SC/69A/GDR/04

Sub-committees/working group name: GDR

WhaleVis: A new visualization tool for the IWC catch database

Ameya Patil, Zoe R. Rand, Trevor A. Branch, Leilani Battle

WhaleVis: A new visualization tool for the IWC catch database

Ameya Patil[1], Zoe R. Rand[2], Trevor A. Branch[3], Leilani Battle[1]

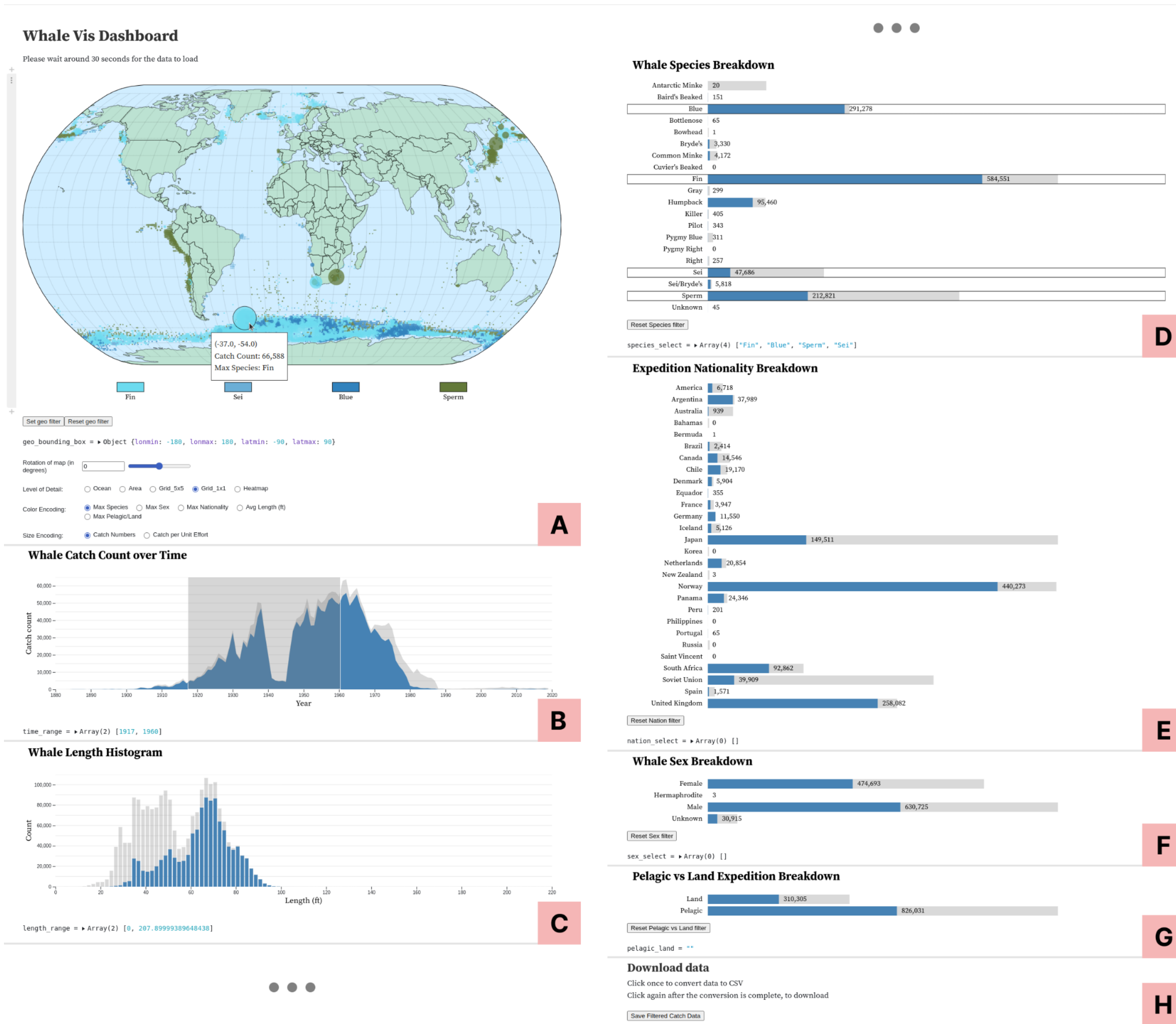

*Fig 1: A snapshot of the WhaleVis dashboard. **(A)** Geographic map visualization for understanding spatial distribution of whaling events or whale populations, with the option to encode different data attributes in the circle color, and view the data in 4 different levels of aggregation (levels of detail). Also acts an interface for setting geographical latitude-longitude filter; **(B)** Area chart for understanding temporal trends in whaling; **(C)** Histogram for understanding distribution of whale lengths; **(B)** and **(C)** also act as interactive widgets to set time and length filters respectively; **(D)** Bar chart for understanding breakdown of catches by species; **(E)** Bar chart for understanding breakdown of catches by expedition nationality; **(F)** Bar chart for understanding breakdown of catches by sex; **(G)** Bar chart for understanding breakdown of catches by expedition type. **(D)**, **(E)**, **(F)** and **(G)** also act as an interface to set a filter for a desired attribute value; **(H)** Option to download the filtered data in .csv format. **(B)** and **(D)** show example filters set on time and species respectively. The unfiltered data is shown in gray, while the filtered data is shown in blue.*

[1] Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle WA, 98195, USA
[2] Quantitative Ecology & Resource Management, University of Washington, Seattle WA, 98195, USA
[3] School of Aquatic and Fishery Sciences, University of Washington, Seattle WA, 98195, USA

Abstract

We present WhaleVis—an interactive visualization and analysis dashboard for the IWC catch database, with the goal to facilitate understanding whaling events and the resulting changes in whale population distribution over the years when whaling was rampant, and prioritize management and conservation efforts for populations under the threat of extinction today. We first canvassed a sample of questions that can be answered using the catch database, and then designed and implemented our dashboard to facilitate answering these questions. Our tool is implemented using the Observable[4] notebook environment, which ensures data and code transparency and facilitates collaborative analysis. We provide a set of visualizations of the catch data, which enable both overviewing the data and filtering down to attributes of interest, using intuitive interactions with the visualization itself. In addition to visualizations and the ability to extract selected portions of the database in comma-delimited (.csv) format, we also introduce users to new perspectives of the data. Notably, we plan to use the information about the timeline and routes of the expeditions contained in the database to (1) model the data in terms of a network graph representation, where the nodes represent locations of whale sightings, and the edges represent whaling expedition routes, and (2) use the network graph representation to plot population maps normalized by search effort i.e. catch-per-unit-effort (CPUE). This would further open up new avenues for performing graph analysis on the data. We demonstrate the use of our dashboard through three example use cases.

1. Introduction

In order to effectively manage and conserve whale populations across the globe, we need a means of knowing where whales were before commercial whaling started disrupting their numbers, and where they are today. The latter question is answered by whale population monitoring efforts carried out via visual (Scarpaci et al., 2008) or acoustic monitoring (Kowarski et al., 2021) and are further augmented with extrapolations from statistical models (Hazen et al., 2017, Bamford et al., 2020). However, we still need to know where whales were caught, which would enable us to establish a clear baseline for whale population distribution before the advent of commercial whaling, as well as improve the accuracy of statistical models of the affected whale populations. The catch database originally collated by the Bureau of International Whaling Statistics (BIWS) and currently maintained by the International Whaling Commission (IWC) (Allison et al., 2020 s, Allison et al., 2020 i), can be used for this purpose.

The IWC catch database has been a topic of interest for many sub-committees at the IWC (Staniland et al., 2021) owing to its comprehensive coverage of whaling events over 140 years starting from the late 19th century, including a variety of whale species across all oceans. It records information about individual whaling expeditions which reported the catches, together with the location, date, species, length, sex, pregnancy status, and fetal information, all of which can be combined to find the distribution of whale catches normalized by the search effort. Currently, the database is available in the format of a series of .csv files, and requires substantial data manipulation to extract subsets, plot locations, and convert data into standard formats (e.g. lengths are recorded in a combination of feet+inches and meters+cm). The data is therefore difficult to analyze due to the lack of an intuitive analysis interface.

In this paper, we present a solution in the form of an interactive visual analysis dashboard, which provides an intuitive interface for understanding the catch data, and mapping catches and whaling search effort. Our dashboard enables users to analyze all whaling events in the current version of the database. Furthermore, the entire data analysis and visualization pipeline is publicly available online, enabling any user to validate our design process and freely reuse any part of the pipeline (data, visualizations, etc.) in future research and communication efforts.

[4] https://www.observablehq.com

Whale mapping visualization dashboards have been created in the past, such as Whale Map[5], Obis-Seamap[6], Happy Whale[7], Whale Test[8] and Pacific Whale[9]. These dashboards visualize the location of whale sightings, but not historical whale catches from the IWC database. Some of these interfaces like Obis-Seamap and WhaleMap can be connected to multiple data sources, but lack the level of transparency to easily enable whale scientists to also plot the complete whaling data. Most of these interfaces also suffer from user experience challenges (1) some fail to provide effective visualization designs, hindering interpretability of the data—Obis-Seamap, WhaleMap; (2) some fail to support easy to use interactive querying of the catch data—HappyWhale, Obis-Seamap, PacificWhale, WhaleTest; and (3) some are too slow to support the full catch data and take many seconds to respond to user interaction events—HappyWhale, Obis-Seamap, WhaleMap;. There also exist generic wildlife mapping interfaces like iNaturalist[10], which can connect to multiple data sources of animal tracking, weather, ground water levels, etc. However, iNaturalist is a generic platform and does not provide a sufficient range of data views and analysis capabilities of interest to whale scientists. Further, the iNaturalist platform is not open to non-experts owing to security concerns for the animals being tracked.

We have tailored our interactive visualization dashboard—WhaleVis—for use by the IWC community. We address the issues of interactivity and intuitiveness or ease of exploration/analysis, while adhering to best practices in data visualization and user interface design. We present a set of visualizations of the catch data which allow both an overview of the data, as well as targeted analysis of select attributes. Further, we leverage the expedition and individual catch details like date and geographic coordinates in the catch dataset to propose a novel representation of the data as a network graph, where nodes in the graph represent locations of whale catches, and edges in the graph represent expedition routes. This kind of modeling can facilitate accurate construction of population maps normalized by the search effort, while also opening up new avenues for performing graph analysis on the data, which can uncover new insights. In the spirit of community values surrounding open access data, we keep the entire pipeline from data cleaning to data transformation to data visualization, transparent so as to make it accessible and reviewable to the users. We do so by using the collaborative visualization and analysis framework of Observable notebook, which allows users to annotate the notebook with their findings or insights for everyone to see. The use of such a dashboard can facilitate new datasets to be added, additional features to be proposed and insights to be shared (Isenberg et al., 2011), through our users. We also demonstrate how IWC researchers may use our dashboard through three example use cases.

Our hope is that we can streamline the process of gathering and understanding the whale catch and catch-per-unit-effort (CPUE) information through WhaleVis. This information can be used to assess which whale populations are depleted to prioritize management actions to conserve them, and increase awareness about whale conservation among the general public.

To summarize, we make the following contributions:

1. WhaleVis, an interactive data visualization tool for the IWC catch database to understand past whaling events and resulting changes in whale population distributions.

2. A transparent and open-access pipeline to enable future development, rigorous scientific analysis, and increased data awareness among both whale experts and non-experts.

