DADOS COM PYTHON E ESTATÍSTICA COM R: ESTUDO DA DESIGUALDADE ESPACIAL DE EQUIPAMENTOS CULTURAIS DO ESTADO DO RIO DE JANEIRO

1

2

3

4

Resumo

Este trabalho propõe a análise da desigualdade de equipamentos culturais no Rio de Janeiro. A partir das informações disponíveis na secretaria de Estado de Cultura e Economia Criativa do Rio de Janeiro, realizou-se a raspagem de dados com a linguagem Python e gerou-se visualizações de dados com o R. A raspagem via *web scraper* automatiza o processo complicado de coleta de dados de muitas páginas da web. Utilizou-se de técnicas que extraem o conteúdo da página da Secretaria para gerar um banco de dados. A abordagem desenvolvida na linguagem R é uma representação gráfica interativa, servindo como uma ferramenta de visualização de dados espaciais, onde os valores individuais são representados por uma mancha de calor. Ela compara as distâncias entre os pontos simultaneamente agrupando-os pelo grau de proximidade. Esta análise sugere que há uma grande disparidade entre os municípios do Rio de Janeiro, todavia, essa concentração dos equipamentos culturais não acontece apenas entres os municípios. Também há evidências de desigualdades espaciais internamente dentro de cada um dos municípios. Além disso, nossos resultados também sugerem que há uma associação entre a renda e a quantidade de equipamentos culturais dos bairros na cidade do Rio de Janeiro. Conclui-se o artigo com exemplos de práticas e políticas públicas que perpetuam a desigualdade do acesso da população aos bens culturais e recomenda-se o uso de equipamentos não faraônicos como o famoso Museu do Amanhã.
Palavras-chave: visualização em mapa, raspagem de dados, equipamentos culturais, desigualdade.

Abstract

This work proposes the analysis of the inequality of cultural equipment in Rio de Janeiro. Based on the information available at the Secretary of State for Culture and Creative Economy of Rio de Janeiro, the data was scraped with the Python language and data visualizations were generated with the R. Scraping via web scraper automates the complicated process of collecting data from many web pages. Techniques were used to extract the contents of the Secretariat page to generate a database. The approach developed in the R language is an interactive graphical representation, serving as a spatial data visualization tool, where individual values are represented by a heat stain. It compares

---

1

2

3

4

distances between points simultaneously grouping them by the degree of proximity. This analysis suggests that there is a great disparity between the municipalities of Rio de Janeiro, however, this concentration of cultural equipment does not happen only among the municipalities. There is also evidence of spatial inequalities internally within each municipality. In addition, our results also suggest that there is an association between the income and the amount of cultural equipment of the neighborhoods in the city of Rio de Janeiro. The article concludes with examples of practices and public policies that perpetuate the inequality of the population's access to cultural goods and it is recommended to use non-pharaonic equipment such as the famous Museum of Tomorrow.
Keywords: data visualization, data scraping, cultural equipment, inequality.

Introdução

É muito comum ouvir que os equipamentos culturais (centros culturais, museus, teatros, bibliotecas, etc.) estão concentrados em determinadas regiões e territórios, em detrimento da ausência desses espaços em grande parte das cidades brasileiras. Este trabalho propõe analisar essas disparidades de equipamentos culturais que acontecem no estado do Rio de Janeiro a partir de dados oficiais. A coleta desses dados permite visualizar no mapa a distribuição espacial desses equipamentos.

Apesar das grandes diversidades artísticas e culturais existentes nos municípios, os equipamentos culturais podem estar concentrados em apenas uma parte do território. Mesmo hoje, ainda é difícil verificar essa afirmação. O Brasil não tem um quadro de informações sobre a cultura disponível para o acesso de qualquer cidadão. O que se tem são tentativas de atender a demanda a partir de informações pontuais. Dentre as pesquisas sobre cultura, cumpre registrar o papel do IBGE com diversas pesquisas como o *Sistema* de Informações e *Indicadores Culturais* - SIIC (2003, 2005) e o perfil de informações básicas municipais – Cultura (Munic Cultura, 2006). Até hoje, essas pesquisas, que não são contínuas, são praticamente as únicas fontes de informações seguras sobre os equipamentos culturais.

Nesse sentido, pretende-se com esse trabalho explorar as informações oficiais sobre os equipamentos culturais. Busca-se criar um mecanismo contínuo de informações sobre os equipamentos culturais do estado do Rio de Janeiro: o Mapa da Cultura - MC. O MC poderá inclusive complementar as pesquisas do IBGE como fonte de pesquisa sobre os equipamentos culturais do Rio de Janeiro. A construção do MC será util para responder alguns questionamentos:

1. Existe uma assimetria na distribuição de equipamentos culturais no Estado do Rio de Janeiro?

2. No caso de existir, essa assimetria aponta uma desigualdade na distribuição dos equipamentos culturais do Estado?

3. Como deixar disponíveis as informações geolocalizadas contidas na secretaria do Estado do Rio de Janeiro?

A partir das informações disponíveis na secretaria de Estado de Cultura e Economia Criativa do Rio de Janeiro, realizou-se a raspagem de dados com a linguagem Python e gerou-se visualizações de dados com o R para a análise dos questionamentos apresentados acima.

Este trabalho está dividido em 04 etapas. Para isso, apresenta-se uma análise da gestão dos equipamentos culturais. Em seguida, descreve-se de forma detalhada as etapas para a raspagem de dados utilizando o Python. Após essa etapa, mostra-se o procedimento para a geração do mapa da cultura utilizando a linguagem R. Com esta exposição parte-se para uma análise acerca da concentração e da desigualdade dos equipamentos culturais. Finalmente, aponta-se alguns indicadores de desigualdade espacial no que tange a distribuição dos equipamentos no Estado do Rio de Janeiro.

