

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)

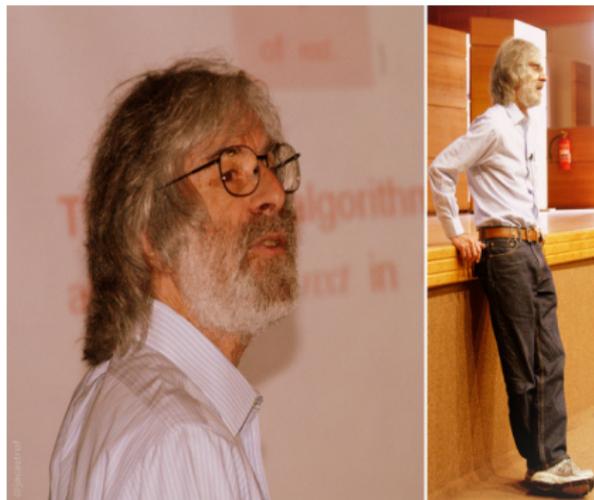


Figura: Leslie Lamport.

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)



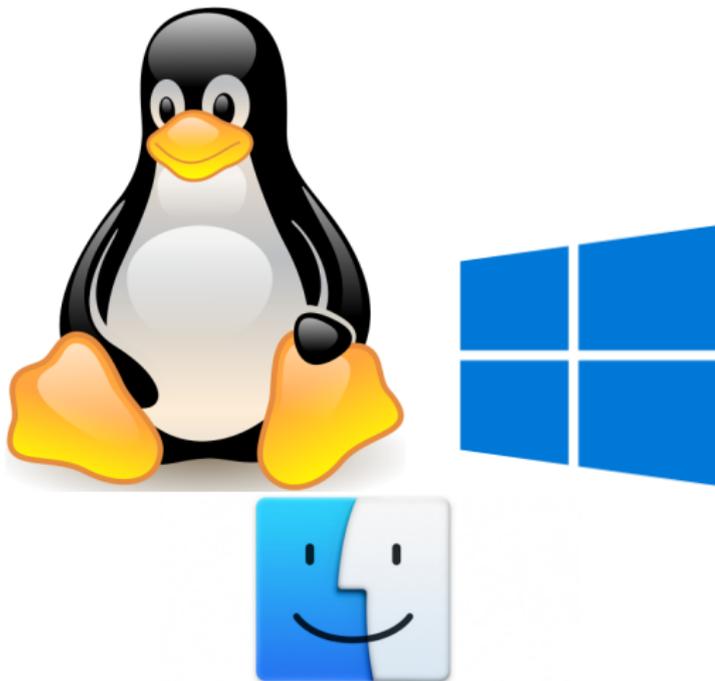
Figura: CTAN lion drawing
by Duane Bibby.

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 .

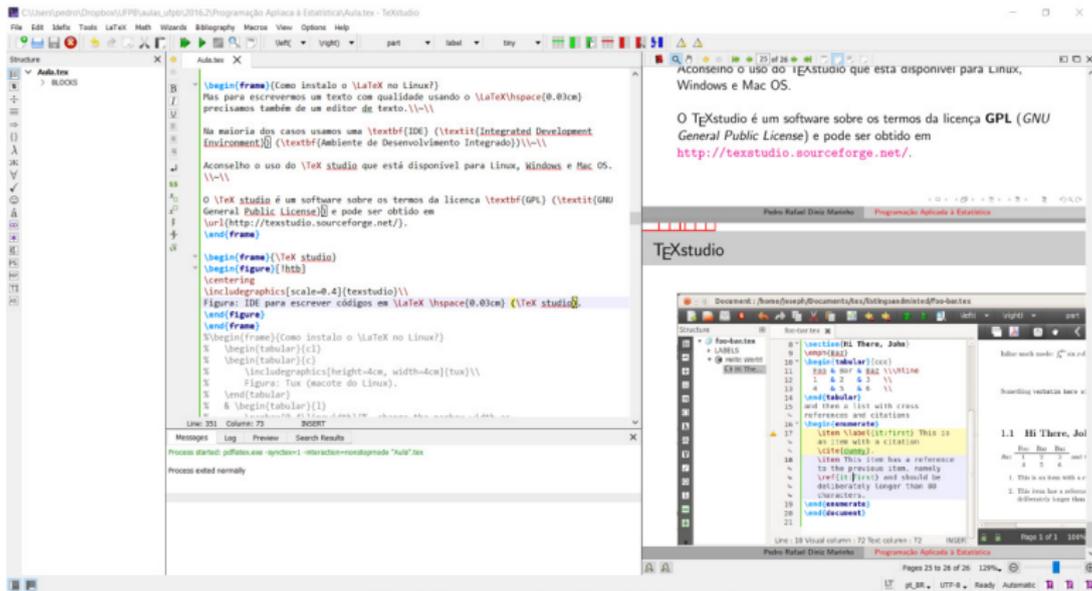


Figura: IDE para escrever cdigos em \LaTeX (\TeX studio).

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.

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 metodologias 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.

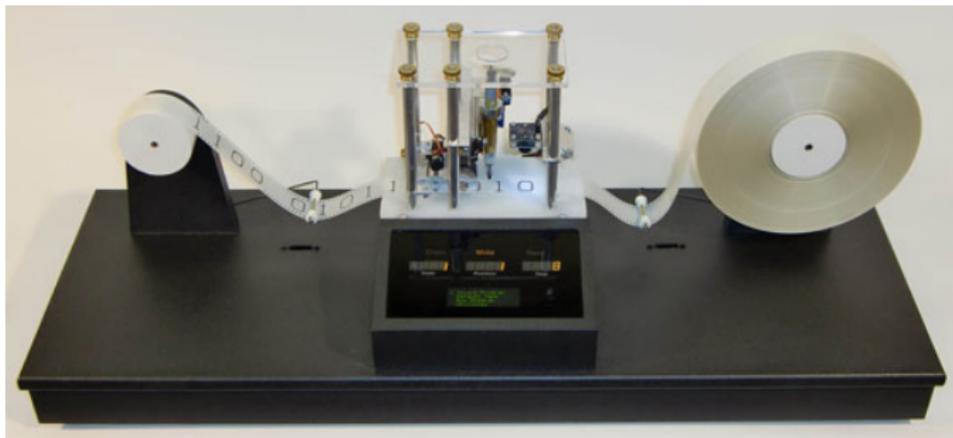


Figura: Exemplo de uma Turing física.

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-Completos fazem parte de nossas vidas...

Na estatística sempre nos deparamos com problemas NP-Completos, 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

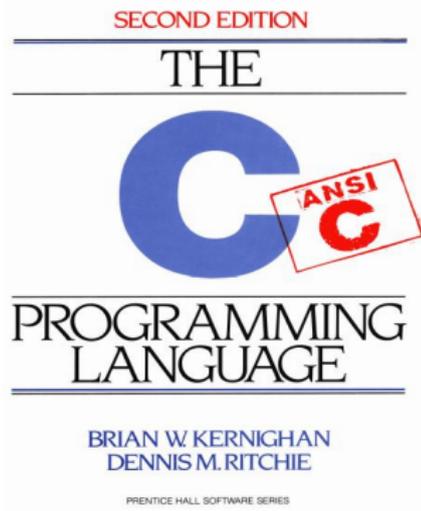


Figura: Ótimo livro sobre a linguagem C.

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

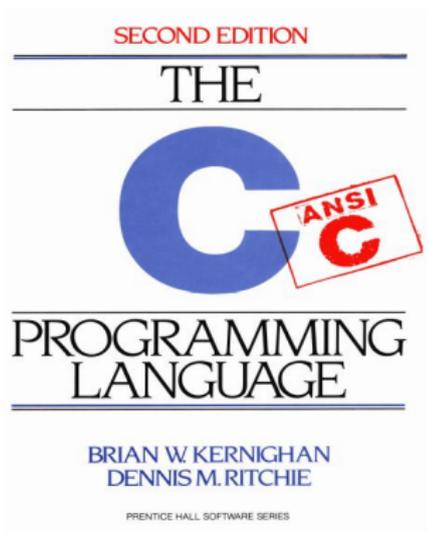


Figura: Ótimo livro sobre a linguagem C.

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

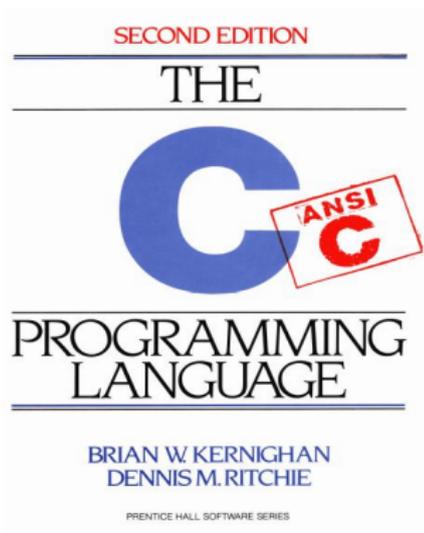


Figura: Ótimo livro sobre a linguagem C.

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.

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%.
E~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

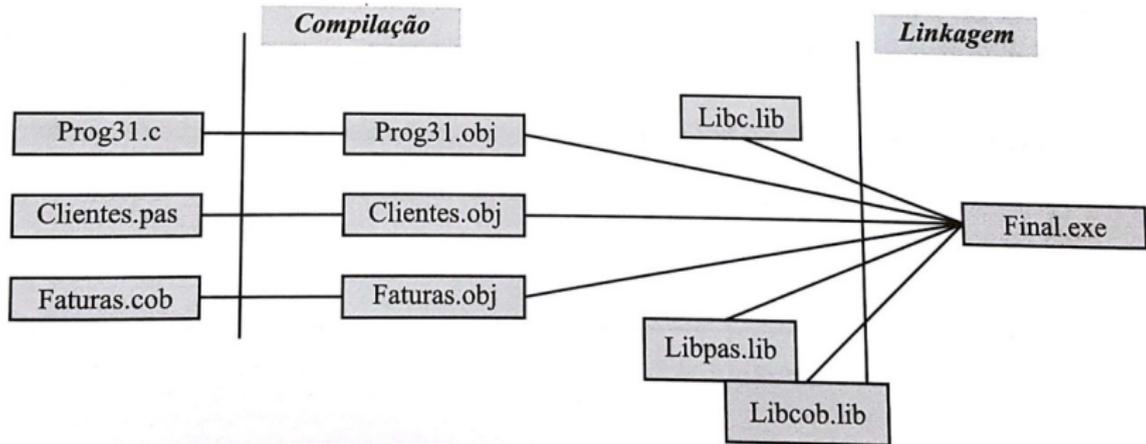


Figura: Diagrama (funcionamento de uma compilador).

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 muitos 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.

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.

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**.

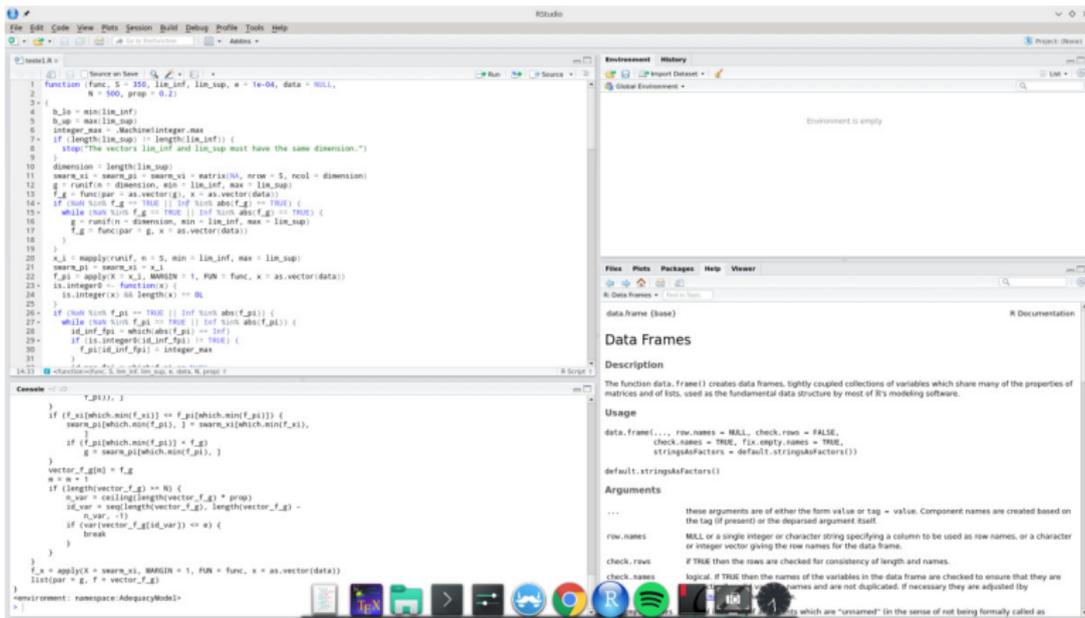


Figura: RStudio (IDE de programação em R).



