# Visual Analytics Report

Davide Rizzello

Mat No.: 12-345-678

2025-11-11

# Table of Contents

# 1 Introduction

This report presents the work developed for Mini-Challenge 2 of the VAST Challenge 2024, proposed as part of the Visual Analytics course.

The challenge revolves around **FishEye**, an organization dedicated to combating illegal and unregulated fishing. Their system, **CatchNet**, aggregates multiple data sources—including vessel movements, harbor reports, and product exports—to support the identification of suspicious fishing activity across Oceanus.

The goal of this project is to design and implement a visual analytics tool that enables analysts to understand, explore, and detect signatures of illegal fishing behavior through the integration of temporal, spatial, and relational data. Analysts must use this system to explain how SouthSeafood Express Corp carried out its illegal activities and then develop a workflow to detect other vessels exhibiting similar patterns.

The challenge poses the following key questions:

1. How can vessel movements, port transactions, and export data be combined to associate vessels with their probable cargo and understand seasonal trends and anomalies?

2. How can visual evidence illustrate the inappropriate or illegal behavior of South-Seafood Express Corp vessels?

3. How can these analytical workflows be generalized to find other vessels with similar suspicious activity patterns?

The proposed system aims to address these questions through a combination of interactive dashboards and coordinated multiple views that guide the analyst through the data exploration process.

# 2 State of the Art

Initial design decisions were inspired by the Martini Glass narrative structure, a visual storytelling method where a guided narrative progressively transitions into open exploration. However, the system developed in this project integrates this concept with a slideshow-based storytelling paradigm, where each analytical question corresponds to a dedicated view (or "slide") that combines narrative guidance with interactive controls.

As a result, the tool supports both structured reasoning (guided story flow) and exploratory analysis (free manipulation of visual parameters), offering a balance between narrative coherence and analytical flexibility.

# 3 Data and Preprocessing

## 3.1 Data Description

The *CatchNet* dataset represents a directed multi-graph of vessel activities, port operations, and commodity flows across Oceanus. It includes roughly 5,600 nodes and 270,000 edges, provided in JSON format via NetworkX's `node_link_data()` structure.

Nodes correspond to four main entity types:

- **Vessels** – described by attributes such as name, company, and tonnage;
- **Locations** – including ports, preserves, and regions with associated fish species;
- **Fish commodities;**
- **Delivery documents.**

Edges capture three categories of events:

- **Transponder pings (OVLS):** link vessels to locations with temporal attributes ( time, dwell);
- **Harbor reports:** record vessel sightings reported by harbor masters;
- **Transactions:** describe fish deliveries and movements between locations and commodities.

This multi-source integration forms a single connected graph that supports spatial–temporal and relational analysis of vessel behavior.

## 3.2 Data Preparation

Given the heterogeneity of the dataset, a preprocessing phase was essential to ensure consistency and usability. The JSON graph was first split into two main components:

1. a **nodes DataFrame** containing all entities;

2. a **links DataFrame** containing all relationships and events.

Each record in the links DataFrame includes a `type` attribute indicating the source subsystem. Using this, the edges were further divided into three specific DataFrames:

1. **Ping Data:** extracted from `Event.TransponderPing`;

2. **Harbor Report Data:** extracted from `Event.HarborReport`;

3. **Transaction Data:** extracted from `Event.Transaction`.

Similarly, nodes were split by entity type into:

- **Vessels;**
- **Fish species;**
- **Delivery documents;**
- **Locations.**

After the split, redundant attributes such as `_date_added` and `_last_edited_by` were removed. Duplicate records were dropped, missing values were filled or filtered, and textual inconsistencies (e.g., naming mismatches) were standardized. Therefore, incorrect records were removed, such as vessels with negative tonnage or pings where the source and target locations were identical. This produced a clean, normalized relational structure ready for merging and analysis.

## 3.3 Enrichment and Integration

Each of the three edge tables was enriched by joining relevant information from the corresponding node tables. For instance:

- **Ping data** were augmented with vessel tonnage, company, and location type.
- **Harbor reports** were integrated with descriptive location attributes and vessel metadata.
- **Transactions** were enriched with fish species details and delivery identifiers.

This enrichment allowed for analytical joins across subsystems—linking vessel behavior, harbor activity, and commodity flow. The resulting unified tables form the analytical backbone of the visualization system, enabling temporal aggregation, spatial mapping, and cross-domain reasoning between vessel activity and fish trade dynamics.

