

Resumo

Este é o relatório final de renovação do Projeto de Pesquisa – PIBIC – Estudo sobre os Sistemas de Informação Geográfica. Na parte inicial deste relatório explicamos o que são os SIG's, baseado em estudos feitos na primeira parte do projeto e falamos da possibilidade de realizar análises de dados espaciais com essa tecnologia. Fizemos algumas considerações sobre o que seja análise espacial e logo em seguida citamos alguns exemplos de análises espacial realizadas por alguns pesquisadores, com o auxílio de SIG's, pois os mesmos se assemelham com a aplicação que trabalhamos. Descrevemos, passo a passo, como realizamos a implementação das novas variáveis; explicamos, uma a uma, as nomenclaturas que criamos para as variáveis; como também, descrevemos, passo a passo, a geração de mapas no SPRING; seguimos com exemplos de mapas possíveis de serem gerados com os dados do nosso banco por aplicações da lógica e de análise exploratória estatística, nos Sistemas SPRING e Tabwin, com o auxílio do “pacote estatístico” SPSS. Por fim, temos a descrição de como foi gerada a página na Internet para os dados deste projeto.

1 – Introdução

O Sistemas de Informação Geográfica (SIG) são ferramentas computacionais que permitem realizar análises complexas, integrando dados de diversas fontes e indexando-os espacialmente (Smith et alli., 1987). Os tipos de dados são referenciados geograficamente e associados à imagens ou mapas para compor planos de informação. Cada plano individualmente pode registrar espacialmente informações sobre uma variável, por exemplo: o número de habitantes das cidades, curvas de altitude, índices de produção agrícola, mapas de vegetação, etc. A própria imagem ou mapa passa a ser um plano de informação. Esses verdadeiros “bancos de dados estatísticos” podem compor espacialmente gráficos multi-dimensionais que não poderiam ser compostos de outra maneira.

Sobre este prisma, foi desenvolvido na primeira parte do projeto uma primeira versão de um mapa cadastral do estado da Paraíba (Teles, 1999). No mapa cadastral estão indexados dados estatísticos de população por município, renda por faixas de salário mínimo e escolaridade por grau, todos obtidos do Censo Demográfico Brasileiro de 1991 (IBGE, 1991), montado sobre o sistema SPRING (INPE, 1998). Por mapa cadastral entende-se a associação de atributos não gráficos – os dados – a um atributo gráfico – um mapa político do estado.

No entanto, alguns procedimentos mais complexos podem ser realizados, como é o caso das análises espaciais. Um dos procedimentos necessários para a realização de uma análise espacial é a distribuição geográfica dos fenômenos. Como um exemplo completo pode ser citado o caso da análise espacial feita pelo Dr. John Snow. Diante de uma epidemia de cólera e do desconhecimento do modo de contágio da doença, ele distribuiu os casos da doença sobre o mapa de Londres. Com isso ele identificou quais as áreas de maior incidência da doença e relacionou essas áreas com os poços de abastecimento de água, desativou-os e descobriu a forma de contágio que era desconhecida naquela época (Câmara, 1999). Diversas outras análises são possíveis com o auxílio de um SIG para resolver problemas em diferentes áreas do conhecimento, entre outras (Câmara, 1999):

- Epidemiologistas coletam dados sobre a ocorrência de doenças . A distribuição dos casos de uma doença forma um padrão no espaço? Existe associação com alguma fonte de poluição? Evidência de contágio?

- A polícia deseja investigar se existe algum padrão espacial na distribuição de roubos. Roubos que ocorrem em determinadas áreas têm alguma correlação com características sócio econômicas dessas área?

- Geólogos desejam estimar a extensão de um depósito mineral em uma região a partir de amostras. Pode-se usar essas amostras para construir um mapa de contaminação?

- Queremos analisar uma região geográfica para fins de zoneamento agrícola. Como escolher as variáveis explicativas (por exemplo, solo, vegetação e a geomorfologia) e determinar qual a contribuição de cada uma delas para a obtenção de um mapa resultante?

- Queremos saber, por exemplo, se numa cidade existem regiões com variações extremas de longevidade, que indicam casos críticos de inclusão ou exclusão social? Com dados sobre números de idosos, renda por faixa salarial, total de domicílios, etc. podemos obter mapas com padrão espacial. O padrão obtido será casual? O que pode está causando esse padrão? Ele muda com o tempo?

- Para auxiliar as secretarias de educação tanto estadual quanto municipal na localização de escolas. Quais os locais onde há maior índice de crianças fora da escola? A quantidade de crianças fora da escola tem alguma relação com a distância de sua casa à escola? Nestas localidades há a presença de trabalho infantil?

Neste trabalho, aprofundamos os estudos sobre os SIG's para a efetiva ampliação do mapa cadastral já existente. A ampliação do banco de dados foi realizada com o acréscimo das seguintes variáveis, área do município, densidade populacional, população por sexo, população segundo cor ou raça, situação de urbanização, condição de ocupação, perfis de idade e população de acordo com a religião.

Como já mencionamos, os SIG's são capazes de realizar procedimentos mais complexos que os normalmente realizados. As análises espaciais estão dentro dessa possibilidade. Assim, buscamos algumas bibliografias que nos deram apoio no que diz respeito à fundamentação teórica, como deram também acesso à exemplos práticos de uso dessa técnica. Esses exemplos de aplicações da técnica de análise espacial, que consultamos e descrevemos neste trabalho, serviram de modelo por se tratarem de aplicações que se assemelham com as aplicações que foram propostas no projeto, que foi a realização de técnicas de análise espacial dos dados de um banco montado no sistema SPRING.

2 – Objetivo

O objetivo principal desse projeto continua sendo o de criar e difundir uma cultura do uso de Sistemas de Informação Geográfica. Para isso, aprofundamos os estudos sobre um SIG, tornando possível a ampliação do banco de dados já implementado.

Faz parte do nosso objetivo também estudar os princípios de análise espacial em SIG's para verificação de quais análises são possíveis nesses sistemas e quais tipos de análises o nosso banco de dados, com a sua estrutura, possibilita. Além disso, como resultado de todo estudo teórico, como também das aplicações das técnicas de análise espacial, faz parte do nosso objetivo, a geração de diversos mapas através do sistema SPRING. Nesses mapas estão espacializadas as informações resultantes das análises.

3 – Metodologia

3.1 – O que é um Sistema de Informação Geográfica (SIG)?

Os SIG's são ferramentas computacionais que auxiliam o Geoprocessamento no tratamento da informação geográfica. Com esse sistema é possível armazenar em uma única base de dados diferentes informações sobre determinado espaço geográfico. Essas informações poderão ser inseridas no computador sob diversas formas, por exemplo, teclado, scanner, mesa digitalizadora, etc.

Os Sistemas de Informações Geográficas realizam a manipulação computacional de dados geo-referenciados com várias capacidades, entre elas a de codificação de dados, armazenagem, recuperação e visualização de dados, geração de mapas, gráficos, relatórios e sobreposição de mapas para utilização especialmente em Análise Espacial (Assunção, 1999).

No entanto, como os SIG's possibilitam aplicações com uma complexidade bem maior, avançamos mais os estudos sobre essa tecnologia e procuramos entender melhor o que seja análise espacial. A tabela 1 ilustra alguns tipos de análises que podem ser feitas através de um Sistema de Informação Geográfica (Câmara e Medeiros, 1996).

TABELA 1 – EXEMPLOS DE ANÁLISE EM UM SISTEMA DE INFORMAÇÕES GEOGRÁFICA

Análise	Pergunta	Exemplo
Condição	“O que está...”	“Qual é a população desta cidade?”
Localização	“Onde está...?”	“Quais as áreas com declividade acima de 20%?”
Tendência	“O que mudou...?”	“Esta terra era produtiva há 5 anos atrás?”
Roteamento	“Por onde ir...?”	“Qual o melhor caminho para se passar um metrô?”
“Padrões	“Qual o padrão...?”	“Qual a distribuição da dengue na Paraíba?”
Modelo	“O que acontece se...”	“Qual o impacto do clima se desmatarmos a Amazônia?”

Fonte: Câmara e Medeiros (1996).

Apesar de já ter disponível inicialmente um banco de dados com informações sobre o estado, não era possível a realização de análises complexas. Portanto ampliou-se o banco de dados para uma aplicação mais proveitosa.

3.2 – Análise Espacial

Dentro dos usos e aplicações dos Sistemas de Informação Geográficas, uma das aplicações mais utilizadas é a representação espacial de informações específicas de um determinado espaço geográfico (Câmara, 1999; cap 11), ou seja, com a obtenção de dados sobre uma determinada área, disposição desses dados em um banco de dados e um mapa de municípios, por exemplo, podemos gerar mapas com o auxílio de um SIG, que representem espacialmente o fenômeno.

A grande maioria dos usuários do SIG limita seu uso a essas operações, quando é possível com essa tecnologia fazer aplicações mais complexas. Um SIG não é apenas um depósito de dados geográficos com a possibilidade de gerar desenhos que representem esses dados. A característica que o distingue de outros sistemas é a possibilidade de gerar novas informações a partir dos dados disponíveis no seu banco de dados (Câmara, 1999). Essas novas informações são adquiridas com o cruzamento dos diversos dados existentes sobre uma área geográfica. Com essas novas informações os dados são espacializados havendo, assim, a possibilidade de inferir relações espaciais que não seriam tão evidentes apenas com a listagem dos dados. Esse processo é um tipo de aplicação do SIG denominado de Análise Espacial de Dados Geográficos. A análise espacial é uma habilidade de se combinar dados espaciais de naturezas diversas, manipulando-os e aplicando-lhes técnicas (matemáticas, estatísticas, etc.) para se obter conclusões adicionais só evidentes com um estudo mais profundo e com uma espacialização dos dados. Segundo Carvalho (1998), “análise espacial é o estudo quantitativo de fenômenos que são localizados no espaço. Utiliza-se a expressão análise de dados espaciais em oposição à análise de dados em geral, quando as técnicas utilizadas consideram explicitamente a localização espacial”.

3.3 - Ferramentas de Análise Espacial

Existem diversas ferramentas lógicas e estatísticas que são utilizadas na análise espacial. As ferramentas que utilizamos no nosso trabalho foram as seguintes:

3.3.1 - Consultas

As consultas ao banco de dados podem ser feitas com a geração de mapas onde podemos visualizar como se efetua no espaço a distribuição das variáveis do banco de dados. Esses mapas são gerados a partir de expressões lógicas sobre as variáveis do banco de dados.

3.3.2 - Análise exploratória

• Estatística descritiva

É o estudo das relações estatísticas que possibilitam a descrição, sob diversos ângulos, do conjunto de dados representado pela amostra (Costa, 1998). Para descrição e análise dos nossos dados utilizamos as relações estatísticas e as representações gráficas a seguir, que nos ajudaram a compreensão dos resultados:

Média – A média aritmética é a soma das observações dividida pelo número delas.

Moda – A moda é o valor do conjunto de valores que aparece mais freqüentemente.

Mediana – Quando os valores do conjunto estão agrupados em ordem, seja crescente ou decrescente, a mediana é a observação que ocupa a posição central, ou seja aquela que apresentar o mesmo número de dados antes e depois de si.

Desvio padrão – Existem métodos que possibilitam a medição das variações dos fenômenos. A medida da variância sintetiza a variabilidade de uma série de dados e nos permite comparar conjuntos que possuem diferentes valores. O princípio básico é medir o quanto os valores se desviam das médias dos conjuntos de valores. A variância é uma medida que expressa um desvio quadrático médio, o que pode causar problemas de interpretação. Para evitar esses problemas usa-se o desvio padrão que é obtido com a raiz quadrada positiva da variância. (Bussab, 1987).

Histograma – O Histograma é um gráfico de barras contíguas, onde a altura é proporcional à freqüência e a base é constituída por um segmento cujos extremos representam os extremos da *i*-ésima classe” (Bussab, 1987).

Qqplot – É uma representação gráfica que mostra a adequação dos dados a uma distribuição normal previamente definida.

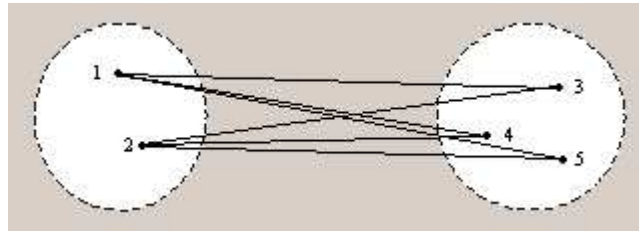
• Classificação

Classificação hierárquica é uma ferramenta de análise estatística que serve para realizar o agrupamento dos dados. Ela não requer a especificação do número de classes e é a mais adequada para quando se tem dados de contagem, como por exemplo: população.

Para realizar a classificação num banco de dados que possui variáveis incompatíveis quanto a unidade e a grandeza, como no caso de dados de renda (valores monetários) e

contagem da população, procede-se transformações, como por exemplo: o logaritmo neperiano (\ln) e a padronização de cada variável. A transformação logarítmica visa a aproximação pela distribuição normal e a padronização visa a obtenção da distribuição normal $N(0,1)$ para cada variável: $Z=(X-\mu)/\sigma$, onde, μ é a média e σ é o desvio padrão da variável X .

Para realizarmos os agrupamentos dos dados no nosso trabalho o método utilizado foi o de ligação média entre os grupos medido pela distância euclidiana ao quadrado: $d_{ij}=\{\sum(x_i - x_j)^2\}^{0,5}$ para $i \neq j$, ou seja, a distância entre dois grupos é a distância média entre todos os pares de pontos de 2 grupos distintos (figura 01).



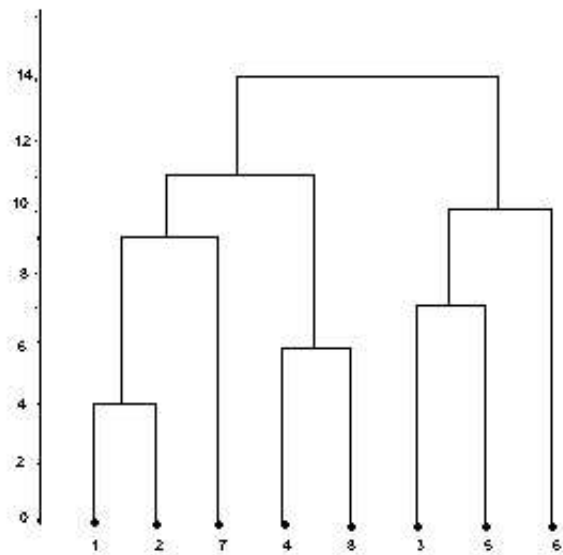
Fonte: Johnson e Wichern,(1992)

Figura 01 Distância média entre todos os pares de pontos de 2 grupos distintos

Os passos para a execução de uma Classificação Hierárquica para N grupos de objetos são os seguintes (Johnson e Wichern, 1992):

- 1) Inicia-se com N agrupamentos, cada qual contendo apenas uma entidade e uma matriz de distâncias (ou similaridade) $N \times N$;
- 2) Busca-se na matriz de distâncias o mais próximo ou mais similar par de agrupamentos U e V .
- 3) Une-se os agrupamentos U e V em um novo agrupamento UV . Recalcula-se a matriz de distâncias eliminando-se as informações dos grupos U e V separadamente e substituindo-as pelas informações do novo grupo UV .
- 4) Repete-se os passos 2 e 3 até que todas as entidades estejam contidas em um único agrupamento.

Podemos visualizar cada fusão desses agrupamentos em um gráfico cartesiano denominado dendograma que está representado na figura 02. No eixo horizontal estão as entidades, que no nosso caso são os municípios e no eixo vertical estão as distâncias do centro dos agrupamentos.



Fonte:

Figura 02 – representação do dendograma

3.4 – Exemplos de análises espaciais

A seguir mostramos alguns exemplos de análises espaciais possíveis de serem realizadas com um SIG. São trabalhos científicos que utilizam as técnicas de análise exploratória que estudamos

Exemplo 1:

Na sua tese de doutoramento, Carvalho (1997) faz a caracterização sócioeconômica de regiões urbanas a partir de informações referentes à microáreas. Com essa caracterização houve a possibilidade de estudo de doenças nas populações, podendo ser analisadas as condições mais gerais de saúde da população. Houve a preocupação com a distribuição geográfica do evento, considerando aspectos da contigüidade espacial das ocorrências. “O corpo principal da tese refere-se à aplicação de métodos de análise de dados espaciais a variáveis sócioeconômicas, comparando classificação multivariada onde se modelou espacialmente os indicadores, com classificação onde se considera cada área independente da sua localização espacial”. Foi feito um levantamento em publicações nacionais dedicadas à saúde pública e constatou-se que há uma pequena utilização de técnicas de análise espacial. Essas técnicas podem ser sistematizadas, a partir do objeto e do tipo de dado disponível em (Carvalho, 1997):

- distribuição de pontos (“*point pattern*”) – quando o objeto da análise é a posição relativa de objetos ou eventos precisamente localizados, sejam estes casos de doenças ou espécies vegetais em estudos de ecologia ambiental;
- - geoestatística – conjunto de técnicas aplicadas que pressupõem a continuidade espacial do objeto, utilizadas na estimativa e interpolação, por exemplo, de fatores ambientais cuja distribuição é contínua;
- dados de áreas – quando a ocorrência do fenômeno em estudo é mensurada a partir de dados agregados por área, como é o caso de taxas de morbi-mortalidade por município;
- deslocamento – quando o objeto de estudo é o acesso e o fluxo entre regiões, inclusive otimizando trajetórias e estudando a localização de equipamentos urbanos.

