



Instituto de
MATEMÁTICA
E ESTATÍSTICA

UFRGS



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DEPARTAMENTO DE ESTATÍSTICA

VISUALIZAÇÃO DE DADOS E SUA IMPORTÂNCIA NA ERA DO *BIG DATA*

FELIPE GARCIA DA COSTA

Porto Alegre
2017

FELIPE GARCIA DA COSTA

VISUALIZAÇÃO DE DADOS E SUA IMPORTÂNCIA NA ERA DO *BIG DATA*

Trabalho de Conclusão de Curso submetido
como requisito parcial para a obtenção do grau de
Bacharel em Estatística.

Orientadora
Professora Dra. Lisiane Priscila Roldão Selau

Porto Alegre
2017

Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Departamento de Estatística

VISUALIZAÇÃO DE DADOS E SUA IMPORTÂNCIA NA ERA DO *BIG DATA*

FELIPE GARCIA DA COSTA

Banca examinadora:

Professora Dra. Lisiane Priscila Roldão Selau
UFRGS

Professora Dra. Liane Werner
UFRGS

AGRADECIMENTOS

Gostaria de agradecer a UFRGS, especialmente ao Instituto de Matemática e Estatística pelo acesso à um ensino de qualidade e gratuito; e aos professores por todos os ensinamentos durante esses anos que estive na universidade. Em especial a professora Lisiane Selau pela confiança, paciência e incentivo durante a orientação e apoio na construção deste trabalho.

À toda minha família, que me deu todo o apoio e amor necessário. Em especial a minha namorada pelo amor, carinho, atenção, generosidade e paciência, à minha mãe pela dedicação incansável e ao meu pai pelo exemplo e inspiração. Muito obrigado pelo carinho, conselhos e amor de sempre.

E ainda, minha gratidão a todos aqueles que de algum modo fizeram parte deste período. Todos têm sua contribuição no meu desenvolvimento profissional e pessoal.

*“You, me, or nobody is gonna hit as hard as life. But it ain't about how hard you hit.
It's about how hard you can get hit and keep moving forward. How much you can
take and keep moving forward. That's how winning is done!”*

Rocky Balboa

RESUMO

Nunca antes na história foram gerados volumes tão altos de dados como hoje. Explorar e analisar os vastos volumes de dados tornou-se cada vez mais difícil e complexo. Muitas vezes as mudanças acontecem a uma velocidade que supera a capacidade de aprendizagem e compreensão. Da mesma forma, há uma necessidade crescente de encontrar mecanismos para comunicar a informação às pessoas de forma eficiente e eficaz, e para ajudar a informá-los sobre processos e conceitos que afetam a vida cotidiana. A resposta para essa crescente necessidade se chama visualização de dados, um conceito que entrelaça um conjunto de técnicas para obter a informação a partir dos dados e da percepção do ser humano. O Big Data pode ser descrito como um conjunto de grandes volumes de dados, complexos, que dificilmente podem ser processados e analisados utilizando programas e processos usuais. Diante disto, o objetivo do presente trabalho é estudar o processo de construção de visualização de dados, desde a obtenção até a análise dos dados e seu uso no *Software R*, além de mostrar uma avaliação qualitativa sobre o tema. O método de visualização foi utilizado para construir estatísticas descritivas sobre um banco de dados de marketing com o objetivo de responder a hipóteses de negócio. Os resultados obtidos foram comparados com os mesmos resultados em formato de tabela a fim de obter uma avaliação qualitativa a partir de um gestor com experiência em visualização de dados e *big data*.

Palavras-chave: Visualização de dados. *Big Data*. *Software R*

ABSTRACT

The amount of data being generated today is unprecedented. Exploring and analyzing such vast amount of information has become a challenge on itself. Changes frequently happen at a pace that overwhelms our own comprehension and cognitive capacity. In addition, there is a growing necessity to efficiently and effectively communicate information to people, and help them understand the key concepts and trends that affect their daily life. The answer to this need is called data visualization, a concept that embraces several techniques to extract information from data and from our own perceptions. Big Data can be described as massive amounts of data that cannot be analyzed or processed using ordinary processes and software. From this standpoint, the objective of this study is to dissect the process of data visualization, from data collection to data analysis and its application on the R Software, as well as present a qualitative review of the subject. The visualization method was used to present descriptive statistics of a marketing database in order to answer business hypothesis. The results obtained were compared with the same results in table format in order to obtain a qualitative evaluation from a manager with experience in data visualization and big data.

Keywords: Data Visualization. Big Data. Software R

SUMÁRIO

1. INTRODUÇÃO.....	10
2. REFERENCIAL TEÓRICO.....	12
2.1 Conceito de <i>Visualização de Dados</i>	12
2.2 <i>Big Data</i>	12
2.3 Percepção Humana.....	14
2.4 Visualizar Padrões ao Longo do Tempo.....	15
2.5 Visualização de Dados Multivariados.....	17
2.6 Conceitos de Interação.....	20
3. METODOLOGIA.....	21
3.1 Hipóteses de Negócio.....	21
3.2 Dados.....	21
3.3 Transformação Numérica.....	21
3.4 Análise de Dados.....	22
3.5 Interpretação Gráfica.....	22
3.6 Interação do Usuário.....	22
3.7 Avaliação do Método.....	23
3.8 O <i>Software</i>	23
4. RESULTADOS.....	24
4.1 Hipóteses para o Problema de Negócio.....	24
4.2 Dados.....	24
4.3 Transformação Numérica.....	25
4.4 Análise de Dados.....	25
4.5 Interpretação Gráfica.....	26
4.6 Interação do Usuário.....	29

4.7 Avaliação do Método.....	31
5. CONSIDERAÇÕES FINAIS	33
REFERÊNCIAS.....	36

1. INTRODUÇÃO

Nunca antes na história os dados foram gerados em volumes tão altos como hoje. Explorar e analisar os vastos volumes de dados tornou-se cada vez mais difícil. A visualização de informações e a mineração de dados podem ajudar a lidar com o fluxo de informações. A vantagem da exploração de dados de forma visual é que o usuário está diretamente envolvido no processo de mineração de dados. Há um grande número de técnicas de visualização de informações que foram desenvolvidas ao longo dos últimos anos para apoiar a exploração de grandes conjuntos de dados (KEIM; WARD, 2003).

O progresso tecnológico, principalmente de *hardware*, permite que os sistemas computacionais de hoje armazenem quantidades muito grandes de dados. Pesquisadores da Universidade de Berkeley estimam que a cada ano são gerados cerca de 1 *exabyte* (1 milhão de *terabytes*) de dados, dos quais uma grande parte está disponível em formato digital. Isto significa que nos próximos três anos mais dados serão gerados do que em toda a história humana até esta data. Os dados são geralmente gravados automaticamente por meio de sensores e sistemas de monitoramento. Mesmo as transações simples da vida cotidiana, como pagar com cartão de crédito ou usar o telefone, são tipicamente registradas pelos computadores. Geralmente muitos parâmetros são gravados, resultando em dados com alta dimensionalidade. Este fenômeno é conhecido como “internet das coisas” (KEIM; WARD, 2003).

As imagens têm sido usadas como um mecanismo de comunicação desde antes da formalização da linguagem escrita. Uma única imagem pode conter uma riqueza de informações, e pode ser processada muito mais rápida quando comparada com páginas cheias de palavras. Isso ocorre porque a interpretação da imagem é realizada em paralelo pelo processo sequencial de leitura. As imagens também podem ser independentes da língua local. Um gráfico ou um mapa pode ser entendido por um grupo de pessoas com diferentes línguas (MATTHEW; GRINSTEIN; KEIM, 2003).

Existe uma crescente necessidade de ferramentas e técnicas para ajudar a fazer uso efetivo deste excesso de informações. Da mesma forma, há uma necessidade crescente de encontrar mecanismos para comunicar a informação às pessoas de

forma eficiente e eficaz, e para ajudar a informá-los sobre processos e conceitos que afetam a vida cotidiana, do aquecimento global à engenharia genética. Em praticamente qualquer domínio, a visualização pode ser, e está se tornando, uma ferramenta eficaz para auxiliar na análise e comunicação (MATTHEW; GEORGES; DANIEL, 2010).

Segundo Card *et al.* (1999), visualização de dados é informação abstraída de modo esquemático, incluindo atributos ou variáveis para unidades de informação. Em outras palavras é apresentar informações quantitativas com uso de recursos gráficos, transformando grandes e pequenos bancos de dados em recursos visuais que facilitam a percepção visual humana auxiliando os processos cognitivos.

Uma das melhores formas de explorar e tentar entender um grande conjunto de dados é pela visualização. Coloque os números em um espaço visual e deixe o cérebro dos leitores encontrar os padrões. As pessoas são boas nisso. Você normalmente encontra informações que pode nunca ter encontrado com apenas métodos estatísticos formais (YAU, 2011).

Além disso, o tempo tem um papel cada vez mais crítico no processo de tomada de decisão. Portanto, tornar a apresentação dos resultados visual, fácil de consumir, permitindo consultas, interação e exploração torna a tomada de decisão mais rápida e eficaz. Nesse sentido, o presente trabalho possui como objetivo principal apresentar utilizações de visualizações de dados, por meio do *Software R*. Como objetivo secundário pretende-se apresentar os resultados a um gestor com intenção de identificar o ganho de velocidade na tomada de decisão.

