

# Uma Ferramenta de Visualização de Dados Georreferenciados e o Turismo Inteligente como Estudo de Caso

Anderson C. K. de Menezes, Antônio A. de A. Rocha,  
Verônica Feder Mayer, Alexandre Plastino

UFF - Universidade Federal Fluminense  
IC - Instituto de Computação

Rua Passo da Pátria, 156-Bloco E - 3º andar - Boa Viagem - CEP 24210-240 Niterói(RJ)

anderson\_menezes@id.uff.br; arocha@ic.uff.br; veronicamayer@id.uff.br;  
plastino@ic.uff.br

**Abstract.** *In this article we propose a georeferenced data visualization tool, based on cell phone data that, after applying data mining techniques, more specifically, Association Rules, generates dynamic and interactive visual representations from which users can quickly extract certain mobility patterns, which can contribute to the generation of knowledge. As a case study, the tool was instantiated for application in smart tourism, based on data from cell phone connections of foreigners in the city of Rio de Janeiro and its Metropolitan Region from 05/2020 to 10/2021.*

**Resumo.** *Neste artigo, propomos uma ferramenta de visualização de dados georreferenciados, baseada em dados de telefonia celular que, após a aplicação de técnicas de mineração de dados, mais especificamente, de Regras de Associação, gera representações visuais dinâmicas e interativas nas quais os usuários podem extrair rapidamente certos padrões de mobilidade, contribuindo para a geração de conhecimento. Como estudo de caso, a ferramenta foi instanciada para aplicação em turismo inteligente, com base em dados de conexões de celulares de estrangeiros ocorridas no município do Rio de Janeiro e na sua Região Metropolitana no período de 05/2020 a 10/2021.*

## 1. Introdução

O desenvolvimento das Tecnologias de Informação e Comunicação (TICs) vem permitindo às cidades tradicionais investir em novas iniciativas para solucionar muitos dos problemas decorrentes do crescimento da urbanização a uma taxa sem precedentes [Cortel 2019]. Foi assim que nasceu o conceito de "Cidades Inteligentes" (*Smart Cities*). Neste sentido, segundo [Lim, Kim and Maglio (2018)], cidades em todo o mundo estão tentando se transformar em *Smart Cities* por meio do uso de grande quantidade de dados urbanos.

As grandes massas de dados atualmente geradas contém um rico conhecimento geográfico e podem ajudar a enfrentar muitos desafios urbanos, como a engenharia de tráfego, problemas de poluição do ar, planejamento turístico, estudos de migração, epidemiologia, ecologia, resposta a desastres, entre outros.

No entanto, um dos grandes desafios dos dias atuais é transformar a crescente e dinâmica massa de dados disponível, que por si só não agrega valor, em conhecimento útil. De acordo com [Han, Pei, and Kamber 2012], esse processo pode ser caracterizado como uma “descoberta de conhecimento” ou *KDD (Knowledge Discovery in Databases)*. O mesmo deve seguir uma sequência iterativa de sete etapas, das quais, destaca-se a de “avaliação de padrões”, decisiva na geração do conhecimento, necessitando de alto grau de cognição.

A simples disponibilização de dados tabulados não é o suficiente para a geração do conhecimento, sendo necessária a adoção de meios complementares de análise, principalmente baseados em técnicas de visualização, uma vez que a visão é o principal sentido da percepção humana. Segundo [Madala, Rajavarman and Venkata Satya Vivek 2018], a visualização de dados é o principal conceito em foco na análise de grandes quantidades de dados devido ao crescimento do tamanho e da complexidade dos dados, principalmente quando se trata da avaliação de padrões.

De acordo com [Elmqvist and Yi 2015], a criação de sistemas de visualização de dados se baseia, muitas vezes, mais no processo de tentativa e erro do que na aplicação de métodos já testados e validados. Normalmente, esses sistemas são projetados para estruturar atividades cognitivas de alto nível, como a compreensão de fenômenos específicos, encontrar percepções sobre um problema ou tomar uma decisão diante de dados complexos ou massivos. Assim, existe a necessidade de se propor novas ferramentas de visualização válidas e que possibilitem o isolamento e a caracterização de padrões no processo de geração do conhecimento nos mais diversos contextos.

A utilização de soluções de visualização de dados, com base em dados de posicionamento móvel, vem sendo cada vez mais discutida na literatura. Em [Kang et al. 2010], são propostas representações gráficas 2D ou 3D de padrões de atividades individuais e sua evolução no espaço e no tempo para derivar padrões de mobilidade agregados e individuais de diferentes grupos. Já em [Watson et al. 2020], com uma proposta semelhante, foram extraídos dados de mobilidade de uma rede de celulares no Senegal, cujo objetivo foi demonstrar os diferentes padrões espaço-temporais da mobilidade humana intermunicipal em uma importante cidade no Senegal. Outra proposta interessante é apresentada em [Gavric et al. 2011], que traz uma abordagem para detectar locais atrativos para turistas e sua dinâmica com base em dados obtidos a partir de imagens georreferenciadas, publicamente registradas em Berlim, Alemanha. Os autores sugerem que um local é considerado atrativo se tiver uma grande quantidade de fotos tiradas por muitas pessoas.

