



# **A IMPORTÂNCIA DA VISUALIZAÇÃO DE DADOS PARA O PROCESSO DE TOMADA DE DECISÃO: UM ESTUDO SOBRE O DESEMPENHO DE ATLETAS DE PARATAEKWONDO**

## ***THE IMPORTANCE OF DATA VISUALIZATION FOR THE DECISION-MAKING PROCESS: A STUDY ON THE PERFORMANCE OF PARATAEKWONDO ATHLETES***

Submetido em: 19/01/2022

Aceito em: 04/03/2022

André Marinho Moreira<sup>1</sup>  
Magali Rezende Gouvêa Meireles<sup>2</sup>

*Esse artigo foi vencedor do III Prêmio de Trabalhos Científicos de Graduação da PUC Minas São Gabriel, na categoria Tecnologia.*

### **RESUMO**

A visualização de dados é de significativa importância para o processo de tomada de decisão, tendo em vista a quantidade considerável de dados que vem surgindo com a evolução tecnológica e a consequente necessidade de sua análise. Este artigo tem como objetivo apresentar relatórios de visualização de dados para uma gestão mais eficaz do desempenho dos atletas de um Centro Esportivo voltado para a modalidade Parataekwondo. A elaboração dos relatórios abrange desde a aderência e limpeza dos dados, seleção e composição dos visuais, considerando aspectos da percepção humana, até sua apresentação. A ideia é que a informação transmitida guie os profissionais para uma tomada de decisão mais assertiva. Na expectativa de validar a metodologia proposta, um profissional da área foi entrevistado para observar os impactos significativos da visualização e da análise dos dados apresentados no processo de acompanhamento do rendimento dos atletas e na definição de estratégias em competições.

**Palavras-chave:** Análise de Dados Esportivos. Visualização de Dados. Percepção Humana. Processos de Tomada de Decisão. Parataekwondo.

### **ABSTRACT**

*The data visualization is significantly important for the decision making process, in view of the considerable amount of data, which has been emerging with the technological evolution and its consequent need for analysis. This article aims to present reports concerning data visualization for a more effective management of athletes' performances in a Sports Center focused on the Parataekwondo modality. The elaboration of the reports covers from the adherence and cleaning of the data, selection and composition of the visuals, considering aspects of human perception, to its presentation. The idea is that the transmitted information will guide the professionals to more assertive decision making. In order to validate the proposed methodology, a professional in the area was interviewed to observe the significant impacts of visualization and analysis of the data presented in the process of monitoring athlete's performances and in the definition of strategies in competitions.*

<sup>1</sup>Graduado em Sistemas de Informação pela PUC Minas, Brasil– andreamarinho@outlook.com

<sup>2</sup> Pesquisadora colaboradora da Queensland University of Technology, Austrália, professora adjunta IV do Instituto de Ciências Exatas e Informática da PUC Minas e doutora em Ciência da Informação pela UFMG, Brasil– magali@pucminas.br

**Keywords:** *Sports Data Analysis. Data Visualization. Human Perception. Decision Making Processes. Parataekwondo.*

## INTRODUÇÃO

A visualização de dados é uma expressão contemporânea que descreve a representação visual dos dados. Mas, desde a pré-história, os povos demonstravam suas manifestações artísticas por meio de desenhos, símbolos e sinais nas paredes das cavernas, que serviam como guias para sobreviver naquele tempo. Mapas das constelações e dos movimentos do sol foram criados pelas antigas civilizações gregas e egípcias com o objetivo de entender o uso do espaço (SILVA, 2019).

Nos dias atuais, com a evolução da tecnologia da informação, são geradas quantidades consideráveis de dados. Todavia, com o aporte dos sistemas de informação, é possível lidar com esses dados com mais facilidade, transformá-los em informação e conhecimento. Diante disso, diversas organizações, indústrias, instituições do ramo financeiro, da saúde, esportivas e entre outras utilizam diferentes tecnologias para gerar visualizações e facilitar a análise dos dados, tornando-se mais assertivas no processo de tomada de decisão.

Um dos ramos que tem se beneficiado dos dados para a tomada de decisão é o esportivo. Em 2004, o jornalista e escritor Michael Lewis publicou "Moneyball: O Homem que Mudou o Jogo", livro em que apresentou um novo método, que preza pela análise de dados para gerar conhecimento e auxiliar na avaliação, captação de talentos e gerenciamento de atletas no âmbito do beisebol (LEWIS, 2004). A obra de Lewis teve um grande peso na indústria esportiva e o método foi aplicado em diferentes campos, como futebol, voleibol, entre outros.

Este trabalho possui como objetivo geral apresentar visualizações de dados relacionadas ao desempenho dos atletas de um centro esportivo paralímpico com foco em Parataekwondo, para gerar conhecimento e auxiliar a tomada de decisão no gerenciamento dos atletas em treinos e competições. Como objetivos específicos, foram apresentados dados por meio de relatórios elaborados com as tecnologias *Power BI* e *Datawrapper*, que serviram como apoio à geração de visualizações dos dados. Estes dados estão relacionados ao desempenho físico dos atletas do centro esportivo durante os treinos e ao estudo dos adversários em competições. Estes relatórios foram elaborados a partir das bases de dados em Excel fornecidas e passaram por um processo de limpeza e estruturação. As visualizações foram geradas e analisadas, selecionando-se as visualizações gráficas adequadas para a avaliação dos atletas do centro esportivo e para o estudo de seus adversários. Por fim, os relatórios foram apresentados ao profissional envolvido para o apoio à tomada de decisão e foi realizada uma entrevista para avaliar a contribuição do processo. Na seção seguinte, é apresentado o referencial teórico, que contém os conceitos de visualização de dados, tipos de visualização, aspectos da percepção humana para visualização de dados e trabalhos relacionados ao entendimento do uso da visualização de dados para tomada de decisão. A Seção 3 se refere a metodologia aplicada a este trabalho. A Seção 4 apresenta a

análise dos resultados obtidos e a Seção 5 diz respeito às considerações finais.

## REFERENCIAL TEÓRICO

Nesta seção, são apresentados os conceitos relacionados a este trabalho. O conceito de visualização de dados é retratado na Subseção 2.1. A Subseção 2.2 aborda a importância da visualização de dados no processo de análise de dados. A Subseção 2.3 aborda as formas de visualização de dados. A Subseção 2.4 apresenta os aspectos da percepção humana na visualização de dados. A Subseção 2.5 apresenta as ferramentas *Power BI* e *Datawrapper* para visualização de dados. Por fim, a Subseção 2.6 apresenta os trabalhos relacionados.

