

Entry Name: **Edwin Rueda MC1**
VAST Challenge 2019
Mini-Challenge 1

Team Members:

Edwin Rueda, Universidade Federal do Pará
Belém-Pará, ejrueda95g@gmail.com

Student Team: YES

Tools Used:

- *Python 3.6.5*
- *BokehJS 1.1.0*
- *Pandas 0.23.0*
- *Numpy 1.14.3*

Approximately how many hours were spent working on this submission in total?

Provide an estimate of the total number of hours worked on this submission by your entire team. 2H, 2h, 3h, 4h, 3h, 4h, 4h, 1h

Questions

1- Emergency responders will base their initial response on the earthquake shake map. Use visual analytics to determine how their response should change based on damage reports from citizens on the ground. How would you prioritize neighborhoods for response? Which parts of the city are hardest hit? Limit your response to 1000 words and 10 images.

As ilustrações a seguir são baseadas na interação com o grafico 1, esta apresenta em cada bairro um diagrama de pizza que indica a magnitude do dano em cada tipo relatado, podendo iterar para cada dia:

registro de acordo com o dia: 6

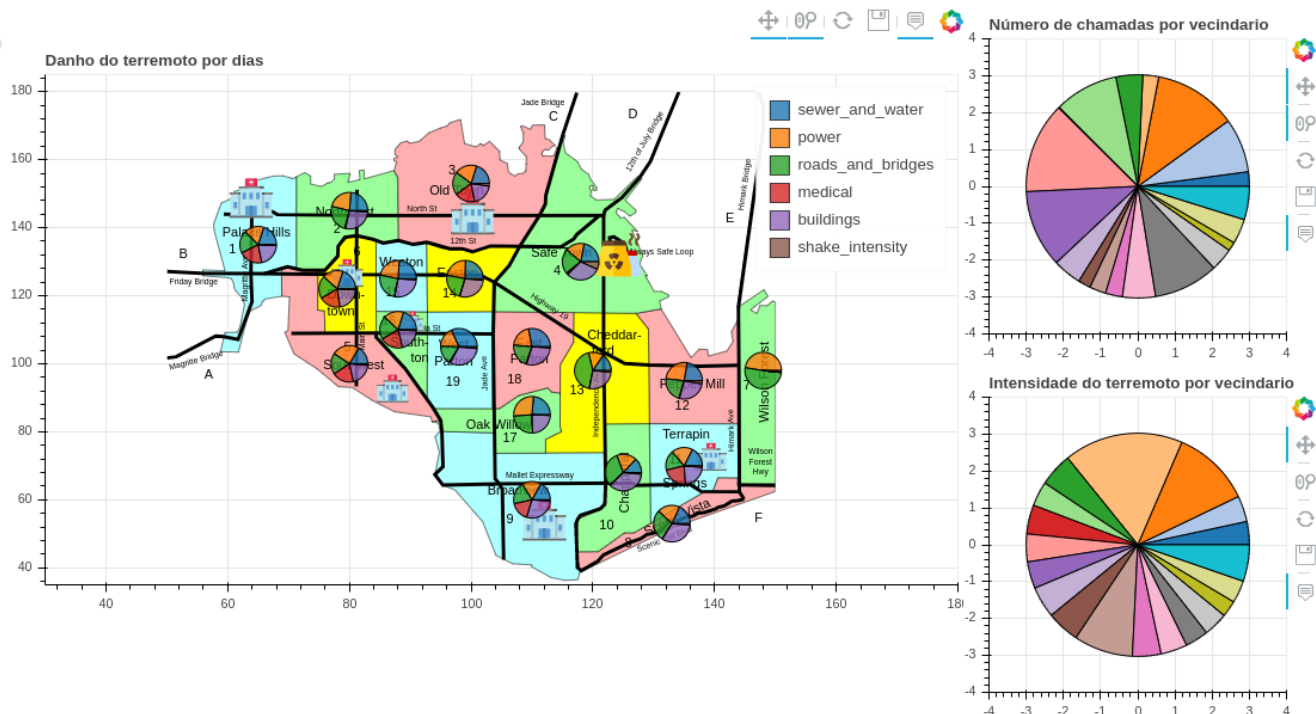


Gráfico 1. Painel de interação diário com dados do terremoto.

Primeiro, uma visão global do mapa é levada em conta, a Figura 1 mostra a magnitude do dano de cada tipo em cada bairro, tornando conhecidas as coisas mais afetadas em cada bairro, sem esquecer que essa informação é a média diária dos danos apresentados.

registro de acordo com o dia: 6

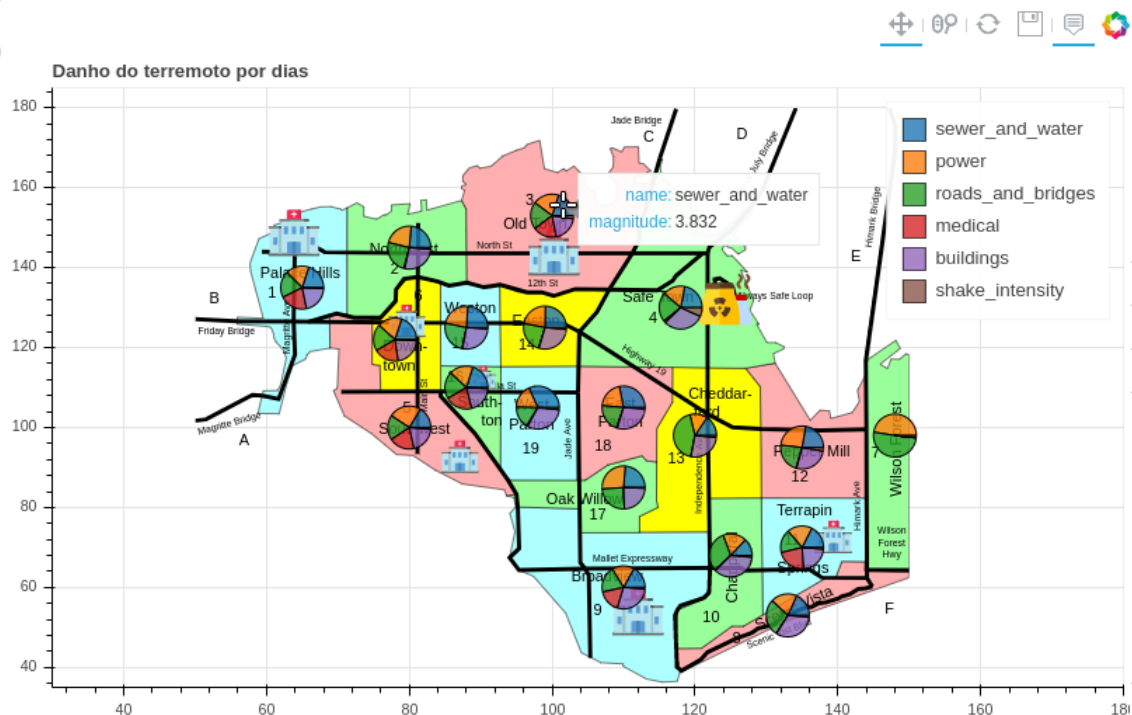


Figura 1. Danos de infraestrutura por dia.

também pode ver nas Figuras 2 e 3 o número de chamadas feitas por cada bairro e a magnitude do terremoto relatado por cada bairro, tendo assim uma proporção de chamadas com a magnitude do dano relatado.

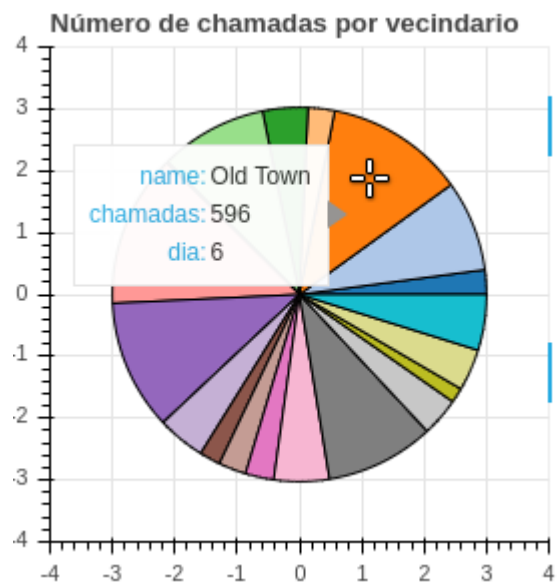


Figura 2. Número de chamadas por bairro.

