

Análise de transferências de recursos da União para os Estados a partir de dados  
do Portal de Transparência: um estudo de caso

*Analysis of financial resource transfers from the Brazilian Federal Government  
to the States based on data from the Transparency Portal: a case study*

Rafael Teogenes da Costa<sup>1</sup>

Luiz Felipe Passos de Oliveira<sup>2\*\*</sup>

Resumo

Na sociedade brasileira, é crescente a conscientização das pessoas em relação aos gastos públicos. Não bastando apenas disponibilizar os dados referentes a estes gastos, é notória a necessidade de uma forma simplificada de analisá-los. A proposta deste trabalho é apresentar uma ferramenta de visualização e análise de dados abertos, fazendo um estudo de caso das transferências federais aos estados da federação brasileira. Concluímos que, após a revisão, a ferramenta se mostrou satisfatória em simplificar a análise de uma considerável quantidade de dados. Analisamos três trabalhos que também tratam de visualização de dados, assim como a ferramenta desenvolvida neste trabalho. Com base na literatura e nas ferramentas relacionadas, foram identificados os momentos exatos em que os conceitos básicos de ETL podem ser observados e, por fim, aplicamos critérios de avaliação das visualizações da ferramenta. Na etapa final, ajustamos o aplicativo de acordo com a referida avaliação, com o intuito de melhorar a experiência de uso. Nesta etapa, atingimos o nosso objetivo, de testar o aplicativo desenvolvido utilizando um sistema formal de avaliação previamente proposto. Objetivando ajustá-lo de forma a oferecer a melhor experiência possível para o usuário final, em nosso estudo de caso específico .

Palavras-chave: Análise de dados. Dados Públicos. Visualização de dados.

Abstract

In Brazillian society, the awareness regarding the resources spent by the government is growing rapidly. It is not enough to just make such data available to the people. A simple way to analyze such data is notoriously needed. The purpose of this work is to present a tool of data analysis and visualization of open data, around a specific case study for the federal transfers made to the states of the Brazilian federation.. For its development, Qlik Sense was used. We reviewed three other works that also approach data visualization . Basing ourselves in such concepts and works, we identified the exact moments in which the ETL concepts can be observed and, after the development, we applied an evaluation method to the visualization of the tool developed. As a final step, we adjusted the application according to the evaluation,

---

<sup>1</sup>Trabalho de conclusão de curso apresentado ao curso de Bacharelado em Sistemas de Informação da Universidade Federal Fluminense como requisito parcial para conclusão do curso.

Graduando do Curso de Sistemas de Informação-UFF; <rafaelteogenes@id.uff.br>

<sup>2\*\*</sup> Graduando do Curso de Sistemas de Informação-UFF; <luizpassos@id.uff.br>

trying to mitigate or eliminate all the relevant aspects, aiming to improve the user experience. In this final step we achieved our initial goal of improving our case study using a formal method previously established. Our goal was to adjust our application in order to offer the best possible experience to the final user, in our specific case study.

Keywords: Data Analysis. Open date. Data Visualization.

**Aprovado em: 10/05/2021 Versão Final em: 10/05/2021**

## 1 INTRODUÇÃO

O surgimento da internet trouxe uma nova era de compartilhamento de informações e dados em um volume nunca visto antes. Segundo Sodré (2016), aproximadamente 2,5 quintilhões de *bytes* são criados diariamente a partir de postagens em redes sociais, upload de fotos e vídeos, registros de transações comerciais, sinais de GPS, rastros de navegação e sensores dos mais diversos tipos. E esse volume só tende a aumentar. Gantz e Reinsel (2012) descrevem que a quantidade de informações existentes no universo digital crescerá em até 50 vezes entre os anos de 2010 e 2020 (VIANNA; DUTRA, 2016).

De tarefas simples, como avisos, a tarefas mais complicadas como registro de coletas de impostos, ter em mãos informações baseadas em dados corretos sempre foi fator decisivo para que as organizações sociais humanas prosperassem. Segundo Betser e Belanger (apud ESPÍNDOLA; ROTH, 2015), o que caracteriza o *Big Data* são fatores como frequência e diversidade das fontes desses dados, além do volume, o que amplia as possibilidades de melhorar a compreensão das previsões para o futuro. Ao estruturar esses dados, é possível traçar visualizações de padrões, relações e correlações inusitadas e com isso criar modelos descritivos, diagnósticos e preditivos de alta performance usados como empoderador do tomador de decisão (SODRÉ, 2016). Atualmente, na sociedade brasileira, há uma preocupação crescente com o uso indevido de recursos públicos, tendo em vista que os mesmos fazem muita falta em diversos setores da sociedade. Nesse Estado gerencial e democrático, a questão da transparência passa a ter destaque como um fator de inibição da corrupção e consolidação democrática. Nesse sentido, "a transparência do Estado se efetiva por meio do acesso do cidadão à informação governamental, o que torna mais democrática a relação entre o Estado e a sociedade civil" (BRESSER-PEREIRA, 2004, p. 23). Diversas notícias de desvios de bilhões de reais em recursos públicos chocam a todos. Enquanto as instituições governamentais brasileiras caem em descrédito, um povo cada vez mais apático e

sem esperança se torna refém da falta de recursos públicos em saúde, educação e segurança. Diversos exemplos na história (e atualmente) mostram o poder de uma sociedade organizada e fiscalizável. Nesse sentido, temos no Brasil a Lei de Acesso à Informação (de 2011, entrando em vigor em 2012, ou seja, muito recente), que garante o acesso do cidadão aos dados produzidos pelas entidades governamentais. Ter acesso a esses dados é um primeiro passo muito importante a ser dado, no caminho de um governo com lisura em sua atuação. Contudo, a forma em que os dados são disponibilizados apenas tem o intuito de cumprir a lei, e não facilitar o entendimento do que ele significa. Ferramentas de transformação e apresentação de dados são usadas de forma extensiva para ficarem sempre conscientes da sua atuação.

A sociedade também tem o direito de usar tais ferramentas para ter consciência da atuação do governo que a representa. Cada vez mais, a população tem cobrado uma postura ética e transparente dos negócios públicos. Segundo Sant'anna (2015), o *Big Data* é a maneira de entender melhor a demanda da sociedade e usar o próprio cidadão no processo de tomada de decisão. De acordo com o autor, a participação civil nos processos de decisão não pode ficar restrita ao processo eleitoral e à militância. “O *Big Data* é um tema recente no mundo acadêmico e no mundo empresarial, no entanto, a velocidade do seu crescimento tem transformado a gestão e a sociedade em todo o mundo” (GONÇALVES, 2015, p. 07). No tocante a um país com dimensões e características do Brasil, muitas vezes estaremos nos referindo a valores que passam facilmente dos “bilhões”. Este trabalho tem por objetivo propor uma ferramenta de visualização de dados públicos e avaliar esta ferramenta de acordo com um método previamente estabelecido, de forma a deixar claro para o cidadão o poder que uma análise crítica pode ter sobre a tomada de suas decisões em sociedade. Com este objetivo, desenvolvemos uma ferramenta que possibilita a análise do caso específico das transferências federais aos estados da república, sendo este o caso de estudo onde será desenvolvido todo o trabalho. Por fim, avaliamos as visualizações desenvolvidas e ajustamos todas de forma a mitigar todos os aspectos da avaliação que não foram avaliados de forma positiva.

Esse trabalho está estruturado em cinco capítulos. O primeiro, “Introdução”, contém informações introdutórias acerca do trabalho, como considerações iniciais e as justificativas do estudo, o problema e escopo, objetivo principal e objetivos gerais, a definição e classificação da metodologia aplicada. O segundo capítulo contém uma revisão bibliográfica sobre os temas abordados, apresentando três dos aspectos considerados principais para o entendimento deste presente estudo. No terceiro capítulo, são apresentados o método utilizado

para a realização do trabalho, assim como a ferramenta que foi usada para analisar e disponibilizar os dados. No quarto capítulo, são apresentados os resultados da pesquisa realizada sobre o nível de entendimento nos conceitos e aplicações da ferramenta. Posteriormente, são avaliados através de verificação de qualidade através de conceitos apresentados anteriormente. No quinto e último capítulo, são tratadas as considerações finais do presente estudo.

