

# Estratégias de Visualização de Dados para auxiliar profissionais da área médica em diagnósticos

Gabriel de P. Félix

Estudante de Ciência da Computação - UFV

Natan S. Garcias

Estudante de Ciência da Computação - UFV

## ABSTRACT

This article addresses the "Challenge One" of the BioVis 2020 conference, where visualization strategies must be created to assist medical professionals in making diagnoses. The available datasets are about families with a high incidence of suicide, living in Utah (USA). Four tasks are proposed in the challenge, so that the visualization strategy to be created meets the criteria of individual analysis of family members and also analyzes the distribution of clinical attributes. Two strategies have been proposed to solve them: Family Trees and Bar Charts. The first deals with individual visualizations and comparisons and the second covers the distribution of clinical attributes in the families studied. Brushing techniques and information on demand were applied for a better visualization.

**Index Terms:** Family Trees, Bar Graphs, BioVis 2020, Brushing and filters;

## 1 INTRODUÇÃO

A saúde de uma pessoa pode ser afetada negativamente por diversos fatores, como por exemplo fatores sociais e hereditários. Tais impactos negativos podem contribuir para o surgimento ou engrandecimento de problemas psicológicos, e até mesmo levar as pessoas a cometerem suicídio. A análise destes fatores torna-se difícil a medida em que o número de indivíduos de uma família aumenta, assim como os seus respectivos atributos clínicos disponíveis para estudo.

Ferramentas de visualização de dados podem auxiliar profissionais da saúde a obter resultados importantes durante a análise. Neste ano, o desafio da conferência *BioVis 2020* [1] trata sobre a criação de estratégias de visualização para aplicar em dados de famílias com altos índices de suicídio.

Para este desafio, times de biomédicos da Universidade de Utah (EUA) foram reunidos para analisar dados multidimensionais de nove famílias locais. São apresentadas quatro tarefas para a visualização, contando com análises de casos de suicídio semelhantes e também de distribuição de atributos clínicos. Foram propostos dois tipos de visualização diferentes: Árvores Genealógicas e Gráficos de Barras, de forma que tanto a análise individual quanto as de distribuição de atributos sejam feitas de forma precisa e facilitada.

## 2 DESCRIÇÃO DOS DATASETS

Foram disponibilizados para o desafio dois datasets: *Family Attributes* e *Family Structure*, referentes à nove famílias da cidade de Utah (EUA).

O dataset *Family Attributes* contém informações de atributos clínicos como a presença de depressão, problemas de personalidade, obesidade, entre outros. Ademais, quando presente, há a descrição de como cada suicídio ocorreu. O dataset *Family Structure* apresenta informações de parentesco de cada indivíduo e informações sobre seu nascimento/falecimento.

Nos dois datasets existem casos especiais como indivíduos com IDs repetidos de e ausência de informações de parentesco. Porém, ainda que com avarias, a estrutura geral das famílias não foi afetada (elas descrevem uma árvore genealógica conexa).

## 3 DESCRIÇÃO DO DESAFIO

Neste desafio, são apresentadas quatro tarefas para a ferramenta de visualização a ser criada. Sendo estas:

- Para um determinado indivíduo-alvo, identifique casos semelhantes, incluindo como eles estão relacionados entre si (por exemplo, se eles co-ocorrem em uma determinada família).
- Caracterizar a distribuição de atributos clínicos para casos de suicídio em famílias com altas taxas de incidência (alto número relativo de casos).
- Caracterizar (ou seja, a relação entre os casos e seus atributos) casos de suicídio em famílias com alta incidência de um determinado atributo clínico (como depressão).
- Compare as informações clínicas para casos de suicídio com seus parentes imediatos (irmãos, pais e filhos).

Em síntese, são necessárias estratégias que permitam uma análise geral do quadro de atributos e de famílias, e também outra análise mais individual para identificação de casos semelhantes e comparação entre parentes imediatos.

## 4 ABORDAGENS

Foram propostas duas abordagens para o problema, utilizando Árvores Genealógicas e Gráficos de Barras. As duas fizeram o uso da biblioteca D3 do JavaScript [2].