Ausência de Gestão Cultural

De acordo com Barros & Ziviani (2009), a área da Cultura sofre com uma baixa institucionalização e despreparo dos órgãos gestores dos municípios quanto à importância das informações das políticas voltadas para o setor. Sem dados torna-se impossível fazer uma alocação de equipamentos culturais de forma eficiente e justa. Dados sobre a Cultura são fundamentais para uma boa política cultural. Todavia, no Brasil, existem muito poucas informações sobre a gestão cultural. Isso se deve, em grande parte, ao jeito informal que o setor da cultura representa no Estado Brasileiro. Assim, “em sua maioria, a gestão da cultura encontra-se acoplada a outro setor, como o turismo, a educação e até mesmo a saúde, como já ocorreu em tempos passados na instância federal” (Barros & Ziviani, 2009).

Nesse sentido, na maioria das vezes, o conceito está associado a noção de direitos culturais como parte dos direitos humanos sendo a dimensão cultural indispensável e, acima de tudo, estratégica para qualquer projeto de desenvolvimento. Por conseguinte, o reconhecimento e a valorização da diversidade cultural estão ligados à democracia cultural,

democratização da gestão cultural, participação da sociedade civil nas decisões políticas e no processo de Gestão Pública, descentralização da produção, entre outros aspectos (Barros & Ziviani, 2009).

Objetivo

O objetivo geral deste trabalho é analisar a visualização em mapa dos instrumentos culturais dos municípios do Estado do Rio de Janeiro.

Os objetivos específicos são:

- Construir a visualização do mapa da cultura com os dados dos endereços georeferenciados dos equipamentos culturais;
- Avaliar a distribuição espacial dos equipamentos culturais;
- Refletir sobre a desigualdade da distribuição espacial dos equipamentos culturais no Estado do Rio de Janeiro.

Material e Método

Sobre a raspagem de dados

Michell (2015) afirma que a internet está repleta de dados não estruturados e semi-estruturados que nunca foram disponibilizados como um banco de dados formal: muitas páginas da *web* trazem conteúdo de texto que é legível por humanos, mas não é facilmente legível por máquina. A raspagem de dados (*webscraping*) preenche essa lacuna e abre um novo mundo de dados para os pesquisadores, extraindo automaticamente conjuntos de dados estruturados de conteúdo legível por humanos.

Já Vargiu & Urru (2012) definem a raspagem de dados como o conjunto de técnicas usadas para obter automaticamente informações de um site, em vez de copiá-las manualmente. De acordo com os autores, o objetivo de um raspador de dados *web* é procurar determinados tipos de informações, extrair e agregar em um banco de dados. Em particular, os raspadores estão focados em transformar dados não estruturados e salvá-los em bancos de dados estruturados.

Desse modo, um raspador de dados acessa páginas da *web*, localiza elementos de dados especificados na página, os extrai e finalmente salva esses dados como um banco de dados estruturados. Esse processo basicamente imita o formato de operação de um navegador da *web*, acessando páginas da *web* e salvando-as em um computador.

Assim, um *webscraper* automatiza o processo complicado de coleta de dados de muitas páginas da *web*. Em resumo, a raspagem de dados da *web* é o processo de extrair e criar um banco de dados estruturado de um site. Este artigo concentrou-se nas técnicas que extraem o conteúdo da página da Secretaria de Estado da Cultura e Economia Criativa do Rio de Janeiro na *web*.

Para realizar essa pesquisa, foi necessário raspar informações de espaços culturais presentes no Estado do Rio de Janeiro através do site mapadecultura.rj.gov.br. Na figura 1 observa-se a disposição destas informações no referido site. Foi utilizado a linguagem de programação Python com as bibliotecas *scrapy* e *rexgen*, para o processo de raspagem das informações desse site, e o R com as bibliotecas *tidyverse* (WICKHAM, 2017), *leaflet* (CHENG et. al. 2017), *ggmap* (KAHLE & WICKHAM, 2013), *readxl* (WICKHAM & BRYAN, 2017), *leaflet.extras* (KARAMBELKAR, 2017) e *rgdal* (BIVAND et. al., 2017), para a elaboração da visualização dos dados em mapas.