A fim de compreender como um contexto afeta a saúde de grupos populacionais através de seleção, distribuição, interação, adaptação e outras respostas, tornou-se necessário medir efeitos em nível de grupo, uma vez que as medidas de nível individual não dão conta desses processos. As pessoas vivem em grupos e a análise individual dos fenômenos não capta os efeitos das interações entre uma pessoa e outras na transmissão de infecções, comportamentos e valores. A perspectiva geral do trabalho foi o aprimoramento do perfil de risco de populações urbanas utilizando uma perspectiva espacial na análise de dados na área de saúde. Fazendo um estudo de diversos métodos de análise espacial aplicados a indicadores do censo demográfico, visando identificar regiões urbanas segundo seu perfil sócioeconômico. O perfil sócio-econômico, no caso foi definido com base em

As técnicas de análise exploratória aplicadas a dados espaciais são essenciais ao desenvolvimento das etapas da modelagem estatística espacial, em geral muito sensível ao tipo de distribuição. As técnicas empregadas são em geral, adaptações das ferramentas usuais. As mais comuns são:

- visualização dos pontos extremos nos mapas;
- Gráficos de médias e medianas segundo linhas e colunas dos pontos amostrados;
- *Boxplots* de indicadores de diferenças entre pares de valores mensurados.

As técnicas de mapeamento permitem descrever de forma gráfica os diversos fenômenos em saúde. Neste trabalho essas técnicas foram utilizadas como ferramentas de análise exploratória e de representação de resultados. Os principais tipos de mapas utilizados nesta parte do trabalho foram:

- localização de pontos – onde, sobre uma base cartográfica, são assinalados os pontos onde se localizam unidades de saúde, fontes poluentes, casos de doenças, etc.;
- padrão – com a subdivisão mais usual sendo geopolítica: municípios, estados, distritos ou bairros, onde os indicadores são calculados para cada subdivisão do mapa, permitindo a divisão da região em classes.

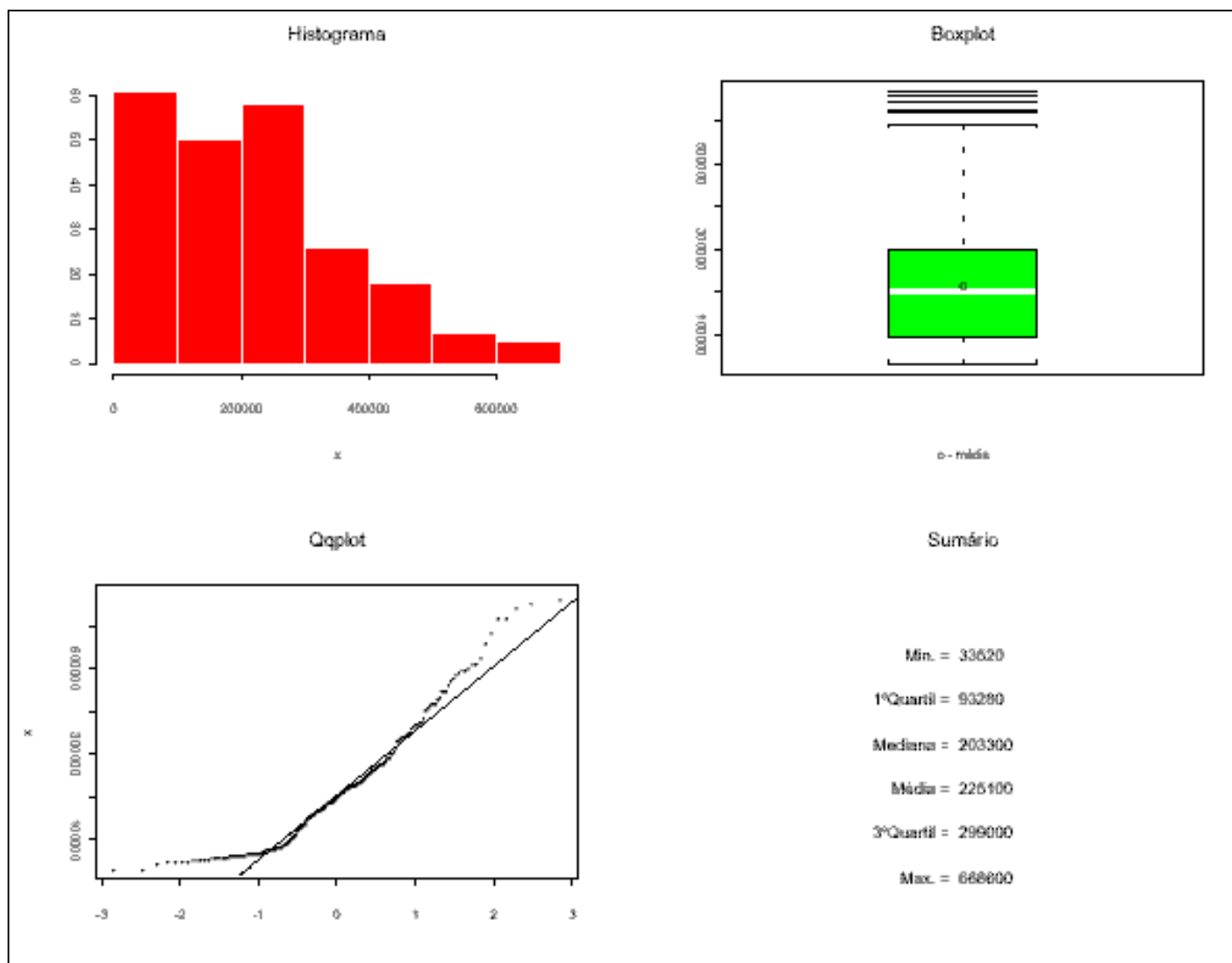
A área estudada foi a Ilha do Governador (figura 03), composta por 14 bairros e 225 setores censitários, onde localizam-se residências de classe média e favelas, indústrias, o Aeroporto Internacional do Galeão e o Campus da Universidade Federal do Rio de Janeiro. A base cartográfica foi digitalizada e convertida para o MAP INFO a partir de plantas originais na escala de 1:5.000, obtidas através de convênio com a FIBGE.



Fonte: Carvalho, (1997)

Figura 03 - Representação da área estudada
Ilha do Governador

Os seis indicadores selecionados para o trabalho foram, a renda média do chefe do domicílio, a proporção de chefes de família com segundo grau completo ou superior, a proporção da população alfabetizada, proporção de domicílios alugados, proporção de domicílios com abastecimento de água proveniente da rede pública geral, proporção de domicílios ligados à rede pública de esgoto. Na figura 04 está representado o resultado da análise exploratória univariada dos indicadores. A renda média do chefe da família, apresenta distribuição com desvio para a esquerda e média e mediana bastante próximas. Pode-se observar valores atípicos no extremo superior da distribuição, onde a renda média do chefe da família é muito superior. O *qqplot* apresenta boa aderência apenas no centro da distribuição.



Fonte: Carvalho (1997)

Figura 04 – Análise exploratória da Renda Média do Chefe da Família (bruta)
Ilha do Governado, Censo Demográfico de 1991

Igual análise foi realizada para as demais variáveis citadas. A partir das variáveis selecionadas foram obtidos indicadores segundo a classificação criada. A figura 05 apresenta os setores segundo a classificação:

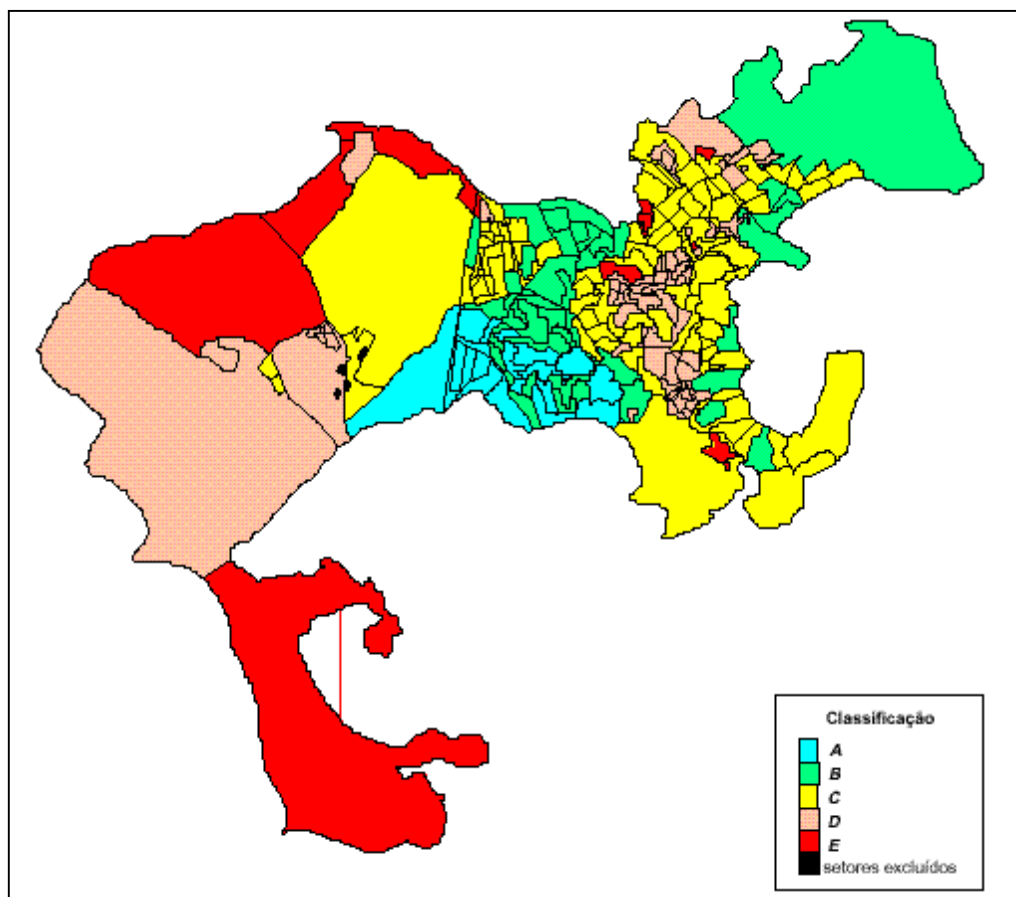
Grupo A – composto por apenas 18 setores, apresenta a maior renda média do chefe da família, melhor situação em relação a todos os indicadores, proporção de casas alugadas mais baixa que a média geral;

Grupo B – com 40 setores, tem a renda média e a escolaridade do chefe da família um pouco menor, demais indicadores semelhantes, maior proporção de casas alugadas que em A;

Grupo C - no maior grupo, com 87 setores, a renda média do chefe da família encontra-se na média geral da população e a proporção de chefes de família com escolaridade igual ou superior ao segundo grau é 52%, superior à média geral; o acesso às redes de água e esgotos e a proporção de população alfabetizada apresentam distribuição idêntica aos grupos anteriores, e a proporção de casas alugadas tem a maior média;

Grupo D – com 66 setores apresenta, junto com o grupo E, as menores renda média do chefe da família e proporção de chefes da família com escolaridade igual ou superior ao segundo grau, diferindo pela proporção de casas ligadas à rede de esgotos; a proporção de população alfabetizada apresenta média inferior à média geral;

Grupo E – com 14 setores é o menor grupo, caracterizando-se principalmente pela proporção de casas ligadas à rede de esgotos no extremo inferior da distribuição; embora a proporção média de casas ligadas à rede pública de água não seja muito diferente do restante, a distribuição inclui os valores mais baixos; proporção da população alfabetizada apresenta menor média e a maior amplitude de distribuição e praticamente inexistem casas alugadas neste grupo.



Fonte: Carvalho (1997)

Figura 05 – Classificação dos Setores Censitários
Ilha do Governador

Exemplo 2:

Os autores Carvalho e Cruz (1998), procuraram investigar a distribuição espacial da mortalidade por acidentes com veículos a motor, nos municípios da região sudeste do Brasil, empregando diversas técnicas exploratórias de análise espacial.

Os dados utilizados foram, dados do Censo Demográfico de 1991, disponibilizado pela FIBGE em formato digital (ano de 1991) e a malha rodoviária de 1996 disponibilizada pela FIBGE em CD-ROM.

Os dados relativos aos óbitos foram coletados através da Internet (<http://www.datasus.gov.br>) e do CD-ROM do Sistema de Informações de Mortalidade (SIM) distribuído pelo DATA SUS.

Os métodos utilizados foram, o cálculo da média trienal dos óbitos (1990-1992) visando estabilizar os indicadores nos locais onde a população é pequena (ver tabela 2). Foram calculados os seguintes indicadores:

- Razão de óbitos por acidentes em veículo a motor segundo município de ocorrência por 100.000 habitantes – total de óbitos ocorridos no município dividido pela população residente, vezes 100.000.

TABELA 2
TOTAL DE ÓBITOS E TAXA DE MORTALIDADE POR UF, NÚMERO MÁXIMO DE ÓBITOS
E MAIOR INDICADOR POR MUNICÍPIO ISOLADO, SEGUNDO CAUSAS ESTUDADAS,
REGIÃO SUDESTE, BR, 1990-92

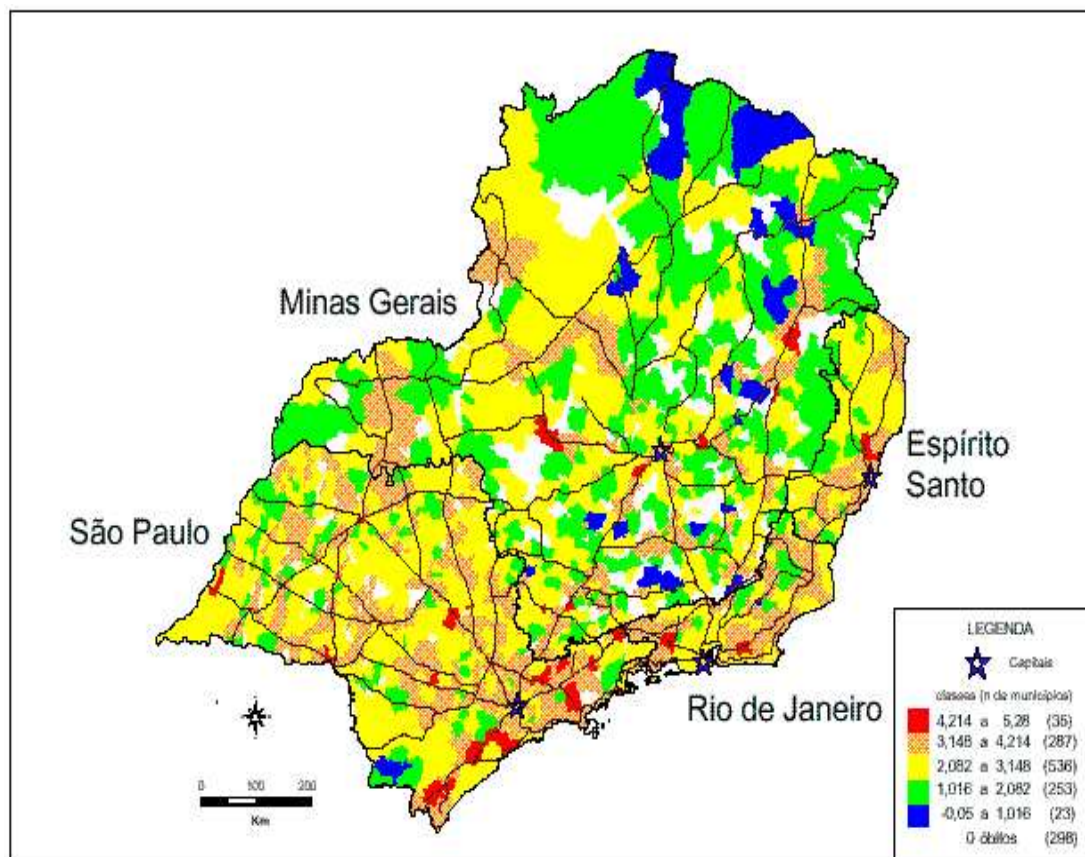
Indicador		UF			
		MG	ES	RJ	SP
Acidentes de trânsito	Índice estadual	14,68	27,11	21,79	23,64
	Nº máximo (p/município)*	1584	496	3742	6940
	Maior índice (p/município)	191,88	99,82	83,95	197,32
Total de municípios		723	87	70	572

* soma dos óbitos nos 3 anos: 90-92.

Fonte: FIBGE, SIM/MS.

“Nos métodos de análise, foram empregadas técnicas de análise exploratória, com ênfase naquelas disponíveis em SIG's, tais como mapas de padrão e superposição de camadas de informação. O MAP INFOTM foi o *software* utilizado. Nos mapas de padrão utilizou-se a transformação logarítmica dos indicadores, apresentados em 5 classes iguais, excluídos os municípios sem registro de óbito pelas causas selecionadas.

Foi analisada a distribuição espacial da mortalidade por acidentes causados por veículos a motor, incluindo pedestres, passageiros e motoristas envolvidos em acidentes automobilísticos, representada no mapa da figura 06. Este mapa foi gerado num SIG e é o resultado da superposição dos dados pesquisados com a malha municipal e a malha rodoviária, ambas em formato digital.



