

Big Data e o Uso da Linguagem R para Análises de Dados Meteorológicos da Cidade de Marília, SP

ALINE DE OLIVEIRA BERTONCINI

RICARDO FONSECA DA CRUZ

LARISSA PAVARINI DA LUZ

Resumo

Introdução: O Big Data é um sistema complexo de fontes de dados, tecnologias e metodologias que resultam em um volume extremo de dados de vários tipos, que podem ser rapidamente coletados, registrados e analisados. Para exemplificar sua aplicação, foi selecionada a meteorologia, que é a ciência que estuda os fenômenos que ocorrem na atmosfera, gerando grande volume de dados. Desta forma, objetivou-se usar a linguagem R para manipulação e análise dos dados coletados por estações meteorológicas. **Metodologia:** Assim, a linguagem R, que possui diversos recursos estatísticos e gráficos, foi usada para manipular e analisar os dados meteorológicos, que foram retirados do site do Instituto Nacional de Meteorologia, da estação automática de Marília, SP, no período de 2020 a 2022, resultando em planilha .xlsx. Dessa planilha foram selecionadas as colunas referentes a: data, hora (UTC) e chuva (mm). **Resultados:** Os dados foram importados para a interface RStudio, onde foram elaborados códigos para a manipulação e análise. Primeiramente, agrupou-se os dados utilizando a função `bind.rows()`, seguida do cálculo da média mensal de cada mês, finalizando com a geração de gráficos de linhas, representando as médias mensais referente às chuvas do período. **Conclusão:** A linguagem R facilitou a realização das análises dos dados meteorológicos para estabelecer as médias mensais da chuva da região de Marília, evidenciando que a linguagem R é eclética para o tratamento de grandes volumes de dados.

Palavras-chave: *big data; linguagem R; análise de dados; meteorologia.*

Big Data and the Use of R Language for Meteorological Data Analysis in the City of Marília, SP

Abstract

Big Data is a complex system of data sources, technologies, and methodologies that result in an extreme volume of data of various types, which can be quickly collected, recorded, and analyzed. To exemplify its application, meteorology was selected, which is the science that studies the phenomena that occur in the atmosphere, generating a large volume of data. Thus, the objective was to use the R language for manipulation and analysis of data collected by meteorological stations. Methodology: The R language, which has several statistical and graphic resources, was used to manipulate and analyze meteorological data, which were taken from the website of the National Institute of Meteorology, from the automatic station of Marília, SP, in the 2020 to 2022 period, resulting in a .xlsx spreadsheet. From this spreadsheet, the following columns were selected: date, time (UTC) and rain (mm). Results: The data were imported into the RStudio interface, where codes were elaborated for manipulation and analysis. First, the data were grouped using the `bind.rows()` function, followed by the calculation of the monthly average of each month, ending with the generation of line graphs, representing the monthly averages referring to the rainfall of the period. Conclusion: The R language facilitated the analysis of meteorological data to establish the monthly averages of rainfall in the region of Marília, showing that the R language is eclectic for the treatment of large volumes of data.

Keywords: *big data; R language; data analysis; meteorology.*

1 INTRODUÇÃO

O Big Data surgiu em decorrência da geração e, conseqüentemente, da necessidade da coleta de grande volume de dados com formatos variados. Como estes dados precisavam ser geridos, Hendler e Fox (2011 apud RIBEIRO, 2014) observaram que a gestão destes recursos possibilitaria a resolução de problemas que nem se sabia que existiam e que prescindiria de ferramentas, pois a capacidade do ser humano de analisar dados e informações com múltiplas facetas seria limitada. Logo, seriam necessários instrumentos que auxiliassem na execução destas tarefas.

O Big Data não representa apenas um grande conjunto de dados, mas um sistema complexo de fontes de dados, tecnologias e metodologias que resultam em um volume extremo de dados de vários tipos que podem ser rapidamente coletados, registrados e analisados.

De acordo com Taurion (2013), o Big Data foi descrito em 5 pilares fundamentais, os 5 V's:

- **Variedade:** engloba o uso dos diferentes tipos de dados, podendo ser estruturados ou não.
- **Volume:** aumento acelerado dos dados, escalando em volumes cada vez maiores criados pela sociedade.
- **Velocidade:** é necessário um grande poder de tempo de resposta com que os dados e informações são analisados, realizando as análises no menor tempo possível.
- **Valor:** é necessário que a análise dos dados compense o custo de sua coleta, armazenamento e processamento. Para a proximidade da exatidão, a qualidade dos dados analisados deve ser percebida pelos usuários.
- **Veracidade:** é necessário ter certeza que os dados fazem sentido e são autênticos, pois quanto maior a porcentagem de autenticidade, menor será a chance de erro nas análises de dados.

