Disciplina

Análise de Dados de Ciência do Solo no R — Aula 4

Introdução ao R

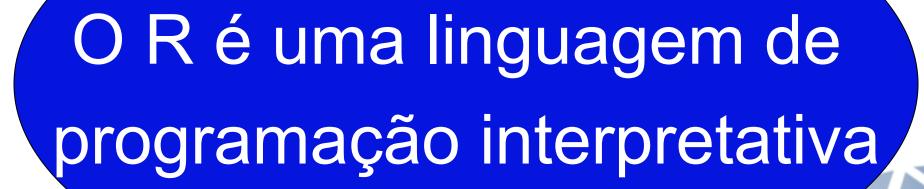


Alan Rodrigo Panosso Departamento de Ciências Exatas

alan.panosso@unesp.br



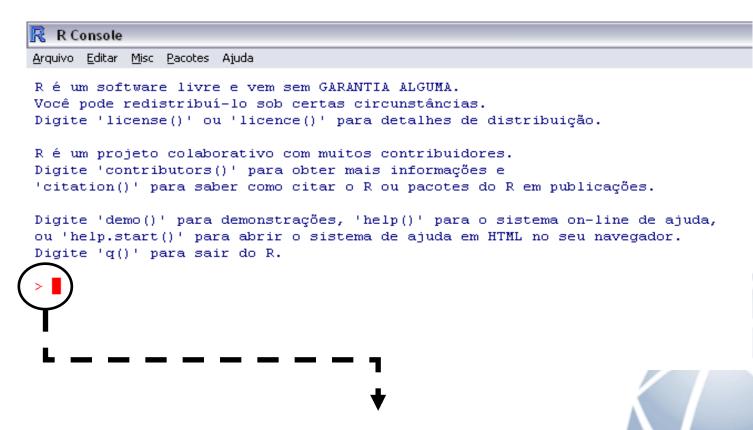
O que é o R



Todos os comandos digitados em seu *prompt* são executados diretamente pelo programa, sem a necessidade de uma compilação prévia.

O que é o R

Tela inicial do sistema R versão xxxxx:



Este símbolo (>) é o *prompt* do R e ele indica que o sistema está pronto para receber seus comandos. Exemplo citation()

O R é um software de domínio público, utilizado para análise de dados em geral

O que é o R

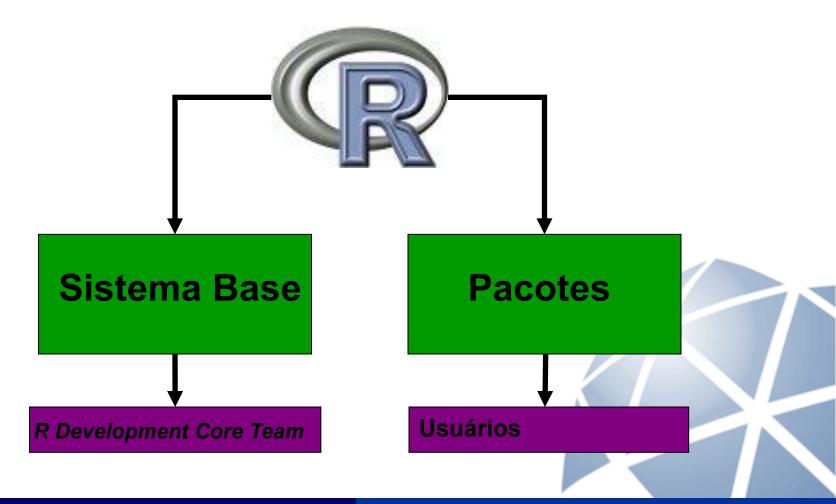
A sintaxe do R é simples e intuitiva, por exemplo, em uma análise de regressão linear, poder-seá utilizar a função lm(y ~ x) o que significa "ajuste um modelo linear contendo y como variável dependente e x como variável independente". Se digitarmos apenas o nome da função, o R mostrará o seu conteúdo, a sua programação, exemplo, função aov, análise de variância.

```
Você pode redistribuí-lo sob certas circunstâncias.
Digite 'license()' ou 'licence()' para detalhes de distribuição.
R é um projeto colaborativo com muitos contribuidores.
Digite 'contributors()' para obter mais informações e
'citation()' para saber como citar o R ou pacotes do R em publicações.
Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda,
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digite 'q()' para sair do R.
[Área de trabalho anterior carregada]
> x < -c(2,4,6,8,10)
> v < -c(3,6,10,15,17)
> lm(v~x)
Call:
lm(formula = y \sim x)
Coefficients:
(Intercept)
                    1.85
      -0.90
```

