

Programação em R

Parte I

Autor: Prof. Dr. Pedro Rafael Diniz Marinho

Universidade Federal da Paraíba
Departamento de Estatística da UFPB

Plano de Curso

A ementa está disponível no site do Departamento de Estatística da UFPB.

O plano de curso é de responsabilidade do professor da disciplina e deve estar de acordo com a ementa da disciplina.

O plano de curso é estabelecido pelo professor e ficará a cargo desse a sua construção e alteração, se necessário, no decorrer do curso, desde que o mesmo esteja de acordo com a ementa do curso.

O rigor e profundidade do assunto ficará a cargo do professor e faz parte do seu plano de curso.

Fiquem atentos ao SIGAA UFPB

É aconselhável que o aluno acesse a disciplina no SIGAA - UFPB (SIGAA) com frequência. Todo tipo de notícia, informações e materiais de apoio serão postados no SIGAA.

Fiquem atentos ao SIGAA UFPB

É aconselhável que o aluno acesse a disciplina no SIGAA - UFPB (SIGAA) com frequência. Todo tipo de notícia, informações e materiais de apoio serão postados no SIGAA.

Caso seja necessário entrar em contato com os alunos sobre reposições de aulas e/ou não ocorrência de alguma aula, estas informações estarão, **SEM FALTA**, registradas no SIGAA.

Fiquem atentos ao SIGAA UFPB

É aconselhável que o aluno acesse a disciplina no SIGAA - UFPB (SIGAA) com frequência. Todo tipo de notícia, informações e materiais de apoio serão postados no SIGAA.

Caso seja necessário entrar em contato com os alunos sobre reposições de aulas e/ou não ocorrência de alguma aula, estas informações estarão, **SEM FALTA**, registradas no SIGAA.

Inclusive, todas as aulas (arquivos PDF's utilizados na projeção) serão postados na plataforma.

Recursos utilizados

Boa parte do curso será apoiada pelo uso de **datashow** o que nos ajudará bastante a decorrer sobre os diversos assuntos contidos na ementa desse curso que é bastante ampla.

O quadro será utilizado para resolução de alguns exemplos bem como complementações em que o professor achar conveniente no momento de aula.

Sobre as avaliações

No curso iremos considerar **três avaliações** (provas).

As datas das avaliações estão registradas no SIGAA.

O aluno não poderá utilizar o computador nas avaliações.

Aos alunos interessados (\LaTeX)



Figura: Donald Knuth.

\TeX é um sistema de tipografia científica desenvolvido por **Donald E. Knuth** que é orientado à produção de textos técnicos e fórmulas matemáticas. A pedido da AMS (*American Mathematical Society*), Donald Knuth desenvolveu uma linguagem de computador para editoração de textos com muitas equações.

Aos alunos interessados (\LaTeX)



Figura: Donald Knuth.

O trabalho de criação do \TeX se estendeu de 1977 a 1998, quando \TeX foi disponibilizado gratuitamente. O \TeX possui aproximadamente **600 comandos** que controlam a construção de uma página.

Pode-se considerar o \TeX como sendo um compilador para textos científicos que produz documentos de alta qualidade tipográfica.

Aos alunos interessados (\LaTeX)

O \TeX atingiu um estado de desenvolvimento em que Beebe (1990 afirmou):

“Meu trabalho no desenvolvimento de \TeX , METAFONT e as fontes Computer Modern chegou ao final. Eu não irei realizar mudanças futuras, exceto corrigir sérios erros de programação.”

Ver em: BEEBE, N. H. Comments on the future of \TeX and METAFONT. TUGboat, v. 11, n. 4, p. 490–494, 1990.

Aos alunos interessados (\LaTeX)

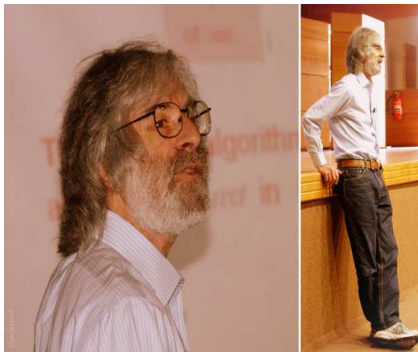


Figura: Leslie Lamport.

Quase que em paralelo foi desenvolvido por Leslie Lamport o \LaTeX . Essas macros definem tipos de documentos, tais como livros, artigos, cartas, entre outros.

Inclusive essa apresentação é um tipo básico de documento que foi produzido em \LaTeX .

Aos alunos interessados (\LaTeX)

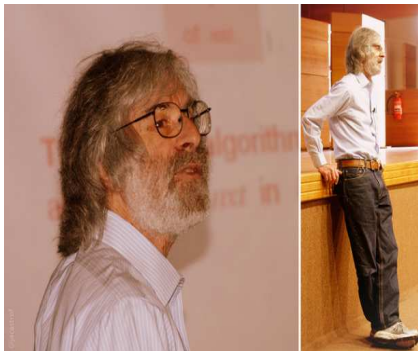


Figura: Leslie Lamport.

Para maiores detalhes leia sobre o pacote beamer que está disponível com a maioria das distribuições \LaTeX . Há diversos documentos disponíveis nos mais variados idiomas na rede.

beamer também está disponível no **The Comprehensive \TeX Archive Network (CTAN)**.

Aos alunos interessados (\LaTeX)



Figura: CTAN lion drawing
by Duane Bibby.

CTAN é o lugar central para todos os tipos de material em torno de \TeX . CTAN tem atualmente **5263 pacotes** e **2409 colaboradores** contribuíram para essa quantidade de pacotes.

O símbolo ao lado foi desenhado pelo artista comercial Duane Bibby. Este leão foi utilizado nas ilustrações para o livro \TeX Book de Donald Knuth e apareceu com grande frequência em outros materiais.

Aos alunos interessados (\LaTeX)



Figura: CTAN lion drawing
by Duane Bibby.

Maiores detalhes sobre o CTAN podem ser encontrados em <https://www.ctan.org/lion/>. Desde dezembro de 1994, a pacote \LaTeX está sendo atualizado pela equipe \LaTeX 3, dirigida por Frank Mittelbach, para incluir algumas melhorias que já vinham solicitadas a algum tempo. A equipe se preocupa também em reunificar todas as versões modificadas que surgiram desde o aparecimento do \LaTeX 2.09.

Aos alunos interessados (\LaTeX)



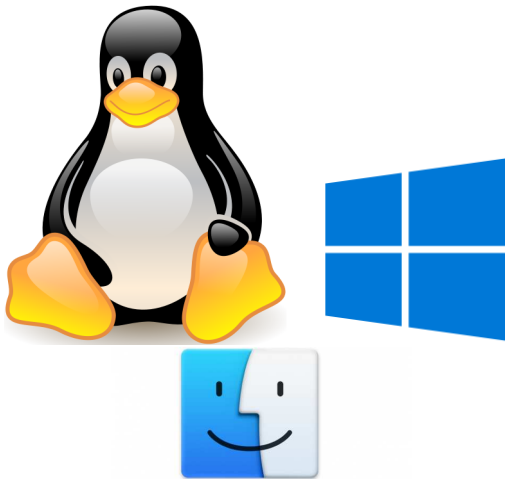
Figura: CTAN lion drawing
by Duane Bibby.

O melhor de tudo, o \LaTeX é um sistema estável mas com crescimento constante, podendo ser instalado em quase todos os sistemas operacionais.

