

R
you
ready?

In a nutshell...

Maria Luiza Mondelli (ou Malu)

malumondelli@gmail.com

Técnica em Informática

Centro de Educação Profissional em Tecnologia da Informação

Tecnóloga em TIC

Faculdade de Educação Tecnológica do Estado do Rio de Janeiro

Mestre e Aluna de Doutorado em Modelagem Computacional

Laboratório Nacional de Computação Científica



DATA
~~DESIGN~~ LED US HERE



E isso porque

VIVEMOS A
ERA DO EXCESSO



HÁ
INFORMAÇÃO
SOBRE
PREVISÃO DO TEMPO,
ESPORTES,
POLÍTICA,
FINANÇAS,
GENTE,
CIÊNCIA,
DIVERSÕES,
ARTE,
RELIGIÃO,
BANCOS,
PREVIDÊNCIA,
TELEFONES,
MERCADO DE AÇÕES,
PUBLICIDADE,
HISTÓRIA,
HORÓSCOPO,
IMPOSTOS,
EDUCAÇÃO,
TEvê,
TECNOLOGIA,
PETRÓLEO ...

INFORMAÇÃO
ACERCA DE
INFORMAÇÃO...

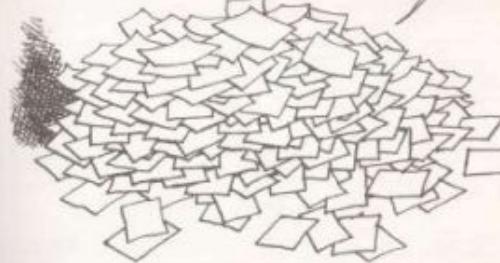


UAU!
QUE BRUTA
ONDA!



TORNA-SE CLARO QUE ESTES TEMPOS
PEDEM UM AGENTE DA TECNOLOGIA
UNICAMENTE VOLTADO PARA
ARMAZENAR, CLASSIFICAR,
QUALIFICAR, COMPARAR, COMBINAR,
E EXIBIR INFORMAÇÃO
EM ALTA VELOCIDADE!

ISSO É MAIS
UMA PA'...

