

UNIVERSIDADE FEDERAL DO PARÁ

INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Anderson Gregório Marques Soares

Expandindo a Visualização de Dados Multidimensionais em Treemaps com Glifos

Belém

2020

Anderson Gregório Marques Soares

Expandindo a Visualização de Dados Multidimensionais em Treemaps com Glifos

Tese de Doutorado apresentado no Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas e Naturais como requisito parcial para obtenção do Título de Doutor em Ciência da Computação.

Universidade Federal do Pará

Orientador: Bianchi Serique Meiguins

Belém

2020

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)

S676e Soares, Anderson Gregório Marques
Expandindo a Visualização de Dados Multidimensionais em Treemaps com Glifos / Anderson Gregório Marques Soares. — 2020.
130 f. : il. color.

Orientador(a): Prof. Dr. Bianchi Serique Meiguins
Tese (Doutorado) - Programa de Pós-Graduação em Ciência da Computação, Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, Belém, 2020.

1. Visualização da informação. 2. Treemap. 3. Glifos. 4. Avaliação. I. Título.

CDD 006

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

ANDERSON GREGÓRIO MARQUES SOARES

EXPANDINDO A VISUALIZAÇÃO DE DADOS MULTIDIMENSIONAIS EM *TREEMAPS* COM GLIFOS

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal do Pará como requisito para obtenção do título de Doutor em Ciência da Computação, defendida e aprovada em 12/06/2020, pela banca examinadora constituída pelos seguintes membros:

Bianchi Serique Meiguins
Bianchi Serique Meiguins (15 de June de 2020 21:23 ADT)

Prof. Dr. Bianchi Serique Meiguins
Orientador – PPGCC/UFPA

Carlos Gustavo Resque dos Santos

Prof. Dr. Carlos Gustavo Resque dos Santos
Membro Interno – PPGCC/UFPA

Nelson Cruz Sampaio Neto.

Prof. Dr. Nelson Cruz Sampaio Neto
Membro Interno – PPGCC/UFPA

Marcelo de Paiva Guimarães

Prof. Dr. Marcelo de Paiva Guimarães
Examinador Externo – UNIFESP

José Gustavo de Souza Paiva
José Gustavo de Souza Paiva (16 de June de 2020 14:17 ADT)

Prof. Dr. José Gustavo de Souza Paiva
Examinador Externo – UFU

Visto: Nelson Cruz Sampaio Neto.

Prof. Dr. Nelson Cruz Sampaio Neto
Coordenador do PPGCC/UFPA

Dedico este trabalho a Deus, a minha esposa Adriele pelo amor, compreensão e companheirismo, a minha filha Agatha pela motivação, aos meus familiares, aos meus mestres e todos aqueles que contribuíram para que eu chegasse até aqui.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por ter me dado sabedoria para conclusão do doutorado, pois sem ele nada teria sido possível.

Agradeço à minha esposa e eterna companheira Mª Adriele Soares, por ter me dado amor, coragem, incentivo, apoio e compreensão nos momentos de grandes dificuldades durante este curso. Amor, obrigado por ter suportado ao meu lado todas as dificuldaes durante esse curso, se eu não tivesse a certeza de que eu poderia contar com teu amor, eu teria enlouquecido nesse trajeto. Eu te amo. Um agradecimento especial a minha filha Agatha Soares, pois ela foi minha fonte de motivação para a conclusão do doutorado.

Agradeço à banca avaliadora, constituída pelos professores José Gustavo de Souza Paiva, Marcelo de Paiva Guimarães, Nelson Cruz Sampaio Neto e Carlos Gustavo Resque dos Santos, por aceitar o convite de compor a banca, pelas contribuições, críticas construtivas e questionamentos, que me fizeram refletir e aprimorar esta tese de doutorado. De maneira especial agradeço ao meu orientador e amigo Prof. Dr. Bianchi Serique Meiguins, pela orientação (sempre disponivel independente da hora), ajuda (especialmente nas quedas de pressão), dedicação, pelas risadas, por todos os momentos de descontração e pelos sábios conselhos.

Aos meus grandes amigos e colegas de trabalho: Elvis Miranda (deve sonhar com glifos até hoje), Rodirgo do Amor Divino (o beldade do Laboratório) e Diego Santos (grande senhor H). Obrigado por me auxiliarem nesta caminhada, e por serem amigos com quem sempre pude confiar. Registro meu agradecimento aos demais amigos do Laboratório de Visualização, Interação e Sistemas Inteligentes (LabVIS), Tiago, Alexandre, Brunelli, Roberto, Walbert, Hugo, e demais membros do laboratório, que contribuiram e que contribuirão para a evolução desta pesquisa. Estendo meu agradecimento ao Prof. Resque (Mestre Gustavo) que além de me auxiliar na pesquisa, me emprestava os ouvidos durante as caronas para a casa.

Agradeço à empresa Gol Software pelo espaço disponibilizado nos finais de semana para o desenvolvimento de pesqusias da equipe do LabVIS, aos participantes dos testes de usabilidade, e aos meus familiares —em especial a meu pai (Valdir Soares), minha mãe (Margarida Soares) e minha madrinha (Margareth Marques)—, obrigado por terem me ensinado a ser perseverante e a ter fé em Cristo. Por fim, agradeço a todos os que direta e indiretamente contribuíram de alguma forma para a conclusão deste trabalho.

"I hated every minute of training, but I said: -Don't quit. Suffer now and live the rest of your life as a champion."

(Muhammad Ali)

RESUMO

A Visualização da Informação (InfoVis) é a área que estuda representações visuais interativas para dados abstratos com objetivo de potencializar o modelo mental dos usuários sobre os dados, ou seja, auxilia na percepção, compreensão, interpretação dos dados, os quais, sem uma visualização, exigiriam maior esforço para a descoberda de novas informações. Entre as técnicas de representação de dados de InfoVis, a técnica Treemap é uma técnica consolidada e difundida, que possui como principais características a visualização de dados hierárquicos e o uso de todo o espaço visual disponível de tela. Tal técnica apresenta alguns desafios para a representação de dados, entre eles destacam-se: poucas opções para representação visual dos dados, limitada a tamanho, cor e rótulo. Para melhorar a navegação no layout da técnica, pode-se adicionar níveis hierarquicos, e com isso, aumenta-se a quantidade de dimensões representadas visualmente, porém a representação visual de um número elevado de hierarquias pode dificultar a percepção dos usuários ao analisarem os dados. No contexto de InfoVis, glifos são objetos visuais que podem codificar vários atributos de uma base de dados a uma ou mais variáveis visuais. Desta forma, o uso de glifos pode aumentar a quantidade de dimensões representadas visualmente e possibilita a representar diversas opções de design. Em virtude do aumento da quantidade de dimensões representadas e da flexibilidade de representações visuais, essa tese adiciona glifos aos itens de dados do Treemap para mitigar os desafios da técnica. Os glifos desta pesquisa são compostos por N camadas empilhadas que se sobrepõem parcialmente, de forma que as camadas superiores possuem área menor que as inferiores, e cada camada mapeia uma dimensão de dados para uma variável visual. Adicionalmente, essa tese propõe o uso de glifos adaptativos, que são glifos que exibem apenas as camadas que possuem informações visuais úteis, ou seja, permanecem as camadas que conseguem representar informações visuais que os usuários conseguem visualizar. Tais glifos são aplicados aos itens do Treemap que possuem áreas muito pequenas. Para avaliar a proposta de glifos em camadas aplicados aos itens do Treemap, realizou-se testes de usuários para comparar cenários de mapeamentos de dados visuais em Treemaps, sendo estes: somente Glifos (G), somente Hierarquia (H), Glifos e Hierarquia (GH) e Glifo Adaptativo (GA). Para tal, foi desenvolvido um protótipo computacional que permite criar os cenários citados para análise de dados multidimensionais. Os resultados apontaram que o cenário GH enriqueceu os recursos de visualização do Treemap, sem prejudicar a legibilidade dos dados. Os participantes do teste cenário GA indicaram que os glifos adaptativos aplicados no Treemap apresentaram Boa usabilidade e melhorou a qualidade visual da informação na visualização.

Palavras-chave: Visualização da Informação, Treemap, Glifo, Avaliação.

ABSTRACT

Information Visualization (InfoVis) is the area that studies interactive visual representations for abstract data to enhance the users' mental model about the data, in other words, it helps in the perception, understanding, interpretation of the data, which, without a visualization, would require more considerable effort to discover new information. Among InfoVis data representation techniques, the Treemap technique is a consolidated and widespread technique, whose main characteristics are the visualization of hierarchical data and the use of all available visual space on the screen. Such technique presents some challenges for the representation of data, among which we highlight: a few options for the visual representation of the data, limited to size, color, and label. We can improve the navigation through the technique layout by adding hierarchical levels, increasing the number of visually represented dimensions. However, the visual representation of a high number of hierarchies can make it difficult for users to perceive when analyzing the data. In the context of InfoVis, glyphs are visual objects that can encode various attributes of a database to one or more visual variables. In this way, the use of glyphs can increase the number of visually represented dimensions and make it possible to represent several design options. This thesis adds glyphs to the Treemap data items to mitigate the challenges of the technique due to the increase in the number of dimensions represented and the flexibility of visual representations. The glyphs in this research are composed of N stacked layers that partially overlap, so that the upper layers have a smaller area than the lower ones, and each layer maps a data dimension to a visual variable. Additionally, this thesis proposes using adaptive glyphs, which are glyphs that display only the layers that have useful visual information, i.e., it remains the layers that can represent visual information that users can visualize. We applied the adaptive glyphs in Treemap items that have tiny areas. To evaluate the proposal of layered glyphs applied to Treemap items, we performed user tests to compare scenarios for visual mapping data in Treemaps, which are: only Glyphs (G), only Hierarchy (H), Glyphs and Hierarchy (GH) and Adaptive Glyph (GA). To this end, we developed a computational prototype to create the scenarios mentioned for the analysis of multidimensional data. The results showed that the GH scenario enriched Treemap visualization resources without impairing the data's legibility. Participants in the GA scenario test indicated that the adaptive glyphs applied in the Treemap showed Good usability. Furthermore, improved the visual quality of the information in the visualization.

Keywords: Information Visualization, Treemap, Glyph, Evaluation.

LISTA DE ILUSTRAÇÕES

Figura 1. A - Treemap representando visualmente 5 atributos de uma base de dados, 1 através do tamanho, 1 através da cor, 1 através do rótulo e 2 níveis hierárquicos. B - Adição de 6 níveis hierarquicos ao Treemap de A, representando visualmente 11 atributos. 21

Figura 2. Modelo do glifo em camadas, no qual as camadas se sobrepoem uma sobre as outras em 70%. A - Visão em perspectiva das camadas (C_1, C_2,..,C_n) do glifo; B - Visão superior das camadas (C_1, C_2,..,C_n) do glifo. 22

Figura 3. Cada variável visual está disposta em uma camada do glifo, de forma que cada camada cause apenas uma sobreposição parcial na camada posterior (B), como em uma torre de hanoi (C). 23

Figura 4. Visão geral da metodologia de pesquisa 26

Figura 5. Modelo de referência para visualização. 29

Figura 6. Treemap e diagrama de árvore. 31

Figura 7. Exemplos de implementação da técnica Treemap. A – Slice and Dice; B – Strip Treemap; C – Squarifield; 33

Figura 8. Exemplo da execução do algoritmo Squarified Treemap. 35

Figura 9. Variáveis visuais preconizadas por BERTIN (1983). 37

Figura 10. Recomendações de BERTIN(1983) para utilização de variáveis visuais. 38

Figura 11. Classificação da codificação de variáveis visuais quantitativas e não quantitativas. 38

Figura 12. Glifos: designs de glifos encontrados na literatura. 41

Figura 13. Percentual dos trabalhos tomando por base o número de dimensões codificadas nos glifos. 42

Figura 14. A - Exemplo de glifo em uma estrutura uniforme. B - Exemplo de glifo em uma estrutura orientada por dados. 43

Figura 15. Exemplo de glifos orientados pela estrutura. A - Estrutura em Espiral. B - Técnica baseada em estrutura de árvore. 43

Figura 16. Etapas ao definir o "porquê" de uma tarefa de InfoVis ser realizada. Este foi o fluxo usado nas tarefas das avaliações desta tese. 45

Figura 17. Exemplo de visualização gerada a partir da ferramenta Glyph Sorting. O design do glifo apresenta sobreposição das variáveis visuais sem prejudicar a leitura dos dados. 48

Figura 18. A - Contexto de aplicação do modelo do glifo; B - Modelo do glifo apresentado por MAGUIRE et al. (2012); C - Exemplo do glifo, mostrando o uso das variáveis visuais cor e forma, e faz uso da sobreposição entre elas. 49
Figura 19. Modelo do glifo apresenta atributos categóricos e ordenados, e utiliza sobreposição das variáveis visuais sem interferir na leitura do glifo. . . 49
Figura 20. A - Contexto de aplicação do modelo do glifo; B - Exemplos de glifos de THERON et al. (2007) baseado em formas e usando sobreposição. . 50
Figura 21. Tela da ferramenta Glyphmaker. 50
Figura 22. Visualização de uma busca no sistema de arquivos, onde os glifos representam resultado da busca (foco) e o Treemap representa o sistema de arquivos geral (contexto) e o design do glifo proposto. 51
Figura 23. Visualização utilizando o ClockMap, combinando Treemap circular e glifos. 52
Figura 24. Treemaps com técnica de agregação para estrutura de arquivos, onde os profile glifos (gráfico de barras) (A) e glifos de gráficos de pizza (B) mostram a distribuição de tamanho dos arquivos. 52
Figura 25. Exemplos de visualizações geradas a partir da ferramenta de avaliação. 56
Figura 26. Valores visuais distintos escolhidos (3, 4 e 5) para cada variável visual estudada. 56
Figura 27. Glifo com duas (2) camadas, sendo que uma a camada com uma variável visual e a outra a camada de sobreposição. A sobreposição poderia variar entre 0%, 50%, 60% e 70%. 57
Figura 28. Visualizações com a camada de sobreposição com um valor de 70% com exemplos de glifos gerados durante a execução do teste (I - exemplo com textura; II - exemplo com cor-matiz; III - exemplo com orientação; IV - exemplo com texto). (I) A - Área da visualização do layout em grade; B - Área de apresentação da tarefa de InfoVis; C - Área de apresentação do alvo para ser procurado na grade da área A. 58
Figura 29. Ranking das variáveis visuais para dados categóricos, considerando os níveis de oclusão (0%, 50%, 60% ou 70%) e valores distintos (3, 4 e 5). 60
Figura 30. Comparação dos resultados com MACKINLAY (1986), que avaliou variáveis visuais em cenários não sobrepostos. 61
Figura 31. Design do glifo em camadas. A - Glifo com as variáveis visuais cor - matiz, forma e texto; B - Neste glifo houve uma troca no posicionamento das camadas cor - matiz e forma comparado com A. 62
Figura 32. Valores das variáveis visuais categóricas e ordinais usadas nas camadas do glifo e o tipo de dados que eles são adequados para representar. . . 63

Figura 33. Treemap com glifos em camadas, onde os glifos possuem três camadas. A camada superior apresenta um profile glifo, que representa dados quantitativos de quatro dimensões. As barras indicam valores positivos (acima do eixo zero) ou negativos (abaixo do eixo zero), sempre respeitando as linhas de referência que indicam o tamanho máximo e mínimo da barra. 65
Figura 34. Aplicação desenvolvida para que os participantes pudessem avaliar a percepção do glifo. A - Exemplo de glifo gerado, variando aleatoriamente tamanho e camadas dos glifos; B - Área de apresentação da tarefa de InfoVis; C - Controles para que o participante pudesse informar quais camadas do glifo de A eram possíveis de ser identificadas. 66
Figura 35. Valores utilizados neste estudo que são referentes às camadas orientação, texto, forma, posição, cor (matiz) e textura das variáveis visuais que compõem os glifos. 66
Figura 36. Visão geral do resultado do pré-teste. Hierarquia do Treemap: Menor lado de um item > Área da última camada (camada N) > Número de camadas do glifo. 67
Figura 37. Árvore de decisão gerada pela biblioteca Scikit-learn. 70
Figura 38. Matriz de confusão exibindo as frequências de classificação para cada classe do modelo. 70
Figura 39. A - Visualização gerada sem aplicação da função adaptativa para construção dos glifos; B - A mesma visualização agora com aplicação do algoritmo adaptativo para construção dos glifos; B1 - Remoção das camadas Letra e Orientação; B2 - Remoção da camada Textura; B3 – Remoção da camada Posição; B4 - Após a execução do algoritmo adaptativo restou apenas a camada cor no glifo. 71
Figura 40. Diagrama de caso de uso do protótipo. 73
Figura 41. Caso de uso da funcionalidade Visualizar glifo. 74
Figura 42. A – Visualização do Treemap; B - Área de configuração das variáveis visuais do glifo; C - Área de mapeamento dos atributos as variáveis visuais; D - Área de exibição das legendas dos glifos. 75
Figura 43. A – Visualização do Treemap; B - Área de configuração das dimensões do profile glifo; C - Área de exibição da legenda do profile glifo. 75
Figura 44. A – Visualização do Treemap com os glifos adaptativos; B - Glifos adaptativos onde as camadas forma e glifo adaptativo foram removidas; C - Área de exibição dos detalhes sob demanda; D - Glifo em camadas sendo exibido nos detalhes sob demanda. 76
Figura 45. Padrão de arquitetura adotado e o relacionamento entre seus componentes. 77
Figura 46. Diagrama de classes usado para implementar os glifos em camadas. . . 78

Figura 47. Fluxo de processo do protótipo. 79
Figura 48. Visualização do cenário G: cenário configurado para uma das tarefas de usuário. Essa visualização não usa hierarquia do Treemap, a ausência de hierarquia é substituída por glifos categóricos. 82
Figura 49. Visualização do cenário GH: cenário configurado para uma das tarefas do usuário. Esta visualização é um Treemap com hierarquia e glifos. . . 83
Figura 50. Visualização do cenário H: cenário configurado para uma das tarefas do usuário. Esta visualização é um squarified Treemap sem glifos. 83
Figura 51. Exemplo de tarefa exibida para o participante, o comando da tarefa e o botão de envio está acima da visualização. 86
Figura 52. Como interpretar os gráficos para os resultados de uma tarefa nas seções a seguir. (a) - Quantidade de respostas corretas e incorretas e um teste em pares usando bootstrapping; (b) - tempo médio de resposta com intervalos de confiança de 95%; (c): comparação do tempo em termos de comparação usando o bootstrap e o Cohen'd. 87
Figura 53. (a) - Respostas certas / erradas e teste aos pares usando bootstrapping; (b) - Intervalo de confiança no tempo de resposta; e (c) - tempo de teste em termos de comparação para tarefas T1, T2, T3 e T4. 88
Figura 54. (a) - Respostas certas / erradas e teste aos pares usando bootstrapping; (b) - Intervalo de confiança no tempo de resposta; e (c) - tempo de teste em termos de comparação para tarefas T5, T6, T7 e T8. 89
Figura 55. (a) - Respostas certas / erradas e teste aos pares usando bootstrapping; (b) - Intervalo de confiança no tempo de resposta; e (c) - tempo de teste em termos de comparação para tarefas T1, T4, T6 e T8. 90
Figura 56. (a) - Respostas certas / erradas e teste aos pares usando bootstrapping; (b) - Intervalo de confiança no tempo de resposta; e (c) - tempo de teste em termos de comparação para a tarefa T2. 91
Figura 57. (a) - Respostas certas / erradas e teste aos pares usando bootstrapping; (b) - Intervalo de confiança no tempo de resposta; e (c) - tempo de teste em termos de comparação para a tarefa T3. 92
Figura 58. (a) - Respostas certas / erradas e teste aos pares usando bootstrapping; (b) - Intervalo de confiança no tempo de resposta; e (c) - tempo de teste em termos de comparação para a tarefa T4. 93
Figura 59. (a) - Respostas certas / erradas e teste aos pares usando bootstrapping; (b) - Intervalo de confiança no tempo de resposta; e (c) - tempo de teste em termos de comparação para a tarefa T5 e T6. 94
Figura 60. (a) - Respostas certas / erradas e teste aos pares usando bootstrapping; (b) - Intervalo de confiança no tempo de resposta; e (c) - tempo de teste em termos de comparação para a tarefa T7 e T8. 95

Figura 61. Total de erros nos cenários. 96
Figura 62. Visualização do cenário GA: cenário configurado para uma das tarefas do usuário. Esta visualização é um squarified Treemap com glifos adaptativos. 98
Figura 63. Tela exibida no manual do protótipo, indicando ao participante onde ler as tarefas, onde ver as legendas e onde estaria o Treemap com glifos. 101
Figura 64. Acurácia das tarefas por camadas ocultas. 102
Figura 65. Tempo de execução das tarefas por camadas ocultas. 102
Figura 66. Acurácia das tarefas por camadas ocultas - Tipo de tarefa Identificação. 103
Figura 67. Acurácia das tarefas por camadas ocultas - Tipo de tarefa Comparação. 103
Figura 68. Escala de adjetivos SUS. A avaliação do glifo adaptativo obteve a nota 72,67, que equivale ao adjetivo Bom e intervalo de aceitação como Aceitável. 105
Figura 69. Pontuação do SUS da avaliação dos glifos adaptativos na técnica Treemap. 106
Figura 70. Qualidade da informação visual entre os cenário GH e GA. Os participantes optaram, em sua maioria, pelo GA. 107
Figura 71. Razões pelas quais os participantes escolheram o cenário GA nas perguntas 1 e 2. 108
Figura 72. Estratégias utilizadas pelos participantes. 108
Figura 73. Alternativas da questão 1 da avaliação qualitativa. A - A área demarcada possui glifos com 6 camadas; B - Os glifos da área demarcada possuem uma variação de 3 a 4 camadas removidas. 109
Figura 74. Alternativas da questão 2 da avaliação qualitativa. A - A área demarcada possui glifos com 4 camadas; B - Os glifos da área demarcada possuem uma variação de 1 a 2 camadas removidas. 110

LISTA DE QUADROS

Quadro 1. Propostas que usaram glifos e propostas que usavam glifos em Treemaps em comparação com esta tese. 53
Quadro 2. Dimensões da base de dados sintética sobre dados meteorológicos. Base de dados com 200 registros. 81
Quadro 3. Lista de tarefas utilizadas durante o teste. 84
Quadro 4. Distribuição dos atributos nos cenários G, GH e H ao realizar a tarefa 4. A tarefa de usuário utiliza os mesmos atributos, variando apenas no mapeamento visual. 85
Quadro 5. Dimensão adicionada a base sintética sobre dados meteorológicos. Base de dados com 1000 registros. 98
Quadro 6. Lista de tarefas utilizadas durante o teste. 99
Quadro 7. Relação do número de camadas ocultas no glifo em cada tarefa com o tipo de tarefa. 101
Quadro 8. Questionário adaptado com base no método SUS 104
Quadro 9. Interpretação da Escala SUS. 105
Quadro 10. Lista de perguntas qualitativas. 107
Quadro 11. Limitações do Treemap e como esta tese buscou mitiga-las. 111

LISTA DE ABREVIATURAS E SIGLAS

2D	Duas Dimensões
3D	Três Dimensões
API	Application Programming Interface (Interface de Programação de Aplicativos)
CSV	Comma-Separated-Values
G	Cenário de teste utilizando apenas glifos sem hierarquias
GA	Cenário de teste utilizando o glifo adaptativo
GH	Cenário de teste utilizando glifos e hierarquias
GUI	Graphical User Interface (Interface Gráfica de Usuário)
H	Cenário de teste sem glifos mas com hierarquias
HD	Hard Disk (Disco Rígido)
InfoVis	Information Visualization (Visualização da Informação)
IP	Internet Protocol (Protocolo de Internet)
JRE	Java Runtime Environment
MVC	Model-View-Controller (Modelo-Visão-Controle)
Px	Pixels
RGB	Sistema de cores Red (Vermelho), Green (Verde) e Blue (Azul)
SUS	System Usability Scale (Escala de Usabilidade do Sistema)
TB	Tera Byte
TCLE	Termo de Consentimento Livre e Esclarecido
TXT	Arquivo de texto

SUMÁRIO

1 INTRODUÇÃO 19
1.1 Objetivos 25
1.2 Contribuições 25
1.3 Metodologia 25
1.4 Organização do trabalho 27

2 VISUALIZAÇÃO DA INFORMAÇÃO 28
2.1 Treemap 30
2.1.1 Squarified Treemap 33
2.2 Variáveis Visuais 36
2.3 Glifos 39
2.4 Tarefas de Visualização da Informação 44

3 TRABALHOS RELACIONADOS 47
3.1 Estudos sobre glifos 47
3.2 Usando glifos em Treemaps 51

4 CONCEPÇÃO E DESIGN DO GLIFO 54
4.1 Configurações do Pré-teste 1 55
4.1.1 Suporte computacional 55
4.1.2 Perfil dos participantes 55
4.1.3 Execução do pré-teste 1 56
4.1.4 Resultado do Pré-Teste 1 59
4.2 Design do glifo 62
4.3 Configurações do Pré-teste 2 65
4.3.1 Resultado do Pré-Teste 2 67
4.3.2 Árvore de decisão 68
4.4 Glifo adaptativo 70

5 PROTÓTIPO DE APOIO 73

6 TESTES DE USUÁRIO 80
6.1 Teste de usuário - glifo em camadas 80
6.1.1 Suporte computacional 80
6.1.2 Perfil dos participantes 80
6.1.3 Descrição da base de dados 81
6.1.4 Hipóteses 81

6.1.5 Cenários do teste . . . 82
6.1.6 Tarefas de usuário . . . 84
6.1.7 Procedimentos do teste . . . 84
6.1.8 Resultado do teste . . . 87
6.1.8.1 H1 – Substituir hierarquias por glifos em camadas . . . 89
6.1.8.2 H2 – Combinar hierarquias com glifos em camadas ao pesquisar o maior valor . . . 91
6.1.8.3 H3 – Usar profile glifos na identificação de valores negativos . . . 91
6.1.8.4 H4 – Usar profile glifos na identificação do valor zero . . . 92
6.1.8.5 H5 – Combinar hierarquias com glifos em camadas ao pesquisar um único valor . . . 93
6.1.8.6 H6 – Usar profile glifos na identificação de padrões em um grupo de dados . . . 94
6.1.8.7 Discussão . . . 96
6.2 Teste de usuário - glifo adaptativo . . . 97
6.2.1 Suporte computacional . . . 97
6.2.2 Perfil dos participantes . . . 97
6.2.3 Descrição da base de dados . . . 97
6.2.4 Hipótese . . . 98
6.2.5 Cenário do teste . . . 98
6.2.6 Tarefas de usuário . . . 99
6.2.7 Procedimentos do teste . . . 99
6.2.8 Resultado do teste . . . 101
6.3 Avaliação de Usabilidade e Qualitativa do Glifo Adaptativo . . . 103
6.3.1 Escala de Usabilidade do Sistema - SUS . . . 104
6.3.2 Avaliação qualitativa - glifo adaptativo . . . 106

7 CONSIDERAÇÕES FINAIS . . . 111

REFERÊNCIAS . . . 115

APÊNDICE A – PARECER DO COMITÊ DE ÉTICA . . . 123

APÊNDICE B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO-TCLE . . . 124

APÊNDICE C – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO-TCLE . . . 125

APÊNDICE D – AVALIAÇÃO GLIFO EM CAMADAS - AMBIENTE G . . . 126

APÊNDICE E – AVALIAÇÃO GLIFO EM CAMADAS - AMBIENTE GH . . . 127

APÊNDICE F – AVALIAÇÃO GLIFO EM CAMADAS - AMBIENTE H 128

APÊNDICE G – AVALIAÇÃO GLIFO ADAPTATIVO - AMBIENTE GA 129

1 INTRODUÇÃO

O termo Visualização da Informação (InfoVis) ou Visualização de Dados engloba os tipos de representação visual que suportam exploração, análise e comunicação de dados abstratos (FEW, 2009). De acordo com CARD; MACKINLAY; SHNEIDERMAN (1999), InfoVis é a representação visual de dados abstratos somada a interação com os mesmos para melhorar a percepção dos usuários sobre o conjunto de dados que está sendo representado.

Durante a construção das técnicas de InfoVis para a visualização de dados abstratos, duas etapas se destacam: a Representação e a Apresentação (WARE, 2004). A Representação define como os itens de dados devem ser codificados visualmente (forma geométrica: barra, linha, círculo, etc), e a Apresentação define como essas representações visuais devem ser organizadas (grade, coordenadas X-Y, ordem crescente, etc). De maneira geral, uma técnica de InfoVis permite codificar visualmente os dados para um determinado *layout* de apresentação (SPENCE, 2014).

Um fato importante é a capacidade do ser humano de notar diferenças em canais visuais, nos quais itens como linhas, formas, orientação, cor, entre outras são distinguidas rapidamente e sem muito esforço, tal fato é denominado de pré-atenção (ROPINSKI; OELTZE; PREIM, 2011; MUNZNER; MAGUIRE, 2015). Boas visualizações fazem uso dos canais visuais que possuem boa pré-atenção para permitir aos usuários realizarem de forma mais fácil tarefas como busca, comparação e identificação de *outliers*, padrões ou tendências nos dados visuais.

Existem várias técnicas de Infovis disponíveis, voltadas a diversos contextos. Algumas representam melhor dados categóricos, enquanto que outras dados contínuos, outras dados hierárquicos, outras permitem a combinação de mais de um tipo de dado. Algumas técnicas utilizam o espaço tridimensional (CARD; MACKINLAY; SHNEIDERMAN, 1999; YASOJIMA; MEIGUINS; MEIGUINS, 2011), sendo o mais comum o espaço bidimensional, utilizando *layouts* em grafo ou em árvores (MUNZNER, 1997; MAGUIRE et al., 2012), ou *layout* específicos como espiral (WARD; GRINSTEIN; KEIM, 2015). Entre as diversas técnicas de InfoVis pode-se citar: Treemap, Coordenadas Paralelas, Diagrama de Dispersão, Sunbust, etc.

A visualização de estruturas de informações hierárquicas é um tópico de pesquisa atual na comunidade de InfoVis, uma vez que muitos dados do mundo real apresentam características de hierarquia. Essas técnicas de InfoVis buscam apresentar cada vez mais dados no espaço visual disponível, mas se preocupando sempre com a qualidade do entendimento dos dados pelos usuários (WANG et al., 2006). Os principais métodos de destaque para esta área de InfoVis são: Diagramas de Node-Link (MAGUIRE et al.,

2012; SAKET et al., 2014) e Visualizações de Preenchimento de Espaço (JOHNSON; SHNEIDERMAN, 1991; BAUDEL; BROEKSEMA, 2012).

A abordagem de Node-Link conecta nós por segmentos de linhas permitindo visualizá-los, por exemplo, em um layout de árvore (JIANU et al., 2014). Contudo, não faz uso eficiente do espaço visual disponível, pois muitas áreas do espaço visual disponível não são utilizadas, e não apresenta uma visão geral efetiva, uma vez que a expansão dos dados nas bordas geralmente é manual. Para mitigar esses problemas a abordagem de preenchimento de espaços tem como objetivo primário a utilização mais eficiente possível do espaço visual disponível para representação dos dados (LONG et al., 2017).

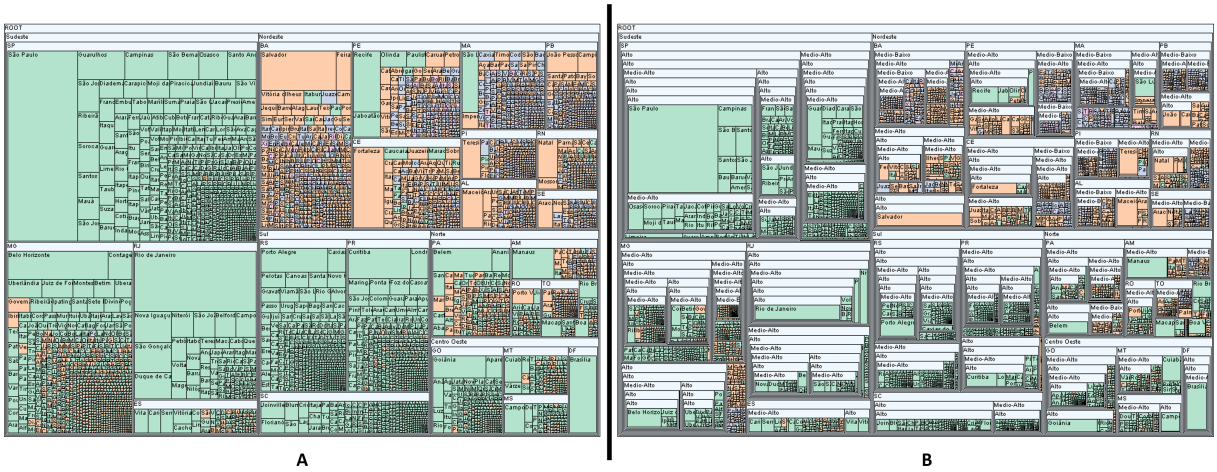
Dentro desta abordagem, destacam-se as técnicas de preenchimento de espaços retangulares, por exemplo, Treemap e suas variações (KONG; HEER; AGRAWALA, 2010; SHNEIDERMAN; WATTENBERG, 2001; BEDERSON; SHNEIDERMAN; WATTENBERG, 2002). O Treemap é uma técnica de visualização, bastante difuldida na área, que foi projetada para visualizar dados hierárquicos com uma abordagem de preenchimento de espaços em áreas retangulares (JOHNSON; SHNEIDERMAN, 1991), ou seja, a técnica se adapta ao espaço disponível para renderização gráfica, ocupando 100% da área disponibilizada.

A técnica Treemap destaca-se por proporcionar bom *overview* sobre o conjunto de dados, com ênfase na distribuição, relacionamentos, quantitativos ou proporções de dados, com base no *layout* hierárquico de grupos que a técnica apresenta (SHNEIDERMAN, 1992). Apesar de muito difundida e utilizada, a técnica Treemap ainda apresenta alguns desafios (HAHN; DÖLLNER, 2017; BLANCH; LECOLINET, 2007), tais como: dificuldade para representar através do tamanho dos itens o valor zero e números negativos; apresenta poucas opções para representação visual dos dados, limitados a tamanho, cor e rótulo.

Para uma melhor navegação no *layout* da técnica, é possível adicionar níveis hierarquicos, e assim, aumenta-se a quantidade de dimensões representadas visualmente. No entanto, a representação visual de mais atributos através de hierarquias influencia na área dos itens —nós folha— do Treemap, pois a área dos itens são reduzidas em detrimento da representação visual das hierarquias, o que influência na qualidade da visualização para análise dos dados.

A Figura 1-A ilustra um Treemap representando cinco (5) atributos de uma base de dados, dois (2) dos atributos são representados através de níveis hierárquicos e os demais são representados através do tamanho, da cor e do rótulo, cada, nesse Treemap ainda é possível fazer a leitura de seus itens —nós folha. Com o intuito de representar visualmente mais atributos da base de dados ao Treemap da Figura 1-A, a Figura 1–B ilustra que foram adicionados mais seis (6) níveis hierarquicos, elevando para onze (11) o número de atributos representados, porém a visibilidade dos itens tornou-se mais difícil, sobrecarregando a leitura dos dados para o usuário.

Figura 1. A - Treemap representando visualmente 5 atributos de uma base de dados, 1 através do tamanho, 1 através da cor, 1 através do rótulo e 2 níveis hierárquicos. B - Adição de 6 níveis hierarquicos ao Treemap de A, representando visualmente 11 atributos.



Fonte: Autor do trabalho.

Ao longo do tempo, a técnica Treemap recebeu várias sugestões de melhoria para enfrentar os desafios ressaltados, tais sugestões variam desde novos algoritmos (BRULS; HUIZING; WIJK, 2000; SHNEIDERMAN; WATTENBERG, 2001) à novos aspectos de design (HAM; WIJK, 2002; BALZER; DEUSSEN, 2005; ZHAO; MCGUFFIN; CHIGNELL, 2005; TU; SHEN, 2007; AUBER et al., 2013; HAHN; DÖLLNER, 2017), visando melhorar o aspecto da qualidade da visualização para muitos dados, considerando os seguintes critérios: itens —nós folha— mapeados para seções muito pequenas no layout, muitos níveis hierarquicos, maior número de dimensões representadas, poucas mudanças no *layout* gerado para auxiliar no processo de memorização e percepção do usuário da mudança ocorrida entre cenários, entre outros.

Visando mitigar os problemas da técnica, alguns autores combinaram o Treemap com outras técnicas, por exemplo: Treemap + Node-Links (ZHAO; MCGUFFIN; CHIGNELL, 2005), Table Lens + Treemap (TELEA, 2006), Treemap + Diagrama de Barras (HUANG; HUANG; ZHANG, 2009), Treemap e Mapas Choropleth (JERN; ROGSTADIUS; ÅSTRöM, 2009), etc, mas de maneira geral as propostas de combinação de técnicas de InfoVis propõem soluções para problemas pontuais, deixando ainda, de maneira geral, os problemas em aberto.

Outra possibilidade de mitigar os problemas apresentados é aumentar o espaço visual para representar os dados, essa abordagem demanda mais investimento em recursos tecnológicos, tais como: monitores, placas gráficas, CPUs, memórias, etc. Outra opção é adicionar e relacionar, através de brushing (SPENCE, 2007), outras técnicas de InfoVis,

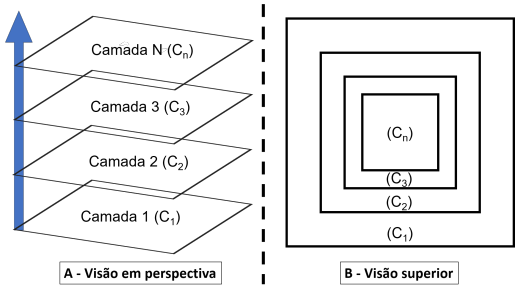
ou a própria técnica Treemap com outra visão de dados (LOURENCO et al., 2012), mas o relacionamento entre várias técnicas de InfoVis ou visão de dados demanda um esforço cognitivo maior do usuário (BALDONADO; WOODRUFF; KUCHINSKY, 2000).

Uma alternativa para a representação de dados multidimensionais em InfoVis é a utilização de glifos. Glifos são objetos gráficos que podem representar um ou vários atributos de uma base de dados através do mapeamento de uma ou várias variáveis visuais, podendo ser: forma, cor, tamanho, orientação, entre outras (FUCHS et al., 2014). De modo geral, os glifos podem compor uma visualização como objetos visuais, sendo utiliziados de forma independente e construtiva no espaço visual para representar atributos de uma base de dados, podendo sofrer variações em seu tamanho (BORGO et al., 2013).

Glifos podem representar dados categóricos e/ou quantitativos, em diversos artigos são considerados como pequenas e compactas representações visuais que são capazes de representar dados multidimensionais (FUCHS; HAUSER, 2009; YATES et al., 2014), e tem sido utilizado amplamente em combinação a outros *layouts* ou com outras técnicas, como Linhas de Fluxo (PILAR; WARE, 2013), Estrutura de Árvores (MAGUIRE et al., 2012), Treemap circulares (FISCHER; FUCHS; MANSMANN, 2012), Layout de Matriz (FLURY; RIEDWYL, 1981), Mapas (TOMINSKI; SCHULZE-WOLLGAST; SCHUMANN, 2005), Glifos 3D (ZHANG et al., 2016), Coordenadas paralelas (OPACH; RØD, 2017), etc.

Assim essa tese objetiva apresentar soluções para mitigar os desafios atuais da técnica Treemap, para isso propõe integrar as boas características da técnica Treemap, de representação hierárquica de grupos, com a boa representatividade de dados dos glifos. Para tal, esta tese apresenta um novo design de glifo, que além de ter sido aplicado ao *layout* do Treemap, poder ser aplicado em combinação a outros *layouts* ou com outras técnicas. No design do glifo proposto as variáveis visuais são apresentadas em camadas parcialmente sobrepostas e podem representar dados categóricos e quantitativos (ver Figura 2).

Figura 2. Modelo do glifo em camadas, no qual as camadas se sobrepoem uma sobre as outras em 70%. A - Visão em perspectiva das camadas (C_1, C_2,..,C_n) do glifo; B - Visão superior das camadas (C_1, C_2,..,C_n) do glifo.

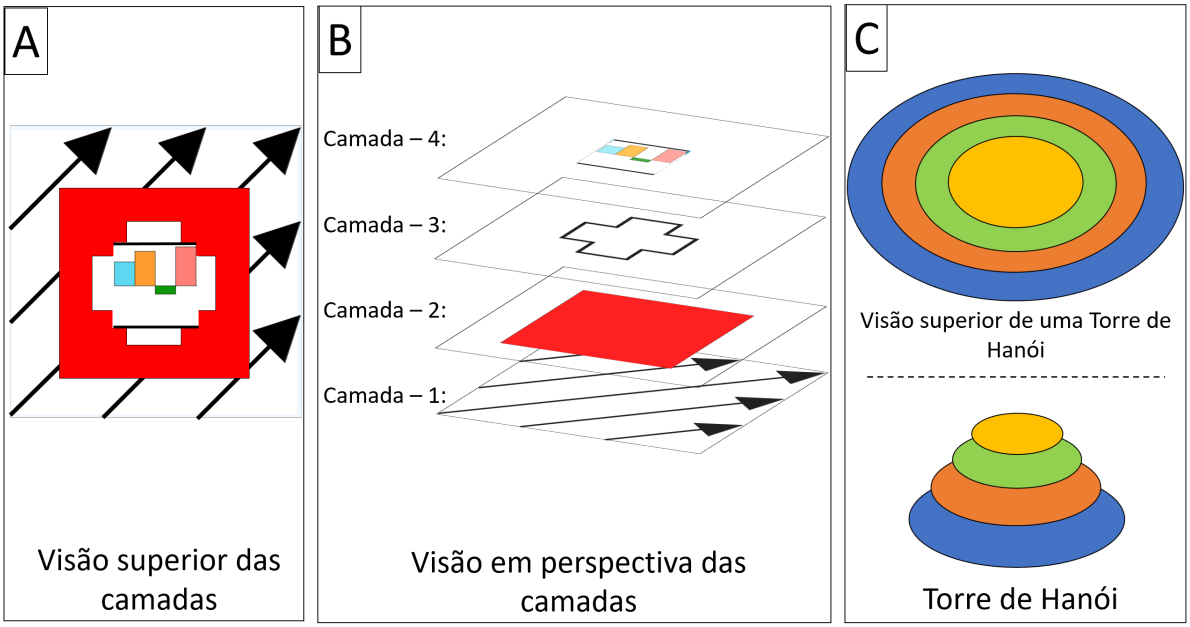


Fonte: Autor do trabalho.

A Figura 2 apresenta um modelo de design de glifo que foi definido com base em dois estudos realizados de percepção do usuário (detalhados no decorrer da tese). Esse modelo pode possuir n camadas concêntricas sobrepostas, para evitar a sobreposição total —sobreposição de 100% de uma camada sobre a outra— entre as camadas, cada uma delas sobrepoe a camada anterior em 70%. Neste modelo de glifo, cada camada é vinculada a uma variável visual, assim é possível mapear, em cada camada, um atributo da base de dados a uma variável visual, consequentemente, cada camada do glifo consegue representar uma dimensão de dados, dando origem a um glifo multidimensional.

Para uma melhor compreensão do modelo de glifo desta tese, pode-se fazer uma alusão a um dos pinos, visto de cima, de uma Torre de Hanói[1], no qual cada disco causa uma uma sobreposição parcial no disco anterior e os discos menores não podem ficar sob os maiores, ou seja, no contexto desta tese, cada disco da Torre de Hanói equivale a uma camada do glifo (ver Figura 3).

Figura 3. Cada variável visual está disposta em uma camada do glifo, de forma que cada camada cause apenas uma sobreposição parcial na camada posterior (B), como em uma torre de hanoi (C).



Fonte: Autor do trabalho.

Adicionalmente, o modelo de glifo desta tese também possui uma abordagem que torna o glifo adaptativo a área de plotagem, isto é, o modelo de glifo, após uma análise

[1] A Torre de Hanói é um quebra-cabeça inventado por Edouard Lucas em 1883. Existem três pinos e n discos de todos os tamanhos diferentes. Inicialmente, os discos são empilhados em um pino em tamanho decrescente de baixo para cima, e durante o jogo os discos menores nunca podemão ficar por baixo dos maiores (PETERSEN, 2019).

prévia da área de plotagem, oculta as camadas que não seriam visíveis pelo usuário. Desta forma, dependendo da quantidade de dados a ser representada pela técnica ou layout, a qual o glifo estará sendo aplicado, e do espaço visual disponível, o glifo adaptativo busca apresentar para o usuário informação visual útil mesmo em regiões de área visual reduzida. Em outras palavras, o glifo adaptativo representa visualmente apenas as informações que são visiveis para o usuário, ocultando as camadas que não seriam vistas e que poderiam causar uma sobrecarga visual na visualização.