Free as in Freedom

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/>.

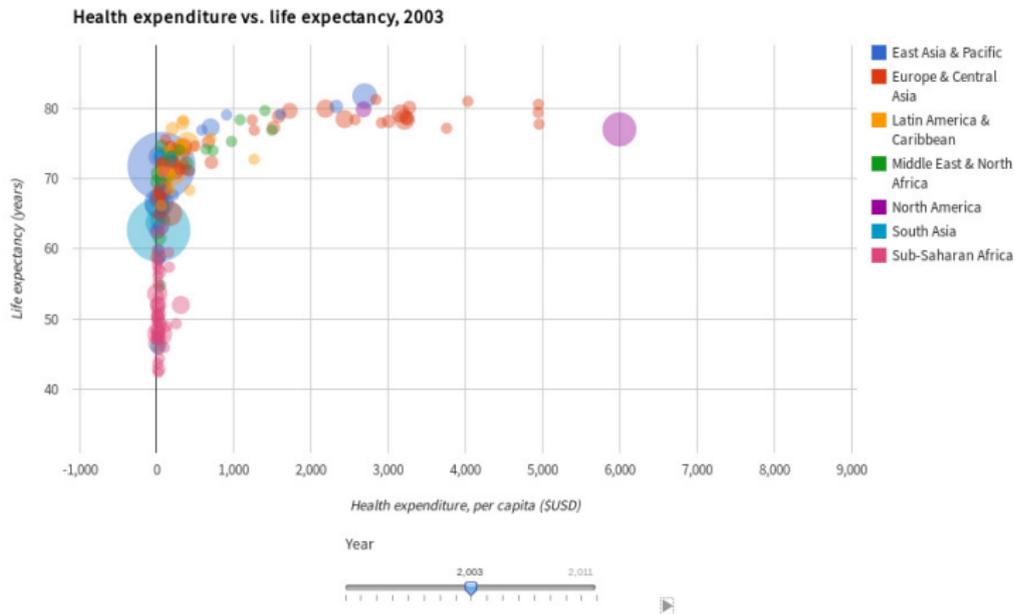


Figura: Exemplo de gráfico interativo usando **Shiny**.

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>.

RCode

The screenshot displays the RStudio IDE interface. The editor window contains R code for data manipulation and plotting. The console shows the execution of the code, including the creation of a data frame and the execution of a gg2 plot. The plot window shows a density plot of the variable 'x'.

```
1 N[is.na(N)] <- 0
2 O[is.na(O)] <- 0
3
4 #X: the total cumulative matrix
5 X <- (N - O)^N*(row(N) <- col(N))
6 X[row(X) < col(X) > n + 1] <- 0
7
8 rownames(X) <- rownames(O) <- rownames(N) <- paste("Year", 1
9
10 colnames(X) <- colnames(O) <- colnames(N) <- paste("Dev", 1
```

```
Call:
gg2(formula = vy ~ ux)
Coefficients:
(Intercept)    ux
      0.1364    0.1960
```

The plot window shows a density plot of the variable 'x' with the following statistics:

Statistic	Value
MEAN	8.500000
SD	4.6721
MIN	1
MAX	15

Figura: IDE RCode para programação em linguagem R.

Estrutura de dados em R

As estruturas das 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.

Estrutura de dados em R

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);

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);

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.

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.

Acessando elementos

Seja x um vetor atômico qualquer. Podemos acessar um elemento, ou conjunto de elementos do vetor x utilizando a seguinte sintaxe.

Sintaxe: $x[\text{posições}]$, em que posições é o vetor de posições ao qual queremos acessar.

Exemplo: Corra o código abaixo:

```
1: x <- c(2, 7, 8, 9, 7)
2: y <- x[c(2,4)]
3: y
#> [1] 7 9
```

Nota: Maiores detalhes serão apresentados mais a frente ao estudarmos operadores de subconjuntos.

Acessando elementos

Nota: Lembre-se que R é uma linguagem vetorial. Sendo assim, em diversas situações não há a necessidade de utilizar loops para acessar cada elemento de um vetor quando precisamos aplicar uma função ou realizar uma operação em seus elementos.

```
1: a <- c(7.1, 2.3, 3L, TRUE)
2: b <- a + 1 # Somando 1 ao vetor "a".
3: c <- a^0.5 # Exponencializando o vetor "a".
4: a
#> [1] 7.1 2.3 3.0 1.0
5: b
#> [1] 8.1 3.3 4.0 2.0
6: c
#> [1] 2.664583 1.516575 1.732051 1.000000
```

Missing Values/Not Available

Valores faltantes (*Not Available*) são especificados por NA. Em R, NA é um vetor de comprimento 1. Além disso, NA sempre será coagido para o tipo correto se utilizado dentro de `c()`.

Também poderemos especificar o tipo de valor faltante utilizando:

Missing Values/Not Available

Valores faltantes (*Not Available*) são especificados por NA. Em R, NA é um vetor de comprimento 1. Além disso, NA sempre será coagido para o tipo correto se utilizado dentro de `c()`.

Também poderemos especificar o tipo de valor faltante utilizando:

- `NA_real_`: um vetor de precisão dupla;

Missing Values/Not Available

Valores faltantes (*Not Available*) são especificados por `NA`. Em R, `NA` é um vetor de comprimento 1. Além disso, `NA` sempre será coagido para o tipo correto se utilizado dentro de `c()`.

Também poderemos especificar o tipo de valor faltante utilizando:

- `NA_real_`: um vetor de precisão dupla;
- `NA_integer_`: um vetor de inteiros;

Missing Values/Not Available

Valores faltantes (*Not Available*) são especificados por `NA`. Em R, `NA` é um vetor de comprimento 1. Além disso, `NA` sempre será coagido para o tipo correto se utilizado dentro de `c()`.

Também poderemos especificar o tipo de valor faltante utilizando:

- `NA_real_`: um vetor de precisão dupla;
- `NA_integer_`: um vetor de inteiros;
- `NA_character_`: um vetor de caracteres.

Missing Values/Not Available

Valores faltantes (*Not Available*) são especificados por NA. Em R, NA é um vetor de comprimento 1. Além disso, NA sempre será coagido para o tipo correto se utilizado dentro de `c()`.

Também poderemos especificar o tipo de valor faltante utilizando:

- `NA_real_`: um vetor de precisão dupla;
- `NA_integer_`: um vetor de inteiros;
- `NA_character_`: um vetor de caracteres.

Observação: Lembre-se que na estatística é muito comum lacunas em conjuntos de dados. Na análise de sobrevivência, por exemplo, é comum a presença de censura nos dados.

Tipos e Testes

Em R é possível verificar o tipo de um vetor como apresentado anteriormente com a função `typeof()`. Além disso, poderemos chegar o tipo específico dos elementos em um vetor utilizando as funções que seguem:

Em R é possível verificar o tipo de um vetor como apresentado anteriormente com a função `typeof()`. Além disso, poderemos chegar o tipo específico dos elementos em um vetor utilizando as funções que seguem:

- **Ex:** `is.character()`. Checa se é um vetor de caracteres;

Tipos e Testes

Em R é possível verificar o tipo de um vetor como apresentado anteriormente com a função `typeof()`. Além disso, poderemos chegar o tipo específico dos elementos em um vetor utilizando as funções que seguem:

- **Ex:** `is.character()`. Checa se é um vetor de caracteres;
- **Ex:** `is.double()`. Checa se é um vetor de números reais;

Tipos e Testes

Em R é possível verificar o tipo de um vetor como apresentado anteriormente com a função `typeof()`. Além disso, poderemos chegar o tipo específico dos elementos em um vetor utilizando as funções que seguem:

- **Ex:** `is.character()`. Checa se é um vetor de caracteres;
- **Ex:** `is.double()`. Checa se é um vetor de números reais;
- **Ex:** `is.integer()`. Checa se é um vetor de inteiros;

Tipos e Testes

Em R é possível verificar o tipo de um vetor como apresentado anteriormente com a função `typeof()`. Além disso, poderemos chegar o tipo específico dos elementos em um vetor utilizando as funções que seguem:

- **Ex:** `is.character()`. Checa se é um vetor de caracteres;
- **Ex:** `is.double()`. Checa se é um vetor de números reais;
- **Ex:** `is.integer()`. Checa se é um vetor de inteiros;
- **Ex:** `is.logical()`. Checa se é um vetor de valores lógicos;

Em R é possível verificar o tipo de um vetor como apresentado anteriormente com a função `typeof()`. Além disso, poderemos chegar o tipo específico dos elementos em um vetor utilizando as funções que seguem:

- **Ex:** `is.character()`. Checa se é um vetor de caracteres;
- **Ex:** `is.double()`. Checa se é um vetor de números reais;
- **Ex:** `is.integer()`. Checa se é um vetor de inteiros;
- **Ex:** `is.logical()`. Checa se é um vetor de valores lógicos;
- **Ex:** `is.atomic()`. Checa se é um vetor atômico.

Tipos e Testes

Exemplo: Execute o código R que segue logo abaixo:

```
1: int_var <- c(1L, 6L, 10L)
2: typeof(int_var)
#> [1] "integer"
3: is.integer(int_var)
#> [1] "TRUE"
4: is.atomic(int_var)
#> [1] TRUE
5:
6: dbl_var <- c(1, 2.5, 4.5)
7: typeof(dbl_var)
#> [1] "double"
8: is.double(dbl_var)
#> [1] TRUE
9: is.atomic(dbl_var)
#> [1] TRUE
```

Observação

Uma forma geral para checar se um vetor é formado por números inteiros ou de precisão dupla (double) é utilizando a função `is.numeric()`.

Exemplo:

```
is.numeric(int_var)
#> [1] TRUE
is.numeric(dbl_var)
#> [1] TRUE
```

Pergunta: Ocorrerá erro se for tentado criar um vetor, em R, com objetos de tipo diferente?

Coerção

Pergunta: Ocorrerá erro se for tentado criar um vetor, em R, com objetos de tipo diferente?



Pergunta: Ocorrerá erro se for tentado criar um vetor, em R, com objetos de tipo diferente?



Resposta direta: Não.

Pergunta: Ocorrerá erro se for tentado criar um vetor, em R, com objetos de tipo diferente?

Uma resposta um pouco maior: Muito embora não ocorrerá erro ao tentarmos criar um vetor com elementos de diferentes tipos, é importante entender que o vetor criado será formado por elementos do mesmo tipo, como dito anteriormente. Neste caso, R irá converter os elementos de um vetor para um tipo específico. Para saber qual é esse tipo, é preciso entender a regra de coerção da linguagem R.

Regra de Coerção

Se um vetor for criado com elementos de tipos básicos diferentes, os elementos serão convertidos para o tipo mais flexível. Logo abaixo segue uma sequência de desigualdades em que $A > B$ significa que o tipo de A é mais flexível que o tipo de B .

Regra de Coerção

Se um vetor for criado com elementos de tipos básicos diferentes, os elementos serão convertidos para o tipo mais flexível. Logo abaixo segue uma sequência de desigualdades em que $A > B$ significa que o tipo de A é mais flexível que o tipo de B .

character > double > integer > logical.

