R: a alternativa ao SPSS e ao NVivo em software livre

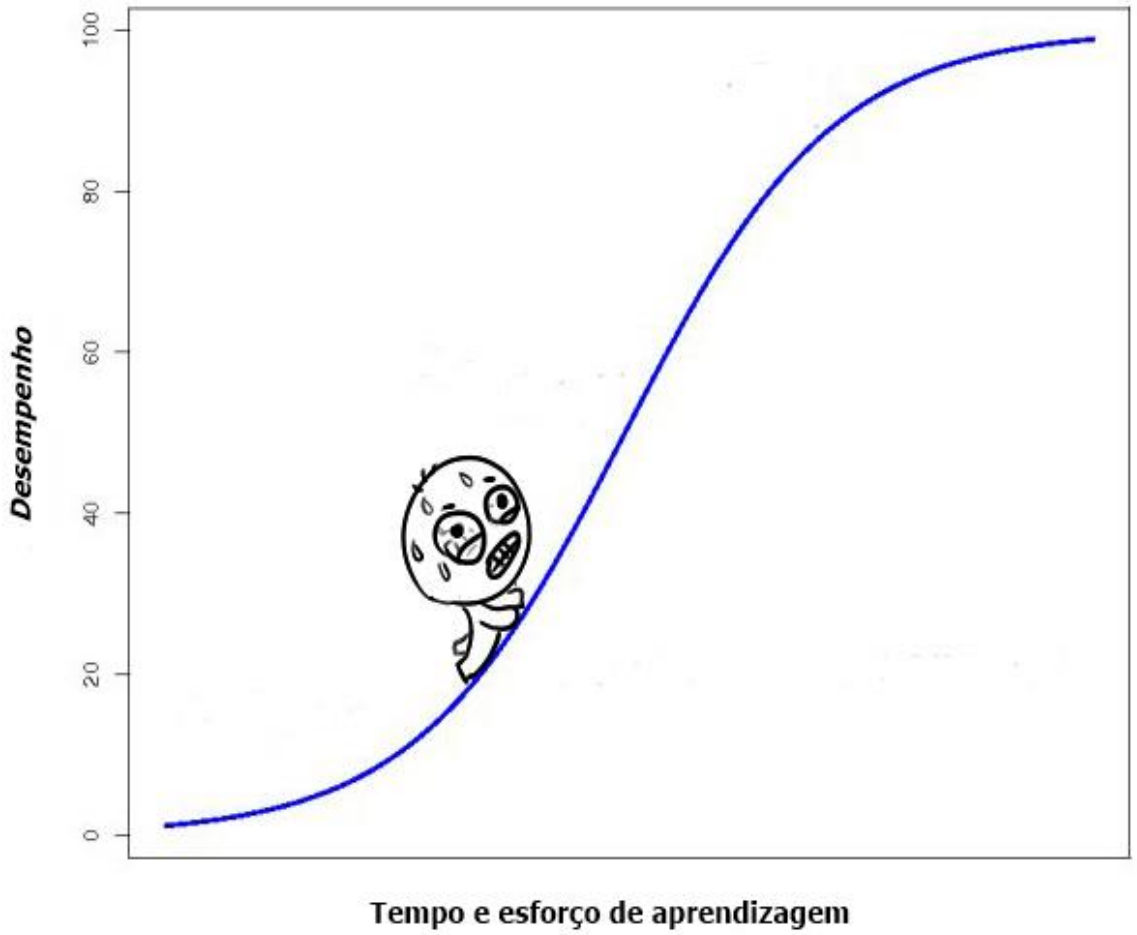
Luis Borges Gouveia
lmbg@ufp.edu.pt
Universidade Fernando Pessoa
Porto, 7 de Dezembro de 2016

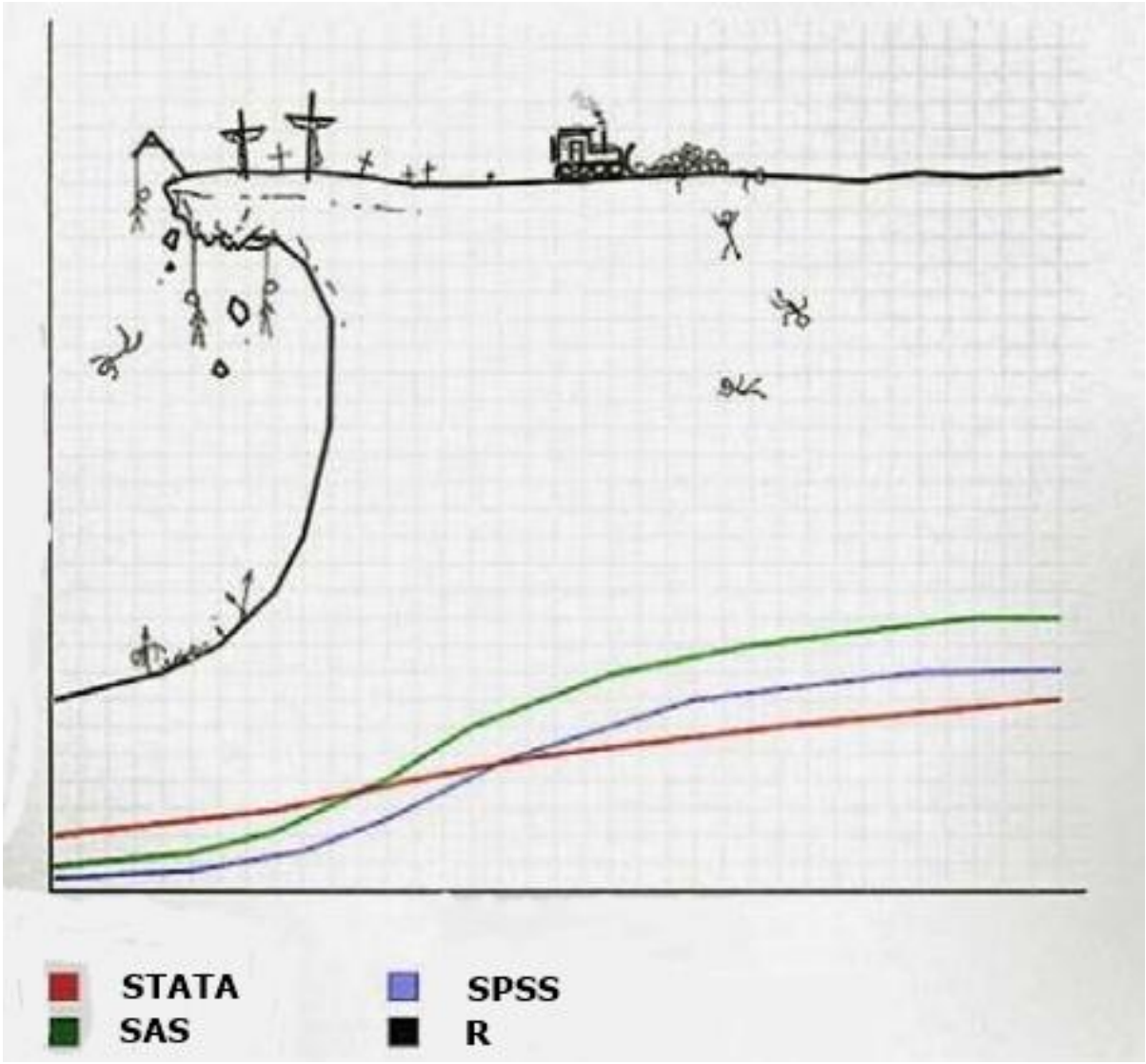
Objetivos da apresentação

- Introduzir o R como um software para análise de dados, estatística, armazenamento e criação de gráficos, que é também uma linguagem de programação
- Introduzir a operação básica do software R e apresentar algumas das suas funcionalidades e potencialidades
- Partilhar um conjunto de recursos que facilitem a aprendizagem e exploração do R
- Apresentar alguns exemplos do uso de R
- Propor uma comunidade de uso local, para explorar o R

Porque esta apresentação?

A curva de aprendizagem do R

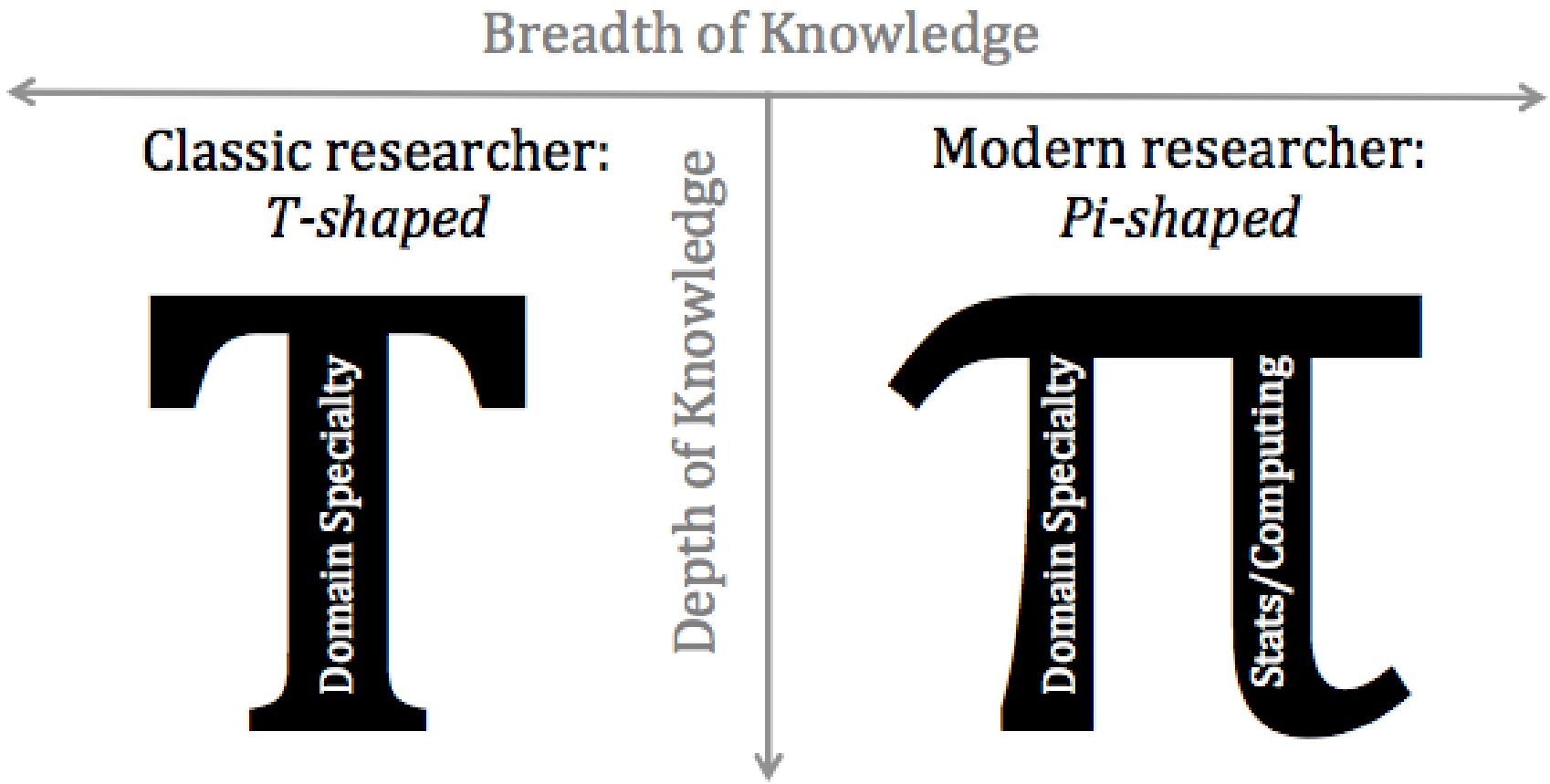




Luis Borges Gouveia, lmbg@ufp.edu.pt

A motivação no contexto da Universidade

extensível a todos os profissionais…



Luis Borges Gouveia, lmbg@ufp.edu.pt

Estrutura da apresentação

- O software R
- Operação básica em R
- Funcionalidades e potencial do R
- Exemplos do uso de R
- Recursos disponíveis sobre o R
- Uma comunidade local de R

