



Bruna Moraes Ferreira

**VIS-SeMa: Apoio ao Design de Visualizações
de Dados para Novatos**

Tese de Doutorado

Tese apresentada como requisito parcial para a obtenção do grau de Doutor pelo Programa de Pós-graduação em Informática, do Departamento de Informática da PUC-Rio.

Orientador: Profa. Simone Diniz Junqueira Barbosa

Rio de Janeiro
Setembro de 2024



Bruna Moraes Ferreira

**VIS-SeMa: Apoio ao Design de Visualizações
de Dados para Novatos**

Tese apresentada como requisito parcial para a obtenção do grau de Doutor pelo Programa de Pós-graduação em Informática. Aprovada pela Comissão Examinadora abaixo:

Profa. Simone Diniz Junqueira Barbosa

Orientador

Departamento de Informática – PUC-Rio

Prof. Alberto Barbosa Raposo

Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio

Profa. Greis Francy Mireya Silva Calpa

Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio

Profa. Milene Selbach Silveira

PUC-RS

Profa. Isabela Gasparini

UDESC

Rio de Janeiro, 30 de Setembro de 2024

Todos os direitos reservados. A reprodução, total ou parcial do trabalho, é proibida sem a autorização da universidade, do autor e do orientador.

Bruna Moraes Ferreira

Graduou-se em Bacharelado em Ciência da Computação pela Universidade Federal do Amazonas (Manaus/AM). Fez mestrado no Departamento Informática da Universidade Federal do Amazonas, especializando-se na área de Engenharia de Software/Interação Humano-Computador. Desenvolveu técnica de Persona para apoiar Elicitação de Requisitos. Trabalhou como Analista de TI/UX na Empresa Processamento de Dados do Amazonas (PRODAM) e Product Owner/ Agile Master na iniciativa Exacta/PUC-Rio.

Ficha Catalográfica

Ferreira, Bruna Moraes

VIS-SeMa: Apoio ao Design de Visualizações de Dados para Novatos / Bruna Moraes Ferreira; orientador: Simone Diniz Junqueira Barbosa. – 2024.

105 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2024.

Inclui bibliografia

1. visualização de dados – Teses. 2. design de visualizações – Teses. 3. apoio à criação de visualizações – Teses. 4. letramento em visualização de dados – Teses. 5. ensino de visualização de dados – Teses. 6. visualização de dados. 7. design de visualizações. 8. apoio à criação de visualizações. 9. letramento em visualização de dados. I. Barbosa, Simone Diniz Junqueira. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

À minha família.

Agradecimentos

Em primeiro lugar agradeço a Deus por ter me dado todas as oportunidades. Obrigada por sempre me guiar e dar forças para vencer os obstáculos e superar os medos. Agradeço também à minha família por sempre me apoiar. Agradeço à minha irmã Sabrina pelas conversas e pela força. Agradeço em especial a minha mãe Ruth por me ajudar durante toda minha vida e por sempre me incentivar nos estudos. Obrigada também carinho e pelo apoio, tudo isso foi e continua sendo muito importante durante toda minha caminhada.

Agradeço à minha orientadora Simone pela confiança, paciência, compreensão, conhecimentos passados e pelas oportunidades, que proporcionaram meu crescimento enquanto pesquisadora e pessoa. Agradeço também aos professores Hélio e Marcos Kalinowski pela oportunidade de participação nos projetos da iniciativa ExACTa, que trouxeram experiências profissionais e acadêmicas e contribuiriam para minha formação.

Obrigada aos amigos (não irei citar nomes para não esquecer ninguém) pelo companheirismo, pelas ajudas, pela amizade, pelos momentos de alegria e aprendizado e pelo apoio e acolhimento que sempre recebi de todos. Agradeço a todo o Bloom Studio, às professoras e amigas do pole pela inspiração e pelos momentos de diversão que se tornaram uma válvula de escape durante as dificuldades do processo. Agradeço aos amigos Walter e Dieinison pelas dicas e pelo apoio. Agradeço Ariane e Gabriel pelas colaborações com a pesquisa.

Agradeço também aos professores Alberto Barbosa Raposo, Greis Silva-Calpa, Isabela Gasparini e Milene Selbach por aceitarem o convite para participação na banca de avaliação desta Tese.

Agradeço ainda a todos os participantes dos estudos que permitiram a coleta de dados para os resultados desta pesquisa de tese. Finalmente, gostaria de agradecer Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo financiamento parcial deste trabalho. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior- Brasil (CAPES)- Código de Financiamento 001.

Resumo

Ferreira, Bruna Moraes; Barbosa, Simone Diniz Junqueira. **VIS-SeMa: Apoio ao Design de Visualizações de Dados para Novatos**. Rio de Janeiro, 2024. 105p. Tese de Doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Interpretar e criar visualizações de dados são habilidades essenciais na sociedade atual, em que o uso de dados é intensivo. Visualizações são ferramentas para extrair e comunicar informações de dados de forma eficiente. No entanto, novatos, com nenhuma ou pouca experiência com visualizações, muitas vezes têm dificuldades com atividades que envolvem mapeamentos visuais. Compreender essas dificuldades e investigar maneiras de melhorar o design de visualização de informação visam a capacitar os novatos a explorar melhor seus dados de forma independente. Nesse contexto, esta pesquisa almeja identificar e compreender lacunas no processo de design de visualização, e fornecer subsídios para que novatos possam planejar suas visualizações de forma adequada. Durante a pesquisa, realizamos dois estudos empíricos. O primeiro foi um estudo de observação do processo de design de visualizações, realizado com cinco profissionais de diferentes empresas. Neste estudo, mapeamos as etapas do processo e o material criado foi avaliado por dois especialistas em visualizações. Identificamos dificuldades e decisões inadequadas dos participantes, principalmente nas atividades de escolha de visualizações e mapeamento de variáveis para estas visualizações. Com base nos resultados e em trabalhos extraídos da literatura, propusemos e avaliamos um material didático estruturado, VIS-SeMa. O material é baseado em *templates* genéricos e exemplos concretos de questões e visualizações para apoiar novatos nas atividades de seleção de visualizações e mapeamento de variáveis. No segundo estudo, avaliamos a viabilidade do VIS-SeMa. Os resultados mostraram que o material apoiou boa parte dos participantes na seleção de visualizações e criação de mapeamentos visuais de forma correta. Nossa pesquisa contribui também com um processo de estudo baseado em utilizar questões para avaliar visualizações criadas e identificar decisões e dificuldades de novatos. Com o material VIS-SeMa, observamos a preferência dos participantes por conteúdos visuais e que a apresentação de exemplos concretos ajudou no entendimento de como fazer o mapeamento das variáveis para as visualizações e que combinar a anatomia (*templates* genéricos) das questões e das visualizações com os exemplos facilitou a compreensão por novatos.

Palavras-chave

visualização de dados; design de visualizações; apoio à criação de visualizações; letramento em visualização de dados.

Abstract

Ferreira, Bruna Moraes; Barbosa, Simone Diniz Junqueira (Advisor). **VIS-SeMa: Supporting Novices in Designing Data Visualizations**. Rio de Janeiro, 2024. 105p. Tese de Doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Interpreting and creating data visualizations are essential skills in the current data-intensive age. Visualizations are a tool to efficiently extract and communicate information from data. However, novices often have difficulties with visual mapping activities. Understanding these difficulties and investigating ways to improve data visualization design aim to empower novices to explore their data better. In this context, this research aimed to identify and understand gaps in the visualization design process, providing support to novices to select and plan their visualizations more appropriately. With this objective, we conducted two empirical studies. First, we performed an observational study in a visualization design process with five practitioners from different companies. In this study, we mapped the steps of the process, and the material created was evaluated by two visualization experts. Based on this analysis, we identified difficulties and decisions during the process, mainly in choosing appropriate visualizations and mapping variables onto these visualizations. Based on the results and works extracted from the literature, we proposed and evaluated a structured material, VIS-SeMa. The material is composed of topics and uses generic question *templates* and concrete examples of questions and visualizations to support novices in selecting a visualization type and mapping variables. In the second study, we evaluated the feasibility of the proposed material. The results of the evaluation showed that the material supported most participants in selecting visualizations and creating visual mappings correctly. From this research, we proposed a study process based on analytical questions to evaluate created visualizations and identify decisions, and difficulties faced by novices. Additionally, we proposed VIS-SeMa, a material to support novices in visualization design. The results of the feasibility study showed that participants preferred visual content and that presenting concrete examples helped them understand how to map variables to visualizations. In addition, we observed that combining the anatomy of the questions and visualizations with the examples improved novices' understanding.

Keywords

data visualization; visualization design; supporting visualization creation; visualization literac.

Sumário

1	Introdução	1
1.1	Motivação	2
1.2	Definição do problema	3
1.3	Metodologia e contribuições esperadas	4
1.4	Organização deste Documento	6
2	Fundamentação	7
2.1	Novatos em Visualização de Dados	7
2.2	Letramento em Visualização de Dados	8
2.3	Processo de design de visualizações	9
2.4	Estudos sobre o Design de Visualizações	9
2.5	Materiais didáticos para apoio ao design de visualizações	11
2.6	Considerações	13
3	Investigando o processo de criação de visualizações	14
3.1	Contexto e Participantes do Estudo	14
3.2	Materiais do Estudo	16
3.3	Etapas do Estudo	19
3.4	Análise dos resultados	20
3.4.1	Análise do Processo de Design	20
3.4.2	Inspeção das Visualizações	22
3.5	Resultados	22
3.5.1	Q1. Quais fatores influenciaram as decisões no processo de criação das visualizações?	23
3.5.2	Q2. Quais desafios encontrados no processo de criação das visualizações?	28
3.5.3	Resultados da Inspeção das Visualizações por Especialistas	30
3.6	Discussão dos Resultados	33
3.6.1	Diferenciação entre as visualizações	33
3.6.2	Mapeamento de variáveis para visualização	33
3.6.3	Escolha de visualizações adequadas	34
3.6.4	Base para decisões	34
3.7	Considerações sobre o Estudo de Observação	35
4	VIS-SeMa	37
4.1	Requisitos para a Elaboração do Material	37
4.1.1	Apoiar Atividades Específicas	38
4.1.2	Apoiar Especialistas em um Domínio, mas Novatos em Visualização	39
4.1.3	Ser Baseado em <i>Templates</i> de Questões Comumente Associadas a Tipos de Visualização	39
4.1.4	Representar Inicialmente Visualizações Comumente Utilizadas	39
4.1.5	Apresentar exemplos concretos	40
4.1.6	Seguir Princípios de Design de Visualização	41

4.2	Organização do VIS-SeMa	42
4.3	Processo de Design de Visualizações utilizando VIS-SeMa	44
4.4	Como Expandir o VIS-SeMa	46
4.5	Considerações	49
5	Estudo de viabilidade do VIS-SeMa	52
5.1	Contexto e Participantes do Estudo	52
5.2	Materiais do Estudo	53
5.3	Etapas do Estudo	54
5.4	Análise dos resultados	54
5.4.1	Análise Quantitativa	55
5.4.2	Análise Qualitativa	56
5.4.3	Análise de Discrepâncias	59
5.5	Discussão dos Resultados	70
5.5.1	Q1. De que forma o material VIS-SeMa apoiou a seleção de visualizações e o mapeamento de variáveis para visualização?	70
5.5.2	Q2. Quais problemas surgiram durante o processo de seleção e mapeamento de variáveis?	71
5.6	Considerações sobre o Estudo de Viabilidade	72
6	Considerações Finais e Trabalhos Futuros	73
6.1	RQ1. Quais decisões são tomadas durante a construção de visualização de informações?	73
6.2	RQ2. Quais dificuldades/problemas são encontrados durante o processo de design de visualização de informações?	74
6.3	RQ3. Como apoiar novatos na melhoria da criação de visualização de informações?	75
6.3.1	Design de Estudos para Avaliar a Criação de Visualizações	76
6.3.2	Diretrizes para a Inspeção de Mapeamentos de Variáveis	77
6.4	Trabalhos Futuros	78
	Referências bibliográficas	79
A	Material do Estudo 1	84
A.1	Tabela com visualizações escolhidas pelos participantes estudo 2	84
B	Material do Estudo 2	85
B.1	Questionário utilizado no estudo 2	85
B.2	Caracterização dos participantes do Estudo de Viabilidade	87
C	Materiais de apoio ao uso da Proposta	88
C.1	Exemplo 1 - Utilizando VIS-SeMa	88
C.2	Exemplo 2 - Inserindo novas visualizações no VIS-SeMa	89

Lista de Figuras

Figura 1.1	Etapas da pesquisa	5
Figura 3.1	Grau de conhecimento dos participantes sobre as Visualizações utilizadas no estudo	16
Figura 3.2	Quadro utilizado pelos participantes para criar as visualizações	17
Figura 3.3	Tabela de dados utilizada no estudo	18
Figura 3.4	Exemplo para questão: Quais os gastos da cidade ao longo dos últimos 5 anos? (Visualização elaborada por P3)	24
Figura 3.5	Exemplo para questão: <i>Quais os gastos da cidade para os últimos 5 anos?</i> (Visualização elaborada por P1)	25
Figura 3.6	Exemplo para questão: <i>Quais setores do serviço público receberam mais investimentos no ano?</i> (Visualização elaborada por P1)	26
Figura 3.7	Exemplo para questão: <i>Quais os dois meses do ano que tiveram mais gastos?</i> (Visualização elaborada por P4)	26
Figura 3.8	Exemplo para questão: <i>Quais os gastos da cidade ao longo dos últimos 5 anos?</i> (Visualização elaborada por P5)	27
Figura 3.9	Exemplo para questão: <i>Quais os dois meses do ano que tiveram maiores gastos?</i>	28
Figura 3.10	Exemplo para questão: <i>Q3. Qual região teve mais gastos no ano?</i>	29
Figura 3.11	Quantidade de problemas encontrados (eixo vertical) para cada diretriz (eixo horizontal)	31
Figura 3.12	Exemplo para a questão: <i>Quais setores do serviço público receberam mais investimentos no ano?</i>	31
Figura 3.13	Exemplo para a questão: <i>Q1. Quais os gastos da cidade ao longo dos últimos 5 anos?</i>	32
Figura 3.14	Exemplo para a questão: <i>Q2. Quais os dois meses do ano que tiveram maiores gastos?</i>	32
Figura 4.1	VIS-SeMa - Requisitos do Material proposto	42
Figura 4.2	VIS-SeMa - Visão geral da estrutura do material proposto	43
Figura 4.3	VIS-SeMa - Etapas para ampliar o material para outras visualizações	47
Figura 4.4	VIS-SeMa - Página 2 do Material onde são organizados os templates e as categorias	48
Figura 4.5	VIS-SeMa - Página 3 do material: Conteúdo - Nome da Visualização e Função para gráfico de barras e gráfico de barras empilhadas	49
Figura 4.6	VIS-SeMa - Página 3 do material: Conteúdo - Anatomia para gráfico de barras e gráfico de barras empilhada	50
Figura 4.7	VIS-SeMa - Página 3 do material: Conteúdo - Exemplos para gráfico de barras e gráfico de barras empilhadas	51

Figura 5.1	Frequência de utilização de cada elemento do VIS-SeMa	55
Figura 5.2	Respostas dos participantes sobre quanto os elementos do material ajudaram nas atividades de escolha do gráfico e no mapeamento das variáveis	56
Figura 5.3	Classificação das visualizações criadas	60
Figura 5.4	Classificação dos mapeamentos criados	61
Figura 5.5	Exemplo 1 - elementos da visualização sem mapeamento de variáveis	62
Figura 5.6	Exemplo 2 - elementos da visualização sem mapeamento de variáveis	62
Figura 5.7	Exemplo 1 para variáveis necessárias que não foram mapeadas na visualização	63
Figura 5.8	Exemplo 1 para variáveis necessárias que não foram mapeadas na visualização	64
Figura 5.9	Exemplo 2 para variáveis necessárias que não foram mapeadas na visualização	64
Figura 5.10	Exemplo 1 para inversão de variáveis	65
Figura 5.11	Exemplo 2 para inversão de variáveis	65
Figura 5.12	Exemplo 1 mapeamento de duas variáveis para o mesmo elemento do gráfico	66
Figura 5.13	Exemplo 1 para mapeamento de uma variável para dois elementos da visualização	67
Figura 5.14	Exemplo 1 para utilização incorreta da visualização	68
Figura 5.15	Exemplo 1 para informações sem contexto	68
Figura 5.16	Exemplo 1 para interpretação diferente do que a questão queria dizer	69
Figura 5.17	Exemplo 1 para exemplo de escolha inadequada de visualização	70
Figura 6.1	VIS-SeMa - Instruções de uso do material - versão atualizada após estudo	75
Figura 6.2	VIS-SeMa - Design de Experimento para avaliar o uso de visualizações	77
Figura B.1	Caracterização dos participantes do estudo de viabilidade	87
Figura C.1	Categoria e templates de questões inseridos no VIS-SeMa	90
Figura C.2	Templates de questões e visualizações inseridos no VIS-SeMa	91
Figura C.3	Exemplos de questões e visualizações inseridos no VIS-SeMa	91

Lista de Tabelas

Tabela 3.1	Experiências dos Participantes do Estudo	15
Tabela 3.2	Exemplo de mapeamento de passos extraídos dos vídeos	21
Tabela 4.1	Templates definidos para cada visualização	40
Tabela A.1	Visualizações escolhidas pelos participantes para cada questão	84

Lista de Abreviaturas

InfoVis – *Information Visualization* (Visualização de Informação)

LI – *Lean Inception*

ML – *Machine Learning* (Aprendizado de Máquina)

PBB - *Product Backlog Building*

UX – *User Experience* (Experiência do Usuário)

1

Introdução

As áreas de ciência de dados e visualização de informação têm trabalhado para ajudar pessoas a entender seus dados, oferecendo abordagens complementares (ferramentas de visualização, plataformas e bibliotecas) para este objetivo (Crisan et al., 2020). Visualizações apoiam o trabalho com dados complexos e em grandes quantidades (Stoiber et al., 2022). Tratam-se de representações visuais (gráficos e mapas, por exemplo) utilizadas para entender, comunicar e apoiar engajamento e tomada de decisões com dados e informações de forma sumarizada (Park et al., 2022; Eberhard, 2023). Esta ferramenta é essencial na análise exploratória de dados (exploração de estruturas de dados, detecção de *outliers* e grupos incomuns, identificação de tendências e *clusters*, identificação de padrões locais, avaliação de saída de modelagem e comunicação de resultados) e ajuda analistas a verificarem a qualidade dos dados e se familiarizarem com a estrutura e os recursos dos dados diante deles (Unwin, 2020).

Crisan et al. (2020) realizaram uma revisão da literatura que descreve a ciência de dados em quatro etapas: preparação dos dados, análise, implantação e comunicação. Dentro de cada uma destas etapas identificaram o papel das visualizações. Na preparação de dados, as visualizações apoiam a análise de qualidade e adequação dos dados. Na etapa de análise, as visualizações ajudam na exploração, modelagem e interpretação de dados. A implantação pode ser apoiada pelas visualizações através do uso de painéis (*dashboards*), que monitoram e ajudam a identificar eventos anômalos. Por fim, a etapa em que as visualizações são consideradas mais importantes é a de comunicação, pois as visualizações fornecem uma interface de informações entre humanos e máquinas, esclarecendo decisões e ajudando na identificação de métricas. Neste trabalho, iremos tratar a etapa de análise, apoiando a seleção e criação de visualizações adequadas a certas necessidades de informação.

Vários aspectos de modelos de visualização têm sido explorados (interação, análise individual e análise em equipe), mas o entendimento de como as pessoas constroem mapeamentos visuais é pouco explorado (Parsons, 2021; Ge et al., 2024; Rezaie et al., 2024) e a importância do letramento em visualização ainda não é destaque nas comunidades de pesquisa (Firat et al., 2022a), apesar

de alguns avanços recentes. Além disso, principalmente no primeiro contato, usuários com baixo letramento em visualização geralmente têm dificuldades em ler e interpretar gráficos (Stoiber et al., 2023; Ge et al., 2023). Neste trabalho focamos em *design* de visualizações. No entanto, é importante citar que as atividades de interpretação e design de visualizações são relacionadas, pois a falta de entendimento sobre o mapeamento de atributos de dados em visualizações pode influenciar no seu design final, possivelmente levando a conclusões inadequadas sobre os dados (Grammel et al., 2010; Vázquez-Ingelmo et al., 2024). Desta forma, é importante investigar formas de apoiar novatos na construção e interpretação de visualizações, além de aumentar seu letramento sobre visualização de informação.

Esta pesquisa tem por objetivo apoiar pessoas que possuem dúvidas e dificuldades no design de visualizações (novatos em visualizações). Para atingir este objetivo, utilizamos estudos empíricos para investigar como novatos criam visualizações e propor formas de apoiar esta atividade. no estudo da atividade de criação (design) de visualizações, investigando desafios e decisões que ocorrem durante esta etapa. Como resultados, identificamos bases para tomada de decisões e dificuldades apresentadas por novatos. A partir desses resultados, elaboramos um material didático para apoiar novatos na seleção de visualizações e mapeamento de atributos nas visualizações. O material é composto por módulos que ajudam a projetar uma visualização utilizando *templates* e exemplos concretos de questões e visualizações. O material elaborado foi útil para apoiar novatos no design de visualizações, pois a maioria das visualizações criadas com uso do material foram selecionadas e construídas corretamente. Além disso, observamos que uma boa forma de apoiar design de visualizações é utilizando *templates* genéricos de questões e visualizações, exemplos de uso concretos, e conteúdos visuais.

1.1

Motivação

Em trabalho anterior (Ferreira et al., 2023), detalhamos o primeiro estudo desta pesquisa, no qual investigamos práticas e dificuldades ocorridas ao longo de projetos de ciência de dados. Entre os resultados desse estudo, identificamos que uma das formas mais utilizadas para representar soluções e comunicar requisitos entre a equipe foram protótipos, na maioria dos casos, de *dashboards* de visualizações. Durante a condução de *workshops* para construção destes protótipos, as visualizações e os dados eram pouco explorados, sendo necessário o apoio de especialistas em visualização para conclusão dos protótipos. No entanto, embora seja importante, nem todas as equipes possuem especialistas

em visualização de informação, tornando-se necessário que novatos¹ adquiram conhecimento suficiente para criar visualizações nos seus protótipos.

Neste contexto, a motivação desta pesquisa é fornecer subsídios para elaboração e melhoria de materiais que possam apoiar o design de visualização por novatos. A partir dos resultados, esperamos contribuir para o ensino de visualização de informação e o design de visualizações em projetos na indústria de software.

1.2

Definição do problema

Com a popularização do uso de visualizações e ferramentas de análise de dados, torna-se necessário que o público em geral e não somente especialistas possam construir e interpretar visualizações com seus próprios dados (Huron et al., 2014). Vários processos e taxonomias têm sido propostos para apoiar o design de visualizações (Keck et al., 2020). No entanto, essas abordagens têm descrição em alto nível, ou seja, fornecem um entendimento geral, mas não apoiam seu uso na prática, não apresentando detalhamentos sobre os passos a serem seguidos para mapear variáveis em visualizações (Keck et al., 2020), exigindo algum grau de aprendizagem e compreensão dos mapeamentos visuais, convenções gráficas, símbolos, marcas, *layouts* e padrões para decodificar adequadamente a representação (Wang et al., 2020). Além disso, vários aspectos de modelos de visualização têm sido estudados (interação, análise individual e análise em equipe), mas o entendimento de como as pessoas constroem mapeamentos visuais é pouco explorado. (Parsons, 2021; Ge et al., 2024; Rezaie et al., 2024).

Neste contexto, investigamos o design de visualizações de informação realizado por novatos com objetivo de fornecer insumos para que novatos possam selecionar e criar visualizações de forma mais adequada, ou seja, minimizando erros e dificuldades nestas atividades. Neste contexto, definimos as seguintes questões de pesquisa para entender o processo de design de visualizações e quais melhorias poderiam ser aplicadas:

RQ1. Quais decisões são tomadas durante a construção de visualização de informações?

RQ2. Quais dificuldades ou problemas são encontrados durante o processo de design de visualização de informações?

¹Neste texto utilizamos o termo *novatos* para nos referirmos à inexperiência somente na área de visualização de informação. Eles podem ser muito experientes em outras áreas de conhecimento e atuação, inclusive computação.

RQ3. Como apoiar novatos na melhoria da criação de visualização de informações?

1.3

Metodologia e contribuições esperadas

Para responder as questões de pesquisa deste trabalho, adotamos uma metodologia baseada em evidências para definição e avaliação de novas tecnologias de software semelhante a proposta por (Shull et al., 2001). Essa metodologia propõe a condução de estudos experimentais para avaliar o funcionamento de uma tecnologia desde de sua definição até sua transferência para a indústria. Na metodologia proposta por Shull et al. (2001) é definido um processo iterativo, onde podem ser realizados quatro tipos de estudos: estudos de viabilidade, estudos de observação, estudos em ciclo de vida e estudos e estudos de caso na indústria. Estudos de viabilidade são realizados para avaliar a utilidade prática da aplicação das tecnologias definidas. Os estudos de observação são aplicados para aprimorar os conhecimentos dos pesquisadores sobre processos de trabalho. Estudos em ciclo de vida examinam a utilização da tecnologia proposta em um ciclo de vida de desenvolvimento de software. O estudos de casos na indústria visam avaliar o uso das tecnologias propostas na indústria de software. Na metodologia proposta nesta pesquisa de tese, utilizamos dois destes tipos de estudo: estudo de observação e estudo de viabilidade. A Figura 1.1 mostra as etapas desta metodologia.

A primeira etapa (etapa 1) da metodologia utilizada nesta pesquisa foi a realização de um **estudo de observação**. Este estudo teve por objetivo entender o processo de design de visualizações por cinco profissionais da indústria de software (representados no texto como P1 a P5) mas sem experiência em visualizações de informação (novatos em visualização). Para a coleta de dados, utilizamos o método de observação e entrevista. Apesar de demandar maior tempo, o método foi escolhido para que pudessem ser coletadas evidências mais detalhadas sobre as atividades realizadas, bases para decisões e dificuldades encontradas durante a criação de visualizações. Esse estudo é apresentado no Capítulo 3.

Após os resultados do primeiro estudo, propusemos o material didático VIS-SeMa (etapa 2), para apoiar novatos na criação de visualizações. O material é composto por explicações textuais e visuais e visa a auxiliar na seleção de visualizações e mapeamento de variáveis. Além do material, disponibilizamos um guia para expansão do material. O detalhamento da proposta é apresentado no Capítulo 4.

Para avaliar a aplicação prática do material proposto, realizamos um

estudo (estudo 3) de viabilidade com o material didático. Os participantes do estudo foram 28 alunos (representados no texto como P1 a P28) do curso de Ciência da Computação, novatos em visualização. O método utilizado neste estudo foi um estudo em sala de aula (presencial); os resultados foram coletados através de análises do material produzido e questionários após a atividade. Esse estudo é apresentado no Capítulo 5.

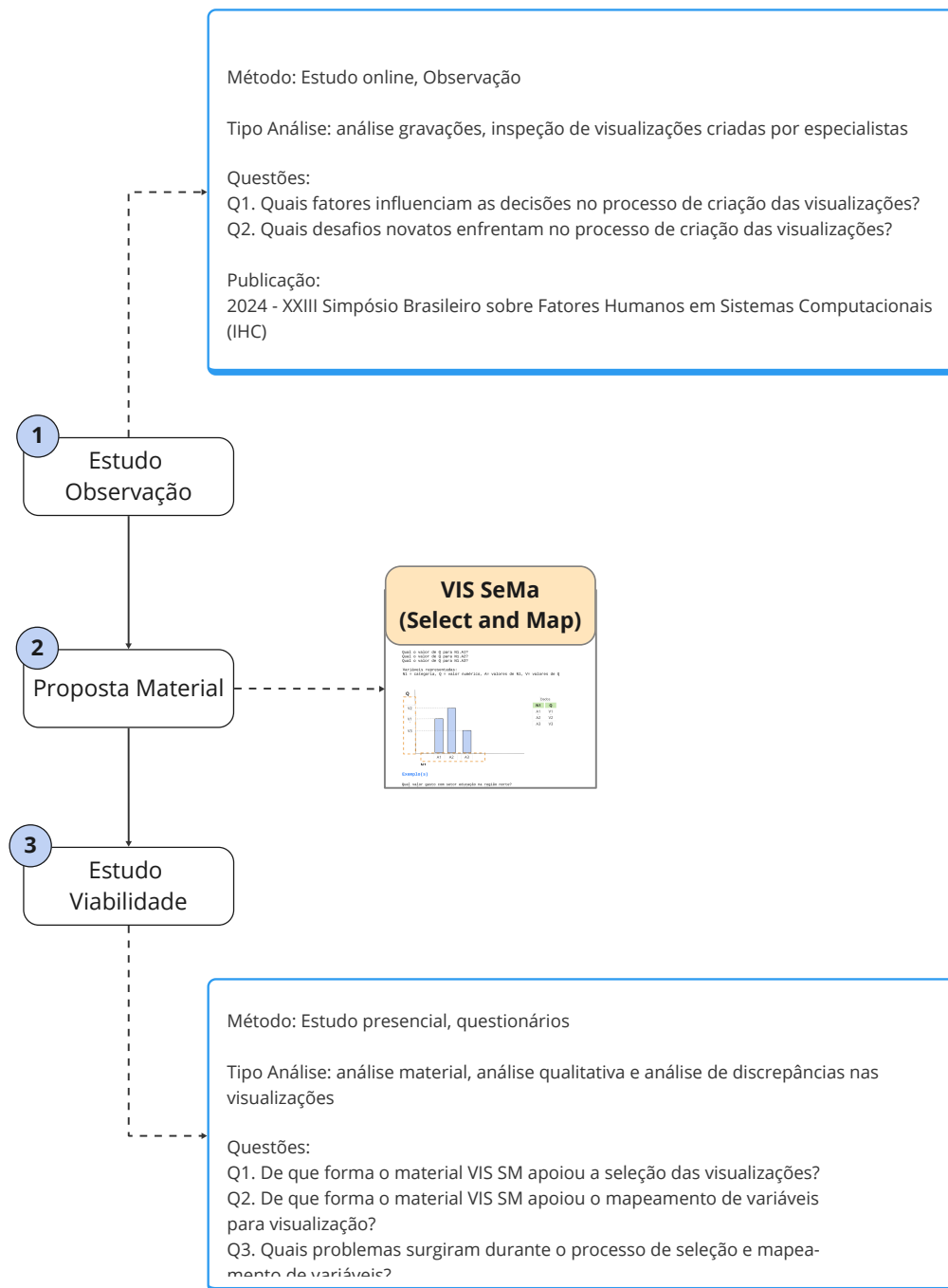


Figura 1.1: Etapas da pesquisa

A partir dos resultados obtidos através da execução dessas etapas, esperamos apoiar novatos na escolha de visualizações mais adequadas e na criação de visualizações mais corretas. Desta forma, buscamos proporcionar maior independência na construção de visualizações para pessoas com pouca experiência. Além disso, através da criação de visualizações mais adequadas pode ser melhorada a interpretação dos dados.