Esta tese aplicou o modelo de glifo proposto ao layout do Treemap Squarified, pois de acordo com BRULS; HUIZING; WIJK(2000) ele utiliza o espaço disponível para desenho de forma mais eficiente, permite que os itens sejam mais fáceis de serem selecionados em ambientes interativos e são mais fáceis de estimar o tamanho dos itens. Essa implementação do algoritmo do Treemap busca criar retângulos mais próximos da forma de um quadrado, ou seja, o aspecto —max(altura/largura; largura/altura)— mais próximo possível de um (1), essa característica visual de maneira geral auxilia em uma melhor comparação visual de itens e grupos de dados (BRULS; HUIZING; WIJK, 2000). Este último fato mencionado é a razão das camadas do modelo do glifo desta tese apresentar-se em forma de quadrados, sendo assim, o modelo de glifo apresentado pode ser adaptado ao layout da técnica.

Como consequência imediata desta pesquisa é possível ter Treemaps representando mais dimensões de dados Outra contribuição desta pesquisa é eliminar a necessidade de utilizar muitas hierarquias em um Treemap, pois os glifos podem representar tanto dados categóricos quanto quantitativos, possibilitando uma possível troca entre uma hierarquia por um item do glifo, assim as visualizações de dados criadas poderiam ser mais limpas, sem as informações de grupos, e com itens maiores para visualização.

Para avaliar a proposta, foram realizados testes com usuários para tarefas de busca e consulta para os seguintes cenários: Treemap com glifos e sem hierarquia (G), Treemap sem glifos e com hierarquia (H), Treemap com glifos e com hierarquia (GH) e Treemap com glifos adaptativos e com hierarquia (GA). Foram conduzidos testes com base em tarefas de busca aplicada a grupos de usuários-especialistas com conhecimento sobre as técnicas do InfoVis adotadas, e os resultado dos cenários comparados entre si.

Foram realizados duas avaliações, primeiramente foram comparados os cenário G, GH e H, em seguida foi realizada a avaliação do cenário GA. O resultado das avaliações mostraram que o cenário GH apresentou uma melhor acurácia para as tarefas realizadas com relação aos cenários G e H. O cenário G obteve o pior desempenho quanto ao tempo e acurácia das tarefas e, com relação aos cenários GH e H, não houve diferença significativa no tempo de resolução das tarefas. Os participantes do teste do cenário GA indicaram que os glifos adaptativos melhoram a qualidade visual da informação no Treemap e que o glifo adaptativo, aplicado ao Treemap, apresentou boa usabilidade.

1.1 Objetivos

O objetivo geral deste trabalho é ampliar a representação visual de dados na técnica Treemap com a utilização de glifos.

Os objetivos específicos são:

- A criação de um modelo de glifo que permita o mapeamento de atributos categóricos e quantitativos;
- A criação de um modelo de glifos que permita a extração de informação visual parcial de acordo com a qualidade de leitura das variáveis visuais do glifo, sem comprometer significativamente a capacidade de extração de informação;
- Implementação da técnica Treemap utilizando glifos em camadas;
- Avaliar a relação do uso de glifos e hierarquias no Treemap para resolução de tarefas de InfoVis;
- Avaliar o uso de glifos de dados quantitativos e categóricos em tarefas de InfoVis na técnica Treemap.

1.2 Contribuições

As principais contribuições deste trabalho são:

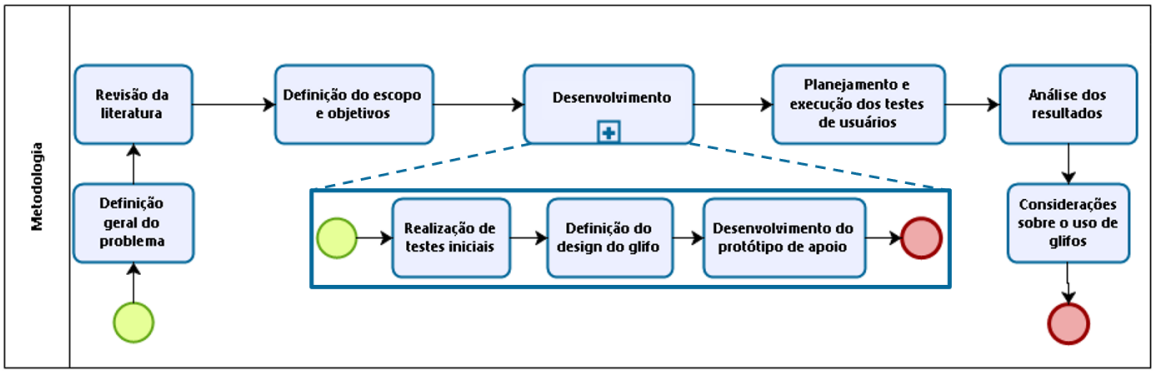
- Melhoramento na capacidade de representação visual da técnica Treemap;
- Mitigar problemas de representações visuais de dados relacionados com a técnica Treemap, tais como: números muito pequenos ou negativos;
- Design e avaliação de glifos em camadas que podem ser adicionados no Treemap e em outras técnicas de visualização ou layouts;
- Implementação e avaliação automática da qualidade de leitura dos glifos, propondo glifos adaptativos de acordo com sua qualidade visual para leitura.

1.3 Metodologia

A visão geral da metodologia de pesquisa dessa tese pode ser observada na Figura 4. A primeira etapa da metodologia foi a definição geral do problema. Em seguida, foi realizada a revisão da literatura em busca de fontes bibliográficas que fundamentassem e justificassem esta tese. Posteriormente, com o direcionamento do uso de glifos para

enfrentar os desafios apresentados pelo Treemap, foram definidos os objetivos e o escopo da pesquisa.

Figura 4. Visão geral da metodologia de pesquisa



Fonte: Autor do trabalho.

O escopo da pesquisa envolve a elaboração, uso e avaliação de glifos em camadas na técnica Treemap. Após a definição dos objetivos e escopo da pesquisa, iniciou-se a etapa de desenvolvimento, no qual envolve três (3) etapas, são elas:

- Realização de testes iniciais: apresenta a avaliação realizada considerando questões iniciais como oclusão das variáveis visuais, a combinação de valores das variáveis visuais, entre outros para subsidiar com dados a etapa de design do glifo;
- Definição do design: apresenta as estratégias utilizadas para a construção do glifo;
- Desenvolvimento do protótipo de apoio: apresenta o protótipo desenvolvido (disponível em labvis.ufpa.br/treemapglyphs) integrando os glifos em camadas à técnica Treemap;

A etapa de planejamento e execução dos testes com usuários apresenta os cenários utilizados nos testes e como os cenários foram combinados para a realização dos testes de usuários. Adicionalmente, nessa etapa são discutidas questões como: onde foram realizados os testes? qual o suporte computacional foi utilizado nos testes? qual foi o perfil dos usuários? quais foram as hipóteses dos testes? quais foram as métricas para análise dos resultados? quais foram as tarefas executadas pelos usuários durante os testes?.

A etapa de análise dos resultados trata da análise quantitativa e qualitativa dos resultados dos testes de usuários, abordando quais técnicas estatísticas foram utilizadas e discute os resultados dos testes. Por fim, são tecidas as considerações finais sobre o uso de glifos em Treemap, apontando as conclusões das avaliações, as limitações desta pesquisa e apontando cenários de futuras pesquisas com glifos em Treemaps.

1.4 Organização do trabalho

Este trabalho está organizado nas seguintes seções: introdução, teoria relacionada, trabalhos relacionados, concepção e design do glifo, protótipo de apoio, testes de usuários, resultados e considerações finais.

1. Introdução: apresenta a visão geral desta proposta, contextualizando a área, justificando a relevância do trabalho e descreve a metodologia de pesquisa utilizada;
2. Visualização da Informação: apresenta conceitos importantes para o entendimento do trabalho;
3. Trabalhos relacionados: apresenta estudos sobre glifos aplicados em várias técnicas de InfoVis e estudos sobre o uso de glifos especificamente em Treemaps;
4. Concepção e design do glifo: apresenta as etapas desenvolvidas para a elaboração do design do glifo;
5. Protótipo de apoio: apresenta o protótipo implementado nesta tese para a expansão das dimensões do Treemap;
6. Testes de Usuários: apresenta as avaliações realizadas e, seus respectivos resultados, para testar a efetividade do glifo proposto;
7. Considerações Finais: finaliza a tese e apresenta os trabalhos futuros.

2 VISUALIZAÇÃO DA INFORMAÇÃO

A Visualização da Informação (InfoVis) estuda o uso de representações visuais e técnicas interativas de dados abstratos com o propósito de ampliar a compreensão do usuário sobre o conjunto de dados de análise, ou seja, é uma área que estuda formas de melhorar a percepção, comunicação e cognição sobre dados abstratos através da representação visual e mecanismos de interação (FEW, 2009; CHEN, 2002). Dentro deste contexto, GÖRG et al. (2007) explica que é através das variáveis visuais que a representação visual consegue transmitir informações aos usuários, diferente da representação sentencial, que faz uso da semântica de palavras e letras. Adicionalmente, WARE (2004) lista algumas vantagens da visualização:

- A identificação de problemas com os dados em si, ou seja, detecção de *outliers* ou anomalias;
- A compreensão de grande ou pequena quantidade de dados;
- A criação de hipóteses variadas relacionadas aos dados.

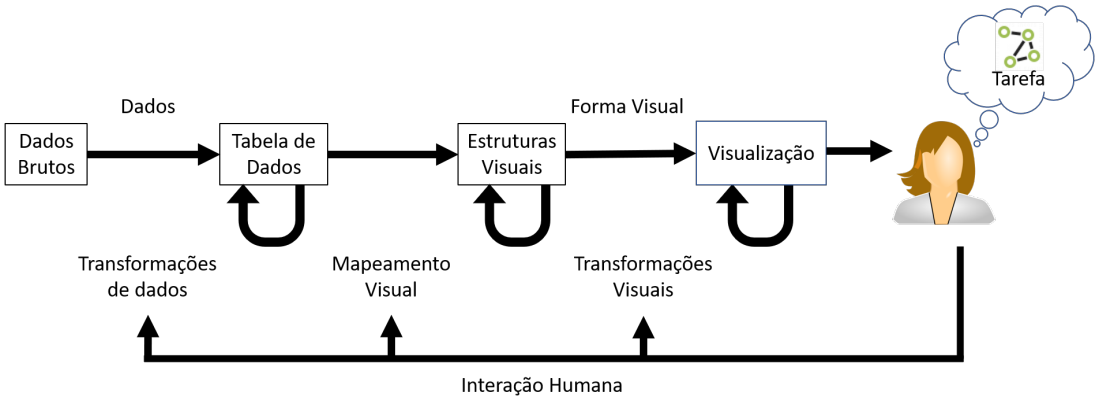
A relevância dessa forma de comunicação dá-se devido à capacidade de uma única figura conter uma grande quantidade de informações e ser interpretada muito mais rapidamente que uma página de texto ou dados tabulares (GÖRG et al., 2007; WARE, 2004). Apesar de suposta simplicidade, ainda é complexo encontrar uma maneira adequada para representar os dados de forma que seja possível promover *insights* e o entendimento eficaz desses dados.

CARR (1999) afirma que visualização deve ser utilizada se existe uma grande quantidade de dados, se os objetivos dos usuários não são facilmente quantificáveis e se não existe um algoritmo simples que permita alcançar os objetivos do usuário. CARD; MACKINLAY; SHNEIDERMAN (1999) preconiza a divisão do processo para gerar uma imagem para um conjunto de dados em três etapas (ver Figura 5):

- Transformação dos dados: processamento (ex: eliminação de dados redundantes) e organização (alocar uma linha para cada dado e uma coluna para cada atributo diferente dos dados) de um conjunto de dados brutos em uma representação lógica mais estruturada (ex: em uma ou mais tabelas);
- Mapeamento Visual: construção de uma estrutura visual que represente os dados da tabela visualmente;

- Transformações Visuais: modificação e aumento das estruturas visuais interativamente através de operações básicas como testes de localização, controles de ponto de vista e distorções da imagem.

Figura 5. Modelo de referência para visualização.



Fonte: Adaptada de (CARD; MACKINLAY; SHNEIDERMAN, 1999)

De acordo com NASCIMENTO; FERREIRA (2005), desenvolver sistemas de visualização baseando-se na capacidade humana de percepção e sistemas de processamento de informações é um desafio. É necessário entender melhor como o ser humano interage com a informação, como ele a percebe visualmente, como a mente trabalha quando está procurando por informações conhecidas e desconhecidas, como a mente resolve os problemas e como o ser humano compreende os dados apresentados. Um projeto de sistemas de visualização deve estabelecer interfaces flexíveis de modo a fornecer ferramentas de navegação e métodos de pesquisa apropriados para cada um dos tipos de usuários existentes, modelos de dados e tarefas.

Para caracterizar melhor uma ferramenta de visualização de informação, SHNEIDERMAN (1996) determinou sete tarefas abstratas que os usuários sempre utilizarão nestas ferramentas:

- Visão geral: os usuários precisam obter uma visão geral dos dados que serão analisados;
- Zoom: o usuário se concentra em um grupo de dados de interesse, aumentando a visualização de detalhes;
- Filtro: o usuário retira de sua visão as informações que não são de seu interesse;
- Detalhes sob demanda: torna possível ao usuário visualizar informações extras acerca de um item ou grupo de itens que não estão representados graficamente, em *popup* ou janela auxiliar;

- Relacionar: permite que o usuário observe relacionamentos entre itens;
- Histórico: o usuário pode necessitar fazer ou desfazer uma ação, assim é necessário manter um histórico das ações realizadas;
- Extração: permite a extração dos resultados das operações de visualização a partir da visualização.

Adicionalmente, SHNEIDERMAN (1996) descreveu sete tipos de dados, sendo eles:

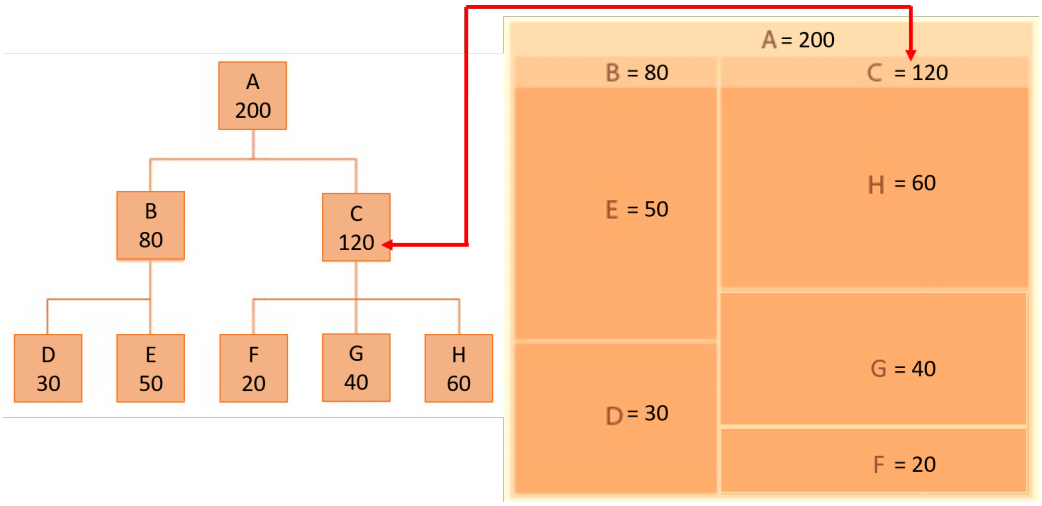
- 1-dimensional: é um tipo de dado linear, como documentos textuais ou qualquer lista organizada de maneira sequencial;
- 2-dimensional: trata de mapas, como mapas geográficos ou plantas de casas;
- 3-dimensional: trabalha com objetos do mundo real como moléculas, objetos que possuem volume além de estabelecer relacionamento com outros objetos;
- Temporal: trabalha com a linha de tempo como em registros médicos ou projetos de gerenciamento; graficamente, em *popup* ou janela auxiliar;
- Multidimensional: refere-se às bases de dados relacionais e estatísticas que são manipuladas como dados multidimensionais em que itens com "n" atributos tornam-se pontos em um espaço n-dimensional;
- Árvore: é uma coleção de itens, na qual cada item pode ter uma ligação para um item pai, com exceção da raiz;
- Rede: é um tipo de dado conveniente quando relacionamentos entre itens não podem ser capturados pela estrutura de árvore, é útil quando é necessária a ligação de um item para um número arbitrário de outros itens.

2.1 Treemap

Treemaps foram originalmente desenvolvidos como uma proposta de visualizar grandes diretórios de arquivos em um computador, utilizando pouco espaço de visualização para exibir hierarquias de arquivos, fornecendo uma visão geral (*overview*) rápida de como estavam organizados os arquivos (estrutura do diretório) (SHNEIDERMAN, 1992).

Treemap é uma técnica de InfoVis que possibilita visualizar a estrutura hierárquica de um diagrama de árvore, representando valores quantitativos dos dados através do tamanho da área. A cada categoria é atribuída uma área retangular com seus retângulos de subcategoria aninhados dentro dela, como pode ser observado na Figura 6.

Figura 6. Treemap e diagrama de árvore.



Fonte: Adaptada de (SPENCE, 2007)

Quando um valor é atribuído a uma categoria, o tamanho da área é exibido proporcionalmente a esse valor. Considerando todas as categorias do conjunto de dados, a Figura 6 mostra que o valor 120 foi atribuído a categoria C, logo sua área é proporcional ao seu valor. Além disso, o tamanho da área da categoria pai é o total de suas subcategorias. Essa relação entre as categorias favorece a comparação entre elas através do tamanho das áreas, quanto maior a área maior o valor de uma categoria.

SHNEIDERMAN(1992) preconizou o algoritmo em questão, que é implementado de acordo com o pseudo-código do algoritmo 1. No pseudo-código, o argumento “root” representa um ponteiro para o nó raiz da árvore ou da sub-árvore; “P” e “Q” são vetores de tamanho dois (2) contendo o par ordenado (x, y) representando os cantos opostos do retângulo, ou seja, o canto superior esquerdo e o canto inferior direito do retângulo; “axis” são valores que variam entre zero (0) e um (1) que indicam a troca a ser feita verticalmente e horizontalmente; “color” indica a cor a ser usada no retângulo corrente. O algoritmo faz uso de mais duas funções (SHNEIDERMAN, 1992):

- *PaintRectangle()*: Função responsável por desenhar o retângulo usando a cor fornecida e redefine a variável cor;
- *Size()*: Função que retorna o número representando o tamanho do retângulo.

Algoritmo 1: Pseudo-código do algoritmo da técnica Treemap. Fonte: Adaptada de SHNEIDERMAN(1992).

```
1  início
2      Função Treemap(root, P [0..1], Q [0..1], axis, color) :
3          PaintRectangle(P, Q, color); // desenha em toda a área
4          width ← Q [axis] - P [axis]; // localização da próxima fatia
5          para i até children.length faça
6              Q [axis] ← P [axis] + (Size(children[i]) / Size(root)) * width;
7              Treemap(children[i], P, Q, 1-axis, color); // repete a cada fatia
8              P [axis] ← Q[axis];
9          fim
10     end
11 fim
```

Ao ler o algoritmo do pseudo-código (algoritmo 1) nota-se que ele desenhará os retângulos da esquerda para a direita e de cima para baixo. Inicialmente, o retângulo é dividido com uma partição vertical de acordo com o número de nós filhos do nó rais da árvore. Posteriormente, cada sub-retângulo é dividido utilizando o mesmo critério, mas com partições horizontais. Este procedimento é repetido para os nós folhas da árvore.

A maneira como os retângulos são divididos e ordenados em sub-retângulos depende do algoritmo utilizado. Para a escolha de um dos algoritmos é essencial considerar duas questões: a estabilidade do *layout* e legibilidade dos dados visuais. A estabilidade do *layout* trata da capacidade de lidar com pequenas alterações nos dados e como tais alterações refletirão no *layout*. Já a legibilidade é a capacidade de gerar itens de fácil interação através de um clique do mouse e que são de fácil reconhecimento pelo olho humano (YANG; CAO, 2019).

Muitos algoritmos de Treemaps foram desenvolvidos, e podem utilizar uma variadade de formas além das rentangulares (ver Figura 7), como por exemplo: circulares (Haisen Zhao; Lu, 2015), poligonais (BALZER; DEUSSEN, 2005) ou geográficas (AUBER et al., 2013). Dentre as implementações baseadas em retângulos, as seguintes são citadas como exemplo:

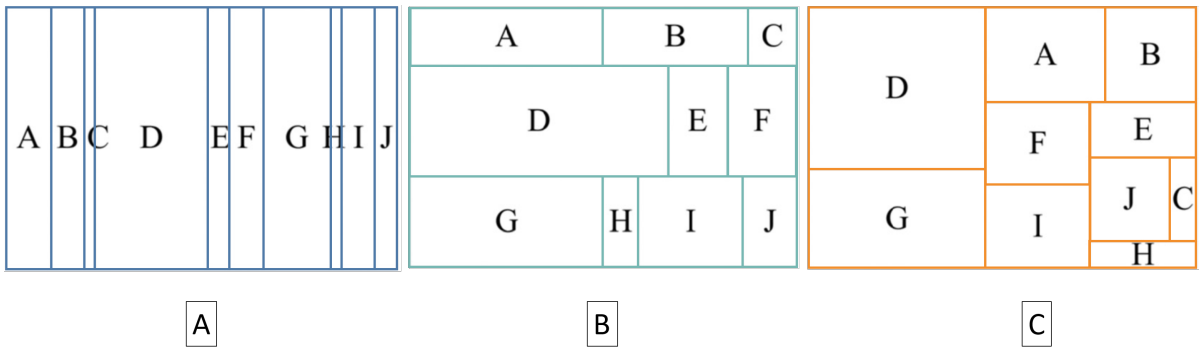
Slice and dice: essa implementação, se utilizada para sub-hierarquias com um grande número de itens, resulta em formas com altas proporções de aspecto e, portanto, baixa legibilidade (HAHN; DÖLLNER, 2017; YANG; CAO, 2019) (ver Figura 7-A).

Strip: essa implementação melhora a estabilidade do *layout*, pois funciona processando os retângulos de entrada na mesma ordem dos dados e dispondo-os na horizontal ou vertical com de espessuras variadas (BEDERSON; SHNEIDERMAN; WATTENBERG, 2002) (ver Figura 7-B), e melhora a legibilidade da técnica. No entanto essa implementa-

ção não é recomendada em situações que necessitem de uma ordenação —crescente ou decrescente— dos dados;

Squarified: essa implementação tenta tornar os retângulos o mais próximo possível dos quadrados, ou seja, possui uma boa legibilidade dos dados. A ordem original dos nós é alterada, o que pode prejudicar a estabilidade da técnica, seguindo a ordenação do maior para o menor (BRULS; HUIZING; WIJK, 2000; HAHN; DÖLLNER, 2017) (ver Figura 7-C).

Figura 7. Exemplos de implementação da técnica Treemap. A – Slice and Dice; B – Strip Treemap; C – Squarifield;



Fonte: Adaptada de (BRULS; HUIZING; WIJK, 2000) (BEDERSON; SHNEIDERMAN; WATTENBERG, 2002) (YANG; CAO, 2019)

2.1.1 Squarified Treemap

O algoritmo *squarified* tem como objetivo subdividir o espaço de apresentação em retângulos que possuam um bom aspecto relacional, ou seja, que possuam comprimentos de largura e altura próximos, diferente das outras implementações dessa técnica, não faz uso de retângulos finos e alongados (BRULS; HUIZING; WIJK, 2000). Essa característica permite gerar retângulos mais próximos da forma de quadrados, ou seja, com o aspecto mais próximo possível de um (1) —max(altura/largura; largura/altura)—, o que facilita a percepção do usuário sobre os itens visuais.

BRULS; HUIZING; WIJK (2000) destacam que o algoritmo em questão utiliza o espaço disponível para desenho de forma mais eficiente, permite que os itens sejam mais fáceis de serem selecionados em ambientes interativos e são mais fáceis de estimar o tamanho dos itens.

O *squarified* Treemap preconizado por BRULS; HUIZING; WIJK (2000) é implementado de acordo com o pseudo-código do algoritmo 2. O algoritmo não considera a subdivisão da área para todos os níveis (hierarquias) simultaneamente. Ao invés disso, são produzidos,recursivamente, retângulos o mais proximos de quadrados para um conjunto de

nós irmãos, ou seja, considerando nós filhos que possuem o mesmo nó pai. É importante destacar a que a lista de nós de entrada do algoritmo apresenta-se em ordem decrescente, ou seja, os nós de maior valor serão representados primeiro.

No pseudo-código (ver algoritmo 2), o argumento *children* representa a lista de nós da árvore com seus respectivos valores a serem representados visualmente; a lista *row* representa uma lista de nós que serão plotados em uma mesma direção (estarão na mesma linha ou coluna), horizontal ou vertical, de forma que eles não impactaram em seus aspectos relacionais, ou seja, estão o mais próximos possíveis da forma de uma quadrado; o argumento w representa o comprimento do lado ao qual o Treemap será plotado.

No decorrer do pseudo-código do algoritmo 2 o símbolo ++ indica que uma dada lista é concatenada; Para indicar que um elemento x pertence a lista, é utilziada a seguinte representação: [x]; e o símbolo [](colhetes vazio) é utilziado para representar que uma lista está vazia. Adicionalmente, o algoritmo também faz uso das seguintes funções:

- $worst()$: Retorna o maior aspecto relacional de uma lista de retângulos, dado o comprimento do lado ao qual devem ser dispostos;
- $width()$: retorna o menor lado do subretângulo restante na qual a linha —$row : Lista < real >$— corrente deve ser disposta;
- $layoutRow()$: retorna uma nova linha —$row : Lista < real >$— de nós a serem representadas.

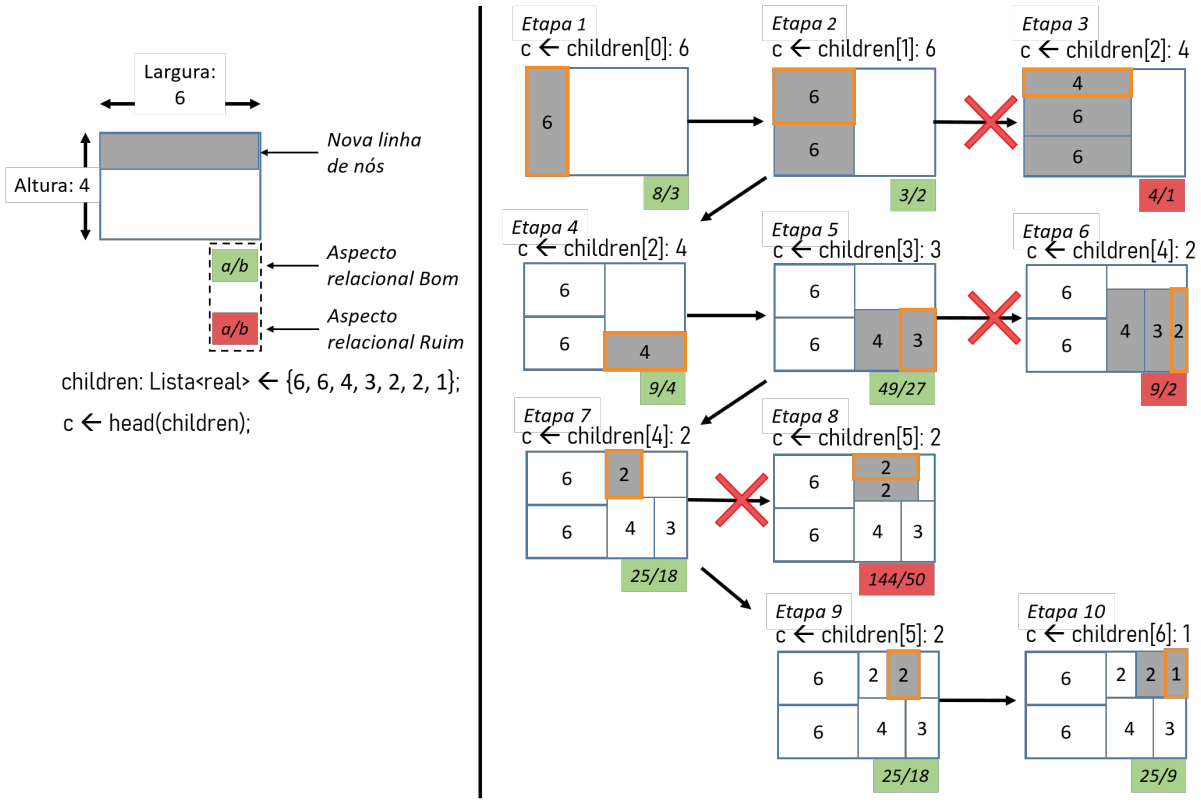
Algoritmo 2: Pseudo-código do algoritmo Squarified Treemap. Fonte: Adaptada de BRULS; HUIZING; WIJK(2000).

```
1  início
2      Função Squarify(children: Lista<real>, row: Lista<real>, w: real) :
3          c ← head(children);
4          se worst(row, w) ≤ worst(row++[c], w) então
5              Squarify(tail(children), row++[c], w);
6          senão
7              layoutRow(row);
8              Squarify(children, [], width());
9          fim
10     end
11 fim
```

BRULS; HUIZING; WIJK(2000) ilustra a execução o algoritmo *squarified* considerando os seguintes valores de entrada para a lista $children : Lista < real >$: 6, 6, 4, 3, 2, 2 e 1, as etapas de sua execução podem ser vistas na Figura 8. A primeira etapa do

algoritmo é dividir o retângulo inicial, considerando o primeiro valor (children[0] ← 6) da lista *children*, criando uma nova linha de nós na horizontal, pois o retângulo no qual será disposto do Treemap possui a largura maior que a altura. Após adicinar o primeiro retângulo, seu aspecto relacional é de 8/3.

Figura 8. Exemplo da execução do algoritmo Squarified Treemap.



Fonte: Adaptada de BRULS; HUIZING; WIJK (2000).

A segunda etapa é adicionar o próximo valor (children[1] ← 6) da lista, para isso é adicionado um novo retângulo na lista de nós já existente. Assim, um segundo retângulo é plotado acima do primeiro melhorando o aspecto relacional para 3/2. A terceira etapa tenta adicionar mais um retângulo a lista de nós, no entanto se o proxímo valor (children[2] ← 4) da lista *children* for adicionado, o aspecto relacional irá piorar, pois será de 4/1. Desta forma, a terceira etapa é descartada.

Na etapa 4, considerando o menor lado do retângulo resultante —*width*()—, é criada uma nova lista de nós —*layoutRow*()—, adicionando o terceiro valor da lista *children* (children[2] ← 4) com um aspecto relacional de 9/4. Em seguida, na etapa 5, é adicionado o quarto valor da lista *children* (children[3] ← 3), obtendo um aspecto relacional de 49/27. Em seguida, tenta-se adicionar o quinto valor da lista *children* (children[4] ← 2) a lista de nós (passo 6), entretanto, o aspecto relacional não atinge o melhor resultado, equivalente a 9/2. Por isso o algoritmo aceita o resultado da etapa 5, e começa a preencher a partição

canto superior direito (passo 7), para tal, é criada uma nova lista de nós —*layoutRow*()— e a ela é adicionada o quinto valor da lista *children* (children[4] ← 2). O passo a passo do algoritmo se repetem até todos os valores da lista *children* terem sido processados.

Esta tese utiliza essa implementação do algoritmo do Treemap devido legibilidade dos dados visuais, ou seja, a escolha se deu em razão do aspecto relacional gerar retângulos que se aproximam muito da forma de um quadrado. Desta forma, é importante destacar que o modelo de glifo desta tese segue o formato quadrado, em virtude do formato dos itens do *squarified* Treemap, o que facilita a leitura e a interação com a técnica.

Ainda com relação a legibilidade do algoritmo, SPENCE (2007) aponta que uma das vantagens da utilização dessa implementação da técnica é a possibilidade de utilizar o espaço interno dos retângulos para adicionar informação à visualização, seja ela textual ou visual. Com isso essa tese utiliza o espaço interno dos itens, gerados pela técnica, com glifos para aumentar a dimensionalidade de dados representados visualmente.

2.2 Variáveis Visuais

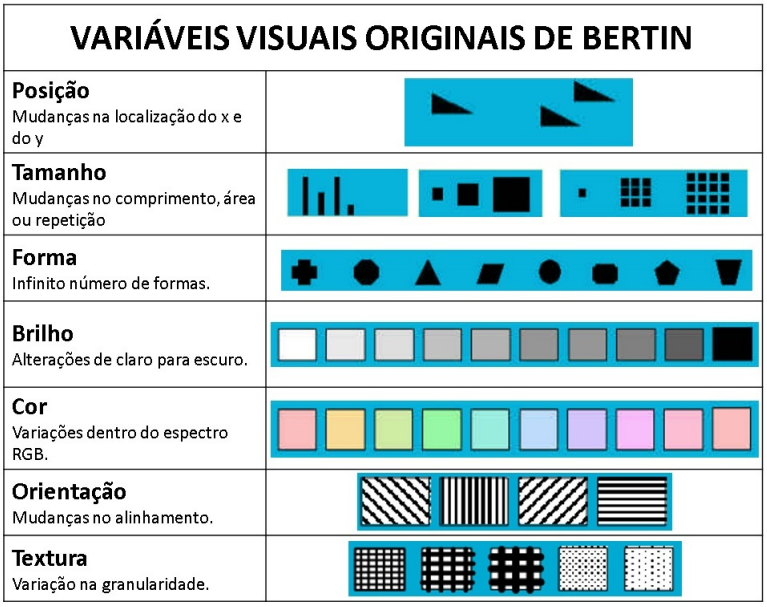
O usuário interage com visualizações de acordo com o que ele vê e interpreta, entender como os usuários veem deveria ajudar a produzir melhores visualizações, ou pelo menos evitar visualizações de baixa qualidade (WARE, 2004; SPENCE, 2007). BERTIN (1983) preconizou o conceito de marcas visuais, que são símbolos gráficos utilizados para representar os itens de dados, como por exemplo pontos, que podem ser figuras geométricas simples, linhas, áreas, volumes e superfícies.

Para BERTIN (1983) as marcas possuem propriedades gráficas que são os atributos visuais que as caracterizam, as variáveis visuais. O autor ainda define sete (7) variáveis visuais, sendo elas (ver Figura 9):

- Posição: refere-se a mudança nas coordenadas x ou y do objeto;
- Tamanho: refere-se a mudança no comprimento, área ou repetição do objeto, tal como a altura de uma coluna;
- Forma: pode ser representado por uma infinidade de formas, como por exemplo, quadrado, círculo, estrela, etc;
- Brilho (Valor): refere-se a alterações de claro para o escuro;
- Cor: refere-se às variações de cor dentro do espectro RGB;
- Orientação: refere-se a mudanças no alinhamento de um padrão, ou seja, a orientação das linhas ou padrões que variam da vertical para a horizontal;

- Textura: refere-se à variação do padrão da textura, como por exemplo a variação dos itens que compõem uma textura, tais como a densidade pontos que compõem a textura.

Figura 9. Variáveis visuais preconizadas por BERTIN (1983).

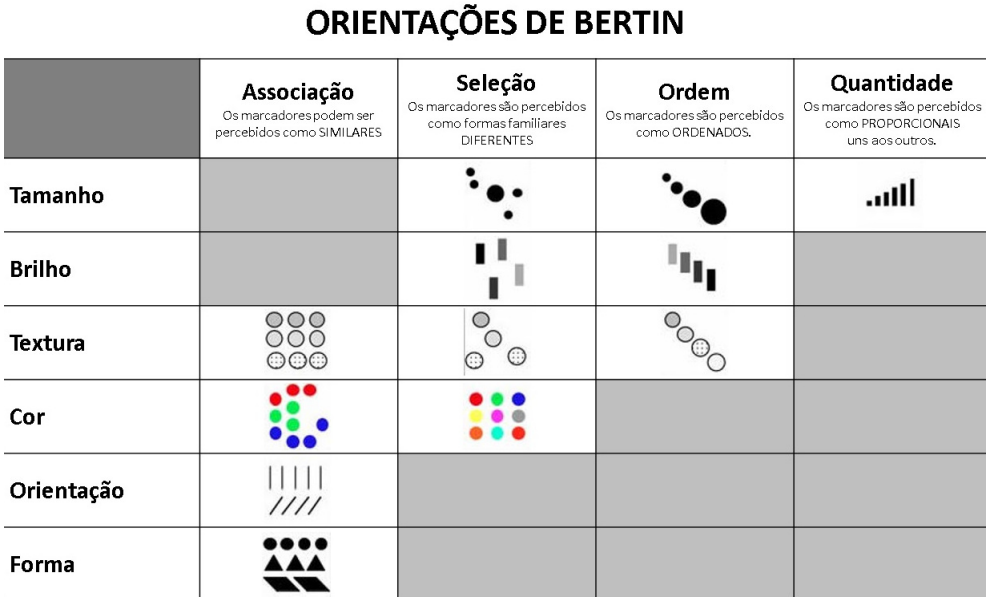


Fonte: Adaptada de (BERTIN, 1983)

No trabalho de BERTIN (1983) ainda são definidas as características mais apropriadas para representar cada aspecto da informação (ver Figura 10), definindo mapeamentos de eficácia entre as percepções e as variáveis visuais, e CARPENDALE(2003) as explica no contexto de InfoVis, dando um foco no tipo de tarefa que as variáveis visuais devem ser utilizadas:

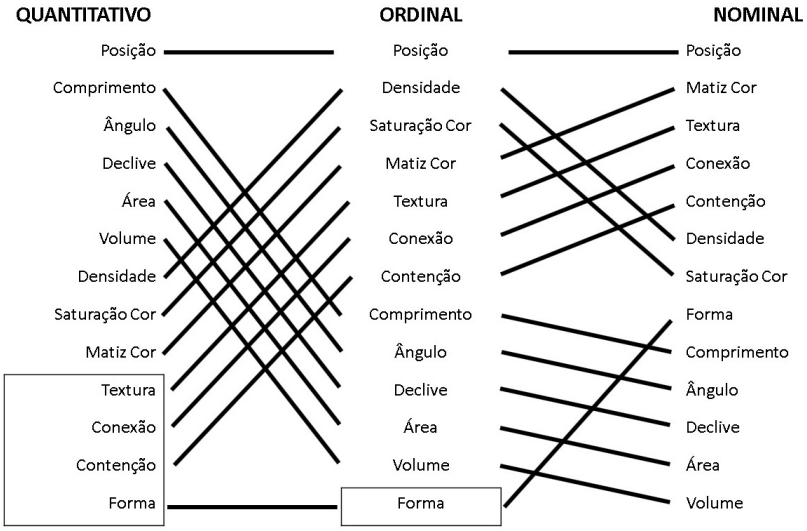
- Seleção: quando ocorre a alteração de uma marca em apenas uma variável, e essa ação facilita a seleção da mesma marca dentre um conjunto de marcas;
- Associação: quando através de mudanças em uma variável visual em comum, é possível agrupar diferentes marcas em um mesmo grupo;
- Quantificação: Duas marcas possuem a mesma variável visual, porém com valores distintos, permitindo uma distinção numérica, ainda que não exata;
- Ordenação: A mudança ocorrida em uma variável visual permite a percepção de uma ordenação na mesma variável;
- Quantidade de Valores (*Length*): a quantidade de variações em uma variável visual que, mesmo retendo suas características, a torna visualmente distinta.

Figura 10. Recomendações de BERTIN(1983) para utilização de variáveis visuais.



Fonte: Adaptada de (BERTIN, 1983)

Figura 11. Classificação da codificação de variáveis visuais quantitativas e não quantitativas.



Fonte: Adaptada de (MACKINLAY, 1986)

MACKINLAY (1986) sugere uma classificação da codificação de variáveis visuais quantitativas e não quantitativas (Figura 11), deixando claro que a pesquisa realizada não foi verificado empiricamente e o estudo foi realizado utilizando estudos psicofísicos e várias análises das diferentes tarefas perceptivas.

BORGO et al. (2013) apontam que foram realizados estudos que analisaram a eficácia sobre a percepção de variáveis visuais básicas, resultando um consenso sobre como estas variáveis se destacam. Os autores ainda apontam a seguinte ordem de percepção das variáveis: cor > tamanho > forma > orientação (o símbolo > deve ser lido como precede).

É importante comentar que as variáveis visuais adotadas nesta tese, para a construção do glifo categórico, foram inspiradas pela classificação de MACKINLAY (1986) e, também foram considerados os apontamentos de BORGO et al. (2013), quanto a ordem de percepção das variáveis visuais cor, forma e textura.

2.3 Glifos

No contexto de InfoVis, um glifo é uma representação visual de um dado ou um conjunto de dados. Suas variáveis visuais (ex: cor) e seus valores (ex: valor da cor - amarelo, vermelho, verde, etc) são mapeados para um ou mais valores dos atributos de uma base de dados (WARE, 2004; MUNZNER; MAGUIRE, 2015). Essa é uma definição bastante ampla para o termo, abrangendo uma grande quantidade de tipos de glifos criados ao longo do tempo, incluindo desde marcadores em um gráfico de dispersão, as barras de um histograma ou até mesmo um gráfico de linha inteiro (MUNZNER, 1997).

Muitos autores desenvolveram listas de variáveis visuais para os quais os valores de dados podem ser mapeados (YOST; NORTH, 2005; FENG et al., 2009; CHLAN; RHEINGANS, 2005). Eles incluem: posição (1, 2 ou 3D), tamanho (comprimento, área ou volume), forma, orientação, material (matiz, saturação, intensidade, textura ou opacidade), estilo da linha (largura, traços ou opacidade) e dinâmica (velocidade de movimento, direção do movimento, taxa de piscar).

Segundo WARD; GRINSTEIN; KEIM (2015) e FUCHS et al. (2017), os principais tipos de mapeamento de variáveis visuais para dados em Glifos são:

- Mapeamentos um-para-um: em que cada atributo de dados é mapeado para um tipo de variável visual distinta e diferente;
- Mapeamentos um-para-muitos: em que cada atributo pode ser mapeado para mais de uma variável visual. Mapeamentos redundantes são usados para melhorar a precisão e a facilidade com que um usuário pode interpretar os valores de dados;
- Mapeamentos muitos para um: em que vários ou todos os atributos de dados são mapeados para um tipo comum de variável visual, separados em espaço, orientação ou outra transformação.

Geralmente, os mapeamentos um para um são projetados para aproveitar o conhecimento do domínio do usuário, usando pares de dados e variáveis visuais intuitivas, tais

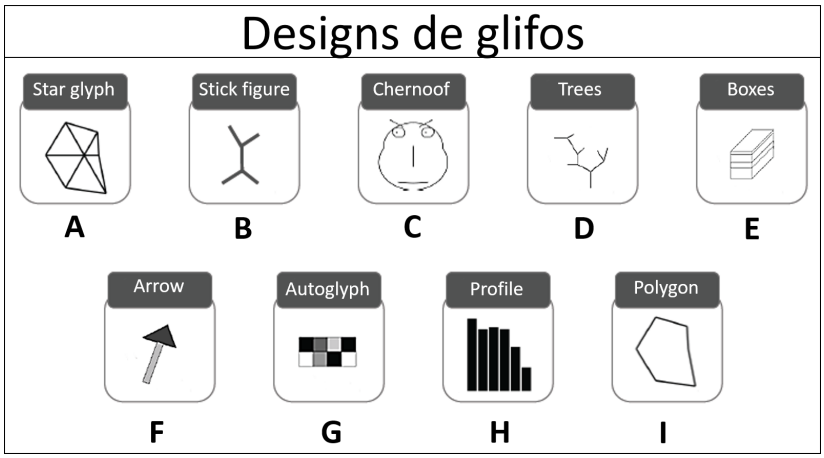
como: temperatura-cor, fluxo-setas, etc. Os mapeamentos redundantes (um-para-muitos) podem ser úteis em situações em que o número de dimensões de dados é baixo e o desejo é reduzir a possibilidade de interpretações erradas. Os mapeamentos muitos para um são usados em situações em que é importante não apenas comparar valores da mesma dimensão para registros separados, mas também comparar dimensões diferentes para o mesmo registro. Por exemplo, valores mapeados para barras, pode proporcionar comparação entre os meses de um mesmo gráfico, agrupadas por ano, pode-se comparar os meses de diversos anos.

