# VAST Challenge 2019
# Reviewer Guide: Mini-Challenge 1

This document provides information to support peer review of submissions to VAST Challenge 2019, Mini-Challenge 1. It covers background about the submission structure, the challenge problem, tasks and questions presented to participants, potential answers, and evidence found in the Challenge data supporting these answers. For a full description of the challenge problems and to access the data provided to the participants, please visit https://vast-challenge.github.io/2019/MC1.html.

## Submissions

Participants are required to submit their entries on a standard answer form, along with a video explaining how visual analytics were used to help solve the challenges. Please consider both parts of the submission in your review. If you have difficulty reading the answer form or playing the video, please contact us at vast_challenge@ieeevis.org for assistance.

## Scenario

The VAST Challenge 2019 presents three mini-challenges and a grand challenge for participants to apply visual analytics research and technologies to help a city grapple with the aftermath of an earthquake that damages their nuclear power plant.

### Overview

St. Himark is a vibrant community located in the Oceanus Sea. Home to the world-renowned St. Himark Museum, beautiful beaches, and the Wilson Forest Nature Preserve, St. Himark is one of the region's best cities for raising a family and provides employment across a number of industries including the Always Safe Nuclear Power Plant. Well, all that was true before the disastrous earthquake that hits the area during the course of this year's challenge. Mayor Jordan, city officials, and emergency services are overwhelmed and are desperate for assistance in understanding the true situation on the ground and how best to deploy the limited resources available to this relatively small community.

### Mini-Challenge 1: Crowdsourcing for Situational Awareness

St. Himark has been hit by an earthquake, leaving officials scrambling to determine the extent of the damage and dispatch limited resources to the areas in most need. They quickly receive seismic readings and use those for an initial deployment but realize they need more information to make sure they have a realistic understanding of the true conditions throughout the city.

In a prescient move of community engagement, the city had released a new damage reporting mobile application shortly before the earthquake. This app allows citizens to provide more timely information to the city to help them understand damage and prioritize their response. In this mini-challenge, participants use app responses in conjunction with shake maps of the earthquake strength to identify areas of concern and advise emergency planners.

With emergency services stretched thin, officials are relying on citizens to provide them with much needed information about the effects of the quake to help focus recovery efforts.

By combining seismic readings of the quake, responses from the app, and background knowledge of the city, participants are to help the city triage their efforts for rescue and recovery.

## Ground Truth

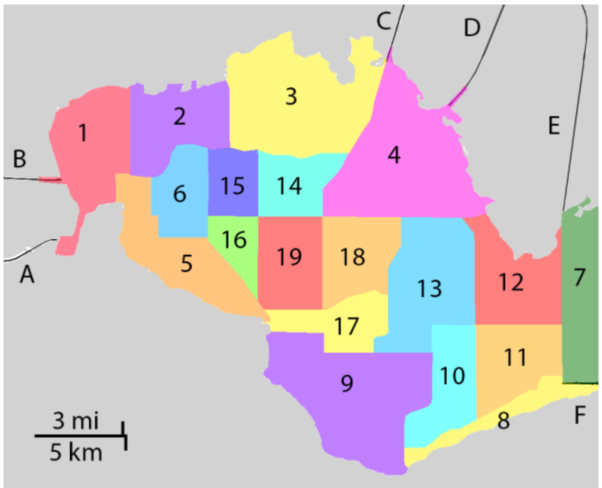



*Figure 1. Neighborhood map*

For reference, the neighborhood locations can be found in Figure 1.

The timeline on the next page shows events the relevant events, ranging from city-wide announcements on social media to earthquakes and their aftermath. This timeline includes the name of each event, whether it is visible within this particular data set, what time it started, and some general notes including effects present in the data. This may provide a convenient snapshot of what is in the data, but we will elaborate further below.

Note that some of the events in the timeline do not manifest themselves directly in data.

| Event | Visible in data? | Start time | Notes/Effects |
| --- | --- | --- | --- |
| Earthquake-2.3 | yes | 4/6/2020 14:33 | Minor earthquake. National Earthquake Prediction Center App releases shakemap, Rumble app reporting rises/falls slightly |
| Announcement to use Rumble app | yes | 4/6/2020 16:00 | City emergency management encourages citizens to use the Rumble app in case of future earthquakes, causing a slight rise in Rumble app reports |
| Earthquake-6.7 | yes | 4/8/2020 8:35 | Major earthquake. National Earthquake Prediction Center App releases shakemap, Rumble app reporting rises/falls sharply |
| Announcement to use shakemap for initial response | yes | 4/8/2020 9:10 | Announcement made, does not affect reporting per se but perhaps shows contestants that damage might correlate with distance from epicenter without being given further information |
| Old Town (Neighborhood 3) damage | yes | 4/8/2020 9:15 | Buildings collapse, driving affected, reports rise/fall on top of immediate earthquake damage |
| Safe Town (Neighborhood 4) damage | yes | 4/8/2020 9:00 | Landslide/liquefaction, buildings collapse, nuc fence/road damaged, driving affected, reports rise/fall on top of immediate earthquake damage |
| Scenic Vista (Neighborhood 8) damage | yes | 4/8/2020 10:05 | Landslide/liquefaction, buildings slide into water, reports rise/fall on top of immediate earthquake damage |
| Announcement of repair procedure and to expect outages | | 4/8/2020 10:05 | Announcement made. Will not affect reporting per se but perhaps preps contestants for blackouts in future reports and which neighborhoods may have had repairs prior to the next event's damage (since damage is all that's reported, a 8->9 seems worse than a 9->6->9 since reporting only shows 9->9) |
| Announcement to use Rumble app | yes | 4/8/2020 10:15 | City system for tracking damage reports is out of service, reports rise/fall on top of immediate earthquake damage |
| Scenic Vista (Neighborhood 8) and Broadview (Neighborhood 9) damage | yes | 4/8/2020 12:00 | Landslide/liquefaction, buildings collapse, residential area damaged, reports rise/fall on top of immediate earthquake damage |
| Palace Hills (Neighborhood 1) accidental misreporting of values | yes | 4/8/2020 8:36 | Up until the 10:15 announcement, a single user drives around uploading damage reports in their neighborhood using an inverted/flipped scale (0 - worst damage, 10 - no damage); overall reports show a bimodal distribution (symmetric around 5) in each infrastructure's damage values, person reports it in social media |
| Cheddarford (Neighborhood 13) and Southton (Nbdh 16) inflated reporting of values | yes | 4/8/2020 8:36 | Throughout the major earthquake and beyond, all reports coming from Neighborhood 13 and Neighborhood 16 are inflated so that a bimodal distribution appears in the total reports, where one peak is the actual mean and the other is 3 points above to coerce the dispatch of help to their regions |
| Blackouts due to power damage at 9 or 10 (out of 10) | yes | ? | Various neighborhoods will have power go out shortly thereafter the damage values Grant provided reach 9 or 10 (out of 10); this is random for now but needs to be decided/fixed; this results in chunks of time where reports are non-existent (which is distinctly different than the non-event times when random reports stream in infrequently) |
| All Neighborhood repairs | yes | 4/9/2020 12:00 | state change in report generation, but no new reports generated per se (essentially separated repairs in as46 from new damage caused by aftershock) |
| Announcement of repairs taking awhile | | 4/9/2020 0:30 | Announcment made. Will not affect reporting per se, but gives contestants background info why some and not all neighborhood damage is improved/blackouts have been handled |
| Aftershock-4.6 | yes | 4/9/2020 15:00 | Major aftershock hits, National Earthquake Prediction Center App does not release a shakemap, but it should look similar to previous one?, Rumble app reporting rises/falls sharply |
| Broadview (Neighborhood 9) damage | yes | 4/9/2020 16:00 | Landslide/liquefaction, reports rise/fall |
| Scenic Vista (Neighborhood 8) damage | yes | 4/9/2020 16:30 | Landslide/liquefaction, reports rise/fall |
| Safe Town, Pepper Mill, Easton, and Oak Willow (Neighborhoods 4, 12, 14, and 17) | yes | ? | 14: 0200, 4: 1100, 12/17: 1130; Power outages, may not show up in rumble app except for complete silence (may not be caught by contestants) |

