



Ferramentas Visuais para o Estudo de Análise de Dados: Um Mapeamento Sistemático

José Antônio Bellato Fonseca, Universidade Federal de Itajubá,
tonibellato@gmail.com, <https://orcid.org/0009-0007-0368-3547>

Rodrigo Duarte Seabra, Universidade Federal de Itajubá,
rodrigo@unifei.edu.br, <https://orcid.org/0000-0002-7465-2963>

Resumo: O mapeamento sistemático conduzido nesta pesquisa identificou 35 ferramentas visuais desenvolvidas para o estudo de análise de dados, com destaque para aquelas que utilizam elementos de gamificação em seus projetos. Os resultados revelaram que a análise exploratória de dados foi o tópico com maior abrangência e que o recurso de tabela de classificação foi o elemento de gamificação mais usado para estimular e aumentar o engajamento dos aprendizes. A principal contribuição do artigo reside no entendimento do estado da arte do desenvolvimento de ferramentas para o estudo de análise de dados, a visualização de um panorama histórico de uso dessas ferramentas e a projeção de possibilidades de concepção de novas aplicações que possam contribuir com o estudo introdutório na área.

Palavras-chave: ferramentas visuais, gamificação, análise de dados, mapeamento sistemático.

Visual Tools for Studying Data Analysis: A Systematic Mapping

Abstract: The systematic mapping conducted in this research identified 35 visual tools developed for studying data analysis, highlighting those that use gamification elements in their projects. The results revealed that exploratory data analysis was the topic with the greatest coverage and that the leaderboard feature was the gamification element most used to stimulate and increase learner engagement. The main contribution lies in understanding the state of the art in the development of tools for studying data analysis, the visualization of a historical overview of the use of these tools, and the projection of possibilities for designing new applications that can contribute to introductory studies in the area.

Keywords: visual tools, gamification, data analysis, systematic mapping.

1. INTRODUÇÃO

A habilidade de analisar grandes volumes de dados possibilita a extração de informações relevantes e a tomada de decisões baseadas em evidências, visando a resolução de problemas complexos ou a descoberta de conhecimento significativo (HICKS; IRIZARRY, 2018). Nesse contexto, fazer descobertas significativas sobre grandes massas de dados tem importância atualmente para as mais diversas áreas do conhecimento. Novos perfis profissionais, como cientistas, analistas e engenheiros de dados, dentre outros, sinalizam que interagir com e analisar dados são habilidades que transcendem as disciplinas quantitativas tradicionais (BURCKHARDT *et al.*, 2021). Dessa forma, a procura mundial por analistas que possam interpretar dados de forma eficaz e indicar melhores decisões cresce exponencialmente, aumentando também o número de cursos e disciplinas que, de alguma forma, se relacionam ou fazem uso da análise de dados (KHACHATRYAN, 2022).



Em princípio, é comum associar o estudo de análise de dados de nível introdutório ao aprendizado de programação. No entanto, argumenta-se que, com base na tecnologia e nos princípios da computação, é possível construir ferramentas que proporcionem aos aprendizes o desenvolvimento de habilidades necessárias à área – como pensamento multivariável, manipulação de dados de vários tipos e complexidades e aprendizagem ativa – e que não tenham a programação como enfoque principal. Tais ferramentas podem ainda introduzir conceitos de análise de dados de forma estruturada e apoiar a construção de habilidades estatísticas mais profundas (BURCKHARDT *et al.*, 2021).

Nesse cenário, argumenta-se que a gamificação tem o potencial de contribuir significativamente para melhoria dos resultados de aprendizagem, quando seus conceitos são aplicados em ferramentas de estudo das mais diversas áreas. Estimulados pelos efeitos que os elementos de jogos podem produzir, pesquisadores têm dedicado esforços no que tange à investigação sobre a influência da gamificação em contexto educativo, obtendo resultados favoráveis, como o aumento do engajamento, do conhecimento e da cooperação (HAKULINEN; AUVINEN, 2014; MAJURI *et al.*, 2018; OLIVEIRA *et al.*, 2022).

Este trabalho conduziu um mapeamento sistemático da literatura, visando identificar, analisar e comparar ferramentas visuais concebidas para o estudo de análise de dados, com destaque para as que incluem elementos de gamificação como recurso pedagógico. Isso posto, pretende-se apresentar um panorama histórico da criação e uso dessas ferramentas ao longo do tempo, bem como compreender o estado da arte neste campo de pesquisa. A identificação sistemática de recursos de estudo, na forma das mais variadas funcionalidades, em ferramentas de variados tipos e contextos de uso, poderá contribuir para futuros trabalhos de seleção desses recursos, representando a principal contribuição da pesquisa. Em adição, o artigo também abrange uma descrição de diversas situações de uso e avaliação das ferramentas identificadas, de forma a destacar os resultados obtidos com suas aplicações e observar desafios e oportunidades nesse contexto de pesquisa.

O restante do trabalho está estruturado da seguinte forma: a Seção 2 apresenta a metodologia de pesquisa empregada para procurar, extrair e classificar informações dos trabalhos encontrados; a Seção 3 discute os resultados; e, por fim, as conclusões do estudo são tecidas na Seção 4.

2. DEFINIÇÃO DA PESQUISA

Com o objetivo de identificar, analisar e classificar ferramentas visuais desenvolvidas para o estudo de análise de dados, destacar as que se beneficiam de elementos de gamificação e ainda examinar as formas como foram utilizadas, o mapeamento sistemático da literatura proposto por este trabalho segue os procedimentos definidos por Petersen *et al.* (2015), sendo estruturado de forma a empregar questões de pesquisa, protocolos de busca, critérios de seleção, classificação e extração de dados.