# 4 Design

## 4.1 Page Layout and Narrative Flow

The system is organized into two interactive investigation pages, each corresponding to one or more of the main analytical questions posed by the challenge. The layout follows a narrative flow inspired by slideshow-based storytelling, where the analyst progresses through successive views while maintaining full interactivity at every stage.

**Investigation 1 – Cargo Association.** This view addresses the question "Which vessels deliver which products and when?". The interface allows users to select a port and a date to filter relevant events. A sequence of coordinated panels then guides the analysis:

1. **Routes & Dwell Time:** A bubble chart visualizes dynamic vessel dwell times, calculated from the vessels' movement data between the city of departure and the arrival port. Bubble size encodes the dwell duration, while color differentiates protected, non-protected, and transit areas.

2. **Declared Exports:** A bar chart displays the exported tons per fish species for the selected day, highlighting prohibited species in red.

3. **Estimation Model:** A textual block explains how cargo weight is proportionally distributed among vessels using a tonnage–dwell weighting scheme.

4. **Cargo–Vessel Distribution:** A horizontal bar chart shows the resulting cargo allocation by vessel, distinguishing legal and prohibited catches.

Together, these panels allow analysts to connect export records to their most probable source vessels.

**Investigation 2 – Suspicious Activity Patterns.** This view focuses on the behavior of vessels belonging to *SouthSeafood Express Corp* and aims to identify signs of illegal activity. The analysis is supported by two main visualizations:

1. **Routine Timeline:** A multi-category timeline visualizes the vessel's movements over the year, with color coding indicating area type (fishing ground, preserve, city, buoy, etc.). Analysts can compare each vessel with the one of SouthSeafood Express Corp that has been caught to detect deviations from normal patterns.

2. **Behavioral Comparison:** A line plot compares the average dwell time per zone against the fleet's baseline. The shaded band represents the normal operating range (25th–75th percentile), while the vessel's own curve highlights deviations that may indicate suspicious or illegal behavior.

**Narrative Structure.** Both investigation pages combine guided storytelling with exploratory flexibility. Users are encouraged to follow the visual sequence top–down (similar to a slideshow) while being able to adjust filters, dates, and vessels interactively. This hybrid design promotes analytical reasoning: each panel introduces a specific aspect of the investigation while contributing to a coherent, progressive narrative across the two pages.

## 4.2 Design Choices

The design of the CatchNet visual analytics system was guided by the need to balance narrative clarity with analytical flexibility. Visual encoding, interaction, and layout were chosen to help analysts interpret complex spatio-temporal relationships without cognitive overload.

**Color.** Colors were selected to maintain semantic consistency across all views. Protected zones are shown in **red**, non-protected in **blue**, and neutral or transit areas in **gray**. Illegal or prohibited catches are highlighted using red tones to draw attention to violations, while legal operations use muted blue tones.

**Shapes and Encodings.** Shape and size encode meaning in a compact and intuitive way. In bubble plots, circle size represents dwell duration, while position corresponds to

vessel–location relationships. Bar lengths indicate export or cargo proportions, and their color encodes legality or zone type.

**Interactions.** Each visualization supports **linked interactions**: selecting a port, date, vessel, or cargo updates all dependent charts dynamically. Hover tooltips reveal contextual details such as tonnage, dwell time, or species legality. A date picker and dropdown filters guide the user's focus while still enabling exploratory control. The interface is designed to encourage hypothesis testing through comparison rather than static observation.

# 5 Detailed Description of the Visualization and Interactions

The CatchNet system integrates multiple coordinated visualizations to support reasoning about vessel behavior, cargo delivery, and potential illegal activities. Each visualization is both explanatory and exploratory, enabling analysts to trace entities, detect anomalies, and validate hypotheses through direct interaction.

## 5.1 Visualization Overview

The dashboard is divided into two investigation pages:

- **Investigation 1 – Cargo Association:** links exports to vessels and explores temporal and spatial relationships between fishing activity and port deliveries.
- **Investigation 2 – Suspicious Behavior:** focuses on abnormal dwell patterns, restricted-area operations, and fleet-level behavioral comparison.

Each page combines filters (date, port, vessel), interactive plots, and narrative explanations arranged top–down to guide the analytical process.

**Visualization Types.**

- **Bubble Chart (Routes & Dwell Time):** Displays the dwell time of each vessel per location. The x-axis lists vessels, the y-axis lists locations, bubble size encodes duration, and color denotes zone type (protected, non-protected, transit). This view helps identify unusually long stays or illegal presence in protected areas.

