



VISGL, UMA FERRAMENTA ONLINE PARA VISUALIZAÇÃO DE DADOS MULTIVARIADOS GEORREFERENCIADOS

Tarsus Magnus Pinheiro

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Sistemas e Computação.

Orientador: Claudio Esperança

Rio de Janeiro
Junho de 2016

VISGL, UMA FERRAMENTA ONLINE PARA VISUALIZAÇÃO DE DADOS
MULTIVARIADOS GEORREFERENCIADOS

Tarsus Magnus Pinheiro

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO
ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE
ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE
JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A
OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE
SISTEMAS E COMPUTAÇÃO.

Examinada por:

Prof. Claudio Esperança, Ph.D.

Prof. Julia Celia Mercedes Strauch, Ph.D.

Prof. Doris Clara Kosminsky, D.Sc.

Prof. Ricardo Guerra Marroquim, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
JUNHO DE 2016

Pinheiro, Tarsus Magnus

VisGL, uma ferramenta online para visualização de dados multivariados georreferenciados/Tarsus Magnus Pinheiro. – Rio de Janeiro: UFRJ/COPPE, 2016.

XIV, 85 p.: il.; 29, 7cm.

Orientador: Claudio Esperança

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2016.

Referências Bibliográficas: p. 71 – 76.

1. Visualização de Informações. 2. Visualização de dados multivariados. 3. Geovisualização. I. Esperança, Claudio. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

*Dedico este trabalho à minha
família, que tanto lutou para que
eu chegasse até aqui. Sem vocês,
nada disto faz sentido. Amo
vocês, irrestritamente.*

Agradecimentos

À minha família, em especial meus pais e irmãos, pela força e solidez que me impedem de fraquejar diante das dificuldades.

À Cris e Bia, meus amores. Vocês são a minha inspiração. Obrigado pela paciência nas horas em que faltei para realizar este trabalho e pela força com a qual sempre posso contar.

Aos professores Claudio Esperança, meu orientador, e Ricardo Marroquim pela luz e apoio irrestrito durante toda esta jornada. Sem isso eu não teria conseguido.

Aos colegas do LCG pela amizade e companheirismo. Com vocês aprendi muito. Esse convívio foi muito mais edificante do que a frieza expressa em uma reles grade curricular pode fazer supor.

Ao IBGE, meu empregador, pela oportunidade concedida de poder me dedicar integralmente a esta pesquisa. Sem isso seria impossível realizar esta etapa tão importante da minha carreira.

Preciso agradecer especialmente à colega Gisele Flores Caldas, que revisou este trabalho e me ajudou a organizar as idéias em uma fase especialmente difícil desta empreitada.

Agradeço também a todos os professores com quem tive o prazer de estudar nesta jornada e aos meus amigos, que tiveram minha companhia furtada nesses últimos três anos, mas que sempre estiveram ao meu lado, torcendo pelo meu sucesso.

É importante registrar um agradecimento especial aos funcionários da secretaria do PESC, especialmente Gutierrez e Solange, que sempre se fizeram solícitos e ajudaram em momentos especialmente complicados.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

VISGL, UMA FERRAMENTA ONLINE PARA VISUALIZAÇÃO DE DADOS MULTIVARIADOS GEORREFERENCIADOS

Tarsus Magnus Pinheiro

Junho/2016

Orientador: Claudio Esperança

Programa: Engenharia de Sistemas e Computação

O presente trabalho descreve uma ferramenta *on-line* desenvolvida com o propósito de apresentar soluções para a visualização de dados multivariados georreferenciados. Trata-se de um mapa temático interativo que permite ações de posicionamento e enquadramento, em diversos níveis de detalhe, e onde são apresentadas simultaneamente duas variáveis escalares. O projeto foi desenvolvido baseado na premissa de que a velocidade na renderização das imagens é elemento chave para a criação de ambientes imersivos. A melhora da experiência de uso pode ser obtida mediante às respostas visuais instantâneas que o sistema é capaz de prover a partir de consultas feitas pelo usuário ao mapa através das ferramentas de manipulação. Para atingir esse objetivo, inovou-se ao se construir o sistema por meio da programação de GPUs a fim de realizar a montagem e a manipulação de ladrilhos previamente rasterizados, contendo as informações de localização gravadas no espaço de cor dos *pixels* - o que permite a implementação de ações animadas de interação e decomposição espacial dos dados.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

VISGL, AN ONLINE TOOL FOR MULTIVARIATE GEOREFERENCED DATA
VISUALIZATION

Tarsus Magnus Pinheiro

June/2016

Advisor: Claudio Esperança

Department: Systems Engineering and Computer Science

This study describes an online tool developed to provide solutions for the visualization of multivariate georeferenced data. It consists of an online interactive thematic map which allows positioning and framing in several levels of detail, presenting simultaneously two scalar variables. The Project has been developed based on the premise that speed in image rendering is a key element for the creation of immersive environments. The user's experience can be improved by means queries posed through manipulation tools. To fulfill this objective, the study proposes an innovative system using GPU programming to assemble and manipulate previously rasterized tiles, with location information recorded in the color space of pixels. This procedure allows the implementation of interactive animated actions and spatial data decomposition.

Sumário

Lista de Figuras	x
Lista de Tabelas	xiv
1 Introdução	1
1.1 Motivação	1
1.2 Objetivos	3
2 Fundamentação Teórica	4
2.1 Visualização de Dados Multivariados	4
2.1.1 Faces de Chernoff	4
2.1.2 Figuras de Palito	6
2.1.3 Simulação de Texturas Naturais	7
2.1.4 Simulação de Pintura a Óleo no Estilo Impressionista	7
2.2 Mapas Temáticos e Semiologia Gráfica	8
2.2.1 Semiologia Gráfica	9
2.2.2 As Requisições Dinâmicas	12
2.3 Representação de Dados através de Mapas Coropléticos	13
2.3.1 Uso das Cores na Classificação dos Dados	23
3 VisGL, uma ferramenta <i>on-line</i> para visualização de dados multi- variados georreferenciados	27
3.1 Mapa <i>on-line</i> interativo	27
3.1.1 Variáveis relativizadas	29
3.1.2 Equalização da imagem conforme o enquadramento	31
3.1.3 Visualização por grade estatística	32
3.1.4 Escalas de classes editáveis	33
4 Implementação	36
4.1 Descrição geral	36
4.2 Detalhes da implementação	38
4.2.1 Experimentos preliminares	39

4.2.2	Montagem do mapa	41
4.2.3	Processamento prévio dos ladrilhos	45
4.2.4	Imagem contendo os dados	50
4.2.5	Composição da grade	52
4.2.6	Classificação	54
4.2.7	Visualização das camadas	56
5	Resultados	59
5.1	Frota de Veículos	59
5.2	Eleições Presidenciais Brasileiras de 2014	60
6	Conclusões e Possíveis Desdobramentos	68
	Referências Bibliográficas	71
A	Levantamento de Informações sobre os Arquivos Contendo a Geometria dos Mapas	77
A.1	Arquivos Rasterizados	81

Lista de Figuras

- 2.1 Exemplo de uso das faces esquemáticas proposta por Chernoff. Imagem retirada de Wikimedia Commons (2010). 5
- 2.2 Variáveis visuais (RODRIGUEZ, 2010). 10
- 2.3 Aspectos evidenciados na Cartografia Temática (RODRIGUEZ, 2010). 11
- 2.4 O mapa acima mostra a população residente nas cidades de Minas Gerais em 2015. Captura de tela do serviço WebCart (IBGE, d). . . . 14
- 2.5 O mapa acima, por sua vez, mostra a quantidade de automóveis nas cidades de Minas Gerais em 2015 e aponta para uma relação entre este dado e a população residente, apresentada no mapa da Figura 2.4. De um modo geral, onde há mais habitantes, há mais automóveis. Captura de tela do serviço WebCart (IBGE, d). 15
- 2.6 O mapa acima, entretanto, mostra a relação entre os dados das Figuras 2.4 e 2.5, organizando os municípios conforme o número estimado de automóveis para cada mil habitantes. O resultado, porém, apresenta um outro ponto de vista sobre a informação, mostrando que a maior quantidade de carros por habitante se dá nos municípios do sul do estado. Captura de tela do serviço WebCart (IBGE, d). 16
- 2.7 Mapas coropléticos retirados de NEWMAN (2012) mostrando os partidos vencedores em cada estado nas eleições presidenciais de 2012 nos EUA. Republicanos em vermelho e Democratas em azul. Mapa à direita distorcido para uniformizar população por unidade de área. . . . 16
- 2.8 Mapas coropléticos retirados de NEWMAN (2012) mostrando os partidos vencedores em cada distrito nas eleições presidenciais de 2012 nos EUA. Republicanos em vermelho e Democratas em azul. Mapa à direita distorcido para uniformizar população por unidade de área. . 17

2.9	Mapas coropléticos retirados de NEWMAN (2012) mostrando os votos por distrito nas eleições presidenciais de 2012 nos EUA. Os Republicanos são representados pela cor vermelha e os Democratas pela cor azul. A cor apresentada no mapa é uma mistura proporcional das duas cores, conforme os votos recebidos por cada partido. Mapa à direita distorcido para uniformizar população por unidade de área.	17
2.10	Mapas coropléticos retirados de NEWMAN (2012) mostrando os votos por distrito nas eleições presidenciais de 2012 nos EUA. Os Republicanos são representados pela cor vermelha e os Democratas pela cor azul. A cor apresentada no mapa é uma mistura das duas cores: vermelho, mais de 70% dos votos para os Republicanos; azul, mais de 70% dos votos para os democratas; roxo, distritos onde nenhum dos partidos recebeu mais de 70% dos votos. Mapa à direita distorcido para uniformizar população por unidade de área.	18
2.11	Recorte da imagem capturada do aplicativo (VIANA <i>et al.</i> , 2015), mostrando um recorte da cidade do Rio de Janeiro e um detalhe para o bairro da Tijuca, em que a mistura aparentemente homogênea apresentada no nível mais baixo não se comprova quando o mapa é mostrado em detalhe.	19
2.12	Exemplo da aplicação de uma grade de população sobre uma imagem de satélite da região de Jaraguá do Sul (RS) para avaliar o efeito de uma eventual inundação sobre a população de parte da cidade. Imagem retirada da apresentação IBGE (2015).	21
2.13	Renderizações de um mapa de densidade populacional gerado no serviço WebCart, do IBGE, mostrando a dificuldade de expressar visualmente a importância de municípios pequenos em tamanho.	22
2.14	Uma versão do gráfico HSV, em formato de cone. Imagem retirada de Wikimedia Commons (2010).	23
2.15	Círculo cromático. Imagem retirada de Wikimedia Commons (TRYGSTAD, 2007).	24
2.16	Escala de matizes.	25
2.17	Escala baseada no matiz amarelo.	25
2.18	Escala de cinzas.	26
2.19	Escala divergente.	26
3.1	Visão geral do aplicativo.	28
3.2	Experimentos preliminares, em que foram testadas escalas baseadas em cores e a utilização de uma textura baseada em uma imagem fotorrealista.	31

3.3	A interpolação dos valores da grade gera limites mais orgânicos para as regiões e o acionamento da transparência e das caixas de texto ajudam a compreender o que acontece naquele ponto.	35
4.1	Interface do sistema: 1 - Mapa; 2 - Variação do nível de detalhe (escala); 3 - Caixa de texto para identificação dos locais; 4 - Tamanho da grade; 5 - Visualização dos limites da grade, interpolação, transparência e mapa no fundo; 6 - Controles e legenda do tema; 7 - Liga/desliga o tema; 8 - Definição da taxa representada no temas; 9 - Classificação do tema;	37
4.2	Matriz de ladrilhos no nível de detalhe 3. Imagem produzida pelo autor baseada em SCHWARTZ (2016).	44
4.3	Exemplos de relações entre os ladrilhos e suas identificações, conforme o nível de detalhe. Imagem produzida pelo autor baseada em SCHWARTZ (2016).	45
4.4	Modificação do nível de detalhe: manutenção da posição e carga de novos ladrilhos, conforme sua <i>quadkey</i>	46
4.5	A imagem mostra o setor censitário nº 330455705250280, localizado no sul da Ilha do Fundão, no Rio de Janeiro, pintado de marrom escuro. Imagem capturada do aplicativo Sinopse por Setores, disponibilizado pelo IBGE em seu portal na Internet (IBGE, 2011).	47
4.6	A figura mostra a associação entre os índices gravados nos <i>pixels</i> da imagem e a posição (u, v) da textura contendo as informações que estão gravadas no espaço RGBA dos <i>pixels</i> das imagens contendo os dados de cada uma das variáveis.	51
4.7	A grade é composta pelo somatório dos valores respectivos à decomposição dos dados pela quantidade de <i>pixels</i> em cada região, naquele determinado nível de detalhe.	52
4.8	A imagem mostra as etapas para calcular os valores mínimos utilizando o método. São necessárias 4 passagens para processar uma textura de 16 x 16 <i>pixels</i>	54
4.9	A imagem mostra as visualizações da grade nas formas simples e interpolada e o resultado após a classificação dos dados.	55
4.10	Fluxo funcional do sistema.	58
5.1	O mapa acima apresenta a quantidade de automóveis para cada 100 mil habitantes e a quantidade de automóveis para cada 10 mil veículos.	59
5.2	O mapa acima apresenta os mesmos dados representados na Figura 5.1. Nesta imagem estão desligadas a imagem do fundo e a camada relativa ao Tema 2.	60

5.3	O mapa acima apresenta os mesmos dados representados na Figura 5.1. Nesta imagem estão desligadas a imagem do fundo e a camada relativa ao Tema 1.	61
5.4	Este mapa é semelhante ao apresentado na Figura 5.1. Aqui, porém, são representadas as quantidade de motocicletas para cada 100 mil habitantes e para cada 10 mil veículos.	62
5.5	A imagem mostra o Tema 1 do mapa representado na Figura 5.4. . .	62
5.6	A imagem mostra o Tema 2 do mapa representado na Figura 5.4. . .	63
5.7	O mapa acima mostra a situação no segundo turno das eleições presidenciais do Brasil em 2014.	63
5.8	A imagem mostra o Tema 1 do mapa apresentado na Figura 5.7. . . .	64
5.9	A imagem mostra o Tema 2 do mapa apresentado na Figura 5.7. . . .	64
5.10	O mapa mostra somente as áreas onde cada candidato recebeu mais da metade dos votos válidos.	65
5.11	Este mapa mostra o Tema 1 do mapa representado na Figura 5.10. . .	65
5.12	Este mapa mostra o Tema 2 do mapa representado na Figura 5.10. . .	66
5.13	O mapa mostra a quantidade de votos recebidos por Dilma Rousseff para cada 100 mil votos recebidos por Aécio Neves. As classes foram organizadas para mostrar onde o candidato venceu ou perdeu a disputa.	66
5.14	O mapa mostra a quantidade de votos recebidos por Aécio Neves para cada 100 mil votos recebidos por Dilma Rousseff. As classes foram organizadas para mostrar onde o candidato venceu ou perdeu a disputa.	67
5.15	O mapa apresenta no Tema 1 a quantidade de votos recebida por Dilma Rousseff para cada 100 mil votos recebidos por Aécio Neves. No Tema 2 é apresentada a situação inversa.	67
A.1	As imagens mostram as diferenças de detalhamento dos polígonos. A imagem acima à esquerda, mostra a renderização do arquivo KML. Ao seu lado é mostrada a imagem gerada pelo arquivo GeoJSON composta por vértices contendo a precisão de 4 casas decimais. As duas imagens abaixo são do mesmo formato, mas com 3 e 2 casas decimais respectivamente. Essas figuras são referentes a capturas de tela do software QGis.	83

Lista de Tabelas

4.1	Tamanho total dos mapas, conforme o nível de detalhe	42
4.2	Significado da codificação do setor censitário <i>nº</i> 330455705090013 . .	46
4.3	Código como identificador de herança	46
A.1	Quantidade de polígonos por Unidades Federativas	78
A.2	Municípios com áreas descontínuas	78
A.3	Tamanho dos arquivos Vetoriais Gerados (kB)	81
A.4	Redução de tamanho promovida nos arquivos	82
A.5	Imagens PNG, RGB de 16 bits contendo apenas os identificadores dos municípios	83
A.6	Imagens JPG (Bing Maps) convertidas para escala de cinzas de 16 bits	84
A.7	Imagens de base (Bing Maps) convertidas para escala de cinzas	84
A.8	Imagens PNG contendo os identificadores de área nos canais R, G e B e a imagem de base (escala de cinzas) no canal A	84
A.9	Imagens PNG contendo os identificadores nos canais B e G e as informações da imagem de base (escala de cinzas) no canal R	85
A.10	Quantidade de imagens necessárias para cobrir a tela e a carga média esperada para o conjunto	85

Capítulo 1

Introdução

O presente trabalho descreve uma ferramenta *on-line* desenvolvida com o propósito de apresentar soluções para a visualização de dados multivariados georreferenciados. Trata-se de um mapa temático interativo que permite ações de posicionamento e enquadramento, em diversos níveis de detalhe, e onde são apresentadas simultaneamente duas variáveis escalares. O projeto foi desenvolvido baseado na premissa de que a velocidade na renderização das imagens é elemento chave para a criação de ambientes que proporcionem uma experiência melhor para o usuário, para que se sinta mais próximo dos dados que está explorando. A melhora da experiência de uso pode ser obtida mediante às respostas visuais instantâneas que o sistema é capaz de prover a partir de perguntas feitas pelo usuário ao mapa através das ferramentas de manipulação. Para atingir esse objetivo, inovou-se ao se construir o sistema por meio da programação em placa gráfica a fim de realizar a montagem e a manipulação de ladrilhos previamente rasterizados, contendo as informações de localização gravadas no espaço de cor dos *pixels* - o que permite a implementação de ações animadas de interação e decomposição espacial dos dados.

1.1 Motivação

A Visualização de Informações é uma área da ciência que estuda as formas de comunicação visual que atuam no registro e comunicação de informações dos mais diversos tipos. Esse procedimento permite a utilização da visão, sentido humano mais desenvolvido, para alcançar novos pontos de vista sobre uma massa de dados e, assim, obter melhor compreensão sobre esse conteúdo.

As representações gráficas, oriundas dos meios impressos, com o advento da Informática se converteram em instrumentos de exploração, utilizando a interação humano-computador para manipular a imagem e, conseqüentemente, o conjunto de dados que lhe dá forma (DIAS e CARVALHO, 2007).

FEW (2013) afirma que “nosso problema não é a falta de dados, mas a incapacidade de compreender as informações disponíveis” e alega que esses dados não fazem sentido sozinhos, isolados em suas posições nas tabelas, mas somente quando convertidos em informação. Por isso, a compreensão, e consequente comunicação, desse conteúdo é um objetivo importante. Para tanto, “os dados precisam ser expressos de modo que os olhos possam facilmente ver e o cérebro facilmente entender (...) Para explorar e entender os dados, precisamos das habilidades corretas potencializadas por boas ferramentas”.

A Visualização de Informações integra dois sistemas de processamento: o computador, que organiza e transcreve dados em imagens, e a mente, que através das capacidades cognitivas da visão, consegue assimilar esse conteúdo. Visualizações eficazes são aquelas capazes de fazer com que o usuário consiga compreender rapidamente grandes volumes de dados, através da revelação de características, padrões e tendências (GERSHON *et al.*, 1998).

Nesse contexto, a visualização das informações georreferenciadas, que são aquelas especificamente relacionadas ao espaço geográfico, assume um papel relevante. Especialmente nos últimos anos, com a popularização do *webmapping* - termo que define sistemas de divulgação de mapas digitais através da Internet. Ferramentas como *Google Maps*, *Bing Maps*, *OpenLayers*, entre outras, viabilizaram uma infinidade de serviços, através dos quais a utilização dos mapas se tornou prática comum.

A compreensão do espaço geográfico se dá também pela assimilação da dimensão humana da Geografia, visto que o homem, com suas ações produtivas e sociais, constrói e modifica o seu meio, integrando elementos naturais e sociais. As informações socioeconômicas registram o desenrolar de fatos e transações que redesenham o espaço, revelando novas realidades. Registrar e visualizar esses limites abstratos são atividades determinantes para que se possa compreender a própria Geografia, para quem é fundamental saber o quê ocorre, onde ocorre, suas causas e consequências (RODRIGUEZ, 2010).

No Brasil, a instituição governamental responsável por pesquisar esse tipo de informações é o IBGE, cuja missão consiste em “Retratar o Brasil com informações necessárias ao conhecimento de sua realidade e ao exercício da cidadania” (IBGE, c). As atividades de mapeamento e divulgação de dados são, portanto, funções essenciais do instituto, e demandam soluções eficientes para a visualização de mapas, cartas e outros produtos relacionados. O principal meio de divulgação das informações levantadas pelo instituto é o seu portal na Internet (IBGE, 2016b), que disponibiliza gratuitamente um gigantesco volume de dados públicos, todos associados ao território.

1.2 Objetivos

Tendo em vista a crescente demanda por visualizações de informações georreferenciadas, existe a necessidade de desenvolvimento de novas soluções para a utilização de mapas digitais interativos. O aumento da capacidade de processamento dos aparelhos móveis, especialmente levando-se em conta o desenvolvimento de novas placas gráficas, menores e mais sofisticadas, juntamente com o advento de tecnologias como o HTML5, o JavaScript e o WebGL, criam um ambiente altamente favorável a inovações.

A Visualização de Informações, nesse cenário, merece destaque, uma vez que promove uma compreensão única do espaço geográfico através da expressão de relacionamentos implícitos e complexos, incapazes de serem assimilados de outro modo.

O objetivo desse estudo é propor uma ferramenta eficiente para visualização de informações georreferenciadas, com base em conhecimentos adquiridos na Computação Gráfica para desenvolver um mapa interativo *on-line*, capaz de promover a visualização de variáveis diversas e simultâneas, com velocidade suficiente para uma interatividade ágil e uma boa experiência de uso para o público.

Capítulo 2

Fundamentação Teórica

2.1 Visualização de Dados Multivariados

Para se chegar ao modelo adotado neste trabalho, vários métodos de visualização de dados multivariados foram analisados. A seguir, são apresentados os mais relevantes para este estudo, ou seja, aqueles com especial ênfase na área estatística, sobre grupos de dados relacionados e com aplicabilidade sobre mapas.

Todos os estudos selecionados para esta seção se baseiam na utilização de glifos para expressar as dimensões dos dados. No contexto da Visualização de Informações, um glifo pode ser descrito como “um objeto gráfico cujas propriedades representam valores associados a variáveis” (KOSARA, 2007) e seu uso aplicado à visualização de mapas pretende aproveitar a capacidade humana de reconhecimento de padrões e texturas para promover uma visualização simultânea de algumas dimensões de dados associadas a um ponto no espaço, assimilando assim suas relações de vizinhança e ocupação.

2.1.1 Faces de Chernoff

Em um artigo escrito em 1973, Herman Chernoff propõe um método capaz de representar muitas dimensões de uma informação associando esses dados às características de uma face esquemática, como o comprimento do nariz e a curvatura da boca (CHERNOFF, 1973). A intenção é que a imagem gerada revele padrões sujeitos a relacionamentos fortes, porém de visualização complexa (Figura 2.1).

O autor salienta que as faces nada mais são do que estimativas grosseiras, mas que o método traz consigo a grande vantagem de revelar uma consciência qualitativa sobre dados relevantes, além de ser promissor para uma análise preliminar em conjuntos de dados, auxiliando na identificação de grupos, separações e mudanças em sequências temporais.

Outra vantagem é que quanto maior o número de variáveis, mais robusto fica o

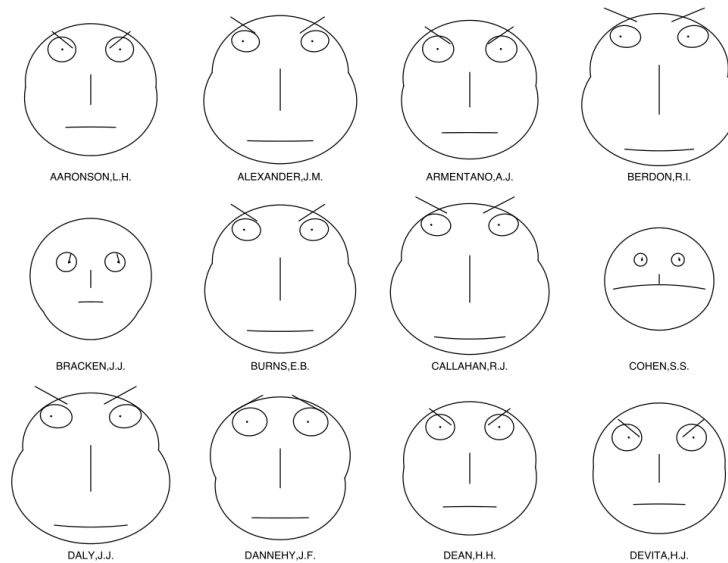


Figura 2.1: Exemplo de uso das faces esquemáticas proposta por Chernoff. Imagem retirada de Wikimedia Commons (2010).

método, conforme comprova um experimento feito com sete variáveis inicialmente e depois com mais cinco. A inclusão destas destacou ainda mais os pontos relevantes, enriquecendo a informação. Em contrapartida, nos casos em que foram analisadas poucas dimensões, a quantidade de ruído diminuiu a capacidade de destacar informações relevantes.

No artigo foram experimentados dados com até 18 dimensões, porém é possível pensar em outras mais, acrescentando novas características à face, como cabelos, orelhas ou novas linhas de expressão.

O método de Chernoff é muito difundido e pode ser encontrado facilmente em diversos exemplos na literatura e em aplicações diversas. Porém, há muita discussão acerca de sua efetividade.

KOSARA (2007) discute a funcionalidade do método alegando que os resultados são questionáveis, sem exemplos de uso sério da técnica. Ele alega que o reconhecimento da face não se dá pela soma de suas características, mas sim por meios inconscientes, focados no reconhecimento do estado de espírito daquele rosto. Além disso, há uma hierarquia rígida na visualização dos detalhes, pela qual olhos e boca são muito mais importantes do que o formato do rosto, por exemplo.

Já WARE (2004) entende que a técnica funciona em alguns casos, porém ressalta que a não linearidade do espaço perceptivo do método exige esforço na comparação entre os detalhes do rosto. Assim, alguns elementos, conforme a sua hierarquia natural, podem ser menos comparáveis do que outros.

Falando especificamente da aplicação do método sobre um mapa, é importante ressaltar que o rosto proposto por Chernoff demanda uma área generosa para que os detalhes sejam legíveis. Um mapa subdividido em regiões muito pequenas pode não

atender a essa limitação, como mostram alguns exemplos disponíveis em HUFFMAN (2010).

2.1.2 Figuras de Palito

PICKETT e GRINSTEIN (1988) propõem a representação de diversas dimensões de um dado utilizando figuras de palitos (*stick figures*). Nelas, os valores são associados diretamente aos ângulos de pequenos traços conectados entre si pelas extremidades, gerando uma coleção de glifos que, organizados sobre um mapa, produzem uma textura capaz de revelar estruturas e relacionamentos implícitos do conjunto.

O trabalho apresenta 12 famílias de ícones, que devem ser selecionadas e testadas, caso a caso. O desafio é elaborar figuras que sejam capazes de criar texturas expressivas. Ademais, a geometria de cada ícone pode favorecer alguns conjuntos e dados e desfavorecer outros, o que exige algumas tentativas até o acerto da melhor solução para cada conjunto.

O glifo em questão é composto por cinco traços, que combinados resultam nas 12 figuras possíveis. Um desses traços é o corpo; os outros, braços e pernas. Cada um é associado a uma variável que determina o ângulo entre ele e o membro ao qual se liga na extremidade anterior. A dimensão determina a rotação do corpo e gira todo o conjunto. Cada ícone produzirá uma textura distinta, assim como a ordem que associa suas características ao conjunto de dados que lhe configura. Dessa forma, as 12 figuras possíveis, combinadas com as 5 variáveis, permitem a composição de 1440 texturas diferentes.

Animar o conjunto, interagindo com ele, pode viabilizar a visualização das diversas possibilidades de composição das texturas e, assim, facilitar a seleção das mais adequadas.

O estudo não apresenta testes massivos para comprovar a eficácia do método, sendo assumidamente uma tentativa de explorar o problema. O experimento apresentado no artigo é a tradução da imagem espectral de um detalhe da região a oeste do Lago Ontário, que separa o Canadá dos EUA. É possível identificar na imagem o formato característico do lago e separá-lo das regiões de terra. O ícone escolhido para a representação foi o que produziu a textura que melhor enfatizava esta diferenciação.

A organização dos dados em classes diminuiria a variedade formal dos ícones e poderia tornar a imagem mais legível. É possível imaginar também que uma aplicação interativa e animada do conjunto permitisse demonstrar com mais clareza as transições entre os conjuntos e reforçar a interpretação da massa de dados representada.

2.1.3 Simulação de Texturas Naturais

INTERRANTE (2000) apresenta um método que utiliza texturas geradas a partir de imagens naturais rasterizadas para representar as classificações de um conjunto multivariado de dados. Dos métodos apresentados nesta dissertação, é o único que trata diretamente das classificações, sem tentar expressar a quantificação através de proporções visuais. As imagens possuem aparência realista e a intenção principal do estudo, mais do que mostrar resultados eficientes, é experimentar uma nova forma de visualização, aproveitando a computação para sair do lugar comum estabelecido pelos métodos tradicionais de representação.

A motivação da escolha de texturas naturais se deu porque são esteticamente agradáveis e adicionam alguma imperfeição à imagem, o que contribui para uma visualização mais confortável. Além disso, há aspectos culturais inerentes às texturas naturalistas que agregam expressividade à imagem.

Os resultados do método apontam para um claro potencial das texturas naturais na representação de dados multivariados, destacando ainda a capacidade de revelarem a natureza precisa/imprecisa dos dados de maneira isenta, ao contrário da cor, que toma partido de juízos de valor. Mas o uso de texturas para visualização ainda depende de muitas questões, como a realização de testes de percepção nos usuários e o desenvolvimento de algoritmos específicos para sua manipulação.

O método também requer que as regiões representadas tenham um tamanho suficiente para a exibição da textura, o que pode ser um entrave na representação de determinados mapas.

Outra questão importante é a garantia de legibilidade para as texturas intermediárias, oriundas de interpolações entre duas texturas vizinhas.

2.1.4 Simulação de Pintura a Óleo no Estilo Impressionista

TATEOSIAN *et al.* (2007) descrevem um método para visualização de dados multivariados através da simulação de uma pintura artística, inspirada nas técnicas usadas pelos pintores impressionistas, cujas características das pinceladas, como cores, relevo, comprimento, largura, curvatura e transparência, derivam das informações contidas na massa de dados.

A intenção principal do método é fazer com que o resultado estético obtido proporcione um maior engajamento dos usuários na visualização, pois os autores apostam no fato de que o aspecto estético agradável e intrigante incentiva à exploração da imagem e, conseqüentemente, das informações contidas nela.

O artigo discute alguns conceitos acerca da expressividade das pinturas artísticas, das técnicas utilizadas pelos artistas e dos algoritmos de renderização não fotorrealista que realizam este tipo de simulação.

Os testes apontaram para o fato de que o método realmente chama a atenção dos usuários e atiza sua curiosidade, indicando uma continuidade da pesquisa com a realização de novos experimentos que procurem quantificar a qualidade estética das imagens perante os usuários, o caráter passional carregado na imagem e as propriedades compositivas dos métodos de renderização.

A complexidade computacional da sua implementação, porém, pode ser considerada um entrave, pois dificulta uma abordagem interativa, o que certamente agregaria bastante valor aos resultados.

2.2 Mapas Temáticos e Semiologia Gráfica

A pesquisa descrita nesta dissertação envolve o desenvolvimento de um mapa interativo, e, por isso, pressupõe informações sobre como a Cartografia aborda o problema da representação visual de informações sobre um mapa temático.

Os mapas temáticos devem cumprir a função de mostrar “o quê, onde e como ocorre determinado fenômeno geográfico” e o fazem através de signos gráficos planejados para levar à compreensão das diferenças, ordenamentos e quantificações dos elementos, possibilitando a visualização de correlações entre os dados (ARCHELA e THÉRY, 2008). Os diversos assuntos que esses mapas são capazes representar fazem com que suas convenções sejam muito mais flexíveis do que as dos mapas topográficos, que possuem uma herança histórica atrelada a áreas específicas de conhecimento em cartografia.