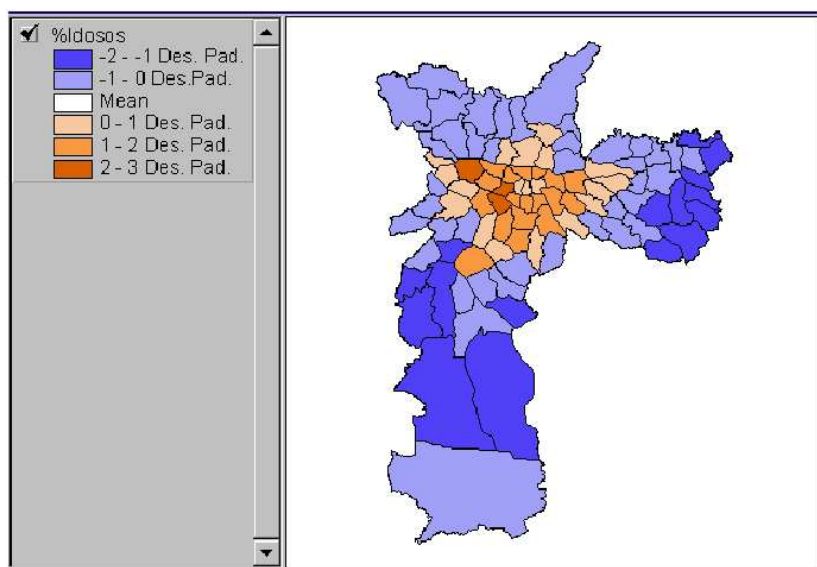
Fonte: FIBGE, SIM/MS

Figura 06 – Logaritmo da Razão de Mortalidade
Por acidentes com veículos a motor
Segundo local de ocorrência por 100.000 residentes,
Região Sudeste-BR,1990-92

Com a distribuição espacial dos dados em forma de mapa os autores chegaram a algumas conclusões que não seriam possíveis com os dados em forma de tabelas. Uma das conclusões dos autores é que os índices de mortalidade por acidentes automobilísticos são altos próximos à algumas rodovias como a São Paulo - Curitiba e a Rodovia Presidente Dutra, justificando suas denominações de “rodovias da morte”.

Exemplo 3:

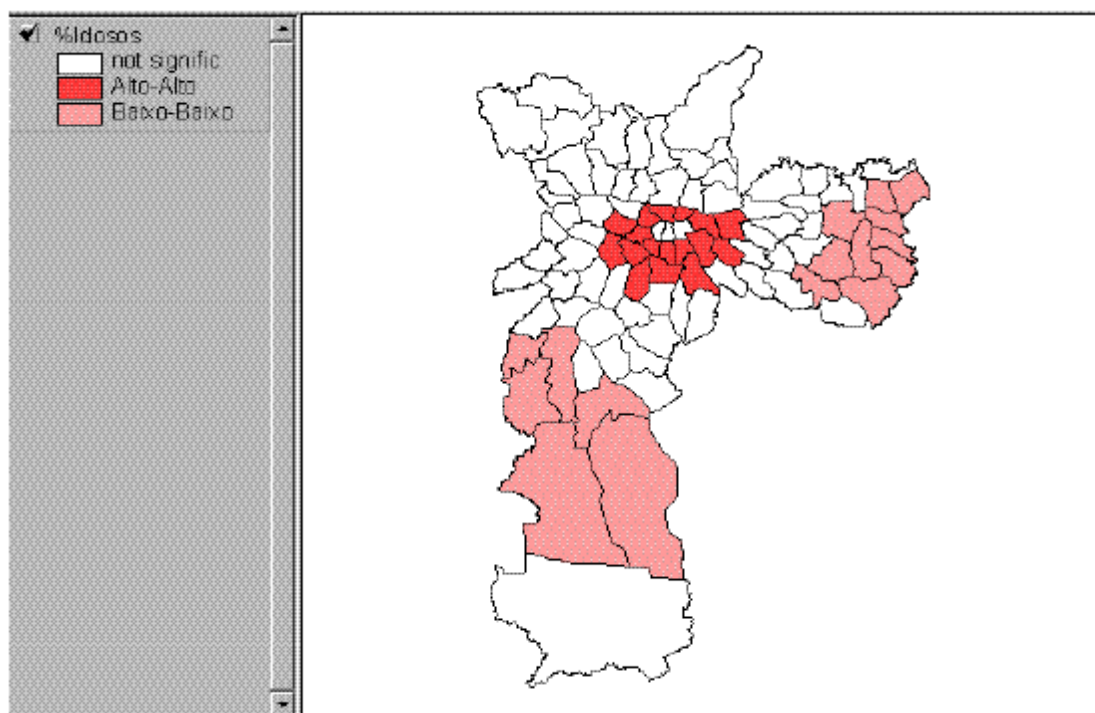
Neste outro trabalho, Gilberto Câmara (Câmara, 2000) analisou dados do município de São Paulo, considerando o percentual dos moradores com mais de 70 anos como indicativo da qualidade de vida e utilizando esse mesmo indicador, buscou obter informações adicionais como: se existem regiões com variações extremas de longevidade, que indicam casos críticos de exclusão ou inclusão social. Foram utilizados os dados de população, número de idosos, renda por faixa salarial, total do domicílios, casas sem esgoto, número de analfabetos, números de crianças na escola.



Fonte: Câmara, (2000).

Figura 07 – Distribuição estatística da porcentagem de idosos por distritos de São Paulo (laranja: acima da média; azul: abaixo da média) Média dos dados = 3,43% e desvio padrão = 2,04.

Inicialmente foram calculados a média e o desvio padrão do percentual de idosos e plotado o mapa de distribuição espacial (Figura 07) que mostra duas regiões com valores muito acima da média (mais de 2 desvios padrões), Jardim Paulista, Consolação e Lapa, com mais de 7,5% de idosos. Este primeiro resultado mostra valores extremos, mas não mostra indicações sobre a dinâmica espacial do processo. Para entender essa dinâmica o autor usou o princípio básico do Geoprocessamento: “o lugar faz a diferença”. Com isso, calculou a média local do número de idosos para cada distrito (obtida a partir da vizinhança de cada região) e comparou o resultado com o valor do atributo para o distrito. Se a diferença for pequena, pode ser indicador de regiões com dinâmica própria, fortemente relacionada e chamadas de aglomerados espaciais e que, neste caso, indicam “bolsões de inclusão ou exclusão social”.



Fonte: Câmara, (2000).

Figura 08 – Aglomerados espaciais na cidade de São Paulo, considerando o percentual de idosos com indicador. Vermelho: “bolsões” de inclusão social; rosa: “bolsões” de exclusão social.

O resultado do indicador baseado na média local está mostrado na figura 08 onde estão destacadas as regiões que formam aglomerados espaciais, com base em critérios de significância estatística (probabilidade de ocorrência aleatória menor que 5%). Percebemos que na zona Leste e na zona Sul de São Paulo há regiões críticas, onde o agravamento das condições sociais resulta numa degradação do nível de vida ainda mais séria que poderíamos esperar. Por exemplo, no caso da zona Sul, os distritos de Jardim Ângela (0,9% de idosos), Parelheiros (1,1%), Capão Redondo (1,2%), Jardim São Luís (1,2%), Cidade Dutra (1,54%) e Grajaú (0,85%) formam um bolsão bem definido de exclusão social.

Continuando a análise, estabeleceu-se uma correspondência entre as variáveis fazendo a hipótese que a quantidade de idosos é condicionada pelas condições de higiene (domicílios com rede de esgotos) e renda (chefes de família com mais de 20 salários mínimos). Mais uma vez o espaço foi levado em conta, supondo-se que a relação entre as variáveis funcione de forma diferente no centro e na periferia. Para comprovar essa idéia, a cidade foi dividida em três grupos: o centro, a periferia e a transição centro-periferia e obteve-se um modelo pelo qual confirma-se a hipótese estabelecida e verifica-se que a relação é condicionada pela localização do idoso na cidade. Assim, a combinação das técnicas de estatística convencional e análise espacial fornece ao usuário meios importantes de “garimpar” padrões e relações em seus dados e calçar suas hipóteses em sólidos fundamentos quantitativos.

Após o estudo das técnicas de análise espacial e dos exemplos estudados, passamos a realização da análise espacial utilizando os dados do nosso banco. Visando uma análise mais sofisticada, inserimos novas variáveis no nosso banco de dados.

4 – Ampliação do Banco de Dados e aplicações

4.1 – Implementação das novas variáveis

A ampliação do banco de dados, prevista no projeto, foi realizada com a inserção das novas variáveis já determinadas. Elas são, a área dos municípios, a densidade populacional, a situação de urbanização, perfis de idade, sexo, religião, cor ou raça e condição de ocupação. Todas obtidas do Censo Demográfico Brasileiro de 1991 (IBGE, 1991). Durante a realização dessa etapa do projeto tivemos alguns problemas com a inserção dessas novas variáveis, pelo fato do Sistema SPRING ter sido criado recentemente, na medida em que vamos utilizando-o é que vamos descobrindo alguns pontos que necessitam de ajustes e modificações. Um dos problemas que detectamos foi o seguinte, depois que montamos o banco de dados da Paraíba com as primeiras variáveis e fomos acrescentar as novas, não conseguimos acrescentá-las, pois o SPRING, naquele momento, não permitia a ampliação. Tentamos várias soluções e a saída encontrada foi usar o programa dBASE para fazer uma espécie de atalho até o SPRING. Entramos no programa dBASE, abrimos o arquivo que contém os dados do nosso projeto (no nosso caso o arquivo é CG000003.dbf) e ao acessar a tabela com esses dados, acrescentamos as novas variáveis a serem inseridas no banco. Foi dessa forma que conseguimos a ampliação do banco de dados sem entrar em conflito com o SPRING. Os procedimentos a seguir indicam passo a passo a maneira que conseguimos para a ampliação do banco de dados:

Entra no programa dBASE

Abrir o arquivo do SPRING que contém os dados do projeto.

[View]

[Table struture] Aparece uma caixa de diálogo perguntando se queremos modificar a estrutura da tabela.

[Open Exclusive] Abre uma tabela com as variáveis já existentes no banco de dados. Na última variável (final da tabela), clicar com o botão esquerdo do mouse na última coluna

[TAB] Agora é só digitar o nome das novas variáveis. Ao fechar o dBASE e abrir o SPRING, as variáveis recém criadas estarão no banco de dados junto com as outras já existentes. Agora é só preencher com os respectivos valores.

4.2 - Nomenclatura das variáveis inseridas no banco de dados

Como já foi mencionado, inserimos novas variáveis ao banco de dados, mas algumas dessas variáveis possuem uma quantidade considerável de tipos e aspectos diferentes. Na variável religião, por exemplo, há 13 tipos diferentes de religiões, cada uma com uma nomenclatura que possui em torno de 15 caracteres. Como o banco de dados só permite a inserção de variáveis com no máximo 8 caracteres, foi necessário criarmos uma nomenclatura específica e abreviada. Vamos

explicar agora como foi nomeada cada uma das variáveis para facilitar a consulta ao banco de dados.

A população residente: POPUL

“A população foi constituída pelos moradores habituais no domicílio, ou seja, as pessoas que tinham o domicílio como local de residência habitual, quer estivessem presentes ou ausentes na data de referência. As pessoas moradoras habituais do domicílio que estavam ausentes na data de referência foram recenseadas, desde que a sua ausência não tenha sido superior a 12 meses em relação àquela data”

O nome do município : NOME_MUNI

Renda da população por faixas de salário mínimo:

R91_A14	Renda ano 1991, até ¼ de salário mínimo
R91_14 A05	Renda ano 1991 de ¼ a ½ salário mínimo
R91_05 A34	Renda ano 1991, de ½ a ¾ de salário mínimo
R91_34 A1	Renda ano 1991, de ¾ a 1 salário mínimo
R91_1 A14	Renda ano 1991, de 1 a 1¼ de salário mínimo
R91_14 A15	Renda ano 1991, de 1 ¼ a 1 ½ de sal. mínimo
R91_15 A02	Renda ano 1991, de 1 ½ a 2 salários mínimos
R91_02 A03	Renda ano 1991, de 2 a 3 salários mínimos
R91_03 A05	Renda ano 1991, de 3 a 5 salários mínimos
R91_05 A10	Renda ano 1991, de 5 a 10 salários mínimos
R91_10 A15	Renda ano 1991, de 10 a 15 salários mínimos
R91_15 A20	Renda ano 1991, de 15 a 20 salários mínimos
R91_M20	Mais de 20 salários mínimos
R91_NÃO	Sem rendimentos
R91_SDE	Sem declaração

A escolaridade por grau:

ES91_PRE	Pre-escolar
ES91_ALAD	Alfabetização de adultos
ES91_PRIM	População com primeiro grau completo
ES91_SEGU	População com segundo grau completo
ES91_SUPE	População com grau superior
ES91_MEDO	População com mestrado e doutorado

A área do município: AREA MUN

A densidade populacional: DENS_MUN

A população por sexo: SEX_MASC
SEX_FEMI

A população segundo cor ou raça:

CR_BRANC	População branca
CR_PRETA	População preta
CR_AMARE	População amarela
CR_PARDA	População parda
CR_INDIG	População indígena
CR_SDE	Sem declaração

População segundo situação de urbanização:

SIT_RURA	População rural
SIT_URBA	População urbana

População segundo a condição de ocupação:

CO_PROPR	Domicílios próprios
CO_ALUGA	Domicílios alugados
CO_CEDID	Domicílios cedidos
CO_OUTRA	Outros

“Considerou-se como particular aquele domicílio que servia de moradia a uma, duas ou no máximo cinco famílias, mesmo que estivessem localizados em estabelecimento comercial, industrial, etc. O prédio em construção onde residiam até cinco pessoas, embora sem laço de parentesco e/ou dependência doméstica, também foi considerado como domicílio particular. Classificou-se o domicílio particular em permanente - assim considerado o construído para fim residencial e improvisado – o que não atendia a referida condição, embora servisse de moradia na data do Censo, tal como o localizado em unidades (lojas, fábricas, etc.) que não possuísse dependências destinadas exclusivamente à moradia, prédios em construção servindo de moradia a pessoal de obra, embarcações, carroças, vagões de estrada de ferro, tendas, barracas, grutas, etc”. (IBGE)

Os perfis de idade:

ID_0 A4	População com idade entre 0 e 4 anos
ID_5 A9	População com idade entre 5 e 9 anos
ID_10 a 14	População com idade entre 10 e 14 anos
ID_15 A19	População com idade entre 15 e 19 anos
ID_20 A24	População com idade entre 20 e 24 anos
ID_25 A29	População com idade entre 25 e 29 anos
ID_30 A34	População com idade entre 30 e 34 anos
ID_35 A39	População com idade entre 35 e 39 anos
ID_40 A44	População com idade entre 40 e 44 anos
ID_45 A49	População com idade entre 45 e 49 anos
ID_50 A54	População com idade entre 50 e 54 anos
ID_55 A59	População com idade entre 55 e 59 anos
ID_60 A64	População com idade entre 60 e 64 anos
ID_65 A69	População com idade entre 65 e 69 anos
ID_70 A74	População com idade entre 70 e 74 anos

ID_75 A79	População com idade entre 75 e 79 anos
ID_80M	População com idade de 80 anos ou mais

A população de acordo com a religião:

REL_CATR	Religião Católica Romana
REL_EVAN	Religião Evangélica*
REL_ESPI	Religião Espírita
REL_OUTR	Outras Religiões**
REL_NÃO	Sem Religião
REL_SDE	Sem declaração

Onde, a variável Religião Evangélica neste caso, é a junção da Religião Evangélica Tradicional com a Pentecostal e a variável Outras Religiões inclui as seguintes Religiões: Outra Cristã Tradicional, Cristã Reformada não determinada, Neo-Cristã, Candomblé e Umbanda, Judaica ou Israelita e Oriental.

4.3 - Como gerar mapas no SPRING

Atualmente no banco de dados existem indexadas 62 variáveis. É um número bastante significativo, levando-se em consideração que atualmente não é possível salvar automaticamente as tabelas do IBGE para o SPRING, pelo fato do sistema só ler arquivos em alguns formatos, como o dBASE, por exemplo. Desse modo a forma de inserção das 62 variáveis dos 223 municípios do estado, foi feita manualmente, o que ocupou uma parte significativa do cronograma do projeto. Com esse número de variáveis dispostas no banco de dados, podemos gerar diversos mapas a partir de suas variáveis. Por exemplo, com a variável população usamos sua nomenclatura - **POPUL** - acrescida do símbolo **>0** (maior que zero), para especificarmos que queremos obter a informação de quais municípios existem com população maior que zero. E assim, podemos proceder com todas as variáveis do banco para visualizar a distribuição espacial no estado dos dados de cada uma. Os mapas são gerados da seguinte forma:

Entrar no SPRING e selecionar no painel de controle:

[Linhas]



Mandar desenhar clicando no botão onde tem um lápis.

Selecionar:

[Objetos]

[Consultar]

Aparece o quadro *Geração e Seleção de Coleção*, onde teremos que criar uma coleção digitando um nome para a coleção no espaço e clicando em:

[Criar]

Criada a coleção, temos acesso aos atributos e sinais com os quais iremos elaborar as operações lógicas. Por exemplo, selecionando o atributo POPUL, o sinal >, digitando um valor qualquer, como o 0 (zero), e mandando em seguida criar, podemos ver embaixo a **expressão lógica: POPUL>0**, ou seja: com essa expressão vamos gerar um mapa onde podemos

visualizar as cidades com população maior a zero, ou seja, a distribuição da população no estado da Paraíba.

[Gerar]
[Aplicar]

Aparecem dois quadros. O primeiro é uma tabela com os municípios que satisfazem a expressão. Minimizando-a, veremos o segundo, que é o quadro *visualização de objetos selecionados*, nele ao lado do nome *município* existe um quadrinho que teremos que selecionar para termos acesso à função editar.

[Editar]
[Agrupamento]

Aparece, então um quadro onde temos os modos de agrupamento, o número de partes, a graduação de cores, etc.

(Modo: Quantil)

O quantil é o modo que nos permitiu a utilização de várias cores num mesmo mapa

(Número de partes: 7)

O amarelo possui 7 cores distintas que são diferenciáveis.

