VAST Challenge 2018 MC2 : Like a Duck to Water

1 INTRODUCTION

This project aims to accomplish the Mini Challenge 2 of VAST Challenge 2018. The project inspects the possible reasons for the environmental damage to the Boonsong Lekagul Wildlife Preserve by answering the questions raised in the Mini Challenge. For this project, we are exploring the hydrology data containing the water sensor readings from rivers and streams in the preserve collected over several years and trying to come up with a visual analytics system. Through this visual analytics system, we will be able to carry out the investigation which had been stalled because of the lack of soil samples.

2 VISUALIZATION DESIGN

The designed system takes the "chemical" to be analyzed as an input from the user and then plots 3 different visualizations corresponding to it i.e. stacked bar charts, stream graphs and scatter plots. Also, in the project we implemented an innovative view which takes into account the location map containing the 10 different locations provided by the mini challenge and then plots a pie chart corresponding to the same. Also, provided in the implementation is a slider along with a play/pause button that helps the user to visualize the change in the concentration of a particular chemical over the years for 10 different locations specified in the maps. Along with the concentration getting plotted on pie chart, there is also a circle whose radius expands or shrinks as the concentration of the chemical changes over the years.

3 DESCRIPTION OF THE VAST MINI CHALLENGE

According to this Mini Challenge, Kasios, a furniture company has been accused of environmental damage to the Boonsong Lekagul Wildlife Preserve for dumping toxic waste and polluting the air with chemicals from its manufacturing process. However, Kasios has been denying these accusations by claiming that they have inspected the area and found it as clean as the rest of the preserve.

To this, the professors of Mistford college from the Ornithology department took to performing soil analysis, but could not come up with any substantial evidence as the top soil was removed because of the recent excavation and building activities.

As the primary evidence against the Kasios company was now gone, the same professors came forward with several years of water sensor readings from rivers and streams in the preserve. These samples were taken from 10 different locations scattered throughout the area and contains measurements of several chemicals. But these samples were never analyzed due to a lack of funding.

Therefore, this project aims to make up for the soil evidence that was destroyed by devising a visual analytics system that can give some useful insights about the contamination at Boonsong Lekagul Wildlife Preserve and finally answer the questions raised in this mini challenge's description (covered in later sections).

4 DATASET DESCRIPTION

The data contains 2 CSV files: Boonsong Lekagul waterways readings and chemical units of measure. There is an image map of the wildlife preserve with marked sampling sites. The Boonsong Lekagul waterways' readings contain 136825 rows with the 5 attributes (columns), the semantics of which are described by the metadata: (Id, Value, Location, Sample Date, Measure). It contains readings from each sampling station over time for several different chemicals and water properties. The Sample Date spans from the year 1998 to 2016. Figure 1 shows the description of attributes.

S.No.	Attribute	Description	Type	Data Type
1.	Id	Identification number for the record. Maintained bookkeeping.	Ordered, Quantitative	Integer
2.	Value	The measured value for the chemical or property in this data point.	Ordered/Quantitative	Float
3.	Location	Name of the location the sample was taken from.	Categorical	String
4.	Sample Date	The date when the sample was taken from the location.	Ordered, Ordinal	Date
5.	Measure	Chemicals (e.g., Methyl Osmolene) or water properties (e.g. Water temperature) are measured in the data point.	Categorical	String

Fig. 1. Table summary of Boonsong Lekagul waterways csv file

The chemical units of measure contain 107 rows and 2 attributes (columns) semantics of which can be described as (Measure, Unit). It describes different units of measurement used in chemicals and water properties. Figure 2 shows the description of attributes. The data preprocessing steps performed in visualization (using python) are:

(1) Convert the values to a single unit of measure for all the chemicals.
(2) Extract the following attributes and store it in the csv files:Year, Measure, Achara, Boonsri, Busarakhan, Chai, Decha, Kannika, Kohsoom, Sakda, Somchair, Tansanee, count of samples, coordinates of region mapping.

