**CST4060 – Visualisation Analytics.**

**Course Work II –**

**Data Visualisation with JavaScript.**

**YAHAYA, Suleyman Olushola.**

**M00736597**

[**Sy423@live.mdx.ac.uk**](mailto:Sy423@live.mdx.ac.uk)

**TASKS**

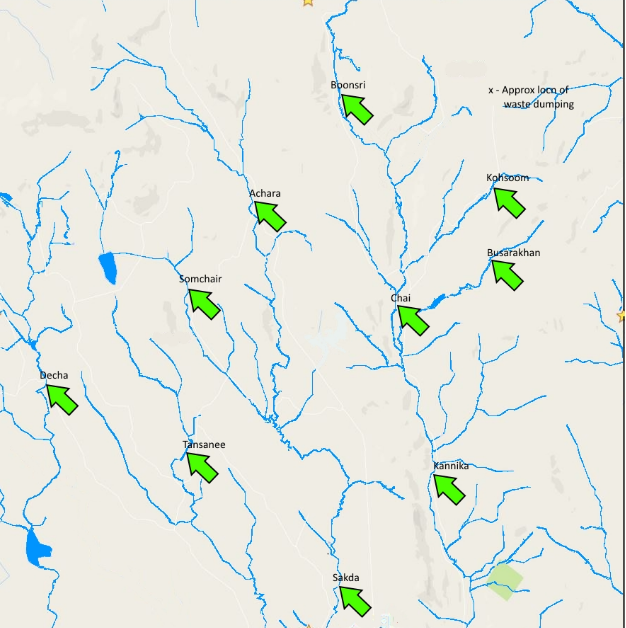
**Use Vega-Lite to create visualisations that answer the following analysis questions:**

1. *Describe trends and anomalies with respect to chemical contamination in the Boonsong Lekagul waterways.*
2. *Describe any data quality and uncertain issues, such as missing data, change in collection frequency, and unrealistic values (e.g., water temperature higher than 100 degrees).*
3. *Describe any particular concern for the Pipit or other Wildlife.*

* *You must use Vega-Lite to create the visualisations.*
* *You are free to apply any pre-processing and/or non-visual analysis to help answer these questions.*

**INTRODUCTION**

Ornithology professors from Mistford College provided eighteen years of hydrology data (1998-2016) of rivers and streams in the Boonsong Lekagul Wildlife Preserve. This data comprises water temperature and Chemical composite levels from different locations across the Preserve to investigate potential environmental pollution claims. Kasios Furniture Company is accused of dumping toxic wastes and polluting the air in and around the Preserve.



The map above indicates the locations were these readings were recorded.

**DATA**

The dataset is a Comma Separated File (CSV) which comprises five columns and 13,6825 data points across 104 measures. Descriptions of the columns for Boonsong Lekagul waterways readings are shown in the table below:

|  |  |
| --- | --- |
| **Id** | Identification number for the record (only for bookkeeping) |
| **Value** | The measured value for the chemical or property in this record |
| **Location** | The location where the readings were recorded as shown in the map. |
| **Sample Date** | Dates readings were recorded. |
| **Measure** | Chemicals (e.g., Sodium) or water properties (e.g., Water temperature) measured in the record |

**FINDINGS**

**TASK 1:**

The measures listed below have been selected from the one hundred and four Measures provided in the data. I chose these because they are some of the most dangerous chemicals known to man. They include:

* Biochemical Oxygen
* Mercury
* Arsenic
* Nickel

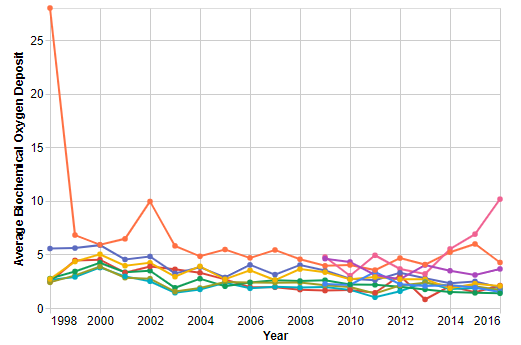
This task investigates the Ornithologists' claims via a data-driven approach rather than an environmental science approach by ascertaining the said chemicals' patterns.