## **2 VISUALIZAÇÃO DE DADOS GOVERNAMENTAIS**

Nesta seção, trabalhos relacionados ao tema de visualização de dados governamentais serão abordados. No primeiro, Barcellos et *al.* (2017), a interpretação dos dados abertos é abordada no contexto de transparência. No segundo, Macedo et *al.* (2020), a recomendação de visualizações de dados governamentais abertos é tratada de forma a guiar os esforços de definição de desenvolvimento. No terceiro, Rolim (2020), é proposta uma arquitetura para interface de sistemas de dados georreferenciados.

Barcellos et *al.* (2017) enunciam que permitir que usuários usem visualizações específicas sobre dados pode gerar uma interpretação mais efetiva desses dados. Este trabalho foi escolhido como referência por ser uma abordagem prática da visualização de dados. Assim sendo, é um conteúdo muito relevante para complementar o trabalho deste artigo. Em concordância, este trabalho propõe visualizações com dimensões e medidas definidas em seu contexto, permitindo que os usuários analisem os seus respectivos comportamentos em visualizações de diversos tipos. O tipo do campo usado é determinante para a escolha da visualização, com o intuito de maximizar a experiência cognitiva do usuário. Assim sendo, podemos notar a concordância deste trabalho com o exposto por Barcellos et *al.* (2017) e a concretização da construção das referidas visualizações dentro do contexto das transferências federais para os estados brasileiros. Ao propor as análises dos dados da cidade de Chicago, notamos a definição de medidas a partir das colunas disponíveis nos datasets disponíveis. Análises usando gráficos de linha, mapas da cidade e heatmaps são, então, construídas. Neste trabalho, usamos uma ferramenta que permite a construção de tais visualizações e muitas outras, conforme foi desenvolvido. A partir de tais visualizações, variações e dependências são então observadas para o caso da cidade. No caso deste trabalho, as visualizações podem ser reutilizadas com o usuário tendo liberdade de fazer suas seleções, elevando as possibilidades de análise para tantos casos quantos forem as dimensões e seus valores.

Transparência, então, refere-se à característica das instituições que possuem credibilidade em seus dados, sendo creditados a estes toda a confiança de serem completos e precisos dentro do escopo em que estão inseridos. Diante da crescente pressão por serviços públicos eficientes e lisura no trato dos recursos públicos, aos dados disponíveis no Portal da Transparência do governo federal do Brasil credita-se a transparência, conforme entendida por Barcellos et al. (2017). Esse trabalho tem por objetivo propor uma ferramenta que teste a transparência dos referidos dados abertos no campo prático do caso de uso dos recursos transferidos pela governo federal. A interpretabilidade dos dados citada por Barcellos et al. (2017) está relacionada diretamente com o conceito de Informação do domínio da Ciência da Informação (2.2.1), visto que a informação é definida como o dado com potencial informacional (significado subjetivo, dependente da interpretação). Este trabalho, então, é o passo prático recomendado por Barcellos et al.(2017), concretizando os referidos conceitos de forma sistemática e orientado à realidade definida pelo escopo dos dados analisados.

Macedo et al. (2020) ressaltam a evidente lacuna que existe entre a disponibilização de uma enorme quantidade de dados em portais públicos e a utilização desses dados pela sociedade. Este trabalho foi escolhido porque, além de ser uma iniciativa prática de visualização de dados, vai além da visualização para propor um *framework* de recomendação. Entendemos que há relevância no trabalho justamente por ter usado a visualização de dados para validar o método proposto em seu conteúdo. Nesta lacuna que ferramentas de visualização atuam. Saber qual tipo de visualização usar dependendo dos dados disponíveis é a tarefa que Macedo et al. (2020) automatizam em sua ferramenta. A solução de Macedo et al. (2020), então, tem um objetivo geral de propor processos de determinação e uso das visualizações mais adequadas para um determinado conjunto de dados, bem como a ferramenta implementada por esses processos. Comparando com a solução proposta neste trabalho, há um aspecto mais geral no método de Macedo et al. (2020), que guia e implementa todos os processos de visualização de dados governamentais. Isto difere deste trabalho, que tem uma ferramenta escolhida (Qlik Sense) e processos definidos de ETL para propor uma solução específica para o caso das transferências governamentais. Em nossa ferramenta, já temos uma grande variedade de visualizações disponíveis e o processo de decisão de uso dessas visualizações vem de experiências de desenvolvimentos anteriores. Apesar disso, Qlik Sense tem disponível um “Insights Advisor” que analisa todo o dataset e retorna as visualizações recomendadas para o usuário.

Não há a pretensão, na solução apresentada neste trabalho, de propor uma solução geral, e sim uma abordagem bem específica e pensada especialmente para o dataset das transferências governamentais. Recomendações automatizadas de visualizações de dados compõem o nosso set de ferramentas disponíveis para propor as melhores visualizações para os dados de transferências governamentais e são importantes para enriquecer o processo de desenvolvimento. Apesar disso, nosso trabalho usa essas recomendações junto com o desenvolvimento concreto para propor objetivamente uma forma prática de analisar os dados. A arquitetura proposta e implementada por Rolim (2020) tem seu escopo delimitado nos dados georreferenciados. No contexto abordado pelo autor (cidades inteligentes, com dados heterogêneos conectados), esse tipo de dados é muito relevante. Em contrapartida, neste trabalho, utilizamos uma arquitetura mais geral (processo de ETL) e uma implementação capaz de tratar dados de diversas naturezas, inclusive os georreferenciados. De tal forma, não há um tipo de dados específico que a ferramenta usada aqui (Qlik Sense) seja mais eficiente em tratar. Além do tipo, ressaltamos aqui o escopo dos dados tratados. Rolim (2020) tem sua proposta baseada no escopo de cidades inteligentes. Em nosso trabalho, estamos tratando de dados de transferências governamentais, tendo o escopo federal, estadual e municipal. Portanto, o escopo dos dados deste trabalho é mais amplo. Apesar disso, a especialização da arquitetura em dados georreferenciados de cidades inteligentes a torna muito eficiente em seu escopo e reutilizável dentro do mesmo

De forma similar a Macedo et al. (2020), um dos objetivos de Rolim (2020) é criar uma interface que faça recomendações de visualizações de dados e auxilie o usuário (desenvolvedor ou não) em sua tarefa de construir suas próprias visualizações. De forma similar ao trabalho citado no parágrafo anterior, a proposta das recomendações de visualizações de dados de Macedo et al. (2020) é válida no contexto deste trabalho justamente por atuar de forma complementar, guiando comparativamente o conjunto de esforços envolvidos no desenvolvimento da ferramenta proposta neste trabalho. O escopo do trabalho de Macedo et al. (2020) é mais amplo que o de Rolim (2020), porém este último aplica a sua arquitetura em um contexto mais específico, de cidades inteligentes. Ou seja, ambos propõem, porém Rolim (2020) propõe a arquitetura e a aplica em seu contexto.

