

Introdução ao R usando RStudio



Objetivos do minicurso

- Apresentar o R usando o ambiente de desenvolvimento (IDE) RStudio
- Identificar seus principais elementos e componentes
- Utilizar algumas das principais funções
- Reconhecer diferentes formas de análises



Conteúdo

- Por que usar o R?
- Baixar e instalar o R, o Rstudio e os principais pacotes
- Tipos de variáveis
- Conjunto de dados nativos e provenientes de pacotes
- Tipos de objetos possíveis de serem criados e explorados no R
 - Vetores
 - Matrizes
 - Data.frames
 - Listas
 - Séries Temporais
 - Fatores

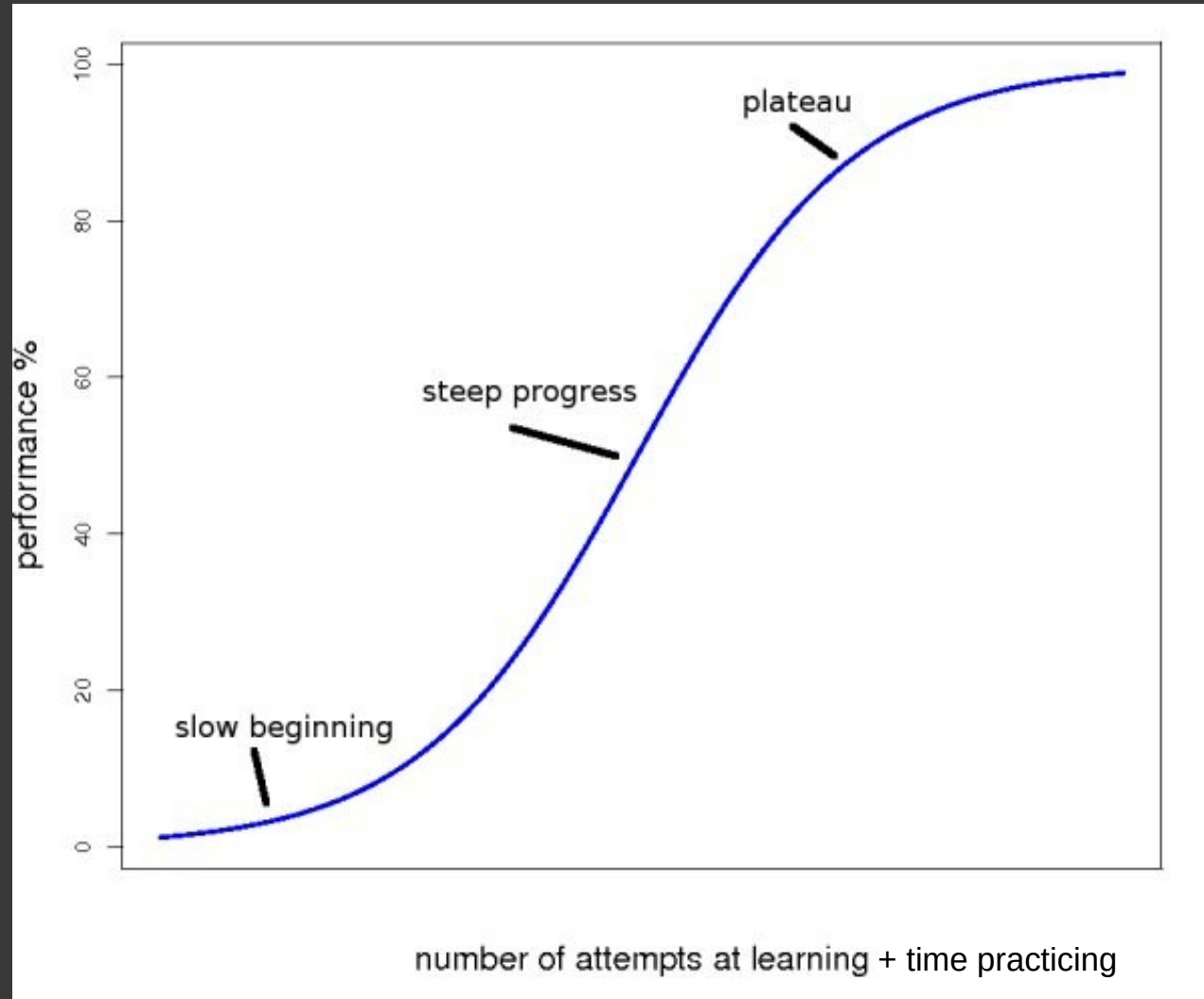
Conteúdo

- Importando dados básicos
- Importando dados genéticos
- Gráficos básicos
- Funções
- Salvando arquivos
- Conversões de tipos de objetos
- Lendo manuais de pacotes
- Rcmdr
- Análises filogenéticas básicas
- Manipulando brevemente filogenias com o ggtree

Por que usar o R?

- Open Source → Gratuito
- Scripts (automatizáveis e replicáveis – GitHub e Dryad)
- Maior grau de liberdade analítica
- Permite integrar diferentes análises (e. g. sistemática filogenética e métodos estatísticos)
- Criação de pacotes e programas próprios e personalizáveis
- Múltiplos recursos e ferramentas em um mesmo software, em constante expansão
- Podem ser igualmente utilizadas em qualquer sistema operacional