(Atributos: POPUL)
(Cor: amarela)

[Agrupar]
[Executar]

No quadro *visualização de objetos*, clicar no triângulo pequeno ao lado do nome município para visualizar a escala de cores do mapa

4.4 - Exemplos de consultas possíveis de serem efetuadas com nosso banco de dados

1º Exemplo (1 variável)

O primeiro exemplo é o mapa que foi gerado seguindo os procedimentos do tópico acima, utilizando expressão lógica **POPUL>0**. Este exemplo foi gerado levando em consideração apenas uma variável, a população. No mapa da figura 09, podemos observar como se configura visualmente a distribuição da população no estado da Paraíba. A faixa de cor que possui valor zero, neste caso, são os municípios novos (criados em 1997) que, evidentemente, não possuíam dados do Censo do ano de 1991. Verifica-se o predomínio de cidades com população maior de 8.000 habitantes na área que abrange a Borborema, o Brejo e o Litoral.

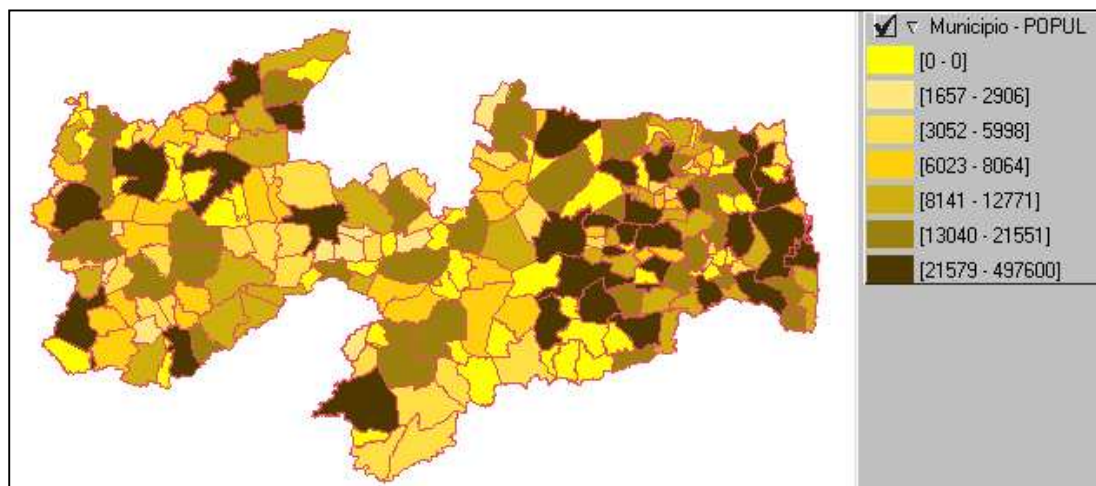


Figura 09 – Distribuição da população do estado da Paraíba

2º Exemplo (1 variável)

Neste segundo exemplo, o mapa foi gerado com os mesmos procedimentos do item 6, a única modificação foi na expressão lógica, que neste caso foi: **ES91_SUPE>0**. A representação espacial dos dados no mapa da figura 10, nos mostra a distribuição da população paraibana com grau superior. Podemos verificar uma concentração dessa população próximo a municípios onde existem universidades.

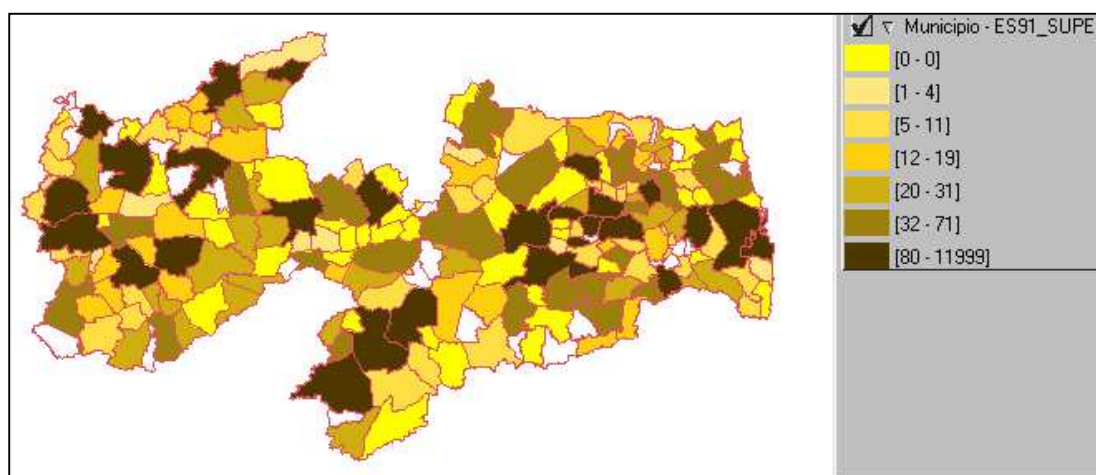


Figura 10 – Distribuição espacial da população paraibana com grau superior

Na tabela 3 temos uma coleção de mapas que foram gerados seguindo os mesmos procedimentos usados para geração dos mapas das figuras 09 e 10. Os mapas da coleção foram gerados com base nas variáveis que estão inseridas no banco de dados. Cada expressão corresponde a um mapa que representa a distribuição espacial dos dados da variável.

TABELA 3 – COLEÇÃO DE MAPAS QUE REPRESENTAM A DISTRIBUIÇÃO ESPACIAL DAS VARIÁVEIS DO BANCO DE DADOS

EXPRESSÃO LÓGICA	FIGURA
POPUL>0	09
R91_A14>0	11
R91_1 A14>0	13
R91_10 A15>0	12
R91_02 A03>0	14
ES91_SUPE>0	15
ES91_SEGU>0	17
ES91_PRIM>0	16
ES91_PRE>0	18
SEX91_MASC>0	21
SEX91_FEMI>0	22
ID_80M>0	31
ID_55 A59>0	33
ID_35 A39>0	32
ID_15 A19>0	34
CR_PRETA>0	25
CR_BRANC>0	26
REL_EVAN>0	23
REL_CATR>0	24
SIT_URBA>0	19
SIT_RURA>0	20
CO_PROPR>0	27
CO_ALUGA>0	28
CO_CEDID>0	29
CO_OUTRA>0	30

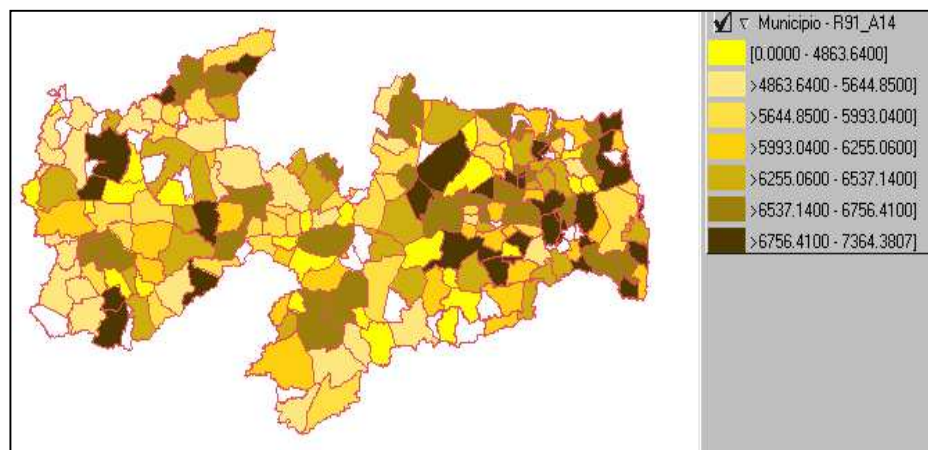


Fig. 11 – Distribuição da população com renda média até ¼ de salário mínimo no ano de 1991.

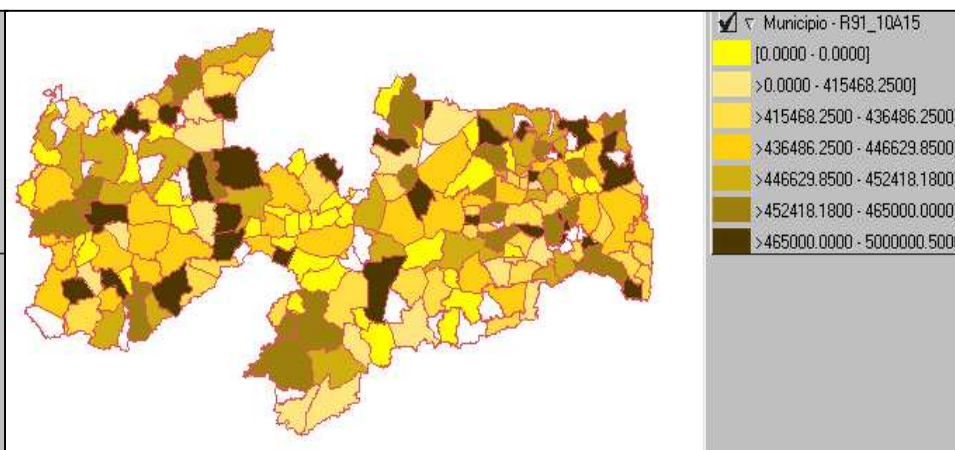


Fig. 12 – Distribuição da população com renda média de 10 a 15 salários mínimos, em 1991.

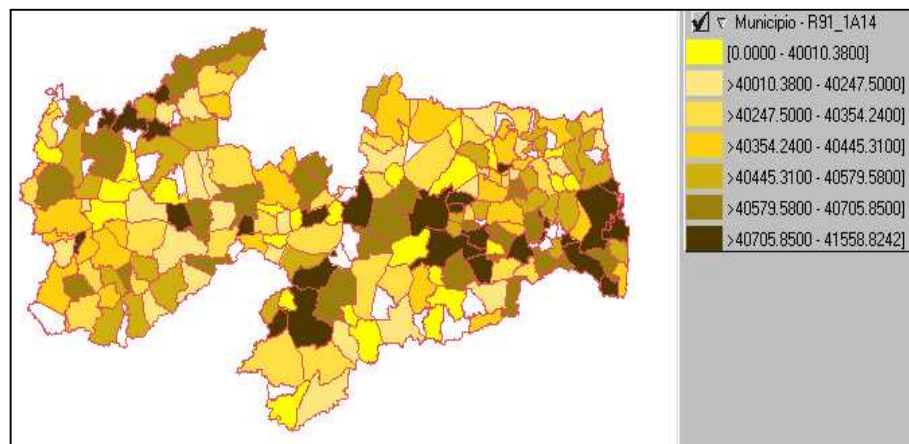


Fig. 13 - Distribuição da população com renda média de 1 a ¼ de salário mínimo, em 1991.

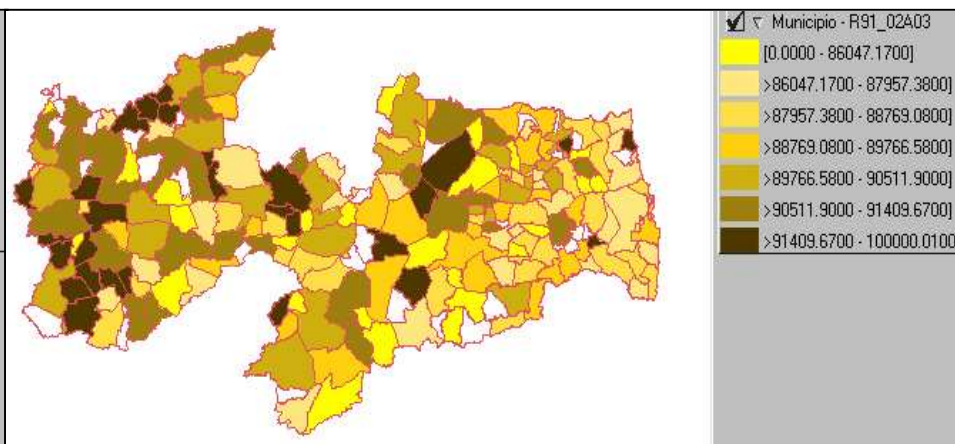


Fig. 14 - Distribuição da população com renda média de 02 a 03 salários mínimos, em 1991.

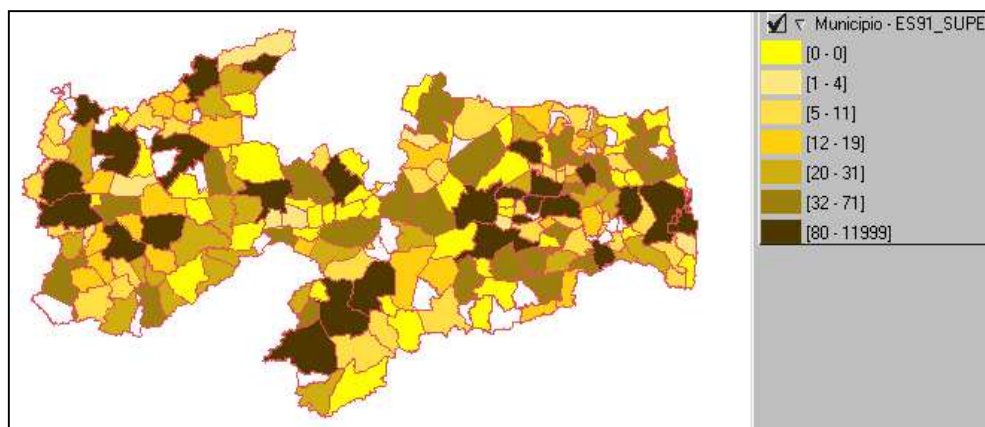


Fig. 15 – Distribuição da população em 1991, com grau superior

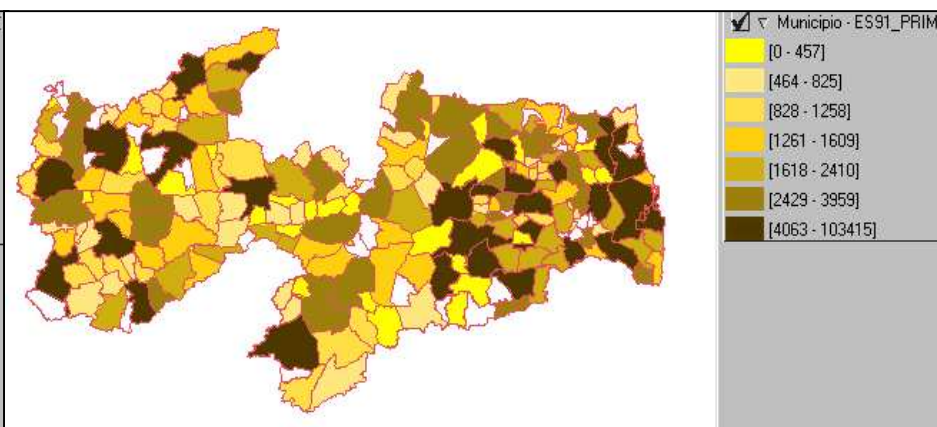


Fig. 16 – Distribuição da população em 1991, com primeiro grau completo.

