



Universidade Estadual de Campinas  
Faculdade de Tecnologia



Rafael Tavares Carvalho Barros

Uso de matrizes reordenáveis e heatmaps para apoio à  
análise de dados de históricos escolares

Limeira  
2019

**Rafael Tavares Carvalho Barros**

**Uso de matrizes reordenáveis e heatmaps para apoio à análise de dados de históricos escolares**

Dissertação apresentada à Faculdade de Tecnologia da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Bacharel em Sistemas de Informação.

**Orientador: Prof. Dr. Celmar Guimarães da Silva**  
**Coorientador: Prof. Thiago Gonçalves Mendes**

Este exemplar corresponde à versão final da Dissertação defendida por Rafael Tavares Carvalho Barros e orientada pelo Prof. Dr. Celmar Guimarães da Silva.

Limeira  
2019



Universidade Estadual de Campinas  
Faculdade de Tecnologia



Rafael Tavares Carvalho Barros

## Uso de matrizes reordenáveis e heatmaps para apoio à análise de dados de históricos escolares

### Banca Examinadora:

- Prof. Dr. Celmar Guimarães da Silva  
FT/UNICAMP
- Profa. Dra. Gisele Busichia Baioco  
FT/UNICAMP
- Profa. Dra. Tania Basso  
COTIL/UNICAMP

A ata da defesa com as respectivas assinaturas dos membros da banca encontra-se no processo de vida acadêmica do aluno.

Limeira, 27 de junho de 2019

# Resumo

Este trabalho foi desenvolvido com o objetivo de criar uma ferramenta usando técnicas de visualização de informação para auxiliar coordenadores de curso na análise de dados sobre históricos escolares. Com um conjunto de históricos escolares como entrada de dados à respeito de disciplinas e alunos, estes são exibidos em forma de um *Heatmap* (mapa de calor), tendo notas representadas com diferentes cores dentro de uma escala que pode ser personalizada pelo usuário. Para auxiliar a análise, é possível realizar diferentes ordenações com base em algoritmos de reordenação existentes ou de acordo com uma disciplina ou aluno em específico, obter uma visão geral ou específica utilizando o recurso de *zoom* e filtragem das turmas que serão exibidas no *Heatmap*, que conta ainda com detalhes sob demanda. A análise do gráfico poderá auxiliar coordenadores de curso na identificação de disciplinas em que os alunos estão tendo um desempenho muito diferente de outras (como, por exemplo, uma disciplina com muitas reprovações) ou em diferenças relevantes entre turmas (de diferentes anos ou cursos) para a mesma disciplina, isso de uma maneira muito mais rápida do que através de uma análise textual dos dados em questão.

# Abstract

This final paper was developed with the objective of creating a tool using visualization techniques to assist course coordinators in the analysis of data on school records. With a set of school records as input for disciplines and students, they are displayed in the form of a Heatmap, having notes represented with different colors within a scale that can be customized by the user. To help the analysis, it is possible to perform different sorts based on existing reordering algorithms or sorts according to a specific discipline or student, get an overview or specific view using the zoom feature and filtering of the disciplines that will be displayed in the Heatmap, which also has details on demand. Analysis of the graph may assist course coordinators in identifying disciplines where students are performing very differently from others (such as a discipline with many failures) or relevant differences between classes (from different years or courses) for the same discipline, this much more quickly than through a textual analysis of the data in question.

# Lista de Figuras

2.1	Nota média em Matemática ao longo do ciclo letivo . . . . .	13
2.2	Mapeando valores numéricos em tamanhos de barras . . . . .	13
2.3	Ferramentas mais utilizadas para fins educacionais . . . . .	14
2.4	Proprietários das ferramentas mais utilizadas para fins educacionais . . . . .	15
2.5	Frequência em escolas por grupo étnico . . . . .	16
2.6	Informações sobre o grupo étnico desejado . . . . .	16
2.7	Desempenho educacional por estado nos EUA . . . . .	17
2.8	Informações complementares sobre o estado . . . . .	17
2.9	Interesse sobre o ENEM no Brasil . . . . .	18
2.10	Conjunto de dados Iris projetado utilizando MDS . . . . .	20
2.11	Exemplo de coordenadas paralelas . . . . .	21
2.12	Doenças que mais causam mortes por país . . . . .	22
2.13	Reordenação de matriz com base na primeira coluna . . . . .	24
4.1	Parte do Heatmap de históricos escolares . . . . .	28
4.2	Disciplinas que envolvem matemática causando mais reprovações . . . . .	29
4.3	Detalhes sob demanda no Heatmap . . . . .	30
4.4	Gráfico gerado com base no arquivo exemplificado . . . . .	32
5.1	Entrada de dados e personalização do gráfico . . . . .	33
5.2	Imagem do sistema após o gráfico ser gerado . . . . .	34
5.3	Gráfico ordenado por aluno e disciplina . . . . .	35
5.4	Gráfico ordenado por aluno e disciplina . . . . .	35
A.1	Formato recebido pelo D3.js . . . . .	42
B.1	Matrizes de distância para o Reorder.js . . . . .	43

# Lista de Tabelas

4.1	Exemplo de arquivo de entrada .csv . . . . .	32
-----	--	----

# Sumário

<b>1</b>	<b>Introdução</b>	<b>9</b>
<b>2</b>	<b>Visualização de Informação</b>	<b>11</b>
2.1	Propriedades visuais . . . . .	14
2.1.1	Tamanho . . . . .	14
2.1.2	Matiz (ou cor) . . . . .	14
2.1.3	Posição espacial . . . . .	15
2.2	Relevância . . . . .	18
2.3	Visualização de dados multidimensionais . . . . .	19
2.3.1	Baseada em pontos . . . . .	19
2.3.2	Baseada em linhas . . . . .	20
2.3.3	Baseada em regiões . . . . .	21
2.3.4	Heatmap . . . . .	22
2.3.5	Matriz reordenável . . . . .	23
<b>3</b>	<b>Mineração de Dados Educacionais</b>	<b>25</b>
3.1	Pesquisas na área . . . . .	26
<b>4</b>	<b>Desenvolvimento</b>	<b>28</b>
<b>5</b>	<b>Resultados</b>	<b>33</b>
<b>6</b>	<b>Conclusão</b>	<b>37</b>
	<b>Referências bibliográficas</b>	<b>38</b>
<b>A</b>	<b>Formato de entrada para o D3.js</b>	<b>42</b>
<b>B</b>	<b>Formato de entrada para o Reorder.js</b>	<b>43</b>

# Capítulo 1

## Introdução

Este trabalho abrange a área da educação. Esta foi escolhida porque possui uma grande quantidade de dados que podem ser utilizados na análise e visualização para auxiliar a identificação de fatores que afetam os processos de ensino e aprendizado, possibilitando melhorá-los, onde os usuários finais da visualização conseguirão compreender os dados e a informação que constituem sem muito treinamento (QU; QING, 2015; RITSOS; ROBERTS, 2014).

Para os coordenadores de curso da Faculdade de Tecnologia da UNICAMP, o processo existente para análise de dados referentes ao curso ou aluno são majoritariamente textuais. É difícil ou demorado para o coordenador verificar se uma disciplina está com uma nota média muito diferente de outras, identificar se houve uma diferença entre turmas de anos ou cursos diferentes para a mesma disciplina, ou até mesmo verificar se um aluno teve um desempenho muito diferente em um determinado semestre.

Com isso, este trabalho tem como objetivo auxiliar coordenadores de curso na análise do curso e no auxílio de alunos através da representação visual de históricos escolares. A visualização é representada através de um gráfico, onde as notas são exibidas em diferentes cores dentro de uma escala que pode ser personalizada pelo usuário, facilitando a percepção de possíveis *outliers* (pontos fora da curva), como uma disciplina em que a maioria dos alunos tem um desempenho ruim.

O gráfico escolhido para esta representação foi o *Heatmap*, por possibilitar a representação de variáveis quantitativas e compactar um grande conjunto de informações em um pequeno espaço, além de ser possível reorganizar linhas e colunas para expor carac-

terísticas e facilitar a visualização de padrões coerentes nos dados (WEINSTEIN, 2008; WARD; GRINSTEIN; D. KEIM, 2010).

O gráfico exibirá as disciplinas e alunos presentes no conjunto de dados, bem como as respectivas notas, para então auxiliar coordenadores de curso na identificação de possíveis problemas do curso referentes às disciplinas (como uma disciplina que tem reprovado muitos alunos nos últimos anos) e também facilitar a gestão de conhecimento, onde novos coordenadores poderão entender alguns aspectos do curso de uma maneira visual e rápida, sem precisar analisar textualmente o histórico de alunos ou das disciplinas.

O *Heatmap* foi desenvolvido para ambiente Web, por ser acessível independente da plataforma que o usuário estará utilizando (dispositivos Android, iOS, Windows, Linux etc.), além de possuir tecnologias *open-source*, com comunidades crescentes, como o D3.js, facilitando o desenvolvimento (JAIN, 2014).

O conteúdo deste trabalho está disposto de maneira a facilitar o entendimento de todo o contexto da visualização e interpretação humana através da visão (Capítulo 2), as propriedades visuais (Seção 2.1), a relevância da visualização (Seção 2.2) e alguns exemplos de visualização (Seção 2.3). Após isto, é apresentada a mineração de dados educacionais e sua relação com a visualização (Capítulo 3), para então abordar os aspectos mais técnicos no desenvolvimento (Capítulo 4).

Após a metodologia, há uma amostra e explicação dos resultados obtidos (Capítulo 5), com imagens do sistema desde a entrada dos históricos escolares até a exibição do gráfico. Por fim, a conclusão diante dos resultados alcançados (Capítulo 6), mencionando a necessidade da utilização da ferramenta para validação da própria e possíveis trabalhos futuros com o objetivo de suprir necessidades possivelmente não contempladas.

## Capítulo 2

# Visualização de Informação

Com a evolução tecnológica, a humanidade passou a gerar e conseqüentemente armazenar cada vez mais dados, possibilitando a utilização desses dados para o entendimento de cenários diversos.

Para especificamente gerenciar esses dados surgiram carreiras como as de analista e mais recentemente a de cientista de dados, essa evolução é impulsionada devido a vantagem estratégica na análise dos mesmos, dado que elas podem promover um melhor entendimento do impacto de ações a serem planejadas ou que já foram realizadas, o impacto de estabelecimento ou não cumprimento de metas, bem como a criação de planos de ações para novos produtos, cenários etc.

Em um exemplo clássico aos olhos do público em geral, lojas online passaram a ter conhecimento sobre o seu cliente com mais facilidade quando comparadas às lojas físicas. É possível saber quais páginas o cliente visitou para identificar o tipo de produto em que ele tem interesse, realizar marketing direcionado através de e-mail ou SMS e saber quais promoções enviadas foram acessadas por ele.

Como exemplo, a rede americana Walmart é capaz de criar estimadamente 2,5 Petabytes de dados por hora em suas transações (1 petabyte equivale a aproximadamente 1 milhão de Gigabytes) e em um cenário mundial estima-se a criação de 2,5 Exabytes de dados por dia (1 exabyte equivale a cerca de 1 bilhão de Gigabytes) (MCAFEE; BRYNJOLFSSON, 2012).

Já em 2016, o tráfego de dados na Internet chegou a 26,600 GB por segundo (cerca de 2,3 Exabytes por dia) (CISCO, 2017). E em outubro de 2018, aconteceu um tráfego

de 63,419 GB de dados na Internet por segundo (aproximadamente 5,5 Exabytes por dia) (INTERNET LIVE STATS, 2018).