A primeira, foi desenvolvida para solucionar as tarefas (1) e (4), das quais são requisitadas análises de casos semelhantes em uma família e também comparar casos de suicídio com seus parentes próximos.

A segunda, foi elaborada para solucionar as tarefas (2) e (3), que requerem a caracterização da distribuição de atributos clínicos entre famílias e também a distribuição destes em uma mesma família.

### 4.1 Árvore Genealógica

Árvores genealógicas são muito úteis para visualizar a estrutura geral de uma família e também explorar individualmente seus integrantes. Famílias quando analisadas somente em um contexto geral podem ocultar informações importantes para auxiliar profissionais em diagnósticos médicos. Tendo em vista estes benefícios, foi proposta uma visualização em árvore para cada família, contando com as seguintes características:

- **Filtragem de atributos clínicos** específicos, onde os nodos filtrados são incrementados em sua opacidade e diâmetro do círculo (Figura 1).
- **Interatividade**, apresentando características e atributos individuais ao deslizar com o mouse sobre os nodos (Figura 2).

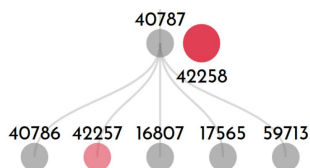


Figure 1: *Brushing* em nodo capturado pela filtragem de atributos.

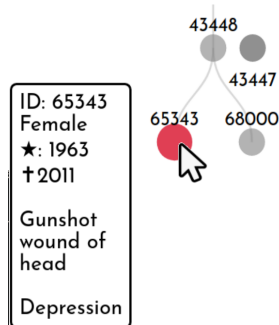


Figure 2: Informações *on demand* em nodo com o deslizar do mouse.

Nesta representação, um indivíduo é representado por um nodo de cor cinza, e quando este possui um cônjuge, é então acompanhado de um nodo extra ao seu lado direito. Caso o indivíduo cometeu suicídio, o nodo possuirá cor vermelha.

Na Figura 3, é apresentada a visualização para "família nove", onde em (1) é possível selecionar os filtros para os atributos clínicos, de forma que caso a pessoa contenha ao menos um destes atributos será então destacada na árvore. Em (2) é apresentada uma breve descrição técnica da família, indicando erros do dataset ou relatando demais indivíduos não apresentados na árvore (como a presença múltiplos cônjuges, limitada pela visualização). Por fim, em (3) é apresentada a árvore genealógica em si para a família selecionada.

## 4.2 Gráfico de Barras

Gráficos de barras são altamente recomendáveis quando se precisa comparar dados, a partir deles nossa cérebro consegue ter uma melhor percepção das diferenças entre os dados apresentados. Ademais, sua implementação permite uma análise de distribuições em geral de maneira fácil e intuitiva.

Seguindo estas vantagens, esta visualização foi aplicada na análise de distribuição de atributos clínicos em uma mesma família e também para analisar a presença destes atributos em todas as famílias, servindo como solução para as tarefas (2) e (3), respectivamente.

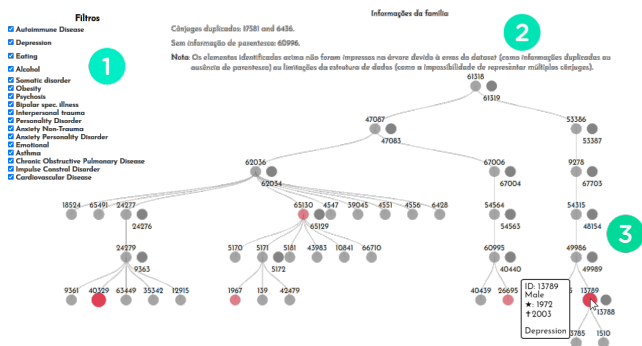


Figure 3: Esquema geral da visualização em Árvores Genealógicas.