Usuários de Unix, Linux, Windows ou Mac OS X podem dispor de todo ferramental para produzir ótimos textos com o \LaTeX .

Nota: Pronuncia-se “leitec” e não “latéx”.

Aos alunos interessados (\LaTeX)



MacTM OS

Como instalo o \LaTeX no Linux?



Figura: Tux (Mascote do Linux).

Como instalo o \LaTeX no Linux?

Inicialmente é preciso instalar o compilador de \LaTeX . Recomento o uso do \TeX Live.

A maioria das distribuições linux (Arch, Ubuntu, Fedora, Mint, Sabayon, entre outros “sabores”) apresentam esse compilador de \LaTeX em seus repositórios.

Por exemplo, no **Arch Linux** e distribuições derivadas que utilizam os mesmos repositórios do Arch como Antergos façam:

```
sudo pacman -S texlive.
```

Como instalo o \LaTeX no Linux?

No **Ubuntu** ou qualquer distribuição que faz uso dos repositórios do Ubuntu façam:

```
sudo apt-get install texlive-full.
```

Já os usuários da distribuição **Fedora** e distribuições derivadas que utilizam-se dos mesmos repositórios devem fazer:

```
sudo dnf -y texlive-scheme-full.
```

Observação: Todos os comandos acima devem ser executados no terminal da respectiva distribuição com permissão de super usuário (usuário que pode fazer alterações no sistema operacional).

Como instalo o \LaTeX no Windows?



Como instalo o \LaTeX no Windows?

Felizmente, há o \TeX Live para Windows que poderá ser obtido no site oficial do projeto \TeX Live.

O usuário de Windows deverá baixar o arquivo

`install-tl-windows.exe`

que possui aproximadamente **13mb**.

Nota: `install-tl-windows.exe` é apenas o instalador do \TeX Live para Windows. Dessa forma, ao final da instalação, o \TeX Live terá muito mais que apenas 13mb instalado em seu computador.

Como instalo o \LaTeX no Linux?

Mas para escrevermos um texto com qualidade usando o \LaTeX precisamos também de um editor de texto.

Na maioria dos casos usamos uma **IDE** (*Integrated Development Environment*) (**Ambiente de Desenvolvimento Integrado**)

Aconselho o uso do \TeX studio que está disponível para Linux, Windows e Mac OS.

O \TeX studio é um software sobre os termos da licença **GPL** (*GNU General Public License*) e pode ser obtido em <http://texstudio.sourceforge.net/>.

Observação: Aperte F6 para compilar o documento e F7 para visualizar o PDF produzido por meio de código \LaTeX .

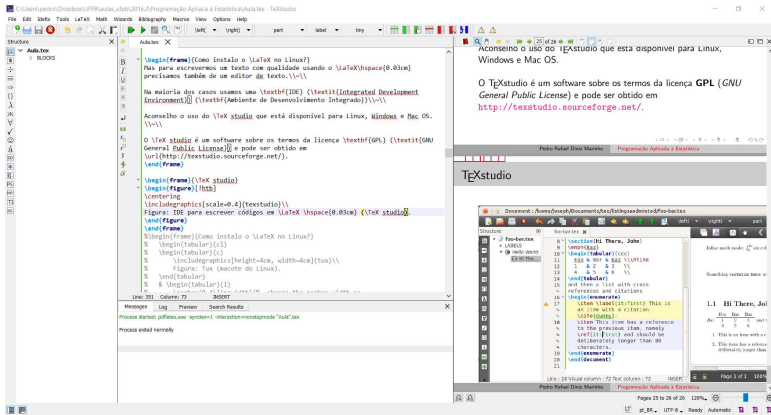


Figura: IDE para escrever códigos em \LaTeX (\TeX studio).

Vantagens do T_EXstudio

Vantagens do T_EXstudio

- ① É uma IDE leve.

Vantagens do T_EXstudio

- ① É uma IDE leve.
- ② O T_EXstudio auto-completa os comandos que vão sendo digitados.

Vantagens do T_EXstudio

- ① É uma IDE leve.
- ② O T_EXstudio auto-completa os comandos que vão sendo digitados.
- ③ É possível ir de um ponto específico do código L^AT_EX para o ponto correspondente no PDF criado.

Vantagens do T_EXstudio

- ① É uma IDE leve.
- ② O T_EXstudio auto-completa os comandos que vão sendo digitados.
- ③ É possível ir de um ponto específico do código L^AT_EX para o ponto correspondente no PDF criado.
- ④ Tem licença **GPL** (*GNU General Public License*), isto é, não pago por ela e tenho acesso ao seu código fonte.

Vantagens do T_EXstudio

- ① É uma IDE leve.
- ② O T_EXstudio auto-completa os comandos que vão sendo digitados.
- ③ É possível ir de um ponto específico do código L^AT_EX para o ponto correspondente no PDF criado.
- ④ Tem licença **GPL** (*GNU General Public License*), isto é, não pago por ela e tenho acesso ao seu código fonte.
- ⑤ O PDF é visualizado ao lado do código L^AT_EX.

Como aprender \LaTeX ?

Pergunta do aluno

Okay professor, o senhor me convenceu em utilizar \LaTeX . Então, como eu posso aprender os seus comandos? Há algum material interessante para se começar a estudar \LaTeX ?

Como aprender \LaTeX ?

Pergunta do aluno

Okay professor, o senhor me convenceu em utilizar \LaTeX . Então, como eu posso aprender os seus comandos? Há algum material interessante para se começar a estudar \LaTeX ?

Resposta do professor legal

Há vários livros e materiais disponibilizados na internet fora a referência considerada nessa disciplina. Há também diversos grupos de discussão sobre \LaTeX a exemplo do grupo \LaTeX -br que poderá ser encontrados nos Grupos do Google. Um bom material em português sobre \LaTeX poderá ser acessado no link <http://www.mat.ufpb.br/lenimar/textos/breve21pdf.zip>.

Como eu sei se o L^AT_EX instalou corretamente em meu computador depois de seguir os passos acima? Reposta: Simples, abra o TeXstudio e digite o comando abaixo e aperte F6 e depois aperte F7 para visualizar o documento.

```
1: \documentclass[a4paper,12pt]{report}
2:
3: \begin{document}
4: Um contato superficial com o \LaTeX. Um amor ao
5: primeiro uso ...
6: \end{document}
```

Observação: Voltaremos a falar sobre L^AT_EX, porém, conversaremos, antes, um pouco sobre programação e a importância que um estatístico deverá dar à esse conhecimento.

O que é programação?

O que é programação?

Resposta: Linguagem de programação é um método padronizado para comunicar instruções para um computador por meio de uma sintaxe.

O que é programação?

Resposta: Linguagem de programação é um método padronizado para comunicar instruções para um computador por meio de uma sintaxe.

Trata-se de um conjunto de regras sintáticas utilizadas para passar instruções para um computador. Por meio dessas regras, é possível que o programador especifique os **tipos de dados** em que o professor irá processar.

Tais dados serão armazenados e/ou transmitidos entre os componentes que forma o computador. Assim, a linguagem também permite especificar quais ações devem ser tomadas e sob quais circunstâncias serão tomadas.

Linguagens



Figura: Diversas linguagens de programação disponíveis para uso.

Por que aprender à programar?

Por que um estatístico deve saber programar?

