University of Pisa

Department of Computer Science

Visual Analytics

Tracking Suspicious Fishing Activity: Solving the VAST Challenge 2024

Matilde Contestabile

Student ID: 639307

m.contestabile1@studenti.unipi.it

Abstract

This report describes the design and implementation of an interactive visual analytics system developed for the VAST Challenge 2024 (Mini-challenge 2). The system enables exploration of vessel activity and fish transactions, allowing users to identify patterns, anomalies, and potential illegal behavior. We detail the dataset, the analytical approaches applied during exploratory data analysis, and the rationale behind the visual design choices, including color, layout, and interaction strategies. Through dynamic visual representations, the system supports both high-level overviews and detailed investigation, facilitating the discovery of insights and supporting analytical tasks in complex data.

Academic Year 2024/2025

Table Of Contents

1 Introduction

1.1 Project Overview and VAST Challenge Context

This project addresses the analytical goals of the **Mini-Challenge 2 (MC2)** from the **VAST Challenge 2024** (available for consultation at `https://vast-challenge.github.io/2024/index.html`). The VAST Challenge is a well-established international competition focusing on advanced visual analytics. The 2024 edition is set in the fictional nation of Oceanus, where a vibrant commercial fishing industry is facing threats from unethical practices by certain actors. The *Challenge Overview* recites as follows:

> Welcome to Oceanus, an island nation with a healthy market for commercial fishing. Most companies in the region are united in following regulations and implementing sustainable fishing practices. But there are a few companies who are willing to cross ethical lines to increase their catch and their profits. Luckily, FishEye International maintains a watchful eye on fishing data. Their dedicated analysts have been processing data from various sources into a knowledge graph that they call CatchNet: the Oceanus Knowledge Graph.

In this context, **Mini-Challenge 2 (MC2)** focuses on analyzing the behavior of vessels operating within Oceanus using multiple datasets, including transponder pings, harbor visit records, and transaction logs. The central goal is to investigate **illegal fishing behaviors** by the company **SouthSeafood Express Corp** and to develop visualizations that support **anomaly detection in vessel movements and supply chains**. Here's the *overview* specific to the mini-challenge:

> In Oceanus, island life is defined by the coming and going of seafaring vessels, many of which are operated by commercial fishing companies. Typically, the movement of ships and goods are a sign of Oceanus's healthy economy. But mundane routines can be disrupted by a major event.
>
> FishEye International has discovered that SouthSeafood Express Corp was engaged in illegal fishing, prompting the need for advanced analysis. As part of this challenge, analysts must investigate this event using CatchNet data to uncover behavioral patterns, track fish product movements, and support future monitoring.

More specifically, the mini-challenge requires to address the following research questions through a series of targeted visual workflows and analytical strategies, with an emphasis on interactivity and clarity:

1. **Cargo Attribution:** Given the lack of direct vessel identifiers in port transaction records (due to wrong purchase of records by FishEye analysts), can we visually and analytically associate cargo deliveries with specific vessels? What seasonal or regional trends emerge in fish exports?
2. **Illegal Behavior Detection:** How do the trajectories and port interactions of SouthSeafood Express vessels differ from compliant vessels? When and where did violations occur?
3. **Behavioral Pattern Matching:** Are there other vessels whose behavior mirrors that of SouthSeafood Express? Can similar illegal activities be inferred?
4. **Post-Incident Trends:** Following the discovery of illegal fishing, have commercial fishing patterns across Oceanus shifted? Are new anomalies emerging?

Answers are available at Section 6.

1.2 Project Repository and Implementation

The visual analytics interface was implemented as a Single Page Application (SPA) using **Vue.js** for the frontend framework, **TailwindCSS** for styling, **Leaflet** for geographic visualizations and **D3.js** for data-driven document manipulation and rendering. The full implementation of the application is publicly available in the GitHub repository: `https://github.com/matildeec/VisualAnalytics2025`. This includes:

- `DataPreprocessing/` – Jupyter Notebooks and Python scripts for data cleaning and inspection of the VAST Challenge datasets
- `src/` – The source code of the web application, developed using Vue.js, D3.js, and TailwindCSS
- `public/` - Static assets including images and preprocessed data files

To install and run the application locally, please refer to the instructions provided in the `README.md` file of the repository.

2 Data Understanding and Preprocessing

This section outlines the data exploration and cleaning process carried out for the VAST Challenge 2024 Mini-Challenge 2. Using a data mining approach and Python libraries, we prepared the knowledge graph dataset for visual exploration and anomaly detection, specifically, in a format suitable for JavaScript-based applications.

2.1 Data Sources and File Structure

The challenge dataset consists of multiple files, with the central one being a JSON-encoded knowledge graph representing vessel activities, harbor transactions, and commercial fishing behaviors. Supplementary files provide metadata and geospatial context.

File Name	Description
`mc2.json`	Main knowledge graph, structured as a multigraph with typed nodes and edges representing entities and events (e.g., vessels, ports, transactions, pings).
`Oceanus Geography Nodes.json`	Metadata for geographic locations (e.g., ports, reefs, regions).
`Oceanus Geography.geojson`	Geospatial file for mapping entities in Oceanus.

Table 1: Overview of provided data files

To support understanding of the knowledge graph, the document `VAST2024 - MC2 Data Description.docx` is provided, detailing the graph structure and entity semantics.

2.2 Graph Structure Analysis and Preprocessing

The knowledge graph contained in `mc2.json` was imported using the `networkx` library, resulting in a directed multigraph. Nodes and edges include a `type` attribute, which was used to categorize and filter graph elements. Both nodes and edges were extracted into separate Pandas DataFrames to enable structured analysis.
Initial preprocessing involved removing metadata fields irrelevant to downstream analysis, specifically:
`[_last_edited_by, _last_edited_date, _date_added, _raw_source, _algorithm]`.
Next, nodes (entities) and edges (events) were grouped by their `type`. Within each group, columns containing only missing values were discarded, and only relevant attributes were retained to support visualization and modeling tasks.
A summary table of the cleaned data files intended for use in the JavaScript implementation is provided at the end of this section.

2.2.1 Node Overview and Distribution

The knowledge graph contains a diverse set of node types. Table 2 summarizes the most relevant node types by category, showing the number of instances and representative attributes retained after cleaning.

Table 2: Updated Node Types with Attributes in the Knowledge Graph

Category	Node Type	Count	Attributes
Document	Entity.Document.DeliveryReport	5307	`qty_tons`, `date`
Vessel	Entity.Vessel.FishingVessel	178	`name`, `flag_country`, `company`, `tonnage`, `length_overall`
	Entity.Vessel.CargoVessel	100	`name`, `flag_country`, `company`, `tonnage`, `length_overall`
	Entity.Vessel.Ferry.Passenger	3	`name`, `flag_country`
	Entity.Vessel.Ferry.Cargo	2	`name`, `flag_country`
	Entity.Vessel.Tour	6	`name`, `flag_country`
	Entity.Vessel.Research	2	`name`, `flag_country`
	Entity.Vessel.Other	5	`name`, `flag_country`, `length_overall`

Continued on next page...

Table 2: Updated Node Types with Attributes in the Knowledge Graph

Category	Node Type	Count	Attributes
Location	Entity.Location.Point	12	`name`, `description`, `activities`, `kind`
	Entity.Location.City	6	`name`, `activities`, `kind`
	Entity.Location.Region	6	`name`, `description`, `activities`, `kind`, `fish_species_present`
Commodity	Entity.Commodity.Fish	10	`name`

2.2.2 Edge Types and Event Semantics

Edges in the knowledge graph represent interactions and events between entities, providing both temporal and relational context to the data. Table 3 summarizes the original edge types, including their counts, source and target entities, and retained attributes.

Event Type	Count	Source	Target	Attributes
Event.TransportEvent.TransponderPing	258,542	Entity.Location	Entity.Vessel	`time`, `dwell`
Event.Transaction	5,307	Entity.Document	Entity.Commodity	`date`
Event.Transaction	5,307	Entity.Document	Entity.Location	`date`
Event.HarborReport	2,487	Entity.Vessel	Entity.Location	`date`, `data_author`

Table 3: Original edge types with counts, sources, targets, and attributes.

From here, two critical issues become apparent that affect readability and understanding of the data. First, for **Event.HarborReport**, the edge direction is counterintuitive: while the information originates from portmasters, the current structure connects vessels (sources) to ports (targets). Reversing this direction would better reflect data provenance and better match with **Event.TransportEvent.TransponderPing**. Second, **Event.Transaction** edges are redundant: each delivery report is linked to both a location and a commodity via two separate edges, resulting in duplication and unnecessary complexity.
To address these issues, we redesigned the schema. Information about commodities, including species and quantities, has been moved into the delivery report entity itself, while **Event.Transaction** edges now exclusively connect delivery reports (sources) to locations (targets). The updated structure is summarized in Table 4.

Event Type	Count	Source	Target	Attributes
Event.TransportEvent.TransponderPing	258,542	Entity.Location	Entity.Vessel	`time`, `dwell`
Event.Transaction	5,307	Entity.Document	Entity.Location	`date`
Event.HarborReport	2,487	Entity.Location	Entity.Vessel	`date`, `data_author`

Table 4: Updated edge types

2.2.3 Cargo Attribution via Suspected Vessels Tagging

To address the challenge of associating cargo deliveries with specific vessels despite missing direct identifiers in the port transaction records (see Question 1 in 1.1), we introduced an attribute called `suspected_vessels` within the transactions dataset which was derived in an analytical way. This attribute records, for each transaction, a list of vessels potentially responsible for the cargo. The goal was to infer likely vessel–cargo associations based on spatiotemporal docking information and the presence of fish species in regions. The logic underlying this inference can be formalized as:

$$
\begin{aligned}
&\textit{Vessel Type} \in \{\text{`CargoVessel', `FishingVessel', `Other'}\} \\
&\land \ \textit{Vessel docked in export harbor} \ (\pm 1 \text{ day}) \\
&\land \ \textit{Vessel Tonnage} \ \geq \ \textit{Cargo Weight} \\
&\land \ \textit{Vessel visited species region within previous 21 days} \\
&\implies \textit{The fish species is likely the vessel's cargo.}
\end{aligned}
$$

Applying this rule, we flagged transactions meeting these conditions and annotated them with the corresponding `suspected_vessels`. A visual summary of the logic is provided in Figure 1.

