

```
---
title: "Espaços de autoria e novas narrativas com Rmarkdown"
author:
- Steven Dutt-Ross
- Alexandre Sousa da Silva
- Thiago de Araujo Severo dos Santos
output:
  html_document:
    theme: cerulean
    highlight: tango
    toc: true
    toc_float: true
    code_folding: hide
---

<style>
body {
  background-color: #ceefee;
}
</style>

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
#install.packages("datasauRus")
library(corrplot)
library(ggmap)
library(ggplot2)
library(datasauRus)
library(wordcloud2)
```

# Introdução

Em relação aos salários, a média global mensal em salários mínimos para o setor cultural em 2010 foi de 4,2, acima da média geral dos demais setores da economia, que situou-se em 3,2. A relação entre as faixas de pessoal ocupado total e salários médios nas atividades culturais tem comportamento similar ao do universo empresarial, quanto maior a faixa de pessoal ocupado das empresas, maior é a média de salários pagos, como pode ser observado para os anos de 2007 e 2010. As empresas maiores, na faixa de 500 pessoas ocupadas ou mais, pagaram em média os maiores salários: 7,3 salários mínimos em 2010 (Gráfico 3).



No que diz respeito ao número total de pessoas ocupadas no setor cultural, nota-se um crescimento de 13,2% entre 2007 e 2010, passando de 1,9 milhão para 2,1 milhões de pessoas ocupadas. O número de sócios e proprietários, que em 2007 somava 534,5 mil, passou para 557,6 mil, em 2010, o que representou um crescimento de 4,3%. No Brasil, as atividades caracterizadas como culturais registraram aumento no número de pessoas ocupadas assalariadas de 1,3 milhão para 1,5 milhão, entre os anos de 2007-2010. O crescimento dos ocupados assalariados do setor cultural (19,0%) foi maior vis-à-vis a taxa de 17,3% referente ao total de pessoas ocupadas assalariadas da economia.

# Tamanho de empresa

Nas estatísticas do Cempre, utiliza-se como proxy para identificar o tamanho das empresas e outras organizações o número de pessoas ocupadas em 31 de dezembro do ano de referência, neste estudo, em cinco faixas de pessoal ocupado total: 0 a 4, 5 a 19, 20 a 99, 100 a 499, 500 e mais (Tabela 2 e Gráfico 1).



A distribuição por faixa de pessoal ocupado total, tanto do número de empresas quanto do pessoal ocupado no setor cultural, se apresentou estável no período de 2007-2010. A faixa entre 0 e 4 pessoas ocupadas apresentou participação de 76,0% do total de empresas culturais e ocupou 12,7% do pessoal ocupado total no setor em 2010. As maiores empresas, com porte de 500 ou mais pessoas ocupadas, representam apenas 0,1% das empresas que atuaram nas atividades culturais, mas foram responsáveis por aproximadamente 22% do pessoal ocupado e 50% do total de salários pagos. Estas empresas registraram um aumento de participação no total de pessoas ocupadas no setor cultural (de 20,5%, em 2007, para 22,1%, em 2010).



Para as empresas e outras organizações de menor porte do setor cultural (faixa de 0 a 4 pessoas ocupadas), observou-se uma maior participação do pessoal ocupado assalariado em relação ao pessoal ocupado total, aumentando de 20,7%, em 2007, para 21,7%, em 2010. Na faixa das maiores empresas relacionadas ao setor cultural (500 e mais pessoas ocupadas) a participação do pessoal ocupado assalariado não variou entre 2007-2010 (99,9%).

Em 2010, a análise da distribuição do pessoal ocupado assalariado por porte da empresa, permite observar que as menores empresas e outras organizações do setor cultural, situadas na faixa de 0 a 4 pessoas, embora em maior número (330,1 mil empresas), ocuparam o menor número de pessoas assalariadas (116,9 mil pessoas assalariadas), conforme Tabela 2. Na outra extremidade, as empresas culturais na faixa de 500 e mais pessoas ocupadas (265 empresas) absorveram cerca de quatro vezes mais pessoal assalariado que as pequenas, ocupando 463,5 mil pessoas assalariadas. O Gráfico 2 mostra que, em 2010, para a faixa de 0 a 4 pessoas ocupadas, o número de sócios e proprietários representava 78,3% do total de pessoas ocupadas e o número de assalariados correspondia a 21,7% desse total. Contudo, na faixa de 500 e mais pessoas ocupadas, o número de sócios e proprietários representava 0,1% do total de pessoas ocupadas, enquanto o número de assalariados representava 99,9% desse total.

# R de revolucionário

O **R** é uma linguagem de programação de **código aberto** e um ambiente para computação estatística.