Os números podem variar de acordo com a fonte, mas o tráfego diário na Internet está atualmente na casa dos Exabytes. Este grande volume de dados e a velocidade em que são gerados torna necessária uma maneira eficaz e rápida de analisá-los.

Analisando outro cenário, de instituições como universidades, escolas de ensino regular e empresas que promovem cursos profissionalizantes ou livres, nota-se que, com a digitalização dos dados para sistemas de gestão escolar, é possível visualizar o desempenho de diferentes alunos em diferentes disciplinas sem precisar procurar por arquivos físicos em papéis ou dados descentralizados em planilhas.

Dependendo do sistema, também é possível gerar diferentes gráficos para diferentes conjuntos de dados, facilitando a interpretação destes pelo usuário final sem exigir do mesmo muito esforço, ou tempo para esta tarefa.

Imagine que um coordenador de uma dessas instituições precise analisar a nota de várias turmas diferentes ao longo dos anos, tendo em sua análise todas as disciplinas. A princípio, o coordenador poderia escolher uma matéria (como Matemática) e visualizar a mudança da nota dos alunos conforme o avanço no ciclo letivo.

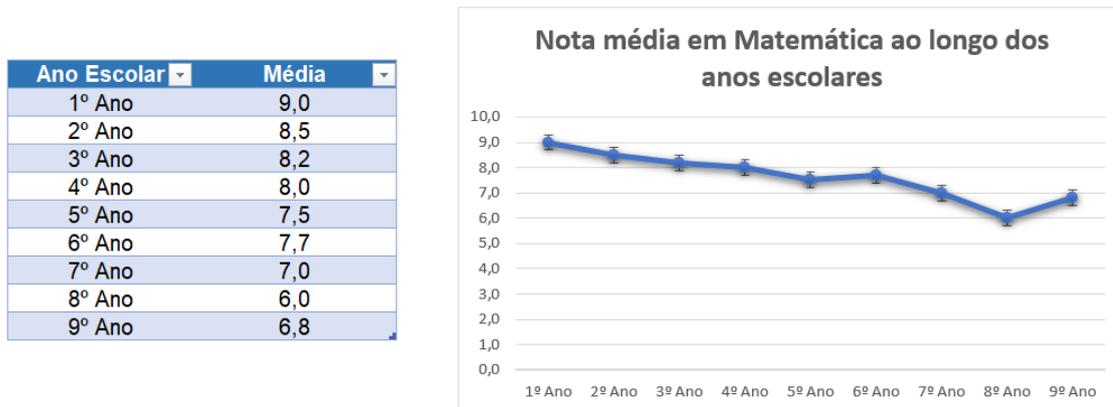
Realizar essa análise tendo que ler diversos boletins seria um trabalho não trivial, pois seria necessário contar quantos alunos já fizeram a disciplina, quais notas eles tiveram e calcular uma média para todos os anos de interesse.

Contudo, em um sistema que é capaz de armazenar e manipular esses dados, seria necessário apenas indicar que você deseja visualizar a nota média em Matemática ao longo do ciclo e então visualizá-lo em um gráfico, por exemplo como o de linhas, conforme ilustrado na Figura 2.1.

De acordo com MAZZA (2009), uma representação visual possui algumas vantagens em relação à visualização textual, pois existem propriedades gráficas que são processadas rápida e eficientemente pelo observador. Essas vantagens podem ser relacionadas com o funcionamento do sistema visual humano.

Um importante resultado inicial sobre como esse sistema funciona, de acordo com WARD, GRINSTEIN e D. KEIM (2010), foi a descoberta de um conjunto limitado de propriedades visuais que podem ser detectados de maneira rápida e precisa, entre 200ms (tempo mínimo de movimento do olho) e 250ms.

Figura 2.1: Nota média em Matemática ao longo do ciclo letivo. Dados fictícios.

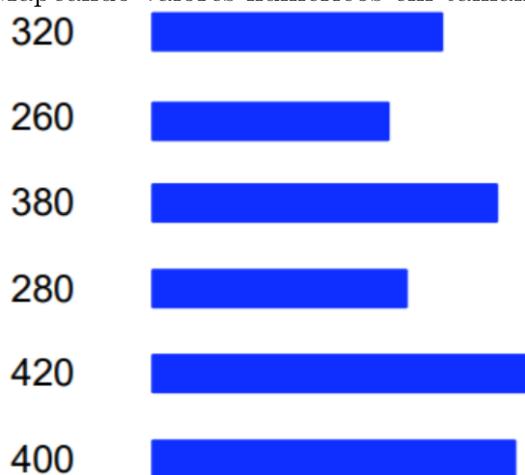


Fonte: Autor.

Essas propriedades visuais são chamadas de *preattentive*. Matiz (*hue*), curvatura (forma), comprimento, largura, tamanho, quantidade, intersecção, proximidade, movimento, direção e cintilação são alguns exemplos de características visuais identificadas como *preattentive* (WARD; GRINSTEIN; D. KEIM, 2010).

No exemplo da Figura 2.2 (MAZZA (2009)) há uma sequência de valores numéricos representados de forma textual e visual, através de barras horizontais com tamanho proporcional ao valor.

Figura 2.2: Mapeando valores numéricos em tamanhos de barras.



Fonte: MAZZA (2009).

Caso o observador deseje identificar o maior e o menor valor do conjunto observando apenas os números, seria necessário ler todos os valores e manter sempre em mente o maior e menor valor lido até o momento. Já com a representação gráfica, é possível perceber

com maior facilidade qual barra possui o maior tamanho e qual o menor, visto que esta é uma característica preattentive (MAZZA, 2009; WARD; GRINSTEIN; D. KEIM, 2010).

## 2.1 Propriedades visuais

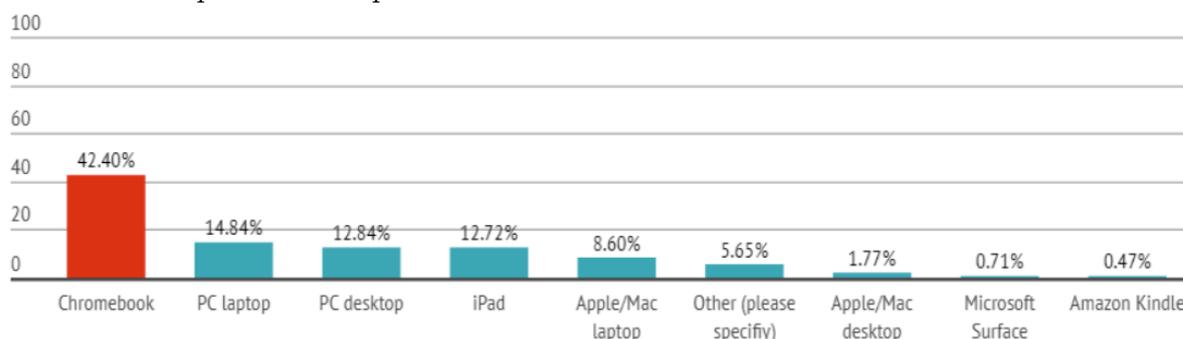
Como os humanos percebem propriedades visuais muito bem, é possível representar diferentes dados de diferentes maneiras, conforme exemplificado nas seções seguintes:

### 2.1.1 Tamanho

Como já visto anteriormente, o tamanho geralmente está diretamente relacionado a um valor numérico. Na Figura 2.3, há uma comparação entre quais das ferramentas fornecidas pelas escolas nos Estados Unidos os educadores e alunos usam com maior frequência para fins educativos (CAVANAGH, 2017).

Com o uso das barras é visível o destaque do Chromebook, com 42,40%, em relação aos demais. A utilização do tamanho neste exemplo e a ordenação dos dados em escala decrescente facilita uma comparação rápida sobre a diferença entre o uso dos dispositivos mencionados.

Figura 2.3: Quais das seguintes ferramentas fornecidas pelas escolas os educadores e alunos usam mais frequentemente para fins educativos.



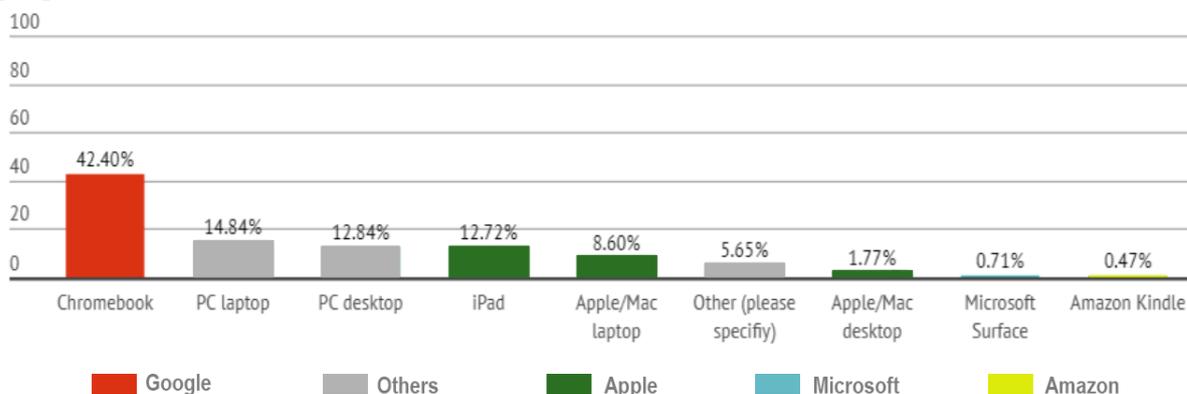
Fonte: CAVANAGH (2017).

### 2.1.2 Matiz (ou cor)

Cores diferentes podem possuir diferentes significados para um mesmo conjunto de dados. No exemplo acima, a cor vermelha foi utilizada para destacar a ferramenta mais utilizada. Na Figura 2.4, as cores do gráfico foram modificadas para corresponderem à empresa

proprietária do equipamento. Os equipamentos genéricos (*PC laptop*, *PC desktop* e *Other*) foram agrupados em *Others* (outras).

Figura 2.4: Gráfico da Figura 2.3 com as cores modificadas de acordo com a empresa proprietária do hardware.



Fonte: CAVANAGH (2017). Adaptada pelo autor.

### 2.1.3 Posição espacial

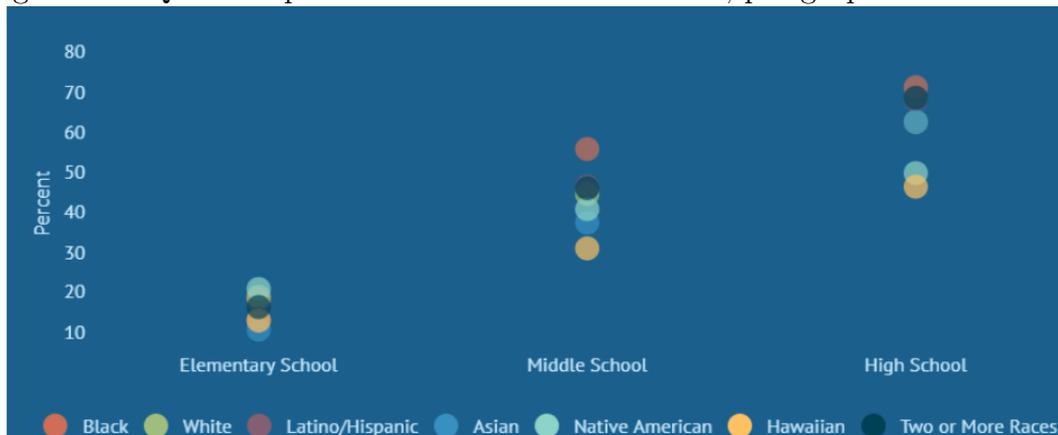
A posição representativa de um dado no gráfico também pode estar associada a outras variáveis. A Figura 2.5, representa qual percentual de cada grupo étnico e racial frequentam escolas com oficiais da lei nos EUA (DECKER, 2017).