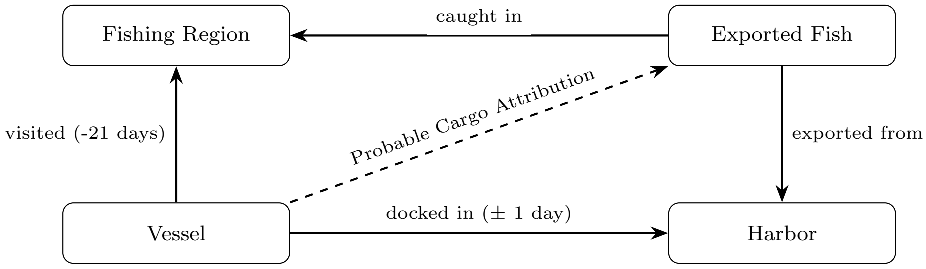


Figure 1: Flowchart showing the spatiotemporal derivation of the `suspected_vessels` attribute.

We set the detection window to 21 days (3 weeks) as a result of reasoning around vessel trip durations derived from trajectory data and the average tonnage of the fleet. The analysis of the statistical distribution of vessel trip durations revealed a highly skewed distribution: while the median trip duration is short ($\approx$ 2 days) and 95% of trips conclude within 7.4 days, the distribution exhibits a significant "long tail", with the 99th percentile extending to 38.2 days. By extending the window beyond the 95th percentile, we aim to capture this specific sub-segment of long-endurance voyages where deep-sea illegal fishing is more likely to occur.
This decision is further supported by the high average tonnage of the fleet (FishingVessel avg: 3,480 GT), which indicates industrial capabilities likely including onboard freezing systems. Unlike artisanal boats, such vessels can operate for weeks without risking spoilage. Furthermore, intermediate stops for logistics (e.g., bunkering, crew changes) do not necessarily imply the offloading of cargo; such behavior may instead serve as a masking strategy to obscure the true origin and timeline of the catch. Consequently, a narrower detection window (e.g., 7 days) would risk the premature exclusion of catches from earlier voyage stages that are still retained onboard.

2.3 Resulting `.json` Files

Table 5 summarizes the cleaned and structured `.json` files generated from the original knowledge graph after preprocessing. These files are optimized for use in the JavaScript-based implementation and support efficient querying since together they construct a database with primary (underlined) and foreign keys.

File Name	**Description**	**Keys**
`vessels.json`	Vessel entities	<u>`id`</u>, `vessel_type`, `name`, `company`, `flag_country`, `tonnage`, `length_overall`
`commodities.json`	Traded fish	<u>`id`</u>, `name`
`documents.json`	Delivery reports	<u>`id`</u>, `commodity`, `qty_tons`
`locations.json`	Locations	<u>`id`</u>, `name`, `location_type`, `kind`, `description`, `activities`, `fish_species_present`
`transponder_pings.json`	Vessel GPS logs	<u>`source`</u>, <u>`target`</u>, `time`, `dwell`
`harbor_reports.json`	Port-exit records	<u>`source`</u>, <u>`target`</u>, `date`, `data_author`
`transactions.json`	Cargo exchanges	<u>`source`</u>, <u>`target`</u>, `date`, `suspected_vessels`

Table 5: Cleaned `.json` files and their key attributes. Underlined keys indicate primary identifiers.

To further facilitate the representation of vessel trajectories, a dedicated file named `trajectories.json` was created. It is structured as a dictionary, where the keys are vessel IDs and the values are ordered lists of transponder pings represented as JSON objects.

3 Exploratory Data Analysis

We conducted an initial exploratory data analysis (EDA) on the cleaned data using a combination of Python libraries such as `pandas`, `matplotlib`, `altair`, and `seaborn`. The purpose of this analysis was to uncover patterns in vessel activity, identify potential anomalies, and provide a foundation for more sophisticated visualizations later on.

3.1 Analytical Analysis of vessel activity

To effectively interpret vessel behavior, it was first necessary to understand the geographical context of the dataset. Certain regions are in fact designated as ecological preserves, where fishing activity is strictly prohibited. Vessels, however, were observed operating across both legal fishing grounds and these protected zones. Additionally, the dataset provides the list of species present in every region, allowing us to identify which species are exclusive to protected preserves. Using this information, we derived an analytical classification of *illegal species*, defined as those found exclusively within protected preserves, and *suspect* ones, which appear in both legal and illegal zones. Table 6 summarizes this mapping: species highlighted in red appear only in illegal zones and therefore serve as indicators of potential illegal fishing activity, while those in yellow are considered suspect due to their presence in both legal and illegal areas.

Fish Species	Cod Table	Wrasse Beds	Tuna Shelf	Ghoti Preserve	Nemo Reef	Don Limpet Preserve
`gadusnspecificatae4ba`	✓					
`piscesfrigus900`	✓	✓	✓		✓	✓
`habeaspisces4eb`	✓	✓	✓	✓	✓	✓
`labridaenrefert9be`		✓		✓	✓	
`piscessatisb87`				✓	✓	✓
`piscisosseusb6d`				✓		
`thunnininveradb7`			✓		✓	✓
`piscesfoetidaae7`						✓
`piscissapidum9b7`			✓			

Table 6: Matrix of fish species presence across regions. Rows in red indicate species found exclusively in ecological preserves (illegal). Shaded columns mark the illegal fishing zones.

After establishing a clear mapping of legal versus illegal zones and species, we turned our attention to temporal patterns in vessel behavior. Specifically, we examined three features: *dwell time* (how long a vessel remained near a given location), *gaps between transponder pings*, and *harbor visit frequencies.* The rationale was straightforward. Prolonged dwell times may signal suspicious activity, such as covert fishing or offloading. Unusual gaps in transponder signals could indicate attempts at evasion, while irregular harbor visits might point to atypical supply or unloading patterns.

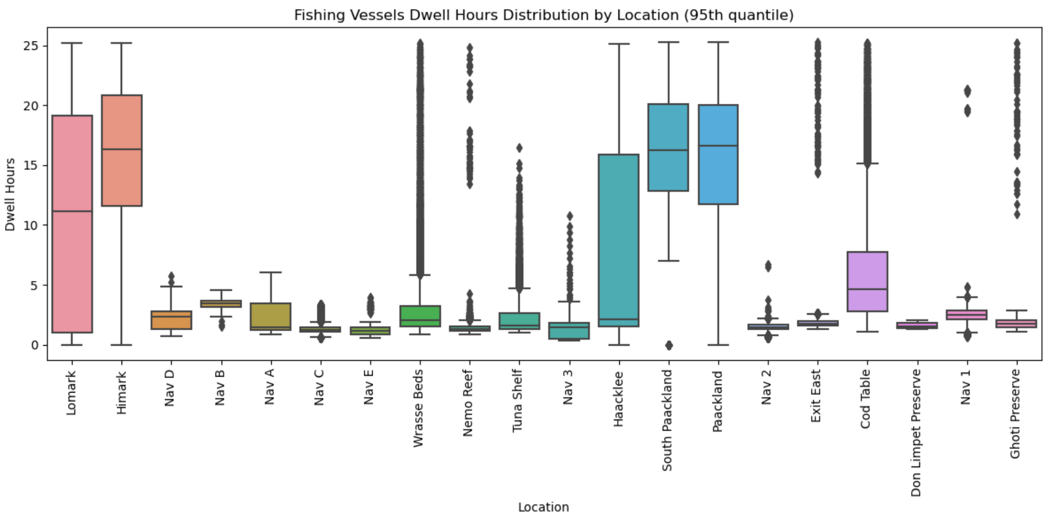


Figure 2: Dwell time of fishing vessels by location.

Of these features, however, the latter two proved less reliable as standalone indicators. Coverage gaps were common but did not consistently align with illegal fishing zones, and variations in harbor visit frequency were not necessarily suspicious when taken in isolation. By contrast, the dwell time analysis

provided the clearest signals of anomalous vessel behavior and thus emerged as the most informative exploratory measure. Figure 2 illustrates these dwell time patterns across different locations.
The boxplots make clear that vessels indeed lingered significantly longer at certain locations than expected. While many values fell within a normal operational range, some extreme outliers stand out and warrant further investigation. This insight guided the direction of subsequent visualizations, where the emphasis shifted toward tracking how vessels move, where they pause, and how these behaviors align with known illegal regions.

3.2 Temporal Analysis of Fish Deliveries

To better grasp the scale of fish deliveries, we aggregated transaction records to track quantities over time. This allowed us to identify seasonal dynamics, peak transaction periods, and potential export trends as requested in Q1. Figure 3 illustrates total transaction volumes for South Paackland as an example case. Noticeable spikes suggest periods of heightened demand, which may align with peak fishing activity or, in some cases, potential illegal operations.
In the visualization, illegal fish species are highlighted in shades of red, while legal (including suspect) species are shown in shades of blue, enabling a direct comparison of their relative contribution to overall trade flows. This perspective not only informed our analysis but also shaped the subsequent design of our visualization approach.

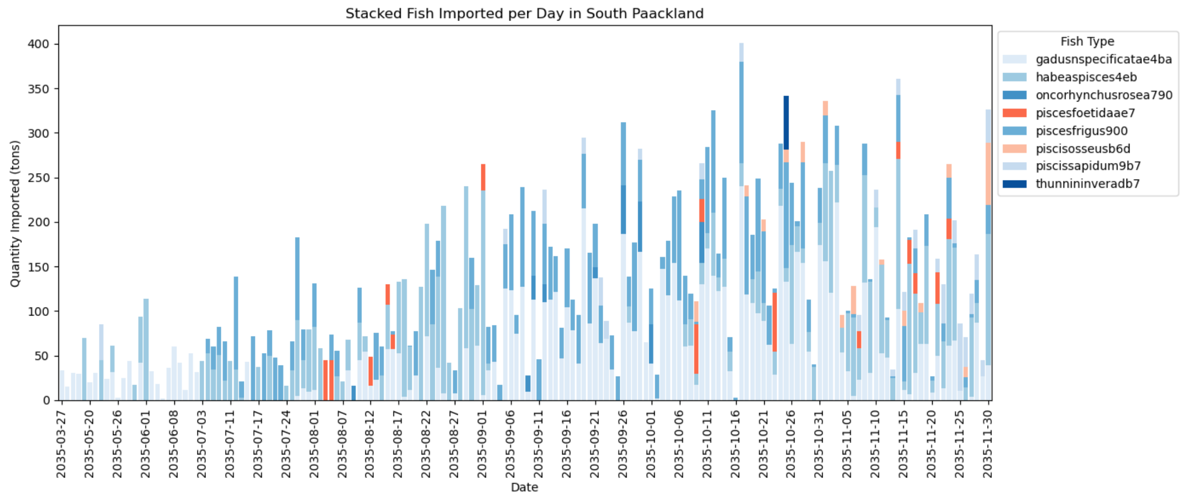


Figure 3: Transaction volume over time, representing fish deliveries. Illegal species are shown in shades of red, while legal species are shown in shades of blue.