## Data

The data for Mini-Challenge 1 includes one comma-separated values (CSV) file spanning the entire length of the timeline, which contains individual reports of shaking/damage by neighborhood over time. Reports are made by citizens at any time; however, they are only recorded in 5-minute batches/increments due to the server configuration. Furthermore, delays in the receipt of reports may occur during power outages.

The data includes the following fields:
- Time: timestamp of incoming report/record
- Location: ID of neighborhood where person reporting is feeling the shaking and/or seeing the damage
- {shake_intensity, sewer_and_water, power, roads_and_bridges, medical, buildings}: reported categorical value of how violent the shaking was/how bad the damage was (0 - lowest, 10 - highest; missing data allowed and designated as NaN)

| | time | sewer_and_water | power | roads_and_bridges | medical | buildings | shake_intensity | location |
|---|---|---|---|---|---|---|---|---|
| **0** | 2020-04-08 17:50:00 | 10.0 | 6.0 | 10.0 | 3.0 | 8.0 | NaN | 1 |
| **1** | 2020-04-09 13:50:00 | 2.0 | 10.0 | 0.0 | 8.0 | 4.0 | 0.0 | 1 |
| **2** | 2020-04-09 00:20:00 | 7.0 | 10.0 | 10.0 | 9.0 | 10.0 | 0.0 | 1 |
| **3** | 2020-04-08 17:25:00 | 1.0 | 1.0 | 2.0 | 10.0 | 7.0 | NaN | 1 |
| **4** | 2020-04-08 02:50:00 | 9.0 | 7.0 | 1.0 | 6.0 | 9.0 | NaN | 1 |

*Figure 2. First five rows of the generated report data*

Also included in the released data are two shakemap (PNG) files which indicate where the corresponding earthquakes' epicenters originate as well as how much shaking can be felt across the city. Figure 3 shows a shakemap of the precursor to the main earthquake on April 6, while Figure 4 shows that of the main earthquake on April 8. There was no map generated for the aftershock on April 9since it was considered similar enough to the main earthquake’s profile.