Para que alcancem seus objetivos, os mapas temáticos são fundamentados na Semiótica, ciência que investiga as linguagens (QUEIROZ, 2011), e são compostos por um sistema de sinais impressos em uma peça de comunicação visual que transcreve as relações entre alguns atributos e sua localização. A Linguagem Cartográfica, portanto, trata da representação gráfica dos fenômenos, operação básica para registro, análise e comunicação do conhecimento geográfico (RODRIGUEZ, 2010).

As informações contidas em um mapa precisam atender a cinco perguntas - **O que? Onde? Quanto? Como? E porquê?** Para respondê-las, o mapa se vale de elementos visuais capazes de expressar essas idéias.

Um mapa com profundidade informacional precisa esclarecer as relações entre seus atributos. Partindo do princípio de que o olho só vê uma forma de cada vez, quando há vários sinais gráficos superpostos, só é possível assimilar a textura do conjunto. Em outras palavras, para identificar um dado, é necessário separá-lo dos outros, pois quanto maior o número de atributos simultâneos, maior a complexidade da visualização.

Ao visualizar os dados contidos em um mapa, o usuário processa essas informações e, por isso, uma representação gráfica original tem que ser capaz de permitir

uma visualização que vá do conjunto ao detalhe, de modo que sejam reveladas as relações que respondam às perguntas feitas à imagem. Para cumprir esse objetivo, é preciso interagir com os dados, questionando quais são os conjuntos de atributos e quais as suas ordens. Só assim serão reveladas as relações que esses grupos estabelecem (ARCHELA, 1999).

Para atender a essas questões de representação, surgiu a Semiologia Gráfica, proposta por Jacques Bertin a partir de estudos das teorias da informação aplicadas à Cartografia. Seus estudos resultaram na criação de uma linguagem sistematizada que explora a percepção humana para a composição de mapas, convertendo-se em uma espécie de gramática da linguagem cartográfica (RODRIGUEZ, 2010).

2.2.1 Semiologia Gráfica

A Semiologia Gráfica pode ser definida como um conjunto de diretrizes que orientam a elaboração de mapas temáticos com o uso de símbolos caracterizadores da informação, tomando por base a semiótica (ARCHELA, 2001).

Segundo Bertin todo mapa é a expressão de uma tabela que relaciona linhas contendo localizações geográficas, (x, y) , e colunas contendo atributos, (z) , ou seja, informações associadas a uma determinada posição no espaço. Fazer um mapa, portanto, é transcrever para um suporte gráfico as relações contidas nessa tabela (ARCHELA, 1999).

Como o ser humano não integra mais de sete conceitos combinatórios sobre o mesmo problema, é preciso reduzir esse complexo conjunto de informações a um número acessível de conceitos. O objetivo se torna, portanto, reduzir esse volume de dados aos agrupamentos espaciais significativos que as relações desse conjunto constroem. Essa deve ser a meta dos tratamentos estatísticos e, particularmente, dos tratamentos multivariados (BERTIN, 1980).

Para isso, é preciso descobrir elementos semelhantes, agrupá-los e classificá-los para formar os grupos que contenham as relações importantes. Assim, todo mapa deve ser capaz de responder quais são as informações contidas nas posições (x, y) e quais são os grupos que esses atributos (z) formam (BERTIN, 1980).

De acordo com Bertin, o modelo tradicional da teoria da comunicação - **emissor, código, receptor** - não é apropriado para a comunicação cartográfica, pois esta possui linguagem própria, que não se baseia em nenhum código arbitrário, e sim na elaboração de regras lógicas que conduzem à transcrição das relações entre os elementos. Não há interpretação pessoal, mas a leitura de uma peça monossêmica, isto é, de uma obra de interpretação única e objetiva, na qual emissor e receptor se encontram do mesmo lado do processo, transformando-se em atores que têm pela frente a mesma questão: identificar as localizações, ordens e proporções entre os ele-

mentos, e assim organizá-los conforme os grupos construídos pelos dados (BERTIN, 1980; RODRIGUEZ, 2010).

A localização é a variável fundamental para a representação cartográfica, afinal sua informação essencial responde à pergunta “onde?”. Sem ela não há mapa. As outras variáveis serão expressas de três maneiras:

- **Qualitativa (ou nominal)** [≠]: Apresenta as relações de diversidade entre os atributos;
- **Ordenada (ou sequencial)** [O]: Apresenta os ordenamentos dos atributos;
- **Quantitativa** [Q]: Apresenta as relações de proporção entre os atributos;

Para isso, o mapa se vale de sinais gráficos capazes de traduzir essas noções. Similaridade, ordem e proporção se transformam, portanto, em variáveis visuais, como **tamanho, valor, textura, cor, orientação e forma**, que são transcritas em um suporte gráfico através das três formas possíveis de implementação - **o ponto, a linha e o plano** (Figura 2.2).

AS VARIÁVEIS VISUAIS		MODOS DE IMPLANTAÇÃO						PROPRIEDADES PERCEPTIVAS	
		PONTUAL			LINEAR				
• IMAGEM a) DUAS DIMENSÕES DO PLANO (XY) b) TAMANHO c) VALOR	(Z)								
• SEPARAÇÃO a) GRANULAÇÃO b) COR c) ORIENTAÇÃO d) FORMA	(Z)								

Figura 2.2: Variáveis visuais (RODRIGUEZ, 2010).

A similaridade representa não apenas as relações de semelhança (e diferença) entre os atributos, mas também a própria existência deles. A ordenação promove

uma relação hierárquica entre os atributos. E a quantificação é a expressão da proporcionalidade entre os atributos (RODRIGUEZ, 2010).

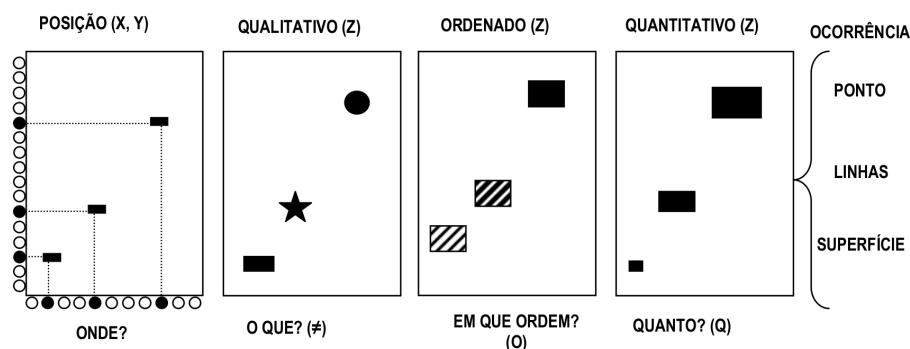


Figura 2.3: Aspectos evidenciados na Cartografia Temática (RODRIGUEZ, 2010).

A representação gráfica não é convencional, pois transcreve uma relação por outra relação. Semelhanças entre os dados são transcritas em semelhanças visuais entre os símbolos; ordenamentos são transcritos por ordens visuais entre os símbolos; e proporções são transcritas por proporções visuais entre os símbolos.

O olho não observa o significado de um único símbolo, mas a comparação entre um grupo deles. Consequentemente, transcrever uma ordem por uma semelhança não significa adotar uma convenção, e, sim, construir falsos agrupamentos. A representação gráfica não é livre, e, por isso mesmo, é universal (BERTIN, 1980).

Realizando essas transcrições, o usuário poderá constatar como se organizam espacialmente os atributos e identificar o que há em cada posição, formulando questões de três níveis para o mapa:

1. **Elementar:** “tal lugar, tal atributo, quanto?”
2. **Médio:** “tal lugar, quais são seus atributos?”
3. **Conjunto:** “como se agrupam espacialmente os atributos?”

Esses agrupamentos só serão possíveis se conservadas as posições espaciais dos atributos e a sua transcrição segundo as relações gráficas equivalentes de semelhança, ordem e quantidade (BERTIN, 1980).

Em um ambiente interativo, manipular o mapa é reorganizar essa tabela, reposicionando linhas e colunas, e aqui reside o cerne das questões de interação dos mapas digitais. A cada reorganização, o resultado aparece na imagem, revelando os grupos (semelhanças), as ordens e as quantidades. Essa manipulação permite mudar os níveis das questões feitas para o mapa, indo do elementar para o conjunto e vice-versa.

Trata-se de um processo onde a velocidade no processamento dos dados e, especialmente, a renderização da imagem são fundamentais para a revelação das relações.

Processar os dados se torna, então, “descobrir os agrupamentos em (x, y) que as relações (z) constroem. Assim, os tratamentos gráficos e os tratamentos multivariados completam a estatística clássica que calcula um coeficiente de correlação, ou uma lei de correspondência entre duas linhas da tabela” (BERTIN, 1980).

Para que tudo isso aconteça de maneira satisfatória, há que se compreender algumas questões ligadas à interatividade e manipulação dos mapas e, por esse motivo, foram pesquisados alguns textos que tratavam do assunto.

2.2.2 As Requisições Dinâmicas

SHNEIDERMAN (1994) escreve sobre as Requisições Dinâmicas¹, que ele define como sendo “o controle interativo de filtros sobre um conjunto de dados, produzindo uma resposta visual imediata em tempo menor que 100 milissegundos” e que seriam uma alternativa intuitiva para as tradicionais *queries* realizadas em bancos de dados, via linguagens como o SQL. Para tanto é preciso converter os signos textuais utilizados nas requisições tradicionais em comandos acionados na tela, acompanhados de respostas imediatas, em uma velocidade equivalente ao limite mínimo perceptivo para a visão humana entender a imagem como uma animação (DEHAAN e MAYHEW, 2011; MACHADO, 2011). Deste modo, os usuários são encorajados a experimentar os dados e procurar padrões que saltem aos olhos, expandindo a sensação de imersão no volume de informações.

As Requisições Dinâmicas são uma aplicação prática dos princípios de manipulação direta em ambientes de dados, descritos em SHNEIDERMAN (1983), e que necessitam de algumas condições essenciais para funcionamento, como um resultado visual, ferramentas que possibilitem filtros, renderização instantânea e animada das transições, cliques no lugar de comandos de texto² e modos de reversão nas ações realizadas.

A sensação de controle sobre a massa de dados, a facilidade com que se faz a pergunta e se recebe uma resposta, a composição de imagens através de controles intuitivos, como barras deslizantes, promovem uma sensação de aprendizagem incidental³ que contrasta com o burocrático padrão de realizar requisições textuais ou outros procedimentos equivalentes.

No caso específico dos mapas, tudo está relacionado mais diretamente às questões de vizinhança, onde as similaridades e diferenças produzirão os padrões visuais que revelam as relações de localização entre os dados: “aplicativos geográficos emergem naturalmente como candidatos para as requisições dinâmicas” (SHNEIDERMAN, 1994).

¹Da expressão “*Dynamic queries*”.

²Da expressão “*pointing than typing*”

³Do termo “*incidental learning*”.

A velocidade também confere confiança ao sistema, o que é fundamental para o efeito lúdico, e incentiva a experimentação nas diversas dimensões dos dados: onde se apresentam densos, esparsos, diferentes, vazios ou com qualquer outro aspecto que faça sentido visualmente. Para obter estes resultados, o autor defende alguns princípios para a manipulação direta de massas de dados, como uma representação visual para todos os tipos de requisições e resultados das ações, que devem ser rápidas, incrementais, reversíveis e operadas através de cliques e não mais por digitação, produzindo respostas visuais imediatas e contínuas.

A percepção humana na leitura de imagens é muito eficiente e respostas rápidas são essenciais para garantir o sucesso do sistema, mas é preciso cumprir algumas exigências, como permitir ações de movimentação e nível de detalhe, permitir o controle dos intervalos entre os dados através de barras deslizantes, implementar a opção por combinações booleanas nas classes e a visualização de muitas dimensões dos dados.

As idéias do autor se afastam da complicação proposta pelas técnicas de visualização simultânea de múltiplas dimensões dos dados e vão ao encontro da cartografia tradicional, com a utilização de métodos simples, porém potencializados pelas possibilidades tecnológicas oferecidas pelos ambientes interativos digitais. É esse o propósito da presente pesquisa e com essa intenção foram investigados alguns trabalhos relacionados à representação simples de dados em mapas coropléticos, de modo que essas soluções pudessem ser replicadas em um ambiente interativo.

2.3 Representação de Dados através de Mapas Coropléticos

A utilização de mapas para representar a quantidades é uma tarefa difícil. Segundo Bertin, o tamanho é a única variável visual capaz de representar adequadamente uma proporção. As cores, em teoria, também poderiam desempenhar este papel, já que se baseiam na quantificação de componentes elementares, que poderiam estar associados às dimensões de uma informação. Porém, o resultado perceptivo não condiz com essa nitidez matemática, exigindo acuidade visual e conhecimento prévio dos métodos de composição cromática por parte dos usuários.

Portanto, a representação de fenômenos quantitativos costuma ser realizada através de classificações, pelas quais os dados são agrupados conforme seu valor, promovendo um resultado gráfico simplificado, porém eficaz.

Um dos métodos mais difundidos para visualização desse tipo de informação são os mapas coropléticos, um tipo específico de cartograma em que os valores representados são agrupados em classes associadas a uma escala cromática que transcreve

algum tipo de diferenciação visual, normalmente um ordenamento (ARCHELA e THÉRY, 2008, p. 8).

Apesar de funcionarem para grandezas escalares em geral, os mapas coropléticos são muito indicados para expor a distribuição de densidades (habitantes por km²), rendimentos (toneladas por hectare), percentagens ou outras grandezas que sejam relacionadas a mais de um elemento (ANDRIENKO, 2001).

As Figuras 2.4, 2.5 e 2.6 mostram como os mapas coropléticos podem expressar de maneira diferente um mesmo conjunto de dados. Os dois primeiros mapas os totais da população e da quantidade de automóveis nos municípios do estado de Minas Gerais em 2015. Os valores são absolutos e estão distribuídos em três classes, compostas de modo que contenham as mesmas quantidades de municípios em cada uma. O terceiro mapa informa uma relação entre as duas variáveis, apresentando a quantidade de automóveis para cada mil habitantes. A diferença é significativa, já que a quantidade de automóveis e de habitantes são variáveis bastante relacionadas entre si, pois onde há mais pessoas, provavelmente há mais carros. Já o último mapa apresenta um outro ponto de vista sobre os mesmos dados, relacionando-os em uma taxa que mostra uma nova perspectiva sobre a informação.

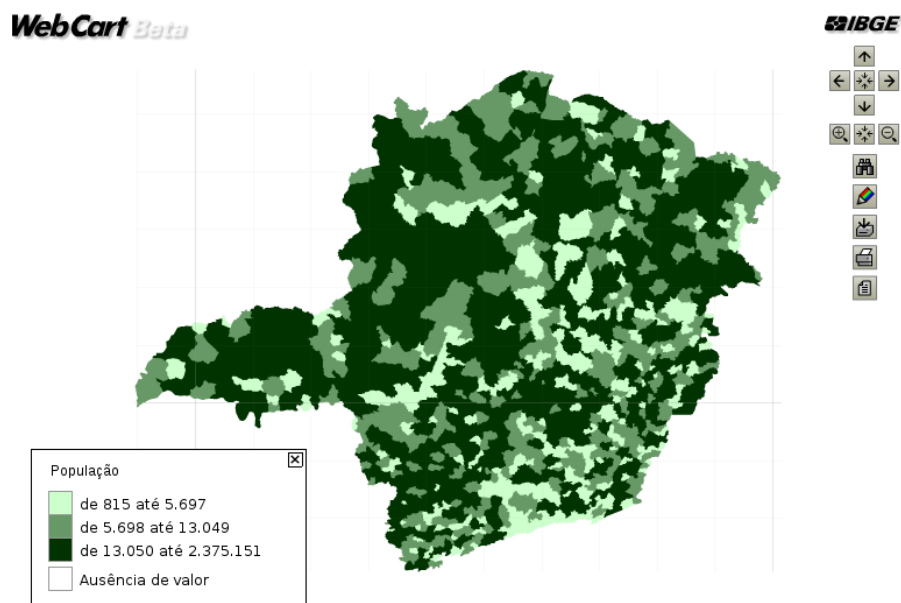


Figura 2.4: O mapa acima mostra a população residente nas cidades de Minas Gerais em 2015. Captura de tela do serviço WebCart (IBGE, d).

ANDRIENKO (2001) discute essa questão e mostra mais exemplos de como classificação dos valores pode mudar completamente o mapa. No artigo, são apresentados quatro mapas das taxas de natalidade dos países da Europa, cada qual com uma classificação para os dados. Um deles tem seus dados organizados em intervalos iguais de valores, desde o mínimo até o máximo dentre os dados. No segundo, os

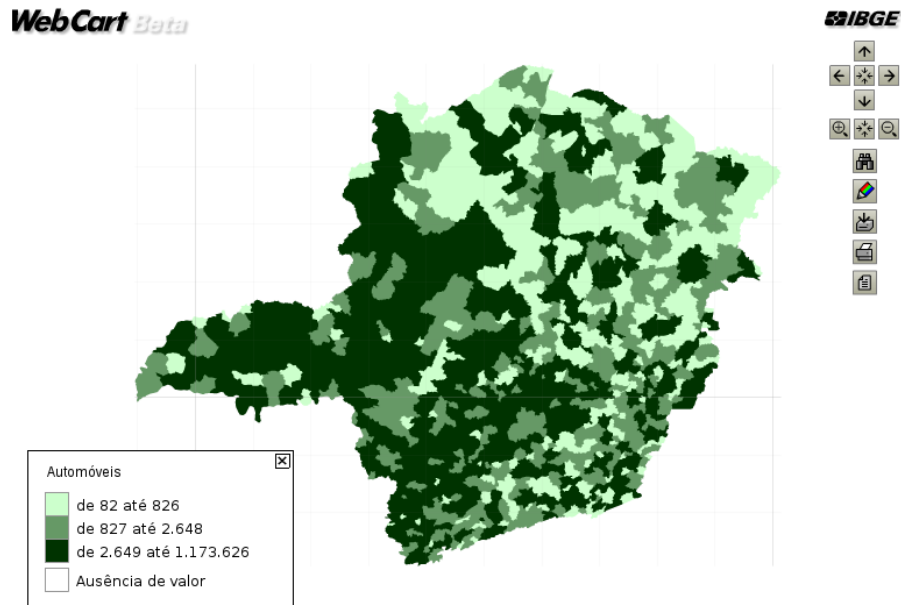


Figura 2.5: O mapa acima, por sua vez, mostra a quantidade de automóveis nas cidades de Minas Gerais em 2015 e aponta para uma relação entre este dado e a população residente, apresentada no mapa da Figura 2.4. De um modo geral, onde há mais habitantes, há mais automóveis. Captura de tela do serviço WebCart (IBGE, d).

dados são ordenados e a classificação é dividida em intervalos iguais, o que promove uma visualização com a mesma quantidade de regiões em cada classe. Já o terceiro foi composto de acordo com uma classificação estatística ótima descrita pelo autor. Os mapas apresentados no artigo são completamente diferentes entre si, apesar de representarem os mesmos valores.

O autor trata também da importância de se desenvolver um conjunto integrado para exploração espacial que ofereça ferramentas de interação e métodos de análise estatística, como medidas de qualidade estatística, métodos tradicionais de classificação, manipulação direta das classes; gráficos que mostrem a distribuição estatística dos valores, como as curvas de frequência, escalas de cor interativas, gráficos auxiliares; resumos das classes, como mínimos, máximos e médias, e métodos de mineração de dados aplicados às classes (ANDRIENKO, 2001).

Em outro artigo, NEWMAN (2012) apresenta uma breve análise sobre a utilização de mapas coropléticos e anamórficos para apresentar os resultados das eleições presidenciais americanas de 2012.

Primeiro, é apresentado um mapa coroplético por estados, colorindo a área conforme o partido vencedor naquela unidade federativa, republicanos em vermelho e democratas em azul. O mapa apresenta quem venceu em cada estado. Um olhar desatento pode interpretar mal e acabar concluindo que os republicanos tenham vencido o pleito, uma vez que as áreas vermelhas são maiores que as azuis. Porém,

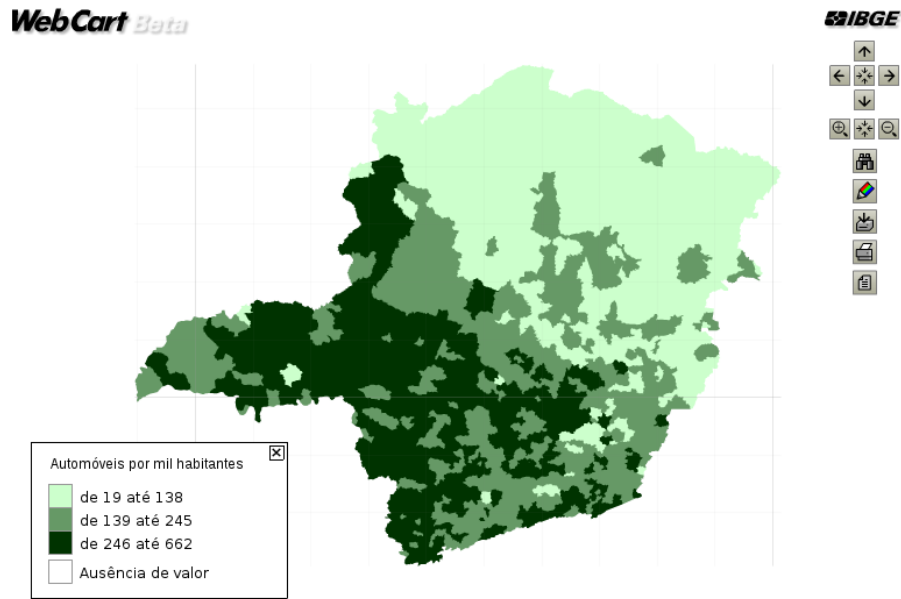


Figura 2.6: O mapa acima, entretanto, mostra a relação entre os dados das Figuras 2.4 e 2.5, organizando os municípios conforme o número estimado de automóveis para cada mil habitantes. O resultado, porém, apresenta um outro ponto de vista sobre a informação, mostrando que a maior quantidade de carros por habitante se dá nos municípios do sul do estado. Captura de tela do serviço WebCart (IBGE, d).

área dos estados não exprime sua população, que é quem decide a eleição, vencida pelos democratas.

Na tentativa de melhorar a visualização, foi aplicado um algoritmo que promoveu uma anamorfose cartográfica sobre o mapa, de modo que as áreas fossem distorcidas para que ficassem com tamanhos proporcionais às suas populações. O resultado é esteticamente interessante, porém confuso, pois embora se veja que azul passou a ocupar uma área maior que o vermelho, traduzindo a vitória dos democratas, há uma distorção que acaba dificultando a identificação dos territórios.

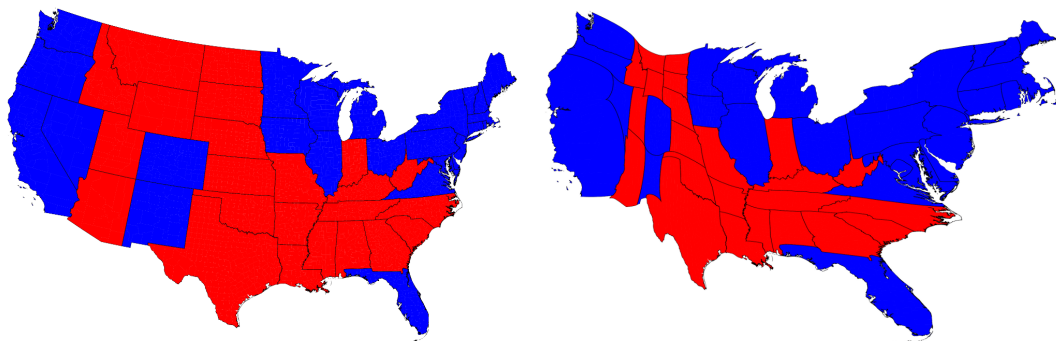


Figura 2.7: Mapas coropléticos retirados de NEWMAN (2012) mostrando os partidos vencedores em cada estado nas eleições presidenciais de 2012 nos EUA. Republicanos em vermelho e Democratas em azul. Mapa à direita distorcido para uniformizar população por unidade de área.

As regras eleitorais dos Estados Unidos definem que a eleição do presidente não se dá por votação direta, mas por uma eleição indireta a partir da votação de um colegiado de representantes eleitos nos distritos. Então, foram confeccionados novos mapas com os números da divisão distrital. Os resultados se mostraram ainda mais confusos, uma vez que a distorção desta nova anamorfose produziu um mapa ainda mais complexo, onde os distritos, bem menores que os estados, ficaram mais difíceis de se identificar.

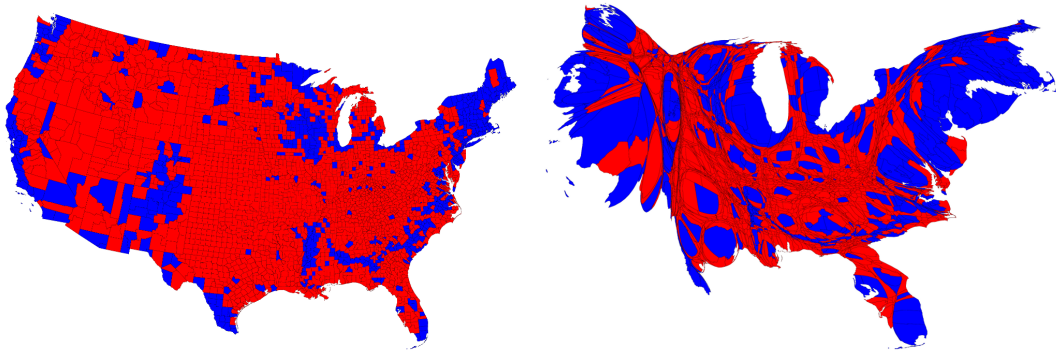


Figura 2.8: Mapas coropléticos retirados de NEWMAN (2012) mostrando os partidos vencedores em cada distrito nas eleições presidenciais de 2012 nos EUA. Republicanos em vermelho e Democratas em azul. Mapa à direita distorcido para uniformizar população por unidade de área.

Aplicou-se, então, uma escala de cor que partia do vermelho, quando 100% dos votos eram republicanos, e ia até o azul, para quando todos os votos eram democratas, passando por tons de roxo resultantes da mistura proporcional das duas cores, conforme os votos de cada partido.

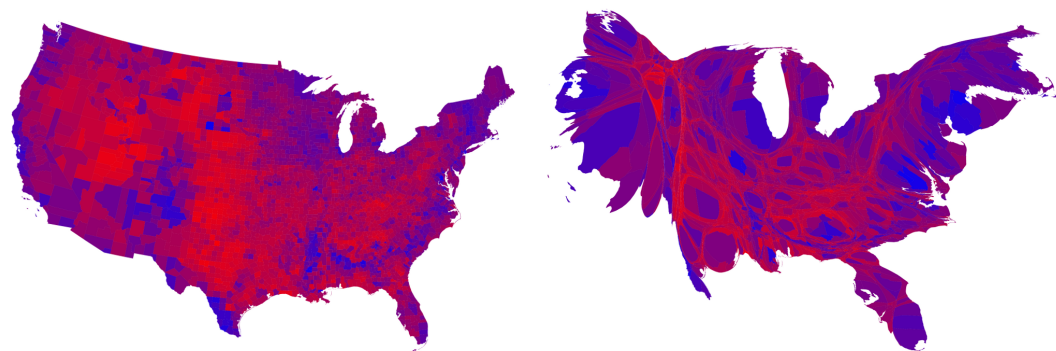


Figura 2.9: Mapas coropléticos retirados de NEWMAN (2012) mostrando os votos por distrito nas eleições presidenciais de 2012 nos EUA. Os Republicanos são representados pela cor vermelha e os Democratas pela cor azul. A cor apresentada no mapa é uma mistura proporcional das duas cores, conforme os votos recebidos por cada partido. Mapa à direita distorcido para uniformizar população por unidade de área.

O resultado é, novamente, uma imagem expressiva porém confusa, onde as cores

transcrevem o comportamento dos números, mas a distorção destrói as referências de localização.

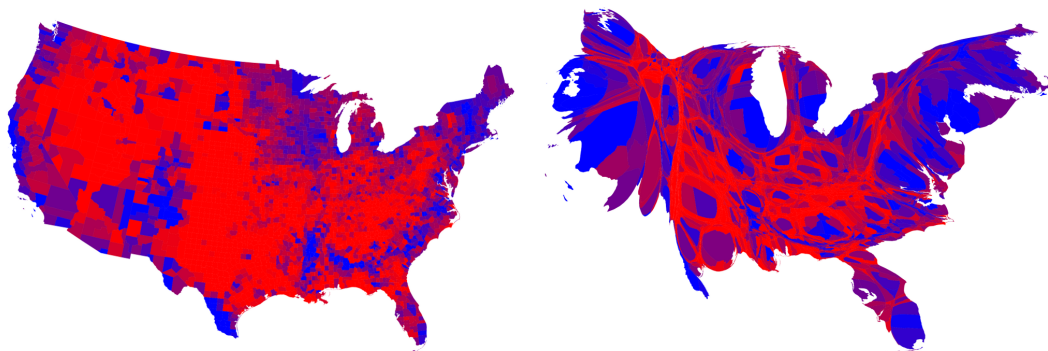


Figura 2.10: Mapas coropléticos retirados de NEWMAN (2012) mostrando os votos por distrito nas eleições presidenciais de 2012 nos EUA. Os Republicanos são representados pela cor vermelha e os Democratas pela cor azul. A cor apresentada no mapa é uma mistura das duas cores: vermelho, mais de 70% dos votos para os Republicanos; azul, mais de 70% dos votos para os democratas; roxo, distritos onde nenhum dos partidos recebeu mais de 70% dos votos. Mapa à direita distorcido para uniformizar população por unidade de área.

Para tentar resolver o problema, foram elaborados dois novos mapas, um anamórfico, outro não, com uma pequena mudança com relação à escala cromática, que passou a se referir a uma classificação que definia que vermelhos seriam os distritos com mais de 70% de votos republicanos, azuis os que tivessem mais de 70% de votos democratas e roxos os que tivessem qualquer outro resultado.

É possível supor que uma boa implementação permita a interatividade e a transição animada entre os estados do mapa, o que poderia garantir a identificação das posições. Porém, trata-se de algoritmos muito complicados, envolvendo otimizações computacionalmente onerosas, o que dificultaria a interação. HEILMANN *et al.* (2004) apresentam um algoritmo de anamorfose cartográfica e discutem sua complexidade.

O método dos pontos de contagem é uma alternativa comum para expressar valores absolutos. Ela funciona organizando pontos sobre o espaço de modo que representem uma fração precisa do total de uma variável. Para que funcione bem, é preciso que seja feita uma boa distribuição espacial dos pontos, de modo que representem o espalhamento dos dados. Sua construção depende, então, de duas decisões: qual valor será atribuído a cada ponto e como estes pontos serão distribuídos dentro da área a ser mapeada? (ARCHELA e THÉRY, 2008).

Um exemplo desta técnica pode ser visto no trabalho de VIANA *et al.* (2015). Assumidamente baseado em um outro aplicativo divulgado pelo US Census Bureau (WELDON COOPER CENTER FOR PUBLIC SERVICE, 2010), o mapa apresenta a distribuição da população do país conforme a declaração de “cor ou raça” constante

nos dados do Censo 2010 do IBGE.

