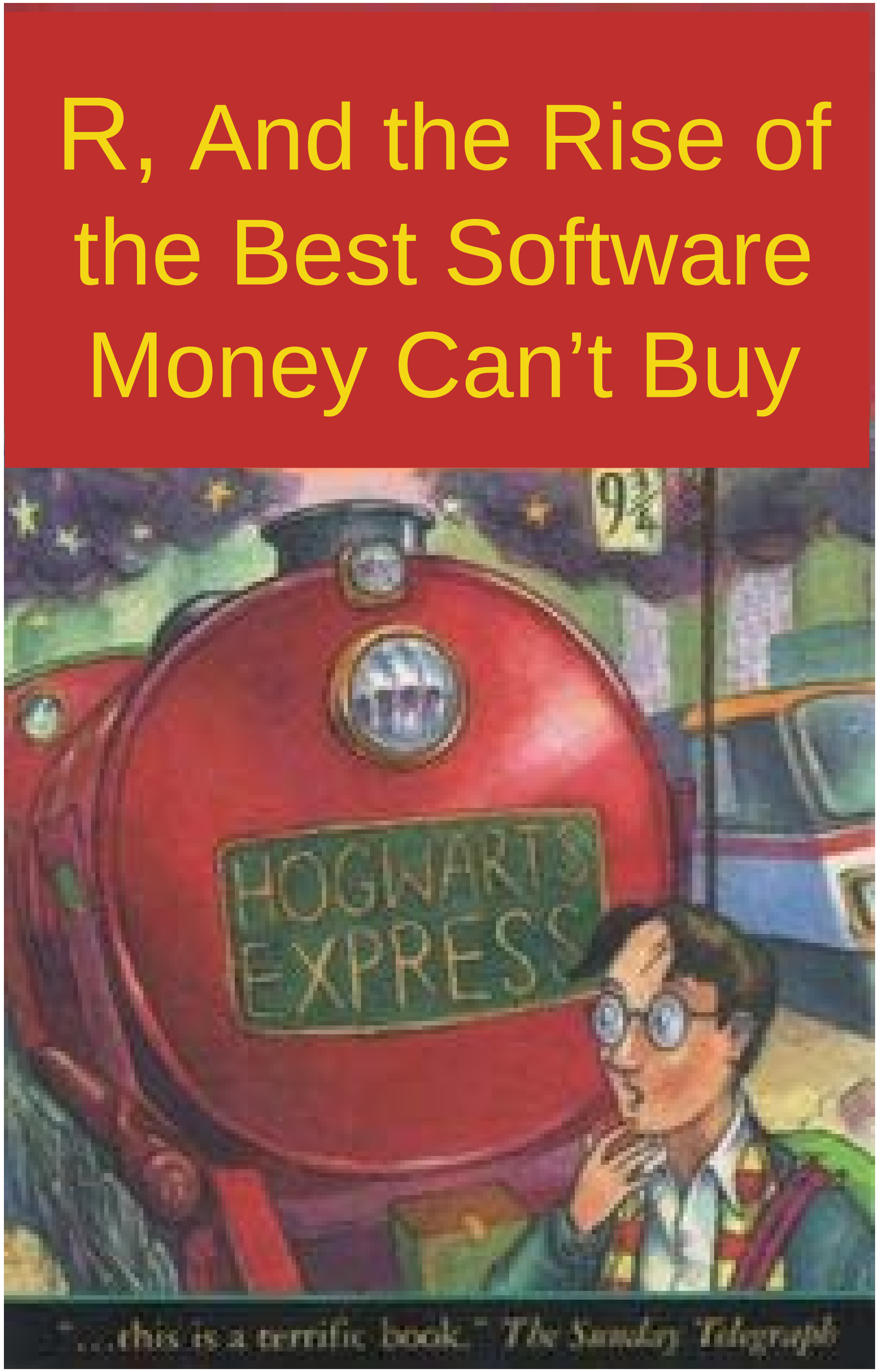
A very brief introduction to R

- Matthew Keller & Steven Boker

Some material cribbed from: UCLA Academic Technology Services Technical Report Series (by Patrick Burns) and presentations (found online) by Bioconductor, Wolfgang Huber and Hung Chen, & various Harry Potter websites

R programming language is a lot like magic... except instead of spells you have functions.