1.4

Organização deste Documento

O restante deste documento está organizado da seguinte maneira: o Capítulo 2 apresenta conceitos e trabalhos relacionados que fundamentam esta pesquisa. Em seguida, o Capítulo 3 apresenta um estudo que investiga o processo de criação de visualizações. O Capítulo 4 apresenta o material proposto, seguido de um estudo de viabilidade para avaliá-lo (Capítulo 5). Por fim, no Capítulo 6 apresentamos as considerações finais da pesquisa e trabalhos futuros.

2

Fundamentação

Neste capítulo, apresentamos conceitos que fundamentam esta pesquisa de tese. Apresentamos o conceito de novatos no design de visualizações – que representa o público alvo a ser apoiado por este trabalho – e conceitos de letramento em visualização. Além disso, discutimos sobre o processo de design de visualizações e apresentamos alguns trabalhos relacionados identificados na literatura que apresentam estudos sobre a criação de visualizações, destacando obstáculos e aspectos que podem apoiar o entendimento deste processo.

2.1

Novatos em Visualização de Dados

Neste trabalho, realizamos uma pesquisa voltada para apoiar novatos em visualizações. No entanto, não realizamos estudos para definir quem seriam os usuários novatos. Neste contexto, utilizamos por base trabalhos da literatura para caracterizar os novatos nesta pesquisa. A seguir apresentamos uma discussão sobre esse de público alvo.

Burns et al. (2023) analisaram 79 trabalhos de visualização para identificar quais características definem um novato. Em seus resultados, observaram que vários termos são utilizados para definir novatos: não experientes, leigos, novatos e público em geral. Identificaram ainda algumas características utilizadas para definir novatos: área profissional, proximidade com áreas de pesquisa específicas, falta de conhecimentos, experiência ou habilidades. Além disso, discutiram algumas barreiras que novatos apresentam ao construir visualizações: como mapear valores para as representações visuais, nomear e interpretar algumas visualizações e ter dificuldades quando o modelo mental de uma visualização não se mostra útil.

Lee et al. (2016) definem novatos como usuários que têm contato com determinada visualização pela primeira vez. Nesse caso, consideram pessoas como novatos em visualizações específicas. Como optamos por trabalhar com várias visualizações, não utilizamos o conceito de Lee et al. (2016) nesta pesquisa. Grammel et al. (2010) definem novatos como usuários que criam visualizações para suas tarefas básicas, mas que tipicamente não têm conhecimentos avançados em análise de dados, visualização de dados e estatística. Nesta pesquisa

de tese, os participantes dos estudos se encaixam nesta última definição, pois possuem conhecimentos em visualizações para atividades simples de trabalho ou na sua rotina, mas não têm especialidade ou conhecimentos aprofundados em análise de dados, visualizações e estatística. Nesse caso, dizemos então que possuem baixo ou nenhum letramento em visualização, conforme definido na seção a seguir.

2.2

Letramento em Visualização de Dados

O Letramento em visualização é uma habilidade essencial para a compreensão e interpretação de representações visuais de dados (Firat et al., 2022b). Boy et al. (2014) definem letramento em visualização como a habilidade de criar e interpretar representações visuais de dados. Segundo Lee et al. (2016), letramento em visualização é a habilidade de ler e interpretar representações visuais de dados representados e extrair informações de visualizações. Embora seja importante, o letramento em visualização ainda não é destaque nas comunidades de pesquisa (Firat et al., 2022a), apesar de alguns avanços recentes. Além disso, principalmente no primeiro contato, usuários com baixo letramento em visualização geralmente têm dificuldades em ler e interpretar gráficos (Stoiber et al., 2023; Ge et al., 2023).

O estudo de Boy et al. (2014) estabelece uma metodologia para avaliar o nível de letramento das pessoas, relacionando as implicações do nível de letramento para pesquisa, design e educação sobre visualizações. Buscando entender o impacto do letramento em visualizações na capacidade de interpretação de dados de indivíduos, Maltese et al. (2015) propuseram uma ferramenta analítica para diferenciar os dois grupos, apresentando as propriedades psicométricas que a fundamentam. Em geral, estudos recentes focam em criar artefatos para mensurar o nível de letramento em visualização, com menos foco em fazer recomendação para aperfeiçoar o ensino de visualizações.

Olhando para o histórico da área, conseguimos identificar trabalhos que também podem ser proveitosos e que serviram de fundamentação para o campo de letramento em visualização como um todo. Card et al. (1999) discutiram o papel do processamento visual de informação na cognição externalizada, fundamento que torna o uso de visualizações de informação como uma estratégia interessante para um melhor entendimento dos dados. Evidentemente, é importante considerarmos os diferentes aspectos da compreensão visual quando discutimos os erros que podem ocorrer na interpretação de gráficos, levando a enganos sobre os dados sendo visualizados. Há também diferentes propostas de sistemas que buscam a construção automática de visualizações com base em

diferentes formatos de dados, como a proposta seminal de Mackinlay (1986). Mesmo que o foco deste trabalho não seja a construção automatizada de gráficos, ainda é possível aprender lições das contribuições teóricas que foram identificadas com a construção e uso dessas ferramentas, como por exemplo, a dependência contextual do mapeamento de dados em atributos visuais.

2.3

Processo de design de visualizações

O design de visualizações envolve o desenvolvimento de um conceito da visualização, onde podem ser elaborados e refinados protótipos (em papel ou em ferramentas) e é realizado o mapeamento das variáveis representadas nos dados (Walny et al., 2019). Segundo Qin et al. (2020), especificar visualizações envolve o trabalho com três elementos: dados, marcações e o mapeamento entre eles. Os dados são as variáveis a serem representadas e as operações que transformam os dados. As marcações são os tipos de representações visuais (barras, linhas, pontos), tamanho (altura, largura), legendas, variações (cores). O mapeamento é a associação entre os dados e as marcações (chamadas neste trabalho de elementos das visualizações). Ao mapear atributos em dados, os designers da visualização utilizam sua percepção para ajudar usuários (leitores da visualização) a discernir e interpretar padrões dentro dos dados (Heer et al., 2012). Desta forma, é importante criar visualizações adequadas para facilitar sua interpretação.

Atualmente, o esforço para criação de visualizações tem sido reduzido através do uso de ferramentas que automatizam a seleção de modelos prontos (Fernandez et al., 2024). As ferramentas são essenciais para agilizar o processamento, análise e comunicação dos dados (Lavanya et al., 2023). Por outro lado, o processo de mapeamento de variáveis em elementos visuais se torna uma caixa preta, ou seja, a criação da visualização que inclui as etapas de seleção e associação entre dados e elementos das visualizações não fica claro para novatos (Fernandez et al., 2024). Alguns trabalhos relacionados ao design de visualizações são apresentados na próxima Seção.

2.4

Estudos sobre o Design de Visualizações

Grammel et al. (2010) apresentaram um estudo em que investigaram como novatos constroem visualizações. No processo de criação das visualizações, os participantes criavam uma prototipação da visualização, analisavam suas criações e faziam refinamentos. O estudo mostrou que novatos em visualização de informações escolhem visualizações de acordo com heurísticas e com sua

familiaridade com os tipos de visualização. As visualizações mais escolhidas foram: gráfico de barras, gráfico de linhas, gráfico de pizza e mapas. Neste estudo, alguns participantes relataram algumas dificuldades: traduzir questões em atributos, criar visualizações adequadas e interpretar as visualizações.

O estudo de Baker et al. (2001) foi realizado com estudantes do ensino fundamental. Eles deveriam interpretar visualizações, selecionar a representação apropriada para diferentes tipos de exibição de dados e gerar diferentes representações. Os resultados do estudo mostraram que os estudantes foram moderadamente bem na interpretação dos gráficos (56% de acertos). Na seleção de visualizações, os estudantes não tiveram resultados bons (25% de acertos). Em relação à geração de representações, os participantes tiveram desempenho baixo: nenhum estudante criou representações totalmente corretas.

Keck et al. (2020) propuseram uma metodologia para facilitar a criação de visualizações. Sua metodologia apresenta elementos para compor uma visualização e ajudar a comunicar decisões para um grupo. Foi desenvolvida para estudantes e é composta por três etapas: (1) análise morfológica com construção de blocos, (2) seleção/fusão e (3) melhorias. Na etapa (1) é realizada uma ideação para a construção de visualizações, na qual os participantes dispõem de um *kit* composto por blocos: grades, estruturas de *layout* e interações. Esses blocos devem ser combinados para criar uma ou vários padrões de visualizações. Na etapa (2), os padrões criados são avaliados para escolher a melhor opção ou criar uma nova visualização a partir da combinação dos padrões criados na etapa (1). Na etapa (3), são realizadas melhorias estéticas nas visualizações. A metodologia foi aplicada em cursos de visualização de informação. Como resultados, descobriram que alguns elementos nos blocos não foram distinguidos (ou seja, foram considerados semelhantes) pelos participantes (por exemplo: ângulo e área). Outro ponto foi que, por utilizarem uma linguagem comum para construir as visualizações, os participantes tiveram facilidade em entender os trabalhos uns dos outros. A utilização da metodologia foi facilitada devido aos exemplos de uso.

Çay et al. (2020) propuseram um *workshop* (ColVis) para apoiar o processo de tomada de decisão colaborativa para criação de visualizações. O workshop foi avaliado no contexto de quatro projetos, cada qual constituído por quatro fases: (i) definição de usuários e objetivos (espaço do problema), (ii) definição de perguntas e tarefas (espaço do problema), (iii) conexão de dados e questões (espaço de solução) e (iv) criação de visualização (espaço de solução). Algumas lições aprendidas nesta pesquisa foram: há dificuldade em manter o foco nas visualizações em dinâmicas no formato de *workshop*; e criar visualizações colaborativamente com pessoas de vários níveis de conhecimento é

uma boa opção para identificar as tarefas mais importantes para a visualização.

A partir da análise destes trabalhos, é possível identificar alguns problemas na criação de visualizações: traduzir questões em atributos, criar visualizações adequadas, interpretar as visualizações e diferenciar elementos que compõem as visualizações. Também foram identificados alguns aspectos não classificados como problemas ou dificuldades, mas que podem ajudar no entendimento do processo de criação das visualizações: as escolhas de visualizações foram baseadas em heurísticas ou na familiaridade dos participantes e o uso de exemplos facilitou a utilização das metodologias que apoiam o design das visualizações. Além disso, foi identificado um formato de criação de visualizações de forma colaborativa, que pode ser uma boa opção para apoiar na definição dos objetivos das visualizações a serem criadas.

2.5

Materiais didáticos para apoio ao design de visualizações

Neste trabalho elaboramos um material para apoiar novatos no design de visualizações, focando nas atividades de seleção de visualizações e mapeamento de variáveis em elementos da visualização. A seguir, são apresentados alguns trabalhos que foram utilizados como base para a construção da proposta.

Wang et al. (2020) elaboraram um material para apoiar a criação de materiais de ensino, aprendizagem e uso de visualizações por público-alvo variado (novatos e especialistas). Seu material é baseado em explicações visuais, histórias em quadrinhos e possui seções explicativas sobre um tipo específico de visualização. As seções foram identificadas a partir de um estudo com professores e alunos de visualizações e ciência de dados: anatomia, construção, padrões visuais, armadilhas, falsos amigos e visualizações semelhantes. A anatomia explica os elementos visuais. A construção explica a ideia geral de uma visualização e seus padrões visuais. As armadilhas mostram erros comuns. Analogias relacionam as visualizações com uma representação mais comumente conhecida e falsos amigos mostram visualizações que parecem semelhantes, mas devem ser interpretadas de forma diferente. Os autores optaram por criar o material para representação de visualizações de construção não trivial: gráficos de coordenadas paralelas, matrizes de adjacência, gráficos de caixas (*boxplots*), *treemaps*, gráficos de confluência e curvas de tempo. O material foi proposto de acordo com princípios de design baseados em seus estudos e na literatura. Os princípios que eles compilaram foram:

- **Modularidade:** divisão em seções para explicar aspectos específicos da visualização;

- **Independência de Contexto:** Não especificar contexto e nem exemplos para facilitar a reutilização em qualquer contexto;
- **Visualizações Limpas:** visualizações representadas utilizam em maior parte linguagem visual e poucas anotações;
- **Neutralidade de Estilo:** não focar em características de um público alvo específico.

Wang et al. (2020) avaliaram o material através de um estudo com 11 participantes. O material foi considerado útil e fácil de usar para novatos e especialistas na construção de visualizações. No entanto, foram citadas duas limitações:

- 1: O estudo não avaliou o desempenho real proporcionado pelo material e quais seções foram mais importantes para a construção das visualizações.
- 2: O estudo não considerou contextos, desta forma seriam necessários estudos explorando domínios e exemplos concretos.

Rodrigues et al. (2021) investigaram quais questões as pessoas geralmente associam a visualizações específicas. Considerando 22 participantes elaboraram um total de 1058 questões associadas a 20 visualizações diferentes. A partir das análises, os autores derivaram *templates*, ou seja, padrões de questões para as visualizações. Os resultados do estudo ajudaram a identificar erros comuns cometidos na interpretação de visualizações e forneceram insumos para apoiar o ensino de visualização de dados. Além disso, forneceram *templates* para apoiar a construção de materiais didáticos.

As visualizações utilizadas por Rodrigues et al. (2021) foram: gráfico de barras (ordenado por categoria), gráfico de barras (ordenado por frequência), gráfico de barras (agrupadas), gráfico de barras (empilhadas), boxplot, heatmap, chord, Sankey, network, linhas (simples), linhas (múltipla), ridge, histograma, gráfico de dispersão, gráfico de dispersão (+ cor), gráfico de bolhas, gráfico de bolhas (+ cor), mapa (cartograma), mapa (coroplético) e tabela. Para padronização, as questões foram definidas como *templates* parametrizados. As regras para definição das questões são explicadas a seguir.

- Os nomes das variáveis das questões são representadas pela inicial dos nomes de seus tipos em inglês: N para variáveis nominais, Q para variáveis numéricas contínuas (quantitativas), T para variáveis temporais, S para variáveis espaciais e Objs para objetos;
- Algumas variações de questões também foram utilizadas como parâmetros. Exemplo: “Qual o maior valor?” e “Qual o maior valor?”, onde os parâmetros são “maior” e “menor”;

- Quando diferentes variáveis do mesmo tipo são representadas na mesma questão, devem ser diferenciadas com acréscimo de um número. Exemplo: três variáveis nominais devem ser representadas como N1, N2 e N3;
- Se aparecer na questão um valor específico para uma variável, deverá utilizar um ponto seguido de outras letras diferentes das letras dos tipos(N, Q, T, S). Exemplo: N= Estado, A1=Rio de Janeiro, A2=Amazonas; estado do Rio de Janeiro -> N.A1; estado do Amazonas -> N1.A2;
- Expressões entre colchetes indicam que podem estar presentes ou não na questão. Exemplo: o template “Qual [menor|maior] valor de Q?” pode ser uma das duas opções -> “Qual menor valor de Q?” e “Qual maior valor de Q?”.

O trabalho de Rodrigues et al. (2021) foi utilizado como base para guiar o conteúdo do material proposto nesta pesquisa. A contribuição deste trabalho para nossa pesquisa foi o uso dos *templates*. Associamos seus *templates* com visualizações genéricas para apoiar o mapeamento de variáveis na visualização. Em relação ao trabalho de Wang et al. (2020), utilizamos a sua estruturação em módulos e o conceito de visualizações limpas, ou seja, com poucas informações textuais. Além disso, nossa proposta contribui para superar problemas relatados por Stoiber et al. (2023), cujo material não apresentava exemplos concretos, como fizemos no material proposto nesta pesquisa.

2.6

Considerações

Este capítulo apresentou alguns conceitos e estudos relacionados ao design de visualizações. Inicialmente apresentamos o conceito de novatos em visualização e discutimos brevemente sobre letramento em visualização. Após os conceitos iniciais, apresentamos a fundamentação sobre design de visualizações de dados e trabalhos existentes na literatura relacionados a estudos em design de visualizações. Além disso, também foram apresentados alguns trabalhos que foram utilizados para embasar o material proposto nesta pesquisa. No próximo capítulo apresentamos um estudo que explora o processo de design de visualizações por profissionais da indústria novatos em visualizações.

3

Investigando o processo de criação de visualizações

Este capítulo apresenta um estudo de observação realizado para aprimorar o conhecimento dos pesquisadores sobre o processo de design de visualizações por novatos. Neste contexto foram analisadas as etapas realizadas neste processo e quais fatores influenciaram na escolha das visualizações. Este estudo foi guiado pelas seguintes questões de pesquisa:

- Q1. Quais fatores influenciaram as decisões no processo de criação das visualizações?
- Q2. Quais desafios novatos enfrentaram no processo de criação das visualizações?

3.1

Contexto e Participantes do Estudo

Em relação ao contexto, foi realizado um estudo de observação. O estudo foi realizado de forma online, via ferramenta Zoom.¹ As atividades de criação de visualizações foram realizadas utilizando o ambiente colaborativo Miro.² Durante todo o processo, o participante compartilhava sua tela com a pesquisadora autora desta tese. A pesquisadora apenas observava o processo sem interferir nas escolhas e sem fornecer explicações sobre visualizações. Apenas respondia dúvidas em relação ao uso do Miro. As atividades foram realizadas de forma individual para cada participante. Os cenários utilizados para o estudo foram criados com base na experiência no ensino e prática de design de visualização de dados da pesquisadora autora e sua orientadora. Durante as atividades, os participantes podiam fazer buscas na internet para tirar suas dúvidas. Não foi recomendada uma ferramenta específica para apoiar a escolha e design de visualizações, ou seja, não houve intervenção na forma como eles realizariam as atividades.

Foram convidados para participar do estudo desenvolvedores de quatro empresas diferentes que tinham alguma experiência em projetos de ciência de dados. O estudo foi realizado com cinco participantes com experiência

¹Ferramenta para reuniões online: <https://zoom.us/>.

²Ferramenta para realização de atividades em grupo de forma online: <https://miro.com>.

entre 2 e 6 anos na indústria de software. Todos os participantes eram desenvolvedores e tinham níveis diferentes de conhecimentos teóricos e práticos em visualização de informação (InfoVis). Para caracterizar o seu conhecimento com visualizações, eles responderam a três questões objetivas:

- 1. Conhecimentos teóricos sobre visualização de informação
- 2. Aplicação prática de técnicas de visualização de informação em projetos de software em sala de aula
- 3. Aplicação prática de técnicas de visualização de informação em projetos de software na indústria

Para cada item, os participantes deveriam indicar qual seu nível de experiência em uma escala de sete pontos, onde 1 significava Nenhuma Experiência/conhecimento e, 7, Muita Experiência/conhecimento. A caracterização dos participantes é apresentada na Tabela 3.1. Além da caracterização inicial, os participantes indicaram se tinham conhecimento sobre algumas das visualizações utilizadas no estudo – para cada visualização, se conhecem e já utilizaram, se conhecem mas nunca utilizaram ou se não conhecem. Esses graus de conhecimento são apresentados na Figura 3.1.

Tabela 3.1: Experiências dos Participantes do Estudo

ID	P1	P2	P3	P4	P5
Experiência na Indústria	3 anos	5 anos	2 anos	4 anos	6 anos
Conhecimentos Teóricos InfoVis	5	5	1	7	5
Prática InfoVis - Aula	5	4	1	4	4
Prática InfoVis - Indústria	4	4	1	7	5

Esta tese faz parte de um projeto de pesquisa submetido e aprovado previamente pela Câmara de Ética em Pesquisa da PUC-Rio (protocolo #97/2020). Antes da realização do estudo, todos os participantes receberam um termo de consentimento livre e esclarecido, no qual aceitaram participar do estudo e ter áudio e vídeo gravados. O termo esclarecia que esta pesquisa não avaliaria os participantes, que seu anonimato seria garantido e que eles poderiam recusar-se a participar da pesquisa ou interrompê-la a qualquer momento.

	P1	P2	P3	P4	P5
Line Chart					
Bar Chart					
Pie Chart					
Histogram					
Scatterplot					
Area Chart					
Stacked Area Chart					
Choropleth Map					
Treemap					
Bubble Chart					

conhece e já usou/aplicou	conhece mas nunca usou/aplicou	não conhece
---------------------------	--------------------------------	-------------

Figura 3.1: Grau de conhecimento dos participantes sobre as Visualizações utilizadas no estudo

3.2

Materiais do Estudo

Conforme explicado anteriormente, as etapas do estudo foram realizadas no formato online utilizando a ferramenta colaborativa Miro, onde foi criado um quadro para cada participante. Cada quadro era composto por um conjunto de visualizações que os participantes poderiam aproveitar, um conjunto de representações de mecanismos de interação sobre visualizações, um cenário textual, uma tabela de dados e um conjunto de seis questões a serem respondidas. Todos os materiais utilizados no estudo foram criados pela pesquisadora que conduziu o estudo e revisados por um segundo pesquisador especialista em visualização de dados. Os elementos do quadro são apresentados na Figura 3.2 e explicados a seguir.

1. **Conjunto de Visualizações:** Este conjunto foi selecionado com base no trabalho de Lee et al. (2016), escolhido por ter sido utilizado em estudos envolvendo compreensão de visualizações. O conjunto é composto pelas seguintes visualizações: gráfico de linhas, gráfico de barras, gráfico de barras empilhadas, gráfico de barras empilhadas 100%, gráfico de pizza, histograma, gráfico de dispersão (*scatterplot*), gráfico de área, gráfico de área empilhada, mapa coroplético, gráfico de bolhas e *treemap*. Para facilitar o trabalho dos participantes, imagens das visualizações foram disponibilizadas no quadro. Cada participante foi instruído a selecionar uma visualização, fazer uma cópia dela e realizar edições indicando como mapeariam cada atributo do dataset a cada variável visual, conforme julgassem correto para responder às questões propostas.
2. **Conjunto de Mecanismos de Interação:** Um conjunto de meca-



Figura 3.2: Quadro utilizado pelos participantes para criar as visualizações

nismos de interação sobre visualizações foi proposto por Keck e Groh (2019). Esse conjunto também foi empregado em estudos anteriores que exploraram a avaliação de criação de visualizações. Da mesma forma que as visualizações, imagens dos mecanismos de interação já estavam disponíveis no quadro, permitindo que os participantes fizessem cópias para sua utilização no estudo. O cenário utilizado é apresentado a seguir.

3. **Cenário:** O cenário visava a ajudar os participantes a entender o contexto no qual as visualizações estariam inseridas. Além disso, apresentava alguns personagens que utilizariam as visualizações.

“VisCity é uma cidade com poucos habitantes. É governada pelo prefeito que gerencia os problemas e os recursos utilizados para manter cada setor de serviços públicos: saúde, educação, meio ambiente, trânsito, lazer e cultura. Cada região da cidade (norte, sul, leste, oeste) possui um agente comunitário que cuida dos seus problemas.

O prefeito registra o dinheiro arrecadado (ganhos) pela cidade e o dinheiro gasto (gastos) para manter os setores de serviços públicos da cidade: saúde, educação, meio ambiente, trânsito, lazer e cultura.

Ator interessado: cidadãos 1 - Os cidadãos precisam acompanhar as movimentações financeiras realizadas no mês e no ano e fazer comparações entre períodos (exemplo: janeiro foi o mês de maior gasto, 2021 teve menos ganhos do que em 2022) através do site de ganhos e gastos da cidade.

Ator interessado: prefeito 2 - Para melhor distribuição dos recursos, o prefeito precisa saber quais regiões (Norte, Sul, Leste, Oeste) precisam de recursos e para quais serviços (exemplo: o Sul precisa de recursos para a saúde, o Norte precisa de mais recursos de lazer e cultura do que o Leste).”

4. **Tabela de Dados:** Uma pequena tabela com dados para que os participantes conhecessem as variáveis a serem utilizadas. A tabela não apresentava um conjunto de dados reais, apenas continha exemplos de dados fictícios descritos no cenário. A Figura 3.3 apresenta a tabela de dados utilizada no estudo.

Exemplos de dados de ganhos e gastos

setor	região	gastos	ganhos	mês/ano
saúde	norte	300.000,23	400.000,18	12/2022
lazer e cultura	norte	50.000,16	100.000,10	12/2022
educação	norte	50.000,16	100.000,11	12/2022
meio ambiente	norte	300.000,23	100.000,12	12/2022
trânsito	norte	200.000,04	400.000,18	12/2022
segurança	norte	100.000,05	100.000,14	12/2022
saúde	norte	300.000,23	200.000,04	01/2023
lazer e cultura	norte	100.000,07	50.000,16	01/2023
...	...	...	...	...

Figura 3.3: Tabela de dados utilizada no estudo

5. **Questões:** No quadro também havia um conjunto de seis questões a serem respondidas. Para cada questão o participante deveria criar uma visualização para ajudar a respondê-la. As questões são apresentadas a seguir:

- Q1. Quais os gastos da cidade ao longo dos últimos 5 anos?
- Q2. Quais os dois meses do ano que tiveram maiores gastos?
- Q3. Qual região teve mais gastos no ano?
- Q4. Quanto foi gasto com cada setor de serviço em cada mês do ano?

Q5. Quais setores de cada região mais gastam?

Q6. Quais setores do serviço público receberam mais investimentos no ano?

3.3

Etapas do Estudo

Antes da realização do estudo, realizamos um estudo piloto para verificar se o entendimento dos materiais estava adequado. Para iniciar a criação das visualizações, os participantes receberam instruções do pesquisador que conduziu o estudo sobre as atividades a serem realizadas, sobre como utilizar a ferramenta Miro e informou que poderiam realizar buscas na internet caso tivessem dúvidas. Nessa etapa, o pesquisador pediu que o participante acessasse o quadro no Miro e testasse as opções de copiar e colar uma visualização para perto da questão a ser respondida e explicou como editar as visualizações para adicionar títulos e variáveis. Além disso, o cenário e tabela de dados foram lidos junto com o participante para que ele pudesse tirar dúvidas. As questões também foram lidas em conjunto, mas não foram explicadas aos participantes, pois a interpretação da questão deveria ser feita pelo participante.

Todos os participantes foram informados que tinham permissão para realizar pesquisas na internet para esclarecer dúvidas ou utilizar ferramentas familiares para auxiliá-los no processo. Apesar dessa liberdade, nenhum deles optou por realizar pesquisas específicas sobre ferramentas de visualização. Somente foram realizadas pesquisas na internet. As buscas realizadas na internet foram sobre diferença entre histograma e gráfico de barras, como e para quê usar algumas visualizações (histograma e *treemap*), além de buscas por exemplos e imagens de visualizações. Destacamos que, durante o estudo, não houve carregamento de dados e nem geração de visualizações com uso de ferramentas. O material gerado era um protótipo não funcional, mostrando um planejamento de como a visualização poderia ser implementada futuramente (a implementação também não fez parte desta pesquisa).

Após as instruções, os participantes utilizaram o quadro para selecionar as visualizações que julgassem mais adequadas para responder cada questão. Para isso, o participante analisava o conjunto de visualizações disponibilizado, selecionava uma visualização, copiava e colava para perto da questão que queria responder e editava a visualização escolhida, indicando quais variáveis eram representadas em cada elemento da visualização. Ao finalizar a criação de suas visualizações, o participante explicava seu raciocínio para o pesquisador, ou

seja, por que achou que a sua escolha era a mais adequada e explicava suas dúvidas e dificuldades e o que quis representar em cada visualização.

Na próxima etapa, os participantes revisitavam as visualizações criadas e decidiam se deveriam adicionar mecanismos de interação para melhorá-las. Nesse momento, também decidiam se mantinham as visualizações da forma que estavam ou se selecionariam outras visualizações para melhorar a resposta às questões. Essa etapa de decisão foi importante para permitir que os participantes revisassem cuidadosamente as questões apresentadas e as visualizações selecionadas inicialmente, corrigindo eventuais equívocos cometidos. Permitiu ainda aos pesquisadores compreender quais fatores influenciaram as escolhas nas substituição ou inclusão de recursos adicionais. Após esta etapa, novamente os participantes explicaram seus raciocínios e decisões e expuseram suas motivações para selecionar outras visualizações (quando era o caso) e adicionar mecanismos de interação. Como citado anteriormente, com os devidos consentimentos, todas as sessões foram gravadas (áudio e vídeo) e transcritas (falas e comportamentos dos participantes).