Luis Borges Gouveia, lmbg@ufp.edu.pt

O software R

Em que consiste o R?

- O R é uma linguagem de computador interpretada, orientada aos objetos
  - O R, é desenvolvido em R, à exceção de um pequeno conjunto de primitivas internas
  - É possível integrar código em C, C++, FORTRAN ou Python, para maior eficiência ou reutilização de trabalho existente
  - Os comandos de sistema são chamados do interior do R
- O R é utilizado para a manipulação de dados, estatística e geração de gráficos e é constituido por:
  - Operadores: (+ - <- * %, entre outros) para cálculos em vetores e matrizes
  - Oferece múltiplos conjuntos, coerentes e integrados, de funções
  - Possui funcionalidades para produzir gráficos de elevada qualidade
  - Permite funções escritas pelo utilizador e conjuntos de funções (pacotes), com uma extensa lista já existente (quase 10000)

Origem do R

- S: linguagem para a análise de dados desenvolvida nos laboratórios Bell, por volta de 1976
  - Licenciada pela *AT&T/Lucent* à *Insightful, que criou o S-plus*, em 2004
  - Em 2008, a TIBCO adquiriu a *Insightful* (*http://www.tibco.com/*)
- R: proposto como software livre, por Ross Ihaka (1954) e Robert Gentleman (1959) na Universidade de Auckland (*https://www.auckland.ac.nz*, Nova Zelândia) em Agosto de 1993 (o R está surge da letra inicial do nome dos seus autores e é também a letra anterior a S…)
  - R foi inspirado na linguagem S e no LISP
  - Desde 1997 o núcleo internacional do R (cerca de 20 pessoas mas uma comunidade de milhares de programadores) transformou o R num software de estatística e análise de dados dos mais sofisticados e completos

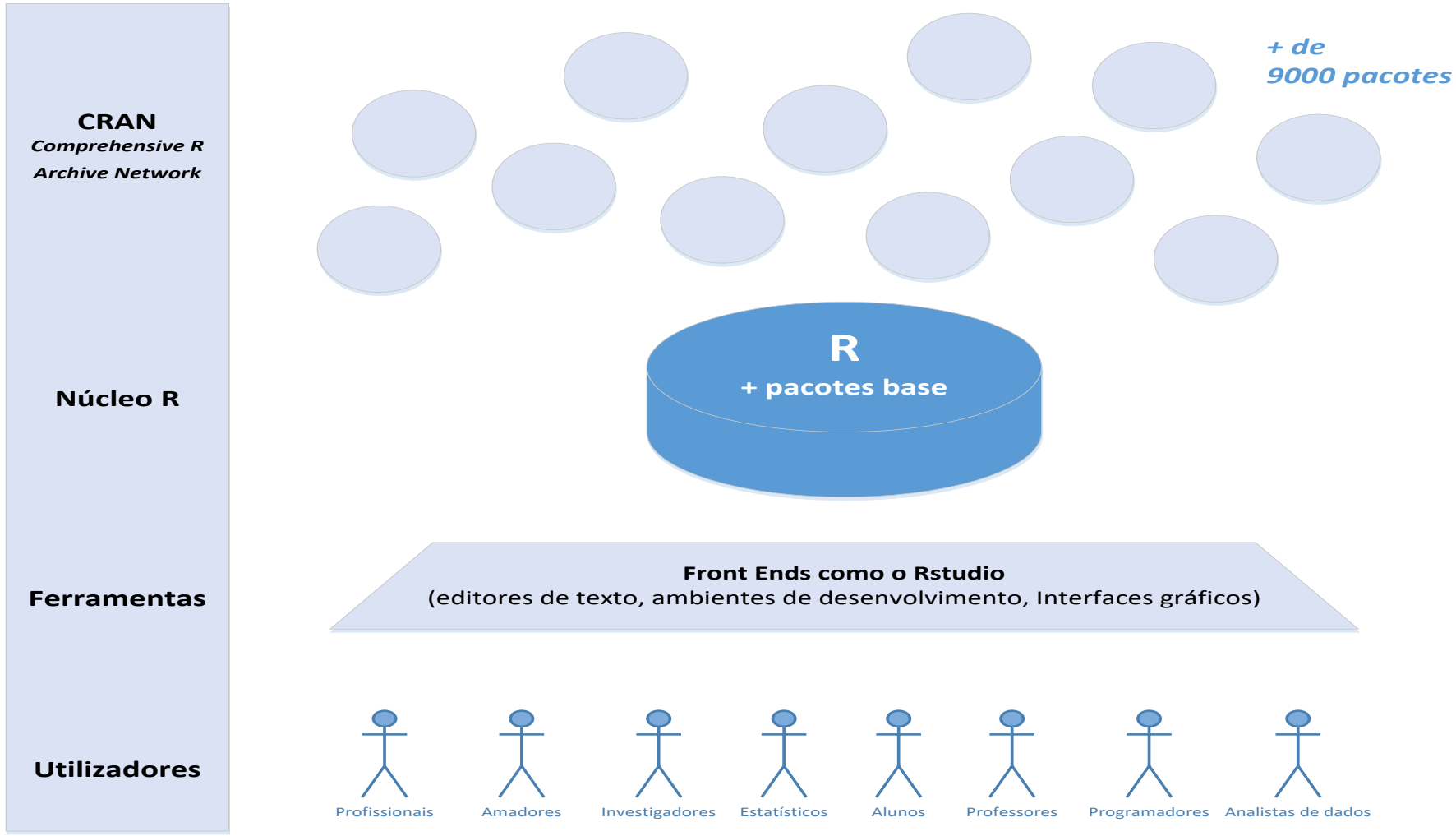
O significado do software livre (*open source*)

- Pode ser obtido sem custos associados à sua aquisição
- Benefícios do R ser *open source*
  - Possibilita o acesso total aos algoritmos e à sua implementação
  - Permite a correção de falhas e a extensão do software existente
  - Proporciona um contexto alargado para explorar e expandir os métodos de análise de dados, estatísticas e tipos de gráficos
  - Torna o produto do esforço colaborativo de milhares de pessoas, à escala global, nas áreas onde são especialistas, proporcionando o estado da arte na análise de dados
  - Possibilita o acesso e posse de software avançado, independentemente do país de origem e dos recursos existentes pelo contexto de quem utiliza o R
  - Promove a investigação reprodutível ao disponibilizar ferramentas acessíveis e abertas
  - Os manuais de muitos dos pacotes do R são verdadeiras introduções operacionais às técnicas que lhes estão asociadas, constituindo um excelente veículo para o ensino e aprendizagem de estatística e da análise de dados

Vantagens e desvantagens do R

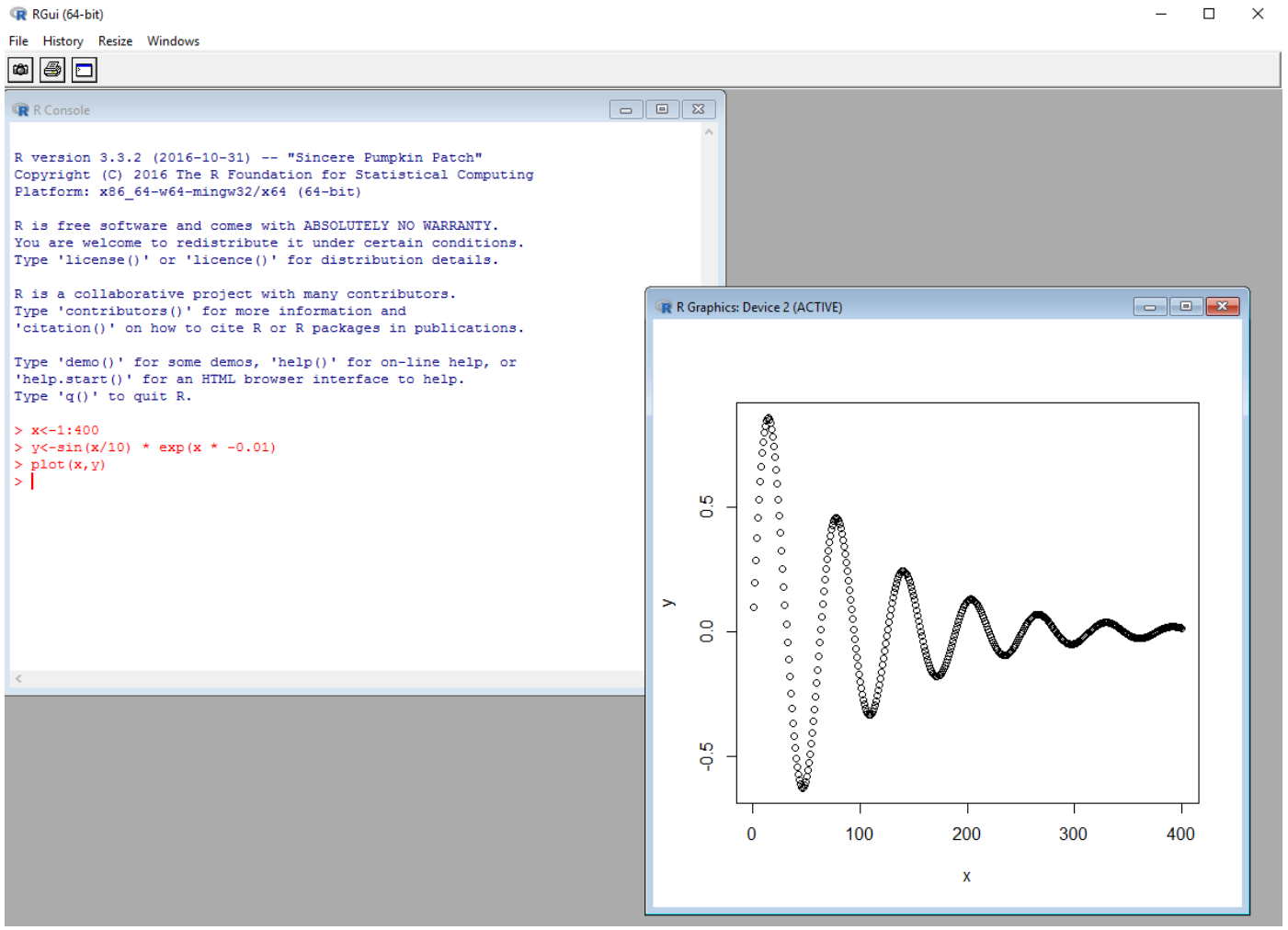
Vantagens	Desvantagens
Rápido e livre	Não é fácil de aprender
Estado da arte em estatística	Não existe suporte comercial a quem recorrer
Gráficos evoluídos	Fácil de errar e não perceber o erro
Comunidade ativa de utilizadores	Trabalhar com conjuntos de dados grandes limitada pela memória principal do computador
Excelente para simulação	A limpeza e arranjo dos dados é mais dificil que em programas comerciais
Força quem usa o R a pensar sobre a análise	
Possui ligação com bases de dados	