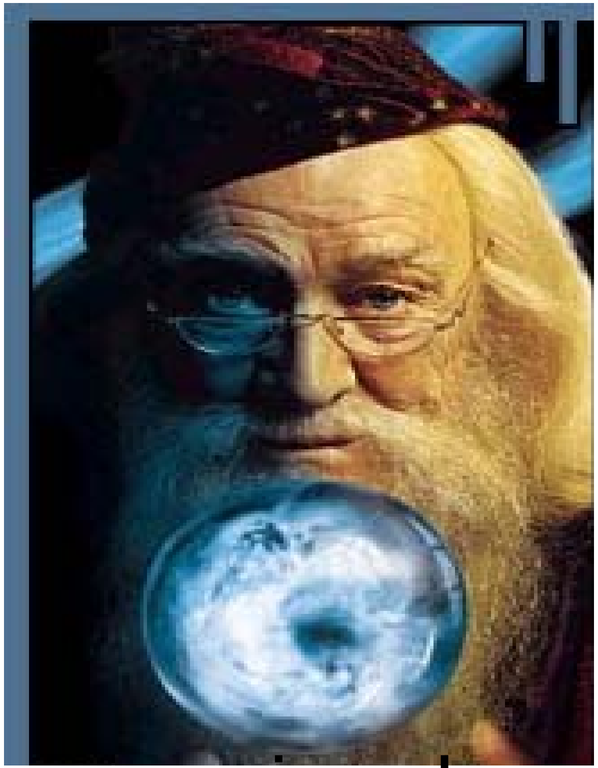
R, And the Rise of the Best Software Money Can't Buy

SPSS =

muggle

Like muggles, users of traditional stats software packages are limited in their ability to change their environment. They have to rely on algorithms that have been developed for them. The way they approach a problem is constrained by how employed programmers thought to approach them. And they have to pay money to use these constraining algorithms.

=



wizard

R users are like wizards. They can rely on functions developed for them by statistical researchers, but they can also create their own. They don’t have to pay for the use of them, and once experienced enough, they are almost unlimited in their ability to change their environment.

R

Advantages

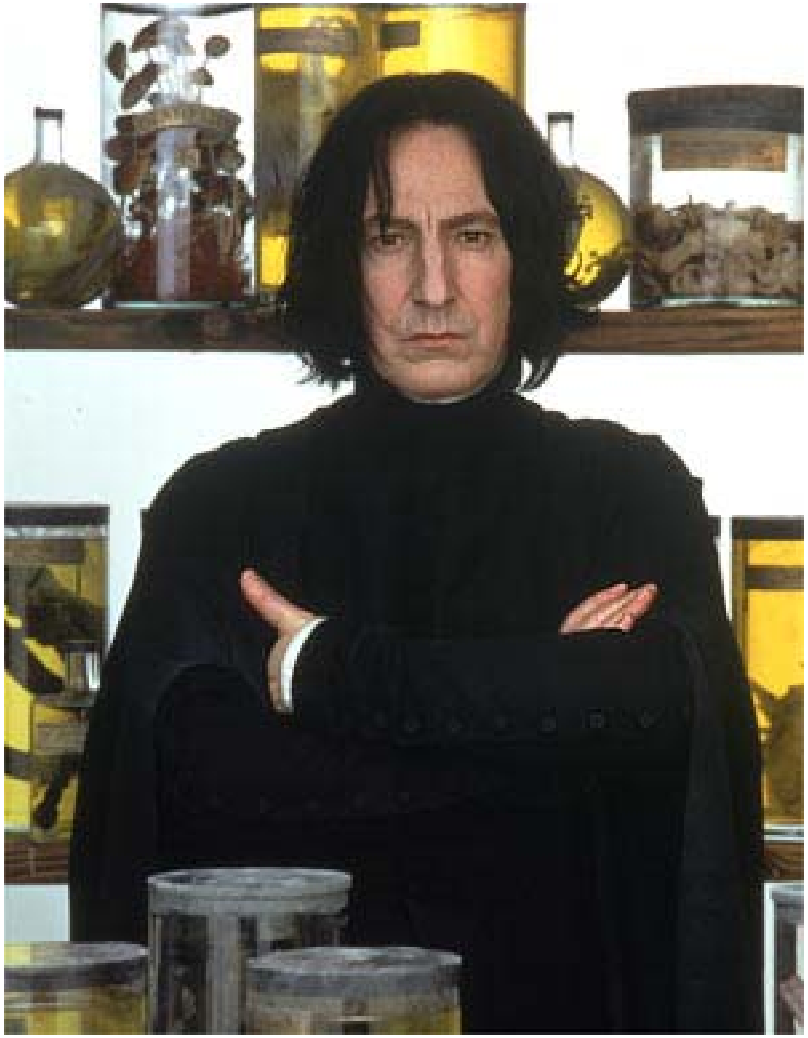
- Fast and free.
- State of the art: Statistical researchers provide their methods as R packages. SPSS and SAS are years behind R!
- 2nd only to MATLAB for graphics.
- Mx, WinBugs, and other programs use R.
- Active user community
- Excellent for simulation, programming, computer intensive analyses, etc.