O eixo horizontal do gráfico representa o nível escolar (Escola Primária, Ensino Fundamental e Colegial, respectivamente) e o eixo y representa o percentual em relação ao total de alunos daquele grupo. Desta forma, a posição da plotagem do ponto remete os dois dados em relação aos respectivos eixos. O gráfico também faz o uso de cores para a representação do grupo étnico e racial.

Representações deste tipo podem apresentar problemas em relação à plotagem de pontos, por exemplo: alguns grupos étnicos e raciais estão sobrepostos na figura, dificultando a identificação da faixa percentual em que eles se encontram; este tipo de gráfico não apresenta uma solução para este problema.

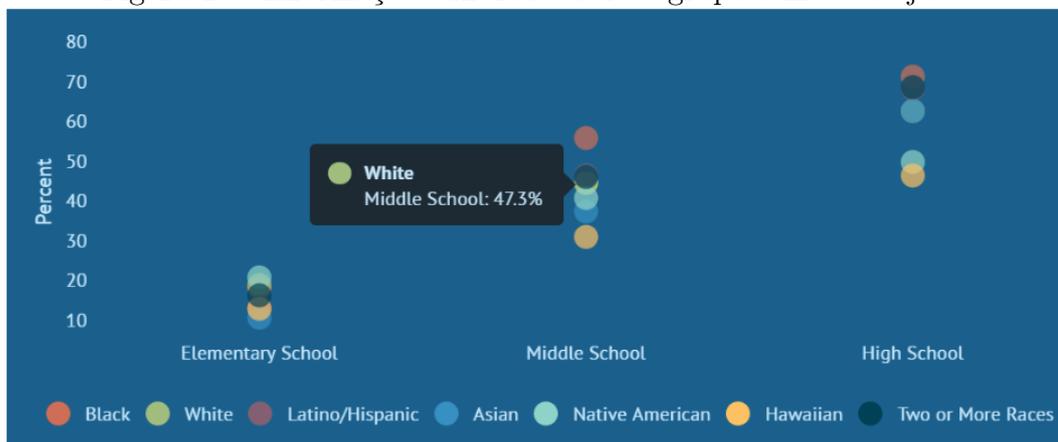
Outra questão é que o observador terá dificuldade em definir com precisão qual é o percentual do grupo étnico e racial desejado. Entretanto, o gráfico original é interativo, permitindo uma análise mais acurada dos dados representados ao posicionar o ponteiro do mouse sobre um ponto plotado no gráfico (Figura 2.6).

Figura 2.5: Quem frequenta escolas com oficiais da lei, por grupo étnico e racial.



Fonte: DECKER (2017).

Figura 2.6: Informações exibidas sobre o grupo étnico desejado.



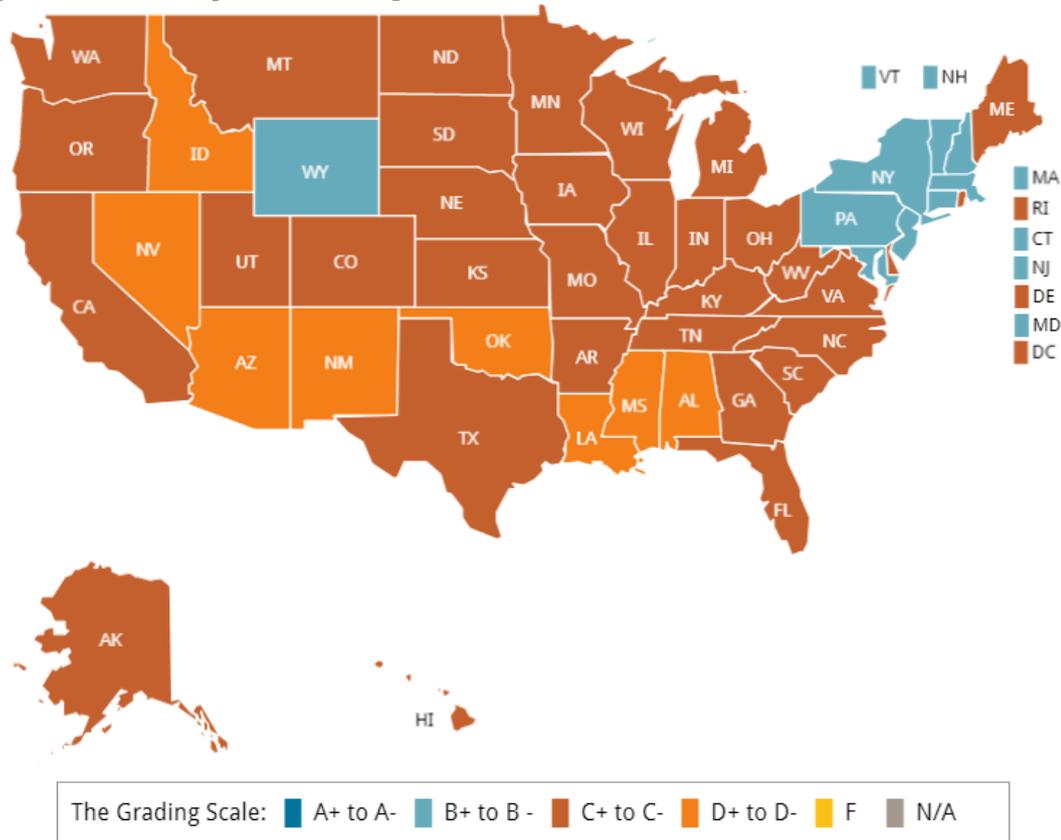
Fonte: DECKER (2017).

A posição de um componente no gráfico pode também estar associada a outro elemento, como na Figura 2.7, onde cada elemento é um estado e deve estar na posição correta para representar os Estados Unidos.

O EDUCATION WEEK (2016) avaliou os EUA e seus estados no desempenho escolar com base em alguns indicadores-chave. O país pontuou 74.2 (nota C) como sendo uma média nacional. Em uma análise secundária, conforme mostrado na Figura 7 abaixo, a maioria dos estados (34) foram avaliados entre C- e C+.

Um problema que este tipo de exibição gera é visível ao nordeste do mapa, onde estados como Massachusetts (MA), Vermont (VT), New Hampshire (NH), Rhode Island (RI) e alguns outros acabaram ficando com a legenda deslocada devido à falta de espaço no território do estado.

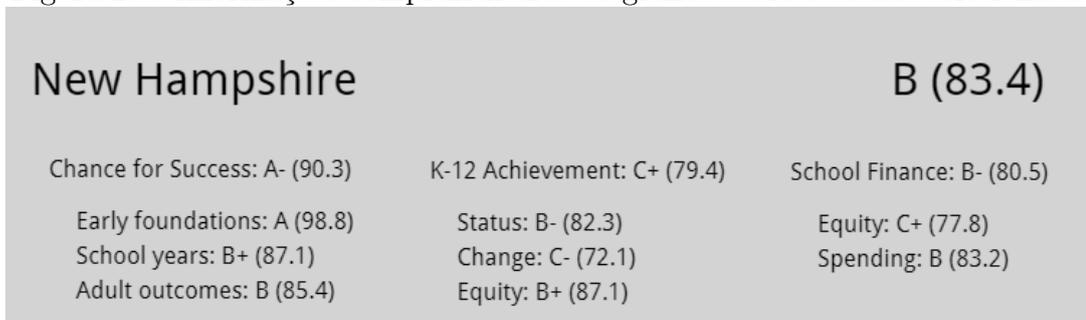
Figura 2.7: Avaliação do desempenho educacional dos estados dos Estados Unidos.



Fonte: EDUCATION WEEK (2016).

Este problema pode ser contornado com uma legenda próxima ao território, como foi feito, ou com uma ação interativa, onde o observador pode selecionar o estado desejado com o mouse e informações complementares aparecem abaixo do mapa, no site do gráfico, em conjunto de alguns indicadores usados para avaliar o estado em questão, conforme Figura 2.8.

Figura 2.8: Informações complementares do gráfico sobre o estado selecionado.



Fonte: EDUCATION WEEK (2016).

## 2.2 Relevância

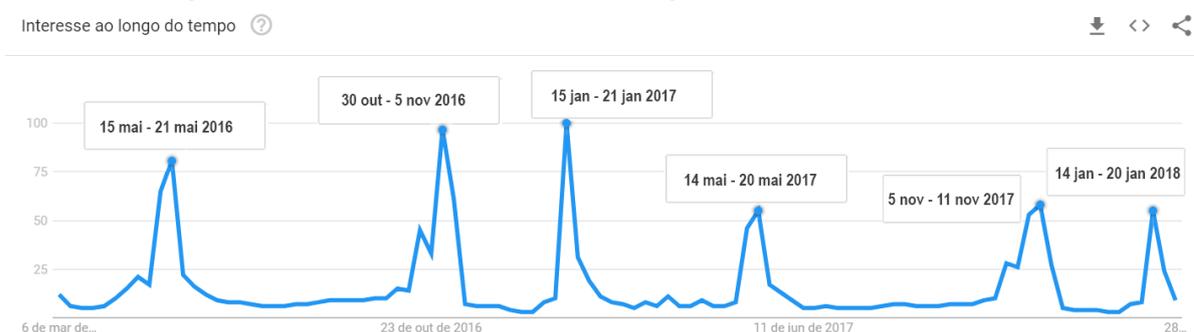
Representações visuais auxiliam na expressão de ideias e conceitos que são mais facilmente entendíveis do que se comparados a uma representação verbal ou textual. Até mesmo em situações mais simples, como em banheiros públicos, é utilizada uma representação visual para que toda a população consiga entender sem a barreira de que somente aqueles que saibam ler consigam compreender o significado da sinalização.

De acordo com MAZZA (2009), em níveis mais profundos, como na análise de dados, a capacidade de análise pela percepção visual e pelo sistema cognitivo humano amplia consideravelmente a eficácia da interpretação através de um conjunto de dados representado de diferentes formas.

A visualização através de gráficos serve para a confirmação de ideias, como por exemplo: em que época as pessoas pesquisam mais no Google sobre o ENEM? O esperado é que seja perto do dia das provas, mas na verdade o ENEM é mais do que um final de semana de provas, e o Google demonstra isso através do interesse das pessoas em diferentes épocas (GOOGLE, 2018).

Na Figura 2.9 pode-se ver que, considerando pesquisas no Google no Brasil, o ENEM possui três picos anuais de interesse, sendo: a época de inscrição (Maio), o período de prova (Outubro ou Novembro) e a liberação dos resultados (Janeiro). A imagem foi modificada para exibir as datas dos picos.

Figura 2.9: Interesse no Brasil ao longo do tempo sobre o ENEM.



Fonte: GOOGLE (2018). Adaptada pelo autor.

Descobrir que existe um maior interesse no ENEM nesses três períodos do ano e identificar os diferentes níveis de relevância entre cada um destes picos seria uma tarefa mais trabalhosa sem uma representação visual dos dados.