O aplicativo usa o Google Maps como suporte para exibir ladrilhos pré-processados, pintados por 190 milhões de pontos (um para cada habitante do país) cujas cores são associadas aos cinco grupos étnicos autodeclarados na pesquisa: brancos, pretos, pardos, indígenas e amarelos (IBGE, 2008). As cores usadas para representar as raças são, respectivamente, azul, vermelho, verde, marrom e amarelo. Os pontos são distribuídos aleatoriamente dentro do polígono referente ao setor censitário ao qual pertencem, representando visualmente a quantidade exata de habitantes da respectiva área. Em níveis de detalhe maiores, onde as áreas contêm mais *pixels* do que habitantes, há o cuidado de não se sobrepor os pontos, de modo que cada pixel possua unicamente uma das cores especificadas. Já nos níveis de detalhe menores, onde a quantidade de *pixels* da área é inferior ao número de habitantes, os *pixels* que acolhem dois ou mais pontos têm suas cores mescladas. Isso produz um resultado visual onde a mistura de cores suaviza as transições. Porém, nesses níveis, não é possível esclarecer com precisão a distribuição das raças; tem-se apenas que naquele espaço há uma mistura. Ao visualizar um município inteiro, com uma cor predominantemente púrpura, pode-se ter a impressão errada de que se trate de uma região com uma distribuição uniforme da população branca e negra por toda a área. Mas isso pode vir a ser desmentido quando se aumenta o nível de detalhe, que pode acabar mostrando uma clara segregação entre as duas populações, como mostra a Figura 2.11.

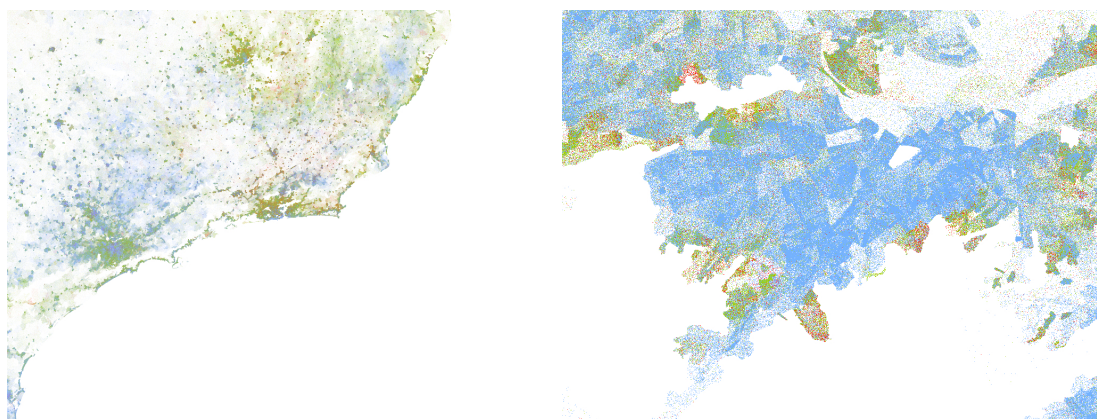


Figura 2.11: Recorte da imagem capturada do aplicativo (VIANA *et al.*, 2015), mostrando um recorte da cidade do Rio de Janeiro e um detalhe para o bairro da Tijuca, em que a mistura aparentemente homogênea apresentada no nível mais baixo não se comprova quando o mapa é mostrado em detalhe.

A utilização do aplicativo mostra que ações de enquadramento e mudança no nível de detalhe, bem como a implementação de um sistema cuja renderização seja eficaz o suficiente para que se consiga explorar o mapa com alguma fluidez, interferem diretamente na capacidade expressiva do mapa, pois a todo momento é exigido que

se façam novas e corretas interpretações sobre o que está sendo visto. A mescla das cores não é uma mentira contada pela imagem mas um ajuste funcional da visualização, que acaba interferindo também na comunicação da informação contida no mapa.

A armadilha na interpretação das distribuições da população é declarada também no projeto que lhes serviu de inspiração, citado no artigo elaborado por CABLE (2013) para o US Census Bureau, onde Mineápolis e Saint Paul são representadas em tons de roxo em um determinado nível de detalhe, demonstrando algumas áreas onde há uma falsa integração social. Mas quando vistas em detalhe, percebe-se que as populações são completamente segregadas nos bairros.

Outro fator levado em conta é a questão da confidencialidade dos dados. Os autores explicam que, no caso de um mapa onde um ponto representa um habitante, pode-se acabar colocando habitantes em campos abertos, fábricas e cemitérios. Dados adicionais, como a localização e forma dessas áreas não residenciais, obtidas em bancos de informações sobre hidrografia, florestas, ruas, podem acrescentar novos detalhes que ajudem no algoritmo de distribuição dos pontos, tornando a interpretação do mapa mais precisa (BUENO e D'ANTONA, 2014). Porém, no caso específico do aplicativo, como tais informações não estavam disponíveis para todo o país, optou-se por ignorá-las, mantendo os *census blocks*⁴ integralmente disponíveis para o posicionamento dos pontos, sem considerar a existência de matas, lagos, parques, cemitérios e outras áreas de improvável adensamento demográfico (CABLE, 2013).

Em um outro trabalho, o IBGE desenvolveu um projeto que trata da elaboração de uma grade estatística, um método de visualização cartográfica recentemente popularizado que vem sendo adotado por institutos oficiais de geografia e estatística de diversos países.

Trata-se de uma subdivisão arbitrária do espaço geográfico no formato de uma grade composta por quadrados de 200 metros de lado nos setores urbanos e 1km nas regiões rurais. Essas células funcionam como ladrilhos, rasterizando o plano e redistribuindo as informações estatísticas de uma forma mais adaptável às necessidades de análise (Figura 2.12) (IBGE, 2015).

A estabilidade formal é uma vantagem importante pois as divisões político-administrativas estão sujeitas a mudanças ao longo do tempo (IBGE, 2010). A grade é uma estrutura regular, que se mantém ao longo do tempo, e, por isso, facilita este tipo de análise.

A utilização da grade pode também melhorar a visualização de regiões que eventualmente desaparecem em um determinado nível de detalhe em um mapa que utiliza

⁴Literalmente, bloco ou quadra, equivale ao setor censitário do IBGE para o Censo Norte Americano.

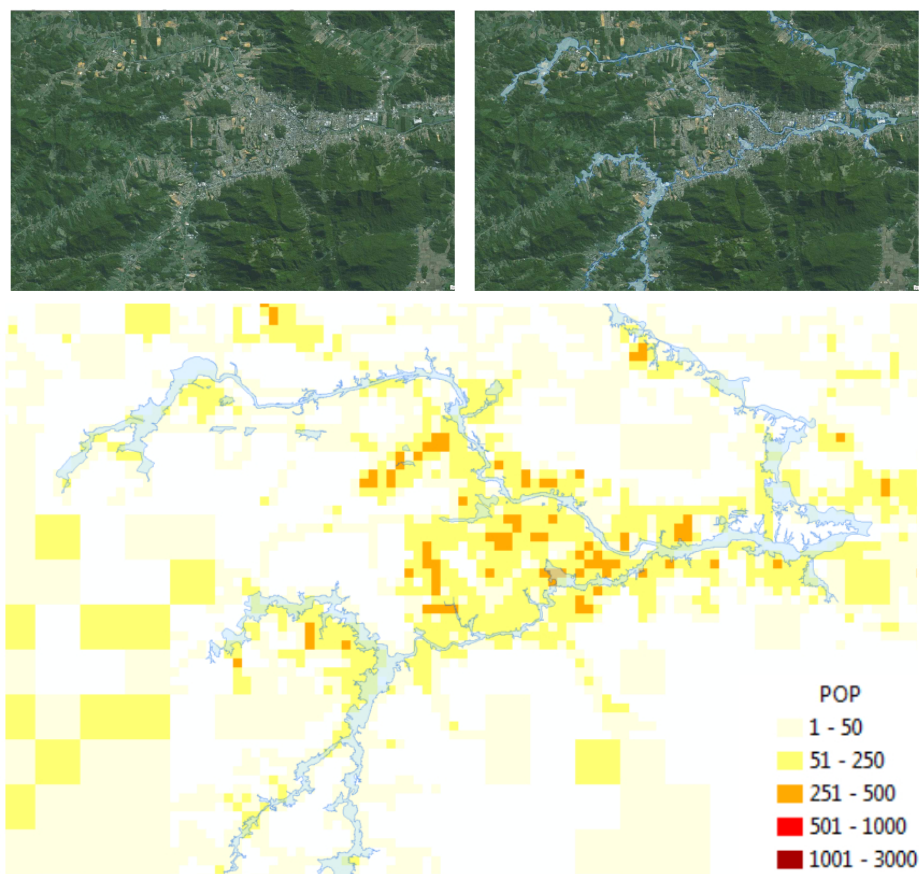


Figura 2.12: Exemplo da aplicação de uma grade de população sobre uma imagem de satélite da região de Jaraguá do Sul (RS) para avaliar o efeito de uma eventual inundação sobre a população de parte da cidade. Imagem retirada da apresentação IBGE (2015).

as divisões políticas originais. Como se pode ver na Figura 2.13, cidades importantes como São Paulo, Rio de Janeiro e Belo Horizonte, que unidas contêm aproximadamente 10% da população total do país (IBGE, a), são representadas por pequenos polígonos, praticamente invisíveis em determinados enquadramentos. A amostragem regular do espaço, promovida pelas células quadradas de uma grade estatística, poderia ajudar nesta visualização, realçando regiões importantes, antes reduzidas em pequenos pontos, mas agora relativizadas sobre uma grade arbitrária regular.

As grades estatísticas podem ser utilizadas em pesquisas para diversos fins, por exemplo quando a região a ser estudada possui um formato que não coincida com as divisões políticas disponíveis ou para análises temporais nas quais se necessite realizar comparações entre bases com registros históricos de dados. O IBGE alerta, porém, para o fato de que os dados são aproximados, ou seja, adequam-se ao objetivos de muitos estudos e análises, mas, por outro lado, podem não se ajustar a outros, pois os métodos de composição podem ser distintos e gerar algum tipo de erro.

O trabalho proposto se inspirou bastante nas possibilidades de visualização ofe-

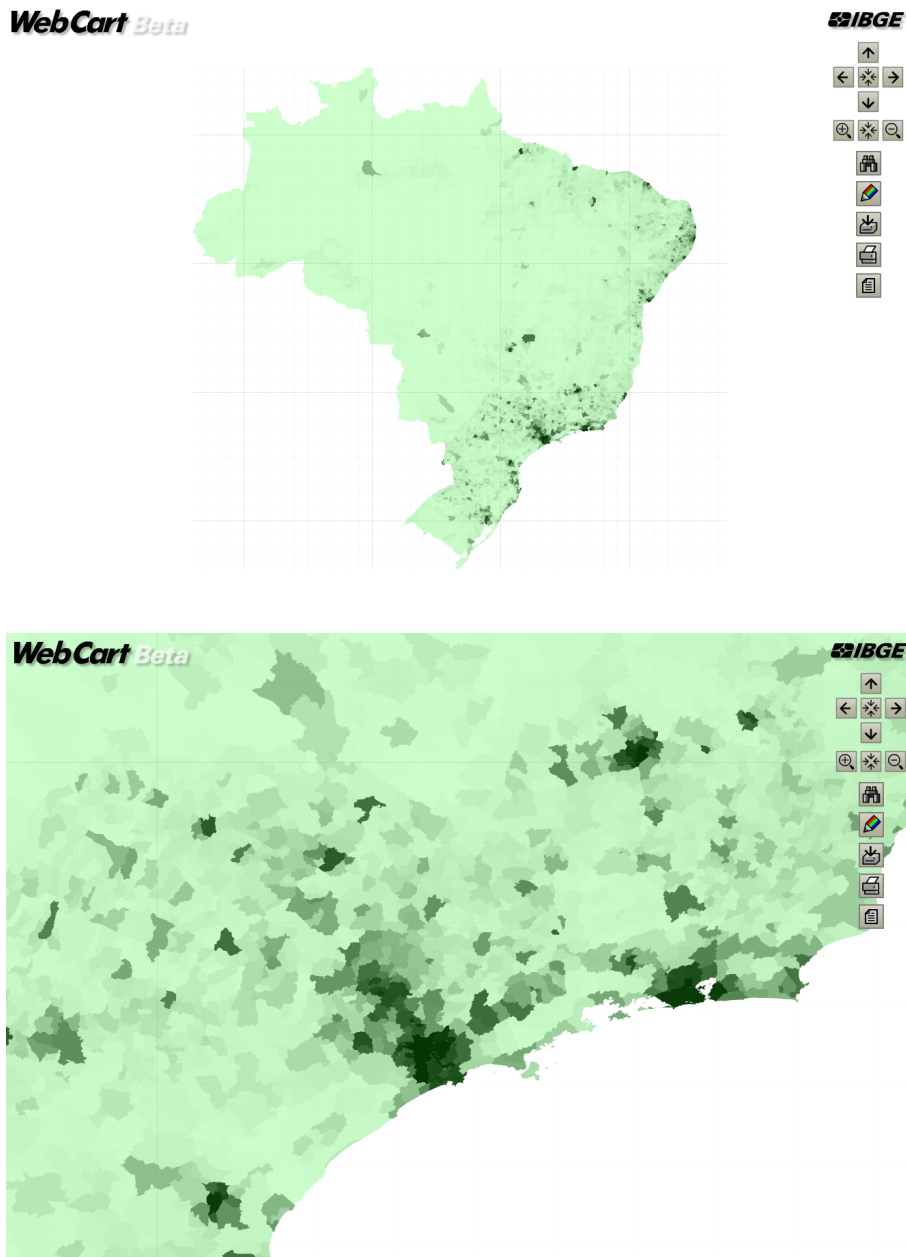


Figura 2.13: Renderizações de um mapa de densidade populacional gerado no serviço WebCart, do IBGE, mostrando a dificuldade de expressar visualmente a importância de municípios pequenos em tamanho.

recidas pela grade estatística. No caso, foi realizada uma decomposição a partir de informações estatísticas associadas às regiões, de modo que os próprios *pixels* se tornassem a grade, o que expande as possibilidades de visualização e interação com os dados. Há, porém, uma série de problemas estatísticos que isso causa, e que não serão discutidos no escopo do presente trabalho. O objetivo aqui é apenas experimentar as possibilidades oferecidas pela manipulação dos *pixels* para a renderização da imagem, e sua conseqüente manipulação e visualização.

2.3.1 Uso das Cores na Classificação dos Dados

Uma das variáveis visuais mais importantes na visualização de dados é a cor. Ela separa, ordena e traz consigo uma forte carga simbólica. Usá-la corretamente significa pensar que paleta pode valorizar a conjunto de informações representadas. Quando bem escolhidas, as cores iluminam a visualização; por outro lado, se mal aplicadas, obscurecem, confundem e atrapalham (STONE, 2006).

O uso mais importante da cor é diferenciar (TUFTE, 1994) e por isso a definição de uma boa escala de cores é fundamental para a visualização de classificações, especialmente em ambientes interativos, onde elas podem evitar consultas constantes à legenda. É importante considerar que preto, branco e cinzas também são cores, e que, além disso, há outros elementos que precisam ser coloridos, como texto, fundo, linhas e molduras.

O sistema mais adequado para compor paletas cromáticas é o HSB⁵, que permite a composição de cores através de três componentes: matiz, saturação e brilho (Figura 2.14) (STONE, 2006).

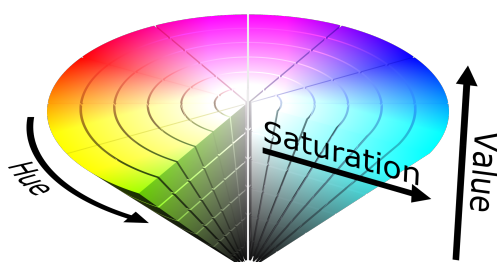


Figura 2.14: Uma versão do gráfico HSV, em formato de cone. Imagem retirada de Wikimedia Commons (2010).

O matiz é a nuance cromática relativa ao comprimento de onda da radiação espectral determinando o que pode ser chamado de cor pura. A saturação é a variação do matiz desde o neutro (cinza) até seu tom mais vibrante. E o valor representa a quantidade de energia refletida, indo de valores desde o branco até o preto, passando pelas nuances relativas àquele matiz.

Os matizes podem ser ordenados conforme sua frequência, formando uma sequência visível, organizada em um gráfico radial, denominado círculo cromático, ilustrado pela Figura 2.15 (RODRIGUEZ, 2010).

No círculo, as cores primárias do sistema aditivo - vermelho, verde e azul - ocupam setores angulares com espaçamento de 120° entre si. As cores secundárias -

⁵*Hue, Saturation and Brightness*, também chamado de HSL (*Hue, Saturation and Lightness*) ou HSV (*Hue, Saturation and Value*).



Figura 2.15: Círculo cromático. Imagem retirada de Wikimedia Commons (TRYGSTAD, 2007).

ciano, magenta e amarelo - formam uma estrutura intercalada. Assim, cores complementares ficam organizadas em lados opostos no círculo e determinam os maiores contrastes entre os matizes. Da mesma forma que as cores adjacentes são similares e menos contrastantes (STONE, 2006).

A legibilidade de uma imagem depende da sua capacidade de separar figura e fundo. Para isso, a luminância (brilho) é a componente visual mais importante. É muito mais fácil distinguir escalas de brilho do que uma escala de matizes com o mesmo nível de luminância (STONE, 2006).

A saturação determina o quão vívida será a cor. Quando mínima, resulta em um cinza com a mesma luminância do matiz; quando máxima, obtém-se a versão mais pura daquele matiz. Cores escuras (próximas ao preto) e os tons pastéis (próximos ao branco) derivam dessas combinações e conseguem produzir paletas sofisticadas. Cores com intensidade máxima se separam e saltam à vista, enquanto tons neutros se agrupam.

As variações na luminância criam a percepção de camadas, com os maiores contrastes saltando aos olhos e os menores se confundindo com o fundo. Contrastes entre matizes com saturações similares não deixam clara a percepção de profundidade (STONE, 2006).

A regra contraste/similaridade deve ser aplicada às três características (HSV), possibilitando a elaboração de paletas que permitam os destaques e agrupamentos necessários para uma boa composição visual. O importante é selecionar os matizes de acordo com a necessidade, usando os contrastes para destacar e as semelhanças para agrupar, variando a luminância para assegurar a legibilidade (STONE, 2006).

Na maioria dos casos, uma boa peça utiliza uma paleta reduzida, contendo não mais que duas ou três cores, valendo-se de uma variação bem planejada de saturações e luminâncias para obter riqueza visual e organização. Minimizar a dependência dos

matizes evita a confusão visual e facilita a hierarquização da composição (STONE, 2006).

Há muitos livros e softwares capazes de ajudar na composição de uma boa paleta. Um exemplo é o sítio ColorBrewer (BREWER e HARROWER, 2016), que oferece auxílio na elaboração de paletas para aplicação em mapas temáticos. Outra referência é o trabalho do professor Israel Pedrosa (PEDROSA, 2003, 2009), de leitura obrigatória para quem pretende aprender sobre a aplicação das cores nas artes e, conseqüentemente, na visualização de informações.

Compor escalas de cor significa pensar sobre a informação que se quer transmitir, pois apesar de se destacarem como elementos rotuladores, as cores também agrupam e ordenam, ou seja, são importantes para organizar visualmente um conjunto de dados. BREWER (1994) apresenta um estudo que relaciona diversos esquemas de cor apropriados para situações em que a organização da informação vai um pouco além da simples separação dos elementos.

Os esquemas qualitativos (Figura 2.16), nos quais o objetivo é destacar a simples separação das classes, sem qualquer eventual conotação de ordenamentos ou sub-grupos, exige apenas uma composição baseada na variação dos matizes, calibrados para terem luminâncias equivalentes (BREWER, 1994).



Figura 2.16: Escala de matizes.

Os esquemas sequenciais (Figura 2.17), nos quais a ordenação das classes é o objetivo, devem ser representados sempre por variações de luminância. De preferência associando o crescimento dos valores com o escurecimento do tom. O inverso também pode funcionar, mas não é usual nem intuitivo, a não ser que o fundo seja escuro (BREWER, 1994).



Figura 2.17: Escala baseada no matiz amarelo.

A escala em preto e branco é a que melhor representa uma grande quantidade de classes (Figura 2.18), pois seu contraste é melhor percebido do que o de qualquer matiz. Nesse caso, é preciso prestar atenção aos outros elementos da imagem, como linhas divisórias e as áreas sem informação, que se pintadas de preto ou branco podem acabar confundidas com algum valor da escala (BREWER, 1994).

Os matizes não possuem luminosidade máxima, logo não conseguem produzir uma escala legível para muitas classes. Uma alternativa é usar saturação máxima como ponto médio e dividir a escala em duas direções, indo para tons mais claros



Figura 2.18: Escala de cinzas.

e escuros deste matiz. A cor mais saturada, no meio da escala, se torna um ponto focal, o que funciona para dar um destaque para esta categoria. Isso é adequado nos casos em que haja algum tipo de marco para os dados, como a média dos valores, por exemplo (BREWER, 1994).

É possível criar escalas entre dois matizes. Porém, essa sequência também deve estar acompanhada de uma variação de luminância, garantindo a percepção de ordenação. É importante observar que os matizes não se comportam da mesma maneira com relação à saturação e à luminosidade; portanto, podem produzir escalas visualmente distintas, porém que causem confusão (BREWER, 1994).

Uma dupla de matizes pode ser aplicada em esquemas divergentes (Figura 2.19), onde, a partir de um ponto de interesse, neutro, a escala seja dividida em direções opostas. O valor médio de uma variável ou o zero de uma função podem ser pontos de referência importantes em escalas de dados e, por isso, o uso de um esquema como esse pode ser interessante. Vale observar que a variação de luminosidade é sempre importante, fazendo com que a classe central seja destacada. É fundamental, também, manter a coerência. Para se evitar o risco de denotar desproporcionalidade, classes que possuem a mesma distância do ponto divergente devem ter luminosidades equivalentes visualmente (BREWER, 1994).



Figura 2.19: Escala divergente.

No estudo são apresentadas também escalas combinadas para produzirem a representação de duas classificações de dados simultaneamente. Porém, o uso das cores para produzir escalas desse tipo pode se tornar complexo, especialmente considerando sua visualização a partir de dispositivos diversos, cuja a qualidade de exibição não garanta a legibilidade. Outro problema importante que precisa ser levado em conta é a capacidade de percepção cromática dos usuários. Muitas pessoas possuem limitações e escalas combinadas em duas ou até três dimensões são matematicamente possíveis, mas nem sempre viáveis. Sendo assim, as escalas mais complexas nem sempre são funcionais e serão ignoradas neste trabalho. Uma alternativa viável é a utilização de um novo sinal gráfico, como hachuras, texturas ou outro identificador qualquer que consiga representar melhor tais diferenças (BREWER, 1994).

Capítulo 3

VisGL, uma ferramenta *on-line* para visualização de dados multivariados georreferenciados

Após a pesquisa por referências que tratassem do assunto, verificou-se que os métodos tradicionais apostam na simultaneidade das representações em um único glifo para a solução do problema. Porém, os resultados, quase que invariavelmente, são questionáveis, no sentido de que a percepção humana, apesar de altamente avançada na assimilação generalista de conjuntos - como intuir emoções a partir de uma face - não é capaz de perceber naturalmente, sem um grande esforço consciente, os detalhes em separado.

Além disso, os métodos apresentam problemas na percepção da linearidade das representações, o que é uma questão importante, discutida em todos os artigos lidos sobre o assunto. A visualização de um detalhe não garante a interpretação intuitiva de que parte do todo ele representa. Trata-se de um problema recorrente também nas implementações dos métodos tradicionais da cartografia, para quem a simples representação de um único tema em um mapa é um desafio.

BERTIN (1980) é categórico ao afirmar que a única variável capaz de representar proporções é o tamanho, e sua aplicação em um mapa não é uma tarefa simples. Exige, no mínimo, espaço para a leitura correta da variável visual ou, então, uma saída diferente, como a visualização do mapa em terceira dimensão, conforme visto no aplicativo WebGL Globe, desenvolvido pela Google.

3.1 Mapa *on-line* interativo

Sendo assim, neste estudo, em vez de se compor um novo glifo, propôs-se uma solução que abordasse a questão a partir do viés da interatividade, pressupondo

que a aplicação da Computação Gráfica poderia potencializar métodos tradicionalmente comprovados pela Semiologia Gráfica para a elaboração de mapas capazes de realizar uma melhor imersão no volume de dados e, assim, promover a desejada compreensão da multidimensionalidade de um conjunto de dados georreferenciados. Essas técnicas, aplicadas em um ambiente digital, *on-line* e interativo, conforme os princípios discutidos por SHNEIDERMAN (1994), possibilitam a criação de um ambiente mais propenso à exploração através de ações simples, como ligar e desligar uma camada, carregar rapidamente um novo tema ou manejar as classes de dados visualizando resultados instantaneamente animados na tela.

Para isso foi desenvolvido um mapa *on-line* interativo do Brasil, em nível municipal, no qual o usuário consegue ver simultaneamente dois índices, calculados a partir de um grupo de variáveis escalares selecionadas, relacionadas com os municípios que compõem o mapa.

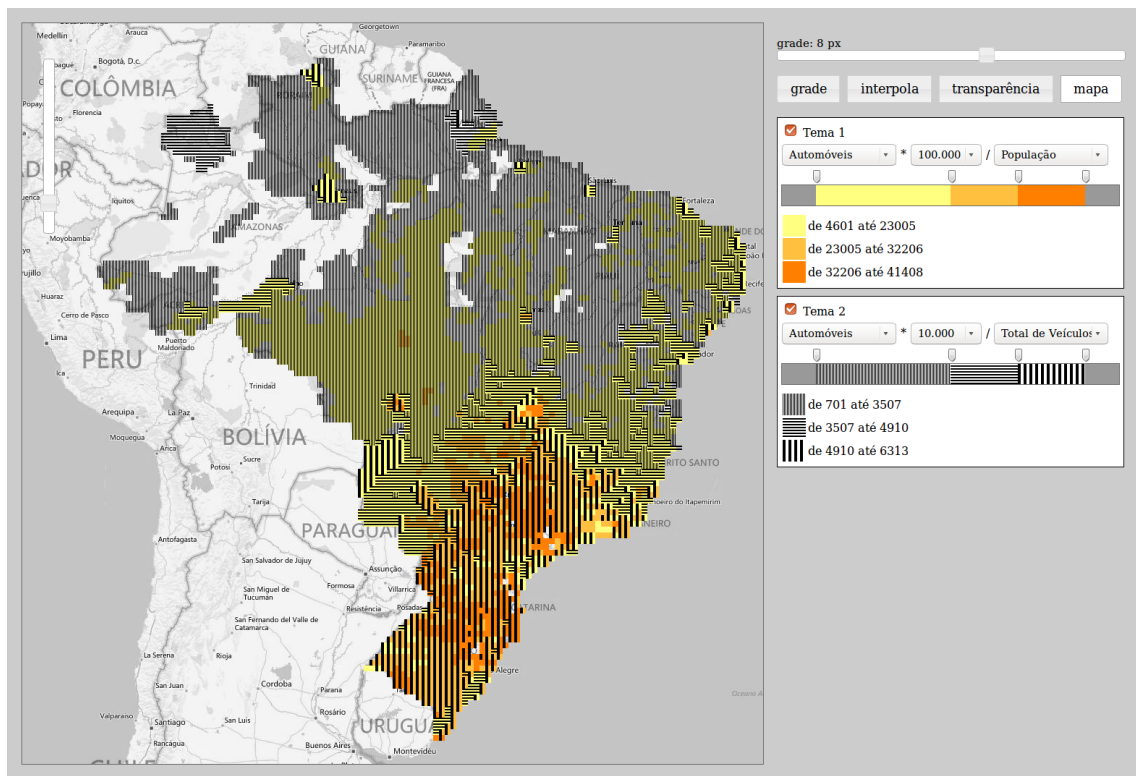


Figura 3.1: Visão geral do aplicativo.

A implementação de uma ferramenta *on-line* foi motivada pela demanda latente por novas soluções de visualização para a Internet, que no Brasil tornou-se um meio popular e democrático para disseminação de informações públicas. Isso se deve, especialmente, à popularização do HTML5, que possibilitou, entre outras coisas, a evolução da linguagem Javascript e ao advento do WebGL, que permite o desenvolvimento de aplicações gráficas avançadas, voltadas para a utilização direta no navegador, produzindo resultados gráficos expressivos antes só viáveis por aplicações

dedicadas.

As informações utilizadas nesse experimento foram obtidas no Portal do IBGE na Internet. As malhas usadas para desenhar os municípios são referentes à Malha Municipal 2013 e estão disponíveis para download no portal. Já as informações socioeconômicas foram obtidas no mesmo portal, mas no canal Cidades e são também referentes ao ano de 2013.

Para calcular o índice, o usuário escolhe as variáveis dispostas em caixas de seleção posicionadas em um menu ao lado direito do mapa, que permitem ao usuário compor uma fórmula simples. São três combos, um para a variável divisora, outro para a variável denominadora e um terceiro para selecionar um fator que multiplica essa divisão. Por razões técnicas, específicas para a implementação desse experimento, os resultados são arredondados. Esse índice é processado e classificado conforme uma escala de valores equidistantes, entre o mínimo e o máximo daquele tema, naquele enquadramento instantâneo.

3.1.1 Variáveis relativizadas

As variáveis socioeconômicas compreendem valores escalares, e são relativizados a partir de ações do usuário, que realiza a composição de uma densidade ou taxa. Em tese, não há diferença teórica na visualização de valores relativos ou brutos. Porém, a prática da representação de informações em mapas coropléticos mostra que há detalhes que precisam ser considerados.

A representação linear de valores brutos, como a população ou a quantidade de automóveis, possui detalhes específicos. As informações socioeconômicas são muitas vezes associadas a outras grandezas, como população ou área de uma subdivisão geográfica. Por exemplo, em uma cidade com população na ordem de milhões de habitantes, certamente haverá muitos milhares de automóveis e também alguns milhares de escolas. Esses valores, quando comparados aos dados relativos à cidades com menor população, ainda serão vultosos e ocuparão as classes maiores da escala. Muitas vezes, porém, a informação que se precisa obter não é a expressão da contagem bruta dos dados, mas a relação destes com outras variáveis. Pode-se usar como exemplo a relação entre automóveis e habitantes ou entre a população em idade escolar e a quantidade de escolas. Essas taxas podem ser capazes de revelar novas percepções sobre os dados, nem sempre capazes de serem expressas através de mapas referentes aos seus valores brutos em separado.

de crianças em idade escolar não revelem uma densidade com a mesma magnitude, mas que muitas vezes é a informação que se deseja obter.

Esse problema remete a uma questão discutida na literatura, que diz respeito à pergunta que se espera que o mapa responda. Em outras palavras, essa pergunta

precisa ser adequada, senão o mapa não conseguirá cumprir seu papel. Do mesmo modo, um mapa que não transcreve as relações que o usuário busca não é necessariamente uma peça ruim; apenas não foi pensado para aquele propósito. BERTIN (1980) afirma que o autor do mapa e seu usuário estão do mesmo lado no processo de leitura dos dados, interpretando as relações de separação, ordenamento e quantificação. Portanto, ao elaborar um mapa, seu autor precisa estar ciente das questões às quais aquele mapa é capaz de responder, pois terá ele mesmo que fazê-las na hora de selecionar e transcrever os dados.

Isso explica porque os mapas coropléticos são considerados ferramentas mais adequadas à visualização de variáveis relativizadas, como taxas e proporções, do que das variáveis brutas que lhes dão origem. Um exemplo claro disso é visto quando se analisa a relação entre as populações das cidades brasileiras. A disparidade dos valores é grande. São Paulo, município mais populoso do Brasil, possui uma área com pouco mais de 1500 km² e uma população estimada de quase 12 milhões de habitantes. Já Altamira (PA), município com maior área territorial (com cerca de 160 mil km² de extensão) abriga pouco menos de 100 mil habitantes. Em suma, o maior município do país em extensão territorial é cerca de 100 vezes maior que o município mais populoso do país, que por sua vez, possui população 120 vezes maior. Diferenças como essas precisam ser levadas em conta na elaboração de uma visualização. Dispostos em um mesmo mapa, São Paulo, com toda sua importância socioeconômica, se torna um pequeno ponto, enquanto municípios com áreas imensas, ganham destaque visual. De modo semelhante, a população de São Paulo é tão maior, que acaba interferindo fortemente na visualização dos outros municípios, que desaparecem na imagem, agrupados na classe mais rasa de uma escala equidistante dentre os dados. A relativização destas variáveis se torna, portanto, quase uma obrigação quando visualizadas em um mapa coroplético.