Disadvantages

- Not user friendly at start - steep learning curve, minimal GUI.
- Easy to make mistakes and not know.
- Working with large datasets is limited by RAM
- Some users complain about hostility on the R listserve

Learning R....

R-help listserve....

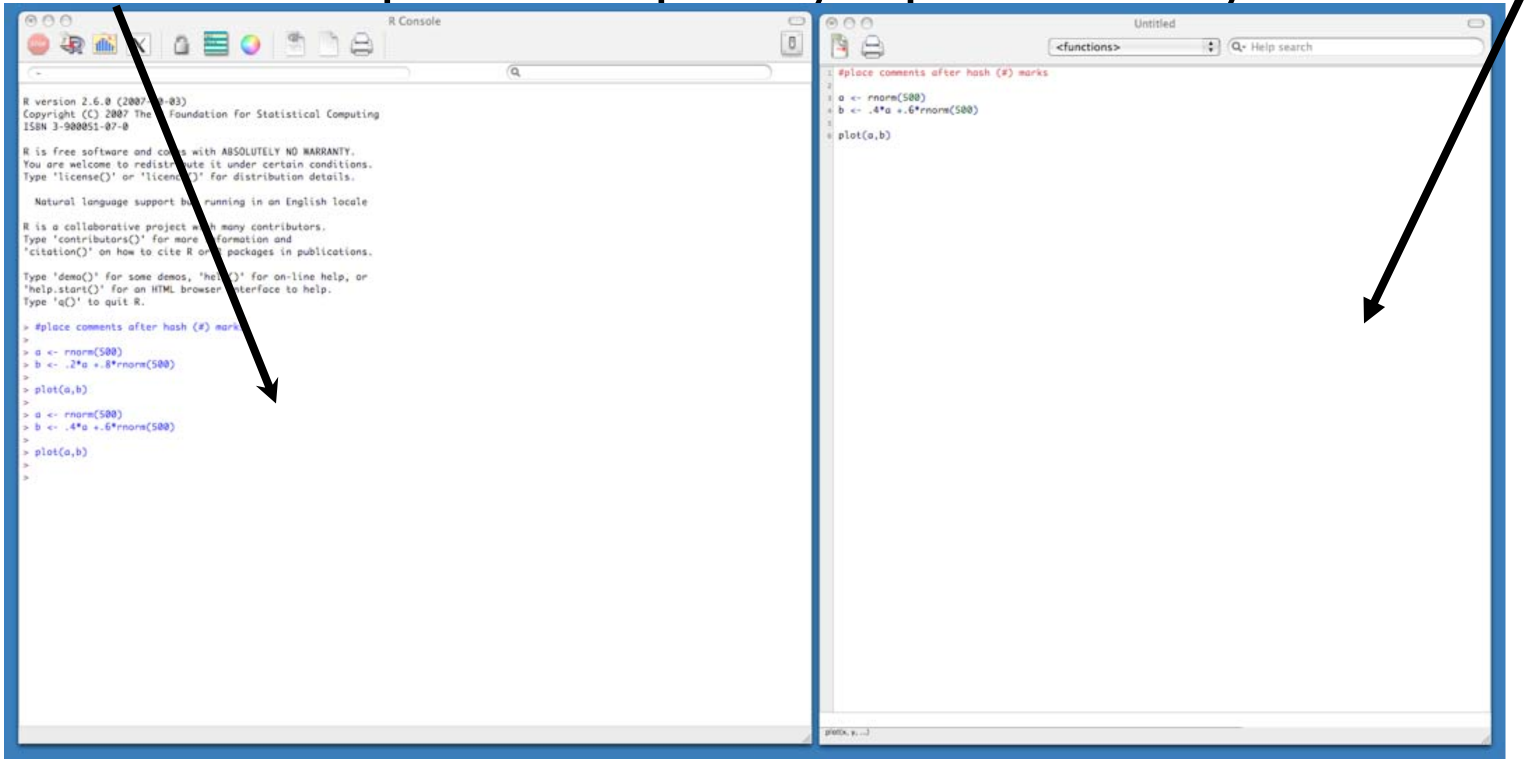


There are over 2000 add-on packages (http://cran.r-project.org/src/contrib/PACKAGES.html)

- This is an enormous advantage - new techniques available without delay, and they can be performed using the R language you already know.
- Allows you to build a customized statistical program suited to your own needs.
- Downside = as the number of packages grows, it is becoming difficult to choose the best package for your needs, & QC is an issue.