Curva de aprendizado

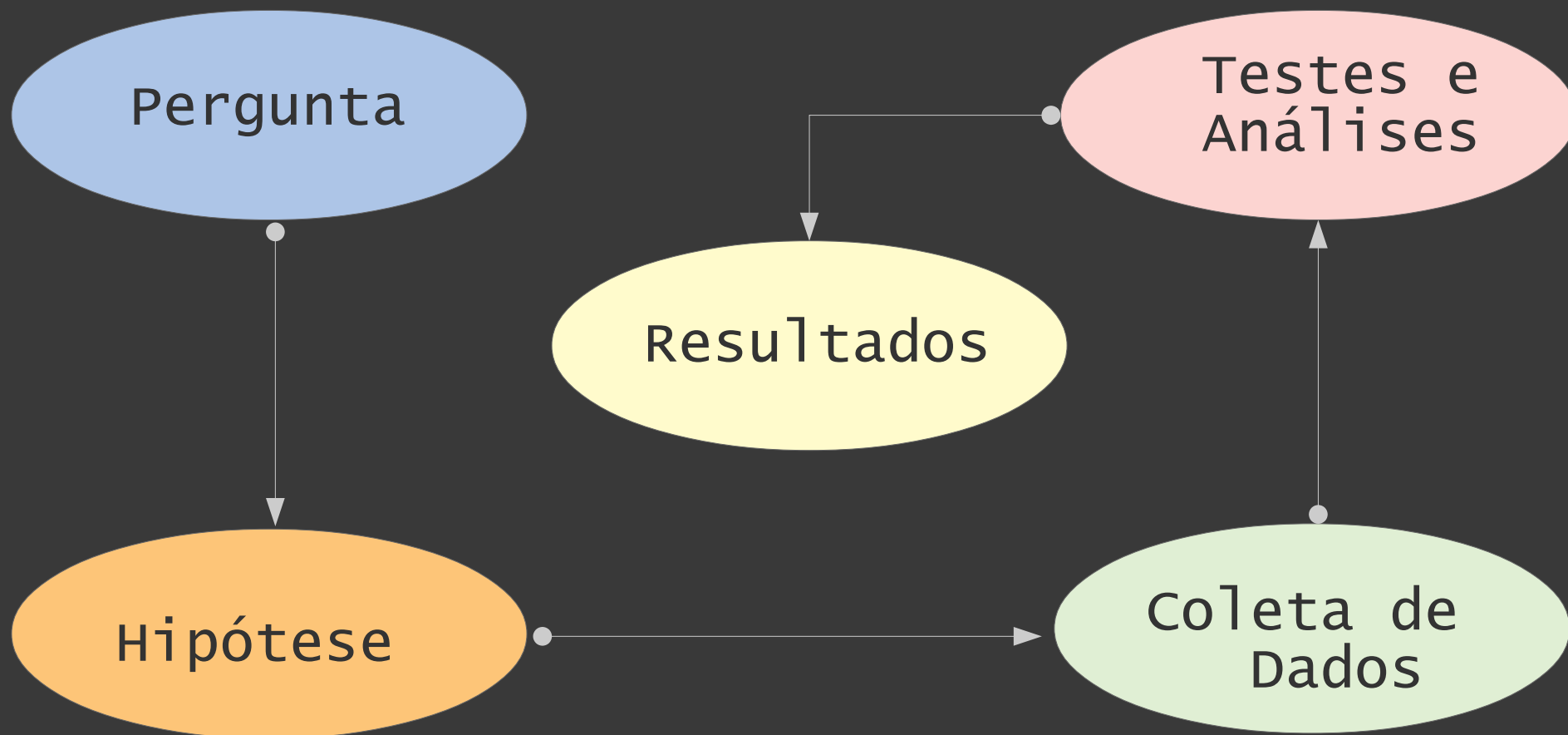


Baixando e instalando o R

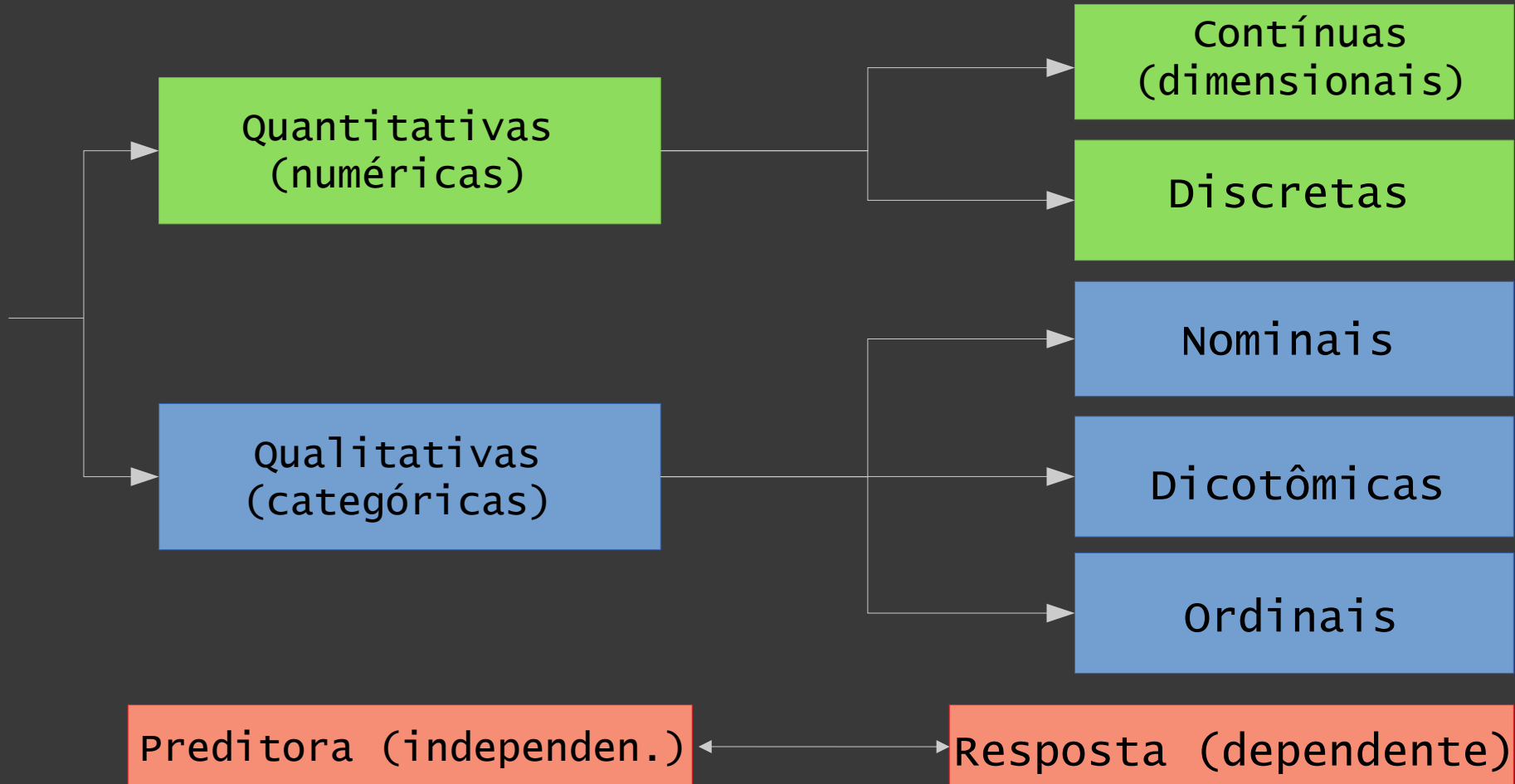
R-Project: www.r-project.org

- Rstudio:
<http://www.rstudio.com/products/rstudio/>
- Instalar alguns pacotes (*vegan*, *ape*, *adegenet*, *phytools*, *phangorn*, *Rcmdr*, *ggtree* e *ggplot2*)

