

Espaços de autoria e novas narrativas com Rmarkdown

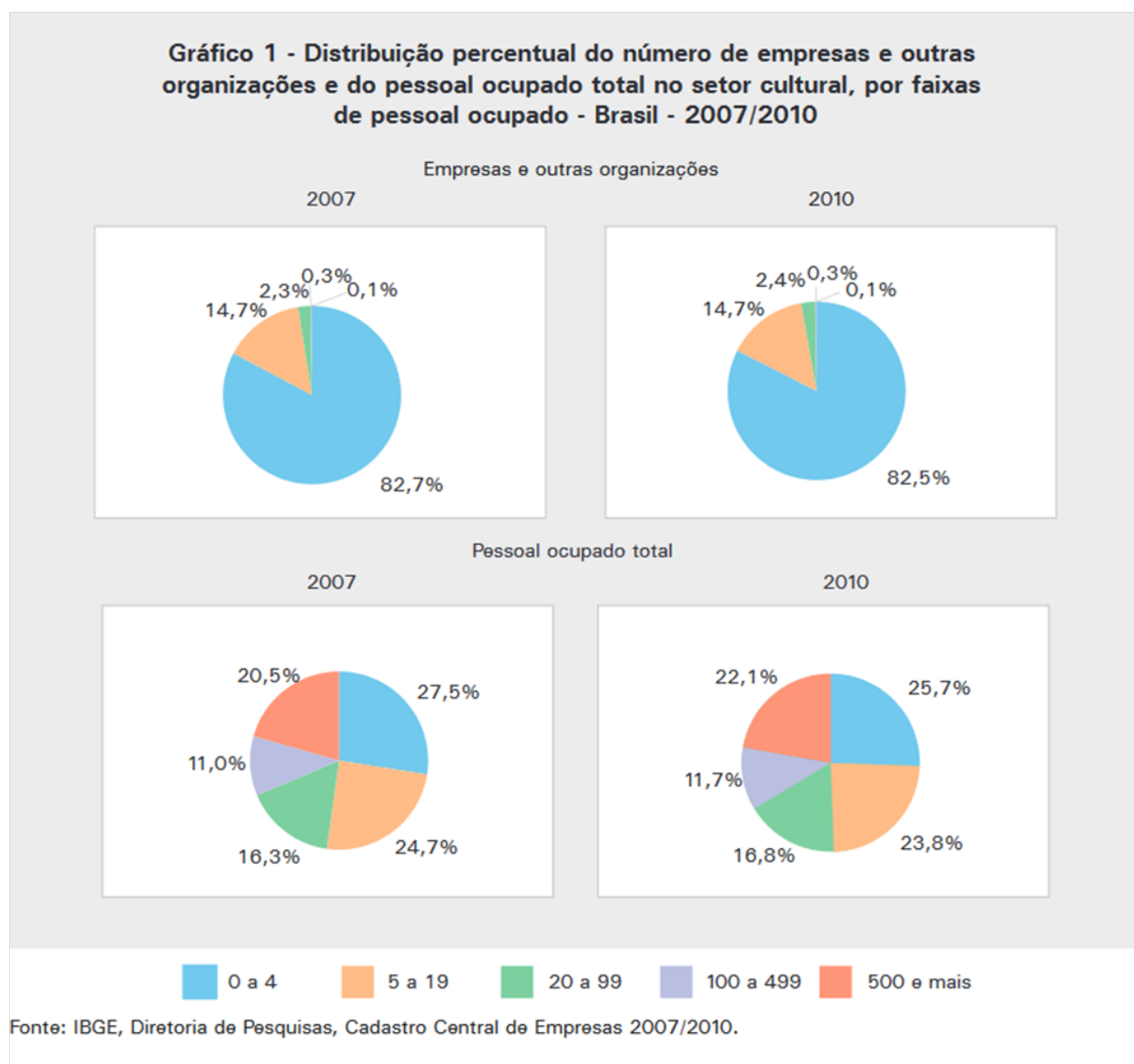
Steven Dutt-Ross

Alexandre Sousa da Silva

Thiago de Araujo Severo dos Santos

latex_engine: xelatex

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).