Todas as etapas do estudo foram realizadas individualmente para cada participante. As sessões duraram em torno de 3 horas. Ao final do estudo, foram geradas 63 visualizações no total. Todo processo gravado e as visualizações geradas foram organizados para análise posterior.

3.4

Análise dos resultados

Após a coleta de dados, realizamos duas análises: (1) análise do processo de design (criação) e (2) inspeção de visualizações. As etapas de cada análise são descritas a seguir.

3.4.1

Análise do Processo de Design

No contexto deste estudo, quando falamos em processo de design, estamos nos referindo ao processo em que os participantes selecionaram as visualizações que julgaram mais adequadas para responder às questões e detalharam quais variáveis representariam em cada visualização. Como citado anteriormente, todo o processo foi gravado. Os vídeos gerados foram transcritos em formato de atividades pelo pesquisador que conduziu o estudo. O processo de transcrição foi realizado da seguinte forma: o pesquisador assistiu cada vídeo (por partes) e segmentou cada trecho do processo realizado pelo participante em atividades. Além disso, também identificou etapas a partir da análise dos comentários dos participantes. Cada atividade identificada foi registrada em uma planilha,

indicando qual participante estava realizando a atividade (P1, P2, P3, P4, P5) e qual questão estava respondendo (Q1, Q2, Q3, Q4, Q5, Q6). Criamos um identificador para identificar a etapa³ e a descrição da atividade. Após o registro de todas as atividades, classificamos as atividades atribuindo rótulos de acordo com a descrição de cada uma, para facilitar sua identificação e análise.

Ao final da rotulação, identificamos 24 atividades realizadas durante o processo: agrupamento de interações, análise de questões, análise de interações, análise de variáveis, análise de visualizações similares, análise tabela de dados, análise visualizações, consulta, cópia, criação de notas explicativas, criação de título, desistência de decisão, relato dúvida, edição, escolha de interação, escolha de visualização, instruções pesquisador, interações miro, identificação de item faltante, leitura, mapeamento de variáveis para visualização, questionamento do pesquisador, remoção e relato de raciocínio. A Tabela 3.2 exemplifica como foram organizadas algumas das etapas extraídas dos vídeos para o participante 1 (P1).

Tabela 3.2: Exemplo de mapeamento de passos extraídos dos vídeos

#	questão	etapa	descrição	classificação
P1	Q3	P1.46	Participante leu a questão 3	leitura
P1	Q3	P1.47	Participante examinou os gráficos	análise visualizações
P1	Q3	P1.48	participante escolheu o treemap	escolha visualização
P1	Q3	P1.49	colocou o título do gráfico como “gasto anual”	criação de título
P1	Q2	P1.50	mudou o título do gráfico da questão 2 para “gasto do último ano”	edição
P1	Q1	P1.51	mudou o título do gráfico da questão 1 para “gasto anual da prefeitura dos últimos 5 anos”	edição

Após a transcrição, organização e a rotulação, revisamos as etapas para identificar em quais delas seria possível identificar as decisões tomadas pelos participantes e quais fatores influenciaram suas decisões, respondendo assim a questão de pesquisa Q1. Além disso, também analisamos se houve dificuldades

³O identificador de uma etapa é formado por ‘P’ + número do participante + ‘.’ + o número da atividade que ele estava realizando. Por exemplo: P1.14 - participante 1, atividade 14.

durante as atividades realizadas no processo, que respondem a questão Q2. Os resultados foram revisados por um segundo pesquisador, especialista em visualização de informação.

3.4.2

Inspeção das Visualizações

Na segunda análise, as 63 visualizações resultantes do estudo foram inspecionadas por dois especialistas em visualização de informação. Um dos inspetores tem doutorado focado em visualização de dados e o outro é coautor de um livro de visualização de dados e já lecionou um curso sobre o tópico. Ambos são coautores de múltiplos artigos científicos sobre visualização de dados publicados em conferências e periódicos internacionais. Os dois especialistas não participaram da condução deste estudo junto aos participantes; participaram apenas da inspeção. Os dois inspetores identificaram alguns problemas na construção e seleção das visualizações. Para examinar o material, solicitamos que eles utilizassem a classificação de tipos de erros de entendimento sobre visualização cometidos por novatos, descrita no trabalho de Rodrigues et al. (2021). A partir dessa classificação de erros, definimos diretrizes para guiar a inspeção, onde os inspetores deveriam identificar:

E1: Se a visualização não responde à questão

E2: Se a visualização responde à questão, mas não é a mais apropriada

E3: Erros ou algo que faltou na visualização

E4: Elementos que podem causar dificuldades no entendimento da visualização

E5: Outros tipos de problemas – problema que não se encaixa nos outros aspectos citados

3.5

Resultados

Nesta seção, apresentamos os resultados obtidos a partir das análises das etapas do design das visualizações e da inspeção. Os resultados estão organizados de acordo com as questões do estudo: Q1, que discute os fatores que influenciaram as decisões dos participantes no design das visualizações, e Q2, que apresenta as dificuldades encontradas durante o processo.

3.5.1

Q1. Quais fatores influenciaram as decisões no processo de criação das visualizações?

A partir da análise das etapas do processo, explicada na Seção 3.4, identificamos alguns fatores que motivaram as decisões tomadas pelos participantes durante o processo de design das visualizações, ou seja, o que os participantes utilizaram como base para fazer escolhas durante a atividade de construção de visualizações. Os fatores que motivaram as decisões, no contexto deste estudo, foram:

- Facilitar a comparação entre informações;
- Permitir o detalhamento de informações;
- Facilitar a identificação ou destaque de informações;
- Considerar a quantidade de informações a serem representadas;
- Considerar as variáveis necessárias para responder à questão; e
- Facilitar a criação da visualização.

Esses fatores são detalhados a seguir. As figuras apresentadas como exemplo são representações retiradas diretamente dos materiais produzidos pelos participantes durante o estudo, sem modificações.

Facilitar comparação entre informações: Este cenário surge quando a pergunta requer a comparação dos dados entre si, como identificar o maior valor, menor valor, entre outros aspectos. Alguns participantes fizeram escolhas de visualizações considerando quais seriam as melhores opções para mostrar a comparação entre os dados, quando julgavam que isso era importante para responder às questões. A Figura 3.4 apresenta um exemplo de visualização em que essa escolha foi feita. Nesse exemplo, P3 justificou a troca dos gráficos de pizza (parte A da Figura 3.4) pelo de barras empilhadas (parte B da Figura 3.4) – “nesse modelo de gráfico fica mais fácil a visualização e comparação de gastos de cada zona da cidade” (etapa P3.24).⁴

⁴A identificação das citações é representada pela identificação do participante (P1, P2 etc.) e por um número sequencial que representa o passo correspondente realizado por este participante, na ordem da transcrição do seu comportamento.

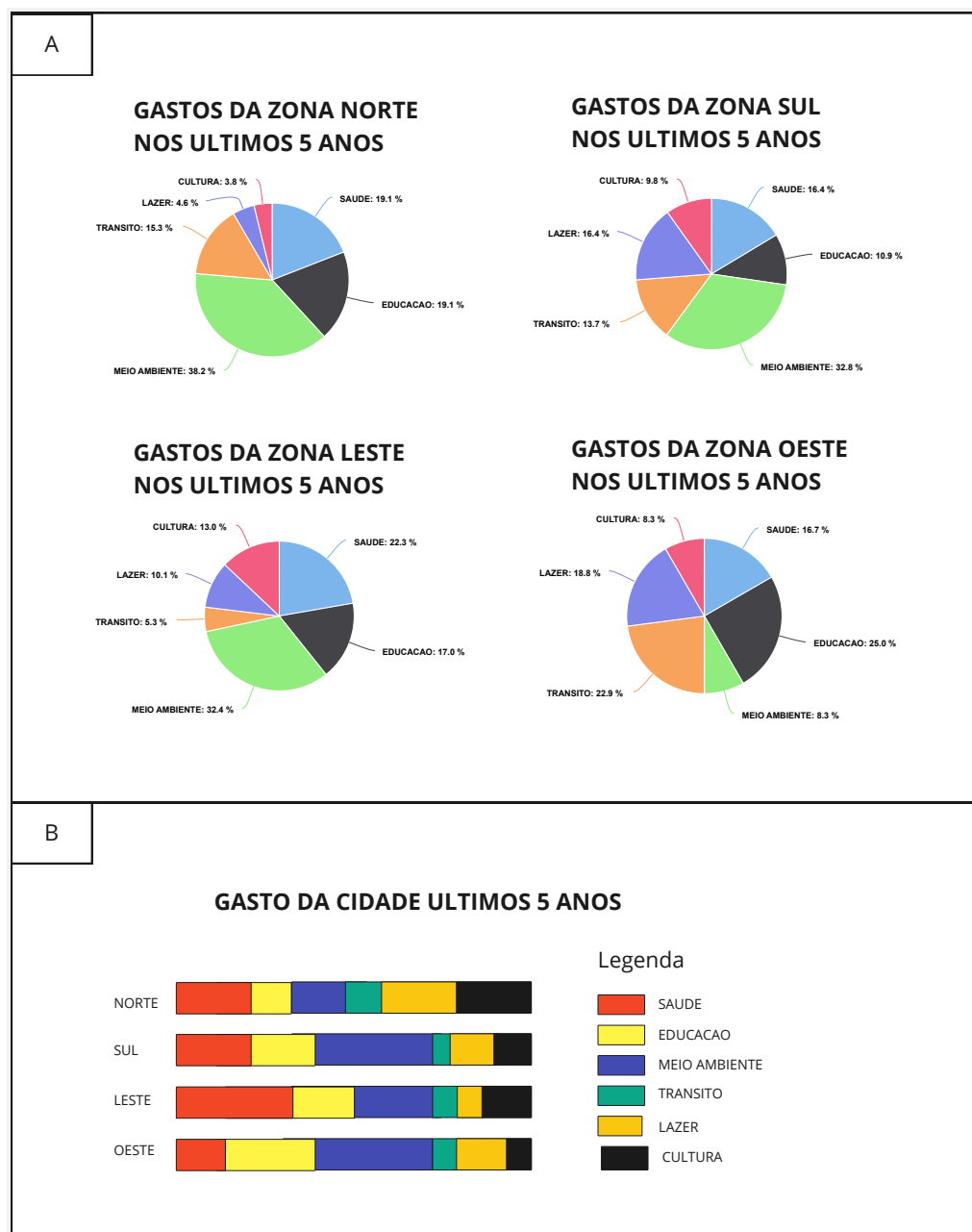


Figura 3.4: Exemplo para questão: Quais os gastos da cidade ao longo dos últimos 5 anos? (Visualização elaborada por P3)

Neste exemplo, o participante percebeu que comparar os gastos no gráfico de pizza seria mais desafiador devido à separação dos quatro gráficos (parte A da Figura 3.4). Em contrapartida, no gráfico de barras empilhadas, os valores foram representados lado a lado em um único gráfico. É evidente que, apesar de considerar o aspecto da comparação, o participante não representou adequadamente os valores ao longo do tempo, conforme solicitado na questão (*Quais os gastos da cidade ao longo dos últimos 5 anos?*).

Permitir o detalhamento de informações O fator de detalhamento se manifesta quando uma visualização apresenta informações sumarizadas,

porém oferece a possibilidade de acesso a detalhes adicionais através de interações. Por exemplo, uma visualização pode inicialmente mostrar o gasto de cada estado do Brasil, mas ao interagir com um estado específico, é possível visualizar os gastos detalhados de cada cidade dentro desse estado. Alguns participantes escolheram adicionar interações nas visualizações para ajudar usuários a acessarem detalhes das informações apresentadas. A Figura 3.5 apresenta um exemplo desse fator. No exemplo, o participante aplicou a interação *navigate* entre histograma e o gráfico de pizza: “usuário clica no histograma, na barra que representa um ano e aparece um *pie chart* que mostra em quais setores eu tive mais gasto nesse ano” (etapa P1.113). Cada barra do histograma representa os gastos de um ano. Ao clicar em um ano específico, é possível visualizar um gráfico de barras que detalha os gastos de cada região da cidade durante o ano selecionado.

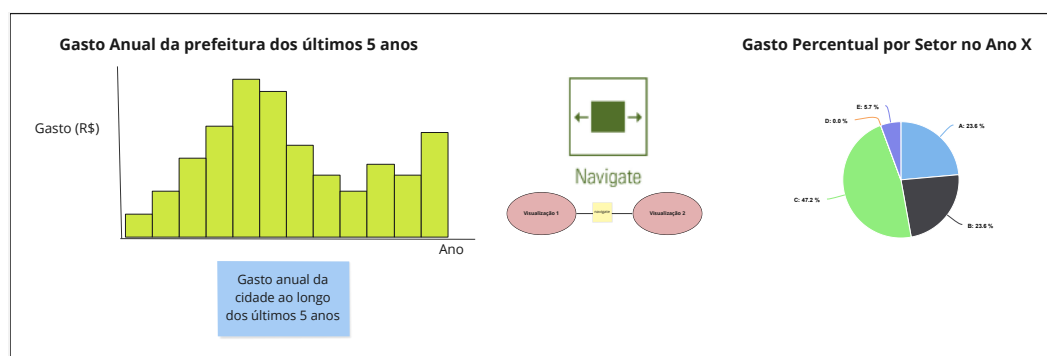


Figura 3.5: Exemplo para questão: *Quais os gastos da cidade para os últimos 5 anos?* (Visualização elaborada por P1)

Facilitar identificação ou destaque de informações: O destaque de informações emergiu como um fator que influenciou na criação das visualizações, especialmente quando os participantes buscavam enfatizar ou facilitar a identificação de dados específicos. Os destaques eram realizados por meio de interações ou recursos tais como: adicionar funcionalidades de interação por *hover* para destacar elementos do gráfico, utilizar cores para enfatizar informações importantes, e empregar ordenação para facilitar a identificação dos maiores e menores valores. A Figura 3.6 mostra um exemplo de visualização em que foi aplicado este fator. No exemplo, P1 explicou que o usuário inspecionaria um setor no gráfico de barras, então mostraria um gráfico de linhas para mostrar como foi o investimento mensal durante aquele ano (etapa P1.136). Neste caso, o destaque ocorre quando o usuário passa o mouse por cima de uma barra do gráfico e esta fica diferenciada das outras (podendo ser através de mudança de cor da borda da barra ou aumentando seu tamanho, por exemplo).

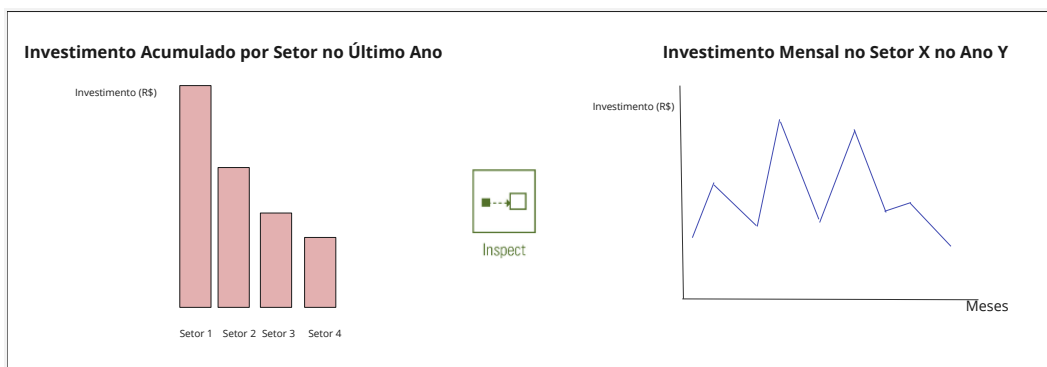


Figura 3.6: Exemplo para questão: *Quais setores do serviço público receberam mais investimentos no ano?* (Visualização elaborada por P1)

Considerar a quantidade de informações a serem representadas:

Neste fator, a decisão é influenciada pelo volume de dados a serem representados na visualização. Por exemplo, representar os dias da semana constitui uma quantidade menor de informações, enquanto representar todos os dias do ano caracteriza uma quantidade maior de informações. Nesse contexto, os participantes consideraram a quantidade de itens a serem representados para escolher a visualização. Como exemplo, P4 pensou em utilizar *treemap*, mas citou que “devido à quantidade de meses, iriam ficar muitos itens na visualização e ficaria difícil identificar no *treemap* os dois meses que teriam mais gastos, caso o tamanho dos itens [as áreas do *treemap*] fosse bem parecido” (etapa P4.27). Sendo assim, ele escolheu o gráfico de barras para representar a questão, tal como apresentado na Figura 3.7.

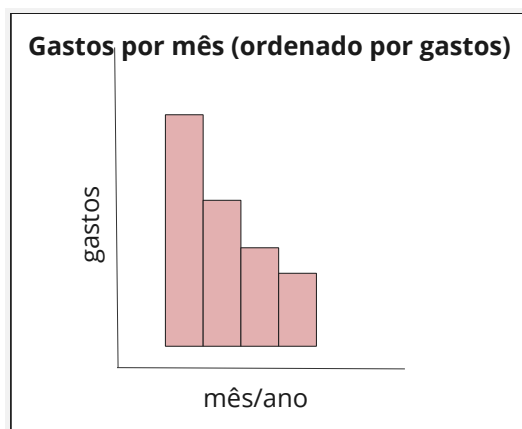


Figura 3.7: Exemplo para questão: *Quais os dois meses do ano que tiveram mais gastos?* (Visualização elaborada por P4)

Considerar as variáveis necessárias para responder à questão:

Nesse fator, os elementos importantes para a decisão foram as variáveis identificadas a partir das questões propostas. Por exemplo, se a questão solicitava a representação da zona da cidade, dos gastos e do tempo, o

participante selecionava uma visualização que incorporasse elementos para representar estas três variáveis. Alguns participantes basearam sua escolha analisando elementos da visualização para representar as variáveis e se a visualização seria uma boa opção para representar uma variável específica. Alguns exemplos destes itens são: P4 pensou que na verdade não poderia utilizar o *treemap* porque não teria três variáveis para representar (etapa P4.29) e P5 citou que: “o gráfico (ver Figura 3.8) ajuda a representar 3 coisas: o tipo de gasto, porque na questão pergunta ‘quais’, [escolheu a cor da linha para representar tipo de gasto]. Quando ele fala durante os cinco anos, esse gráfico tem uma coisa que ele dá para representar um histórico (eixo x); e no eixo y os valores dos gastos” (etapa P5.18). Analisando o exemplo, é possível observar a associação feita pelo participante entre as variáveis e os elementos do gráfico, como linhas, cores e eixos.

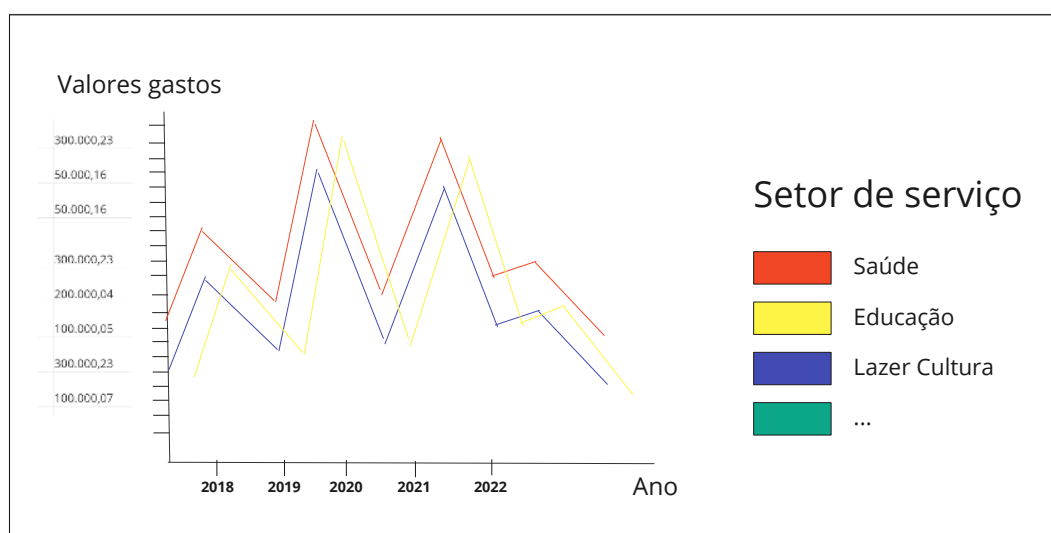


Figura 3.8: Exemplo para questão: *Quais os gastos da cidade ao longo dos últimos 5 anos?* (Visualização elaborada por P5)

Facilitar a criação da visualização: Outro fator que influenciou a decisão de escolha foi que os participantes escolheram visualizações com base na sua familiaridade. Alguns participantes desistiram de utilizar algumas visualizações, pois não sabiam como utilizá-las. Alguns exemplos deste fator foram: P4 afirmou que não daria para usar *treemap* com datas pois não sabia como representar (etapa P4.63) e escolheu gráfico de linhas porque tinha mais familiaridade com ele (etapa P5.17). Neste exemplo, observamos que o participante optou por uma determinada visualização não necessariamente por considerá-la mais apropriada para a tarefa, mas sim por sua familiaridade com o seu uso.

3.5.2

Q2. Quais desafios encontrados no processo de criação das visualizações?

Ao analisar as etapas do processo de criação e os relatos dos participantes durante o estudo, identificamos alguns desafios na diferenciação entre visualizações, na escolha de visualizações adequadas e no mapeamento de variáveis para visualização. Esses desafios são descritos a seguir.

Dúvidas sobre como diferenciar visualizações: Um dos participantes (P1) teve dúvidas sobre a diferença entre histograma e gráfico de barras. Realizou pesquisa na internet para entender as diferenças, mas não chegou a uma conclusão. Escolheu histograma na lista de visualizações, mas durante a criação, configurou a visualização como se fosse um gráfico de barras (ver Figura 3.9). Outro participante citou que todas as visualizações pareciam ter a mesma função. P2 também teve dúvida entre usar o histograma ou o gráfico de barras para responder a Q1. Pesquisou sobre diferença entre histograma e gráfico de barras e selecionou o gráfico de barras. Escolheu gráfico de barras para facilitar comparação e porque achou que fica mais fácil identificar onde se encontram os valores, do que em um gráfico de linhas.

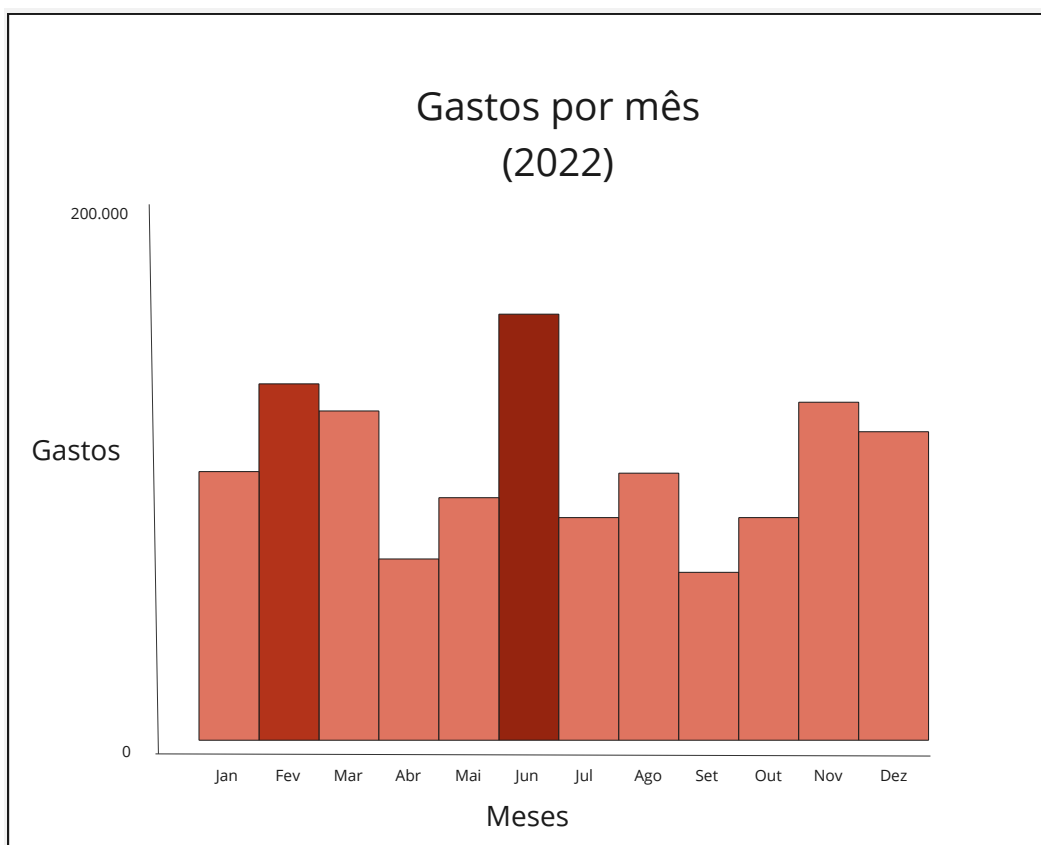


Figura 3.9: Exemplo para questão: *Quais os dois meses do ano que tiveram maiores gastos?*

Dificuldade no mapeamento de variáveis na visualização: P1

relatou que ficou em dúvida ao escolher *treemap* (ver exemplo de visualização na Figura 3.10) pois esta visualização tem três elementos: a cor, intensidade e tamanho; e P1 não ficou confortável com a escolha. P4 pesquisou sobre mapa coroplético, analisou as imagens resultantes da busca, decidiu não utilizá-lo pois não sabia como representar variável tempo nesta visualização. Além disso, P4 não conseguiu utilizar *treemap* nas questões Q4 e Q6, pois não sabia representar a variável mês; P4 pesquisou como representar variável tempo no *treemap*, visualizou as imagens, mas não entendeu como utilizar e desistiu de utilizar a visualização.

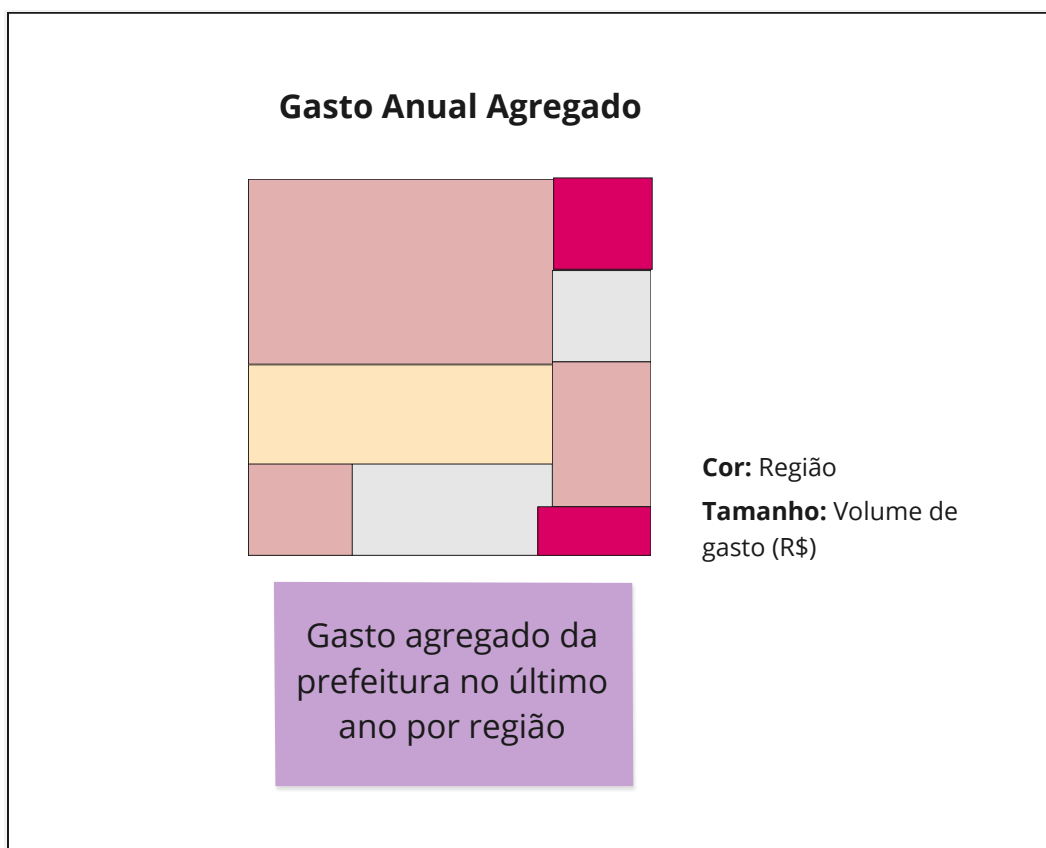


Figura 3.10: Exemplo para questão: Q3. Qual região teve mais gastos no ano?

Dificuldade na escolha da visualização: P1 relatou que poderia ter algo que o ajudasse a lembrar a utilidade de cada visualização: “Os gráficos também poderiam relembrar um pouco pra que cada um deles serve” - P1. P2 disse que teve dúvidas sobre qual visualização escolher e sentiu falta de saber qual era o propósito de cada visualização e o que ela pode representar: “acho que é mais a questão mesmo de decidir qual gráfico utilizar, que às vezes dá uma certa dificuldade. Tem que ver qual objetivo e saber para quê serve cada um desses gráficos e o que ele pode representar. A dificuldade foi isso e o desconhecimento de alguns gráficos mesmo” - P2. Além disso, disse que não sabe dizer se fez as melhores escolhas. P3 também citou que não sabe como

algumas visualizações funcionam e que seria bom ter exemplos de uso: “Em cada tipo de gráfico demonstrar como ficaria os dados no gráfico, para ter uma ideia de visualização (...). Não sei se talvez o uso deles poderia ser melhor para usar ou não. Eu não tinha uma relação para eu olhar os dados, analisar os gráficos e dizer “talvez esse daqui seja melhor” - P3 e “Talvez ter exemplos de uso dos gráficos, alguns gráficos não sei como de fato funcionam” - P3. P5 relatou que não sabia como usar alguns gráficos e escolheu gráfico de linhas por ter mais familiaridade com ele. Em quatro das seis questões, P5 utilizou gráfico de linhas.