2.1 Questões de Interesse

Estabelecido o objetivo principal de compreender o estado da arte acerca das ferramentas concebidas para o estudo de análise de dados, evidenciando nelas recursos de gamificação, foram definidas quatro questões de pesquisa:

Q1: Quais ferramentas visuais, de contexto educacional, existem para o estudo de conceitos de análise de dados?



Q2: Quais conceitos de análise de dados são abordados por essas ferramentas?

Q3: Para quais públicos ou contextos essas ferramentas são voltadas?

Q4: Dentre as ferramentas identificadas, quais são e de que forma elas empregam elementos de gamificação?

A primeira questão permite identificar nos trabalhos mapeados quais ferramentas foram empregadas em cada estudo. A segunda questão possibilita o reconhecimento dos conceitos de análise de dados incorporados nas ferramentas. A terceira questão identifica o contexto ou o público principal para o qual a ferramenta foi desenvolvida. A última questão tem o objetivo de reconhecer quais são e de que forma essas ferramentas empregam elementos de gamificação.

2.2 Execução da Pesquisa

As buscas foram executadas nas principais bases de documentação científica: *ACM Digital Library*, *ERIC – Education Resource Information Center*, *IEEE Xplore*, *ScienceDirect*, *Scopus*, *Springer Link*, *Web of Science* e *Wiley Online Library*, acessadas por meio do Portal Capes. Foi utilizado ainda o *Google Scholar*, de forma a ampliar a abrangência de resultados e minimizar o risco de omissões. Também foram incluídos trabalhos obtidos por meio de *snowballing*, referenciados pelos artigos iniciais localizados.

Foram executados vários testes nas bases com o objetivo de ajustar a calibragem das *strings* de busca, de forma a otimizar os resultados da pesquisa. Devido às informações retornadas nessas buscas iniciais, decidiu-se por não empregar o termo ‘gamificação’ ou sinônimos nas *strings* definitivas, no caso das principais bases científicas. Essa decisão se deveu ao fato de ter sido observado que o uso desses termos restringia significativamente a abrangência de trabalhos retornados. Nas buscas iniciais, foi observado que vários trabalhos não mencionavam termos relacionados com ‘gamificação’ em seus títulos, palavras-chave ou mesmo em seus *abstracts*, porém por meio de uma leitura pormenorizada dos textos, pôde-se verificar que estavam presentes elementos de gamificação nas publicações.

Visando minimizar o risco de omissões e maximizar o número de resultados com potencial de resposta às questões de interesse da pesquisa, optou-se por utilizar os termos ‘gamificação’ ou ‘*gamified*’ somente na *string* definitiva do *Google Scholar*. Isso foi feito com a intenção de delimitar a quantidade e ajustar o escopo do conjunto de trabalhos retornados por essa plataforma, pois a busca sem esses termos retornava mais de 6000 trabalhos; com o uso deles, o quantitativo foi reduzido para 290, considerado um resultado satisfatório pelos autores deste trabalho. Vale destacar que não foram incluídas restrições de datas no mapeamento, de forma a possibilitar o desenho de um panorama histórico e cronológico da concepção e uso das ferramentas. As *strings* completas de busca, adaptadas para cada plataforma, podem ser consultadas no Quadro 1.

Quadro 1 – Strings de busca nas bases digitais (continua). Fonte: Os autores.

Fonte	Strings de Busca
ACM Digital Library	Title:(("teaching data analytics" OR "teaching data science" OR "teaching data analysis" OR "data analysis education" OR "data analytics education" OR "data science education") AND ("tool" OR "environment" OR "application" OR "platform")) OR Abstract:(("teaching data analytics" OR "teaching data science" OR "teaching data analysis" OR "data analysis education" OR "data analytics education" OR "data science education") AND ("tool" OR "environment" OR "application" OR "platform")) OR Keyword:(("teaching data analytics" OR "teaching data science" OR "teaching data analysis" OR "data analysis education" OR "data analytics education" OR "data science education") AND ("tool" OR "environment" OR "application" OR "platform"))
ERIC	("teaching data analysis" OR "teaching data analytics" OR "teaching data science" OR "data



	analysis education" OR "data analytics education" OR "data science education") AND ("tool" OR "environment" OR "application" OR "platform")
IEEE Xplore	((("All Metadata":"teaching data analysis" OR "All Metadata":"teaching data science" OR "All Metadata":"teaching data analytics" OR "All Metadata":"data analysis education" OR "All Metadata":"data analytics education" OR "All Metadata":"data science education") AND ("All Metadata":"tool" OR "All Metadata":"application" OR "All Metadata":"environment" OR "All Metadata":"platform"))))
Google Scholar	("teaching data analysis" OR "teaching data analytics" OR "teaching data science" OR "data analysis education" OR "data analytics education" OR "data science education") AND ("tool" OR "environment" OR "application" OR "platform") AND ("gamification" OR "gamified")
ScienceDirect	("teaching data analysis" OR "teaching data analytics" OR "teaching data science" OR "data analysis education" OR "data analytics education" OR "data science education")
Scopus	TITLE-ABS-KEY (("teaching data analytics" OR "teaching data science" OR "teaching data analysis" OR "data analysis education" OR "data analytics education" OR "data science education") AND ("tool" OR "software" OR "environment" OR "application"))
Springer Link	("teaching data analytics" OR "teaching data science" OR "teaching data analysis" OR "data analysis education" OR "data analytics education" OR "data science education") AND ("tool" OR "environment" OR "application" OR "platform")
Web of Science	("teaching data analysis" OR "teaching data analytics" OR "teaching data science" OR "data analysis education" OR "data analytics education" OR "data science education") AND ("tool" OR "application" OR "environment" OR "platform") (Title) or ("teaching data analysis" OR "teaching data analytics" OR "teaching data science" OR "data analysis education" OR "data analytics education" OR "data science education") AND ("tool" OR "application" OR "environment" OR "platform") (Abstract) or ("teaching data analysis" OR "teaching data analytics" OR "teaching data science" OR "data analysis education" OR "data analytics education" OR "data science education") AND ("tool" OR "application" OR "environment" OR "platform") (Author Keywords)
Wiley Online Library	("teaching data analysis" OR "teaching data analytics" OR "teaching data science" OR "data analysis education" OR "data analytics education" OR "data science education") AND ("tool" OR "application" OR "environment" OR "platform") anywhere

