

Análise de Dados Ambientais com R

Jônatan Tatsch

2018-04-23

Contents

Apresentação

Este material é uma composição das notas de aula da disciplina **Análise de Dados Ambientais com** do curso de Graduação em METEOROLOGIA oferecido no Departamento de Física da Universidade Federal de Santa Maria (UFSM).

O livro é designado para quem não tem experiência em programação, ou qualquer um com interesse em aprender o para manipular dados ambientais. O objetivo é prover uma material para ensinar os conceitos básicos de programação necessários para o processamento, a visualização e a análise de dados ambientais com o sistema computacional . Estes procedimentos são potencializados com o uso do software RStudio, uma interface de desenvolvimento integrado (IDE) para o .

Neste livro o leitor aprenderá a sintaxe básica da linguagem (?), a importação e exportação de dados, a criação de gráficos, funções, a padronização e organização de conjunto de dados ambientais; e finalmente, a confecção de relatórios dinâmicos e reproduzíveis.

O material do livro inclui o uso de dados ambientais de diferentes áreas (meteorologia, climatologia, hidrologia, sensoriamento remoto) em exemplos práticos e em exercícios, para estimular a prática da programação.

O texto é intercalado com trechos de códigos que podem ser reproduzidos e os resultados visualizados no computador do leitor.

Após a introdução ao apresenta-se as capacidades específicas do para manipulação de dados. Baseado na experiência do autor são empregados os pacotes mais adequados para cada finalidade, como **dplyr** e **tidyr** para o processamento de dados e o **ggplot2** para visualização de dados.

A intenção do livro é que após a leitura, o leitor tenha o conhecimento suficiente para desenvolver códigos que automatizem tarefas repetitivas, assim reduzindo o tempo gasto na etapa de preparação de dados. Esta programação mais efetiva permitirá focar mais na análise de dados e na comunicação dos resultados, seja ela na forma de documentos acadêmicos, ou relatórios técnicos em empresas públicas e privadas.

O texto está em formato html para tirar o melhor proveito de recursos de multimídia, da capacidade de busca de texto e links para websites.

O texto é organizado em 5 capítulos:

- ?? Introdução
- ?? Instalação do e Rstudio
- ?? Interface do Usuário
- ?? Rstudio
- ?? Operações Básicas
- ?? Tipos de dados

Chapter 1

Introdução

Breve intro.

1.1 Análise de dados meteorológicos

Processo pelo qual adquire-se conhecimento, compreensão e percepção dos fenômenos meteorológicos a partir de observações (dados) qualitativas e quantitativas.

1.2 Ciência de dados

1.3 Etapas para abordagem de um problema

1. **Questão científica/problema**
2. **Obtenção de dados:** coleta/medida do(as) estado/condições da atmosfera
 - Instrumentos e sensores
3. **Processamento de dados:** *download* —> limpeza —> formatação —> transformação —> controle de qualidade
 - ferramenta/software
 - conhecimento em programação
4. **Análise de dados**
 - ferramenta/software
 - conhecimento em programação
5. **Solução para o problema**
 - Proposta de um modelo
 - estatístico, empírico, ou fisicamente baseado
 - conhecimento em programação
6. **Apresentação/divulgação/publicação**

1.4 Programação computacional

1.5

- R é o termo usado para se referir a linguagem de programação e ao software que interpreta os scripts escritos usando esta linguagem.
- Comunidade fantástica
- Contribuidores (R-core Team)
- milhares de pessoas usam o R diariamente e ajudam outras pessoas
- **Software Livre** (GPL), Código aberto e multiplataforma
- Ambiente para Análise de dados interativa

1.6 Por que o R?

- R não é uma GUI (Interface gráfica do usuário) e isso é bom
 - há uma natural resistência e dificuldade ao uso de códigos e scripts
 - scripts favorecem a **automatização** e **reprodutibilidade**
 - força você a ter um conhecimento mais aprofundado do que está fazendo
- Reprodutibilidade
 - qualquer pessoa (inclusive você mesmo no futuro) pode obter os mesmos resultados do mesmo conjunto de dados
 - R é integrado com outras ferramentas de que permitem atualizar seus resultados, figuras e análises automaticamente
- Relatório dinâmicos e interativos
- Acesso ao estado da arte da ciência de dados (*Big Data*, *Data Mining*, *Machine Learning*)
- é um software livre, de código fonte aberto e funciona em diversos sistemas operacionais (Linux, Windows e MacOS).
- Interface com Fortran, C, C++, Python
- Visualização de dados
- R produz gráficos de alta qualidade
- R trabalha com dados de todas formas e tamanhos
- Extensões para Manipulação de dados

1.7 Pacotes da comunidade do R

Evolução do nº de pacotes disponíveis no CRAN

1.8 Por que um meteorologista usaria o R?

A meteorologia é 4D:

```
meteorologia <- function(x, y, z, t){  
  ...muita coisa para caber em um slide...  
}
```

Logo, requer ferramentas específicas para:

- manipulação de dados espaciais
- análise de séries temporais
- importação e ferramentas de SIG
- leitura de dados em formatos específicos (netcdf, binários, grib2, ...)

1.9 R não é perfeito!

- Muitos códigos em R são escritos para resolver um problema;
 - foco nos resultados e não no processo
 - usuários não são programadores
 - códigos deselegantes, lentos e difíceis de entender
- Como o nosso idioma, há muitas exceções para serem lembradas
- R não é muito rápido e códigos mal escritos serão lentos
- São apenas ~20 anos de evolução
- Há muito o que melhorar

1.10 Para saber mais sobre o R

Documentação oficial - Manuais do R traduzidos

Lista de Livros relacionados ao R

- Livros gratuitos (em inglês)

Fóruns:

- lista Brasileira de discussão do programa R: **R-br**
- stackoverflow

Chapter 2

Instalação do R e RStudio

A interação do usuário com o `R` é por meio da linha de comando. Essa interação pode ser facilitada com o uso do software RStudio *Desktop*.

A seguir descreve-se como:

- instalar o `R` no Windows e no Linux Ubuntu
- manter o `R` sempre atualizado no Linux Ubuntu
- configurar um diretório para armazenar os pacotes do `R` instalados
- instalar o Rstudio *Desktop*

Neste livro, o maior foco na instalação do `R` é dada para o SO Linux Ubuntu, pelo fato de assim como o `R`, ser um software livre e de código aberto. Como o Linux Ubuntu é baseado no Debian o procedimento de instalação também se estende a essa distribuição Linux e as versões derivadas do Ubuntu que são oficialmente reconhecidas.

A instalação no SO Windows é igual a instalação de qualquer outro *software* e pode ser facilmente encontrada na internet. Por esta razão, somente indicou-se o caminho de instalação, sem as instruções detalhadas de instalação para este SO.



Ao instalar `R` e RStudio recomenda-se optar por instalar na língua inglesa. Assim quando surgir uma mensagem de erro durante o uso do software, basta usá-la numa pesquisa na internet para solucionar o problema. As chances de resolver o problema serão muito maiores se sua pesquisa for realizada em inglês.

2.1 Instalando o

O `R` pode ser instalado a partir dos binários pré-compilados ou do código fonte. Aqui, descreve-se a instalação do `R` a partir dos arquivos binários.

2.1.1 Windows

O binário executável do `R` para o Windows está disponível na **Rede Abrangente de Arquivos do** (CRAN) e pode ser baixado aqui.

Abra o executável e siga instruções de instalação do `R` mantendo todas as opções padrões.