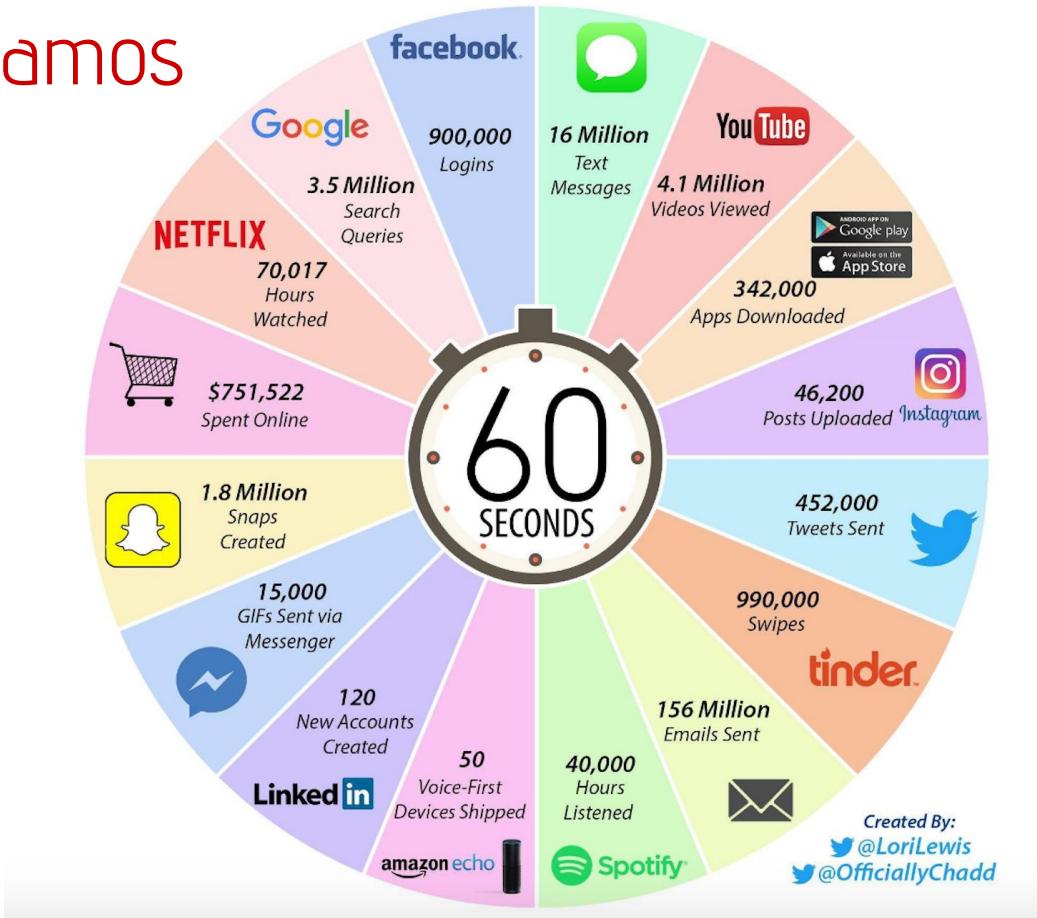


E ESTE AGENTE É O COMPUTADOR.

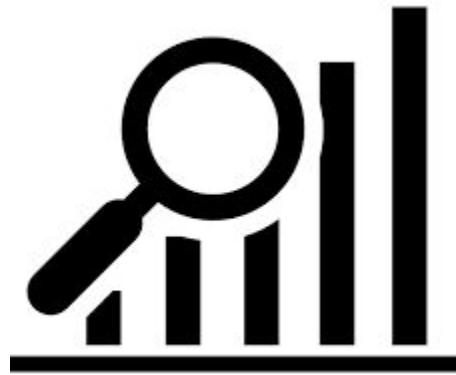




Onde estamos



E por quê isso é importante?



Data Science

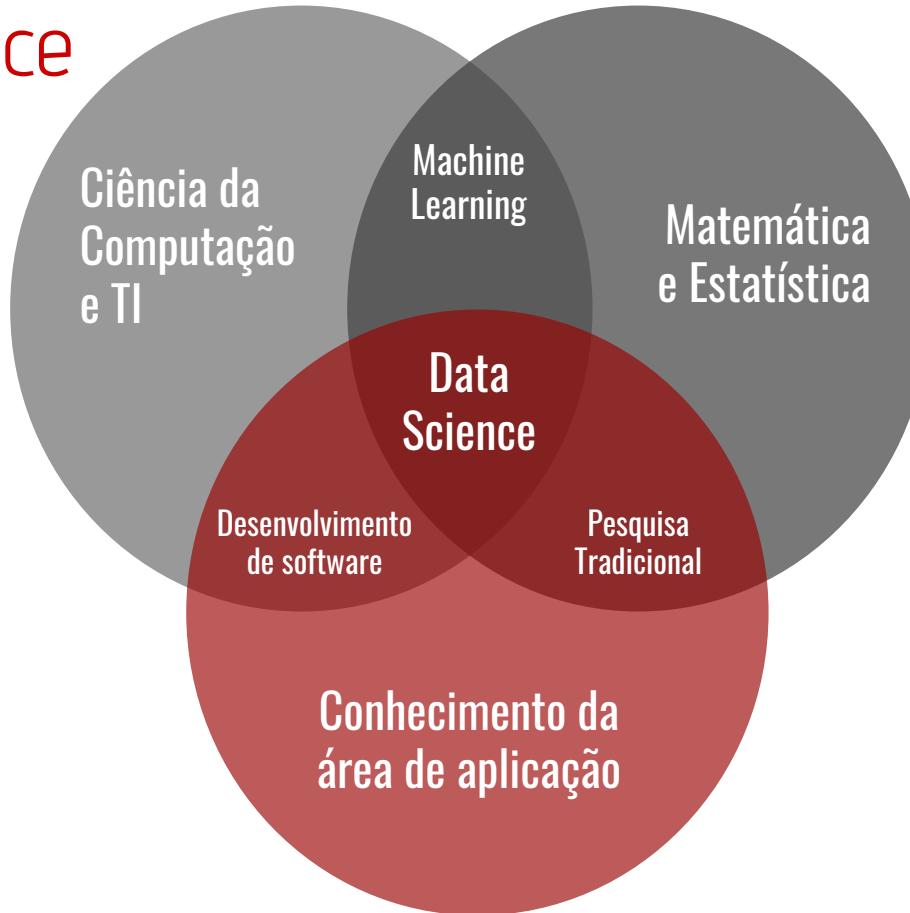
- Termo novo?
- Extrair conhecimento a partir de um conjunto de dados
 - Entender fenômenos
 - Apoiar tomada de decisão

Matemática

Estatística

Computação

Data Science



Um pouco da história

- 1990: Início do desenvolvimento
- 1993: Divulgado na lista de e-mails de novidades do S
- 1995: Disponibilizado como software livre
- 1997: Foi criado um time inicial de desenvolvimento
- 2000: Versão 1.0.0 foi lançada
- O sistema é composto por
 - Uma base (CRAN)
 - O resto



Por que R?