As buscas com as *strings* definitivas foram executadas pelos autores deste trabalho em dezembro de 2023, sendo que o resultado obtido totalizou 1115 artefatos, considerando todas as fontes: *ACM Digital Library* – 21; *ERIC* – 101; *IEEE Xplore* – 124; *Google Scholar* – 290; *ScienceDirect* – 19; *Scopus* – 162; *Springer Link* – 184; *Web of Science* – 45; *Wiley Online Library* – 169. Após esse processo, os trabalhos duplicados (83) foram eliminados. Os remanescentes (1032) foram, em seguida, analisados, de forma a encontrar informações que pudessem atender às questões de pesquisa. Além da leitura dos resumos, também foi feita uma busca, nos textos completos, por palavras que pudessem corresponder às informações desejadas. Essa etapa resultou em um total de 192 artefatos, que foram submetidos, então, a critérios de exclusão (Quadro 2).

Quadro 2 – Critérios de inclusão (CI) e exclusão (CE). Fonte: Os autores.

Critério	Descrição
CI-01	Trabalho menciona ferramenta para o estudo de conceitos de análise de dados.
CI-02	Incluído por referência de algum estudo (<i>snowballing</i>).
CE-01	Idioma diferente de português, inglês ou espanhol.
CE-02	Ferramenta mencionada não foi concebida para estudo.
CE-03	Ferramenta mencionada não é visual.
CE-04	Publicação contém somente o <i>abstract</i> ou uma única página.
CE-05	Trabalho não disponível para leitura completa.

Focando especificamente em ferramentas de contexto educacional, concebidas para o estudo de conceitos de análise de dados, foram excluídos trabalhos relacionados a ferramentas desenvolvidas para outros propósitos, como planilhas eletrônicas, linguagens de programação, plataformas para *business intelligence* ou para competição de análise de dados. Também foram excluídos trabalhos relativos a *frameworks* e que não apresentavam ferramentas concretas. Publicações que se referiam a ferramentas não computacionais, como jogos com cartões e atividades realizadas em papel e outros tipos de materiais, também foram excluídas.



A aplicação dos critérios de exclusão reduziu o conjunto a 36 publicações. Executado o processo de *snowballing*, foram incluídos sete trabalhos, e o conjunto final resultante totalizou 43 publicações para extração de dados.

3. RESULTADOS E DISCUSSÃO

Visando a responder as questões de pesquisa propostas no artigo, foram analisados os trabalhos identificados no mapeamento conduzido. No que tange à primeira questão de pesquisa – **Q1: Quais ferramentas visuais, de contexto educacional, existem para o estudo de conceitos de análise de dados?**, foram identificadas 35 ferramentas visuais desenvolvidas para o estudo de análise de dados. O Quadro 3 exibe os seguintes elementos de cada ferramenta: i) identificador ou nome; ii) publicação de referência; iii) tipo da ferramenta quanto ao ambiente de execução; iv) linguagens de *scripting* – código usado na ferramenta para completar as tarefas propostas; e v) tipo de *dataset* utilizado – fornecido em conjunto com a ferramenta, carregado pelo aprendiz, ou com suporte a ambos os tipos. Algumas publicações se referem a um conjunto de aplicações que interagem e se completam para atingir os objetivos de estudo (BOUKOUVALA *et al.*, 2018; DI IORIO; VANTINI, 2021), sendo computadas nessa pesquisa como apenas uma ferramenta.

Quadro 3 – Catálogo das ferramentas identificadas (continua). Fonte: Os autores.

Ferramenta	Referências	Tipo	Linguagens de Scripting	Tipo de Dataset
Blockly-DS	(BARBOZA <i>et al.</i> , 2023)	Aplicação web	Python, JavaScript	Ambos
CODAP – Common Online Data Analysis Platform	(FRISCHEMEIER <i>et al.</i> , 2021; RUBIN, 2021; ERICKSON; CHEN, 2021; HIGGINS <i>et al.</i> , 2023; NOLL <i>et al.</i> , 2023; ZAPATA-CARDONA, 2023)	Aplicação web	-----	Ambos
dAn-oNo – Data Analytics Onboarding Notebook	(STOIBER <i>et al.</i> , 2023)	Aplicação web	Python	Ambos
Data Desk Student Version	(WEISBERG; HENNESSY, 1991)	Aplicação desktop	-----	Ambos
Data Store / Data Visualization Sub-System	(IMAZAWA; NAOE, 2018)	Aplicação desktop	-----	Fornecido pela ferramenta
Data World	(HOYT; WANGIA-ANDERSON, 2018)	Aplicação web	R, Python, SQL, SPARQL	Ambos
DataCamp	(BAUMER <i>et al.</i> , 2020)	Aplicação web	R, Python, SQL	Não mencionado
DataLab	(ZHANG <i>et al.</i> , 2017)	Aplicação web	R, Python	Ambos
DS Lab Notebook	(IONASCU; STEFANIGA, 2020)	Aplicação web/mobile	Python	Não mencionado
DSLE – Data Science Lab Environment	(ATTANASIO <i>et al.</i> , 2020)	Aplicação web	Python	Ambos
Fathom	(BIEHLER, 2003)	Aplicação desktop	-----	Fornecido pela ferramenta
Shiny Application	(KONKOL'OVÁ; PARALIČ, 2018)	Aplicação web	-----	Fornecido pela ferramenta
Habitat Explorer	(SCHWARTZ <i>et al.</i> , 2016)	Aplicação web	-----	Fornecido pela ferramenta
Horses for Courses	(LEGAKI <i>et al.</i> , 2020)	Aplicação web	-----	Fornecido pela ferramenta
Hypothesis Testing Learning Tool	(DELGADO-GÓMEZ <i>et al.</i> , 2020)	Aplicação desktop	-----	Fornecido pela ferramenta
InSTEP	(LEE <i>et al.</i> , 2023)	Aplicação web	-----	Não mencionado
iNZight	(WILD <i>et al.</i> , 2021)	Aplicação desktop	-----	Carregado pelo estudante
ISLE – Integrated Statistics Learning Environment	(BURCKHARDT <i>et al.</i> , 2020; BURCKHARDT <i>et al.</i> , 2021)	Aplicação web	R, Markdown	Ambos
IVN (Interactive Visual Narratives)	(SAUNDAGE <i>et al.</i> , 2016)	Aplicação desktop	-----	Fornecido pela ferramenta