3.5.3

Resultados da Inspeção das Visualizações por Especialistas

Na segunda análise, dois inspetores, especialistas em visualização de informação, examinaram as 63 visualizações criadas e identificaram alguns problemas. O problema mais comumente identificado foi a seleção de visualizações inadequadas para representar as respostas para as questões propostas. O resumo dos resultados quantitativos da inspeção são apresentados na Figura 3.11. Como citado anteriormente, as diretrizes (apresentadas no gráfico da Figura 3.11) utilizadas para guiar a inspeção foram:

- E1: Se a visualização não responde à questão
- E2: Se a visualização responde à questão, mas não é a mais apropriada
- E3: Erros ou algo que faltou na visualização
- E4: Elementos que podem causar dificuldades no entendimento da visualização
- E5: Outros tipos de problemas – problema que não se encaixa nos outros aspectos citados

Em relação à inspeção pelos especialistas, a maioria dos problemas identificados estavam relacionados à escolha de visualizações que respondem à questão, mas não eram as mais apropriadas (E2), sendo este problema encontrado em 49 (77.8%) visualizações. A Figura 3.12 mostra um exemplo de problema classificado como E2. A visualização foi criada para responder à questão: “Quais setores do serviço público receberam mais investimentos no ano?”. O inspetor comentou que: “ [A visualização] responde mas eu teria que somar todos os meses de cada região de cada mês [para obter a resposta]”.

Os inspetores também identificaram 29 problemas relacionados a erros cometidos no uso das visualizações e/ou elementos faltantes (E3). Para o tipo E4 (elementos que podem causar dificuldade na interpretação da visualização),

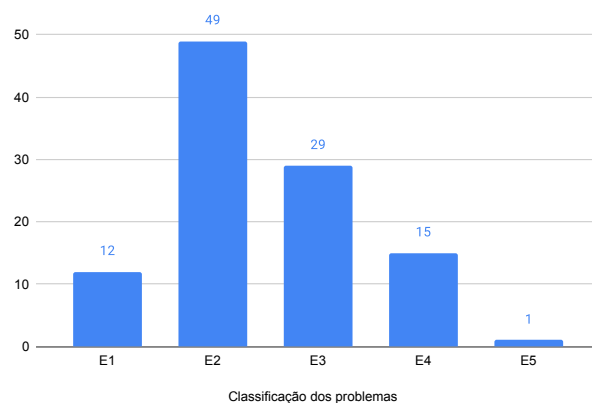


Figura 3.11: Quantidade de problemas encontrados (eixo vertical) para cada diretriz (eixo horizontal)

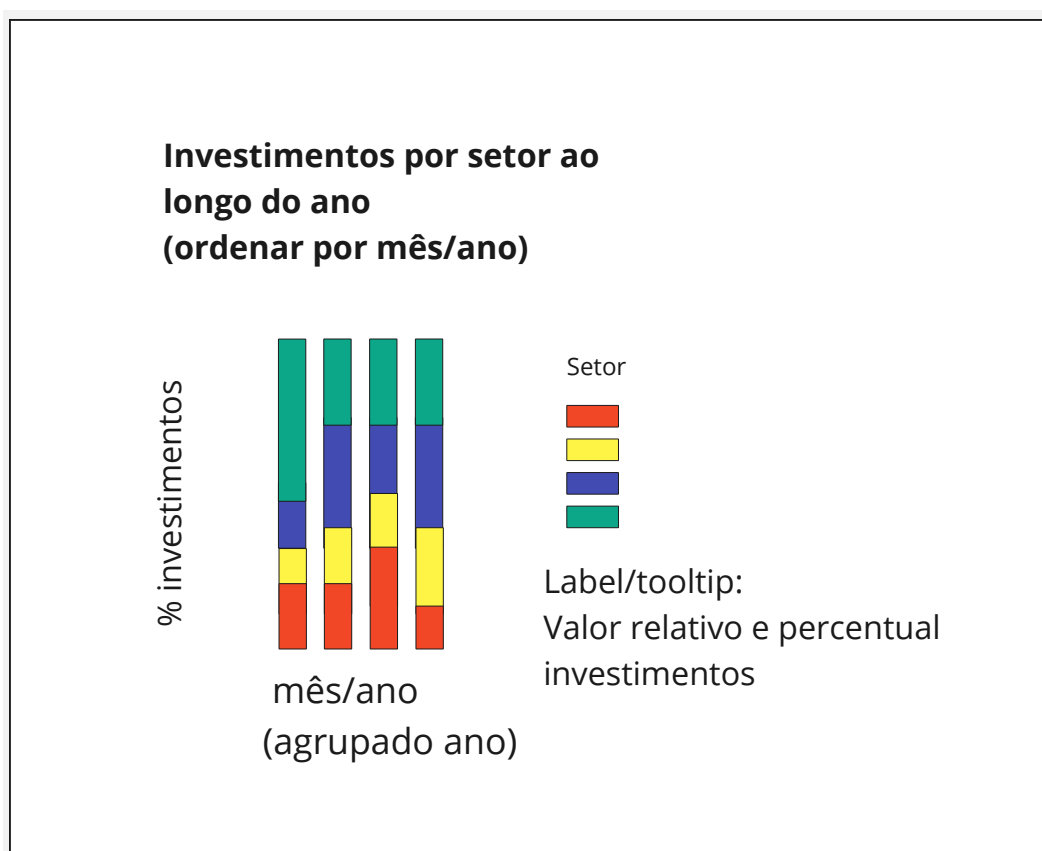


Figura 3.12: Exemplo para a questão: *Quais setores do serviço público receberam mais investimentos no ano?*

eles identificaram 15 erros e, para criação de visualizações que não respondiam às questões (E1), 12 casos dentre as 63 visualizações. Um exemplo de problema identificado na inspeção foi: “A visualização (ver Figura 3.13) não especifica quais os gastos, e sim o total de gastos por ano.” – inspetor2 e “Talvez um gráfico de barras seja mais adequado para a tarefa de *information retrieval* (retornar valor)” – inspetor2. Pode-se observar ainda que o participante

utilizou um histograma equivocadamente, como sendo gráfico de barras.

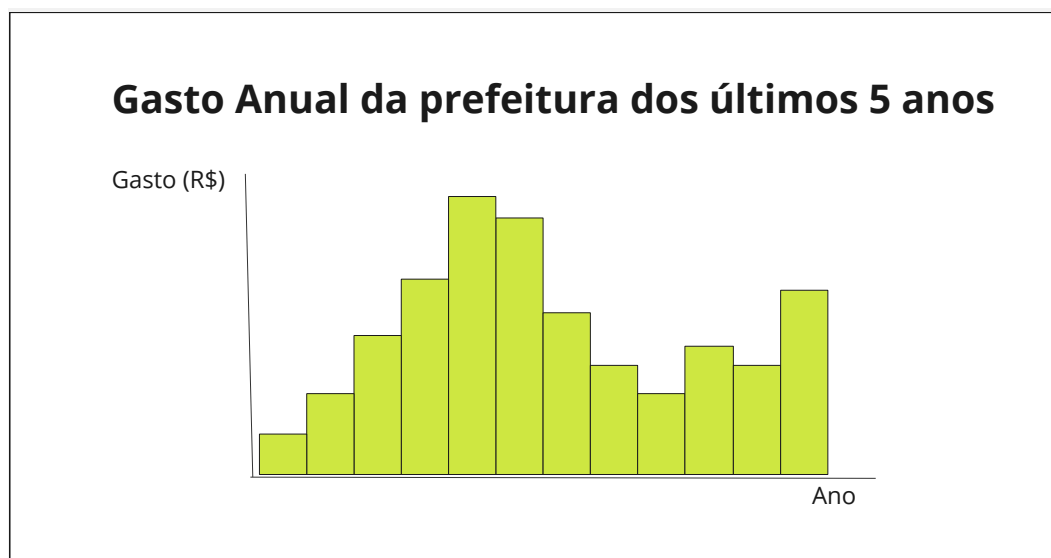


Figura 3.13: Exemplo para a questão: Q1. *Quais os gastos da cidade ao longo dos últimos 5 anos?*

Para o item E5, encontraram apenas um problema: “O gráfico (ver Figura 3.14) foi reordenado, mas o eixo X não foi reordenado corretamente”. Este comentário foi relacionado à ordenação em gráfico de barras, em que o participante decidiu ordenar as barras de acordo com os gastos e não de acordo com os meses (que ficaram fora de ordem).

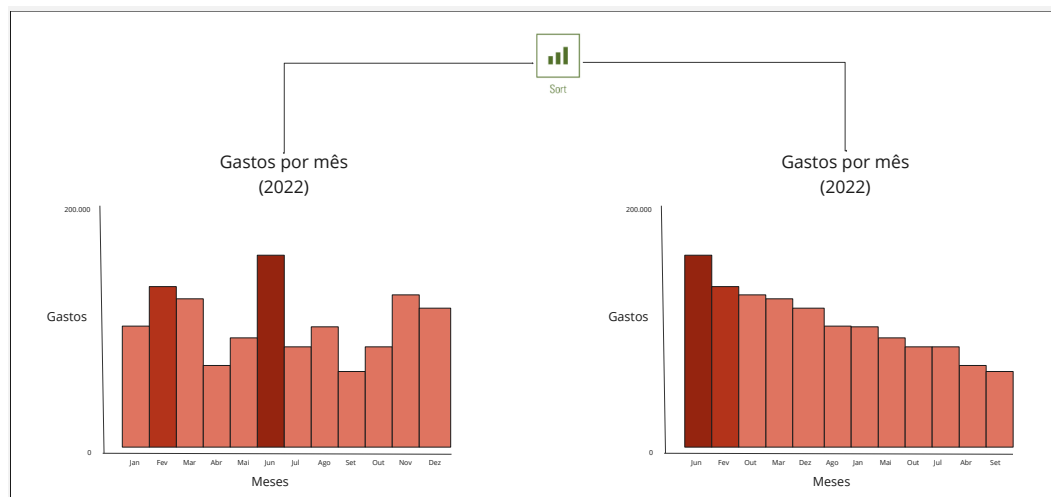


Figura 3.14: Exemplo para a questão: Q2. *Quais os dois meses do ano que tiveram maiores gastos?*

3.6

Discussão dos Resultados

A partir dos resultados das análises, identificamos alguns aspectos que geraram problemas durante o processo de criação de visualizações. Esses aspectos foram: diferenciação entre visualizações, mapeamento de variáveis para visualização e escolha de visualizações adequadas. Além disso, identificamos alguns elementos que ajudam a entender o que novatos utilizam como base para tomada de decisões durante representação dos dados, no contexto deste estudo.

3.6.1

Diferenciação entre as visualizações

Em relação à **diferenciação entre as visualizações**, os participantes tiveram dificuldades em diferenciar histogramas e gráficos de barras. Alguns escolheram histogramas, mas utilizaram o gráfico de barras. Esse problema foi identificado na análise das etapas e também através da inspeção dos especialistas em visualização. Esse aspecto também foi identificado no estudo de Keck et al. (2020), cujos participantes tiveram dificuldades em identificar diferenças entre os elementos de construção das visualizações.

Nesse contexto, argumentamos que é importante apoiar novatos no entendimento das estruturas (exemplo: suas formas, representações, nomenclaturas) das visualizações de forma que consigam diferenciar qual visualização estão utilizando para representar seus dados de forma correta. Ao analisarmos a literatura sobre letramento em visualização, é comum observar confusão entre visualizações esteticamente similares. Isto ocorre porque as agregações subjacentes aos gráficos nem sempre são evidentes para indivíduos não familiarizados com as possibilidades existentes.

3.6.2

Mapeamento de variáveis para visualização

No **mapeamento de variáveis**, os participantes apresentaram dúvidas sobre como mapear variáveis em uma visualização. Outra dificuldade encontrada foi saber se todas as variáveis necessárias para responder uma questão poderiam ser representadas em uma única visualização. A maioria dos participantes não pensou em combinar visualizações. Este resultado está em linha com o estudo realizado por Grammel et al. (2010), em que também identificaram dificuldades dos participantes em traduzir perguntas em atributos de dados e variáveis visuais.

Observamos que a dificuldade em mapear os dados para uma determinada visualização pode levar novatos a optar pela escolha de outra visualização

mais familiar, mesmo que não seja a mais adequada para representar os dados. A partir destes resultados, notamos que o mapeamento de variáveis para a visualização (ou seja, a associação entre elementos da visualização e os dados) é uma atividade que gera dificuldades no design de visualizações, sendo necessário buscar formas de apoiar esta tarefa.

3.6.3

Escolha de visualizações adequadas

Sobre a **escolha de visualizações**, os participantes sentiram falta de algo que lhes ajudasse a entender a utilidade de cada visualização e o que elas poderiam representar. Alguns participantes realizaram consultas à internet, mas não chegaram a conclusões e alguns não ficaram confiantes com sua escolha.

Os especialistas que realizaram a inspeção das 63 visualizações criadas identificaram 12 escolhas erradas de visualizações e 49 casos em que as escolhas respondiam às questões, mas não eram a escolha mais apropriada. Essa dificuldade de escolher a visualização adequada também foi identificada nos trabalhos de Baker et al. (2001) e Grammel et al. (2010). Pudemos observar que os participantes apresentaram dificuldades na escolha de visualizações adequadas, seja por uma preferência por visualizações familiares ou por um desconhecimento de outras opções. Do ponto de vista do letramento em visualização, a preferência por visualizações familiares é compreensível, dado que o conhecimento sobre diferentes tipos de gráficos é frequentemente limitado. Uma manifestação deste ponto em nosso estudo foi que a maioria dos participantes escolheu gráficos de barras para responder às questões, mesmo não sendo o tipo de gráfico mais adequado para aquela questão. Destacamos que todos os participante relataram conhecer gráficos de barras. Tendo em vista que, apesar de, na caracterização os participantes citarem conhecer algumas visualizações e terem conhecimentos práticos e teóricos em visualizações, eles cometeram vários erros no uso e na escolha dessas mesmas visualizações.

3.6.4

Base para decisões

Outro item analisado durante o estudo foram os aspectos que os participantes consideraram para tomar suas decisões. Identificamos cinco fatores: permitir detalhamento das informações, facilitar comparações, quantidade de informações a serem apresentadas, análise de variáveis e familiaridade com a visualização. Profissionais experientes em visualizações podem considerar outros critérios que influenciam suas decisões.

Permitir detalhamento de informações: um dos fatores considerados

pelos participantes foi facilitar o detalhamento de informações para o usuário final. Um dos recursos que ajudou nessa tomada de decisão foi a adição de interações. Neste contexto, quando solicitados a representar dados com informações complementares entre si (por exemplo: estados → cidades ou anos → meses), os participantes empregavam mecanismos de interação como clicar ou dar zoom em elementos da visualização para exibir detalhes relacionados.

Facilitar comparações: Outro ponto considerado importante pelos participantes foi facilitar a comparação entre valores nas visualizações (por exemplo: maior gasto ou menor valor). Este item motivou a troca de uma visualização por outra considerada mais adequada. Um dos recursos empregados para facilitar comparações foi a utilização de uma única visualização que apresentasse os dados a serem comparados lado a lado.

Quantidade de informações a serem apresentadas: A quantidade de informações a serem representadas também foi um fator que motivou a escolha de visualizações (por exemplo: ter de representar os 365 dias do ano em uma visualização). No estudo, um exemplo foi a não utilização de gráfico de pizza, pois havia grande quantidade de informações para representar e, por isso, essa visualização foi considerada inadequada.

Análise de variáveis: A análise de variáveis foi uma estratégia que os participantes utilizaram para ajudar na criação de visualizações. No entanto, pudemos notar pelos resultados que houve dificuldade em traduzir estas variáveis em visualizações, levando a escolhas inadequadas ou incorretas.

Familiaridade com a visualização: No contexto deste estudo, este fator se relacionou com a análise de variáveis e a incerteza na escolha das visualizações, pois este critério foi utilizado quando os participantes não optaram por visualizações que já dominavam, mesmo que outras opções fossem mais adequadas, devido à falta de familiaridade com determinadas técnicas de visualização. Esse fato também foi observado no trabalho de Grammel et al. (2010).

3.7

Considerações sobre o Estudo de Observação

Neste capítulo, apresentamos um estudo empírico realizado para investigar o processo de criação de visualização por cinco profissionais da indústria de software que eram novatos em visualização de informação. O estudo objetivou analisar as atividades que os participantes executariam caso estivessem em um projeto, com questões definidas, uma base de dados e consultas à internet. Durante o processo, não houve intervenção com uso de ferramentas existentes para facilitar a criação, embora os participantes fossem livres para realizar

buscas na internet e encontrar ferramentas ou quaisquer outras formas de apoio à construção de visualizações. Os participantes criaram um total de 63 visualizações, que foram inspecionadas por dois especialistas para identificação de problemas. Além disso, analisamos as etapas de criação realizadas pelos participantes.

A partir dos resultados, conseguimos identificar alguns problemas na criação das visualizações do ponto de vista dos especialistas que não participaram da condução do estudo mas analisaram as visualizações criadas pelos participantes (inspeção), dos pesquisadores que conduziram e planejaram o estudo e acompanharam e revisaram o processo de design, e dos próprios participantes (relatos de raciocínio). Além disso, identificamos alguns aspectos que os participantes utilizaram como base para decisões que tomaram durante o processo. Esses aspectos foram: permitir detalhamento das informações, facilitar comparações, quantidade de informações a serem apresentadas, análise de variáveis e familiaridade com a visualização. Com isso, conseguimos compreender o que novatos consideraram importante avaliar durante a criação de visualizações no contexto desta pesquisa.

Além dos aspectos citados, também identificamos três principais obstáculos que levaram os participantes à criação de representações incorretas ou inadequadas para seus dados: diferenciação entre visualizações, mapeamento de variáveis para visualização e escolha de visualizações adequadas.

Os resultados corroboram os desafios previamente discutidos na literatura (*cf.* (Baker et al., 2001; Grammel et al., 2010; Keck e Groh, 2019)). Além disto, esta pesquisa identificou fatores adicionais que contribuem para a compreensão das decisões de novatos no design de visualizações, os quais não foram abordados nos estudos anteriores, com exceção da familiaridade com visualizações mencionada por Grammel et al. (2010).

No próximo capítulo, apresentamos um material didático que elaboramos para apoiar novatos no processo de criação de visualizações. O material proposto foca especificamente em duas dificuldades apresentadas neste estudo: mapeamento de variáveis e escolha de visualizações adequadas. O problema de diferenciação de elementos da visualização não foi considerado no material, pois os problemas encontrados nesse item foram menos citados e mais sutis.

4

VIS-SeMa: Material para apoiar a seleção de visualizações de dados e o mapeamento de variáveis

Esta pesquisa tem por objetivo contribuir para apoiar novatos no design de visualizações de dados. Elaboramos o VIS-SeMa, um material de consulta estruturado, que guia novatos, através de *templates* de questões, a decidirem sobre quais visualizações utilizar e como mapear variáveis para estas visualizações. O material pode ser utilizado quando o seu “usuário” (criador/autor das visualizações) já possui questões a serem respondidas, mas não sabe qual visualização é mais adequada para responder suas questões; ou quando tem dúvidas sobre como utilizar uma visualização.

A partir dos obstáculos enfrentados por novatos durante o processo de design de visualizações, identificados no estudo anterior e em estudos da literatura, organizamos um conjunto de características que seriam importantes apoiar esse processo. A partir dessas características, definimos alguns requisitos para elaboração do material didático VIS-SeMa (*Select and Map*). A seguir explicamos e apresentamos as evidências para justificar cada um desses requisitos.

4.1

Requisitos para a Elaboração do Material

Como citado anteriormente, para proposta do VIS-SeMa (*Select and Map*), utilizamos como base as evidências identificadas a partir de estudos encontrados na literatura e resultados do estudo de observação realizado nesta pesquisa. Com base na análise dos estudos, definimos alguns requisitos para guiar a elaboração do material: apoiar atividades específicas; apoiar público-alvo de novatos em visualizações. Como requisitos relacionados ao conteúdo do material, temos: ser baseado em *templates* (modelos) de questões, apresentar visualizações simples, apresentar exemplos concretos e seguir princípios de design (ser modularizado e conter visualizações limpas).

4.1.1

Apoiar Atividades Específicas

Crisan et al. (2020), em sua revisão da literatura definiram quatro fases para ciência de dados, explicadas anteriormente: preparação de dados, análise, implantação e comunicação. O VIS-SeMa pode ser utilizado na fase de análise, após os dados serem explorados e a pessoa que analisou já tiver ideia de quais questões deseja responder. Dessa forma, o material apoia o planejamento de qual visualização utilizar (seleção) e como utilizá-la (mapeamento) para responder cada questão, gerando protótipos de visualizações.

Decidimos apoiar atividades específicas neste momento para não deixar o material muito extenso e não dificultar o seu uso. Destacamos que o material proposto não trata das etapas de carregamento e exploração de dados e nem de geração de gráficos automaticamente. O objetivo do material é o planejamento de visualizações para questões pré-definidas. Todo conteúdo gerado será criado em ferramenta auxiliar pelos próprios usuários. Dentre as três atividades que apresentaram dificuldades aos novatos no estudo de observação (Capítulo 3) – diferenciar visualizações semelhantes, seleção de visualizações e mapeamento de variáveis em visualizações e diferenciação de visualizações com anatomia semelhante (gráfico de barras e histograma) –, selecionamos as duas em que os novatos mostraram ter mais dificuldades: (i) seleção de visualizações adequadas e (ii) mapeamento de atributos do dado para variáveis visuais.

Seleção de visualizações adequadas: Esta atividade consiste em selecionar tipos de visualização (*e.g.*, gráfico de barras, linhas, dispersão) adequados para cada contexto, dependendo de quais dados se quer representar e quais questões se quer responder. Para a seleção com apoio do VIS-SeMa, o autor da visualização deve já ter definido as questões que deseja responder. Além de a seleção de visualizações adequadas ter sido uma dificuldade que identificamos no estudo de observação, Baker et al. (2001) também relataram essa dificuldade.

Mapeamentos de atributos do dado (variáveis) para elementos visuais: Trata-se da atividade de identificar quais os atributos dos dados devem ser representados de forma visual e como associar cada um desses atributos às variáveis visuais (elementos e aspectos gráficos da visualização). Exemplos de elementos das visualizações são: linhas, barras, cores, símbolos. Além de o mapeamento de variáveis em elementos visuais ter sido uma dificuldade que identificamos no estudo de observação, também foi relatada em alguns trabalhos na literatura de diferentes formas: dificuldades em identificar variáveis nas questões (Grammel et al., 2010); erros ao associar variáveis com elementos da visualização (Baker et al., 2001) e não conseguir distinguir alguns

elementos semelhantes que compõem as visualizações (Keck et al., 2020).

4.1.2

Apoiar Especialistas em um Domínio, mas Novatos em Visualização

Em relação ao público-alvo, definimos o material para apoiar novatos *em visualização de informação*, ou seja, pessoas que podem ser especialistas em outras áreas, mas não possuem conhecimentos avançados na construção de visualizações. Esta decisão foi baseada em resultados de nosso estudo anterior (Ferreira et al., 2023), em que novatos apresentaram dificuldades na construção de visualizações, necessitando recorrer a especialistas.

Tendo em vista que nem sempre as equipes possuem especialistas em visualização, embora seja o mais adequado, optamos por elaborar um material que apoie o uso de visualizações por novatos. Esperamos ainda que, através desse material, os novatos adquiram conhecimento suficiente para dispensar o material no futuro. Entretanto, deixamos para trabalho futuro realizar estudos longitudinais para avaliar este efeito.

4.1.3

Ser Baseado em *Templates* de Questões Comumente Associadas a Tipos de Visualização

Para apoiar a seleção da visualização, o material proposto utiliza os *templates* de questões identificados por (Rodrigues et al., 2021) como perguntas que novatos associaram a visualizações que lhes foram apresentadas. Os *templates* apresentados são definidos como questões parametrizadas, cujos parâmetros representam os tipos de dados (*e.g.*, N: nominal, O: ordinal, Q: quantitativo contínuo) das questões.

O *template* associado com a anatomia da visualização (representação abstrata da estrutura visualização) visa a auxiliar o mapeamento de variáveis para os elementos da visualização e fornece uma linguagem padronizada para o design de visualizações. Segundo Keck et al. (2020), utilizar uma linguagem comum facilita a comunicação de informações. A Tabela 4.1 apresenta alguns *templates* de questões identificados por Rodrigues et al. (2021) e as visualizações correspondentes que são disponibilizadas no material VIS-SeMa.

4.1.4

Representar Inicialmente Visualizações Comumente Utilizadas

Para a primeira versão do material, escolhemos apenas algumas visualizações e seus respectivos *templates* de questões: Gráficos de barras e gráficos de linhas e algumas de suas variações (gráfico de barras simples, gráfico de barras

Tabela 4.1: Templates definidos para cada visualização

templates	visualização
- Qual o valor de Q para N1.A1?	Gráfico de Barras
- Qual N1 tem [maior menor] Q?	Gráfico de Barras
- Quais Ns têm [mais menos] que A objetos?	Gráfico de Barras
- Qual é a [média moda mediana] do número de objetos(Qs)	Gráfico de Barras
- Qual é o valor de Q para [N1 e N2 N1.A e N2.B]?	Gráfico de Barras Agrupadas
- Qual N1 tem o [maior menor] percentual em N2.B?	Gráfico de Barras Empilhadas
- Qual N1 tem o [maior menor] percentual por N2?	
- Como Q [se comporta/cresce/diminui] em T.A?	Gráfico de Linha Simples
- Em qual T, Q atinge [maior/menor] valor?	
- Em que período T, Q [cresce/diminui] [mais/menos]?	
- Qual N tem [maior/menor] variação de Q no período de tempo T?	
- Como Q se comporta [em cada N1] (ao longo de T.A) ?	Gráfico de Linhas Múltiplas
- Como Q se comporta [em cada N1.B] (ao longo de T.A) ?	
- Em que N1 Q [aumenta/diminui] [mais/menos] ao longo de T?	
- Qual N1 tem o [maior menor] Q em um período T?	

agrupadas, gráfico de barras empilhada, gráfico de linhas com uma linha e gráfico de barras com várias linhas).

Neste primeiro momento, selecionamos visualizações mais comumente utilizadas para compor o material. Esta escolha se justifica pois, apesar de terem sido relatadas pelos participantes do estudo de observação como mais conhecidas e mais utilizadas, eles enfrentaram diversos problemas ao tentar utilizá-las, tanto na seleção de visualizações como no mapeamento de variáveis. Após avaliar a efetividade do material com essas visualizações e fazer ajustes, será possível agregar outros tipos de visualização em trabalhos futuros. Para auxiliar a expansão do VIS-SeMa, fornecemos um guia para adicionar mais conteúdos ao material (Seção 4.4).

4.1.5

Apresentar exemplos concretos

Neste material, incluímos exemplos associados com as questões e com as visualizações para facilitar o entendimento pelos novatos. Escolhemos utilizar exemplos pois, no estudo apresentado no Apêndice A, alguns participantes

pesquisaram por exemplos na internet para ajudar no design das visualizações.

Além disso, alguns trabalhos da literatura observaram que o uso de exemplos contribuem para esse processo. Keck et al. (2020) citaram que a utilização da sua metodologia foi facilitada devido ao uso de exemplos. Wang et al. (2020) optaram por não utilizar exemplos, elaborando uma metodologia livre de contexto. No entanto, citaram como limitações que seria necessário explorar resultados com uso de exemplos concretos e domínios específicos. Bako et al. (2024) apresentaram um estudo em que observou que o uso de exemplos ajuda designers a desenvolverem mais ideias de visualizações.

4.1.6

Seguir Princípios de Design de Visualização

Para elaboração do material, utilizamos alguns princípios de design definidos por Wang et al. (2020), selecionados por terem sido utilizados para a elaboração de outro material para design de visualização já avaliado. Os princípios definidos foram: modularidade, independência de contexto, visualizações limpas e neutralidade de estilo. Cada um dos princípios é explicado no Capítulo 2. Não utilizamos a independência de contexto pois optamos por apresentar exemplos concretos no material e não utilizamos a neutralidade de estilo pois definimos um público-alvo específico para o material: novatos em visualização. Sendo assim, os princípios utilizados na elaboração do VIS-SeMa foram a modularidade e visualizações limpas, conforme explicado a seguir.

Estrutura modularizada: O material foi estruturado de forma modularizada, ou seja, o conteúdo explicativo das visualizações foi dividido em tópicos específicos para facilitar o entendimento e aplicação do conteúdo, bem como a localização eficiente de um conteúdo particular em caso de se utilizar o material como referência rápida. Os tópicos apresentados são: nome da visualização recomendada, função, anatomia e exemplos. Esses itens foram organizados de forma semelhante a como é feito em padrões de design de interfaces. O material proposto apresenta informações sobre “o que” deve ser utilizado, “quando” utilizar, “como” utilizar e “exemplos” de utilização. “O que” é representado através da recomendação de qual visualização utilizar. “Quando” é representado através do item função, que apresenta a explicação sobre a utilidade da visualização e o “Como” é representado pela anatomia, formada pelos *templates* de questão que guiam o entendimento de quando a visualização deve ser utilizada, através da análise das questões e seus tipos de variáveis. Os gráficos representados na anatomia que explicam como deve ser o mapeamento de variáveis para os elementos da visualização. Por fim, o material apresenta “Exemplos”, que são os exemplos concretos da visualização elaborados para

uma questão associada.

