A Study on Data Visualization for Fishery Management

by

Volodymyr Kozyr

B. Sc., Taras Schevchenko National University of Kyiv, 2017

Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the

School of Computing Science

SIMON FRASER UNIVERSITY
Fall 2021

Declaration of Committee

Name:	**Volodymyr Kozyr**
Degree:	**Master of Science**
Thesis title:	**A Study on Data Visualization for Fishery Management**
Committee:	**Chair:** Binay Bhattacharya Professor, School of Computing Science
	Frederick Popowich Senior Supervisor Professor, School of Computing Science
	Steven Bergner Supervisor University Research Associate, School of Computing Science
	Kay Wiese Examiner Professor, School of Computing Science

Abstract

Fishery data visualization plays a vital role for companies in countries with developed fishery infrastructure. It is important to identify trends, anomalies, year-by-year comparison (drill-down) of fishing catches and their monetary value. Showing correlation between these two parameters for each Canadian province and fish species is a prominent part of the work. Our focus is to provide more user-friendly ways to show all necessary data for non-domain expert users in a web browser.

We introduce a novel tool called FishPlots that can help users see/identify fish amounts in specific regions and help fishery managers add or remove restrictions on fish quotas. It helps analysts answer questions about the number of caught species or possible threats to biological diversity. Having the visualization tool in a web browser is convenient for users to allow them to perform their analyses more quickly and more accurately, plus they do not have to install software on their computers that may not be compatible with their systems. The proposed web interface will be a highly accessible multi-platform tool that will help users interested in the fishery domain analyze table data quickly and effectively.

Keywords: fishery; web-interface; visualization; online platform

Dedication

Dedicated to my Grandfather, George Kozyr, the strongest man I know.

Acknowledgements

I would like to acknowledge and say many thanks to my supervisors Fred Popowich and Steven Bergner. I cannot imagine how hard this work would have been without your help and advice. The value of our discussions is unmeasurable.

I am extremely grateful to my family for support. Thanks to my wife Zhenya, sister Lena, dad Volodymyr and mother Olena who constantly pushed me to move on my work and helped when I was in doubt or confusion.

Table of Contents

List of Figures

List of Acronyms

DFO	Department of Fishery and Oceans

Chapter 1. Introduction

This project addresses the challenge of making it easier for users to make work involving data analysis faster and more productive. While there has been a great deal of effort devoted to the development of software that allows users to access to raw table data, support data processing, analysis, and report generation, getting value from data can still be a very time consuming activity. Of course, digging into raw data might give the expected results, but it usually takes significant time to analyze a big amount of data, especially if users are not familiar with computer software.

Visual data representation plays a crucial role in data analysis. It can condense vast amounts of data into several plots and labels, giving information about trends; it is also much easier to compare pictures than data rows. For example, it can help users to make life-saving decisions [1] using a tool, designed for doctors to monitor the state of people with diabetes which will be discussed more in chapter 2, or it can allow financial analysts to make better decisions more quickly [2].

While visual representations can be used across a wide range of data sources, we are interested in how they can be used in the analysis of data associated with fisheries. Fishery data, like data from other domains, deals with observations across time, and across space. As such, many of the visualization techniques that have been applied to temporal or spatial data can be considered in our work. What is particularly interesting about fisheries data is that depending on decisions made after analysis it can lead to different economic and environmental consequences. Also, there are a large number of open data sources that we can draw upon.

This work will help people who are domain experts but are not very familiar with data analysis tools. The tool can help fishery management regulate fishery catching in certain Canadian provinces and help decide which policies or fishing quotas for specific fish types should be applied.

In our project, we will start by introducing some related work in section 2. Then in section 3, we will discuss visualization tasks related to the marine industry, what kind of data is important for the fishery, visualization tool requirements and four use cases where the tool might be used. Chapter 4 will introduce our tool named FishPlots:

https://github.com/vladymyrkozyr/FishPlots, giving details of the implementation of the system and its usage. And finally, in chapter 5 there will be conclusions and discussions on how the FishPlots may be improved.