Simulator				
LAVADE – Latent Variable Demonstrator	(SCHAEFFER; BRAATZ, 2022)	Aplicação web	----	Fornecido pela ferramenta
Learn2Mine	(ANDERSON <i>et al.</i> , 2014; TURNER <i>et al.</i> , 2014)	Aplicação web	R, Python	Fornecido pela ferramenta
LiveDataLab	(GREEN; ZHAI, 2019)	Aplicação web	----	Carregado pelo estudante
MetaData	(NAIR <i>et al.</i> , 2020)	Aplicação desktop	----	Carregado pelo estudante
Milo	(RAO <i>et al.</i> , 2018)	Aplicação web	JavaScript, Python	Não mencionado
Módulo PCA – mobile application	(JUNG <i>et al.</i> , 2023)	Aplicação mobile	----	Fornecido pela ferramenta
Módulo 1 Correlações; Módulo 2 Regressão; Módulo 3 Otimização	(BOUKOUVALA <i>et al.</i> , 2018)	Aplicação web	----	Fornecido pela ferramenta
Playdata	(FERNANDEZ <i>et al.</i> , 2023)	Aplicação web	Scratch	Carregado pelo estudante
Playmeans	(KHACHATRYAN, 2023)	Aplicação web	----	Fornecido pela ferramenta
Point Blank 2D; Point blank revolution; Dimensions for descent; Deadly premonition	(DI IORIO; VANTINI, 2021)	Aplicação web	----	Ambos
Stat2Labs	(KUIPER; STURDIVANT, 2015)	Aplicação web	----	Fornecido pela ferramenta
statistics4school	(RIZOU <i>et al.</i> , 2021)	Aplicação web	----	Ambos
Stock Market	(LIU; WEI, 2020)	Aplicação web	R	Fornecido pela ferramenta
MYSTAT	(WEISBERG; HENNESSY, 1991)	Aplicação desktop	----	Fornecido pela ferramenta
TinkerPlots	(KOPARAN, 2019; DVIR; BEN-ZVI, 2022; DAHLSTROM-HAKKI; WALLACE, 2022)	Aplicação desktop	----	Ambos
Tuva	(ERICKSON, 2016)	Aplicação web	----	Fornecido pela ferramenta

Com relação às datas de publicação dos trabalhos, ressalta-se que é também um objetivo do estudo em tela visualizar uma linha do tempo (Figura 1), situando a época de uso e estudo das ferramentas. Com isso, é possível comparar recursos de diferentes épocas tecnológicas em aplicações que possuem os mesmos objetivos de estudo.

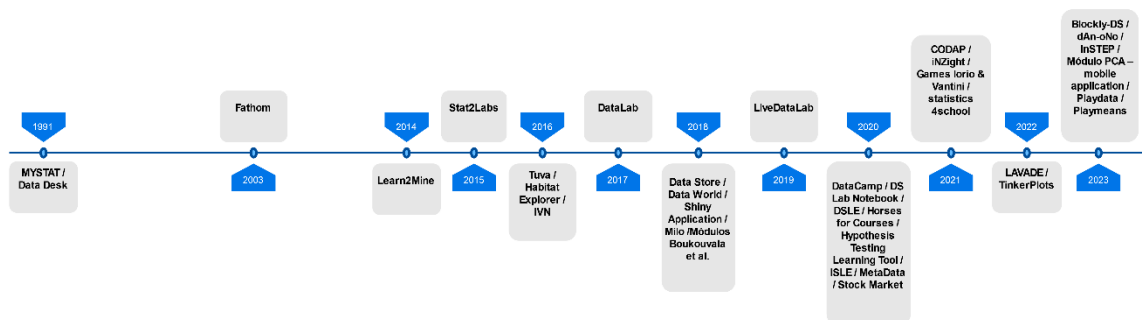


Figura 1 – Linha do tempo das ferramentas selecionadas. Fonte: Os autores.

A maioria das ferramentas encontradas foi desenvolvida e aplicada nos últimos cinco anos (21 ferramentas), sendo que elas incorporam recursos de interação recentes e têm capacidade de adaptação a vários tipos de arquiteturas, ou seja, são responsivas. Por outro lado, merecem destaque ferramentas como as citadas por Weisberg e Hennessy (1991), pois trata-se de aplicações criadas para estudo de análise de dados há mais de 30 anos. Refletindo o desenvolvimento tecnológico da época (início da década de 90), elas



fazem uso intenso dos recursos gráficos então disponíveis nas plataformas *Apple Macintosh* e *IBM PC*.

Quanto ao ambiente de execução, 25 ferramentas operam por meio de um navegador (*browser*), não requerendo qualquer tipo de instalação na máquina do aprendiz; e nove são aplicações *standalone (desktop)*, executando diretamente sob um sistema operacional. Foi identificada uma ferramenta (JUNG *et al.*, 2023) concebida especificamente para uso em dispositivo móvel, utilizando recursos de *hardware* e *software* do aparelho (*smartphone*) para coletar dados de interesse que serão analisados posteriormente pelo aprendiz. Cabe aqui a observação de que o ambiente de execução é uma variável importante a ser considerada na seleção de uma ferramenta de estudo, devido à grande variedade de plataformas possíveis de serem usadas pelos aprendizes.