Typical R session

- Start up R via the GUI or favorite text editor
- Two windows:
  - 1+ new or existing scripts (text files) - these will be saved
  - Terminal – output & temporary input - usually unsaved

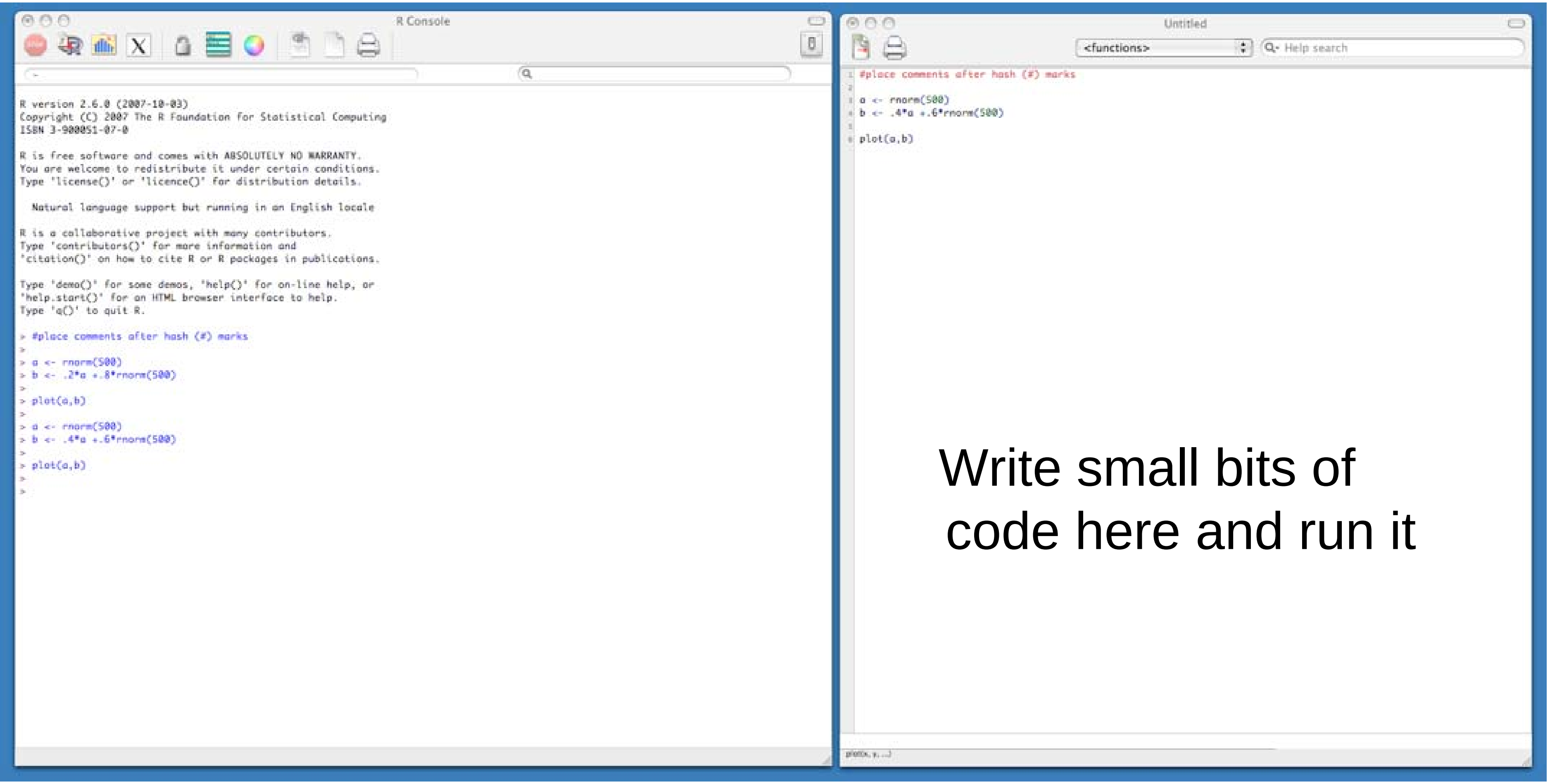


```
R version 2.6.0 (2007-10-03)
Copyright (C) 2007 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

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.

> #place comments after hash (#) marks
>
> a <- rnorm(500)
> b <- .2*a +.8*rnorm(500)
>
> plot(a,b)
>
> a <- rnorm(500)
> b <- .4*a +.6*rnorm(500)
>
> plot(a,b)
>
>
```

```
#place comments after hash (#) marks

a <- rnorm(500)
b <- .4*a +.6*rnorm(500)

plot(a,b)
```