A divisão do trabalho está feita em cinco seções, sendo esta a primeira seção com a introdução do tema e apresentação dos objetivos. A segunda seção vai apresentar o que é visualização e seus conceitos, contextualizando o *Big Data*, como se estabelece o processo de visualização de dados e as novidades de uso da técnica. A terceira seção trata da metodologia, primeiramente apresentando o processo de visualização de dados e o *software R* que será utilizado para obter os resultados da quarta seção. A quarta seção conta com a aplicação de algumas técnicas de visualização, mostrando os passos da análise e os resultados obtidos. Na última seção constarão as conclusões deste trabalho e suas limitações.

2. REFERENCIAL TEÓRICO

2.1 Conceito de visualização de dados

A visualização é a apresentação gráfica da informação, com o objetivo de proporcionar ao espectador uma compreensão qualitativa dos conteúdos informativos. É também o processo de transformar objetos, conceitos e números em uma forma que é visível aos olhos humanos. Quando se diz "informação", pode-se referir a dados, processos, relações ou conceitos (MATTHEW; GRINSTEIN; KEIM, 2003).

Visualização de dados não se trata de entender individualmente os números, mas sobre compreender os padrões, tendências e relações que existem em grupos de dados. Do ponto de vista do analista, envolve detecção, medição e comparação, e é melhorada por meio de técnicas interativas que fornecem a informação de diferentes formas e visões (PARSAYE; CHIGNELL, 1993).

Segundo Matthew, Grinstein e Keim (2010), o processo de visualização se divide em cinco formas:

Modelagem de dados: Os dados a serem visualizados, seja a partir de um arquivo ou de um banco de dados, devem ser estruturados para facilitar sua visualização. O nome, tipo, intervalo e semântica de cada atributo ou campo de um registro devem estar disponíveis em um formato que garanta acesso rápido e de fácil modificação.

Seleção dos dados: A seleção de dados envolve a identificação do subconjunto dos dados que serão potencialmente visualizados. Isso pode ocorrer totalmente sob o controle do usuário ou através de métodos algorítmicos, como ciclos de tempo ou detecção automática de características de interesse do usuário.

Dados para mapeamentos visuais: O coração do *pipeline* de visualização é a realização do mapeamento de valores de dados para entidades gráficas ou seus atributos. Assim, um componente de um registro de dados pode mapear o tamanho de um objeto, enquanto outros podem controlar a posição ou a cor do objeto. Esse mapeamento geralmente envolve o processamento dos dados antes do mapeamento, como escalonamento, deslocamento, filtragem, interpolação ou subamostragem.

Ajuste de parâmetros visuais: Como nos gráficos tradicionais, o usuário deve especificar vários atributos da visualização que são relativamente independentes dos dados. Estes incluem a seleção do mapa de cores (para diferentes domínios, certas cores têm significado definido), seleção de mapa de som (caso os canais auditivos também forneçam informações) e especificações de iluminação (para visualizações em 3D).

Renderização ou geração da visualização: A projeção ou renderização específica dos objetos de visualização varia de acordo com o mapeamento usado; Técnicas como sombreado ou mapeamento de textura podem estar envolvidas, embora muitas técnicas de visualização requerem apenas linhas de desenho e polígonos uniformemente sombreados. Além de mostrar os dados em si, a maioria das visualizações também inclui informações suplementares para facilitar a interpretação, como eixos, chaves e anotações.

A visualização geralmente é parte de um processo maior, que pode ser a análise exploratória de dados, descoberta de conhecimento, apresentação de resultados de técnicas ou análise visual. Visualização e análise andam lado a lado, com o objetivo de construir um modelo que representa ou aproxima os dados. A visualização na exploração de dados é usada para transmitir informações, descobrir novos conhecimentos e identificar estruturas, padrões, anomalias, tendências e relacionamentos (MATTHEW; GRINSTEIN; KEIM, 2010).

2.2 *Big Data*

Big Data é um termo utilizado para descrever grandes volumes de dados e ganha cada vez mais relevância à medida que a sociedade se depara com um aumento sem precedentes no número de informações geradas a cada dia (IBM GROUP, 2014).

O termo *Big Data* já foi utilizado em diversos congressos e eventos especializados de diferentes áreas, porém sem muita precisão quanto a sua definição. Ele tem sido utilizado para descrever: i) os fenômenos sociais, como a mudança de comportamento gerada pela informação em tempo real; ii) as características específicas dos dados, como grandes volumes gerados rapidamente, e de fontes e

formas variadas; iii) as técnicas de análises e de armazenamento, que precisam ser pensadas e planejadas de modo a superar limites das técnicas utilizadas hoje em dia; e, ainda, iv) as necessidades tecnológicas, como dispositivos de captura e de processamento. A propagação rápida desse conceito tem levado a sua ampliação sem um núcleo claro de significado de o que é *Big Data* (MAURO; GRECO; GRIMALDI, 2016).

Diebold (2012) realizou uma busca pela origem etimológica do termo sem chegar a conclusões definitivas. Contudo, em meados dos anos 1990, apontou possíveis origens do uso do termo com o sentido que se conhece hoje, que é descrever soluções para a armazenagem e processamento de grandes volumes de dados.

A consolidação do fenômeno do *Big Data* enquanto atributo dos dados veio no início dos anos 2000 com o relatório de Douglas Laney (2001), em que estuda os desafios que o incremento do *e-commerce* trouxe para o mercado de gestão de dados, e afirma que as condições e meios de negócios atuais estão empurrando os princípios tradicionais da gestão de dados para os seus limites, fazendo surgir novas abordagens.

Os desafios da gestão de dados mencionados pelo autor são apresentados em três dimensões: Volume, Velocidade e Variedade. Laney (2001) não cunhou nem utilizou o termo *Big Data*, mas estabeleceu seus fundamentos: os “3 vês”, que são referidos constantemente como núcleo do conceito de *Big Data*. Sua caracterização tornou-se, assim, referência (MAURO, 2014; DIEBOLD, 2012).

Nesse cenário, para a Gartner Group (2012), *Big data* é definido como ativos de alto volume, velocidade e variedade de informação que exigem custo-benefício, de formas inovadoras de processamento de informação para maior visibilidade e tomada de decisão.

Mesmo assim, o desconhecimento desse conceito, a não estruturação e um padrão dos dados e a falta de uso completo das ferramentas de análise ainda é o cenário vivido pela maioria das empresas. Dessa forma, parece muito distante para a maioria dos gestores das organizações a tomada de decisão com as informações

certas. Segundo Ferguson (2013), em uma pesquisa realizada pela empresa de consultoria McKinsey, somente 65% das organizações são eficientes na disseminação da informação. Esses dados ainda são mais preocupantes quando o autor diz que apenas 4% das organizações usam todos os dados que coletam.

2.3 Percepção Humana

Ver e entender imagens é um dos instintos naturais do ser humano, já para entender dados numéricos são anos de treinamento para desenvolver essa habilidade nas escolas, e mesmo assim, muitas pessoas ainda não são boas com dados numéricos (PARSAYE; CHIGNELL, 1993).

Em um gráfico bem desenhado, é muito mais fácil encontrar as tendências e relações, porque a apresentação visual da informação tira proveito da capacidade vasta, e frequentemente subutilizada, do olho humano para detectar a informação das imagens e das ilustrações. A visualização dos dados desloca a carga do raciocínio numérico para o raciocínio visual. Obtenção de informações através de imagens é muito mais rápida do que olhar através de texto e números - é por isso que muitos decisores preferem ter informações apresentadas a eles em forma gráfica, em oposição a uma forma textual (YAU, 2000).

A visualização de dados é uma forma de apresentar e exibir informações de uma forma que incentive a interpretação, seleção e associação. Utiliza habilidades humanas para reconhecimento de padrões e análise de tendências e explora a capacidade das pessoas de extrair uma grande quantidade de informações em um curto período de tempo a partir de imagens apresentadas em um formato padronizado (YAU, 2000).

Por meio do sistema visual, a percepção humana desempenha um papel importante na área da visualização, auxiliando os processos cognitivos. Assim, considerar fatores da percepção visual humana no desenvolvimento de ferramentas computacionais de visualização de dados complexos e elevada dimensão, assume um papel fundamental com grande interesse científico e até para o público em geral (MATTHEW; GRINSTEIN; KEIM, 2010).

Já Robertson, Card e Mackinlay (1993), define o conceito de visualização de forma mais específica como sendo o uso de representações visuais de dados abstratos, suportadas por computador e interativas para ampliar a cognição. Destas definições entende-se que uma visualização pode funcionar como uma ferramenta cognitiva para a construção de conhecimento utilizando as capacidades perceptivas e cognitivas do ser humano.

Informações representadas graficamente, por outro lado, tendem a ser processadas de maneira mais automática pela visão, em um processo mais superficial, paralelo (no sentido de obter várias informações simultaneamente), rápido e de capacidade elevada. Dessa forma, representar graficamente os dados a serem analisados é interessante do ponto de vista da obtenção de informação, pois faz com que não apenas mecanismos computacionais sejam usados para a análise de dados, mas também recursos da visão e da cognição humana (MATTHEW; GRINSTEIN; KEIM, 2010).

