

## CepespData: Procedimentos de Georreferenciamento dos Locais de Votação

### 1. Preparação dos dados: criação de código único para os locais de votação entre 2006 e 2018. (Esta etapa foi necessária pois não havia nas tabelas do TSE um código de local de votação unificado para todas as eleições)

Procedimentos:

- i. Comparação exata entre os nomes dos locais de votação e os endereços nas mesmas seções em diferentes anos;
- ii. Utilização da tabela de remanejamento para análise de mudanças nos locais de votação em cada seção ao longo dos anos;
- iii. Comparação entre nomes dos locais na mesma zona-município em diferentes anos e seções;
- iv. Comparação entre endereços dos locais na mesma zona-município em diferentes anos e seções;
- v. Atribuição de código numérico de 4 dígitos (padrão “5---”) para locais para os quais não foi possível encontrar correspondência;
- vi. Criação de painel de locais de votação e tabela de locais de votação únicos (Mun-Zona-LocVot).

### 2. Georreferenciamento:

- **Estado de São Paulo e município do Rio de Janeiro:**

- a. Juntar a tabela de locais de votação com as relações disponibilizadas pelo CEM, com coordenadas dos locais de votação na região metropolitana de São Paulo entre 1994 e 2008, pelo CEPESP (GeoLV), com coordenadas dos locais de votação em São Paulo para as eleições de 2014, pelo TRE-SP e pelo TSE, com coordenadas de alguns locais de votação para as eleições de 2018.
- b. **No município de São Paulo:** pesquisa manual das coordenadas dos locais que sem coordenadas [44] no site do googlemaps (por endereço e/ou nome do local);
- c. **Demais municípios:** Georreferenciamento dos locais sem coordenadas pelo endereço dos locais de votação únicos utilizando a API do googlemaps;
- d. **Consistência 1:** testar se os locais georreferenciados nas etapas anteriores estão dentro dos limites de seus municípios;
- e. Novo georreferenciamento dos locais que não passaram no teste anterior [135] agora pesquisando as coordenadas pela API do googlemaps utilizando o nome do local de votação;
- f. Pesquisar manualmente as coordenadas não alteradas, não encontradas ou que estão em centros de bairro e de municípios [22] no site do googlemaps (por endereço e/ou nome do local);

#### Critérios:

- i. Se encontro correspondência do endereço e/ou do nome do local: atribuo a coordenada do ponto encontrado;
- ii. Se não encontro correspondência nem de endereço, nem de nome do local: atribuo a coordenada do centro do bairro indicado ou, no caso de endereço incompleto (sem número), procuro encontrar o local exato do local e atribuo a coordenada do ponto indicado;

- iii. Se não encontro correspondência nem de endereço, nem de nome do local e nem do bairro, procuro em outras fontes, como em planos diretores dos municípios e secretarias municipais e estaduais, alguma correspondência para aproximação das coordenadas.
- b. **Teste de clusterização** dos locais por zona
  - i. O teste consiste em três critérios. Primeiro, verificamos se a diferença entre a média da distância (euclidiana) de um local em relação aos locais da sua zona e a média da distância entre todos os locais da mesma zona é maior do que 5km. Quando a primeira etapa é verdadeira, verificamos se o local está em um município com mais de uma zona. Se esta segunda etapa também for verdadeira, verificamos se a distância média de um local em relação aos locais da sua zona é maior do que a distância média do local em relação aos locais de pelo menos uma outra zona. Se esta terceira etapa também for verdadeira, avaliamos manualmente a localização;
  - ii. Pesquisar manualmente as coordenadas dos locais cujas coordenadas não passaram no teste [43](por endereço e/ou nome do local);

[Critérios descritos acima]