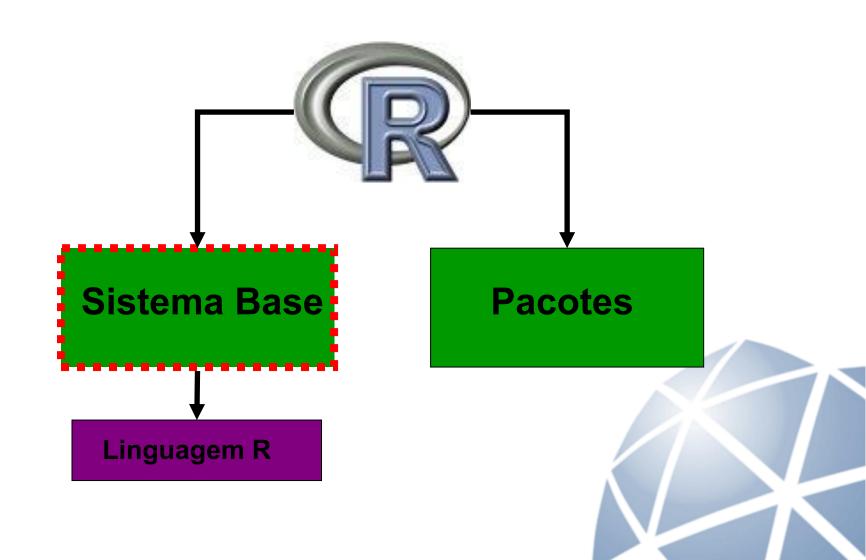
```
function (formula, data = NULL, projections = FALSE, gr = TRUE,
   contrasts = NULL, ...)
   Terms <- if (missing(data))
       terms (formula, "Error")
   else terms(formula, "Error", data = data)
   indError <- attr(Terms, "specials") $Error
   if (length(indError) > 1L)
        stop(sprintf(ngettext(length(indError), "there are %d Error terms: only$
            "there are %d Error terms: only 1 is allowed"), length(indError)),
           domain = NA)
   lmcall <- Call <- match.call()</pre>
   lmcall[[1L]] <- as.name("lm")</pre>
   lmcall$singular.ok <- TRUE
   if (projections)
       gr <- lmcall$gr <- TRUE
   lmcall$projections <- NULL</pre>
   if (is.null(indError)) {
       fit <- eval(lmcall, parent.frame())
       if (projections)
           fit$projections <- proj(fit)
       class(fit) <- if (inherits(fit, "mlm"))</pre>
           c("maov", "aov", oldClass(fit))
       else c("aov", oldClass(fit))
       fit$call <- Call
        return(fit)
   else {
       if (pmatch("weights", names(match.call()), OL))
            stop("weights are not supported in a multistratum aov() fit")
       opcons <- options("contrasts")
       options(contrasts = c("contr.helmert", "contr.poly"))
       on.exit(options(opcons))
       allTerms <- Terms
       errorterm <- attr(Terms, "variables")[[1 + indError]]
        eTerm <- deparse(errorterm[[2L]], width.cutoff = 500L,
            backtick = TRUE)
```

Apresenta código fonte aberto, ou seja, pode ser modificado ou implementado por usuários

O que é o R



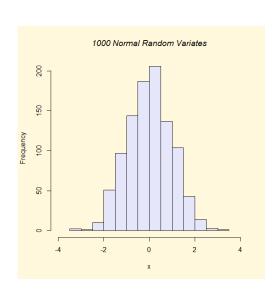
O que é o R

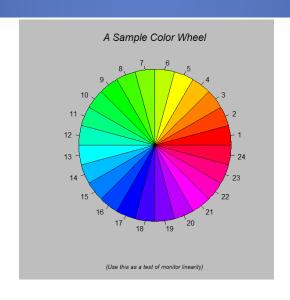


"R é uma linguagem de alto nível e um ambiente para análise de dados e gráficos"

