Graduação em Ciência da Computação

Mariama Celi Serafim de Oliveira

UM ESTUDO SOBRE O IMPACTO DA INTERATIVIDADE NA COMPREENSÃO DE VISUALIZAÇÕES POR USUÁRIOS NÃO ESPECIALISTAS

Trabalho de Graduação

Universidade Federal de Pernambuco
secgrad@cin.ufpe.br
www.cin.ufpe.br/~secgrad

RECIFE

2018

Universidade Federal de Pernambuco
Centro de Informática
Graduação em Ciência da Computação

Mariama Celi Serafim de Oliveira

UM ESTUDO SOBRE O IMPACTO DA INTERATIVIDADE NA COMPREENSÃO DE VISUALIZAÇÕES POR USUÁRIOS NÃO ESPECIALISTAS

Trabalho apresentado ao Programa de Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: *Nivan Roberto Ferreira Júnior*

RECIFE
2018

Resumo

A evolução das tecnologias de aquisição e disseminação de informação trouxe um aumento no uso de recursos de visualização de dados para o público em geral. Todavia, a maioria destes recursos foi projetada para os chamados usuários especialistas. Para as grandes massas, que não possuem experiência prévia com análise visual de dados, o processo de compreensão de visualizações muitas vezes não é uma tarefa simples. Entre os vários recursos presentes em visualizações que podem auxiliar neste processo destaca-se a interatividade. Esta é uma característica onipresente em sistemas de visualização modernos e é reconhecidamente vital no processo de exploração visual de dados. Enquanto que o uso de interatividade é amplamente advogado, a eficiência deste recurso no entendimento de visualizações para usuários não especialistas ainda não foi caracterizada. O presente trabalho tem como objetivo contribuir nesta direção por meio de uma avaliação quantitativa e qualitativa do impacto da interatividade no entendimento de visualizações por pessoas não especialistas. Para tanto, realizamos um estudo de usuário com estudantes universitários. Este contou com a participação de cinquenta e duas pessoas no estudo a distância e oito no presencial. Dada uma visualização, os estudantes deveriam responder perguntas referentes a informações contidas nela. Os participantes foram submetidos de forma independente a três situações (tipo 0 visualização estática, tipo 1 visualização estática, com auxílio de tabela e tipo 2 visualização interativa). Segundo os resultados, verificamos que o tipo interativo não apresentou resultado superior no processo de entendimento da visualização em relação aos outros dois. No entanto, observamos que ele executou papel importante da função de refinamento do mapa mental criado sobre a visualização. Além dos aspectos cognitivos, o trabalho ressalta a importância dos princípios de design para que se possa criar visualizações mais efetivas.

Palavras-chave: Visualização de dados, IHC, Interatividade, Usuários não especialistas

Abstract

The evolution of the acquisition and dissemination of information has led to an expansion in the use of data visualization resources for the general public. Nevertheless, most of these features are designed for expert users. For the general public who do not have prior experience with visual data analysis, the process of understanding visualizations is not often a simple task. However, there are now various interactive tools which can assist in the visualization process that have become ubiquitous features in modern visualization systems. These features have proliferated, and are now recognized as vital in the process of visual data exploration. Despite the use of interactivity being widely advocated, the effectiveness of these features toward the understanding of visualizations for novice users has yet to be characterized. The present study aims to contribute toward this direction, conducting a quantitative and qualitative evaluation of the impact of interactivity toward the understanding of visualizations by non-experts. In order to achieve this goal, we conducted a user study with university students. The study involved fifty-two remote participants, and eight participants directly. Given a visualization, students should answer questions in relation to the information contained in it. Participants were submitted independently to three situations (type 0 - static visualization, type 1 - static visualization, with aid of a chart and type 2 - interactive visualization). According to the results, we verified that the interactive type did not present superior results in the students understanding regarding how visualizations encode information. However, we observed that interaction performed an important role in the refinement of the participants visualization mental model. In addition to the cognitive aspects, the study emphasizes the importance of following certain design principles in order to create more effective visualizations.

Keywords: Data visualization, HCI, Interactivity, Novice users

Lista de Figuras

3.1 Tela Experimento: Parte 1 - Tutorial. 18
3.2 Tela Experimento: Parte 2 - Dados da Universidade Federal de Pernambuco (UFPE). 18

4.1 Boxplot dos números questões certas por tipo de experimento. 21
4.2 Gráfico de barras representando o percentual de acerto por questão e tipo de experimento. 22
4.3 Boxplot dos números clicks por tipo de experimento. 23
4.4 Boxplot dos eventos de movimento do mouse registrado por tipo de experimento. 24
4.5 Gráfico que relaciona acertos, grau de certeza, clicks por participante do experimento. Eixo X: grau de certeza, Eixo Y: acertos, tamanho da bolha: número de clicks. 25

Lista de Tabelas

4.1 Cálculos do teste ANOVA e Tukey das amostras do teste a distância. Consideramos significativamente diferente valores onde p<0,05. 21

7.1 Cronograma de atividades. 35
7.2 Perguntas apresentadas ao participante durante o tutorial. 40
7.3 Perguntas apresentadas ao participante durante o experimento. 41
7.4 Perguntas do experimento presencial. 42

Lista de Acrônimos

FC	*Frame of Content*
FVE	*Frame of Visual Encoding*
GCP	Gráfico de Coordenadas Paralelas
IHC	Interação Humano Computador
IIQ	intervalo interquartil
Infovis	Visualização de Informação
UFPE	Universidade Federal de Pernambuco

Sumário

1 Introdução **9**

2 Trabalhos Relacionados **11**

2.1 Visualizações de dados por não especialistas 11

2.2 *Sensemaking* 11

2.3 Interatividade 13

2.4 Interatividade e suas técnicas 13

3 Design do Experimento **15**

3.1 Objetivo 15

3.2 Hipóteses 15

3.3 Metodologia 16

3.3.1 Seleção dos participantes 16

3.3.2 O Experimento 17

3.3.2.1 Definição da visualização, dados e perguntas 19

4 Resultados Experimentais **20**

4.1 Estudo a distância 20

4.1.1 Acertos e tipo de experimento 21

4.1.2 Tempo de experimento 22

4.1.3 Grau de certeza e tipo de experimento 23

4.1.4 Interatividade (mouse *log*) e tipo de experimento 23

4.1.5 Correlação entre acertos, certeza e clicks 24

4.2 Estudo presencial 25

4.2.1 Considerações Gerais 25

4.2.2 Tipo 0 - Visualização sem interatividade e sem auxílio de tabela 26

4.2.3 Tipo 1 - Visualização sem interatividade e com auxílio de tabela 26

4.2.4 Tipo 2 - Visualização com interatividade e sem auxílio de tabela 27

5 Discussão **29**

5.1 Interatividade e FVE 30

5.2 Limitações 31

6 Conclusão **32**

Referências **33**

7 Apêndice **35**

7.1 Cronograma 35

7.2 Aspectos Éticos 35

7.3 Termo de Consentimento Livre e Esclarecido 36

7.4 Convite para experimento - Estudo a distância 38

7.5 Convite para experimento - Estudo presencial 39

7.6 Perguntas do tutorial 39

7.7 Perguntas do experimento 40

7.8 Perguntas do experimento presencial - pós uso do sistema 41

1

Introdução

Representar dados por meio de visualizações é uma maneira bastante eficiente de transmitir informação. Nos últimos anos, com o avanço das tecnologias de aquisição e iniciativas de disseminação de dados, houve crescimento considerável na quantidade do uso de recursos de visualização de dados a fim de difundir informação para o grande público. Um respeitado pioneiro nesta direção é o jornal *The New York Times* que apresenta notáveis visualizações em suas reportagens com o intuito de enriquecer a experiência de leitura. Um outro exemplo é o projeto Subspotting BAUR; GODDEMEYER (2016). O qual consiste em um aplicativo de celular que, por intermédio de gráficos interativos, mapeia o sinal de internet em estações de metrô da cidade de Nova Iorque. Similarmente, ferramentas como Tableau TABLEU (2017), Plotly PLOTLY (2017), Many Eyes VIEGAS et al. (2007) e bibliotecas tal qual D3.js BOSTOCK (2012) popularizaram a criação de visualizações.

No entanto, a grande maioria dessas técnicas e sistemas foi projetada para os chamados especialistas. Estes indivíduos são habituados em sua formação e cotidiano a tratarem e manejarem determinados tipos de representações para explicar e entender os problemas que eles lidam. Para as grandes massas, que não possuem experiência prévia com análise visual de dados, a compreensão de visualizações muitas vezes não é uma tarefa fácil. Um exemplo real deste fato é como a má interpretação de uma visualização, projetada para especialistas da área de meteorologia, gerou uma situação perigosa à população afetada pelo furacão Irma [1]. Este problema de compreensão se dá devido ao fato que à utilização de artefatos cognitivos, como visualizações, não é uma habilidade inata. O processo de compreensão de uma visualização envolve aprendizado. Este pode ser significativamente diferente de uma pessoa com conhecimento prévio para uma leiga LIU; STASKO (2010). Pode-se dizer que este processo é similar a de assimilação do funcionamento de um sistema e conforme Victor VICTOR (2011) a melhor forma de ganhar insights de um sistema seria navegar por suas múltiplas facetas, para isso, ele advoga que a interatividade é um poderoso instrumento que viabiliza tal ação.

A interatividade juntamente com a representação compõem as características básicas de sistemas de Visualização de Informação (Infovis). Sendo a representação aquilo que o usuário

[1] `http://bit.ly/2xW5Icj`

consegue visualizar em um display, enquanto interatividade é o dialogo que o usuário realiza com o sistema YI; KANG; STASKO (2007). A interatividade propicia que a representação possa ser moldada de acordo com as necessidades do usuário durante o ato de investigação da informação BAUR (2017); MUNZNER (2014). Em algumas visualizações como Gráfico de Coordenadas Paralelas (GCP) INSELBERG (1985) e *Treemap* SHNEIDERMAN (1992), a aplicação de interatividade é essencial devido ao volume de informação codificada ou a dificuldade de navegação da visualização.