### Conceito de Visualização de Dados

A visualização de dados é a apresentação e a representação visual dos dados com o objetivo de facilitar o entendimento (KIRK, 2019). Representação envolve tomar decisões sobre como apresentar os dados visualmente para que a compreensão do assunto abordado seja acessível para o público. Está relacionado aos gráficos e ao ato de selecionar o gráfico correto para mostrar as características mais relevantes dos dados. A apresentação diz respeito ao modo no qual um trabalho de visualização de dados é planejado para ser transmitido ao público, o que independe do meio ou do método de divulgação.

Para Wilke (2019), "a visualização de dados é parte arte e parte ciência. O desafio é fazer a arte certa sem errar na ciência e vice-versa". Para o autor, uma visualização de dados deve transmitir os dados com precisão, não deve enganar ou distorcer e deve ser esteticamente agradável.

Kirk (2019) destaca o processo de entendimento de uma visualização de dados, que é composto por três fases, quais sejam: percepção, interpretação e compreensão. A primeira fase é a percepção, que é sobre o ato de ler os gráficos. Nesta fase, o espectador observa como os dados estão representados e cria uma ideia inicial das principais características apresentadas. A segunda fase é a interpretação, que é "a tradução das observações iniciais em conhecimento qualitativo e/ou quantitativo". Segundo o autor, a interpretação envolve assimilar o que o espectador observou com o que ele sabe acerca do contexto. A terceira e última fase é a compreensão, que diz respeito ao ato do espectador descobrir o significado da interpretação feita e extrair o que é importante para ele.

### 1. Visualização, Análise e História com Dados

A visualização de dados tem uma grande relevância no processo de análise de dados e constitui uma de suas etapas, seja ela exploratória ou explanatória.

A análise exploratória é a análise inicial dos dados brutos, que tem o objetivo de extrair informações relevantes e formular as primeiras hipóteses acerca do conjunto analisado. A

visualização de dados contribui de maneira expressiva neste contexto, pois auxilia a visualizar padrões que podem ser relevantes ou descartáveis na primeira análise (GOMES; TAVARES, 2011).

A análise explanatória é a apresentação das informações extraídas dos dados para o público-alvo para auxiliar o processo de tomada de decisão. Este tipo de análise conta com o apoio da visualização de dados, haja vista que ela facilita o entendimento. Nesta análise, é aplicada a técnica de *Data Storytelling*, que consiste em conduzir o público por meio de uma história com início, meio e fim, por meio dos dados. A história deve conter um ou mais objetivos centrais e terminar com uma recomendação específica (KNAFLIC, 2019).

## 1.1 Tipos de Visualização de Dados

Nesta subseção, são apresentados os principais tipos de visualização de dados usados neste trabalho: texto simples e gráficos. Gráficos possibilitam visualizar facilmente padrões, picos, alterações de resultado e tomar decisões com mais agilidade.

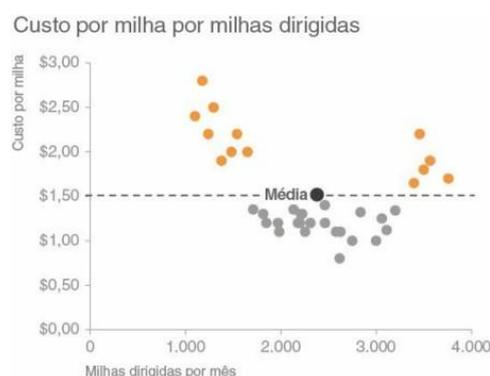
### 1.1.1 Texto Simples

O texto simples é uma visualização que é usada quando há apenas um ou dois números para compartilhar com o espectador. Ademais, o fato de se ter apenas alguns números não significa que seja necessário um gráfico. Entretanto, aconselha-se ter cautela sempre que se reduz vários números a apenas um, pois o contexto pode ser perdido ao tomar esta decisão (KNAFLIC, 2019).

### 1.1.2 Gráfico de Dispersão

O gráfico de dispersão apresenta o relacionamento entre duas ou mais variáveis quantitativas, que podem ser de diferentes categorias (KIRK, 2019). Este tipo de visual é muito útil para mostrar a relação entre duas variáveis, pois é permitido codificar dados simultaneamente em um eixo x horizontal e um eixo y vertical para identificar a relação existente (KNAFLIC, 2019). A Figura 2 exemplifica este tipo de visual.

**Figura 1 – Exemplo de Gráfico de Dispersão**

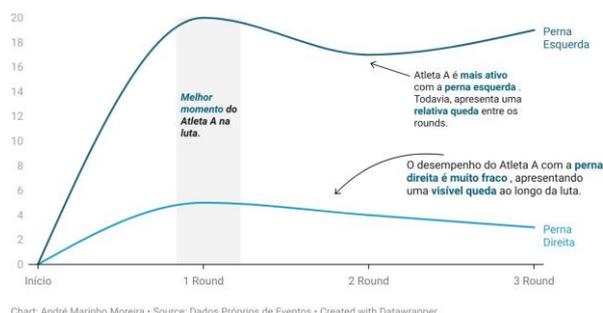


Fonte: KNAFLIC (2019, p. 40).

### 1.1.3 Gráfico de Linhas

Os gráficos de linha são usados para mostrar como os dados quantitativos se comportam ao longo do tempo (KIRK, 2019). O tempo é representado ao longo do eixo x horizontal do gráfico e a variável de estudo no eixo y vertical, com intervalos coerentes (KNAFLIC, 2019). A Figura 3 é um exemplo desta visualização e apresenta os dados do desempenho das duas pernas de um atleta ao longo de uma luta.

Figura 2 – Exemplo de Gráfico de Linha



Fonte: Elaborada pelos autores.

### 1.1.4 Gráfico de Barras

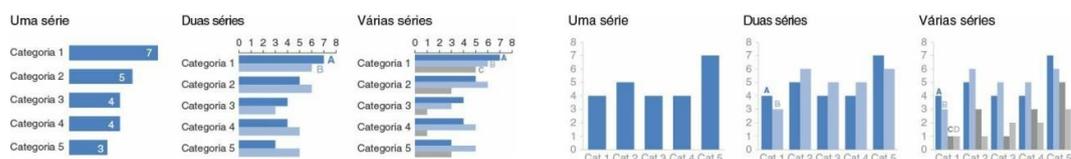
O gráfico de barras compara valores quantitativos de diferentes itens. O comprimento das barras é proporcional ao valor de cada item em estudo (KIRK, 2019).