Como forma de contribuir com o avanço dessa área, propomos neste trabalho uma ferramenta de visualização de dados georreferenciados provenientes de conexões de celulares, que procura extrair padrões visuais de ocorrência de conexões ou de dispositivos. Pretendemos demonstrar a sua aplicabilidade por meio de um estudo de caso relacionado ao turismo inteligente. Por meio da visualização de mapas interativos, gerados com base na extração de regras de associação, uma técnica de mineração de dados, e outras ferramentas estatísticas, extraímos padrões visuais com grande potencial de fornecer conhecimento útil a ser aplicado no planejamento de ações relacionadas não só ao turismo inteligente, mas também em outras áreas, que poderiam se beneficiar do conhecimento de padrões de mobilidade.

## 2. Proposta de Ferramenta de Visualização Baseada em Dados Georreferenciados

A ferramenta de visualização apresentada neste artigo foi desenvolvida em linguagem Python suportada, principalmente, pelas bibliotecas Pandas e Folium. Teve como fonte de dados um repositório de conexões de celulares de uma operadora (TIM), coletados em cerca de 1350 antenas distribuídas no município do Rio de Janeiro e suas imediações. O volume de dados gerado, que representa aproximadamente 15% do total de conexões na região, é da ordem de 13 GB/dia, ou cerca de 130 milhões de conexões diárias. Os arquivos do repositório possuem os seguintes dados: data e hora da conexão; um identificador anonimizado do dispositivo (atendendo à LGPD); DDD do dispositivo (ausente nos celulares de estrangeiros); *Mobile Country Code (MCC)*, um identificador do país de origem do dispositivo; e, latitude e longitude, coordenadas de georreferenciação da antena que registrou a conexão. Estes dados passam por um processo de transformação com o acréscimo de atributos complementares para enriquecer as informações geradas pelas representações visuais da ferramenta.

Disponibilizamos uma interface gráfica onde é possível escolher qual visualização se deseja executar e o como os dados serão apresentados, por conexões ou por dispositivos. Além disso, estão disponíveis filtros temporais e categóricos, baseados nos atributos dos dados, permitindo uma ampla variedade de combinações para as análises desejadas. Após a escolha das opções, a ferramenta processa milhões de conexões em poucos segundos e renderiza os dados.

O objetivo da ferramenta é oferecer ao usuário uma forma de identificar padrões de conexões ou de dispositivos, por meio de visualizações, que apresentam dados estatísticos, baseados nas escolhas feitas através da interface gráfica. São mapas do tipo coroplético, que destacam as diferenças de magnitude dos dados de cada Região Administrativa da cidade do Rio de Janeiro e de cada Município da sua Região Metropolitana. Assim, as cores das localidades variam conforme o valor da variável estatística retratada. As antenas de celular também são representadas nos mapas, por meio de marcadores circulares que variam de tamanho e de cor, conforme a magnitude dos dados.

As visualizações são interativas, pois quando o usuário desliza o cursor sobre cada área, esta é destacada pela intensificação da tonalidade de sua cor, sendo apresentadas informações estatísticas de interesse, conforme o objetivo de cada mapa. Por sua vez, os marcadores das antenas apresentam informações similares sendo identificados pelas suas coordenadas geográficas. Existe a possibilidade de customização da tela com opções alternativas de apresentação do mapa. Também estão disponíveis botões de zoom para aumentar ou diminuir o detalhamento da exibição.

Foram desenvolvidas três propostas de representação visual, conforme a estratégia de análise pretendida: (i) Magnitude de ocorrências de conexões ou dispositivos; (ii) Análise comparativa de dois conjuntos de dados; e (iii) Representação visual de regras de associação na busca de relacionamentos ou padrões entre conjuntos de dados analisados. Estas estratégias de análise serão mais detalhadas em seguida.

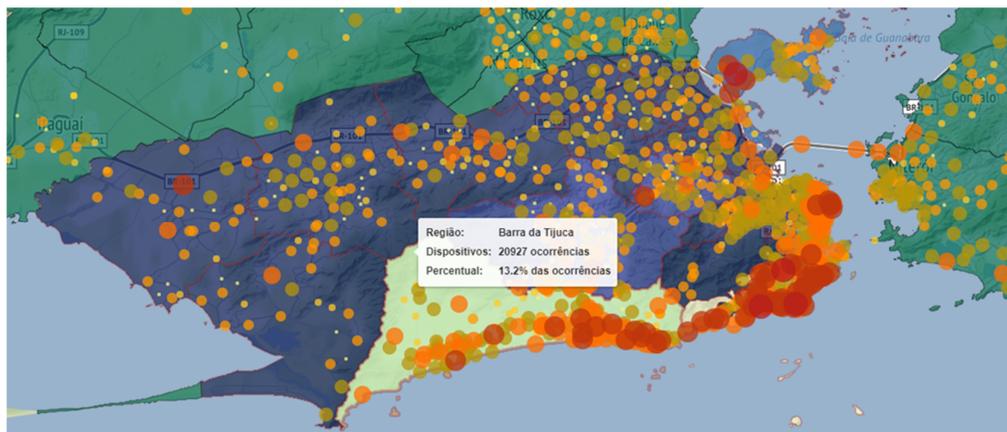
### 2.1. Magnitude de Ocorrências de Conexões ou Dispositivos

