Entry Name: **"LCEE-MC3"**

VAST Challenge 2019

Mini-Challenge 3

Team Members:
Ana Larissa Dias, Federal University of Pará, larissa.engcomp@gmail.com
Cassio Batista, Federal University of Pará, cassio.batista.13@gmail.com
Edwin Rueda, Federal University of Pará, ejrueda95g@gmail.com
Erick Campos, Federal University of Pará, erick.c.modesto@gmail.com

Student Team: YES

Tools Used:

- Python (v3.5.3)
- Bokeh (v1.2.0)
- NLTK (v3.4.1)
- TextBlob (v0.15.3)
- NumPy (v1.16.0)
- pandas (v0.24.2)
- Matplotlib (v3.0.2)
- word_cloud (v1.5.0)
- scikit-learn (v0.20.2)
- Node.js (v10.16.0)
- LibreOffice Calc (v5.2.7.2)
- grep (v2.27)

Approximately how many hours were spent working on this submission in total? 114 hours

May we post your submission in the Visual Analytics Benchmark Repository after VAST Challenge 2019 is complete? YES

Video: https://youtube.com/?v=vaietefuder

Questions

The City has been using Y*INT to communicate with its citizens, even post-earthquake. However, City officials needs additional information to determine the best way to allocate emergency resources across all neighborhoods of St. Himark. Your task, using your visual analytics on the community Y*INT data, is to determine the types of problems that are occurring across the St. Himark. Then, advise the City on how to prioritize the distribution of resources. Keep in mind that not all sources on Y*INT are reliable, and that priorities may change over time as the state of neighborhoods also changes.

1. Using visual analytics, characterize conditions across the city and recommend how resources should be allocated at 5 hours and 30 hours after the earthquake. Include evidence from the data to support these recommendations. Consider how to allocate resources such as road crews, sewer repair crews, power, and rescue teams. 1000 words, 12 images.

   In order to discern between reliable and unreliable messages, we have considered a bar chart of frequency of users that tweeted the most with a hover tool to show the most frequent words spoken by each user. Figure 1 shows the unfiltered distribution, in which the red bars represent users that are probably sellers since their messages include too many words like "*deal*", "*offer*", "*opportunity*", "*chances*", etc. All tweets from those users have therefore been discarded.

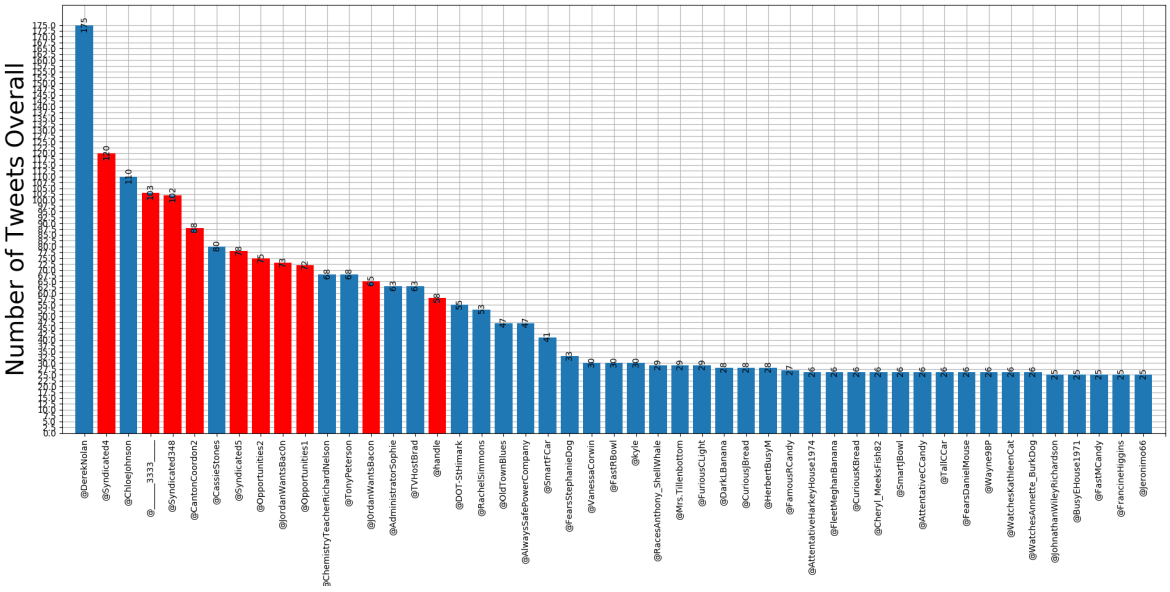


Figure 1: Most frequent tweeters (usernames). Red bars reresent sellers.

