

第二届中国R语言会议介绍



中国人民大学统计学院

陈堰平

2009年12月5日



UseR eveRywheRe !



目录

- 主办方
- 会议背景
- 参会情况
- 展望



主办方

- 北京会场（2009.12.5~6）

主办：应用统计科学研究中心
中国人民大学统计学院

协办：统计之都

- 上海会场（2009.12.12~13）

主办：华东师范大学资源与环境科学学院
华东师范大学金融与统计学院

协办：统计之都

会议背景

- 第一届中国R语言会议

<http://cos.name/2008/12/1st-chinese-r-conference-summary/>

2008年12月13日~14日，由教育部重点研究基地中国人民大学应用统计科学研究中心与中国人民大学统计学院主办的“第一届中国R语言会议”在中国人民大学明德商学楼0402教室成功召开。国内R语言的各位先锋如中国人民大学统计学院吴喜之教授、中国科学院上海生命科学研究院丁国徽博士等以及国外R语言专业人士如澳大利亚国立大学John Maindonald教授也通过网络连线会场参加了本次会议；会议围绕R语言的众多应用领域展开了广泛的交流和讨论，为R语言在中国的应用和推广开辟了一条崭新的道路。

第一届中国R语言会议的会议纪要也发表在第一期R Journal上了！

http://journal.r-project.org/2009-1/RJournal_2009-1.pdf



R上了《纽约时报》

2009年1月7号，《纽约时报》科技版登了一篇注定要引起四方瞩目的文章，[*Data Analysts Captivated by R's Power*](#)（1月6号就有网络版），作者是该报的记者[Ashlee Vance](#)。这大概是开源统计软件包R，自1996年诞生以来，第一次出现在公众视野，而且是出现在《纽约时报》这样的主流媒体。

SPSS 创立者 Norman Nie 加入 R 社区

2009年10月20日 REvolution 消息，REvolution 获得来自 North Bridge 和英特尔的 900 万美元的风险投资。最为轰动的是：SPSS 软件的创始人之一，前 SPSS 的 CEO——Norman Nie 正式成为 REvolution 的 CEO！



参会单位

大学

北京林业大学、北京大学、北京师范大学、南京财经大学、复旦大学、浙江大学、北京工业大学、中国人民大学、华中科技大学、中央民族大学、中国石油大学（北京）、东南大学、山东大学、中南大学、北方工业大学、上海海洋大学、中国农业大学、上海电机学院、江苏大学、上海大学、华东师范大学、苏州大学、西北师范大学、上海财经大学、上海中医药大学、西藏民族学院、北京协和医科大学、南京人口管理干部学院等



科研所

上海植生所、中科院西双版纳植物园、中科院北京植物所、上海生命科学研究所、复旦大学遗传所、中科院生态中心、上海生科院计算所、有色金属技术经济研究院、中国林科院环森保所、苏州大学系统生物学中心、中科院计算生物所、南京地质古生物研究所、中科院植物研究所、中科院研究生院等



公司

赛诺菲巴斯德（中国）、中科软科技股份有限公司、中芯科技、方正科技、**CUPL**、上海源略数据、思创软件、好耶、新华信国际信息咨询（北京）、上海汇众汽车、拓普公司、**SPSS China**、**SAS China**、摩托罗拉（中国）等

政府事业单位

国家统计局、国家知识产权局专利局、宁波市疾控中心、北京市质量协会



展望

- 组织形式
- R相关资料的编辑出版
-

吃水不忘挖井人！

R在中国的繁荣昌盛不会太远了，但同志们需要大大努力。

——谢益辉

Department of Statistics,
Iowa State University

<http://yihui.name/>





祝贺第二届中国R语言大会
胜利召开！

开放、绿色、功能强大、具
有源源不断巨大资源的R不
仅有必要而且一定能够在中
国推广和发展。

——吴喜之



感谢大家的热情参与!