Família 3 - Incidência de atributos clínicos em casos de suicídio

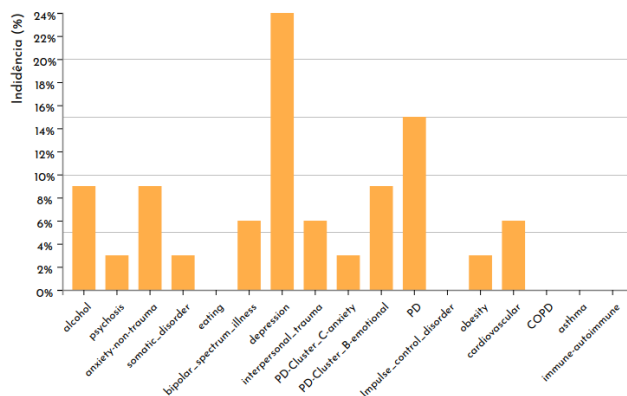


Figure 4: Gráfico que ilustra a presença de atributos clínicos para casos de suicídio de uma mesma família.

Depression - Incidência em casos de suicídio

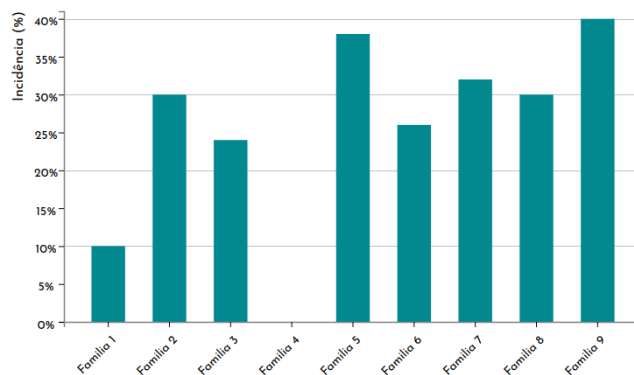


Figure 5: Gráfico que ilustra a incidência do atributo depressão nos casos de suicídio das famílias.

mente.

A primeira aplicação de Gráficos de Barras trata-se de gráficos por família, onde são apresentadas a porcentagem de incidência de atributos clínicos em casos de suicídio da família. Dessa forma, além de possibilitar a análise da distribuição de atributos clínicos em uma família, é possível identificar os mais presentes e também possíveis tendências (Figura 4).

A segunda aplicação trata-se de gráficos por atributos clínicos, que apresentam a porcentagem de incidência de um atributo clínico nos casos de suicídio de todas as famílias do dataset. Dessa forma, permite-se ao usuário visualizar quais atributos clínicos mais afetam uma família, além de possibilitar a identificação de atributos hereditários (Figura 5).

## 5 WEBSITE DA FERRAMENTA

As visualizações aqui apresentadas foram publicadas em um website e podem ser acessadas por [este link](#). Na página inicial, o tema é apresentado após uma suavização devido à delicadeza do assunto tratado. Em "Desafios & Abordagens" são descritos estes dois itens de forma detalhada, e os datasets estão presentes na página "Datasets". Por fim, as ferramentas apresentadas neste artigo fazem parte das páginas "Árvores Genealógicas" e "Gráficos".

## 6 CONCLUSÕES

As técnicas de visualização implementadas neste trabalho foram divididas em duas partes principais: Árvores Genealógicas e Gráficos

de Barras, de modo que uma funciona de maneira complementar à outra, possibilitando uma análise tanto individual quanto geral dos membros de cada família. Dessa forma, agora é possível explorar de forma rápida e eficiente as características das famílias, concluindo os objetivos de cada tarefa do desafio. Para trabalhos futuros, é tido como objetivo a criação de gráficos em que há a classificação de indivíduos pelo sexo, além de permitir a fragmentação das árvores em partes menores.

## REFERENCES

- [1] IEEE VIS 2020. Disponível em <http://biovis.net/2020/ieevvis/>, acesso em 15 de Outubro de 2020.
- [2] Data-Driven Documents - D3. Disponível em <https://d3js.org/>, acesso em 15 de Outubro de 2020.