Regra de Coerção

Se um vetor for criado com elementos de tipos básicos diferentes, os elementos serão convertidos para o tipo mais flexível. Logo abaixo segue uma sequência de desigualdades em que $A > B$ significa que o tipo de A é mais flexível que o tipo de B .

character > double > integer > logical.

Exemplo:

```
1: str(c("a", 1))  
#> chr [1:2] "a" "1"  
2: str(c(T, T, F, 2))  
#> num [1:4] 1 1 0 2
```

Exemplo: Nesse exemplo é possível observar que se o tipo mais flexível é **integer** ou **double** e o tipo menos flexível é **logical**, os valores booleanos **TRUE** será convertido para **1** e **FALSE** será convertido para **0**.

Essa característica de converter tipos booleanos para **0** (se **FALSE**) ou **1** (se **TRUE**) permite, por exemplo, com que possamos contar o número de **TRUE**'s ou calcular a proporção destes.

Exemplo:

```
1: x <- c(FALSE, FALSE, TRUE)
2: as.numeric(x)
#> [1] 0 0 1
3:
4: # Total de TRUEs
5: sum(x)
#> [1] 1
6:
7: # Proporção de TRUEs
8: mean(x)
#> [1] 0.3333
```

Nota: Três novas funções foram apresentadas no código logo acima:

Nota: Três novas funções foram apresentadas no código logo acima:

- ① `as.numeric()`: Esta função aplicada ao vetor `x` forçará o vetor a coagir para um tipo de dados específico;

Nota: Três novas funções foram apresentadas no código logo acima:

- ① `as.numeric()`: Esta função aplicada ao vetor `x` forçará o vetor a coagir para um tipo de dados específico;
- ② `sum()`: Esta função aplicada ao vetor `x` fornecerá a soma do vetor;

Nota: Três novas funções foram apresentadas no código logo acima:

- ① `as.numeric()`: Esta função aplicada ao vetor `x` forçará o vetor a coagir para um tipo de dados específico;
- ② `sum()`: Esta função aplicada ao vetor `x` fornecerá a soma do vetor;
- ③ `mean()`: Esta função aplicada ao vetor `x` fornecerá a média aritmética do vetor.

Nota Importante: `as.numeric()` é um objeto mais geral e converterá os elementos de um vetor para o tipo de dados `double`. Porém, também é possível utilizar o objeto `as.double()` para produzir o mesmo resultado. No momento, o que pode ser dito para não parecer estranho ao iniciante da linguagem R é que R possui, basicamente, dois sistemas de orientação a objeto que são mais utilizados, sendo eles o sistema **S3** ao qual o objeto `as.double()` pertence e o sistema **S4** ao qual o objeto `as.numeric()` pertence.

Dica: Na grande maioria dos casos, qualquer uma das opções será uma escolha conveniente.

Coerção

Da mesma forma que podemos converter os elementos de um vetor para o tipo específico `double`, também podemos fazer isto com os outros tipos de dados.

Da mesma forma que podemos converter os elementos de um vetor para o tipo específico `double`, também podemos fazer isto com os outros tipos de dados.

- `as.character()`: convertendo os elementos do vetor para o tipo básico **character**;

Coerção

Da mesma forma que podemos converter os elementos de um vetor para o tipo específico `double`, também podemos fazer isto com os outros tipos de dados.

- `as.character()`: convertendo os elementos do vetor para o tipo básico **character**;
- `as.integer()`: convertendo os elementos do vetor para o tipo básico **integer**;

Da mesma forma que podemos converter os elementos de um vetor para o tipo específico `double`, também podemos fazer isto com os outros tipos de dados.

- `as.character()`: convertendo os elementos do vetor para o tipo básico **character**;
- `as.integer()`: convertendo os elementos do vetor para o tipo básico **integer**;
- `as.logical()`: convertendo os elementos do vetor para o tipo básico **logical**.

Lembre-se

A coerção ocorre, frequentemente de forma automática. A maioria das funções matemáticas (+, log, abs, etc.) coagirá para um valor do tipo double ou inteiro. Já os operadores lógicos (&, |, any, etc.) coagirão para um valor lógico. Mas lembre-se que é possível forçar a coação com as funções `as.logical()`, `as.character()`, `as.integer()` e `as.numeric()`.

Nota: Podem existir casos de surgimento de mensagens de avisos (Warnings) em que se a coerção implicar em perda de informação.

Algumas funções matemáticas

Tabela: Algumas funções matemáticas/estatísticas aplicadas ao vetor x .

Função em R	Finalidade
<code>sum(x)</code>	Retorna a soma do vetor x
<code>mean(x)</code>	Retorna a média do vetor x
<code>abs(x)</code>	Retorna o vetor com os valores absolutos
<code>sqrt(x)</code>	Retorna o vetor com as raízes quadradas
<code>sort(x)</code>	Retorna o vetor x ordenado de forma crescente
<code>sort(x, decreasing=TRUE)</code>	Retorna o vetor x ordenado de forma decrescente
<code>exp(x)</code>	Retorna o vetor exponencial
<code>log(x)</code>	Retorna o vetor logaritmo natural
<code>log(x, base=10)</code>	Retorna o vetor com o logaritmo na base 10
<code>sin(x), cos(x), tan(x)</code>	Retorna o vetor com resultados trigonométricos
<code>length(x)</code>	Cardinalidade do vetor x
<code>factorial(x)</code>	Fatorial dos elementos de x
<code>choose(3,2)</code>	Combinação três, dois a dois
<code>summary(x)</code>	Medidas descritivas do vetor x



Figura: Procure ajuda!

Até o momento foram apresentadas algumas funções. Muitas outras funções irão surgir no decorrer das aulas. A melhor forma de conseguir ajuda sobre uma determinada função além de estudar pelas aulas ministradas é ler a documentação destas funções. Na documentação apresenta exemplos do uso da função o que poderá ajudar bastante.



Figura: Procure ajuda!

Para algumas funções da tabela acima, por exemplo, vimos o emprego de alguns argumentos. Maiores detalhes sobre estes argumentos bem como outros argumentos suportados por uma função poderão ser encontrados nas documentações das funções.

Existem basicamente duas formas equivalentes de acessar a documentação de uma função disponibilizada na linguagem base ou empacotadas que pode ser `?nome_da_funcao` ou `help("nome_da_funcao")`.

Exercício: Descubra, lendo a documentação, o que fazem as funções `ls` e `rm`.

Uma outra forma de encontrar informações sobre uma determinada função é utilizando a função `find()`.

```
1: find("ls")  
#> [1] "package:base"
```

Nota: A função `find("funcao")` irá procurar a existência da função `funcao()` nos pacotes instalados e carregados na linguagem.

Nota: A função `find("funcao")` irá procurar a existência da função `funcao()` nos pacotes instalados e carregados na linguagem.

A função `find(nome_da_funcao)` fornece o pacote ao qual o nome da função pertence. No caso do exemplo logo acima, a função `ls` pertence ao pacote **base** da linguagem R. O pacote **base** está contido em toda instalação básica de R.

Nota: A função `find("funcao")` irá procurar a existência da função `funcao()` nos pacotes instalados e carregados na linguagem.

A função `find(nome_da_funcao)` fornece o pacote ao qual o nome da função pertence. No caso do exemplo logo acima, a função `ls` pertence ao pacote **base** da linguagem R. O pacote **base** está contido em toda instalação básica de R.

Há outros pacotes, sob os termos da licença $\text{GPL} \geq 2$ com diversas funções disponíveis para os programadores de R, porém, daremos mais detalhes mais a frente.

Um outro comando bastante útil, principalmente nos casos em que lembramos parcialmente o nome de uma função R é o comando `apropos()`.

Uso: `apropos("nome_da_funcao")`.

Exemplo:

```
1: apropos("ps")
#> [1] ".standard_regexps"      "cairo_ps"      "dev.copy2eps"
#> [4] "fileSnapshot"            "Ops"           "Ops.data.frame"
#> [7] "Ops.Date"                 "Ops.difftime" "Ops.factor"
#> [10] "Ops.numeric_version"     "Ops.ordered"  "Ops.POSIXt"
#> [13] "ps.options"              "psigamma"     "psignrank"
#> [16] "pso"                      "setEPS"       "setPS"
#> [19] "supsmu"                  "zapsmall"
```

Como mencionado anteriormente, as documentações de uma função R de funções presentes no pacote **base** ou empacotadas em outros pacotes que são distribuídos pela comunidade R possuem exemplos de utilização. Podemos acessar a documentação e copiar e executar um exemplo específico no prompt de comando ou podemos executar todos exemplos de uso de uma função por meio do comando `example()`.

Exemplo:

```
1: example(rm)
```

```
rm> tmp <- 1:4
```

```
rm> ## work with tmp and cleanup
```

```
rm> rm(tmp)
```

E como faço para listar todos os objetos de um determinado pacote?

E como faço para listar todos os objetos de um determinado pacote?

Resposta: Corra o comando:

```
objects(grep("base", search()))
```

Nesse exemplo, será retornado uma lista com todos os objetos contidos no pacote **base**. Atualmente, o pacote **base** apresenta 1218 objetos.

E como faço para listar todos os objetos de um determinado pacote?

Resposta: Corra o comando:

```
objects(grep("base", search()))
```

Nesse exemplo, será retornado uma lista com todos os objetos contidos no pacote **base**. Atualmente, o pacote **base** apresenta 1218 objetos.

Exercício: O que faz o código abaixo?

```
length(objects(grep("base", search())))
```

Exercício: Para que serve o argumento `ignore.case` da função `grep()`? Lembre-se, agora você já sabe consultar a documentação da linguagem R. Apresente um exemplo do uso da função `grep()` modificando o argumento `ignore.case`.

Exercício: Para que serve o argumento `ignore.case` da função `grep()`? Lembre-se, agora você já sabe consultar a documentação da linguagem R. Apresente um exemplo do uso da função `grep()` modificando o argumento `ignore.case`.

Entendendo melhor o código acima, note que `search()` é uma função passada como argumento para a função `grep()` que por sua vez é uma função que é passada como argumento para a função `objects` e que esta é uma função que é passada para a função `length()`.

Exercício: Para que serve o argumento `ignore.case` da função `grep()`? Lembre-se, agora você já sabe consultar a documentação da linguagem R. Apresente um exemplo do uso da função `grep()` modificando o argumento `ignore.case`.

Entendendo melhor o código acima, note que `search()` é uma função passada como argumento para a função `grep()` que por sua vez é uma função que é passada como argumento para a função `objects` e que esta é uma função que é passada para a função `length()`.

Pergunta: O que cada uma dessas funções faz?

Respostas:

- 1 `search()`: Lista um vetor de caracteres com os nomes dos pacotes carregados.

Respostas:

- ① `search()`: Lista um vetor de caracteres com os nomes dos pacotes carregados.
- ② `grep("base", search())`: Busca "base" no vetor obtido como retorno da função `search()`. A função `grep()` retorna as posições do vetor retornado por `search()` que contém "strings" que contenham a palavra "base".

