VAST Challenge 2017:
Mini Challenge 2
‘Project Gaia’

ASSIGNMENT 2 – VISUAL ANALYTICS – COMP 5048

Project Gaia Team Members		
Dan BRIDGMAN (4604454256)	Dan ELIAS (470293722)	Kristopher LOPEZ (470236420)
Matthew BURGESS (430173596)	Andrew HUANG (470327371)	Cameron WASILEWSKY (470267240)

Declaration: We declare that we have read and understood the University of Sydney Academic Dishonesty and Plagiarism in Coursework Policy, and except where specifically acknowledged, the work contained in this assignment/project is our own work, and has not been copied from other sources or been previously submitted for award or assessment.

We understand that failure to comply with the Academic Dishonesty and Plagiarism in Coursework Policy, can lead to severe penalties as outlined under Chapter 8 of the University of Sydney By-Law 1999 (as amended). These penalties may be imposed in cases where any significant portion of my submitted work has been copied without proper acknowledgement from other sources, including published works, the internet, existing programs, the work of other students, or work previously submitted for other awards or assessments.

We realise that each individual may be asked to identify those portions of the work contributed by themselves and required to demonstrate their knowledge of the relevant material by answering oral questions or by undertaking supplementary work, either written or in the laboratory, in order to arrive at the final assessment mark.

Signed:

Dan BRIDGMAN

Matthew BURGESS

Dan ELIAS

Andrew HUANG

Kristopher LOPEZ

Cameron WASILEWSKY

Report Table of Contents

1. AIMS AND CONTRIBUTION

In March 2017, the Institute of Electrical and Electronics Engineers ('IEEE') for Visual Analytics Science and Technology ('VAST') announced an annual competition for the Visual Analytics community to design interactive systems to help solve conceptual environmental problems.

A fictional environmental problem, called 'Mini Challenge 2' ('MC2'), was the focus of this project, code named 'Gaia'.

The project requires the design and development of a Visual Analytics (VA) system. The system is designed to enable a new user to visually manipulate data and gain the insights needed to solve the problem questions.

1.0 Problem Background

MC2 is based on the town of Mistford and its neighbouring industrial zone. Four factories are potentially emitting up to four types of gaseous chemicals, concentrations of which are being picked up by nine different sensors.

Tables 1 and 2 respectively summarize the chemical pollutants and factories in this study. In all the visualizations in this report and on the web app, the same colour coding is used to distinguish the chemical pollutants.

Chemical and colour
Methylosmolene
Chlorodinine
Appluimonia
AGOC-3A

Table 1 - Chemical Pollutants and colour coding

Factory company	**Factory**
'**Roadrunner** Fitness Electronics'	Roadrunner
'**Kaisos** Office Furniture'	Kaisos
'**Radiance** ColorTek'	Radiance
'**Indigo** Sol Boards'	Indigo

Table 2 - Factories

Nine chemical sensors are located near the factories as shown in Figure 1.

Figure 1 - Map of sensor and factory locations

1.1 Aims

The aims of this project were two-fold:

- Build a dedicated VA system that would allow users to (1) explore and interact with the data in a scientific and information visualisation environment and (2) enable users to form hypotheses and derive insights.
- Through using our VA system, provide compelling answers for the three questions in the MC2 challenge:
    1. Characterize the sensors' performance and operation. Are they all working properly at all times? Can you detect any unexpected behaviors of the sensors through analyzing the readings they capture?
    2. Now turn your attention to the chemicals themselves. Which chemicals are being detected by the sensor group? What patterns of chemical releases do you see, as being reported in the data?
    3. Which factories are responsible for which chemical releases? Carefully describe how you determined this using all the data you have available. For the factories you identified, describe any observed patterns of operation revealed in the data.

1.2 Contribution

The table below summarises the contribution of each team member to the project:

Project Gaia Team Member	Roles	Contribution
Dan BRIDGMAN (4604454256)	Project Management Lead, meeting chair, report and presentation author, presenter	16.6%
Matthew BURGESS (430173596)	Unity developer, web developer	16.6%
Dan ELIAS (470293722)	Development, analysis	16.6%
Andrew HUANG (470327371)	Research, VA UX design, report author	16.6%
Kristopher LOPEZ (470236420)	Developer, designer and mathematician	16.6%
Cameron WASILEWSKY (470267240)	Developer, presentation author, presenter	16.6%

Table 3 - Contributions

2. DATA SET

The data set provided for the challenge included meteorological and sensor reading data from the months of April, August and December 2016.

Meteorological data was supplied in three-hourly frequencies, which included wind speed (in metres per second) and wind direction (in 360/000 north-referenced azimuth format) at an assumed constant elevation of 370 metres above ground.

Sensor data was available at one-hourly frequencies. Each data point referred to a single reading of a chemical from a single sensor and its detected chemical concentration (in parts per million).

The dates for which chemical and wind data are supplied are illustrated in the scatterplot below.

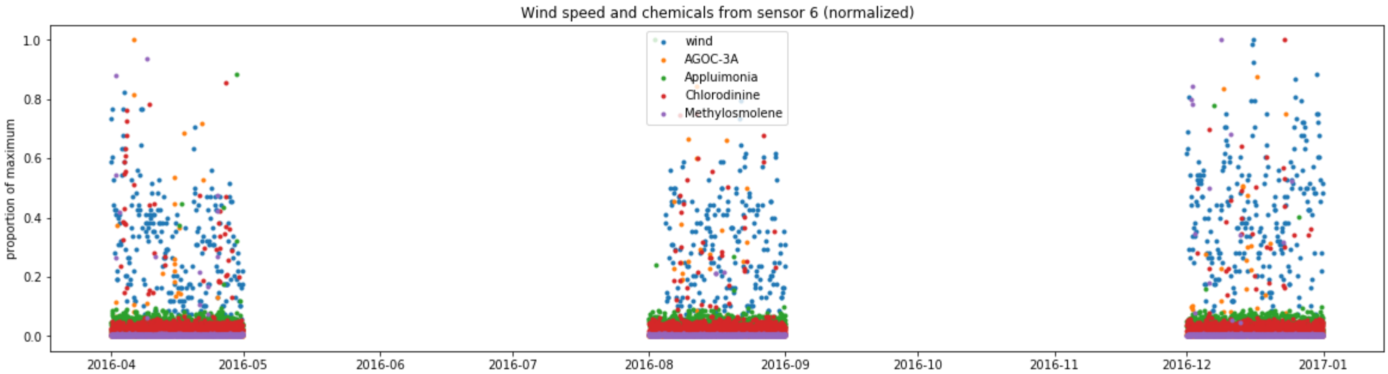


Figure 2: Hourly sensor readings (all chemicals)

2.1 - Data cleaning and transformation

The raw dataset required pre-processing so that a user could derive more meaningful analysis and insights from visualisations in the VA system. Cleaning, processing and transformation of the raw dataset included:

Feature separation and conversion. Temporal categories (month, day, weekday, weekend, hour, hour range categories) were separated for ease of future analysis. Boolean flags were created to indicate possible quality issues (missing sensor data, missing wind data, duplicate data) whilst distances were converted to kilometers and wind speed to kilometres per hour.

Data corrections. To output accurate analysis and visualisation some cleansing was required. Duplicate readings for AGOC-3A were observed to correspond to missing readings for Methylosmolene. The duplicates were flagged as having ambiguous chemical labelings.

The visual impact of large outliers were denoised by applying a square root or log transform to the reading values used for the time lapse scientific visualisation that was later implemented in advanced visualisations in Unity.

Wind data augmentation. Linear interpolation was applied to the wind velocity data set, which was provided at three-hourly frequencies to simulate 'missing' wind velocities on an hourly time scale. This enabled alignment with sensor data, which was reported at hourly frequencies. Interpolation required rotation of the wind vector in a direction (clockwise or anticlockwise) that results in the smallest hourly directional change (e.g. a rotation from 10 to 340 in north-referenced azimuth format is best resolved by 10 degrees per hour in the anticlockwise direction, rather than 110 degrees per hour in the clockwise direction). Linear interpolation of wind velocities is equivalent to assuming piecewise constant wind acceleration vectors every three hours.

3. DESIGN

3.1 Analysis

With cleaned and de-duplicated data the team began an iterative design process to determine, assess and implement design and analytical features:

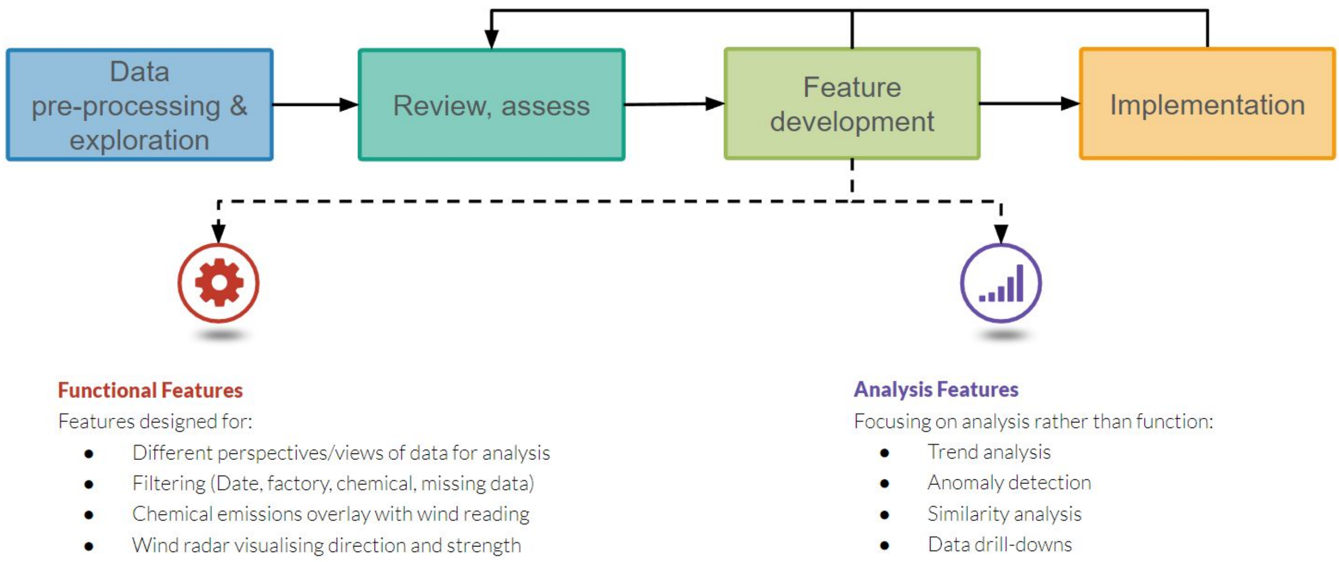


Figure 3: Iterative design process

The below section provides low level detail for each of the functions that were implemented into VA system.

3.1.2 Implemented features

The set of planned analytical and functional features in implementation of our VA system are based on common and effective approaches observed among winning VAST entries for each question combined with approaches constructed by the group.

Prior to implementation, data was cleaned, processed and enhanced with granular features to enable flexibility in design of various aspects of the VA system. As described in 2.1 - Data Corrections, this included the following steps:

- Preprocessing - The data was summarised by time to reveal rows that were missing and required to be added to dataset. This allowed missing data to be correctly represented and provide the foundations for more thorough analysis of sensor behaviour.
- Granular feature separation - Data was parsed and enhanced into additional, separate features including: (1) temporal criteria (month, calendar dates, days of week, hours of day) and (2) sensor criteria (subgroups of sensors) or chemicals of interest (selection of chemicals).

1. Python notebooks were used for:
    - [data preparation]
        - initial data exploration and cleaning
        - Interpolation of wind data
        - data matching and joining
    - [static visualizations]
        - diagrams
        - frames for animation

Exploratory Python analysis and prototyping was used to answer Questions 1 and 3 in the challenge. Question 1 required diagnosis of sensor behaviour, missing data and detection of sensor behavioural anomalies. Question 3 required attribution of chemical emissions to responsible factories and discover emission operating patterns.

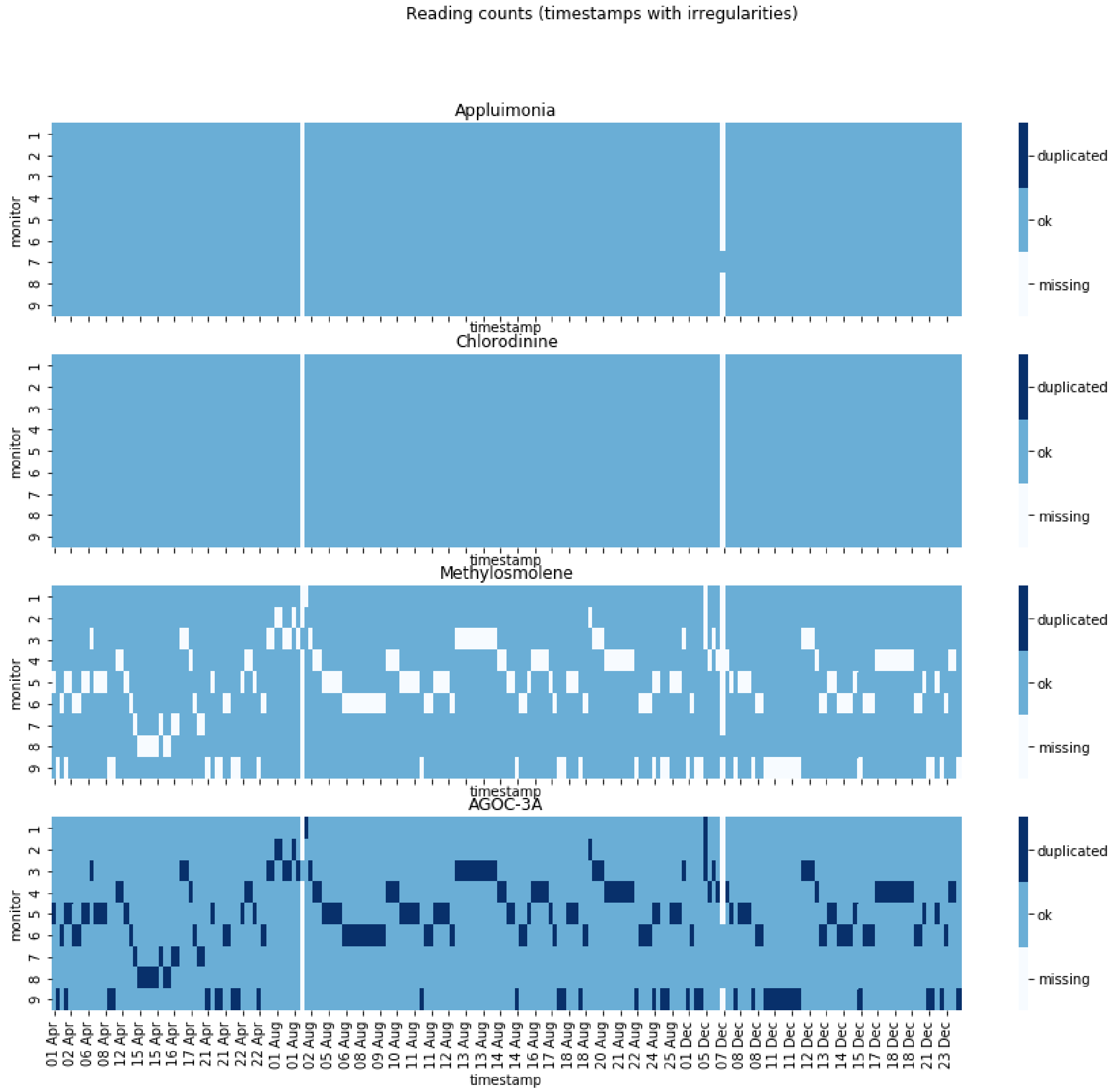
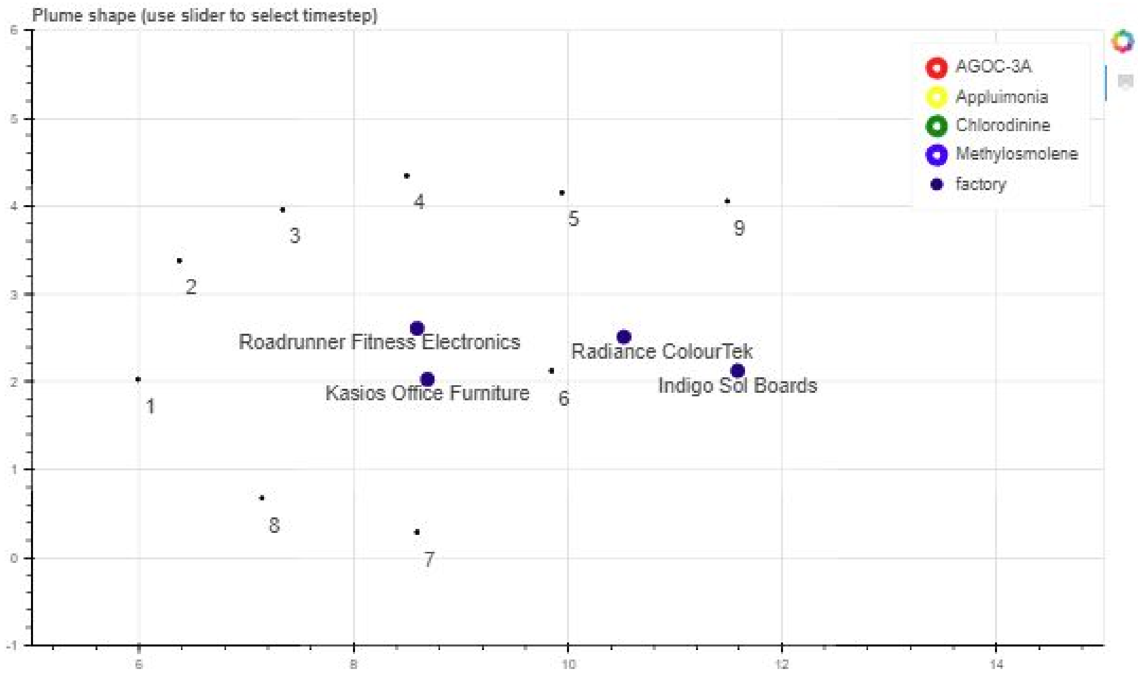