* R é a plataforma mais importante para o **desenvolvimento de novos métodos estatísticos** e para a prática de ciência dos dados. Se você busca os mais recentes métodos estatísticos, a melhor aposta é que eles são implementados em R.
* É fácil utilizar o R para fazer qualquer análise de dados ou atividade estatística. **Não custa nada**, então universidades, professores, estudantes e *startups* encontram poucas barreiras para adotá-lo. Entretanto, o R não enfatiza a perspectiva *point&click* como um software tradicional. Em vez disso, ele usa linha de comando que interage com ferramentas para desenvolvedores
* Existem **mais de 11.000 pacotes para R** - todos disponíveis gratuitamente - que estendem a sua utilidade. Esses pacotes disponibilizam os mais recentes métodos estatísticos e fornecem as bases para o desenvolvimento futuro.
```

```
<!--
## Pense em programação junto com a estatística

R é uma linguagem de programação projetada para trabalhar com dados e implementar algoritmos estatísticos. Uma implicação óbvia é que, se você não gosta de programação, você não vai gostar de trabalhar com R. R não enfatiza a perspectiva *point&click* como um software tradicional de escritório; Em vez disso, ele usa linha de comando que interage com ferramentas para desenvolvedores, como um ambiente de desenvolvimento integrado.

Isso significa que aprender R exige um investimento significativo e contínuo. Nenhum idioma - humano ou computador - pode ser aprendido fazendo uma breve oficina ou memorizando algumas palavras e estruturas de linguagem. Em vez disso, para se tornar fluente, **você deve praticar o idioma com frequência** e usá-lo para tarefas importantes.

## Por que você deve usar R?

1. **O R é a Ferrari dos softwares estatísticos!** Para mim, o R faz cerca de 99% das coisas que eu preciso fazer, mas, infelizmente, quando eu preciso falar com a minha esposa, eu ainda tenho que pegar o telefone.
2. O R é a **principal ferramenta para a estatística e análise de dados**. É mais do que um *software* estatístico; é uma linguagem de programação, assim você pode criar seus próprios objetos, funções e pacotes.
3. Falando de pacotes, **existem mais de 11.000 pacotes** de contribuições dos usuários disponíveis no SITE do R (CRAN). Muitos pacotes são colocados pelos membros mais proeminentes de seus respectivos campos.
4. Você pode **usá-lo facilmente em qualquer lugar**. É independente de plataforma, de modo que você pode usá-lo em qualquer sistema operacional.
5. E **é grátis**, assim você pode usá-lo em qualquer empresa, escola, em qualquer local sem ter que convencer o chefe a comprar uma licença.
6. R tem uma **grande (e crescente) comunidade de usuários**. As listas de discussão fornecer acesso a muitos usuários e autores de pacotes que são especialistas em seus respectivos campos. Além disso, existem várias conferências R cada ano.

-->

** A maioria dos softwares é construída para centralizar e controlar os dados, mas não para democratizá-los. Como resultado, as pessoas ficam reféns dos especialistas para responder questões básicas. ** O R é diferente! É democrático! É de graça!

## Para fazer citação.

> Que mal fiz eu aos deuses todos?
> Se têm a verdade, guardem-na!
> Sou um técnico, mas tenho técnica só dentro da técnica.
Fora disso sou doido, com todo o direito a sê-lo.
Com todo o direito a sê-lo, ouviram?
> Queriam-me casado, fútil, quotidiano e [tributável?] (http://www.jornaldepoesia.jor.br/facaml5.html)

## Para fazer tópicos/bullets.

> "Se nos ferem, não sangramos?"

1. Hamlet,
2. MacBeth,
3. Oteló,
4. O Mercador de Veneza,
5. Romeu & Julieta.

## Também podemos criar uma lista não-ordenada

* Séries
+ Game of Thrones
+ Vikings
+ Mr. Robot
* Estudos
+ Estatística
+ Data Science
* Músicas
+ Pop
+ Rock
+ Rock Nacional
+ Rock Progressivo
+ Rock Sinfônico

## Ênfase com Game of Thrones

Mudando o tamanho da fonte

<!--
#### "Winter Is Coming"

Ficou pequeno? que tal

# "Winter Is Coming"
-->

# "Winter Is Coming"
## "Winter Is Coming"
### "Winter Is Coming"
#### "Winter Is Coming"
##### "Winter Is Coming"
##### "Winter Is Coming"

# Colocando hiperlink

Esse é o [site do GAE] (http://gae.uniriotec.br/)

# Como colocar imagem local?

Basta adicionar o comando:
...
![] (C:/enderenco/da/sua/imagem.png)
...
ou o comando:
...
![] (https://www.r-project.org/Rlogo.png)
...