Ainda que seja reconhecida a importância da interatividade nos sistemas de Infovis LAM (2008), pouca pesquisa foi feita neste âmbito uma vez que, por anos, o estudo da interatividade dentro desta área foi relegado em favor de estudos sobre representação YI; KANG; STASKO (2007). Embora em anos recentes, trabalhos relevantes acerca do assunto tenham sido publicados, a maioria destes foca em estudar técnicas de interação YI; KANG; STASKO (2007); HEER; SHNEIDERMAN (2012); LAM (2008). Poucos se posicionam em relação ao papel da interatividade no processo de compreensão de visualizações. Entre estes poucos trabalhos, pode-se ressaltar Z. Liu e J. T. Stasko LIU; STASKO (2010) defensores da análise da interatividade não só como uma série de técnicas, mas como uma forma de amplificar processos cognitivos e auxiliar a criação e reconstrução de mapas mentais de visualizações. Logo, entende-se que dada a expansão do uso de sistema de Infovis para diversas audiências é de fundamental importância o estudo de seus pilares, representação e interatividade.

O presente trabalho tem como objetivo contribuir nesta direção por meio de uma avaliação quantitativa e qualitativa do impacto da interatividade no entendimento de visualizações por pessoas não especialistas. Para tanto, foram realizados estudos de usuário que visaram medir o impacto da interatividade na acurácia e velocidade de resposta com respeito a questões analíticas. Nestes estudos usaremos o GCP pelo seu amplo estudo teórico e uso em sistemas de visualização.

2

Trabalhos Relacionados

2.1 Visualizações de dados por não especialistas

Com a crescente expansão do uso de recursos de visualizações para transmitir informação para o grande público, estudos começaram a ser publicados neste âmbito. Pousman et al. POUSMAN; STASKO; MATEAS (2007), por exemplo, exploram de que formas esta nova audiência faz uso destes recursos. Segundo estes autores, os recursos de visualizações que este público utiliza possuem aplicações diferentes do que os utilizados pelos especialistas. Com efeito, os recursos e sistemas projetados para os especialistas são geralmente focados em explorar, gerar, refinar e testar hipóteses da informação que eles têm em mãos. Enquanto que para a grande população, que pode ser composta de especialista ou leigos, o uso de visualizações ocorre de forma mais casual como, por exemplo, para visualizar informações pessoais como fotos, e-mails e músicas. Este tipo de uso apresenta algumas diferenças quando comparada com Infovis tradicional, como o tipo de dados e os *insights* que usuários buscam ao fazer uso da visualização. Preocupados com este público, que em sua grande maioria não é familiarizado em interpretar representações visuais, estudos recentes LEE et al. (2016); GRAMMEL; TORY; STOREY (2010) abordam como pessoas não especialista lidam com recursos de Infovis. O *framework NOVIS (NOvices information VIsualization Sensemaking)* LEE et al. (2016) foi proposto por Lee et al. tentando entender melhor como ocorre este processo de entendimento de visualizações por não especialistas. A partir de experimentos, este trabalho identifica as principais atividades cognitivas que ocorrem enquanto um indivíduo tenta compreender uma visualização. A compreensão destas fases é bastante importante pois permite que futuras pesquisas, como o presente trabalho, possam levantar e estudar estratégias que possam ser implementadas para construir um processo mais fluido de entendimento de visualizações.

2.2 *Sensemaking*

Sensemaking é um termo que vem do inglês e bastante utilizado no campo de Interação Humano Computador (IHC) e Infovis. O termo é definido como o processo pelo qual as pessoas

passam ao coletar, organizar e estruturar informação com intuito de entender um problema RUSSELL et al. (1993). Em Infovis este termo é utilizado quando se investiga como ocorre o processo de compreensão de uma visualização ou da informação codificada nela. Portanto, ao se projetar sistemas de Infovis é importante observar como este processo ocorre.

Dentro dos estudos de *Sensemaking*, Moore et al. MOORE; HOFFMAN (2011) propuseram a *data/frame theory*. Nesta teoria eles apresentam o conceito de *frame*, uma metáfora utilizada para explicar como as pessoas compreendem algo. Este *frame* pode ser histórias, mapas, scripts, diagramas etc. Nesta teoria, o *frame* estaria constantemente evoluindo a partir da aquisição de outras informações e exposição deste a situações no qual ele pode ser questionado. Assim sendo, o processo de *Sensemaking* é cíclico. O ciclo envolve a criação do *frame*, seu refinamento e questionamento dado sua interação com novas informações. Este ciclo, que não necessariamente é ordenado, acontece até que se consiga fazer sentido de algo. Fundamentado neste conceito Lee et al. LEE et al. (2016) propuseram o *framework NOVIS*. Lee et al. argumentam que dentro do processo de compreensão da visualização de dados acontece a formação do *frame* da visualização, o qual evoluirá de acordo com a exploração dos dados por parte do usuário utilizando a ferramenta de visualização. O *frame* definido por Lee et al. divide-se em dois, *Frame of Content* (FC) e *Frame of Visual Encoding* (FVE). O primeiro define a maneira como a pessoa entende a informação que está codificada dentro da visualização. Por exemplo, um gráfico que mostra dados sobre países em desenvolvimento, o entendimento das informações destes países define o FC. Já o segundo *frame* FVE define a compreensão de como o gráfico codifica informação. Tamanhos diferentes, formas e cores em visualizações são bons exemplos de como codifica-se informação. Yalçn et al. YALÇIN; ELMQVIST; BEDERSON (2016) também sugerem um *framework* dos estágios cognitivos no processo de exploração de uma visualização. Neste *framework*, os autores focam em ressaltar quais custos e barreiras cognitivas ocorrem durante o processo de exploração de uma visualização. Outra abordagem interessante sobre o entendimento de visualizações é a argumentada por Liu e Stasko LIU; STASKO (2010). Os autores discutem em seu trabalho o inter-relacionamento entre a representação interna e externa de uma visualização. Conforme estes autores, a representação interna poderia ser compreendida como o mapa mental que as pessoas têm de uma visualização. O conceito de mapa mental definido pelos autores relembra o conceito já apresentado anteriormente de *frame* definido por Lee et al. LEE et al. (2016). Especificamente o conceito de *frame* relacionado ao modo como a visualização codifica a informação. O trabalho de Liu e Stasko se mostra bastante relevante devido ao fato de explorar o papel da interatividade dentro da visualização e o processo de *Sensemaking*. A interatividade para os autores possui papel essencial neste processo, posto que ela auxilia a superar os problemas cognitivos que aparecem ao se expandir as barreiras do mapa mental para a representação externa.

2.3 Interatividade

Dentro de IHC o papel da interação é constantemente discutido. Para Dix et al. DIX et al. (2003), interação pode ser considerada a comunicação que ocorre entre o ser humano e o computador. Por ser algo bastante complexo, diversos modelos tentam explicar como este fenômeno ocorre. Entre estes, está o modelo de Norman NORMAN (2013) intitulado como *"execution-evaluation cycle"*. Neste modelo, a interação é descrita da seguinte forma: a pessoa formula um plano, que posteriormente é colocado em prática na interface do computador e depois de executado a pessoa avalia o resultado para então planejar novas ações.

Já dentro de Infovis, a interatividade pode ser considerada uma das bases da visualização YI; KANG; STASKO (2007). Muitos trabalhos advogam que ela é essencial dentro da visualização de dados VICTOR (2011); YI; KANG; STASKO (2007); PIKE et al. (2009); ELMQVIST et al. (2011). Segundo Munzner MUNZNER (2014), a interatividade permite que as capacidades das ferramentas de visualização sejam amplificadas. A autora afirma que dentro da imensa quantidade de informação lidada por estes sistemas é quase impossível para o usuário compreender os dados sem a ajuda da interação. De maneira análoga argumenta Bret VICTOR (2011), defensor da interatividade a fim de que o usuário possa navegar por diferentes facetas de um problema. No entanto, Scaife e Rogers alertam em seu trabalho SCAIFE; ROGERS (1996) que muitos estudos que elogiam as qualidades da interatividade recaem no que eles chamam de *"recemblace fallacy"*. *Recemblace fallacy* acontece quando se pressupõe que a conexão externa e interna de uma representação ocorre de forma intuitiva devido as qualidades e características desta representação. Todavia, não há nenhum embasamento teórico que fundamente tal certeza. Um exemplo disto seria o fato de se afirmar que visualizações interativas são mais eficientes que estáticas pelo fato delas apresentarem funções mais avançadas e simularem comportamentos mais parecidos aos que ocorrem no cérebro de uma pessoa. Para Scaife e Rogers tal afirmação é completamente equivocada pois a eficiência do sistema dependerá da configuração deste. Um estudo que tenta entender não só as vantagens, mas também os custos da interação é o realizado por Lam LAM (2008). Neste trabalho a autora cria um *framework* apoiado no modelo citado anteriormente de Norman NORMAN (2013). Este *framework* tenta compreender em que momentos a interação em uma visualização pode ser falha. Ao final de seu trabalho a autora sugere uma necessidade de identificar os custos cognitivos na interação. Para isso, ela defende o foco em todo o ciclo cognitivo ao invés do foco somente nos problemas relacionados a fase de avaliação do sistema.

2.4 Interatividade e suas técnicas

Como foi descrito na seção anterior (seção 2.3), interatividade é um conceito bastante amplo. O simples fato do usuário focar com os olhos em partes do gráfico pode ser considerado ações de interatividade LIU; STASKO (2010), pois neste processo há comunicação entre a

pessoa e visualização. No entanto, de acordo com Yi et al. YI; KANG; STASKO (2007) técnicas de interação são menos difíceis de serem caracterizadas. Segundo os autores, técnicas de interação são qualidades apresentas pela visualização que permitem o usuário manipular diretamente ou indiretamente representações. Partindo deste conceito, podemos considerar que visualizações estáticas ou animadas não possuem técnicas de interação, ainda que o usuário interaja com elas.

Há diversos trabalhos que realizaram uma taxonomia de técnicas interativasYI; KANG; STASKO (2007). Todavia, estudos mais recentes YI; KANG; STASKO (2007); BREHMER; MUNZNER (2013) defendem a caracterização do objetivo do usuário ao interagir com a visualização, em oposição aos meios utilizados (técnicas de interação). Yi et al.YI; KANG; STASKO (2007) utilizando este tipo de classificação categorizou a interação em sete diferentes tipos: **selecionar**: marcar algo como interessante e conseguir rastrear isso, **explorar**: permitir que o usuário explore diferentes subconjuntos da representação, **reconfigurar**: mudar o arranjo da representação a fim de ver outra perspectiva, **codificar**: mudar o modo como a informação é codificada, **abstrair/elaborar**: mudar os níveis de abstração alternando entre algo abrangente para específico, **filtrar**: mudar a informação apresentada dada algumas condições, **conectar**: realizar conexões entre diferentes variáveis representadas. Esta abordagem é bastante interessante pois ressalta o relacionamento dos objetivos do usuário e os impactos da interatividade.