### **3 FERRAMENTA PROPOSTA**

O principal propósito da ferramenta é a disponibilização de uma interface de fácil utilização, que permita ao usuário, com um mínimo conhecimento de sistemas de informação, analisar os dados de transferências governamentais, aumentando a participação popular na fiscalização dos recursos públicos. As etapas necessárias para o desenvolvimento da ferramenta são: (i) identificação das fontes de dados, (ii) extração dos dados, (iii) transformação dos dados e (iv) carregamento dos dados transformados. Nossa abordagem considera o processo de ETL conforme o diagrama (notação BPMN) da Figura 4, a seguir:

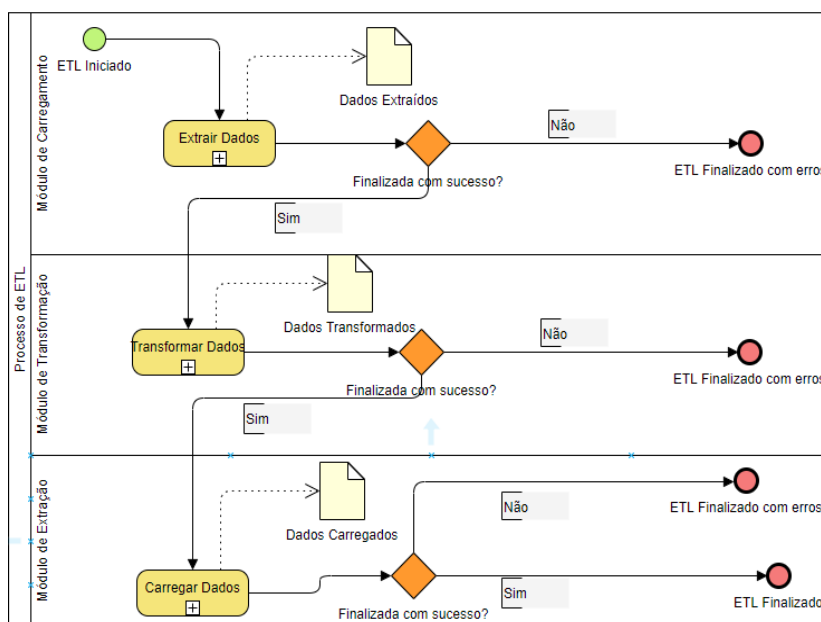


Figura 1: Processo de ETL em notação BPMN

Para a identificação das fontes dos dados, realizamos uma busca exploratória nos portais federais. A busca teve como objetivo principal encontrar a fonte mais completa de dados do governo federal do Brasil. Primeiramente, foi escolhido um poder da República para análise: o Poder Executivo. Por ter características de administração direta dos interesses e serviços governamentais, a análise dos seus dados tem um potencial impacto direto nas relações humanas da sociedade brasileira. A fonte considerada para o governo federal é o site oficial do Portal da Transparência, que tem por objetivo cumprir o requerimento legal de disponibilizar dados públicos a todos os cidadãos da república<sup>3</sup>. Os dados são exportados em formato tabular (arquivo de extensão “xlsx”). Isto atende ao critério de que dados abertos devem ser legíveis por máquinas segundo OKF(2021).

<sup>3</sup> Link: <http://www.portaltransparencia.gov.br/download-de-dados/transferencias>

Para a etapa de extração dos dados, um processo manual de download dos arquivos do portal foi implementado, dado que o tamanho dos arquivos (variando de 17 MB a 137 MB, totalizando mais de 4 GB) torna o uso de APIs muito limitado e suscetível a erro. Dado que o aplicativo tem a característica de analisar dados históricos, este não é um sistema que precisa de dados em tempo real.

Para a etapa de transformação dos dados e carregamento dos dados, utilizamos o Qlik Sense Desktop<sup>4</sup>, software de Business Intelligence que é de livre uso para alunos e professores de universidades e conta com funcionalidades compatíveis e suficientes com as necessidades de análise dos dados.

Para se ter uma visão unificada de todas as etapas, pode-se fazer um paralelo dos conceitos de dado, informação e conhecimento com o conceito de ETL, sendo a etapa de Extract correspondendo ao dado, a etapa de Transform correspondendo à informação e etapa de Load possibilitando a ocorrência de conhecimento.

Este paralelo proposto (Padrão Estratificado) entre os dois conceitos auxiliou no desenvolvimento, durante o qual separamos o nosso código em camadas, cada camada sendo responsável por uma etapa do ETL. Assim, facilitamos a manutenção e a adição de novas funcionalidades ao código em si. Propomos a nomenclatura e relacionamento das etapas e as relacionamos aos conceitos de Ciência da Informação conforme a Tabela 1.

Tabela 1: Comparação entre os conceitos utilizados.  
Fonte: própria.

		Padrões	
		Ciência de Dados	Ciência da Informação
Etapas	Extrair	Dado	ODS
	Trasformar	Informação	TDS QMART
	Carregar	Conhecimento	Visualização

Na etapa 1 (ODS, Original Data Source), tem-se apenas a extração dos dados para o modelo de dados em memória do Qlik Sense, em memória principal. Apenas poucas operações de transformação são feitas nesta etapa: apenas o suficiente para unificar todas as fontes, de forma a termos uma etapa de transformação única.

<sup>4</sup> Link: <https://www.qlik.com/pt-br/company/academic-program>



Na etapa 2 (TDS, Transformed Data Source), todas as operações de transformações são feitas. Essas operações incluem, por exemplo, operações como: mudanças de formato, renomeações, agregações, concatenações de campos ou tabelas e etc.

Na etapa 3 (Q-Mart, Qlik Sense Data Mart), operações de transformação mais complexas e diretamente relacionadas aos requisitos funcionais são implementadas. Todas as regras de negócio complexas são desenvolvidas nesta camada. No aplicativo do nosso projeto, adaptamos e desenvolvemos um calendário, que nos dá análises de quarter (QTH - trimestre), Year to date (YTD - agregação de todas as transferências de um ano), monthly (MTH - a visão comum mês a mês) e Moving Annual Turnover (MAT, a agregação dos últimos 12 meses, independente do ano) . As operações de otimização de desempenho também entram nesta etapa.

Na Etapa 4 (Apresentação), as agregações de campos em medidas e a definição das dimensões é feita. Neste trabalho, utilizamos agregações comuns em quaisquer ferramentas de visualizações: sum (denotando soma), distinct count (denotando contagem distinta) dentre outras. A ferramenta de visualização usada, Qlik Sense, disponibiliza diversos objetos de visualização para serem utilizados. Para cada um deles, medidas (usando as agregações disponíveis) e dimensões (usando as dimensões vindas das fontes de dados).

Cada uma das visualizações desenvolvidas foi avaliada de acordo com o método proposto por Barcellos et al. (2018), em onze heurísticas. A seguir, apresentamos as heurísticas de avaliação das visualizações propostas por Barcellos et al. (2018).

A ferramenta, em seu estado pré avaliação<sup>5</sup> e pós avaliação<sup>6</sup> estão disponíveis no ambiente virtual disponibilizado pela Qlik. De acordo com o programa de incentivo ao uso do Qlik Sense, conforme dito anteriormente.

Tabela 2: Definição das Heurísticas propostas por Barcellos et al. (2018).

<b>Heurística</b>	<b>Descrição</b>
Heurística H1	“O avaliador deve verificar se a visualização de dados mantém o contexto quando impressa sem cores.”
Heurística H2	O avaliador deve verificar se há pouca informação, confusão ou ruído.”

<sup>5</sup> <https://rafaeltcosta.us.qlikcloud.com/sense/app/74aa3d18-5f1a-49b0-a5e0-3e31dd35f56a/overview>

<sup>6</sup> <https://rafaeltcosta.us.qlikcloud.com/sense/app/f9d3d8ee-78b6-4a17-8e68-d7d20b2619c4/overview>

Heurística H3	“O avaliador deve verificar a direção de leitura do usuário.”
Heurística H4	“O avaliador deve verificar itens relacionados próximos entre si.”
Heurística H5	“O avaliador deve verificar questões de espaço e tempo - eixos X e Y.”
Heurística H6	“O avaliador deve verificar paletas e ícones não visíveis para daltônicos.”
Heurística H7	“O avaliador deve verificar se o tamanho dos objetos pode dificultar a leitura.”
Heurística H8	“O avaliador deve verificar se a cor dos objetos pode dificultar a leitura.”
Heurística H9	“O avaliador deve verificar se há espaços em branco que podem dificultar a leitura.”
Heurística H10	“O avaliador deve verificar se há lacunas de comunicação com o usuário.”
Heurística H11	“O avaliador deve evitar perspectivas 3D.”

Os problemas encontrados são classificados pelas seguintes escalas, cada indivíduo define o tipo de problema de acordo com sua percepção:

- Problema cosmético(Cosmetic problem): não há necessidade de consertar se não houver tempo.
- Problema menor(Minor Problem) : corrija se o problema for de prioridade mais baixa.
- Problema maior(Major Problem) : importante para resolver o problema, alta prioridade.
- Desastre(Disaster) : repare imediatamente.

