

Análise de Dados Ambientais com R

Jônatan Tatsch

2018-04-17

Contents

Apresentação	5
1 Introdução	7
1.1 Análise de dados meteorológicos	7
1.2 Ciência de dados	7
1.3 Etapas para abordagem de um problema	7
1.4 Programação computacional	8
1.5 R	8
1.6 Por que o R?	8
1.7 Pacotes da comunidade do R	8
1.8 Por que um meteorologista usaria o R?	8
1.9 R não é perfeito!	9
1.10 RStudio	9
1.11 Para saber mais sobre o R	9
2 Instalação do R e RStudio	11
2.1 Instalação	11
2.2 Diretório para instalação de pacotes	14
2.3 Testando o <i>R</i>	15
2.4 Rstudio no Ubuntu	15
3 Interface do Usuário	21
3.1 Pluviometria	21
4 Operações básicas	23
4.1 Formação do solo	23
4.2 Propriedades do solo	23
4.3 Textura e Estrutura	23
4.4 Movimento da água no solo	23
4.5 Infiltração	23
5 Tipos de dados	25
5.1 Evapotranspiração de referência	25
5.2 Evapotranspiração da cultura	25
5.3 Evapotranspiração real	25

Apresentação

Este material é uma composição das notas de aula da disciplina **Análise de Dados Ambientais com R** 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 R 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 R. Estes procedimentos são potencializados com o uso do software RStudio, uma interface de desenvolvimento integrado (IDE) para o R.

Neste livro o leitor aprenderá a sintaxe básica da linguagem R (R Core Team, 2018), a importação e exportação de dados, a criação de gráficos, funções, a padronização e organização de conjunto de dados; 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 R apresenta-se as capacidades específicas do R 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 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:

- 1 Introdução
- 2 Instalação do R e Rstudio
- 3 Interface do Usuário
- 4 Operações Básicas
- 5 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

- 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*)
- 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 n^o 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 RStudio

RStudio é um **ambiente de desenvolvimento integrado** livre e de código aberto.

- Multiplataforma (Windows, Linux e Mac)
- ênfase da sintaxe do R, auto-preenchimento de código, indentação inteligente
- execução do R diretamente do editor
- manejo de diretórios e projetos
- histórico de gráficos, zoom, atalhos para exportar imagens
- Integrado com knitr
- Integrado com GitHub para controle de versões