2.4 Visualizar Padrões ao Longo do tempo

Ao analisar dados orientados pelo tempo, os usuários estão geralmente interessados na evolução de seu comportamento ao longo do tempo. Para atingir esse objetivo, a tarefa principal dos usuários é comparar dados localizados em diferentes posições do eixo do tempo. Detectar tendências e padrões são objetivos de segunda ordem que levam à percepção e à compreensão dos dados (WOLFGANG, *et al.* 2007).

De acordo com Yau (2011), o mais comum que se procura em séries temporais são tendências nos dados. Alguma coisa está aumentando ou diminuindo? Existem ciclos sazonais? Para encontrar esses padrões, deve-se olhar para além dos pontos de dados individuais para obter uma imagem do todo.

Muitos tipos diferentes de dados estão relacionados ao tempo. Pode-se pensar em dados da bolsa de valores, dados do censo, dados de simulação, entre outros. Mas também artigos de notícias, coleções de fotos ou planos de projetos podem conter informações temporais. Do ponto de vista teórico, todos esses dados são orientados para o tempo e devem ser representáveis com uma mesma abordagem. (MATTHEW; GRINSTEIN; KEIM, 2010).

Segundo (Matthew; Grinstein; Keim, 2010), a comunicação da dependência dos dados no tempo requer principalmente adequar uma boa colocação do eixo do tempo. A grande variedade de técnicas visuais expõe sobre abordagens muito diferentes. Os detalhes sutis desta variedade são descritos em dois critérios fundamentais:

Mapeamento do tempo: Existem duas opções para o mapeamento do tempo: o mapeamento de tempo no espaço e o mapeamento de tempos em tempos. Ao falar de um mapeamento de tempo no espaço, quer dizer que o tempo e os dados são apresentados em uma única representação visual. Essa representação não se altera automaticamente ao longo do tempo, razão pela qual é chamado de visualizações de dados orientados para tempo estático. Por outro lado, as representações dinâmicas utilizam a dimensão física do tempo para transmitir a dependência dos dados ao longo do tempo. Isso resulta em visualizações que mudam automaticamente ao longo do tempo. Na Figura 1 pode-se observar sugestões para o eixo tempo.

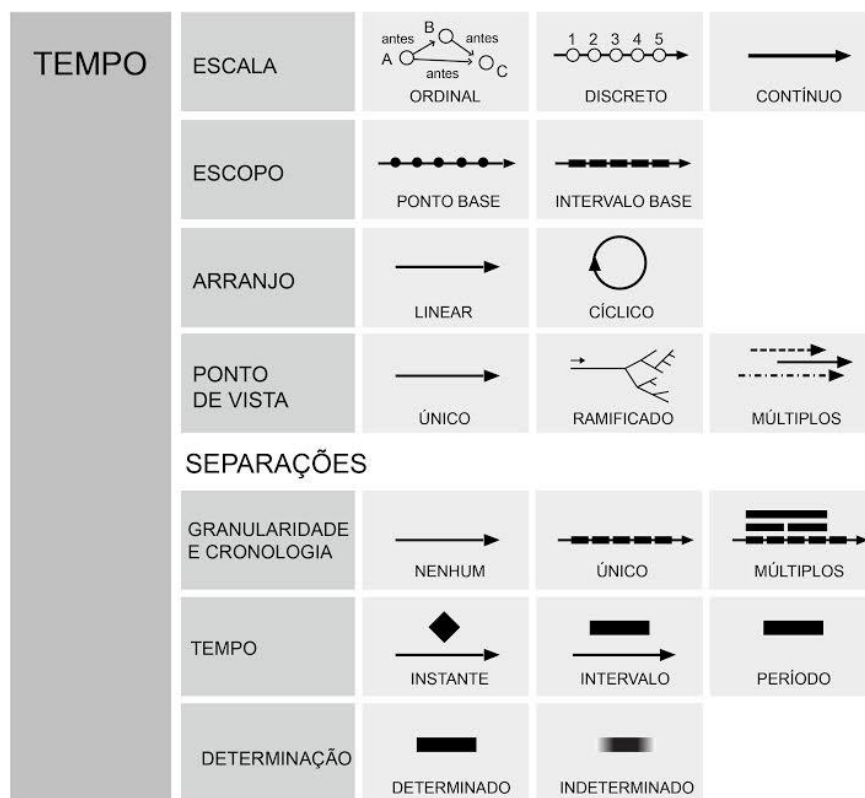


Figura 1 – Eixo tempo para gráficos

Fonte: MATTHEW; GEORGES; DANIEL, 2010

Dimensionalidade de apresentação do espaço: Pode-se diferenciar as apresentações 2D e 3D de dados orientados para o tempo. As abordagens de visualização que usam

um espaço de apresentação 2D têm que garantir que o eixo do tempo seja enfatizado, pois as dimensões de tempo e dados muitas vezes têm que compartilhar as duas dimensões de exibição disponíveis. No caso de representações em 3D, uma terceira dimensão de exibição é alocável. Na verdade, muitas técnicas o utilizam como uma dimensão dedicada para o eixo do tempo, separando claramente o tempo de outras dimensões.

2.5 Visualização de Dados Multivariados

A visualização de dados multivariados, como um tipo específico de visualização de informação, é um campo de pesquisa ativo com inúmeras aplicações em diversas áreas, desde comunidades científicas e design de engenharia até indústria e mercados financeiros, nas quais as correlações entre muitos atributos são de interesse vital (FRIENDLY, 2000).

Abaixo são listados alguns tópicos que deve-se observar para construção de gráficos multivariados.

Mapeamento: Encontrar um mapeamento adequado de dados multivariados de alta dimensão em uma forma visual 2D nunca é uma tarefa simples. Geralmente, depende da natureza dos conjuntos de dados para serem visualizados e está mais relacionado à percepção humana. Além disso, a associação de atributos de dados a entidades gráficas exige extrema cautela para evitar superar a capacidade de visualização do observador. A junção de vários elementos nas representações pode induzir a sobrecarga de cognição dos usuários e, portanto, os atributos gráficos devem ser cuidadosamente selecionados, de modo que sejam de fácil entendimento. É importante que os diferentes atributos possam ser vistos de forma holística e, ao mesmo tempo, cada dimensão possa ser julgada pelos usuários separadamente e de forma independente (HEALEY, 2005).

Dimensionalidade: Os dados multivariados são muitas vezes de grande tamanho e alta dimensionalidade, resultando em uma estrutura densa. Portanto, é difícil apresentar esses dados em uma única exibição visual, o que torna desafiador permitir que os usuários explorem o espaço de dados de forma intuitiva e interativa, bem como discriminando dimensões individuais. Além disso, o ordenamento das dimensões tem um grande impacto na capacidade da visualização (KEIM, 2000).

Design Tradeoffs: A visualização pode fornecer uma visão geral qualitativa de conjuntos de dados grandes e complexos para que os usuários possam procurar estrutura, recursos, padrões, tendências e relacionamentos de forma mais eficaz (GRINSTEIN; WARD, 2001). Devido à alta dimensionalidade dos dados multivariados, inevitavelmente sacrifica-se a capacidade de mostrar os detalhes de cada atributo, pois há menos atributos gráficos para codificação. Esta situação pode não ser harmonizada quando é necessária uma análise quantitativa. Para a visualização de dados multivariados, existe sempre uma compensação entre quantidade de informação, simplicidade e precisão.

Avaliação da Eficácia: O objetivo final da visualização de dados multivariados é obter informações sobre os dados e mostrar a possível correlação entre os diferentes atributos. Na maioria dos casos, certas correlações ainda não foram descobertas antes de olhar para a exibição visual, e eles são exatamente o que se quer obter após a visualização. Segundo Wijk (2006), é um paradoxo que proíbe a avaliação da eficácia de uma técnica de visualização de informação. Não se sabe o conhecimento valioso presente nos dados, por isso espera-se obter informações ao visualizá-lo. No entanto, se não se sabe nada sobre o padrão ou relação a ser mostrado na representação de dados, nunca se pode avaliar a eficácia de uma técnica particular de visualização.

2.6 Conceitos de Interação

A interação dentro do contexto de visualização de dados e informações é um mecanismo para modificar o que os usuários vêem e como eles o vêem. Existem muitas classes de técnicas de interação (YI, *et al.* 2007), tais como:

Navegação: O usuário tem o controle para alterar a posição da câmera e para dimensionar a vista (o que é selecionado para a tela), como deslocamento, rotação e zoom.

Seleção: Controles de usuário para identificar um objeto, uma coleção de objetos ou regiões de interesse para ser objeto de alguma operação, como destaque, exclusão e modificação.

Filtragem: Controles do usuário para reduzir o tamanho dos dados sendo apresentados na tela, eliminando registros, dimensões ou ambos.

Reconfiguração: Controles de usuário para alterar a forma como os dados são apresentados para entidades ou atributos gráficos, como reordenar os dados ou *layouts*, fornecendo assim uma maneira diferente de visualizar um subconjunto de dados.

Codificação: Controles de usuário para alterar os atributos gráficos, como o tamanho do ponto ou a cor da linha, para potencialmente revelar diferentes recursos.

