



Universidade Federal
do Rio de Janeiro

Escola Politécnica

ASPECTOS FORMAIS DE INTENÇÃO DE MÍDIA E CARACTERÍSTICAS DE DESIGN EM VISUALIZAÇÕES DE DADOS

Lucas Barcellos Oliveira

Projeto de Graduação apresentado ao Curso de Engenharia de Computação e Informação da Escola Politécnica da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários à obtenção do título de Engenheiro.

Orientador: Claudio Esperança

Coorientadora: Doris Kosminsky

RIO DE JANEIRO, RJ - BRASIL

DEZEMBRO DE 2019

ASPECTOS FORMAIS DE INTENÇÃO DE MÍDIA
E CARACTERÍSTICAS DE DESIGN
EM VISUALIZAÇÕES DE DADOS

Lucas Barcellos Oliveira

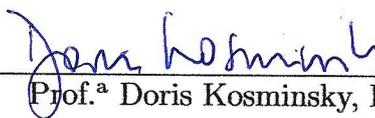
PROJETO DE GRADUAÇÃO SUBMETIDO AO CORPO DOCENTE DO CURSO DE ENGENHARIA DE COMPUTAÇÃO E INFORMAÇÃO DA ESCOLA POLITÉCNICA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE ENGENHEIRO DE COMPUTAÇÃO E INFORMAÇÃO



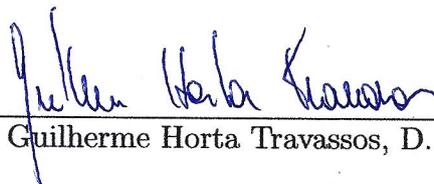
Lucas Barcellos Oliveira



Prof. Claudio Esperança, Ph. D.



Prof.ª Doris Kosminsky, Ph. D.



Prof. Guilherme Horta Travassos, D. Sc.

RIO DE JANEIRO, RJ - BRASIL

DEZEMBRO DE 2019

OLIVEIRA, LUCAS BARCELLOS

Aspectos formais de intenção de
mídia e características de design em
visualizações de dados / Lucas Barcellos
Oliveira. - Rio de Janeiro: UFRJ /
Escola Politécnica, 2019.

XIV, 64 p.: il.; 29,7 cm.

Orientador: Claudio Esperança.

Coorientadora: Doris Kosminsky.

Projeto de Graduação - UFRJ / Escola
Politécnica / Curso de Engenharia de
Computação e Informação, 2019.

Referências Bibliográficas: p.59-64.

1.Visualização de Dados 2.Design
da Informação 3.Computação Gráfica
4.Avaliação 5.Web

I.Universidade Federal do Rio de
Janeiro, Escola Politécnica, Curso de
Engenharia de Computação e Informação.

II.Título (série)

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Escola Politécnica - Departamento de Eletrônica e de Computação

Centro de Tecnologia, bloco H, sala H-217, Cidade Universitária

Rio de Janeiro - RJ CEP 21949-900

Este exemplar é de propriedade da Universidade Federal do Rio de Janeiro, que poderá incluí-lo em base de dados, armazenar em computador, microfilmear ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do autor.

DEDICATÓRIA

A todos, sem exceção.

AGRADECIMENTOS

Agradeço genuinamente a todos que, de alguma forma e por algum momento, compartilharam dessa jornada. Em especial, agradeço a meus pais, Carla Cristina e José Maurício, por todo o apoio dado desde o princípio, e aos membros do LabVis, Claudio Esperança, Doris Kosminsky e Luana da Silva, com os quais tive a alegria de trabalhar e sem os quais este trabalho não se concretizaria.

Resumo do Projeto de Graduação apresentado à Escola Politécnica / UFRJ como parte dos requisitos necessários para a obtenção do grau de Engenheiro de Computação e Informação.

Aspectos formais de intenção de mídia e características de design em visualizações de dados

Lucas Barcellos Oliveira

Dezembro/2019

Orientador: Claudio Esperança

Co-orientadora: Doris Kosminsky

Curso: Engenharia de Computação e Informação

Este trabalho apresenta uma discussão sobre elementos relevantes de mídias de visualização de dados e propõe uma lista de aspectos formais a serem considerados por engenheiros e designers quanto à forma de publicação e escolhas de design dessas mídias. É definido cada item proposto e são apresentados exemplos de visualizações de dados disponíveis online que ilustrem tais conceitos.

Em seguida, foi realizada uma atividade na qual a lista proposta foi validada junto a estudantes de visualização de dados. Posteriormente, os comentários coletados são analisados e melhorias a essa proposta são sugeridas e discutidas.

Palavras-Chave: visualização de dados, design da informação, computação gráfica, avaliação, web.

Abstract of Undergraduate Project presented to POLI/UFRJ as a partial fulfillment of the requirements for the degree of Engineer.

Formal aspects of media intention and design characteristics in data visualization

Lucas Barcellos Oliveira

December/2019

Advisor: Claudio Esperança

Co-advisor: Doris Kosminsky

Course: Computing and Information Engineering

This work presents a discussion about relevant elements to data visualization media and proposes a list of formal aspects to be considered by engineers and designers concerning the form of publication and design choices of these media. Each proposed item is defined and examples of online data visualizations illustrating such concepts are presented.

Next, an activity was performed in which the proposed list was validated with students of a data visualization course. Subsequently, all gathered comments are analyzed and improvements to this project are raised and discussed.

Keywords: data visualization, information design, computer graphics, evaluation, web.

SIGLAS

UFRJ – Universidade Federal do Rio de Janeiro

LabVis – Laboratório da Visualidade e Visualização

INPE – Instituto Nacional de Pesquisas Espaciais

DBLP – *Digital Bibliography & Library Project*

XML – *Extensible Markup Language*

KNIME – *Konstanz Information Miner*

Sumário

1	Introdução	1
1.1	Tema	1
1.2	Delimitação	2
1.3	Justificativa	2
1.4	Objetivos	4
1.5	Metodologia	4
1.6	Organização do Trabalho	5
2	Panorama Atual	7
2.1	Referencial Teórico	7
2.2	Trabalhos Relacionados	8
3	Lista de Aspectos Formais de Intenção e Design	13
3.1	Introdução à lista	13
3.2	Intenção da mídia de visualização	14
3.2.1	Familiaridade do público	14
3.2.2	Grau de explorabilidade	17
3.2.3	Frequência de visitaçào	18
3.3	Características de design	20
3.3.1	Forma de manipulação dos dados	20

3.3.2	<i>Feedforward</i>	21
3.3.3	<i>Dashboards</i>	22
3.3.4	Número de layouts	23
3.3.5	Continuidade temporal	25
3.3.6	Animação	25
3.3.7	Elementos explicativos adicionais	27
3.3.8	Integração com narrativa	30
3.3.9	<i>Tooltips</i> e/ou <i>popups</i>	30
4	Exemplo de Aplicação	34
4.1	Descrição da visualização	34
4.2	Decisões de implementação	35
4.3	Quanto à intenção da mídia	38
4.3.1	Quanto à familiaridade do público	38
4.3.2	Quanto ao grau de explorabilidade	38
4.3.3	Quanto à frequência de visitação	38
4.4	Quanto aos aspectos de design	39
4.4.1	Quanto à forma de manipulação dos dados	39
4.4.2	Quanto à existência de <i>feedforward</i>	39
4.4.3	Quanto à presença de <i>dashboards</i>	39
4.4.4	Quanto à existência de mais de um layout	39
4.4.5	Quanto à continuidade temporal	40
4.4.6	Quanto às animações	40
4.4.7	Quanto aos elementos explicativos adicionais	41
4.4.8	Quanto à existência de <i>tooltips</i> e <i>popups</i>	41
4.5	Discussão	42

5	Avaliação de Viabilidade	44
5.1	Descrição do estudo	44
5.2	Preparação	45
5.3	Primeira Atividade	45
5.4	Segunda Atividade	46
5.5	Discussão dos Resultados	51
5.5.1	Aferindo Concordância: Distância média	53
5.5.2	Aferindo Relações: Matriz de Correlação	55
6	Conclusões	57
	Bibliografia	59

Lista de Figuras

3.1	Captura de tela de <i>How long are characters in each episode?</i>	16
3.2	Captura de tela de <i>Narrative Chart</i>	16
3.3	Exemplo de interação por manipulação direta em <i>Spyglass on the Past: New York City 1836 and Today</i>	21
3.4	Exemplo de <i>feedforward</i> em <i>U.S. Gun Deaths in 2013</i>	22
3.5	Captura de tela de <i>Census Data Visualization</i>	23
3.6	Layouts apresentados por <i>GOV—DNA</i>	24
3.7	Layouts apresentados por <i>U.S. Gun Deaths in 2013</i>	24
3.8	Exemplo de uso de animação inicial em <i>U.S. Gun Deaths in 2013</i>	27
3.9	Exemplo de elemento explicativo adicional em <i>Out of Sight, Out of Mind</i>	28
3.10	Exemplo de <i>tooltip</i> com detalhes do dado em <i>Selfiecity</i>	32
3.11	Exemplo de <i>tooltip</i> com dica de interação em <i>Project Ukko</i>	33
4.1	Exposição de <i>Visualizing Visualizers</i>	36
4.2	Captura de tela de <i>Visualizing Visualizers</i>	37
4.3	Teclado utilizado em <i>Visualizing Visualizers</i>	40
5.1	Captura de tela da Seção 2 formulário aplicado.	49
5.2	Página de referência no website do LabVis.	50
5.3	Fluxo de trabalho de processamento e análise dos dados no KNIME.	53

Lista de Tabelas

5.1	Questões e tipos de resposta do formulário aplicado: Seção 1	47
5.2	Questões e tipos de resposta do formulário aplicado: Seção 2	47
5.3	Questões e tipos de resposta do formulário aplicado: Seção 3	48
5.4	Questões e tipos de resposta do formulário aplicado: Seção 4	49

Capítulo 1

Introdução

1.1 Tema

Esta produção trata de um aspecto que recentemente passou a ser mais investigado no campo de design da informação: como combinar visualizações de dados, que despertam crescente interesse e uso em diversos setores, com mídias mais convencionais, como texto? De reportagens jornalísticas a ferramentas de monitoramento, páginas web representam uma crescente demanda por compor diferentes mídias em um só produto de modo a oferecer maior facilidade de uso e de leitura.

Por ser um tema de interesse ainda muito incipiente, esse ainda é um desafio em aberto e não trivial. É necessário, então, formalizar um conjunto de aspectos de projeto e de design capazes de caracterizar mídias de visualização de dados, atentando-se para como tais fatores se relacionam entre si e para o contexto no qual visualizações são inseridas.

Com isso em mente, ensejam-se, também, maneiras de avaliar como esse tipo de mídia se encaixa nas expectativas de seus criadores e de seus usuários, compreendendo como fatores tais quais público-alvo e natureza dos dados impactam nas características da mídia de visualização de dados.

1.2 Delimitação

Para este trabalho, será considerado o estudo de páginas web acessíveis através de navegadores em computadores *desktop*. Visto que dispositivos móveis, como *smartphones*, e softwares específicos, como aplicativos, podem contar com peculiaridades muito distintas das observadas no cenário aqui considerado, não seria possível tentar cobrir tais casos neste trabalho devido à grande extensão dos tópicos.

Ademais, atuais limitações percebidas em dispositivos móveis, como a dificuldade em utilizar interações de arraste em telas de tamanho muito limitado, e trunfos neles encontrados, como a possibilidade de utilizar diferentes orientações da tela para diferentes *layouts* de visualização com praticidade, estão sujeitos a variações entre diferentes aparelhos e versões, assim limitando o potencial de generalização e replicação dos resultados encontrados e discutidos neste trabalho quando comparado a restringir o objeto de estudo a mídias de visualização consumidas em computadores de mesa.

Vale notar que fogem do escopo desta produção análises acerca dos tipos de visualização, tecnologias e bibliotecas utilizadas para desenvolvê-las. No presente trabalho, serão considerados fatores relacionados ao projeto de mídias de visualização, como o contexto no qual dada visualização se insere e a intenção com que determinada página web é concebida. Isso considera aspectos como público-alvo, tempo esperado para o consumo da informação, velocidade de mudança dos dados, estrutura e organização da informação, entre outros, no lugar de consumo de memória, tempo de execução, tempo de desenvolvimento e outros aspectos mais atrelados à codificação.

1.3 Justificativa

Dados são produzidos constantemente em volume e variedade recordes, introduzindo na prática unidades de ordens de grandeza até então nunca utilizadas, como o *exabyte*. Diversas fontes, sejam equipamentos ou indivíduos, em diferentes contextos e formatos, coletam dados sobre o mundo ao nosso redor e geram dados sobre suas atividades. [1]

Tal abundância permite cada vez mais cruzamentos e descobertas com base na análise e observação de dados. Por outro lado, o mesmo fenômeno que fortalece o estudo de tais dados também suscita um questionamento: como representar visualmente essa avalanche de dados e informação, em especial com o advento de Big Data, de modo a embasar novas análises e descobertas?

A visualização de dados trata de representações visuais com o objetivo de ampliar a cognição e favorecer a aquisição de conhecimentos que possam colaborar com a tomada de decisões. [2] Esse campo foca-se na pesquisa das melhores práticas em exibir dados de maneira, precisa, acessível e objetiva, permitindo que um observador seja capaz de assimilar informações a partir deles.

Desde um infográfico presente em uma reportagem jornalística até um painel interativo de exploração em uma empresa, a Visualização de Dados combina conhecimentos sobre comunicação visual, interação humano-computador, cognição e outras áreas correlatas a fim de alimentar essa demanda crescente e essencial para o desenvolvimento dos mais variados campos. É um domínio interdisciplinar em que, enquanto alguns conhecimentos, como os usos de cores, já se encontram sedimentadas, ainda há um vasto espaço para novas contribuições.

Por se tornarem cada vez mais presentes e representarem uma disciplina-chave para o futuro, muitas perguntas em aberto fomentam temas para pesquisa relacionadas a visualizações de dados. A necessidade de produzir visualizações mais intuitivas e claras, que permitam um entendimento rápido e acurado de leitores dos mais variados setores da sociedade e de sociedades distintas, ensejam e justificam a necessidade de estudos sob esta ótica.

Também consideremos que, a cada ano, o número de internautas e, consequentemente, o tráfego e consumo de dados na web crescem sem qualquer sinal aparente de desaceleração [3]. Tendo em vista essa tendência, especial atenção deve ser dada a mídias de visualização de dados que agregam tais visualizações a outras mídias online.

1.4 Objetivos

Este trabalho tem como objetivos principais o levantamento e discussão dos principais aspectos na criação conceitual e design de páginas que incluem visualizações de dados em seu corpo e a proposição de um conjunto resumido de elementos a serem observados nesse tipo de projeto a fim de avaliar exemplos já existentes bem como de projetar e validar novas páginas. Isto é, almeja-se levantar uma lista de itens que caracterizem bem as mais diversas mídias de visualização de dados para web existentes.

É importante salientar que o atingimento de tais objetivos deve considerar a vasta e diversa miríade de profissionais envolvidas na área de visualização de informação. Ou seja, busca-se refletir não apenas a perspectiva de implementação e codificação, mas também elaborar quesitos sob a ótica de comunicação e design de informação. [4]

Em suma, busca-se obter uma lista de aspectos formais relevantes como primeiro passo para engenheiros e designers ao projetar novas mídias de visualização para clientes. Adicionalmente, almeja-se verificar como tais características se correlacionam, descobrindo ligações frequentes encontradas em mídias de visualização existentes.

Assim, é possível, por exemplo, conhecer melhor a relação, se houver, entre tópicos relativos à mídia de visualização, como público-alvo e grau de explorabilidade e escolhas de design, como o uso de animações e elementos explicativos adicionais.