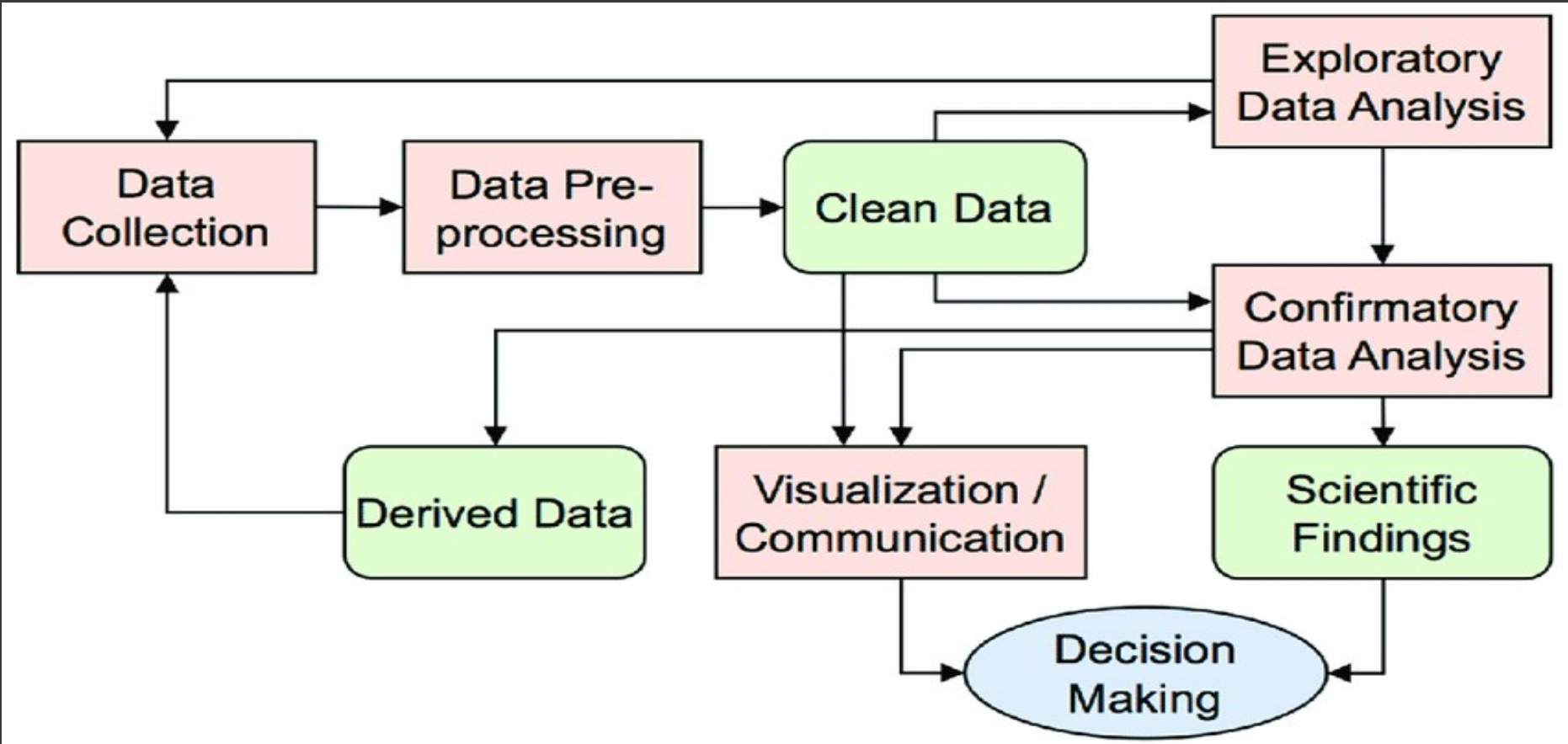
Passo-a-passo do trabalho científico

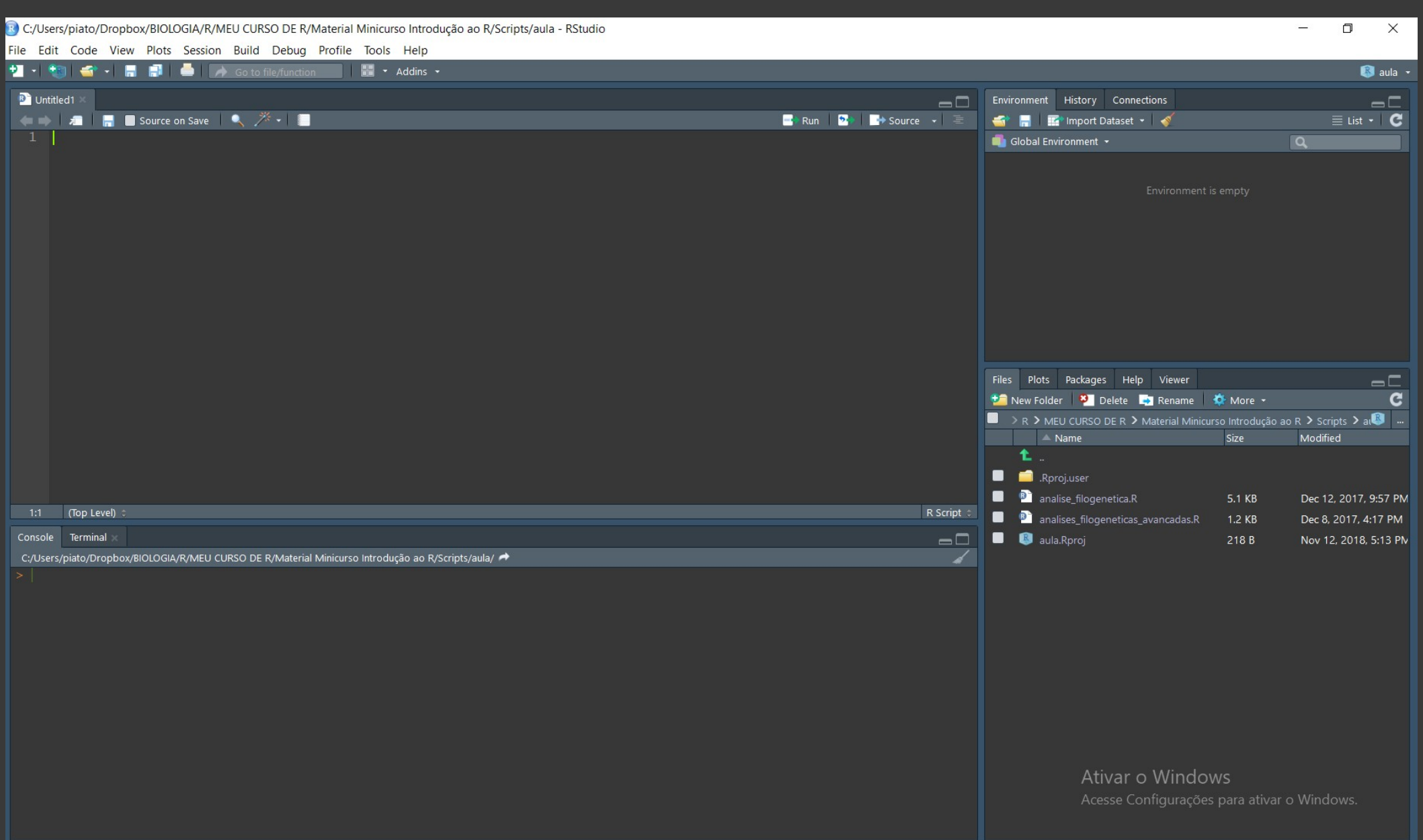


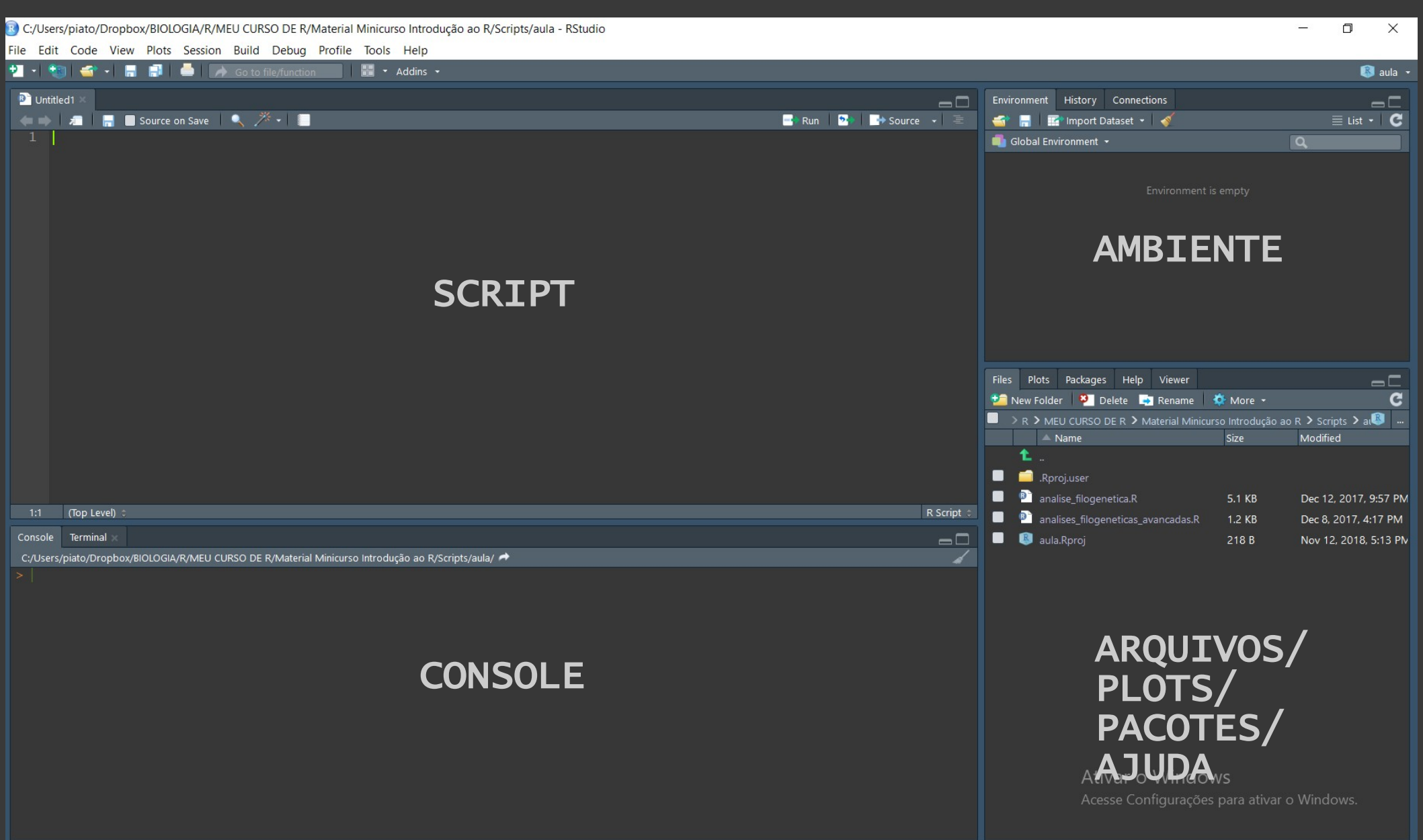
Tipo de variáveis

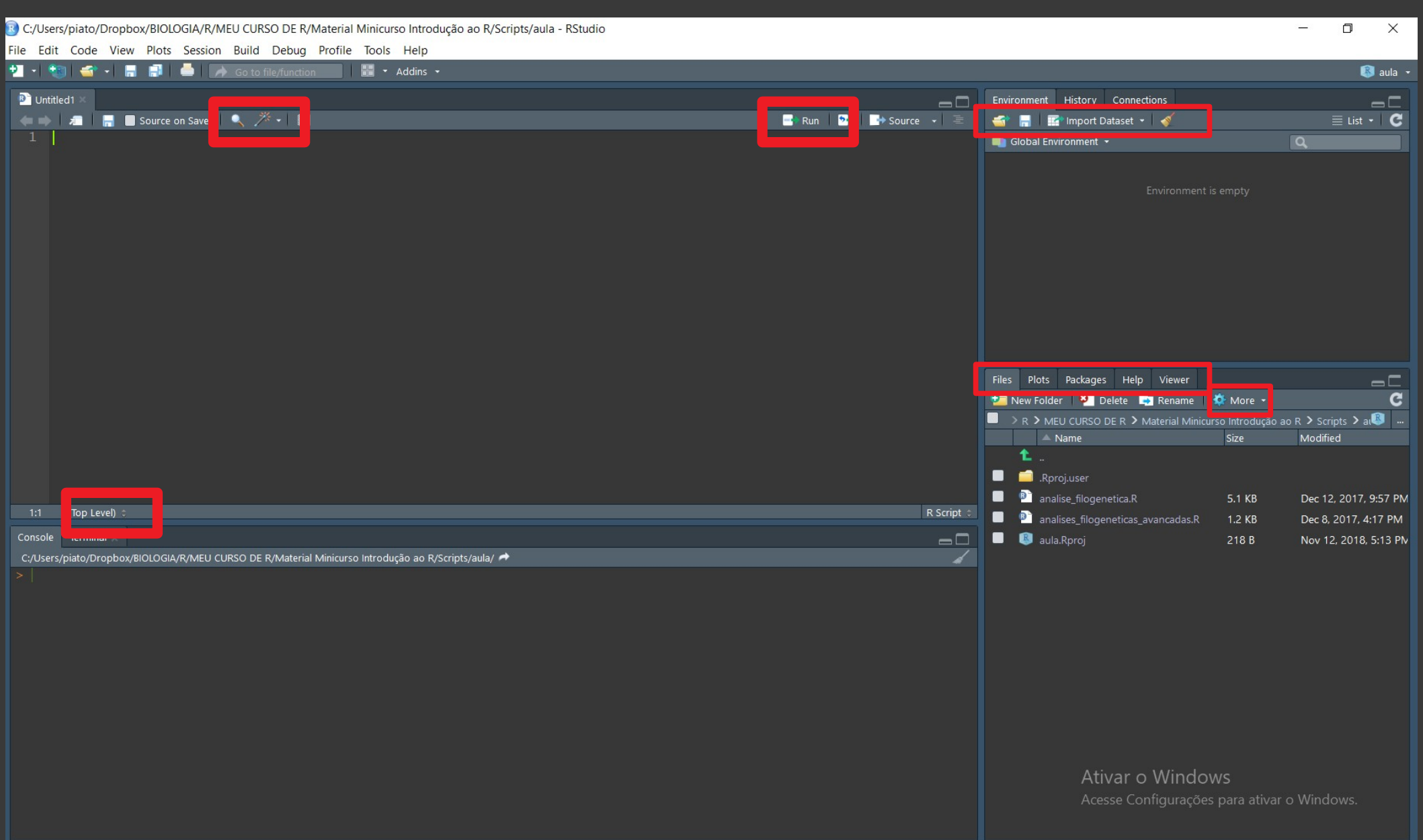


Passo-a-passo da análise de dados









Principais tipos de objetos no R

- Vetores (*vector*)
- Matrizes e arrays (*matrix; array*)
- Data frames (*data.frame*)
- Listas (*list*)
- Séries Temporais (*time series*)
- Fatores (*factor*)
- Função (*function*)
- Todos os objetos têm: comprimento e tipo básico (*mode*)
 - *Character; numeric (integer ou double); logical*

Pacotes

- ♦ Pacotes são conjuntos de funções com determinada finalidade