Conexão: Controles de usuário para acionar diferentes visualizações ou objetos para mostrar itens relacionados.

Resumo/elaboração: Controles de usuário para modificar o nível de detalhe.

Híbrido: Controles de usuário combinando vários dos itens acima em uma técnica, por exemplo, aumentando o espaço de tela atribuído a uma ou mais áreas de foco para permitir que os usuários vejam detalhes, enquanto mostra as outras áreas de dados em um espaço menor, de uma maneira que preserva o contexto.

3. METODOLOGIA

O estudo foi estruturado com base nas etapas de criação de visualização, sugerido por Kaid (2011), que propõe a partir da obtenção dos dados aplicar transformações numéricas com objetivo de ajustar as escalas das visualizações dentro do escopo de apresentação, utilizar técnicas de análise de dados, criar representações gráficas e criar a opção de interagir com a visualização. Na Figura 2 são apresentadas as etapas para construção de visualização de dados, que não sequência são detalhadas.

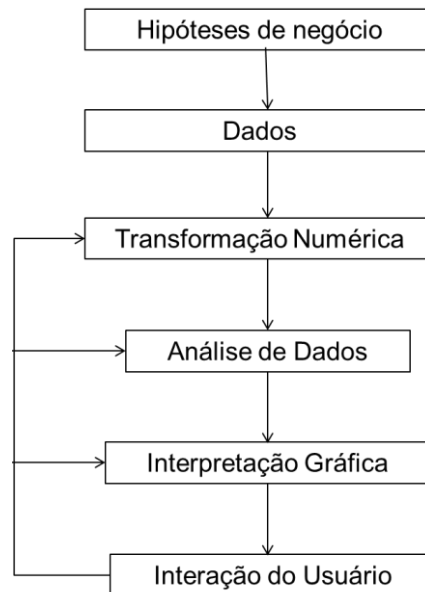


Figura 2 – Etapas para construção de visualização de dados

Fonte: Adaptado de Kaid (2011)

3.1 Hipóteses de Negócio

Esta etapa consiste em elaborar as hipóteses de negócio que se deseja responder. Pode-se começar com uma lista de problemas, identificar novas oportunidades ou até mesmo rediscutir decisões já estabelecidas.

3.2 Dados

Nesta etapa, uma vez decididas as hipóteses de negócio, deve-se definir quais variáveis e indivíduos entrarão no conjunto de dados. É importante compreender os dados, incluindo seu tamanho e cardinalidade. Além disso, sugere-se cuidar o esforço necessário para a preparação dos dados, para que nenhuma informação seja esquecida ou que informação desnecessária seja incluída.

3.3 Transformação Numérica

Segundo Kaid (2011), a visualização é uma espécie de transformação de dados numéricos em representação como pontos e linhas, e para isso requer uma transformação. As transformações incluem: (i) alterar a distribuição dos dados: modificar a distribuição para que eles sejam mais adequados para análise ou apresentação visual. Algumas transformações frequentemente utilizadas são:

transformação linear, logarítmica, normalização, arco seno, raiz quadrada, inversa; (ii) redefinir o contexto: ajuste dos números de forma que sejam mais significativos ou mais representativos dentro do contexto que o analista esteja interessado. Por exemplo, evidenciar mudanças repentinas de comportamento e grandes variações.

3.4 Análise de Dados

A análise de dados é o processo de aplicação de vários métodos aos dados para auxiliar na interpretação. Alguns dos métodos de análise de dados exploratórios mais utilizados são: métodos descritivos, análise de agrupamento, dimensionamento multidimensional e análise fatorial. A análise de dados pode ser utilizada para transformar dados ou para resumir os dados em si.

3.5 Interpretação Gráfica

A interpretação gráfica consiste em algumas atividades-chave, como compreender a magnitude, proporções, tendências, inclinações e agrupamentos. Também pode-se usar algumas formas, tais como:

- Usar escalas e deslocamento para ajustar ao escopo.
- Usar valores derivados (resíduos, *logs*) para enfatizar as mudanças.
- Usar projeções e combinações para compactar informações e obter estatísticas.
- Usar formatos diferentes para separar sobreposições. Utilizar várias visualizações para lidar com relações ocultas e alta dimensionalidade.
- Usar conexões efetivas, chaves e rótulos para auxiliar o entendimento

3.6 Interação do Usuário

Além da técnica de visualização, para uma efetiva exploração dos dados é necessário usar uma ou mais técnicas de interação. Segundo Daniel e Matthew (2002), as técnicas de interação permitem que o analista de dados interaja diretamente com as visualizações e mude dinamicamente as visualizações de acordo com os objetivos de exploração. Além disso, eles também permitem relacionar e combinar várias visualizações independentes.

As técnicas de interação podem ser categorizadas com base nos efeitos que eles exibem. As técnicas de navegação se concentram na modificação da projeção dos dados na tela, usando métodos manuais ou automatizados. Os métodos de aprimoramento de exibição permitem aos usuários ajustar o nível de detalhe em parte ou a totalidade da visualização ou modificar o mapeamento para enfatizar algum subconjunto dos dados. As técnicas de seleção fornecem aos usuários a capacidade de isolar um subconjunto dos dados exibidos para operações como o destaque, a filtragem e a análise quantitativa. A seleção pode ser feita diretamente na visualização (manipulação direta) ou através de caixas de diálogo ou outros mecanismos de consulta (manipulação indireta). Alguns exemplos de técnicas de interação são: Ajuste dinâmico do mapeamento, Filtragem interativa, *Zooming*, *Distorção*, *Brushing* e *Linking* (KEIM; WARD, 2002).

3.7 Avaliação do Método

A fim de ter uma percepção inicial sobre o método, foi aplicado no trabalho uma entrevista semi-estruturada construída a partir das hipóteses de negócios. Desta forma, estruturou-se o questionário com 5 perguntas dissertativas visando a objetividade da entrevista.

Para validação do método foi selecionado um profissional com experiência em *big data* e visualização de dados. O objetivo é mostrar a visão desse gestor sobre visualização de dados, além da importância de utilizar recursos visuais para comunicar dados e auxiliar a tomada de decisão.

3.8 O Software

O *software* utilizado para esse trabalho foi o R versão 3.3.2, sendo escolhido por ser um *software* que contém recursos para criação de análises visuais, por se gratuito e por conter um extenso material e comunidade ativa. O *Software* R pode ser baixado diretamente na internet pelo site '<https://www.r-project.org/>'. Além disso, o *Software* R tem amplo conjunto de funções e pode ser aperfeiçoado com o uso de novos pacotes, sendo assim, é um programa bastante poderoso quanto a análises estatísticas.

O pacote '*ggplot2*' possui as principais funções utilizadas para visualização de dados. Também foi utilizado o pacote "*plotly*" que é uma das melhores ferramentas de

visualização de dados disponíveis, construído com base na biblioteca de visualização *D3.js*, *HTML* e *CSS*. Ele é criado usando o *Python* e o *framework Django*.

4. RESULTADOS

Nesta seção apresentam-se resultados obtidos em cada passo descrito na metodologia descrita na seção 3.

4.1 Hipóteses para Problemas de Negócio

O objetivo desta seção é estruturar questões a serem respondidas tendo em vista os dados de vendas para 1559 itens em 10 lojas da rede *Big Market*. O objetivo é criar hipóteses que auxiliem a construir visualizações a fim de responder questões relacionadas às vendas de cada item em determinada loja. Utilizando essas visualizações se tentará entender as propriedades de itens e lojas que são relevantes para o aumento das vendas.

A fim de facilitar o entendimento, dividiu-se as hipóteses de negócio em dois grupos. Hipóteses de negócio relacionadas as lojas (as duas primeiras) e outras relacionadas aos itens (as duas últimas). O negócio em questão trata-se de informações de vendas de itens em lojas com diferentes localizações e tamanhos.

Tipo de cidade: lojas localizadas em cidades urbanas ou de nível 1 devem ter vendas mais elevadas devido aos maiores níveis de renda das pessoas de lá.

Tamanho das lojas: lojas que são maiores em tamanho possuem vendas mais elevadas à medida em que as pessoas prefeririam obter todos os tipos de itens em um lugar só.

Área de exibição: os itens que recebem prateleiras maiores na loja provavelmente atrairão a atenção primeiro e venderão mais.

Ofertas promocionais: os itens com ofertas mais atrativas de descontos venderão mais.

4.2 Dados

As informações utilizadas são de uma base de dados disponível no site <https://www.analyticsvidhya.com>. O conjunto de dados possui quatro variáveis numéricas contínuas: peso do item, percentual da área de exibição, preço máximo de varejo do item e receita total do item. Além disso, há também variáveis categóricas: categoria do item, tipo de cidade, tipo de loja e categoria de peso do item. A Figura 3 mostra um resumo descrevendo as variáveis. São 12 variáveis de 1559 itens.