3.3 Analytical Analysis of SouthSeafood Express Corp vessels

As part of our EDA, we conducted a focused investigation of the company identified as engaging in illegal activities, *SouthSeafood Express Corp*, as requested by the challenge. Table 7 summarizes the two vessels associated with this company, including their vessel IDs, names, and tonnage.

Vessel ID	**Name**	**Tonnage**
snappersnatcher7be	Snapper Snatcher	100
roachrobberdb6	Roach Robber	11,700

Table 7: Vessels associated with SouthSeafood Express Corp.

Both vessels were carefully examined for unusual patterns in their activity, including extended dwell times, abnormal harbor visit frequency, and irregular transaction volumes.
We first established the timeframe of activity prior to the discovery of illegal behavior: *SouthSeafood Express Corp* vessel pings ranged from `2035-02-01` to `2035-05-14`. This period serves as a reference for analyzing changes in behavior, relevant to questions such as those in Q4 regarding post-incident

trends: whether commercial fishing patterns across Oceanus have shifted and whether new anomalies are emerging.
To assess whether dwell time and other features are indicative of illegal activity, we generated rankings of top dwellers, vessels with the longest ping gaps, and other relevant metrics. In the ranking of dwell times at illegal locations, `snappersnatcher7be` (Snapper Snatcher) ranked 76th, while `roachrobberdb6` (Roach Robber) did not appear among the top vessels. This suggests that although Snapper Snatcher spent time at potentially suspicious locations, it was not among the vessels with the longest dwell times, indicating that further analysis is required to determine whether this behavior is significant.
Similarly, in the analysis of ping gaps, Snapper Snatcher ranked 173rd and Roach Robber 175th. Neither vessel exhibited substantial gaps in their transponder signals, implying that ping gaps alone are not a reliable indicator of evasion or suspicious activity in this context.
More interesting insights emerged from examining vessel routes based on transponder pings. Snapper Snatcher made multiple visits to *Exit East*, a region designated for deep-sea fishing. While this does not immediately stand out compared to general vessel activity, it raises questions about whether this route reflects typical operational behavior or unusual activity. Further investigation, particularly with complete route visualizations, will help clarify these patterns.
While detailed considerations of vessel behavior will be presented later as the full visualizations are presented, we provide the initial plot that was generated using Altair to explore these routes.

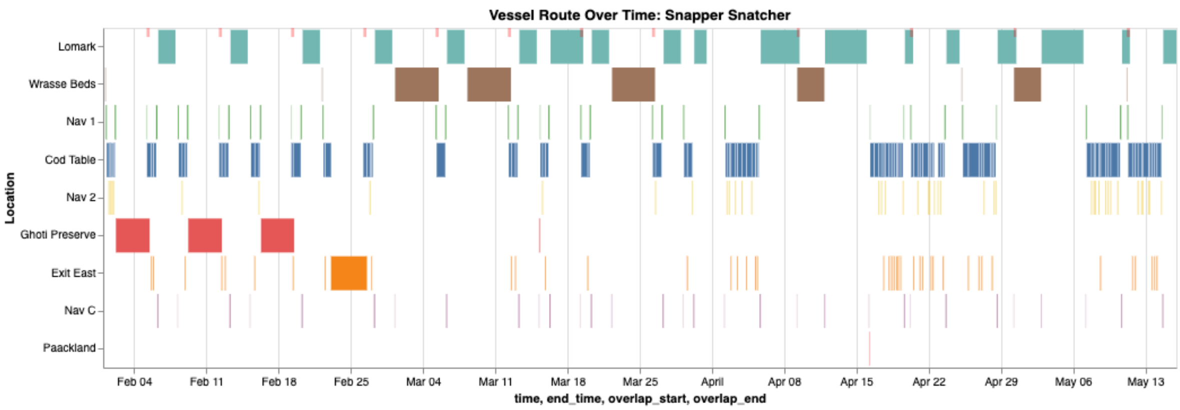


Figure 4: Route of `Snapper Snatcher`

4 Design and Architecture of the Visual Analytics Interface

The visual analytics interface was designed to support anomaly detection, pattern recognition, and hypothesis generation with the ultimate goal of answering the four research questions reported in Section 1.1. Guided by established principles of visual encoding, interaction design, and narrative visualization, the design process consisted of three main stages. First, we surveyed the state-of-the-art tools to gain inspiration for tracking fishing vessels. Next, we identified the key visual variables necessary to address the research questions posed by the VAST Challenge. Finally, these design decisions were translated into a prototype using Figma, which then served as the blueprint for the implementation of the interactive system in JavaScript.
This section provides a detailed account of the resulting design choices for the visual analytics system developed.

4.1 State-of-the-Art Interfaces

To first gain inspiration, we explored existing platforms addressing similar challenges:

- **VesselFinder**[1]: a real-time vessel tracking platform that visualizes vessel locations worldwide using AIS data, providing an intuitive overview of maritime activity.

[1] Available at `https://www.vesselfinder.com/`

- **GlobalFishingWatch**[2]: a platform for visualizing global fishing activity, allowing users to monitor vessel movements and identify potential illegal fishing operations.

These platforms shaped our initial thinking, particularly in how vessel routes could be represented on a geographic map. However, the structure of our dataset imposed important constraints that made it impossible to reproduce the same level of detailed route visualizations. Although we had geographic coordinates available to reconstruct a map of Oceanus (via the provided `.geojson` file of Oceanus), the vessel GPS pings did not form a continuous trajectory. Instead, they resembled a series of discrete points – often multiple pings from the same location – linked only by timestamps and dwell times. As a result, we lacked the granular path data necessary to accurately reconstruct "real ocean routes" between locations *A, B*, and *C*.
Rather than forcing an incomplete or misleading geographic reconstruction of exact travel paths, we chose to design visualizations conceptually similar to the static plots created in our Jupyter Notebooks as part of EDA (see Section 3), focusing on patterns and trends in vessel pings – such as timing, frequency, and distribution. This approach allowed us to emphasize meaningful behavioral patterns over potentially inaccurate (and ultimately insignificant) geographic details.

4.2 Interface Structure and Pattern to Communicate

To manage the complexity of the data and the multi-faceted research questions, we adopted a multi-page architecture. The application is structured into three distinct views: **Traffic Explorer**, **Harbor Inspector**, and **Trajectory Analyzer**. Each view serves a distinct stage in the analytical pipeline, progressively revealing deeper insights. We provide a brief overview of each page below.
As a design choice, we adopted a **reader-driven narrative visualization** paradigm, empowering users to explore data at their own pace while providing **author-driven** entry points (such as pre-selected filters for known violators like SouthSeafood Express Corp) to jumpstart the investigation.
A consistent global navigation bar allows seamless switching between views. To support deep investigation, the interface integrates interactive techniques including **zooming**, **brushing**, and "details-on-demand" **tooltips**, enabling micro-level analysis without cluttering the macro-level visualizations.

4.2.1 Traffic Explorer

The **Traffic Explorer** serves as the primary entry point. It fulfills a dual purpose: establishing geographical context via an **Interactive Map** and visualizing temporal patterns via a **Traffic Activity Chart**. This combination allows users to identify macroscopic anomalies in traffic density and spatial distribution.

Interactive Map. The spatial component renders the geography of Oceanus using Leaflet and the provided GeoJSON data. It visually distinguishes between functional areas, such as Ecological Preserves (green), Fishing Zones (blue), and landmasses (beige).
To assist in understanding the complex regulations of the region, the map is linked to a dynamic **Info Panel**. When a user selects a zone (e.g., *Nemo Reef*), it turns ocre to indicate selection, and the panel populates with semantic metadata, including the zone type, permitted activities (e.g., Recreation, Tourism), and a list of endemic species colored according to their legal status (see Section 4.3.2). This context is critical for identifying illegal fishing activities in protected waters.

Traffic Activity Chart. Upon selecting a zone, the application generates a **temporal strip plot** to visualize vessel presence over the entire dataset duration. In this visualization:

- The **Y-axis** lists individual vessels.
- The **X-axis** represents the timeline.
- **Vertical bars** represent discrete "pings" or presence events within the selected zone.

Visually resembling a barcode, this density strip allows investigators to instantly recognize patterns, such as regular entry intervals or continuous loitering.
The chart is equipped with robust filtering capabilities:

[2] Available at `https://globalfishingwatch.org/map`

(a) Interactive Map with zone selection

(b) Info Panel showing zone details

Figure 5: The interactive map.

- **Time-of-Day Filtering:** Recognizing that illicit activity often occurs during specific windows, a range slider allows users to filter pings by time of day (00:00–24:00). Presets for *Morning*, *Day*, and *Night* allow for rapid comparison of diurnal vs. nocturnal traffic.
- **Visual Highlighting:** To aid the narrative, vessels belonging to *SouthSeafood Express Corp* are highlighted in yellow by default, contrasting with the color of general traffic (colors are assigned based on the zone selected, as explained in Section 4.3.3). Each vessel is also annotated with its type (e.g., Fishing, Cargo) via a colored tag next to its name.
- **Pinning:** Users can "pin" specific vessels of interest, isolating their trajectories for a focused comparison with others.