Para a visualização das variáveis brutas originais, como a população ou a quantidade de carteiras escolares, um método adequado para a composição das classes é o da distribuição regular dos dados por uma escala obtida com o cálculo de um histograma. O efeito prático disso é a organização das áreas em grupos equivalentes de municípios, o que produz uma imagem mais equilibrada, mesmo levando em conta a disparidade dimensional das cidades.

Como o cálculo dos mínimos e máximos depende do enquadramento instantâneo do mapa, as ações de movimentação e mudança no nível de detalhe se tornam essenciais para a visualização dos resultados. O usuário pode mover o mapa, arrastando-o com o cursor, e mudar seu nível de detalhe, acionando um pequeno controle vertical posicionado no canto superior esquerdo do mapa.

3.1.2 Equalização da imagem conforme o enquadramento

Neste trabalho, a decisão de realizar uma equalização automática do mapa conforme seu enquadramento foi tomada porque, quando detalhadas na tela, muitas das regiões apresentam dados que, quando comparados aos mínimos e máximos globais, podem acabar enquadrados em apenas uma classe, gerando regiões inteiras da mesma cor ou textura. A equalização instantânea a partir dos máximos e mínimos locais impede que isso aconteça, tornando o mapa mais dinâmico e propício a análises detalhadas. Ademais, fazendo uso dos controles laterais, o usuário pode manipular a imagem e obter as relações que desejar.

Optou-se por fazer um mapa com apenas duas variáveis por vez. Isso se deu após a realização de alguns experimentos preliminares, nos quais foi testada a visualização de três temas (variáveis) simultâneos codificados usando escalas de cores, baseadas em vermelhos, verdes e azuis. A idéia foi descartada depois de constatar-se que o mapa tornou-se de difícil leitura, mesmo usando recursos auxiliares tais como a manipulação da imagem ou o desligamento de um dos três canais de cor.

Além do uso das escalas de cor, foi experimentada também a substituição de uma dessas escalas cromáticas pela representação composta por uma textura que variava de tamanho conforme a progressão da escala. A textura usada nessa versão foi obtida a partir de uma imagem rasterizada que simulava um trançado. A imagem era fotorrealista, em preto, branco e cinzas, sendo que o valor da cor (luminância) determinava sua transparência, isto é, o branco se tornava totalmente transparente, o preto, totalmente opaco, e os cinzas translúcidos. A imagem acabou ficando complexa demais, causando confusão, como mostrado na Figura 3.3.

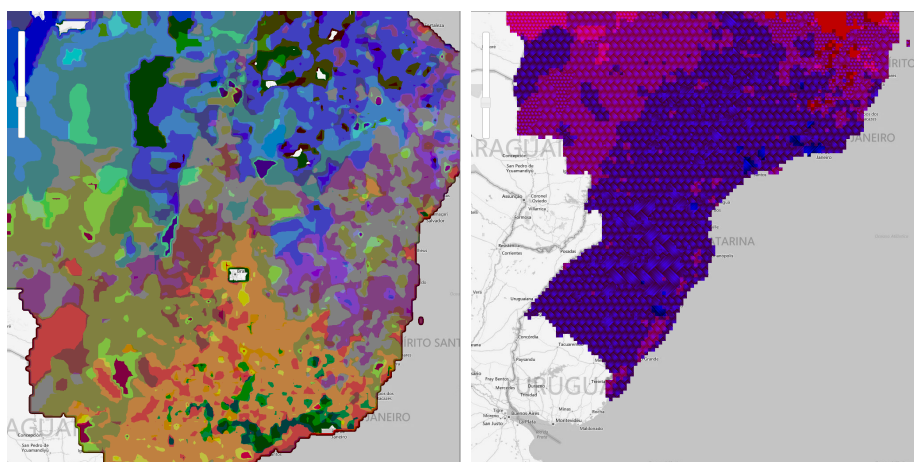


Figura 3.2: Experimentos preliminares, em que foram testadas escalas baseadas em cores e a utilização de uma textura baseada em uma imagem fotorrealista.

Esses experimentos preliminares serviram para comprovar a eficácia da técnica usada para compor o mapa, renderizando as imagens em tempo satisfatório, mesmo em máquinas com processamento limitado. Isso foi suficiente para que se apostasse

em uma troca rápida, quase instantânea, dos temas, carregados a partir de pequenas imagens contendo os dados. Por isso, decidiu-se visualizar simultaneamente apenas dois temas por vez, um representado por uma escala de cores, e o outro associado a um conjunto de texturas simples que varia em tamanho e rotação, trocando os temas quando necessário.

3.1.3 Visualização por grade estatística

O presente trabalho se inspira na visualização por grade estatística, desenvolvida pelo IBGE e disponibilizada recentemente para o público em geral (IBGE, 2015). Trata-se de um sistema inovador que utiliza uma grade regular quadrada para apresentar as informações socioeconômicas georreferenciadas.

Partindo do princípio de que os *pixels* também são uma subdivisão regular de uma representação espacial, aplicou-se o mesmo raciocínio da grade estatística para promover uma divisão dos dados pelos *pixels* da imagem. Dessa maneira, cada um deles passa a estar associado a uma fração dos dados relativos à subdivisão geográfica que lhe diz respeito.

É preciso ressaltar, porém, que para a realização desse experimento foram desconsiderados muitos dos problemas discutidos naquele projeto, especialmente com relação às questões sobre a decomposição espacial dos dados, discutida por BUENO e D'ANTONA (2014), bem como dos problemas oriundos da distorção provocada pela projeção Webmercator, utilizada aqui. Essa projeção teve sua escolha motivada tanto pelas questões de implementação quanto pelo fato dela ser o padrão dos serviços mais populares de *webmapping*, como o Google Maps, o Bing Maps e OpenLayers. Apesar de sua popularidade, essa projeção distorce significativamente o mapa, promovendo a criação de *pixels* que representam porções variantes do espaço conforme a latitude. O aplicativo realiza a divisão igualitária dos dados pela quantidade de *pixels* referente a sua área respectiva, o que gera um erro. Neste trabalho, porém, o problema foi abstraído em prol da implementação prática da idéia. Apesar dos resultados serem promissores, a ferramenta não pretende propor algo que contrarie o bom uso das informações estatísticas, cabendo a ela apenas o papel de experimentar novas técnicas de visualização e, especialmente, de utilizar na prática alguns conceitos de computação gráfica aplicados na composição de mapas para a visualização de dados. O problema há pouco mencionado demanda reflexão e novas pesquisas acerca de sua adaptação, considerando a questão da distorção, ou mesmo da criação de uma projeção que faça com que a grade regular coincida com a grade estatística proposta.

A despeito do problema, o sistema realiza a decomposição dos dados em uma grade formada pelos *pixels* da tela, retalhando as áreas originais, juntamente com

seus dados, por essa amostragem arbitrária do espaço. A grade ajuda na solução de algumas questões de visualização, granularizando o espaço e dando mais liberdade para o desenho de regiões e representações. Entretanto, a relativização dos dados pelos *pixels* promove alguma complexidade, pois a taxa se refere na verdade, à densidade dos dados em cada um desses pequenos fragmentos. Por isso, mesmo a visualização de informações brutas associadas às áreas, no fim das contas são relativizadas pelos *pixels*, o que promove uma densidade não muito fácil de compreender a princípio.

A visualização dessa grade é feita de dois modos. Um deles oferece uma visão pura e simples das células quadradas, coloridas ou preenchidas pelos padrões das classes que lhe dizem respeito. O outro modo é a visualização da região interpolada, ou seja, o desenho dos limites estipulados pelos controles, aplicados diretamente sobre os valores dos *pixels*, obtidos pela interpolação linear baseada nos valores das grades vizinhas.

Interpolação dos valores da grade

A interpolação é feita da seguinte forma: o ponto central da célula determina o seu valor, assim como os pontos centrais das células vizinhas. O valor referente a cada *pixel* da grade é, portanto, obtido pela interpolação linear desses quatro pontos. O limite determinado pela classe opera, portanto, *pixel* a *pixel* na imagem, determinando fronteiras mais orgânicas a partir dessa suavização.

A opção pela visualização interpolada da grade é acionada em um controle no menu e o resultado da ação é instantâneo, permitindo que o usuário ligue e desligue o mecanismo quando achar necessário. A granularidade da grade também é definida por um controle no menu. Ele é do tipo *slider* e permite que os usuários determinem o tamanho das células. A princípio, seria possível criar células de formas retangulares e tamanhos variados. Porém, para efeito de otimização computacional, foi definido que elas serão quadradas de tamanhos derivados das potências de 2, variando progressivamente entre 1×1 (nível 0) e 32×32 *pixels* (nível 5).

3.1.4 Escalas de classes editáveis

A edição das escalas que definem as classes também é realizada interativamente, através de controles semelhantes. Esses controles são acionados pelo arrasto do cursor e permitem a modificação dos valores, redefinindo os limites para as classes e promovendo uma renderização imediata da imagem. A atualização instantânea do mapa é essencial para o funcionamento do sistema, pois produz uma animação que confere uma exploração mais intuitiva do volume de dados, permitindo, assim, que o usuário averigüe os padrões de ocupação espacial revelados pelo tema, não

apenas a partir da análise visual de uma imagem estática, mas pela compreensão da movimentação das fronteiras. Dessa maneira, é possível identificar vales, picos, concentrações e outras regiões de interesse no mapa sem a necessidade de muitas classes, por exemplo.

O sistema permite duas operações para a classificação dos valores relativos a cada um dos temas. O controle deslizante permite que o usuário defina um recorte nos dados, estabelecendo os valores máximos e mínimos que deseja representar no mapa, e defina uma escala entre os valores intermediários, fazendo com o que o mapa se comporte conforme essa classificação.

A opção por utilizar três classes visíveis para organizar as informações sobre o mapa se deu porque o uso de muitas classes não proporciona uma visualização de fácil identificação. Considerando a visualização de dois temas simultaneamente, cada qual dividido em três classes visuais, o resultado são nove possibilidades de composição dos sinais escolhidos para representá-los, o que ultrapassa o limite de sete elementos simultâneos que a cognição humana consegue lidar (BERTIN, 1980).

Além disso, a possibilidade da interação animada dos valores permite identificar na imagem as distribuições das informações sem que haja necessidade do estabelecimento de muitos níveis.

Sendo assim, com o objetivo de simplificar o resultado visual para ampliar a experimentação, a utilização de três classes visíveis e editáveis se mostrou suficiente para gerar recortes minimamente organizados visualmente, sem causar confusão.

Representação visual dos temas

Os temas, devidamente organizados em classes, precisam ser representados por variáveis visuais adequadas que transcrevam corretamente a percepção de separação e ordenamento. Para isso, foram selecionadas duas escalas. Um tema é representado por uma escala de cores, composta por uma progressão simples de matizes, acompanhada de uma variação singela de luminância, que inicia com um amarelo claro e termina com um vermelho um pouco mais escuro, passando por uma laranja médio na classe central. A escolha dos matizes foi determinada pelo fato de que, de acordo com BREWER (1994), essa é, dentre as escalas formadas por matizes vizinhas no círculo cromático, a que produz melhor distinção entre os tons. A variação de luminância, por sua vez, é pequena e foi definida visto que essa camada será visualizada sob uma outra, identificada por uma textura que, mesmo simples, acabaria promovendo uma sensação de escurecimento do tons e poderia causar confusão. Como se trata de apenas três classes, o uso de matizes como identificadores da ordenação da escala - mesmo que não seja a escolha ideal, uma vez que a variação de luminância é mais indicada - para este estudo, foi a melhor opção.

Para representar o segundo tema, foi utilizada uma escala de texturas simples,

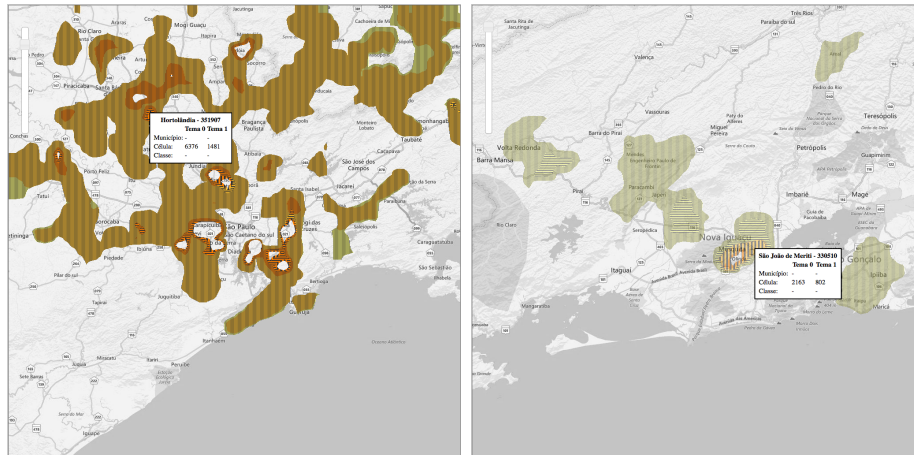


Figura 3.3: A interpolação dos valores da grade gera limites mais orgânicos para as regiões e o acionamento da transparência e das caixas de texto ajudam a compreender o que acontece naquele ponto.

compostas por linhas contínuas pretas, regularmente espaçadas por vazios transparentes de mesma espessura. A textura varia em tamanho e rotação. O tamanho promove a impressão de ordenamento e a rotação reforça a diferenciação entre as classes. Chegou-se a esse formato após algumas experimentações, em especial com a utilização de uma outra textura, mais complexa, oriunda de imagens naturais rasterizadas, mas que não funcionou muito bem.

O sistema permite ligar e desligar as camadas, permitindo que o usuário visualize o comportamento de cada tema separadamente para analisá-lo com mais precisão, conforme recomendado por QUEIROZ (2011). É possível ligar e desligar também a imagem de fundo. Para facilitar a visualização da referência espacial, é possível acionar a transparência dos temas, permitindo visualizar a imagem de fundo sem perder totalmente a noção do que acontece nas camadas acima. O usuário também pode ligar ou desligar a visualização da grade em um botão do menu.

A legenda é uma parte essencial de um mapa e, de acordo com ARCHELA e THÉRY (2008), é na construção da legenda que acontece a transcrição da linguagem escrita para a gráfica e há uma delas para cada tema.

Ao parar o cursor sobre o mapa, aparece uma pequena caixa de diálogo que mostra as informações básicas daquele ponto, como o nome do município ao qual pertence e os valores das variáveis, para a cidade, para a célula e para as classes.

Uma vez apresentada a proposta e discutidos brevemente a motivação e os problemas referentes à idéia do projeto, o próximo capítulo passa a detalhar a implementação do sistema.

Capítulo 4

Implementação

4.1 Descrição geral

O sistema implementado apresenta ao usuário a interface ilustrada na Figura 3.1. Nesta seção os elementos dessa interface são descritos em maior detalhe, em particular os controles mostrados no painel à direita, como mostra a Figura 4.1.

O cartograma aqui apresentado é um mapa interativo de municípios, renderizado a partir da Malha Municipal 2013 (IBGE, b), disponibilizada pelo IBGE em seu portal na Internet e que contém a divisão política municipal mais recente.

1. Mapa: Área para renderização do mapa. É possível movimentá-lo clicando e arrastando o cursor sobre a imagem.

2. Definição do nível de detalhe: Um pequeno controle, posicionado no canto superior esquerdo do mapa, permite a mudança do nível de detalhe, o que, neste sistema, trata da mudança de uma variável global que controla a carga das imagens que formarão o mapa. Isso será melhor explicado em breve.

3. Caixa de diálogo: Ao se manter o cursor parado mais de três segundos sobre o mapa, aparece uma pequena caixa de diálogo contendo os valores das taxas referentes à célula e ao município em questão e as classes referentes a cada um dos dois temas.

4. Controlar o tamanho da grade: Controle do tipo barra deslizante, que determina a granularidade da grade, ou seja, define o tamanho de suas células. Essas células são quadradas, com tamanhos que variam conforme potências de dois entre os expoentes 0 e 5, resultando em uma variação de dimensões entre (1×1) e (32×32) *pixels*.

5. Opções de visualização: Quatro botões do tipo "liga/desliga" determinam modos de visualização da imagem: a visualização da grade, onde os limites das células são pintados de azul; a transparência dos temas, diminuindo a opacidade dessas camadas para mostrar a imagem de fundo com o objetivo de facilitar a localização;

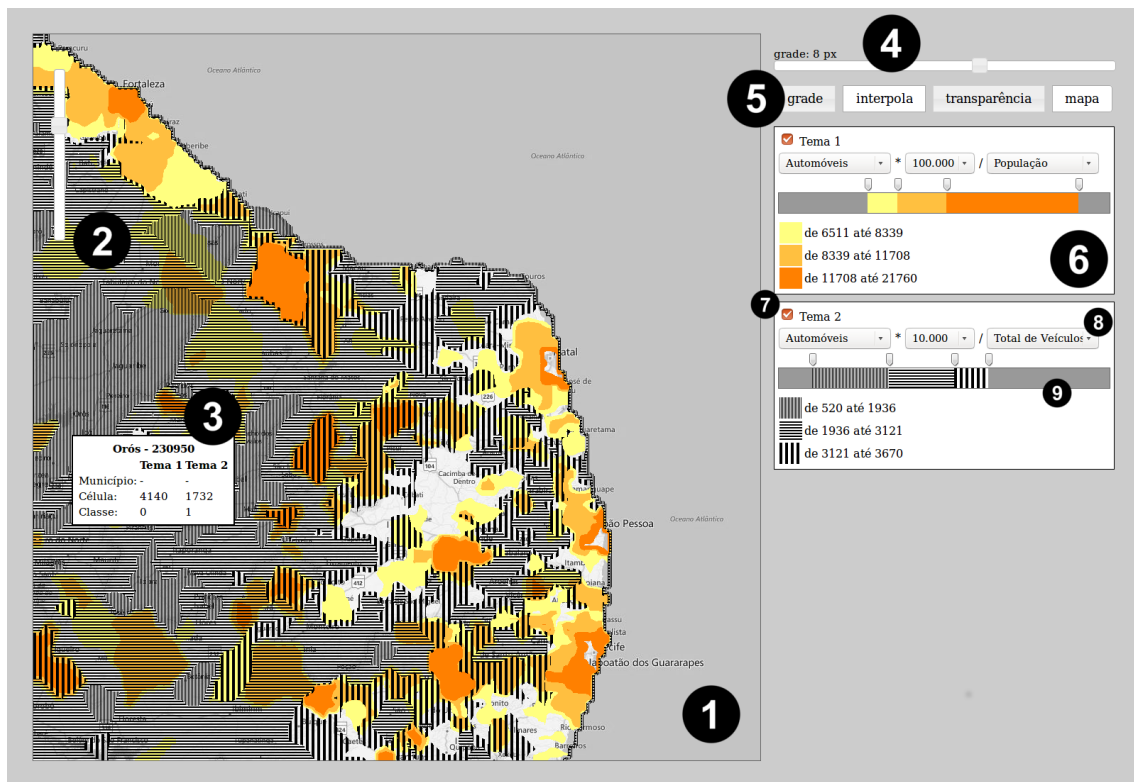


Figura 4.1: Interface do sistema: 1 - Mapa; 2 - Variação do nível de detalhe (escala); 3 - Caixa de texto para identificação dos locais; 4 - Tamanho da grade; 5 - Visualização dos limites da grade, interpolação, transparência e mapa no fundo; 6 - Controles e legenda do tema; 7 - Liga/desliga o tema; 8 - Definição da taxa representada no temas; 9 - Classificação do tema;

a interpolação dos valores da grade, o que promove a visualização de limites definidos por um recorte mais orgânico das classes; e a apresentação, ou não, do mapa topográfico de fundo.

6. Definição dos temas: Neste sistema é realizada a visualização de dois temas. A legenda não permite interação, mas é reativa às ações do usuário, atualizando em tempo real as representações definidas para a imagem do mapa.

7. Ligar ou desligar os temas: Botão que determina a visualização ou não de cada um dos temas.

8. Definição da taxa: Cada um dos temas consiste de uma expressão definida pelo usuário, que seleciona duas dessas variáveis e compõe uma fórmula, que resulta em uma taxa, que pode ser definida como: sendo T a taxa referente a um dos temas, V_1 e V_2 as variáveis selecionadas, e F o fator selecionado, o valor da taxa é definido como $T \leftarrow \frac{V_1 \times F}{V_2}$.

As variáveis disponíveis no sistema para realizar essa tarefa foram obtidas no portal do IBGE na Internet, no canal Cidades@, que disponibiliza um banco de dados com informações oficiais sobre os municípios brasileiros contendo mais de 10 mil variáveis de naturezas diversas. Em particular, foram selecionadas 140 dessas

variáveis que consistem em valores numéricos inteiros obtidos em contagens de indicadores como população, quantidade de veículos e número de matrículas em escolas públicas.

9. Classificação dos temas: A classificação de cada tema é definida pelo usuário através de dois controles do tipo barra deslizante capazes de organizar cinco divisões entre os valores máximo e mínimo de cada tema, naquele enquadramento do mapa, o que significa que os valores analisados dizem respeito especificamente às regiões representadas naquela instância da imagem.

Esses controles determinam dois estados para a apresentação do tema. As classes que compreendem os dois extremos da escala, definem o recorte do que não estará visível, ou seja, tornam transparentes os lugares cujos valores estejam compreendidos dentro dos limites dessas duas classes. As três classes intermediárias realizam a organização do que está visível, através de escalas visuais de cor e textura para os temas 1 e 2, respectivamente.

O primeiro controle, relativo ao tema 1, indica a representação cromática, variando a escala entre o amarelo-claro e o alaranjado-escuro, passando por um laranja-médio na classe do meio. O segundo controle, referente ao tema 2, indica a representação através de uma escala composta por hachuras, que variam de tamanho e orientação.

4.2 Detalhes da implementação

Uma vez definida a proposta, é preciso explicar alguns detalhes da sua implementação. Afinal, por mais bem elaborada que seja a ideia, há decisões importantes que dependem da programação. O objetivo é construir um sistema que consiga colocar em prática o dinamismo gráfico necessário para uma boa interatividade.

Considerando que se trata de uma ferramenta *on-line*, o sistema é implementado em HTML5, usando o Javascript como linguagem para a programação da interface e o WebGL, através da programação dos *shaders*, para explorar o potencial da placa gráfica com o objetivo de obter renderizações velozes dentro dos navegadores padrão, sem a necessidade de *plugins* ou componentes não padrão em navegadores modernos.

Essa velocidade no processamento gráfico é oriunda de uma arquitetura construtiva determinada para o processamento paralelo de vértices e *pixels* na placa de vídeo, que organiza um fluxo rígido de operações que só permite algum tipo de intromissão através de programas específicos, denominados *shaders*. No caso do WebGL, há dois tipos de *shaders*: os *shaders* de vértice, que realizam o processamento independente de cada um dos vértices da geometria do objeto; e os *shaders* de fragmento, que realizam o processamento uma vez para cada *pixel* da imagem rasterizada dessa geometria. O sistema aqui desenvolvido se baseia na criação uma

sequência desses programas para realizar o processamento de todas as etapas da composição do mapa.

O fluxo funcional da placa de vídeo não permite nenhuma interação além da inserção desses programas, que atuam especificamente em duas posições nesse *pipeline*, que opera, basicamente, recebendo geometrias (vértices) e texturas (imagens rasterizadas) de um lado, processando esses dados e dando saída em uma textura processada, do outro. Normalmente, essa saída é a tela do computador. Porém, para realizar a aplicação de uma sequência desses *shaders*, há a opção de direcionar essa saída para a memória do aparelho, salvando a textura resultante no sistema, de modo que possa ser novamente processada e assim permitir a realização de procedimentos sequenciais mais elaborados. Essa técnica é denominada, em uma tradução livre, “renderização para textura” (*render to texture*) ou simplesmente RTT, e é usada no processamento deste sistema.

Dessa forma, este experimento realiza a montagem de pequenas imagens, previamente renderizadas e organizadas em uma estrutura de dados que permite a composição do mapa em diversos enquadramentos e níveis de detalhe. Essa arquitetura é apropriada para as condições da Internet, sendo adotada pelos principais serviços de *webmapping* por sua escalabilidade e adequação às transmissões de arquivos pela rede.

Antes de chegar a essa solução, porém, outras ideias foram experimentadas, com especial destaque para o uso de arquivos vetoriais para desenhar os mapas.

4.2.1 Experimentos preliminares

Os arquivos gráficos de natureza vetorial são bastante apropriados para a elaboração de mapas. Os mapas coropléticos, por exemplo, possuem uma estrutura funcional adequada para a adoção de arquivos vetoriais, pois nada mais são do que um conjunto de regiões associadas a um conjunto de dados, cuja representação se dá pela colorização de polígonos. Isto pode ser realizado com naturalidade usando uma estrutura de dados como o SVG, por exemplo.

Há várias ferramentas disponíveis que facilitam essa programação, oferecendo soluções prontas para muitos problemas. Bibliotecas como D3 (BOSTOCK, 2015b) e jQuery (THE JQUERY FOUNDATION, 2016), entre outras, e especificações como o GeoJSON (BUTLER *et al.*, 2008) e o TopoJSON (BOSTOCK, 2015a), permitem explorar o potencial do SVG, tornando-o uma poderosa solução para a criação de mapas interativos *on-line*.

Esse tipo de desenvolvimento, porém, apresenta um importante gargalo que precisa ser considerado. Devido à sua natureza geométrica complexa, os mapas, quando convertidos em arquivos vetoriais, produzem em uma coleção de polígonos, em ge-

ral, muito detalhados, o que os torna excessivamente pesados, tanto para realizar a transmissão desses dados pela rede, quanto para seu processamento por parte dos renderizadores.

Há métodos conhecidos para lidar com essa complexidade, como o algoritmo de generalização de polígonos proposto por DOUGLAS e PEUCKER (2011), que reduz a quantidade de vértices, preservando os mais significativos para a representação de seu formato.

O resultado, porém, limita a aplicabilidade desses arquivos otimizados a níveis de detalhe menores, pois um mapa formado por polígonos muito simplificados, quando visualizado em detalhe, perde a sua identidade formal. Por outro lado, arquivos detalhados, contendo milhões de vértices, são inviáveis para aplicações em níveis de detalhes menores, pois muitos polígonos precisam ser carregados e processados.

O uso de imagens rasterizadas também é bastante utilizado para a elaboração de mapas *on-line*. Na verdade, os serviços de mapas para a Internet mais populares se baseiam na montagem de imagens desse tipo. A idéia central é montar uma imagem a partir de pequenos pedaços de um mapa maior, previamente renderizado, através em um sistema arquitetado para otimizar a carga dos arquivos conforme a demanda, criando uma estrutura escalonável e dimensionada de acordo com a tela do dispositivo que apresentará a imagem.

Esses mapas rasterizados, porém, são mais difíceis de editar em uma aplicação interativa, uma vez que a manipulação de seus *pixels* é computacionalmente onerosa, o que torna sua aplicação inviável, por exemplo, para um aplicativo baseado na colorização de regiões da imagem.

Placas gráficas modernas, porém, permitem resolver o problema usando tecnologias que viabilizam a programação direta das suas CPUs¹, que são também conhecidas como GPUs². Por meio de programas conhecidos como *shaders*, é possível manipular *pixels* individuais da tela com grande rapidez. Isso permite o desenvolvimento de um sistema interativo, valendo-se do espaço de cor de cada *pixel* para identificar as áreas. Essa é a chave do funcionamento do mapa aqui proposto: a montagem de imagens rasterizadas em cujos *pixels* está gravada a identificação do polígono que lhe deu origem, tornando possível a associação entre ele e seus dados.

Para efeito de análise e comparação entre as características dos arquivos utilizados até aqui, foi realizado um levantamento que ajuda a esclarecer as decisões tomadas. Trata-se de um detalhamento extenso, com diversas tabelas, e por isso optou-se por disponibilizá-lo no Apêndice A, ao final desta dissertação.

¹ *Central Processing Unit*.

² *Graphics Processing Unit*.

4.2.2 Montagem do mapa

A montagem do mapa é feita exatamente como mostrado por SCHWARTZ (2016), que explica a arquitetura utilizada no serviço Bing Maps. Essa estrutura é semelhante às adotadas por outros serviços similares e funciona baseada na montagem de pequenas imagens quadradas (*tiles*, em inglês), ou ladrilhos, recortadas de mapas previamente renderizados conforme a projeção cartográfica Webmercator, em vários níveis de detalhe sequenciados. Isso permite realizar desde o enquadramento do mundo todo em uma pequena tela de aparelho móvel, até a visualização de detalhes como um campo de futebol em uma tela de TV. Essa arquitetura resolve também o problema da carga de informações desnecessárias à visualização, promovendo uma requisição enxuta e estritamente adequada ao tamanho da tela onde é exibida.

A projeção Webmercator é baseada na projeção clássica de Mercator (AITCHISON, 2011), mas com algumas pequenas diferenças, que visam a uma otimização para utilização em aplicações *on-line*. Apesar de distorcer significativamente a escala e as áreas, especialmente as localizadas nas latitudes mais extremas, ela possui duas importantes características: preserva a forma de objetos pequenos, o que evita distorções na forma dos municípios, e mantém as orientações Norte, Sul, Leste e Oeste.

Para simplificar os cálculos, foi usada a forma esférica dessa projeção, em detrimento da alternativa elipsoidal, que é mais precisa, porém mais complexa de calcular. Como o objetivo é apenas visualizar o mapa, sem utilizá-lo para realizar operações refinadas de localização e distância, a projeção esférica é suficiente. Segundo o autor, o erro dessa projeção em relação à elipsoidal é de cerca de 0,33% na escala do eixo Y , o que é desprezível para o funcionamento da visualização proposta neste sistema.

Os ladrilhos descritos, são recortes quadrados do mapa total, medindo 256×256 *pixels*. O mapa, por sua vez, possui dimensões ($mapWidth \times mapHeight$) que variam conforme o nível de detalhe (lod):

$$mapWidth = mapHeight = 256 \times 2^{lod}$$

Os mapas são quadrados, com lados de tamanhos múltiplos de 256, o que permite a organização de uma estrutura piramidal iniciada no nível 0 - em um mapa composto por um único ladrilho - dobrando de tamanho a cada novo nível, até chegar ao máximo definido para o sistema, que é 32. A Tabela 4.1 apresenta a progressão de tamanho e quantidade de ladrilhos, conforme o nível de detalhe.

Devido à distorção promovida pela projeção, a escala do mapa depende não só do nível de detalhe, mas também da latitude. Considerando que o raio da Terra

Tabela 4.1: Tamanho total dos mapas, conforme o nível de detalhe

LOD	Grade	Mapa (<i>pixels</i>)
0	1 x 1	256 x 256
1	2 x 2	512 x 512
2	4 x 4	1024 x 1024
3	8 x 8	2048 x 2048
4	16 x 16	4096 x 4096
5	32 x 32	8192 x 8192
6	64 x 64	16384 x 16384
7	128 x 128	32768 x 32768
8	256 x 256	65536 x 65536
9	512 x 512	131072 x 131072
10	1024 x 1024	262144 x 262144
11	2048 x 2048	524288 x 524288
12	4096 x 4096	1048576 x 1048576
13	8192 x 8192	2097152 x 2097152
14	16384 x 16384	4194304 x 4194304
15	32768 x 32768	8388608 x 8388608
16	65536 x 65536	16777216 x 16777216
(...)		
30	1073741824 x 1073741824	274877906944 x 274877906944
31	2147483648 x 2147483648	549755813888 x 549755813888
32	4294967296 x 4294967296	1099511627776 x 1099511627776

(*earthCircunference*) mede aproximadamente 6378137 metros, a resolução do mapa (*groundResolution*) pode ser calculada da seguinte maneira:

$$\begin{aligned}
 groundResolution &= \cos\left(\frac{\pi \times lat}{180}\right) \times \frac{earthCircunference}{mapWidth} \\
 &= \cos\left(\frac{\pi \times lat}{180}\right) \times \frac{6378137}{256 \times 2^{lod}}
 \end{aligned}$$

Em uma tela com 96 dpi (pontos por polegada) de resolução (*screenDpi*) e considerando que uma polegada vale 0,0254 metros, a escala do mapa (*mapScale*) pode ser obtida:

$$\begin{aligned}
 mapScale &= 1 : \frac{groundResolution \times screenDpi}{0.0254} \\
 &= 1 : \frac{\cos\left(\frac{lat \times \pi}{180}\right) \times (2 \times \pi \times 63787 \times 96)}{256 \times 2^{lod} \times 0.0254}
 \end{aligned}$$

Uma vez conhecido o tamanho do mapa, é possível calcular a coordenada de cada *pixel* (x, y) a partir de sua posição, mapeando-o entre os pontos (0, 0) no canto

superior esquerdo e $(mapWidth - 1, mapHeight - 1)$ no canto inferior direito da imagem. Da mesma forma, dado um ponto no espaço, definido por uma latitude (lat) e uma longitude (lng), é possível projetá-lo sobre o mapa e calcular qual *pixel* o representa:

$$\begin{aligned}
 x &= \left(\frac{lng + 180}{360} \right) \times mapWidth \\
 &= \left(\frac{lng + 180}{360} \right) \times (256 \times 2^{lod}) \\
 \\
 y &= \left(0.5 - \frac{\log \left(\frac{1 + \sin Lat}{1 - \sin Lat} \right)}{4 \times \pi} \right) \times mapWidth \\
 &= \left(0.5 - \frac{\log \left(\frac{1 + \sin \left(\frac{lat \times \pi}{180} \right)}{1 - \sin \left(\frac{lat \times \pi}{180} \right)} \right)}{4 \times \pi} \right) \times (256 \times 2^{lod})
 \end{aligned}$$

Para organizar a montagem, é necessário que os ladrilhos possuam uma identificação capaz de informar sobre sua posição e nível de detalhe. Para isso, considere-se que a montagem do mapa é organizada por uma matriz quadrada, de tamanho $(m \times m)$, onde $m = 2^{lod}$.