Figure 4: Exploratory analysis of missing and duplicate data (Python)




```
In [7]: def play_animation(sens_1=True, sens_2=True, sens_3=True, sens_4=True, sens_5=True, sens_6=True, sens_7=True, sens_8=True, sens_
        9=True, AGOC=True, Appluimonia=True, Chlorodinine=True, Methylosmolene=True, save_frames=False, start_snapshot=0,
        stop_snapshot=len(adj_sensor)):
            kwargs = locals()
            del kwargs['start_snapshot']
            del kwargs['stop_snapshot']
            for snapshot in range(start_snapshot, stop_snapshot):
                update_plumes(snapshot=snapshot, **kwargs)
            print('done')

        #play_animation(save_frames=True)  # WARNING: this takes a long time
```

Figure 5: Exploratory prototype for time lapse scientific visualisation (Python)

Analytic and visual functionality prototyped in Python was deployed across four separate user interfaces in development: (1) Sensor Diagnostics dashboard (via Tableau), (2) Sensor Heat Maps dashboard (via Tableau), (3) Dynamic spatio-temporal polar plots (via D3.js) and (4) Immersive scientific visualisation (via Unity).

2. The Sensor Diagnostics dashboard emphasizes anomaly detection for sensor and chemical combinations, while allowing the user to explore settings that minimise their visual impacts. The screen shows emission concentrations detected by each sensor for each chemical.

This was used to answer Question 1 and 2 by diagnosing sensor behaviour, missing data and detect behavioural sensor anomalies.

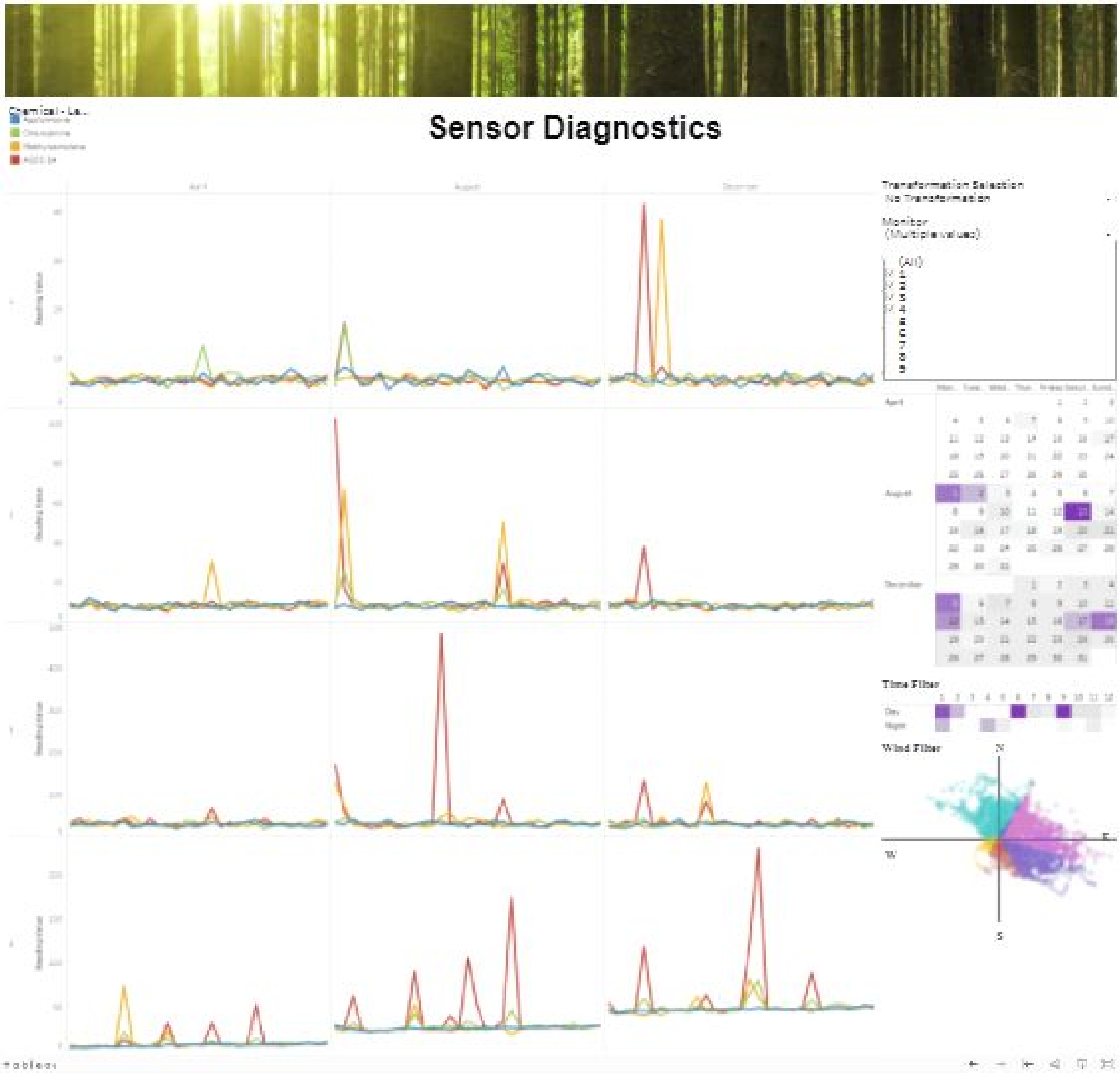


Figure 6: Sensor diagnostics dashboard (Tableau)

(Link: https://public.tableau.com/profile/cameron.wasilewsky5814#!/vizhome/GaiaTableauVisualisation-SensorAnalysis/Gaia?publish=yes)

Analytic functionality of this dashboard was enhanced by:

- Enabling transformation and display of sensor reading data. Transformation options include log and square root to denoise visual impact of outliers.

Visual functionality was further enhanced with filtering capabilities to:

- Filter by dataset: monitor sensor and chemical time series to display.
- Filter by observations: calendar day, time and wind velocity.
- Visual enhancements: heatmaps embedded into the calendar filter to guide user selection.
- Dynamic tooltips displaying data readings and statistics.

3. The [Sensor Heat Maps dashboard](#) assists a user to reveal both seasonal patterns and longer term trends of interest. The Chemicals Intensity and Sensor Intensity heat maps display average concentrations for a given day and hour in a month. The Wind Intensity heatmap (bottom left) displays hourly wind speed patterns , while the Missing Sensors heatmap indicate days with missing chemical readings or wind velocity data.

This was used for answering Question 2, for detecting patterns of chemical release in the data.

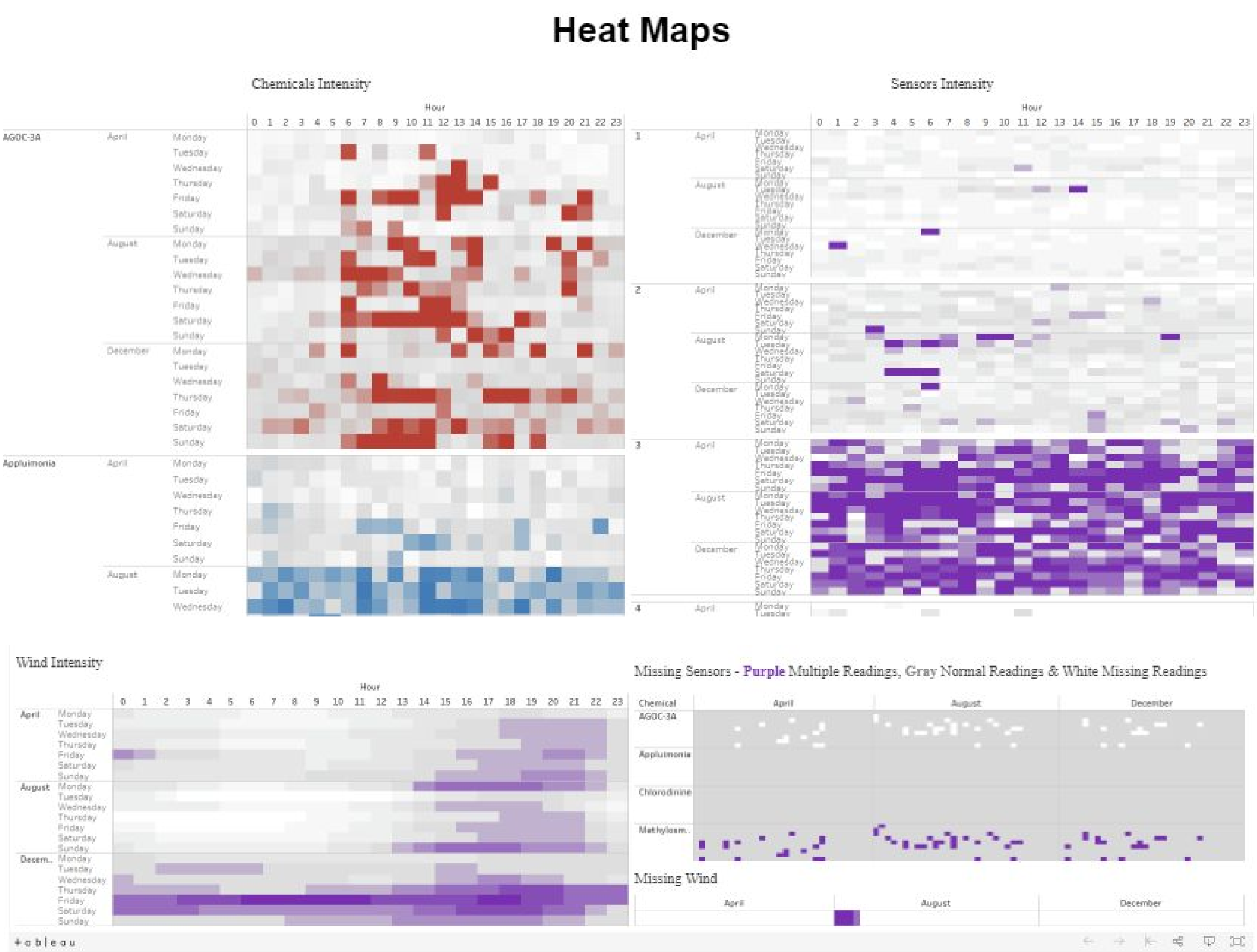


Figure 7: Sensor chemical heat maps with advanced filtering (Tableau)

(Link: https://public.tableau.com/profile/cameron.wasilewsky5814#!/vizhome/GaiaTableauVisualisation-HeatMap/HeatMaps?publish=yes)

Visual functionality was further enhanced with dynamic tooltips displaying data readings and statistics.

4. The Dynamic spatial-temporal polar plots assist a user to attribute responsibility of certain chemical emissions detected by the sensor array to a single factory. The emissions data are represented by polar coordinate plots for each sensor. The angle parameter represents wind direction at time of emission, while the distance from center represents the significance of the chemical reading. The Chemicals toggle on the bottom left enables selection of a single sensor to see a more detailed view. The Sensor toggle enables a user to see more detailed chemical detection statistics for a single sensor see Figure 7 below.

This was used for answering Questions 2 and 3. Question 2 was regarding detection of patterns of chemical release in the data. Question 3 required attribution of chemical emissions to responsible factories and discover emission operating patterns.

