UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas
Programa de Especialização em Estatística

Constantino Veríssimo dos Santos Filho

UM APLICATIVO *WEB* PARA ANÁLISE DE DADOS PÚBLICOS

Belo Horizonte
2022

Constantino Veríssimo dos Santos Filho

UM APLICATIVO *WEB* PARA ANÁLISE DE DADOS PÚBLICOS

Monografia apresentada ao Programa de Especialização em Estatística da Universidade Federal de Minas Gerais (UFMG), como requisito parcial para a obtenção do título de Especialista em Estatística.

Orientador: Prof. Dr. Fábio Nogueira Demarqui

Belo Horizonte
2022

:

Santos Filho, Constantino Veríssimo dos.

S237a Um aplicativo web para análise de dados públicos [manuscrito] / Constantino Veríssimo dos Santos Filho. — 2022.
46.f. il.

Orientador: Fábio Nogueira Demarqui.
Monografia (especialização) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística.
Referências 44-45.

1. Estatística. 2. Analise por conglomerados. 3. Sistemas interativos. I. Demarqui, Fábio Nogueira. II. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatísitca .III.Título.

CDU 519.2 (043)

Ficha Ficha catalográfica elaborada pela bibliotecária Belkiz Inez Rezende Costa CRB 6/1510 Universidade Federal de Minas Gerais - ICEx

Universidade Federal de Minas Gerais **E-mail:**
r
Instituto de Ciências Exatas Tel: 3409-
9-5924
Departamento de Estatística
P Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 – Belo Horizonte – MG

ATA DO 244ª. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DE CONSTANTINO VERÍSSIMO DOS SANTOS FILHO.

Aos dezenove dias do mês de julho de 2022, às 13:30 horas, com utilização de recursos de videoconferência a distância, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística, para julgar a apresentação do trabalho de fim de curso do aluno **Constantino Veríssimo dos Santos Filho**, intitulado: "*Um Aplicativo Web para análise de dados públicos*", como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, o Presidente da Comissão, Professor Fabio Nogueira Demarqui – Orientador, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Após a defesa, os membros da banca examinadora reuniram-se sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: o candidato foi considerado Aprovado condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje por unanimidade. O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 19 de julho de 2022.

Prof. Fabio Nogueira Demarqui (Orientador)
Departamento de Estatística / UFMG

Prof. Cristiano de Carvalho Santos
Departamento de Estatística / UFMG

Walmir dos Reis Miranda Filho
Doutor em Estatística Pelo Departamento de Estatística / UFMG

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br
Tel: 3409-5923 – FAX: 3409-5924

DECLARAÇÃO DE CUMPRIMENTO DE REQUISITOS PARA CONCLUSÃO DO CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA.

Declaro para os devidos fins que Constantino Veríssimo dos Santos Filho**,** número de registro 2019705430, cumpriu todos os requisitos necessários para conclusão do curso de Especialização em Estatística, entregando a versão final do trabalho de conclusão de curso para seu orientador, o professor Fabio Nogueira Demarqui, que aprovou a versão final. O trabalho foi apresentado no dia 19 de julho de 2022 com o título "Um Aplicativo Web para análise de dados públicos".

Belo Horizonte, 17 de agosto de 2022

Prof. Roberto da Costa Quinino
Coordenador da Comissão
do Curso de Especialização
em Estatística

Prof. Roberto da Costa Quinino
Coordenador do curso de
Especialização em Estatística
Departamento de Estatística / UFMG

Dedicatória: aos meus pais, onde estiverem, nesta longa viagem, que é a vida, tenho imensa gratidão e amor. Sigam em paz.

AGRADECIMENTOS

Ao universo, pelo ensinamento contínuo e harmonia absoluta sobre a vida.

Aos meus filhos Gabriel e Maria Clara, pela compreensão da minha ausência física em muitos momentos durante esta caminhada.

À minha esposa Edléia, pela compreensão nos momentos em que estive distante.

A todos os colegas desta jornada que tornaram esse caminho mais leve.

À dupla Fernanda e Diego, pelas trocas de conhecimento, por me apresentarem a concepção do grupo sexta meia-noite.

Ao João Lucas, pela camaradagem, pelo poder de síntese e pelas discussões.

A todos os servidores do departamente de estatística, pelo acolhimento cortez.

Ao Professor Roberto, que sempre esteve presente nas demandas necessárias do curso.

Aos professores, gratidão pelo conhecimento compartilhado e pelo profissionalismo.

Ao Professor Fábio, pela disponibilidade durante todo o processo de orientação com os questionamentos e a parceria no decorrer desta monografia.

RESUMO

A aplicação do monitoramento e avaliação na gestão pública tem sido cada vez mais utilizada. Para tanto, ferramentas tecnológicas são utilizadas para identificação de possíveis melhorias na gestão, permitindo, desta maneira, que ações preventivas, ou mesmo corretivas, possam ser aplicadas em menor tempo e com maior segurança. Dentre as ferramentas, temos os *softwares* interativos, capazes de promover tanto a análise dos dados com maior segurança e utilizando uma gama cada vez maior de dados, quanto à visualização dos dados por meio de gráficos dinâmicos. O presente trabalho tem como objetivo propor um aplicativo interativo, amigável e que forneça informações suficientes para a promoção de estudos, tais como: análises descritivas e análise de agrupamento. O aplicativo foi construído usando o pacote *Shiny* do ambiente R de computação estatística, com a proposta de fornecer informações suficientes para contribuir na formulação de ações a serem aplicadas. A vantagem da utilização de um aplicativo criado neste tipo de concepção é que o usuário pode explorar as informações contidas nos dados, sem que seja necessário conhecimento de programação. No que tange à gestão, tem-se a possibilidade de comparação dos dados ao longo de um período, bem como a visualização do agrupamento de entidades que possuam características similares. Isso permite economia de recursos financeiros e de tempo para análise das razões das incidências de anomalias, propiciando a criação de estratégias de melhorias direcionadas por grupos de parcerias.

Palavras-chave: Análise de Agrupamento. Interatividade. Gráficos Dinâmicos. *Shiny*

ABSTRACT

Monitoring and evaluating results in public management has been constantly increasing. To this end, technological tools are used to identify possible improvements in management, thus allowing preventive or even corrective actions to be applied in less time and with greater security. Among those tools, there are interactive softwares which are capable of promoting data analysis with greater security and using an increasing range of data, as well as data visualization through dynamic graphics there are interactive software capable of promoting data analysis with greater security and using an increasing range of data, as well as visualizing data through dynamic graphics. The present work aims to propose an interactive and friendly application that provides sufficient information for the development of studies, such as: descriptive analysis and cluster analysis. The application was built using the Shiny package of the statistical computing environment R, with the purpose of providing enough information to contribute to the conceptualization of actions to be applied. The advantage of using an application created based on this structure is that the user can explore the information contained in the data, without the need for programming knowledge. With regard to management, there is the possibility of comparing data over a period, as well as viewing the grouping of entities that have similar characteristics. This allows financial and time savings because it enables the creation of improvement strategies directed by groups of partnerships as well as the analysis of the reasons for the incidence of anomalies.

Keywords: Cluster Analysis. Interactivity. Dynamic graphics. Shiny.

LISTA DE FIGURAS

SUMÁRIO

1 INTRODUÇÃO

Diariamente é gerada uma grande quantidade de dados[1] provenientes dos mais diversos ambientes, sendo a origem desses dados bastante diversa: redes sociais bem consolidadas, repositórios de mídia audiovisual, organizações governamentais (IBGE, Ipea, Tesouro Nacional, dentre outros) que, a cada dia mais, aprimoram a disponibilização de dados para os cidadãos. Em muitos casos, é possível verificar a transformação desses dados em informações direcionadas para os que desconhecem a linguagem estatística envolvida.

O aumento considerável dos dados está diretamente relacionado com o crescente desenvolvimento tecnológico, que possibilita o acesso a uma maior quantidade de pessoas, seja pelo fato de possuírem maior facilidade de uso, seja pelo acesso no que tange ao preço. Por trás da integração da tecnologia com o usuário, estão *softwares* comerciais que incorporam no seu conceito, cada vez mais, a filosofia que envolve a inteligência artificial.

Paralelo a toda essa avalanche tecnológica, é possível observar, mesmo que de forma tímida, o crescimento de comunidades de desenvolvedores de *softwares* de código aberto, que na sua maioria possuem a mesma, e até melhor, qualidade daqueles destinados ao uso comercial. Dentre os diversos *softwares* de código aberto têm-se os de ambiente estatístico, tais como o R, que teve a sua primeira fase de desenvolvimento realizada pelos professores Ross Ihaka e Robert Getleman, do departamento de estatística da University of Auckland, da Nova Zelândia (IHAKA e GENTLEMAN, 1996). Muitos estatísticos aderiram a essa comunidade de desenvolvedores, pois vislumbraram a oportunidade da criação de ferramentas capazes de utilizarem a interatividade com o usuário sem a perda dos fundamentos estatísticos.

É fato que existem excelentes ferramentas tecnológicas para o campo da análise de dados, bem como algumas que merecem cuidados. Porém, a ferramenta por si só não fornece as informações que se busca na análise exploratória. Conforme Tukey (1977), não é mais possível a continuidade dos procedimentos que se aplicavam à análise exploratória, que por analogia se parecia mais com um trabalho de detetive que não se detinha nem nas suas ferramentas de trabalho, tampouco no seu conhecimento sobre o que estava pesquisando. Desta maneira, quando se pretende realizar uma análise exploratória, é necessário apropriar-se devidamente

1 Na reportagem de 01/10/2015, a Forbes afirmou que "[...]até 2020 cerca de 1,7 megabyte de novas informações serão criadas por segundo para cada uma das pessoas no planeta. Disponível em: <https://forbes.com.br/fotos/2015/10/20-fatos-sobre-a-internet-que-voce-provavelmente-nao-sabe/#foto2> acesso em 18 jun 2022.

dos preceitos da temática em questão, fato que possibilita determinar qual ferramenta será a mais adequada para investigar os dados de maneira apropriada.

Assim sendo, para o melhor entendimento do objetivo desse trabalho é importante compreender alguns instrumentos jurídicos da administração pública que foram utilizados como referenciais, tais como: (i) Constituição Brasileira, 1988[2] no seu "*Art. 37. A administração pública direta e indireta de qualquer dos Poderes da União, dos Estados, do Distrito Federal e dos Municípios obedecerá aos princípios de legalidade, impessoalidade, moralidade, publicidade e eficiência"*; (ii) a Lei 13.019 de 2014[3], cuja finalidade é estabelecer o regime jurídico para regular as parcerias entre a administração pública e as organizações da sociedade civil em todo território nacional; e (iii) o Decreto Municipal nº 16.746[4] de 2017, instrumento onde estão o regime jurídico das parcerias entre a administração pública do Município de Belo Horizonte e as organizações da sociedade civil, doravante denominadas *parceiras*.

No Decreto Municipal 16.746 constam todas as diretrizes para a efetivação à implementação da Lei 13.019 no Município de Belo Horizonte, perpassando pelos aspectos da formalização das parcerias e da transparência das informações inerentes à elas; das prestações de contas; e das orientações aos órgãos e entidades da administração pública municipal, no que concerne à materialização e viabilização jurídica das parcerias com as organizações da sociedade civil.

Após a análise realizada dos termos jurídicos citados acima, e dos dados publicizados referentes aos gastos das parcerias firmadas com o município em questão, no âmbito da Secretaria Municipal de Educação, especificamente na rede própria, foi constatado a possibilidade de construção de uma ferramenta capaz de contribuir com o processo, já existente, do monitoramento e avaliação das parcerias.

Desta maneira, o objetivo deste trabalho foi desenvolver um aplicativo para subsidiar os gestores públicos no monitoramento e avaliação das parceiras para a detecção daquelas que necessitem de possíveis melhorias no seu processo da gestão financeira. Para tal, será realizada a análise exploratória dos dados referentes aos gastos com pagamentos de juros e multas, sendo mais um dos indicadores para os setores de monitoramento e avaliação das prestações de contas, o que corrobora na melhoria das ações preventivas já existentes e na

2 Constituição da República Federativa do Brasil, promulgada em 5 de outubro de 1988. Estabelece os princípios para instituir o Estado democrático Brasileiro. Disponível em: <http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm>. Acesso em: 03 Jul. 2022.