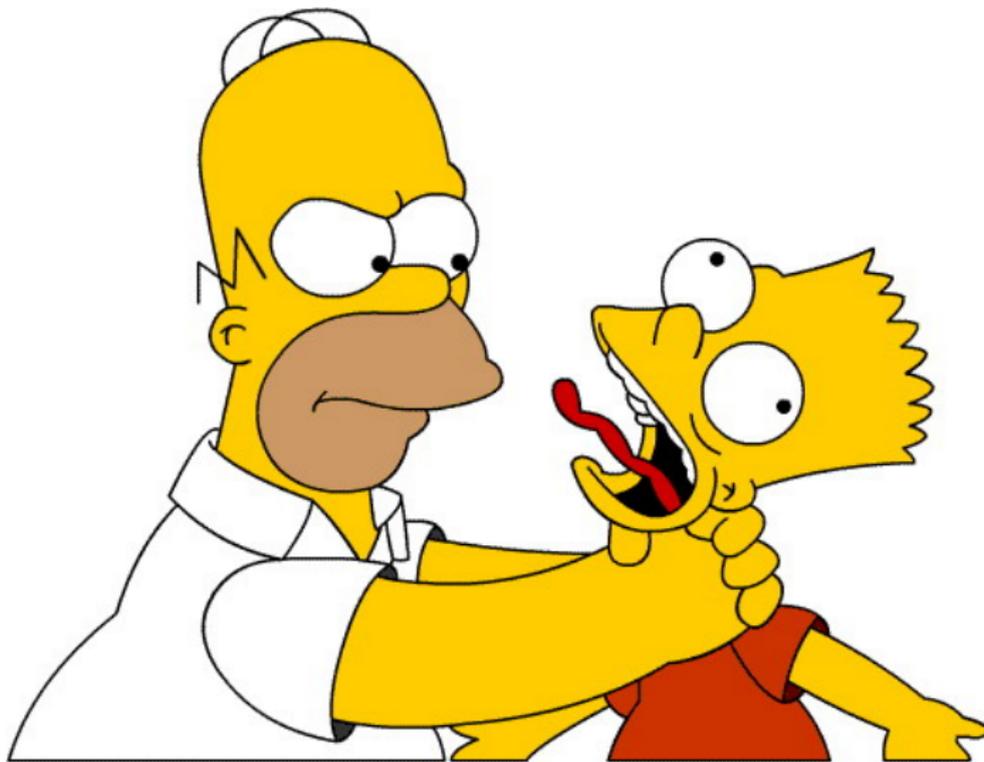
Respostas:

- ① `search()`: Lista um vetor de caracteres com os nomes dos pacotes carregados.
- ② `grep("base", search())`: Busca "base" no vetor obtido como retorno da função `search()`. A função `grep()` retorna as posições do vetor retornado por `search()` que contém "strings" que contenham a palavra "base".
- ③ `objects()`: Lista todos os objetos do pacote correspondente.

Exercício: Explique passo a passo a sintaxe abaixo. Certifique-se que tenha entendido de fato o código abaixo.

```
search()[grep("gr", search())]
```

E se eu misturar letras maiúsculas e minúsculas do nome da função?



Sério mesmo? Não há nenhuma forma de tentar encontrar informações sobre uma função caso troque letras maiúsculas por minúsculas ou o contrário?

Resposta: Claro que tem! Mas não queira mais do que isso.

Resposta: Claro que tem! Mas não queira mais do que isso.



Nestas situações em que não compreendemos se uma determinada letra é escrita em maiúscula ou minúscula, utilize a função `??`.

Uso: `??nome_da_funcao`.

Exemplo: Corra o código `??Mean` e `??mean` ou `??MeaN`. Diversas funções que fazem dessas expressões regulares irão ser retornadas.

Nestas situações em que não compreendemos se uma determinada letra é escrita em maiúscula ou minúscula, utilize a função `??`.

Uso: `??nome_da_funcao`.

Exemplo: Corra o código `??Mean` e `??mean` ou `??MeaN`. Diversas funções que fazem dessas expressões regulares irão ser retornadas.

Dica: Na internet também é possível obter diversas informações sobre a linguagem R. Há diversos materiais (documentação oficial, livros, apostilas, blogs, grupos, etc) focado para o aprendizado da linguagem R.

Listas são também vetores não-atômicos. As listas diferem de vetores atômicos pelo fato de poder conter elementos de tipos diferentes, incluindo listas. Para se construir uma lista, deve ser utilizado o comando `list()` ao invés de `c()`.

Exemplo: Criando uma lista em R.

```
1: x <- list(1:3, "a", c(TRUE, FALSE, TRUE), c(2.3, 5.9))
2: str(x)
#> List of 4
#> $ : int [1:3] 1 2 3
#> $ : chr "a"
#> $ : logi [1:3] TRUE FALSE TRUE
#> $ : num [1:2] 2.3 5.9
```

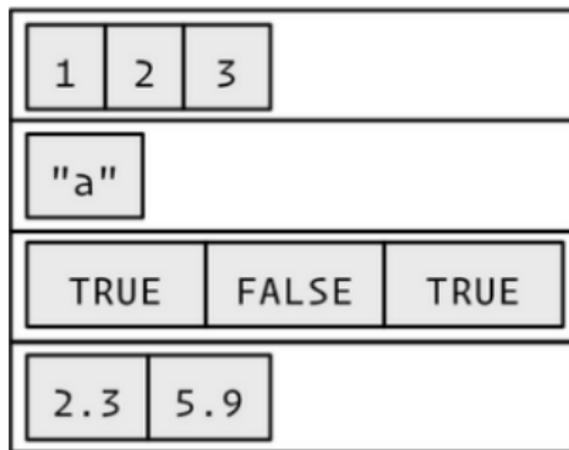


Figura: Estrutura hierárquica do objeto x.

Nota: No exemplo anterior foi passado para a lista `x` a expressão `1:3`. Tal expressão, em R, significa que estamos criando um vetor com **sequência** de valores de intervalos unitários de 1 a 3. Perceba que os limites são inclusos nesse caso.

Exemplo: Corra o código que segue:

```
1: seq1 <- 1.7:7.7
2: str(seq1)
#> num [1:7] 1.7 2.7 3.7 4.7 5.7 6.7 7.7
3:
4: seq2 <- 1.7:6.1
5: str(seq2)
#> num [1:5] 1.7 2.7 3.7 4.7 5.7
```

Nota: Caso haja o interesse de criar uma sequência de valores de intervalos não unitários, isso é possível utilizando a função `seq()` e alterando o parâmetro `by` utilizado para essa finalidade. Para maiores detalhes sobre a função e seus outros argumentos, basta consultar a documentação, isto é, faça `help(seq)` ou `?seq`. **Ex:** Corra o código `seq(from = 1, to = 3)`.

Exercício: Construa um vetor de números pares de 2 a 100 e um outro vetor de números ímpares de números de 1 a 99.

Exercício: Construa uma sequência de números no intervalo fechado $[1, 100]$ por intervalos de comprimento 0.1. Construa uma outra sequência no intervalo fechado $[1, 100]$ com 37 elementos.

Assim com nos vetores atômicos, as listas também podem apresentar recursividade. Corra o exemplo que segue.

Exemplo:

```
1: x <- list(list(list(list())))
2: str(x)
#> List of 1
#> $ :List of 1
#> ..$ :List of 1
#> .. ..$ : list()
3: is.recursive(x)
#> [1] TRUE
```

```
1: x <- list(list(list(1)))  
2: str(x)
```

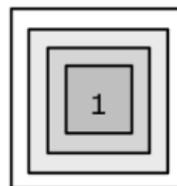


Figura: Exemplo de uma lista recursiva (concatenação de listas).

Nota: Ao contrário de listas, note que a concatenação de um vetor não fornece uma estrutura recursiva.

Uso do comando `c()` em listas

O comando `c()` poderá ser utilizado para combinar várias listas em uma. Caso venha ser dada uma combinação atômica de vetores e listas, a função `c()` irá coagir os vetores para uma lista.

Exemplo: Compare as listas armazenadas nas variáveis `x` e `y` no exemplo que segue.

```
1: x <- list(list(1, 2), c(3, 4))
2: y <- c(list(1, 2), c(3, 4))
3: str(x)
#> List of 2
#> $ :List of 2
#> ..$ : num 1
#> ..$ : num 2
#> $ : num [1:2] 3 4
```

Exemplo: Compare as listas armazenadas nas variáveis `x` e `y` no exemplo que segue.

```
4: str(y)
#> List of 4
#> $ : num 1
#> $ : num 2
#> $ : num 3
#> $ : num 4
```

Exemplo: Compare as listas armazenadas nas variáveis `x` e `y` no exemplo que segue.

```
4: str(y)
#> List of 4
#> $ : num 1
#> $ : num 2
#> $ : num 3
#> $ : num 4
```

Nota: Seja `x` uma estrutura de dados qualquer. No R podemos utilizar a função `View(x)` para visualizar de forma tabelada tal estrutura de dados. Equivale a clicar na aba **Environment** no **RStudio** e selecionar o objeto.

Exercício: Da mesma forma que o exemplo anterior, corra o código abaixo e assimile as saídas.

```
1: l <- c(list(c(3,2),"a"),c(1,2))
2: str(l)
#> List of 4
#> $ : num [1:2] 3 2
#> $ : chr "a"
#> $ : num 1
#> $ : num 2
3: m <- list(list(c(3,2),"a"),c(1,2))
4 : str(m)
#> List of 2
#> $ :List of 2
#> ..$ : num [1:2] 3 2
#> ..$ : chr "a"
#> $ : num [1:2] 1 2
```

Nota: É possível checar se um objeto é um lista. Para isso, utiliza-se a função `is.list()`. Nesse caso, será retornado um valor booleano (**TRUE** ou **FALSE**). Além disso, o comando `typeof()` também poderá ser utilizado, porém, neste caso será retornado uma string informando o tipo do objeto. No caso do objeto venha a ser uma lista, o retorno observado será "list".

Exemplo: Corra o código que segue:

```
is.list(mtcars)
```

```
#> [1] TRUE
```

```
mod <- lm(mpg ~ wt, data = mtcars)
```

```
is.list(mod)
```

```
#> [1] TRUE
```

Algumas observações para o exemplo anterior:

- ① O objeto `mtcars` nada mais é do que um conjunto de dados disponível em qualquer instalação básica da linguagem R no pacote **datasets**. Para maiores detalhes, faça `help(mtcars)`;
- ② A função `lm` é responsável com o ajustamento de modelos lineares. No caso do exemplo, foi ajustado um modelo linear simples de regressão com as variáveis **mpg** e **wt** pertencentes ao conjunto de dados **mtcars**. Para maiores detalhes, faça `help(lm)`.

Nota: Também é possível converter uma lista para um vetor atômica utilizando o comando `unlist()`. Porém, lembre-se, os elementos da lista serão coagidos para o tipo de dados mais flexível.

Exemplo: Corra o código que segue:

```
1: m <- list(list(c(3,2),"a"),c(1,2))
2: unlist(m)
#> [1] "3" "2" "a" "1" "2"
```

Exercícios: Corra os códigos abaixo e explique o por quê das saídas obtidas.

- Qual a saída esperada para os códigos `c(1, FALSE)`, `c("a", 1)`, `c(list(1), "a")`, `c(TRUE, 1L)`? Explique.
- Por que `1 == "1"` é igual à **TRUE**? Além disso, responda o por quê `-1 < FALSE` é igual à **TRUE**.

Muitas vezes é útil nomear um vetor. Em R, é possível nomear um vetor de três maneiras:

Muitas vezes é útil nomear um vetor. Em R, é possível nomear um vetor de três maneiras:

- 1 Nomeando no momento em que criamos o vetor `x`: `x <- c(a = 1, b = 2, c = 3)`;

Muitas vezes é útil nomear um vetor. Em R, é possível nomear um vetor de três maneiras:

- ① Nomeando no momento em que criamos o vetor `x`: `x <- c(a = 1, b = 2, c = 3)`;
- ② Nomeando um vetor `x` previamente criado: `x <- 1:3`;
`names(x) <- c("a", "b", "c")`;

Muitas vezes é útil nomear um vetor. Em R, é possível nomear um vetor de três maneiras:

- ① Nomeando no momento em que criamos o vetor `x`: `x <- c(a = 1, b = 2, c = 3)`;
- ② Nomeando um vetor `x` previamente criado: `x <- 1:3`;
`names(x) <- c("a", "b", "c")`;
- ③ Utilizando a função `setNames()`: `x <- setNames(1:3, c("a", "b", "c"))`.

Exemplo: Corra o código que segue:

```
1: y <- c(a = 1, 2, 3)
2: names(y)
#> [1] "a" "" ""
3:
4: z <- c(1, 2, 3)
5: names(z)
#> NULL
```

Nota: Nem todos os elementos de um vetor precisa ter um nome. Caso algum deles não sejam será retornado uma string varia. Caso nenhum elemento venha ter um nome, NULL será retornado.