Variáveis	Descrição
<i>Item_Identifier</i>	Id do item
<i>Item_Weight</i>	Peso do produto
<i>Item_Fat_Content</i>	Categoria de produto: leve ou pesado
<i>Item_Visibility</i>	Percentual da área de exibição de um produto na loja
<i>Item_Type</i>	Categoria de cada produto pertence
<i>Item_MRP</i>	Preço máximo de varejo do produto
<i>Outlet_Identifier</i>	Id da loja
<i>Outlet_Establishment_Year</i>	Ano de lançamento da loja
<i>Outlet_Size</i>	Tamanho da loja em relação a área coberta
<i>Outlet_Location_Type</i>	Tipo de cidade em que a loja está localizada
<i>Outlet_Type</i>	Tipo de loja: Merceria, supermercado, etc
<i>Item_Outlet_Sales</i>	Receita total por produto em determinada loja

Figura 3 – Descrição das variáveis

4.3 Transformação Numérica

Para a variável categoria do peso do item, foi realizada uma nova categorização com base nas regras apresentadas na Figura 4. Assim, os itens que apresentassem peso até 9 serão incluídos na categoria leve, acima de 9 e menor ou igual a 12,8 entra para a categoria regular e acima de 12,8 é classificado como pesado. Essa transformação foi necessária, pois a variável original possui apenas duas categorias (leve e regular) e pode-se haver uma relação com o percentual de visibilidade do item na loja.

peso do item \leq 9	Leve
9 > peso do item \leq 12,8	Regular
peso do item > 12,8	Pesado

Figura 4 – Critérios para nova categorização do peso do item.

4.4 Análise de dados

O intuito desta seção é explicar brevemente como será construída a estratégia de análise, tendo em vista que o objetivo principal é construir visualizações com métodos estatísticos que forneçam informações entre vendas dos itens das lojas, confrontado as características dos itens e das próprias lojas, relacionados com as quatro hipóteses apresentadas na seção 4.1. As análises apresentadas foram realizadas por meio do *Software R*.

Primeiro, para hipótese de tamanho de lojas utilizou-se um *boxplot*, devido a características das variáveis e para comparar o total de receita entre os diferentes tamanhos de lojas. Na Figura 5 tem-se a interpretação gráfica da análise.

Para a hipótese para o tipo de cidade utilizou-se também um *boxplot*, devido a característica das variáveis e para comparar o total de receita entre os diferentes tipos de lojas. Na Figura 6 tem-se a interpretação gráfica da análise.

Para a hipótese da área de exibição dos itens, tem-se duas variáveis contínuas: percentual da área de visibilidade do item. Utilizou-se um *scatterplot* para representar a relação entre essas variáveis. Na Figura 7 tem-se a representação gráfica da análise.

Para a hipótese de ofertas promocionais, assim como para análise da área de exibição de itens, utilizou-se um *scatterplot*. As variáveis utilizadas foram: preço máximo de varejo do item e receita total de item. Na Figura 8 tem-se a representação gráfica da análise.

4.5 Interpretação Gráfica

A Figura 5 apresenta uma paleta de cores que é um grupo de cores que foi utilizado para tornar o gráfico mais atraente e ajuda a criar distinções visuais nos dados, neste caso, entre as categorias de tamanho de loja. A interpretação mais intuitiva é que não existe uma diferença aparente entre as medianas de receita entre

os diferentes tamanhos de loja e verifica-se a presença de valores de pontos extremos (*outliers*), para cada tamanho de loja.

O *boxplot* apresentado na Figura 6 representa a receita total de item para cada tipo de cidade. Analisando visualmente não se percebe uma diferença aparente entre a mediana da receita total por tipo de cidade, demonstrando que as distribuições do total de receita nos três tipos de cidades parecem as mesmas.

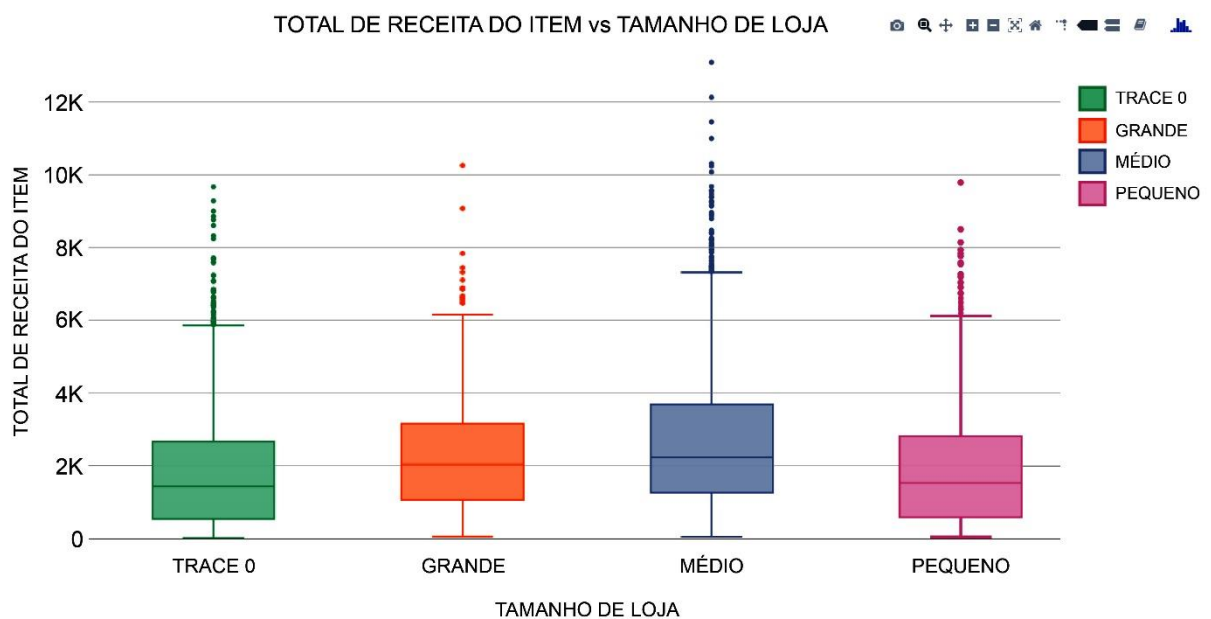


Figura 5 – Boxplot das variáveis Total de Receita do Item e Tamanho de Loja

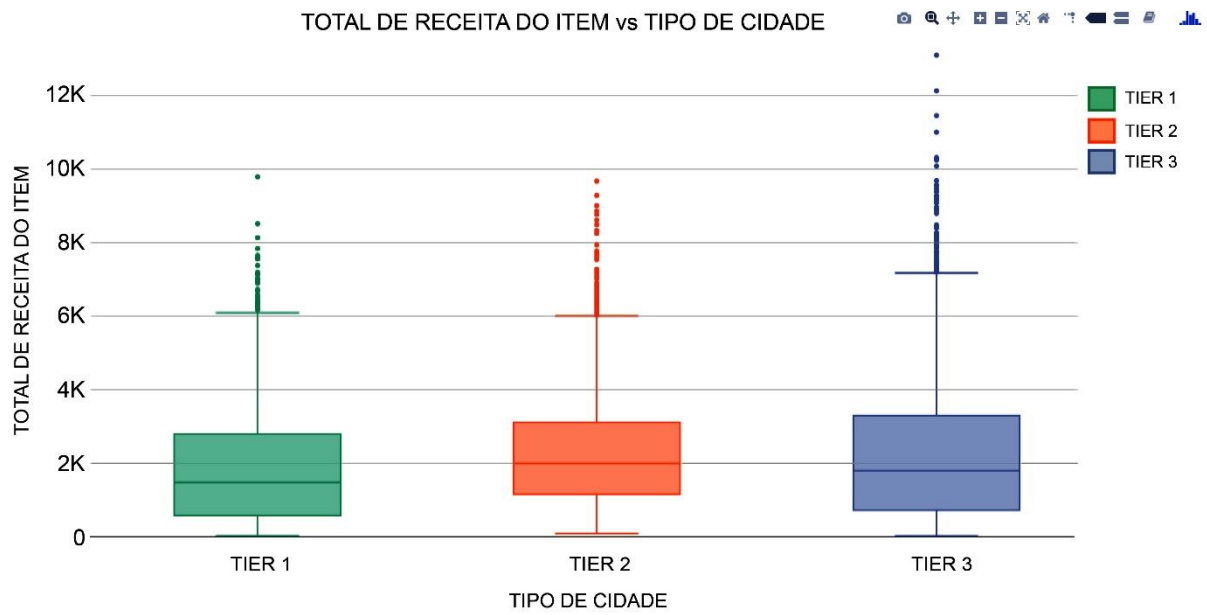


Figura 6 – Boxplot das variáveis Total de Receita do Item e Tipo de Cidade

A Figura 7 apresenta o *scatterplot* das variáveis total de receita e percentual de visibilidade do item. Percebe-se que não existe uma relação clara entre a receita total e aumento percentual de visibilidade do item na loja.

