

## REPRESENTAÇÃO DE DADOS MULTIVARIADOS ATRAVÉS DAS FACES DE CHERNOFF

Maurício Scagliante Bomtempo<sup>1</sup>

### RESUMO

Esse artigo aborda a representação gráfica como valioso instrumento de apoio na avaliação de fatores, variáveis e fenômenos observáveis, concentrando-se num modelo especial, que procura superar a restrição de representação de dados multidimensionais, sem perda da veracidade e da facilidade de compreensão: apresenta as Faces de Chernoff, ilustra sua aplicação em alguns trabalhos publicados, discorre sobre suas características e particularidades, expõe suas limitações e críticas identificadas em revisão bibliográfica, e aborda aspectos de sua construção – desde a associação dos atributos da face às variáveis do estudo, até alguns dos recursos computacionais disponíveis. O artigo conclui destacando sua conveniência em análises de dados multivariados, estimulando sua exploração.

**PALAVRAS-CHAVE:** Gráficos estatísticos. Representação de dados multivariados. Faces de Chernoff.

### INTRODUÇÃO

A apresentação de dados estatísticos através de gráficos é tema relativamente recente, quando comparado com a evolução da geometria, álgebra e teoria das probabilidades. O desenvolvimento de teorias estatísticas e a evolução computacional, especialmente a partir da popularização da microinformática, abriram espaço para o crescimento da representação gráfica como um poderoso recurso de análise de dados e transmissão de informação, já que a manipulação de alta quantidade de informação tem sido o grande desafio a partir da disponibilidade que oferecem.

---

<sup>1</sup> Mestre em Administração de Empresas (FECAP) e Bacharel em Estatística (Univ. Federal de São Carlos). Gerente de Informática da FEBRABAN – Federação Brasileira de Bancos e professor de Estatística em cursos de MBA da Estácio UniRadial.

---

William Playfair, considerado atualmente um dos pais dos gráficos estatísticos, expressou em 1801 que “nenhum estudo é menos sedutor ou mais tedioso que a Estatística”, e desenvolveu representações visuais de dados para torná-los um pouco mais “palatáveis” (FIENBERG, 1979).

Desde então, os gráficos vêm sendo utilizados intensamente, em publicações científicas, jornais e revistas populares, e na comunicação em geral. O “apelo visual” cria uma atmosfera favorável para a leitura, análise e interpretação dos dados.

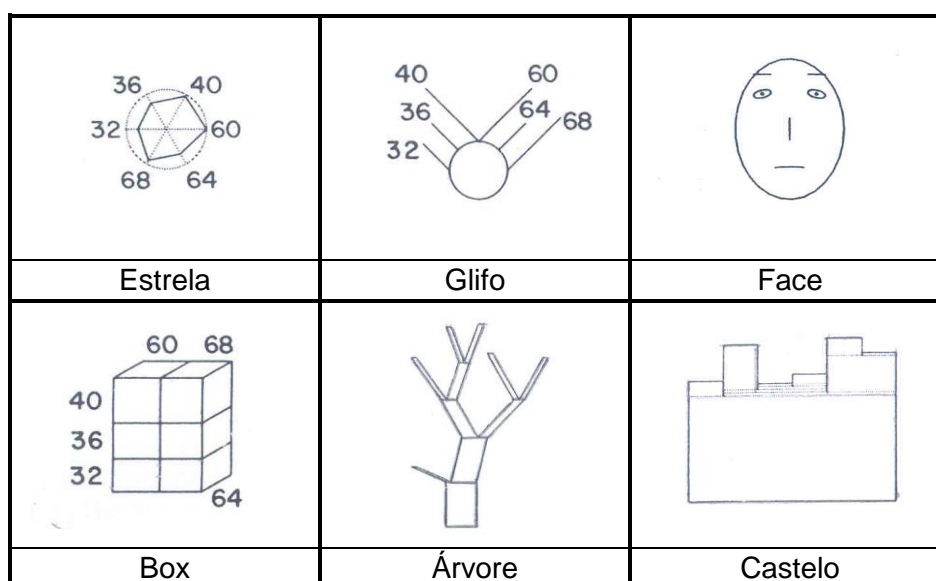
Gráficos atendem a diversas finalidades: são apresentados para sintetizar informações, apoiar na análise de dados, exemplificar conclusões, mostrar relacionamentos de variáveis, detectar padrões, identificar grupos e comportamentos, diagnosticar. Permitem visualmente a compreensão de fenômenos e fatos que muitas vezes ficariam obscurecidos apenas com a representação numérica, facilitam a memorização, estimulam a investigação e o pensamento científicos. Em especial, servem para o tratamento de grande volume de dados, como mecanismo de extração do conhecimento (abordagens de Visualização da Informação e Mineração de Dados).

## **A REPRESENTAÇÃO DE DADOS MULTIVARIADOS**

Há, contudo, um desafio especial quando se avaliam dados multivariados. A diversidade de variáveis dificulta uma representação no espaço no máximo tridimensional que estamos acostumados.

Com foco nessa necessidade, surgiram na década de 1970, diversas propostas de representação visual, compondo um grupo de técnicas denominado “iconográficas”. Sua propriedade fundamental é o mapeamento das variáveis em características particulares de ícones. Cada elemento do ícone representa um atributo dos dados multidimensionais, de forma tal que a aparência geral de cada objeto, criada a partir das diferentes configurações, atribui-lhe uma identidade visual que pode ser percebida pelo observador, facilitando o agrupamento de indivíduos e a descoberta de relações e interações entre variáveis.