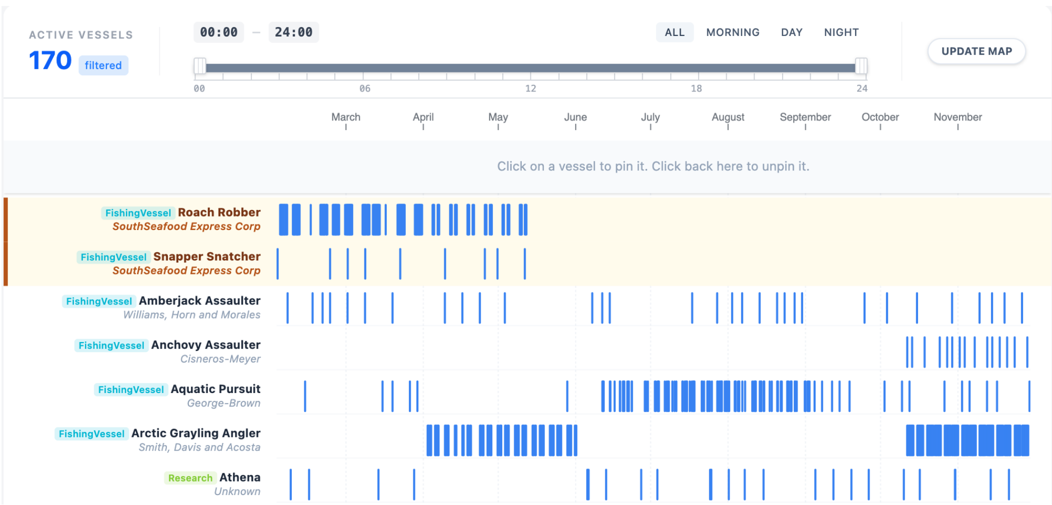
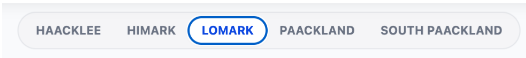


Figure 6: The Traffic Activity Chart.

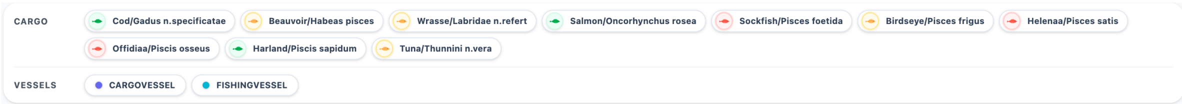
4.2.2 Harbor Inspector

The **Harbor Inspector** page focuses on port dynamics, enabling users to analyze cargo exports and harbor visits by specific location. This view is designed to address the core challenge of cargo attribution, which directly addresses Q1. To achieve this, we developed a composite **Dual-Axis Mirror Plot**, which visually correlates the volume of exports with vessel arrival dynamics on a shared temporal axis.

Data Filtering. To manage the high data volume, the interface provides a coordinated filtering system where users can refine the view by Harbor, Cargo Species, and Vessel Type. These filters operate globally, updating the Mirror Plot to help users correlate specific cargo types with relevant vessels.



(a) Harbor selection panel



(b) Filter panel for cargo species and vessel type

Figure 7: Harbor selection and filtering panels.

The Mirror Plot. This central visualization serves as the analytical engine of the view, split into two aligned components that share a common time axis:

- **Exported Cargo (Top Half):** A **stacked bar chart** represents the daily volume of fish exports (in tons), color-coded by legal status of species. This allows analysts to instantly detect seasonal anomalies and peaks in illegal shipments.
- **Docked Vessels (Bottom Half):** An **inverted lollipop chart** visualizes vessel arrivals by gross tonnage (GT) over time. Vessels are colored by type, and the dataset is filtered to include only vessels with registered gross tonnage (> 0) to focus the analysis on significant maritime activity (this excludes small boats and recreational vessels that are unlikely to carry large cargo loads like research or tour boats).

By vertically aligning these two datasets, the user can perform visual correlation: a spike in illegal cargo in the top chart can be vertically traced down to identify which heavy-tonnage vessels docked immediately prior to the export. Those analytically linked to illegal shipments are highlighted in red.

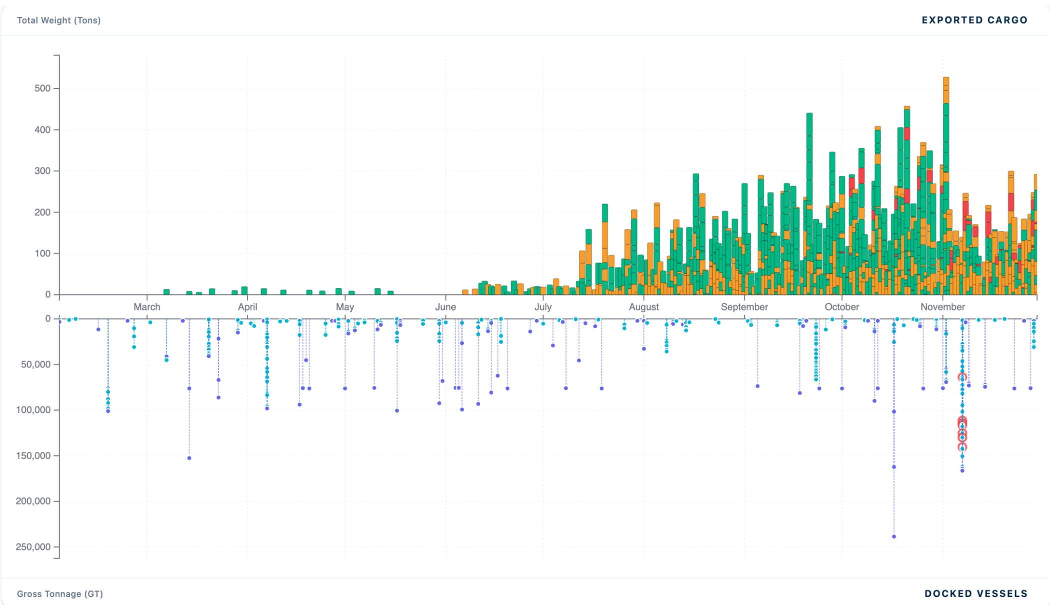
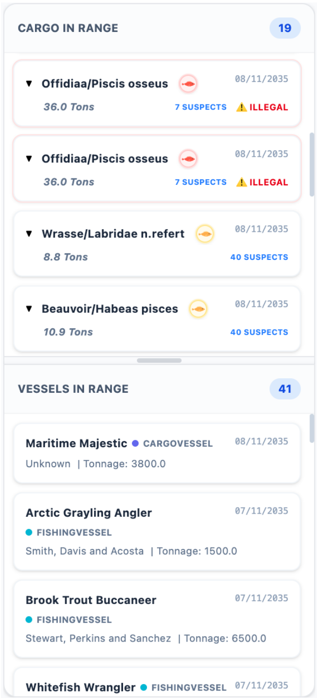


Figure 8: The Mirror Plot.

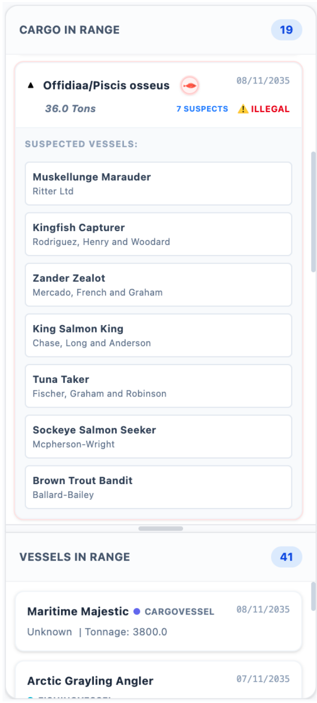
Interactive Filtering and Suspect Identification. To move from visual correlation to specific attribution, the interface employs a **temporal brushing** interaction. Selecting a time window on the chart updates the **Details Sidebar**, which lists:

1. **Cargo in Range:** A list of all shipments during the brushed period. Illegal shipments are explicitly flagged with a warning icon.

2. **Suspect Generation:** Expanding a specific cargo entry reveals a computed list of **"suspected vessels"** (ships that were docked in the harbor within a relevant time window of the export, have sufficient tonnage to carry the cargo, and have been observed fishing in protected zones).
3. **Vessels in Range:** A list of all vessels docked during the brushed period, providing details such as company and exact tonnage.



(a) The Details Sidebar closed.



(b) The Details Sidebar showing suspected vessels for an illegal cargo.

Figure 9: The Details Sidebar.

4.2.3 Trajectory Analyzer

Finally, the **Trajectory Analyzer** page enables granular comparisons between specific vessel trajectories, addressing Q2 and Q3, about illegal behavior detection and pattern matching. This view consists of a juxtaposed layout, allowing users to select any two vessels via distinct control panels and visualize their movements side-by-side.

Vessel select. The vessel selection panels allow users to filter the fleet by company, type, and tonnage. Once a vessel is selected, key metadata (e.g., company, type, tonnage) is displayed. This contextual information aids in selecting vessels of interest for comparison.

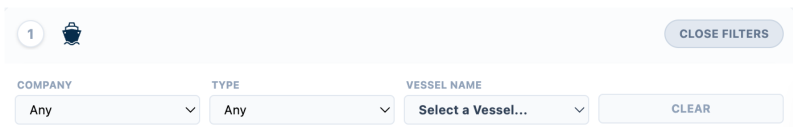


Figure 10: The Vessel Selection Panel.

Multi-Layered Spatiotemporal Timelines. Trajectories are visualized using a location-based timeline chart, inspired by the static plots developed during EDA (see Section 3). This design effectively conveys complex spatiotemporal patterns while allowing for direct comparison between two vessels.

- The **Y-axis** lists all key locations in Oceanus (fishing grounds, ecological preserves, cities, and buoy markers).

- The **X-axis** represents the timeline.
- **Data Layers** are superimposed to reveal discrepancies between reported and actual behavior:
  1. **Transponder Pings Layer**: pings are shown in solid bars colored according to the zone type. These represent self-reported locations via AIS data.
  2. **Harbor Reports Layer**: harbor reports are shown as striped bars colored according to the zone type. These indicate official harbor logs of vessel presence. Discrepancies between Pings and Reports immediately signal "dark ship" activity or falsified logs.
  3. **Cargo Attribution Layer**: cargo attribution is visualized via fish icons overlaid on the timeline at harbor locations. To handle multiple shipments associated with a single docking event, transactions are aggregated into a single representative icon featuring a red numeric badge that indicates the count of distinct cargo batches. The icon is color-coded based on the highest-risk species present (e.g., illegal), allowing analysts to instantly correlate movement through protected zones with the volume and nature of the transported catch.