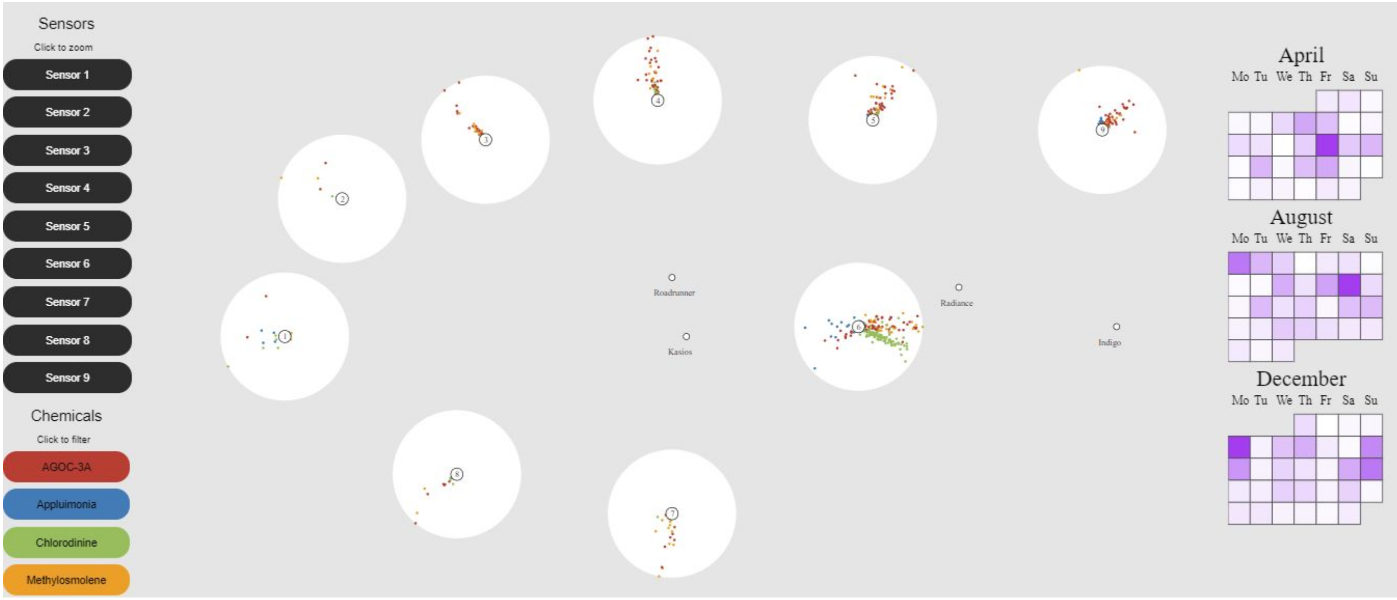


Figure 8: Dynamic spatial-temporal polar plots (d3.js)

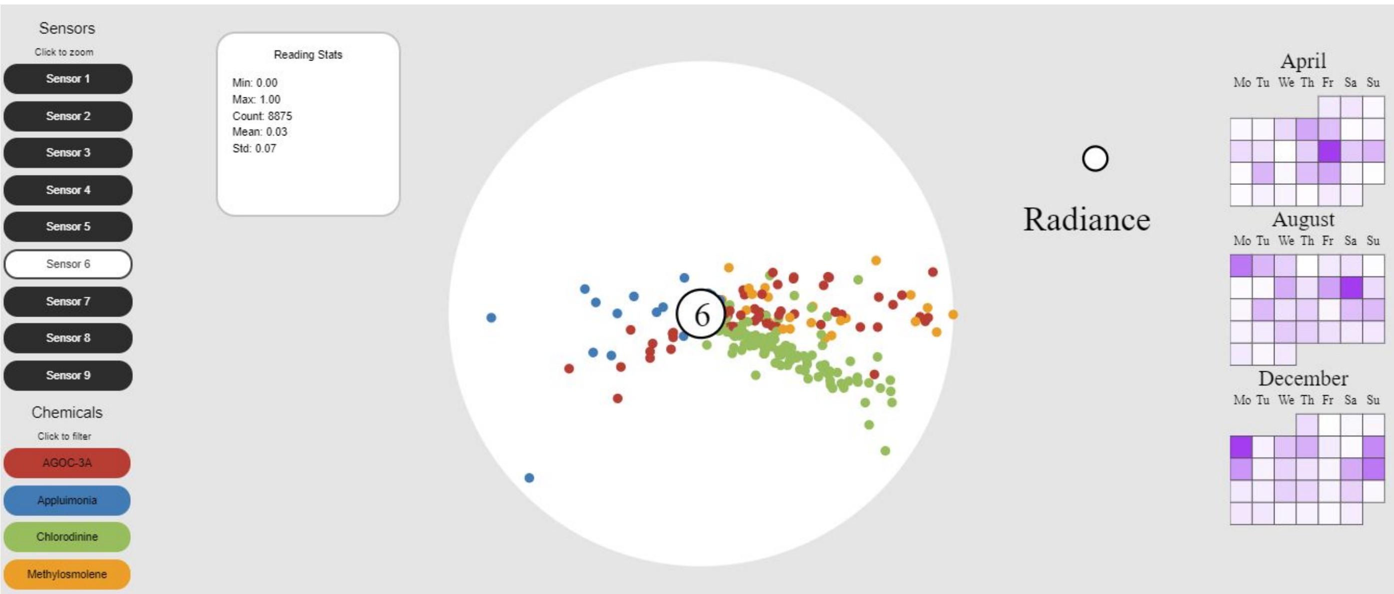


Figure 9: Single sensor view (d3.js)

(Link: http://www.gaiaanalytics.website/)

Visual functionality was enhanced by:

- Preloading normalised data: outliers were denoised at the sensor level as a default setting to provide better user experience.
- Tooltips: displaying sensor data cards with a five figure summary of chemical concentration statistics.
- Filter by chemical: enables user to focus on a single chemical to carry out attribution analysis.
- Filter by calendar day: with embedded heatmap into the calendar filter to guide user selection
- Visual enhancements: heatmaps embedded into the calendar filter to guide user selection.

5. The [Immersive scientific visualisation] was designed to enable a user to 'experience' the data, while blending non-intrusive elements of information visualisation concepts (sensor chemical detection heatmaps) to further augment the experience. This has been designed to be a user intuitive platform which assists users to form visceral hypotheses about the data in a realistic context. Toggles on the left side enable choice of dataset to experience: selection by month, factory, sensor and chemical. Toggles on the top right enable a user to adjust their experience.

Scientific visualisation elements include the implied physical modelling of pollutant gas dispersion in wind turbulence with an accurate geo-spatial representation. Information visualisation elements include the presence of concentric rings surrounding a sensor column. The width of the rings/bars increases with higher chemical emissions detected.

This was used for answering Question 3, which required attribution of chemical emissions to responsible factories and discover emission operating patterns.

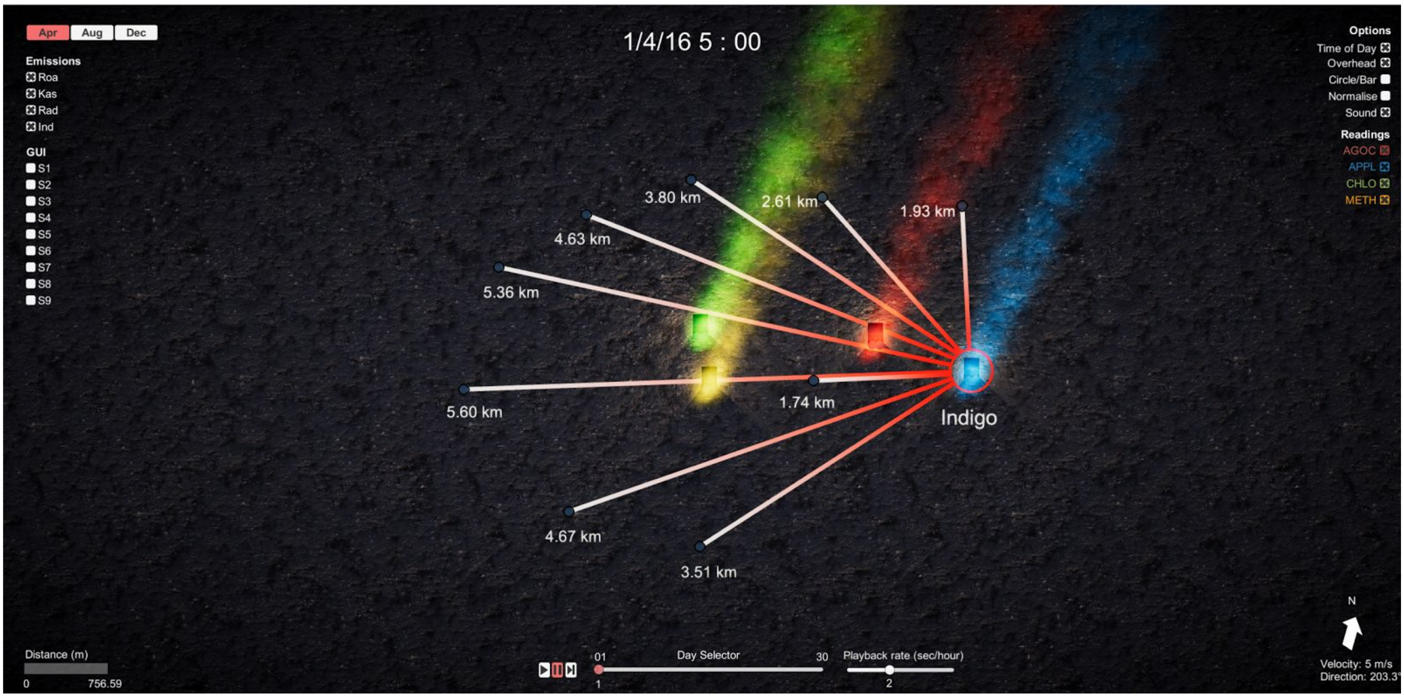


Figure 10: Immersive SciViz (Unity) - Hover UI/Dynamic lighting

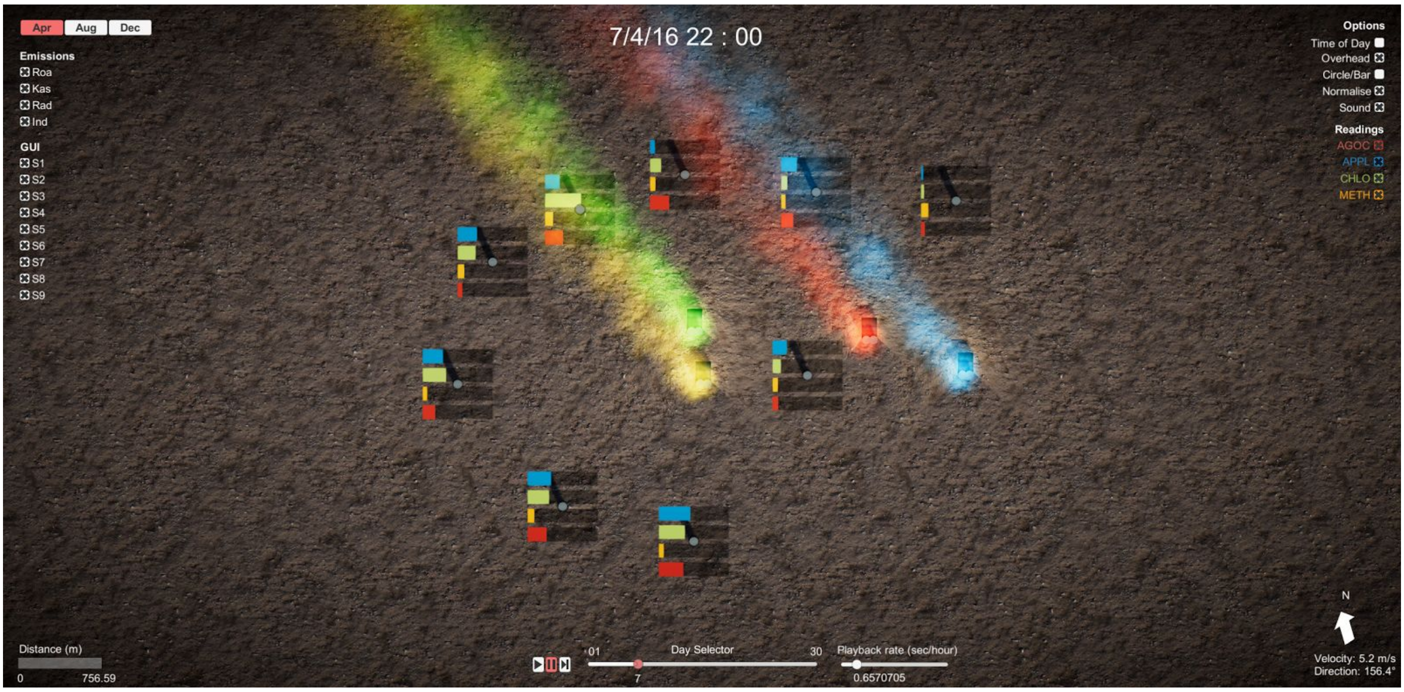


Figure 11: Immersive SciViz (Unity) - Overhead view/Normalised bar graphs

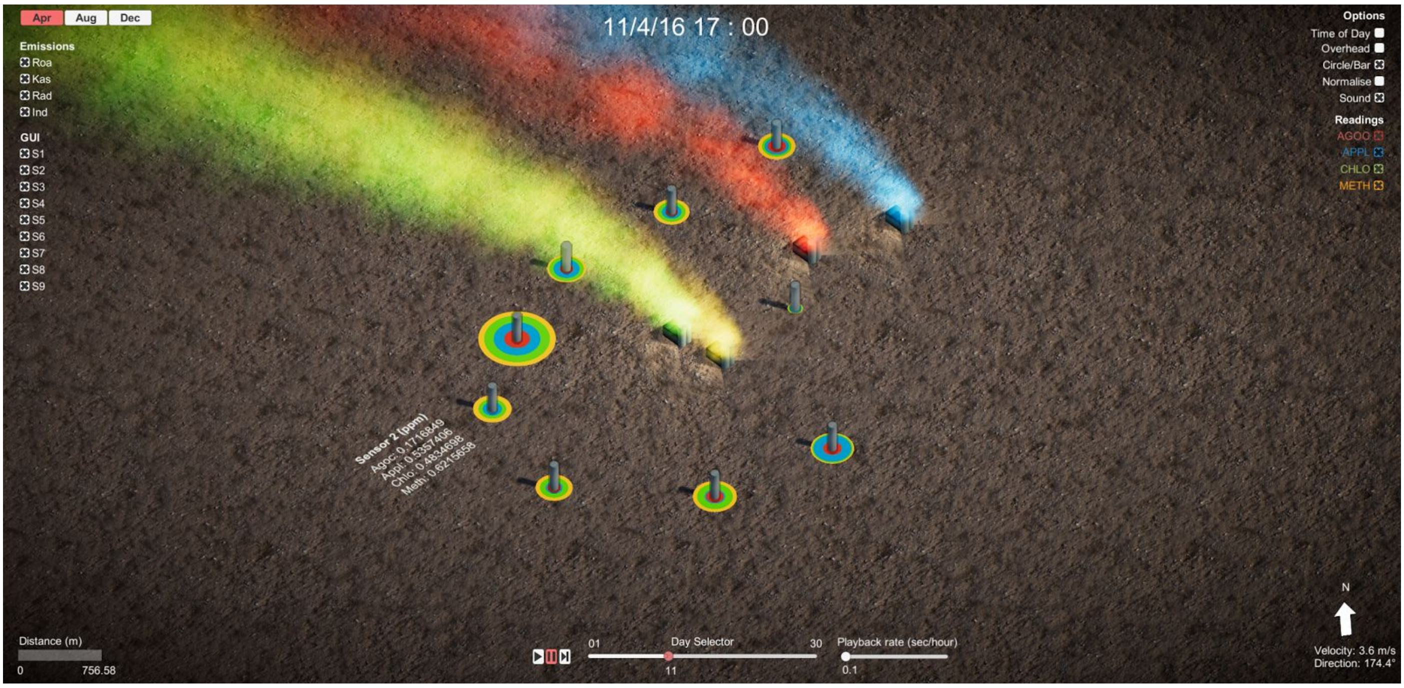


Figure 12: Immersive SciViz (Unity) - Radial graphs/Chemical readings

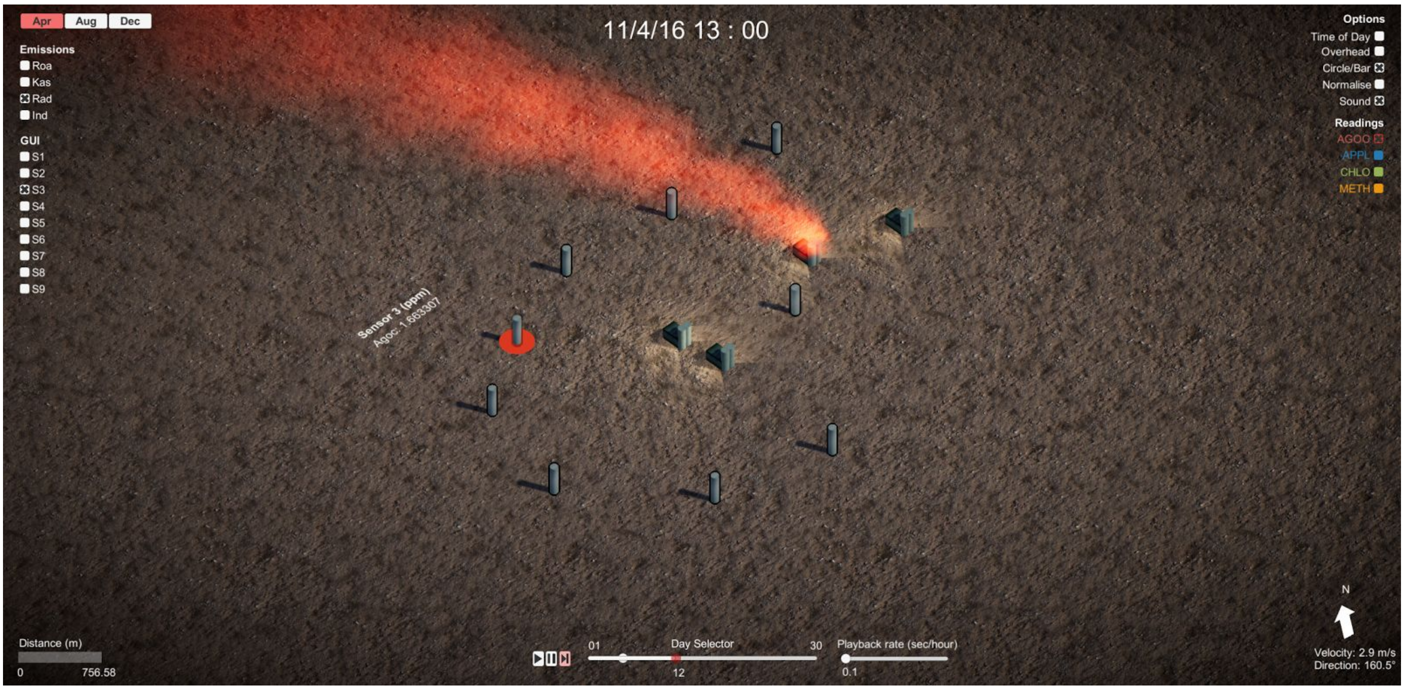


Figure 13: Immersive SciViz (Unity) - Isolated readings

(Link: http://www.gaiaanalytics.website/Unity/ or download from:https://drive.google.com/uc?export=download&id=0B7AOtUDisKwcSUswbE54SWJxZzA)