Alguns dos ícones desenvolvidos com essa finalidade foram: Estrelas, Glifos, Faces, Boxes, Árvores e Castelos (KLEINER e HARTIGAN, 1981). A Figura 1 apresenta alguns exemplos dessas representações.



**Figura 1:** Exemplos de representações iconográficas, indicando o percentual de votos republicanos em seis eleições presidenciais (1932, 1936, 1940, 1960, 1964, 1968) no estado do Missouri - EUA

Fonte: Elaborado a partir de Kleiner e Hartigan (1981)

Dentro desse grupo, uma representação em especial será abordada nesse artigo: as Faces de Chernoff.

## AS FACES DE CHERNOFF

Herman Chernoff (1973) apresentou de forma inovadora um método de representação de dados multivariados, associando cada ponto de um espaço k-dimensional a um esboço de face humana, e cada dimensão representada por uma característica da face.

Assim, por exemplo, num estudo comparativo sobre a cobrança de tarifas em instituições financeiras, cada instituição poderia ser representada por uma face, e cada tarifa por um elemento da face: valor para renovação de cadastro pela *inclinação da sobrancelha*, valor de um saque em caixa eletrônico pelo *tamanho dos olhos*, custo de emissão de extrato pelo *tamanho do nariz*, tarifa para realização de uma transferência de valor representada pela *curvatura da boca*, fornecimento de uma folha de cheque pelo *formato do rosto*, e assim sucessivamente. A composição final de cada face permite avaliar no âmbito geral a posição de cada instituição.

O artigo em que a proposta é exposta originalmente ilustra sua aplicação com exemplos onde a nova representação visual é usada para criar grupos por semelhança do conjunto geral de variáveis.

O autor afirma que esse método gráfico “fornece uma promissora abordagem para uma observação inicial de dados multivariados, eficiente em revelar relações menos complexas nem sempre visíveis a partir de correlações simples baseadas em teorias lineares bi-dimensionais”. De fato, o uso de faces permite observar de maneira holística um conjunto muitas vezes complexo de variáveis, facilitando o reconhecimento de relações ou padrões entre os elementos.

A representação por meio de faces pressupõe a habilidade da mente humana em identificar e assimilar rapidamente uma multiplicidade de variações faciais, e conseguir criar agrupamentos consideradas as diversidades. Ou seja, o ser humano apresenta uma alta sensibilidade às variações nas expressões faciais. Nas palavras de Chernoff: “as pessoas crescem estudando e reagindo às faces o tempo todo”, e “pequenas diferenças são facilmente detectadas e evocam reações emocionais de um longo catálogo armazenado na memória”. Subconscientemente, a mente humana filtra as feições visuais e foca-se nas potencialmente importantes.

O fato da representação ser expressa por caricaturas da face, e portanto por imagens pouco realísticas, não indicam, para Chernoff, uma grande perda na capacidade de identificação e sensibilização, podendo potencialmente aprimorar a habilidade dos usuários em detectar e compreender fenômenos importantes e servir como um recurso mnemônico para recordar conclusões principais.

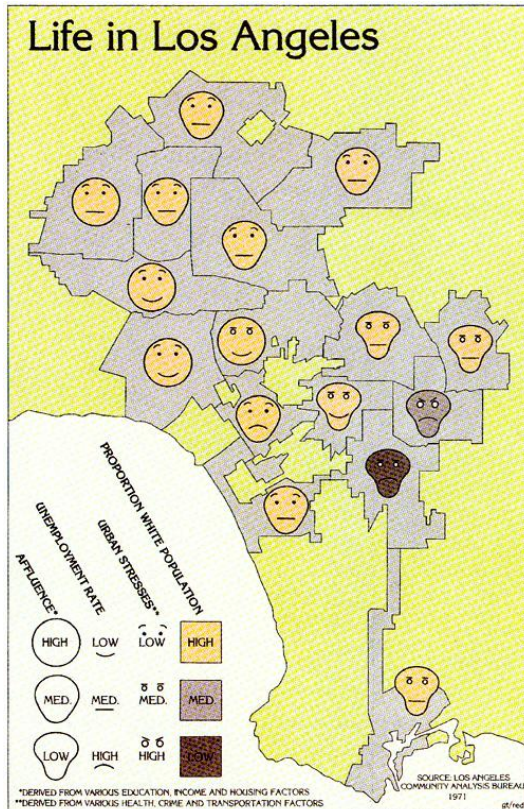
Um benefício específico, também salientado quando se faz uso de representações por meio de faces, é sua facilidade na identificação de *outliers* (pontos aberrantes). Um elemento da face que em magnitude apresente-se muito diferente é prontamente detectado e diagnosticado. Lott e Durbridge (1990), nesse sentido, ao usarem as Faces de Chernoff para analisar tendências em dados laboratoriais, constatam que esse tipo de representação visual permite reconhecer mais rapidamente a existência de anomalias. Em especial, destacam a situação onde são realizados muitos estudos laboratoriais, facilitando a comparação de resultados de exames em diferentes momentos - caracterizando a evolução de uma anomalia.