Por que aprender à programar?

Por que o estatístico deve saber programar?

Resposta: Simplesmente pelo fato de que **não** dá para fazer muita coisa na estatística (principalmente no mercado de trabalho) se o profissional não é capaz de fazer com que o computador resolva os seus problemas.

Por que aprender à programar?

Por que o estatístico deve saber programar?

Resposta: Simplesmente pelo fato de que **não** dá para fazer muita coisa na estatística (principalmente no mercado de trabalho) se o profissional não é capaz de fazer com que o computador resolva os seus problemas.

Um aluno questiona...

Mas professor, temos o SPSS, Excel, SAS, Statistica e outros softwares em que podemos chamar nosso conjunto de dados e apertar centenas de botões e ter **alguns** resultados.

“Ciência da computação tem tanto a ver com o computador como a Astronomia com o telescópio, a Biologia com o microscópio, ou a Química com os tubos de ensaio. A Ciência não estuda ferramentas, mas o que fazemos e o que descobrimos com elas.- **Edsger Dijkstra** (Prêmio Turing em 1972)



Figura: Edsger Dijkstra.

A frase empregada pelo Edsger Dijkstra seria perfeitamente válida se substituirmos a expressão **“Ciência da computação”** por **“Estatística Computacional”**. Porém, não há como se utilizar Estatística Computacional se não dominarmos alguma linguagem de programação e em muitas situações fica difícil desvincular as metologias da estatística computacional da teoria da programação.

Por essa dificuldade de se separar a teoria da estatística computacional da teoria de linguagem de programação é que faz com que a ementa da disciplina contemple o tópico de programação em R, não só aqui, como em diversos outros cursos de estatística computacional.

Máquina de Turing

A máquina de Turing é um dispositivo teórico conhecido como máquina universal, que foi concebido pelo matemático britânico Alan Turing (1912-1954), muitos anos antes de existirem os modernos computadores digitais.

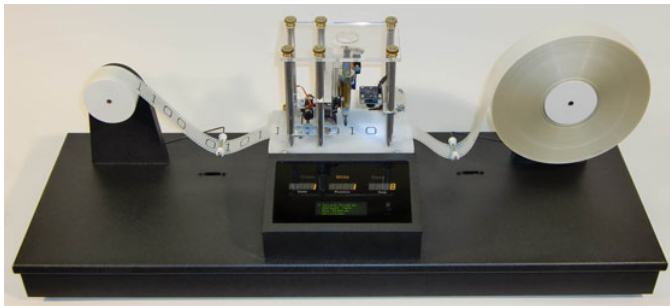


Figura: Exemplo de uma Turing física.

Máquina de Turing

Na Teoria da Computabilidade, um problema é solúvel se há uma Máquina de Turing para aquele problema

Em seu artigo original, Turing demonstra a existência de um **problema insolúvel**.

Existe basicamente dois tipos de máquinas de Turing:

- ① Máquina de Turing **Determinística**: (Se 'A', então 'B')
- ② Máquina de Turing **Não-Determinística**: (Se 'A", então 'B' ou 'C' ou 'D' ou ...)

Teoria da Computabilidade

A computabilidade é a Teoria da Complexidade Computacional que estudam os **limites** da computação:

- ① Quais problemas jamais poderão ser resolvidos por um computador, independente de sua velocidade ou memória?
- ② Quais problemas podem ser resolvidos por um computador, mas requerem um período tão extenso de tempo para completar a ponto de tornar a solução impraticável?
- ③ Em que situações podem ser mais difícil resolver um problema do que verificar cada uma das soluções?

Teoria da Computabilidade

Das três perguntas anteriores, a última é referente às **classes de paradigmas**, são elas:

- Classe **P**: De tempo polinomial determinístico. Os algoritmos pertencentes à esta classe são chamados de **algoritmos eficientes**.
- Classe **NP**: De tempo polinomial não-determinístico.

O conjunto de problemas que não podem ter solução em tempo polinomial mas candidatos a solução podem ser checados em tempo polinomial são problemas pertencentes à classe NP.

Somente uma máquina de Turing não-determinística podem resolver esses problemas. Eles são resolvidos em tempo polinomial por uma máquina de Turing não-determinística que acerta em todos os passos.

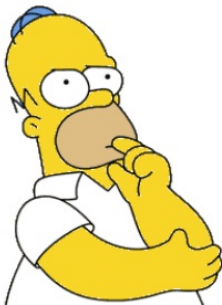
Teoria da Computabilidade

Na matemática, a questão a respeito de $P = NP$ ou $P \neq NP$ é um problema em aberto.

A grande importância dessa classe (NP) de problemas se baseia no fato de que ela contém muitos problemas de busca e otimização para os quais gostaríamos de saber se há uma solução.

Exemplos: Problema do Caixeiro Viajante, problema da mochila, problemas de otimização descritos na ementa do curso como é o caso do método BFGS.

E se $P = NP$?



- ① Diga adeus a criptografia;
- ② Soluções matemáticas não complicadas;
- ③ Previsão do tempo, terremotos e tsunamis;

Problemas NP (precisa-se de linguagens eficientes)

Problemas NP-Completo fazem parte de nossas vidas...

Na estatística sempre nos deparamos com problemas NP-Completo, isto é, sempre lidamos com problemas que não possuem soluções em tempo polinomial.

Exemplo: Constantemente precisamos estimar parâmetros de um modelo probabilístico por meio do método de máxima verossimilhança, isto é, maximizamos a função de log-verossimilhança de um modelo probabilístico. A otimização global é um problema NP-Completo.

Observação: Atualmente está cada vez mais complicados realizar tais otimizações uma vez que os modelos cada vez mais estão adicionando parâmetros extras o que torna a função extremamente complicada em alguns casos.

Breve História da Linguagem C

Embora possua um nome estranho quando comparada com outras linguagens de programação da terceira geração, como FORTRAN, PASCAL, ou COBOL, a linguagem C é uma das linguagens mais importantes até hoje criada.

Curiosidade: O nome da linguagem (e a própria linguagem) resulta da evolução de uma outra linguagem de programação, desenvolvida pelo programador **Ken Thompson** nos Laboratórios Bell, chamada de B.

Por que aprender à programar?

O professor continua respondendo...

A maioria desses software não possuem as técnicas estatísticas mais recentes e muitas vezes não são capazes de se adequar aos problemas específicos que nos deparamos ao tentar resolver um problema.

Muitas vezes precisamos modificar uma função programada por um outro programador para que ela venha a funcionar como solução ao nosso problema.

Observação: Diversas outras vezes precisamos programar para realizar simulações. É muito comum na estatística estudar propriedades de algumas estatística ou modelo estatístico e querer simular o seu comportamento em diversos cenários diferentes.

Breve História da Linguagem C



Figura: Ken Thompson (sentado) jogando xadrez com um colega.

B trata-se de uma simplificação da linguagem BCPL (*Basic Combined Programming Language*). Assim como BCPL, B só possuía um tipo de dados.

A linguagem B também foi recebeu contribuições do Dennis Ritchie (criador de C).

Breve História da Linguagem C



Figura: Dennis Ritchie (criador da linguagem C).

A linguagem C foi criada em 1972 nos *Bell Telephone Laboratories* por Dennis Ritchie com a finalidade de permitir a escrita do sistema operacional Unix.

Desejava-se uma linguagem de alto nível de modo a evitar o uso do Assambly.