There appears to have occurred three earthquakes, but the first one seems to have started at 2 PM of the first day (April 6th), as can be seen at the heatmap of Figure 3a. The heatmap is divided by a time interval of one hour and by neighbourhood location. It was generated by considering a list of keywords that were tweeted and are similar in meaning (synonyms) and are directly related to the word "earthquake". The list is: "*shake*", "*shudder*", "*vibrate*", "*wobble*", "*tremor*", "*tremble*", "*quaver*", "*quiver*", "*hazard*", "*disaster*", "*destruction*", and "*rubble*".

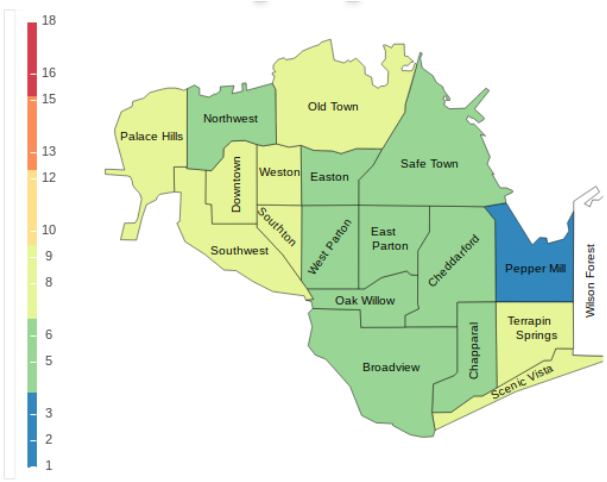
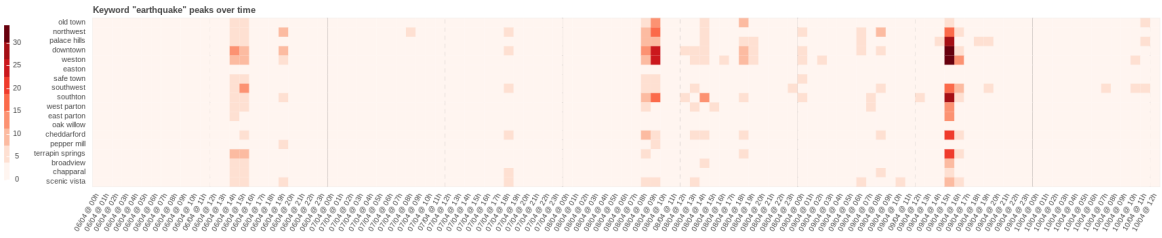


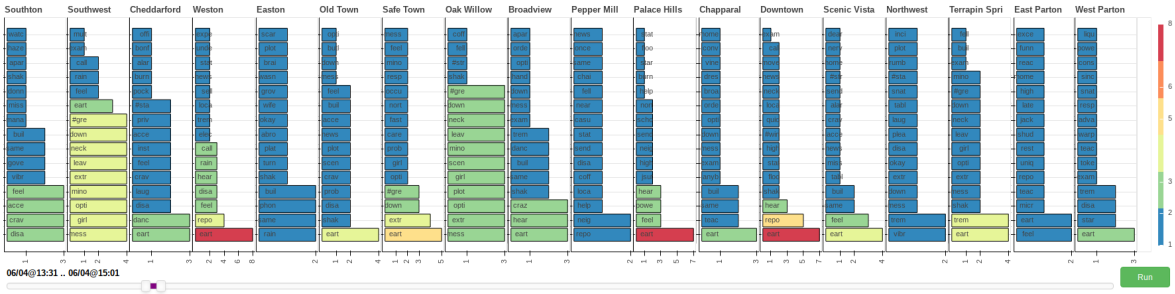
Figure 2: St. Himark's map 5h after the first earthquake. Lighter shades of green represent a higher frequency of messages for each location.