3. Demonstrative use cases of analyzing catch data using WhaleVis, showing the potential for its use among the IWC community.

Readers are welcome to interact with WhaleVis through the following links (open only one link at a time):
1. Data wrangling - https://observablehq.com/@whales/iwc-dataset
2. Visualization dashboard - https://observablehq.com/@whales/whale-vis-dashboard

---

[5] https://whalemap.org/whalemap/
[6] https://seamap.env.duke.edu/
[7] https://happywhale.com/browse
[8] https://michellemodest.shinyapps.io/whaletest
[9] https://www.pacificwhale.org/research/hawaii-wdt-sightings-map
[10] https://www.inaturalist.org

2. Design and Development

Following best practices in conducting design studies and building visualization application tools (Munzner, 2009, Sedlmair et al., 2012, Meyer et al., 2019, Ye et al., 2020), we worked with our co-authors (T.A.B. and Z.R.R.) who have extensive experience in aquatic and fishery sciences and in the use of the IWC database, throughout the development process - from understanding the data, to gathering requirements, to getting feedback on the implementation. In this section, we describe the dataset, our development process, the design goals, and the tool design which is informed by these goals.

2.1 Dataset Description

We use the commercial whaling data collected and maintained by the IWC (Allison et al., 2020 s, Allison et al., 2020 i). The current version of the dataset (V7.1), which is actively being updated, covers the time period from 1880 to 2020. It covers all the oceans and has details about whale catches which includes the date, geographic coordinates, species, length, sex of the whales, pregnancy status etc., and details of the corresponding expedition which caught the whale, which includes the expedition code, company name, nationality of the expedition, expedition type, etc. This source data is available as individual catch records and a summary database.

Individual Catch Records V7.1 (Allison et al., 2020 i)

This database contains separate files for each oceanic region. These records were merged together, and records with missing dates, where either the Day or Month were marked as '0', and records with missing geographic coordinates, where the latitude degrees, latitude minutes, longitude degrees and longitude minutes were marked as '0', were removed. This resulted in the removal of 31,366 records, leaving a total of 2,148,279 individual whale catch records, which is 80-90% of all whales caught during the period of modern whaling (late 1800s to present).

Summary Database V7.1 (Allison et al., 2020 s)

It contains details about the expeditions, and additional information on demarcation of different oceanic regions, together with annual totals of the number of whales of each species caught by each expedition, in each whaling season (either a calendar year or the season starting on 1 July each year). The summary database includes all known catches, including many in the earliest years that do not have information about individual catches.

Both the individual and summary databases were joined using the expedition code to gather all the details of the catch record and the corresponding expedition.

2.2 Domain Tasks & Abstractions

A central goal of this work is to provide a visualization tool that meets the analytical needs of whale scientists. One of our coauthors (T.A.B.) was interested in piecing together the paths followed by whaling expeditions, to estimate the number of days each expedition spent in different parts of the ocean, so that catches can be divided by effort to create plots of catch-per-unit-effort (CPUE). These calculations are underway. In addition, through discussions among the coauthors, we identified common tasks that many researchers seek to perform with the IWC data, which informed the design of WhaleVis.

(T1) Understand the distribution of whale catches with respect to space (geographic coordinates) and time.

(T2) Understand how whaling spread from one region to another over time.

(T3) Identify areas where whales were caught in the past, but recent surveys have not sighted any whales of that species in recent years, suggesting a possible extirpation event.

(T4) Identify examples of vagrant whales: whale catches reported far away from clusters of other catches and the known range of existing populations.

(T5) Understand the distribution of whale catches with respect to space and time, normalized by the search effort across space and time, both across all expeditions and for a given expedition.

(T6) Inspect the characteristics of the whales caught in a given region and time period (species, sex and length) or by a particular whaling expedition or country.

(T7) Understand how whale sex ratios vary across time and space.

(T8) Save the data with a given state of filters applied for sharing or further offline analysis, for example to be used in models to assess the current status of particular whale populations.

The above list is not exhaustive but merely represents the different kinds of questions that IWC users may be interested in, based on the dataset. We abstracted these example tasks into the following visual and interactive constructs, which we later use to inform our design goals and consequently the interface design.

C1: Geographic Map based Visualization.

Whale population distribution only makes sense in the context of the geographic locations of the whales. Geographic maps play a central role in visualizing location based data, and form a core part of the WhaleVis interface. These maps partially or completely answer many of the aforementioned tasks. The density of the whale population can be represented as a 2D scatter plot or as a smoothed heatmap.

C2: Navigating the Time Dimension.

Most of the domain tasks (*T1, T2, T5, T6, T7*) require inspecting the population distribution or other attributes within a time-window. Having the ability to set a time stamp or a time window, the data during which can be visualized, would thus be useful.

C3: Details on Demand.

The listed tasks also show a spectrum of analysis questions from *overview* to *details*. For example, *T1, T2, T3 and T5* are concerned with *overview* information as they require a zoomed out or big picture of the data. This includes understanding whale catch distribution, the CPUE information, and the spread of whaling over time over the entire globe. On the other hand, tasks T4 through T7 involve looking at *details* i.e. a small subset of values of select attributes of the dataset. This includes inspecting vagrant whale catches (a sub-region of the globe), catches of a certain species in a particular time window, etc. This calls for interaction constructs which enable the user to view an overview of the data, filter attributes and select data points of choice, to view more details as needed. This is in accordance with how humans generally browse information - "overview first, zoom and filter, details on demand" (Shneiderman 1996). Our interface should facilitate this approach of browsing the data.

C4: Graph Modeling.

Tasks *T2, T5 and T6* ask questions about the whaling expeditions, and the routes taken by them. While the answers to such questions can be pieced together using the aforementioned constructs, modeling the catch locations as nodes and the expedition routes as edges in a network graph can better convey this information. For example, catch counts and CPUE (T5) can be estimated using the map visualization with appropriate filters for geographic coordinates, expedition and time, however this approach misses out on regions through which the expeditions traveled but with no catches recorded. With the data modeled as a graph, the expedition routes can convey the same information in addition to showing the actual routes along with the direction of travel, and any regions through which expeditions traveled but reported zero catches, thus giving a complete picture of CPUE. Graph modeling also opens up new possibilities for analyzing this data; for example, computing centrality measures (measure of importance for a node) for the nodes (catch locations) can tell us quantitatively which regions were hotspots for whaling within a certain time window.

2.3 Design Goals

Based on the overarching theme of analyzing historical whaling data, and the range of tasks that researchers want to perform using the IWC data (T1-T8), we identified the key mechanisms that must be implemented to support completion of these tasks through a user interface (C1-C4). The next step of our implementation process is to integrate these mechanisms with best practices in data science and data visualization to form concrete goals for how WhaleVis should be designed.

G1: Transparency & Sustainability.

The whaling dataset maintained by IWC is public and has been meticulously put together from multiple sources going back as early as the late 19th century, well before the advent of digital bookkeeping. With each iteration of the dataset, it has been consistently shared for further validation and improvement. In the same spirit, we strive to build a system around this dataset that keeps the data and the code similarly transparent. This becomes critical also considering the future scope of additionally using this tool to visualize and analyze whale monitoring data, and sustaining the tool in terms of additional use cases, and scalability in terms of dataset size and computation capabilities. We thus aim to facilitate user participation in improving the tool (Isenberg et al., 2011).

G2: Enable Spatial & Graph Analysis.

This design goal is motivated by the aspect of understanding the spatial distribution of whales and whaling activities, and the resulting visual and interactive constructs C1 and C4. Recognizing the potential of the use of graph analysis on the whaling dataset for computing accurate estimates of CPUE as described in C4, and for enabling new kinds of analysis of the catch data, we propose graph modeling constructs and aim to facilitate basic graph analysis in our tool, in addition to the spatial analysis constructs.

G3: Balance Inclusivity & Usability.

The requirement to support as many different kinds of analyses as possible, and consequently support as many researchers as possible, was a recurring theme in our design planning discussions. With each new use case, we had to accommodate, or enable via interactions, a new layer of information on the limited screen space (inclusivity). This implied providing a large number of visualizations, or trying to encode multiple attributes in a single visualization, with interactive widgets to control the encodings (C3). However, for a visualization dashboard to be truly usable, clutter or overload both in terms of visual marks and in terms of UI elements must be avoided (usability) (Hutchins et al., 1985). We strive to achieve this sweet spot between inclusivity and usability by adhering to best practices of visualization design (Sarikaya et al., 2019, Bach et al., 2022).

2.4 Implementation Details

We now provide implementation details for WhaleVis, its working and usage.

G1: Transparency & Sustainability.

We use the collaborative visualization and analysis environment of Observable notebook for our implementation. This publicly available, online notebook environment makes both the data and the code visible and accessible to the audience for their scrutiny. This includes transformations applied to the data for cleaning and reformatting it for visualization, code for modeling the data as a network graph, and the visualization dashboard implementation itself. We create separate notebooks for the data wrangling and the visualization dashboard, for each analysis approach - spatial and graph. The IWC data files are also attached to the data wrangling notebook and can be downloaded for further offline analysis, Fig 2(a). Similarly, data with the desired filters applied during the analysis/exploration in the dashboard notebook, can be downloaded for offline analysis, Fig 1(H), and corresponding visualization snapshots can be downloaded for sharing or included in reports for dissemination, Fig 2(b). Users can annotate the dashboard by adding publicly visible comments with the insights generated from using the dashboard, or seek help by posting questions, Fig 2(c). The data wrangling notebook also allows for running SQL queries interactively to browse the data, Fig 2(d).

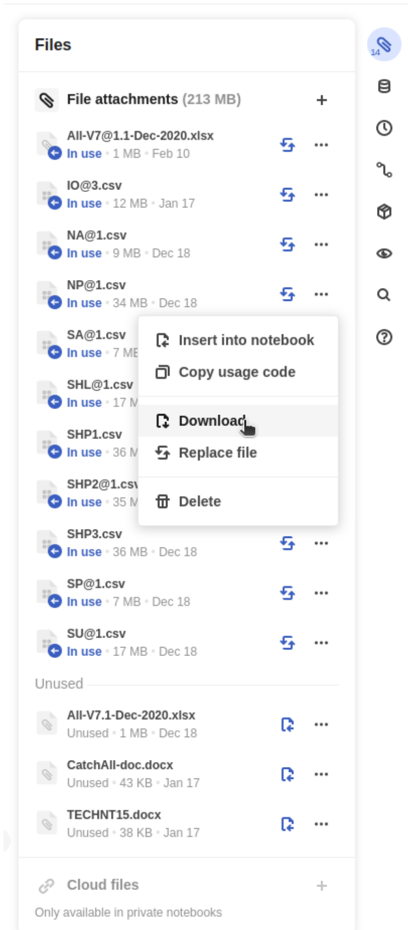

Fig 2(a): Downloading raw data files from the data wrangling notebook.

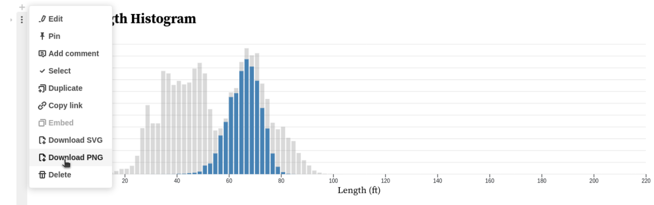

Fig 2(b): Downloading visualization snapshots from the Whale Vis dashboard notebook.

