DV PROJECT REPORT - 2021
2019 VAST Mini-Challenge 1: Crowdsourcing for Situational Awareness

Team members:

Hemanth Kumar Sinde - 1219489830 - hsinde@asu.edu
Sai Teja Tatikonda - 1222715819 - statiko1@asu.edu
Sandeep Somu - 1220154260 - ssomu1@asu.edu
Saad Anwer - 1222307476 - sanwer@asu.edu
Dhrithi Chidananda - 1219433826 - dchidana@asu.edu
Saumya Gangwar - 1219377939 - sgangwa4@asu.edu

1. INTRODUCTION

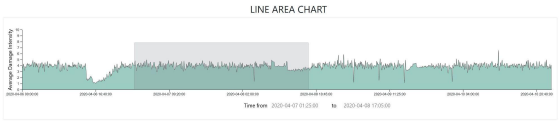
This paper proposes a visualization tool that can be used to answer the three tasks posed in the 2019 VAST Mini-Challenge 1: Crowdsourcing for Situational Awareness. The reports from the residents were collected, examined, and depicted in an interactive visualization. This application may be used to identify the regions that are the most affected so that emergency assistance can be given there first. There are five visualizations in the tool with one interactive and innovative. A Geo-Spatial chart showing all the neighborhoods based on damage intensities over a selected time period, a line area plot showing the damage intensities over a period of time, a Scatter plot displaying top 4 cities with damage intensities for a selected dimension over time, a Sunburst chart which depicts top 5 reported cities based on number of reports for a particular dimension and a Grid-map displaying the uncertainty of all regions over a selected time period and when clicked on a particular neighbourhood a histogram portrays variance of all individual dimensions.

2. VISUALIZATION DESIGNS

There are five visualizations to depict the data in a productive manner. The five visualizations are listed below.

2.1: Line Area chart

We have implemented this line area chart to show how conditions change over time. So we took a timeline of all data('time' column from the dataset) and plotted it over x axis and average densities of all dimensions over y axis. X-axis starts from '06-04-2020 00:00'' and ends by '11-04-2020 00:00' whereas the y axis has a range between 0 and 10. Users have a privilege to select the timeframe that he/she requires for insights.



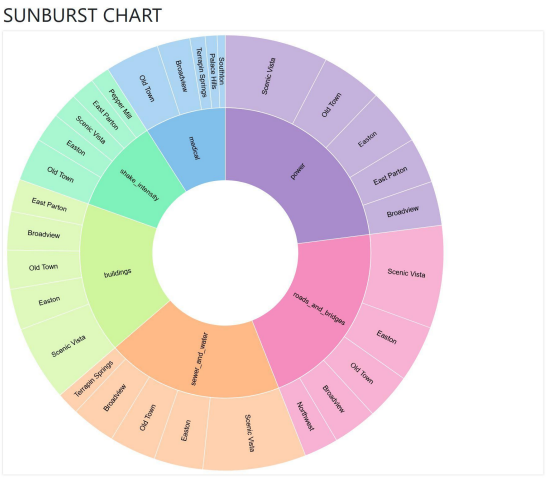
Additionally, this line area chart is linked to other visualizations like sunburst, geospatial and innovative grid map. In this way we can manage all the charts over a particular time that the user has chosen.

The following are the marks and channels for the above visualization:
Marks: Line, area
Channels: Position, color hue

2.2: Sunburst chart (INTERACTIVE)

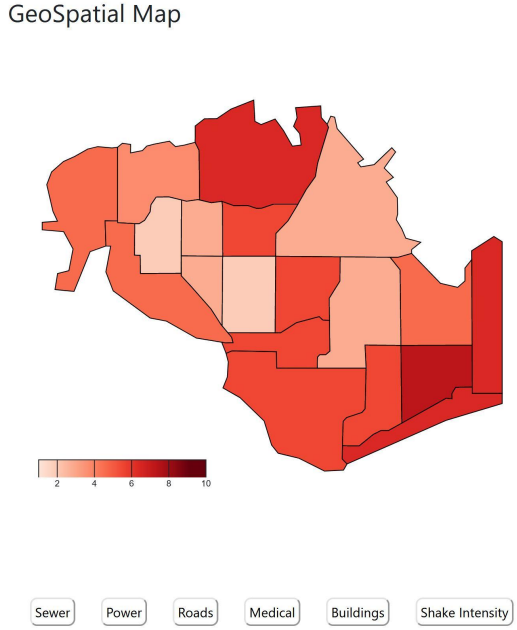
We developed a Sunburst chart to show the top five cities with the highest number of reports for a specific dimension over a selected time period. The Sunburst is divided into two levels. Shake intensity, Medical, Power, Buildings, Sewer and Water, Power, Roads and Bridges are the six dimensions of the inner level. The top five cities with the highest number of reports are located on the outer level. When we hover

