

# Coursework 2 – Data Visualisation with JavaScript (20%)

Kinza Shabbir M00736576

## Question No. 1: Characterize the past and most recent situation with respect to chemical contamination in the Boonsong Lekagul waterways.

### Finding (Pattern over time and location)

Below interactive dashboard shows one bar plot and two line plot. Plot1 shows bar chart of the total number of chemicals recorded over the time. Plot2 shows line chart for mean value of selected chemical for all locations over time. Plot3 which is also line chart shows average value for selected chemical over time. We can clearly see from plot1 shows there is no consistency at all in recording all chemicals. So, It is very difficult to compare all chemicals for their contamination in water over time. Plot 2 and plot 3 show a pattern of slightly high and low for all chemicals on all locations. However, Tansanee, Kohsoom and Achara shows abrupt peaks and low after 2009. It also shows that sampling started at Tansanee, Achara and Decha in 2009. There is a strange peak which can be seen in Plot2 and Plot 3 for all locations from 2002-2004 except Boonsri.

### 1. Narrowing down on chemicals with interesting findings

**Ammonium:** Plot3 shows decreasing pattern over time for Ammonium.

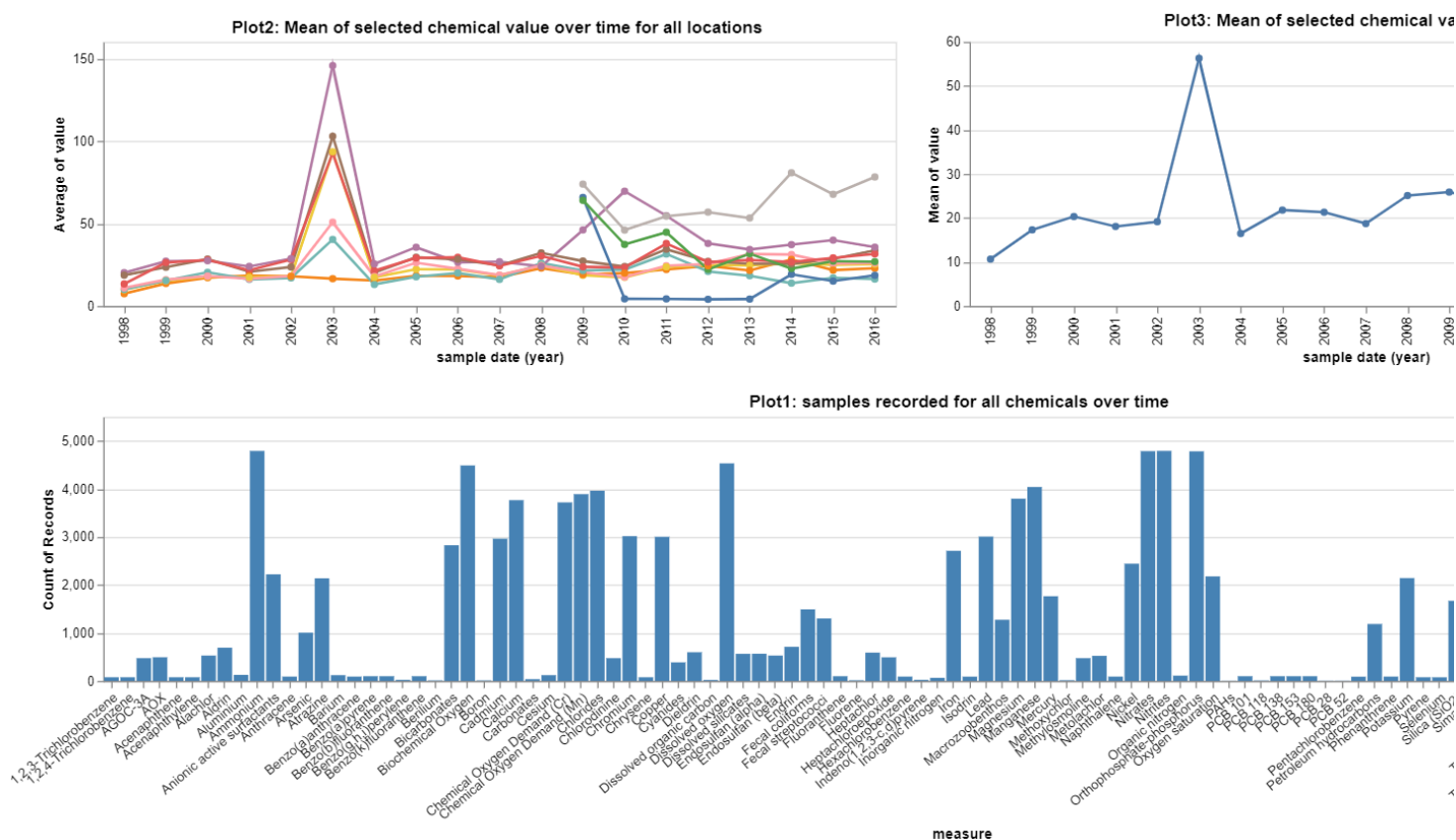
**Arsenic:** On selecting Arsenic on plot1, Plot3 shows increasing pattern for Arsenic suffered highest peak very recently in 2015.

**Lead:** Lead shows very sharp increase in 2000 recording highest peak. After 2000, it decreased suddenly from average value 30 to around 3 by the end of period.

**Nitrates:** Overall there is decreasing pattern comparing start and end of period.

**Orthophosphate-phosphorus:** Overall this chemical decreased but suffered highest peak between 2005 and 2015. Kohsoom and Boonsri shows abrupt increase and decrease for this chemical over time. Kohsoom having highest peak in 2014. Achara saw huge decrease in Orthophosphate-phosphorus.