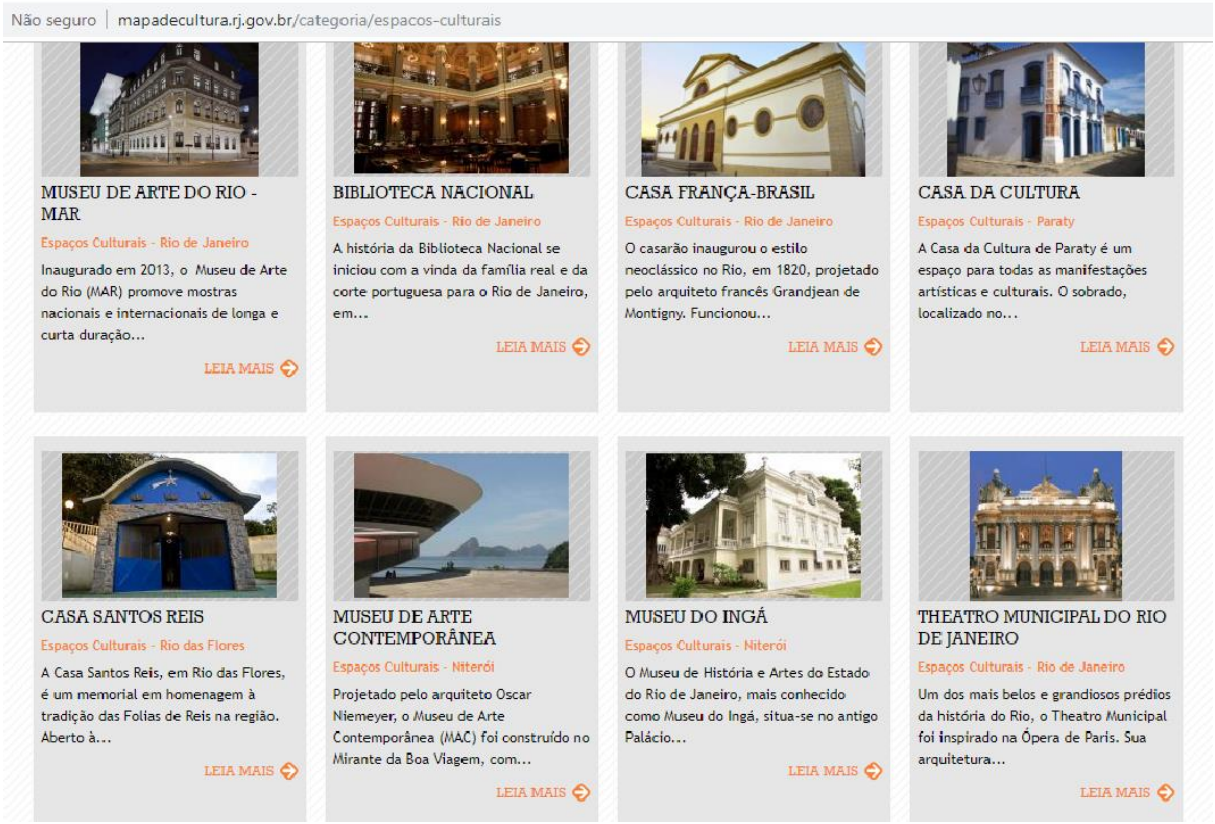

Figura 1 – Dados dos equipamentos culturais no site do mapa de cultura do Estado do Rio de Janeiro

Fonte: http://mapadecultura.rj.gov.br/, 2019

Sobre a visualização de dados

Para apresentar os resultados de bancos de dados são utilizados métodos de comunicação, como por exemplo, tabelas, gráficos ou mapas. Algumas ferramentas gerais de visualização têm sido propostas, como o Box-plot (TUKEY 1977, p. 41). O Mapa de árvores (JOHNSON, 1992), barômetro da sustentabilidade (KRONEMBERGER et al., 2008), entre muitos outros. Entre as novas metodologias de visualização de dados espaciais, destaca-se o mapa anamórfico. O mapeamento anamórfico visa adaptar a forma do mapa não à realidade física, mas sim à realidade percebida. Essa distorção do espaço pode ser construída por um modelo matemático a partir de dados quantitativos reais. O mapa não é mais considerado como um modelo de realidade geográfica, mas principalmente como um documento de comunicação (DENAIN & LANGLOIS, 1998). Cada uma dessas abordagens tem metodologia totalmente distinta e é utilizada para apresentar diferentes resultados dependendo do contexto de comunicação onde a visualização está inserida.

A abordagem aqui utilizada é o mapa de calor, uma representação gráfica interativa, servindo como uma ferramenta de visualização de dados espaciais, onde os valores individuais são representados por um ponto e uma mancha de calor. Ele compara as distâncias entre os pontos simultaneamente agrupando-os pelo grau de proximidade dos locais. O termo mapa de calor foi originalmente cunhado e registrado pelo designer de software Cormac Kinney em 1991 (GUIMARÃES et al., 2014).

Mapas interativos com Leaflet

Mapas estáticos são úteis para criar figuras para relatórios e apresentações. Todavia, às vezes, deseja-se interagir com seus dados. Para isso, pode-se utilizar o pacote chamado *leaflet* (CHENG et. al. 2017) que permite sobrepor os dados espaciais sobre mapas interativos. A ideia dessa abordagem é utilizar o *google maps* com dados espaciais sobrepostos. O *leaflet* (CHENG et. al. 2017) é uma biblioteca *JavaScript* de código aberto que pode ser usada para criar mapas interativos compatíveis com dispositivos móveis.

Procedimento com o Python

Foi criado uma "*spider*" utlizando o pacote *scrapy* do *Python* que "rasteja" pelo site do mapa de cultura na aba espaços culturais. O processo começa com o link da primeira aba da categoria espaços culturais e com isso o *Python* irá guardar o link de todos os *posts* de espaços culturais presentes na aba. Com os *links* de cada *post*, o código irá para o site do *post* e em seguida irá salvar em uma variável o código *html* de todo o corpo de texto, e utilizando o pacote *re*, irá filtrar o endereço. Em seguida irá extrair o nome do lugar e salvar

em um documento no formato *txt* no formato "título"; "endereço". Com todos os *posts* da primeira aba extraídos, o código irá analisar se existe um *link* para a próxima aba, se existir, o processo irá se repetir para aquela aba.

Com todas as informações necessárias extraídas, gera-se um documento *txt* para prosseguir com a análise na linguagem R.

Procedimento de visualização de dados com o R

Com o arquivo *txt* gerado anteriormente, foi preciso um pequeno tratamento manual pois algumas poucas vezes houve erro de *enconding* por falta de padrão presente no site do mapa cultural que apresentava problemas para o R. Também observou-se que quando há uma rede de lugares, por exemplo cinemas, o site agrupava os endereços em um só *post*. Então, separamos esses lugares de modo que cada um possua seu próprio nome e endereço.