Visualizações Limpas: Em relação ao conteúdo explicativo das visualizações, optamos por utilizar em maior parte elementos gráficos para ser um material de consulta rápida e fácil de utilizar. Explicações textuais foram utilizadas somente quando necessário. A maior parte textual foi utilizada para explicar o processo de uso do material e tipos de dados básicos.

A Figura 4.1 apresenta os requisitos do material proposta organizado pelos respectivos tópicos.

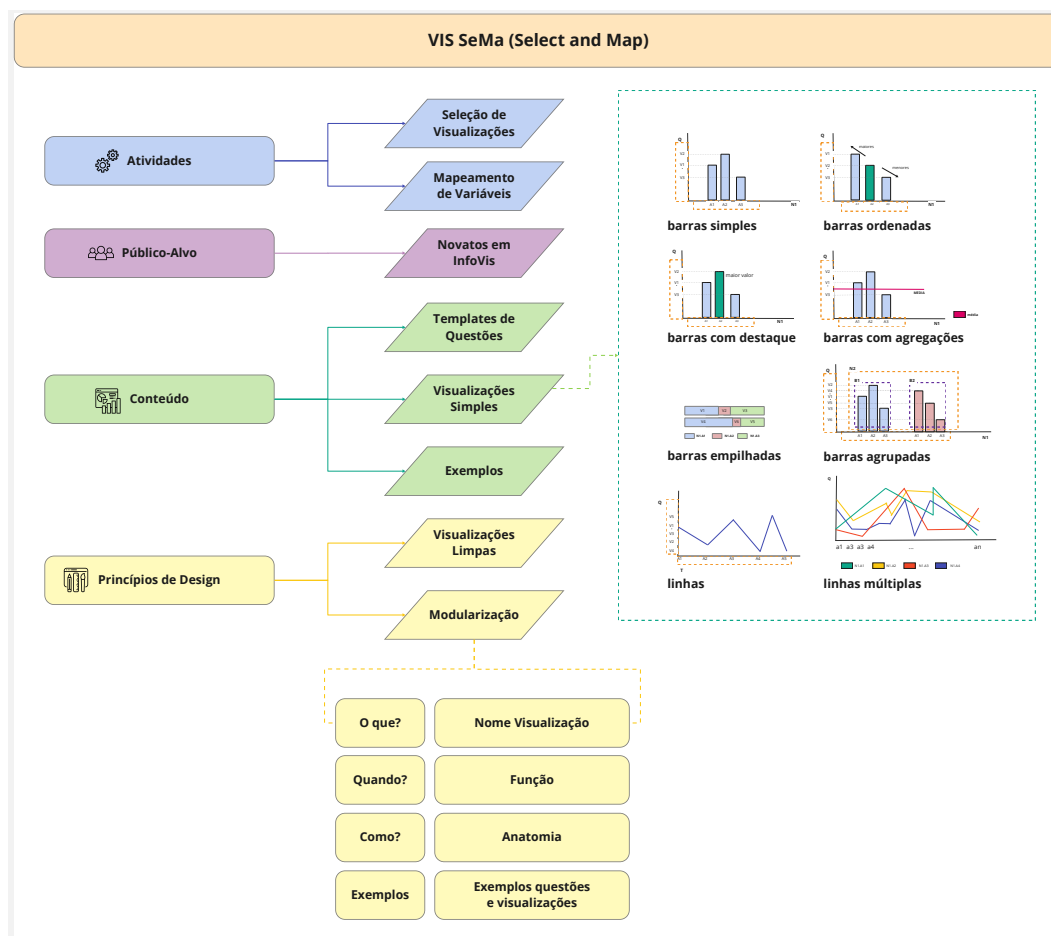


Figura 4.1: VIS-SeMa - Requisitos do Material proposto

4.2

Organização do VIS-SeMa

Na seção anterior foram definidos os requisitos para elaboração do material. Nesta seção detalhamos como é organizado o VIS-SeMa e apresentamos sua estrutura. O material pode ser acessado através de um navegador web. Para construir o VIS-SeMa, o pesquisador utilizou a ferramenta Figma,¹ onde foram escritas as instruções, imagens e conteúdos do material.

¹<https://www.figma.com>

A Figura 4.2 apresenta as páginas que compõem o material. Ao acessar o [VIS-SeMa](#), é apresentada a página inicial, com instruções para uso do material (número 1 da Figura 4.2). Nesta página de instruções, há um botão “Iniciar” que, ao ser selecionado, direciona o usuário para a página 2 do material (número 2 da Figura 4.2).

A página 2 apresenta os *templates* de questões organizados por categorias, onde o usuário deve selecionar uma questão. As categorias descrevem brevemente o objetivo do *template*, ou seja, o que o *template* retorna como resposta final. Ao clicar na questão escolhida, o usuário é direcionado para uma página semelhante à página 3 (número 3 da Figura 4.2). Ali o usuário tem acesso ao conteúdo explicativo sobre como realizar o mapeamento das variáveis para a visualização correspondente ao *template* selecionado na página 2.

O material proposto nesta pesquisa é apenas para consulta de conteúdos; o design das visualizações de fato não é realizado dentro do VIS-SeMa. Nos estudos que conduzimos nesta pesquisa, utilizamos a ferramenta Miro para apoiar a prototipação, mas ao utilizar o VIS-SeMa para selecionar a visualização e entender como utilizá-la, o autor das visualizações pode utilizar outras ferramentas para criar suas visualizações, como por exemplo, ferramentas de prototipação, ferramentas de criação de visualizações ou linguagens de programação e pacotes gráficos.

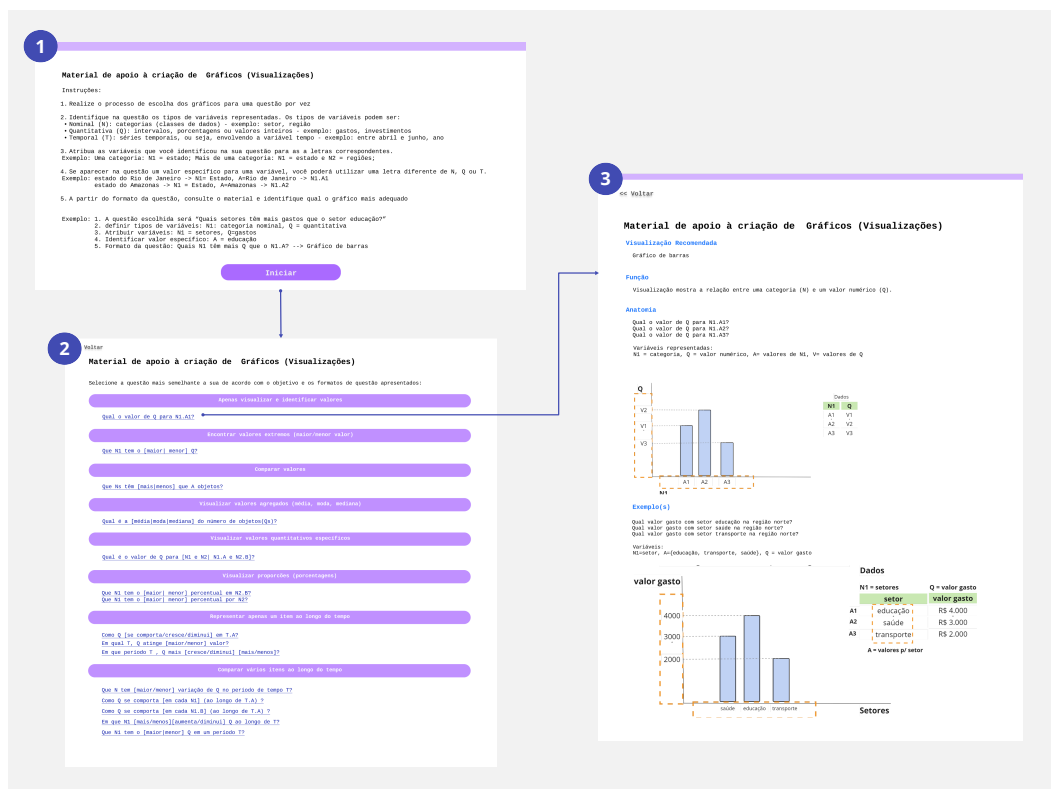


Figura 4.2: VIS-SeMa - Visão geral da estrutura do material proposto

4.3

Processo de Design de Visualizações utilizando VIS-SeMa

Para utilizar o material VIS-SeMa são necessárias algumas etapas anteriores e posteriores ao seu uso. Definimos um processo para apoiar o planejamento das visualizações que inclui a seleção de visualizações e mapeamento de variáveis para visualizações. Um cenário de uso do guia é apresentado na Seção C.1. A seguir são explicadas as etapas para uso do material.

Em primeiro lugar, enfatizamos que esse processo pressupõe que os dados já tenham sido explorados e que a questão de análise que se deseja responder já tenha sido definida.

E1. Análise da Questão: A primeira etapa do processo de design é a análise da questão. Nesta etapa, as questões definidas previamente são transformadas em *templates* parametrizados de questões. Para concluir esta etapa são necessárias as seguintes atividades: E1.1. identificação de variáveis na questão, E1.2. transformação das variáveis em parâmetros e E1.3. definição do *template*. Todas estas atividades são explicadas e exemplificadas resumidamente na página 1 do material VIS-SeMa. Apesar de conter instruções, a análise das questões não é realizada dentro do material, devendo ser realizada de forma manual em documentos auxiliares como editores de texto, anotações em ferramentas de prototipação (de livre escolha) ou como anotações em um quadro na ferramenta Miro (que foi a ferramenta utilizada nesta pesquisa para análise das questões).

E1.1. Identificação de Variáveis na Questão: Nesta atividade, o criador da visualização deve fazer a seguinte reflexão: “Quais informações desta questão devem estar representadas na visualização a ser criada?”. Como produtos dessa reflexão devem ser extraídas as variáveis das questões que representam os dados a serem trabalhados. Por exemplo, na questão “Quais os gastos das regiões do Brasil ao longo do ano de 2024?”, as variáveis a serem representadas são “gastos”, “regiões” e “anos”.

E1.2. Transformação de Variáveis em Parâmetros: A segunda atividade está relacionada a identificar o tipo das variáveis extraídas na atividade E1.1 e transformá-las em parâmetros. Os parâmetros são letras, definidas de acordo com seu tipo. Nesta versão inicial do material, incluímos visualizações que representam três tipos de dados; por esse motivo, consideraremos apenas esses três tipos como parâmetros: nominal, quantitativa (contínua) e temporal. O parâmetro que representa cada tipo de dado é representado por uma letra: Nominal - N, Quantitativa - Q e Temporal - T. Cada variável identificada na questão deve ser atribuída à letra correspondente ao seu tipo. Por exemplo, na questão “Quais os gastos das regiões do Brasil ao longo do ano de 2024?”,

a parametrização das variáveis é: N = regiões (variável nominal), Q = gastos (variável quantitativa) e T = ano (variável temporal). Caso a questão tenha mais de uma variável do mesmo tipo, devem ser utilizados números para diferenciá-las, como por exemplo: N1 = estado, N2 = região. Caso sejam especificados valores específicos para as variáveis na questão, eles devem ser atribuídos a outras letras (A1, A2, B1, B2, etc.), por exemplo: o valores para a variável região são Norte e Sul, as atribuições seriam N1 = Região, A1 = Norte e A2 = Sul; neste caso, o valor Região Sul seria representado por N1.A2.

E1.3. Definição do template: Nesta etapa, as letras identificadas na etapa anterior devem ser substituídas na questão. Desta forma, são criados os *templates* parametrizados das questões. Por exemplo, na questão “Quais os gastos das regiões do Brasil ao longo do ano de 2024?”, onde N = regiões do Brasil (variável nominal), Q = gastos (variável quantitativa) e T = ano (variável temporal), o *template* da questão seria: “Quais os Q de N em T?”, onde os parâmetros são Q, N e T.

E2.Seleção da visualização: Nesta etapa realizamos a atividade de seleção de visualização mais adequada para responder determinada questão. A etapa tem como insumos os *templates* parametrizados gerados na etapa anterior a partir da análise das questões. Para facilitar o processo, recomendamos repetir esta etapa para uma questão por vez. Para selecionar a visualização, a pessoa deve acessar a página 2 do VIS-SeMa e comparar o seu *template* com os templates de questão apresentados no material VIS-SeMa. A comparação é feita manualmente pelo criador da visualização. Após identificar o *template* semelhante ao seu, com mesmo objetivo e mesmos parâmetros (tipos de dados). Ele deve então selecionar, através de clique, o *template* escolhido, para em seguida visualizar a recomendação de visualização que deve utilizar, com informações para guiar o mapeamento de suas variáveis nos elementos da visualização. Na próxima etapa, explicamos como realizar este mapeamento.

E3.Mapeamento de variáveis para visualização: No mapeamento de variáveis, já com a visualização selecionada na etapa anterior, o criador da visualização terá acesso ao conteúdo do material para ajudá-lo a mapear as suas variáveis para os elementos da visualização. Conforme explicado anteriormente, o material é modularizado, sendo dividido nos seguintes tópicos: nome da visualização recomendada, função, anatomia e exemplos. O nome e a função da visualização são informações para identificar a visualização e seus objetivos de forma resumida. Na anatomia são apresentados os *templates* parametrizados das questões e como associar cada parâmetro a cada elemento da visualização. Também é mostrada uma tabela de dados para que o usuário entenda quais dados estão representados nos *templates* da questão e da

visualização, facilitando o mapeamento de variáveis e valores do dado para os elementos correspondentes. Além da anatomia, o material também apresenta um exemplo de uso na prática da visualização, respondendo questões concretas. Como mencionado, o VIS-SeMa apoia o mapeamento das variáveis através dos conteúdos visuais e explicativos, mas o protótipo (“desenho”) da visualização não é feito no VIS-SeMa. Nos estudos deste trabalho, utilizamos a ferramenta Miro para fazer os protótipos: fornecemos um conjunto de visualizações que o participante poderia copiar, colar e editar para fazer seu mapeamento. Apesar de utilizarmos o Miro, destacamos que os protótipos podem ser construídos com a ferramenta e da forma que o usuário preferir.

Relacionando as páginas do material VIS-SeMa com as etapas do processo de design das visualizações, detalhamos que a página 1 (instruções) apresenta instruções (explicações textuais e exemplos) sobre como realizar a etapa E1. A página 2 (categorias e templates) lista os *templates* de questões organizadas por categorias e auxilia na escolha da visualização, estando associada com a etapa E2. Por fim, a página 3 do material, apresenta o conteúdo modularizado sobre a visualização que irá apoiar o mapeamento das variáveis na visualização, associado à etapa E3.

4.4

Como Expandir o VIS-SeMa

Como explicado anteriormente, a versão inicial do VIS-SeMa é composta por visualizações simples e mais conhecidas. No entanto, outros pesquisadores ou instrutores podem utilizar a estrutura deste material para criar conteúdo sobre outros tipos de visualização. Apresentamos, a seguir, um guia para apoiar a adição de outras visualizações ao material. Um cenário de uso do guia é apresentado na Seção C.2. As etapas descritas são ilustradas na Figura 4.3.

A primeira etapa para utilizar o material para representar outras visualizações é escolher quais visualizações se deseja incluir. Em seguida, deve-se consultar o material proposto por Rodrigues et al. (2021) e verificar se há questões para a visualização que deseja adicionar. Caso haja, deve-se selecionar as questões correspondentes. Caso não haja, a pessoa deve criar seus *templates* seguindo as seguintes regras definidas por Rodrigues et al. (2021) para definição dos parâmetros, explicadas anteriormente:

- Os nomes das variáveis das questões são representados pela inicial dos nomes de seus tipos em inglês: N para variáveis nominais, Q para variáveis numéricas contínuas (quantitativas), T para variáveis temporais, S para variáveis espaciais e Objs para objetos (instâncias);

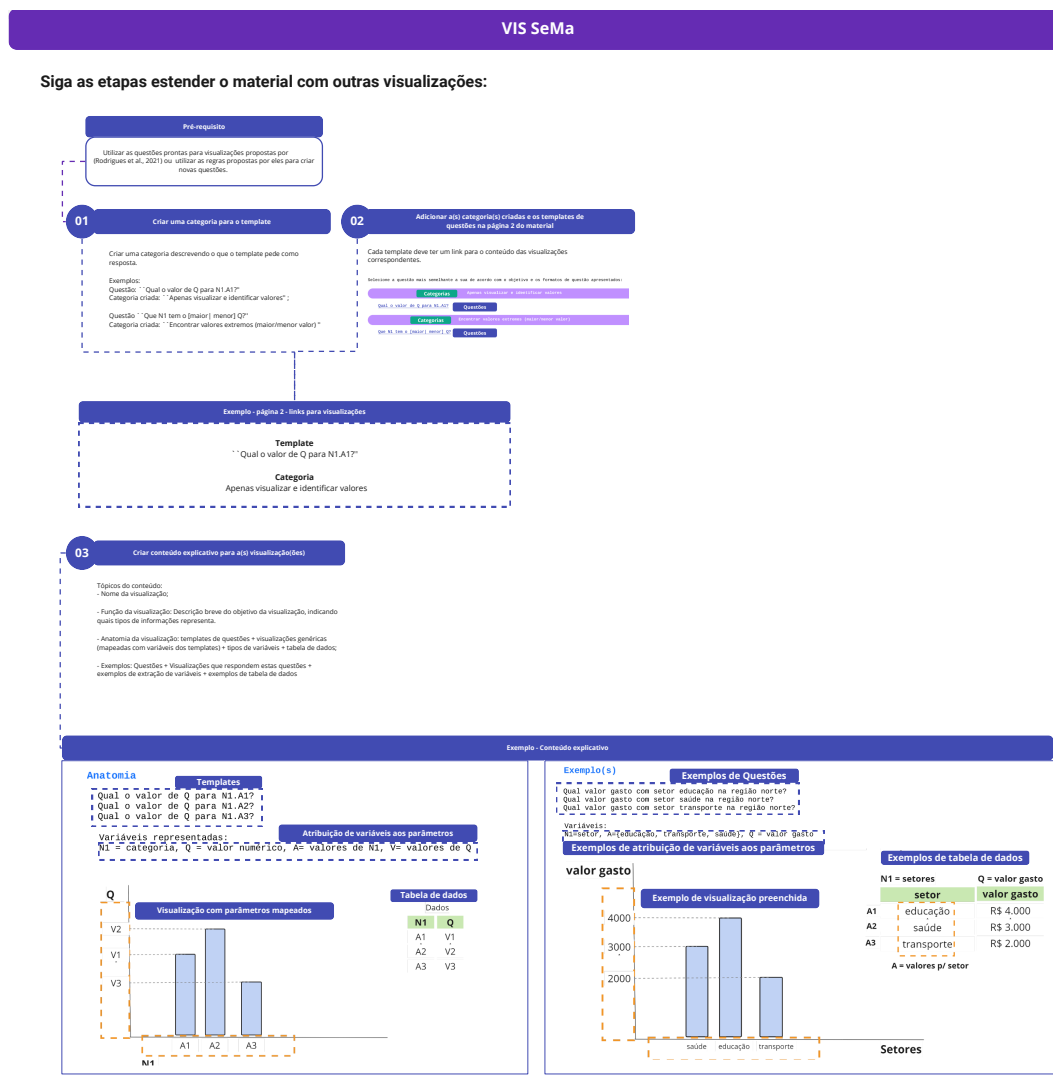


Figura 4.3: VIS-SeMa - Etapas para ampliar o material para outras visualizações

- Algumas variações de questões também podem ser utilizadas como parâmetros. Exemplo: “Qual o maior valor?” e “Qual o menor valor?”, onde os parâmetros são “maior” e “menor”;
- Quando diferentes variáveis do mesmo tipo são representadas na mesma questão, devem ser diferenciadas com acréscimo de um número. Exemplo: três variáveis nominais devem ser representadas como N1, N2 e N3;
- Se aparecer na questão um valor específico para uma variável, deverá utilizar um ponto seguido de outras letras diferentes das letras dos tipos (N, Q, T, S). Exemplo: N=Estado, A1=Rio de Janeiro, A2=Amazonas; estado do Rio de Janeiro → N.A1; estado do Amazonas → N1.A2;
- Expressões entre colchetes indicam opções dentre as quais o usuário deve selecionar uma para completar a questão. Exemplo: o template “Qual

[menor|maior] valor de Q?” pode ser uma das duas opções: “Qual menor valor de Q?” e “Qual maior valor de Q?”.

Após criado o *template*, deve-se criar uma categoria para esse *template* de questões descrevendo rapidamente seu objetivo, ou seja, o que a questão representada pede como resposta. Exemplo: Para a questão “Qual o valor de Q para N1.A1?” a categoria criada foi “Apenas visualizar e identificar valores” ; Para a questão “Qual N1 tem o [maior| menor] Q?” a categoria criada foi “Encontrar valores extremos (maior/menor valor) ”. Por fim, a categoria e o *template* devem ser adicionados à listagem de seleção das questões (página 2 do material - ver Figura 4.4).

VIS SeMa

Selecione a questão mais semelhante a sua de acordo com o objetivo e os formatos de questão apresentados:

Categorias	Questões
Apenas visualizar e identificar valores	Qual o valor de Q para N1.A1?
Encontrar valores extremos (maior/menor valor)	Que N1 tem o [maior menor] Q?
Comparar valores	Que Ns têm [mais menos] que A objetos?
Visualizar valores agregados (média, moda, mediana)	Qual é a [média moda mediana] do número de objetos(Qs)?
Visualizar valores quantitativos específicos	Qual é o valor de Q para [N1 e N2 N1.A e N2.B)?
Visualizar proporções (porcentagens)	Que N1 tem o [maior menor] percentual em N2.B? Que N1 tem o [maior menor] percentual por N2?

Diagrama de conexão:

- As sugestões de texto "Apenas visualizar e identificar valores", "Encontrar valores extremos (maior/menor valor)" e "Comparar valores" estão conectadas ao campo "Categorias".
- As sugestões de texto "Visualizar valores agregados (média, moda, mediana)", "Visualizar valores quantitativos específicos" e "Visualizar proporções (porcentagens)" estão conectadas ao campo "Questões".

Figura 4.4: VIS-SeMa - Página 2 do Material onde são organizados os templates e as categorias

A próxima etapa é a criação do conteúdo do material, onde devem ser preenchidas as informações sobre a visualização. A seguir explicamos como preencher cada tópico.

1. Visualização recomendada: informe o nome da visualização (ver Figura 4.5).
2. Função: Descreva brevemente o objetivo da visualização, indicando os tipos de dados que devem ser nela representados (ver Figura 4.5).
3. Anatomia: Adicione o(s) *template(s)* de questões criados, explique o que deve ser preenchido em cada parâmetro, adicione uma tabela de dados ilustrativa, preenchida com os parâmetros, adicione uma imagem da visualização e indique na imagem qual parâmetro deve ser mapeado para cada elemento da visualização (ver Figura 4.6).

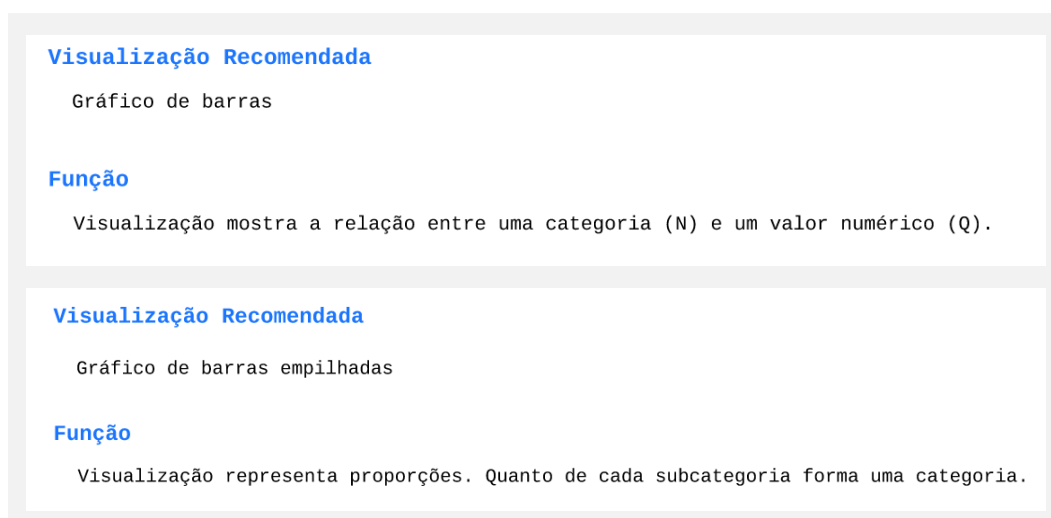


Figura 4.5: VIS-SeMa - Página 3 do material: Conteúdo - Nome da Visualização e Função para gráfico de barras e gráfico de barras empilhadas

4. Exemplos: Adicione exemplos de questões, mostre qual variável da questão deve ser atribuída a cada parâmetro, adicione uma tabela com exemplos de dados que se tem disponível para a visualização e adicione uma imagem da visualização com as variáveis representadas (ver Figura 4.7).

4.5 Considerações

Este capítulo apresentou o material VIS-SeMa (*Select and Map*) que elaboramos para apoiar novatos no processo de design de visualizações, com foco nas atividades de seleção de visualizações e mapeamento de variáveis em visualizações. O material é baseado em *templates* parametrizados de questões e possui conteúdos para consulta sobre escolha e uso de algumas visualizações. Na versão apresentada, tratamos apenas alguns tipos de visualizações, pois o foco da pesquisa é avaliar visualizações mais simples que, apesar de serem mais conhecidas, ainda apresentam erros em sua criação, de acordo com resultados do nosso estudo de observação. Apesar disso, apresentamos um guia para que outros pesquisadores possam adicionar conteúdo de outras visualizações ao material.

O material é estruturado de forma modularizada e apresenta a combinação de *templates* genéricos de questões e visualizações, exemplos e representações visuais concretas para apoiar a compreensão do processo por novatos. Em seu conteúdo, VIS-SeMa apresenta poucas instruções visuais associadas às visualizações para que seja lida e utilizada de forma simples e rápida. No próximo capítulo, apresentamos um estudo de viabilidade que realizamos para avaliar o material proposto.

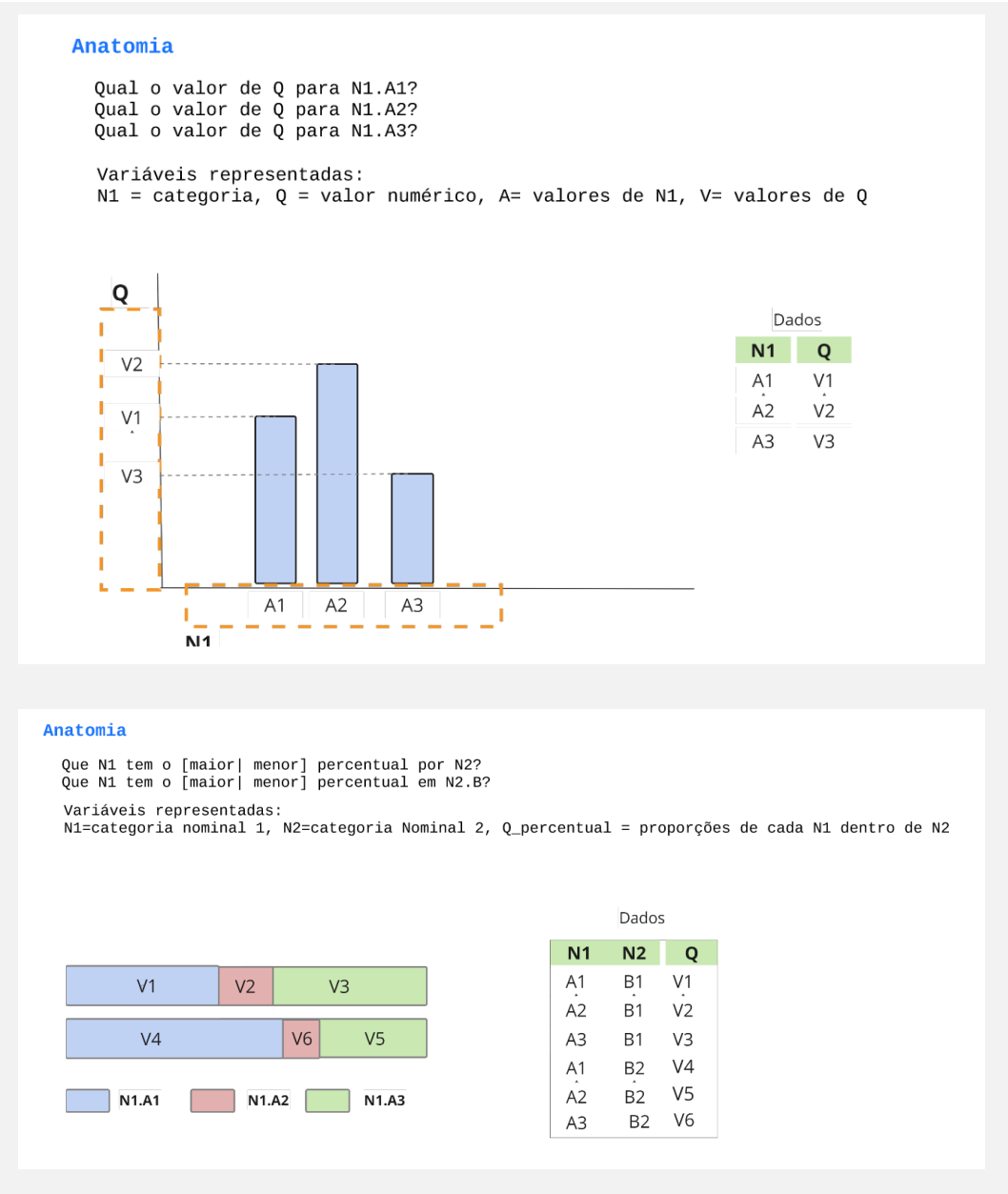


Figura 4.6: VIS-SeMa - Página 3 do material: Conteúdo - Anatomia para gráfico de barras e gráfico de barras empilhada



Exemplo(s)

Que setor tem [maior/menor] percentual de gastos por região?
Que setor tem [maior/menor] percentual na região Norte?