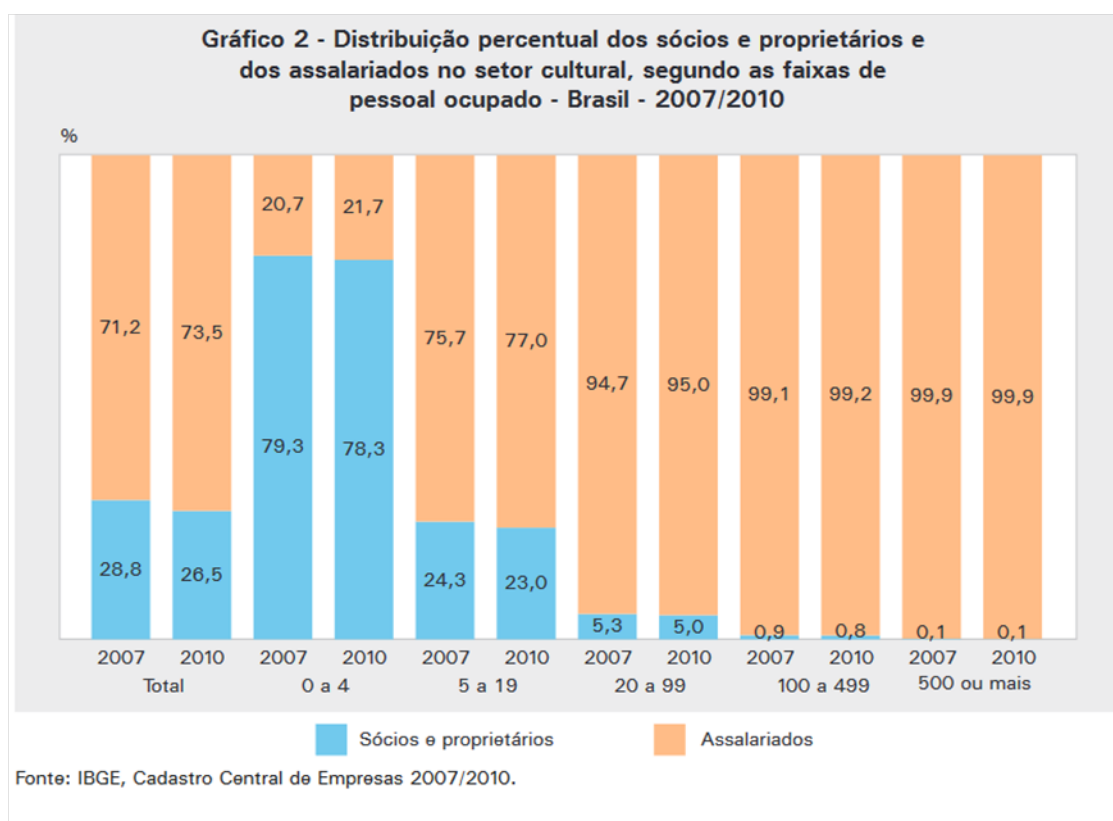


No que diz respeito ao número total de pessoas ocupadas no setor cultural, nota-se um crescimento de 13,2%

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.