Para Knaflic (2019), os gráficos de barras devem ser usados justamente por serem comuns, o que diminui a curva de aprendizado para o público e permite que se tenha facilidade em extrair informação da apresentação. A autora também salienta a importância deste tipo de visualização sempre ter uma linha de base zero (cruzamento do eixo x e eixo y em zero), caso contrário, a comparação visual é falsa.

É necessário ter atenção à ordem das barras. Se elas representam categorias não ordenadas, então, se faz necessário ordená-las de forma crescente ou decrescente de acordo com os valores dos dados (WILKE, 2019).

Existem algumas variações do gráfico de barras, que pode ser representado tanto na horizontal quanto na vertical, podendo apresentar uma, duas ou diversas séries (KNAFLIC, 2019). A Figura 4 exemplifica este tipo de gráfico na horizontal e na vertical respectivamente.

Figura 3 – Exemplo de Gráfico de Barras Horizontal e Vertical

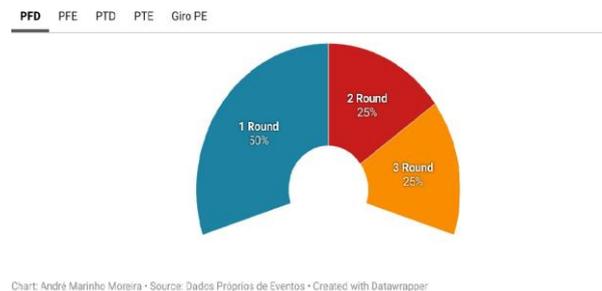


Fonte: KNAFLIC (2019, p. 52).

### 1.1.5 Gráfico de Rosca

O gráfico de rosca é utilizado para mostrar as partes proporcionais de um todo. Segundo o tutorial do Datawrapper (2020), não é recomendado que um gráfico de rosca tenha muitas fatias, o que prejudica sua leitura. A Figura 5 é um exemplo desta visualização com a proporção de chutes com a perna direita de um atleta ao longo de três *rounds*.

**Figura 4 – Exemplo de Gráfico de Rosca**



Fonte: Elaborada pelos autores.

## 1.2 Aspectos da Percepção Humana na Visualização de Dados

Após a escolha do visual adequado para a apresentação das informações, se faz necessário pensar nos aspectos da percepção humana para que a informação seja captada com facilidade pelo público. Para Gomes e Tavares:

[...] uma visualização, para além de ser uma construção interna da mente, pode funcionar como uma ferramenta cognitiva, tornando-se um artifício externo de apoio ao processo de tomada de decisão e construção de conhecimento, que utilize e maximize as capacidades perceptivas e cognitivas do ser humano (GOMES; TAVARES, 2011).

Ao elaborar a visualização de dados, é necessário considerar dois aspectos: sinal e ruído. O sinal é a principal informação a ser passada com o visual e ela deve ser considerada e identificada. O ruído diz respeito aos excessos que dificultam a leitura do gráfico para o espectador e estes devem ser eliminados para facilitar o entendimento. Para isso, é necessário considerar os Princípios da Gestalt de Percepção Visual (KNAFLIC, 2019). Para a autora, estes princípios facilitam o entendimento de como as pessoas veem, o que pode ser usado para identificar elementos desnecessários e facilitar o processamento das comunicações visuais.

Os princípios da Gestalt foram descritos por Max Wertheimer, que os considerava como leis da organização perceptiva. Para os gestaltistas, o todo é criado a partir da relação das partes

constituintes (GOMES; TAVARES, 2011). Os princípios que podem ser usados na visualização de dados são: proximidade, similaridade, acercamento, fechamento, continuidade e conexão.

A proximidade é quando o espectador é levado a pensar que objetos próximos fazem parte de um grupo. Já o princípio da similaridade dita que objetos semelhantes em forma, cor, orientação ou tamanho são percebidos como relacionados ou pertencentes à parte de um grupo (KNAFLIC, 2019).

O princípio de acercamento dita que "objetos fisicamente delimitados fazem parte de um grupo (KNAFLIC, 2019)". Pode-se aproveitar tal princípio na visualização de dados por meio da distinção dos dados, por exemplo.

O princípio do fechamento dita que elementos são agrupados caso se completem, o que provoca o o cérebro humano a formar os contornos sem a existência deles. No que diz respeito ao princípio da continuidade, ele dita que os objetos, ao serem observados, provocam a busca de um caminho e criam uma continuidade natural. A continuidade pode ser feita por meio de cores e formas (KNAFLIC, 2019).

O princípio da conexão dita que objetos conectados fazem parte de um grupo. Segundo Knafllic (2019), "a propriedade conectiva normalmente tem um valor associativo mais forte do que cor, tamanho ou forma similar". Na visualização de dados, a conexão é usada no gráfico de linhas para conectar valores e facilitar a observação da mudança ao longo do tempo.

Gomes e Tavares (2011) dão ênfase a um importante fator da percepção visual humana na visualização de informação, que é o uso elementos pré-atentivos. Os elementos pré-atentivos são:

[...] um conjunto de propriedades visuais que são detectadas de forma muito rápida e precisa pelo sistema visual de baixo nível [...] traduz a noção de velocidade e da facilidade com que certas propriedades são identificadas pelos humanos nas representações visuais (GOMES; TAVARES, 2011).

Essas propriedades visuais incluem cor, forma, movimento e localização espacial, delimitados por um espaço de visualização. Tais elementos são utilizados na visualização de dados quando se faz necessário atrair o foco da atenção do público para uma informação.

### **1.3 Ferramentas de Visualização de Dados**

Existe uma significativa variedade de ferramentas de visualização de dados. É necessário escolher a ferramenta que melhor atende as necessidades do problema em questão. Neste trabalho, são utilizadas as ferramentas *Power BI* e *Datawrapper*.

#### **1.3.1 Power BI**

O *Power BI* é uma ferramenta de visualização e análise de dados desenvolvida pela *Microsoft*. É uma ferramenta que dispõe de uma variedade de funcionalidades que auxiliam no

processo de criação de diversos tipos de visualização de dados.

Segundo o Blog Microsoft Power BI (2021), a ferramenta é líder do quadrante gartner no que diz respeito ao processo de análise de dados. Se trata de uma tecnologia bastante acessível, pois conta com uma ampla documentação, comunidade e tutoriais gratuitos na Internet, disponibilizados pela própria *Microsoft*. Além disso, a tecnologia conta com licenças gratuitas e licenças pagas.