A seguir, serão apresentadas todas as visualizações resultantes do processo descrito nos parágrafos anteriores desta seção, bem como a avaliação, de acordo com a Tabela 2. Ou seja, toda Heurística listada na Tabela 2 considerada violada por cada visualização, será listada separadamente com sua respectiva figura.

### CAPITAL vs INTERIOR ANALYSIS

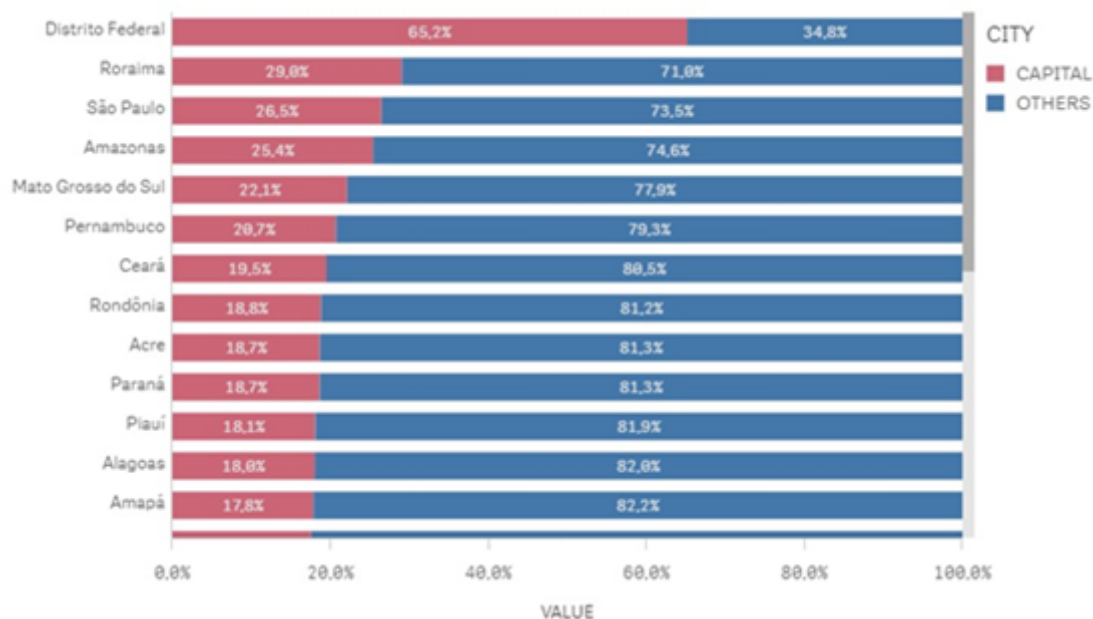


Figura 2: Gráfico de barra mostrando a relação de investimento capital vs interior.

### CAPITAL vs INTERIOR ANALYSIS

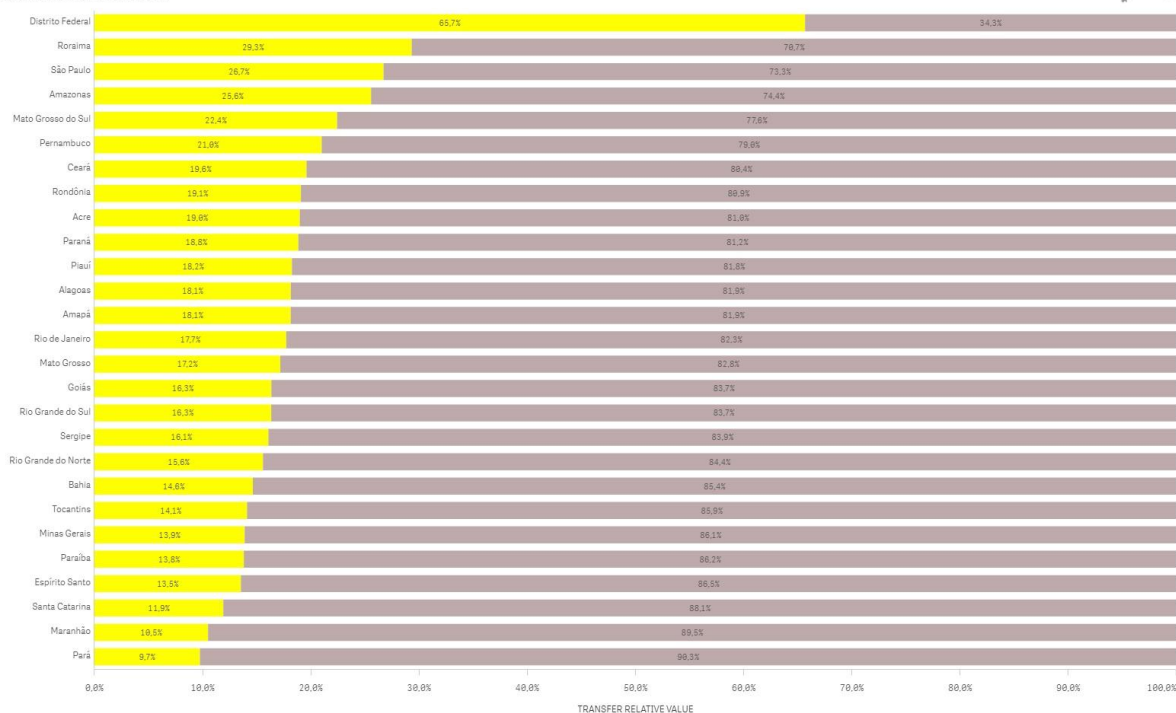


Figura 3. referente a Figura 2, após os ajustes.

Na Figura 2, temos uma visualização (gráfico de barra horizontal) que compara a quantidade de recursos transferidos à capital dos estados da federação (em amarelo) e a quantidade de recursos transferidos para as outras cidades do estado (em cinza). O objetivo

desta visualização é evidenciar o quão concentradas podem estar as transferências. Diversos fatores podem fazer com que estados tenham recursos mais concentrados em sua capital.

A seguir, as melhorias aplicadas na Figura 2, tendo como resultado a Figura 3:

Tabela 3: Melhorias da visualização correspondente a Figura 2.

Heurística Violada	Impacto	Ação/Melhoria
H5 A legenda do eixo x se encaixa no problema “variáveis mal descritas” apontado na heurística H5. O valor representado pode ter melhor descrição do que apenas “value” a fim de melhor o leitor. Eixo x mal utilizado. Não dá para saber o que seria esse “value”.	Major Problem: Quanto à frequência do problema, ocorrerá sempre que o usuário não possuir informações sobre o contexto da visualização. Alto impacto, pode dificultar a interpretação da visualização. Alto impacto, certamente dificulta a interpretação da visualização pelo usuário.	Mudança do nome para “TRANSFER RELATIVE VALUE”, denotando de forma mais explícita o conteúdo da medida da visualização.
H6 Apresenta uma combinação de azul e vermelho.	Major Problem: Baixa frequência, impossível de ser contornado pelo usuário e alto impacto.	Mudança das cores para valores manualmente escolhidos e atrelados ao valor “CAPITAL” e “OTHERS”, dimensão criada na camada de transformação.

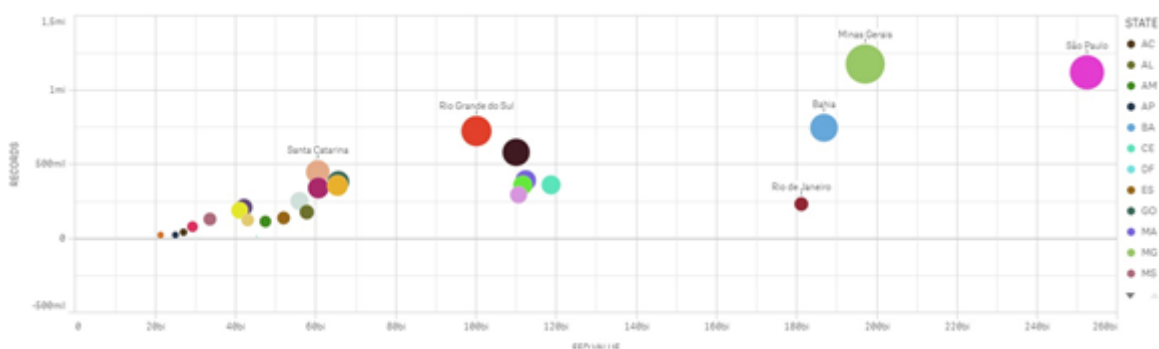


Figura 4: Gráfico de bolha mostrando o total de recursos transferidos e o total de linhas processadas por estado.

