

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DA CIÊNCIA DA COMPUTAÇÃO

ANTONIO FIUZA DE SOUSA LANDIM

VISUALIZAÇÃO DE INFORMAÇÃO COMO APOIO À GESTÃO PÚBLICA

BRASÍLIA

2019

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO
ESPECIALIZAÇÃO EM INFORMÁTICA

**VISUALIZAÇÃO DE INFORMAÇÃO COMO APOIO A GESTÃO
PÚBLICA**

por

Antonio Fiuza de Sousa de Landim

Monografia de final de curso

Profa. Dra. Raquel C. de Melo Minardi
Orientadora

BRASÍLIA

2019

ANTONIO FIUZA DE SOUSA LANDIM

VISUALIZAÇÃO DE INFORMAÇÃO COMO APOIO À GESTÃO PÚBLICA

Monografia apresentada ao Curso de Especialização em Informática do Departamento de Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito para a obtenção do grau de Especialista em Informática.

Orientadora: Prof^a Dra Raquel C. de Melo Minardi

BRASÍLIA

2019

Landim, Antonio Fiuza de Sousa

L257v Visualização de informação como apoio à gestão pública /
Antonio Fiuza de Sousa Landim. – Brasília, 2019.
xi, 73 f. : il.

Monografia (especialização) – Universidade Federal de Minas
Gerais. Departamento de Ciência da Computação.

Orientador: Raquel C.de Melo Minardi

1. Computação – Monografias. 2. Gestão de Dados. 3.
Visualização de Informação. 4. Administração pública.I. Orientador.
II. Título.

CDU 519.6*



UNIVERSIDADE FEDERAL DE MINAS GERAIS

INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO
ESPECIALIZAÇÃO EM INFORMÁTICA: ÁREA DE CONCENTRAÇÃO GESTÃO EM
TECNOLOGIA DA INFORMAÇÃO

Visualização de Informação como Apoio à Gestão Pública

ANTÔNIO FIUZA DE SOUZA LANDIM

Monografia apresentada aos Senhores:

Profa. Raquel Cardoso de Melo Minardi
Orientadora
DCC - ICEx - UFMG

Prof. José Nagib Cotrin Arabe
DCC - ICEx - UFMG

Prof. José Marcos Silva Nogueira
DCC - ICEx - UFMG

Belo Horizonte, 14 de março de 2019

AGRADECIMENTOS

Agradeço a Deus, pela bênçãos e dons recebidos.

Agradeço à minha família, pela paciência e compreensão ao longo desta jornada

À minha orientadora, pelo apoio e conhecimentos adquiridos.

E finalmente aos colegas de curso pela convivência.

RESUMO

Este trabalho apresenta uma forma útil de extrair dos dados informações relevantes para subsidiar a tomada de decisões pela gestão pública, através do uso de técnicas de visualização de informação. Também descreve, de modo o mais formal possível, o processo de escolha das respectivas visualizações de maneira a facilitar a disseminação do seu uso em diferentes conjuntos de dados.

Adicionalmente, este trabalho destaca a importância e os desafios de mensurar a efetividade de uma visualização, mapeada com base em uma análise da literatura científica da área, e na provocação dos gestores públicos sobre a importância desta atividade. Neste sentido, apresenta um estratégia para avaliação da efetividade das visualizações utilizadas apropriadas para o contexto da gestão pública.

Palavras-chave: Visualização de Informação, Avaliação, Painel, Visualização

ABSTRACT

This work presents a useful way of extracting from the data relevant information to subsidize decision making by public management through the use of information visualization techniques. It also describes, as formally as possible, the process of choosing the respective visualizations in order to facilitate the dissemination of their use in different datasets.

Additionally, this work highlights the importance and the challenges of measuring the effectiveness of a visualization, mapped based on an analysis of the scientific literature of the area, and on the provocation of the public managers on the importance of this activity. In this sense, it presents a strategy for evaluating the effectiveness of the visualizations used appropriate to the context of public management.

Keywords: Information Visualization, Evaluation, Dashboard, Visualization

LISTA DE FIGURA

FIGURA 1 – Exemplo de visualização: Tipo de sanção – CEIS	13
FIGURA 2 – Exemplo de visualização: Proporção dos gastos em relação ao total do órgão	14
FIGURA 3 - Exemplo de visualização: Custos controláveis	14
FIGURA 4 - Exemplo de visualização: Média/Mediana de preços por período.	15
FIGURA 5 - Exemplo de visualização: Gráfico de Dispersão por item de compra	15
FIGURA 6 – Exemplo de gráfico de pareto	33
FIGURA 7 - Classificação dos itens identificados para a questão A	37
FIGURA 8 - Bullet graph - Quantidade atual e percentual alcançado de servidores e pensionistas, em relação ao total possível, que estão utilizando o SIGAC	38
FIGURA 9 - Quantidade de usuários do SIGAC, em relação ao potencial do seu público alvo	40
FIGURA 10 - Classificação dos itens identificados para a questão B	41
FIGURA 11 - Histograma representando a distribuição dos usuários SIGAC por faixa etária e sexo	42
FIGURA 12 - Histograma representando a distribuição dos usuários SIGAC cadastrados com situação funcional "ATIVOS", segmentado por faixa etária e sexo	43
FIGURA 13 - Histograma representando a distribuição dos usuários SIGAC cadastrados com situação funcional "INATIVOS", segmentado por faixa etária e sexo	44
FIGURA 14 - Histograma representando a distribuição dos usuários SIGAC cadastrados com situação funcional "PENSIONISTAS", segmentado por faixa etária e sexo	45
FIGURA 15 - Pequenos múltiplos representando a distribuição dos usuários SIGAC de acordo com a situação funcional, segmentado por faixa etária e sexo	46
FIGURA 16 - Pequenos múltiplos representando a distribuição dos usuários SIGAC de acordo com a situação funcional, segmentado por faixa etária e sexo. Nesta visualização foi incluído a representação de todas as situações funcionais agrupadas	47
FIGURA 17 – Classificação dos itens identificados para a questão C	48
FIGURA 18 - Técnica de visualização sparkline, associada a outros indicadores, utilizada para representar o comportamento da identidade do usuário no sistema SIGAC	49
FIGURA 19 - Classificação dos itens identificados para a questão D	51
FIGURA 20 - Gráfico de linhas representando a quantidade de acessos diários, picos, vales e outliers no ano de 2018	52
FIGURA 21 - Gráfico de linhas representando a quantidade de acessos diários, picos, vales e outliers no ano de 2017	52
FIGURA 22 - Gráfico de linhas representando a quantidade de acessos mensais no ano de 2017 e 2018	53
FIGURA 23 - Gráfico de linhas representando a quantidade de acessos trimestrais no ano de 2017 e 2018.	54
FIGURA 24 - Diagrama de caixa representando a quantidade de acessos diários, consolidados por mês, no período de 2017 e 2018.	55
FIGURA 25 - Classificação dos itens identificados para a questão E	56
FIGURA 26 - Visualização treemap agregada com indicadores numéricos e filtros de variáveis.	57
FIGURA 27 - Visualização treemap, com seleção de filtro que exclui o módulo "SERVIDOR".	58
FIGURA 28 - Visualização treemap, com seleção de filtro em um módulo específico do sistema SIGEPE	58
FIGURA 29 - Classificação dos itens identificados para a questão F	59
FIGURA 30 – Gráfico de linhas representando a média diária de usuários no SIGAC realizando acessos por diferentes canais (Web e Mobile) no período de 2017 e 2018 e início de 2019.	60
FIGURA 31 - Gráfico de barras empilhadas representando a consulta ao contracheque por meio de diferentes canais, consolidado por mês	61

FIGURA 32 - Gráfico de barras empilhadas representando a consulta à prévia do contracheque por meio de diferentes canais, consolidado por dia do mês.	62
FIGURA 33 - Diferentes graus de ênfase visual estão associados a diferentes regiões de um painel. 59 esta faltando a figura	63
FIGURA 34 - Visão geral do painel com um conjunto de visualizações representando dados do SIGAC	64
FIGURA 35 - FP-tree modificado para anotar técnicas de visualização.	72

LISTA DE TABELAS

TABELA 1 - Descrição dos dados, e respectivas fontes, utilizadas para compor as visualizações.	28
TABELA 2 – Possíveis relacionamentos quantitativos aplicados aos dados	29
TABELA 3 - Coleção de termos utilizado no estudo de FRANCISCANI JUNIOR., Geraldo <i>et al</i> [5]	71

SUMÁRIO

1	INTRODUÇÃO	10
1.1	OBJETIVOS.....	11
1.2	JUSTIFICATIVA	12
1.3	RESULTADOS E CONTRIBUIÇÕES ESPERADAS	16
2	TRABALHOS RELACIONADOS.....	17
3	MÉTODO	23
4	ENTENDER O NEGÓCIO	24
5	QUESTÕES DE PESQUISA	26
6	MAPEAMENTO E PREPARAÇÃO DOS DADOS.....	27
7	RELACIONAMENTOS QUANTITATIVOS	29
7.1	SÉRIES TEMPORAIS	30
7.2	RANKING E PARTE-TODO.....	32
7.3	DISTRIBUIÇÃO	33
8	SELEÇÃO DA TÉCNICA DE VISUALIZAÇÃO	35
9	DISPONIBILIZAÇÃO DO PAINEL.....	63
10	ESTRATÉGIA DE AVALIAÇÃO	65
11	CONCLUSÕES	67
	REFERÊNCIAS BIBLIOGRÁFICAS.....	69
	APÊNDICE A	71
	APÊNDICE B	72

1 INTRODUÇÃO

A evolução das tecnologias de comunicação e informação vem promovendo diversas mudanças na sociedade como um todo. Uma delas está relacionada ao crescimento exponencial do volume de informações que nos é apresentado. Devido à essa enorme sobrecarga de informação a qual somos submetidos, enfrentamos, em geral, grande dificuldade de lidar com esse fenômeno e de extrair informação relevante e conhecimento de dados. Infelizmente, nossa capacidade de compreender e comunicar não tem evoluído na mesma velocidade do volume de dados gerados e disponíveis nos mais diversos cenários.

Os computadores aceleram o processo de manipulação de dados e informações, mas ainda não são capazes de nos ajudar de forma automática a entender esses dados, extrair informação ou comunicar seu significado aos tomadores de decisão. Essas habilidades não são intuitivas e, em geral, mesmo o ensino superior não nos dá os subsídios adequados para aquisição das competências suficientes para análise de dados.

Um modo de facilitar e potencializar a análise de dados e obtenção de *insights* é utilizar técnicas de visualização de informação. Elas fornecem um meio poderoso tanto para dar sentido aos dados quanto para comunicar o que descobrimos aos outros. Através destas técnicas é possível reunir milhares de dados em uma imagem e revelar padrões visuais eventualmente ocultos neles em forma tabular/textual.

Apesar de seu potencial, os benefícios da visualização de informação são atualmente afetados por uma falta geral de compreensão sobre o funcionamento do sistema de percepção e cognição humanos. Muitas das tendências atuais na visualização de informação estão produzindo o oposto do efeito pretendido, confusão em vez de compreensão. Para usufruir dos benefícios da visualização, temos que utilizar o que realmente funciona em termos de percepção e abrir mão de utilizar objetos visuais e funcionalidades simplesmente decorativos e que apenas confundem e prejudiquem.

Conforme palavras de FEW, Stephen C. [1], devemos utilizar "mais ciência do que arte, mais simplicidade do que barulho" (tradução livre).

Por fim, importante ressaltar, conforme WARE, Colin [4], que "o sistema visual tem suas próprias regras. Podemos ver facilmente os padrões apresentados de certas maneiras, mas se eles são apresentados de outras formas, eles se tornam invisíveis. Se pudermos entender como funciona a percepção, nosso conhecimento pode ser traduzido em regras para exibir informações".

1.1 OBJETIVOS

O objetivo geral deste trabalho é aplicar técnicas de visualização de informação em um conjunto de dados de tal forma a dar sentido à informação e não somente apresentá-la, demonstrando o seu potencial como mecanismo de apoio à tomada de decisões e melhoria da eficiência da gestão pública.

Para a consecução do objetivo geral serão propostos os seguintes objetivos específicos:

1. Criação de um painel (*dashboard*¹) com um conjunto de visualizações representando dados do Sistema de Gestão de Acesso - SIGAC, associados a dados do Sistema de Gestão de Pessoas do Governo Federal - SIGEPE/SIAPE;
2. Descrever de modo o mais formal possível o processo de escolha das respectivas visualizações de maneira a facilitar a disseminação do seu uso em outros conjuntos de dados.
3. Fazer uma revisão da literatura com objetivo de compreender o estado da arte das estratégias de avaliação de efetividade das visualizações.
4. Apresentar uma estratégia para avaliação da efetividade das visualizações utilizadas.

1

Dashboard, definido por FEW, Stephen C. [1] como "exibição visual das informações mais importantes necessárias para atingir um ou mais objetivos, consolidados e organizados em uma única tela, para que as informações possam ser monitoradas rapidamente."

1.2 JUSTIFICATIVA

Considerando somente o contexto do Governo Federal, há iniciativas que visam prover mecanismos para tomada de decisões, promover controle social, transparência das ações públicas, entre outros.

Como exemplo, pode-se citar a INDA, Infraestrutura Nacional de Dados Abertos, que "é um conjunto de padrões, tecnologias, procedimentos e mecanismos de controles necessários para atender às condições de disseminação e compartilhamento de dados e informações públicas no modelos de dados abertos em conformidade com o disposto na e-PING" [2]. O foco principal da iniciativa de dados abertos é tornar público os dados governamentais. Um dos desafios colocados de forma tácita é como extrair informações úteis e relevantes a partir dos dados governamentais disponibilizados.

Além da iniciativa de dados abertos, mais recentemente vários órgãos do Governo Federal tem disponibilizado portais e/ou painéis para consulta de informações.

O Painel de Preços [3], de acordo com sua página principal, "disponibiliza de forma clara e de fácil leitura, dados e informações de compras públicas....Tem como objetivo auxiliar os gestores nas execuções de processos de compras.... e estimular o controle social".

O Portal de Transparência (<http://www.portaltransparencia.gov.br/>), "lançado pelo Ministério da Transparência e Controladoria-Geral da União em 2004, é um site de acesso livre, no qual o cidadão pode encontrar informações sobre como o dinheiro público é utilizado, além de se informar sobre assuntos relacionados à gestão pública do Brasil." Além de disponibilizar vários dados para que os usuários possam analisar, o portal possui uma área de painéis, onde dados são apresentados visualmente (vide exemplo da FIGURA 36).

Por meio de uma pesquisa no Google encontramos vários outros portais/painéis disponibilizados por diferentes órgãos/secretarias do Governos Federal, por exemplo:

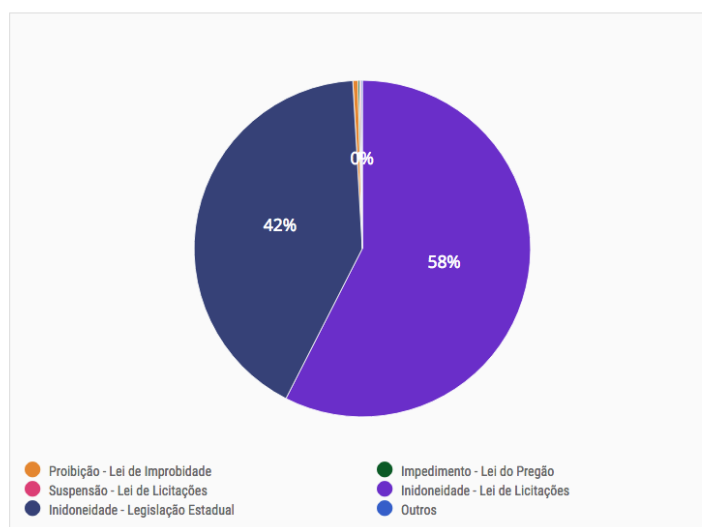
- Painel Estatístico de Pessoal (<http://painel.pep.planejamento.gov.br/>);
- Painel de Obras (<http://paineldeobras.planejamento.gov.br/>);
- Painel de Transferências Abertas
(<http://www.transferenciasabertas.planejamento.gov.br/>);
- Painel de Gastos de TI (<http://paineis.cgu.gov.br/gastosti/index.htm>);
- Painel de Governança de TI na Administração Pública (<https://painel3.tcu.gov.br/>);
- Portal de Custos (<http://www.tesourotransparente.gov.br/visualizacoes/portal-de-custos>);

Além dos painéis públicos, vários outros estão em uso privativo dos órgãos da administração pública com objetivo de disponibilizar dados aos gestores para subsidiar a tomada de decisões, elaboração de políticas públicas, etc.

Em todos os exemplos de portais/painéis citados são utilizados recursos de visualização com a intenção de facilitar o entendimento e/ou exploração dos dados/informações (vide exemplos FIGURA 37, FIGURA 38, FIGURA 39 e FIGURA 40).

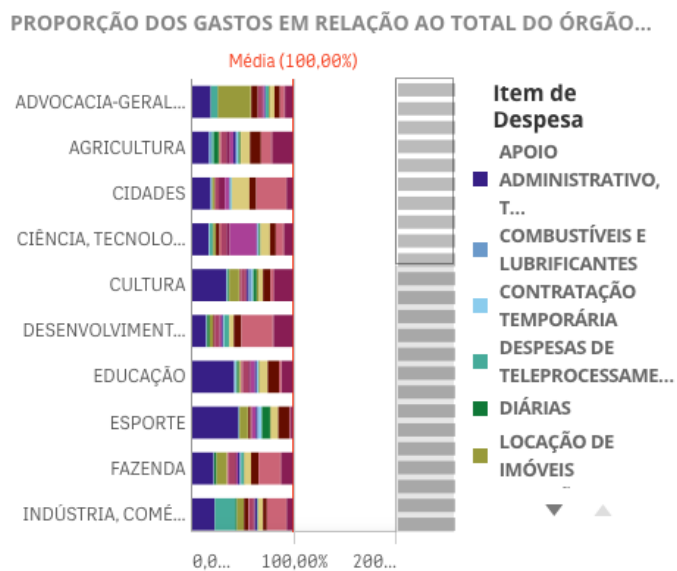
FIGURA 36 – Exemplo de visualização: Tipo de sanção – CEIS

TIPO DE SANÇÃO - CEIS



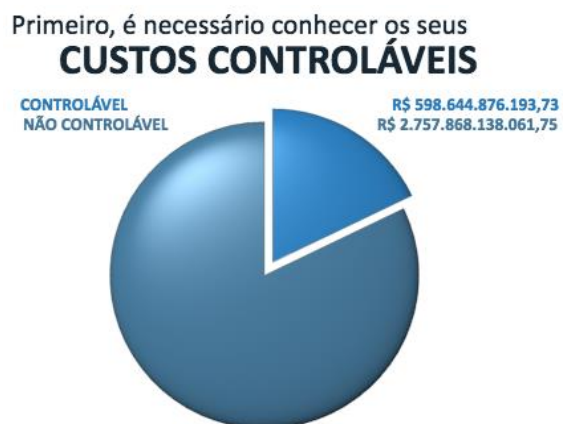
FONTE: Portal da Transparência.