**Water Temperature:** Overall there is slight increase in the temperature at the end as compared to start value.



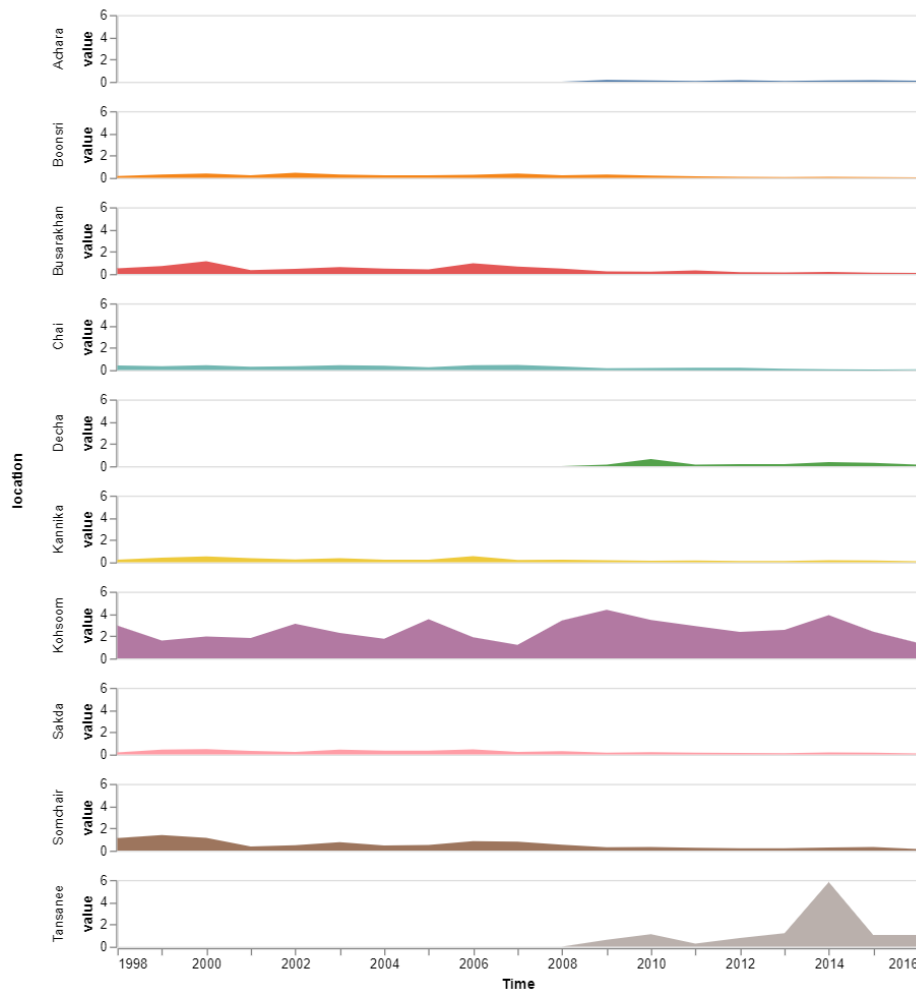
### 2. Narrowing down on chemicals across location

Faceting (Trellis Area) advanced feature has been used to display multiple plots for subgroups of same data with different location.

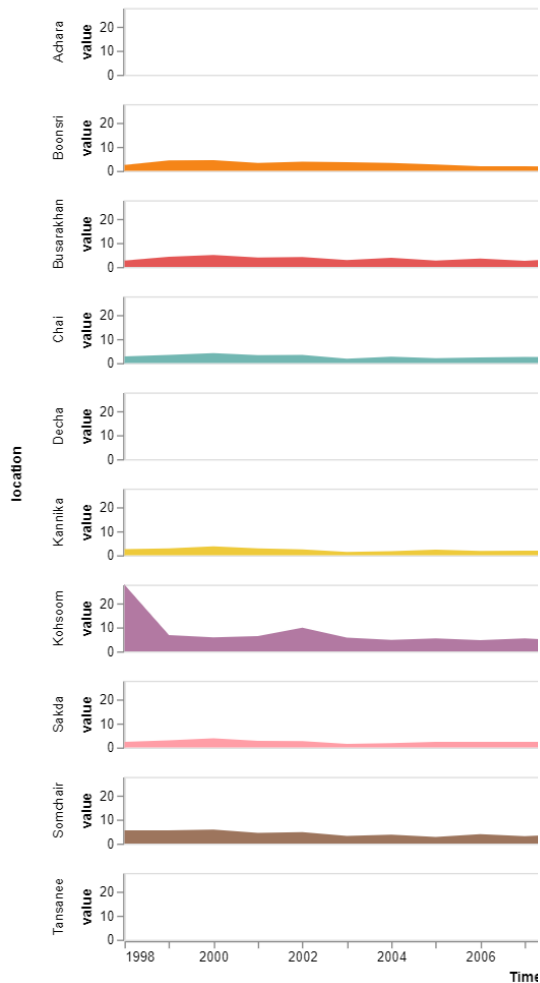
**Ammonium:** Plot4 shows Ammonium contamination had highest spike for Tansanee in 2014 and Kohsoom shows ups and down over the time.

**Arsenic:** Plot5 shows also decreasing pattern for this chemical on all location however, Tansanee has highest spike by the end of period in 2016.