Analytic functionality of this visualisation was enhanced by enabling a user to:

- Denoise the display, but filtering only for sensors and chemicals of interest to experience.
- Vary the playback speed, to quickly decide on time periods of interest.
- Step forward hour by hour, to consider possible explanations for detected sensor changes.
- Select display experience of normalised or raw data for each sensor.

Visual functionality was further enhanced with capabilities to:

- Pan and zoom, to enable users to closely inspect geo-spatial patterns.
- Dynamic tooltips for distance calculations and sensor triangulation when hovering over a factory.
- Simulate and select modelled gas plumes for individual or multiple factories.
- Experience day and night cycles to potentially attribute factory responsibility for emission patterns.

4. IMPLEMENTATION

4.1 Platform and Tools

The 'Gaia' visual analytics system was hosted on the public cloud provider, Amazon Web Services. WIX was in turn used to blend the visualisations created in the system with a web interface. Gaia was designed to provide accessible tools for new users to explore, analyse and visualise the data. Gaia is publically available and accessible at: https://www.gaiaanalytics.com/

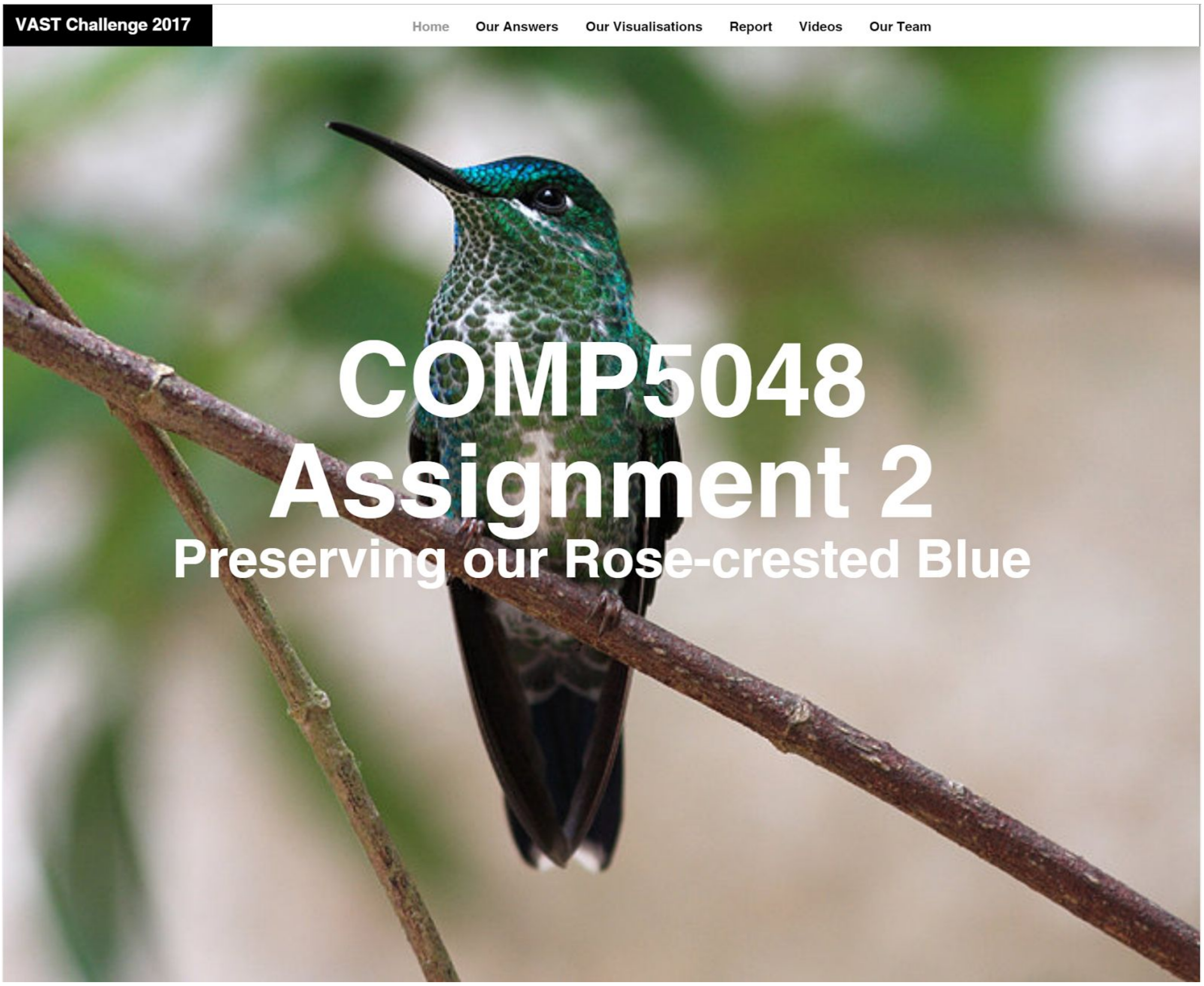


Figure 14: Gaia - Front Page

For the reasons outlined below, the project team agreed on the following implementation tools to deliver the range of VA system functionality described in Section 3:

- **HTML/CSS/JS - Front end for VA system:**
  - Web development stack allowed for flexibility in implementing the proposed interactive visualisation system whilst providing agility for further iterations of development and enhancement.
  - D3 blocks to reinforce desired narrative of analysis
  - Javascript to add interaction and transition elements to the system. Examples of implementation may include on-click events to create buttons and change data; hover events to provide additional info as necessary.
  - CSS ensures that the elements are placed, sized and styled appropriately.

- **Tableau:**
    - Detailed dashboards and visualisations for lower-level analysis and visualisation
    - Powerful way to display data with respect to trends and patterns for chemicals and sensors
    - Exploratory analysis enabled through interactivity and filtering
    - Complex filters across visualisations emphasising cause and connection between the data
- **Python - Data preparation, animation frames and static images:**
    - Detect missing sensor data and add to dataset with flag to enable it to be visualised
    - Clean and create wind data to match the increased frequency of sensor data
    - Used to merge datasets
    - Static visualizations and images of frames for animation
- **Unity/C# - Advanced visualisation of the factories release of chemicals:**
    - Unity enabled advanced, interactive 3D simulations for the emissions from each factory
    - Interactive control of timeline and playback rates
    - Multiple perspectives of the factories and surrounding sensors
    - Full capability to zoom and pan around environment
    - Light cycle for more noticable time of day
    - Switch between radial and bar graphs of current sensor readings
    - Turn on or off normalisation of sensor data
    - Playback controls to play, pause or step through the simulation
    - Selective display of factory emissions, graphs of sensor readings and individual chemical readings
    - Dynamic distance scales with perspective and zoom
    - Hover UI for descriptions and easier navigation
    - Wind vector showing current direction and velocity
- **Utilities:**
    - ffmpeg: Combining animation frames into videos

4.2 Development stack

The figure below illustrates all components of the VA system, including the necessary data flow, layers and intended visual elements for output:

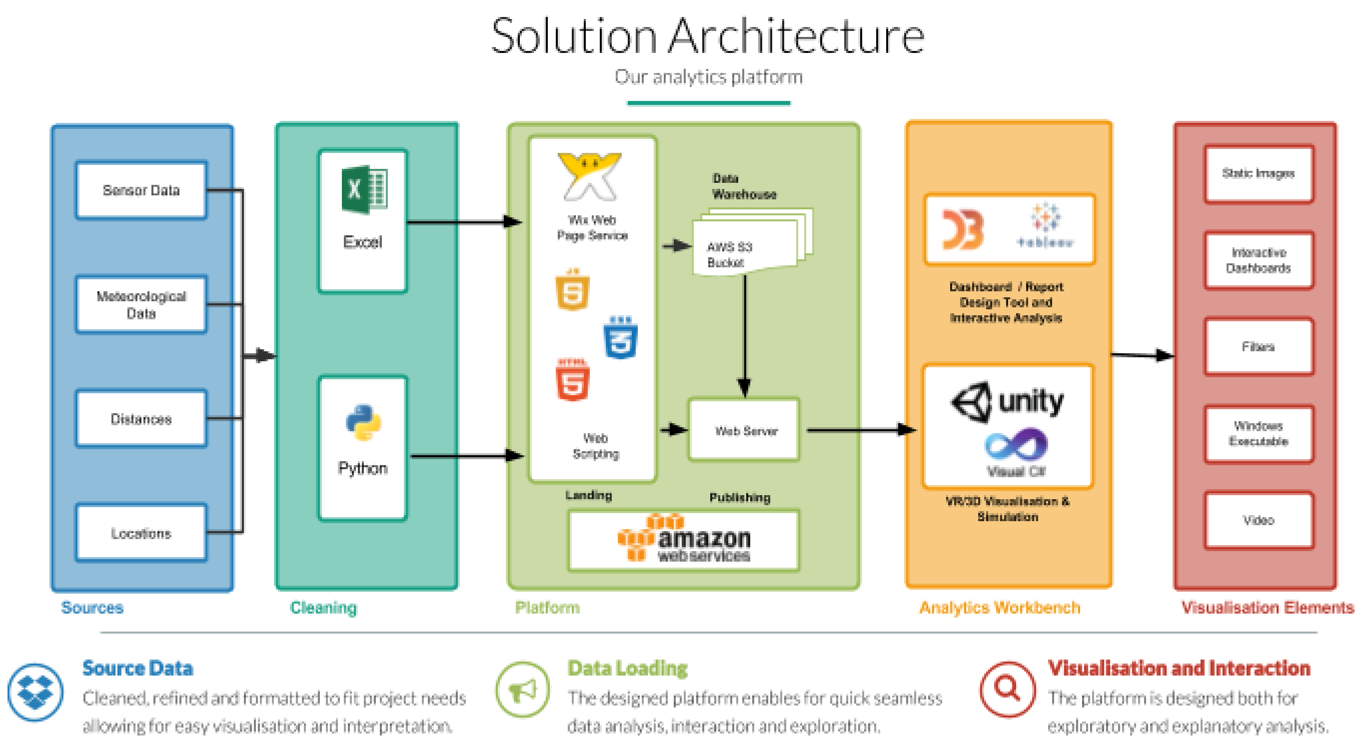


Figure 15: Development stack

5. EVALUATION

5.1 Question 1

Characterize the sensors' performance and operation. Are they all working properly at all times? Can you detect any unexpected behaviors of the sensors through analyzing the readings they capture?

Results:

<u>Missing and duplicate data</u>

During initial data cleansing on Python, it was discovered that there were several instances of missing and duplicate entries of sensor chemical readings. A temporal heatmap across all sensors and chemicals was constructed to look for potential patterns for when these occur.

The heatmap below shows the timing of when missing entries (white) and duplicate entries (blue) occur for all sensor monitors (vertical axis) over date/time (horizontal axis). It becomes readily apparent that duplicate readings for AGOC-3A correspond to missing readings for Methoylosmolene suggesting that missing Methylosmolene readings were mislabelled as AGOC-3A. Further, there is a strong coincidence of missing data on the dates of 1 August and 6 December across sensors.

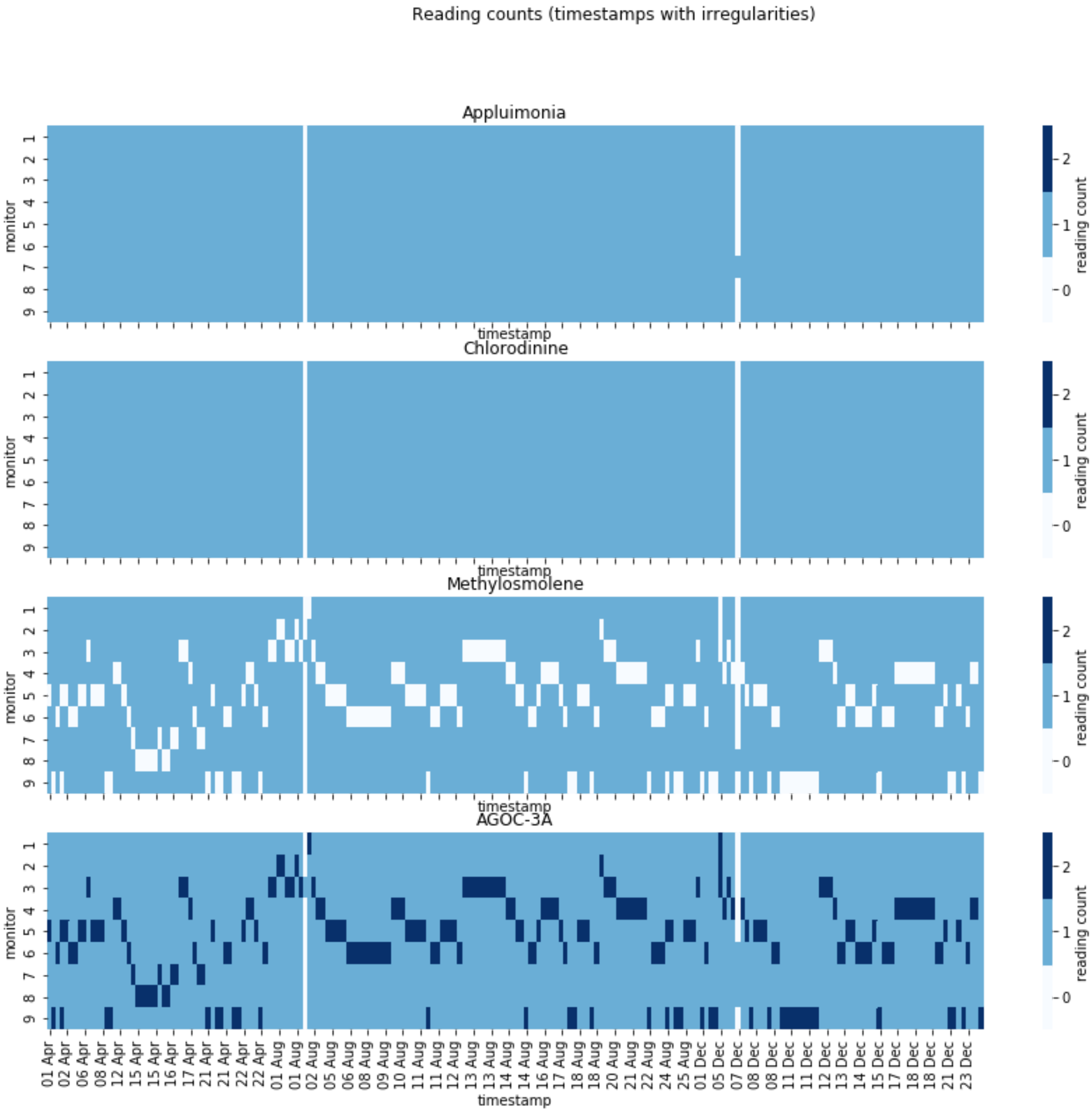


Figure 16: Missing and duplicate data

The above visualisation was also implemented in Tableau in Figure 17 below. This provides the user with dynamic tooltips to query instances of missing or duplicate sensor chemical concentration data. In addition, the Tableau implementation also shows a period of duplicative wind reading data along the bottom. This corresponds to the dates/times between 1-4 August.