A arquitetura do R



Luis Borges Gouveia, lmbg@ufp.edu.pt

O Interface do R

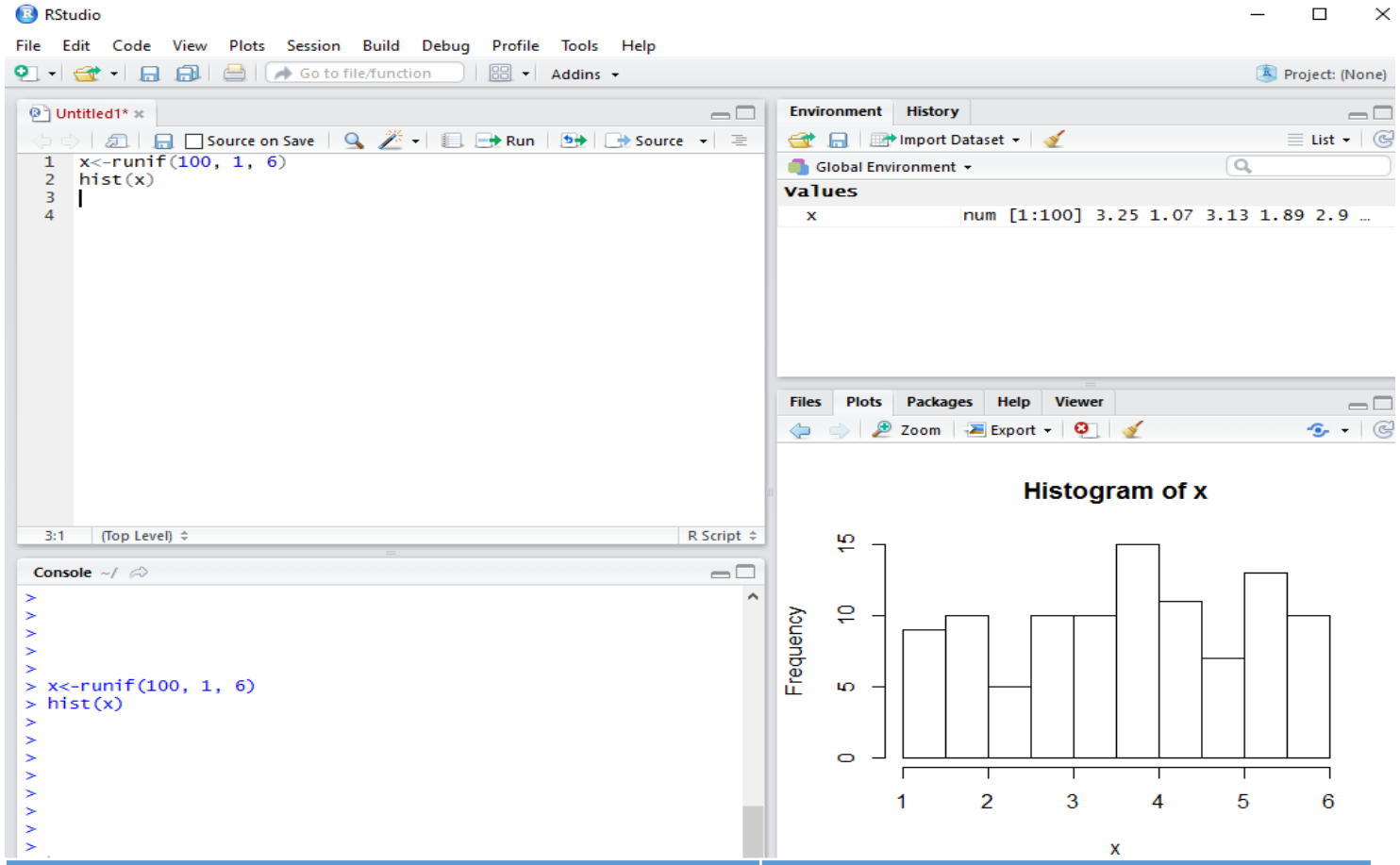


Interfaces de desenvolvimento e suporte

o R-Studio como ambiente de trabalho

Muitos editores e ambientes que facilitam o desenvolvimento em R

- o mais conhecido e usado é o R-Studio
  - *https://www.rstudio.com/*
- Eclipse com StaET for R
  - *http://www.walware.de/goto/statet*
- ESS, *Emacs Speaks statistics*
  - *http://ess.r-project.org/*
- Togaware
  - *http://rattle.togaware.com/*
- Rgedit
  - *http://rgedit.sourceforge.net/*



Editor de código	Ambiente (objetos) e histórico
Consola do R	Ficheiros, Gráficos, Pacotes e Ajuda

Luis Borges Gouveia, lmbg@ufp.edu.pt

Operação básica em R

Numa sessão R

1. Ler os dados de outras fontes
2. Utilizar pacotes, bibliotecas e funções
3. Escrever funções e código, quando necessário
4. Conduzir uma análise de dados estatística
5. Produzir os gráficos necessários para evidenciar os resultados
6. Gravar os resultados para ficheiros, as tabelas de dados e os gráficos gerados
7. Gravar o espaço de trabalho R, se necessário

Ler os dados: alternativas

- como tabelas, a partir de:
  - ficheiros de texto, txt (delimitados por *tab*), csv (com informação separada por vírgulas)
  - diretamente dos conjuntos de dados, nos sites Web
  - via *clipboard* (*copy* e *paste*) ou introduzido por teclado
- dados em formato texto que podem ser lidos de:
  - páginas Web (*webscraping*)
  - ficheiros de texto
  - ficheiros em formato pdf (*portable document format*)
- dados de som e de imagem que podem ser lidos e processados

Objetos em R

- A maior parte das coisas no R são objetos
  - Funções, conjuntos de dados, resultados, …
  - Os gráficos são produzidos e não são guardados como objetos
- Um *script* pode ser pensado como um modo para produzir objetos
  - O objetivo é criar um *script* que crie os objetos para obter os resultados estatísticos e os gráficos pretendidos
- Os objetos podem ser classificados de acordo com duas categorias:
  - modo (*mode*): como os objetos que são armazenados no R
    - numérico, inteiro, complexo, lógico, caracter, fator, lista, função
  - classe (*class*) : como os objetos são tratados pelas funções
    - vetor, matriz, lista, data.frame e centenas de outras classes especiais, criadas por funções específicas

Objetos em R

```
> M <- matrix(c(2, 3, 5, 6, 4, 2, 1, 8, 5), nrow=3, ncol=3)
> M
     [,1] [,2] [,3]
[1,]    2    6    1
[2,]    3    4    8
[3,]    5    2    5
```

- O **modo** da matriz M é determinado de forma automática pelos tipos de valores guardados em M, neste caso números inteiros (caso sejam uma mistura de tipos, o modo é lista)
- A **classe** da matriz M pode ser definida por defeito (dependendo de como foi criada) ou de forma explícita pelo utilizador. Podemos verificar a classe de um objeto e modificá-la. A classe determina como as funções vão lidar com M

Um pequeno exercício em R…

- Gerar uma distribuição normal de 100 valores, com uma média de 62 e um desvio padrão de 25
  - x <- rnorm(100, mean=62, sd=25)
- Gerar dados que simulam 20 lançamentos de uma moeda equilibrada
  - x <- sample(1:2,20,TRUE,prob=c(1/2,1/2))
- Gerar dados que simulam 100 lançamentos de um dado equilibrado e com valores de 1 a 6, nas faces
  - x <- sample(1:6,100,TRUE, prob=c(1/6,1/6,1/6, 1/6, 1/6, 1/6))

Exercícios...

1. Qual é o maior valor? $\log \sqrt{\pi}$ ou $\sqrt{\log \pi}$
2. O que é que a função *rep* faz?
3. Crie um vetor contendo 50 vezes o valor 1 e chame ao vetor *grupo2*
4. Utilize um ciclo *for* para calcular os primeiros 50 números de Fibonacci. Armazene estes valores no vetor grupo2 (obs: Fn = F(n-1) + F(n-2), F1 = 1, F2 = 1
5. O que faz a função *table*?
6. Quantos dos 50 números de Fibonacci são divisíveis por 3? (obs: a %% b)
7. Qual é a media dos primeiros 15 números de Fibonacci?