A primeira representação visual desenvolvida tem como objetivo identificar os locais que se destacam pela alta concentração de conexões ou de dispositivos, por meio do

mapeamento por Região/Município e por antena de celular. Quando os dados são apresentados por dispositivos, o seu quantitativo é calculado com base na ocorrência de celulares distintos que se conectaram em uma localidade ou em uma antena, independente do número de conexões feitas por cada um. No caso da apresentação por conexões, o seu quantitativo é calculado com base na soma das suas ocorrências.

A diferenciação é estabelecida de acordo com a proporção das ocorrências. Quanto maior a participação no total das ocorrências, mais clara tende a ser a representação da localidade, de acordo com uma escala de cores gerada com base nos dados coletados. No caso das antenas, buscando um contraste com as localidades, estas são representadas por marcadores circulares que variam de cor e tamanho com base em uma escala logarítmica, aumentando o tamanho e a intensidade da cor conforme a magnitude das ocorrências.

A Figura 1 apresenta um mapa retratando o total de dispositivos de estrangeiros que, no período de 05/2020 a 10/2021, se conectaram às antenas localizadas nas Regiões Administrativas da cidade do Rio de Janeiro (em azul) e nos municípios da Região Metropolitana (em verde), com exceção do litoral, destacado com cores mais claras por conta da grande concentração da ocorrência de dispositivos. Nesta Região, a maior parte das antenas está sendo representada com grandes marcadores em tonalidades de um vermelho intenso devido ao grande número de dispositivos com conexões.



**Figura 1. Visualização da ocorrência de dispositivos**

A quantidade de ocorrências, acompanhada do percentual relativo ao total de dispositivos ou conexões da base de dados, aparece no mapa quando se posiciona o cursor sobre cada uma das localidades. Na Figura 1 encontra-se destacado um quadro com as informações da Região da Barra da Tijuca, por exemplo. Por sua vez, os marcadores das antenas apresentam informações sobre suas coordenadas geográficas, o número de ocorrências e a classe de grandeza, com a respectiva faixa de valores a qual pertence.

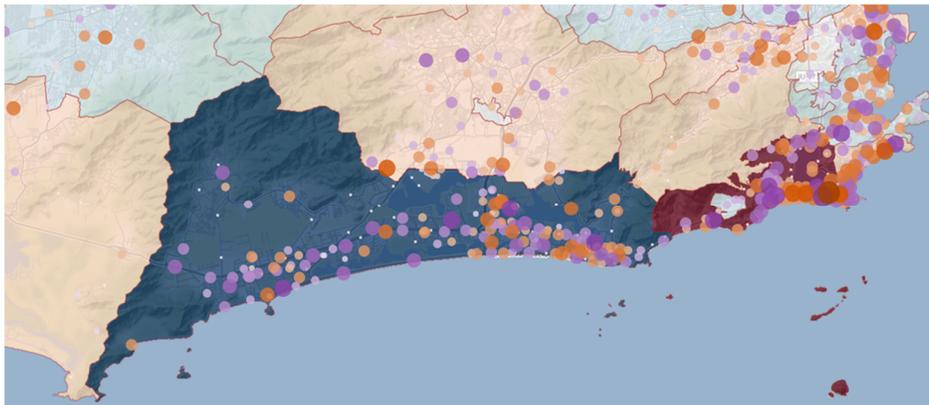
## **2.2. Análise Comparativa de Dois Conjuntos de Dados**

Esta representação visual demonstra a diferença de ocorrências de conexões ou de dispositivos entre duas consultas distintas na base de dados, subtraindo-se a segunda da primeira. Um exemplo prático poderia ser a comparação do volume de conexões de

2020 contra as de 2021. O objetivo é destacar as principais variações para possibilitar a identificação de padrões de mudanças nos dados e possíveis tendências. Mostra-se particularmente útil quando é utilizada na comparação entre períodos temporais distintos, embora seja possível utilizar parâmetros categóricos para a filtragem dos dados.

Nesta representação, as localidades são diferenciadas de acordo com a magnitude e o sinal das diferenças apuradas. Existem duas classes de cores, uma com tons avermelhados para as diferenças negativas e outra com tons azulados para diferenças positivas, cujas intensidades são proporcionais aos valores apurados e à grandeza dos dados coletados. Assim, as diferenças pouco significativas tendem a ser representadas pela cor branca. No caso das antenas, seus marcadores variam em tamanho de acordo com os valores das diferenças, cuja escala é logarítmica. Os valores negativos são representados em tons amarronzados e os positivos em tons arroxeados.