Portanto, o conceito de Big Data permite que as informações sejam tratadas de forma eficiente para que os processos se tornem mais eficazes e com menores perdas produtivas, buscando melhorar os processos de trabalho dos seus utilizadores, pois possibilita interpretações rápidas e valiosas sobre as tendências de mercado, comportamento de consumo e oportunidades potenciais.

Para exemplificar o uso cotidiano da big data, selecionou-se a área da meteorologia, que é responsável pela geração de grande volume de dados.

A meteorologia é definida como a ciência que estuda os fenômenos que ocorrem na atmosfera, e está relacionada ao estado físico, dinâmico e químico da atmosfera, as interações entre elas e a superfície terrestre subjacente (INMET, 2023).

O tempo é o resultado desses elementos e outros mais com a influência dos fatores astronômicos e fatores geográficos, que podem estar distribuídos em um número infinito de padrões no espaço e no tempo e em constante modificação.

Os instrumentos meteorológicos são equipamentos utilizados para adquirir dados meteorológicos (termômetro/temperatura do ar, pressão atmosférica/barômetro, higrômetro/umidade relativa do ar, entre outros). A reunião desses instrumentos em um mesmo local, é denominada estação meteorológica.

As estações meteorológicas captam dados da região em que está instalada, resultando em milhões de dados abertos, ou seja, informações digitais, disponíveis para download pela internet, que conforme Luz (2021) descreve, podem ser processadas por computadores para serem amplamente exploradas. Implicando que um dado aberto possa ser utilizado e

redistribuído, gerando novas informações através de mistura de outros conjuntos de dados e distribuição de resultados.

Desta forma, para analisar os dados coletados pelas estações meteorológicas, foi selecionada a Linguagem R, que é uma linguagem de programação que conta com diversos recursos estatísticos e gráficos, que vem se especializando na manipulação, análise e visualização dos dados, sendo considerada uma das melhores ferramentas para essa finalidade (HUSSAN, CAMPBELL et al., 2018 apud COSTA, 2022).

Portanto, objetivou-se o uso da linguagem R para analisar, manipular e tratar os dados coletados de repositórios nacionais, para obter médias mensais de chuva, buscando criar gráficos dos dados manipulados.

2 DESENVOLVIMENTO

2.1 Metodologia

Foi realizado levantamento bibliográfico em repositórios, bases científicas e governamentais para obter informações de dados meteorológicos e uso de metodologia aplicada para tratamento dos dados obtidos.

A previsão do tempo no Instituto Nacional de Meteorologia (INMET) se baseia em dados atmosféricos (precipitação, ventos, umidade relativa do ar, pressão atmosférica, entre outros) observados em estações meteorológicas, em modelos numéricos de previsão, em imagens de satélite e em radares meteorológicos.

As estações automáticas captam dados a cada hora. As estações convencionais, três vezes ao dia. Esses dados são recebidos nos distritos, que os processa e os envia para a Sede em Brasília-DF, que, por sua vez, armazena os dados em um banco de dados oficial, os processa, os integra e os envia por satélite para todo o mundo.

Os dados meteorológicos foram retirados do site do INMET, no endereço disponível em: <https://tempo.inmet.gov.br/TabelaEstacoes/A001>.

Para o recorte temporal, no campo Tabela de Dados das Estações, foram coletados os dados da estação automática de Marília (A763), com seleção do primeiro ao último dia de cada mês, dos anos de 2020, 2021 e 2022, resultando em 12 planilhas por ano.

O site do INMET gera planilhas CSV com as informações pesquisadas. Cada planilha retorna colunas com data, hora, temperatura, umidade, ponto de orvalho, pressão, vento, radiação e chuva (Figura 1), com medições referentes a cada hora de todos os dias do mês.