3 Lei nº 13.019, de 31 de Julho de 2014. Estabelece o regime jurídico das parcerias entre a administração pública e as organizações da sociedade civil. Disponível em <https://www2.camara.leg.br/legin/fed/lei/2014/lei-13019-31-julho-2014-779123-normaatualizada-pl.pdf>. Acesso em: 03 Jul. 2022

4 Decreto nº 16.746, de 10 de Outubro de 2017. Dispõe sobre as regras e procedimentos do regime jurídico das parcerias celebradas entre a administração pública municipal e as organizações da sociedade civil. Disponível em <http://portal6.pbh.gov.br/dom/iniciaEdicao.do?method=DetalheArtigo&pk=1185204>. Acesso em: 03 Jul. 2022

possível criação de novas metodologias que contribuam para minimização das causas. Considera-se o aspecto interativo propiciado pelo *Shiny*, em que torna-se possível, aos gestores, realizar a investigação dos dados contidos nos gráficos e tabelas. Outro aspecto desejado é a percepção da importância da garantia da qualidade dos dados fornecidos pelas parceiras, que poderá ser demonstrado nos encontros entre os gestores públicos e os representantes das parceiras.

A organização desse trabalho perpassa pela apresentação na Seção 1 do objetivo do trabalho, demonstrando a vantagem do uso do aplicativo e descrevendo os recursos utilizados para a criação dele. Na Seção 2, são descritos os conjuntos de dados e a respectiva análise exploratória, bem como a apresentação detalhada do aplicativo. Na Seção 3 é discutido o resultado dos testes no aplicativo. Na Seção 4, têm-se as considerações finais.

1.1 SOFTWARE ESTATÍSTICO R

O *Software* Estatístico R, doravante denotado como R, é uma linguagem e ambiente para computação estatística e gráficos, o qual incorpora as características de software livre, seguindo a tônica dos termos do projeto *GNU (General Public License) da Foundation Free Software*, na forma de código-fonte. Após a publicação de Ross Ihaka e Robert Getleman, que relatava sobre o R em 1996, surge nos meados de 1997 a equipe central (*R Core Team)*[5], a qual tinha acesso à fonte do R, e passa a trabalhar em sua melhoria, até que em 2000 tem-se a primeira versão, ocasião na qual assume característica de um projeto colaborativo internacional de desenvolvimento e pesquisa, passando a ser mantida pela *R Foundation,* e sua distribuição realizada no Comprehensive R Archive Network (*CRAN)*.

Em 2009 é fundada a empresa *Rstudio,* desenvolvedora do ambiente de desenvolvimento integrado (*IDE*)**,** homônimo à empresa, para a linguagem R, por *J.J.Allaire.* Contendo ferramentas robustas, o *Rtudio* é capaz de suportar as técnicas para a criação de análise de dados confiáveis e de alta qualidade.

1.2 PACOTES DO *SOFTWARE* ESTATÍSTICO R

5 Mais detalhes sobre a evolução do R pode ser obtido no site <https://www.r-project.org/>

O R é fornecido com alguns pacotes residentes, suficientes para a realização mínima de análises estatísticas. Em virtude do crescimento exponencial da comunidade do R, atualmente o número de pacotes já chega à casa dos milhares. Devido à quantidade expressiva de contribuintes envolvidos no aprimoramento dos pacotes do R e do perfil colaborativo da comunidade de *software* livre, a testagem dos pacotes é intensa, o que permite uma confiabilidade cada vez maior.

1.2.1 Pacotes do R utilizados no Aplicativo

1.2.1.1 Pacote *tidyverse*

O pacote *tidyverse,* segundo Wickham et al.(2019), tem a finalidade de contribuir com a arrumação dos dados. Na composição do seu núcleo, existe um conjunto de pacotes para atender as premissas da Ciência de Dados: *ggplot2, dplyr, readr, tibble, string, forcats, tidyr, e purrr,* que subsidiam nas ações de *importar, arrumar* e *visualizar dados*.

A sequência de apresentação dos pacotes encontrados no núcleo do *tidyverse* seguirá o modelo de ferramentas necessárias proposto pelos autores:

- *readr*: projetado para ler e analisar diversos tipos de dados retangulares originados de arquivos delimitados, rápido e amigavelmente. Dentre suas vantagens tem-se o fornecimento de relatório com informações dos problemas apresentados nos resultados anômalos durante análise dos dados.

  A seguir, destacam-se algumas funções utilizadas para realizar a leitura de arquivos específicos, do tipo retangulares de textos simples, em *data frames*: separados por vírgula (*read_csv())*; separados por tabulação (*read_tsv())*; que possui qualquer delimitador (*read_delim())*; cuja largura é fixa *(read_fwf());* que possui uma variação de arquivos**,** cuja largura é fixa, onde suas colunas são separadas por espaços em branco (*read_table());* cujo registro possui o estilo apache(*read_log())*.

  Exemplo de uso: *read_csv("data/dados_originais.csv")*

- *tibble*: são objetos originalmente do tipo *data frame* cujas características originais foram ajustadas, e que agora possuem as seguintes restrições: não alteram nomes ou tipos de variáveis; não fazem correspondência parcial; não modificam o tipo das entradas (por exemplo, nunca converte *strings* em fatores); e não criam nomes de linhas. Em um *tibble*: os nomes dos cabeçalhos podem utilizar termos "não sintáticos", ou seja, é possível criar nomes das variáveis com caracteres incomuns, como um espaço, números etc.; porém tais nomes devem ser cercados por acento grave. Exemplo: `1500`.

  Quando se deseja transpor os dados do *tibble*, o termo utilizado para classificá-lo é *tribble*, onde os cabeçalhos são iniciados por ~ , e as entradas são separadas por vírgula, como pode ser observado na Figura 1.

Figura 1 - Exemplo Tribble()

```
tribble(
  ~x, ~y, ~z,
  #--|--|----
  "a", 2, 3.6,
  "b", 1, 8.5
)
#> # A tibble: 2 x 3
#>   x         y     z
#>   <chr> <dbl> <dbl>
#> 1 a         2   3.6
#> 2 b         1   8.5
```

Fonte: Livro R para Ciência de Dados

Outros dois aspectos de destaque são: impressão e subconjuntos. *Impressão*: *Tibbles* tem um método de impressão refinado que mostra apenas as primeiras 10 linhas e todas as colunas que cabem na tela. Isso facilita lidar com dados grandes. Também é possível a utilização da função *print()*, com os descritores: n = quantidade de linhas e *width*, onde *width = Inf* exibirá todas as colunas. A função *options(tibble.print_max = n, tibble.print_min=m)* determina a impressão de colunas. *Subconjuntos*: antes do *tibble*, as ferramentas utilizadas trabalhavam com um *data.frame* completo. Com esta nova função, para acessar uma única variável no *tibble***,** tem-se as ferramentas $ e [[, onde $ extrai a variável apenas pelo nome e [[extrai pelo nome e pela posição, vide Figura 2.

Figura 2 - Exemplo - Subconjunto do tibble()

```
df <- tibble(
  x = runif(5),
  y = rnorm(5)
)

# Extract by name
df$x
#> [1] 0.73296674 0.23436542 0.66035540 0.03285612 0.46049161
df[["x"]]
#> [1] 0.73296674 0.23436542 0.66035540 0.03285612 0.46049161

# Extract by position
df[[1]]
#> [1] 0.73296674 0.23436542 0.66035540 0.03285612 0.46049161
```

Fonte: Livro R para Ciência de Dados

- *dplyr*: é formado por funções que dão suporte nas situações diretamente relacionadas à manipulação de dados, tais como *mutate()*: cria variáveis como funções de variáveis já existentes; *select()*: seleciona variáveis por meio dos seus nomes; *filter()*: seleciona observações por meio de seus valores; *arrange()* ordena as linhas; *group_by()*: realiza operações por grupos; *summarise()*: em conjunto com a função *group_by()* constrói resumos agrupados, podendo ser compostos pelas medidas de localização, dispersão, classificação, posição**,** bem como pela contagem: *count()*, e valores lógicos: *mean(y==0)*. Existem três famílias de verbos projetados para trabalharem com dados relacionais como *Mutating joins*: verbos que adicionam novas variáveis a um *data.frame***,** a partir de observações que possuem correspondência num outro *data.frame; Filtering joins*: realizam filtros nas observações que se encontram num *data.frame*, considerando se combinam ou não com uma observação presente num outro *data.frame; Set operations*: as observações são tratadas como se fossem um conjunto de elementos. A família de verbos que foi mais utilizada na preparação dos dados desse trabalho foi a *Mutating joins.* As funções que a compõem, além de fazerem a junção dos dados de um par de tabelas, a partir da combinação de observações por suas *keys*, adicionam as variáveis à direita no *data.frame.* As funções dessa família são: *left_join*(); *right_join(); inner_join(); full_join()*.

- *tidyr*: fornece subsídios para a obtenção dos dados arrumados, cujas características são: cada coluna é uma variável; cada linha é uma observação; cada célula é um único valor. De fato, a análise de dados organizados possui mais celeridade em relação aos dados desorganizados. Assim, esta é uma tarefa que deve ser contínua na análise de dados. As funções relacionadas com a pivotagem buscam a adequação dos bancos de dados, os quais foram preparados para facilitar a entrada de dados num determinado software, e que, por vezes, causam entraves no tratamento dos dados. A família de

pivotagem corresponde às funções: *pivot_longer()*, que reduz o número de colunas e aumenta o número de linhas; *pivot_wider()*, que faz o oposto.

- *purr*: consiste num aprimoramento das ferramentas que compõem a *programação funcional (FP) do R,* em relação ao trabalho com funções e vetores.

- *stringr*: composto por ferramentas para manipulação de *strings*, cujas finalidades perpassam pela determinação do tamanho; combinação de duas ou mais *strings*; detecção; visualização; pesquisa de expressões regulares; determinação do maiúsculo e minúsculo em função da localidade; classificação em função da localidade; pesquisa da *string* à partir de um referencial; substituição; divisão; quantidade e extração de correspondência; regras de agrupamento; e pesquisa de objetos comuns no ambiente global.

- *forcats*: trata-se de variáveis categóricas, que possuem grupo de valores. Composto por ferramentas que alteram a ordem dos níveis ou dos valores.

- *ggplot2*: é um pacote criado para produzir gráficos estatísticos ou de dados, possui a peculiaridade de construir gráficos utilizando o conceito da gramática de gráficos, concebida por Leland Winkinson (WICKHAM, 2016). Esta se caracteriza pela composição de um conjunto de componentes independentes, os quais também podem ser compostos de diversas maneiras. Essas características tornam o *ggplot2* capaz de criar gráficos que atendam a especificidade de cada usuário. Ele pode funcionar de forma interativa como iniciar a demonstração gráfica dos dados na forma bruta e, por meio de camadas, adicionar preceitos estatísticos na forma de anotações ou resumos. É uma ferramenta capaz de estabelecer uma conexão do gráfico planejado pelo analista de dados à sua manifestação na tela.

1.2.1.2 Pacote plotly

Conforme *Sievert* (2020), o pacote *plotly* é alimentado pela biblioteca *plotly.js* (*JavaScript*), o que permite uma interface direta entre a função *plot_ly*() e *plotly.js*, pacote que per-

mite a criação de gráficos interativos. O *plotly* possui conexão com o *ggplot2*, por meio da função *ggplotly()* que converte os gráficos estáticos em uma versão interativa baseada na web. Essa função possui também argumentos ditos como de alto nível, por exemplo, a *dynamicTicks* que dialoga com a *plolty.js* no intuito de recalcular os eixos dos gráficos dinamicamente, quando for o caso. Outro aspecto importante é a possibilidade da utilização de qualquer função do pacote R, tais como *layout(), add_traces()*. Outra função importante neste pacote é a *style()* utilizada para modificar os atributos de dados.