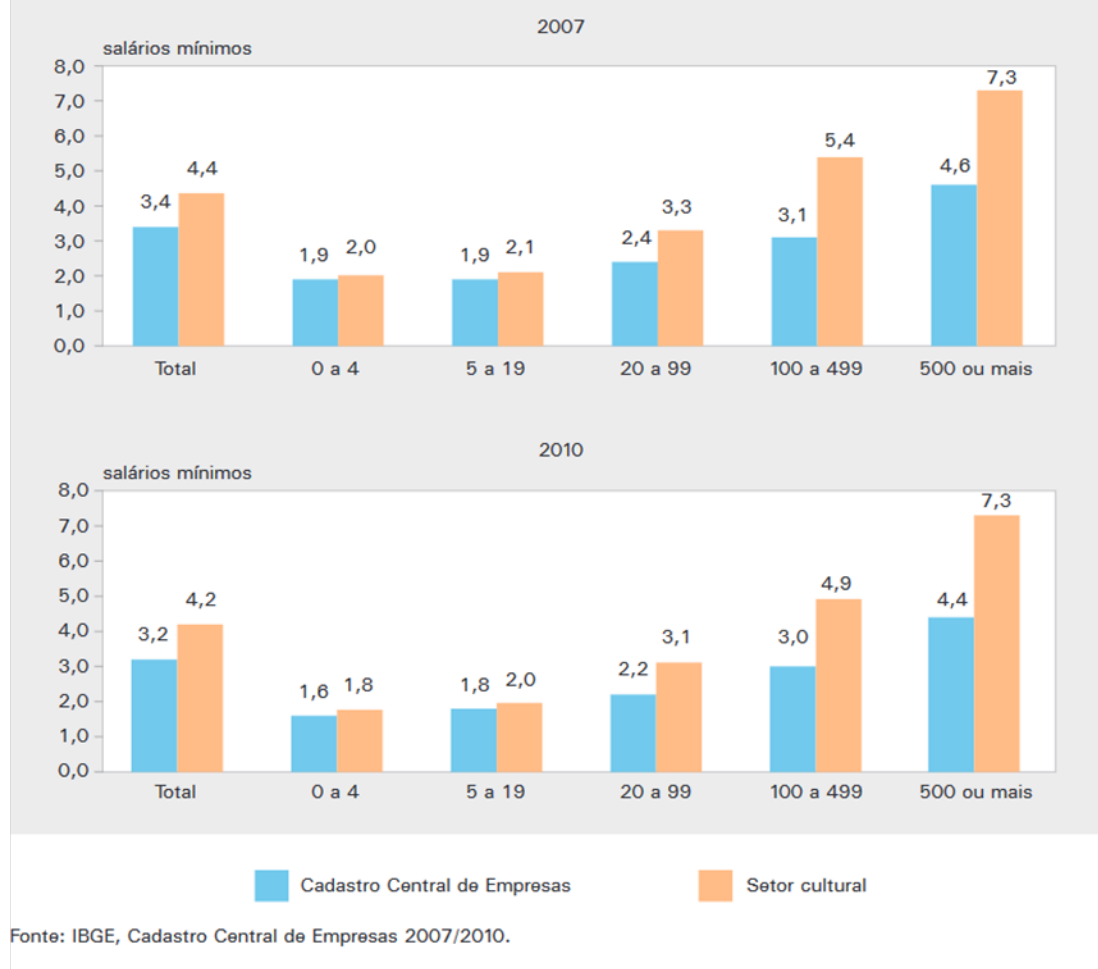
Tamanho de empresa

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, agrupadas, 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).

Gráfico 3 - Salário médio mensal no Cadastro Central de Empresas e no setor cultural, segundo as faixas de pessoal ocupado - Brasil - 2007/2010



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.

** 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?

Para fazer tópicos/bullets.

“Se nos ferem, não sangramos?”

1. Hamlet,
2. MacBeth,
3. Otelo,
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”

“Winter Is Coming”

“Winter Is Coming”

“Winter Is Coming”

“Winter Is Coming”

“Winter Is Coming”

Colocando hiperlink

Esse é o site do GAE

Como colocar imagem local?

Basta adicionar o comando:

```

```

ou o comando:

```

```

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

```
<br>
```

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

YAML

YAML é uma linguagem que pretende simplificar a codificação de dados. No caso do R Markdown, ela define informações (metadados) no início do documento, inclusive sua estilização. 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: * Para gerar um PDF: **output: pdf_document**

- Para gerar um documento do Word: **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**? A. Nada. Nunca ouvi falar, B. Reamostragem, C. Customização (Aparência e Estilo)

Bootstrap: temas válidos “default”, “cerulean”, “journal”, “flatly”, “readable”, “spacelab”, “united”, “cosmo”, “lumen”, “paper”, “sandstone”, “simplex”, and “yeti”.