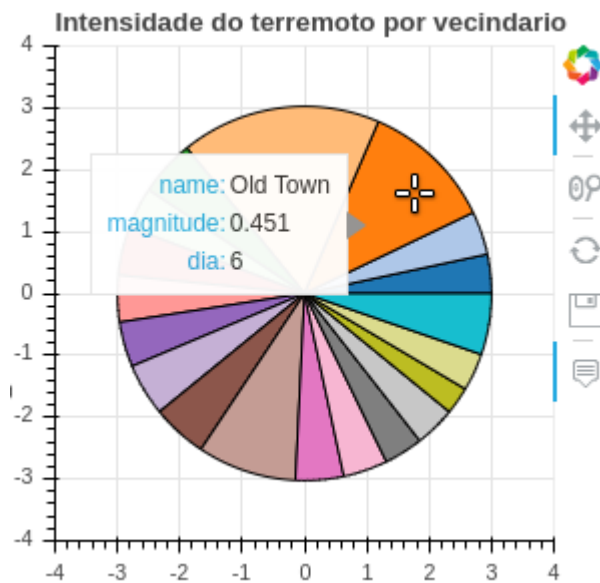


Figura 3. Magnitude do terremoto relatado pelo bairro.

fazer zoom todos os dias, como mostrado na Figura 4, pode-se ver que cada diagrama de pizza representa o dano causado em cada tipo, assim, a relação de danos causados entre cada um dos tipos na vizinhança.

registro de acordo com o dia: 8

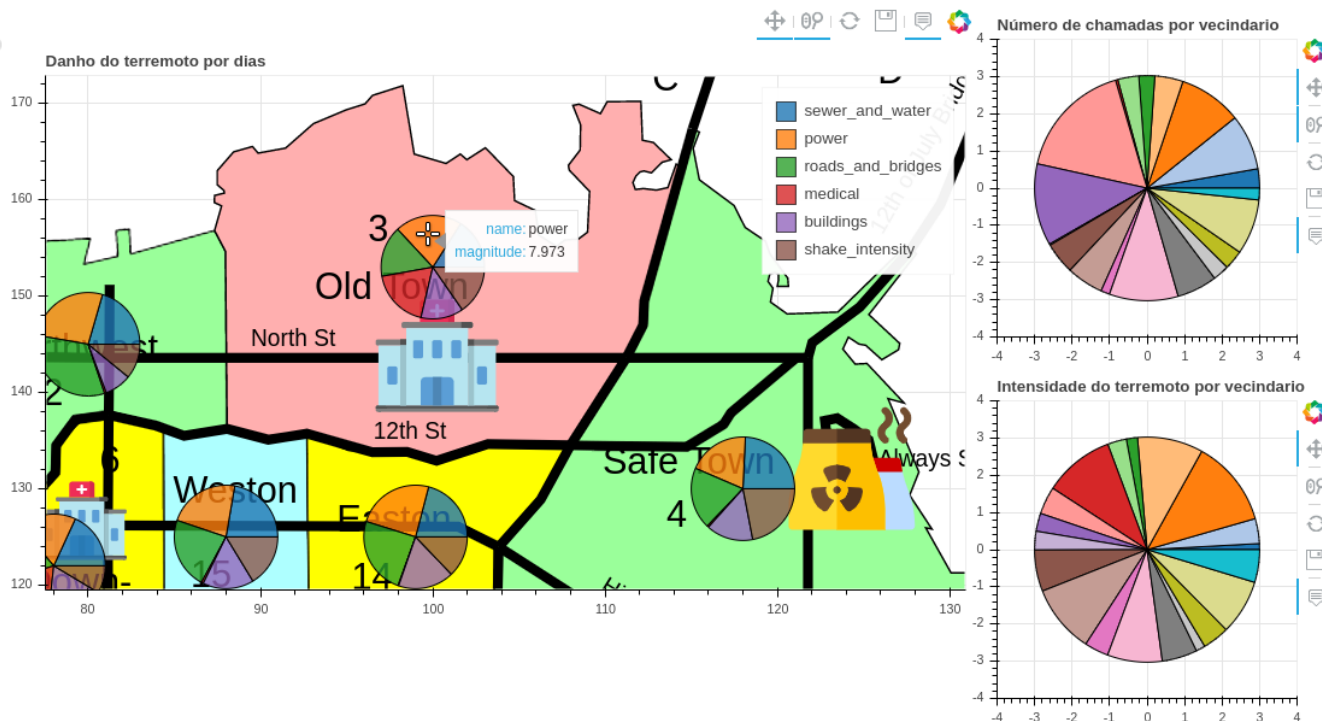


Figura 4. Magnitude de dano por tipo em cada bairro.

Inicialmente, o número de chamadas poderia ser associado à urgência de ser atendido, a Figura 2 mostra essa proporção, mas, ainda assim, pode ser visto na figura 5 que, embora no *Wilson Forest* a magnitude do terremoto relatado tenha sido alta, não houve muitas chamadas, pois é uma reserva natural, então pode-se dizer que não há muitas pessoas, portanto a prioridade é menor.

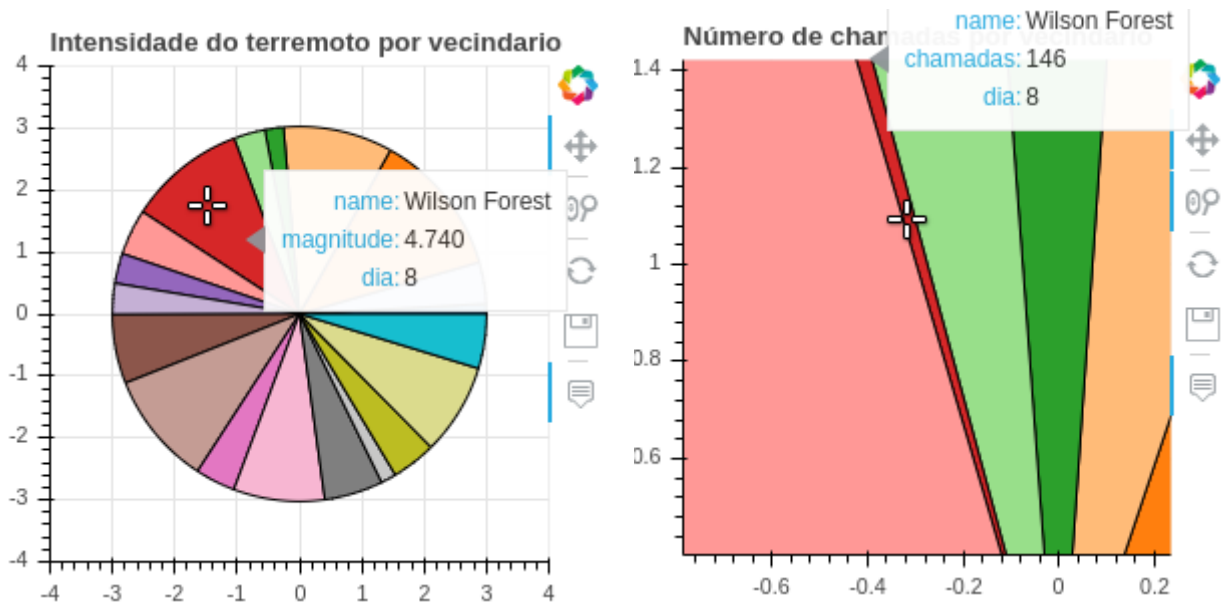


Figura 5. Número de chamadas e magnitude do terremoto para o Wilson Forest no dia 8.

Também pode ser visto na Figura 6 que algumas áreas, como em Safe Town, embora o número de chamadas não tenha sido tão alto quanto em relação a outros bairros, pode-se dizer que ele tem uma importância alta, já que a magnitude do terremoto lá foi alto e lá há uma usina nuclear.