Plot4: Ammonium Chemical Trend over time



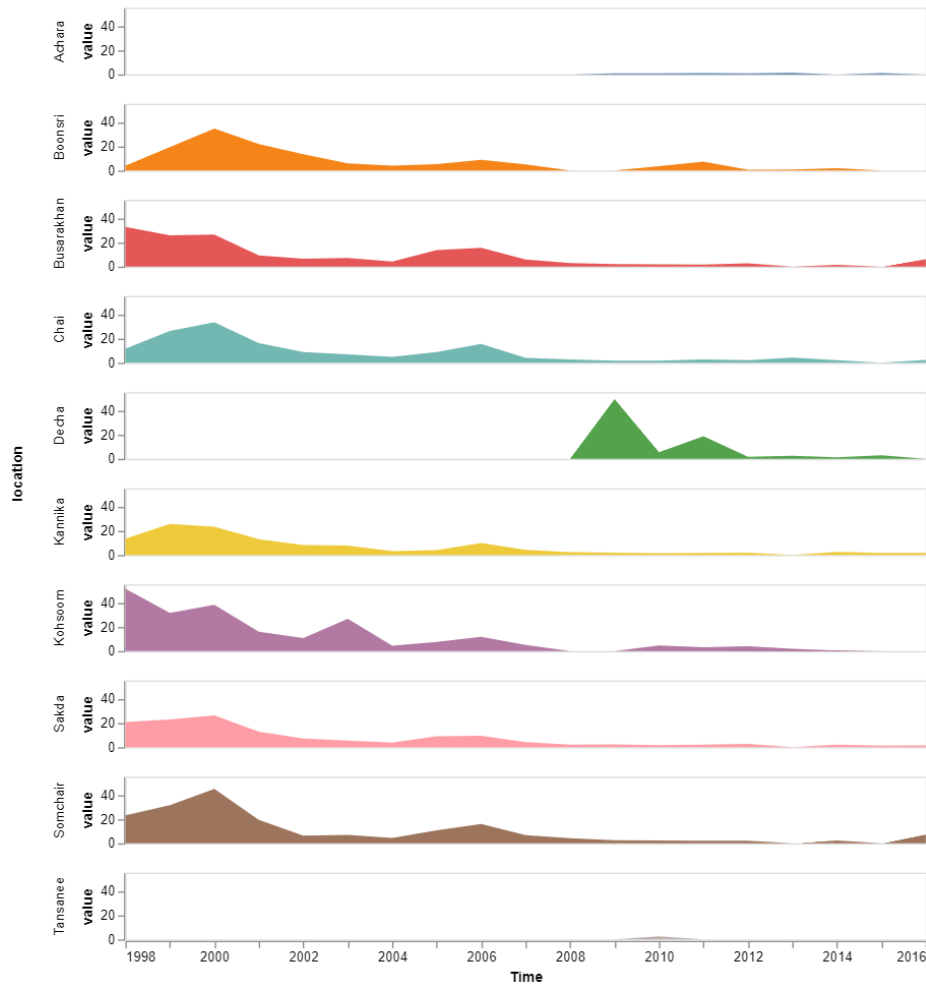
Plot 5: Arsenic Chemical Trend over time



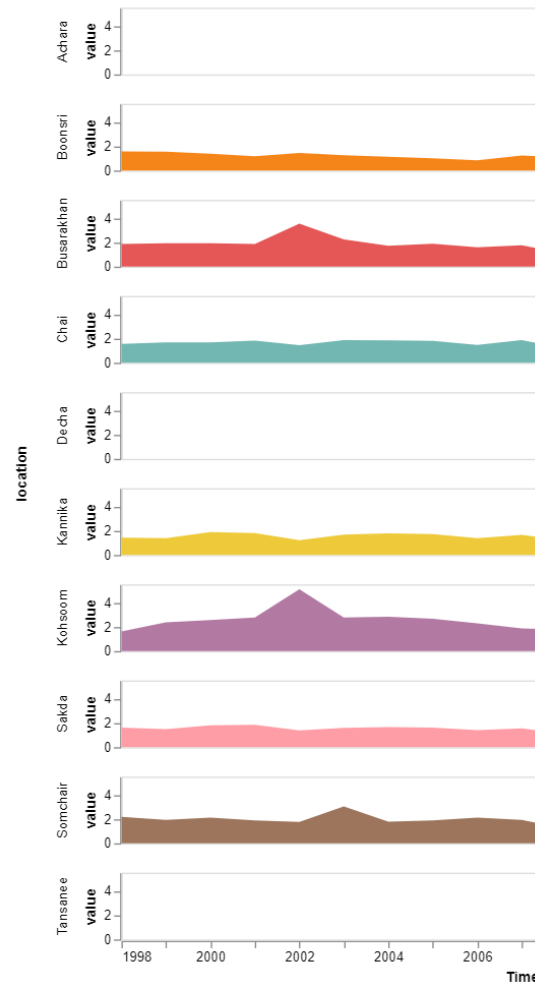
**Lead:** ALL location shows decreasing pattern for Lead. However, Busarakhan and Somchairr shows slightly increase by the end of period in 2016

**Nitrates:** Highest value of Nitrates is observed at Kohsoom. However, Kohsoom, Busarakhan, Somchair has spikes during 2001-2004 and Tansanee shows high spike in 2014

Plot 6: Lead Chemical Trend over time



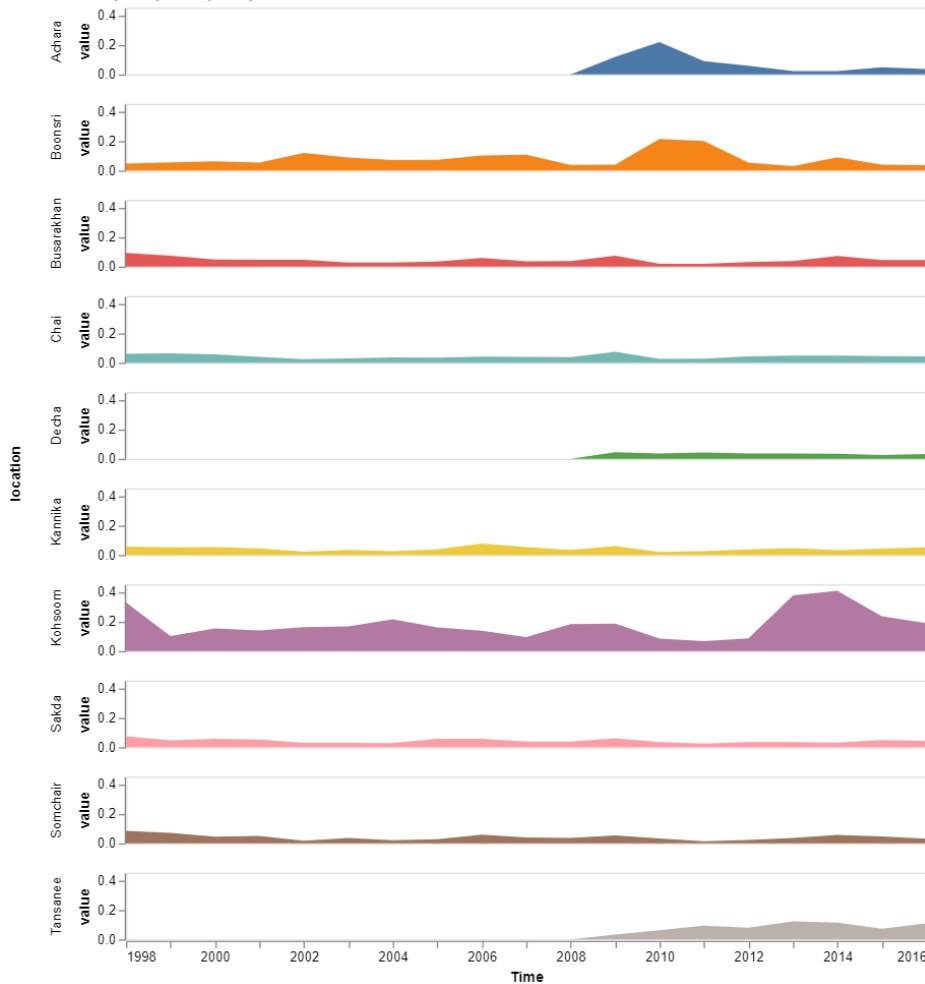
Plot 7: Nitrates Chemical Trend over time