Moreover, to directly answer Q4, the timeline includes a **Contextual Event Layer**. A vertical red dashed line marks the exact date of the *SouthSeafood Express Corp* ban. This visual anchor allows analysts to instantly assess whether a vessel's behavior (e.g., stopping illegal fishing) changed in response to the regulatory enforcement.
To manage visual complexity, a "Layer Control" toolbar (visible at the bottom of Figure 11) allows users to toggle specific layers on or off, isolating specific variables such as cargo or location pings for clearer inspection.

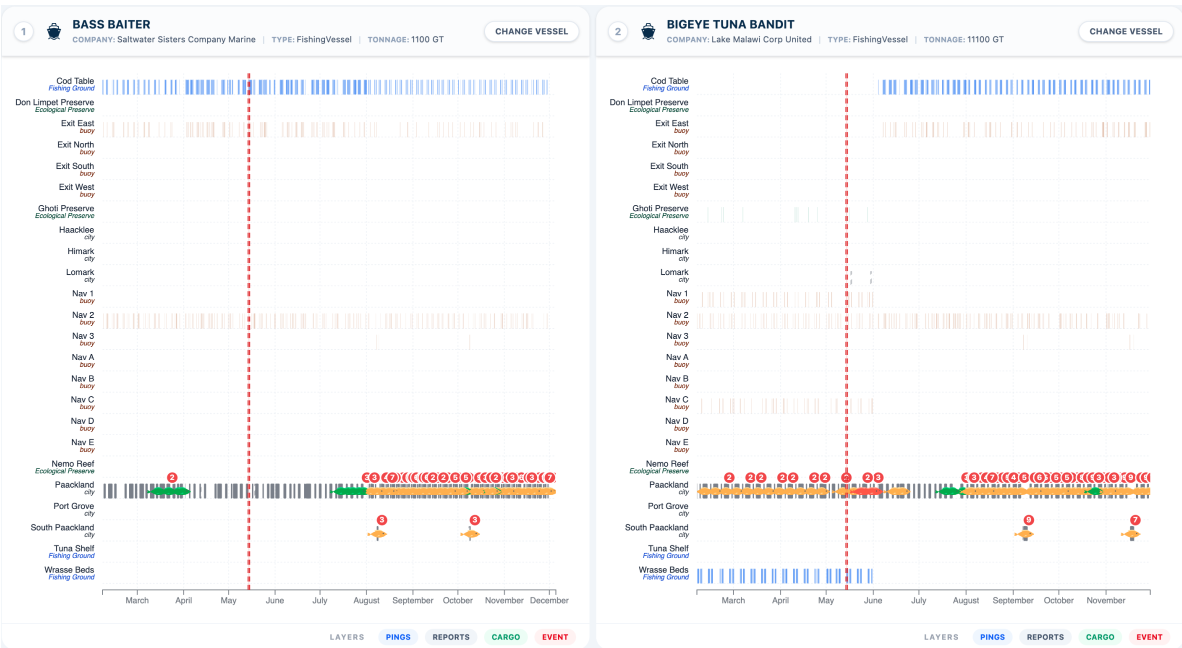


Figure 11: The Trajectory Analyzer View.

4.3 Design Choices and Visual Encoding

The interaface was designed with a focus on clarity, usability, and effective communication of complex data patterns. It employs a clean, minimalist aesthetic prioritizing white space and high-contrast color encodings to enhance data legibility. To maintain a neutral and distraction-free environment, the color palette was carefully curated to ensure that visual encodings for key data variables (commodities, zones, vessel types) stand out distinctly against the background. The following sections detail the specific design choices and visual encoding strategies employed.

4.3.1 Overall Aesthetic and Color Palette

The interface background and layout were designed to minimize visual clutter and enhance focus on data visualizations. The primary background color is a **clean white** (`#FFFFFF`), chosen to maximize

contrast and maintain a neutral canvas for complex data visualizations. This design choice minimizes visual noise, ensuring that chromatic encodings for data variables remain the focal point. To establish a strong visual hierarchy, the top navigation bar (Figure 12) utilizes a **dark blue** (`#092C4C`) anchor, which contrasts effectively with the lighter content areas. Dark blue was selected for its associations with trustworthiness and professionalism, aligning with the investigative nature of the application, as well as its thematic connection to maritime contexts.
The navigation controls incorporate a semi-transparent "glassmorphism" effect, allowing the background to subtly show through; this adds a modern aesthetic while highlighting the navigation elements without overwhelming the user.
Consistent with this philosophy, buttons and controls throughout the remainder of the interface adhere to a **minimalist, white-on-white aesthetic**. By keeping these secondary interactive elements subtle, we prevent them from competing for visual attention, ensuring that the user's focus remains fixed on the high-contrast data visualizations.

Figure 12: The navigation bar with glassy buttons.

4.3.2 Commodities (Fish Species)

Commodities are encoded primarily through color and iconography to facilitate rapid identification. Guided by the classification established in Section 3, we applied a semantic color scale to the fish species: *legal* species are rendered in green, *suspect* in orange, and *illegal* in red. This scheme leverages pre-attentive processing and common cultural associations (e.g., red for danger) to intuitively convey the risk level of a cargo. Throughout the interface, this color coding extends to specific icons, effectively distinguishing between species even in dense visualizations.

Status	Color Encoding	Icon
Legal	(`#10B981`)	
Suspect	(`#F59E0B`)	
Illegal	(`#EF4444`)	

Table 8: Commodity status color encoding.

4.3.3 Map Zones

To reduce cognitive load, the geographic zones within the Oceanus map employ a **semantic color mapping** strategy. This approach aligns visual encoding with the user's natural mental model of the environment, enabling intuitive recognition without constant reference to a legend. For example, a vessel crossing into a "green" zone instinctively signals entry into a protected nature preserve, while a "blue" zone implies standard fishing grounds.
Crucially, the palette was calibrated to ensure high contrast against the white background of the **Trajectory Analyzer** page. To distinguish the active focus during analysis, selected zones are highlighted in a warm amber, providing immediate visual feedback against the cool tones of the base map.

4.3.4 Vessel Types

A categorical color scheme was employed to differentiate between vessel types. The primary design constraint for this variable was to ensure distinctness from the commodity palette used in the **Harbor Inspector** page to avoid visual confusion. Consequently, we selected a set of high-contrast hues that are visually separable from the red-orange-green spectrum used for fish species, ensuring that vessel types remain easily distinguishable even when overlaid with cargo data. The assigned colors are as follows:

Zone Type	Color Encoding
Fishing Ground	(fill: #3B82F6, border: #1d4ed8)
Ecological Preserve	(fill: #0C875E, border: #044431)
Buoy	(fill: #A94700, border: #732100)
City	(fill: #575F6C, border: #2d3642)
Island	(fill: #DCD5C5, border: #9a8c73)
Selected Zone (on map)	(fill: #fbbf24, border: #f59e0b)

Table 9: Color encoding for geographic zones.

Vessel Type	Color Encoding
Cargo	(#6366F1)
Fishing	(#06B6D4)
Ferry (Passenger)	(#D946EF)
Ferry (Cargo)	(#8B5CF6)
Tour	(#FB7185)
Research	(#84CC16)
Other	(#334155)

Table 10: Vessel type color encoding.

5 Use Case Example

This section presents a detailed use case example demonstrating how an analyst might utilize the visual analytics system to investigate illegal fishing activities. Given that the starting point for analysis can vary widely based on the specific questions and hypotheses an analyst wishes to explore, we outline a hypothetical scenario that showcases the system's capabilities across all views. We address the following research question:

> *Following recent enforcement actions against known illegal fishing fleets, has the illegal fishing activity effectively ceased?*

Phase 1: Anomaly Detection (Harbor Inspector). The investigation begins in the **Harbor Inspector** view. By analyzing the seasonal trends in cargo exports and vessel arrivals across all harbors, the analyst notices anomalies in the export volumes of regulated fish species post-May 2035, a period following the enforcement actions against known illegal fleets (e.g., SouthSeafood Express Corp).
By filtering the cargo types to focus on *regulated species* (specifically *Offidiaa/Pisces osseus* as shown in Figure 13), the analyst observes that despite the interdiction, significant exports of these species persist. To pinpoint the source, the analyst zooms into **South Paackland**, a port historically less associated with illegal operations. Brushing over an export peak in mid-October reveals a corresponding arrival event. The **Suspected Vessels** sidebar panel immediately flags a new set of IDs associated with this high-risk cargo. The analyst selects the primary suspect from the list: the *Sockeye Salmon Seeker*.

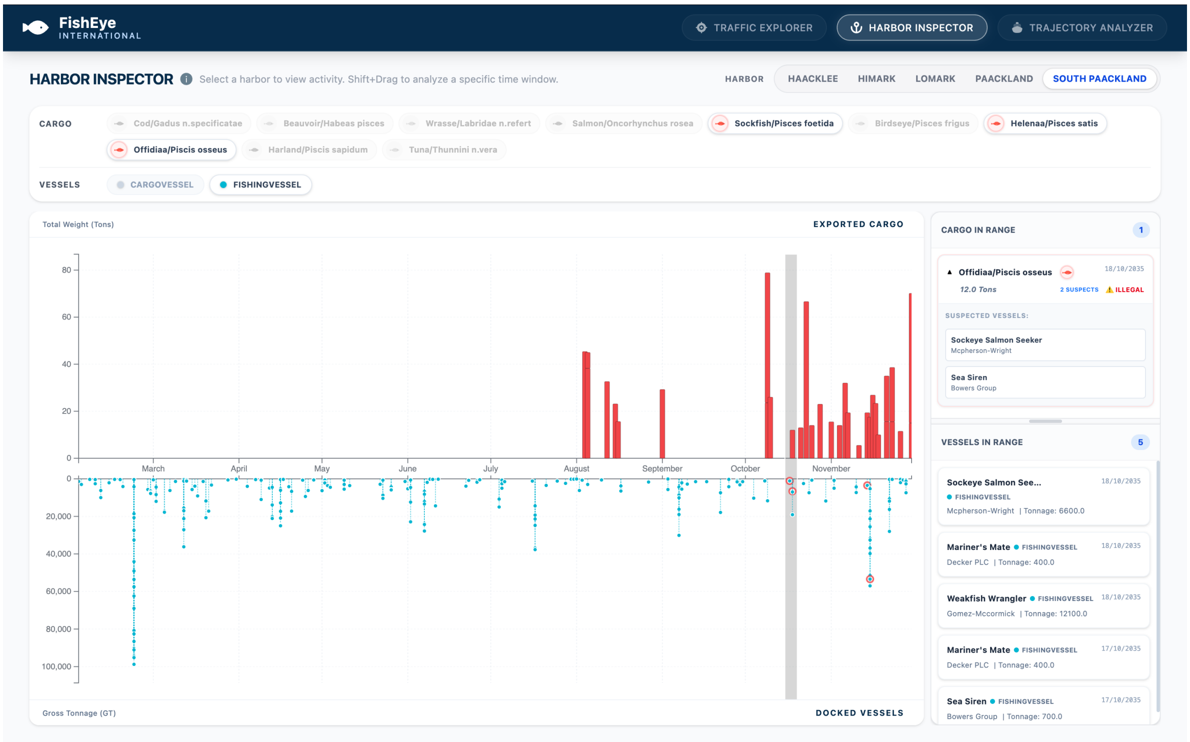


Figure 13: Harbor Inspector view for South Paackland. Filtering for regulated *Offidiaa* reveals spikes in illegal exports (red bars) in late 2035. The system identifies *Sockeye Salmon Seeker* as the carrier.

