



# Linguagem R

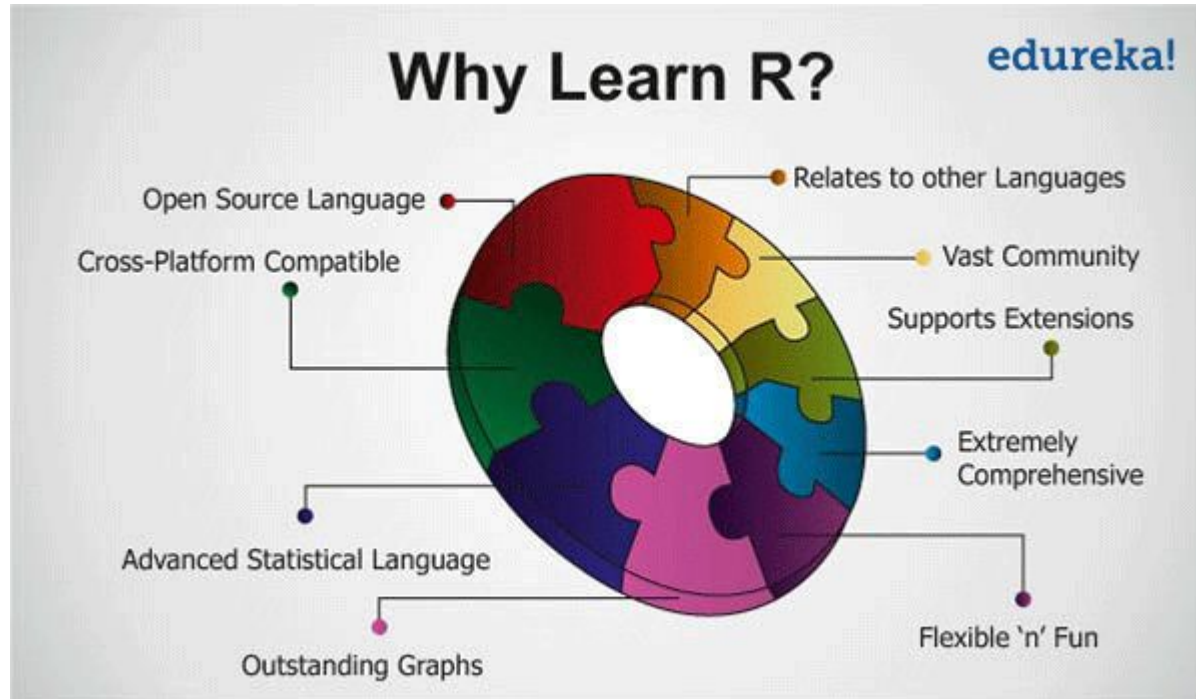


Aluna: Lays F. E. Popova

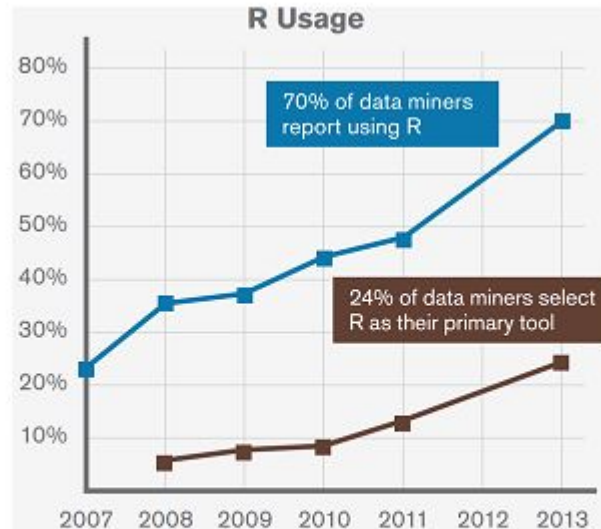
Matrícula: 201210322411

Professor: Francisco Sant'anna

# Por que aprender R?



# Uso da linguagem R na Mineração de dados



[Click on image for larger view.]