## **ALGUNS EXEMPLOS**

Para melhor ilustrar a representação com esses recursos, selecionou-se alguns exemplos de estudos focados nas Faces de Chernoff, usando-as para identificação de fatos e conclusões, ou mesmo como apoio para diagnósticos. O primeiro é um exemplo amplamente referenciado no tema. Os demais foram apurados junto a publicações científicas e são escassos casos de uso do recurso. Considerando que o objetivo é apenas ilustrar aplicações das Faces, não são feitos comentários aprofundados ou críticas à sua construção.

### ***Exemplo 1: Vida da População em Los Angeles (EUA)***

Spinelli e Zhou (2004) citam construção de Eugene Turner, da California State University (EUA), em 1979, de um mapa intitulado “Vida em Los Angeles”. Nele, as condições de vida da população, por área distrital, são representadas através de quatro variáveis: nível de riqueza (que considera condições educacionais, familiares e de renda), taxa de desemprego, stress urbano e porcentagem de população branca. A essas variáveis foram associados elementos das Faces de Chernoff: formato da face, curvatura da boca, inclinação da sobrancelha e cor da face, respectivamente (figura 2). A representação ganhou grande destaque e o sucesso foi atribuído exatamente pelo seu simbolismo e porque, segundo o autor, “as expressões evocam uma associação emocional com os dados”.



**Figura 2:** Condições de Vida da População de Los Angeles (EUA) – 1971

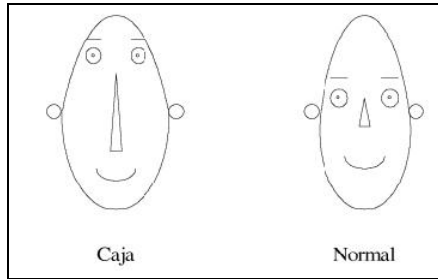
Fonte: Spinelli e Zhou (2004)

### **Exemplo 2: Aprovação da bebida alcoólica a partir da polpa de cajá**

Dias, Schwan e Lima (2003) apresentaram um trabalho propondo o uso do mosto da polpa de cajá para a produção de uma bebida alcoólica. Para constatar a aprovação da bebida pelos provadores, foram utilizadas as Faces de Chernoff: uma representando a aceitação do fermentado de cajá, e outra representando uma escala padrão, “normal”, que não expressa valores (figura 3). Segundo os autores, “este teste tornou mais nítidas as diferenças obtidas entre os atributos da bebida por utilizar um reconhecimento visual, não numérico”, e concluíram pela aceitação da qualidade da bebida – baseada na aparência, aroma e aspectos gerais.

#### Associações:

- Aspectos gerais = largura da face;
- Aroma = comprimento do nariz;
- Sabor = curvatura da boca;
- Aparência = altura dos olhos.



**Figura 3:** Representação da aceitação do fermentado de cajá

Fonte: Dias, Schwan e Lima (2003)

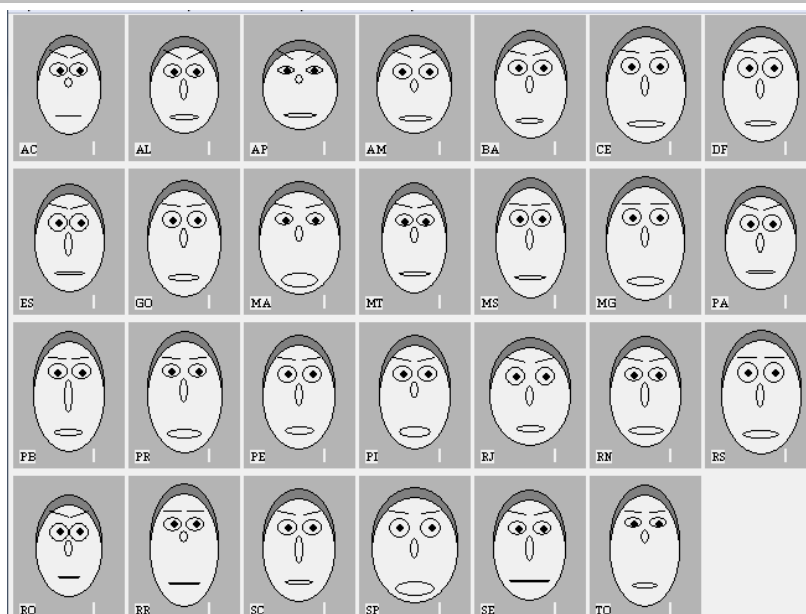
### ***Exemplo 3: Comparação da produção científica dos estados brasileiros***

Silva (2011), em sua dissertação de mestrado, usa as Faces de Chernoff para representar e analisar dados multivariados da produção científica nacional, justificando a característica de facilidade de associação e entendimento proporcionada por esse recurso gráfico.

Para comparar os diferentes estados nacionais, usa a representação dos volumes de produção por elementos da face:

- quantidade de periódicos nacionais por “altura do rosto”;
- periódicos internacionais por “largura do rosto”;
- trabalhos completos publicados em Anais e eventos pelo “tamanho do nariz”;
- capítulos de livros pela “altura dos olhos”;
- resumos em periódicos especializados pela “abertura da boca”; e
- resumos em anais por “tamanho da boca”.





