

## E-book – A História da Linguagem R

### Como tudo começou: a linguagem S.

S é a linguagem que foi desenvolvida por *John Chambers* nos laboratórios da Bell Labs, originalmente parte da AT&T. S foi inicialmente desenvolvida em 1976 como um ambiente de análise estatística, originalmente implementada usando bibliotecas da linguagem de programação Fortran. Entretanto, as primeiras versões da linguagem não continham funções de modelagem estatística.

Em 1988 o sistema foi reescrito em linguagem C e começou a ter o formato que existe atualmente (já era a versão 3 da linguagem). O livro *Statistical Models in S* de *Chambers* e *Hastie* documenta as funcionalidades de análise estatística. A versão 4 de S foi liberada em 1998, sendo a versão que existe atualmente. *Chambers* escreveu o livro *Programming with Data* para documentar esta versão da linguagem.

Em 1993, a Bell Labs concedeu à empresa Insightful Corp (que à época tinha o nome de StatSci) uma licença exclusiva para desenvolver e vender a linguagem S. Em 2008 a Insightful foi adquirida pela TIBCO por 25 milhões de dólares. A TIBCO é atualmente a proprietária da linguagem S. Os fundamentos da linguagem S praticamente não mudaram desde a versão lançada em 1998. Em 1998 aliás, S ganhou um prêmio por sua preciosa contribuição no campo da ciência da computação. O prêmio concedido a S, dizia:

“...irá mudar para sempre a forma como as pessoas analisam, visualizam e manipulam dados. S é elegante, amplamente aceita e possui os conceitos principais de sistemas de software, como integridade....obrigado John Chambers por esta ideia e esforço...”

Outros softwares que ganharam o mesmo prêmio: Unix, TeX, TCP/IP, Web, Apache, Make.

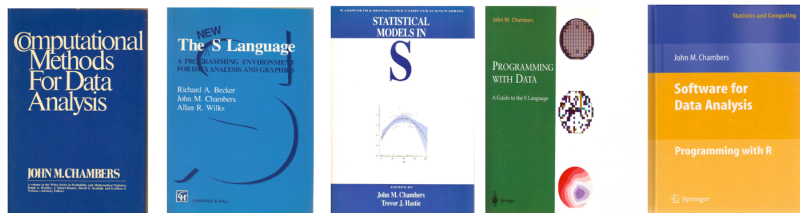
Entender a filosofia da linguagem S, ajuda a entender as raízes que geraram a linguagem R. O fator mais importante a notar, é que a linguagem S nasceu como uma ferramenta de análise de dados e não como uma linguagem de programação tradicional. Os inventores de S tinham como objetivo tornar a análise de dados mais fácil, inicialmente para eles e então para os outros.

Veja o que disse *John Chambers* no livro *Stages in the Evolutions of S*:

*“Nós queremos que os usuários tenham um ambiente interativo, onde eles não pensem conscientemente como se estivessem programando. À medida que as necessidades se tornam mais claras e a sofisticação aumenta, eles devem estar aptos a gradualmente iniciar a programação, quando a linguagem e aspectos de sistema se tornarem mais importantes.”*

Em resumo, o objetivo era criar um ambiente que servisse tanto a usuários, quanto a desenvolvedores. A linguagem deveria ser interativa para análise de dados (através de linha de comando) e útil para escrever programas que poderiam se tornar sistemas de produção.

A história de R através de livros



Chambers, *Computational Methods for Data Analysis*. Wiley, 1977.

Earliest publication of what became S.

Becker, Chambers, and Wilks. *The New S Language*. Chapman & Hall, 1988.

Introduced what is now known as S version 3 (S3)

Chambers and Hastie. *Statistical Models in S*. Chapman & Hall, 1992.

Statistical modeling in S; S3 version of classes and methods.

Image label

Chambers. *Programming with Data*. Springer, 1998.

Version 4 of S, a major revision of S.

Chambers. *Software for Data Analysis: Programming with R*. Springer, 2008

The R version of S4 and other R techniques.

