

Explorando Dados: Ferramentas e Métodos para Apoiar a Aprendizagem na Construção de Visualizações de Dados

Cassia de Oliveira Fernandez¹,

Roseli de Deus Lopes¹ (orientadora), Paulo Blikstein² (coorientador)

¹Escola Politécnica – Universidade de São Paulo (USP)

²Teachers College, Columbia University – New York, NY – EUA

`cassia.fernandez@usp.br, roseli.lopes@poli.usp.br, pb@tc.columbia.edu`

Resumo. *Este trabalho apresenta contribuições na interseção entre Ciência da Computação, Educação e Visualização de Dados, propondo: (1) um framework para o design de ferramentas educacionais, (2) um ambiente de programação em blocos para construção de visualizações (PlayData), e (3) um método automatizado para caracterização de trajetórias de aprendizagem usando técnicas de aprendizado de máquina e Epistemic Network Analysis.*

Abstract. *This work presents contributions at the intersection of Computer Science, Education, and Data Visualization, proposing: (1) a framework for educational tool design, (2) a block-based programming environment for constructing visualizations (PlayData), and (3) an automated method for characterizing learning trajectories using machine learning techniques and Epistemic Network Analysis.*

1. Introdução

A crescente disponibilidade de grandes volumes de dados impõe à Educação Básica o desafio de preparar estudantes para compreender e comunicar informações quantitativas criticamente. O letramento em visualização de dados (*visualization literacy*) é central nessa formação [Bach et al., 2023; Börner et al., 2019]. Ferramentas educacionais existentes neste campo tipicamente oferecem apenas *templates* predefinidos que, embora acelerem a criação, ocultam o mapeamento entre dados e representações visuais e limitam a criatividade [Rubin, 2020]. Ambientes de programação permitem maior expressividade e o desenvolvimento de competências meta-representacionais [diSessa e Sherin, 2000], mas exigem conhecimentos técnicos elevados para um uso educacional efetivo.

Diante desse cenário, a tese avançou no *design* de ambientes de aprendizagem para aprimorar o letramento em visualização de dados na Educação Básica, investigando: (QP1) Como a literatura de Visualização da Informação pode orientar o design de novas ferramentas educacionais; (QP2) Quais são as principais ações dos aprendizes ao construir visualizações; (QP3) Que padrões emergem nas trajetórias de programação; e (QP4) Quais são os desafios e potencialidades ao construir visualizações de dados utilizando a ferramenta de programação em blocos PlayData.

2. Contribuições

2.1 Framework para Ferramentas Educacionais de Visualização de Dados

Desenvolvemos um *framework* de análise e design para ferramentas educacionais de visualização de dados estruturado em quatro dimensões [Fernandez et al., 2025b]: (i) objetivos educacionais suportados: exploração ou comunicação; (ii) expressividade: amplitude de representações que a ferramenta permite criar; (iii) abstração: nível de ocultação dos detalhes computacionais; e (iv) transparência para mapeamentos de dados: grau em que a relação entre dado e elemento visual é explícita para o aprendiz. Para sua construção, foram revisadas taxonomias da Visualização da Informação e analisadas sete ferramentas educacionais similares existentes. Sua aplicação revelou que a maioria das ferramentas se concentra em baixa expressividade e alta abstração, evidenciando um espaço de design pouco explorado que motivou o desenvolvimento do PlayData.

2.2 PlayData: Ambiente de Programação em Blocos

O PlayData é um ambiente open-source de programação em blocos, baseado no Scratch, desenvolvido para a construção de visualizações de dados por estudantes da Educação Básica. Seu desenvolvimento adotou a metodologia de Design-Based Research (DBR) [Barab e Squire, 2004], com cinco ciclos iterativos de design, implementação, análise e redesign realizados entre 2021 e 2023. O PlayData incorpora blocos especializados para iteração sobre listas de dados, mapeamento de atributos a propriedades visuais (posição, cor, tamanho, transparência, entre outros), transformações de dados (média, ordenação, filtragem) e importação de *datasets* em tempo real via Google Sheets. O sistema registra logs de cada interação com granularidade de evento, viabilizando análises posteriores de trajetórias. A ferramenta está disponível publicamente em playdatalab.github.io, traduzida para português, inglês, espanhol e japonês.

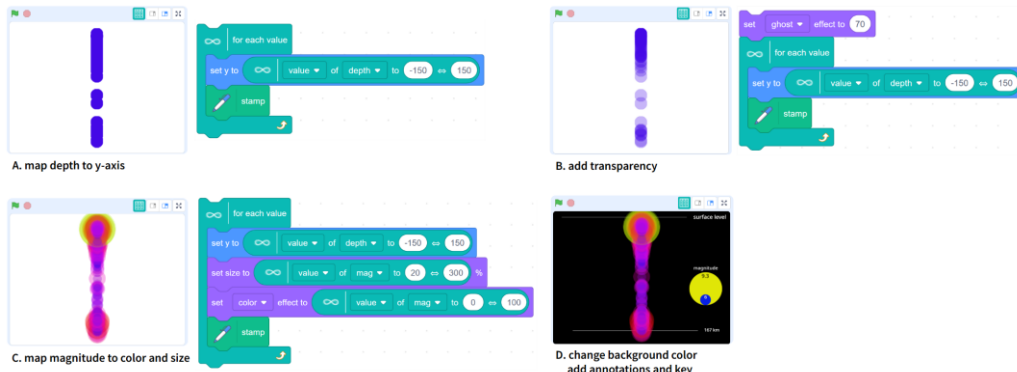


Figura 4. Construção de uma visualização de dados sísmicos com PlayData: (A) mapeamento de profundidade ao eixo Y, (B) transparência, (C) cor e tamanho, (D) anotações e background.

2.3 Método Automatizado para Caracterização de Trajetórias de Programação

Por fim, desenvolvemos uma metodologia de *learning analytics* que combina análise qualitativa de vídeo, *clustering* e *Ordered Network Analysis* (ONA) [Shaffer et al., 2016] para caracterizar automaticamente trajetórias a partir de logs coletados no ambiente [Fernandez et al., 2025a]. O desenvolvimento seguiu quatro etapas: (1) microanálise qualitativa de vídeos de captura de tela para identificação indutiva das ações centrais; (2) automação via *k-means* sobre vetores de contagem de eventos segmentados em janelas de 1 minuto, atingindo índice de concordância de 0,77 com a codificação humana manual; (3) aplicação de ONA para mapear conexões cronológicas entre eventos dentro de cada ação; e (4) análise agregada e temporal das trajetórias entre sessões. A abordagem é generalizável a qualquer ambiente de programação com logging granular.

2.4 DVP Moves: Ações Centrais na Construção de Visualizações

Por meio da análise combinada de vídeo e logs identificamos seis ações centrais, denominadas *Data Visualization Programming (DVP) moves*, cada uma com assinatura distinta de eventos. Estas ações foram identificadas indutivamente a partir da microanálise de vídeo e posteriormente validados e automatizados via *k-means*: (1) *Program Creation*: elaboração do código inicial; predominante no início das sessões, (2) *Program Rearrangement*: reorganização e depuração do código, (3) *Data Inspection*: análise dos dados na tabela de origem, (4) *Output Inspection*: observação do resultado visual gerado, (5) *Selection of Parameters*: ajuste iterativo de parâmetros; dominante na fase intermediária, (6) *Visual Design*: inclusão de títulos, legendas e anotações; dominante na fase final. Alguns dos resultados obtidos são apresentados na Figura 2.

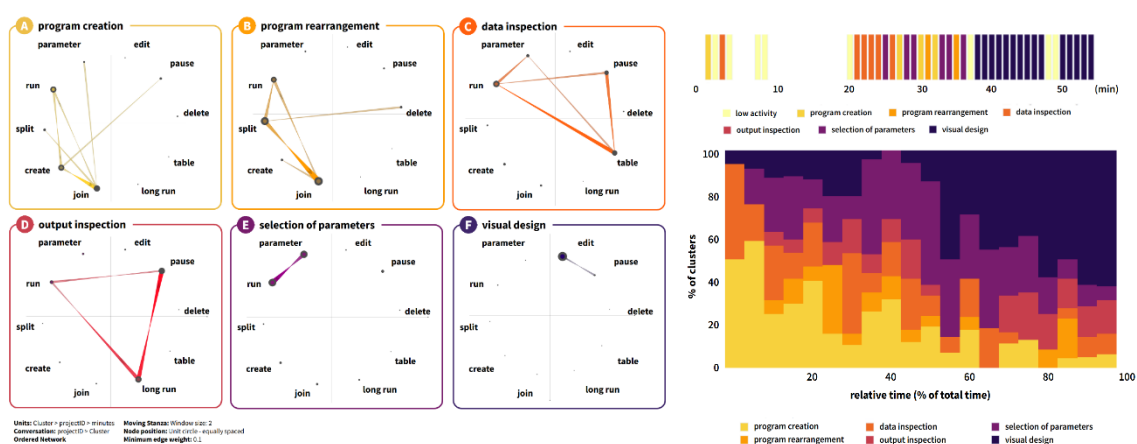


Figura 2. Esquerda: ONAs para cada ação (DVP move) identificada. Direita: trajetória de um projeto por meio das ações realizadas ao longo do tempo (acima) e visualização agregada para todos os projetos analisados (abaixo).

3. Conclusão e Perspectivas Futuras

A tese abre caminhos concretos para a integração da programação voltada à visualização de dados em currículos de Computação e Ciência de Dados na Educação Básica. O PlayData pode apoiar a criação de visualizações originais e complexas, com múltiplos atributos visuais mapeados simultaneamente, com alta expressividade e abstração intermediária. A análise temporal das trajetórias evidenciou que as sessões se iniciam com inspeção de dados e criação de código; a fase intermediária é dominada pela seleção iterativa de parâmetros; e a fase final converge para o design visual. Ao longo de um workshop de três dias, o tempo dedicado à depuração reduziu de 21% para 3%, enquanto design visual e inspeção de output aumentaram, sugerindo crescente autonomia e foco comunicativo. Do ponto de vista técnico, a metodologia de caracterização automatizada de trajetórias via aprendizado de máquina e redes epistêmicas representa uma contribuição inovadora para o campo de *Educational Data Mining* e *Learning Analytics*, generalizável a qualquer ambiente de programação com logging granular. Os DVP moves fornecem vocabulário teórico para análise de tarefas abertas de programação.

As perspectivas futuras incluem: (1) estudos em maior escala para validação dos padrões identificados; (2) desenvolvimento de *dashboards* para professores baseados nos DVP moves; (3) exploração de algoritmos de segmentação dinâmica de trajetórias em substituição às janelas fixas de tempo; e (4) extensão do PlayData para acessibilidade e colaboração em tempo real.

Referências

- Bach, B. et al. (2023) “Challenges and opportunities in data visualization education”, *IEEE Transactions on Visualization and Computer Graphics*, v. 29, n. 1, p. 700–710.
- Barab, S. e Squire, K. (2004) “Design-based research: Putting a stake in the ground”, *Journal of the Learning Sciences*, v. 13, n. 1, p. 1–14.
- Börner, K. et al. (2019) “A multi-level typology of abstract visualization tasks”, *IEEE TVCG*, v. 25, n. 1, p. 1197–1202.
- diSessa, A. A. e Sherin, B. L. (2000) “Meta-representation: An introduction”, *Journal of Mathematical Behavior*, v. 19, n. 4, p. 385–398.
- Fernandez, C., Blikstein, P., & de Deus Lopes, R. (2025a). A Multi-Method Approach for Exploring Programming Trajectories Through Log Data: Insights from Data Visualization Tasks. *Journal of Science Education and Technology*, 34(5), 994-1019.
- Fernandez, C., Freitas, J. A., Blikstein, P., & de Deus Lopes, R. (2025b). The design space of visualization tools for data science education: Literature review and framework for future designs. *International Journal of Child-Computer Interaction*, 43, 100698.
- Rubin, A. (2020) “Learning to reason from samples”, *Educational Studies in Mathematics*, v. 103, n. 2, p. 201–222.
- Shaffer, D. W. et al. (2016) “Epistemic network analysis: A worked example of theory-based learning analytics”. In: *Handbook of Learning Analytics*. SoLAR.