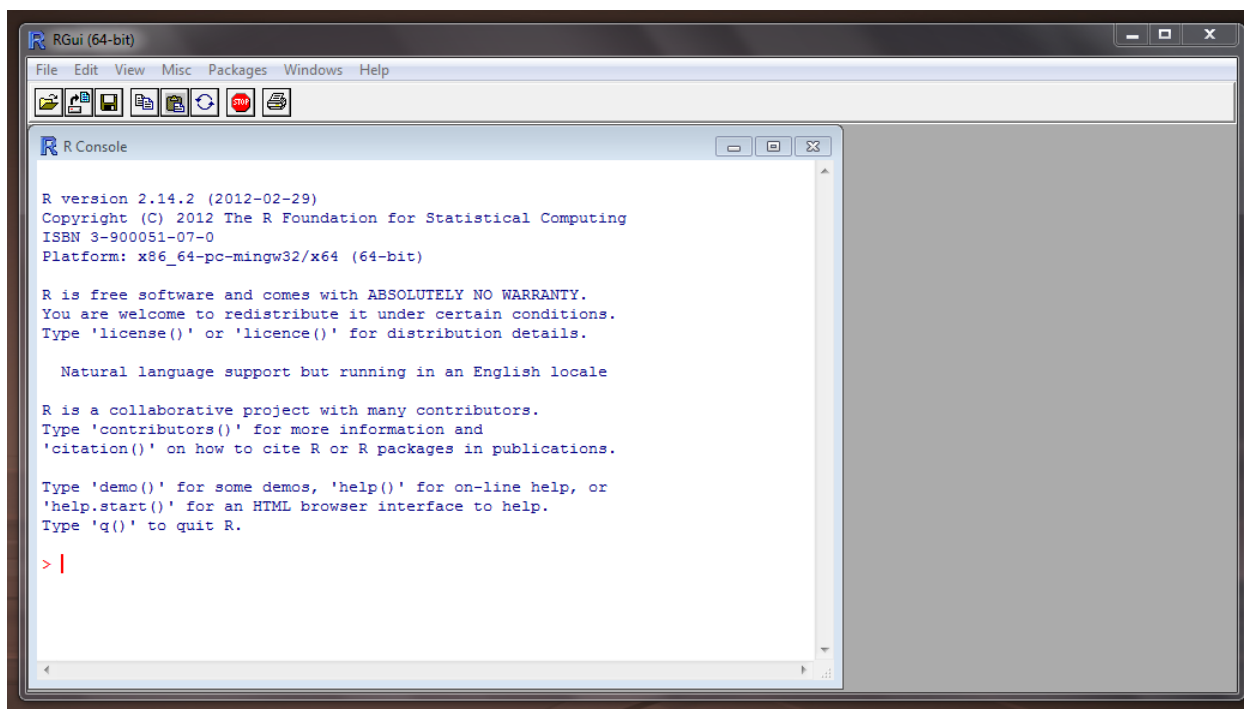


Figure 2.1: Interface gráfica do usuário no R para Windows.

No Windows a instalação do inclui uma Interface Gráfica do Usuário (GUI) acessível pelo executável `RGui.exe` (Figura ??). Um atalho para esse executável é gerado por *default* na área de trabalho com o símbolo do .

2.1.1.1 Atualização do no Windows

Novas versões do R são disponibilizadas em geral com frequência de 5 vezes por ano. Recomenda-se manter o R atualizado, pois as novas versões incluem aperfeiçoamentos e a correção de *bugs*.

As novas versões do vem com os pacotes padrões do R. Os demais pacotes instalados pelo usuário na versão anterior precisam ser reinstalados na nova versão do .

Para atualizar o no Windows, ao invés de baixar o executável a cada nova versão e repetir o processo da seção anterior, você pode utilizar o pacote **installr**. A instalação de pacotes no será vista na seção ??.

2.1.2 Linux

2.1.2.1 Ubuntu

Há várias formas de instalar o no Ubuntu, mas geralmente a versão compilada no repositório *default* do Ubuntu não é a última. Se isso não for problema para você então basta executar:

```
sudo apt-get install r-base
```

Entretanto, os pacotes do recém lançados são compilados para última versão do . Então você pode ter restrições ao uso de pacotes novos, os quais geralmente incluem o estado da arte de análise de dados. Por esta razão, abaixo mostra-se como instalar o de forma que seja atualizado automaticamente pelo sistema.

2.1.2.2 R sempre atualizado

Se você quer trabalhar sempre com a última versão estável do `R`, é possível configurar o Linux Ubuntu para atualizar automaticamente o `R`. O procedimento de instalação requer senha de superusuário do sistema ou de privilégios `sudo`. Caso não tenha, consulte o administrador do sistema.

Ao utilizar distribuições Linux Ubuntu é importante optar por versões estáveis¹. As versões de Suporte de longo prazo (LTS) mais recentes são:

- 14.04 (abril de 2014, *codename* `trusty`)
- 16.04 (abril de 2016, *codename* `xenial`)

A versão mais atual é a `R` version 3.4.4 (2018-03-15). Para que ele seja atualizado automaticamente no Ubuntu você precisa adicionar o endereço do repositório do `R` mais próximo de sua região à lista de repositórios do Linux. No exemplo deste livro, o repositório mais próximo é o da UFPR (<http://cran-r.c3sl.ufpr.br/>).

2.1.2.2.1 Incluindo repositório do na Lista de repositórios do Ubuntu

A lista de repositórios do sistema é armazenada no arquivo `/etc/apt/sources.list`. Mas primeiro, você precisa descobrir ou verificar o nome da versão do sistema operacional. Para isso, você pode utilizar o seguinte comando²:

```
$ lsb_release --codename | cut -f2
trusty
```

Precisamos incluir no arquivo `sources.list` o espelho do repositório do `R` mais próximo. Veja a lista de espelhos de repositórios do `R` aqui. Assim o gerenciador de pacotes `apt`³ fará a atualização do `R` quando uma nova versão estiver disponível. Ou seja, você estará utilizando sempre versão mais atual do `R`.

O endereço do repositório da UFPR será inserido na última linha do arquivo `sources.list` usando alguns comandos linux. Essa tarefa requer privilégios de superusuário. Vamos trocar do seu usuário para o superusuário.

```
$ sudo su
```

Vamos definir no terminal uma variável com o endereço do repositório (da UFPR nesse caso) e o nome de versão do Ubuntu.

```
# repos="deb http://cran-r.c3sl.ufpr.br/bin/linux/ubuntu `lsb_release --codename | cut -f2`/"
```

Note que a variável `repos` é uma sequência de caracteres com as seguintes informações:

```
deb `linkRepositorioSelecionado`/bin/linux/ubuntu `versaoUbuntu`/
```

O valor da variável `repos` é mostrado pelo comando: `echo $repos`. Certifique-se de que a última palavra corresponde ao nome da sua versão Ubuntu.

Para acrescentar essa informação no final do arquivo `sources.list` digite no terminal linux:

```
# echo $repos >> /etc/apt/sources.list
```

Feito isso, você pode retornar a sessão de usuário comum, usando o comando abaixo:

```
# exit
```

¹Clique aqui para saber mais sobre as versões do Ubuntu.

²Se o comando `lsb_release` não funcionar você precisa instalar o pacote `lsb-release` no sistema. Para isso, digite no terminal Linux `sudo apt-get install lsb-release`.

³o gerenciador de pacotes `apt` é usado para instalação, atualização e remoção de pacotes em distribuições Debian GNU/Linux.