- **Demais capitais estaduais brasileiras:**

- a. Juntar a tabela de locais de votação com a relação de 2018 disponibilizada pelo TSE com algumas coordenadas;
- b. Georreferenciamento dos locais sem coordenadas pelo endereço dos locais de votação únicos utilizando a API do googlemaps;
- c. **Consistência 1:** testar se os locais georreferenciados na etapa anterior estão dentro dos limites de seus municípios;
- d. Novo georreferenciamento dos locais que não passaram no teste anterior agora pesquisando as coordenadas pela API do googlemaps utilizando o nome do local de votação;
- e. **Consistência 2:** testar se os locais georreferenciados na etapa anterior estão dentro dos limites de seus municípios;
- f. Pesquisar manualmente as coordenadas dos locais que não passaram no teste anterior no site do googlemaps (por endereço e/ou nome do local);

**Critérios:**

1. Se encontro correspondência do endereço e/ou do nome do local: atribuo a coordenada do ponto encontrado;
2. Se não encontro correspondência nem de endereço, nem de nome do local: atribuo a coordenada do centro do bairro indicado ou, no caso de endereço incompleto (sem número), procuro encontrar o local exato do local e atribuo a coordenada do ponto indicado;
3. Se não encontro correspondência nem de endereço, nem de nome do local e nem do bairro, procuro em outras fontes, como em planos diretores dos municípios<sup>1</sup>, alguma correspondência para aproximação das coordenadas.

---

<sup>1</sup><https://sempog.portovelho.ro.gov.br/uploads/arquivos/2017/12/25051/1513184688anuario-2010-2016.pdf> [Exemplo de arquivo utilizado – aqui, para o município de Porto Velho]

- g. **Consistência 3:** selecionar para novo geoprocessamento os locais que, nas etapas anteriores, foram georreferenciados no centro de seus respectivos municípios ou bairros;
  - i. Procurar padrões como “RO, Brazil” e “Stateof”, que indicam endereço incompleto, nas descrições dos endereços retornados pelo googlemaps nas pesquisas anteriores;
- h. Novo georreferenciamento dos locais que não passaram no teste anterior pesquisando as coordenadas pela API do googlemaps utilizando o nome do local de votação;
- i. **Consistência 4:** testar se os locais georreferenciados na etapa anterior estão dentro dos limites de seus municípios;
- j. Pesquisar manualmente as coordenadas dos locais que não passaram no teste anterior no site do googlemaps (por endereço e/ou nome do local);

[Critérios descritos acima]

- k. **Consistência 5:** testar se entre os locais que foram georreferenciados utilizando-se o endereço há locais diferentes (nome e/ou endereço) mas que foram georreferenciados no mesmo ponto;
- l. Novo georreferenciamento dos locais que não passaram no teste anterior pesquisando as coordenadas pela API do googlemaps utilizando o nome do local de votação;
- m. **Consistência 6:** testar se os locais georreferenciados na etapa anterior estão dentro dos limites de seus municípios;
- n. Pesquisar manualmente as coordenadas dos locais que não passaram no teste anterior no site do googlemaps (por endereço e/ou nome do local);

[Critérios descritos acima]

- o. **Consistência 7:** selecionar para novo geoprocessamento os locais que, nas etapas anteriores, foram georreferenciados no centro de seus respectivos municípios ou bairros;

[Procurar padrões descritos acima]

- p. Pesquisar manualmente as coordenadas dos locais que não passaram no teste anterior no site do googlemaps (por endereço e/ou nome do local);

[Critérios descritos acima]

- q. **Consistência 8:** testar se os locais georreferenciados pelo TSE estão dentro dos limites de seus municípios;
- r. **Teste de clusterização** dos locais por zona
- s. O teste consiste em três critérios. Primeiro, verificamos se a diferença entre a média da distância (euclidiana) de um local em relação aos locais da sua zona e a média da distância entre todos os locais da mesma zona é maior do que 5km. Quando a primeira etapa é verdadeira, verificamos se o local está em um município com mais de uma zona. Se esta segunda etapa também for verdadeira, verificamos se a distância média de um local em relação aos locais da sua zona é maior do que a distância média do local em relação aos locais de pelo menos uma outra zona. Se esta terceira etapa também for verdadeira, avaliamos manualmente a localização;