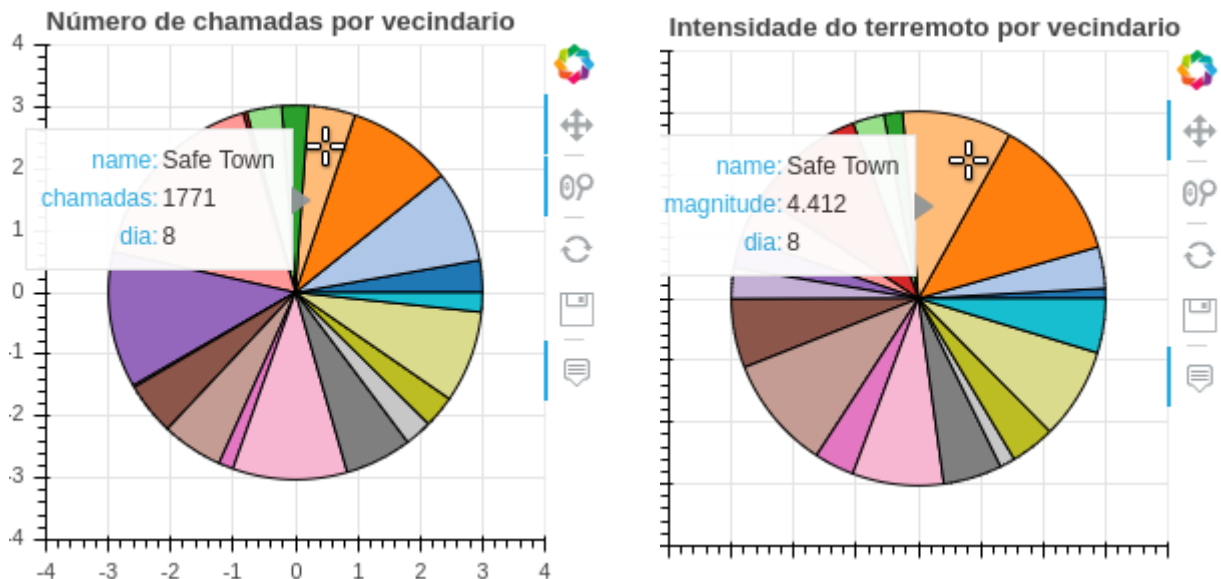


Figura 6. Lista de chamadas com a magnitude do terremoto em Safe Town.

Finalmente, a figura 7 mostra o dano médio de cada medida para cada bairro, revelando, como mostra a figura 8, que a "Old Town" era o bairro mais afetado.

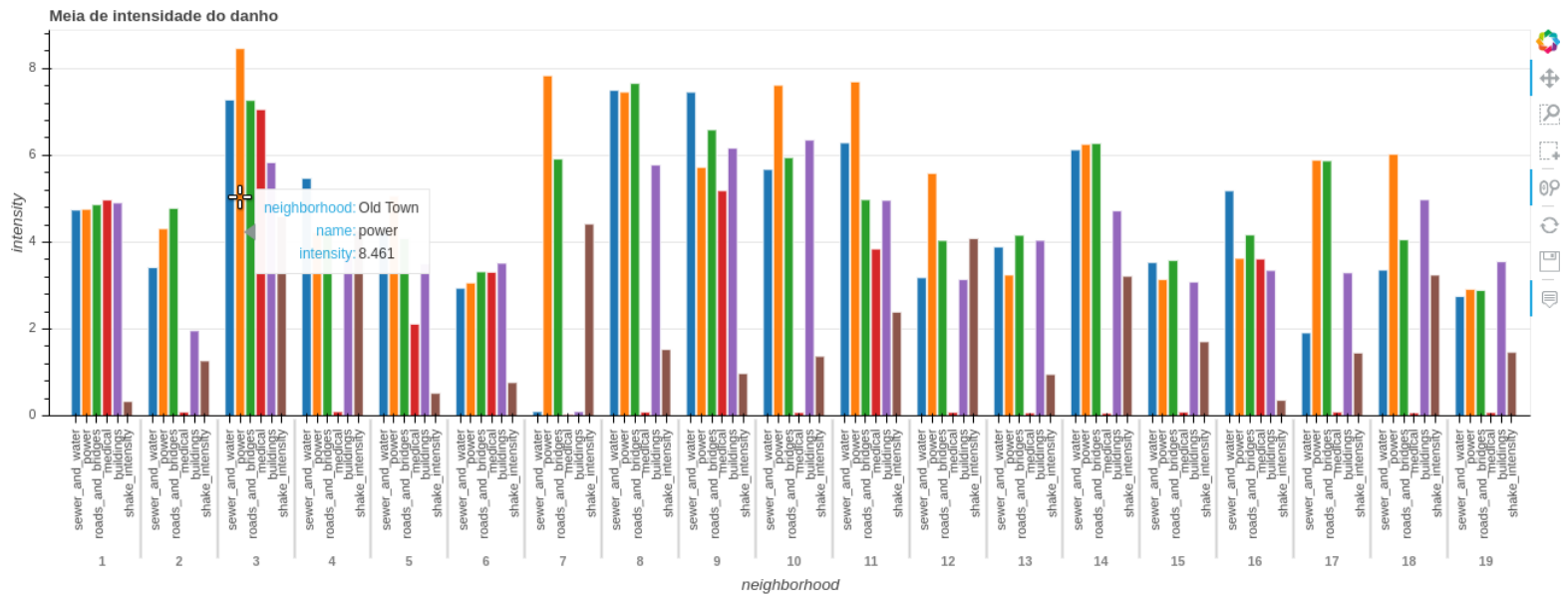


Figura 7. Média de danos para cada medição em cada bairro.

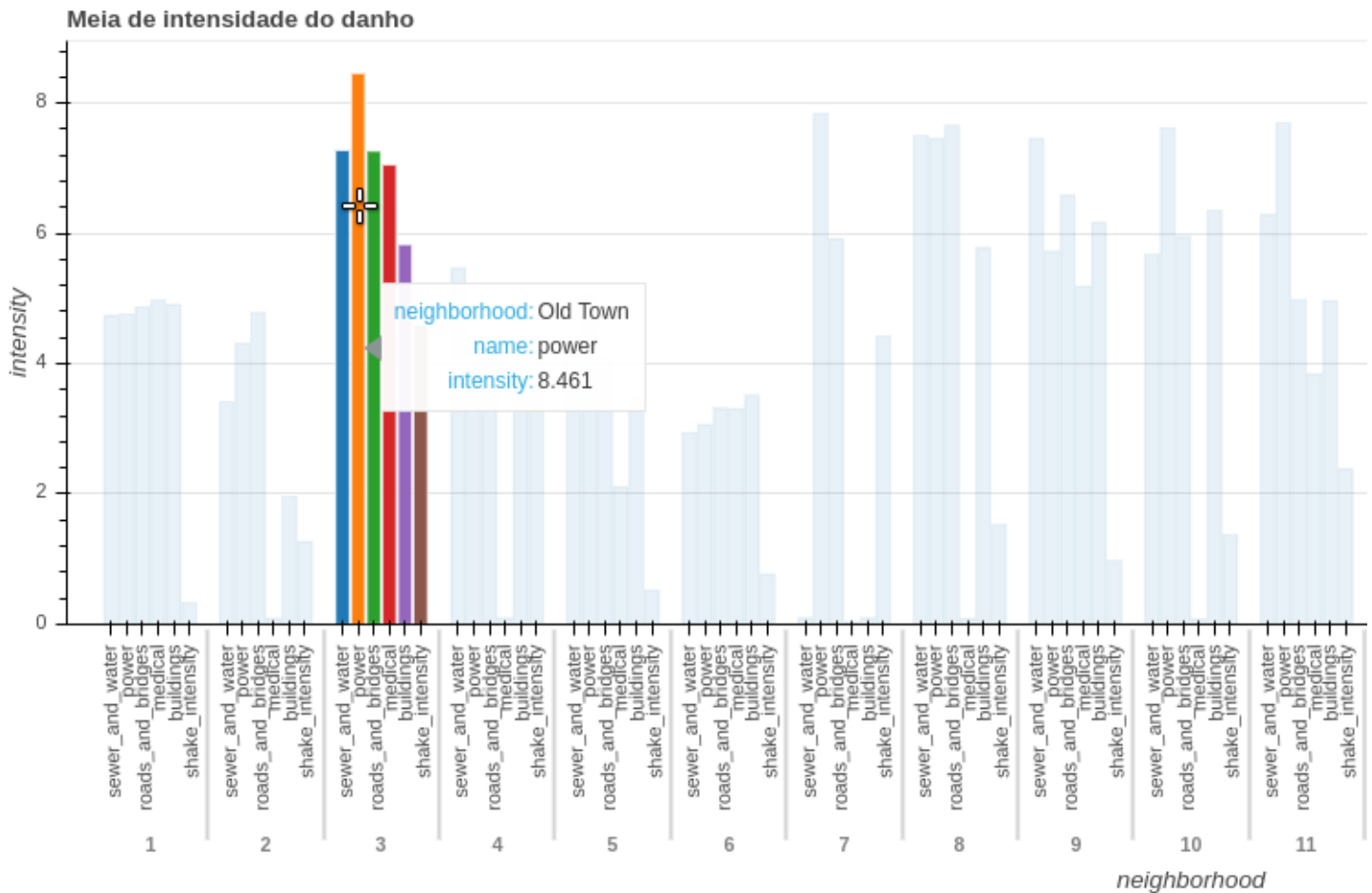


Figura 8. Old Town o qual apresenta os maiores danos.