1.5 Metodologia

Inicialmente, foi realizado um levantamento do cenário atual de pesquisa sobre o tema de mídias de visualização de dados. Em um primeiro momento, buscou-se a familiarização com os fundamentos e conceitos já estabelecidos sobre o tema. Consecutivamente, almejou-se compreender metodologia e resultados de publicações recentes bem como os caminhos apontados como desenvolvimentos futuros de suas produções e que poderiam se alinhar aos objetivos deste trabalho.

Em um segundo momento, foram levantados e discutidos possíveis aspectos relevantes ao projeto e desenvolvimento de mídias de visualização de dados junto ao grupo de pesquisa do Laboratório de Visualidade e Visualização (LabVis) da Escola de Belas Artes da UFRJ. Foram também selecionadas visualizações de dados que permitiam ilustrar com clareza cada aspecto formal levantado, de forma a ser incluído em um resumo produzido em seguida, contendo a descrição de cada aspecto.

Posteriormente, foi realizada uma atividade para ratificar os atributos elencados na lista e obter impressões acerca de seu uso na avaliação de mídias de visualização de dados. Dessa maneira, também foi possível obter sugestões de outros possíveis aspectos a serem incluídas em novas versões da lista e de aspectos que poderiam ser omitidos em alguns casos e procurar relações frequentes entre eles.

1.6 Organização do Trabalho

No Capítulo 2, serão abordados os avanços atuais no tema estudado, visando prover um panorama geral sobre o que se sabe acerca de mídias de visualização e seus elementos.

O Capítulo 3, por sua vez, discute um conjunto de aspectos relevantes para a elaboração de mídias de visualização de dados, estruturado no formato de uma lista de checagem. São apresentadas descrições e oferecidos exemplos para os itens de tal lista, firmando os conceitos que serão usado de base para argumentações e desenvolvimentos dos capítulos posteriores.

O Capítulo 4 aplica os conceitos apresentados no Capítulo 3 em um caso real de mídia de visualização, exemplificando seu uso. Discutem-se, também, as conclusões obtidas pela análise utilizando tais aspectos, embasando melhorias no exemplo de mídia estudado.

Já o Capítulo 5 apresenta o experimento no qual buscou-se a validação da lista de aspectos formais apresentadas no Capítulo 3. São descritos o formato desse teste, o público junto ao qual ele foi realizado e seu conteúdo, além da análise dos resultados obtidos.

Por fim, no Capítulo 6, são tomadas conclusões acerca do trabalho, recapitulando o progresso feito e os resultados obtidos. São sugeridos, ainda, trabalhos futuros como próximos passos dessa pesquisa.

Capítulo 2

Panorama Atual

2.1 Referencial Teórico

Consideremos o trabalho de Boy et al. [5], em que são estudadas diferentes formas das chamadas “interatividades sugeridas” (do original, em inglês, “*suggested interactivities*”). Nesse estudo, interatividades sugeridas representam as formas de interação planejadas pelos criadores de uma visualização de dados e, conseqüentemente, as maneiras pelas quais espera-se que seus usuários atuem sobre ela.

De modo similar ao adotado neste trabalho, o autor determina uma lista de atributos que caracterizem tais interatividades sugeridas e classifica um conjunto de exemplos segundo tais critérios. Em tal artigo, no entanto, são considerados apenas elementos e características apresentados como parte da visualização de dados, com maior enfoque no estudo de interações. Não é também ponderada a mídia como um todo na qual a visualização encontra-se inserida.

Tal publicação dialoga com a temática de *affordances*. A discussão acerca da definição de *affordance* elaborada no trabalho de Boy é de grande ajuda no entendimento desse conceito, que será de suma importância em discussões no decorrer desta produção.

Affordances podem ser compreendidas por diferentes definições e aplicações. Numa de suas definições mais clássicas, segundo Gibson [6] em 1977, elas são propriedades presentes em certos objetos capazes de incitar interações.

Por sua vez, Norman [7], em 2002, introduz a noção de que *affordances* são atributos de objetos que permitem que pessoas entendam como os mesmos devem ser manipulados. Outras distinções posteriores foram propostas por ele, como entre *affordances* reais, propriedades físicas de objetos, e *affordances* percebidas, a compreensão obtida pelo usuário sobre qual interação pode ser realizada.

Os conceitos de *affordances* estão presentes em uma variedade de aplicações, de diferentes campos de estudo como, por exemplo, no design de produtos [8]. Na área de visualização de dados, utilizamos a definição de *affordances* como elementos gráficos que sugerem que determinada interação é possível.

Outro relevante conceito levantado no artigo é o de *feedforward*. Segundo ele, *feedforward* refere-se a propriedade de uma visualização de oferecer uma prévia do resultado de uma interação pelo usuário antes que ela efetivamente ocorra. Assim, por exemplo, quando uma interação de clique em um botão altera o estado de uma visualização, tal novo estado pode ser exibido previamente – tipicamente usando uma imagem miniatura (*thumbnail*) – quando o mouse é posicionado sobre o botão antes de pressioná-lo. *Feedforward*, conforme apresentado por Boy [5], tem grande eficácia na facilitação do uso de visualizações interativas.

2.2 Trabalhos Relacionados

Na comunidade acadêmica, especialmente na última década, algumas publicações abordaram o tema aqui discutido e visaram objetivos similares aos aqui propostos, porém sob outros enfoques. Fazemos um breve percurso por alguns desses trabalhos correlatos de modo a conhecer melhor o cenário existente e os esforços atuais de caracterização de mídias de visualização.

Em 2004, Plaisant [9] já levantava questionamentos sobre como adequar diferentes ferramentas a diferentes usuários, tarefas e problemas do mundo real. Seu estudo, direcionado a ferramentas para avaliação de visualizações de dados, argumenta sobre a necessidade de considerar cenários realistas, tendo em mente os contextos diversos em que visualizações são inseridas, não se limitando tão somente a experimentação em laboratório.

Apesar de se referir diretamente a testes de usabilidade, Plaisant salienta a compreensão acerca das necessidades de diferentes públicos, meios de distribuição e objetivos das visualizações e seu impacto no “sucesso” da tarefa de visualização como um desafio para a área.

Já abordagens como as de North [10] e Saraiya et al. [11] analisam a concepção de visualizações de dados a partir de experimentos e comparações. Em seu artigo, North conceitua e avalia visualizações com base em *insights*, denotando então a importância de estudar o que ocorre além dos limites da visualização de dados.

Keller et al. no livro *Visual Cues: Practical Data Visualization* [12] expõe a relevância de estudar a mídia em que visualizações de dados são inseridas. Os autores analisaram as diferenças nas escolhas de design, como cor, resolução e proporções, entre os meios popularmente utilizados na época, como papel, slides, vídeo. Também são levantadas observações acerca do propósito, nível de detalhamento, quantidades de gráficos e textos, tempo para produção, equipamento de projeção, equipe, conveniência, impressão, tamanho do público e custo, dentre outros aspectos.

Apesar da detalhada lista e bem desenvolvida iniciativa, a passagem do tempo e desenvolvimento de novas formas de publicizar visualizações, além de avanços nos processos de produção, ensejam uma atualização de tal proposta considerando fatores como a incorporação em páginas web pela Internet.

Wassink et al. [13], em capítulo de título *Applying a User-centered Approach to Interactive Visualisation Design* do livro *Trends in Interactive Visualization: State-of-the-Art Survey* [14], comentam a importância de considerar quem são os usuários esperados de uma visualização de dados. Sua proposta é, especificamente, colocar o usuário no foco do processo de design de visualizações interativas.

É mencionado, além das necessidades diferentes entre grupos diferentes de usuários, que habilidades de percepção e fatores muitas vezes considerados secundários, como idade, gênero e cultura, desempenham uma grande influência na forma como um sistema de visualização é utilizado. E, isso posto, tais questões devem ter, de algum modo, papel importante na concepção de visualizações de dados.

Finalmente, trabalhos como o de Segel e Heer [15], em 2010, elaboram conjuntos de atributos propostos para a classificação de exemplos de visualizações. Nesse caso, tais exemplos são avaliados segundo gênero de mídia no qual a visualização está inserida, táticas de narrativa visual e táticas de estrutura narrativa. As possibilidades de gênero listadas em tal sistema de classificação são vídeo, slide, charge, fluxograma, pôster, mapa ou artigo de revista. Já narrativa visual e estrutura narrativa são subdivididas em grupos mais específicos de atributos.

O critério narrativa visual consiste em: *visual structuring* ou “estruturação visual”, que avalia as formas utilizadas para comunicar ao leitor a estrutura da narrativa trazida pela visualização; *highlighting*, que pode ser traduzido como “realce” ou “destaque”, composto por mecanismos que direcionam a atenção do leitor a um determinado conjunto de elementos visuais; e *orientation guidance* ou “condução de transição”, que engloba técnicas para transitar de maneira fluida e natural entre dois layouts da visualização causando o menor impacto negativo possível na compreensão e atenção do leitor.

Por fim, o eixo de estrutura narrativa é composto por: *clustering of different ordering structures* ou “agrupamento de diferentes estruturas de ordenação”, que se refere à maneira como a mídia de visualização organiza o conteúdo sequencialmente; *consistency of interaction design* ou “consistência do design de interação”, que examina se as decisões de projeto quanto à interação entre usuário e visualização permanecem as mesmas ao longo de toda a sua utilização; e *under-utilization of narrative messaging* ou “subutilização de comunicação narrativa”, que avalia o quão utilizadas ou relegadas são as técnicas narrativas mais convencionais e externas à visualização, como textos explicativos.

O grau de explorabilidade da visualizações de dados também é abordado. Os autores fazem uma distinção entre abordagens *author-driven* (ou “orientadas pelo autor”) e *reader-driven* (ou “orientadas ao leitor”). A primeira seguiria um caminho majoritariamente linear de leitura, com pouca interatividade e narrativa fortemente presente. Já a última refletiria casos com muita liberdade de exploração pelo usuário, sem narrativa principal ou ordem pré-definida de leitura.

Vale salientar que essas definições são extremos de um espectro. Uma visualização de dados, em geral, se localiza em algum lugar intermediário dentro desse contínuo. Ademais, em tal publicação, levanta-se a tendência de que cada vez mais frequentemente profissionais da área buscam criar visualizações que se encaixem próximo do centro desse espectro, mostrando as duas abordagens de forma balanceada.

Já em 2016, Lu et al. [16] estabelecem um panorama sobre publicações recentes na área de visualização de dados ao analisar cerca de 70 artigos publicados na IEEE VIS, uma das maiores e mais tradicionais conferências do meio, naquele ano.

Em tal estudo, os autores propõem classificações para visualizações dessas publicações em relação ao que chamam de “filosofia de design”, que concerne às formas de representação dos dados, e “filosofia de exploração”, que concerne às formas de interação entre usuário e aplicação.

Quanto à filosofia de design, a publicação categoriza cada obra em: *multi-view*, consistindo de múltiplas formas interligadas de visualizar a informação; *stand-alone*, consistindo de uma única forma de visualização disponível; *mixed-in*, que se trata de uma visualização híbrida oriunda da combinação de duas ou mais visualizações; *add-on*, definida como uma forma de aprimorar um formato já existente adicionando visualizações auxiliares ainda mantendo a familiaridade da visualização original; e *physical*, na qual a visualização é manifesta em um objeto físico e não em formato digital.

Já em relação à filosofia de interação, as categorias possíveis descritas foram: *overview-detail*, que se norteia em permitir uma exploração global em primeiro momento e depois de seus detalhes sob demanda do usuário; *brush-link*, que conecta explorações em diferentes layouts de visualizações *multi-view*; *exploration-recommendation*, que toma a exploração dos usuários como subsídio para recomendar interações para cada caso específico; *progressive exploration*, que modifica o recorte dos dados exibido iterativamente à medida que a exploração ocorre; *query*, presente em sistemas focados na consulta de dados; *interaction enhanced exploration*, que engloba formas de exploração focadas nas interações registradas e *immersive*, que imerge os usuários no ambiente da visualização aliando-se à realidade virtual.

Conforme indicado por tal estudo, a maioria das publicações analisadas eram de múltiplos layouts interligados ou de layout único, quando considerada a filosofia de design. No que concerne a filosofia de interação, *overview-detail* e *brush-link* foram as escolhas mais populares daquele ano.

Capítulo 3

Lista de Aspectos Formais de Intenção e Design

3.1 Introdução à lista

Páginas web que divulgam dados optam, cada vez mais frequentemente, por exibi-los de modo visual, visando uma mais leitura mais clara e fácil. Para isso, incorporam visualizações de dados junto a textos, imagens e outros formatos em um mesmo tecido semântico. Os diferentes elementos que compõem a página, assim como as interações estabelecidas entre eles, precisam ser melhor compreendidos de forma a garantir a transmissão do conteúdo e possibilitar a proposição de direti-vas capazes de nortear concepção, projeto e desenvolvimento de novas mídias de visualização de dados.

Esta lista visa desenvolver um estudo preliminar de principais eixos que organi-zam essas mídias, explicitando suas definições e seus fundamentos além de prover uma coleção de exemplos de páginas web disponíveis na Internet para ilustrá-los. Assim, espera-se obter um guia de fácil acesso aos principais conceitos de mídias de visualização de dados para a web.

É importante salientar, por outro lado, que o escopo deste trabalho não inclui o design de tipos de visualização, como abordado em outras iniciativas, como nos websites *The Data Visualization Catalogue* [17] e *From Data To Viz* [18], que também oferecem exemplos de implementação de cada formato com base na natureza do conjunto de dados sobre o qual se trabalhará.

Também não serão discutidos escolhas de implementação e tecnologias. Neste momento, busca-se um levantamento conceitual sobre visualizações de informação e os meios nos quais se inserem, um passo anterior à codificação.

3.2 Intenção da mídia de visualização

Os tópicos relativos à intenção da mídia de visualização refletem não somente a visualização de dados, mas também o contexto no qual ela se insere. Assim, podemos observar que outros elementos dão suporte a uma visualização e quais são por ela suportados.

3.2.1 Familiaridade do público

O público-alvo de uma visualização de dados consiste no conjunto de indivíduos que busca-se alcançar com a publicação da visualização. A diversidade de *backgrounds* (sociais, culturais) entre diferentes públicos impacta diretamente no sucesso de uma visualização devido a fatores como:

1. **Familiaridade com os dados apresentados**

Diferentes usuários da visualização podem apresentar diferentes graus de conhecimento prévio sobre a temática abordada e os próprios dados exibidos. Públicos mais versados no assunto desenvolvido podem apresentar maior facilidade em entender uma visualização e obter conclusões com base nela. Por outro lado, para públicos que tiveram menor contato anterior com o mesmo assunto, textos adicionais de ajuda na interpretação dos dados são bem-vindos.

2. Familiaridade com formato da visualização

Além da familiaridade com os dados, o conhecimento prévio dos usuários de formatos de visualização pode lhes conferir uma sensação de maior facilidade de leitura e, do ponto de vista da interação, de maior facilidade de uso. Por exemplo, públicos que já consumiram visualizações de dados com recursos de clique e arraste no passado podem receber com maior naturalidade e interagir com maior rapidez ao serem apresentados a uma nova visualização do mesmo tipo. O mesmo se aplica a formatos de representação de dados, tais como gráficos de barra, mapas e grafos.

Dessa forma, conhecer o público-alvo ao qual uma visualização de dados se destina é parte fundamental para seu desenvolvimento pois embasa decisões de design como a densidade de informação e detalhamento exibidos na página bem como a forma de exibição de textos explicativos e *affordances*.