- É open-source
- Linguagem simples
- Versátil para resolver problemas em diversos domínios
- Possui mais de 7800 pacotes, para tudo o que você pode imaginar
- Grande comunidade de desenvolvedores e usuários

A close-up photograph of a person's hands typing on a laptop keyboard. The laptop is open and angled towards the viewer. The person is wearing a silver mesh bracelet on their left wrist. The background is a plain, light-colored wall.

Hands on!



Primeiro, sobre a instalação

Vamos utilizar o RStudio, então:

- Acesse [aqui](#)
- Em *Installers for Supported Platforms*, escolha o instalador do R Studio com base no seu sistema operacional.
- Next - Next - Finish
- Inicie o RStudio

O ambiente

script
ctrl+1

```
2 setwd('/home/mondelli/Dropbox/LNCC/Disciplinas/Evolução - Fábio Custódio/Mutação/results_ok/')
3 library(ggplot2)
4 library("reshape2")
5 library(plyr)
6
7
8 ## Unidirecional #####
9
10 # 0.25
11 u25.df <- read.csv("mUnidirecional_0001_1000000", sep = " ")
12 colnames(u25.df) <- c('Gerações', '0.25 (p)', '0.75 (q)')
13 u25.df <- melt(u25.df, id = "Gerações")
14 colnames(u25.df) <- c('Gerações', 'Valores', 'Frequência')
15 u25.plot <- ggplot(data = u25.df) +
  geom_line(aes(x = Gerações, y = Frequência, colour = Valores)) +
  theme_minimal() +
  scale_colour_manual(values=c("black", "orange"))
16
17
18
```

environment
ctrl+4

Console
ctrl+2

```
+           p = freq1_dom1[2],
+           Dominância = '1 1 0.5')
> getwd()
[1] "/home/mondelli/Dropbox/LNCC/Disciplinas/Evolução - Fábio Custódio/Seleção/results_selecao"
> setwd('/home/mondelli/Dropbox/LNCC/Disciplinas/Evolução - Fábio Custódio/Mutação/results_ok/')
> library(ggplot2)
> library("reshape2")
> library(plyr)
> u25.df <- read.csv("mUnidirecional_0001_1000000", sep = " ")
> colnames(u25.df) <- c('Gerações', '0.25 (p)', '0.75 (q)')
> u25.df <- melt(u25.df, id = "Gerações")
> colnames(u25.df) <- c('Gerações', 'Valores', 'Frequência')
> u25.plot <- ggplot(data = u25.df) +
+   geom_line(aes(x = Gerações, y = Frequência, colour = Valores)) +
+   theme_minimal() +
+   scale_colour_manual(values=c("black", "orange"))
> u25.plot
> |
```

plots
ctrl+3

Pacotes

- É o diferencial do R
- Tudo no R é feito com
 - Pacotes
 - Funções base
- Para instalar:

```
install.packages("nome do pacote")
```

Tutorial

Download:

<https://github.com/mmondelli/courses>





Spotify

“[...]to tell data stories about music in an entertaining, informative way, while maintaining a scientific approach.

Everything here will be real, in other words, based on actual data and respect for the scientific process.”

Spotify Insights



Operação Serenata do Amor

“[...] inteligência artificial capaz de analisar cada pedido de reembolso dos deputados e identificar a probabilidade de ilegalidade.”

8.276
Reembolsos suspeitos

R\$ 3,6 mi

629
Denúncias

216
Deputados

Se você gostou, isso pode ser útil:

- Sites para encontrar datasets
- Mais datasets...
- Livros - iniciante ao avançado
- Curso da Universidade de John Hopkins, EUA
- Podcasts de Data Science e Machine Learning



Thank
you!

Malu Mondelli

malumondelli@gmail.com

github.com/mmondelli

<http://lattes.cnpq.br/9518178216296481>