2.1.2.2.2 APT protegido

Os arquivos binários do para Ubuntu na CRAN são assinados com uma chave pública⁴ Para adicionar essa chave ao seu sistema digite os seguintes comandos:

```
$ gpg --keyserver hkp://keyserver.ubuntu.com:80 --recv-keys E084DAB9
```

e então use essa informação como entrada no `apt-key` com

```
$ gpg -a --export E084DAB9 | sudo apt-key add -
```

Se aparecer a mensagem de que a chave pública foi importada, então não há necessidade de executar os comandos abaixo. Mas caso seja impresso alguma mensagem de erro, outra alternativa pode ser usada para obter a chave, via os comandos:

```
$ gpg --keyserver keyserver.ubuntu.com --recv-key E084DAB9
```

```
$ gpg -a --export E084DAB9 | sudo apt-key add -
```

2.1.2.2.3 Atualização da lista de repositórios do Ubuntu e instalação do

Após fazer as configurações da lista de repositórios e adicionar a chave é necessário fazer a atualização dessa lista (requer poderes de super usuário):

```
$ sudo apt-get update
```

Agora, pode instalar o binário do R:

```
$ sudo apt-get install r-base
```

2.1.2.2.4 Testando o

Para iniciar o no Ubuntu, digite R no cursor do terminal:

```
$ R
```

A partir desse momento já começamos uma sessão no . Vamos gerar uma sequência numérica de 1 a 10 e plotá-la.

```
> 1:10
[1] 1 2 3 4 5 6 7 8 9 10
> plot(1:10)
```

Você pode sair do , sem salvar os dados da seção, com o código a seguir:

```
> q(save = "no")
```

2.1.2.3 Diretório para instalação de pacotes

Uma boa prática é definir um diretório para armazenamento dos pacotes utilizados. Isso lhe dá mais controle sobre os pacotes do instalados no sistema. Um local sugerido é o `/home/usuario/.R/libs`. O seu `home` ou `pasta pessoal` pode ser obtido com o comando `echo $HOME`. Para criar o diretório você pode digitar o comando abaixo:

```
$ mkdir -p `echo $HOME`/.R/libs/
```

Para informar ao onde procurar os pacotes instalados, você precisa criar um arquivo chamado `.Renviron`, no diretório `$HOME`, contendo a expressão `R_LIBS=/home/usuario/.R/libs/`. Você pode fazer isso em um terminal com os comandos:

⁴Chave pública de autenticação é um meio alternativo de se logar em um servidor ao invés de digitar uma senha. É uma forma mais segura e flexível, mas mais difícil de ser configurada. Esse meio alternativo de fazer login é importante se o computador está visível na internet. Para saber mais veja aqui.

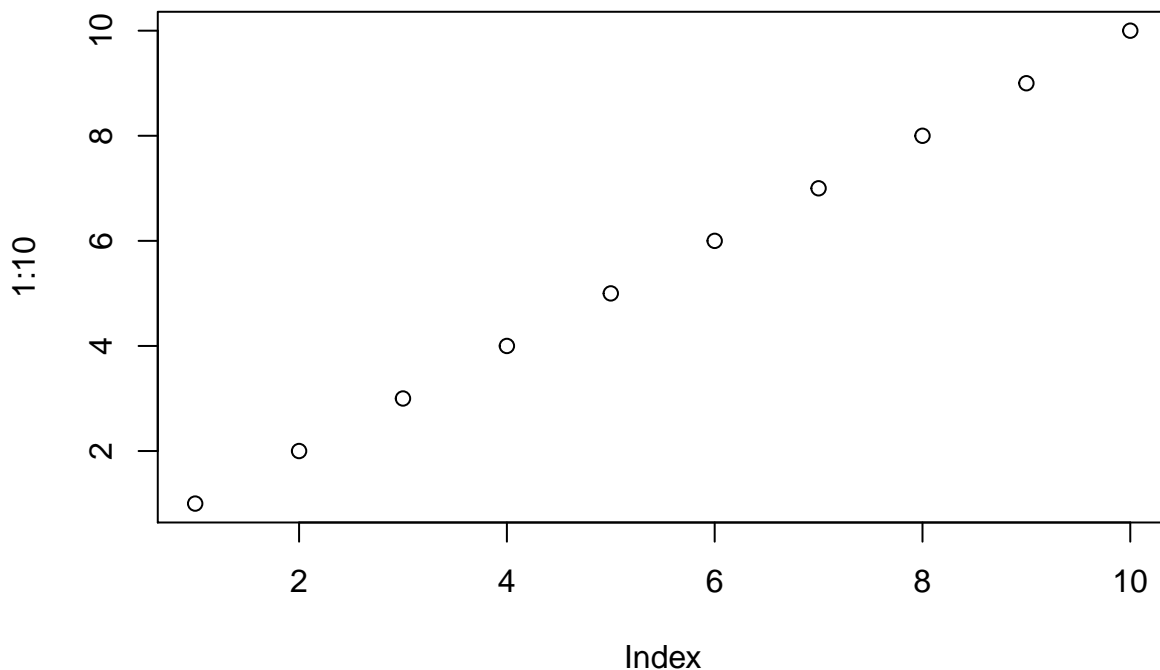


Figure 2.2: Gráfico da sequência de 10 números.

```
$ R_LIBS=`echo $HOME/.R/libs/`
$ echo $R_LIBS >> `echo $HOME/.Renviro`
```

Esse caminho fica então visível ao `ls`, o que pode ser verificado executando a função `.libPaths()` na linha de comando do `R`.

Abra o `R`:

```
$ R
```

e ao digitar:

```
> .libPaths()
[1] "/home/pqgfapergs1/.R/libs"      "/usr/local/lib/R/site-library"
[3] "/usr/lib/R/site-library"        "/usr/lib/R/library"
```

o seu diretório `/home/usuario/.R/libs`⁵ deve aparecer em primeiro lugar. Indicando que este local tem prioridade para instalação dos pacotes. Caso o diretório deixe de existir os seguintes diretórios serão usados.

2.2 Pacotes do R

2.2.1 Da internet

2.2.1.1 CRAN

A forma mais fácil de instalar uma pacote do R é através da função `install.packages("nome_do_pacote")`.

Por *default* o pacote informado é instalado a partir da (CRAN)

⁵Diretórios precedidos por “.” no Linux são diretórios ocultos. O diretório `/home/usuario/.R` é um diretório oculto, para visualizá-lo no Ubuntu, na interface gráfica do sistema, acesse *View > Show Hidden Files* (ou *Visualizar > Mostrar arquivos ocultos*). No terminal utilize `ls -a` para listar os arquivos ocultos.

Por exemplo, para instalar o pacote **devtools**:

```
install.packages("devtools")
```

A função automaticamente resolverá as dependências do pacote, de forma que qualquer pacote dependente também será instalado.

Para ter acesso as funções disponibilizadas com o pacote você precisa carregar o pacote:

```
library(devtools)
```

Para desinstalar um pacote você pode usar a função `remove.packages("nome_do_pacote")`.

2.2.1.2 GitHub e R-forge

Nem todos pacotes são disponíveis na CRAN. Muitos desenvolvedores disponibilizam seus pacotes em plataformas como o GitHub e R-forge. As vezes um pacote pode estar em ambos CRAN e GitHub (ou R-forge), mas a última versão - a de desenvolvimento - é somente disponibilizada no GitHub (ou R-forge).

Para instalar um pacote de um repositório do GitHub usa-se a função `install_github()` do pacote **devtools**. Portanto, o pacote **devtools** precisa ser instalado primeiro.

Antes de instalar o pacote **devtools**, usuários Windows precisam instalar o programa Rtools.

A função para instalar um pacote do GitHub requer como argumento o nome do usuário e do repositório. Por exemplo, para instalar o pacote **inmetr** do repositório mantido pelo autor deste livro, usa-se:

```
# install.packages("devtools")
# carrega o pacote devtools
library(devtools)
# instala o pacote inmetr do repositório
# https://github.com/lhmet/inmetr
install_github("lhmet/inmetr")
```

Para um repositório do R-forge, por exemplo o repositório do pacote raster, usa-se:

```
install.packages("raster", repos = "http://R-Forge.R-project.org")
```

2.2.1.3 Arquivo fonte local

Códigos fonte de pacotes do R são armazenados como arquivos com a extensão **.tar.gz**. Binários compilados são armazenados com a extensão **.zip**. Exemplo de arquivos como estes podem ser baixados manualmente da CRAN (veja a seção Downloads em <https://cran.r-project.org/web/packages/ggplot2/index.html>), GitHub ou R-forge.