Luis Borges Gouveia, lmbg@ufp.edu.pt

Exercícios... resolução do 1

1. Qual é o maior valor? $\log \sqrt{\pi}$ ou $\sqrt{\log \pi}$

```
> log(sqrt(pi))
[1] 0.5723649
> sqrt(log(pi))
[1] 1.069921
```

Resposta: O valor associado à segunda expressão é maior

Exercícios... resolução do 2 e 3

2. O que é que a função *rep* faz?

?rep

Resposta: repete os elementos num vetor ou lista

3. Crie um vetor contendo 50 vezes o valor 1 e chame ao vetor *grupo2*

```
grupo2 <- rep(1, 50)
grupo2
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[26] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

Exercícios... resolução do 4

4. Utilize um ciclo *for* para calcular os primeiros 50 números de Fibonacci. Armazene estes valores no vetor grupo2 (obs: Fn = F(n-1) + F(n-2), F1 = 1, F2 = 1

```
grupo2 <- rep(1,50)
for(i in 3:50) {
  grupo2[i] <- grupo2[i-1] + grupo2[i-2]
}
```

```
> grupo2
 [1] 1 1 2 3
 [5] 5 8 13 21
 [9] 34 55 89 144
[13] 233 377 610 987
[17] 1597 2584 4181 6765
[21] 10946 17711 28657 46368
[25] 75025 121393 196418 317811
[29] 514229 832040 1346269 2178309
[33] 3524578 5702887 9227465 14930352
[37] 24157817 39088169 63245986 102334155
[41] 165580141 267914296 433494437 701408733
[45] 1134903170 1836311903 2971215073 4807526976
[49] 7778742049 12586269025
```

Exercícios... resolução 5, 6 e 7

5. O que faz a função *table*?

?table

Resposta: usa os fatores de classificação cruzada para a criação de uma tabela de contingência que conta cada combinação dos níveis de fatores

6. Quantos dos 50 números de Fibonacci são divisíveis por 3? (obs: a %% b)

```
> table(grupo2 = grupo2 %% 3 ==0)
grupo2
FALSE  TRUE
   38    12
```

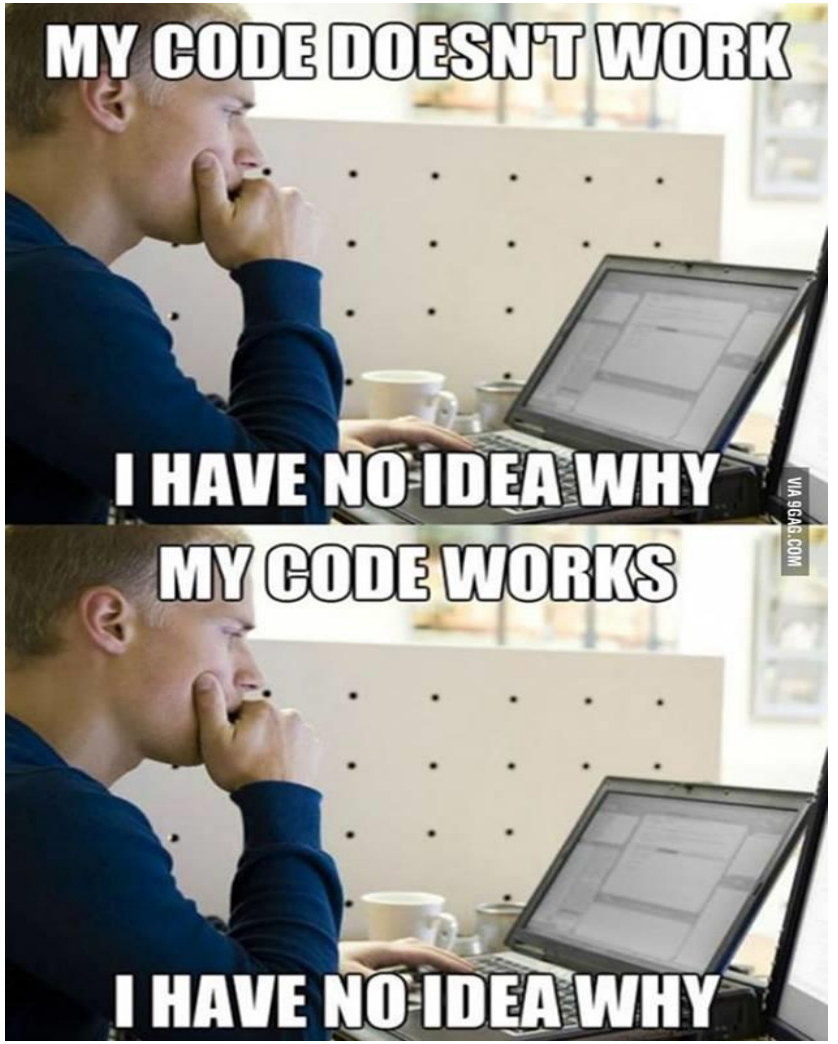
Resposta: até ao quinquagésimo número de Fibonacci, existem 12 valores divisíveis por 3

7. Qual é a media dos primeiros 15 números de Fibonacci?

```
> mean(grupo2[1:15])
[1] 106.4
```

Típico de uma linguagem de programação…
Tempo e experiência

- usar um caderno (*logbook*) para apoio
- recorrer a uma carta de referência da linguagem R: *https://cran.r-project.org/doc/contrib/Short-refcard.pdf*
- consultar índice de referência do R (*R: A Language and Environment for Statistical Computing*), manual com 3518 páginas: *https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf*
- Recorrer a apoios especializados como o catálogo de cores em R: *http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf*



Luis Borges Gouveia, lmbg@ufp.edu.pt

Funcionalidades e potencial do R

Como comparar a oferta de software na área da análise de dados? Como escolher?

Luis Borges Gouveia, lmbg@ufp.edu.pt

Comparar o R com o SAS e o SPSS
(adaptado de Bob Muenchen)

Tópico	SAS	SPSS	R
Modelos avançados	SAS/STAT	SPSS advanced models	stat, MASS, muitos outros
Base (básico)	SAS	SPSS base	R
Análise conjunta	SAS/STAT: Transreg	SPSS Conjoint	homals, psychoR, bayesm
Análise de correspondência	SAS/STAT: Corresp	SPSS Categories	homals, MASS, FactoMineR, ade4, PTAk, cocorresp, vegan, made4, PsychoR
Tabelas personalizadas	SAS Base Report, SQL, Tabulate	SPSS Custom Tables	reshape
Acesso a dados	SAS/ACCESS	SPSS Data Access Pack	DBI, foreign, RODBC
Exploração de dados	Enterprise Miner	Clementine	rattle, arules, FactoMineR
Preparação de dados	Vários procedimentos	Vários procedimentos SPSS Data Preparation	dprep, várias funções

Comparar o R com o SAS e o SPSS
(adaptado de Bob Muenchen)

Tópico	SAS	SPSS	R
Testes exatos	SAS/STAT: vários	SPSS exact tests	coin, elrm, exactLoglinTest, exactmaxsel, exactRankTest, outros
Genética	SAS/ Genetic SAS/ Microaaray Solution, JMP Genomics	Não tem	Bioconductor, com cerca de 200 pacotes associados
Mapas/Sistemas de Informação Geográfica	SAS/GIS SAS/Graph	SPSS Maps	maps, mapdata, mapproj, GRASS via spgrass6, RColorBrewer, outros
Interface gráfico de utilizador	Enterprise Guide	SPSS Base	JGR, R Commander, pmg, Sciviews
Gráficos interativos	SAS/Graph	SPSS Base	ggplot, gplots, graphics, grid, gridBase, hexbin, lattice, plotrix, scatterplot3d, vcd, vioplot, geneplotter, Rgraphics, plotly

Luis Borges Gouveia, lmbg@ufp.edu.pt

Comparar o R com o SAS e o SPSS
(adaptado de Bob Muenchen)

Controlo de qualidade	SAS/QC	SPSS Base	R
Análise orientada	SAS/LAB	Não tem	Não tem
Matrizes, álgebra linear	SAS/LML, SAS/IML Workshop	SPSS Matrix	R, matlab, Matrix, sparseM
Afetação de valores em falta	SAS/STAT, MI	SPSS Missing Values Analysis	aregimpute (Hmisc), EMV, fit.mult.impute (Design), mice, mitools, mvnmle
Investigação operacional	SAS/OR	Não tem	glpk, linprog, LowRankQP, TSP
Análise fatorial	SAS Power e Sample Size Application, SAS/STAT: Power, GLM Power	SamplePower	asypow, powerpkg, pwr, MBESS
Controlo de qualidade	SAS/QC	SPSS Base	qcc, spc
Modelos de regressão	SAS/Base	SPSS Base	R, Hmisc, Design, lasso, VGAM, pda

Comparar o R com o SAS e o SPSS
(adaptado de Bob Muenchen)

Tópico	SAS	SPSS	R
Amostragem, complexidade e sondagens	SAS/Stat, surveymeans, outros	SPSS Complex Samples	pps, sampfling, sampling, spsurvey, survey
Equações estruturais	SAS/Stats, Calis	Amos	Lavaan, sem, OpenMX
Análise de texto	Text Miner	SPSS Text Analysis for Surveys, Text Miner for Clementine	Rstem, lsa, tm
Séries temporais	SAS/ETS	SPSS Trends Expert Modeler	ts, astsa, vários
Séries temporais automáticas	SAS Forecast Studio	SPSS trends, DecisionTime / Whatif	forecast
Árvores, decisão ou regressão	Enterprise Miner	SPSS Classification Trees, AnswerTree	ada, adabag, BayesTree, boost, GAMboost, gbev, gbm, maptree, vários

Análise qualitativa em R

Pacote	URL	Observações
RQDA	http://rqda.r-forge.r-project.org/	O mais conhecido
tm	http://tm.r-forge.r-project.org/	Orientado para a exploração de dados (*text mining*)
QCA, QCA3	https://cran.r-project.org/web/packages/	Análise qualitativa por comparação
CNA	https://cran.r-project.org/web/packages/cna	Análise de coincidências
qdap	https://github.com/trinker/qdap	Análise de discurso (misto qualitativo e quantitativo)
ca	http://www.carme-n.org/	Análise de correspondência
textreg	https://cran.r-project.org/web/packages/textreg/	Síntese comparativa
Aquad	http://www.aquad.de/en	Análise qualitativa, com integração com R (programa autónomo)
sentiment	https://github.com/timjurka/sentiment	Análise do sentimento (emoção) positiva ou negativa

Obter e tratar dados das redes sociais

Pacote	URL	Observações
twitteR	https://cran.r-project.org/web/packages/twitteR/index.html	Obter dados da rede social Twitter
RSQLite	https://github.com/rstats-db/RSQLite	Permite embeber a base de dados SQLite no R
Rfacebook	https://github.com/pablobarbera/Rfacebook	Obter dados da rede social Facebook
igraph	http://igraph.org/r/	Análise de redes sociais (SNA, *Social Network Analysis*)
sna	http://www.statnet.org/	Análise de redes sociais (SNA, *Social Network Analysis*)
ndtv	https://github.com/statnet/ndtv	Análise de redes dinâmicas
xml	http://www.omegahat.net/RSXML/	Obter dados da World Wide Web
SocialMediaLab	https://cran.r-project.org/web/packages/SocialMediaLab/index.html	Ferramentas para construir redes e recolher dados das redes sociais

A multiplicidade e abrangência das soluções em R pacotes: *https://cran.r-project.org/web/packages/*

- A maior vantagem do R: novas técnicas e algoritmos ficam disponíveis para explorar por todos e sem custos acrescidos (modular e atual)
- Permite o desenvolvimento e a personalização de um programa estatístico para contextos específicos (flexibilidade)
- Desvantagem: à medida que o número de pacotes cresce, torna-se mais dificil escolher o que melhor se adequa às necessidades e até mesmo ter conhecimento da sua existência e tempo para aprender a explorar a oferta existente
- Desenvolver pacotes:
  - devtools (*https://www.rstudio.com/products/rpackages/devtools/*): ferramentas de desenvolvimento
  - testhat (*https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf*): teste do código
  - roxygen2 (*https://www.rdocumentation.org/packages/roxygen2/versions/5.0.1?*): documentação e criação de manuais a partir do código
  - Um livro sobre o desenvolvimento de pacotes para R: R packages, *http://r-pkgs.had.co.nz/*

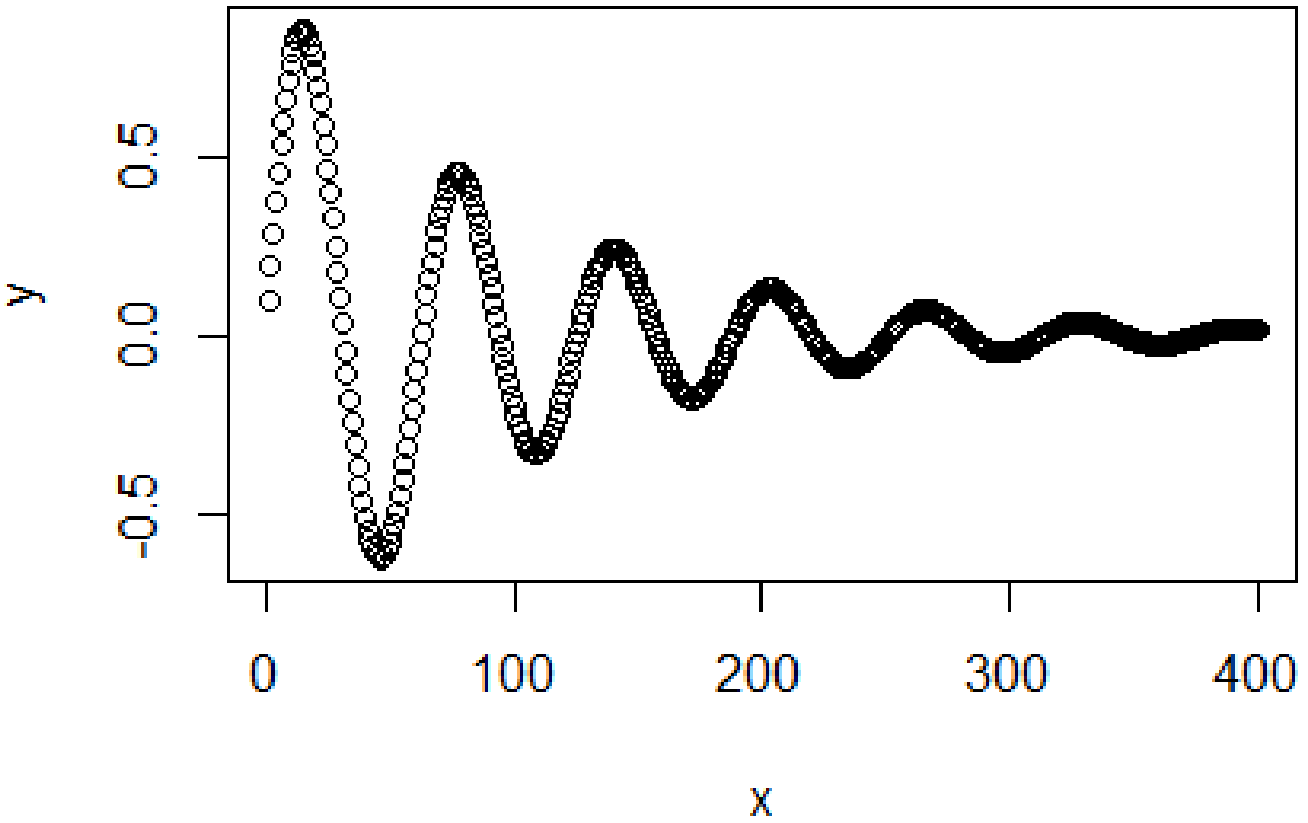
Luis Borges Gouveia, lmbg@ufp.edu.pt

Exemplos do uso de R

Uso de gráficos em R

- Um exemplo de mapeamento de uma função em 2D