# Esse não é o unico tipo de imagem
Podemos colocar memes, gifs, e outros formatos de imagens externas<br>
(é só referenciar o link da imagem).
```

```
...\n
![] (https://media.giphy.com/media/jZtEA0nyCrVxm/giphy.gif)\n
...\n
![] (https://media.giphy.com/media/jZtEA0nyCrVxm/giphy.gif)\n
<br>\n
<br>\n
<br>\n
<br>\n
![] (https://raw.githubusercontent.com/thiagoassantos/r-spotify-popularity/master/wordcloud.png)\n
<br>\n
<br>\n
<br>\n
<br>
```

Isso dá uma oportunidade para criar...

Entre o Van Gogh e o Esher.

```
#### Van Gogh\n
![] (http://big.assets.huffingtonpost.com/ohmyvangosh.gif)\n
<br>\n
<br>
```

```
#### Esher\n
![] (https://media.giphy.com/media/KhD7MU5Kt7F16/giphy.gif)\n
<br>\n
<br>
```

Quebra de linhas
Podemos adicionar quebras de linha em textos.

Por exemplo:
> R Markdown is a file format for making dynamic documents with R. An R Markdown document is written in markdown (an easy-to-write plain text format) and contains chunks of embedded R code, like the document below.

Pode ser formatado dessa maneira:
> R Markdown is a file format for making dynamic documents with R.
 An R Markdown document is written in markdown (an easy-to-write plain text format) and contains chunks of embedded R code, like the document below.

```
Basta adicionar a tag\n
```\n

\n
...`
```

Na posição onde você deseja quebrar a linha (semelhante ao Enter ↵ do seu teclado).<br>

```
YAML\n
<!-- desenvolver algo sobre o YAML-->\n
YAML é uma linguagem que pretende simplificar a codificação de dados.
\n
No caso do R Markdown, ela define informações (metadados) no início do documento, inclusive sua estilização.
\n
Por padrão, o documento é gerado em uma página HTML, mas podemos escolher outros formatos. Os mais usados são PDF e .Docx (Word).
```

```
Basta alterar a opção de `output`. Exemplo:\n
* Para gerar um PDF:\n
`output: pdf_document`\n

\n
* Para gerar um documento do Word:\n
`output: word_document`
```

Para consultar outros formatos de saída, veja: <http://rmarkdown.rstudio.com/lesson-9.html>

## Aparência e Estilo

Existem diversas opções de controle da aparência dos documentos em HTML. O que vem a sua cabeça quando eu falo a palavra **\*\*BOOTSTRAP\*\***? <br>  
A. Nada. Nunca ouvi falar,<br>  
B. Reamostragem,<br>  
C. Customização (Aparência e Estilo)<br>

```
Bootstrap: temas válidos "default", "cerulean", "journal", "flatly", "readable", "spacelab", "united", "cosmo", "lumen", "paper", "sandstone", "simplex", and "yeti".\n

\n
Highlight style: estilos possíveis "default", "tango", "pygments", "kate", "monochrome", "espresso", "zenburn", "haddock", and "textmate". Pass null to prevent syntax highlighting.
```

```
Exemplo 1:\n
...`\n
output:\n
 html_document:\n
 theme: cerulean\n
 highlight: tango\n
...`
```

```
Exemplo 2:\n
...`\n
output:\n
 pdf_document:\n
 toc: true\n
 highlight: zenburn\n
 always_allow_html: yes\n
 word_document:\n
 fig_width: 5\n
 fig_height: 5\n
 fig_caption: true\n
 always_allow_html: yes\n
 html_document:\n
 theme: cerulean\n
 highlight: tango\n
 toc: true\n
 toc_float: true\n
 code_folding: hide
```

```

...

Programando em R dentro do Markdown

Gráficos

Diagrama de Dispersão

```{r cars, echo=TRUE}
#```{r cars, echo=FALSE}
#```{r cars, results='hide'}
#```{r cars, warning=FALSE}
#```{r cars, message=FALSE}
plot(cars$speed, cars$dist)
plot(cars$speed, cars$dist, col="#0d6887",pch=19,main="Diagrama de Dispersão", xlab="Velocidade", ylab ="o que é disp?" )
abline(lsfilt(cars$speed, cars$dist),col="#032435", lty=2, lwd=4)
...

```{r}
data(mtcars)
library(DT)
datatable(mtcars)
...

```

Correlação....

```

```{r datasaurus, echo=TRUE}
dados <- datasaurus_dozen[1:142,]
cor(dados$x,dados$y)
...

```

Vamos ver o banco de dados para nos certificar...

```

```{r datasaurus2, echo=TRUE}
datatable(dados)
...