Breve História da Linguagem C



Figura: Dennis Ritchie (criador da linguagem C).

Devido às capacidades e através da divulgação do sistema Unix pelas universidades dos Estados Unidos, a linguagem C deixou cedo as portas dos laboratórios Bell.

C disseminou-se e tornou-se conhecida por todos os tipos de programadores, independentemente dos projetos em que estivessem envolvidos.

Breve História da Linguagem C

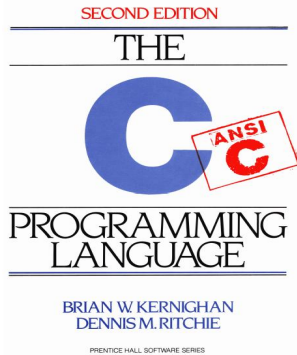


Figura: Ótimo livro sobre a linguagem C.

Essa dispersão de diferentes projetos utilizando a linguagem C levou a que diferentes organizações desenvolvessem e utilizassem diferentes versões da linguagem C criando assim alguns problemas de compatibilidade, entre diversos outros.

O material ao lado é um livro sobre a linguagem C escrito pelo seu criador.

Breve História da Linguagem C

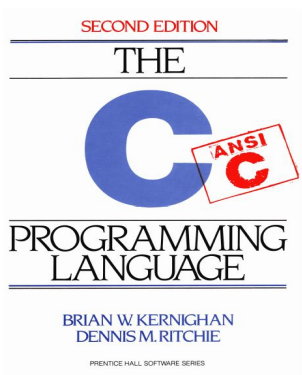


Figura: Ótimo livro sobre a linguagem C.

Devido ao fenômeno que foi a linguagem C e aos problemas de compatibilidade que existiam na época, o **American National Standards Institute** (ANSI) formou em 1983 um comitê para a definição de um padrão para a linguagem C.

Breve História da Linguagem C

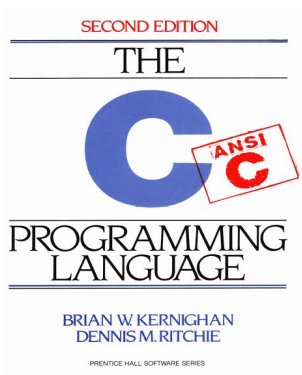


Figura: Ótimo livro sobre a linguagem C.

O padrão tem como objetivo o funcionamento semelhante de todos os compiladores da linguagem, com especificações muito precisas sobre aquilo que a linguagem deve ou não fazer, seus limites, definições, dentre outras coisas.

Mil e Uma Razões para Programar em C

Devido à enorme quantidade de linguagens de programação disponíveis no mercado, seria necessário que uma delas se destacasse muito em relação às outras para conseguir interessar tantos programadores.

A maior parte das linguagens de programação tem um objetivo específico a atingir:

- PASCAL - Ensino de Técnicas de Programação.
- FORTRAN - Cálculo Científico.
- LISP e PROLOG - Vocacionadas para as áreas de Inteligência Artificial.

Mil e Uma Razões para Programar em C

Pergunta do aluno

Certo, entendi, temos linguagens que se destinam a um objetivo específico como as que foram listadas a cima. Mas quanto à C, a que área de desenvolvimento se destina?

Mil e Uma Razões para Programar em C

Pergunta do aluno

Certo, entendi, temos linguagens que se destinam a um objetivo específico como as que foram listadas a cima. Mas quanto à C, a que área de desenvolvimento se destina?

Resposta do professor legal

NENHUMA 😊 .

Por que C?

Qual o porquê de se aprender a linguagem C

C é uma linguagem bastante interessante devido ser uma linguagem de programação de propósito geral, isto é, C pode ser utilizada para os mais variados fins.

Alguns outros motivos para se aprender C são:

Por que C?

Qual o porquê de se aprender a linguagem C

C é uma linguagem bastante interessante devido ser uma linguagem de programação de propósito geral, isto é, C pode ser utilizada para os mais variados fins.

Alguns outros motivos para se aprender C são:

- ① C é uma linguagem extremamente eficiente.

Por que C?

Qual o porquê de se aprender a linguagem C

C é uma linguagem bastante interessante devido ser uma linguagem de programação de propósito geral, isto é, C pode ser utilizada para os mais variados fins.

Alguns outros motivos para se aprender C são:

- ① C é uma linguagem extremamente eficiente.
- ② Diversas outras linguagem foram programas utilizando C.

Por que C?

Qual o porquê de se aprender a linguagem C

C é uma linguagem bastante interessante devido ser uma linguagem de programação de propósito geral, isto é, C pode ser utilizada para os mais variados fins.

Alguns outros motivos para se aprender C são:

- ① C é uma linguagem extremamente eficiente.
- ② Diversas outras linguagem foram programas utilizando C.
- ③ Na estatística muitas vezes precisamos fazer simulações que são computacionalmente intensivas.

Por que C?

Qual o porquê de se aprender a linguagem C

C é uma linguagem bastante interessante devido ser uma linguagem de programação de propósito geral, isto é, C pode ser utilizada para os mais variados fins.

Alguns outros motivos para se aprender C são:

- ① C é uma linguagem extremamente eficiente.
- ② Diversas outras linguagem foram programas utilizando C.
- ③ Na estatística muitas vezes precisamos fazer simulações que são computacionalmente intensivas.
- ④ C é uma ótima linguagem para se começar a programar.

Por que C?

Qual o porquê de se aprender a linguagem C

C é uma linguagem bastante interessante devido ser uma linguagem de programação de propósito geral, isto é, C pode ser utilizada para os mais variados fins.

Alguns outros motivos para se aprender C são:

- ① C é uma linguagem extremamente eficiente.
- ② Diversas outras linguagem foram programas utilizando C.
- ③ Na estatística muitas vezes precisamos fazer simulações que são computacionalmente intensivas.
- ④ C é uma ótima linguagem para se começar a programar.
- ⑤ Códigos em C podem ser importados para R, isto é, R “conversa” com C.

Por que C?

Observação: Pelo fato de C ser uma linguagem de propósito geral (*general purpose*), esta linguagem pode ser utilizada nos mais variados fins, como sistemas operacionais, interfaces gráficas, etc.

Importante

Há uma falácia de que C é uma linguagem extremamente difícil. Na verdade ocorre que muitas pessoas começam a estudar programação por meio de C, momento este em que é somado dificuldades em aprender à programar (lógica de programação) com as dificuldades de se aprender uma sintaxe de uma linguagem de programação.

Observação: Porém, é verdade que muitas coisas temos que fazer nós mesmo em C. Por isso que ela é a linguagem adotada na maioria dos cursos de introdução à programação espalhados pelo mundo.

Por que C?

Outras razões para se utilizar C

C é utilizado quando a velocidade, espaço e portabilidade são importantes. A maioria dos sistemas operacionais das outras linguagens e de grande parte dos softwares e games são escritas em C.

Observação: Há basicamente três padrões de C que podem ser encontrados por aí. São eles:

- ① **ANSI C** que é do fim dos anos de 1980 e é utilizado para códigos mais antigos;
- ② Muitas coisas foram consertadas no padrão C99 de 1999;
- ③ Algumas novidades foram acrescentadas no atual padrão C11 lançado em 2011.

Observação: Não existem grandes diferenças entre as versões de C. Iremos destacá-las ao longo do caminho.