### **1.3.2 Datawrapper**

O *Datawrapper* é uma ferramenta de visualização de dados na WEB que permite a criação de visuais gráficos interativos de uma forma simplificada. A ferramenta foi criada por uma equipe de dezessete pessoas, composta por *designers*, jornalistas e desenvolvedores que trabalharam em grandes empresas, como The New York Times, *Bloomberg* e entre outras<sup>2</sup>.

O site do *Datawrapper* conta com uma variedade de materiais gratuitos de treinamento para que os usuários consigam usá-lo de maneira efetiva. Além disso, a ferramenta conta com planos gratuitos e pagos, ambos com funcionalidades que atendem as necessidades básicas dos usuários no que diz respeito à criação de visualizações de dados como gráficos, mapas e tabelas.

## **1.4 Trabalhos Relacionados**

Com o objetivo de entender como é trabalhada a visualização de dados em diferentes estudos, foram considerados alguns trabalhos. O primeiro trabalho é de Lemes (2019), que aborda a visualização de dados para o acompanhamento dos dados gerados pela Bengala Enativa Afetiva, que é um circuito eletrônico embarcado em uma bengala. O segundo trabalho é de Moroço e Noman (2020), que aborda a visualização de dados acerca do desempenho do sistema nacional de saúde do Brasil. No terceiro, Andrade et al. (2017) realizam uma análise e visualização de dados de rastreamento da logística urbana. Por fim, Santos e Freitas (2016) discorrem sobre a visualização de dados com o uso da biblioteca D3 (Data-Driven Documents) para facilitar o processo de análise de dados e tomada de decisão do Exame Nacional do Ensino Médio (ENEM).

Lemes (2019) em seu trabalho junto ao Laboratório de Arte e Tecnociência da Universidade de Brasília, que desenvolve uma aplicação ubíqua na área da saúde, a Bengala Enativa Afetiva, propõe uma visualização dos dados gerados pelo equipamento, que são sinais captados que permitem avaliar problemas de saúde. O trabalho tem como objetivo facilitar o monitoramento e o acompanhamento dos sintomas decorrentes de diversas patologias. Foi desenvolvida uma aplicação web disposta de 5 diferentes formas de visualização de dados,

---

<sup>2</sup> DATAWRAPPER. **About datawrapper - our team, product, and the background.** [S.l.], 2020. Disponível em: <<https://www.datawrapper.de/about-us>>. Acesso em: 24 de ago. 2021.

como gráfico de linhas, barras, histogramas e outros, que foi disponibilizada *online* para os usuários. É possível acessar cada forma de visualização, com permissão de acesso à base de dados para possíveis alterações e possibilidade de inserção de um novo arquivo para atualizar os dados exibidos em cada gráfico.

O trabalho de Moroço e Noman (2020) visa desenvolver visualizações gráficas com o objetivo de melhorar a visibilidade dos problemas encontrados no desempenho do sistema nacional de saúde do Brasil, cujos dados são coletados pelo próprio IBGE. Os autores desenvolveram uma aplicação em *Power BI* disposta em uma grande variedade de visualizações, que transmitem informações relacionadas a atendimentos domiciliares, procura de atendimento, diagnósticos de doenças, grau de limitação a atividades habituais decorrentes de alguns problemas como depressão, colesterol alto, câncer e entre outros. Foi observado que o trabalho simplificou o entendimento das informações do desempenho nacional do sistema de saúde, possibilitando que a tomada de decisão seja tomada com mais eficiência.

No artigo de Andrade e outros (2017), foi proposta uma visualização de dados para entender a logística urbana nas principais capitais do país. É estudado o comportamento do transporte urbano de carga por meio de dados logísticos como distribuição de volume de caminhões ao longo do tempo, distribuição dos pontos de entrega, etc. Os autores, por meio de visualizações ao longo do tempo, como gráficos de barras e mapas de calor, conseguiram observar padrões da distribuição dos caminhões por dia da semana. O trabalho realizado foi de grande contribuição para a tomada de decisão, haja vista que as informações extraídas das análises e visualizações feitas podem ser aplicadas para identificar um diagnóstico da logística urbana e propor políticas públicas mais eficazes neste sentido.

Por fim, Santos e Freitas (2016) desenvolvem uma aplicação WEB de visualização de dados como apoio à análise de dados e à tomada de decisão acerca da educação brasileira. Os autores utilizam os dados abertos disponibilizados pelo MEC, que contêm informações acerca do desempenho geral dos estudantes no Exame Nacional do Ensino Médio (ENEM). O trabalho é de grande relevância, haja vista que foram descobertos padrões relevantes e de utilidade pública por meio da análise comparativa, como por exemplo, a confirmação do fato de que os estudantes de escolas particulares ou que cursam maior tempo nestas terem uma média geral maior do que aqueles que estudaram em escolas públicas.

## **2 METODOLOGIA**

Nesta seção, é apresentada a metodologia deste trabalho. A Subseção 3.1 apresenta o estudo em que a metodologia foi aplicada e a Seção 3.2 aborda as etapas metodológicas.

### **2.1 O campo de estudo: o parataekwondo**

Sediado na Universidade Federal de Minas Gerais (UFMG), o Centro de Treinamento Esportivo é uma instituição pública do Governo Federal que fomenta o treinamento esportivo de

alto rendimento, olímpico e paralímpico. O Parataekwondo é uma das modalidades ofertadas nessa instituição. As atividades são ofertadas para qualquer pessoa que queira se desenvolver como atleta e são realizadas de segunda a sexta-feira em diversos horários conforme a disponibilidade de cada modalidade esportiva. Segundo o *site* Info Escola (2008), o Taekwondo é uma arte marcial de origem coreana e que teria surgido há aproximadamente 2000 anos atrás. O termo Taekwondo significa "caminho dos pés, das mãos e do espírito" e tem a proposta de desenvolver o atleta fisicamente e mentalmente, prezando pelo respeito, disciplina e humanidade. O Parataekwondo é uma adaptação desta arte marcial para pessoas que possuam deficiência física, visual, motora ou intelectual. Vale ressaltar que o Brasil ficou em primeiro lugar nesta modalidade, que teve sua primeira edição nos Jogos Paralímpicos em 2021.

O trabalho é realizado por um professor que acompanha o rendimento dos atletas durante os treinos e competições e realiza um estudo dos adversários de seus atletas. Este acompanhamento e estudo envolve uma quantidade considerável de dados relacionados às avaliações físicas que são realizadas e ao estudo dos adversários. Estes dados são coletados pelo próprio professor e sua equipe e armazenados em planilhas de avaliação para serem analisados posteriormente.