O exemplo da Figura 2 apresenta um mapa retratando a diferença de conexões de estrangeiros às quintas-feiras, no período de 05/2020 a 10/2021, com aquelas ocorridas às sextas-feiras no mesmo período. Estão em destaque as regiões da Barra da Tijuca e da Lagoa, com as maiores diferenças. A primeira, em azul, teve mais conexões às sextas-feiras e a segunda, mais conexões às quintas-feiras. Com base nesta informação, é possível inferir uma possível tendência migratória entre estas localidades, de acordo com o dia da semana, merecendo uma investigação mais aprofundada de suas possíveis causas. Os marcadores em roxo representam as antenas que aumentaram o volume de conexões e aqueles em vermelho, as antenas que diminuíram o número de conexões.



**Figura 2. Visualização da Diferença de Dois Conjuntos de Dados**

### **2.3. Representação Visual de Regras de Associação**

Em mineração de dados, as regras de associação representam relacionamentos entre itens de dados de um mesmo domínio de aplicação que são observados em uma determinada base de dados [Han, Pei, *and* Kamber 2012]. Estas regras possuem a forma  $X \rightarrow Y$ , onde  $X$  (chamado de antecedente) e  $Y$  (denominado conseqüente ou alvo) são subconjuntos disjuntos do conjunto de itens do domínio da aplicação (ou seja,  $X \cap Y = \emptyset$ ). Dentre as métricas utilizadas para avaliar as regras e identificar quais são interessantes se destacam o suporte, a confiança e o Lift.

O suporte, dada uma regra  $X \rightarrow Y$ , representa a porcentagem de registros da base de dados que contém os itens de X e Y, indicando a relevância da mesma. Seu cálculo é dado por:  $Supporte(X \rightarrow Y) = \frac{Freq(X,Y)}{N}$ , onde  $Freq(X, Y)$  é o número de registros onde os itens de X e de Y aparecem e N é número total de registros da base. A medida de confiança representa, dentre os registros que possuem os itens de X, a porcentagem de registros que possuem também os itens de Y, indicando a validade da regra. Sua equação é dada por:  $Confiança(X \rightarrow Y) = \frac{Freq(X,Y)}{Freq(X)}$ , onde  $Freq(X)$  é o número de registros onde os itens de X aparecem.

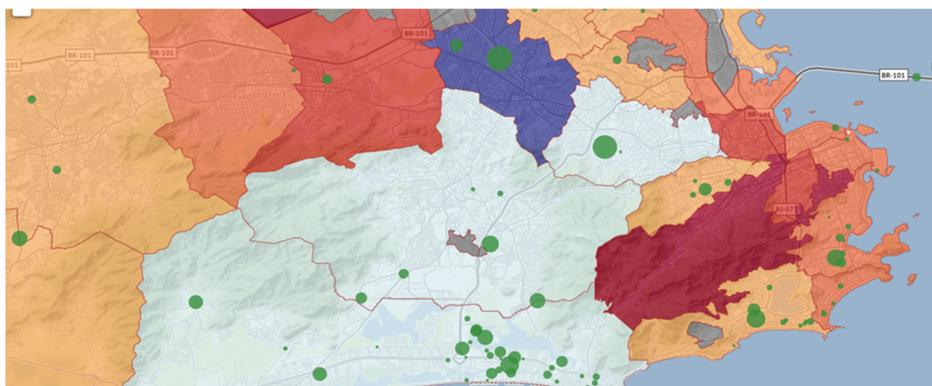
O Lift é uma métrica que compara a probabilidade do consequente da regra ocorrer com a probabilidade condicional do consequente da regra ocorrer dado que o antecedente ocorreu. É obtido pela equação:  $lift(X \rightarrow Y) = \frac{Confiança(XY)}{Supporte(Y)}$ , onde  $Supporte(X \rightarrow Y)$  é o suporte do consequente. Valores acima de 1 (correlação positiva) indicam que a ocorrência do antecedente aumenta as chances da ocorrência do consequente. Em contrapartida, valores abaixo de 1 (correlação negativa) indicam que a ocorrência do antecedente diminui as chances de ocorrência do consequente. Valores próximos de 1 indicam que antecedente e consequente são independentes.

A ferramenta proposta apresenta uma representação visual desta técnica no intuito de identificar se uma localidade aumenta ou reduz as chances de ocorrência de determinado grupo (alvo) quando comparada com a sua participação desse mesmo grupo no total da amostra dos dados. É particularmente útil quando procuramos identificar padrões específicos de correlação ou de comportamento das variáveis, quando consideradas em conjunto, mais especificamente, quando a intenção é identificar rapidamente se uma localidade é um ponto de atração ou de repulsão do alvo.

Nesta visualização, as áreas geográficas são diferenciadas de acordo com a proporção de aumento ou redução das chances de ocorrência do alvo. Existem duas classes de cores, em tons avermelhados e em tons azulados, representando os locais de redução ou aumento das chances de ocorrência, respectivamente. Quão maior for o aumento/redução das probabilidades, mais intensa é a tonalidade da cor. Variações pouco significativas são apresentadas na cor branca. Localidades cuja magnitude dos dados seja inferior a um mínimo pré-estabelecido, recebem uma tonalidade acinzentada.