To ensure the heatmap is providing a reliable information, a horizontal bar chart, shown in Figure 3b, was used from 1:30 PM to 3:00 PM. It counts isolated words and discards words such as adverbs, pronouns, adjectives, articles and some nouns and verbs that were considered to be useless such as “anyone”, “make”, “know”, “food”, “hate”, etc. Then the Porter Stemmer from the `nltk` package was used to clip the words by its invariant parts (word root), and that root was further reduced to 4-chars only. A heat-like colormap from blue to red was also included to enhance frequency distinction.

The bar chart shows some interesting other words such as “*feel*”, “*hear*”, “*report*” apart from the keywords aforementioned, in which “*earthquake*” is the most frequent one in accordance with the red color.



(a) Heatmap considering cluster of similar keywords (synonyms)



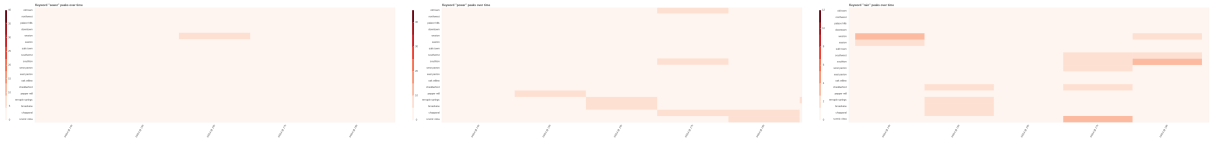
(b) Bar chart with blue-to-red colormap considering frequency of words from 1:00 PM to 3 PM of April 6th.

Figure 3: Earthquake start

Figure 2 shows a dynamically-colored SVG map of the city, where the neighbourhood inner colors range from blue to red in a heatmap fashion according to the average mean of the colors of the five bigger bars of the horizontal bar chart for each location.

Figure 4 shows the heatmap per keyword in a five-hour-time interval from 2:00 PM to 6:59 PM. The top shows 3 blank graphs for the keywords “building”, “medical”, and “road”, which means these resources do not appear to be requested by any neighbourhood. On the other hand, the 3 graphs at the bottom show the number of mentions for keywords related to “sewer and water” (Figure 4d), “power” (Figure 4e), and “rain” (Figure 4f).

(a) Building (b) Medical (c) Roads and Bridges



(d) Sewer and water (e) Power (f) Rain

Figure 4: Conditions after 5h of the first earthquake

Suggestions for crew allocation is detailed as follows:

- *Sewer and water:* A crew must be sent only to Weston between 4:00 PM and 4:59 PM.
- *Power*: Issues have occurred in Pepper Mill, **Terrapin Springs**, **Broadview**, Chapparal, **Southton**, **Old Town** and Scenic Vista, but we'll consider only the locations highlighted in bold because they have hospitals.
  - A crew must be sent to Terrapin Springs between 3:00 PM and 3:59 PM. Broadview also has a power demand at this time interval but the tweet frequency is much lower considering the five-hour period.
  - Two crews must be sent to Old Town and Southton between 4:00 PM and 4:59 PM.
  - Lastly, the crew from Terrapin Springs can be reallocated to Chapparal between 5:00 PM and 5:59 PM. Although Chapparal does not have hospitals, it has been nearly two hours with electrical issues.
- *Rescue, sewer and water*: A crew must be sent to Southton only because there have been small issues in Weston, Southton and Downtown, and therefore Southton is geographically in the middle of such neighbourhoods.

Looking at the useful-words colormap over the SVG map of St. Hirmak, it can be inferred right at the outset that the neighbourhoods that are in most need are Downtown, Southton, Old Town and Weston.