1.11 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

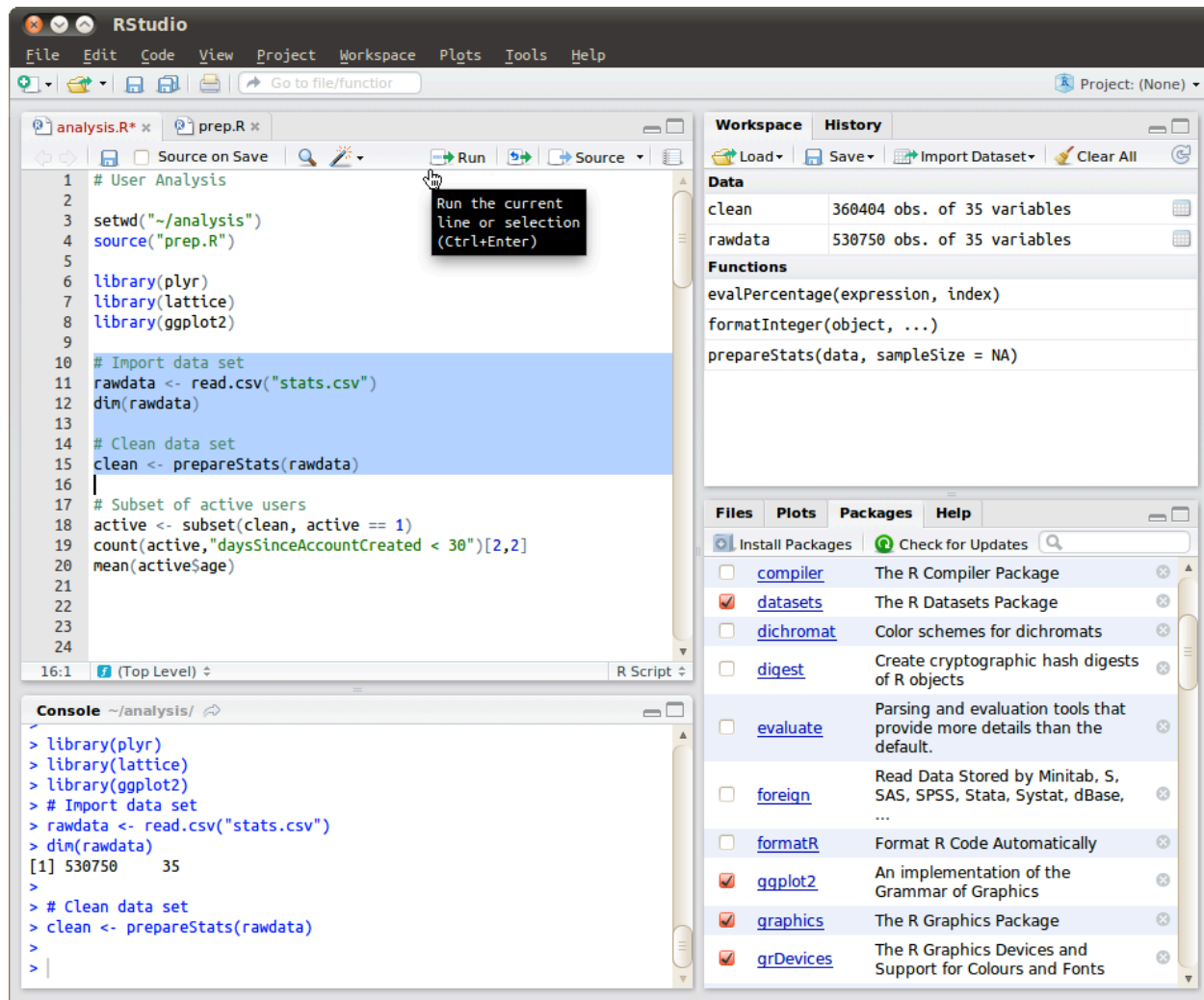


Figure 1.1: RStudio IDE

Chapter 2

Instalação do R e RStudio

O *R* é um software livre, de código fonte aberto e funciona em diversos sistemas operacionais (Linux, Windows e MacOS). O usuário interage com o *R* pela linha de comando. Essa interação pode ser facilitada por meio de uma interface gráfica como o RStudio.

A seguir descreve-se como instalar o *R* no Windows e no Linux Ubuntu. A forma de instalação do *R* no Linux tenta ser mais didática do que prática. Alguns comandos linux básicos serão utilizados, mas mesmo quem não é usuário linux será capaz de entendê-los.

2.1 Instalação

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 binários

2.1.1 Windows

A forma de instalar o *R* no Windows é baixar o binário executável da **Rede Abrangente de Arquivos do R** (CRAN). Depois clicar em *Download R for Windows* e *install R for the first time*. Quando este tutorial foi escrito a última versão foi a *R* 3.4.4.

A instalação do *R* para Windows a partir do executável acima incluirá na instalação uma GUI chamada *RGui.exe*, mostrada abaixo.

2.1.2 Linux

2.1.2.1 Ubuntu

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

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

2.1.2.2 R sempre atualizado

Se você prefere trabalhar com a última versão estável do *R*, precisamos configurar o Linux Ubuntu para atualizar automaticamente o *R*. Também uma boa prática definir um diretório para armazenamento dos pacotes utilizados.