No caso das antenas, são destacadas somente aquelas que representam um aumento das chances do grupo alvo, sendo representadas pela cor verde com o seu tamanho proporcional ao aumento das chances de ocorrência, a escala é linear. Nas localidades são apresentadas as informações relacionadas com o aumento ou a redução das chances de ocorrência do seu grupo alvo. Já os marcadores das antenas apresentam as mesmas estatísticas, acrescentadas das suas coordenadas geográficas.

O exemplo ilustrado na Figura 3 apresenta um mapa retratando as chances de ocorrência de conexões de italianos no período de 05/2020 a 10/2021. A localidade destacada em azul, representa a região de Madureira. Já a área de tom de vermelho forte representa as regiões da Tijuca e de Santa Teresa. Na primeira, os italianos representam mais de 50% das conexões dos estrangeiros, quando o esperado seria menos que 9%, um pólo de atração deste grupo. Já nas outras áreas em destaque, as conexões dos italianos não chegam a 2% do total, sendo claramente evitadas por este grupo.



**Figura 3. Visualização de Regras de Associação**

### **3. Estudo de Caso: Turismo Inteligente**

Segundo [Bismart 2019], o turismo é benéfico para as cidades pois contribui para a economia, a geração de empregos e para a criação de oportunidades de negócios, especialmente nos setores de serviço, varejo e restaurantes. Um conceito que vem surgindo nos últimos anos é o de “turismo inteligente”, como um componente de uma *Smart City* [Tripathy et al. 2018], contemplado com a aplicação de ferramentas e abordagens inovadoras de tecnologias de informação para sua evolução contínua. Para tal, conforme [Andrade, Gama and Cancela 2019], existe uma riqueza de infraestruturas tecnológicas fornecendo grandes quantidades de dados espaço-temporais alimentando continuamente, em tempo real, enormes repositórios de dados de mobilidade humana que estão associados a muitos aspectos da vida contemporânea. Ainda, segundo [Xu et al. 2021], dados de mobilidade turística podem ser vinculados a abordagens de ciência de dados para melhor compreender os destinos turísticos e suas interações.

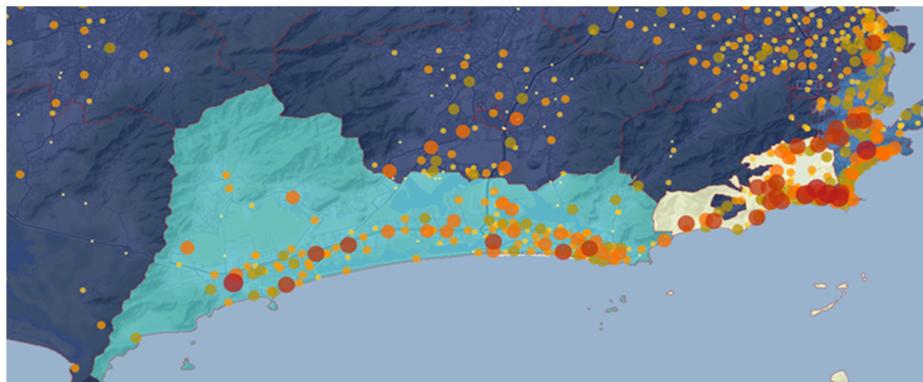
Diante do exposto e para demonstrar a utilidade do nosso trabalho, foi desenvolvido um estudo de caso direcionado ao turismo inteligente envolvendo cada estratégia analítica proposta. O objetivo é verificar se a ferramenta possibilita inferir rapidamente certos padrões de comportamento, localização e mobilidade indicando possíveis tendências. Mesmo sendo necessárias investigações mais aprofundadas, caso confirmadas, esse conhecimento poderá ser utilizado na construção de estratégias governamentais de turismo inteligente e de promoção de destinos, permitindo um melhor aproveitamento desta atividade econômica.

#### **3.1. Estudo da Magnitude das Conexões de Grupos Distintos**

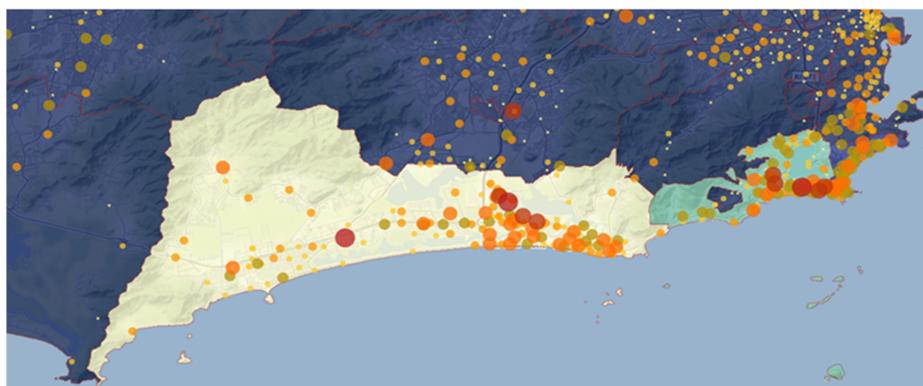
Foi feito um breve estudo analítico envolvendo a ocorrência de conexões de franceses e italianos no período de maio de 2020 a outubro de 2021. Foram escolhidas estas nacionalidades por estarem entre as que tiveram a maior representatividade estatística no período estudado. A análise da magnitude de conexões tende a ser mais completa do que a de dispositivos, pois está relacionada não só com o volume de usuários, mas também com o seu tempo de permanência. Quanto maior a estadia mais conexões tendem a ser realizadas.