Evidenciando a dimensão dessa questão, Min Lu et al. [16] lista a capacidade de adaptação de sistemas de visualização de informação ao conhecimento previamente acumulado dos usuários como um dos grandes problemas ainda não solucionados na pesquisa de visualização.

Imports and Exports to and from Canada [19], desenvolvida pelo Canada Energy Regulator do Governo do Canadá, é um exemplo de visualização desenvolvida tendo em mente um público-alvo já familiarizado com o tema, neste caso, de importação e exportação de energia no Canadá.

Já *The Deadliest Jobs in America* [20], veiculada no portal de notícias Bloomberg, é um exemplo de visualização destinada a um público mais abrangente e sem conhecimento específico sobre o tema dos dados, dado o seu desenvolvimento visando ampla circulação em um portal de notícias. Também que não demanda múltiplas visitas por um usuário, visto que sua base de dados refere-se a um momento passado específico e não sofrerá alterações.

Quanto à familiaridade com o formato da visualização, duas visualizações podem ter como tema central e fonte do conjunto de dados o mesmo assunto, mas utilizar representações gráficas distintas. Por exemplo, *How long are characters in each episode?* [21], mostrada na Figura 3.1, e *Narrative Chart* [22], mostrada na Figura 3.2,

são duas visualizações desenvolvidas por Jeffrey Lancaster acerca do seriado Game of Thrones. No entanto, enquanto a primeira delas utiliza um gráfico de barras, formato amplamente utilizado para representação de dados, seja na web ou em jornais e televisão, a segunda visualização opta por uma representação em fluxo, que pode ser pouco intuitiva especialmente para visitantes não habituados a esse formato.

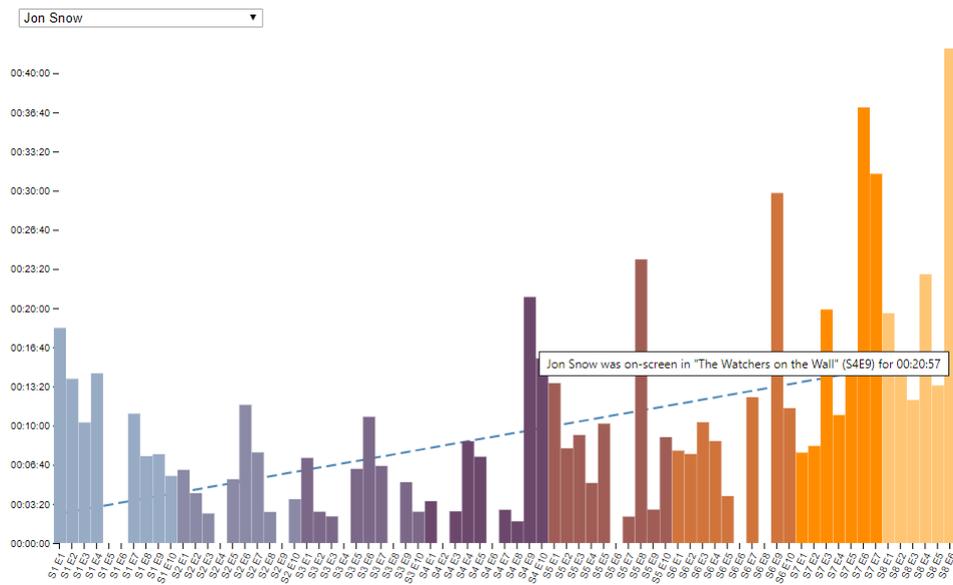


Figura 3.1: Captura de tela de *How long are characters in each episode?*.

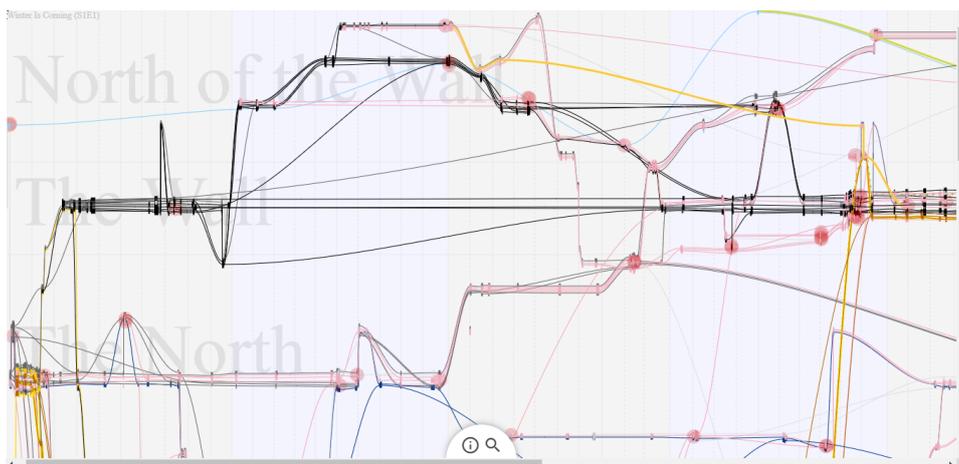


Figura 3.2: Captura de tela de *Narrative Chart*.

3.2.2 Grau de explorabilidade

Visualizações de dados são desenvolvidas com diferentes propósitos, podendo atender a diferentes demandas. Consideramos duas extremidades de um espectro de quão aberta à exploração uma visualização pode ser. São elas:

1. **Explicação**

De um lado, podemos conceber visualizações que possuem como principal finalidade ilustrar análises já consolidadas como por exemplo, em um gráfico que acompanha o texto de uma reportagem jornalística. Nesse caso, a visualização atua como o fio condutor de uma narrativa a ser transmitida ao leitor. Deste modo, visualizações explanatórias não oferecem recursos que fomentem outras análises dos dados, dificultando a produção de diferentes conclusões ou interpretações sobre eles.

2. **Exploração**

No outro extremo do espectro de explorabilidade, existem visualizações oferecidas como um espaço aberto para manipulação interativa, conforme demanda do leitor. Por exemplo, filtros e codificações podem ser aplicados sobre o conjunto de dados, gerando um recorte específico do mesmo, viabilizando desde descobertas e conclusões específicas sobre o conjunto de dados, até a percepção de novas tendências em séries temporais ou, ainda, favorecendo a validação de hipóteses. [23]

Uma mesma mídia de visualização pode trazer elementos de explicação e exploração simultaneamente. Portanto, devemos pensar cada caso como não apenas dentro desse espectro como também em lados diferentes desse espectro dependendo do uso que um visitante faz dela.

Vale ressaltar a distinção entre explorabilidade e interatividade. Apesar de, em muitos casos, tais conceitos poderem andar lado a lado, uma página e suas visualizações de dados podem contar com diversas formas de interação sem de fato oferecer ao usuário controle sobre a manipulação de dados e a abertura para diversas conclusões e narrativas a eles subjacentes, como apresentado e discutido no livro de

Alberto Cairo, *The Functional Art: An introduction to information graphics and visualization* [24].

Esse é o caso de mídias de visualização como as do projeto *Writing Without Words*, de Stefanie Posavec [25]. Composto por uma coleção de imagens estáticas que permitem visualizar a estrutura textual do romance *On The Road* [26], de Jack Kerouac, o trabalho oferece ao leitor um grande potencial de geração de *insights* e descobertas sobre os dados, sendo portanto uma mídia exploratória, mesmo sem oferecer formas de interação.

A importância dada a esse aspecto é espelhada em outras publicações. Por exemplo, na discussão de Segel e Heer (2010) [15], distingue-se abordagens *author-driven* de *reader-driven* em visualizações, análogas aos referenciais “estática” e “explorável” aqui propostos.

Já Heer et al. (2014) relacionam exploração e familiaridade do público e discutem profundamente as diferentes demandas de cada grupo em relação a softwares de visualização [27].

Um exemplo de mídia de visualização de dados que oferece grande capacidade de exploração ao usuário é o BibViz Project [28], que permite analisar inconsistências encontradas em qualquer texto bíblico interativamente e a partir de várias fontes.

No extremo oposto desse espectro, podemos mencionar a visualização *World's Best Hangover Cure?* [29] que é exclusivamente dedicada a leitura e tem pouco a oferecer em termos de *insights* ao visitante, tornando o viés exploratório mais ausente.

3.2.3 Frequência de visitaç o

Uma visualizaç o de dados pode conter um conjunto de dados que permanece o mesmo ao longo do tempo. Esse   o caso de muitas visualizaç es de dados desenvolvidas com base em uma pesquisa de opini o ou censo, por exemplo.

Por outro lado, uma mesma visualização pode também receber continuamente novos conjuntos de dados atualizados em tempo real. Esse é o caso de sistemas de monitoramento que incluem, por exemplo, interfaces que acompanham o deslocamento de veículos em uma cidade ou que registram focos de incêndio em um mapa de uma área florestal.

Assim, a depender do propósito de uma visualização e das características dos dados utilizados, uma visualização pode estar sujeita desde a uma única visita até a visitas recorrentes de um mesmo leitor, refletindo a frequência com o que os dados mostrados por ela se modificam e como isso se reflete no sistema.

Podemos estabelecer uma escala de classificação formada de: “visitação única”, “visitação ocasional” e “visitação recorrente”. Páginas de visita única encontram-se relacionadas ao emprego de dados que não sofrem alterações ao longo do tempo e que, portanto, são compreendidos em uma única visita. Visitação ocasional pode tratar, por exemplo, do monitoramento de dados econômicos e/ou censitários, que demandam mais de uma visita. E páginas de visita recorrente tratam de dados que necessitam de acompanhamento constante, como é o caso, do monitoramento de queimadas em tempo real, que oferece atualização recorrente dos dados.

Por exemplo, o portal da situação atual do Programa Queimadas do Instituto Nacional de Pesquisas Espaciais (INPE) [30] oferece uma visualização, de origem nacional e destinada ao monitoramento de queimadas em tempo real, que enseja a visita recorrente por parte de seus usuários, tendo em vista a rápida atualização de sua base de dados.

Já visualizações como o Atlas da Complexidade Econômica [31] e o Observatório da Complexidade Econômica [32], desenvolvidos a partir de dados econômicos ou censitários, demandam mais de uma visita do usuário, porém em menor frequência que as do exemplo anterior. Podemos, então, considerar este um exemplo de visualização com visita ocasional.

3.3 Características de design

Os seguintes tópicos apresentados dão enfoque às escolhas relativas à forma de apresentação do conteúdo que será exibido na página. Nesta etapa, não se considera o contexto, já analisado anteriormente. Serão discutidas formas de representação dos dados e formas de interação com a aplicação.

3.3.1 Forma de manipulação dos dados

Em uma página web com uma visualização de dados, frequentemente os dados exibidos podem ser manipulados de forma interativa. Podemos subdividir essa forma de interação em duas grandes abordagens:

1. Convencional

Inclui formas de interação como menus e botões. Nessa categoria, a interação se dá com um elemento gráfico apartado do local onde os dados são exibidos visualmente. Dessa forma, o usuário não age diretamente sobre os dados.

2. Manipulação direta

É uma forma de interação mais complexa que a anterior e que exige mais informação sobre a interação. Nesse caso, a manipulação é feita pelo arraste dos elementos na tela, que têm sua posição alterada, influenciando na forma com que os dados são apresentados [33].

Enquanto formas de manipulação convencionais são muito comumente encontradas em uma variedade de visualizações, como nos exemplos anteriores, o uso de manipulação direta é menos frequente. Um exemplo de visualização que utiliza manipulação direta é o mapa interativo para comparação das paisagens contemporânea e histórica da cidade de Nova Iorque chamado *Spyglass on the Past: New York City 1836 and Today* [34], em que a ação de arraste sobre o próprio mapa é necessária para revelar as informações de interesse no segundo mapa para o visitante. Uma captura de tela dessa mídia é exibida na Figura 3.3.



Figura 3.3: Exemplo de interação por manipulação direta em *Spyglass on the Past: New York City 1836 and Today*.

3.3.2 *Feedforward*

O termo *feedforward* refere-se à capacidade de uma visualização de dados ou um de seus elementos gráficos transmitir ao usuário o que ocorrerá após determinada interação ser realizada. [5] Ou seja, a presença de *feedforward* dá ao usuário a noção de qual será o resultado obtido como fruto de sua interação e, portanto, oferece um entendimento antecipado das funcionalidades disponíveis.

Em *How Will Facebook's Stock Fare?* [35], uma visualização veiculada no The New York Times acerca do preço esperado das ações do Facebook em seu primeiro dia de capital aberto, o número de comentários colhidos em determinada faixa de preço é mostrado ao passar o ponteiro do mouse sobre a barra no topo da visualização antes de efetivamente selecionar tal faixa. Esse é um caso de *feedforward*.

Nas visualizações sobre controle de armas da Periscopic [36, 37], o botão à esquerda da tela que permite transitar entre os dois layouts disponíveis é dotado de *feedforward*, como visto na Figura 3.4. Ao posicionar o cursor do mouse sobre tal botão, sem efetivamente pressioná-lo, o layout alternativo aparece sobreposto ao layout vigente, adiantando a mudança que ocorrerá como resultado de sua interação.

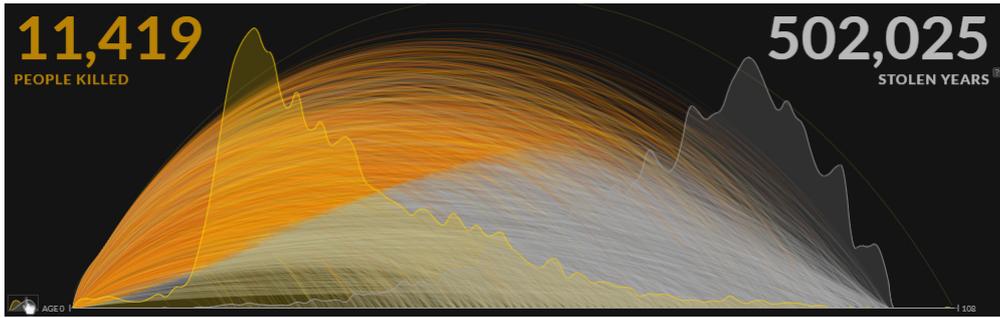


Figura 3.4: Exemplo de *feedforward* em *U.S. Gun Deaths in 2013*.

3.3.3 *Dashboards*

Podemos considerar que uma *dashboard* é constituída por várias visualizações complementares apresentadas em uma mesma página. Esse tipo de aplicação atua como um painel, geralmente destinado a públicos especializados e com alguma familiaridade com o tema, e tem como objetivo dar suporte a tarefas de supervisão e controle, proporcionado pela composição de várias visualizações correlatas em uma mesma tela. Suas aplicações mais populares são encontradas no mercado financeiro, no setor de energia e no monitoramento de tráfego.

A visualização de dados *Census Data Visualization* [38], que reúne dados censitários de cada estado estadunidense, é um exemplo de *dashboard*. Em uma mesma tela, várias visualizações correlatas são apresentadas ao visitante simultaneamente. A Figura 3.5 exibe a tela inicial de tal mídia.