```
x<-1:400
y<-sin(x/10) * exp(x * -0.01)
plot(x,y)
```

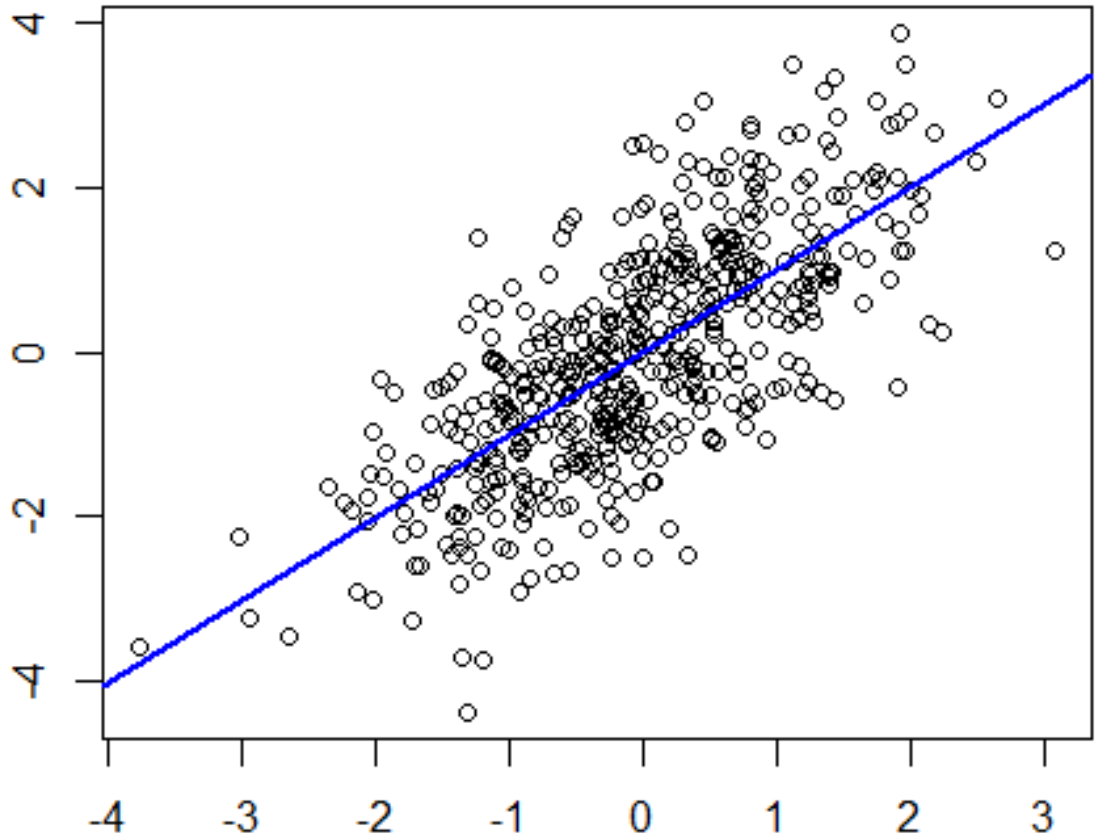


Luis Borges Gouveia, lmbg@ufp.edu.pt

Uso de gráficos em R

- Um exemplo de um gráfico de pontos com visualização da linha de regressão entre as variáveis X e Y

```
n <- 500
x <- rnorm(n)
y <- x + rnorm(n)
plot(x, y)
abline( lm(y ~ x), col = "blue ", lwd=2)
```



Luis Borges Gouveia, lmbg@ufp.edu.pt

Uso de gráficos em R

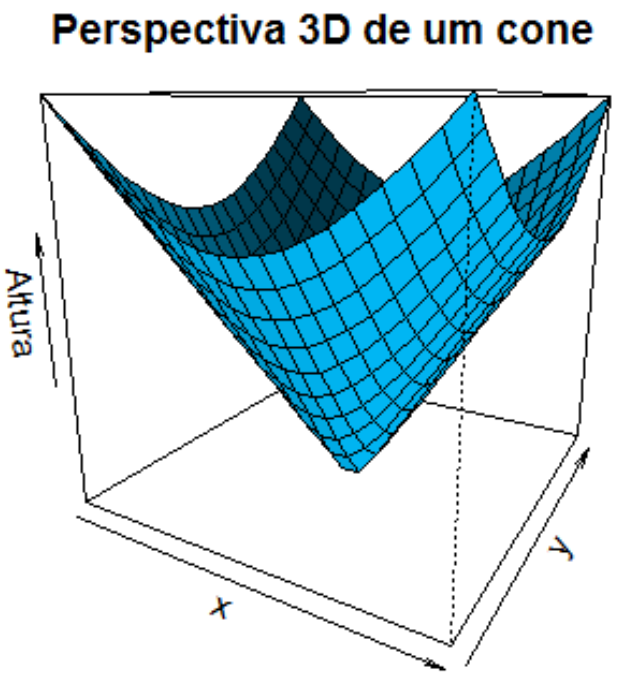
- Um exemplo de mapeamento de uma função em 3D

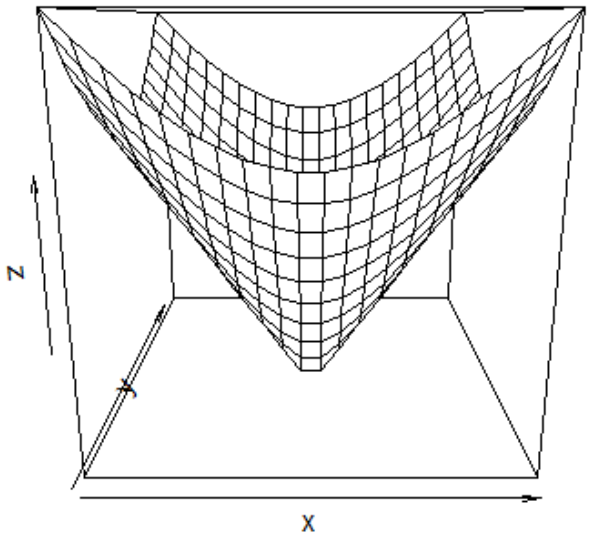
```
cone <- function(x, y){
  sqrt(x^2+y^2)
}