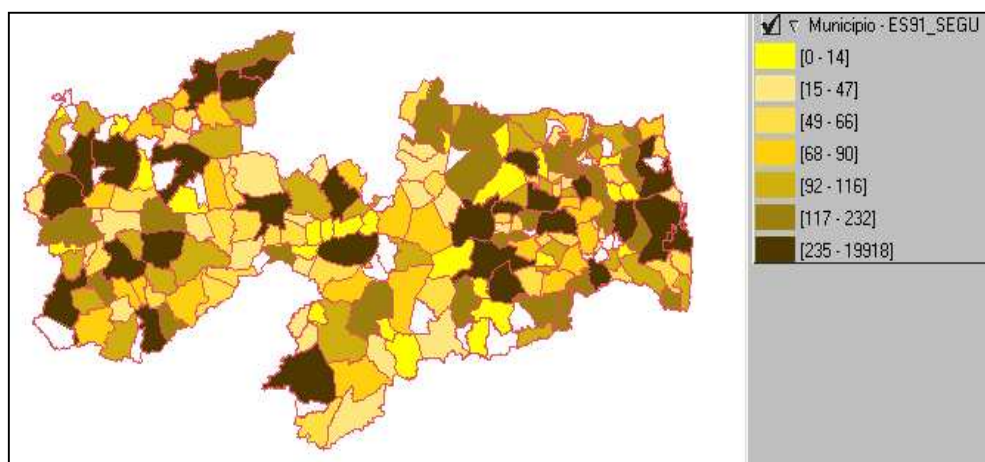


Fig. 17 – Distribuição da população em 1991 com segundo grau completo

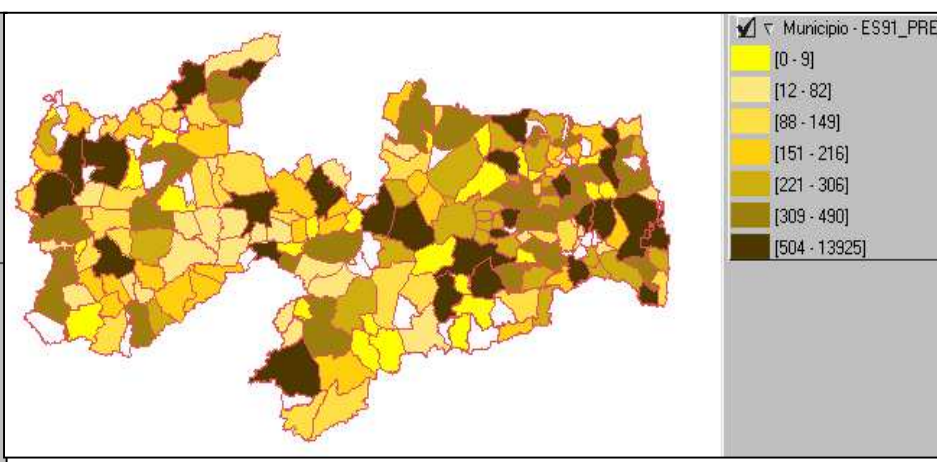


Fig. 18 – Distribuição da população em 1991 com pre-Escolar

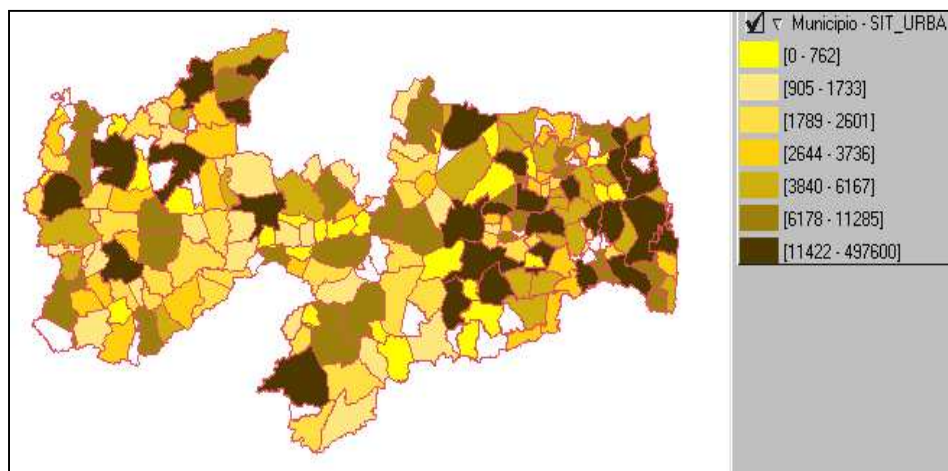


Fig. 19 – Distribuição da população urbana do estado em 1991

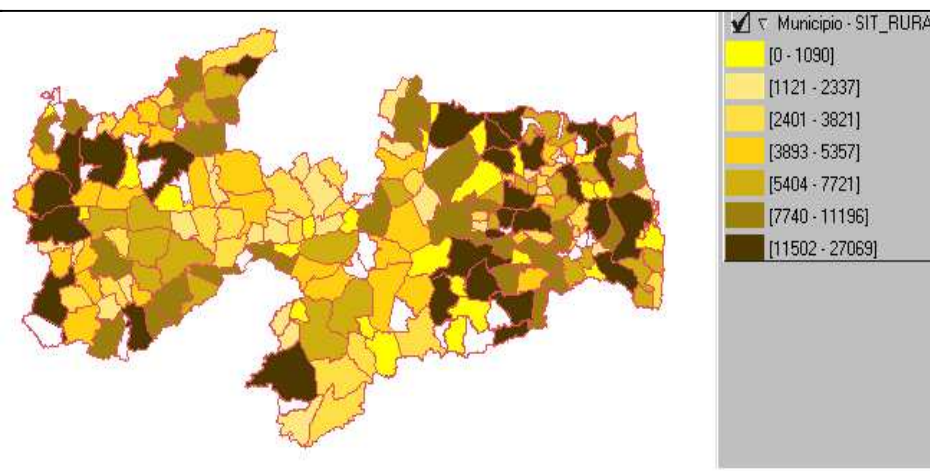


Fig. 20 – Distribuição da população rural do estado em 1991.

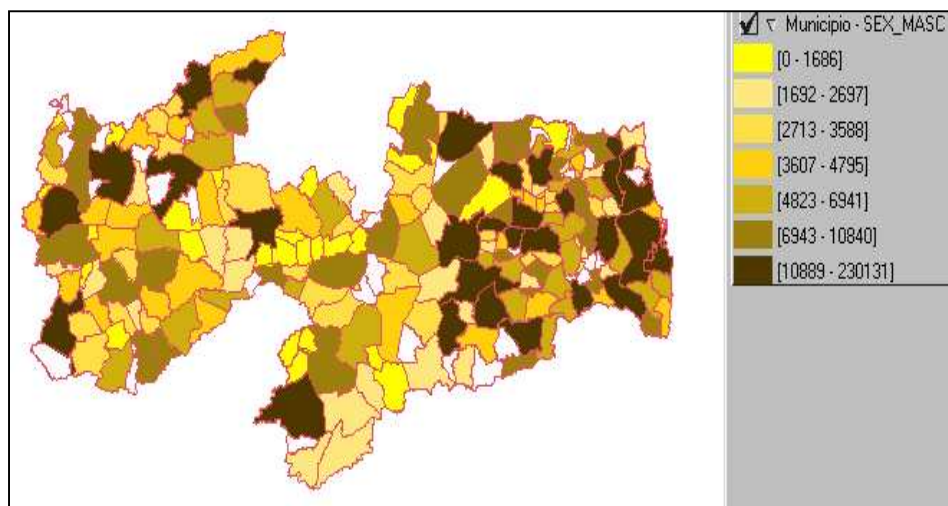


Fig. 21 – Distribuição espacial da população do sexo masculino no estado em 1991.

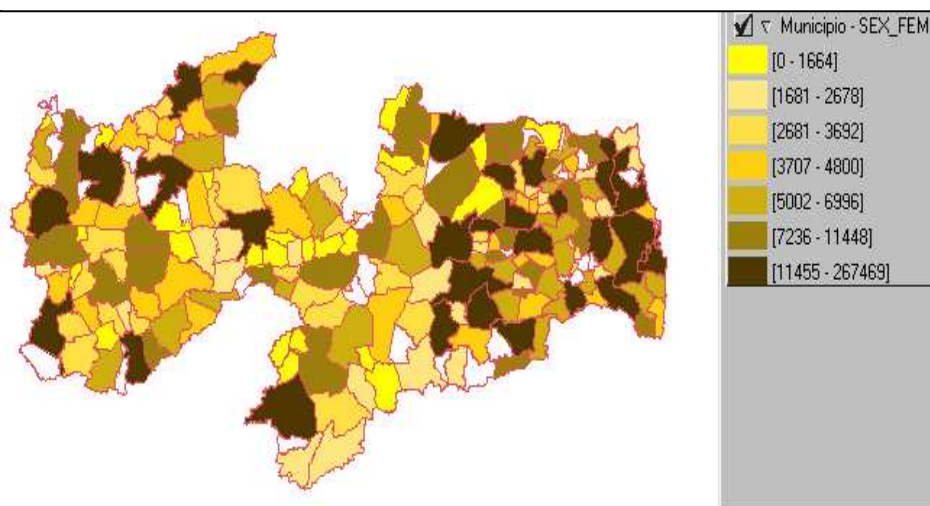


Fig. 22 – Distribuição espacial da população do sexo feminino no estado, em 1991.

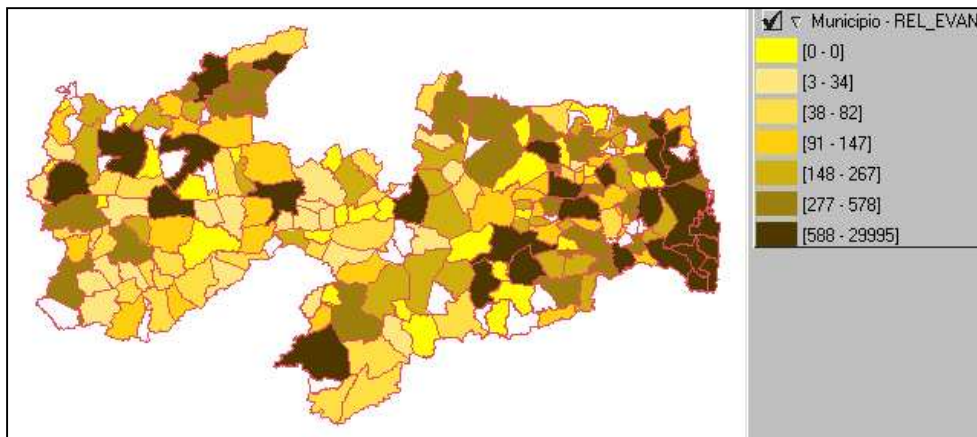


Fig. 23 – População do estado adeptos da Religião Evangélica, em 1991.

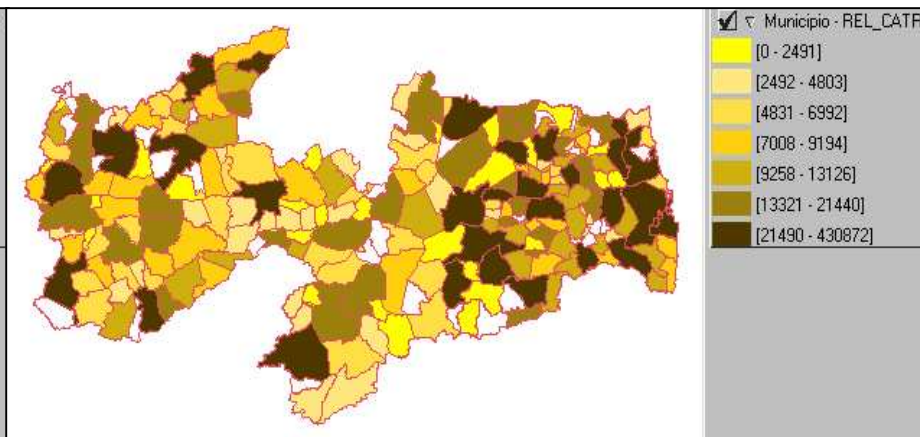


Fig. 24 – População do estado adeptos da Religião Católica Romana, em 1991.

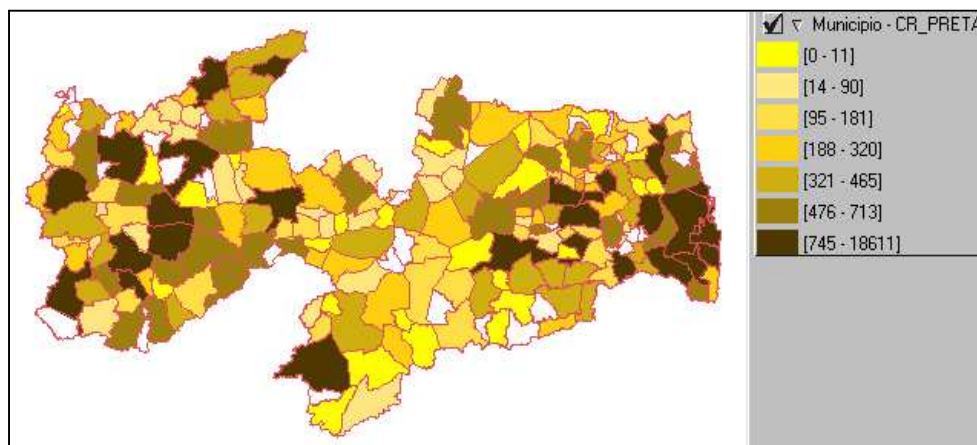


Fig. 25 – Distribuição da população de cor preta no estado em 1991.

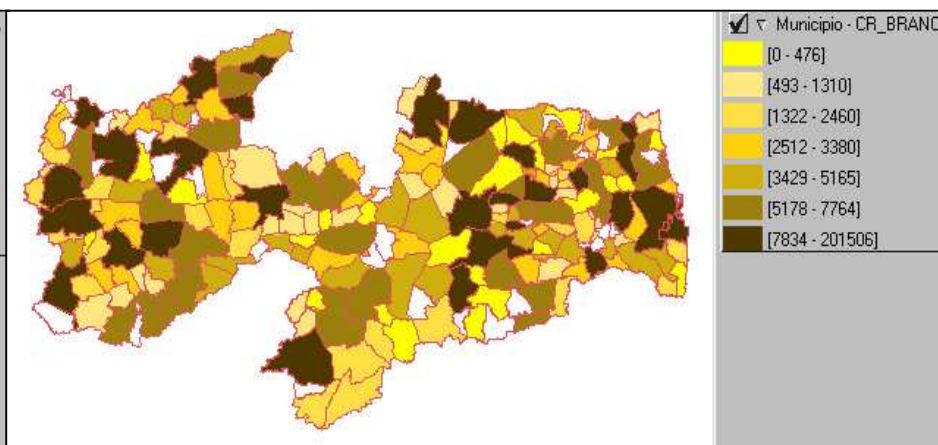


Fig. 26 – Distribuição da população de cor branca no estado em 1991.

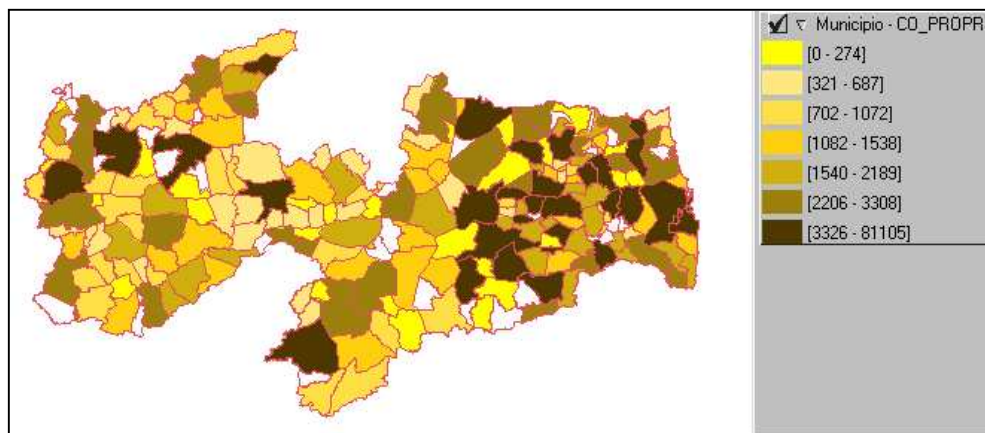


Fig. 27 – Distribuição da população com domicílios próprios, no ano de 1991.

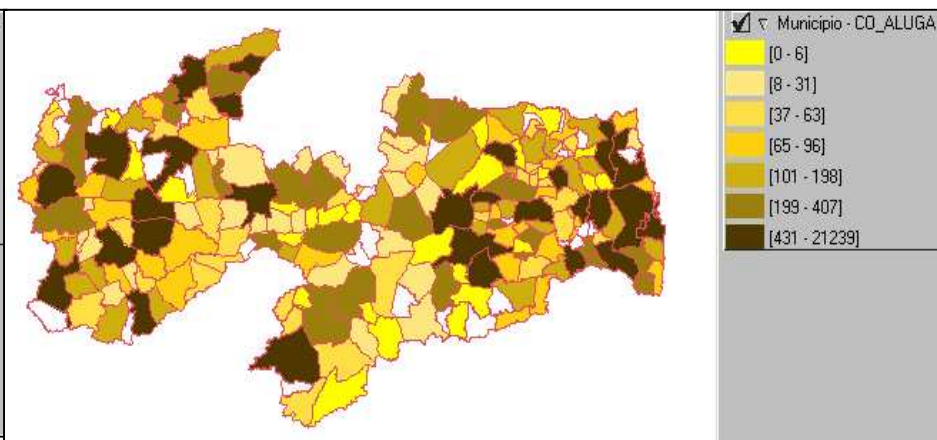


Fig. 28 – Distribuição da população residente em domicílios alugados, no ano de 1991.

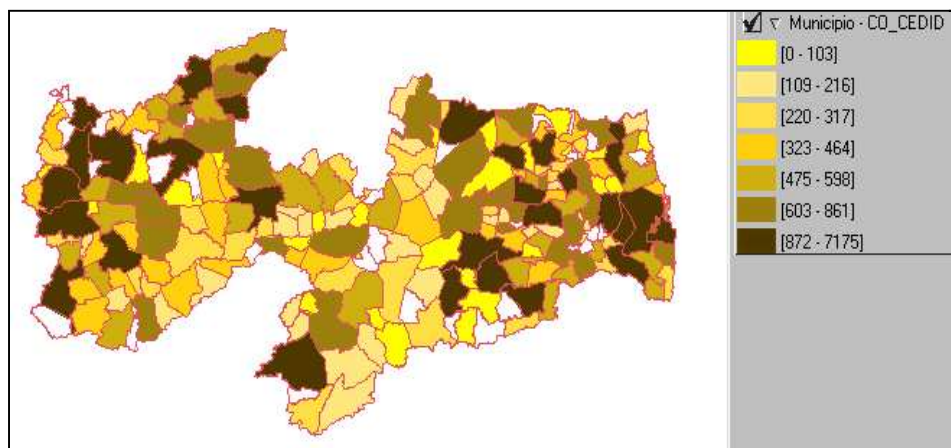


Fig. 29 – Distribuição da população residente em domicílios cedidos, no ano de 1991.

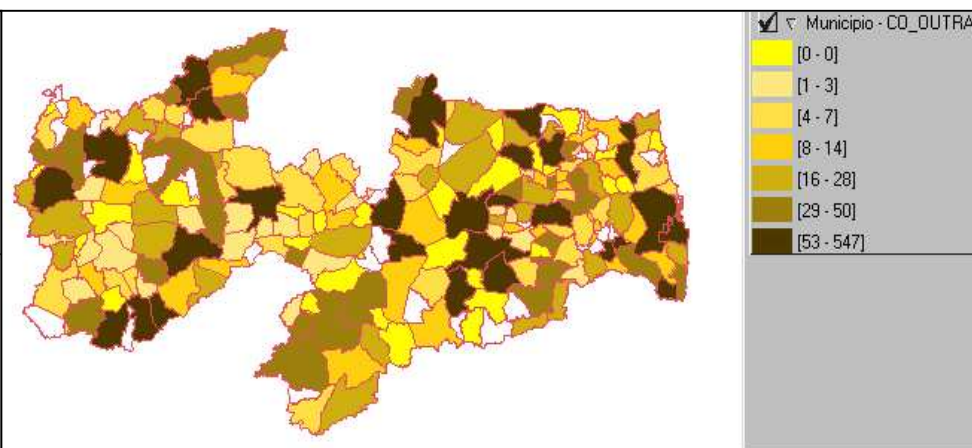


Fig. 30 – Distribuição da população residente, em 1991, em domicílios que não sejam alugados, próprios ou cedidos.