Foi utilizado o pacote *readxl* (WICKHAM & BRYAN, 2017) para que fosse importado o arquivo *txt* para o R em formato de um *tibble* (MÜLLER & WICKHAM, 2018) e em seguida foram criadas colunas chamadas de "latitude" e "longitude" preenchidas todas com 999 para posteriormente serem preenchidas com os dados geográficos de cada endereço do banco de dados.

Por falta de padrão de endereços no site, foi criado um processo utilizando o pacote *stringr* (WICKHAM, 2018) em que cada endereço foi padronizado para incluir, no caso de ausência das palavras "RJ" ou "Rio de Janeiro", a inclusão da palavra "Rio de Janeiro" no final.

Após essa etapa, foi necessário adquirir acesso a uma *Application Programming Interface* - API do *google* que permitisse acessar aos dados do *google map.,* Para isso, realizou-se um cadastro na plataforma *google cloud* e configurou-se a API para que fosse possível o uso da linguagem R para obter dados de latitude e longitude.

Com o pacote *ggmap* (KAHLE & WICKHAM, 2013), cadastrou-se a chave da API a fim de utilizar a função *geocode* para que ela retornasse a latitude e a longitude de cada endereço disponível no banco de dados. Alguns poucos lugares apresentaram falhas na sua geolocalização por causa de possíveis erros de endereço no site ou até mesmo por estarem incompletos. Foi feito uma rápida pesquisa para que esses problemas fossem resolvidos e assim criou-se um processo utilizando a função *geocode* em cada endereço do banco de dados anteriormente criado e salvando sua latitude e longitude nas devidas colunas. Assim obtive-se um *tibble* (MÜLLER & WICKHAM, 2018) de todos os espaços culturais presentes

no site *mapadecultura.rj.gov.br* com as colunas: nome do equipamento, endereço, latitude e longitude

Com o *tibble* (MÜLLER & WICKHAM, 2018) completo de informações, utilizou-se a função *readOGR* do pacote *rgdal* (BIVAND et. al., 2017) para ler o *shapefile* com a malha dos municípios do estado do Rio de Janeiro, para em seguida utilizar o pacote *leaflet (CHANG* et. al. 2017*)*, que gerou um mapa global com o RJ marcado. Configurou-se o Mapa da Cultura - MC para que todos os endereços do banco de dados sejam marcados. Para melhor visualização, criou-se o mapa de marcadores de círculos e o mapa de calor. Há possibilidade de selecionar dois tipos diferentes de mapas, o *Esri.WorldImagery* ou o *CartoDB*.

Resultados e Discussão

O Mapa Cultural

O mapa está disponível no endereço: https://rpubs.com/NOMEDOAUTOR/Mapa-Cultural. A sua navegação é intuitiva e totalmente interativa, possibilitando que o usuário, com apenas alguns cliques, tenha informação de qual município está visualizando e seus respectivos aparelhos culturais.

No canto superior direito, pode-se trocar o *layout* do mapa e seus marcadores. É possível visualizá-lo com formato "*Esri.WorldImagery*", que apresenta uma visão mais realística, ou com "*CartoDB*" que lembra o formato clássico de mapas. Para a visualização dos marcadores, pode-se escolher entre "Calor" que possibilita a visualização de um mapa de calor, deixando visível a concentração de aparelhos por região, ou "Círculos", que torna possível a visão mais exata da localização de cada equipamento cultural junto com seu nome.

Avaliação da distribuição espacial dos equipamentos culturais

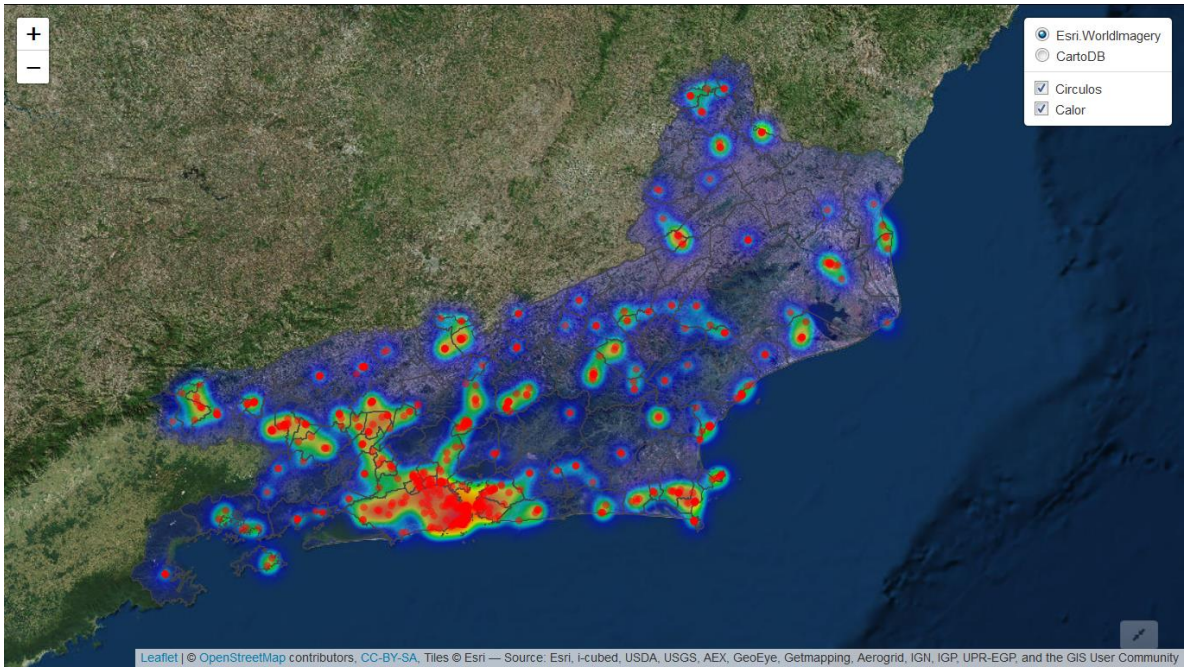

Figura 2 – Visão Geral do mapa de calor do estado do rio de janeiro.