x <- y <- seq(-1, 1, length= 20)
z <- outer(x, y, cone)

persp(x, y, z)

persp(x, y, z,
      main="Perspectiva 3D de um cone",
      zlab = "Altura",
      theta = 30, phi = 15,
      col = "deepskyblue1", shade = 0.5)
```



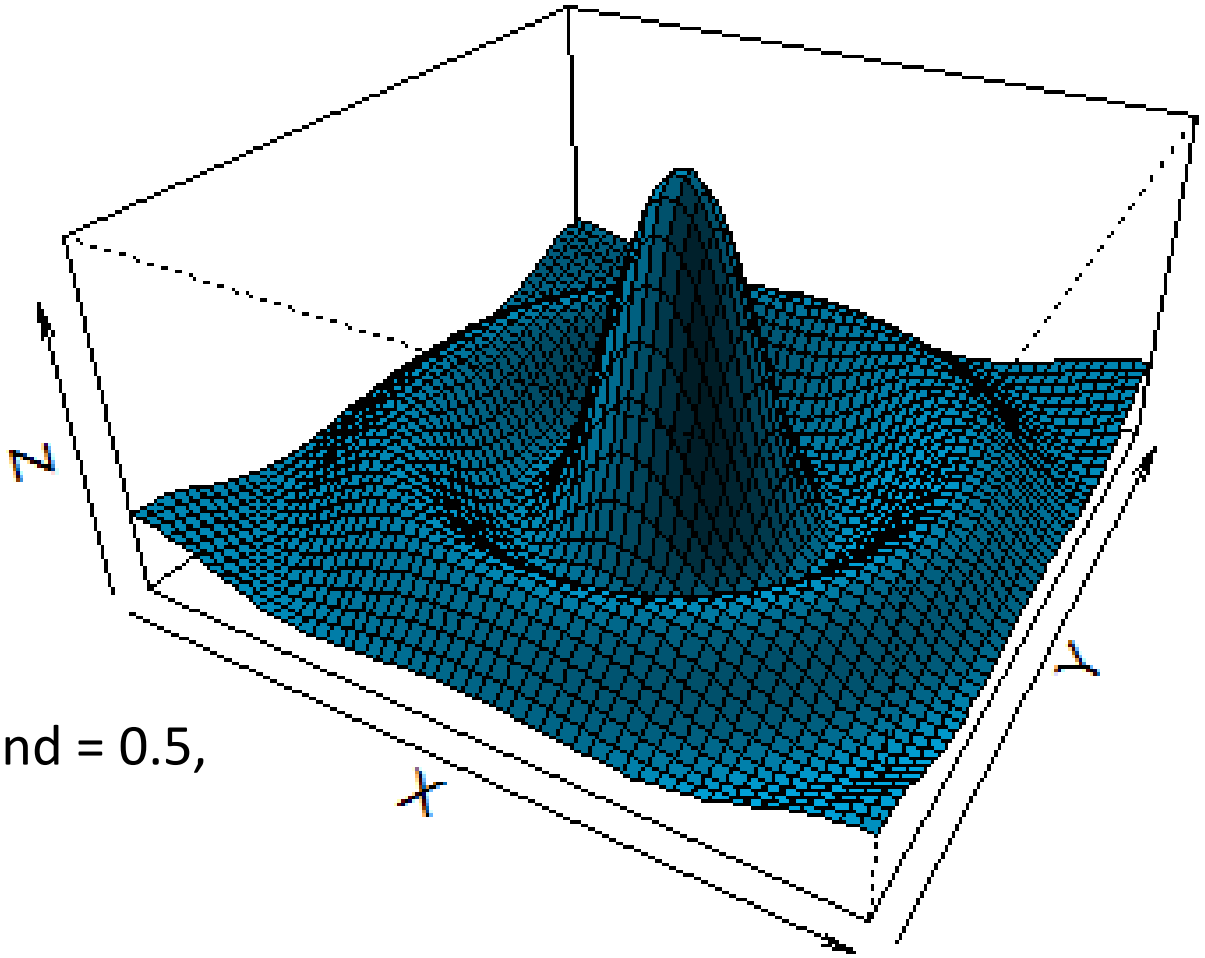


Luis Borges Gouveia, lmbg@ufp.edu.pt

Uso de gráficos em R

- Um clássico…

```
x <- seq(-10, 10, length=50)
y <- x
f <- function(x,y) {
  r <- sqrt(x^2+y^2)
  10 * sin(r)/r
}
z <- outer(x, y, f)
z[is.na(z)] <- 1
persp(x, y, z, theta = 30, phi = 30, expand = 0.5,
      col = “deepskyblue",
      shade=.5,
      xlab = "X", ylab = "Y", zlab = "Z")
```



Luis Borges Gouveia, lmbg@ufp.edu.pt

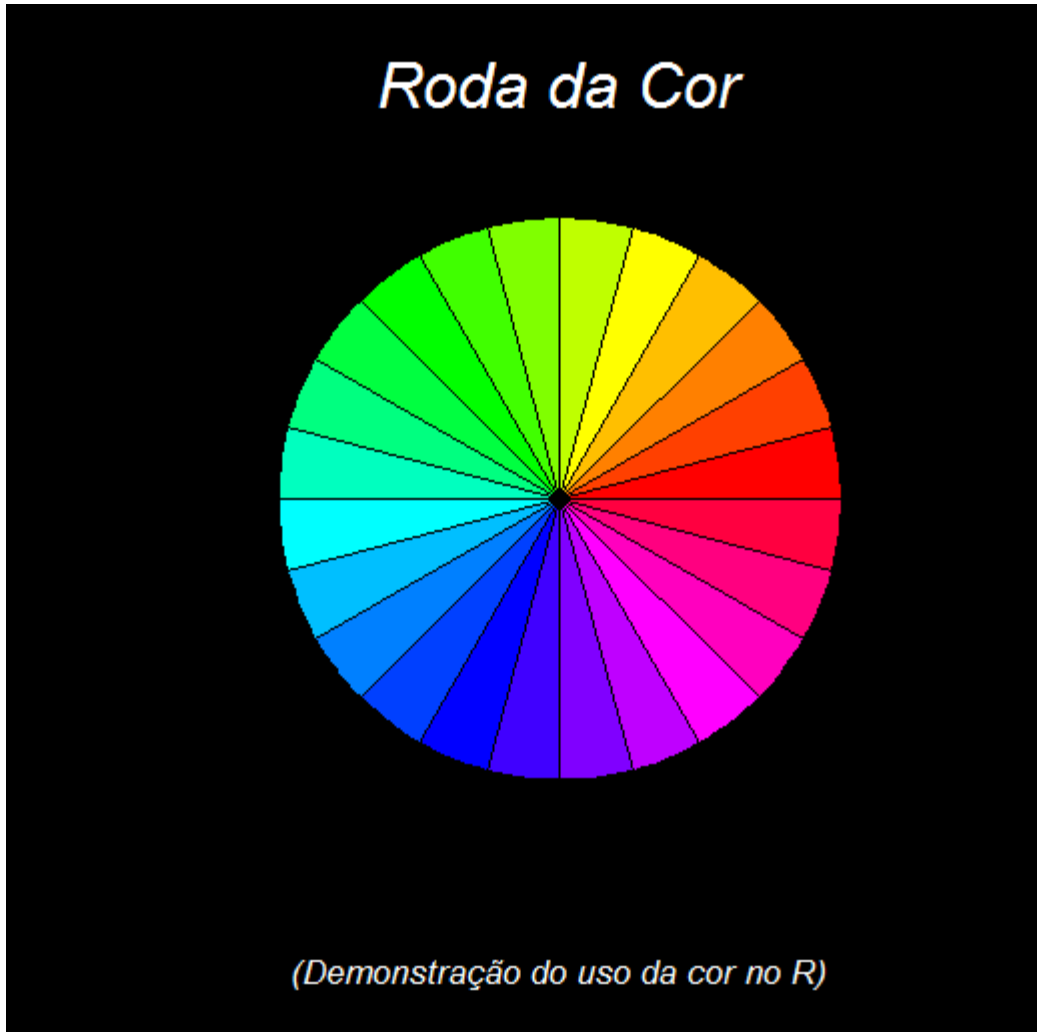
O uso de cor

```
par(bg = "black", col.main= "white ", col.lab= " white")
pie(rep(1,24), col = rainbow(24), radius = 0.9)
title(main = "Roda da Cor", cex.main = 2.0, font.main = 3)
title(xlab = "(Demonstração do uso da cor no R)",
                cex.lab = 1.1, font.lab = 3)
```

Este exemplo foi retirado de *demo(graphics)*

Os códigos de cor no R, estão disponíveis num catálogo com os nomes e cores, em
http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf

Para ajudar a saber os códigos Hexadecimais de cor:
http://www.color-hex.com/

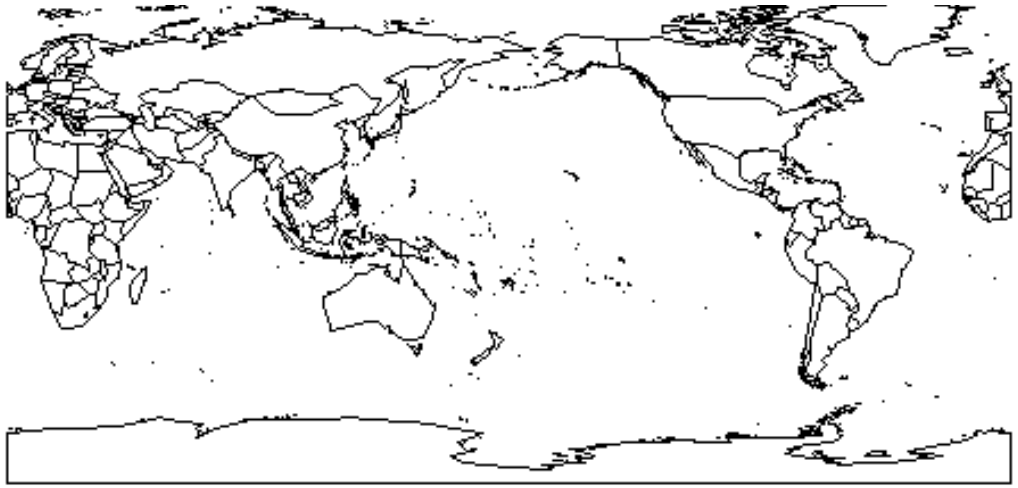


Luis Borges Gouveia, lmbg@ufp.edu.pt

Package maps

