Reconstructing Whaling Voyages to Build Maps of Whale Density by Species

AMEYA PATIL

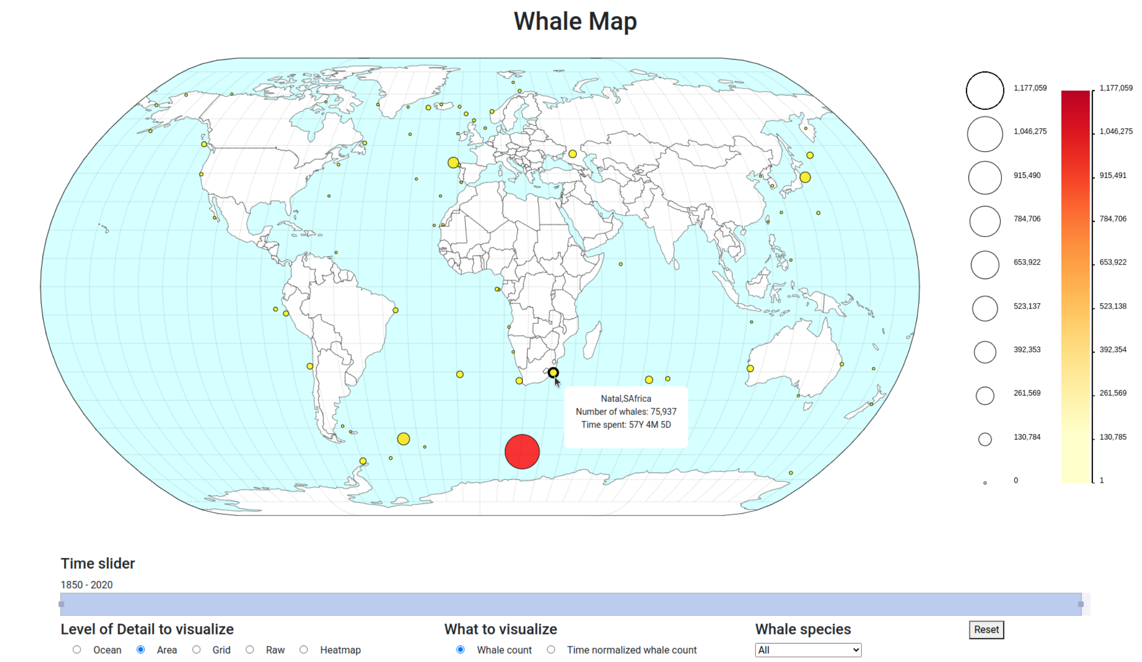


Fig. 1. Our proposed Whale Map interface

We create density maps of whale populations of different species based on a historical commercial whale hunting dataset, with the aim to create a pipelined system to facilitate whale conservation efforts. Whales are an important part of the oceanic ecosystem. However commercial whaling till the early 20th century, and fallouts of human activities both within and outside the oceans, have severely threatened whale populations. It thus becomes important to take efforts towards conservation of whales. However, we first need to understand where the whale populations were before, and where they are today, i.e., we need to understand how their numbers and dwellings have changed over time. To this end, we create a system to build visualizations of whale population densities, along with the routes taken by whaling voyages. We model the whaling data as network data, where the nodes represent locations of whale sightings and the edges represent whaling voyages. We are building a large scale data analysis system for visualizing and analyzing general purpose geographic network data which can be used to create the visualizations for whale population densities and voyage routes. The population density map coupled with the oceanic transport routes data can help us plan management activities for the conservation of whales.

Author's address: Ameya Patil, ameyap2@cs.washington.edu.

Manuscript submitted to ACM

CCS Concepts: • **Applied computing → Earth and atmospheric sciences**; • **Human-centered computing → Visualization systems and tools**.