Alguns exemplos de design de glifos e variáveis visuais são apresentados por FUCHS et al. (2017) e WARD; GRINSTEIN; KEIM (2015), e podem ser observados na Figura 12. Cada design de glifo possui características específicas e podem ser detalhados da seguinte forma:

- *Star* glifo (A) - construído a partir de raios que emanam uniformemente a partir do centro do glifo (SIEGEL et al., 1972).
- *Stick Figures* glifo (B) - baseado em variações ligadas ao comprimento, ângulos e cores das ramificações que são criadas em sua estrutura (PICKETT; GRINSTEIN, 1988).
- *Face* glifo (C) - formado pela combinação e variação dos elementos tamanho e posição dos olhos, nariz, boca, curvatura da boca e ângulo das sobrancelhas (CHERNOFF, 1973).
- *Tree* glifo (D) - formado a partir de combinações dos elementos comprimento, espessura, ângulos dos ramos criados, sendo que a estrutura de ramo é derivada da análise de relações entre dimensões (KLEINER; HARTIGAN, 1981).
- *Boxes* glifo (E) - sua construção tem por base a combinação dos elementos altura, largura, profundidade da primeira caixa e a altura de caixas sucessivas (HARTIGAN, 1975).
- *Arrow* glifo (F) - estrutura que tem por base a variação do comprimento, largura, conicidade e cor da base do corpo e da forma utilizada para caracterizar a cabeça do glifo WITTENBRINK; PANG; LODHA.
- *Autoglyph* glifo (G) - sua estrutura é baseada na utilização de caixas e cores para representação dos dados (BEDDOW, 1990).
- *Profile* glifo (H) - construído a partir da manipulação do comprimento e das cores das barras que o formam (TOIT; STEYN; STUMPF, 1986).

- *Polygon* glifo (I) - sua base estrutural é formada a partir da transmissão de deformação local em um campo vetorial por meio de mudanças na orientação e na forma (SCHROEDER; VOLPE; LORENSEN, 1991).

Figura 12. Glifos: designs de glifos encontrados na literatura.



Fonte: Adaptada de FUCHS et al. (2017) e WARD; GRINSTEIN; KEIM (2015)

Considerando questões voltadas a percepção, o uso de glifos em InfoVis requer atenção para alguns pontos e limitações, são elas:

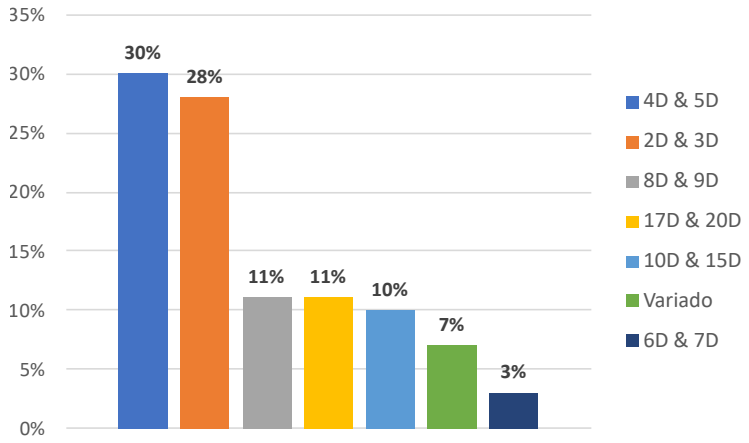
- Em relação ao uso das variáveis visuais, há variáveis visuais que se pode julgar com mais precisão do que outros, como orientação ou cor;
- Relacionamento entre variáveis visuais adjacente ou próximas são mais fáceis de perceber;
- Tarefas de comparação são mais fáceis para glifos próximos um do outro na tela;
- O número de dimensões e registros de dados que podem ser manipulados com glifos é limitado.

Em relação às dimensões que podem ser representadas por um glifo, FUCHS et al. (2017) apresenta o mapeamento de dimensões codificadas em glifos (ver Figura 13), no qual aproximadamente 50% dos trabalhos levantados codifica até 5 dimensões em glifos. Outra questão importante é a ordenação das dimensões disponíveis que podem ser mapeadas no glifo, algumas estratégias são comentadas a seguir:

- As dimensões podem ser ordenadas de acordo com sua correlação, de modo que dimensões similares sejam mapeadas adjacentes umas às outras. Isso pode ajudar a revelar padrões de tendências, bem como expor alguns outliers;

- As dimensões podem ser mapeadas de forma a promover glifos de formato simétrico, que podem ser mais fáceis de perceber e lembrar. Formas menos simétricas em relação a glifos vizinhos também se destacam.
- Dimensões podem ser classificadas de acordo com seus valores de um registro. Por exemplo, se os dados representam uma série temporal multivariada, a classificação das dimensões com base no primeiro registro pode criar um padrão de tendência ao longo do tempo mais perceptível.

Figura 13. Percentual dos trabalhos tomando por base o número de dimensões codificadas nos glifos.



Fonte: Adaptada de FUCHS et al. (2017).

As dimensões podem ser classificadas manualmente com base no conhecimento do usuário sobre o domínio. Assim, dimensões semanticamente semelhantes podem ser agrupadas ou usadas para características de glifos simétricos, o que pode simplificar a interpretação.

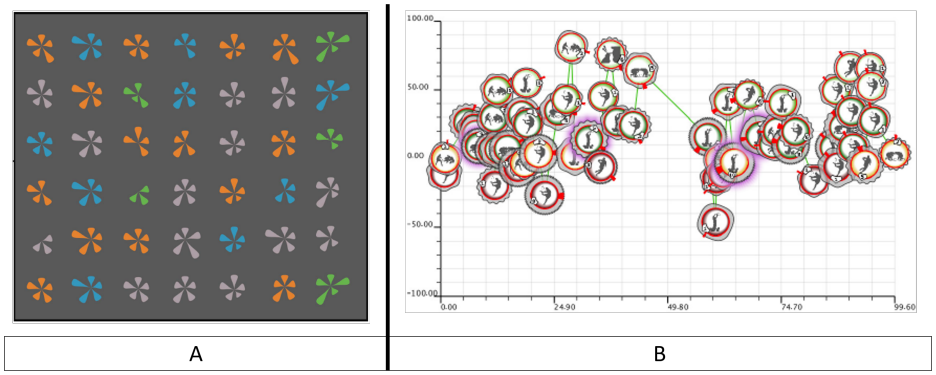
Uma consideração final importante no projeto de uma visualização baseada em glifos é com relação a onde os glifos serão dispostos na tela, ou seja, o seu *layout*. De acordo com WARD (2002), existem três classes gerais de estratégias de *layout*:

1. Uniformes - os glifos são dimensionados e posicionados com espaço igual entre eles para ocupar toda a tela (ver Figura 14-A). Essa estratégia elimina sobreposições, enquanto faz uso eficiente do espaço da tela. Diferentes ordenações de valores de registros podem expor diferentes características dos dados, assim como diferentes ordenações das variáveis visuais (dimensões).
2. Orientado por Dados - valores de dados são usados para controlar o posicionamento dos glifos, como em um gráfico de dispersão de dados (ver Figura 14-B). Abordagens

matemáticas para o cálculo da posição do glifo podem minimizar a questão de sobreposição dos mesmos, separando aqueles que estão muito próximos.

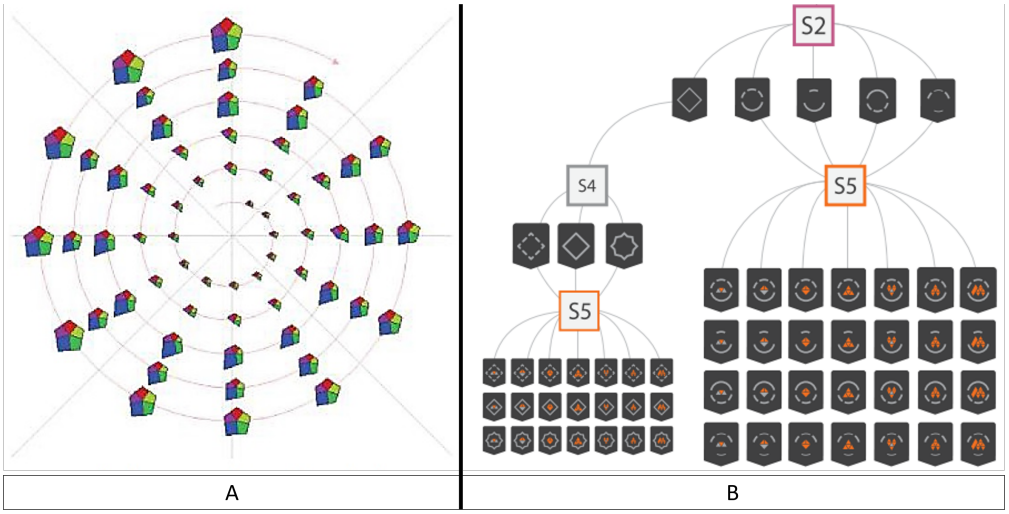
3. Orientado pela Estrutura - se os dados tiverem uma estrutura implícita ou explícita, como cíclica ou hierárquica, isso pode ser usado para controlar o posicionamento. Por exemplo, os glifos podem ser dispostos em uma espiral ou grade para enfatizar padrões cíclicos (Figura 15-A) da mesma forma, qualquer uma das técnicas de desenho de árvores pode ser usada para posicionar os glifos para ajudar na comunicação de relações hierárquicas (Figura 15-B).

Figura 14. A - Exemplo de glifo em uma estrutura uniforme. B - Exemplo de glifo em uma estrutura orientada por dados.



Fonte: Adaptada de KECK et al. (2017) e CHUNG et al. (2013).

Figura 15. Exemplo de glifos orientados pela estrutura. A - Estrutura em Espiral. B - Técnica baseada em estrutura de árvore.



Fonte: Adaptada de WARD; GRINSTEIN; KEIM (2015) e MAGUIRE et al. (2012).

FUCHS et al. (2017) aponta cinco tipos de *layouts* utilizados, sendo o mais frequente o *small-multiples grid* (65,22%), seguido por mapas geográficos (17,39%), diagramas de dispersão (6,52%), diagramas Node-Link (4,35%), e outros *layouts* (6,52%), como ambientes 3D. FUCHS et al. (2017) destaca ainda que *layouts* do tipo Treemap com glifos ainda não foram investigados, e considera uma oportunidade de pesquisa.

Assim, esta tese investiga glifos em Treemaps —*layout* orientado pela estrutura— pela oportunidade de pesquisa a ser explorada apontada por FUCHS et al. (2017), possibilitando aumentar a representação visual de dados na técnica, e mitigar problemas conhecidos, tais como: valores negativos, zeros e pequenos, em virtude da variedade de representações visuais dos glifos. A ideia principal é utilizar o poder de agrupamento e comparação do Treemap com a diversidade de representações visuais do glifo, como é destacado por FUCHS et al. (2017).

O modelo de glifo desta tese pode representar visualmente um número de dimensões variável entre 1 e 14. Para tal, utiliza as variáveis visuais textura, cor-matiz, forma, texto, orientação e posição para representar visualmente dados categóricos, através de uma mapeamento um-para-um, e para a representação de dados quantitativos é utilizado um profile glifo, através de um mapeamento um-para-muitos.

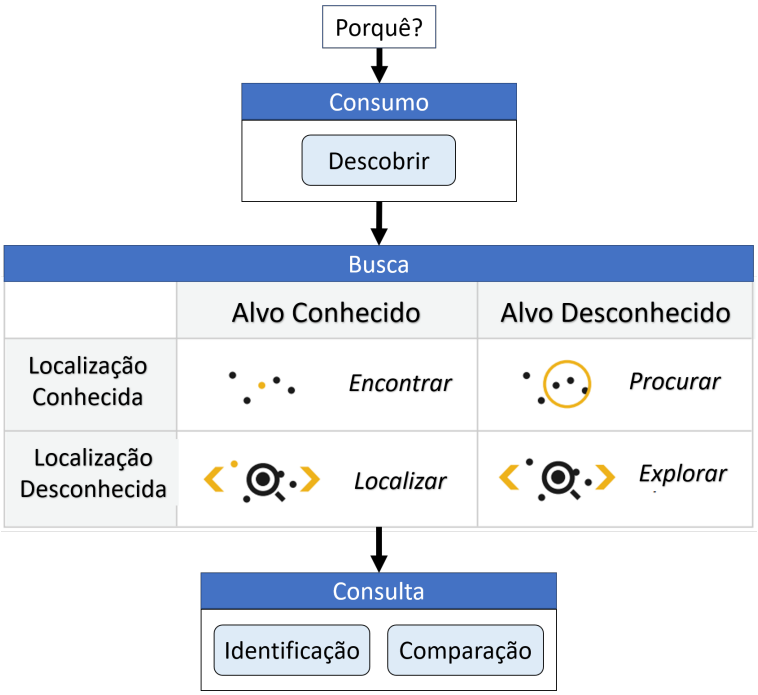
2.4 Tarefas de Visualização da Informação

Uma ferramenta de InfoVis pode ser utilizada com diversos propósitos, como análise e exploração de dados, comunicação, geração de novos conhecimentos, entre outros. As taxonomias de tarefas de InfoVis permitem conduzir experimentos para avaliar as ferramentas InfoVis nos cenários propostos. Dentre as tarefas de InfoVis mais encontradas nas taxonomias da literatura destacam-se: identificação, localização, comparação, configuração, inferência, correlação, ordenação, associação, entre outras (PILLAT; VALIATI; FREITAS, 2005; WEHREND; LEWIS, 1990).

Para MUNZNER; MAGUIRE (2015) ao definir o "porquê" de uma tarefa de InfoVis ser realizada, os autores descrevem que é necessário incluir vários níveis de detalhamento, partindo de um escopo mais alto nível (Consumo), passando para um nível médio (Busca) e, por fim, um nível mais baixo (Consulta). A Figura 16 ilustra uma versão simplificada e adaptada da tipologia de MUNZNER; MAGUIRE (2015). O escopo de mais alto nível, o Consumo, é definido quando as visualizações são utilizadas para consumir informações de diferentes contextos. O "Consumo"é motivado pela necessidade de Descobrir informação, que por sua vez trata-se da geração e verificação de hipóteses.

De acordo com MUNZNER; MAGUIRE (2015), no nível médio, as tarefas de Busca, podem ser subdivida em Encontrar (Lookup), Localizar (Locate), Procurar (Browse), e Explorar (Explore), de acordo com a posição e identificação do item visual de análise ser

Figura 16. Etapas ao definir o "porquê" de uma tarefa de InfoVis ser realizada. Este foi o fluxo usado nas tarefas das avaliações desta tese.



Fonte: Adaptada de MUNZNER; MAGUIRE (2015).

conhecido ou não (Figura 16). As características de cada uma das tarefas são descritas a seguir:

- Encontrar: se os usuários já souberem o que estão procurando e onde devem procurar, o tipo de busca é simples. Por exemplo, pode-se chegar rapidamente na categoria primata lembrando-se da classe mamíferos;
- Localizar: é caracterizada por conhecer o item visual e desconhecer a sua localização, isto é, descobrir onde está o objeto específico;
- Procurar: é caracterizada por possuir localização conhecida, mas não sua identidade visual. Nesse caso, os usuários estão procurando por um ou mais itens que se ajustam a algum tipo de especificação, por exemplo utilizando um determinado intervalo de valores de um atributo.
- Explorar: é caracterizada por não possuir localização e identidade conhecida. Implica procurar por características sem ter em conta a sua localização, muitas vezes a partir da visão geral. Exemplos incluem a busca por *outliers* em um gráfico de dispersão, por picos de anômalos ou padrões periódicos em um gráfico de linha de dados de séries temporais, entre outros.

Por fim, o nível mais baixo, a Consulta pode ser dividida em Identificação, e Comparação. A diferença entre identificar ou comparar é definida pela saída do nível de busca. Desta forma, se a Busca retornou mais de um alvo, a tarefa é de comparação, caso contrário a tarefa será de identificação. Essa tese utiliza tarefas de descoberta (nível mais alto), encontrar, localizar e procurar (nível intermediário), identificação e comparação (nível mais baixo), conforme o fluxo da Figura 16.

3 TRABALHOS RELACIONADOS

Neste capítulo foi realizada uma análise do estado da arte envolvendo estudos com Treemaps e glifos. Serão apresentadas duas categorias de estudo que influenciaram este trabalho: estudos sobre glifos aplicados em várias técnicas de InfoVis e estudos sobre o uso de glifos especificamente nos Treemaps.

3.1 Estudos sobre glifos

FUCHS et al. (2017) realizou uma revisão sistemática envolvendo 64 trabalhos relacionados a avaliação de visualizações de dados utilizando glifos para auxiliar pesquisadores sobre o ganho da informação no entendimento de glifos, bem como seu design. O estudo também aponta os designs utilizados, questões, dados e tarefas utilizadas nos artigos. Algumas considerações apontadas foram:

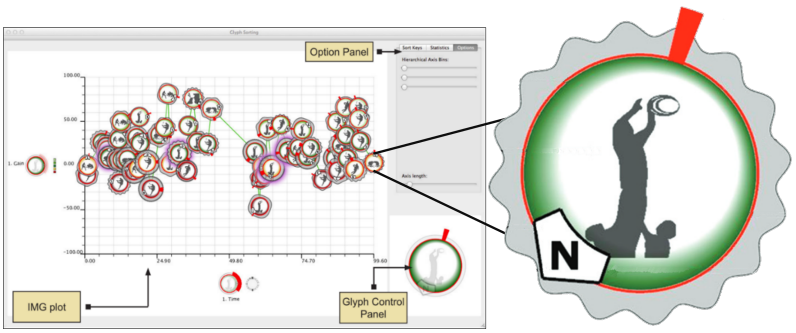
- Dimensões: a maioria dos estudos utilizam glifos codificados para menos de 10 dimensões;
- Tarefas: a maioria dos estudos utilizam tarefas de similaridade entre glifos e tarefas de busca para avaliar o design do glifo;
- Influência do *layout* e *background*: poucos estudos, mas estes apontam que o arranjo dos glifos e seus vizinhos não influenciam em seu entendimento;
- Quantidade de dados e dimensões: o aumento no número de glifos na visualização e o número de dimensões representada no glifo afeta negativamente a leitura dos glifos;
- Uso de metáforas: alguns estudos sugerem que o uso de metáforas no design do glifo pode auxiliar em seu entendimento;
- Novos *layouts* para glifos: muitos estudos utilizam os glifos como *small multiples* em um *layout* de grade, não há estudos de glifos em *layouts* como Treemaps, outros *layouts* encontrados na literatura utilizando glifos são mapas, dispersão de dados, node-link, etc.;
- Avaliações: a necessidade de mais estudos avaliativos sobre glifos pode ajudar a gerar boas práticas para utilização de glifos em visualização de dados.

Em BORGO et al. (2013) foi realizado um levantamento do estado da arte focado em visualizações baseadas em glifos. As revisões incluem considerações de design e implementações de glifos, bem como seu uso em diferentes aplicações de visualizações. Destaca-se

estudos avaliativos sobre a percepção de variáveis visuais em glifos, onde a ordem e variáveis mais comuns de pré-atenção em glifo são cor, tamanho, forma e orientação.

CHUNG et al. (2013) apresentam uma estrutura de classificação baseada em glifos para orientar e facilitar a classificação interativa dos dados. Os autores apresentam uma ferramenta interativa baseado em glifos para controlar a classificação de alta dimensão e visualizar os resultados classificados. A ferramenta, chamada de ***Glyph Sorting***, utiliza como base um layout de dispersão de dados, onde os itens do gráfico são glifos circulares, que mapeiam até nove atributos. Um pictograma no centro do círculo representa os eventos da partida. Seus glifos usam a sobreposição de variáveis visuais sem comprometer a compreensão dos dados que eles representam (ver Figura 17).

Figura 17. Exemplo de visualização gerada a partir da ferramenta Glyph Sorting. O design do glifo apresenta sobreposição das variáveis visuais sem prejudicar a leitura dos dados.

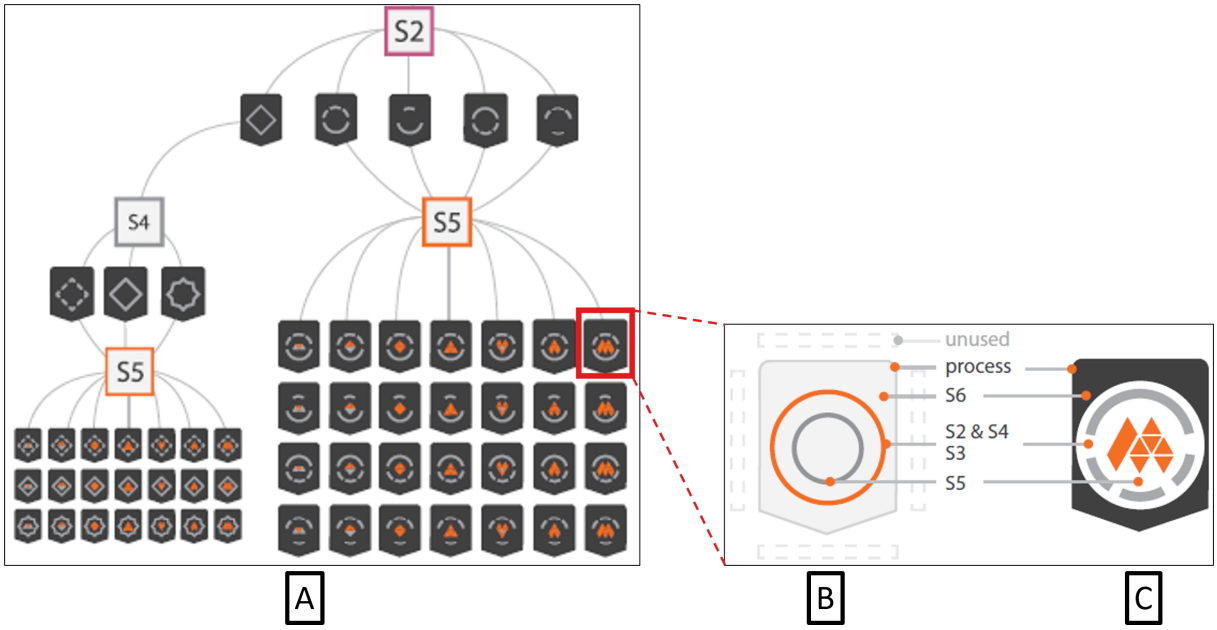


Fonte: Adaptada de CHUNG et al. (2013).

MAGUIRE et al. 2012 utilizou o campo experimental da biologia como plataforma de teste para o desenvolvimento de uma metodologia sistemática para criar um mapeamento visual baseado em glifos para visualizar projetos e processos experimentais. O artigo destaca que durante a coleta de evidências de dados biológicos, a representação dessas informações é limitada à descrição detalhada ou a representações ineficazes. De forma geral, os fluxos de trabalho são tradicionalmente desenhados como diagramas baseados em texto, tais diagramas são particularmente ineficazes quando é necessário comparar entre diferentes fluxos de trabalho ou identificar componentes incomuns ou ausentes em um fluxo de trabalho. Como alternativa MAGUIRE et al. 2012 propôs substituir os rótulos textuais por glifos, permitindo acesso interativo a descrições detalhadas, explorando os conceitos de hierarquia de dados e a capacidade de distinguir variáveis visuais através da sobreposição entre elas (ver Figura 18).

LEGG et al. (2012) apresentaram uma solução de visualização para a análise de dados de desempenho esportivo em tempo real. Os autores usaram um design de glifos para permitir que a equipe de treinamento e os analistas visualizassem ações e

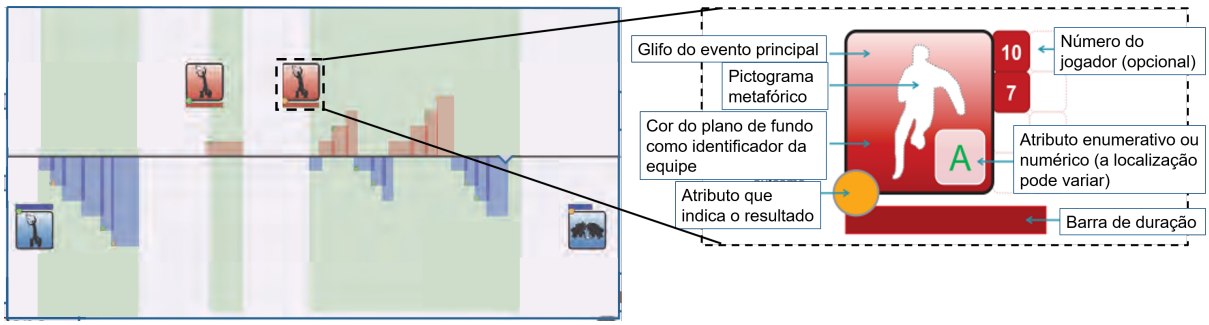
Figura 18. A - Contexto de aplicação do modelo do glifo; B - Modelo do glifo apresentado por MAGUIRE et al. (2012); C - Exemplo do glifo, mostrando o uso das variáveis visuais cor e forma, e faz uso da sobreposição entre elas.



Fonte: Adaptada de MAGUIRE et al. (2012).

eventos rapidamente. O artigo apresentou um glifo composto por sete atributos, tanto categóricos quanto ordenados. Um pictograma representa o evento principal do glifo e seu design apresenta a sobreposição de variáveis visuais sem prejudicar a leitura do glifo (ver Figura 19).

Figura 19. Modelo do glifo apresenta atributos categóricos e ordenados, e utiliza sobreposição das variáveis visuais sem interferir na leitura do glifo.

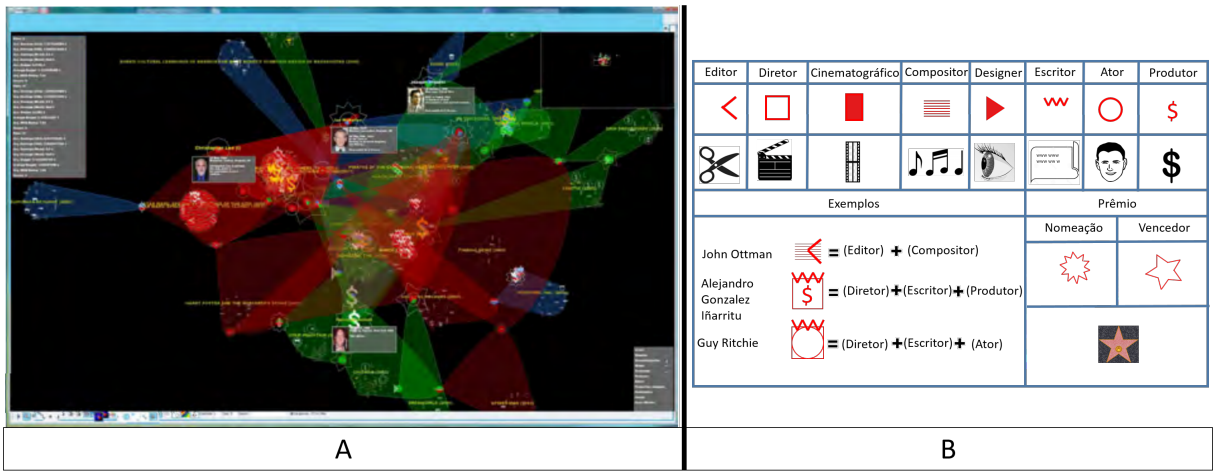


Fonte: Adaptada de LEGG et al. (2012).

Em THERON et al. (2007) foi apresentada uma ferramenta que emprega múltiplas técnicas do InfoVis, incluindo coordenadas paralelas, Treemap e glifos. O estudo foi centrado na recuperação de informação de obras cinematográficas. Eles usaram glifos

baseados na variável visual forma, variando-a para representar o perfil da pessoa envolvida em um filme. Destaca-se que no design do glifo a sobreposição das formas nunca causa oclusão total desta variável (ver Figura 20).

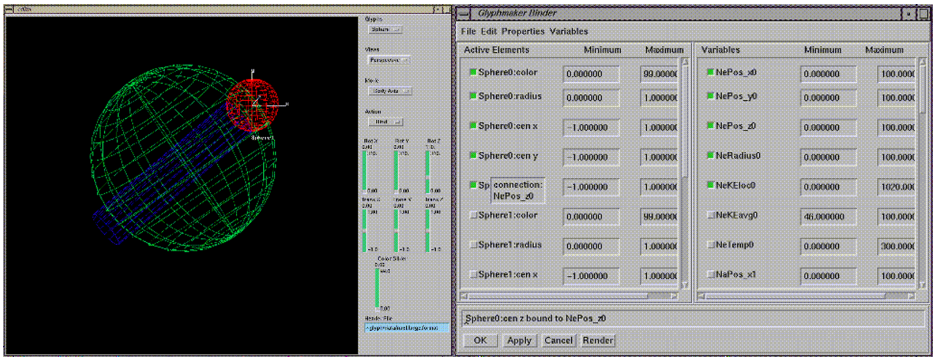
Figura 20. A - Contexto de aplicação do modelo do glifo; B - Exemplos de glifos de THERON et al. (2007) baseado em formas e usando sobreposição.



Fonte: Adaptada de THERON et al. (2007).

No trabalho de RIBARSKY et al. (1994) foi implementado o Glyphmaker, uma ferramenta para facilitar a criação de glifos para usuários não especialistas. O Glyphmaker permite que os usuários personalizem propriedades gráficas usando um editor de glifos simples e uma interação de ligação de apontar e clicar (ver Figura 21). A ferramenta oferece flexibilidade e dinamismo ao usuário na tarefa de criação de glifos.

Figura 21. Tela da ferramenta Glyphmaker.



Fonte: Adaptada de RIBARSKY et al. (1994).

A estratégia de construção dos glifos desta tese foi pensada considerando a sobreposição das variáveis visuais, de forma que seja possível a análise de todas partes do glifo,

ou seja, não ocorrendo oclusão total de qualquer variável visual. O modelo de glifo desta tese permite que o usuário possa escolher o mapeamento entre os atributos e as variáveis visuais, permiti a escolha da quantidade e a ordem entre as variáveis a serem utilizadas. Com isso, o usuário tem a flexibilidade para definir qual o melhor mapeamento do glifo para seus dados.

3.2 Usando glifos em Treemaps

A pesquisa conduzida por KHAN et al. (2014) apresenta a técnica de visualização envolvendo squarified Treemap com glifos para representar um sistema de arquivos de forma a melhorar o processo de recuperação de informações a partir dos arquivos disponíveis no sistema (Figura 22). A pesquisa foi centrada na visualização de foco + contexto, onde o resultado (o "foco") de uma consulta usou glifos sobre os itens do Treemap (o "contexto"), causando uma oclusão parcial deles. O design do glifo representa quatro atributos: três categóricos e um quantitativo; no entanto, o usuário não pode modificar o design do glifo.

Figura 22. Visualização de uma busca no sistema de arquivos, onde os glifos representam resultado da busca (foco) e o Treemap representa o sistema de arquivos geral (contexto) e o design do glifo proposto.

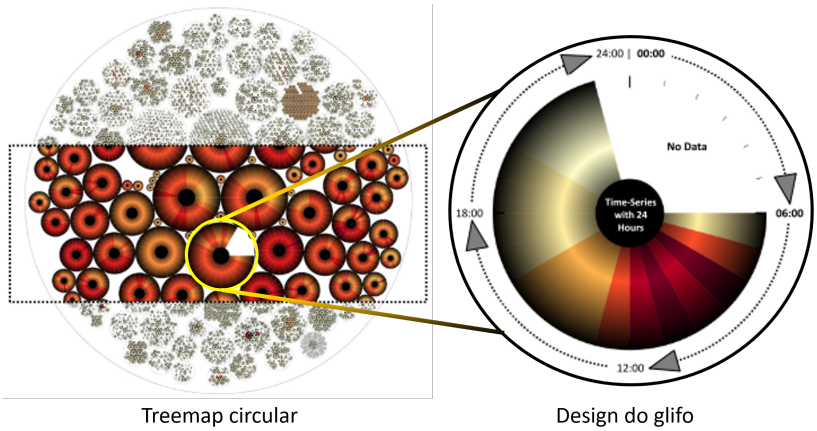


Fonte: Adaptada de KHAN et al. (2014).

Na pesquisa de FISCHER; FUCHS; MANSMANN (2012) foi desenvolvida uma ferramenta combinando Treemap circulares como layout para um grande número de glifos temporais que representam dados de uma série temporal (ver Figura 23). O glifo do estudo, chamado de ***clockeyes***, é baseados em relógio (*clock-based*), representando 24h e cada segmento circular indica 1h. Além disso, eles usam cores para indicar a intensidade de um evento. O design do glifo emprega uma estrutura estática que não permiti modificações.

ELMQVIST; FEKETE (2010) desenvolveram um modelo para construção, visualização e interação, baseados em técnicas de InfoVis usando agregação hierárquica. O trabalho foi desenvolvido com o foco de tornar as representações visuais mais escaláveis

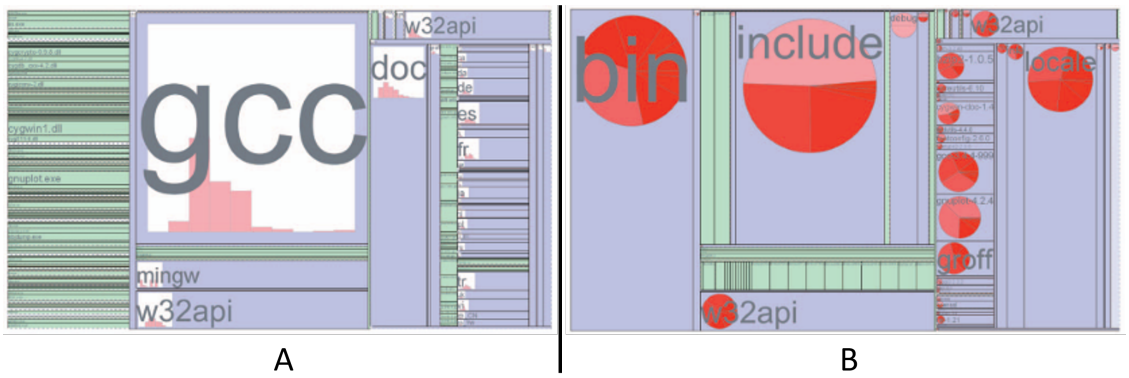
Figura 23. Visualização utilizando o ClockMap, combinando Treemap circular e glifos.



Fonte: Adaptada de FISCHER; FUCHS; MANSMANN (2012).

e menos confusas. O modelo proposto em questão permite o melhoramento de técnicas existentes com funcionalidades multiescalares, assim como projetar novas técnicas de InfoVis, como diagramas de dispersão, coordenadas paralelas e diagramas node-link. Os autores propuseram o uso de profile glifos e glifos de pizza para representar a agregação de itens do Treemap (ver Figura 24), com o objetivo de melhorar o entendimento das visualizações. A investigação dos autores resultou em um conjunto de diretrizes para o design de glifos.

Figura 24. Treemaps com técnica de agregação para estrutura de arquivos, onde os profile glifos (gráfico de barras) (A) e glifos de gráficos de pizza (B) mostram a distribuição de tamanho dos arquivos.



Fonte: Adaptada de ELMQVIST; FEKETE (2010).

Entre os trabalhos relacionados citados, que mesclam Treemaps e glifos, nenhum deles ataca todos os problemas da técnica, considerando principalmente a representação de

zero, valores negativos e pequenos. Além disso, apenas alguns artigos realizaram estudos com usuários para avaliar a aplicação de glifos no Treemap.

De forma geral, a tabela 1 resume as principais diferenças entre as abordagens relacionadas, das seção 3.1 e seção 3.2, e esta tese.

Quadro 1. Propostas que usaram glifos e propostas que usavam glifos em Treemaps em comparação com esta tese.

	*1	*2	*3	*4	*5	*6	*7	*8	**Tese**
Usa glifo na visão geral	✓	✓	✓	✓	X	X	✓	X	✓
Aumenta a quantidade de dimensões	✓	✓	✓	✓	✓	✓	✓	X	✓
Número de variáveis visuais	8	2	6	1	3	4	2	1	7
Glifo é flexível	X	X	X	X	✓	X	X	X	✓
O glifo usa sobreposição	✓	✓	✓	✓	X	X	X	X	✓
Usa glifo quantitativo	✓	X	✓	X	✓	✓	✓	✓	✓
Usa glifo categórico	✓	✓	✓	✓	✓	✓	X	X	✓
O glifo é independente de domínio	X	X	X	X	✓	X	✓	✓	✓
Usa o layout do Treemap	X	X	X	X	X	✓	✓	✓	✓
Realizou estudos com usuários	X	X	X	X	X	✓	X	X	✓

X → Não; ✓ → Sim.

*1: **(CHUNG et al., 2013)** ; *2: **(MAGUIRE et al., 2012)**; *3: **(LEGG et al., 2012)** ; *4: **(THERON et al., 2007)**; *5: **(RIBARSKY et al., 1994)**; *6: **(KHAN et al., 2014)** ; *7: **(FISCHER; FUCHS; MANSMANN, 2012)**; *8: **(ELMQVIST; FEKETE, 2010)**;

4 CONCEPÇÃO E DESIGN DO GLIFO

Ao elaborar o modelo do glifo desta tese foram estabelidos alguns requisitos para seu design. Primeiramente, definiu-se que o glifo precisaria ser capaz de mapear atributos categóricos e quantitativos, pois assim seria possível representar uma variedade maior de dados. Outro requisito era que o glifo deveria ser de independente de domínio, pois sem um domínio poder-se-ia representar qualquer base de dados.

Um requisito importante era com relação a flexibilidade do glifo, e para isso o glifo deveria permitir que o usuário tivesse um certo grau de configuração sobre eles, pois com isso o glifo seria mais flexível a alterações realizadas pelos usuários a medida que fosse necessário durante uma análise. Por fim, o glifo precisaria adaptar-se ao layout da técnica aplicada, e com isso o glifo poderia variar sua forma e tamanho, assim o glifo poderia ser aplicado diversas técnicas de InfoVis.

Para atender ao requisito de tornar o glifo flexível para o usuário, definiu-se que o glifo deveria possuir uma estrutura em camadas, assim o usuário poderia escolher a quantidade de camadas, a ordem e o que iria conter em cada camada. Com a definição da estrutura em camadas, definiu-se que o mapeamento entre as variáveis visuais e os atributos de uma base de dados deveriam ser um-para-um, ou seja, cada variável visual irá repesentar apenas um atributo da base de dados (WARD; GRINSTEIN; KEIM, 2015; FUCHS et al., 2017).

Considerando a estrutura em camadas, definiu-se que as camadas seriam empilhadas uma sobre as outras de forma concêntrica. Com relação a organização das camadas e tendo em mente o requisito da representação de dados categóricos e quantitativos, definiu-se que a última camada —a camada mais acima— seria responsável por representar dados quantitativos, utilizando mapeamento varios-para-um, e as demais camadas —camadas da base do glifo— representariam dados categóricos.

Como um dos requisitos era que o glifo fosse independente de domínio, pensou-se em uma estrutura que fosse flexível com relação ao uso de metáforas, ou seja, glifos que possuem representações metáforicas do mundo real para facilitar sua compreensão (FUCHS et al., 2017), portanto o uso ou não de metáforas fica a critério do usuário ao criar o glifo. Assim, definiu-se que o modelo de glifo em camadas permitiria a inclussão de metáforas na última camada —a camada mais acima— do glifo, no entanto é importante destacar que esta tese não aplicou a utilização de metáforas no design dos glifos utilizados.

Para permitir que o glifo se adapta-se ao layout da técnica de InfoVis aplicada, foi definido que o glifo precisaria submeter-se a uma uma análise automática de uma técnica de aprendizado de máquina, para que fossem geradas regras inteligentes para a

sua adaptação a técnica de InfoVis.

Para auxiliar na compreensão das decisões tomadas com relação ao design do glifo dessa tese, neste capítulo serão apresentados os pré-testes realizados que auxiliaram na concepção do glifo. Serão abordados aspectos sobre o design e uso do glifo em camadas aplicado em um Treemap. Também será apresentado o teste realizado para avaliar a visibilidade do glifo e será apresentado o modelo de glifo adaptativo.

4.1 Configurações do Pré-teste 1

A literatura, de forma geral, apresenta trabalhos que buscam elimiar a oclusão entre itens visuais das técnicas de InfoVis (LEUNG; APPERLEY, 1994; SHNEIDERMAN, 1996; KEIM, 1997), pois a oclusão de itens visuais pode trazer como consequências erros de interpretações sobre um conjunto de dados ou a perda de informações durante uma análise de dados. Desta forma, faz-se necessário o estudo sobre o quanto uma variável visual poderia ser notada dado um valor de oclusão (sobreposição) para que se possa mitigar o problema de perda de informações.

Para tal, foi realizado um pré-teste com o objetivo de avaliar a percepção visual sobre variáveis visuais, considerando a variação dos níveis de sobreposição parcial e aumento gradativo da quantidade de valores distintos para codificação de dados categóricos.

4.1.1 Suporte computacional

Para a execução do teste foi utilizado um computador com configuração de 8GB de memória RAM, HD de 1TB e processador Intel Core i7, um *display* de 21" com resolução de 1920x1080 px e posicionado em orientação paisagem. Destaca-se que durante a execução do pré-teste os participantes ficaram sentados confortavelmente e com olhos a uma distância do monitor entre 60cm e 70cm.

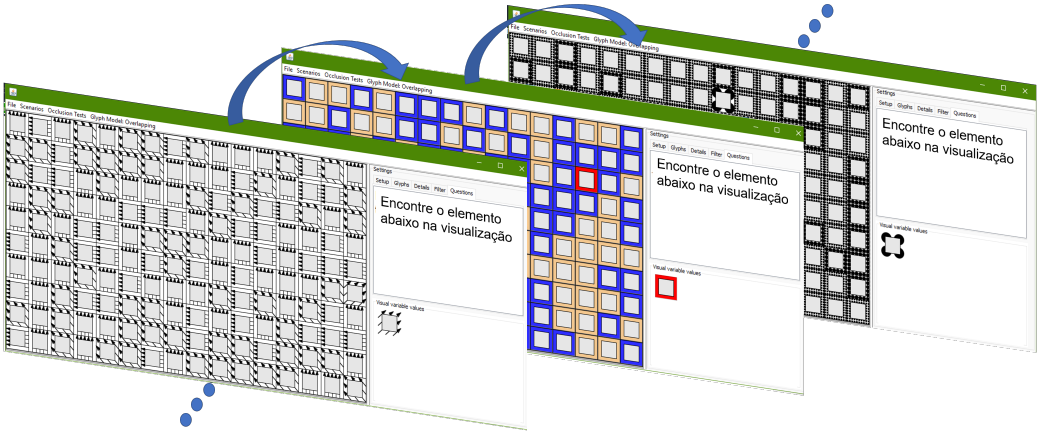
4.1.2 Perfil dos participantes

Todos os participantes do teste, um total de 48 voluntários, possuíam ou estavam cursando o nível superior, pertencendo a diversas áreas de atuação, como por exemplo: administração, biomedicina, ciências naturais, computação, entre outras. Os participantes tinham idades entre 18 e 42 anos, sendo 39 homens e 09 mulheres. Nesse pré-teste não era necessário que os participantes fossem especialistas, pois o intuito deste pré-teste era avaliar a percepção e visibilidade das variáveis visuais, ou seja, o pré-teste não exigia um conhecimento específico dos participantes.

4.1.3 Execução do pré-teste 1

Com a preocupação principal sobre a oclusão das variáveis visuais, foi desenvolvida uma aplicação que apresentava, em um *layout* de grade, um conjunto de variáveis visuais sobrepostas parcialmente e coleta as respostas dos participantes com relação a percepção das variáveis visuais (ver Figura 25).