Arquitetura de von Neumann



Figura: John von Neumann.

A arquitetura de von Neumann é uma arquitetura de um computador digital que possibilita uma máquina digital armazenar os seus programas no mesmo espaço de memória que os dados, podendo assim manipular tais programas.

“Não há sentido em em ser preciso quando não se sabe sobre o que está a falar-von Neumann.

Arquitetura de von Neumann

A máquina proposta por von Neumann possui as seguintes componentes:

- ① Uma **memória**;
- ② Uma **unidade aritmética e lógica**;
- ③ Uma **unidade central de processamento** (CPU), composto por diversos registradores;
- ④ Uma **unidade de controle**, cuja função é a mesma da tabela de controle de uma Máquina de Turing universal (estabelece as mudanças de estado por meio das entradas).

O que é um compilador?

O que é um compilador?

Resposta: Um compilador é um programa de computador ou mesmo um grupo de programas que é responsável por traduzir um código fonte escrito em uma linguagem compilada à um programa equivalente do ponto de vista semântico.

O que é um compilador?

Resposta: Um compilador é um programa de computador ou mesmo um grupo de programas que é responsável por traduzir um código fonte escrito em uma linguagem compilada à um programa equivalente do ponto de vista semântico.

O compilador traduz o código fonte de uma linguagem de programação de médio/alto nível para uma linguagem de programação de baixo nível (a exemplo da linguagem Assembly ou código de máquina).

Bytecode?

Bytecode?

Alguns compiladores traduzem o código para um formato intermediário, denominado de **bytecode** que é um código de baixo nível. Sendo assim, o bytecode não é imediatamente um arquivo executável.

Bytecode?

Alguns compiladores traduzem o código para um formato intermediário, denominado de **bytecode** que é um código de baixo nível. Sendo assim, o bytecode não é imediatamente um arquivo executável.

Observação: Chamamos de linguagem de baixo nível as linguagem que trabalhando próximo ao hardware. Baixo nível, médio nível ou alto nível em nada em a ver com a qualidade da linguagem de programação.

Bytecode?

Alguns compiladores traduzem o código para um formato intermediário, denominado de **bytecode** que é um código de baixo nível. Sendo assim, o bytecode não é imediatamente um arquivo executável.

Observação: Chamamos de linguagem de baixo nível as linguagem que trabalhando próximo ao hardware. Baixo nível, médio nível ou alto nível em nada em a ver com a qualidade da linguagem de programação.

Importante: Jamais confunda bytecode com código de máquina. Bytecode é um formato intermediário que irá ser interpretado em uma máquina virtual que fará a execução.

A vantagem do bytecode é que o código torna-se mais **portável**, isto é, podemos com o resultado da compilação executar o código proveniente de um processo de compilação em diversas arquiteturas distintas. Dessa forma, o bytecode irá produzir o mesmo resultado esperado em qualquer arquitetura que possua uma máquina virtual que execute o código intermediário.

Exemplos de linguagem que converte o código fonte para bytecode: Java que corre o código sobre a máquina virtual Java, .NET que corre o código sobre a *Common Language Runtime*.

Código Objeto e Código de Máquina

O **código de máquina** é um código binário (0 e 1) que poderá ser executado diretamente pela CPU.

Se abrirmos um arquivo de código de máquina em um editor de texto, veríamos um emaranhado de caracteres sem sentido. É possível ter acesso ao código de máquina em formato hexadecimal por meio de softwares adequados.

O **código objeto** é a saída de um processo de compilação e trata-se de uma parte do código de máquina que ainda não foi vinculado em um programa completo por meio de um **linker**.

Abrindo um Código de Máquina

Ao tentarmos abrir um código de máquina em um editor de texto comum visualizamos algo sem sentido como a sequência de caracteres abaixo:

Abrindo um Código de Máquina

Ao tentarmos abrir um código de máquina em um editor de texto comum visualizamos algo sem sentido como a sequência de caracteres abaixo:

```
MZÀ$PÿvèŠÿ]Ë3ÀP,FëfF,< uè2ÀëäÀtBa
ÀuC†à2Àüã¬I,"t"<\u€<"u¬IöÃ□□éîYÊ.<å%.
£~ã%v,vüÿvÿv□èÅfÄÿvpÿvüèüêYY<V<F
|ëRÿvpÿvüèWífÄ<å]ËU<ìfìHVW<~<F
```

Porém, é possível ter acesso ao código de máquina utilizando editores próprios que apresentam o código em hexadecimal, como o exemplo que segue no *frame* seguinte.

Compilador

```
C:\Utility>debug v.exe
```

```
-d 0 100
```

```
0E3D:0000  CD 20 FF 9F 00 9A F0 FE .....0.....
0E3D:0010  F0 07 17 03 F0 07 DF 07 .....
0E3D:0020  FF FF FF FF FF FF FF FF .....L.
0E3D:0030  D0 0C 14 00 18 00 3D 0E .....=.....
0E3D:0040  05 00 00 00 00 00 00 00 .....
0E3D:0050  CD 21 CB 00 00 00 00 00 .!.....
0E3D:0060  20 20 20 20 20 20 20 20 .....
0E3D:0070  20 20 20 20 20 20 20 20 .....
0E3D:0080  00 0D 76 2E 65 78 65 0D ..v.exe.DOWS\sys
0E3D:0090  74 65 6D 33 32 5C 64 6F tem32\dosx..da r
0E3D:00A0  65 64 65 20 28 63 61 72 ede (carregar an
0E3D:00B0  74 65 73 20 64 6F 20 64 tes do dosx.exe)
```

Compilador

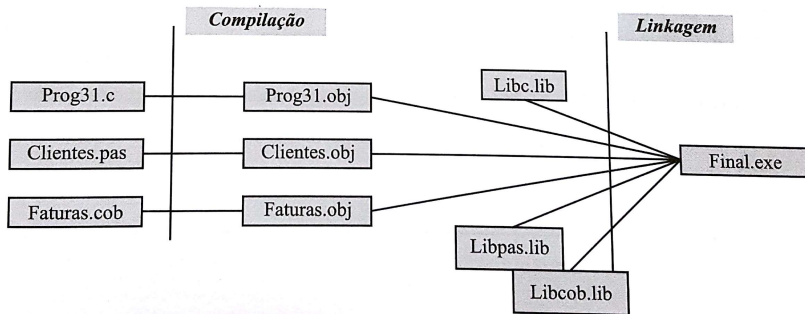


Figura: Diagrama (funcionamento de uma compilador).

Alguns autores citam linguagens compiladas em que a tradução do código gera código em C.

Maiores detalhes na referência abaixo

Cooper, Torczon. Engineering a Compiler (em inglês). San Francisco: Morgan Kaufmann, 2003. p. 2. ISBN 1-55860-698-X.

Importante

É importante não confundir **compilador** com **tradutor** ou **filtro** que também pode ser chamado de **conversor de linguagem**.

O conversor de linguagem é responsável por converter o código de uma linguagem de alto nível para o código de uma outra linguagem de médio/alto nível.

Observação: Um programa que traduz uma linguagem de programação de baixo nível para uma linguagem de programação de alto nível é denominado de **descompilador**.



Figura: Grace Hopper.

O primeiro compilador foi escrito por Grace Hopper no ano de 1952 para a linguagem de programação A-0.