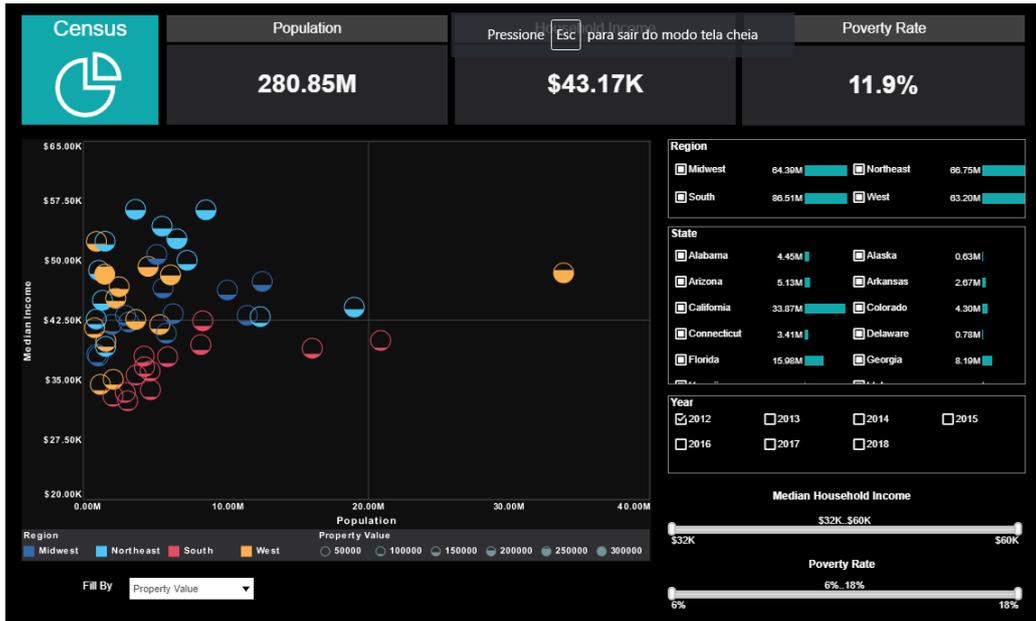


Figura 3.5: Captura de tela de *Census Data Visualization*.

3.3.4 Número de layouts

Em algumas visualizações, observam-se representações gráficas distintas a partir de um mesmo conjunto de dados. Layouts diferentes permitem diferentes perspectivas sobre os dados, a partir da comparação de registros, criação de rankings ou localização em um mapa.

Por exemplo, a visualização GOV—DNA [39] contém quatro diferentes layouts para exibição do mesmo conjunto de dados, apresentados na Figura 3.6. Já as visualizações sobre controle de armas da Periscope [36, 37], dois layouts complementares são oferecidos, como na Figura 3.7.

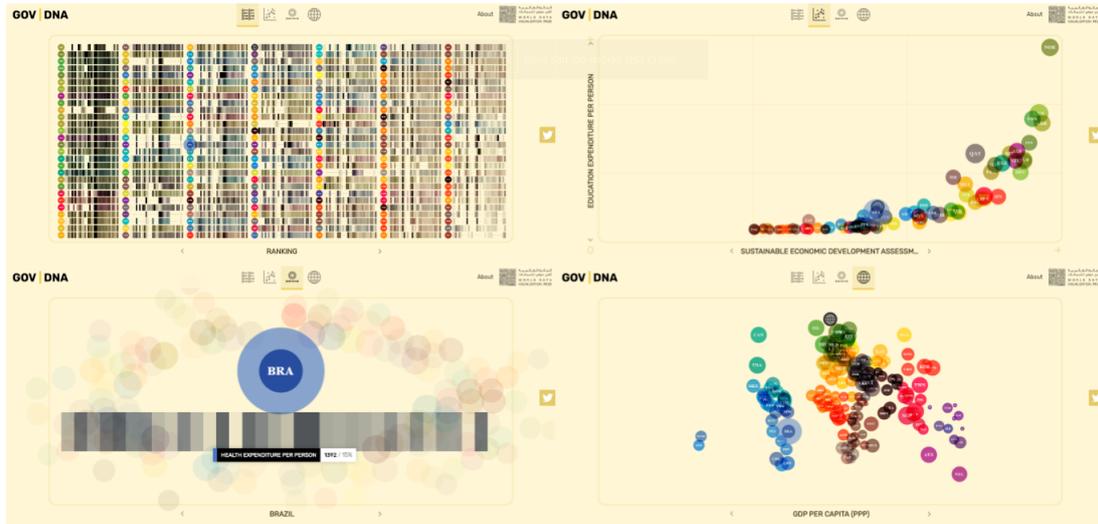


Figura 3.6: Layouts apresentados por *GOV—DNA*.

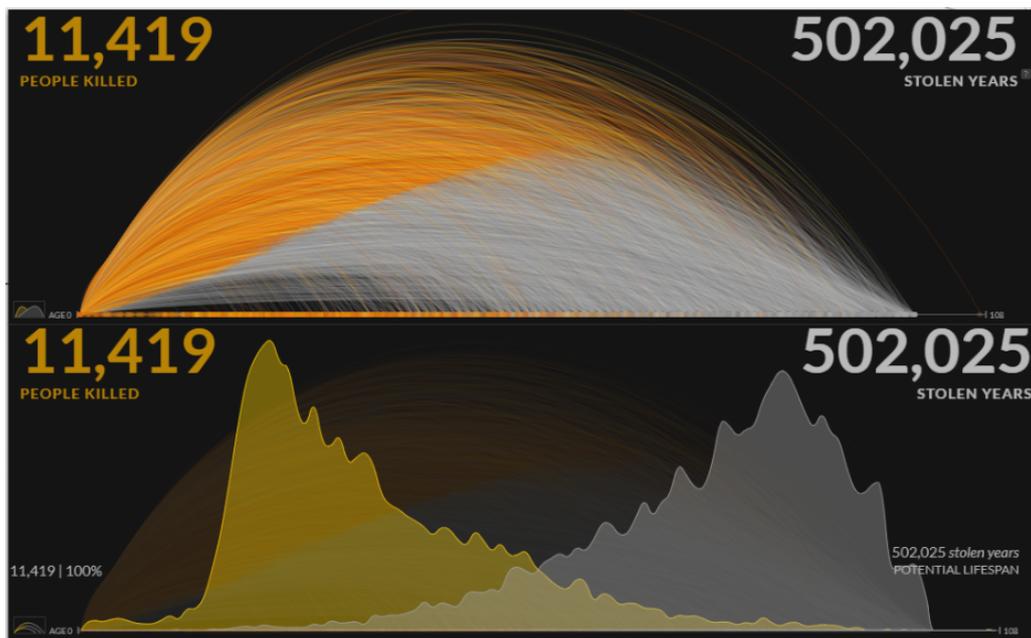


Figura 3.7: Layouts apresentados por *U.S. Gun Deaths in 2013*.

3.3.5 Continuidade temporal

Quando uma visualização possui mais de um layout para os mesmos dados, esses layouts podem ser substituídos na página de forma abrupta ou contínua. Nesta última, ao alterar o layout, as representações dos dados são animadas de forma suave de maneira a ser acompanhada pelo olho. Assim, é possível acompanhar um registro nos diferentes modos de visualização.

GOV—DNA [39] também é um exemplo de visualização que exhibe continuidade entre cada layout oferecido. Animações de transição que preservam a seleção feita pelo usuário no layout anterior garantem uma experiência mais fluida e natural.

3.3.6 Animação

Animações são um importante recurso em visualizações, porém podem adquirir especial utilidade ao assumirem o papel de *affordances*. Ao animar um ou mais elementos de uma visualização, transmite-se ao usuário sua natureza dinâmica, implicando que tal objeto ou seus atributos, como posição e cor, podem se transformar após uma determinada ação.

Dessa forma, animações não só atraem a atenção do leitor como podem alertá-lo sobre uma possibilidade de interação. Além disso, quando a visualização está incluída em outra mídia, como um texto, as animações podem convidar o leitor a interagir, evitando que visualizações sejam confundidas com imagens estáticas e passem despercebidas por ele.

Nesse contexto, podemos categorizar o uso de animações em uma visualização de dados de três principais formas:

1. Ausente

Neste cenário, não são utilizadas animações na visualização. Todos os elementos são mostrados estáticos.

2. Inicial

Animações iniciais ocorrem no momento em que uma visualização (ou parte dela) é carregada e exibida ao leitor pela primeira vez. Uma vez concluída a animação, após um período definido de tempo ou após uma primeira interação (como um clique), tal animação não volta a ser exibida. Sua presença é utilizada comumente ao apresentar novas possibilidades de interação. Adicionalmente, quando essa animação sugere o que irá ocorrer após a interação, tal caso também caracteriza *feedforward*.

3. Em resposta a uma interação

Animações podem ocorrer acionadas por uma ação provocada pelo usuário. Por exemplo, mudanças de layout que possuem animações garantem uma transição suave ao usuário.

4. Persistente

Finalmente, animações persistentes permanecem ocorrendo ao longo do uso da visualização. Em geral, sua presença recorrente garante que determinado elemento visual receba a atenção do leitor.

Por exemplo, a visualização *The Most Violent Cities in The World* [40] é apresentada na forma de uma imagem estática, não oferecendo qualquer animação. Assim, podemos dizer que esse é um caso de visualização de dados com animações ausentes.

Já páginas como *U.S. Gun Deaths in 2010* [36] e *U.S. Gun Deaths in 2013* [37], da Perisopic, possuem exemplos de visualizações com animação inicial, que também incluem explicações para a leitura dos dados e *feedforward*, como na Figura 3.8.

Por fim, *Migrations in Motion* [41], da organização *The Nature Conservancy* mostra movimentos de migração de animais por todo o mundo. Nela, as animações que representam os fluxos migratórios são reproduzidos em *loop*, persistentemente.

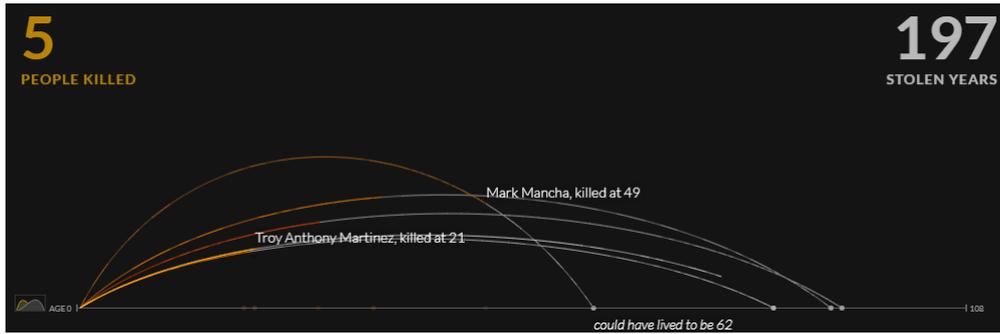


Figura 3.8: Exemplo de uso de animação inicial em *U.S. Gun Deaths in 2013*.

3.3.7 Elementos explicativos adicionais

Elementos explicativos adicionais são elementos (comumente textos e ícones) que não fazem parte da interface primária da visualização, mas que dão suporte à compreensão do leitor. Um elemento explicativo é algo que se retirado não afeta o funcionamento da página.

Com base nessa definição, existem três tipos de explicação de uso quanto a seu teor:

1. De exibição/de leitura

Elementos explicativos de exibição e leitura dão ao usuário informações para interpretar corretamente um formato utilizado na visualização, como um gráfico ou um mapa. Assim, busca-se dar clareza às escolhas de representação utilizadas na página.

2. Do conjunto de dados

Elementos explicativos podem, também, especificar informações como a fonte dos dados, a data de coleta, a metodologia utilizada em sua obtenção e outras características sobre o conjunto de dados como um todo. Isso permite ao usuário conhecer mais sobre a proveniência (*provenance*) dos dados, garantindo a confiabilidade do que está sendo visualizado. Em geral, explicações acerca do conjunto de dados são oferecidas de forma textual.

3. De uso/de interação

Essa classificação se inclui, de maneira geral, no conceito de *affordances* anteriormente mencionado. São elementos que indicam que uma forma de interação existe e qual é o resultado esperado na realização da interação.

Tomemos como exemplo a visualização *Out of Sight, Out of Mind* [42], que traz uma reflexão sobre o uso de *drones* em ofensivas militares no Paquistão a partir de 2004 e inspirada nas visualizações sobre mortes por arma de fogo da Periscope [36, 37]. Nela, são exibidos, na seção "INFO", textos explicativos dos três tipos descritos – de exibição/de leitura, do conjunto de dados e de uso/de interação – acerca dos elementos gráficos utilizados e do conjunto de dados que serviu de base para seu desenvolvimento e das interações possíveis, além de também contar com uma animação inicial de apresentação. Um trecho deles é mostrado na Figura 3.9.

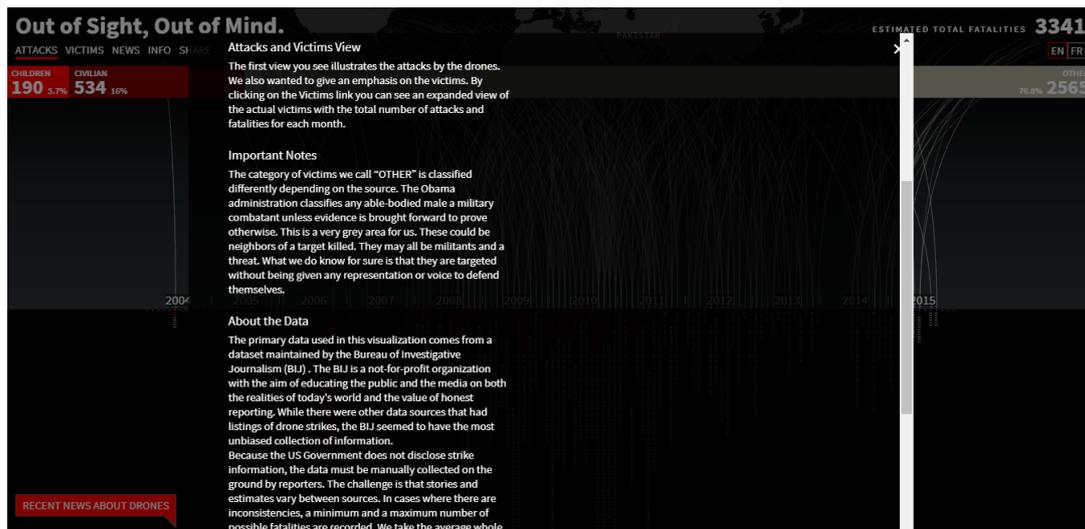


Figura 3.9: Exemplo de elemento explicativo adicional em *Out of Sight, Out of Mind*.

O mesmo ocorre com *A Day in the Life of Americans* [43], uma página que combina visualização de dados e explicações em texto interpretando o que é visto e trazendo detalhes ao leitor. É essencial notar a importância do bom uso de elementos explicativos para assegurar a compreensão da mídia de visualização por todos os públicos. Esse aspecto é especialmente relevante em visualizações que serão amplamente veiculadas ou que utilizam formatos com os quais o público-alvo pode não estar completamente familiarizado.

Paralelamente, também podemos classificar explicações de uso quanto a sua exibição na tela da seguinte maneira:

1. **Ausente**

Não há elemento explicativo visível na tela. Nesse caso, o usuário deve explorar de forma independente a visualização ou possuir conhecimento prévio sobre ela para entendê-la.

2. **Sob demanda**

Os elementos explicativos não estão visíveis por padrão. No entanto, o usuário pode realizar uma ação, como posicionar o mouse sobre um local ou clicar em um ícone, para revelar a explicação de uso. Uma vez lida, a explicação de uso volta a se tornar oculta. Isso contribui para interfaces com menos *cluttering* (poluição visual causada por grande número de elementos sendo exibidos simultaneamente) e um melhor aproveitamento do espaço disponível para a exibição dos dados.

3. **Persistente**

Um elemento explicativo pode estar constantemente disponível para leitura junto à visualização. Apesar de aumentar as chances de ser visto pelo usuário, pode comprometer a experiência de uso de usuários que já conheçam tal explicação. É frequentemente encontrada na forma de subtítulos e legendas.

Earth: a global map of wind, weather, and ocean conditions [44] é um exemplo em que explicações de uso estão ausentes. Nessa visualização, o usuário é convidado a explorar por conta própria uma combinação de dados atualizados várias vezes ao dia acerca da atmosfera e dos oceanos em todo o globo.