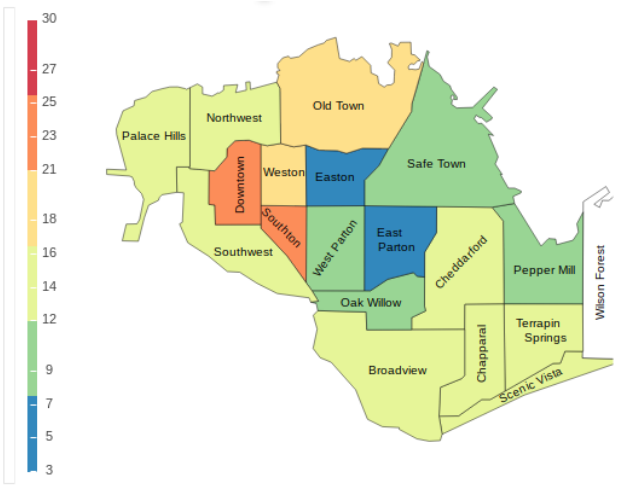
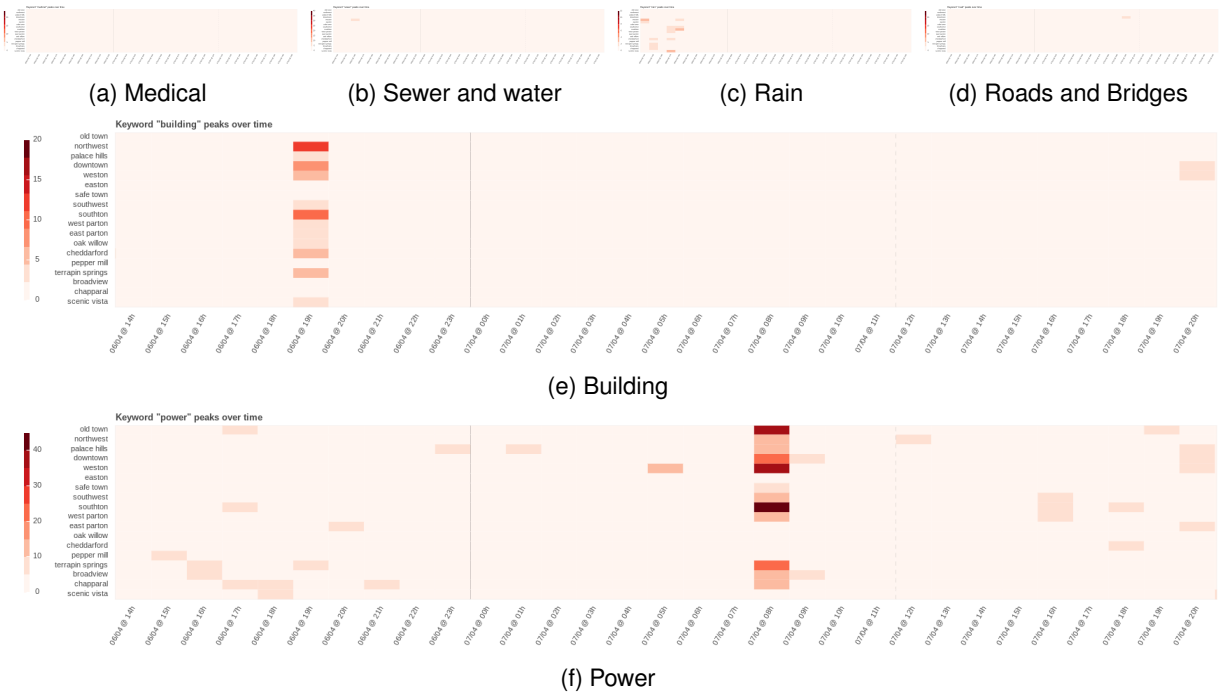


Figure 5: St. Himark's map 30h after the first earthquake. Yellow and orange shades represent a higher frequency of messages for each location.

By looking at the heatmap of Figure 6 for each keyword, it can be seen that there have been no occurrences for medical, and the ones related to sewer/water and rain have already been attended within the first five hours. With respect to roads and bridges in particular, there have been 3 occurrences from 10:00 AM to 11:00 AM of April 8th at Downtown, but this neighbourhood is under resurfacing maintenance, which implies a road crew is already working there.

Suggestions for crew allocation is detailed as follows:

- *Building*: On April 7th from 7:00 PM to 7:59 PM there have been multiple casualties on almost all locations so we would prioritize the dark-red-colored ones according to the heatmap of Figure 6e: Northwest, Southton, Downtown, and Weston. All four have a high density of buildings and people, apart from being geographically close to each other, which can be an advantage for an eventual reallocation of crews in the following hours. Terrapin Springs and Cheddarford also have some less-intense occurrences, but they must be ignored due to the absence of high buildings.
- *Power*: According to Figure 6f on April 7th there have been sporadic, less-intense occurrences that could be solved by sending small units to individual locations, but from 8:00 AM to 9:00 AM an energy disaster appear to have affected almost all neighbourhoods. Again we would prioritize regions where the keywords were mentioned the most: Southton, Old Town, and Weston. The other can be later attended in the following hours.



