

## ESTUDOS AVANÇADOS EM R: VISUALIZAÇÃO DE DADOS

### Introdução

Em uma época onde o volume de dados cresce de maneira exponencial, fica nítida a importância da visualização de dados. Além da quantidade destes dados, o número de dimensões e variáveis que um pesquisador pode analisar tornaram algumas representações de dados insuficientes para lidarem com a complexidade dos fenômenos investigados. Para lidar com esta complexidade, elaborou-se uma “gramática” que descreve como mapear variáveis a diferentes características visuais, de modo que relações possam ser exploradas visualmente de maneira intuitiva e que façam mais sentido à cognição humana que uma sequência de matrizes e tabelas (WILKINSON, 2005). Com algumas das novas ferramentas desenvolvidas neste campo, permite-se um controle maior do usuário final por meio de representações interativas, como, por exemplo, a ampliação para visualizar os dados em escala micro, ou o afastamento para visualizar os mesmos dados em escala macro. De forma lúdica, a exploração de dados por meio deste mapeamento e controle do usuário abre as portas para uma abordagem em que é possível dar ao usuário o poder de produzir narrativas com os dados de modo mais personalizado.

Para entrar neste campo e apresentar algumas destas novas ferramentas, utilizaremos nesse trabalho o software R (R Core Team, 2017) através do RStudio (RStudio Team, 2015). A escolha do R se dá principalmente por ser um software livre e oferecer um conjunto de funções e bibliotecas estatísticas para a elaboração de técnicas gráficas simples e sofisticadas dentro dessa gramática dos gráficos. Desta forma, para a criação de uma aplicação web se faz necessário o estudo de alguns pacotes para a visualização que ajudem a revelar informações sobre os dados.

### Objetivos

O objetivo mediato deste trabalho é aprofundar os estudos das funções de visualizações de dados através de gráficos claros e eficientes, utilizando o R para a criação de apresentações interativas e dinâmicas. Com isto, procuramos oferecer alternativas atraentes para a exploração dinâmica dos dados, de modo a integrar uma base de dados do Instituto Brasileiro de Geografia e Estatística (IBGE) com as novas formas de visualização utilizando o software R.

## **Material e Método**

Para este trabalho foi necessário estudar os principais pacotes para confecção de gráficos interativos para visualizar as características importantes de um conjunto de dados. Os principais pacotes utilizados para representações gráficas são:

(a) ggplot2 (WICKHAM & CHENG, 2016), que permite a criação de gráficos por camadas de forma que seja menos trabalhoso e mais fácil de entender o código para uma grande quantidade de tipos e estilos de representações visuais.

(b) plotly (SIEVERT ET ALII, 2017), que produz ou traduz gráficos do ggplot2 interativos e também cria gráficos online através de uma conta “plotly”.

(c) highcharter (KUNST, 2017), que permite a criação de gráficos interativos com a função hchart() dentro do R.

Foi desenvolvida uma aplicação web com o uso do pacote shiny (CHANG ET ALII, 2017), sistema para desenvolvimento de aplicações web através do R, de forma a integrar a base de dados da Pesquisa Nacional por Amostra de Domicílios (PNAD), por meio do Sistema IBGE de Recuperação Automática (SIDRA), para obter uma visualização de dados interativa através de gráficos utilizando os pacotes citados. Para a obtenção dos dados utilizamos o API SIDRA que permite a extração de dados do Banco Sidra, com o pacote sidrar (SIQUEIRA, 2017) que permite a conexão com o IBGE e disponibiliza os dados agregados.

## **Resultados e Discussão**

Com o levantamento dos pacotes necessários para visualização de dados e implementação de representações dinâmicas obtivemos um documento feito com o uso rmarkdown (ALLAIRE ET ALII, 2018), que contém opções para gerar documentos em vários formatos de maneira mais aprimorada, intitulado “Guia do ggplot” com suas principais funções e recursos estudados. Esse documento aborda desde a sua forma mais simples os conceitos do uso da função ggplot() até os mais complexos, de forma que uma pessoa com conhecimentos básicos em R consiga utilizar e aplicar aos seus dados os mais diversos gráficos. Também se encontram nesse guia explicações e exemplos com os pacotes plotly e highcharter, além de uma introdução a utilização do shiny com esses pacotes. E com a aplicação web criada é possível extrair informações, escolher as variáveis de interesse a serem analisadas além de personalizar a aparência do gráfico de forma que seja mais compreensível e atrativa por apresentar gráficos interativos da base de dados da PNAD.

## **Conclusão**

A importância desse trabalho é a aplicação web como uma fonte de acesso para quem busca uma forma mais sofisticada e detalhada de gráficos interativos e análise dos dados da PNAD. E dada à importância crescente da visualização de dados e comunicação ou apresentação dos resultados, faz-se necessária à elaboração destas aplicações e documentos de referência para a representação gráfica no R.

## **Referências**

[1] Carson Sievert, Chris Parmer, Toby Hocking, Scott Chamberlain, Karthik Ram, Marianne Corvellec, Pedro Despouy (2017). Plotly: R package version 4.7.1. <https://cran.r-project.org/package=plotly>

[2] Hadley Wickham, Winston Chang (2016). Ggplot2: R package version 2.2.1. <https://cran.r-project.org/package=ggplot2/>

[3] JJ Allaire, Jeffrey Horner, Vicent Marti and NatachaPorte (2018). Markdown: R package version 1.9. <https://CRAN.R-project.org/package=markdown>

[4] Joshua Knust (2017). Highcharter: R package version 0.5.0. <https://cran.r-project.org/package=highcharter>

[5] R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org/>.

[6] Renato Prado Siqueira (2017). Sidrar: R package version 0.2.4. <https://cran.r-project.org/package=sidrar>

[7] RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>.

[8] Wilkinson, L. (2005) The Grammar of Graphics (Statistics and Computing). Springer.

[9] Winston Chang, Joe Cheng, JJ Allaire, YihuiXie and Jonathan McPherson (2017). Shiny: R package version 1.0.5. <https://CRAN.R-project.org/package=shiny>