Estimação dos tempos de viagem distrito a distrito

O produto final do programa pretende ser uma matriz dos distritos de são paulo em relação a eles mesmos com indicação do tempo médio de viagem entre eles e o desvio padrão relacionado. A imagem esquemática abaixo ilustra a estrutura:

|  |  |  |  |
| --- | --- | --- | --- |
| Distritos partida/chegada | 1 | 2 | ... |
| 1 | Tempo | Desv | Tempo | Desv | ... |
| 2 | Tempo | Desv | Tempo | Desv | ... |
| ... | ... | ... |  |

Para montar essa estrutura a idéia é montar três processos: Sorteio das coordenadas nos distritos e conversão das coordenadas; Cálculo das viagens e das médias entre distritos.

Sorteio das coordenadas nos distritos

Para sortear as coordenadas dos endereços é necessário ponderar os endereços por alguma informação espacial que dificulte ou impeça endereços impossíveis para as partidas e chegadas das viagens. Atribuindo a cada setor censitário do censo demográfico um peso relativo a proporção da população do distrito que o habita, é possível sortear os endereços de forma a privilegiar as regiões mais densamente povoadas. A princípio serão sorteados ao menos 30 coordenadas para cada distrito e será produzido um vetor contendo a latitude e a longitude e a indicação do distrito a que pertence a coordenada. No processo de incluir as coordenadas no vetor é preciso também convertê-las de UTM (da projeção do mapa usado para sortear as coordenadas) para Latitude/Longitude, que é o input necessário para a função que calcula os tempos de viagem.

A implementação dessa etapa envolve o uso do shapefile dos setores censitários, disponível no portal Geosampa(). A tabela de atributos do shapefile será operada a partir das seguintes etapas:

1º) Retirar as informações desnecessárias da tabela

2º) Classificar os setores censitários pelos distritos a que pertencem, em uma nova coluna da tabela de atributos

3º) Calcular, em uma nova coluna, o peso da população de cada setor censitário na população do distrito a que pertencem

Com esse novo shapefile já modificado é possível realizar os sorteios dos endereços usando os pesos calculados para cada setor censitário. Para isso será iniciado um vetor (*Endereços*) com as seguintes “variáveis”:

|  |  |  |
| --- | --- | --- |
| Código do distrito | Latitude | Longitude |

O preenchimento do vetor seguirá as seguintes etapas:

1º) Para o distrito número 1, transformar os pesos dos setores censitários em uma “pixel image”(?)

2º) Usar a função rpoint para sortear 30(?) pontos no distrito e armazenar as coordenadas sorteadas em um vetor auxiliar

3º) Converter as coordenadas do vetor auxiliar para LatLong e preencher o vetor final de coordenadas

4º) Repetir o processo para os 95 distritos seguintes

O vetor *Endereços* conterá as coordenadas de Latitude e Longitude dos 2880 endereços sorteados. O próximo passo será calcular as viagens entre todos os endereços. Para isso será usada, como primeira saída, a função google\_distance() do pacote googleway (que usa a API do Google Distance Matrix) para o software R, que é capaz de calcular uma matriz de distâncias e tempos para 10 endereços de saída e 10 endereços de entrada. A partir desta função serão calculados os tempos médios e desvios padrão dos deslocamentos entre distritos. Isso será feito da seguinte forma:

1º) Parear dois distritos em um loop duplo:

for(distrito1 entre 1 e 96):

For (distrito2 entre 1 e 96):

If distrito1 == distrito2:

Return (É uma possibilidade não necessária excluir as viagens interdistritais)

Else:

Rodar o programa

2º) Com os dois distritos pareados, a função googleway será chamada nove vezes entre os três grupos de dez endereços e os tempos de viagens serão armazenados em um vetor auxiliar de tamanho 900. A partir desse vetor se calcula a média do tempo entre os distritos e os desvio padrão e se armazena os dois valores na Matriz final na célula referente a linha do distrito de saída e coluna do distrito de chegada.

3º) Se repete o processo para todos os distritos