- ♦ Repositórios:
 - CRAN
 - GitHub
 - DevTools
- ♦ *> install.package("nomedopacote")*

Pacotes

- *vegan, ape, phangorn, phytools, adegenet, Rcmdr, ggtree, ggplot2, dplyr e tidyr*
- *>pacotes <- c("vegan"; "ape";
"phangorn"; "phytools"; "adegenet";
"Rcmdr"; "ggtree"; "ggplot2";
"dplyr"; "tidyr")

>install.package(pacotes)*

Vetores

- São conjuntos de dados unidimensionais (lineares)
- 1,2,3,4,5,6
- a,b,c,d,e,f
- TRUE, TRUE, FALSE, FALSE, FALSE, FALSE
- Uma série de cálculos e manipulações podem ser realizadas com vetores
- Função “combine”: *c(“conjuntodedados”)*

Vetores

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]					
[2,]					
[3,]					
[4,]					
[5,]					

Vetores

Vetor

Vetor

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]					
[2,]					
[3,]					
[4,]					
[5,]					

Matrizes

- Conjuntos de dados bidimensionais do mesmo tipo

	[,1]	[,2]	[,3]
[1,]	a	b	c
[2,]	d	e	f
[3,]	g	h	i

- Função `matrix()`: `matrix(c("a","b","c","d","e","f","g","h","i"), byrow=T, nrow=3, ncol=3)`
- Outras funções: `rbind()` e `cbind()`
- Arrays são conjuntos tridimensionais de dados do mesmo tipo (matrizes sobrepostas)

Data.frames

- Conjuntos de dados bidimensionais com mistura de tipos de dados

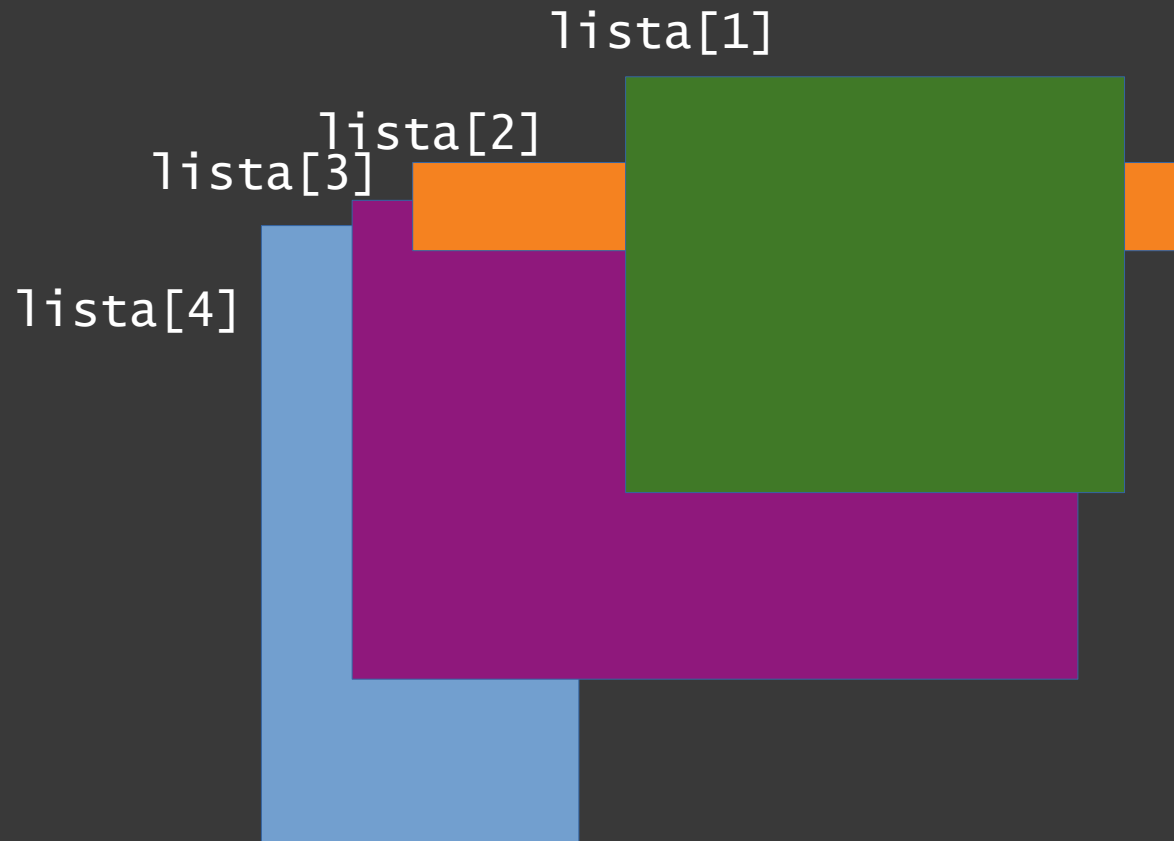
	[,1]	[,2]	[,3]
[1,]	a	1	T
[2,]	b	2	F
[3,]	c	3	T