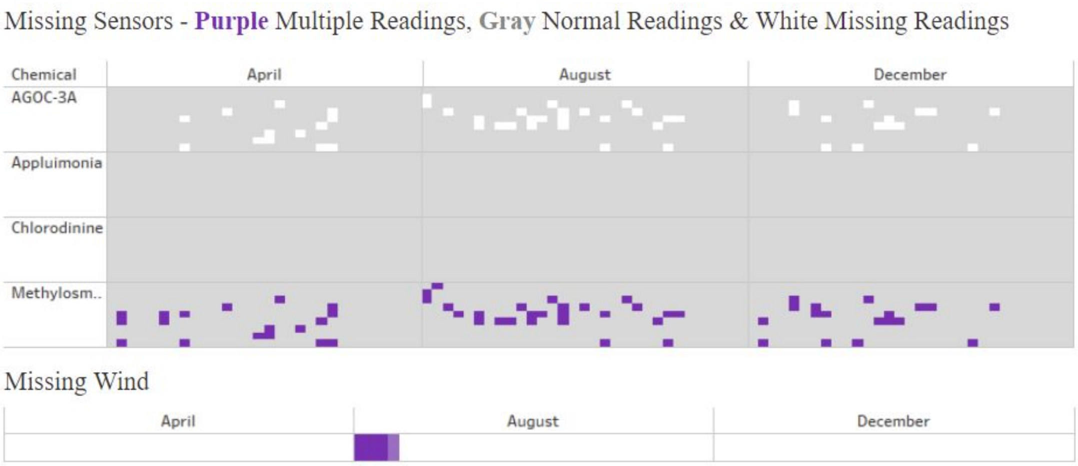


Figure 17: Missing sensor reading and wind data

<u>Diagnosing sensor behaviour</u>

When considering the performance of sensors, the [Sensor Diagnostics](#) screen was used to plot chemical reading time series across sensors. A sample is shown in the figure below. Sensor 4 is of particular interest and its readings compared to its nearest geographical neighbours (Sensors 3 and 5). Using a log transform of the data to denoise the visual impact of outliers, we see that Sensor 4 (in the middle row) records systematically increasing readings, suggesting malfunction. Sensor 3 does not exhibit similar trending tendencies, while Sensor 5 does to a lesser extent.

Figure 18: Implied malfunction of Sensor 4

Figure 19 below implies that Sensors 5 and 9 appear more sensitive than others (versus a control group, using Sensor 1). Sensor 1 is located west of all factories. Sensors 5 and 9, which are near neighbours located north to north-east of all factories appear to have increasingly noisy readings over time.

Sensor 1 (control) is represented by the top series of line graphs, while sensors 5 and 9 are the middle and bottom series of line graphs respectively.

Ignoring chemical AGOC-3A series (in red), the chart on the left hand side exhibits noisier readings across all sensors, as the range of readings for other chemicals grows progressively over the three months. Sensors 5 and 9 continue to exhibit increasing noise even when selecting instances when wind is moving away from both the control group (Sensor 1) and the group of interest (Sensors 5 and 9).

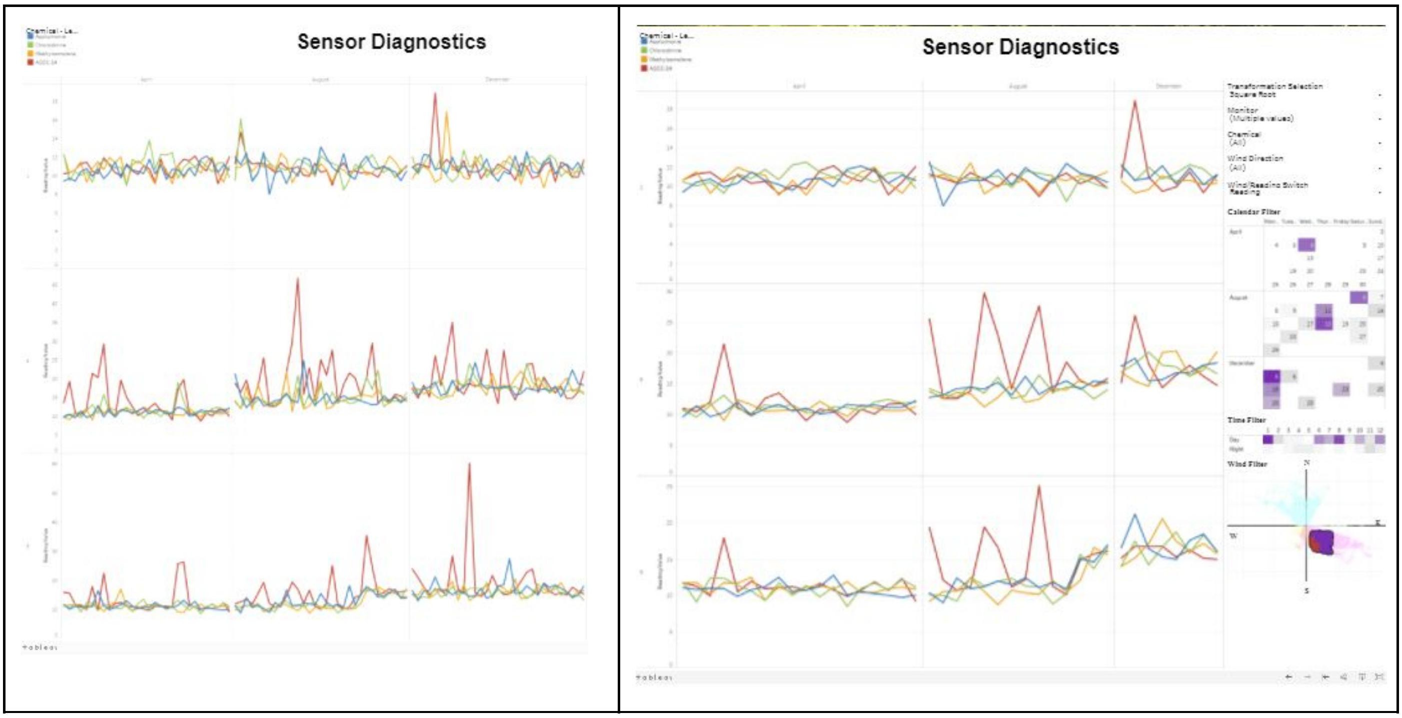


Figure 19: Increasingly noisy readings of sensors 5 and 9

Discussion:

The process for evaluating missing and duplicate sensor behaviour was straightforward via the use of heatmaps. Also, visually inspecting for trends using a combination of data transformations (log / square root) and filters provided exploratory flexibility.

Additional analytical features were identified in the initial design scope which would have further enhanced user functionality to yield similar insights.

These would include a wider range of data transformations (e.g. cumulative sum to assess rates of change for relatively small readings), immediate availability of statistics (e.g. standard deviation of readings data which has already been standardised for specific sensors) and other denoising options (e.g. winsorization of outliers and linear or seasonal trend removal)

5.2 Question 2

Which chemicals are being detected by the sensor group? What patterns of chemical releases do you see?

Results:

The table of images below are from Gaia's Sensor reading dashboard in Tableau and represent line graph readings for each chemical across the period of each of the three months. Options are available to de-noise through square root or logarithmic means:

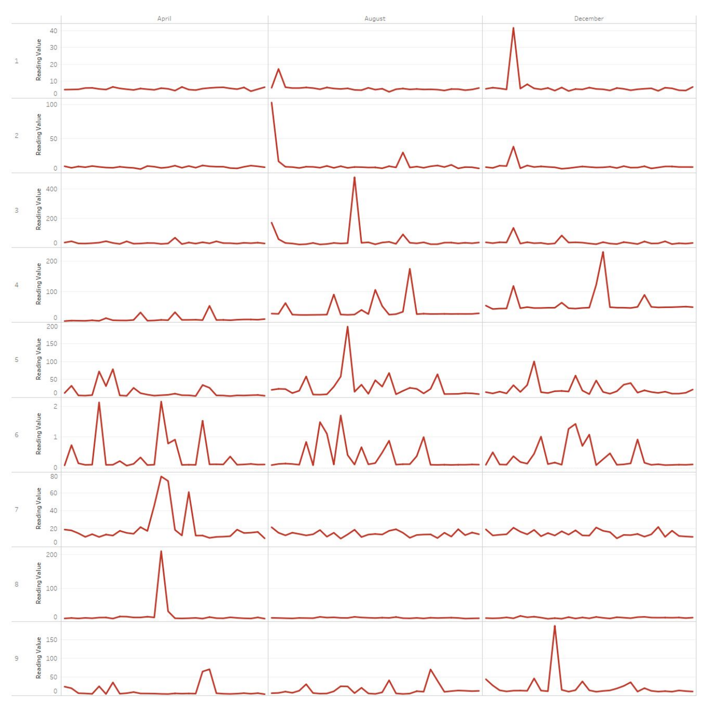
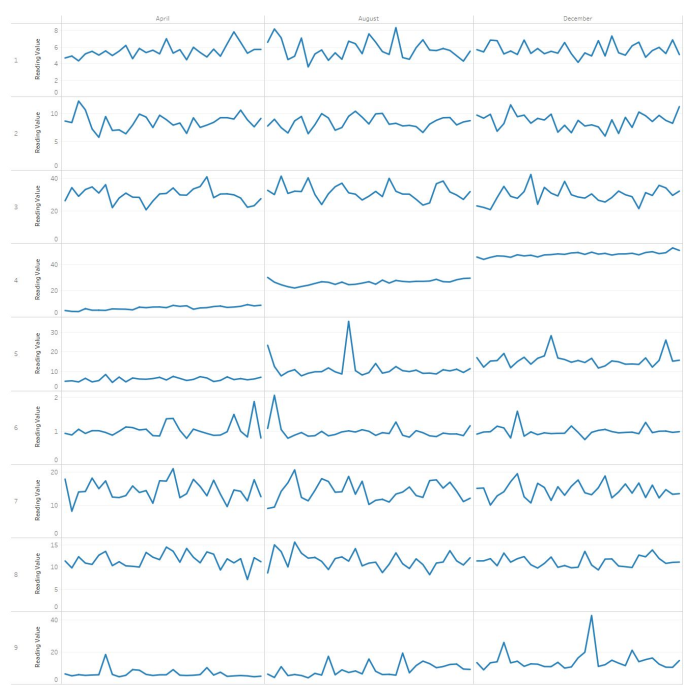
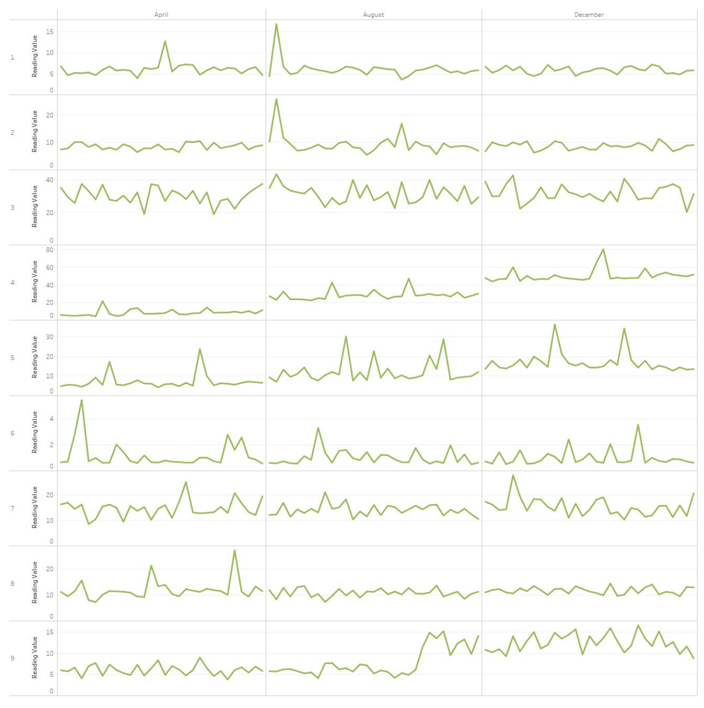
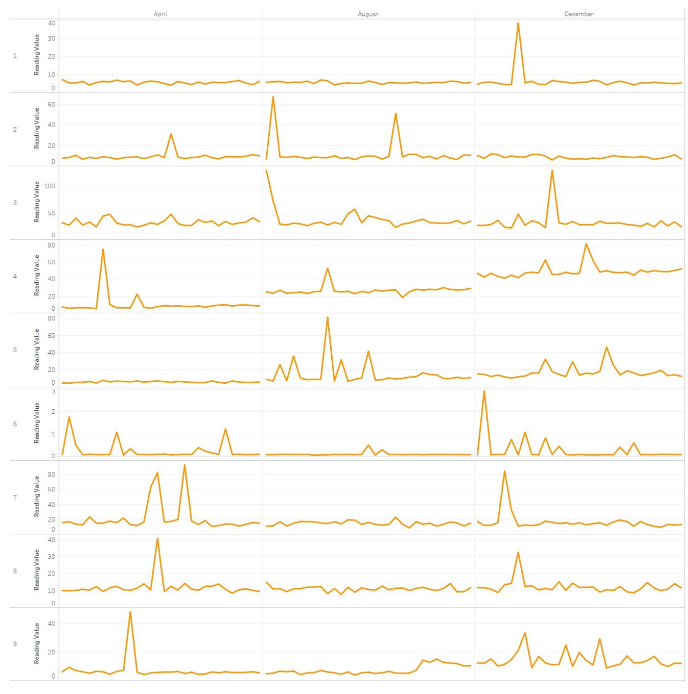
AGOC-3A	Appulomonia
	
Chlorodinine	Methylosomolene
	

Figure 20: Temporal chemical reading behaviour

<u>Detected chemicals</u>
By applying a threshold of 10 parts per million (ppm) and filtering to each chemical in turn on sensor reading Tableau dashboard:

- AGOC-3A is detected by sensors 3, 5 and 7 with consistency. Sensor 9 also indicate consistent but often the readings were below the minimal threshold.
- Appulomonia displays readings of greater than 10 ppm by sensors 3, 5, 7, 8 and 9.
- Cholorodinine is detected by sensors 3, 5, 7 and 8. From approximately the 21st of August Cholordinine readings increase in sensor 9 to cross the threshold of 10 ppm. These readings remain above 10ppm for the remainder of the year.
- Methylosomolene is detected by sensors 3, 5, 7, 8 and 9. The readings for Methylosomolene are perceived to be less noisy than others and on sensors 5 and 9 the readings increase with time throughout the year.

