A bit of History of R

R Intro

Gaston Sanchez

STAT 33B, Fall 2025

About R

R entails both:

- An Environment for Statistical Computing, Data Analysis and Graphics
- A Programming Language

Why R

- Allows custom analysis
- High-level scripting language
- Statistical programming language
- Interactive exploratory data analysis
- Easy to replicate analysis
- Sound numerical methods
- Large community of contributors
- Free and open source

Some Comments

- R is a free implementation of a dialect of the **S** language
- Cosmetically, R was developed to look a lot like S
- Under the hood, R and S have important differences

A little bit about S
(on which R is based)

About S

- S is the statistics and graphics environment created in 1970s by John Chambers and colleagues at Bell Labs
- S was designed to blur the distinction between users and programmers
- S is a system for **interactive** data analysis

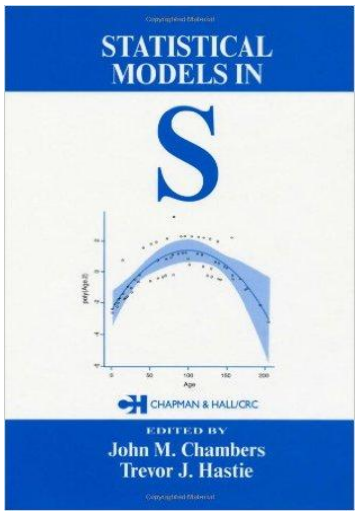
About S

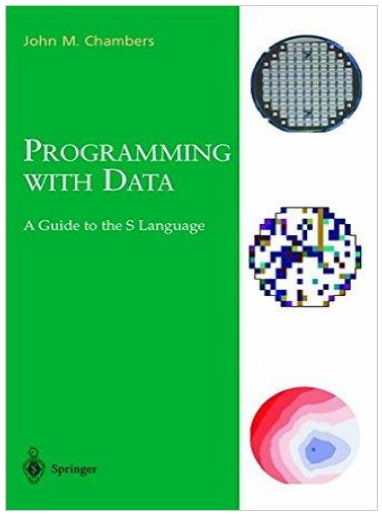
S is the statististics and graphics environment created by John Chambers and colleagues

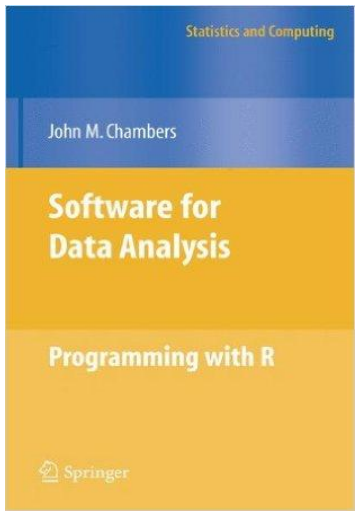
John Chambers
(main creator of S)
Harvard PhD

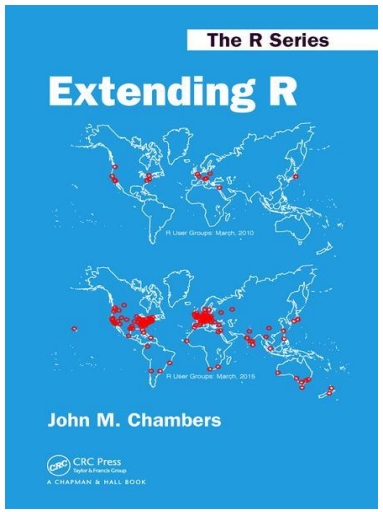
https://en.wikipedia.org/wiki/John_Chambers_(statistician)

Main books by John Chambers









Some 1970s inventions at Bell Labs AT&T

- More and more computer related inventions as part of the personal computer revolution
- 1970s Dennis Ritchie and Ken Thompson developed the **UNIX** operating system
- 1972 Dennis Ritchie developed the compiled programming language **C**
- 1973 Ken Thompson developed **GREP** (regular expressions)
- 1974 Lee McMahon developed **SED** (stream editor)
- 1977 Aho, Weinberger and Kernighan developed **AWK** (text processing)

History of S

- May 5, 1976 at Bell Labs
- Group of 5 researchers brainstormed designing a system for statistical computing
- System designed to serve the needs of the statistics research group at Bell Labs
- Preliminary version available by the end of 1976
- This is why you'll often hear that *"S was created by statisticians for statisticians"*

History of S

- No agreement on the suggested names
- Although they all contained the letter **S**
- Inspired by the recently designed language **C**, they reached an agreement with **S**
- Portability was a main concern: it was decided to make a UNIX version of S, **portable** wherever UNIX was

History of S

- Due to an antitrust case, AT&T was forbidden to enter the software industry
- AT&T had to license any software pattents
- 1973 AT&T licensed UNIX to educational institutions
- It also started to license S to universities and research laboratories worldwide

Philosophy of S

Keys to keep in mind:

- S was designed as an environment for statististics and graphics
- S was designed to blur the distinction between users and programmers
- S was designed as a system for interactive data analysis

Creators of R

Ross Ihaka
(New Zealand)
Berkeley PhD

Robert Gentleman
(Canada)
UW PhD

- https://en.wikipedia.org/wiki/Ross_Ihaka
- https://en.wikipedia.org/wiki/Robert_Gentleman_(statistician)

History of R

Ross Ihaka's papers (1998) about some of the history of R:

- Genesis
  https://cran.r-project.org/doc/html/interface98-paper/paper_1.html
- A Free Software Project
  https://cran.r-project.org/doc/html/interface98-paper/paper_2.html
- The Future
  https://cran.r-project.org/doc/html/interface98-paper/paper_3.html

The University of Auckland, New Zealand

https://www.calendar.auckland.ac.nz/en/info/campuses.html

History of R

- Ross Ihaka and Robert Gentleman (RR) became colleagues at The University of Auckland (New Zealand).
- They had an interest in statistical computing and saw a common need for a better software environment that could be used for teaching statistics.
- Because they saw no suitable commercial environment for their teaching of statistics purposes, they began to experiment with developing one by themselves.

History of R

- RR really liked **S**, and started to implement a Scheme-like interpreter—written in **C**—to have a language with an S-like syntax.
- Despite the similarity between R and S, the main key differences are in terms of memory management and scoping rules.
- RR began to use the first experimental versions for their teaching purposes, and released (1993) some binary copies of R in the S-news mailing list
- And yes, **R** is named after RR

Beginnings of R

- August 1993: Binary copies of R shared on the s-news mailing list
- Martin Maechler (ETH Zurich) encouraged R&R to release R source code as free software
- June 1995: R released under GNU general license
- Kurt Hornik of *TU Wien* established the main archive (CRAN) in Austria
- mid-1997: R Core Group (developers) was established

Interactive Use

- R also follows the idea of **interactive** data analysis
- *Interactive*: as having a dialogue with the computer
- You type one or more commands, execute them, and get the results, i.e. ask questions, get answers

In Summary

- Building of S reflected many UNIX features
- 1980s saw a redesigning of S
- 1990s RR designed a system compatible with S
- RR joined the open-source movement
- Self-managing group of volunteers took over the development of R: **R Core group**

In Summary

- R provides an environment for statistical computing
- R also has a programming language
- Free Software
- Open Source
- Extensible with packages