*The growth of R*

*(source: Rexer Analytics 2013 Data Miner Survey, provided by Revolution Analytics)*

# Integração do R com outras plataformas

- ❖ Se integra bem com outras linguagens de programação como C ++, Java e C ;
- ❖ De forma geral O R possuiu ferramentas implementadas em pacotes para acesso aos seguintes sistemas gerenciadores:
  - .txt;
  - .CSV e .XLS
  - .REC do Epi-Info
  - "Minitab Portable Worksheet"
  - OCTAVE (um clone do MATLAB)
  - Oracle, Microsoft SQL server, MySQL, PostgreSQL, Microsoft Access, dentre outros.



Minitab® 17





Propósito	
R	Python
análise de dados	Produtividade
estatística e modelos gráficos	código fácil de ser entendido (Readability)
Uso	
R	Python
Utilizado no meio acadêmico	Programadores e / ou profissionais de TI que já possuem conhecimento prévio em
Utilização crescente no meio corporativo.	programação.
Comunidade	
R	Python
Grande comunidade que suporta a linguagem	Bom suporte dos seus seguidores
Fóruns (Stackoverflow)	Stackoverflow
Contribuição com documentação	Boa documentação
Estatísticos, pesquisadores, analistas de dados	Programadores / desenvolvedores



### Usability

R	Python
Modelos estatísticos podem ser escrito em poucas linhas	Codificação e debugging são mais fáceis
uma mesma funcionalidade pode ser escrita de várias maneiras.	Indentação afeta o significado
	Mesmo qualquer pequena funcionalidade será escrita da mesma forma.

### Jobs And Salary?

2014 Dice Tech Salary Survey:  
Average Salary For High Paying Skills and Experience



\$ 115,531



\$94,139

# Código

## Importando CSV

R

```
nba <- read.csv("nba_2013.csv")
```

Python

```
import pandas  
nba = pandas.read_csv("nba_2013.csv")
```

## Calculando erro de modelos de predições

R

```
mean((test["ast"] - predictions)^2)
```

Python

```
from sklearn.metrics import mean_squared_error  
mean_squared_error(test["ast"], predictions)
```