(a) Medical (b) Sewer and water (c) Rain (d) Roads and Bridges

(e) Building

(f) Power

Figure 6: Conditions after 30h of the first earthquake

2. Identify at least 3 times when conditions change in a way that warrants a re-allocation of city resources. What were the conditions before and after the inflection point? What locations were affected? Which resources are involved? Limit your response to 1000 words and 10 images.

   As already mentioned, there have been 3 independent earthquakes, as could be seen in Figure 3a: Apr 6th 2:00 PM, Apr 8th 7:00 AM and Apr 9th 3:00 PM, approximately. Those two last earthquakes are emphasized in Figure 7.

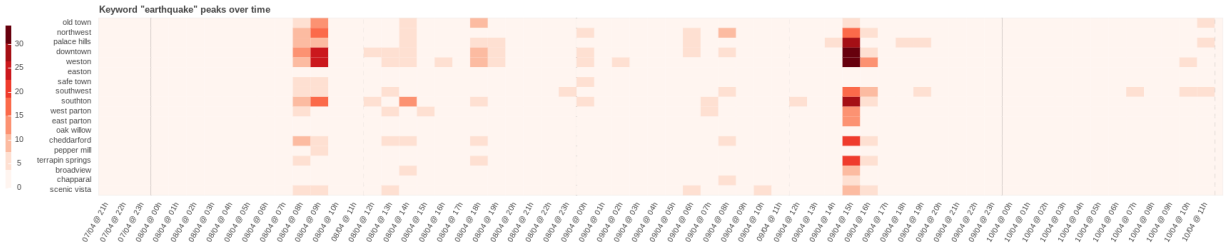


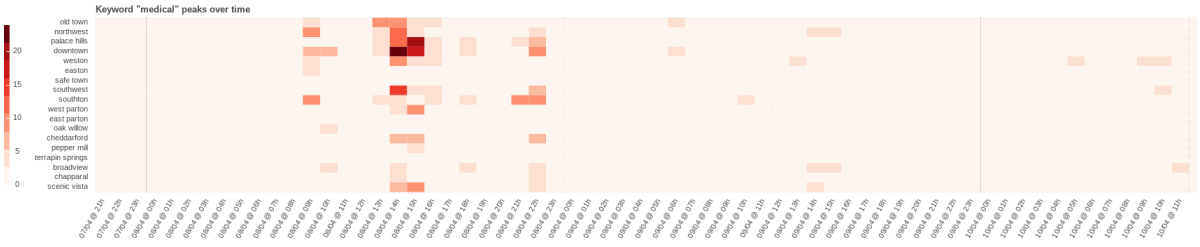
Figure 7: Heatmap for the two last earquakes using keywords related to the word “earthquake”.

3. Take the pulse of the community. How has the earthquake affected life in St. Himark? What is the community experiencing outside the realm of the first two questions? Show decision makers summary information and relevant/characteristic examples. Limit your response to 800 words and 8 images.

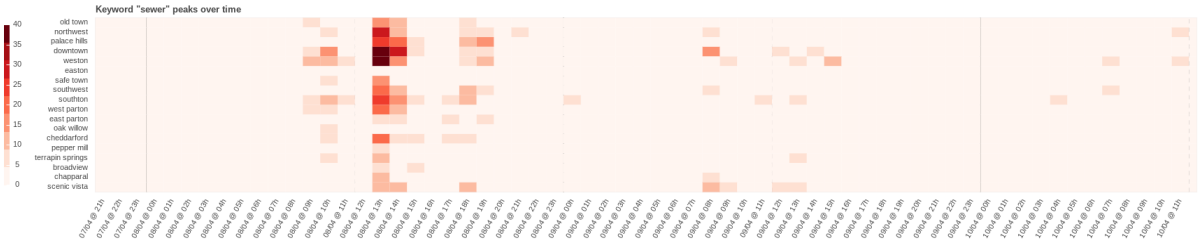
   pay no cy da birguesia

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.

