

Welcome to the First Korean R Meeting!

Duncan Murdoch

Department of Statistical and Actuarial Sciences
University of Western Ontario

October 28, 2011

- Chel Hee Lee has asked me to give a short talk to welcome you to the conference.
- I will talk about R's past, present, and future.

- Chel Hee Lee has asked me to give a short talk to welcome you to the conference.
- I will talk about R's past, present, and future.
- But first, what is R?

R is Free

R is a free software environment for statistical computing and graphics.

- A GNU project distributed under the GPL.
- Runs on a wide variety of platforms, including UNIX-like ones (Linux, Solaris, Mac OS, etc.) as well as MS Windows.
- It is highly extensible, with thousands (!) of user-contributed packages available.

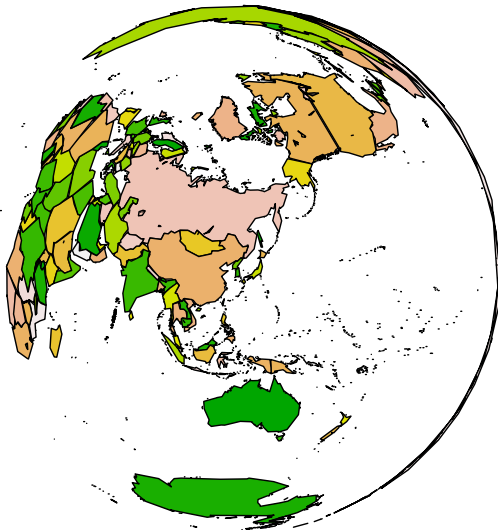
R is an Environment for Statistical Computing and Graphics

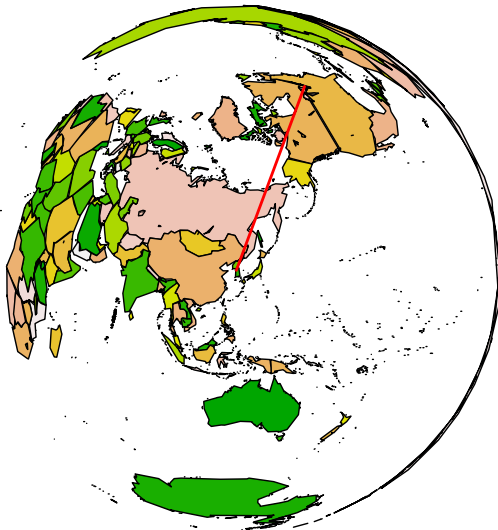
- Data handling and storage
- Calculations on vectors, matrices and more general arrays and structures
- Tools for data analysis
- Graphical support for interactive display and publication quality printing
- A well-developed programming language
- Software development support, including documentation and testing

Example: Maps

```
> library(maps)
> map("world",
+     projection = "azequalarea",
+     orientation = c(37.57, 126.98, 0),
+     col = sample(terrain.colors(100)),
+     fill =TRUE)
>
```

The maps package is by Ray Brownrigg, based on S code by Rick Becker and Allan Wilks.





R History

R is nearly 20 years old. How did it start?

First There Was S...

S began as a project at Bell Laboratories in 1976, involving John Chambers, Rick Becker, Doug Dunn, Paul Tukey, and Graham Wilkinson.

We wanted users to be able to begin in an interactive environment, where they did not consciously think of themselves as programming. Then as their needs became clearer and their sophistication increased, they should be able to slide gradually into programming, when the language and system aspects would become more important.

— John Chambers, in [Stages in the Evolution of S](#)

The Birth of R

Around 1992, Ross Ihaka and Robert Gentleman at the University of Auckland needed statistical software for a teaching lab. S-PLUS (a commercial implementation of S) didn't run on MacOS; they decided to write their own. Ross was very impressed with Scheme (a Lisp dialect); they based the initial code on the design of Scheme interpreters.

To make the interpreter useful, we had to add data structures to support statistical work and to choose a user interface. We wanted a command driven interface and, since we were both very familiar with S, it seemed natural to use an S-like syntax.