All chemicals are picked up by sensor 4, however, the veracity of the data is sub-optimal due to the aforementioned [“systematically increasing readings”](#) when compared to other sensor readings.

<u>Patterns of chemical release</u>
The “Chemicals” heatmap (bottom left of the [heatmaps](#) page) show that from a temporal perspective:

- There were greater emissions of Appluimonia and Chlorodinine in December than in the other months:

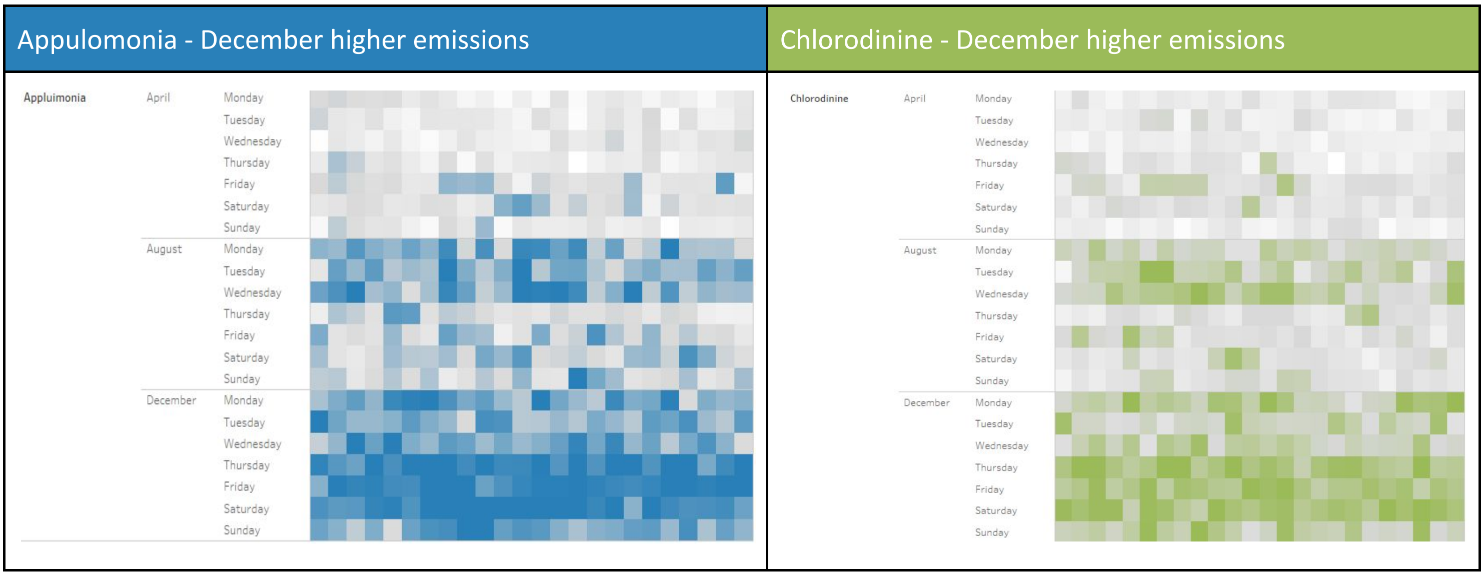


Figure 21: December - Temporal heatmap of Appulomonia and Cholordinine

In addition, returning to the Time Filter heat maps on the [sensor diagnostic](#) page in Tableau:

- AGOC-3A is shown to have consistent detection throughout approximately the business hours of the day, beginning from 6am until recordings finish 2pm.
- Appulomonia illustrates a relatively sporadic release schedule throughout the day. Often starting at 7am with recordings until 4/5am.
- Chlorodinine illustrates a behaviour of release throughout the entire day. From very early morning and through the night with no explicit true day/night pattern as demonstrated by AGOC-3A or Methylosomolene.
- Methylosomolene is shown to have consistent detection overnight. Specifically between the hours of 10pm and 5am.

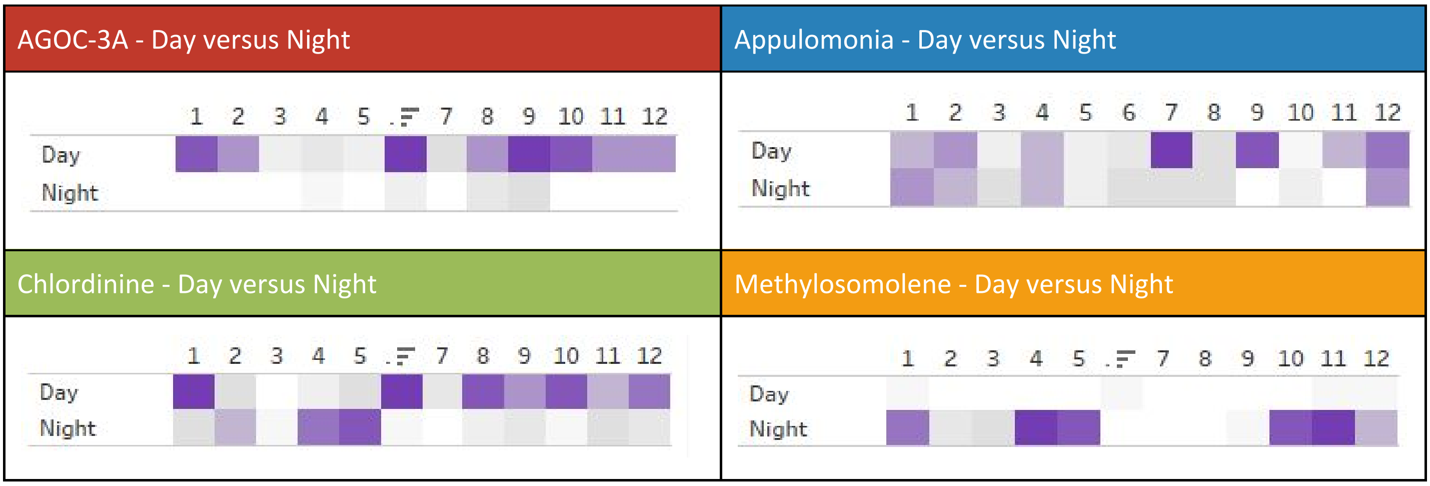


Figure 22: Temporal heatmap - Day/Night

Discussion:

The realisation and discovery of chemical patterns quickly and succinctly for question 2 was predominantly powered by the Tableau temporal heatmaps. Through visual analysis it is relatively easy to see the pattern of releases throughout the day/night and time of year. Appulomonia and Chlorodinine show distinctive increases in reading throughout the year, specifically in December. In tandem with the increase in sensor reading of these chemicals is an increase in wind strength in December. Whilst no exact details of the chemical compound weight/structure is provided it is hypothesised that each chemical does directly advect to the strength and direction of the wind.

The sensor readings when filtered and with the correct data transformation applied give the user a sensor of the behavior of the chemical releases throughout each of the three months. Interestingly sensor 6 has the weakest readings of all 9 of the sensors, yet it is in fact often the closest sensor. It is hypothesised that whilst the chemicals do advect to the wind that due to the height of the factories (370m) that sensor 6 is actually too close to provide accurate readings of the chemicals themselves.

The choice of colour for the heatmaps was changed later in the project to reflect a true change of saturation in one colour. It is suggested that this reinforces the insight the reader is able to gain from visual inspection.

One additional analytical feature was identified in the initial design scope which would have further enhanced user functionality to yield similar insights. This would include a trend line conveying the r^2 and pattern of sensor readings over a given period of time.

5.3 Question 3

Which factories are responsible for which chemical releases? Carefully describe how you determined this using all the data you have available. For the factories you identified, describe any observed patterns of operation revealed.

Results:

Identification of polluters

The VA system provides two primary ways to identify polluters:

- Summary across time: interactive sensor reading visualization
- Visualization through time: simulation and animation

The table below shows stills from the interactive sensor reading visualization switched to each chemical. This provides a view of the sensor readings throughout the study period and can be filtered by any combination of chemical and sensor. The coloured dots near each sensor location indicate chemical readings. The distance of the dot from the center is the amount of the chemical, and the angle indicates the wind direction. The calendars at the right indicate the distribution of chemical readings by date.

These images indicate that for each sensor, the largest readings for each chemical occurred when the wind was blowing away from a specific factory (specific to the chemical). The only exception to this is AGOC-3A which appears to come from two factories. The indicated emitting factories for each chemical are:

- AGOC-3A: Radiance Colour Tek and Kasios Office Furniture
- Applumonia: Indigo Sol Brands
- Chlorodinine: Roadrunner Fitness Electronics
- Methylosmolene: Kasios Office Furniture

Figure 23: Sensor readings by chemical

These conclusions are supported by the table below which shows incriminating stills from the sensor readings animation. This is an animated overhead view of the of the same area covered by the interactive sensor reading visualization showing chemical readings and wind movements as they occurred. The sensor readings are indicated by the concentric circles around each sensor (area corresponds to the reading amount). Movement of air from the factory locations is indicated by the grey streamlines.

An extra piece of information that comes from these snapshots is that the two chemicals emitted by Kasios Office Furniture are typically detected simultaneously. This would be consistent with the chemicals being released at the same time.

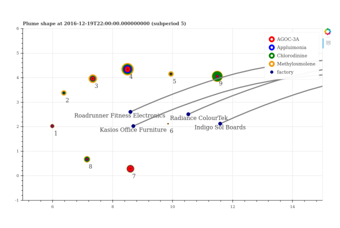
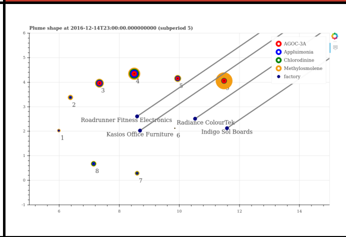
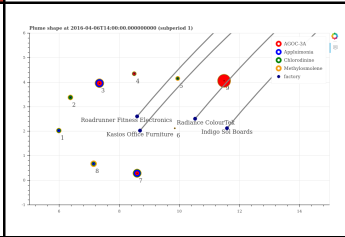
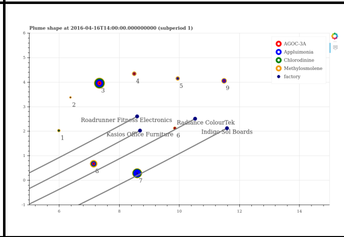
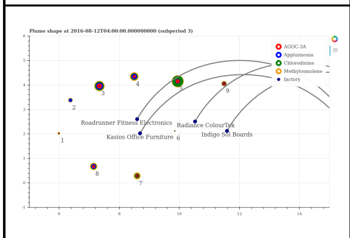
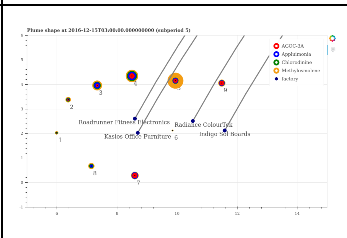
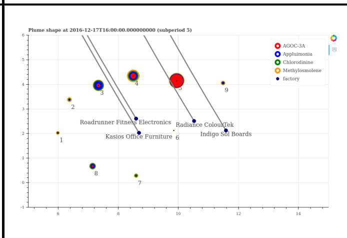
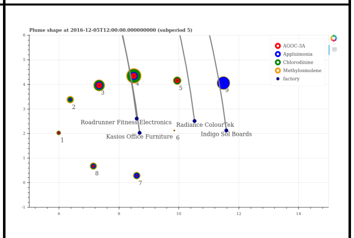
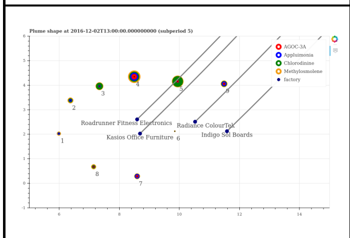
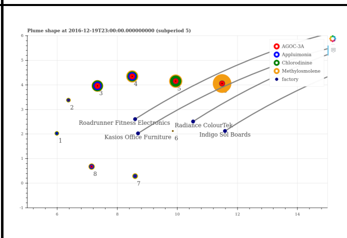
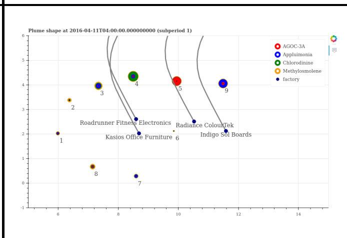
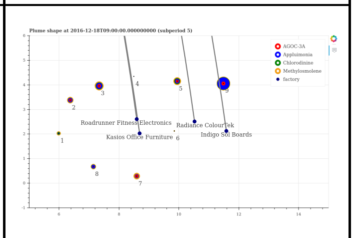
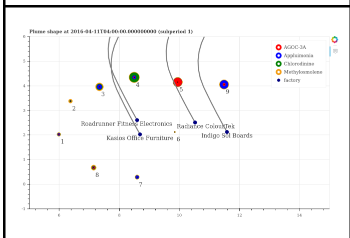
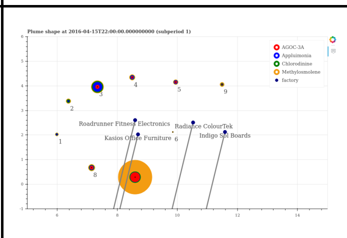
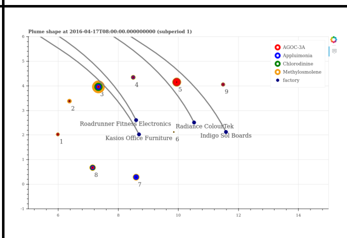
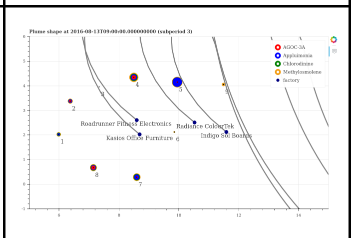
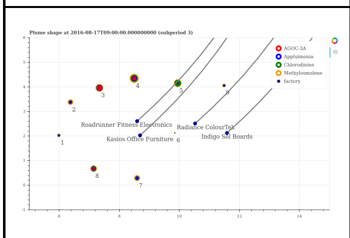
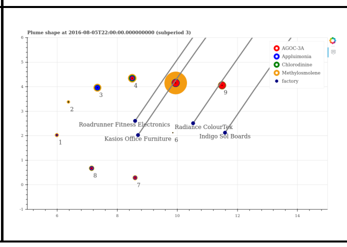
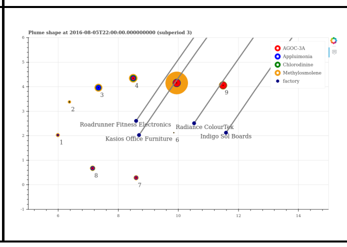
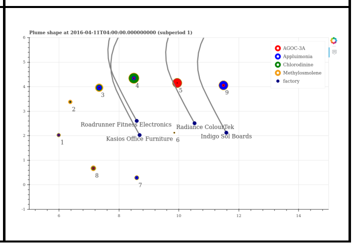
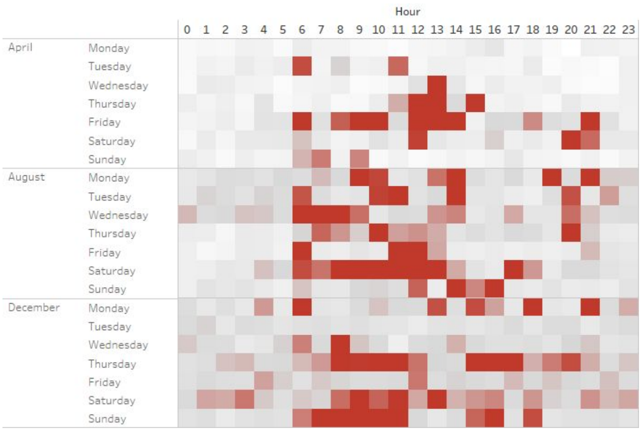
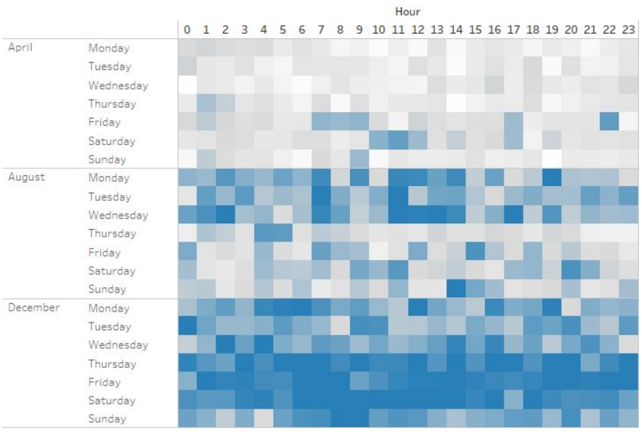
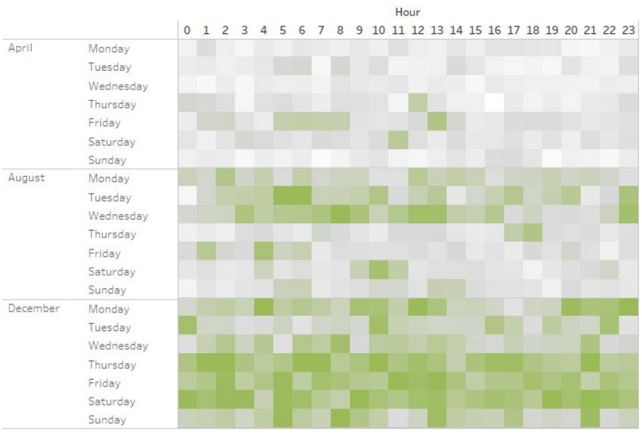
Roadrunner Fitness Electronics	Kasios Office Furniture	Radiance Colour Tek	Indigo Sol brands
Chlorodinine	Methylosmolene	AGOC-3A	Applumonia
	AGOC-3A		
			