Grace Hopper foi analista de sistemas da Marinha dos Estados Unidos. Ela também criou o primeiro compilador para a linguagem COBOL.

Curiosidade: É atribuído à Grace Hopper o termo **bug** utilizado para designar uma falha no código fonte.



Figura: Grace Hopper.

Grace Hopper é graduada em matemática e física em 1928 e em 1930 concluiu seu mestrado na Yale University. Em 1934, na mesma Universidade, ela obteve o seu PhD em matemática.

Muitos compiladores incluem um **pré-processador** que é um programa separado mas invocado pelo compilador antes do início do processo de tradução.

Normalmente é pre-processador responsável por mudanças no código fonte destinadas de acordo com decisões tomadas em tempo de compilação.

Em programas em C há diversas diretivas para inclusão de novos códigos disponíveis em bibliotecas ou código a parte escrito pelo programador que é informado sua existência por meio de diretivas para o pre-processador.

Exemplo: É o pré-processador que substitui os comentários do código fonte por espaços em branco. Ou seja, o compilador não “enxerga” nenhum comentário.

Exemplos de linguagens compiladas: C, C++, Fortran, Objecte-C, Ocaml, BASIC, COBOL, Ada, D, entre outras.

O que é um interpretador?

Resposta: Interpretadores são programas de computador que tem a finalidade de ler o código-fonte de uma linguagem de programação interpretada e o converte em um código executável.

O que é um interpretador?

Resposta: Interpretadores são programas de computador que tem a finalidade de ler o código-fonte de uma linguagem de programação interpretada e o converte em um código executável.

Seu funcionamento pode ser, em geral, de duas formas:

- O interpretador lê linha-por-linha e converte o código fonte em código objeto (ou bytecode) a medida que vai executando o programa.

O que é um interpretador?

Resposta: Interpretadores são programas de computador que tem a finalidade de ler o código-fonte de uma linguagem de programação interpretada e o converte em um código executável.

Seu funcionamento pode ser, em geral, de duas formas:

- O interpretador lê linha-por-linha e converte o código fonte em código objeto (ou bytecode) a medida que vai executando o programa.
- O interpretador converte o código fonte por inteiro e depois o executa.

O que é um interpretador?

Resposta: Interpretadores são programas de computador que tem a finalidade de ler o código-fonte de uma linguagem de programação interpretada e o converte em um código executável.

Seu funcionamento pode ser, em geral, de duas formas:

- O interpretador lê linha-por-linha e converte o código fonte em código objeto (ou bytecode) a medida que vai executando o programa.
- O interpretador converte o código fonte por inteiro e depois o executa.

Exemplos de linguagens interpretadas: R, Perl, Python, Haskell, Lua, Ruby, Lisp, JavaScript, entre outras.

Agora que conhecemos a diferença entre **linguagem de programação compilada** para **linguagem de programação interpretada** e sabemos que R é uma linguagem interpretada, jamais fale:

“MEU CÓDIGO R ESTÁ COMPILANDO ...”

Dizer o que está destacado acima é uma prova contundente do pouco conhecimento de computação e/ou da linguagem R.

Por que aprender à programar?

Uma linguagem de programação bastante utilizada na estatística é a linguagem R.

Linguagem R

R é uma linguagem de programação para computação estatística e gráficos. R é uma parte oficial do projeto GNU da Free Software Foundation's.

Curiosidade: A linguagem R foi criada originalmente por Ross Ihaka e Robert Gentleman no **Departamento de Estatística** da Universidade de Auckland, Nova Zelândia em agosto de 1993.

Nota: É muito importante que um estatístico saiba programar na linguagem R. Alguns empregos exigem isso. Porém, se não exigirem, o R te ajudará bastante.

Por que aprender à programar?



Figura: Criadores da linguagem R [Robert Gentleman (foto à esquerda) e Ross Ihaka (foto à direita)].

Por que aprender à programar?



Figura: Logo da linguagem R.

Um dos grandes motivos da grande popularidade da linguagem R se deve a grande quantidade de pacotes disponíveis para os usuários da linguagem.

Atualmente há mais de 12750 pacotes para R com o foco nas mais variadas áreas: estatística, matemática, biologia, economia, entre outras.

Por que aprender à programar?



Figura: Logo da linguagem R.
Obtenha a linguagem R em
<https://www.r-project.org/>.

Observação: Para programar em R **não é suficiente entender alguns pacotes específicos**. É preciso entender a sintaxe base da linguagem que nos permite inclusive criar outros pacotes e melhorar os existentes.

Por que aprender à programar?



Figura: Logo da linguagem R.

Obtenha a linguagem R em

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

Observação: As novas metodologias estatísticas chegam mais rapidamente em R do que em outros softwares estatísticos pelo fato do R ser uma linguagem livre (código aberto e gratuita).

Por que R?

Algumas respostas:

- ① R é uma linguagem de código-aberto;
- ② Muitos estatísticos espalhados pelo mundo utilizam R;
- ③ Grandes empresas utilizam a linguagem R;
- ④ Há muitas metodologias estatísticas empacotadas;
- ⑤ R se comunica com a linguagem C/C++ e outras;
- ⑥ É possível escrever textos tipográficos de qualidade usando \LaTeX e inserir códigos R no texto para produção de gráficos, tabelas e resultados de metodologias estatística.

Observação: Maiores detalhes sobre download e instalação da linguagem R nos diferentes sistemas operacionais e arquiteturas poderão ser obtidos em <https://cloud.r-project.org/>.

Programação em R

É importante entender que R antes de ser um ambiente para computação estatística e gráficos trata-se de uma linguagem de programação.

É importante entender que R antes de ser um ambiente para computação estatística e gráficos trata-se de uma linguagem de programação.

Muito Importante

Para que seja possível dominar a linguagem R é fundamental esquecermos o paradigma de estudo da linguagem por meio de livros e tutoriais que tentam ensinar R por meio de soluções de problemas estatísticos. Esses materiais são úteis quando se tem um domínio razoável da linguagem. Tentar aprender a linguagem R como um software estatístico e não como uma linguagem de programação irá impor diversas deficiências de programação ao usuário/programador.

Para termos uma boa experiência de programação em R, é importante instalar um editor ou IDE (*Integrated Development Environment*) de programação.

Uma boa IDE para programação em R é o **RStudio**. **RStudio** vem evoluindo bastante com o passar dos anos e hoje é a melhor IDE para se trabalhar com R.

Para maiores detalhes e download do **RStudio**, acessar: <https://www.rstudio.com/>.



Figura: Algumas empresas que utilizam o **RStudio**.

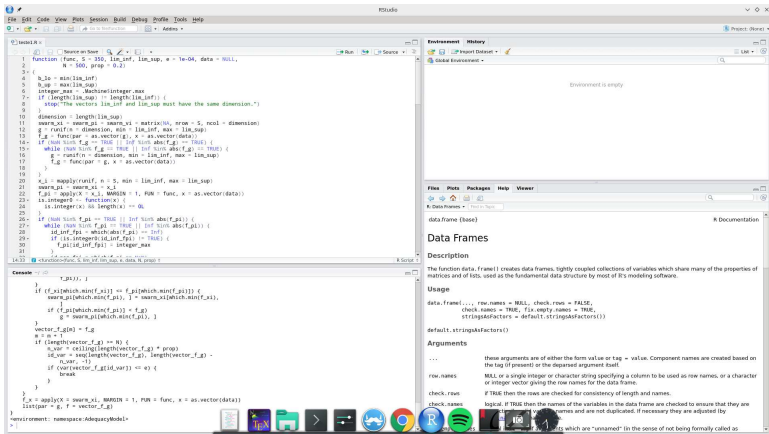


Figura: RStudio (IDE de programação em R).