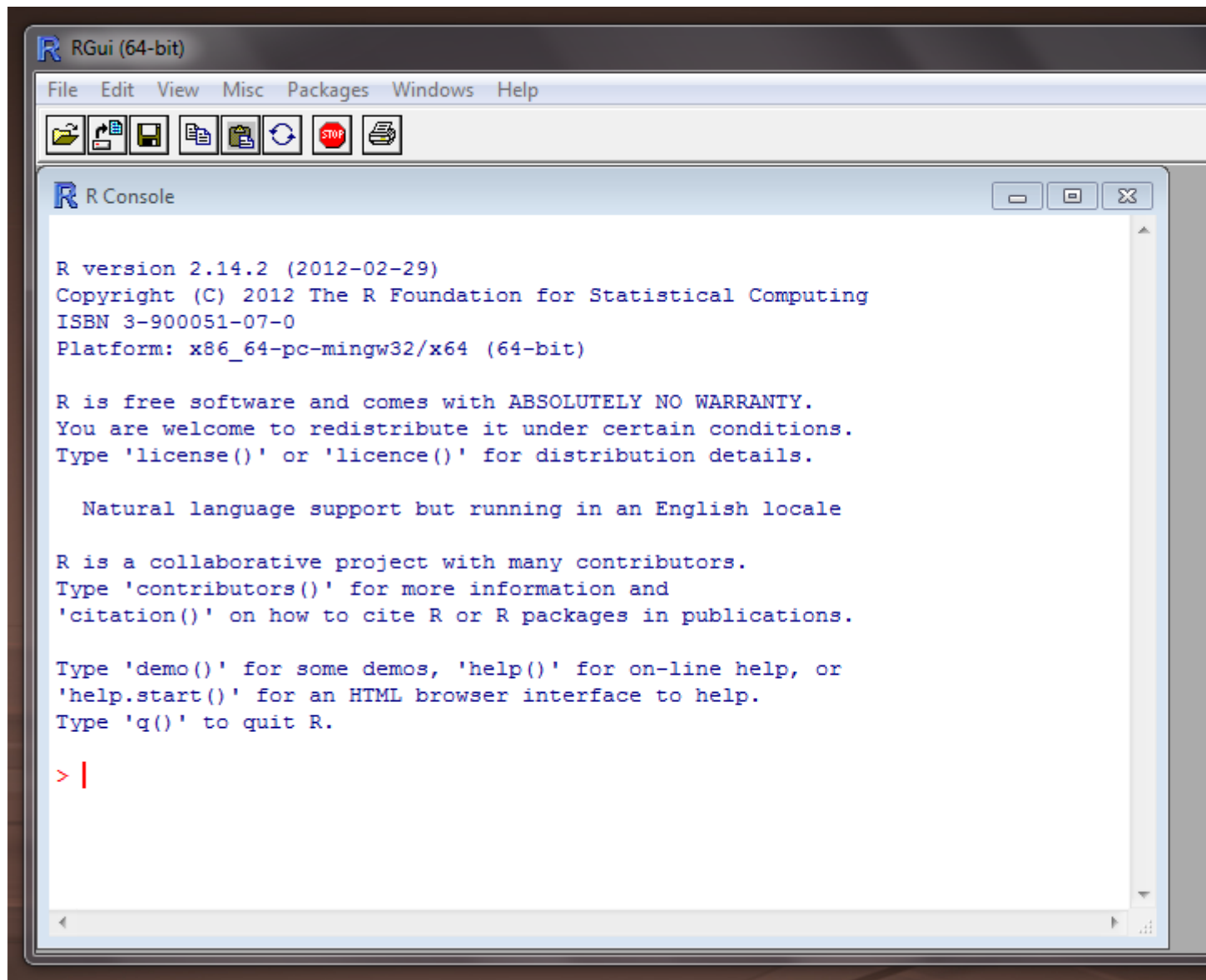


Figure 2.1: Interface gráfica do usuário no R para Windows.

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`)

O R é distribuído na CRAN. Geralmente há duas atualizações ao ano. A versão mais atual é a R version 3.4.4 (2018-03-15). Para que ele seja atualizado automaticamente no Ubuntu precisamos adicionar o repositório do R mais próximo da nossa região à lista de repositórios do Linux. No nosso caso, o repositório mais próximo é o da UFPR (<http://cran-r.c3sl.ufpr.br/>).