			
			
			
			

Figure 24: Pollutants detected downwind of factories

<u>Operational patterns</u>

Patterns are apparent in the detection times of each of the chemicals. These are summarized in the table below.

Chemical	Detection time heatmap	Pattern in detection times
AGOC-3A		AGOC-3A is detected from 6am onwards. However, it is released by two different factories - Kasios Office Furniture and Radiance Colour Tek. The information from the animation and the snapshots above indicates that Kasios's emissions of AGOC-3A happen at the same time as its emissions of methylosmolene (which are mostly in the evening). Therefore, Radiance Colour Tek appears to emit AGOC-3A in the morning, and Kasios in the evening.
Applumonia		These emissions display an Increasing trend month-to-month. Also, applumonia is detected throughout the day and night. Possible explainations for this are: ● Applumonia is accumulating in the environment, or ● something (perhaps equipment or procedure) at Indigo Sol Brands is deteriorating, causing a steadily increasing emission rate.
Chlorodinine		Chlorodinine shows similar behaviour to applumonia (increasing detections and detection ultimately increasing to all hours of the day). This raises the same possibilities (accumulation in the environment or an increasing emission rate) that it did for applumonia.

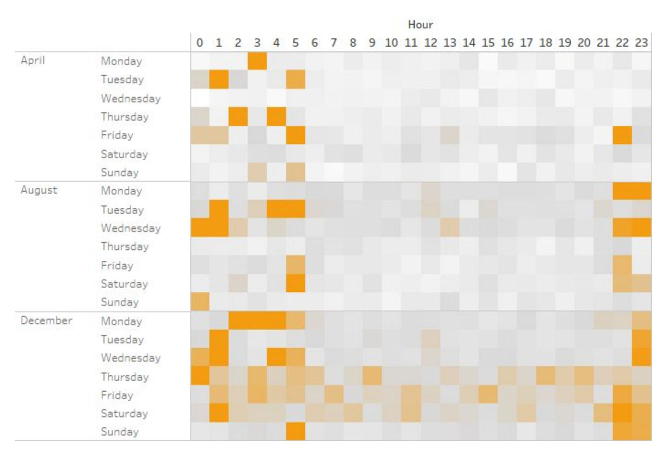
Methylosmolene		Methylosmolene is detected from the evening onwards, into the early morning. Kasios office furniture appears to be releasing it in the evenings. This might be a deliberate attempt at stealth, but it also might be due to the production routine (maybe a piece of equipment is cleaned in the evenings and releases methylosmolene then). Levels are also increasing indicating either acculation or increased emission

Figure 25: Chemical Operational patterns

Discussion

- AGOC-3A is released by Radiance Colour Tek in the mornings and by Kasios Office Furniture in the evenings.
- Applumonia and Chlorodinine are released by Indigo Sol Brands and Roadrunner Fitness Electronics respectively. They are detected throughout the day and their levels are increasing, suggesting either accumulation in the environment or increasing emissions.
- Methylosmolene is released by Kasios Office Furniture, mostly in the evenings, but in December it began being detected in appreciable quantities throughout the day.

5.3 HCI Evaluation

The VA system was designed to provide users with the functionality to visually explore the data and complete tasks corresponding to the three MC2 challenge questions outlined in Section 1.1. During the early stages of development, regular feedback was provided by the internal team to guide development of a more simple, aesthetically pleasing and effective VA environment.

In the following sections the DECIDE[8] framework illustrates how the evaluation framework was set up and deployed as well as presenting results from the survey and hence improvements made to the system.

To effectively evaluate the Gaia VA system, an appropriate framework was required. The group agreed on using the **DECIDE** framework to guide setting of evaluation goals and determine processes for seeking evaluation of potential design changes.

- **Determine the goals.**

The high level goals of the evaluation were to elicit quantitative and qualitative responses regarding the usability, quality and depth of the system/visualisations from a sample of analytic experts known to the members of the group. It is suggested that by choosing analytical experts a walk through/interactive session can then prepare the evaluators for the quantitative and qualitative parts of the evaluation.

Following completion of a full working VA system prototype, evaluation was sought from external testers for areas of improvement. 10 external testers with analytical experience volunteered to trial the VA system for 30 minutes each. This was after the interviewer had set the context of the interview and introduced each part of the VA system.

- **Explore the questions**

Testers were required to attempt to complete the three MC2 challenge questions by manipulating settings and views within the VA system. Following testing, testers completed a qualitative and quantitative survey to assess (1) the effectiveness of the VA system for solving each of the three questions; (2) ease of navigation and usability and (3) aesthetic design. A second section proposed two open ended questions regarding suggestions for further improvement and design as well as general feedback. The survey questions are contained in Appendix C - Survey Questions.

- **Choose the evaluation approach and methods.**

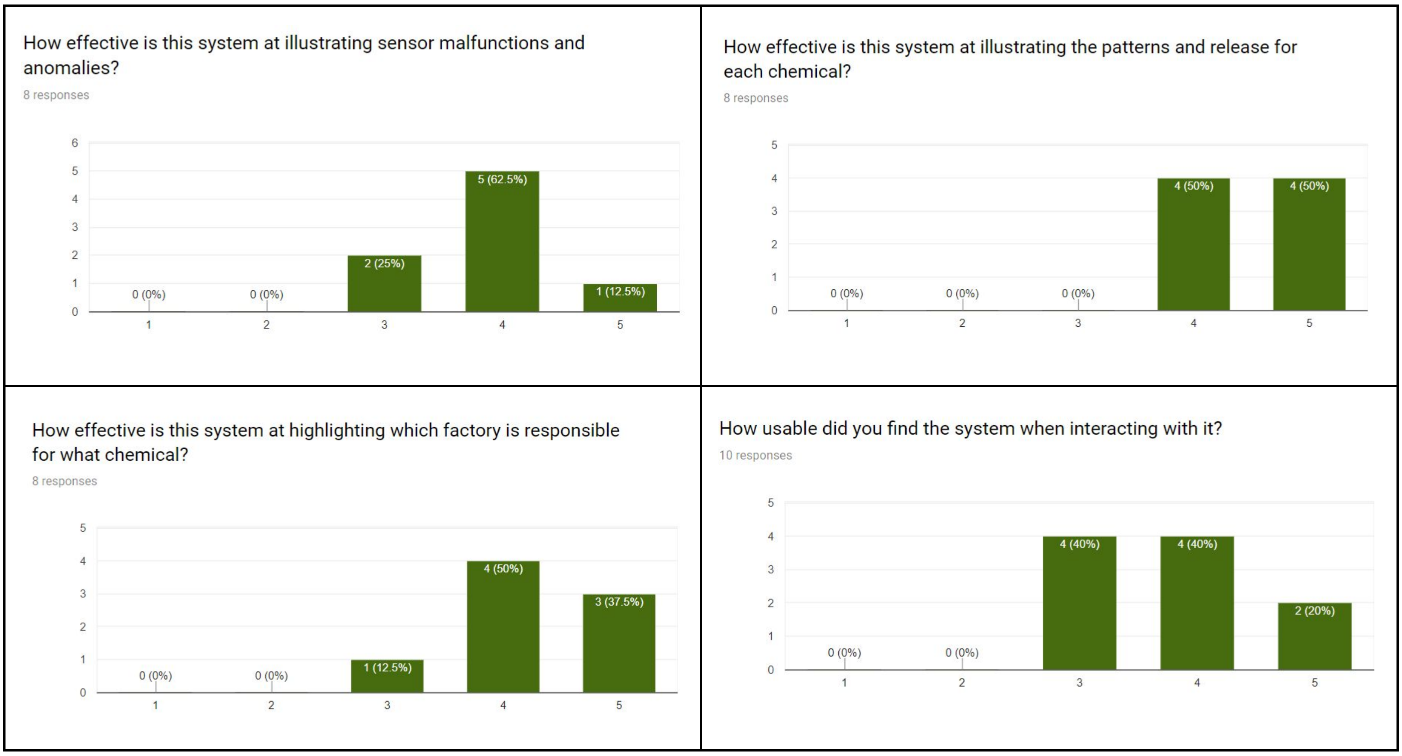
Each evaluator was introduced to the Gaia VA system over the course of 30 minutes to one hour and was handed control to interact with Gaia. After testing, the above survey was completed online by participants, which was followed up with an individual in person interview.

- **Identify the practical issues.**
  - Survey results needed to be secure. Users carrying out the survey had to have a legitimate Google email and be invited to fill out the survey.
  - Different approaches might be taken by different group members carrying out interviews/walk throughs. To mitigate this, prior to evaluations being carried out, the members of the group all agreed on a common strategy for delivery, approximate time frame for each evaluation and expectations for each person to conduct the evaluation.
  - TIme management. Each member of the group collecting results had a pre-defined schedule and time limit to adhere to that was agreed in person at team meeting (See minutes in Section 8.5).
- **Decide how to deal with the ethical issues.**

Due to the fictional nature of the data and relative low level of confidentiality of the data there were little ethical issues to deal with. However, it was decided to remove the usernames for each response to keep it objective and not personalised or open to any form of bias.

- **Evaluate, analyze, interpret and present the data**

As shown in Figure 24 below, the results for the quantitative part were on average positive, however, the results for 'How effective is this system at illustrating sensor malfunctions and anomalies' showed that there was room for improvement with 20% of responses giving 3/5 and only one user giving 5/5 for the system.



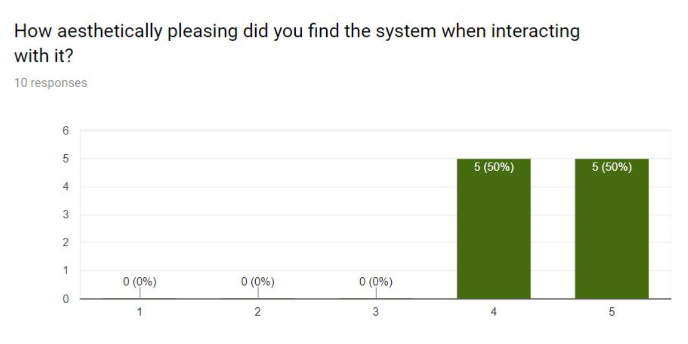


Figure 26: Quantitative survey question results

Table 4 in [APPENDIX C: Quantitative Results] suggests that the usability of the system as the area the evaluators found the least performant in the use of Gaia. Questions 2 and the aesthetics of the system were well received with a low standard deviation of responses and a very high average of 4.5/5. It must also be noted that questions 1 and 2 only have 8 of 10 answers due to earlier amendments by the project team.

APPENDIX C: Survey Questions & Results contains some example qualitative comments that were taken into consideration for the final stages of development of the VA system. As a whole, poor labelling and inconsistent/confusing colouring was consistently identified across respondents and survey questions. The project team found that access to respondents who were analytically minded but not involved in project delivery was of great benefit for further improvements to Gaia. Most of the respondent feedback was incorporated into design and development, which improved the quality of VA system implementation

6. CONCLUSIONS

6.1 Measures of Success

It is the belief of the project team that each of the aims in section 1.1 were successfully addressed through the design and implementation of the visual analytics system, Gaia. Specific visualisations were created to address the sensor performance and operation data as well as chemical pattern data in Tableau. Custom visualisations were built from scratch in D3 to answer question 3 regarding the factory responsibility over time, and Unity whilst not addressing any one particular question provides an excellent overview and ability to simulate all areas of data for all questions. In a way that many standard data visualisation applications cannot match.

The best measure of success was however deemed to be the results shown in the aforementioned HCI evaluation section. As per the evaluation section the results from each of the 10 users of the system shows that 50% of the users found the system to be 'beautiful' to visualise the data. See figure 27 below:

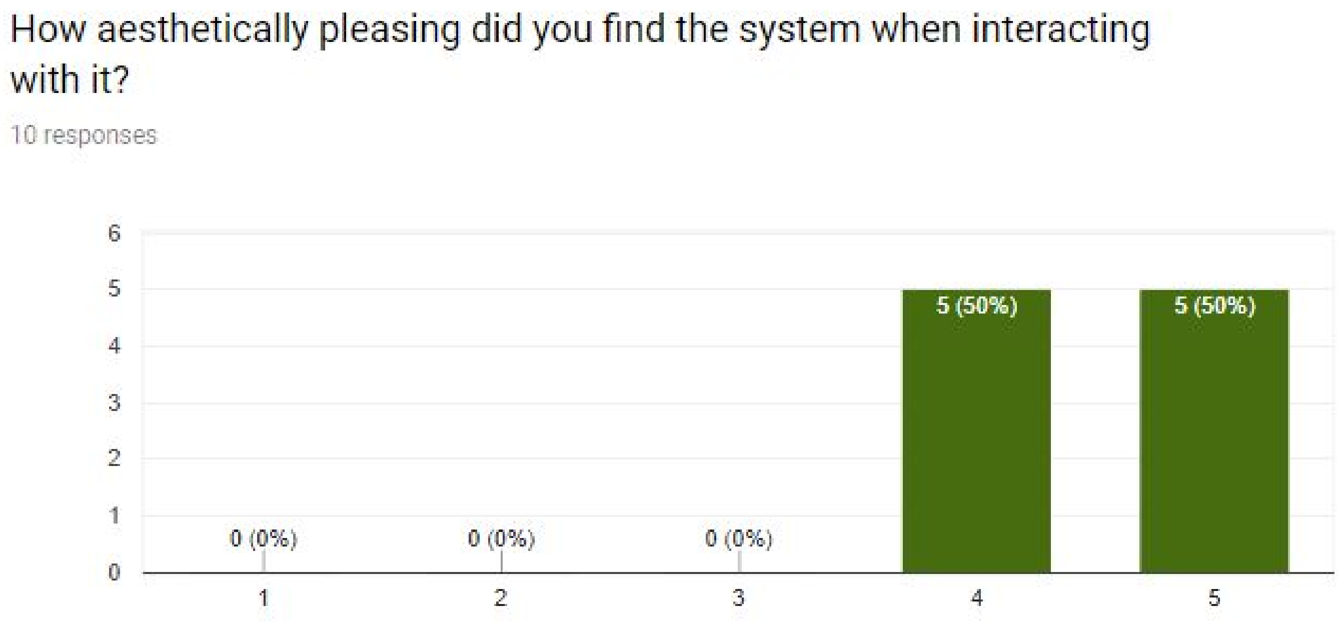


Figure 27: Users' aesthetic satisfaction rating

In addition an average score of 4.23 was received for the three quantitative questions for the mini challenge. Through the external user testing and the results that were received the project team identified and remedied the majority of the issues into the final product.