**Figura 4:** Representação da produção científica nos estados brasileiros

Fonte: Silva (2011)

Apoiado pela aparência das imagens da Figura 4, o autor analisa as similaridades entre os estados, e destaca a grande quantidade de Unidades da Federação com baixa produção dos resumos (abertura e tamanho das bocas). Também salienta a conveniência do uso de cores para melhor visualização de comportamentos regionais.

#### **Exemplo 4: Avaliação da qualidade de serviços prestados ao consumidor**

Nel, Pitt e Webb (1994) usam as Faces de Chernoff como uma técnica para retratar dados sobre percepção da qualidade de serviços prestados.

Nesse estudo, os autores levantam a percepção de 2.700 consumidores ingleses com a qualidade de serviços de uma companhia elétrica, classificados segundo quatro grupos distintos: usuários residenciais, não-usuários residenciais, usuários comerciais e não-usuários comerciais. A atribuição de não-usuário foi dada àqueles que não usaram os serviços da companhia nos últimos cinco anos.

As dimensões de qualidade focadas são: *confiabilidade, segurança, empatia, tangibilidade e responsividade*.

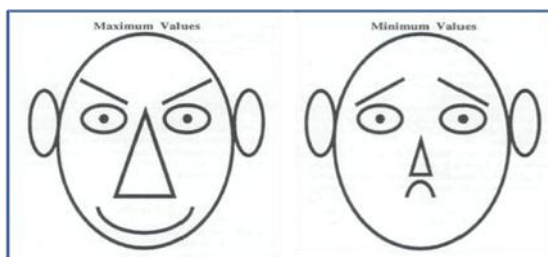
A atribuição dessas variáveis às características faciais segue, segundo os autores, uma ordem de importância dos aspectos avaliados comparativamente às



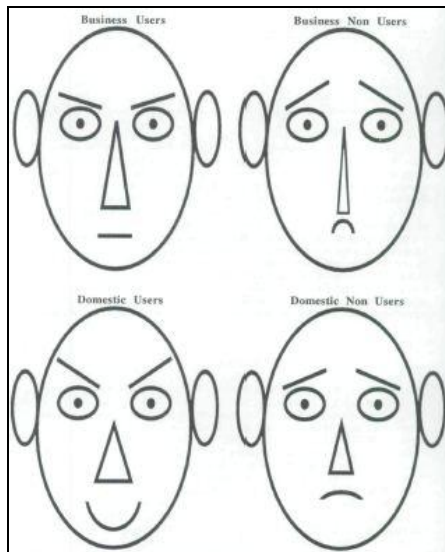
características faciais entendidas como “mais visíveis”. Assim, *confiabilidade* – destacada como a dimensão mais importante – é representada pela curvatura da boca. As demais associações são: *segurança* representada pela inclinação da sobrancelha, *empatia* pela largura do nariz, *tangibilidade* pelo comprimento do nariz, *responsividade* pelo comprimento da boca, e acrescenta-se um *índice de qualidade de serviço composto*, representando-o pela altura do centro da boca. Surpreendentemente não é associada nenhuma variável aos olhos, situação incomum nos estudos que adotam as Faces de Chernoff.

Para ilustrar as relações básicas entre as características faciais e também para fornecer medidas referenciais para comparação entre qualidades de serviços em pontos extremos, os autores constroem inicialmente duas faces contrastantes (figura 5). Na sequência, representam os resultados para os quatro grupos (figura 6), baseados nos quais concluem que:

- usuários, tanto residenciais quanto comerciais, apresentam scores significativamente mais favoráveis que não-usuários;
- todas categorias apresentam perfil similar com relação à tangibilidade, indicando ser essa a característica menos problemática; e
- os usuários residenciais são os que têm o índice de qualidade de serviço mais favorável. Esse grupo, observado de forma holística, é o que demonstra uma feição com maior satisfação.



**Figura 5:** Níveis extremos de satisfação com serviços prestados por companhia elétrica inglesa  
Fonte: Nel, Pitt e Webb (1994)



**Figura 6:** Resultado de pesquisa de satisfação com serviços prestados por companhia elétrica inglesa, segundo quatro grupos

Fonte: Nel, Pitt e Webb (1994)

## **ESTRUTURA DE CORRELAÇÕES E IMPORTÂNCIA DAS VARIÁVEIS**

As faces, conforme observam Kleiner e Hartigan (1981), têm alguma estrutura de correlação intrínseca entre as características que representam as variáveis. Assim, por exemplo, o comprimento das sobrancelhas é altamente correlacionado com a extensão dos olhos, e o comprimento do nariz é altamente correlacionado com a altura da face. “É, portanto, importante arranjar as variáveis de forma que a adequação à estrutura de correlação seja a melhor possível”.

Morris, Ebert e Rheingans (1999), sob a alegação que ainda não é claro como as pessoas reconhecem as faces humanas, e se há algum conjunto específico de regras que comandam esse processo, conduziram uma pesquisa de campo, e apontam o tamanho dos olhos e a inclinação da sobrancelha como os elementos mais precisos para tempos de visualização mais longos.

Em estudo específico, Dickinson (2001) estabelece que o “formato do rosto” é o atributo mais importante na composição da Face, justificando que esse elemento marca os limites dos atributos internos e contribui grandemente para a impressão inicial pelo leitor. O atributo “cabelo” é colocado pela autora como essencial para a Face, pois acrescenta interesse visual e configura-se como um padrão, fazendo com que a detecção de casos não usuais seja feita mais facilmente. Já as orelhas são

---

vistas pela autora como um atributo “periférico” à Face, que pode ser removido sem uma perda significativa de informação.