2.1.2.2.1 Incluindo repositório do R na Lista de repositórios do Ubuntu

A lista de repositórios do sistema é armazenada no arquivo `/etc/apt/sources.list`. Vamos visualizar o conteúdo desse arquivo. Em um terminal linux (use o atalho `Ctrl+Alt+t`), digite o seguinte comando:

```
$ cat /etc/apt/sources.list | head -15

## # deb cdrom:[Ubuntu 14.04.2 LTS _Trusty Tahr_ - Release amd64 (20150218.1)]/ trusty main restricted
##
## # See http://help.ubuntu.com/community/UpgradeNotes for how to upgrade to
## # newer versions of the distribution.
## # deb http://br.archive.ubuntu.com/ubuntu/ trusty main restricted
## # deb-src http://br.archive.ubuntu.com/ubuntu/ trusty main restricted
## deb http://ubuntu.c3sl.ufpr.br/ubuntu/ trusty main restricted
## deb-src http://ubuntu.c3sl.ufpr.br/ubuntu/ trusty main restricted
## ## Major bug fix updates produced after the final release of the
## ## distribution.
## deb http://br.archive.ubuntu.com/ubuntu/ trusty-updates main restricted
## deb-src http://br.archive.ubuntu.com/ubuntu/ trusty-updates main restricted
##
## ## N.B. software from this repository is ENTIRELY UNSUPPORTED by the Ubuntu
## ## team. Also, please note that software in universe WILL NOT receive any
```

Para descobrir o nome da versão do sistema operacional, digite na terminal linux o seguinte comando²:

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

trusty
```

Precisamos incluir no arquivo `sources.list` o repositório da UFPR. 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 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:

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.

```
deb `linkRepositorioSelecioneado`/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
```

2.1.2.2.2 APT protegido

Os arquivos binários do *R* 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 *R*

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
```

2.2 Diretório para instalação de pacotes

Para termos melhor controle sobre os pacotes do *R* instalados no sistema é recomendado criar um diretório (`/home/usuario/.R/libs`) no seu home. O diretório do home pode ser obtido com o comando `echo $HOME`.

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

Para informar ao *R* onde procurar os pacotes que instalamos criamos um arquivo chamado `.Renviron` no diretório `$HOME` contendo a linha `R_LIBS=/home/usuario/.R/libs/`, o que pode ser feito com o comando:

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

```
$ echo $R_LIBS >> `echo $HOME`/.Renviron`
```

Esse caminho fica então visível ao *R*, o que pode ser verificado executando a função `.libPaths()` no console do *R*.

```
$ R
```

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.

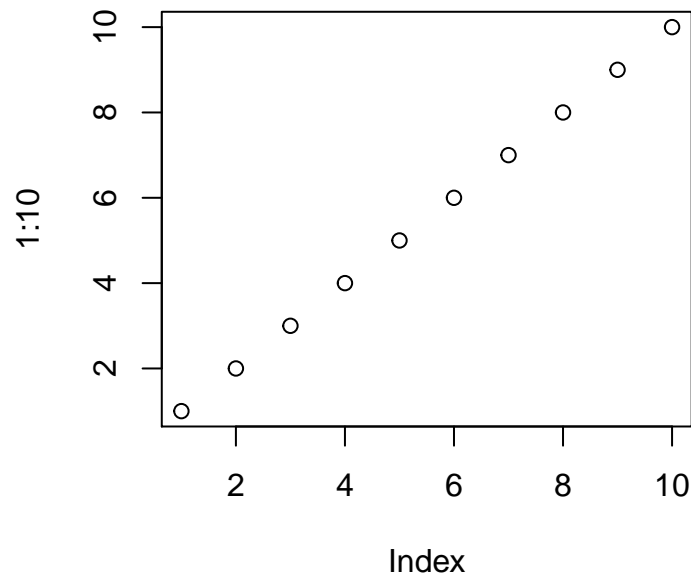


Figure 2.2: Gráfico da sequência de 10 números.