At the outset of the project the team took the approach of looking at the 'art of the possible' given the time limits. Whilst the team were not able to achieve all goals we set ourselves we were able to reach a considerable stretch goal in the use of Unity to display advanced visualisation of data. The team were able to blend the top features of winning systems from VAST 2017 to create the foundations of a visual analytics system we are all uniquely proud of.

6.2 Key Learnings

The Gaia team's implementation of a conceptual solution to an environmental data problem was successful for a number of reasons, which included:

- Leveraging the talents, diverse skillsets and innovative ideas across all team members

- Clarity in target state for the VA system, its functionality and user experience
- Effective project management, resource allocation and coordination

To establish a clear target state, a review of previous work conducted by other MC2 challenge winners and real world applications of similar technology was extremely valuable for getting the team up the 'design learning curve'. This enabled rapid ideation of concepts and visualisations that would be more likely to work in practice and had the additional benefit of eliminating unnecessary testing cycles. Two key insights from this exercise included (1) maximising data feature creation at the outset and at the back end to build a VA system that could solve 'general classes of problems', which also maximised downstream development options and end user experience; and (2) improved conceptual front end designs targeting more effective visualisations and intuitive user experiences.

The research team worked closely with the development team to plan execution towards the target state, which included feature release schedules and managing implementation risk. Innovative ideas were contributed by all members to improve all aspects of design and development at the outset and along the way.

Project management was managed effectively with key milestones and fall back execution plans to manage project risk. This enabled the possibility to explore implementation of the Immersive scivis with reduced project risk. Governance and accountability was enforced by weekly meetings with planned agendas, clear accountabilities and deliverables agreed to by each team member. Key design, development or implementation issues were surfaced and solved during meetings, which allowed for a continuous development cadence. Feedback from self-testing and external evaluation was also used to adjust project deliverables within the project management plans.

6.3 Future Work

With respect to future work the project team suggest two further areas of development:

6.3.1 Further feature design and implementation:

While the majority of features in the ideal target state were implemented and sufficient to answer the MC2 challenge questions, a number of advanced features with longer development cycles were omitted given the timeframe available for the project. In terms of increasing complexity, these included:

- An even wider range of analytical functionality by identifying more valuable features to implement beyond the scope of the three questions, expanding the range of data transformations and filtering a user could select from, user options to adjust data ingested by the VA system on the fly and presentation of more informative, dynamic statistics to aid a user in problem solving.
- Enhancing the information visualisation experience further, by providing more relevant key indicators ('KPIs') and visual overlays such as trend lines in Tableau, D3 or Unity environments.
- Potential interoperability between screens for a streamlined user experience.
- More advanced physical and statistical modelling of gaseous dispersion in turbulent wind conditions, which could aid user prediction with distributional, rather than mean path tracebacks and enabling emission density overlays for more accurate scientific visualisation and exploration.

6.3.2 Real world utilisation and implementation:

Many assumptions were made in this study, mostly due to lack of data, relative complexity with time provided or knowledge on the area of study. With additional time the group speculate that the visualisation and user experience concepts in Gaia could be extended to real world industrial applications that are relevant to turbulence theory.

The conceptual experience of the Gaia VA system could be applied to improving manufacturing processes for base materials and chemicals, advanced manufacturing for aerospace, defense and automotive industries and potentially in weather forecasting, oceanography and climate modelling.

7. REFERENCES

[1] Ecke, R. (2005). The Turbulence Problem: An Experimentalist's Perspective. *Los Alamos Science*, No. 29, p 124-6.

[2] Chen, J. (2017). Analytics Inspired Visualisation: a Holistic In-situ Scientific Workflow at Extreme Scale. *VIS 2017 Keynote Synopsis.*

[3] Jiang, R. et al (2017). Peking University, *VAST 2017 submission.*
Retrieved from http://hcil2.cs.umd.edu/newvarepository/VAST%20Challenge%202017/challenges/Mini-Challenge%202/entries/Peking%20University/

[4] Tang, H. et al (2017). Purdue University, *VAST 2017 submission.*
Retrieved from http://hcil2.cs.umd.edu/newvarepository/VAST%20Challenge%202017/challenges/Mini-Challenge%202/entries/Purdue%20University/
and https://va.tech.purdue.edu/vast2017/mc2/windnebula.mp4.

[5] Gu, Q. et al (2017). The Hong Kong University of Science and Technology, *VAST 2017 submission.*
Retrieved from
http://hcil2.cs.umd.edu/newvarepository/VAST%20Challenge%202017/challenges/Mini-Challenge%202/entries/The%20Hong%20Kong%20University%20of%20Science%20and%20Technology/ and https://www.youtube.com/watch?v=2MBZLHq4Hvo&feature=youtu.be

[6] Castor, J. et al (2017). University of Illinois at Chicago, *VAST 2017 submission.*
Retrieved from:
http://hcil2.cs.umd.edu/newvarepository/VAST%20Challenge%202017/challenges/Mini-Challenge%202/entries/University%20of%20Illinois%20at%20Chicago/
and https://www.youtube.com/watch?v=2MBZLHq4Hvo&feature=youtu.be.

[7] Gaia Analytics Platform (2017) - University of Sydney, https://www.gaiaanalytics.com/

[8] Huang, Tony (2014). University of Tasmania - http://www.cs.usyd.edu.au/~shhong/tony.pdf

8. APPENDIX A: Meeting Minutes

Team members (initials): Dan Bridgman (DB), Matthew Burgess (MB), Dan Elias (DE), Andrew Huang (AH), Kristopher Lopez (KL), Cameron Wasilewsky (CW).

8.1 Group minutes - 13/09

Apologies: MB

Purpose: Discuss approach to assignment and decide on project strategy and objectives.

Outcomes:

The group agreed on the following project goals in order of value and risk:

Basic: Minimum requirements with static images to answer questions
Intermediate: Website embedding D3.js code to visualise analysis and data
Advanced: (1) Use Unity; (2) Animation with wind showing chemicals/pollution dispersion

The group agreed on the following implementation tools: D3.js, css, Tableau and Unity, html

Required actions:

(1) Clarify on Edstem about Evaluation section and also question on the questions.
(2) Discussed and decided on roles & responsibilities of group
   - Coding, data prep: KL, DE, CW
   - Report structure/presentation: AH, DB, (CW assisting)
   - Dedicated unity development for advanced visualisations: MB

Next meeting agenda:
Design & approach - hard stop by 17/09/2017

8.2 Group minutes - 20/09

Apologies: DE, CW

Purpose: Discuss ideas for analysis requirements to meet project objectives.

Status update: All required actions from the previous meeting were met.

Discussion:

Visualisation ideas:

- Heatmap for sensor performance, to inspect if working properly at all times or for unexpected behaviour.
- Change the line graph for mockup to heatmap.
- Changing heat for pollution over night
- Dendrogram for pollution from origin: Cluster dendrogram for visualisation of origin of chemicals released: https://bl.ocks.org/mbostock/4063570
- Boxplots for outliers - https://bl.ocks.org/mbostock/4061502

Questions ideas:

- Q2: Which chemicals are being detected by the sensor group? Heatmap
- Q2: What patterns of chemical releases do you see, as being reported in the data? Heatmap
- Q3: Probability % of chemical from factories - Stretch goal - Final report as part of analysis

Presentation ideas:

- Online presentation using Wix

Outcomes:

The group agreed to determine a list of features required for development. Concept features to start 'wide' with analysis ideas and then narrow down as insight/interesting patterns found.

The group agreed to host the VA system on Wix in web based format.

Required actions:

(1) Create starting feature list and framework, circulate over Slack, Google Docs (Owner: AH – due EoP 22/9)
(2) Review / add additional ideas: functionality, analysis, visualisations required to answer the questions (Owner: All members – due EoP 23/9)
(3) Create web based presentation structure on Wix (Owner: CW – progress update at Oct 4 meeting)

Next meeting agenda:
Post submission briefing; progress stocktake and re-prioritisation

8.3 Group minutes – 05/10 (One week to Presentation)

Apologies: None

Purpose: Discuss progress to date across Development. Walk through of design slides. Discuss presentation and questions regarding.

Status update: Change role for Matt to 100% development on Unity. Firming up development efforts for different objectives.

Discussion:

Kris - Focus on visual elements for system
Dan - Focus on data engineering/pre-processing for trend removal, outliers for Kris
Risk raised: Web responsiveness, scaling for different browsers.

Outcomes:

Change role of Matt development in Unity.

Required actions:

(1) Cam to set up hosting for - due end of week
(2) Kris to upload assets for hosting - due end of week
(3) Cam & Dan to finalise presentation and carry out practice sessions as required

Next meeting agenda
None - Presentation - No minutes to be taken.

8.5 Group minutes – 19/10

Apologies: AH

Purpose: Consolidate development efforts, discuss evaluation implementation of system.

Status update: Final report summary (DB and AH),

Discussion: Discussed evaluation framework using DECIDE. Discuss final submission of project and who is doing what. Discussed results on questions & aims. Dan B & Andy to answer questions with assistance from Dan E for review.

Evaluation - Introduction & methodology. Andy - Heuristics, Dan - Interviews, Cam - Survey.

Discussed deadlines for carrying out Evaluation of the system:
Structure/detail - Sunday EoP
Data collected - Wednesday EoP

Outcomes:

Agreed on structure of evaluation framework and roles/number of evaluators.

Required actions:

(1) Kris L & Matt to upload latest version of Gaia analytics platform by Sunday EoP for survey/interview collection of system.
(2) Cam to create Evaluation - Survey by Sunday EoP
(3) Dan B to create Evaluation - Interview by Sunday EoP
(4) Dan B, Dan E & Kris to carry out surveys and interviews by Wednesday EoP

Next meeting agenda:
Final wrap up of project prior to submission. Review any last minute code changes otherwise stop development of the Gaia system.

8.6 Group minutes – 26/10 - One week until final report due

Apologies: CW and MB

Purpose: Review evaluation results from survey interviewees. Consolidation of report and development efforts for final report hand in on 02/11.

Status update: Discuss final changes of system, agreed on cut off time for code cut. Andy & Dan catch up on report updates.

Discussion:

- Final wrap up of project any remaining business or concerns.

Outcomes:

Code cut time - Tuesday - 31/10 8pm. Final code changes agreed to be communicated to CW and MB (KL present)

Required actions:

1. Matt to fix issues with wind directions and plumes reversed - Due 29/10
2. Cam to update heatmap colours as per evaluation - Due 28/10
3. Kris to update D3 elements to remove performance counters for sensors and lengthen chemical compound name labels - Due 29/10

Next meeting agenda:
None, day of submission.

9. APPENDIX B: Code

Unity:

- Web: accessible from http://www.gaiaanalytics.website/Unity/, run in full screen (bottom located bottom right) for best results. Press escape key to exit full screen.
- Windows: Direct download of Zip file from https://www.gaiaanalytics.com/explore-yourself, simply extract the folder and run the executable file inside. This version has the highest visual fidelity and best performance.
- Operation: zoom using scroll wheel, pan with arrow keys or WSAD keys. Hover over UI elements for descriptions of their functions.

D3:

- Web accessible from http://www.gaiaanalytics.website
- To load as localhost please follow the below steps
    1) Extract the zip file onto your machine
    2) Open console/terminal and navigate to the folder containing index.html
    3) Load localhost using Python Simple HTTP Server
        a) Windows: python -m http.server 8888
        b) Mac: python -m SimpleHTTPServer 8888
    4) Navigate to localhost:8888

Tableau:

- Web accessible from https://www.gaiaanalytics.com/explore-yourself
- Navigate to either Sensor Dashboard or Heat map Dashboard

Python:

- Web accessible from https://www.gaiaanalytics.com/explore-yourself
- Navigate to notebooks used to carry out data cleansing and analysis

10. APPENDIX C: Survey Questions & Results

Survey Questions

Quantitative:

How effective is this system at illustrating sensor malfunctions and anomalies? (Aimed at Question 1)

How effective is this system at illustrating the patterns and release for each chemical? (Aimed at Question 2)

How effective is this system at highlighting which factory is responsible for what chemical? (Aimed at Question 3)

How usable did you find the system when interacting with it? (Usability testing)

How aesthetically pleasing did you find the system when interacting with it? (How 'good' are the visualisations and what kind of impact does it have on the user interacting with the VA system?)

Qualitative:

Are there any improvements or design changes you feel would make the system better? (How can the system be improved)

Are there any final comments or feedback you would like to pass back to the system developers about Gaia? (Rounding out feedback and open feedback)

Quantitative results summary:

Question	No. of responses	Average	Standard Deviation
How effective is this system at illustrating sensor malfunctions and anomalies?	8	3.9	0.22
How effective is this system at illustrating the patterns and release for each chemical?	8	4.5	0.19
How effective is this system at highlighting which factory is responsible for what chemical?	10	4.3	0.25
How usable did you find the system when interacting with it?	10	3.8	0.25
How aesthetically pleasing did you find the system when interacting with it?	10	4.5	0.17

Table 4: Quantitative results

Example qualitative comments:

Are there any improvements or design changes you feel would make the system better?

"The circles on the sensor of the Unity visualisation were confusing in terms of size. Maybe label the sensors in this visualisation. In tableau visualisation, the colour on heat map should be one colour diverging. Green and red makes it confusing. Also maybe pick a colour that is different for all the other colours picked for chemicals. In D3 visualisation , performance percentage is not very clearly understandable."

“Can you label the sensors in the unity display in the same way as the d3 chart? The simulation was really good, but it was quite difficult to get the data I wanted out of it. There was also a bug with sensor 1 - ticking the checkbox disabled it, rather than enabling it like the others.”

“....4. Heat Map. a. Colours are confusing - I associated the colours with the chemicals on previous pages. Should be a single colour that has not been used before. There should be a legend to show what the colour is LOW--> HIGH. b. Bottom right sensor graph- words Monday- Sunday are squished.”

Are there any final comments or feedback you would like to pass back to the system developers about Gaia?

”Important to keep colours consistent throughout all four visualisations.”

“Labelling data is essential to know what each measure represents, and to effectively use each visualisation. If further information is required which cannot be communicated via labels and is essential for the visualisation, please provide a prominent link to it in the visualisation.”

“Among all three systems, I like Tableau the most specially the filter capabilities.”