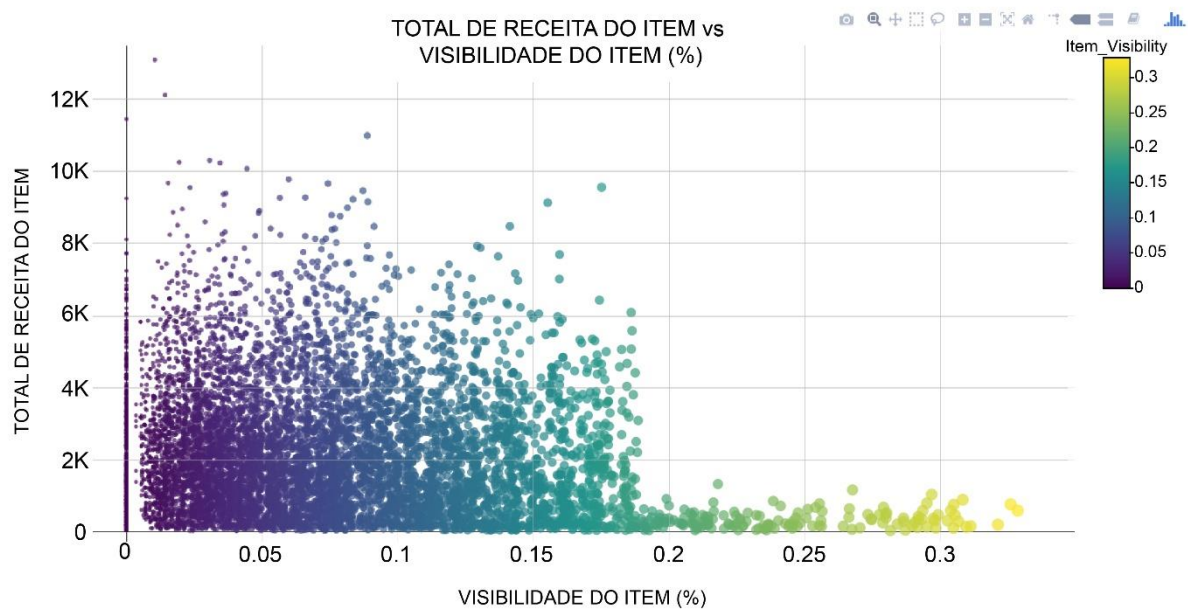


Figura 7 – *Scatterplot* das variáveis Total de Receita por Item e Percentual de Visibilidade do Item na Loja.

A Figura 8, a seguir, apresenta o *scatterplot* resultante das variáveis total de receita e preço máximo de varejo do item. Percebe-se, intuitivamente, que existe uma relação entre o aumento da receita total e o preço máximo de varejo do item.

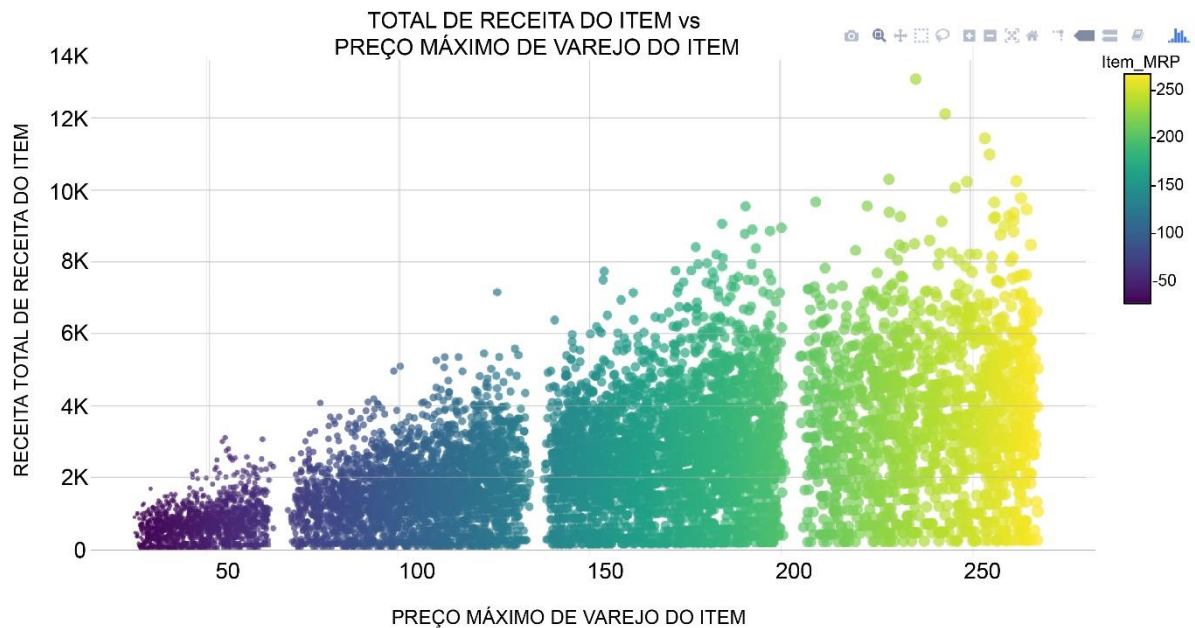


Figura 8 – *Scatterplot* das variáveis Total de Receita do Item e Preço Máximo de Varejo do Item

4.6 Interação do Usuário

Nesta seção, serão apresentadas as interações que o usuário pode realizar com a visualização. O pacote “*plotly*” permite ao usuário interagir utilizando as seguintes funções: zoom, redefinir a escala, voltar a seleção inicial, mostrar as coordenadas, *mouseover* para exibir os dados, seleção de dados e destacar seleção.

Através do clique na opção desejada é possível realizar determinada ação no gráfico. Na Figura 9 tem-se destacado o menu com as opções de interação com o gráfico e as estatísticas da loja de tamanho pequena. Esta opção de interação ajuda o usuário a observar os dados obtidos através da análise além da visualização.

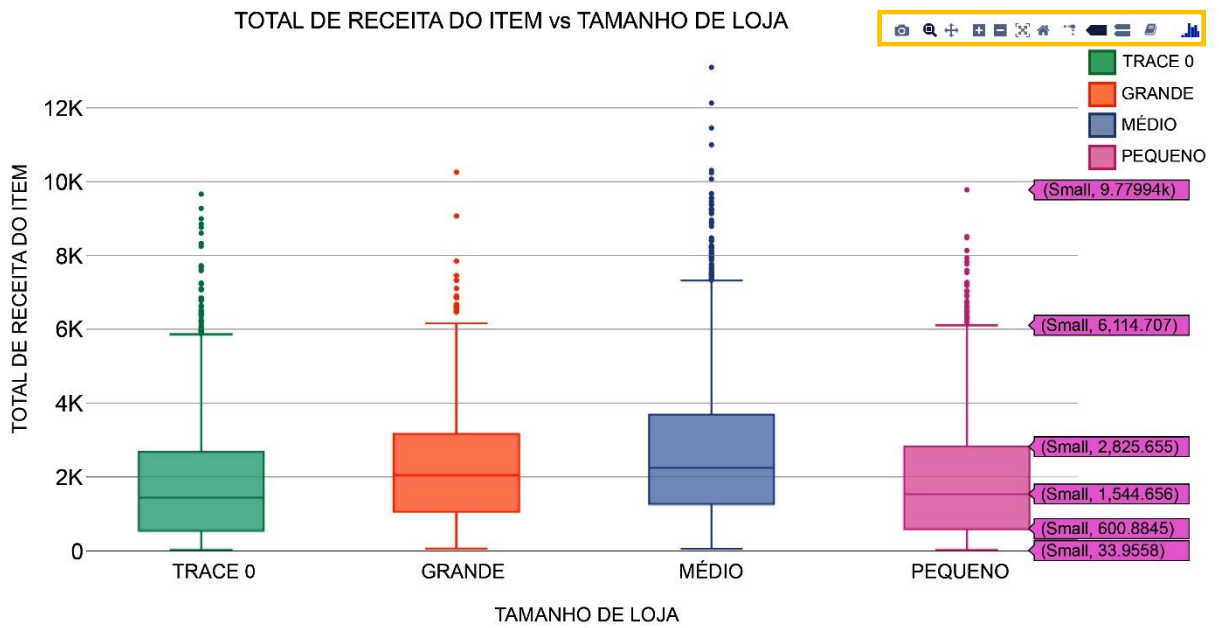


Figura 9 – Boxplot das variáveis Total de Receita do Item e Tamanho de Loja com *mouseover*

Na Figura 10 foi utilizado a opção de seleção para destacar os pontos no gráfico que apresentam uma distribuição diferente dos demais. Pode-se observar os pontos que apresentam visibilidade acima de 20% e o total de receita.