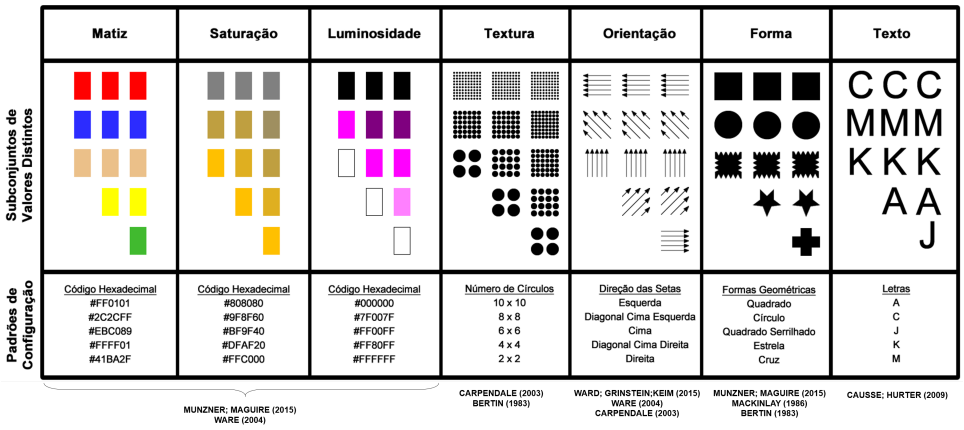
Figura 25. Exemplos de visualizações geradas a partir da ferramenta de avaliação.



Fonte: Autor do trabalho.

O pré-teste utilizou as variáveis visuais cor-matiz, cor-saturação, cor-luminosidade, textura, orientação, forma e texto (ver Figura 26). A ferramenta desenvolvida para o pré-teste sorteia uma das variáveis visuais, já citadas, e a apresentava aos participantes com uma sobreposição parcial, que pode ser de 0%, 50%, 60% e 70%.

Figura 26. Valores visuais distintos escolhidos (3, 4 e 5) para cada variável visual estudada.



Fonte: Adaptado de SANTOS et al. (2019).

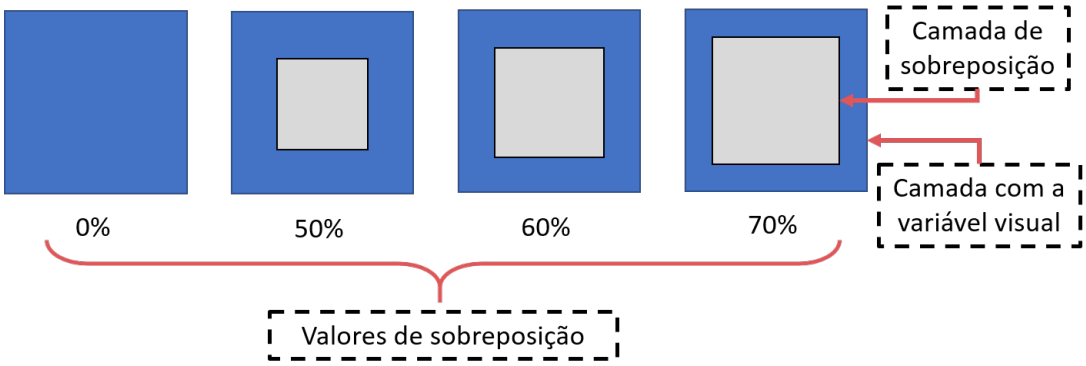
A ferramenta, em questão, também sorteia a quantidade de valores das variaveis visuais, variando entre 3, 4 ou 5 (ver Figura 26). Desta forma, o pré-teste foi caracterizado

como um *Mixed-design* (FORSELL, 2010) de 3 (três) variáveis independentes: variável visual; nível de sobreposição parcial e quantidade de valores distintos de uma variável visual.

Após o sorteio da variável visual e de sua quantidade de valores distintos, a ferramenta apresenta em cada célula do *layout* de grade um valor da variável visual com uma sobreposição parcial de um quadrado concentrico. Tal quadrado é uma camada adicionada acima da variável visual reprensentando sua sobreposição.

Em outras palavras, a ferramenta implementa o conceito dos glifos em camadas, e para o pré-teste em questão convencionou-se que os glifos teriam no máximo duas (2) camadas, ou seja, uma delas deveria ser a camada chamada de **camada de sobreposição** e a outra deveria representar uma variável visual (ver Figura 27).

Figura 27. Glifo com duas (2) camadas, sendo que uma a camada com uma variável visual e a outra a camada de sobreposição. A sobreposição poderia variar entre 0%, 50%, 60% e 70%.



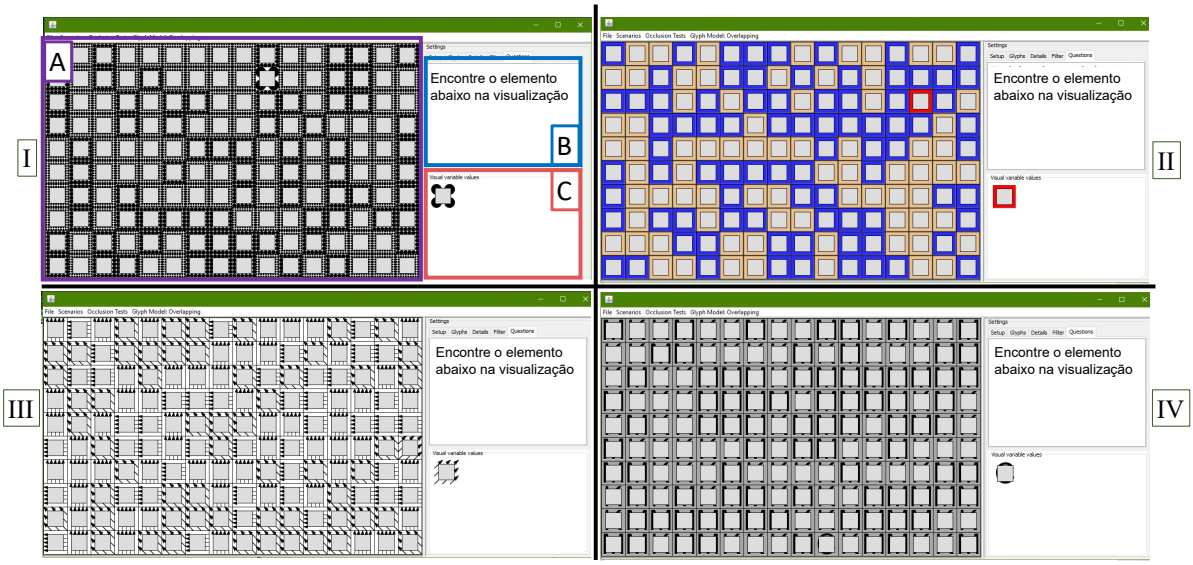
Fonte: Autor do trabalho.

Tal convenção foi adotada pois o escopo do pré-teste era avaliar o comportamento da variável visual com sobreposição, e para isso, bastaria sobrepô-la com uma camada. A camada de sobreposição sofria a variação de sua área, com relação a camada da variável visual, entre 0%, 50%, 60% e 70% e foi adicionada ao glifo para simular a oclusão da variável visual e, no escopo do pré-teste, não mapeava nenhum valor de uma base de dados ou variável visual.

É importante destacar que a ferramenta ao ser carregada gerava todas as células da grade com o mesmo tipo de variável visual e nível de oclusão (por exemplo, com um valor de 70% de sobreposição na variável visual forma), diferindo apenas nos valores da variável visual (por exemplo, variável visual forma: cruz, estrela e círculo), a Figura 28 ilustra como eram apresentados os glifos no *layout* de grade. A ferramenta gerava uma grade de 16 x 10 elementos (480 x 768 pixels), sendo que apenas uma ocorrência da instância a ser procurada constava no *layout*. Assim, o teste consistia em localizar e selecionar o elemento

único da grade apresentada.

Figura 28. Visualizações com a camada de sobreposição com um valor de 70% com exemplos de glifos gerados durante a execução do teste (I - exemplo com textura; II - exemplo com cor-matiz; III - exemplo com orientação; IV - exemplo com texto). (I) A - Área da visualização do layout em grade; B - Área de apresentação da tarefa de InfoVis; C - Área de apresentação do alvo para ser procurado na grade da área A.



Fonte: Autor do trabalho.

Durante a realização do teste a ferramenta exibia a seguinte tarefa para os participantes: "Encontre o elemento abaixo na visualização". A tarefa era exibida na área B da ferramenta (ver Figura 28-(I)B), e o elemento que deveria ser encontrado na grade (ver Figura 28-(I)A) era apresentado na área C (ver Figura 28-(I)C).

Para a execução da tarefa foi estabelecido 30 segundos, tempo definido após a realização de testes pilotos. Assim, se o participante não conseguisse encontrar o elemento único na grade, uma janela *pop-up* aparecia informando que o tempo havia terminado, nesse caso a ferramenta registrava a resposta como "tempo excedido" para diferenciar das respostas incorretas.

Os participantes foram divididos em quatro grupos compostos por 12 indivíduos, como sugerido em FORSELL (2010), em um design *Between-Subjects*[1], cuja variável independente é o nível de sobreposição parcial. Dessa forma, todas as visualizações da grade que um participante visse tinha o mesmo nível de sobreposição em seus elementos, o que definia o grupo do participante.

As variáveis independentes variável visual e quantidade de Valores Distintos foram

[1] Quando a comparação é entre grupos de participantes, e suas pontuações não são relacionadas

permutadas em 18 combinações (Matiz - 3, Matiz - 4, Matiz - 5, texto - 3, texto - 4, ...) e organizadas em um design *Within-Subject*, no qual todos os participantes realizaram todas as combinações. Para cada combinação foram apresentados 6 (seis) grades, totalizando 108 visualizações da grade por participante. Adicionalmente, a cada combinação a ferramenta coletava as respostas dos participantes automaticamente em um arquivo CSV (*Comma-Separated-Values*), gravando os seguintes dados:

- Quantidade de valores distintos por variável visual;
- Identificação de qual variável visual estava em análise;
- Identificação do percentual de oclusão que estava sendo aplicado na visualização em questão, variando entre 0% (sem oclusão), 50%, 60% ou 70% de área ocluída;
- Resultado da resposta do participante: clique no elemento único da grade (certo), não clicou no elemento único (errado), ou tempo excedido;
- Tempo de resolução da tarefa por visualização medido em segundos.

4.1.4 Resultado do Pré-Teste 1

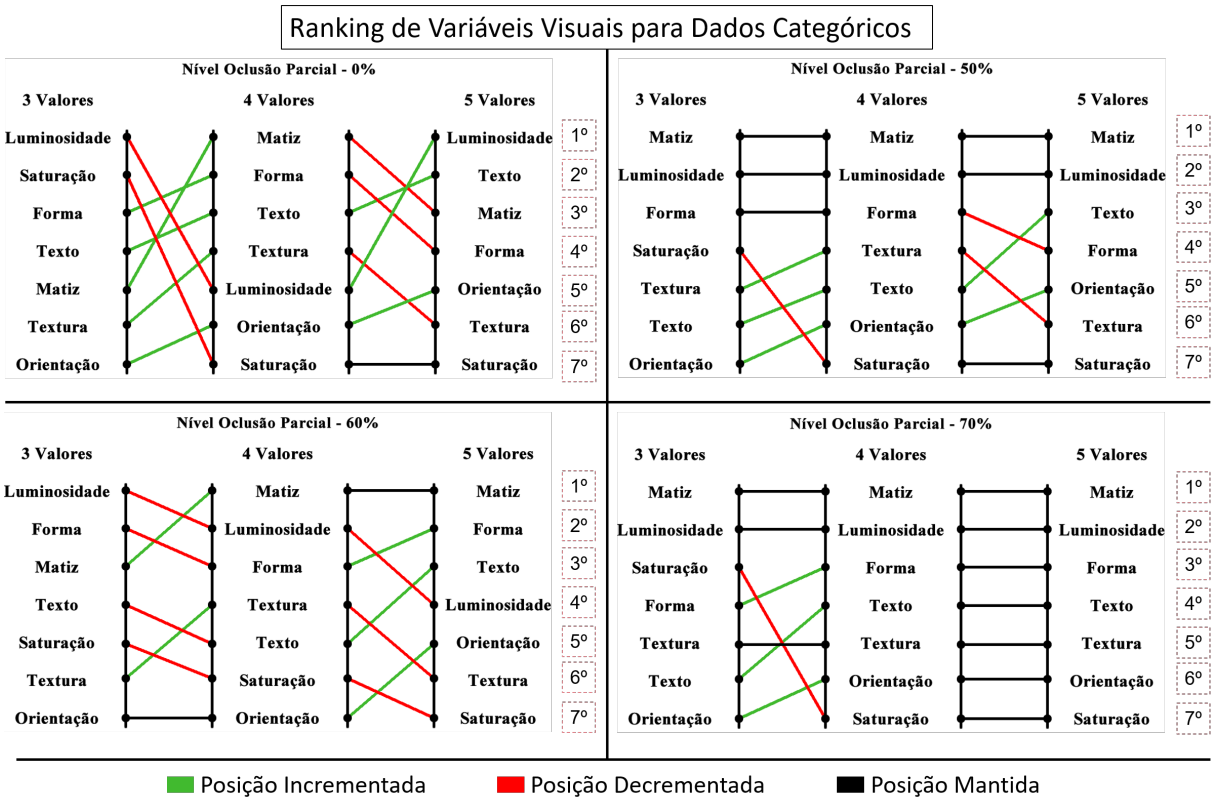
O resultado do pré-teste (ver Figura 29) que avaliou a percepção das variáveis visuais com sobreposição mostrou que quando considerados apenas 3 valores distintos para as variáveis visuais, as variáveis cor - matiz, cor - luminosidade e forma apresentaram os melhores desempenhos quanto a tempo e acurácia, considerando os níveis de oclusão de 50%, 60% ou 70%.

É importante destaca-se que, na configuração de 3 valores distintos, a variável visual cor - saturação obteve um resultado melhor que a variável forma, nos níveis de oclusão de 0% e 70%. Não obstante, as variáveis visuais textura e orientação obtiveram os piores resultados considerando qualquer um dos níveis de oclusão testados.

Quando considerados 4 valores distintos para as variáveis visuais, independente do nível de oclusão (0%, 50%, 60% ou 70%) as variáveis visuais orientação e saturação obtiveram os piores desempenhos. Já as variáveis cor - matiz, cor - luminosidade e forma foram as que apresentaram melhores resultados de tempo e acurácia, quando considerados níveis de oclusão de 50%, 60% e 70%, pois ao considerar o nível de oclusão de 0%, a variável visual texto obteve um resultado melhor do que a variável cor - luminosidade.

Quando foram considerados 5 valores distintos para as variáveis visuais, independente do nível de oclusão (0%, 50%, 60% ou 70%), as variáveis visuais orientação, textura e saturação obtiveram os piores resultados de tempo e acurácia. Assim, as variáveis cor - matiz, cor - luminosidade, forma e texto foram as que apresentaram melhores desempenhos.

Figura 29. Ranking das variáveis visuais para dados categóricos, considerando os níveis de oclusão (0%, 50%, 60% ou 70%) e valores distintos (3, 4 e 5).



Fonte: Adaptado de SANTOS et al. (2019).

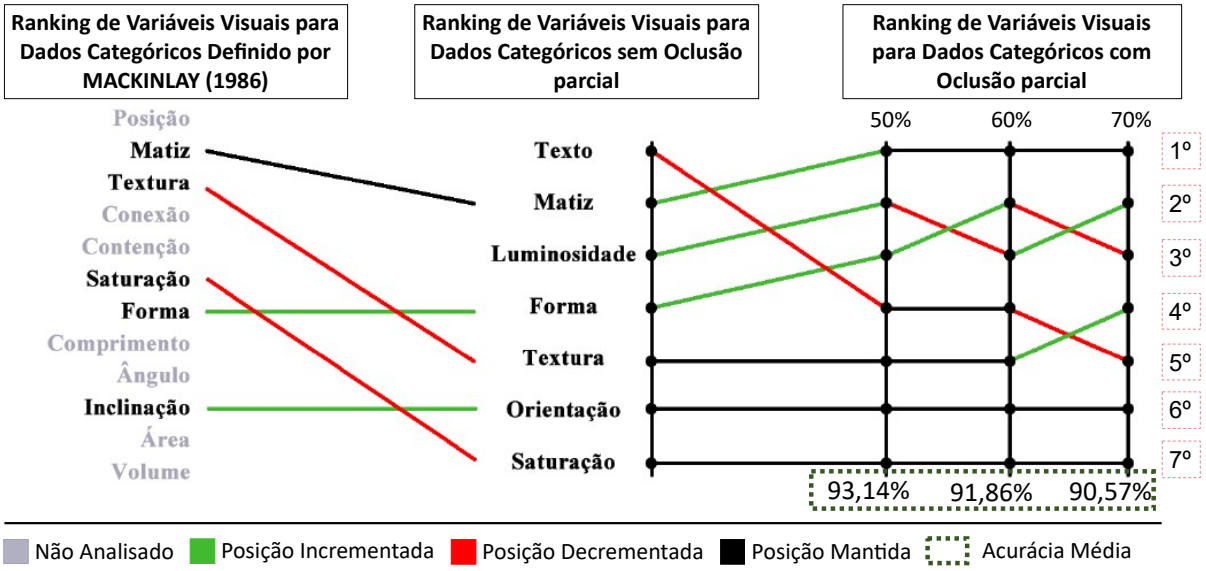
Apesar do estudo desenvolvido neste teste não incluir todas as variáveis visuais utilizadas por MACKINLAY (1986) e não utilizar o mesmo *layout* de visualização aplicado para as análises, foi realizado uma comparação com o ranking apresentado pelo autor sobre percepção visual de variáveis visuais quando empregadas sobre dados categóricos (ver Figura 30).

Primeiramente, foi realizada a comparação das variáveis utilizadas neste estudo sem oclusão (nível de oclusão de 0%) com o ranking de variáveis visuais para dados categóricos de MACKINLAY (1986). É importante destacar que a variável texto, apesar de não ser mencionada no ranking de MACKINLAY (1986), foi a variável visual com melhor desempenho.

É importante destacar que o desempenho entre as variáveis texto, cor - matiz, luminosidade e forma foram aproximados, sendo texto com uma acurácia de 100% e as outras três 99%. O desempate entre as variáveis cor - matiz, cor - luminosidade e forma considerou o tempo de resolução das tarefas realizadas.

Em seguida, foi realizado o ranking das variáveis a partir do nível de oclusão.

Figura 30. Comparação dos resultados com MACKINLAY (1986), que avaliou variáveis visuais em cenários não sobrepostos.



Fonte: Adaptado de SANTOS et al. (2019).

Assim, pode-se destacar que a variável visual cor - matiz mantém-se como a variável com o melhor desempenho considerando os nível de oclusão 50%, 60% e 70%. A variável visual luminosidade aparece logo em seguida, com uma pequena variação no nível de oclusão de 60%, onde a variável forma apresentou um desempenho superior. Esta última, por sua vez, apresentou-se como uma boa opção após a variável luminosidade.

Ressalta-se que a variável que apresentou a maior perda a partir do aumento do nível de sobreposição foi a variável texto. Adicionalmente, a variável textura manteve sua colocação para os níveis de oclusão de 50% e 60%, apresentando uma ligeira melhora no desempenho com 70% de oclusão. Fato semelhante ocorre com as variáveis visuais orientação e saturação, que mantiveram suas respectivas colocações independente do nível de oclusão.

É importante destacar que, dado o cenário de aplicação do teste, a variável visual saturação obteve o pior desempenho e portanto, para o modelo do glifo, a variável visual em questão foi descartada. Para o modelo do glifo a variável visual posição foi utilziada em substituição a variável saturação. Tal troca foi realizada, pois de acordo com o ranking de MACKINLAY (1986) a variável posição é a melhor variável visual para a representação de dados categóricos.

De forma geral, foi calculado a acurácia média entre os níveis de oclusão parcial das variável visuais, a fim de ter uma métrica de avaliação para guiar esta pesquisa sobre qual nível de sobreposição utilizar no modelo de glifo proposto. Desta forma, obteve-se

os 93,14%, 91,86% e 90,57% de acurácia para os níveis de oclusão parcial de 50%, 60% e 70%, respectivamente (ver Figura 30).

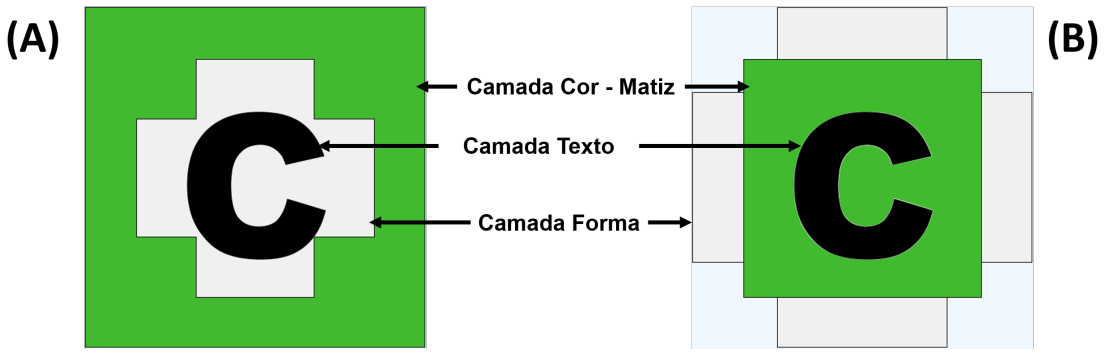
Considerando que todos os níveis de oclusão parcial tiveram uma acurácia média acima de 90% e a necessidade de representação do maior número de camadas possível, de modo a não prejudicar a leitura do glifo, adotou-se o nível de 70% de oclusão parcial no modelo de glifo desta tese.

Adicionalmente, sobre a quantidade de valores distintos a ser representados por uma variável visual, no nível de 70% de oclusão parcial houve uma pequena variação no desempenho entre as variáveis visuais (ver Figura 29) considerando três (3) valores distintos, pois a única variável visual que sofreu uma piora em seu desempenho foi a saturação, as demais mantiveram seu desempenho ou tiveram uma singela melhora. Desta forma, adoutou-se 5 valores distintos entre as variáveis visuais categóricas do modelo do glifo, pois entre quatro (4) e cinco (5) valores distintos não houveram alterações nos desempenhos das variáveis visuais.

4.2 Design do glifo

O glifo desta tese possui forma quadrada, é composto por camadas alinhadas no centro, e que se sobrepõem parcialmente umas às outras. Considerando uma camada N posicionada acima de uma camada N-1, a camada superior N se sobrepõe a N-1 em 70%, visando sobrepor a maior área possível, mas considerando evitar a oclusão total (ver Figura 31).

Figura 31. Design do glifo em camadas. A - Glifo com as variáveis visuais cor - matiz, forma e texto; B - Neste glifo houve uma troca no posicionamento das camadas cor - matiz e forma comparado com A.



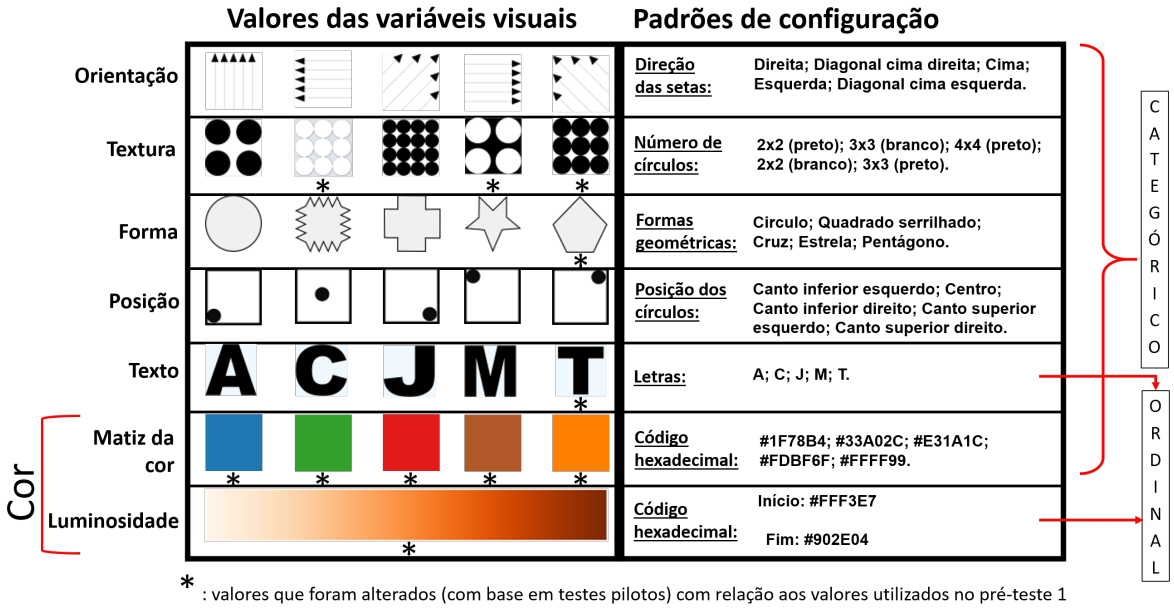
Fonte: Autor do trabalho.

As decisões para a contrução do modelo de glifo foi construido com base nos resultados da subseção 4.1.4, tais como a utilização de 70% de oclusão parcial entre

as variáveis visuais, a adoção das variáveis visuais de melhor desempenho para dados categóricos (segundo o resultado do pré-teste 1), a utilização de cinco (5) valores distintos entre as variáveis visuais para a representação. É importante destacar que durante a implementação deste modelo de glifo alguns testes pilotos foram realiziados, os quais evidenciaram a necessidade de adaptações de alguns valores de variáveis visuais com relação aos utilizados no pré-teste 1 (ver valores marcados com um * na Figura 32).

Para a representação de atributos categóricos, cada camada é mapeada para uma variável visual, que pode ser cor-matiz, forma, textura, texto e posição. É importante mencionar que para a representação de atributos ordinais é necessário que o usuário entre com os dados já ordenados. Os valores das variáveis visuais usadas nas camadas do glifo e o tipo de dados que eles são adequados para representar são apresentados na Figura 32.

Figura 32. Valores das variáveis visuais categóricas e ordinais usadas nas camadas do glifo e o tipo de dados que eles são adequados para representar.



Fonte: Autor do trabalho.

Ao observar a Figura 31-A, nota-se o uso das variáveis visuais na seguinte ordem: cor, forma e letra, sendo que cada variável visual é mapeada para uma camada do glifo. A Figura 31-B ilustra uma troca entre as camadas forma e cor, quando comparada a Figura 31-A apresentando a seguinte ordem: cor, forma e letra. O modelo de design em camadas foi utilizado para que o usuário possa, de forma mais flexível, configurar o glifo que representará seus dados.

A posição das camadas é totalmente permutável e o usuário tem a liberdade de escolher como será realizado o mapeamento visual entre as variáveis visuais e os atributos,

ou seja, é possível escolher variáveis visuais de fácil percepção para dados mais importantes, de acordo com a preferência do usuário.

Para representar dados quantitativos, foi considerado o uso do profile glifo na camada superior do glifo proposto. Um profile glifo representa observações n-dimensionais com n barras (WARD; GRINSTEIN; KEIM, 2015)), a altura das barras é proporcional aos valores dos dados selecionados e as barras são dispostas próximas umas das outras (TOIT; STEYN; STUMPF, 1986).

No design do glifo desta tese, o profile glifo é um tipo específico de camada que codifica, através da variável visual tamanho, informações quantitativas com uma miniatura de gráfico de barras. Considerando que qualquer sobreposição pode prejudicar a compreensão do profile glifo, ele é sempre posicionado na camada de nível superior para ficar totalmente visível.

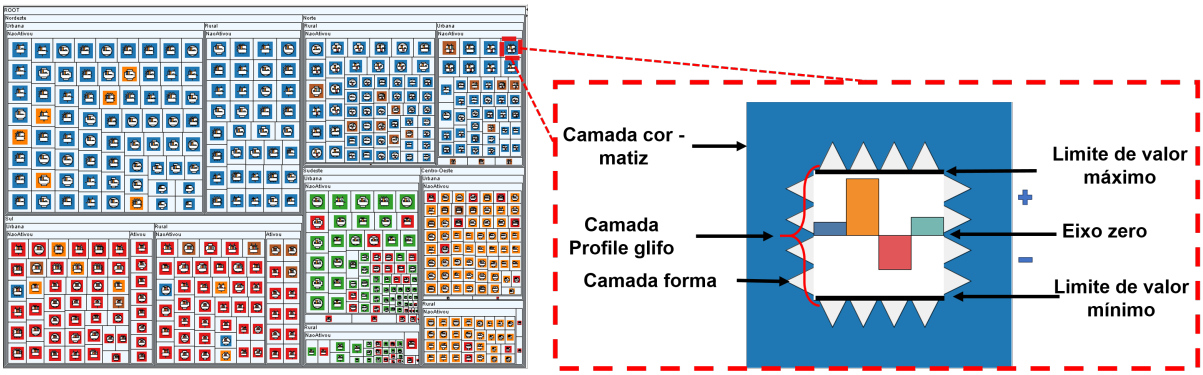
O usuário pode codificar muitas dimensões quantitativas na camada do profile glifo usando barras. Para facilitar a distinção entre as dimensões, representadas pelas barras, foi adotado o uso da variável visual cor - matiz. No design deste glifo, o usuário pode adicionar oito dimensões, ou seja, oito barras coloridas. Considerando o contexto desta tese, é importante destacar que o profile glifo pode adicionar mais de oito barras, no entanto, sua visibilidade pode ser prejudicada devido à área dos itens do Treemap. Os profile glifos podem se beneficiar de pesquisas adicionais para quantificar sua eficácia em diferentes espaços de tela, mas esse tópico está fora do escopo desta tese.

O profile glifo permite o mapeamento de valores negativos, usando uma barra inversa que inicia na origem e cresce para baixo. Durante o desenvolvimento do design do profile glifo optou-se por adicionar linhas de referência marcando a altura mínima e máxima que as barras podem alcançar. O uso de tais linhas de referência foram adotadas para permitir que o usuário possa comparar cada barra do profile glifo, ou seja, as linhas de referência servem de guia para que o usuário possa notar o quão próximo do valor máximo do atributo da base de dados, ao qual a barra está mapeada, uma dada barra pode estar.

Considerando o contexto desta pesquisa, antes de plotar um profile glifo, os dados eram normalizados, pois cada item do Treemap não possui o mesmo tamanho ou área. A Figura 33 apresenta um exemplo de Treemap com glifos, onde cada glifo possui três camadas (variáveis visuais), com a última camada — camada N — tendo um profile glifo com cada barra representando uma dimensão quantitativa dos dados.

Ao observar o Treemap da Figura 33 é possível notar que o glifo aumenta a quantidade de informações presente na técnica Treemap. No entanto, nos itens do Treemap que possuem área reduzida, nos quais já era complexo visualizar a informação, a adição dos glifos a visualização dos dados tornou-a ainda mais densa. Dessa forma, foi necessário

Figura 33. Treemap com glifos em camadas, onde os glifos possuem três camadas. A camada superior apresenta um profile glifo, que representa dados quantitativos de quatro dimensões. As barras indicam valores positivos (acima do eixo zero) ou negativos (abaixo do eixo zero), sempre respeitando as linhas de referência que indicam o tamanho máximo e mínimo da barra.



Fonte: Autor do trabalho.

o desenvolvimento de uma solução que buscasse amenizar essa sobrecarga visual, exibindo os glifos até os itens cuja a percepção do glifo não seria afetada. Nas próximas seções será apresentado o pré-teste realizado para avaliar a percepção do glifo e o glifo resultante deste pré-teste.

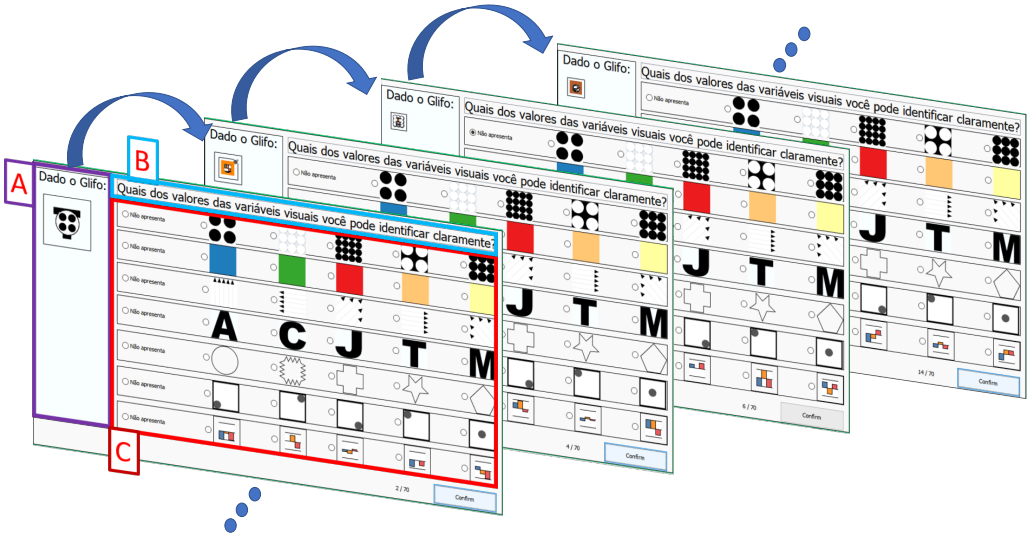
4.3 Configurações do Pré-teste 2

Para avaliação da percepção sobre glifos foi desenvolvida uma aplicação que apresentava um glifo e coletava as respostas dos participantes a respeito da identificação das variáveis visuais que compunham o glifo (Figura 34).

É importante informar que, para a realização dos testes em questão, o suporte computacional e o perfil dos participantes foram os mesmos da seção 4.1.1 e 4.1.2, respectivamente. Os participantes receberam um treinamento a respeito das variáveis visuais utilizadas, sendo orientação, texto, forma, posição, cor (matiz), textura e tamanho, e seus respectivos valores (ver Figura 35). O participante era informado que as configurações do glifo seriam geradas aleatoriamente, ou seja, o glifo iria variar aleatoriamente de tamanho (lado de um glifo sofreu variação de 5px a 125px), número de camadas (variáveis visuais) e a ordem entre elas.

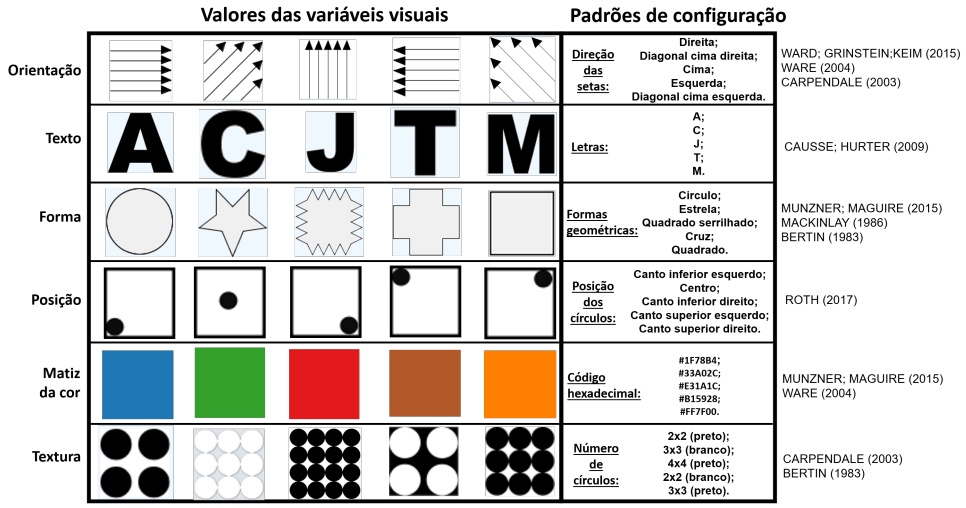
Em seguida, o participante era informado de que deveria visualizar o glifo gerado na área A da Figura 34 e, dada a configuração do glifo, responder a tarefa da área B da Figura 34. Para informar sua resposta o participante selecionava as opções dos valores das variáveis visuais na área C da Figura 34.

Figura 34. Aplicação desenvolvida para que os participantes pudessem avaliar a percepção do glifo. A - Exemplo de glifo gerado, variando aleatoriamente tamanho e camadas dos glifos; B - Área de apresentação da tarefa de InfoVis; C - Controles para que o participante pudesse informar quais camadas do glifo de A eram possíveis de ser identificadas.



Fonte: Autor do trabalho.

Figura 35. Valores utilizados neste estudo que são referentes às camadas orientação, texto, forma, posição, cor (matiz) e textura das variáveis visuais que compõem os glifos.



Fonte: Adaptado de (SANTOS et al., 2019; ROTH, 2017; CAUSSE; HURTER, 2009; CARPENDALE, 2003; WARD; GRINSTEIN; KEIM, 2015; WARE, 2004; MUNZNER; MAGUIRE, 2015; MACKINLAY, 1986; BERTIN, 1983)

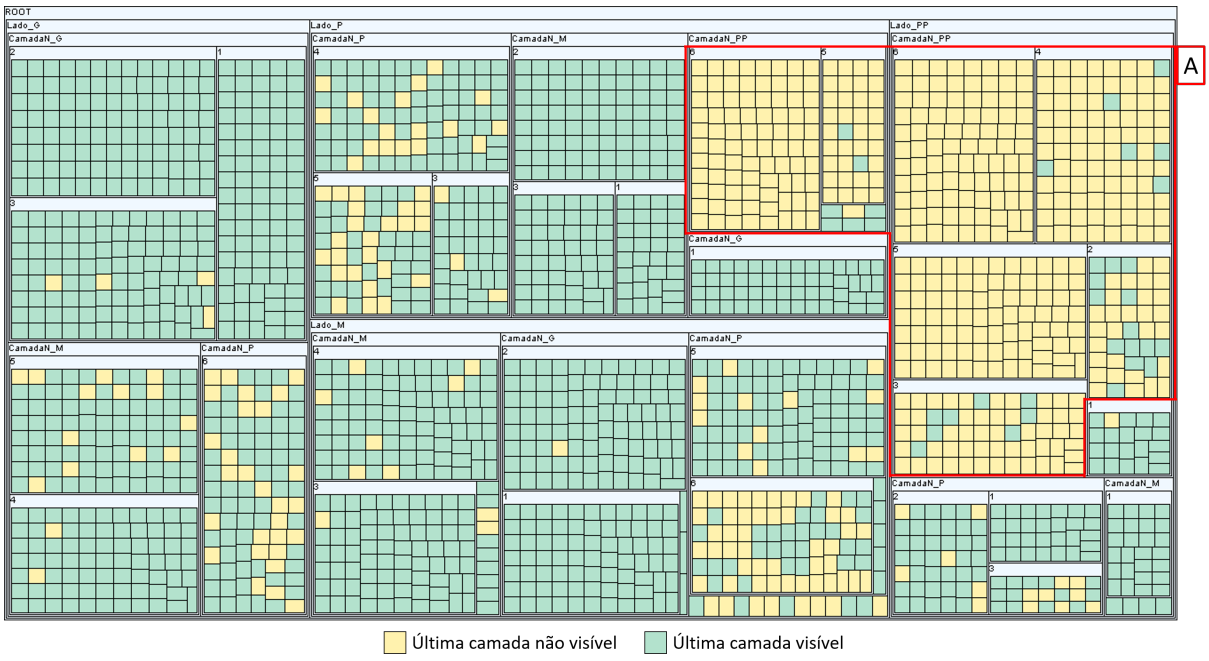
Durante a realização dos testes os participantes executavam uma única tarefa, sendo esta: *"Quais dos valores das variáveis visuais você pode identificar claramente?"* (área B da Figura 34). Para a execução desta tarefa não foram estabelecidos tempo mínimo ou máximo, deixando o participante livre para informar quais valores das variáveis visuais (camadas) ele conseguia identificar. No total a ferramenta gerava 70 glifos. Durante esse teste participaram 30 voluntários. Ao final era gerado um arquivo CSV (*Comma-Separated-Values*) com os seguintes dados:

- Área: item do Treemap; orientação; texto; forma; posição; cor (matiz); textura; tamanho (profile glifo).
- Presença da camada no glifo: orientação; texto; forma; posição; cor (matiz); textura; tamanho (profile glifo).
- Ordem das camadas do glifo.

4.3.1 Resultado do Pré-Teste 2

A análise dos resultados do pré-teste que avaliou a percepção sobre glifos considerou a área da última camada do glifo (camada N), o menor lado de um Treemap item, e o número de camadas de um glifo, a Figura 36 apresenta uma visão geral do resultado.

Figura 36. Visão geral do resultado do pré-teste. Hierarquia do Treemap: Menor lado de um item > Área da última camada (camada N) > Número de camadas do glifo.



Fonte: Autor do trabalho.

Para auxiliar a interpretação dos dados do pré-teste em questão, foi realizada a classificação das colunas "áreaCamadaN"e "MenorLado", que representam a área da última camada de um glifo (em pixels) e o menor lado de um Treemap item (em pixels), respectivamente. Tais colunas foram categorizadas tendo como critégio seus respectivos quartis:

- áreaCamadaN:
  - CamadaN_PP (1º quartil): $[1; 81[$;
  - CamadaN_P (2º quartil): $[81; 256[$;
  - CamadaN_M (3º quartil): $[256; 1072,75[$;
  - CamadaN_G (4º quartil): $[1072,75; 7569]$;
- MenorLado:
  - Lado_PP (1º quartil): $[5; 34[$;
  - Lado_P (2º quartil): $[34; 66[$;
  - Lado_M (3º quartil): $[66; 96[$;
  - Lado_G (4º quartil): $[96; 124]$.

Dado o pré-processamento realizado nos dados gerados no pré-teste em questão, a Figura 36-A ilustra que quando o glifo está em um item cujo o menor lado é de tamanho PP ($[5; 34[$), a área da última camada do mesmo glifo for também PP ($[1; 81[$) e este glifo apresentar mais de uma (1) camada, os usuários não conseguiram identificar o glifo, pois estes já não são mais visiveis. O mesmo ocorre se o glifo estiver em um item do Treemap onde o menor lado é de tamanho P ($[34; 66[$), a área da última camada do glifo for PP ($[1; 81[$) e se o glifo tiver mais de quatro (4) camadas.

De forma geral, os usuáios apresentaram mais dificuldade de visuzaliar os glifos quando eles possuíam a área da última camada PP ($[1; 81[$) ou P ($[81; 256[$) e quando possuíam mais de 3 (três) camadas, em especial se o menor lado do item do Treemap fosse PP ($[5; 34[$) ou P ($[34; 66[$). Dado este cenário, houve a necessídade de criar um mecanismo inteligente que fosse possível automatizar a criação de regras para entregar para o usuário apenas as camadas que seriam visiveis, impedindo que visualização se torne carregada visualmente.

4.3.2 Árvore de decisão

Dada a necessidade de criar regras que permitam que a visualização com glifos se torne menos carregada, com os dados gerados pelo pré-teste 2 (seção 4.3), que buscava

avaliar a percepção dos participantes no que tange a visibilidade das variáveis visuais, número de camadas e o tamanho do glifo, foi treinada uma árvore de decisão.

É importante destacar que de posse dos resultados do pré-teste 2 poder-se-ia gerar uma análise manual e com isso obter regras para melhorar a visualização dos glifos, porém o volume de regras poderia ser muito grande, a análise manual tornar-se-ia difícil e custosa, o que gerou a necessídade de uma ferramenta de análise automática. Adicionalmente, não faz parte do escopo desta tese a análise de diversas técnicas de aprendizado de máquina. Para está tese a árvore de decisão foi escolhida pois permite uma rápida interpretação e implementação do modelo (BARBIERI, 2011).

Com o pré-teste 2 foram obtidas 2100 avaliações de glifos em camadas, na qual 80% das amostras foram utilizadas como base para o treinamento dea árvore de decisão e 20% das amostras foram deixadas para realização do teste de acurácia para aferição do correto treinamento da árvore.