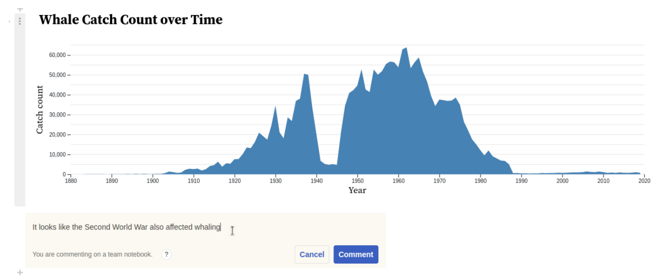

Fig 2(c): Adding comments in the dashboard.

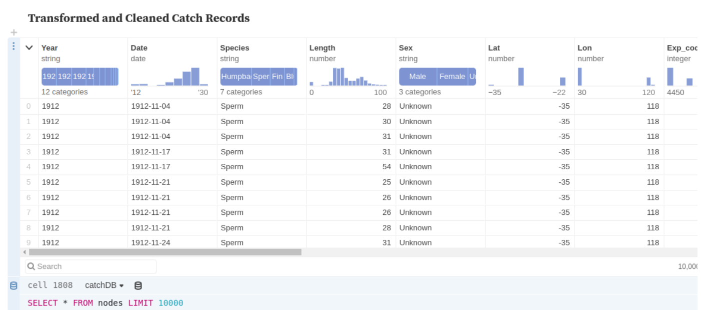

Fig 2(d): Running SQL queries on the cleaned data in the data wrangling notebook.

Fig 2: Enabling open-access of raw and filtered data and easy dissemination of analysis, supporting design goal G1.

We use DuckDB-Wasm (Kohn et al., 2022) for data pre-processing and as the data store for the dashboard. DuckDB-Wasm is a relational database engine like PostgreSQL, but designed specifically for fast interactive analytics in the web browser.

G2: Enable Spatial & Graph Analysis.

To enable spatial analysis, we visualize catch locations on a geographic map with scatterplots and heatmap (Fig 3, and 4). The map visualization supports pan and zoom actions to browse the globe. In order to support different kinds of comparative analyses at various levels of granularity, we support 4 different levels of aggregation or *levels of detail*, and provide a UI widget to control the level of aggregation of the data. The following *levels of detail* are provided in the dashboard:

- *Ocean* - aggregating all catches in an oceanic region,
- *Area* - aggregating all catches in oceanic sub-areas defined in the database,
- *Grid_5x5* - aggregating all catches in 5x5∘ lat-lon grids, Fig 3(a) and
- *Grid_1x1* - for aggregating all catches in 1x1∘ lat-lon grids, Fig 3(b).

We also provide options to choose which data attribute to represent in the scatterplot using the circle color encoding—majority species, Fig 3(a), majority nationality, Fig 4(a), majority sex, Fig 4(b), majority pelagic vs land, Fig 4(c), avg length—and initialize the visualization with reasonable defaults. We encode the catches with the circle size (area), and are working on enabling the option to encode CPUE with the circle size. Fig 1(A) shows the UI widgets to set the level of detail and the color encodings. We set semantic default color encodings to visualize over 20 different *species* and *nation* attribute values (Lin et al., 2013) in the dataset. These colors are chosen in accordance with proximity between species based on taxonomy, and in accordance with spatial proximity and parent continents, for the *nation* attribute. We also provide the heatmap encoding to view a continuous representation of the distribution as opposed to the discrete scatterplot, Fig 3(c).

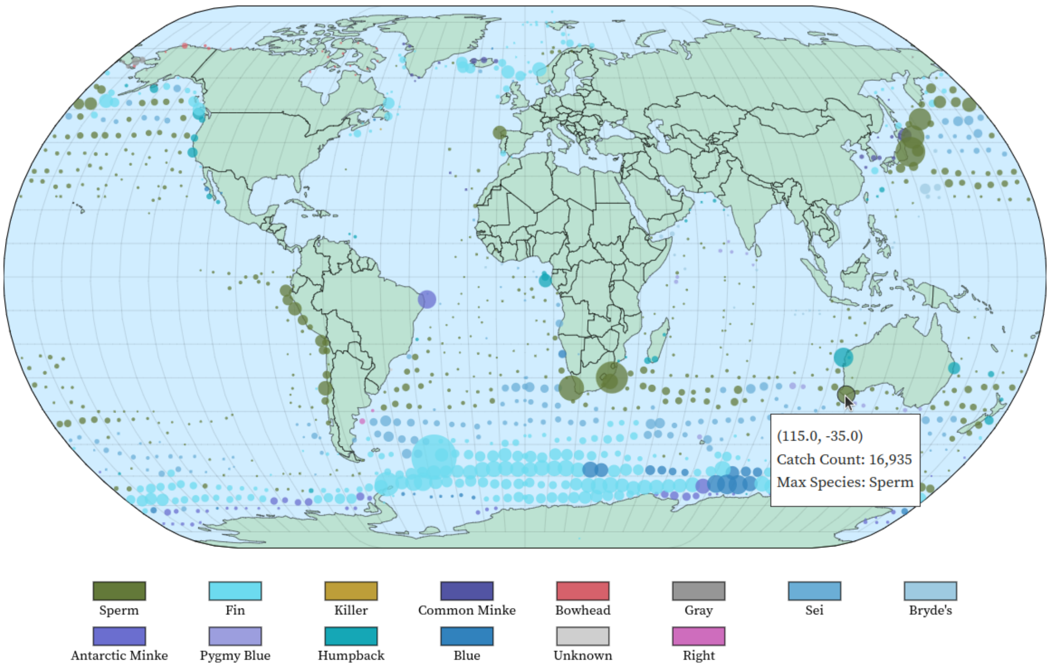

Fig 3(a): Grid_5x5 level of detail, with circle colors encoding the majority species caught in each 5x5◦ lat-lon grid.

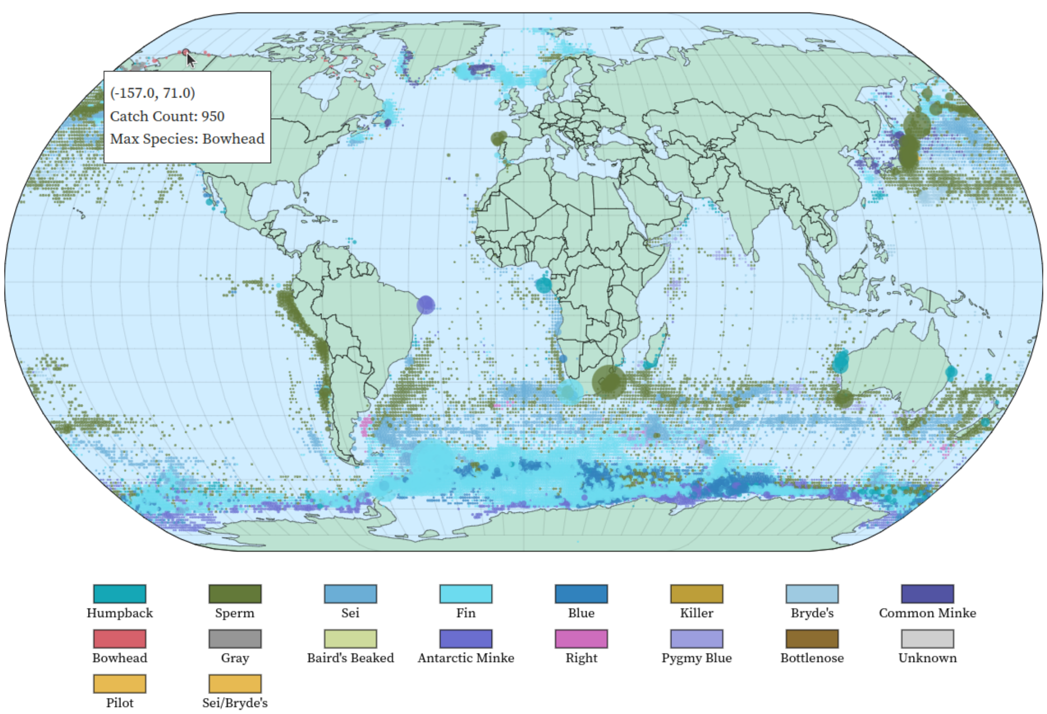

Fig 3(b): Grid_1x1 level of detail, with circle colors encoding the majority species caught in each 1x1◦ lat-lon grid.

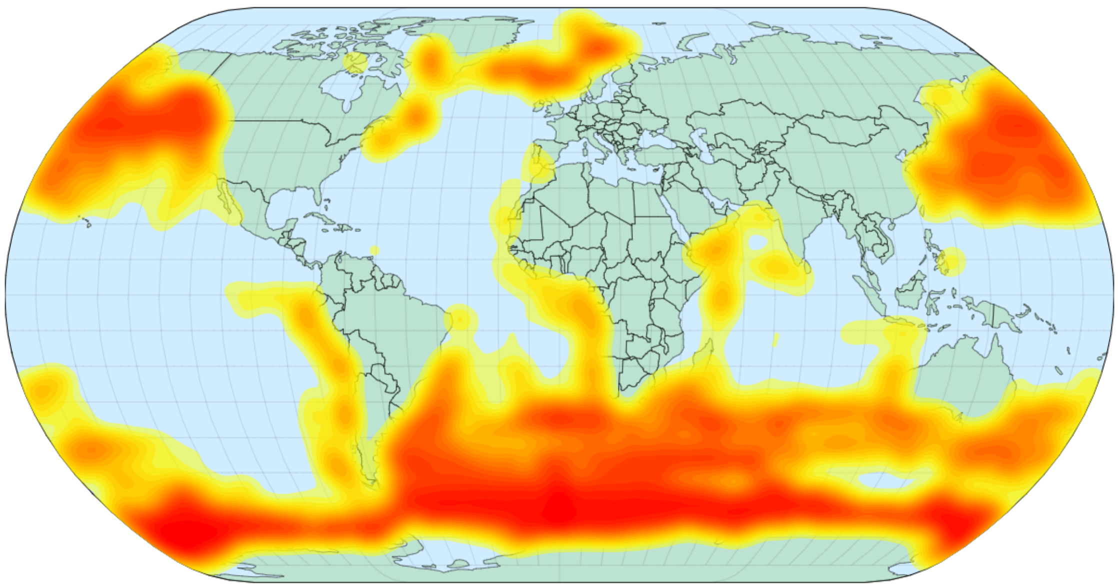

Fig 3 (c): Heatmap of whaling events.

Fig 3: Whaling data shown in different levels of detail i.e. different levels of aggregation. For the scatterplots, mark size encodes the catches, and mark color encodes the whale species.