S.No.	Attribute	Description	Type	Data Type
1.	Measure	Chemicals (e.g., Methyl Osmolene) or water properties (e.g. Water temperature) are measured in the data point.	Categorical	String
2.	Unit	Unit of measure	Categorical	String

Fig. 2. Table summary of chemical units of measure csv file

(3) For a certain chemical, calculate the value in every location (in case of null, replace it by 0) and sort the Sample Date (only year) in ascending order from 1998-2016.
(4) The year for which there is no reading of chemical concentrations, mark it by 0 value.

5 MC'S QUESTIONS/TASKS

Following 3 questions were raised in the mini challenge and along with the questions is specified the solution to those questions using our visual analytics system:

(1) **Characterize the past and most recent situation with respect to chemical contamination in the Boonsong Lekagul waterways. Do you see any trends of possible interest in this investigation?**

The stream graph in our visualization helps answer this question as after taking the particular chemical as an input from the user (through a drop down menu) we plotted 10 different streams for 10 different locations over our year range i.e. from 1998 to 2016. This stream graph visualizes the change in the concentration of a chemical and our using of dynamic tooltip provides information to the user as to what is the "Location", "Year" and "Value" of the concentration of a chemical (in µg/l). Action in this visualization is Analyze and Identify whereas the Targets include Trends and Paths.
The trends of interest shown using this visualization are sudden spikes in the concentration of a chemical for a particular year depicting some contaminating activity/manufacturing process that might have caused a sudden increase in the concentration for that particular year.

(2) **What anomalies do you find in the waterway samples dataset? How do these affect your analysis of potential problems to the environment? Is the Hydrology Department collecting sufficient data to understand the comprehensive situation across the Preserve? What changes would you propose to make in the sampling approach to best understand the situation?**

To answer this question we have stacked bar chart and scatter plots in our visualization. The stacked bar chart shows how many samples (of the selected chemical) were taken month-wise for the complete year range i.e. from 1998 - 2016. Anomaly analyzed here is that the samples are not collected regularly. For instance, for the chemical "ammonium" in January 2008 only 5 samples were collected from the location "Kohsoom", for February 2008 only 6 samples were collected of the same chemical from the location "Chai" and so on. The tooltip for the stacked bar chart depicts the dates in the month at which these samples were collected. Action in this visualization is Analyze, Search, Identify and Compare whereas the Targets include Trends, Distribution and Outliers.
The anomaly detected here is that the samples collection was not consistent and regular. Through the use of scatter plots, we compare the sample collection for 10 different locations of the selected chemical over the complete year range. This provides us a comparative study for all the locations along with the possible contamination trends.

(3) **After reviewing the data, do any of your findings cause particular concern for the Pipit or other wildlife? Would you suggest any changes in the sampling strategy to better understand the waterways situation in the Preserve?**

Through the innovative visualization we show the change in the concentration of a chemical over the years for different locations (which gets visualized on the pie chart as well). This visualization shows which particular basin is getting contaminated more (owing to the chemical and locations lying on that basin) and will not be suitable for the Pipit or other wildlife. Action in this visualization is Discover, Present, Summarize and Compare whereas the Targets include Trends and Topology.
Also, as specified in the answer to the previous question, since the collection of samples over the years for different locations is irregular, we can only get an estimate as to which locations/basin is getting contaminated more.

6 DATASET EXPLORATION

To analyze the data provided better, it was wrangled to different formats. This helped in plotting the various visualizations of the dashboard. Firstly, the chemical concentration values and their corresponding units were read from the separate files and merged. Since a few readings were measured in mg/l and few others in μg/l, all these data points were normalized to the μg/l unit.