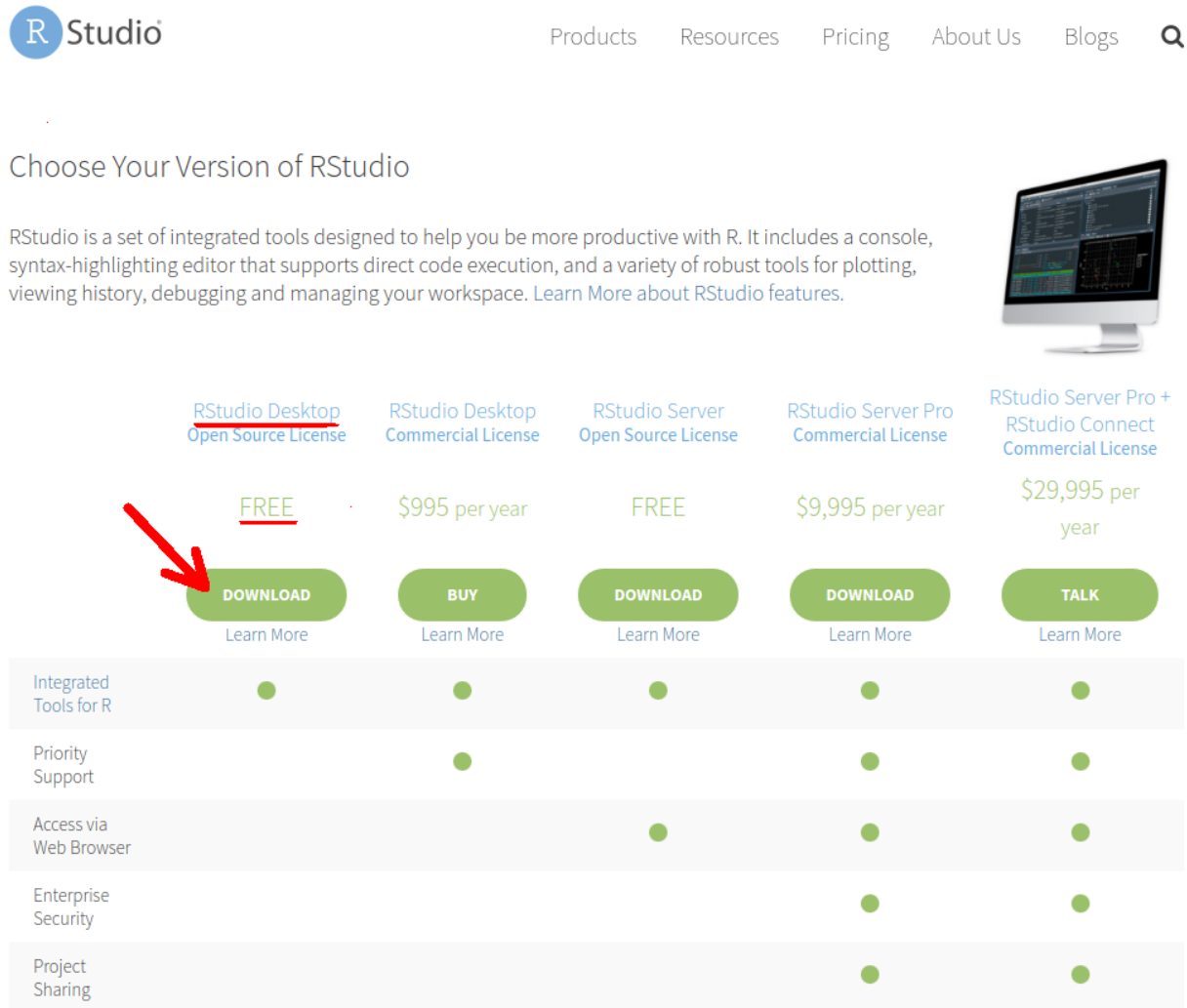
Eventualmente um usuário pode instalar um pacote a partir desses arquivos localmente. Isto pode também ser feito com a função `install.packages()`, especificando o argumento `repos = NULL` e o argumento `pkgs` com o caminho do arquivo. Por exemplo:

```
install.packages("ggplot2_2.1.0.tar.gz", repos=NULL)
```

2.3 RStudio no Ubuntu

é uma empresa que desenvolve ferramentas gratuitas para o e produtos pagos para empresas.

Uma de suas ferramentas gratuitas é o software RStudio *Desktop* que consiste em um ambiente integrado de desenvolvimento (IDE) construído especificamente para o , consequentemente, também é multiplataforma.



Choose Your Version of RStudio

RStudio is a set of integrated tools designed to help you be more productive with R. It includes a console, syntax-highlighting editor that supports direct code execution, and a variety of robust tools for plotting, viewing history, debugging and managing your workspace. [Learn More about RStudio features.](#)

	<u>RStudio Desktop</u> Open Source License	RStudio Desktop Commercial License	RStudio Server Open Source License	RStudio Server Pro Commercial License	RStudio Server Pro + RStudio Connect Commercial License
	FREE	\$995 per year	FREE	\$9,995 per year	\$29,995 per year
	DOWNLOAD Learn More	BUY Learn More	DOWNLOAD Learn More	DOWNLOAD Learn More	TALK Learn More
Integrated Tools for R	●	●	●	●	●
Priority Support		●		●	●
Access via Web Browser			●	●	●
Enterprise Security				●	●
Project Sharing				●	●

Figure 2.3: Opção para baixar o RStudio *Desktop*.

Para instalação da versão do RStudio para *Desktop*, você precisa saber se seu SO é 64 ou 32-bit e a versão do Linux Ubuntu. Essas informações podem ser obtidas, respectivamente, pelos comandos:

```
$ arch
```

```
x86_64
```

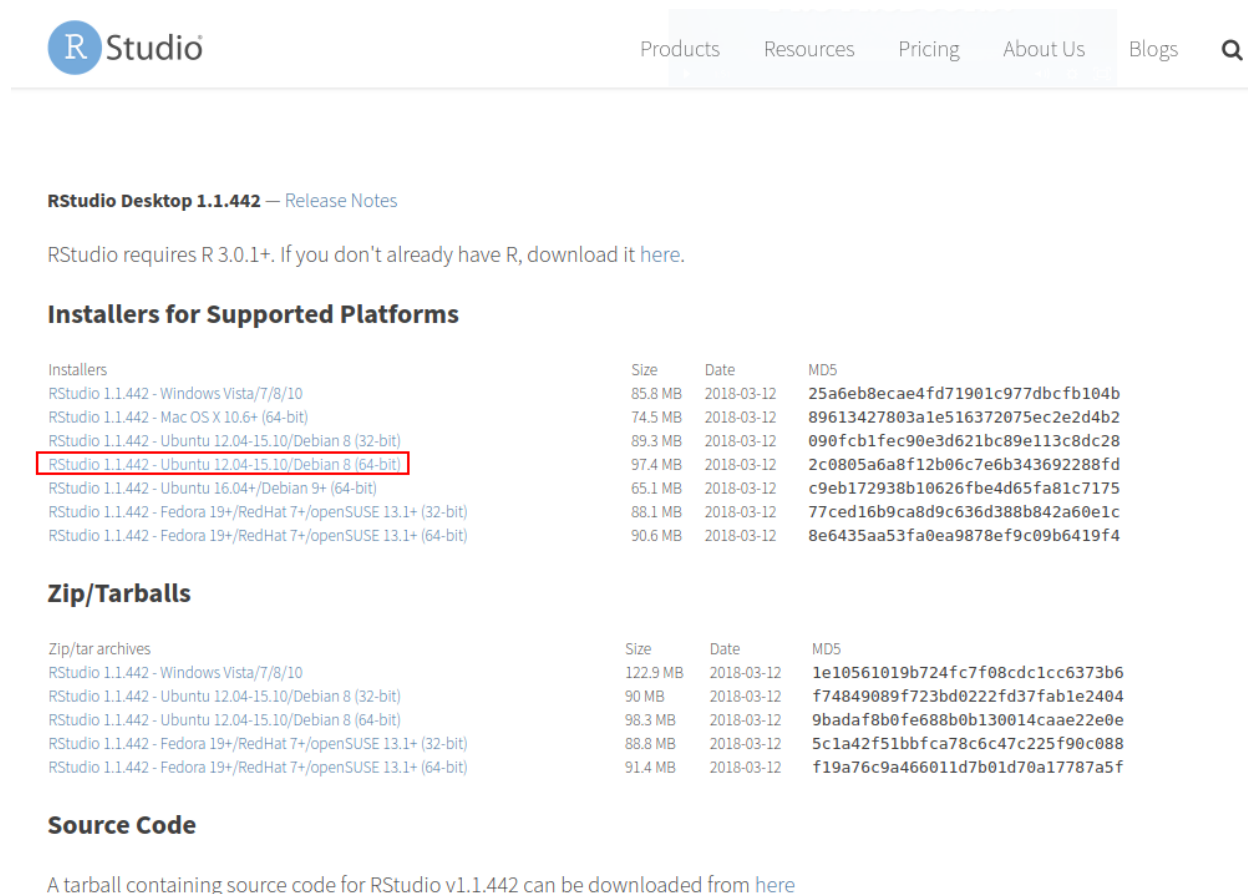
Se retornar **x86_64** sua máquina é 64-bit.

```
$ lsb_release --release | cut -f2
```

```
14.04
```