Chapter 2. Related Work

2.1. Role of the Fishery Visualization

Visualization can play an important role in decision support in fisheries information systems. It can give a person who is working with fishery information more insights about data. As in other domains, it can save time for making correct decisions about the fishing company's business logic as in [3], because it is easier to see trends and outliers while using charts and interactive diagrams rather than just looking through a spreadsheet, sometimes located even across multiple files.

There are different groups of users in the fishery domain like fishery management companies, environmental policy issuers and simple fishermen etc. All of them need to get different data and analyze it in some way. For example, marine environmentalists analyze data about the quantity of fish in a particular region, they issue quotas for fishery companies. After that companies analyze the fish stock market, plan, and distribute information to their employees (fishermen) in a way that it is optimized and profitable.

Papers discussed in this chapter are oriented primarily for fishery management, because papers for environmentalists, for example, include more information about biological and ecological perspectives rather than focusing on visualization methods. However, the content is complicated for readers who are not data analysts or/and experienced computer users.

So, the interest and primary goal of our work and research are in the design and use of tools that will be easily accessible for fishery management users as it needs to have more charts, visualizations, etc. than other groups. The tools need to support the decision-making process, so that appropriate conclusions can be made based on the available data. It needs to support data from various sources for effective marine environment management. In this chapter, we will examine some of the previous work in this area, and their approaches to the use of visualization.

2.2. Marine Environmental Management

FishCAM2000 (FC) [4] is a computer-based integrated information system for fisheries management and marine environmental monitoring. It incorporates visualization of the fishery activity over the same spatial area of interest for a particular type of fish. It has a user-friendly and straightforward interface implemented in Windows Forms. Similar topics of special fishery visualizations are discussed in papers Spatial and Temporal Differences in the Reproductive Traits of Skipjack Tuna [5] by Hiroshi Ashida and Evaluation of a US West Coast Groundfish Habitat Conservation Regulation via Analysis of Spatial and Temporal Patterns of Trawl Fishing Effort [6] by Marlene Bellman and others.

Figure 2.2.1. below shows the output after the user passes seven screens of settings of the query wizard. While entering so many screens of data may suggest a large amount of effort before one obtains any results, this process does allow the system to be as generic as possible but requires a great deal of effort and time from the user. One important advantage of the system is that it presents complex geodata, which includes the amount of fish caught in the geographical zone (polygon area shapes with latitudes and longitudes as data points) on the map instead of the data table, which is a considerable timesaving for users to understand the query output.

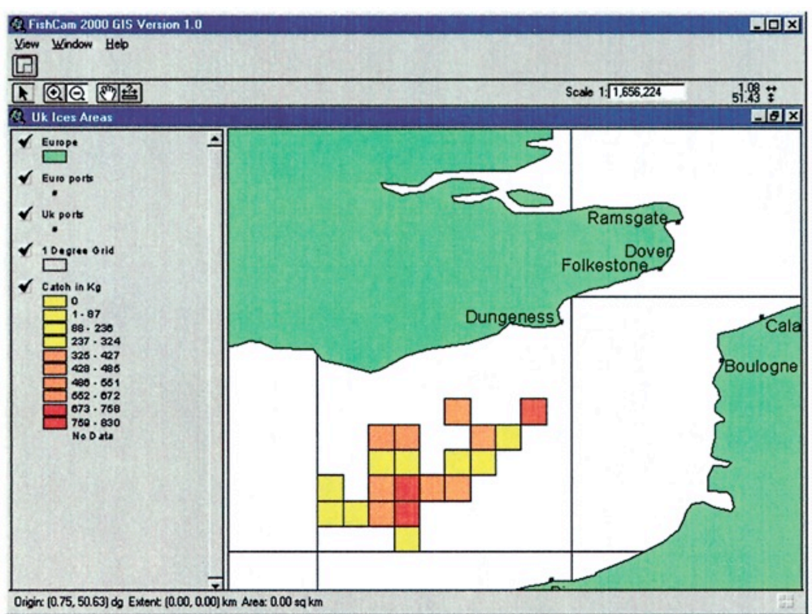


Figure 2.2.1. FishCAM2000 user interface; Source: [4]