- t. Novo georreferenciamento dos locais que não passaram no teste anterior [excluindo os locais processados pelo TSE] pesquisando as coordenadas pela API do googlemaps utilizando o nome do local de votação;
- u. Pesquisar manualmente as coordenadas dos locais cujas coordenadas não foram atualizadas no processamento anterior (por endereço e/ou nome do local);

**[Critérios descritos acima e incorporando, manualmente, coordenadas disponibilizadas por secretarias municipais e pelo INEP (Censo Escolar)]**

- v. **Novo teste de clusterização** dos locais por zona
- w. O teste consiste em três critérios. Primeiro, verificamos se a diferença entre a média da distância (euclidiana) de um local em relação aos locais da sua zona e a média da distância entre todos os locais da mesma zona é maior do que 5km. Quando a primeira etapa é verdadeira, verificamos se o local está em um município com mais de uma zona. Se esta segunda etapa também for verdadeira, verificamos se a distância média de um local em relação aos locais da sua zona é maior do que a distância média do local em relação aos locais de pelo menos uma outra zona. Se esta terceira etapa também for verdadeira, avaliamos manualmente a localização;
- x. Verificar manualmente (colocando no mapa cada zona com locais problemáticos em cada estado) se há alguma coordenada aparentemente destoante de sua zona e pesquisar no site do googlemaps (por endereço e/ou nome do local) se há, de fato, alguma inconsistência.
  - i. Contatou-se que grande parte dos locais problemáticos, em especial no município de Porto Velho (RO), foram exaustivamente checados e aperfeiçoados em etapas anteriores, não havendo razões para novas modificações – talvez extensão longitudinal atípica da zona pode estar enviesando o teste.
  - ii. Dois locais georreferenciados pelo TSE (em GO e DF) foram corrigidos.

**Tabelas:**

Resumo de fontes (SP e Rio de Janeiro)	
Fonte	Locais Únicos
CEM	174
geolv	996
google_manual	5042
Secretaria_Educ_SP <sup>2</sup>	8
tre_sp	6529
tse	1509

Resumo de fontes (Demais Capitais)	
Fonte	Locais Únicos
censo_escolar <sup>3</sup>	74
google_maps	6368
pref_porto_velho <sup>4</sup>	21
seduc_salvador <sup>5</sup>	15
seudc_manaus <sup>6</sup>	11
TSE_2018	4766

Resumo dos Testes e Correções (Demais Capitais)			
		Correções	
	Locais selecionados para correções	API_GoogleMaps	Pesquisa Manual
Consistência 1	137	137	0
Consistência 2	8	0	8
Consistência 3	928	928	0
Consistência 4	12	0	12
Consistência 5	419	419	0
Consistência 6	3	0	3
Consistência 7	36	0	36
Consistência 8	0	0	0
Teste Clusters 1	351	125	92
Teste Clusters 2	341	0	6

<sup>2</sup><http://catalogo.governoaberto.sp.gov.br/dataset/enderecos-de-escolas>