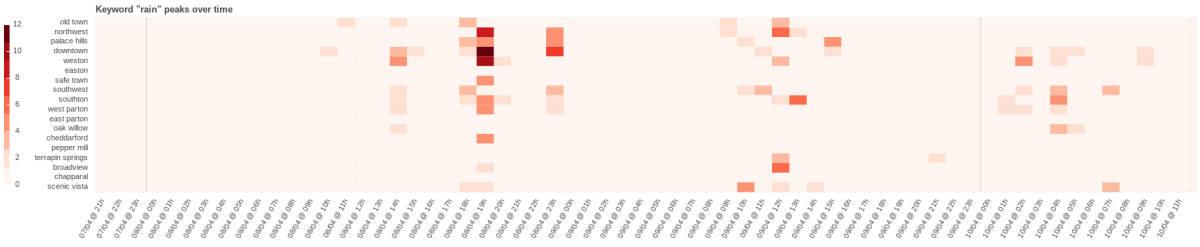
   lyla libre



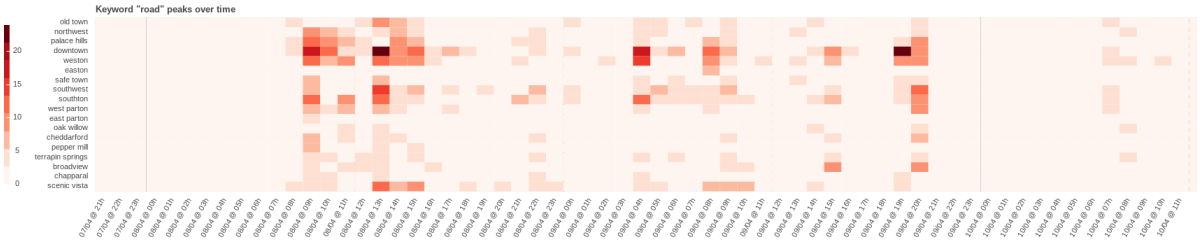
(a) Medical



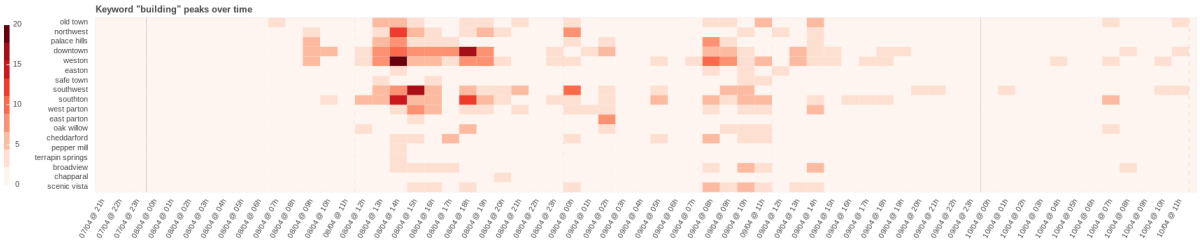
(b) Sewer and water



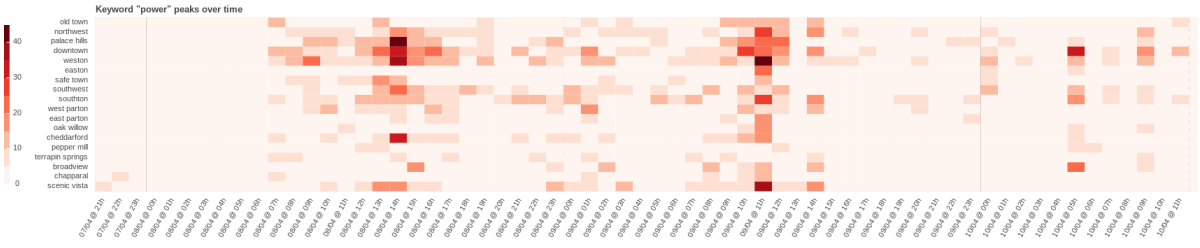
(c) Rain



(d) Roads and Bridges

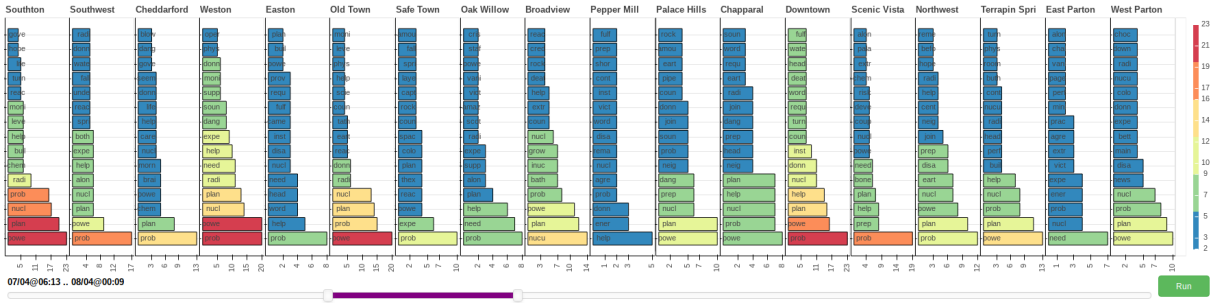