## 2.3 Visualização de dados multidimensionais

O objetivo da análise e visualização dos dados é evidenciar informações relevantes e apoiar a tomada de decisão. No ramo educacional, por exemplo, isso pode ajudar o coordenador de curso a identificar correlações entre disciplinas de acordo com as notas dos alunos, reconhecer a melhor grade curricular para o aluno no semestre escolhido, visualizar o desempenho de diferentes turmas em diferentes disciplinas e, com essas informações disponíveis, tomar decisões que possam tornar a experiência dos alunos melhor e mais proveitosa.

Isto possibilita uma análise geral de como o curso pode ser para os alunos, quais as maiores dificuldades ou facilidades e quais disciplinas afetam negativamente o desempenho do aluno em outras disciplinas. Por exemplo, se um aluno tem maus resultados em uma disciplina A, provavelmente terá maus resultados também na disciplina A II, mas e numa disciplina B ou C? Existe uma correlação relevante entre essas disciplinas?

Entretanto, para uma análise neste sentido, será necessário lidar com um conjunto de variáveis multidimensionais (por exemplo, a quantidade de alunos, notas e disciplinas). Para isso, existem quatro grandes tipos de visualizações de dados multidimensionais: baseadas em pontos, em linhas, em regiões e híbridas.

### 2.3.1 Baseada em pontos

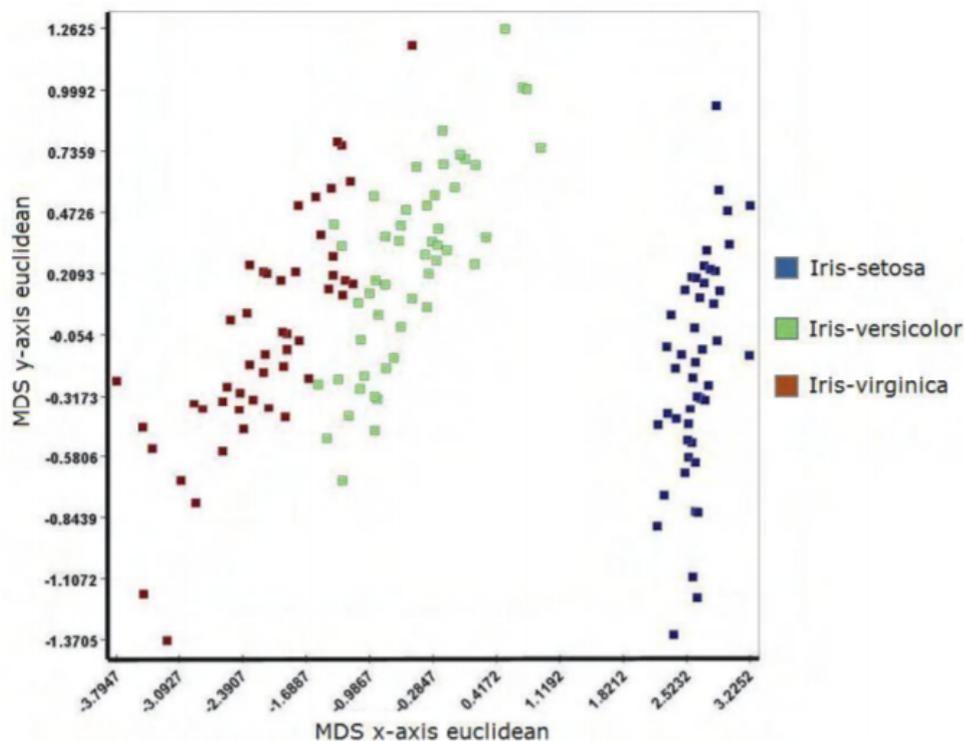
De acordo com WARD, GRINSTEIN e D. KEIM (2010), os gráficos de pontos são introduzidos como visualizações que projetam registros de um espaço de dados n-dimensional para um espaço de exibição arbitrariamente k-dimensional. Os gráficos de pontos podem ser definidos para exibir registros individuais ou registros de resumo e podem ser estruturados por várias técnicas de projeção.

Um exemplo de gráfico de pontos é o MDS (*multidimensional scaling*), que é a busca por um espaço dimensional baixo, geralmente euclidiano, no qual os pontos no espaço representam os objetos, de forma que as distâncias entre os pontos no espaço combinam, tanto quanto possível, as diferenças originais (T. F. COX; M. A. A. COX, 2000).

A Figura 2.10 mostra um exemplo do conjunto de dados Iris, sobre três tipos de flores (Íris Setosa, Íris Virgínica e Íris Versicolour), que contém atributos em quatro dimensões numéricas, referentes à largura e comprimento das pétalas e sépalas das flores. Estes

atributos são projetados usando MDS para representação no eixo euclidiano. Quanto mais próximos estão os pontos, mais similares eles são.

Figura 2.10: Conjunto de dados Iris projetado com a utilização do MDS.



Fonte: WARD, GRINSTEIN e D. KEIM (2010).

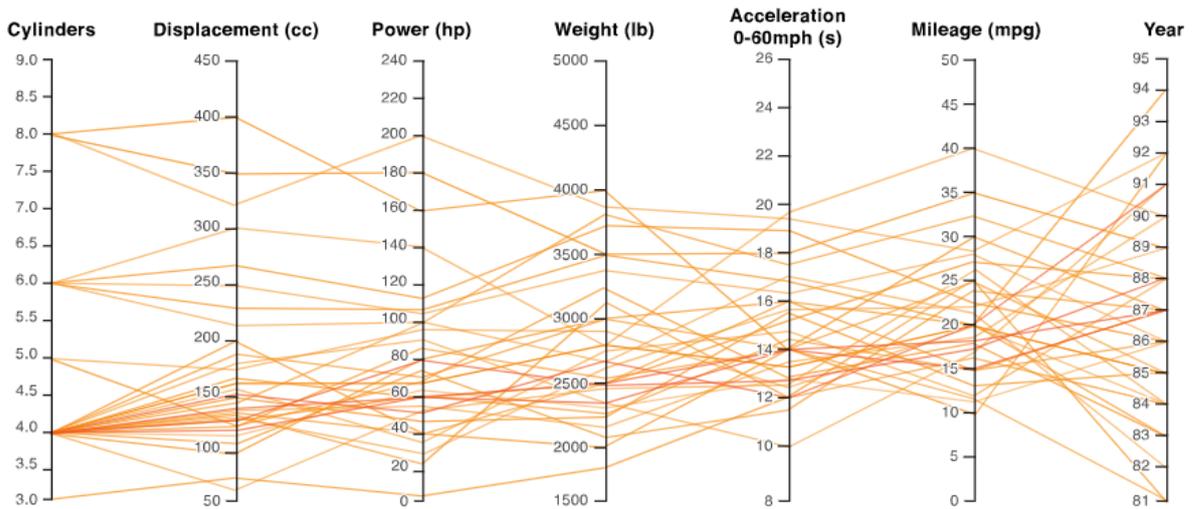
### 2.3.2 Baseada em linhas

De acordo com WARD, GRINSTEIN e D. KEIM (2010), nos métodos baseados em linhas, os pontos correspondentes a um registro ou dimensão particular são vinculados a linhas retas ou curvas. Essas linhas reforçam as relações entre os valores de dados e transmitem características perceptíveis dos dados por meio de inclinações, curvatura, cruzamentos e outros padrões de linhas.

Um tipo de gráfico baseado em linhas é o de coordenadas paralelas (*parallel coordinates*), que produz representações gráficas de relações multidimensionais em vez de apenas conjuntos de pontos finitos. Os eixos são paralelos, com linhas horizontais ou verticais uniformemente espaçadas representando uma ordenação particular das dimensões. Um ponto de dado é plotado como uma polilinha que cruza cada eixo em uma posição proporcional ao seu valor para a dimensão representada (INSELBERG; DIMSDALE, 1990; WARD; GRINSTEIN; D. KEIM, 2010).

A Figura 2.11 representa um conjunto de dados de automóveis, onde cada linha é um automóvel, cada eixo é uma dimensão (cilindros, deslocamento, potência etc.) e o ponto que a linha cruza no eixo representa o valor que aquele automóvel possui naquela dimensão.

Figura 2.11: Conjunto de dados de automóveis projetado com coordenadas paralelas.



Fonte: Catalogue (2019).

### 2.3.3 Baseada em regiões

De acordo com WARD, GRINSTEIN e D. KEIM (2010), nas técnicas baseadas em regiões os polígonos preenchidos são usados para transmitir valores, com base em seu tamanho, forma, cor ou outros atributos. Para alguns gráficos, o objetivo não é mostrar os dados brutos em si, mas sim resumos ou distribuições dos valores. Algumas técnicas baseadas em regiões foram projetadas para dados univariados, como gráficos de pizza e gráficos de barras, outras, foram projetadas para várias dimensões, como visualizações tabulares.

Dados multivariados são frequentemente armazenados em tabelas, e várias técnicas de visualização foram modeladas nessa estrutura. Heatmap é um exemplo de visualização tabular, ele exibe a tabela de valores de registro usando cores ao invés de texto. Para essa técnica de visualização, todos os valores de dados são mapeados para o mesmo espaço de cores normalizado e cada um é renderizado como um quadrado ou retângulo colorido (WARD; GRINSTEIN; D. KEIM, 2010).

### 2.3.4 Heatmap

De acordo com WEINSTEIN (2008), uma forma de visualização gráfica popular é o “mapa de calor agrupado” (*clustered heat map*), pois consegue representar grande quantidade de dados em um pequeno espaço para a análise de possíveis padrões no conjunto.

Esse gráfico é composto de uma tabela onde cada linha e coluna representa uma variável diferente, e cada célula representa uma terceira variável, que pode ser quantitativa, como na figura Figura 2.12, ou ordinal, como o nível de perigo em se tomar uma decisão (baixo, médio ou alto).

A cor da célula varia conforme um conjunto de possibilidades. No caso da expressão de número de ocorrências, varia da menor quantidade para a maior, ou no caso de variáveis ordinais, varia de acordo com a ordem escolhida. Essa variação pode ser representada através de uma escala de tons de cor (do azul claro para o escuro, por exemplo) ou através de cores diferentes que podem ter um significado.

Figura 2.12: Classificação de doenças que mais causam mortes por país.

	China	Indonesia	Fiji	Kazakhstan	Poland	Ukraine	Japan	Australia
Cardiovascular diseases	1	1	1	1	1	1	2	1
Neoplasms	2	2	3	2	2	2	1	2
Chronic respiratory	3	6	4	5	5	8	6	4
Neurological disorders	4	8	8	6	3	3	3	3
Diabetes & CKD	5	3	2	10	7	11	7	5
Unintentional inj	6	11	6	7	8	6	8	7
Digestive diseases	7	5	10	3	4	4	5	6
Transport injuries	8	10	13	9	11	10	11	11
Respiratory infections & TB	9	4	5	8	6	7	4	8
Self-harm & violence	10	17	11	4	9	5	9	9
Other non-communicable	11	12	9	13	12	13	10	10
Maternal & neonatal	12	9	7	12	13	14	18	14
Substance use	13	20	18	11	10	9	16	12

Fonte: INSTITUTE FOR HEALTH METRICS AND EVALUATION (2017).

Na Figura 2.12, como exemplo de um heatmap, as linhas representam doenças e as colunas representam países. Há também a classificação da doença em relação ao país, representada através de número e cor, sendo 1 (em vermelho) a doença que mais causa mortes e 20 (em azul) a que menos causa mortes (o gráfico original apresenta 21 doenças no total), Neste caso, as linhas estão ordenadas conforme a coluna China.