Figura 1 – Modelo de planilha gerada pelo site INMET, referente ao mês de janeiro de 2021

Instituto Nacional de Meteorologia																			
Data de Referência: 01/01/2021 - 31/01/2021 Estação: MARILIA A763 *Dados disponíveis em tempo real (sem controle de qualidade). Baixar CSV																			
Data	Hora	Temperatura (°C)			Umidade (%)			Pto. Orvalho (°C)			Pressão (hPa)			Vento			Radiação	Chuva	
		UTC	Inst.	Máx.	Min.	Inst.	Máx.	Min.	Inst.	Máx.	Min.	Inst.	Máx.	Min.	Vel. (m/s)	Dir. (°)	Raj. (m/s)	Kj/m²	mm
01/01/2021	0000		19,3	21,3	19,2	98,0	98,0	96,0	19,0	20,9	18,9	939,9	940,2	939,1				0,30	25,0
01/01/2021	0100		19,0	19,3	19,0	99,0	99,0	98,0	18,8	19,0	18,7	939,8	939,9	939,7					1,0
01/01/2021	0200		19,1	19,2	19,0	99,0	99,0	99,0	18,9	18,9	18,8	939,2	939,8	939,2					0,0
01/01/2021	0300		19,8	19,8	19,0	99,0	99,0	99,0	19,6	19,6	18,8	938,4	939,3	938,4					0,0
01/01/2021	0400		19,8	20,0	19,7	98,0	99,0	98,0	19,5	19,7	19,4	938,3	938,4	938,1					0,0
01/01/2021	0500		19,8	19,9	19,7	97,0	98,0	97,0	19,4	19,5	19,3	938,4	938,4	938,3					0,0

Fonte: Dados da pesquisa (2023).

Da planilha original, foram selecionadas três colunas: data, hora (UTC) e chuva (mm/hora), que foram salvas em formato .xlsx para importação na ferramenta utilizada.

Assim, devido ao grande volume de dados levantados, para análise e manipulação, escolheu-se a linguagem R, usando as bibliotecas ggplot2, readxl, dplyr e o ambiente de desenvolvimento integrado RStudio, discriminados a seguir.

2.2 Linguagem R

Criada em 1995, surgiu derivada da linguagem S, da Bell Labs, com sintaxes muito parecidas, o que possibilitou a difusão da linguagem para fins estatísticos, devido a facilidade em manusear o software e gerar resultados satisfatórios (HUSSAN, CAMPBELL et al., 2018 apud COSTA, 2022).

A linguagem R é um software que, diferentemente de outras linguagens, é interpretada, ou seja, necessita de um interpretador para que as informações, linhas de código, sejam lidas e mostradas para o usuário.

Além disso, uma das suas principais características é ser uma linguagem orientada a objetos. “*Objetos*” é o método que o R possui para guardar os valores, funções e resultados que são produzidos. Como foi posto por Adler (2010 apud FARIA, 2022), todo código do R busca utilizar, manipular ou modificar, de alguma forma, um objeto do R. Logo, quando se trabalha com dados no R, aplica-se constantemente operações e transformações sobre os objetos onde os dados estão guardados, de uma forma interativa e dinâmica.

Para seu melhor desempenho, foram utilizadas bibliotecas específicas, como: ggplot2, dplyr, readxl.

2.2.1 Ggplot2

O ggplot2 é um pacote da linguagem R voltado para a visualização de dados, ou seja, para a construção de gráficos (DIDÁTICA TECH, 2022).

O ggplot2 é um pacote desenvolvido por Hadley Wickham. A ideia do pacote, ainda que com algumas modificações, advém de uma obra chamada The Grammar of Graphics, que é uma maneira de descrever um gráfico a partir dos seus componentes. Dessa forma, teoricamente, ficaria mais fácil entender a construção de gráficos mais complexos (GUERRA et al., 2021).

Portanto, é uma biblioteca de alta flexibilidade para a elaboração de diferentes tipos de gráficos, para exibir a análise de dados e realizar estudos pontuais sobre determinado dado específico.

2.2.2 Dplyr

O dplyr é o pacote mais útil para realizar transformação de dados, aliando simplicidade e eficiência de uma forma elegante.

As principais vantagens de se usar o dplyr são: maior facilidade de manipular os dados; escrita mais intuitiva e facilidade de leitura do código.

O pacote dplyr utiliza C e C++ por trás da maioria das funções, o que geralmente torna o código mais rápido. É possível trabalhar com diferentes fontes de dados, como bases relacionais SQL (Structured Query Language) e data.table (DAMIANI et al., 2022).