## E como nasceu a linguagem R?

A linguagem R surgiu um pouco depois da linguagem S. Uma das limitações de S era o fato da linguagem estar disponível apenas através do pacote comercial S-PLUS. Em 1991, R foi criada por Ross Ihaka e Robert Gentleman no Departamento de Estatística da Universidade de Auckland. Em 1993 a linguagem R foi anunciada em público pela primeira vez. Em 1995, Martin Mahler fez uma importante contribuição, convencendo Ross e Robert a usarem a licença GNU General Public e tornar R um software livre. Isso permitiu que o código fonte da linguagem R se tornasse disponível para toda a comunidade.

Em 1996 as listas públicas R-help e R-devel foram criadas e em 1997 foi formado o grupo R Core, com profissionais associados ao S e S-PLUS, estatísticos e cientistas da computação. Atualmente o grupo R Core controla o código fonte de R. Em 2000, finalmente a versão 1.0.0 do R foi liberada ao público. Em 21 de Junho de 2016, foi lançada a versão mais recente do R, a versão 3.3.1, que usaremos aqui no curso.

## Limitações

R também possui limitações. R foi baseada em tecnologia desenvolvida muitas décadas atrás, que deram origem a linguagem S. Praticamente não existe

suporte nativo a gráficos dinâmicos e 3D. Outra limitação de R é o gerenciamento de memória física. Devido regras de escopo da linguagem, R geralmente perde no quesito gerenciamento de memória, para outros pacotes estatísticos comerciais. Muitos avanços têm ocorrido nesta área nos últimos anos e embora o problema de memória ainda exista, foi minimizado pela capacidade dos novos hardwares com alta capacidade de memória.

Dirk

### **O que é o R afinal?**

O R é um software livre, o que significa que não tem custos e que seu código fonte está acessível para qualquer usuário. Ele pode ser obtido em seu site oficial:

<http://www.r-project.org>

Basta acessar o link ‘CRAN’ no lado esquerdo da tela, selecionar um dos endereços para fazer download, e selecionar o sistema operacional. Supondo que você está usando um computador com Windows, clique em ‘Download R for Windows’, em seguida em ‘base’ e, por fim, no link para download que aparecerá no alto da tela. O R recebe atualizações constantes, e se você desejar instalar uma versão mais recente, saiba que ela não substituirá a versão anterior, que deve ser desinstalada manualmente pelo usuário (ou mantida, o que pode ser útil para utilizar pacotes antigos que não existam nas versões mais novas).

Nunca se esqueça de que o R possui uma quantidade imensa de material sobre seu uso nas mais diversas áreas da ciência e buscar material é bastante importante como parte do processo de aprendizado. Na página inicial da linguagem R há alguns manuais disponíveis (no link ‘Manuals’, no lado esquerdo), que podem ser

um bom ponto de partida. Quando desejar buscar na internet por algum material específico (como uma análise ou conjunto de análises de uma determinada área), um caminho inicial é buscar por: CRAN “análise desejada”.

Site da documentação oficial do R: <https://cran.r-project.org/manuals.html>

O R disponibiliza uma grande variedade de métodos estatísticos (modelagem linear e não-linear, testes estatísticos clássicos, séries temporais, classificação, métodos multivariados, etc) e técnicas gráficas. Um dos pontos fortes do R é a facilidade com que gráficos bem delineados e de alta qualidade para impressão podem ser produzidos com possibilidade de inclusão de fórmulas e símbolos matemáticos quando necessário.

R se presta a diversas funções, desde uma calculadora científica, passando pela integração e derivação de funções matemáticas, até a realização de complexas análises estatísticas. Além disso, o R também apresenta uma série de recursos gráficos que permitem a descrição detalhada de todos os aspectos que se pode querer personalizar em um gráfico, como cor, tipo e tamanho de letra, símbolos, títulos e subtítulos, pontos, linhas, legendas, planos de fundo e muito mais.

Mais que um software que realiza análises estatísticas, R é um ambiente e uma linguagem de programação orientada a objeto. Nele, números, vetores, matrizes, arrays, data frames e listas podem ficar armazenados em objetos. Pode-se entender objeto como uma caixinha onde você pode guardar o que quiser. A partir daí todas as operações matemáticas podem ser feitas usando esses objetos. Isso torna as coisas mais simples.