3

Design do Experimento

3.1 Objetivo

Dado o expressivo crescimento do acesso a sistemas Infovis por múltiplas audiências, o presente trabalho tem por objetivo analisar o impacto da interatividade no entendimento de visualizações por pessoas não especialistas. Dando ênfase na relação da interatividade e a compreensão de como a informação é mapeada em uma visualização.

3.2 Hipóteses

Baseado no objetivo descrito acima, as seguintes hipóteses foram criadas:

H1: Visualizações que utilizam técnicas interativas possuem uma performance melhor quando comparadas as estáticas em relação ao FVE formado, em outras palavras, usuários de visualizações interativas aprendem em menos tempo e de forma mais correta como a informação é codificada dentro da visualização.

H2: O uso de técnicas interativas ajuda usuários a encontrarem anomalias no FVE formado, caso este tenha sido construído de modo incorreto.

H3: Visualizações que possuem técnicas interativas engajam por mais tempo o usuário quando comparadas com visualizações estáticas.

Apesar das hipóteses elencadas fazerem afirmações em relação as visualizações de forma geral, neste trabalho iremos apenas explorar os efeitos iteratividade na visualização GCP. Logo, conclusões alcançadas a partir destas hipóteses se restringirão à visualização escolhida e as técnicas de interatividade presentes nesta.

Para que as hipóteses fossem verificadas elaboramos um experimento. Este contou com a presenças de pessoas e um sistema de visualização. Informações mais detalhadas estão contidas na próxima seção deste capítulo, Metodologia. Na Metodologia, será detalhado as decisões tomadas para o experimento e como este ocorreu.

3.3 Metodologia

Com a intenção de verificar as hipóteses estabelecidas no presente trabalho definimos uma abordagem quantitativa e qualitativa. A primeira abordagem, a distancia, teve um cunho exploratório. Já a segunda, presencial, serviu para que pudéssemos observar os processos cognitivos que ocorrem durante a interação (pessoa/visualização) como a literatura recomenda LIU; STASKO (2010); LAM (2008).

3.3.1 Seleção dos participantes

Os participantes escolhidos para fazerem parte dos experimentos foram estudantes da Universidade Federal de Pernambuco (UFPE). Estes estudantes fazem parte da graduação e pós-graduação de 26 cursos da universidade nas áreas de exatas, saúde e humanas. Sendo eles homens e mulheres de diferentes idades. Além disso, todos os participantes não tiveram contato prévio com o tipo de visualização escolhida para o estudo, no caso, GCP. A escolha desta população deve-se ao fato de constituir uma amostra de pessoas com habilidades e conhecimentos diversificados devido as suas formações distintas.

Por consequência, foram convidados a participar do estudo as pessoas que atenderam aos seguintes critérios de inclusão:

a) Matriculados em algum curso de graduação da Universidade Federal de Pernambuco no semestre 2018.1.

b) Maiores de 18 anos.

c) Não tiveram contato prévio com a visualização GCP.

d) Aceitaram participar da pesquisa mediante a assinatura do Termo de Consentimento Livre Esclarecido (TCLE) (Apêndice A estudo presencial ou Apêndice B estudo a distância).

O recrutamento dos participantes para o experimento foi realizado das seguintes formas:

a) Envio de convites para listas de e-mail direcionado a alunos da UFPE.

b) Envio de convites para redes sociais em comunidades dedicadas a alunos da UFPE.

Neste convite existia uma breve descrição do projeto, seus riscos e benefícios, as etapas da coleta de dados e tempo de duração da coleta conforme o Apêndice 7.3.

3.3.2 O Experimento

Neste experimento indivíduos que nunca tiveram contato com GCP foram submetidos de forma independente a três tipos de situação. As três modalidades são:

- **Tipo 0 - estática sem auxílio de tabela:** A visualização desta categoria pode ser comparada ao que uma pessoa encontraria caso a visualização fosse impressa em uma folha de papel. Esta modalidade não contou com nenhum tipo de interação entre o usuário e a representação. Ao explorar a visualização, os participantes também não contaram com nenhum tipo de auxílio textual além do que o próprio gráfico possui.

- **Tipo 1 - estática com auxílio de tabela:** A representação desta modalidade possui as mesmas características do tipo 0. No entanto, durante o experimento houve a presença de uma tabela que apresentava as mesmas informações contidas no gráfico.

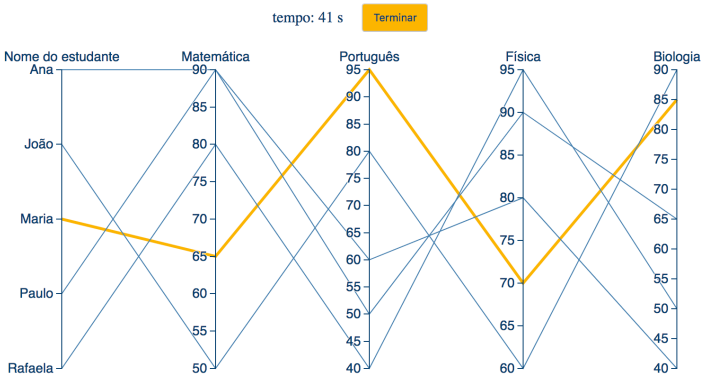
- **Tipo 2 - interativa sem auxílio de tabela:** Este tipo de visualização possuía elementos interativos. Sendo ela igual ao experimento do tipo 1, mas com um gráfico interativo. O gráfico apresentado possuía duas de cinco capacidades interativas descritas por Siirtola et al. SIIRTOLA (2000) em artigo explorando GCP. Estas duas capacidades são rearranjo dos eixos ao realizar a tarefa de arrastar com o mouse e *brushing* ao selecionar partes dos eixos. Estas capacidades possibilitam que objetivos como selecionar, reconfigurar, filtrar e conectar (ver seção 2.4) ocorram de maneira mais fácil durante a exploração de informação.

O experimento ocorreu por meio de um sistema web [1] que conteve as características descritas no decorrer desta seção. Preliminarmente, antes de iniciar o experimento com os dados da UFPE, os participantes tinham oportunidade de interagir com a visualização GCP. Esta etapa que chamamos de tutorial (Figura 3.1), consistiu de um intervalo de no máximo 1 minuto, no qual os participantes tinham acesso a uma visualização de GCP com dados mais simples e uma tabela que possuía a mesma informação da visualização. Após 1 minuto, os participantes responderam 4 perguntas (ver Apêndice 7.6) e receberam o *feedback* da resposta correta. Na segunda parte do experimento (Figura 3.2), os participantes eram apresentados a visualização dos dados da UFPE.

Baseado nesta visualização estes deveriam responder 12 perguntas (ver Apêndice 7.7) e atribuir um valor de 0 a 4 quanto ao grau de certeza ao responder a pergunta; em seguida, o experimento era encerrado.

Cada indivíduo só foi apresentado uma vez a uma das modalidades. Em cada uso do sistema foi coletada para cada questão a resposta do participante, grau de certeza, tempo de

[1] `experimenttcc.firebaseapp.com`



Nome do estudante	Matemática	Português	Física	Biologia
Maria	65	95	70	85
João	50	80	60	90
Paulo	90	50	90	65
Rafaela	80	40	95	50
Ana	90	60	80	40

Figure 3.1 Tela Experimento: Parte 1 - Tutorial.

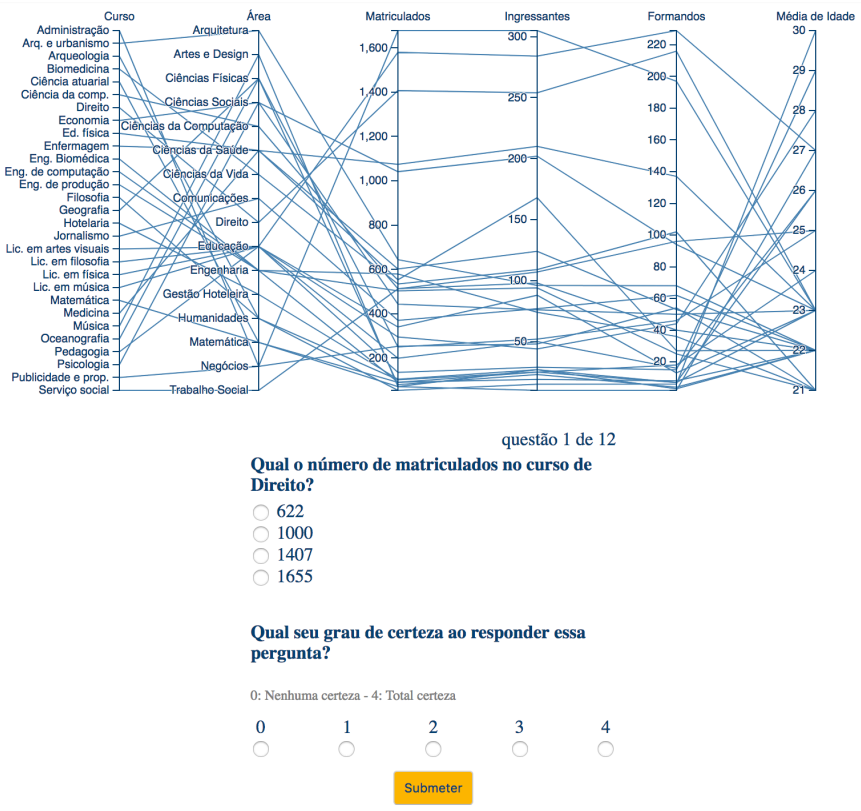


Figure 3.2 Tela Experimento: Parte 2 - Dados da UFPE.

resposta, movimento (evento disparado pela função *onmousemove* [2] de javascript) e click de mouse (coordenada e *timestamp*).

O design descrito acima ocorreu tanto no estudo a distância quanto presencial. En-

[2] `https://developer.mozilla.org/en-US/docs/Web/API/GlobalEventHandlers/onmousemove`

tretanto, o estudo presencial conteve uma etapa de entrevista com perguntas semiestruturadas (ver Apêndice 7.8) com o objetivo de coletar qualitativamente a opinião dos usuários. Também foram observadas as ações do participante dentro do sistema e suas reações. Além disso, foi requerido ao participante que este descrevesse suas ações e intenções dentro do sistema. Este é um método conhecido como "*Think-Aloud*" RUBIN; CHISNELL (2008) que apesar de apresentar interferências cognitivas consegue extrair o processo cognitivo dos participantes LEE et al. (2016).