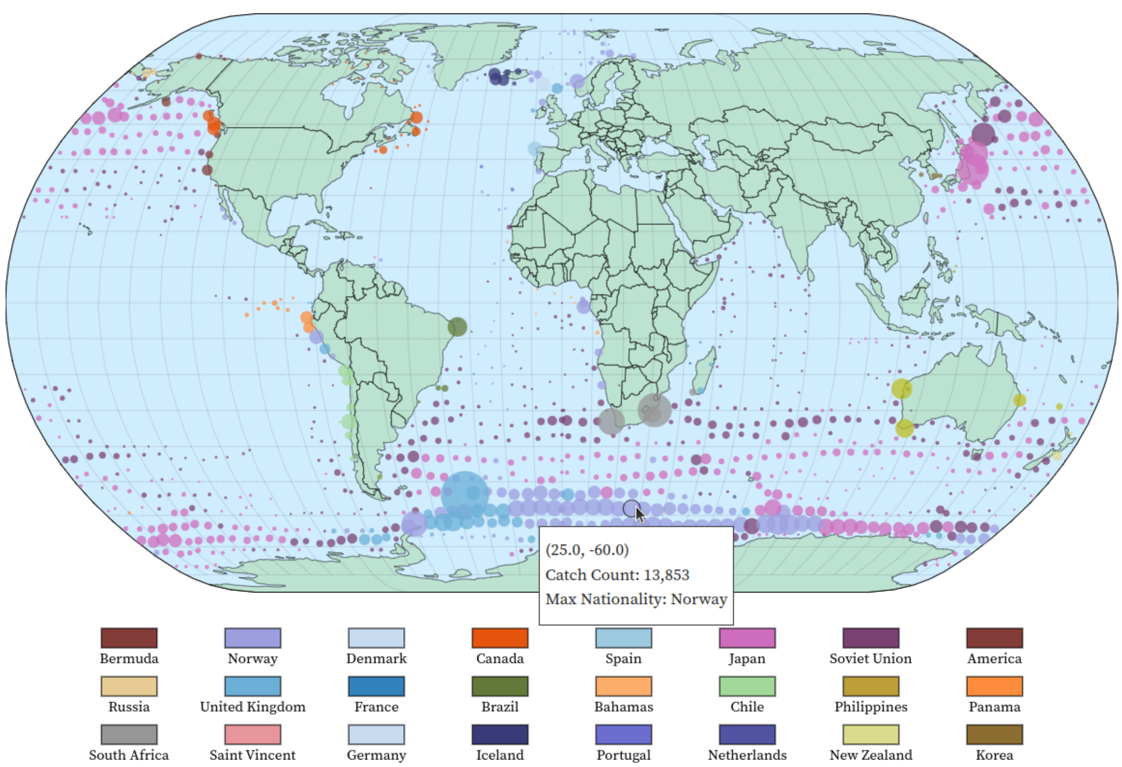

Fig 4(a): Colors encode the nation catching the most whales in each 5x5° lat-lon grid.

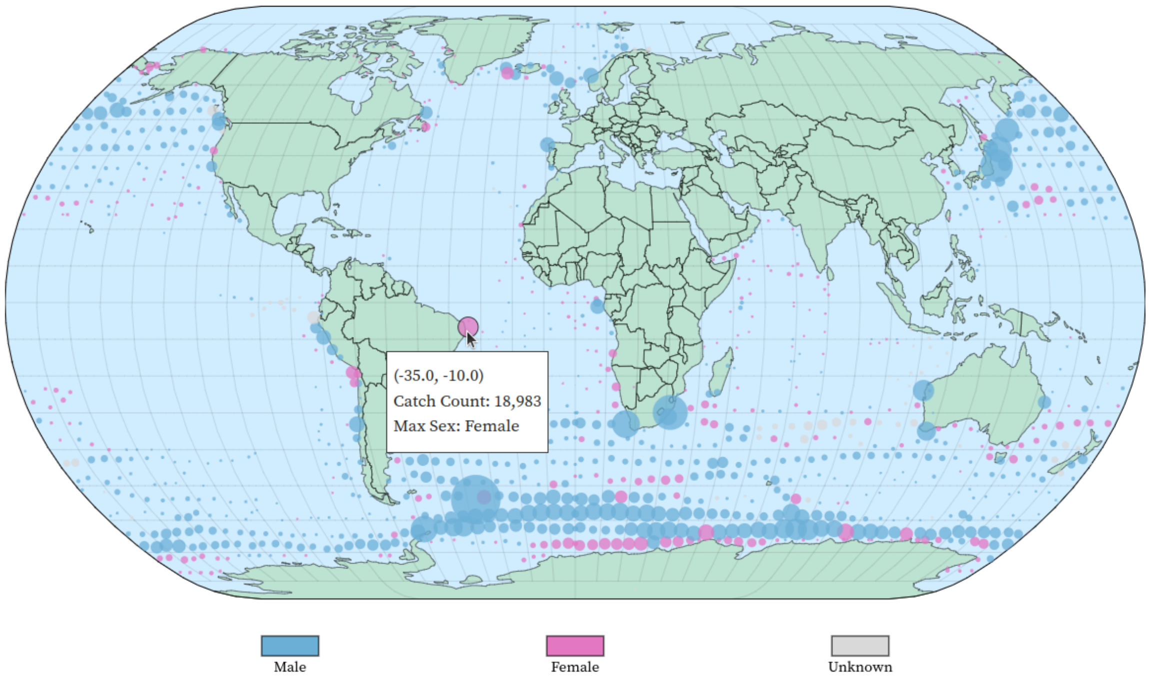

Fig 4 (b): Color encoding the sex of the majority of the catches in each 5x5∘ lat-lon grid.

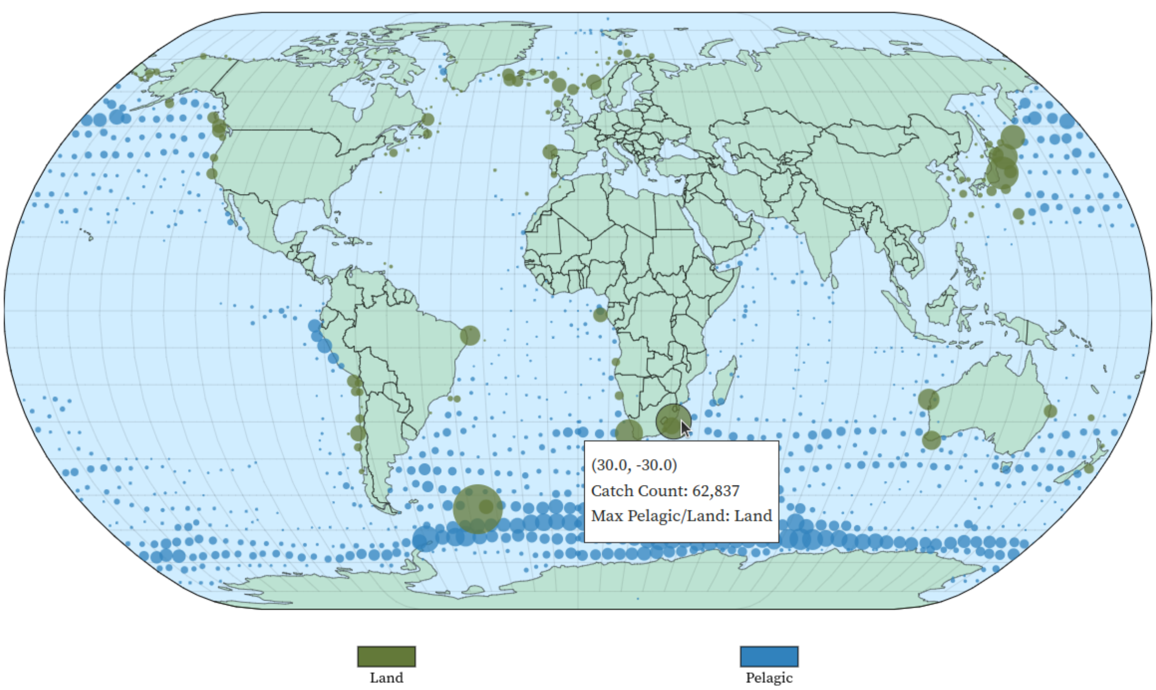

Fig 4 (c): Color encoding whether a majority of the expeditions were based from a land station or operating from a floating expedition ("pelagic") in each 5x5∘ lat-lon grid.

Fig 4: Examples of the different size encodings at the Grid_5x5 level of detail; the size of the circle encodes the number of whales caught at that location.

While the data already has the required attributes for spatio-temporal visualization and analysis, we perform additional precomputation to model it as a graph. We reconstruct all the expedition routes from the timestamps of successive catch locations for each expedition. This process was done only for pelagic whaling expeditions; the land whaling expeditions record all their catches as coming from a single coastal location. For each catch record and its corresponding expedition, we find the corresponding expedition start and end date, since some expeditions had multiple start and end dates. We scanned the catch records for a particular instance of an expedition within its start and end dates in a chronological sequence; a route (graph edge) was created whenever the catch location over successive records changed, indicating movement of the expedition. These expedition routes are visualized on a geographical map similar to the whaling events scatterplots, Fig 3 and 4. We have created separate Observable notebooks for graph modeling computation and expedition route visualization[11,12], and it is currently work in progress.

G3: Balance Inclusivity & Usability

We implemented charts to view the details for the most relevant attributes in the dataset, viz. catch trends over the years (*timeline chart*), detailed breakdown of catches with respect to whale species, sex, length, expedition nationalities, and expedition types (pelagic vs. land). To provide the capability to filter the scatterplot visualization on the geographic map, or the other detailed breakdown information of catches, we make all of the visualizations interactive. Data filters on attributes represented in a chart can be set by interacting with the chart itself, and these filters affect all the visualizations apart from the one used to set the filter, Fig 5. This approach, called *cross-filtering*, follows the direct manipulation paradigm (Hutchins et al., 1985) which helps reduce visual clutter in a visualization dashboard by reducing the amount of UI widgets, and also makes the interactions more intuitive. On similar lines, the timeline chart also serves as a time slider widget to set a desired time window filter (C2), Fig 1(B), 5(a). Filters, one or many, can be set/reset on the bar charts by clicking on individual bars, Fig 5(b), while the timeline chart and length histogram can be filtered by a click and drag action (*brushing*), Fig 5(a). Unfiltered data is colored gray, whereas the filtered data, which is a subset of the unfiltered data, is rendered on top of the unfiltered data, and is colored blue. The numbers on filtered bar charts represent the filtered data. A filter on the geographic map visualization (latitude-longitude filters) can be set with a simple button click, after zooming in to the region of choice, Fig 5(c). The viewport state determines the geographic filter bounds. Fig 1 also shows another example of cross-filtering in WhaleVis.

We thus provide a default configuration of as many important views of the data as possible, while also making it easier to interact with them. In the current state, our dashboard has a restricted linear layout of visualizations, which is common in collaborative notebook environments, Fig 1. However we recognise that different analyses may require a different set of visualizations, or a different layout which improves user experience. Identifying this requirement, we also plan to make the dashboard configurable as a future work.

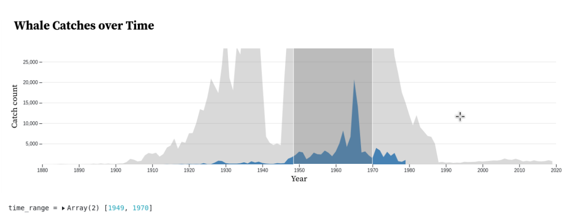

Fig 5(a): Filter set on time, and the effect of the species (5(b)) and geo filter (5(c)) visible in the timeline chart.

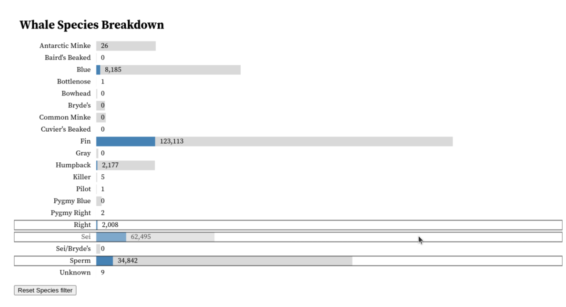

Fig 5(b): Filter set on species, and the effect of time (5(a)) and geo filter (5(c)) visible in the species breakdown chart.