R语言的发展

R语言是从S语言演变而来。S和R在统计学中都有着广泛的应用。

S语言是二十世纪70年代由贝尔实验室的Rick Becker, John Chambers, Allan Wilks等人开发的用于统计绘图的语言。S语言可在商业软件Splus上运行。

1995年由Auckland大学统计系的**R**obert Gentleman和**R**oss Ihaka开始编制R软件，目前由一批志愿者组成的R核心开发小组负责。



The R Project for Statistical Computing

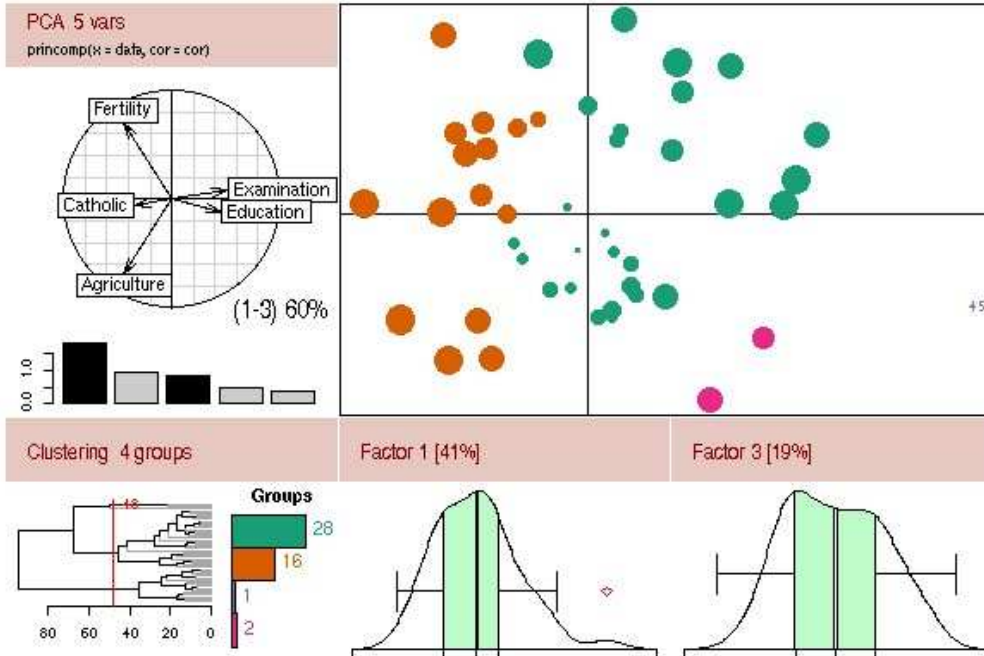
About R
[What is R?](#)
[Contributors](#)
[Screenshots](#)
[What's new?](#)

Download, Packages
[CRAN](#)

R Project
Foundation
[Members & Donors](#)
[Mailing Lists](#)
[Bug Tracking](#)
[Developer Page](#)
[Conferences](#)
[Search](#)

Documentation
[Manuals](#)
[FAQs](#)
[The R Journal](#)
[Wiki](#)
[Books](#)
[Certification](#)
[Other](#)

Misc
[Bioconductor](#)
[Related Projects](#)



Getting Started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).
- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

R软件首页 <http://www.r-project.org/>



CRAN Mirrors

The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some statistics on the status of the mirrors can be found [here](#).

About R

[What is R?](#)
[Contributors](#)
[Screenshots](#)
[What's new?](#)

Download, Packages CRAN

[R Project
Foundation](#)
[Members & Donors](#)
[Mailing Lists](#)
[Bug Tracking](#)
[Developer Page](#)
[Conferences](#)
[Search](#)

[Documentation](#)
[Manuals](#)
[FAQs](#)
[The R Journal](#)
[Wiki](#)
[Books](#)
[Certification](#)
[Other](#)

Argentina	http://cran.patan.com.ar/	Patan.com.ar, Buenos Aires
	http://mirror.cricyt.edu.ar/r/	CONICET, Mendoza
Australia	http://cran.ms.unimelb.edu.au/	University of Melbourne
Austria	http://cran.at.r-project.org/	Wirtschaftsuniversitaet Wien
Belgium	http://www.freeststatistics.org/cran/	K.U.Leuven Association
Brazil	http://cran.br.r-project.org/	Universidade Federal do Parana
	http://cran.fiocruz.br/	Oswaldo Cruz Foundation, Rio de Janeiro
	http://www.vps.fmvz.usp.br/CRAN/	University of Sao Paulo, Sao Paulo
	http://brieger.esalq.usp.br/CRAN/	University of Sao Paulo, Piracicaba
Canada	http://cran.stat.sfu.ca/	Simon Fraser University, Burnaby
	http://probability.ca/cran/	University of Toronto
	http://cran.parentinginformed.com/	iWeb, Montreal
Chile	http://dirichlet.mat.puc.cl/	Pontificia Universidad Catolica de Chile, Santiago
China		