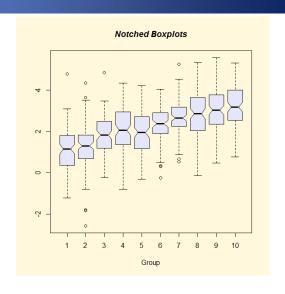
Crawley, 2007

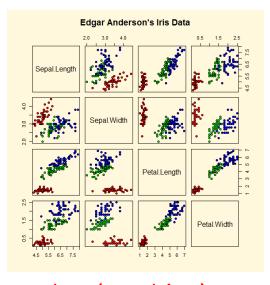
O que é o R

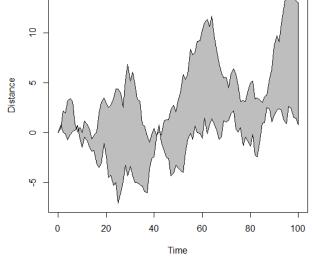


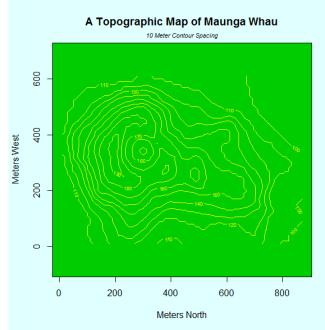


Distance Between Brownian Motions









demo(graphics)

Adequação geral ao fluxo de trabalho:

- 1. Adequado a trabalhos que necessitam de tradicionais linguagens de programação como *C*, *Fortran* e *Pascal*.
- Similar a softwares matemáticos como MATHEMATICA e MAPLE, com várias bibliotecas para análises simbólicas.
- 3. Desempenho de softwares estatísticos com *SPlus* e *SAS* com várias bibliotecas para análise de dados.

Confiabilidade

1960

1988

1996

Atualmente

Linguagem S (John Chambers *et al;* Lucent Technologies) S-Plus. Software proprietário de análise de dados que contém a linguagem S

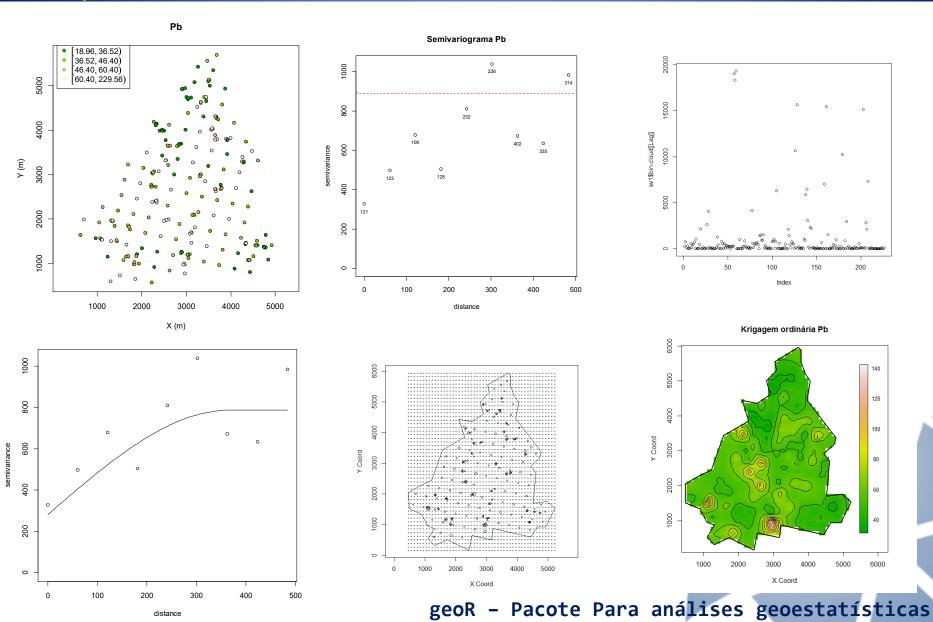
Linguagem orientada a objetos R (Ross Ihaka e Robert Gentleman)

Crescente utilização no mercado de trabalho e principal ferramenta para pesquisas estatísticas reprodutíveis.