Typical R session

- R sessions are *interactive*



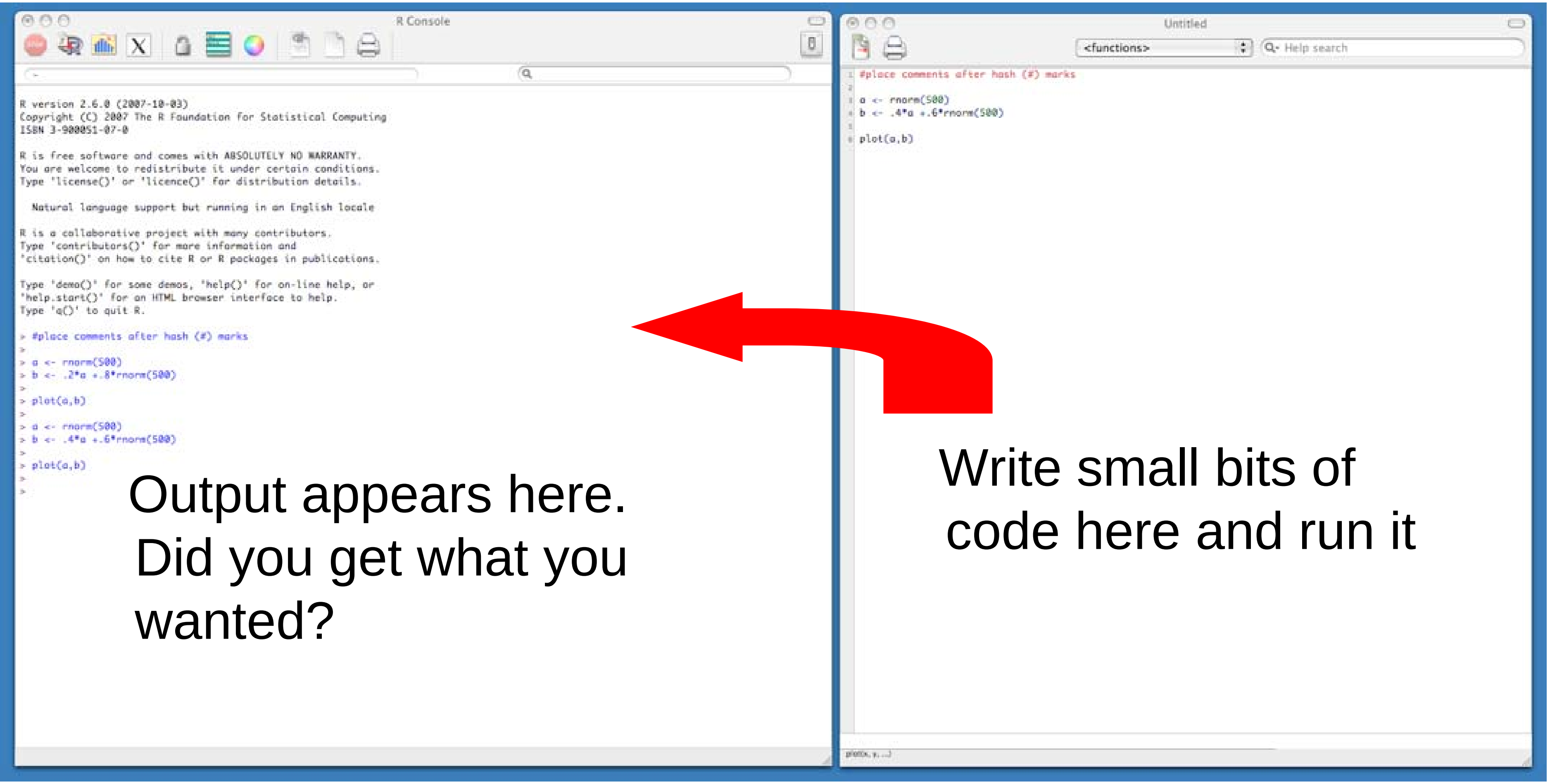
```
R version 2.6.0 (2007-10-03)
Copyright (C) 2007 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

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.

> #place comments after hash (#) marks
>
> a <- rnorm(500)
> b <- .2*a +.8*rnorm(500)
>
> plot(a,b)
>
> a <- rnorm(500)
> b <- .4*a +.6*rnorm(500)
>
> plot(a,b)
>
>
```

```
#place comments after hash (#) marks

a <- rnorm(500)
b <- .4*a +.6*rnorm(500)

plot(a,b)
```

Write small bits of code here and run it

Typical R session

- R sessions are *interactive*



```
R version 2.6.0 (2007-10-03)
Copyright (C) 2007 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

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.

> #place comments after hash (#) marks
>
> a <- rnorm(500)
> b <- .2*a +.8*rnorm(500)
>
> plot(a,b)
>
> a <- rnorm(500)
> b <- .4*a +.6*rnorm(500)
>
> plot(a,b)
>
>
```