Por sua vez, *The Daily Routines of Famous Creative People* [45], que compara as rotinas de personalidades de carreiras criativas, é uma visualização que apresenta textos explicativos de exibição de exibição/de leitura à medida que o visitante da página seleciona diferentes opções na barra de menu ou posiciona o mouse sobre as barras do gráfico. O mesmo ocorre com a comparação de índices de felicidade entre diversos países do mundo *What do the happiest countries in the world have in common?* [46]. Portanto, suas explicações podem ser classificadas como acionadas sob demanda.

3.3.8 Integração com narrativa

Este item aborda o grau de integração entre a visualização e a narrativa que a acompanha na página. Por exemplo, visualizações integradas à narrativa podem ser complementares a grandes textos ou apresentadas sequencialmente para ilustrar uma narrativa. Esse caso é frequentemente encontrado em textos jornalísticos e institucionais.

A visualização *Women's Pockets are Inferior* [47] é um bom exemplo de visualização profundamente integrada com uma narrativa textual que a acompanha na página, neste caso, para comparar tamanho e desenho de bolsos de roupas masculinas e femininas.

3.3.9 Tooltips e/ou popups

Visualizações de dados podem contar com elementos adicionais exibidos em sobreposição à tela original em resposta a uma interação do visitante. Apesar de não serem elementos indispensáveis ao uso da página, esses recursos ainda assim são amplamente empregados em especial para a redução da desordem visual (*cluttering*) na tela.

Tooltips, ou dicas de contexto, referem-se a molduras flutuantes exibidas sobre um outro elemento gráfico quando o usuário mantém o ponteiro do mouse sobre este último. Em geral, são utilizadas para prover mais informações relacionadas à representação e podem ser fechadas ao retirar o ponteiro do local. É importante destacar que as *tooltips* constituem uma das formas de interação mais básicas, por não demandarem cliques, arrastes, menus ou quaisquer outras ações que exijam mais atenção e foco do visitante.

Já *popups* são janelas auxiliares criadas sobre a página original. Diferentemente das *tooltips*, *popups* requerem uma interação adicional por parte do usuário para o seu fechamento. Além disso, comumente possuem o diferencial de serem interfaces modais, impedindo ações sobre a página ao fundo. Por isso, essas peculiaridades geralmente implicam num maior nível de atenção do leitor à informação por elas trazida.

Podemos classificar o emprego de *tooltips* e *popups* em mídias de visualização de dados como:

1. Ausente

Páginas com visualizações de dados podem não utilizar recursos de *tooltips* e *popups*, em geral, por julgarem que a funcionalidade dessas páginas não é comprometida.

2. Detalhes do dado

Tooltips e *popups* podem detalhar a informação e apresentar dados adicionais aos apresentados anteriormente, evitando o preenchimento excessivo da tela.

3. Dica de interação

Além disso, podem ser apresentadas instruções para uso e resultado esperados das funcionalidades interativas da página, complementando o papel de *affordances*.

4. Metadados

Esses elementos podem apresentar informações acerca do conjunto de dados, ainda que não inclua novos dados como no item 2. Tais adições podem incluir unidades de medida e formas de leitura dos dados, por exemplo.

Selfiecity [48] é uma visualização que investiga o estilo das fotografias de auto-retrato em diferentes lugares do mundo. Em sua exibição inicial, a visualização oferece um panorama geral de todo o conjunto de fotografias estudado. É através de *tooltips* que o usuário se torna capaz de visualizar um foto específica em tamanho maior, como mostrado na Figura 3.10. Por isso, podemos dizer que essa é um *tooltip* de detalhes do dado.

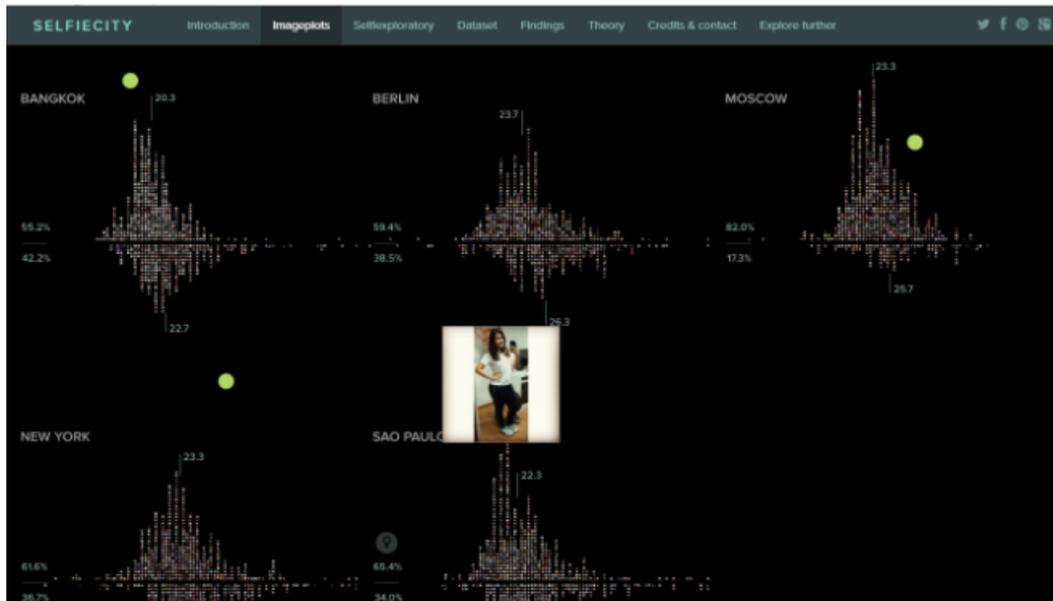


Figura 3.10: Exemplo de *tooltip* com detalhes do dado em *Selfiecity*.

Já o *Project Ukko* [49] é uma visualização de ventos sazonais ao redor do mundo, codificando dados como velocidade, variação e capacidade de geração de energia eólica em cada ponto aferido. Ao posicionar o cursor do mouse sobre um dos pontos, uma *tooltip* contendo o texto “Click to see details about this region.” é exibida. Ao clicar, novos dados sobre observações passadas, previsões e valores numéricos dos dados são exibidos. Esse é um exemplo de *tooltip* de dica de interação e pode ser visto na Figura 3.11.

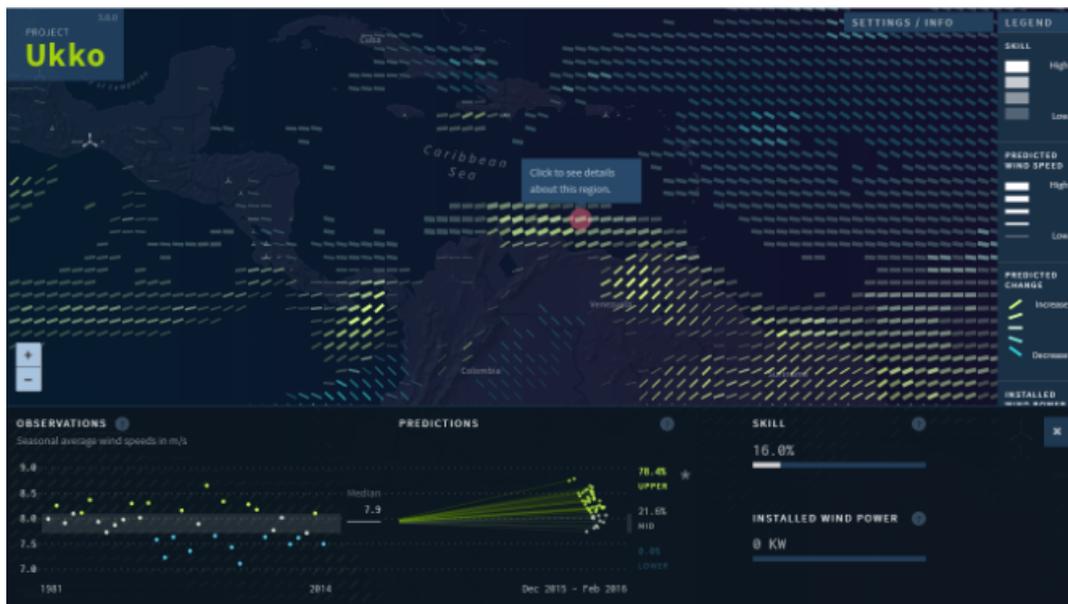


Figura 3.11: Exemplo de *tooltip* com dica de interação em *Project Ukko*.

Capítulo 4

Exemplo de Aplicação

De modo a ilustrar a aplicação da lista apresentada no Capítulo 3 em sua totalidade em um mesmo exemplo e examinar a aderência do objeto de estudo – os aspectos formais – a um caso real, podemos analisar a visualização interativa *Visualizing Visualizers* [50], de autoria de Doris Kosminsky, Lucas Barcellos Oliveira e Claudio Esperança, categorizando-a segundo cada item discutido. Assim, podemos observar o uso de cada aspecto formal na caracterização de uma mesma mídia de visualização e em um cenário no qual a mídia estudada já é familiar, verificando num primeiro teste sua aplicabilidade.

4.1 Descrição da visualização

Tal visualização foi exibida na IEEE VIS Arts Program 2019, ou IEEE VISAP 2019. Esse é o programa de artes da conferência IEEE VIS, que ocorreu junto ao evento principal em Vancouver, Canadá. Nela, é exibida a rede de colaboração a partir de um autor, com dados extraídos do DBLP [51].

O nome do autor a partir da qual a rede é visualizada aparece em destaque na parte superior da tela. Nomes de pesquisadores com os quais tal autor já escreveu alguma publicação descem pela tela verticalmente, enquanto título, revista/*proceedings* e ano de tais publicações cortam a tela em faixas horizontais laranjas.

Os nomes de co-autores param suas quedas ao colidir com a faixa referente ao artigo no qual colaboraram e continuam sua queda até sair da tela uma vez que a respectiva faixa termina de atravessar a tela. As cores utilizadas nos nomes dos colaboradores codificam qual foi a colaboração mais recente com o autor em destaque, apresentando cores mais intensas nos colaboradores mais recentes.

A visualização codifica, também, alguns atributos das conexões entre o autor em destaque e cada colaborador nessa rede. Primeiramente, cada nome de colaborador pode ser acompanhado por repetições translúcidas que oscilam ao seu redor. O número total de réplicas de dado nome é o número de publicações co-escritas por ele e pelo autor em destaque.

Além disso, colaboradores dos colaboradores, ou seja, os autores com dois graus de separação do autor em destaque também descem pela tela em cor cinza e tamanho da fonte menor, ligados por uma linha elástica aos colaboradores de primeiro grau que os conectam ao autor de destaque na rede. A Figura 4.1 mostra a mídia de visualização exposta no evento, utilizando um monitor. Já a Figura 4.2 mostra uma captura de tela da visualização.

4.2 Decisões de implementação

A visualização foi desenvolvida em JavaScript com a biblioteca Matter.js como *engine* física e utilizando a plataforma online reativa Observable. A apresentação foi realizada com um monitor LCD montado em orientação retrato e visitantes do evento puderam inserir seus nomes em um teclado numa tela sensível ao toque posicionada ao lado para torná-los o nome em destaque da visualização e ver sua própria rede de colaboração.

Os dados brutos do DBLP são referentes a 8 de Setembro de 2019. Eles foram obtidos em formato XML e carregados através de um *script* na linguagem de programação Python para uma base de dados MySQL, cujo servidor era acessado localmente por *proxy*. No total, mais de 1,5 milhão de autores e mais de 20 milhões de publicações podiam ser visualizados.

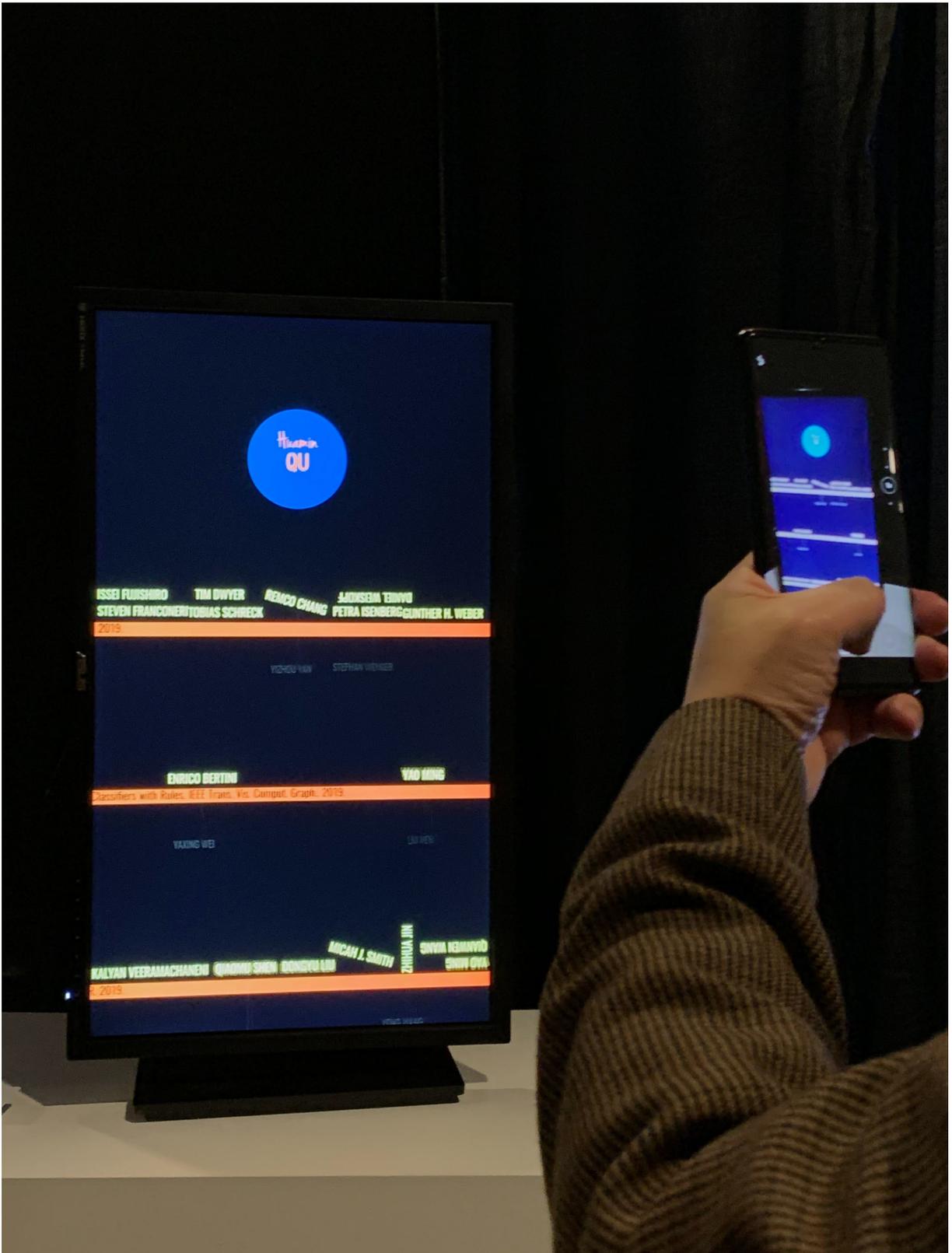


Figura 4.1: Exposição de *Visualizing Visualizers*.



Figura 4.2: Captura de tela de *Visualizing Visualizers*.

4.3 Quanto à intenção da mídia

4.3.1 Quanto à familiaridade do público

O público para o qual a visualização foi voltada é composto por pesquisadores que compareceram à IEEE VIS 2019. Assim, podemos dizer que o público-alvo da visualização é altamente familiar com visualizações de dados. Ademais, tendo em vista que os dados abordam as produções científicas do próprio público-alvo, consideramos também que ele é altamente familiar ao conjunto de dados.