Additional Key Words and Phrases: whaling, conservation, data visualization, interactive data analytics

ACM Reference Format:
Ameya Patil. 2022. Reconstructing Whaling Voyages to Build Maps of Whale Density by Species. 1, 1 (December 2022), 9 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Whales are a crucial cog in the oceanic ecosystem — because of their sheer size, whales are one of the apex predators in the oceanic food chain and thus regulate the populations of many other marine life; they share symbiotic relationship with many marine species e.g., barnacles living on the skin of whales [16], or remora fish swimming close to whales or sticking to whales [10], and most importantly, they contribution towards transferring nutrients across the depths of the ocean, and also across different regions at the same depth [14]. Thus whales, like any other species in an ecosystem, contribute to the health of the oceanic ecosystem.

In spite of the importance of whales, humans have threatened their existence for a long time now. Whales have been hunted as a source of food, and since the 17th century, for commercial purposes to extract blubber and oil for industrial use. The industrial revolution gave further impetus to the race for whale hunting. Owing to this over-exploitation, whale populations were adversely affected by the end of the 20th century; e.g., the right whale population in the south-western Atlantic ocean dropped from around 58,000 before commercial hunting began in the early 17th century, to 2000 in the 1830s - a 29 times decrease over 2 centuries. Today the number is estimated to be around 4000, a mere 2x increase over 90 years [15], blue whales in 2014, were reduced to 1% of their historical numbers in the southern hemisphere [9]. The International Whaling Comission (IWC) [1] which was formed in 1946 issued a moratorium to pause commercial whaling starting 1986. Even after this moratorium, whales have faced threats from ship strikes due to commercial transport activities carried out in the oceans by humans. It has thus become important to work towards the conservation of whales.

Effectively protecting whales requires understanding their population distribution, movements, and vulnerabilities to human activities. However, whale tracking is a resource intensive and low-yield task, since we simply do not have the resources to scan entire oceans looking for whales. For example, consider Figure 2, which shows a snapshot from the Right Whale Sighting Advisory System maintained by the National Oceanic and Atmospheric Administration (NOAA) [2]. Note how in this interface called WhaleMap, we only have data for the coastal regions showing around 4000 sightings along the east coast of North America. To fill in the gaps and get a baseline for deep sea data, we could use historical whaling data to investigate what has happened to the right whale population densities and those of other endangered whale species in a data driven way.

We use commercial whaling data collected and maintained by IWC (International Whaling Commission) and BIWS (Bureau of International Whaling Statistics) from 1880 to 2020. Since this dataset consists of records from commercial hunting expeditions which were more exhaustive in their search across the oceans, along with the details of the expeditions themselves, it gives us a better picture of the deep sea numbers of whale populations. Instead of only looking at a population density map of different whale species, we are also interested in the time spent by whaling expeditions searching for whales and the routes they traversed during their search, as they give us a better idea of the search

[1] https://iwc.int/en/
[2] https://apps-nefsc.fisheries.noaa.gov/psb/surveys/MapperiframeWithText.html

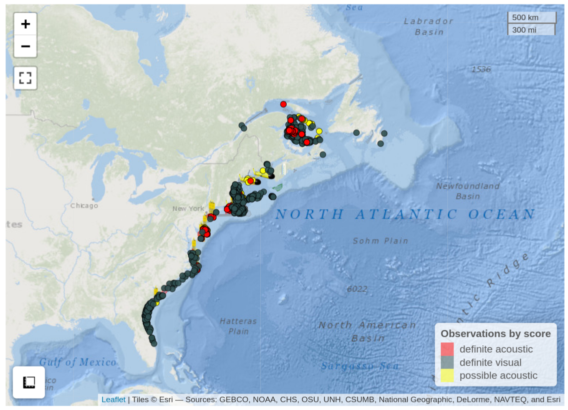


Fig. 2. WhaleMap interface used by the Right Whale Sighting Advisory System maintained by the National Oceanic and Atmospheric Administration. The snapshot shows the details of whale sightings from November 2021 to November 2022 in the north Atlantic Ocean

efforts and the population density normalized by the search efforts. This hints towards a network like representation in which nodes represent the locations of whale sightings, while edges represent the voyage routes between locations. Realizing the suitability of this dataset to be modeled as network data, we visualize the population density map and the voyage routes as network visualization, thus enabling analysis of whaling data in an intuitive way in terms of graph relationships, and temporal and geographic information. To this end, we are building an interactive data visualization and analysis system for large scale geographic graph or network data. Our hope is that the population density heatmap coupled with the oceanic transport routes data can enable advanced analysis and help us to better plan management activities for the conservation of whales.