CRAN The Comprehensive R Archive Network



The Comprehensive R Archiv

Frequently used pages

CRAN

[Mirrors](#)

[What's new?](#)

[Task Views](#)

[Search](#)

About R

[R Homepage](#)

[The R Journal](#)

Software

[R Sources](#)

[R Binaries](#)

[Packages](#)

[Other](#)

Documentation

[Manuals](#)

[FAQs](#)

[Contributed](#)

Download and Install R

Precompiled binary distributions of the base system packages, **Windows** and **Mac** users most likely want c
R:

- [Linux](#)
- [MacOS X](#)
- [Windows](#)

Source Code for all Platforms

Windows and Mac users most likely want the precompil
the upper box, not the source code. The sources have
you can use them. If you do not know what this means
want to do it!

- **The latest release** (2009-10-26): [R-2.10.0.tar](#)
the latest version).
- Sources of [R alpha and beta releases](#) (daily sna

CRAN 北京



R for Windows

This directory contains binaries for a base distribution and packages to run on i386/x64 Windows.

Note: CRAN does not have Windows systems and cannot check these binaries for viruses. Use the normal pre-downloaded executables.

Subdirectories:

base	Binaries for base distribution (managed by Duncan Murdoch)
contrib	Binaries of contributed packages (managed by Uwe Ligges)

Please do not submit binaries to CRAN. Package developers might want to contact Duncan Murdoch or Uwe Ligges if questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Last modified: April 4, 2004, by Friedrich Leisch

选择base: 下载基础包

选择contrib: 应用程序包

R-2.10.0 for Windows

[Download R 2.10.0 for Windows](#) (31 megabytes)

[Installation and other instructions](#)

New features in this version: [Windows specific](#), [all platforms](#).

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](#) ; [command line versions](#) are available.

Frequently asked questions

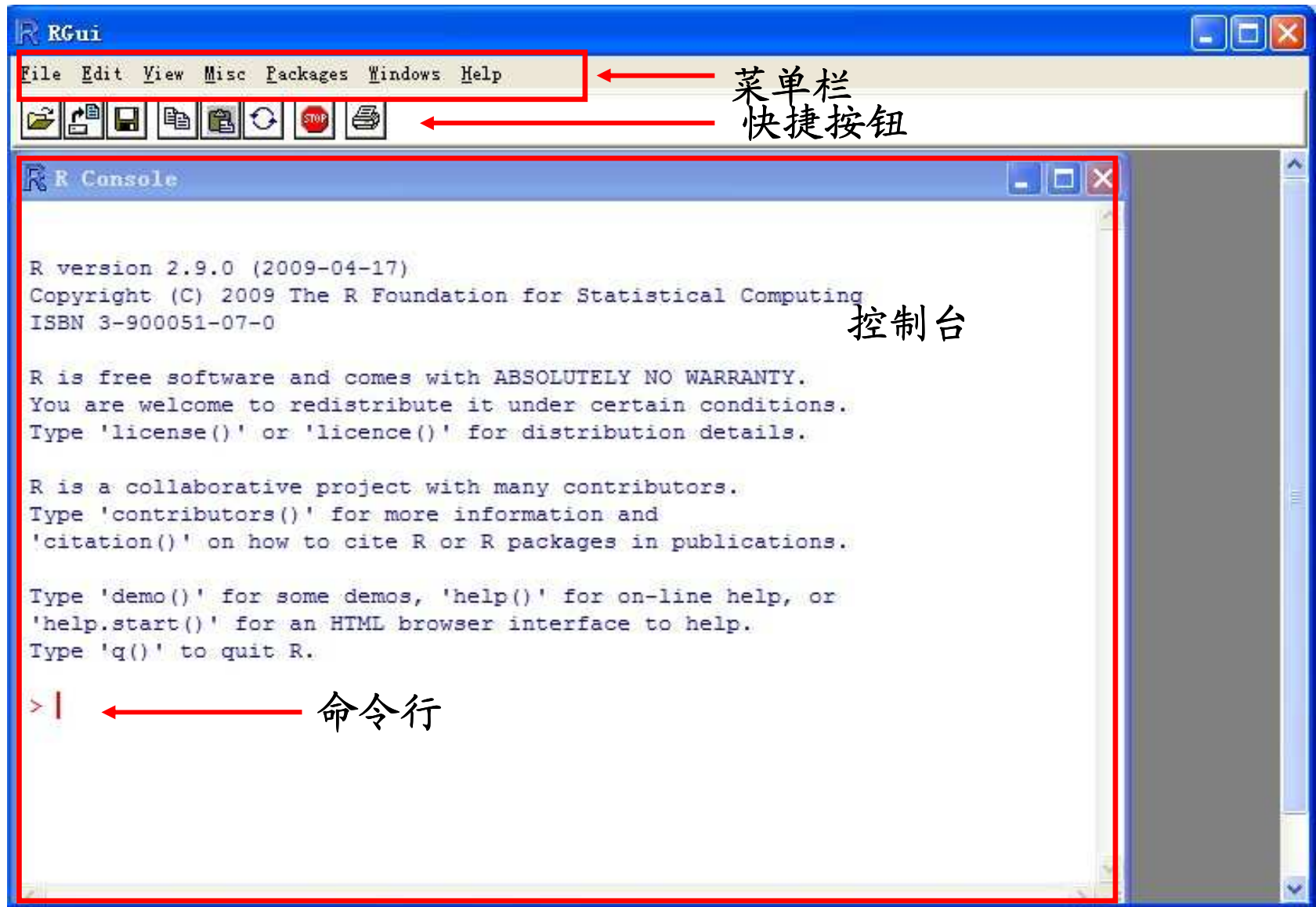
- [How do I install R when using Windows Vista?](#)
- [How do I update packages in my previous version of R?](#)

Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

Other builds

- Patches to this release are incorporated in the [r-patched snapshot build](#).

R-2.10.0 for windows 下载界面





R资源

R主页: <http://www.r-project.org>

统计之都论坛 S-plus & R 语言版

<http://cos.name/bbs/thread.php?fid=15>

R程序包

CRAN

[Mirrors](#)

[What's new?](#)

[Task Views](#)

[Search](#)

About R

[R Homepage](#)

[The R Journal](#)

Software

[R Sources](#)

[R Binaries](#)

[Packages](#)

[Other](#)

Documentation

[Manuals](#)

[FAQs](#)

[Contributed](#)

Download and Install R

Precompiled binary distributions of the base system and packages, **Windows and Mac** users most likely want one of R:

- [Linux](#)
- [MacOS X](#)
- [Windows](#)

Source Code for all Platforms

Windows and Mac users most likely want the precompiled binary from the upper box, not the source code. The sources have to be compiled by you can use them. If you do not know what this means, you want to do it!

- **The latest release** (2009-10-26): [R-2.10.0.tar.gz](#) (the latest version).
- Sources of [R alpha and beta releases](#) (daily snapshots during time periods before a planned release).

CRAN 主页 <http://cran.r-project.org/>



CRAN Task Views

Bayesian	Bayesian Inference
ChemPhys	Chemometrics and Computational Physics
Cluster	Cluster Analysis & Finite Mixture Models
Distributions	Probability Distributions
Econometrics	Computational Econometrics
Environmetrics	Analysis of Ecological and Environmental Data
ExperimentalDesign	Design of Experiments (DoE) & Analysis of Experimental Data
Finance	Empirical Finance
Genetics	Statistical Genetics
Graphics	Graphic Displays & Dynamic Graphics & Graphic Devices & Visual
gR	gRaphical Models in R
HighPerformanceComputing	High Performance and Parallel Computing
MachineLearning	Machine Learning & Statistical Learning
MedicalImaging	Medical Image Analysis
Multivariate	Multivariate Statistics
NaturalLanguageProcessing	Natural Language Processing
Optimization	Optimization and Mathematical Programming
Pharmacokinetics	Analysis of Pharmacokinetic Data
Psychometrics	Psychometric Models and Methods
Robust	Robust Statistical Methods
SocialSciences	Statistics for the Social Sciences
Spatial	Analysis of Spatial Data
Survival	Survival Analysis

CRAN Task Views: 对程序包的分类介绍



R程序包安装

(1)函数`install.packages()`，将自动下载程序包并将其安装好。

例如: `install.packages("quantreg")`

(2)安装本地zip包

路径: `Packages>install packages from local files`

选择光盘或者本地磁盘上存储zip包的文件夹。



调用程序包

在控制台中输入命令:

```
library(quantreg)
```



帮助系统

(1) ?