11 https://observablehq.com/@whales/iwc-dataset-expedition-routes

12 https://observablehq.com/@whales/whale-vis-dashboard-expedition-routes

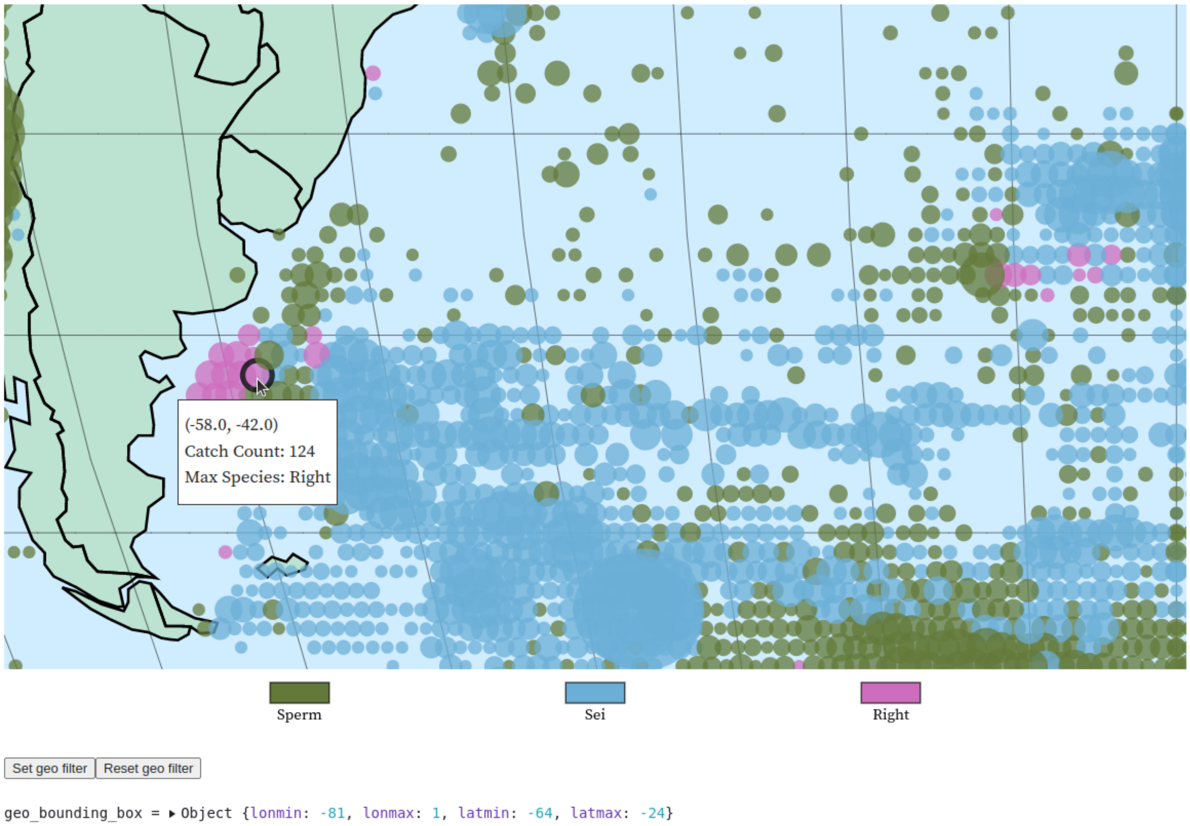

Fig 5(c): Filter set on latitude and longitude range determined by the viewport state, and the effect of time (5(a)) and species (5(b)) filter visible in the geographic map visualization.

Fig 5: Demonstrating the working of cross-filtering between visualizations.

We intend this notebook to be used both for data exploration and analysis or explanation (Ynnerman et al., 2018). The Observable notebook environment helps us towards this goal by enabling transparency and collaborative analysis. Whale scientists can use the notebook to gather their insights and findings, and then also annotate the notebook with comments to share their findings with others. As a future work, we also plan to improve insight management in-house in the tool to make information communication more efficient; more on this is discussed in Section 4.2.

3. Use cases

We illustrate how the database can be used to answer particular questions related to our ongoing research.

3.1 Vagrant Blue Whales

Blue whales include many separate populations, each with well known regions of high abundance. But sometimes individuals are identified far from known centers of distribution, and it is not clear whether these individuals are rare vagrants or represent an understudied population. We used WhaleVis to identify isolated blue whale catches far from catch centers, possibly caught during operations targeting other species. Extracting the relevant attributes like date, location and physical attributes like sex and length of such whale catches could help us understand if blue whales have been straying away from their usual nesting places, and any patterns therein.

Since we are interested in only blue whales, we set the corresponding species filter by clicking on the blue whales bar in the Whale Species Breakdown bar chart. In response, the geographic visualization automatically updates to show only blue whale catches, allowing us to see where they were mostly hunted. Since we are looking for vagrant blue whales, we need to focus on regions with low catches. We can do this using the *Grid_1x1* level of detail in the map.

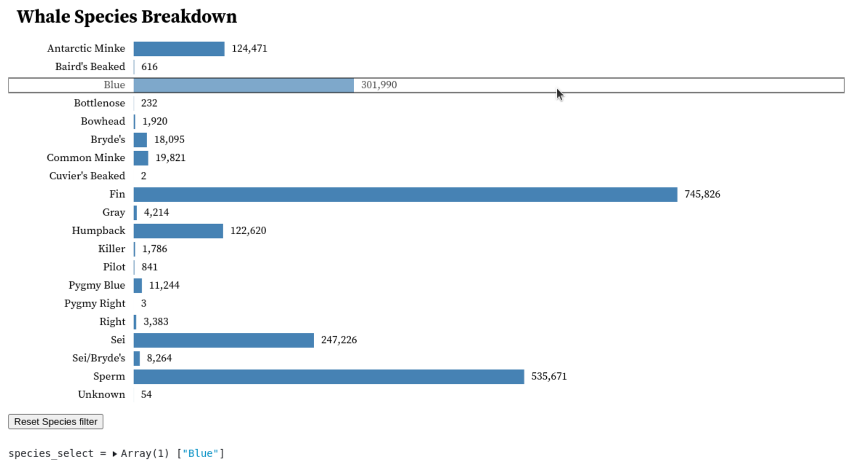

Fig 6(a): Filter set on Blue whales.

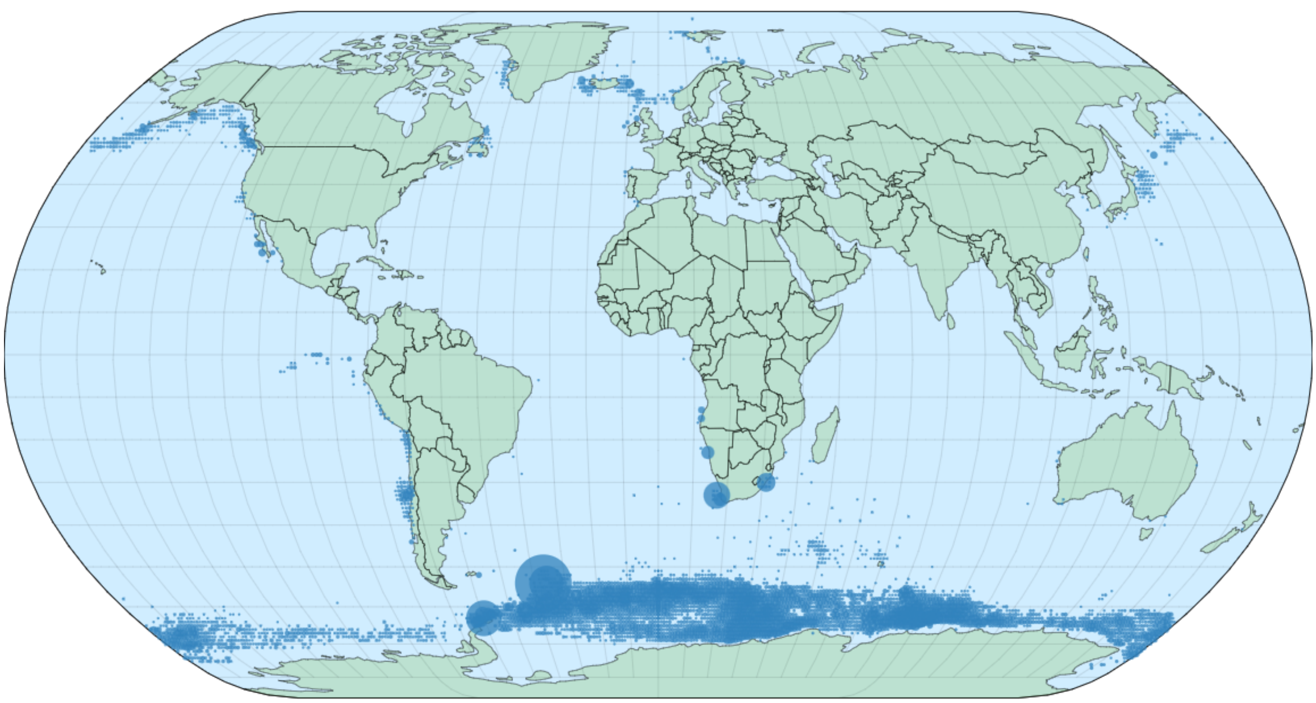

Fig 6(b): Geographic map showing Blue whale catches at the Grid_1x1 level of details.

Fig 6: Investigating vagrant blue whale catches using WhaleVis.

From Fig 6(b) we observe that most of the blue whales were caught in the Southern Hemisphere, with major catches also recorded in the Norwegian Sea, Labrador Sea, Bering Sea, the Gulf of Alaska, Gulf of California, along the eastern coast of South America, and in southern Africa. After zooming into the map, Fig 7, we can see some isolated catches in the southern Atlantic Ocean, along the coasts of Brazil. We can change the color encoding of the geographic map visualization to inspect other attributes of those catches. A geo filter can be set and the resulting vagrant catch data can be downloaded as a .csv file for further analysis.

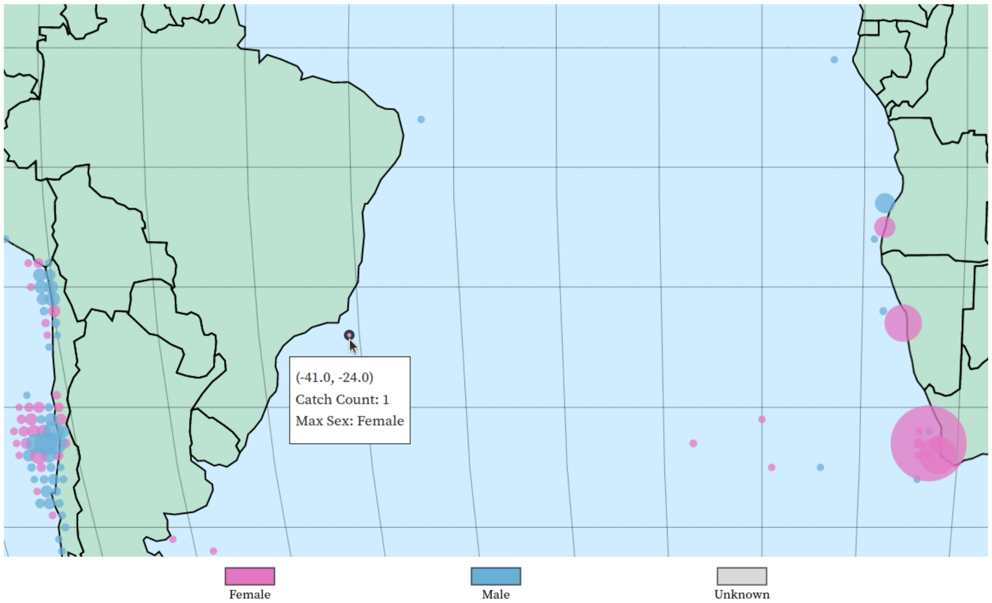

Fig 7(a): Blue whale catches at the Grid_1x1 level of detail encoding majority sex with color, in the South Atlantic ocean.