1.2.1.3 Pacote *Shiny*

Shiny é um um pacote do R, cuja versão beta foi apresentada em julho de 2012 na *Joint Statistical Meetings (JSM)*, e o lançamento da primeira versão foi disponibilizada em dezembro**,** do mesmo ano, por um grupo de profissionais da *Rstudio* (Chang et al., 2021). Destaca-se, dentre eles, *Winston Chang*, um dos autores [*aut*] e o criador [*cre*] deste pacote destinado para estatísticos e cientistas de dados.

Shiny é um *framework* que possibilita**,** aos desenvolvedores do R**,** a criação de aplicações interativas na *web*, não sendo necessário ao programador ter conhecimento de *HTML, CSS ou Javascript*. Ele possui um conjunto de funções destinadas a promover a interface com o usuário (*ui*), onde são coletadas informações fornecidas pelo mesmo, e enviadas para o *server*, cujo papel é processar as informações coletadas e retorná-las para *ui* , que dará o *feedback* ao usuário. O detalhamento de instalação de um aplicativo nas nuvens não será discutida neste trabalho, mas as orientações de como realizá-la será disponibilizado no endereço eletrônico[6].

1.2.1.3.1 Estrutura do *Shiny*

6 Neste livro eletrônico também será possível aprender sobre o funcionamento do shiny, disponível em: <https://mastering-shiny.org/>, acesso em 08 ago. 2022.

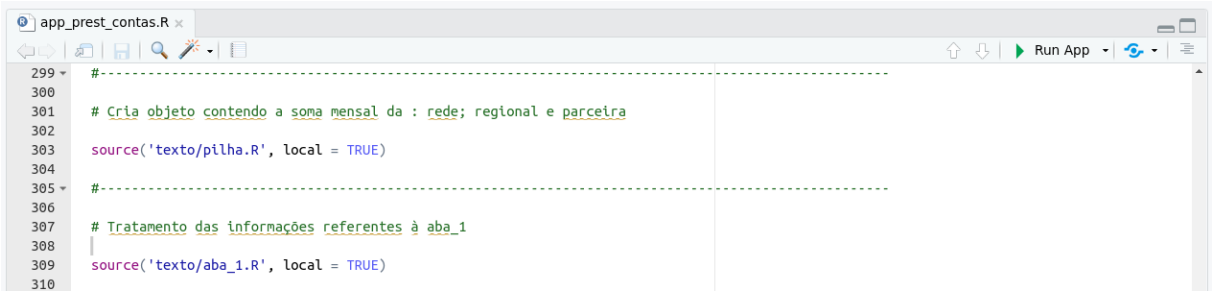
Como já mencionado acima, existem duas funções: *ui* (*User Interface*) e *server* (*work*), que podem ser escritas num mesmo *script* ou serem separadas em dois *scripts*. Se a escolha for criá-las separadamente, deve-se armazená-las em dois arquivos distintos com os seguintes nomes: *ui.R* e *server.R*.

Os parâmetros criados na função *ui s*ão transformados em *HTML,* a tela *web* criada para o aplicativo é compartilhada com o usuário. A função *server*, por sua vez, possui a missão mais complexa, que é trabalhar a informação colhida pela *ui* e retornar, de maneira particular, o resultado.

As ações mencionadas acima podem ser visualizadas no *script* de criação do *aplicativo_teste.R* no APÊNDICE A, cujo objetivo é a construção de um histograma denominado por "*hist*", o qual é estruturado utilizando a amostragem normal criada pela função *rnorm()*, sendo o quantitativo de elementos definido pelo usuário na variável *n*.

Para executar o *script* pode-se clicar no botão "Run App", conforme Figura 3, sendo também possível fazer esta ação via linha de comando do R.

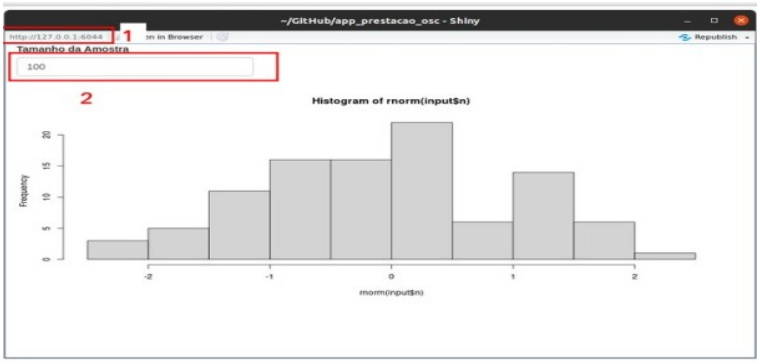
Figura 3 - Recorte RStudio (Run App)



Fonte: Elaborado pelo autor

O resultado desse *script* pode ser visualizado na Figura 4. Nela é possível observar a interação com o usuário**,** o local onde o usuário informa a quantidade de elementos da amostra e a apresentação do histograma. O usuário, quando acessar o *aplicativo_teste.R,* observará que este foi previamente programado para iniciar o processo com a quantidade de 100 elementos, vide item 2 da Figura 4, quantidade que pode ser alterada a qualquer momento pelo usuário.

Figura 4-Resultado do aplicativo_teste.R



Fonte: Elaborado pelo autor

1.2.1.3.2 Programação Reativa

Conforme Wickham (2021), a lógica que envolve os cálculos executados na função *server* é expressa pela Programação Reativa, da qual a essência está em especificar a lógica que construirá um "gráfico de dependências" para garantir que quando ocorrer a modificação de um *input*, todas os *outputs* que possuem relação com ela sejam automaticamente atualizados.

Quando se observa as atividades realizadas pelas funções *ui* e *server*, percebe-se que *server* possui maior complexidade, pois, além dos cálculos, ela precisa fazer um entrega de versão independente para o usuário. Quando se utiliza o aplicativo na *web*, o usuário deve ter a sensação de ser o único naquele momento.

O objeto *input* recebe as entradas vindas do navegador, onde está sendo rodado o aplicativo. Esse objeto atua como uma lista que contém todas as entradas devidamente identificadas por um ID. Existem tipos de *input* diferenciados para o recebimento da entrada. Se esta for numérica, por exemplo, usa-se o *numericInput(variável, label, ...)*. A leitura do objeto *input* acontece num ambiente reativo, criado por funções do tipo *render** ou *reactive(),* ambas ficam localizadas no interior da função *server.*

Existem diversas funções do tipo *render** que possuem finalidades distintas, mas com um mesmo cerne que é capturar uma expressão R e realizar um pré-processamento, conforme *Xie(*2022). O *Shiny*, ao detectar uma ou mais alterações solicitadas pelo usuário, encaminha os pedidos para a função *render** que realiza o processamento e retorna uma resposta usando a

relação entre o **Output* e a respectiva *render**. Para o melhor entendimento desta função, segue o exemplo do funcionamento da Aba “Análise Descritiva”, Figura 7,do aplicativo construído neste trabalho. Ao ser aberta, ela inicia com todas as regionais escolhidas, vide Figura 5.

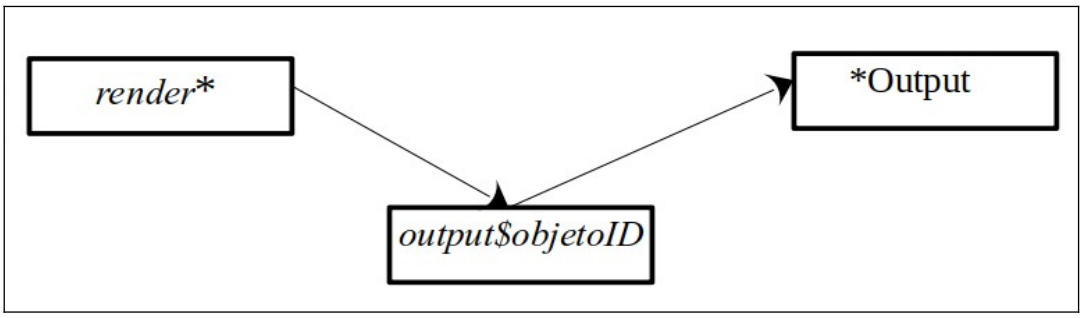
Figura 5-Tela de Escolha das Regionais - Aba "Análise Descritiva"

Fonte: Elaborado pelo autor

Ao abrir esta aba, além deste conjunto de regionais, estão os gráficos BoxPlot e uma tabela de dados (DataTable) demonstrando todos os gastos realizados por regional. Quando o usuário desmarcar alguma regional, o *Shiny*, que a cada microsegundo verifica se alguma alteração foi realizada, ativa as funções reativas relacionadas com esta aba, porém, as ações dessas funções não serão discutidas neste momento, uma vez que a intenção, neste exemplo, é demonstrar o funcionamento das funções *render**. No esboço do script da Figura 7 existem dois tipos dessas funções**,** que são ativadas quando o usuário realiza uma modificação na escolha das regionais, são elas: *renderDataTable()* e *renderPlotly*, tendo respectivamente a finalidade de gerar a estrutura de saída de uma tabela de dados (DataTable) e de um gráfico(s), ambos reativos.

Na figura 6, é possível observar como a atualização que uma função render pode ser visualizada pelo usuário**:** (i) formatação dos dados atualizados (*render**); (ii) armazenamento da formatação dos dados num objeto (*output$objetoID)*; (iii) transporte do objeto com a formatação gerada no item (i) para ser exibida ao usuário (**Output)*.

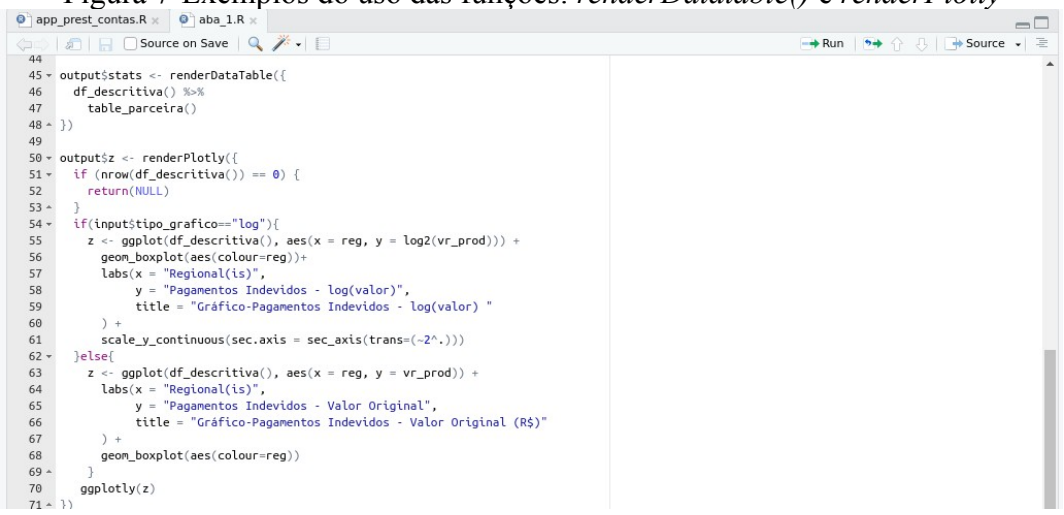
Figura 6-Fluxo de Atualização da Saída - Funções render*



Fonte: Elaborado pelo autor

Considerando a descrição do fluxo acima e o esboço do *script***,** na Figura 7 para a atualização da Tabela de Dados (Data Table), pode-se observar que: (i) a função de renderização é a *renderDataTable()*; (ii) o objeto criado para armazenar os critérios da renderização é *"stats"*, logo tem-se *output$stats;* a função que exibirá o objeto "*stats"* é *dataTableOutput("stats").* No que diz respeito à atualização dos gráficos BoxPlot, temos: (i) a função de renderização é a *renderPlotly()*; (ii) o objeto criado para armazenar os critérios da renderização é *"z"*, logo tem-se *output$z;* a função que exibirá o objeto "z*"* é *plotlyOutput(z)*. Um aspecto importante é que as funções *renderDataTable()*, *output$stats, renderPlotly()* e *output$z* estão no interior da função *server,* Figura 7, enquanto as funções *dataTableOutput("stats")* e *plotlyOutput(z)* estão no interior da função *ui,* Figura 8.