Fonte: Autores, 2019

De acordo com o mapa da figura 2, verifica-se uma grande concentração dos equipamentos culturais na região metropolitana do Rio de Janeiro e na capital. Observam-se áreas com muitos equipamentos culturais e outras áreas sem nenhum. Esta comparação sugere que há uma grande disparidade entre os municípios do Estado do Rio de Janeiro, todavia, essa concentração dos equipamentos culturais não acontece apenas entres os municípios.

Há desigualdades espaciais também internamente dentro de cada um dos municípios. Um exemplo disso pode ser observado no mapa da figura 3, no município do Rio de Janeiro. Os equipamentos culturais estão concentrados na Zona Sul da cidade nos bairros do Centro, Glória, Flamengo, Botafogo, Copacabana, Ipanema, e Leblon. Está é a região mais rica da cidade do Rio de Janeiro e também a região com maior volume de equipamentos culturais oficiais. À exceção da Zona Sul, todo o resto do município do rio de janeiro parece esquecido pelo poder público.

Todavia, para uma comparação justa, deve-se levar em consideração também, o tamanho da população da cidade. O Rio de Janeiro tem cerca de 6 milhões de habitantes enquanto o seu município vizinho Niterói, tem algo em torno de 500 mil. Ao observar municípios com aproximadamente a mesma população como é o caso de Niterói e São João de Meriti constata-se que enquanto Niterói tem mais de 30 equipamentos, São João de Meriti na baixada fluminense tem apenas 07 aparelhos.

A relevância desta pesquisa se deve não somente por observar essas desigualdades, mas também no sentido de servir como ferramenta de gestão. A alocação do próximo equipamento cultural deve obrigatoriamente incorporar as problemáticas geradas pelas desigualdades espaciais apresentadas no trabalho.

Cumpre registrar que essa análise leva em consideração apenas os equipamentos culturais oficiais, isso é, aqueles divulgados pela Secretaria de Cultura e Economia Criativa do Estado do Rio de Janeiro. Destaca-se que no site oficial da secretaria, os municípios de Italva, Guapimirim e Cardoso Moreira não apresentam nenhum aparelho cultural. Para esses municípios, há uma urgência de equipamentos culturais.

Além disso, no mapa da figura 4, destaca-se a diferença entre a capital e o interior do Estado. Enquanto na região metropolitana têm-se muitos equipamentos culturais, o interior do Estado está em local com a inexistência de equipamentos, isto é, um deserto de equipamentos culturais.