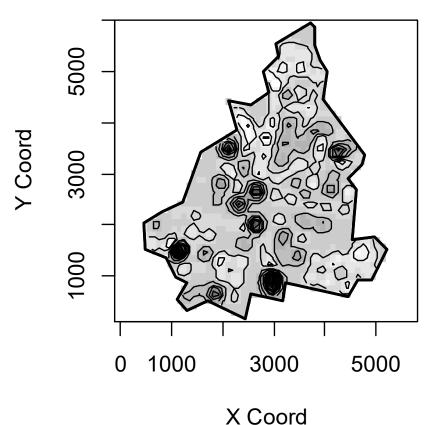
- O sistema R é um ambiente que incorpora uma implementação da linguagem de programação S, que é poderosa, flexível e possui excelentes facilidades gráficas (R Development Core Team);
- R é um projeto *open-source* e está disponível na internet sobre a *General Public License* (www.gnu.org/copyleft/gpl.html e www.fsf.org);
- Home-page oficial: www.r-project.org

Longevidade

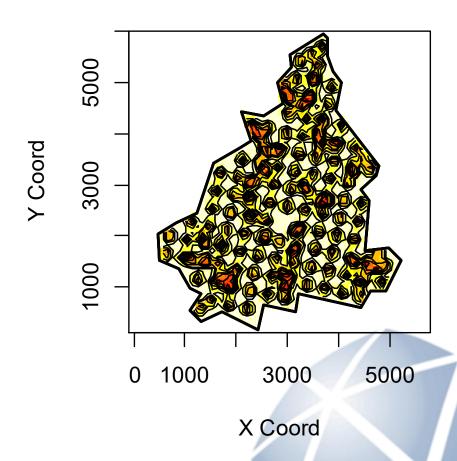
- A Abrangência e disponibilidade das mais variadas aplicações no campo da análise de dados, das simples às sofisticadas técnicas.
- Mais e mais autores tem utilizado o software em suas pesquisas, a compreensão da literatura é essencial.
- Grandes pesquisadores em estatística já migraram para o R, observe em sua área.
- Contribuição: Esses mesmo pesquisadores, são aqueles que dispensam seu tempo e esforços para o desenvolvimentos de pacotes específicos para as mais variadas aplicações.



Krigagem ordinária Pb



Krigagem erro da estimativa Pb



geoR - Pacote Para análises geoestatísticas, desenvolvido pelo Prof. Paulo Justiniano Ribeiro Júnior (UFPR)

Reprodutibilidade

- Princípio fundamental do método científico.
- A reprodução de resultados em diferentes configurações de softwares e hardwares, torna-se impossível sem a utilização de programas de código aberto.
- Políticas de disponibilização de partes do código utilizado na pesquisa.
- Que atualmente têm sido adotadas por periódicos como Biostatistics,
 Science e Geoscientific Model Development.

Liberdade

O R é distribuído sobre os termos de licença pública GNU.

Existem razões ideológicas para apoiar um programa por uma fundação acadêmica pública, que distribui o seu trabalho grátis para todo o mundo.

O termo código fonte aberto significa que existem vários cientistas trabalhando para a melhoria contínua do programa.

Instalação

O sistema base e os pacotes são distribuídos via *Comprehensive R Archive Network (CRAN)* acessível sobre

http://cran.r-project.org

O analista de dados precisa baixar a distribuição binária pré-compilada e instalá-la localmente. Para os usuários do *Windows* segue o link

http://cran.r-project.org/bin/windows/base/R-3.1.1-win.exe

O usuário irá baixar o arquivo correspondente (atualmente chamado *R-3.1.0-win.exe*), execute-o localmente e siga as instruções dada pelo instalador.

Após a instalação procure o ícone do R criado pelo instalador e clique nele



Instalação

```
Arquivo Editar Misc Pacotes Ajuda

R é um software livre e vem sem GARANTIA ALGUMA.

Você pode redistribuí-lo sob certas circunstâncias.

Digite 'license()' ou 'licence()' para detalhes de distribuição.

R é um projeto colaborativo com muitos contribuidores.

Digite 'contributors()' para obter mais informações e
'citation()' para saber como citar o R ou pacotes do R em publicações.

Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda, ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.

Digite 'q()' para sair do R.
```

```
> 25+35
[1] 60
```

Poderemos nos deparar com o símbolo (+) na parte esquerda do **prompt**, ao invés do (>). Isto indica que o sistema está esperanto pelo término do comando

```
> 25+(35+53
```

Neste caso podemos completar o comando digitando), e o R retornará o resultado 113. No caso de erros, pressione a tecla ESC para (>) retornar na linha do prompt.