Os mapas das Figuras 4 e 5 ilustram as análises realizadas com dados de conexões de franceses e italianos, respectivamente. As regiões com cores mais claras

indicam as maiores ocorrências desses estrangeiros e, dentro das mesmas, os maiores marcadores de antenas indicam pontos específicos de concentração.



**Figura 4. Ocorrências de conexões de franceses de 05/2020 a 10/2021**



**Figura 5. Ocorrências de conexões de italianos de 05/2020 a 10/2021**

Observa-se nos dois mapas que, embora ambos os grupos sigam um padrão geral de concentração ao longo da orla do Rio de Janeiro, houve uma concentração maior principalmente nas regiões da Barra da Tijuca e da Lagoa. Porém, existem perfis específicos de ocorrência quando os mapas são analisados mais detalhadamente. No caso dos franceses (Fig.4), fica evidenciada a preferência pela região da Lagoa e ao analisar o padrão de marcadores das antenas, verifica-se uma concentração maior de conexões na orla de Ipanema. Já no caso dos italianos (Fig.5), que preferem a Barra da Tijuca, embora o padrão dos marcadores das antenas indique uma distribuição um pouco mais equilibrada, existe uma tendência de concentração na região mais comercial da Barra (próximo aos grandes shoppings), além de um ponto próximo ao cruzamento de duas grandes avenidas na região (Av. das Américas com a Av. Salvador Allende).

Tendo em vista os cenários apresentados, é possível levantar algumas questões relacionadas às preferências de cada grupo, as quais poderiam ser usadas como ponto de partida para solução de problemas envolvendo turismo orientado à origem. Isso porque, o padrão de ocorrências de um tipo de turista em uma região pode estar relacionado com variáveis específicas – infraestrutura, recursos humanos, recursos naturais, monumentos, museus etc. – que influenciam o comportamento desses grupos. Dessa forma, um estudo mais aprofundado poderia estabelecer uma relação de causa e efeito cujo conhecimento poderia ser utilizado para se investir em melhorias visando aumentar o potencial turístico desses grupos nas regiões onde se deseja atingir tal resultado.

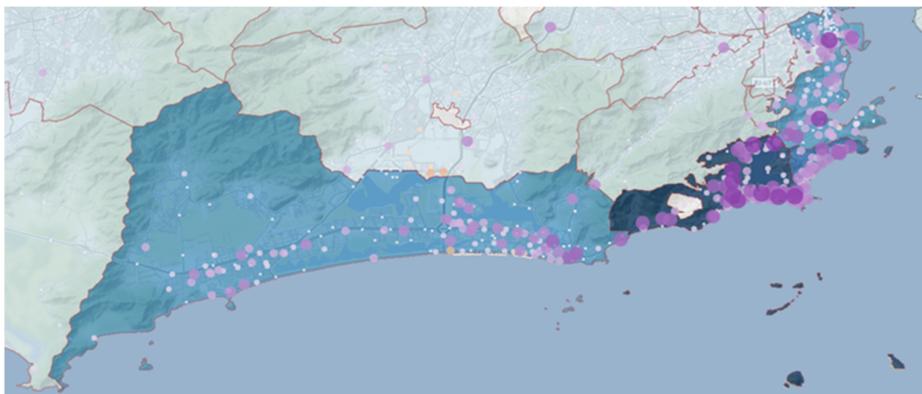
### 3.2. Estudo Comparativo de Dois Conjuntos de Dados

Um breve estudo analítico, envolvendo a ocorrência de dispositivos de norte-americanos e de franceses em dois períodos distintos, foi realizado. Foram comparadas as ocorrências de maio a outubro de 2020 (auge das restrições sanitárias) contra o mesmo período em 2021 (quando a vacinação no Brasil começou a evoluir). Foram escolhidos estes dois grupos por serem os mais representativos nos períodos estudados.

Os mapas das Figuras 6 e 7 ilustram as análises realizadas com dados de norte-americanos e franceses, respectivamente. As regiões com tons intensos de vermelho registram grandes diminuições no número de dispositivos entre os dois períodos, ao contrário daquelas com tons intensos de azul que registram fortes aumentos. No caso dos norte-americanos (Fig.6), o mapa indica uma migração de boa parte das ocorrências saindo da Região da Barra da Tijuca para a Região de Copacabana, quase na mesma proporção. Esta mudança poderia sugerir uma possível troca do turismo de negócios pelo turismo de lazer, sugerindo uma investigação mais aprofundada. Já no caso dos franceses (Fig.7), o mapa indica um forte aumento da ocorrência de dispositivos deste grupo ao longo de toda a orla, com destaque para a região da Lagoa. Esta análise poderia corroborar a tendência já observada no estudo anterior da preferência dos franceses pelo turismo de lazer, desta vez, impulsionado pelo avanço da vacinação da população contra a Covid-19.