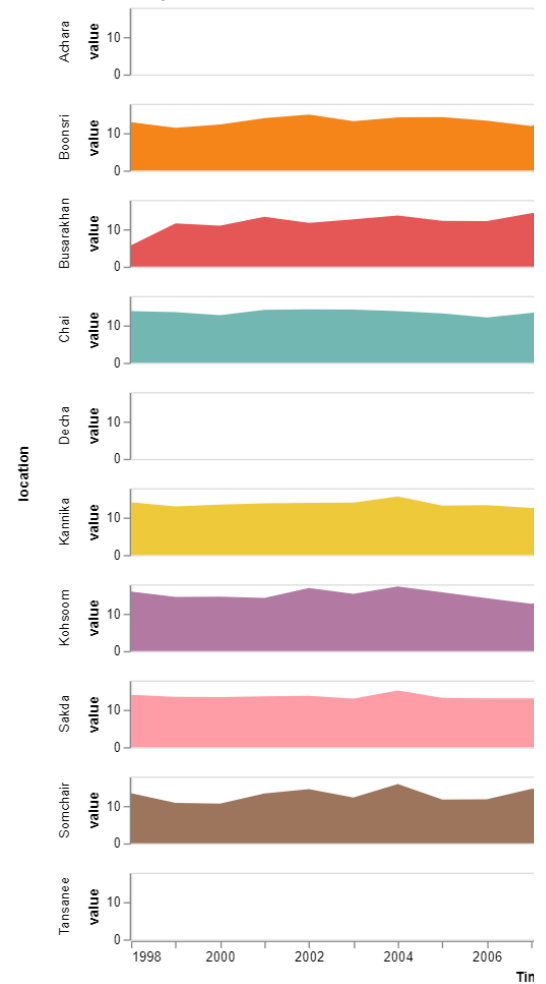
**Orthophosphate-phosphorus:** Kohsoom shows highest value for this chemical over the time recorded highest peak from 2012-2014. Kohsoom and Tansanee shows increase in value by the end of period. Achara and Boonsri had peaks 2008-2012

**Water temperature:** Overall there is very slight increase in water temperature on all locations.

Plot 8: Orthophosphate-phosphorus Chemical Trend over time



Plot 9: Water Temperature Trend over time



## 2. Marks and Channels

### 1. Plot1: Bar chart: **Mark** 1D Line **Channel**:

Position ( Horizontal position: categorical, Vertical position (quantitative)) and length **Why Effective?**

Because the purpose was to find the total samples recorded for each column. It is discrete analysis and clearly shows count for different chemicals.

### 2. Plot2 and Plot3: Line Chart **Mark**

1D: Line, Points **Channel**

Size (length) and Color (Categorical Attribute)

**Why Effective?** To find increasing or decreasing pattern, line chart is best option considering temporal variable at x-axis. It is suitable to look for trend over time

### 3. Plot 4-9: Trellis Area **Mark**

2D: Area **Channel**

color (Location) and Size Area (value attribute)

**Why effective** Used area chart to instead of line chart to see the clearly the shaded area to check the average value of each chemical measured on locations.

## 3. Advanced Features

**1. Interactive Dashboard with crosshighlight** Plot1, 2, 3 uses Dashboard to horizontally concatenate two line charts and one bar chart vertically. Dashboard uses selection of one bar and also changes other two line charts.

**2. Tooltip** to display values on place mouse.

**3. Facets (Trellis Area)** Plot 4- 9 uses repeat area chart for all locations

## Question#2 Anomalies

What anomalies do you find in the waterway samples dataset? Is the Hydrology Department collecting sufficient data to understand the comprehensive situation across the reserve?

### Finding(Inconsistency in data collection)

#### 1. Number of samples for all chemical on all locations

The below heatmap shows inconsistency in taking chemicals samples on different locations. Achara, Decha and Tansanee shows more white space for all chemicals which indicates the hydrology department didn't take samples on these location for most of chemicals. To check contamination of chemicals in water, the sensors on all should take samples consistently of all chemicals.