Para sair do R, use a função: q() ou a opção de *Menu*: *arquivo / sair*

Funções matemáticas (calculadora)

OBS.: Para adicionar um comentário no R basta colocar o símbolo

Cálculos apenas com os sinais de operação aritmética:

- > 1 + 7 #cálculo da soma destes dois números
- > 1 2 * 10 #prioridade da multiplicação
- > 3 ** 7 #potências utilizando ** ou ^
- > 10/3 #divisão

Cálculos com funções encontradas em calculadoras científicas:

- > sqrt(9) #raiz quadrada
- > sin(pi) #seno de Pi radianos é zero
- > factorial(4)#4!=4*3*2*1
- $> \exp(5)$ #exponencial de 5, e^5

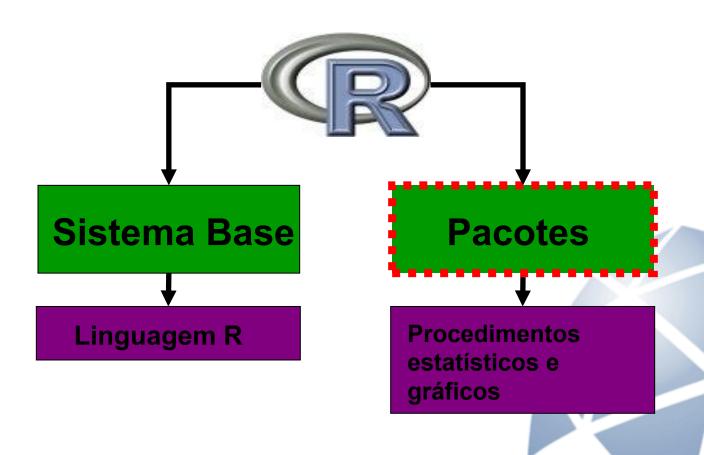
Outros operadores disponíveis:

```
%% # módulo
```

%/% #divisão inteira



"Um <u>pacote</u> é uma coleção de <u>funções, exemplos e documentação</u>. A funcionalidade de um pacote é frequentemente <u>focada</u> em uma <u>metodologia</u> <u>estatística especial</u>" (Everitt & Hothorn).



Os procedimentos básicos para a instalação de pacotes são:

- I) Download a partir do CRAN e instalação manual
- a) Entrar no site: http://www.r-project.org
- b) Selecione a opção CRAN
- c) Selecione o *CRAN* do Brasil
- d) No menu da esquerda em Software, selecione Packages
- e) Selecione a opção *Table of available packages, sorted by name* (ou *Table of available packages, sorted by date of publication*)
- f) Clique no pacote desejado e na próxima página, selecione para o download a versão específica do pacote, de acordo com o seu Sistema Operacional.
- g) No R, clicar em *Pacotes / Instalar Pacote(s) a partir de arquivos zip locais* e selecione o pacote para instalação.

Os procedimentos básicos para a instalação de pacotes são:

- II) Instalação a partir do Menu do R
- a) No R, clicar em *Pacotes / Instalar Pacote(s)*.
- b) Selecione o *CRAN* do Brasil.
- c) Posteriormente aparecerá uma janela com todos os pacotes disponíveis em ordem alfabética.
- d) Selecione o pacote desejado e clique em *OK*, o R fará o *download* e a instalação automaticamente.

Os procedimentos básicos para a instalação de pacotes são:

III) Instalação por meio de linha de comando

Essa é a opção mais simples e rápida para a instalação de pacotes no R, portanto a mais utilizada entre os usuários.

Sintaxe: > install.packages("nome_do_pacote")

Após a execução do comando é solicitado o *CRAN* onde ele será acessado.

Exemplo:

install.packages("agricolae")

```
#listando os pacotes disponíveis no seu computador
.packages(all.available = TRUE)
#listando os pacotes carregados no seu ambiente R
(.packages())
#instalando um pacote, especificamente, o "sandwich"
install.packages("sandwich")
#carregando o pacote instalado no seu ambiente R
require("sandwich")
#ou
library("sandwich")
#veja a lista atualizada de pacotes carregados
(.packages())
```

Lista de pacotes oficiais:

Na página http://cran.r-project.org/clique no link "Packages".