2.2.3 Readxl

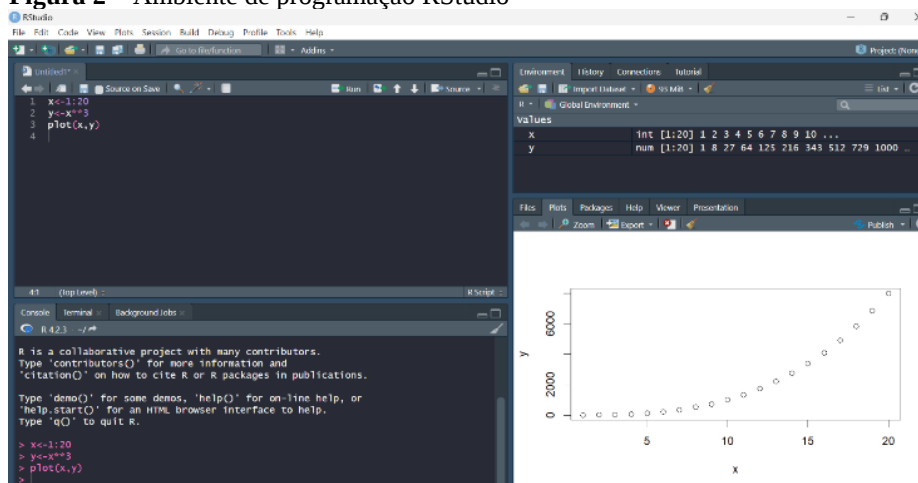
A biblioteca readxl é utilizada para ler planilhas do Excel (arquivos .xlsx ou .xls), basta utilizar a função read_excel() do pacote readxl. É necessário instalá-lo antes de usar. O pacote disponibiliza 5 exemplos de arquivos com formato .xls e .xlsx (DAMIANI et al., 2022).

2.2.4 RStudio

O RStudio é um ambiente de desenvolvimento integrado (Integrated Development Enviroment - IDE) para a programação da linguagem R. Com este ambiente de programação pode-se combinar vários componentes como: console, gráficos, editor de códigos, entre outros.

Para a programação em R foi utilizado a interface, conforme figura 1, para análise e manipulação de dados, assim como, visualização dos gráficos.

Figura 2 – Ambiente de programação RStudio



Fonte: Dados da pesquisa (2023).

2.2.6 RMarkdown

O R Markdown é um sistema de uso fácil, que permite combinar a análise estatística com a produção de um documento, ou seja, processar um script do R e gerar um documento com gráficos e texto estilizado (BAUMER et al., 2014 apud KONRATH et al., 2018).

Sendo assim, possibilita a criação de um relatório técnico, juntamente com a explicação do código, para facilitar a reprodução dos resultados alcançados, sendo uma ferramenta que realiza a transformação dos dados em um documento.

3 RESULTADOS

Para o tratamento das informações, os dados foram analisados mensalmente durante o período de um ano, sendo selecionados os anos de 2020 a 2022.

Para fins de demonstração, o ano de 2020 foi o escolhido, mas se realizou os mesmos procedimentos aos anos de 2021 e 2022.

O primeiro passo foi importar os arquivos .xlsx para a IDE Rstudio, conforme código disponível no quadro 1.

Quadro 1 – Importação de dados do ano de 2020 para o Rstudio

```
DJaneiro2020 <-
read_excel("/Users/Desktop/TCC/DadosR/2020/Janeiro2020.xlsx")
DFevereiro2020 <-
read_excel("/Users/Desktop/TCC/DadosR/2020/Fevereiro2020.xlsx")
DMarco2020 <- read_excel("/Users/Desktop/TCC/DadosR/2020/Marco2020.xlsx")
DAbril2020 <- read_excel("/Users/Desktop/TCC/DadosR/2020/Abril2020.xlsx")
DMAio2020 <- read_excel("/Users/Desktop/TCC/DadosR/2020/Maio2020.xlsx")
DJunho2020 <- read_excel("/Users/Desktop/TCC/DadosR/2020/Junho2020.xlsx")
DJulho2020 <- read_excel("/Users/Desktop/TCC/DadosR/2020/Julho2020.xlsx")
DAgosto2020 <- read_excel("/Users/Desktop/TCC/DadosR/2020/Agosto2020.xlsx")
DSetembro2020 <-
read_excel('/Users/Desktop/TCC/DadosR/2020/Setembro2020.xlsx')
DOutubro2020 <-
read_excel("/Users/Desktop/TCC/DadosR/2020/Outubro2020.xlsx")
DNovembro2020 <-
read_excel("/Users/Desktop/TCC/DadosR/2020/Novembro2020.xlsx")

DDezembro2020 <-
read_excel("/Users/Desktop/TCC/DadosR/2020/Dezembro2020.xlsx")
```

Fonte: Elaborado pelos autores (2023)

Para manipulação dos dados, a função *bind.rows()*, da biblioteca *dplyr*, foi usada para agrupar os dados em uma única base de dados, chamada de Dados2020, para uma melhor tratativa. E a função *mutate()* foi usada para alterar o quadro de dados adicionando novas colunas ou substituindo as existentes.