- **Bar Chart (Declared Exports):** Shows exported tons by fish species for a given day and port. Colors distinguish legal versus prohibited species, allowing a quick assessment of potential violations.
- **Cargo–Vessel Distribution:** Depicts how the estimated cargo weight is distributed among candidate vessels, based on the tonnage–dwell model.
- **Timeline (Routine Activity):** Represents vessel presence across different zones throughout the year. Each line corresponds to a zone; colored segments indicate type of area (fishing ground, preserve, city, buoy). Enables direct comparison of normal and suspicious trajectories.
- **Behavioral Comparison Plot:** A line chart compares the selected vessel's dwell time per zone against fleet averages. The shaded area (25th–75th percentile) defines the "safe zone," while deviations above or below suggest abnormal behavior.

## 5.2 Interactions

All visualizations are **linked and coordinated**, allowing analysts to propagate selections across views:

- Selecting a **port or date** filters all related datasets and updates the corresponding charts.
- Tooltips reveal contextual metadata (e.g., vessel name, company, tonnage, dwell time, protected status).

This interactivity allows users to seamlessly move between overview and detail, focusing on emerging anomalies without losing global context.

## 5.3 Analytical Use Case Example

To illustrate the workflow, consider an analyst investigating a suspected illegal delivery at the **City of Paackland** on November 15, 2035.

1. The analyst selects the port "City of Paackland" and the date `15/11/2035` in the filter panel. The bubble chart updates (Figure: 1), revealing multiple vessels with

prolonged dwell in surrounding areas. One vessel shows a long presence inside a protected zone (red bubbles).

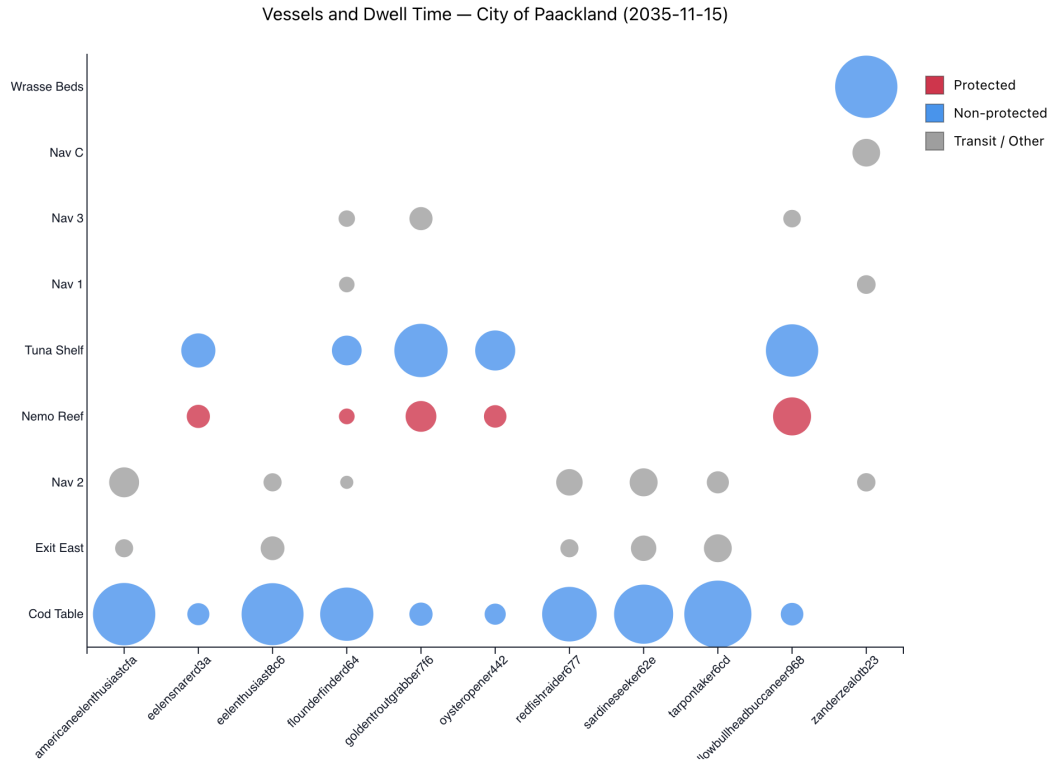


Figure 1: Dwell time for each zone in the route to the port

2. The "Declared Exports" bar chart (figure 2) shows that *Prohibited Fish* were unloaded the next day, suggesting possible illegal capture.