over the particular dimension, it will show the sum of reports of top 5 cities.



When we click on any of the dimensions, the chart collapses into a single level that displays the top 5 cities for that dimension. To return to the previous level, click in the center of the chart.
Marks: Area
Channels: Color hue

2.3: Geospatial Map



The geospatial map was developed to show the average intensity of a specific dimension over time chosen. The geospatial map is affected by two parameters, one being the dimension that is Sewer, Power, Roads, Buildings, Medical and Shake Intensity and the other being time. The dimensions each have a button on select of which the color luminance that to light to dark representation. The lighter the color shade the least affected is the city over time. When the timeframe is extended/ selected/ decreased over a particular time in the Line Area chart the geospatial map adapts the visualization accordingly showing the affected cities over the chosen timeframe.

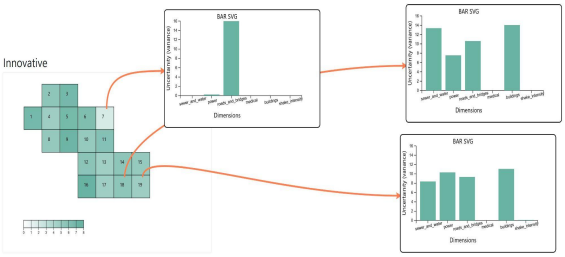
In addition, on hover over the geospatial map it displays the neighbourhood name and intensity rate from 1-10 which was reported.
Marks:Area
Channels: Color luminance and Color hue

2.4: Scatterplot

The scatterplot was developed to show the average intensity over a particular time period on the time slider. It shows the intensity of a particular dimension at a particular time on the time slider for top 4 cities. The plot has intensity on the y-axis and time on the x-axis .For the period selected on the time slider the dimension is selected. There is a button for each dimension Sewer, Power, Roads, Buildings, Medical and Shake Intensity. Average intensity of that particular dimension for that particular time period selected is calculated for the top 4 cities. The plot for each city is represented by a different color hue.
Marks:Points
Channels:Color hue

2.5: Grid + Bar chart (Innovative)

This visualization is a combination of a Grid Map and a Bar Chart. In this, the Grid map depicts the uncertainty of all cities for a certain time period selected in the Line Area chart. Uncertainty is obtained by calculating the variance of all dimensions. When we hover over the grid, the name of the city and the variance are shown. When we click on any city on the grid map, we get a Bar chart with six dimensions that shows the variance for each dimension on the X-axis.



To generate a bar chart, we may choose up to three cities from the grid map. In order to reduce clutter, we can move the bar chart in any way that is suitable for us. With the time period specified in the Line Area chart, the grid map and bar chart are updated.
Marks: Area
Channels: Color Luminance, Size, Position

3. DESCRIPTION OF VAST MINI CHALLENGE

As the title “VAST Challenge 2019: Disaster at St. Himark!” suggests that there’s a disaster at the fictional location St. Himark, in particular there’s been an earthquake at that location. And because of this there has been a lot of chaos around the city which includes but is not limited to damages.

Health care facilities, Nuclear power plants and so on. There were in total 3 Mini Challenges. Out of which we choose the very first one.

In the Mini Challenge 1 of the VAST Challenge 2019 it states that the city of St. Himark has released an app just before the earthquake so that the citizens can successfully upload real time issues they were facing be it power disruption or sewer blockage. With the help of the data obtained from the citizens combined with the earthquake disruption shake intensities the authorities can prioritize their responses to issues in an effective manner.