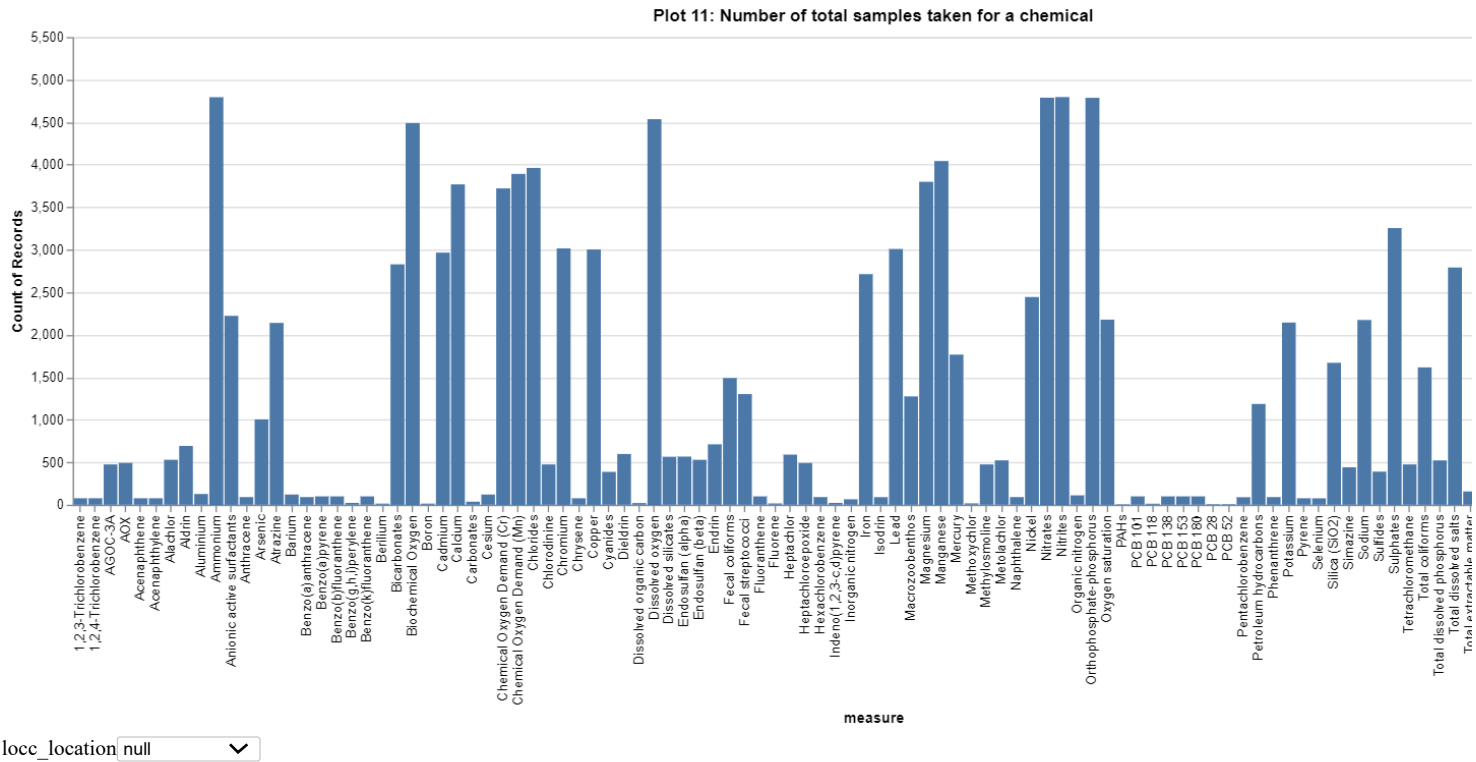
Plot 10: Heatmap with labels show sampling count for all locations over time

1,2,3-Trichlorobenzene		60					16
1,2,4-Trichlorobenzene		60					16
AGOC-3A		91	31	104		92	32
AOX		146	21	90		92	49
Acenaphthene		60					16
Acenaphthylene		60					16
Alachlor		135	28	148		82	28
Aldrin		210	35	175		93	48
Aluminium			10	30		39	
Ammonium	135	976	286	1233	148	701	216
Anionic active surfactants	14	665	100	339	58	362	164
Anthracene		60	2	5		2	16
Arsenic	62	137	52	170	21	232	65
Atrazine	45	382	123	507	22	408	108
Barium			12	28		31	
Benzo(a)anthracene		60	2	5		2	16
Benzo(a)pyrene		60	3	7		4	17
Benzo(b)fluoranthene		60	3	7		4	17
Benzo(g,h,i)perylene			3	7		4	1
Benzo(k)fluoranthene		60	3	7		4	17
Berilium			2	3		3	
Bicarbonates	29	832	122	447	41	471	202
Biochemical Oxygen	117	906	279	1178	92	674	216
Boron			2	3		3	
Cadmium	70	584	165	686	45	547	168
Calcium	89	863	215	720	86	665	217
Carbonates	1		7	9		6	
Cesium		21	3	36		36	6
Chemical Oxygen Demand (Cr)	120	815	192	801	93	604	211
Chemical Oxygen Demand (Mn)	104	857	202	827	94	666	215
Chlorides	88	871	215	872	80	674	217
Chlorodinine		91	31	104		92	32
Chromium	57	603	157	674	68	568	172
Chrysene		60					16
Copper	62	602	157	669	53	566	172
Cyanides	12		28	96	34	92	
Dieldrin		154	34	161		91	31
Dissolved organic carbon			1			3	
Dissolved oxygen	120	884	278	1215	99	682	216
Dissolved silicates	88	87	48	162	30	36	12
Endosulfan (alpha)		152	30	160		80	27
Endosulfan (beta)		133	29	154		77	27
Endrin		219	34	183		99	49
Fecal coliforms		461	71	284		240	141
Fecal streptococci		417	64	239		190	127
Fluoranthene		60	3	7		4	17
Fluorene		12					3
Heptachlor		148	30	168		93	32
Heptachloroepoxide		124	26	131		78	27
Hexachlorobenzene		60	2	5		2	18
Indeno(1,2,3-c,d)pyrene			3	7		4	1
Inorganic nitrogen	28	1			12		
Iron	8	575	148	652	52	479	146
Isodrin		60	2	5		2	17
Lead	70	592	165	695	47	558	170
Macrozoobenthos	2	354	63	267	8	238	58
Magnesium	89	864	218	732	83	680	217
Manganese	20	943	198	880	103	690	253
Mercury	80	340	91	395	35	353	51
Methoxychlor		4	1	6			2
Methylosmoline		91	31	104		92	32
Metolachlor		125	27	150		85	27
Naphthalene		60	2	5		2	16
Nickel	64	492	127	578	52	443	129
Nitrates	135	976	286	1233	150	701	216
Nitrites	135	976	286	1232	150	701	216
Organic nitrogen	89	4		17			
Orthophosphate-phosphorus	135	976	286	1233	148	696	217
Oxygen saturation	36	560	86	488	53	326	126
PAHs		1	1	1		1	1
PCB 101		60	3	7		4	19
PCB 118			2	5		2	
PCB 138		60	3	7		4	19
PCB 153		60	3	7		4	19
PCB 180		60	3	7		4	19
PCB 28			1	2		2	1
PCB 52			1	2		2	1
Pentachlorobenzene		60	2	5		2	16
Petroleum hydrocarbons	37	234	76	187	31	218	89
Phenanthrene		60	2	5		2	16
Potassium	18	759	91	296	30	329	209
Pyrene		60					16
Selenium			8	19		19	
Silica (SiO2)	12	408	84	502		246	86
Simazine		103	24	121		73	24
Sodium	22	768	91	302	40	329	213
Sulfides		59	23	117		77	10
Sulphates	22	860	168	619	77	551	213
Tetrachloromethane		91	31	104		92	32
Total coliforms	3	507	76	298	21	256	142
Total dissolved phosphorus		121	50	132		72	24
Total dissolved salts	28	569	151	708	80	437	141
Total extractable matter		27	12	36		31	8
Total hardness	30	521	91	229	23	272	124
Total nitrogen	133	227	196	777	113	407	58
Total organic carbon	20	137	13	54		50	48
Total phosphorus	134	958	274	1165	145	662	217
Trifluralin			2	5		2	
Water temperature	119	984	283	1559	96	697	217
Zinc	75	609	158	681	51	537	172
alpha-Hexachlorocyclohexane		154	32	167		96	31
beta-Hexachlorocyclohexane		173	34	174		95	32
gamma-Hexachlorocyclohexane	50	526	151	577	25	451	153
p,p-DDD		161	34	179		94	32
p,p-DDE		175	34	186		102	32
p,p-DDT	48	471	123	485	25	386	138
	Adhara	Boonsri	sarakhan	Chai	Decha	Kannika	Chosoom