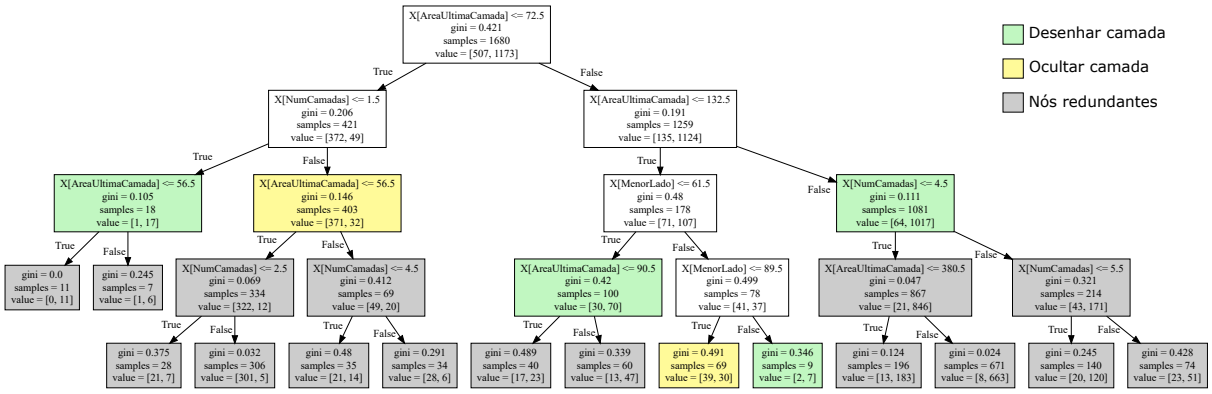
Para a construção da árvore de decisão foram utilizadas as seguintes *features*: número de camadas do glifo, o menor lado do *Treemap* item, a área da última camada do glifo. Para treinar a árvore de decisão foi utilizada a classe DecisionTreeClassifier da biblioteca da biblioteca Scikit-learn[2] (PEDREGOSA et al., 2011). Os parâmetros utilizados no treinamento foram: max_depth = 4 (profundidade máxima da árvore), min_samples_split = 10 (número mínimo de amostras para considerar a realização de uma divisão).

A árvore gerada é apresentada na Figura 37. Após uma análise da árvore gerada foram removidos alguns nós redundantes, ou seja, dois nós irmãos com a mesma resposta, tais nós estão representados na cor cinza na Figura 37. Os nós na cor verde indicam que a última camada do glifo deve ser desenhada, enquanto que os nós na cor amarela indicam que a última camada do glifo deve ser ocultada.

O modelo de árvore de decisão gerada obteve 0.919 de acurácia, 0.91 de precisão e 0.919 de F1-score, quando testada com a base de teste (20% das amostras). Tais métricas quantitativas sugerem que a árvore de decisão de fato aprendeu as regras expressadas empiricamente pelos participantes, como pode-se observar na matriz de confusão ilustrada na Figura 38. A matriz mostra que o modelo previu corretamente 116 ocorrencias que uma camada N de um glifo não era visível e que está deveria ser ocultada (verdadeiro negativo), assim como o modelo previu que 270 ainda eram visiveis e que deveriam ser exibidas (verdadeiro positivo).

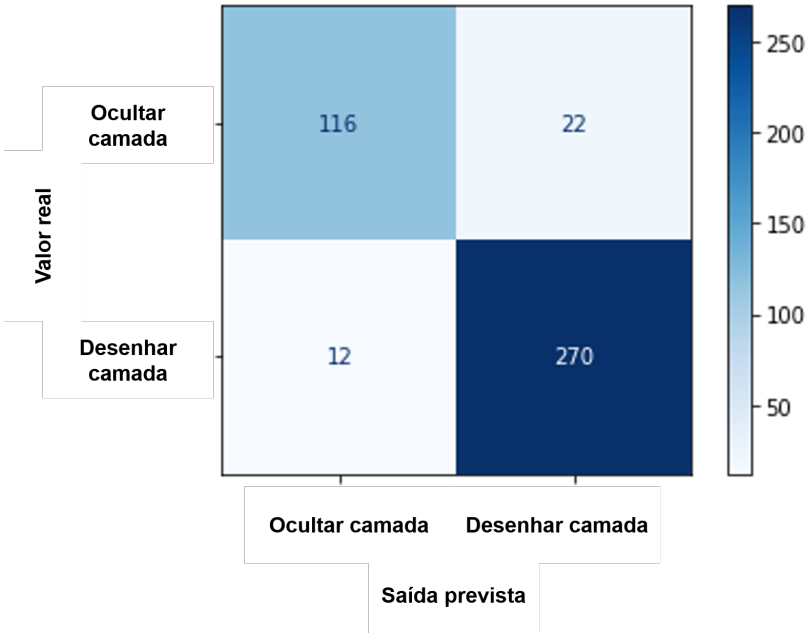
[2] http://scikit-learn.org

Figura 37. Árvore de decisão gerada pela biblioteca Scikit-learn.



Fonte: Autor do trabalho.

Figura 38. Matriz de confusão exibindo as frequências de classificação para cada classe do modelo.



Fonte: Autor do trabalho.

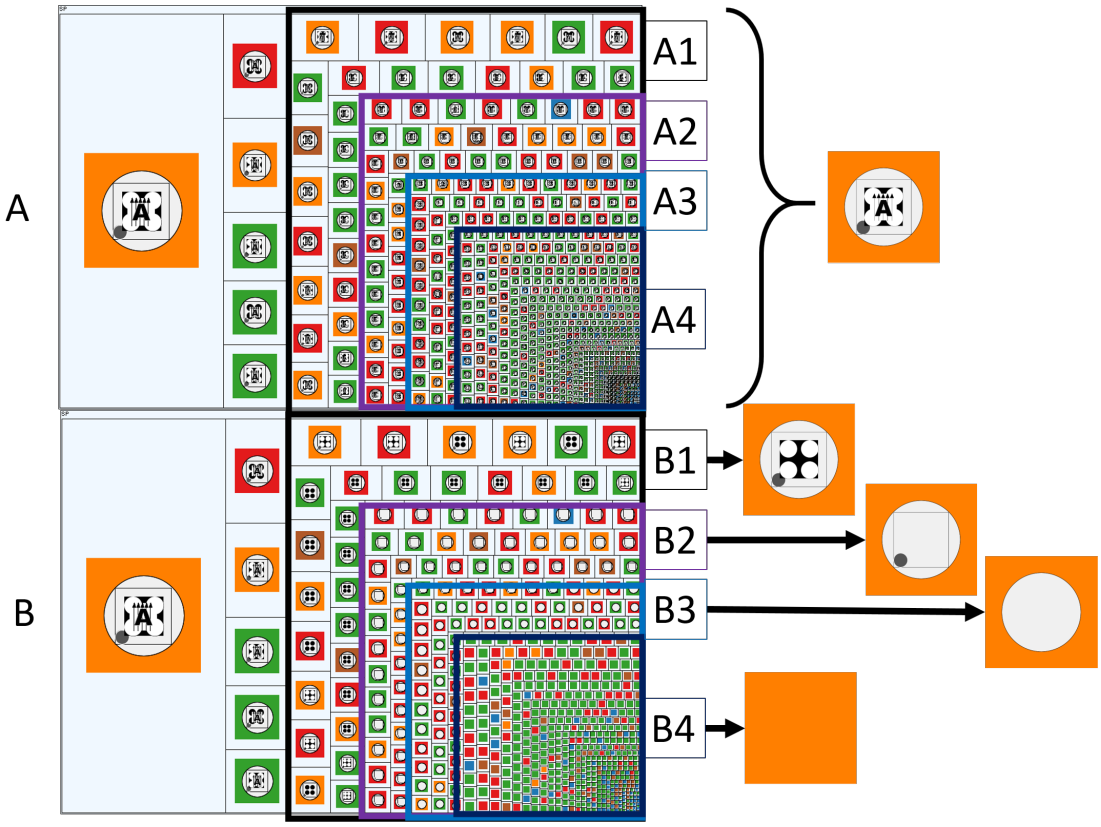
4.4 Glifo adaptativo

Os itens do Treemap gerados podem apresentar áreas pequenas, o que influencia diretamente no tamanho dos glifos criados (ver Figura 39 - A1, A2, A3, A4), gerando o problema de perda de percepção das variáveis visuais que compõem um glifo. Contudo algumas variáveis visuais continuam mantendo seu poder de comunicação mais que outras conforme a área vai diminuindo.

Ao invés de se retirar totalmente o glifo da visualização de dados quando a área

for pequena e perder totalmente o poder da análise multidimensional, o objetivo dos glifos adaptativos é avaliar se ainda há alguma variável visual que mantém seu poder de comunicação e deixá-la na visualização (ver Figura 39 – B1, B2, B3, B4). Com isso, não só se mantém dados visuais úteis para tarefas de busca, comparação, descoberta de tendências, etc, mas melhora-se de maneira geral a qualidade da visualização dos dados.

Figura 39. A - Visualização gerada sem aplicação da função adaptativa para construção dos glifos; B - A mesma visualização agora com aplicação do algoritmo adaptativo para construção dos glifos; B1 - Remoção das camadas Letra e Orientação; B2 - Remoção da camada Textura; B3 – Remoção da camada Posição; B4 - Após a execução do algoritmo adaptativo restou apenas a camada cor no glifo.



Fonte: Autor do trabalho.

Para obtenção das regras que determinam quais glifos devem ser selecionados e quais variáveis visuais devam ser ocultadas do glifo, foi utilizada a árvore de decisão treinada com dados do pré-teste 2 (subseção 4.3.2).

Após a verificação da eficácia da árvore de decisão, esta foi colocada em produção na ferramenta como uma opção para o usuário selecionar quando os glifos deveriam ser construídos de forma adaptativa ou não. Os parâmetros considerados pela árvore de decisão foram: número de camadas do glifo, o menor lado de um item do Treemap e a área da última camada do glifo.

Para cada item do Treemap é realizada a avaliação de sua área, do número de camadas do glifo e da área da última camada (camada N) do glifo na árvore de decisão. Assim, a árvore retorna o resultado se o glifo, pertencente ao item do Treemap em análise, deve ou não plotar sua camada N. De posse do resultado, se a camada N do glifo foi ocultada, este é submetido a uma nova avaliação. As avaliações do glifo ao modelo de aprendizado de máquina é concluída quando não há mais camadas para serem removidas, ou quando a árvore de decisão retorna o resultado informando que a área da camada N do glifo ainda é visível.

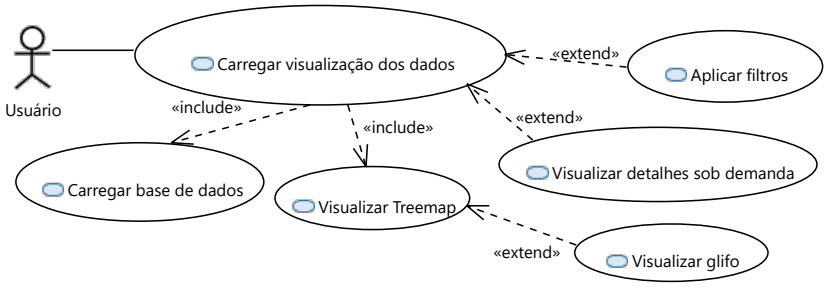
5 PROTÓTIPO DE APOIO

A abordagem desta pesquisa trata do uso de glifos tendo como layout a técnica treemap. O protótipo[1] possui como principais funcionalidades básicas:

- Carregar bases de dados;
- Carregar visualização dos dados;
- Visualizar treemap;
- Configurar glifos;
- Visualizar detalhes sob demanda;
- Aplicar filtros

A Figura 40 ilustra o diagrama de caso de uso do protótipo, onde pode ser observada que o caso de uso principal trata da visualização dos dados. Para tal, o usuário precisará “Carregar a base de dados” para enfim poder “Visualizar treemap”. A configuração da técnica treemap fica a cargo do usuário, ou seja, a definição dos atributos que definirão o tamanho dos itens, cores, rótulos, hierarquias e Visualizar glifo” são opcionais ao usuário. A funcionalidade “Aplicar filtros” permite ao usuário filtrar itens da visualização, sendo uma funcionalidade opcional, assim como “Visualizar detalhes sob demanda”.

Figura 40. Diagrama de caso de uso do protótipo.



Fonte: Autor do trabalho.

Após o usuário realizar a funcionalidade "Visualizar Treemap", o usuário pode executar a funcionalidade "Visualizar glifo"(ver diagrama de caso de uso em Figura 41), para isso a funcionalidade "Criar glifo em camadas" deverá ser executada. Desta forma,

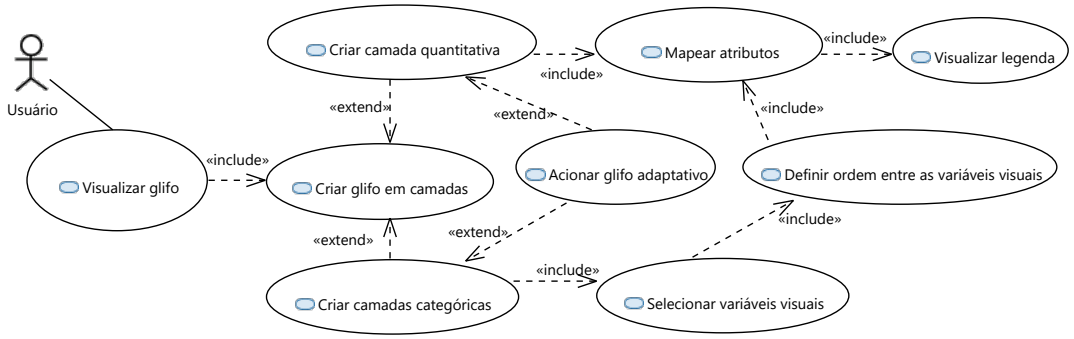
[1] Link para download do protótipo: http://labvis.ufpa.br/treemapglyphs

o usuário poderá "Criar camadas categóricas" para poder representar dados categóricos, e/ou "Criar camada quantitativa"para representar dados quantitativos.

Ao "Criar camadas categóricas", o usuário deverá "Selecionar variáveis visuais", podendo ser textura, cor, forma, texto, orientação e posição, e então "Definir a ordem entre as variáveis visuais". Uma vez já escolhida quais são as variáveis visuais e a ordem de representação, o usuário deverá executar a funcionalidade "Mapear atributos", que consiste em mapear atributos da base de dados às variáveis visuais já selecionadas. O usuário pode verificar o resultado do mapeamento visual ao executar a funcionalidade "Visualizar legenda", que ilustra o mapeamento dos atributos aos valores das variáveis visuais selecionadas.

Ao "Criar camada quantitativa" o usuário deverá "Mapear atributos", e em seguida ordenar os atributos quantitativos. Por fim, o usuário poderá "Visualizar legenda". O caso de uso "Acionar glifo adaptativo" pode ser ativado pelo usuário após a criação de um glifo com camadas categóricas ou quantitativas, ou ambas.

Figura 41. Caso de uso da funcionalidade Visualizar glifo.

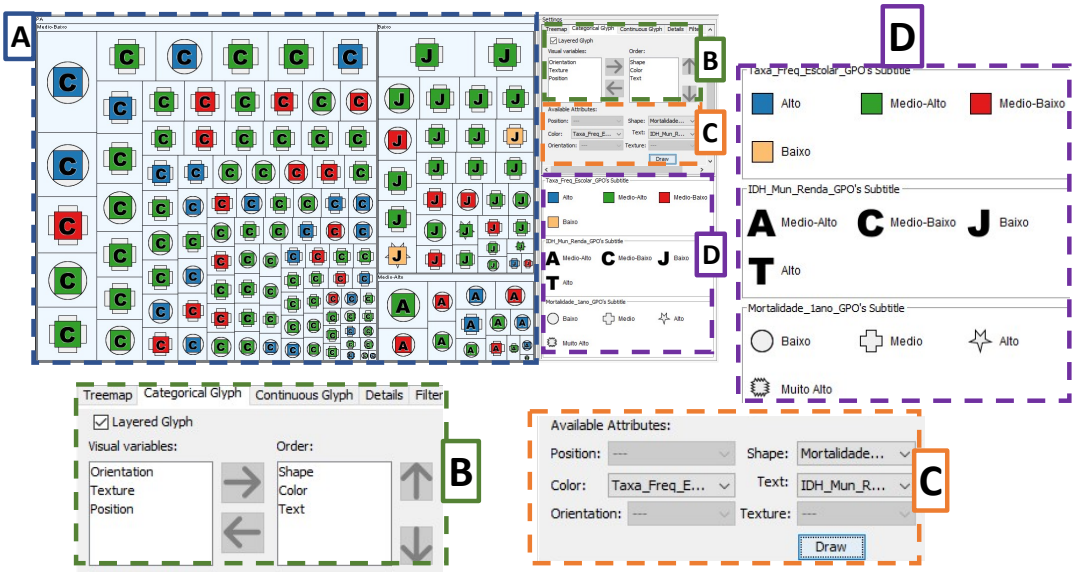


Fonte: Autor do trabalho.

A Figura 42 ilustra uma visão geral do protótipo de apoio desenvolvido. Na área A é possível observar a técnica squarified Treemap, já plotada com os glifos em camadas representando os atributos mapeados de uma base de dados. Na área B, nota-se a área de configuração dos glifos categóricos, local onde é realizada a seleção das variáveis visuais e a ordenação das camadas do glifo. No exemplo da figura em questão, apenas três variáveis visuais foram escolhidas para compor o glifo.

Na área C da Figura 42 é exibida a região onde é realizado o mapeamento entre as variáveis visuais selecionadas e os atributos da base de dados. A figura mostra que o mapeamento dos atributos às variáveis visuais só ocorrerá para as três variáveis visuais selecionadas anteriormente. Por fim, é exibida a legenda dos glifos na área D da Figura 42, ilustrando o mapeamento entre os valores das variáveis visuais aos valores dos atributos da base de dados, respeitando as configurações definidas pelo usuário.

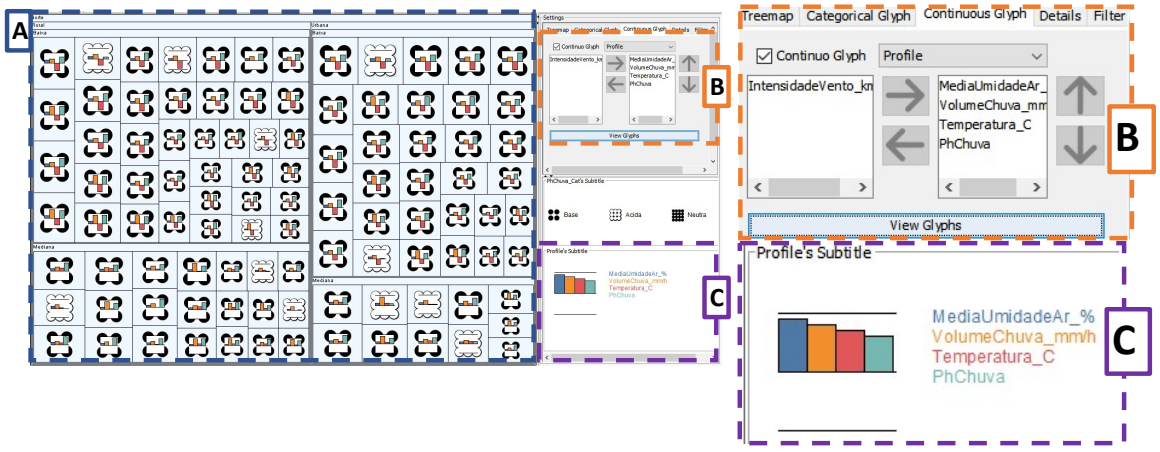
Figura 42. A – Visualização do Treemap; B - Área de configuração das variáveis visuais do glifo; C - Área de mapeamento dos atributos as variáveis visuais; D - Área de exibição das legendas dos glifos.



Fonte: Autor do trabalho.

A Figura 43 ilustra uma visão geral do protótipo, apresentando os glifos em camadas no Treemap, sendo que a última camada do glifo está representando dados quantitativos através do profile glifo (Figura 43 - A). Na área B é possível observar onde é realizada o mapeamento entre os atributos da base de dados e as dimensões do profile glifo. A área C apresenta a legenda do profile glifo, que utiliza cores para ajudar a identificar as dimensões mapeadas em cada barra.

Figura 43. A – Visualização do Treemap; B - Área de configuração das dimensões do profile glifo; C - Área de exibição da legenda do profile glifo.

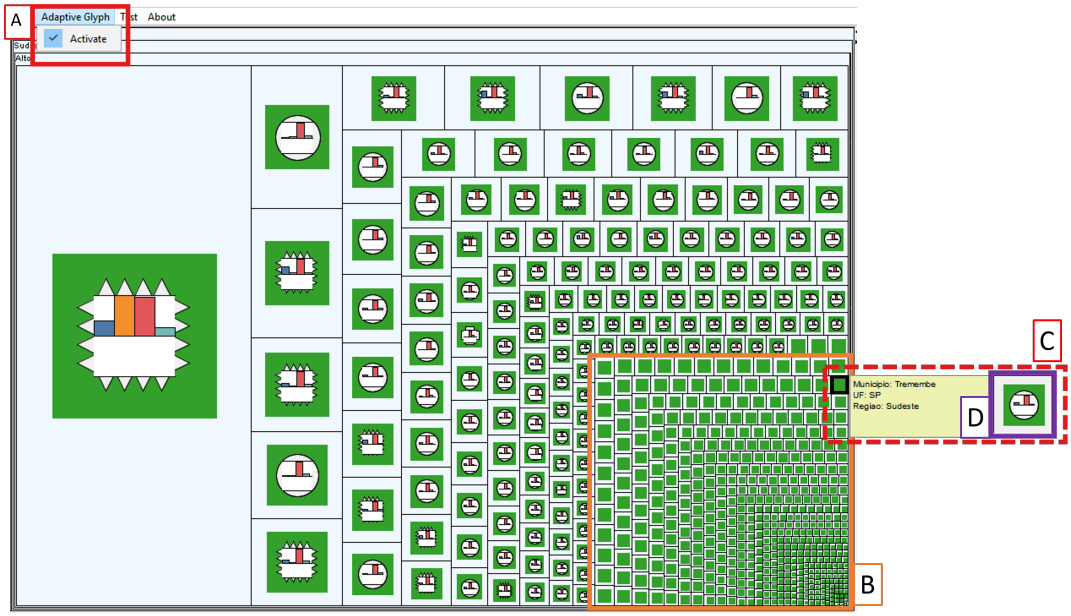


Fonte: Autor do trabalho.

O usuário pode ativar a funcionalidade glifo adaptativo clicando no menu "Glifo adaptativo"(ver Figura 44 - A). Ao ativar os glifos adaptativos, a ferramenta realiza a verificação de cada item do Treemap, visando avaliar a área dos itens. Essa avaliação é realizada consultando a árvore de decisão (ver subseção 4.3.2) que retorna se o glifo deve ser desenhado ou não, baseada no número de camadas presentes no glifo, na área da última camada e no menor lado do glifo.

Com o resultado da árvore de decisão a ferramenta remove a última camada do glifo e submete novamente o item com os glifos ao algoritmo da árvore de decisão. As consultas à árvore de decisão terminam quando não há mais camadas a serem removidas ou quando o algoritmo retorna que a área da última camada é visível (ver Figura 44 - B). Para auxiliar na visualização dos glifos adaptativos, os detalhes sob demanda (Figura 44 - C) exibem o glifo com todas as camadas (Figura 44 - D).

Figura 44. A – Visualização do Treemap com os glifos adaptativos; B - Glifos adaptativos onde as camadas forma e glifo adaptativo foram removidas; C - Área de exibição dos detalhes sob demanda; D - Glifo em camadas sendo exibido nos detalhes sob demanda.



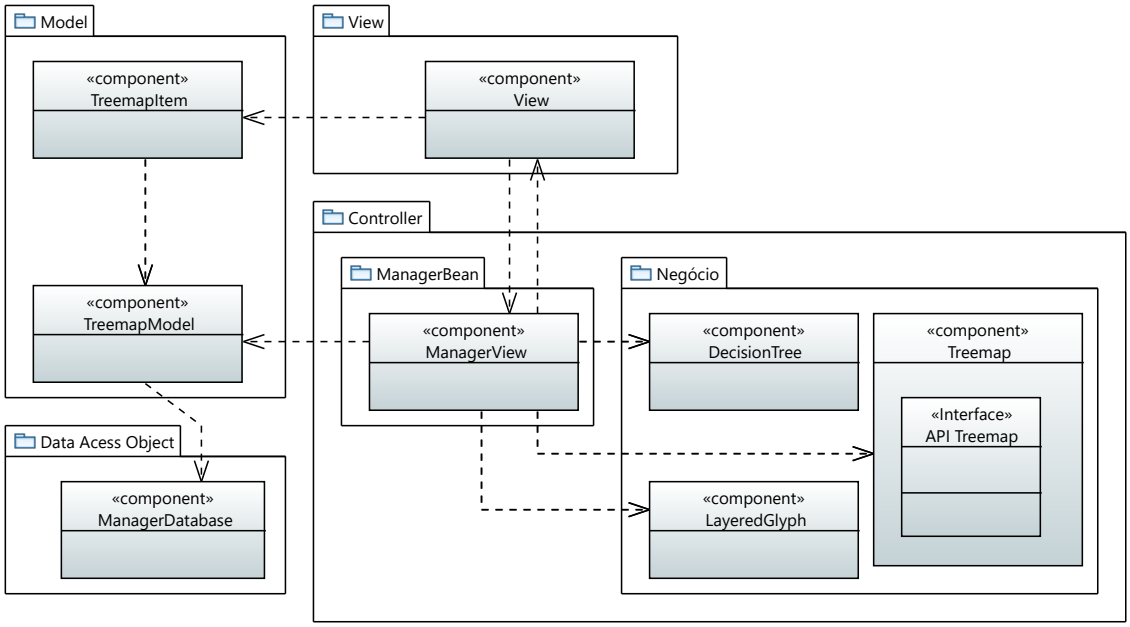
Fonte: Autor do trabalho.

Durante a implementação do protótipo foi adotado o padrão de arquitetura MVC (*Model-View-Controller*), a fim de proporcionar ao protótipo um baixo grau de acoplamento entre seus componentes, permitindo isolar as regras de negócio da lógica de apresentação (GUI). A Figura 45 ilustra o padrão de arquitetura adotado e o relacionamento entre seus componentes.

O componente “View” é responsável por plotar as visualizações, assim como pro-

cessar qualquer evento de mudança na GUI. O componente “ManagerView” é responsável por gerenciar a lógica de desenho vinda dos componentes de negócio e as envia a “View”. Para tal o “ManagerView” busca informações no componente do modelo, que por sua vez é montado com dados oriundos do componente “ManagerDatabase”. Adicionalmente, o componente “ManagerView” processa as interações do usuário e as repassa aos componentes de negócio correspondente, podendo ser: “Treemap”, “LayeredGlyphs” e “DecisionTree”.

Figura 45. Padrão de arquitetura adotado e o relacionamento entre seus componentes.



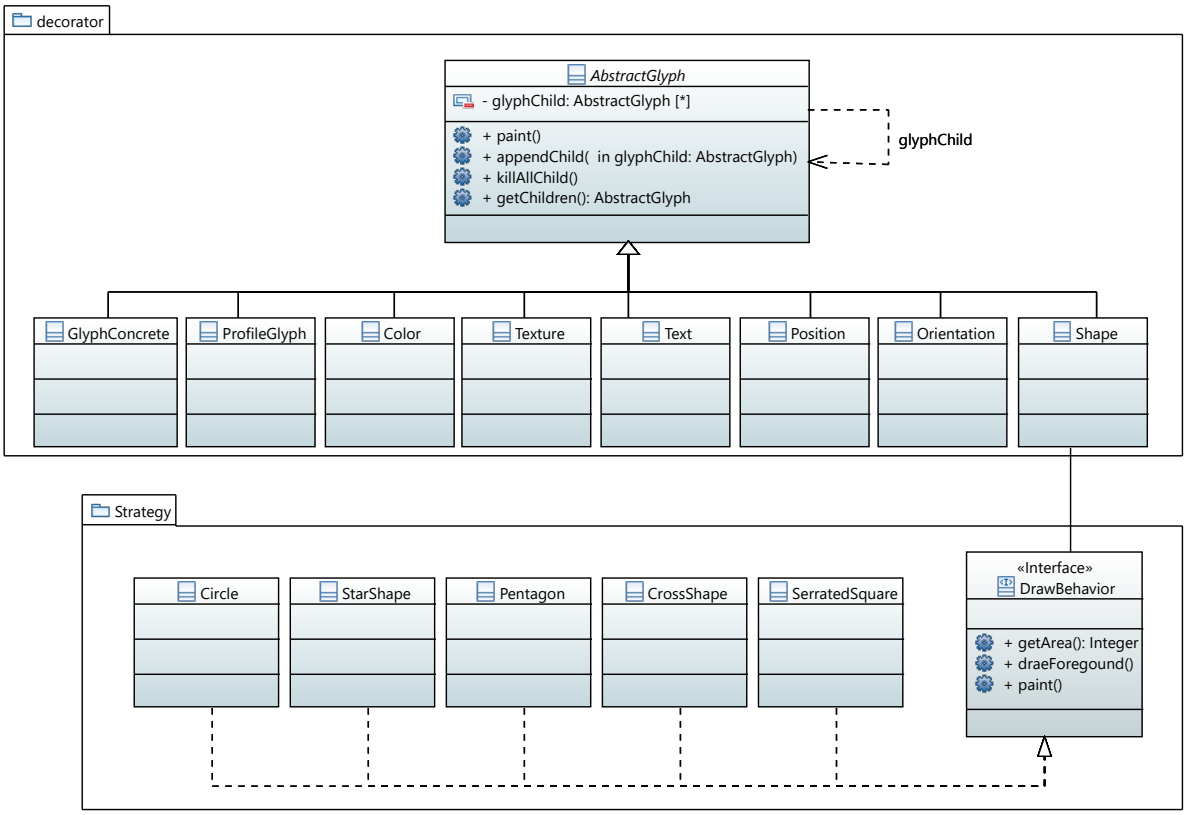
Fonte: Autor do trabalho.

O componente “Treemap” é responsável por criar uma instância da biblioteca TreemapAPI (BOUTHIER, 2001) e repassar a ela as configurações da técnica Treemap definidas pelo usuário (definidas na “View”). O componente “LayeredGlyphs” monta o modelo de glifo em camadas, seguindo as configurações definidas na “View”, tais como variáveis visuais utilizadas e sua respectiva ordem, e quais atributos serão representados pelas variáveis visuais. Por fim, o componente “DecisionTree” é responsável por acionar a árvore de decisão que analisa o tamanho dos itens do treemap e a área visível dos glifos.

A Figura 46 ilustra o diagrama de classes utilizado para a criação dos glifos multidimensionais em camadas. Para a implementação dos glifos foi utilizado o *decorator design pattern* (FREEMAN et al., 2004), a fim de permitir que os glifos pudessem variar a quantidade e a ordem das variáveis visuais. Para tal, cada item do Treemap possui um objeto “AbstractGlyph”, a partir daí cada variável visual, que também é um “AbstractGlyph”, é criada compondo a lista de glifos filho do objeto “AbstractGlyph” de um item do treemap. O glifo proposto possui as variáveis visuais textura, cor, forma, letra, posição

e orientação, para mapear atributos categóricos, representados pelas respectivas classes Texture, Color, Shape, Text, Position e Orientation. A classe Shape possui o modelo das primitivas visuais, tais como círculo, estrela, pentagono, etc. Para mapear atributos quantitativos foi implementado o glifo Profile glifo.

Figura 46. Diagrama de classes usado para implementar os glifos em camadas.

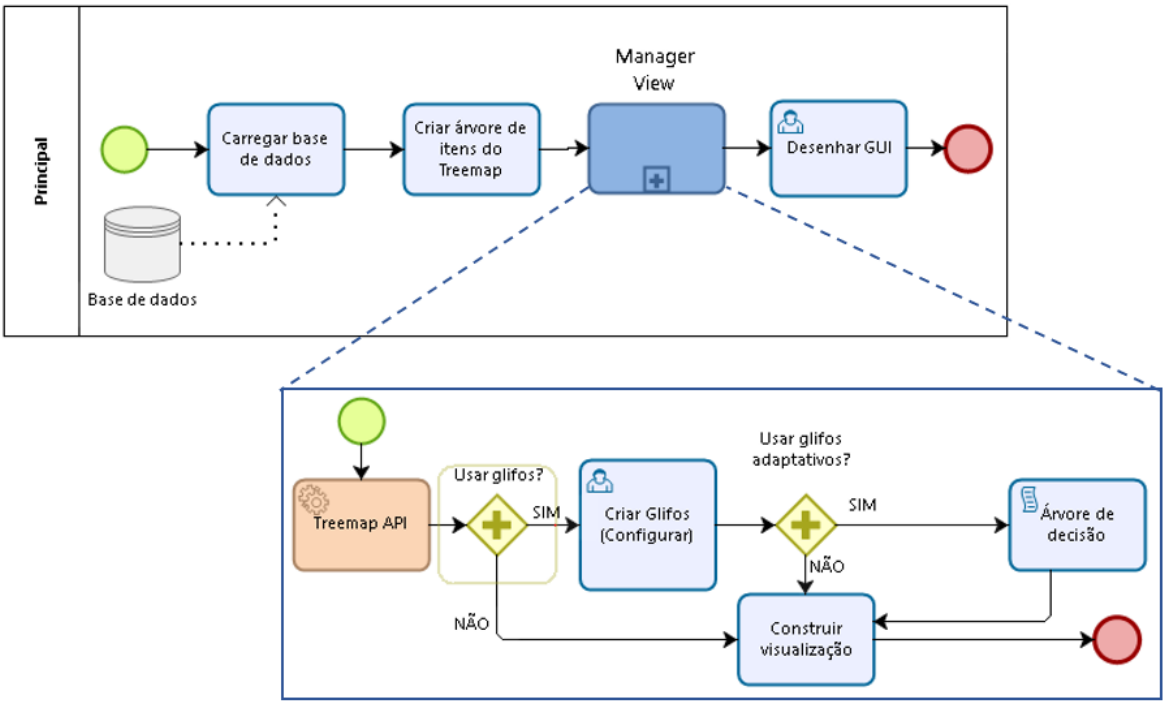


Fonte: Autor do trabalho.

De forma geral, o fluxo do processo para a construção de uma visualização no protótipo pode ser visto Figura 47. O primeiro passo é selecionar os dados a serem analisados, cuja fonte pode ser arquivo texto ou uma view de um banco de dados. Posteriormente é criada uma estrutura de dados em árvore que será utilizada para construção do layout do Treemap. Essa árvore é encaminhada ao módulo “Manager View” que é o responsável por acionar o módulo do Treemap para construção da camada de visualização Treemap.

Por opção do usuário, pode ser realizada a inserção de uma camada de visualização de glifos à camada de visualização do Treemap, totalmente configurável pelo usuário. Adicionalmente, o usuário pode escolher a utilização do algoritmo de glifo adaptativo, que dependendo da área dos itens do Treemap, poderá ter camadas não renderizadas.

Figura 47. Fluxo de processo do protótipo.



Fonte: Autor do trabalho.

6 TESTES DE USUÁRIO

Este capítulo apresenta os testes de usuário baseados em tarefas para investigar a eficiência (teste do tempo de conclusão da tarefa) e eficácia (teste de acurácia) de glifos em camadas e glifos adaptativos incorporados aos itens do Treemap. Os procedimentos éticos da pesquisa e das avaliações envolvendo usuários foram aprovados pelo Comitê de Ética em Pesquisa do Instituto de Ciências da Saúde da Universidade Federal do Pará (CAAE: 15522319.2.0000.0018 sob o parecer no: 3.618.227 (ver Apêndice A).

6.1 Teste de usuário - glifo em camadas

O objetivo do teste de usuário dos glifos em camadas foi avaliar se a adição de glifos nos itens do Treemap torna mais eficaz (análise da acurácia) e eficiente (análise do tempo) a realização de tarefas de busca e análise pelo usuário. Para tal, foi proposto uma comparação entre os cenários do Treemap com e sem glifos para executar as tarefas. Além disso, outro objetivo do teste é investigar se os glifos em camadas podem ser usados para substituir as hierarquias do Treemap, a fim de minimizar a sobrecarga visual ao usar muitas hierarquias. Finalmente, procurou-se investigar se o glifo em questão é adequado para a representação do valor zero (0), valores pequenos e valores negativos.

6.1.1 Suporte computacional

A sala para os testes possuía uma mesa para o computador, duas cadeiras, um monitor de 21", um notebook, boa iluminação e temperatura agradável. O ambiente computacional utilizado para os testes possuía a seguinte configuração: processador Intel Core i7, 8 GB de RAM, monitor de 21 polegadas com resolução de 1920 x 1080 px. A tela era regulada para estar entre 60 cm e 70 cm dos olhos dos usuários, bem como o centro da tela na direção horizontal natural do olhar do usuário. Durante o treinamento e os testes, apenas um condutor e um participante estavam presentes na sala.

6.1.2 Perfil dos participantes

Participaram do teste 36 voluntários com conhecimento em técnicas do InfoVis, com idades entre 18 e 42 anos, entre homens (33) e mulheres (3), graduados (31) e graduandos (5) em tecnologia da informação. Para esta pesquisa, definiu-se que a experiência em InfoVis deveria ser caracterizada minimamente por cursar e ter sido aprovado em uma disciplina de InfoVis, seja na graduação ou em nível de pós-graduação, ou estar atuando

como pesquisador em visualização a pelo menos seis meses e que conheça o conceito de técnicas hierárquicas.

6.1.3 Descrição da base de dados

A base de dados utilizada era sintética, possuía 200 registros e 13 dimensões, sendo 8 categóricas e 5 quantitativas (ver Quadro 2). Buscou-se utilizar uma base de dados sintética com o intuito de facilitar o controle e as características dos dados para as tarefas do teste.

Quadro 2. Dimensões da base de dados sintética sobre dados meteorológicos. Base de dados com 200 registros.

Dimensão	**Tipo de dado**	**Dados**
EstacaoAno	Categórico	primavera, verão, outono, inverno
Regiao	Categórico	norte, nordeste, centro-oeste, sudeste, sul
PhChuva	Quantitativo	0 a 14
MediaUmidadeAr_%	Quantitativo	1 a 70
Temperatura_C	Quantitativo	-15 a 35
Temperatura_CAT	Categórico	baixa, média e alta
Localidade	Categórico	rural, urbana
VolumeChuva_mm/h	Quantitativo	1 a 30
IntensidadeVento_km/h	Quantitativo	1 a 120
IntensidadeVento_CAT	Categórico	moderado, massivo, furacão, forte, muito forte
AlertaDefesaCivil	Categórico	ativou, não ativou
PhChuva_Cat	Categórico	ácida, base, neutra
IntensidadeChuva	Categórico	fraca, moderada, forte,muito forte

Fonte: Autor do trabalho.

6.1.4 Hipóteses

Os principais cenários propostos para o teste têm o Treemap como base e consideram o uso ou não de hierarquias e glifos em camadas, ou incluem o uso combinado deles. Desta forma, pretende-se avaliar as seguintes hipóteses:

- Hipótese 1 (H1): substituir hierarquias do Treemap por glifos em camadas torna mais eficaz a realização de tarefas de busca.
- Hipótese 2 (H2): combinar hierarquias do Treemap com glifos em camadas torna mais eficaz a realização de tarefas de busca quando comparado ao Treemap apenas com hierarquias.
- Hipótese 3 (H3): combinar hierarquias do Treemap com profile glifos torna mais eficaz a identificação de valores negativos, quando comparado ao Treemap apenas com hierarquias.

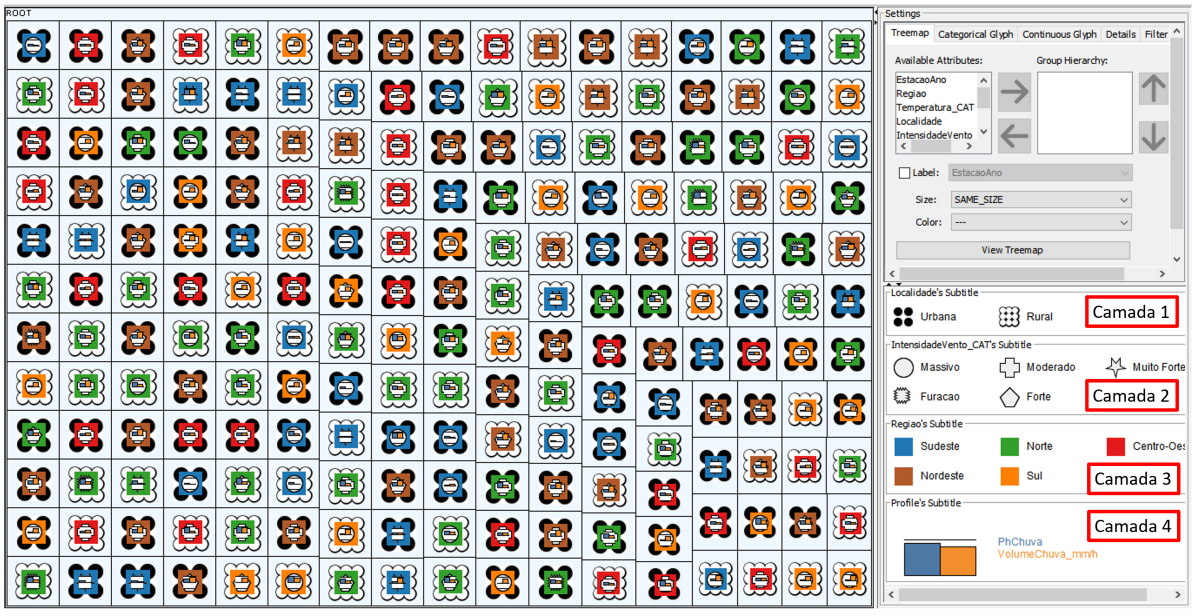
- Hipótese 4 (H4): combinar hierarquias do Treemap com profile glifos torna mais eficaz a realização de tarefas de busca ao identificar o valor zero, quando comparado ao Treemap apenas com hierarquias.
- Hipótese 5 (H5): combinar hierarquias do Treemap com glifos em camadas torna mais eficaz a realização de tarefas ao pesquisar um único valor, quando comparado ao Treemap apenas com hierarquias.
- Hipótese 6 (H6): combinar hierarquias do Treemap com profile glifos torna mais eficaz a realização de tarefas ao identificar padrões em um grupo de dados, quando comparado ao Treemap apenas com hierarquias.

6.1.5 Cenários do teste

Foram definidos três cenários de teste para avaliar a eficácia do uso de glifos no Treemap, de acordo com as hipóteses da subseção 6.1.4:

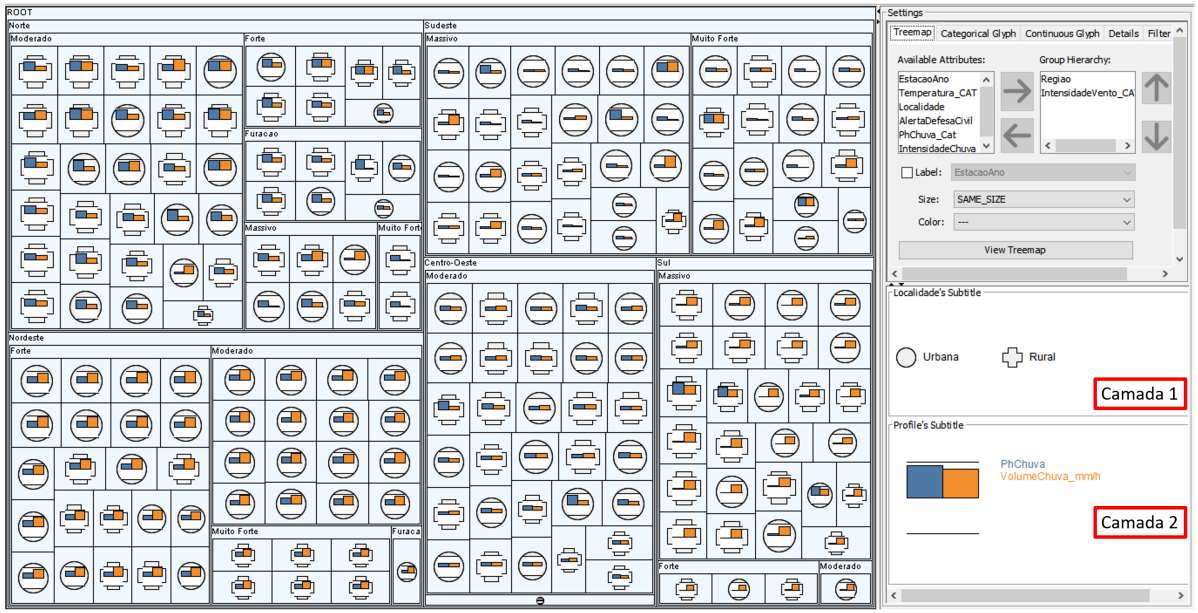
- Cenário **G**: Treemap com **glifos** e sem hierarquia (ver Figura 48);
- Cenário **GH**: Treemap com **glifos** e com **hierarquia** (ver Figura 49);
- Cenário **H**: Treemap sem glifos e com **hierarquia** (ver Figura 50).