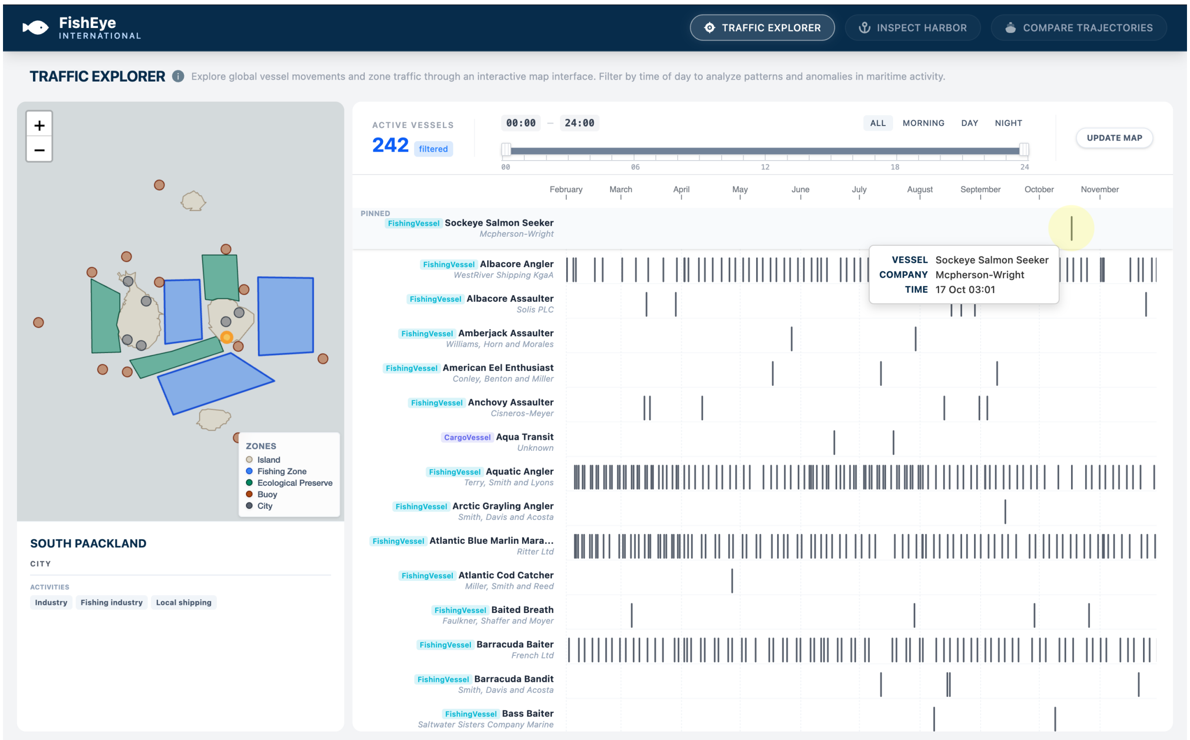


Figure 14: Traffic Explorer view, highlighting the singular and anomalous docking event of the *Sockeye Salmon Seeker* vessel at South Paackland on October 17.

Phase 2: Spatial Filtering (Traffic Explorer). To investigate the movement patterns of the *Sockeye Salmon Seeker*, the analyst switches to the **Traffic Explorer**. By pinning the vessel ID, the

timeline immediately isolates a single, anomalous docking event in South Paackland on October 17 at 03:00 A.M. (Figure 14).
By selecting other zones on the map while keeping the vessel pinned allows to have an idea of spatial trajectory. This reveals a striking operational anomaly. While the vessel concentrates its fishing efforts in the **Cod Table**, it consistently bypasses the nearest logical ports (South Paackland or Paackland) to dock at **Lomark**, located on the opposite side of the ocean region.
This route represents a significant economic inefficiency, increasing transit time and fuel costs without a clear commercial justification. Crucially, this specific trajectory forces the vessel to traverse the restricted **Ghoti Preserve**. Unlike compliant vessels that dock locally to maximize efficiency, this vessel appears to use the pretext of long-distance transit to justify its presence in the restricted zone. This hypothesis is strengthened by the data: pings inside the Ghoti Preserve cluster intensely starting in late August, suggesting that the “transit” is merely a cover for opportunistic illegal fishing during peak season.

Phase 3: Behavioral Verification (Trajectory Analyzer). To distinguish navigational necessity from intentional poaching, the analyst transitions to the **Trajectory Analyzer** view. They load the *Sockeye Salmon Seeker* alongside a reference compliant vessel, the *Aquatic Angler*, chosen for its historically consistent and predictable navigation patterns through Cod Table and South Paackland. As illustrated in Figure 15, the contrast is stark. Unlike the reference vessel, which maintains a direct and predictable route between Cod Table and the nearest ports without entering restricted zones, the *Sockeye Salmon Seeker* displays a recurring illicit pattern between August and November. The vessel leaves Lomark and ostensibly heads toward the Cod Table but diverts course to loiter within the **Ghoti Preserve** for days and the **Exit East** buoy. This loitering behavior mirrors the exact *modus operandi* of the banned SouthSeafood Express Corp fleet. The pattern is consistent across multiple trips, confirming that the incursions are not navigational errors but calculated illegal fishing operations.

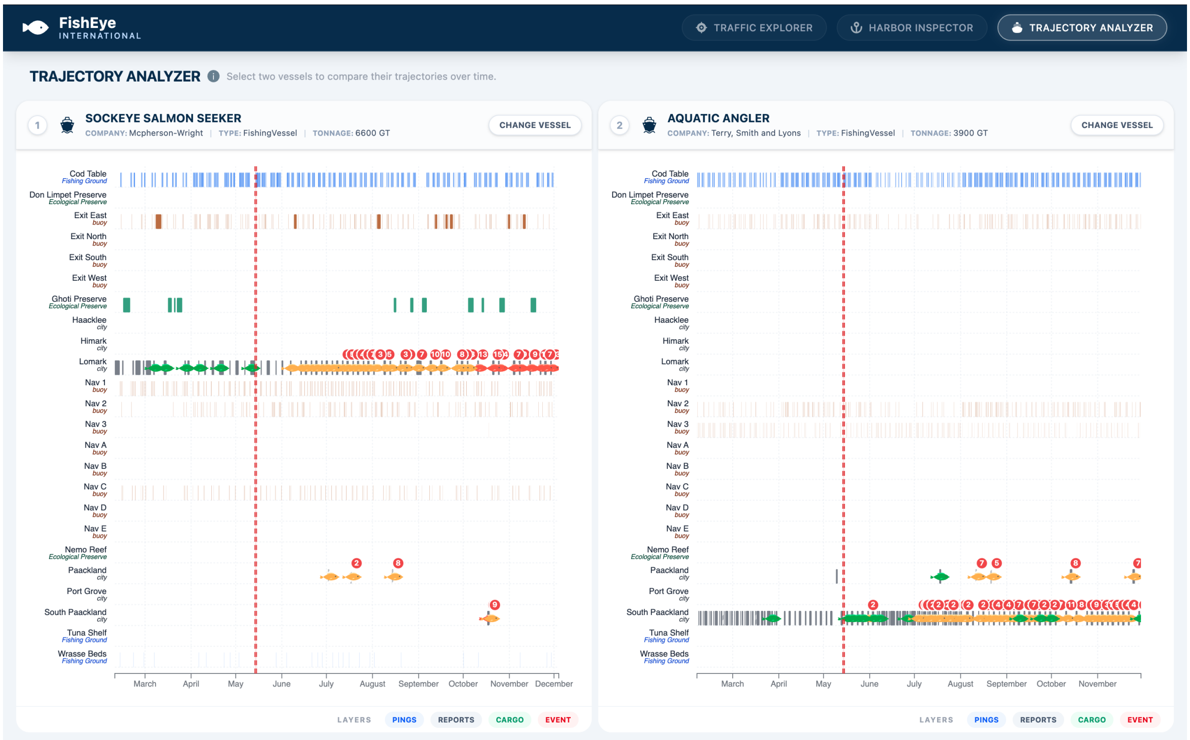


Figure 15: Trajectory Analyzer view. Left: The suspect *Sockeye Salmon Seeker* shows repeated incursions into Gothi Preserve (green) and loitering at Exit East (brown). Right: The compliant *Aquatic Angler* sticks strictly to the legal Cod Table fishing grounds (blue).

Conclusion. By triangulating cargo attribution data in the Harbor Inspector view, spatial anomalies in the Traffic Explorer, and temporal behavioral signatures in the Trajectory Analyzer, the analyst concludes that illegal fishing has not ceased. Instead, it temporarily paused during the early summer before resuming under new vessel identities (such as the *Sockeye Salmon Seeker*) during the peak autumn fishing season.