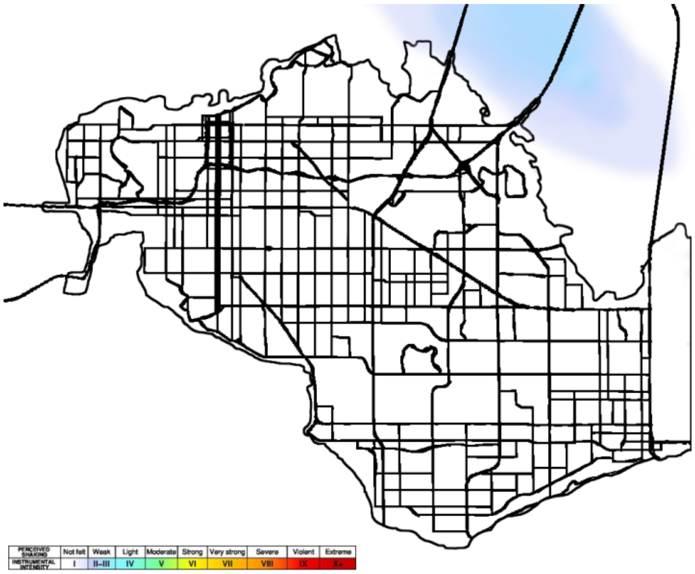



*Figure 3. Shakemap of 2.3 pre-quake on April 6.*

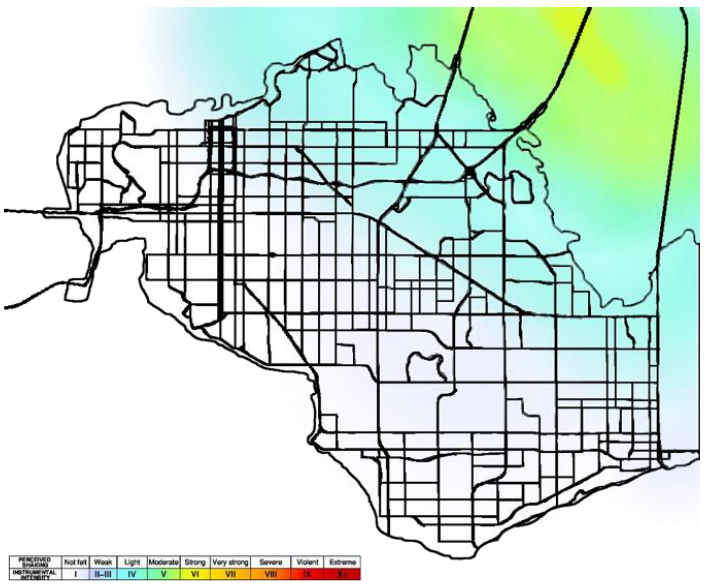



*Figure 4. Shakemap of 6.7 earthquake on April 8*

Total number of reports collected by the app over time gives a good overview of how much absolute and relative reporting happened, both between events and across neighborhoods. Figure 5 shows generally that each neighborhood reports each of the three major events (pre-quake, main quake, and aftershock, respectively in time), and that the shape of the curve is that of a gamma distribution nearing an exponential (i.e. a sharp increase as the shaking begins followed by a slower decay in overall reports).

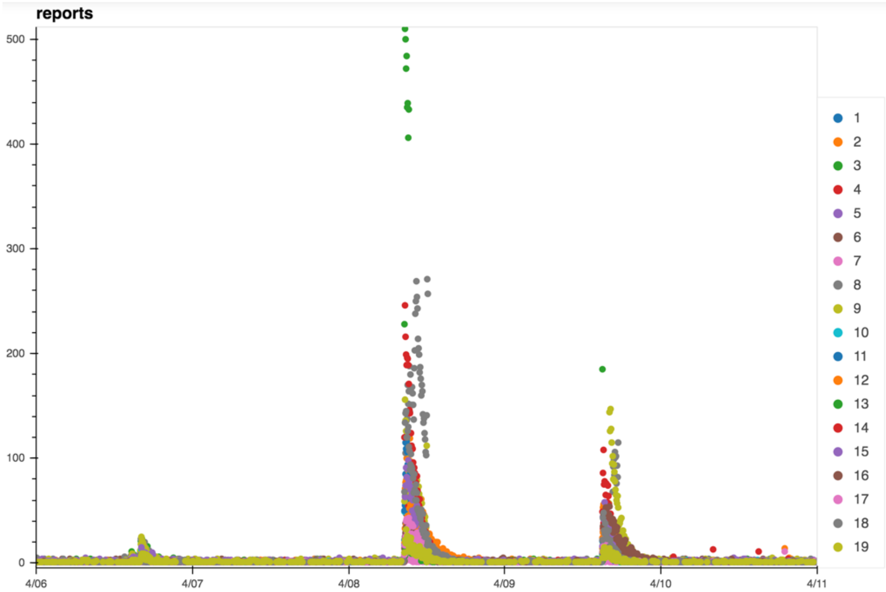



*Figure 5. Total number of reports by neighborhood*

Aside from the main events, there are a number of minor events that are visible when zooming in and/or separating these total reporting plots. For instance, Figure 6 and Figure 7 show that while Palace Hills (Neighborhood 1), Northwest (Neighborhood 2), and Safe Town (Neighborhood 4) seem to have constant reporting (mostly noise outside of the major events), Old Town (Neighborhood 3) has some windows of time where all reporting is absent. This is an indicator that the power went out. The large spikes in reporting just after such a time lapse is due to a majority of the backlogged reports being processed all at once by the Rumble App server once power is restored. (We made some simplifying assumptions like that each neighborhood has a cell tower that loses its ability to send data when the corresponding neighborhood loses power, and that a neighborhood is said to have lost power when in actuality a large portion may have only lost power.)

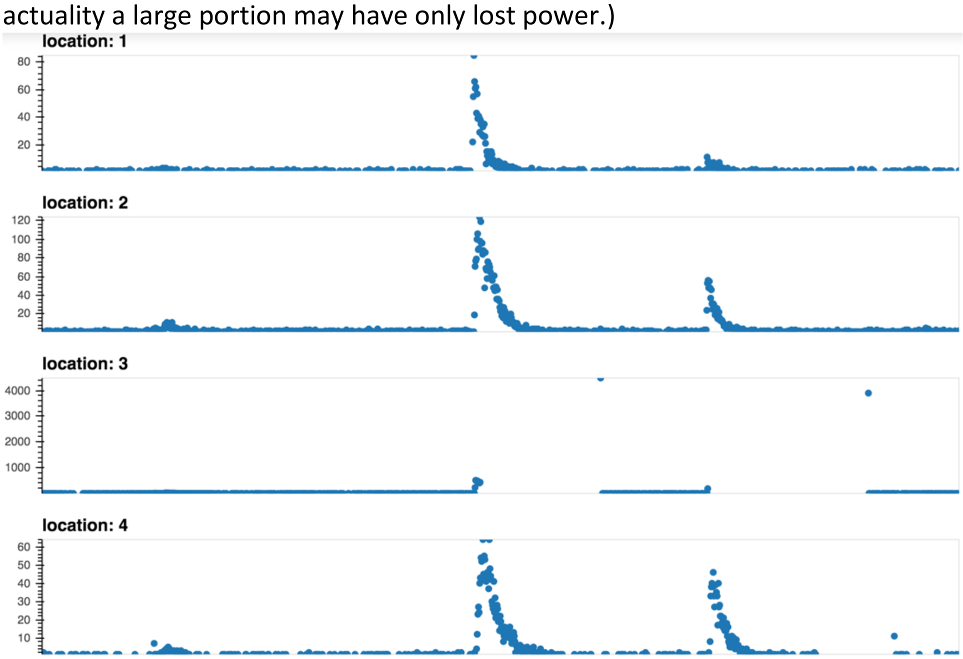



*Figure 6. Total reporting in Palace Hills (1), Northwest (2), Old Town (3), and Safe Town (4)*


*Figure 7. Total reporting in Safe Town (4), zoomed into the y-axis*

Similarly, note that Scenic Vista (Neighborhood 8), Broadview (Neighborhood 9), Capparal (Neighborhood 10), and Terrapin Springs (Neighborhood 11) have missing data for various time windows (see Figure 10 for each zoomed into their y-axes). In contrast, while Old Town (Neighborhood 3) had an initial spike in reports just after the main earthquake and then a window of missing data, Scenic Vista (Neighborhood 8) and Broadview (Neighborhood 9) have a second spike as the initial spike decays and only then is there a window of missing data. This indicates that the power outage was most likely due to a secondary event (and indeed both neighborhoods experience a landslide/liquefaction shortly after the main quake hits). The remaining neighborhoods have a few similar signatures, where either a window of data is missing, a secondary spike occurs after a major event, or both. None of these plots – especially the ones without any of these signatures – really capture the reported magnitude of the events, or if a switch in reported magnitude has occurred (which would further hint at a secondary event).