Figura 48. Visualização do cenário G: cenário configurado para uma das tarefas de usuário. Essa visualização não usa hierarquia do Treemap, a ausência de hierarquia é substituída por glifos categóricos.



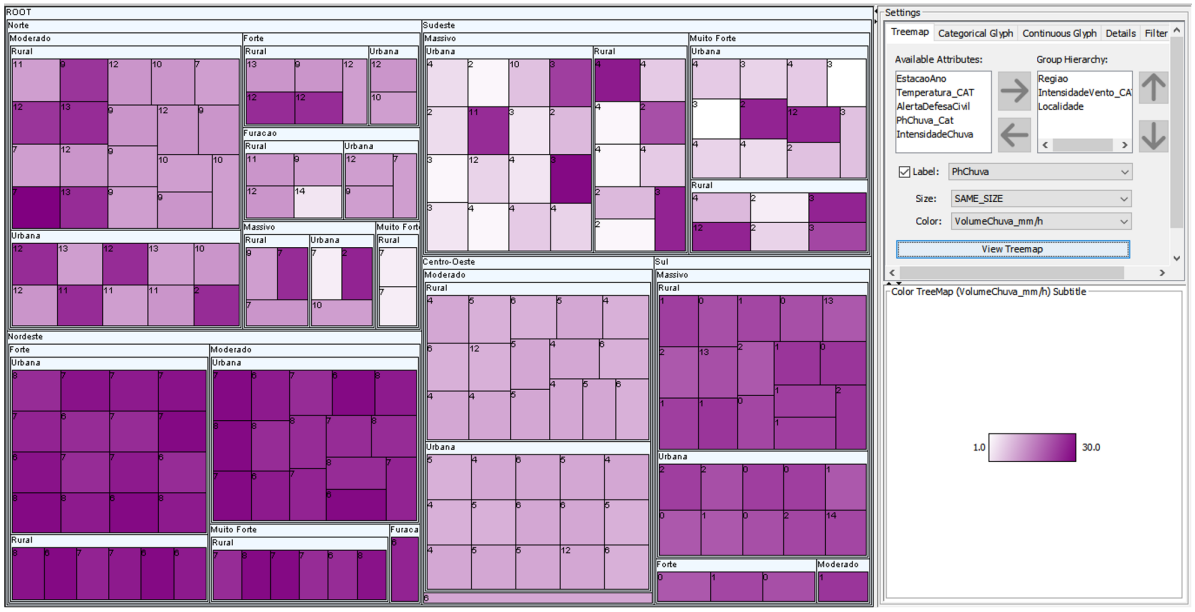
Fonte: Autor do trabalho.

Figura 49. Visualização do cenário GH: cenário configurado para uma das tarefas do usuário. Esta visualização é um Treemap com hierarquia e glifos.



Fonte: Autor do trabalho.

Figura 50. Visualização do cenário H: cenário configurado para uma das tarefas do usuário. Esta visualização é um squarified Treemap sem glifos.



Fonte: Autor do trabalho.

O cenário G foi proposto para analisar casos em que os glifos substituem todas as hierarquias do Treemap. O cenário GH avalia o uso de glifos e hierarquias juntos no Treemap. Finalmente, o cenário H é o uso tradicional do squarified Treemap com

hierarquias e sem o uso de glifos. Este último cenário é a linha de base para comparar os resultados com outros cenários.

6.1.6 Tarefas de usuário

Foram elaboradas oito (8) tarefas de usuário para guiar a análise das hipóteses propostas na subseção 6.1.4. O Quadro 3 mostra uma descrição das tarefas, as hipóteses as quais estão relacionadas e o objetivo que cada tarefa pretende avaliar. As configurações dos atributos do teste estão em Apêndice D, Apêndice E e Apêndice F.

Quadro 3. Lista de tarefas utilizadas durante o teste.

Tarefa	Hipótese	O que está sendo avaliado?
T1—Localize a região que ocorreu o registro de chuva base com ventos massivos e que tenham ativado o alerta da defesa civil.	H1 e H5	O uso de glifos para substituir hierarquias.
T2—Na região norte durante o verão, selecione as três maiores intensidades de vento.	H1 e H2	Facilidade de localizar o maior valor.
T3—Na região nordeste,selecione o único registro de temperatura negativa.	H1 e H3	Facilidade de localizar valores negativos.
T4—Selecione o único registro da região sul com ph da chuva igual a Zero (0), com a ocorrência de vento forte, na localidade rural.	H1 e H4	Facilidade de localizar o valor zero (0).
T5—Localize a região em que registrou a ocorrência de apenas um furacão.	H1 e H5	Facilidade de localizar um valor único.
T6—Localize a região em que houve o registro de apenas uma chuva ácida com temperatura baixa.	H1 e H5	Facilidade de localizar um valor pequeno e único.
T7—Identifique o padrão na intensidade do vento na localidade Rural da região centro-oeste.	H1 e H6	Facilidade de identificar padrões.
T8—Durante a primavera, identifique a região que apresentou as menores médias de umidade do ar (%).	H1 e H6	Facilidade de identificar e comparar padrões entre grupos.

Fonte: Autor do trabalho.

Cada tipo de cenário possui mapeamentos visuais pré-configurados para cada tarefa. Os atributos visíveis são consistentes em todos os cenários, ou seja, são os mesmos atributos de dados, usando mapeamento visual diferente, como hierarquia, forma, cor, etc. Para ilustrar o uso dos mesmos atributos de dados em todos os cenários, observe o mapeamento visual para cada um deles no Quadro 4 relacionados à tarefa 4 do Quadro 3 para o Cenário G (Figura 48), Cenário GH (Figura 49) e no Cenário H (Figura 50).

6.1.7 Procedimentos do teste

Esta seção descreve os procedimentos para realizar os testes. Antes de tudo, um conjunto de pessoas recebeu um convite para participar do procedimento de avaliação.Todos os participantes que aceitaram o convite eram voluntários e não receberam nenhuma compensação por isso. Cada participante realizou o teste individualmente. Antes do início dos testes, era entregue aos participantes o Termo de Consentimento Livre e Esclarecido (ver Apêndice B) e o condutor do procedimento informava ao participante que o uso dos

Quadro 4. Distribuição dos atributos nos cenários G, GH e H ao realizar a tarefa 4. A tarefa de usuário utiliza os mesmos atributos, variando apenas no mapeamento visual.

Mapeamento Visual	**G**	**GH**	**H**
Treemap - Tamanho	—	—	—
Treemap - Rótulo	—	—	PhChuva
Treemap - Cor	—	—	VolumeChuva_mm/h
Treemap - Hierarquia	—	Região, IntensidadeVento_CAT	Região, IntensidadeVento_CAT, Localidade
Glifo - Textura	Localidade	—	—
Glyph - Cor	Região	—	—
Glyph - Forma	IntensidadeVento_CAT	Localidade	—
Glyph - Texto	—	—	—
Profile Glifo	PhChuva, VolumeChuva_mm/h	PhChuva, VolumeChuva_mm/h	—
Total de atributos usados	5	5	5

dados coletados durante os testes seria anônimo. Além disso, se desejassem, poderiam interromper o teste a qualquer momento e não haveria um prazo para a execução das tarefas.

Depois disso, o condutor mostrou um vídeo de treinamento. O uso de um vídeo foi importante para padronizar o treinamento dos participantes. O vídeo incluiu os seguintes tópicos:

- uma breve explicação das técnicas Treemap e glifos;
- o design dos glifos propostos;
- uma visão geral do conjunto de dados usado;
- exemplos de tarefas, semelhantes às do teste, mas usando um conjunto de dados sintético diferente;
- informações sobre como os participantes forneceriam as respostas (por exemplo, clicando em um item do Treemap).

Após o vídeo de treinamento, foi apresentado as tarefas preliminares de visualização de informação com um conjunto de dados diferente. Foi ressaltado que os testes foram iniciados somente quando o participante respondeu corretamente a pelo menos duas tarefas na etapa preliminar: uma tarefa contendo dados quantitativos e outra contendo dados quantitativos e categóricos. Além disso, os participantes tiveram que informar o condutor do teste quando eles estavam preparados para realizar o teste.

Quando o usuário inicia o teste, um tipo de cenário (G, GH ou H) é escolhido aleatoriamente, respeitando o máximo de 12 participantes por grupo sem repetições. Assim, todos os 36 participantes formaram três grupos de 12, um grupo para cada cenário de visualização.

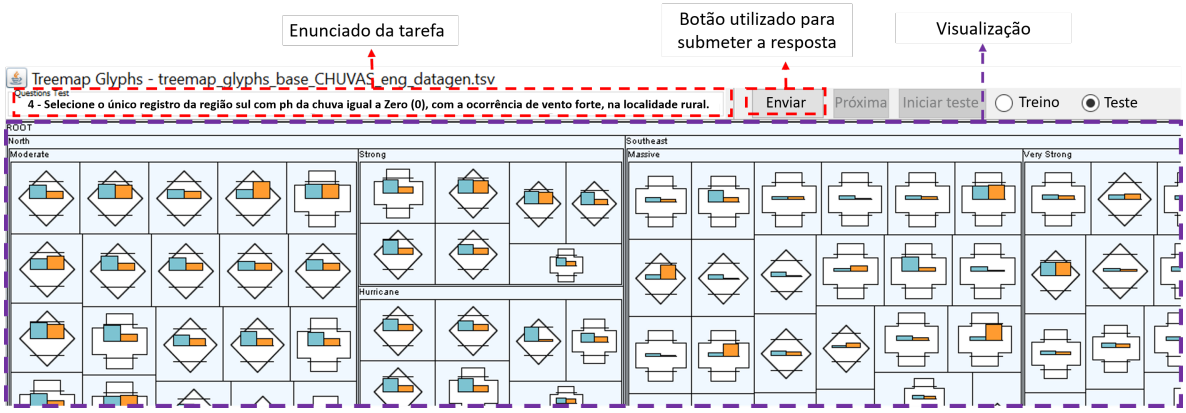
Cada participante executou oito tarefas - de T1 a T8, sorteadas aleatoriamente - usando apenas a visualização do grupo (G, GH ou H) ao qual o participante estava

associado, ou seja, nenhum participante executou tarefas em mais de um tipo de cenário. Foram pré-configurados todos os cenários de visualização para garantir a equivalência entre eles. Assim, os participantes não precisaram configurar os cenários de visualização, para que pudessem se concentrar em analisar a visualização e encontrar a resposta.

Como a única diferença entre os cenários (isto é, a variável independente) é a presença ou ausência de glifos ou hierarquias, o efeito pode ser medido. No estudo, foi coletado dados sobre tempo, erro e interação, e esse processo é automático na ferramenta. O participante precisa apenas clicar nos itens do Treemap para responder a uma tarefa, para que todos os dados sejam registrados. Nenhuma informação é coletada da fase de treinamento, pois seu objetivo é tentar garantir que os participantes estejam no mesmo nível de conhecimento.

Foi mostrado as tarefas na área superior da tela do usuário, imediatamente acima da visualização (veja a Figura 51).

Figura 51. Exemplo de tarefa exibida para o participante, o comando da tarefa e o botão de envio está acima da visualização.



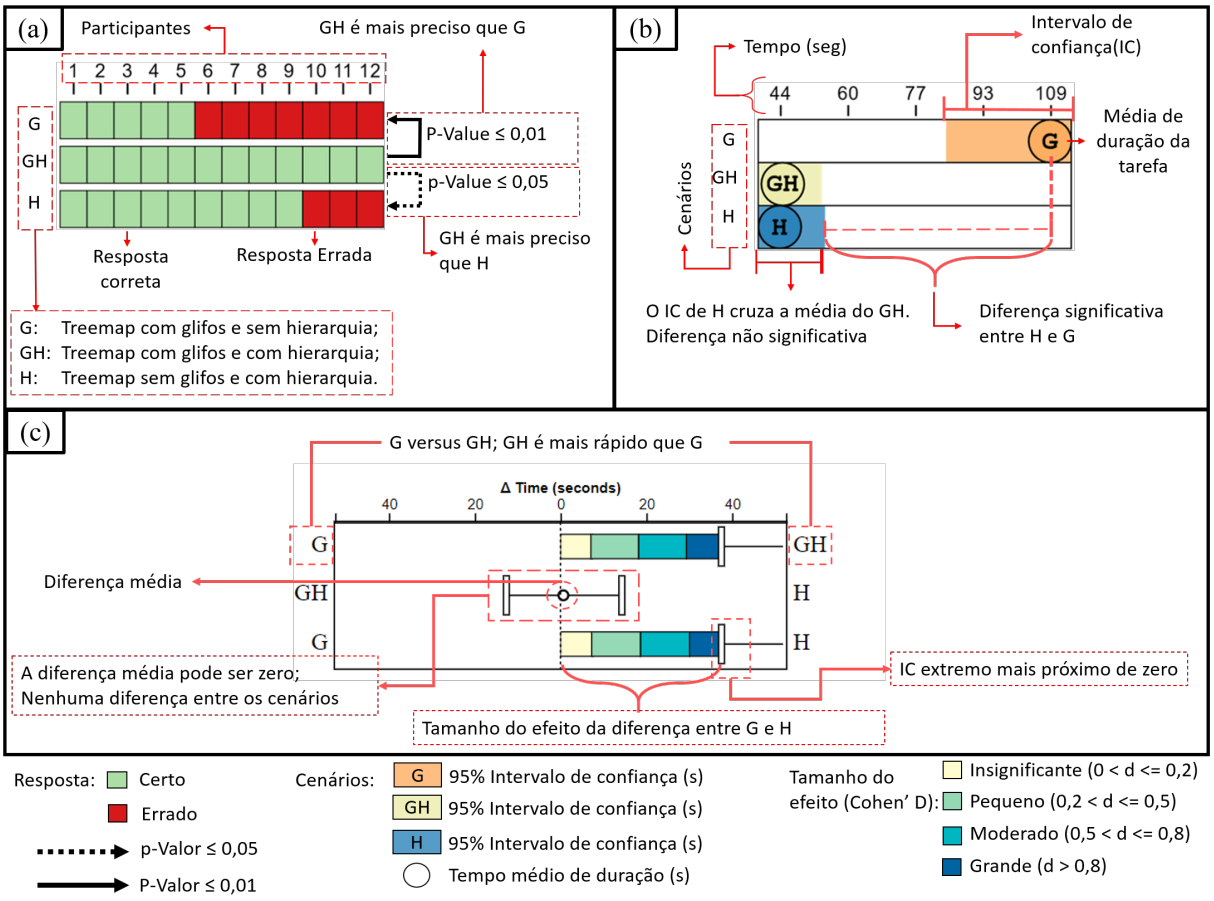
Fonte: Autor do trabalho.

Para o teste, os participantes realizaram oito tarefas em ordem aleatória. Eles só poderiam passar para a próxima tarefa depois de responder à atual. Um participante inicia uma tarefa clicando no botão Iniciar e a lê em voz alta; assim, o participante fica livre para entender a solicitação da tarefa e começar a analisar a visualização. Depois disso, o participante fornece uma resposta clicando nos itens do Treemap e depois no botão Enviar. O protótipo armazena o tempo entre os botões iniciar e enviar e também a resposta do participante em um arquivo CSV. Após repetir esse processo para todas as oito tarefas, a sessão de teste terminou.

6.1.8 Resultado do teste

Três aspectos dos dados coletados foram analisados para comparar o desempenho entre os grupos: a proporção de respostas corretas e erradas, o tempo médio de resposta com intervalos de confiança e uma análise de tamanho de efeito de tempo em pares (KELLEY; PREACHER, 2012; NAKAGAWA; CUTHILL, 2007). A Figura 52 mostra como interpretar os gráficos que apresentarão esses aspectos nas seções a seguir.

Figura 52. Como interpretar os gráficos para os resultados de uma tarefa nas seções a seguir. (a) - Quantidade de respostas corretas e incorretas e um teste em pares usando bootstrapping; (b) - tempo médio de resposta com intervalos de confiança de 95%; (c): comparação do tempo em termos de comparação usando o bootstrap e o Cohen'd.



Fonte: Autor do trabalho.

A Figura 52(a) mostra a quantidade de respostas corretas e incorretas de 12 participantes para uma tarefa em cada cenário e um teste em pares usando bootstrapping para calcular o intervalo de confiança de 95% da diferença entre cada par de grupos. A Figura 52(b) mostra o tempo médio de resposta de 12 participantes e seu intervalo de confiança de 95%, ambos para a mesma tarefa. Se os intervalos de confiança dos dois grupos estiverem distantes, de modo a não haver intersecção entre eles, haverá uma diferença

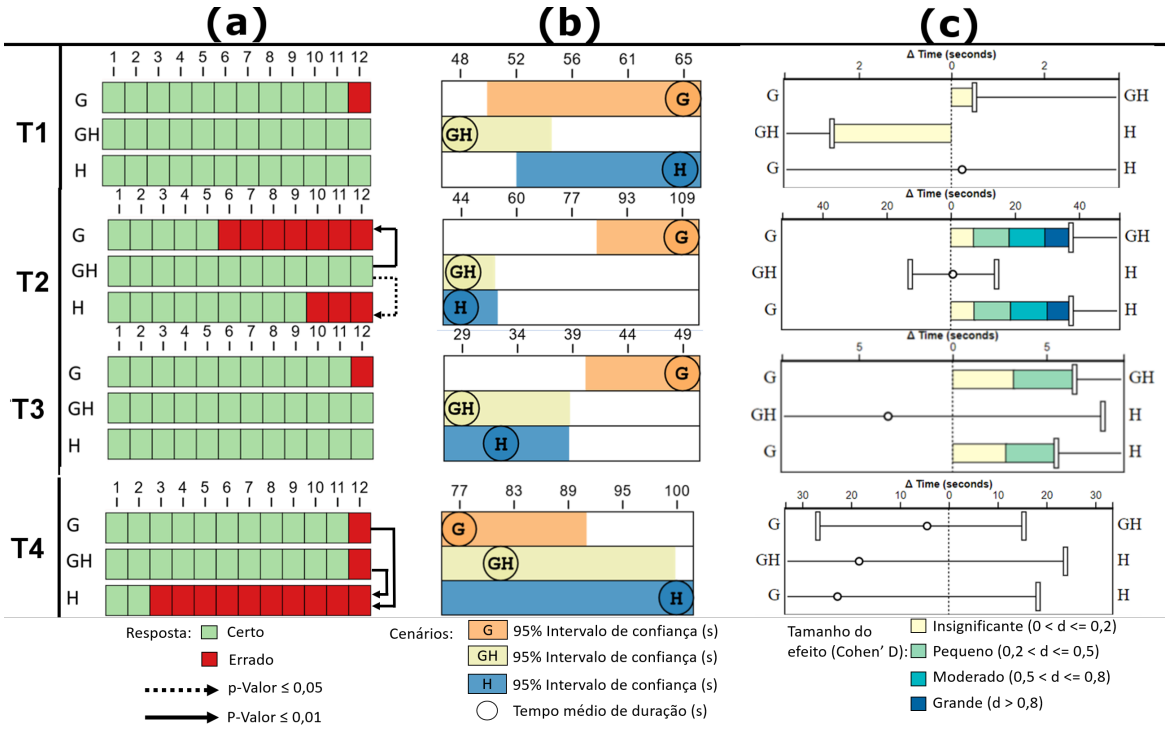
significativa entre eles. Por exemplo, na Figura 52(b), não há diferença significativa entre H-GH, mas há entre GH-G e H-G.

A Figura 52(c) mostra a comparação do tempo em pares entre os grupos. Foi utilizado o *bootstrapping* (BRUCE; BRUCE, 2017) para calcular o intervalo de confiança de 95% da diferença entre cada par de grupos. Se o intervalo ultrapassar zero, não pode-se rejeitar a hipótese (nula) de que os grupos são iguais. Se o intervalo não cruzar o zero, pode-se rejeitar a hipótese nula em favor da hipótese (alternativa) de que um grupo é melhor que o outro (em $\alpha < 0.05$).

Além disso, a Figura 52(c) mostra o tamanho do efeito em termos do Cohen'd (DIENER, 2010) usando uma barra multicolorida que está posicionada no centro e cresce em direção ao cenário mais rápido. As cores indicam os limites dos tamanhos de efeito de Cohen (efeito insignificante, pequeno, moderado e grande). Assim, na Figura 52(c), o cenário GH é pelo menos 38 segundos mais rápido que o cenário G, que é um tamanho de efeito grande (a marca 38 segundos atinge a cor azul escuro).

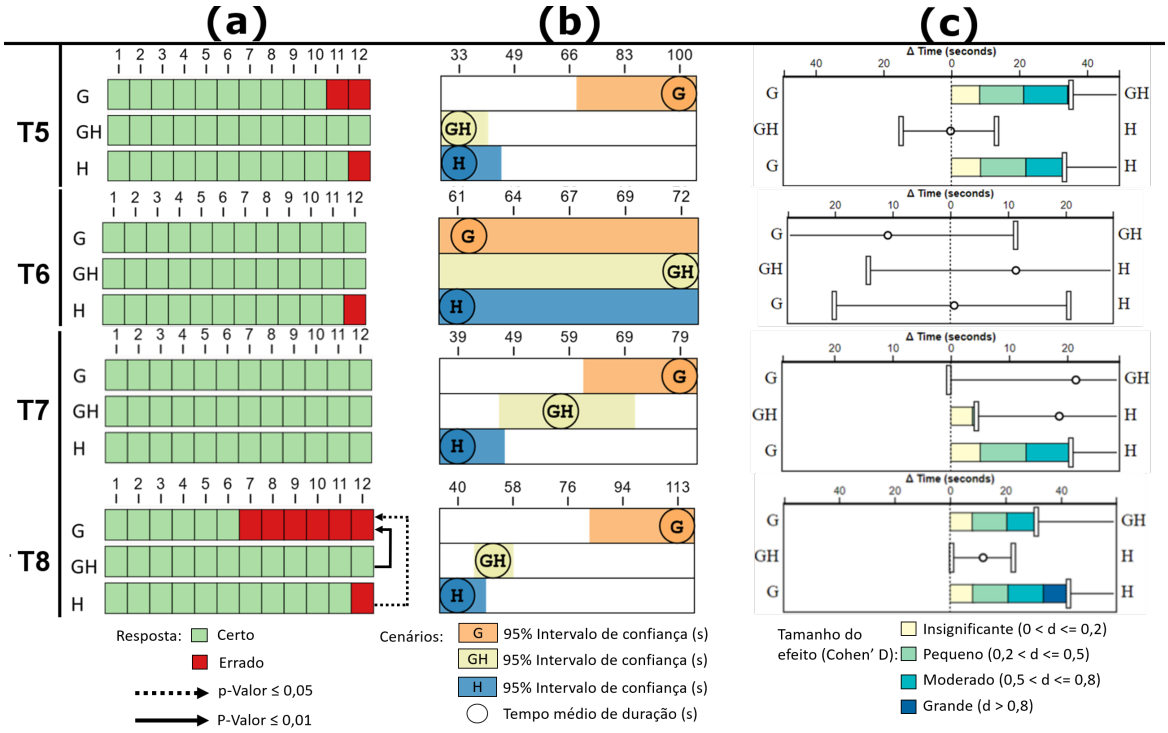
As Figura 53 e Figura 54 apresentam uma visão geral dos resultados de acurácia e tempo das tarefas executadas por todos os grupos de participantes, incluindo a análise estatística em pares entre os grupos.

Figura 53. (a) - Respostas certas / erradas e teste aos pares usando bootstrapping; (b) - Intervalo de confiança no tempo de resposta; e (c) - tempo de teste em termos de comparação para tarefas T1, T2, T3 e T4.



Fonte: Autor do trabalho.

Figura 54. (a) - Respostas certas / erradas e teste aos pares usando bootstrapping; (b) - Intervalo de confiança no tempo de resposta; e (c) - tempo de teste em termos de comparação para tarefas T5, T6, T7 e T8.



Fonte: Autor do trabalho.

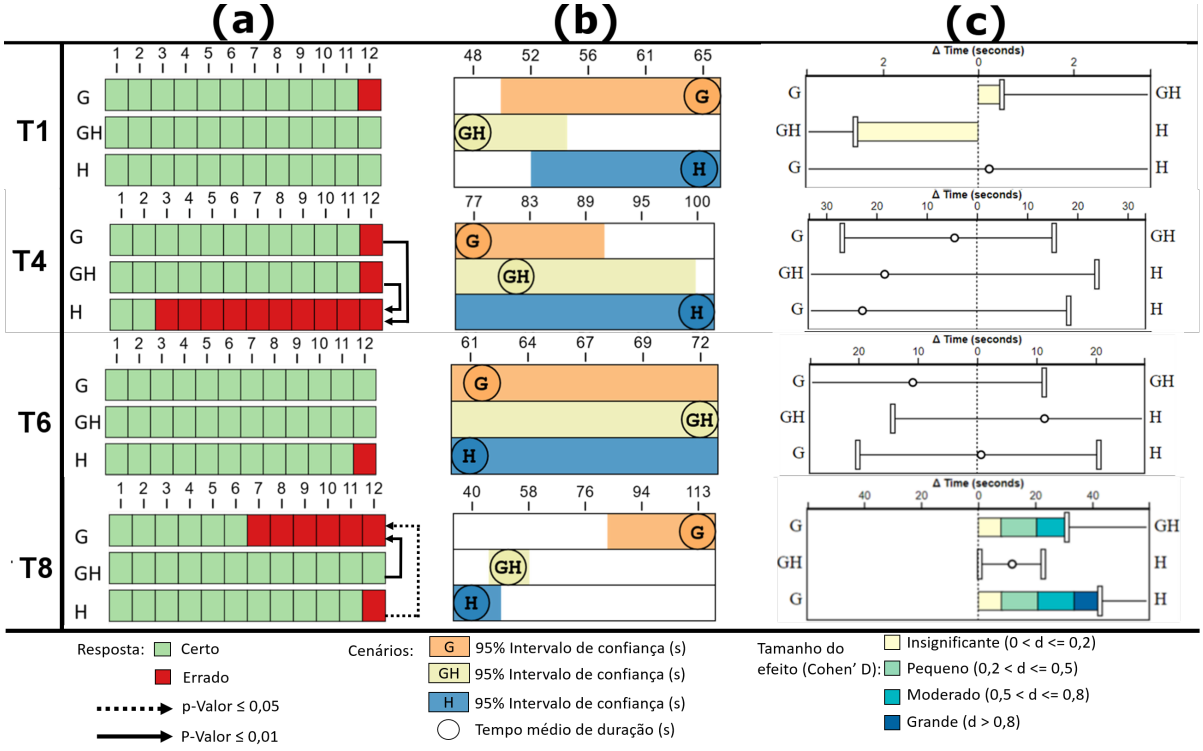
As seguintes seções guiarão a análise dos resultados, a explanação será realizada por hipóteses (ver subseção 6.1.4) e seguida da discussão dos resultados.

6.1.8.1 H1 – Substituir hierarquias por glifos em camadas

A hipótese H1 visa confrontar os cenários G e H. Tal hipótese foi objeto de estudo de todas as tarefas de usuário do teste. Em geral, os cenários G e H não mostraram diferenças estatísticas significativas em relação a análise da acurácia. No entanto, destacam-se as seguintes tarefas:

- T4 (localizar o valor zero) (ver Figura 55(a)) mostrou que o cenário G tem uma acurácia estatísticamente significante com relação ao cenário H (valor de p-valor ≤ 0.01 e valor de p-valor ≤ 0.01, respectivamente);
- T8 (identificar padrões entre grupos) (ver Figura 55(a)) mostrou que H têm acurácia estatísticamente significante com relação ao cenário G (valor de p-valor ≤ 0.01 e valor de p-valor ≤ 0.05, respectivamente).

Figura 55. (a) - Respostas certas / erradas e teste aos pares usando bootstrapping; (b) - Intervalo de confiança no tempo de resposta; e (c) - tempo de teste em termos de comparação para tarefas T1, T4, T6 e T8.



Fonte: Autor do trabalho.

Em uma análise geral, o uso de glifos sem hierarquia (cenário G) reduziu a acurácia quando comparado ao Treemap tradicional (cenário H). Em outras palavras, substituir todas as hierarquias por glifos, em alguns casos, não foi adequado para resolver as tarefas.

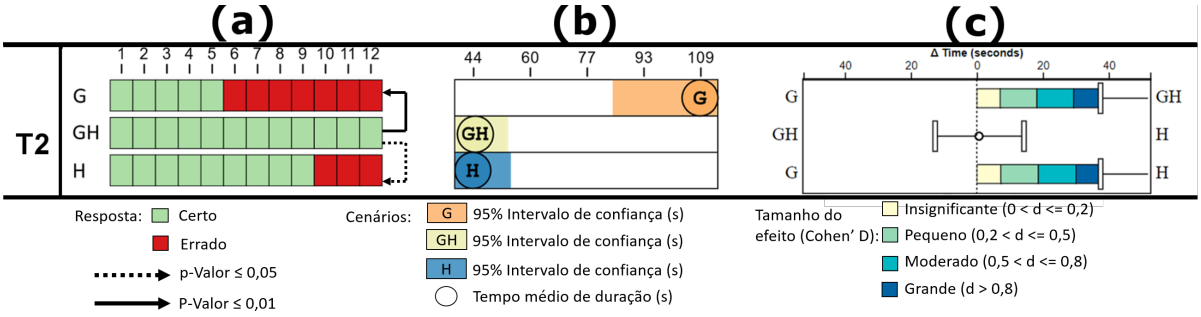
Com relação a análise da variação do tempo, está mostra que não houve diferença estatística a favor do cenário G, ou seja, o cenário H foi estatisticamente significante com relação ao cenário G, e dependendo da tarefa, com o tamanho de efeito (Cohen'D) efeito variando de pequeno ($0.2 < d \leq 0.5$) a amplo ($d > 0.8$). As únicas tarefas que não houveram significancia estatística na variação do tempo entre os cenários em questão foram T1, T4 e T6 (ver Figura 55(b) e (c)).

Assim, considerando a análise da acurácia não é possível confirmar a hipótese H1, pois o cenário G só se mostrou eficaz na tarefa T4, evidenciando que o cenário H não é recomendado para a localização do valor zero (0). Já com relação a variação do tempo das tarefas, os resultados mostraram que houve um tamanho de efeito pequeno ou grande a favor do cenário H.

6.1.8.2 H2 – Combinar hierarquias com glifos em camadas ao pesquisar o maior valor

A hipótese H2 visa confrontar os cenários GH e H. Tal hipótese foi objeto de estudo da tarefa T2, que visava localizar os maiores valores dado um contexto. Com relação a análise da acurácia, pode-se observar na Figura 56(a) que o cenário GH obteve 100% de respostas corretas. Assim, destaca-se que este cenário apresenta uma diferença estatística significante com relação ao cenário H (p-valor ≤ 0.05).

Figura 56. (a) - Respostas certas / erradas e teste aos pares usando bootstrapping; (b) - Intervalo de confiança no tempo de resposta; e (c) - tempo de teste em termos de comparação para a tarefa T2.



Fonte: Autor do trabalho.

Com relação a variação do tempo entre os cenários na tarefa T2, pode-se observar na Figura 56(b) e (c) que não houve significancia estatística entre os cenários GH e H. Desta forma, não pode-se confirmar a hipótese H2 com relação a variação do tempo. No entanto, H2 pode ser confirmada com relação a acurácia da tarefa, mostrando que o cenário GH é mais recomendado, com relação ao cenário H, para tarefas que busquem localizar o maior valor.

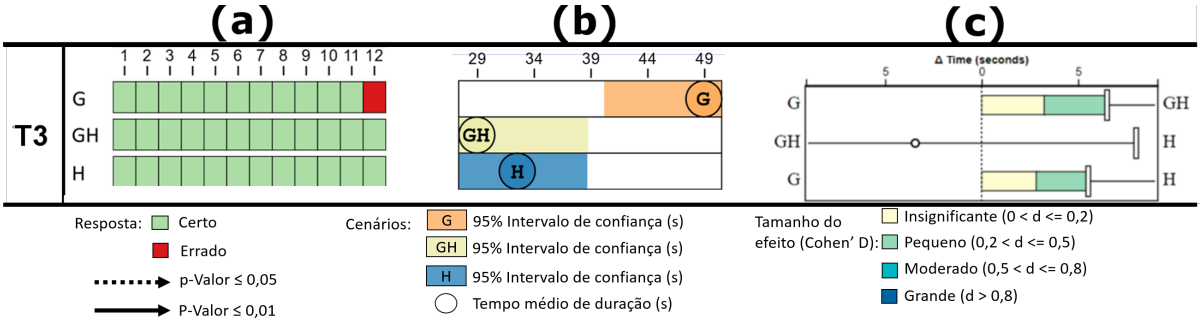
6.1.8.3 H3 – Usar profile glifos na identificação de valores negativos

A hipótese H3 visa confrontar os cenários GH e H. Tal hipótese foi objeto de estudo da tarefa T3, que visava à localização dos valores negativos dado um contexto. A hipótese H3 tinha como objetivo estudar se o uso de profile glifos traria algum ganho na eficácia ao representar visualmente o valores nagativos do que um Treemap sem glifos. É importante destacar que a técnica Treemap não é indicada para a representação de valores nagativos com a variável tamanho. Assim, para a tarefa em questão, no cenário H, um rótulo dentro dos itens do Treemap representou visualmente os valores negativos.

Ao analisar a acurácia da tarefa T3 (ver Figura 57(a)) pode-se observar que ambos os cenários se mostraram eficazes para essa de tarefa, pois tanto o cenário GH quanto o cenário H obtiveram 100% de respostas corretas. No entanto, não obteve-se diferenças

estatísticas quanto a acurácia entre os cenários já citados, ou seja, quanto a acurácia a hipótese H3 não pode ser confirmada.

Figura 57. (a) - Respostas certas / erradas e teste aos pares usando bootstrapping; (b) - Intervalo de confiança no tempo de resposta; e (c) - tempo de teste em termos de comparação para a tarefa T3.



Fonte: Autor do trabalho.

Quanto ao tempo médio de resolução da tarefa, o cenário GH apresentou a melhor média (ver Figura 57(b)), mas sem diferença estatística em relação a H (ver Figura 57(c)). Portando, quanto a variação do tempo, também não foi possível confirmar a hipótese H3.

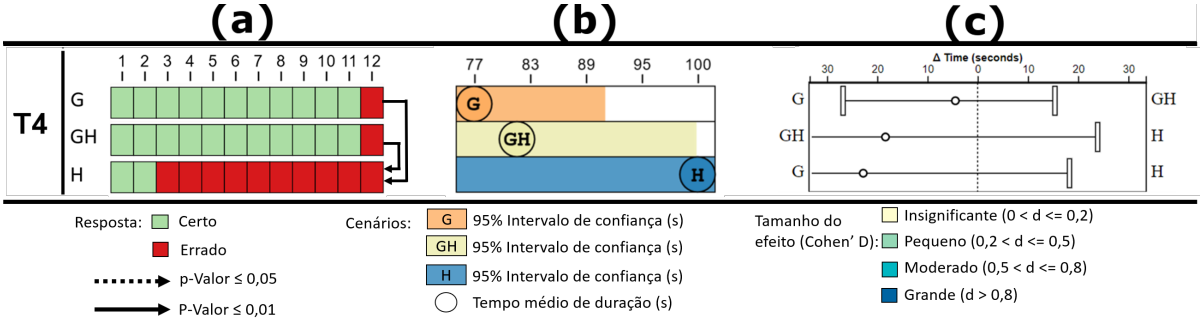
6.1.8.4 H4 – Usar profile glifos na identificação do valor zero

A hipótese H4 visa confrontar os cenários GH e H. Tal hipótese foi objeto de estudo da tarefa T4, que visava a localização do valor zero dado um contexto. A hipótese H4 tinha como objetivo estudar se o uso de profile glifo traria algum ganho na eficácia ao representar visualmente o valor zero (0) do que a representação visual apenas com um Treemap sem glifos. Destaca-se que a técnica Treemap não é indicada para a representação do valor zero (0) com a variável tamanho. Para a tarefa em questão, no cenário H, um rótulo dentro dos itens do Treemap representou visualmente o valor zero (0).

A análise da acurácia da tarefa T4 mostrou que o cenário GH apresentou uma significancia estatística com relação ao cenário H (p-valor ≤ 0.01), o que confirmaria a hipótese H4 quanto a acurácia (ver Figura 58(a)). Porém, ao analisar a variação do tempo para resolver a tarefa T4, pode-se notar que não houve diferenças estatisticamente significativas entre os cenários (ver Figura 58(b) e (c)).

De forma geral, esta tarefa destacou a dificuldade do Treemap em representar o valor zero (0), evidenciando que o uso de profile glifos em Treemaps torna a técnica mais eficaz.

Figura 58. (a) - Respostas certas / erradas e teste aos pares usando bootstrapping; (b) - Intervalo de confiança no tempo de resposta; e (c) - tempo de teste em termos de comparação para a tarefa T4.



Fonte: Autor do trabalho.

6.1.8.5 H5 – Combinar hierarquias com glifos em camadas ao pesquisar um único valor

A hipótese H5 visa confrontar os cenários GH e H. Tal hipótese foi objeto de estudo das tarefas T1, T5 e T6, que visavam a localização de um valor único dado um contexto. A hipótese H5 tinha como objetivo estudar se combinar hierarquias com glifos em camadas ao pesquisar por um valor único na visualização teria algum ganho na eficácia com relação ao uso de um Treemap sem glifos.

Ao analisar a acurácia das tarefas T1, T5 e T6, nota-se que o cenário GH obteve 100% de acertos (ver Figura 59(a)). Entretanto, não foi observado diferença estatística com relação a acurácia entre os cenários (GH e H) nas tarefas analisadas. Com relação a análise da variação do tempo para resolver as tarefas, a tarefa T1 apresentou uma significancia estatística a favor do cenário GH (ver Figura 59(b) e (c)), com um tamanho de efeito insignificante ($0 < d \leq 0.2$). Em outras palavras, o cenário GH é, aproximadamente, no mínimo 2,5 segundos mais rápido que o cenário H. No entanto, ao analisar as tarefas T5 e T6 não foram observadas diferenças estatisticamente significativas entre os cenários (ver Figura 59(b) e (c)).

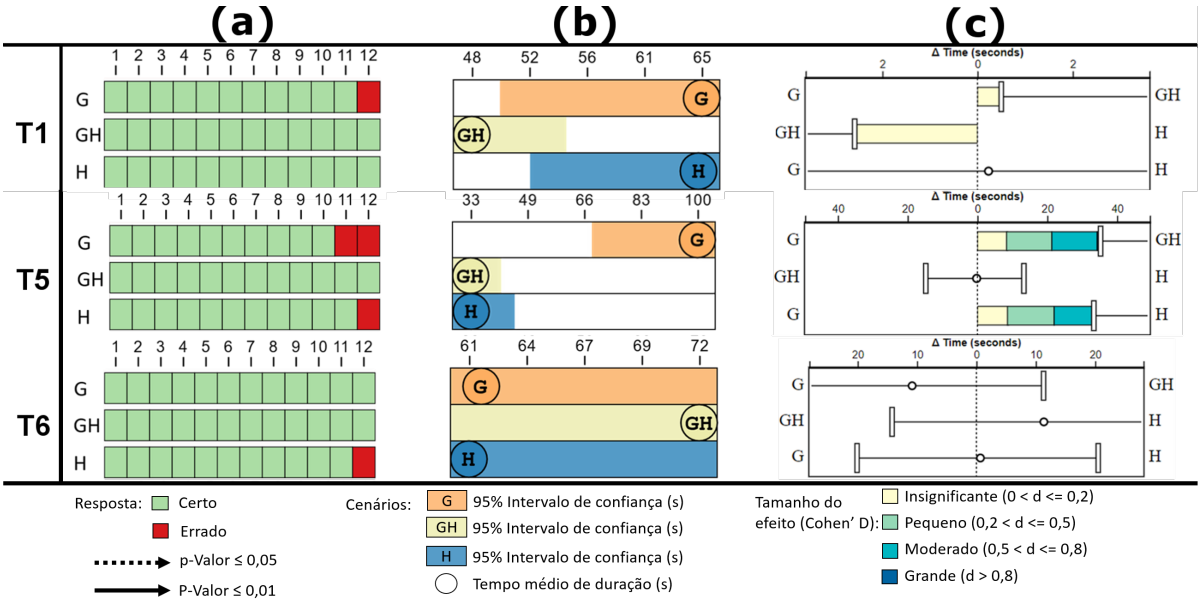
Considerando que não foram observadas diferenças estatísticas significativas entre os cenários GH e H, nem na análise da acurácia e nem na análise da variação do tempo, pois apenas a tarefa T1 apresentou uma diferença estatística a favor do cenário GH, sendo que com um tamanho de efeito insignificante ($0 < d \leq 0.2$), não foi possível confirmar a hipótese H5. A fim de compreender a razão deste fato, faz-se necessário analisar as comparações entre os cenários GH e H com o cenário G.

Na tarefa T1 e T5, ao observar a variação do tempo entre os cenários G e GH, nota-se que houve uma diferença estatística a favor de GH, com tamanho de efeito insignificante ($0 < d \leq 0.2$) e moderado ($0.5 < d \leq 0.8$), respectivamente ver Figura 59(b) e (c).

Destaca-se que na tarefa T6 não houve diferença estatística em nenhuma das comparações de variação de tempo entre os cenários analisados.

Entre os cenários G e H, na tarefa T1 não houve diferença estatística, porém na tarefa T5 houve uma significancia estatística a favor do cenário H, cujo o tamanho do efeito foi moderado ($0.5 < d \leq 0.8$). Tais resultados demostram que o uso de profile glifo em um fasto espaço de busca visual prejudica a eficiência da tarefa, ou seja, o profile glifo deve ser utilziado em conjunto com hierarquias, o que diminui o espaço de busca visual durante a pesquisa de um único valor.

Figura 59. (a) - Respostas certas / erradas e teste aos pares usando bootstrapping; (b) - Intervalo de confiança no tempo de resposta; e (c) - tempo de teste em termos de comparação para a tarefa T5 e T6.



Fonte: Autor do trabalho.

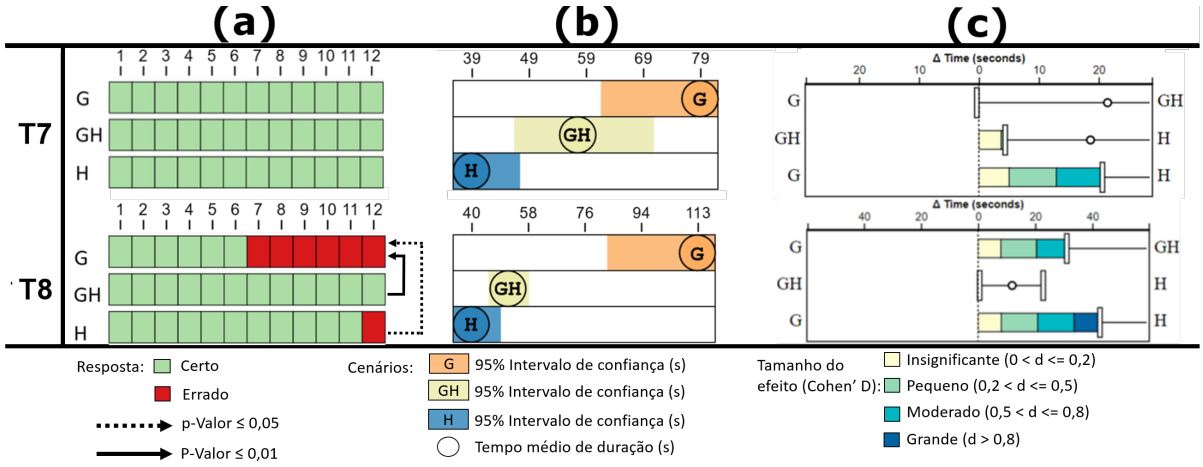
6.1.8.6 H6 – Usar profile glifos na identificação de padrões em um grupo de dados

A hipótese H6 visa confrontar os cenários GH e H. Tal hipótese foi objeto de estudo das tarefas T7 e T8, que visavam a facilidade de identificar padrões. A hipótese H6 tinha como objetivo estudar se o uso de profile glifos na identificação de padrões em um grupo de dados seria mais eficaz do que utilizar apenas um Treemap sem glifos.

Ao análisar a acurácia das tarefas T7 e T8 observa-se que o cenário GH obteve 100% de acertos. Porém, não houve diferença estatística com relação ao cenário H (ver Figura 60(a)). Sobre a variação do tempo entre os cenários, na tarefa T7 o cenário H apresentou uma significancia estatística com relação ao cenário GH, tal diferença possui um tamanho de efeito pequeno ($0.2 < d \leq 0.5$), ou seja, o cenário H é, aproximadamente,

no mínimo 5 segundos mais rápido que o cenário GH. Tal diferença estatística não ocorreu na análise da tarefa T8, que apesar de ter a hipótese H6 como objeto de estudo, focava em estudar a localização de padrões entre grupos.