Com base nos números mencionados acima e interagindo com eles, pode-se dizer que os bairros mais afetados de acordo com os danos médios relatados são Old Town, Safe Town e Broadview. Por outro lado, de acordo com estas medidas, é criada a seguinte ordem de atenção às chamadas, tendo como prioridade os bairros em que as infraestruturas médicas tiveram um dano relevante:

- *Safe Town, Old Town, Palace Hills, Broadview, Terrapin Springs, Southon, Downtown, Southwest.*

Depois de dar prioridade aos bairros com maior número de danos médicos, seria dada prioridade àqueles bairros com altos danos à sua iluminação.

- *Northwest, Scenic Vista, Chapparral, Easton, Pepper Mill, CheddardFord, East Parton, Oak Willow, West Parton, Weston, Wilson Forest*

2 - Use visual analytics to show uncertainty in the data. Compare the reliability of neighborhood reports. Which neighborhoods are providing reliable reports? Provide a rationale for your response. Limit your response to 1000 words and 10 images.

As ilustrações a seguir são baseadas na interação com o gráfico 2, que apresenta uma ilustração na forma de um *boxplot* no qual você pode ver os relatórios que estão longe dos valores médios. Por outro lado, o gráfico de barras à esquerda representa o número de *outliers* por tipo de dano.

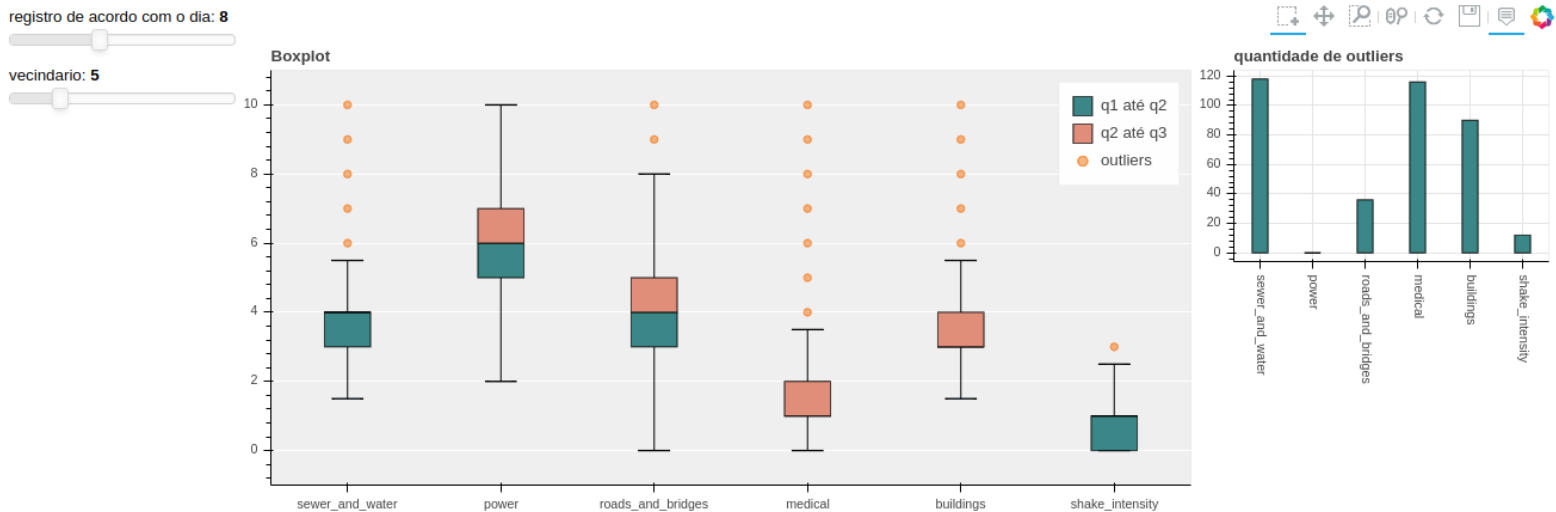


Gráfico 2. Painel de interação diária por bairro com dados do terremoto.

Em princípio, os dados têm valores "nan" que foram considerados como se o usuário não quisesse relatar qualquer dano desse tipo, portanto eles foram considerados como zero.

Alguns bairros relatam danos em infraestrutura médica sem que eles tenham hospitais, a Figura 9 é um exemplo disso, portanto, os usuários não estão fornecendo dados verdadeiros.

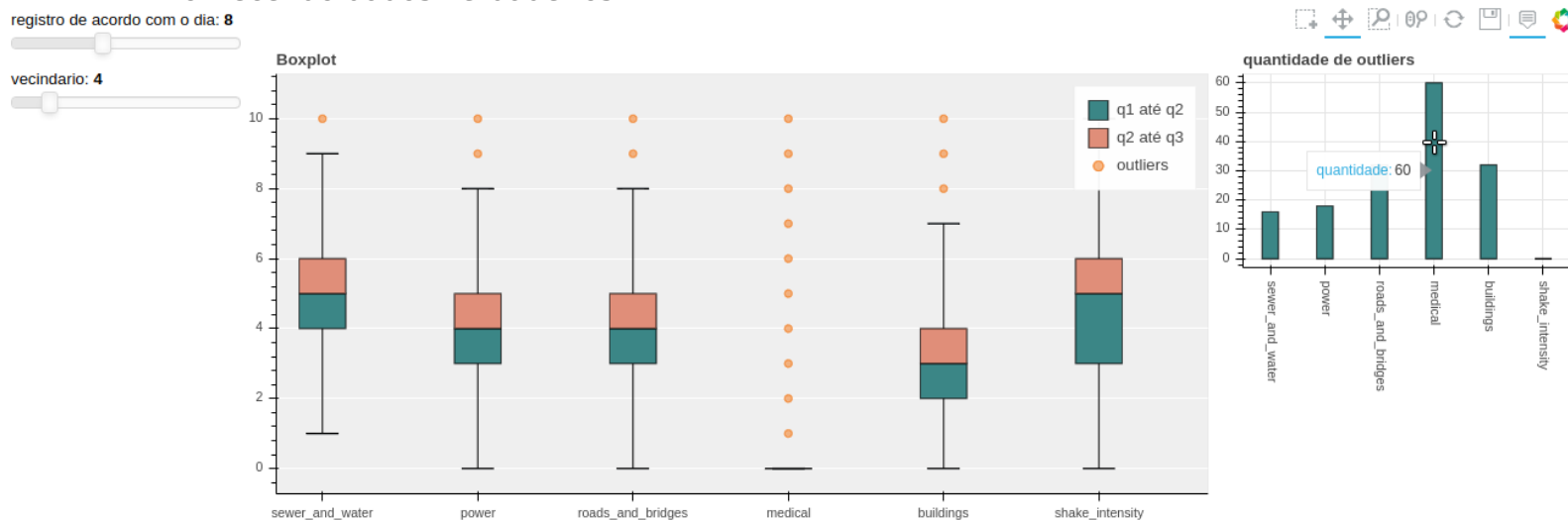


Figura 9. registro do bairro 4 o dia 8.

Por outro lado, existem bairros que apresentam valores muito semelhantes ao longo do tempo, apresentam poucos dados com alta dispersão, a figura 10 mostra uma das vizinhanças, que não apresenta tantos dados inconsistentes.

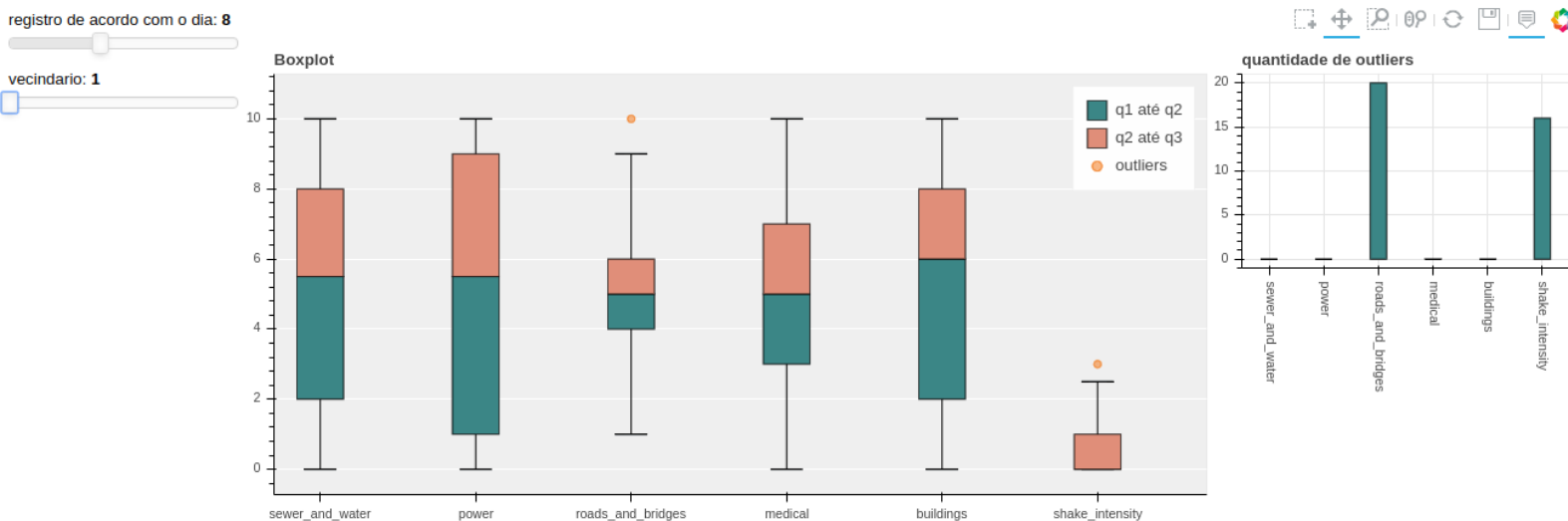


Figura 10. registro do bairro um o día oito, muito poucos outliers.