2. Number of samples taken for each chemical

Bar plot below shows total number of samples collected for each chemical. This plot shows difference in number of samples collected for each chemical.

- 1.Minimum number of samples collected for a chemical= 7
- 2.Maximum number samples for one chemical=5031
- 3.For 18 chemicals, samples count >3000 have been collected
- 4.For 35 chemicals, samples count <100



locc\_location null

3. Number of samples recorded in a year on all locations

The line chart below shows the inconsistency of samples taken on each location.

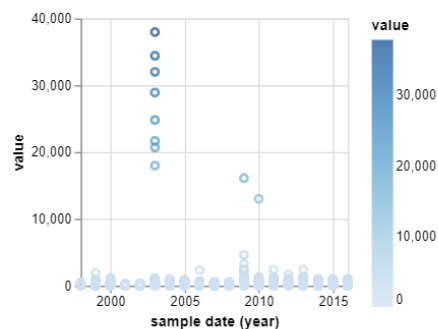
1. Highest number of samples taken on Boonsri and chai during 2004-2009.
2. Almost same number of samples taken on Kannika and Sakda and highest number of samples taken during 2004-2011.
3. Somchair, Busarakhan and Kohsoom shows same samples count over the period having peak from 2005-2009
4. The hydrology department started sampling at Decha, Achara and tansanee in 2009 and the sample count is lowest as compared to other locations. The reason to start taking samples late could be because these three locations are far from dumping location.

4. Strange Sampling of Chemicals during 2002-2004

The below dot plot13 and 14 with brushing to output data shows totally same values for two different chemicals Iron and zinc samples. The value is outlier as compared to other values , that might be happened because of sensors not working properly.

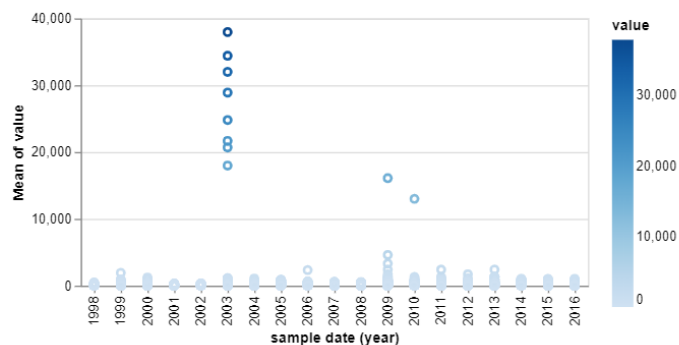
Happened on 15th August, 2003 on Kohsoom, Chai, Somchair, Busarakhan, Kannika and Sakda.

Plot 13: Anomaly in Iron sampling



Location	sample date	value
Boonsri	Jan 11, 1998	2
Boonsri	Jan 11, 1998	9.1
Boonsri	Jan 11, 1998	0.33
Boonsri	Jan 11, 1998	0.01
Boonsri	Jan 11, 1998	1.47
Boonsri	Jan 11, 1998	0.06
Boonsri	Jan 11, 1998	0.09
Boonsri	Jan 11, 1998	13.9
Boonsri	Jan 11, 1998	3.5
Boonsri	Jan 11, 1998	52
Boonsri	Jan 11, 1998	9.7
Boonsri	Jan 11, 1998	19.2
Boonsri	Jan 11, 1998	36.7
Boonsri	Jan 11, 1998	4.95
Boonsri	Jan 11, 1998	0.01
Boonsri	Jan 11, 1998	21
Boonsri	Jan 11, 1998	22
Boonsri	Jan 11, 1998	17
Boonsri	Jan 11, 1998	0

plot 14: Anomaly in Zinc sampling



Location	sample date	value
Boonsri	Jan 11, 1998	2
Boonsri	Jan 11, 1998	9.1
Boonsri	Jan 11, 1998	0.33
Boonsri	Jan 11, 1998	0.01
Boonsri	Jan 11, 1998	1.47
Boonsri	Jan 11, 1998	0.06
Boonsri	Jan 11, 1998	0.09
Boonsri	Jan 11, 1998	13.9
Boonsri	Jan 11, 1998	3.5
Boonsri	Jan 11, 1998	52
Boonsri	Jan 11, 1998	9.7
Boonsri	Jan 11, 1998	19.2
Boonsri	Jan 11, 1998	36.7
Boonsri	Jan 11, 1998	4.95
Boonsri	Jan 11, 1998	0.01
Boonsri	Jan 11, 1998	21
Boonsri	Jan 11, 1998	22
Boonsri	Jan 11, 1998	17
Boonsri	Jan 11, 1998	0

## 2. Marks and Channels

1. Plot 10:Heatmap: **Mark Area Channel** : color saturation by quantitative attribute(count) and two categorical attribute (measure, location)

**Why effective** It is best suitable for more than two variables analysis as I used heatmap to show total samples count for all chemicals and for all locations as bar chart shows between one categorical and one quantitative.