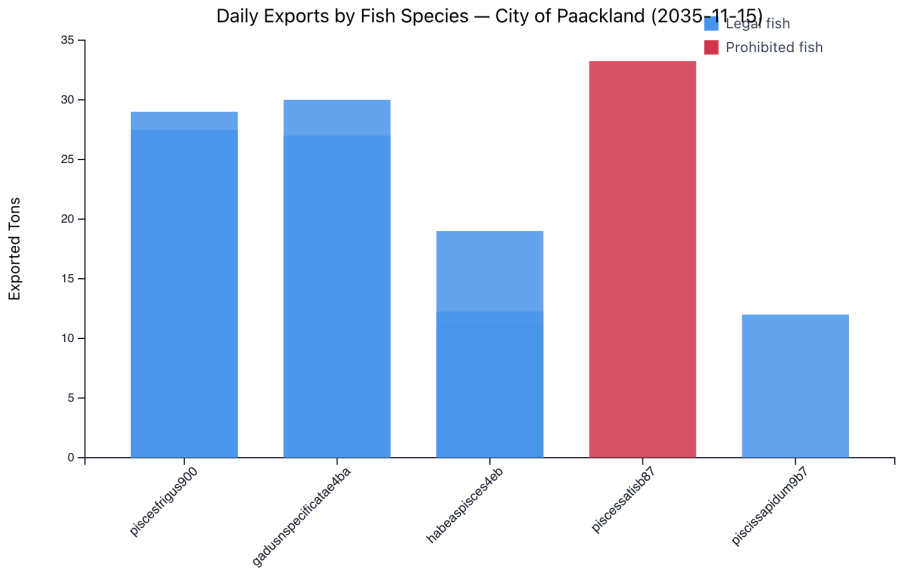


Figure 2: Exported tons by each fish_id

3. The analyst scrolls to the "Estimation Model" section, which allocates cargo to vessels based on tonnage and dwell time. The algorithm indicates that the same vessel (figure 3) from the protected area contributed the majority of the exported volume.

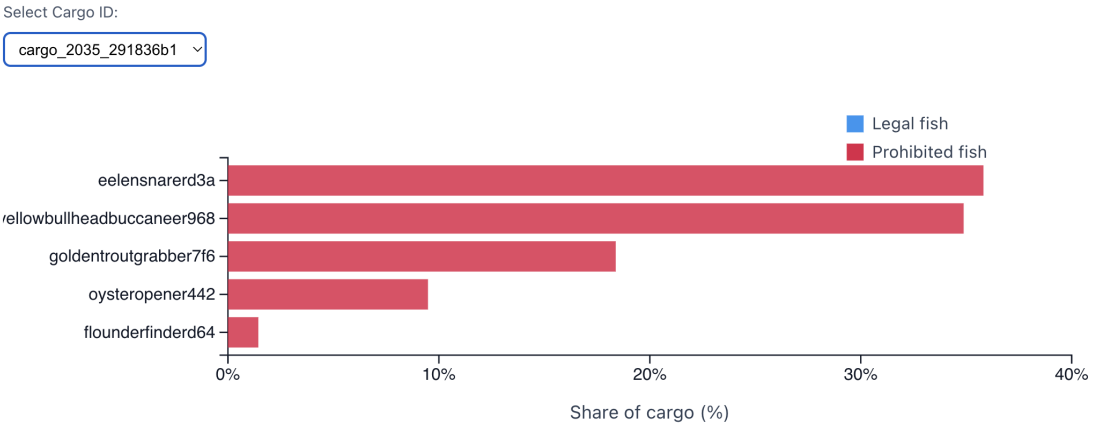


Figure 3: Share (probability) of selected Cargo by each vessel

4. Moving to **Investigation 2**, the analyst selects that vessel from the dropdown. The timeline (figure 4) view shows recurring activity within ecological preserves across multiple months, while the comparison plot highlights abnormally high dwell durations relative to fleet averages.