To better help the city officials in sending the help of the limited resources to the places in most need, the goal of Mini Challenge 1 of the VAST Challenge 2019 is to characterize nearby cities in order to deliver an immediate reaction based on two metrics: damage and uncertainty. To achieve this task we are equipped with three tasks. One task is to use Visual analytics to better understand how responses are coming through with respect to the earthquake shake map and how to better send responders to the locations of the most need. In the second task there could be responses which might not require immediate response and there could be uncertainties in the data being sent to the authorities. We have to visually compare the reliability of neighbourhoods. And which neighbourhoods are providing reliable data and the uncertainty in the data. And in the third and final task we have to visually show how the conditions change over time and how uncertainties change over time.

4. Dataset Description:

In this Mini challenge the dataset is provided in csv file format in which it has 8 attributes namely Time, Sewer and Water, Power, Roads and Bridges, Medical, Buildings, Shake Intensity, Location. This data is reported by people from the city of St. Himark. The above Data types are quantitative, and ordered. There are about 11 levels for each dimension. We have 83070 records in total. The datatypes of each column is as follows: Ordered data - time, Categorical - sewer & water, power, roads & bridges, medical, shake intensity, buildings, Quantitative - location.

To calculate the damage intensity, we used mean for all dimensions and to find uncertainty we used variance as a measure.

Damage Intensity = $\frac{1}{DR}\sum_{i=1}^{R}\sum_{j=1}^{D} x_{ij}$

Uncertainty = $\frac{1}{DR}\sum_{i=1}^{R}\sum_{j=1}^{D}(x_{ij} - \bar{x}_j)^2$

$$\bar{x}_j = \frac{1}{R}\sum_{i=1}^{R}(x_{ij})$$

Where
D = Number of Dimensions
R = Number of Records in selected time period
x_{ij} = Value of Single dimension
$\bar{x}_j$ = Mean of all values in that dimension

We modified the data to meet the task at hand. We sorted the data by location attribute and calculated the number of attribute items, Mean of each categorical attribute, and Variance of each categorical attribute in relation to timestamp. While working on averages, We assumed and replaced 0 inplace of Null data.

To create the Geospatial map, we used GeoJson Spatial data with the attribute location id and with spatial fields x and y coordinates.

For scatterplot, we fed the sorted data of selected dimension and timeframe which was grouped by in a specific time interval like if the time frame selected was 60 minutes then, the time interval grouping was of 5 minutes.

5. HOW TO USE THE SYSTEM TO ANSWER THE MC'S QUESTIONS

5.1: 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?

• Task: Neighborhoods should be prioritized according to the amount of damage reported. For this task, we outlined the key action-target pairs.
{Present -> Explore -> Summarize, Compare}
• Present - We used the data from the damage reports given by people.
• Explore - We examined the data as the targets i.e., highest damaged cities should be known.
• Summarize, Compare - Summarized damage intensities of all categories, all neighborhoods using GeoSpatial map, Summarized damage intensities of top 4 locations of all categories using scatter plot.
• We used GeoJson spatial data to plot the Geospatial Map

5.2: Use visual analytics to show uncertainty in the data. Compare the reliability of neighborhood reports. Which neighborhoods are providing reliable reports?

• Task: The task here is to derive the uncertainty in data and compare reliability of reports.
{Present -> Explore -> Summarize, Compare}
• Derive - We derive uncertainty in data by calculating variance which depends on data distribution over the given time range.
• Explore- Exploring data as the location and targets(uncertain data locations) are unknown.
• Summarize, Compare- Summarizing data through innovative charts and comparing the uncertainty for various dimensions across various neighbourhoods.
• The uncertainty in data is shown in an innovative chart. It is a grid chart, embedded with a bar chart.It provides the visualization of uncertainty of damage intensity reported. On the click of each neighbourhood city the bar

chart shows the uncertainty rate of all the damage dimensions for the selected neighbourhood. Uncertainty is calculated by taking the variance of mean of intensity. Higher the variance higher is the intensity. Many locations indicate high damage but some of them are not certain. From the innovative chart we can interpret that most cities/ neighbourhoods provide certain data during quake time and only some locations have missing data during quake time.