2 RELATED WORKS

The need to be able to understand the population maps of whales in the past and present times has already been identified as a requirement for working towards their conservation. Accordingly, prior work has been done to be able to track or keep record of whales across the oceans.

To make up for the difficulty of monitoring deep seas, Hazen et al. [11] and Bamford et al. [8] propose the use satellite imagery coupled with synthetically generated data modeled on the limited whale population data. Whale tracking using under-water acoustics [13] has also been used recently to supplement the satellite data. However, most modern datasets are very recent, they do not give us any idea of the baseline numbers, or in other words they do not show the scenario before the modern data collection started. This baseline context is made available to some extent by the historic hunting data collected by Allison [6, 7] and maintained by IWC over 140 years from 1880 to 2020. This dataset is also comprehensive with respect to space simply because more oceanic regions were scanned for better economic gains. We thus use this dataset to build our interface.

In terms of building data visualization and analysis interfaces for whale tracking data, there have been both generic animal tracking data visualization interfaces like iNaturalist [2], and dedicated whale tracking interfaces like WhaleMap [4] (Figure 2), HappyWhale [1], Pacific Whale [3] and WhaleTest [5]. All the aforementioned options inform the user about the population density of whales, along with filters to choose time ranges, whale species, and regions of interest. However, there are perceptible limitations to the interactivity of these interfaces. The density map does not normalize the sighting data with the amount of time spent at each location, thereby not presenting a true picture of whale population densities. Further, most of these interfaces directly plot the raw data points on the map which can result in visual clutter or overlapping marks if the number of data points is not controlled using appropriate filters. Finally, these interfaces only visualize the data along with some options to filter the data, but no algorithmic analysis capabilities. We address these shortcomings in our interface by reducing the response time for interactions within reasonable thresholds of latency - 500ms [12], improving visual cognition by using the details on demand paradigm [17] to manage visual clutter or overplotting of marks, and by modeling the data as network data, open up new avenues for performing graph analysis tasks on the data to get insights.

3 DATA PREPROCESSING

This section describes the dataset used, and discusses the intuition behind behind and process of modeling the data as network data.

3.1 Dataset

We use the commercial whaling data collected by Allison [6, 7] and maintained by the IWC. The dataset covers the time period from 1880 to 2020, and covers all the oceans. It has details about whale catches which includes the timestamp, lat-lng coordinates and whales species, length, sex, etc., and details of the corresponding shipping expedition which caught the whale which includes the expedition code, the owning company and the nationality of the expedition. Records with missing timestamp and geographic coordinates were removed from the dataset to avoid noisy density maps. We thus got rid of 13k noisy records, and were left with a total of 2.1M records.

3.2 Modeling the Data as a Network

Given the details about the expeditions which hunted for whales, we can also map the routes taken by these expeditions, in addition to the population density. A network or graph like representation would serve well for visualizing this kind of data. Further, it also opens up new avenues of graph like exploration to be performed on the whaling data, examples of which are discussed in subsection 5.2.

We thus model this whaling dataset as a graph where the individual whale sighting locations represent nodes in the graph, and the routes taken by the shipping expeditions represent the edges in the graph. Each edge connects two nodes i.e., whale catch locations consecutive in time, for a single expedition. Since there are multiple whale catches recorded at a single location, we aggregate the details of the whale catches for each location, with a user configurable level of granularity. In case the level of granularity requires aggregation over multiple whale catch locations, the corresponding edges i.e., shipping routes are also aggregated. This graph data is visualized on a geographical map of the earth for spatial context.