Para Golden e Sirdesai (1992) as variáveis importantes devem ser associadas à curvatura da boca, altura da face, olhos e sobrancelhas, o que permitirá uma melhor compreensão da imagem criada.

Flury e Riedwyl (1981) recomendam que, quando houver menos variáveis que parâmetros da face, deve-se associar mais de uma característica da face à mesma variável. Nesse caso, as variáveis que forem consideradas mais importantes para o experimentador terão um maior peso na representação visual – dependendo, naturalmente, de quais elementos faciais serão escolhidas.

Uma especial recomendação para melhor representar atributos como sexo e estado civil é o uso de cores na visualização (RABELO, 2007).

## **A CONSTRUÇÃO E O USO DE SOFTWARES COMPUTACIONAIS**

Se, por um lado, a transmissão de conhecimento e entendimento é favorecida pela imagem gráfica, por outro lado uma construção indevida ou sem os devidos cuidados pode conduzir o leitor a interpretações enganosas e resultados distorcidos. No caso das faces, a dificuldade operacional de construção é um fator que agrava essa preocupação.

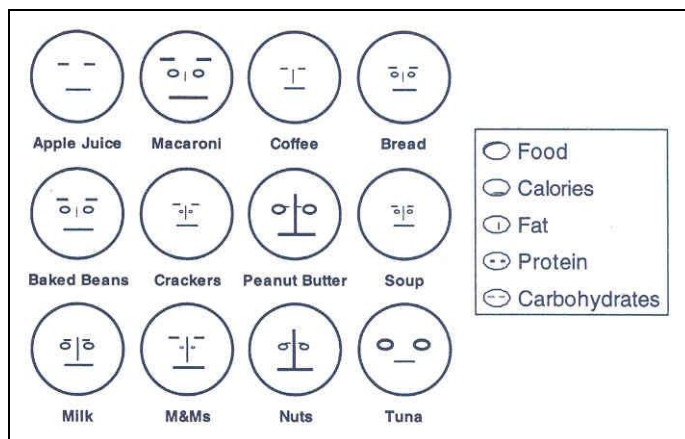
Em seu artigo, Chernoff faz uma reflexão: “Se esta simples idéia é tão boa, por que não foi pensada antes?”. E ele próprio responde: a aplicação eficiente desta idéia requer uma tecnologia computacional que só naquele momento (1973) passava a ser disponibilizada.

No período que imediatamente sucedeu a apresentação da proposta por Chernoff, as faces serviram de base e inspiração para novos estudos, que incluíam o desenvolvimento de algoritmos e programas computacionais específicos, tais como: FACE - desenvolvido em Fortran IV (Flury e Riedwyl, 1981), ArcGIS (Spinelli e Zhou, 2004), Pacote TeachingDemos – linguagem R (Rabelo, 2007), MATLAB (Silva Neto, 2008), dentre outros.

Não houve, contudo, adesões significativas a nenhuma proposta em especial. A restrição de acesso aos produtos computacionais desenvolvidos aliada à

dificuldade de construção manual, desestimulou a popularização das Faces de Chernoff.

Hunt (2004) apresenta uma forma de construir as Faces de Chernoff através do software Excel, da Microsoft, usando o modelo “Bolhas”. Embora esse produto seja de intensa penetração no mercado e a proposta de construção tenha sido apresentada de forma didática, o processo é artesanal e as imagens devem ser vistas, como expresso pelo autor, “com um pouco de ingenuidade”. O exemplo publicado é o apresentado na Figura 7, comparando características nutricionais entre diferentes tipos de alimentos.



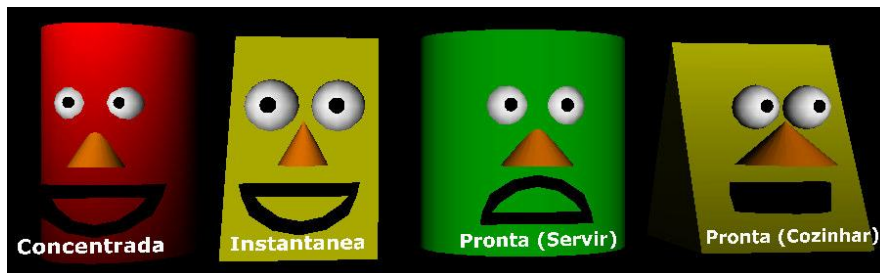
**Figura 7:** Exemplo de Representação das Faces de Chernoff através do Gráfico de Bolhas do software Excel, mostrando características nutricionais de alguns alimentos

Fonte: Hunt (2004)

Bueno (2005) usa as Faces de Chernoff e duas outras técnicas de visualização (coordenadas paralelas e coordenadas paralelas extrusivas) como base no desenvolvimento, em linguagem de programação Java, de uma ferramenta de visualização de dados multivariados num espaço tridimensional.

O autor ilustrou o uso desse recurso num estudo de caso sobre informações nutricionais de sopas servidas numa cafeteria. Ao invés das tradicionais cabeças propostas por Chernoff, representa as latas e pacotes de sopa respectivamente por elementos em formatos cilíndricos e paralelepípedos triangulares. Outras associações feitas no estudo: uso de cores para simbolizar os sabores das sopas (vermelho para tomate, amarelo para galinha, verde para legumes); custo, representado pela curvatura da boca; nível de sódio, atribuído ao tamanho dos olhos; teor de calorías, indicado pela largura da imagem; gordura, representada pela

largura do nariz; e calorias da gordura, associadas à tonalidade da cor da imagem. A Figura 8 mostra exemplos das imagens propostas.



