

VAST Challenge 2019

Mini Challenge 1

Visualización de la Información

Pablo Alejandro Costesich - 50109

Introducción

El Visual Analytics Science and Technology (VAST) Challenge es una competencia anual en el área de analítica visual. Consta de una serie de desafíos que deben ser resueltos a lo largo de una temática central.

El objetivo de este año fue sobre los daños de un desastre natural en St. Himark, un lugar ficticio. Se presentaron tres mini-desafíos y un desafío central. El objetivo de este trabajo es analizar el primero.

El análisis apuntó a brindar información para que los servicios de emergencia puedan priorizar su actuar y entender los principales focos de interés.

Las preguntas guía planteadas fueron las siguientes:

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.
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.
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.

Los datos se encuentran disponibles en los siguientes links:

- [Enunciado MC1](#)
- [Enunciado MC2](#)

Análisis de los datos

Preanálisis

Primero se analizó la data disponible en los dos archivos descargables del VAST Challenge, para los tres mini-desafíos. El objetivo fue encontrar data útil entre los tres.

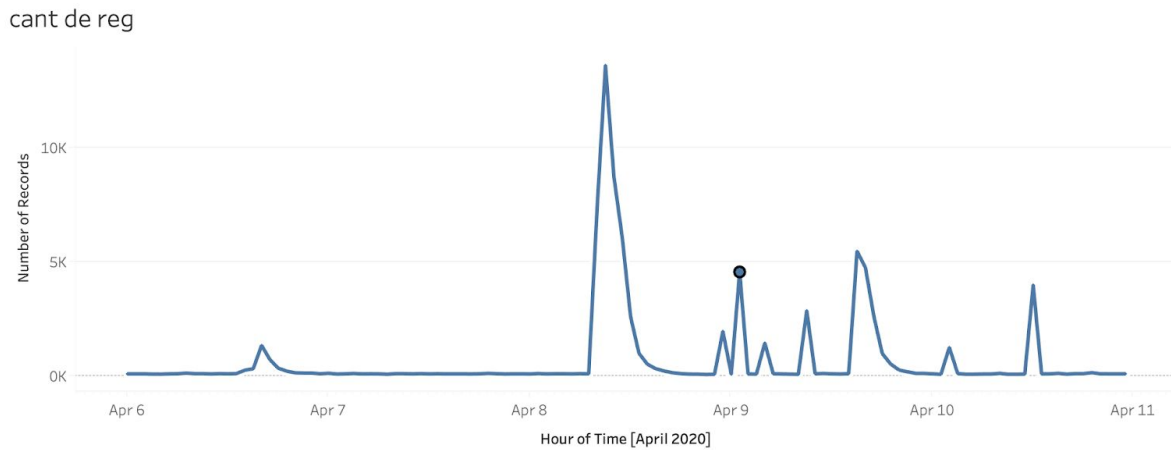
El mini-desafío 1 posee un archivo mc1-reports-data.csv con información sobre reportes de los daños causados a los edificios, infraestructura de caminos, desagües, intensidad de movimiento y otros. En el MC2 se encontró un shapefile con la definición de los barrios correspondientes.

Los valores de intensidad pueden variar de 0 a 10, al igual que los reportes de daños.

Pre-procesamiento

Ambos archivos mencionados anteriormente son formatos comunes, por lo que se optó utilizar Tableau para procesarlos. Se vinculó mediante el id de barrio ambas fuentes de datos, lo que nos proveyó de los nombres de las zonas geográficas y visualizaciones del mapa.

Distribución de los datos



Para entender correctamente con qué datos nos encontramos, se plotó en gráficos de líneas la cantidad de registros vs el momento temporal en el que fueron reportados. Similarmente, se plotearon esos mismos registros como un promedio para cada campo de reporte de daños (shake intensity, medical, entre otros) vs su momento temporal.