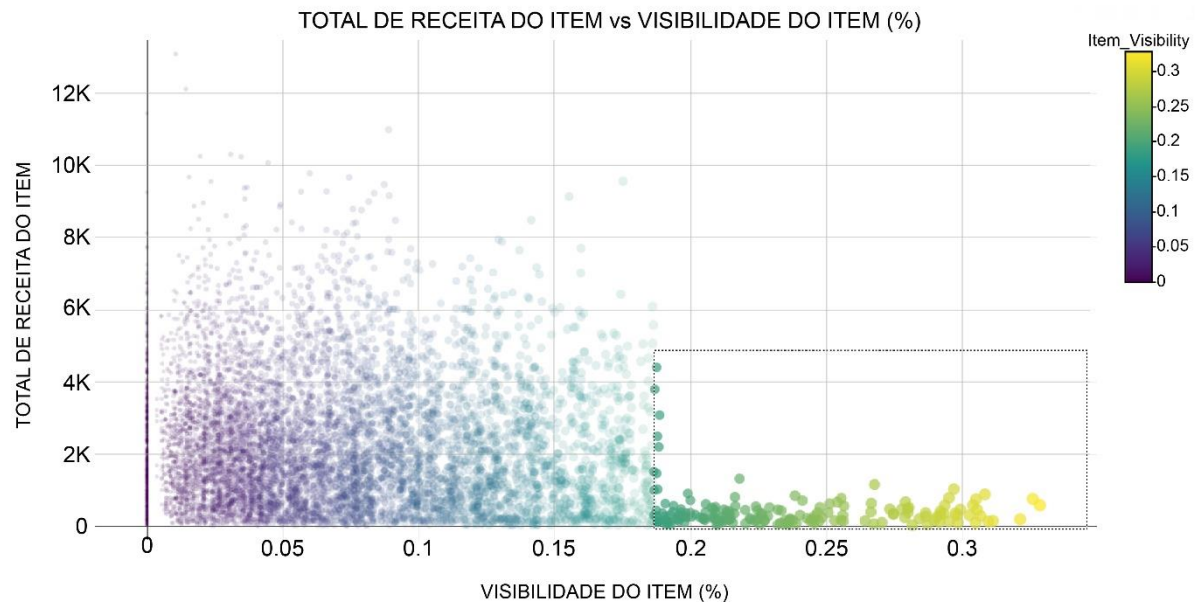


Figura 10 – *Scatterplot* das variáveis Total de Receita do Item e Percentual de Visibilidade do Item na loja.

Na Figura 11 pode-se observar o ponto destacado com as coordenadas nos eixos e as linhas que conectam cada eixo. Essa funcionalidade auxilia a identificar o ponto que se destaca no gráfico.

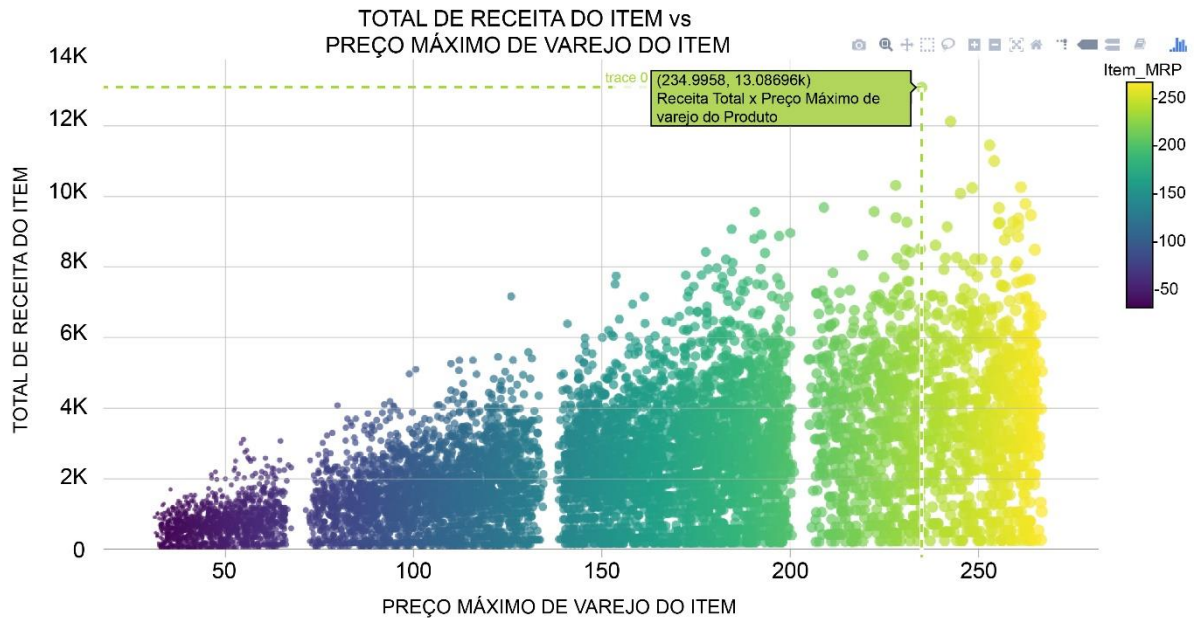


Figura 11 – Scatterplot das variáveis Total de Receita do Item e Preço Máximo de Varejo do Item.

Abaixo, na Figura 12, tem-se a representação da funcionalidade do Zoom. Selecionou-se uma parte dos pontos do gráfico e na sequência é apresentado apenas os pontos selecionados.

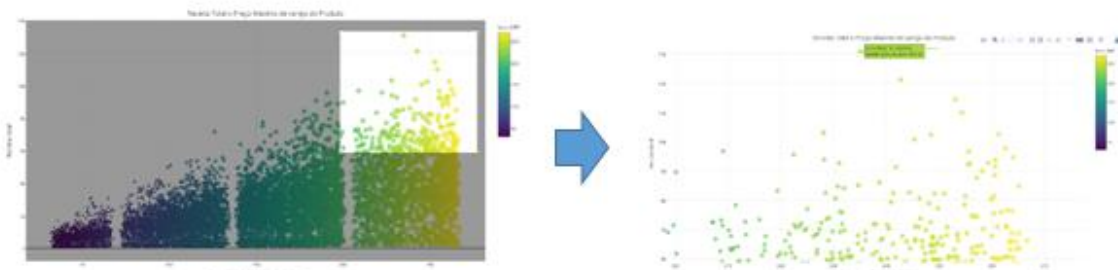


Figura 12 – Scatterplot das variáveis Total de Receita do Item e Preço Máximo de Varejo do Item com Zoom.

4.7 Avaliação Qualitativa

Com objetivo de ter uma avaliação inicial da percepção de um gestor ao utilizar as visualizações criadas com base nas análises realizadas e apresentadas na seção 4.1. Em seguida, foram feitas perguntas visando a objetividade da entrevista na aplicação do trabalho.

Para validação desse método, foi selecionado um profissional que trabalha com Big Data diariamente. É formado em publicidade, com mais de 10 anos de experiência na área e gestor da área de business intelligence & Insights de uma importante companhia. Foi escolhido por ter contato direto com o tema e com o cenário de comunicação visual de dados para o negócio.

Comparou-se os resultados com tabelas a fim de propor perspectivas diferentes sobre a mesma análise, conforme apresentado na Figura 13. Essa etapa teve duração de aproximadamente 30 minutos, em que foram apresentadas as visualizações e foram ouvidas sugestões de mudanças e comentários relativos às dificuldades de interpretação dos resultados.

A análise da entrevista está dividida entre análise geral da percepção sobre os resultados, que diz respeito a perguntas sobre as hipóteses de negócio, e a análise sobre as vantagens na comunicação dos números através das visualizações, a fim de obter uma percepção qualitativa sobre o tema.

Rótulos de Linha	Tier 1	Tier 2	Tier 3	(vazio)	Total Geral
Supermarket Type1	R\$ 4.302.364,98	R\$ 6.472.313,70	R\$ 2.142.663,57		R\$ 12.917.342,24
Supermarket Type3			R\$ 3.453.926,04		R\$ 3.453.926,04
Supermarket Type2			R\$ 1.851.822,83		R\$ 1.851.822,83
Grocery Store (vazio)	R\$ 179.694,09		R\$ 188.340,17		R\$ 368.034,27
Total Geral	R\$ 4.482.059,07	R\$ 6.472.313,70	R\$ 7.636.752,61		R\$ 18.591.125,38

Rótulos de Linha	Tier 1	Tier 2	Tier 3	(vazio)	Total Geral
Supermarket Type3	R\$ 2.313,10	R\$ 2.323,99	R\$ 3.694,04		R\$ 3.694,04
Supermarket Type1			R\$ 2.239,00		R\$ 2.239,00
Supermarket Type2			R\$ 1.995,50		R\$ 1.995,50
Grocery Store (vazio)	R\$ 340,33		R\$ 339,35		R\$ 339,83
Total Geral	R\$ 1.976,91	R\$ 2.323,99	R\$ 2.279,63		R\$ 2.181,29

Rótulos de Linha	Tier 1	Tier 2	Tier 3	(vazio)	Total Geral
Baking Goods	7,23%	6,36%	7,16%		6,92%
Breads	6,33%	6,32%	6,68%		6,63%
Breakfast	9,88%	8,35%	7,32%		8,57%
Canned	7,21%	6,50%	6,74%		6,81%
Dairy	7,92%	6,37%	7,42%		7,24%
Frozen Foods	6,88%	6,13%	6,70%		6,56%
Fruits and Vegetables	7,28%	6,59%	6,77%		6,85%
Hard Drinks	6,38%	6,26%	6,77%		6,49%
Health and Hygiene	5,72%	5,27%	5,58%		5,32%
Household	6,78%	5,48%	6,22%		6,13%
Meat	6,76%	4,97%	6,75%		6,23%
Others	6,26%	5,42%	6,32%		6,02%
Seafood	10,01%	5,08%	7,74%		7,50%
Snack Foods	7,42%	8,24%	8,54%		8,89%
Soft Drinks	7,09%	5,89%	6,37%		6,40%
Starchy Foods	6,97%	5,81%	7,47%		6,76%
Total Geral	7,12%	6,10%	6,68%		6,61%

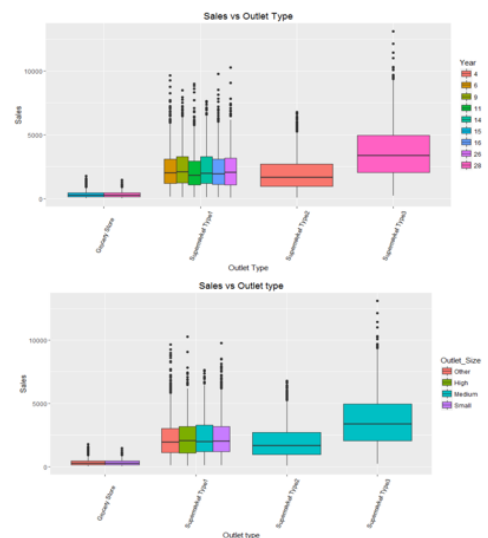


Figura 13 – Comparação entre tabelas e gráficos

Inicialmente, a entrevista abordou os resultados obtidos sobre as hipóteses de negócio comparando diferentes formas de exibir a mesma informação. Diante disto, a percepção do gestor é que de forma geral o tempo de resposta é mais rápido quando se utilizou representações gráficas, principalmente, por concentrar volume maior de informações em um espaço menor.