3.3.2.1 Definição da visualização, dados e perguntas

Com a intenção de conseguir o maior número de pessoas leigas em uma visualização, escolhemos GCP para o estudo. A escolha desta representação seguiu o mesmo raciocínio discutido por Lee et al. LEE et al. (2016), onde os autores similarmente conduzem experimentos com pessoas leigas em visualizações de dados. Eles argumentam que GCP é uma visualização pouco conhecida para a maioria das pessoas visto que ela é pouco ensinada durante a vida acadêmica. Como conjunto de dados a ser utilizado durante o experimento, usamos informações relativas a universidade, como por exemplo, número de estudantes matriculados e taxa de formandos.

Estes dados foram extraídos da plataforma de dados abertos DataViva [3] que possui dados socioeconômicos de diversos setores do Brasil. Esperamos que este tipo de informação diminua possíveis barreiras cognitivas que venham a ocorrer durante o experimento, visto que esta informação é supostamente de interesse e conhecimento dos participantes escolhidos. Quanto às perguntas a serem realizadas durante o experimento (ver Apêndice 7.7) sobre as informações apresentadas, elas foram selecionadas de acordo com as capacidades e atividades que GCP possibilitam. A determinação das características das GCP foi baseada em estudos da literatura sobre esta visualização MUNZNER (2014); SIIRTOLA (2000); HEER; SHNEIDERMAN (2012). Além disso, para a elaboração das perguntas utilizou-se a definição de atividades analíticas elaborada por Amar et al. AMAR; EAGAN; STASKO (2005).

[3] `http://dataviva.info/pt/about/`

4

Resultados Experimentais

O experimento contou com duas modalidades, presencial e a distância. Em ambas modalidades o usuário deveria interagir com um sistema web construído para capturar respostas do usuário baseado na visualização e questões apresentadas. Destas questões foram salvos resposta, tempo, click e movimento do mouse (com sua respectiva coordenada e *timestamp*), além do grau de certeza para as perguntas relativas a UFPE. Antes da condução dos dois tipos de experimento foi realizado um experimento piloto. O piloto contou com a participação de um voluntário. As respostas coletadas no piloto não entraram para os resultados do experimento, elas serviram somente para avaliar o sistema do experimento e as tarefas propostas neste.

4.1 Estudo a distância

O estudo a distância ocorreu entre os dias 28 de maio a 14 de junho de 2018. Ao total 76 pessoas participaram do experimento, sendo aproveitadas 52 respostas, de acordo com os parâmetros estabelecidos no experimento. Aos participantes aleatoriamente foi apresentado um dos três tipos de experimento (ver seção 3.3.2). Sendo 15 pessoas do tipo 0, 18 pessoas do tipo 1 e 19 pessoas do tipo 2, todos eles independentes entre si.

A situação na qual cada participante respondeu às perguntas é desconhecida, visto que cada voluntário acessou a URL do experimento quando lhe foi conveniente. Com efeito, informações referentes a tipo de aparelho utilizado, local e situação de uso são incógnitas. Somente podemos afirmar que a tela do aparelho possuía tamanho mínimo de 750 *pixels* de largura e 500 *pixels* de altura, uma vez que estas eram restrições impostas no sistema para permitir o seu uso.

O experimento a distância possibilitou verificar o comportamento e relacionamento das variáveis de acerto, tempo, grau de certeza, click e movimento. Primeiramente, ao analisar as variáveis realizamos o teste ANOVA e Tukey (tabela 4.1) para determinar a diferença entre os grupos. Depois aumentamos a granularidade e observamos como cada variável se comportou por questão. Por fim, verificamos a correlação entre as variáveis.

	ANOVA		Tukey					
			tipo 0 vs tipo 1		tipo 0 vs tipo 2		tipo 1 e tipo 2	
	F	p	Q	p	Q	p	Q	p
Acertos	5,745	0,006	3,778	0,027	4,546	0,006	0,757	0,84
Tempo	2,490	0,093	-	-	-	-	-	-
Grau de certeza	9,709	0,0003	5,534	0,001	5,441	0,001	0,171	0,9
Click	9,433	0,0003	0,103	0,9	5,064	0,001	5,428	0,002
Movimento	2,193	0,122	-	-	-	-	-	-

Table 4.1 Cálculos do teste ANOVA e Tukey das amostras do teste a distância. Consideramos significativamente diferente valores onde p<0,05.

4.1.1 Acertos e tipo de experimento

No experimento, havia quatro alternativas para cada tarefa, foi considerada correta somente uma alternativa [1].

Os níveis de acerto dos participantes por tipo de experimento foram representados no textitboxplot (Figura 4.1).

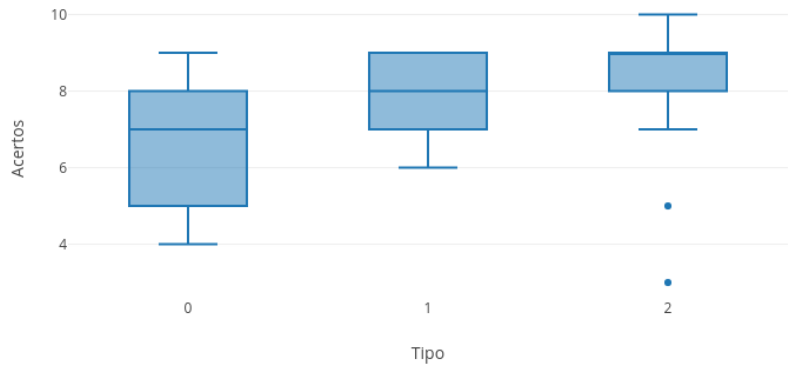


Figure 4.1 Boxplot dos números questões certas por tipo de experimento.

As caixas do gráfico se estendem do primeiro quartil (25%) ao terceiro quartil (75%). Este intervalo denota os valores de acertos ordenados que estão entre 25% e 75% intervalo interquartil (IIQ) das amostras. As linhas do gráfico se estendem caso existam valores dentro do intervalo de 1,5xIIQ da mediana e os círculos são amostras *outliers*.

Como pode ser visto no gráfico, o experimento do tipo 0 (estático, sem tabela) apresentou uma menor performance em relação aos outros dois. O que pode ser pode confirmado pelo teste ANOVA (F = 5,745; p = 0,006; α = 0,05) e teste de Tukey no qual apresentou valores de p = 0,027 em relação ao tipo 1 e p = 0,006 em relação ao tipo 2. Enquanto o experimento de

[1] A questão 8 foi desconsiderada de todos os resultados a distância. Esta questão não possuía resposta correta, o que pode ter causado confusão ao participante.

tipo 1 e tipo 2 não apresentaram diferenças significativas quando comparados entre si (Tukey: p = 0,84).

Quando observado o nível de acertos por questão (Figura 4.2), como está ilustrado no gráfico de barras, percebemos uma diferença entre os tipos. As questões que apresentam significativamente menor grau de acerto (abaixo de 25%) independente do tipo são a 9 e 10. Ambas

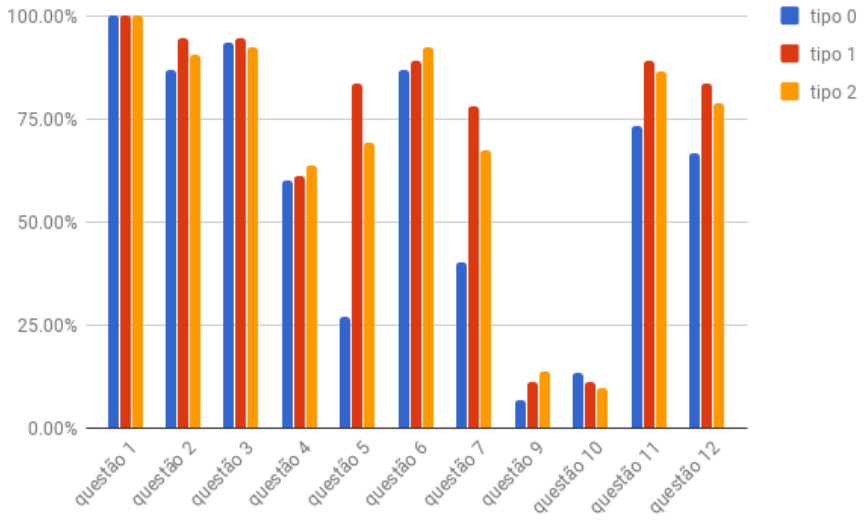


Figure 4.2 Gráfico de barras representando o percentual de acerto por questão e tipo de experimento.

as questões abordam a capacidade do usuário em perceber o significado da inclinação no gráfico.

No caso do experimento tipo 0, como esperado, este apresenta performance inferior em várias questões. Uma vez que a quantidade de dados era extremamente grande e quase impossível de responder sem auxílio de interação ou tabela.

4.1.2 Tempo de experimento

Quando analisado o tempo total para responder às questões da fase de avaliação não encontramos uma diferença significativa entre os 3 tipos de acordo com teste ANOVA (F = 2,490; p = 0,093; $\alpha = 0,05$).

No entanto, ao examinar o tempo de resposta por questão, constatamos que estas diferem razoavelmente entre si. Com a questão 6 apresentando tempo médio geral de 22,12 segundos, enquanto a questão 3 apresenta tempo médio geral de 80,67 segundos. Este fato deve-se provavelmente ao nível de dificuldade da questão, visto que a questão de menor média (questão 6) apresentava resposta de forma direta, não necessitando fazer correlação entre variáveis, contagem de retas ou estudo da inclinação de reta.

4.1.3 Grau de certeza e tipo de experimento

De acordo com o teste ANOVA calculado, existe diferença significativa entre os tipos de experimento (F = 9,709; p = 0,000281; α = 0,05) quanto ao grau de certeza. A diferença foi encontrada em relação ao tipo 0 (estático, sem tabela) e os outros dois tipos. Apresentando o valor de p = 0,001 tanto para o tipo 1 quanto para tipo 2 no teste de Tukey. Enquanto tipo 1 e tipo 2 não possuem diferenças significativas. Confirmado o que já havia sido observado quando analisado o nível de acertos por tipo. O experimento do tipo 0 se tornou muito mais difícil de responder quando comparado aos outros dois experimentos. Contudo, devemos atentar ao fato deste ser um critério subjetivo haja vista que o nível de confiança é relativo para cada participante.