Figura: GNU Affero General Public License - GNU AGPL, uma das licenças do **RStudio**.

Uma das licenças do **RStudio** é a licença GNU AGPL. Com o **RStudio** sob os termos dessa licença poderemos ter acesso ao código fonte do **RStudio** e este poderá ser executado atrás de um servidor web, do mesmo modo como funciona uma aplicação web.

Observação: Há também uma versão paga do **RStudio** custando \$995 dólares por ano. Essa versão trás novas funcionalidades ao **RStudio** como, por exemplo, executar várias análises em paralelo, a possibilidade de executar várias versões do R lado a lado, monitora sessões ativas, CPU e utilização de memória, bem como outras funcionalidades administrativas além do suporte técnico.

A empresa **RStudio** com sede em Boston, Massachusetts, USA também fornece outros serviços relacionados com a linguagem R e com a ferramenta **RStudio** como é o caso do **Shiny** que é um pacote open-source de R que fornece uma estrutura para construção de aplicativos web usando R.

Shiny ajuda o programador transformar as análises em aplicativos web interativos e não requer conhecimento de HTML, CSS ou JavaScript.

Observe no link alguns exemplos de análises interativas: <http://shiny.rstudio.com/gallery/>.

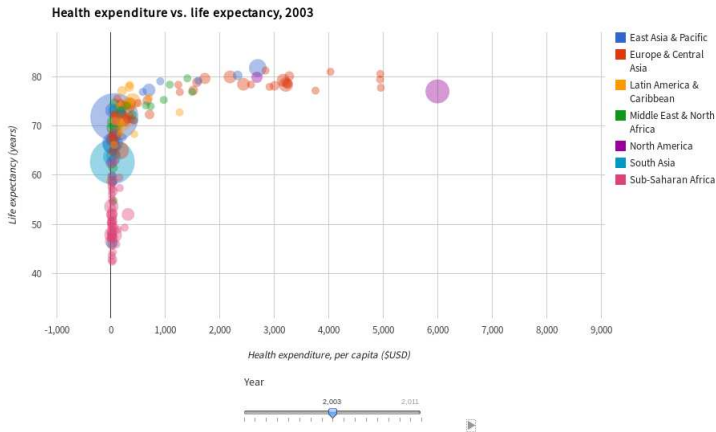


Figura: Exemplo de gráfico interativo usando **Shiny**.

A equipe do **RStudio** também vem produzindo alguns pacotes importantes que são bastante úteis para programadores de R além do **Shiny**, como é o caso do **rmarkdown**, **knitr**, **ggplot2**, entre outros.



Figura: Alguns pacotes produzidos pela equipe do **RStudio**.

No primeiro quadrante da IDE é onde escrevemos nosso código com extensão `.R`. Para que o interpretador de R possa agir sobre um trecho do código, este deverá ser selecionado e em seguida deverá ser teclado **Ctrl** + **Enter**.

No segundo quadrante é onde observamos os objetos criados pelo programador que basicamente são matrizes, listas, dataframes, variáveis, entre outros objetos suportados pela linguagem.

O terceiro quadrante é o prompt de comando de R. Toda a saída encontra-se nesse quadrante.

O quarto quadrante é onde podemos consultar o help das funções e onde observamos as saídas gráficas.

Recentemente surgiu uma nova IDE para programação em R com o nome de **RCode**.

Recentemente surgiu uma nova IDE para programação em R com o nome de **RCode**.

Algumas características do **RCode** são:

- Tem como objetivo ser otimizada e o mais leve possível, permitindo trabalhar bem como máquinas mais modestas.

Recentemente surgiu uma nova IDE para programação em R com o nome de **RCode**.

Algumas características do **RCode** são:

- Tem como objetivo ser otimizada e o mais leve possível, permitindo trabalhar bem como máquinas mais modestas.
- Não há necessidade de baixar manualmente o **RCode** cada vez que é atualizado.

Recentemente surgiu uma nova IDE para programação em R com o nome de **RCode**.

Algumas características do **RCode** são:

- Tem como objetivo ser otimizada e o mais leve possível, permitindo trabalhar bem como máquinas mais modestas.
- Não há necessidade de baixar manualmente o **RCode** cada vez que é atualizado.
- Fornece automaticamente o tempo de execução de cada comando.

Recentemente surgiu uma nova IDE para programação em R com o nome de **RCode**.

Algumas características do **RCode** são:

- Tem como objetivo ser otimizada e o mais leve possível, permitindo trabalhar bem como máquinas mais modestas.
- Não há necessidade de baixar manualmente o **RCode** cada vez que é atualizado.
- Fornece automaticamente o tempo de execução de cada comando.
- Preenchimento automático de comandos.

O **RCode** possui versões para Linux, macOS e Windows. Maiores detalhes em <https://www.pgm-solutions.com/rcode>.

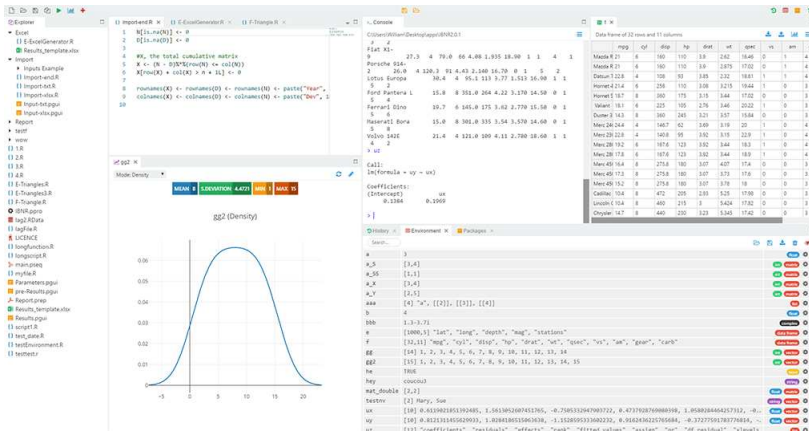


Figura: IDE RCode para programação em linguagem R.

Estrutura de dados em R

As estruturas da base de dados em R podem ser organizadas em diversas dimensões (1d, 2d, nd) que podem ser homogêneas (mesmo tipo de dados) ou heterogêneas (diferentes tipos de dados).

Tabela: Estrutura de dados em R.

	Homogênea	Heterogênea
1d	Atomic vector	List
2d	Matrix	Data Frame
nd	Array	

Observação: Quase todos os outros objetos da linguagem R são construídos sobre essas fundações.

Muito Importante

R não possui strings nem escalares propriamente ditos. Na verdade estes são tratados como um vetor de cardinalidade 1.

Muito Importante

R não possui strings nem escalares propriamente ditos. Na verdade estes são tratados como um vetor de cardinalidade 1.

Nota: Dado um objeto qualquer em R, a melhor forma de entender sua estrutura de dados é utilizando a função `str()`. A função `str()` apresenta um resumo de forma compacta e legível do objeto em questão.

Muito Importante

R não possui strings nem escalares propriamente ditos. Na verdade estes são tratados como um vetor de cardinalidade 1.