We parse the data and generate two separate csv files, one each for nodes and edges representation of the whale catches and expeditions data respectively. For the 2.1M whale catch records, we were able to extract 0.5M shipping routes i.e., edges. Table 1 and Table 2 show a few samples of the data.

4 SYSTEM DESIGN

In this section, we discuss the current state of the framework, interface and the visualization, (Figure 1) and its implementation.

Node id	Lat	Lon	Date	Ocean	Area	Whale Species	Expedition code	Owning company	Land station/ Floating factory	Nationality
0	-35	118	1912-11-04	Indian	Australia W	Sperm	5030	Spermacet (Nielsen)	Vasco da Gama	Norway
1	24	-112	1924-11-28	North Pacific	Mexico	Humpback	5475	A/S Vega	Kommandoren I	Norway
2	-1	-85	1926-05-25	South Pacific	Chile/ Peru/ Ecuador	Fin	5409	Cia.Ballenera Del Ecuador Ltd	Whale	Equador
3	38	-28	1979-08-24	North Atlantic	Azores	Sperm	4260	NA	Azores	Portugal
4	-57	62	1961-12-04	Southern Hemisphere	Antarctic	Fin	6300	NA	Slava	Russia

Table 1. Whale catch data represented as nodes

Edge id	Src Node	Dst Node	Expedition type	Expedition code	Owning company	Nationality
0	282	283	Commercial	5030	Spermacet (Nielsen)	Norway
1	389	390	Aboriginal	1990	NA	US
2	694	695	Illegal	2470	NA	Korea
3	287	288	Special permit	6605	Research	Japan
4	89	90	No catch	1810	NA	Russia

Table 2. Shipping routes data represented as edges

4.1 Framework

We use a standard backend - server/middleware - client machine framework for our system. With millions of data points each having multiple attributes, it is reasonable to assume that the dataset may not fit within the main memory of the user machine. This implies that frequent accesses need to be made to the secondary memory during interactive analysis. Access to secondary memory usually entails high latency resulting in bad user experience. We thus use a standard Postgres database as the backend which allows for an efficient storage and also provides an optimized query engine which can return data to be visualized by the user as per filters set by the user.

Each user interaction in the interface is translated to a query in the server and sent to the backend. The resulting data is processed in the server and sent back to the client which renders the visualization.

Using a backend like Postgres also enables the use of spatial indexing on the location attribute for each node, which helps in efficient pan/zoom based browsing. Our current implementation does not utilise the indexing capability of the backend, which we leave for future work, along with offloading data preprocessing/precomputation to a distributed setup for speed and scalability.

4.2 Interface

Figure 1 shows a snapshot of our proposed Whale Map interface. The interface consists of the whale catch records visualized on a world map for geographical/spatial context. The visualization allows for pan/zoom interactions for the user to investigate certain regions in detail. Below the visualization, interaction widgets are provided for the user to set required filters. Filters can be set on the time period, and whale species to investigate. The user can also control the level of detail or the level of aggregation to visualize the data in; these levels of detail are described in subsection 4.3. The user can also control the encoding of the visualization by choosing between two options - visualize the raw whale catch count per node, or visualize the time normalized whale catch count i.e., whales caught per day.

4.3 Visualization

We adopt the details on demand paradigm [17] wherein the visualization is initialized with a coarse level or highly aggregated overview of the data. This allows the user to visually spot regions of interest which they can then choose to dig deeper into. Thus as required, the user can view less aggregated or fine level views of certain portions of the data by zooming into those regions.

The 'Ocean' level of detail shown in Figure 3a aggregates data for each of the 6 oceanic regions, resulting in 6 visual marks in the visualization. The 'Area' level of detail shown in Figure 1 aggregates data on the sub-region level specified for each whale catch as shown in Table 1. The 'Grid' level of detail shown in Figure 3b aggregates data for each 1x1 lat-lon grid, whereas the 'Raw' level of detail shown in Figure 3c simply plots all the data points as is without any aggregation. Finally, the 'Heatmap' level of detail shown in Figure 3d plots a kernel density estimate of the whale catches across the oceans.