4.1.4 Interatividade (mouse *log*) e tipo de experimento

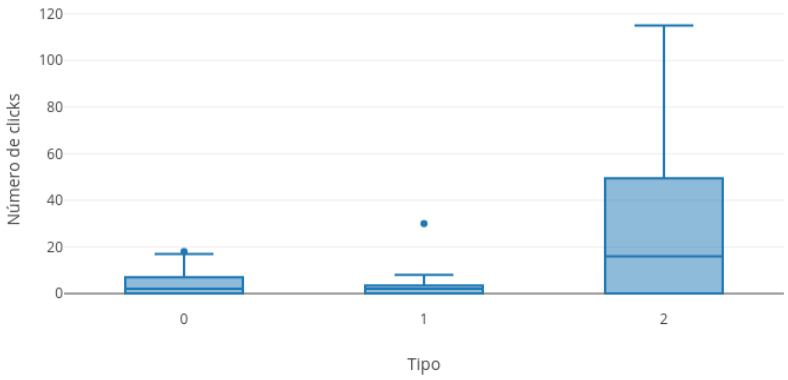


Figure 4.3 Boxplot dos números clicks por tipo de experimento.

Neste trabalho, decidimos considerar interatividade no experimento a atividade de movimento e clicks do mouse sobre a visualização. No entanto, temos ciência que interatividade vai além destas duas variáveis. Entretanto, para termos práticos consideraremos interatividade os *logs* do mouse (movimento e click).

Realizamos o teste ANOVA tanto para a variável de click quanto de movimento. Detectamos que a variável de movimento não apresentou diferença significativa entre os grupos (F = 2,193; p = 0,122; α = 0,05). Conseguimos verificar, no entanto, algo também observado no experimento presencial: o fato de que vários participantes utilizaram o cursor e movimento do mouse como ferramenta de auxílio na interpretação do gráfico, independente do tipo de experimento.

Já para clicks observamos uma diferença significativa entre o experimento tipo 2 (interativo) e os outros dois tipos (Figura 4.3), de acordo com o teste de Tukey (p = 0,002 com o tipo

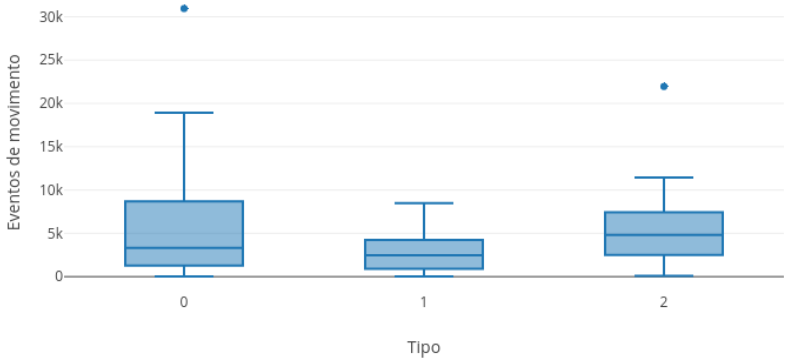


Figure 4.4 Boxplot dos eventos de movimento do mouse registrado por tipo de experimento.

0 e p = 0,001 com o tipo 1). Este resultado está dentro do esperado, visto que se imagina que ao interagir com a visualização o usuário realize mais clicks em relação à visualização estática.

Também é importante ressaltar a variância de clicks do experimento tipo 2 (interativo) quando comparada com os outros tipos (tipo 0: σ^2 = 34,22; tipo 1: σ^2 = 45,99,22; tipo 2: σ^2 = 2049,04). Como pode ser observado no *boxplot* (Figura 4.4), houve participantes que realizaram bastante clicks enquanto outros não realizaram tal interação.

4.1.5 Correlação entre acertos, certeza e clicks

No gráfico de bolha (Figura 4.5), exploramos a correlação entre acertos, certeza e clicks por participante. Gostaríamos de saber se quanto maior o número de clicks (interação), maior o número de acertos e confiança ao responder as perguntas. Cada bolha representa um participante, o eixo Y número de acerto, o eixo X a média do grau de certeza, o tamanho bolha representa o número de clicks e a cor o tipo do experimento.

Neste gráfico podemos observar que participantes do tipo 2 (interativo) que apresentaram valores baixo de click tiveram performance bastante similar às encontradas pelos participantes do tipo 0 (estático, sem tabela). Sinalizando que estes participantes não perceberam que se tratava de uma visualização interativa. Verificamos que estas bolhas apresentaram menor grau de certeza em comparação às outras, mas não necessariamente menor número de acertos.

Outra característica interessante da visualização encontra-se na parte superior direita do gráfico, na qual podemos perceber uma concentração de atividade para os experimentos dos tipos 1 e 2. Diferem somente no tamanho da bolha, ou seja, quantidade de clicks. Levando a concluir que experimentos do tipo 1 e tipo 2 (que apresentaram clicks) tiveram valores de acertos e graus de certeza maiores (com coeficiente de correlação de Pearson, r = 0,71). Tal comportamento sugere que apesar de não interferir diretamente na quantidade de acertos, o auxilio da interatividade e da tabela aumentaram relativamente o grau de certeza dos participantes.

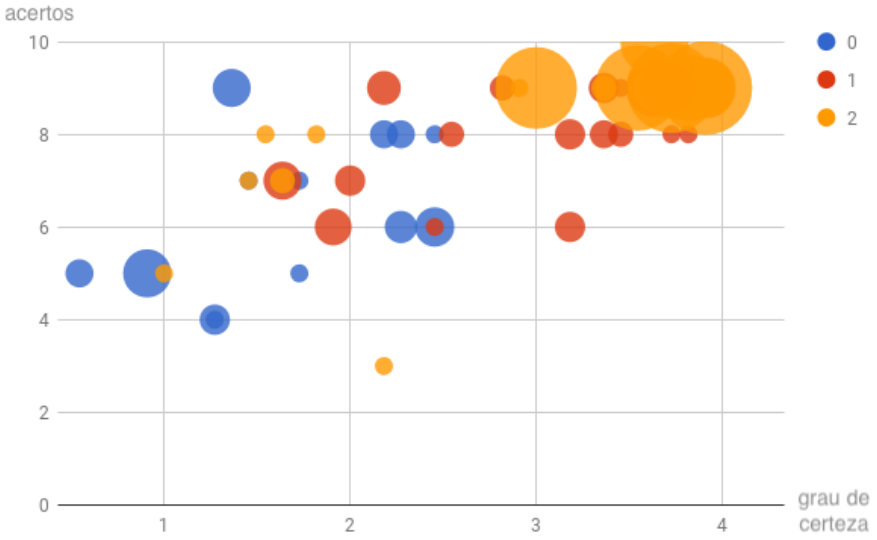


Figure 4.5 Gráfico que relaciona acertos, grau de certeza, clicks por participante do experimento. Eixo X: grau de certeza, Eixo Y: acertos, tamanho da bolha: número de clicks.

4.2 Estudo presencial

O estudo presencial ocorreu no dia 5 de junho de 2018 nas salas de estudo das bibliotecas do Centro de Ciências da Saúde (CCS) e Centro de Ciências Sociais e Aplicadas (CCSA). Este estudo tinha como principal meta observar e entender de que forma usuários noviços lidam com os diferentes tipos de configurações de visualização. Principalmente, desejávamos compreender aspectos de comportamento e opiniões dados os diferentes cenários aos quais estes usuários eram submetidos.

Neste estudo, foram recrutados de forma aleatória oito pessoas, quatro mulheres e quatro homens, de 5 cursos da UFPE das áreas de exatas, saúde e humanas. O tempo médio do experimento foi de 30 a 35 minutos. Sendo este dividido em duas partes. A primeira parte consistiu do usuário interagindo com o sistema (similar ao a distância) falando o que ele estava tentando realizar neste, de acordo com a técnica de "*Think-Aloud*". Na segunda parte, o participante respondeu uma série de perguntas semiestruturadas sobre a situação que ele vivenciou. Todos os experimentos foram gravados em vídeo para possibilitar análises posteriores.

4.2.1 Considerações Gerais

Ao entrar em contato pela primeira vez com a visualização (tutorial), em geral, os participantes compreenderam rapidamente o que a visualização codificava, independente do tipo de experimento. No entanto, devemos ressaltar que no tutorial havia juntamente com o gráfico uma tabela por no máximo 1 minuto. Em alguns casos, o participante explicou sua percepção sobre o gráfico (como este codificava informação) e logo após olhou para a tabela, de modo a confirmar o seu entendimento prévio sobre a visualização. Desta forma, os participantes con-

seguiram responder as perguntas do tutorial sem problemas.

Todavia, os participantes em geral enfrentaram dificuldades ao tentar responder perguntas relativas a UFPE. Eles relataram que o gráfico era bastante denso e que a quantidade de linhas tornava quase impossível extrair certos tipos de informação dentro da visualização. Como pode ser visto por alguns extratos do que os participantes falaram:

Participante C: "Eu achei a visualização desse gráfico péssima, porque é tudo da mesma cor, é um bocado de reta tudo junta, não ajuda muito."

Participante D: "Eu vou passando o mouse por cima aqui pra ver se não me perco...agora me perdi porque tem um bocado de linha aqui."

Outro fato interessante, foi o uso do cursor e movimento do mouse como guia. No experimento, os participantes constantemente utilizaram o mouse para guiar o olhar ou acompanhar uma linha dentro do gráfico. Desse modo, servindo como ferramenta de filtro e foco.

Apesar de cada tipo de experimento apresentar suas particularidades, os pontos descritos acima foram comum a todos os tipos de experimento.

4.2.2 Tipo 0 - Visualização sem interatividade e sem auxílio de tabela

Nesta modalidade de experimento o participante foi submetido a visualização de GCP que não possuía nenhum tipo de auxílio de interatividade. No experimento presencial, duas pessoas foram submetidas a este tipo de visualização. Durante seu uso, observamos que os participantes enfrentaram uma série de dificuldades ao interpretar o gráfico. Contudo, isto ocorreu somente no gráfico com informações relativas a UFPE, passando pelo tutorial sem complicações.

No gráfico com informações da UFPE, os participantes relataram confusão ao tentar encontrar informação onde a densidade de retas era muito grande. O participante E expressou este fato quando perguntado o que achou mais difícil na visualização, *"Eu não conseguia acompanhar o caminho de uma linha. Por exemplo, eu começava acompanhando ela mas como era um número de informação muito grande ela se juntava com outra linha e daqui a pouco eu não conseguia entender o caminho dela."*. Talvez estes problemas justifiquem as médias mais baixas do grau de certeza por tipo. Quanto ao tempo e quantidade de acertos este grupo não diferiu dos outros que realizaram o experimento presencialmente.