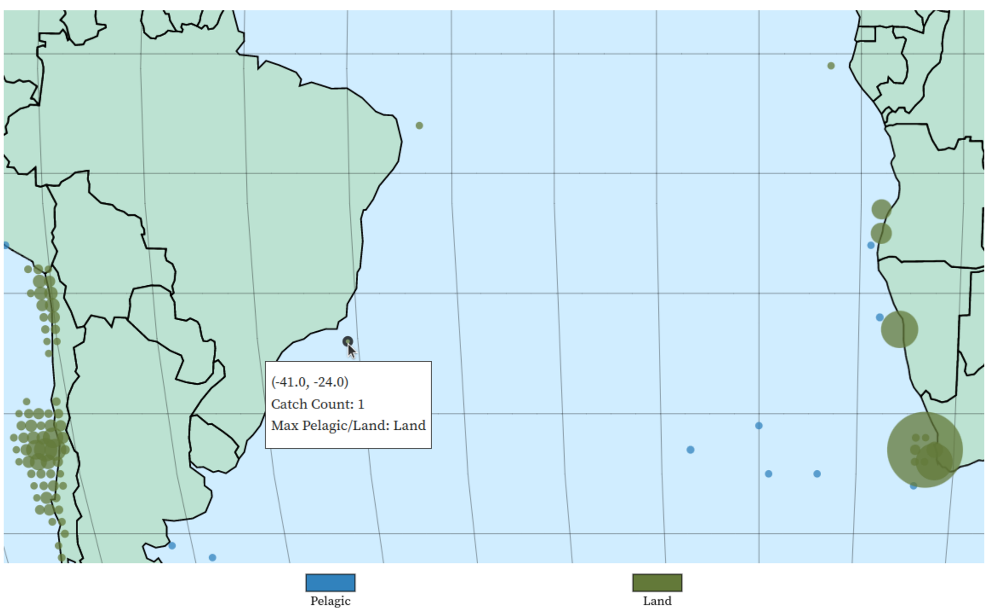

Fig 7(b): Blue whale catches at the Grid_1x1 level of detail encoding majority pelagic/land expedition with color, in the South Atlantic ocean.

Fig 7: Using the geographic map visualization to understand characteristics of vagrant blue whale catches.

3.2 Progression of Hunted Whale Species

Over time, different whale species were sequentially depleted depending on their profitability: abundance, ease of hunting, and the amount of oil that could be obtained from each species. The IWC database can be used to understand how humans switched from hunting one species to the other from 1880 to the present day. To illustrate this progression, we can set filters on individual species and inspect the time window over which that species was hunted, and also the length histogram to understand the possible commercial gains of hunting that particular species.

From Fig 8, we can see that Blue and Fin whales, in that order, were two of the most highly sought after species at the start of the 20th century. Blue whale hunting was reduced by the year 1960. Around 1965, as Fin whale hunting started declining, Sei whale hunting peaked. This was followed by a peak in Minke (both Common and Antarctic) whale hunting. The length histogram shows one possible reason behind this kind of a progression - Blue whales at around 80 ft were the longest and thus had the potential of yielding high commercial gains, compared to Fin whales at around 70 ft, followed by Sei whales at around 50 ft and Minke whales at around 30 ft.

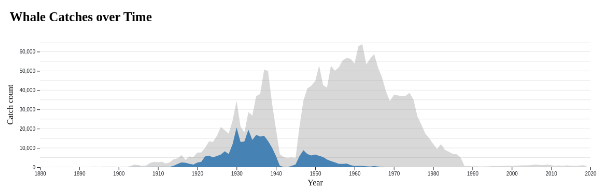

Fig 8(a): Blue whale catches over time.

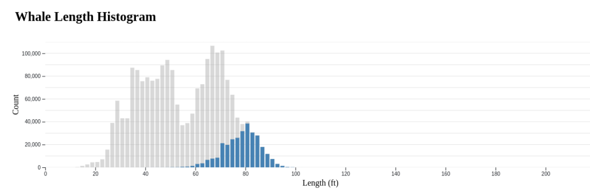

Fig 8(b): Blue whale length histogram.

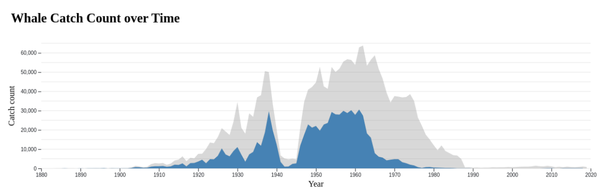

Fig 8(c): Fin whale catches over time.

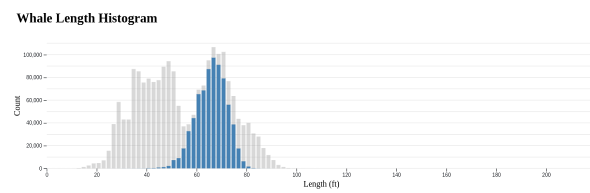

Fig 8(d): Fin whale catches over time.

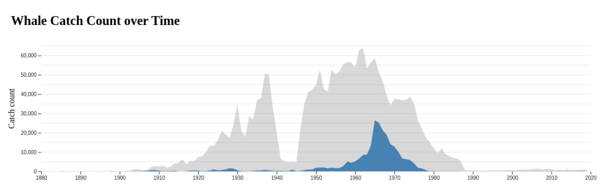

Fig 8(e): Sei whale catches over time.

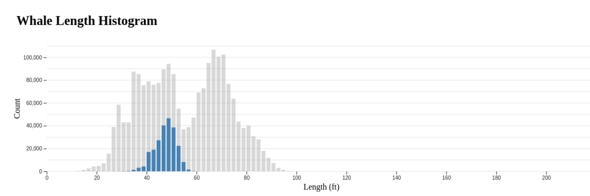

Fig 8(f): Sei whale length histogram.

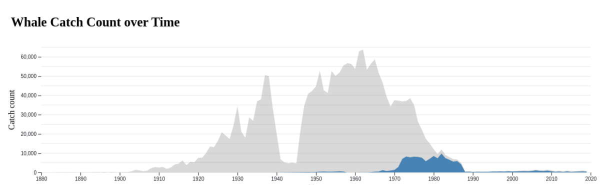

Fig 8(g): Minke whale catches over time.

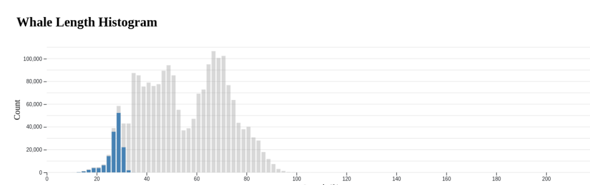

Fig 8(h): Minke whale length histogram.

Fig 8: Catches over time and length histograms for Blue, Fin, Sei and Minke (Antarctic and Common) whales.

Another pattern can be observed in the whaling expeditions before and after 1925, Fig 9. Most of the expeditions before 1925 were based out of land stations. While land based whaling continued thereafter, pelagic whaling exploded. This allowed for Blue and Fin whale hunting to sustain, while also resulting in many different species like Sperm, Bryde's and Minke being caught, after the Blue and Fin whales started declining in numbers.

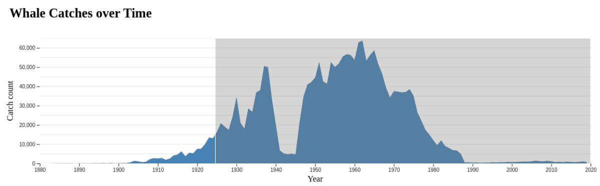

Fig 9(a): Time filter set to cover whaling events before 1925.

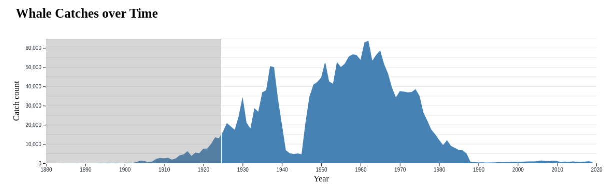

Fig 9(b): Time filter set to cover whaling events after 1925.

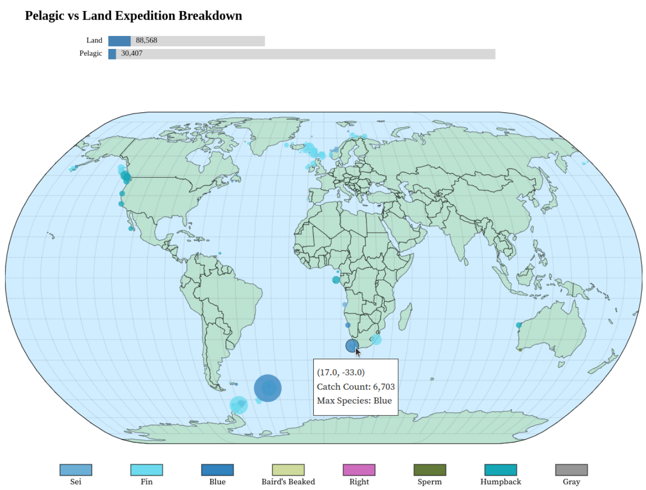

Fig 9(c): Proportion of land vs pelagic whaling expeditions before 1925.

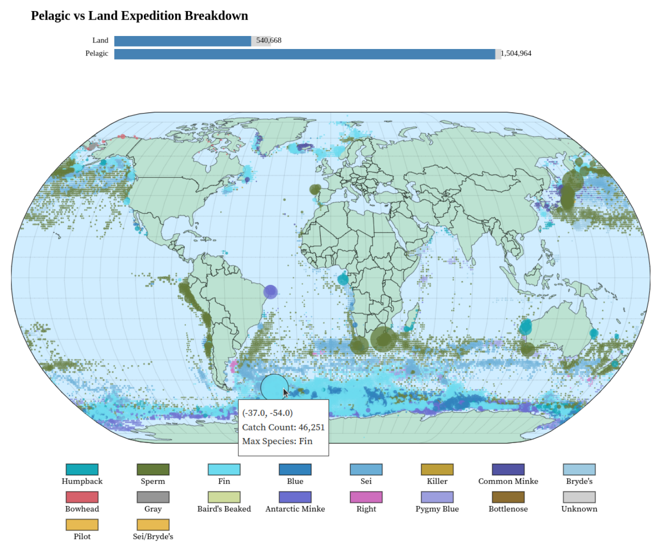

Fig 9(d): Proportion of land vs pelagic whaling expeditions after 1925.

Fig 9: Land vs Pelagic whaling expeditions before and after 1925.