Along with easing visual cognition of the data by visualizing limited number of data points, the details on demand paradigm also reduces the amount of data fetched from the backend thereby improving interactivity. However, in our current state of the interface, we have only provided the option to view the data in different levels of detail, it is not coupled with the zoom action. We plan to address this aspect as our future work, to truly harness the benefits of details on demand, for visual cognition and interactivity.

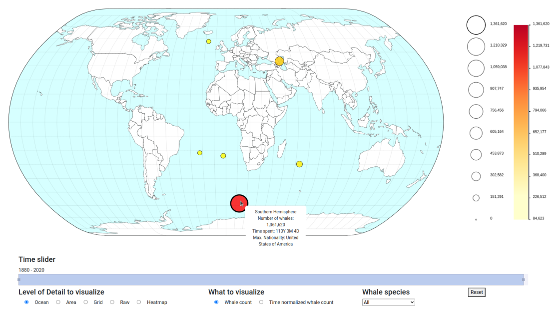
Both the whale count and time normalized whale count data is encoded using both a color scale and size of the marks. The heatmap is encoded only using the color scale.

5 USE CASES

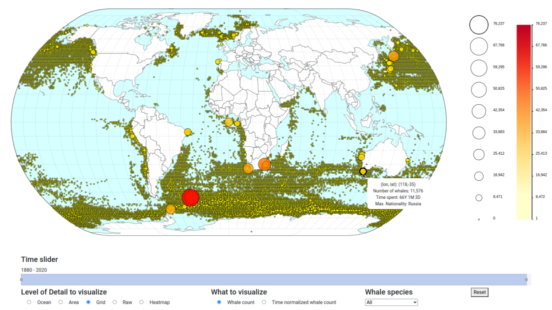
We now discuss some ideas on how this interface may be used for conservation purposes. These ideas came out of our meetings with our collaborator Dr. Trevor Branch.

5.1 Understanding past hunting patterns, events and their consequences

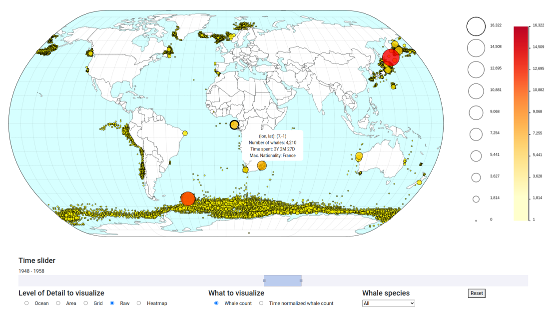
Historical whale catch records can help us understand how humans scanned the oceans for whales over time, and what happened to different populations of whale species as a result of this. Here we use the tool to look at Blue whale populations as an example, refer Figure 4