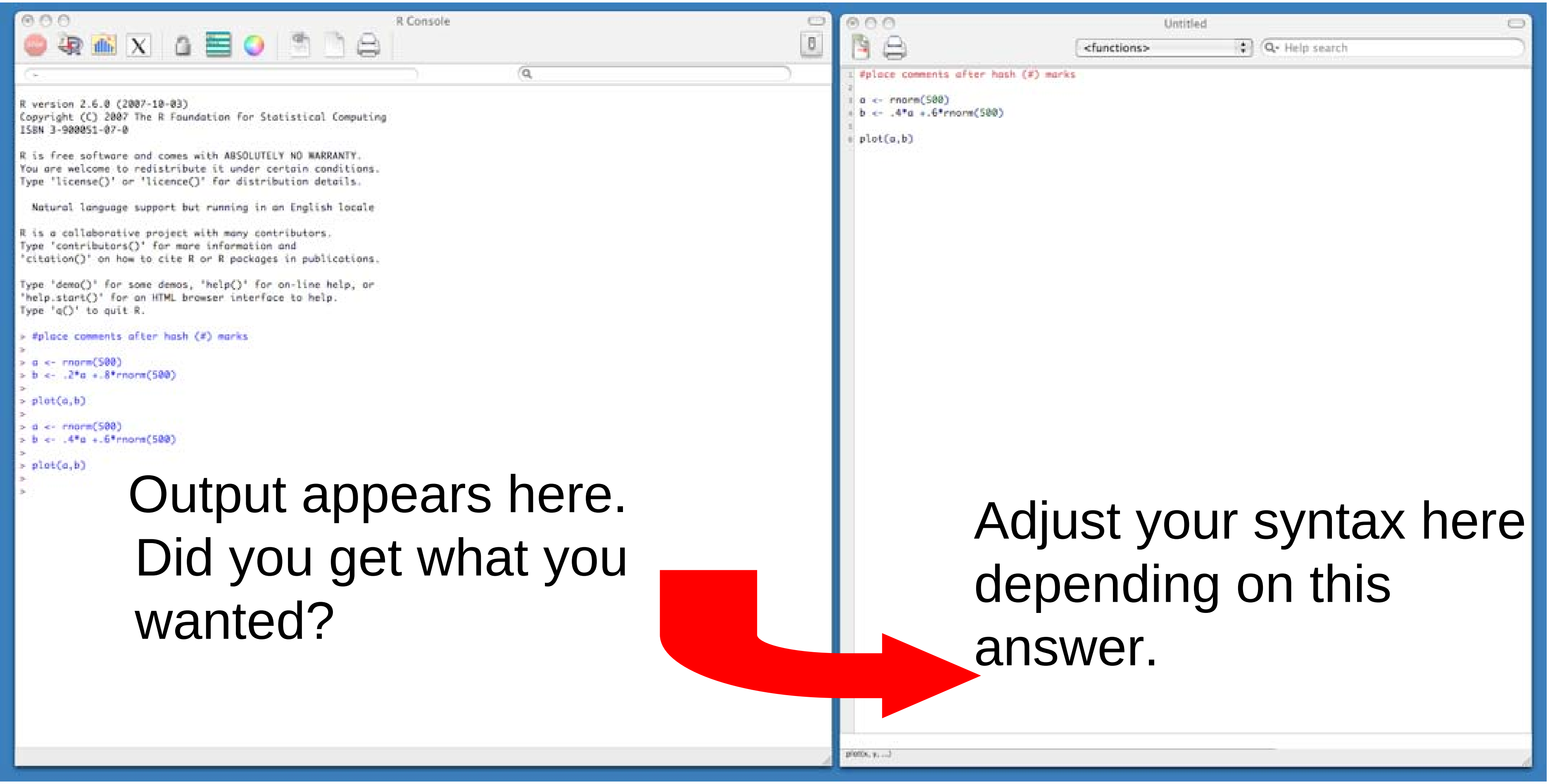
```
#place comments after hash (#) marks

a <- rnorm(500)
b <- .4*a +.6*rnorm(500)

plot(a,b)
```

Output appears here. Did you get what you wanted?