No console do R:

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

2.3 Testando o R

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

```
$ R
```

A partir desse momento já começamos uma sessão no R. 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)
```

Vamos sair do R sem salvar os dados da sessão.

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

2.4 Rstudio no Ubuntu

O RStudio é um ambiente integrado de desenvolvimento (IDE) construído especificamente para o R. Assim como o R, o RStudio pode ser baixado gratuitamente e é multiplataforma.

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

```
$ arch  
x86_64
```

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

14.04

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


Com essa informação e versão do sistema operacional, siga os seguintes passos:

1. acesse RStudio
2. clique em *Download RStudio*
3. Procure a opção *RStudio Desktop (FREE)* e clique *download*

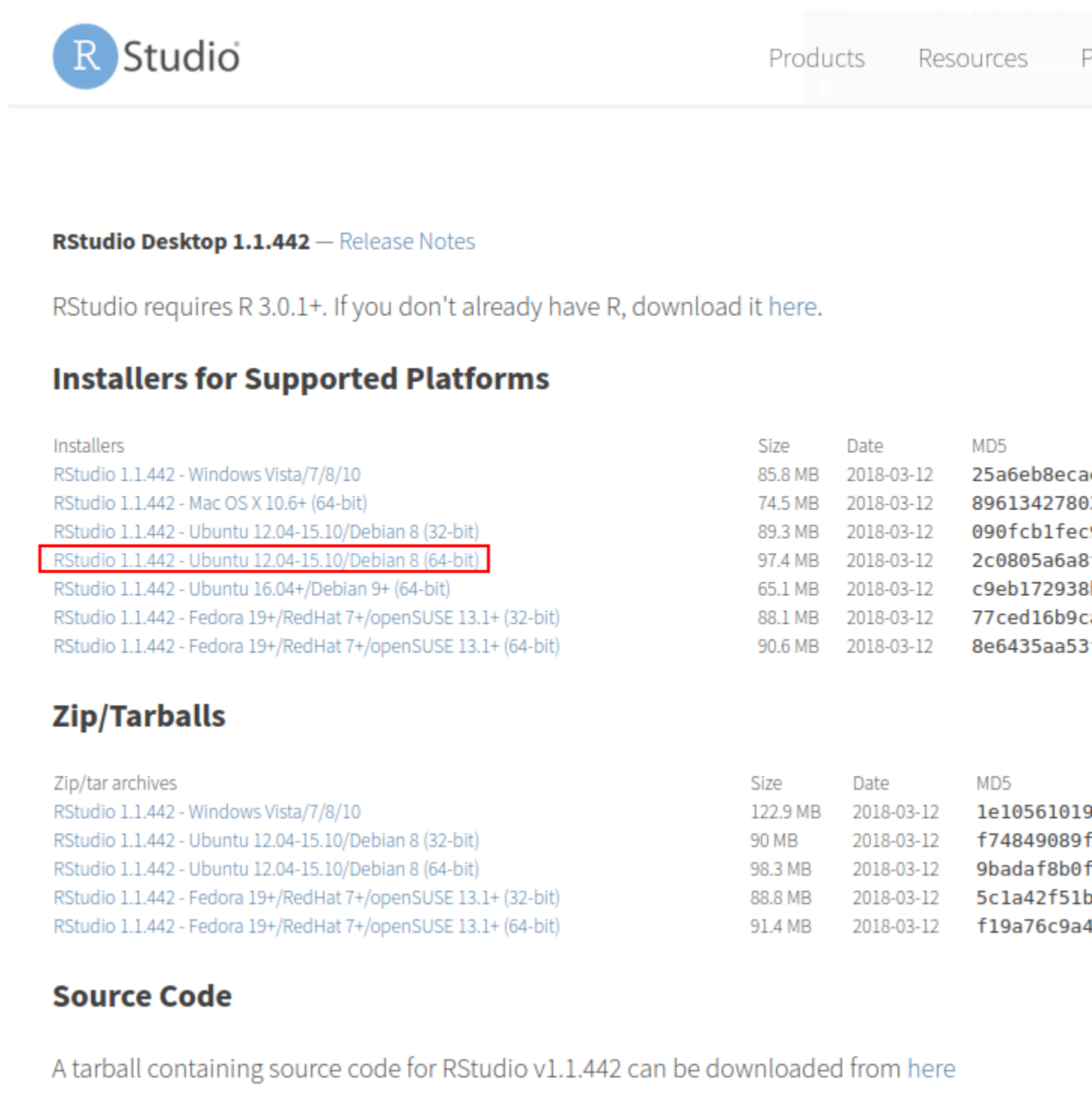
[Products](#)[Resources](#)[Pricing](#)

Choose Your Version of RStudio

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

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

5. Selecione sua plataforma



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	25a6eb8ecac
RStudio 1.1.442 - Mac OS X 10.6+ (64-bit)	74.5 MB	2018-03-12	8961342780
RStudio 1.1.442 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	89.3 MB	2018-03-12	090fcb1fec
RStudio 1.1.442 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	97.4 MB	2018-03-12	2c0805a6a8
RStudio 1.1.442 - Ubuntu 16.04+/Debian 9+ (64-bit)	65.1 MB	2018-03-12	c9eb172938
RStudio 1.1.442 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.1 MB	2018-03-12	77ced16b9c
RStudio 1.1.442 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	90.6 MB	2018-03-12	8e6435aa53

Zip/Tarballs

Zip/tar archives	Size	Date	MD5
RStudio 1.1.442 - Windows Vista/7/8/10	122.9 MB	2018-03-12	1e10561019
RStudio 1.1.442 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	90 MB	2018-03-12	f74849089f
RStudio 1.1.442 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	98.3 MB	2018-03-12	9badaf8b0f
RStudio 1.1.442 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.8 MB	2018-03-12	5c1a42f51b
RStudio 1.1.442 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	91.4 MB	2018-03-12	f19a76c9a4

Source Code

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

clique sobre o link da sua plataforma, p.ex.: *RStudio x.xx.xxx - Ubuntu 12.04-15.10/Debian 8 (64-bit)*

6. Dependendo da sua versão Ubuntu, ao clicar sobre o sobre o arquivo baixado com o botão direito, há a 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**⁵ com os seguintes

⁵digite 'Ctrl+Alt+t' para abrir um terminal no Linux Ubuntu

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 *R* aproveitando as facilidades que o RStudio oferece.

Chapter 3

Interface do Usuário