Then the data was grouped together on the chemicals(measure) attribute, for each year and for each location. This was done by averaging the values measured for a specific chemical, in a specific year, in a specific location. The headers in this format of data, were

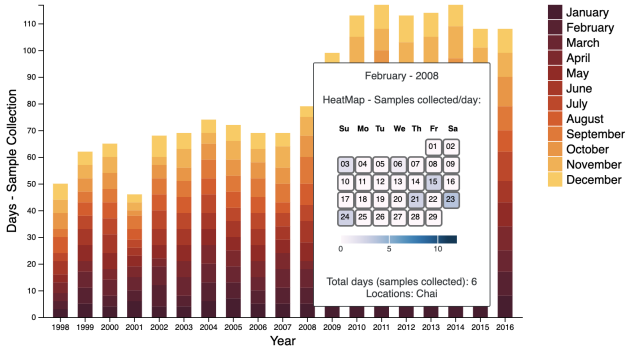

Fig. 3. Stacked Bar Charts along with Heatmap representation

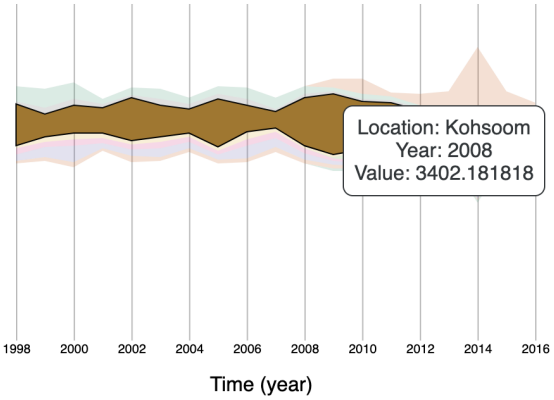

Fig. 4. Streamline Graphs

the ten different location, along with the measure, the year, the value and its units. If any of these data points did not exist, they were defaulted to zero. Data in this format helped in plotting the stream graph, the scatter-plots, the interactive graph and also the pie-chart.

For stacked bar graph and the heatmap, since we require the individual sampling dates, we created a data structure which was a nested array of dictionaries. This had "years" as its first set of keys, which were mapped to months, which were in turn mapped to the individual set of dates. These dates contained the number of samples taken on that date. This is computed on-the-fly as and when the chemical is changed in the dropdown.

7 INTERACTIONS

- On clicking any data point on the stacked bar graph or stream graph, the corresponding year's visualization is seen on the wildlife preserve map and on the pie-chart.
- On hovering over the stacked bar graph, a heatmap of the samples collected per day in the form of a tooltip can be seen.
- On hovering over the stream graph, the stream specific to the corresponding location is highlighted and the remaining are greyed out. Also, a tooltip with the data related to the stream is displayed.