**Figura 6. Ocorrência de dispositivos de norte-americanos em 2020 vs. 2021**



**Figura 7. Ocorrência de dispositivos de franceses em 2020 vs. 2021**

Estes casos de uso expõem a possibilidade de aplicação desse tipo de mapa para detectar aumentos ou diminuições de ocorrências de turistas de determinados perfis

entre dois períodos de tempo e assim, posteriormente, buscar entender as causas – políticas, sanitárias, segurança, estratégias de marketing, grandes eventos, entre outros – que possam ter contribuído para influenciar a reputação da área, dentro dos períodos de tempo estudados.

### 3.3. Estudo da Visualização das Regras de Associação

Foi realizado ainda um breve estudo analítico envolvendo as chances de ocorrência de dispositivos de norte-americanos e franceses no período de maio de 2020 a outubro de 2021, utilizando a técnica de regras de associação. Nos mapas apresentados, as localidades com tonalidades avermelhadas apontam uma participação do grupo-alvo inferior ao que seria esperado considerando a sua média geral, indicando que o mesmo tende a evitar a região. Em contrapartida, tons azulados indicam ocorrências acima do esperado, apontando a localidade como um ponto de atração do grupo-alvo. Esta análise pode ser utilizada em conjunto com a magnitude de ocorrências para a obtenção de um diagnóstico mais preciso dos dados quando procuramos identificar padrões específicos de correlação ou de comportamento das variáveis quando consideradas em conjunto.

Os mapas das Figuras 8 e 9 demonstram padrões bem distintos de concentração de norte-americanos e franceses. Os primeiros (Fig.8), que têm um padrão relativamente homogêneo de distribuição na orla do Rio de Janeiro, se destacaram nas regiões da Barra da Tijuca e da Lagoa (azul intenso), com um percentual de aproximadamente 28,5% na ocorrência de dispositivos nestas regiões, quando a probabilidade esperada seria um pouco menor do que 22%, um aumento de cerca de 30% nas suas chances de ocorrência. Já no caso dos franceses (Fig.9), embora estes também sejam atraídos para a zona sul, a região que chamou mais atenção no mapa foi a de Santa Teresa (azul intenso) a qual, embora tenha uma baixa participação no total das ocorrências (aprox. 0,5%), tem uma concentração de franceses muito acima do que seria esperado (23% contra 9%), aumentando suas chances de conexões em 160% nesta região.

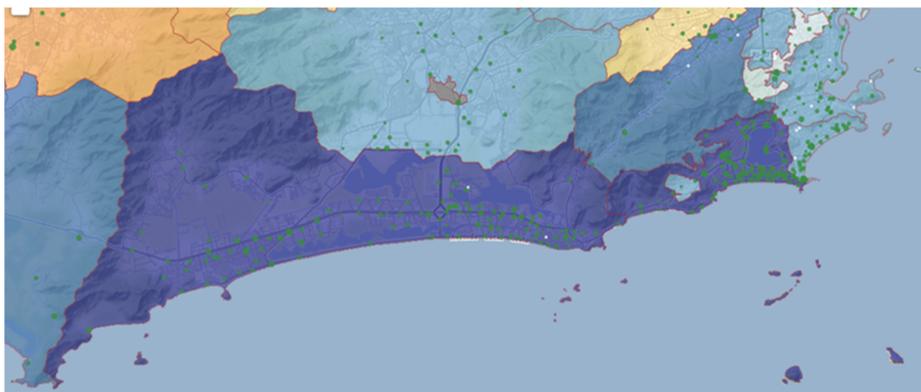
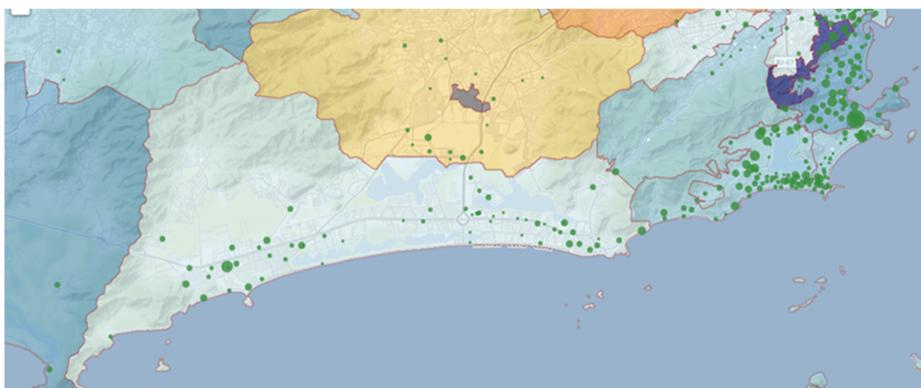


Figura 8. Regras de Associação - Concentração de dispositivos de norte-americanos



**Figura 9. Regras de Associação - Concentração de dispositivos de franceses**

Em suma, nas Regiões/Municípios destacados nos mapas houve maior ocorrência de usuários do que o esperado quando são consideradas as duas variáveis (Região/Nacionalidade) como independentes, indicando que é uma região de forte atratividade deste grupo específico, quando comparado com os outros grupos de estrangeiros. A ocorrência desses padrões, mesmo em pequena escala em alguns casos, poderia indicar a existência de um ou mais atrativos de interesse turístico e cultural na localidade, especialmente atraentes para o grupo estudado. Sob a ótica do turismo inteligente, uma investigação mais apurada poderia indicar as razões deste sucesso, com possibilidade de ampliação ou replicação em outras regiões da cidade.