Highlight style: estilos possíveis “default”, “tango”, “pygments”, “kate”, “monochrome”, “espresso”, “zenburn”, “haddock”, and “textmate”. Pass null to prevent syntax highlighting.

Exemplo 1:

```
output:
  html_document:
    theme: cerulean
    highlight: tango
```

Exemplo 2:

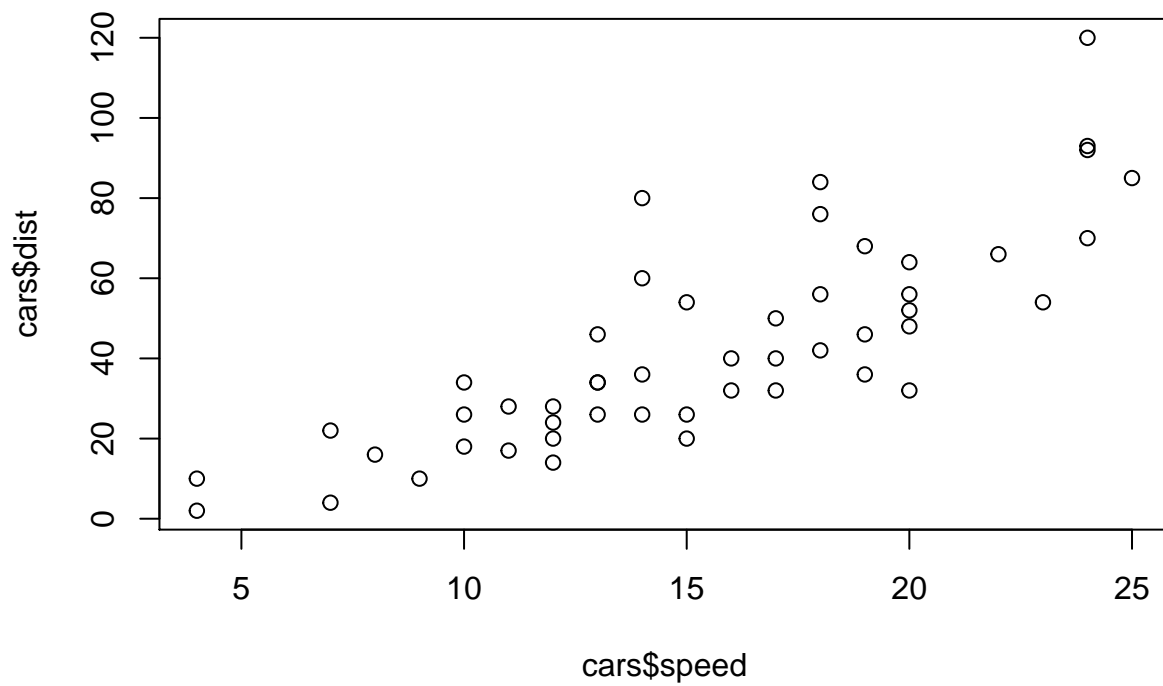
```
output:
  pdf_document:
    toc: true
    highlight: zenburn