A posição do ladrilho é definida, portanto, no espaço que compreende as coordenadas $(0, 0)$ e $(m - 1, m - 1)$ da matriz. Por exemplo, para o nível de detalhe 3, as posições dos ladrilhos variam entre as coordenadas $(0, 0)$ e $(7, 7)$, como mostra a Figura 4.2.

Nessa matriz, é simples determinar em que ladrilho (T) se encontra o *pixel* (x, y) do mapa:

$$T_{(col, row)} = \left(\left\lfloor \frac{x}{256} \right\rfloor, \left\lfloor \frac{y}{256} \right\rfloor \right)$$

Para identificar as os ladrilhos e facilitar a sua organização são usadas as *quadkeys*, *strings* unidimensionais compostas a partir de uma combinação das coordenadas da matriz. Cada *quadkey* é uma chave que identifica unicamente um ladrilho e contém as informações necessárias para posicionamento na matriz que organiza a montagem do mapa, bem como da sua altura na pirâmide que organiza os níveis de detalhe.

Para compor as *quadkeys* é preciso converter as posições (col, row) da matriz para números binários, intercalar seus algarismos para formar um novo número, também binário, que finalmente é convertido para a base 4 e assim obtém-se a identificação. Um detalhe importante: é preciso preservar a quantidade de zeros à

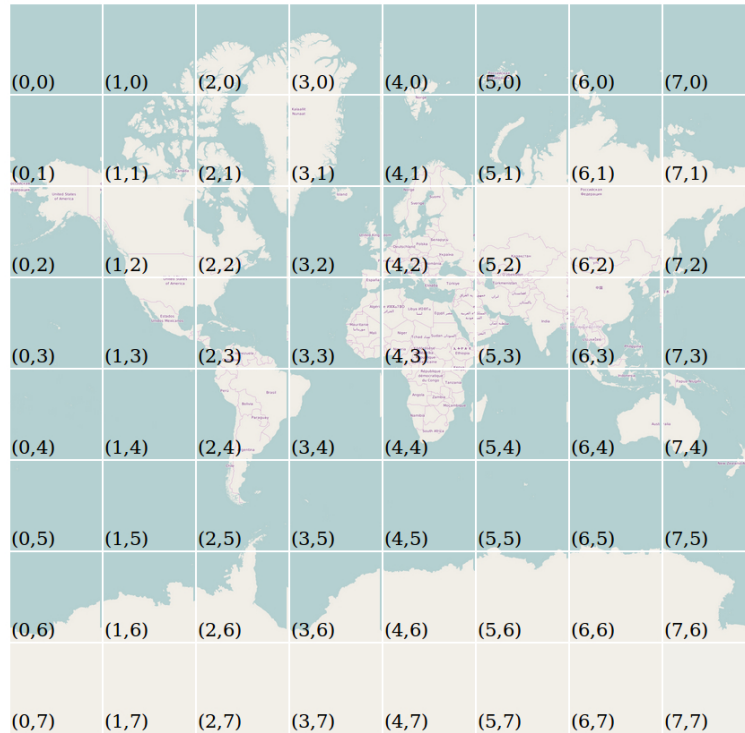


Figura 4.2: Matriz de ladrilhos no nível de detalhe 3. Imagem produzida pelo autor baseada em SCHWARTZ (2016).

esquerda, pois é a quantidade algarismos que identifica o nível de detalhe. Abaixo, o exemplo do cálculo da *quadkey* referente ao ladrilho (3, 5) do nível de detalhe 3:

$$T_{(3,5)} = (3_{10}, 5_{10}) = (011_2, 101_2) = 100111_2 = 213_4$$

As *quadkeys* possuem características interessantes e essenciais para o funcionamento do sistema, pois a quantidade de algarismos identifica o nível de detalhe e sua ordem determina a hierarquia da árvore, conforme mostra a Figura 4.3.

A montagem do mapa, então, é realizada. O tamanho da tela define a quantidade de ladrilhos necessários para a exibição daquela porção do mapa. O centro é definido pelo usuário e pode ser projetado no espaço em *pixels* do mapa a partir de sua posição (*lat, lng*) e do seu nível de detalhe, que também é definido pelo usuário. Uma vez conhecido o centro projetado no espaço do mapa, pode-se obter a localização dos *ladrilhos* que cobrem os extremos da tela e assim fazer uma caminhada pelas *quadkeys* até que se preencha toda a tela. Para o ajuste de posição, deve-se empregar o vetor de deslocamento obtido a partir do resto dessa divisão. Movimentar o mapa, então, se torna simplesmente modificar sequencialmente esse centro. A mudança do nível de detalhe usa essa mesma sequência, apenas atualizando essa variável no sistema, de modo que sejam carregados os ladrilhos daquele nível.

A etapa de carga e montagem dos ladrilhos é realizada duas vezes. Uma monta

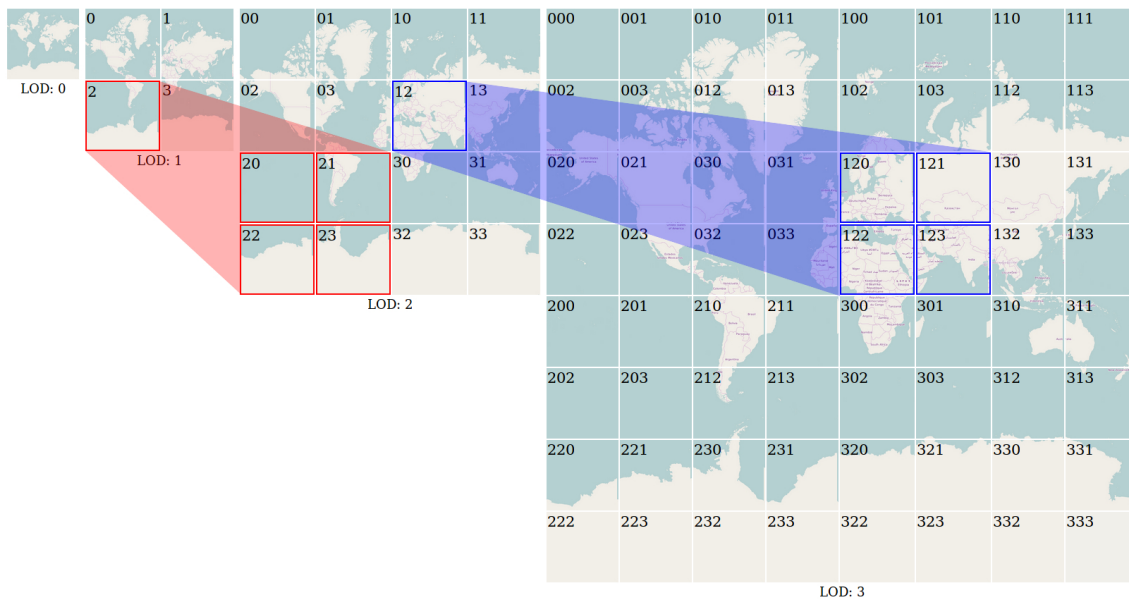


Figura 4.3: Exemplos de relações entre os ladrilhos e suas identificações, conforme o nível de detalhe. Imagem produzida pelo autor baseada em SCHWARTZ (2016).

o mapa do fundo, trazendo consigo a referência topográfica para a visualização da localização. A outra, mais complexa, é formada por um conjunto de ladrilhos especialmente preparados para conter em seus *pixels* as identificações dos municípios, o que permite fazer o relacionamento entre o mapa e as variáveis.

4.2.3 Processamento prévio dos ladrilhos

A chave para a conexão entre os municípios representados no mapa e as variáveis respectivas é o Código de Área, uma identificação oficial que organiza as subdivisões geográficas oficiais do país. As informações estatísticas georreferenciadas divulgadas pelo IBGE também são identificadas por esses mesmos códigos, o que permite o relacionamento entre elas e as áreas. No presente sistema esse código ficará registrado no espaço de cor dos *pixels*, de modo que cada uma dessas unidades mínimas da imagem poderá ser processada em paralelo, direto na placa gráfica, através dos *shaders*.

Funcionamento dos códigos de área

O IBGE estabelece um código identificador numérico único para todas as subdivisões político-administrativas do país. É denominado Código de Área e funciona como uma chave para identificação da área e de sua posição na hierarquia.

O Brasil é composto por cinco Grandes Regiões, subdivididas em 27 Unidades Federativas. Cada Unidade Federativa é subdividida em mesorregiões, que, por sua vez, são subdivididas em microrregiões e estas em municípios. Os municípios são

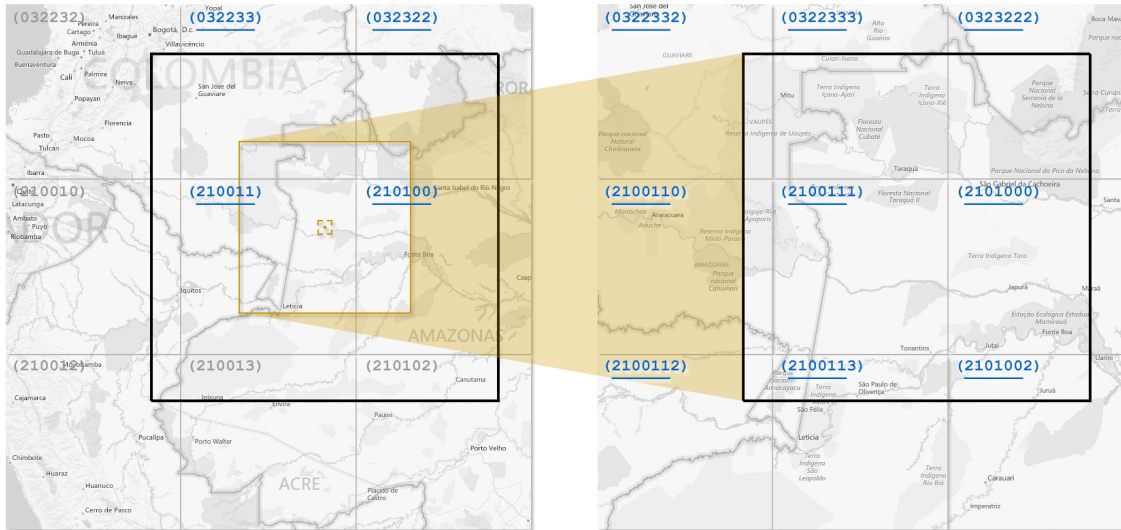


Figura 4.4: Modificação do nível de detalhe: manutenção da posição e carga de novos ladrilhos, conforme sua *quadkey*.

subdivididos em distritos e estes em subdistritos. Os subdistritos são divididos em setores censitários, que são as menores unidades usadas para a divulgação pública de dados pelo IBGE.

Estas subdivisões formam uma hierarquia rígida, na qual cada ponto no espaço pertence exclusivamente a um setor, que pertencerá a um único subdistrito e assim sucessivamente até chegar a uma das Grandes Regiões contidas nos limites políticos do Brasil. As Tabelas 4.2 e 4.3 mostram como o código funciona.

Região	UF	Mesorregião	Microrregião	Município	Distrito	Subdistrito	Setor
3	3	0	4	557	05	09	0013

Tabela 4.2: Significado da codificação do setor censitário n^o 330455705090013

Nível	Nome	Código
Grande região	Sudeste	3
UF	Rio de Janeiro	33
Mesorregião	Metropolitana do Rio de Janeiro	330
Microrregião	Rio de Janeiro	3304
Município	Rio de Janeiro	3304557
Distrito	Rio de Janeiro	330455705
Subdistrito	Ilha do Governador	33045570525
Setor censitário	-	330455705250280

Tabela 4.3: Código como identificador de herança

Para ilustrar melhor, a Figura 4.5 mostra um setor censitário localizado no sul da Ilha do Fundão, na cidade do Rio de Janeiro. O código de área de um setor censitário possui quinze algarismos.

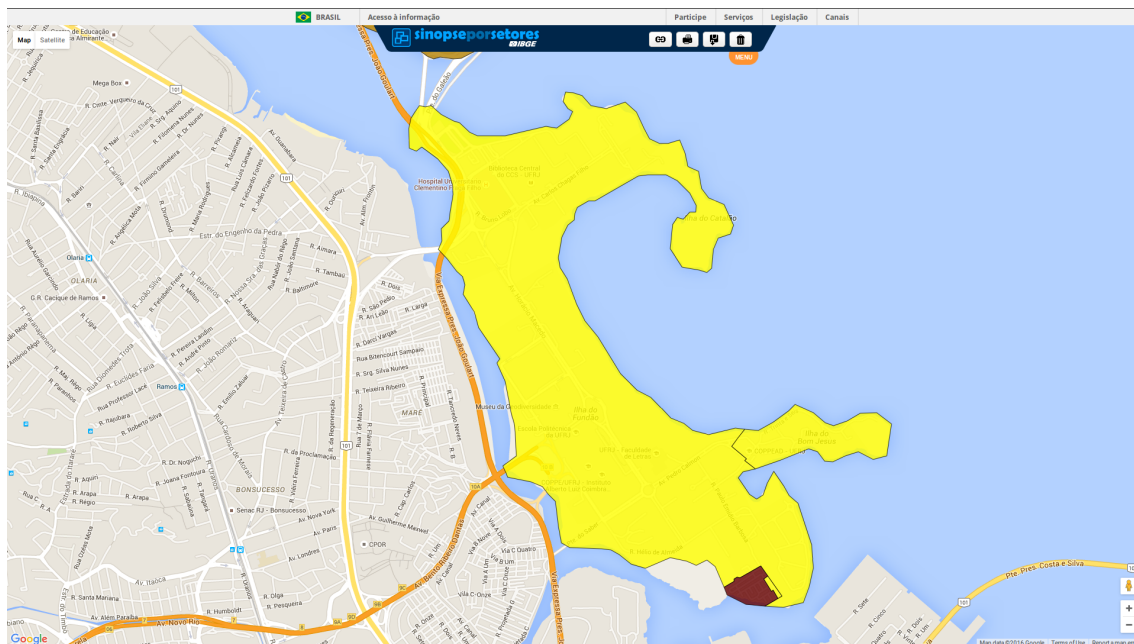


Figura 4.5: A imagem mostra o setor censitário n° 330455705250280, localizado no sul da Ilha do Fundão, no Rio de Janeiro, pintado de marrom escuro. Imagem capturada do aplicativo Sinopse por Setores, disponibilizado pelo IBGE em seu portal na Internet (IBGE, 2011).

Nos municípios, os códigos de área possuem seis algarismos e poderiam perfeitamente ser registrados no espaço de cor RGB de um *pixel*, ao contrário da codificação dos setores, que exige um espaço muito maior.

Previendo uma implementação futura para visualização dos setores censitários, a solução foi mapear esses códigos identificadores para um novo índice, obtido pela posição na lista ordenada dos códigos das regiões para cada nível geográfico. Dessa forma, os três canais (RGB) são suficientes para guardar os pouco menos de 6 mil municípios e os cerca de 320 mil índices dos setores censitários.

Rasterização dos ladrilhos

Como o sistema realiza a visualização somente do território brasileiro, foram preparadas apenas as imagens necessárias para cobri-lo entre o nível de detalhe 5 - o suficiente para enquadrar toda extensão continental brasileira em uma tela com 1024×1024 *pixels* de tamanho - e o nível 10 - a ampliação suficiente para visualizar em um tamanho legível nessa mesma tela o menor município do país, a cidade de Santa Cruz de Minas (MG) (IBGE, 2016a).

A confecção desses ladrilhos foi realizada por um programa, desenvolvido em Python, que lê os arquivos vetoriais originais e rasteriza os polígonos, colorindo os *pixels* conforme os índices das áreas que lhes dizem respeito. As imagens são salvas no formato PNG, que preserva as informações precisas das cores e consequentemente

dos códigos.

Dessa forma, são gerados todos os ladrilhos necessários para compor a imagem do Brasil no maior nível de detalhe previsto. Os ladrilhos dos níveis inferiores são produzidos a partir dos de nível maior. Esse procedimento exige um algoritmo que garanta a integridade dos *pixels* limítrofes, que não podem ter suas informações de cor misturadas com as dos vizinhos. Como o *pixel* funciona como uma unidade de identificação, ou ele pertence a um município apenas ou é vazio.

A imagem dobra de tamanho a cada novo nível. Logo, para gerar esses novos ladrilhos, basta acessar os quatro ladrilhos respectivos à sua posição nos níveis superiores e decidir, para cada *pixel* do novo ladrilho, qual dos quatro *pixels* superiores se manterá registrado. Essas reduções, porém, suscitam uma discussão importante, acerca da preservação de todas as identificações necessárias ao mapa.

Representação única de todas as regiões

No presente sistema, o *pixel* codificado é a chave para o processamento das informações. Logo, é preciso desenvolver um algoritmo que garanta a sobrevivência de todas as áreas durante as reduções de níveis de detalhe, uma vez que há municípios pequenos que podem desaparecer da imagem quando reduzidos.

Especificamente para o caso dos municípios brasileiros, a partir do nível de detalhe 6, é garantido que todas as cidades possuem algum *pixel* representante no mapa. Porém, em níveis menores, algumas cidades desaparecem e ficam sem nenhum registro.

A algoritmo usado para gerar os ladrilhos foi o mais simples possível e não garantiu a manutenção das áreas. As reduções se deram da seguinte maneira: conforme a arquitetura proposta, a nova imagem é uma redução de 50% dos 4 ladrilhos filhos, no nível de detalhe superior. Isso significa que a cor de cada *pixel* desta nova imagem, é decidida entre as cores dos 4 *pixels* correspondentes no nível acima, conforme o seguinte pseudocódigo:

```
Lê os 4 pixels da imagem superior;
Se houver cor mais representada:
    grava esta cor;
senão:
    sorteia uma das cores;
    grava esta cor;
```

Esse algoritmo não garante a sobrevivência de nenhum *pixel* específico, que pode ser o último representante de uma área, deixando-a sem identificador a partir daqui. Por isso, é necessário pensar em maneiras de resolver essa questão, desenvolvendo

um algoritmo que seja capaz de garantir a permanência de, pelo menos, um *pixel* para cada área que se precisa representar.

É possível estimar um limite mínimo para o problema. Considerando uma imagem quadrada, por sobre onde são organizados os 5570 municípios, cada qual exigindo pelo menos um *pixel* representante, é possível afirmar que essa imagem deveria ter, pelo menos, o tamanho de 75×75 *pixels* para abrigar todos eles. Obviamente, seria um limite tão estreito que as referências de posição e vizinhança das áreas se perderiam completamente. Pode-se prever, porém, que quanto maior a imagem, melhor a distribuição, até que se chegue ao tamanho equivalente ao nível de detalhe 6, no qual, todas as unidades territoriais estão representadas, mantendo relações de posição e vizinhança.

A quantidade de áreas representadas no mapa determina, portanto, um limite mínimo para o tamanho da imagem. Levando isso em consideração, é possível modificar o algoritmo descrito anteriormente, e produzir um melhor, simplesmente modificando a decisão sobre o *pixel* persistente, dessa vez selecionando o menos representado. Este procedimento melhoraria a manutenção das áreas menores, porém, ainda assim, haveria a chance de duas áreas como essas estarem destinadas a um mesmo grupo, o que forçaria um sorteio que faria uma delas desaparecer:

```
Lê os 4 pixels da imagem superior;  
Se houver cor menos representada:  
    grava esta cor;  
senão:  
    sorteia uma das cores menos representadas;  
    grava esta cor;
```

Aumentando a área de varredura, talvez se obtivesse uma melhora. Ou então, buscando uma estrutura de dados que garantisse a manutenção conforme a quantidade de *pixels* restantes em cada área. Uma outra idéia seria iniciar a rasterização dos novos ladrilhos a partir dos *pixels* mais representativos de cada uma das áreas - como o seu centroide ou a localização de sua sede, por exemplo. O problema é que, mesmo assim ainda poderiam ocorrer colisões que poriam em risco o funcionamento perfeito do algoritmo.

A questão é que, em uma primeira análise, trata-se de um problema de otimização, sobre o qual, por mais que se tente, não há como não levar em conta dinamicamente o contexto global. Estudos como o de HEILMANN *et al.* (2004), que trata de um algoritmo para a realização de um anamorfismo cartográfico, podem ser levantados para ver se ajudam na elaboração de uma solução para o problema aqui descrito.

Essa breve reflexão se encontra aqui descrita apenas para registrar o problema e tentar definir seu escopo. A questão, porém, não foi melhor desenvolvida neste trabalho. Neste experimento se encontra implementado o algoritmo mais elementar, referente ao primeiro pseudocódigo descrito nesta seção, o que, conseqüentemente, produziu imagens com algum nível de erro nos níveis de detalhe menores. Como essa visualização foi limitada ao nível 5, não ocorreram problemas graves na visualização. É necessário, entretanto, que essa questão seja melhor investigada em eventuais desdobramentos do presente trabalho.

Processamento dos ladrilhos de fundo

Para as imagens de fundo, que servirão como referência topográficas para as camadas visualizadas acima, foram usadas as imagens originais padrão do serviço Bing Maps. Imagens equivalentes poderiam ter sido geradas a partir de outros insumos públicos, porém, a utilização das imagens oriundas de um serviço externo ajudaram a validar a correção do algoritmo utilizado para a rasterização. Essas imagens foram salvas e convertidas para preto, branco e cinzas, de modo não atrapalhem a visualização das cores que são visualizadas nas camadas acima. Essas imagens foram salvas no formato JPG, que permitiu uma melhor compactação dos arquivos.

4.2.4 Imagem contendo os dados

O mapa é composto, portanto, por um conjunto de ladrilhos contendo imagens de referência, e outro, contendo a codificação necessária para estabelecer o relacionamento entre dados e áreas.

A idéia inicial era carregar os dados a partir de arquivos textuais, em formato XML ou JSON, contendo os dados associados aos códigos. Porém, mediante a naturalidade com que a placa de vídeo lê e processa imagens, optou-se por uma saída diferente.

O espaço nos canais RGBA das imagens é discreto, dividido em valores inteiros compreendidos entre 0 e 255. Ele determina o valor das três cores primárias do sistema aditivo (RGB) e seu canal de transparência (A). Esse espaço, permite uma combinação que promove 256^4 (4294967296) possibilidades de cor.

Pensando nisso, optou-se por gravar os valores das 140 variáveis disponíveis no sistema diretamente em imagens com espaço de cor RGBA. Foi gravada uma imagem para cada variável. Assim, uma imagem correspondente a uma variável V possui um *pixel* por município, que codifica o valor daquela variável.

Há vantagens práticas na adoção das imagens para esse fim, como a sua inserção direta na memória da placa de vídeo; mas há também limitações, como a restrição ao uso a valores inteiros e menores que 256^4 (pouco mais do que 4 bilhões). Uma

maneira de lidar com essa limitação seria usar uma imagem maior para registrar os dados, utilizando dois ou mais *pixels* para guardar cada valor, estendendo o raciocínio para a gravação de eventuais valores decimais. Porém, os dados municipais não chegam a essa ordem de grandeza e no sistema aqui descrito foram usados apenas valores inteiros que coubessem no espaço de um *pixel*.

Foi decidido também que a largura dessa imagem seria de 256 *pixels*, coincidindo com a dimensão de um canal de cor. Logo, para guardar 5570 valores distintos em uma imagem com esta largura, são necessárias $\lceil \frac{5570}{256} \rceil = 22$ linhas de *pixels*. Para organizar os dados de modo que pudessem ser recuperados, utilizou-se o mesmo índice identificador definido para os municípios, ou seja, um valor ordenado iniciado em 0 e com valor máximo de 5569 para enfileirar as informações pelos *pixels* da imagem, linha por linha, coluna por coluna. Assim, o índice pode ser convertido na posição (u, v) da textura, onde estará localizado o valor desejado. Dessa forma é feito o relacionamento entre o *pixel* identificado no ladrilho e o *pixel* contendo o valor da textura referente à variável, conforme mostra a Figura 4.6.

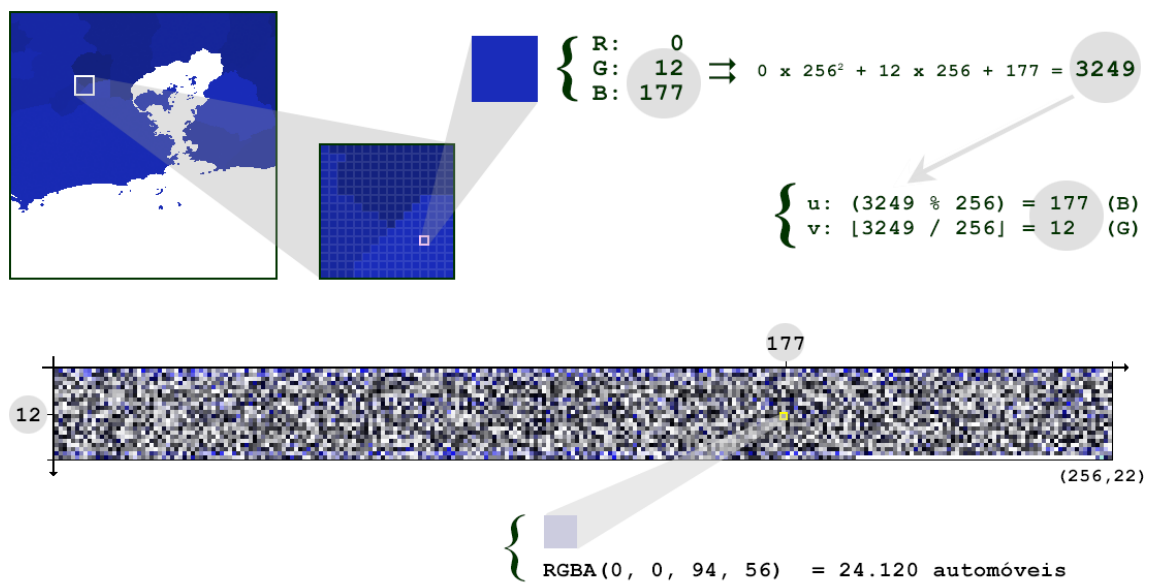


Figura 4.6: A figura mostra a associação entre os índices gravados nos *pixels* da imagem e a posição (u, v) da textura contendo as informações que estão gravadas no espaço RGBA dos *pixels* das imagens contendo os dados de cada uma das variáveis.

Como a largura da imagem possui o mesmo tamanho dos canais de cor da imagem, a coordenada (u, v) dos municípios possui exatamente os valores dos canais B e G do seu respectivo índice, conforme mostra a Figura 4.6.

Outra informação importante que precisou ser processada e guardada em imagens como essas é o tamanho, em *pixels*, de cada município, em cada nível de detalhe. Essa informação é essencial para o funcionamento do sistema, uma vez que ele se baseia na decomposição dos valores das variáveis por unidade de área, aqui representadas

por *pixels*, naquele determinado nível de detalhe.

4.2.5 Composição da grade

Para realizar a visualização proposta pelo sistema, é necessário criar uma grade, composta por células quadradas variando, conforme o desejo do usuário, entre 1×1 e 32×32 *pixels*. Cada célula tem seu valor calculado a partir do somatório das frações referentes aos valores estimados das variáveis municipais para cada *pixel* que lhe pertence. Em outras palavras, isso significa que é realizada uma divisão simples dos valores referentes às variáveis municipais pela quantidade de *pixels* referentes a cada um dos municípios que compõem a imagem (naquele determinado nível de detalhe) (Figura 4.7).

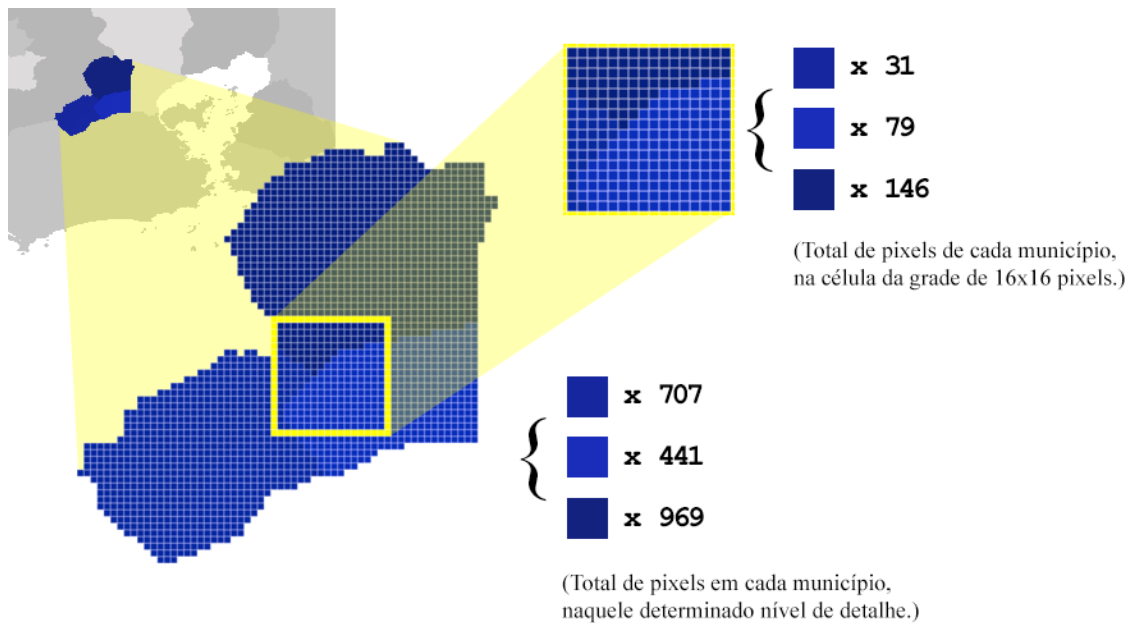


Figura 4.7: A grade é composta pelo somatório dos valores respectivos à decomposição dos dados pela quantidade de *pixels* em cada região, naquele determinado nível de detalhe.

Para facilitar o processamento, o mapa também é quadrado, com lados de tamanho igual a uma potência de 2, de modo que sua divisão pelo tamanho das células resulte em um número inteiro. Isso facilita o procedimento para o cálculo dos valores mínimos e máximos da grade, informação necessária para a etapa de classificação dos dados.