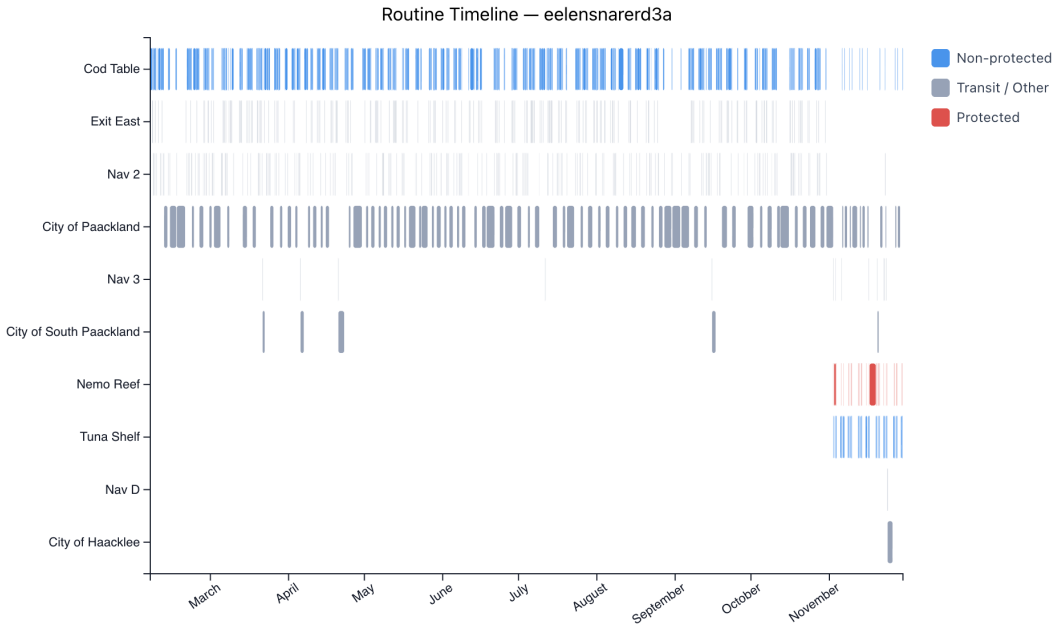


Figure 4: Annual travel pattern of vessel eelensnarerd3a, illustrating dwell times across zones and protection types.

5. Combining these visual cues, the analyst concludes that this vessel exhibited behavior consistent with illegal fishing, both spatially (presence in restricted areas) and temporally (extended dwell far beyond fleet norms).

# 6 Answers to the Challenge Questions

This section summarizes the main analytical findings derived from the CatchNet visual analytics system in response to the four challenge questions.

## 6.1 Vessel–Cargo Association and Seasonal Trends

The first investigation addressed the question: *"Which vessels deliver which products and when?"*

Because the available port–exit records lacked explicit vessel identifiers, the system estimated vessel–cargo associations through a weighted model based on **tonnage** and **dwell time** within relevant fish habitats. For each exported cargo, all vessels that operated in compatible fishing zones prior to the export date were considered potential contributors.

The cargo's total exported weight was then distributed proportionally according to:

$$estimated\_tons_{v,f} = exports\_tons_f \times \frac{(tonnage_v \times dwell_v)}{\sum_i (tonnage_i \times dwell_i)}$$

This approach permitted assigning a share or probability of contribution for each vessel–cargo pair, effectively linking vessels to their most likely exports despite the absence of direct identifiers.

## 6.2 SouthSeafood Express Corp – Illegal Behavior Analysis

The second investigation focuses on *SouthSeafood Express Corp*, a company repeatedly linked to suspicious activities. Through the timeline and dwell-comparison views, several behavioral anomalies were detected:

- **Spatial violations:** their vessels, such as *snappersnatcher7be*, repeatedly entered ecological preserves (Ghoti preserves in February) and remained there for extended periods, despite legal restrictions.
- **Temporal irregularities:** dwell durations far exceeded the fleet's safe-zone interval (25th–75th percentile) with visible peaks.

Visual evidence therefore confirms that SouthSeafood Express vessels engaged in systematic illegal fishing.

## 6.3 Discovering Other Suspicious Vessels

To extend the workflow beyond known offenders, the system allows analysts to filter the fleet by similarity in **dwell distributions**, **operational zones**, and **export composition**. By comparing vessels whose dwell profiles deviate from fleet norms or overlap with prohibited species exports, several candidates emerged with comparable behavior.

Examples include:

- vessels showing recurrent presence in protected habitats with high dwell
- vessels whose estimated export patterns are dominated by illegal fish types.

This workflow provides a reproducible method for identifying fleets exhibiting potential Illegal activity, supporting continuous surveillance within CatchNet.

## 6.4 Summary

Across all analyses, the integrated visual system enabled analysts to:

1. Reconstruct plausible vessel–cargo linkages despite missing identifiers;
2. Expose concrete evidence of illegal activity by SouthSeafood Express vessels;
3. Generalize detection methods to find other anomalous fleets;