Além disso, o entrevistado comentou que a visualização de dados auxilia a aprofundar o conhecimento sobre o conjunto de dados, mas principalmente ajuda a explorar e descobrir padrões que o usuário poderia não estar enxergando com os dados em formato usual de tabela. Por outro lado, a visualização de dados não elimina a necessidade de observar os dados em uma tabela, por exemplo, para obter resultados mais precisos, principalmente no momento de tomar a decisão.

O gestor também comentou sobre o esquema de cores, legendas e sumarização da informação deve estar bastante claro e objetivo para que seja intuitiva a interpretação. Por outro lado, a interação não é fundamental, mas é desejável, pois ajuda a ter mais profundidade na análise, sem necessariamente aumentar a complexidade. Pode-se aprofundar o conhecimento sobre o conjunto de dados sem ter elementos que distraiam o usuário da essência da análise.

Sobre as vantagens na comunicação, o entrevistado citou a facilidade no acesso a informação, compreensão e compartilhamento. Segundo o gestor, a visualização de dados tem uma função muito importante dentro dos negócios, principalmente, em departamentos empresariais, pois ajuda a interpretar e compartilhar de maneira fácil e ágil.

5 CONSIDERAÇÕES FINAIS

A visualização de dados é um tema contemporâneo no trabalho dos estatísticos e cientistas de dados. Inúmeras soluções vêm sendo criadas para facilitar a sua utilização, principalmente nos departamentos empresariais, para os quais pode trazer vantagens significativas aos negócios. Paradoxalmente na era do Big Data, quando as mudanças acontecem a uma velocidade que supera a capacidade de aprendizagem, precisa-se de mais simplicidade e rapidez.

Diante deste contexto, o objetivo principal deste artigo foi utilizar o método de visualização de dados. De forma geral, pode-se adotar qualquer plataforma para auxiliar a visualização. O ponto principal da pesquisa não é apenas “como visualizar” os dados, mas também como tornar a visualização útil e compreensível para o usuário. Desde a visualização de métodos simples de baixa dimensionalidade, até os métodos mais complexos e de alta dimensionalidade, o algoritmo para construção das visualizações pode ser mantido.

O profissional de estatística pode trabalhar em todas as áreas de estudo da atividade acadêmica ou em diversos setores do mercado de trabalho, utilizando-se de métodos estatísticos que permitam o planejamento adequado e a análise eficiente de dados. Em função disso, deve ter boa capacidade de interagir com equipes multidisciplinares e ser criativo para transmitir eficientemente a informação gerada pelos números. Também deve-se apropriar da comunicação como sua principal virtude e, sendo seu trabalho permeado por números, a forma mais intuitiva e direta de comunicação é através de imagens. Por este motivo, a visualização de dados é um poderoso meio para compreender e democratizar a informação.

Este trabalho proposto teve algumas limitações nos resultados finais apresentados:

- 1) A pesquisa para avaliação qualitativa do método proposto foi realizada com apenas um gestor, o que pode ter gerado resultados distorcidos ou viesados. Além de envolver mais gestores, esta avaliação poderia envolver uma etapa de experimentação mais elaborada com hipóteses mais complexas.
- 2) Utilizou-se apenas estatísticas descritivas para análise e construção das visualizações, o que era suficiente para atender à hipóteses propostas neste trabalho, mas poderiam ser utilizados métodos mais sofisticados de análise, obtendo-se múltiplas comparações.

Conforme demonstrado neste trabalho, o cenário atual e futuro mostra que a visualização de dados é um tema que ganha cada vez mais relevância, tanto no contexto acadêmico, quanto no mercado de trabalho. Por ser um tema ainda pouco explorado, sugere-se estudos de caracterização do tema, utilizando técnicas estatísticas mais sofisticadas, com o intuito de comunicar e apresentar os resultados.

Por fim, sugere-se a construção de um aplicativo ou plataforma que possibilite o agrupamento das informações, com filtros, seleção de variáveis e entrada de dados. A interação do usuário neste contexto é muito mais rica e propicia uma experiência ampla e intuitiva.

REFERÊNCIAS

- CARD, S. K.; MACKINLAY, J. D.; SHNEIDERMAN, B. *Using Vision to Think, chapter 1: Information Visualization*, p. 1–34. MORGAN KAUFMANN, 1999.
- CHEN, C., Top 10 unsolved information visualization problems, *IEEE Computer Graphics and Applications*, 2005.
- DE MAURO, A.; GRECO, M.; GRIMALDI, M. *What is big data? A consensual definition and a review of key research topics*. Disponível em: < <http://big-data-fr.com/wp-content/uploads/2015/02/aip-scitation-what-is-bigdata.pdf> >. Acesso em: 12/07/2017.
- DIEBOLD, D. X. *A Personal Perspective on the Origin(s) and Development of “Big Data”: The Phenomenon, the Term, and the Discipline*, p. 35-45, 2012.
- FERGUNSON, R. *Data Analytics and the Information Transfer Gap*. < <http://sloanreview.mit.edu/article/data-analytics-and-the-information-transfer-gap/> >. Acesso em: 12/07/2017.
- FRIENDLY, M. *Visualizing Categorical Data*. Cary, NC: SAS Institute. ISBN 1-58025-660-0, 2000.
- GARTNER IT GLOSSARY: *Big Data*. Disponível em: < <http://www.gartner.com/it-glossary/big-data/> >. Acesso em: 12/07/2017.
- IMB: *O que é Big Data*. Disponível em: < https://www.ibm.com/midmarket/br/pt/infografico_bigdata.html >. Acesso em 12/07/2017.
- KAIDI, Z. *Data Visualization*. Disponível em: < https://www.cs.uic.edu/~kzhao/Papers/00_course_Data_visualization.pdf >. Acesso em: 12/07/2017.
- KEIM, D.: Information Visualization and Visual Data Mining, *IEEE transactions on visualization and computer graphics*, v. 7, no. 1, jan-mar 2002.
- KEIM, D.; WARD, M. *Visual Data Mining Techniques*, Johnson/Hansen: The Visualization Handbook, 2003.
- LANEY, D. *3D Data Management: Controlling Data Volume, Velocity, and Variety*, Meta Group. Disponível em: < <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> >. Acesso em: 12/07/2017.
- MATTHEW O. W.; GRINSTEIN, G.; KEIM, D. *Interactive Data Visualization Foundations Techniques and Applications*. 2nd Edition, CRC Press, 2010.

PARSAYE, K.; CHIGNELL, M. *Intelligent Database Tools & Applications: Hyperinformation access, data quality, visualization, automatic Discovery*. John Wiley & Sons Inc. 541p.1993.

ROBERTSON, G.; CARD, K.S.; MACKINLAY, D.J. Information visualization using 3D interactive animation. *Magazine Communications of the ACM - Special issue on graphical user interfaces*, p. 57-71, 1993.

SHNEIDERMAN, B. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. *In: Proceedings of IEEE Symposium on Visual Languages*, p. 336-343, 1996

SYNNESTVEDT, M.; CHEN, C. Design and Evaluation of the Tightly Coupled Perceptual-Cognitive Task in Knowledge Domain Visualization. *The 11th International Conference on Human-Computer Interaction*, 2005.

VLIEGEN, R.; WIJK, J. J. V.; LINDEN, E. J. V. D. Visualizing Business Data with Generalized Treemaps. *IEEE Transactions on Visualization and Computer Graphics* 12, 789-796, 2006.

WARD, M. *Overview of Data Visualization*, Disponível em: < www.cs.wpi.edu >. Acesso em: 12/07/2017

WIJK, J.J. *Image Based Flow Visualization*. Proc. SIGGRAPH'02, Transaction on Graphics 21:3, 2002.

YAU, N. *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*. Ed. Willey. 384p. 2011.

YI, J. S.; YOUN, K.; STASKO, J. T.; JACKO, J. A. *Toward a Deeper Understanding of the Role of*, IEEE Symposium on Information Visualization. Disponível em: < <http://www.cc.gatech.edu/~stasko/papers/infovis07-interaction.pdf> >. Acesso em: 12/07/2017.