- Função `data.frame()`: `data.frame(c("a", "b", "c"), 2=c(1,2,3), 3=c(T, F, T))`
- Funções `rbind()` e `cbind()` também são aplicáveis

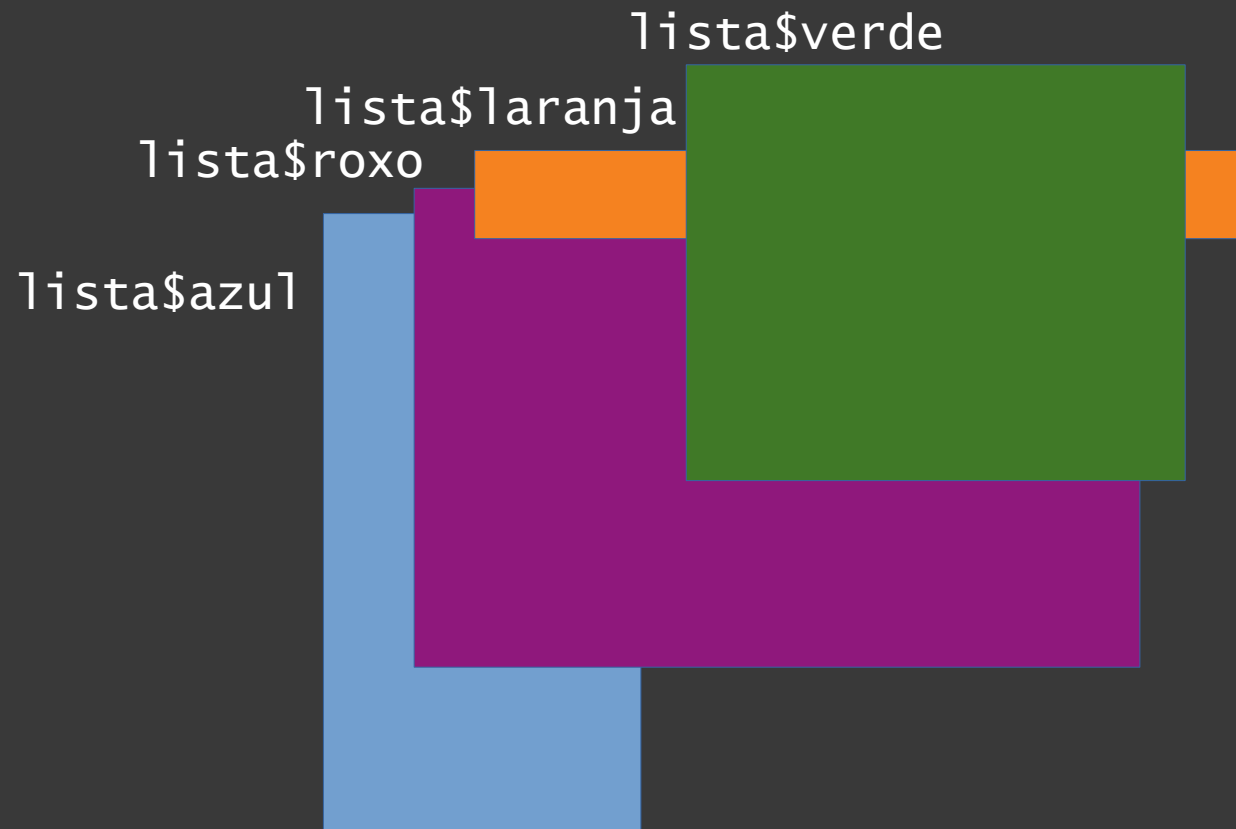
Listas

- Reúnem múltiplos objetos de diferentes classes em um único objeto
- Os diferentes objetos listados podem ser acessados através do comando: `nomedaLista$nomedoobjeto`
- Listas são extremamente úteis e comuns
- Maior parte das análises filogenéticas usam listas
- Função `list()`: `list("e"=e, "f"=f, "iris"=iris, "q"=q, "r"=r)`

Listas



Listas



Séries temporais

- São conjuntos de dados temporalmente ordenados
- São extremamente similares a matrizes, mas têm poderes gráficos um pouco diferentes
- Podem ser geradas pela função `ts()`: `ts(c(1:48), start=c(2010,1), end=c(2013,12), frequency=12)`

Fatores

- São variáveis categóricas, que agrupam os dados
 - Exemplo: “grupo controle” e “grupo teste” são dois níveis de um fator
- Podem ser ordenados ou não-ordenados
- Podem ser usados como fatores: vetores, linhas ou colunas de matrizes e data.frames, etc.
- Para transformar um determinado vetor em fator usamos a função *as.factor()*, que usará como níveis do fator, cada uma das entradas no vetor:
as.factor(c(1,1,1,2,3,3,2,1,2,3,2,1,1))

Funções

- É através delas que praticamente todas as atividades do R são desempenhadas
- Há funções simples, como somatório
- Há funções para cálculo de árvores filogenéticas
- Há funções para cálculos estatísticos
- Podemos criar nossas próprias funções
- Tem sintaxe geral *function(data, args)*
- Argumentos (parâmetros) são os valores e objetos a serem atribuídos em uma função

Importando e manipulando dados

- Praticamente todo tipo de dados pode ser importado no R
- Os tipos mais comuns são matrizes, `data.frames` e listas (fotos e sons são derivações destes tipos de dados)
- Dados importados podem ser manipulados dentro do R, mantendo os originais
- Funções como `read.csv()` e `read.table()` são as mais usadas