FIGURA 37 – Exemplo de visualização: Proporção dos gastos em relação ao total do órgão.



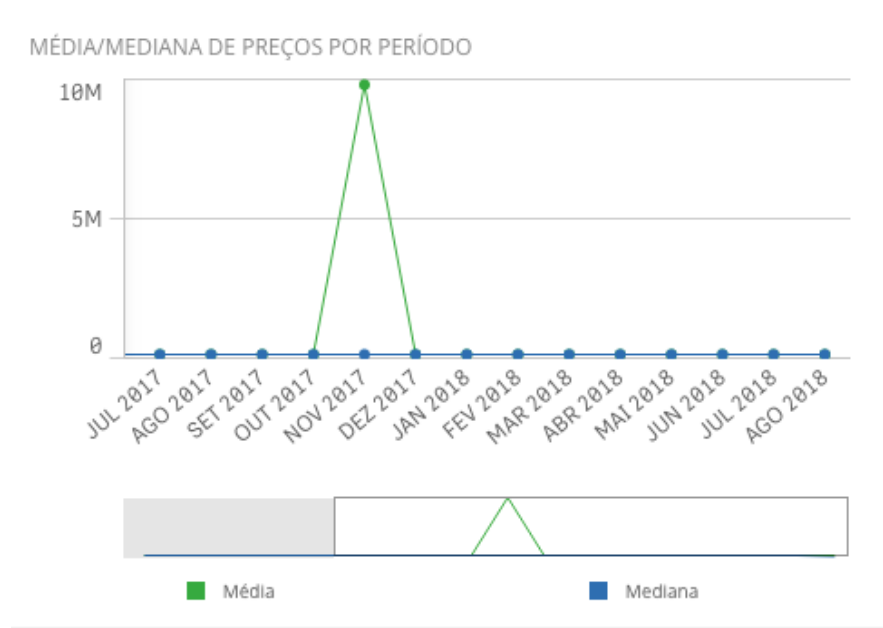
FONTE: Painel de custeio

FIGURA 38 - Exemplo de visualização: Custos controláveis



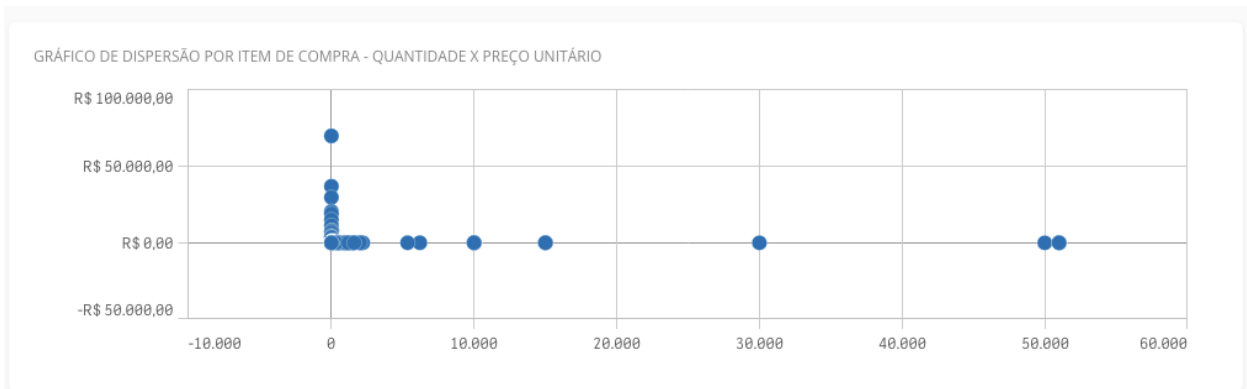
FONTE: Portal de Custos.

FIGURA 39 - Exemplo de visualização: Média/Mediana de preços por período.



FONTE: Painel de Preços.

FIGURA 40 - Exemplo de visualização: Gráfico de Dispersão por item de compra



FONTE: Painel de Preços.

Estas iniciativas são louváveis e extremamente importantes, mas analisando do ponto de vista do potencial da visualização de informação, pode-se questionar se, de uma maneira geral, utilizam as técnicas mais adequadas para o respectivo dado e pergunta específica. Ou se as técnicas de visualização utilizadas promovem o adequado estímulo para a descoberta de informações e/ou conhecimento existentes nos dados

disponíveis. A partir da pequena amostragem coletada, observa-se que em alguns casos é complicado extrair conhecimento a partir da visualização disponibilizada.

O que tem ocorrido é a aquisição de pacotes sofisticados para análise e visualização de dados sem a devida capacitação das pessoas que as operam em análise e visualização de dados.

Adicionalmente, não foram encontradas evidências de utilização de mecanismos de avaliação e/ou mensuração da efetividade das visualizações utilizadas. Assumindo esta constatação como uma realidade, há uma boa probabilidade de que realmente parte das visualizações disponibilizadas pelos diferentes painéis não estejam alcançando os objetivos esperados ou potenciais e que não há nenhum processo sistematizado que tenha como propósito aprimorar as soluções inicialmente disponibilizadas. Neste contexto essas soluções acabam consumindo um volume considerável de recursos públicos para sua criação e manutenção sem que haja uma contribuição mensurável para melhoria da gestão pública.

1.3 RESULTADOS E CONTRIBUIÇÕES ESPERADAS

Como resultado deste trabalho espera-se apresentar um conjunto de técnicas de visualização e os respectivos processos para suas escolhas, selecionados dentro de várias opções disponíveis atualmente, cujo objetivo é ser utilizado como referência para as iniciativas da administração pública de utilização de visualizações de informação como mecanismo efetivo para tomada de decisões, promoção do controle social, transparência das ações públicas, entre outros.

Este trabalho também traz uma contribuição no sentido de apresentar os desafios de mensurar a efetividade de uma visualização, mapeada com base em uma análise da literatura científica da área, e na provocação dos gestores públicos sobre a importância desta atividade.

2 TRABALHOS RELACIONADOS

Um desafio relacionado à visualização de informação é organizar o conhecimento dos métodos e capacidades de cada técnica de visualização, com o foco em tornar o desenvolvimento da visualização fácil, acessível e efetivo (FRANCISCANI JR., Geraldo *et al.* [2014] [5]). Essa organização do conhecimento deve promover o desenvolvimento de visualizações que apoiem os usuários na escolha de visualizações para problemas específicos e representação efetiva dos dados

Porém, o que tem sido observado na prática é um cenário diferente. Na medida do crescimento da área de visualização de informação, muitas técnicas de visualização vêm sendo desenvolvidas e não necessariamente essas técnicas estão sendo utilizadas de forma adequada. Como exemplo, no estudo elaborado por FRANCISCANI JR., Geraldo *et al.* [2014] [5] foram mapeadas 53 técnicas de visualizações diferentes. Em função desta grande quantidade e diversidade de técnicas, a escolha da visualização adequada para o respectivo conjunto de dados têm se tornado uma atividade complexa tanto para especialistas quanto para os usuários finais.

Neste contexto, umas das abordagens adotadas pela comunidade científica é a proposição de taxonomias/ontologias e modelos de classificação para categorizar as técnicas de visualização e apoiar os usuários e especialistas na seleção da respectiva visualização a ser utilizada. Como resultado, há uma diversidade de estudos que abordam aspectos e contextos distintos, tais como, tarefas por tipos de dados (TTT), apresentado por SHNEIDERMAN, Ben. [6], as tarefas executadas pelos usuários (HEER e SHNEIDERMAN [8]), etapas operacionais de processamento dos dados (CHI [7]), algoritmo de visualização (TORY e MÖLLER [9]), combinação de diferentes aspectos relacionados à visualização de informação (dados, tarefas, interações, contexto e capacidade de cognição humana), proposto por PFITZNER *et al.* [10], e ontologia baseada nas propriedades das representações gráficas e objetos visuais (GILSON, O *et al.* [11]).

Esta pequena amostra de estudos indica que não há atualmente uma taxonomia unificada, tampouco um método único a ser seguido no sentido de conduzir o usuário para a escolha de uma visualização ao problema específico. Há alguns estudos e iniciativas no sentido de propor uma "unificação" das diferentes iniciativas. Um exemplo a ser citado é o estudo realizado por FRANCISCANI JR., Geraldo *et al.* [2014] [5], que propõe um processo de anotação que consiste na definição de uma coleção de termos e conceitos relacionados a um conjunto de técnicas de visualização de dados. O índice visual na forma de uma árvore de anotação, proposto no estudo (FIGURA 70), ajuda a visualizar todo o conjunto de técnicas de visualização analisadas (53) e os termos e atributos associados a cada uma delas (TABELA 6).

Outro exemplo, é a tese de mestrado de BARROS [12], que também propõe " uma taxonomia unificada que pode ser usada para descrever e comparar os modelos existentes". A partir da análise de publicações sobre o tema, esse trabalho identificou três focos ou eixos principais de classificação, a saber:

- Dados: caracterização dos tipos de dados sendo visualizados. Ex: quantitativos ou qualitativos ou sua dimensionalidade: unidimensional (1D), bidimensional (2D), tridimensional (3D) e multidimensional.
- Tarefas Analíticas: descrição das tarefas dos usuários durante a análise de dados. Ex: identificar padrões desconhecidos ou comparar dados de acordo com suas características.
- Atributos Visuais: atributos relacionados à representação visual. Ex: cor, forma, tamanho ou textura.

O escopo deste trabalho não é propor uma nova taxonomia ou analisar as taxonomias existentes. E, como não há um padrão definido de facto ou adotado universalmente pela comunidade científica, durante a descrição do processo de escolha do conjunto de visualizações a ser proposto, buscar-se-á utilizar termos comumente utilizados. Como referência, será utilizado a lista de 74 termos mapeados no estudo de FRANCISCANI JR., Geraldo *et al.* [2014] [5]. Além da utilização destes termos em comum, será utilizada como direcionador os eixos de classificação de BARROS [12].

Outro desafio relacionado à visualização de informação é avaliar a efetividade das visualizações que é um processo naturalmente complexo. Esta avaliação não se restringe simplesmente à avaliação da visualização, mas também envolve outros aspectos como desempenho, eficiência de algoritmos utilizados e avaliação do alcance da comunicação realizada por meio da visualização. Conforme citado por LAM, Heidi *et al.*[2012] [13], "pode ser assustador para os avaliadores identificar as questões de avaliação certas a serem feitas, escolher as variáveis certas a serem avaliadas, escolher as tarefas certas, usuários ou conjuntos de dados a serem testados e escolher os métodos de avaliação apropriados."

Em seu estudo, em que realizou uma extensa revisão de literatura de mais de 800 publicações sobre visualização de informação, LAM, Heidi *et al.*[2012] [13] apresenta uma abordagem diferente. Ao invés de propor um método prescritivo de avaliação, propõe uma abordagem descritiva. Essa visão é baseada em cenários de avaliação, categorizados em dois grupos: entender processo de análise de dados e avaliar as próprias visualizações. No primeiro grupo, entender processo de análise de dados, são propostos 4 cenários:

- Avaliando ambientes e práticas de trabalho;
- Avaliando análise de dados visuais e raciocínio;
- Avaliando a comunicação através da visualização;
- Avaliando a análise colaborativa de dados;

No segundo grupo (avaliar as próprias visualizações), são propostos 3 cenários:

- Avaliando o desempenho do usuário;
- Avaliando a experiência do usuário;
- Avaliando algoritmos de visualização;

O aspecto interessante dessa abordagem está no fato de que os cenários distinguem diferentes objetivos de estudo e tipos de perguntas de pesquisa, guiando o avaliador para explorar diferentes opções de avaliação. Essa abordagem tem o mérito de estimular a reflexão sobre metas e questões de avaliação antes de escolher os métodos específicos.

Adicionalmente, o estudo mostra um contínuo aumento no percentual de artigos que reportam pelo menos um tipo de avaliação (o estudo avaliou artigos no período de 1995 a 2010). Também identifica que a maior parte das avaliações se concentram no segundo grupo de cenários (avaliar as próprias visualizações), com um total de 85% de avaliações.

ISENBERG, Tobias et al [\[14\]](#), apresentou um estudo utilizando e ampliando o esquema de codificação estabelecido previamente por LAM, Heidi et al.[2012] [\[13\]](#). A avaliação foi realizada num conjunto de artigos diferentes, publicados na conferência IEEE Visualization (o estudo anterior foi focado em artigos publicados em outras conferências da área). Foi adicionado apenas um novo cenário (inspeção de resultado qualitativo) no grupo de avaliação das próprias visualizações. Esse novo cenário inclui avaliações através de discussões qualitativas e avaliações dos resultados das visualizações. Ele não envolve usuários finais na avaliação, e sim apenas o visualizador de uma imagem resultante, que é instado a avaliar por conta própria a respectiva visualização.

De forma semelhante aos resultados do estudo de LAM, Heidi et al.[2012] [\[13\]](#), a maior parte dos achados de avaliação se concentraram no grupo de avaliação das próprias visualizações. A principal diferença observada entre os estudos foi a concentração de avaliações em cenários diferentes (avaliação da experiência de usuários - LAM vs avaliação de algoritmo - ISENBERG).

O estudo de ISENBERG também identificou que 97% dos artigos analisados, utilizando a codificação proposta, apresentaram pelo menos um tipo de avaliação (utilizando a codificação de LAM, o índice é de 76%). Outro fato interessante foi a constatação de que, embora tenha melhorado nos últimos anos, o nível geral de rigor das avaliações reportados nos artigos ainda é muito baixo. Isto porque, na maioria dos estudos, a avaliação é realizada apenas pelo próprio autor da visualização, representando um viés importante. Idealmente as visualizações deveriam ser avaliadas por outros usuários independentes.

Em seu artigo NORTH, C.[15] também identifica a tendência recente de maior presença na literatura de estudos relacionados à avaliação de visualização. Faz uma crítica em relação a uma grande quantidade de avaliações que se restringem a mapear se o usuário gostou ou não da visualização. Algo muito raso, considerando a grande quantidade de opções e itens de avaliação passíveis de avaliação.

Neste mesmo artigo, NORTH, C.[15], aproveitando esta tendência recente de estudos relacionados à avaliação de visualização, propõe reabrir uma questão a respeito de uma importante finalidade da visualização: gerar *insights*. Ele define *insight* como uma observação individual sobre os dados pelo usuário, uma unidade de descoberta, e caracteriza cada *insight* como sendo complexo, profundo, qualitativo, relevante e inesperado. Partindo desse ponto, o desafio de uma avaliação é avaliar a quantidade de insights obtidas a partir da visualização. O foco de discussão são os métodos de avaliação aplicados: avaliação controladas baseada em tarefas de benchmarks ou avaliação eliminando as tarefas de benchmarks.

O uso de experimentos controlados baseados em tarefas de *benchmarks* agrega rigor científico e profundidade na análise de pontos específicos. Em contrapartida há possível perda de capacidade de avaliação de identificar um maior número e diversidade de *insights*.

Adotando a estratégia de avaliar sem utilizar tarefas de *benchmarks*, ou seja, utilizando um protocolo "aberto" e uma análise qualitativa de *insights*, há o potencial de revelar os *insights* que os usuários conseguiram obter a partir da visualização. Também cria a oportunidade de comparação entre a expectativa de *insights* e aqueles efetivamente identificados pelos usuários. Em contrapartida, a execução deste método, apesar de mais rico no potencial de identificar *insights*, possui alguns desafios, tais como demandar maior esforço para capturar os resultados, maior motivação dos avaliadores (não estão seguindo apenas instruções) e necessidade de especialistas no domínio da informação para ajudar a codificar os achados juntamente com os especialistas em visualização.

O autor conclui que, na prática, os dois métodos são necessários, pois atingem objetivos diferentes.

Em outro trabalho, STASKO, John [2014][17] também enfatiza uma noção chave do trabalho de LAM, Heidi *et al.*[2012] [13], que é entender o propósito de uma avaliação. O autor propõe "uma avaliação de visualizações orientada por valor, na qual uma pessoa ilustra o valor de um sistema através de quatro recursos importantes: minimizar o tempo para responder a diversas perguntas, estimular a geração de insights e perguntas perspicazes, transmitir a essência dos dados e gerar confiança e conhecimento sobre o domínio e contexto dos dados." Baseado nestes recursos, desenvolveu a equação:

$$V = T + I + E + C$$

V: Valor T: Tempo I: *Insight* E: Essência C: Confiança

Importante destacar que a proposição da "equação" para identificar o valor da visualização é uma métrica qualitativa. Ou seja, a equação serve mais como uma ajuda descritiva que prescritiva

Embora a abordagem dos estudos de LAM, Heidi *et al.*[13], SENBERG, Tobias *et al.*[14] e STASKO, John [2014][17] sejam diferentes da abordagem de NORTH, C.[15], os mesmos são complementares. Convergem no destaque da importância de avaliar a visualização e nos desafios de executar este processo.

Na próxima seção será apresentado o método utilizado para construir o conjunto de visualizações, que servirão de base para o alcance dos resultados e contribuições esperadas desse trabalho.

3 MÉTODO

A fim de construir um painel e um conjunto de visualizações com dados do Sistema de Gestão de Acesso - SIGAC, associados a dados do Sistema de Gestão de Pessoas do Governo Federal - SIGEPE/SIAPE, estamos propondo percorrer um conjunto de etapas, a saber:

- Entender o negócio.
- Definir as perguntas que se deseja responder e as hipóteses a verificar.
- Mapear e preparar os tipos de dados.
- Definir os relacionamentos quantitativo que potencialmente podem responder cada pergunta.
- Definir as técnicas de visualização adequadas para cada pergunta e relacionamento quantitativo e descrever de modo o mais formal possível o processo de escolha destas visualizações.
- Construção do painel
- Apresentar uma estratégia para avaliação das visualizações utilizadas, seguindo propostas existentes na literatura científica que tratam do tema.