A seguir, as melhorias aplicadas na Figura 4, tendo como resultado a Figura 5:

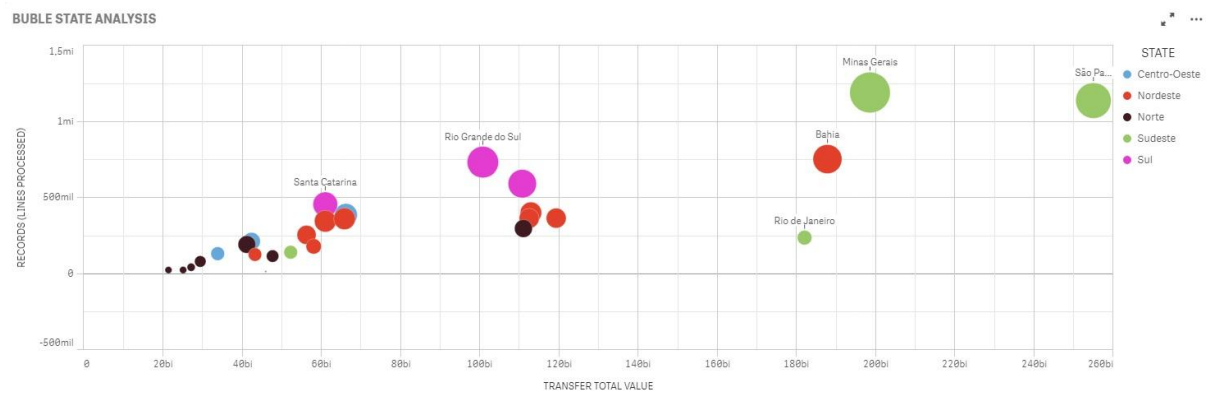


Figura 5. referente a Figura 4, após os ajustes.

A seguir, a avaliação da Figura 4:

Tabela 4: Melhorias da visualização correspondente a Figura 4.

Heurística Violada	Impacto	Ação/Melhoria
H1 Em escala de cinza fica impossível distinguir algumas cores.	Minor Problem: Baixa frequência/facilmente contornado. Não há como interpretar corretamente, se as cores das regiões são indistinguíveis.	As cores foram alteradas para representarem a região em que os estados estão inseridos.
H5 Eixos x e y mal utilizados. Não dá para entender sobre o que se trata a visualização	Major Problem: Alto impacto, certamente impossibilita a interpretação da visualização pelo usuário	As legendas dos eixos foram alteradas para denotar de forma mais explícita os valores que eles representam.
H6 Apresenta combinação de vários tons de verde, azul e vermelho, que podem impossibilitar a distinção das categorias (estados) por daltônicos.	Major Problem: Baixa frequência, impossível de ser contornado pelo usuário e alto impacto.	As cores foram alteradas para representarem a região em que os estados estão inseridos. Ainda haverá a variação de cor, porém a legenda (e o <i>tooltip</i> do <i>mouse hover</i> ) dizem exatamente que valor da dimensão está representado.

Na Figura 5, temos uma visualização (gráfico de bolha ou *scatter plot*) que compara os estados quanto ao valor monetário de recursos transferidos (eixo x), a quantidade de transações relacionadas a esse valor monetário (eixo y). O tamanho do círculo varia de acordo com a quantidade de cidades relacionadas às transações em cada estado. A cor do círculo representa a região em que ele está situado. O objetivo desta visualização é comparar os

estados não somente em relação ao valor monetário das transferências, mas também evidenciar a quantidade de cidades e transações envolvidas relacionadas. Ou seja, notamos, por exemplo, que o Rio de Janeiro conta com um alto valor monetário (mais a direita no eixo x). Apesar disto, o círculo é pequeno, então há poucas cidades envolvidas. E a quantidade de transações que gerou este comportamento foi pequena, comparativamente (RJ está mais abaixo que a maioria dos estados, no eixo y).

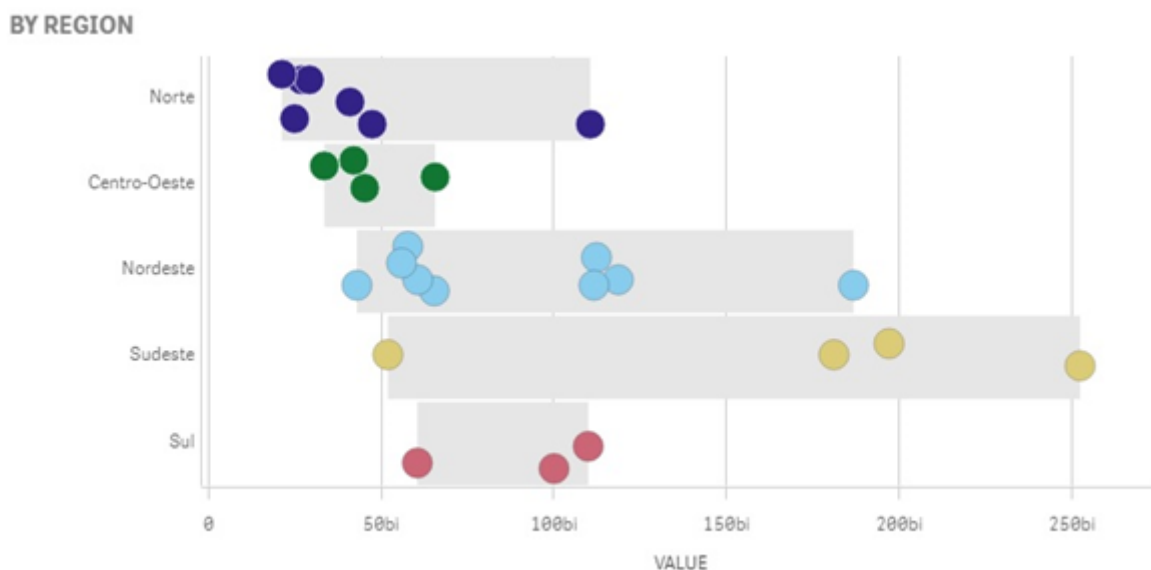


Figura 6: Gráfico de distribuição mostrando o total de recursos transferidos por região geográfica.



Figura 7: referente a Figura 6, após os ajustes.

A seguir, a avaliação da Figura 6:

Tabela 5: Avaliação da visualização correspondente a Figura 6.

Heurística Violada	Impacto	Ação/Melhoria
H1 Em escala de cinza pode ser um pouco difícil distinguir as bolhas do nordeste (azuis) e sudeste (amarelas), dificultando a interpretação, porém ainda é possível distinguir pela posição em relação à legenda do eixo x e pela parte destacada originalmente em cinza.	Cosmetic problem: Problema de baixa frequência, que pode ser facilmente contornado. Ocorre apenas em impressão em cinza e tela em escala de cinza (ex: telas E-ink), que são menos comuns. Impacto baixo pois pode atrapalhar, mas não impede a interpretação da visualização.	Nenhuma ação. As cores representam a região em que os estados estão inseridos. Ainda haverá a variação de cor, porém a legenda (e o <i>tooltip</i> do <i>mouse hover</i> ) dizem exatamente que valor da dimensão está representado.
H5 Eixo x mal utilizado. Não dá para saber o que seria esse “value”	Major Problem: Alto impacto, certamente dificulta a interpretação da visualização pelo usuário.	Mudança do nome para “TRANSFER VALUE”, denotando de forma mais explícita o conteúdo da medida da visualização.
H6 Apresenta combinação de verde, azul e vermelho e, para daltônicos, os tons das categorias podem parecer próximos. Entretanto, ainda é possível distinguir pela posição em relação à legenda do eixo x e pela região originalmente em cinza.	Cosmetic Problem: Problema de baixa frequência (considerando % de pessoas com daltonismo), que influencia, mas não impossibilita a interpretação, devido a outros recursos que facilitam a interpretabilidade. Problema pode ser contornado; de baixo impacto.	Nenhuma ação. As cores representam a região em que os estados estão inseridos. Ainda haverá a variação de cor, porém a legenda (e o <i>tooltip</i> do <i>mouse hover</i> ) dizem exatamente que valor da dimensão está representado.