O problema, portanto, pode ser definido sobre um mapa, *map*, consistindo de uma imagem quadrada, de tamanho $(m \times m)$, composta pelos *pixels* P_m , tal que m é uma potência de 2. A grade, por sua vez, é composta por células quadradas, *cell*, de tamanho $(c \times c)$, tal que $1 \leq c \leq 32$. Os valores referentes a cada célula são

guardados em uma textura, *grid*, quadrada, de tamanho $(g \times g)$. Uma vez que c e m são potências de 2 e $m > c$, o tamanho da textura *grid* é definido por: $g = \frac{m}{c}$. Cada *pixel*, P_g de *grid* corresponde a uma célula *cell*, que, por sua vez, corresponde a um conjunto quadrado de tamanho $(c \times c)$ de *map*.

O valor T de um tema representado neste sistema é uma taxa calculada a partir de duas variáveis, V_1 e V_2 , e um fator numérico, F , selecionados pelo usuário, de modo que o valor do tema, portanto, é dado por $T \leftarrow \frac{V_1 \times F}{V_2}$. Sendo assim, essa grade pode ser criada em um processamento descrito pelo seguinte pseudocódigo:

Sejam:

map a textura do mapa;

dBase₁ a textura com os dados da variável 1;

dBase₂ a textura com os dados da variável 2;

dBase_{px} a textura com a quantidade de *pixels* de cada município, naquele nível de detalhe;

grid a textura que guardará as informações da grade.

Para cada *pixel* T_g de *grid*, faça:

$V_1, V_2 \leftarrow 0$;

Para cada *pixel* P_m de *map*, referente à célula definida por P_g , faça:

$V_1 \leftarrow \left(V_1 + \frac{\text{valor contido em } dBase_1}{\text{valor contido em } dBase_{px}} \right)$;

$V_2 \leftarrow \left(V_2 + \frac{\text{valor contido em } dBase_2}{\text{valor contido em } dBase_{px}} \right)$.

$T_g \leftarrow \left(\frac{V_1 \times F}{V_2} \right)$.

Cada *pixel* da grade se refere a um pedaço quadrado da imagem do mapa (que contém as referências). Para encontrar a posição inicial (x_{mapa}, y_{mapa}) de cada um desses *pixels*, basta multiplicar sua posição (x_{grid}, y_{grid}) na grade pelo tamanho da célula, c :

$$(x_{mapa}, y_{mapa}) = (x_{grid}, y_{grid}) \times c$$

Esse procedimento é realizado uma vez para cada tema (lembrando que são dois temas), gerando duas texturas, contendo as taxas calculadas para as células das grades de cada tema, que, por sua vez, são novamente processadas e convertidas nas camadas a serem visualizadas sobre o mapa.

Os valores finais, resultantes de somatórios de frações, são arredondados, o que produz um pequeno erro. É preciso cuidado, portanto, na composição da fórmula por parte do usuário. Valores muito pequenos podem desaparecer. É preciso pensar em implementações futuras que contemplem métodos para lidar com valores decimais.

4.2.6 Classificação

Essas texturas são classificadas segundo escalas determinadas pelos usuários, que são definidas conforme os valores máximo e mínimo de cada grade. É preciso realizar um novo procedimento para obter esses valores.

Esse processamento é realizado por um algoritmo recursivo, que grava, a cada ciclo, uma nova textura com um quarto do tamanho da anterior (metade do tamanho do lado), onde são gravados os máximos e mínimos de cada grupo de 4 *pixels* da textura precedente. Dessa forma, é construída uma estrutura piramidal que converte um problema linear em logarítmico, aproveitando o paralelismo da GPU para melhorar consideravelmente o desempenho. A Figura 4.8 ilustra esse procedimento.

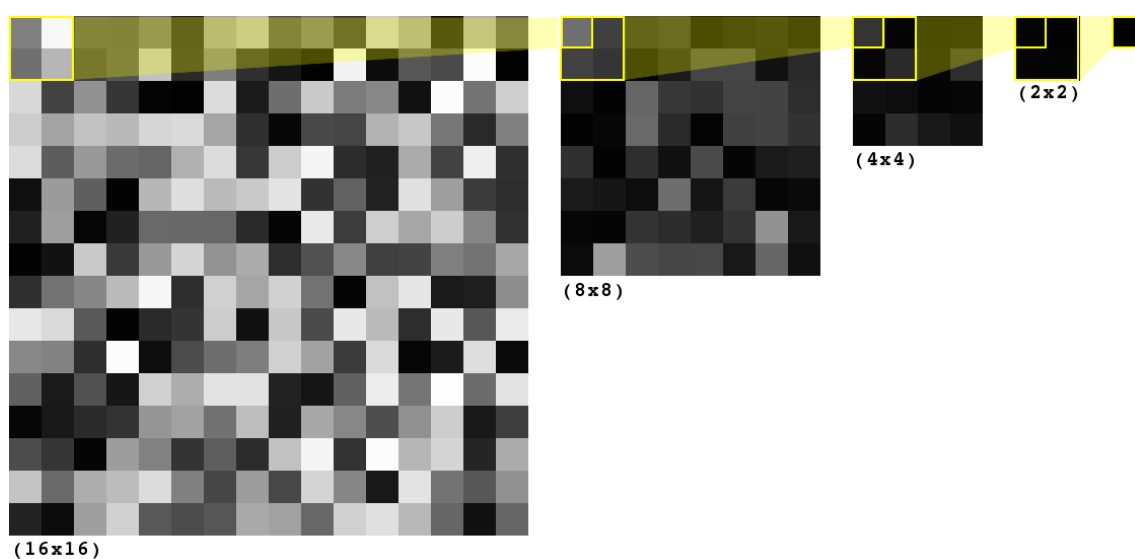


Figura 4.8: A imagem mostra as etapas para calcular os valores mínimos utilizando o método. São necessárias 4 passagens para processar uma textura de 16 x 16 *pixels*.

Interpolação

Antes de realizar a classificação, cabe ao usuário optar por dois tipos de visualização: a renderização simples dessa grade, mostrando as células quadradas; ou a renderização interpolada dos valores, produzindo uma variação gradual entre os valores dos *pixels*, resultando em uma imagem aparência mais suave.

Para entender esse procedimento, é necessário considerar que o que ocorre, na prática, é a ampliação das texturas contendo os valores da grade, de modo que ocupem novamente o tamanho do mapa. Essa ampliação pode ser realizada de dois modos: uma projeção simples dos *pixels*, sem processamento algum de suavização, onde cada um deles terá seu o valor exato replicado em sua respectiva posição na grade, o que produz uma imagem de aspecto quadriculado; ou pode ser feita uma ampliação suave, através de um processo onde os valores dos *pixels* são definidos por

uma interpolação linear dos valores da grade. Dessa maneira, a imagem resultante apresenta uma progressão suave entre valores mínimos e máximos, mais propícia à visualização de picos e vales na informação, produzindo uma classificação com desenhos curvilíneos de aspecto mais natural. A Figura 4.9 ilustra o resultado.

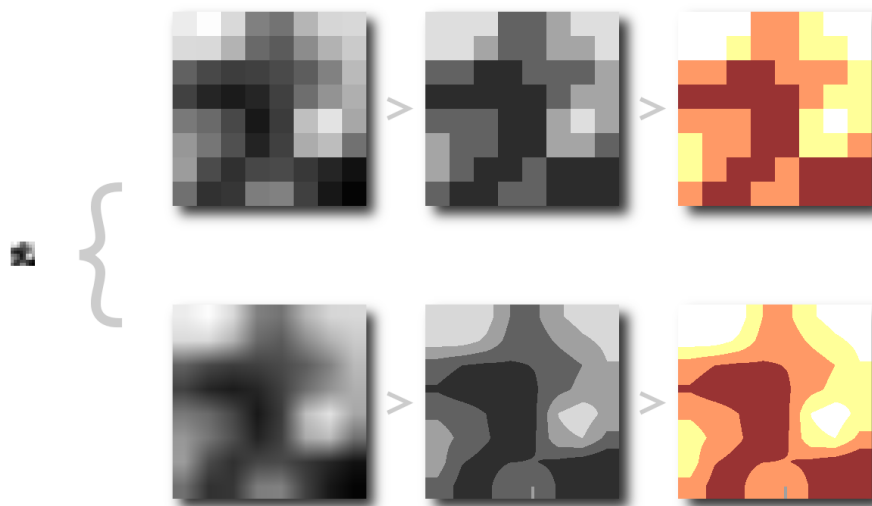


Figura 4.9: A imagem mostra as visualizações da grade nas formas simples e interpolada e o resultado após a classificação dos dados.

O objetivo dessa interpolação é estabelecer um recorte mais orgânico na imagem. Considerando a interação com o usuário, que pode interferir e animar os resultados na tela, essa interpolação promove uma visualização diferente, estabelecendo um novo ponto de vista sobre a informação. A observação da grade, sem a aplicação da interpolação, determina limites precisos, cartesianos, expressando com exatidão os valores contidos nas células (mesmo que se tratem apenas de estimativas). A versão interpolada, entretanto, permite a expressão visual dessa incerteza, onde os desenhos, curvilíneos, expressam limites que não dizem respeito nem às divisões políticas oficiais - arbitradas por leis e pela topografia - nem às linhas retas da grade, fazendo revelar as fronteiras estabelecidas pelas informações abstratas representadas no mapa.

Uma vez ampliadas - lembrando que são duas texturas de grade para dois temas - essas novas texturas passam por um processo de classificação conforme as escalas definidas pelo usuário, o que consiste em mapear cada um de seus *pixels* em uma das cinco classes relativas, em um procedimento definido da seguinte forma:

Seja T um dos temas computados pelo sistema, para o qual se determinou os valores mínimo T_{min} e máximo T_{max} . Sejam também $q_i, i \in 0, 1, \dots, 4$, os valores limítrofes relativos entre uma classe e a seguinte, conforme especificados pelo usuário na interface. Então, os valores limítrofes absolutos Q_i para a variável correspondem

a

$$Q_i = T_{min} + (T_{max} - T_{min}) \times q_i$$

Um *pixel* cujo valor computado é V pertence, portanto, à classe i se $Q_{i-1} < V \leq Q_i$.

Assim, é gerada uma nova textura, que mapeia em cada um de seus *pixels* a sua condição: se é transparente ou se pertence a uma das três classes visíveis.

4.2.7 Visualização das camadas

Agora é necessário realizar o último processamento, que envolve a mistura dessas duas texturas contendo as classificações relativas aos dois temas representados no mapa e a textura de fundo, composta pela montagem dos ladrilhos contendo a topografia de referência. O algoritmo implementado é detalhado no pseudocódigo descrito abaixo:

Sejam:

F a cor do *pixel* da imagem de fundo;

i e j as classes relativas às duas variáveis, tais que $0 \leq i, j \leq 5$;

Pt_1 , Pt_2 e Pt_3 as cores dos *pixels* das texturas representativas das classes do Tema 2, tais que Pt_1 , Pt_2 e Pt_3 in [Preto, Vazio];

P a cor final do *pixel* final;

Se a imagem de fundo está ligada, então:

$P \leftarrow F$;

Senão:

$P \leftarrow$ Branco;

Conforme o valor de i , determine:

$P_1 \leftarrow$ [Vazio, Amarelo, Laranja, Vermelho, Vazio];

Conforme o valor de j , determine:

$P_2 \leftarrow$ [Vazio, Pt_1 , Pt_2 , Pt_3 , Vazio];

Se a transparência não estiver habilitada pelo usuário, então:

Se P_2 é Preto, então:

$P \leftarrow P_2$;

Senão:

$P \leftarrow P_1$;

Senão:

Se P_2 é Preto, então:

$P \leftarrow \left(\frac{P+P_2}{2} \right)$;

Senão:

$P \leftarrow \left(\frac{P+P_1}{2} \right)$;

O pseudocódigo descrito acima não trata das ações de ligar ou desligar cada tema, nem da visualização dos limites da grade.

Para ligar ou desligar as camadas, basta inserir uma decisão ao selecionar a cor do *pixel* referente ao tema. Se não estiver habilitado, pinta-se o *pixel* do outro tema (caso ele próprio não esteja desabilitado. Se estiver, sobra apenas a cor do fundo).

A visualização da grade é mais simples. Basta decidir, antes da execução desse algoritmo, se a posição do *pixel* coincide com o primeiro *pixel* da grade:

Sejam:

x e y as posições horizontal e vertical do *pixel* P na imagem e;

c o tamanho em *pixels* da célula quadrada da grade.

Se $\lfloor \frac{x}{c} \rfloor$ é igual a 0 ou $\lfloor \frac{y}{c} \rfloor$ é igual a 0:

$P \leftarrow$ Cinza;

Senão:

(realiza o procedimento descrito no início desta seção.)

Dessa forma, a montagem do mapa é composta em uma operação deflagrada a cada novo *frame* da área de exibição. Como a renderização é instantânea, o resultado na tela são transições animadas conforme as ações do usuário. A Figura 4.10 apresenta um fluxo esquemático do funcionamento do sistema, desde a seleção e enquadramento dos ladrilhos até a fusão das texturas finais.

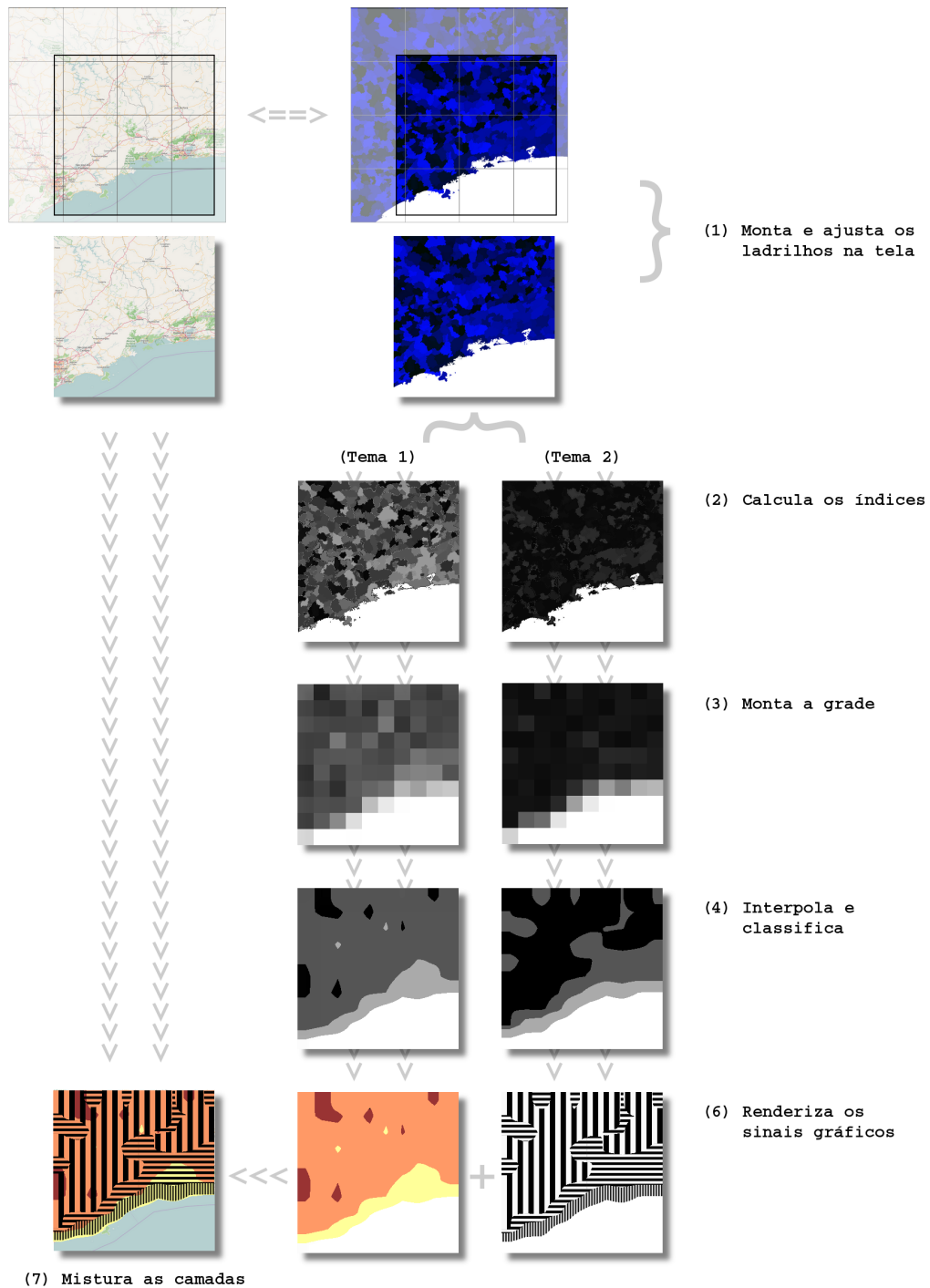


Figura 4.10: Fluxo funcional do sistema.

Capítulo 5

Resultados

Para experimentar o sistema e mostrar que tipo de resultados ele é capaz de apresentar, foram feitos dois estudos de caso com o objetivo de explorar as capacidades da ferramenta na análise dos dados disponíveis.

5.1 Frota de Veículos

O primeiro estudo utiliza as informações sobre a Frota de Veículos. Nele são apresentados os mapas representados nas Figuras 5.1, 5.2, 5.3, 5.4, 5.5 e 5.6.

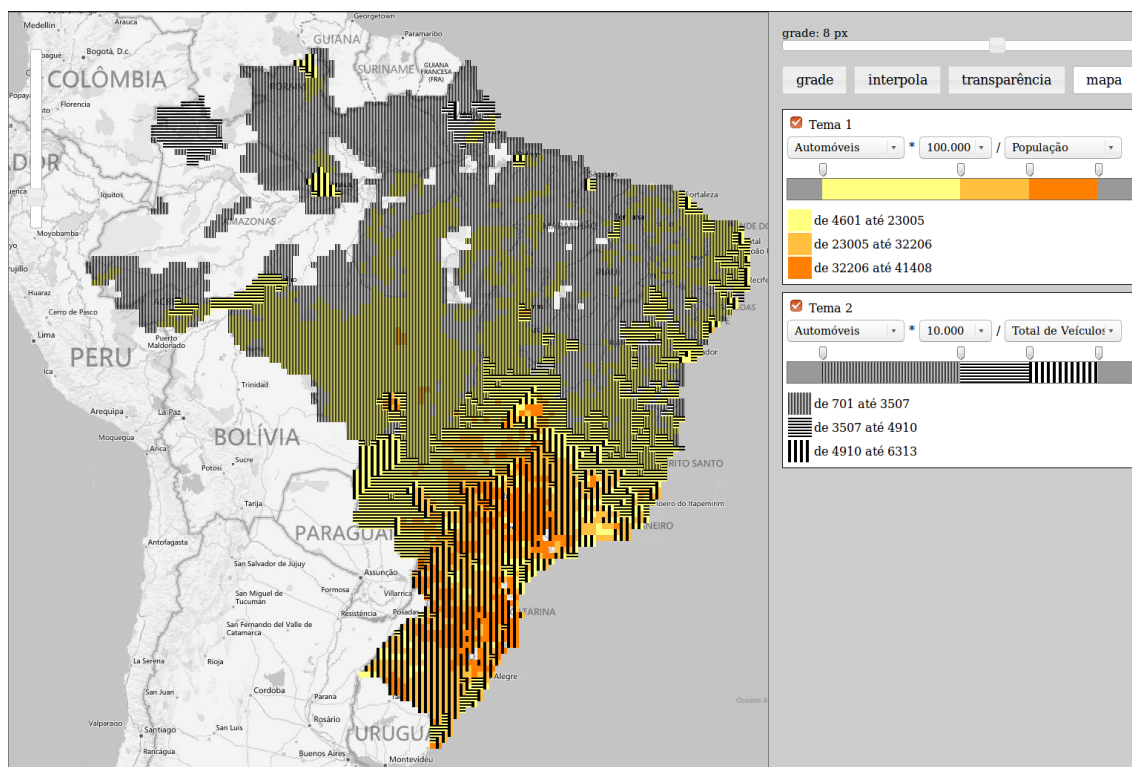


Figura 5.1: O mapa acima apresenta a quantidade de automóveis para cada 100 mil habitantes e a quantidade de automóveis para cada 10 mil veículos.

As três primeiras imagens (Figuras 5.1, 5.2 e 5.3) mostram a situação dos automóveis com relação à população e ao total de veículos automotores. É possível ver com clareza que em ambos os casos, as Regiões Sul e Sudeste do país concentram a maior proporção de carros. As Figuras 5.2 e 5.3 exemplificam como as ações de visualização individual das camadas facilitam a compreensão do mapa.

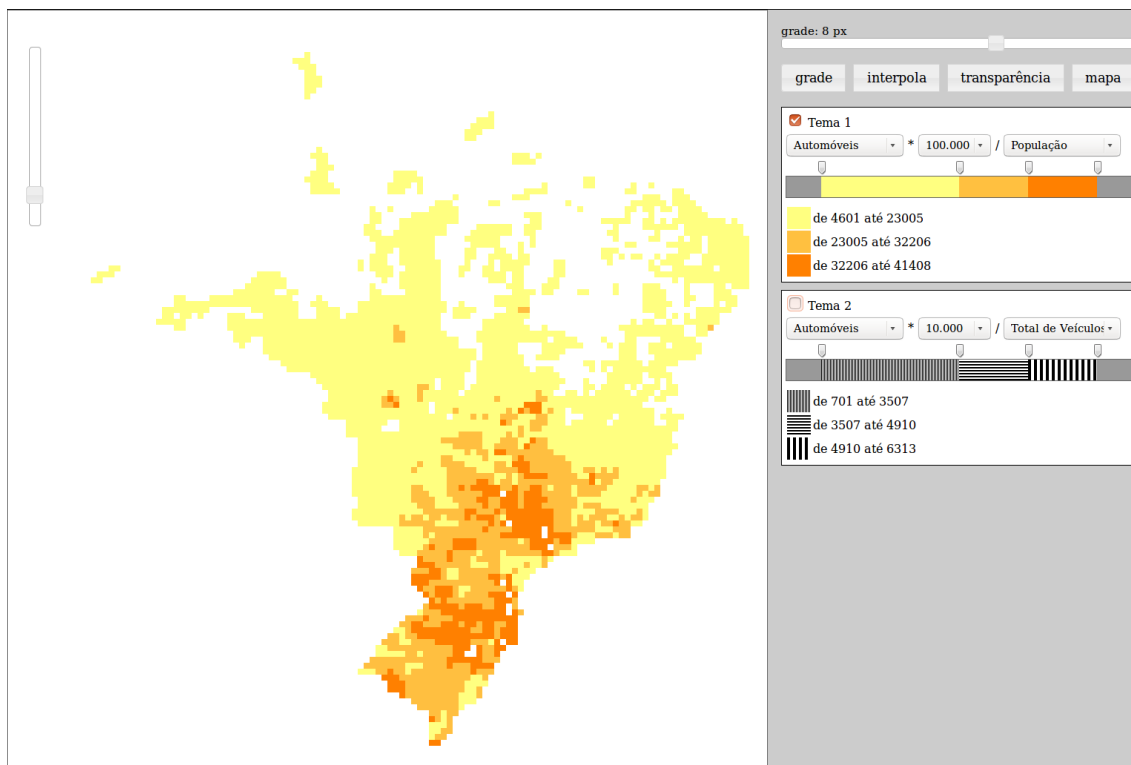


Figura 5.2: O mapa acima apresenta os mesmos dados representados na Figura 5.1. Nesta imagem estão desligadas a imagem do fundo e a camada relativa ao Tema 2.

As três imagens seguintes (Figuras 5.4, 5.5 e 5.6) são semelhantes, e apresentam a situação das motocicletas com relação à população e ao total de veículos automotores. Nelas é possível fazer um contraponto às informações respectivas aos automóveis, mostrando que, no caso das motocicletas, são as Regiões Norte e Nordeste do país que se destacam pela utilização mais intensa deste tipo de veículo.

5.2 Eleições Presidenciais Brasileiras de 2014

No segundo estudo de caso, foram selecionadas três variáveis respectivas aos resultados do segundo turno das eleições presidenciais no Brasil em 2014: Total de votos recebidos pela candidata Dilma Rousseff, Total de votos recebidos pelo candidato Aécio Neves e Total de votos válidos.

As Figuras 5.7, 5.8 e 5.9 apresentam as quantidades de votos recebidos por cada candidato para cada 100 mil votos válidos. O Tema 1 representa a quantidade rece-

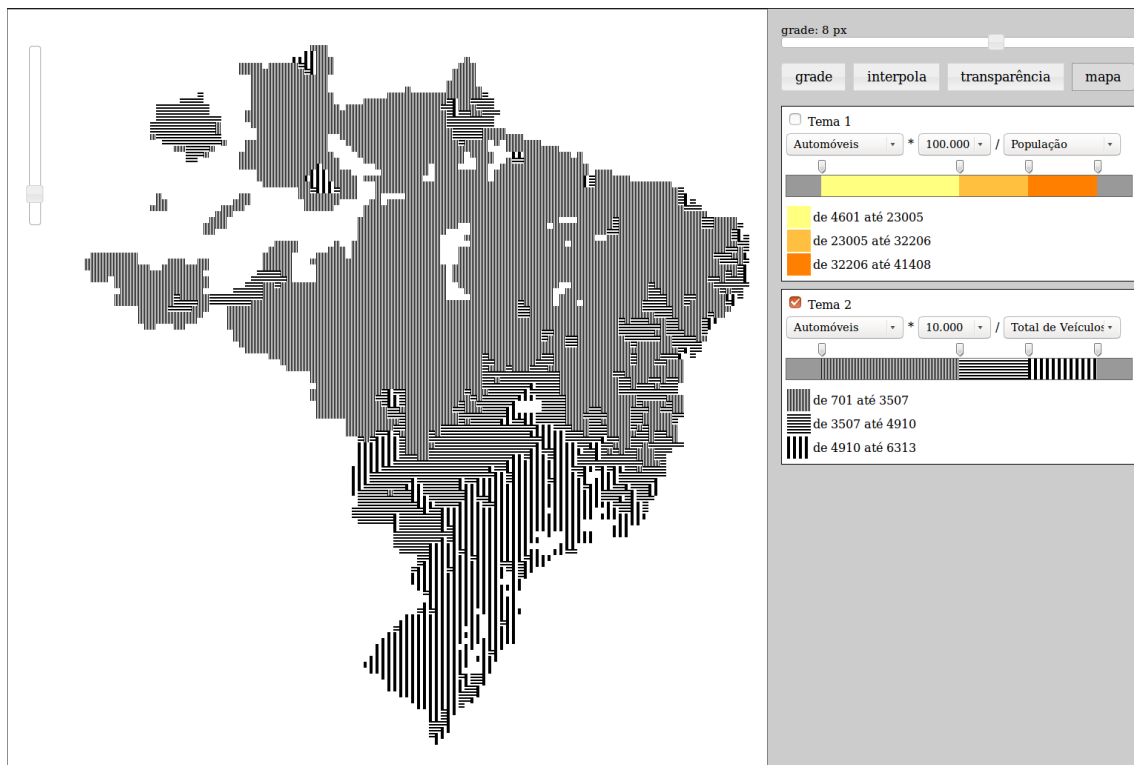


Figura 5.3: O mapa acima apresenta os mesmos dados representados na Figura 5.1. Nesta imagem estão desligadas a imagem do fundo e a camada relativa ao Tema 1.

bida pela candidata vencedora, Dilma Rousseff. O Tema 2 representa a quantidade de votos recebidos pelo candidato derrotado, Aécio Neves.

As Figuras 5.8 e 5.9 mostram o mapa modificado, separando os dois temas e deixando mais claro em que áreas se concentraram os votos para cada candidato. Dilma Rousseff se destaca nas Regiões Norte e Nordeste do país. Já Aécio Neves obteve maiorias expressivas nas Regiões Sudeste, Sul e Centro-Oeste.

O mapa apresentado na Figura 5.10 mostra somente as áreas onde cada candidato obteve maioria dos votos válidos. Os dois seguintes (Figuras 5.11 e 5.12) mostram os dos temas isoladamente, mostrando as regiões onde cada candidato venceu seu adversário.

As Figuras 5.13 e 5.14 são semelhantes e mostram dois mapas contendo apenas um tema cada, que relaciona os votos dos dois candidatos e apresenta as áreas onde cada um deles venceu ou foi derrotado no pleito.

O mapa apresentado na Figura 5.15 teve a escala modificada para mostrar como se distribuíram os votos onde cada candidato obteve vitória.

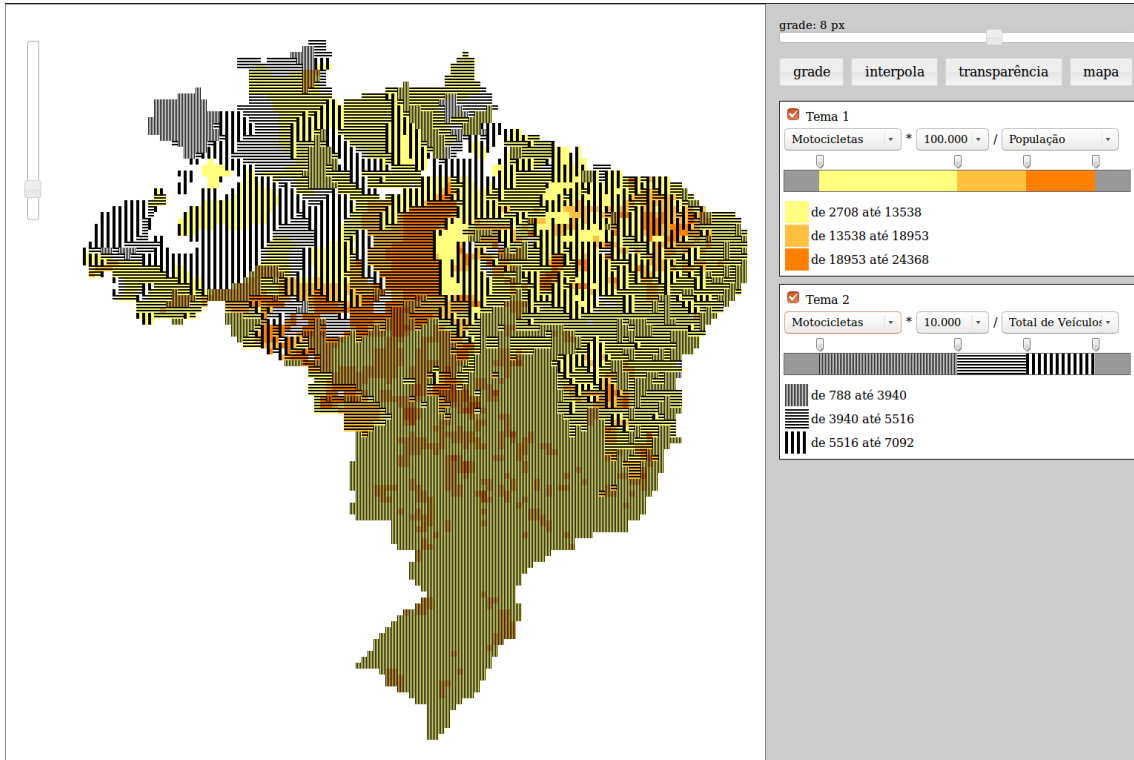


Figura 5.4: Este mapa é semelhante ao apresentado na Figura 5.1. Aqui, porém, são representadas as quantidade de motocicletas para cada 100 mil habitantes e para cada 10 mil veúlos.