4 ENTENDER O NEGÓCIO

O primeiro passo é buscar entender o contexto em que está inserido o conjunto de dados a ser trabalhado a partir de uma perspectiva de negócios. Este trabalho está baseado nos dados oriundos do Sistema de Gestão de Acesso - SIGAC e respectivos sistemas integrados.

O SIGAC é um sistema de autenticação e autorização dos usuários de aplicação em um modelo padronizado e único. Para toda funcionalidade ou processo executado pelo sistema (por exemplo, autenticação), são gerados arquivos de dados que permitem rastrear diversas ações (quem, quando, o que acessou, de onde, etc).

Os seguintes sistemas já utilizam o SIGAC como solução de gestão de acesso:

- Sigepe, novo sistema de gestão de pessoas do governo federal que está substituindo, gradativamente, os sistemas da família Siape (Siape, SiapeCad, SiapeNet e Extrator). Atualmente já possui 25 módulos distintos integrados;
- Sigepe mobile: aplicativo para dispositivos móveis que permite aos servidores e pensionistas da APF acesso a informações pessoais e financeiras;
- Sigepe Banco de Talentos: aplicativo para dispositivos móveis que permite aos servidores públicos registrarem seus conhecimentos, habilidades e experiências.
- Sigepe Gestor: aplicativo para dispositivos móveis que disponibilizam aos gestores funcionalidades relacionadas a gestão de sua equipe (homologação de férias, afastamentos, etc)

O público alvo destes sistemas e aplicativos são principalmente servidores públicos do executivo federal (ativos e aposentados) e pensionistas.

Atualmente o Sigac já possui mais de 1,2 milhões de usuários cadastrados e média mensal de 2,6 milhões de acessos.

O volume de dados gerados tanto no SIGAC, como nos demais sistemas e aplicativos integrados ao SIGAC é muito elevado. E, potencialmente, a partir deste conjunto de

dados, podem ser extraídas informações importantes para a gestão do ambiente, tomada de decisão de novos projetos, justificativa de investimentos, entre outros.

5 QUESTÕES DE PESQUISA

Esse trabalho tem como objetivo demonstrar que o uso de técnicas de visualização de informação aplicadas a um conjunto de dados pode ser uma poderosa ferramenta no apoio a tomada de decisões e melhoria da eficiência da gestão pública.

Para alcançar este objetivo, uma atividade necessária é definir um conjunto de perguntas que inicialmente se deseja responder e/ou hipóteses a verificar. A importância dessa atividade é estimular o desenvolvimento de visualizações que efetivamente possam agregar valor aos usuários. E, potencialmente, os *insights* obtidos pela análise das visualizações iniciais poderão estimular novas questões, que por sua vez, podem demandar aprimoramentos e/ou novas visualizações.

Para o exemplo trabalho com dados do SIGAC, as perguntas iniciais e/ou hipóteses apresentadas foram:

- A. Qual a quantidade de servidores (ativos e inativos) e pensionistas já utilizam o SIGAC, e o potencial de usuários ainda a ser alcançado?
- B. Qual a faixa etária e sexo dos usuários do SIGAC e respectivos quantitativos?
- C. Qual o comportamento dos status das identidades dos usuários (ativa, bloqueada, expirada, desabilitada, etc) ao longo do tempo?
- D. Qual a quantidade e padrão de acesso de usuários ao longo do tempo?
- E. Qual é quantidade de usuários segmentado por sistema e respectivos módulos?
- F. Qual é o canal de acesso mais utilizado para acessar os sistemas: web ou mobile?

6 MAPEAMENTO E PREPARAÇÃO DOS DADOS

Nesta etapa devem ser mapeadas as fontes de dados que possam alimentar as visualizações com objetivo de responder as questões e hipóteses de pesquisa.

É a etapa que demanda maior esforço e dedicação de tempo, principalmente as atividades de limpeza e normalização dos dados.

Importante destacar que o processo de construção de visualizações, de que o mapeamento e preparação também fazem parte, é uma maneira interessante de identificar problemas incorporados no conjunto de dados, como valores ausentes, errôneos ou incompletos.

No escopo deste trabalho foram utilizados os dados descritos na TABELA 4, que correspondem ao período de 2017 até o presente momento.

Importante registrar que nenhum dado pessoal sigiloso está sendo utilizado para compor as visualizações. Um dos tratamentos realizados na etapa de preparação dos dados foi justamente descaracterizar os dados e/ou agregá-los de tal maneira a permitir construir visualizações que representem o comportamento dos dados reais, porém com o cuidado de não expor de nenhuma maneira dados pessoais e/ou sigilosos.

TABELA 4 - Descrição dos dados, e respectivas fontes, utilizadas para compor as visualizações.

Descrição	Fonte
Dados de autenticação nos sistemas e/ou aplicativos integrados ao SIGAC (data, usuário, origem, status da ação, tipo de registro)	Logs diários de auditoria do SIGAC
Dados de uso das principais funcionalidades utilizadas pelos servidores (acesso ao contracheque, autorização de consignatária, solicitação de férias, entre outros)	Logs diários de utilização do Sigepe/Siape
Dados dos status da identidade do usuário no SIGAC	Relatório do SIGAC de status diário do status das identidades dos usuários
Estatísticas de utilização dos sistemas e aplicativos	Google analytics ² e Google Firebase ³
Dados de situação funcional dos servidores	Sigepe/Siape

FONTE: Elaborada pelo autor.

Não é escopo deste trabalho discorrer sobre as transformações, técnicas e/ou ferramentas adotadas para superar esta etapa.

2

Google analytics (<https://analytics.google.com/analytics/web/#/>): estatísticas de acesso ao site do SIGAC (<https://gestaodeacesso.planejamento.gov.br>).

3

Google firebase (<https://firebase.google.com/>): estatísticas de acesso do aplicativo sigepe mobile (<https://www.servidor.gov.br/servidor/sigepe-mobile>)

7 RELACIONAMENTOS QUANTITATIVOS⁴

A análise envolve relacionamento entre os dados e a visualização de informação pode apoiar e facilitar a análise por meio de padrões visuais, juntamente com o uso de técnicas de análise particulares (adaptado de FEW, Stephen [16]).

Neste sentido faz-se necessário identificar os possíveis relacionamentos quantitativos associados aos questionamentos e hipóteses. A partir deste mapeamento e agregando outros componentes (Ex: tarefas e atributos), será pavimentado o caminho para seleção de uma técnica de visualização adequada.

Os principais relacionamentos quantitativos são:

- Séries temporais
- Ranking e parte-todo
- Desvio
- Distribuição
- Correlação
- Multivariados

TABELA 5 – Possíveis relacionamentos quantitativos aplicados aos dados

Questões	Possíveis Relacionamentos Quantitativos					
	Série Temporal	Ranking	Desvio	Distribuição	Correlação	Multivariado
A	x	x		x		
B		x		x		
C	x			x		
D	x	x		x		
E		x		x		
F	x	x				

FONTE: Elaborada pelo autor.

Para o conjunto de dados em análise, dentre os principais relacionamentos quantitativos, serão explorados os relacionamentos de séries temporais, ranking e parte-todo e distribuição.

7.1 SÉRIES TEMPORAIS

Conforme destacado por FEW, Stephen [\[16\]](#), a análise das mudanças dos valores ao longo do tempo é o tipo de relacionamento quantitativo que recebe maior atenção e por tal motivo estão onipresentes no dia dia de diversas publicações. Um dos motivos prováveis está relacionado ao tempo, que mais que qualquer outra variável, agrega um contexto para o entendimento dos dados em análise.

Seis padrões visuais básicos são bastante significativos quando suas mudanças são analisadas ao longo do tempo: tendência, variação, taxa de mudança, covariância, ciclos e exceções. E as diferentes técnicas de visualização aplicadas para representar séries temporais ajudam a revelar estes padrões ocultos no conjunto de dados.

Para os dados em análise neste trabalho, o uso de séries temporais é quase que mandatório. Para quase todos os questionamentos apresentados envolve direta ou indiretamente o aspecto temporal.

Há diversas técnicas de visualizações que podem ser utilizadas para representar séries temporais. As principais delas são:

- Gráfico de linhas;
- Gráfico de barras;
- Gráfico de pontos;
- Gráficos de radar;
- Mapa de calor;
- Diagrama de caixas (*box plot*);
- Gráficos de dispersão;

Nas próximas seções do documento, na medida em que a técnica for utilizada, maiores detalhamentos das características de cada uma delas serão apresentadas.

Adicionalmente, alguns aspectos podem ser observados com objetivo de potencializar o uso e análise de séries temporais através das técnicas de visualização. Conforme descrito por FEW, Stephen [\[16\]](#), "em função da importância da análise de séries temporais para a maioria das organizações....algumas diretrizes devem ser seguidas para evitar erros nas visualizações e análises de séries temporais":

- Agregar os dados em diferentes intervalos de tempo: isto permite identificar diferentes padrões nos dados a depender da granularidade utilizada (hora, dia, mês, trimestre, ano);
- Visualizar o período de tempo dentro do contexto: importante permitir a análise em diferentes intervalos de tempo, de tal forma a direcionar a uma análise sob diferentes perspectivas. Quando uma análise temporal é realizada considerando somente um período curto de tempo, há o risco de identificar padrões que não representem efetivamente o conjunto de dados. O que num curto intervalo (uma semana, por exemplo) representa uma tendência de incremento, se for analisado no intervalo de um mês, pode representar um comportamento inverso (tendência de queda).
- Otimizar a relação de proporção de apresentação do gráfico: trata-se da proporção entre o tamanho do eixo X e a altura do eixo Y. Segundo Cleveland [\[19\]](#), "se a proporção entre os eixos for muito grande, não poderemos mais discriminar duas inclinações positivas ou duas inclinações negativas, porque as orientações ficam muito próximas. De forma semelhante, a constatação é válida quando a proporção é muito pequena." Porém, não há uma regra universal. Deve ser avaliado caso a caso, buscando uma relação que não cause distorções na apresentação. Buscando uma regra prática, Tufte [\[18\]](#) propôs que "a proporção deve ser tal que os gráficos de série temporal tendem para um perfil irregular, em vez de um perfil de pico ou um perfil plano".
- Utilizar escala logarítmica ou percentual para comparar taxas de variação: o uso de escala logarítmica permite que taxas de variações semelhantes sejam apresentadas de forma semelhantes na visualização, independentemente da diferença de escala entre

os dados que estão sendo comparados. De forma alternativa, porém com efeito semelhante, é comparar dois conjuntos de dados utilizando no eixo Y as taxas de variação expressas em valores percentuais.

- Sobrepor escalas de tempo com objetivo de comparar padrões cíclicos: ao sobrepor os gráficos de uma mesma escala de tempo (por exemplo, período de um ano), potencializa-se identificar e comparar padrões de comportamento entre os diferentes períodos.
- Empilhar gráficos de linhas para comparar múltiplas variáveis: nem sempre é possível ou desejável apresentar num mesmo gráfico diferentes variáveis, seja pela diferença de escala ou de unidade de medida, por exemplo. Nestes casos, criar uma visualização com vários gráficos de linhas empilhados, utilizando a mesma escala de tempo no eixo X, permite identificar padrões de mudanças entre as variáveis. Isto é possível mesmo nos cenários em que as escalas quantitativas dos diferentes gráficos são diferentes, não permitindo comparações de magnitudes entre os valores.

7.2 RANKING E PARTE-TODO

Provavelmente o tipo análise mais frequente e relativamente simples efetuada envolve a comparação de uma parte em relação ao todo e a criação de um ranking.

As técnicas de visualizações mais comuns para representar ranking e parte todo são:

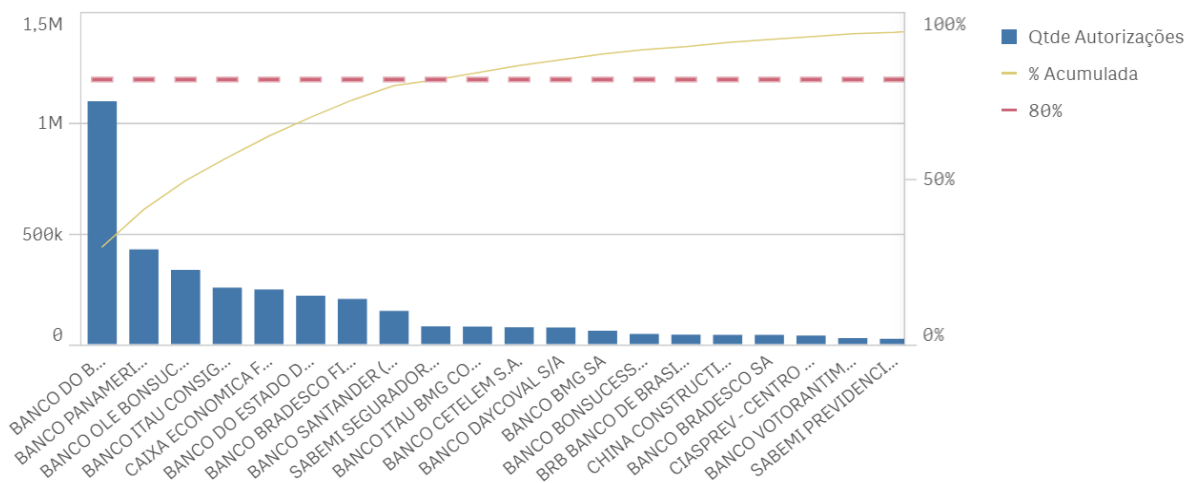
- Gráfico de pizza;
- Gráficos de barras;
- Gráficos de pontos;
- Gráfico de Pareto;

Apesar de ser uma das representações parte todo mais populares, o uso do gráfico de pizza deve ser evitado. A mesma informação pode ser apresentada com maior clareza e facilidade de interpretação utilizando um gráfico de barras da horizontal.

O gráfico de pareto é uma poderosa técnica de visualização, especialmente para analisar a contribuição cumulativa das partes em relação ao todo. A ideia da técnica é

aplicar o conhecido princípio de pareto (80/20) e identificar o conjunto de partes que possui maior relevância no contexto dos dados em análise. A Figura 6 é um exemplo de utilização do gráfico de pareto. No caso em específico, há 576 empresas diferentes no eixo x. Porém apenas 9 empresas são responsáveis por 80% do volume de transações. A depender do objetivo da análise, esta técnica de visualização direciona rapidamente o usuário para o conjunto de dados de maior representatividade.

FIGURA 41 – Exemplo de gráfico de pareto



FONTE: Elaborada pelo autor.

As demais técnicas de ranking e parte todo são bastante conhecidas e utilizadas. Na medida da necessidade, análises pontuais serão feitas em outras seções do documento.

7.3 DISTRIBUIÇÃO

Entender como o conjunto de valores estão distribuídos sob diferentes perspectivas é uma atividade de análise muito relevante. Para tal, diversas técnicas de visualizações foram desenvolvidas com objetivo de representar distribuição. As principais delas são:

- Distribuição de uma única variável:
 - Histogramas;
 - Polígonos de frequência
 - *Strip plots*

- Gráficos de caule e folhas
- Distribuição de múltiplas variáveis:
- Diagrama de caixas (*box plot*);
- Gráfico de múltiplas *strip plots*
- Polígonos de frequência;
- Gráficos de desvio de distribuição

Nas próximas seções do documento, na medida em que a técnica for utilizada, maiores detalhamentos das características de cada uma delas serão apresentadas.

Adicionalmente, alguns aspectos podem ser observados com objetivo de potencializar o uso e análise de distribuições através das técnicas de visualização.

- Manter consistência entre os intervalos das variáveis categóricas: é algo simples, mas essencial para que seja possível realizar análises comparativas.
- Selecionar o intervalo adequado: o desafio neste item é encontrar um equilíbrio entre a quantidade de intervalos em que os dados serão agrupados. Quando é utilizado um número excessivo de intervalos, são apresentados muitos detalhes e dificulta-se a busca por padrões essenciais no conjunto de dados. Por outro lado, quando utilizado poucos intervalos, há uma generalização muito grande dos dados, podendo desta forma resultar na perda de significado nas variações do formato da distribuição.
- Utilizar medidas que são resistentes a *outliers*: a depender do conjunto de dados, o uso de medidas estatísticas como média e desvio, que são bastante influenciadas por dados "fora da curva", podem causar distorções na descrição do padrão da distribuição como um todo apresentado por meio de uma visualização. Neste casos, o uso de medidas como mediana e percentis, que são resistentes a *outliers*, é mais apropriado.

8 SELEÇÃO DA TÉCNICA DE VISUALIZAÇÃO

A partir de uma pré seleção dos relacionamentos quantitativos a serem utilizados, a próxima etapa é selecionar as técnicas de visualizações adequadas.

Conforme mencionado anteriormente, muitas técnicas de visualizações foram desenvolvidas na medida do incremento de interesse em visualização de informação. Somente no estudo de FRANCISCANI JR., Geraldo *et al.* [2014] [5] foram mapeadas 53 técnicas de visualizações diferentes (Figura 35). Isto torna desafiador o processo de selecionar a técnica adequada para o conjunto de dados e tipo de análise desejada.

A identificação dos possíveis relacionamentos quantitativos a serem associados aos questionamentos e hipóteses, realizados anteriormente, auxiliou na seleção de um subconjunto de técnicas de visualizações candidatas (ex: diagrama de caixas, gráfico de barras, linhas, etc).

Porém, de forma a permitir uma seleção de uma técnica mais adequada, alguns aspectos adicionais devem ser considerados. Conforme mencionado anteriormente, a taxonomia proposta por BARROS [12] identificou três focos ou eixos principais de classificação: Dados, Tarefas Analíticas e Atributos Visuais. A identificação dos relacionamentos quantitativos abrange apenas um dos aspectos do eixo de "Dados", de acordo com esta proposta. E, para auxiliar a identificação destes "aspectos adicionais", foi utilizada a ferramenta UTIL⁵, que faz parte do estudo de BARROS [12] .