Abaixo, o código para agrupar os dados para calcular a média mensal, por hora, da chuva (Quadro 2).

Quadro 2 – Código agrupamento dos dados para calcular a média de chuva, por mês, usando o bind.rows()

```
Dados2020 <- bind_rows(DJaneiro2020 %>% mutate(Mes = "Janeiro"),
DFevereiro2020 %>% mutate(Mes = "Fevereiro"),
DMarco2020 %>% mutate(Mes = "Marco"),
DAbril2020 %>% mutate(Mes = "Abril"),
DMAio2020 %>% mutate(Mes = "Maio"),
DJunho2020 %>% mutate(Mes = "Junho"),
DJulho2020 %>% mutate(Mes = "Julho"),
DAgosto2020 %>% mutate(Mes = "Agosto"),
DSetembro2020 %>% mutate(Mes = "Setembro"),
DOutubro2020 %>% mutate(Mes = "Outubro"),
DNovembro2020 %>% mutate(Mes = "Novembro"),
DDezembro2020 %>% mutate(Mes = "Dezembro") )
```

```
ordem_meses <- c("Janeiro", "Fevereiro", "Março", "Abril", "Maio",
  "Junho", "Julho", "Agosto", "Setembro", "Outubro", "Novembro",
  "Dezembro")
```

Fonte: Elaborado pelos autores (2023)

Após a realização da junção dos dados, foi necessário calcular a média de chuva mensal, milímetros/hora. Para isso, utilizou-se o *group_by* para unir as linhas pertencentes ao mesmo mês e a função *summarise* para realizar o cálculo de média da chuva, como representado no quadro 3.

Quadro 3 – Código para cálculo de média mensal, milímetros/hora, com a função *summarise()*

```
#cálculo de média chuva mensal
MediaChuvaMensal2020 <- Dados2020 %>%
  group_by(Mes) %>%
  summarise(MediaChuvaMensal = mean(`Chuva (mm)`, na.rm = TRUE), .groups =
  "drop") %>%
  mutate(Mes = factor(Mes, levels = ordem_meses))
```

Fonte: Elaborado pelos autores (2023)

Para comparação das médias obtidas dos anos de 2020, 2021 e 2022, a Tabela 1 foi criada.

Como os dados são referentes a cada hora, optou-se por não fazer o arredondamento de casas decimais, para que ficasse visível as mínimas oscilações existentes.

Pode-se observar que o mês mais chuvoso em 2020 foi fevereiro e o mais seco foi julho. Em 2021 e 2022, o mês mais chuvoso foi janeiro e o mais seco foi julho.

Ainda é possível verificar que o ano mais chuvoso foi 2020, seguido de 2022 e 2021 foi o ano com menos chuvas.

Tabela 1 – Valores médios mensais dos milímetros/hora de chuva de 2020, 2021 e 2022

MESES	ANO		
	2020	2021	2022
Janeiro	0,175	0,303	0,360
Fevereiro	0,644	0,205	0,195
Março	0,0613	0,191	0,232
Abril	0,0767	0,0533	0,0733
Maio	0,0672	0,0336	0,0804
Junho	0,138	0,0844	0,0969
Julho	0,00941	0,00430	0,000269
Agosto	0,164	0,0124	0,106
Setembro	0,0106	0,0767	0,158
Outubro	0,0817	0,207	0,178
Novembro	0,0992	0,206	0,0586
Dezembro	0,431	0,158	0,250

Fonte: Elaborada pelos autores (2023)

Para mais demonstrações, usando a tabela resultada pelo código da média, com o auxílio da biblioteca *ggplot*, foi possível configurar os dados e o tipo de gráfico.

O gráfico escolhido para a representação das médias das chuvas foi o de linhas (Quadro 4).