— Ross Ihaka, in [R: Past and Future History](#)

GNU S

- Ross and Robert announced R on the `s-news` mailing list in 1993; Martin Maechler convinced them to release it as an open source project under the GPL, which they did in 1995.
- In 1997 the R Core group was established, consisting of Doug Bates, Peter Dalgaard, Robert Gentleman, Kurt Hornik, Ross Ihaka, Friedrich Leisch, Thomas Lumley, Martin Maechler, Paul Murrell, Heiner Schwarte, and Luke Tierney.
- Since then John Chambers, Seth Falcon, Stefano Iacus, Guido Masarotto, Duncan Murdoch, Martyn Plummer, Brian Ripley, Duncan Temple Lang, Uwe Ligges, Deepayan Sarkar and Simon Urbanek have joined R Core. Guido Masarotto and Heiner Schwarte have left.

The R Community

- The main form of communication is through the `R-help` and `R-devel` mailing lists. There are also a number of specialized mailing lists dedicated to particular topics.
- Since 1999, R has dominated the DSC series of conferences in Vienna, Seattle, Auckland and Copenhagen.
- There have been 7 useR! conferences in Vienna, Ames, Dortmund, Rennes, Gaithersburg and Coventry.
- Bioconductor has held numerous conferences, courses and workshops on using R for statistical genomics and related topics.

Contributed Packages!

- There are about 3350 contributed packages on CRAN, 1100 packages in Bioconductor, many other packages in private use.
- Publishing an R package has become an accepted (almost expected) way to disseminate statistical research.

Why Has R Succeeded?

Many things have helped R:

- The initial design of S was good, and R is better: this has kept some smart people interested from the beginning.

Why Has R Succeeded?

Many things have helped R:

- The initial design of S was good, and R is better: this has kept some smart people interested from the beginning.
- Statistical packages like SPSS and SAS are hard to extend: researchers prefer S and R.

Why Has R Succeeded?

Many things have helped R:

- The initial design of S was good, and R is better: this has kept some smart people interested from the beginning.
- Statistical packages like SPSS and SAS are hard to extend: researchers prefer S and R.
- Many people (especially Ross Ihaka, Robert Gentleman and Brian Ripley) have given a huge amount of time to R, and we all benefit.

Why Has R Succeeded?

- Licenses for technical software are expensive and restrictive.

Why Has R Succeeded?

- Licenses for technical software are expensive and restrictive.
- R is managed by developers, not sales managers. We help users to contribute.

Why Has R Succeeded?

- Licenses for technical software are expensive and restrictive.
- R is managed by developers, not sales managers. We help users to contribute.
- R users are very knowledgeable; the online community is very helpful.

Why Has R Succeeded?

- Licenses for technical software are expensive and restrictive.
- R is managed by developers, not sales managers. We help users to contribute.
- R users are very knowledgeable; the online community is very helpful.
- R has always tried to run universally: Unix, Linux, Windows, Mac OS, ...

Why Has R Succeeded?

- Licenses for technical software are expensive and restrictive.
- R is managed by developers, not sales managers. We help users to contribute.
- R users are very knowledgeable; the online community is very helpful.
- R has always tried to run universally: Unix, Linux, Windows, Mac OS, ...
- We've been lucky.

Resources

R development is almost entirely by volunteers.

- The R core team are all volunteers. Most of us are academics, but several work in industry.
- Most contributed packages are written by volunteers. There are a few written by commercial entities (e.g. Revolution Analytics).
- Translation teams make R available in many languages: Chinese, Danish, Dutch, French, German, Greek, Italian, Japanese, Korean, Norwegian, Portuguese, Russian, Spanish, Turkish. The Korean one is led by Chel Hee Lee, who has also organized this conference.
- The R Foundation was set up to hold copyrights and to fund R development.

R's Future

What can we expect to see from R in the future?

Core Team Plans

- Support for multi-core machines in package *parallel* in R 2.14.0 in a few days.
- Byte code compilation of base packages will happen by default in 2.14.0, and will be expanded and made more efficient over time.
- Support for bigger vectors is coming. Currently vectors are limited to 2^{31} entries; this will likely be increased to 2^{51} entries or more in 2.15.0 next spring.
- After 2.15.0, major releases will happen once per year.

Contributions from the Community

- Contributed packages will continue to be the easiest way to contribute to the R project.
- There are more than 100 books on R; there will be more!
- Translation teams like Chel Hee Lee's will make R available to more people around the world.

More meetings like this!

The R Project depends on enthusiastic participants. On behalf of the R Core Team, I would like to thank Chel Hee Lee and all of you for making this meeting a success.