Figura 60. (a) - Respostas certas / erradas e teste aos pares usando bootstrapping; (b) - Intervalo de confiança no tempo de resposta; e (c) - tempo de teste em termos de comparação para a tarefa T7 e T8.



Fonte: Autor do trabalho.

É importante destacar que as tarefas T7 e T8 utilizam o profile glifo nos cenários G e GH. Desta forma, para que se possa chegar a um entendimento melhor sobre o uso de profile glifos ao identificar padrões, faz-se necessário olhar para a análise de acurácia e de variação do tempo dos cenários GH e H quando comparado ao cenário G.

Na tarefa T7 não houve diferença estatística entre os cenários G, GH e H, a acurácia dos três cenários foi de 100%. Com relação a variação do tempo, apesar de não haver diferença estatística entre os cenários G e GH, houve uma significancia estatística em favor do cenário H com relação ao cenário G, cujo o tamanho do efeito foi moderado ($0.5 < d \leq 0.8$), ou seja, o cenário H é mais rápido que o cenário G, aproximadamente, no mínimo 20 segundos.

Na tarefa T8 os cenários GH e H apresentaram acurácia estatisticamente significate com relação ao cenário G, com valor de p-valor < 0.01 e p-valor < 0.05, respectivamente. Com relação a variação do tempo, os cenários GH e H são mais rápidos que G com efeito de tamanho moderado ($0.5 < d \leq 0.8$) e grande ($d > 0.8$), respectivamente.

Desta forma, o cenário G apresentou a pior variação de tempo para T7 e a pior variação de tempo e precisão para T8. Esses resultados sugerem que o profile glifo exige mais tempo para identificar o alvo de busca se eles forem usados sozinhos, ou seja, quando estão em um *layout* em que o espaço de busca visual é vasto. Do contrário, quando o profile glifo é usado junto com hierarquias, diminuindo o espaço de busca visual, os resultados

são aprimorados e aproximam-se dos resultados do cenário H. No entanto, não é possível confirmar a hipótese H6.

6.1.8.7 Discussão

Considerando os valores das respostas corretas das tarefas e comparado a outros cenários da mesma tarefa, o cenário GH se destaca com a maior acurácia, sempre com o melhor ou compartilhando o melhor número de respostas corretas. Um fato que merece destaque é a ocorrência de uma baixa acurácia no cenário G, especialmente nas tarefas identificação e comparação. Tal fato ocorreu devido o cenário G possuir a característica do não uso de hierarquias no *layout* do Treemap, influenciando nas respostas em função de um maior espaço visual de busca quando comparado aos outros cenários.

O cenário H merece destaque quanto a baixa acurácia na tarefa T4 (ver Figura 61), uma vez que a tarefa visava encontrar um elemento da visualização com o valor zero (0), tal fato ocorreu devido a dificuldade que o Treemap possui ao representar o valor zero (0) com a variável visual tamanho, em outras palavras, é possível representar o valor zero (0) na técnica Treemap, porém a técnica apresenta apenas as opções cor ou rótulo, sendo que a variável visual cor pode se tornar imprecisa para localizar um valor em específico entre dados contínuos (CARPENDALE, 2003) e o label pode se tornar ilegível (BLANCH; LECOLINET, 2007).

Figura 61. Total de erros nos cenários.

	Erros		
Tarefas	**G**	**GH**	**H**
T1	1	0	0
T2	7	0	3
T3	1	0	0
T4	1	1	10
T5	2	0	1
T6	0	0	1
T7	0	0	0
T8	6	0	1
Total	18	1	16

Fonte: Autor do trabalho.

A Figura 61 mostra uma visão geral de todos os erros de tarefas por cenário. É possível notar que o uso de glifos em um Treemap sem hierarquias (cenário G) acaba influenciando na acurácia, em outras palavras, substituir todas as hierarquias por glifos, em alguns casos, não foi adequado para resolver algumas tarefas. Enquanto que a adição de glifos em um Treemap com hierarquias (cenário GH) mostrou que auxilia o usuário a fazer uma leitura mais precisa dos itens visuais sem a sobrecarga de muitas hierarquias e, em alguns tipos de tarefa, mais eficiente do que o uso de um Treemap (cenário H).

Destaca-se ainda que o uso de profile glifo para identificar um único valor ou para identificar um padrão em um grupo de dados, é recomendado sempre em um *layout* que permita a redução do espaço de busca visual. No caso desta tese o espaço de busca foi reduzido com as hierarquias da técnica Treemap. De forma geral, os resultados apontam que o cenário GH é a melhor opção, uma combinação de hierarquias e glifos em camadas.

6.2 Teste de usuário - glifo adaptativo

O objetivo do teste de usuário com os glifos adaptativos é avaliar a sua eficacia na técnica Treemap em tarefas de busca, sendo identificação ou comparação, principalmente quando o alvo de busca estiver em regiões da visualização cujos itens apresentam áreas reduzidas.

6.2.1 Suporte computacional

O teste em questão foi realizada de forma remota. Assim, para a execução do teste, foi solicitado aos participantes voluntários que utilizassem um computador com no mínimo 4GB de RAM e que tivessem a JRE (*Java Runtime Environment*) instalado em sua máquina. As resoluções da telas dos participantes variaram em 1280x800, 1366x768, 1440x900 e 1920x1080, todos no formato 16:9.

6.2.2 Perfil dos participantes

O teste contou com um total de 15 participantes, com idades entre 20 e 45 anos, entre homens (12) e mulheres (3), todos com nível superior. Para este estudo, definiu-se que todos os participantes deveriam conhecer o conceito de técnicas hierárquicas e já ter utilizado a técnica squarified Treemap pelo menos uma vez para análise de dados.

6.2.3 Descrição da base de dados

Para a realização deste teste foi utilizada a mesma estrutura da base sintética da seção 6.1.3, ou seja, foram utilizados os mesmos atributos e tipos de dados. No entanto, para este teste foram utilizados 1000 (mil) registros e foi adicionada mais uma dimensão quantitativa, conforme o Quadro 5.

Considerando que todos os participantes que participaram do teste em questão também participaram do teste anterior (seção 6.1), a adição desta dimensão quantitativa foi realizada com o intuito de evitar que os participantes, tenham aprendido algum padrão ou tendência com a base de dados da seção 6.1.3. Deve-se considerar que o intervalo entre os testes de usuários foi superior a seis (6) meses, os registros sintéticos utilizados entre os

testes eram diferentes, apenas seguiam a mesma estrutura de criação. Desta forma, o teste foi realizado com 14 dimensões, sendo 8 categóricas e 6 quantitativas.

Quadro 5. Dimensão adicionada a base sintética sobre dados meteorológicos. Base de dados com 1000 registros.

Dimensão	Tipo de dado	Dados
NumeroVitimas	Quantitativo	50 a 599

Fonte: Autor do trabalho.

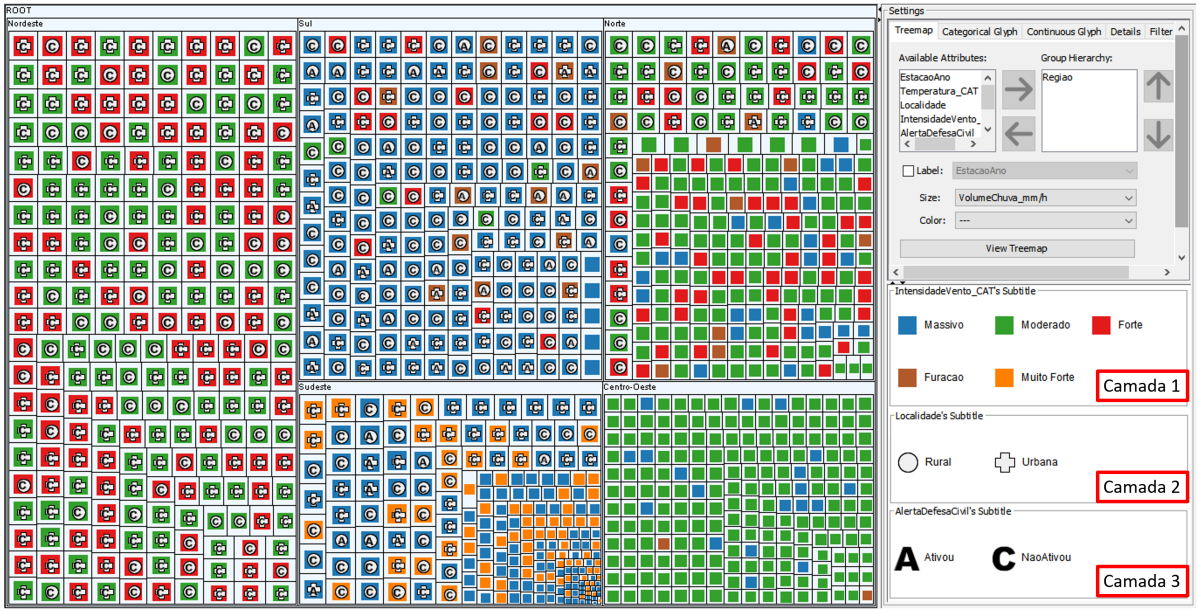
6.2.4 Hipótese

A hipótese deste teste, utiliza como base a combinação de hierarquias do Treemap com glifos em camadas, sendo que os glifos fazem o uso da função adaptativa. Desta forma, pretende-se avaliar a seguinte hipótese: o cenário GA melhora a qualidade visual geral da visualização e leitura das camadas do glifo na técnica Treemap.

6.2.5 Cenário do teste

A combinação entre hierarquias do Treemap com **glifos** usando a função **adaptativa** (Cenário **GA** - ver Figura 62), foi definida como único um cenário de teste para avaliar se os glifos adaptativos conseguiam auxiliar os usuários em tarefas de localização, conforme as hipóteses da seção 6.2.4.

Figura 62. Visualização do cenário GA: cenário configurado para uma das tarefas do usuário. Esta visualização é um squarified Treemap com glifos adaptativos.



Fonte: Autor do trabalho.

6.2.6 Tarefas de usuário

Foram elaboradas um conjunto de cinco (5) tarefas de usuário. O Quadro 6 ilustra a descrição das tarefas e o objetivo que cada tarefa pretende avaliar, a configuração dos atributos do teste estão no Apêndice G.

Quadro 6. Lista de tarefas utilizadas durante o teste.

Tarefa	**O que está sendo avaliado?**
T1—Localize uma das regiões que o alerta da defesa civil foi ativado em registros de furacão.	Capacidade de localizar valores que tiveram 1 camada removida.
T2—Localize o único registro que apresentou intensidade da chuva muito forte, temperatura mediana, e ativou o alerta da defesa civil.	Capacidade de localizar valores que tiveram 2 camadas removidas.
T3—Localize o único registro de furacão que não ativou o alerta da defesa civil, com temperatura mediana na localidade rural.	Capacidade de localizar valores que tiveram 3 camadas removidas.
T4—Na região sul, identifique o único registro que apresenta intensidade do vento e da chuva moderado, com temperatura alta, na localidade urbana, e que não ativou o alerta da defesa civil.	Capacidade de localizar valores que tiveram 4 camadas removidas.
T5—Localize a região que registrou apenas ventos Massivos e Muito Fortes, somente temperaturas Mediana, e não registrou chuvas de PH neutro.	Capacidade de localizar valores que tiveram 5 camadas removidas.

Fonte: Autor do trabalho.

6.2.7 Procedimentos do teste

Um conjunto de pessoas recebeu um convite, via e-mail, para participar do procedimento de avaliação. Estes eram informados que o teste seria dividido em duas partes, uma avaliação quantitativa e uma qualitativa, e que o teste seria realizado remotamente[1]. Os participantes também eram informados que caso aceitassem o convite, seriam voluntários a pesquisa e não receberiam nenhuma recompensa. Após os participantes responderem o e-mail aceitando o convite, eram enviados os seguintes itens para que o participante pudesse realizar o teste:

- Vídeo de treinamento;
- Termo de Consentimento Livre e Esclarecido - TCLE (ver Apêndice C)
- Manual de instruções do protótipo;
- Protótipo de apoio;
- Base de dados.

[1] É importante destacar que o teste foi realizado remotamente em virtude da pandemia do novo Coronavírus SARS-CoV-2, que causa a doença COVID-19

No corpo do e-mail que enviou os itens para a realização do teste, era recomendado que os participantes assistissem ao vídeo treinamento, e em seguida lessem o TCLE. Caso aceitassem prosseguir com o teste, os participantes deveriam ler o manual de instruções da ferramenta, executar o protótipo, e ao final do teste, enviar um email com TCLE em anexo com seguinte conteúdo "concordo com TCLE em anexo", e o arquivo CSV, que continham as respostas das tarefas, gerado pelo protótipo.

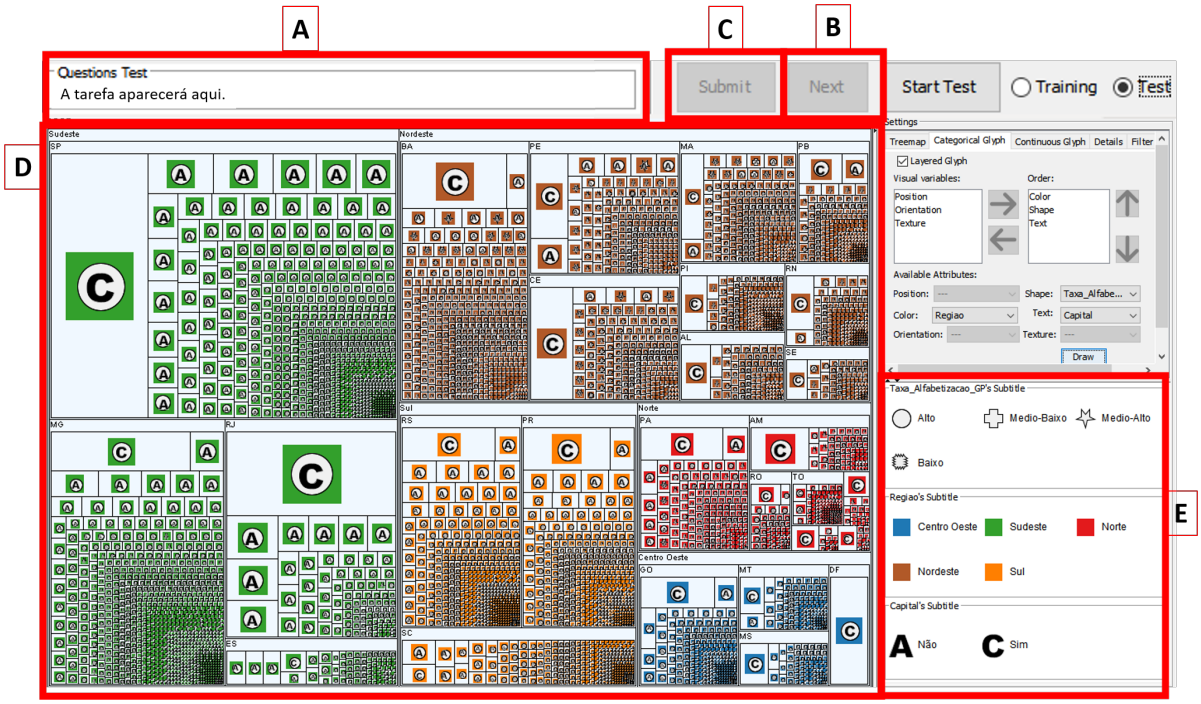
O vídeo de treinamento fazia um breve apresentação sobre a técnica squarified Treemap e sobre glifos. Em seguida, era apresentado o design do glifo em camadas e explicava-se o que era a funcionalidade adaptativa do glifo. Também era dado uma visão geral sobre a base de dados que seria utilizada, apresentando os atributos e tipos de dados. O vídeo fazia uma breve apresentação do protótipo, dando foco apenas nas regiões que o usuário iria utilizar durante o teste, tais como um simples clique nas hierarquias ou nos itens do Treemap para selecioná-los.

O vídeo buscava deixar claro para os participantes que não seria necessário configurar as técnicas, pois o protótipo já iria realizar, automaticamente, o mapeamento visual dos atributos da base de dados as variáveis visuais do Treemap e do glifo. Antes da conclusão, o vídeo de treinamento solicitava aos participantes que realizassem o teste sem interrupções, e mostrava exemplos de tarefas semelhantes às que seriam realizadas durante o teste, e mostrava como responder as tarefas do teste, por exemplo, clicando em uma hierarquia do Treemap.

Considerando que o participante leu e aceitou os termos do TCLE, este deveria ler o manual do protótipo que mostrava quais seriam os passos para que o teste pudesse ser iniciado. A Figura 63 ilustra a tela, apresentada no manual, de como seria o ambiente do teste, a fim de instruir os participantes onde seriam apresentadas as tarefas, onde o Treemap, os glifos e as respectivas legendas apareceriam.

Ao inicializar o protótipo, era exibida uma tarefa de treinamento para o participante, para que este pudesse conhecer e se ambientalizar com o teste. É importante destacar que os dados da tarefa de treinamento não foram contabilizados no resultado deste teste. Durante a execução do teste, os participantes realizaram, em ordem aleatória, cinco (5) tarefas, sem configurar o mapeamento visual das técnicas, sem aplicar filtro, e sem aplicar *drill-down/up*. Após encontrar a resposta da tarefa, clicando em uma ou mais hierarquias ou itens do Treemap, o participante deveria submeter sua resposta, e ir a próxima tarefa. O protótipo salvava o tempo entre o início (*Start*) e o término de uma tarefa (*Submit*), e a resposta do participante em um arquivo CSV. Após conclusão da tarefa 5, o protótipo conduzia o participante a avaliação qualitativa (seção 6.3).

Figura 63. Tela exibida no manual do protótipo, indicando ao participante onde ler as tarefas, onde ver as legendas e onde estaria o Treemap com glifos.



Fonte: Autor do trabalho.

6.2.8 Resultado do teste

Para a análise dos resultados foram considerados a acurácia e o tempo de duração das tarefas pela quantidade de camadas ocultas em cada tarefa. O Quadro 7 apresenta a relação entre a quantidade de camadas presentes no glifo, a quantidade de camadas ocultas (dos glifos a serem identificados) e o tipo de cada tarefa.

Quadro 7. Relação do número de camadas ocultas no glifo em cada tarefa com o tipo de tarefa.

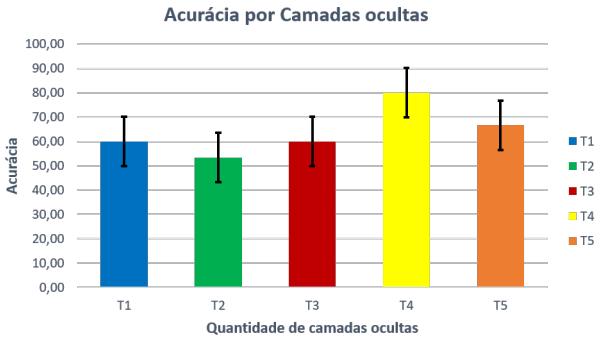
	Camadas		
Tarefas	**Total**	**Ocultas**	**Tipo da tarefa**
T1	2	1	Comparação
T2	3	2	Identificação
T3	4	3	Identificação
T4	5	4	Identificação
T5	6	5	Comparação

Fonte: Autor do trabalho.

Os resultados dos testes mostraram que a média da acurácia entre as tarefas foi de 64%, com um desvio padrão de 10,11 (ver Figura 64). Outro fator a ser destacado é o tempo de resolução das tarefas, que apresentou média de 2,73 minutos, com um desvio padrão de 1,58. Ao analisar a Figura 65 é possível observar que a ocultação maior que 2 camadas gerou uma variação relevante quanto ao tempo de execução, no entanto, cabe

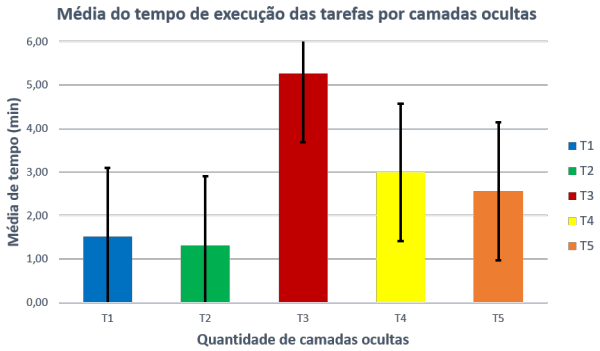
observar que, mesmo os participantes tendo sido instruídos para executarem as tarefas sem interrupções, não houve como o avaliador ter a garantia de que o participante não foi interrompido durante a resolução de uma das tarefas, uma vez que o teste foi realizado remotamente.

Figura 64. Acurácia das tarefas por camadas ocultas.



Fonte: Autor do trabalho.

Figura 65. Tempo de execução das tarefas por camadas ocultas.

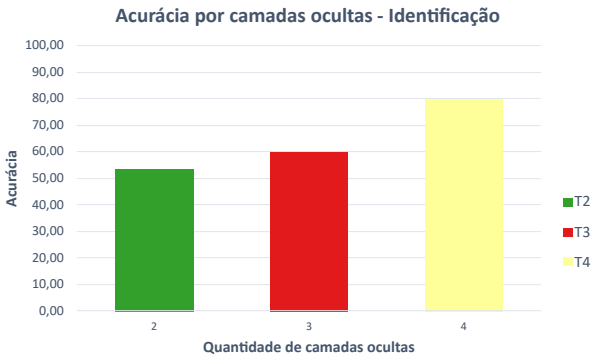


Fonte: Autor do trabalho.

Ao observar a Figura 66 é possível notar que a acurácia entre as questões é crescente, assim como o número de camadas ocultas nos glifos. Tal fato pode ser atribuído a camada que permanecia visível mesmo nos itens de área reduzida, pois nas três tarefas — T2, T3 e T4 (tarefa do tipo de identificação) — a variável visual cor pertencia a camada visível, e por ser a variável visual de maior pré-atenção juntamente com os detalhes sob demanda, auxiliaram os participantes até a resposta das tarefas.

Ao analisar as tarefas de comparação (ver Figura 67), nota-se que os glifos da tarefa 1 (T1) (ver Quadro 7), considerando os que faziam parte da resposta da tarefa, tiveram uma camada ocultada, enquanto que o glifo da tarefa 5 (T5) tiveram cinco camadas ocultadas (ver Quadro 7).

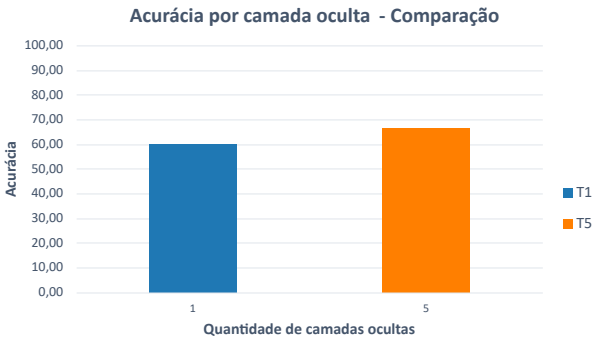
Figura 66. Acurácia das tarefas por camadas ocultas - Tipo de tarefa Identificação.



Fonte: Autor do trabalho.

É importante destacar que, considerando a tarefa 5 (T5), região que continha a resposta da tarefa e após a execução do algoritmo adaptativo, os itens do Treemap de maior área tiveram 4 camadas removidas, ou seja, restaram duas camadas para análise inicial da tarefa, e os de menor área tiveram as 5 camadas removidas. Desta forma, a tarefa mencionava um atributo que estava mapeado a segunda camada, que por sua vez, nos itens de maior área ainda não tinha sido removida. Sendo assim, os itens de maior área que possuíam duas camadas, com o auxilio dos detalhes sob demanda, guiavam os participantes até a resposta certa, o que favoreceu a acurácia da tarefa com relação a T5.

Figura 67. Acurácia das tarefas por camadas ocultas - Tipo de tarefa Comparação.



Fonte: Autor do trabalho.

6.3 Avaliação de Usabilidade e Qualitativa do Glifo Adaptativo

A avaliação qualitativa foi realizada através de um questionário on-line, que foi dividido em duas etapas. A primeira etapa consistia em um questionário com base no método SUS (Escala de Usabilidade do Sistema)(BROOKE et al., 1996; SAURO; LEWIS,

2016), e a segunda coletava os dados sobre a experiência do usuário ao utilizar o glifo adaptativo na resolução de tarefas de InfoVis, e a opinião dos participantes sobre o uso do Treemap com glifos adaptativos, destacando ainda pontos positivos e negativos sobre o uso dos mesmos.

6.3.1 Escala de Usabilidade do Sistema - SUS

Como parte da avaliação dos glifos adaptativos foi utilizada e adaptada a Escala de Usabilidade do Sistema (System Usability Scale - SUS) (BROOKE et al., 1996). O SUS é um método de averiguação do nível de usabilidade de um sistema, ou seja, é utilizada para quantificar o quão usável é um determinado produto. O método SUS auxilia a avaliar critérios como: efetividade, eficiência e satisfação. Para tal, o SUS utiliza dez (10) perguntas seguindo um modelo de questionário, que pode ser adaptado para encaixar melhor no contexto da pesquisa. Este questionário deve ser respondido utilizando a escala *Likert* (1 - Discordo, 5 - Concordo). O Quadro 8 exibe o questionário adaptado com base no método SUS.

Quadro 8. Questionário adaptado com base no método SUS

	Discordo				Concordo	
Tarefas		1	2	3	4	5	
1 - Se eu precisasse usar glifos no Treemap, eu usaria os glifos adaptativos.							
2 - Eu acho que o uso dos glifos adaptativos deixou o Treemap com glifos desnecessariamente complexo.							
3 - Eu acho que o uso dos glifos adaptativos tornou a representação dos dados no Treemap mais fácil.							
4 - Eu precisaria do apoio de um técnico para poder usar o Treemap com glifos adaptativos.							
5 - Eu acho que a qualidade da representação visual dos dados melhorou com o glifo adaptativo.							
6 - Eu achei que haviam muitas inconsistências no glifo adaptativo.							
7 - Eu imagino que a maioria das pessoas aprenderia a usar o glifo adaptativo muito rápido.							
8 - Eu achei o glifo adaptativo muito complicado para o uso.							
9 - Eu me senti muito confiante ao usar o glifo adaptativo.							
10 - Eu precisei aprender muita coisa antes de poder utiliza o glifo adaptativo.							

Fonte: Autor do trabalho.

Para essa avaliação o SUS foi escolhido devido as seguintes características (BROOKE et al., 1996): as escalas são simples de se administrar aos participantes da avaliação; pode ser utilizado em amostras pequenas com resultados confiáveis; e permite diferenciar efetivamente sistemas utilizáveis e inutilizáveis.

Para BROOKE et al.(1996) um sistema pode ser considerado usável se apresentar 68 pontos ou mais na escala do SUS, cujo intervalo é de 0 - 100, ou equivalente a uma nota C. O Quadro 9 exibe a interpretação da escala SUS. De acordo com BROOKE et al.(1996), um resultado com valor menor que 51 pontos, ou nota F o sistema precisará ser remodelado ou refeito. Já resultados no intervalo entre 51 - 68 o sistema precisará de melhorias antes de seu uso. A partir de 68 pontos o sistema requer menos melhorias, até

chegar na percepção de usabilidade excelente ou estupenda, que equivale a faixa de valores entre 85 e 100.

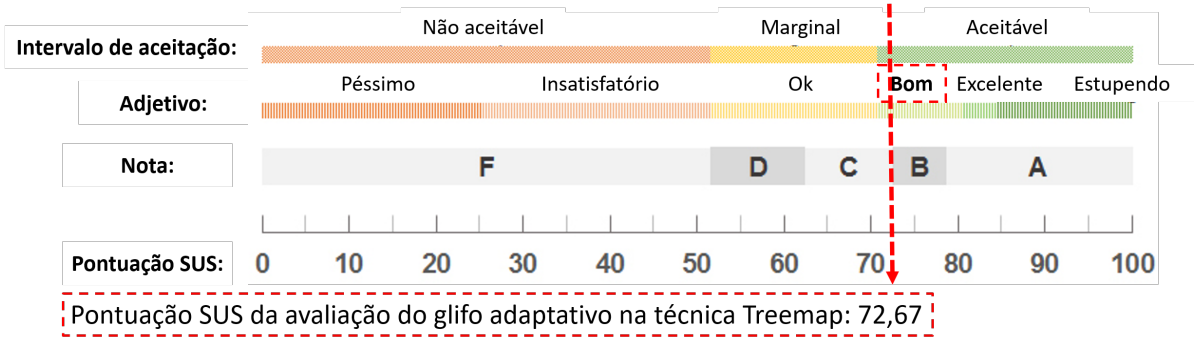
Quadro 9. Interpretação da Escala SUS.

Intervalo SUS	Nota	Porcentagem
84,1 - 100	A+	96 - 100
80,8 - 84,0	A	90 - 95
78,9 - 80,7	A-	85 - 89
77,2 - 78,8	B+	80 - 84
74,1 - 77,1	B	70 - 79
72,6 - 74,0	B-	65 - 69
71,1 - 72,5	C+	60 - 64
65,0 - 71,0	C	41 - 59
62,7 - 64,9	C-	35 - 40
51,7 - 62,6	D	15 - 34
0,0 - 51,6	F	0 - 14

Fonte: Adaptado de SAURO; LEWIS (2016)

BANGOR; KORTUM; MILLER (2009), avaliaram e associaram 1.000 (mil) pontuações (avaliações) do SUS a uma escala de adjetivos, para tornar possível usar palavras em vez de números para descrever a experiência do usuário. A escala contém adjetivos, incluindo "Bom", "OK", "Insatisfatório"e "Excelente". A Figura 68 ilustra a escala mencionada.

Figura 68. Escala de adjetivos SUS. A avaliação do glifo adaptativo obteve a nota 72,67, que equivale ao adjetivo Bom e intervalo de aceitação como Aceitável.

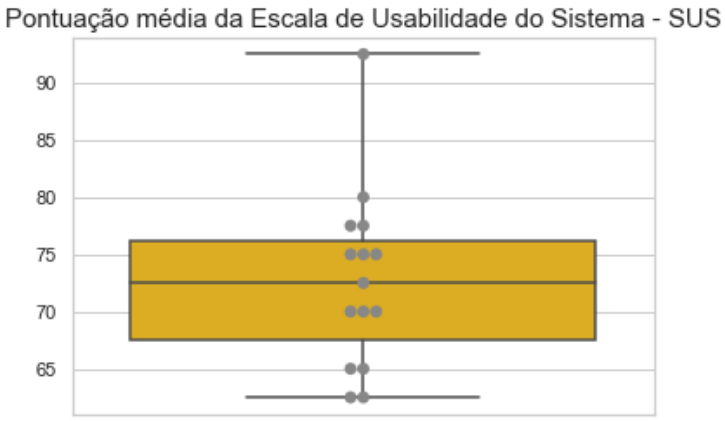


Fonte: Adaptada de BANGOR; KORTUM; MILLER (2009).

As tarefas propostas não permitiam o uso de configurações, filtro ou caminhamento nos grupos da técnica Treemap, apenas os detalhes sob demanda. Deste modo o foco principal da avaliação do usuário foi analisar os dados com a técnica proposta. Aplicando-se o SUS após os participantes realizarem as tarefas propostas, a avaliação do glifo adaptativo na técnica Treemap obteve uma pontuação média de 72,67, conforme Figura 69, que equivale a nota B-. O cálculo realizado para obter esse valor do SUS ocorre da seguinte forma:

- para cada pergunta de índice par, ou seja negativa, é atribuído à resposta um número resultado do seguinte cálculo: valor = 5 - resposta;
- todos os valores encontrados para perguntas par devem ser somados;
- para cada pergunta de índice ímpar, ou seja positiva, é atribuído à resposta um número através do seguinte cálculo: valor = resposta - 1;
- todos os valores encontrados para perguntas ímpar devem ser somados;
- por fim, soma-se os resultados obtidos e multiplica-se pelo fator 2,5.

Figura 69. Pontuação do SUS da avaliação dos glifos adaptativos na técnica Treemap.



Fonte: Autor do trabalho.

A pontuação média do SUS para o glifo adaptativo na técnica Treemap indica que o participante teve uma percepção da usabilidade em um nível Bom, apesar de muito próximo do OK, ou regular, e intervalo de aceitação Aceitável, de acordo com a escala de adjetivos. Desta forma, é possível que sejam realizados testes para verificar se a experiência do usuário com o glifo adaptativo no Treemap pode ser melhorada considerando que as tarefas fossem realizadas com as interações disponíveis no protótipo de apoio, tais como aplicação de *drill-down/up*, configuração do glifo e uso de filtros, pois assim os participantes poderiam eliminar possíveis dúvidas durante a visualização dos glifos, ou simplesmente reduzirem a análise dos dados a cenários GH.

6.3.2 Avaliação qualitativa - glifo adaptativo

A avaliação qualitativa foi realizada em um questionário *on-line* e buscou coletar a opinião e a experiência dos participantes com relação ao uso da técnica Treemap com glifos adaptativos. O Quadro 10 apresenta as questões aplicadas durante a avaliação.

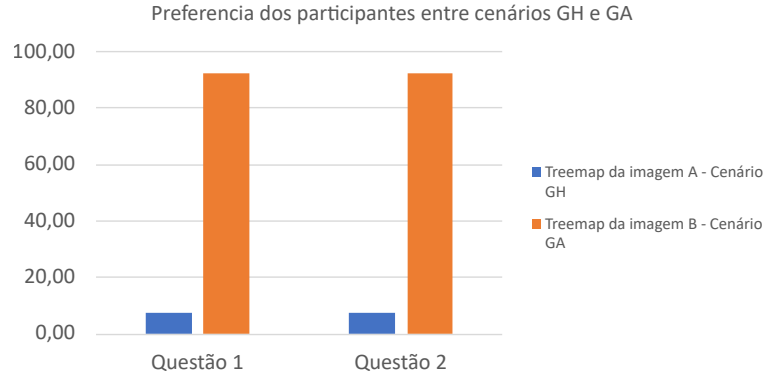
Quadro 10. Lista de perguntas qualitativas.

Questão	Opção/Resposta
1 - As Imagens A e B estão representando o mesmo conjunto de dados referentes ao IDH (índice de Desenvolvimento Humano) do ano 2000 referentes a região Norte e estados utilizando a técnica de visualização Treemap. Considere a área destacada em vermelho: qual dos Treemaps apresentados (imagem A ou imagem B) você consideraria para analisar os referidos dados?	• Figura 73 Imagem A • Figura 73 Imagem B
1.1 - Justifique sua escolha entre as imagens A e B. (Ex: Se você escolheu A, justifique a escolha do Treemap da imagem A)	
1.2 - Informe a razão de você não ter escolhido o Treemap da outra imagem. (Ex: Se você escolheu A, informe porque você NÃO escolheu B).	
2 - As Imagens A e B estão representando o mesmo conjunto de dados referentes ao IDH (índice de Desenvolvimento Humano) do ano 2000 referentes ao estado do Ceará, IDH municipal e taxa de alfabetização, utilizando a técnica de visualização Treemap. Considere a área destacada em preto: qual dos Treemaps apresentados (imagem A ou imagem B) você consideraria para analisar os referidos dados?	• Figura 74 Imagem A • Figura 74 Imagem B
2.1 - Justifique sua escolha entre as imagens A e B. (Ex: Se você escolheu A, justifique a escolha do Treemap da imagem A).	
2.2 - Informe a razão de você não ter escolhido o Treemap da outra imagem. (Ex: Se você escolheu A, informe porque você NÃO escolheu B)	
3 - Durante a realização do teste, quais recursos você mais utilizou para encontrar uma informação que estava faltando?	• Detalhes sob demanda; • Caminhar nas hierarquias; • Legendas; • Caminhar nos itens; • Outros
4 - Cite cenários de utilização para o glifo adaptativo	
5 - Cite cenários de uso onde o glifo adaptativo não deveria ser aplicado	
6 - Destaque pontos positivos para o uso do glifo adaptativo.	
7 - Destaque pontos negativos para o uso do glifo adaptativo	
8 - Você teria sugestões de melhoria para o glifo adaptativo?	

Fonte: Autor do trabalho.

As perguntas 1 e 2 (ver Quadro 10) realizavam uma comparação entre os cenários GH e GA, a fim de saber dos participantes qual cenário, dentre as visualizações do Treemap, seria considerado para a analisar os dados de um referido contexto. A Figura 70 mostra que os participantes optariam pelo Treemap do cenário GA, tanto na questão 1 quanto na questão 2.

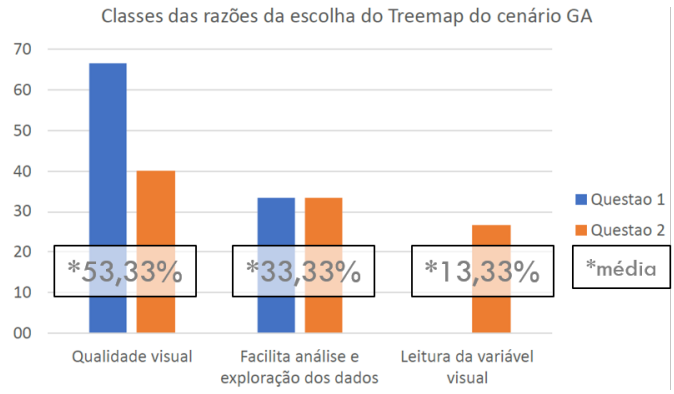
Figura 70. Qualidade da informação visual entre os cenário GH e GA. Os participantes optaram, em sua maioria, pelo GA.



Fonte: Autor do trabalho.

As perguntas que seguiam essas tarefas buscavam obter o motivo da escolha dos participantes. Desta forma, entre as perguntas 1 e 2, uma média de 53,33% dos participantes informaram que optaram pelo cenário GA devido a qualidade visual do glifo, seguido de 33,33% que informaram que o cenário GA facilitou a análise e exploração dos dados, e 13,33% informou que o cenário GA favorecia a leitura das variáveis visuais que permaneciam na visualização (ver Figura 71). Assim, pode-se afirmar que os participantes consideram que o cenário GA melhorou a qualidade visual do Treemap, com relação ao cenário GH, confirmando a hipótese proposta neste teste (subseção 6.2.4).

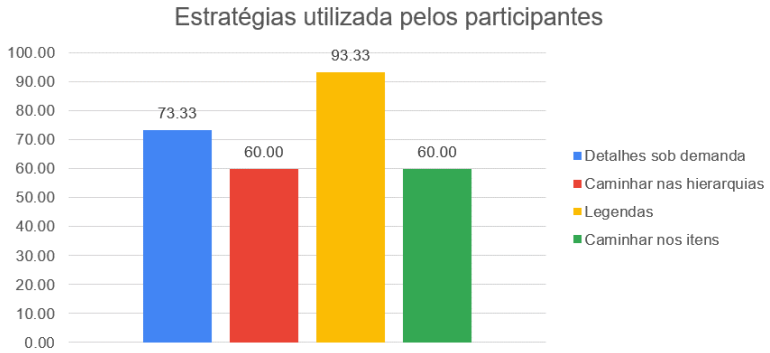
Figura 71. Razões pelas quais os participantes escolheram o cenário GA nas perguntas 1 e 2.



Fonte: Autor do trabalho.

A pergunta 3 (ver Quadro 10) permitiu avaliar a estratégia utilizada pelos participantes para identificar a solução das tarefas de usuário propostas. Foi utilizado como critério de ranking dos recursos a soma das resposta com valor 5 (muito) e 4 (razoável). Os participantes, em um total de 93,33%, destacaram a legenda como principal recurso utilizado na busca pela solução da tarefa (ver Figura 72).

Figura 72. Estratégias utilizadas pelos participantes.



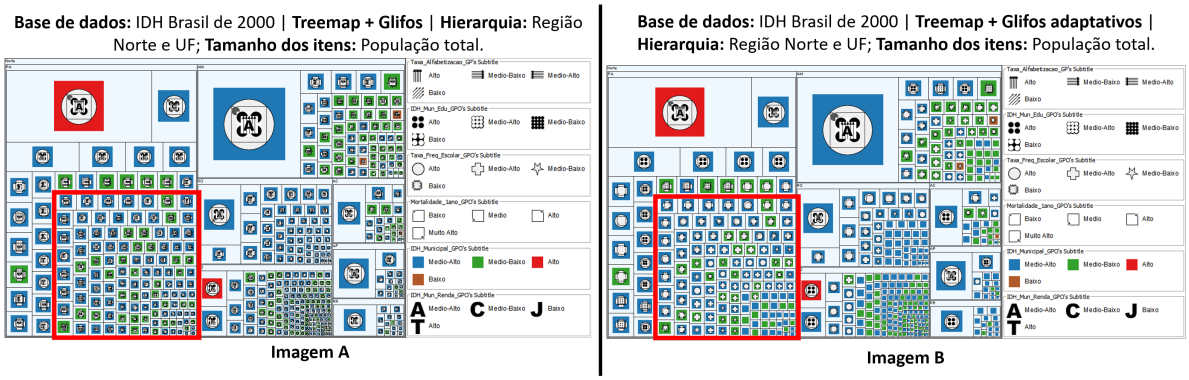
Fonte: Autor do trabalho.

O segundo recurso mais utilizado foi o detalhe sob demanda, com 73,33% de uso. Por fim, empatados com 60% de uso pelos participantes o caminhar(exploração manual) por itens/glifos e hierarquias da visualização. Pode-se inferir, que antes de qualquer busca visual, o participante realizava um filtro visual de que regiões não deveriam ser exploradas, tendo como base as próprias hierárquicas e variáveis visuais visíveis, diminuindo o espaço de busca, e diminuindo o tempo para completar a tarefa. Assim, o glifo adaptativo colaborou com informação útil na busca da solução das tarefas propostas.

Os participantes sugeriram cenários de uso para os glifos adaptativos (perguntas 4 e 5 — Quadro 10) e aspectos de melhoria (perguntas 6, 7 e 8 — Quadro 10). Entre os cenários mais adequados, destacam-se: análise de conjunto de dados com poucas linhas e muitas dimensões, principalmente categóricas, permitindo a correlação entre as dimensões de dados, também destacaram a necessidade de aplicar glifos adaptativos quando a visualização precisa melhorar a qualidade visual dos dados, diminuindo a oclusão de itens e melhorando a leitura das variáveis visuais.

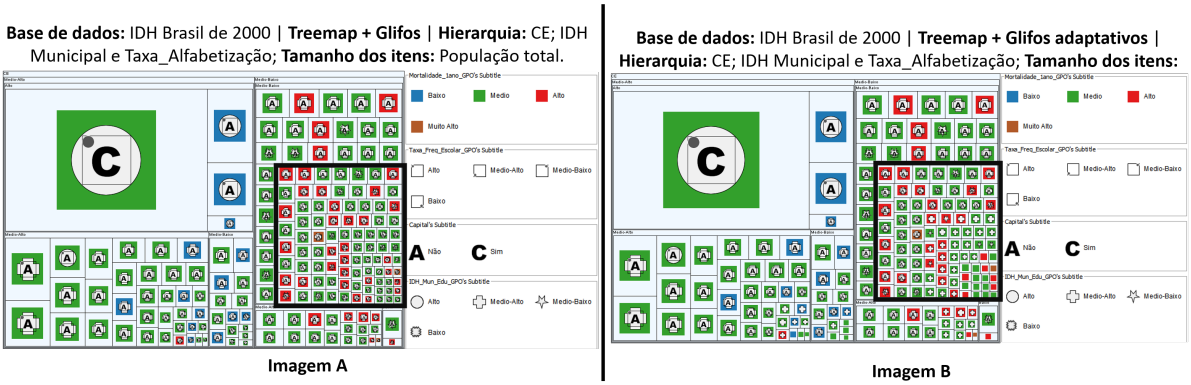
Por outro lado, cenários não favoráveis para uso dos glifos adaptativos foram: quando o conjunto de dados apresenta poucas dimensões de dados, muito atributos numéricos contínuos, muitos valores por dimensão de dado, necessidade de mais variáveis visuais efetivas como a cor. Como sugestões de melhoria, destaca-se: melhorar a qualidade das variáveis visuais, *brushing* entre detalhes sob demanda e legenda para uma rápida leitura do glifo, e sugerir automaticamente a construção do glifo de acordo com os dados selecionados. Outros pontos levantados foram possibilidade de configurar o glifo e utilização de filtros, mas essas demandas já estão implementados no protótipo.