Na Figura 7, temos uma visualização (gráfico de distribuição ou *distribution plot*) que compara os estados quanto ao valor monetário de recursos transferidos (eixo x), separando-os em regiões de distribuição (eixo y). O tamanho do círculo é fixo. O objetivo desta visualização é comparar os estados dentro de sua região e entre regiões, ao mesmo tempo. Desta forma, fica evidente a região que tem estados com muitas transferências, bem como a região que tem estados com poucas transferências.

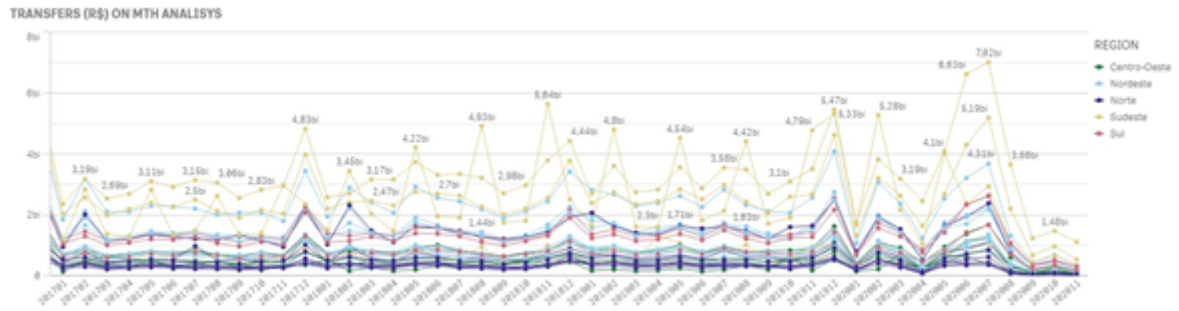


Figura 8: Gráfico de linha mostrando o total de recursos transferidos (últimos 12 meses) por estado, cor determinada por região.

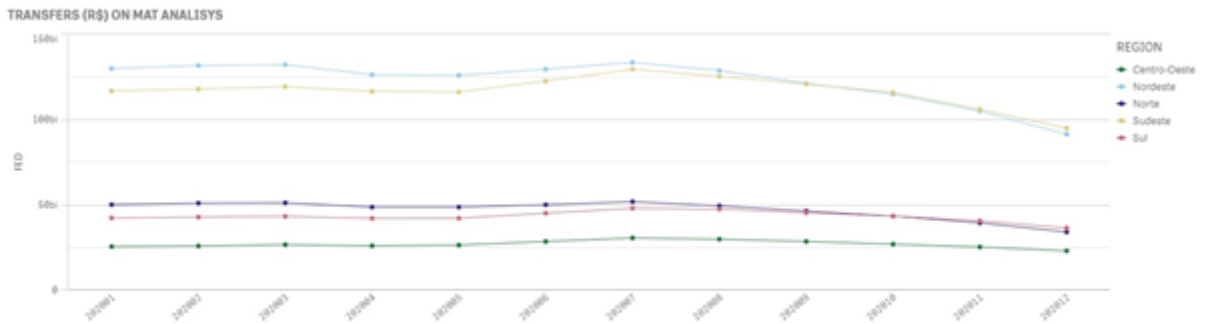


Figura 9: Gráfico de linha mostrando o total de recursos transferidos (últimos 12 meses) por região após melhorias . Cor por região.

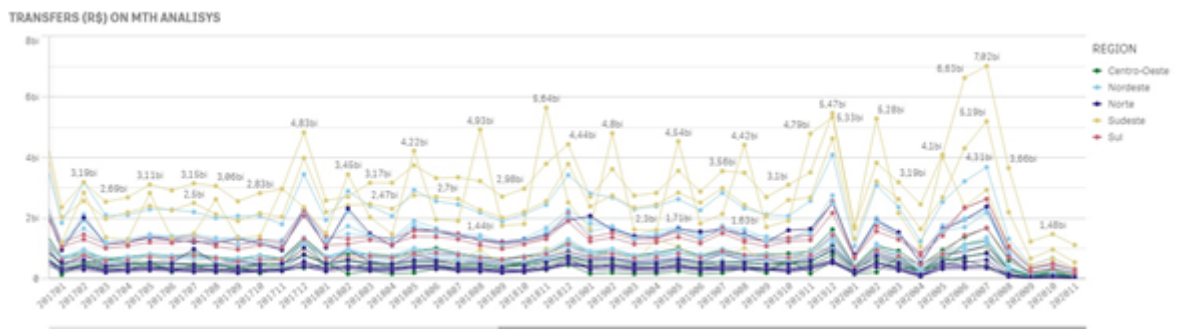


Figura 10: Gráfico de linha mostrando o total de recursos transferidos (por mês) por estado, cor determinada por região.



Figura 11: Gráfico de linha mostrando o total de recursos transferidos (por mês) por região após melhorias. Cor por região.



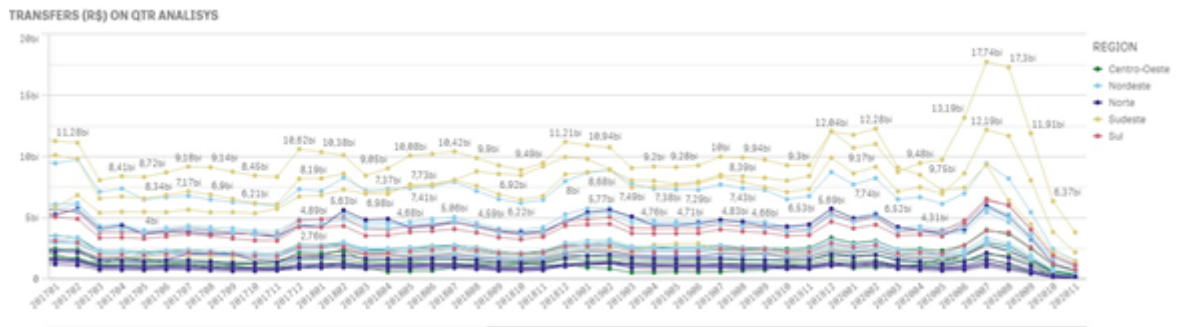


Figura 12: Gráfico de linha mostrando o total de recursos transferidos (últimos 3 meses) por estado, cor determinada por região.

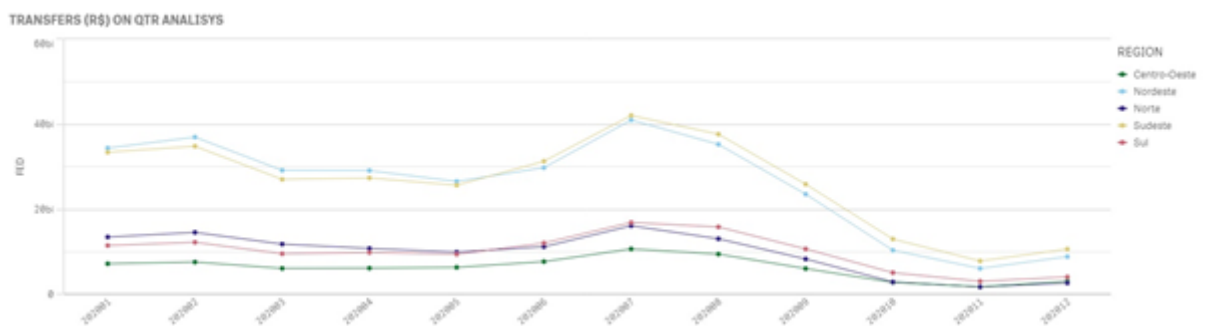


Figura 13: Gráfico de linha mostrando o total de recursos transferidos (últimos 3 meses) por região após melhorias . Cor por região.