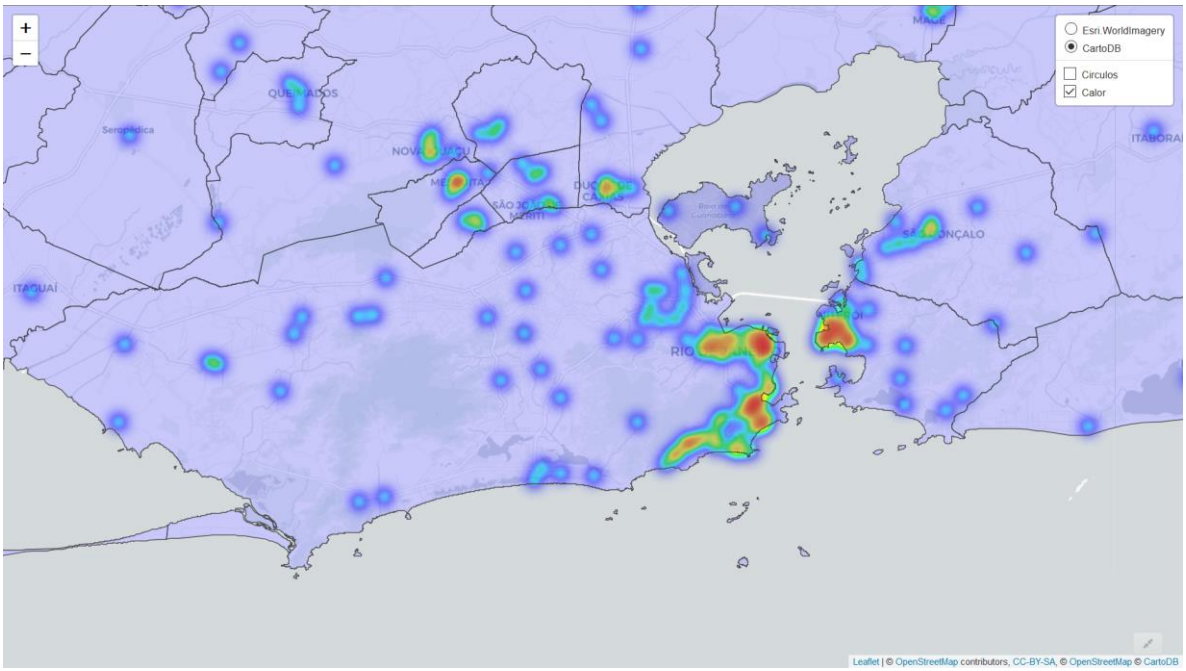

Figura 3 – Mapa de calor dos equipamentos da cidade do Rio de Janeiro

Fonte: Autores, 2019

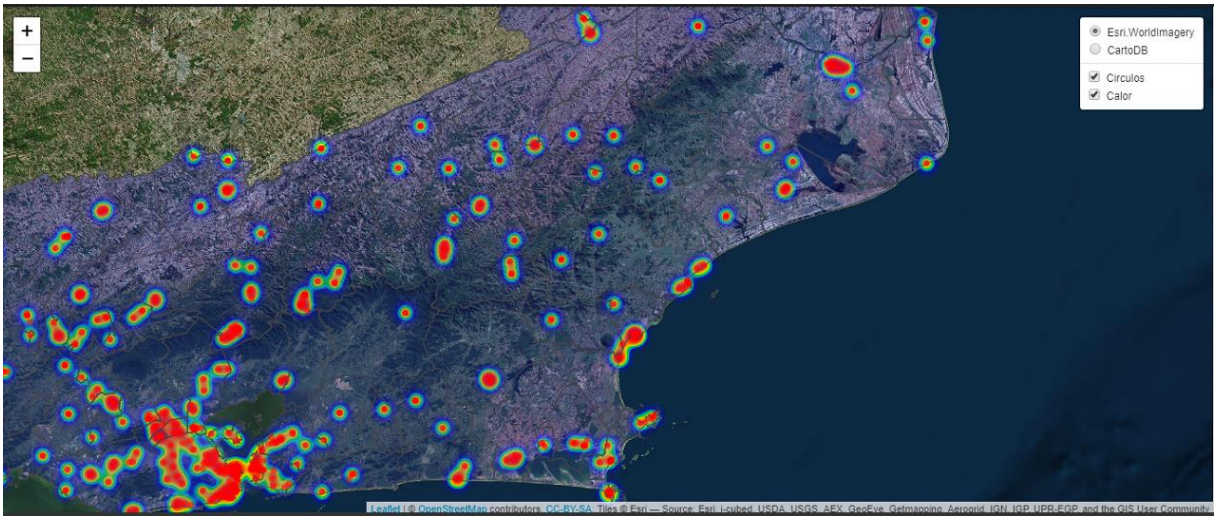

Figura 4 – Mapeamento dos locais com inexistência de equipamentos culturais

Fonte: Autores, 2019

Conclusão

Conforme observado, a intenção deste trabalho foi visualizar o conjunto de equipamentos culturais dos municípios do Estado do Rio de Janeiro, listados no site da Secretaria de Cultura e Economia Criativa do Governo do Estado do Rio de Janeiro, no primeiro trimestre de 2019. A visualização do mapa foi construída utilizando-se de forma integrada as linguagens Python e R, que possibilitou de forma eficiente a raspagem dos dados e a construção do mapa. Todo o código para o desenvolvimento desse trabalho e reprodução dos resultados pode ser encontrado neste endereço: https://github.com/NOMEDOAUTOR/Mapa-Cultural

A partir deste Mapa da Cultura, pode-se avaliar como os equipamentos culturais se distribuem pelos municípios, evidenciando-se três regiões do Estado com uma grande área marcada notadamente pela ausência destes equipamentos, áreas estas que foram denominadas de deserto cultural. A primeira área compreende a região de Guapimirim; a segunda área compreende a região de Italva e a terceira área Cardoso Moreira.

Diante deste cenário, fez-se uma reflexão quanto à desigualdade da distribuição espacial destes equipamentos culturais, tomando como base o Museu do Amanhã que foi inaugurado em dezembro de 2015 com um custo de construção de aproximadamente Duzentos e quinze milhões de reais. Supondo que este montante fosse direcionado para a construção de equipamentos culturais de aproximadamente um milhão de reais, seria possível agregar pelo menos mais dois equipamentos por município. Evidencia-se neste exemplo a prática de uma política pública que perpetua a desigualdade do acesso da

população aos bens culturais ao concentrar a aplicação de recursos numa única área que inclusive já possui uma grande quantidade de equipamentos culturais. As lonas culturais ou mesmo equipamentos culturais que possuem formatos não faraônicos, por terem menores custos, podem ser uma alternativa para reduzir essa desigualdade cultural presente no Estado do Rio de Janeiro.

Referências

BARROS, José Márcio & ZIVIANI, Paula. **Equipamentos, meios e atividades culturais nos municípios brasileiros: indicadores de diferenças, desigualdades e diversidade cultural** IN: CALABRE, Lia Políticas culturais : reflexões e ações. São Paulo : Itaú Cultural ; Rio de Janeiro : Fundação Casa de Rui Barbosa, 2009.

BIVAND, R. KEITT, T. e ROWLINGSON, B. **rgdal: Bindings for the 'Geospatial' Data Abstraction Library**. R package version 1.2-16. URL https://CRAN.R-project.org/package=rgdal, 2017.

CHENG, J. KARAMBELKAR, B. E XIE, Y. (2017). **leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library**. R package version 1.1.0.9000. URL http://rstudio.github.io/leaflet/, 2017.

DENAIN, J. C ; LANGLOIS, P. **Cartographie em anamorphose, Mappe Monde**, Avignon/França, v.49, 1998.

IBGE. **Sistema de informações e indicadores culturais 2003-2005**. v. 22. Rio de Janeiro: IBGE, 2007. 129 p.

_____. **Pesquisa dos municípios brasileiros – Cultura 2006**. Rio de Janeiro: IBGE, 2007. 268 p.

_____. **Sistema de informações e indicadores culturais 2003**. V. 18. Rio de Janeiro: IBGE, 2006. 126 p.

GUIMARÃES, J. T. F. et al. **Palynology of the Middle Miocene**—Pliocene Novo Remanso Formation, Central Amazonia, Brazil. Asociación Paleontológica Argentina: BioOne 2014.

JOHNSON, B. **TreeViz: treemap visualization of hierarchically structured information**. 1992, ACM. p.369-370.