<sup>3</sup><http://inep.gov.br/web/guest/dados/catalogo-de-escolas>

<sup>4</sup><https://transparencia.portovelho.ro.gov.br/Site/Principal/#>

<sup>5</sup><http://www.transparencia.salvador.ba.gov.br/Modulos/Escolas.aspx>

<sup>6</sup><http://semed.manaus.am.gov.br/escolas-cmeis-e-creches/>

## Considerações Finais:

- Dos quase 13.000 locais cujas coordenadas foram fornecidas pelo TSE e pelo TRE-SP, **menos de 20 foram alteradas por serem consideradas inconsistentes** em nossos testes.
- Alguns passos dos procedimentos descritos anteriormente (especialmente no geoprocessamento das capitais) podem parecer redundantes ou realizados de forma pouco eficiente. Como não havia uma rotina de testes e procedimentos pré-estabelecida, ainda mais diante da indisponibilidade de nossa principal ferramenta de geoprocessamento (o GeoLV), os passos descritos foram adotados em concordância com os problemas que pareciam surgir a cada progressão. **Para as próximas rotinas, a revisão e a consolidação dos procedimentos serão indispensáveis para maximizar a eficiência do processamento e reduzir as intervenções manuais.**
- Os testes de consistência descritos acima não garantem que os locais georreferenciados pela API do googlemaps correspondam ao exato local de seus endereços, nem foi criado parâmetro para indicação de algum grau de confiança sobre essas coordenadas. Ainda assim, testes como os de clusterização da zona e de checagem de coordenadas repetidas dão algum grau de confiança para os pontos encontrados. **Futuros testes podem e devem ser discutidos.**

## **Resumo dos procedimentos de georreferenciamento dos locais de votação (Fluxo de Trabalho):**

### **Etapa 1:** Criação do código único entre eleições

0. Correspondência entre códigos do TSE;
1. Comparação entre nomes dos locais na mesma zona-município em diferentes anos e seções;
2. Comparação entre endereços dos locais na mesma zona-município em diferentes anos e seções;
3. Utilização da tabela de remanejamento para análise de mudanças nos locais de votação em cada seção ao longo dos anos;
4. Atribuição de código numérico de 4 dígitos (padrão “5---”) para locais para os quais não foi possível encontrar correspondência;
5. Criação de painel de locais de votação e tabela de locais de votação únicos (Mun-Zona-LocVot);
6. Testes de Consistência.

### **Etapa 2:** Georreferenciamento

1. Busca de coordenadas em diferentes fontes (TSE, TRE, google, censo escolar, CEM ...);
2. Busca na API do GoogleMaps (ou GeoLV) a partir do endereço do local de votação.

### **Etapa 3:** Teste de consistência

1. Verificar se os locais estão no limite dos municípios;
2. Verificar coordenadas que estão em centros de bairros e de municípios (procurar padrões como “RO, Brazil” e “Stateof”, que indicam endereço incompleto, nas descrições dos endereços retornados pelo googlemaps nas pesquisas anteriores);
3. Teste de clusterização dos locais por zonas. O teste consiste em três critérios. Primeiro, verificamos se a diferença entre a média da distância (euclidiana) de um local em relação aos locais da sua zona e a média da distância entre todos os locais da mesma zona é maior do que 5km. Quando a primeira etapa é verdadeira, verificamos se o local está em um município com mais de uma zona. Se esta segunda etapa também for verdadeira, verificamos se a distância média de um local em relação aos locais da sua zona é maior do que a distância média do local em relação aos locais de pelo menos uma outra zona. Se esta terceira etapa também for verdadeira, temos um problema;
4. Testar se entre os locais que foram georreferenciados utilizando-se o endereço há locais diferentes (nome e/ou endereço) mas que foram georreferenciados no mesmo ponto.

#### **Etapas 4:** Correção dos erros e repetição dos testes de consistência

Para as coordenadas que não passaram pelos testes da etapa anterior, refazemos a geolocalização:

1. Busca na API do GoogleMaps a partir do nome do local de votação (escola) (e não pelo endereço);
2. Busca manual no GoogleMaps
3. Se não encontro correspondência nem de endereço, nem de nome do local: atribuo a coordenada do centro do bairro indicado ou, no caso de endereço incompleto (sem número), procuro encontrar o local exato do local e atribuo a coordenada do ponto indicado;
4. Se não encontro correspondência nem de endereço, nem de nome do local e nem do bairro, procuro em outras fontes, como em planos diretores dos municípios, alguma correspondência para aproximação das coordenadas.
5. Repetição da **Etapas 3** para as novas coordenadas e da **Etapas 4**, caso problemas persistam (algum grau de tolerância a ser definido).

#### **Scripts de procedimentos consolidados a serem criados para cada etapa:**

- 01script\_codigounico.R
- 02script\_georreferenciamentoR
- 03script\_consistencia.R
- 04script\_correcao.R