After the IWC passed a moratorium on commercial whaling starting in 1986, whaling activities reduced drastically. Since 2000, only a handful of species have been hunted in a regulated manner; these include the Minke, Gray, Sei and Bowhead whales, with opportunistic catches of other depleted species, Fig 10. Most of these whaling activities are conducted for subsistence for the residents of the arctic regions, or for research purposes.

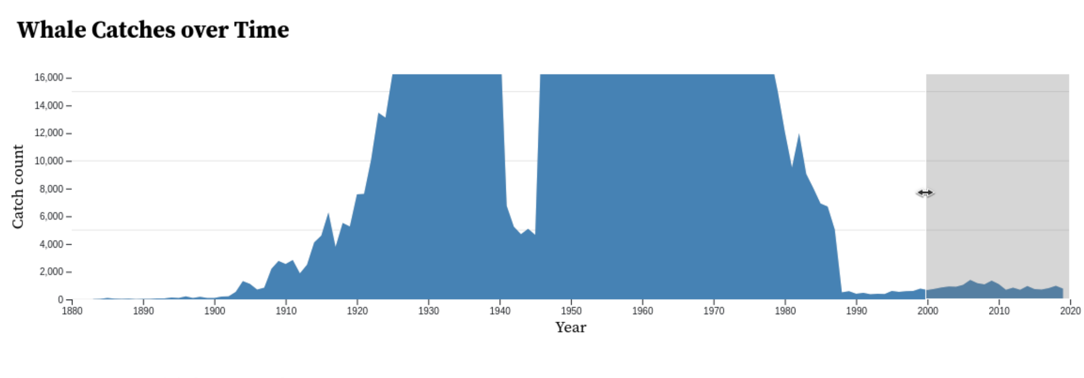

Fig 10(a): Time filter set to cover whaling events after 2000.

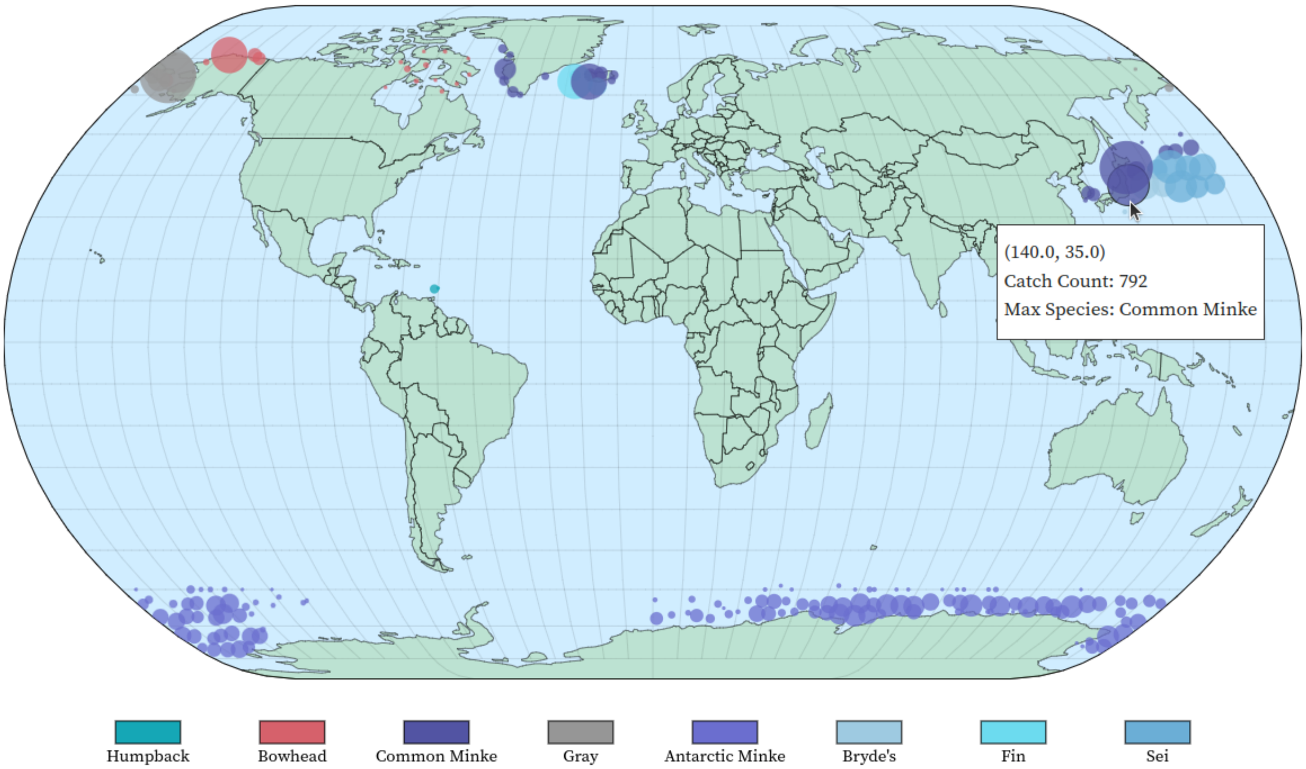

Fig 10(b): Whaling events after 2000.

Fig 10: Whales hunted since the start of the 21st century.

3.3 Verifying Catch Data

The IWC catch database will continue being updated over the coming years, and every new update may contain certain erroneous records. WhaleVis can be used to perform sanity checks on the data, and report any such errors.

As an example of this, we describe how one of our coauthors (T.A.B.) was able to find two such possible errors while casually exploring the data. The length histogram domain (X-axis) is determined by the range of length values in the database. Since even the largest whales (Blue) have lengths up to 110 ft, a maximum value of 220 ft hinted towards possibly erroneous catch records. Setting a length filter for catches reporting lengths above 104 ft, Fig 11(a), revealed 11 catch records.

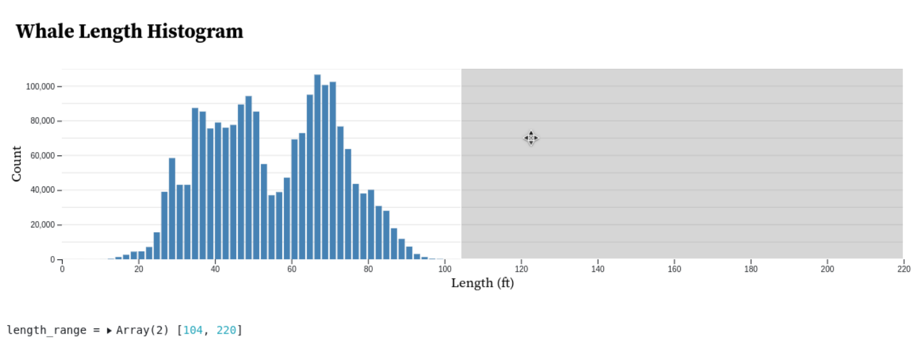

Fig 11(a): Filter set on lengths above 104 ft.

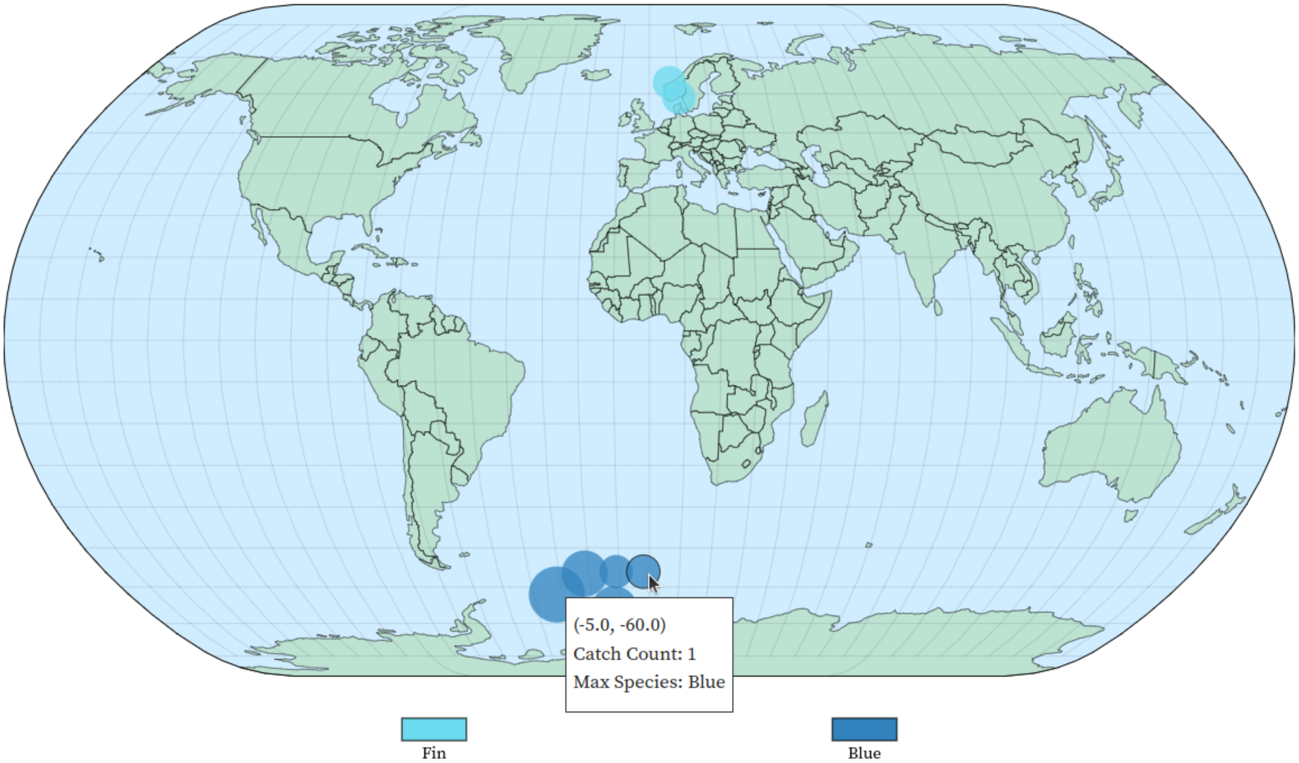

Fig 11(b): Geographic map visualization showing the catches recorded to have lengths above 104 ft - Blue whales and Fin whales.