```
install.packages("maps")
library("maps")
map("world", "Portugal")

map("world")
map("mworld2")
```



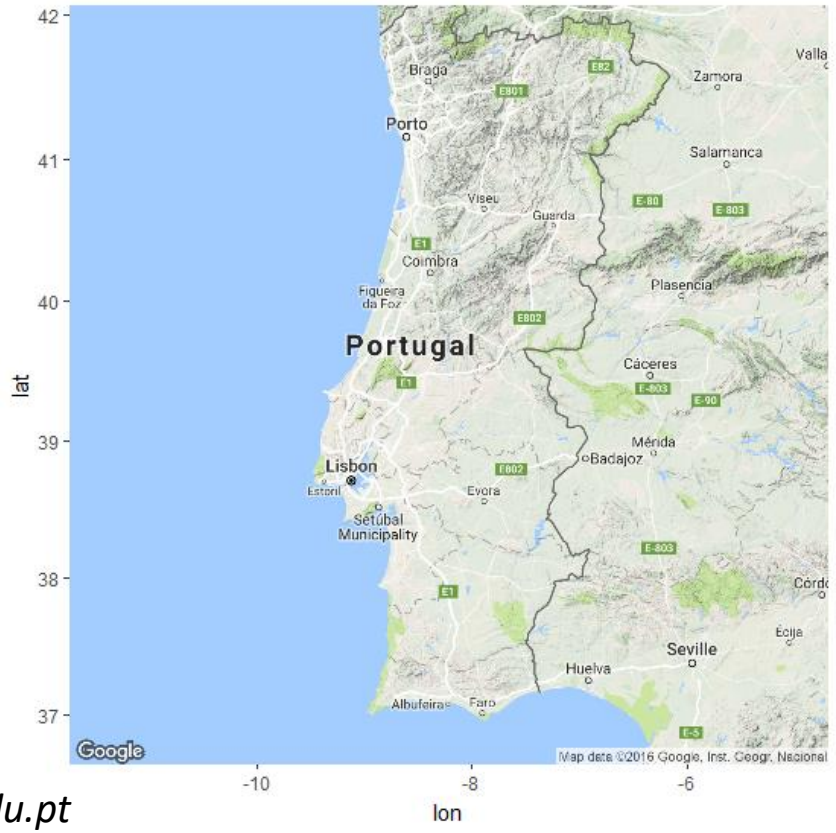
Luis Borges Gouveia, lmbg@ufp.edu.pt

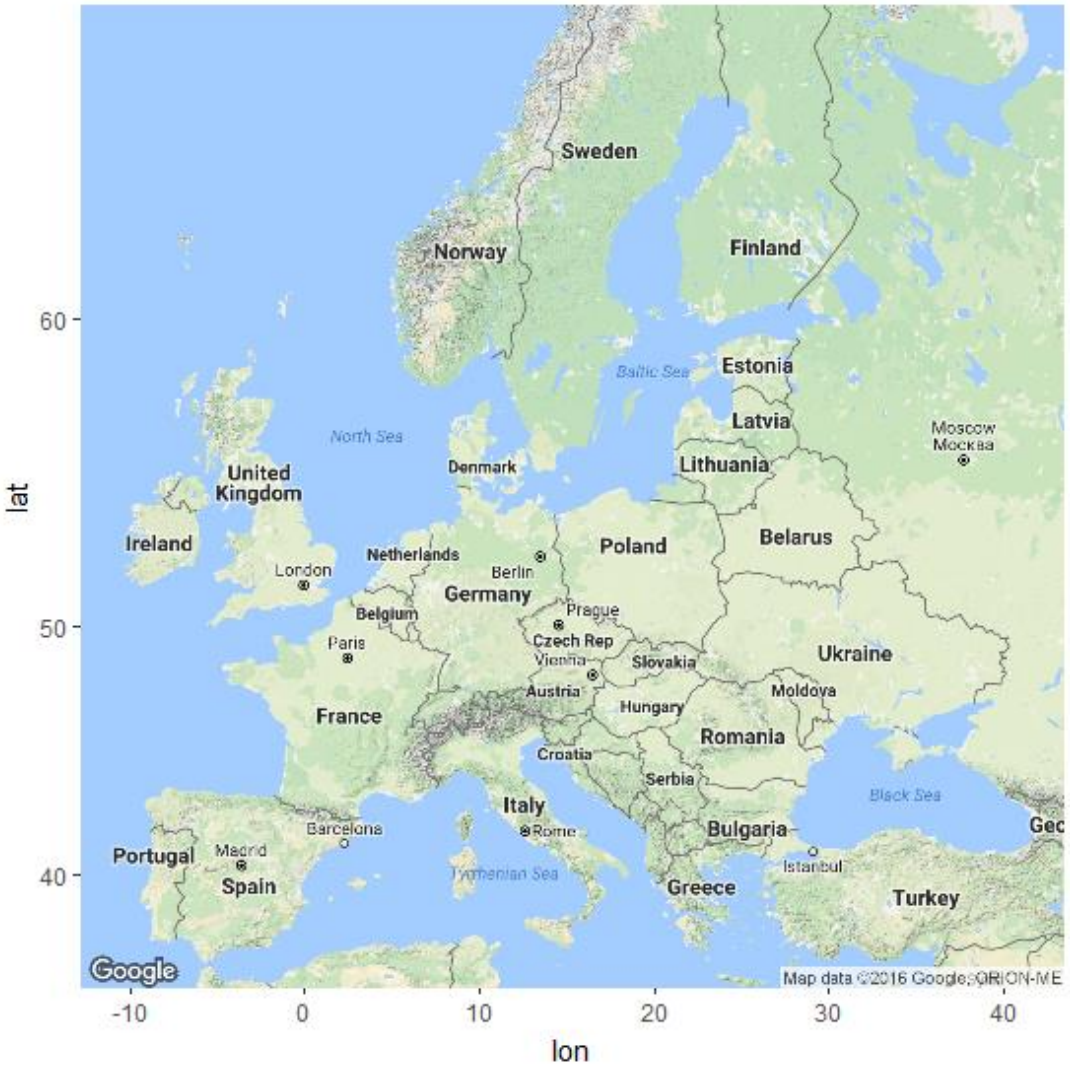
Package ggmap

usa como fonte, o Google Maps

```
library(ggmap)
library(mapproj)
map <- get_map(location = ‘Europe?', zoom = 4)
map2 <- get_map(location = 'Portugal?', zoom = 7)
ggmap(map)

ggmap(map2)
```

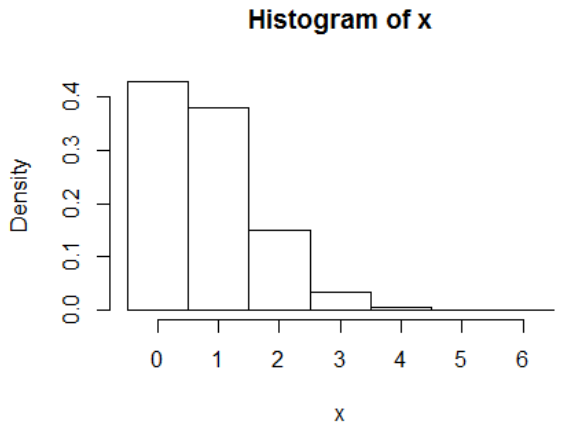


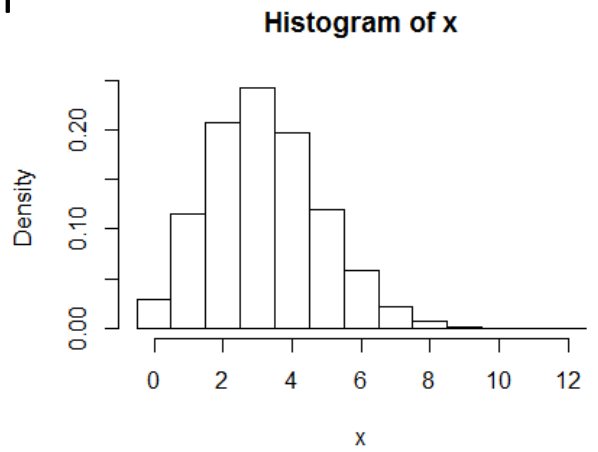


Luis Borges Gouveia, lmbg@ufp.edu.pt

O problema dos aniversários: imagine-se que estão n pessoas numa sala e que pretendemos contar quantas pessoas fazem anos no mesmo dia, considerando a contagem para todos os diferentes dias do ano

```
# Exemplo de simulação: o problema dos aniversários
# m define as iterações da simulação e n, o número de pessoas nas sala
m = 100000; n = 25 # o segundo gráfico, considera n = 50
# vetor para o número de coincidência de datas
x = numeric(m)
for (i in 1:m) {
  b = sample(1:365, n, repl=T)  # número aleatório para o aniversário da iteração i
  x[i] = n - length(unique(b))     # número de coincidências para a iteração i
}
mean(x == 0); mean(x)          # aproxima P{x=0}; E(x)
corteprob = (0:(max(x)+1)) - .5       # pontos de quebra para o histograma
hist(x, breaks=corteprob, prob=T) # frequencia relativa para o histograma
```





Luis Borges Gouveia, lmbg@ufp.edu.pt

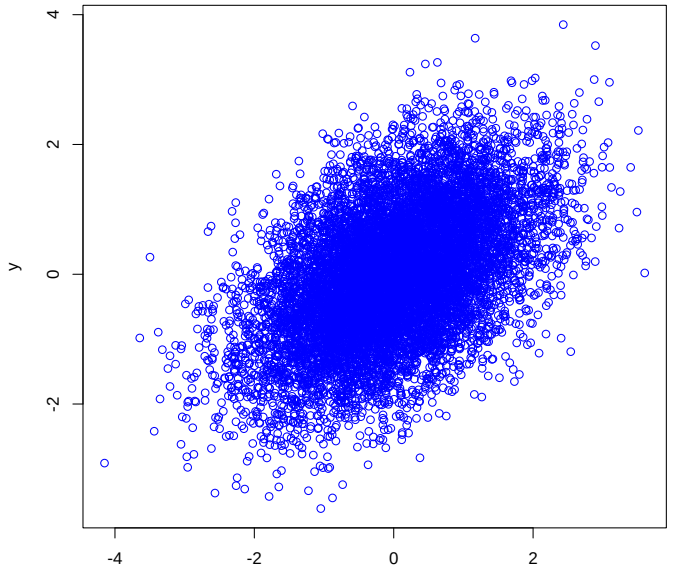
Gerar dados com duas variáveis correlacionadas
(desenvolvimento de funções próprias)