Diante do exposto, a visualização de dados é de importância significativa, haja vista que é possível manipular estes dados, gerar visualizações relacionadas ao desempenho dos atletas, dos adversários e analisá-las com o objetivo de facilitar o acompanhamento dos atletas para auxiliar a tomada de decisão.

## **2.2 Fases da Metodologia**

A metodologia deste trabalho é composta por 5 etapas, descritas a seguir: adesão à base de dados, estruturação e limpeza dos dados, elaboração dos relatórios, apresentação dos relatórios e avaliação do processo por meio da aplicação de entrevista.

### **2.2.1 Adesão à Base de Dados**

Esta etapa consiste da adesão da base de dados a ser estudada. O principal objetivo desta etapa é a identificação dos tipos de dados da base. Para isso, o professor responsável pela geração desses dados foi consultado. Desse modo, buscou-se entender os dados e as necessidades quanto à visualização e à análise dos dados disponíveis e planejar os relatórios a serem elaborados.

A base de dados é composta por dados quantitativos antropométricos dos atletas que facilitam o acompanhamento do alto rendimento dos mesmos, considerando-se que estes tipos de dados ajudam a entender o desempenho em treinos e lutas. Esta base é composta por um total de 23 variáveis para estudo e cerca de 60 registros distintos dessas variáveis. São registrados dados como nome, peso, massa, estatura, índice de massa corporal, percentual de gordura, massa de gordura, massa muscular, quantidade de saltos, quantidade de chutes, entre outros.

Além disso, há dados quantitativos relacionados ao desempenho de possíveis adversários dos atletas em diferentes competições. Dados considerados apropriados como chutes com determinada perna e quantidade de vezes que uma técnica foi utilizada são registrados para preparar os atletas para as competições.

### 2.2.2 *Limpeza e Estruturação de Dados*

Após a adesão da base de dados, é preciso que os dados sejam tratados. A etapa de limpeza e estruturação de dados é fundamental, pois, a partir dela, atinge-se a qualidade dos dados necessária no processo de visualização e tomada de decisão. Esta etapa tem como objetivo eliminar anomalias dos dados e estruturá-los para que as visualizações não sejam prejudicadas posteriormente.

Segundo Kumar e Khosla (2018), dados não estruturados podem ser compostos de dados faltantes, dados duplicados, dados incorretos e dados não integrados. Como a base de dados em estudo é gerida e alimentada por profissionais que acompanham os atletas, ela não dispõe de dados incorretos ou duplicados, mas possui dados faltantes, que são dispersos em diferentes planilhas. Dessa maneira, os dados faltantes foram tratados e foi realizada a integração dos dados por meio do *Power Query*, que é uma ferramenta da *Microsoft* para limpeza e manipulação de dados inclusive dentro no *Power BI*.

### 2.2.3 *Elaboração dos Relatórios*

Com os dados tratados, finalmente, é possível elaborar os relatórios. Entende-se por relatório uma interface composta por uma ou mais páginas, que contém um conjunto de visualizações que auxiliam a tomada de decisão<sup>3</sup>. Esta etapa tem como objetivo utilizar o *Power BI* para a elaboração de relatórios mais abrangentes e interativos, com diversas visualizações que permitam acompanhar o desempenho de todos os atletas durante as avaliações físicas em um único local. Já com o auxílio do *Datawrapper*, foi criado um relatório mais pontual que consiste em analisar as características de um adversário específico em uma competição. Este relatório foi feito no formato de apresentação e foi possível exportar os visuais criados no *Datawrapper* e analisá-los, destacando os principais pontos visualizados.

A composição das visualizações, que fazem parte dos relatórios, foi feita de acordo com as etapas propostas por Knaflic (2019): seleção dos visuais adequados, remoção de excessos e focalização da atenção do público.

A seleção dos visuais adequados foi feita por meio de uma consulta ao professor que acompanha os atletas, com o objetivo de verificar quais informações são importantes visualizar. Deste modo, foi possível selecionar os tipos de gráficos adequados para cada situação e implementá-los.

---

<sup>3</sup> DOCUMENTAÇÃO MICROSOFT POWER BI. **Relatórios no Power BI**. [S.l.], 2020. Disponível em: <<https://docs.microsoft.com/pt-br/power-bi/consumer/end-user-reports>>. Acesso em: 29 de mai. 2021.

Os gráficos no *Power BI* e no *Datawrapper*, por padrão, vêm com informações que podem ser desnecessárias para a comunicação, fator que provoca uma carga cognitiva excessiva e dificulta a interpretação por parte do público. Segundo Knaflíc (2019), carga cognitiva é o esforço mental exigido para aprender novas informações e é preciso minimizá-la. Dessa maneira, foi feita a remoção dos excessos considerando os princípios da Gestalt, que consistem em identificar e eliminar o que não está acrescentando valor informativo.

Por fim, foi trabalhada a atenção do público por meio de técnicas como a seleção de atributos pré-atentivos, para atrair o foco do público para a principal informação a ser transmitida. O objetivo principal é fazer com que as informações dos gráficos sejam processadas de maneira objetiva e clara.

Finalizada a composição das visualizações, o relatório é finalizado e compartilhado. No *Power BI*, é feita uma publicação na WEB e é planejado seu compartilhamento posterior junto aos profissionais do Centro Esportivo. Já em um relatório pontual, há a possibilidade de exportar para o formato .pdf ou fazer o compartilhamento via *email*.

#### **2.2.4 Apresentação dos Relatórios**

Com os relatórios gerados e devidamente compartilhados, inicia-se a etapa de apresentação. Faz parte da apresentação a análise explanatória, em que é feito o destaque das principais informações perceptíveis, interpretadas e compreendidas a partir das visualizações que compõem os relatórios. Nesta apresentação, foi aplicada a técnica de *Data Storytelling*, que consistiu em elaborar uma narrativa por meio dos dados apresentados e engajar o público para que a informação fosse transmitida da melhor maneira possível e se tornasse mais fácil de ser compreendida (KNAFLIC, 2019).

A expectativa desta etapa é que sejam destacados os pontos mais importantes das visualizações, para que a apresentação se torne mais concisa e consiga oferecer *insights* valiosos para o professor. É uma tentativa de guiá-lo para uma tomada de decisão a respeito do desenvolvimento de seus atletas, para a condução dos treinos e para o planejamento para competições.