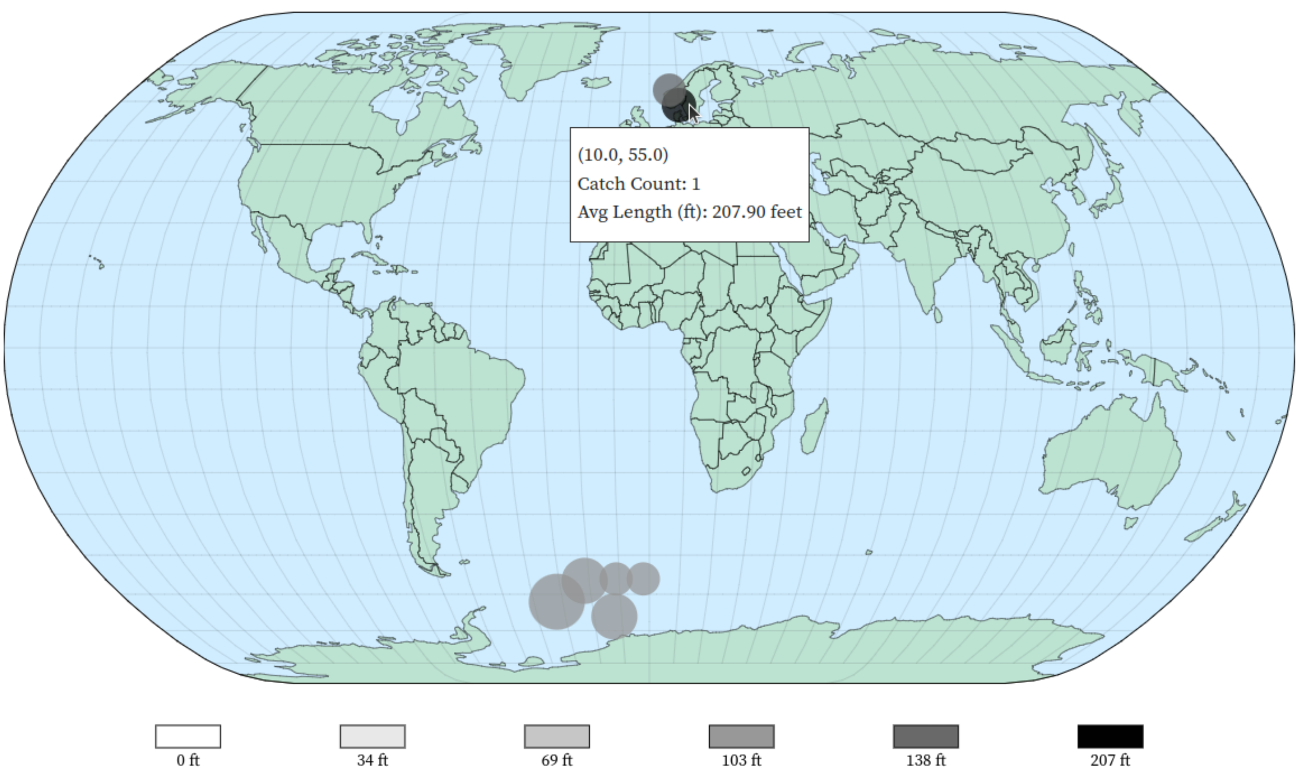

Fig 11(c): Geographic map visualization showing Fin whale catches having lengths higher than 120 ft.

Fig 11: Investigating possible erroneous catch records.

Investigating the resulting geographic map visualization with circle color encoding species and average length, Fig 11(b) and 11(c) respectively, revealed that two Fin whales were reported to have lengths of 138.6 ft and 207.9 ft, hinting towards possible errors in recording those catches. These records can be downloaded in a .csv format to investigate other attributes of these records.

4. Future Work

Although in its current state, Whale Vis can be used to answer many exploratory/explanatory questions, we envision more improvements and new features in our tool.

4.1 Expedition Route Visualization & Graph Modeling

In the current state, the data processing for expedition route computation is complete. We are yet to complete the implementation of route visualization. We anticipate visual scalability issues like visual clutter due to a high number of nodes and edges drawn in a limited screen space ('hairball problem' in graph visualizations) (Pienta et al., 2015), and compute scalability issues like running graph algorithms on a large number of nodes and edges, which may take prohibitively high time for interactive exploration and analysis of the data. We would be working towards addressing these problems. With this addition, Whale Vis would enable graph analysis capabilities on the whaling expeditions data e.g. path finding analysis, centrality measure computation, etc., thus facilitating a novel approach to analyze the IWC catch dataset.

4.2 Insight Management

Motivated by our design goal (G1) and the exploration/explanation paradigm (Ynnerman et al., 2018), we would also like to build better infrastructure around recording and sharing insights generated by users during their analysis workflow in WhaleVis. For example, as described in Section 3.3, one of the coauthors (T.A.B.) was able to download the possibly erroneous catch records to report it to the IWC. This process of correcting possible errors in the source catch data can be made more efficient if such findings can be managed within the tool.

The current Observable notebook environment facilitates adding public comments, and downloading screenshots of visualizations, or filtered data. However to make it truly accessible, even for the non-expert audience, we need better ways of communicating expert insights in an easy to follow manner, e.g. walkthroughs of the analysis workflow for better clarity of insight generation.

4.3 Persistent data storage

The current implementation of Whale Vis using the Observable notebook environment does not provide us flexible data storage options. We envision a dedicated database server for the IWC catch database to which our dashboard can connect. This would reduce the current requirement of running compute-heavy data pre-processing on users' laptops, by pushing it to the high compute capability server machines. It would also enable the implementation of optimizations for interactive data analysis like data caching and prefetching (Battle et al. 2016), thus improving user experience by reducing the response time for user interactions.

5. Conclusion

We created an interactive data visualization and analysis dashboard for the IWC catch database, to understand historical whaling events and inform our future efforts for conservation and management of whales. Our guiding principles were based on community values around open access data, and best practices in visualization and interaction design, and visualization application theory. We shared our goals for designing WhaleVis, prototyped WhaleVis using the Observable notebook environment, and demonstrated its usage through three real life examples. We plan to further improve our tool to facilitate a novel graph modeling perspective on the whaling expedition routes data, which would enable computing accurate estimates of catch-per-unit-effort (CPUE) in addition to running graph analysis algorithms on the data, and better insight management for sharing of knowledge among the users, among many other changes to better support usability and scalability. Our hope is that our interactive visual data analysis pipeline can streamline the process of gathering and understanding the whale catch and CPUE information, in addition to the current efforts of monitoring whale populations. WhaleVis can hopefully be used to better organize management actions to conserve whales, and help spread awareness about whales.

Acknowledgements

This project was funded in part by an award from the UW Computing for the Environment Initiative, which is supported in part by the Allen School, the College of the Environment, the College of Engineering, the eScience Institute, the Department of Civil & Environmental Engineering, the Paul G. Allen School of Computer Science & Engineering Endowed Fund for Excellence, and a gift from Google.

References

Scarpaci, C., Parsons, E.C.M. and Lück, M., 2008. Recent advances in whale-watching research: 2007–2008. Tourism in Marine Environments, 5(4), pp.319-336.

Kowarski, K.A. and Moors‐Murphy, H., 2021. A review of big data analysis methods for baleen whale passive acoustic monitoring. Marine Mammal Science, 37(2), pp.652-673.

Hazen, E.L., Palacios, D.M., Forney, K.A., Howell, E.A., Becker, E., Hoover, A.L., Irvine, L., DeAngelis, M., Bograd, S.J., Mate, B.R., Bailey, H., 2017. WhaleWatch: a dynamic management tool for predicting blue whale density in the California Current. Journal of Applied Ecology 54, 1415–1428. https://doi.org/10.1111/1365-2664.12820

Bamford, C.C.G., Kelly, N., Dalla Rosa, L., Cade, D.E., Fretwell, P.T., Trathan, P.N., Cubaynes, H.C., Mesquita, A.F.C., Gerrish, L., Friedlaender, A.S., Jackson, J.A., 2020. A comparison of baleen whale density estimates derived from overlapping satellite imagery and a shipborne survey. Sci Rep 10, 12985. https://doi.org/10.1038/s41598-020-69887-y

Allison, C. 2020. IWC summary large whale catch database Version 7.1 released 23 December 2020. Available from the International Whaling Commission, 135 Station Road, Impington, Cambridge, CB24 9NP UK.

Allison, C. 2020. IWC individual catch database Version 7.1 released 23 December 2020. Available from the International Whaling Commission, 135 Station Road, Impington, Cambridge, CB24 9NP UK.

Staniland I J., Tandy M, Webster I, Ki Sow M, and Smith S. 2021. Databases and data holdings in the IWC: A review of the current status, issues and future plans. IWC paper SC/68C/GDR/01

Isenberg, P., Elmqvist, N., Scholtz, J., Cernea, D., Ma, K.-L., & Hagen, H. 2011. Collaborative visualization: Definition, challenges, and research agenda. Information Visualization, 10(4), 310–326. https://doi.org/10.1177/1473871611412817

Munzner, T., 2009. A Nested Model for Visualization Design and Validation. IEEE Transaction on Visualization and Computer Graphics Graphics 15, 921–928. https://doi.org/10.1109/TVCG.2009.111

Sedlmair, M., Meyer, M., Munzner, T., 2012. Design Study Methodology: Reflections from the Trenches and the Stacks. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis) 18, 2431–2440.

Meyer, M., Dykes, J., 2019. Criteria for Rigor in Visualization Design Study. IEEE Transaction on Visualization and Computer Graphics Graphics 1–1. https://doi.org/10.1109/TVCG.2019.2934539

Ye, Y., Sauer, F., Ma, K.-L., Aditya, K., Chen, J., 2020. A User-Centered Design Study in Scientific Visualization Targeting Domain Experts. IEEE Transaction on Visualization and Computer Graphics Graphics 26, 2192–2203. https://doi.org/10.1109/TVCG.2020.2970525

Shneiderman, B., 1996. The eyes have it: a task by data type taxonomy for information visualizations, in: Proceedings 1996 IEEE Symposium on Visual Languages. Presented at the 1996 IEEE Symposium on Visual Languages, IEEE Comput. Soc. Press, Boulder, CO, USA, pp. 336–343. https://doi.org/10.1109/VL.1996.545307

Hutchins, E.L., Hollan, J.D., Norman, D.A., 1985. Direct Manipulation Interfaces. Human–Computer Interaction 1, 311–338. https://doi.org/10.1207/s15327051hci0104_2

Sarikaya, A., Correll, M., Bartram, L., Tory, M., Fisher, D., 2019. What Do We Talk About When We Talk About Dashboards? IEEE Trans. Visual. Computer. Graphics 25, 682–692. https://doi.org/10.1109/TVCG.2018.2864903

Bach, B., Freeman, E., Abdul-Rahman, A., Turkay, C., Khan, S., Fan, Y., Chen, M., 2022. Dashboard Design Patterns. IEEE Transaction on Visualization and Computer Graphics Graphics 1–11. https://doi.org/10.1109/TVCG.2022.3209448

Kohn, A., Moritz, D., Raasveldt, M., Mühleisen, H., Neumann, T., 2022. DuckDB-wasm: fast analytical processing for the web. Proc. VLDB Endow. 15, 3574–3577. https://doi.org/10.14778/3554821.3554847

Sharon Lin, Julie Fortuna, Chinmay Kulkarni, Maureen Stone, and Jeffrey Heer. 2013. Selecting semantically-resonant colors for data visualization. In Proceedings of the 15th Eurographics Conference on Visualization (EuroVis '13). The Eurographs Association & John Wiley & Sons, Ltd., Chichester, GBR, 401–410.

Ynnerman, A., Lowgren, J., Tibell, L., 2018. Exploranation: A New Science Communication Paradigm. IEEE Comput. Grap. Appl. 38, 13–20. https://doi.org/10.1109/MCG.2018.032421649

Pienta, R., Abello, J., Kahng, M., Chau, D.H., 2015. Scalable graph exploration and visualization: Sensemaking challenges and opportunities, in: 2015 International Conference on Big Data and Smart Computing (BIGCOMP). Presented at the 2015 International Conference on Big Data and Smart Computing (BigComp), IEEE, Jeju, South Korea, pp. 271–278. https://doi.org/10.1109/35021BIGCOMP.2015.7072812

Leilani Battle, Remco Chang, and Michael Stonebraker. 2016. Dynamic Prefetching of Data Tiles for Interactive Visualization. In Proceedings of the 2016 International Conference on Management of Data (SIGMOD '16). Association for Computing Machinery, New York, NY, USA, 1363–1375. https://doi.org/10.1145/2882903.2882919