This is done by taking an average reading of the measures over the years for the Locations. These locations are colour coded, as seen below.



**Biochemical Oxygen**

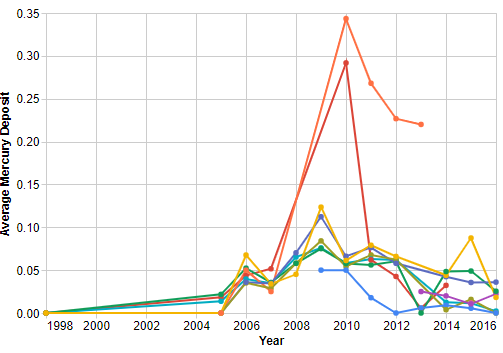
Demand for Biochemical Oxygen is the quantity of liquefied Oxygen needed by micro-organisms to decompose the organic matter in a stream, river or any body of water.



Judging by the visualisation above, an astronomically high Biochemical Oxygen level was present in the Kohsoom region in 1998 and in 2002, albeit significantly less before flattening, while other areas maintained reasonably low levels of Biochemical Oxygen until 2013, where Tansee began to see a linear rise.

**Mercury**

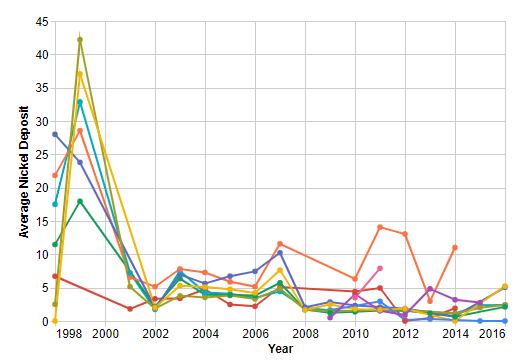
Mercury occurs naturally in the earth's crust, but human activities, such as mining and fossil fuel combustion, could lead to Mercury Pollution in the Bekagul Waterways. A high level of Mercury is hazardous to the Wildlife present in the Preserve.



All locations maintained similar trends in Mercury's slow rise until 2006, where a significant upward trend emerged. Somchair and Busarakhan saw higher Mercury levels in 2009, which is way less than the abnormal rise witnessed by Boonsrii and Kohsoom in 2009. Busarakhan also saw an uptick in 2015 before dipping back to normal levels in 2016.

**Nickel**

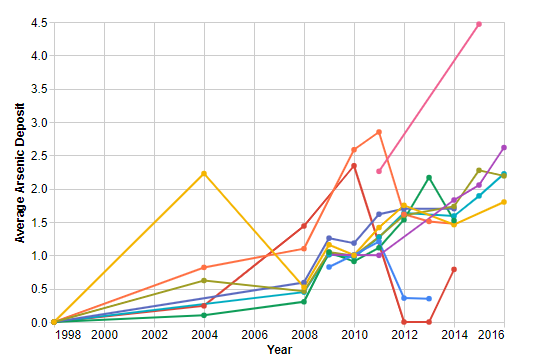
Nickel is a chemical element available in abundance and can be found in the air, soil, sediments, and water. A high Nickel concentration in the water or air in the Wildlife Preserve is dangerous to its inhabitants.



A very significant presence of Nickel is seen across all locations from 1998 to 2001. These numbers dropped to back to normal levels where continuous and similar trends are observed until 2016, except for the Kohsoom region which saw an uptick in 2006 and maintained its trend before dipping considerately in 2013.

**ARSENIC**

Arsenic is a chemical element present in the earth and can seep into groundwater. The presence of a high Arsenic concentration over prolonged periods is detrimental to the Wildlife in the Preserve.