4.2.3 Tipo 1 - Visualização sem interatividade e com auxílio de tabela

Neste tipo de experimento, o usuário foi apresentado a uma visualização estática e a uma tabela. Neste tipo de situação gostaríamos de verificar se os participantes preferiam a tabela ao gráfico.

Durante o experimento presencial, três voluntários se submeteram a este tipo de situação. Inicialmente, no tutorial, os participantes fizeram uso majoritariamente do gráfico para

responder às perguntas propostas. Uma vez que se tratava de um experimento sobre visualização e estes não sabiam que poderiam utilizar a tabela. Entretanto ao saberem que poderiam utilizar a tabela sem nenhum prejuízo ao experimento, estes a utilizaram constantemente.

Foi notado que a tabela serviu, principalmente, como instrumento de apoio ao gráfico. Seja para encontrar uma resposta que não foi possível verificar no gráfico ou como forma de certificar que a resposta já encontrada no gráfico estava realmente correta. A participante A, por exemplo, esboçou durante o experimento que para responder uma pergunta, a tabela era mais exata, *"Eu acho mais fácil olhando pela tabela, porque é mais exato. Aqui (no gráfico) você tem que procurar onde a linhazinha tá chegando..."*.

Apesar da tabela servir como apoio, muitas vezes durante o experimento presencial não foi notado diferença no nível de certeza em relação às respostas quando comparado com os outros grupos. Observamos também que assim como nos outros grupos a tabela não ajudou os participantes a responderem perguntas que fazem correlação entre as variáveis. Os participantes desta modalidade descreveram este tipo de pergunta como as mais difíceis que eles tiveram que responder. As variáveis de acerto e tempo também não apresentam diferença quando comparadas os outros grupos.

4.2.4 Tipo 2 - Visualização com interatividade e sem auxílio de tabela

O experimento do tipo 2 é o único entre os três tipos que apresentou interatividade. Na fase presencial, três participantes utilizaram este tipo de visualização. Somente a participante B percebeu desde o início que se tratava de uma visualização interativa (apesar do cursor do mouse mudar de forma ao passar por cima das partes interativas do gráfico). Por este motivo, durante o experimento, o moderador perguntou para os outros dois participantes se a mudança de formato no cursor do mouse significava alguma coisa para eles. Dado este *prompt*, o participante H descobriu a interatividade da visualização, enquanto a participante C respondeu que aquele formato não significava nada para ela.

Considerando os dois participantes que utilizaram o experimento de forma interativa, observamos que o único tipo de capacidade interativa utilizada foi o *brushing* nos eixos. Os participantes não tentaram rearranjar os eixos apesar de terem realizado um breve movimento dos eixos ao tentarem fazer o*brushing*. Não sabemos se estes não realizaram tal ação por não saber que era possível ou por não relacionar a utilidade desta função com a capacidade de correlacionar dados.

Quanto ao textitbrushing do eixo que possibilitava a seleção de linha no gráfico, os participantes B e H expressaram opiniões adversas quanto à utilidade da interação:

Participante B: "...Eu ainda achei um pouco difícil, mas ainda essa coisa de você selecionar deu uma facilitada"

Participante H: "Agora com essa ferramenta dá pra ter certeza do que você está marcando"

No entanto, percebemos que a função de seleção não ocorreu de forma descomplicada durante o experimento. Notamos que o usuário, ao tentar realizar a seleção, clicava diretamente na linha ao invés de realizar *brushing* no eixo. Os dois participantes expressaram frustração quanto ao ato de selecionar. A participante B reclamou que a seleção "enganchava" e o participante H sugeriu como melhoramento poder selecionar ao clicar em uma parte textual do gráfico.

Em suma, os experimentos da modalidade interativa não apresentaram performance melhor quando comparados com os outros. Uma vez que os participantes tiveram dificuldade em descobrir a interatividade e de utilizar estas funções.

5

Discussão

Baseado nos resultados das duas modalidades de experimento (presencial e a distância) conseguimos sugestões de como responder às hipóteses propostas neste trabalho (H1, H2 e H3). Apesar de GCP ser uma visualização majoritariamente utilizada por especialistas, os participantes noviços, em sua grande maioria, não apresentaram dificuldade em entender como a informação foi codificada. Este fato ficou bastante evidente durante o experimento presencial, no qual participantes descreveram seu entendimento sobre o gráfico sem dificuldade.

Todavia, o mesmo não pode ser afirmado em relação aos participantes extraírem informação do gráfico quando este apresentou um número elevado de informação. Isto ocorreu na segunda fase do experimento, situação na qual o gráfico possuía uma densidade de linhas muito maior em comparação ao tutorial. Nesta circunstância, 66,67% dos usuários do experimento tipo 0 (estático, sem tabela) tiveram performance inferior em relação a média geral (7,63 acertos). Enquanto para os outros tipos esse número decai para 27,78% e 21,05% para o tipo 1 (estático, com tabela) e tipo 2 (interativo) respectivamente. Este contexto comprova a necessidade do uso da interatividade ou auxílios textuais, como a tabela, para grandes quantidades de informação em visualizações.

No entanto, neste experimento, pudemos verificar que estes auxílios cognitivos como a interatividade devem ser implementados respeitando os princípios de design e usabilidade. Pois, como já foi advogado por Lam LAM (2008), o usuário apresenta expectativas ao interagir com a visualização. Caso estas expectativas não sejam atendidas, pode causar a pessoa dificuldade e frustração. Este cenário foi observado durante o experimento presencial, todos os participantes da modalidade interativa demonstraram dificuldade em descobrir as características interativas da visualização (nem todos descobriram). Os que lograram em descobrir a interatividade, a utilizaram de forma parcial e com dificuldade. O mesmo foi constatado no experimento a distância, no qual alguns participantes desta modalidade apresentaram comportamento similar aos do tipo 0 (estático, sem tabela).

Em relação ao tempo não foi encontrado diferença significativa entre os grupos (ANOVA com p = 0,093). Logo, não podemos afirmar que a hipótese H3, que relaciona tempo e engajamento, é verdadeira. Apesar de usuários da versão interativa apresentarem uma média de tempo

superior (647,36s) a versão estática sem tabela (462,77s) e com tabela (627,5s).

5.1 Interatividade e FVE

Baseado nos fatos reportados acima, podemos discutir as hipóteses H1 e H2. Estas hipóteses discutem diretamente a relação da interatividade e formação do FVE.

Com os achados do experimento pudemos notar que o entendimento da codificação da informação do gráfico não está diretamente ligado a presença da interatividade. Pois, como foi relatado, na fase do tutorial os participantes conseguiram entender o funcionamento do gráfico independente da modalidade do experimento. Todavia, durante o experimento conseguimos notar que o uso da interatividade ajudou a confirmar o frame criado pelo o usuário, posto que este recebia um *feedback* da informação buscada. Este fenômeno se relaciona diretamente ao que Lee at al. LEE et al. (2016) discutem em seu artigo sobre como noviços compreendem visualizações. Similar ao que foi descrito em seu artigo, os participantes no experimento do presente trabalho utilizaram a interatividade para confirmar o entendimento da codificação do gráfico FVE e assim aperfeiçoar o mapa mental sobre o funcionamento da visualização e sobre a informação contida nesta. O que talvez justifique os níveis de certeza mais altos em relação ao experimento do tipo 0 (estático, sem tabela). Entretanto, vale ressaltar que níveis superiores de performance também ocorreram na modalidade com tabela, provando que o usuário necessita de "âncoras cognitivas" LIU; STASKO (2010) no processo do entendimento de um sistema e informação. Esta afirmação se relaciona diretamente no que Liu e Stasko LIU; STASKO (2010) defendem quanto ao papel da interação em visualizações. Stasko et al., que utilizam exatamente o termo "âncora" em seu trabalho, discutem o uso da interatividade e suas características no processo de formação do mapa mental de uma visualização.

Baseado no trabalho de Stasko et al., pudemos observar que as ferramentas e condições impostas na visualização interferiram diretamente como o processo de compreensão ocorreu para o participante. Participantes, por exemplo, que não sucederam em arrastar os eixos do gráfico ficaram impossibilitados de fazer análises mais amplas sobre correlacionar variáveis e provavelmente não fizeram conexão da utilidade disto e o funcionamento do gráfico. Logo, podemos sugerir que a interatividade pode ser um amplificador do processo cognitivo, mas também um limitador, caso este não funcione de forma fluída entre o processo mental e o mundo físico.

Portanto, não seria correto afirmar que as hipóteses H1 e H2 são verdadeiras, pois, ao mesmos tempo que a interatividade é essencial para a visualização MUNZNER (2014) como forma de amplificar suas capacidades, ela também pode restringir o entendimento do funcionamento da visualização FVE e informação FC representada nesta. Isto ocorre em casos que sua comunicação com o mapa mental do usuário não ocorre da maneira esperada. Em suma o experimento demonstrou a importância do uso consciente da interatividade dentro da visualização, uma vez que serve como guia para as ações do usuário dentro do sistema. Além disso, é

fundamental seguirmos certos princípios de design e usabilidade NORMAN (2013); NIELSEN (1994); MACE (1997) afim de evitarmos que a comunicação entre o usuário e o sistema apresente ruído. Do mesmo modo, não devemos esquecer da ajuda que elementos textuais podem adicionar a visualização, como foi o caso da tabela no experimento do tipo 1. Desse modo, conseguiremos que o processo de formação e refinamento do FVE ocorra mais facilmente.

5.2 Limitações

Este estudo apresenta algumas limitações quanto às situações as quais os participantes foram submetidos, tanto no experimento presencial quanto no experimento a distância. Dado a população de usuários escolhida, noviços, devemos considerar que seria improvável que estes encontrem este tipo de visualização na situação imposta pelo experimento. Muito provavelmente, os participantes entrariam em contato com visualizações em contextos mais casuais como sites de notícias ou materiais de estudo. Logo, não podemos afirmar que o comportamento dos participantes ao interpretar o gráfico seria similar, posto que diversas variáveis como, por exemplo, tempo disponível, texto de apoio e objetivo poderiam influenciar nos resultados. Além disso, no tutorial, o experimento contou por 1 minuto com uma tabela que continha a mesma informação do gráfico. Até então não sabemos a influência desta no aprendizado da visualização.