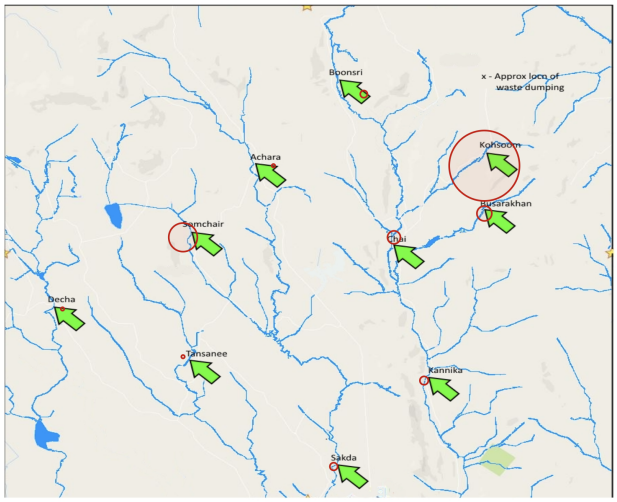

Fig. 5. Location Map Visualizations

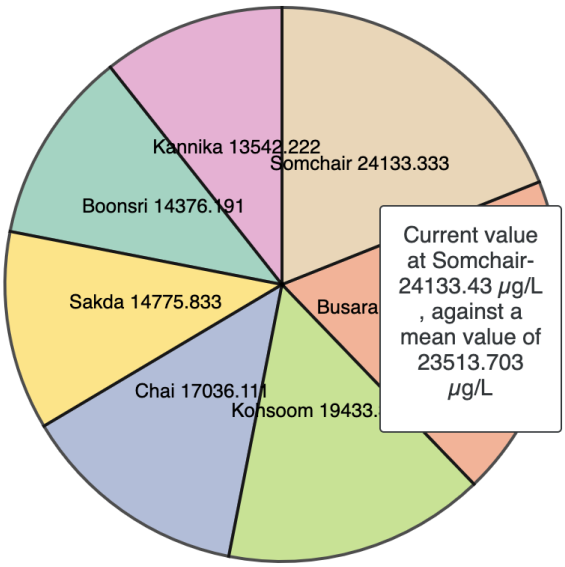

Fig. 6. Pie charts

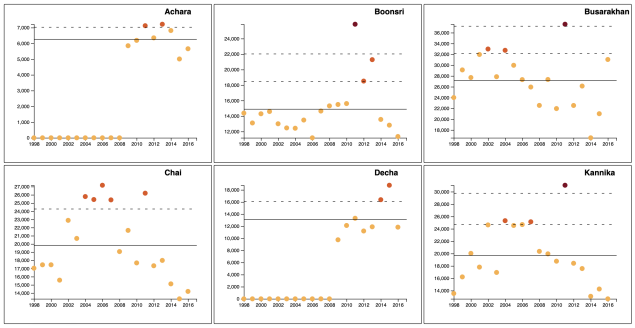

Fig. 7. Scatter plots (1)

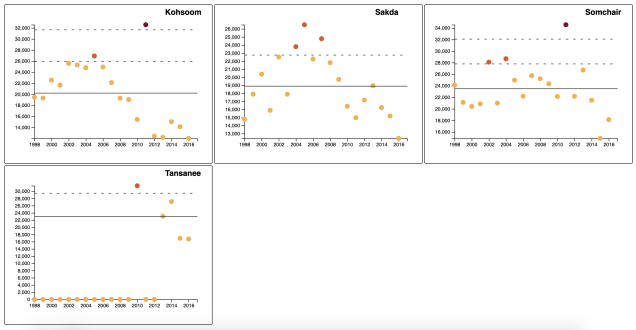

Fig. 8. Scatter Plots (2)

- On hovering over the pie-chart, a tooltip appears describing the current value of the measure at the location, comparing it to its mean over the years.

8 TEAM MEMBERS

(1) **Raja Sowmith Reddy Chityala** - rchityal@asu.edu, ASU ID: 1219447580
(2) **Krishna Teja Kalaparty** - kkalapa1@asu.edu, ASU ID: ID: 1219473892
(3) **Mahima Sandeep Bakshi** - mbakshi1@asu.edu, ASU ID: 1220205428
(4) **Nithya Vardhan Reddy Veerati** - nveerati@asu.edu, ASU ID: 1219450375
(5) **Prateek Srivastava** - psriva21@asu.edu, ASU ID: 1219492469
(6) **Utkarsh Sharma** - usharm13@asu.edu, ASU ID: 1219454509

9 REFERENCES

[1] https://www.d3-graph-gallery.com/graph/streamgraph_template.html
[2] https://www.d3-graph-gallery.com/graph/scatter_basic.html
[3] https://observablehq.com/@d3/scatterplot
[4] https://www.d3-graph-gallery.com/backgroundmap
[5] https://www.d3-graph-gallery.com/graph/barplot_stacked_basicWide.html
[6] https://www.d3-graph-gallery.com/graph/interactivity_tooltip.html
[7] https://www.d3-graph-gallery.com/connectedscatter
[8] https://www.d3-graph-gallery.com/graph/barplot_stacked_hover.html