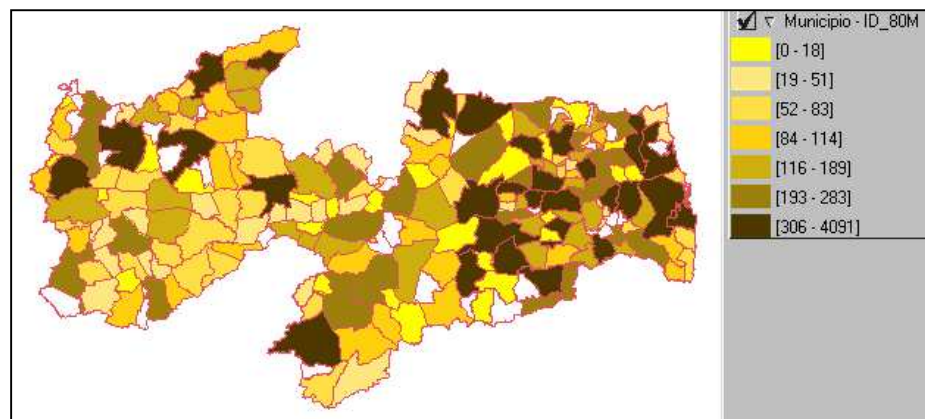


Fig. 31 – Distribuição da população em 1991 com oitenta anos ou mais

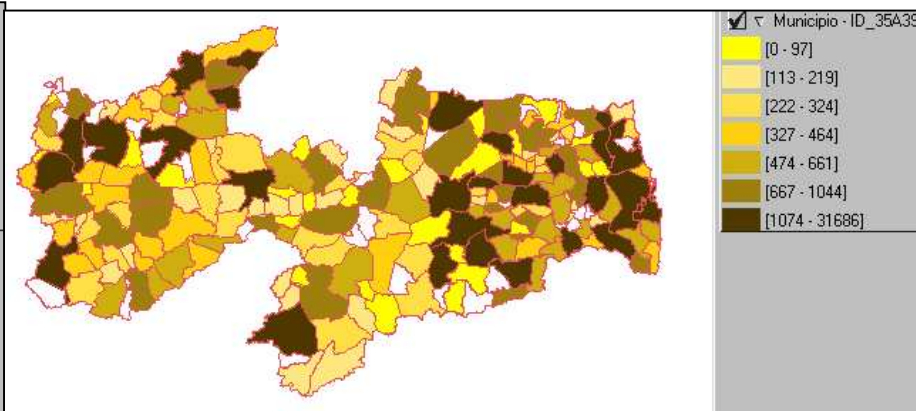


Fig. 32 – Distribuição da população em 1991 com idade de 35 a 39 anos.

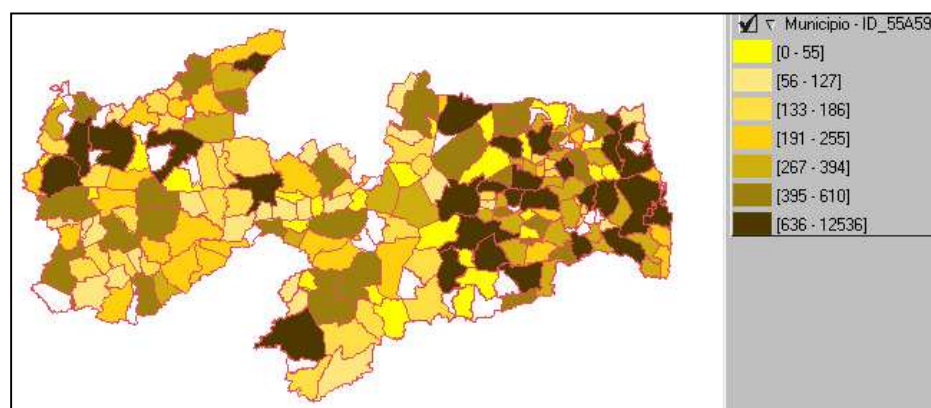


Fig. 33 – Distribuição da população em 1991 com idade de 55 a 59 anos

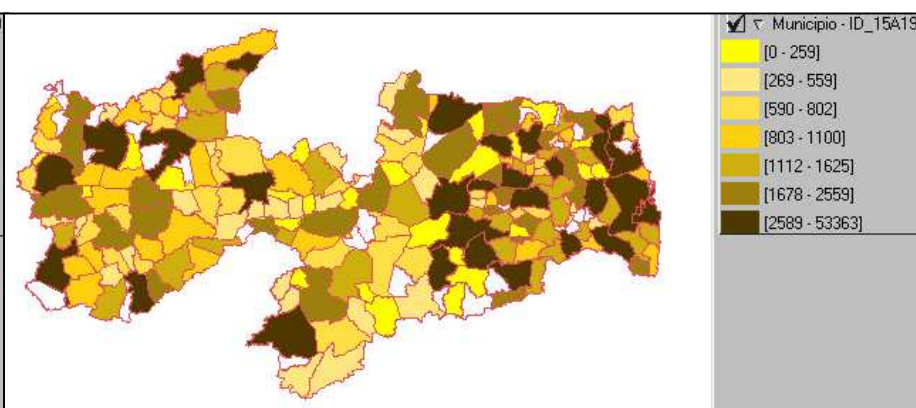


Fig. 34 – Distribuição da população em 1991 com idade de 15 a 19 anos.

Vamos em seguida citar exemplos de utilização prática do nosso banco de dados. Todos as situações dos exemplos são hipotéticas e estão servindo para ilustrar quais as consultas que poderiam ser feitas e quais tipos de consultas são possíveis levando em consideração os dados disponíveis no banco.

3º - Exemplo (hipotético) com 2 variáveis

Uma empresa, instalada no estado da Paraíba, que trabalha com a distribuição de produtos, tinha em seu estoque uma quantidade de lâminas de barbear disponíveis ao mercado. Mas, devido ao resumido quadro de representantes dos seus produtos, precisou organizar de forma mais produtiva suas vendas no estado. A empresa decidiu que as vendas deveriam ser concentradas em alguns municípios. Esses municípios deveriam possuir uma população com características que satisfizessem o perfil do consumidor do produto.

Os critérios estabelecidos para a escolha das cidades que seriam visitadas pelos vendedores foram os seguintes: seriam visitadas aquelas cidades onde a população que ganhasse até ¼ de salário mínimo recebesse em média mais de Cr\$ 2.000,00 (moeda vigente no ano de 1991, ano do Censo considerado) e que o número de pessoas do sexo masculino (que seriam as principais consumidoras do produto) residentes na cidade, estivesse entre 5.000 e 15.000 pessoas. A expressão lógica elaborada e utilizada nos procedimentos do item 6 para elaboração de um mapa que seleciona as cidades com esse perfil, foi a seguinte:

R91_A14>2000 AND
SEX_MASC>5000 AND
SEX_MASC<15000.

No mapa apresentado na figura 35, está a representação da distribuição espacial das cidades que possuem uma tendência em ser mercado consumidor para o produto em questão.

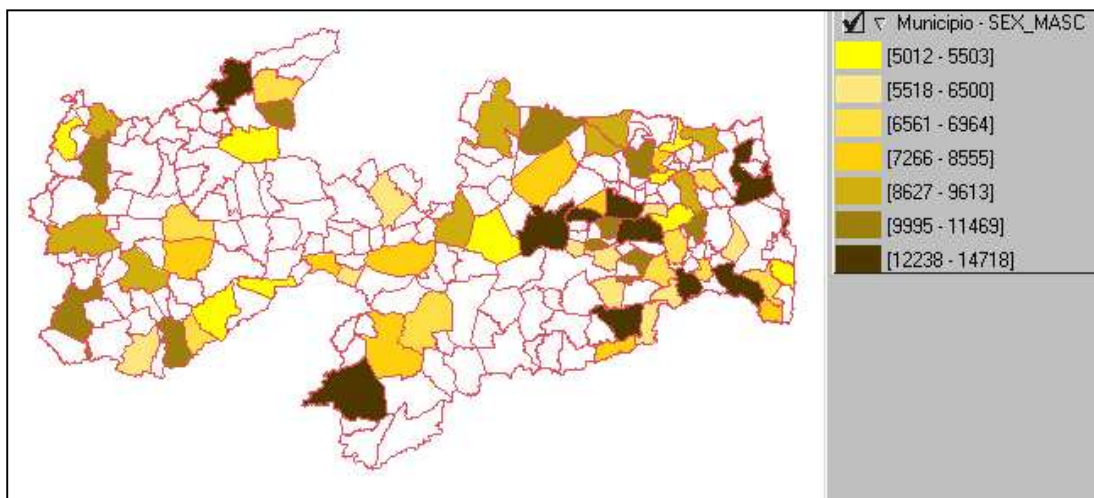


Figura 35 – Seleção dos municípios do estado com as características que satisfazem a expressão elaborada.

4º - Exemplo (hipotético) com 2 variáveis

Uma fábrica de brinquedos, para fugir de encargos fiscais, precisou vender sua produção dentro do próprio estado. Para isso, fez uma pesquisa utilizando o banco de dados, identificando quais as cidades que seriam incluídas no roteiro de vendas dos produtos. Essas cidades deveriam ter uma população maior de 7.500 crianças na faixa etária em que o produto é indicado, ou seja, de 10 a 14 anos e onde a renda média das famílias que recebem entre 10 a 15 salários mínimos seja de Cr\$ 400.000,00 (moeda vigente no ano de 1991). A expressão utilizada que seleciona os municípios com essas características foi a seguinte:

R91_10 A15>400.000 AND
ID_10 A14>7.500

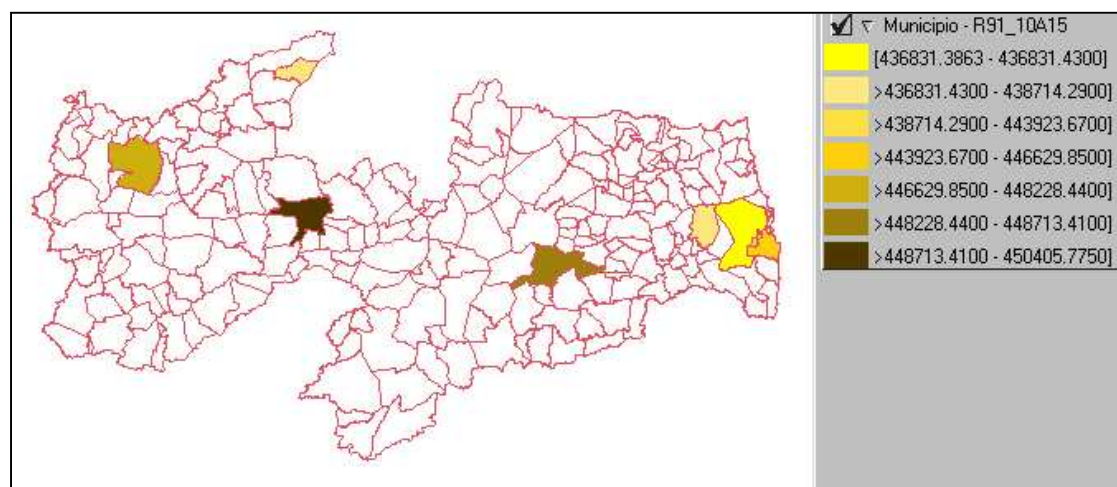


Figura 36 – Municípios com as características especificadas a partir da expressão lógica elaborada.

No mapa da figura 36, observamos que cada faixa de cor corresponde a um único município. Apenas 7 municípios do estado contemplam as especificações da empresa. São eles: João Pessoa, Santa Rita, Sapé, Campina Grande, Patos, Cajazeiras e São José do Brejo do Cruz.

5º - Exemplo (hipotético) com 3 variáveis

O Governo do Estado está elaborando um projeto social que abrange diversas áreas como, saúde, educação, promoção de emprego e renda, etc. Na parte que diz respeito à educação, uma das medidas é uma campanha de alfabetização que seria voltada principalmente às populações da área rural dos municípios. Os critérios estabelecidos para uma primeira seleção das cidades que seriam contempladas com o programa, seriam os seguintes: as cidades precisavam apresentar um baixo índice de alfabetização de adultos, ou seja, aquelas que possuísem menos de 20 pessoas alfabetizadas com esse método. Além disso, deveriam ter uma população com menos de 400 pessoas com o pré-escolar completo, como também, possuir um número maior de 600 famílias morando na zona rural. A expressão lógica usada para geração de um mapa onde poderemos visualizar os municípios que se enquadram nesses critérios foi a seguinte:

ESC91_ALAD<20 OR
ESC_91PRE<400 AND

SIT_RURA>600.

A figura 37 representa espacialmente os dados e nos mostra uma quantidade considerável de municípios que se enquadram com as características especificadas. Podemos identificar, analisando os dados e observando o mapa, a presença de uma quantidade significativa de população rural e um alto índice de analfabetismo no estado.

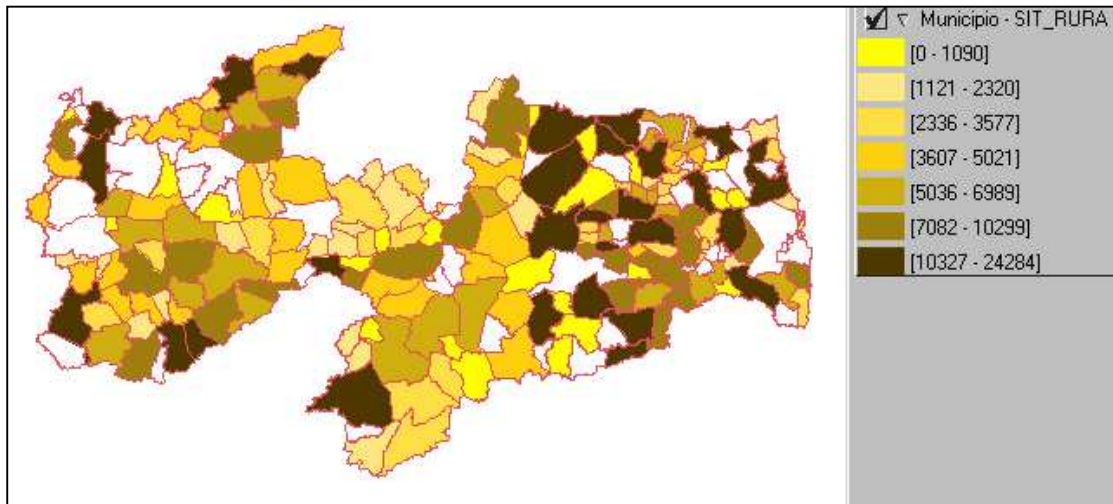


Figura 37 – Visualização dos municípios que se enquadram dentro das especificações estabelecidas na expressão lógica.

6º Exemplo (hipotético) com 3 variáveis

Com o aumento considerável nos casos de gravidez na infância, a Secretaria de Saúde do Estado está promovendo uma campanha de prevenção à maternidade infantil em todo o estado. Em alguns municípios porém, essa campanha vai ser mais intensificada. Para isso, ficou estabelecido que os municípios seriam escolhidos de acordo com alguns critérios. Esses critérios deveriam selecionar quais as cidades que teriam uma campanha mais cuidadosa. Essas cidades teriam que ter uma população entre 50.000 e 250.000 habitantes, com um número de habitantes do sexo feminino maior que 20.000, além de ter uma população com idade entre 10 e 14 anos maior que 5.000. A expressão elaborada para ilustrar num mapa os municípios com essas características foi a seguinte:

POPUL_>50.000 AND
POPUL_<250.000 AND
SEX_FEMI>20.000 AND
ID_10 A14>5.000

O mapa da figura 38 ilustra um fenômeno particular na distribuição populacional do estado, que é o fato de não haver cidades que apresentem população entre 94.413 e 250.000 habitantes, ilustrando, dessa forma, uma disparidade na distribuição populacional do estado.

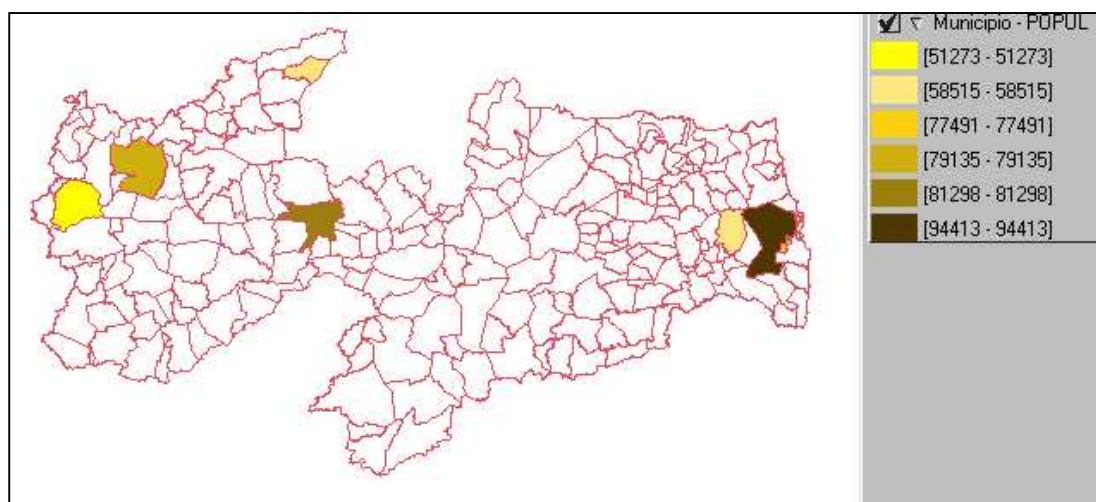


Figura 38 – Municípios selecionados com a utilização da expressão lógica onde haverá uma campanha de prevenção à maternidade infantil mais intensa.