Write small bits of code here and run it

Typical R session

- R sessions are *interactive*



```
R version 2.6.0 (2007-10-03)
Copyright (C) 2007 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

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.

> #place comments after hash (#) marks
>
> a <- rnorm(500)
> b <- .2*a +.8*rnorm(500)
>
> plot(a,b)
>
> a <- rnorm(500)
> b <- .4*a +.6*rnorm(500)
>
> plot(a,b)
>
>
```

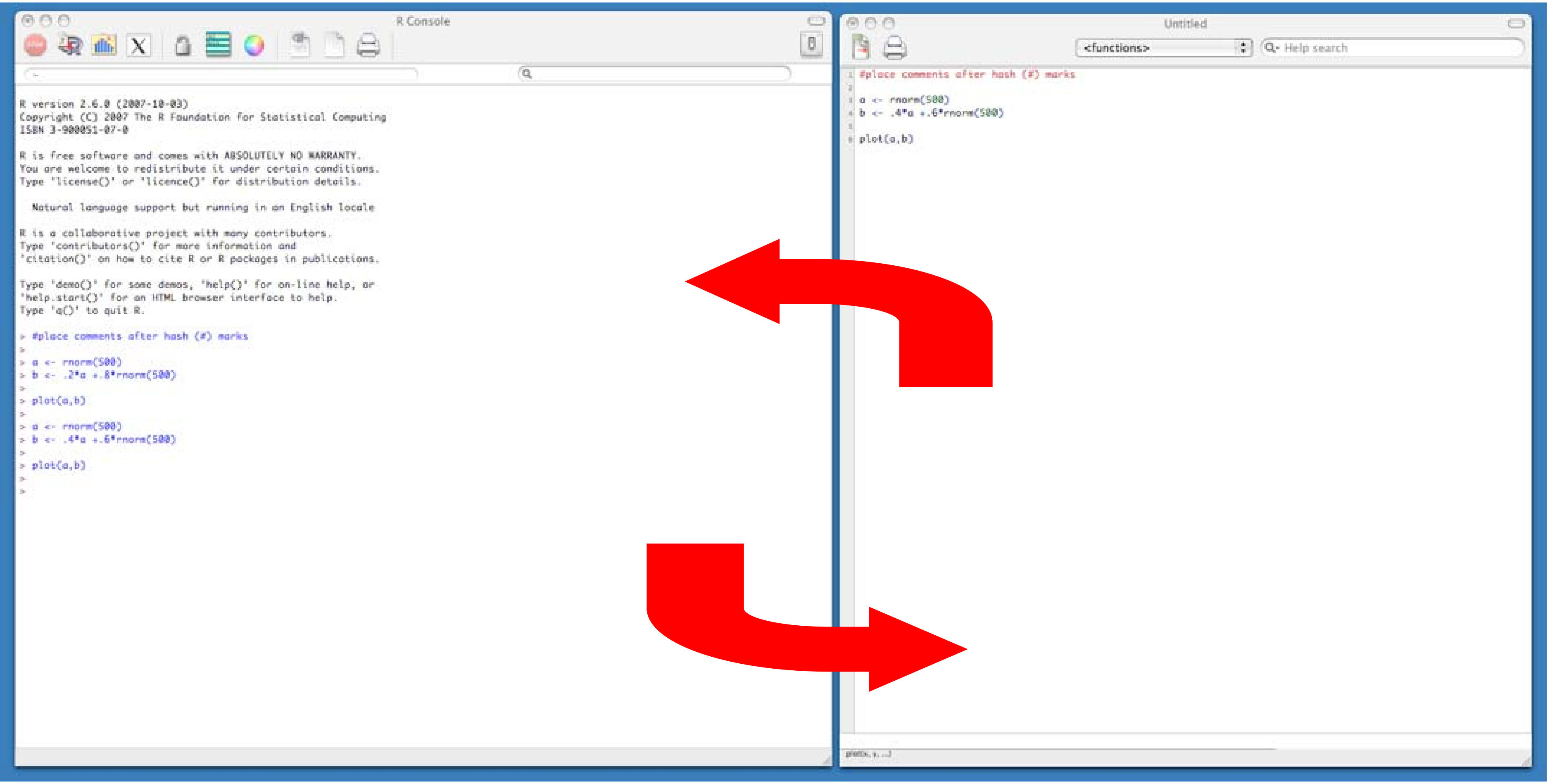
```
#place comments after hash (#) marks

a <- rnorm(500)
b <- .4*a +.6*rnorm(500)

plot(a,b)
```

Output appears here. Did you get what you wanted?

Adjust your syntax here depending on this answer.