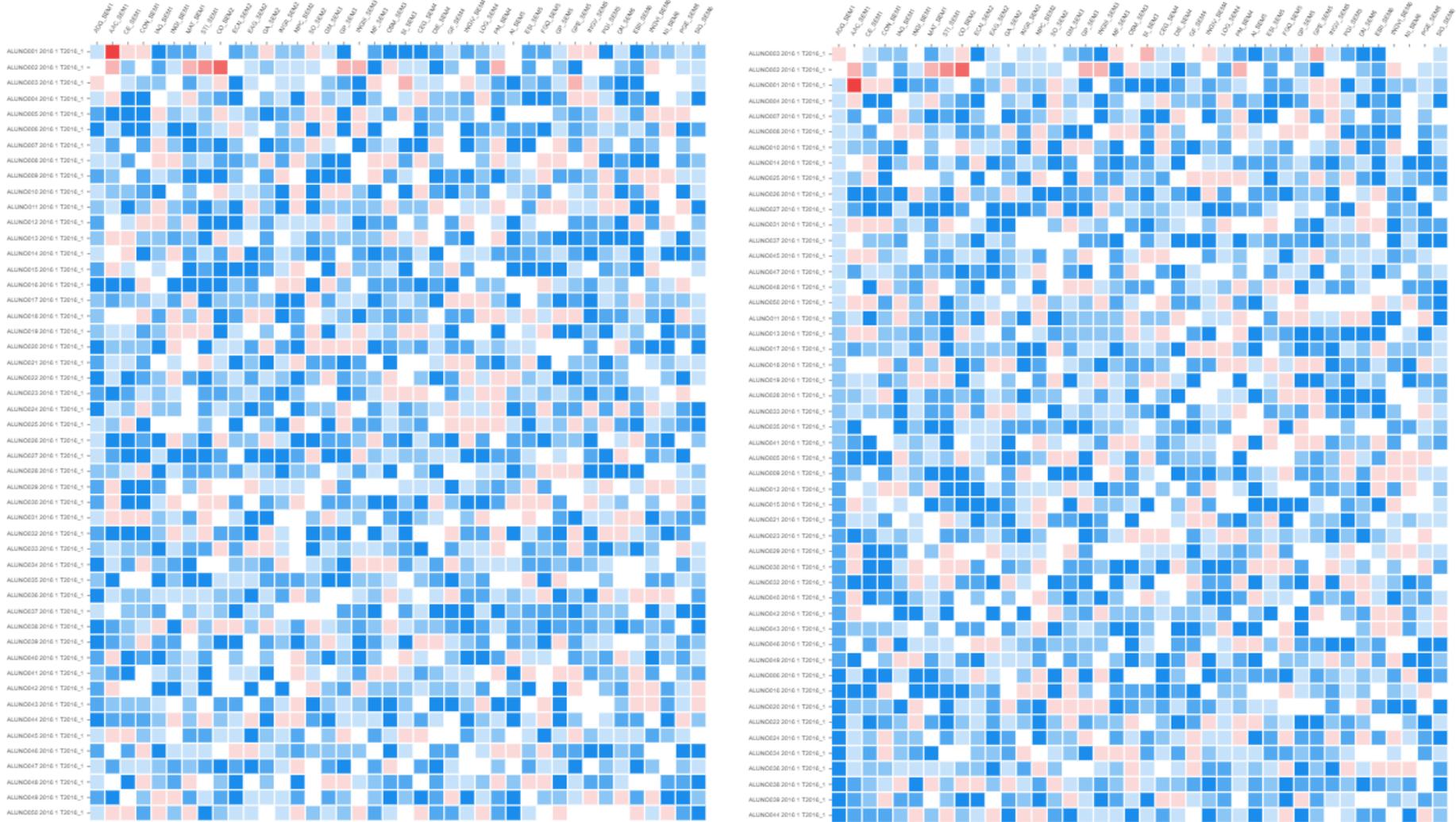
### 2.3.5 Matriz reordenável

A matriz reordenável é um método de visualização para exploração de dados tabulares. A ideia básica é transformar um conjunto de dados multidimensional em um gráfico interativo bidimensional (2D). As linhas e colunas que são representadas podem ser permutadas, permitindo diferentes visualizações do conjunto de dados. Ao interagir com a apresentação visual, o usuário tem a chance de detectar padrões na apresentação e obter informações sobre os dados. Esse tipo de reconhecimento de padrões é algo que a visão humana é conhecida por fazer notavelmente bem (SIIRTOLA; MÄKINEN, 2005).

A reordenação da matriz pode ser feita manualmente, onde o usuário permuta as linhas e colunas, ou de maneira automatizada, com a utilização de algoritmos que ordenam a matriz inteira ou apenas a linha ou coluna desejada. O sistema desenvolvido realiza apenas as reordenações automatizadas. A Figura 2.13 é uma visualização gerada pelo sistema desenvolvido neste trabalho. A matriz original de entrada (com dados aleatórios) está à esquerda e ao lado está a matriz reordenada com base na primeira coluna.

O princípio da reordenação visual automática é que as linhas e colunas são reordenadas de acordo com sua similaridade visual. Para isso, é necessário que o algoritmo de reordenação receba os dados normalizados e as representações visuais utilizadas devem garantir que as diferenças são aproximadamente proporcionais às diferenças numéricas. Deste modo, parece que o algoritmo está operando visualmente (PERIN; DRAGICEVIC; FEKETTE, 2014).

Figura 2.13: À esquerda, a matriz original, à direita, a matriz reordenada com base na primeira coluna.



Fonte: Autor.

## Capítulo 3

# Mineração de Dados Educacionais

De acordo com HAND (2007), a mineração de dados (MD) é a descoberta de estruturas interessantes, inesperadas ou valiosas em grandes conjuntos de dados. Um dos aspectos da MD diz respeito às estruturas "globais", de grande escala, com o objetivo de modelar as formas, ou características das formas, das distribuições. O outro diz respeito às estruturas "locais", de pequena escala, com o objetivo de detectar essas anomalias e decidir se são ocorrências reais ou aleatórias.

Esse processo tem sido utilizado por empresas, cientistas e governos, para auxílio na tomada de decisão, para entender melhor o mercado, clientes, ou mesmo o que os censos indicam. De forma geral, a mineração tende a facilitar a análise de diferentes conjuntos de dados.

De acordo com FEKETE et al. (2008), quando um conjunto de dados se torna muito grande para ser visualizado diretamente, uma automatização da análise é necessária para escolher os fatos mais interessantes. Ao explorar um conjunto de dados em busca de novas visões, a visualização da informação deve ser usada, possivelmente em conjunto de técnicas de mineração de dados, se o conjunto de dados for muito grande.

De acordo com D. A. KEIM (2002), para que a mineração de dados seja eficaz, é importante incluir o humano no processo de exploração de dados e combinar a flexibilidade, a criatividade e o conhecimento geral do ser humano com a capacidade de armazenamento e o poder computacional dos computadores. A exploração de dados visuais visa integrar o humano no processo de exploração de dados, aplicando suas habilidades perceptivas aos grandes conjuntos de dados disponíveis nos sistemas de computadores atuais.

De acordo com SCHEUER e MCLAREN (2012), a mineração de dados educacionais (EDM, do inglês *Educational Data Mining*) está preocupada em desenvolver, pesquisar e aplicar métodos computadorizados para detectar padrões em grandes conjuntos de dados educacionais - padrões que seriam difíceis ou impossíveis de analisar devido ao grande volume de dados que existem.

### 3.1 Pesquisas na área

Como já citado, a área de mineração e visualização estão intimamente ligadas, já que a primeira proporciona um conjunto de dados interessantes para o usuário e a segunda explora estes dados de uma maneira visual para facilitar o entendimento. A seguir, algumas pesquisas que relacionam EDM com a visualização são apresentadas.

De acordo com QU e QING (2015), MOOCs (*Massive Open Online Courses*, cursos livres online destinados à massa) possuem uma grande quantidade de dados, visto que é comum um curso ter mais de 10,000 inscritos. Os dados podem ser estruturados (como nome, gênero, data de nascimento) ou não-estruturados (como os textos dos fóruns), contendo informações espaciais (como a localidade do aluno) e temporais (como quando o aluno estudou e por quanto tempo).

Os usuários finais dos sistemas analíticos, como instrutores de cursos, pesquisadores em educação e estudantes, geralmente têm pouco ou nenhum conhecimento sobre técnicas de mineração de dados. Portanto, técnicas de visualização devem ser usadas para fornecer uma interface intuitiva, mas devem ser bem escolhidas de modo que os usuários finais possam facilmente compreendê-las sem muito treinamento (QU; QING, 2015).

QU e QING (2015) enfatizam que, para enfrentar estes desafios, são necessários métodos avançados de mineração de dados para revelar padrões de dados em MOOC e técnicas de visualização para transmitir os resultados analíticos aos usuários finais e permitir que eles explorem livremente os dados por si mesmos.

Para RITSOS e ROBERTS (2014), a análise e visualização de dados coletados por instituições educacionais têm o potencial de identificar os fatores que afetam os processos de ensino e aprendizado, possibilitando melhorá-los. A dinâmica de análise e visualização dos dados pode também ajudar os usuários conforme eles aprendem.

De acordo com RITSOS e ROBERTS (2014), a análise e a visualização oferecem aos usuários a capacidade de adquirir novas visões sobre como o processo de aprendizado pode ser melhorado através de ações, possibilita intervenções, autopercepção do usuário e, portanto, adaptação ao seu comportamento, determinação de seu desempenho e um melhor mapeamento do currículo.

Como exemplo, ASIF et al. (2017) realizaram uma análise sobre o desempenho de estudantes de graduação, usando EDM em conjunto de diferentes tipos de visualização. Uma das visualizações foi feita em forma de Heatmap, agrupando os estudantes como uma ajuda visual para verificar os indicadores de alto e baixo desempenho.

# Capítulo 4

## Desenvolvimento

O intuito deste trabalho é desenvolver uma ferramenta usando técnicas de visualização de informação para auxiliar coordenadores de curso. Para isso, o *Heatmap* foi escolhido como forma de visualização, bem como o meio online, por ser um ambiente mais acessível se comparado com programas de computador ou aplicativos de dispositivos móveis, onde o software acaba restringindo-se ao tipo de dispositivo (computador ou dispositivo móvel) e, dependendo da linguagem de programação, ao sistema operacional.

O *Heatmap* consegue representar três atributos do conjunto de dados: cada coluna do gráfico representa uma disciplina, cada linha representa um aluno e cada célula possui uma cor correspondente à nota do aluno na disciplina, conforme Figura 4.1.

Figura 4.1: Parte do Heatmap gerado pelos históricos escolares inseridos no sistema, com base em dados aleatórios.