Caso tenhamos o interesse de eliminar os nomes dos elementos de um vetor poderemos utilizar a função `unname()` ou atribuir `NULL` à `names(x)`.

Exemplo: Corra o código que segue:

```
1: x <- c(a = 1, b = 2, c = 3)
2: x
#> a b c
#> 1 2 3
4: x <- unname(x)
5: x
#> [1] 1 2 3
6: x <- c(a = 1, b = 2, c = 3)
7: names(x) <- NULL
```

O que é um fator?

Fator é um vetor que pode conter apenas valores predefinidos, e é usado para armazenar dados categóricos.

Exemplo:

```
1: x <- factor(c("a", "b", "b", "a"))
2: x
#> [1] a b b a
#> Levels: a b
3: class(x)
#> [1] "factor"
4: levels(x)
#> [1] "a" "b"
```

Nota: Caso o programador tente atribuir um elemento que não corresponde aos níveis admitidos pelo fator ocorrerá uma mensagem de erro e na posição do elemento atribuído irá conter NA.

Exemplo: Corra o código que segue:

```
1: x <- factor(c("a", "b", "b", "a"))
2: x[2] <- "c"
3:
#> Warning: invalid factor level, NA generated
4: x
#> [1] a <NA> b a
#> Levels: a b
```

Importante: Não podemos combinar fatores.

Exemplo: Corra o código que segue:

```
# Obs: Não é possível combinar fator. Algo inesperado
# será retornado.
1: c(factor("a"), factor("b"))
#> [1] 1 1
```

Podemos utilizar a função `table()` para construir uma tabela de frequência com um objeto da classe **factor**.

Exemplo: Corra o código abaixo:

```
1: sex_char <- c("m", "m", "m", "f", "f", "f", "f")
2: sex_factor <- factor(sex_char, levels = c("m", "f"))
3: table(sex_factor)
#> sex_factor
#> m f
#> 3 4
```

Exemplo: Podemos atribuir os níveis em que um fator poderá assumir após sua criação:

```
1: sex_char <- c("f", "f" ,"f", "f", "f")
2: sex_factor <- factor(sex_char)
3: levels(sex_factor) <- c("m", "f")
4: table(sex_factor)
#> sex_factor
#> m f
#> 5 0
```

Fator

Exercício: O que faz a função `rev()` e o objeto `letters`?
Discuta as diferenças de `f1`, `f2` e `f3`.

```
1: f1 <- factor(letters)
2: levels(f1) <- rev(levels(f1))
3: f2 <- rev(factor(letters))
4: f3 <- factor(letters, levels = rev(letters))
5: f1
#> [1] z y x w v u t s r q p o n m l k j i h g f e d c
#> Levels: z y x w v u t s r q p o n m l k j i h g f e d
6: f2
#> [1] z y x w v u t s r q p o n m l k j i h g f e d c
#> Levels: a b c d e f g h i j k l m n o p q r s t u v w
7: f3
#> [1] a b c d e f g h i j k l m n o p q r s t u v w x
#> Levels: z y x w v u t s r q p o n m l k j i h g f e d
```

Removendo objetos

Na estatística trabalhamos com objetos que, em geral, consomem muita memória. Dessa forma, parece bem conveniente entendermos como remover tais objetos da memória. Porém, precisamos antes identificar quais objetos estão carregados na memória. Para isso utilize a função `ls()` ou `objects()`.

Exemplo:

```
1: s1 <- "Eu"  
2: s2 <- "amo"  
3: s3 <- "a"  
4: s4 <- "ESTATÍSTICA."  
5: s5 <- paste(s1, s2, s3, s4, sep = " ")  
6: ls()  
7: objects.size(ls())
```

Removendo objetos

Exercício: Explique resumidamente a utilidade da função `paste()`. Obtenha detalhes na documentação da linguagem. Apresente dois exemplos do uso da função `paste()` fazendo uso dos argumentos `sep` e `collapse`, respectivamente. Perceba a diferença do uso desses argumentos.

É possível remover todos objetos com um único comando? Como?

Resposta: Sim. Utilize o comando `rm(list = ls(all = TRUE))`.

Exercício: Explique o código `rm(list = ls(all = TRUE))`.

Removendo objetos

Importante

Em situações em que desejamos remover objetos “**grandes**” (objetos que ocupam muita memória) é sempre interessante executar a função `gc()`. Ao removermos objetos R utilizando a função `rm()` automaticamente o R irá devolver ao sistema operacional a memória ocupada por esse objeto. Porém, não necessariamente isso ocorrerá de forma imediata.

Importante

Em situações em que desejamos remover objetos “**grandes**” (objetos que ocupam muita memória) é sempre interessante executar a função `gc()`. Ao removermos objetos R utilizando a função `rm()` automaticamente o R irá devolver ao sistema operacional a memória ocupada por esse objeto. Porém, não necessariamente isso ocorrerá de forma imediata.

Uma chamada da função `gc()` faz com que uma coleta de lixo ocorra

Removendo objetos

Exemplo: Corra o exemplo que segue.

```
1: x <- rnorm(1e6) # 1e6 = 1 * 10^6
2: # Busque nas documentações detalhes sobre as funções
3: # print() e object.size().
4: print(object.size(x), units = "Mb")
5: rm(x)
6: gc()
```

Removendo objetos

Exemplo: Corra o exemplo que segue.

```
1: x <- rnorm(1e6) # 1e6 = 1 * 10^6
2: # Busque nas documentações detalhes sobre as funções
3: # print() e object.size().
4: print(object.size(x), units = "Mb")
5: rm(x)
6: gc()
```

Nota: Observe e entenda o código abaixo.

```
1: 1e6 == 1 * 10^6
#> TRUE
2: 1E6 == 1 * 10^6
#> TRUE
```

Matrices



Matrizes e Arrays

Adicionar um atributo `dim()` à um vetor atômico permite que ele se comporte como uma matriz multidimensional. O caso especial do **array** é o que chamamos de **matriz** que tem apenas duas dimensões.

Matrizes e Arrays

Adicionar um atributo `dim()` à um vetor atômico permite que ele se comporte como uma matriz multidimensional. O caso especial do **array** é o que chamamos de **matriz** que tem apenas duas dimensões.

Matrizes são usadas comumente como parte da “maquinaria” matemática e estatística. O uso do array é um pouco mais raro, mas é fundamental estar ciente do seu uso.

Adicionar um atributo `dim()` à um vetor atômico permite que ele se comporte como uma matriz multidimensional. O caso especial do **array** é o que chamamos de **matriz** que tem apenas duas dimensões.

Matrizes são usadas comumente como parte da “maquinaria” matemática e estatística. O uso do array é um pouco mais raro, mas é fundamental estar ciente do seu uso.

Exercício: Leia a documentação das funções `matrix()` e `array()`. Procure entender os seus argumentos.

Adicionar um atributo `dim()` à um vetor atômico permite que ele se comporte como uma matriz multidimensional. O caso especial do **array** é o que chamamos de **matriz** que tem apenas duas dimensões.

Matrizes são usadas comumente como parte da “maquinaria” matemática e estatística. O uso do array é um pouco mais raro, mas é fundamental estar ciente do seu uso.

Exercício: Leia a documentação das funções `matrix()` e `array()`. Procure entender os seus argumentos.

Exercício: Leia a documentação da função `dim()` e construa um vetor atômico de tamanho 10 e em seguida construa uma M matriz de dimensão 5 por 2. Utilizando a matriz M , construa o vetor `myvector`.

Matrizes e Arrays

Matrizes e arrays são criados com as funções `matrix` e `array`, respectivamente, ou utilizando a forma de atribuição com o comando `dim()`.

Exemplo: Corra o exemplo que segue:

```
1: # Dois argumentos escalares são passados
2: # representando linha e coluna
3: a <- matrix(1:6, ncol = 3, nrow = 2)
4:
5: # Passando o vetor de dimensões do array.
6: b <- array(1:12, dim = c(2, 3, 2))
7:
8: # Modificando o objeto c com a função dim().
9: c <- 1:6
10: dim(c) <- c(3,2)
```

Matrizes e Arrays

Nota: Podemos utilizar as funções `length()`, `nrow()` e `ncol()` para obter a quantidade de elementos, número de linhas e número de colunas, respectivamente, de uma matriz ou array.

Exemplo: Corra o exemplo abaixo:

```
1: m <- array(1:12, dim = c(2,2,3))
2: length(m)
#> [1] 12
3: ncol(m)
#> [1] 2
4: nrow(m)
#> [1] 2
```

Nota: A função `dim()` poderá ser utilizada para obter o vetor com as dimensões linha e coluna, respectivamente. Faça `dim(m)` no exemplo anterior.

Matrizes e Arrays

Muitas vezes é útil nomear cada uma das dimensões de uma matriz ou array. Inclusive, em situações em que queremos acessar um subconjunto de um vetor, matriz ou array, podemos levar em conta tais nomes. Veremos isto mais a frente. No momento, focaremos apenas em atribuir nomes às dimensões.

No caso de matrizes (array de dimensão dois) utilizaremos as funções `rownames()` e `colnames()` para atribuir nomes às dimensões linha e coluna de uma matriz, respectivamente.

Exemplo: Corra o código seguinte:

```
1: a <- matrix(1:6, ncol = 3, nrow = 2)
2: rownames(a) <- c("A", "B")
3: colnames(a) <- c("a", "b", "c")
4: a
#>   a b c
#> A 1 3 5
#> B 2 4 6
```

Nota: Caso o array possua mais de duas dimensões, utilize a função `dimnames()` aplicado ao array. Deverá ser passado uma lista com os nomes de cada uma das dimensões do objeto.

Exemplo: Corra o código:

```
1: b <- array(1:12, c(2, 3, 2))
2: dimnames(b) <- list(c("one", "two"), c("a", "b", "c"),
3:                      c("A", "B")); b
#> , , A
#>   a b c
#> one 1 3 5
#> two 2 4 6

#> , , B
#>   a  b  c
#> one 7  9 11
#> two 8 10 12
```

Importante observar

Fazer `b <- array(1:12, c(2, 3, 2))` equivale a fazer `b <- array(data = 1:12, dim = c(2, 3, 2))`.

Notas:

Importante observar

```
Fazer b <- array(1:12, c(2, 3, 2)) equivale a fazer b <- array(data = 1:12, dim = c(2, 3, 2)).
```

Notas:

- 1 A diferença é que no primeiro caso não foi especificado os nomes dos argumentos, preocupação esta que foi tomada no segundo caso. O primeiro caso funcionará apenas se a ordem de passagem dos argumentos respeitar a ordem dos argumentos especificada em sua documentação. A ordem poderá ser alterada desde que os nomes dos argumentos venha estar devidamente especificado.

Importante observar

Fazer `b <- array(1:12, c(2, 3, 2))` equivale a fazer `b <- array(data = 1:12, dim = c(2, 3, 2))`.

Notas:

- ① A diferença é que no primeiro caso não foi especificado os nomes dos argumentos, preocupação esta que foi tomada no segundo caso. O primeiro caso funcionará apenas se a ordem de passagem dos argumentos respeitar a ordem dos argumentos especificada em sua documentação. A ordem poderá ser alterada desde que os nomes dos argumentos venha estar devidamente especificado.
- ② **É sempre uma boa prática de programação especificar os nomes dos argumentos.**

Importante

Os arrays são úteis quando necessitamos armazenar uma sequência de matrizes que poderão ser utilizadas posteriormente no código, por exemplo para fazer um produto de duas matrizes.

Importante

Os arrays são úteis quando necessitamos armazenar uma sequência de matrizes que poderão ser utilizadas posteriormente no código, por exemplo para fazer um produto de duas matrizes.