Embora também existam bairros em que foram reportados valores de dano muito altos, que vão além do padrão normal de relato, são considerados dados exagerados. Como mostrado na Figura 11, muitos dos dados relatados são os mesmos, então o um ou dois quartis não são observados, uma vez que não há variação significativa do quartil para o quartil.

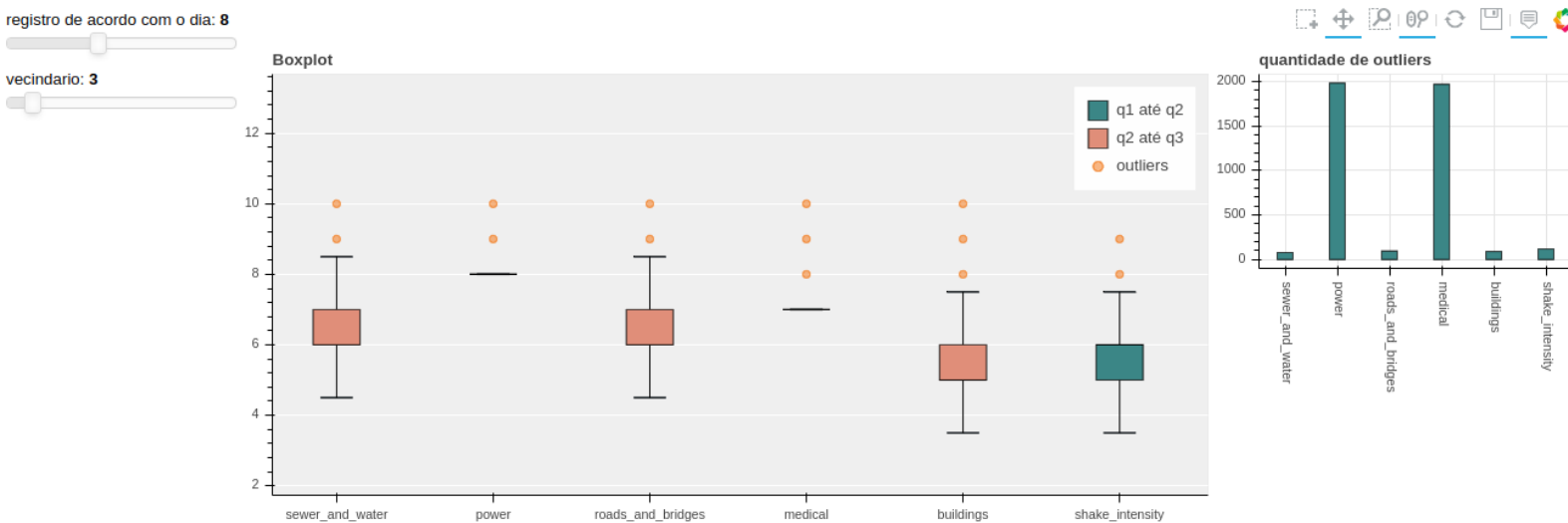


Figura 11. Outliers para o bairro tres o día oito.

Pode-se observar também que, como há muito poucos registros a partir do dia 11 e estes não são muito dispersos, nenhum outlier é apresentado em nenhum bairro. A Figura 12 ilustra o acima mencionado.

registro de acordo com o dia: 11

vecindario: 2

Boxplot

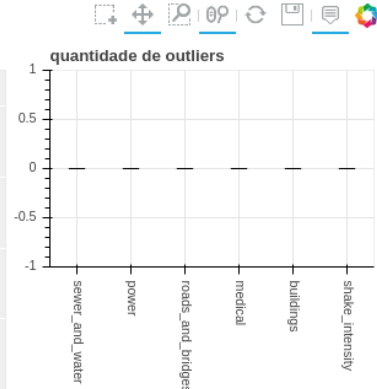
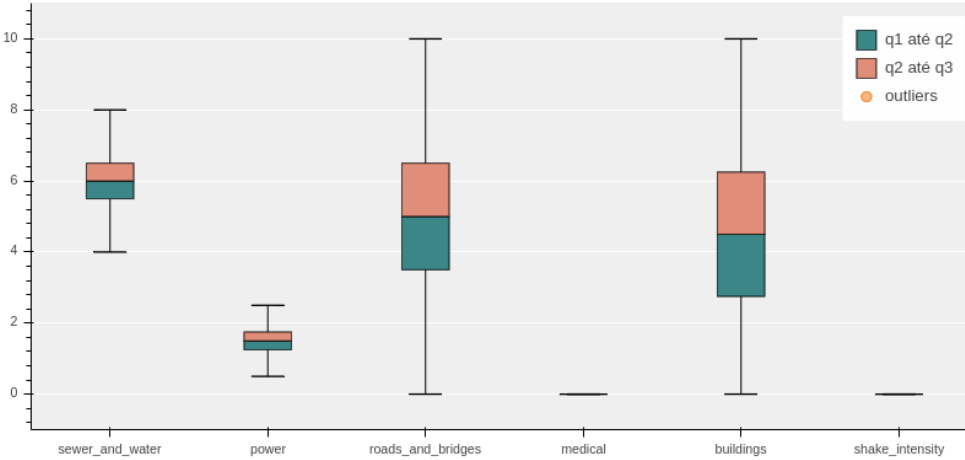


Figura 12. Registro do dano do bairro dois no dia onze, não tem outliers.

Após a interação com o gráfico 2, pode-se encontrar um padrão, que mostra que os dados que possuem um maior nível de dispersão são encontrados nos dias 7 e 8, nestes dias a magnitude relatada para o terremoto também foi maior portanto, pode-se dizer que muitas pessoas ficaram assustadas e relataram dados um pouco altos, apenas os bairros 1 e 10 mantiveram relatórios menos dispersos, como mostram as figuras 13 e 14.

registro de acordo com o dia: 8

vecindario: 1

Boxplot

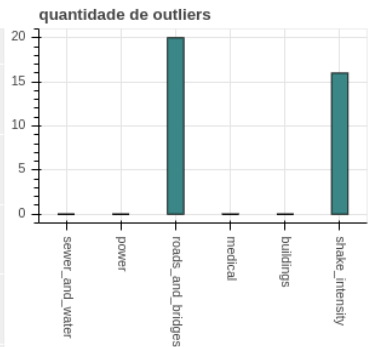
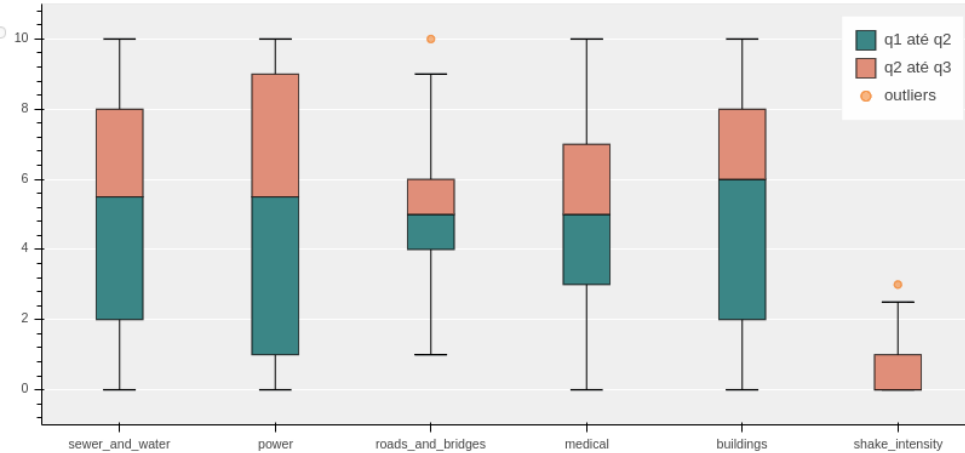


Figura 13. Registro do dano do bairro um o dia oito, só apresenta 36 outliers no total.