Figura 7-Exemplos do uso das funções: *renderDatatable()* e *renderPlotly*



Fonte: Elaborado pelo autor

Figura 8-Exemplo de uso das Funções: *dataTableOutput()* e *plotlyOutput(z)*

Fonte: Elaborado pelo autor

Como citado anteriormente, existem outras funções *render** no *Shiny*, que são apresentadas no Quadro 1, juntamente com as respectivas funções ***Output* correspondentes e suas finalidades.

Quadro 1 – Exemplos de funções *render**

Função *server*	**Função *ui***	**Finalidade**
renderDataTable()	dataTableOutput()	DataTable – cria uma tabela de dados reativa
renderImage()	imageOutput()	Images
renderPlot()	plotOutput()	Plots
renderTable()	tableOutput()	Data frames, matrix outras tabelas como estruturas
renderText()	TextOutput()	Cadeias de strings
renderText()	VerbatimTextOutput()	Excelente para apresentar texto formatado, como código.

Fonte: Tutorial do *Shiny* no Rstudio (modificado)[7]

É possível que algumas rotinas do *Shiny* necessitem dos valores gerados por uma determinada função. Com intuito de evitar a repetição de um mesmo processamento dessas funções no *server,* é prudente a utilização da função *reactiveEvent({}),* pois ao executá-la, na primeira vez, o resultado é armazenado na memória *cache.* Para a utilização do resultado**,** que está na memória cache, basta acionar o objeto reativo gerado pela função *reactiveEvent({}),* enquanto não houver uma nova atualização.

7 Tutorial do Shiny. disponível em:<https://shiny.rstudio.com/tutorial/written-tutorial/lesson4/>. acesso em 08 ago. 2022.

2 MATERIAL E MÉTODOS

2.1 MINERAÇÃO DO DADOS

Os dados que serão utilizados no aplicativo fazem parte de uma base de dados formada pelos lançamentos de todos os gastos das parceiras, no período de 2017 a 2020. Dentre os diversos tipos de gastos, optou-se por utilizar**,** no aplicativo**,** os dados referentes aos pagamentos de juros e ou de multas desse período. Tal escolha deve-se ao fato de que todos os gastos seguem um planejamento prévio, que é fornecido pelo poder público quando envia a receita para a parceira. Desta maneira, os pagamentos de juros e multas, quando ocorrem, na maioria das vezes devem-se às duas razões: (i) quando o poder público não repassa o valor pactuado para as parceiras ou (ii) por uma falha na gestão da parceira que, de posse da receita, não realiza o pagamento em tempo hábil.

Desta maneira, os dados selecionados para serem estudados no aplicativo correspondem ao tipo (ii), em função do objetivo do trabalho; sendo isso possível em função da identificação de cada um dos lançamentos que não atenderam à diretriz contida nos termos jurídicos já citados. A identificação desses tipos de gastos é feita com a indicação de 1 dos 8 tipos de inconsistências disponíveis ao usuário quando o lançamento é realizado no sistema: 1-Pagamento de tarifas não permitidas (bancárias e outras); 2-Pagamento de impostos com multa/juros; 3-Falta de orçamento; 4-Quantidade insuficiente de orçamentos; 5-Pagamento à maior que o portal de preços e/ou orçamentos; 6-Comprovante de despesa referente aos juros, multas ou pago à maior; 7-Aquisição realizada sem o comprovante de despesa; 8-Despesas indevidas à parceira.

O resultado do processo de filtragem**,** realizada pelo sistema utilizando os indicadores acima**,** foi armazenado numa planilha eletrônica de extensão *.xls*, contendo 22 (colunas) e 18.150 linhas. De posse dos dados, fez-se necessário a *mineração de dados*, que corresponde a processos utilizados para realização da análise exploratória nos dados para descobrir padrões que corroborem a detecção das informações**,** que possam consubstanciar o conhecimento que servirá de suporte para as tomadas de decisões necessárias, como afirmam Silva, L. A.; Peres, S. M.; Boscarioli, C. (2016).

Na *mineração de dados*, optou-se pelo uso do formato de *dados arrumados* caracterizado pela padronização na estruturação dos dados, o que facilita o trabalho do analista de dados, quando se busca pelas informações que necessita. Para garantir que a base de dados está enquadrada neste formato, é necessário observar: "(i) cada variável forma uma coluna ; (ii) cada observação forma uma linha; (iii); cada valor deve ter a sua própria célula." (WICKHAM, 2014, p.4).

Considerando as diretrizes acima, foi elaborado um script com o nome de *prepara_dados.R* cujo algoritmo realiza os seguintes passos:

1. Criação de um objeto, de nome *matriz_risco,* para receber os dados da planilha gerada;
2. Alteração dos nomes das colunas do objeto *matriz_risco*: antes existiam nomes longos, com acentos e espaços entre as palavras, mas todos foram substituídos por nomes mais curtos; e os que eram compostos, além da substituição dos nomes, foram separados por *underlines,* também conhecidos por *underscore ou subtraço*;
3. A variável *incons* (inconsistências) continha dados alfanuméricos agrupados identificando o(s) tipo(s) de inconsistência(s) - *Exemplo*: *[1][2][3][5].*
    - A variável *incons* foi decomposta em 8 novas variáveis que ficaram à direita da última coluna do *data.frame*: inc_1; inc_2; inc_3; inc_4; inc_5; inc_6; inc_7; inc_8. Isso foi possível com o uso da função *mutate()*, que provém do pacote *dplyr,* em conjunto com a função *str_count()*, do pacote *stringr,* utilizada para armazenar**,** nas novas variáveis**,** a quantidade de lançamentos referente a cada uma.
4. A partir da variável *emissao* (data de emissão)**,** foram criadas 2 novas variáveis: ano e mês;
5. Criação de um novo objeto *df*: o objeto *matriz_risco* contém dados que estão ligados aos 8(oito) tipos de inconsistências já apresentados, incluindo aqueles relacionados com os gastos**,** *juros e multas*. Desta maneira, será gerado um novo objeto denotado por *df***,** a partir do *matriz_risco,* porém, contendo apenas informações relacionadas com *juros e multas***,** que serão selecionadas quando da transferência, utilizando as funções *filter()* e *str_detect(),* dos pacotes *stringr* e *dplyr* respectivamente;
6. No último passo de preparação dos dados, foram realizadas as seguintes ações:
    a) Sorteio de números aleatórios sem reposição, usando a função *sample(),* na mesma quantidade de parceiras da rede própria, que servirão para proteção dos dados das parceiras, possibilitando a retirada dos respectivos nomes;

b) Sorteio de números aleatórios sem reposição, usando a função *sample(),* na mesma quantidade de regionais do município de Belo Horizonte, que servirão para proteção dos dados das parceiras, possibilitando a retirada da identificação das regionais, passando a ser representadas em ordem aleatória e com a composição do nome: reg_*número-sorteado*;

c) Geração do objeto *parceiras*, criado a partir da planilha *parceiras.xlsx*, que possui as variáveis *reg* e *parceiras*. Foi inserida uma variável, *id_p,* que passa a ser utilizada como identificador das parceiras, quando recebe os números criados no item (a). Assim, o objeto *parceiras* passa a ser composto pelas seguintes variáveis: *reg, parceira* e *id_p*;

d) O objeto *parceiras* contribuirá na geração/complementação de três objetos:

- *parceiras_pilha*: todos os dados do objeto *parceiras* são transferidos, exceto a variável *parceiras*, utilizando o comando *select(-parceiras)*). Ressalta-se que a utilização da função *select()* não desfaz a estrutura do objeto. A sua funcionalidade possibilita manipular o objeto, neste caso, com a retirada de uma variável, e utilizando esse resultado para execução de diversas ações;
- *parceiras_df*: todos os dados do objeto *parceiras* são transferidos, exceto a variável *reg* , utilizando o comando *select(-reg)*);
- *df*: Os dois objetos, *df* e *parceiras_df*, foram concatenados usando a função *left_join(),* tendo a variável *parceira* como a *key* para interligá-los. As variáveis do objeto *parceiras_df* serão inseridas à direita da última coluna do objeto *df.*

Após a aplicação do algoritmo da preparação dos dados, os objetos *df* e *parceiras_df* foram gravados em arquivos com o mesmo nome de cada objeto e com extensão *.RDS,* formato de arquivo binário nativo do R, compactando os dados e diminuindo o tempo para leitura.

Tais arquivos foram a base para a elaboração de funções reativas que atenderam especificidades do aplicativo em cada aba. As análises realizadas na aba_2 (“Análise Gastos Parceiras versus Regiões”) e na aba_3 (“Análise Gastos entre Regiões”), possuem como objetivo, respectivamente na linha do tempo, o comportamento do gasto médio com juros e multas: (i) de uma parceira em relação a sua regional e em comparação com todas parceiras da rede municipal de educação; (ii) entre todas as regionais e com o gasto médio de todas parceiras da rede municipal de educação.

Para cada aba**,** foi elaborado um *script* contendo expressões reativas do *Shiny* com o objetivo de criar para cada mês/ano dos gastos, três camadas, cada uma como uma *observação,* contendo os dados: (i) média dos gastos de uma parceira; (ii) média dos gastos da regional onde está localizada a parceira; e (iii) média dos gastos de todas as parceiras da rede municipal de educação. Esses dados foram armazenados num objeto denotado como *pilha.* A identificação das camadas é realizada pela variável *nivel,* sendo nela armazenado, em cada linha, a sua identificação: *id_p (id da parceira); regiao; ou rede.* Este formato de estrutura, segue a concepção de redução de colunas e aumento de linhas, facilitando ao analista de dados a localização mais rápida das variáveis necessárias**,** em função do padrão na estruturação da base de dados, conforme Wickham (2019).

2.2 UM APLICATIVO *WEB* PARA ANÁLISE DE DADOS PÚBLICOS[8]

Para o desenvolvimento do aplicativo web, produto deste trabalho, foram seguidos os pressupostos de Knaflic (2019, p.17) quando afirma que "[…] *antes de começar a criar uma apresentação ou comunicação de dados, a atenção e o tempo devem estar voltados a entender o contexto a necessidade de se comunicar*". Ressalta-se, também, a importância da identificação do público-alvo, pesquisando no intuito de apreender as informações necessárias que possibilitarão uma comunicação mais limpa e objetiva.

Desta maneira, o primeiro passo foi entender o público-alvo, neste caso os gestores públicos com a função de promover a melhoria na tomada de decisões na gestão das parceiras. O aplicativo foi concebido para apresentar os dados fornecidos pelas parceiras no formato de gráficos e tabelas, possibilitando dessa maneira indicar possíveis fragilidades das parceiras, bem como fornecer parâmetros aos gestores para embasar a tomada de decisões necessárias.

2.2.1 Estrutura do Aplicativo

O aplicativo foi desenvolvido utilizando o *Shiny*, onde o formato da página contendo uma barra de navegação de nível superior (navbarPage()) com um conjunto de painéis separa-

8 Acesso ao aplicativo: https://vsfconstantino.shinyapps.io/app_prestacao_osc/

dos (*tabPanel( )*)**,** referentes a 5(cinco) abas temáticas: Análise Descritiva; Análise Gasto Parceria versus Regiões; Análise Gasto entre Regiões; Análise de Agrupamentos; Análise das Justificativas.