Variáveis:
N1=setor, N2=região, Q_percentual = percentual de gastos

Dados

N1 = setor	N2 = região	Q = percentual gasto
setor	região	percentual gasto
educação	norte	40%
saúde	norte	20%
transporte	norte	40%
educação	sul	65%
saúde	sul	10%
transporte	sul	25%

A = valores p/ setor B = valores p/ setor

Região	Setor	Percentual
Norte	educação	40%
	saúde	20%
	transporte	40%
Sul	educação	65%
	saúde	10%
	transporte	25%

Figura 4.7: VIS-SeMa - Página 3 do material: Conteúdo - Exemplos para gráfico de barras e gráfico de barras empilhadas

5

Estudo de viabilidade do VIS-SeMa

Após a elaboração da proposta, realizamos um estudo de viabilidade para avaliar o uso do material VIS-SeMa (*Select and Map*) por novatos em visualização de informação e sua estrutura. Esse estudo foi guiado pelas seguintes questões de pesquisa:

- Q1. De que forma o material VIS-SeMa apoiou a seleção das visualizações e o mapeamento de variáveis para visualização?
- Q2. Quais problemas surgiram durante o processo de seleção e mapeamento de variáveis?

5.1

Contexto e Participantes do Estudo

O estudo foi realizado de forma presencial em laboratório de Informática, durante uma aula da disciplina de Interação Humano-Computador lecionada por um professor não associado a este trabalho. As atividades do estudo foram realizadas no computador, as atividades de criação de visualizações foram realizadas na ferramenta Miro (a mesma utilizada no estudo anterior). Todo o estudo foi acompanhado pelo professor da turma. O professor recebeu instruções prévias para a condução das atividades em sala de aula. O estudo durou em torno de 2 horas (tempo de duas aulas da disciplina). O cenário utilizado no estudo foi criado pela autora desta tese. O professor informou que tinha ministrado aulas de noções básicas de tipos de dados. Os participantes criaram as visualizações individualmente.

Os participantes do estudo foram 28 alunos do terceiro período do curso de Ciência da Computação. Os alunos estavam cursando a disciplina de Interação Humano-Computador. Antes da realização do estudo, todos os participantes receberam um termo de consentimento livre e esclarecido, no qual registraram sua aceitação em participar do estudo nas condições oferecidas. No termo, estava esclarecido que esta pesquisa não avaliaria os participantes, que seu anonimato seria garantido e que eles poderiam recusar-se a participar da pesquisa ou interrompê-la a qualquer momento.¹

¹Este estudo faz parte de um projeto de pesquisa submetido e aprovado previamente pela Câmara de Ética em Pesquisa da PUC-Rio (protocolo #97/2020).

5.2

Materiais do Estudo

Como citado anteriormente, os participantes receberam materiais para o trabalho e realizaram as atividades utilizando um quadro criado na ferramenta Miro.² Além do quadro no Miro, os participantes também receberam um link para acessar o material VIS-SeMa. Após a realização da atividade, os participantes responderam a um questionário de *feedback*. Dos 28 participantes, 15 responderam ao questionário. O quadro criado no Miro era composto por um conjunto de visualizações, um cenário textual, um conjunto de cinco questões e um exemplo de como realizar a atividade.

- **Cenário e Questões:** O cenário descrevia um contexto onde uso de visualizações seria necessário e um conjunto de cinco questões a serem respondidas com as visualizações; O contexto estava relacionado à gestão dos produtos de uma fazenda.

“FarmCity é uma fazenda que produz vários tipos de suprimentos. Seu dono precisa administrar os recursos utilizados e suprimentos produzidos para manter cada setor de produção funcionando da melhor forma.

Na fazenda existem 3 setores de produção: animal, vegetal e mineral.

Cada setor de produção fornece produtos do tipo 1, tipo2, tipo 3, tipo 4.

O dono registra o dinheiro arrecadado pela fazenda e o dinheiro gasto (gastos) para manter os setores de produção. Para ajudá-lo a administrar a produção da fazenda, o dono precisa elaborar gráficos para responder algumas questões, apresentadas a seguir:”

- Q1. Qual a quantidade de suprimentos produzidos de cada setor?
- Q2. Qual tipo de produto tem maior percentual de produção para cada um dos setores?
- Q3. Qual setor teve mais aumento de quantidade de produção ao longo do ano?
- Q4. Qual a quantidade de suprimentos produzidos de cada tipo para cada setor de produção?
- Q5. Qual é a média de quantidade de produção de cada setor?

- **Visualizações:** O quadro criado no Miro apresentava o mesmo conjunto de visualizações apresentado no VIS-SeMa para facilitar aos participantes fazer cópias das visualizações e fazer edições para criar seus protótipos. Desta forma, não precisariam desenhar as visualizações manualmente ou em uma ferramenta externa.

²Miro: <https://miro.com>

- **VIS-SeMa:** O material de apoio à criação de visualizações elaborado nesta pesquisa.

5.3

Etapas do Estudo

O estudo foi dividido em duas atividades principais: seleção da visualização para responder à questão e mapeamento de variáveis das questões para os elementos da visualização. Para apoiar a escolha da visualização e o mapeamento das variáveis, os participantes utilizaram o material VIS-SeMa.

Antes do estudo, os participantes receberam do professor da disciplina todas as instruções para a realização da atividade. Para iniciar a atividade, os participantes receberam acesso ao quadro do miro e cada participante fez uma cópia do quadro para que pudesse utilizar para criar suas visualizações. No quadro, os participantes deveriam ler o cenário e as questões e, para cada questão dada, deveriam seguir as instruções da primeira página do VIS-SeMa. Na primeira página havia uma explicação sobre como identificar os tipos de dados e transformar as questões em *templates*. Após seguirem as instruções, os participantes deveriam ter gerado um *template* para cada questão. Em seguida, os participantes deveriam prosseguir para a segunda página do material e verificar qual *template* do material mais se assemelha ao *template* criado por eles. Após selecionar o *template*, o participante era direcionado para o conteúdo da visualização associado ao *template* selecionado. A partir desta etapa, se inicia a atividade de mapeamento de variáveis para a visualização.

Na atividade de mapeamento, o participante deveria utilizar o conteúdo explicativo da visualização para associar as variáveis das questões dadas aos elementos da visualização. Para isso, eles deveriam consultar o material VIS-SeMa a fim de entender como mapear dados na visualização escolhida e registrar sua escolha e o mapeamento que julgasse correto no quadro do Miro. Ao final do processo, os participantes responderam a um questionário de *feedback* (disponível no Apêndice B) sobre o material utilizado. Somente 15 dos 28 participantes responderam ao questionário. O material gerado após o estudo foi um conjunto de 140 visualizações e 139 mapeamentos (um dos participantes não descreveu um dos seus mapeamentos).

5.4

Análise dos resultados

A análise dos resultados foi realizada de três formas: (1) análise quantitativa e (2) análise qualitativa do *feedback* dos usuários, e (3) análise de discrepâncias nas visualizações produzidas pelos participantes.

5.4.1 Análise Quantitativa

Para esta análise utilizamos as questões 1 e 2 do questionário de *feedback* disponível no Apêndice B. O questionário foi aplicado em um momento posterior às aulas e respondido por 15 dos 28 participantes do estudo.

A primeira questão era relacionada à frequência com que os participantes utilizaram cada elemento do material VIS-SeMa durante a realização da atividade. O resumo dos resultados é apresentado na Figura 5.1. De acordo com os resultados, pudemos observar que os itens mais utilizados pelos participantes durante as atividades foram: os exemplos de gráficos, os gráficos apresentados (com as variáveis) e os exemplos de questões. Os elementos menos consultados pelos participantes foram: as instruções na página inicial e os conteúdos explicativos. Por falar nisso, os elementos que representavam informações visuais foram mais utilizados que os elementos que apresentaram informações textuais.

Com que frequência foi utilizado cada item do material VIS SeMa?

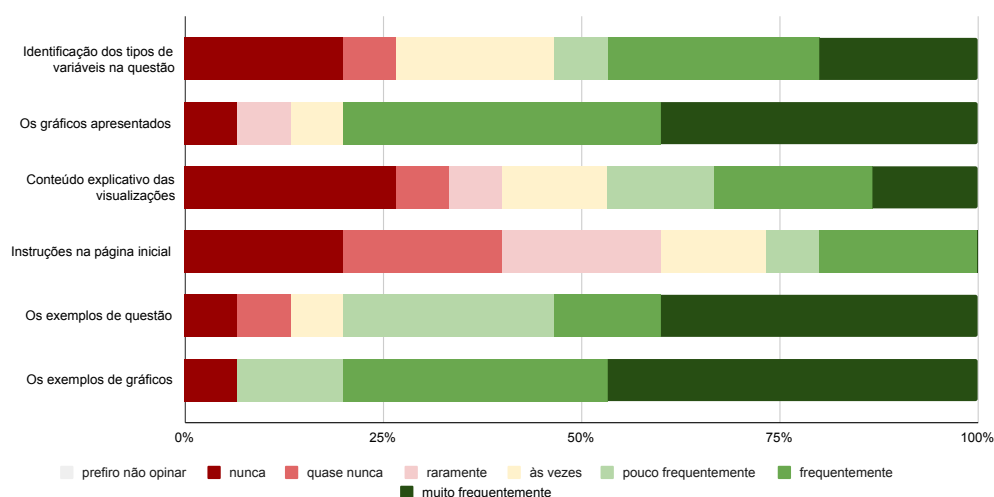


Figura 5.1: Frequência de utilização de cada elemento do VIS-SeMa

A segunda questão era relacionada a quanto cada tópico do material ajudou na escolha das visualizações mapeamento das variáveis. Um resumo das respostas dos participantes é apresentado na Figura 5.2. A partir destes resultados, observamos que, novamente, os exemplos de gráficos, os gráficos apresentados (com as variáveis) e os exemplos de questões foram os itens que mais ajudaram nas atividades. E os itens que não ajudaram muito foram as instruções na página inicial, os conteúdos explicativos e a identificação dos tipos de variável. Em relação à organização do material e os termos utilizados, a maioria dos participantes marcou que não teve dificuldades.

Quais itens do material VIS SeMa mais ajudaram a construir as visualizações?

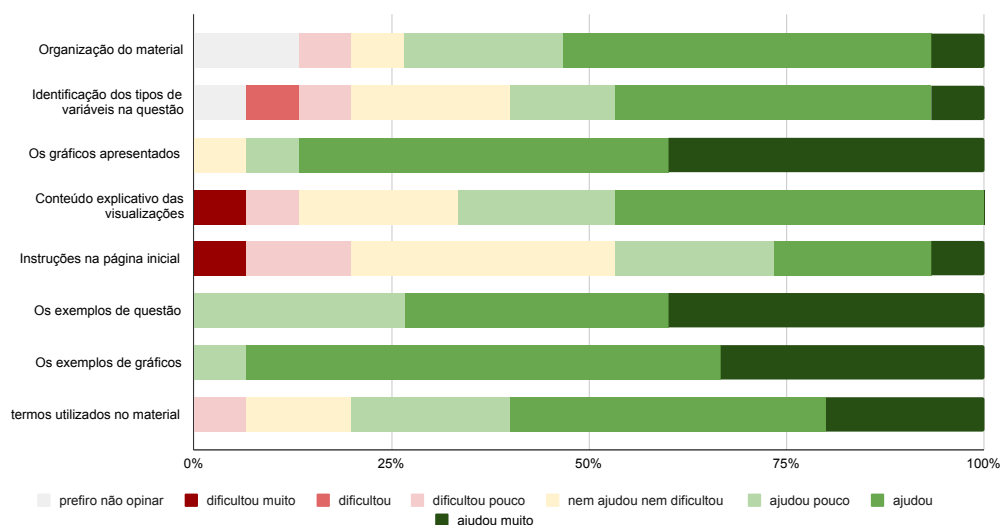


Figura 5.2: Respostas dos participantes sobre quanto os elementos do material ajudaram nas atividades de escolha do gráfico e no mapeamento das variáveis

5.4.2

Análise Qualitativa

Para a análise qualitativa, utilizamos as respostas às questões abertas do questionário. A partir disto, coletamos respostas sobre os tópicos do material. O pesquisador transcreveu as respostas ao questionário e agrupou de acordo com as opiniões relacionadas a cada elemento trabalhado no material: organização do material, gráficos apresentados, conteúdo explicativo dos gráficos, utilidade das instruções, exemplos de questões e exemplos de visualizações. A seguir apresentamos a análise realizada.

O primeiro item investigado foi a **organização do material**. Em relação a esse item, os participantes disseram que o conteúdo do material estava bem organizado e simples, mas a interface utilizada necessita de melhorias:

“A forma que o material estava organizado facilitou muito na hora da procura do gráfico ideal” - P24

“Bem organizado e direto ao ponto” - P02

“Bem simples, sem muita complexidade para uso” - P06

“Está bem organizado, cada tópico refere ao que está sendo abordado, mas senti falta de alguns elementos mais avançados para deixar o material mais intuitivo” - P11

“Eu não gostei muito da paleta de cores da página (...) Mas o material é organizado, você facilmente consegue achar o que procura com palavras chaves” - P17

Em relação à **explicação dos tipos de variáveis** apresentada na página de instruções, alguns participantes citaram dificuldades no entendimento, mas outros acharam que ajudou no mapeamento. As citações de alguns participantes são apresentadas a seguir.

“Relativamente difícil de relacionar os tipos das variáveis com o que o enunciado pedia” - P2

“(...) sugere bons tipos de variáveis, aplicáveis a vários cenários” - P8

“As variáveis estão bem definidas, ainda mais que estão sendo demonstradas com um quadrado pontilhado nos gráficos” - P11

“Os tipos de variáveis são bem explicativas e diretas, possuindo uma facilidade de compreensão” - P15

Sobre a **linguagem utilizada no material**, os participantes não apresentaram dificuldades:

“Achei que estava escrito em uma linguagem de fácil entendimento” - P17

“Foram termos simples e básicos, não fiquei confuso com relação a isso durante o trabalho” - P24

“linguagem simples, compreensível” - P01

Outro item avaliado foram os gráficos apresentados no material, presentes no tópico **anatomia**. Os participantes citaram que **as visualizações estavam bem organizadas e ajudaram a encontrar a opção mais adequada para representar a questão**. As citações de alguns participantes são apresentadas a seguir.

“Foram bem organizados mas eu gostaria de ter mais de 1 exemplo por gráfico, acho que facilitaria mais o entendimento” - P06

“Os gráficos estão demonstrados de forma clara para entendimento de maneira simples” - P11

“Ajudaram bastante a ter uma ideia sobre qual gráfico utilizar para cada tipo de dado” - P20

“Os gráficos eram bem abrangentes e, mesmo possuindo mais de uma resposta certa, sempre foi possível encontrar o gráfico ideal” - P24

Em relação ao **conteúdo explicativo dos gráficos** os participantes citaram que:

“ajuda bastante na hora de compreender os exemplos genéricos” - P01

“Sem o conteúdo já teria uma leve compreensão, somado com o conteúdo facilita mais ainda” - P15

“Elas ajudam a entender por que x gráfico foi escolhido para representar uma questão, eloquente o suficiente que até mesmo inicialmente não concordando com a escolha do gráfico, eu compreenda por quê ele foi escolhido depois

da leitura” - P17

“Não utilizei muito, para mim a parte visual foi mais útil” - P24

Outro ponto investigado foi a **utilidade das instruções**. Os participantes não acharam as instruções atrativas por serem textuais; preferiram trabalhar utilizando as visualizações e os exemplos do material. As citações de alguns participantes são apresentadas a seguir.

“Muito texto, seria interessante ter mais imagens ou algo mais intuitivo” - P08

“As instruções estão enumeradas de forma que ajuda a compreender, no entanto sinto que falta algo que ajude ainda mais a entender as instruções de forma hierárquica e seja mais intuitivo para ler” - P11

“Vou admitir que inicialmente eu não li. Era muito texto em pouco espaço, ficou um bloco. Achei melhor aprender na prática do que ler” - P17

“Não visualizei as instruções iniciais” - P18

“só usei pra ver o nome das variáveis, não li mais nada” - P28

Sobre **os exemplos das questões** apresentados no material, os participantes citaram que ajudaram nas atividades de criação das visualizações. As citações de alguns participantes são apresentadas a seguir.

“Ajudaram bastante para o desenvolvimento das atividades” - P15

“Os exemplos me ajudaram muito na hora de encontrar o gráfico ideal, pois pude comparar a minha situação com o exemplo” - P24

“facilitaram para fazer a atividade” - P28

Por fim, coletamos a **opinião dos participantes sobre os exemplos de visualização**. Observamos que os exemplos ajudaram a selecionar as visualizações mais adequadas e a entender os tipos de variáveis, e não causaram dificuldades no entendimento. Além disso, um dos participantes citou que gostaria que fosse apresentado mais de um exemplo. As citações de alguns participantes são apresentadas a seguir.

“são bons gráficos, abrangem uma boa quantidade de informações” - P01

“Bons exemplos, mas eu gostaria de mais de 1 exemplo de uso por categoria” - P06

“Gráficos bons, apresentados genericamente pra associar com outros” - P8

“Parece que falta alguma informação para complementar os gráficos, no entanto todos eles estão demonstrados de forma clara” - P11

“Eles me ajudaram a entender os tipos de variáveis da questão” - P17

“Ajudaram bastante a ter uma ideia sobre qual gráfico utilizar para cada tipo de dado” - P20

“tinham gráficos que nunca vi na vida, mas consegui entender bem” - P21

“Os exemplos de gráfico me ajudaram muito pois pude selecionar o gráfico mais parecido dentre os possíveis no trabalho” - P24

5.4.3

Análise de Discrepâncias

Nesta etapa, analisamos todas as visualizações criadas pelos participantes para verificar discrepâncias na escolha das visualizações e nos mapeamentos criados. O material foi avaliado pelo pesquisador (chamaremos aqui de primeiro pesquisador) que planejou o estudo e elaborou o material e a análise foi revisada por um segundo pesquisador (chamaremos aqui de segundo pesquisador) que apoiou na concepção do material. No total foram elaboradas 140 visualizações e 139 mapeamentos (um dos participantes não fez o mapeamento de uma das suas visualizações). Consideramos como discrepâncias escolhas de visualizações e mapeamentos que não apresentavam o que era esperado que os participantes criassem no contexto das perguntas apresentadas no estudo. Esclarecemos que, independentemente da forma como as visualizações e mapeamentos fossem registrados, caso estivessem corretos, não foram considerados discrepâncias. O primeiro pesquisador organizou as informações das visualizações e mapeamentos criados em uma planilha, onde registrou os seguintes itens: a identificação do participante (P1, P2, P3 etc.), a questão correspondente (Q1, Q2, Q3 etc.), a visualização escolhida, a classificação da visualização (correto, incorreto, incompleto, parcialmente correto), o mapeamento relacionado à visualização e a classificação do mapeamento (correto, incorreto, incompleto, parcialmente correto). Após a classificação inicial, o primeiro pesquisador selecionou as visualizações e mapeamentos classificados como incorretos, incompletos e parcialmente corretos e identificou que tipos de problemas apresentaram. Esses problemas deram origem aos grupos de discrepâncias. As classificações do material e das discrepâncias foram apresentadas para o segundo pesquisador, para revisão dos resultados.

A classificação utilizada para avaliar o material é explicada a seguir: correto - para seleção de visualizações e mapeamentos que não apresentaram discrepâncias; incorreto - para seleção de visualizações e mapeamentos que continham erros; incompleto: quando participantes não completavam a elaboração do mapeamento mas que o tinham elaborado não continha erros; e parcialmente correto: quando os mapeamentos criados foram finalizados e apenas parte dele continha erros ou quando escolhiam uma visualização que não era a mais adequada mas era suficiente para representar os dados. As Figuras 5.3 e 5.4 resumem a classificação das visualizações e dos mapeamentos. A partir dos resultados, podemos notar que a maioria das visualizações e ma-

peamentos foram adequados para responder às questões e não apresentaram discrepâncias. Das 140 visualizações criadas, 85 estavam corretas e 55 apresentaram discrepâncias. Em relação aos mapeamentos, dos 139 mapeamentos criados 82 não apresentaram discrepâncias, 34 estavam totalmente incorretos, 15 estavam incompletos e 8 estavam parcialmente corretos. O participante que cometeu acima de 4 discrepâncias não tinha conhecimentos em visualização e nem experiência em projetos de software na indústria. A questão que gerou mais erros foi a questão “Q5. Qual é a média de quantidade de produção de cada setor?”, os participantes não levaram em consideração a criação de uma visualização para “cada setor” e representaram a média de produção de todos os setores juntos.

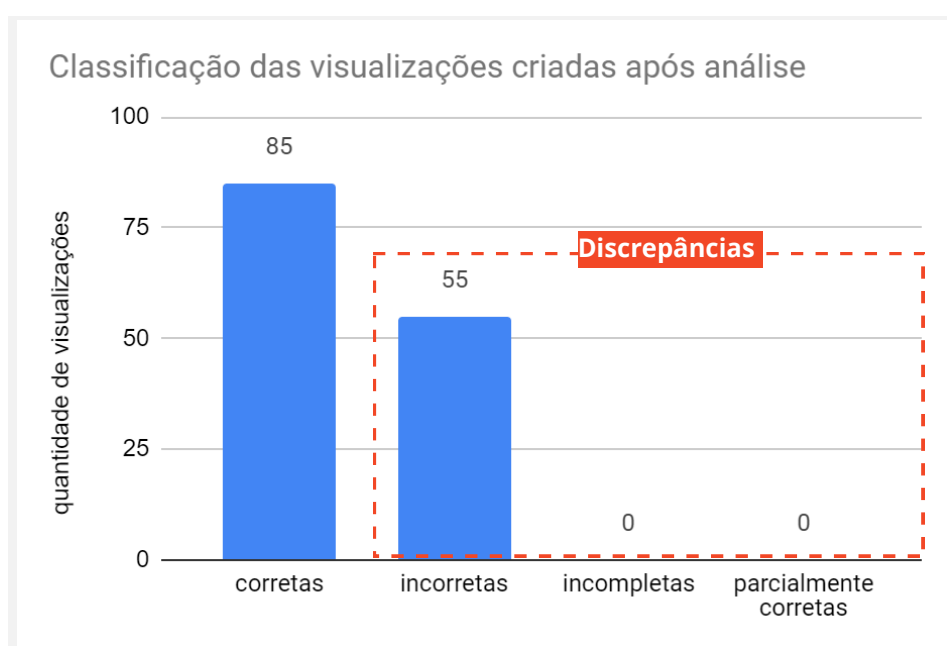


Figura 5.3: Classificação das visualizações criadas

Em relação às discrepâncias identificadas, como citado anteriormente, realizamos uma análise do material classificado (incorreto, parcialmente correto e incompleto), onde identificamos grupos de problemas que chamamos de discrepâncias. Esses grupos foram: elementos da visualização sem mapeamento de variáveis; falta de mapeamento de algumas variáveis; inversão no mapeamento das variáveis; mapeamento de duas variáveis para o mesmo elemento da visualização; mapeamento de uma variável para dois elementos da visualização; utilização incorreta da visualização, informações sem contexto; interpretação diferente da questão e escolha de visualização inadequada. A seguir, apresentamos os grupos de discrepância e alguns exemplos de visualizações criadas e associadas a cada grupo.

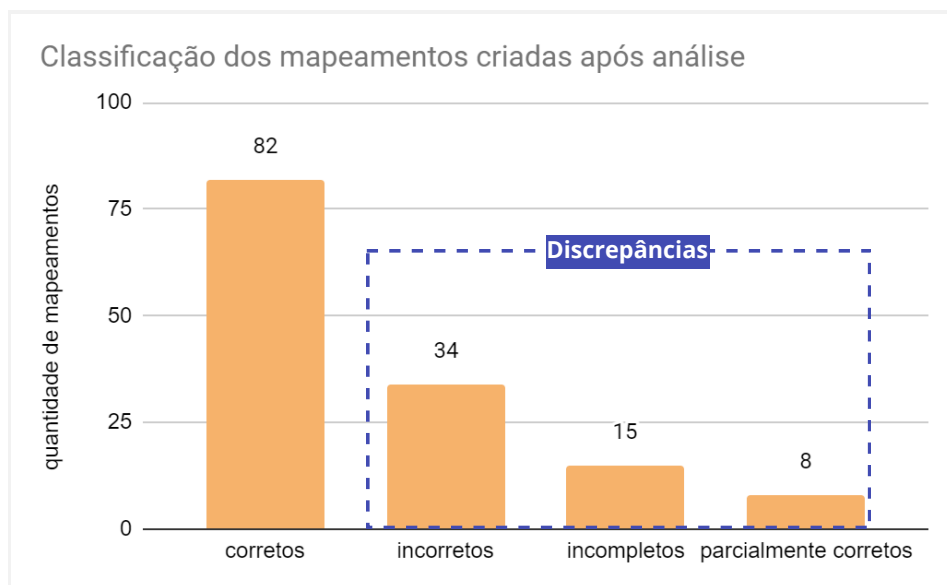


Figura 5.4: Classificação dos mapeamentos criados

Elementos da visualização sem mapeamento de variáveis: Ocorre quando um participante escolhe uma determinada visualização, mas não utiliza todos os seus elementos para representar os dados, ou seja, não representa o mapeamento dos atributos dos dados nas variáveis visuais. Por exemplo: alguns só usam os eixos horizontal e vertical, mas não atribuem nenhuma variável para as linhas em um gráfico de linhas. A seguir são apresentados alguns exemplos de visualizações em que ocorreu esta discrepância (veja as Figuras 5.5, 5.6).

- Questão: Q4. Qual a quantidade de suprimentos produzidos de cada tipo para cada setor de produção? (veja a Figura 5.5)
- Visualização: Gráfico de linhas
- Discrepância: P11 utilizou gráfico de linhas, mas não explicou qual variável representava o elemento linha.

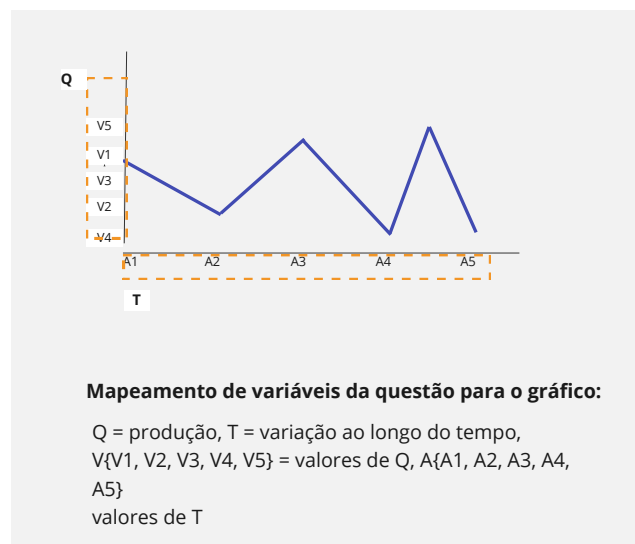


Figura 5.5: Exemplo 1 - elementos da visualização sem mapeamento de variáveis

- Questão: Q3. Qual setor teve mais aumento de quantidade de produção ao longo do ano? (veja a Figura 5.6)
- Visualização: Gráfico de linha
- Discrepância: P6 utilizou somente uma linha na visualização, sendo que a visualização pedia uma comparação entre vários setores, e não explicou qual variável está mapeada para esta linha.

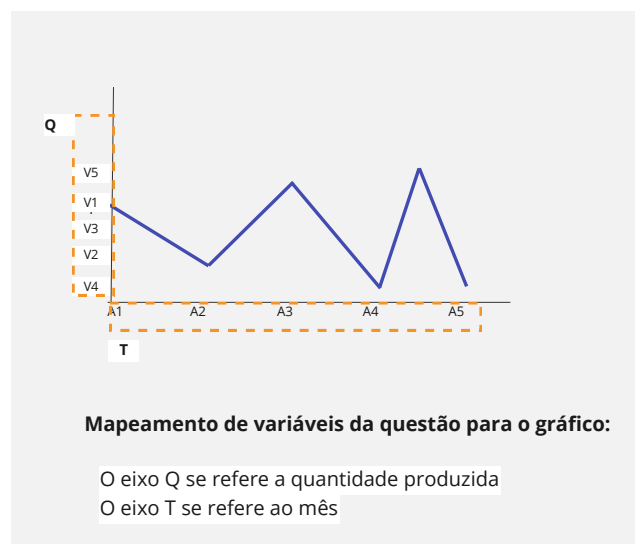


Figura 5.6: Exemplo 2 - elementos da visualização sem mapeamento de variáveis

Falta de mapeamento de variáveis necessárias para responder a questão: Neste item, os participantes não indicaram em seu mapeamento

algumas variáveis que eram necessárias para responder as questões. A seguir, apresentamos alguns exemplos deste tipo de discrepância.

- Questão: Q3. Qual setor teve mais aumento de quantidade de produção ao longo do ano? (veja a Figura 5.7)
- Visualização: Gráfico de linha
- Discrepância: P2 não mapeou a variável quantitativa para a visualização.