#### **2.2.5 Realização de Entrevista**

Esta etapa tem como objetivo avaliar o processo desenvolvido por meio da aplicação da técnica de entrevista. A escolha desta técnica se justifica pelo contato mais próximo com o profissional e com os atletas que fazem parte da pesquisa. O sujeito da pesquisa foi o professor que acompanha os atletas durante os treinos e as competições, que é um profissional graduado em Educação Física com conhecimento técnico compatível com o teor das perguntas.

A entrevista aplicada foi do tipo estruturado, constituída por perguntas previamente definidas no que diz respeito aos relatórios apresentados. A expectativa foi avaliar se o processo contribuiu para a apropriação dos dados pelos envolvidos na modalidade Parataekwondo, bem como a utilização e interpretação dos dados de modo assertivo durante os treinamentos e

competições. No primeiro tópico da entrevista, são feitas perguntas específicas a respeito do processo de visualização dos dados apresentado nos relatórios, com o intuito de averiguar se a composição apresentada facilitou a transmissão da informação. No segundo tópico, são apresentadas perguntas referentes ao processo de apropriação dos dados e suas aplicabilidades. A ideia é que o entrevistado possa compartilhar sua experiência a partir do que foi apresentado, informando se as visualizações influenciaram as decisões tomadas e seu modo de agir.

### **3 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS**

As cinco etapas da metodologia proposta foram desenvolvidas e apresentadas nesta seção. As duas primeiras etapas, descritas na Subseção 4.1, consistem em aderir à base de dados, realizar o processo de limpeza e estruturá-la. As três últimas etapas são apresentadas na Subseção 4.2 e dizem respeito à elaboração e à apresentação dos relatórios. Por fim, a Subseção 4.3 descreve a realização de entrevista e a análise das respostas.

#### **3.1 Adesão, Limpeza e Estruturação dos Dados**

Os experimentos iniciais foram realizados com o objetivo de identificar a base de dados relacionada a este trabalho. Para obtê-la, foi realizada uma reunião com o professor que acompanha os atletas e foram obtidos arquivos em Excel contendo os dados de avaliações físicas e de estudos de adversários. Foi criado, a partir dos arquivos, um dicionário de dados, que contém as informações relacionadas aos dados coletados. Além disso, foram levantados os requisitos relacionados aos relatórios elaborados e informado o que o usuário deseja visualizar.

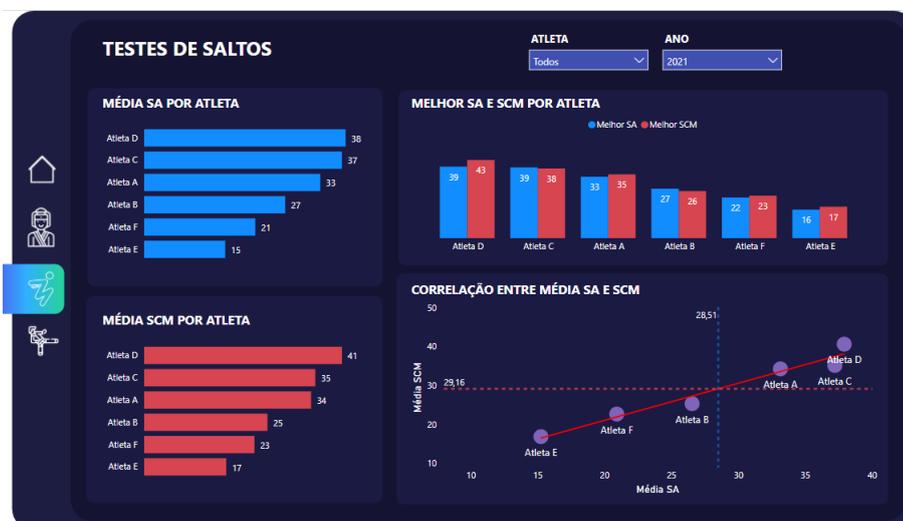
Após a adesão à base de dados, foi realizada sua estruturação e sua limpeza. Para isso, foi criado um banco de dados em MySQL, que é um sistema gerenciador de banco de dados gratuito, onde foram carregados os dados das planilhas Excel relacionados às avaliações físicas. Nesta transição, foi possível realizar a limpeza dos dados, eliminar valores vazios e criar os relacionamentos entre as tabelas por meio da linguagem SQL (Linguagem de Consulta Estruturada), usada para manipulação e definição dos dados. Para finalizar a estruturação, foi realizada a conexão do *Power BI* com o banco de dados. O tratamento final foi feito no *Power Query*. Foi possível identificar com mais clareza os valores ausentes, haja vista que a ferramenta faz esta detecção automática. Para eliminar estes valores, foi observado se as colunas com os valores ausentes não tinham impacto na análise final e, depois desta avaliação, as colunas foram excluídas. Por fim, os dados foram carregados de fato para o *Power BI*, para a elaboração do relatório de avaliações físicas. No que diz respeito ao *Datawrapper*, os dados relacionados ao estudo dos adversários foram exportados do banco de dados e carregados diretamente de uma planilha, considerando-se que a ferramenta apenas permite conexão com planilhas e arquivos csv.

### 3.2 Elaboração e Apresentação dos Relatórios

Os relatórios foram elaborados com o auxílio das tecnologias *Power BI* e *Datawrapper*. Foi criado um relatório em que o professor consegue acompanhar o alto rendimento de seus atletas por meio dos resultados das avaliações físicas. Além disso, foi criado um relatório de análise de adversário, em que o professor consegue realizar um estudo de um possível adversário de um de seus atletas e orientar melhor sua decisão no que diz respeito a condução dos treinos. O relatório de avaliações físicas pode ser acessado por meio do *link* <https://bit.ly/2YIKzo4>. Já o relatório de análise de adversário está disponível no *link* <https://bit.ly/3FPv3YH>. A seguir, é apresentado o relatório de avaliações físicas.

Na primeira página do relatório, são apresentados os dados relativos à composição corporal dos atletas, quais sejam, percentual de gordura, massa magra e massa gorda. Foram utilizados os tipos de visuais em texto simples, haja vista que a comparação é entre os dados do próprio atleta. É possível visualizar os dados dos anos de 2020 e de 2021 e a respectiva comparação entre os dados dos dois anos para que o professor possa visualizar a evolução de cada atleta individualmente. No caso do Atleta A, por exemplo, a massa magra, em 2020, era de 37,50 Kg. Em 2021, passou para 44,59 Kg, gerando um aumento percentual de 18,91%. A Figura 5 apresenta a segunda página do relatório.