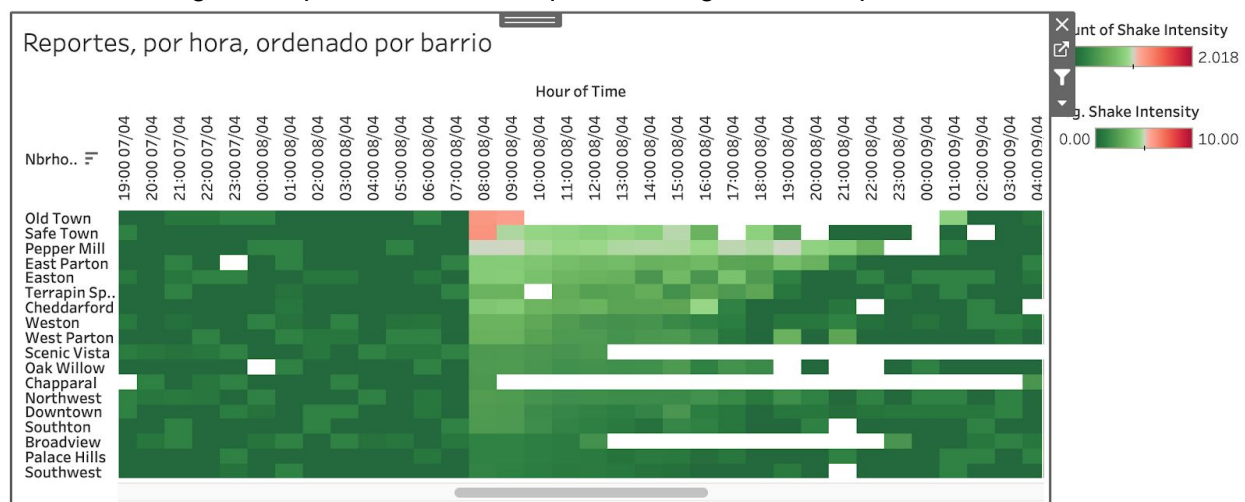
Podemos ver que existen tres picos muy definidos, y picos menores. El primero, ubicado entre el 6 y 7 de Abril, puede considerarse ruido: existe mayor cantidad de reportes, pero no vemos cambios en la intensidad de movimiento.

Luego se detectan una serie de picos menores el 9 de Abril, que en el análisis resultaron ser poco confiables, hasta una posterior réplica el mismo día a las 15hs. Por último, el 11 de Abril a las 12hs se detectó el tercer pico de interés.

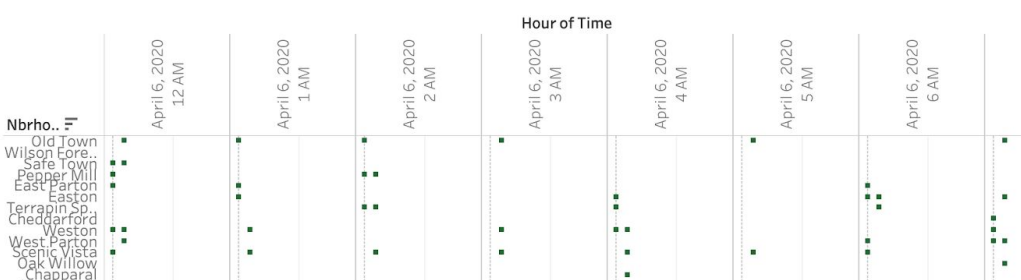
Pregunta 1 - Priorización

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.

Se optó por un mosaico que pudiera ser ordenado por la intensidad global de daños, junto a un plot que diferenciara la distribución de cada uno de esos reportes. En el primer gráfico el color indica el daño global reportado, mientras que en el segundo es el promedio.



Distribución interna por barrio

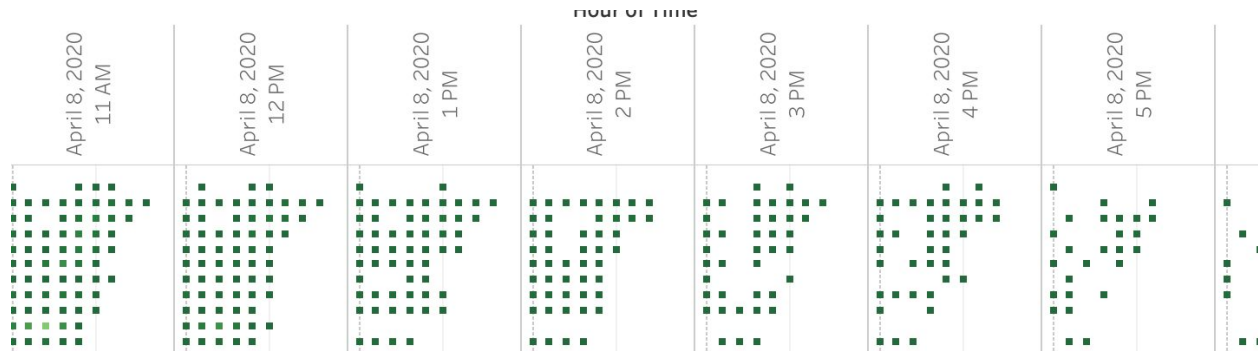


El mismo tiene un doble conjunto de filtros que permite priorizar los barrios por la suma de los daños reportados y un enlace con el gráfico inferior, que nos permite seleccionar en ambos un set específico de datos para hacer drill-down.

Primero se ordenó por daños globales. Luego, analizando para los casilleros con colores particulares (outliers) se analizó la distribución de los datos. Se encontró que en algunos reportes existían distribuciones multimodales, correspondientes con data atrasada de reportes, por lo que fueron filtrados.