Apesar de apresentar um alto nível de abstração, optou-se por apoiar-se na ferramenta UTIL em função de sua proposta de descrever os diferentes modelos de classificação de técnicas de visualização encontrados na literatura utilizando a mesma taxonomia, com a virtude de abranger os aspectos destes respectivos modelos.

5

Além dos aspectos de classificação, importante também considerar no processo de seleção da técnica outros pontos:

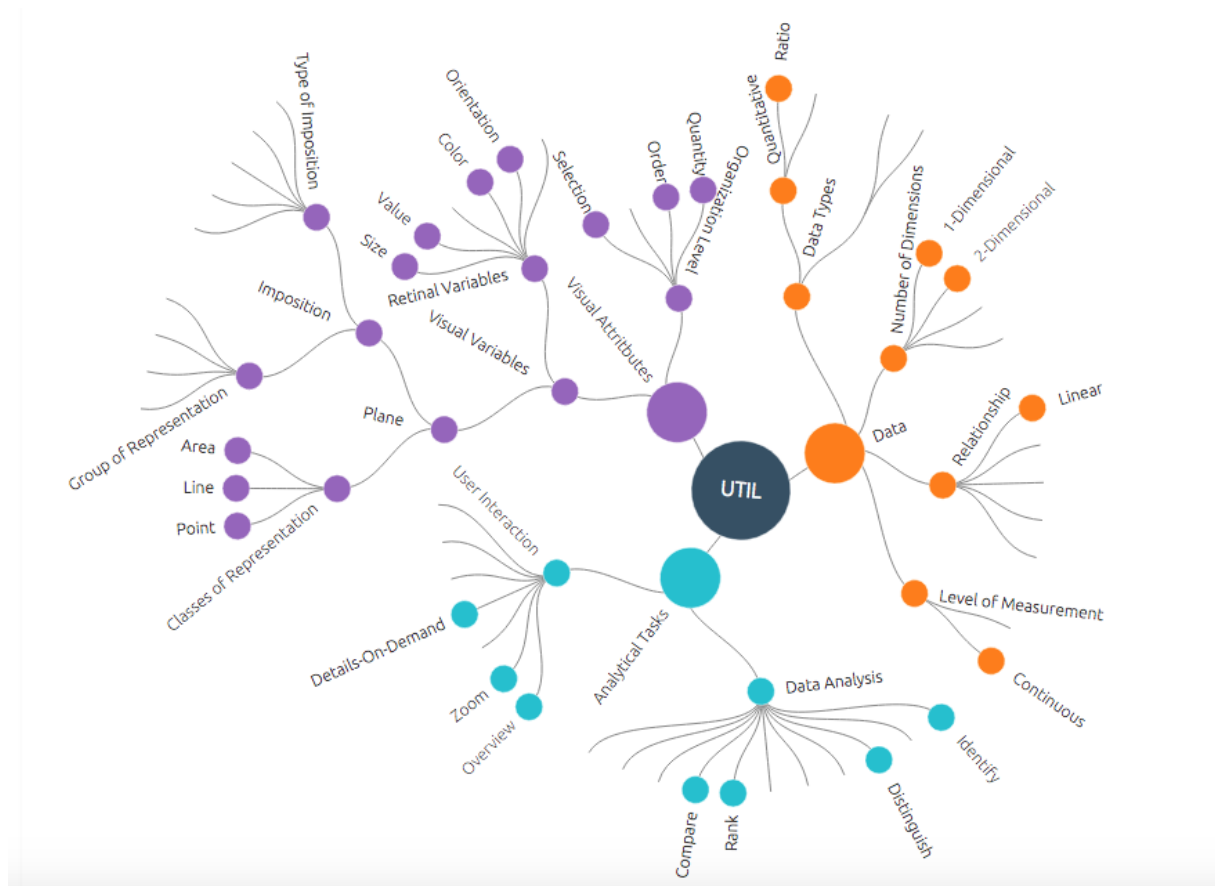
- Não perder o foco do resultado a ser alcançado com o uso da técnica de visualização no contexto em que o trabalho estiver sendo desenvolvido. Por exemplo, para este trabalho, espera-se que as visualizações de informação possam ser utilizadas como mecanismo efetivo para tomada de decisões no contexto de gestão do ambiente SIGAC/SIGEPE/SIAPE.
- Realizar uma reflexão prévia dos aspectos da visualização que serão avaliados após sua construção. Por exemplo, no contexto dos resultados esperados para este trabalho, propõe-se que a avaliação seja realizada seguindo os aspectos levantados no estudo de STASKO, John [2014][17], com o propósito de mensurar o valor agregado alcançado através das "variáveis": tempo, *insight*, essência e confiança.

Considerando o exposto anteriormente, para cada questão de pesquisa apresentada foi realizado uma classificação de técnicas de visualização baseadas nas características com potencial de serem utilizadas para alcançar o resultado desejado. Associado à reflexão sobre a avaliação, selecionou-se uma ou mais técnicas de visualização de informações para representar o conjunto de dados em análise.

A. Qual a quantidade de servidores (ativos e inativos) e pensionistas já utilizam o SIGAC, e o potencial de usuários ainda a ser alcançado?

Com o auxílio da ferramenta UTIL, os itens (eixos de dados, tarefas analíticas e atributos visuais) apresentados na FIGURA 42 podem fazer parte da técnica de visualização a ser utilizada para responder a questão apresentada.

FIGURA 42 - Classificação dos itens identificados para a questão A.



FONTE: Elaborada pelo autor

Sobre a dimensionalidade, a visualização pode apresentar a informação de forma unidimensional, na forma de um indicador ou em duas dimensões, associando os usuários à respectiva situação funcional (ativo, inativo, pensionista) ou até mesmo ao aspecto temporal (aquisição de usuários ao longo do tempo).

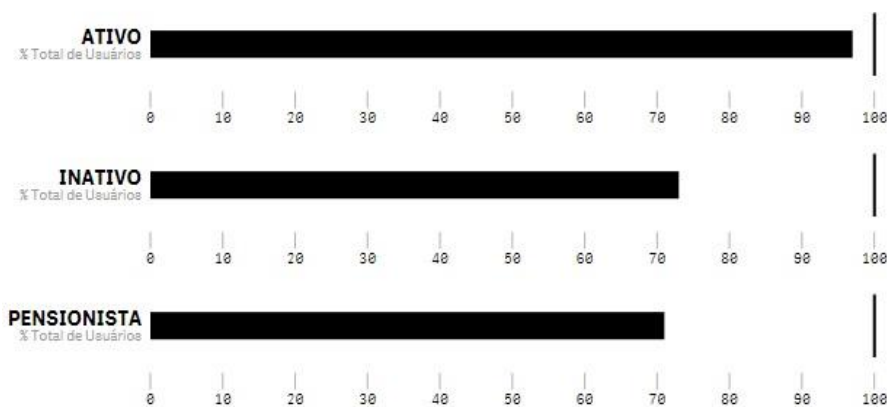
Em relação aos atributos visuais, importante destacar o item cor. No contexto dos dados em análise, há três situações funcionais bem definidas. Neste sentido, pode ser útil a atribuição de uma cor que caracterize cada uma destas situações, de tal forma a ser utilizada em diferentes visualizações, criando uma associação entre da variável com a respectiva cor, auxiliando a interpretação das informações por parte dos usuários.

Ainda sobre o tema cor, é recomendável que seja adotado um esquema de cores que possa ser identificado por pessoas com quaisquer deficiências visuais. Aproximadamente 8% da população masculina e 0,5% da população feminina possuem algum tipo de daltonismo.⁶

Considerando o resultado a ser alcançado com esta questão, a utilização de uma técnica em que a informação seja interpretada de forma rápida alcança um maior valor agregado para o usuário, em contrapartida a técnicas que disponibilizem recursos diversos (por exemplo, detalhes sob demanda e zoom), mas que não privilegiam o requisito de tempestividade.

A partir desta visão, uma técnica bastante interessante, proposta por FEW, Stephen C. [1], e que pode ser utilizada para responder à questão, é a *bullet graph*. Esta técnica é uma variação do gráfico de barras. Na FIGURA 43 está apresentado o resultado a visualização proposta.

FIGURA 43 - Bullet graph - Quantidade atual e percentual alcançado de servidores e pensionistas, em relação ao total possível, que estão utilizando o SIGAC



FONTE: Elaborada pelo autor

O *bullet graph* apresenta uma única medida (no exemplo, a quantidade de usuários), compara essa medida ao um valor alvo (100% dos usuários, no exemplo), além de permitir comparar a outros contextos (situações funcionais diferentes).

A esquerda da figura são apresentados os rótulos dos textos (Ex: ATIVO, INATIVO e PENSIONISTA). Há uma escala quantitativa, de 0% a 100%. A barra em cor preta representa o desempenho da respectiva medida e o marcado de cor preta na vertical, o valor alvo a ser alcançado.

Um dos benefícios desta técnica é possuir um design otimizado, sem perder granularidade de informação.

De forma a complementar a informação desejada, pode ser acrescentado à visualização um indicador do quantitativo atual do total de usuários do ambiente e indicadores de quantitativo segmentado pelas respectivas situações funcionais.

Um dos desafios de utilização do *bullet graph* é a sua disponibilidade de forma nativa nas ferramentas comerciais. Para algumas ferramentas de construção de painéis, o *bullet graph* está disponível somente através de extensões desenvolvidas por terceiros, o que nem sempre é suportado oficialmente pelo fornecedor, tornando impeditivo o seu uso no dia a dia. Porém, os conceitos utilizados por FEW, Stephen C. [\[1\]](#) no desenvolvimento do *bullet graph* podem ser agregados algumas técnicas de visualizações mais comuns e alcançar resultados semelhantes.

Por exemplo, no contexto do exemplo da questão em análise, foi utilizado um gráfico de barras horizontais, sendo cada barra representada por uma situação funcional. A magnitude da barra representa o percentual de usuários que utilizam o SIGAC em cada uma das respectivas situações funcionais. O gráfico foi ordenado de forma a indicar um ranking decrescente da quantidade de usuários. Logo acima do gráfico, foram utilizados indicadores do quantitativo atual de cada uma das respectivas situações funcionais e um valor global. O resultado desta proposta está apresentado na FIGURA 44.

Adicionalmente, optou-se por utilizar uma diferenciação de cores para cada situação funcional.

FIGURA 44 - Quantidade de usuários do SIGAC, em relação ao potencial do seu público alvo.



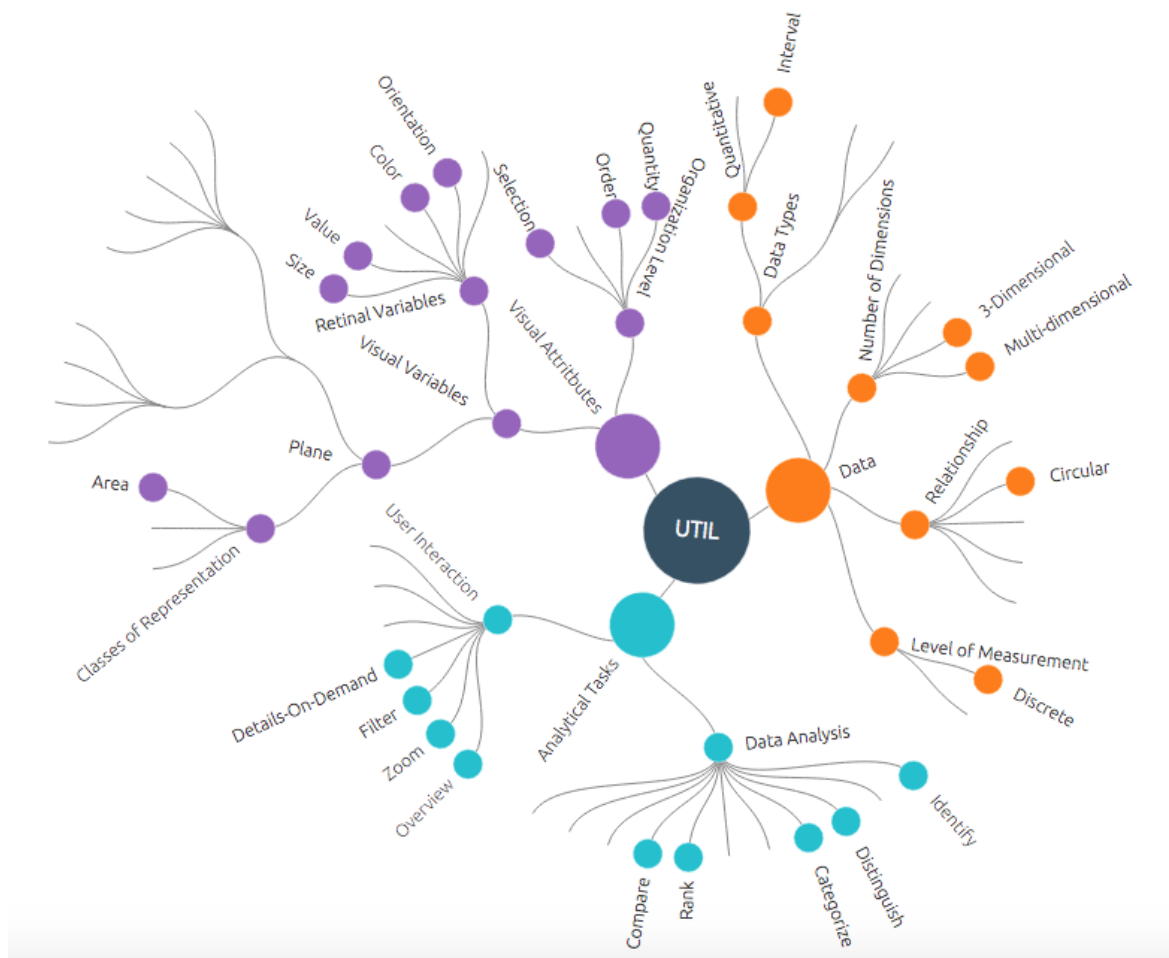
FONTE: Elaborada pelo autor

A visualização de informações apresentada na FIGURA 44 permite, de forma intuitiva, transmitir ao usuário, por exemplo, que a situação funcional que possui a maior quantidade de usuários são os "ATIVOS" e também a maior penetração (97,03%).

B. Qual a faixa etária e sexo dos usuários do SIGAC e respectivos quantitativos?

A FIGURA 45 apresenta a classificação dos itens que podem fazer parte da técnica de visualização a ser utilizada para responder a questão apresentada.

FIGURA 45 - Classificação dos itens identificados para a questão B



FONTE: Elaborada pelo autor

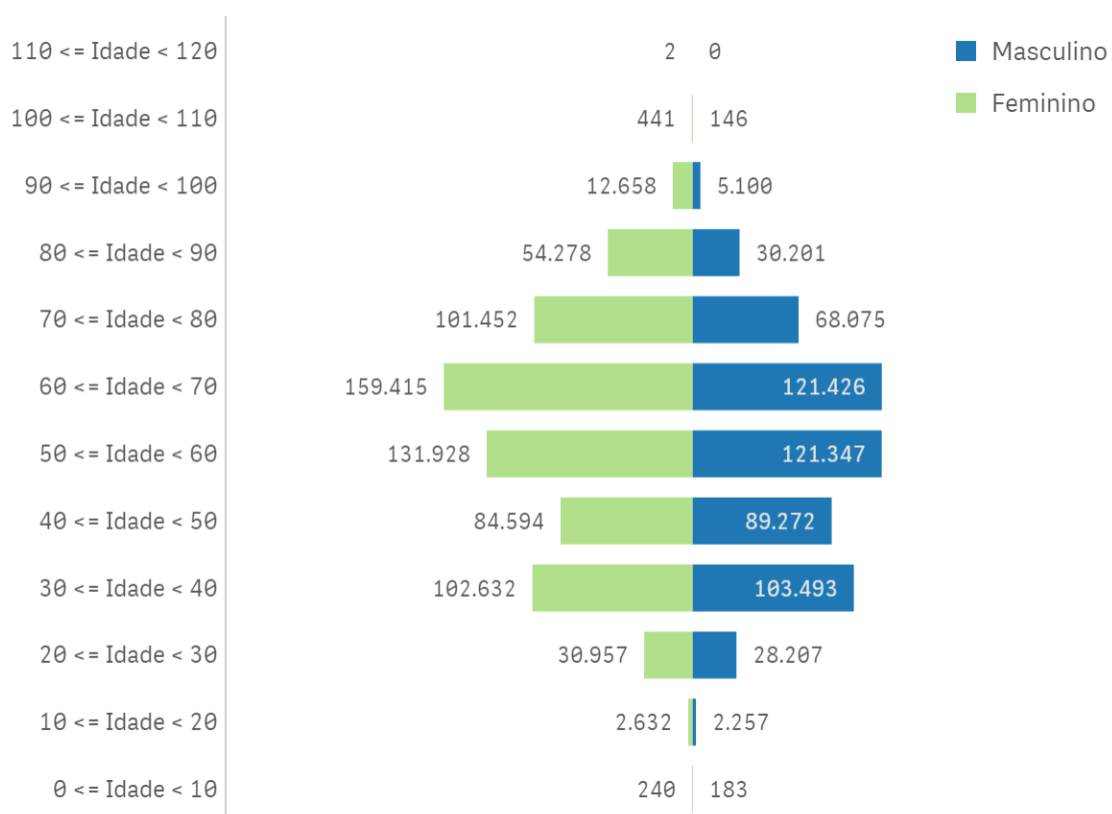
A questão proposta remete a uma análise de distribuição envolvendo a quantidade de usuários, segmentado por faixa etária e sexo (masculino/feminino). Adicionalmente, considerando o contexto do conjunto de dados, pode-se agregar na visualização a informação de situação funcional, com potencial de enriquecer o rol de análises possíveis.

Dentre as visualizações voltadas à representação de distribuição, o histograma é uma opção adequada para responder aos questionamentos apresentados. Por meio dele é possível representar a distribuição de diferentes grupos etários de usuários do SIGAC, cuja altura das barras é proporcional à quantidade que representa em relação ao todo.

A diferenciação por sexo é realizada utilizando dois conjuntos de barras, aplicando-se cores distintas.

A informação de situação funcional pode compor a visualização na forma de opção de filtros, estimulando o usuário a interagir com o gráfico. A FIGURA 46 é o resultado desta técnica de visualização.