Ademais, no experimento presencial, utilizamos o método "*Think-Aloud*". Este método tem impacto no processo cognitivo do participante RUBIN; CHISNELL (2008); NIELSEN; CLEMMENSEN; YSSING (2002). Uma vez que falar pode ajudar o usuário a construir ideias, fato que não ocorreria no contexto real. Também, nem tudo que o usuário pensou provavelmente foi verbalizado, pois o processo cognitivo provavelmente é mais rápido e complexo NIELSEN; CLEMMENSEN; YSSING (2002).

Também, existe a limitação referente ao número de participantes tanto do estudo presencial quanto a distância. Para que pudéssemos encontrar padrões mais precisos no comportamento das diferentes modalidades de experimento o ideal seria testar com uma população maior.

Por fim, as discussões alcançadas no âmbito de interatividade são somente relativas a visualização escolhida para o experimento, GCP, e as técnicas de interação utilizadas na visualização.

6

Conclusão

Neste estudo, discutimos a importância da interatividade dentro dos sistemas de Infovis. Além disso, apresentamos os resultados de um experimento conduzido com estudantes que não conheciam o GCP. Este estudo, que contou com duas modalidades presencial e a distância, conseguiu delinear os efeitos que a interatividade causa no processo de compreensão de visualizações. Com efeito, o presente trabalho sucede em descrever o processo de compreensão dos participantes, além de ressaltar alguns pontos positivos e algumas "armadilhas" causadas pela interatividade.

Acreditamos que há muitas oportunidades dentro do campo de Infovis para aprofundar este assunto. Como por exemplo, outros fatores que podem afetar a compreensão de uma visualização desconhecida, como formação e conhecimento prévio do usuário, interesse na informação e contexto. Trabalhos que explorem o desenvolvimento de heurísticas para avaliar a usabilidade e efetividade de uma visualização também julgamos que gerariam resultados relevantes.

Sabemos que este trabalho não apresenta respostas definitivas sobre o efeito da interatividade dentro de visualizações. Mas representa uma contribuição para que sistemas de Infovis possam cada vez mais incluir diferentes audiências.

Referências

AMAR, R.; EAGAN, J.; STASKO, J. Low-level components of analytic activity in information visualization. In: INFORMATION VISUALIZATION, 2005. INFOVIS 2005. IEEE SYMPOSIUM ON. **Anais...** [S.l.: s.n.], 2005. p.111–117.

BAUR, D. **The death of interactive infographics? Medium, 2017**. Acesso em: 09 de setembro de 2017, "https://medium.com/@dominikus/the-end-of-interactive-visualizations-52c585dcafcb".

BAUR, D.; GODDEMEYER, D. **SUBSPOTTING**. Acesso em: 09 de setembro de 2017, "https://do.minik.us/blog/subspotting".

BOSTOCK, M. D3. js. **Data Driven Documents**, [S.l.], v.492, p.701, 2012.

BREHMER, M.; MUNZNER, T. A multi-level typology of abstract visualization tasks. **IEEE Transactions on Visualization and Computer Graphics**, [S.l.], v.19, n.12, p.2376–2385, 2013.

DIX, A. J. et al. **Human-Computer Interaction**. 3.ed. [S.l.]: Prentice Hall, 2003.

ELMQVIST, N. et al. Fluid interaction for information visualization. **Information Visualization**, [S.l.], v.10, n.4, p.327–340, 2011.

GRAMMEL, L.; TORY, M.; STOREY, M.-A. How information visualization novices construct visualizations. **IEEE transactions on visualization and computer graphics**, [S.l.], v.16, n.6, p.943–952, 2010.

HEER, J.; SHNEIDERMAN, B. Interactive dynamics for visual analysis. **Queue**, [S.l.], v.10, n.2, p.30, 2012.

INSELBERG, A. The plane with parallel coordinates. **The visual computer**, [S.l.], v.1, n.2, p.69–91, 1985.

LAM, H. A framework of interaction costs in information visualization. **IEEE transactions on visualization and computer graphics**, [S.l.], v.14, n.6, 2008.

LEE, S. et al. How do people make sense of unfamiliar visualizations?: a grounded model of novice's information visualization sensemaking. **IEEE transactions on visualization and computer graphics**, [S.l.], v.22, n.1, p.499–508, 2016.

LIU, Z.; STASKO, J. Mental models, visual reasoning and interaction in information visualization: a top-down perspective. **IEEE transactions on visualization and computer graphics**, [S.l.], v.16, n.6, p.999–1008, 2010.

MACE, R. What is universal design. **The Center for Universal Design at North Carolina State University. Retrieved Retrieved November**, [S.l.], v.19, p.2004, 1997.

MOORE, D. T.; HOFFMAN, R. R. Data-frame theory of sensemaking as a best model for intelligence. **American Intelligence Journal**, [S.l.], v.29, n.2, p.145–158, 2011.

MUNZNER, T. **Visualization analysis and design**. [S.l.]: CRC press, 2014.

NIELSEN, J. **Usability engineering**. [S.l.]: Elsevier, 1994.

NIELSEN, J.; CLEMMENSEN, T.; YSSING, C. Getting access to what goes on in people's heads?: reflections on the think-aloud technique. In: NORDIC CONFERENCE ON HUMAN-COMPUTER INTERACTION. **Proceedings...** [S.l.: s.n.], 2002. p.101–110.

NORMAN, D. **The design of everyday things**: revised and expanded edition. [S.l.]: Basic Books (AZ), 2013.

PIKE, W. A. et al. The science of interaction. **Information Visualization**, [S.l.], v.8, n.4, p.263–274, 2009.

PLOTLY. **Plotly**. Acesso em: 12 de setembro de 2017, "`https://plot.ly`".

POUSMAN, Z.; STASKO, J.; MATEAS, M. Casual information visualization: depictions of data in everyday life. **IEEE transactions on visualization and computer graphics**, [S.l.], v.13, n.6, 2007.

RUBIN, J.; CHISNELL, D. **Handbook of usability testing**: howto plan, design, and conduct effective tests. [S.l.]: John Wiley & Sons, 2008.

RUSSELL, D. M. et al. The cost structure of sensemaking. In: INTERACT'93 AND CHI'93 CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS. **Proceedings...** [S.l.: s.n.], 1993. p.269–276.

SCAIFE, M.; ROGERS, Y. External cognition: how do graphical representations work? **International journal of human-computer studies**, [S.l.], v.45, n.2, p.185–213, 1996.

SHNEIDERMAN, B. Tree visualization with tree-maps: 2-d space-filling approach. **ACM Transactions on graphics (TOG)**, [S.l.], v.11, n.1, p.92–99, 1992.

SIIRTOLA, H. Direct manipulation of parallel coordinates. In: INFORMATION VISUALIZATION, 2000. PROCEEDINGS. IEEE INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2000. p.373–378.

TABLEU. **Tableu**. Acesso em: 12 de setembro de 2017, "`https://www.tableau.com`".

VICTOR, B. Up and Down the Ladder of Abstraction: a systematic approach to interactive visualization. **Kill Math**, [S.l.], 2011.

VIEGAS, F. B. et al. Manyeyes: a site for visualization at internet scale. **IEEE transactions on visualization and computer graphics**, [S.l.], v.13, n.6, 2007.

YALÇIN, M. A.; ELMQVIST, N.; BEDERSON, B. B. Cognitive Stages in Visual Data Exploration. In: BELIV WORKSHOP: BEYOND TIME AND ERRORS-NOVEL EVALUATION METHODS FOR VISUALIZATION. **Proceedings...** [S.l.: s.n.], 2016.

YI, J. S.; KANG, Y. ah; STASKO, J. Toward a deeper understanding of the role of interaction in information visualization. **IEEE transactions on visualization and computer graphics**, [S.l.], v.13, n.6, p.1224–1231, 2007.

7

Apêndice

7.1 Cronograma

As atividades realizadas nesse trabalho seguiram o seguinte cronograma. Declaro que a coleta de dados só foi iniciada após a aprovação do Comité de Ética.

Atividades	**Fev 2018**	**Mar 2018**	**Abr 2018**	**Mai 2018**	**Jun 2018**
Levantamento bibliográfico	x	x	x	x	x
Estudo de Tecnologias para experimento		x	x		
Design e Implementação do experimento		x	x	x	
Obtensão de dados para o experimento		x			
Estudo piloto*				x	
Realização do experimento*					x
Análise dos resultados					x
Escrita do trabalho final					x
Elaboração da Apresentação					x
Defesa do trabalho de graduação					x

Table 7.1 Cronograma de atividades.

7.2 Aspectos Éticos

Este estudo foi planejado respeitando a resolução nž 466, de 12 de dezembro de 2012 da CONEP.

Os indivíduos que participaram do estudo estavam sujeitos aos seguintes riscos: situações durante o estudo no qual o indivíduo se sinta constrangido por não saber manusear o sistema ou não saber a resposta para uma pergunta proposta. Como forma de amenizar esta

situação, orientamos o participante a utilizar o sistema no momento em que se sentiu seguro e ele pôde levar quanto tempo ele julgou necessário para interagir com o sistema e responder as perguntas.

Durante o experimento foram realizadas perguntas sobre um tipo de visualização que será apresentada ao indivíduo. No entanto, no sistema utilizado pelo usuário não houve necessidade deste entrar com informações pessoais que possam identifica-lo. Além disso, foi cumprido o seu assentimento livre e esclarecido.

No decorrer do experimento foram realizadas gravações de áudio, filmagem que serão arquivadas no computador do pesquisador por um período mínimo de 5 anos. O pesquisador garante a segurança e integridade destes dados.

Ao fazer parte do experimento, o participante foi beneficiado na medida em que conheceu e aprendeu um tipo de visualização de dados, no caso do experimento coordenadas paralelas. Ademais, este estudo contribuiu para a comunidade de Infovis entender melhor como ajudar pessoas não especialistas a fazerem uso de sistemas Infovis.