## Download de webpages

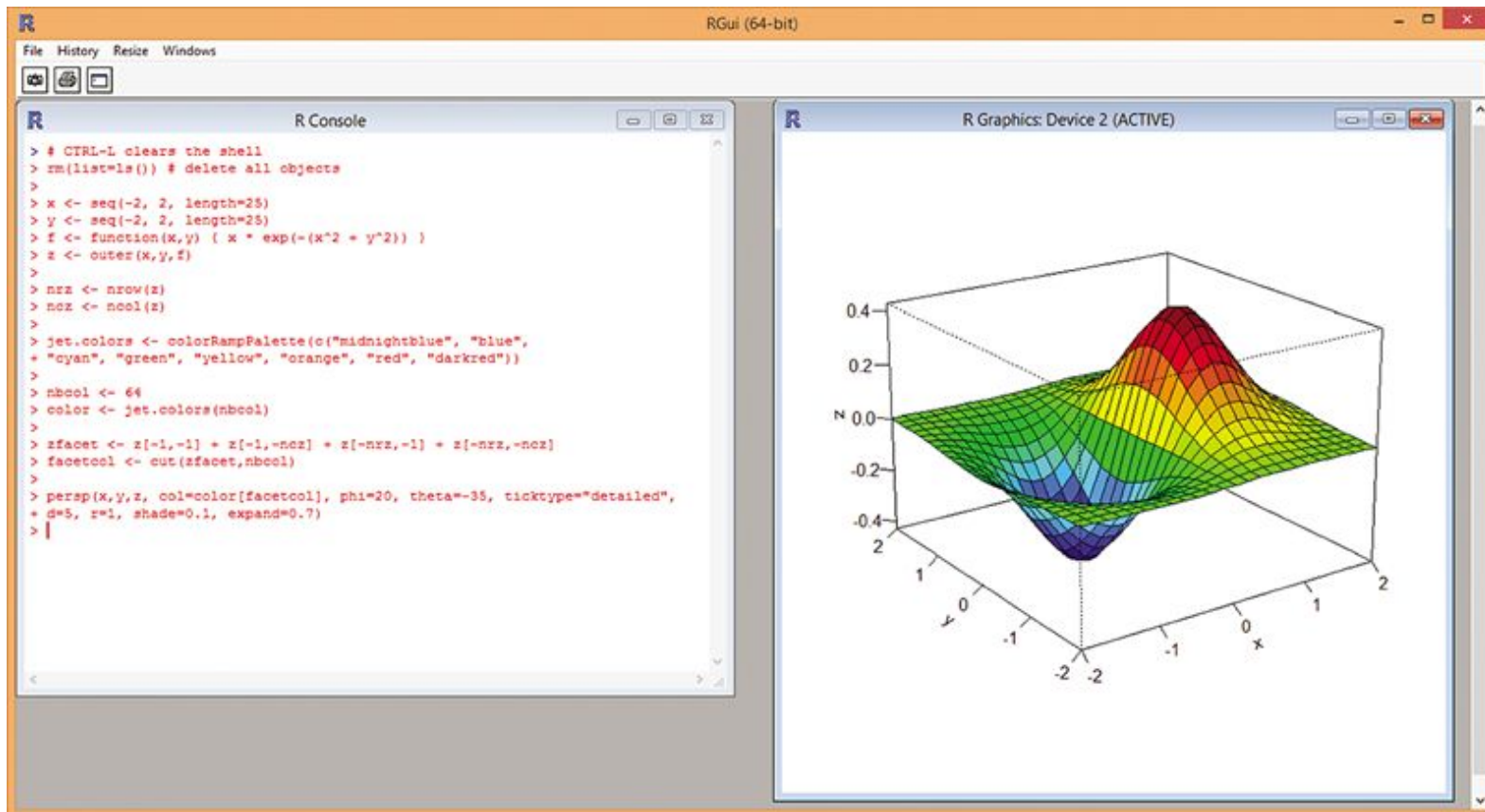
R

```
library(RCurl)  
url <- "http://www.basketball-reference.com/boxscores/201501150WAS.html"  
data <- readLines(url)
```

Python

```
import requests  
url = "http://www.basketball-reference.com/boxscores/201501150WAS.html"  
data = requests.get(url).content
```

# R Studio





# Como programar no R

RStudio interface showing a script, console, and environment pane.

**Script:**

```
1  
2  
3
```

**Console:**

```
> barplot(jogadores, col= "darkgreen",main="Esportes Coletivos", xlab="Esportes", ylab="Numero de Jogadores", names.arg=coletivo$Esportes, border="black");  
Hit <Return> to see next plot:  
> coletivo  
Esportes Numero.Jogadores  
1 Futebol 11  
2 Volei 6  
3 Basquete 6  
4 rugby 15  
5 handebol 7  
6 polo 4  
> sapply(coletivo, mean, na.rm=TRUE);  
Esportes Numero.Jogadores  
NA 8.166667  
warning message:  
In mean.default(X[[i]], ...) :  
argumento não é numérico nem lógico: retornando NA  
>  
> sapply(coletivo$Numero.Jogadores, mean, na.rm=TRUE);  
[1] 11 6 6 15 7 4  
>  
> print(coletivo$Esportes);  
[1] Futebol volei Basquete rugby handebol polo  
Levels: Basquete Futebol handebol polo rugby volei  
>
```

**Environment:**

Object	Class	Value
Esportes	Factor w/ 6 levels	Basquete, Futebol, ...
g	Factor w/ 10 levels	"1", "2", "3", "4", ...
i	9L	
jogadores	int [1:6]	11 6 6 15 7 4
lev	num [1:12]	90 100 110 120 130 140 150 160 170 180 ...
media.aves	26.7	
moeda	chr [1:2]	"CARA" "COROA"
n	100	
n.aves	10L	
nJog	15L	

**Viewer:**

**Esportes Coletivos**

Esportes	Numero de Jogadores
Futebol	11
Volei	6
Basquete	6
rugby	15
handebol	7
polo	4

Obrigado!