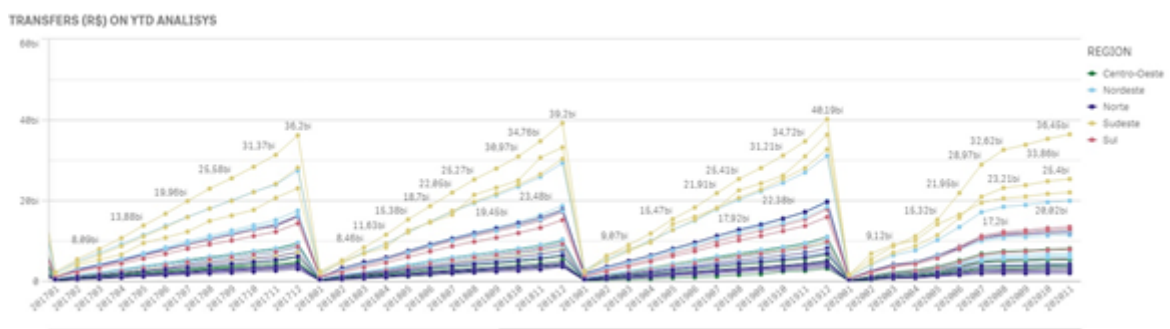


Figura 14: Gráfico de linha mostrando o total de recursos transferidos (total do ano até o mês da dimensão) por estado, cor determinada por região.

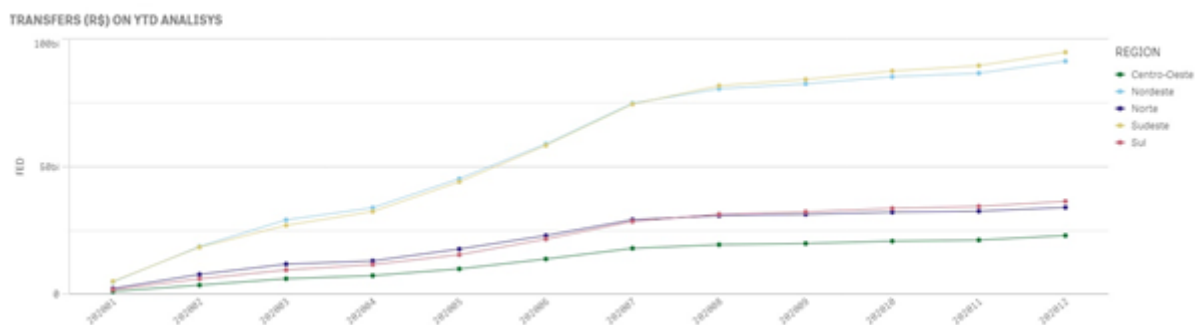


Figura 15: Gráfico de linha mostrando o total de recursos transferidos (total do ano até o mês da dimensão) por região após melhorias . Cor por região.

Nas Figuras 9, 11, 13 e 15, temos uma visualização (gráfico de linha) que compara as regiões quanto a quantidade de recursos transferidos (eixo y) ao longo do tempo (períodos no eixo x). As 4 visualizações podem ser acessadas apenas uma por vez, visto que ocupam o mesmo espaço no layout. A única diferença entre elas é a medida. Todas as medidas têm em comum o fato de levarem em conta espaços diferentes de tempo em suas fórmulas.

Na Figura 9, a medida é o valor agregado das transferências nos últimos 12 meses, sendo o mês do gráfico o último dos 12. Na Figura 11, a medida é o valor agregado das transferências no mês mostrado no eixo x. Esta é a forma mais comum de análise, em que cada valor do eixo x representa o referencial do total correspondente mostrado. Na Figura 13, a medida é o valor agregado das transferências nos últimos 3 meses, sendo o mês do gráfico o último dos 3. Ou seja, esta visualização pretende comparar o desempenho das transferências em relação aos trimestres correntes. Na Figura 15, a medida é o valor agregado das transferências no ano corrente. Ou seja, o mês de um ano acumula o total das transferências do ano em que está inserido até o mês mostrado no gráfico. Esta visualização mostra de forma clara a evolução anual e a comparação entre os anos do total das transferências.

A seguir, apresentamos a avaliação das Figuras 8, 10, 12 e 14. Esses objetos são quase idênticos, tendo como única diferença a medida apresentada. Assim sendo, o resultado da avaliação das visualizações em questão é idêntico. O resultado da aplicação das melhorias pode ser visualizado nas Figuras 9, 11, 13 e 15 respectivamente.

Tabela 6: Avaliação da visualização correspondente às Figuras: 8;10;12 e 14.

Heurística Violada	Impacto	Ação/Melhoria
H1 Cores que dificultam a	Minor Problem: Baixa frequência/facilmente	Nenhuma ação.As cores representam a região em que

interpretação, quando a visualização é impressa sem cores ou em baixa qualidade.	contornado. Não há como interpretar corretamente, se as cores das regiões são indistinguíveis.	os estados estão inseridos. Ainda haverá a variação de cor, porém a legenda (e o <i>tooltip</i> do <i>mouse hover</i> ) dizem exatamente que valor da dimensão está representado.
H2 Excesso de chartjunk	Major Problem: Alta frequência/Alto impacto. Grids desnecessários, excesso de informação. Poderia condensar os dados.	A dimensão de estado foi alterada para região e a dimensão de mês foi alterada para ser dimensão <i>Drill Down</i> , em cujo primeiro nível temos o ano e, quando um dos anos é selecionado, o segundo nível (mês) é mostrado. Valores das transferências nos pontos foram retirados.
H3 Texto do eixo x na diagonal pode dificultar levemente a leitura.	Minor Problem: Problema de alta frequência, que pode ser facilmente contornado e tem impacto baixo. Ocorre sempre, dificultando levemente a leitura	Nenhuma ação. A única alternativa disponível seria forçar o texto a ficar na horizontal, o que dificultaria ainda mais a leitura.
H6 Apresenta combinação de verde, azul e vermelho que, somada ao fato de as linhas do gráfico se entrelaçarem, pode dificultar a distinção das categorias por daltônicos.	Problema de baixa frequência (considerando % de pessoas com daltonismo), impossível de ser contornado pelo usuário no momento, e com alto impacto (impossibilita a leitura correta)	Nenhuma ação. As cores representam a região em que os estados estão inseridos. Ainda haverá a variação de cor, porém a legenda (e o <i>tooltip</i> do <i>mouse hover</i> ) dizem exatamente que valor da dimensão está representado.

Ao separar logicamente a camada de extração das demais camadas, facilitamos a manutenção e expansão futura do aplicativo, pois, caso as fontes mudem, sabemos exatamente onde atualizar. Não há o risco de alterarmos inadvertidamente alguma funcionalidade das outras camadas, conforme a Tabela 1. Outra vantagem dessa separação de camadas é a reusabilidade do modelo de dados. Em um contexto de um modelo servindo de fonte de dados para várias aplicações, os dados que foram extraídos e transformados podem ser guardados para serem posteriormente usados por outros aplicativos. Apenas a camada de apresentação

precisaria ser desenvolvida pois todo o esforço computacional investido nas outras camadas pode ser perfeitamente aproveitado para outros aplicativos. Em nosso processo de ETL, processamos aproximadamente 4.5 GB de dados. O nosso aplicativo tem 128 MB de tamanho. Uma redução extrema, que envolveu, entre outras coisas, filtro dos dados desnecessários. O tempo de compilação e reload é de 4 minutos aproximadamente. Levando-se em conta a quantidade de dados processados, vemos como muito positivo o resultado final do processo de ETL.