As linguagens de *scripting* são utilizadas nas ferramentas para completar alguma tarefa ou estudo específico. Algumas ferramentas identificadas não fazem uso de qualquer tipo de codificação para completar as tarefas, sendo totalmente visuais no que se refere à operação. Dentre as aplicações que fazem uso de algum tipo de codificação, foi verificado que as linguagens mais utilizadas são Python (9) e R (6), sendo que algumas têm a capacidade de fazer uso de ambas. Essa evidência não se revela uma surpresa, sabendo-se que essas linguagens são largamente utilizadas para fins de análise de dados.

Foi verificado que a maior parte das ferramentas identificadas (16) opera com um ou mais *datasets* (conjuntos de dados) próprios, não permitindo que o aprendiz faça uso de outros tipos de dados de sua escolha. Essa condição pode, por um lado, permitir a redução da complexidade do ambiente de execução e da própria ferramenta, mas, por outro lado, pode limitar significativamente as opções de contextualização que o aprendiz teria por meio da livre seleção dos dados. Entretanto, uma quantidade proporcional de ferramentas (15) permite que o aprendiz ou o instrutor façam o carregamento de *datasets* de sua escolha, possibilitando uma customização mais ampla da experiência de uso e maior adaptação aos objetivos de estudo.

Em relação à segunda questão de pesquisa – **Q2: Quais conceitos de análise de dados são abordados por essas ferramentas?**, cabe salientar que a análise exploratória de dados tem como principal objetivo promover um reconhecimento inicial sobre os dados que serão estudados. Ao iniciar o estudo, por meio da observação de anomalias ou *outliers* (valores discrepantes) e da distribuição dos dados, pode-se começar a formulação de testes de hipóteses específicos (KOMOROWSKI *et al.*, 2016). Tipos de gráficos variados como histogramas, *scatter plots*, *boxplots*, barras, dentre outros, conforme elencados por Barboza *et al.* (2023), dão suporte à visualização e à análise exploratória de dados. Nas ferramentas identificadas nessa pesquisa, os tópicos de análise de dados estudados variam da simples visualização de dados até a abordagem de conceitos estatísticos avançados, como a *Principal Component Analysis* (PCA).

De acordo com Seltman (2018), qualquer método de análise de dados que não inclua modelagem estatística formal e inferência, se enquadra no termo análise exploratória de dados. Dessa forma, pode-se afirmar que todas as ferramentas relacionadas no mapeamento conduzido fornecem ao menos uma funcionalidade para visualização e análise exploratória de dados. Outros conceitos que podem ser estudados por meio dessas ferramentas são: análise descritiva e preditiva de dados (Blockly-DS, CODAP, LiveDataLab); correlações (dAn-oNo, Tuva, MYSTAT), PCA (MetaData, Data Desk, Módulo PCA), *K-means clustering* (Milo, Playmeans, Learn2Mine), estudo de tipos de dados (Data World, DataCamp, iNZight), dentre outros. Com relação ao estudo de conceitos específicos, destaca-se aqui o trabalho de Jung *et al.* (2023), no qual é descrita uma aplicação para dispositivo móvel com capacidade de captura de



movimentos corporais por meio do *hardware* do aparelho. Como declarado pelos autores, a ferramenta implementa, como único objeto de estudo, a técnica de redução de dimensionalidade PCA. Dessa forma, o aprendiz usa o grande volume de dados gerados para estudar em detalhes as várias etapas que compõem essa técnica, por meio das funcionalidades da aplicação.

No âmbito da terceira questão de pesquisa – **Q3: Para quais públicos ou contextos essas ferramentas são voltadas?**, grandes volumes de dados são gerados nos mais variados contextos de atividades. O gerenciamento adequado e a compreensão dos significados desses dados tornaram-se atualmente indispensáveis, porque a tomada de decisões em áreas como negócios, política, sociedade, dentre outras, está apoiada em evidências baseadas em dados (FRISCHEMEIER *et al.*, 2021). Isso posto, a demanda por educação em análise de dados pode ser notada em públicos bastante distintos e com interesses sem nenhum ponto em comum, a não ser a necessidade de analisar, de forma adequada e significativa, os dados concernentes à sua própria área de atuação.

Nos trabalhos mencionados nessa pesquisa notou-se a divisão dos públicos-alvo em duas grandes áreas: aprendizes STEM e aprendizes não-STEM. STEM, que é o acrônimo para *Science, Technology, Engineering and Mathematics*, pode ser visto como um grande conjunto de cursos e disciplinas da grande área das ciências exatas. Dessa forma, foram identificadas ferramentas desenvolvidas para atender demandas de estudo de áreas STEM, tais como engenharia, matemática e estatística, e de áreas não-STEM, como medicina, negócios, jornalismo, políticas sociais, dentre outras. Destaca-se aqui o estudo conduzido por Burckhardt *et al.* (2020), no qual sete residentes em medicina estudaram conceitos estatísticos em análise de dados por meio da ferramenta ISLE (*Integrated Statistics Learning Environment*), ao mesmo tempo testando a usabilidade da aplicação. Os aprendizes executaram exercícios de visualização e análise exploratória de dados, testes de hipóteses e modelagem preditiva sobre um *dataset* com dados de indicadores de problemas cardíacos e outros atributos físicos de 202 pacientes.

Outra segmentação que ficou evidenciada foi a por faixas etárias e níveis de ensino. A maior parte das ferramentas foi mencionada em estudos realizados com aprendizes dos níveis fundamental e médio, seguidos pelos níveis superior e pós-graduação. Como exemplo dessa segmentação, pode ser citada a ferramenta CODAP (*Common Online Data Analysis Platform*), que destaca na página principal do *site* que se trata de uma plataforma de estudo de análise de dados voltada para estudantes dos ensinos básico e secundário. Merece também destaque, devido ao contexto e tipo de público bastante específico, o estudo realizado por Dahlstrom-Hakki e Wallace (2022). Os autores utilizaram a ferramenta TinkerPlots para demonstrar elementos de visualização e exploração de dados a estudantes com deficiência de aprendizado, Transtorno de Déficit de Atenção e Hiperatividade (TDAH) e autismo.

Na quarta questão de pesquisa – **Q4: Dentre as ferramentas identificadas, quais são e de que forma elas empregam elementos de gamificação?**, conforme observado por Deterding *et al.* (2011), há elevado nível de subjetividade e contextualidade na identificação da gamificação, sendo necessário conhecer as intenções do *designer* e a experiência do usuário para se conseguir diferenciar uma aplicação gamificada de um jogo. Ainda de acordo com os autores, o projeto de uma ferramenta gamificada inclui elementos de *design* de jogos, mas sem o objetivo de construir um jogo completo. Considerando o exposto e os elementos de jogos elencados por Feichas e Seabra (2022), foram identificadas 13 ferramentas, a partir dos trabalhos mapeados nesta pesquisa, que podem ser caracterizadas como aplicações gamificadas.

O Quadro 4 relaciona os elementos encontrados com as respectivas ferramentas que os implementam, ficando evidente que o recurso de gamificação mais utilizado foi a



tabela de classificação ou *leaderboard*, sendo o único elemento de gamificação incorporado em várias ferramentas. Fora desse conjunto, onde as aplicações principais implementam elementos de gamificação, localiza-se a plataforma CODAP. Nesta ferramenta existe a possibilidade de adicionar jogos para o estudo de dados, na forma de *plugins*. Ainda, dentre as aplicações identificadas, destacam-se duas que, conforme Schwartz *et al.* (2016) e Di Iorio e Vantini (2021), evidenciam, nas publicações referentes, que foram concebidas para serem jogos educacionais. A intenção de compor um jogo completo pode ser percebida nos nomes dados às ferramentas *Habitat Explorer*, *Point Blank 2D*, *Point Blank Revolution*, *Dimensions for Descent* e *Deadly Premonition*.

Quadro 4 – Elementos de gamificação identificados. Fonte: Os autores.

Elemento	Ferramentas
Tabela de classificação	DataCamp; DataLab; DSLE; Horses for Courses; Hypothesis Testing Learning Tool; Learn2Mine; LiveDataLab
Pontos	Horses for Courses; Hypothesis Testing Learning Tool; Learn2Mine
Quiz	Shiny Application; ISLE
Níveis	Horses for Courses; Learn2Mine
Desafios	Horses for Courses; Stat2Labs
Simulações	Stat2Labs; Stock Market
Competição	Learn2Mine
Feedback constante	Learn2Mine
Badges (emblemas)	Learn2Mine
Narrativa	IVN (Interactive Visual Narratives)
Interação audiovisual e emoções	Playmeans

O elemento de gamificação mais utilizado – tabela de classificação – foi implementado em sete ferramentas. Em seguida, está o recurso de pontos, utilizado em três aplicações. *Quiz*, níveis, desafios e simulações foram implementados em duas ferramentas cada. A ferramenta identificada por esta pesquisa que apresenta a maior diversidade de recursos de gamificação é a *Learn2Mine*. Tal evidência encontra paralelo na premissa de seus autores (ANDERSON *et al.*, 2014), isto é, a chave para o ambiente *Learn2Mine* é a sua confiança na noção de gamificação, a aplicação de elementos típicos de jogos.

Finalmente, os autores apontam duas questões como ameaças à validade desta pesquisa: seleção das bases científicas pesquisadas e pesquisa manual dos trabalhos. No primeiro caso, apesar de terem sido selecionadas nove bases renomadas em pesquisas de estudos primários, outros veículos de publicações poderiam ser incluídos no estudo, o que possivelmente permitiria a identificação de mais trabalhos relevantes. No segundo caso, apesar de as buscas terem sido bem planejadas e executadas, algum trabalho relacionado ao tema pode não ter sido incluído devido a alguma falha humana. Como forma de mitigar essa última ameaça, os autores desta pesquisa revisaram os trabalhos identificados, visando resolver possíveis divergências. Além disso, a restrição de pesquisa relativa ao enfoque apenas em ferramentas desenvolvidas para contexto educacional constitui uma limitação deste estudo, reduzindo o número de trabalhos selecionados.

4. CONSIDERAÇÕES FINAIS

Este trabalho conduziu um mapeamento sistemático de pesquisas e ferramentas utilizadas para o estudo de análise de dados, tendo descoberto, analisado e classificado essas ferramentas, além de obter um panorama temporal de uso delas. O estudo foi iniciado com uma busca abrangente em bases científicas, que resultou em 1115



publicações. Então, após a aplicação de critérios de seleção e de exclusão, e ainda executado o processo de *snowballing*, obteve-se um conjunto final de 43 publicações. Dados foram extraídos desse conjunto, resultando em 35 ferramentas visuais identificadas para o estudo de análise de dados.

Em consonância com o objetivo principal da pesquisa e respondendo às questões específicas formuladas, as ferramentas identificadas foram classificadas quanto aos conceitos estudados, aos contextos em que foram utilizadas e aos elementos de gamificação nelas incorporados. Foi observado que ferramentas computacionais visuais para o estudo de conceitos de análise de dados vêm sendo utilizadas desde as últimas décadas do século XX, acompanhando o desenvolvimento dos recursos disponibilizados nos computadores pessoais, em especial no que tange às interfaces gráficas. Nos últimos 10 anos, como resposta ao aumento no volume de dados gerados globalmente, observou-se também aumento na quantidade e na variedade de ferramentas concebidas para essa finalidade, bem como na diversidade de públicos-alvo beneficiados por elas.

Com o objetivo de estimular a participação dos aprendizes e de melhorar os resultados em seus estudos, uma quantidade significativa dessas ferramentas têm incorporado conceitos e elementos empregados inicialmente em jogos em seus projetos, adaptados para funcionar como recurso pedagógico. Os resultados dessa pesquisa indicam que ferramentas para o estudo de conceitos de análise de dados podem ser construídas para aplicações nas mais diversas áreas, contextos de interesse e para uso dos mais variados tipos de públicos. Essa constatação, somada ao que se sabe sobre a demanda global atual para conhecimento relevante extraído a partir de grandes volumes de dados, leva à conclusão de que são virtualmente ilimitadas as possibilidades de desenvolvimento de novas ferramentas, bem como de investigação científica no que tange ao estudo de análise de dados.

Trabalhos futuros poderão considerar estudos relacionados com o uso de ferramentas desenvolvidas originalmente para outras finalidades, como linguagens de programação, *frameworks*, plataformas de *business intelligence*, planilhas eletrônicas, dentre outras. Dessa forma, poderão ser investigadas outras tecnologias e tipos de aplicações utilizadas no estudo de análise de dados, bem como outros contextos de dados e públicos interessados. Outra possibilidade dessa abrangência de estudos é a visualização de uma linha temporal mais longa de pesquisas na área, com a descoberta de trabalhos ainda mais distantes no passado e, possivelmente, outros que abordem temas com projeção no futuro, como Inteligência Artificial e *Machine Learning*.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDERSON, P. E. *et al.* An extensible online environment for teaching data science concepts through gamification. In: **2014 IEEE Frontiers in Education Conference (FIE) Proceedings**, p. 1-8, 2014.

ATTANASIO, G. *et al.* DSLE: A smart platform for designing data science competitions. In: **2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)**, p. 133-142, 2020.

BARBOZA, L. *et al.* Blockly-DS: Blocks programming for data science with visual, statistical, descriptive and predictive analysis. In: **13th International Learning Analytics and Knowledge Conference**, p. 644-649, 2023.



BAUMER, B. S. *et al.* Teaching introductory statistics with DataCamp. **Journal of Statistics Education**, v. 28, n. 1, p. 89-97, 2020.

BIEHLER, R. Interrelated learning and working environments for supporting the use of computer tools in introductory classes. In: **Proceedings of the IASE Satellite Conference Statistics Education and the Internet**, 2023.

BOUKOUVALA, F. *et al.* Teaching data-analytics through process design. **Computer Aided Chemical Engineering**, v. 44, p. 1663-1668, 2018.

BURCKHARDT, P. *et al.* Teaching statistical concepts and modern data analysis with a computing-integrated learning environment. **Journal of Statistics and Data Science Education**, v. 29, S61-S73, 2021.

BURCKHARDT, P. *et al.* Informing the design of e-learning lessons on data science through think-aloud experiments with healthcare professionals. In: **2020 IEEE International Conference on Healthcare Informatics (ICHI)**, p. 1-7, 2020.

DAHLSTROM-HAKKI, I.; WALLACE, M. L. Teaching statistics to struggling students: Lessons learned from students with LD, ADHD, and autism. **Journal of Statistics and Data Science Education**, v. 30, n. 2, p. 127-137, 2022.

DELGADO-GÓMEZ, D. *et al.* Improving the teaching of hypothesis testing using a divide-and-conquer strategy and content exposure control in a gamified environment. **Mathematics**, v. 8, n. 12, 2244, 2020.

DETERDING, S. *et al.* From game design elements to gamefulness: defining "gamification". In: **Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments**, p. 9-15, 2011.

DI IORIO, J.; VANTINI, S. How to get away with statistics: Gamification of multivariate statistics. **Journal of Statistics and Data Science Education**, v. 29, n. 3, p. 241-250, 2021.

DVIR, M.; BEN-ZVI, D. Students' actual purposes when engaging with a computerized simulation in the context of citizen science. **British Journal of Educational Technology**, v. 53, n. 5, p. 1202-1220, 2022.

ERICKSON, T., CHEN, E. Introducing data science with data moves and CODAP. **Teaching Statistics**, v. 43, S124-S132, 2021.

ERICKSON, T. Practical public online data: Introducing TUVa and CODAP. In: **Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE)**, 2016.

FEICHAS, F. A.; SEABRA, R. D. PGE-UML: Uma plataforma *web* gamificada para o estudo da Unified Modeling Language. **RENOTE**, v. 20, n. 2, p. 249-258, 2022.

FERNANDEZ, C. *et al.* (2023). Programming representations: Uncovering the process of constructing data visualizations in a block-based programming environment. In:



Proceedings of the 2023 Symposium on Learning, Design and Technology, p. 11-20, 2023.

FRISCHEMEIER, D. *et al.* A first introduction to data science education in secondary schools: Teaching and learning about data exploration with CODAP using survey data. **Teaching Statistics**, v. 43, S182-S189, 2021.

GREEN, A.; ZHAI, C. Livedatalab: A cloud-based platform to facilitate hands-on data science education at scale. In: **Proceedings of the Sixth (2019) ACM Conference on Learning@ Scale**, p. 1-2, 2019.

HAKULINEN, L.; AUVINEN, T. The effect of gamification on students with different achievement goal orientations. In: **2014 International Conference on Teaching and Learning in Computing and Engineering**, p. 9-16, 2014.

HICKS, S. C.; IRIZARRY, R. A. A guide to teaching data science. **The American Statistician**, v. 72, n. 4, p. 382-391, 2018.

HIGGINS, T. *et al.* Students' approaches to exploring relationships between categorical variables. **Teaching Statistics**, v. 45, S52-S66, 2023.

HOYT, R.; WANGIA-ANDERSON, V. An overview of two open interactive computing environments useful for data science education. **JAMIA Open**, v. 1, n. 2, p. 159-165, 2018.

IMAZAWA, A.; NAOE, N. View of a study to foster students' interest in data analysis using air temperature data. In: **2018 IEEE 10th International Conference on Engineering Education (ICEED)**, p. 148-153, 2018.

IONASCU, A.; STEFANIGA, S. A. DS Lab Notebook: A new tool for data science applications. In: **22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)**, p. 310-314, 2020.

JUNG, S. *et al.* Learning data science methods through a mobile device and full body motion data. In: **Proceedings of the Human Factors and Ergonomics Society Annual Meeting**, v. 67, n. 1, p. 277-282, 2023.

KHACHATRYAN, D. Playmeans: Inclusive and engaging data science through music. **Journal of Statistics and Data Science Education**, v. 31, n. 2, p. 151-161, 2023.

KOIVISTO, J.; HAMARI, J. The rise of motivational information systems: A review of gamification research. **International Journal of Information Management**, v. 45, p. 191-210, 2019.

KOMOROWSKI, M. *et al.* Exploratory data analysis. In: **Secondary Analysis of Electronic Health Records**. MIT Critical Data. Springer Open, p. 185-203, 2016.

KONKOL'OVÁ, V.; PARALIČ, J. Active learning in data science education. In: **16th International Conference on Emerging eLearning Technologies and Applications (ICETA)**, p. 285-290, 2018.



KOPARAN, T. Examination of the dynamic software-supported learning environment in data analysis. **International Journal of Mathematical Education in Science and Technology**, v, 50, n. 2, p. 277-291, 2019.

KUIPER, S.; STURDIVANT, R. X. Using online game-based simulations to strengthen students' understanding of practical statistical issues in real-world data analysis. **The American Statistician**, v. 69, n. 4, p. 354-361, 2015.

LEE, H. S. *et al.* Online teacher professional learning: An approach to foster personalized pathways. **Research report presented at the Annual Research Conference of the National Council of Teachers of Mathematics**, October, Washington, DC, 2023.

LEGAKI, N. Z. *et al.* The effect of challenge-based gamification on learning: An experiment in the context of statistics education. **International Journal of Human-Computer Studies**, v. 144, 102496, 2020.

LIU, Y.; WEI, X. How to use stock data for data science education: a simulated trading platform in classroom. In: **IEEE 2nd International Conference on Computer Science and Educational Informatization (CSEI)**, p. 5-8, 2020.

MAJURI, J. *et al.* Gamification of education and learning: A review of empirical literature. In: **Proceedings of the 2nd International GamiFIN Conference**, p. 11-19, 2018.

NAIR, R. *et al.* Metadata: A tool to supplement data science education for the first year undergraduates. In: **Proceedings of the 8th International Conference on Information and Education Technology**, p. 153-160, 2020.

NOLL, J. *et al.* Introduction to rethinking learners' reasoning with nontraditional data. **Teaching Statistics**, v. 45, p. S1-S4, 2023.

OLIVEIRA, W. *et al.* Tailored gamification in education: A literature review and future agenda. **Education and Information Technologies**, v. 28, n. 1, p. 373-406, 2023.

PETERSEN, K. *et al.* Guidelines for conducting systematic mapping studies in software engineering: an update. **Information and Software Technology**, v. 64, p. 1-18, 2015.

RAO, A. *et al.* Milo: A visual programming environment for data science education. In: **2018 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)**, p. 211-215, 2018.

RIZOU, O. *et al.* Supporting statistical literacy with ICT-based teaching scenario. **International Journal of Education**, v. 9, p. 59-78, 2021.

RUBIN, A. What to consider when we consider data. **Teaching Statistics**, v. 43, p. S23-S33, 2021.



SAUNDAGE, D. *et al.* Teaching data analysis with interactive visual narratives. **Journal of Information Systems Education**, v. 27, n. 4, p. 233-248, 2016.

SCHAEFFER, J.; BRAATZ, R. D. Latent variable method demonstrator – software for understanding multivariate data analytics algorithms. **Computers & Chemical Engineering**, v. 167, p. 108014, 2022.

SCHWARTZ, A. *et al.* Habitat Explorer: Designing educational games for collaborative learning on interactive surfaces. In: **Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces**, p. 397-401, 2016.

SELTMAN, H. J. **Experimental Design and Analysis**, 2018. Disponível em: <<https://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf>>. Acesso em: 28 de mar. 2024.

STOIBER, C. *et al.* dAn-oNo: Learning environment for data journalists teaching data analytics principles. In: **2023 IEEE VIS Workshop on Visualization Education, Literacy, and Activities (EduVis)**, p. 41-48, 2023.

TURNER, C. A. *et al.* Learn2Mine: Data science practice and education through gameful experiences. **International Journal of e-Education, e-Business, e-Management and e-Learning**, v. 4, n. 3, p. 243, 2014.

WEISBERG, H. F.; HENNESSY, L. V. For the classroom: Teaching data analysis in an interactive graphics environment. **PS: Political Science & Politics**, v. 24, n. 3, p. 505–510, 1991.

WILD, C. J. *et al.* On democratizing data science: some iNZights into empowering the many. **Harvard Data Science Review**, v. 3.2, 2021.

ZAPATA-CARDONA, L. The possibilities of exploring nontraditional datasets with young children. **Teaching Statistics**, v. 45, S22-S29, 2023.

ZHANG, Y. *et al.* DataLab: Introducing software engineering thinking into data science education at scale. In: **2017 IEEE/ACM 39th International Conference on Software Engineering: Software Engineering Education and Training Track (ICSE-SEET)**, p. 47-56, 2017.