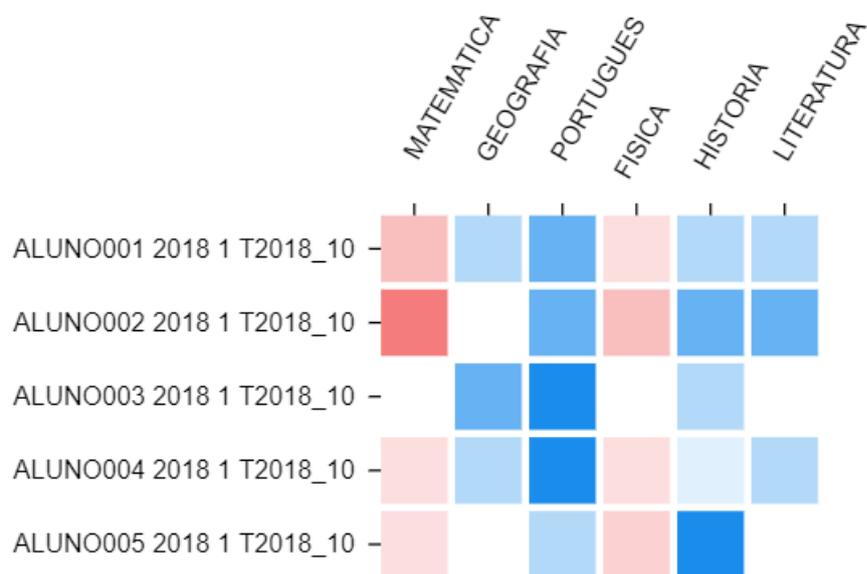


Fonte: Autor.

Este tipo de gráfico foi escolhido para possibilitar uma análise das disciplinas do curso com bases nos históricos escolares de algumas turmas de maneira visual, sendo assim um processo mais fácil e rápido quando comparado à análise textual dos dados.

Como exemplo, na Figura 4.2, vermelho representa uma nota ruim, branco representa uma nota média e azul uma nota boa. É visível que disciplinas que envolvem matemática causaram muitas reprovações dentre os alunos em questão e que Português teve as melhores notas.

Figura 4.2: Exemplo com dados fictícios no qual as disciplinas referentes à matemática causam mais reprovações.



Fonte: Autor.

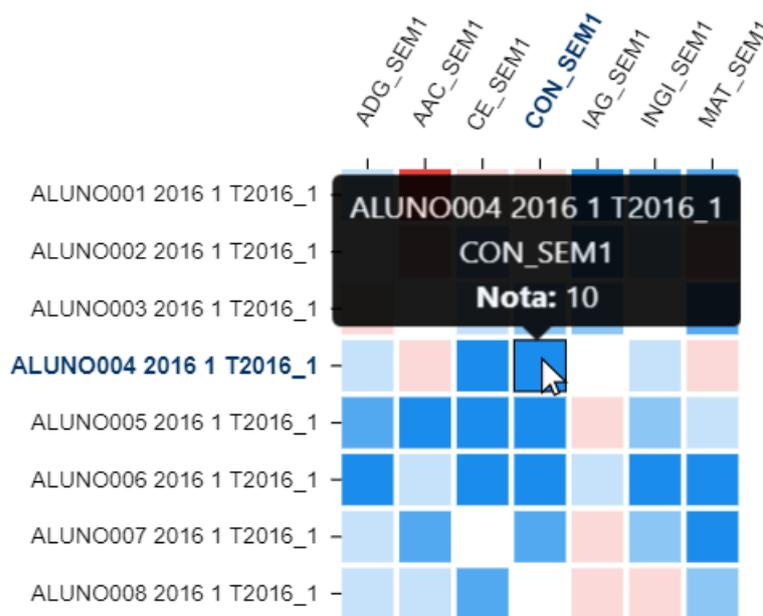
Para possibilitar uma melhor análise do gráfico, algumas funcionalidades foram adicionadas ao sistema com base no mantra de SHNEIDERMAN (2003). Conforme abordagens visuais começaram a ser exploradas, novas oportunidades surgiram, e então o *Visual Information-Seeking Mantra* (mais conhecido como mantra de Shneiderman) foi desenvolvido conforme uma análise das ações as quais os usuários podem executar em busca de determinadas informações.

De acordo com SHNEIDERMAN (2003), existem várias diretrizes visuais, porém o princípio básico pode ser resumido como: visão geral em primeiro lugar, seguido por *zoom* e filtro, e por fim detalhes sob demanda.

Com esse mantra em mente, o sistema foi desenvolvido de modo a possibilitar a filtragem de turmas para exibir apenas as desejadas, uma visão geral do gráfico através

da redução do zoom e um foco nos itens de interesse com a ampliação do zoom, e os detalhes sob demanda conforme o ponteiro do mouse é colocado sob alguma célula, como na Figura 4.3, onde a disciplina, aluno e nota são exibidos.

Figura 4.3: Detalhes sob demanda ao mover o ponteiro do mouse sob alguma célula do Heatmap.



Fonte: Autor.

Para deixar o sistema personalizável, também é possível escolher diferentes cores para as três principais notas (menor, média e maior), com valores padrão para caso o usuário não queira personalizar (vermelho para a nota 0, branco para a nota 6 e azul para a nota 10). A escala de cores é calculada de acordo com a biblioteca Chrome.js (AISCH, 2016).

A ferramenta escolhida para gerar o gráfico foi o D3.js, uma biblioteca JavaScript para manipulação de documentos com base em dados. O D3 auxilia na exibição dos dados em forma de gráficos utilizando HTML (para o conteúdo da página), SVG (para gráficos vetoriais), CSS (para a estética) e JavaScript (para interação) (OGIEVETSKY; HEER; BOSTOCK, 2011).

Para gerar o gráfico, entretanto, foi necessário um tratamento dos dados para o formato de entrada do D3.js, conforme explicado no Apêndice A.

Com uma ênfase em padrões da Web, oferecendo recursos dos navegadores modernos sem se vincular a um framework proprietário, combinando componentes de visualização e uma abordagem orientada a dados para manipulação do DOM (*Document Object*

*Model*, uma interface de programação que permite a alteração do documento HTML) (BOSTOCK, 2017).

O D3 foi escolhido pois, como já dito, foi construído utilizando tecnologias padrões da web, é gratuito e *open-source* (código-fonte aberto), com uma crescente comunidade de apoio e o código é executado no próprio navegador da Web, além de ser relativamente fácil de aprender, sendo inclusive considerado uma maneira prática de abordar a programação de computadores (JAIN, 2014).

Os algoritmos de ordenação aplicados no sistema foram desenvolvidos e disponibilizados por FEKETE (2015a), como uma biblioteca JavaScript chamada *Reorder.js*. A decisão por utilizar a biblioteca de Fekete foi tomada com o objetivo de otimizar o tempo disponível para o desenvolvimento do projeto sem se preocupar em desenvolver códigos já existentes e que estão disponíveis abertamente (FEKETE, 2015b).

Para conseguir realizar a reordenação do *Heatmap* com a utilização do *reorder.js*, foi necessário um tratamento de dados diferente do utilizado no *D3.js*. Esse tratamento está explicado no Apêndice B.

Para facilitar o desenvolvimento do design e da codificação em JavaScript, foram utilizados frameworks e bibliotecas como o Bootstrap, JQuery, Fontawesome e Bootstrap Colorpicker, que são amplamente conhecidos e utilizados para o desenvolvimento de sites.

O arquivo de entrada para gerar o gráfico teve o formato decidido com base numa discussão com o grupo de pesquisa SEIS. O arquivo possui extensão *.csv*, com três linhas representando metadados, seguidas pela linha de cabeçalho dos dados e então pelos dados em si.

Na primeira linha do arquivo é indicado o tipo de separador dos valores (em arquivos *.csv* pode variar entre vírgula ou ponto e vírgula). Na segunda linha, a quantidade de colunas a serem agrupadas para representar o aluno. Na terceira linha, o nome da coluna que representa a turma, para permitir a aplicação do filtro de turmas. Na quarta linha está o cabeçalho dos valores e as linhas restantes são os históricos dos alunos.

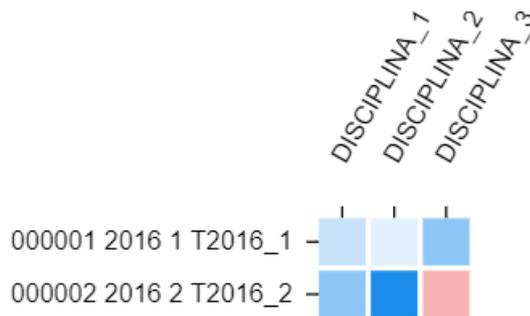
A Tabela 4.1 representa um exemplo do arquivo de entrada. Neste caso, o separador é o caractere ponto e vírgula, quatro colunas serão agrupadas para representar o aluno (RA, ano, semestre e turma), existem três disciplinas e as notas de dois alunos nestas disciplinas. O gráfico gerado com base neste arquivo pode ser visualizado na Figura 4.4.

```

;
4
TURMA
  RA   ; ANO ; SEMESTRE ; TURMA ; DISCIPLINA_1 ; DISCIPLINA_2 ; DISCIPLINA_3
000001 ; 2016 ; 1 ; T2016_1 ; 7 ; 6.5 ; 8
000002 ; 2016 ; 2 ; T2016_2 ; 8 ; 10 ; 4
    
```

Tabela 4.1: Exemplo de arquivo um possível arquivo de entrada .csv

Figura 4.4: Gráfico gerado com base no arquivo exemplificado.



Fonte: Autor.

Os gráficos exibidos são referentes à históricos escolares fictícios pois o foco deste trabalho é na área da educação, porém outros tipos de dados poderiam ser também representados com pequenas adaptações na interface da ferramenta, desde que o conjunto de dados esteja no formato adequado para a leitura.

# Capítulo 5

## Resultados

O sistema foi desenvolvido de acordo com o planejamento: a utilização do *Heatmap* para visualização dos históricos escolares; o meio Web por ser de mais fácil acesso; a possibilidade de personalização da legenda por parte do usuário; o filtro por turma; e a reordenação do gráfico.

Conforme Figura 5.1, o usuário indicará o arquivo .csv que possui os históricos escolares, podendo também escolher as cores de escala da legenda, bem como a menor, média e maior nota para a legenda, contando ainda com o filtro de turmas a serem exibidas no gráfico.

Figura 5.1: Arquivo de entrada de dados e personalização de cor e filtro de turmas.

**Arquivo de Dados**

Escolher arquivo Simulação.csv

▼ Legenda ?

Menor nota	<input type="text" value="1"/>	<input type="color" value="#800080"/>
Média	<input type="text" value="7"/>	<input type="color" value="#D3D3D3"/>
Maior nota	<input type="text" value="10"/>	<input type="color" value="#008080"/>

▼ Filtro por turma ?

- T2016\_1
- T2016\_2
- T2017\_1

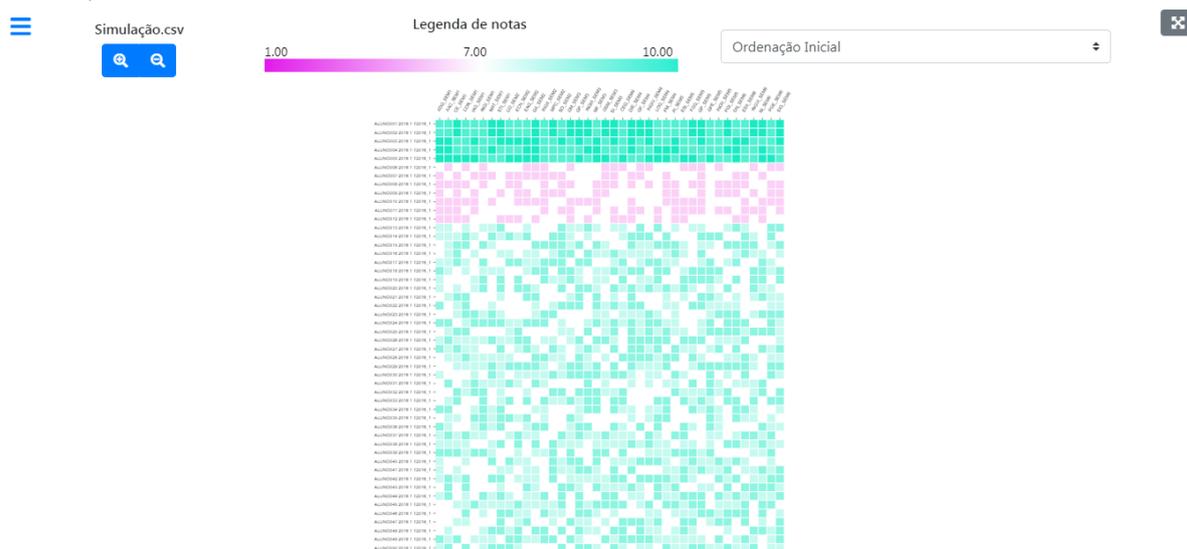
Fonte: Autor.