(2) 菜单 帮助>Html帮助

(3) 网络求助

R论坛, <http://www.rseek.org>, 邮件列表,.....



编辑器

- R的自带编辑器
- Emacs+ESS
- Notepad++ & NppToR插件
- Tinn-R
- Vim
- ...

CHAMPION File: *R*

File Edit Options Buffers Tools iESS Comp

ch4.r stolz.r

```
> n <- 5000
> x <- 1:n
> y <- numeric(n)
> tmp <- x^(1/x)
> for(i in x){
+   y[i] <- sum(tmp[1:i])/i
+ }

> plot(x,y,type='l')
> abline(h=1,col="red")

a
```

stolz.r All L1 (E)

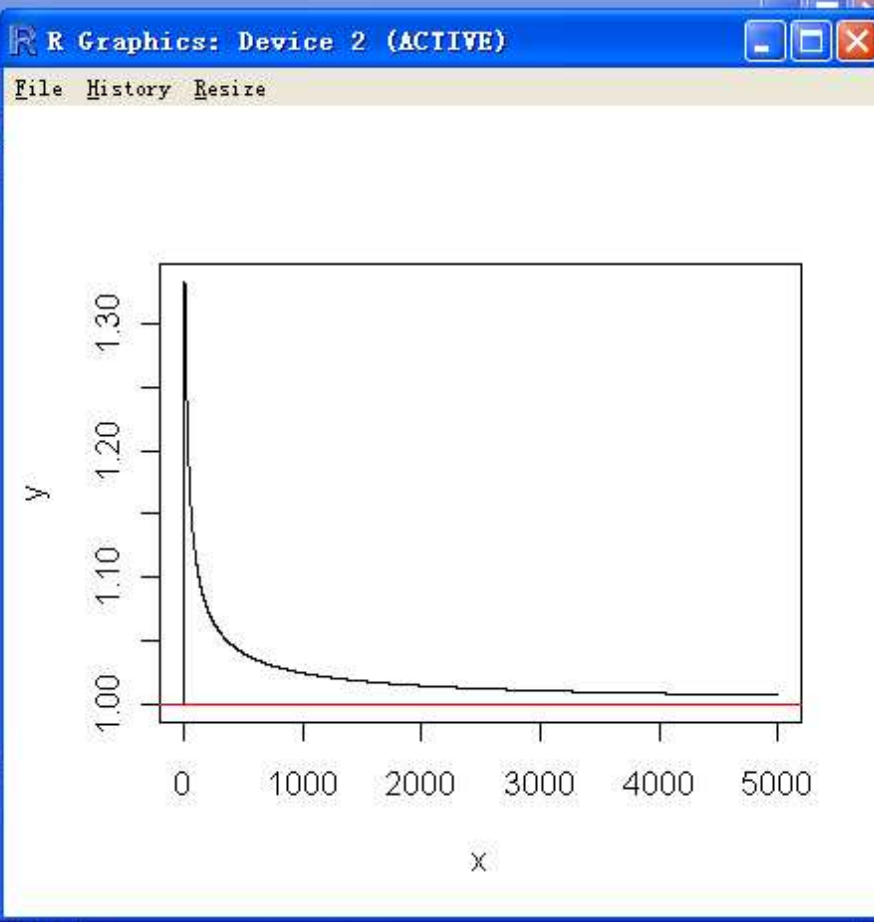
R

Type 'q()' to quit R.