Exemplo: Corra o código abaixo.

```
1: myarray <- array(data = NA, dim = c(2, 2, 3))
2: a1 <- matrix(data = c(1,2,3,4), nrow = 2, ncol = 2)
3: a2 <- 2 * a1
4: a3 <- 3 * a2
5: myarray[ , , 1] <- a1
6: myarray[ , , 2] <- a2
7: myarray[ , , 3] <- a3
8: myarray
```

Exemplo: Corra o código que segue. O código que segue justifica bem o porquê é uma boa prática de programação considerar os nomes dos argumentos de uma função.

```
1: # Não tente rodar o código comentado abaixo
2: # Você poderá não ter memória =).
3: # b <- array(c(2, 3, 2), 1:12)
4:
5: c <- array(dim = c(2, 3, 2), data = 1:12)
```

Nota: Como apresentado anteriormente, a função `dim()` fornece um vetor com as dimensões linha e coluna de um array. Se o interesse é obter especificamente a dimensão da linha ou coluna, utilizamos a função `nrow()` ou `ncol()`, respectivamente.

É possível concatenar matrizes por linha ou coluna utilizando as funções `rbind()` e `cbind()`, respectivamente.

Exemplo: Corra o código abaixo:

```
1: v1 <- matrix(1:12, ncol = 4, nrow = 3)
2: w1 <- matrix(1:8, ncol = 4, nrow = 2 )
3: rbind(v1,w1)
4:
5: v2 <- matrix(1:12, ncol = 2, nrow = 4)
6: w2 <- matrix(1:12, ncol = 3, nrow = 4 )
7: cbind(v2,w2)
```

Observação Importante

Os elementos de uma matriz poderão ser listas. Trata-se de uma estrutura de dados “exotérica” mas que pode ser útil se desejarmos organizar nossas listas em uma estrutura semelhante à uma grade.

Exemplo: Corra o código que segue:

```
1: l <- list(1:3, "a", TRUE, 1.0)
2: dim(l) <- c(2, 2)
3: l
#>      [,1]      [,2]
#> [1,] Integer,3 TRUE
#> [2,] "a"      1
```

Exercícios: Para complementar alguns detalhes, responda o exercício abaixo.

- 1 O que a função `dim()` retorna quando aplicado a um vetor?

Exercícios: Para complementar alguns detalhes, responda o exercício abaixo.

- ① O que a função `dim()` retorna quando aplicado a um vetor?
- ② Se `is.matrix(x)` retornar `TRUE`, o que irá retornar `is.array(x)`?

Exercícios: Para complementar alguns detalhes, responda o exercício abaixo.

- 1 O que a função `dim()` retorna quando aplicado a um vetor?
- 2 Se `is.matrix(x)` retornar `TRUE`, o que irá retornar `is.array(x)`?
- 3 Como você descreveria os três objetos a seguir? Sem executar o código, o que você espera como resultado?
 - `x1 <- array(1:5, c(1, 1, 5))`
 - `x2 <- array(1:5, c(1, 5, 1))`
 - `x3 <- array(1:5, c(5, 1, 1))`

Exercícios: Responda os exercícios abaixo:

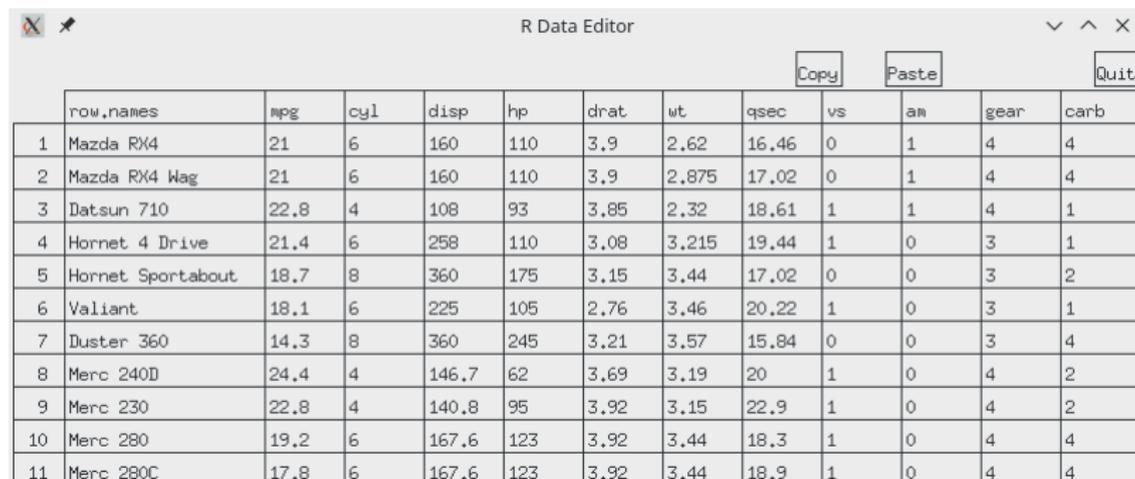
- ① Crie três vetores com três elementos **inteiros**, sejam eles os objetos x , y , z . Combine os três vetores para se tornar uma matriz 3×3 em que cada coluna representa um vetor. Altere os nomes das linhas para a , b e c , respectivamente.

Exercícios: Responda os exercícios abaixo:

- ① Crie três vetores com três elementos **inteiros**, sejam eles os objetos x , y , z . Combine os três vetores para se tornar uma matriz 3×3 em que cada coluna representa um vetor. Altere os nomes das linhas para a , b e c , respectivamente.
- ② Crie um vetor v com 15 valores **inteiros** e converta-o em uma matriz M (5×3). Faça isso de duas formas diferentes. Além disso, nomeie as linhas de $l1$ à $l5$ e as colunas de $c1$ à $c3$. Depois, crie a matriz N a partir da matriz M em que N **não** possui nomes nas linhas e colunas.

Data Frames

Um **quadro de dados (data frame)** é a forma mais comum de armazenar dados em R.



The image shows a screenshot of the R Data Editor window. The window title is "R Data Editor". At the top right, there are window control buttons (minimize, maximize, close) and three buttons labeled "Copy", "Paste", and "Quit". The main area of the window displays a data frame with 11 rows and 13 columns. The columns are labeled: row.names, mpg, cyl, disp, hp, drat, wt, qsec, vs, am, gear, carb. The rows contain data for various car models, including Mazda RX4, Datsun 710, and Mercedes-Benz models.

	row.names	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
1	Mazda RX4	21	6	160	110	3.9	2.62	16.46	0	1	4	4
2	Mazda RX4 Wag	21	6	160	110	3.9	2.875	17.02	0	1	4	4
3	Datsun 710	22.8	4	108	93	3.85	2.32	18.61	1	1	4	1
4	Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
5	Hornet Sportabout	18.7	8	360	175	3.15	3.44	17.02	0	0	3	2
6	Valiant	18.1	6	225	105	2.76	3.46	20.22	1	0	3	1
7	Duster 360	14.3	8	360	245	3.21	3.57	15.84	0	0	3	4
8	Merc 240D	24.4	4	146.7	62	3.69	3.19	20	1	0	4	2
9	Merc 230	22.8	4	140.8	95	3.92	3.15	22.9	1	0	4	2
10	Merc 280	19.2	6	167.6	123	3.92	3.44	18.3	1	0	4	4
11	Merc 280C	17.8	6	167.6	123	3.92	3.44	18.9	1	0	4	4

Figura: Data frame em R.

Data Frames

Nota: Por se tratar de uma estrutura bidimensional assim como matrizes, todos os comandos (`dim()`, `names()`, `colnames()`, `ncol()`, `nrow()`, etc) apresentados anteriormente ao tratarmos de matrizes poderão ser considerado à um data frame.

Nota: Por se tratar de uma estrutura bidimensional assim como matrizes, todos os comandos (`dim()`, `names()`, `colnames()`, `ncol()`, `nrow()`, etc) apresentados anteriormente ao tratarmos de matrizes poderão ser considerado à um data frame.

Observação Importante

A forma mais conveniente de organizar um base de dados de modo a facilitar as análises estatísticas é considerar um data frame. No artigo **Tidy Data** publicado em 2014 por **Hadley Wickham** no **Journal of Statistical Software**. No artigo o autor aborda a dificuldade de se organizar uma base de dados e como tornar a limpeza (tratamento dos dados) tão fácil e eficaz quanto possível.

Nota: Por se tratar de uma estrutura bidimensional assim como matrizes, todos os comandos (`dim()`, `names()`, `colnames()`, `ncol()`, `nrow()`, etc) apresentados anteriormente ao tratarmos de matrizes poderão ser considerado à um data frame.

Observação Importante

A forma mais conveniente de organizar um base de dados de modo a facilitar as análises estatísticas é considerar um data frame. No artigo **Tidy Data** publicado em 2014 por **Hadley Wickham** no **Journal of Statistical Software**. No artigo o autor aborda a dificuldade de se organizar uma base de dados e como tornar a limpeza (tratamento dos dados) tão fácil e eficaz quanto possível.

Nota: Hadley Wickham é estatístico e pesquisador chefe da equipe do **RStudio**.

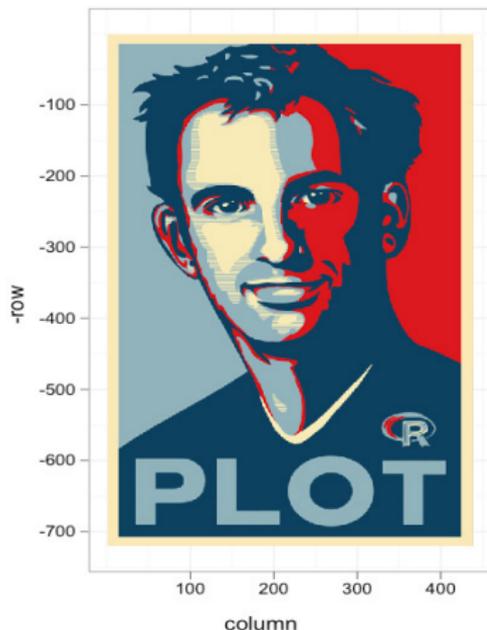


Figura: Hadley Wickham.

Hadley Wickham também é criador da biblioteca **ggplot2** para construção de gráficos em R. A imagem ao lado é produzida pela biblioteca **ggplot2**.

Há o livro **ggplot2: Elegant Graphics for Data Analysis** escrito por ele e complementa a documentação da biblioteca. O livro é do projeto **UseR!** e publicado pela editora Springer.

Data Frames

Para criar um data frame utilizamos a função `data.frame()` que possui como entrada vetores nomeados.

Exemplo: Corra o código que segue:

```
1: df <- data.frame(x = 1:3, y = c("a", "b", "c"))
2: str(df)
#> 'data.frame':    3 obs. of 2 variables:
#> $ x: int 1 2 3
#> $ y: Factor w/ 3 levels "a","b","c": 1 2 3
```

Nota: Observe que por padrão um vetor de strings em um fator. Esse comportamento poderá ser alterado se usamos o argumento `stringAsFactors = FALSE` na função `data.frame()`.

Exemplo: Execute o código abaixo:

```
1: df <- data.frame(x = 1:3, y = c("a", "b", "c"),
2:                 stringsAsFactors = FALSE)
3: str(df)
#> 'data.frame': 3 obs. of 2 variables:
#> $ x: int 1 2 3
#> $ y: chr "a" "b" "c"
```

Data Frames

Como um `data.frame` é uma classe **S3**, seu tipo reflete o vetor subjacente usado para construí-lo: a **lista**.

Exemplo: Utilizando a função `class()` ou explicitamente a função `is.data.frame()`.

```
typeof(df)
#> [1] "list"
class(df)
#> [1] "data.frame"
is.data.frame(df)
#> [1] TRUE
```

Nota: Veja que o tipo de um data frame é uma lista.

Data Frames

Assim como no caso de matrizes, podemos combinar data frames usando as funções `rbind()` e `cbind()`.

Exemplo: Corra o código seguinte:

```
1: df <- data.frame(x = 1:3, y = c("a", "b", "c"))
2: cbind(df, data.frame(z = 3:1))
#>   x y z
#> 1 1 a 3
#> 2 2 b 2
#> 3 3 c 1
3: rbind(df, data.frame(x = 10, y = "z"))
#>   x y
#> 1 1 a
#> 2 2 b
#> 3 3 c
#> 4 10 z
```

Importante

É um erro comum achar que concatenar dois vetores usando a função `cbind()` irá criar um data frame. Na verdade, fazer isto acarretará na criação de uma matriz. Para que um data frame seja criado, considere envolver a função `cbind()` pela função `data.frame()`.