Com essas informações, siga os seguintes passos:

1. acesse RStudio
2. clique em *Download* (Figura ??)
3. Clique na sua plataforma (de acordo com seu SO, arquitetura e versão da distribuição) (Figura ??), no exemplo deste livro *RStudio 1.1.447 - Ubuntu 12.04-15.10/Debian 8 (64-bit)*
4. Dependendo da sua versão Ubuntu, ao clicar sobre o sobre o arquivo baixado com o botão direito, há a



RStudio Desktop 1.1.442 — Release Notes

RStudio requires R 3.0.1+. If you don't already have R, download it [here](#).

Installers for Supported Platforms

Installers	Size	Date	MD5
RStudio 1.1.442 - Windows Vista/7/8/10	85.8 MB	2018-03-12	25a6eb8ecae4fd71901c977dbcfb104b
RStudio 1.1.442 - Mac OS X 10.6+ (64-bit)	74.5 MB	2018-03-12	89613427803a1e516372075ec2e2d4b2
RStudio 1.1.442 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	89.3 MB	2018-03-12	090fcb1fec90e3d621bc89e113c8dc28
RStudio 1.1.442 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	97.4 MB	2018-03-12	2c0805a6a8f12b06c7e6b343692288fd
RStudio 1.1.442 - Ubuntu 16.04+/Debian 9+ (64-bit)	65.1 MB	2018-03-12	c9eb172938b10626f6be4d65fa81c7175
RStudio 1.1.442 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.1 MB	2018-03-12	77ced16b9ca8d9c636d388b842a60e1c
RStudio 1.1.442 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	90.6 MB	2018-03-12	8e6435aa53fa0ea9878ef9c09b6419f4

Zip/Tarballs

Zip/tar archives	Size	Date	MD5
RStudio 1.1.442 - Windows Vista/7/8/10	122.9 MB	2018-03-12	1e10561019b724fc7f08cdc1cc6373b6
RStudio 1.1.442 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	90 MB	2018-03-12	f74849089f723bd0222fd37fab1e2404
RStudio 1.1.442 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	98.3 MB	2018-03-12	9badaf8b0fe688b0b130014caae22e0e
RStudio 1.1.442 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.8 MB	2018-03-12	5c1a42f51bbfca78c6c47c225f90c088
RStudio 1.1.442 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	91.4 MB	2018-03-12	f19a76c9a466011d7b01d70a17787a5f

Source Code

A tarball containing source code for RStudio v1.1.442 can be downloaded from [here](#)

Figure 2.4: Escolha da plataforma em que será usada o RStudio *Desktop*.

opção de abrir com *Ubuntu Software Center* e então clicar em **instalar**. Se na versão de seu *Desktop* não há esta opção ao clicar com botão direito sobre o arquivo, instale via **terminal**⁶ com os seguintes comandos:

```
$ cd /local/do/arquivo/baixado
$ sudo dpkg -i arquivoBaixado.deb
$ sudo apt-get install -f
```

Abra o RStudio digitando no terminal:

```
$ rstudio &
```

Agora você está pronto para começar a programar em aproveitando as facilidades que o RStudio oferece.

⁶digite 'Ctrl+Alt+t' para abrir um terminal no Linux Ubuntu

Chapter 3

Interface do Usuário

Na maior parte do tempo você provavelmente usará o no **modo interativo**: rodando comandos e vendo os resultados.

Eventualmente esse processo pode ser inconveniente. Por exemplo, no caso de uma análise com um código bem extenso e que precisa ser repetida com dados atualizados semanalmente. Nessa situação, recomenda-se a criação de um script, ou seja, um arquivo texto, com a extensão `.R`, contendo o código de sua análise.

Esse *script* pode ser executado pelo R no **modo de processamento em lote** (do termo em inglês *Batch Processing*) através de um terminal do SO Linux, ou via o Prompt de comando (`cmd.exe`) do SO Windows.

Nesta seção apresenta-se ao leitor estes dois modos de execução do .

3.1 no modo interativo

No Linux o pode ser aberto simplesmente digitando em um terminal a letra R.

```
$ R

R version 3.4.4 (2018-03-15) -- "Someone to Lean On"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```

A janela com a linha de comando do apresenta o *prompt* do (>). Após este símbolo digitamos os comandos, pressionamos a tecla <enter>, o interpreta o comando e retorna o resultado.

Os comandos digitados na linha de comando são chamados de expressões. Esse é o modo iterativo do `R`. Portanto, a linha de comando é a mais importante ferramenta do `R`, pois todas expressões são avaliadas através dela.

```
> 62 + 38
[1] 100
```

A expressão é avaliada pelo `R`, o resultado é mostrado, mas o seu valor é perdido.

O número entre colchetes que aparece como resultado da operação (“[1]” no caso acima) indica o conteúdo resultante da operação iniciando na posição 1 desse objeto. O significado dessa informação torna-se mais óbvio quando trabalhamos com objetos maiores, como por exemplo com vetores. Observe os valores nos colchetes para uma sequência de 100 até 1.

```
> 100:1
 [1] 100  99  98  97  96  95  94  93  92  91  90  89  88  87  86  85  84
[18]  83  82  81  80  79  78  77  76  75  74  73  72  71  70  69  68  67
[35]  66  65  64  63  62  61  60  59  58  57  56  55  54  53  52  51  50
[52]  49  48  47  46  45  44  43  42  41  40  39  38  37  36  35  34  33
[69]  32  31  30  29  28  27  26  25  24  23  22  21  20  19  18  17  16
[86]  15  14  13  12  11  10   9   8   7   6   5   4   3   2   1
```

O elemento [18] da sequência de 100 até 1 é o número 83.

Pode ocorrer da expressão digitada na linha ser muito extensa e ir além de uma linha. Se a expressão estiver incompleta o `R` mostra um sinal de +.

```
> 1 * 2 * 3 * 4 * 5 *
+ 6 * 7 * 8 * 9 * 10
[1] 3628800
```

Execute a expressão abaixo até o sinal de menos e tecle `<enter>`. Enquanto a expressão não estiver completa o sinal de + se repetirá. Até que você digite o número que deseja subtrair de 4.