```
knitr::include_graphics('images/SatCloud.gif')
```

3.1 Pluviometria

Em construção.

Chapter 4

Operações básicas

Os solos são caracterizados por uma típica sequência de horizontes que constituem o perfil do solo.

4.1 Formação do solo

4.2 Propriedades do solo

4.2.1 Materiais

4.2.2 Hídricas

4.3 Textura e Estrutura

4.4 Movimento da água no solo

4.5 Infiltração

4.5.1 Modelo de frente de umedecimento de Green-Ampt

asd

Chapter 5

Tipos de dados

A EVAPORAÇÃO ocorre quando uma superfície úmida está exposta a ar relativamente mais seco. Quando parcelas de ar movem-se sobre a superfície elas carregam umidade daquela superfície.

A água evaporada da superfície aumenta a quantidade de vapor d'água no ar. Quando o ar está saturado com vapor d'água a evaporação cessa.

A TRANSPIRAÇÃO é a evaporação da água das folhas das plantas quando esta move-se do solo, através das plantas e através das folhas para o ar.

As plantas consomem grandes quantidades de água durante o crescimento. Um lavoura de milho de 4000 m² pode consumir 10000-15000 litros de água (2,5 - 3.75 mm) em um dia. Uma árvore sem restrição hídrica pode transpirar 100-150 litros por dia. Os processos meteorológicos próximo a superfície controlam a EVAPORAÇÃO e TRANSPIRAÇÃO. A TRANSPIRAÇÃO é também regulada pela fisiologia das plantas. Quando as plantas cobrem uma pequena porção do solo a evaporação do solo é o fluxo dominante. A TRANSPIRAÇÃO torna-se mais importante com o aumento da cobertura de área das plantas aumenta. Entretanto é difícil distinguir EVAPORAÇÃO de TRANSPIRAÇÃO e os dois termos são frequentemente combinados em EVAPOTRANSPIRAÇÃO.

EVAPORAÇÃO TRANSPIRAÇÃO EVAPOTRANSPIRAÇÃO

5.1 Evapotranspiração de referência

5.2 Evapotranspiração da cultura

5.3 Evapotranspiração real

Bibliography

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.