O formato escolhido para a visualização não é convencional, tendo sido pensando especificamente para essa visualização. Por isso, não é possível afirmar que todos os visitantes possuíam afinidade com tal formato.

4.3.2 Quanto ao grau de explorabilidade

O elemento de exploração oferecido pela visualização é a seleção de diferentes nomes como nomes em destaque através do teclado. Assim, um visitante pode percorrer toda a rede de colaboração inferida dos registros do DBLP.

No entanto, uma vez selecionado um nome, a visualização é predominantemente explanatória. Não é oferecido ao usuário, por exemplo, a aplicação de filtros ou a ordenação dos dados, o que reduz a chance de alcançar novas conclusões e interpretações a partir deles.

Isso ocorre, neste caso, devido à própria natureza dos dados. As informações sobre artigos e colaborações são públicas e não são frequentemente alteradas. Com isso, elas podem já ser previamente conhecidas pelo visitante antes de interagir com a visualização. A visualização analisada dá enfoque a uma forma diferenciada de expressar visualmente tais dados.

4.3.3 Quanto à frequência de visitação

Dado o ambiente de exposição e a proposta de que vários visitantes interajam com a visualização, espera-se que a frequência de utilização seja única.

No entanto, a ordem de exibição dos artigos e colaboradores é selecionada aleatoriamente a cada execução. Dessa forma, caso o mesmo nome seja inserido no teclado como nome em destaque mais de uma vez, o visitante pode ter uma experiência totalmente distinta a cada utilização.

4.4 Quanto aos aspectos de design

4.4.1 Quanto à forma de manipulação dos dados

A única forma de interação disponível é a de inserção de outros nomes utilizando o teclado. Não há controle direto sobre os colaboradores e artigos exibidos na tela. Por isso, podemos considerar que a forma de manipulação é convencional, feito indiretamente.

4.4.2 Quanto à existência de *feedforward*

Na animação principal, não há elementos que forneçam ao visitante alguma forma de *feedforward*. Por outro lado, podemos observar também a forma de entrada de novos nomes. Nota-se que o teclado exibe, à medida que o usuário digita, sugestões de nomes presentes na base de dados. Veja a Figura 4.3. Essa funcionalidade de autocomplemento (*autocompleting*) dá ao visitante a noção de que nomes poderão ser selecionados por ele e se enquadra como um recurso dotado de *feedforward*.

4.4.3 Quanto à presença de *dashboards*

Não há qualquer *dashboard* presente nesta visualização. Ela não associa diversos layouts em uma única tela.

4.4.4 Quanto à existência de mais de um layout

A visualização oferece apenas um layout para sua leitura. Tal escolha se mostra associada às circunstâncias da apresentação ao público, num evento expositivo.



Figura 4.3: Teclado utilizado em *Visualizing Visualizers*.

4.4.5 Quanto à continuidade temporal

Devido à resposta ao item anterior, de que a visualização possui apenas um layout para os dados, o aspecto de continuidade temporal não se aplica a este caso.

4.4.6 Quanto às animações

As animações centrais desta visualização, que são a “queda” de nomes de colaboradores e a passagem de faixas horizontais relativas a publicações, ocorrem persistentemente, por um período indeterminado de tempo. Além disso, uma vez exibidos todos os nomes de colaboradores e dados de publicações disponíveis, a animação volta a exibir os dados desde o início, em *loop*.

Essas animações podem ser consideradas animações persistentes. Elas atuam no decorrer de toda a visualização e não cessam, mesmo se permanecer ociosa, sem qualquer nova interação de visitantes, indeterminadamente.

Esta visualização também conta com uma animação assim que é carregada e toda a vez em que um novo nome é selecionado no teclado para permanecer em destaque.

Nela, o nome em destaque anterior (se houver) desaparece gradativamente, o círculo azul no qual ele é posicionado se expande, cobrindo totalmente a tela, e depois se retrai, quando o novo nome em destaque aparece gradativamente. Tal animação tem como objetivo garantir uma transição suave entre a visualização de um nome e de outro, cobrindo a tela para permitir a remoção dos elementos antigos e revelar um novo espaço ainda não preenchido por nomes e publicações. Ela também torna imperceptível o tempo de carregamento dos novos dados a serem visualizados.

Já essas animações atuam, simultaneamente, como animações iniciais e animações em resposta a uma interação. O mesmo conjunto de animações é executado no momento do primeiro carregamento da visualização, quando um nome *default* é colocado em destaque, e no momento em que um visitante seleciona um nome diferente no teclado.

4.4.7 Quanto aos elementos explicativos adicionais

Não há elementos explicativos adicionais associados à visualização. No evento em que foi apresentada, informações adicionais sobre o conjunto de dados e sobre a codificação adotada na visualização foram providas no próprio local de exibição. Entretanto, visitantes que a acessam em data posterior, através da página na plataforma Observable, não dispõem de quaisquer explicações adicionais.

4.4.8 Quanto à existência de *tooltips* e *popups*

Visualizing Visualizers não conta com qualquer *tooltip* ou *popup*. Isso está relacionado, mais uma vez, com a forma de apresentação da visualização no evento no qual foi exibida.

Primeiramente, visto que a única forma de interação prevista é a inserção de outros nomes utilizando o teclado, não é possível a ação de um usuário de forma a selecionar um elemento gráfico na tela, como um clique ou o posicionamento do mouse sobre um nome ou publicação. Assim, não se faz viável exibir dicas de contexto ou outras janelas que façam referência a um elemento gráfico específico, como para a exibição de detalhes ou metadados, usos comuns de *tooltips* e *popups*.

Ademais, visto que a proposta da visualização é expositiva, esse tipo de recurso acarretaria na sobreposição de pelo menos uma parte da tela com o conteúdo da *tooltip* ou *popup*. Dessa forma, a compreensão de visitantes no evento em que a visualização foi exposta poderia ser prejudicada caso um dos elementos na tela estivesse oculto por uma mensagem adicional.

4.5 Discussão

Uma vez realizada a caracterização da visualização segundo os aspectos descritos, podemos levantar seus pontos fortes e fracos, além de propor melhorias para suas versões futuras e revisões. Além do entendimento sobre como é o caso estudado, conhecer os aspectos formais nos permite criticá-lo construtivamente a fim de melhorá-lo, numa espécie de análise prescritiva.

Com base na análise feita, é possível perceber uma questão-chave: a visualização foi pensada no contexto de uma exposição em um evento, não para o uso pessoal acessando-a pela internet. Com isso, ela apresenta grande refinamento do formato escolhido e das animações nele empregadas, bem como especial atenção ao conjunto de dados utilizado e na definição de seu público-alvo. Porém, não conta com elementos explicativos, dicas de contexto ou janelas adicionais, nem com aspectos de exploração mais diversos.

Essas características podem desestimular, dificultar - ou até impossibilitar - o uso da visualização fora de eventos, por um usuário desacompanhado. É essencial, para alcançar um número maior de visitantes, despertar neles um maior interesse e garantir uma boa compreensão dos dados, sugerir alterações com enfoque nos aspectos que podem ser mais aprofundados.

Observemos as intenções da mídia, especificamente, no que tange o grau de explorabilidade. Se buscarmos oferecer uma mídia de visualização com um aspecto de exploração do conjunto de dados mais forte, seria bem-vinda a inclusão de filtros, que permitissem ao visitante selecionar diferentes tipos de publicação, como artigos, livros e capítulos de livros, ou anos. Assim, seria possível visualizar recortes dos dados nos quais o usuário pode ter maior interesse, por exemplo, visualizando apenas

artigos publicados a partir de 2015.

Já em relação às características de design, é importante adicionar elementos explicativos adicionais sobre como realizar a leitura da visualização, informando como cada atributo foi codificado, visto que o significado das cores e das repetições de cada nome podem não ser intuitivas para todos os usuários. Outrossim, *tooltips* podem ser utilizadas para informar o ano da publicação mais recente e o número de publicações co-escritas, dados relativos às cores e repetições de cada nome, respectivamente, quando o visitante da página clicar sobre um dos nomes em descida.

Capítulo 5

Avaliação de Viabilidade

5.1 Descrição do estudo

De modo a ratificar os itens levantados e discutidos no capítulo anterior, além de verificar outros aspectos não representados na lista, foi planejada uma atividade junto a um público de estudantes na área de visualização. Tal estudo tem como objetivo aferir, através de respostas a um questionário, a percepção do uso dos aspectos formais estabelecidos por um público familiarizado ao projeto de mídias de visualizações de dados e em um conjunto de mídias mais amplo. Busca-se compreender a completude, inteligibilidade e objetividade dos aspectos formais e a viabilidade de sua utilização por mais pessoas e em mais mídias de visualização.

Os respondentes do questionário desenvolvido são alunos da disciplina “Visualização de Dados” no período de 2019.2. O grupo entrevistado representa uma diversidade de cursos da UFRJ, contando com alunos de graduação de Engenharia de Computação e Informação e de Engenharia Eletrônica e de Computação, oferecidos pela Escola Politécnica, Comunicação Visual e Design, oferecido pela Escola de Belas Artes, e Economia, oferecido pelo Instituto de Economia. Além disso, incluem-se também estudantes do Programa de Pós-Graduação em Artes Visuais e do Programa de Pós-Graduação em Design, ambos oferecidos pela Escola de Belas Artes da UFRJ, e do Programa de Pós-Graduação em Ciência da Informação, oferecido em uma parceria entre o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e a UFRJ.

5.2 Preparação

A consulta foi elaborada no formato de um questionário. Foi utilizada a ferramenta de criação, publicação e acompanhamento de formulários online Google Forms, devido à funcionalidade de organização das perguntas em seções. Além de uma seção inicial para apresentação e identificação do respondente, o formulário contou com mais três seções, uma para cada parte da lista de aspectos formais e uma seção final contendo um campo aberto convidando o estudante a avaliar se a mesma foi capaz de caracterizar bem a visualização de dados que selecionou e, caso contrário, a deixar críticas e sugestões à lista proposta de modo a melhorá-la no futuro. A figura 5.1 mostra uma captura de tela do formulário.

Cada item da lista foi mapeado diretamente para uma pergunta do questionário. Para as perguntas referentes à primeira parte (Intenção da Mídia de Visualização) cada pergunta admitiu respostas numa escala de 1 a 5, de modo a considerar o fato de cada mídia de visualização existir num espectro de familiaridade, explorabilidade e frequência de visitação. Já as demais questões, da segunda parte da lista, admitiam respostas objetivas. Vide as tabelas 5.1, 5.2, 5.3 e 5.4.

Foram adicionadas páginas contendo a lista de aspectos, suas definições e seus exemplos ao website do LabVis [52]. Dessa forma, os alunos foram capazes de consultar mais informações sobre cada item durante a realização da atividade (Figura 5.2).

5.3 Primeira Atividade

Neste primeiro momento, os alunos não foram apresentados aos critérios e suas definições anteriormente. Dessa forma, o primeiro contato com os conceitos utilizados na atividade ocorreu através da leitura da página de referência disponibilizada. O objetivo desta primeira etapa foi capturar a percepção dos participantes frente aos aspectos formais.

Cada respondente foi convidado a realizar a atividade em casa, com um prazo máximo de uma semana. Primeiramente, o aluno foi orientado a visitar a página que sintetizava a lista de aspectos e seus exemplos, como um resumo para recapitular a discussão realizada em aula e servir de consulta rápida durante a atividade (Figura 5.2). Em seguida, ele deve selecionar uma visualização de dados de sua escolha, desde que não tenha sido citada como um exemplo anteriormente no documento de suporte. Tal visualização seria utilizada para responder todas as questões do subsequente formulário.

5.4 Segunda Atividade

Neste segundo momento, a lista de aspectos formais aqui elaborada e descrita foi apresentada e discutida com estudantes em sala de aula. Dessa forma, o público já se encontraria familiarizado com os conceitos abordados no momento da realização da atividade. Foram também apresentados exemplos de cada item da lista, de forma similar ao exposto no Capítulo 3.

Diferentemente da primeira atividade realizada, os alunos foram convidados a responder às mesmas perguntas do questionário tendo como base visualizações de dados selecionadas por colegas de classe na atividade anterior. Cada respondente recebeu duas visualizações distintas da que analisou anteriormente. A distribuição de mídias de visualização entre os alunos se deu de modo que, no total após as duas atividades, cada visualização fosse analisada ao menos três vezes por pessoas diferentes.

Assim como na primeira atividade, o prazo dado aos respondentes para submissão das respostas foi de uma semana, contado a partir da data da discussão em sala de aula.

Tabela 5.1: Questões e tipos de resposta do formulário aplicado: Seção 1
Seção 1 (Introdução)

Enunciado	Resposta
“Qual é o seu nome?”	Campo aberto
“Assinale a opção no qual se insere o seu curso atual.”	Lista de cursos de graduação e programas de pós-graduação
Endereço da visualização de dados escolhida	Campo aberto

Tabela 5.2: Questões e tipos de resposta do formulário aplicado: Seção 2
Seção 2 (Intenção da mídia de visualização)

Enunciado	Resposta
“Considerando o tema abordado pela visualização escolhida, a que tipo de público ela é direcionada?”	Escala de 1 (Menos especializado em relação ao tema abordado) a 5 (Mais especializado em relação ao tema abordado)
“Quanto à facilidade de uso, o quão familiarizado com interfaces de visualizações de dados o público-alvo deve estar?”	Escala de 1 (Nada familiarizado com o uso de visualizações) a 5 (Muito familiarizado com o uso de visualizações)
“Qual é o grau de explorabilidade da interface?”	Escala de 1 (Voltada à explanação) a 5 (Voltada à exploração)
“Com que frequência espera-se que um visitante acesse a página?”	Escala de 1 (Visitação única) a 5 (Visitação recorrente)

Tabela 5.3: Questões e tipos de resposta do formulário aplicado: Seção 3
Seção 3 (Aspectos de design)

“A forma de interação da visualização sugerida se aplica de maneira?”	Convencional x Manipulação direta
“Há alguma forma de <i>feedforward</i> ?”	Sim x Não
“A visualização se dá em formato de <i>dashboard</i> ?”	Sim x Não
“Possui mais de um layout para o mesmo conjunto de dados?”	Sim x Não
“Se há mais de um layout, os dados apresentam continuidade temporal?”	Sim x Não x Minha visualização não tem mais de um layout
“As animações presentes são classificadas como: (Marque todos que se apliquem.)”	Ausentes, Iniciais, Persistentes, Em resposta a uma interação
“Podemos classificar as explicações na interface como: (Marque todos que se apliquem.)”	Ausentes, Sob demanda, Persistentes
“Caso não sejam ausentes, os elementos explicativos buscam explicar: (Marque todos que se apliquem.)”	Forma de exibição/leitura, O conjunto de dados, Modo de uso/Formas de interação, Não há elementos explicativos (ausentes)
“Os dados se integram a alguma narrativa?”	Sim x Não
“De que maneira elementos como <i>tooltips</i> e <i>popups</i> são usados na visualização? (Marque todos que se apliquem.)”	Ausentes, Detalhes do dado, Dicas de interação, Metadados

Tabela 5.4: Questões e tipos de resposta do formulário aplicado: Seção 4
Seção 4 (Conclusão)

<p>“Os aspectos listados permitem caracterizar bem os diversos tipos de mídia de visualização que você encontra na web? Caso contrário, que aspectos adicionais você incluiria nessa lista?”</p>	<p>Campo aberto</p>
--	---------------------