**Figura 8:** Representação gráfica de sopas usando formatos diferenciados

Fonte: Bueno (2005)

Raciborski (2009) propõe um algoritmo para geração das Faces de Chernoff usando o software STATA como alternativa para contornar alguns inconvenientes do método de construção proposto por Chernoff. As imagens geradas foram ilustradas pelas figuras representadas na figura 9.



**Figura 9:** Exemplos de faces criadas com o uso do software STATA.

Fonte: Raciborski (2009)

Alguns pacotes estatísticos que alcançaram popularidade e reconhecimento internacional já incluem recursos para construção das Faces de Chernoff. Nuñez (2010) destaca: STATISTICA, S-PLUS e SYSTAT.

Embasados nas Faces de Chernoff, e já sob um desenvolvimento tecnológico mais evoluído para construções pictóricas, Loizides e Slater (2002) criaram o Algoritmo de Visualização Empática (*Empathic Visualisation Algorithm - EVA*), para representar um conjunto de dados através de imagens faciais que permitem uma visão holística do estado financeiro de quatro diferentes empresas. A proposta apresenta duas importantes características:

- representação visual “natural”, próxima de nossa realidade diária; e
- mapeamento automático, que representa as características semanticamente “importantes” associadas a elementos também “importantes” da estrutura visual, no sentido de transmitir a emoção ou sentimento humano. Os graus de importância são pré-determinados pelo pesquisador e são informações de entrada do algoritmo.

Os autores salientam que seu algoritmo, fazendo uso de técnicas de Programação Genética, considera o impacto da face na emoção do observador, associando aos dados um “sistema de valor” que lhe reflete o interesse ou importância dos dados. O exemplo apresentado está reproduzido na Figura 10.



**Figura 10:** Exemplos de faces construídos pelo algoritmo EVA

Fonte: Loizides e Slater (2002)

## LIMITAÇÕES E CRÍTICAS ÀS FACES DE CHERNOFF



---

Em seu artigo, Chernoff mencionou algumas limitações e fez uma autocrítica a sua proposta. Uma das questões levantadas foi sobre a importância relativa das características estudadas. Tendo obtido opiniões diversas de pesquisadores com quem trabalhou, acabou por concluir que essa questão requeria estudos mais aprofundados. O autor alertou ainda para algumas limitações que requerem cuidado no relacionamento entre variáveis, ilustrando que olhos de tamanho muito pequeno dificultam a detecção da posição da pupila. Chamou também a atenção para a representação não usual do “ponto zero” em elementos como curvatura da boca. Mas entendeu serem questões facilmente contornáveis. Para o pesquisador que sentir que algumas características são mais discriminantes, Chernoff sugere tentar a permutação de variáveis na representação de elementos faciais e eventualmente associar a soma de variáveis a alguns elementos.

Silva Neto (2008) menciona que, embora as Faces de Chernoff sejam bastante úteis para mostrar tendências em dados multidimensionais, os valores dos dados propriamente ditos devem ser apresentados adicionalmente, uma vez que as faces não transmitem informação sobre os reais valores com as quais se relacionam.

Huff, Mahajan e Black (1981) alertam para a importância relativa de certas variáveis, que podem ser exageradas em razão do elemento facial escolhido para representá-las. Comentam também a transmissão de sentimento das imagens (por exemplo, uma boca em formato de sorriso), quando na representação não se intencionou transmitir nenhum significado emocional. E indicam um terceiro potencial inconveniente: para uma determinada variável, a existência de um valor extremo comprime a escala de representação das demais observações, de forma que as diferenças entre estas fiquem mais difíceis de serem constatadas nas respectivas representações faciais. Mas os autores reconhecem que essa técnica oferece algumas vantagens sobre as demais técnicas gráficas, dentre as quais a facilidade de reconhecimento e percepção como um *gestalt*, a apuração da sensibilidade de cada variável em relação às outras, a identificação das dimensões chaves, e especialmente a atratividade em análises exploratórias de dados. E ainda: ser um indicador de performance pictórico para avaliar a eficácia das decisões sobre gestão de recursos ou para monitoramento de estratégias concorrenciais.

Raciborski (2009) cita, além da restrição gerada por valores extremos de certos parâmetros na variação de outros, o fato que, para manter as faces com um



mesmo tamanho, a largura e o comprimento de cada face precisam ser normalizados, o que compromete o efeito das variáveis atribuídas a essas duas características.

Rabelo (2007) faz uma avaliação das técnicas de visualização para representação dos resultados obtidos com a aplicação de técnicas de mineração de dados. O autor compara, seguindo um modelo de análise específico, seis métodos de representação gráfica segundo sete características básicas. As Faces de Chernoff, um dos métodos analisados, foram pontuadas como não satisfatórias nos quesitos de *dimensionalidade* (suporte à quantidade de atributos), *escalabilidade* (suporte à quantidade de dados) e *relacionamento entre os atributos*. Em contrapartida, mostraram destaque positivo na *facilidade de interpretação*.