    always_allow_html: yes
  word_document:
    fig_width: 5
    fig_height: 5
    fig_caption: true
    always_allow_html: yes
  html_document:
    theme: cerulean
    highlight: tango
    toc: true
    toc_float: true
    code_folding: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

Programando em R dentro do Markdown

Gráficos

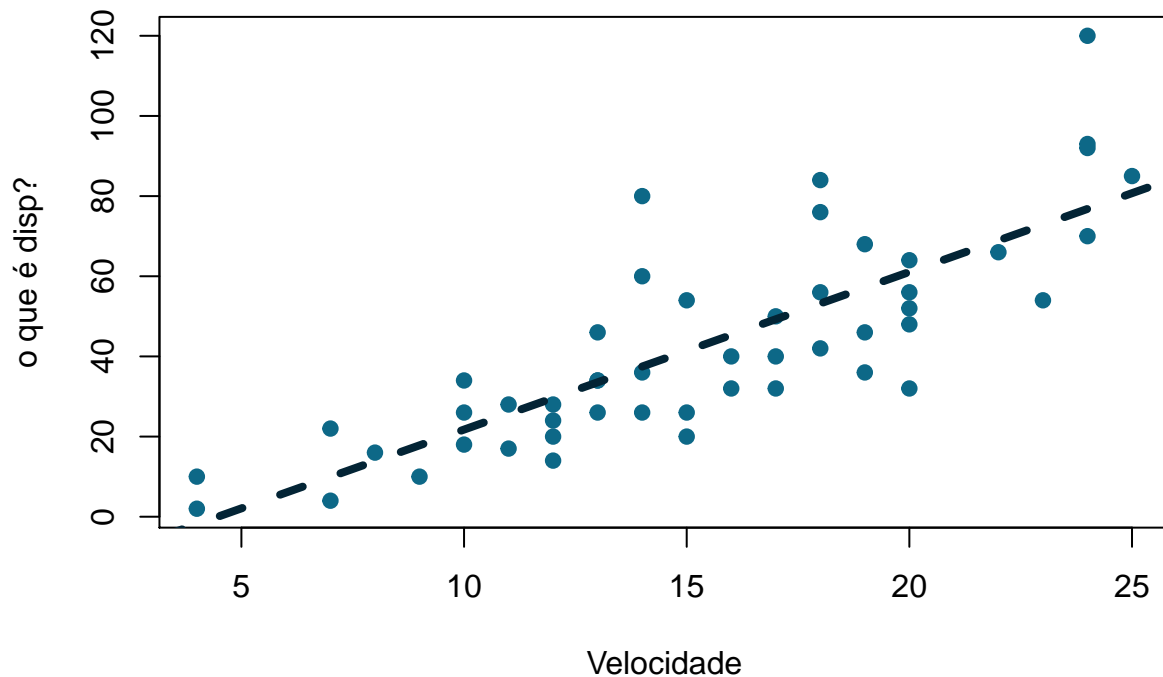
Diagrama de Dispersão