Quadro 4 – Código para criação de gráficos com as informações das médias, milímetros/ hora, de chuva

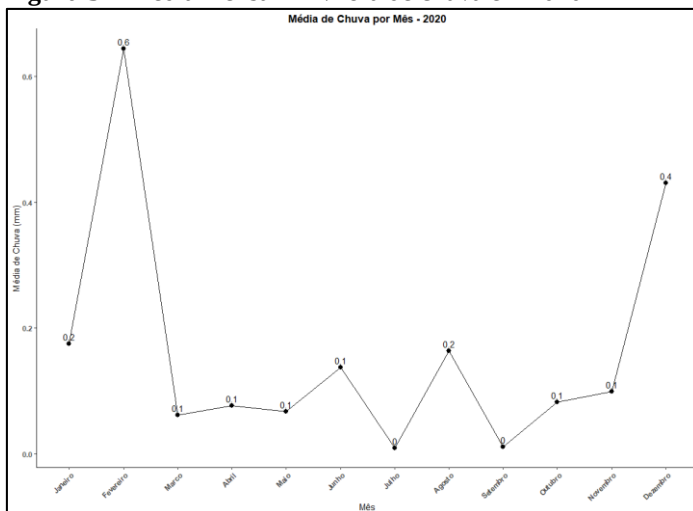
```
#gráfico da chuva mensal do ano 2020
ggplot(MediaChuvaMensal2020, aes(x = Mes, y = MediaChuvaMensal, group =
1)) +
```

```
geom_line(color = "black") +  
geom_point(color = "black", size = 2.5) +  
geom_text(aes(label = round(MediaChuvaMensal, 1)), vjust = -0.5, color  
= "black", size = 4) +  
labs(title = "Média de Chuva por Mês - 2020",  
x = "Mês",  
y = "Média de Chuva (mm)") +  
theme_classic() +  
theme(  
plot.title = element_text(hjust = 0.5, face = "bold"),  
axis.text.x = element_text(angle = 45, hjust = 1, face = "bold"),  
axis.text.y = element_text(face = "bold")  
)
```

Fonte: Elaborado pelos autores (2023).

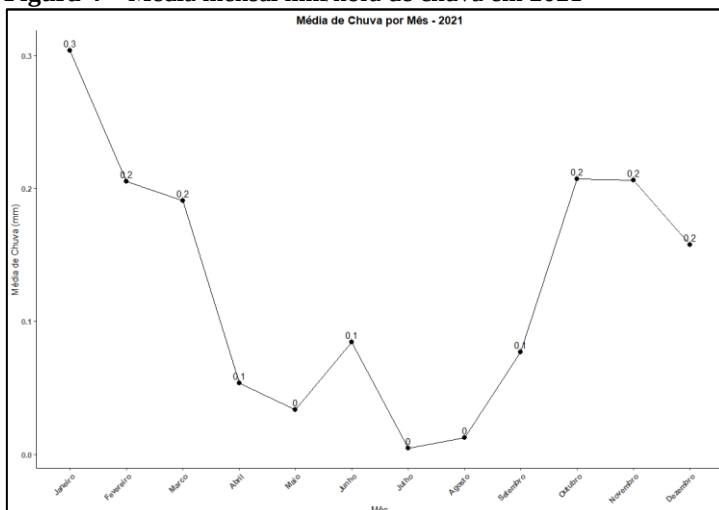
Os gráficos de linhas (Figuras 3, 4 e 5) representam os anos de 2020, 2021 e 2022, comparando os resultados obtidos mensalmente de chuvas.

Figura 3 – Média mensal mm/hora de chuva em 2020



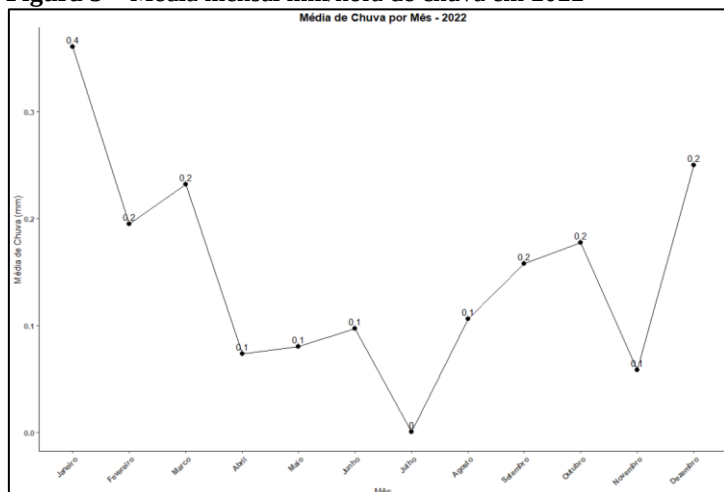
Fonte: Elaborada pelos autores (2023).

Figura 4 – Média mensal mm/hora de chuva em 2021



Fonte: Elaborada pelos autores (2023).