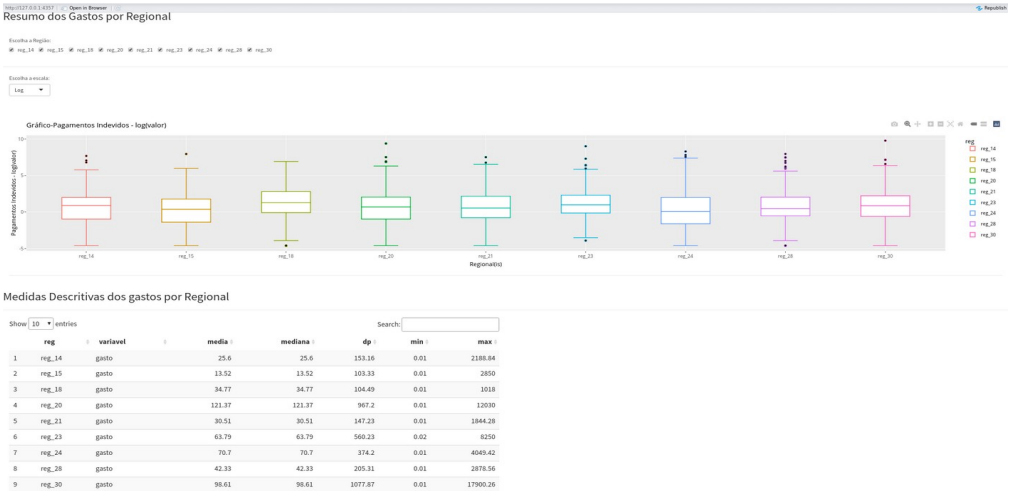
Em cada aba, tem-se o contato com a interatividade e reatividade, características do *Shiny*. As ações inerentes a essas duas ferramentas podem ser percebidas pelo usuário quando solicita um ou mais itens específicos que serão obtidos ao posicionar o ponteiro do mouse sobre a opção do filtro desejada, e clicar. A resposta a essa ação é percebida instantaneamente na tela. Outro aspecto de visualização é a exibição de informações ao deslizar o ponteiro do mouse sobre o gráfico (*mouseover)*.

Os dados utilizados no aplicativo são provenientes dos lançamentos dos gastos realizados pelas parceiras**,** no período de 2017 a 2020. Na planilha de texto (*csv*) existiam 22 colunas e 18.149 linhas, compreendendo respectivamente a quantidade de variáveis e lançamentos de gastos com insumos. O objetivo do quantitativo de variáveis era fornecer o máximo de informação dos dados inerentes aos comprovantes de despesas, tais como nome do fornecedor, produto, data de emissão do comprovante fiscal, data do pagamento, tipo de pagamentos, dentre outros. Segue abaixo a descrição de cada aba que compõe o aplicativo:

2.2.1.1 Aba 1 - Análise Descritiva

Nesta aba têm-se as medidas descritivas dos gastos com juros e multas por região, no período compreendido entre os anos de 2017 a 2020, conforme a Figura 9. A apresentação dos dados está disposta em dois ambientes: visualização gráfica - gráficos *boxplot,* que faz parte do pacote *ggplot2* (WICKHAM, 2016)**,** e na parte inferior estão os resumos.

Figura 9 - Estudo da Análise Descritiva

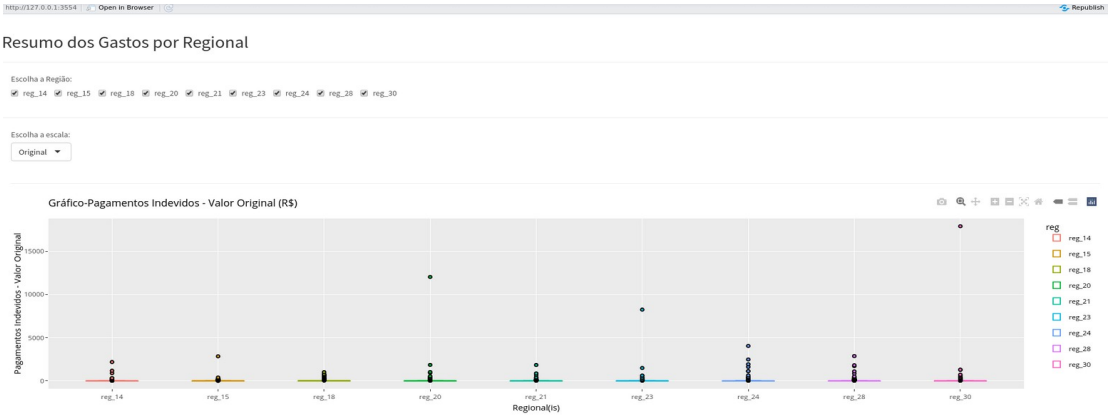


Fonte: Elaborado pelo autor

Na parte superior desta aba é possível detectar um grupo de caixas de seleção, cujo título é "Escolha a Região", que possibilita selecionar as opções das regionais de forma independente. O *input* para a confecção, tanto dos gráficos**,** quanto do quadro resumo que contêm as medidas da Estatística Descritiva, consiste em um vetor de caracteres formados pelos valores selecionados do grupo de caixas de seleção, indicando qual(is) regiões serão utilizadas, quando for o caso, para comparações.

O *Shiny* (CHANG et.al., 2021) permite a visualização gráfica dos dados, tanto na escala "original", quanto na "log". Isto foi necessário, pois na escala original dos dados**,** a visualização das medidas descritivas é comprometida porque a maioria dos valores dos gastos é próxima de zero, enquanto outros se aproximam da casa dos milhares ou estão nela, fatos que podem ser comprovados visualizando a Figura 10.

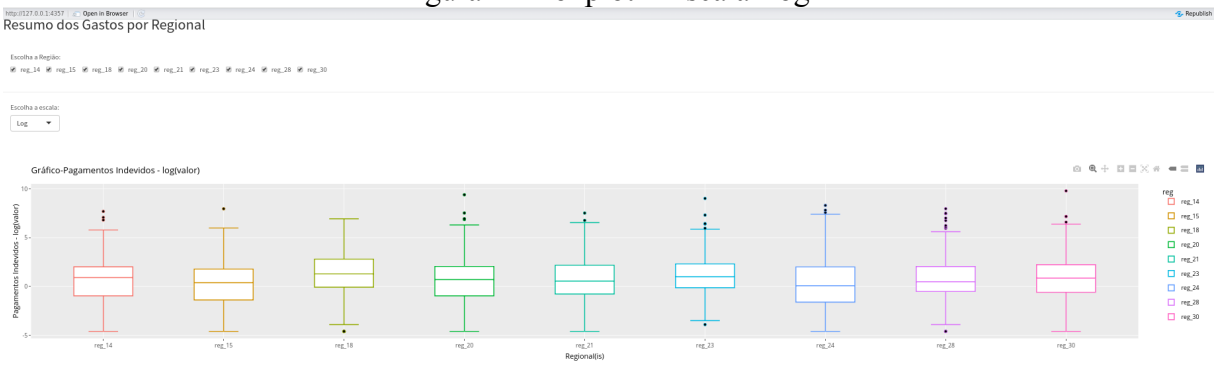
Figura 10-Boxplot – Valores Originais



Fonte: Elaborado pelo autor

Para visualização das duas escalas foram criados dois gráficos, um para cada tipo, que são acessados quando se escolhe o seletor “Escolha a escala”, localizado acima do gráfico, à esquerda. A melhoria na visualização é significativa, como pode ser visto na Figura 11.

Figura 11-Boxplot – Escala Log



Fonte: Elaborado pelo autor

Comparando as Figuras 10 e 11 é possível constatar que a escala “log” contribuiu para melhoria na visualização gráfica das medidas estatísticas, permitindo a comparação visual das regionais. Em complementação, na Figura 12 existe um quadro de resumo composto pelas medidas: média, mediana, desvio-padrão, valor mínimo e valor máximo. As constatações visuais, juntamente com os dados numéricos disponíveis, fornecem aos gestores maior segurança para fundamentar as tomadas de decisões.

Figura 12-Medidas Descritivas

Medidas Descritivas dos gastos por Regional

Show 10 entries Search:

	reg	variavel	media	mediana	dp	min	max
1	reg_14	gasto	25.6	25.6	153.16	0.01	2188.84
2	reg_15	gasto	13.52	13.52	103.33	0.01	2850
3	reg_18	gasto	34.77	34.77	104.49	0.01	1018
4	reg_20	gasto	121.37	121.37	967.2	0.01	12030
5	reg_21	gasto	30.51	30.51	147.23	0.01	1844.28
6	reg_23	gasto	63.79	63.79	560.23	0.02	8250
7	reg_24	gasto	70.7	70.7	374.2	0.01	4049.42
8	reg_28	gasto	42.33	42.33	205.31	0.01	2878.56
9	reg_30	gasto	98.61	98.61	1077.87	0.01	17900.26

Showing 1 to 9 of 9 entries Previous 1 Next

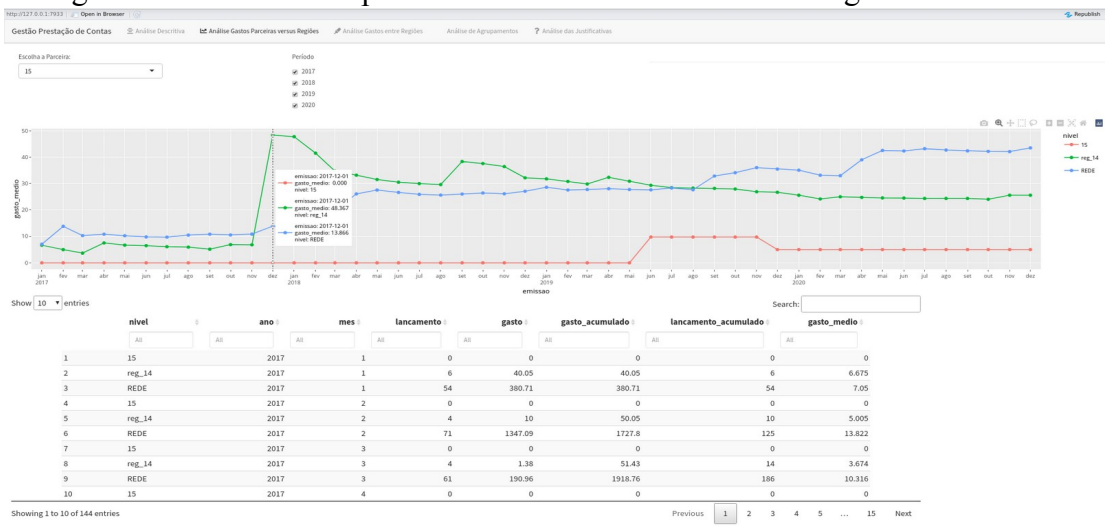
Fonte: Elaborado pelo Autor

2.2.1.2 Aba 2 - Análise Gastos Parcerias versus Regiões

A Figura 13 possibilita a comparação das médias mensais de gastos indevidos de uma parceira referência *versus* as médias mensais de gastos indevidos de todas as parceiras que estão na mesma região onde está localizada *versus* as médias mensais de gastos indevidos de toda a rede de parceiras do município.

Na parte superior existem dois grupos de botões para seleção de referenciais para estabelecer os dados que serão utilizado na comparação: i) Escolha da parceira usando uma lista de seleção (*selectInput*); ii) Seleção de um ou mais anos, no período de 2017 a 2020, por meio de um grupo de caixas de seleção (*checkboxGroupInput*), sendo possível selecionar opções de forma independente. Para realizar a comparação dos dados, existe um gráfico composto por três linhas que representam cada nível utilizado para comparação**,** e uma tabela, conforme a Figura 13.