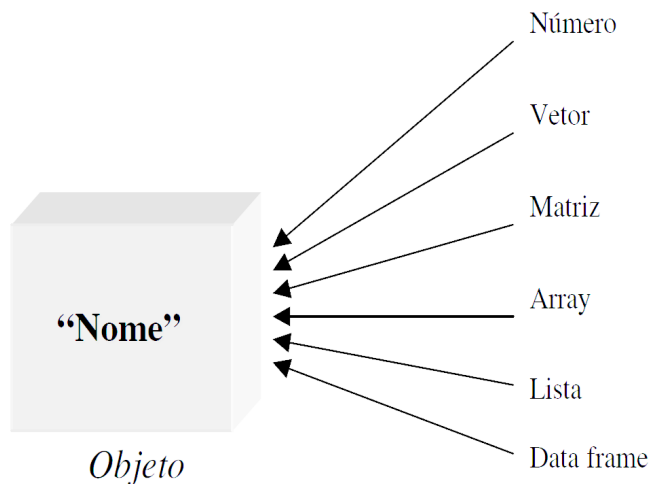


Image label

O ideal para aprender a usar o R é "usá-lo!". Então, a melhor forma de se familiarizar com os comandos do R é assistir os vídeos que iremos disponibilizar e ao mesmo tempo ir digitando os comandos no R e observando os resultados, gráficos, etc. Apenas assistir os vídeos talvez não o ajude a fazer progressos no seu aprendizado do R. É preciso usar e sentir a linguagem. Aprender a usar o R pode ser difícil e trabalhoso, mas lembre-se, o investimento será para você!

*John Chambers* escreveu no prefácio de seu livro: "Será que é proveitoso gastar tempo para desenvolver e estender habilidades em programação? Sim, porque o investimento pode "contribuir" com sua habilidade em formular questões e na confiança que você terá nas respostas".

### Em resumo, R é?

- Software e ambiente de análise de dados;
- Linguagem de programação estatística desenvolvida por estatísticos para estatísticos;
- Um ambiente interativo com grande variedade de bibliotecas e algoritmos de acesso a dados, manipulação de dados, análise e gráficos;

- Um projeto open-source: livre, aberto e ativo;
- Uma comunidade de milhões de usuários (estimada em 2 milhões) e pessoas em todo o mundo contribuindo nas mais diversas áreas de conhecimento;
- As funções do R são escritas em ... R! Isto permite verificar facilmente o que as funções realmente fazem

### E por que aprender R?

*Se você estiver interessado em uma carreira na qual os serviços estarão em alta demanda, você deve buscar alguma forma de prover um serviço escasso, que seja complementar a algo que esteja crescendo em quantidade e de baixo custo. O que está crescendo em quantidade, sendo de baixo custo? **Dados**. Qual o serviço escasso e complementar aos dados? **Análise**!*

Fonte: Hal Varian, Freakonomics blog,

<http://freakonomics.blogs.nytimes.com/2008/02/25/hal-varian-answers-your-questions/>

E mais:

*A profissão de Cientista de Dados será a mais sexy nos próximos 10 anos. Acha que é uma piada? Pois quem diria que os engenheiros de computação teriam o trabalho mais sexy dos anos 90?*

Fonte: Hal Varian, The McKinsey Quarterly,

[http://www.mckinseyquarterly.com/Hal\\_Varían\\_on\\_how\\_the\\_Web\\_challenges\\_managers\\_2286](http://www.mckinseyquarterly.com/Hal_Varían_on_how_the_Web_challenges_managers_2286)

Seja bem-vindo ao curso de Big Data Analytics com R e Microsoft Azure Machine Learning!

Equipe Data Science Academy

Links úteis:

Manuais oficiais do R: <https://cran.r-project.org/manuals.html>

Listas de E-mail do R: <https://www.r-project.org/mail.html>

R-bloggers: <http://www.r-bloggers.com>

R-Seek: <http://rseek.org>

CRAN (Comprehensive Archive R Network): <https://cran.r-project.org>