Exemplo: Corra o código que segue:

```
1: good <- data.frame(a = 1:2, b = c("a", "b"),
2:                   stringsAsFactors = FALSE)
3: str(good)
#> 'data.frame': 2 obs. of 2 variables:
#> $ a: int 1 2
#> $ b: chr "a" "b"
```

Nota: É possível que cada coluna de um data frame seja uma lista. Essa característica em um data frame é bastante importante.

Exemplo: Corra o código:

```
1: df <- data.frame(x = 1:3)
2: df$y <- list(1:2, 1:3, 1:4)
3: df
#>   x      y
#> 1 1      1, 2
#> 2 2      1, 2, 3
#> 3 3      1, 2, 3, 4
```

Exercício: Um professor irá criar um data frame para armazenar as notas de 5 alunos de sua turma em que cada aluno poderá ter realizado no máximo 3 avaliações. Crie um data frame com duas colunas sendo a primeira com os nomes dos alunos e a segunda com as notas das avaliações.

Solução:

```
1: avaliacoes <- list(c(8.12,9.81,8.20), c(8.58,9.28,
                        8.22), c(7.13,9.35,9.25),
                        c(7.13,9.35,9.25),c(9.61,7.99,
                        8.04))
2: alunos <- c("Wanusa", "Juliana", "Leticia", "Bianca",
              "Claryce")
3: notas_alunos <- data.frame(nomes = alunos,
                              avaliacoes = NA)
4: notas_alunos$avaliacoes <- avaliacoes
5: # Trataremos melhor sobre as funções da família apply
   # mais adiante.
6: lapply(X = notas_alunos[, 2], FUN = mean)
```

Nota: É possível acessar cada vetor (coluna) de um data frame usando o nome do objeto seguido do símbolo \$ e do nome da coluna. No exemplo logo acima, para acessar a primeira coluna faríamos `df$x` equivalente a fazer `df[,1]`. Analogamente para a segunda coluna deveríamos fazer `df$y` que equivale a fazer `df[,2]`.

Importante

Tentar atribuir diretamente uma lista como argumento da função `data.frame()` ocorrerá um erro.

Exemplo: Corra o código que segue:

```
1: data.frame(x = 1:3, y = list(1:2, 1:3, 1:4))
#> Error: arguments imply differing number of rows:
#> 2, 3, 4
```

Data Frames

O exemplo logo abaixo contorna mostra como contornar tal problema. Para isto, bastará envolver a função `list()`, passada como argumento da função `data.frame()`, com a função `I()`.

O exemplo logo abaixo contorna mostra como contornar tal problema. Para isto, bastará envolver a função `list()`, passada como argumento da função `data.frame()`, com a função `I()`.

Exemplo:

```
1: df1 <- data.frame(x = 1:3, y = I(list(1:2, 1:3, 1:4)))
2: # Acessando a linha dois e coluna de nome "y".
3: # O comando abaixo poderia ser substituído por
4: # df1[2,2].
5: df1[2, "y"]
#> [[1]]
#> [1] 1 2 3
```

Exercício: Com base no exercício anterior referente às notas dos 5 alunos, construa o data frame com os nomes e notas dos alunos atribuindo a lista de notas como argumento da função `data.frame()`.

Solução:

```
1: avaliacoes <- list(c(8.12,9.81,8.20), c(8.58,9.28,
                        8.22), c(7.13,9.35,9.25),
                        c(7.13,9.35,9.25),c(9.61,7.99,
                        8.04))
2: alunos <- c("Wanusa", "Juliana", "Leticia", "Bianca",
              "Claryce")
3: notas_alunos <- data.frame(nomes = alunos,
                              avaliacoes = I(avaliacoes))
4: notas_alunos$avaliacoes <- avaliacoes
5: # Trataremos melhor sobre as funções da família apply
   # mais adiante.
6: lapply(X = notas_alunos[, 2], FUN = mean)
```

De forma análoga, é possível ter uma coluna de um data frame sendo uma matriz, desde que o número de linhas da matriz seja o mesmo da dimensão linha do data frame.

Exemplo: Corra o código abaixo:

```
1: dfm <- data.frame(x = 1:3,  
2:                   y = I(matrix(1:9, nrow = 3)))  
3: dfm[2, "y"]  
#>      [,1] [,2] [,3]  
#> [1,]    2    5    8
```

Exercício: Refaça o exercício das notas dos 5 alunos considerando que a segunda coluna do data frame criado para armazenar os nomes e avaliações é uma matriz de dimensão 5×3 .

Solução:

```
1: avaliacoes <- list(c(8.12,9.81,8.20), c(8.58,9.28,
      8.22), c(7.13,9.35,9.25), c(7.13,
      9.35,9.25), c(9.61,7.99,8.04))
2:
3: avaliacoes <- matrix(data = unlist(avaliacoes),
      nrow = 5, ncol = 3, byrow = TRUE)
4:
5:
6: alunos <- c("Wanusa", "Juliana", "Leticia", "Bianca",
      "Claryce")
7:
8: notas_alunos <- data.frame(nomes = alunos,
      avaliacoes = I(avaliacoes))
```

Data Frames

Exercício: Realize uma estatística descritiva das notas dos alunos utilizando a função `summary()`.

Exercício: Realize uma estatística descritiva das notas dos alunos utilizando a função `summary()`.

Solução:

```
1: apply(X = notas_alunos$aavaliacoes, MARGIN = 1,  
        FUN = summary)
```

Exercício: Realize uma estatística descritiva das notas dos alunos utilizando a função `summary()`.

Solução:

```
1: apply(X = notas_alunos$avaliacoes, MARGIN = 1,  
        FUN = summary)
```

	[,1]	[,2]	[,3]	[,4]	[,5]
Min.	8.120	8.220000	7.130000	7.130000	7.990000
1st Qu.	8.160	8.400000	8.190000	8.190000	8.015000
Median	8.200	8.580000	9.250000	9.250000	8.040000
Mean	8.710	8.693333	8.576667	8.576667	8.546667
3rd Qu.	9.005	8.930000	9.300000	9.300000	8.825000
Max.	9.810	9.280000	9.350000	9.350000	9.610000

Acessando subconjunto - Vetor

Os operadores de subconjunto da linguagem R são poderosos e rápidos. O domínio de tais operadores permite que o programador expresse sucintamente operações complexas de uma maneira que outras linguagens podem não fazer.

Considere o vetor atômico `x <- c(2.1, 4.2, 3.3, 5.4)`. Os exemplos seguintes deverão ser replicados e considerarão esse vetor.

Acessando subconjunto - Vetor

Exemplo: Inteiros positivos retornam os elementos nas respectivas posições e números reais serão truncados.

```
1: x[c(3, 1)]
#> [1] 3.3 2.1
2: x[order(x)] # Faça ?order para detalhes.
#> [1] 2.1 3.3 4.2 5.4
3:
# Duplicated indices yield duplicated values
4: x[c(1, 1)]
#> [1] 2.1 2.1
# Real numbers are silently truncated to integers
5: x[c(2.1, 2.9)]
#> [1] 4.2 4.2
```

Acessando subconjunto - Vetor

Exemplo: Números inteiros negativos omite os elementos omite a respectiva posição e valores reais negativos serão truncados.

```
1: x[-c(3, 1)]
```

```
#> [1] 4.2 5.4
```

```
2:
```

```
3: x[c(-1, 2)]
```

```
#> Error: only 0's may be mixed with negative subscripts
```

Acessando subconjunto - Vetor

Exemplo: Vetores lógicos selecionam elementos onde o valor lógico corresponde à verdade. Assim, uma expressão lógica também poderá ser utilizada para selecionar os elementos de um vetor atômico.

```
1: x[c(TRUE, TRUE, FALSE, FALSE)]
```

```
#> [1] 2.1 4.2
```

```
2: x[x > 3]
```

```
#> [1] 4.2 3.3 5.4
```

Acessando subconjunto - Vetor

Importante: Se o vetor lógico tiver um comprimento menor que o vetor ao qual queremos extrair um subconjunto, o vetor lógico será reciclado para ter o mesmo comprimento.

```
1: x[c(TRUE, FALSE)]
```

```
#> [1] 2.1 3.3
```

```
2: # Equivalent to
```

```
3: x[c(TRUE, FALSE, TRUE, FALSE)]
```

```
#> [1] 2.1 3.3
```

Acessando subconjunto - Vetor

Exemplo: Um valor faltante no vetor de índices irá produzir uma informação faltante na respectiva posição do vetor saída.

```
x[c(TRUE, TRUE, NA, FALSE)]  
#> [1] 2.1 4.2 NA
```

Acessando subconjunto - Vetor

Exemplo: Nada retorna o vetor original. Isto não é útil para vetores mas é bastante útil para arrays e data frames.

```
1: x[]  
#> [1] 2.1 4.2 3.3 5.4
```

Exemplo: Zero retornará um vetor de comprimento zero.

```
1: a <- x[0]  
2: a[1] = 2; a[2] = 7  
3: a  
#> [1] 2 7
```

Acessando subconjunto - Vetor

Exemplo: Vetor de caracteres retornará os elementos de nomes correspondentes.

```
1: y <- setNames(x, letters[1:4])
#>  a    b    c    d
#> 2.1 4.2 3.3 5.4
2: y[c("d", "c", "a")]
#>  d    c    a
#> 5.4 3.3 2.1
```

Nota: Faça `?letters` e leia o manual.

Exercício: De que outra forma poderíamos criar e nomear o objeto `y` sem utilizar a função `setNames()`?

Acessando subconjunto - Vetor

Exemplo: A não correspondência exata de um nome não provocará erro. Nesse caso, a linguagem R retornará NA.

```
1: z <- c(abc = 1, def = 2)
2: z[c("a", "d")]
#> <NA> <NA>
#>   NA   NA
```

Acessando subconjunto - Matrizes e Array

Exemplo: Corra o código logo abaixo:

```
1: a <- matrix(1:9, nrow = 3)
2: colnames(a) <- c("A", "B", "C"); a[1:2, ]
#>      A B C
#> [1,] 1 4 7
#> [2,] 2 5 8
3: a[c(T, F, T), c("B", "A")]
#>      B A
#> [1,] 4 1
#> [2,] 6 3
4: a[0, -2]
#> A C
```

Acessando subconjunto - Matrizes e Array

Exemplo: Também é possível passar uma matriz como vetor de índices. Procure nas documentações da linguagem entender o funcionamento das funções `outer()` e `paste()`.

```
1: vals <- outer(1:5, 1:5, FUN = "paste", sep = ",")
#>      [,1] [,2] [,3] [,4] [,5]
#> [1,] "1,1" "1,2" "1,3" "1,4" "1,5"
#> [2,] "2,1" "2,2" "2,3" "2,4" "2,5"
#> [3,] "3,1" "3,2" "3,3" "3,4" "3,5"
#> [4,] "4,1" "4,2" "4,3" "4,4" "4,5"
#> [5,] "5,1" "5,2" "5,3" "5,4" "5,5"
2:
3: vals[c(4, 15)]
#> [1] "4,1" "5,3"
```

Acessando subconjunto - Matrizes e Array

```
1: vals <- outer(1:5, 1:5, FUN = "paste", sep = ",")
2: select <- matrix(ncol = 2, byrow = TRUE,
3:                 c(1, 1, 3, 1, 2, 4))
4: vals[select]
#> [1] "1,1" "3,1" "2,4"
```

Notas:

- 1 O argumento `byrow = TRUE` faz com que o preenchimento da matriz seja realizado por linha. Por padrão `byrow = FALSE`.
- 2 Obviamente a matriz com `select` deverá possuir duas colunas, uma vez que a matriz possui uma estrutura bidimensional. Os elementos na primeira coluna refere-se ao elemento linha e o elemento da coluna refere-se ao elemento coluna da matriz ao qual será obtido o subconjunto.