**Figura 5 – Segunda Página do Relatório de Avaliações Físicas**

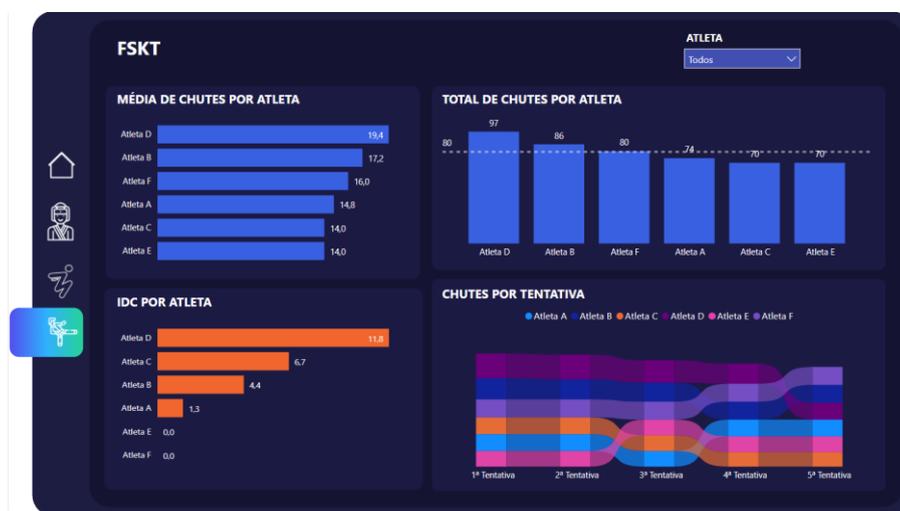


Fonte: Elaborada pelos autores.

Na segunda página, são apresentados os dados dos testes de saltos normais e contra movimento. Durante este teste, os atletas realizam três tentativas de saltos normais (SA) e saltos contra movimento (SCM). O SCM é realizado quando o salto do atleta contraria o movimento de descida do corpo. Em seguida, é obtida a média das tentativas e a melhor das médias. No primeiro gráfico, é possível visualizar a média de saltos por atleta e aquele que obteve a melhor média. O gráfico localizado no canto inferior esquerdo apresenta a média de SCM por

atleta. Os atletas D e C, por exemplo, obtiveram as melhores médias. No canto superior direito, é apresentado o melhor SA e o SCM por atleta e é possível verificar quem obteve o melhor desempenho por meio de um gráfico de barras na horizontal. Neste caso, os atletas D e C obtiveram o melhor desempenho. Por fim, o gráfico de dispersão apresenta a correlação entre a média de SA e SCM, onde foi traçada uma linha de tendência e linhas médias. Com a linha de tendência, foi possível observar que houve uma correlação positiva entre as tentativas realizadas no teste realizado em 2021 pois, à medida que a média de SA cresceu, a média de SCM cresceu também, ou seja, os atletas apresentaram um bom equilíbrio neste teste no ano de 2021. A Figura 6 apresenta a terceira e última página do relatório.

**Figura 6 – Terceira Página do Relatório de Avaliações Físicas**



Fonte: Elaborada pelos autores.

No que diz respeito a última página do relatório, é possível visualizar os resultados do *Frequency Speed of Kick Test*, que é o Teste de Frequência de Chute Rápido (FSKT), realizado para avaliar a frequência de chutes de um determinado atleta (SANTOS et al., 2016). Por meio das visualizações, o professor consegue perceber qual o atleta que possui o melhor chute. O primeiro gráfico mostra a média de chutes por atleta e é possível verificar que o Atleta D possui a melhor média, com 19,4. O segundo ilustra o total de chutes por atleta, com a possibilidade de visualizar quais atletas ficaram acima e abaixo da média geral. Os atletas D e B, por exemplo, ficaram acima da média geral. O terceiro gráfico apresenta o Índice de Decréscimo de Chute (IDC), métrica em que é possível verificar qual atleta teve um melhor desempenho ao final do teste. Os atletas D e C possuem os melhores índices, com 11,8 e 6,7 respectivamente. A última visualização da página é um gráfico de faixas, que possui a mesma ideia do gráfico de linhas, ou seja, apresentar o comportamento de uma variável ao longo do tempo. Nesta visualização, é possível identificar o desempenho de cada atleta individualmente ao longo das cinco etapas do teste, cujo gráfico ilustra a evolução do desempenho por meio da quantidade de chutes por etapa. Foi possível observar que os atletas D e B obtiveram o melhor desempenho ao longo do

teste, mantendo-se acima dos demais na maioria das tentativas.

No que se refere ao relatório de análise de adversários, foram analisados os dados de desempenho relacionados aos totais de técnicas que foram utilizadas ao longo do combate e as diferentes posições de cada perna e as variações das duas pernas. Dessa forma, foi possível observar padrões, identificar os pontos fracos e fortes do adversário e as oportunidades a serem exploradas. Por fim, os relatórios foram apresentados ao professor que acompanha os atletas. Foram destacados os pontos importantes de cada relatório, visando guiar a tomada de decisão, no que diz respeito ao desempenho dos atletas.

### **3.3 Realização da Entrevista e Análise das Respostas**

Após a apresentação dos relatórios, a entrevista foi realizada com o professor que acompanha os atletas. Foram feitas perguntas a respeito do processo de visualização dos dados relacionados aos atletas e da importância dos relatórios para as tomadas de decisão diárias no âmbito do Parataekwondo, no que diz respeito aos treinamentos e à elaboração de estratégias para competições, buscando compreender a importância dos dados e de sua apropriação.

No que diz respeito ao processo de visualização de dados, observou-se que não houve dificuldade por parte do espectador no que diz respeito às fases de percepção, interpretação e compreensão dos dados apresentados nos relatórios. A maneira como estes foram tratados, analisados e expostos facilitou a leitura e a interpretação das informações transmitidas. Verificou-se também que não houve a presença de excessos ou acúmulo de informações que tenham dificultado o processo de visualização. Além disso, segundo o professor entrevistado, os tipos de visualizações utilizados foram adequados ao processo e proporcionaram uma significativa contextualização.

No que se refere ao processo de apropriação dos dados e do seu uso, observou-se que a forma visual do processo apresentada no relatório de avaliações físicas foi de significativa importância para as tomadas de decisão diárias. No que diz respeito aos treinamentos, possibilitou novos olhares no que se refere ao acompanhamento do desenvolvimento dos atletas, permitindo a observação de nuances e padrões que antes não eram possíveis de serem observados. O relatório facilitou a comunicação do professor com os atletas durante a condução dos treinos e foi possível explicar como está a evolução de cada um e o que é preciso melhorar ou potencializar, já que o relatório possui uma organização visual que propicia isto.