#### **4 CONCLUSÃO**

Neste trabalho, desenvolvemos e apresentamos o resultado da obtenção, transformação e análise de dados públicos referentes às transferências federais para os estados, estimulados pela crescente relevância da fiscalização e conscientização do bom uso dos recursos públicos. O tamanho dos dados é consideravelmente grande, porém, a nossa abordagem de desenvolvimento e a ferramenta escolhida mitigaram os possíveis problemas referentes ao tamanho dos dados. Permitimos ao usuário obter conhecimento sobre as transações inter-governamentais, encapsulando todo o conhecimento necessário de modelagem de dados em camadas inferiores a da visualização, como se espera de um aplicativo de Data Science e visualização de dados. Definimos um processo de ETL, codificamos todas as etapas necessárias para a visualização dos dados e buscamos correlacionar os conceitos exercitados com a teoria de Ciência da Informação. Esta correlação, por sua vez, tem o objetivo de apoiar a reflexão sobre como os dados sistematizados podem influenciar os indivíduos e organizações. Pois, conforme afirmam Beira *et al.* (2020), o conhecimento, sendo subjetivo, depende do indivíduo e do seu compromisso em agir de acordo com as informações disponíveis.

O universo dos dados públicos é muito extenso. Então, decidimos começar por um tipo específico de dados, para futuramente poder expandir o escopo dos dados a serem tratados. Para tal, nos baseamos em uma situação específica, as transferências, para poder futuramente abstrair um modelo de análise. Segundo Meter(2020), no conhecimento reside o consenso de conceitos da realidade. Assim sendo, pretendemos oferecer uma ferramenta que oferece ao cidadão informações relevantes para a formação do consenso. Para o futuro, pretendemos trabalhar para aprimorar todas as etapas do processo, agregando regras de negócio pertinentes a uma análise cidadã dos recursos que pertencem à sociedade. Diversos tipos de recursos de análise ainda podem ser adicionados. Cada um com uma aplicação e utilidade particular. Além dos aspectos citados, é ampla a quantidade de implementações de visualizações de

dados, seja em instituições públicas ou privadas, ou ainda para estudos particulares. Portanto, a tarefa de encontrar soluções para as limitações da implementação aqui proposta deve apresentar muitos resultados positivos.

## REFERÊNCIAS

- BARCELLOS, Raissa & VITERBO, Jose & BERNADINI, Flavia & TREVISAN, Daniela. (2018). **An Instrument for Evaluating the Quality of Data Visualizations**. 10.1109/iV.2018.00038.
- BARCELLOS, Raissa. **Transparency in practice: using visualization to enhance the interpretability of open data**. In: Proceedings of the 18th Annual International Conference on Digital Government Research. 2017. p. 139-148.
- BEIRA, Joana & OLIVEIRA, Marlene & MOREIRA DA SILVA, Cristiano. (2020). **Um olhar dos estudantes do curso de Biblioteconomia acerca do que são dados, informações e conhecimentos**. Informação & Informação. 25. 484-508. 10.5433/1981-8920.2020v25n2p484.
- BRESSER-PEREIRA, Luis Carlos. **"As formas de responsabilização na gestão pública"**. In: BRESSER PEREIRA, Luiz Carlos & GRAU, Nuria Cunill (Coord.). Responsabilização na administração pública. São Paulo: Clad/Fundap, 2006.
- ESPINDOLA, A.M.S; ROTH, L. **Big Data e Inteligência Estratégica: Um Estudo de Caso Sobre a Mineração de Dados como Alternativa de Análise**. Revista Espacios, v. 37, n. 4, p. 16, out. 2015.
- GONÇALVES, Joanna. **Business Intelligence Suportado por Tecnologia Big data na Gestão Estratégica das Empresas – Estudo de Caso**. 2015. 15 f. Dissertação (Mestrado) - Curso de Mestrado em Gestão Empresarial, Utad: Universidade de Trás-os-montes e Alto Douro, Vila Real, 2015.
- MACEDO, Daiane; BARCELOS, Raissa; BERNARDINI, Flavia; VITERBO, José. **Uma ferramenta para recomendação de visualização de dados governamentais abertos**. In: Workshop de computação aplicada em governo eletrônico (Wcge), 8. , 2020, Cuiabá. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2020 . p. 96-107.
- Meter, Heather. (2020). **Revising the DIKW Pyramid and the Real Relationship Between Data, Information, Knowledge and Wisdom**. Law, Technology and Humans. 2. 69-80. 10.5204/lthj.1470.
- Open Knowledge Foundation (OKF) (2021). **Open data handbook**. Disponível em [https://opendatahandbook.org/guide/pt\\_BR/what-is-open-data](https://opendatahandbook.org/guide/pt_BR/what-is-open-data). Acessado em 09/02/2021.
- ROLIM, Douglas Arthur de Abreu. **Dashboards para desenvolvimento de aplicações e visualização de dados para plataformas de cidades inteligentes**. 2020. 87f. Dissertação (Mestrado em Sistemas e Computação) - Centro de Ciências Exatas e da Terra, Universidade Federal do Rio Grande do Norte, Natal, 2020.
- SANT'ANNA, Yan. **Big data: O Uso da Tecnologia da Informação como Provedora da Gestão Pública Integrada**. 2015. 60 f. TCC (Graduação) - Curso de Gestão de Comércio Internacional, Faculdade de Ciências Aplicadas, Universidade Estadual de Campinas, Limeira, 2015.

SODRÉ, L. **Big Data Estratégico: Um Framework para Gestão Sistêmica do Ecossistema Big Data**. 125 f. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia de Produção, COPPE, Universidade Federal do Rio de Janeiro, abr. 2016.

## APÊNDICE A – Lista de ilustrações

FIGURA 1	Processo de ETL em notação BPMN	7
FIGURA 2	Gráfico de barra mostrando a relação de investimento capital vs interior.	11
FIGURA 3	Gráfico referente a Figura 5, com melhorias aplicadas	11
FIGURA 4	Gráfico de bolha mostrando o total de recursos transferidos e o total de linhas processadas por estado.	12
FIGURA 5	Gráfico referente a Figura 7, com melhorias aplicadas	13
FIGURA 6	Gráfico de distribuição mostrando o total de recursos transferidos por região geográfica.	14
FIGURA 7	Gráfico referente a Figura 9, com melhorias aplicadas	14
FIGURA 8	Gráfico de linha mostrando o total de recursos transferidos (últimos 12 meses) por estado, cor determinada por região.	16
FIGURA 9	Gráfico de linha mostrando o total de recursos transferidos (últimos 12 meses) por região após melhorias . Cor por região.	16
FIGURA 10	Gráfico de linha mostrando o total de recursos transferidos (por mês) por estado, cor determinada por região.	16
FIGURA 11	Gráfico de linha mostrando o total de recursos transferidos (por mês) por região após melhorias. Cor por região.	16
FIGURA 12	Gráfico de linha mostrando o total de recursos transferidos (últimos 3 meses) por estado, cor determinada por região.	17
FIGURA 13	Gráfico de linha mostrando o total de recursos transferidos (últimos 3 meses) por região após melhorias . Cor por região.	17
FIGURA 14	Gráfico de linha mostrando o total de recursos transferidos (total do ano até o mês da dimensão) por estado, cor determinada por região.	17
FIGURA 15	Gráfico de linha mostrando o total de recursos transferidos (total do ano até o mês da dimensão) por região após melhorias . Cor por região.	18



## APÊNDICE B – Lista de tabelas

TABELA 1	Comparação entre os conceitos utilizados.	8
TABELA 2	Definição das Heurísticas propostas por Barcellos et al. (2018).	9
TABELA 3	Tabela 3: Melhorias da visualização correspondente a Figura 2, resultando na Figura 3.	12
TABELA 4	Tabela 4: Melhorias da visualização correspondente a Figura 4, resultando na Figura 5.	13
TABELA 5	Tabela 5: Avaliação da visualização correspondente a Figura 6 e suas melhorias gerando a Figura 7.	15
TABELA 6	Tabela 6: Avaliação da visualização correspondente às Figuras: 8;10;12 e 14, e suas respectivas melhorias gerando as Figuras: 9,11,13,15.	18