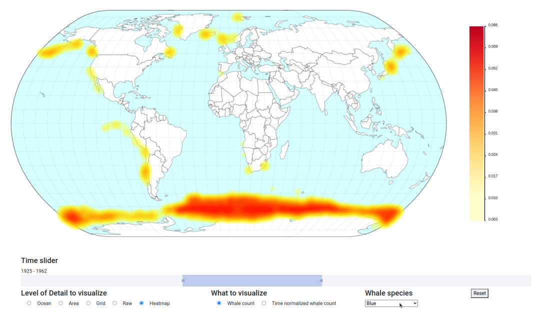
(a) 'Ocean' level of detail



(b) 'Grid' level of detail

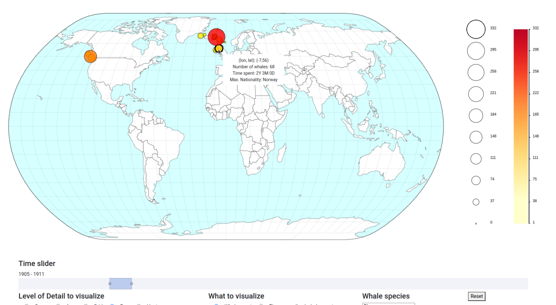


(c) 'Raw' level of detail

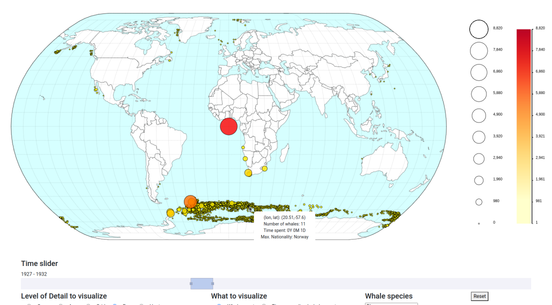


(d) 'Heatmap' level of detail

Fig. 3. Different levels of detail available in the interface



(a) Blue whale catches in the 1900s



(b) Blue whale catches in the 1930s

Fig. 4. Blue whale catches before and after the invention of steam powered explosive harpoons and the method to keep dead whales afloat

Blue whales swim very fast and sink when they die and thus it was difficult to hunt for them. These whales were being caught only along coastal regions, because of the difficulty in dragging a sinking whale from the deep sea to the coast Figure 4a. A Norwegian whaler named Svend Foyen invented steam powered explosive harpoons and the method of pumping air in whales to keep them afloat in the 1860s. With this it became easier to hunt for these whales even in deep seas. Consequently, the number of Blue whale catches kept creeping up gradually from around hundreds of whales in 5 year periods in the 1880s to around thousands of whales around 1910s, as whalers began hunting in deep seas. In the decade starting 1920, we see an explosion in both Blue whale catches Figure 4b, where the number went up to around 10k catches in such 5 year periods, as hunters then started exploring the Antarctic waters armed with the ability to keep whales afloat. This continued till the IWC moratorium to pause whaling came about in 1986, after which no more Blue whale catches were recorded.

5.2 Rerouting oceanic transport to avoid interference with whale populations

We have built this tool such that any new means of tracking whaling data can be used as a source and a real-time population density map can be generated, given the IWC data format is adhered to. With this map as one layer of information, we can overlay another layer of information, that of oceanic transport routes and hubs. We can then perform a shortest path finding graph analysis task to reroute oceanic transport around regions of high whale densities.

6 FUTURE WORK

The current state of the tool only visualizes the whale density maps with the ability to set desired filters on time, whale species, and set the desired level of detail and encoding.

To realise the potential of performing graph analysis tasks on this data, we need to compute aggregated edges corresponding to different levels of detail. These edge aggregations are non-trivial operations which cannot be expressed as simple SQL queries for the Postgres backend. Hence we need to precompute edge aggregations offline. Keeping in mind the speed and scalability aspect, we plan to offload this task to a remote distributed setup using the Apache Spark GraphX framework [3]. This framework also facilitates efficient graph algorithm implementations and graph data structures which we can use for in-house analysis.

Further, the current state of the tool does not leverage the spatial indexing capabilities of Postgres to facilitate efficient pan/zoom browsing, thereby missing out on true details on demand style visualization. We plan to address this aspect in future iterations of the tool.

We also plan to evaluate our tool at the hands of fishery science domain experts, while also incorporating their feedback alongside. Further, to make both the tool and expert insights accessible for use to the public, we plan to allow users to capture screenshots of the tool and annotate it with their insights so that domain experts can communicate their findings better, all in-house.

Our hope is that such a collaborative exploration tool can be extended to cover more animal species and can make people more aware about the changes in their populations and lead to better conservation efforts.

ACKNOWLEDGMENTS

Special thanks to Trevor Branch (University of Washington, College of the Environment - School of Aquatic & Fishery Sciences) for the collaboration opportunity, access to the International Whaling Commission Dataset and his guidance and feedback for developing the tool; to Leilani Battle (University of Washington, Paul Allen School of Computer Science and Engineering) for the collaboration opportunity; and to Matt Ziegler and Kurtis Heimerl for their feedback and inputs, and the engaging discussions throughout the duration of the course.