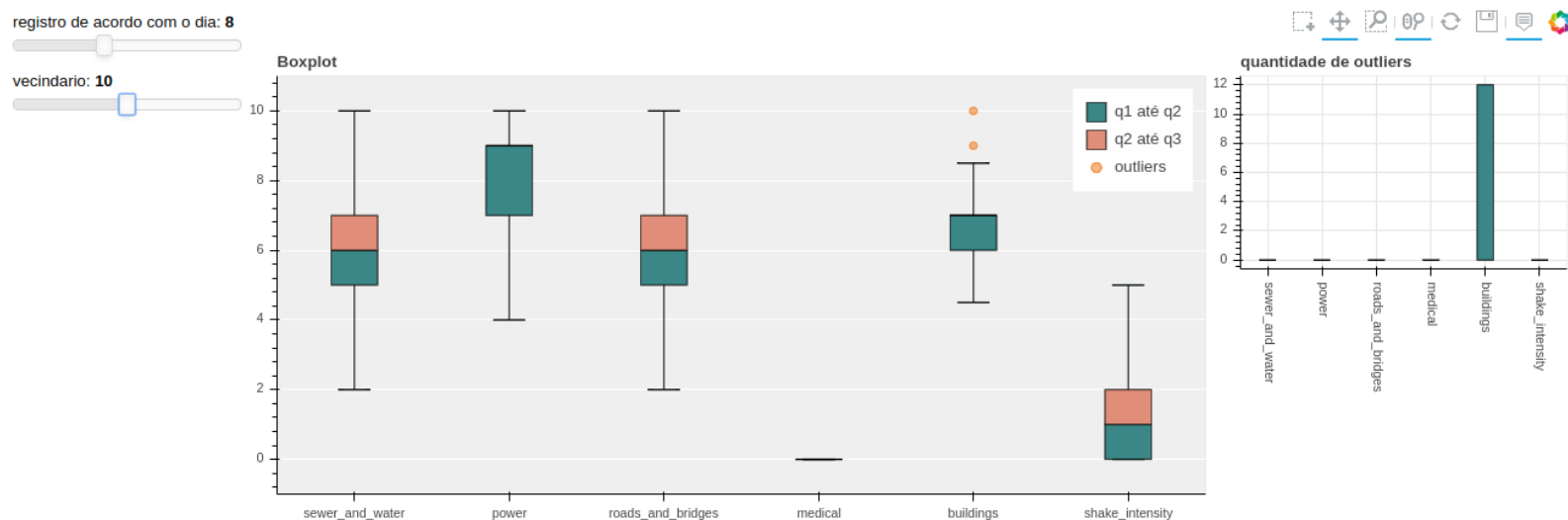


Figura 14. Registro do dano do bairro dez no dia oito, só apresenta 12 *outliers* no total.

Após analisar todos os casos possíveis, pode-se afirmar que os seguintes bairros são aqueles que relatam dados mais confiáveis, entendendo como confiáveis que os valores reportados apresentam menor dispersão.

- Palace Hills, Safe Town, Southwest, Wilson Forest, Chapparal, Cheddarford

3 – How do conditions change over time? How does uncertainty in change over time? Describe the key changes you see. Limit your response to 500 words and 8 images.

As ilustrações a seguir são baseadas na interação com o gráfico 3, que apresenta uma ilustração central que contém todos os relatórios por dia de cada bairro e para cada tipo de dano, podendo assim observar como os relatórios se alteram ao longo do tempo. Também ao redor estão as distribuições que cada variável leva ao longo do tempo. Na parte superior esquerda o usuário pode interagir selecionando o bairro, o dia e o intervalo de horas que deseja analisar, por sua vez, clicando no "rótulo" do gráfico central, esse atributo ficará oculto para melhorar a visualização.

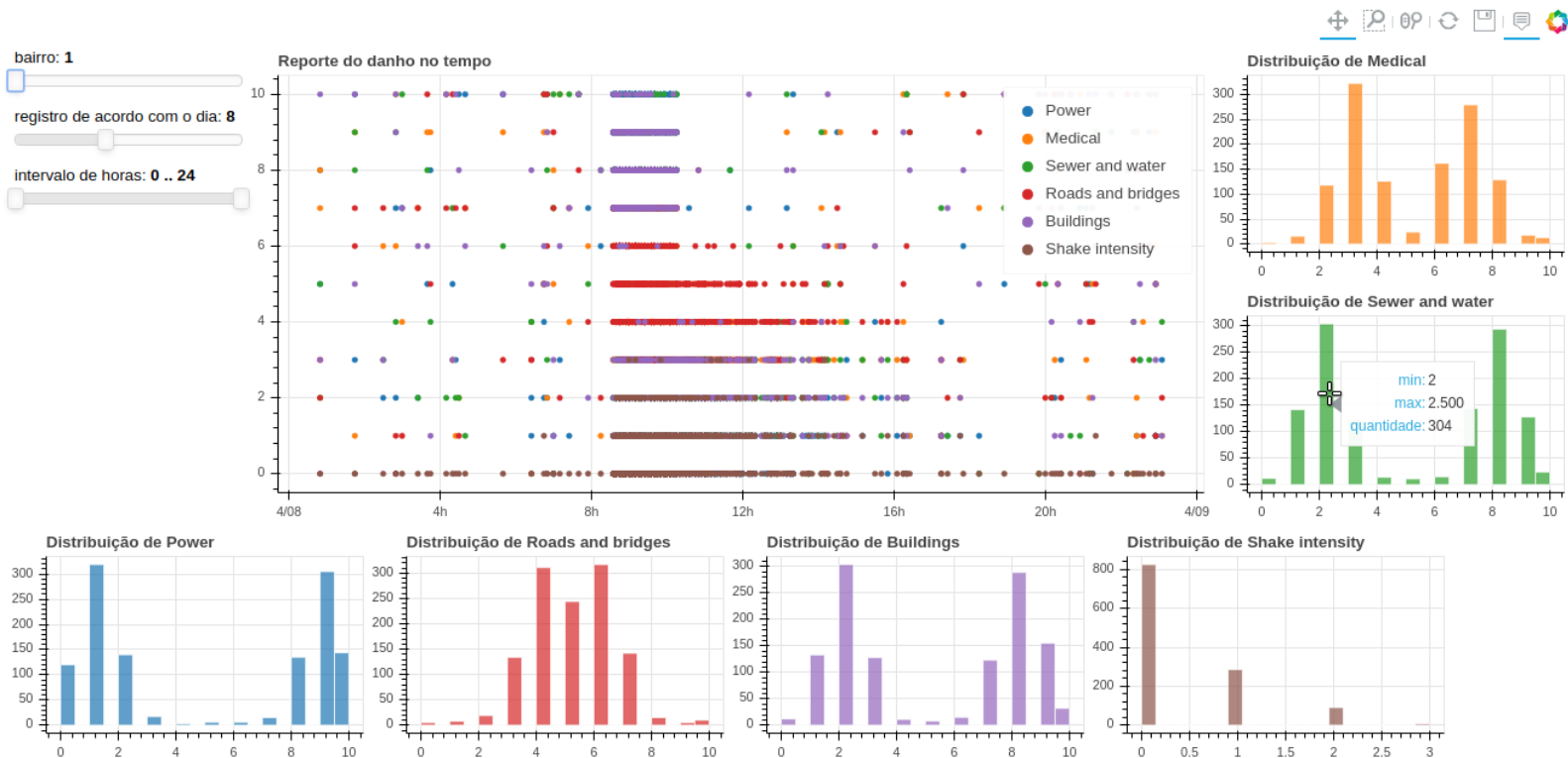


Gráfico 3. Paineis de interação diária por bairro e intervalos de tempo.

Uma das primeiras coisas que podem ser observadas, é que os dias 8 e 9 são aqueles em que ocorre o pico do terremoto, estes dias a intensidade mais relatada do terremoto, sendo que para o dia 8 o tempo entre 8 da manhã e 4 da tarde, onde as pessoas perceberam o terremoto. Enquanto no dia 9 os relatórios de intensidade estavam concentrados entre as 15h e as 19h. As Figuras 15 e 16 mostram isso respectivamente. Além disso, eles mostram que certos tipos de danos têm a mesma distribuição.