```
cor(x,y)
[1] 0.4945964
```

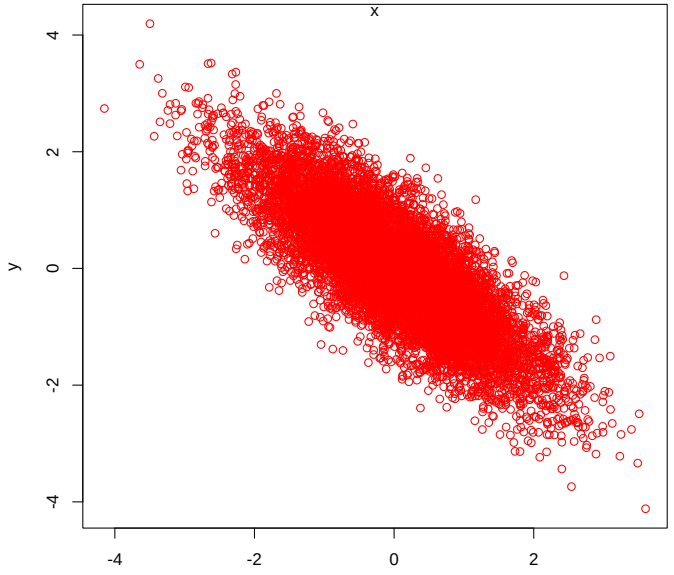
```
vCorrelacionado = function(x, r) {
   r2 = r**2
   ve = 1-r2
   SD = sqrt(ve)
   e  = rnorm(length(x), mean=0, sd=SD)
   y  = r*x + e
   return(y)
}
```

```
cor(x,y)
[1] -0.8029628
```

```
set.seed(5)
x = rnorm(10000)
y = vCorrelacionado (x=x, r=.5)
plot(x,y, col="blue")
```




```
set.seed(5)
x = rnorm(10000)
y = vCorrelacionado (x=x, r=-.8)
plot(x,y, col="red")
```



Luis Borges Gouveia, lmbg@ufp.edu.pt

Exemplos de uso do R em áreas específicas

- utilização do R para investigação em Psicologia: *http://personality-project.org/r/*
- Six Sigma e sistemas de qualidade: *http://www.sixsigmawithr.com/*
- *High-Performance and Parallel Computing with* R (*Big Data*): *https://CRAN.R-project.org/view=HighPerformanceComputing*
- *Web Technologies and Services* (Web e *cloud*): *https://CRAN.R-project.org/view=WebTechnologies*
- CRAN *views*: compilação de pacotes por áreas de interesse e aplicação, *https://cran.r-project.org/web/views/*
  - *Inclui, Química, Testes Clínicos, Equações Diferenciais, Econometria, Design Experimental, Ambiente, Finanças, Genética, Imagem Médica, Processamento de Linguagem Natural, Farmácia, Psicometria, Ciências Sociais, Informação Geográfica*

Recursos disponíveis sobre R

Excesso de informação obriga a escolher: menos é mais

O uso de recursos deve ser progressivo e orientado para explorar e responder a problemas concretos

Luis Borges Gouveia, lmbg@ufp.edu.pt

Existe sempre o (santo) Google…

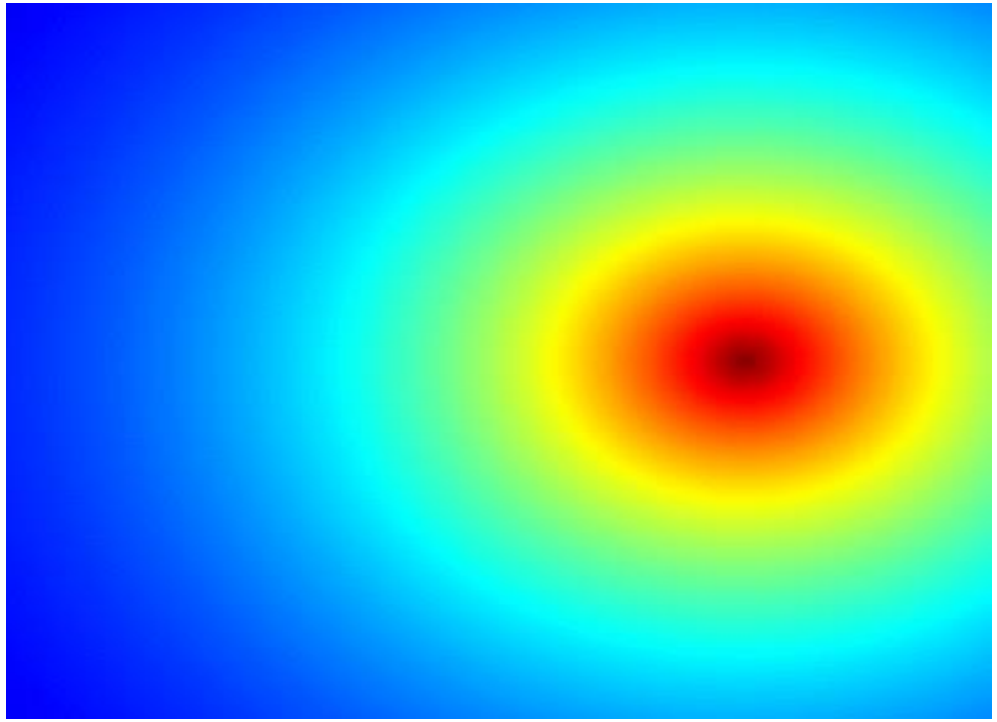
- O uso do Google pode complicar mais, do que ajudar: excesso de informação que exige um esforço considerável para o seu entendimento
- Aplica-se a lei da distância ao conhecimento…
  - ***A Web possui informação útil em função do quadrado do conhecimento de quem o usa e da sua capacidade de fazer as perguntas corretas***

Nem sempre o código funciona (Wikipedia)

https://pt.wikipedia.org/wiki/R_(linguagem_de_programa%C3%A7%C3%A3o)

```
library(caTools)
jet.colors<-colorRampPalette(c("#00007f", "#0000ff", "#007fff", "#00ffff", "#7fff7f", "#ffff00", "#ff7f00", "#ff0000",
"#7f0000"))
m<-1200
C <-complex(real=rep(seq(-1.8, 0.6, length.out=m), each=m), imag=rep(seq(-1.2, 1.2,length.out=m), m))
C<-matrix(C, m, m)
Z<-0
X<-array(0, c(m, m, 50))
for (k in 1:50) {
    Z<-Z^2+C
    X[,,k]<-exp(-abs(Z))
}
write.gif(X, "Mandelbrot.gif", col=jet.colors, delay=100)

# cria 2 matrizes com 1,44 Milhões de elementos (22 MB cada)
# cria um array com 72 Milhões de elementos (549 MB)
# cria um Gif animado com 11 MB
```



Luis Borges Gouveia, lmbg@ufp.edu.pt

7 motores de pesquisa para recursos em R

- **RSeek *http://www.rseek.org/***
- R *Documentation*: *https://www.rdocumentation.org/*
- R *Site Search*: *http://finzi.psych.upenn.edu/search.html*
- *Search the R statistical language*: *http://www.dangoldstein.com/search_r.html*
- **R-Bloggers: *https://www.r-bloggers.com/***
- Nabble R Forum *http://r.789695.n4.nabble.com/*
- R *mailing lists archive*: *http://tolstoy.newcastle.edu.au/R/*

Luis Borges Gouveia, lmbg@ufp.edu.pt

Alguns recursos (consumir com moderação)

- uma lista de exemplos de análise de dados, IDRE, UCLA *http://www.ats.ucla.edu/stat/dae/*
- sobre criação de documentos em R
  - knitr, criação de documentos dinâmicos em R: *http://personality-project.org/r/*
  - Sweave, integrar R em documentos Latex: *http://www.statistik.lmu.de/~leisch/Sweave/*
  - R Markdown, produz documentos com integração de R e outros linguagens: *http://rmarkdown.rstudio.com/*
- Swirl: aprender de forma interativa R e programar em R: *http://swirlstats.com/*
- vídeos de 2 minutos sobre temas do R: *http://www.twotorials.com/*
- (mais) um tutorial sobre R: *http://www.r-tutor.com/*
- um tutorial em Português: *http://ecologia.ib.usp.br/bie5782*

Aprender e usar o R

- Instalar o R (*http://www.r-project.org/*) e o RStudio (*https://www.rstudio.com/*)
- Experimentar, resolver problemas, partilhar e experimentar de novo…
- Explorar o CRAN (*The Comprehensive R Archive Network*): *https://cran.r-project.org/*
- Participar em grupos de discussão, aprender com os outros e voltar a experimentar... O mais completo é o R-Bloggers (*https://www.r-bloggers.com/*)
- Usar o motor de pesquisa RSeek *http://www.rseek.org/* em vez do Google
- Para saber mais dos objetos: class(x)
- Para saber mais das funções: ?x ou help(x)
- Para correr demonstrações ou exemplos de funções: demo(x) ou example(x)
- Como o R é interativo, os erros constituem uma ajuda para o código a criar
- Quanto mais experimentar o R, mais confortável é o seu uso. A experiência adquire-se essencialmente da realização de projetos concretos

Luis Borges Gouveia, lmbg@ufp.edu.pt

Uma comunidade local de R

A criação de uma comunidade local R
vale a pena?

- O R é fácil de instalar e não possui custos associados
- O R possui um potencial de utilização em diferentes contextos e por diferentes utilizadores
- O R exige um esforço de aprendizagem e o conhecimento especializado de conceitos de estatística e da exploração e análise de dados
- Partilhando dúvidas, desafios e soluções, fica mais fácil e rápido a aprendizagem e o aprofundamento do R
- O R pode ser útil em diferentes contextos
  - áreas de conhecimento (com pacotes específicos e especializados)
  - níveis de ensino graduado e pós-graduado
  - em apoio à investigação e exploração de dados quantitativa, qualitativa ou mista

Luis Borges Gouveia, lmbg@ufp.edu.pt

Luís Borges Gouveia

http://homepage.ufp.pt/lmbg/ | lmbg@ufp.edu.pt

- Professor Associado com Agregação da Universidade Fernando Pessoa, Coordenador do Doutoramento em Ciências da Informação, ramo de Sistemas, Tecnologias e Gestão da Informação. Autor de 15 livros. Possui a Agregação em Engenharia e Gestão Industrial pela Universidade de Aveiro e Doutoramento em Ciências da Computação, pela Universidade de Lancaster (UK).
- Nos últimos anos, tem estudado os temas do digital e como este impacta o dia-a-dia da atividade humana e a gestão da informação para organizações e indivíduos. Defensor das plataformas digitais para suporte à cooperação entre pessoas e organizações, aspeto onde também se insere o *open source*.