Figura 5 – Média mensal mm/hora de chuva em 2022



Fonte: Elaborada pelos autores (2023).

O RStudio possibilita inúmeras possibilidades, incluindo gráficas, para representação dos dados. Assim, no quadro 5, foi apresentado o código para criação de um gráfico de colunas, que compara os anos de 2020, 2021 e 2022 simultaneamente.

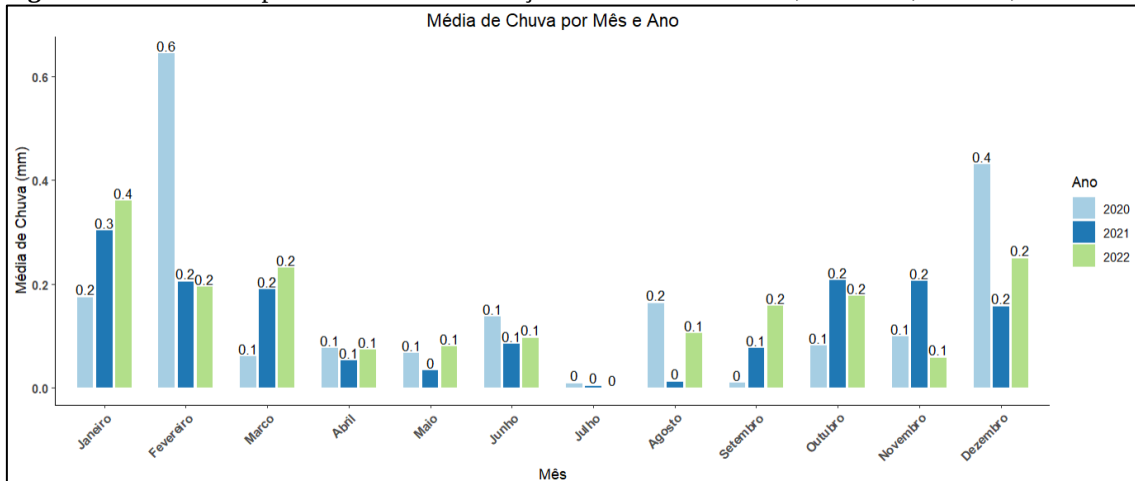
Quadro 5 – Código para criação de gráfico comparativo com as informações das médias de chuva, milímetros/hora, de 2020, 2021 e 2022

```
# Criar o gráfico comparativo
ggplot(MediaChuvaMensal, aes(x = Mes, y = MediaChuvaMensal, fill = as.factor(Ano))) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.7), width = 0.6) +
  geom_text(aes(label = round(MediaChuvaMensal, 1)), position = position_dodge(width = 0.7),
    vjust = -0.25, size = 3.5) +
  labs(title = "Média de Chuva por Mês e Ano",
    x = "Mês",
    y = "Média de Chuva (mm)",
    fill = "Ano") +
  scale_fill_brewer(palette = "Paired") +
  theme_classic() +
  theme(
    plot.title = element_text(hjust = 0.5),
    axis.text.x = element_text(angle = 45, hjust = 1, face = "bold"),
    axis.text.y = element_text(face = "bold")
  )
```

Fonte: Elaborado pelos autores (2023).

A figura 6 apresenta o gráfico de colunas comparativo do anos analisados, possibilitando uma apresentação visual mais nítida das informações obtidas, como meses mais chuvosos ou secos.

Figura 6 – Gráfico comparativo com as informações das médias de chuva, mm/ hora, de 2020, 2021 e 2022



Fonte: Elaborada pelos autores (2023).

Assim, foi possível verificar que fevereiro e dezembro de 2020 foram os meses mais chuvosos. Em julho não houve chuvas.

Dezembro, janeiro e fevereiro são meses que chovem mais, pois são meses de verão.

Os meses de inverno (junho, julho, agosto) tem pouquíssimas chuvas.

Do período analisado, o ano de 2020 foi o ano mais chuvoso e 2021 o mais seco.

Essas foram algumas constatações, mas seria possível realizar outras variadas interações com os dados obtidos e o R, analisando inúmeras possibilidades, gerando e/ou confirmando informações importantes para tomada de decisão da população envolvida, tanto urbana quanto rural.

4 CONSIDERAÇÕES FINAIS

O fluxo de dados é volumoso, mas transformá-los em informação útil demanda trabalho e, às vezes, o uso de ferramentas específicas.

Muitos dados estão disponíveis para download na internet, inclusive abertos, mas podem estar dispersos em vários ambientes, que precisam ser compilados e analisados.