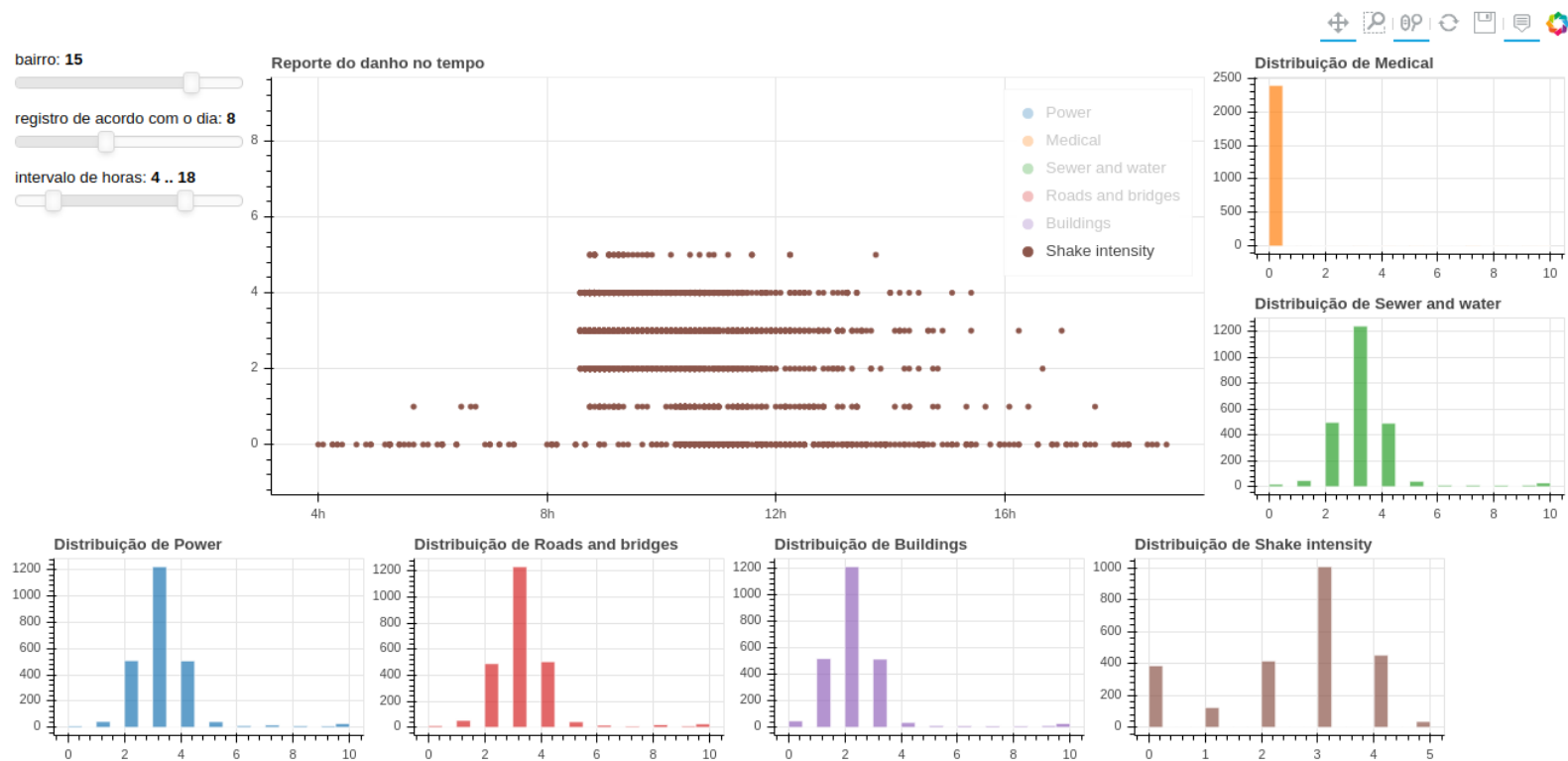


Figura 15. Intensidade do terremoto informada pelo bairro 15 no dia 8, todos os bairros têm esse intervalo horário de relato, embora apenas alguns, assim, tenham seus tipos de danos relacionados, você poderia dizer que eles são os mesmos.

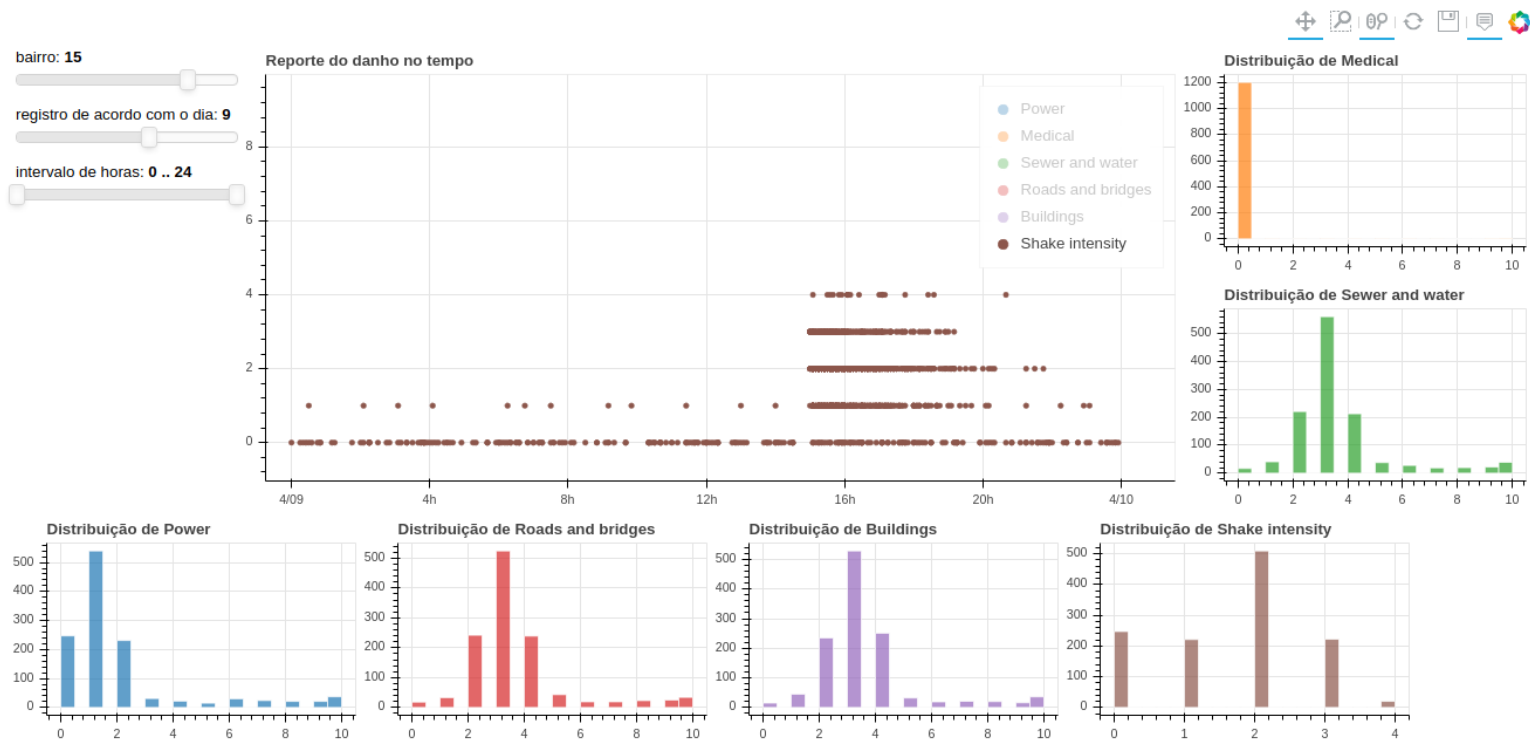


Figura 16. Intensidade do terremoto relatado pelo bairro 15 no dia 9, todos os bairros têm esse intervalo horário de relato, embora apenas alguns, assim, tenham seus tipos de danos relacionados, você poderia dizer que eles são os mesmos.

Ao interagir com o gráfico 3, pode-se observar que certas anomalias aparecem nos dados, uma vez que apresentam uma grande incerteza, a figura 17 mostra como para os danos relatados em "*buildings*" há dois picos de concentração, que inicialmente, acreditar-se-ia que os dados deveriam seguir uma distribuição normal, embora para aquele dia naquele bairro, a maioria seguisse essa anomalia.