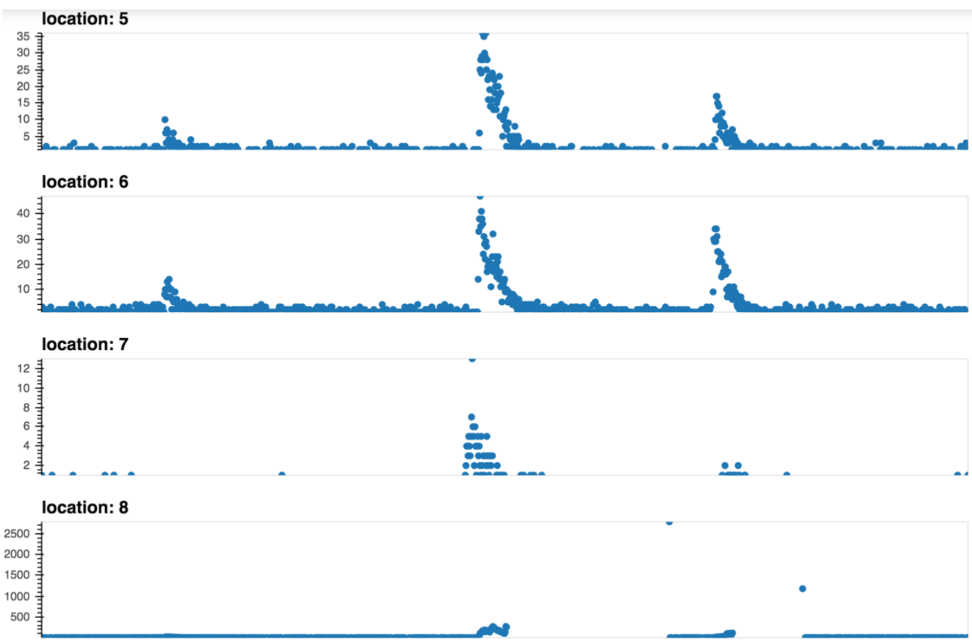



*Figure 8. Total reports in Southwest (5), Downtown (6), Wilson Forest (7), and Scenic Vista (8)*

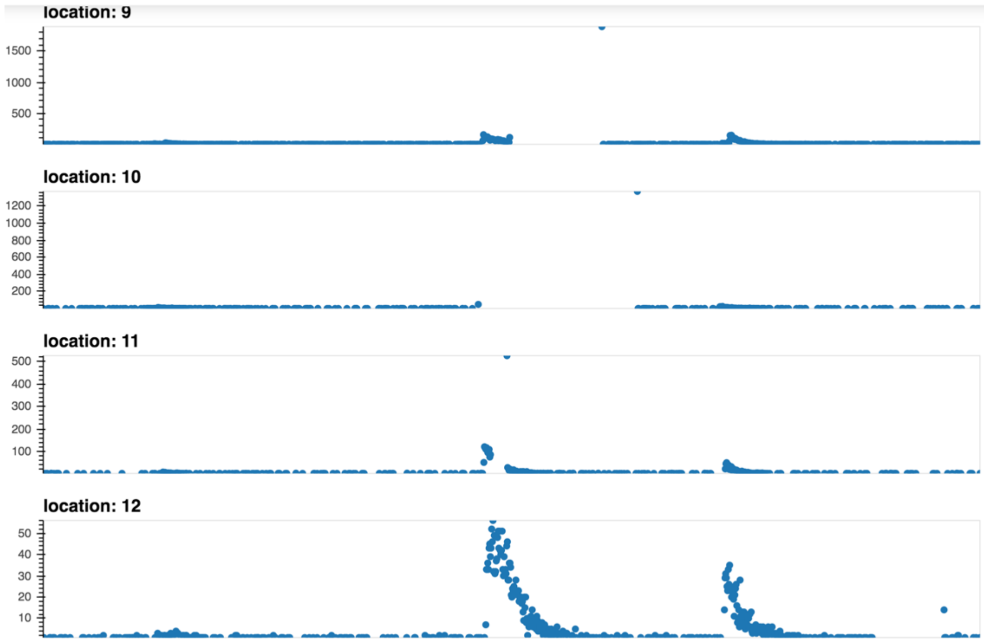



*Figure 9. Total reports in Broadview (9), Chapparal (10), Terrapin Springs (11), and Pepper Mill (12)*