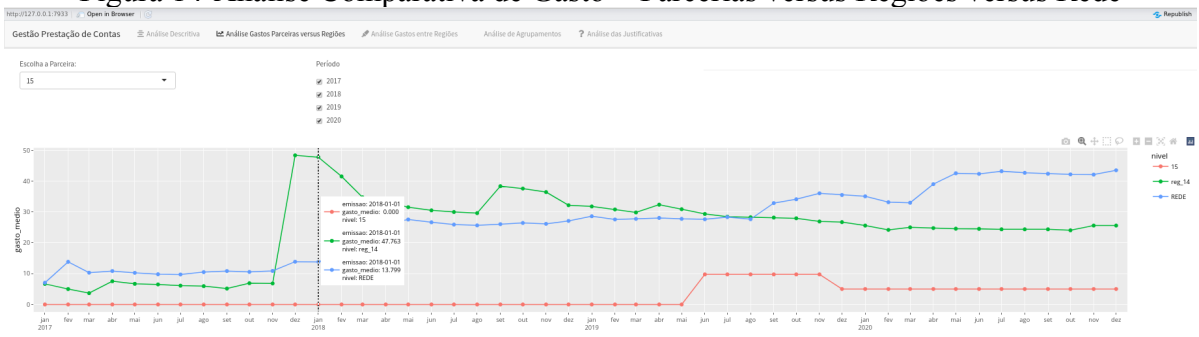
Figura 13 - Análise Comparativa de Gasto Parcerias versus Regiões versus Rede



Fonte: Elaborado pelo autor

A Figura 14 apresenta o gráfico que aparece na parte superior da aba, cuja finalidade é a mesma da tabela que aparece na parte inferior da aba, ambos podem ser vistos na Figura 13. O aspecto de visualização da comparação por meio de gráficos tende a fazer mais sentido quando se busca localizar a região temporal com maior discrepância do gasto médio, seja por parceira, região ou de todo município.

Figura 14-Análise Comparativa de Gasto - Parcerias versus Regiões versus Rede



Fonte: Elaborado pelo autor

Na Figura 15, encontra-se a tabela que está na parte inferior da aba, que possui a mesma finalidade do gráfico, comparação numérica mensal dos gastos médios entre os três níveis, os quais correspondem: primeira linha à parceira, a segunda a todas parceiras da região e a terceira a todas as parceiras do município. A tabela é composta pelas variáveis: "nivel"; "ano"; "mês"; "lancamento": quantitativo total de lançamentos de gastos, de cada nível, no mês; "gasto": total de gastos, de cada nível, a cada mês; "gasto_acumulado": total de gastos, de cada nível, acumulado desde janeiro de 2017 até dezembro de 2020; "lancamento_acumulado": quantitativo total de lançamentos de gastos, de cada nível, acumulado desde janeiro de 2017 até dezembro de 2020; "gasto_medio": resultado da divisão "gasto_acumulado" / "lancamento_acumulado", a cada mês. Uma característica a ser ressaltada nesta tabela é a possibilidade de filtragem, que pode ser feita em todas as variáveis.

Figura 15-Análise Comparativa de Gasto Parcerias versus Regiões versus Rede

Show 10 entries Search:

	nivel	ano	mes	lancamento	gasto	gasto_acumulado	lancamento_acumulado	gasto_medio
	All	All	All	All	All	All	All	All
1	15	2017	1	0	0	0	0	0
2	reg_14	2017	1	6	40.05	40.05	6	6.675
3	REDE	2017	1	54	380.71	380.71	54	7.05
4	15	2017	2	0	0	0	0	0
5	reg_14	2017	2	4	10	50.05	10	5.005
6	REDE	2017	2	71	1347.09	1727.8	125	13.822
7	15	2017	3	0	0	0	0	0
8	reg_14	2017	3	4	1.38	51.43	14	3.674
9	REDE	2017	3	61	190.96	1918.76	186	10.316
10	15	2017	4	0	0	0	0	0

Showing 1 to 10 of 144 entries Previous 1 2 3 4 5 … 15 Next

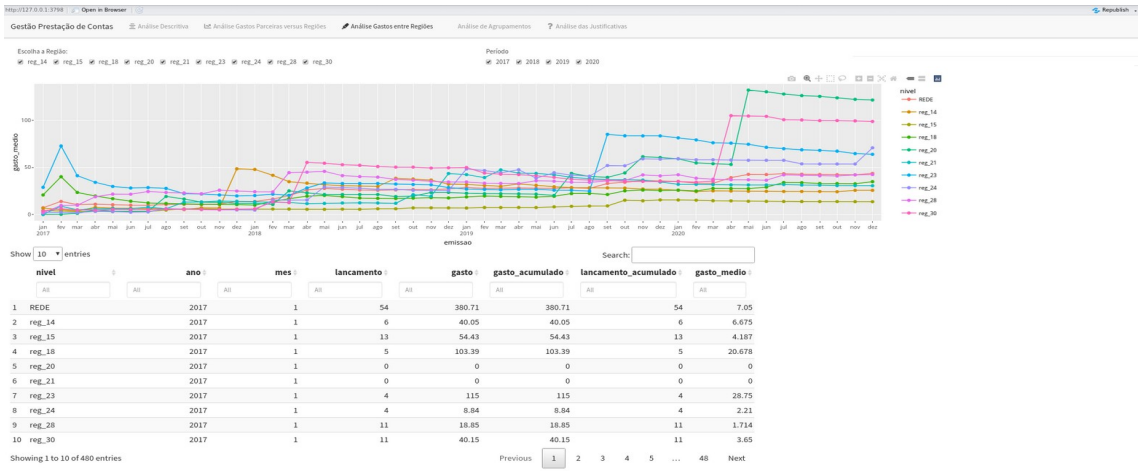
Fonte: Elaborado pelo autor

2.2.1.3 Aba 3 - Análise Gastos Médios Indevidos entre Regiões

A análise dos gastos médios indevidos entre as regionais tem o objetivo de identificar as regiões onde a ocorrência de gastos com juros e multas ganha destaque, seja com maior quantidade de lançamentos e valor, seja na menor quantidade. Esses fatos são importantes**,** pois o objetivo do monitoramento não está pautado no aspecto punitivo, e sim em detectar os fatores que desencadeiam tais fatos e estabelecer diretrizes que possam contribuir para a melhoria da gestão. Para tanto, foram criados dois instrumentos, como pode ser visto na Figura 16: a representação gráfica e uma tabela, ambas referentes aos gastos mensais por região.

Na parte superior desta aba existem dois grupos de caixas de seleção (*checkboxGroupInput*) que possibilitam selecionar referenciais que serão utilizados na comparação: a) Grupo 1: Escolha da(s) região(ões); b) Grupo 2: Escolha de um ou mais anos, no período de 2017 a 2020.

Figura 16-Análise Comparativa de Gasto indevidos por Regiões



Fonte: Elaborado pelo autor

A Figura 17 fornece a comparação por meio da visualização gráfica das regiões, porém de maneira panorâmica onde é possível a visualização dos gastos médios indevidos por região, que foram selecionadas nas opções existentes na parte superior da aba. Nesse tipo de visualização é possível verificar o período do ano no qual determinada(s) região(ões) teve(tiveram) o destaque no gasto médio, seja por ser acima dos demais**,** ou mesmo abaixo.

Figura 17-Análise Comparativa de Gasto indevidos por Regiões

Fonte: Elaborado pelo autor

A partir da percepção comparativa realizada na Figura 17**,** é possível realizar **a** pesquisa na Tabela da Figura 18, composta por filtros dos dados gerados especificamente para a comparação entre as regionais. Os dados foram distribuídos mensalmente por regiões, o que permite, neste caso**,** o isolamento do ano/mês da variação que se deseja realizar em um estudo mais pontual.

Figura 18-Análise Comparativa de Gastos Indevidos por Regiões

emissao

Show 10 entries Search:

	nivel	ano	mes	lancamento	gasto	gasto_acumulado	lancamento_acumulado	gasto_medio
	All	All	All	All	All	All	All	All
1	REDE	2017	1	54	380.71	380.71	54	7.05
2	reg_14	2017	1	6	40.05	40.05	6	6.675
3	reg_15	2017	1	13	54.43	54.43	13	4.187
4	reg_18	2017	1	5	103.39	103.39	5	20.678
5	reg_20	2017	1	0	0	0	0	0
6	reg_21	2017	1	0	0	0	0	0
7	reg_23	2017	1	4	115	115	4	28.75
8	reg_24	2017	1	4	8.84	8.84	4	2.21
9	reg_28	2017	1	11	18.85	18.85	11	1.714
10	reg_30	2017	1	11	40.15	40.15	11	3.65

Showing 1 to 10 of 480 entries Previous 1 2 3 4 5 … 48 Next

Fonte: Elaborado pelo autor

2.2.1.4 Aba 4: Análise de Agrupamento

A análise de agrupamentos, conforme Hair (2005, p.430), consiste no uso de técnicas multivariadas que une elementos com as mesmas características. Num mesmo agrupamento**,** os elementos são similares uns aos outros, tendo como referência as características contidas nas variáveis, escolhidas previamente pelo pesquisador**,** e uma considerável heterogeneidade entre os agrupamentos.

2.2.1.4.1 Análise Exploratória das Variáveis Referenciais

A base de dados df.RDS, utilizada para a realização da Análise de Agrupamento, possui 33 variáveis e um total de 3265 registros. Após o tratamento dos dados, onde foram selecionadas as parceiras que possuíam somente gastos referentes a juros e multas, foi observado a redução das variáveis para 11: *id_p* (identificador de cada parceira); *reg* (regional); *emissao* (data da emissão do comprovante de despesas); inc*_1* (pagamento de tarifas não permitidas (bancárias e outras)); inc*_2* (pagamento de impostos com multa/juros); inc*_3* (falta de orçamentos); inc*_4* (quantidade insuficiente de orçamentos); inc*_5* (pagamento a maior que o portal de preços e/ou orçamentos) e inc*_6* (comprovante de despesa referente a juros, multas ou pago a maior); *vr_prod* (valor do gasto realizado); inc*_7* (aquisição realizada sem o comprovante de despesa); inc*_8* (despesas indevidas à parceira). O quantitativo de observações também reduziu, passando de 3265 para 235 registros, resultado da somatória dos valores das variáveis quantitativas de cada parceira.

Logo em seguida, foi realizada a soma de cada uma das variáveis quantitativas por parceira, e constatou-se que: a variável *inc_4* correspondia a apenas dois lançamentos de uma única parceira; as variáveis *inc_3, inc_5, inc_7 e inc_8* não possuíam lançamentos; e as variáveis *inc_1, inc_2* e *inc_6* possuem basicamente os mesmos valores em cada observação. Dessa maneira, ao final desta análise**,** optou-se pela permanência 3 (três) variáveis: *id_p, inc_6 e vr_prod*.

Quando se observam os dados referentes às variáveis quantitativas *inc_6* e *vr_prod***,** é possível notar a diferença nítida entre as unidades de medidas, vide Figura 19, pois a variável *inc_6* corresponde às quantidades de lançamentos (números inteiros), enquanto *vr_prod* corresponde aos valores pagos (números decimais). Tais diferenças entre as unidades de medidas, conforme Mingoti(2005, p.200), podem causar distorções nos resultados finais de semelhança. Assim sendo, foi utilizado a padronização das variáveis, por meio da Equação 1, transportando-as para uma grandeza adimensional (sem unidades de medida), tornando-as comparáveis entre si:

Equação 1- Padronização

$$z = \frac{(x - \bar{X})}{dp(x)}$$

Os resultados da padronização das variáveis *inc_6 e vr_prod* correspondem respectivamente às *v1* e *vr* da Figura 19. Comparando os valores das variáveis *x1* e *vr*, é possível notar a redução acentuada nas discrepâncias, antes percebidas nas variáveis *inc_6* e *vr_prod*. Desta maneira, a base de dados a ser utilizada na Análise de Agrupamento denotada como *df-cluster.rds*, após a padronização das variáveis quantitativas, passa a ser formada por 6 variáveis: “id_p”; “reg”; “inc_6”; “vr_prod”; “x_6”; “vr”.