El valor de este gráfico para los servicios de emergencia se encuentra en:

1. Decidir qué barrio reportó mayor movimiento
2. Entender si ese movimiento se tradujo en daños estructurales de importancia.



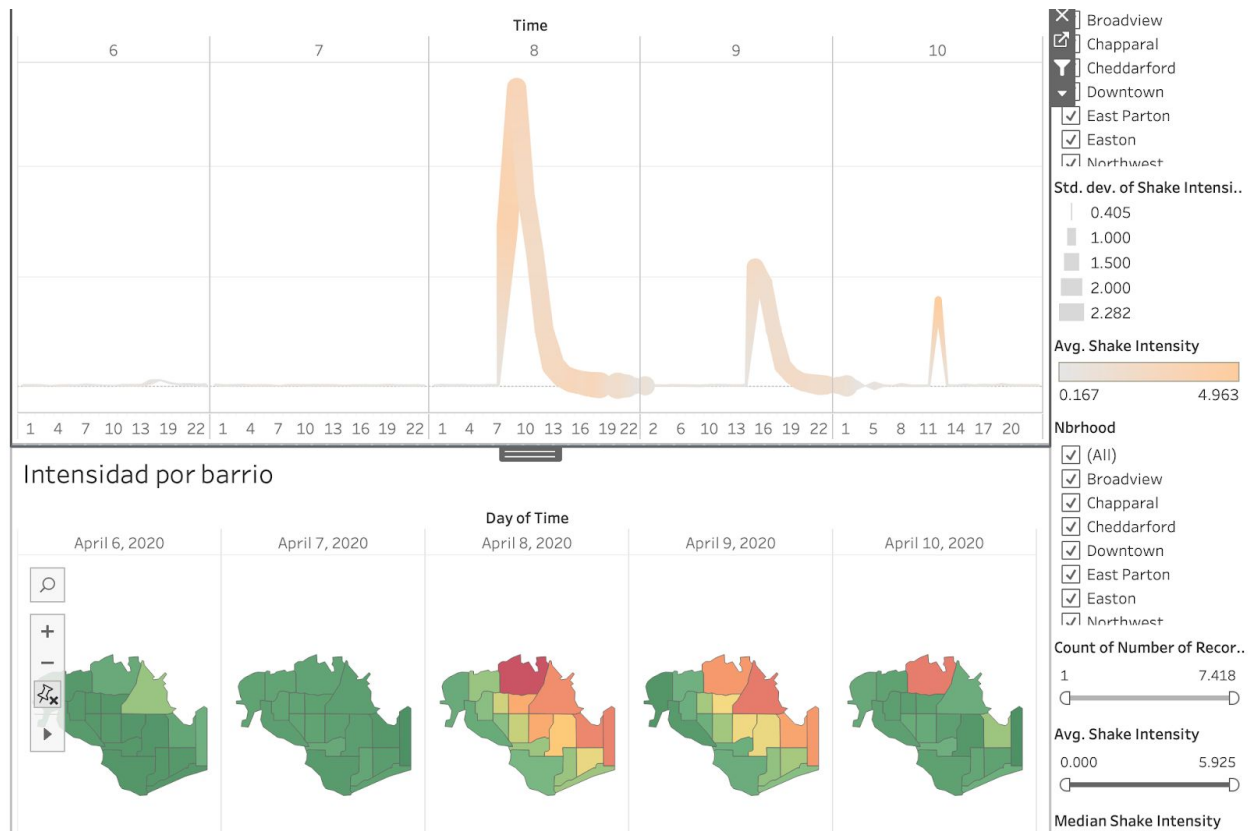
Se pueden apreciar distribuciones bimodales en ciertos tipos de reporte

Pregunta 2 - Incertidumbre

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.

Para responder a esta pregunta, se planteó un dashboard que nos permitiera separar visualmente la incertidumbre geográfica como a nivel reporte. Primero se partió por plotear los reportes con respecto a su distribución estadística en función del tiempo, es decir, su valor medio vs el momento. El tamaño de los círculos está dado por el desvío estándar: a mayor desvío, mayor es su tamaño, generando un “halo de error” dentro del plot.

Con esta visualización se pudo filtrar una serie de datos muy poco confiables, correspondientes a los picos de ruido explicados con anterioridad.



El gráfico tiene ya filtrados los outliers del preanálisis

En la parte inferior podemos ver la distribución del movimiento reportado por día, por barrio. Esto nos da una idea de qué barrios se vieron más afectados y si fue un suceso generalizado o puntual a ciertas áreas.

Se pudo ver en la curva (en la versión interactiva del workbook) que Old Town tiene picos ascendentes marcados, pero su data desaparece para la caída. Se estimó que es por la infraestructura vieja del pueblo. No es el caso con Safe Town, que tiene curvas mucho más suaves.