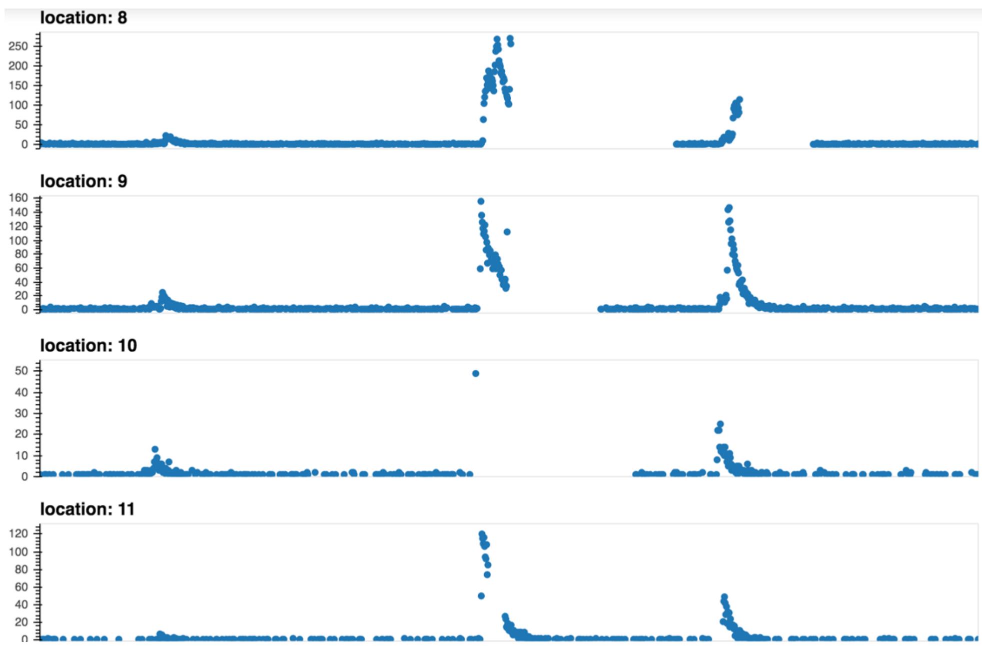
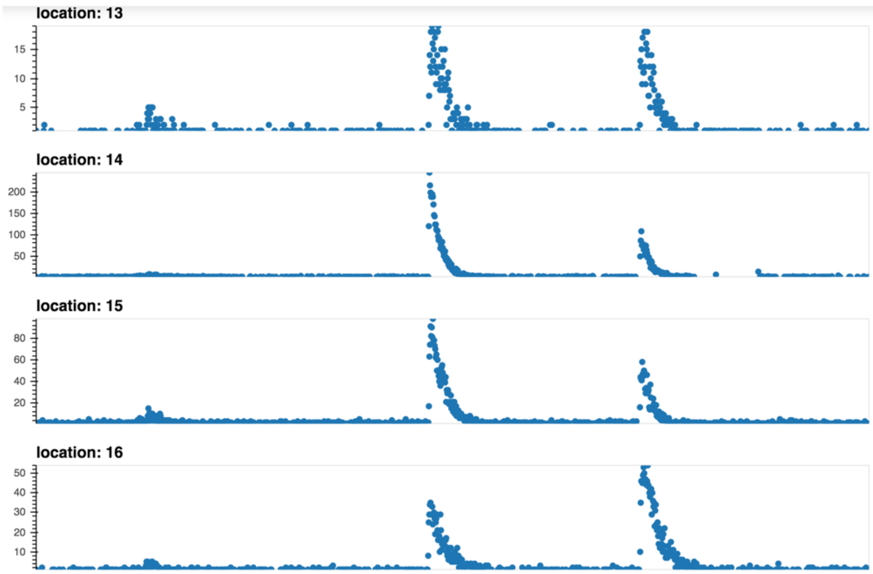



*Figure 10. Total reports in Scenic Vista (Neighborhood 8), Broadview (Neighborhood 9), Chapparal (Neighborhood 10), and Terrapin Springs (Neighborhood 11), zoomed into the respective y-axes*




When breaking down total reporting within a neighborhood by infrastructure type (as well as shake intensity), then we start to see what has been damaged, where, and when.  Figure 11 shows that Downtown (Neighborhood 6) has reported values no higher than 3 most of the time, while Figure 12 shows that Terrapin Springs (Neighborhood 11) has its power most greatly affected during the main quake (which dwarfs the other events' reported values).

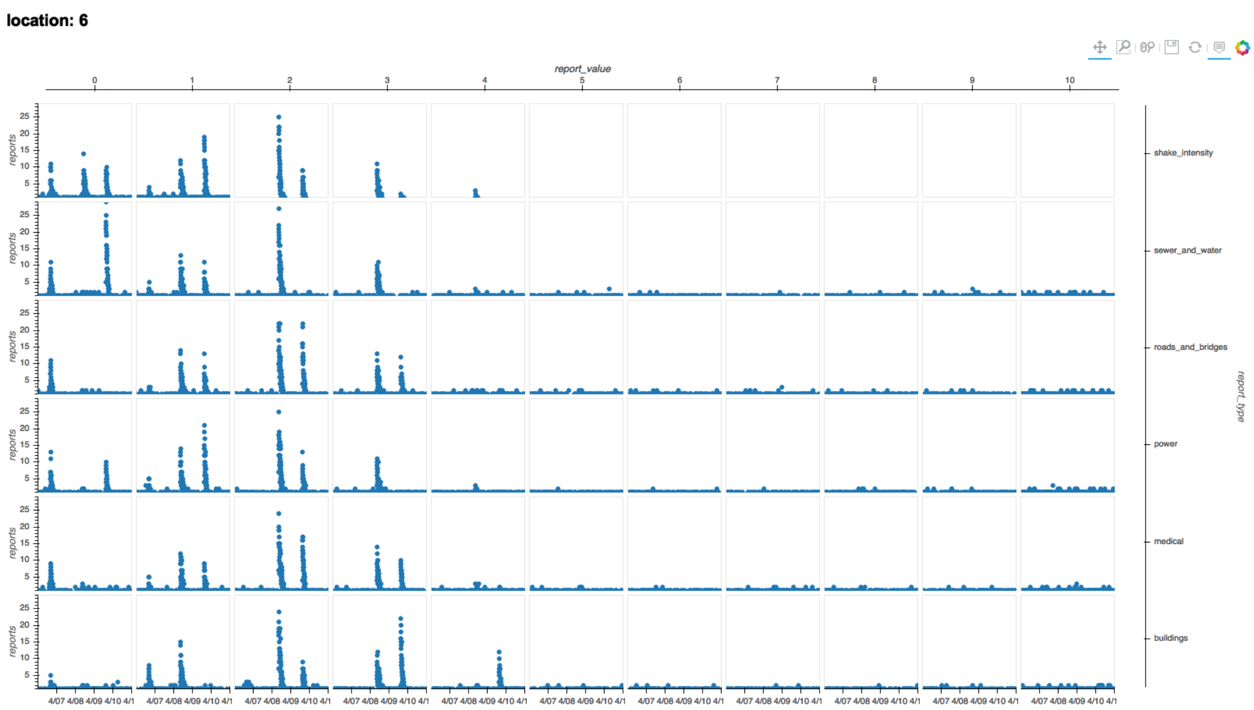



*Figure 11. Total reports by infrastructure type (and shake intensity) for Downtown (neighborhood 6)*