```
> 4 -
+
+ 3
[1] 1
```

3.1.1 Expressões em sequência

Podemos executar todas expressões anteriores em apenas uma linha, usando o ponto e vírgula ; para separar as expressões:

```
> 62 + 38; 100:1; 1 * 2 * 3 * 4 * 5 * 6 * 7 * 8 * 9 * 10; 4 - 3
[1] 100
 [1] 100  99  98  97  96  95  94  93  92  91  90  89  88  87  86  85  84
[18]  83  82  81  80  79  78  77  76  75  74  73  72  71  70  69  68  67
[35]  66  65  64  63  62  61  60  59  58  57  56  55  54  53  52  51  50
[52]  49  48  47  46  45  44  43  42  41  40  39  38  37  36  35  34  33
[69]  32  31  30  29  28  27  26  25  24  23  22  21  20  19  18  17  16
[86]  15  14  13  12  11  10   9   8   7   6   5   4   3   2   1
[1] 3628800
[1] 1
```

3.1.2 Navegação entre as expressões já avaliadas

Você pode usar as teclas `e` e `p` para navegar entre as expressões já avaliadas pelo `R`. O que é útil quando precisamos repetir um comando anterior com alguma mudança ou para corrigir um erro de digitação ou a omissão de um parentêses.

Quando a linha de comando é usada por muito tempo a sua tela pode ficar poluída com a saída das expressões anteriores. Para limpar a tela, tecla `Ctrl+L`. Assim o console aparece na parte superior do terminal.

```
> 15 + 4
[1] 19
> 100:1
[1] 100 99 98 97 96 95 94 93 92 91 90 89 88 87 86 85 84
[18] 83 82 81 80 79 78 77 76 75 74 73 72 71 70 69 68 67
[35] 66 65 64 63 62 61 60 59 58 57 56 55 54 53 52 51 50
[52] 49 48 47 46 45 44 43 42 41 40 39 38 37 36 35 34 33
[69] 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16
[86] 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1
> #tecle <Ctrl + L>
```

Para parar ou cancelar a execução de uma expressão utilize as teclas `Ctrl + C`. As teclas `Ctrl + L` tem o efeito de limpar a tela.

3.1.3 Comentários

No `R`, a cerquilha `#` (hashtag) é um caracter especial. Qualquer coisa após esse caracter será ignorada pelo `R`. Somente as expressões antes da `#` são avaliadas. Por meio desse símbolo de comentário podemos fazer anotações e comentários no código sem atrapalhar a interpretação das expressões pelo `R`.

```
> # comentário antes do código

> 17 + 3 # comentário ao lado do código: adicionando 17 e 3
[1] 20
```

3.1.4 Auto preenchimento de funções

O `R` inclui o preenchimento automático de nomes de funções e arquivos por meio da tecla `<tab>`. Uma lista de possíveis funções que começam com as letras inicialmente digitadas aparecerão.

```
> read#<tab> pressione <tab> para ver as opções de comandos que iniciam com o termo read
```

3.1.5 Primeiro script

O trecho de código abaixo apresenta nas primeiras linhas algumas expressões do executadas anteriormente. Mas há também, na segunda parte, códigos para salvar um gráfico de pontos num arquivo *pdf*. Na última parte do trecho, define-se uma variável `x` que contém aquela mesma sequência numérica usada no gráfico.

```
# Primeiro script no R
#-----
# cálculos básicos
15 + 4
1:100
1 * 2 * 3 * 4 * 5 * 6 * 7 * 8 * 9 * 10
4-3
#-----
```

```
# salvando um gráfico em um arquivo pdf
arquivo_pdf <- "plot-script1.pdf"
pdf(arquivo_pdf)      # cria e abre um arquivo pdf
plot(1:100)           # faz o gráfico
dev.off()             # fecha o arquivo pdf
#-----
# definindo uma variável x
x <- 1:100
x
```

Este conjunto de linhas de código, quando inseridos em um arquivo texto¹ formam um primeiro *script*. Este *script* pode ser executado pelo através da função `source()`, usando como argumento o caminho para o local do *script*.

```
> source("/home/usuario/adar/script1.R")
```

Este *script* produzirá como saída o arquivo `/home/usuario/adar/plot-script1.pdf`. Você pode visualizar o arquivo para conferir o gráficos de pontos gerado.

3.2 no modo de processamento em lote

Para rodar um *script* no modo de processamento em lote do através do seguinte comando no terminal Linux:

```
$ R CMD BATCH opcoes arqentrada arqsaida
```

Onde: **arqentrada** é o nome do script (arquivo com a extensão `.R`) a ser executado; **arqsaida** é o arquivo (com a extensão `.Rout`) com as saídas dos comandos executados no R; **opcoes** é a lista de opções que controlam a execução.

Vamos rodar como exemplo, o `script1.R` da seção ??.

```
$ R CMD BATCH /home/usuario/adar/script1.R
```

O comando acima, produzirá dois arquivos de saída:

1. `script1.Rout`² criado por *default* quando o **arqsaida** não é especificado, e;
2. arquivo "plot-script1.pdf".

Você pode especificar o nome do **arqsaida** como desejar. No exemplo abaixo, mostra-se como salvar o arquivo de saída incluindo a data em que ele foi gerado, `script1-saida-adatadehoje.log`.