O usuário pode voltar para este menu a qualquer momento, seja para escolher outro conjunto de históricos escolares a ser visualizado, seja para filtrar turmas ou modificar a legenda. Quando o usuário fecha o menu, o sistema verifica se houve alguma mudança na entrada de dados (arquivo, legenda e filtro) para decidir se é necessário ou não exibir um novo gráfico.

Após o gráfico ser gerado, Figura 5.2, ele é exibido de maneira que seja possível ver todo o *Heatmap* sem precisar da rolagem da página, variando a sua escala entre 10% (menor *zoom* possível) e 100% (tamanho original). A escala da legenda é exibida em conjunto de outras opções para o usuário, como escolher uma ordenação, alterar o *zoom* ou deixar o sistema em modo tela cheia.

Uma vez que os alunos tenham um viés a ser mostrado, isso ficaria claro para o usuário na construção visual inicial, pois a visão geral permite uma identificação rápida do viés através das cores exibidas. Na Figura 5.2 é visível que existe um grupo de alunos no início do gráfico que foi muito bem em todas as matérias e, abaixo destes, existe outro grupo de alunos que não foi bem, tendo todas as notas iguais ou inferiores a sete, de acordo com a legenda.

Figura 5.2: Gráfico ao centro, *zoom* no canto superior esquerdo, legenda ao lado direito do *zoom*, seguida pela ordenação e opção de alternar o modo da tela (tela cheia ou modo normal).



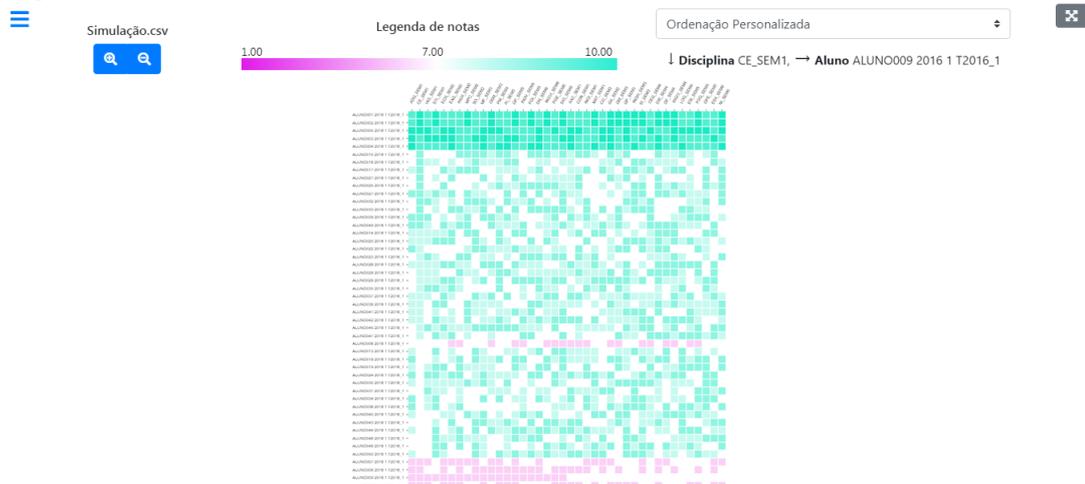
Fonte: Autor.

O sistema também permite a reordenação do gráfico conforme aluno ou disciplina escolhidos, indicando a possibilidade de ação ao usuário ressaltando o nome da disciplina

ou aluno conforme o mouse é posicionado em cima destes, além de mudar o cursor para *pointer*, que indica que aquele local é clicável.

Para informar ao usuário qual a ordenação atual, a seleção de ordenação no canto superior direito é atualizada para "Ordenação Personalizada", seguida por ícones de seta indicando se é uma ordenação ascendente ou descendente, o nome do aluno e disciplina, conforme Figura 5.3.

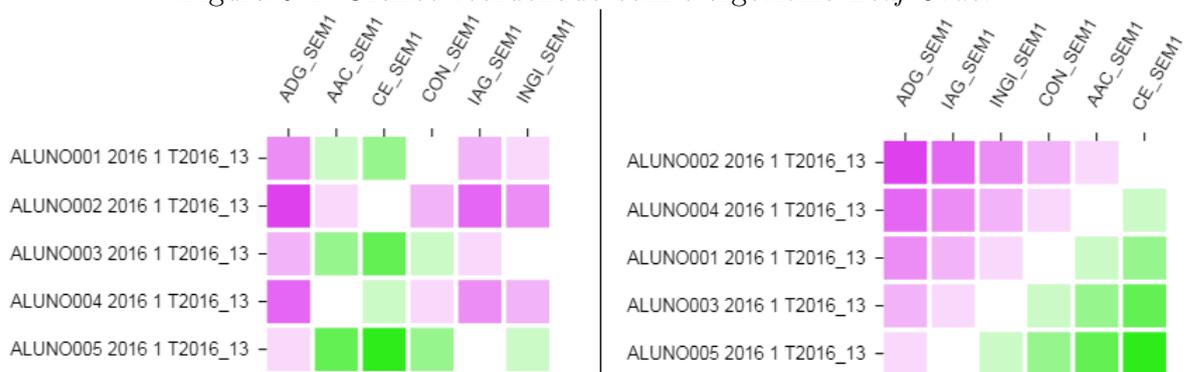
Figura 5.3: Gráfico ordenado por aluno e disciplina, informando abaixo da seleção de ordenação qual é o tipo de ordem corrente.



Fonte: Autor.

Na Figura 5.4 é possível visualizar à esquerda o gráfico original, e à direita o gráfico após a aplicação da reordenação *Leaf Order*, algoritmo disponibilizado por FEKETE (2015a). O conjunto de dados utilizado nesta situação é menor para facilitar a visualização das permutações de alunos e disciplinas realizadas.

Figura 5.4: Gráfico reordenado com o algoritmo *Leaf Order*.



Fonte: Autor.

Todas funcionalidades citadas foram implementadas com o objetivo de facilitar a interpretação dos históricos escolares por parte do coordenador de curso. O embasamento no mantra de SHNEIDERMAN (2003) foi importante para nortear o desenvolvimento da ferramenta, visto que esta não foi utilizada por coordenadores de curso para recepção de *feedback*.

Uma discussão sobre a ferramenta é válida e encorajada, bem como o uso deste em casos reais para verificar a real utilidade desta e elencar melhorias para a mesma, como funções novas ou conversões na entrada de dados para o modelo de arquivo que é interpretado, abrangendo um número maior de universidades ou institutos que poderiam utilizar a ferramenta desenvolvida.

Algumas funcionalidades que podem ser incrementadas são: identificação e agrupamento das disciplinas de acordo com suas similaridades (por exemplo, disciplinas relacionadas à matemática, física, sociologia, programação); exibir informações referentes à disciplina ao selecioná-la; identificar e informar ao usuário sobre disciplinas em que houve um número de reprovações acima do normal; novas reordenações.

As principais dificuldades encontradas no desenvolvimento foram relativas à falta de exemplos implementados sobre o *Heatmap* de uma matriz com dados diferentes nas linhas e colunas. Os exemplos encontrados foram em grande maioria sobre matrizes de correlação.

A integração dos algoritmos entre os frameworks d3.js responsável pela geração de gráficos, e o reorder.js usado para a reordenação, mostrou-se complexa, porém o ganho dado os recursos dos mesmos se mostrou satisfatório.

# Capítulo 6

## Conclusão

Foi criada uma ferramenta que realiza a leitura de históricos escolares de acordo com um formato específico, explicado no Capítulo 4, e gera uma visualização em forma de *Heatmap*, permitindo uma percepção visual melhor e mais rápida sobre os históricos em questão, quando comparada à análise textual.

Com base no mantra de SHNEIDERMAN (2003), a ferramenta possibilita a filtragem de turmas, visão geral e específica do gráfico com a manipulação do zoom, detalhes sob demanda quando o ponteiro do mouse é posicionado sobre um quadrado do gráfico que representa a nota e reordenação do gráfico. Essas funcionalidades foram desenvolvidas com o objetivo de melhorar a experiência do coordenador, facilitando a observação dos dados.

A utilização da ferramenta desenvolvida é necessária para ser possível afirmar se o gráfico escolhido, bem como os recursos implementados, auxilia em uma melhor análise dos históricos escolares para os coordenadores. Outras funcionalidades podem ser adicionadas ao sistema conforme demanda em trabalhos futuros, com o objetivo de suprir necessidades ainda possivelmente não contempladas de modo a torná-lo mais completo e eficiente como ferramenta.

Algumas funcionalidades que podem ser desenvolvidas e agregariam valor à ferramenta foram citadas no Capítulo 5, sendo elas: identificação e agrupamento de disciplinas conforme similaridades; exibir informações referentes à disciplina ao selecioná-la; identificar e informar ao usuário sobre disciplinas em que houve um número de reprovações acima do normal; novas reordenações; formatação de outros arquivos para o formato necessário para gerar o gráfico.

## Referências bibliográficas

AISCH, G. *Chroma.js*. [S.l.: s.n.], 2016. Disponível em: <<https://github.com/gka/chroma.js/>>. Acesso em: 3 mar. 2019.

ASIF, R. et al. Analyzing undergraduate students' performance using educational data mining. *Elsevier*, v. 113, p. 177–194, out. 2017. ISSN: 0360-1315. DOI: 10.1016/j.compedu.2017.05.007. Disponível em: <<https://doi.org/10.1016/j.compedu.2017.05.007>>. Acesso em: 26 maio 2019.

BOSTOCK, M. *Data-Driven Documents*. [S.l.: s.n.], 2017. Disponível em: <<https://d3js.org/>>. Acesso em: 21 nov. 2018.

CATALOGUE, T. D. V. *Parallel Coordinates Plot*. [S.l.: s.n.]. [https://datavizcatalogue.com/methods/parallel\\_coordinates.html](https://datavizcatalogue.com/methods/parallel_coordinates.html). Acesso em: 26 maio 2019.

CAVANAGH, S. *Amazon, Apple, Google, and Microsoft Battle for K-12 Market, and Loyalties of Educators*. [S.l.], maio 2017. Disponível em: <<https://marketbrief.edweek.org/special-report/amazon-apple-google-and-microsoft-battle-for-k-12-market-and-loyalties-of-educators/>>. Acesso em: 26 out. 2018.

CISCO. *The Zettabyte Era: Trends and Analysis*. [S.l.], jun. 2017. Disponível em: <<https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html>>. Acesso em: 25 out. 2018.

COX, T. F.; COX, M. A. A. *Multidimensional Scaling*. 2. ed. [S.l.]: CRC Press, set. 2000. p. 1–2.