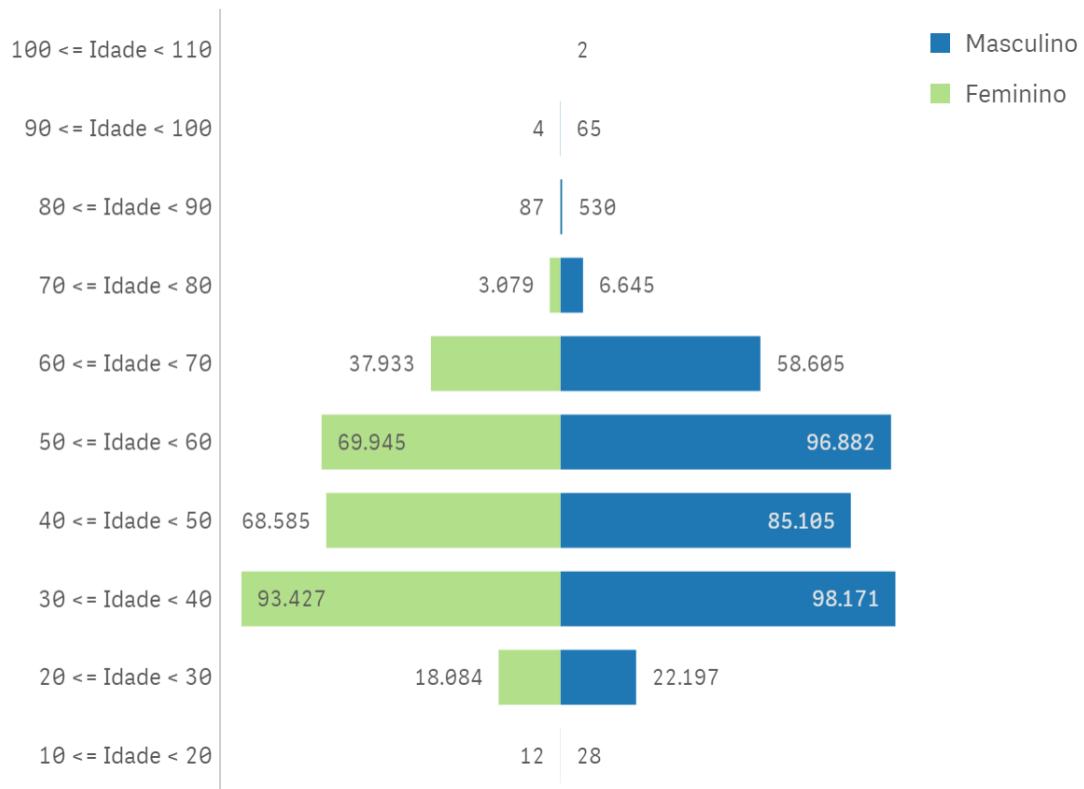
FIGURA 46 - Histograma representando a distribuição dos usuários SIGAC por faixa etária e sexo.



FONTE: Elaborada pelo autor.

Quando aplicado os filtros de situação funcional, por exemplo ativos, a distribuição do histograma apresenta outro comportamento, conforme apresentado na FIGURA 47.

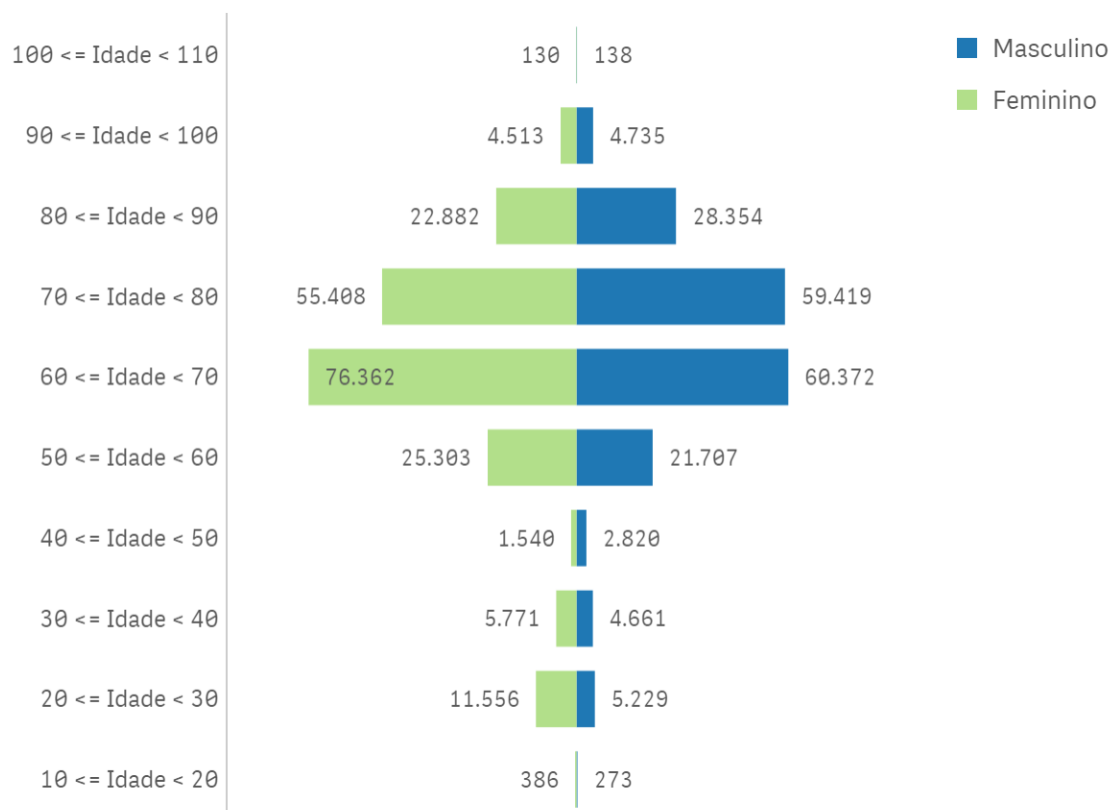
FIGURA 47 - Histograma representando a distribuição dos usuários SIGAC cadastrados com situação funcional "ATIVOS", segmentado por faixa etária e sexo.



FONTE: Elaborada pelo autor.

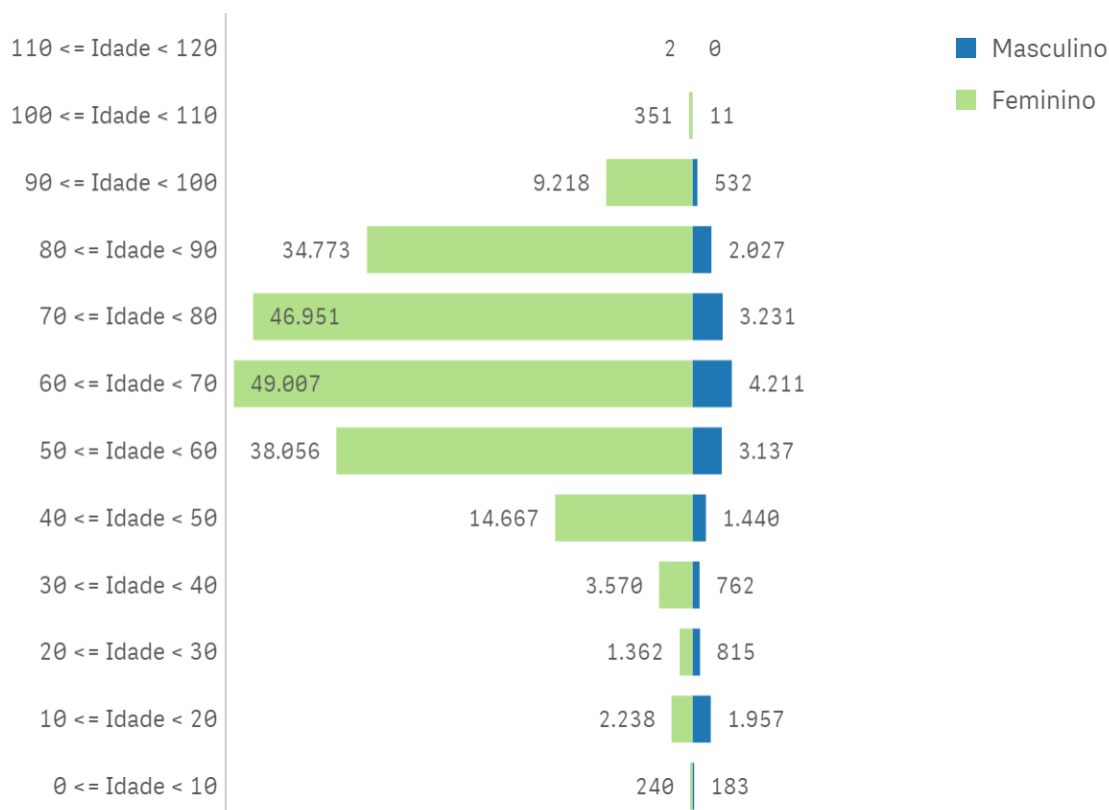
De forma semelhante, quando aplicados os filtros de situação funcional "INATIVOS" e "PENSIONISTAS", FIGURA 48 e FIGURA 49, respectivamente.

FIGURA 48 - Histograma representando a distribuição dos usuários SIGAC cadastrados com situação funcional "INATIVOS", segmentado por faixa etária e sexo



FONTE: Elaborada pelo autor.

FIGURA 49 - Histograma representando a distribuição dos usuários SIGAC cadastrados com situação funcional "PENSIONISTAS", segmentado por faixa etária e sexo.



FONTE: Elaborada pelo autor.

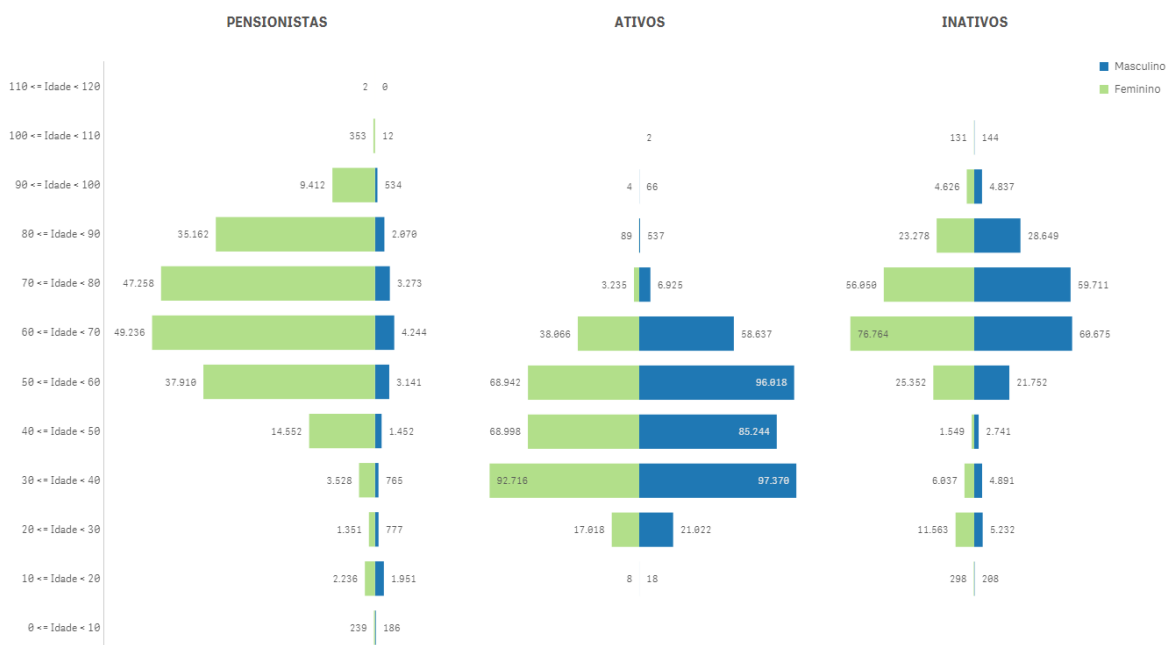
A representação do histograma anterior é também conhecida como pirâmide etária, e é utilizado com maior frequência em outros contextos, não apenas para monitorar a estrutura de sexo e idade, mas como um complemento aos estudos da qualidade de vida de uma população, já que é possível, por exemplo, visualizar a média do tempo de vida, a taxa de mortalidade e a regularidade, ou não, da população ao longo do tempo.

No contexto do conjunto de dados do SIGAC, a visualização contribui, entre outros, para um mapeamento de como estão distribuídos os seus usuários (e particularidades de cada grupo em específico) e instiga a investigação sobre a existência de eventuais inconsistências cadastrais (por exemplo, usuários na faixa etária de 110 e 120 anos).

De forma alternativa, ao invés de optar pela utilização de filtros, pode-se utilizar a representação por pequenos múltiplos. Nesta visualização os dados são separados em

múltiplos gráficos de forma que estejam disponíveis no painel ao mesmo tempo, próximos um do outro, permitindo a análise dos gráficos mais facilmente. Desta maneira, a comparação de valores e padrões entre os gráficos é relativamente mais simples.

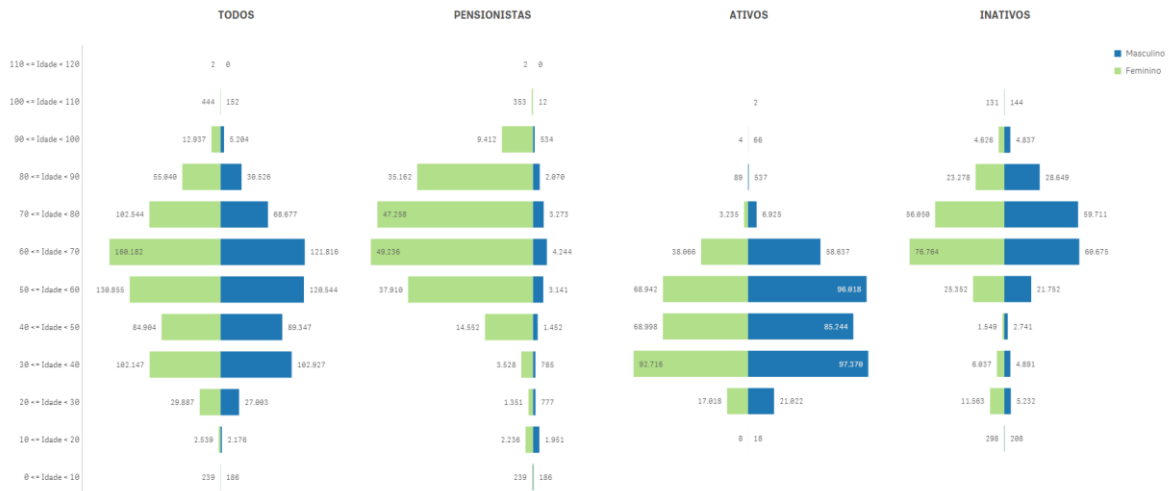
FIGURA 50 - Pequenos múltiplos representando a distribuição dos usuários SIGAC de acordo com a situação funcional, segmentado por faixa etária e sexo.



FONTE: Elaborada pelo autor.

A FIGURA 51 é uma representação de pequenos múltiplos semelhante à FIGURA 50, porém acrescentando a pirâmide etária que agrupa todas as situações funcionais (ativo, inativo e pensionista) simultaneamente.

FIGURA 51 - Pequenos múltiplos representando a distribuição dos usuários SIGAC de acordo com a situação funcional, segmentado por faixa etária e sexo. Nesta visualização foi incluído a representação de todas as situações funcionais agrupadas,

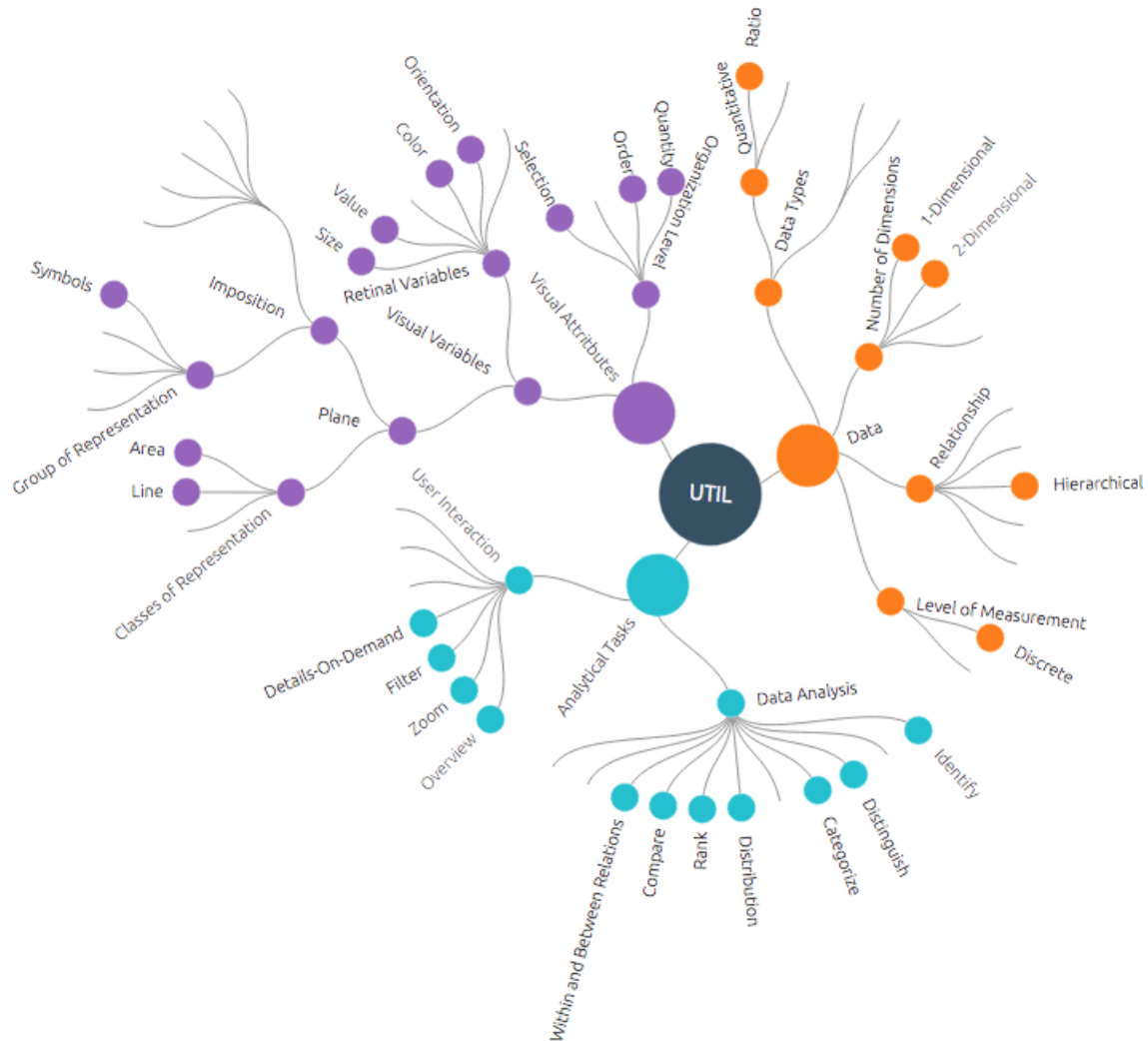


FONTE: Elaborada pelo autor.