KAHLE, D. e WICKHAM, H. ***ggmap: Spatial Visualization with ggplot2.*** The R Journal, 5(1), 144-161. URL http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf, 2013.

KARAMBELKAR, B. ***leaflet.extras: Extra Functionality for 'leaflet' Package***. URL https://github.com/bhaskarvk/leaflet.extras, 2017.

KRONEMBERGER, D. M. P. et al. **Desenvolvimento sustentável no brasil: uma análise a partir da aplicação do barômetro da sustentabilidade**. Sociedade & Natureza. 20: 25-50 p. 2008.

MITCHELL R.. **Web Scraping with Python: Collecting Data from the Modern Web.** Sebastopol, CA: O'Reilly Media, 2015.

MÜLLER, K. e WICKHAM, H. ***tibble: Simple Data Frames.*** *R package version 1.4.2.* URL https://CRAN.R-project.org/package=tibble, 2018.

PYTHON Software Foundation. **Python Language Reference**, version 3.7.1. URL http://www.python.org, 2019.

R Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/, 2018.

SCRAPY developers. **A high-level Web Crawling and Web Scraping framework**. Python library version 1.6.0. URL https://pypi.org/project/Scrapy, 2019.

TUKEY, J. W. Box-and-Whisker Plots. 2C in: **Exploratory Data Analysis**. Reading, MA: Addison-Wesley, pp. 39-43, 1977.

VARGIU, Eloisa; URRU, **Mirko Exploiting web scraping in a collaborative filtering- based approach to web advertising.** Artificial Intelligence Research, 2013, Vol. 2, No. 1 ISSN 1927-6974 E-ISSN 1927-6982 44. Online Published: December 5, 2012 DOI: 10.5430/air.v2n1p44 URL: http://dx.doi.org/10.5430/air.v2n1p44

WICKHAM, H. ***tidyverse: Easily Install and Load the 'Tidyverse'***. R package version 1.2.1. URL https://CRAN.R-project.org/package=tidyverse, 2017.

_____. ***stringr: Simple, Consistent Wrappers for Common String Operations***. *R package version 1.3.0.* URL https://CRAN.R-project.org/package=stringr 2018.

WICKHAM, H e BRYAN, J. ***readxl: Read Excel Files***. *R package version 1.0.0.* URL https://CRAN.R-project.org/package=readxl, 2017.

Anexo 1

Código Python: https://github.com/NOMEDOAUTOR/mapaCultural.py

```
import scrapy
from scrapy.crawler import CrawlerProcess
import re
arquivo = open('Casas.txt','w')
regex = re.compile(r'Endereço: (.*)<br>')
class mapaCultura(scrapy.Spider):
    name = "cultura"
    # Começar
    def start_requests( self ):
        site = 'http://mapadecultura.rj.gov.br/categoria/espacos-
culturais?page=1#ancora'
        yield scrapy.Request(url = site, callback = self.parse)
#Primeira etapa consiste em pegar todos as casas que estão presentes na segunda
caixadDe apresentação do site
    def parse(self, response):
        segundoquadro =
response.xpath('//*[@id="conteudo"]/div[5]/ul[2]//li/h3/a/@href').extract()
        proxima = response.css('.proxima > a::attr(href)').extract_first()
        for link in segundoquadro :
```

```
                yield response.follow(url = link, callback = self.parse2)
            if(proxima):
                yield response.follow(url = proxima, callback = self.parse)
            if(proxima is  None):
                yield response.follow(url =
    'http://mapadecultura.rj.gov.br/categoria/espacos-culturais', callback=
    self.Comeco )
    #Por final, eu volto para primeira página e pego as 12 que sempre se repetem
        def Comeco(self, response) :
            primeiroQuadro =
    response.xpath('//*[@id="conteudo"]/div[5]/ul[1]//li/h3/a/@href').extract()
            for link in primeiroQuadro :
                yield response.follow(url = link, callback = self.parse2)
        def parse2(self, response):
            pagina = response.css('div#container').extract_first().strip()
            endereco = regex.search(pagina).group(1)
            titulo = response.css('h1::text').extract_first().strip()
            arquivo.write(titulo+' ; '+endereco+'\n')
            arquivo.write('\n')
    process = CrawlerProcess()
    process.crawl(mapaCultura)
    process.start()
    arquivo.close()
```