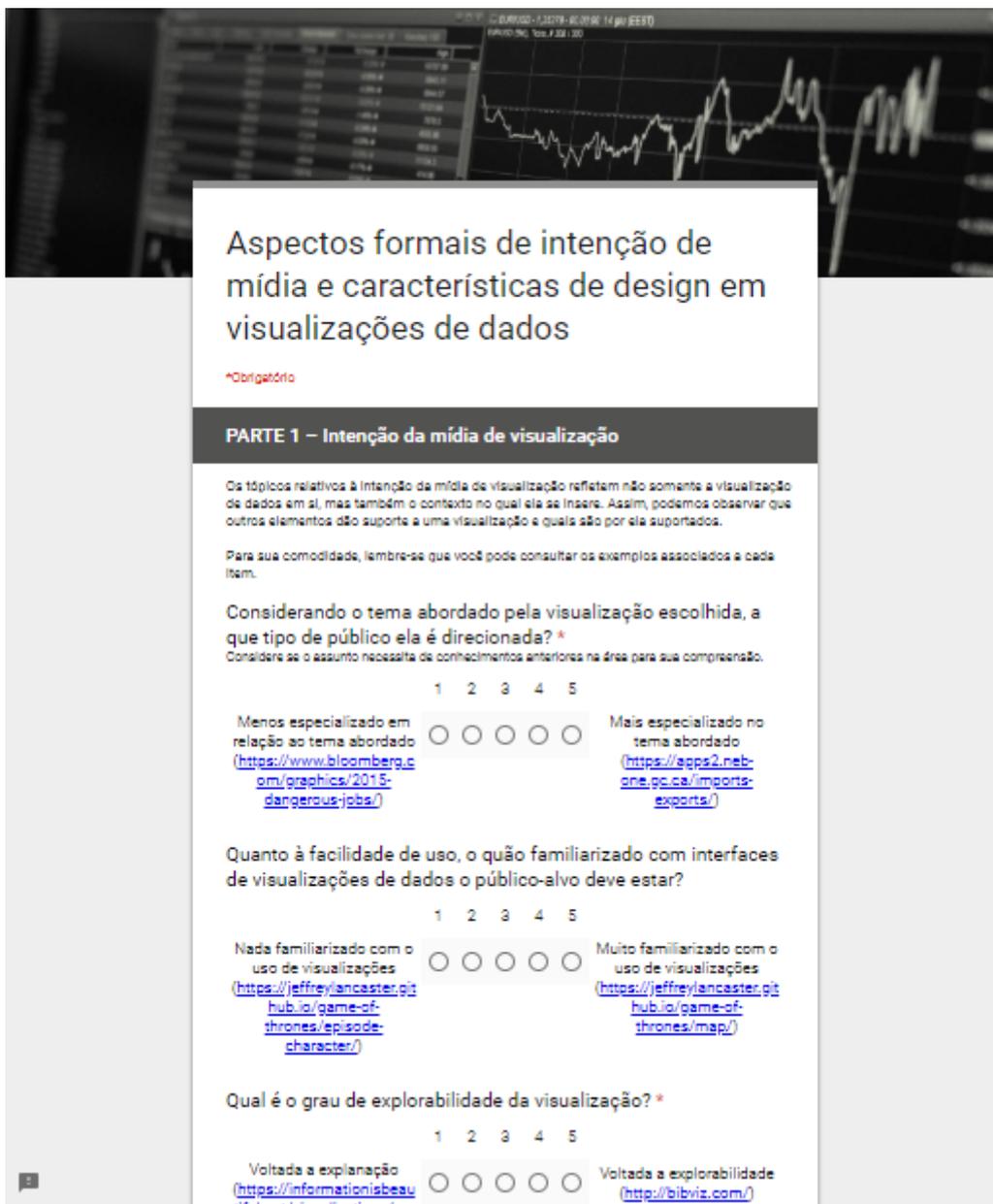


Figura 5.1: Captura de tela da Seção 2 formulário aplicado.

Aspectos formais de intenção de mídia e características de design em visualizações de dados – Parte 1

Blog - 23 de set de 2019



Páginas web que divulgam dados optam, cada vez mais frequentemente, por exibi-los de modo visual. Para isso, incorporam visualizações de dados junto a textos, imagens e outros formatos em um mesmo tecido semântico. Os diferentes elementos que compõem a página, assim como as interações estabelecidas entre eles, precisam ser melhor compreendidos de forma a garantir a transmissão do conteúdo.

O presente trabalho visa desenvolver um estudo preliminar de principais eixos que organizam essas mídias, explicitando suas definições e seus fundamentos além de prover uma coleção de exemplos de páginas web disponíveis na Internet para ilustrá-los. Assim, espera-se obter um guia de fácil acesso aos principais conceitos de mídias de visualização de dados para a web.

Intenção de mídia de visualização

1. Familiaridade (com os dados / com visualizações)

O público-alvo de uma visualização de dados consiste no conjunto de indivíduos que se busca alcançar com a publicação da visualização. A diversidade de *backgrounds* (sociais, culturais) entre diferentes públicos impacta diretamente no sucesso de uma visualização devido a fatores como:

a. Familiaridade com os dados apresentados

Diferentes usuários da visualização podem apresentar diferentes graus de conhecimento prévio sobre a temática abordada e os próprios dados exibidos. Públicos mais versados no assunto desenvolvido podem apresentar maior facilidade em entender uma visualização e obter conclusões com base nela. Por outro lado, para públicos que tiveram menor contato anterior com o mesmo assunto, textos adicionais de ajuda na interpretação dos dados são bem-vindos.



The Deadliest Jobs in America

The U.S. Department of Labor tracks how many people die at work, and who. The latest numbers were released in April and cover the last seven years through 2013. Some of the results may surprise you.

Figura 5.2: Página de referência no website do LabVis.

5.5 Discussão dos Resultados

Na primeira atividade, foram recebidas respostas de 45 alunos, analisando 42 mídias de visualização de dados distintas. Em relação à última questão, sobre opiniões em relação à lista de aspectos formais desenvolvida, todas as respostas foram positivas.

Apesar de não terem sido levantadas sugestões para remover qualquer aspecto da lista, alguns alunos propuseram outros aspectos que poderiam ser incluídos na lista proposta. Um deles foi a ocupação do espaço de tela pela visualização, que reflete não só as dimensões e proporções da visualização como também o modo como ela se posiciona dentro da mídia e divide espaço com outros elementos, como textos.

Outro aspecto proposto pelos alunos foi a dimensionalidade dos dados. Essa característica aborda quantas variáveis são representadas no conjunto de dados e precisam ser codificadas visualmente na mídia.

Sugestões sobre aspectos já incluídos e sobre o questionário também foram trazidas. Por exemplo, discutiu-se que o item “Integração com a narrativa” (3.4.8) poderia ser expresso através de uma escala, assim como os itens relativos a intenção da mídia, em lugar de unicamente opções “Sim” e “Não”. Foi mencionada também a importância de clarificar como são classificados casos como os de visualizações que apresentam animações dentro de *tooltips* ou a rolagem de página como uma forma de interação.

Em relação à segunda atividade conduzida, foram recebidas mais 82 respostas, referentes às mídias de visualização trazidas pelos alunos na etapa anterior. As respostas foram combinadas às respostas da primeira atividade e analisadas utilizando a plataforma gratuita *KNIME Analytics Platform* [53], uma iniciativa de software livre sediada na Suíça, em sua versão 4.0.2.

Após a concatenação dos dois conjuntos de respostas, organizados em tabelas, e filtragem de modo a remover colunas referentes à identificação do respondente, foram realizados passos de pré-processamento dos dados. Cada questão foi codificada em variáveis numéricas e, no caso de questões referentes a características de design, posteriormente vetorizadas em variáveis binárias. A sequência de operações realizadas sobre as respostas é representada pelo fluxo de trabalho mostrado na Figura 5.3.

Também foi necessária uma etapa adicional para perguntas que, no formulário, admitiam a marcação de múltiplas opções como resposta, por exemplo de modo categorizar múltiplas animações ou elementos explicativos adicionais. O passo nomeado *Unpack* inclui um fluxo de trabalho secundário que expande cada coluna relativa a tais perguntas em múltiplas colunas, uma para cada possível opção de resposta, condizendo com semântica original da questão.

O conjunto total de respostas foi obtido pela leitura de arquivos *.tsv* (*tab-separated values*) pela Internet de planilhas atualizadas pela própria plataforma de criação de formulários utilizada. A tabela processada teve dimensões de 127 linhas, uma para cada resposta registrada, e 48 colunas, uma para cada atributo após as etapas de codificação.

Com isso, tornou-se possível efetuar análises de modo a mensurar duas informações relevantes para ponderar a aplicabilidade dos aspectos formais propostos: a distância entre respostas referentes a uma mesma mídia de visualização, de modo a medir o quão consistente é a aplicação dos atributos formais listados; e as correlações entre diferentes questões, que possibilitam perceber como diferentes aspectos no desenvolvimento de visualizações de dados podem se relacionar. A seguir, são descritos os resultados obtidos em cada abordagem.

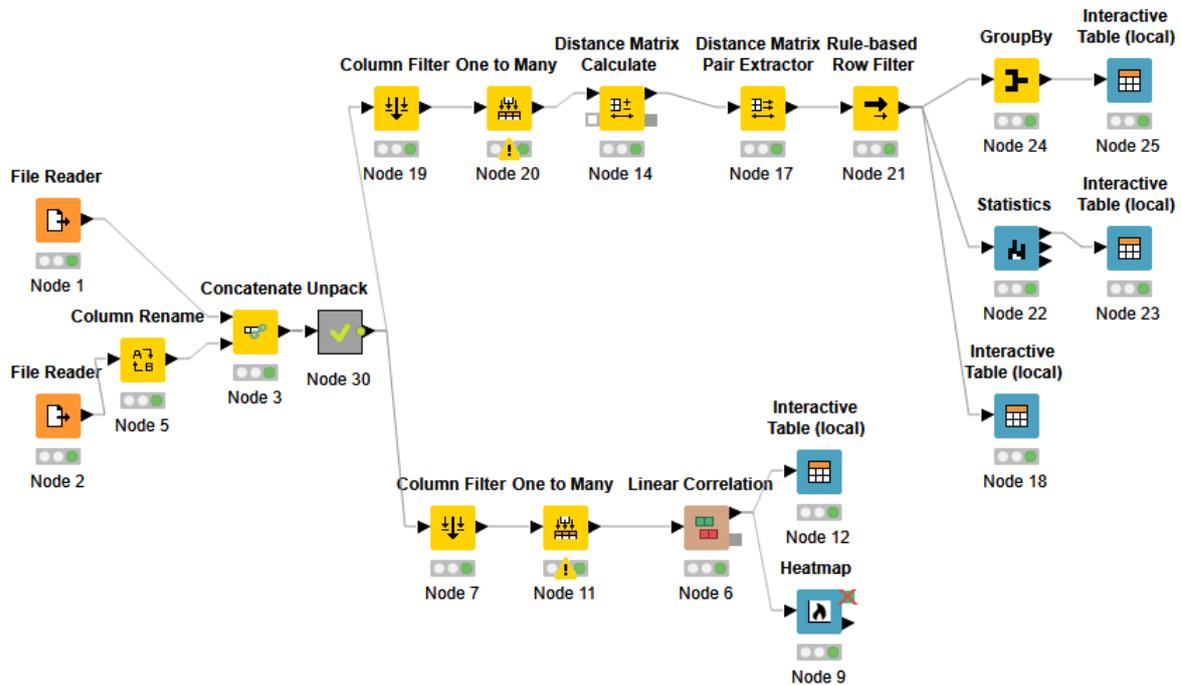


Figura 5.3: Fluxo de trabalho de processamento e análise dos dados no KNIME.

5.5.1 Aferindo Concordância: Distância média

Após o tratamento inicial dos dados, a matriz de distâncias foi computada. Em seguida, apenas as distâncias entre respostas para a mesma visualização foram selecionadas através de uma nova filtragem. Os resultados foram agrupados por visualização com suas médias calculadas. Foi utilizada a distância euclidiana para o cálculo das distâncias entre respostas.

A distância média encontrada entre respostas relativas a uma mesma mídia de visualização foi de $4,16 \pm 0,96$ de uma distância máxima possível de 60. Considerando apenas a primeira parte dos aspectos formais, quanto à intenção da mídia de visualização, a distância média obtida entre respostas de uma mesma visualização foi de aproximadamente $2,40 \pm 1,00$ de 15 de distância possível. Já apenas entre questões relativas à configuração de design, foi encontrada uma distância média de cerca $3,57 \pm 0,83$ de 45 pontos possíveis de distância.

Tendo em vista que a distância média de 4,16 representa menos de 7% da discrepância total possível entre respostas, vemos que os aspectos formais propostos podem ser aplicados por diferentes pessoas, inclusive de áreas de formação distintas, e apresentar resultados consistentes entre si. A baixa divergência entre análises de uma mesma mídia de visualização indica que os aspectos formais são atrelados ao caso analisado e menos suscetível a diferenças pessoais.

Pode-se notar, também, uma concordância maior entre respostas na segunda parte dos aspectos formais, que abordam configurações de design, possivelmente por consistirem de verificações mais objetivas e pragmáticas, por serem diretamente observáveis, do que, por exemplo, questões sobre a familiaridade do público-alvo, como as presentes na seção sobre a intenção da mídia.

Devemos considerar, no entanto, que as questões referentes à intenção da mídia de visualização, que utilizam escalas intervalares de 1 a 5, podem ter seus valores intermediários compreendidos de maneira diferente por cada pessoa. Por exemplo, o valor central (3) da escala pode ter sido interpretado por alguns respondentes como o meio-termo entre os dois extremos do espectro e por outros como “Não sei”. Tendo em vista a ocorrência menos frequente desse valor em comparação com os demais, podemos realizar alguns ajustes em nossa análise.

De modo a minimizar o impacto desses casos na validade dos resultados, uma nova análise foi realizada descartando os valores centrais de todas essas questões e combinando as quatro outras alternativas em apenas duas respostas, uma referente a cada extremo da escala. Por exemplo, os valores 1 e 2 da questão pertinente ao grau de explorabilidade da mídia de visualização foram agrupados em uma só resposta “Voltada à explanação” enquanto os valores 4 e 5 foram agrupados em uma resposta “Voltada à exploração” e o valor 3 foi desconsiderado. Passando a utilizar essa metodologia, a distância média encontrada entre respostas relativas a uma mesma mídia de visualização foi de $3,79 \pm 0,97$ de 56, também representando aproximadamente 7% da distância total possível compatível de maneira com o observado na análise anterior.

5.5.2 Aferindo Relações: Matriz de Correlação

Ao computar as correlações de Pearson entre cada atributo fornecido pelas respostas, foi possível verificar a existência de algumas correlações moderadas (de módulo igual ou superior a 0,5 e inferior a 0,7) e fortes (de módulo igual ou superior a 0,7). Foram elas:

1. **Animação \neq ausente \Leftrightarrow Animação em resposta a uma interação**

Correlação forte: 0,80

Visualizações de dados que apresentam alguma animação costumam apresentar alguma animação em resposta a uma interação do visitante.

2. ***Tooltips/popups* \neq ausentes \Leftrightarrow *Tooltips/popups* com detalhes dos dados**

Correlação forte: 0,77

Visualizações de dados que apresentam alguma *tooltip* ou *popup* usam esse recurso para exibir detalhes dos dados.

3. **Elemento explicativo adicional \neq ausente \Leftrightarrow Elemento explicativo adicional sobre o conjunto de dados**

Correlação moderada: 0,60

Visualizações de dados que apresentam algum elemento explicativo adicional costumam apresentar algum elemento explicativo adicional a respeito do conjunto de dados.