C. Qual o comportamento dos status das identidades dos usuários (ativa, bloqueada, expirada, desabilitada, etc) ao longo do tempo?

A FIGURA 52 apresenta a classificação dos itens que podem fazer parte da técnica de visualização a ser utilizada para responder a questão apresentada.

FIGURA 52 – Classificação dos itens identificados para a questão C



FONTE: Elaborada pelo autor.

A questão proposta envolve a análise de variação da situação do usuário no SIGAC ao longo do tempo, a partir de dados extraídos diariamente. O objeto de interesse neste caso é identificar padrões e alterações de tendências ao longo do tempo. Apenas para contextualizar, no SIGAC, a identidade de um usuário pode assumir diferentes status (ativo, certificado digital, bloqueio por erro de senha, bloqueio por erro de dados cadastrais, expirada, bloqueio manual, novo acesso, primeiro acesso, desabilitada e excluída), a depender da ação executada pelo usuário (por exemplo, seguidas tentativas de acesso com senha incorreta altera o status para bloqueio por erro de senha).

Para identificar padrões e alterações de tendências ao longo do tempo, a utilização de um gráfico de linhas é apropriado. Para responder a questão proposta com maior completude, é também de interesse que seja permitido comparar estes padrões e tendências entre as respectivas variáveis (status da identidade).

Uma técnica de visualização que agrega muito destes atributos é a *sparkline*. É uma representação de um gráfico de linha, agrupado de tal forma a permitir comparação entre variáveis num mesmo intervalo de tempo. Associado a um contexto, esta representação pode ser construída de tal forma a instigar a descoberta de diferentes características no conjunto de dados, tais como magnitude dos valores, magnitude da mudança, taxa de mudança e grau de variação.

FIGURA 53 - Técnica de visualização sparkline, associada a outros indicadores, utilizada para representar o comportamento da identidade do usuário no sistema SIGAC.

Status	Min	Últimos 12 meses	Max	Atual	%
Ativa	1.037.809		1.124.284	1.103.538	80,53%
Certificado	14.175		20.049	19.788	1,44%
Bloqueada	5.234		8.262	7.894	0,58%
Bloqueada	16.442		22.071	21.751	1,59%
Bloqueada	4		33	16	0,00%
Expirada	16.199		131.553	67.940	4,96%
Novo Acesso	4.700		9.217	9.217	0,67%
Primeiro	28.839		33.384	28.839	2,10%
Desabilitada	65.583		98.636	98.636	7,20%
Excluída	5.759		12.768	12.768	0,93%

FONTE: Elaborada pelo autor.

Na FIGURA 53, a técnica *sparkline* representa o comportamento de cada variável ao longo de 12 meses. A disposição dos gráficos estimula a comparação entre as variáveis e permite identificar padrões de mudanças ao longo do tempo. No caso em específico,

a grandeza das variáveis são bastante distintas, algumas representando dezenas e outras milhares. A estratégia adotada neste caso foi compor a visualização com indicadores numéricos, identificando os valores mínimo, máximo e atual de cada uma das variáveis.

Por fim, com objetivo de agregar informação adicional de distribuição, acrescentou-se indicadores percentuais da presença de cada uma das variáveis em relação ao todo.

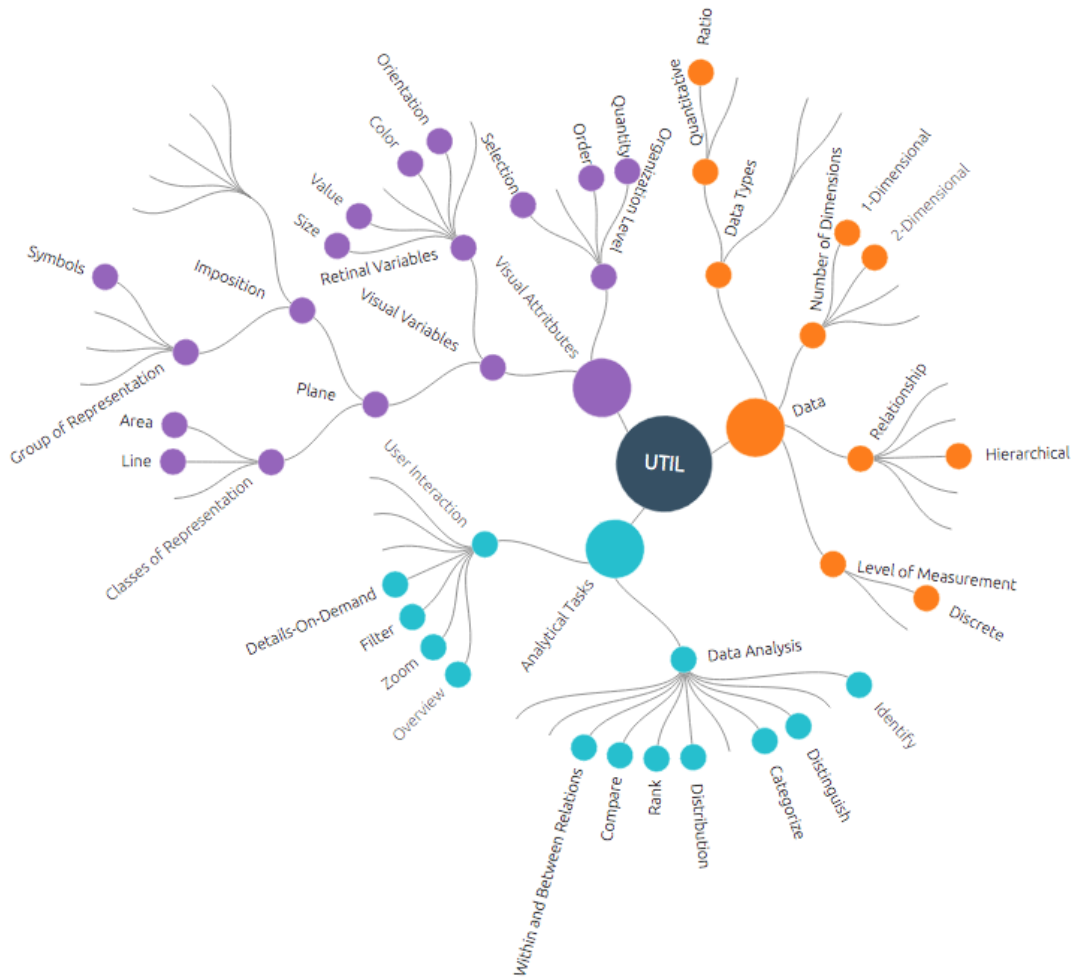
No caso em específico também foi aplicado o recurso de diferenciação de cores, de tal forma a agrupar variáveis com significados próximos. Por exemplo, as variáveis "excluída" e "desabilitada" englobam usuários que não mais acessam o ambiente.

Apesar de não ser o objetivo principal desta visualização, da forma que ela foi construída no exemplo da FIGURA 53 é possível, navegando o cursor do mouse ao longo do gráfico sparkline, identificar os valores da respectiva variável categórica em um período específico, sendo possível aplicar a seleção de forma instantânea para os gráficos das demais variáveis. Este recurso de interatividade pode ser interessante para uma análise exploratória por parte do usuário.

D. Qual a quantidade e padrão de acesso de usuários ao longo do tempo?

A FIGURA 54 apresenta a classificação dos itens que podem fazer parte da técnica de visualização a ser utilizada para responder a questão apresentada. Os itens são basicamente os mesmos explorados para a questão C. O diferencial da abordagem adotada está no contexto do conjunto de dados, que, por exemplo, envolve menos variáveis: usuário e período de acesso.

FIGURA 54 - Classificação dos itens identificados para a questão D



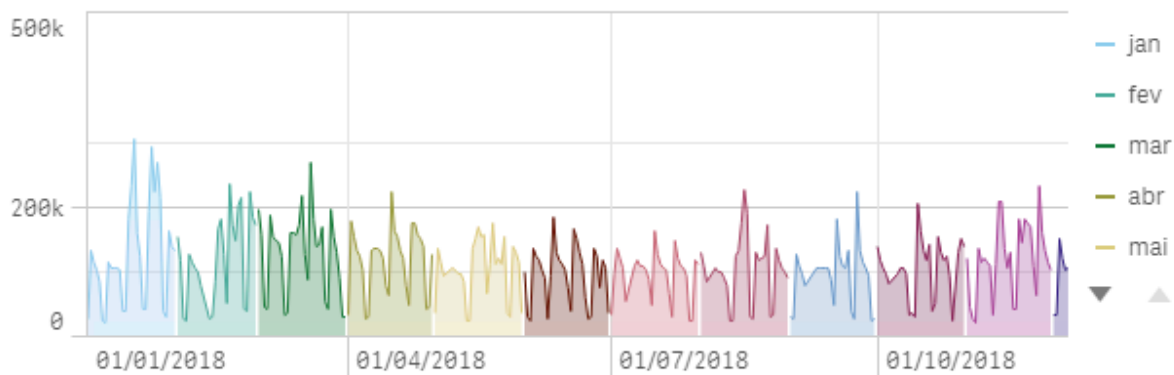
FONTE: Elaborada pelo autor.

Neste caso, novamente o gráfico de linhas é bastante apropriado e intuitivo para representar o comportamento ao longo do período. O que torna poderoso o uso desta técnica, e ao mesmo tempo simples, é representar o mesmo conjunto de dados sobre diferentes perspectivas.

A FIGURA 55 e FIGURA 56, que apresentam visualmente o quantitativo e comportamento do acesso diário de usuários em diferentes anos, permite identificar a existência de picos e vales e determinados dias. A técnica de visualização permite ao usuário interagir com o gráfico (selecione dias específicos, por exemplo), estimulando análises e hipóteses dentro do respectivo contexto.

FIGURA 55 - Gráfico de linhas representando a quantidade de acessos diários, picos, vales e outliers no ano de 2018.

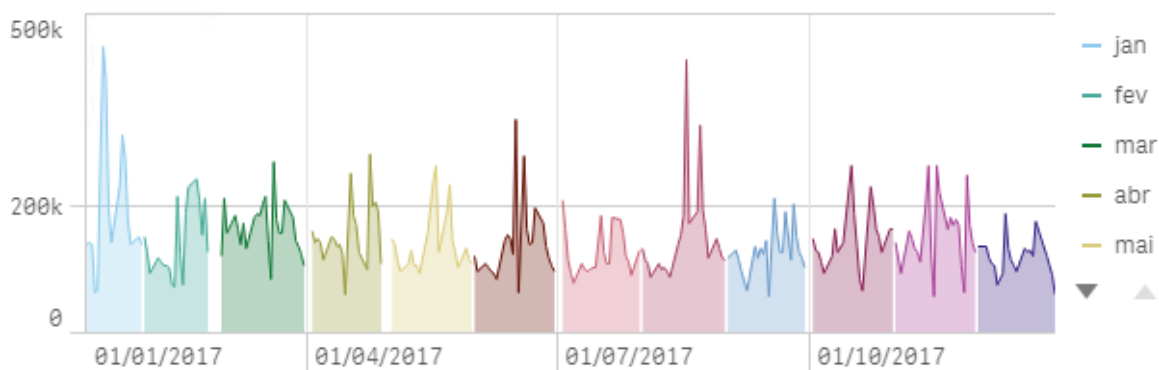
2018 - Qtde de Acessos por Dia



FONTE: Elaborada pelo autor.

FIGURA 56 - Gráfico de linhas representando a quantidade de acessos diários, picos, vales e outliers no ano de 2017.

2017 - Qtde de Acessos por Dia

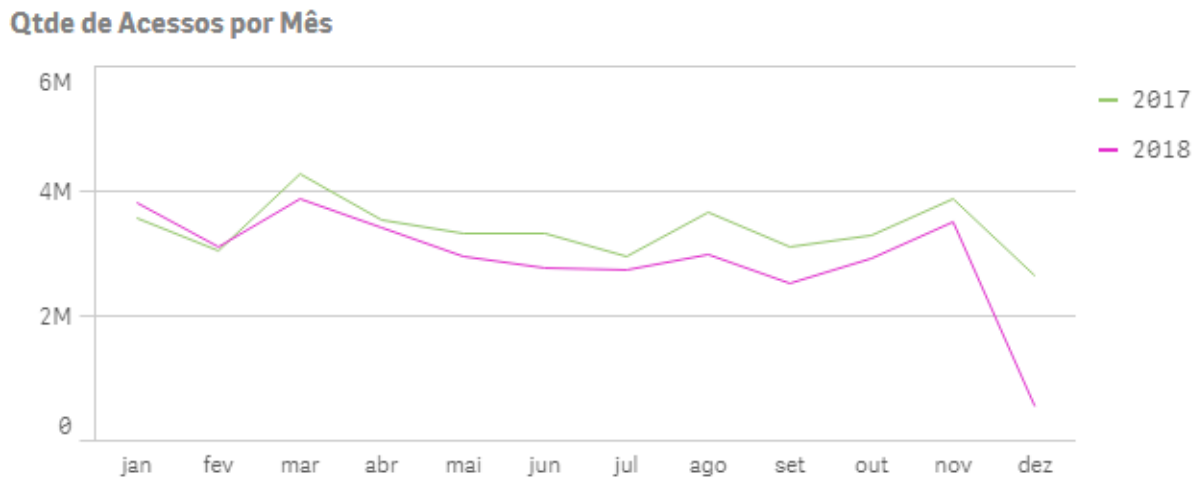


FONTE: Elaborada pelo autor.

A disposição próxima dos dois gráficos permite identificar a existência de um comportamento cíclico em períodos semelhantes. Analisando a escala dos gráficos, há um indicativo que em 2018 há um volume menor de acesso comparativamente à 2017. Também é possível identificar valores fora do padrão (*outliers*), como por exemplo nos meses de janeiro/2017, junho/2017 e agosto/2017.

Alterando a perspectiva de representação diária para uma perspectiva mensal (FIGURA 57) e representando os acessos dos dois anos distintos (2017 e 2018) numa mesma visualização, há um forte estímulo para potencializar a identificação de comportamentos cíclicos no conjunto de dados.

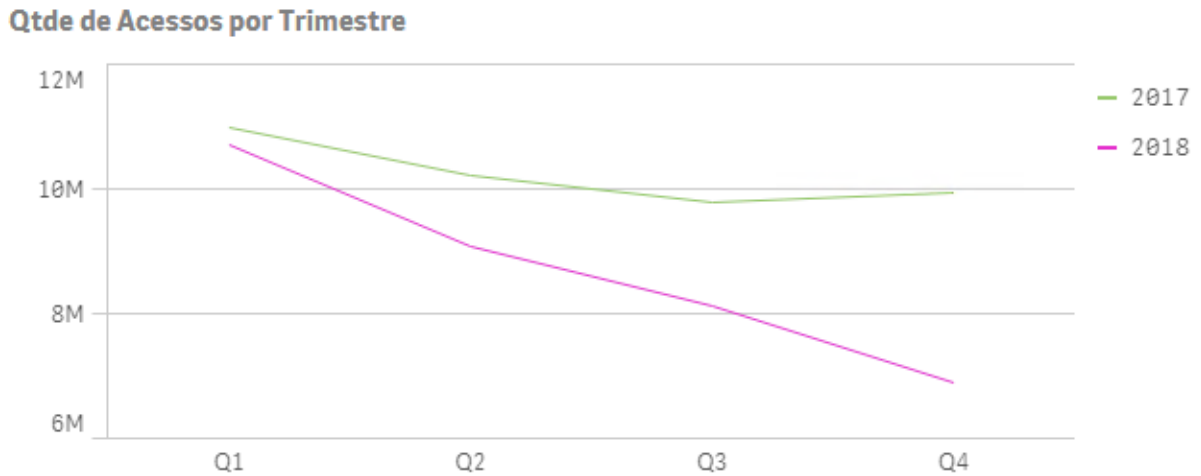
FIGURA 57 - Gráfico de linhas representando a quantidade de acessos mensais no ano de 2017 e 2018.



FONTE: Elaborada pelo autor.

Da mesma forma, agrupando o conjunto de dados numa escala trimestral (FIGURA 58), abre a perspectiva de uma análise diferenciada. No caso em específico, observa-se uma tendência de redução do quantitativo de acessos, que numa análise por meio da representação das FIGURA 55 e FIGURA 56, não é tão intuitivo identificar.