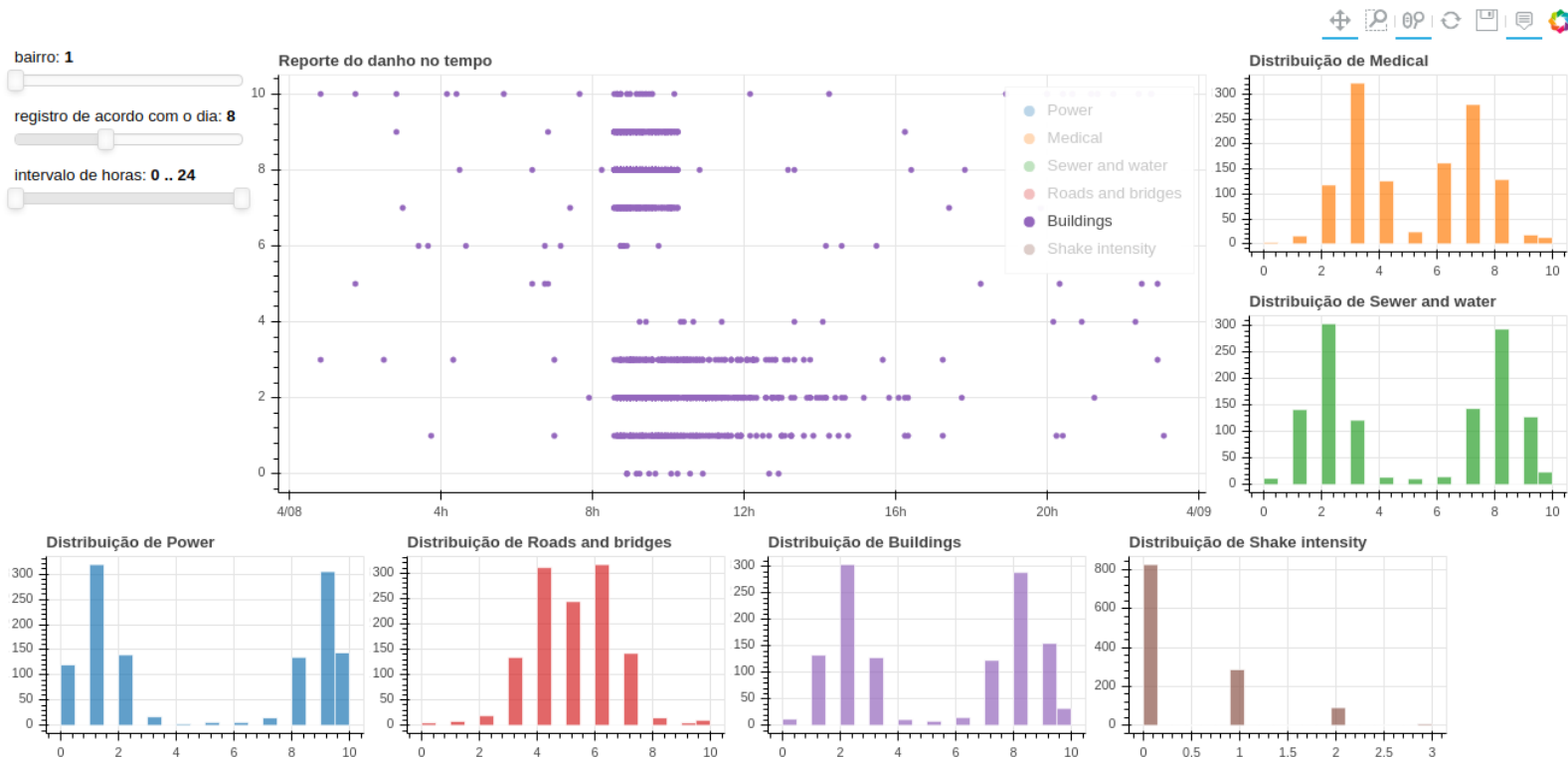


Figura 17. Relatório de danos em edifícios no bairro 1 no dia 8. Pode ser visto que quase todos os danos relatados seguem uma distribuição para dois picos, alguns com picos mais distantes do que outros.

Pode ser visto que o dano significativo relatado pelo bairro 7 (Wilson Forest) ocorreu no dia 8, como é mostrado pela Figura 18. Também pode ser visto que o dano está apenas no “power” e “Roads and Bridges”, embora haja alguns outliers como indicado no gráfico, que pertence a um relatório de danos na infraestrutura médica, sendo que esse bairro não possui infraestrutura médica.

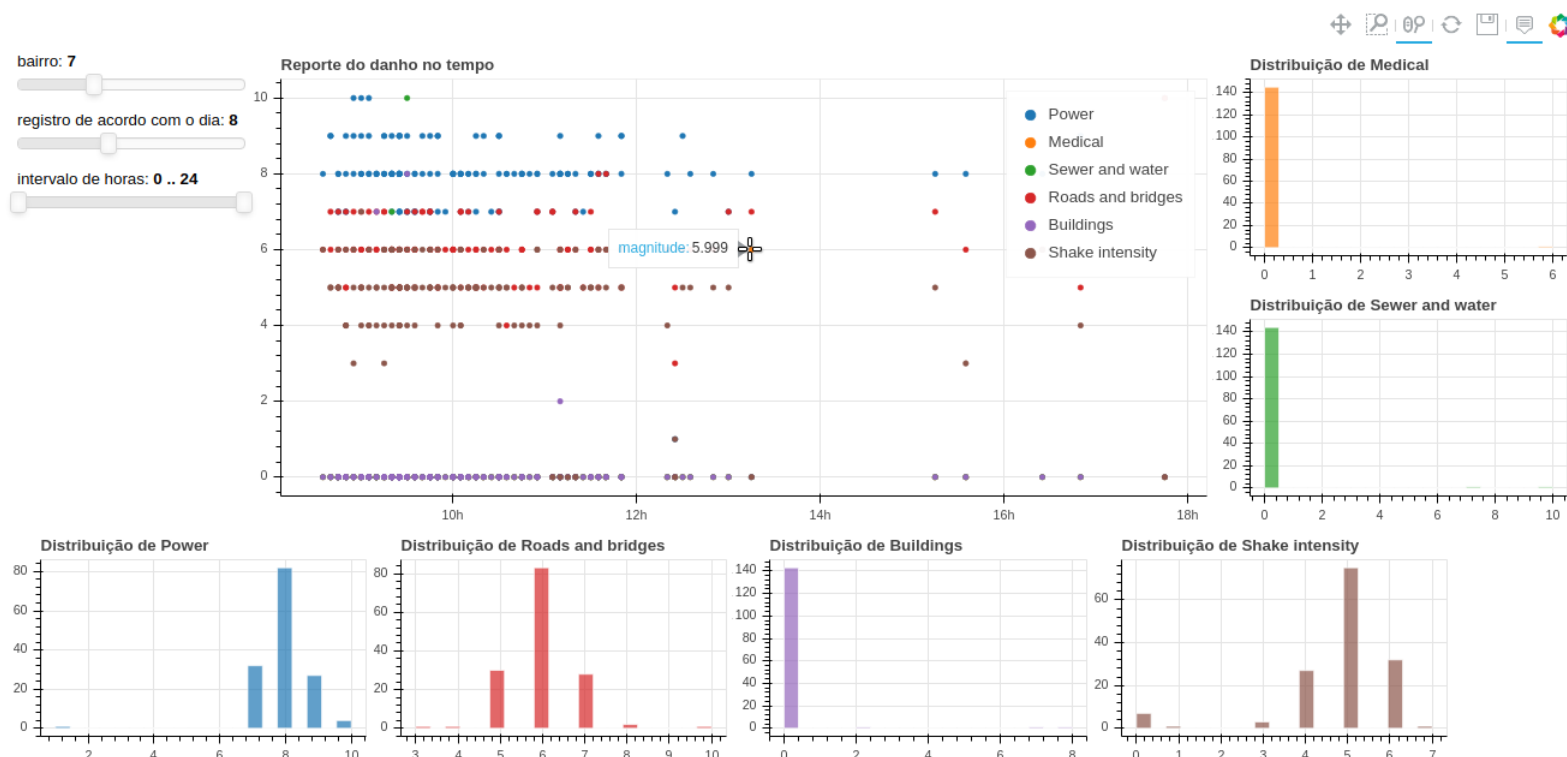


Figura 18. Relatório de danos no *Wilson Forest*. Danos são evidentes apenas em *power* e *Roads and Bridges*, embora pequenos relatórios médicos sejam vistos.

4–The data for this challenge can be analyzed either as a static collection or as a dynamic stream of data, as it would occur in a real emergency. Describe how you analyzed the data - as a static collection or a stream. How do you think this choice affected your analysis? Limit your response to 200 words and 3 images.

Os dados foram analisados como se fossem uma coleta estatística, por exemplo, no primeiro ponto as decisões foram tomadas com base em todos os dados observados ao longo do tempo, outro resultado poderia ter acontecido se os dados tivessem sido tomados como um fluxo, e que ações devem ser tomadas de acordo com um curto intervalo de informações que podem ou não ser verdadeiras.

Ao analisar os dados como uma coleção, foi possível obter estatísticas mais precisas e ter uma visão global do que aconteceu no terremoto.