**location: 11**

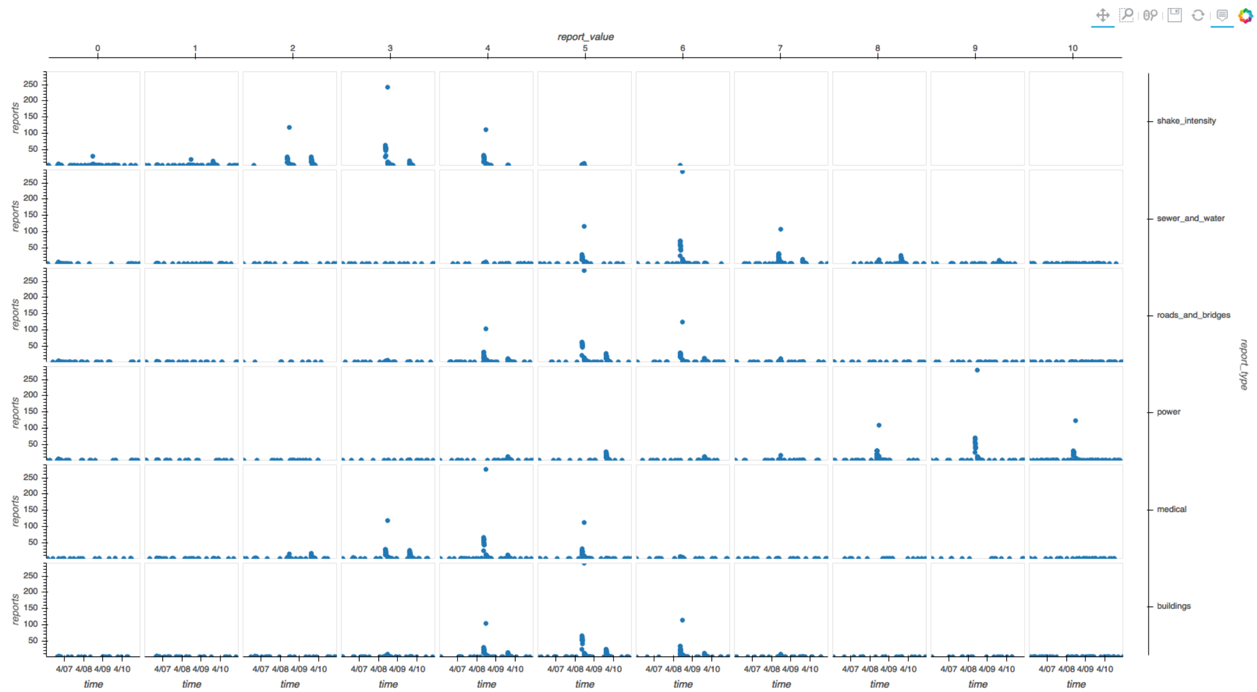



*Figure 12. Total reports by infrastructure type (and shake intensity) for Terrapin Springs (neighborhood 11)*

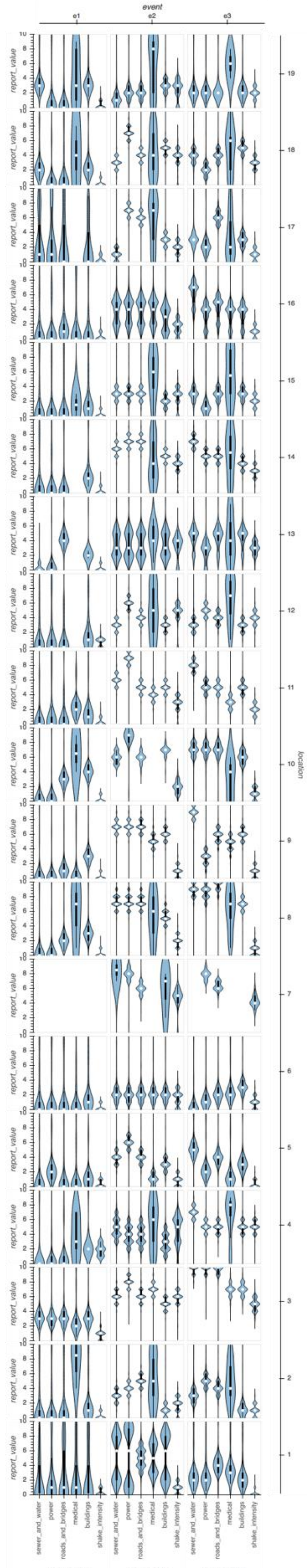


*Figure 13. Violin plots of each events damage values for each neighborhood. E1 is for the initial mini-quake on April 6, E2 is the large quake on April 8, and E3 is the aftershock on April 9.*

Figure 13 show the same thing but in terms of distributions separated further by the three main events (plus background, which is everything else outside of event start times and some duration thereafter). Each report type (such as sewer and water status) for a given neighborhood typically has a commonly reported value in what like a normal distribution, if the data were continuously interpolated.

The anomalies that can be seen are those in Cheddarford (Neighborhood 13) and Southton (Neighborhood 16), which show bimodal distributions in the main earthquake (e2); this is due to inflating report values to get help faster since they feel neglected.  On the other hand, the extreme, nearly symmetric about the midpoint of our 0-10 scale, bimodal distribution found in Palace Hills' (Neighborhood 1's) e2 (April 8 major earthquake) is odd; this is due to someone flipping their reported values and constantly reporting these erroneous numbers as he/she drives around the city trying to be a good Samaritan.