#### **4. Conclusão e trabalhos futuros**

Neste trabalho, foi proposta uma ferramenta de visualização de dados georreferenciados provenientes de conexões de celulares, que procura extrair padrões visuais de ocorrência de conexões ou de dispositivos. Além disso, foi demonstrada a sua aplicabilidade por meio de um estudo de caso relacionado ao turismo inteligente, baseado nas estratégias analíticas propostas. Foi possível verificar que a utilização de dados georreferenciados de mobilidade associados à ferramenta proposta possibilitou a interpretação e apreensão visual de determinados padrões. Estes últimos sugerem certas tendências que, se confirmadas por estudos mais aprofundados, poderiam auxiliar gestores públicos e privados no planejamento de ações relacionadas não só ao turismo inteligente, mas em um contexto muito mais amplo.

Assim, conclui-se que a ferramenta proposta pode auxiliar a atores do turismo diante dos grandes desafios atualmente enfrentados no contexto das cidades inteligentes. Além disso, embora existam iniciativas semelhantes na literatura, a aplicação de técnicas de mineração de dados, mais especificamente de regras de associação, em soluções de visualização de dados georreferenciados ainda é um assunto muito pouco explorado. Este fato corrobora o potencial de contribuição deste trabalho.

Como trabalhos futuros, propomos a ampliação do escopo da ferramenta com a inclusão de novas propostas de visualização, e de novas fontes de dados georreferenciados. Além disso, sugerimos a elaboração de novos estudos de caso, no contexto de cidades inteligentes, tais como: o controle epidemiológico, o planejamento urbano, o controle de tráfego, entre outros.

## **Agradecimentos**

Trabalho parcialmente apoiado pelo CNPq (processo 315750/2021-9) e FAPESP/MCTIC/CGI.br (processo 2020/09835-1) no escopo do CPA-IA IARA - Inteligência Artificial Recriando Ambientes. Agradecemos à operadora TIM pela cessão dos dados utilizados no estudo de caso.

## **Referências**

- Cortel, F.L. (2019) Smart Cities and Mobility: What Are The Main Challenges? Disponível em: <https://www.plugandplaytechcenter.com/resources/smart-cities-and-mobility>. Acesso em: 14 de Maio de 2022
- Lim, C., Kim, K.J. and Maglio, P.P. (2018) Smart cities with big data: Reference models, challenges, and considerations. *Cities*, v. 82, p. 86-99.
- Han, J., Pei, J. and Kamber, M. (2012) *Data mining: concepts and techniques*. 3rd. ed. Elsevier.
- Madala, S.R., Rajavarman, V.N. and Venkata Satya Vivek T. (2018) Analysis of Different Pattern Evaluation Procedures for Big Data Visualization in Data Analysis. In: *Data Engineering and Intelligent Computing*. Springer, Singapore, p. 453-461.
- Elmqvist N. and Yi J.S. (2015) Patterns for visualization evaluation. *Information Visualization*, v. 14, n. 3, p. 250-269.
- Kang, C. et al. (2010) Analyzing and geo-visualizing individual human mobility patterns using mobile call records. In: *IEEE 18th International Conference on Geoinformatics*. p. 1-7.
- Watson, J.R. et al. (2020) Identifying multiscale spatio-temporal patterns in human mobility using manifold learning. *Peer J Computer Science*, v. 6, p. e276.
- Gavric, K.D. et al. (2011) Detecting attractive locations and tourists' dynamics using geo-referenced images. In: *IEEE 10th International Conference on Telecommunication in Modern Satellite Cable and Broadcasting Services (TELSIKS)*., p. 208-211.
- Bismart (2019) Smart City Tourism: The Future of Tourism. Disponível em: <https://blog.bismart.com/en/smart-city-tourism-future-tourism>. Acesso em: 14 de Maio de 2022.
- Tripathy, A.K. et al. (2018) iTour: The future of smart tourism: An IoT framework for the independent mobility of tourists in smart cities. *IEEE consumer electronics magazine*, v. 7, n. 3, p. 32-37.
- Andrade, T., Gama, J. and Cancela, B. (2019) Mining human mobility data to discover locations and habits. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, Cham. p.390-401.
- Xu, Y. et al. (2021) Characterizing destination networks through mobility traces of international tourists—A case study using a nationwide mobile positioning dataset. *Tourism Management*, v. 82, p. 104195.