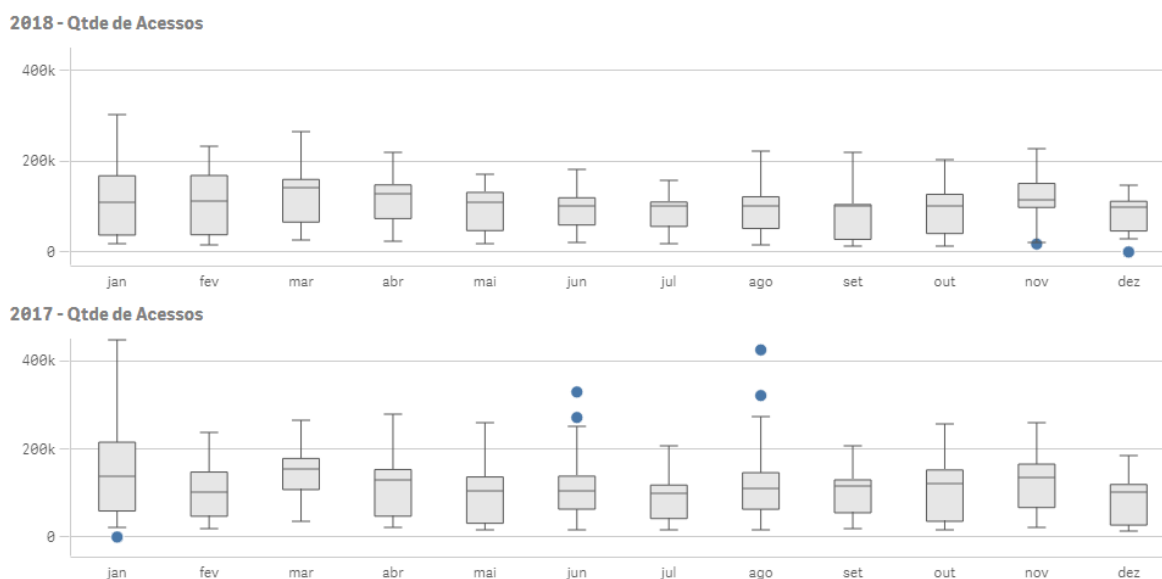
FIGURA 58 - Gráfico de linhas representando a quantidade de acessos trimestrais no ano de 2017 e 2018.



FONTE: Elaborada pelo autor.

De forma alternativa, uma técnica de visualização bastante utilizada em estatística descritiva também poderia ser utilizada para compor às visualizações anteriores. Trata-se do diagrama de caixas ou *box plot* (FIGURA 59). Este diagrama permite identificar rapidamente onde estão localizados 50% dos valores do conjunto de dados, a mediana e os valores extremos. Os espaços entre as diferentes partes da caixa indicam o grau de dispersão, a obliquidade (assimetria) nos dados e os outliers⁷.

FIGURA 59 - Diagrama de caixa representando a quantidade de acessos diários, consolidados por mês, no período de 2017 e 2018.



FONTE: Elaborada pelo autor.

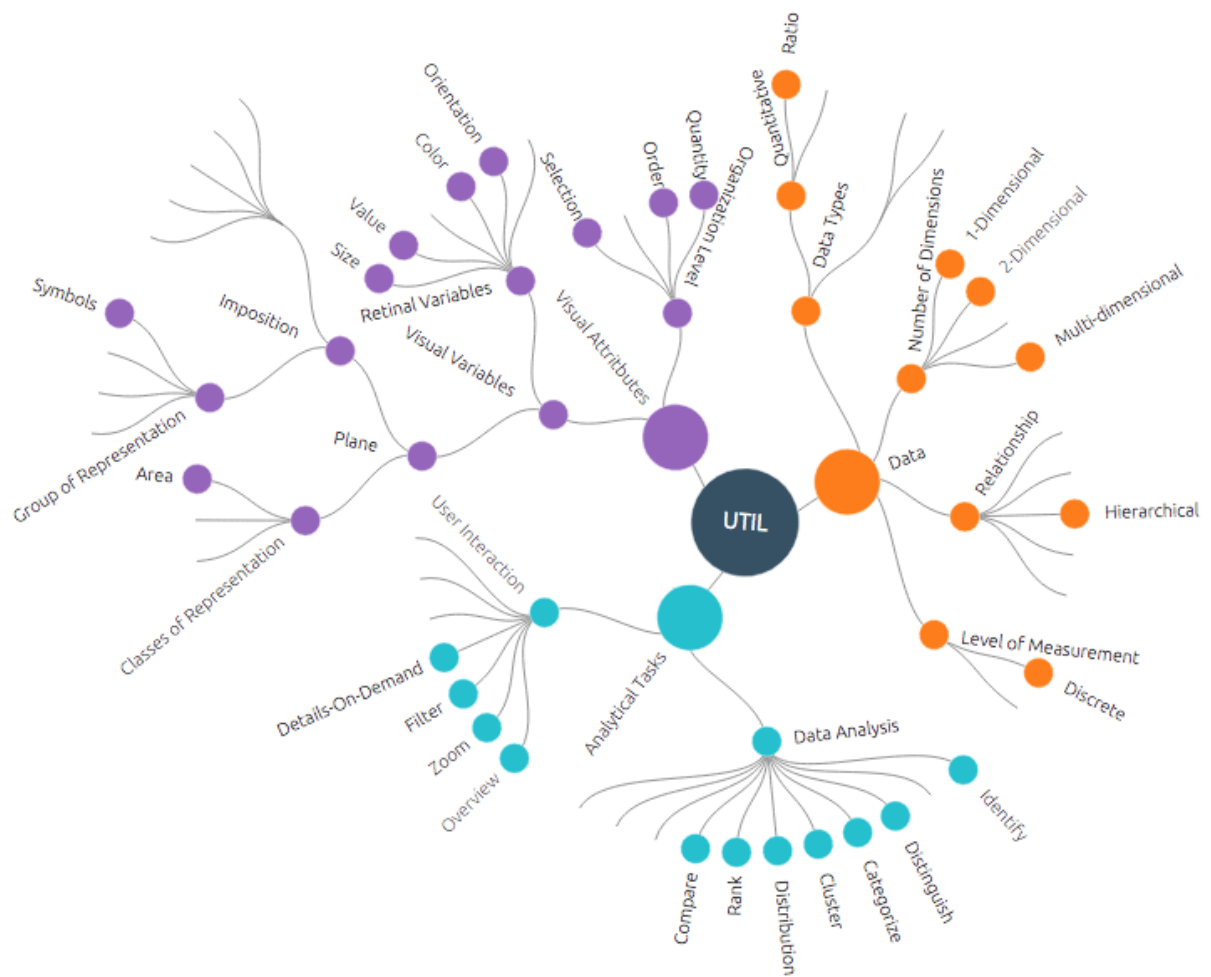
De forma mais rápida, é possível identificar períodos em que ocorreram acessos incomuns (outliers), como por exemplo em junho/2017 e agosto/2017. Ao contrário da percepção inicial após análise da FIGURA 56 de que em janeiro/2017 ocorreram acessos incomuns considerados como outliers, a FIGURA 59 mostra que, apesar um maior volume de acessos, não foram identificados valores outliers.

Ao mesmo tempo que o diagrama de caixa é uma técnica poderosa e bastante utilizada por algumas áreas de conhecimento, demanda conhecimento prévio de como interpretar seus parâmetros. Por tal, é importante conhecer previamente o público alvo antes de utilizá-la no respectivo painel de visualizações.

E. Qual é quantidade de usuários segmentado por sistema e respectivos módulos?

A FIGURA 60 apresenta a classificação dos itens que podem fazer parte da técnica de visualização a ser utilizada para responder a questão apresentada

FIGURA 60 - Classificação dos itens identificados para a questão E.



FONTE: Elaborada pelo autor.

Considerando o contexto e volume do conjunto de dados, o questionamento apresentado e a quantidade de itens que podem ser explorados, uma técnica a ser utilizada é o *treemap*.

De acordo com FEW, Stephen C. [1], o propósito do *treemap* não é realizar comparações quantitativas precisas ou apresentar ranks de itens, mas sim destacar condições particulares de interesse em um grande conjunto de valores.

Complementando os *insights* gerados por meio da utilização do *treemap* com alguns filtros de variáveis e indicadores numéricos, foi construída uma visualização que

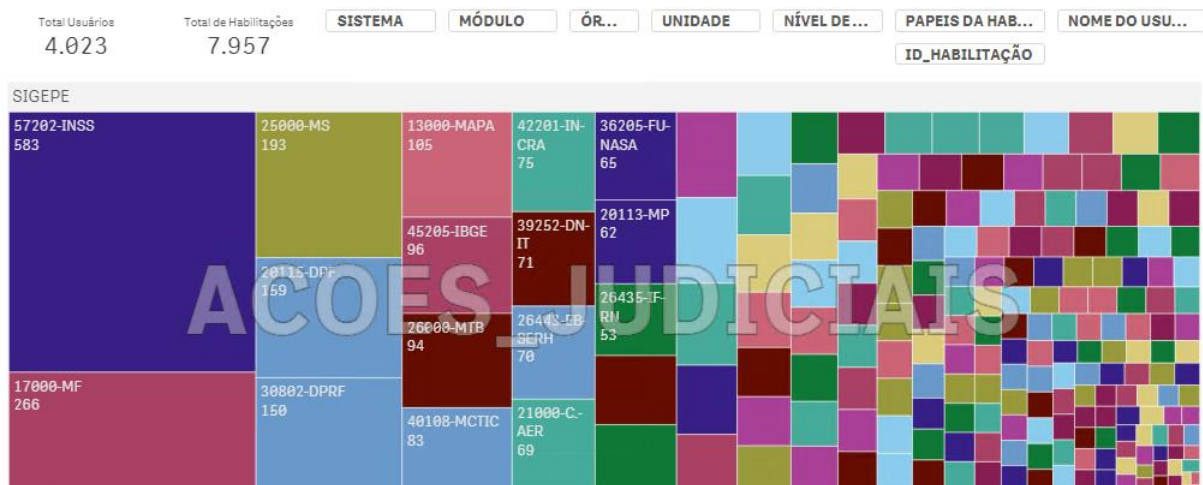
FIGURA 62 - Visualização treemap, com seleção de filtro que exclui o módulo "SERVIDOR".



FONTE: Elaborada pelo autor.

De forma semelhante, quando é aplicado um filtro mais específico, por exemplo, selecionando o módulo "AÇÕES JUDICIAIS", um novo conjunto de informações são apresentados, conforme FIGURA 63.

FIGURA 63 - Visualização treemap, com seleção de filtro em um módulo específico do sistema SIGEPE.

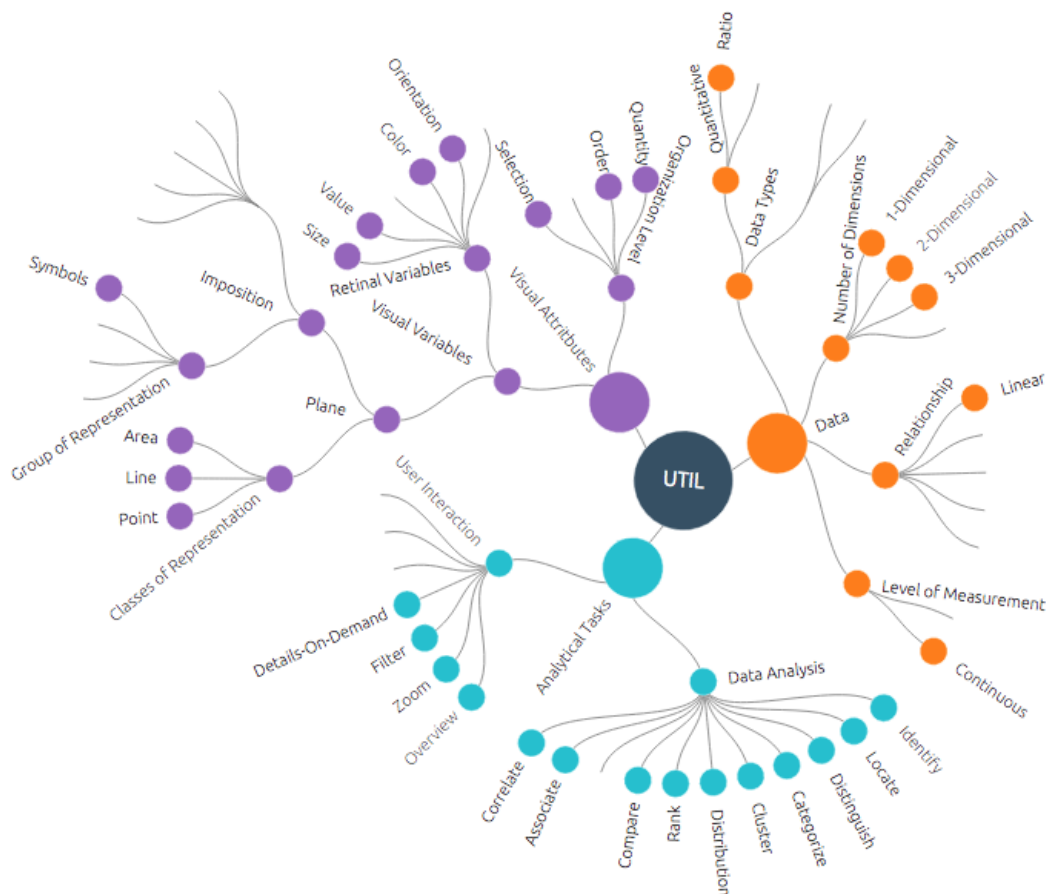


FONTE: Elaborada pelo autor.

F. Qual é o canal de acesso mais utilizado para acessar os sistemas: web ou mobile?

A FIGURA 64 apresenta a classificação dos itens que podem fazer parte da técnica de visualização a ser utilizada para responder a questão apresentada.

FIGURA 64 - Classificação dos itens identificados para a questão F.



FONTE: Elaborada pelo autor.

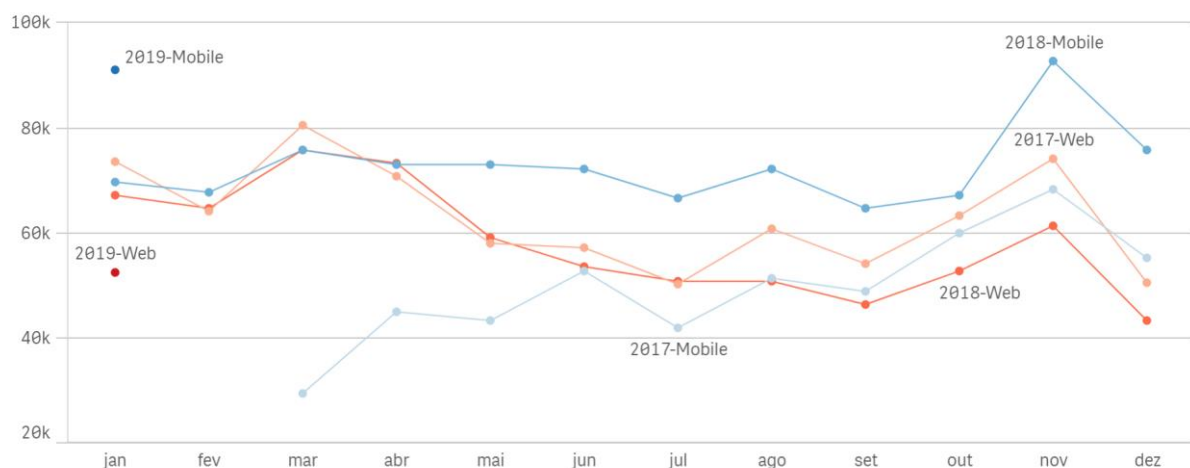
Analisando a classificação da FIGURA 64, observa-se que há o potencial de utilização de diferentes técnicas de visualização. Na etapa anterior, para este questionamento foram pré selecionados os relacionamentos quantitativos de séries temporais e ranking e parte todo.

Uma estratégia inicial foi utilizar um gráfico de linhas sobreposto de forma estimular a comparação das diferentes variáveis em um período semelhante de tempo (anual). Observe que a partir de uma rápida análise da FIGURA 65 é possível extrair várias

informações de forma intuitiva, mesmo sem conhecimento prévio do contexto dos dados. Por exemplo, observa-se que no início de 2017 os acessos Web eram predominantes. Este comportamento foi se modificando, culminando numa inversão ao final do ano de 2017 e sendo acentuado ao longo de 2018 (gráfico em tons de laranja - acessos web vs gráfico em tons de azul - acessos mobile).

FIGURA 65 – Gráfico de linhas representando a média diária de usuários no SIGAC realizando acessos por diferentes canais (Web e Mobile) no período de 2017 e 2018 e início de 2019.

Média diária de usuários

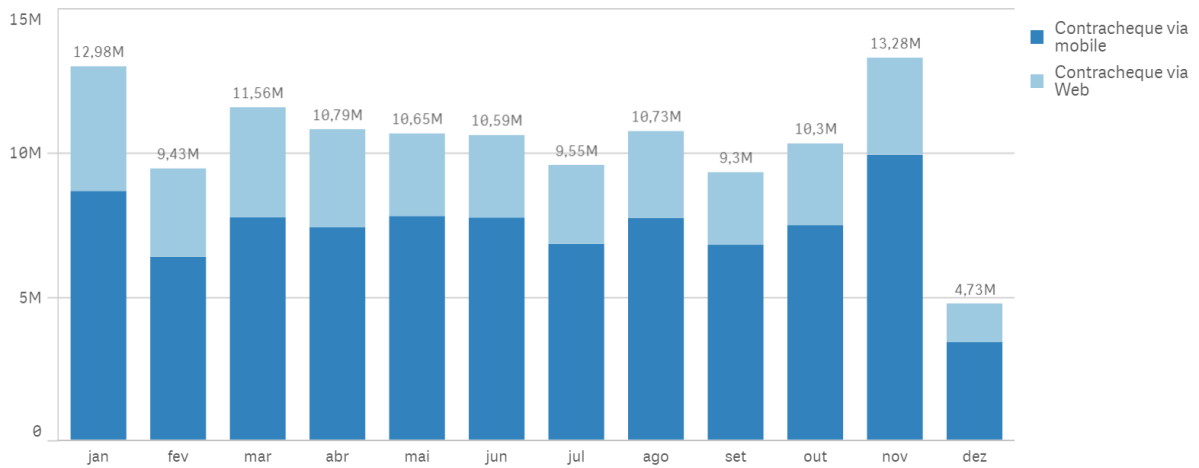


FONTE: Elaborada pelo autor.

A FIGURA 66 mostra a utilização da técnica de gráfico de barras empilhadas representando o acesso a uma funcionalidade do sistema (consulta ao contracheque) por meio de diferentes canais (Web e mobile). Por meio desta técnica, rapidamente é possível intuir, por exemplo, a predominância o uso do canal mobile de forma sistemática ao longo dos diferentes meses.

FIGURA 66 - Gráfico de barras empilhadas representando a consulta ao contracheque por meio de diferentes canais, consolidado por mês.