```
$ R CMD BATCH script1.R script1-saida-`date +%Y%m%d`\.log
```

Após a execução do último comando, os mesmos arquivos resultantes do comando anterior serão gerados, exceto pelo primeiro (`.Rout`), que será nomeado `script1-saida-20180423.Rout`.

Para mais opções do comando `R CMD BATCH` digite no terminal do Linux `R --help`.

¹Para fazer isso, você pode usar um editor de texto qualquer (p.ex.: `gedit` no SO Linux, ou `Notepad` no SO Windows).

²Você pode notar que este arquivo tem o mesmo nome do **arqentrada**, exceto que a sua extensão foi alterada para `.Rout`.

Chapter 4

RStudio

O RStudio *Desktop* é um ambiente integrado de desenvolvimento (IDE) para o R. Portanto, o RStudio depende da instalação prévia do R. Ele funciona como uma interface gráfica do usuário (GUI), mas com muito mais potencialidades.

O RStudio é uma ferramenta que potencializará sua interação com o R :

- na produção de gráficos
- na organização de seu código na forma de projetos
- na reprodutibilidade de seu trabalho ou pesquisa
- na manutenção e criação de seus próprios pacotes do R
- na criação e compartilhamento de seus relatórios
- no compartilhamento de seu código e a colaboração com outros

Nessa seção você terá uma visão geral do RStudio *Desktop*.

4.1 Visão geral do RStudio

Assumindo que o RStudio tenha sido instalado (seção ??), ao abri-lo e clicar em *File > New File > R script* você verá uma tela com aspecto similar ao da Figura ??.

O RStudio possui 4 painéis principais:

1. Editor para scripts e visualização de dados
 - abrir e criar scripts
 - rodar scripts
 - código com sintaxe realçada
 - rodar partes do código <Ctrl+enter>
 - rodar todo script <Ctrl+Shift+S>
 - autopreenchimento das funções <tab>
 - comentar linhas <Ctrl+Shift+C>
 - desfazer <Ctrl+Z>
 - refazer <Ctrl+Shift+Z>
 - referência para teclas de atalho <Alt+Shift+K>
 - abrir script com <Ctrl+Click>
 - encontrar e substituir Ctrl+F
2. Console do R

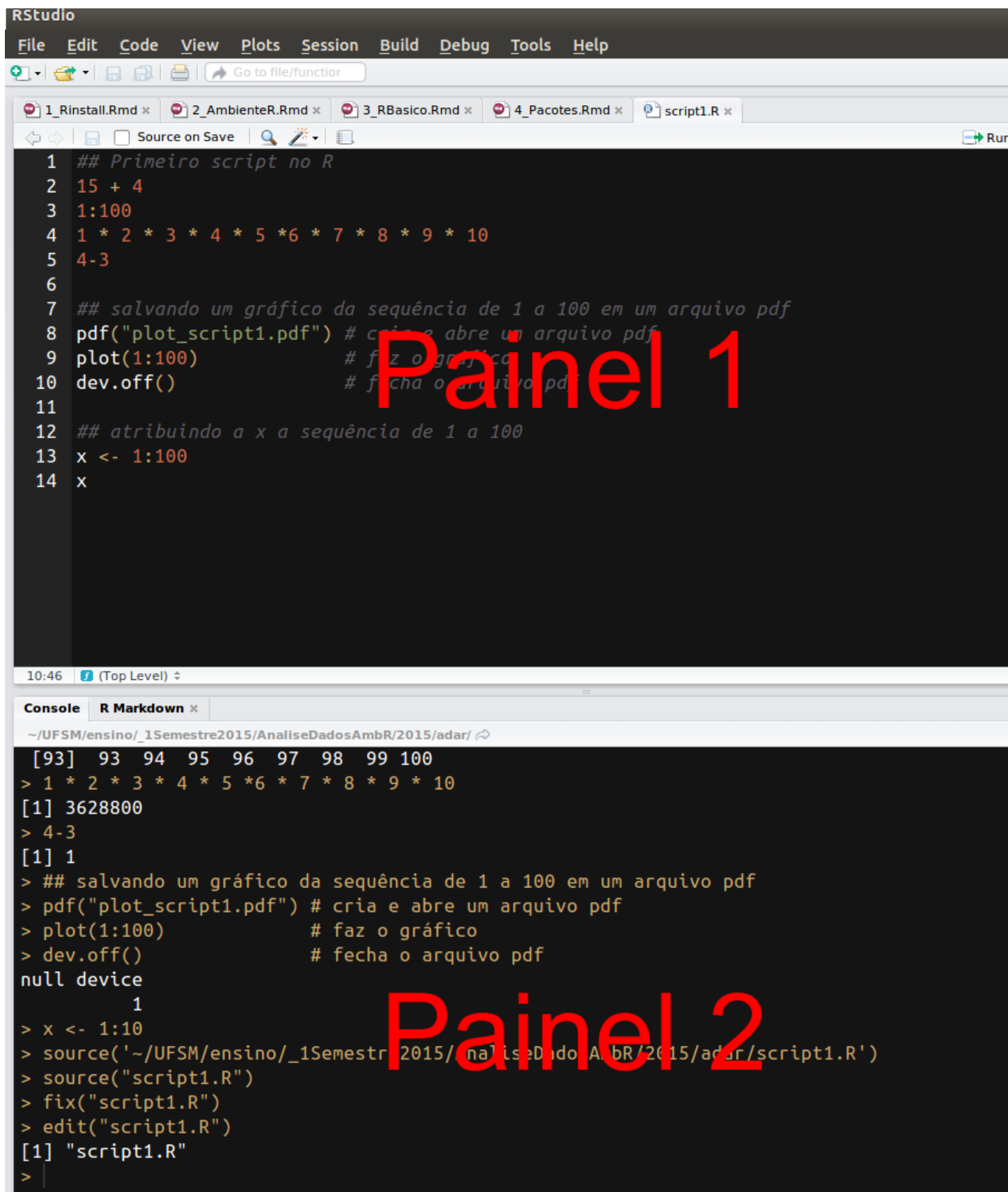


Figure 4.1: Rstudio



- Menus
 - Tools > global Options > Appearance
 - * mostrar linhas, alterar realce da sintaxe
 - Session
 - Plots

Folha de referência do RStudio

Chapter 5

Operações básicas

Nesta seção veremos:

- operações aritméticas básicas com R
- a atribuição de valores a uma variável
- o uso de funções matemáticas internas do R
- valores numéricos especiais do R
- os cuidados ao nomear variáveis

5.1 Convenção

A partir deste capítulo, os códigos a serem avaliadas no R terão o prompt do R (>) omitidos. Essa convenção é para tornar mais fácil a ação de copiar e colar os códigos na linha de comando do R. O resultado da avaliação das expressões será mostrado precedido do símbolo (#>). Esses valores são os resultados que esperam-se sejam reproduzidos pelo leitor na sessão do R em seu computador. Por exemplo:

```
1:5  
#> [1] 1 2 3 4 5
```

No trecho de código acima, a primeira linha contém o código a ser copiado pelo leitor para execução em seu computador. A segunda linha é a saída do código avaliado pelo R.

5.2 Calculadora

O R é uma calculadora turbinada com diversas funções matemáticas disponíveis. Para quem não conhece o R, essa é uma forma de familiarizar-se com a linha de comandos do R.

5.2.1 Aritmética básica

Todas operações feitas em uma calculadora podem ser realizadas na linha de comandos do R.

```
10 + 2 + 4  
#> [1] 16  
# Exemplo de divisao  
(5 + 14)/2  
#> [1] 9.5  
# exponenciação
```

```

2^3
#> [1] 8
4^0.5
#> [1] 2
# operador aritmético para se determinar o resto de uma divisao
10 %% 2
#> [1] 0
2001 %% 2
#> [1] 1
# o inteiro do quociente
11 %/% 2
#> [1] 5

```



Note que no R, o separador decimal é o ponto “.”, ao invés da vírgula “,” usada na notação brasileira. As vírgulas tem a finalidade de separar os argumentos nas chamadas de funções, tal como em `log(x = 10, base = 10)`.

Conheça mais operadores aritméticos, digitando na linha de comando:

```
? "Arithmetic"
```

A janela que se abrirá mostrará o texto que faz parte do manual de ajuda do R.

5.2.2 Constantes

O R possui algumas constantes pré-definidas, como o a constante pi (π).

```

pi
#> [1] 3.141593

```

O R também tem vetores de caracteres pré-definidos, são eles:

```

LETTERS
#> [1] "A" "B" "C" "D" "E" "F" "G" "H" "I" "J" "K" "L" "M" "N" "O" "P" "Q"
#> [18] "R" "S" "T" "U" "V" "W" "X" "Y" "Z"
letters
#> [1] "a" "b" "c" "d" "e" "f" "g" "h" "i" "j" "k" "l" "m" "n" "o" "p" "q"
#> [18] "r" "s" "t" "u" "v" "w" "x" "y" "z"
month.abb
#> [1] "Jan" "Feb" "Mar" "Apr" "May" "Jun" "Jul" "Aug" "Sep" "Oct" "Nov"
#> [12] "Dec"
month.name
#> [1] "January" "February" "March" "April" "May"
#> [6] "June" "July" "August" "September" "October"
#> [11] "November" "December"

```

Note que caracteres estão sempre entre aspas: “”.

“caracteres são entre aspas”

```

aeiou
#> Error in eval(expr, envir, enclos): object 'aeiou' not found

"aeiou"
#> [1] "aeiou"

```

5.2.3 Funções matemáticas internas

Existem diversas funções internas do R que permitem, por exemplo, sortear números aleatoriamente, arredondar números, calcular o fatorial, calcular o seno, cosseno de um ângulo e etc. A sintaxe para chamar uma função no R é:

`função(argumento)`

Por exemplo:

```
# funções trigonométricas
sin(pi/6)
#> [1] 0.5
cos(pi)
#> [1] -1
# raiz quadrada
sqrt(100)
#> [1] 10
# exponencial
exp(1)
#> [1] 2.718282
# fatorial
factorial(4)
#> [1] 24
```

No R você verá que parênteses são frequentemente utilizados. Eles são sempre associados à funções. Qualquer palavra antecedendo um parênteses é uma função.

Para ver a lista completa de funções trigonométricas:

```
? "Trig"
```

5.2.4 Valores numéricos especiais

Um caso particular sobre operação aritméticas no R, são os valores numéricos `Inf` e `NaN` que resultam de operações como:

```
2/0
#> [1] Inf
-12/0
#> [1] -Inf
exp(-Inf)
#> [1] 0
log(0)
#> [1] -Inf
0/Inf
#> [1] 0
(0:3)^Inf
#> [1] 0 1 Inf Inf
log(-0.5)
#> Warning in log(-0.5): NaNs produced
#> [1] NaN
sqrt(-1)
#> Warning in sqrt(-1): NaNs produced
#> [1] NaN
0/0
#> [1] NaN
```

Table 5.1: Tabela 1. Operações com NA.

operação	resultado
NA + 5	NA
sqrt(NA)	NA
NA^2	NA
NA/NaN	NA