Conversões de objetos

- Por vezes precisaremos converter objetos internos do R para usar funções e pacotes específicos
- Conversões de vetores em fatores são corriqueiras
- Conversões de matrizes em dataframes
- Conversões entre classes especiais para análises filogenéticas

Manipulando dados

- ♦ Pacotes *dplyr* e *tidyr*
- ♦ Principais funções
 - *group_by()*, *mutate()*, *arrange()*,
filter(), *sample_n()*, *distinct()*
 - *gather()*, *separate()*, *spread()*,
unite()

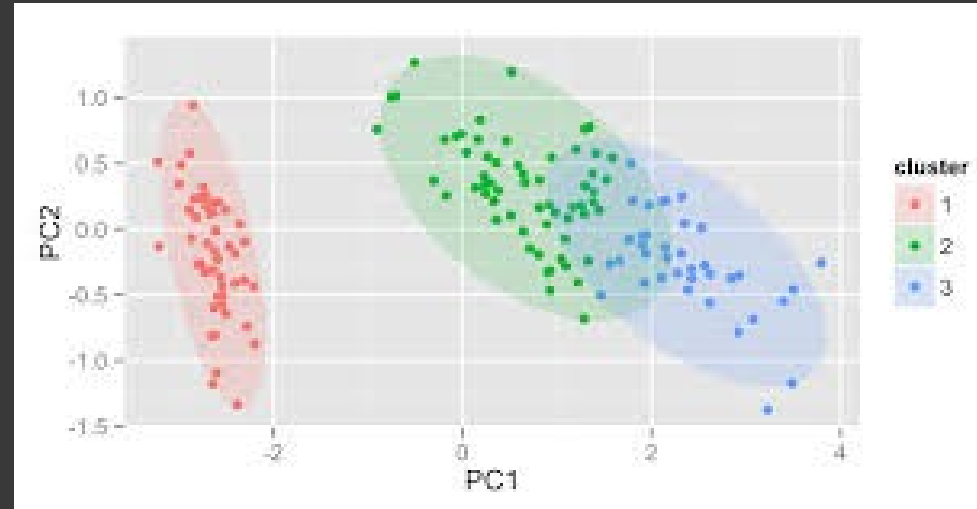
Lendo manuais de pacotes

- Todo pacote do R possui um manual, um paper de apresentação e vinhetas
- Manual apresenta todas as funções e argumentos
- Vinhetas trazem exemplos de uso e tutoriais
- Material extra disponível no GitHub

Salvando objetos em arquivos

- Para salvar objetos simples, como tabelas e vetores, usar função `write.txt()` ou `write.csv()`: `write.txt(objeto, "nomedoarquivo.txt")`
- Para salvar gráficos, por exemplo, devemos abrir o dispositivo gráfico do R destinando um arquivo, adicionar o que será gravado, e fechar o dispositivo

```
pdf("iris.pdf", width=800,  
height=600)  
plot(iris$Sepal.Length)  
dev.off()
```



Importando dados moleculares

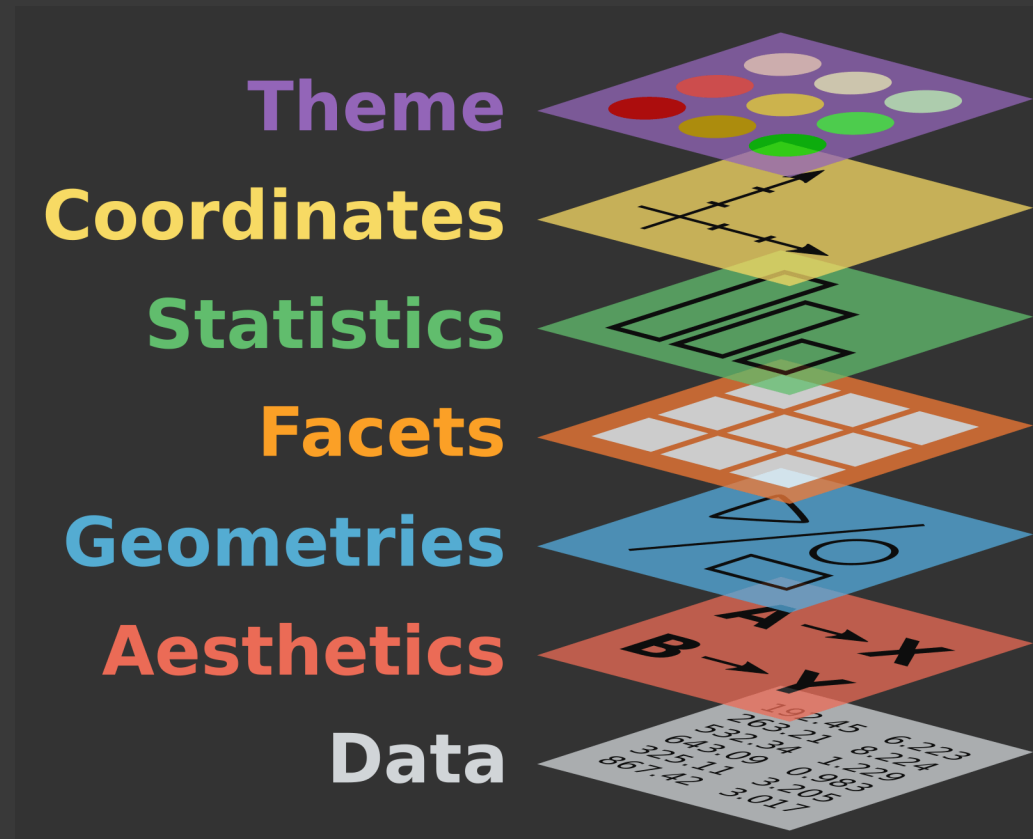
- Diferentes pacotes possuem funções para importação de tipos específicos de dados moleculares
- Principais pacotes: *ape*, *adegenet*, *pegas*, *phangorn* e *phytools*
- Principais funções: *read.FASTA()*, *read.phyDat()*, *read.nexus()* e *read.dna()*
- Criam objetos de classes específicas

Gráficos

- Função básica *plot()*: *plot(x~y)*
- Muitos tipos de gráficos podem ser gerados com esta função: gráficos de dispersão, de barras, de pizza, histogramas, boxplots, etc.
- Há pacotes específicos e aprimorados, como o *ggplot2*