Em relação ao relatório de análise de adversários, observou-se uma significativa contribuição deste para a tomada de decisão no que diz respeito à elaboração de estratégias, pois este foi utilizado na preparação de um atleta para uma competição importante realizada em 2021. A visualização dos dados do adversário analisado propiciou uma maior compreensão da análise do comportamento do mesmo. Os treinos preparatórios para a competição junto ao atleta cujo relatório foi apresentado foram realizados a partir dos padrões identificados nas análises das visualizações do adversário. Como resultado, o atleta teve um bom desempenho na competição

e conseguiu obter o melhor resultado contra o adversário analisado até então, competindo em alto nível. O atleta já o enfrentara antes, mas sempre era superado com grande desvantagem.

#### 4 CONSIDERAÇÕES FINAIS

O uso da tecnologia para a visualização e para a análise de dados é de significativa importância para o contexto esportivo nos dias atuais. Esportes com repercussão midiática como futebol, basquete, vôlei, entre outros, não medem esforços no desenvolvimento de uma cultura orientada a dados, com métricas e análises que permitam tomadas de decisão mais objetivas em diversos contextos. Todavia, durante o desenvolvimento deste trabalho, foi possível observar que esportes com pouca repercussão midiática como o parataekwondo, carecem neste sentido, tendo em vista a tamanha dificuldade na procura por fontes de dados, na coleta e no alcance da qualidade dos dados para que seja facilitada a análise.

As fases de adesão, limpeza e estruturação dos dados foram importantes para o desenvolvimento do trabalho, pois permitiram refinar os dados trabalhados e, conseqüentemente, atingir uma visão mais abrangente do contexto relacionado ao estudo realizado, possibilitando um processo de análise mais preciso. Além disso, é importante ressaltar a relevância do processo de seleção e elaboração das visualizações adequadas, considerando os aspectos da percepção humana, reforçados, na entrevista, pelos comentários do professor que acompanha os atletas.

A realização de entrevista foi importante para avaliar o processo, pois as respostas demonstram a importância do estudo realizado. Foi possível observar que as decisões no que diz respeito ao acompanhamento dos atletas e à elaboração de estratégia para competições, foram tomadas com mais precisão a partir dos relatórios apresentados.

Como sugestões para trabalhos futuros, propõe-se a criação de um sistema de visão computacional, que atue na captação de movimentos de atletas durante treinos e competições, com o armazenamento simultâneo de dados em um banco de dados. Desse modo, espera-se que a coleta dos dados seja realizada com mais precisão e de forma automatizada, evitando inconsistências e facilitando o processo de visualização e de análise das informações que possam ser extraídas.

#### REFERÊNCIAS

ANDRADE, Patrícia Faias Laranjeiro de et al. Análise e visualização de dados de rastreamento para caracterização da logística urbana. **Revista Transportes**, São Paulo, v. 1, n. 1, p. 24–35, 2017.

BLOG MICROSOFT POWER BI. **Microsoft named a Leader in the 2021 Gartner Magic Quadrant for Analytics and BI Platforms**. [S.l.], 2021. Disponível em:

<<https://powerbi.microsoft.com/pt-br/blog/microsoft-named-a-leader-in-2021-gartner-magic-quadrant-for-analytics-and-bi-platforms/>>. Acesso em: 21 de abr. 2021.

DATAWRAPPER. **How to create a donut chart**. [S.l.], 2020. Disponível em: <<https://academy.datawrapper.de/article/25-how-to-create-a-donut-chart>>. Acesso em: 24 de ago. 2021.

GOMES, Leandro Filipe Oliveira; TAVARES, João Manuel R.S. Percepção humana na visualização de grandes volumes de dados. **10º Congresso Ibero-americano de Engenharia Mecânica (CIBEM 10)**, Porto, Portugal, v. 10, n. 10, p. 2–18, 2011. Disponível em: <<https://hdl.handle.net/10216/56574>>.

INFO ESCOLA. **Taekwondo**. [S.l.], 2008. Disponível em: <<https://www.infoescola.com/artes-marciais/taekwondo/>>. Acesso em: 21 de abr. 2021.

KIRK, Andy. **Data visualization: a handbook for data-driven design**. 2. ed. London, England: SAGE Publishing LTD, 2019. 242 p.

KNAFLIC, Cole Nussbaumer. **Storytelling com dados: um guia de visualização de dados para profissionais de negócio**. 2. ed. Rio de Janeiro: Alta Books, 2019. 312 p.

KUMAR, Virender; KHOSLA, Cherry. Data cleaning-a thorough analysis and survey on unstructured data. **2018 8th International Conference on Cloud Computing, Data Science Engineering (Confluence)**, p. 305–309, 2018. Disponível em: <<https://ieeexplore.ieee.org/document/8442950>>.

LEMES, João Paulo Borges. **Visualização de dados para bengala enativa afetiva**. 2019. 34f. Monografia (Conclusão de curso) — Universidade Federal de Ouro Preto, João Monlevade.

LEWIS, Michael. **Moneyball: the art of winning an unfair game**. 1. ed. New York: W.W.Norton Company, 2004. 316 p.

MOROÇO, Carlos Eduardo Versiani; NOMAN, Lucas Calhau. Visualização de dados acerca do desempenho do sistema nacional de saúde do Brasil. **Revista ABAKÓS**, Belo Horizonte, v. 1, n. 1, p. 2–25, 2020.

SANTOS, César Henrique Cardoso dos; FREITAS, Maykon Carlos de. **VISDADOSENEM: Visualização de dados do ENEM com a biblioteca D3**. 2016. 75f. Monografia (Conclusão de curso) — Universidade Federal de Santa Catarina, Araranguá.

SANTOS, Jonatas Ferreira da Silva et al. Influence of half-squat intensity and volume on the subsequent countermovement jump and frequency speed of kick test performance in taekwondo athletes. **Kinesiology**, v. 48, n. 1, p. 95–102, 2016. Disponível em: <<https://hrcak.srce.hr/160777>>.

SILVA, Fabiano Couto Corrêa da. Visualização de dados: passado, presente e futuro. **Liinc Em Revista**, Rio Grande do Sul, v. 15, n. 2, p. 2–18, 2019. Disponível em: <<https://doi.org/10.18617/liinc.v15i2.4812>>.

WILKE, Claus O. **Fundamentals of data visualization: a primer on making informative and compelling figures**. United States: O'Reilly, 2019. 370 p.