Loading Tcl/Tk interface ... done
Error in winMenuAdd(menuName, NULL, NULL) :
Menu functions can only be used in the GUI
R> options(chmhelp = FALSE)
R> options(STERM='iESS', editor='gnuclient.exe')
R> n <- 5000
R> x <- 1:n
R> y <- numeric(n)
R> tmp <- x^(1/x)

R Graphics: Device 2 (ACTIVE)

File History Resize



Emacs + ESS

*J:\work\citics\aug10.r - Notepad++

```

16 d <- dist(tcdata)
17 #png("aug11.png")
18 fit <- hclust(d, method="complete")
19 plot(fit) # display dendrogram
20 groups <- cutree(fit, k=5) # cut tree into 5 clusters
21 # draw dendrogram with red borders around the 5 clusters
22 rect.hclust(fit, k=5, border="red")
23 #dev.off()
24
25 meanWeight <- colMeans(weight[,2:23])
26
27 odbcClose(con)
28

```

nb char : 643 nb line : 30 Ln : 23

R Console

```

R>
R> ##聚类
R> cdata <- indust[,2:23]
R> tcdata <- t(cdata)
R>
R> d <- dist(tcdata)
R> #png("aug11.png")
R> fit <- hclust(d, method="complete")
R> plot(fit) # display dendrogram
R> groups <- cutree(fit, k=5) # cut tree into 5 clusters
R> # draw dendrogram with red borders around the 5 clusters
R> rect.hclust(fit, k=5, border="red")
R>

```

R Graphics: Device 2 (ACTIVE)

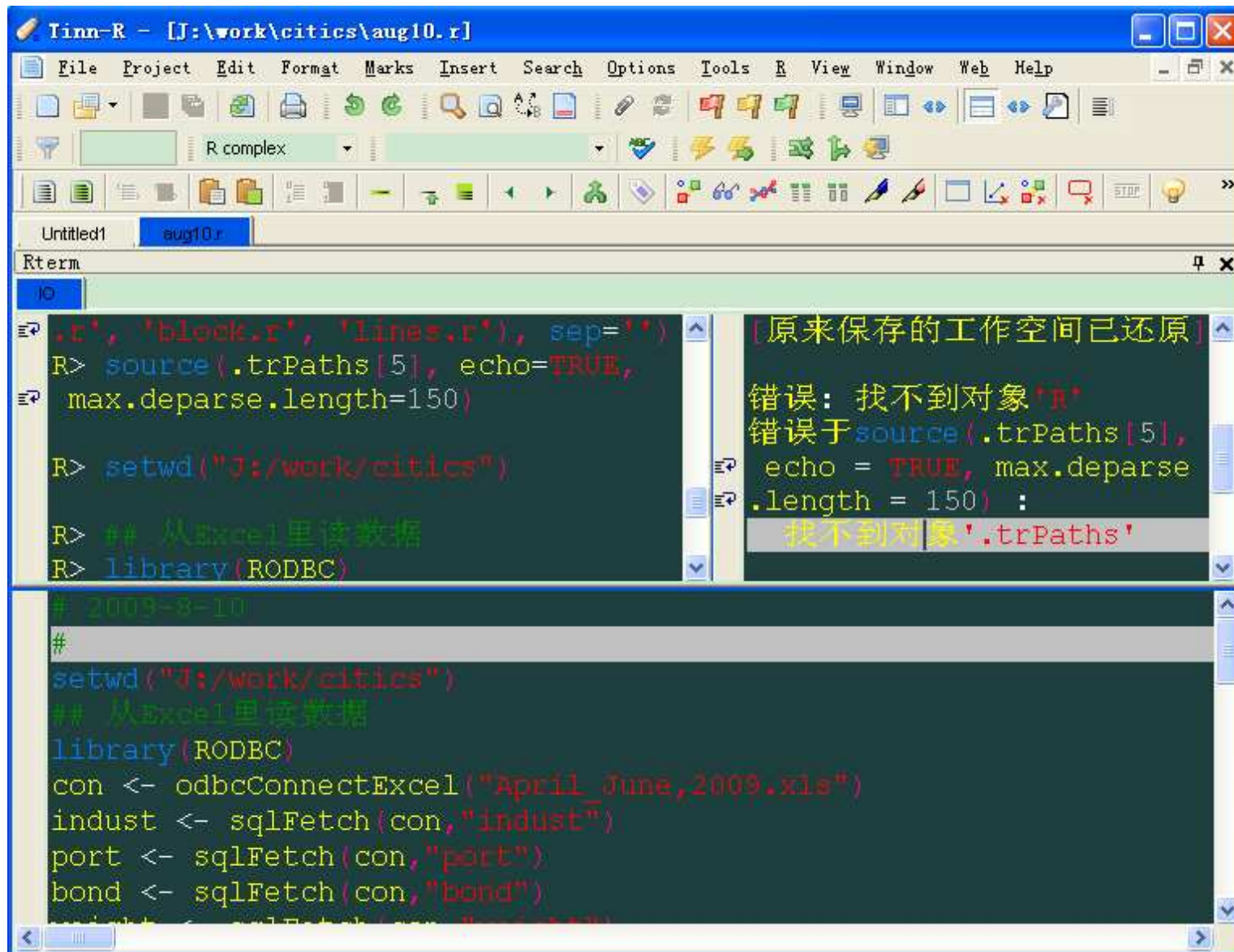
文件 历史 重设大小

Cluster Dendrogram

Height

d
hclust (*, "complete")