Figura 73. Alternativas da questão 1 da avaliação qualitativa. A - A área demarcada possui glifos com 6 camadas; B - Os glifos da área demarcada possuem uma variação de 3 a 4 camadas removidas.



Fonte: Autor do trabalho.

Figura 74. Alternativas da questão 2 da avaliação qualitativa. A - A área demarcada possui glifos com 4 camadas; B - Os glifos da área demarcada possuem uma variação de 1 a 2 camadas removidas.



Fonte: Autor do trabalho.

7 CONSIDERAÇÕES FINAIS

Esta tese apresentou um novo modelo de design de glifo, chamado de glifo em camadas, e o aplicou no *layout* da técnica Treemap. Os glifos adicionados ao Treemap apresentam base quadrada, representam atributos categóricos e quantitativos em camadas, onde cada camada possui apenas uma variável visual representada, e as camadas são sobrepostas uma sobre as outras com uma sobreposição parcial de 70% de uma camada N-1 por uma camada N.

Para a representação de atributos categóricos cada camada pode ser mapeada para um tipo de variável visual, sendo textura, cor, forma, texto, posição e orientação. Para a representação dos atributos quantitativos, o glifo permite a adição de um profile glifo, adicionado sempre na última camada superior.

Apesar da técnica Treemap ser bastante difundida para análise de dados hierárquicos, essa tese buscou mitigar alguns dos problemas que a técnica apresenta, tais como: representação de números de valor muito pequeno, zero ou negativo, perda de qualidade da visualização ao utilizar muitas hierarquias, e poucas opções de mapeamento visual dos dados. O Quadro 11 apresenta como esta tese mitigou alguns desses problemas.

Quadro 11. Limitações do Treemap e como esta tese buscou mitiga-las.

Limitações do Treemap	**Solução implementada com o glifo**
(I) Poucas opções de mapeamento visual dos dados	Representa até 7 tipos mapeamentos visuais
(II) Perda de qualidade da visualização ao utilizar muitas hierarquias	Representa 6 camadas com variáveis visuais categóricas
(III) Representação de valores negativos	Uso de profile glifo
(IV) Representação do valor zero	Uso de profile glifo
(V) Representação de números de valores muito pequenos	Uso de profile glifo
(VI) Representação visual de dados nos itens de área pequena	Criação de um glifo adaptativo as áreas pequenas

Fonte: Autor do trabalho.

O Treemap apresenta como uma de suas limitações poucas opções para representar dados visualmente, para tal a técnica utiliza apenas cor, rótulo e tamanho. Para mitigar essa limitação (ver Quadro 11(I)), essa tese apresenta um modelo de glifos que permite o mapeamento visual de dados através de sete (7) variáveis visuais. Desfa forma, tem-se como objetivo aumentar a dimensionalidade da técnica Treemap com o uso do modelo de glifo proposto.

A técnica Treemap permite adicionar níveis hiearquicos para melhorar a navegação do layout e, por consequência, aumenta-se a quantidade de dimensões representadas visualmente. No entanto, o aumento dos níveis hierarquicos acaba prejuducando a percepção dos usuários durante a análise de dados. Para mitigar essa limitação (ver Quadro 11(II)), esta tese propôs o uso de glifos em camadas. O modelo de glifo proposto apresenta até 6 camadas que podem representar seis (6) variáveis visuais categóricas. O intuito é permitir

que o usuário possa realizar uma troca entre níveis hierarquicos e as camadas categóricas do glifo.

Outra limitação da técnica Treemap é a dificuldade de representação visual, na variável tamanho, de valores negativos, valor zero (0) e valores pequenos (ver Quadro 11(III),(IV) e (V)). Para mitigar essa limitação esta tese propôs o uso de profile glifos na camada N do glifo em camadas (ver seção 4.2).

Quanto a limitação de representação visual de dados nos itens de área pequena (BLANCH; LECOLINET, 2007) (ver Quadro 11(VI)), esta tese apresentou um modelo de glifo adaptativo que oculta suas camadas a medida que área dos itens vão reduzindo, plotando apenas as camadas que ainda forem visíveis. Desta forma, mesmo nos itens do Treemap de área pequena é possível entregar para o usuário uma informação visual, na qual poderá ser utilizada para uma tomada de decisão.

Para a construção do modelo de glifo em camadas foi realizado um estudo inicial sobre a percepção de variáveis visuais parcialmente ocluídas (pré-teste 1), a fim de avaliar qual nível de oclusão deveria ser aplicado para uma boa leitura das variáveis visuais, quais seriam as variáveis visuais mais robustas a oclusão, e a quantidade de valores distintos que as variáveis visuais poderiam assumir. Ao final do estudo sobre a percepção de variáveis visuais, observou-se a seguinte ordem entre as variáveis: cor-matiz, luminosidade, forma, textura, texto, orientação, e por fim, saturação.

Para a elaboração do design do glifo foram consideradas as variáveis do estudo sobre a percepção das variáveis visuais, considerando o nível de 70% de oclusão com cinco valores distintos para cada variável, pois estas seriam as mais robustas a sobreposição, com exceção da variável saturação. A variável visual saturação foi substituída pela variável posição, uma vez que no ranking de variáveis visuais para dados categóricos de MACKINLAY(1986) esta é a variável de melhor ranking. Adicionalmente, foi adicionado um profile glifo na última camada, para a representação de atributos quantitativos.

Para a validação do estudo, foi desenvolvido um protótipo de apoio que permitia a aplicação dos glifos em camadas no layout da técnica Treemap (disponível em <http://labvis.ufpa.br/treemapglyphs/>). Tal protótipo permite ao usuário visualizar dados na técnica Treemap e dá a possibilidade do usuário adicionar glifos a ela. Os glifos em camadas são inseridos em cada item do Treemap, sendo que cada camada pode mapear uma variável visual, e os valores das variáveis visuais são relacionados aos valores dos atributos da base de dados.

Para mitigar o problema de leitura do glifo em itens de área muito pequena foi proposto o glifo adaptativo. Para tal, foi realizado um teste de usuário para avaliar a percepção sobre as variáveis visuais do glifo (pré-teste 2). Após a análise dos resultados do teste, notou-se a necessídade de um mecanismo inteligente que automatizasse a criação de

regras para deixar na visualização apenas as camadas que seriam visiveis no glifo.

Desta forma, foi gerada uma árvore de decisão que informa quando as variáveis visuais que compõem o glifo ainda são visíveis em função da área disponível para desenho. Uma vez de posse dessa informação, é possível remover uma camada do glifo e submete-lo novamente a análise da árvore de decisão, até o glifo estar visível ou quando não houver mais camadas para a remoção.

Para avaliar o glifo proposto na técnica Treemap, foram realizadas duas avaliações, uma realizando a comparação entre os cenários G (Glifo), GH (Glifo e Hierarquias) e H (Hierarquias), e a outra para avaliar o cenário GA (Glifo Adaptativo). A análise dos resultados da primeira avaliação mostrou que o cenário GH apresentou a melhor acurácia dentre os cenários, independente do tipo de tarefa aplicada.

Do contrário, o cenário G foi o cenário com pior acurácia, especialmente em tarefas cujo o tipo sugeria indicar onde procurar o alvo (item) de busca. Como uma das características do cenário G é não utilizar hierarquias, as tarefas do tipo mencionadas se tornavam mais demoradas ou mais difíceis durante sua resolução, em função de um espaço de busca visual maior.

Cabe destaque o elevado número de erros no cenário H na tarefa 4, que tinha como objetivo a procura de um item que possuía um valor único, o valor zero (0), confirmando assim a dificuldade do Treemap em representar o valor zero (0). De maneira geral, não houve diferença significativa quanto a variação do tempo entre GH e H, e o cenário GH apresentou uma acurácia melhor em relação a realização dos testes com as tarefas propostas, demonstrando que a abordagem de Treemap com hierarquia e glifos apresenta grande potencial na resolução de tarefas de InfoVis, representação visual e multidimensional de números pequenos, valores zero, números negativos, etc.

O nível de usabilidade do cenário GA foi avaliado, e obteve uma nota B- (72,67), o que indica que os participantes tiveram uma percepção da usabilidade em um nível Bom e intervalo de aceitação como Aceitável. Por fim, foi realizada uma avaliação qualitativa do cenário GA, que de modo geral, indicou que o glifo adaptativo melhorou a qualidade visual da informação no Treemap com glifos. Contudo, uma investigação maior sobre o uso de GAs deve ser realizada, no que tange a realização de uma avaliação comparativa em cenários GH e GA.

É importante destacar algumas limitações desta pesquisa, tais como:

- Como o modelo de design do glifo é independente de domínio, os usuários podem apresentar uma curva de aprendizado maior para o entendimento do glifo;
- O modelo de glifo só consegue representar dados ordinais se o usuário os fornecer já ordenados;

- Com o usuário podendo adicionar várias camadas ao glifo, o número de dimensões representadas visualmente no glifo pode afetar negativamente sua leitura;
- No glifo adaptativo, ao ocultar uma camada pode-se estar retirando da visualização um dado importante.
- O modelo do glifo adaptativo tende a apresentar uma dificuldade para representar dados quantitativos, representados por um profile glifo, uma vez que este, quando utilizado, sempre fica na camada N do glifo.

Como trabalhos futuros, pode-se destacar: sugerir e avaliar outras variáveis visuais; o uso de novos design e avaliações de glifos, considerando critérios de pré-atenção, memorização, qualidade visual para pequenas áreas; realizar um estudo para sugerir automaticamente a construção do glifo de acordo com os dados selecionados; realizar estudo para a adoção de novos modelos de aprendizado de máquina para a remoção de camadas do glifo adaptativo; realizar um estudo e sugerir um modelo de glifo que agregue as informações visuais dos itens de área reduzida da técnica Treemap; reaplicar as avaliações dos cenários, comparando os cenários GH e GA, considerando o uso de diferentes tamanhos de bases de dados, se possível, com bases de dados reais e permitindo a interação dos participantes com o protótipo de apoio com todas as suas funcionalidades de interação; e por fim, há um potencial quanto a expansão do modelo de glifos em camadas para outras técnicas de visualização, principalmente em árvore, tais como: *voronoi*, *bubble tree*, *sunburst*, entre outras.

REFERÊNCIAS

AUBER, D. et al. Gospermap: Using a gosper curve for laying out hierarchical data. *IEEE Transactions on Visualization and Computer Graphics*, IEEE, v. 19, n. 11, p. 1820–1832, 2013. Citado 2 vezes nas páginas 21 e 32.

BALDONADO, M. Q. W.; WOODRUFF, A.; KUCHINSKY, A. Guidelines for using multiple views in information visualization. In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. New York, NY, USA: Association for Computing Machinery, 2000. (AVI '00), p. 110–119. ISBN 1581132522. Disponível em: <https://doi.org/10.1145/345513.345271>. Citado na página 22.

BALZER, M.; DEUSSEN, O. Voronoi treemaps. In: *IEEE Symposium on Information Visualization*. [S.l.: s.n.], 2005. p. 49–56. ISSN 1522-404X. Citado 2 vezes nas páginas 21 e 32.

BANGOR, A.; KORTUM, P.; MILLER, J. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, Citeseer, v. 4, n. 3, p. 114–123, 2009. Citado na página 105.

BARBIERI, C. *BI2–Business intelligence: Modelagem & Qualidade*. [S.l.]: Elsevier Editora, 2011. Citado na página 69.

BAUDEL, T.; BROEKSEMA, B. Capturing the design space of sequential space-filling layouts. *IEEE Transactions on Visualization and Computer Graphics*, v. 18, n. 12, p. 2593–2602, Dec 2012. ISSN 2160-9306. Citado na página 20.

BEDDOW, J. Shape Coding of Multidimensional Data on a Microcomputer Display. In: *Proceedings of the 1st Conference on Visualization '90*. Los Alamitos, CA, USA: IEEE Computer Society Press, 1990. (VIS '90), p. 238–246. ISBN 0-8186-2083-8. Citado na página 40.

BEDERSON, B. B.; SHNEIDERMAN, B.; WATTENBERG, M. Ordered and quantum treemaps: Making effective use of 2d space to display hierarchies. *ACM Trans. Graph.*, Association for Computing Machinery, New York, NY, USA, v. 21, n. 4, p. 833–854, out. 2002. ISSN 0730-0301. Disponível em: <https://doi.org/10.1145/571647.571649>. Citado 3 vezes nas páginas 20, 32 e 33.

BERTIN, J. Semiology of graphics: Diagrams. *Networks, Maps*, v. 10, n. 00690805.1987, p. 10438353, 1983. Citado 5 vezes nas páginas 9, 36, 37, 38 e 66.

BLANCH, R.; LECOLINET, E. Browsing zoomable treemaps: Structure-aware multi-scale navigation techniques. *IEEE Transactions on Visualization and Computer Graphics*, Institute of Electrical and Electronics Engineers (IEEE), v. 13, n. 6, p. 1248–1253, nov 2007. Citado 3 vezes nas páginas 20, 96 e 112.

BORGO, R. et al. Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. In: SBERT, M.; SZIRMAY-KALOS, L. (Ed.). *Eurographics 2013 - State of the Art Reports*. [S.l.]: The Eurographics Association, 2013. ISSN 1017-4656. Citado 3 vezes nas páginas 22, 39 e 47.

BOUTHIER, C. *Treemap java library*. Projektinternetseite, 2001. Disponível em: <http://treemap.sourceforge.net/>. Citado na página 77.

BROOKE, J. et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, London–, v. 189, n. 194, p. 4–7, 1996. Citado 2 vezes nas páginas 103 e 104.

BRUCE, P.; BRUCE, A. *Practical statistics for data scientists: 50 essential concepts*. [S.l.]: "O'Reilly Media, Inc.", 2017. Citado na página 88.

BRULS, M.; HUIZING, K.; WIJK, J. J. van. Squarified Treemaps. In: *Eurographics*. [S.l.]: Springer Vienna, 2000. p. 33–42. Citado 5 vezes nas páginas 21, 24, 33, 34 e 35.

CARD, S. K.; MACKINLAY, J. D.; SHNEIDERMAN, B. *Readings in Information Visualization: Using Vision To Think*. [S.l.]: Morgan Kaufmann, 1999. 686 p. ISBN 978-1-55860-533-6. Citado 3 vezes nas páginas 19, 28 e 29.

CARPENDALE, M. S. T. Considering visual variables as a basis for information visualisation. University of Calgary, 2003. Citado 3 vezes nas páginas 37, 66 e 96.

CARR, D. Guidelines for designing information visualization applications. In: *Proceedings of the 1999 Ericsson Conference on Usability Engineering :*. [S.l.: s.n.], 1999. Godkänd; 1999; 20080426 (ysko). Citado na página 28.

CAUSSE, M.; HURTER, C. The physiological user's response as a clue to assess visual variables effectiveness. In: KUROSU, M. (Ed.). *Human Centered Design*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. p. 167–176. ISBN 978-3-642-02806-9. Citado na página 66.

CHEN, C. Information Visualization. *Information Visualization*, Springer Nature, v. 1, n. 1, p. 1–4, 2002. Citado na página 28.

CHERNOFF, H. The Use of Faces to Represent Points in k-Dimensional Space Graphically. *Journal of the American Statistical Association*, Informa UK Limited, v. 68, n. 342, p. 361–368, jun 1973. Citado na página 40.

CHLAN, E. B.; RHEINGANS, P. Multivariate glyphs for multi-object clusters. In: *IEEE Symposium on Information Visualization*. [S.l.]: IEEE, 2005. Citado na página 39.

CHUNG, D. H. et al. Glyph sorting: Interactive visualization for multi-dimensional data. *Information Visualization*, SAGE Publications, v. 14, n. 1, p. 76–90, nov 2013. Citado 3 vezes nas páginas 43, 48 e 53.

DIENER, M. J. Cohen's d. In: _____. *The Corsini Encyclopedia of Psychology*. [S.l.]: American Cancer Society, 2010. p. 1–1. ISBN 9780470479216. Citado na página 88.

ELMQVIST, N.; FEKETE, J.-D. Hierarchical Aggregation for Information Visualization: Overview, Techniques, and Design Guidelines. *IEEE Transactions on Visualization and Computer Graphics*, Institute of Electrical and Electronics Engineers (IEEE), v. 16, n. 3, p. 439–454, 2010. Citado 3 vezes nas páginas 51, 52 e 53.

FENG, D. et al. Evaluation of glyph-based multivariate scalar volume visualization techniques. In: *Symposium on Applied Perception in Graphics and Visualization*. [S.l.]: ACM Press, 2009. Citado na página 39.

FEW, S. *Now You See It*. [S.l.]: Analytics Press, 2009. ISBN 0970601980. Citado 2 vezes nas páginas 19 e 28.

FISCHER, F.; FUCHS, J.; MANSMANN, F. Clockmap: Enhancing circular treemaps with temporal glyphs for time-series data. In: MEYER, M.; WEINKAUF, T. (Ed.). *Proceedings of the Eurographics Conference on Visualization (EuroVis 2012 Short Papers)*. Vienna, Austria: [s.n.], 2012. p. 97–101. Citado 4 vezes nas páginas 22, 51, 52 e 53.

FLURY, B.; RIEDWYL, H. Graphical representation of multivariate data by means of asymmetrical faces. *Journal of the American Statistical Association*, Taylor & Francis, v. 76, n. 376, p. 757–765, 1981. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/01621459.1981.10477718>. Citado na página 22.

FORSELL, C. A guide to scientific evaluation in information visualization. In: *2010 14th International Conference Information Visualisation*. [S.l.: s.n.], 2010. p. 162–169. ISSN 1550-6037. Citado 2 vezes nas páginas 57 e 58.

FREEMAN, E. et al. *Head first design patterns*. [S.l.]: "O'Reilly Media, Inc.", 2004. ISSN 9780596007126. Citado na página 77.

FUCHS, J. et al. The influence of contour on similarity perception of star glyphs. *IEEE Transactions on Visualization and Computer Graphics*, v. 20, n. 12, p. 2251–2260, Dec 2014. ISSN 2160-9306. Citado na página 22.

FUCHS, J. et al. A Systematic Review of Experimental Studies on Data Glyphs. *IEEE Transactions on Visualization and Computer Graphics*, Institute of Electrical and Electronics Engineers (IEEE), v. 23, n. 7, p. 1863–1879, jul 2017. Citado 7 vezes nas páginas 39, 40, 41, 42, 44, 47 e 54.

FUCHS, R.; HAUSER, H. Visualization of multi-variate scientific data. *Computer Graphics Forum*, v. 28, n. 6, p. 1670–1690, 2009. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8659.2009.01429.x>. Citado na página 22.

GÖRG, C. et al. Visual representations. In: _____. *Human-Centered Visualization Environments: GI-Dagstuhl Research Seminar, Dagstuhl Castle, Germany, March 5-8, 2006, Revised Lectures*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 163–230. ISBN 978-3-540-71949-6. Disponível em: <https://doi.org/10.1007/978-3-540-71949-6_4>. Citado na página 28.

HAHN, S.; DÖLLNER, J. Hybrid-treemap layouting. In: EUROGRAPHICS ASSOCIATION. *Proceedings of the Eurographics/IEEE VGTC Conference on Visualization: Short Papers*. [S.l.], 2017. p. 79–83. Citado 4 vezes nas páginas 20, 21, 32 e 33.

Haisen Zhao; Lu, L. Variational circular treemaps for interactive visualization of hierarchical data. In: *IEEE Pacific Visualization Symposium*. [S.l.: s.n.], 2015. p. 81–85. ISSN 2165-8765. Citado na página 32.

HAM, F. v.; WIJK, J. J. v. Beamtrees: compact visualization of large hierarchies. In: *IEEE Symposium on Information Visualization*. [S.l.: s.n.], 2002. p. 93–100. ISSN 1522-404X. Citado na página 21.

HARTIGAN, J. A. Printer graphics for clustering. *Journal of Statistical Computation and Simulation*, Informa UK Limited, v. 4, n. 3, p. 187–213, jan 1975. Citado na página 40.

HUANG, M. L.; HUANG, T.-H.; ZHANG, J. TreemapBar: Visualizing Additional Dimensions of Data in Bar Chart. In: *13th International Conference Information Visualisation*. [S.l.]: IEEE, 2009. Citado na página 21.

JERN, M.; ROGSTADIUS, J.; ÅSTRöM, T. Treemaps and Choropleth Maps Applied to Regional Hierarchical Statistical Data. In: *13th International Conference Information Visualisation*. [S.l.]: IEEE, 2009. Citado na página 21.

JIANU, R. et al. How to display group information on node-link diagrams: An evaluation. *IEEE Transactions on Visualization and Computer Graphics*, v. 20, n. 11, p. 1530–1541, Nov 2014. ISSN 2160-9306. Citado na página 20.

JOHNSON, B.; SHNEIDERMAN, B. Tree-maps: a space-filling approach to the visualization of hierarchical information structures. In: *Proceeding Visualization 91*. [S.l.]: IEEE Comput. Soc. Press, 1991. Citado na página 20.

KECK, M. et al. Towards glyph-based visualizations for big data clustering. In: *Proceedings of the 10th International Symposium on Visual Information Communication and Interaction*. New York, NY, USA: Association for Computing Machinery, 2017. (VINCI '17), p. 129–136. ISBN 9781450352925. Disponível em: <https://doi.org/10.1145/3105971.3105979>. Citado na página 43.

KEIM, D. A. Visual techniques for exploring databases. In: *Knowledge Discovery in Databases (KDD'97)*. [S.l.: s.n.], 1997. Citado na página 55.

KELLEY, K.; PREACHER, K. J. On effect size. *Psychological Methods*, American Psychological Association (APA), v. 17, n. 2, p. 137–152, 2012. Citado na página 87.

KHAN, S. et al. A Study on Glyph-based Visualisation with Dense Visual Context. In: BORGO, R.; TANG, W. (Ed.). *Computer Graphics and Visual Computing (CGVC)*. [S.l.]: The Eurographics Association, 2014. ISBN 978-3-905674-70-5. Citado 2 vezes nas páginas 51 e 53.

KLEINER, B.; HARTIGAN, J. A. Representing Points in Many Dimensions by Trees and Castles. *Journal of the American Statistical Association*, Informa UK Limited, v. 76, n. 374, p. 260–269, jun 1981. Citado na página 40.

KONG, N.; HEER, J.; AGRAWALA, M. Perceptual guidelines for creating rectangular treemaps. *IEEE transactions on visualization and computer graphics*, IEEE, v. 16, n. 6, p. 990–998, 2010. Citado na página 20.

LEGG, P. A. et al. MatchPad: Interactive glyph-based visualization for real-time sports performance analysis. *Computer Graphics Forum*, Wiley, v. 31, n. 3pt4, p. 1255–1264, jun 2012. Citado 3 vezes nas páginas 48, 49 e 53.

LEUNG, Y. K.; APPERLEY, M. D. A review and taxonomy of distortion-oriented presentation techniques. *ACM Transactions on Computer-Human Interaction (TOCHI)*, ACM New York, NY, USA, v. 1, n. 2, p. 126–160, 1994. Citado na página 55.

LONG, L. K. et al. A study on the effectiveness of tree-maps as tree visualization techniques. *Procedia Computer Science*, Elsevier BV, v. 124, p. 108–115, 2017. ISSN 1877-0509. 4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia. Citado na página 20.

LOURENCO, R. A. de_Moraes et al. Exploring, comparing and coordinating multiple datasets in an information visualization tool. In: *2012 16th International Conference on Information Visualisation*. [S.l.: s.n.], 2012. p. 613–618. ISSN 1550-6037. Citado na página 22.

MACKINLAY, J. Automating the design of graphical presentations of relational information. *ACM Trans. Graph.*, Association for Computing Machinery, New York, NY, USA, v. 5, n. 2, p. 110–141, abr. 1986. ISSN 0730-0301. Disponível em: <https://doi.org/10.1145/22949.22950>. Citado 7 vezes nas páginas 10, 38, 39, 60, 61, 66 e 112.

MAGUIRE, E. et al. Taxonomy-based glyph design—with a case study on visualizing workflows of biological experiments. *IEEE Transactions on Visualization and Computer Graphics*, Institute of Electrical and Electronics Engineers (IEEE), v. 18, n. 12, p. 2603–2612, 2012. Citado 8 vezes nas páginas 10, 19, 20, 22, 43, 48, 49 e 53.

MUNZNER, T. H3: laying out large directed graphs in 3d hyperbolic space. In: *Proceedings of VIZ '97: Visualization Conference, Information Visualization Symposium and Parallel Rendering Symposium*. [S.l.: s.n.], 1997. p. 2–10. ISSN null. Citado 2 vezes nas páginas 19 e 39.

MUNZNER, T.; MAGUIRE, E. *Visualization analysis and design*. Boca Raton, FL: CRC Press, 2015. (AK Peters visualization series). Disponível em: <https://cds.cern.ch/record/2001992>. Citado 5 vezes nas páginas 19, 39, 44, 45 e 66.

NAKAGAWA, S.; CUTHILL, I. C. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, Wiley, v. 82, n. 4, p. 591–605, nov 2007. Citado na página 87.

NASCIMENTO, H.; FERREIRA, C. Visualização de informações – uma abordagem prática. In: _____. [S.l.]: Sociedade Brasileira de Computação, 2005. (A universidade da computação: Um agente de Inovação e Conhecimento, 62), cap. 2, p. 1262–1309. XXV Congresso da Sociedade Brasileira de Computação. Citado na página 29.

OPACH, T.; RØD, J. K. Augmenting the usability of parallel coordinate plot: The polyline glyphs. *Information Visualization*, SAGE Publications, v. 17, n. 2, p. 108–127, feb 2017. Citado na página 22.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, v. 12, n. Oct, p. 2825–2830, 2011. Citado na página 69.

PETERSEN, T. K. Introduction to this book. In: *Inquiry-Based Enumerative Combinatorics*. Springer, Cham, 2019. p. 1–14. Disponível em: <https://doi.org/10.1007/978-3-030-18308-0_0>. Citado na página 23.

PICKETT, R. M.; GRINSTEIN, G. G. Iconographic Displays For Visualizing Multidimensional Data. In: *Proceedings of the 1988 IEEE International Conference on Systems, Man, and Cybernetics*. [S.l.]: IEEE, 1988. Citado na página 40.

PILAR, D. H. F.; WARE, C. Representing flow patterns by using streamlines with glyphs. *IEEE Transactions on Visualization and Computer Graphics*, v. 19, n. 8, p. 1331–1341, Aug 2013. ISSN 2160-9306. Citado na página 22.

PILLAT, R. M.; VALIATI, E. R.; FREITAS, C. M. Experimental study on evaluation of multidimensional information visualization techniques. In: *Proceedings of the 2005 Latin American conference on Human-computer interaction*. [S.l.: s.n.], 2005. p. 20–30. Citado na página 44.

RIBARSKY, W. et al. Glyphmaker: creating customized visualizations of complex data. *Computer*, Institute of Electrical and Electronics Engineers (IEEE), v. 27, n. 7, p. 57–64, jul 1994. Citado 2 vezes nas páginas 50 e 53.

ROPINSKI, T.; OELTZE, S.; PREIM, B. Visual computing in biology and medicine: Survey of glyph-based visualization techniques for spatial multivariate medical data. *Comput. Graph.*, Pergamon Press, Inc., USA, v. 35, n. 2, p. 392–401, abr. 2011. ISSN 0097-8493. Disponível em: <https://doi.org/10.1016/j.cag.2011.01.011>. Citado na página 19.

ROTH, R. E. Visual variables. In: _____. *International Encyclopedia of Geography*. American Cancer Society, 2017. p. 1–11. ISBN 9781118786352. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118786352.wbieg0761>. Citado na página 66.

SAKET, B. et al. Node, node-link, and node-link-group diagrams: An evaluation. *IEEE Transactions on Visualization and Computer Graphics*, v. 20, n. 12, p. 2231–2240, Dec 2014. ISSN 2160-9306. Citado na página 20.

SANTOS, D. H. dos et al. Analyzing the effect of different partial overlap sizes in perceiving visual variables. In: *23th International Conference Information Visualisation (IV)*. Paris, France: [s.n.], 2019. p. 37–43. Citado 4 vezes nas páginas 56, 60, 61 e 66.

SAURO, J.; LEWIS, J. R. *Quantifying the user experience: Practical statistics for user research*. [S.l.]: Morgan Kaufmann, 2016. Citado 3 vezes nas páginas 103, 104 e 105.

SCHROEDER, W. J.; VOLPE, C. R.; LORENSEN, W. E. The stream polygon-a technique for 3D vector field visualization. In: *Proceeding Visualization 91*. [S.l.]: IEEE Comput. Soc. Press, 1991. Citado na página 41.

SHNEIDERMAN, B. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics*, Association for Computing Machinery (ACM), v. 11, n. 1, p. 92–99, jan 1992. Citado 4 vezes nas páginas 20, 30, 31 e 32.

SHNEIDERMAN, B. The eyes have it: a task by data type taxonomy for information visualizations. In: *Proceedings 1996 IEEE Symposium on Visual Languages*. [S.l.]: IEEE Comput. Soc. Press, 1996. Citado 3 vezes nas páginas 29, 30 e 55.

SHNEIDERMAN, B.; WATTENBERG, M. Ordered treemap layouts. In: *Proceedings of the IEEE Symposium on Information Visualization 2001 (INFOVIS'01)*. USA: IEEE Computer Society, 2001. (INFOVIS '01), p. 73. ISBN 0769513425. Citado 2 vezes nas páginas 20 e 21.

SIEGEL, J. H. et al. The surgical implications of physiologic patterns in myocardial infarction shock. In: *Thirty-third Annual Meeting of the Society of University Surgeons*. San Francisco, Calif.: [s.n.], 1972. p. 10–12. Citado na página 40.

SPENCE, R. *Information Visualization: Design for Interaction*. Second. [S.l.]: Pearson, 2007. ISBN 0132065509. Citado 3 vezes nas páginas 21, 31 e 36.

SPENCE, R. *Information Visualization: An Introduction*. Third. [S.l.]: Springer, Cham, 2014. ISBN 978-3-319-07340-8. Citado na página 19.

TELEA, A. Combining Extended Table Lens and Treemap Techniques for Visualizing Tabular Data. In: *Joint Eurographics/IEEE VGTC Conference on Visualization*. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2006. (EUROVIS'06), p. 51–58. ISBN 3-905673-31-2. Citado na página 21.

THERON, R. et al. Overlapper: movie analyzer. *Infovis Confererence Compendium*, p. 140–141, 2007. Citado 4 vezes nas páginas 10, 49, 50 e 53.

TOIT, S. H. C. D.; STEYN, A. G. W.; STUMPF, R. H. *Graphical exploratory data analysis*. [S.l.]: Springer Science & Business Media, 1986. Citado 2 vezes nas páginas 40 e 64.

TOMINSKI, C.; SCHULZE-WOLLGAST, P.; SCHUMANN, H. 3d information visualization for time dependent data on maps. In: *Ninth International Conference on Information Visualisation (IV'05)*. [S.l.: s.n.], 2005. p. 175–181. ISSN 2375-0138. Citado na página 22.

TU, Y.; SHEN, H. Visualizing changes of hierarchical data using treemaps. *IEEE Transactions on Visualization and Computer Graphics*, v. 13, n. 6, p. 1286–1293, Nov 2007. ISSN 2160-9306. Citado na página 21.

WANG, W. et al. Visualization of large hierarchical data by circle packing. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2006. (CHI '06), p. 517–520. ISBN 1595933727. Disponível em: <https://doi.org/10.1145/1124772.1124851>. Citado na página 19.

WARD, M. O. A Taxonomy of Glyph Placement Strategies for Multidimensional Data Visualization. *Information Visualization*, SAGE Publications, v. 1, n. 3-4, p. 194–210, 2002. ISSN 1473-8716. Citado na página 42.

WARD, M. O.; GRINSTEIN, G.; KEIM, D. *Interactive Data Visualization - Foundations, Techniques, and Applications*. Second. [S.l.]: A K Peters/CRC Press, 2015. ISBN 9780429173226. Citado 8 vezes nas páginas 19, 39, 40, 41, 43, 54, 64 e 66.

WARE, C. *Information Visualization: Perception for Design*. Second. San Francisco, CA: Elsevier, 2004. ISBN 9781558608191. Citado 5 vezes nas páginas 19, 28, 36, 39 e 66.

WEHREND, S.; LEWIS, C. A problem-oriented classification of visualization techniques. In: IEEE. *Proceedings of the First IEEE Conference on Visualization: Visualization90*. [S.l.], 1990. p. 139–143. Citado na página 44.

WITTENBRINK, C. M.; PANG, A. T.; LODHA, S. K. Glyphs for visualizing uncertainty in vector fields. *IEEE Transactions on Visualization and Computer Graphics*, Institute of Electrical and Electronics Engineers (IEEE), v. 2, n. 3, p. 266–279, 1996. Citado na página 40.

YANG, B.; CAO, W. The ordered treemap of weight divided layout algorithm. *Journal of Computers*, v. 30, n. 5, p. 31–45, 2019. Citado 2 vezes nas páginas 32 e 33.

YASOJIMA, E. K. K.; MEIGUINS, B. S.; MEIGUINS, A. S. o. Collaborative augmented reality application for information visualization support. In: *Proceedings of the 2011 15th International Conference on Information Visualisation*. USA: IEEE Computer Society, 2011. (IV '11), p. 170–175. ISBN 9780769544762. Disponível em: <https://doi.org/10.1109/IV.2011.44>. Citado na página 19.

YATES, A. et al. Visualizing multidimensional data with glyph sploms. *Computer Graphics Forum*, v. 33, n. 3, p. 301–310, 2014. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12386>. Citado na página 22.

YOST, B.; NORTH, C. Single complex glyphs versus multiple simple glyphs. In: *CHI 05 extended abstracts on Human factors in computing systems*. [S.l.]: ACM Press, 2005. Citado na página 39.

ZHANG, C. et al. Glyph-based comparative visualization for diffusion tensor fields. *IEEE Transactions on Visualization and Computer Graphics*, v. 22, n. 1, p. 797–806, Jan 2016. ISSN 2160-9306. Citado na página 22.

ZHAO, S.; MCGUFFIN, M. J.; CHIGNELL, M. H. Elastic hierarchies: combining treemaps and node-link diagrams. In: *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* [S.l.: s.n.], 2005. p. 57–64. ISSN 1522-404X. Citado na página 21.

APÊNDICE A . PARECER DO COMITÊ DE ÉTICA

PARECER DO COMITÊ DE ÉTICA

DADOS DA VERSÃO DO PROJETO DE PESQUISA

Título da Pesquisa: Avaliações Quantitativas e Qualitativas em Técnicas de Visualização da Informação
Pesquisador Responsável: BIANCHI SERIQUE MEIGUINS
Área Temática:
Versão: 1
CAAE: 15522319.2.0000.0018
Submetido em: 07/06/2019
Instituição Proponente: Universidade Federal do Pará
Situação da Versão do Projeto: Aprovado
Localização atual da Versão do Projeto: Pesquisador Responsável
Patrocinador Principal: Financiamento Próprio



Comprovante de Recepção: PB_COMPROVANTE_RECEPCAO_1374065

DOCUMENTOS DO PROJETO DE PESQUISA

LISTA DE APRECIAÇÕES DO PROJETO

Apreciação	Pesquisador Responsável	Versão	Submissão	Modificação	Situação	Exclusiva do Centro Coord.	Ações
PO	BIANCHI SERIQUE MEIGUINS	1	07/06/2019	03/10/2019	Aprovado	Não	

APÊNDICE B . TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO- TCLE

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Expandindo a visualização de dados multidimensionais em Treemaps com glifos
TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

(Conforme Resolução 510/16 do Conselho Nacional de Saúde)

Esta pesquisa tem por objetivo avaliar a técnica treemap com glifos em camadas em tarefas de busca e comparação. Adicionalmente, este trabalho avalia um modelo em camadas de glifos multidimensionais, no qual cada uma das camadas é formada por uma variável visual, considerando o fator de oclusão e a quantidade de valores distintos que a variáveis visuais podem assumir.

Informamos que:

a) Sua participação nesta pesquisa será voluntária, não havendo qualquer tipo de remuneração;

b) Você é livre para aceitar ou recusar responder a qualquer pergunta, bem como para interromper a coleta dos dados a qualquer momento, se assim lhe for conveniente;

c) Os riscos por sua participação nesta pesquisa serão mínimos, pois não causará qualquer tipo de transtorno à sua saúde e/ou integridade física e psíquica;

d) Os resultados deste estudo serão divulgados com finalidade científica e social, porém, os nomes dos participantes não serão revelados, nem haverá qualquer referência de caráter individual, ou seja, seus dados estarão protegidos por sigilo e anonimato;

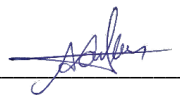
Declaro que fui adequadamente esclarecido (a) sobre a natureza desta pesquisa e da minha participação, nos termos deste documento. Declaro, ainda, que concordo em participar por livre e espontânea vontade e que não sofri qualquer tipo de pressão para tomar esta decisão.

Belém, ____ de _____ de 2020.

_____

Assinatura do participante

Responsáveis pela Pesquisa: Anderson Gregório Marques Soares (anderson.gmarques@gmail.com).



_____

Assinatura do Pesquisador

Nota: Este Termo de Consentimento Livre e Esclarecido será elaborado em duas vias. Depois de assinadas, uma ficará com o participante e a outra com o pesquisador.

APÊNDICE C . TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO- TCLE

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Expandindo a visualização de dados multidimensionais em Treemaps com glifos
TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

(Conforme Resolução 510/16 do Conselho Nacional de Saúde)

Esta pesquisa tem por objetivo avaliar a técnica treemap com glifos adaptativos em tarefas de busca e comparação.

Informamos que:

a) Sua participação nesta pesquisa será voluntária, não havendo qualquer tipo de remuneração;

b) Você é livre para aceitar ou recusar responder a qualquer pergunta, bem como para interromper a coleta dos dados a qualquer momento, se assim lhe for conveniente;

c) Os riscos por sua participação nesta pesquisa serão mínimos, pois não causará qualquer tipo de transtorno à sua saúde e/ou integridade física e psíquica;

d) Os resultados deste estudo serão divulgados com finalidade científica e social, porém, os nomes dos participantes não serão revelados, nem haverá qualquer referência de caráter individual, ou seja, seus dados estarão protegidos por sigilo e anonimato;

Declaro que fui adequadamente esclarecido (a) sobre a natureza desta pesquisa e da minha participação, nos termos deste documento. Declaro, ainda, que ao executar o arquivo **TreemapGlyph_v21.jar** concordei em participar por livre e espontânea vontade e que não sofri qualquer tipo de pressão para tomar esta decisão.

Belém, ____ de _____ de 2020.

_____

Assinatura do participante

Responsáveis pela Pesquisa: Anderson Gregório Marques Soares (anderson.gmarques@gmail.com).

_____

Assinatura do Pesquisador

APÊNDICE D . AVALIAÇÃO GLIFO EM CAMADAS - AMBIENTE G

Configuração Testes

Configuração do ambiente treemap sem hierarquias e com glifos - Ambiente G:

Configurações do Treemap:

	Tamanho	Hierarquia	Cor	Label
T1	SAME_SIZE		Temperatura_C	
T2	SAME_SIZE			
T3	SAME_SIZE		região	
T4	SAME_SIZE			
T5	PhChuva		Temperatura_C	
T6	UmidadeAr_%		Regiao	
T7	SAME_SIZE			
T8	SAME_SIZE			

Configurações dos glifos:

	Textura	Cor	Forma	Letra	Glifo quantitativo
T1	Ph_Chuva_CAT	Regiao	AlertaDefesaCivik	IntensidadeVento_CAT	
T2		Região	PhChuva_CAT		
T3					TemperaturaC; intensidade vento km/h
T4	Localidade	Regiao	IntensidadeVento_CAT		PhChuva; VolumeChuva_mm/h
T5		Região		IntensidadeVento_CAT	
T6			PhChuva_CAT	Temperatura_CAT	
T7		Regiao	Localidade		IntensidadeVento; VolumeChuva
T8		Regiao	EstacaoAno		VolumeChuva_mm/h; UmidadeAr_%

Ordem entre as camadas do glifo: Textura > cor> forma > Letra ou glifo quantitativo (lê-se o símbolo “>” como precede).

APÊNDICE E . AVALIAÇÃO GLIFO EM CAMADAS - AMBIENTE GH

Configuração Testes

Configuração do ambiente treemap com hierarquias e com glifos - Ambiente GH:

Configurações do Treemap:

	Tamanho	Hierarquia	Cor	Label
T1	SAME_SIZE	Regiao	Temperatura_C	
T2	SAME_SIZE	Região		
T3	SAME_SIZE	Região		
T4	SAME_SIZE	Regiao /IntensidadeVento_CAT-		
T5	PhChuva	Regiao	Temperatura_C	
T6	UmidadeAr_%	Regiao		
T7	SAME_SIZE	Regiao-Localidade		

Configurações dos glifos:

	Textura	Cor	Forma	Letra	Glifo quantitativo
T1	Ph_Chuva_CAT	IntensidadeVento_CAT	AlertaDefesaCivil		
T2			PhChuva_CAT		
T3					TemperaturC; intensidade vento km/h
T4			Localidade		PhChuva; VolumeChuva_mm/h
T5				IntensidadeVento_CAT	
T6			PhChuva_CAT	Temperatura_CAT	
T7					IntensidadeVento; VolumeChuva
T8					VolumeChuva_mm/h; UmidadeAr_%

Ordem entre as camadas do glifo: Textura > cor> forma > Letra ou glifo quantitativo (lê-se o símbolo “>” como precede).

APÊNDICE F . AVALIAÇÃO GLIFO EM CAMADAS - AMBIENTE H

Configuração Testes

Configuração do ambiente treemap com hierarquias e sem glifos - Ambiente H:

Configurações do Treemap:

	Tamanho	Hierarquia	Cor	Label
T1	SAME_SIZE	Regiao-Ph_Chuva_CAT; IntensidadeVento_CAT-	Temperatura_C	AlertaDefesaCivil
T2	SAME_SIZE	Região	PhChuva_CAT	
T3	SAME_SIZE	Região	intensidadeVentoKM/h	temperaturaC
T4	SAME_SIZE	Regiao; IntensidadeVento_CAT-Localidade	VolumeChuva_mm/h	PhChuva
T5	PhChuva	Regiao	Temperatura_C	IntensidadeVento_CAT
T6	UmidadeAr_%	PhChuva_CAT	Regiao	Temperatura_CAT
T7	SAME_SIZE	Regiao-Localidade	IntensidadeVento	VolumeChuva_mm/h

Configurações dos glifos:

	Textura	Cor	Forma	Letra	Glifo quantitativo
T1					
T2					
T3					
T4					
T5					
T6					
T7					
T8					

Ordem entre as camadas do glifo: Esse cenário não utilizou glifos.

APÊNDICE G . AVALIAÇÃO GLIFO ADAPTATIVO - AMBIENTE GA

Configurações Treemap

Configuração do ambiente treemap com hierarquias e com glifos adaptativos - Ambiente GA:

Configurações do Treemap:

	Hierarquia	Tamanho	Cor	Rótulo
T1	Regiao	MediaUmidadeAr_%	---	---
T2	Regiao	NumeroVitimas	---	---
T3	Localidade; Temperatura_CAT;	NumeroVitimas	---	---
T4	Regiao	NumeroVitimas	---	---
T5	Regiao;	VolumeChuva_mm/h	---	---

Configurações glifo:

	Textura	Cor	Forma	Texto	Posição	Orientação
T1	---	AlertaDefesaCivil	IntensidadeVento_CAT	---	---	---
T2	---	IntensidadeChuva	AlertaDefesaCivil	Temperatura_CAT	---	---
T3	IntensidadeVento_CAT	Regiao	IntensidadeChuva	AlertaDefesaCivil	---	---
T4	Localidade	IntensidadeVento_CAT	Temperatura_CAT	AlertaDefesaCivil	---	IntensidadeChuva
T5	IntensidadeChuva	IntensidadeVento_CAT	PhChuva_Cat	AlertaDefesaCivil	Estacao Ano	Temperatura_CAT

Ordem das camadas do glifo (lê-se o símbolo “>” como precede):

	Ordem
T1	Cor > Forma
T2	Cor > Forma > Texto
T3	Cor > Textura > Forma > Texto
T4	Cor > Forma > Textura > Orientação > Texto
T5	Textura > Orientação > Cor > Posição > Forma > Texto