Typical R session

- R sessions are *interactive*

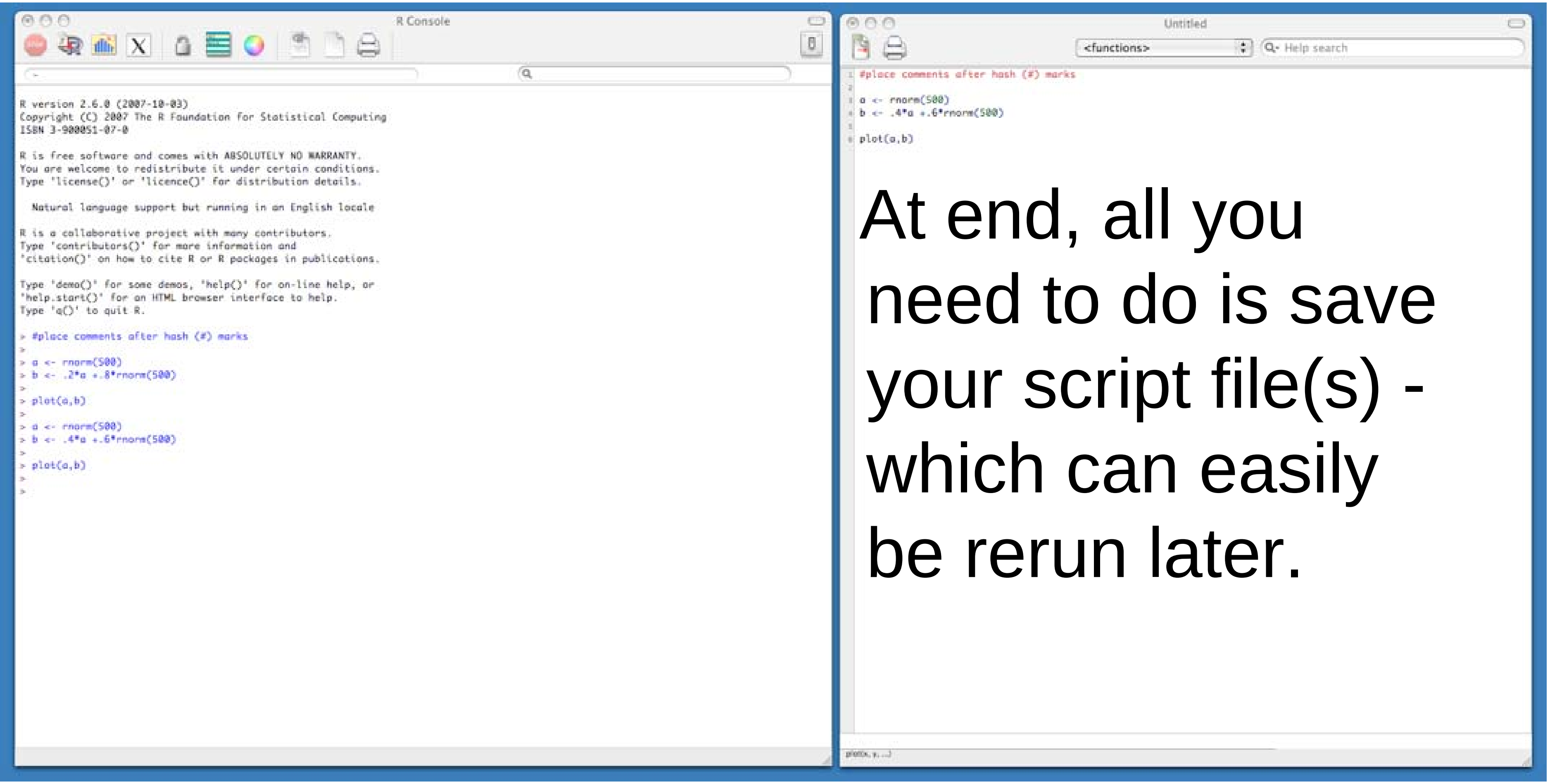


```
R version 2.6.0 (2007-10-03)
Copyright (C) 2007 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

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.

> #place comments after hash (#) marks
>
> a <- rnorm(500)
> b <- .2*a +.8*rnorm(500)
>
> plot(a,b)
>
> a <- rnorm(500)
> b <- .4*a +.6*rnorm(500)
>
> plot(a,b)
>
>
```

```
#place comments after hash (#) marks

a <- rnorm(500)
b <- .4*a +.6*rnorm(500)

plot(a,b)
```

Typical R session

- R sessions are *interactive*



```
R version 2.6.0 (2007-10-03)
Copyright (C) 2007 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

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.

> #place comments after hash (#) marks
>
> a <- rnorm(500)
> b <- .2*a +.8*rnorm(500)
>
> plot(a,b)
>
> a <- rnorm(500)
> b <- .4*a +.6*rnorm(500)
>
> plot(a,b)
>
>
```

```
#place comments after hash (#) marks

a <- rnorm(500)
b <- .4*a +.6*rnorm(500)

plot(a,b)
```

At end, all you need to do is save your script file(s) - which can easily be rerun later.

Learning R

- Use http://www.rseek.org/ instead of google
- Because R is interactive, errors are your friends!
- ?lm gives you help on lm function. Reading help files can be very… helpful
- MOST IMPORTANT - the more time you spend using R, the more comfortable you become with it. After doing your first real project in R, you won't look back. I promise.