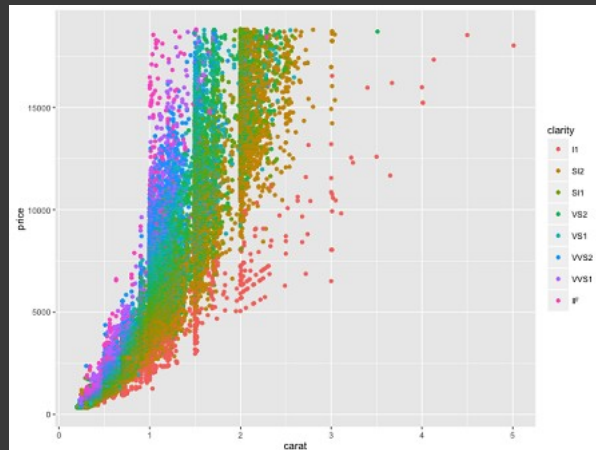
Gráficos

- O *ggplot2* trabalha com a sintaxe Grammar of Graphics



Gráficos

- `ggplot(data=dados, aes(x=eixox, y=eixoy))` # Primeira camada da função. Somente a informação dos dados a serem usados
- `ggplot(data=dados, aes(x=eixox, y=eixoy)) + geom_point()` # Ao adicionar os geométricos temos o gráfico mais básico possível



Rcmdr

- Pacote de interface gráfica para análises estatísticas mais comuns
- Implementa diversos outros pacotes e plug-ins
- Ótimo pacote para aprendizado
- Permite criar scripts, já que disponibiliza as linhas usadas em cada análise

FileEditCodeViewPlotsSessionBuildDebugProfileToolsHelp

Go to file/function

Addins

aula

EnvironmentHistoryConnections

Import Dataset

List

Global Environment

Environment is empty

HelpViewer

eteRenameMore

R > Material Minicurso Introdução ao R > Scripts > aula

SizeModified

5.1 KBDec 12, 2017, 9:57 PM

1.2 KBDec 8, 2017, 4:17 PM

218 BNov 12, 2018, 5:13 PM

Ativar o Windows

Acesse Configurações para ativar o Windows.

Untitled1*

Source on Save

RunSource

1library(Rcmdr)

2

2:1 (Top Level)

ConsoleTerminal

C:/Users/piato/Dropbox/BIOLOGIA/R/

Carregando pacotes exigidos: sandwich

Carregando pacotes exigidos: effects

lattice theme set by effectsTheme()

See ?effectsTheme for details.

RcmdrMsg: [1] NOTA: Versão do R Commander 2.5-1: Mon Nov 12 17:31:53 2018

Versão do Rcmdr 2.5-1

Attaching package: 'Rcmdr'

The following object is masked from 'package:car':

Confint

>

R Commander

ArquivoEditarDadosEstatísticasGráficosModelosDistribuiçõesFerramentasAjuda

Conjunto de Dados: <Não há conjunto de dados ativo>

Editar conjunto de dados

Ver conjunto de dados

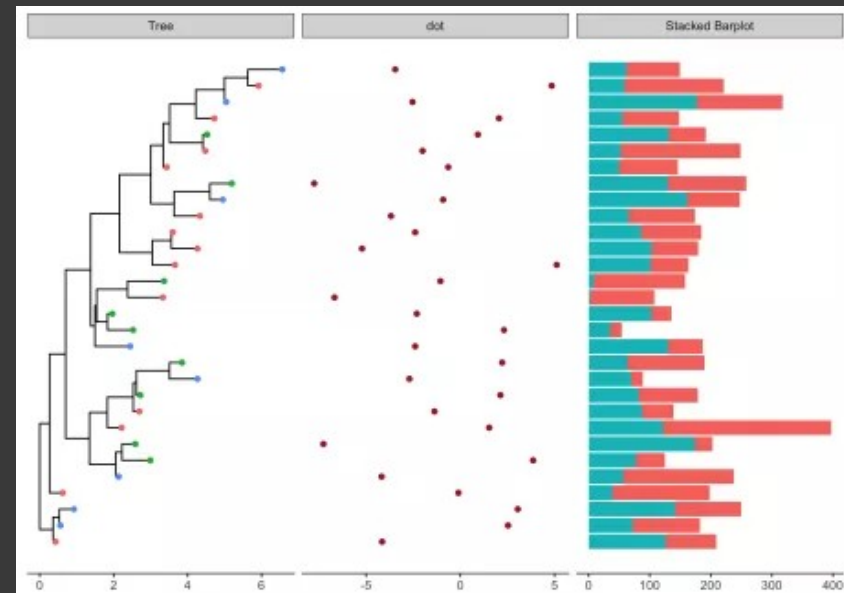
Modelo: <sem modelo ativo>

R ScriptR Markdown

Submeter

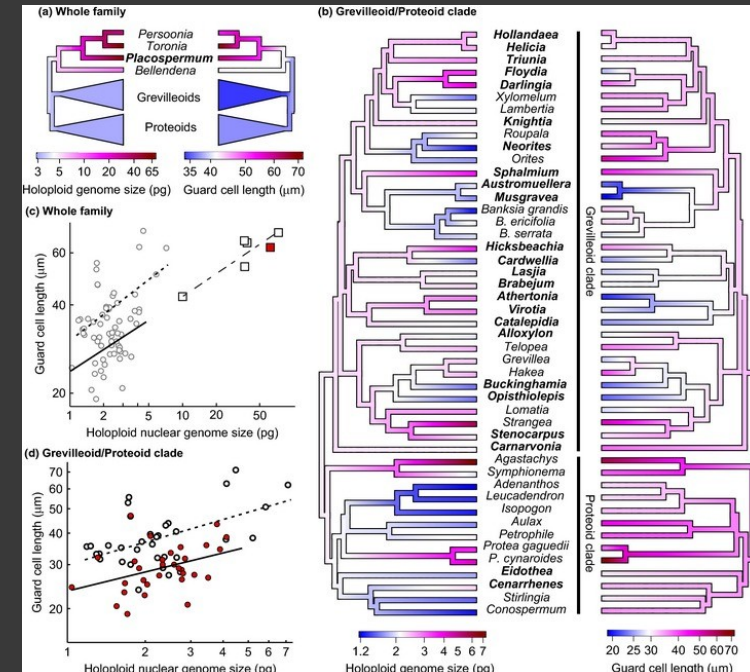
Análises filogenéticas básicas

- Análises filogenéticas no R são feitas através de etapas
- Análises baseada em distância, em máxima verossimilhança e em máxima parcimônia são as mais comuns
- São personalizáveis e possibilitam a integração com análises estatísticas
- Envolvem uma série de passos



Manipulando árvores

- Diversos pacotes permitem a manipulação gráfica de árvores filogenéticas
- Substituem programas como Adobe Illustrator
- Alta liberdade de personalização
- Reduz erros gráficos por integrar análises



MUITO OBRIGADO !