Figura 19-Comparativo entre unidades de medidas

id_p	reg	inc_6	vr_prod	x6	vr
296	reg_15	37	391.59	1.080061600	-0.118431342
297	reg_21	1	0.42	-0.602686307	-0.331127246
301	reg_24	2	16.22	-0.555943310	-0.322536108
302	reg_20	8	291.47	-0.275485325	-0.172870879
309	reg_28	19	25.85	0.238687646	-0.317299864
310	reg_20	4	12070.69	-0.462457315	6.231996176
314	reg_30	4	10.81	-0.462457315	-0.325477757
318	reg_24	6	51.66	-0.368971320	-0.303265860
320	reg_18	4	2.59	-0.462457315	-0.329947324
324	reg_28	28	3984.37	0.659374623	1.835117217
329	reg_24	2	8.54	-0.555943310	-0.326712054
335	reg_15	5	102.87	-0.415714317	-0.275420787
336	reg_30	32	66.06	0.846346613	-0.295435963

Fonte: Elaborado pelo autor

2.2.1.4.2 Técnicas para Construção de Agrupamentos

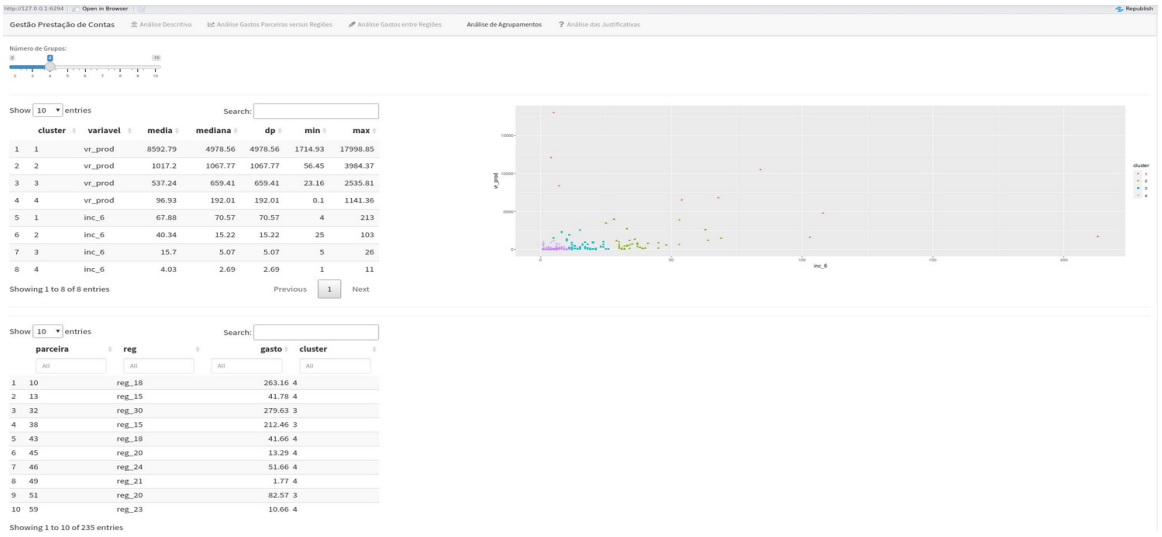
Para realização da análise de agrupamento**,** os dois métodos mais utilizados são: método hierárquico e método não hierárquico. Os métodos hierárquicos são aglomerativos e divisivos, capazes de identificar o provável número de grupos, bem como os objetos que pertencem a cada grupo. Nos métodos não hierárquicos, é necessário que o pesquisador identifique as prováveis quantidades de grupos (Mingot, 2005, p. 164).

Considerando que o aplicativo será para uso dos gestores públicos, sem formação estatística, optou-se pela utilização dos métodos não hierárquicos, uma vez que a obtenção do quantitativo provável de agrupamentos**,** pelo método hierárquico**,** requer conhecimento específico. Para obtenção da provável quantidade de agrupamentos, foi disponibilizado uma ferramenta interativa, Figura 20, pelo *sliders* (controles deslizantes) compostos por números inteiros de 2 a 7. Ao realizar a opção por quantitativo de grupos, tem-se a demonstração, por meio gráfico, de duas tabelas: (i) contendo a estatística descritiva de cada agrupamento, observando os dados das variáveis “vr_prod” e “inc_6”; (ii) identificação das parceiras que fazem parte de cada um dos agrupamentos.

O método K-means, conforme Mingoti(2005, p.192)**,** é provavelmente um dos métodos mais utilizado e conhecido, no que tange a problemas práticos. Tal método consiste em um algoritmo, que a partir do valor de K (quantidade de agrupamentos), julga-se pertinente. Ele inicia a escolha aleatória de K centros de grupos, para em seguida agrupar cada um dos objetos que estão na base de dados. Para que cada objeto possa pertencer a um determinado grupo, é necessário que a distância entre este objeto e o centro do grupo escolhido seja mínima.

Com o deslocamento dos objetos para os seus grupos, o algoritmo refaz o cálculo para localizar o novo centro de cada grupo, em função das mudanças ocorridas. A verificação da distância entre o objeto e o centro do grupo onde ele está no momento e a nova previsão de distância dos demais grupos serão realizadas continuamente até que os objetos não tenham mais necessidade de se deslocar para um local**,** onde a sua distância ao centro seja a escolha ótima, ou seja, a menor.

Figura 20 - Análise de Agrupamento



Fonte: Elaborado pelo autor

A utilização dos métodos não hierárquicos, quando se possui ferramentas interativas como as que compõem esta aba, permite que o usuário possa realizar a exploração dos dados como um detetive que detém o conhecimento prévio, neste caso sobre a gestão pública. MacQueen (1967, p.288) relata que o uso desses métodos possibilitam a visualização das informações qualitativa e quantitativa que podem ser adquiridas quando do tratamento de grandes quantidades de dados N-dimensionais, o que possibilita a obtenção de agrupamentos com similaridades razoáveis.

Nesta aba, o gestor ao realizar simulações para detectar o número mais adequado de grupos, observa simultaneamente a resposta gráfica e as medidas estatísticas dos agrupamentos que são apresentados nas duas tabelas geradas a partir do método k-means. Ao realizar a simulação, o gestor deverá estar atento ao seu objetivo: construir grupos com as parceiras que possuam maior similaridade, seja pelo maior quantitativo de lançamentos, seja pela observação dos valores de juros e multas. Para complementar essa análise, o gestor tem a possibilidade de detectar as possíveis razões dos gastos indevidos, utilizando os dados que estão na "*Aba Justificativa"* .

2.2.1.5 Análise das Justificativas

Considerando que todas as parceiras que pagaram juros e multa lançaram também as respectivas justificativas na forma de textos, foram realizadas pesquisas em artigos sobre o uso da mineração de textos, o que possibilitou a constatação do seu uso em diversas áreas**,** tais como: verificação dos aprendizados de estudantes, conforme DePaolo and Wilkinson (2014); facilitação da leitura dos dados estatísticos coletados durante a pesquisa de satisfação, no campo da educação em saúde, para identificar melhorias na adequação dos materiais utilizados no treinamento, conforme Bletzer(2015); estudo de entrevistas realizadas com atletas olímpicos na busca da singularidade, origem social e cultural de cada um, conforme Freitas, Neves & Gonçalves (2018); pesquisa em jornais científicos, conforme Waghmare(2021) .

Diante da constatação do uso da mineração de textos em diversas áreas, optou-se por realizar o tratamento das justificativas**,** utilizando a *mineração de texto*. Esse termo foi utilizado por Gaikwad, Chaugule e Patil (2014), por entenderem a existência de uma relação da *mineração de texto* com a *mineração de dados*, uma vez que por meio dessa técnica é possível ter como resultado a extração de informações relevantes de um texto[9].

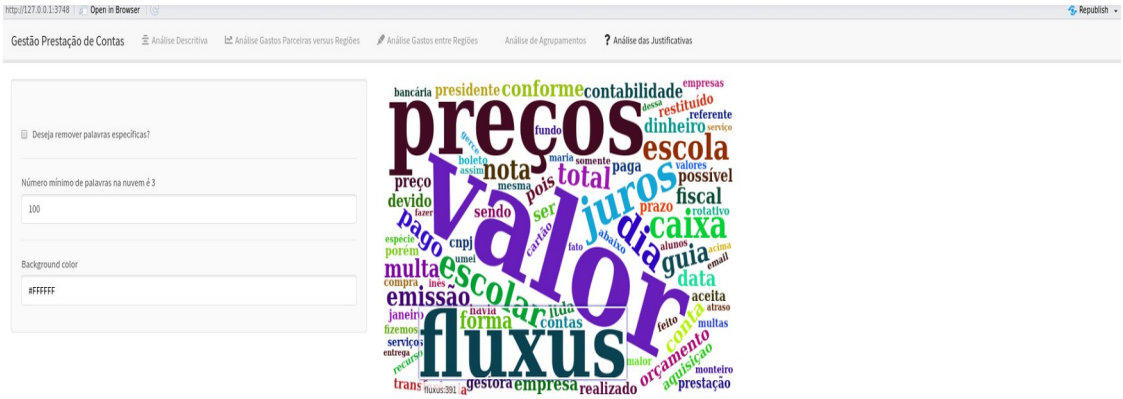
2.2.1.5.1 Mineração de Texto e Criação da Nuvem de Palavras

9 Cada vez maior de textos, tanto em órgãos governamentais, quanto não governamentais. No site do Tesouro Nacional pode ser observada a utilização de *storytelling* para promover uma aproximação do cidadão e os dados quantitativos. https://medium.com/tchiluanda/e-se-forrest-gump-fosse-servidor-p%C3%BAblico-82d2bec0beb0 visualizado em 16/06/2022.

Antes da criação da nuvem de palavras**,** foi feita a mineração do texto utilizando o pacote *tm* (Feinerer, 2008) do R, que preparou o texto realizando as seguintes ações: *tolower* que transforma todas as palavras para minúsculas; *removePunctuation* que remove pontuações; removeNumbers que remove números; stripWhitescape refere-se aos espaços em branco extras; *stopwords* que, conforme Feinerer,Hornik e Meyer(2008), remove palavras que nos seus idiomas não possuem valor informativo, sendo habitual a retirada antes da realização de análise mais aprofundada; *TermDocumentMatrix* que cria uma matriz do texto**,** contendo os termos como linhas**,** e documentos como colunas. Após a criação da matriz, foi necessário realizar tratamento final, transformando-a em um *data frame* contendo as palavras e as respectivas frequências.

Finalmente, para visualização da *nuvem de palavra***,** foi necessário o pacote *wordcloud2* (Lang and Chien, 2018) e a função com o mesmo nome. O aplicativo fornece a opção de interatividade com o usuário, dando opção para retirada de palavras**,** indicar a quantidade de palavras desejadas para comporem a nuvem**,** e alterar a cor do *background,* como pode ser observado na Figura 21.

Figura 21 - Análise das Justificativas



Fonte: Antoine Soetewey[10] (Modificado)

Observando a nuvem de palavras, na Figura 21, provenientes das justificativas declaradas pelas parceiras, tem-se a expectativa de que os usuários possam identificar os possíveis motivos que levaram ao pagamento de juros e multas. A partir daí, espera-se que seja possível aprimorar as ações junto às parceiras, no intuito de contribuir para melhoria da gestão financeira, consequentemente para a minimização desse tipo de pagamento.

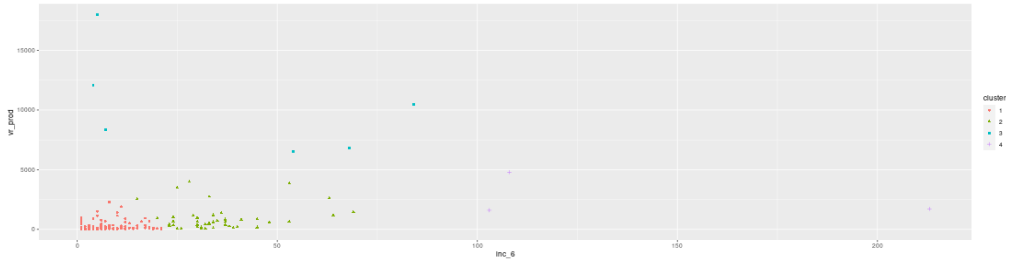
10 Script modificado do disponível no blog de Antoine Soetewey, disponível em: <https://statsandr.com/blog/draw-a-word-cloud-with-a-shiny-app/>, acesso em: 10 jun. 2022.