Kosara (2007) reconhece que as Faces de Chernoff permitem uma identificação visual geral rápida, pois as pessoas não identificam isoladamente nariz, olhos, bocas, sobrancelhas. Há o reconhecimento do rosto como um todo, inclusive com aquilo que se imagina ser a personalidade da pessoa. Mas os elementos faciais sofrem certa priorização na atenção. A comparação e a diferenciação de características através das faces são muito difíceis, o que as torna, no entender do autor, uma escolha ruim para visualização.

Lee, Reilly e Butavicius (2003) mencionam que as Faces de Chernoff podem ser mais eficientes que muitas outras técnicas na medida em que permitem ao observador despender menos tempo na análise das informações. Mas ao compará-las com algumas técnicas de visualização espacial, especificamente para dados binários, obtiveram que as faces levam o observador a concluir menos rapidamente e de forma mais imprecisa.

Outra questão que se apresenta é sobre o erro a que se está sujeito nas classificações baseadas em identificações visuais. Pouco após a proposição das Faces, Chernoff e Rizvi (1975) conduziram um experimento gerando permutações aleatórias de 18 características da face provenientes de duas populações multivariadas. Após submissão a um grupo de pessoas, mediram o efeito da habilidade visual para classificar as observações em dois grupos separados, correspondentes às populações originais. Os autores concluíram que as permutações aleatórias afetaram a taxa de erro nesta classificação em cerca de

---

25%. Fienberg (1979), contudo, observa que esse estudo não avaliou a eficácia de características específicas, como por exemplo, os olhos ou a boca.

Em que pesem as restrições do experimento, a taxa constatada estimula o uso do recurso.

## **CONSIDERAÇÕES FINAIS**

Embora haja alguma variedade de recursos para representação visual de dados, tem havido pouca evolução rigorosa de suas eficácias. Muito espaço ainda há para avaliação dessas técnicas o que, acredita-se, só ganhará a devida atenção com a utilização e interesse dos pesquisadores.

Duas características fundamentais para os gráficos em geral, que incentivam sua adoção em grande escala e facilitam seu entendimento, são: simplicidade e familiaridade. As Faces de Chernoff apresentam algumas limitações, mas têm forte apelo na representação de variáveis exatamente pelo símbolo que adota: faces humanas são imagens de extrema familiaridade aos observadores. E são particularmente úteis no reconhecimento de padrões, identificação de pontos aberrantes e no agrupamento de unidades amostrais.

As Faces de Chernoff, como qualquer outro gráfico, não devem ser vistas como um fim em si próprio, mas como um instrumento que permite a análise do fenômeno que se propõe identificar. Também não devem ser usadas como recurso único na análise de dados, nem devem concorrer com outros gráficos já tradicionalmente adotados e eficientes dentro de sua finalidade. Sua grande colaboração está em complementar outras ferramentas analíticas de tratamento de dados. Em especial, sua visualização auxilia na detecção de padrões, formação de clusters, identificação de pontos aberrantes e na verificação de tendências temporais.

Operacionalmente, algumas recomendações para construção foram oferecidas, mas não esgotam o tema. Ao contrário, devem ser encaradas como ponto inicial para aprofundar o conhecimento do recurso, de maneira a permitir-lhe a exploração de potencialidades reconhecidamente de grande valia no tratamento de dados multidimensionais.

---

Espera-se com esse artigo, não só transmitir conhecimento sobre uma forma gráfica pouco explorada, mas também encorajar e estimular estudantes, professores, pesquisadores e profissionais em geral a fazer uso de métodos gráficos em seus trabalhos, qualquer que seja sua natureza ou campo de atuação. Em especial a representação de dados multidimensionais através das Faces de Chernoff deve ser alvo de atenção para proporcionar uma maior facilidade de compreensão e análise dos fenômenos associados. Aos pesquisadores da área, desafia-se o maior desenvolvimento de teorias sobre gráficos, que estruturam, disciplinem, padronizem, popularizem e desenvolvam novos dispositivos e recursos nesse segmento.

## **MULTIVARIATE DATA REPRESENTATION USING THE CHERNOFF FACES**

### **ABSTRACT**

This paper approaches the graphical representation as a valuable resource to support the assessment of factors, variables and observable phenomena, focusing on a particular model, which seeks to overcome the restriction of the representation of multidimensional data, without loss of accuracy and ease of understanding: it presents the "Chernoff Faces," illustrates its application in some published works, discusses its characteristics and peculiarities, exposes its limitations and criticisms identified in the literature review, and discusses aspects of its construction - from the association of the attributes of the face to the variables of the study until some of the computational resources available. The paper concludes by highlighting your convenience in the analysis of multivariate data, stimulating their exploration.

**KEY WORDS:** Statistical graphics, Multivariate data representation, Chernoff Faces.

### **REFERÊNCIAS BIBLIOGRÁFICAS:**

BUENO, Márcio Augusto Silva. **Meta3D** – Uma ferramenta para visualização de informações em 3D. 2005. Dissertação (Mestrado em Ciências da Computação) – Universidade Federal de Pernambuco – UFPE, Recife, 2005.

CHERNOFF, Herman. **The use of faces to represent points in k-dimensional space graphically.** Journal of the American Statistical Association, v. 68, n. 342, p. 361–368, jun. 1973.