REFERENCES

[1] 2022. HappyWhale. https://happywhale.com/browse
[2] 2022. iNaturalist. https://www.inaturalist.org/

[3] https://spark.apache.org/graphx/

[3] 2022. Pacific Whale. https://www.pacificwhale.org/research/hawaii-wdt-sightings-map/

[4] 2022. WhaleMap. https://whalemap.org/WhaleMap/

[5] 2022. WhaleTest. https://michellemodest.shinyapps.io/whaletest/

[6] C Allison. 2020. IWC individual catch database Version 7.1. Date: 23 December 2020.

[7] C Allison. 2020. IWC summary catch database Version 7.1. Date: 23 December 2020.

[8] C. C. G. Bamford, N. Kelly, L. Dalla Rosa, D. E. Cade, P. T. Fretwell, P. N. Trathan, H. C. Cubaynes, A. F. C. Mesquita, L. Gerrish, A. S. Friedlaender, and J. A. Jackson. 2020. A comparison of baleen whale density estimates derived from overlapping satellite imagery and a shipborne survey. *Scientific Reports* 10, 1 (Dec. 2020), 12985. https://doi.org/10.1038/s41598-020-69887-y

[9] Line Bang Christensen. 2006. Marine mammal populations: reconstructing historical abundances at the global scale. (2006).

[10] Brooke E. Flammang, Simone Marras, Erik J. Anderson, Oriol Lehmkuhl, Abhishek Mukherjee, David E. Cade, Michael Beckert, Jason H. Nadler, Guillaume Houzeaux, Mariano Vázquez, Haley E. Amplo, John Calambokidis, Ari S. Friedlaender, and Jeremy A. Goldbogen. 2020. Remoras pick where they stick on blue whales. *Journal of Experimental Biology* 223, 20 (10 2020). https://doi.org/10.1242/jeb.226654 arXiv:https://journals.biologists.com/jeb/article-pdf/223/20/jeb226654/1981426/jeb226654.pdf jeb226654.

[11] Elliott L. Hazen, Daniel M. Palacios, Karin A. Forney, Evan A. Howell, Elizabeth Becker, Aimee L. Hoover, Ladd Irvine, Monica DeAngelis, Steven J. Bograd, Bruce R. Mate, and Helen Bailey. 2017. WhaleWatch: a dynamic management tool for predicting blue whale density in the California Current. *Journal of Applied Ecology* 54, 5 (Oct. 2017), 1415–1428. https://doi.org/10.1111/1365-2664.12820

[12] Zhicheng Liu and Jeffrey Heer. 2014. The Effects of Interactive Latency on Exploratory Visual Analysis. 20, 12 (2014), 2122–2131. https://doi.org/10.1109/TVCG.2014.2346452

[13] NOAA. 2021. Tracking technology: The Science of Finding Whales. https://www.fisheries.noaa.gov/feature-story/tracking-technology-science-finding-whales

[14] Joe Roman, James Estes, Lyne Morissette, Craig Smith, Daniel Costa, James McCarthy, James Nation, Stephen Nicol, A. Pershing, and Victor Smetacek. 2014. Whales as marine ecosystem engineers. *Frontiers in Ecology and the Environment* 12 (09 2014). https://doi.org/10.1890/130220

[15] María Romero, Mariano Coscarella, Grant Adams, Juan Carlos Pedraza, Raúl González, and Enrique Crespo. 2022. Historical reconstruction of the population dynamics of southern right whales in the southwestern Atlantic Ocean. *Scientific Reports* 12 (02 2022), 3324. https://doi.org/10.1038/s41598-022-07370-6

[16] Adolf Seilacher. 2005. Whale barnacles: exaptational access to a forbidden paradise. *Paleobiology* 31, S2 (2005), 27–35. https://doi.org/10.1666/0094-8373(2005)031[0027:WBEATA]2.0.CO;2

[17] B. Shneiderman. 1996. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*. 336–343. https://doi.org/10.1109/VL.1996.545307