Assim, não basta somente coletá-los, é necessário limpar, analisar e manipular para gerar valor.

A linguagem R é uma ferramenta útil para manipulação e análise, pois permite trabalhar com codificação para manipular dados, gerar visualização gráfica e relatório técnico, através do RMarkdown, que pode ser disponibilizado via web.

Desta forma, a linguagem R facilitou a realização das análises dos dados meteorológicos, disponibilizados pelo INMET, para estabelecer as médias mensais, por milímetros/hora, de chuva da região de Marília.

Portanto, comprovou-se que a linguagem R é eclética para o tratamento de grandes volumes de dados.

E, por ser um software de código aberto, há várias informações disponíveis gratuitamente, sendo mantido atualizado através da sua própria comunidade de usuários, contribuintes e pesquisadores.

Concluindo, verificou-se que há inúmeras possibilidades de uso com a linguagem R, sendo uma ferramenta amigável para análise, manipulação e visualização de dados. Acessível para todos os níveis de usuários, de iniciantes a especialistas.

REFERÊNCIAS

- COSTA, Diogo Ricardo da. Linguagem R: conceitos e aplicações. Ponta Grossa: Aya, 2022. Disponível em: <https://ayaeditora.com.br/wp-content/uploads/2022/02/L110.pdf>. Acesso em: 18 mar 2023.
- DAMIANI, Athos; MILS, Beatris, LENTE, Caio; FALBEL, Daniel; CORREA, Fernando; TRECENTINI, Julio; LUDUVICE, Nicole; LACERDA, Tereza; AMORIM, William. Ciência de dados em R. Curso-R, 2022. Disponível em: <https://livro.curso-r.com/index.html>. Acesso em: 10 maio 2023.
- DIDÁTICA TECH. Inteligência Artificial e Data Science. O pacote ggplot2, função ggplot: linguagem R. 2022. Disponível em: <https://didatica.tech/o-pacote-ggplot2-linguagem-r/>. Acesso em: 17 abr 2023.
- FARIA, Pedro Duarte. Introdução à Linguagem R: seus fundamentos e sua prática. 4. ed. Belo Horizonte: [s.n.], 2022. ISBN 978-65-00-57872-0. Disponível em: https://pedro-faria.netlify.app/pt/publication/book/introducao_linguagem_r/. Acesso em: 10 mar. 2023.
- GUERRA, Saulo; OLIVEIRA, Paulo Felipe; MACDONNELL, Robert; GONZAGA, Sillas. Visualizações de dados (ggplot2). In: GUERRA, Saulo; OLIVEIRA, Paulo Felipe; MACDONNELL, Robert; GONZAGA, Sillas. Ciência de dados com R: introdução. IBPAD: 2021. Disponível em: <http://sillasgonzaga.com/material/cdr/ggplot2.html>. Acesso em: 03 abr. 2023.
- INMET (Instituto Nacional de Meteorologia). Sobre meteorologia. Disponível em: <https://portal.inmet.gov.br/sobre-meteorologia>. Acesso em: 03 abr. 2023.
- KONRATH, Andréa Cristina; SILVA, Silvio Aparecido da; HENNING, Elisa; SANTOS, Luciane Mulazani dos; MIRANDA, Rodrigo Gabriel de; SAMOHYL; Robert Wayne. Desenvolvimento de Aplicativos Web Com R e Shiny: inovações no ensino de estatística. Abakos, Belo Horizonte, v. 6, n. 2, p. 55-71, maio 2018.
- LUZ, Larissa Pavarini. Framework para publicação de dados com ênfase em enriquecimento e mapeamento semântico. 2021. 278 f. Tese (Doutorado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Universidade Estadual Paulista, Marília, 2021. Disponível em: https://repositorio.unesp.br/bitstream/handle/11449/204711/luz_lp_dr_mar.pdf?sequence=3&isAllowed=y. Acesso em: 20 fev. 2023.
- RIBEIRO, Claudio Jose Silva. Big data: os novos desafios para o profissional da informação. Informação & Tecnologia (ITEC): João Pessoa/Marília, 1(1): 96-105, jan./jun., 2014. Disponível em: https://www.brapci.inf.br/_repositorio/2015/12/pdf_616d318e35_0000015108.pdf. Acesso: 10 abr. 2023.

TAURION, Cesar. Big Data. Rio de Janeiro: Brasport, 2013. Disponível em: <https://goo.gl/MVNFQ3>. Acesso em 3 de novembro de 2023.