7.3 Termo de Consentimento Livre e Esclarecido

Convidamos o (a) Sr. (a) para participar como voluntário (a) da pesquisa UM ESTUDO SOBRE O IMPACTO DAINTERATIVIDADE NA COMPREENSÃO DEVISUALIZAÇÕES POR USUÁRIOS NÃO ESPECIALISTAS, da aluna Mariama Celi Serafim de Oliveira (e-mail: mcso@cin.ufpe.br) e que está sob a responsabilidade do pesquisador Nivan Roberto Ferreira Junior (Endereço: Centro de Informática (CIn) da UFPE Av. Jornalista Aníbal Fernandes, s/n - Cidade Universitária (Campus Recife) - CEP: 50.740-560 | Telefone: (81) 2126-8430, e-mail: nivan@cin.ufpe.br). Todas as suas dúvidas podem ser esclarecidas com o responsável por esta pesquisa. Apenas quando todos os esclarecimentos forem dados e você concorde com a realização do estudo, pedimos que rubrique as folhas e assine ao final deste documento, que está em duas vias. Uma via lhe será entregue e a outra ficará com o pesquisador responsável. Você estará livre para decidir participar ou recusar-se. Caso não aceite participar, não haverá nenhum problema, desistir é um direito seu, bem como será possível retirar o consentimento em qualquer fase da pesquisa, também sem nenhuma penalidade.

INFORMAÇÕES SOBRE A PESQUISA

Dado o expressivo crescimento do acesso a sistemas Infovis por múltiplas audiências, o presente trabalho tem por objetivo analisar o impacto da interatividade no entendimento de visualizações por pessoas não especialistas.

Neste experimento serão realizadas perguntas e observações do participante interagindo com uma visualização. Além disso, esta sessão será gravada para possibilitar que análises fu-

turas possam ser realizadas sobre o experimento.

RISCOS: Pode ocorrer situações nas quais você se sinta constrangido ou tímido por não saber a resposta para uma pergunta proposta. Há também a possibilidade que você não saiba o que fazer no sistema, visto que se trata de uma visualização desconhecida para você. Durante a utilização do sistema, você poderá ficar à vontade caso não se sinta confortável em falar e poderá utilizar o sistema no momento que se sentir confortável, levando quanto tempo precisar.

BENEFÍCIOS: Participar possibilitará que você conheça e aprenda um tipo de visualização de dados, no caso do experimento coordenadas paralelas.

Todas as informações desta pesquisa serão confidenciais e serão divulgadas apenas em eventos ou publicações científicas, não havendo identificação dos voluntários, a não ser entre os responsáveis pelo estudo, sendo assegurado o sigilo sobre a sua participação. Os dados coletados nesta pesquisa (gravações de áudio, entrevistas, fotos e filmagens), ficarão armazenados em computador pessoal, sob a responsabilidade do pesquisador no endereço acima informado, pelo período de mínimo 5 anos. Nada lhe será pago e nem será cobrado para participar desta pesquisa, pois a aceitação é voluntária, mas fica também garantida a indenização em casos de danos, comprovadamente decorrentes da participação na pesquisa, conforme decisão judicial ou extra-judicial. Se houver necessidade, as despesas para a sua participação serão assumidas pelos pesquisadores (ressarcimento de transporte e alimentação). Em caso de dúvidas relacionadas aos aspectos éticos deste estudo, você poderá consultar o Comitê de Ética em Pesquisa Envolvendo Seres Humanos da UFPE no endereço: **(Avenida da Engenharia s/n 1ž Andar, sala 4 - Cidade Universitária, Recife-PE, CEP: 50740-600, Tel.: (81) 2126.8588 e-mail: cepccs@ufpe.br)**.

_____

(assinatura do pesquisador)

CONSENTIMENTO DA PARTICIPAÇÃO DA PESSOA COMO VOLUNTÁRIO (A)

Eu,_____, CPF _____, abaixo assinado, após a leitura (ou a escuta da leitura) deste documento e de ter tido a oportunidade de conversar e ter esclarecido as minhas dúvidas com o pesquisador responsável, concordo em participar do estudo UM ESTUDO SOBRE O IMPACTO DAINTERATIVIDADE NA COMPREENSÃO DEVISUALIZAÇÕES POR USUÁRIOS NÃOESPECIALISTAS, como voluntário (a). Fui devidamente informado (a) e esclarecido (a) pelo pesquisador sobre a pesquisa,

os procedimentos nela envolvidos, assim como os possíveis riscos e benefícios decorrentes de minha participação. Foi-me garantido que posso retirar o meu consentimento a qualquer momento, sem que isto leve a qualquer penalidade.

Local e data:_____

Assinatura do participante:_____

Presenciamos a solicitação de consentimento, esclarecimentos sobre a pesquisa e o aceite do voluntário em participar. (02 testemunhas não ligadas à equipe de pesquisadores):

Nome:	Nome:
Assinatura:	Assinatura:

7.4 Convite para experimento - Estudo a distância

Estamos realizando um experimento on-line parte fundamental do trabalho de conclusão de curso da aluna Mariama Oliveira na área de visualização de dados. Este experimento visa analisar o impacto da interatividade no entendimento de visualizações por pessoas não especialistas. Este possui duração de aproximadamente 20 minutos e consiste no participante interagir com um sistema web e responder em seguida algumas perguntas.

Todas as informações desta pesquisa serão confidenciais e serão divulgadas apenas em eventos ou publicações científicas, não havendo identificação dos voluntários, a não ser entre os responsáveis pelo estudo, sendo assegurado o sigilo sobre a sua participação. Qualquer comentário, orientação ou dúvida pode ser submetida através dos e-mails (mcso@cin.ufpe.br ou nivan@cin.ufpe.br).

Obrigada por seu tempo e cooperação.

Informações Adicionais da pesquisa

RISCOS: Pode ocorrer situações nas quais você se sinta constrangido ou tímido por não saber a resposta para uma pergunta proposta. Há também a possibilidade que você não saiba oque fazer no sistema, visto que se trata de uma visualização desconhecida para você. Durante a utilização do sistema, você poderá ficar à vontade em não responder caso não se sinta confortável e poderá utilizar o sistema no momento que se sentir confortável, levando quanto tempo precisar.

BENEFÍCIOS: Participar possibilitará que você conheça e aprenda um tipo de visualização de dados, no caso do experimento coordenadas paralelas.

7.5 Convite para experimento - Estudo presencial

Estamos realizando um experimento presencial parte fundamental do trabalho de conclusão de curso da aluna Mariama Oliveira na área de visualização de dados. Este experimento visa analisar o impacto da interatividade no entendimento de visualizações por pessoas não especialistas. Este possui duração de aproximadamente 40 minutos e consiste no participante interagir com um sistema web e responder em seguida algumas perguntas. Durante o experimento o usuário será pedido para expressar suas ações no sistema verbalmente. Além disso, gravações do participante serão realizadas durante o experimento para futuras análises.

Todas as informações desta pesquisa serão confidenciais e serão divulgadas apenas em eventos ou publicações científicas, não havendo identificação dos voluntários, a não ser entre os responsáveis pelo estudo, sendo assegurado o sigilo sobre a sua participação. Qualquer comentário, orientação ou dúvida pode ser submetida através dos e-mails (mcso@cin.ufpe.br ou nivan@cin.ufpe.br).

Obrigada por seu tempo e cooperação.

Informações Adicionais da pesquisa

RISCOS: Pode ocorrer situações nas quais você se sinta constrangido ou tímido por não saber a resposta para uma pergunta proposta. Há também a possibilidade que você não saiba o que fazer no sistema, visto que se trata de uma visualização desconhecida para você. Durante a utilização do sistema, você poderá ficar à vontade caso não se sinta confortável em falar e poderá utilizar o sistema no momento que se sentir confortável, levando quanto tempo precisar.

BENEFÍCIOS: Participar possibilitará que você conheça e aprenda um tipo de visualização de dados, no caso do experimento coordenadas paralelas.

7.6 Perguntas do tutorial

Pergunta
1. A nota de quantos estudates é apresentada na visualização?
2. Quantos estudantes possuem nota entre 65 e 85 em Física?
3. Qual a pessoa com a menor nota em Matemática?
4. Estudantes com nota boa em Biologia também tiram nota boa em qual matéria?

Table 7.2: Perguntas apresentadas ao participante durante o tutorial.

7.7 Perguntas do experimento

Pergunta	**Função**	**Atividade de interação com a ferramento**	**Atividade de interação (intensão)**
1. Qual o número de matriculados no curso de Direito?	Overview sobre todos os atributos	overview	selecionar
2. Quantos cursos possuem mais de 250 alunos ingressantes?	Overview sobre todos os atributos	overview	selecionar
3. Quais cursos possuem entre 1000 e 1200 matriculados?	Selecionar uma faixa de itens	brushing	filtrar, selecionar
4. Quantos cursos possuem até 20 formandos?	Selecionar uma faixa de itens	brushing	filtrar
5. Qual curso que possui a maior média de idade?	Detecção de outliers	overview	-
6. Qual curso com o maior número de alunos matriculados?	Detecção de outliers	overview	-
7. Qual curso possui até 400 matriculados e idade média de 21 anos?	Selecionar uma faixa de itens	brushing	filtrar
8. O que cursos da área de Educação possuem em comum?	Checar por correlação entre atributos	overview e brushing	filtrar

9. Dentre os 3 cursos com o maior número de matriculados, qual curso não obedece a relação de proporcionalidade de matriculados e formandos?	Overview sobre todos os atributos, Checar por correlação entre atributos,	overview, brushing, rearrajar os eixos	filtrar, reconfigurar
10. Qual curso que proporcionalmente ao número de entrada possui o menor número de formandos?	Checar por correlação entre atributos	brushing, rearrajar os eixos	-
11. Qual o intervalo da média de idade dos cursos da universidade?	Overview sobre todos os atributos	overview	-
12. O que cursos que possuem entre 1000 e 1500 matriculados possuem comum?	Checar por correlação entre atributos	brushing	filtar

Table 7.3: Perguntas apresentadas ao participante durante o experimento.

7.8 Perguntas do experimento presencial - pós uso do sistema

Pergunta
1. Você tinha algum conhecimento prévio destes dados?
2. Que característica desta visualização você achou mais útil? Por que?
3. Que característica desta visualização você achou mais gostou? Por que?
4. Que característica desta visualização você achou menos gostou? Por que?
5. Que coisas você achou mais difícil na visualização? Por que?
6. Que coisas você achou mais fácil na visualização? Por que?
7. Quais perguntas você teve mais dificuldade em responder? Por que?
8. Quais perguntas você teve menos dificuldade em responder? Por que?
9. Que tipo de ferramenta você sente falta nesta visualização?
10. Você acha que algum conhecimento prévio lhe ajudou a entender esta visualização?
11. Qual você utilizou mais a tabela ou o gráfico? Por que?*

*Caso o participante tenha utilizado o sistema do tipo 1 (estático com tabela).

Table 7.4 Perguntas do experimento presencial.