DECKER, S. *U.S. Education in 2017 in 10 Charts*. [S.l.: s.n.], dez. 2017. Disponível em: <<https://www.edweek.org/ew/section/multimedia/us-education-in-2017-in-10-charts.html>>. Acesso em: 26 out. 2018.

- EDUCATION WEEK. *Quality Counts 2017: State Report Cards Map*. v. 36. [S.l.: s.n.], dez. 2016. Disponível em: <<https://www.edweek.org/ew/qc/2017/2017-state-education-grades-map.html>>. Acesso em: 26 out. 2018.
- FEKETE, J.-D. *Reorder.js*. [S.l.]: GitHub, 2015. Disponível em: <<https://github.com/jdfekete/reorder.js>>. Acesso em: 26 maio 2019.
- *Reorder.js: A JavaScript Library to Reorder Tables and Networks*, out. 2015. Disponível em: <<https://hal.inria.fr/hal-01214274>>. Acesso em: 26 maio 2019.
- FEKETE, J.-D. et al. The Value of Information Visualization. *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, v. 4950, p. 1–18, 2008. DOI: 10.1007/978-3-540-70956-5\_1. Disponível em: <[https://doi.org/10.1007/978-3-540-70956-5\\_1](https://doi.org/10.1007/978-3-540-70956-5_1)>. Acesso em: 20 nov. 2018.
- GOOGLE. *Assunto: Exame Nacional do Ensino Médio (ENEM)*. [S.l.: s.n.], 2018. Disponível em: <<https://trends.google.com/trends/explore?date=2016-03-01%5C%202018-01-31&geo=BR&q=%5C%2Fm%5C%2F047v0j1>>. Acesso em: 27 out. 2018.
- HAND, D. J. Principles of Data Mining. *Drug Safety*, v. 30, p. 621–622, jul. 2007. DOI: 00002018-200730070-00010. Disponível em: <<https://doi.org/10.2165/00002018-200730070-00010>>. Acesso em: 9 nov. 2018.
- INSELBERG, A.; DIMSDALE, B. Parallel Coordinates: A Tool for Visualizing Multi-dimensional Geometry. In: *Proceedings of the 1st Conference on Visualization '90*. San Francisco, California: IEEE Computer Society Press, 1990. (VIS '90), p. 361–378. ISBN: 0-8186-2083-8. Disponível em: <<http://dl.acm.org/citation.cfm?id=949531.949588>>.
- INSTITUTE FOR HEALTH METRICS AND EVALUATION. *GBD Compare | Viz Hub*. [S.l.: s.n.], 2017. Disponível em: <<https://vizhub.healthdata.org/gbd-compare/heatmap>>. Acesso em: 13 nov. 2018.
- INTERNET LIVE STATS. [S.l.: s.n.], out. 2018. Disponível em: <<http://www.internetlivestats.com/one-second/#traffic-band>>. Acesso em: 25 out. 2018.
- JAIN, A. Data Visualization with the D3.JS Javascript Library. *J. Comput. Sci. Coll.*, Consortium for Computing Sciences in Colleges, USA, v. 30, n. 2, p. 139–141, dez. 2014. ISSN: 1937-4771. Disponível em: <<http://dl.acm.org/citation.cfm?id=2667432.2667451>>. Acesso em: 21 nov. 2018.

- KEIM, D. A. Information Visualization and Visual Data Mining. *IEEE Transactions on Visualization & Computer Graphics*, v. 8, n. 1, p. 1–8, jan. 2002. DOI: 10.1109/2945.981847. Disponível em: <<https://doi.ieeecomputersociety.org/10.1109/2945.981847>>. Acesso em: 20 nov. 2018.
- MAZZA, R. *Introduction to Information Visualization*. [S.l.]: Springer, London, 2009. cap. 1,3. ISBN: 978-1-848002-18-0. DOI: 10.1007/978-1-84800-219-7. Disponível em: <<https://doi.org/10.1007/978-1-84800-219-7>>. Acesso em: 20 ago. 2018.
- MCAFEE, A.; BRYNJOLFSSON, E. Big Data: The Management Revolution. *Harvard Business Review*, v. 90, n. 10, p. 59–68, out. 2012. Disponível em: <<https://hbr.org/2012/10/big-data-the-management-revolution>>. Acesso em: 6 set. 2018.
- OGIEVETSKY, V.; HEER, J.; BOSTOCK, M. D<sup>3</sup> Data-Driven Documents. *IEEE Transactions on Visualization & Computer Graphics*, v. 17, p. 2301–2309, set. 2011. ISSN: 1077-2626. DOI: 10.1109/TVCG.2011.185. Disponível em: <<https://doi.ieeecomputersociety.org/10.1109/TVCG.2011.185>>. Acesso em: 21 nov. 2018.
- PERIN, C.; DRAGICEVIC, P.; FEKETTE, J.-D. Revisiting Bertin Matrices: New Interactions for Crafting Tabular Visualizations. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, v. 20, n. 12, p. 2082–2091, dez. 2014. ISSN: 1941-0506. DOI: 10.1109/TVCG.2014.2346279. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/6875988>>. Acesso em: 25 maio 2019.
- QU, H.; QING, C. Visual Analytics for MOOC Data. *IEEE Computer Graphics and Applications*, v. 35, p. 69–75, nov. 2015. DOI: 10.1109/MCG.2015.137. Disponível em: <<https://doi.org/10.1109/MCG.2015.137>>. Acesso em: 20 nov. 2018.
- [S.l.]. *Towards more Visual Analytics in Learning Analytics*. Edição de M. Pohl e J. Roberts. [S.l.]: The Eurographics Association, 2014. ISBN: 978-3-905674-68-2. DOI: 10.2312/eurova.20141147. Disponível em: <<http://dx.doi.org/10.2312/eurova.20141147>>. Acesso em: 20 nov. 2018.
- SCHEUER, O.; MCLAREN, B. M. Educational Data Mining. In: — . *Encyclopedia of the Sciences of Learning*. Edição de Norbert M. Seel. Boston, MA: Springer US, 2012. p. 1075–1079. ISBN: 978-1-4419-1428-6. DOI: 10.1007/978-1-4419-1428-6\_618. Disponível em: <[https://doi.org/10.1007/978-1-4419-1428-6\\_618](https://doi.org/10.1007/978-1-4419-1428-6_618)>. Acesso em: 9 nov. 2018.

SHNEIDERMAN, B. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. *Institute for Systems Research*, p. 364–371, 2003. DOI: 10.1016/B978-155860915-0/50046-9. Disponível em: <<https://doi.org/10.1016/B978-155860915-0/50046-9>>. Acesso em: 26 maio 2019.

SIIRTOLA, H.; MÄKINEN, E. Constructing and reconstructing the reorderable matrix. *Palgrave Macmillan Ltd.*, p. 32–48, 2005. ISSN: 1473-8716. Disponível em: <<https://journals.sagepub.com/doi/pdf/10.1057/palgrave.ivs.9500086>>. Acesso em: 25 maio 2019.

WARD, M.; GRINSTEIN, G.; KEIM, D. *Interactive Data Visualization: Foundations, Techniques, and Applications*. [S.l.]: A K Peters, Ltd., 2010. ISBN: 978-1-568814-73-5.

WEINSTEIN, J. N. A Postgenomic Visual Icon. *Science*, v. 319, p. 1772–1773, mar. 2008. DOI: 10.1126/science.1151888. Disponível em: <<https://doi.org/10.1126/science.1151888>>. Acesso em: 10 nov. 2018.

# Apêndice A

## Formato de entrada para o D3.js

Para conseguir exibir o *Heatmap* utilizando o D3.js, foi necessário realizar um tratamento dos dados para um formato compatível com o D3. O formato final, Figura A.1, é uma lista (*array*) de objetos, onde cada objeto possui um atributo indicando a linha correspondente (*row*), a coluna (*column*) e o valor (*value*), que no conjunto de dados tratado neste trabalho representa a nota.

Figura A.1: Formato recebido pelo D3.js.

```
▶0: {row: 0, column: 0, value: "3"}
▶1: {row: 0, column: 1, value: "7"}
▶2: {row: 0, column: 2, value: "8"}
▶3: {row: 0, column: 3, value: "6"}
▶4: {row: 0, column: 4, value: "4"}
▶5: {row: 0, column: 5, value: "5"}
▶6: {row: 1, column: 0, value: "1"}
▶7: {row: 1, column: 1, value: "5"}
▶8: {row: 1, column: 2, value: "6"}
▶9: {row: 1, column: 3, value: "4"}
▶10: {row: 1, column: 4, value: "2"}
▶11: {row: 1, column: 5, value: "3"}
▶12: {row: 2, column: 0, value: "4"}
▶13: {row: 2, column: 1, value: "8"}
▶14: {row: 2, column: 2, value: "9"}
▶15: {row: 2, column: 3, value: "7"}
▶16: {row: 2, column: 4, value: "5"}
▶17: {row: 2, column: 5, value: "6"}
▶18: {row: 3, column: 0, value: "2"}
▶19: {row: 3, column: 1, value: "6"}
▶20: {row: 3, column: 2, value: "7"}
▶21: {row: 3, column: 3, value: "5"}
▶22: {row: 3, column: 4, value: "3"}
▶23: {row: 3, column: 5, value: "4"}
▶24: {row: 4, column: 0, value: "5"}
▶25: {row: 4, column: 1, value: "9"}
▶26: {row: 4, column: 2, value: "10"}
▶27: {row: 4, column: 3, value: "8"}
▶28: {row: 4, column: 4, value: "6"}
▶29: {row: 4, column: 5, value: "7"}
```

Fonte: Autor.

# Apêndice B

## Formato de entrada para o Reorder.js

Para conseguir exibir o *Heatmap* utilizando o Reorder.js, foi necessário realizar um tratamento dos dados para um formato compatível com o método de reordenação que foi utilizado (*optimal\_leaf\_order*, referente ao algoritmo *Leaf Order*). Este método recebe duas matrizes de distância, uma matriz referente às linhas e outra às colunas, conforme Figura B.1.

Figura B.1: Matrizes de distância utilizadas pelo Reorder.js para o *Leaf Order*.

```
▶0: (5) [0, 4.898979485566356, 2.449489742783178, 2.449489742783178, 4.898979485566356]
▶1: (5) [4.898979485566356, 0, 7.3484692283495345, 2.449489742783178, 9.797958971132712]
▶2: (5) [2.449489742783178, 7.3484692283495345, 0, 4.898979485566356, 2.449489742783178]
▶3: (5) [2.449489742783178, 2.449489742783178, 4.898979485566356, 0, 7.3484692283495345]
▶4: (5) [4.898979485566356, 9.797958971132712, 2.449489742783178, 7.3484692283495345, 0]

▶0: (6) [0, 8.94427190999916, 11.180339887498949, 6.708203932499369, 2.23606797749979, 4.47213595499958]
▶1: (6) [8.94427190999916, 0, 2.23606797749979, 2.23606797749979, 6.708203932499369, 4.47213595499958]
▶2: (6) [11.180339887498949, 2.23606797749979, 0, 4.47213595499958, 8.94427190999916, 6.708203932499369]
▶3: (6) [6.708203932499369, 2.23606797749979, 4.47213595499958, 0, 4.47213595499958, 2.23606797749979]
▶4: (6) [2.23606797749979, 6.708203932499369, 8.94427190999916, 4.47213595499958, 0, 2.23606797749979]
▶5: (6) [4.47213595499958, 4.47213595499958, 6.708203932499369, 2.23606797749979, 2.23606797749979, 0]
```

Fonte: Autor.