2. Plot 11: Bar Chart **Mark Line 1DChannel** Length and size

**Why effective** Same as question No.1 bar chart plot

3. Plot 12: line chart **Mark 1D**: line **Channel** color (Location) and length is used to show trend for count of samples.

4. Plot 13 & 14: Dot plot **Mark 0D**: point **Channel** Vertical and horizontal position, Color Saturation(value)

**Why effective?** I used dot plot specifically to display outliers in sampling for chemicals and using brushing to see data for these outliers.

## 3. Advanced Features

1. **Layered plot(Labeling and Annotation)** Plot 10 : Heatmap has layering of text which shows sampling count of a chemical for a location.

2. **Tooltip** to display values on place mouse.

3. **Query Widget** Plot 11 bar chart has Query Widget for selecting one location.

4. **Brushing to output data in table** Plot 13 and 14 is dot plot with brushing to show data in table form for anomalous data in the plot.

## Question No. 3.

**After reviewing the data, do any of your findings cause particular concern for the Pipit or other wildlife?**

### Finding(Concern for Wildlife)

There are some chemicals which are dangerous for wildlife.

Ammonium, Arsenic, Lead, Nitrites, Orthophosphate-phosphorus, Selenium and rising water temperature[1].

Ammonium , Lead, Nitrates, Orthophosphate-phosphorus decreased over time. So these are not concern. However, water temperature, Arsenic and Selenium is of concern because they have increasing pattern and more pollutants for wildlife.

## 1. Rising temperature of water over time

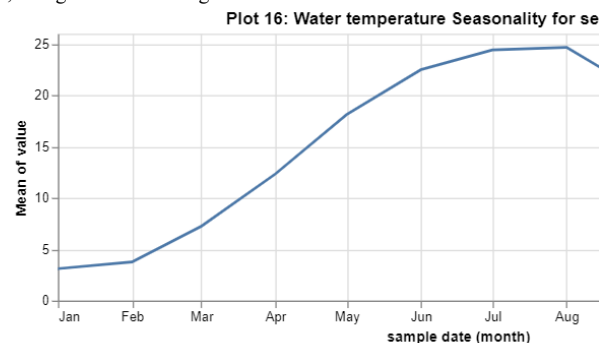
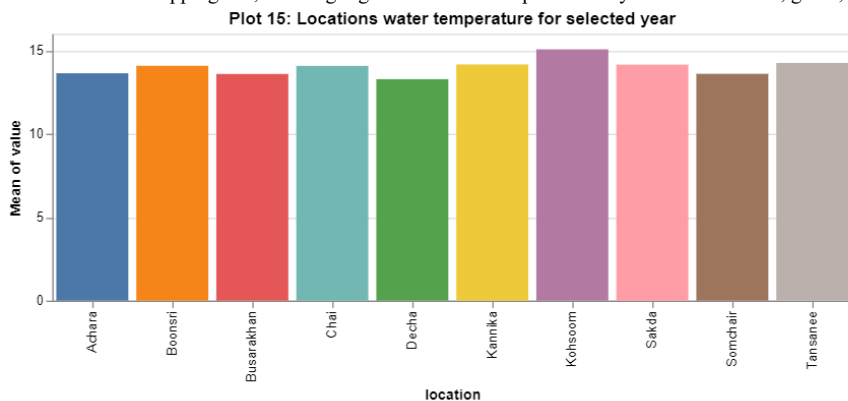
Plot 17 shows higher average temperature recorded in 2003, 2004 and end of period in 2016.

Narrowing on 2016, plot 15 shows average temperature for all location in 2016. Boonsri, Achara, kohsoom and Tansanee shows average temperature above 15c.

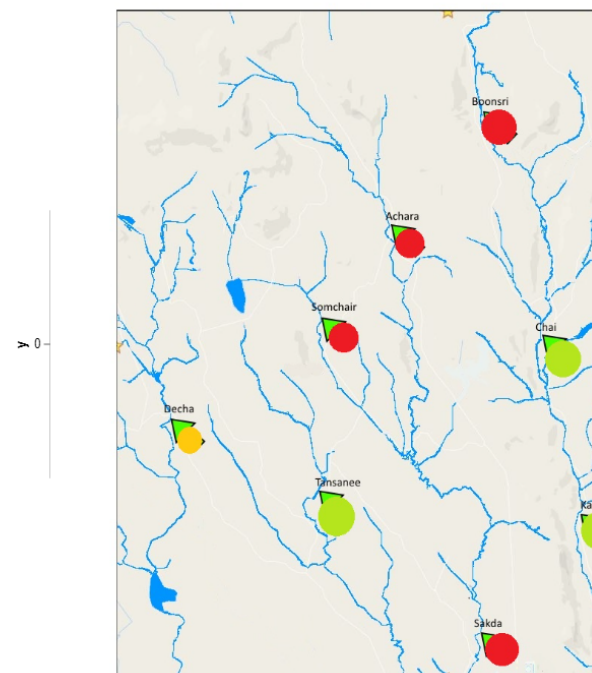
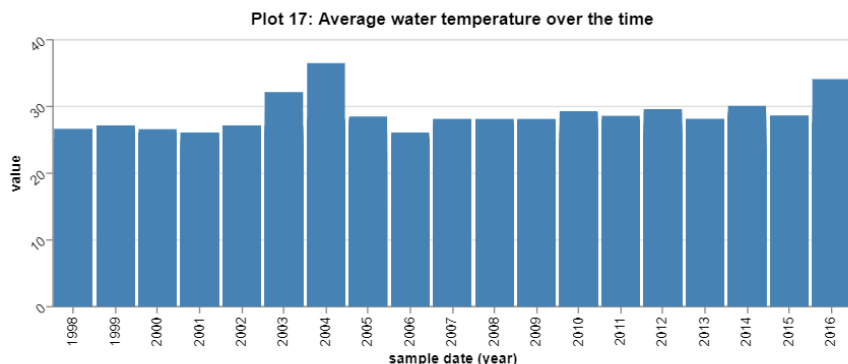
For 2016, plot 16 shows the average temperature for all months. After feb, there is sudden increase in temperature recorded peak in august and decreased till december.