Notepad++ & NppToR



The screenshot shows the Tinn-R R editor interface. The window title is "Tinn-R - [J:\work\citics\aug10.r]". The menu bar includes File, Project, Edit, Format, Marks, Insert, Search, Options, Tools, R, View, Window, Web, and Help. The toolbar contains various icons for file operations and editing. The main editor area shows the following R code:

```
.r', 'block.r', 'lines.r'), sep='')
R> source(.trPaths[5], echo=TRUE,
max.deparse.length=150)

R> setwd("J:/work/citics")

R> ## 从excel里读数据
R> library(RODBC)

# 2009-8-10
#
setwd("J:/work/citics")
## 从excel里读数据
library(RODBC)
con <- odbcConnectExcel("April_June,2009.xls")
indust <- sqlFetch(con,"indust")
port <- sqlFetch(con,"port")
bond <- sqlFetch(con,"bond")

```

The Rterm window shows the following error message:

```
[原来保存的工作空间已还原]
错误: 找不到对象'.r'
错误于source(.trPaths[5],
echo = TRUE, max.deparse
.length = 150) :
找不到对象'.trPaths'
```

Tinn-R



简单计算

在控制台中键入如下命令

```
2+2
```

```
a<-2
```

<-也可用=, ->代替

```
b<-2
```

```
a+b->c # 注释
```




对象

向量 (vector)

因子 (factor) 分类变量, 如
"red", "blue", "green"

矩阵 (matrix)

数据框 (dataframe) 是由一个或几个向量和 (或) 因子构成, 它们必须是等长的, 但可以是不同的数据类型。

列表 (list) 列表可以包含任何类型的对象。



运算符

数学运算

$+, -, *, /, ^,$

比较运算

$>, <, <=, >=, ==, !=$

逻辑运算

! 非

& 且

| 或



R中的函数

R是一种解释性语言，不用编译，语法简单而直观。

函数形式

`function (选项)`

平均值 `mean()`

线性回归 `lm(y~x1+x2, data=data1)`

R处理的所有数据、变量、函数和结果都以**对象**的形式保存。

```
result <- lm(y~x1+x2, data=data1)
```



编程基础

R语言编程时无需声明变量的类型，这与C, C++等语言不同。

函数定义的基本格式

```
myfunction <-function(x, y)
{
    x+y #表达式
}
```

程序流程控制 if

if 语句的调用格式

if(条件) 表达式

if(条件) 表达式1 **else** 表达式2

举例:

```
if(score<60)  
    print("不及格")
```

```
if(score<60)  
    print("不及格") else  
    print("及格")
```



for循环

for 用法

```
for(变量 in 向量) { 表达式 }
```

举例

```
for(i in 1:10) {  
  t<-sqrt(i)  
  print(t)  
}
```



while 循环

while 用法

```
while(条件) { 表达式 }
```

举例

```
sum<-0; i <- 1  
while (i<=100) {  
  sum <- sum + i  
  i<-i+1  
}  
sum
```



函数举例

定义函数:

```
myfun<-function(x,y)
{
  z <- x + y^2
  result<-z*sqrt(z)
  result
}
```

调用函数:

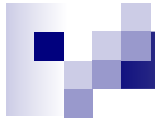
```
myfun(3,4)
```




推荐的教材

- 自带文档
 - An Introduction to R* 有中文版
- *153 分钟学会 R*
<http://cos.name/bbs/read.php?tid=7673>
- *Quick R* <http://www.statmethods.net/>
- 汤银才 《R语言与统计分析》
-

- 其它
 - <http://cran.r-project.org/manuals.html>
 - <http://www.r-project.org/doc/bib/R-books.html>



谢谢 😊