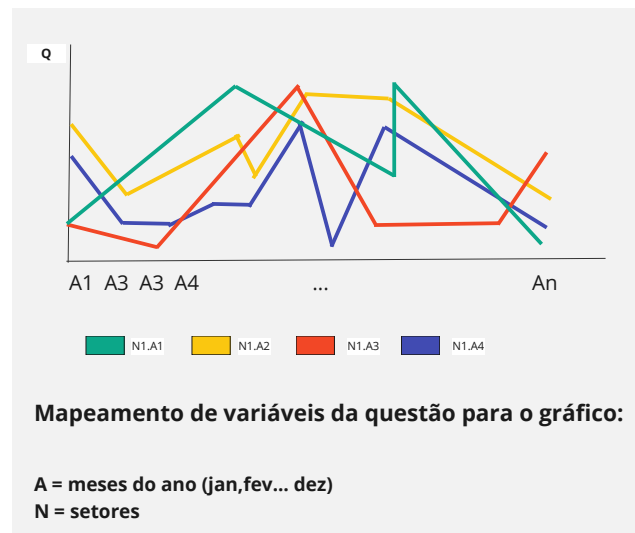


Figura 5.7: Exemplo 1 para variáveis necessárias que não foram mapeadas na visualização

- Questão: Q4. Qual a quantidade de suprimentos produzidos de cada tipo para cada setor de produção? (veja a Figura 5.8)
- Visualização: Gráfico de barras
- Discrepância: P3 não mapeou os tipos para cada setor, conforme pedia a questão.

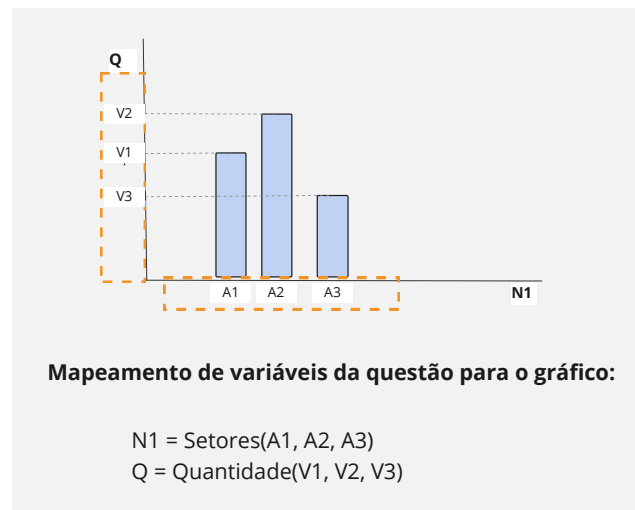


Figura 5.8: Exemplo 1 para variáveis necessárias que não foram mapeadas na visualização

- Questão: Q2. Qual tipo de produto tem maior percentual de produção para cada um dos setores? (veja a Figura 5.9)
- Visualização: Gráfico de barras empilhadas
- Discrepância: P21 não mapeou todas as variáveis para a visualização, somente tipos de suprimentos

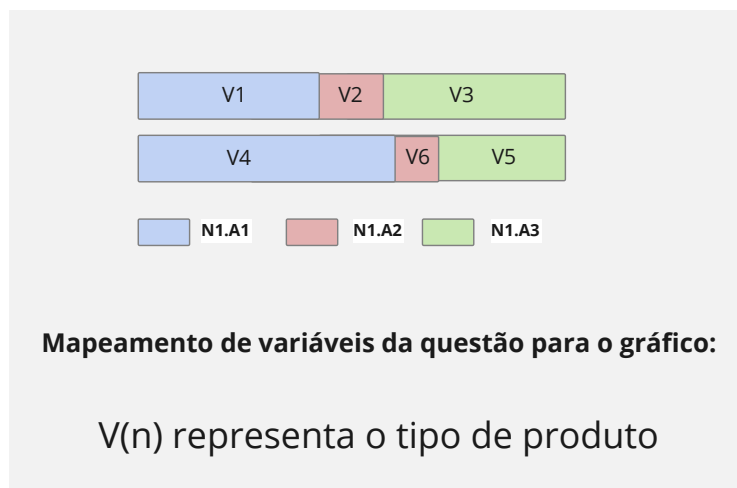


Figura 5.9: Exemplo 2 para variáveis necessárias que não foram mapeadas na visualização

Inversão no mapeamento das variáveis: Identificamos a inversão quando os participantes mapearam variáveis em elementos diferentes do que era esperado para responder questões específicas. Por exemplo: quando mapeavam a variável A para tamanho da barra e B para o eixo horizontal, mas o mapeamento correto seria variável A para eixo X e variável B para o tamanho da barra.

- Questão: Q2. Qual tipo de produto tem maior percentual de produção para cada um dos setores? (veja a Figura 5.10)
- Visualização: Gráfico de barras empilhadas
- Discrepância: P1 mapeou suprimentos onde deveria colocar os setores e vice e versa.

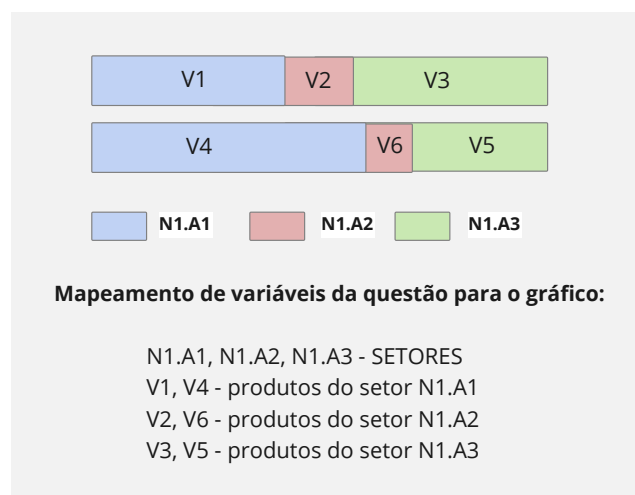


Figura 5.10: Exemplo 1 para inversão de variáveis

- Questão: Q4. Qual a quantidade de suprimentos produzidos de cada tipo para cada setor de produção? (veja a Figura 5.11)
- Visualização: Gráfico de barras agrupadas
- Discrepância: P12 mapeou setores para o eixo horizontal e tipos para os agrupamentos de barras mas, de acordo com a questão, deveria ser o contrário.

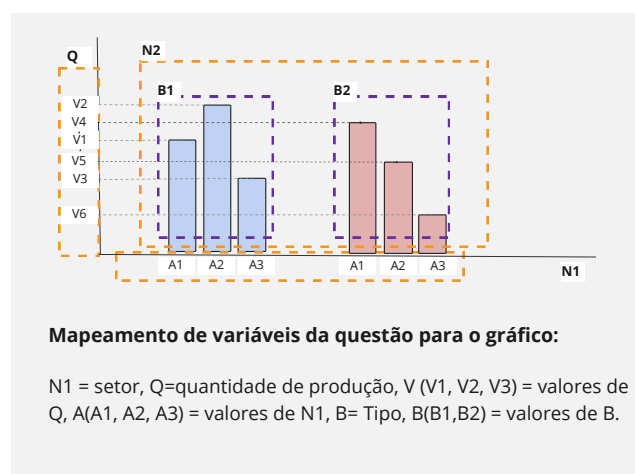


Figura 5.11: Exemplo 2 para inversão de variáveis

Mapeamento de duas variáveis para o mesmo elemento da visualização: Algumas vezes, os participantes mapearam duas variáveis distintas para o mesmo elemento da visualização, exemplo: mapeavam setores e suprimentos para o eixo x de um determinado gráfico.

- Questão: Q2. Qual tipo de produto tem maior percentual de produção para cada um dos setores? (veja a Figura 5.12)
- Visualização: Gráfico de barras empilhadas
- Discrepância: P25 mapeou setores e suprimentos para o mesmo elemento da visualização – as divisões por cores

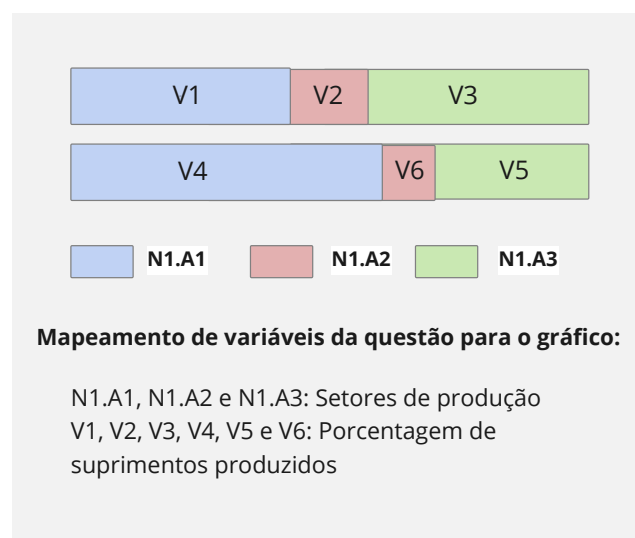


Figura 5.12: Exemplo 1 mapeamento de duas variáveis para o mesmo elemento do gráfico

Mapeamento de uma variável para dois elementos da visualização: Neste item, os participantes mapearam equivocadamente uma variável para dois elementos das visualizações, exemplo: mapeamento da variável setores para o eixo horizontal e para os valores das barras em gráfico de barras.

- Questão: Q3. Qual setor teve mais aumento de quantidade de produção ao longo do ano? (veja a Figura 5.13)
- Visualização: Gráfico de linhas múltiplas
- Discrepância: P17 mapeou setor aos dois eixos do gráfico de linhas.

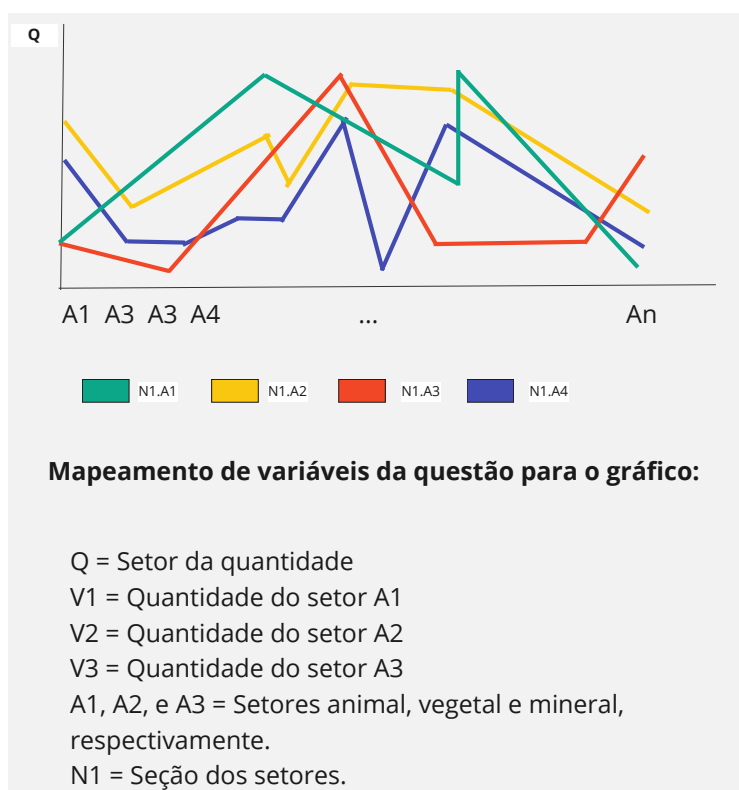


Figura 5.13: Exemplo 1 para mapeamento de uma variável para dois elementos da visualização

Utilização incorreta da visualização: esta discrepância ocorreu quando os participantes utilizavam os elementos da visualização de forma incorreta, mesmo que tivessem escolhido a visualização correta. Por exemplo, deixaram de utilizar variáveis quantitativas quando era necessário, ou utilizaram gráfico de linhas simples para representar várias variáveis (que deveriam ser representadas com várias linhas).

- Questão: Q5. Qual é a média de quantidade de produção de cada setor? (veja a Figura 5.14)
- Visualização: Gráfico de Barras
- Discrepância: P7 associou setores (eixo horizontal) com suprimentos (eixo vertical) e não utilizou a variável quantitativa.

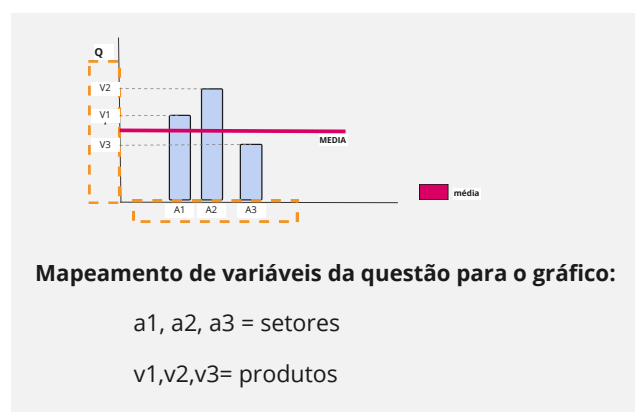


Figura 5.14: Exemplo 1 para utilização incorreta da visualização

Informações sem contexto: em alguns casos, os participantes mapearam para a visualização variáveis que não eram pedidas na questão. Alguns exemplos desta discrepância são apresentados a seguir.

- Questão: Q4. Qual a quantidade de suprimentos produzidos de cada tipo para cada setor de produção? (veja a Figura 5.15)
- Visualização: Gráfico de barras agrupadas
- Discrepância: P21 adicionou uma variável que não existia na questão - N2=empresa.

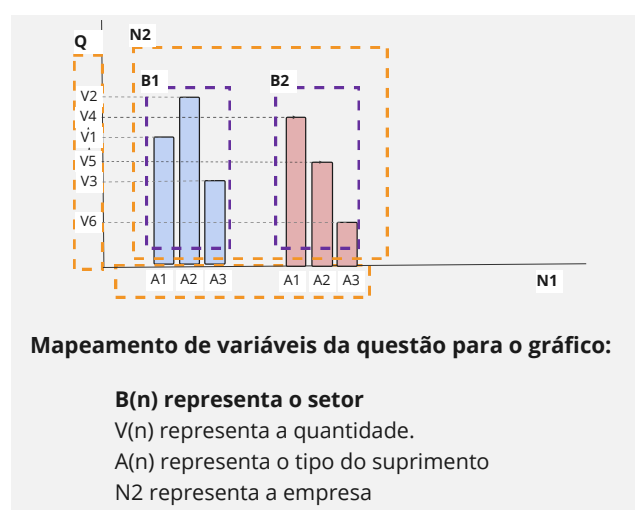


Figura 5.15: Exemplo 1 para informações sem contexto

Interpretação diferente da questão: Neste item, os participantes interpretaram mal o que a questão pedia. Desta forma, construíram uma visualização que não era adequada para a questão proposta.

- Questão: Q5. Qual é a média de quantidade de produção de cada setor? (veja a Figura 5.16)

- Visualização: Gráfico de Barras
- Discrepância: P5 fez um gráfico para representar a média de todos os setores e não de cada setor como pedido na questão.

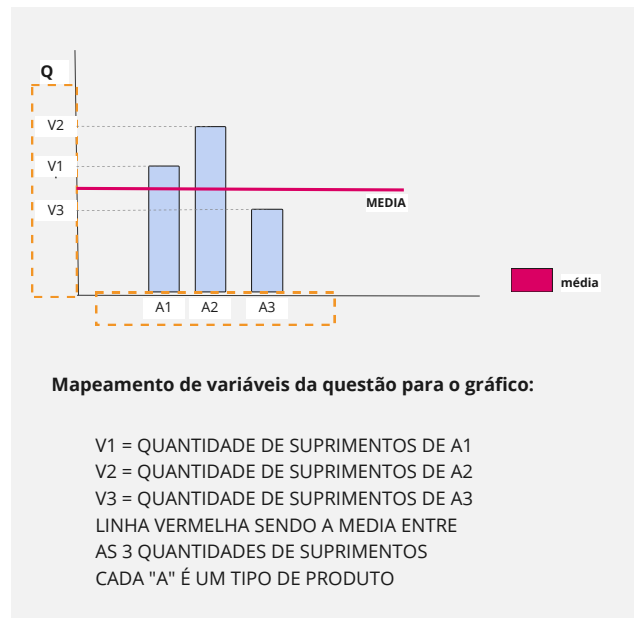


Figura 5.16: Exemplo 1 para interpretação diferente do que a questão queria dizer

Escolha de visualização inadequada: Alguns participantes escolheram visualizações que não eram as melhores opções para responder as questões propostas. Um exemplo dessa discrepância é apresentado a seguir.

- Questão: Q3. Qual setor teve mais aumento de quantidade de produção ao longo do ano? (veja a Figura 5.17)
- Visualização: Gráfico de barras
- Discrepância: P19 utilizou gráfico de barras para representar evolução ao longo do tempo, mas a opção mais adequada seria gráfico de linhas.

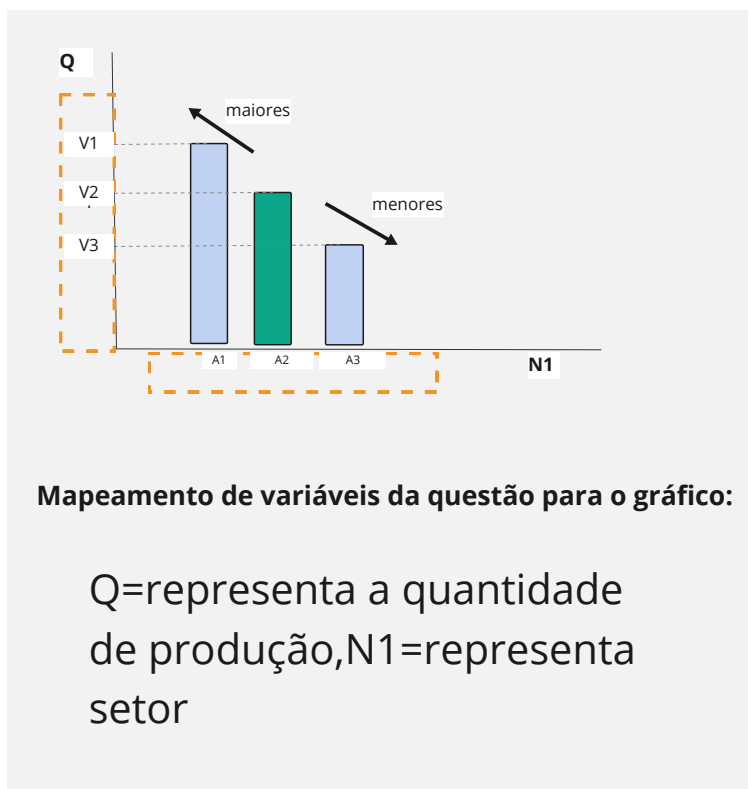


Figura 5.17: Exemplo 1 para exemplo de escolha inadequada de visualização

5.5

Discussão dos Resultados

Nesta seção apresentamos os resultados consolidados para responder as questões que guiaram o estudo.

5.5.1

Q1. De que forma o material VIS-SeMa apoiou a seleção de visualizações e o mapeamento de variáveis para visualização?

O objetivo desta questão foi identificar se o material proposto apoiou os participantes na criação de visualizações, focando nas atividades de seleção de visualizações e no mapeamento de variáveis na visualização.

Identificamos que, em relação à organização do material, os participantes não tiveram dificuldades. A organização do material foi considerada simples e os tópicos ajudaram a identificar o que estava sendo abordado nas questões.

Em relação aos padrões de visualização, conforme explicado anteriormente, a parte do material que detalha cada visualização era composta pelos seguintes padrões: visualização recomendada (o que), anatomia (quando/como usar) e exemplos. Em relação a estes itens, identificamos que a maioria dos participantes utilizou com maior frequência as representações visuais (anatomia do gráfico e seus exemplos). As explicações textuais foram pouco utilizadas.

Além disso, a apresentação de exemplos concretos ajudou no entendimento de como utilizar as visualizações, ou seja, como fazer o mapeamento das variáveis para as visualizações. Outro item que ajudou na atividade foram os exemplos de questões.

Em relação ao mapeamento dos tipos de variáveis, alguns participantes apresentaram dificuldades no entendimento. Apesar disso, as questões apresentadas ajudaram a entender qual tipo de visualização escolher. Em outras palavras, somente apresentar os tipos de variáveis, como feito em outras ferramentas, pode gerar ainda mais dificuldades. No entanto, combinar a anatomia das questões e das visualizações com os exemplos facilita a compreensão por novatos.

5.5.2

Q2. Quais problemas surgiram durante o processo de seleção e mapeamento de variáveis?

Esta questão tinha por objetivo apresentar dificuldades dos participantes ao utilizarem o material para criação das visualizações. Essas dificuldades são resumidas a seguir.

Houve algumas dificuldades diretamente relacionadas ao mapeamento das variáveis nas visualizações, ou seja, qual variável deve ser representada em que elemento da visualização. Neste sentido, identificamos que o material proposto necessita reforçar essa identificação de elementos e a associação destes elementos com as variáveis. As dificuldades observadas foram:

- Elementos da visualização que ficaram sem variáveis mapeadas;
- Falta de mapeamento de algumas variáveis importantes para responder às questões;
- Inversão no mapeamento de variáveis, ou seja, uma variável foi mapeada para o elemento que outra variável deveria ter sido mapeada e vice versa;
- Mapeamento de duas variáveis para o mesmo elemento;
- Mapeamento de uma variável para elementos distintos;
- Informações sem conteúdo, ou seja, variáveis que não existiam na questão foram inseridas na visualização.

Além das dificuldades relacionadas ao mapeamento, identificamos também que houve algumas interpretações erradas de questões e também escolhas inadequadas da visualização. Na maioria dos casos, os participantes escolheram visualizações que não eram adequadas para representar determinado tipo de variável.

5.6

Considerações sobre o Estudo de Viabilidade

Neste capítulo apresentamos um estudo de viabilidade para entender de que forma o material VIS-SeMa poderia ajudar e como poderia ser melhorado para apoiar novatos no design de visualizações. Como resultados, identificamos que uma forma de melhorar o entendimento sobre o uso das visualizações é utilizando a combinação de *templates* de questões e de gráficos (anatomia) e exemplos concretos das questões e dos gráficos. Adicionalmente, foi possível observar que conteúdos textuais não são uma boa opção para apoiar o mapeamento de variáveis.

Em relação às dificuldades, identificamos que o material proposto necessita reforçar a relação entre os elementos das visualizações e as variáveis a serem representadas. Além disso, identificamos alguns tipos de dificuldades no mapeamento de variáveis que, futuramente, podem ser utilizadas como diretrizes para o design ou inspeção de mapeamentos em visualizações. Apesar de algumas dificuldades, os resultados mostraram que a maioria dos participantes selecionou visualizações e criou mapeamentos adequadamente.

6

Considerações Finais e Trabalhos Futuros

Esta pesquisa teve como foco investigar o processo de design de visualizações realizado por novatos. Conforme explicado anteriormente, consideramos novatos os usuários que utilizam visualizações para tarefas simples de sua rotina mas não possuem especialidades em visualizações, análise de dados e estatística. Além disso, restringimos este grupo a profissionais da área de computação. Iniciamos a pesquisa investigando aspectos gerais do processo de criação de visualizações, onde pudemos identificar dificuldades e analisar decisões tomadas durante o processo. Após a análise geral, focamos em duas atividades principais do processo, nas quais encontramos maiores dificuldades por parte dos participantes. As atividades foram: seleção de visualizações e mapeamento de variáveis. Neste contexto, elaboramos a proposta de um material, o (VIS-SeMa), para apoiar essas atividades. O material foi baseado nos resultados de estudos anteriores e de características identificadas na literatura. A avaliação do material permitiu identificar problemas na criação de visualizações e algumas oportunidades de melhoria para o próprio material. Além disso, mostrou que, com seu uso, os participantes do estudo acertaram na escolha e no mapeamento as variáveis de grande parte das visualizações criadas. A seguir são apresentadas as contribuições da pesquisa e trabalhos futuros.

6.1

RQ1. Quais decisões são tomadas durante a construção de visualização de informações?

O processo de tomada de decisões é importante para entender um pouco sobre o que novatos pensam ao construir visualizações. A partir disto, podemos analisar quais fatores influenciam estas decisões e se são fatores adequados para a guiar a construção da visualização. Caso não sejam adequados, os instrutores ampliam seu conhecimento sobre quais erros novatos cometem na tomada de decisões e podem ajustar o seu material de acordo. As decisões identificadas são resumidas a seguir.

No contexto desta pesquisa, as decisões foram identificadas no estudo de observação descrito no Capítulo 3. A partir dos resultados, apresentamos cinco fatores que contribuíram para as decisões:

- Necessidade de permitir detalhamento das informações: Foco em identificar se as informações a serem representadas na visualização precisam de detalhes adicionais para melhor entendimento;
- Necessidade de facilitar comparações: Foco em identificar se a questão pede que informações sejam comparadas;
- Quantidade de informações a serem apresentadas: Foco em analisar se há muitas variáveis a serem representadas nos elementos das visualizações;
- Análise de variáveis: Fazer análise de quais variáveis devem ser representadas em quais elementos da visualização;
- Familiaridade com a visualização: Escolha baseada no conhecimento do usuário sobre a visualização.

6.2

RQ2. Quais dificuldades/problemas são encontrados durante o processo de design de visualização de informações?

Em nossas pesquisas anteriores (Ferreira et al., 2023, 2021), identificamos a dificuldade de profissionais da indústria, novatos em visualização, em explorar o uso de visualizações durante *workshops* de prototipação. A partir disto, optamos por focar este estudo em apoiar novatos em visualização. Durante esta pesquisa, executamos outros estudos para entender as dificuldades de novatos no design de visualizações e fornecer subsídios para apoiá-los.

Como resultados do estudo de observação (Capítulo 3), identificamos duas principais dificuldades: escolha de visualizações adequadas e mapeamento de variáveis em visualizações. Nesse contexto, propusemos e avaliamos, através de um estudo de viabilidade, um material para apoiar essas atividades. Observamos que o material proposto necessita reforçar a relação entre os elementos das visualizações e as variáveis a serem representadas. Além disso, identificamos alguns tipos de dificuldades no mapeamento de variáveis que, futuramente, podem ser utilizadas como diretrizes para o design ou inspeção de mapeamentos em visualizações. Apesar das dificuldades encontradas, a abordagem utilizada no material de associar *templates* parametrizados de questões com os *templates* de visualizações e com exemplos ajudou nas atividades de mapeamento e seleção, pois a maioria das visualizações e mapeamentos criados com utilização do material foram elaboradas sem discrepâncias, sendo 85 de 140 (61%) visualizações escolhidas corretamente e 82 de 139 (59%) mapeamentos elaborados sem erros.

6.3

RQ3. Como apoiar novatos na melhoria da criação de visualização de informações?

Para apoiar novatos no design de visualizações, propusemos o material VIS-SeMa (*Select and Map*). A partir da avaliação, observamos que as explicações textuais do material foram pouco utilizadas. No entanto, a apresentação de exemplos concretos ajudou no entendimento de como utilizar as visualizações, ou seja, como fazer o mapeamento das variáveis para as visualizações.

Outro item que ajudou na atividade foram os exemplos de questões. As questões apresentadas ajudaram a entender qual tipo de visualização escolher. Conforme observado nos resultados, combinar a anatomia das questões e das visualizações com os exemplos facilita a compreensão por novatos. Como os participantes mostraram preferência por conteúdos visuais e exemplos, as instruções de uso do material foram reestruturadas conforme mostrado na Figura 6.1. A atualização da página de instruções passou a apresentar um conteúdo mais reduzido e com as etapas de uso divididas em módulos para facilitar o uso das instruções.

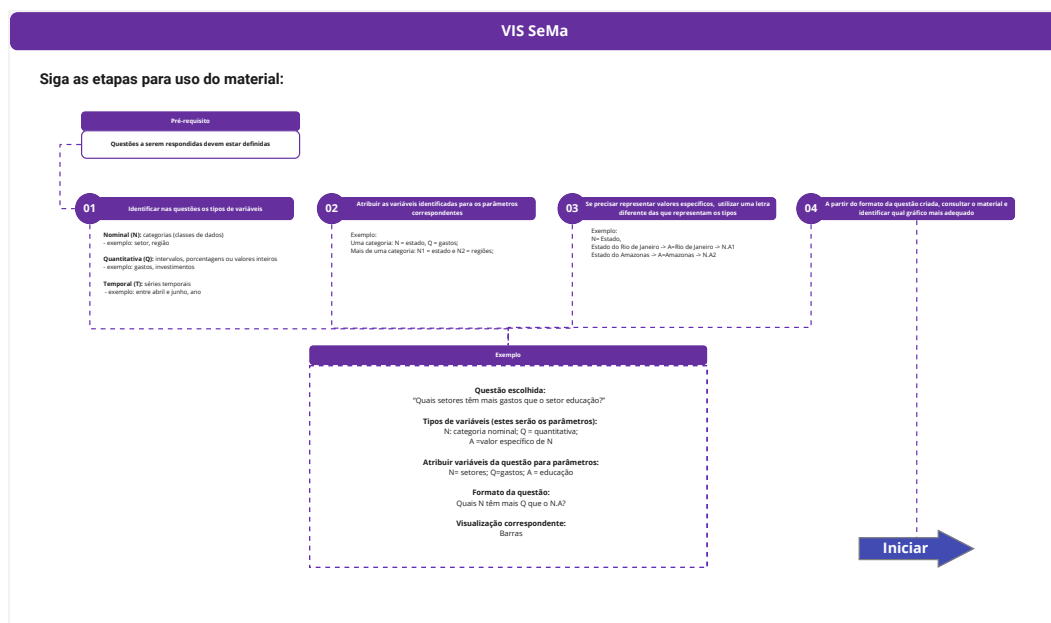


Figura 6.1: VIS-SeMa - Instruções de uso do material - versão atualizada após estudo

Além do material proposto, a partir das dificuldades identificadas podem ser elaboradas diretrizes que podem servir como forma de guiar novatos no mapeamento de variáveis para visualizações. Outra contribuição para o apoiar um instrutor na identificação de dificuldades de um novato é utilizar o design definido no estudo para identificar fatores que influenciam nas decisões dos

novatos. Dessa forma, podem entender se o novato está sendo influenciado por fatores corretos e direcioná-lo melhor.