7º Exemplo (hipotético) com 4 variáveis

A consulta ao banco de dados pode auxiliar o Governo no desenvolvimento de um programa de geração de emprego e renda para cidades da Paraíba. Para isso foi feita uma consulta, onde foram selecionados alguns municípios que apresentavam alguns indicativos que evidenciavam a necessidade de implementação do programa. As características que foram definidas para que a cidade se enquadrasse nas condições foram: possuir uma população que ganhasse até $\frac{1}{4}$ de salário mínimo recebendo em média Cr\$ 6.000,00 (moeda vigente no ano de 1991) e apresentasse uma situação de urbanização onde os índices estejam entre 4.000 e 7.000 famílias residindo na área urbana, além de um número maior de 50 e menor que 700 casas alugadas e um número entre 200 e 1.000 unidades habitacionais cedidas. A expressão que podemos utilizar para a geração do mapa foi a seguinte:

```
R91_A1/4<6.000 AND
SIT_URBA>1000 AND
SIT_URBA<9000 AND
CO_ALUGA>50 AND
CO_ALUGA<700 AND
CO_CEDID>200 AND
CO_CEDID<1000
```

Após a definição dos critérios, a elaboração da expressão e a sua utilização nos procedimentos para a geração de mapas, foi obtido um mapa temático representado na figura 39, onde estão os municípios com maior aptidão para implantação do programa. Os critérios escolhidos para a consulta ao banco de dados exemplificam uma forma de encontrar tendências para uma melhor alocação de recursos e equipes de trabalho para realização de qualquer tipo de projeto.

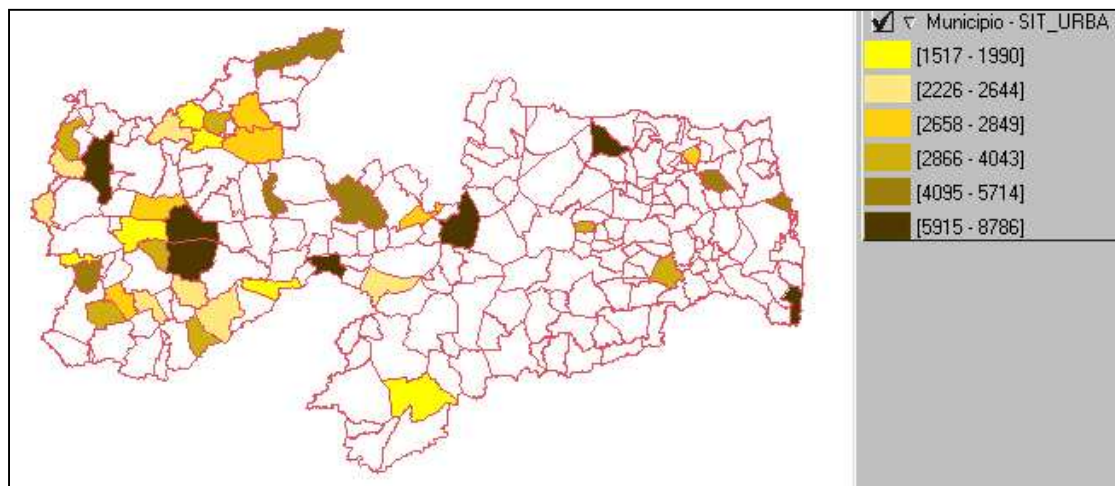


Figura 39 – Municípios selecionados a partir da expressão lógica elaborada para definição de tendências para implantação do projeto

8º Exemplo (hipotético) com 4 variáveis

Em todo o estado a defasagem na quantidade de habitações para as populações de baixa renda fez com que o Governo desenvolvesse um projeto de política habitacional sob forma de mutirão. O programa teria um caráter descentralizador onde a verba destinada seria repassada às prefeituras. Na primeira etapa do projeto, fez-se um levantamento preliminar para selecionar os municípios que, a princípio, seriam assistidos. A consulta foi feita utilizando o banco de dados e as condições determinadas para inclusão da cidade no programa foram as seguintes: as cidades deveriam ter uma população maior que 1.000 pessoas com faixa etária de 25 a 35 anos, além de ter uma população maior que 400 famílias morando em casas cedidas e população maior que 200 famílias morando em casas alugadas. A expressão utilizada nos procedimentos do item que nos orienta como realizar a geração de mapas, foi a seguinte:

```
ID_25 A29>1.000 OR
ID_30 A34>1.000 AND
CO_CEDID>400 OR
CO_ALUGA>200
```

O mapa resultante, representado na figura 40, serve como suporte na decisão a ser tomada para solução dos problemas habitacionais do estado, na medida em que nos dá uma visão espacial dos municípios que, a princípio, precisam de soluções urgentes para o problema habitacional. Estando, portanto, dentro das especificações estabelecidas pelos técnicos do governo.

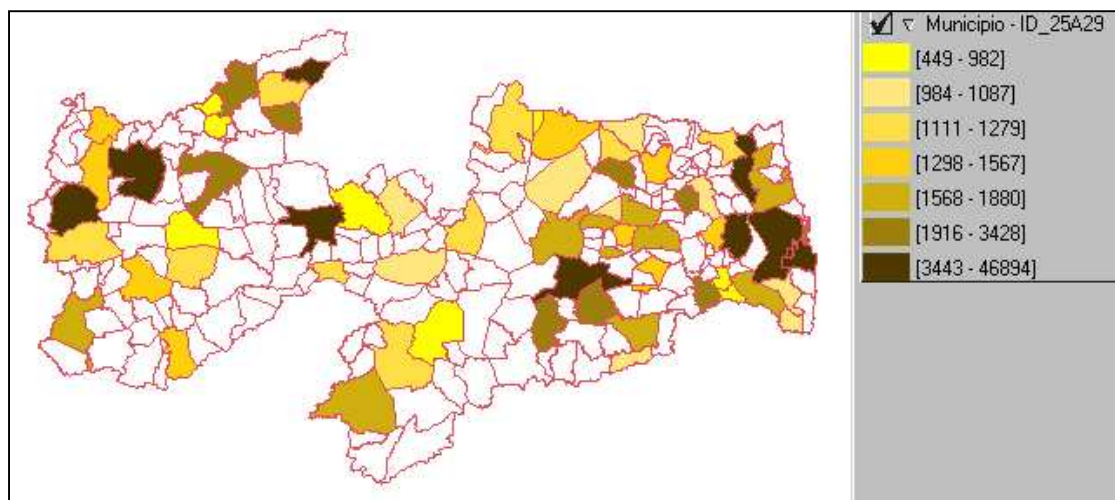


Figura 40 – Municípios selecionados para a implementação do projeto de mutirão habitacional

9º Exemplo (hipotético) com 4 variáveis

Uma editora evangélica publicou uma revista voltada principalmente para etnias consideradas minoritárias. Resolveu distribuí-la em cidades do estado e para descobrir em quais municípios a revista teria uma boa aceitação, fez uma consulta ao banco de dados. Para isso, considerou que as cidades mais propícias à leitura daquela revista, seriam aquelas que apresentassem uma população evangélica maior que 150 pessoas, uma população de cor preta maior que 150 e das demais etnias uma população maior que 20 habitantes. Para gerar o mapa que representasse a distribuição espacial das cidades com essas características foi elaborada a seguinte expressão lógica:

```
REL_EVAN>150 AND
CR_PRETA>140 OR
CR_AMARE>20 OR
CR_INDIG>20
```

O mapa da figura 41 representa a distribuição dos municípios com tendência a leitura desta revista evangélica. Podemos notar uma tendência mais significativa nas cidades com uma população maior, o que pode ser o indicativo de uma forte influência da Religião Católica em pequenas cidades do estado.

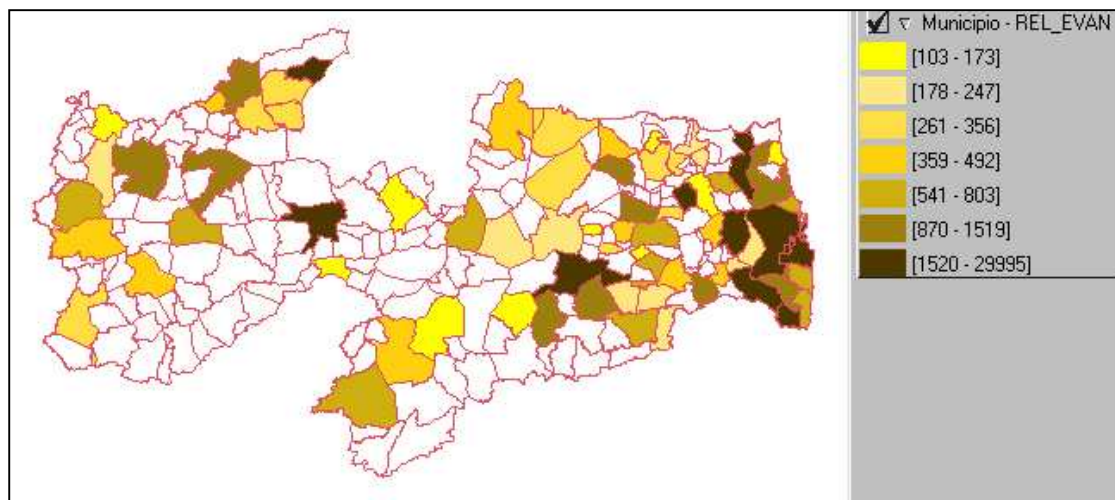


Figura 41 – Seleção dos municípios com tendência a consumo de revistas evangélicas

10º Exemplo (hipotético) com mais de 4 variáveis

Com o número crescente de casos de câncer em mulheres na Paraíba, a Secretaria de Saúde do Estado desenvolveu uma campanha voltada à saúde da mulher. Nessa campanha serão realizados, além de palestras informativas, exames preventivos e vacinação contra o tétano. Para facilitar a seleção preliminar dos municípios que serão assistidos pela campanha foi feita uma consulta ao banco de dados, onde puderam ser visualizados os municípios que estavam dentro das características estabelecidas pelos técnicos da secretaria de saúde. Essas características deveriam ser: possuir uma população do sexo feminino na faixa etária entre 15 e 55 anos, com a preferência que resida na zona rural. O banco de dados foi consultado utilizando a expressão a seguir para viabilizar a geração e a visualização, no mapa, dos municípios que serão atendidos pela campanha:

```
SEX_FEMI>0 AND
ID_15 A19>0 OR
ID_20 A24>0 OR
ID_25 A29>0 OR
ID_30 A34>0 OR
ID_35 A39>0 OR
ID_40 A44>0 OR
ID_45 A49>0 OR
ID_50 A54>0 AND
SIT_RURA>5.000
```

O mapa da figura 42 foi gerado a partir da expressão acima e representa a distribuição dos municípios que serão assistidos pelo programa de saúde da mulher.

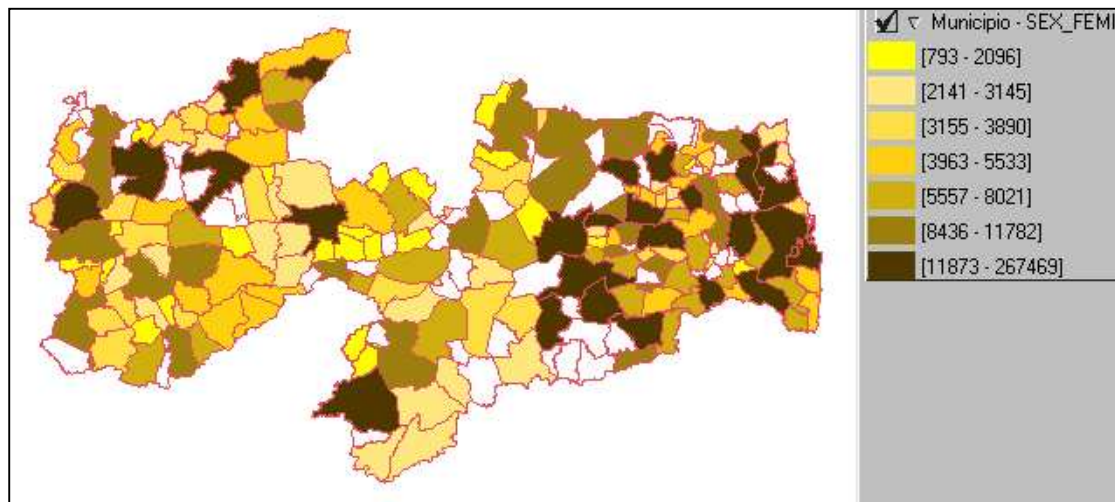


Figura 42 – Municípios selecionados para implantação do programa de saúde da mulher.

4.5 - A análise exploratória que realizamos com o nosso banco de dados

Na realização da análise exploratória de dados, efetuamos a estatística descritiva para todas as variáveis do banco, gerando medidas de tendência central e variabilidade, como a média, moda, mediana e desvio padrão. E para cada variável, geramos também, os gráficos de representação, como histogramas e *qqplots*.

TABELA 04 – Estatística descritiva de algumas variáveis

		POPUL	R91_A14	ES91_SEGU	CO_PROPR	ID_25A29
N	Válidos	223	192	192	192	192
	Omitidos	0	31	31	31	31
Média		14354,77	5494,4010	294,40	2556,46	1234,55
Mediana		7257,00	6136,0350	75,50	1241,50	512,50
Moda		0	,00	0	0	0
Desvio Padrão		41246,99	2016,0453	1576,57	7124,22	4014,51
Intervalo		497600	7364,38	19918	81105	46894
Mínimo		0	,00	0	0	0
Máximo		497600	7364,38	19918	81105	46894
Soma		3201114	1054925,00	56524	490840	237034

Podemos observar pelos valores da tabela 04 que a média da população do estado é de 14354,77; apresentado uma mediana que equivale a metade da média e um desvio padrão alto, indicando uma disparidade na distribuição da população no estado.

Na variável renda, a mediana possui valor próximo à média, um desvio padrão alto e diferença entre os valores máximo e mínimo identificando grandes diferenças salários dentro dessa faixa de renda.

A variável segundo grau completo, apresenta uma média de 294,40 indivíduos, muito distante da mediana e um desvio padrão alto. O número máximo de pessoas que estão cursando 2º Grau é de 19.918 e está em João Pessoa, demonstrando um afunilamento da educação, entre o primeiro e segundo graus no estado em 1991.

Na variável condição de ocupação, onde temos a quantidade de famílias residindo em domicílios próprios, o valor da média não se encontra próximo ao da mediana e o desvio padrão se apresenta alto, indicando uma assimetria da distribuição .

Na distribuição da idade de 25 a 29 anos, observamos que a mediana tem o valor que eqüivale a metade da média e um desvio padrão alto, chegando a ser quase três vezes maior que a média, nos mostrando que a variável não segue uma distribuição normal.

O histograma da figura 43, gerado pelo SPSS, representa a variável renda até ¼ de salário mínimo, apresentando uma distribuição normal (se excluirmos os valores nulos), com os valores concentrados situados entre Cr\$ 5.000,00 e 7.000,00

No qqplot da figura 44, gerado pelo SPSS, temos a representação da distribuição das idades de 25 a 29 anos, onde podemos observar os valores fora da reta do gráfico, evidenciando que a variável não segue uma distribuição normal.

Para a realização da classificação hierárquica, realizamos a normalização dos dados, através transformação logarítmica e a padronização das variáveis do nosso banco de dados. A normalização e a padronização dos dados foi feita, porque algumas variáveis eram incompatíveis quanto a unidade e a grandeza (por exemplo: unidades monetárias e contagem). O método de análise de agrupamento que foi utilizado é medido pela distância euclideana ao quadrado e tem como objetivo reunir dentro de um mesmo agrupamento elementos que tenham características semelhantes. Durante a realização da classificação houve a necessidade da exclusão de 16 variáveis do banco de dados, bem como dos 52 municípios criados após o Censo de 1991 e do município de Tacima, por apresentarem carência de dados, ou seja, muitas variáveis com valor nulo. Além desses, outros 16 municípios foram excluídos pelo próprio método de agrupamento.