Pregunta 3 - Cambio Temporal

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.

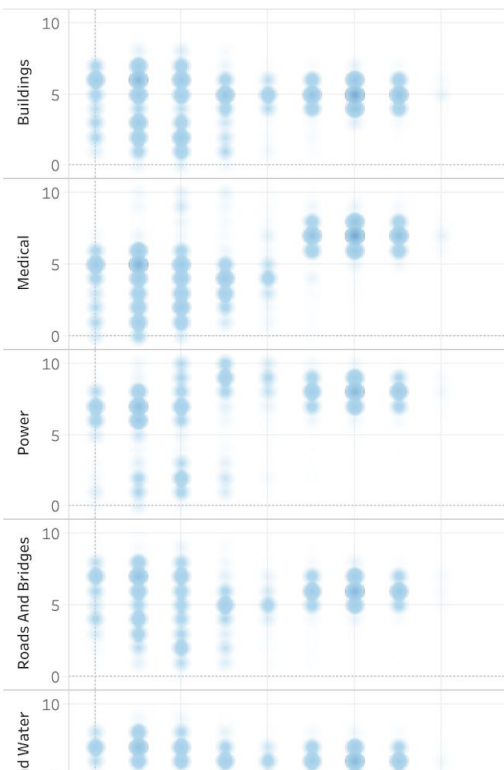
Para este análisis se estableció un tercer dashboard en Tableau con una vista de flow. En los gráficos anteriores es posible indagar sobre las condiciones particulares de la infraestructura y servicios, en particular usando el primer dashboard. Sin embargo, se decidió encarar una vista de flow de estos datos. Para ello, se dividió en páginas la visualización para que fuera mostrado como una animación interactiva.

Se separó verticalmente en dos zonas: la izquierda, con un heatmap por cada tipo de reporte (medical, sewer and water, bridges and roads, y power) comparando el shake intensity vs los daños reportados.

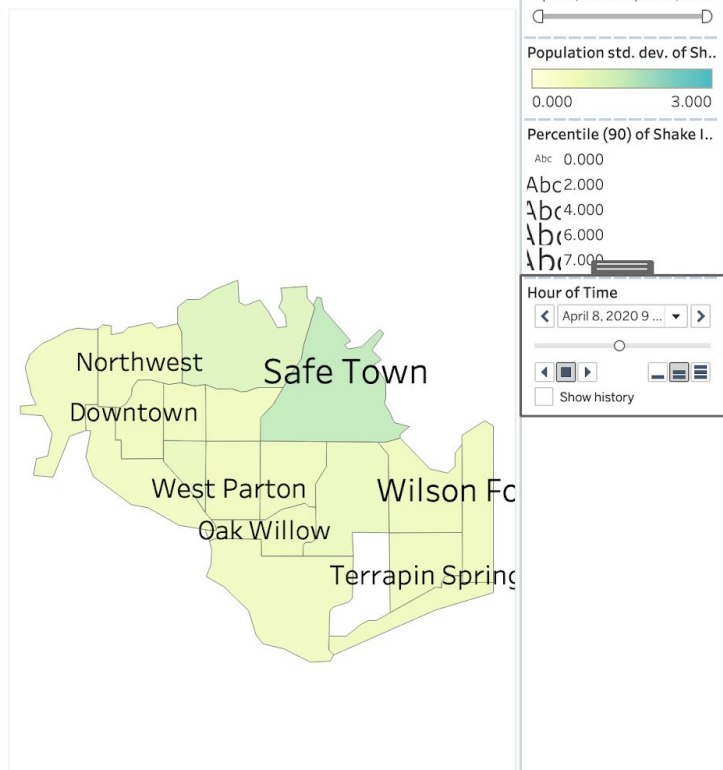
Para el lado derecho, se planteó una vista de mapa con las siguientes salvedades:

- El tamaño de las etiquetas en los barrios es su shake intensity promedio
- El color, con una escala de amarillo a celeste, indica en orden creciente la incertidumbre de los datos.

Dispersión del shake intensity vs daños



Atención a barrios- April 8, 2020 9 AM



Este gráfico hace evidente la confiabilidad de los datos para cada categoría y su evolución temporal como animación. Es un sintetizado de lo que verían los servicios de respuesta en tiempo real.

Bonus Track

The data for the individual mini-challenges can be analyzed either as a static collection or as a dynamic stream of data, as it would occur in a real emergency. Were you able to bring analysis on multiple data streams together for the grand challenge using the same analytic environment?

Describe how having the data together in one environment, or not, affected your analysis of the grand challenge. Limit your response to 500 words and 10 images.

Los datos fueron analizados principalmente como colecciones estáticas de datos. Sólo para el último punto se planteó una vista de flujo.