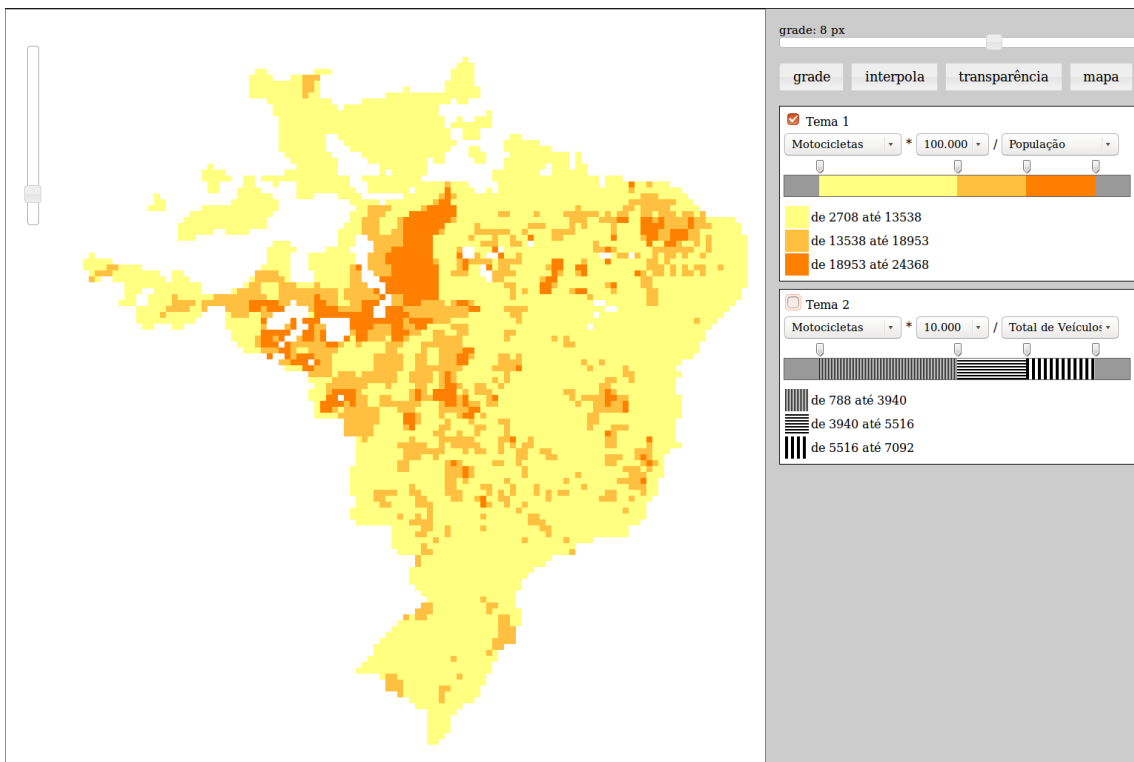


Figura 5.5: A imagem mostra o Tema 1 do mapa representado na Figura 5.4.

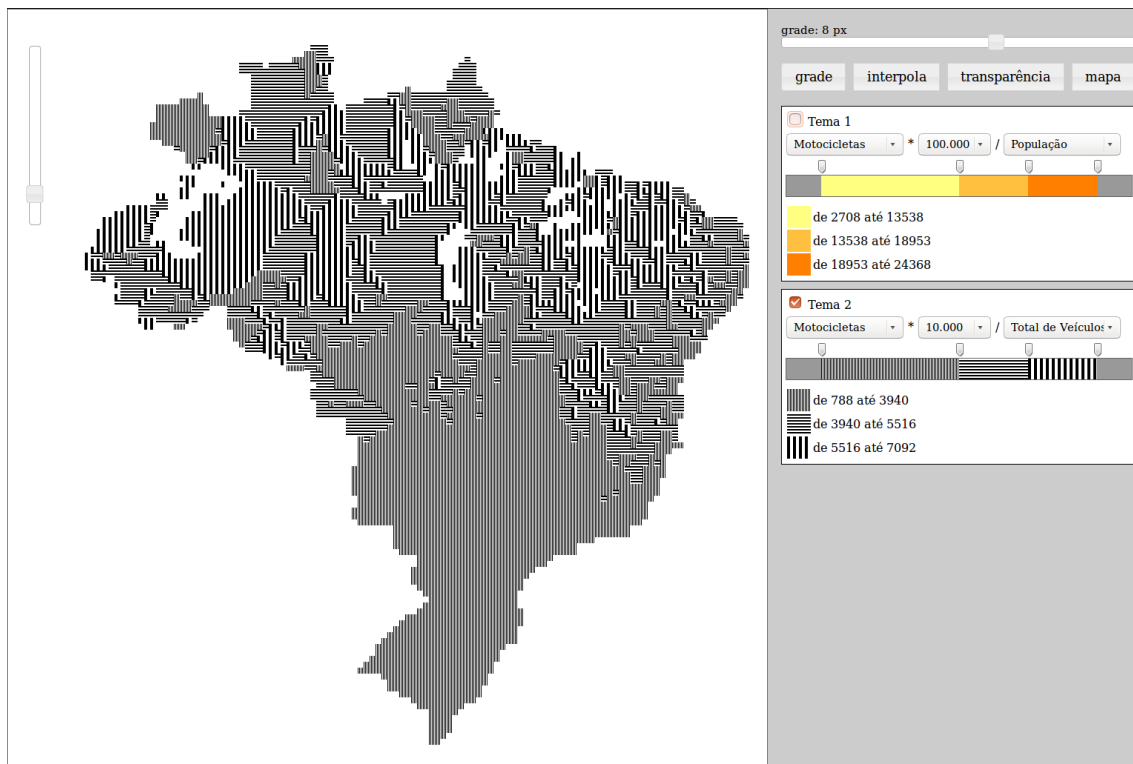


Figura 5.6: A imagem mostra o Tema 2 do mapa representado na Figura 5.4.

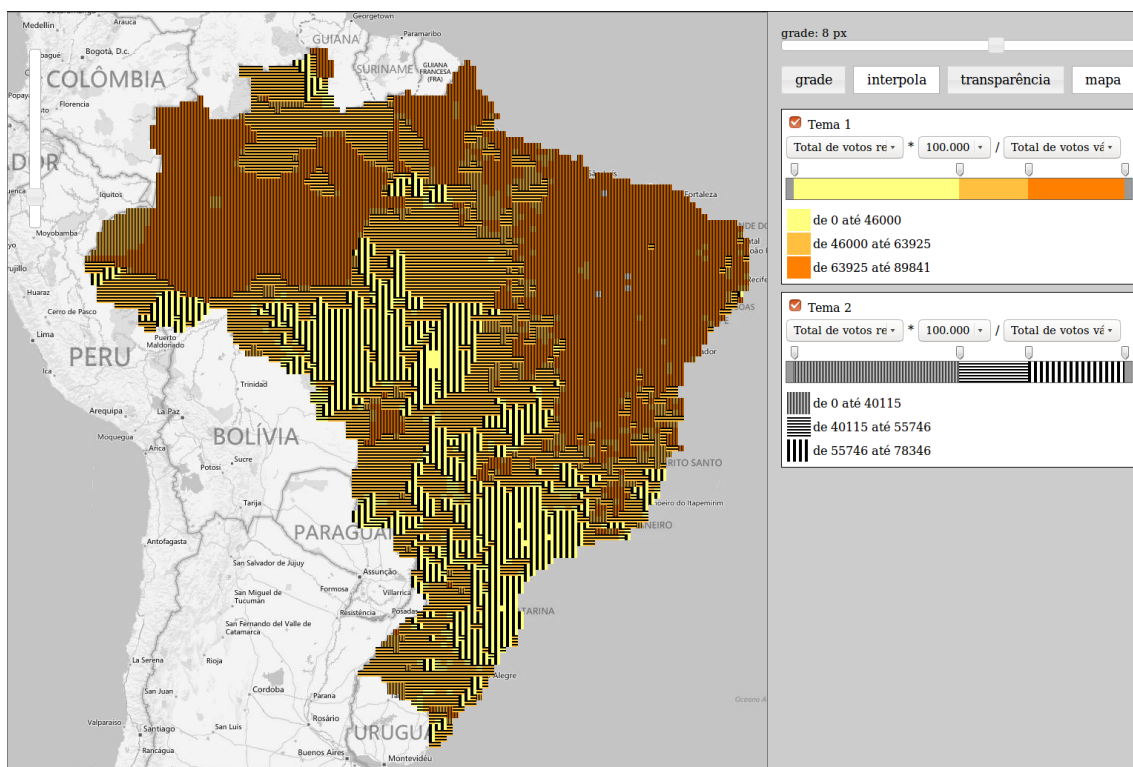


Figura 5.7: O mapa acima mostra a situação no segundo turno das eleições presidenciais do Brasil em 2014.

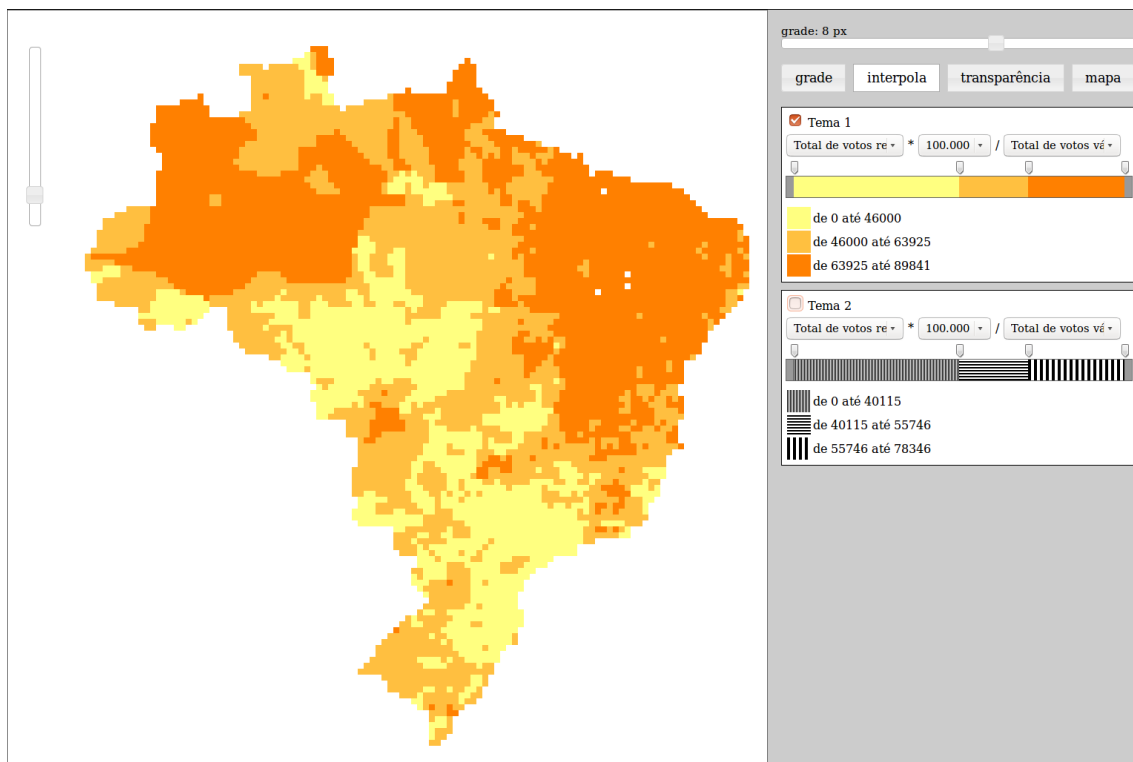


Figura 5.8: A imagem mostra o Tema 1 do mapa apresentado na Figura 5.7.

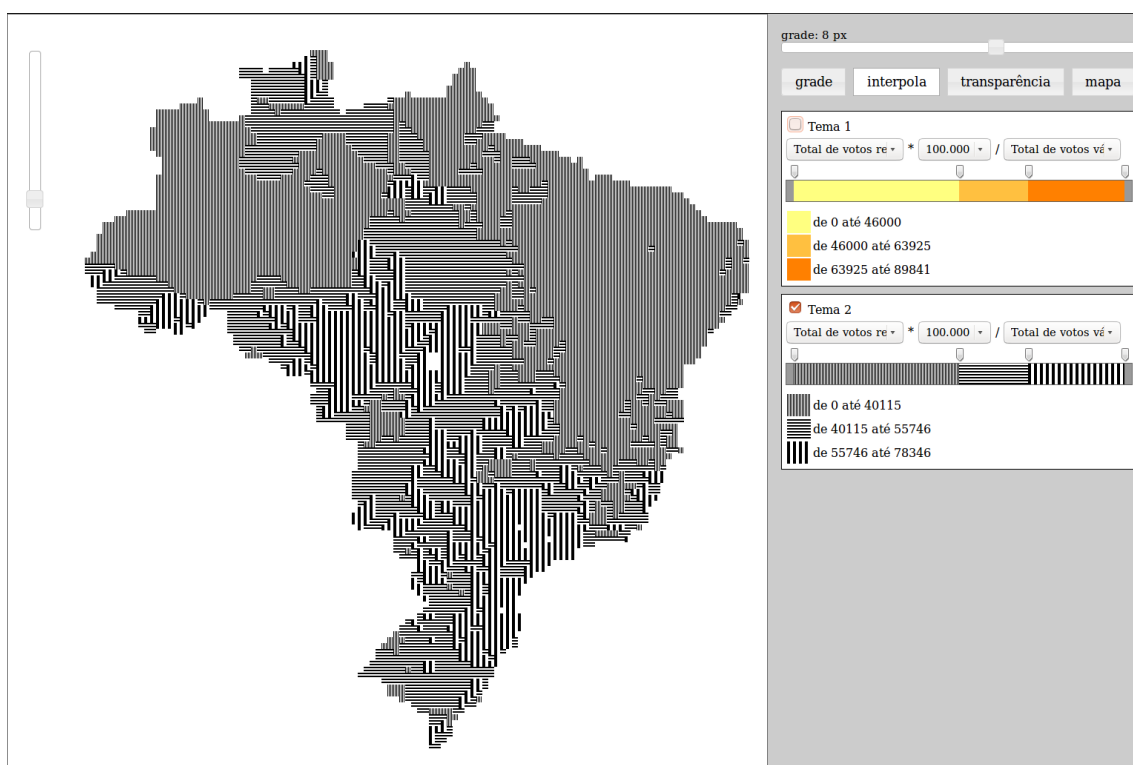


Figura 5.9: A imagem mostra o Tema 2 do mapa apresentado na Figura 5.7.

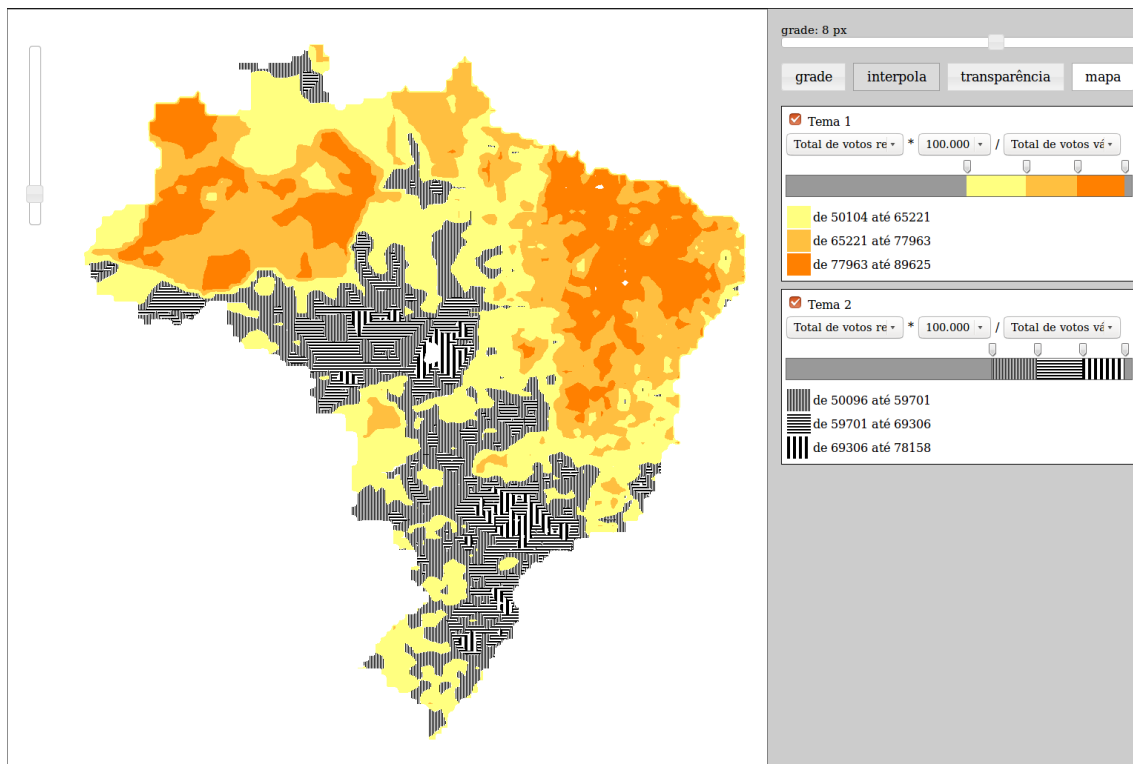


Figura 5.10: O mapa mostra somente as áreas onde cada candidato recebeu mais da metade dos votos válidos.

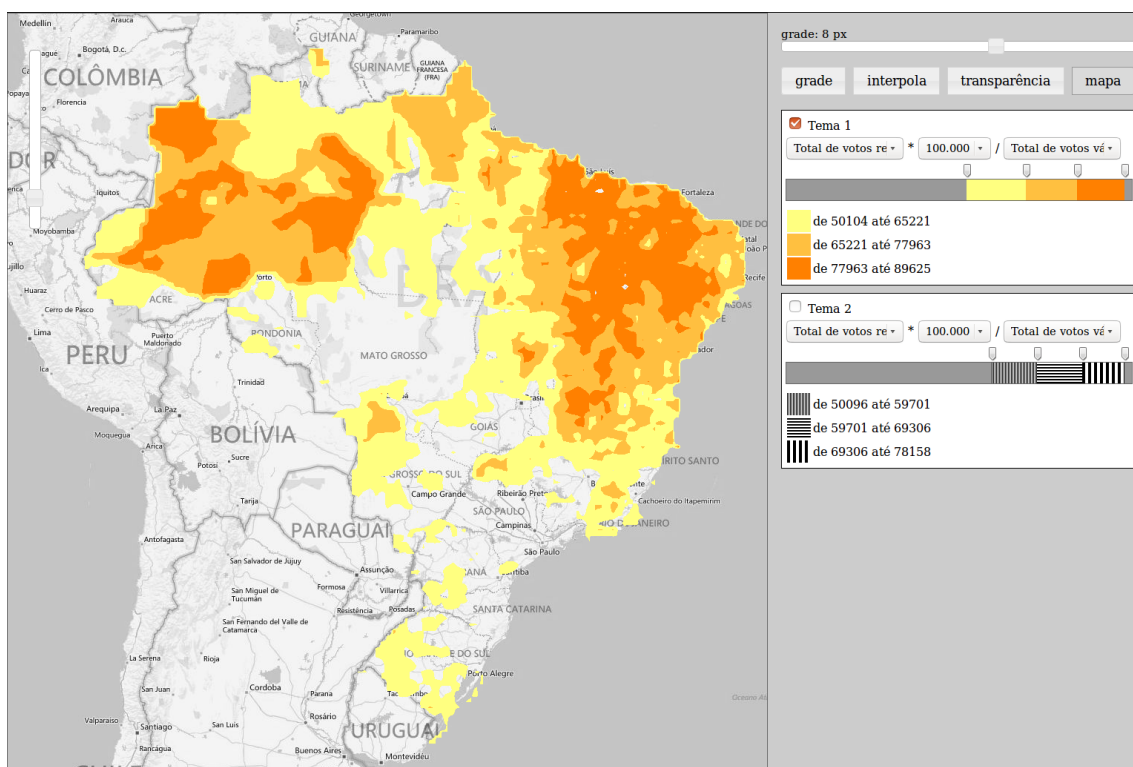


Figura 5.11: Este mapa mostra o Tema 1 do mapa representado na Figura 5.10.

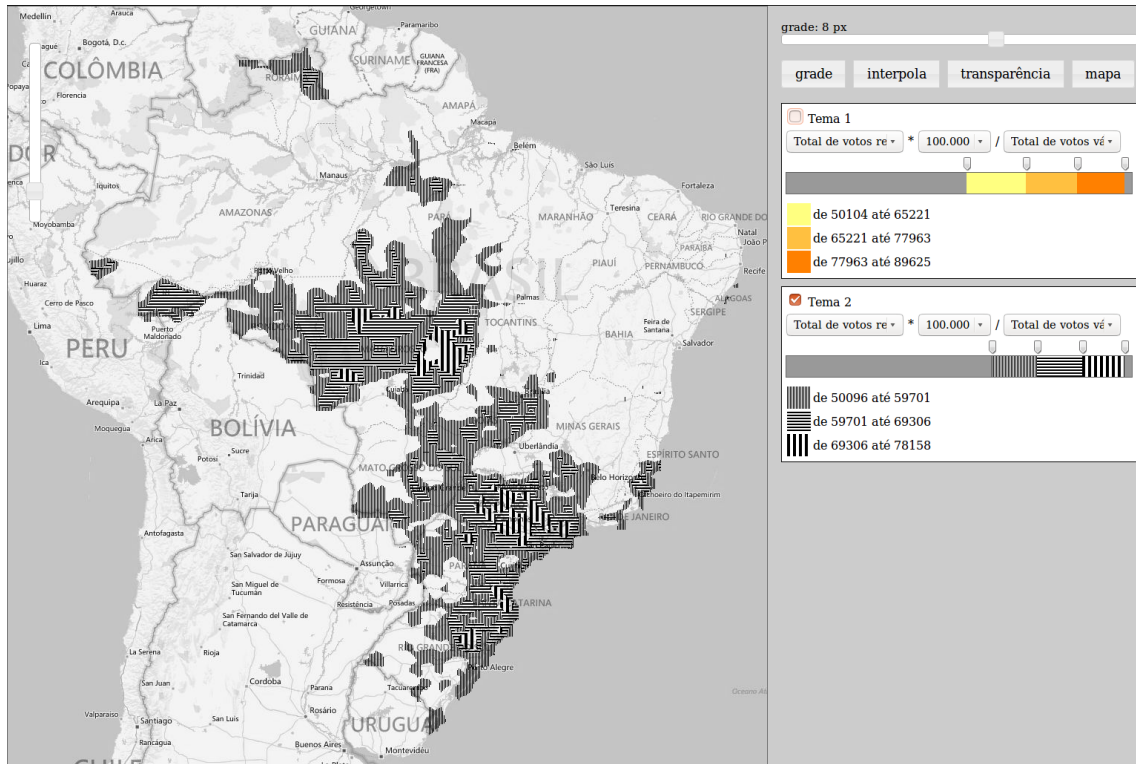


Figura 5.12: Este mapa mostra o Tema 2 do mapa representado na Figura 5.10.

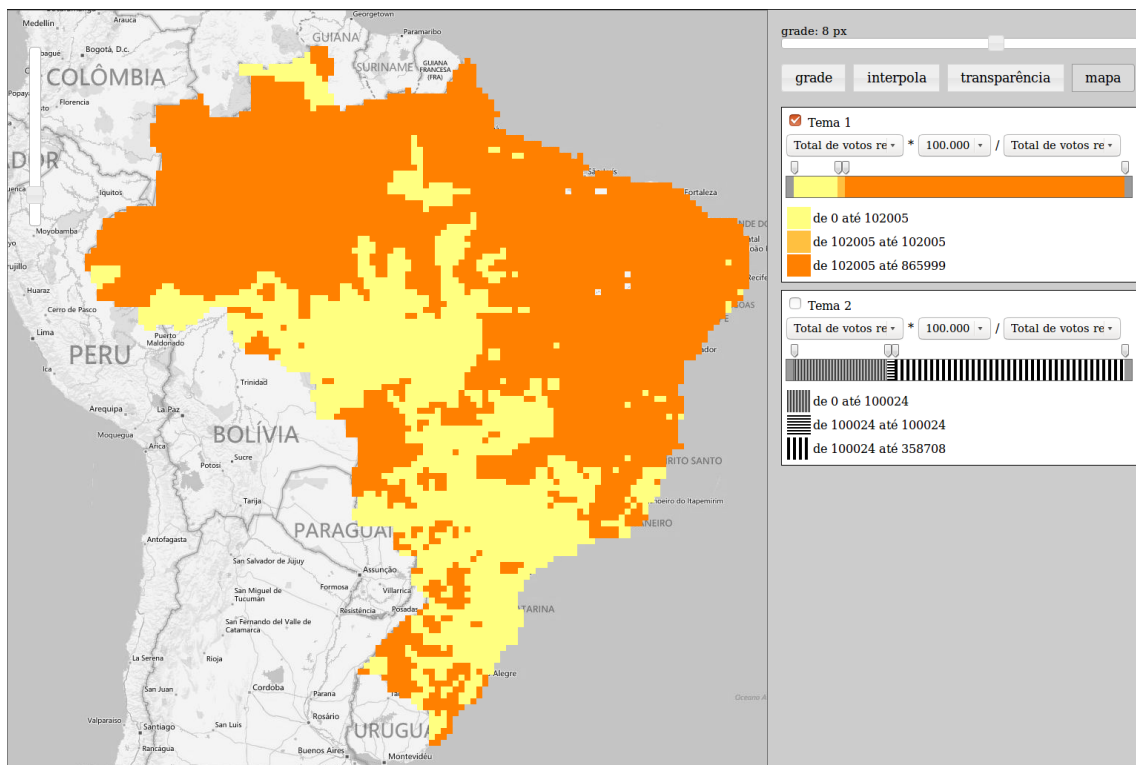


Figura 5.13: O mapa mostra a quantidade de votos recebidos por Dilma Rousseff para cada 100 mil votos recebidos por Aécio Neves. As classes foram organizadas para mostrar onde o candidato venceu ou perdeu a disputa.

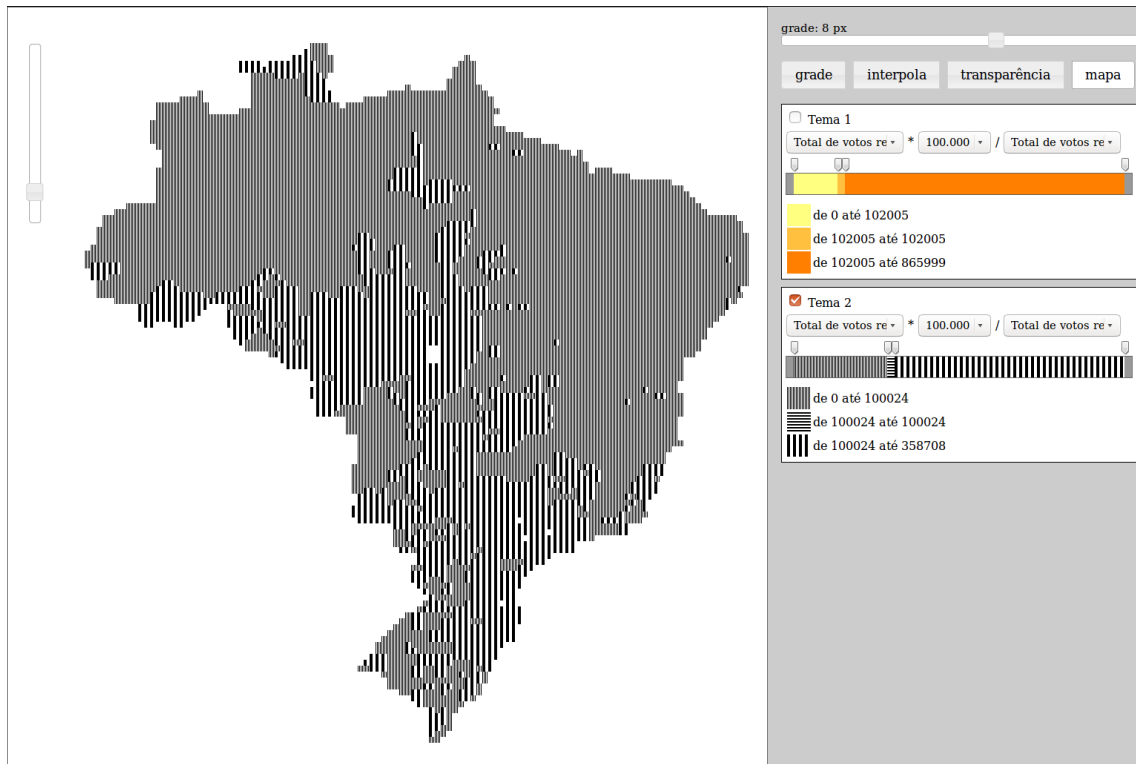


Figura 5.14: O mapa mostra a quantidade de votos recebidos por Aécio Neves para cada 100 mil votos recebidos por Dilma Rousseff. As classes foram organizadas para mostrar onde o candidato venceu ou perdeu a disputa.

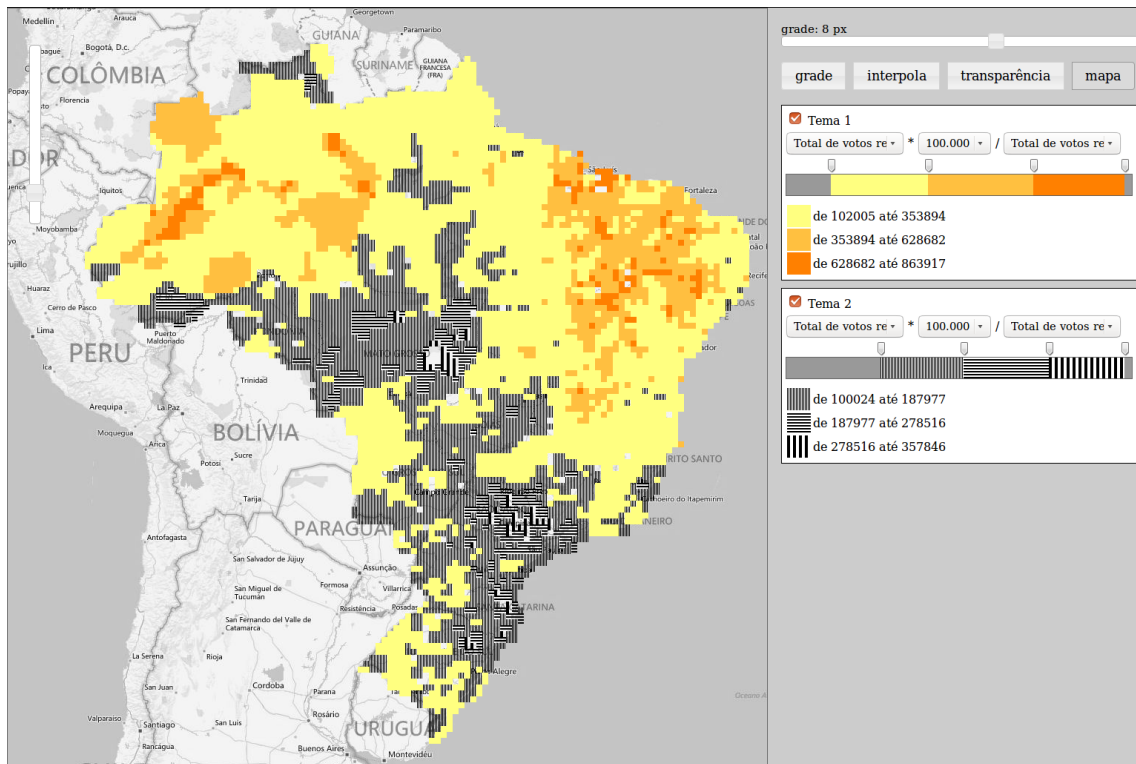


Figura 5.15: O mapa apresenta no Tema 1 a quantidade de votos recebida por Dilma Rousseff para cada 100 mil votos recebidos por Aécio Neves. No Tema 2 é apresentada a situação inversa.

Capítulo 6

Conclusões e Possíveis Desdobramentos

A visualização de mais de uma variável simultaneamente sobre um mapa, que era o objetivo inicial da pesquisa, se mostrou um desafio difícil, pois, por mais que se tente expor diversas informações sobre um mesmo ponto, acaba-se caindo nas limitações encontradas na literatura acerca da visibilidade dos níveis médio e de conjunto, onde ou se atende à pergunta “o que há nesse lugar?” ou se responde às questões do tipo “esse caractere, onde está?”. Realizar essa simultaneidade é uma tarefa complexa. A interação se apresenta como um poderoso aliado nessa empreitada, facilitando as transições entre as ações de experimentação da ocupação do espaço por cada uma das variáveis em separado e a visualização da movimentação do conjunto. Mas a verdade é que possivelmente o melhor resultado da presente experiência foi o sistema em si, que exigiu a elaboração de uma técnica específica que abre espaço para novas ideias.