```

Vamos fazer o diagrama de dispersão para nos certificar...

```

```{r datasaurus3, echo=TRUE}
#plot(dados$x,dados$y, col="#0d6887",pch=19)
...

```

Esse é um banco de dados para te mostrar que **sempre** precisamos visualizar a estatística!

Todas essas figuras tem a mesma correlação (rho=-0.06).

```

```{r datasaurus4, echo=TRUE}
#ggplot(datasaurus_dozen, aes(x = x, y = y, colour = dataset)) +
geom_point() +
theme_void() +
theme(legend.position = "none") +
#facet_wrap(~ dataset, ncol = 3)
...

```

## Matriz de Correlação

```

```{r correlacao, echo=TRUE}
cor(mtcars[,c("disp","drat","hp","mpg","qsec","wt")])
...

```

Mas que coisa sem graça!

```

```{r correlacao2, echo=TRUE}
M <- cor(mtcars[,c("disp","drat","hp","mpg","qsec","wt")])
corrplot(M, method="circle")
...

```

Agora sim!

## Mapas

```

Mapa do Rio de Janeiro
```{r mapa, echo=TRUE, warning=FALSE, message=FALSE}
#library(ggmap)
#qmap(location = "rio de janeiro")
#qmap(location = "rio de janeiro", zoom = 14)
#qmap(location = "rio de janeiro", zoom = 15, maptype = "satellite")
...

```

Nuvem de palavras

Você sabe o que é uma nuvem de palavras?

Já se perguntou como as pessoas criam “wordclouds” (nuvens de palavras)? Nuvens de palavras são imagens compostas por palavras de várias cores e tamanhos e, opcionalmente, organizadas em direções distintas. Pois agora você saberá como e o melhor de tudo: aprenderá a criar a sua nuvem de palavras!

Como criar a minha nuvem de palavras?

```

```{r wordcloud, echo=TRUE}
wordcloud2(data = demoFreq)
...

```

## LaTeX

Temos duas formas de **colocar** equações do LaTeX no RMarkdown

A primeira....

```

$$r = \frac{COV(x,y)}{S_x S_y} $$

```

A segunda...

Que tal colocar a **função de densidade da Normal**?

```
<!--Colocar a formula no quadro antes-->
\begin{equation}
f(x) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2} (x - \mu)^2}
\end{equation}
```

**Um modelo de regressão**

```
`r regressao, echo=TRUE}
data(mtcars)
modelol<-lm(mpg~hp, data=mtcars)
modelol
`r
```

A estimativa do  $\beta_1$  desse modelo linear simples é  
`r coef(modelol)[2]`.

# E se fosse uma apresentação? como devemos fazer?

<br>  
Existem diversas forams de apresentações no Rmarkdown. As mais utilizadas são:<br>

- a. ioslides\_presentation<br>
- b. slidy\_presentation<br>
- c. beamer\_presentation<br>
- d. slidify <br>

Exemplos:

```
`r
output: ioslides_presentation
output: slidy_presentation
`r
```

Para maiores informações, [esse link é bem útil](http://www.rstudio.com/wp-content/uploads/2016/03/rmarkdown-cheatsheet-2.0.pdf)

#### Para estudar mais o R Markdown

- \* [Tutorial do RStudio](http://rmarkdown.rstudio.com/lesson-1.html)
- \* [Manual de referência](https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf)
- \* [Guia de dicas](http://www.rstudio.com/wp-content/uploads/2016/03/rmarkdown-cheatsheet-2.0.pdf)

\* [Chapman,C. What You Should Know About R](https://www.ama.org/publications/MarketingInsights/Pages/what-you-should-know-about-r.aspx)

---

#### Biografia dos autores:

**Steven Dutt-Ross**<br>

[Steven Dutt-Ross](https://stevendutt.shinyapps.io/pagina/) é um entusiasta do R, professor e pesquisador do Departamento de Métodos Quantitativos da UNIRIO - DMQ/UNIRIO e coordenador do projeto de extensão [DATAUNIRIO](https://dataunirio.github.io/).<br>

**Alexandre Sousa da Silva**<br>

[Alexandre Sousa da Silva](http://gae.uniriotec.br/2/index.html) é professor e pesquisador do Departamento de Métodos Quantitativos da UNIRIO - DMQ/UNIRIO e coordenador do projeto de extensão chamado [Grupo de Apoio Estatístico - GAE](http://gae.uniriotec.br/).<br>

**Thiago de Araujo Severo dos Santos**<br>

[Thiago de Araujo Severo dos Santos](https://github.com/thiagoassantos/) é estudante de Sistemas de Informação da UNIRIO e bolsista do projeto de extensão [DATAUNIRIO](https://dataunirio.github.io/).