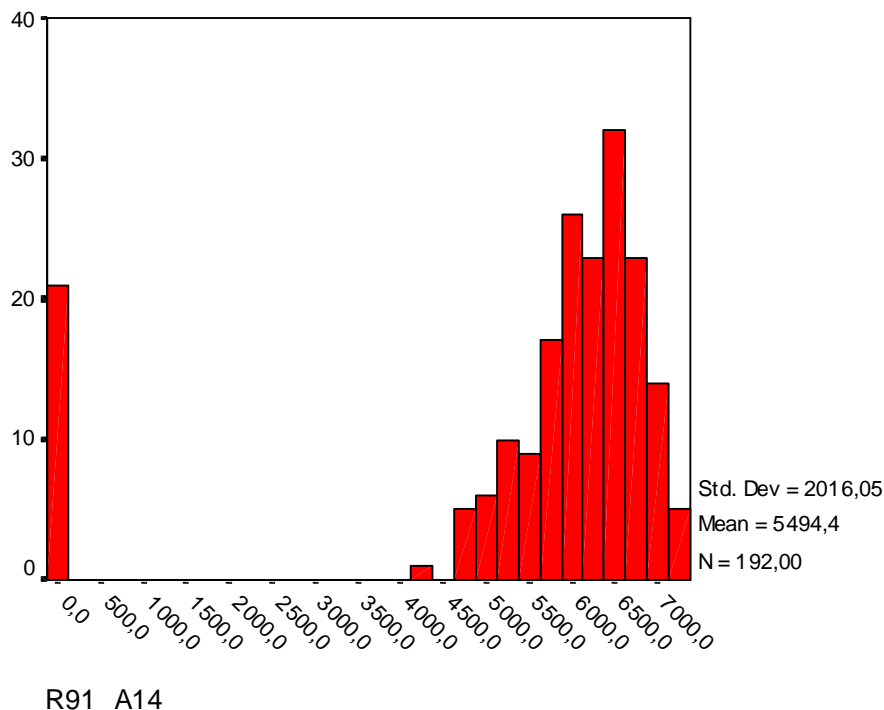


FIGURA 43 – Histograma que representa a variável renda até 1/4 de salário mínimo

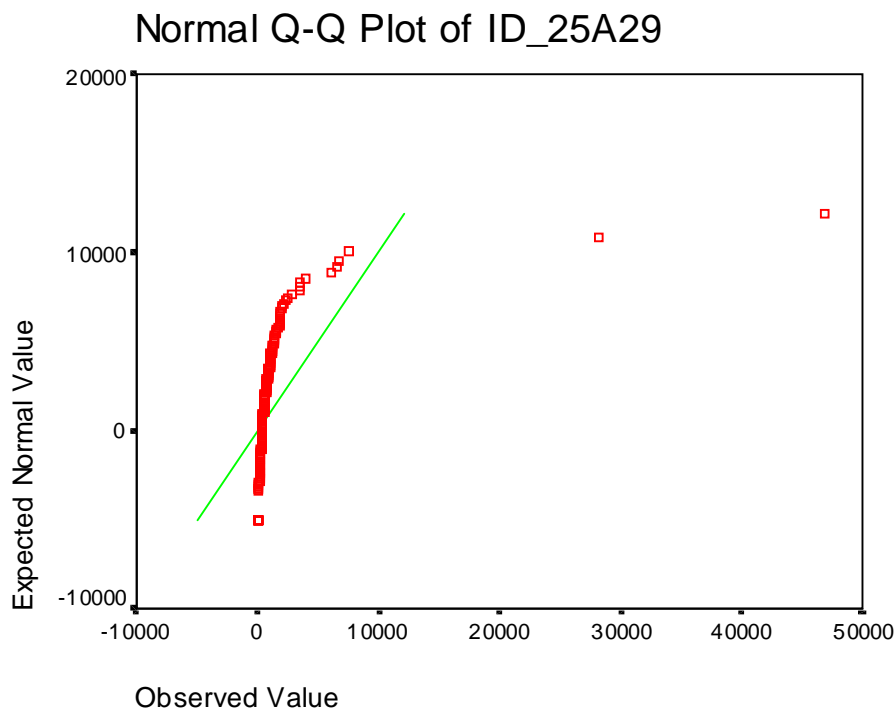


FIGURA 44 – Qqplot que representa a distribuição das idades de 25 a 29 anos

As 16 variáveis excluídas da análise por apresentarem poucas informações por município são as seguintes:

CR_AMARE	População amarela
CR_INDIG	População indígena
CR_SDE	População sem declaração
ES91_ALAD	Alfabetização de adultos
ES91_MEDO	População com mestrado ou doutorado
ES91_SUPE	População com grau superior
R91_15 A20	Renda média de 15 a 20 salários mínimos
R91_10 A15	Renda média de 10 a 20 salários mínimos
R91_M20	Mais de 20 salários mínimos
R91_SDE	sem declaração
R91_NÃO	Sem rendimentos
REL_ESPI	Religião espírita
REL_OUTR	Outras religiões
REL_SDE	Sem declaração
SIT_RURA	Situação rural
CO_OUTRA	Outras condições de ocupação

Foram retirados da análise os municípios criados após o Censo de 1991, mesmo constando nos mapas digitalizados (SPRING de 1996 e TAB WIN de 1997). Os municípios novos são:

Alcantil, Algodão de Jandaíra, Amparo, Aparecida, Areia de Baraúna, Assunção, Baraúnas, Barra de Santana, Bernadino Batista, Boa Vista, Cacimba, Cajazeirinhas, Capim, Caraúbas, Casserengue, Caturité, Coxixola, Cuité de Mamanguape, Curral de Cima, Damião, Gado Bravo, Logradouro, Marcação, Marizópolis, Matinhas, Mato Grosso, Maturéia, Parari, Pedro Régis, Poço Dantas, Poço José de Moura, Riacho de Santo Antônio, Riachão do Bacamarte, Riachão do Poço, Riachão, Santa Cecília de Umbuzeiro, Santa Inês, Santarém, Santo André, Sertãozinho, Sobrado, Sossego, São Bento de Pombal (São Bentinho), São Domingos de Pombal, São Domingos do Cariri, São Francisco, São José de Princesa, São José do Brejo do Cruz, São José dos Ramos, Tenório, Vieirópolis, Zabelê.

Foram excluídos da análise pelo método de agrupamento, os seguintes municípios:

Bom Jesus, Camalaú, Carrapateira, Lastro, Montadas, Olho d'Água, Pilõezinhos, Salgadinho, São José de Caiana, São José dos Cordeiros, Seridó, Serra da Raiz, Serra Grande, Serra Redonda, Várzea, Vista Serrana.

Deste modo, realizamos uma análise de agrupamento (*clusters*) no SPSS, sendo obtidos sete agrupamentos (onde cada agrupamento é composto por municípios que foram classificados como possuindo características semelhantes) e dois agrupamentos de municípios excluídos, totalizando nove agrupamentos.

Geramos então, um mapa onde pode ser observado o comportamento espacial dos *clusters* no estado da Paraíba. Este mapa foi gerado através do Tabwin e pode ser observado na figura 45.

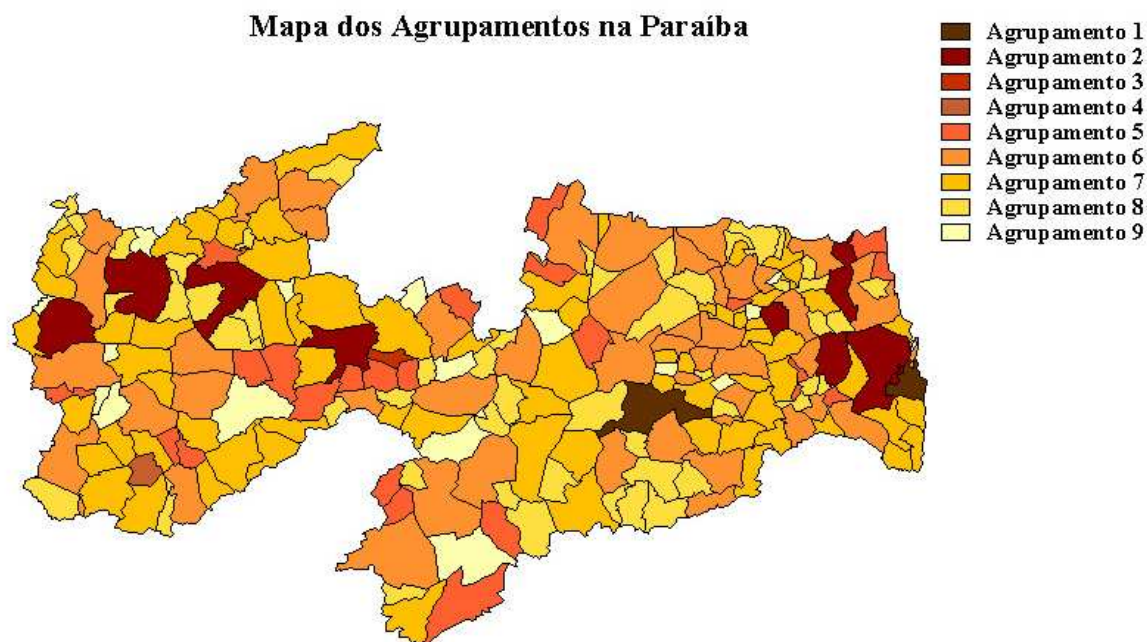


FIGURA 45 – Mapa dos nove agrupamentos no estado da Paraíba

Agrupamento 1 - Incluídos 2 municípios : João Pessoa e Campina Grande.

No agrupamento 1 estão os 2 maiores municípios do estado que se identificam como sendo as principais cidades do estado e apresentam um desenvolvimento sócio-econômico semelhante.

Agrupamento 2 - Incluídos 9 municípios: Sousa, Patos, Sapé, Mamanguape, Pombal, Guarabira, Cajazeiras, Santa Rita, Bayeux ;

O agrupamento 2 é composto dos principais municípios do estado, depois da capital e da cidade de Campina Grande. São municípios que estão localizados próximos às principais rodovias do estado, a BR 230 e a BR 101 e os municípios se localizam, principalmente no litoral e no sertão. Se destacam por serem cidades que possuem uma boa condição sócio-econômica em relação a outros municípios com população inferior a deles.

Agrupamento 3 - Incluído 1 município: Quixabá;

Agrupamento 4 - Incluído 1 município: Curral Velho;

O agrupamento 3 e 4 é composto por municípios que possuem as menores populações do estado, sendo Quixabá o 1º e Curral Velho o 4º menor em população no ano de 1991. Os dois municípios diferenciam-se por apresentarem rendas semelhantes à cidades de porte médio. No caso de Curral Velho, com 2.537 habitantes, por exemplo, sua renda média até ¼ de salário mínimo é comprável à de municípios com população de mais de 50.000 habitantes.

Agrupamento 5 – Incluído 23 municípios: Prata, Nova Olinda, Lagoa, Emas, São João do Tigre, Frei Martinho, Cacimba de Areia, Passagem, Monte Horebe, Nova Palmeira, Congo, São José do Bonfim, Ouro Velho, Mataraca, São José do Sabugi, Catingueira, São Miguel de Taipu, Mãe d'Água, Caldas Brandão, Borborema, Pedra Branca, Oivedos, Baía da Traição.

O agrupamento 5 possui características bem predominantes, com população entre 3.129 a 6.198 habitantes. Vários de seus municípios estão localizados bem próximo às divisas do estado, onde existem rodovias estaduais de acesso a outros estados. Outra característica, é que 6 dos seus 23 municípios estão localizados próximos ao município de Patos, bem como da BR 230.

Agrupamento 6 – Incluídos 46 municípios: Cabedelo, Cacimba de Dentro, Juazeirinho, Uiraúna, São João do Rio do Peixe, Conceição, Jacaraú, Araçagi, Piancó, Itaporanga, Sumé, Serra Branca, Santa Luzia, Teixeira, São José de Piranhas, Coremas, Pilar, Pocinhos, Brejo do Cruz, Taperoá, Remígio, Belém, Barra de Santa Rosa, Umbuzeiro, Araruna, Aroeiras, Queimadas, Rio Tinto, Itabaiana, Pedras de Fogo, Areias, Marí, Ingá, Princesa Isabel, São Bento, Picuí, Lagoa Seca, Cuité, Bananeiras, Alagoa Nova, Esperança, Catolé do Rocha, Boqueirão, Solânea, Monteiro, Alagoa Grande;

O agrupamento 6 é composto por municípios com uma população que varia de 13.773 a 33.255 habitantes com um número razoável de crianças cursando o primeiro grau (entre 2460 a 6.214). Estes municípios encontram-se espalhados por todo o estado.

Agrupamento 7 – Incluídos 72 municípios: Água Banca, Juru, Pitimbu, Lucena, Juarez Távora, Juripiranga, Cuitegi, Caaporã, Conde, Alhandra, Mulungu, Fagundes, Mogeiro, Massaranduba, Itapororoca, Santana dos Garrotes, São Mamede, Jericó, Paulista, Belém do Brejo do Cruz, Gurinhém, Imaculada, Cruz do Espírito Santo, Triunfo, Tavares, Desterro, Salgado de São

Félix, Arara, Soledade, Natuba, Manaíra, Pirpirituba, Puxinanã, Caiçara, São Sebastião de Lagoa de Roça, Itatuba, Serraria, Dona Inês, Alagoinha, Riacho dos Cavalos, Diamante, Santana de Mangueira, Santa Helena, Cubati, São Sebastião do Umbuzeiro, Junco do Seridó, Pedra Lavrada, Condado, São José de Espinharas, Santa Terezinha, Barra de São Miguel, São José da Lagoa Tapada, Cachoeira dos Índios, Bonito de Santa Fé, Bom Sucesso, Livramento, Boa Ventura, Pilões, Nazarezinho, São João do Cariri, Duas Estradas, Gurjão, Cabaceiras, Santa Cruz, Brejo dos Santos, Areial, Lagoa de Dentro, Nova Floresta, Malta, Ibiara, Igaracy, Aguiar.

O agrupamento 7 é o *cluster* com maior número de municípios. Seus municípios possuem uma população entre 4.448 a 15.007 habitantes com um número pequeno de crianças cursando o primeiro grau (entre 838 e 2.757). Quanto à localização de seus municípios, não se apresentam aglomerados, estando distribuídos por todo estado

O agrupamento 8 é formado pelos municípios criados após 1991. Esses municípios foram excluídos da análise por apresentarem valores nulos em várias das suas variáveis. Notamos que a microrregião do Cariri, com exceção e alguns municípios dos grupos 5, 6 e 7, é formada quase que totalmente por municípios pertencentes ao agrupamento 8.

No agrupamento 9 temos os municípios excluídos pelo método de agrupamento hierárquico, por apresentarem valores indeterminados em algumas das variáveis transformadas pelo logaritmo neperiano.

Geração da página na Internet

Utilizamos o Netscape Composer para gerar a página html do projeto contendo os relatórios finais da iniciação científica – PBIC de 1998/1999 e de 1999/2000. Essa página encontra-se disponível em <http://www.de.ufpb.br/~ronei/monica/>

Conclusões finais

Implementamos, na primeira etapa do projeto, um banco de dados da Paraíba no Sistema SPRING, que continha inicialmente, um mapa digitalizado do estado, dados sobre população residente de cada município, renda por faixa de salário mínimo e escolaridade por grau.

Nesta etapa do projeto estudamos, dentro de um SIG, a melhor forma de ampliação do um banco de dados e inserimos, no banco de dados da Paraíba, outras variáveis, totalizando agora 62 variáveis inseridas. Utilizando as variáveis do banco realizamos consultas ao banco de dados, com a geração de mapas onde podemos visualizar como se distribui espacialmente cada variável no estado. Também geramos, a partir de expressões lógicas, mapas com novas informações resultantes do cruzamento entre diferentes variáveis.

Fizemos estudos sobre análise espacial com a leitura de trabalhos realizados com a aplicação da análise espacial. A análise exploratória do banco de dados foi feita com a realização da estatística descritiva das variáveis, onde foi efetuado o cálculo de medidas de tendência central e variabilidade, além de obtermos também histogramas e *qqplots*. Efetuamos transformações nas variáveis e realizamos uma classificação hierárquica utilizando o método de ligação média entre os grupos medido pela distância euclidiana ao quadrado. Como resultado da classificação foram obtidos nove clusters que foram representados sob a forma de um mapa gerado no Tabwin. E por fim realizamos a geração da página na Internet com os dados do nosso trabalho.

Bibliografia

Assunção, R.M.; *Estatística Espacial com aplicações em epidemiologia*. 31ª Reunião Regional, Associação Brasileira de Estatística, Natal, 1999.

Bussab, W. O. e Morettin, P. A .; *Estatística básica*. 4ª edição, São Paulo: Atual, 1987.

Câmara, G. e Medeiros, J.S.; *Geoprocessamento para Projetos ambientais*. VIII Simpósio Brasileiro de Sensoriamento Remoto, 1996.

Câmara, G; *Mineração de dados espaciais: a busca de pepitas de ouro*. Obtido em <http://www.dpi.inpe.br/gilberto/infogeo9.pdf>, em 26/01/2000.

Câmara, G. et al (ed); *Geoprocessamento – teoria e aplicação*, 1999. Obtido em <http://www.dpi.inpe.br/gilberto/livro/> , em 26/01/2000.

Carvalho, M. S.; *Aplicação de métodos de análise espacial na caracterização de áreas de risco à saúde*. Rio de Janeiro: Tese de Doutorado em Engenharia Biomédica – COOPE/UFRJ, 1997.

Carvalho, M. S. e Cruz, O.G.; *Mortalidade por causas externas – Análise exploratória espacial, Região Sudeste/Brasil*. in anais XI Encontro Nacional de Estudos Populacionais, Caxambu, 19 a 23 de outubro de 1998, CD ROM.

Costa, S. F.; *Introdução ilustrada à Estatística*. 3ª edição, São Paulo: Editora Harbra, 1998.

Instituto Brasileiro de Geografia e Estatística (IBGE); Censo Demográfico Brasileiro de 1991. <http://www.ibge.gov.br/>, 18 de março de 1999.

Instituto Nacional de Pesquisas Espaciais (INPE); Manual do SPRING – v 2.0.4. [online]. <http://www.dpi.inpe.br/~spring/>, 14 de março de 1998.

Johnson, R. A. e Wichern, D. W.; *Applied Multivariate Statistical Analysis*. New Jersey, Prentice Hall, 1992, 3rd. ed.

Teles, M. M. F. e Moraes, R. M.; *Estudos sobre os sistemas de informação geográfica*. Relatório PIBIC/CNPq. João Pessoa: UFPB, 1999.