6 Research Questions

Question 1: *FishEye analysts have long wanted to better understand the flow of commercially caught fish through Oceanus's many ports. But as they were loading data into CatchNet, they discovered they had purchased the wrong port records. They wanted to get the ship off-load records, but they instead got the port-exit records (essentially trucks/trains leaving the port area). Port exit records do not include which vessel that delivered the products. Given this limitation, develop a visualization system to associate vessels with their probable cargos. Which vessels deliver which products and when? What are the seasonal trends and anomalies in the port exit records?*

Answer

To resolve the ambiguity caused by the absence of direct off-load records, we implemented the analytic data association method described in Section 2.2.3. This technique creates a "suspected_vessels" attribute by correlating the timing and tonnage of port-exit records with vessel arrivals, forming the backbone of our Harbor Inspector view.
The central component of this view is the Mirror Plot, designed to simultaneously visualize seasonal trends and isolate probable vessel-cargo connections. The upper axis aggregates daily cargo exports, revealing a distinct seasonal pulse where volumes surge in late summer and peak in October and November. Notably, anomalies in regulated species exports frequently align with these trade peaks, suggesting that illicit actors strategically mask their activities within high-volume periods – a pattern particularly evident in Lomark and South Paackland. Additionally, the visualization captures regional specializations, such as Haacklee's distinct dominance in the Tuna market.
The lower axis addresses the attribution challenge by plotting vessel arrivals aligned with these export events. By integrating a color-coded alert system, the visualization distinguishes standard traffic from high-risk vessels (highlighted in red). This visual linkage bridges the granular data gap, allowing analysts to trace a specific cargo batch back to a limited set of suspect vessels, which can then be forensically verified in the Trajectory Analyzer view (Section 4.2.3).
Figure 16 illustrates the Mirror Plot for South Paackland Harbor, showcasing both the seasonal export surges and the corresponding vessel associations.

Question 2: *Develop visualizations that illustrate the inappropriate behavior of SouthSeafood Express Corp vessels. How do their movement and catch contents compare to other fishing vessels? When and where did SouthSeafood Express Corp vessels perform their illegal fishing? How many different types of suspicious behaviors are observed? Use visual evidence to justify your conclusions.*

Answer

To analyze the behavior of SouthSeafood Express Corp vessels, we utilized the Traffic Explorer for broad spatial verification (Section 4.2.1) and the Trajectory Analyzer view for detailed temporal analysis (Section 4.2.3).
While the automated cargo attribution did not flag SouthSeafood vessels as prominently as other illicit actors, a manual inspection of their trajectories reveals distinct suspicious behaviors when compared to compliant vessels. We premise our analysis on the baseline that compliant fishing vessels generally maintain direct routes, avoiding restricted areas unless strictly necessary for transit.
Snapper Snatcher clearly breaks this compliant pattern. As illustrated in Figure 17, unlike the control vessel (*Whiting Wrangler*), Snapper Snatcher records confirmed presence within the Ghoti Preserve (indicated by the distinct green markers in February and March). Furthermore, the vessel displays erratic routing, abruptly switching fishing grounds from the Cod Table to Wrasse Beds and making unusually extended stops at the "Exit East" buoy. These deviations are inconsistent with standard fishing operations and suggest potential poaching or transshipment activities.
Roach Robber presents a more subtle case (Figure 18). While its route appears more predictable than the one of Snapper Snatcher, it exhibits suspicious loitering at Nav 1, a buoy situated directly

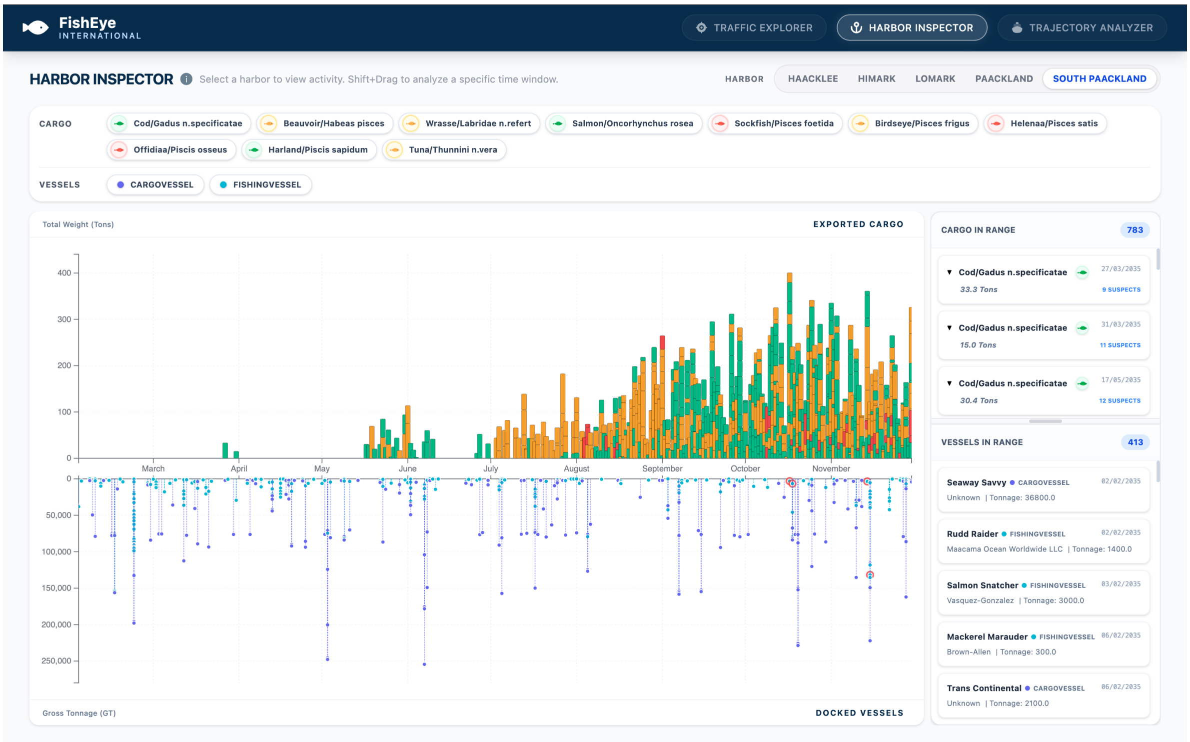


Figure 16: Mirror Plot for South Paackland Harbor showing seasonal export trends and vessel-cargo associations.

over the Gothi Preserve. A compliant vessel (e.g., *Tiger Muskellunge Master*) typically bypasses such markers or adheres to standard navigation channels (e.g., Nav C). This proximity to the preserve boundary, combined with the lack of clear fishing activity elsewhere during those windows, could indicate "edge-fishing" or coordination with others vessels operating inside the restricted zone.
Although the lack of granular catch data prevents a direct comparison of the cargos (as neither Snapper Snatcher nor Roach Robber was attributed an illegal cargo), the spatial anomalies (specifically the preserve incursions and buoy loitering) provide sufficient visual evidence to warrant the deeper investigation detailed in Question 3.

Question 3: *To support further Fisheye investigations, develop visual analytics workflows that allow you to discover other vessels engaging in behaviors similar to SouthSeafood Express Corp's illegal activities? Provide visual evidence of the similarities.*

Answer

To discover other vessels mimicking SouthSeafood Express Corp's illicit strategies, we developed a behavioral profiling workflow based on the "signatures" identified in Question 2: (1) direct incursions into restricted zones (Ghoti Preserve) and (2) non-transit loitering at specific buoys (Exit East).
As a first step, we employed the Traffic Explorer to filter the fleet for vessels exhibiting spatial overlaps with the known SouthSeafood offenders. By pinning the **Snapper Snatcher** as a reference (Figure 19), we visually scanned for vessels that frequented the same sectors during identical timeframes. This broad filtering process highlighted the **Catfish Capturer** as a high-priority suspect due to its overlapping presence in both the Ghoti Preserve and Exit East buoy areas.
We then validated this suspect using the Trajectory Analyzer view (Figure 20). The visual alignment is striking. Prior to the interdiction (left of the red dotted line), Catfish Capturer mirrors Snapper Snatcher with high precision. Both vessels record presence in the Ghoti Preserve (green markers) during late February and May, followed by identical loitering patterns at the Exit East buoy (brown markers), suggesting a coordinated protocol or shared illegal route.
Finally, we examined post-interdiction behavior to assess the impact of enforcement actions. With the banning of SouthSeafood Express Corp in May, Catfish Capturer drastically altered its behavior. As seen in the right half of Figure 20, the vessel immediately ceased all Ghoti Preserve and Exit East activity. Instead, it shifted entirely to legitimate operations, recording dense fishing activity in the

Figure 17: Comparison of Snapper Snatcher (SouthSeafood Express Corp) showing incursions into Ghoti Preserve vs. the compliant Whiting Wrangler.

Wrasse Beds (indicated by the continuous blue blocks at the bottom of the chart). This may suggest that the Catfish Capturer vessel was successfully deterred and returned to compliant fishing.

Question 4: *How did fishing activity change after SouthSeafood Express Corp was caught? What new behaviors in the Oceanus commercial fishing community are most suspicious and why?*

Answer

The Trajectory Analyzer view enables a temporal analysis of fishing activity before and after the interdiction of SouthSeafood Express Corp (marked by the vertical red dotted line). Post-interdiction, we observe a notable shift in fishing patterns across the Oceanus fleet. Several vessels exhibited increased caution, avoiding previously frequented areas such as the Ghoti Preserve. However, traffic in the broader region did not diminish; instead, it redistributed, suggesting that illegal activities persisted and adapted to the new enforcement landscape. Two primary suspicious behaviors emerged: displacement to less-monitored protected areas and the adoption of "edge-fishing" tactics.

As shown in Figure 21, vessels that previously ignored the Ghoti Preserve suddenly began targeting the Nemo Reef Ecological Preserve immediately after the crackdown. The **White Marlin Master** (Figure 21, left) and the **Marlin Master** (Figure 22, right) both show a distinct lack of preserve activity prior to the event, followed by repeated incursions into Nemo Reef afterward. This suggests a coordinated shift to exploit a less-monitored protected area.

Other vessels adopted tactics to avoid inspection. The **Blue Marlin Bandit** (Figure 21, right) completely ceased docking at ports (Himark/Haacklee) after the interdiction. Instead, it shifted to the Wrasse Beds and began loitering at Nav C, likely awaiting transshipment vessels to offload catch without entering a harbor.

Similarly, the **Bluefin Tuna Bandit** (Figure 22, left) adopted "edge fishing" tactics. After the crackdown, it began frequenting Nav 1, a buoy positioned directly on the boundary of the Ghoti Preserve. This suggests the vessel is skimming the edge of the restricted zone, dipping in only briefly or luring fish out, essentially testing the limits of the new enforcement protocols.