3 DISCUSSÃO DOS RESULTADOS

Considerando que o objetivo do aplicativo é fornecer subsídios para que o gestor público possa detectar os motivos que levaram às parceiras ao pagamento de despesas com juros e multas, os passos para detecção foram:

(i) na aba Análise de Agrupamento, quando se testa a realização de 4 cluster**,** é possível visualizar, Figura 22, uma distribuição coerente no que diz respeito ao quantitativo de lançamentos e valores gastos por grupo;

Figura 22-Análise Gráfica de Agrupamento



Fonte: Elaborado pelo autor

(ii) na tabela ao lado do gráfico, Figura 23, existem as informações dos intervalos dos valores e a quantidade de lançamentos das parceiras em cada cluster. Exemplo: no cluster 1, os valores variam de R$ 0,10 a R$ 2.301,01**,** e os lançamentos por parceira variam de 1 a 21.

Figura 23-Análise Descritiva do Agrupamento

Show 10 entries Search:

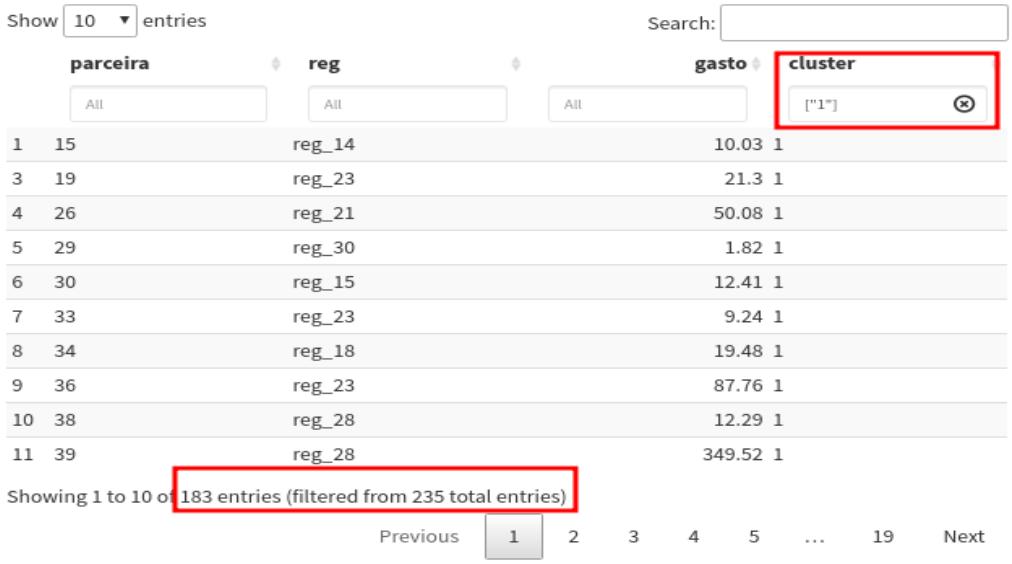
	cluster	variavel	media	mediana	dp	min	max
1	1	vr_prod	177.4	367.11	367.11	0.1	2301.01
2	2	vr_prod	939.56	1020.4	1020.4	49.52	3984.37
3	3	vr_prod	10372.96	4306.25	4306.25	6500.81	17998.85
4	4	vr_prod	2701.71	1809.09	1809.09	1600.58	4789.63
5	1	inc_6	6.04	4.98	4.98	1	21
6	2	inc_6	35.19	11.65	11.65	15	69
7	3	inc_6	37	35.98	35.98	4	84
8	4	inc_6	141.33	62.12	62.12	103	213

Showing 1 to 8 of 8 entries Previous 1 Next

Fonte: Elaborado pelo autor

(iii) a tabela abaixo do gráfico à esquerda, Figura 24, possibilita a identificação das parceiras por cluster. Em complementação ao exemplo citado no item (ii) acima, para identificar o quantitativo de parceiras, bem como os seus id_p's, basta clicar no filtro « cluster » da tabela e escolher 1. O resultado da quantidade de parceiras foi de 183 parceiras neste cluster.

Figura 24-Demonstrativo dos Elementos do Agrupamento



Fonte: Elaborado pelo autor

O cluster 1 possui aproximadamente 78% das parceiras que realizaram gastos com juros e multas, ele é um candidato com ótimas características para pesquisa dos motivos que levam aos gastos indevidos. Por outro lado, o cluster 3 é um candidato para pesquisa dos motivos que levaram as parceiras ao pagamento de juros e multas com os maiores valores.

A complementação dessa pesquisa estão nas demais abas, pois possibilitam determinar o mês nos quais ocorreram estes pagamentos e as justificativas utilizadas.

Desta maneira, é possível afirmar que o aplicativo cumpre o objetivo do trabalho**,** pois possibilitará ao gestor público a detecção das parceiras do período nos quais ocorreram e os motivos dos gastos indevidos.

4 CONSIDERAÇÕES FINAIS

O objetivo desse trabalho foi o desenvolvimento do aplicativo que contribuísse no Monitoramento da Gestão das Parcerias. A construção do aplicativo atendeu ao objetivo desse trabalho, pois já nessa primeira versão**,** é perceptível a contribuição para identificação das parceiras que possuem irregularidades similares, bem como das possíveis ações que as causaram. Esse fato possibilitará aos gestores públicos**,** responsáveis pelo monitoramento e avaliação das prestações de contas**,** o incremento nas ferramentas já existentes no monitoramento.

No que se refere à elaboração desse aplicativo, foi possível sedimentar muitos conceitos aprendidos durante a especialização, assim como vivenciar a importância no tratamento dos dados, uma das etapas da construção do aplicativo que demandou maior tempo e dedicação. Ao terminar esse trabalho, com o aplicativo na sua primeira versão, foi possível realizar simulações para constatação das parceiras que realizaram pagamentos das despesas com juros e multas, e dos possíveis agrupamentos das parceiras com características aproximadas. Como a expectativa é de que ele seja utilizado pelos gestores públicos, esses poderão demandar as melhorias não percebidas durante a construção.

Desta maneira, tem-se como meta futura a continuidade na melhoria do aplicativo, numa maior imersão que permita detectar novas áreas para utilização de ferramentas estatísticas. Outro aspecto a ser trabalhado é a adptação do aplicativo para ser utilizado por outros munícipios que tenham interesse.

5 REFERÊNCIAS

BLETZER, K V. Visualizing the qualitative: making sense of written comments from an evaluative satisfaction survey. **Journal of educational evaluation for health professions**, Chuncheon, v.12, n.12, p. 1-8, 2015. Disponível em: <https://doi.org/10.3352/jeehp.2015.12.12>. Acesso em: 18 jun. 2022.

BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil de 1988**. Brasília: Senado Federal, Centro Gráfico, 292 p, 1988.

CHANG, W; CHENG, J; ALLAIRE, JJ; SIEVERT, C; SCHLOERKE, B; XIE, Y; ALLEN, Jeff; MCPHESON, J; DIPERT, A; BORGES, B**. Shiny: web application framework for r. R package version 1.7.1** ,2021. Disponível em: <https://CRAN.R-project.org/package=shiny> Acesso em: 18 jun. 2022.

DEPAOLO, C.; WILKINSON, K. **Get Your Head into the Clouds: Using Word Clouds for Analyzing Qualitative Assessment Data**. TechTrends, Statesboro, 58, p.38-44, 2014. Disponível em: <10.1007/s11528-014-0750-9> Acesso em: 28 mai. 2022

FREITAS, R; NEVES, R; GONÇALVES, V. (2018). **Utilizando as técnicas de "nuvem de palavras" e clusterização aplicadas as entrevistas dos atletas olímpicos da cidade de São Carlos. Olimpianos**. Journal of Olympic Studies, Fullerton, 2, p.423-434, 2018. Disponível em: <10.30937/2526-6314.v2n2.id41. > Acesso em: 05 abr. 2022

FEINERER, I; HORNIK, K; MEYER, D. **Text Mining Infrastructure in R**. **Journal o Statistical Software**, Austria, 25, 5 Mar, 2008. Disponível em <https://www.jstatsoft.org/article/view/v025i05>. Acesso em: 16 jun 2022.

HAIR JR, Joseph F et al**. Análise multivariada de dados**. Tradução Adonai Schlup Sant'Anna.Porto Alegre:Bookman, 2009.688 p.

IHAKA, ROSS and GENTLEMAN, ROBERT. R : A Language for Data Analysis and Graphics. Journal of Computacional and Graphical Statistics, v. 5, n. 3, pages 29 9-314. Disponivel em : <https://www.jstor.org/stable/1390807>, acesso em 15 jun. 2022.

LANG, D; CHIEN, G **Wordcloud2: create word cloud by 'htmlwidget'. R package version 0.2.1**, 2018. Disponível em: < https://CRAN.R-project.org/package=wordcloud2> Acesso em: 25 jun. 2022.

MACQUEEN, J. B. S**ome methods for classification and analysis of multivariate observations**. In: BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTIC AND PROBABILITY,5., 1967, Berkley. *Proceedings...* Berkley: University of California Press, Berkley, 1967. p.281-297

MINGOTI, Sueli Aparecida. **Análise de Dados através de métodos de estatística multivariada**: uma abordagem aplicada. Belo Horizonte: Editora UFMG, 2005.

SIEVERT, C. **Interactive Web-Based Data Visualization with R, plotly, and shiny**, Chapman and Hall/CRC, 2020.

SIEVERT, c., C. Parmer, T. Hocking, S. Chamberlain, K. Ram, M. Corvellec, e P. Despouy, 2021 plotly: Create Interactive Web Graphics via 'plotly.js'. R package version 4.10.0

SILVA, L. A.; PERES, S. M.; BOSCARIOLI, C. **Introdução à Mineração de Dados: Com Aplicações em R** . 1ª Ed. Rio de Janeiro: Elsevier Editora Ltda. 2016. 277 p

TUKEY, J W. **Explority Data Analysis**. Readings: Addison-Wesley Publishing Company.Inc. 506p, 1977.

WAGHMARE, P. **Text mining and word cloud analysis of the articles pblished in selected library and information science journal over the past decade.** International Journal of Information Dissmination and Techology, v.11, n.), p.138-142, 2021.

WICKHAM, Hadley. Tidy Data.**Journal of Open Source Software**. Innsbruck, v. 59, n. 10. Aug. 2014. Disponível em <https://www.jstatsoft.org/article/view/v59i10>. Acesso em: 16 jun 2022.

WICKHAM, Hadley. **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2016

WICKHAM, Hadley, GROLEMUND, Garrett. **R para Data Science***:* Importe, Arrume, Transforme, Visualize e Modele Dados. Tradução de Samantha Batista. Rio de Janeiro: Alta Books, 2019. 528 p. Original ingês.

WICKHAM et al., (2019). **Welcome to the tidyverse**. Journal of Open Source Software, 4(43), 1686, Disponível em <https://doi.org/10.21105/joss.01686>. Acesso em 10 jun. 2022.

WICKHAM, Hadley. **Mastering Shiny:** *Build Interactive Apps, Reports, and Dashboards Powered by R.* Sebastopol, CA, United States of America. Ed. O'Reilly Media, Inc., 2021. 348 p.

XIE, yihui, CHENG, Joe and TAN, Xianying (2022). **DT: A Wrapper of the JavaScript Library 'DataTables'**. R package version 0.23. https://CRAN.R-project.org/package=DT

5.1 APÊNDICE A – *SCRIPT* EXEMPLO

aplicativo_teste.R

```
library(shiny)

ui<-fluidPage( # fluidpage é uma das Funções Layout utilizadas para customizar a interface com o usuário na ui

  # [1] numericInput() - Função para entrada numérica
  # [2] input = "n" - identifica a variável que recebe o valor da entrada
  # [3] value = 100 - estabelece o valor inicial de entrada
  numericInput(input="n",
  label="Tamanho da Amostra", value=100),

  # [4] plotOutput() - Função para Saída Gráfica
  plotOutput(outputId="hist")
)

# [5] Chama a função Server para renderizar as saídas e respostas para as entradas com R

server<-function(input, output, session) {
  # [6] output$hist
  # Envolve o código de resposta à entrada, usando a renderização, com as funções render*().
  # Após a renderização, a saída é enviada ao usuário.
  # No caso abaixo, a renderização é para saída de um gráfico
  output$hist<-renderPlot({

    hist(rnorm(input$n))
  })
}
[7] chama shinyApp() para combinar ui e server numa aplicação interactiva!
shinyApp(ui = ui, server = server)
```