Nota: Dado um objeto qualquer em R, a melhor forma de entender sua estrutura de dados é utilizando a função `str()`. A função `str()` apresenta um resumo de forma compacta e legível do objeto em questão.

Porém, espere um pouco antes de utilizar esta função. Nós ainda não sabemos como criar objetos em R.

Estrutura de dados em R

A estrutura básica de dados da linguagem R é o vetor. Por isso que dizemos que R é uma **linguagem vetorial**.

Existem dois tipos de vetores em R:

Estrutura de dados em R

A estrutura básica de dados da linguagem R é o vetor. Por isso que dizemos que R é uma **linguagem vetorial**.

Existem dois tipos de vetores em R:

- ① Vetores atômicos;

Estrutura de dados em R

A estrutura básica de dados da linguagem R é o vetor. Por isso que dizemos que R é uma **linguagem vetorial**.

Existem dois tipos de vetores em R:

- ① Vetores atômicos;
- ② Listas.

Estrutura de dados em R

A estrutura básica de dados da linguagem R é o vetor. Por isso que dizemos que R é uma **linguagem vetorial**.

Existem dois tipos de vetores em R:

- ① Vetores atômicos;
- ② Listas.

Lembre-se: Lista em R é um vetor não-atômico. Enquanto um vetor atômico possui elementos do mesmo tipo, uma lista poderá ser heterogênea.

Existem três funções úteis para obter as propriedades de um vetor. Estas são:

Estrutura de dados em R

Existem três funções úteis para obter as propriedades de um vetor. Estas são:

- ① `typeof()` (retorna o tipo de dados do vetor);

Existem três funções úteis para obter as propriedades de um vetor. Estas são:

- ① `typeof()` (retorna o tipo de dados do vetor);
- ② `length()` (retorna a cardinalidade do vetor);

Estrutura de dados em R

Existem três funções úteis para obter as propriedades de um vetor. Estas são:

- ① `typeof()` (retorna o tipo de dados do vetor);
- ② `length()` (retorna a cardinalidade do vetor);
- ③ `attributes()` (retorna metadados adicionais que possam existir).

Faça no prompt do R:

```
1: # Inicializando um vetor em R.  
2: x <- c(1,7,9,7) # x é um vetor atômico.  
3: length(x)  
4: typeof(x)
```


Comentando um código R

Observação: Comentários na linguagem R são realizados utilizando o símbolo #. Um código bem implementado sempre possui uma quantidade razoável de comentários. Há quem diga que um código deverá ter mais comentários do que linhas implementadas. De fato, isto não é ruim!

```
1: # Sempre é bom comentar
2: # um código
3:
4: código
5:
6: # Mais comentários
```

Checando um vetor

Podemos utilizar duas formas para checar se um objeto em R é um vetor (atômico ou não-atômico).

- **Forma 1:** `is.vector(x)`, em que `x` é um vetor ou lista;
- **Forma 2:** `is.atomic(x) || is.list(x)`.

Adianto que a melhor forma de fazer isso é a segunda, pois esta forma funciona nas situações em que o vetor possui algum atributo.

Checando um vetor

Podemos utilizar duas formas para checar se um objeto em R é um vetor (atômico ou não-atômico).

- **Forma 1:** `is.vector(x)`, em que `x` é um vetor ou lista;
- **Forma 2:** `is.atomic(x) || is.list(x)`.

Adianto que a melhor forma de fazer isso é a segunda, pois esta forma funciona nas situações em que o vetor possui algum atributo.

Importante: Os **tipos atômicos** (*atomic types*) da linguagem R são basicamente: `logical`, `integer`, `numeric`, `complex`, `character`, `NULL` e `raw`. Este último tipo não é importante para nós no momento.

Checando um vetor

Exemplo: Note a diferença da saída de `is.atomic(y)` e da obtida por `is.atomic(y) || is.list(y)`.

```
1: x <- c(a=1,b=2)
2: y <- c(a=1,b=2)
3: attr(y,'ano') <- '2017'
4: # Dois comandos em uma linha.
5: is.vector(x); is.vector(y)
6: is.atomic(y) || is.list(y)
```

Se algo ainda parece complicado, na medida que formos avançando na programação em R, as dúvidas serão sanadas.

Vetores atômicos

Vetores atômicos são criados com o comando `c()`. Alguns exemplos:

Vetores atômicos

Vetores atômicos são criados com o comando `c()`. Alguns exemplos:

- **Ex:** `dbl_var <- c(1,2.5,4.5)`. Vetor de variáveis de precisão dupla;

Vetores atômicos

Vetores atômicos são criados com o comando `c()`. Alguns exemplos:

- **Ex:** `dbl_var <- c(1,2.5,4.5)`. Vetor de variáveis de precisão dupla;
- **Ex:** `int_var <- c(1L, 6L, 10L)`. Aqui estamos criando um vetor de inteiros quando colocamos o sufixo `L`;

Vetores atômicos

Vetores atômicos são criados com o comando `c()`. Alguns exemplos:

- **Ex:** `dbl_var <- c(1,2.5,4.5)`. Vetor de variáveis de precisão dupla;
- **Ex:** `int_var <- c(1L, 6L, 10L)`. Aqui estamos criando um vetor de inteiros quando colocamos o sufixo `L`;
- **Ex:** `log_var <- c(TRUE, FALSE, T, F)`. Diferentemente da linguagem C, os valores lógicos em R são representados por `TRUE` ou `T` e `FALSE` ou `F`;

Vetores atômicos

Vetores atômicos são criados com o comando `c()`. Alguns exemplos:

- **Ex:** `dbl_var <- c(1,2.5,4.5)`. Vetor de variáveis de precisão dupla;
- **Ex:** `int_var <- c(1L, 6L, 10L)`. Aqui estamos criando um vetor de inteiros quando colocamos o sufixo `L`;
- **Ex:** `log_var <- c(TRUE, FALSE, T, F)`. Diferentemente da linguagem C, os valores lógicos em R são representados por `TRUE` ou `T` e `FALSE` ou `F`;
- **Ex:** `chr_var <- c("string", "outra_string2")`. Vetor de sequências de caracteres.

Vetores atômicos

Importante

Note que atribuições da da forma `objeto <- valor` faz com que **objeto** seja um vetor de uma única posição.

Vetores atômicos

Importante

Note que atribuições da da forma `objeto <- valor` faz com que **objeto** seja um vetor de uma única posição.

```
1: a <- 7L
```

```
2: a[1]
```

```
#> 7
```

```
# Note que o objeto "a" é um vetor. Pergunte se ele  
# é um vetor:
```

```
3: is.vector(a) # Já sabia que "a" não iria "list".
```

```
#> TRUE
```

```
# Poderia ter feito a comparação acima da seguinte forma:
```

```
4: is.atomic(a) || is.list(a)
```

```
#> TRUE
```

Concatenando um vetor

Importante

Vetores atômicos são sempre “planos” mesmo que tenhamos a ideia de aninhar vários vetores atômicos.

Exemplo: Veja que não existe nenhuma diferença entre executar o código `c(1, c(2, c(3, 4)))` e executar `c(1, 2, 3, 4)`.

Nota: Muito embora para o fim do exemplo acima concatenar vetores não haveria necessidade uma vez que poderíamos criar o vetor de uma única vez utilizando `c(1, 2, 3, 4)`, saber concatenar vetores é bastante necessário.