A técnica desenvolvida para o experimento se mostrou muito promissora e abre diversas portas. O sistema funcionou bem, atingindo velocidade suficiente de renderização para uma boa interação em diversos dispositivos, inclusive nos móveis, onde a capacidade de processamento é limitada, mesmo considerando as diversas etapas de *shaders* necessários para o funcionamento do conjunto.

Seria interessante experimentar a ferramenta para a visualização de subdivisões políticas mais desafiadoras, como a dos setores censitários, cuja geometria complexa dificulta muito a implementação de sistemas baseados na renderização de arquivos vetoriais. No sistema aqui proposto, é possível, por exemplo, aproveitar essas mesmas imagens para apresentar dados referentes não só ao seu nível geográfico, mas também aos níveis superiores, como subdistritos, distritos, municípios e outros nessa ordem, uma vez que seu código de área já identifica sua paternidade. Como o sistema se baseia integralmente no uso de imagens rasterizadas, a maior complexidade das geometrias não é um problema para ele, o que se mostra uma grande vantagem

sobre as técnicas que utilizam a renderização em tempo de execução de arquivos vetoriais.

A inspiração no método da visualização por grade estatística também se mostrou bastante promissora. Há que se levantar o problema das distorções promovidas pela projeção geográfica utilizada no mapa, que amostra os dados igualmente pelos *pixels* da área. Isso gera uma distorção, uma vez que a projeção WebMercator gera *pixels* que representam áreas diferentes, diminuindo conforme a latitude. Deve-se pensar inclusive na aplicação de algum método que considere áreas de ocupação improvável, como florestas, descampados e espelhos d'água, para estimar melhor a amostragem dos dados. Porém, para efeitos práticos do mecanismo de visualização, a decomposição dos dados pelos *pixels* em tempo real se mostrou bastante satisfatória, possibilitando imaginar desdobramentos do trabalho para explorar idéias mais avançadas, como a renderização interativa do método proposto por TATEOSIAN *et al.* (2007) ou até em uma visualização em terceira dimensão, utilizando o mapa em perspectiva para visualizar as alturas relativas das células como elemento de transcrição visual das informações.

O sistema permite vislumbrar também a implementação de uma amostragem por células de formatos diferentes, ou até mesmo irregulares, como os *honeycombs* e as decomposições espaciais de Voronoi, com os pontos geradores baseados, por exemplo, nos centróides das subdivisões convexas dos polígonos que definem as regiões.

Uma questão a ser amadurecida diz respeito a como lidar com a exibição de resultados não-inteiros. Hoje o sistema trabalha somente com dados arredondados. Esse problema poderia ser resolvido, por exemplo, com texturas de tamanho maior, com células dedicadas para a guarda dos valores decimais.

No que tange aos dispositivos móveis, a usabilidade do aplicativo precisa ser revista, já que está totalmente voltada para os modelos de interação tradicionais, orientados pelo clique do mouse e pela precisão de um *pixel*. A movimentação do mapa não funciona fora desse ambiente, pois depende do arrasto do mouse, que conflita com a interação padrão implementada pelos navegadores para acionarem a rolagem das páginas em dispositivos móveis. Problema semelhante ocorre com a manipulação das classes através das barras deslizantes, que não funcionaram bem nos dispositivos móveis onde foram testados.

Outra questão séria que precisa ser revista é a composição das imagens que contêm as informações de codificação das áreas. É preciso que haja pelo menos um *pixel* para que cada área sobreviva à rasterização. Pesquisar um algoritmo que resolva bem o problema de eventuais desaparecimentos torna-se necessário.

Por conta da amostragem pela grade, especialmente quando acionada a visualização interpolada, as fronteiras externas do mapa literalmente transbordam para além dos limites informados pela imagem de referência embaixo. É preciso pensar

em como resolver esse problema de visualização. Talvez diluir os dados no “vazio” seja uma alternativa, mas é preciso testar.

No geral, portanto, os resultados foram bastante positivos. Partiu-se da experimentação de uma técnica simples, baseada em imagens rasterizadas, promovendo interatividade através da identificação das localidades a partir dos *pixels*. Trata-se de uma forma original de encarar o problema e que revelou uma boa aplicabilidade para a visualização.

Há ainda muito que pode ser feito no campo da visualização multivariada georreferenciada. Este trabalho não pretende fechar a questão; pelo contrário, espera-se, com este estudo, que as técnicas aqui discutidas possam ser aperfeiçoadas para a solução de problemas na área e em áreas afins e que as eventuais lacunas possam ser preenchidas em futuras pesquisas.

Referências Bibliográficas

- “QGIS - A Free and Open Source Geographic Information System”. Disponível em: <<https://www.qgis.org/en/site/index.html>>. Acessado em: 12/06/2016.
- AITCHISON, A., 2011. “The Google Maps / Bing Maps Spherical Mercator Projection”. Disponível em: <<https://alastaira.wordpress.com/2011/01/23/the-google-maps-bing-maps-spherical-mercator-projection/>>. Acessado em: 10/08/2015.
- ANDRIENKO, G., 2001, “Choropleth Maps: Classification Revisited”, Disponível em: <<http://geoanalytics.net/and/papers/ica01.pdf>>. Acessado em: 03/07/2016.
- ARCHELA, R. S., 1999, “Imagem e representação gráfica”, *Geografia (Londrina)*, v. 8-1, pp. 5–11. Disponível em: <<http://www.uel.br/revistas/uel/index.php/geografia/article/view/10198/9004>>. Acessado em: 03/07/2016.
- ARCHELA, R. S., 2001, “Contribuições da Semiologia Gráfica para a Cartografia Brasileira”, *Geografia*, v. 10, n. 1, pp. 45–50.
- ARCHELA, R. S., THÉRY, H., 2008, “Orientação metodológica para construção e leitura de mapas temáticos”, *Confins*, v. 3, n. 3. ISSN: 1958-9212. doi: 10.4000/confins.3483.
- BERTIN, J., 1980, “O Teste de Base da Representação Gráfica”, *Revista Brasileira de Geografia*, pp. 160–182. Disponível em: <http://biblioteca.ibge.gov.br/visualizacao/monografias/GEBIS-RJ/RBG/RBG1980v42_n1.pdf>. Acessado em: 03/07/2016.
- BOSTOCK, M., 2012. “Let’s Make a Map”. Disponível em: <<http://bost.ocks.org/mike/map/>>. Acessado em: 22/12/2015.

- BOSTOCK, M., 2015a. “TopoJSON”. a. Disponível em: <<https://github.com/mbostock/topojson/wiki>>. Acessado em: 12/06/2016.
- BOSTOCK, M., 2015b. “D3 - Data Driven Documents”. b. Disponível em: <<https://d3js.org/>>. Acessado em: 12/06/2016.
- BREWER, C., 1994, “Color Use Guidelines for Mapping and Visualization”, *Visualization in Modern Cartography*, v. 3, n. Dx, pp. 123–147.
- BREWER, C., HARROWER, M., 2016. “Color Brewer 2.0 - Color advice for cartography”. Disponível em: <<http://colorbrewer2.org/>>. Acessado em: 04/08/2015.
- BUENO, M. D. C. D., D’ANTONA, A. D. O., 2014, “Avaliação de métodos de desagregação para geração de grades de população”, *Revista Espinhaço*, v. 3, n. 1, pp. 127–137. Disponível em: <<http://www.cantacantos.com.br/revista/index.php/espinhaco/issue/view/22/showToc>>. Acessado em: 03/07/2016.
- BUTLER, H., DALY, M., DOYLE, A., et al., 2008. “The GeoJSON Format Specification”. Disponível em: <<http://geojson.org/geojson-spec.html>>. Acessado em: 12/06/2016.
- CABLE, D., 2013. “The Racial Dot Map: One Dot Per Person for the Entire United States”. Disponível em: <<http://www.coopercenter.org/demographics/Racial-Dot-Map>>. Acessado em: 13/12/2015.
- CHERNOFF, H., 1973, “The use of faces to represent points in k-dimensional space graphically.” *Journal of the American Statistical Association*, v. 68, n. 342, pp. 361–368. ISSN: 0162-1459. doi: 10.1080/01621459.1973.10482434.
- DEHAAN, J., MAYHEW, J., 2011. “Animation Learning Guide for Flash: Frame rates”. Disponível em: <http://www.adobe.com/devnet/flash/learning_guide/animation/part02.html>. Acessado em: 13/01/2016.
- DIAS, M. P., CARVALHO, J. O. F., 2007, “A Visualização da Informação e a sua contribuição para a Ciência da Informação”, v. 8, n. 5 (10). Disponível em: <http://www.datagramazero.org.br/out07/Art_02.htm>. Acessado em: 15/01/2016.
- DOUGLAS, D., PEUCKER, T., 2011, *Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line Or Its Caricature*. Canadian Cartographer.

- FEW, S., 2013. “Data Sensemaking: An Interaction of Eyes and Mind”. Disponível em: <<https://business.uc.edu/content/dam/business/centers/analytics/docs/Few-UniversityofCincinnatiPresentation.pdf>>. Acessado em: 03/07/2016.
- GERSHON, N., EICK, S. G., CARD, S., 1998, “Information Visualization”, *Interactions*, v. 5, n. 2, pp. 9–15. ISSN: 10725520. doi: 10.1145/274430.274432.
- HEILMANN, R., KEIM, D., PANSE, C., et al., 2004, “RecMap: Rectangular Map Approximations”, *IEEE Symposium on Information Visualization*, pp. 33–40. ISSN: 1522-404X. doi: 10.1109/INFVIS.2004.57.
- HUFFMAN, D., 2010. “On the Abuse of Chernoff Faces”. Disponível em: <<https://cartastrophe.wordpress.com/2010/06/16/on-the-abuse-of-chernoff-faces/>>. Acessado em: 02/06/2016.
- IBGE, 2008. “Características Étnico-Raciais da População - um estudo das categorias de classificação de cor ou raça 2008”. Disponível em: <http://www.ibge.gov.br/home/estatistica/populacao/caracteristicas_raciais/default_raciais.shtm>. Acessado em: 14/05/2016.
- IBGE, 2015. “Grade Estatística - Guia de Utilização”. Disponível em: <ftp://geoftp.ibge.gov.br/malhas_digitais/censo_2010/grade_estatistica/ge_guiia_utilizacao.pdf>. Acessado em: 29/11/2015.
- IBGE, 2016a. “Cidades@ - Minas Gerais - Santa Cruz de Minas”. a. Disponível em: <<http://cidades.ibge.gov.br/xtras/perfil.php?lang=&codmun=315733>>. Acessado em: 26/08/2015.
- IBGE, a. “Cidades@”. a. Disponível em: <<http://cidades.ibge.gov.br/xtras/home.php>>. Acessado em: 22/12/2015.
- IBGE, b. “Malhas Digitais - Municípios 2013”. b. Disponível em: <ftp://geoftp.ibge.gov.br/malhas_digitais/municipio_2013/>. Acessado em: 16/11/2015.
- IBGE, 2010. “Evolução da Divisão Territorial do Brasil 1872 - 2010”. Disponível em: <ftp://geoftp.ibge.gov.br/organizacao_territorial/divisao_territorial/evolucao_da_divisao_territorial_do_brasil_1872_2010/evolucao_da_divisao_territorial_mapas.pdf>. Acessado em: 27/12/2015.

- IBGE, c. “A Instituição”. c. Disponível em: <<http://www.ibge.gov.br/home/disseminacao/eventos/missao/>>. Acessado em: 03/07/2016.
- IBGE, 2016b. “Instituto Brasileiro de Geografia e Estatística”. b. Disponível em: <<http://www.ibge.gov.br>>. Acessado em: 02/07/2015.
- IBGE, 2011. “Censo 2010 - Sinopse por Setores”. Disponível em: <<http://www.censo2010.ibge.gov.br/sinopseporsetores/?nivel=st>>. Acessado em: 11/06/2016.
- IBGE, d. “WebCart”. d. Disponível em: <<http://www.ibge.gov.br/webcart/>>. Acessado em: 03/07/2016.
- INTERRANTE, V., 2000, “Harnessing Natural Textures for Multivariate Visualization”, *IEEE Computer Graphics and Applications*, v. 20-6, pp. 6–11. ISSN: 0029-6570. Disponível em: <[http://www-users.cs.umn.edu/~sim\\$interran/papers/visview00.pdf](http://www-users.cs.umn.edu/~sim$interran/papers/visview00.pdf)>. Acessado em: 03/07/2016.
- KOSARA, R., 2007. “A Critique of Chernoff Faces”. Disponível em: <<https://eagereyes.org/criticism/chernoff-faces>>. Acessado em: 29/06/2015.
- MACHADO, J. D., 2011. “O que são Frames por Segundo?” Disponível em: <<http://www.tecmundo.com.br/video/10926-o-que-sao-frames-por-segundo-.htm>>. Acessado em: 13/01/2016.
- NEWMAN, M., 2012. “Maps of the 2012 US presidential election results”. Disponível em: <[http://www-personal.umich.edu/~sim\\$mejn/election/2012/](http://www-personal.umich.edu/~sim$mejn/election/2012/)>. Acessado em: 21/07/2015.
- PEDROSA, I., 2003, *O universo da cor*. SENAC Nacional. ISBN: 9788574581262.
- PEDROSA, I., 2009, *Da cor à cor inexistente*. Senac Nacional. ISBN: 9788574582672.
- PICKETT, R., GRINSTEIN, G., 1988, “Iconographic Displays For Visualizing Multidimensional Data”, v. 1. doi: 10.1109/ICSMC.1988.754351.
- QUEIROZ, D., 2011. “A semiologia e a cartografia temática”. Disponível em: <<http://eduem.uem.br/ojs/index.php/BolGeogr/article/view/12866>>. Acessado em: 03/07/2016.
- RODRIGUEZ, A. C. M., 2010, *A Cartografia temática digital do estado de São Paulo (SP)*. Tese de Doutorado, Universidade de São Paulo.

- SCHWARTZ, J., 2016. “Bing Maps Tile System”. Disponível em: <<https://msdn.microsoft.com/en-us/library/bb259689.aspx>>. Acessado em: 12/08/2015.
- SHARKD, 2010. “HSV color solid cone chroma gray”. Disponível em: <https://commons.wikimedia.org/wiki/File:HSV_color_solid_cone_chroma_gray.png>. Acessado em: 11/06/2016.
- SHNEIDERMAN, B., 1983, “Direct Manipulation: a Step Beyond Programming Languages.” *Computer*, v. 16, n. 8, pp. 57–69. ISSN: 00189162. doi: 10.1109/MC.1983.1654471.
- SHNEIDERMAN, B., 1994, “Dynamic queries for visual information seeking”, *IEEE Software*, v. 11, n. 6, pp. 70–77. ISSN: 07407459. doi: 10.1109/52.329404.
- STONE, M., 2006, “Choosing Colors for Data Visualization”, *Perceptual Edge*, pp. 1–10. Disponível em: <http://www.perceptualedge.com/articles/b-eye/choosing_colors.pdf>. Acessado em: 03/07/2016.
- TATEOSIAN, L. G., HEALEY, C. G., ENNS, J. T., 2007, “Engaging Viewers Through Nonphotorealistic Visualizations”, *Proceedings of the Fifth International Symposium on Non-Photorealistic Animation and Rendering*, pp. 93–102. doi: <http://doi.acm.org/10.1145/1274871.1274886>.
- THE JQUERY FOUNDATION, 2016. “jQuery, Write Less Do More”. Disponível em: <<https://jquery.com/>>. Acessado em: 12/06/2016.
- TRYGSTAD, R., 2007. “Blue Yellow Red Color Wheel”. Disponível em: <https://commons.wikimedia.org/wiki/File:BYR_color_wheel.svg>. Acessado em: 11/06/2016.
- TUFTE, E., 1994, *Envisioning Information*. Graphics Press. ISBN: 9781930824140.
- TUFTE, E., 2001, *The Visual Display of Quantitative Information*, v. 4. ISBN: 0961392142.
- VIANA, R., APOLINÁRIO, J. P., GAMA, T., 2015. “Mapa Racial do Brasil”. Disponível em: <<http://patadata.org/maparacial/>>. Acessado em: 10/12/2015.
- WARE, C., 2004, *Information visualization: perception for design*. ISBN: 1558608192. doi: 10.1016/B978-0-12-381464-7.00018-1.

WELDON COOPER CENTER FOR PUBLIC SERVICE, 2010. “2010 Census Block Data”. Disponível em: <<http://demographics.coopercenter.org/DotMap/>>. Acessado em: 10/12/2015.

WIKIMEDIA COMMONS, THE FREE MEDIA REPOSITORY, 2010. “Chernoff faces for evaluations of US judges”. Disponível em: <https://commons.wikimedia.org/wiki/File:Chernoff_faces_for_evaluations_of_US_judges.svg>. Acessado em: 11/06/2016.

Apêndice A

Levantamento de Informações sobre os Arquivos Contendo a Geometria dos Mapas

Com o objetivo de avaliar as condições dos arquivos usados no sistema, foram levantadas algumas informações. Segue abaixo uma pequena descrição do que foi feito, acompanhada de uma série de tabelas contendo dados a respeito de tamanho, quantidade de polígonos, vértices entre outros.

Todas as imagens geradas para o protótipo tiveram como origem os arquivos referentes à Malha Municipal 2013, que representa as divisões político-administrativas dos municípios do país na data de 01/07/2013, divulgada pelo IBGE.

De lá foram baixados os arquivos em formato ShapeFile (SHP), que, em conjunto com um grupo de arquivos associados, nos formatos CPG, DBF, PRJ e SHX, compõem um conjunto adequado para edição e visualização de mapas no software QGis Desktop 2.8.1 (QGi). Através desse programa, os arquivos foram abertos, conferidos, editados e exportados para os formatos KML, GeoJSON e TopoJSON, que constituíram os insumos utilizados nos experimentos realizados.

As malhas originais são bastante detalhadas, possuindo coordenadas compostas por muitos algarismos, o que resulta em arquivos complexos e pesados para o desenvolvimento de aplicações para a Internet. A Tabela A.1 apresenta o problema. Os arquivos originais estão divididos por Unidade Federativa.

Constam na tabela o total do país e os valores de cada uma das Unidades Federativas. Quase todas possuem mais polígonos do que cidades, o que significa que há regiões descontínuas, ilhas em sua maior parte. Na tabela estão registradas também as quantidades médias de vértices por polígono. No Brasil, o arquivo registra as 5572 áreas em 6162 polígonos que somam um total de 7490899 vértices. Isso significa que é estimada uma quantidade média de 1216 vértices para cada polígono. A Tabela

Tabela A.1: Quantidade de polígonos por Unidades Federativas

UF	Municípios	Polígonos	Vértices	Média
11 - RO	52	52	54266	1044
12 - AC	22	22	96342	4379
13 - AM	62	62	50396	813
14 - RR	15	15	25836	1722
15 - PA	144	161	131151	815
16 - AP	16	17	53520	3148
17 - TO	139	139	126040	907
21 - MA	217	282	150540	534
22 - PI	224	224	57478	257
23 - CE	184	185	91935	497
24 - RN	167	167	25477	153
25 - PB	223	225	45928	204
26 - PE	185	185	67332	364
27 - AL	102	102	34826	341
28 - SE	75	79	34628	438
29 - BA	417	424	889727	2098
31 - MG	853	853	2786687	3267
32 - ES	78	103	241256	2342
33 - RJ	92	379	226950	599
35 - SP	645	702	819659	1168
41 - PR	399	399	310020	777
42 - SC	295	416	337546	811
43 - RS	499	499	190611	382
50 - MS	79	79	104026	1317
51 - MT	141	142	289210	2037
52 - GO	246	248	237100	956
53 - DF	1	1	12412	12412
Brasil (total)	5572	6162	7490899	1216

A.2 lista a ordem decrescente dos 64 municípios cujo desenho é composto por mais de um polígono.

Tabela A.2: Municípios com áreas descontínuas

Início da Tabela A.2			
Município	Polígonos	Vértices	Média
3300100 - ANGRA DOS REIS	106	8063	76
3304557 - RIO DE JANEIRO	102	54082	530
4205407 - FLORIANÓPOLIS	43	7209	168
3303807 - PARATY	37	4526	122
2103703 - CURURUPU	23	3506	152
4216206 - SÃO FRANCISCO DO SUL	22	1502	68
3302601 - MANGARATIBA	21	1623	77

Continuação da Tabela A.2

Município	Polígonos	Vértices	Média
3555406 - UBATUBA	19	5406	285
4202453 - BOMBINHAS	17	13016	766
3205200 - VILA VELHA	14	2701	193
3518701 - GUARUJÁ	13	2761	212
3520400 - ILHABELA	11	5458	496
4206009 - GOVERNADOR CELSO RAMOS	11	5723	520
2104305 - GODOFREDO VIANA	10	1451	145
3550704 - SÃO SEBASTIÃO	10	19298	1930
2100832 - APICUM-AÇU	9	1741	193
3202405 - GUARAPARI	9	2570	286
3303302 - NITERÓI	9	2994	333
4202008 - BALNEÁRIO CAMBORIÚ	9	1099	122
2102606 - CÂNDIDO MENDES	8	1882	235
2106201 - LUÍS DOMINGUES	8	744	93
2906907 - CARAVELAS	8	2483	310
3302007 - ITAGUAÍ	7	1190	170
1500909 - AUGUSTO CORRÊA	6	515	86
2109056 - PORTO RICO DO MARANHÃO	6	1231	205
4207304 - IMBITUBA	6	1529	255
1506112 - QUATIPURU	5	387	77
1508308 - VISEU	5	1319	264
2800407 - ARAUÁ	5	360	72
3205309 - VITÓRIA	5	1344	269
3304904 - SÃO GONÇALO	5	1972	394
4205704 - GAROPABA	5	4688	938
4209409 - LAGUNA	5	1393	279
3300233 - ARMAÇÃO DOS BÚZIOS	4	8113	2028
3510500 - CARAGUATATUBA	4	5126	1282
3537602 - PERUÍBE	4	3316	829
4212502 - PENHA	4	713	178
1501709 - BRAGANÇA	3	596	199
2100204 - ALCÂNTARA	3	692	231
2104909 - GUIMARÃES	3	508	169
3300258 - ARRAIAL DO CABO	3	652	217
3302403 - MACAÉ	3	1647	549
4208450 - ITAPOÁ	3	934	311
1502905 - CURUÇÁ	2	359	180

Continuação da Tabela A.2

Município	Polígonos	Vértices	Média
1507805 - SENADOR JOSÉ PORFÍRIO	2	1982	991
1600105 - AMAPÁ	2	3142	1571
2102903 - CARUTAPERA	2	844	422
2109452 - RAPOSA	2	425	213
2112407 - TURIAÇU	2	2247	1124
2302107 - BATURITÉ	2	466	233
2503209 - CABEDELO	2	337	169
2508901 - MAMANGUAPE	2	378	189
3506359 - BERTIOGA	2	3234	1617
3509908 - CANANÉIA	2	1416	708
4202057 - BALNEÁRIO BARRA DO SUL	2	224	112
4202107 - BARRA VELHA	2	654	327
4208203 - ITAJAÍ	2	1496	748
4211306 - NAVEGANTES	2	977	489
4212304 - PAULO LOPES	2	1385	693
4213500 - PORTO BELO	2	1539	770
5101704 - BARRA DO BUGRES	2	3821	1911
5213103 - MINEIROS	2	2958	1479
5220702 - SÍTIO D'ABADIA	2	894	447

Fim da Tabela A.2

Os números comprovam o problema: para visualizar o mapa inteiro do Brasil, é preciso desenhar sete milhões e meio de vértices. Esses arquivos precisavam de otimização.

Os arquivos KML foram gerados no software QGis a partir dos arquivos SHP. Os arquivos GeoJSON também. Os arquivos aqui denominados GeoJSON (2) são versões obtidas a partir da generalização dos arquivos GeoJSON originais, baixando a precisão das informações de localização para 4 casas decimais e eliminando os vértices coincidentes do polígono. Os arquivos TopoJSON foram gerados a partir dos arquivos GeoJSON originais, utilizando o programa Topojson (BOSTOCK, 2012). A segunda versão de arquivos, TopoJSON (2), foi gerada a partir dos arquivos TopoJSON originais, generalizados pelo algoritmo de Ramer-Douglas-Peucker, arbitrando um valor de tolerância de 0,001, o que resultou em uma imagem ruim. Não foram feitos mais testes pois, mesmo com essa demasiada tolerância, não se produziram arquivos muito menores que os TopoJSON iniciais, conforme mostra a Tabela A.3.

A Tabela A.4 apresenta os percentuais de redução obtidos nos arquivos em re-

Tabela A.3: Tamanho dos arquivos Vetoriais Gerados (kB)

UF	KML	GeoJSON	GeoJSON (2)	TopoJSON	TopoJSON (2)
11	2072	2232	1140	212	116
12	3604	3880	1972	296	72
13	1888	2028	1028	196	180
14	940	1020	504	116	68
15	4900	5296	2668	456	300
16	1936	2108	1036	176	64
17	4776	5148	2636	468	240
21	5656	6088	3056	540	244
22	2220	2344	1184	232	128
23	3472	3724	1752	320	148
24	1012	1048	528	120	64
25	1788	1876	940	200	108
26	2552	2720	1292	220	112
27	1336	1416	652	124	48
28	1348	1440	724	144	48
29	33872	36656	17588	1468	420
31	84456	94332	59128	4264	1076
32	9188	9912	4912	480	112
33	8672	9324	3976	616	116
35	28092	30604	16744	1992	568
41	11864	12716	6604	1072	436
42	12432	13420	6844	1060	276
43	7456	7900	4112	716	488
50	3984	4284	2196	384	200
51	10996	11900	6104	904	404
52	9100	9788	5008	792	384
53	472	512	220	76	8
Total (Brasil)	260084	283716	154548	17644	6428

lação ao arquivo GeoJSON original. Estão também registradas as quantidades de polígonos e vértices por Unidade Federativa, para demonstrar a relação entre tamanho dos arquivos, a quantidade de polígonos e a quantidade total de vértices.

A versão generalizada promoveu uma sensível diminuição no tamanho, especialmente nos estados onde a geometria é mais complexa. Para efeito de comparação, a Figura A.1 mostra a renderização de um detalhe da Ilha do Fundão, no Rio de Janeiro (RJ), obtida por arquivos distintos, mostrando o quanto essa generalização interfere no atendimento das necessidades de exibição.

A.1 Arquivos Rasterizados

A adoção de uma arquitetura baseada na montagem de imagens rasterizadas exigiu novas experimentações. Para tanto, foram geradas séries de imagens a partir

Tabela A.4: Redução de tamanho promovida nos arquivos

UF	Polígonos	Vértices	GeoJSON (2)	TopoJSON	TopoJSON (2)
11	52	54266	49	91	95
12	22	96342	49	92	98
13	62	50396	49	90	91
14	15	25836	51	89	93
15	161	131151	50	91	94
16	17	53520	51	92	97
17	139	126040	49	91	95
21	282	150540	50	91	96
22	224	57478	49	90	95
23	185	91935	53	91	96
24	167	25477	50	89	94
25	225	45928	50	89	94
26	185	67332	52	92	96
27	102	34826	54	91	97
28	79	34628	50	90	97
29	424	889727	52	96	99
31	853	2786687	37	95	99
32	103	241256	50	95	99
33	379	226950	57	93	99
35	702	819659	45	93	98
41	399	310020	48	92	97
42	416	337546	49	92	98
43	499	190611	48	91	94
50	79	104026	49	91	95
51	142	289210	49	92	97
52	248	237100	49	92	96
53	1	12412	57	85	98
Total (Brasil)	6162	7490899	46	94	98

dos arquivos KML originais, nos níveis de detalhe entre 4 e 10, conforme o padrão adotado para o sistema. As Tabelas A.5, A.6, A.7, A.8 e A.9 demonstram os resultados.

Os números mostram que a maneira mais econômica é carregar dois ladrilhos distintos, um para a renderização das divisões político-administrativas indexadas e outro com as informações topográficas para localização.

Considerando que o mapa seja renderizado em uma tela de tamanho 1920 x 1080 *pixels*, será necessária a carga de 40 ladrilhos, compondo uma tela com 8 colunas e 5 linhas de imagens, duplicando essa carga para o caso de se utilizar duas imagens. A Tabela A.10 apresenta estimativas para a montagem de tamanhos comuns de telas.

A opção pelo uso de duas camadas, uma contendo as imagens em formato PNG com espaço RGB com os índices das áreas e outra com as informações de base em formato JPG, se mostrou a mais adequada. Elas foram geradas a partir de um con-



Figura A.1: As imagens mostram as diferenças de detalhamento dos polígonos. A imagem acima à esquerda, mostra a renderização do arquivo KML. Ao seu lado é mostrada a imagem gerada pelo arquivo GeoJSON composta por vértices contendo a precisão de 4 casas decimais. As duas imagens abaixo são do mesmo formato, mas com 3 e 2 casas decimais respectivamente. Essas figuras são referentes a capturas de tela do software QGis.

Tabela A.5: Imagens PNG, RGB de 16 bits contendo apenas os identificadores dos municípios

LOD	Files	Size
4	9	0,06 Mb
5	25	0,14 Mb
6	72	0,30 Mb
7	255	0,71 Mb
8	990	1,8 Mb
9	3900	5,2 Mb
10	15.351	16,3 Mb
Total	20.609	24,6 Mb

O tamanho médio de um ladrilho é de 0,91 kb

junto de imagens originais do serviço Bing Maps, que eram originalmente coloridas mas foram dessaturadas para que pudessem ser aplicadas no presente experimento. A utilização das imagens originais também seria possível, uma vez que, em termos de tamanho, a diferença é pouca entre estas e as originais: em uma tela *full HD*, seriam necessários 180,3 kB para 40 ladrilhos coloridos contra 144,0 kB na versão dessaturada. Uma diferença nada desprezível de 36,3 kB. Ainda assim, a utilização de imagens em escala de cinzas para a base do mapa se apresenta como melhor solução porque facilita a visualização mais adequada das camadas coloridas acima.

Tabela A.6: Imagens JPG (Bing Maps) convertidas para escala de cinzas de 16 bits

LOD	Files	Size
4	9	0,07 Mb
5	25	0,21 Mb
6	72	0,57 Mb
7	255	1,9 Mb
8	990	5,9 Mb
9	3900	18,7 Mb
10	15.351	64,8 Mb
Total	20.609	92,9 Mb

O tamanho médio de um ladrilho é de 4,51 kb

Tabela A.7: Imagens de base (Bing Maps) convertidas para escala de cinzas

LOD	Files	Size (JPEG)	Size (PNG)
4	9	0,05 Mb	0,15 Mb
5	25	0,17 Mb	0,49 Mb
6	72	0,46 Mb	1,4 Mb
7	255	1,6 Mb	4,6 Mb
8	990	4,6 Mb	14,5 Mb
9	3900	14,4 Mb	48,3 Mb
10	15.351	52,0 Mb	188,8 Mb
Total	20.609	74,2 Mb	259,0 Mb

Tamanho médio dos ladrilhos: 3,6 kb (JPG) e 12,6 kb (PNG)

Tabela A.8: Imagens PNG contendo os identificadores de área nos canais R, G e B e a imagem de base (escala de cinzas) no canal A

LOD	Files	Size
4	9	0,28 Mb
5	25	0,86 Mb
6	72	2,4 Mb
7	255	7,6 Mb
8	990	23,4 Mb
9	3900	76,8 Mb
10	15.351	294,7 Mb
Total	20.609	407,0 Mb

O tamanho médio de um ladrilho é de 19,7 kb

Tabela A.9: Imagens PNG contendo os identificadores nos canais B e G e as informações da imagem de base (escala de cinzas) no canal R

LOD	Files	Size
4	9	0,25 Mb
5	25	0,78 Mb
6	72	2,1 Mb
7	255	6,9 Mb
8	990	21,3 Mb
9	3900	70,2 Mb
10	15.351	270,1 Mb
Total	20.609	372,8 Mb

O tamanho médio de um ladrilho é de 18,1 kb

Tabela A.10: Quantidade de imagens necessárias para cobrir a tela e a carga média esperada para o conjunto

Tela	ladrilhos	RGB + JPG	RGB	RGBA
1920x1080	40	191,8 kb	723,6 kb	789,9 kb
1366x768	24	115,1 kb	434,14 kb	474,0 kb
1024x600	15	71,9 kb	271,34 kb	296,2 kb