Another interesting phenomenon is the slightly wider distributions, like in Safe Town's (Neighborhood 4's) case; they experience a secondary event which does not cause the power to go completely out (and therefore reporting is seemingly constant), thus the increase in damage causes increased report values to around a new mean and shows up in this bimodal fashion.  Similarly, Scenic Vista (Neighborhood 8) had a slightly wider distribution, and this becomes more apparent when looking at the time series heatmap in Figure 14, where after noon on the day of the earthquake, just before power goes out, the mean value of reports jumps by 1 point or so across 4 of the infrastructure types.

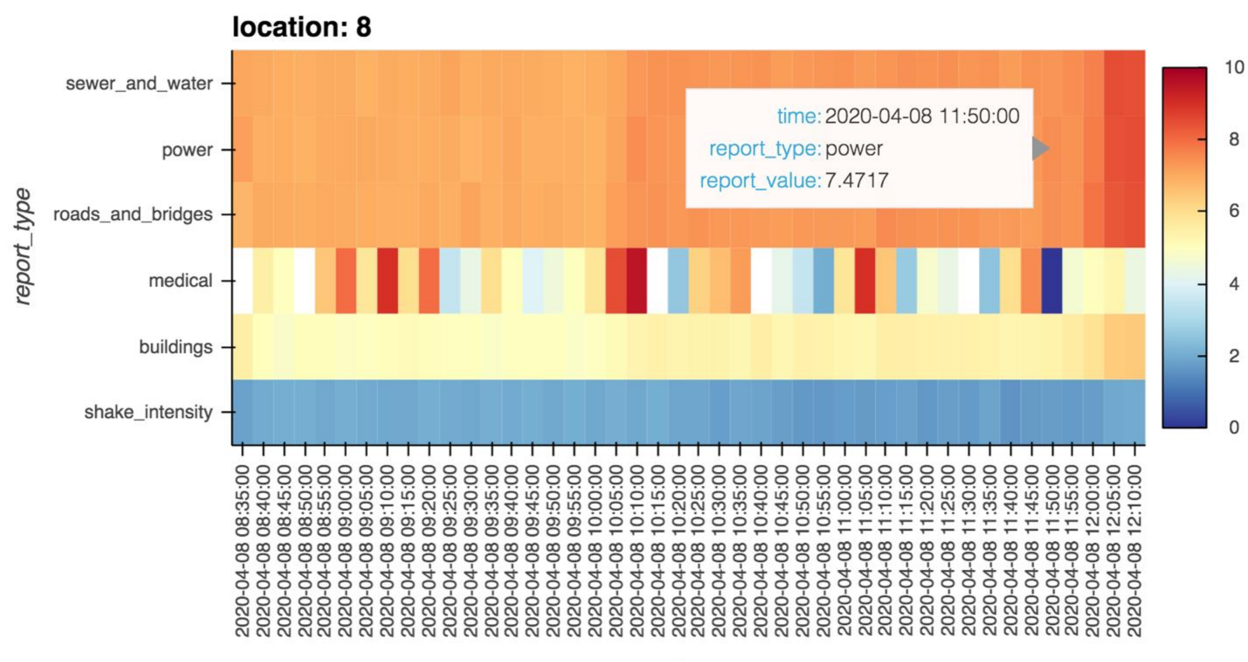



*Figure 14. Timeseries heatmap of average report values for Scenic Vista (neighborhood 8)*

Heatmaps show how each neighborhood's infrastructure damage compares to the others, by event. Figure 15  shows that (in somewhat decreasing order) Terrapin Springs (Neighborhood 11), Chapparal (Neighborhood 10), Old Town (Neighborhood 3), and Wilson Forest (Neighborhood 7), have the highest reported Power damage values, while Wilson Forest (Neighborhood 7) appears to be suffering from sewer and water damage as well.  Similarly, Figure 16 shows reports for the aftershock event, from which we see that (in somewhat decreasing order) Old Town (Neighborhood 3), Scenic Vista (Neighborhood 8), Broadview (Neighborhood 9), Wilson Forest (Neighborhood 7), and Terrapin Springs (Neighborhood 11) are hit hardest in various infrastructures. On the other hand, Wilson Forest (Neighborhood 7) is sparsely populated so the noise may outweigh the actual damage reports (yellow is midpoint of 0-10 scale, while white indicates no reports for that category either due to not having, e.g. medical facilities, or no one reported due to the random generator never being called enough times given a small population).

