Coursework 2 - Group Work

Coursework 2 Group Work

- Group size – 5 students
- 25% of the total module mark
- Due at 5pm on the Friday of Week 12 (see the submission page for date and time).
- The marking scheme and dataset are on the submission page

Dataset

- VAST challenge 2018 Mini Challenge 2:

  http://www.vacommunity.org/VAST+Challenge+2018+MC2
- VAST challenge is an annual international visual analytics competition
- It provides a dataset and a few analysis questions
- Each entry needs to present:
  - The visual analytics tool they developed and
  - How they found the answers using the tool

Background

(All the people, places, groups, technologies, contained therein are fictitious.)

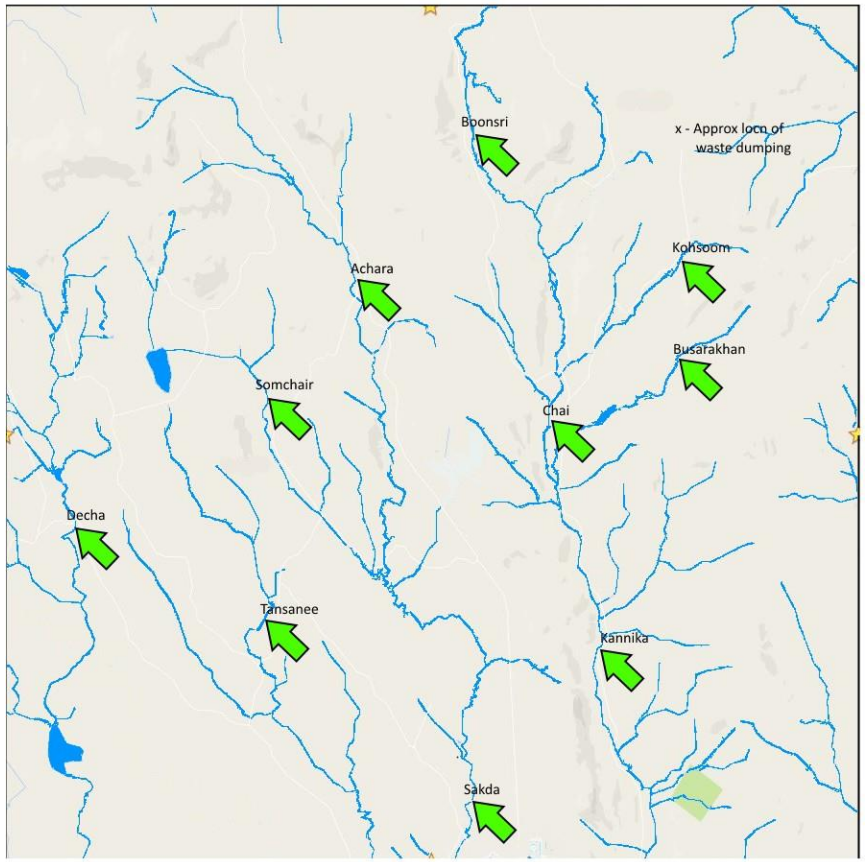
- Mistford is a mid-size city to the southwest of the Boonsong Lekagul Wildlife Preserve.
  - The city has a small industrial area with four light-manufacturing endeavors.
- Mistford and the wildlife preserve are struggling with the possible endangerment of the Rose-Crested Blue Pipit, a locally loved bird.
  - The bird's nesting pairs seem to have decreased alarmingly

Background

- An investigation last year (VAST challenge 2017) indicated that the Kasios Office Furniture, a Mistford manufacturing firm, may be linked to this
    - Though there is no firm evidence.
- Now the company insists that they have done nothing wrong.
- It is time for more visual analytics investigation.