4. **Grau de explorabilidade elevado \Leftrightarrow Frequência de visitação elevada**

Correlação moderada: 0,57

Visualizações de dados que aparentam ter intenção de serem visitadas com maior frequência por seus usuários tendem a oferecer um grau de exploração maior.

5. **Presença de mais de um *layout* \Leftrightarrow Ausência de continuidade temporal**

Correlação moderada: 0,57

Visualizações de dados com mais de um *layout* não costumam apresentar continuidade temporal entre eles.

6. **Grau de explorabilidade elevado** \Leftrightarrow **Forma de manipulação direta**

Correlação moderada: 0,55

Visualizações de dados que visam oferecer uma maior exploração do visitante utilizam mais frequentemente a manipulação direta como forma principal de manipulação.

7. **Elemento explicativo adicional de uso/interação** \Leftrightarrow ***Tooltips/popups***
com dicas de interação

Correlação moderada 0,52

Visualizações de dados que apresentam elementos explicativos adicionais sobre uso/interação costumam também apresentar *tooltips* ou *popups* com dicas de interação.

8. **Público familiar com conjunto de dados** \Leftrightarrow **Público familiar com formato/interface da visualização**

Correlação moderada: 0,52

Mídias de visualização de dados projetadas para um público-alvo mais familiar com o conjunto de dados representado costumam também apresentar formatos de visualização e interfaces mais familiares.

As correlações encontradas corroboram alguns dos conhecimentos tácitos já bem estabelecidos, como entre alto grau de explorabilidade e manipulação direta. No entanto, outras correlações esperadas como aquela entre alta familiaridade do público e uso de *dashboards*, por sua vez, não foram encontradas nesse momento.

Esse fenômeno pode ter sua causa no número relativamente pequeno de mídias de visualização de dados considerado. Para resultados mais acurados, confirmando as correlações encontradas e/ou possibilitando a detecção de outras, bem como obtendo valores mais precisos de distância média, um terceiro teste com um número maior de alunos e de mídias de visualização é imprescindível.

Capítulo 6

Conclusões

Tendo em vista os resultados analisados no capítulo anterior, é possível concluir que a lista de aspectos formais descrita neste trabalho é capaz de caracterizar, com sucesso, os diferentes elementos relevantes de publicação e design de mídias de visualização de dados.

A lista foi testada junto a estudantes universitários de visualização de dados, tendo sido bem recebida junto a eles: um grupo composto por alunos de Engenharia, Design, Artes Visuais e Economia, em níveis de graduação e pós-graduação. Assim, foi possível estabelecer um conjunto de parâmetros que pode ser facilmente compreendido por profissionais de diferentes áreas, um fator importante especialmente em um campo interdisciplinar como o de Visualização de Dados.

Uma variedade de projetos futuros podem ser elaborados em continuidade a este trabalho. Aqui, foi investigado especificamente o uso de uma lista de aspectos formais de intenção e design na avaliação de mídias de visualização de dados já existentes. No entanto, tal lista pode também servir para guiar a criação de novas mídias de visualização. Assim, é bem-vinda a realização de novos testes para perceber a aceitação desses aspectos por parte de engenheiros e designers no contexto do desenvolvimento de novos projetos de visualização. Possivelmente, os estudos conduzidos podem fomentar e contribuir para a elaboração de *checklists* e outras ferramentas de avaliação a serem incorporadas a processos de desenvolvimento desse tipo de software.

Além disso, é especialmente valioso buscar compreender quais são os principais desafios na comunicação entre engenheiros e designers no projeto de uma nova mídia de visualização. Isso é, devemos explicitar causas e propor soluções para empecilhos que ocorrem na interface entre a concepção visual e a implementação de novas visualizações. Os atributos formais seriam um passo inicial em construir uma comunicação mais eficiente entre profissionais de diferentes áreas num projeto de visualização.

Outro possível desdobramento deste trabalho é a criação de um sistema de análise prescritiva para a elaboração de novas mídias de visualização de dados. Aplicando técnicas de aprendizado de máquina sobre um número suficientemente grande de mídias de visualização caracterizadas segundo uma lista de aspectos, seria viável uma aplicação que, com base em algumas variáveis de entrada já conhecidas pelo engenheiro ou designer, aconselhe sobre as demais decisões de projeto.

Uma abordagem similar pode ser realizada para o desenvolvimento de um sistema de recomendação. A lista de aspectos formais (ou uma variação dela) pode ser utilizada para caracterizar um grande conjunto de mídias de visualização. Nesse sistema online, o visitante, com base em uma mídia de seu interesse, poderia encontrar mídias também acessadas por outros usuários com interesses próximos ou mídias similares às que já acessou.

Finalmente, é importante sugerir também a adaptação dos aspectos formais para mídias de visualização sob a ótica dos dispositivos móveis. Outras características relevantes para esse tipo de dispositivo, como responsividade, suporte a interações por toque e uso de recursos como o giroscópio para rotação da tela, enriquecem a diversidade do espaço de mídias de visualizações, com experiências de uso que se distinguem das contempladas por este trabalho e que necessitam de estudos subsequentes.

Referências Bibliográficas

- [1] Raconteur. A Day in Data. In *Future of Data Special Report. The Sunday Times*, 2019. (Disponível em <https://www.raconteur.net/infographics/a-day-in-data>) (Acesso em 27 Setembro 2019).
- [2] Colin Ware. *Information visualization: perception for design*. Elsevier, 2012. (Acesso em 11 Dezembro 2019).
- [3] Domo Inc. Data Never Sleeps 7.0 Report, 2019. (Acesso em 27 Setembro 2019). URL: <https://www.domo.com/learn/data-never-sleeps-7>.
- [4] Alex Bigelow, Steven Drucker, Danyel Fisher, and Miriah Meyer. Reflections on how designers design with data. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces, AVI '14*, pages 17–24, New York, NY, USA, 2014. ACM. URL: <http://doi.acm.org/10.1145/2598153.2598175>, doi:10.1145/2598153.2598175.
- [5] Jeremy Boy, Louis Eveillard, Françoise Detienne, and Jean-Daniel Fekete. Suggested interactivity: Seeking perceived affordances for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 22:639–648, 2016.
- [6] James J Gibson. *The ecological approach to visual perception: classic edition*. Psychology Press, 1977.
- [7] Joel Norman. Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. *Behavioral and brain sciences*, 25:73–96, 2002.

- [8] Donald A. Norman. *The Design of Everyday Things*. Basic Books, Inc., New York, NY, USA, 2002.
- [9] Catherine Plaisant. The challenge of information visualization evaluation. In *Proceedings of the working conference on Advanced visual interfaces*. ACM, pages 109–116, 2004.
- [10] Chris North. Toward measuring visualization insight. *IEEE computer graphics and applications*, 26(3):6–9, 2006.
- [11] Purvi Saraiya, Chris North, Vy Lam, and Karen A Duca. An insight-based longitudinal study of visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1511–1522, 2006.
- [12] Peter R Keller, Mary M Keller, Scott Markel, A John Mallinckrodt, and Susan McKay. Visual cues: practical data visualization. *Computers in Physics*, 8(3):297–298, 1994.
- [13] Ingo Wassink, Olga Kulyk, Betsy van Dijk, Gerrit van der Veer, and Paul van der Vet. *Applying a User-centered Approach to Interactive Visualisation Design*, pages 175–199. Springer London, London, 2009. URL: https://doi.org/10.1007/978-1-84800-269-2_8, doi:10.1007/978-1-84800-269-2_8.
- [14] Elena Zudilova-Seinstra, Tony Adriaansen, and Robert van Liere. *Trends in Interactive Visualization: State-of-the-Art Survey*. Springer Publishing Company, Incorporated, 2014.
- [15] Edward Segel and Jeffrey Heer. Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics*, 16(6):1139–1148, 2010.
- [16] Min Lu, Siming Chen, Chufan Lai, Lijing Lin, and Xiaoru Yuan. Frontier of information visualization and visual analytics in 2016. *Journal of Visualization*, 20(4):667–686, 2017.
- [17] Severino Ribecca. The data visualization catalogue, 2017. (Acesso em 01 Outubro 2019). URL: <https://datavizcatalogue.com/>.

- [18] Yan Holtz and Conor Healy. From data to viz, 2018. (Acesso em 14 Outubro 2019). URL: <https://www.data-to-viz.com/>.
- [19] Canada Energy Regulator, Government of Canada. Imports and exports of energy products to and from canada, 2019. (Acessado em 17 Setembro 2019). URL: <https://apps2.neb-one.gc.ca/imports-exports/>.
- [20] Christopher Cannon, Alex McIntyre, and Adam Pearce. The deadliest jobs in america, 2015. (Acesso em 17 Setembro 2019). URL: <https://www.bloomberg.com/graphics/2015-dangerous-jobs/>.
- [21] Jeffrey Lancaster. Game of Thrones: The Land of Ice and Fire - How long are characters in each episode?, 2019. (Acesso em 26 Setembro 2019). URL: <https://jeffreylancaster.github.io/game-of-thrones/episode-character/>.
- [22] Jeffrey Lancaster. Game of Thrones: The Land of Ice and Fire - Narrative Chart, 2019. (Acesso em 26 Setembro 2019). URL: <https://jeffreylancaster.github.io/game-of-thrones/map/>.
- [23] M. Adil Yalçın, Niklas Elmqvist, and Benjamin B. Bederson. Cognitive stages in visual data exploration. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, BELIV '16, pages 86–95, New York, NY, USA, 2016. ACM. URL: <http://doi.acm.org/10.1145/2993901.2993902>, doi:10.1145/2993901.2993902.
- [24] Alberto Cairo. *The Functional Art: An introduction to information graphics and visualization*. Voices That Matter. Pearson Education, 2012.
- [25] Stefanie Posavec. Writing without words, 2012. (Acesso em 11 Dezembro 2019). URL: <http://www.stefanieposavec.com/writing-without-words>.
- [26] Jack Kerouac. *On the Road*. Penguin Classics, 2011. (Reimpressão).
- [27] Jeffrey Heer, Frank Ham, Sheelagh Carpendale, Chris Weaver, and Petra Isenberg. Creation and collaboration: Engaging new audiences for information visualization. In Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and Chris North, editors, *Information Visualization*, pages 92–133.

- Springer-Verlag, Berlin, Heidelberg, 2008. URL: http://dx.doi.org/10.1007/978-3-540-70956-5_5, doi:10.1007/978-3-540-70956-5_5.
- [28] Daniel G. Taylor. BibViz Project - Bible Contradictions, Misogyny, Violence, Inaccuracies interactively visualized, 2013. (Acesso em 26 Setembro 2019). URL: <http://bibviz.com/>.
- [29] David McCandless. World's Best Hangover Cure?, 2019. (Acesso em 26 Setembro 2019). URL: <https://informationisbeautiful.net/visualizations/worlds-best-hangover-cure/>.
- [30] INPE – Instituto Nacional de Pesquisas Espaciais. Situação atual - programa queimadas, 2012. (Acesso em 17 Setembro 2019). URL: <http://queimadas.dgi.inpe.br/queimadas/portal-static/situacao-atual/>.
- [31] Center for International Development, Harvard University. The atlas of economic complexity, 2017. (Acesso em 17 Setembro 2019). URL: <http://atlas.cid.harvard.edu/>.
- [32] Collective Learning Group, MIT Media Lab. Oec: The observatory of economic complexity, 2012. (Acesso em 17 Setembro 2019). URL: <https://oec.world/>.
- [33] Edwin L Hutchins, James D Hollan, and Donald A Norman. Direct manipulation interfaces. *Human-computer interaction*, 1(4):311–338, 1985.
- [34] Smithsonian Magazine ESRI Environmental Systems Research Institute and David Rumsey Map Collection. Spyglass on the past: New york city 1836 and today, 2013. (Acesso em 08 Novembro 2019). URL: <http://storymaps.esri.com/stories/2013/then-and-now/?appid=97ae55e015774b7ea89fd0a52ca551c2&webmap=60a6d8d2ab754250918e7fc21d7ac999>.
- [35] Joseph Plambeck, Jacky Myint, and Marc Lavalee. How will facebook's stock fare?, 2012. (Acesso em 30 Setembro 2019). URL: <http://archive.nytimes.com/www.nytimes.com/interactive/2012/05/17/business/how-will-facebook-stock-fare.html>.

- [36] Perisopic. U.S. Gun Deaths in 2010, 2010. (Acessado em 17 Setembro 2019). URL: <https://guns.perisopic.com/?year=2010>.
- [37] Perisopic. U.S. Gun Deaths in 2013, 2013. (Acessado em 17 Setembro 2019). URL: <https://guns.perisopic.com/?year=2013>.
- [38] InetSoft Technology Corp. Census Data Visualization, 2019. (Acesso em 26 Setembro 2019). URL: https://www.inetsoft.com/evaluate/bi-visualization_gallery/dashboard.jsp?dbIdx=6.
- [39] Frontwise. GOV—DNA, 2019. (Acessado em 17 Setembro 2019). URL: <https://govdna.frontwise.com/>.
- [40] Federica Fragapane. The most violent cities in the world, 2018. (Acesso em 07 Novembro 2019). URL: <https://www.behance.net/gallery/70033395/The-Most-Violent-Cities>.
- [41] The Nature Conservancy. Migrations in motion, 2016. (Acesso em 07 Novembro 2019). URL: <http://maps.tnc.org/migrations-in-motion/>.
- [42] Pitch Interactive. Out of Sight, Out of Mind: A visualization of drone strikes in Pakistan since 2004, 2015. (Acesso em 26 Setembro 2019). URL: <http://drones.pitchinteractive.com/>.
- [43] Nathan Yau. A Day in the Life of Americans, 2015. (Acesso em 26 Setembro 2019). URL: <https://flowingdata.com/2015/12/15/a-day-in-the-life-of-americans/>.
- [44] Cameron Beccario. earth :: a global map of winds, weather and ocean conditions, 2019. (Acesso em 26 Setembro 2019). URL: <https://earth.nullschool.net/>.
- [45] Citrix Podio. The Daily Routines of Famous Creative People, 2019. (Acesso em 26 Setembro 2019). URL: <https://podio.com/site/creative-routines>.
- [46] Rebecca Barter. What do the happiest countries in the world have in common?, 2019. (Acesso em 26 Setembro 2019). URL: <http://infobeautiful-wdvp.s3.amazonaws.com/20190109095235/index.html>.

- [47] Jan Diehm and Amber Thomas. Women’s Pockets are Inferior, 2018. (Acesso em 26 Setembro 2019). URL: <https://pudding.cool/2018/08/pockets/>.
- [48] DigitalThoughtFacility. Selfiecity, 2014. (Acesso em 26 Setembro 2019). URL: <http://selfiecity.net/>.
- [49] Future Everything and Barcelona Supercomputing Center. Project Ukko - Visualizing seasonal wind predictions, 2015. (Acesso em 26 Setembro 2019). URL: <http://project-ukko.net/map.html>.
- [50] Doris Kosminsky, Lucas Barcellos Oliveira, and Claudio Esperança. Visualizing visualizers, 2019. (Acesso em 02 Outubro 2019). URL: <https://observablehq.com/@lucasbarcellosoliveira/visualizing-visualizers>.
- [51] University of Trier and Schloss Dagstuhl Leibniz Center for Informatics. Dblp computer science bibliography, 1993. (Acesso em 11 Novembro 2019). URL: <https://dblp.uni-trier.de/>.
- [52] Laboratório da Visualidade e Visualização - UFRJ. Site oficial do Laboratório da Visualidade e Visualização, 2019. (Acesso em 25 Setembro 2019). URL: <https://labvis.eba.ufrj.br/>.
- [53] KNIME AG. Knime analytics platform 4.0.2, 2019. (Acesso em 26 Outubro 2019). URL: <https://www.knime.com/>.