(e) Building

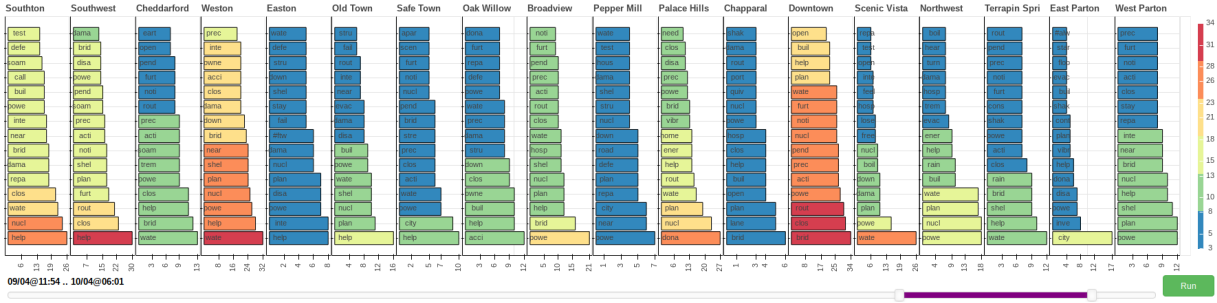


(f) Power

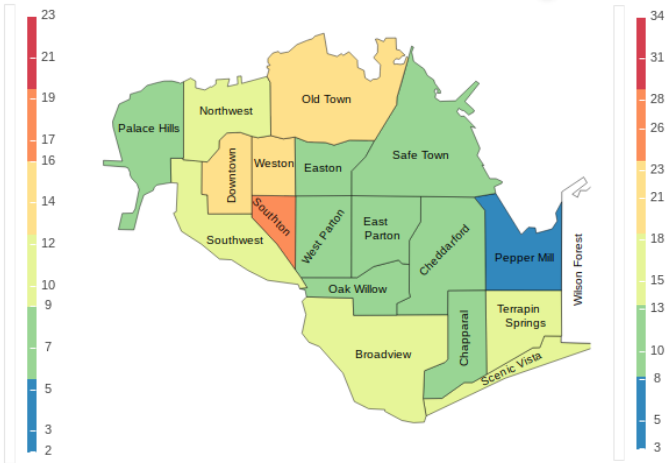
Figure 8: Conditions after 30h of the first earthquake



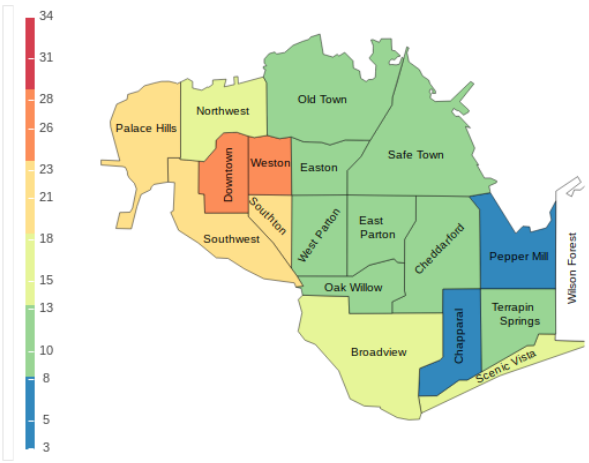
(a) HBar



(b) HBar



(a) SVG



(b) SVG

Figure 10: SVG