Help on-line

```
#ajuda sobre um comando específico do R
help (mean) #o pacote o qual esta função pertença deve estar carregado
#ou simplesmente
?mean
#busca informação do argumento em todos os pacotes instalados
help.search("mean")
#ou simplesmente
??mean
#obter ajuda sobre um particular pacote (instalado)
help(package = survival)
#mostrar exemplos de um comando específico do R
example (mean)
```

Help on-line

```
#lista as funções e operações contidas no pacote "base"
  do R
ls("package:base")
#carregar documento informativo adicional de um pacote
vignette("sandwich") #nem todos os pacotes possuem este
  documento
#iniciar ajuda no browser padrão instalado
help.start()
#busca ajuda sobre frases ou palavras-chave na lista de
  arquivos ou documentação do R (<a href="http://search.r-project.org">http://search.r-project.org</a>)
RSiteSearch ("statistical multivariate")
#verificar o status da sessão atual do R
sessionInfo()
```

Help on-line

Para uma documentação eletrônica mais ampla, acesse:

http://cran.r-project.org/manuals.html

Para o principiante é indicado a leitura dos 2 primeiros manuais:

"An Introduction to R" e "R Data Import/Export"

Uma outra fonte de documentação disponível eletronicamente

http://cran.r-project.org/doc/Rnews/

Para acessar a documentação eletrônica contendo as respostas das questões frequentemente perguntadas

Na página http://cran.r-project.org/ clique no link "FAQs".

```
#A free book on probability and statistics with R
install.packages("IPSUR")
library(IPSUR)
read(IPSUR)
#seguir o twitter do R
install.packages("twitteR")
require("twitteR")
Rtweets()
#R via interface
install.packages("Rcmdr")
require(Rcmdr)
```

```
#A free handbook of statistical analysis using R (Everitt & Hothorn
require(HSAUR)
vignette()
vignette("Ch_multiple_linear_regression")
vignette("Ch_principal_components_analysis")
browseURL("http://www.r-project.org")
browseURL("http://www.feis.unesp.br")
```

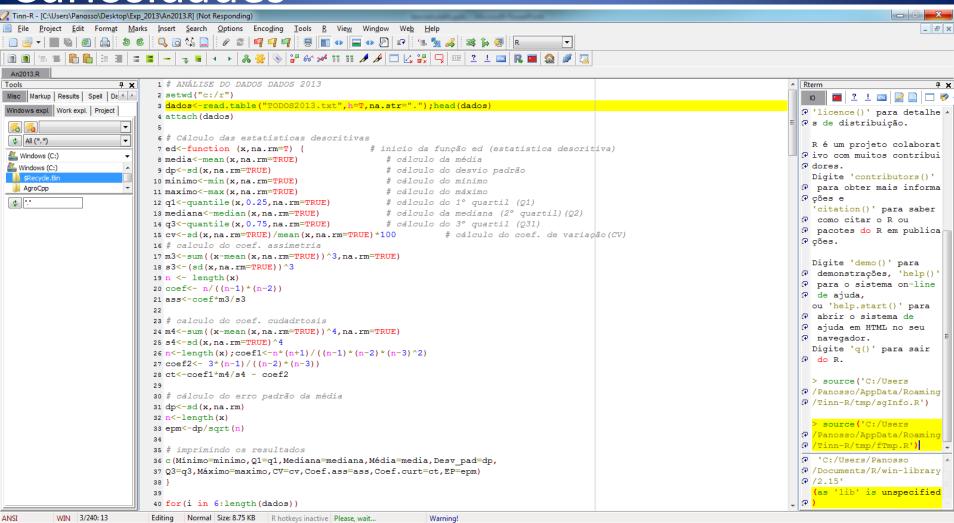
Para usuários Windows recomendo a instalação o Tinn-R ou do RStudio:

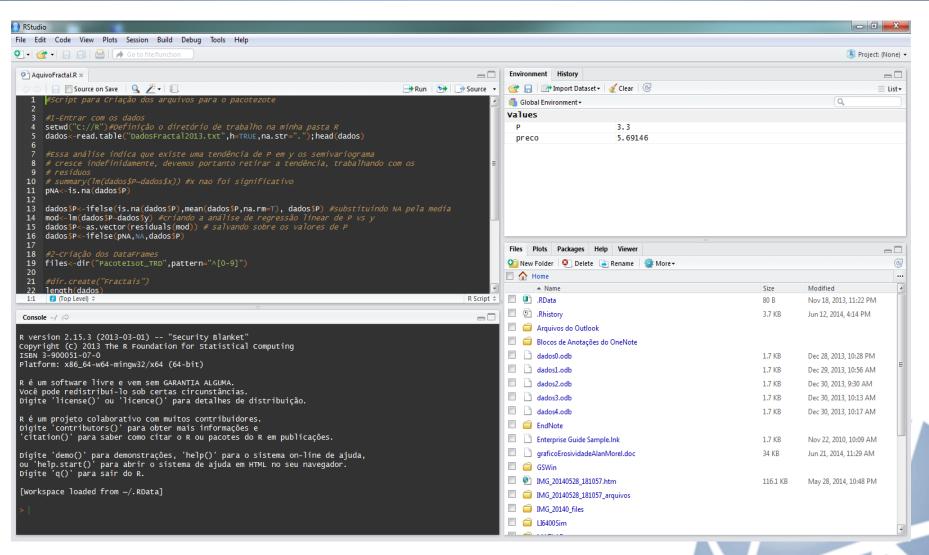
Baixar o executável do Tinn-R ou Rstudio a partir do seguinte link

http://sourceforge.net/projects/tinn-r/

http://download1.rstudio.org/RStudio-0.98.1028.exe

Seguir a instalação recomendada.





Livro Base

"Conhecendo o R. Uma visão estatística."

Autores: Luiz Alexandre Peternelli e Márcio Pupin de Mello

Editora UFV

Publicado em 2007



Operações com escalares

```
#OBS.: A linguagem do R é case-sensitive
x<-5 #o objeto x receberá o valor 5
5->y #o valor 5 será armazenado no objeto y
X=8 #o objeto X receberá o valor 8, não usual
x=sqrt(100) #x receberá o valor da operação indicada
× #011
#ao construir funções o comando print deverá ser
  usado explicitamente
print(x)
ls() #lista todos os objetos existentes na memória
lista = ls() #armazena a lista de objetos
rm(x,y) #remove os objetos x e y
rm(list=ls()) #remove TUDO do projeto ativo!
  CUIDADO!
```

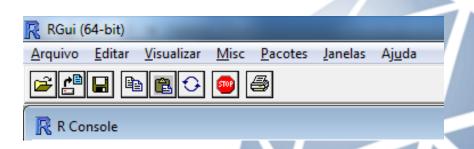
Quando executamos uma análise no R, geralmente, utilizamos várias funções, executando várias linhas de comandos, nesse caso, consideramos útil a utilização de um editor de texto (script) para armazenar os diversos comandos, ao invés de executá-los diretamente no prompt.

O R possui seu próprio editor, que pode ser acessado pela opção *Arquivo / Novo script*. Será apresentada uma nova janela chamada **Editor R**.

Clique na opção *Janelas* e escolha a melhor forma de apresentação das janelas no programa, por exemplo: *Janelas / Dividir Lado a Lado*.

OBS: As opções de *Menu* e os *Ícones* apresentados na interface do R serão alterados de acordo com a janela ativa no programa (**Editor R** ou **R Console**).





No Editor R, poderemos digitar e editar os comandos, para executar uma linha do Script, basta deixarmos o cursor nessa linha e pressionar Ctrl+R ou F5, para executarmos um grupo de linhas basta selecionálas e pressionar Ctrl+R ou F5.

Alternativamente a execução de linhas, ou grupos de linhas, também pode ser realizada por meio do ícone abaixo (ou opção *Editar / Executar linha ou seleção*).



Símbolos e comandos importantes

Serão apresentados os principais símbolos e comandos para o início de nossos trabalhos no R.

Gerais:

Símbolo Significado/ação

Linha de comentário, não será executado pelo

programa.

NA Dado ausente.

NULL Nulo, elemento ou componente inexistente.

Não é um número (exemplo: 0/0).

Inf Infinito (exemplo: 1/0).

-Inf Menos infinito (exemplo: -12/0).

criação de sequências (exemplo: 1:50).

Representa igualdade na formulação de modelos.