Plot 18: shows circle size to represent higher temperature value and color grouping locations based on temperature value.

Kohsoom is near dupping site, showing highest value for temperature in yellow color. Then, green, red, orange in order of large to small value.



**Plot 18: Water Temperature on Map**



## 2. Arsenic and Selenium increased contamination in Water

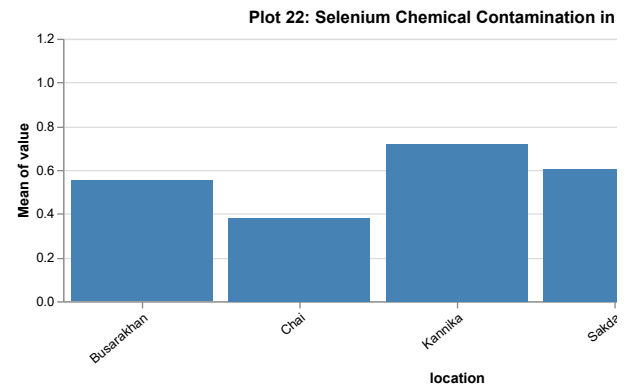
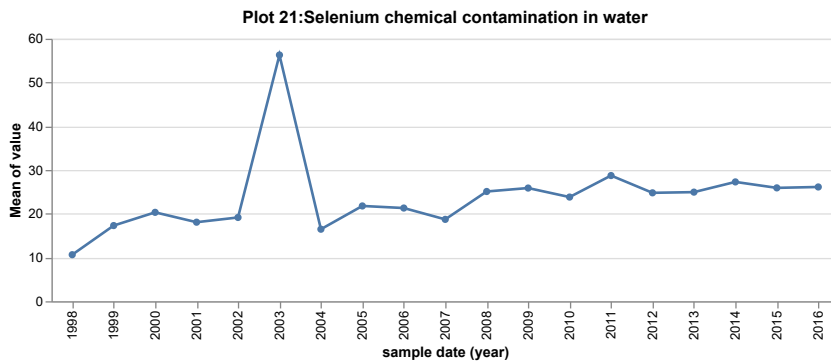
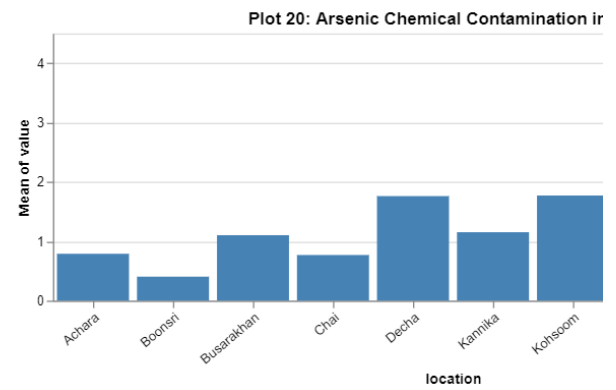
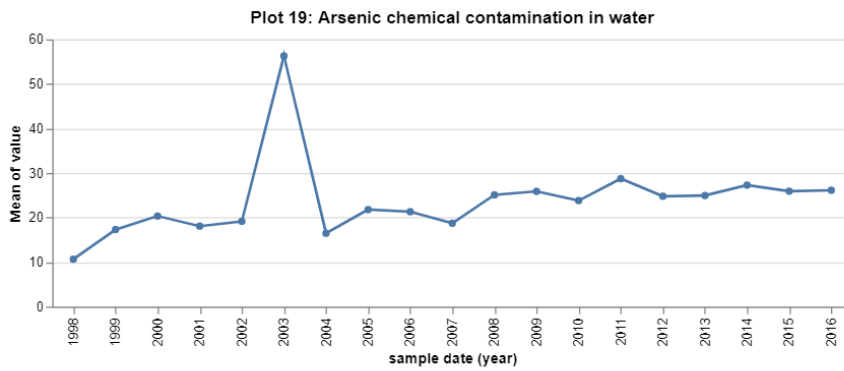
As mentioned above Arsenic is concern for wildlife. Its contamination should be reduced.

Plot 19 shows Arsenic value has slightly increased from start by the end of period.

Plot 20 shows that Tansanee has highest value above 4 which is the location of concern.

Plot 21 shows slight increase by the end of the period in Selenium value. plot 22 shows Somchair shows highest contamination of this chemical. This chemical is recorded orobserved only on five locations





## 2. Marks and Channels

1. Plot 15, 17, 20, 22: Bar Chart **Mark** Line 1D **Channel** Length and size,color is used for because of one categorical and one quantitative data.
2. Plot 12: line chart **Mark** 1D: line **Channel** length is used to show trend for chemicals and water temperature seasonality.
3. Map image display for showing the locations which has high temperature are near dumping location.

## 3. Advanced Features

1. **Interactive dashboard with selection** Plot 15, 16, 17 uses dashboard with crossfilter.
2. **Tooltip** to display values on place mouse.
3. **Filtering** Plot 19-22 uses filter the two chemicals out of all.

## Conclusion

1. After detailed analysis, the first question raise more concerns for Kohsoom and Tansanee which are having high contamination of chemicals as compared to other locations.
2. The inconsistency in taking the samples of chemicals on different locations. On all locations sampling should be done in sequence with the flow of water in the reserve.
3. There is decreasing pattern in some dangerous chemicals level for wildlife but some new dangrous are increasing in pattern leaving concern for wildlife.
4. Sampling should be taken first on location near dumping and then where it flows to other location to check contamination for same chemical.

**Reference**1. <https://www.sepa.org.uk/media/120299/chemistry-of-water-pollution.pdf>