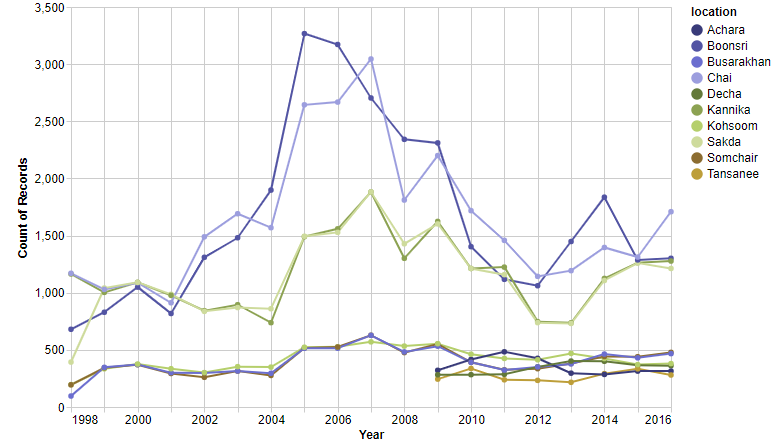
A slow but steady ready in Arsenic levels can be seen across most of the locations except for Busarakhan which saw an astronomic rise up until 2004 before dipping to normal levels and Boonsri whose significant linear surge began in 2004 before a sharp decline in 2012. Other regions maintained a sinusoidal pattern from 2008 to 2016.

Tansanee saw an anomalous from 2011 to 2016.

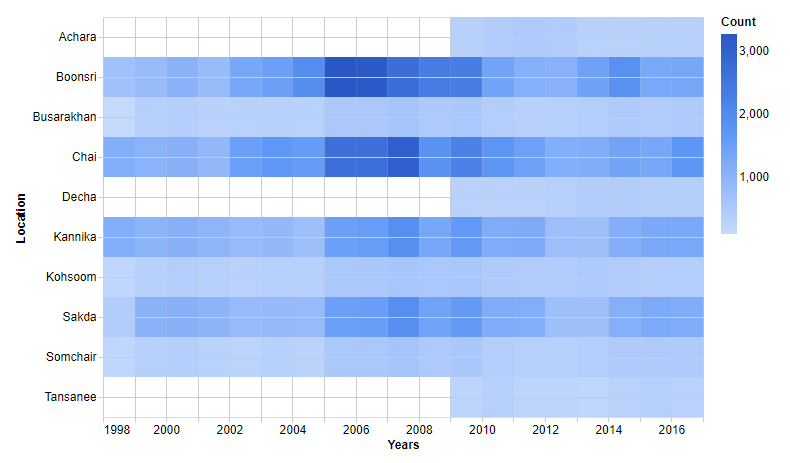
**TASK 2:**

One of the most crucial aspects of the Data Science lifecycle is the Data Exploration phase. This phase involves understanding everything about the data.

While exploring the data, it was discovered that the Measure column had ten distinct locations, and the readings for some of these locations did not commence in 1998. The graph below shows the collection of data over the years across all the areas.



Readings in Dacha, Achara and Tansanee began 2009 while the others started in 1998. Boonsri, Chai and Sadka have the highest data collection frequency. The Readings rose significantly between 2005 to 2010, as displayed above in the graph.



The Heatmap above chose the density of the collection of data for these locations over the years.

While investigating the trends, some anomalies were also discovered in the visualisations. These anomalies were corrected by computing for every measure's fourth quartile range and replacing any values greater or equal to the 4th Quartile with Average on Microsoft Excel.

The readings' inconsistencies could be down to a lack of resources to track these readings or a malfunction in the water sensors.

**TASK 3:**

**Concern For The Pipit and other Wildlife in the Preserve**

* Nickel, Arsenic and Mercury in high compositions, (whether individually or collectively) over a prolonged period, would have made inhabiting the Preserve for the Pipit and other Wildlife present in the Preserve very difficult.
* And also, increased demand for Biochemical Oxygen would have led to a rapid depletion of oxygen levels in the water, which could have had detrimental effects on survival in the Preserve.

**CONCLUSIONS**

**What and Why**

* The Line Chart with Dots utilised for Task 1, is the most suitable tool for analysing trends and Anomalies and it can be as seen killing two birds with one stone.
* The Sample Dates(Year) were encoded in the X-axes for Tasks 1 and 2, respectively, as they are temporal.
* The Heatmap is very appropriate when analysing data density as it shows the composition of data.
* In Task 2, the Value measure was in the colour marks to show the values' density.
* In Task 1, Locations were encoded in colour Marks to differentiate and analyse the visualisations' trends properly as they are nominal.