第1章

基本观念

1-1：Big Data的起源

1-2：R语言之美

1-3：R语言的起源

1-4：R的执行环境

1-5：R的扩展

1-6：本书学习目标

1-1：Big Data的起源

Big Data一词，有人解释为大数据，也有人解释为巨量资料，其实都OK，本书则以大数据为主要用法。

2012年世界经济论坛在瑞士达沃夫(Davos)有一个主要议题”Big Data, Big Impact.”，同年纽约时报(The New York Times)的一篇文章，How Big Data Became So Big，清楚揭露大数据时代已经降临，它可以用在商业、经济和其他领域中。



本图片取材自The New York Times

1-2：R语言之美

大数据需处理的数据是广泛的，基本上可分成二大类，有序数据与无序数据，对于有序数据，目前许多程序语言已可处理。但对于无序数据，例如，地理位置信息，脸书讯息，视讯数据….等，是无法处理。而R语言正可以解决这方面的问题，自此R已成为有志成为信息科学家(Data Scientist)或大数据工程师(Big Data Engineer)所必需精通的计算机语言。

Google首席经济学家Hal Ronald Varian有一句经典名言形容R。



本图片取材自Wikipedia

The Great beauty of R is that you can modify it to do all sorts of things. And you have a lot of prepackaged stuff that’s already available, so you’re standing on the shoulders of giants.

大意是，R语言之美在于，你可以透过修改很多高手已经写好的套件程序，解决各式各样的问题。因此，当你使用R语言时，你已经站在巨人的肩膀了。

1-3：R语言的起源

提到R语言，不得不提John Chambers。他是加拿大多伦多大学毕业，然后拿到哈佛大学统计硕士和博士。



John Chambers本图片取材自网络

John Chambers在1976年于Bell实验室工作时，为了节省使用SAS和SPSS软件经费，以Fortran为基础，开发了S语言。这个S语言主要是处理，向量(Vector) 、矩阵(Matrix) 、数组组(Array)以及进行图表和统计分析，初期只是可以在Bell实验室的系统上执行，随后这个S语言被移植至早期的UNIX系统下执行。然后Bell实验室以很低廉价格授权各大学使用。

R语言主要是以S语言为基础，开发完成。

1993年新西兰University of Auckland大学统计系的教授Ross Ihaka和Robert Gentleman两位R先生，为了方便教授统计学以S语言为基础开发完成一个程序语言，因为他两人名前缀字皆是R，于是他们所开发的语言就称R语言。

Ross Ihaka本图片取材自网络 Robert Gentleman本图片取材自网络



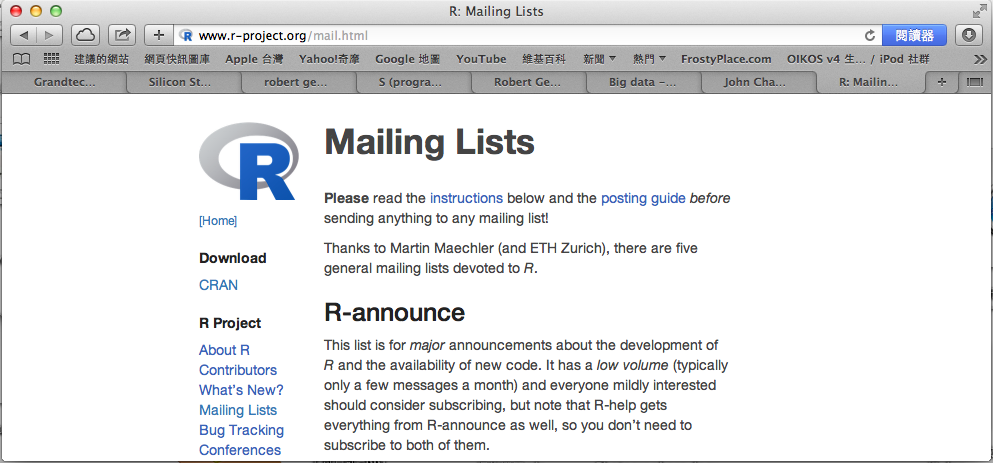
R语言标准logo

现在的R语言则有一个R核心开发团队负责，当然Ross Ihaka和Robert Gentleman是这个开发团队成员，另外，S语言开发者John Chambers也是这个R语言开发团队成员。目前这个开发团队共有18个成员，这些成员拥有修改R核心代码的权限。下列是R语言开发的几个有意义的时间点。

1：1990年代初期R语言开发

2：1993年Ross Ihaka和Robert Gentleman 开发了R语言软件，在S-news邮件中发表。吸引的一些人关注并和他们合作，自此一组针对R的邮件被建立。