Consulta ao Contracheque por Canal

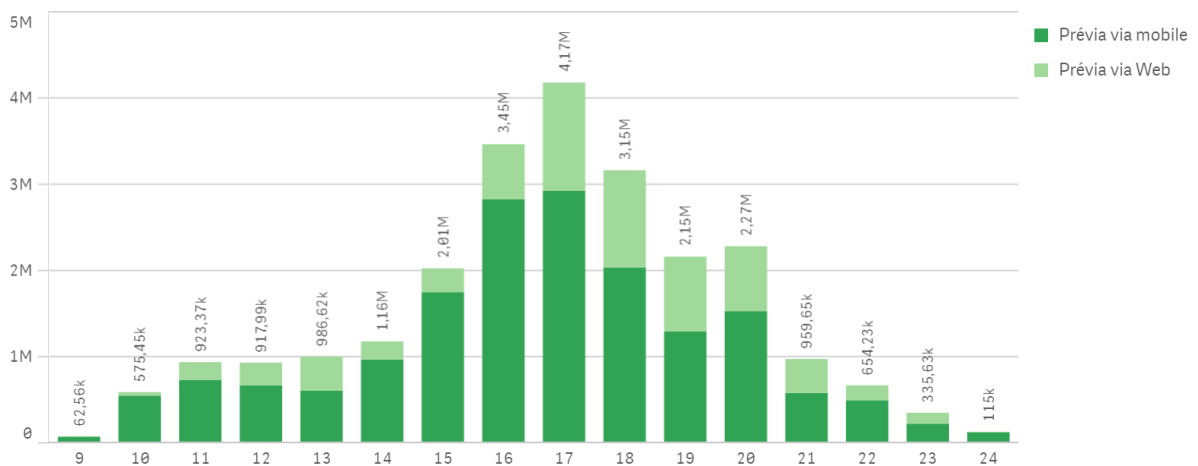


FONTE: Elaborada pelo autor.

A FIGURA 67 mostra a importância de permitir a alternância da escala de análise. Neste caso, ao invés de consolidar as informações por mês, a granularidade aplicada foi o período de um dia. Observe que novas perspectivas de análise são obtidas com esta simples mudança.

FIGURA 67 - Gráfico de barras empilhadas representando a consulta à prévia do contracheque por meio de diferentes canais, consolidado por dia do mês.

Consulta à Prévia do Contracheque por Canal



FONTE: Elaborada pelo autor.

Importante mencionar que várias ferramentas disponibilizam aos próprios usuários recursos para interagirem com a visualização e de forma exploratória, aplicarem estas mudanças, estimulando desta forma a descoberta de novos *insights*.

9 DISPONIBILIZAÇÃO DO PAINEL

As visualizações de informações propostas com o objetivo de responder aos questionamentos fazem parte um subconjunto de visualizações possíveis. Acredita-se que este subconjunto pode servir como referência para uma grande parte das iniciativas da administração pública.

De maneira a verificar a hipótese mencionada anteriormente é necessário que as propostas sejam avaliadas. O primeiro passo, naturalmente é disponibilizá-las em um painel.

FEW, Stephen C. [1], descreve que a localização da informação na tela (*layout*) é um aspecto que não deve ser alterado dinamicamente. Isto porque os visitantes, após utilizarem algumas vezes, tem a expectativa de encontrar dados específicos em locais específicos. Além da questão da localização, alguns aspectos relacionados ao design visual enfatizam de forma diferente a informação disponibilizada no painel.

FIGURA 68 - Diferentes graus de ênfase visual estão associados a diferentes regiões de um painel.

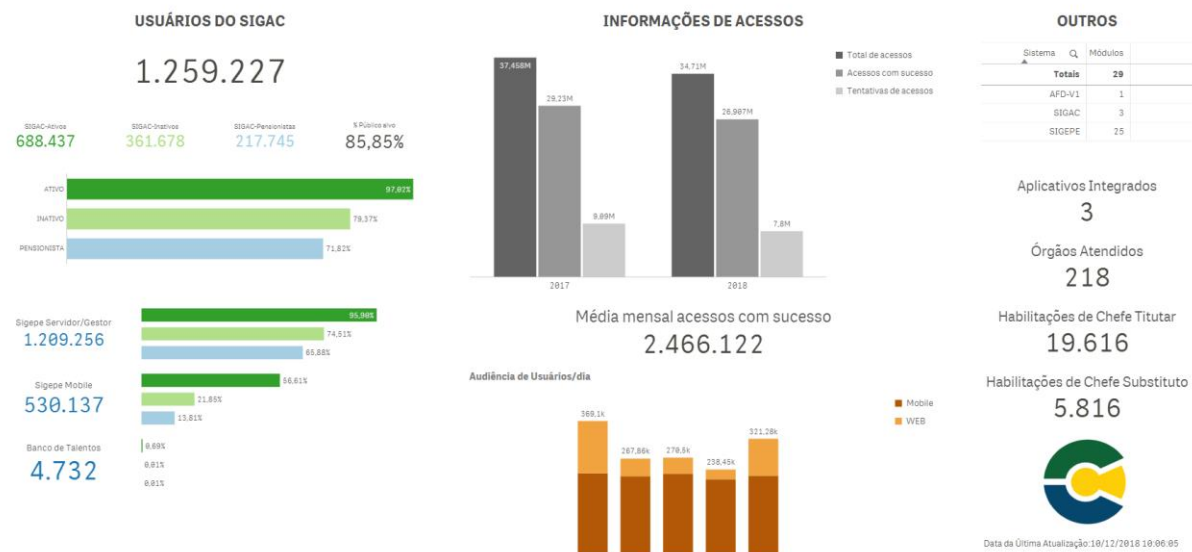
FONTE: Adaptado de FEW, Stephen C.[1].

Desta forma, de acordo com o exposto na FIGURA 68, no quadrante superior esquerdo (região de maior ênfase) deve estar localizado o conteúdo que se deseja maior destaque.

Como exemplo, a FIGURA 69 apresenta uma visão geral do painel com um conjunto de visualizações representando dados do SIGAC, associados a dados SIGEPE/SIAPE.

FIGURA 69 - Visão geral do painel com um conjunto de visualizações representando dados do SIGAC.

Visão Geral



FONTE: Elaborada pelo autor.

10 ESTRATÉGIA DE AVALIAÇÃO

Baseado nos dados consolidados na revisão bibliográfica sobre avaliação, entende-se que a estratégia mais adequada para alcançar a finalidade de utilização de visualização de informação no âmbito da administração pública como mecanismo efetivo para tomada de decisões, promoção do controle social e transparência das ações públicas, é utilizar a abordagem proposta por STASKO, John [2014][\[17\]](#). Ou seja, avaliar em que medida a visualização proposta atingiu o propósito desejado.

O desafio posto então é como operacionalizar a avaliação dos quatro recursos avaliados na equação " $V = T + I + E + C$ ". Para o caso do Painel SIGAC, a avaliação foi realizada por 5 (cinco) usuários chaves, com o acompanhamento de um especialista na construção da visualização e com conhecimento do contexto dos dados, seguindo basicamente duas etapas:

1ª etapa: utilizando um protocolo "aberto" e uma análise qualitativa sobre a interação de usuários chaves com as visualizações propostas.

2ª etapa: direcionando questionamentos aos usuários, com foco principal nas questões de pesquisas e respostas esperadas a partir da interação com as visualizações.

O protocolo "aberto" da 1ª etapa consistiu em observar cada um dos usuários individualmente interagindo com cada visualização disponibilizada no painel. O objetivo foi identificar quais informações foram extraídas, padrões e relacionamentos encontrados, dificuldades enfrentadas, lacunas de informação, entre outros. Isto sem uma explicação prévia e direcionamento sobre o objetivo inicial da visualização.

Em seguida (2ª etapa), com objetivo de analisar a efetividade da visualização, foram realizados os questionamentos que motivaram a construção da visualização e coletadas as respostas e descobertas apresentadas.

A cada ciclo de avaliação englobando as duas etapas com um usuário, os resultados foram avaliados e adequações nas visualizações e/ou configurações do painel foram executados. Desde simples ajustes, como alterações de títulos da visualização ou

representação da legenda, até a utilização de uma técnica de visualização diferente. Após os ajustes, novos ciclos de avaliação foram executados com novos usuários.

A chave do processo de avaliação foi manter o foco na questão a ser respondida e ao mesmo tempo estimular o usuário obter novas informações a partir da visualização.

Por exemplo, analisando a Figura 34 com objetivo de responder a questão A - "Qual a quantidade de servidores (ativos e inativos) e pensionistas já utilizam o SIGAC, e o potencial de usuários ainda a ser alcançado?", sob a ótica da equação de STASKO, John [2014][17], pode-se ilustrar a contribuição de valor de cada variável da equação.

Na variável tempo (T), o gráfico de barras com o indicativo de percentual responde quase de imediato ao questionamento. Em conjunto com os indicadores numéricos, relacionados pela associação de cores, informam aos usuários a grandeza de cada umas das categorias de usuários sem a necessidade de muita interação.

Na variável *insight* (I), as informações entregues pelo conjunto de visualizações estimularam os usuários a desenvolver questionamentos adicionais (por exemplo utilização dos canais web e mobile). Alguns questionamentos foram respondidos com este mesmo conjunto de visualizações. Outros demandaram complemento com visualizações adicionais.

Para a variável essência (E), observou-se que esta parte do painel transmitiu aos usuários a essência geral do conjunto de dados, o contexto que o SIGAC está inserido, a sua abrangência (por exemplo, 218 órgãos, diferentes sistemas) e estimulou a busca por informações adicionais.

A valor da variável confiança (C) foi melhor percebida durante o processo de construção desta parte do painel. Neste período as inconsistências, erros e valores incompletos foram identificados e soluções foram aplicadas. A consequência deste processo foi a apresentação de um conjunto de visualizações que agregou maior confiança para os usuários finais.

11 CONCLUSÕES

A Organização para a Cooperação e o Desenvolvimento Econômico (OCDE) desenvolveu um modelo de habilidades para a inovação do setor público, que inclui 6 áreas de competências essenciais⁸. Entre essas áreas de competências é citada a "alfabetização de dados: garantir que as decisões sejam orientadas por dados e que estes não sejam utilizados apenas no final de um processo. O letramento em dados significa que, sempre que possível, as decisões devem ser baseadas em dados, não intuições ou suposições. Os dados não são apenas para "geeks", os não especialistas devem entender sua importância."

Neste trabalho foi apresentado uma forma útil de extrair dos dados informações relevantes para subsidiar a tomada de decisões pela gestão pública, através do uso de técnicas de visualização de informação. Foram apresentadas um conjunto de técnicas e os respectivos processos para suas escolhas, selecionados dentro de várias opções disponíveis atualmente.

Uma das contribuições deste trabalho é mostrar um processo para a escolha de uma visualização, que deve abranger os dados, as tarefas analíticas e atributos visuais, de acordo com a taxonomia proposta por BARROS [12].

Adicionalmente, o trabalho também traz uma contribuição no sentido de destacar a importância e os desafios de mensurar a efetividade de uma visualização. Faz uma proposta de mensurar em que medida a visualização proposta atingiu o propósito desejado baseada na abordagem proposta por STASKO, John [2014][17]. As variáveis da equação de valor (Tempo, *Insight*, Essência e Confiança) são bem apropriadas para o contexto da gestão pública.

Como trabalhos futuros, há a oportunidade de desenvolver uma escala quantitativa para a equação de valor proposta, estabelecendo uma escala de contribuição de cada uma das quatro variáveis.

REFERÊNCIAS

- [1] FEW, Stephen C. *Information Dashboard Design: Displaying data for at-a-glance monitoring*. [S.l.]: Analytics Press, 2013.
- [2] BRASIL. MINISTÉRIO DO PLANEJAMENTO, SECRETARIA DE TECNOLOGIA DA INFORMAÇÃO, Desenvolvimento e Gestão. *Bem vindo - Portal Brasileiro de Dados Abertos*. Disponível em: <<http://dados.gov.br/>>. Acesso em: 7 set. 2018.
- [3] BRASIL MINISTÉRIO DO PLANEJAMENTO. ORÇAMENTO E GESTÃO . *Painel de Preços*. Disponível em: <<http://paineldeprescos.planejamento.gov.br/>>. Acesso em: 7 set. 2018.
- [4] WARE, Colin. *Information Visualization: Perception for Design*. [S.l: s.n.], 2012.
- [5] FRANCISCANI JÚNIOR., Geraldo *et al.* An Annotation Process for Data Visualization Techniques. *Proceedings of the 3rd International Conference on Data Analytics*, 2014. Disponível em: <<https://pdfs.semanticscholar.org/7438/0aa87f9c77a95a8227e8532d041d2a2004bf.pdf>> . Acesso em: 13 nov. 2018.
- [6] SHNEIDERMAN, Ben. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. . [S.l: s.n.], [S.d.]. Disponível em: <<https://www.cs.umd.edu/~ben/papers/Shneiderman1996eyes.pdf>>. Acesso em: 13 nov. 2018.
- [7] CHI, Ed H. *A Taxonomy of Visualization Techniques using the Data State Reference Model*. . [S.l: s.n.], 2000. Disponível em: <<https://pdfs.semanticscholar.org/e353/cd93eefa2a959ea725308dcebb59f870b2b9.pdf>> . Acesso em: 15 nov. 2018.
- [8] HEER, Jeffrey; SHNEIDERMAN, Ben. Interactive dynamics for visual analysis. *Queue*, v. 10, n. 2, p. 30, 2012. Disponível em: <<https://idl.cs.washington.edu/files/2012-InteractiveDynamics-CACM.pdf>>. Acesso em: 15 nov. 2018.
- [9] TORY, Melanie; MÖLLER, Torsten. Rethinking visualization: A high-level taxonomy. 2004, [S.l.]: IEEE, 2004. p. 151–158. Disponível em: <<http://ieeexplore.ieee.org/document/1382903/>>. Acesso em: 15 nov. 2018.
- [10] PFITZNER, Darius; HOBBS, Vaughan; POWERS, David M. W. A Unified Taxonomic Framework for Information Visualization. 2003, [S.l.]: InVis.au, 2003. p. 57–66. Disponível em: <<https://www.semanticscholar.org/paper/A-Unified-Taxonomic-Framework-for-Information-Pfitzner-Hobbs/353e39ec948ccb434fea79d5369384df8a532110>>. Acesso em: 15 nov. 2018.

- [11] GILSON, O *et al.* From Web data to visualization via ontology mapping. *Computer Graphics Forum*, v. 27, n. 3, p. 959–966, 2008. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.143.1630&rep=rep1&type=pdf>>. Acesso em: 15 nov. 2018.
- [12] BARROS, Diego Augusto de Faria. *UTIL: UMA TAXONOMIA UNIFICADA PARA VISUALIZAÇÃO DE INFORMAÇÃO*. 2015. Universidade Federal de Minas Gerais, 2015. Disponível em: <<https://www.dcc.ufmg.br/pos/cursos/defesas/1812M.PDF>>. Acesso em: 15 nov. 2018.
- [13] LAM, Heidi *et al.* *Empirical studies in information visualization: Seven scenarios*. *IEEE Transactions on Visualization and Computer Graphics*. [S.l.: s.n.]. Disponível em: <<http://ieeexplore.ieee.org/document/6095544/>>. Acesso em: 15 nov. 2018. , set. 2012
- [14] ISENBERG, Tobias *et al.* A systematic review on the practice of evaluating visualization. *IEEE Transactions on Visualization and Computer Graphics*, v. 19, n. 12, p. 2818–2827, dez. 2013. Disponível em: <<http://ieeexplore.ieee.org/document/6634108/>>. Acesso em: 15 nov. 2018.
- [15] NORTH, C. Toward measuring visualization insight. *IEEE Computer Graphics and Applications*, v. 26, n. 3, p. 6–9, maio 2006. Disponível em: <<http://ieeexplore.ieee.org/document/1626178/>>. Acesso em: 15 nov. 2018.
- [16] FEW, Stephen. *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. [S.l.: s.n.], 2009. Disponível em: <<http://portal.acm.org/citation.cfm?id=1611401>>.
- [17] STASKO, John. Value-driven evaluation of visualizations. 2014, [S.l.: s.n.], 2014. p. 46–53. Disponível em: <<http://dx.doi.org/10.1145/2669557.2669579>>. Acesso em: 27 nov. 2018.
- [18] EDWARD R. TUFTE. *Beautiful evidence / Edward R. Tufte*. [S.l.]: Graphics Press; 1st edition edition (July 1, 2006), 2006.
- [19] CLEVELAND, William. *Elements of Graphing Data (selections)*. [S.l.]: Hobart Press; 2 edition (October 1, 1994), 1994.

APÊNDICE A

TABELA 6 - Coleção de termos utilizado no estudo de FRANCISCANI JR., Geraldo *et al* [5]

Barras	Caixas	Células
Círculos de Seção	Linhas	Pontos
<i>Ring Sector</i>	Forma	Trilhas
Movimento	Direção	Posição espacial 2D representando quantidades
Posição de agrupamento espacial que representa categorias	Blur	Variação de cores
Curvatura	Enclosure	Variação de orientação
Variação da forma	Variação de Tamanho	Variação de textura
Variação de valores	1D (Dimensional)	2D (Dimensional)
3D (Dimensional)	Multidimensional	Correlação
Desvio	Distribuição	Multivariada
Parte todo	Classificação	Série temporal
Diferenças alternadas	Centro	Covariância
Ciclos	Exceções	Lacunas
<i>Increasingly different</i>	Diferenças não uniformes	Taxa de variação
Espalhamento	Tendência	Uniformidade
Diferenças uniformes	Variabilidade	Direcionado (navegação analítica)
Exploratória (navegação analítica)	Hierárquica (navegação analítica)	Acesso aos detalhes sob demanda
Adicionando variáveis	Agregação	Anotar
<i>Bookmarking</i>	<i>Brushing and linking</i>	Comparação
<i>Drilling</i>	Filtrando	Foco e contexto juntos
Destacando	Re-expressando	Redimensionamento
Re-visualizando	Classificação	<i>Zooming and panning</i>
Agrupando itens por similaridade	Comparação de valores individuais e cumulativos	<i>Multiple concurrent views and brushing</i>
Escala de tempo sobrepostas	Classificação de itens por similaridade	Regiões e linhas de referência
Pequenos múltiplos		

FONTE: Estudo de FRANCISCANI JR., Geraldo *et al*. [2014] [5]