```
Inf-Inf
#> [1] NaN
Inf/Inf
#> [1] NaN
mean(c(NA, NA), na.rm = TRUE)
#> [1] NaN
```

NaN é a abreviação para *Not a Number*. Geralmente surge quando um cálculo não tem sentido matemático ou não pode ser propriamente realizado.

A demonstração das diferentes formas de se obter essas constantes especiais é importante para entender a origem delas durante a execução de um script mais extenso.

Outra constante especial do R é o NA (*Not Available*) que representa valor faltante, um problema comum em análise de dados. Qualquer operação envolvendo NA resultará em NA (Tabela 1).

5.2.5 Notação científica e número de dígitos

Na maioria das vezes precisamos trabalhar com números grandes e consequentemente acabamos usando uma notação científica ou exponencial. No R há diferentes formas de representar números com expoentes:

```
1.2e-6
#> [1] 1.2e-06
# expressões equivalentes
1.2E6; 1.2*10^6
#> [1] 1200000
#> [1] 1200000
```

Os resultados dos cálculos no R são mostrados com 7 dígitos significativos, o que pode ser verificado pela `getOption()`. É possível mudar para `n` dígitos usando a função `options()`, conforme exemplo abaixo.

```
# opção de dígitos padrão
getOption("digits")
#> [1] 7
exp(1)
#> [1] 2.718282
# alterando para 14
options(digits = 14)
exp(1)
#> [1] 2.718281828459
getOption("digits")
#> [1] 14
# redefinindo para o número de casas decimais padrão
options(digits = 7)
getOption("digits")
#> [1] 7
```


5.3 Variáveis

5.3.1 Formas de atribuição

5.3.1.1 Variável recebe valor

Até agora nós usamos expressões para fazer uma operação e obter um resultado. O termo "expressão" significa uma sentença de código que pode ser executada. Se a avaliação de uma expressão é salva usando o operador `<-`, esta combinação é chamada "atribuição". O resultado da "atribuição" é armazenado em uma variável e pode ser utilizado posteriormente. Então uma variável é um nome usado para guardar os dados.

```
variavel <- valor
```

```
p <- 1013
# para mostrar a variável digite o nome da variável
p
#> [1] 1013
# ou use a função print()
print(p)
#> [1] 1013
```

O R diferencia letras maiúsculas de minúsculas. Portanto p e P são variáveis diferentes.

```
p
#> [1] 1013
P
#> Error in eval(expr, envir, enclos): object 'P' not found
```

Como criamos apenas a variável p, P não foi encontrada.

A variável p pode ser utilizado para criar outras variáveis.

```
p_pa <- p * 100
# pressão em Pascal
p_pa
#> [1] 101300
```

A seta de atribuição pode ser usada em qualquer sentido. Parênteses, além de estarem sempre acompanhando uma função, também são usados para indicar a prioridade dos cálculos.

```
7/3 + 0.6 -> y1
y1
#> [1] 2.933333
7/(3 + 0.6) -> y2
y2
#> [1] 1.944444
```

Os espaços em torno do símbolo de atribuição (`<-`) não são obrigatórios mas eles ajudam na legibilidade do código.

```
x <- 1
x < -1
# atribuição ou menor que?
x<-1
```

Vamos criar uma variável chamada `ndias3` que recebe o nº de dias no mês de Março e `ndias4` que recebe o nº de dias no mês de Abril.

```
nd3 <- 31
nd4 <- 30
```

O total de dias nos meses de março e abril será armazenado na variável `totdias`:

```
totd <- nd3 + nd4
totd
#> [1] 61
```

A atribuição de um mesmo valor para diferentes variáveis pode ser feita da seguinte forma:

```
# número de dias em cada mês
jan <- mar <- mai <- jul <- ago <- out <- dez <- 31
abr <- jun <- set <- nov <- 30
fev <- 28
# verificação
jan; jul
#> [1] 31
#> [1] 31
jun; set
#> [1] 30
#> [1] 30
fev
#> [1] 28
```

Nós estamos definindo a variável, digitando o nome dela na linha de comando e teclando enter para ver o resultado. Há uma forma mais prática de fazer isso e mostrar o resultado cercado a atribuição por parênteses:

```
# ao invés de
# tar <- 20
# tar
# é mais prático
(tar <- 20)
#> [1] 20
```

Se desejamos calcular e já visualizar o valor da pressão de vapor de saturação obtida com a equação de Tetens, podemos fazer:

```
(es <- 0.611 * exp((17.269 * tar)/(tar + 237.3)))
#> [1] 2.338865
```

Quando usamos a mesma variável numa sequência de atribuições o seu valor é sobrescrito. Portanto não é bom usar nomes que já foram usados antes, exceto se a intenção for realmente essa. Para saber os nomes das variáveis já usados use a função `ls()`¹ para verificar as variáveis existentes:

```
ls()
#> [1] "abr"      "ago"      "dez"      "es"       "fev"      "jan"
#> [7] "jul"      "jun"      "mai"      "mar"      "nd3"      "nd4"
#> [13] "nov"      "oper_nas" "out"      "p"        "pcks"     "p_pa"
#> [19] "rblue"    "set"      "tar"      "totd"     "y1"       "y2"

totd <- jan*7; totd <- totd + fev; totd <- totd + 4*abr
totd
#> [1] 365
```

5.3.1.2 Atribuição com a função `assign()`

Outra forma de atribuição é através da função `assign()`:

¹Essa lista de variáveis também é mostrada no painel *Environment* do RStudio (canto direito superior, aba *Environment*).

```
es
#> [1] 2.338865
assign(x = "es_hpa", value = es/10)
es_hpa
#> [1] 0.2338865
# usando função assign sem nome dos parâmetros
assign("u", 2.5)
u
#> [1] 2.5
```

Um exemplo mais elaborado de uso da função `assign()` para criar várias variáveis pode ser visto aqui.

5.3.2 Removendo variáveis

Para remover variáveis usa-se a função `rm()`.

```
# lista de variáveis existentes
ls()
#> [1] "abr"      "ago"      "dez"      "es"      "es_hpa"   "fev"
#> [7] "jan"      "jul"      "jun"      "mai"     "mar"     "nd3"
#> [13] "nd4"      "nov"      "oper_nas" "out"     "p"       "pcks"
#> [19] "p_pa"     "rblue"    "set"      "tar"     "totd"    "u"
#> [25] "y1"      "y2"
```

Vamos remover a variável `u` criada previamente e ver a lista de objetos no espaço de trabalho.

```
rm(u)
# lista de variáveis existentes, sem u
ls()
#> [1] "abr"      "ago"      "dez"      "es"      "es_hpa"   "fev"
#> [7] "jan"      "jul"      "jun"      "mai"     "mar"     "nd3"
#> [13] "nd4"      "nov"      "oper_nas" "out"     "p"       "pcks"
#> [19] "p_pa"     "rblue"    "set"      "tar"     "totd"    "y1"
#> [25] "y2"
```

Podemos remover mais de uma variável ao mesmo tempo.

```
rm(es_hpa, es, tar, y1, y2)
# lista de variáveis existentes, sem es_hpa, es, tar, y1, y2
ls()
#> [1] "abr"      "ago"      "dez"      "fev"     "jan"     "jul"
#> [7] "jun"      "mai"      "mar"      "nd3"     "nd4"     "nov"
#> [13] "oper_nas" "out"      "p"        "pcks"    "p_pa"    "rblue"
#> [19] "set"      "totd"
```

Para remover todas variáveis do espaço de trabalho (use com cautela):

```
# apagando tudo
rm(list = ls())
ls()
#> character(0)
```

5.3.3 Nomeando variáveis

É preciso ter cuidado ao nomear variáveis no R porque existem algumas regras: