



Frameworks de Big Data

UNIDADE 07

A arte de transformar dados em histórias: introdução à visualização de dados e sua importância em projetos de big data

Bem-vindo à semana dedicada à visualização de dados em big data! Nesta jornada, mergulharemos na arte e na ciência por trás da interpretação visual de conjuntos massivos de dados. Desde a importância da visualização na compreensão de insights até as técnicas mais avançadas para representar informações complexas de maneira acessível, esta semana oferecerá uma exploração abrangente do papel crucial que a visualização desempenha na

análise de big data. Prepare-se para descobrir como transformar montanhas de dados em visualizações claras e impactantes, que impulsionam as tomadas de decisões informadas. Vamos começar!

Em um mundo inundado por dados, a capacidade de transformá-los em histórias impactantes é para tomar decisões assertivas. Aqui, surge a visualização de dados, como ferramenta essencial para essa jornada, cuja origem já existe há algum tempo, tendo suas raízes no desenvolvimento de mapas e diagramas no século XVII e na invenção do gráfico de pizza em meados do século XIX.

Para você ter uma ideia da importância desse assunto, foi nesse contexto que o engenheiro civil francês Charles Minard eternizou seu nome, criando um dos exemplos mais célebres de visualização de dados: *o mapa da invasão de Napoleão à Rússia*. Combinando informações sobre o tamanho do exército e a rota de retirada de Moscou, Minard não se limitou a números e gráficos. Ele incorporou a temperatura e a escala temporal ao seu mapa, tecendo uma narrativa visual rica em detalhes e *insights*.

A obra de Minard, representada na Figura 1, vai além de um mero mapa. Ela narra uma história de ambição, sofrimento e resiliência. As linhas grossas que representam o exército francês diminuem gradativamente à medida que se afastam de Moscou, simbolizando o declínio das tropas diante das intempéries e da resistência russa. A temperatura, indicada por faixas coloridas ao longo do mapa, revela as condições extremas enfrentadas pelos soldados, enquanto a escala temporal, na parte inferior, contextualiza a narrativa.

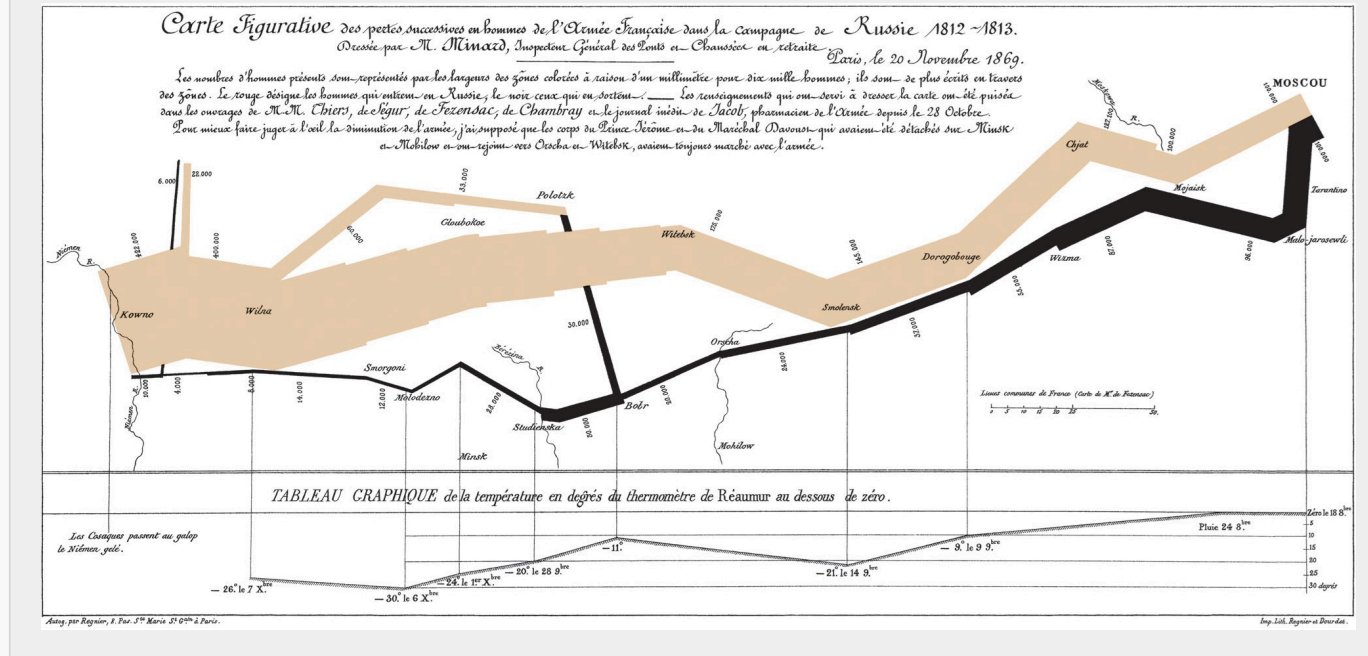


Figura 1: Mapa de Minard. Fonte: Cunha Lima (2015).

O mapa de Minard transcende a mera visualização de dados, tornando-se um relato pungente da história. Ele nos convida a navegar por um passado distante, desvendando os eventos através da lente da informação visual. Sua obra serve como um farol para aqueles que buscam transformar dados em histórias, utilizando-os para iluminar o caminho para decisões mais conscientes e impactantes.

Nesta aula, aprenderemos como podemos representar os dados que estão alocados nos grandes bancos de dados. Como podemos representar a imensa quantidade de dados e transformar em informação útil que possa gerar conhecimentos ou tomadas de decisões do ponto de vista dos dados?

Você está pronto. Vamos nessa!

Antes de entrarmos em conceitos e definições sobre como podemos visualizar os dados, observe o cenário a seguir.



EXEMPLO

O curso de Análise de Big Data possui duas turmas: a Turma A e a Turma B. A Turma A optou por realizar os exames das provas em uma

escala de 0 a 100 pontos, enquanto a Turma B decidiu adotar uma escala de 0 a 137 pontos. A maior nota da Turma A foi de 72 pontos, enquanto a da Turma B foi de 96 pontos. Diante desse contexto, qual turma obteve um desempenho superior?

É uma escolha difícil, não é mesmo? No entanto, alguns estudantes já devem ter percebido que uma das maneiras de comparar é dividir o valor da nota pelo total de pontos: 72/100 e 96/137. Assim, é possível concluir que a Turma A teve um desempenho superior em relação à Turma B!

Você concorda que a apresentação do resultado de desempenho é uma informação baseada em dados? Agora, imagine se a mesma informação fosse apresentada conforme a Figura 2.

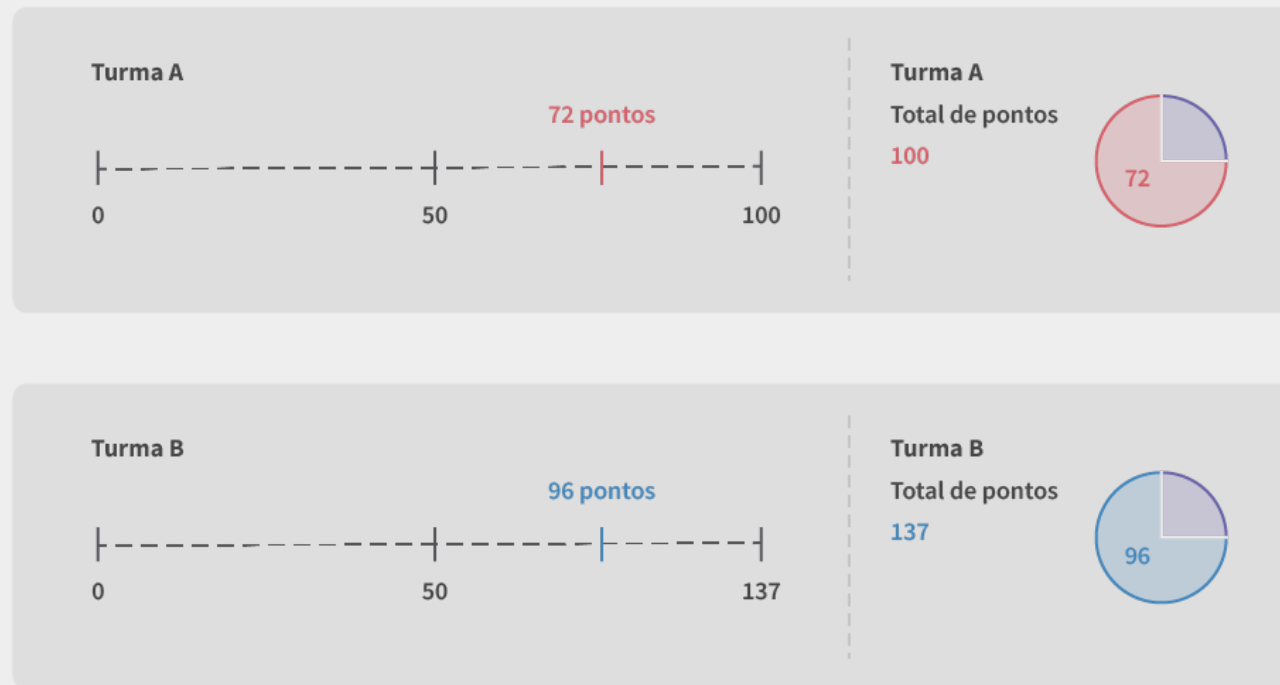


Figura 2: Relação de notas dos estudantes da Turma A e B. Fonte: O autor (2024).

Agora fica mais fácil de entender, certo? E por quê? Talvez, porque, agora, estamos visualizando os dados. Você se lembra daquela frase “uma imagem fala mais do que mil palavras”? Na visualização de dados, isso conta muito.

Voltando ao contexto da aula, dentro de projetos de *big data*, na maioria das vezes, o resultado será uma apresentação de todo o trabalho realizado, e isso inclui desde a captura, o tratamento e até a ingestão dos dados. Por fim, pode ser que você, estudante, seja chamado pela diretoria da empresa para apresentar um resultado de trabalho. Como você faria essa apresentação? Os dados precisam falar por si próprios e é aqui que entram as técnicas de visualização de dados!

A visualização de dados trata do processo de representar visualmente conjuntos de dados massivos e complexos, com o objetivo de extrair *insights* e facilitar a compreensão.

O que é visualização de dados?

Você já deve ter ouvido que, com a explosão de dados em nossa era digital, deparamo-nos com desafios significativos que demandam habilidades específicas para interpretar e utilizar as informações de forma eficaz.

É aqui que surge uma ferramenta importante para fornecer formas de como comunicar informações por meio de representações visuais: a visualização de dados! Ela pode, ainda, envolver diversos tipos de dados, como números e texto, apresentados de diferentes formas, como gráficos de barras ou mapas.

Na Figura 3, apresentamos o processo de ingestão de dados e de visualização de ponta a ponta. A grande maioria dos desafios em visualização de dados seguirá o fluxo apresentado a seguir.

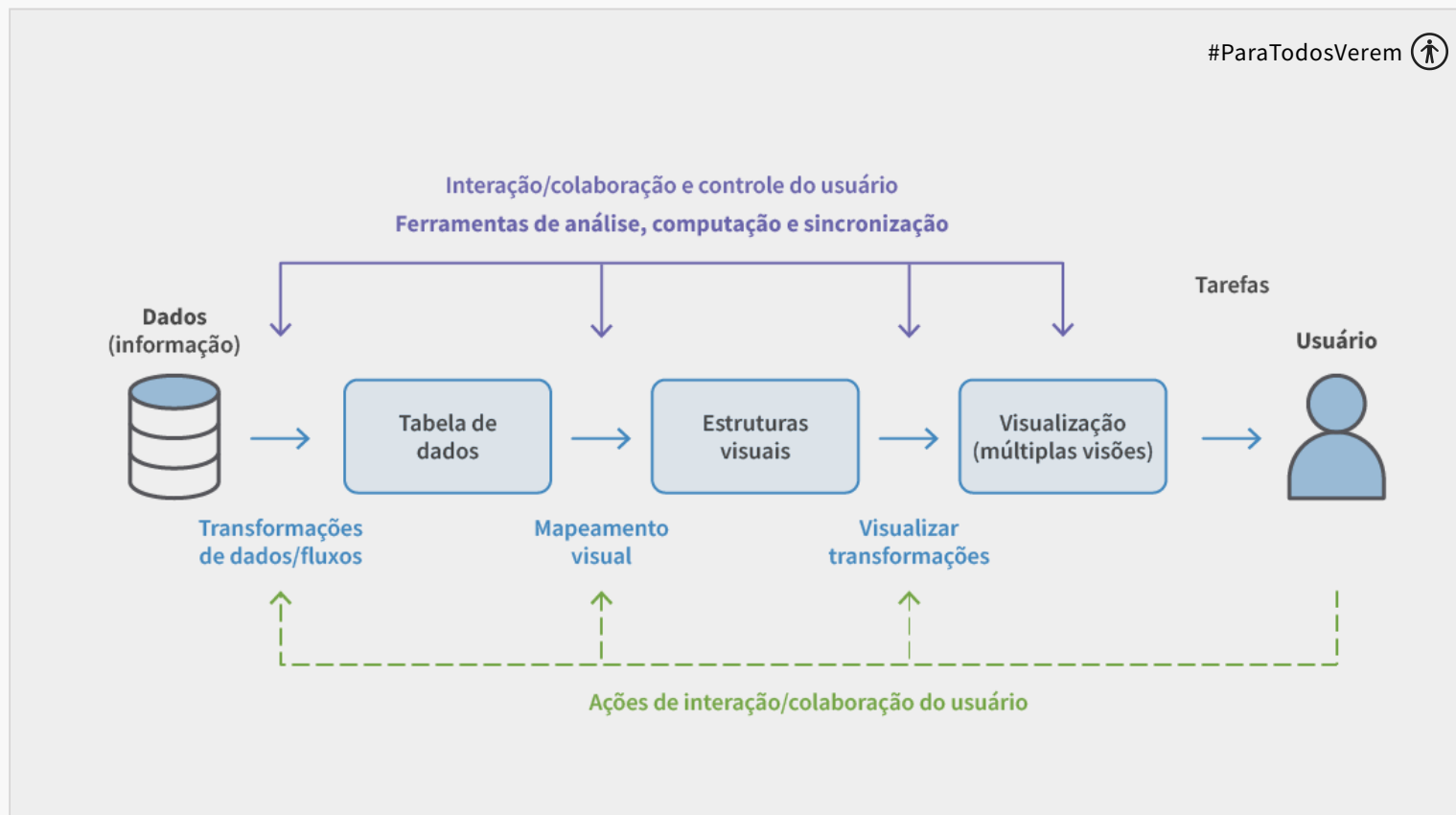


Figura 3: Processo de ingestão de dados e de visualização de ponta a ponta. Fonte: Adaptado de Pereira (2020).

É importante que haja uma grande quantidade de dados sobre o clima ou sobre o mercado financeiro. No entanto, muito mais importante é saber como você entenderá e tirará conclusões rapidamente apenas olhando para uma tabela cheia de números. É um pouco difícil de fazer correto?

A visualização de dados permite transformar esses dados em imagens nas quais nosso cérebro pode processar mais facilmente e, assim, tomar decisões de forma mais rápida e eficiente. Mas por que isso é tão importante? No Quadro 1, apresentamos uma lista de benefícios ao utilizar os recursos de visualização de dados e como as empresas relatam os resultados obtidos.

Quadro 1: Benefícios do uso de recursos de visualização de dados

Benefício	Descrição	Empresas que relatam benefícios significativos
Melhor tomada de decisão	Permite que os usuários acessem e analisem dados de forma rápida e fácil, para que possam tomar decisões mais informadas e baseadas em dados.	77%
Melhor análise <i>ad hoc</i> de dados	Fornece ferramentas para que os usuários explorem dados por conta própria.	43%
Aprimoramento da colaboração/compartilhamento de informações	Facilita o compartilhamento de informações e <i>insights</i> entre diferentes departamentos e equipes.	41%
Fornecimento de recursos de autoatendimento aos usuários finais	Permite que os usuários acessem os dados e as análises de que precisam sem a necessidade de solicitar ajuda à equipe de TI.	36%
Maior retorno do investimento (ROI)	Pode levar a uma melhor tomada de decisão, o que pode resultar em maior eficiência, produtividade e lucratividade.	34%
Economia de tempo	Reduz o tempo que os usuários gastam procurando e analisando dados, liberando-os para se concentrarem em tarefas mais importantes.	20%

Fonte: adaptado de Pereira (2020).

É importante compreender que a visualização de dados ajuda a apresentar os resultados de análises de forma clara e compreensível. Além disso, ela pode ser extremamente útil quando não temos questões específicas em mente ao iniciar a análise dos dados.

Às vezes, simplesmente não sabemos por onde começar ou quais perguntas fazer. Nessas situações, a visualização pode nos guiar e, até mesmo, revelar padrões ou problemas nos próprios dados, ajudando-nos a formular novas perguntas e hipóteses.

Dessa forma, podemos entender a **visualização de dados como um processo**, ou seja, uma sequência de etapas, em que cada etapa tem a missão de entregar um valor perceptível, a depender do tipo de usuário. No entanto, a ação de visualização de dados, na maioria das vezes, só fará sentido para os usuários finais.

De acordo com Milani (2020), **o processo de visualização de dados é composto por etapas**. Em cada etapa há ações que, quando somadas, resultam em dados disponibilizados na forma visual. Esse processo envolve quatro etapas:

+ Fase 1

Inicialmente, ocorre a coleta e o armazenamento dos dados.

+ Fase 2

Em seguida, é realizado um pré-processamento para organizar logicamente os dados.

+ Fase 3

A etapa seguinte aborda o *hardware* e os algoritmos gráficos que geram visualizações.

+ Fase 4

Por fim, há o envolvimento da percepção humana e do sistema cognitivo, facilitando a compreensão dos dados e, conseqüentemente, aprimorando a tomada de decisões e a formulação de hipóteses com base neles.

Por fim, uma das grandes vantagens da visualização é que ela permite que nosso cérebro processe informações de forma paralela, ou seja, podemos entender várias coisas ao mesmo tempo apenas olhando para uma imagem, enquanto a leitura de texto é um processo sequencial, mais lento. Além disso, as visualizações podem ser compreendidas por pessoas de diferentes idiomas, uma vez que imagens, muitas vezes, transcendem barreiras linguísticas.

| Importância da visualização de dados em big data

No contexto do uso de soluções de *big data*, a visualização de dados assume um papel importante. Considerando os muitos dados disponíveis em grandes bancos de dados, a missão, agora, é transformar esses dados em informações significativas e de fácil compreensão. Observe a Figura 4 e veja que é mais fácil entender o que os dados querem informar do que os ver representados em várias tabelas de banco de dados.

Essa é a vantagem de utilizar as técnicas de visualização de dados, em especial no contexto de projetos de *big data*, ao apresentar o resultado da exploração, da ingestão e do tratamento dos dados e, em seguida, as possíveis relações entre esses dados. Por isso, podemos utilizar os recursos da visualização de dados para comunicar os resultados obtidos.

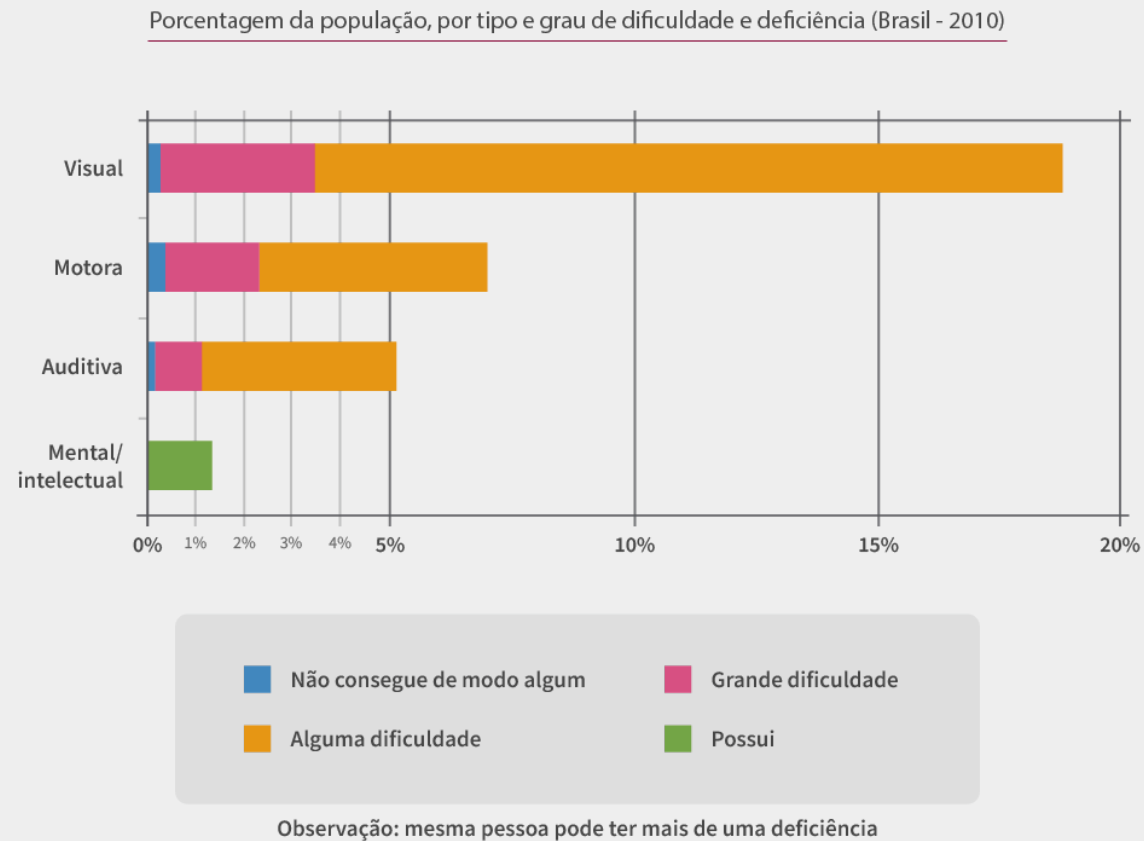


Figura 4: Exemplo de visualização de dados. Fonte: Adaptado de Milani (2020).

Os recursos a serem utilizados, ao representar os dados, podem ser nos formatos de gráficos ou mapas. A ideia é tornar padrões complexos e tendências de modo mais acessíveis, facilitando a interpretação até mesmo para os não especialistas.

Vamos adiante com mais um exemplo: vocês se lembram do jogo Pokemon Go?



EXEMPLO

Clareando um pouco a mente, o Pokémon Go foi um jogo de realidade aumentada para dispositivos móveis lançado em 2016. Desenvolvido pela Niantic, em colaboração com a Pokémon Company e a Nintendo, o jogo se tornou um fenômeno mundial instantâneo, uma vez que ao utilizar a tecnologia de realidade aumentada, permitia aos jogadores a captura de Pokémons no mundo real, apenas usando seus *smartphones*.

A mecânica do jogo era simples: os jogadores exploravam suas cidades e seus arredores em busca de Pokémons, que apareciam em locais específicos do mundo real, dependendo do tempo, da hora, do dia e de outros fatores. Os jogadores podiam capturar, treinar e batalhar com seus Pokémons, havendo, também, locais como "ginásios", onde os jogadores podiam competir entre si.

Pokémon Go foi amplamente elogiado por sua capacidade de motivar as pessoas a saírem de suas casas, explorarem seu ambiente e interagirem com outros jogadores. No entanto, também enfrentou críticas e preocupações com questões de segurança e privacidade, bem como com o uso excessivo do telefone celular.

Como o contexto de visualização de dados e de projetos de *big data* e o jogo Pokémon Go estão relacionados? Pense em como seria contar o número de pokémons capturados ou na quantidade de pokémons contabilizados em uma tabela de banco de dados.

Com base no cenário apresentado, a visualização de dados pode ajudar os amantes desse jogo de forma muito peculiar. Neste site, você poderá acessar um exemplo sobre visualização dos dados, utilizando os dados do jogo Pokémon Go, relacionando os sentimentos e as emoções de mais de 43.000 *tweets* sobre o jogo, capturados e representados, no mapa dos EUA, em forma de Pokébola.

| A importância da compreensão dos dados

Como podemos entender os dados em um gráfico?

Na Figura 6, para um determinado tipo de investidor em bolsa de valores, talvez seja a hora certa para comprar ou para vender um conjunto de ativos em carteira. A decisão se dará na compreensão baseada nos dados apresentados. Dessa forma, ao representar os dados visualmente, como em gráficos ou mapas, podemos simplificar conceitos complexos e facilitar a interpretação mesmo para aqueles sem experiência técnica.

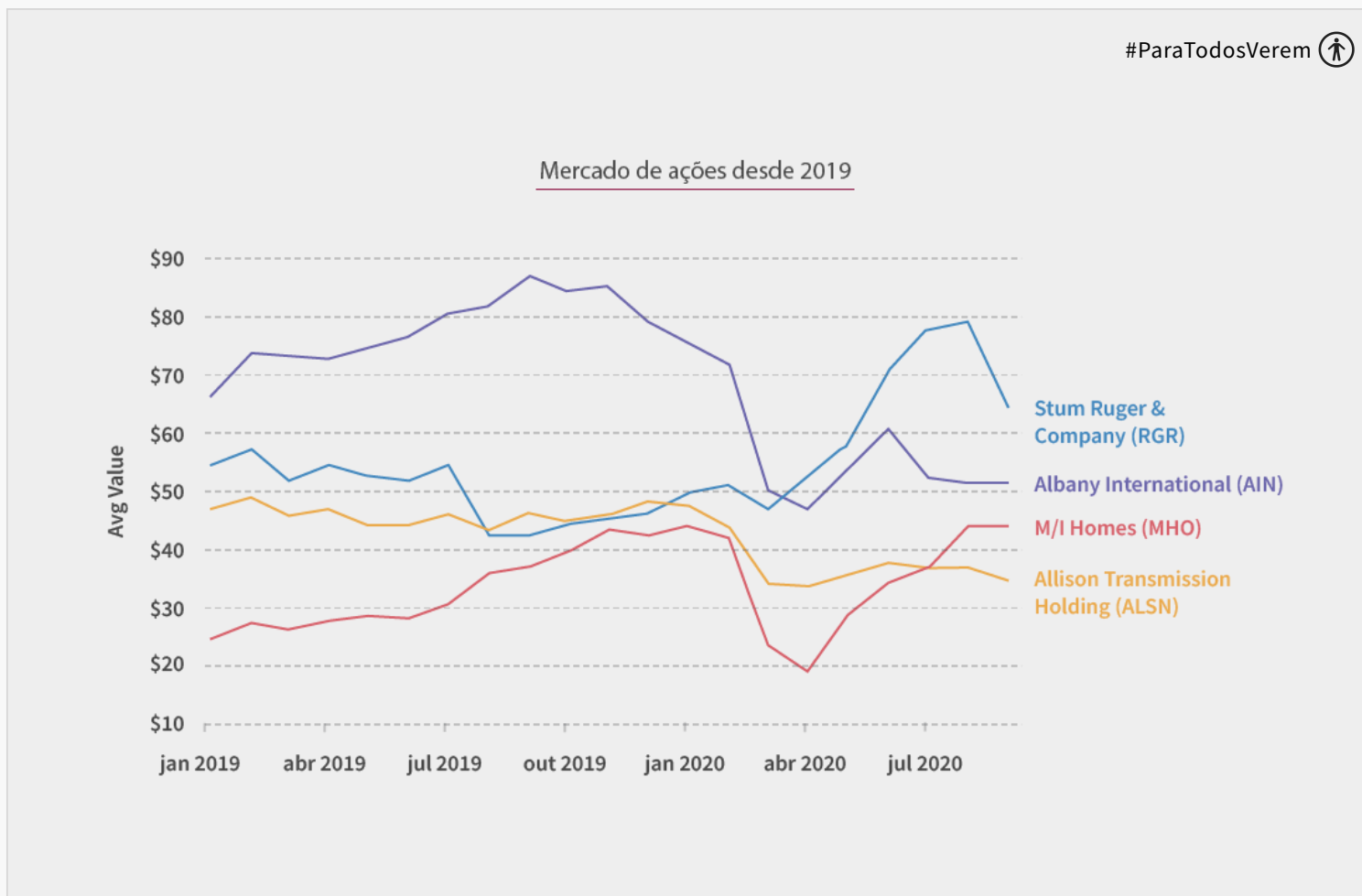


Figura 5: Bolsa de Valores 2019- 2020. Fonte: Data Hackers (2020).

Identificação de padrões e tendências

Por meio de gráficos e outras formas de visualização, podemos identificar padrões, tendências e correlações que poderiam passar despercebidos em conjuntos de dados brutos. Por exemplo, um gráfico de linha pode revelar a evolução de uma tendência ao longo do tempo, enquanto um gráfico de dispersão pode mostrar a relação entre duas variáveis. Essas descobertas são essenciais para compreender o comportamento dos dados e orientar análises mais aprofundadas.

| Frameworks e linguagens para visualização de dados

Existem diversos *frameworks* disponíveis para ajudar na visualização de dados, que facilitarão a criação de gráficos e *dashboards* interativos. No Quadro 2, trazemos exemplos de *frameworks* amplamente utilizados para visualização de dados em ambientes de *big data*. Essas ferramentas oferecem recursos avançados para análise e apresentação visual de grandes volumes de dados.

Quadro 2: Exemplos de *frameworks* para visualização de dados

Framework	Linguagem	Recursos	Vantagens	Desvantagens
Plotly	Python	<ul style="list-style-type: none">- Cria visualizações interativas em 2D e 3D.- Suporta diversos tipos de gráficos.- Integra-se com Pandas e NumPy.	<ul style="list-style-type: none">- Altamente personalizável.- Grande variedade de recursos.- Fácil de aprender e usar.	<ul style="list-style-type: none">- Pode ser lento para grandes conjuntos de dados.- Requer conhecimento básico de Python.
Bokeh	Python	<ul style="list-style-type: none">- Cria visualizações interativas de alta qualidade.- Dimensionamento automático, <i>zoom</i> e <i>pan</i>.- Legendas personalizáveis.	<ul style="list-style-type: none">- Ideal para criar painéis e aplicativos interativos.- Altamente personalizável.- Suporta grandes conjuntos de dados.	<ul style="list-style-type: none">- Curva de aprendizado mais íngreme.- Requer conhecimento de Python e HTML/CSS.
Holoviews	Python	<ul style="list-style-type: none">- Cria visualizações hierárquicas e multidimensionais.- Ferramentas para combinar e explorar conjuntos de dados.- Visualizações flexíveis e adaptáveis.	<ul style="list-style-type: none">- Altamente personalizável.- Ideal para análise de dados complexa.- Suporta grandes conjuntos de dados.	<ul style="list-style-type: none">- Curva de aprendizado mais íngreme.- Requer conhecimento de Python.
Tableau	Multiplataforma	<ul style="list-style-type: none">- Interface intuitiva arrastar e soltar.- Ampla gama de recursos de visualização.	<ul style="list-style-type: none">- Não requer conhecimento de programação.	<ul style="list-style-type: none">- Modelo de licenciamento caro.

		<ul style="list-style-type: none">- Análise e compartilhamento de dados fáceis.	<ul style="list-style-type: none">- Ideal para usuários não técnicos.- Poderosos recursos de análise.	<ul style="list-style-type: none">- Menos personalizável que algumas opções de código aberto.
Power BI	Multiplataforma	<ul style="list-style-type: none">- Integração com produtos da Microsoft (Excel, SQL Server).- Recursos poderosos de análise de negócios.- Boa opção para usuários do ecossistema Microsoft.	<ul style="list-style-type: none">- Não requer conhecimento de programação.- Ideal para usuários do Windows.- Integração perfeita com outros produtos da Microsoft.	<ul style="list-style-type: none">- Modelo de licenciamento caro.- Menos personalizável que algumas opções de código aberto.
Qlik Sense	Multiplataforma	<ul style="list-style-type: none">- Capacidade de lidar com grandes conjuntos de dados.- Interface de usuário intuitiva.- Poderosos recursos de análise.	<ul style="list-style-type: none">- Ideal para grandes volumes de dados.- Altamente personalizável.- Suporta análise em tempo real.	<ul style="list-style-type: none">- Modelo de licenciamento caro.- Curva de aprendizado mais íngreme.

Fonte: O autor (2024).

Linguagens de programação para visualização de dados

Além dos *frameworks*, há diversas linguagens de programação utilizadas para criar visualizações de dados. Entre as mais populares estão o *Python*, com bibliotecas como *Matplotlib*, *Seaborn* e *Plotly*, e o *JavaScript*, com ferramentas como *D3.js* e *Chart.js*. Essas linguagens fornecem flexibilidade e recursos poderosos para criar uma variedade de visualizações interativas e personalizadas.

| Conclusão

Agora você já sabe um pouco sobre visualização de dados e como podemos comunicar os dados utilizando recursos visuais, em especial no contexto de projetos de *big data*. Retomando a história do mapa, criado por Charles Minard, você já pensou se Napoleão tivesse tomado a decisão de invadir a Rússia do ponto de vista de dados ou se tivesse confiado no que a visualização de dados sobre o exército dele estava indicando, qual teria sido o destino da humanidade?

CUNHA LIMA, R. O. da. O que é infografia jornalística? **InfoDesign**: Revista brasileira de design da informação, v. 12, n. 1, p. 111–127, 2015. Disponível em: <https://doi.org/10.51358/id.v12i1.312>. Acesso em: 18 jun. 2024.

| Referências bibliográficas

DATA HACKERS. Data visualization: Uma arte objetiva. **Medium**, 2020. Disponível em: <https://medium.com/data-hackers/data-visualization-uma-arte-objetiva-6d28cd92e981>. Acesso em: 18 jun. 2024.

MICROSOFT. Data visualization: The past, present, and future. **Microsoft News**, 2024. <https://news.microsoft.com/stories/data/>. Acesso em: 1 maio 2024.

MILANI, A. M. P. *et al.* **Visualização de Dados**. Porto Alegre: SAGAH, 2020.

PEREIRA, M. A. *et al.* **Framework de Big Data**. Porto Alegre: SAGAH, 2020.



© PUCPR - Todos os direitos reservados.