6.3.1

Design de Estudos para Avaliar a Criação de Visualizações

Os estudos realizados nesta pesquisa definiram como tarefa criar visualizações para responder questões específicas. A partir desses estudos, conseguimos identificar bases para decisões e problemas na criação de visualizações. Os designs utilizados nestes estudos podem ser replicados posteriormente para outros estudos sobre a criação de visualizações e para avaliar a extensão do material para outros tipos de visualização. As etapas utilizadas nos estudos com visualizações (Capítulos 3 e 5) são apresentadas a seguir.

Estudo de Observação - Capítulo 3:

1. Criar um cenário com questões específicas a serem respondidas e uma tabela de dados com exemplos;
2. Disponibilizar alguns modelos de visualizações que os participantes do estudo possam editar para facilitar a atividade;
3. Pedir que os participantes escolham visualizações que julguem mais adequadas para responder as questões;
4. Pedir para participantes explicarem suas visualizações e o que quiseram representar. Nesta etapa, conseguimos identificar algumas de suas decisões.

Estudo de Viabilidade - Capítulo 5:

1. Criar um cenário com questões específicas a serem respondidas e uma tabela de dados com exemplos;
2. Disponibilizar alguns modelos de visualizações que os participantes do estudo possam editar para facilitar a atividade;
3. Pedir que os participantes escolham visualizações que julguem mais adequadas para responder as questões;
4. Pedir para os participantes representarem através do mapeamento explicado no material VIS-SeMa quais variáveis extraíram das questões e o que representam nas visualizações. Esta etapa ajuda a identificar as dificuldades em associar os dados com a visualização.

A partir das etapas e resultados desses dois estudos, elaboramos um processo consolidado para avaliação de estudos com visualizações e extensões do material utilizando o guia. O processo consolidado é apresentado na Figura 6.2. No caso de avaliar uma extensão do material VIS-SeMa, as avaliações editáveis citadas na Figura 6.2 devem ser as que foram adicionadas no material.

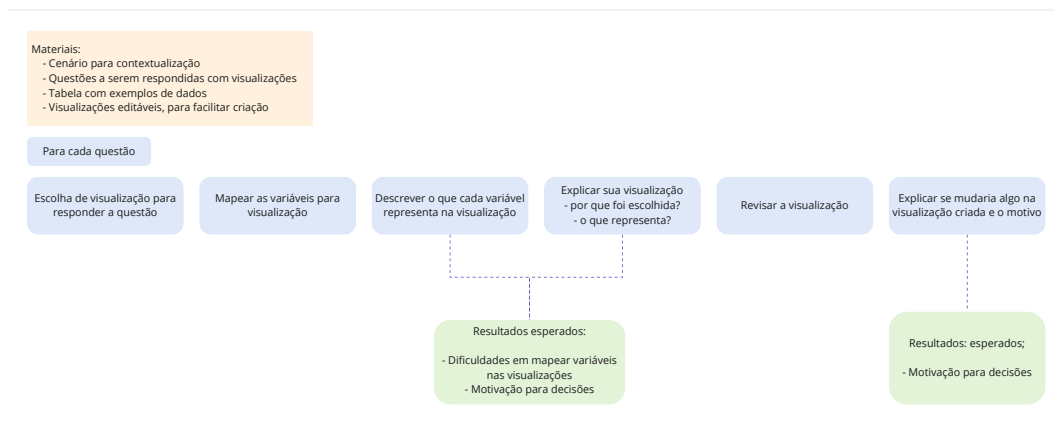


Figura 6.2: VIS-SeMa - Design de Experimento para avaliar o uso de visualizações

6.3.2

Diretrizes para a Inspeção de Mapeamentos de Variáveis

Além do guia e do design de estudo para avaliar o material, outra contribuição deste trabalho é a identificação de algumas diretrizes que podem ajudar a inspecionar visualizações criadas a fim de identificar possíveis discrepâncias. As diretrizes identificadas foram escritas em formato de perguntas-guia e são apresentadas a seguir.

- M1. Algum elemento da visualização ficou sem mapeamento de variáveis?
Exemplo: em alguns casos foram utilizados gráficos de linhas, mas não foi definida qual variável cada linha representava na visualização;
- M2. Falta mapear alguma variável importante para responder as questões?
Exemplo: em alguns casos era pedida a evolução de uma variável ao longo do tempo, mas essa variação não foi apresentada na visualização;
- M3. Há variáveis que poderiam inverter suas posições na visualização para melhorar a resposta à questão?
Exemplo: Na questão: “2- Qual tipo de produto tem maior percentual de produção para cada um dos setores?” - com a utilização do gráfico de barras empilhadas, uma opção melhor seria comparar o percentual dos

produtos dentro de cada setor separadamente, ou seja, as barras seriam os setores e as divisões de cada barra seriam os produtos. No entanto, alguns participantes inverteram, mapearam os produtos para a barra e as divisões para os setores;

- M4. Variáveis diferentes foram sendo mapeadas para o mesmo elemento?
Exemplo: atribuição da mesma variável ao eixo horizontal;
- M5. Uma variável está sendo mapeada de forma equivocada para mais de um elemento na visualização?
Exemplo: atribuição de uma variável ao eixo horizontal do gráfico e ao tamanho da barra, em um gráfico de barras;
- M6. Há variáveis “inseridas” (que não foram citadas na questão) na visualização?
Exemplo: Em alguns casos foi citada a variável fazenda, que não fazia parte dos dados.

6.4

Trabalhos Futuros

Como trabalhos futuros, podem ser realizados estudos para identificar novas diretrizes para apoiar o mapeamento de variáveis em visualizações. Além disso, as diretrizes propostas nesta pesquisa podem ser avaliadas para verificar sua viabilidade.

Em relação ao material proposto, foi definida uma estrutura padronizada que demonstrou apoiar principalmente o mapeamento de variáveis para visualizações. Esta estrutura pode ser utilizada como base para o desenvolvimento de outras abordagens que apoiem a criação de questões, criação e inspeção de visualizações. Além disso, a estrutura proposta pode ser utilizada para expandir os conteúdos criados para outros tipos de visualizações. Como citado por Bach et al. (2023), há dificuldades de professores em encontrar, avaliar e adaptar materiais. Para apoiar a expansão e avaliação do material, apresentamos um guia na Seção 4.4. O guia pode ajudar instrutores a construir seu próprio material utilizando a estrutura do VIS-SeMa.

Outros trabalhos podem envolver propostas para apoiar a construção de questões a partir de um conjunto de dados. Em vez de criar os *templates* extraíndo os dados de questões, pode ser criado um material para ajudar a criar questões nos formatos dos *templates* utilizando os dados disponíveis. Esse material pode ser útil para apoiar a comunicação entre equipes técnicas, clientes e usuários finais através da discussão das questões que se quer responder.

Outra possível abordagem para evoluções do material VIS-SeMa envolve a realização de encadeamento dos *templates* de questões, gerando vários caminhos possíveis para selecionar uma visualização de acordo com o cenário da questão. Além disso, também pode ser desenvolvido um sistema que ajude ativamente na construção da visualização, guiando novatos através da indicação de discrepâncias e apresentação dicas de como melhorar a visualização criada. Pode-se ainda adicionar outros aspectos no material, como por exemplo, apoiar o uso de interações e marcações nas visualizações, indicando suas funções, anatomia e exemplos, conforme proposto no VIS-SeMa. Adicionalmente, na seção de anatomia, os elementos das visualizações podem ser mais explorados visando minimizar as discrepâncias que ocorrem no mapeamento de variáveis para visualização.

Em relação à identificação de decisões que novatos tomam durante o processo de criação de visualizações, o estudo do Capítulo 3 pode ser reproduzido em outros contextos, por exemplo, ajudando instrutores a identificar se seus alunos estão tomando decisões baseadas em itens errados e dar um melhor direcionamento a eles, podendo traçar perfis de dificuldades. Futuramente, podem ser realizados outros estudos visando a avaliar se o material contribui para o letramento em visualização, sendo avaliado por instrutores de disciplinas de visualizações.

Referências Bibliográficas

- Bach, B., Keck, M., Rajabiyazdi, F., Losev, T., Meirelles, I., Dykes, J., Laramée, R. S., AlKadi, M., Stoiber, C., Huron, S., et al. (2023). Challenges and opportunities in data visualization education: A call to action. *IEEE Transactions on visualization and computer graphics*.
- Baker, R. S., Corbett, A. T., e Koedinger, K. R. (2001). Toward a model of learning data representations. In *Proceedings of the 23rd annual conference of the Cognitive Science Society*, pages 45–50. Mahwah NJ.
- Bako, H. K., Liu, X., Ko, G., Song, H., Battle, L., e Liu, Z. (2024). Unveiling how examples shape visualization design outcomes. *IEEE Transactions on Visualization and Computer Graphics*.
- Boy, J., Rensink, R. A., Bertini, E., e Fekete, J.-D. (2014). A Principled Way of Assessing Visualization Literacy. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1963–1972. Conference Name: IEEE Transactions on Visualization and Computer Graphics.
- Burns, A., Lee, C., Chawla, R., Peck, E., e Mahyar, N. (2023). Who do we mean when we talk about visualization novices? In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–16.
- Card, S. K., Mackinlay, J., e Shneiderman, B. (1999). *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann. Google-Books-ID: wdH2gqWfQmgC.
- Çay, D., Nagel, T., e Yantac, A. E. (2020). Colvis: collaborative visualization design workshops for diverse user groups. In *2020 24th International Conference Information Visualisation (IV)*, pages 550–558. IEEE.
- Crisan, A., Fiore-Gartland, B., e Tory, M. (2020). Passing the data baton: A retrospective analysis on data science work and workers. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1860–1870.
- Eberhard, K. (2023). The effects of visualization on judgment and decision-making: a systematic literature review. *Management Review Quarterly*, 73(1):167–214.

- Fernandez, C., Freitas, J. A., Blikstein, P., e de Deus Lopes, R. (2024). The design space of visualization tools for data science education: literature review and framework for future designs. *International Journal of Child-Computer Interaction*, page 100698.
- Ferreira, B., Kalinowski, M., Gomes, M., Marques, M., Lopes, H., e Barbosa, S. (2021). Investigating problem definition and end-user involvement in agile projects that use lean inceptions. In *Proceedings of the 20th Brazilian Symposium on Software Quality*.
- Ferreira, B., Marques, S., Kalinowski, M., Lopes, H., e Barbosa, S. D. (2023). Lessons learned to improve the ux practices in agile projects involving data science and process automation. *Information and Software Technology*, 155:107106.
- Firat, E. E., Joshi, A., e Laramee, R. S. (2022a). Interactive visualization literacy: The state-of-the-art. *Information Visualization*, 21(3):285–310.
- Firat, E. E., Joshi, A., e Laramee, R. S. (2022b). Vislite: visualization literacy and evaluation. *IEEE Computer Graphics and Applications*, 42(3):99–107.
- Ge, L. W., Cui, Y., e Kay, M. (2023). Calvi: Critical thinking assessment for literacy in visualizations. In *Proceedings of the 2023 CHI conference on human factors in computing systems*, pages 1–18.
- Ge, L. W., Hedayati, M., Cui, Y., Ding, Y., Bonilla, K., Joshi, A., Ottley, A., Bach, B., Kwon, B. C., Rapp, D. N., et al. (2024). Toward a more comprehensive understanding of visualization literacy. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–7.
- Grammel, L., Tory, M., e Storey, M.-A. (2010). How information visualization novices construct visualizations. *IEEE transactions on visualization and computer graphics*, 16(6):943–952.
- Heer, J., Shneiderman, B., e Park, C. (2012). A taxonomy of tools that support the fluent and flexible use of visualizations. *ACM queue*, 10(2):1–26.
- Huron, S., Jansen, Y., e Carpendale, S. (2014). Constructing visual representations: Investigating the use of tangible tokens. *IEEE transactions on visualization and computer graphics*, 20(12):2102–2111.
- Keck, M. e Groh, R. (2019). A construction kit for visual exploration interfaces. In *EuroVis (Short Papers)*, pages 79–83.

- Keck, M., Groh, R., e Vosough, Z. (2020). A didactic methodology for crafting information visualizations. In *2020 IEEE Visualization Conference (VIS)*, pages 186–190. IEEE.
- Lavanya, A., Sindhuja, S., Gaurav, L., e Ali, W. (2023). A comprehensive review of data visualization tools: features, strengths, and weaknesses. *Int. J. Comput. Eng. Res. Trends*, 10(01):10–20.
- Lee, S., Kim, S.-H., e Kwon, B. C. (2016). Vlat: Development of a visualization literacy assessment test. *IEEE transactions on visualization and computer graphics*, 23(1):551–560.
- Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *ACM Trans. Graph.*, 5(2):110–141.
- Maltese, A. V., Harsh, J. A., e Svetina, D. (2015). Data Visualization Literacy: Investigating Data Interpretation Along the Novice—Expert Continuum. *Journal of College Science Teaching*, 45(1):84–90. Publisher: National Science Teachers Association.
- Park, S., Bekemeier, B., Flaxman, A., e Schultz, M. (2022). Impact of data visualization on decision-making and its implications for public health practice: a systematic literature review. *Informatics for Health and Social Care*, 47(2):175–193.
- Parsons, P. (2021). Understanding data visualization design practice. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):665–675.
- Qin, X., Luo, Y., Tang, N., e Li, G. (2020). Making data visualization more efficient and effective: a survey. *The VLDB Journal*, 29(1):93–117.
- Rezaie, M., Tory, M., e Carpendale, S. (2024). Struggles and strategies in understanding information visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 30(6):3035 – 3048.
- Rodrigues, A. M. B., Barbosa, G. D. J., Lopes, H. C. V., e Barbosa, S. D. J. (2021). What questions reveal about novices' attempts to make sense of data visualizations: Patterns and misconceptions. *Computers & Graphics*, 94:32–42.
- Shull, F., Carver, J., e Travassos, G. H. (2001). An empirical methodology for introducing software processes. *ACM SIGSOFT Software Engineering Notes*, 26(5):288–296.

- Stoiber, C., Wagner, M., Grassinger, F., Pohl, M., Stitz, H., Streit, M., Potzmann, B., e Aigner, W. (2022). Visualization onboarding grounded in educational theories. *Visualization Psychology*, page 139.
- Stoiber, C., Wagner, M., Grassinger, F., Pohl, M., Stitz, H., Streit, M., Potzmann, B., e Aigner, W. (2023). Visualization onboarding grounded in educational theories. In *Visualization Psychology*, pages 139–164. Springer.
- Unwin, A. (2020). Why is data visualization important? what is important in data visualization? *Harvard Data Science Review*, 2(1):1.
- Vázquez-Ingelmo, A., García-Holgado, A., Verdugo-Castro, S., Theron, R., e García-Penalvo, F. J. (2024). Data visualization and domain knowledge: Insights through focus groups of researchers in spain. *Computers in Human Behavior*, 155:108162.
- Walny, J., Frisson, C., West, M., Kosminsky, D., Knudsen, S., Carpendale, S., e Willett, W. (2019). Data changes everything: Challenges and opportunities in data visualization design handoff. *IEEE transactions on visualization and computer graphics*, 26(1):12–22.
- Wang, Z., Sundin, L., Murray-Rust, D., e Bach, B. (2020). [16]cheat sheets for data visualization techniques. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13.

A
Material do Estudo 1

A.1
Tabela com visualizações escolhidas pelos participantes estudo 2

Tabela A.1: Visualizações escolhidas pelos participantes para cada questão

#P	Q1	Q2	Q3	Q4	Q5	Q6
P1	Histo- grama	Histo- grama	Tree Map	Gráfico de barras	Gráfico de barras empi- lhadas 100%	Gráfico de barras
P2	Gráfico de barras	Gráfico de barras	Gráfico de barras	Gráfico de barras empi- lhadas	Gráfico de barras	Gráfico de barras
P3	Gráfico de pizza	Histo- grama	Gráfico de pizza	Gráfico de barras empi- lhadas 100%	Gráfico de pizza	Gráfico de pizza
P4	Gráfico de linhas	Gráfico de barras	Gráfico de barras	Gráfico de linhas	Gráfico de barras	Gráfico de barras
P5	Gráfico de linhas	Gráfico de linhas	Gráfico de barras	Gráfico de linhas	Gráfico de pizza	Gráfico de linhas

B

Material do Estudo 2

B.1

Questionário utilizado no estudo 2

Q1. Marque com que frequência você utilizou os elementos do material VIS-SeMa:

- 1- nunca
- 2- quase nunca
- 3- raramente
- 4- às vezes
- 5- pouco frequentemente
- 6- frequentemente
- 7- muito frequentemente

Elementos do material:

- Identificação dos tipos de variáveis na questão [1][2][3][4][5][6][7]
- Os gráficos apresentados [1][2][3][4][5][6][7]
- Conteúdo explicativo das visualizações [1][2][3][4][5][6][7]
- Instruções na página inicial [1][2][3][4][5][6][7]
- Os exemplos de questão [1][2][3][4][5][6][7]
- Os exemplos de gráficos [1][2][3][4][5][6][7]

Q2. Marque quanto cada elemento do material VIS-SeMa ajudou a construir as visualizações para cada questão:

- 1- dificultou muito
- 2- dificultou
- 3- dificultou pouco
- 4- nem ajudou nem dificultou
- 5- ajudou pouco
- 6- ajudou
- 7- ajudou muito

Elementos do material:

- Organização do material [1][2][3][4][5][6][7]
- Identificação dos tipos de variáveis na questão [1][2][3][4][5][6][7]
- Os gráficos apresentados [1][2][3][4][5][6][7]
- Conteúdo explicativo das visualizações [1][2][3][4][5][6][7]
- Instruções na página inicial [1][2][3][4][5][6][7]
- Os exemplos de questão [1][2][3][4][5][6][7]
- Os exemplos de gráficos [1][2][3][4][5][6][7]
- Termos utilizados no material [1][2][3][4][5][6][7]

Q3. Qual sua opinião (pontos positivos e negativos) sobre a forma de organização do material ?

Q4. Qual sua opinião (pontos positivos e negativos) sobre a identificação dos tipo de variáveis da questão?

Q5. Qual sua opinião (pontos positivos e negativos) sobre os termos utilizados no material?

Q6. Qual sua opinião (pontos positivos e negativos) sobre os gráficos apresentados?

Q7. Qual sua opinião (pontos positivos e negativos) sobre o conteúdo explicativo das visualizações ?

Q8. Qual sua opinião (pontos positivos e negativos) sobre as instruções na página inicial do material ?

Q9. Qual sua opinião (pontos positivos e negativos) sobre os exemplos de questões?

Q10. Qual sua opinião (pontos positivos e negativos) sobre os exemplos de gráficos apresentados?

Q11. Além das citadas anteriormente, mais alguma parte do material VIS-SeMa gerou dificuldades? Quais dificuldades?

Q12. Você sentiu falta de alguma informação no material? Qual?

B.2
Caracterização dos participantes do Estudo de Viabilidade

Os números na Figura B.1 representam o nível de experiência indicado pelos participantes em uma escala de 1 a 7, onde 1 indica nenhuma experiência/conhecimento e 7 indica muita experiência/conhecimento. Os participantes que não possuem dados, não responderam ao questionário de caracterização, aplicado antes do estudo.

ID	Experiência Indústria	Conhecimentos teóricos sobre a elaboração de gráficos	Experiência prática em ler gráficos	Experiência prática em construir gráficos em sala de aula	Experiência prática em construir gráficos em projetos de software em indústria	Conhecimentos básicos em estatística - tipos de variável (nominal, ordinal, quantitativo etc)
P1	nenhuma	2	3	1	1	1
P2	nenhuma	2	1	2	1	2
P3	Zero	2	3	1	1	2
P4	Não respondeu caracterização					
P5	nenhuma	1	2	1	1	1
P6	Nenhuma experiência	2	2	2	2	1
P7	nenhuma	1	1	1	1	1
P8	Sem experiência.	5	3	1	1	2
P9	Não respondeu caracterização					
P10	Nenhuma experiência	3	3	2	1	2
P11	Até o momento não atuei na área	2	2	2	1	3
P12	2 anos - desenvolvedor de uma empresa	3	3	3	3	2
P13	1 mês - estagiário desenvolvimento web	1	2	1	1	2
P14	Sem experiência.	1	1	2	1	2
P15	4 meses - Analista de Redes	2	2	1	1	3
P16	sem experiência	3	3	2	1	3
P17	Não tenho nenhuma experiência	1	1	1	1	2
P18	nenhuma	1	2	1	1	1
P19	1 ano - programador.	2	2	2	2	2
P20	Nao tenho experiência	2	2	1	1	2
P21	3 meses - Desenvolvedor Back-End	2	2	1	1	1
P22	Sem experiência	2	2	1	1	1
P23	Não respondeu caracterização					
P24	nenhuma	2	3	2	1	1
P25	nenhuma	2	3	2	1	2
P26	nenhuma	1	1	1	1	1
P27	Nenhuma	2	2	2	2	3
P28	Não respondeu caracterização					

Figura B.1: Caracterização dos participantes do estudo de viabilidade

C

Materiais de apoio ao uso da Proposta

C.1

Exemplo 1 - Utilizando VIS-SeMa

Persona 1: Ana Maria, estudante de Design, está cursando a disciplina de visualização de informação pela primeira vez. Não tem experiência com uso de visualizações.

Persona 2: Paulo, estudante de Ciência da Computação, cursou a disciplina de Informação de visualização no semestre passado. Sua experiência com Visualização foi através da disciplina.

Cenário 1: Utilizando VIS-SeMa em Disciplinas de visualização

Ana Maria está cursando a disciplina de Visualização de Informação pela primeira vez, está com dificuldades em escolher e utilizar visualizações para responder algumas questões que a professora pediu em um exercício. Ela comentou com Paulo sobre sua dificuldade, ele então indicou que ela utilizasse um material de apoio que ele havia utilizado no semestre passado quando cursou a disciplina: O VIS-SeMa. Paulo então passou o [link](#) do material para Ana.

Ao chegar em casa, Ana acessou o link do material e começou a ler as instruções de uso. Conforme necessário para uso do material, ela já tinha questões pré-definidas para responder. Primeiramente, selecionou uma de suas questões “Qual tipo de produto (vegetal, mineral, animal) produzido no país X tem maior percentual de produção para cada região (norte, sul, leste, oeste)?”, e seguiu as etapas descritas nas instruções:

- E1. Iniciou a análise da questão, “Qual tipo de produto (vegetal, mineral, animal) produzido no país VisCity tem maior percentual de produção para cada região (norte, sul, leste, oeste)?”. Ela realizou esta análise fazendo anotações textuais na ferramenta Miro. Realizou as seguintes etapas:
 - E1.1. Identificou as variáveis na questão: tipo de produto, [maior] percentual de produção, região;

- E1.2. Associou as variáveis com tipos de dados (Nominal, Quantitativo, Temporal): N1 (A1, A2, A3) - tipo de produto (vegetal, mineral, animal), Q - percentual de produção, N2 (B1, B2, B3, B4) - região (norte, sul, leste, oeste);
- E1.3. Definiu o template parametrizado para sua questão: Qual N1 (A1, A2, A3) produzido no país VisCity tem maior Q para cada N2 (B1, B2, B3, B4)?
- E2. Após criar seu template utilizando o quadro que criou no miro, Ana voltou para a ferramenta VIS-SeMa e acessou a página 2 do material. Nesta página, comparou o template da sua questão com os templates apresentados no material. Ao fazer a comparação identificou que o template semelhante ao seu era da categoria “Visualizar proporções (porcentagens)” e o template semelhante ao seu era “Que N1 tem o [menor|maior] percentual Q em N2?”; Ela clicou no template escolhido no material e foi direcionada para a página 3, onde indicava qual visualização ela deveria utilizar: gráfico de barras empilhadas.
- E3. Na página 3 do material, Ana encontrou conteúdo explicativo sobre como mapear suas variáveis para os elementos da visualização e um exemplo de uso para ajudá-la. Utilizando as instruções, Ana fez um rascunho da sua visualização no quadro do Miro, onde havia feito a análise da sua questão.

C.2

Exemplo 2 - Inserindo novas visualizações no VIS-SeMa

Persona 1: Júlia é professora de visualização. Encontrou o material VIS-SeMa durante suas pesquisas, mas o material não tem algumas das visualizações que ela gostaria de apresentar aos alunos. Ao estudar o material, ela percebeu que ele possui um guia para extensão, ou seja, para ajudar a adicionar novas visualizações ao VIS-SeMa. Ela então utilizou o guia para adicionar as visualizações que deseja trabalhar com seus alunos ao material. A seguir são apresentadas as etapas que ela realizou para inserir uma destas visualizações: o Gráfico de Barras Empilhadas.

Cenário 2: Utilizando VIS-SeMa em Disciplinas de visualização

Conforme indicado no guia para adicionar novas visualizações ao material, Júlia consultou o material de (Rodrigues et al., 2021) e encontrou o seguinte template para o gráfico de barras empilhadas: “Que N1 tem o [maior|menor] em N2.B2?” e “Que N1 tem o [maior|menor] por N2?”.

Após encontrar o template, ela criou uma categoria que descreve o resultado que esperava da visualização: “Visualizar proporções e porcentagens”. Então, ela adicionou seu template e a categoria na página 2 do material VIS-SeMa, conforme apresentado na Figura C.1.

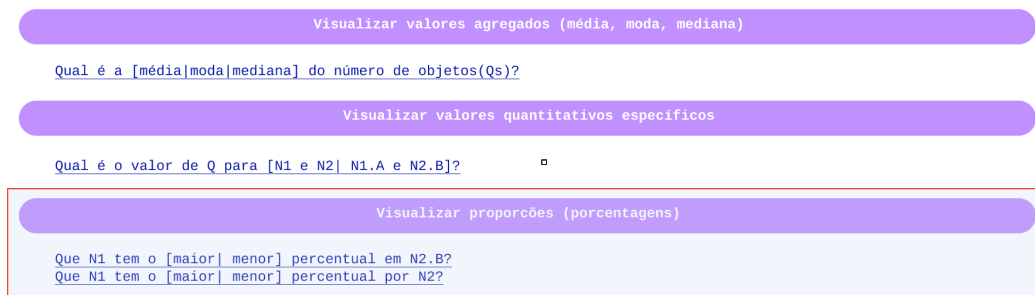


Figura C.1: Categoria e templates de questões inseridos no VIS-SeMa

De acordo com o guia, a próxima etapa realizada foi adicionar os templates de questões na página 3 do material, explicando qual o tipo de cada variável do template e indicando em quais elementos da visualização cada variável deveria ser mapeada. Além disso, inseriu uma tabela de dados preenchida com as variáveis para facilitar o entendimento de como as variáveis estavam representadas em uma tabela. Júlia criou todas as imagens e textos a serem inseridos no material na ferramenta Miro. O conteúdo criado é apresentado na Figura C.2.

Após o detalhamento dos templates, ainda na página 3 do material, ela inseriu um exemplo concreto de questão e visualização. Conforme Figura C.3.

Material de apoio à criação de Gráficos (Visualizações)

Visualização Recomendada

Gráfico de barras empilhadas

Função

Visualização representa proporções. Quanto de cada subcategoria forma uma categoria.

Anatomia

Que N1 tem o [maior| menor] percentual por N2?
Que N1 tem o [maior| menor] percentual em N2.B?
Variáveis representadas:
N1=categoria nominal 1, N2=categoria Nominal 2, Q_percentual = proporções de cada N1 dentro de N2

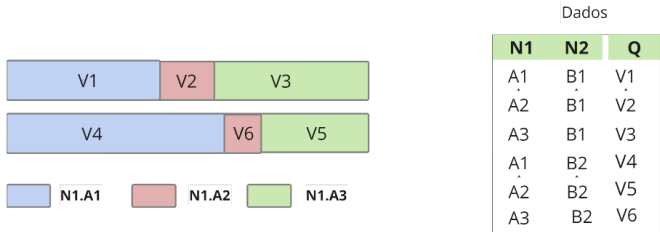


Figura C.2: Templates de questões e visualizações inseridos no VIS-SeMa

Exemplo(s)

Que setor tem [maior/menor] percentual de gastos por região?
Que setor tem [maior/menor] percentual na região Norte?
Variáveis:
N1=setor, N2=região, Q_percentual = percentual de gastos

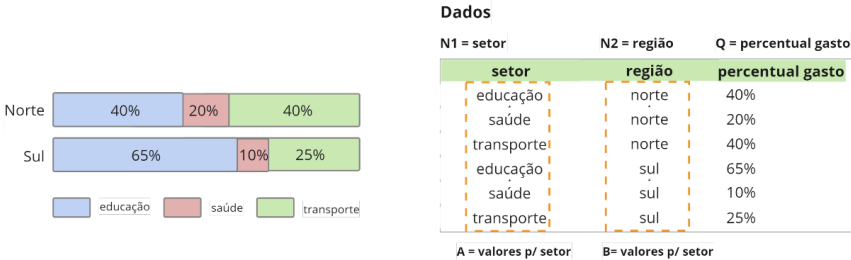


Figura C.3: Exemplos de questões e visualizações inseridos no VIS-SeMa