如果你想了解更多这方面的讯息可参考下列网址。



3：1995年6月在Martin Maechler等人的努力下，这个R语言同意免费供使用，同时遵守自由软件基金会(Free Software Foundation)的GNU General Public License(GPL) Version 2的协议。

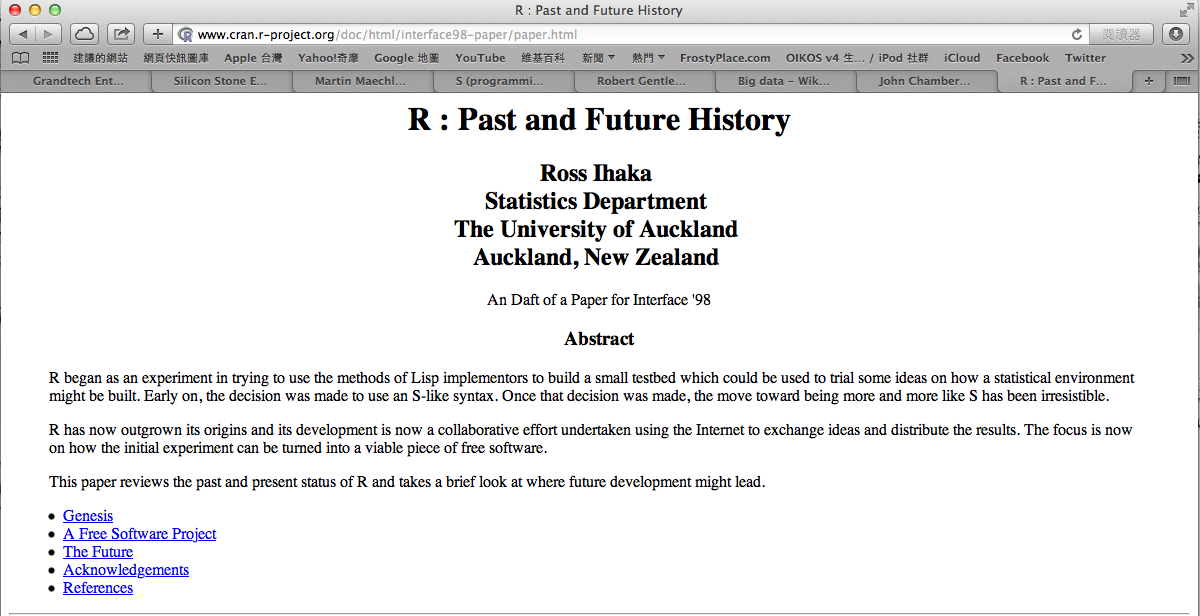


Dr. Martin Maechler取材自stat.ethz.ch/people/maechler

4：1997年R语言核心开发团队成立。

5：2000年第1版R1.0.0正式发表。

Ross Ihaka有将R开发简史记录下来，可参考下列网址。



1-4：R的执行环境

在R语言核心开发团队的努力下，目前R语言已可以在常见的各种操作系统下执行。例如，Windows、Mac OS、Unix和Linux。

1-5：R的扩展

R的一个重要优点是，R是Open Source License，这表示任何人均可下载并修改，因此许多人在撰写增强功能的套件，同时供应他人免费使用。

1-6：本书学习目标

不容否认，不论是S语言或R语言皆是统计专家所开发的，因此，R具有可以处理各种统计的工具。但已有越来越多的程序设计师开始加入学习R，使得R也开始可以处理非统计方面的工作，例如，数据处理、图形处理、心理学、遗传学、生物学 …. 市场调查等等。

本书在撰写时，尽量将读者视为初学者，辅以丰富实例，期待读者可以用最轻松方式学会R语言。

本章习题

**一：是非题**

( O )1：要成为大数据工程师(Big Data Engineer)，学习R语言是一件很重要的事。

( X )2：脸书(Face book)讯息、视讯数据是可排序的数据。

( X )3：R语言目前只能在Windows和Mac OS系统下执行。

( O )4：R语言是免费软件。

**二：选择题**

( C )1：R语言无法在以下那一个系统下执行。

A：Linux

B：Unix

C：Android

D：Mac OS

( A )2：下列那一个人对R语言的开发比较没有贡献。

A：Steve Job

B：Ross Ihaka

C：John Chambers

D：Robert Gentleman

( B )3：R语言是以那一个语言为基础开发完成。

A：SAS

B：S

C：SPSS

D：C

**三：复选题**

( A, B, C )1：我们现在可以免费使用R语言，下列那些人是有贡献的。(选择3项)

A：Martin Maechler

B：Ross Ihaka

C：Robert Gentleman

D：Tim Cook

E：Marissa Mayer