Operadores:

aritméticos:

- + Adição.
- Subtração.
- * Multiplicação.
- / Divisão (exemplo: 119/3 R retorna 39.6667).
- %/% Quociente Inteiro da divisão (exemplo: 119%/%3 R retorna 39).
- % Módulo, resto da divisão (exemplo: 119%%3 R retorna 2).
- ^ Potenciação (exemplo: 2^10 retorna 1024)
- ** Potenciação (exemplo: 2**10 retorna 1024).

```
Relacionais:
              Verdadeiro.
T ou TRUE
F ou FALSE
              Falso.
       igual a (exemplo: 225^0.5==15 ).
       menor (exemplo: 4/8<3).
<
       maior (exemplo: 78/56>48/78).
       menor ou igual (exemplo: 1/8<=0.250).
<=
       maior ou igual (exemplo: 1/8>=0.125).
>=
       diferente (exemplo: 0.3!=3/10).
!=
Lógicos:
       Não (!TRUE ou !T)
       e (exemplo: 3==30/10 & 50/0.1==50).
&
       ou (exemplo: 3==30/10 | 50/0.1==50).
```

Funções:

No R todas as funções têm a seguinte forma:

```
função(argumento(s)obrigatórios , argumento(s)opcional(is))
```

Sendo que os argumentos opcionais podem apresentar um valor padrão.

Os argumentos estarão sempre entre parênteses, separados por vírgula. Entretanto, algumas função não terão argumentos.

A atribuição de parâmetros aos argumentos será realizada com o sinal de igual "= ".

Algumas funções do R.

```
Função
                                  Significado/ação
q()
                                  Sair do programa.
                                  Salva o trabalho realizado.
save.image()
                                  Lista todos os objetos da área de trabalho atual.
ls()
                                  Remove o objeto x.
rm(x)
                                  Remove os objetos x e y.
rm(x,y)
rm(list=ls(all=TRUE))
                                  Remove todos os objetos (R Console – Misc/Remover todos os objetos).
                                  Verdadeiro se existir dado(s) ausente(s) no objeto x.
is.na(x)
sqrt(x)
                                  raiz quadrada de x.
log(x,n)
                                  logaritmo de x na base n.
log(x)
                                  logaritmo neperiano de x.
                                  logaritmo decimal de x.
log10(x)
exp(x)
                                  antilogaritmo - ex.
                                  seno de x (em radianos).
sin(x)
asin(x)
                                  arco-seno de x.
abs(x)
                                  modulo(x).
factorial(x)
                                  x!.
floor(x)
                                  maior inteiro < x.
ceiling(x)
                                  menor inteiro >x.
trunc(x)
                                  inteiro de x, descartando seus decimais.
round(x, digits=0)
                                  arredondando o valor x para um inteiro.
signif(x, digits=6)
                                  apresentar 6 dígitos significativos de x.
                                  gera n números aleatórios entre 0 e 1 a partir de uma distribuição uniforme.
runif(n)
                                  concatenação, criação de vetores.
c()
```

```
X = C(1,3,5) #o objeto x recebe a concatenação dos valores numéricos
  1, 3 e 5, ou seja, x é um vetor.
class(x); mode(x); length(x) #classe, tipo e tamanho do objeto
is.numeric(x) #x tem atributo numérico?
# Outros possíveis atributos (tipos). Como, por exemplo:
  vetor, fator, array, matriz, data.frame, ts, lista.
x=pi
round(x,3)#arredondando o valor de pi para 3 casas decimais
trunc(pi)#o valor de pi truncado
trunc(22.98)
signif(pi,3)#o valor de pi truncado
x=0.000256981
signif(x,3)
ceiling(pi)# aproxima para o menor inteiro superior ao
  valor especificado (teto)
floor(pi)# aproxima para o maior inteiro inferior ao
  valor especificado (piso)
```

```
X = 1:10 #cria a sequência crescente de inteiros de 1 a 10
y = 50:30 #cria a sequência decrescente de inteiros de 50 a 30
#outra forma de criar sequências é através do comando "seq"
seq(1,10,1)# o mesmo que 1:10
seq(1,10,2)
seq(10,1,3) #ordem inversa, produzirá ERRO. Veja!
seq(10,1,-3)
#para criar valores repetidos
rep(1,10)#cria um vetor com o valor 1 repetido 10 vezes
rep(c(1,8),10)
c(rep(0,10),rep(1,5))
a=8
rep(c(1,2),a)#repetição de "c(1,2)" a vezes
rep(4:1,1:4)#exemplo interessante
rep(c(10,20,30),c(3,1,2))#para cada valor, um número de
  repetições
```