7 Conclusion and Reflections

This report has detailed the design, implementation, and application of a visual analytics system aimed at investigating illegal fishing activities within the fictional nation of Oceanus. Through a

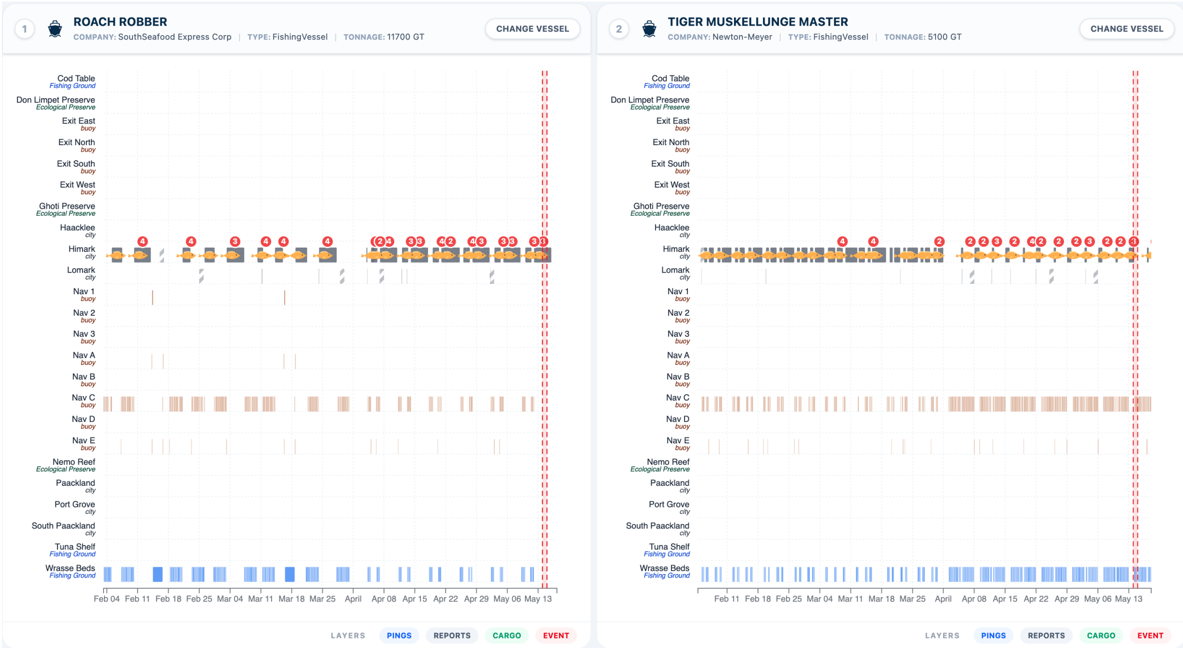


Figure 18: Comparison of Roach Robber (SouthSeafood Express Corp) showing suspicious loitering at Nav 1 vs. a compliant vessel.

combination of novel visualization techniques and a user-centered design approach, we have developed a platform that effectively addresses the complex questions posed by the VAST Challenge 2024.

Reflecting on the development process, the primary challenge lay in the ambiguity and scale of the provided datasets. The lack of direct vessel-to-cargo linkages required not just data visualization, but active data reconstruction, which we addressed through inferred associations as described in Section 2.2.3. Furthermore, the sheer volume of vessel traffic and port transactions necessitated a design that could distill vast amounts of information without overwhelming the analyst. Our solution, splitting the interface into distinct conceptual views (Traffic, Harbor, Trajectory), proved essential. This compartmentalization established a workflow that moves from broad surveillance to granular forensic analysis, managing the data's complexity effectively.

Looking forward, the platform could be enhanced through tighter integration of these views. Currently, identifying the "geographical justification" for a vessel's stop requires the user to mentally bridge the gap between the static Traffic Explorer map and the linear Compare Trajectories timeline. Future iterations should implement a unified geospatial-temporal interface to visualize the proximity, periodicity, and spatial context of stops directly within the trajectory view, reducing the cognitive load of cross-referencing.

Additionally, the system's capabilities could be augmented by predictive analytics. By integrating machine learning-based clustering, the system could automatically categorize vessel behaviors, while outlier detection algorithms could flag anomalies in port-exit records without requiring manual filtering by the user.

Finally, this project offered a significant learning opportunity. While the initial complexity of the dataset presented a steep learning curve, the process of untangling these records was highly rewarding. It reinforced a the fundamental principle that effective visual analytics is not just about creating appealing graphics, but about crafting tools that empower users to derive insights from complex data. And I must say, it was quite fun to play detective in the process of uncovering illicit fishing operations in Oceanus!

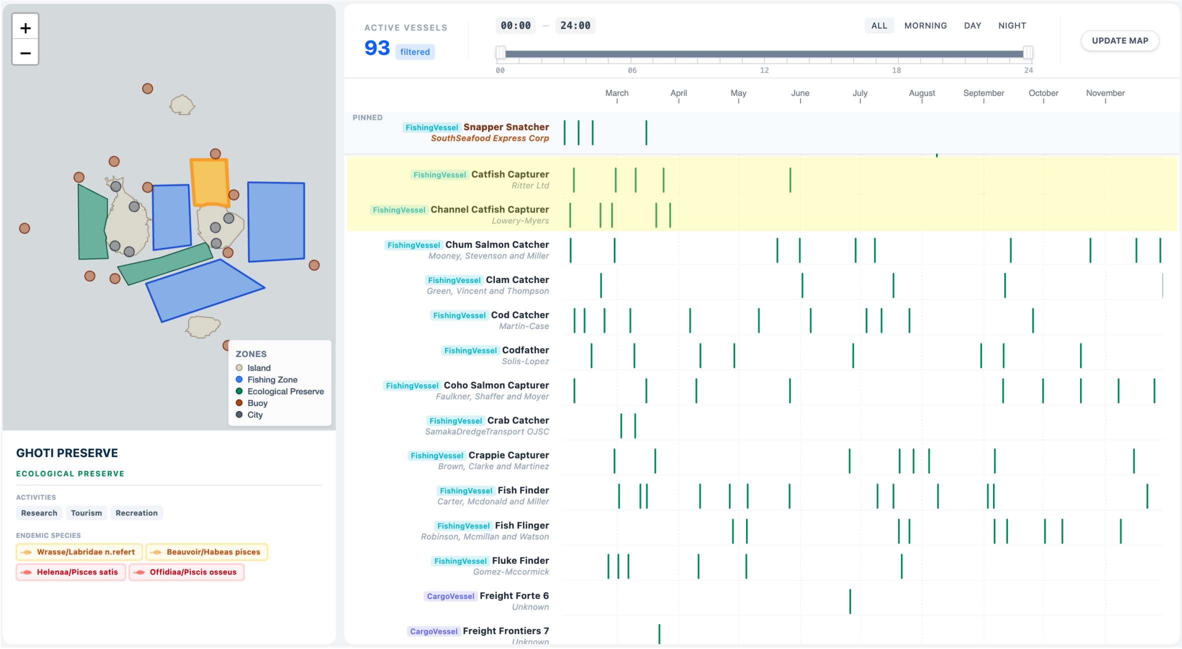


Figure 19: Traffic Explorer filtering process. Highlighting Catfish Capturer (yellow) based on spatial overlaps with the pinned Snapper Snatcher.

Figure 20: Visual Evidence of Similarity and Reform. Left: Snapper Snatcher (illegal reference). Right: Catfish Capturer mirrors the illegal Ghoti/Exit East stops (green/brown markers) before the crackdown, but shifts to legal Wrasse Beds fishing (blue bars) immediately after.

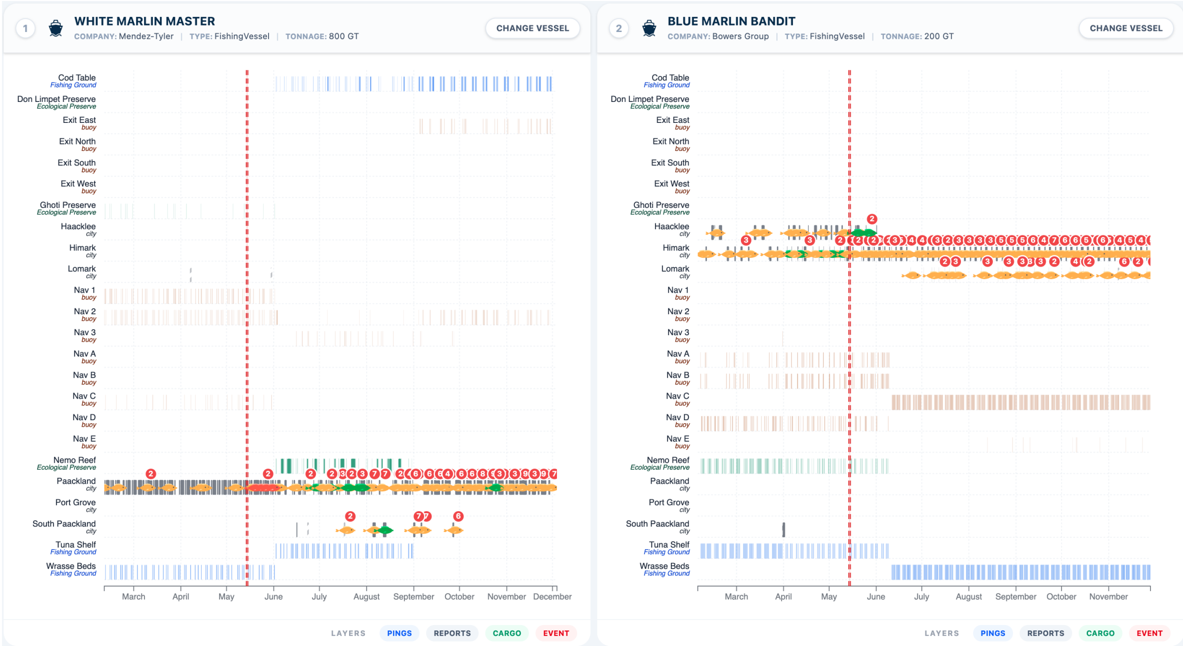


Figure 21: Behavioral divergence post-interdiction. Left: White Marlin Master shifts operations into Nemo Reef. Right: Blue Marlin Bandit ceases all port calls to loiter at Nav C.

Figure 22: Evasive tactics. Left: Bluefin Tuna Bandit begins "edge fishing" at Nav 1 (boundary of Ghoti Preserve). Right: Marlin Master confirms the fleet-wide shift to poaching in Nemo Reef.