```
#```${r cars, echo=FALSE}
#```${r cars, results='hide'}
#```${r cars, warning=FALSE}
#```${r cars, menssage=FALSE}
plot(cars$speed, cars$dist)
```



```
plot(cars$speed, cars$dist, col="#0d6887",pch=19,main="Diagrama de Dispersão", xlab="Velocidade", ylab = "Distância",  
abline(lsfit(cars$speed, cars$dist),col="#032435", lty=2, lwd=4))
```

Diagrama de Dispersão



Correlação....

```
dados <- datasaurus_dozen[1:142,]  
cor(dados$x,dados$y)
```

```
## [1] -0.06447185
```

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

```
library(DT)  
datatable(dados)
```

```
## PhantomJS not found. You can install it with webshot::install_phantomjs(). If it is installed, please
```

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

```
#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 ($\rho=-0.06$).

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

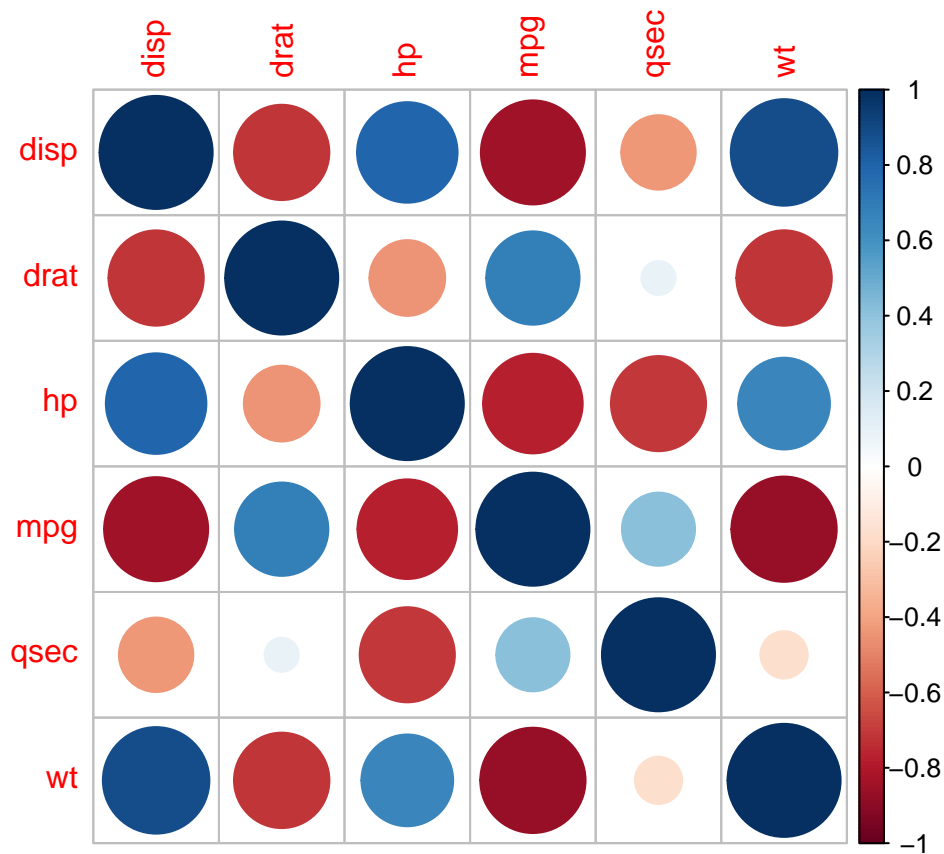

Matriz de Correlação

```
cor(mtcars[,c("disp", "drat", "hp", "mpg", "qsec", "wt")])
```

```
##           disp      drat      hp      mpg      qsec      wt
## disp  1.0000000 -0.7102139  0.7909486 -0.8475514 -0.43369788  0.8879799
## drat -0.7102139  1.0000000 -0.4487591  0.6811719  0.09120476 -0.7124406
## hp    0.7909486 -0.44875912  1.0000000 -0.7761684 -0.70822339  0.6587479
## mpg  -0.8475514  0.68117191 -0.7761684  1.0000000  0.41868403 -0.8676594
## qsec -0.4336979  0.09120476 -0.7082234  0.4186840  1.00000000 -0.1747159
## wt    0.8879799 -0.71244065  0.6587479 -0.8676594 -0.17471588  1.0000000
```

Mas que coisa sem graça!

```
M <- cor(mtcars[,c("disp", "drat", "hp", "mpg", "qsec", "wt")])
corrplot(M, method="circle")
```



Agora sim!

Mapas

Mapa do Rio de Janeiro

```
#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?

```
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?**

$$f(x) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad (1)$$

Um modelo de regressão

```
data(mtcars)
modelo1<-lm(mpg~hp, data=mtcars)
modelo1

##
## Call:
## lm(formula = mpg ~ hp, data = mtcars)
##
## Coefficients:
## (Intercept)          hp
##    30.09886      -0.06823
```

A estimativa do β_1 desse modelo linear simples é -0.0682283.

E se fosse uma apresentação? como devemos fazer?

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

- ioslides_presentation
- slidy_presentation
- beamer_presentation
- slidify

Para maiores informações, esse link é bem útil

Para estudar mais o R Markdown

- Tutorial do RStudio
 - Manual de referência
 - Guia de dicas
 - Chapman,C. What You Should Know About R
-

Biografia dos autores:

Steven Dutt-Ross Steven Dutt-Ross é um entusiasta do R, professor e pesquisador do Departamento de Métodos Quantitativos da UNIRIO - DMQ/UNIRIO e coordenador do projeto de extensão DATAUNIRIO.

Alexandre Sousa da Silva Alexandre Sousa da Silva 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.

Thiago de Araujo Severo dos Santos Thiago de Araujo Severo dos Santos é estudante de Sistemas de Informação da UNIRIO e bolsista do projeto de extensão DATAUNIRIO.