- CHERNOFF, Herman; RIZVI, M. Haseeb. **Effect on classification error or random permutations of features in representing multivariate data by faces.** Journal of American Statistical Association, v. 70, n. 351, p. 548-554, sep. 1975.
- DIAS, Disney R.; SCHWAN, Rosane F.; LIMA, Luiz Carlos O. **Metodologia para elaboração de fermentado de cajá (*Spondias mombin* L.).** Ciência e Tecnologia de Alimentos, Campinas, v. 23, n. 3, Set./Dez. 2003.
- DICKINSON, Wendy B. Escaping flatland: **Chernoff's faces revisited.** Proceedings of the Twenty-sixth SAS Users Group International Conference (Paper 195). Cary, NC: SAS Institute. Florida, 2001. Disponível em <http://www2.sas.com/proceedings/sugi26/p195-26.pdf>. Acesso em 03.02.2009.
- FIENBERG, Stephen E. **Graphical methods in statistics.** The American Statistician, v. 33, n. 4, p. 165-178, nov. 1979.
- FLURY, Bernhard; RIEDWYL, Hans. **Graphical representation of multivariate data by means of asymmetrical faces.** Journal of the American Statistical Association, v. 76, n. 376, p. 757-765, dec. 1981.
- GOLDEN, Linda L.; SIRDESAI, Mayur. **Chernoff Faces: A Useful Technique for Comparative Image Analysis and Representation.** Advances in Consumer Research, v. 19, p. 123-128, 1992.
- HUFF, David L.; MAHAJAN, Vijay; BLACK, William C. **Facial representation of multivariate data.** Journal of Marketing, v. 45, n. 4, p. 53-59, fall 1981.
- HUNT, Neville. **Chernoff faces in Microsoft Excel.** Teaching Statistics. v. 26, n. 3, p. 75-77, aut. 2004.
- KLEINER, Beat; HARTIGAN, John A. **Representing points in many dimensions by trees and castles.** Journal of the American Statistical Association, v. 76, n. 374, p. 260-269, jun. 1981.
- KOSARA, Robert. **A critique of Chernoff faces.** 2007. Disponível em <http://eagereyes.org/viscrit/chernofffaces.html>. Acesso em 02.02.2009.
- LEE, Michael D.; REILLY, Rachel E.; BUTAVICIUS, Marcus A. **An empirical evaluation of Chernoff faces, star glyphs, and spatial visualizations for binary data.** ACM International Conference Proceeding Series, v. 142, 2003. Adelaide. Disponível em <http://portal.acm.org/citation.cfm?id=857080.857081>. Acessado em 02.03.2009.

- LOIZIDES, Andreas; SLATER, Mel. **The Empathic Visualisation Algorithm (EVA) — An automatic mapping from abstract data to naturalistic visual structure.** Sixth International Conference on Information Visualisation, p. 705, 2002. Disponível em [http://www.cs.ucl.ac.uk/staff/a.loizides/eva\\_paper.pdf](http://www.cs.ucl.ac.uk/staff/a.loizides/eva_paper.pdf). Acessado em 07.01.2009.
- LOTT, John A.; DURBRIDGE, Timothy C. **Use of Chernoff faces to follow trends in laboratory data.** Journal of Clinical Laboratory Analysis, v. 4, n. 1, p. 59-63, 1990.
- MORRIS, Christopher J.; EBERT, David S.; RHEINGANS, Penny. **An experimental analysis of the pre-attentiveness of features in Chernoff faces.** Proceeding of Applied Imagery Pattern Recognition '99: 3D Visualization for Data Exploration and Decision Making. Oct. 1999. Disponível em [http://www.research.ibm.com/people/c/cjmorris/publications/Chernoff\\_990402.pdf](http://www.research.ibm.com/people/c/cjmorris/publications/Chernoff_990402.pdf). Acessado em 06/01/2009.
- NEL, Deon; PITT, Leyland; WEBB, Trevor. **Using Chernoff Faces to Portay Service Quality Data.** Journal of Marketing Management, n. 10, p. 247-255, 1994.
- NUÑEZ, José Jesús Reyes. **Ideas para el uso de las caras de Chernoff en la cartografía escolar.** Boletim de Geografia, v. 28, n. 1, p. 5-15, 2010.
- RABELO, Emerson. **Avaliação de técnicas de visualização para mineração de dados.** 2007. Dissertação (Mestrado em Ciências da Computação) – Universidade Estadual de Maringá – UEM, Maringá, 2007.
- RACIBORSKI, Rafal. **Graphical representation of multivariate data using Chernoff faces.** The Stata Journal, v. 9, n. 3, p. 374-387, 2009.
- SILVA, Levi Alã Neves dos. **Contribuição da mineração de dados e da otimização heurística para a interpretação dos dados da produção científica brasileira.** 2011. Dissertação (Mestrado em Ciências da Informação) – Universidade Federal da Bahia, Salvador, 2011.
- SPINELLI, Joseph G.; ZHOU, Yu. **Mapping quality of life with Chernoff faces.** Ohio, 2004. Disponível em <http://gis.esri.com/library/userconf/educ04/papers/pap5000.pdf>. Acessado em 14.01.2009.
- SILVA NETO, Marco Aurélio. **Mineração visual de dados: extração do conhecimento a partir das técnicas de visualização da informação e mineração de dados.** 2008. Dissertação (Mestrado em Ciências) – Universidade Federal do Paraná – UFPR, Curitiba, 2008.