Dataset

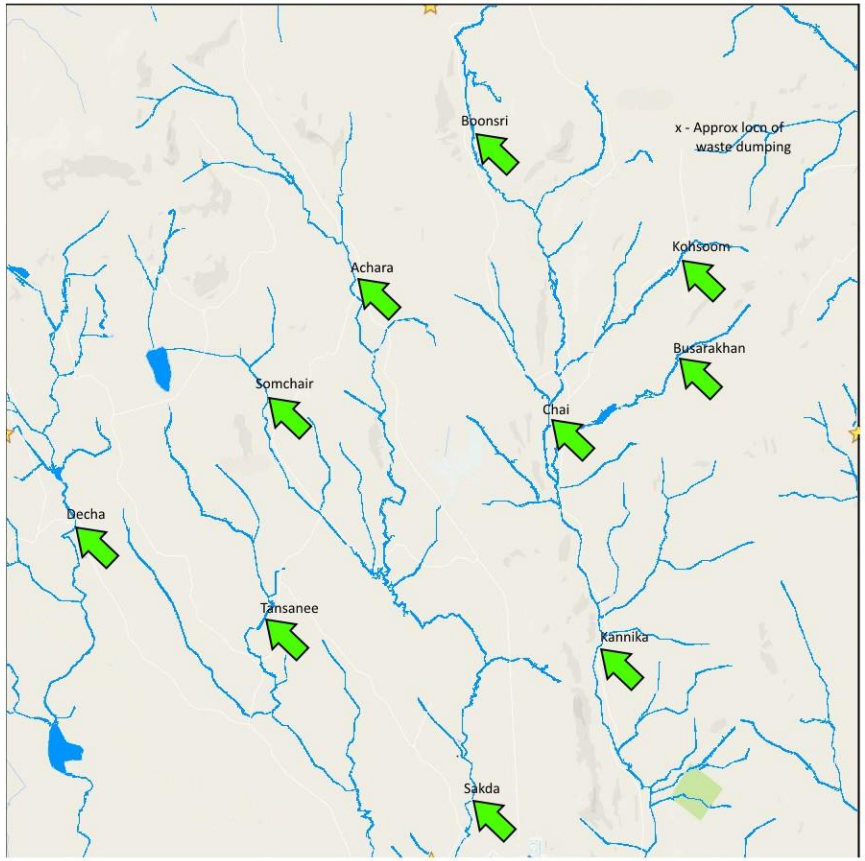
- Several years of water sensor readings from rivers and streams in the preserve.
- These samples were taken from different locations scattered throughout the area
    - Contain measurements of several chemicals of possible interest
- Your task is to investigate the sensor readings to find possible link to the bird population deduction.



Dataset

- The map
- The sensor reading data

```
id,value,location,sample date,measure
2221,2,Boonsri,11-Jan-98,Water temperature
2223,9.1,Boonsri,11-Jan-98,Dissolved oxygen
2227,0.33,Boonsri,11-Jan-98,Ammonium
2228,0.01,Boonsri,11-Jan-98,Nitrites
2229,1.47,Boonsri,11-Jan-98,Nitrates
2231,0.09,Boonsri,11-Jan-98,Total phosphorus
2232,13.9,Boonsri,11-Jan-98,Sodium
2233,3.5,Boonsri,11-Jan-98,Potassium
```



Dataset

- There is no need to introduce additional dataset, such as
  - The dataset from VAST 2017
  - What chemical is toxic to wild life and at what level
- Answer the questions with only the information provided.

Analysis questions:

1. Describe any data quality and uncertain issues, such as
    i. missing data,
    ii. change in collection frequency, and
    iii. unrealistic values (e.g. water temperature higher than 100 degrees).

2. Describe trends and anomalies with respect to chemical contamination
    i. Trends: changes over time and/or sensor site
    ii. Anomalies: sudden change over time or one site significantly different from others.

Use Altair to create visualisations

- You must use Altair to create the visualisations;
  - Tableau or other visualisation library is not allowed.
- You are free to apply any pre-processing and/or non-visual analysis to help answer these questions.
  - These can be done in a separate tool such as Excel/R/Jupyter
  - Or using Python

Requirements

- There should be at least one visualisation for each analysis undertaken.
- Usually, 3-5 visualisations (including dashboard) is expected for each questions:
    - There can be more than one trend or anomaly in each question;

Requirements

Besides the visualisations, the answer to each question should include discussion:

- What the finding is (a pattern, an anomaly, etc.);
- How the finding can be seen from the visualisation;
- How the visualisation design support the analysis, i.e. what the data and analysis task are and how the visualisation is designed to match and support them.
- Any advanced Altair visualisation features used, such as multi-layer, chart concatenation, and interaction.
- Any additional (non-visual) analysis used and how it contributed to the answer.

Report Marking scheme (total 20%)

The quality of the findings, i.e., how insightful is the finding (6%)

- What is the finding, i.e., what message the visualisation aims to convey;
- Insightful finding receives higher mark:
  - for example, findings that considers multiple aspects of the data, such as time, location, and measurements is more interesting than those with less aspects;
- Visualisation that clearly shows the intended finding receives higher mark.

The effectiveness of the visualisation design (8%):

Why such visual mapping is effective for the given data (*what*) and analysis (*why*), e.g.,

- Why is the chart type most appropriate for the analysis?
- Why are the choice of mark and channel the most effective?
- Is there any additional feature, such as sorting/filtering, dashboard or interactions, is used to improve the visualisation?

The quality of the visualisation and analysis code (6%):

- The quality of Altair code;
- Usage of advanced features (which contribute to the analysis) such as multiple
  - views/dashboard and
  - interaction receives
- Usage of additional analysis
  - such as statistical analysis that
  - contributes to the analysis

Submission

- A single 8pg report explaining the work of your group (20%)
    - **Name and student number of all your group members**
    - Including images/screenshots/code snippets
    - Do not include full code.
- A 5min group video of your presentation explaining the work done (5%)
- A combined Jupyter notebook including the Altair visualisations and answer to the analysis questions.
- All the necessary files should be included in a zip file (max 100MB), including the dataset
    - Please include the (original) data files used in the visualisations, especially if you change the date (reformatted, adding new analysis results, etc.)
- You are free to use any third-party library or API to help with the analysis.
    - Make sure the files are included if any additional local file is needed.

Questions?