5.3: How do conditions change over time? How does uncertainty change over time?

Task: Insights on how damages impact over time and uncertainty over time.

{Derive -> Explore -> Trends, Outliers, Features}

Trends, Outliers and Features are summarized below

• Line chart is the center of this problem as it has all the time period over the X-axis.

• The peak in the line area chart shows the time in which average damage intensities are high whereas the crest shows the time in which least damage occurred.

• Users can change the range of time according to his/her need. Based on this time change, remaining visualisations change and show data for that time period.

• On analysing GeoSpatial maps in the given time range for the shake intensities, intensity of the damage in the records will show higher with the locations nearer to earthquakes than the locations farther from it.

• Sunburst provides us the analysis on top 5 locations which are having more reported with single dimension in a particular time period.

• Scatter plot is enabling us to observe the damage changing over time in the top 4 locations.

• Innovative chart give us insights on which location has more uncertainty. Further clicking on the single location provides us the bar chart which gives uncertainty over different dimensions for that particular location.

6. DISCUSSION

Initially it looked interesting to pick this MC - 1 because of the issues the VAST MC - 1 stresses itself upon. Later discussing among ourselves and self analyzing our strengths and weaknesses we realized this is the one we want to do out of the pool of Challenges.

6.1 LESSONS LEARNED

We have learnt a lot from this course from the very beginning like How to effectively represent the data to the audience.

• The first and utmost important thing we have learnt is to properly investigate the domain and what their purpose is from the Visualization. Everything revolves around this central theme according to us.

• With the discussions and the late nights we had among ourselves we concluded to use Line Area chart, Sunburst chart (Interactive), Geospatial, scatterplot, and grid + bar chart (Innovative)

• We have effectively learnt the technical aspects of the Visualization like implementing Javascript, D3 JS, and flask

• We have also realized that D3 is more robust compared to the python data wrangling.

• Not only the technical side we have also learnt about the collaborativeness and teamwork skills which are of utmost importance in the present days.

• Even though we proposed a ribbon chart, going forward we realised scatter plot is more intuitive for top 4 cities data.

• We were able to effectively learn and use GitHub commands for various activities including merging individual codes.

• Because of the challenges, we were able to explore and code complex visualisations like time slider, sunburst, arrow pointing to bar chart, draggable feature in innovative bar chart.

6.2 SYSTEM FURTHER IMPROVED IN FUTURE

• Few of the data wrangling has been done in python flask. This can be done with complete code developing in D3 js itself.
• Scatterplot can be more interactive such as hovering over one particular location, we could have intensified it's color while lowering the other plots improving more focus on the data on which the user is hovering.
• Data is taken from human interaction which sometimes leads to uncertainties. To further improve the system, we can install some metric tools which can give accurate earthquakes damage reports.
• Here we are limiting these visualisations to St. Himark city. We can extend this feature to countries and also worldwide.
• Because of the limited scope, we limited charts to one interactive and one innovative which can be further extended to all interactive and innovative.

7. ACKNOWLEDGMENTS

This course is quite different from all of the other ones we had in the field of computer science. With having little to no experience in the field of Visualization and the concepts surrounding it like D3 and Javascript, we can confidently say that we have risen from initial to Intermediate level in this field. And all of this couldn’t have been possible without the proper guidance and mentoring from our beloved professor Dr. Chris Bryan. We would like to sincerely thank him for correcting us when we made mistakes and properly guiding us all the way through and his state of the art lectures about the current times made us aware of the current skills in this field.

8. REFERENCES

Scatter plot:
https://www.d3-graph-gallery.com/graph/scatter_basic.html
Sunburst chart:
https://bl.ocks.org/vasturiano/12da9071095fbd4df434e60d52d2d58d
Grid Map:
https://bl.ocks.org/cagrimmett/07f8c8daea00946b9e704e3efcbd5739
Leader line:
https://github.com/anseki/leader-line
Plain Draggable:
https://github.com/anseki/plain-draggable
Flask with JS:
https://towardsdatascience.com/talking-to-python-from-javascript-flask-and-the-fetch-api-e0ef3573c451
Plotting:
https://bl.ocks.org/curran/3f2ff2e32652397de94d406460e240ce