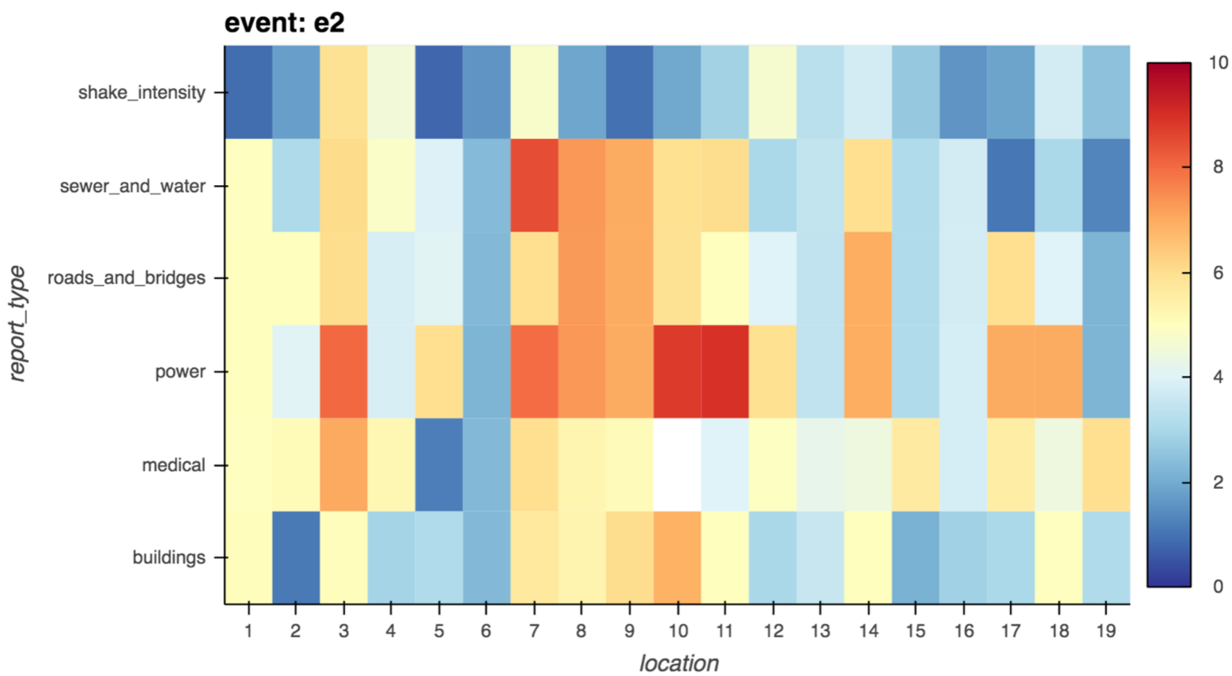



*Figure 15. Main earthquake event (e2) heatmap of infrastructure damage report values*

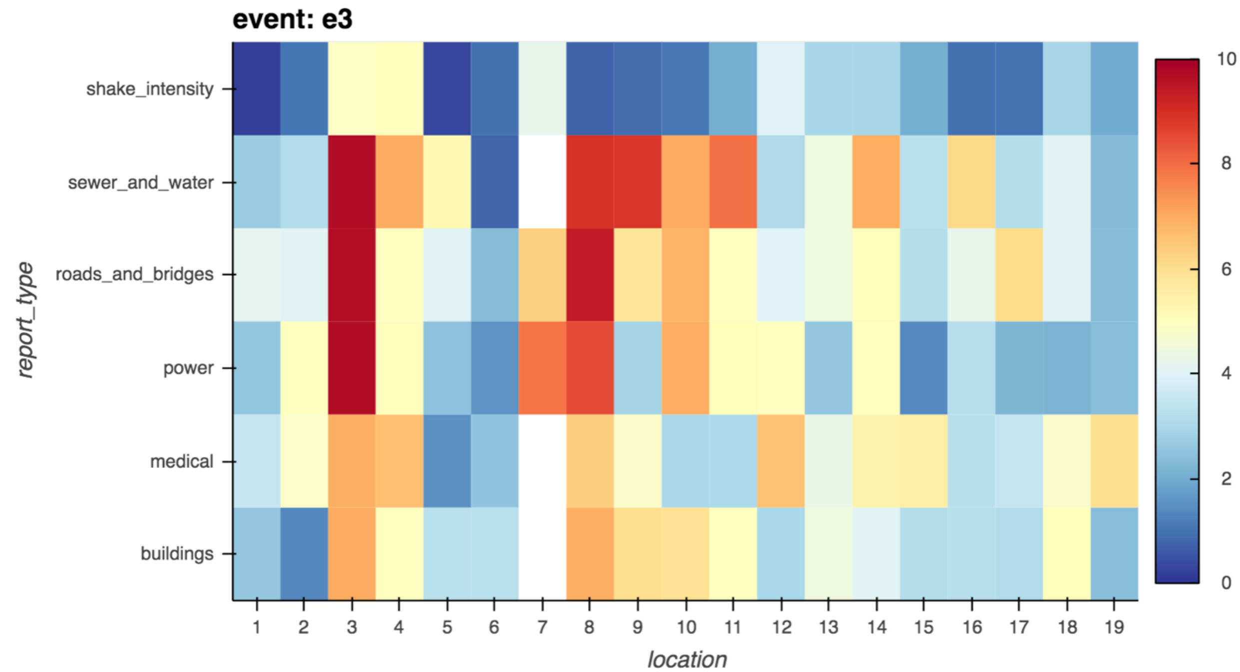



*Figure 16. Aftershock event (e3) heatmap of infrastructure damage values*

## Questions for Participants

The following are the questions they were asked to respond to. Following each question is information to support your evaluation of their answers.

As you review the submissions, keep in mind that we are looking for submissions that go above and beyond rudimentary visualization techniques and shallow analyses. We are looking for visual analytics approaches that provide context for the data, support interactive exploration to identify and understand patterns. We encourage creativity. We are also particularly interested in representations of uncertainty in the visualization. Analyses should provide both answers and evidence or rationales, and should demonstrate critical thinking about the data. It is even better if assumptions, multiple competing hypotheses are documented as well.

As you develop your review of the submission, please also consider the following questions.

- If there were hypotheses that remained unresolved, did the submission specify actions to be taken to resolve them?
- Did the team develop an innovative visual analytic tool? Alternatively, did they use an existing tool in an innovative way?
- Did visualizations enable the analysis process? Or did they merely illustrate conclusions? Did the submission rely more heavily on non-visual analytic approaches?
- Did their tool allow useful interactions?
- Did they use all the available data?
- Was the submission clear?

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?***

   As mentioned above, a handful of neighborhoods are hit hard in various infrastructure types. Participants may would prioritize by restoring power first so that more reports can come in as events unfold, followed by those affected by water/sewer damage (since that appears to be hit next hardest). The plots above combined with the timeline show exactly who gets hit when, and how hard (the mean plotted in the heatmaps or the major "hump" in the violin plots is a good indicator of the actual damage in most cases).

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.***

   As mentioned above, the uncertainty unfolds in various ways, from missing reports, to spikes in total reports before/after the missing data periods, to bimodal distributions in the report values.

3. ***How do conditions change over time? How does uncertainty in change over time? Describe the key changes you see.***

The conditions worsen for some neighborhoods, which may indicate the lack of help sent before another shake event happens, while others improve, hinting that damages were fixed and new damages brought the new levels down to something still above what previous damage had brought.

The uncertainty changes from one event to the next in a few neighborhoods, i.e. neighborhood 1 has one resident flipping the scale of reporting but correct his/her mistake during the next event (or just not reporting). Cheddarford (Neighborhood 13) and Southton (Neighborhood 16) inflate reported values in the main earthquake event (e2) but stop during the aftershock event (e3). Mean report values jump up over time indicating further damage; such a hypothesis is strengthened when power goes out shortly thereafter (especially if report values for such are within those values increased).