Acessando subconjunto - Data Frame

Exemplo: Data frames possuem as mesmas características de matrizes para acessar seus subconjuntos. Corra o código que segue:

```
1: df <- data.frame(x = 1:3, y = 3:1, z = letters[1:3])
2: df[df$x == 2, ]
#>   x y z
#> 2 2 2 b
3: df[c(1, 3), ]
#>   x y z
#> 1 1 3 a
#> 3 3 1 c
4: # There are two ways to select columns from a data
5: # frame
```

Acessando subconjunto - Data Frame

```
5: # Like a list:
6: df[c("x", "z")]
#>   x z
#> 1 1 a
#> 2 2 b
#> 3 3 c
7: # Like a matrix
8: df[, c("x", "z")]
#>   x z
#> 1 1 a
#> 2 2 b
#> 3 3 c
```

Nota Importante: Passar um vetor como índices de um objeto da classe data frame fará com que a linguagem R selecione as colunas correspondentes ao índices do vetor.

Acessando subconjunto - Data Frame

Exercício: Inicialmente, crie a lista abaixo:

Exercício: Inicialmente, crie a lista abaixo:

```
notas <- list(aluno_1 = c(7.1, 3.2, NA),
              aluno_2 = c(2.7, 8.8, 10.0),
              aluno_3 = c(0.0, NA, NA),
              aluno_4 = c(7.7, 8.4, 6.3),
              aluno_5 = c(3.6, 6.6, 8.1),
              aluno_6 = c(NA, NA, NA),
              aluno_7 = c(7.4, 7.1, 7.3),
              aluno_8 = c(10.0, NA, 7.0),
              aluno_9 = c(1.6, 3.2, 5.3),
              aluno_10 = c(8.8, 9.2, 8.0))
```

Acessando subconjunto - Data Frame

- a) Crie o vetor `status` contendo o status dos dez alunos.
Considere: **A** (aprovado), **REP** (reprovado), **F** (final). **Dica:**
Considere as regras de nossa instituição.

Acessando subconjunto - Data Frame

- a) Crie o vetor `status` contendo o status dos dez alunos. Considere: **A** (aprovado), **REP** (reprovado), **F** (final). **Dica:** Considere as regras de nossa instituição.
- b) Crie o vetor `alunos` com os nomes dos alunos. Considere os nome fictício **Aluno_1** para o primeiro aluno e de forma análoga para os demais alunos. Crie esse vetor utilizando a função `paste()`.

Acessando subconjunto - Data Frame

- a) Crie o vetor `status` contendo o status dos dez alunos. Considere: **A** (aprovado), **REP** (reprovado), **F** (final). **Dica:** Considere as regras de nossa instituição.
- b) Crie o vetor `alunos` com os nomes dos alunos. Considere os nome fictício **Aluno_1** para o primeiro aluno e de forma análoga para os demais alunos. Crie esse vetor utilizando a função `paste()`.
- c) Construa um data frame de nome **histórico** com as variáveis **nomes**, **notas** e **status**.

Acessando subconjunto - Data Frame

- d) Com base no data frame de nome **historico**, construa um outro data frame com o nome **aprovados** com os alunos aprovados. De forma análoga para os demais status.

Acessando subconjunto - Data Frame

- d) Com base no data frame de nome **historico**, construa um outro data frame com o nome **aprovados** com os alunos aprovados. De forma análoga para os demais status.
- e) Suponha que o professor está interessado em saber quais alunos foram ou tem alguma chance de assumir o status de aprovado. Construa o data frame de nome **bons_alunos** com estes alunos.

Acessando subconjunto - Data Frame

- d) Com base no data frame de nome **historico**, construa um outro data frame com o nome **aprovados** com os alunos aprovados. De forma análoga para os demais status.
- e) Suponha que o professor está interessado em saber quais alunos foram ou tem alguma chance de assumir o status de aprovado. Construa o data frame de nome **bons_alunos** com estes alunos.
- f) Modifique os nomes das linhas do data frame **historico** colocando **id_1** na primeira linha e respectivamente no mesmo padrão para as demais linhas.

Acessando subconjunto - Data Frame

- g) Obtenha por meio do data frame de nome **historico** um novo data frame (**historico_na**) com os alunos que deixaram ao menos uma prova para repor.

Acessando subconjunto - Data Frame

- g) Obtenha por meio do data frame de nome **historico** um novo data frame (**historico_na**) com os alunos que deixaram ao menos uma prova para repor.

- h) Apenas para os alunos que fizeram as três avaliações, obtenha uma média aritmética das avaliações. Acrescente a variável de nome **media** no data frame **historico**.

Acessando subconjunto - Data Frame

Exercício: Utilizando o conjunto de dados de nome **state.x77** responda os itens abaixo:

a) Construa o data frame **dados** a partir de **state.x77**.

Acessando subconjunto - Data Frame

Exercício: Utilizando o conjunto de dados de nome **state.x77** responda os itens abaixo:

- a) Construa o data frame **dados** a partir de **state.x77**.
- b) Obtenha um data frame de nome **dados_1** com as observações de **dados** que possua população maior que 4246, isto é, com os estados estadunidenses que possua uma população maior que 4246 (quatro milhões duzentos e quarenta e seis mil).

Acessando subconjunto - Data Frame

Exercício: Utilizando o conjunto de dados de nome **state.x77** responda os itens abaixo:

- a) Construa o data frame **dados** a partir de **state.x77**.
- b) Obtenha um data frame de nome **dados_1** com as observações de **dados** que possua população maior que 4246, isto é, com os estados estadunidenses que possua uma população maior que 4246 (quatro milhões duzentos e quarenta e seis mil).
- c) Obtenha o data frame **dados_2** com as observações população maior que 4246 e menores que 8 milhões, isto é, menor que 8000.

Acessando subconjunto - Data Frame

Exercício: Utilizando o conjunto de dados de nome **state.x77** responda os itens abaixo:

- Construa o data frame **dados** a partir de **state.x77**.
- Obtenha um data frame de nome **dados_1** com as observações de **dados** que possua população maior que 4246, isto é, com os estados estadunidenses que possua uma população maior que 4246 (quatro milhões duzentos e quarenta e seis mil).
- Obtenha o data frame **dados_2** com as observações população maior que 4246 e menores que 8 milhões, isto é, menor que 8000.
- Obtenha o vetor **estados_c** com os nomes dos estados que obedecem os critérios do item c.

Acessando subconjunto - Data Frame

- e) Construa o data frame **dados_3** com os estados estadunidenses com população maior que 1.5 vezes a média dos 50 estados considerados. Obtenha o vetor com o nome dos estados que obedecem essa regra.

Acessando subconjunto - Data Frame

- e) Construa o data frame **dados_3** com os estados estadunidenses com população maior que 1.5 vezes a média dos 50 estados considerados. Obtenha o vetor com o nome dos estados que obedecem essa regra.

- f) Construa o data frame **dados_4** com os estados estadunidenses com população maior que duas vezes a mediana dos 50 estados e que tenha uma população com expectativa de vida maior que 71.84 anos.

Acessando subconjunto - Data Frame

- e) Construa o data frame **dados_3** com os estados estadunidenses com população maior que 1.5 vezes a média dos 50 estados considerados. Obtenha o vetor com o nome dos estados que obedecem essa regra.
- f) Construa o data frame **dados_4** com os estados estadunidenses com população maior que duas vezes a mediana dos 50 estados e que tenha uma população com expectativa de vida maior que 71.84 anos.
- g) Obtenha o data frame **dados_5** com os estados estadunidenses com renda maior que a média nacional, expectativa de vida maior que 72 anos.

Acessando subconjunto - Data Frame

- e) Construa o data frame **dados_3** com os estados estadunidenses com população maior que 1.5 vezes a média dos 50 estados considerados. Obtenha o vetor com o nome dos estados que obedecem essa regra.
- f) Construa o data frame **dados_4** com os estados estadunidenses com população maior que duas vezes a mediana dos 50 estados e que tenha uma população com expectativa de vida maior que 71.84 anos.
- g) Obtenha o data frame **dados_5** com os estados estadunidenses com renda maior que a média nacional, expectativa de vida maior que 72 anos.
- h) Adicione ao data frame **dados** duas linhas com a média de todas as variáveis e variâncias, respectivamente.

Acessando subconjunto - Lista

Vamos fazer uso de uma metáfora que pode nos ajudar. Em listas é muito comum o uso dos operadores \$ e []. Entender da seguinte forma poderá ajudar:

Acessando subconjunto - Lista

Vamos fazer uso de uma metáfora que pode nos ajudar. Em listas é muito comum o uso dos operadores \$ e []. Entender da seguinte forma poderá ajudar:

- ① []: é usado para extrair itens únicos.

Acessando subconjunto - Lista

Vamos fazer uso de uma metáfora que pode nos ajudar. Em listas é muito comum o uso dos operadores \$ e []. Entender da seguinte forma poderá ajudar:

- 1 []: é usado para extrair itens únicos.
- 2 \$: é uma sintaxe útil à sintaxe acima quando os elementos de uma lista estão nomeados.

Metáfora

Se a lista x é um trem que transporta objetos, então $x[[5]]$ é o objeto no carro 5; e $x[4:6]$ é um trem de carros 4 à 6.

Acessando subconjunto - Lista

Metáfora

Se a lista x é um trem que transporta objetos, então $x[[5]]$ é o objeto no carro 5; e $x[4:6]$ é um trem de carros 4 à 6.

```
x <- list(1:3, "a", 4:6)
```

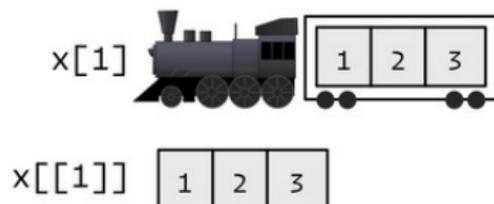


Acessando subconjunto - Lista

Continuando com a metáfora, ao extrair um subconjunto de uma lista temos duas opções. Você poderá criar um trem menor ou extrair o conteúdo de um dos vagões. Essa é de forma metafórica a diferença entre os operadores `$` e `[[`, respectivamente.

Acessando subconjunto - Lista

Continuando com a metáfora, ao extrair um subconjunto de uma lista temos duas opções. Você poderá criar um trem menor ou extrair o conteúdo de um dos vagões. Essa é de forma metafórica a diferença entre os operadores `$` e `[[`, respectivamente. A imagem abaixo mostra isso de uma forma mais lúdica.

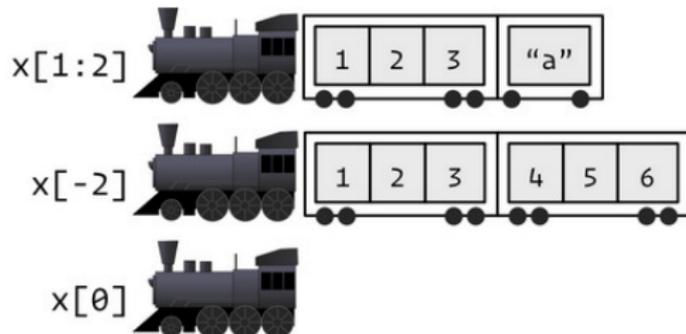


Acessando subconjunto - Lista

Entendida a metáfora, as figuras abaixo são autoexplicativas:

Acessando subconjunto - Lista

Entendida a metáfora, as figuras abaixo são autoexplicativas:



Acessando subconjunto - Lista

Exemplo: O operador `[[` é similar ao operador `[`, exceto pelo fato que `[[` retornará apenas um único valor. O operador `$` é uma abreviação do operador `[[`.

```
1: a <- list(a = 1, b = 2)
```

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

```
#> [1] 1
```

```
3: a[["a"]]
```

```
#> [1] 1
```

Acessando subconjunto - Lista

```
4: # If you do supply a vector it indexes recursively
5: b <- list(a = list(b = list(c = list(d = 1))))
6: b[[c("a", "b", "c", "d")]]
#> [1] 1
7: # Same as
8: b[["a"]][["b"]][["c"]][["d"]]
#> [1] 1
9: # Same as
10: a$b$c$d
#> [1] 1
```