Código R: https://github.com/NOMEDOAUTOR/Scriptfinal.R

```
library(tidyverse); library(leaflet); library(ggmap); library(readxl);
library(leaflet.extras); library(rgdal)
#Registrar a chave da API do google, confidencial.
  register_google(key = "-")
  casas = read_csv2("F://GitHub//Mapa-Cultural//Casas.txt"
                    ,col_names = F,locale = locale(encoding = 'ISO-8859-1'))
  casas = casas %>%
    rename(lugar = X1, endereco = X2)
  casas = casas %>%
    mutate(
      latitude = "999",
      longitude = "999")
  qtd = dim(casas)[1]
  for(i in 1:qtd){
    if(!str_detect(casas$endereco[i],c("Rio de Janeiro")) &
!str_detect(casas$endereco[i],c("RJ"))  ){
      casas$endereco[i] = paste(casas$endereco[i],", Rio de Janeiro")
    }
```

```
  }
#Procedimento que busca latitude e longitude dos endereços presentes no banco de
dados(está retornando 2 endereços com NA
  for ( i in 1:qtd){

    aux = geocode(as.character(casas$endereco[i]))
    longitudeAux = aux$lon
    latitudeAux = aux$lat
    casas$longitude[i] = longitudeAux
    casas$latitude[i] = latitudeAux
  }
Correções
  errados = casas %>% filter(lugar == "Museu de Cera de Petrópolis" |
                             lugar=="Centro Cultural Bernardino Lopes"|
                             lugar=="Museu do Cárcere" |
                             lugar =="Biblioteca Leonor Leite Bastos de Souza" |
                             lugar=="Associação Sociocultural e Ambiental de Triunfo"|
                             lugar == "Biblioteca Córrego do Ouro" |
                             (lugar == "Casa de Cultura" & endereco == "Praça Orlando
de Barros Pimentel, s/nº, Centro. , Rio de Janeiro"))
  errados$endereco = if_else(errados$lugar =="Museu de Cera de Petrópolis",
                      "Rua Barão do Amazonas, 35 - Centro, Petrópolis - RJ, 25685-070",
                      errados$endereco)
  errados$endereco = if_else(errados$lugar == "Associação Sociocultural e Ambiental de
Triunfo",
                     "R Abdo Felix, Sn Santa Maria Madalena - RJ",
                      errados$endereco)
  errados$endereco = if_else(errados$lugar == "Museu do Cárcere",
                      "RR95+35 Dois Rios, Angra dos Reis - RJ",
                       errados$endereco)
  errados$endereco = if_else(errados$lugar == "Biblioteca Leonor Leite Bastos de
Souza",
                       "35JJ+2M Eldorado, Maricá - RJ",
                       errados$endereco)
  errados$endereco = if_else(errados$lugar == "Centro Cultural Bernardino Lopes",
                      "Rua Alexandre Pereira Dos Santos, sn - Boa Esperança, Rio Bonito
- RJ, 28810-000",
                       errados$endereco)
  errados$endereco = if_else(errados$lugar == "Biblioteca Córrego do Ouro",
                      "Rua Prefeito Antônio Curvelo Benjamin, 226-232,Macaé, RJ",
                       errados$endereco)
  errados$endereco = if_else(errados$lugar == "Casa de Cultura",
                      " R. Álvares de Castro, 154 - Centro, Maricá - RJ, 24942-395",
                             errados$endereco)
```

```
#Correção dos pontos que deram NA na longitude e latitude.
  nas = casas %>%
    filter(is.na(latitude))
  nas$endereco = if_else(nas$lugar == "Ateliê Artimpério",
                    "Rua Barão de Vassouras, 19,Casario Shopping , Rio de Janeiro",
                     nas$endereco)
  nas$endereco = if_else(nas$lugar == "Talentos da Roça, Cultura e Cidadania",
                    "Rua Geraldino Silva, Porciúncula - RJ",
                     nas$endereco)
#Vamos juntar a base de dados dos pontos errados.
  BD_corrigido = full_join(nas,errados)

#Vamos gerar o geocode para o BD_corrigido.
  for ( i in 1:dim(BD_corrigido)[1]){
    aux = geocode(as.character(BD_corrigido$endereco[i]))
    longitudeAux = aux$lon
    latitudeAux = aux$lat
    BD_corrigido$longitude[i] = longitudeAux
    BD_corrigido$latitude[i] = latitudeAux
  }
#Terei um tratamento diferenciado para Casa de cultura pois há mais de 1 lugar chamado
assim no Banco de dados
  nomes_corrigir = BD_corrigido %>%
    filter(lugar != "Casa de Cultura") %>%
    select(lugar) %>%
    pull()

  casas = casas[-538,]
  casas = casas %>%
    filter(!(lugar %in% nomes_corrigir))
#Agora iremos juntar todos os lugares em um só banco de dados todo corrigido e salva-
lo, pois é custoso rodar o código porque utilizamos a API do google.
  casas_corrigidas = full_join(casas,BD_corrigido)
  write_rds(casas_corrigidas,"F://GitHub//Mapa-Cultural//casas_corrigidas.rds")
  casas_corrigidas = read_rds("F://GitHub//Mapa-Cultural//casas_corrigidas.rds")
#Ler a malha dos municípios do RJ
  RiodeJaneiro = readOGR(dsn="F://GitHub//Mapa-
Cultural//Malha_shp",layer="33MUE250GC_SIR",
                         use_iconv = TRUE,
                         encoding = "UTF-8")
#Agora iremos plotar o mapa final
  mapaGeral = leaflet(RiodeJaneiro,
                      options = leafletOptions(minZoom = 1)) %>%
    addTiles() %>%
```

```
    setView(lng=-42.5303, lat=-22.1, zoom = 9) %>%
    addPolygons(color = "#444444", weight = 1, smoothFactor = 0.5,
                opacity = 1.0, fillOpacity = 0.2,
                fillColor = 'blue',
                highlightOptions = highlightOptions(color = "white", weight = 2,
                                                    bringToFront = FALSE),
                popup = RiodeJaneiro$NM_MUNICIP) %>%
    addCircles(lng = as.numeric(casas_corrigidas$longitude),
               lat =  as.numeric(casas_corrigidas$latitude),
               popup = casas_corrigidas$lugar,color = "Red",
               group = "Circulos") %>%
    addHeatmap(lng = as.numeric(casas_corrigidas$longitude),
               lat =  as.numeric(casas_corrigidas$latitude)
               ,radius = 10,
               group = "Calor") %>%
    addProviderTiles(providers$Esri.WorldImagery,group = "Esri.WorldImagery") %>%
    addProviderTiles(providers$CartoDB,group = "CartoDB") %>%
    addLayersControl(overlayGroups = c("Circulos","Calor"),
                     baseGroups = c("Esri.WorldImagery","CartoDB"),
                     options = layersControlOptions(collapsed = FALSE)) %>%
    hideGroup("Calor")
library(htmlwidgets)
saveWidget(mapaGeral, file="F://GitHub//Mapa-Cultural//Mapa//index.html")
```