While the work mentioned above was focused on summarized data, there is subsequent work related to a tool that gets and analyzes data that is directly coming from vessels [7]. It also depicts geographical data and the number of fish caught by a particular vessel in a specific region. The authors show not only the geographical spread of fish but also the amount caught using colour coding (Figure 2.2.2.). In this way, user can easily analyze both quantities at the same time. Also, charts are presented for different months for the user to know how the fishing season was completed.

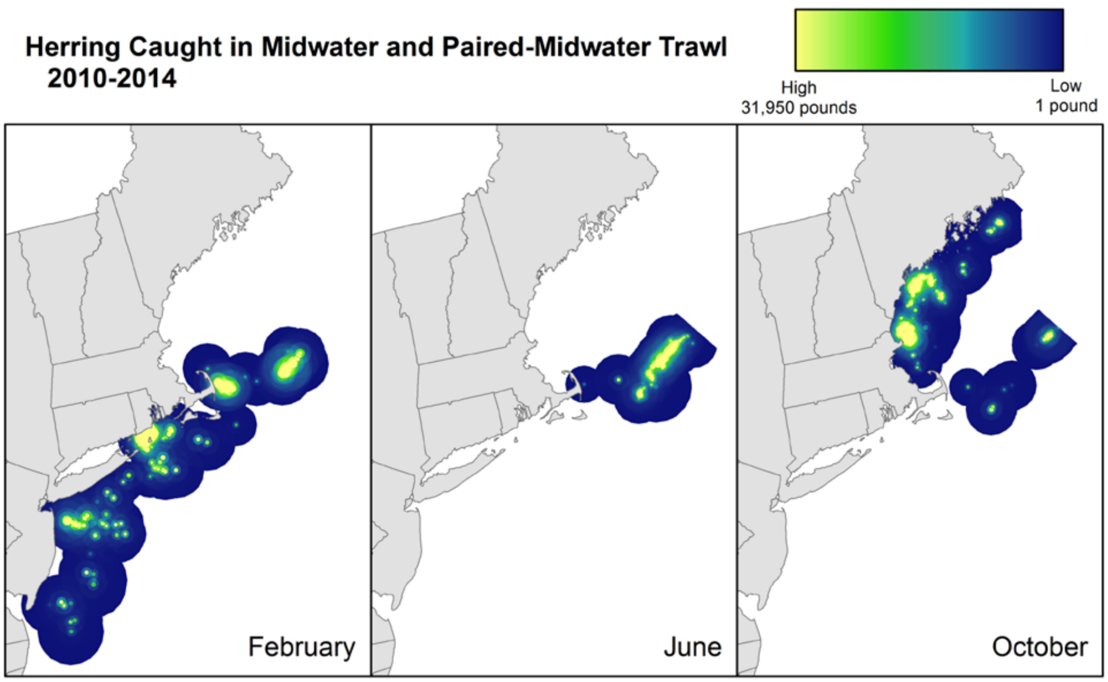


Figure 2.2.2. Heatmap fish trawl visualization; Source: [7]

Whereas the work we have discussed so far focussed on the visualization of geographical and vessel-related data, the proposal of da Silva, Charles Fulcher [8] allows the investigation of land-sea connections. It shows the human impact on the sea from land, and vice versa. The maps show the connection between vessels and ports and depict the distribution of gear types used in different regions using descriptive labels, lines, and appropriate legends that any person can easily understand (Figure 2.2.3.).

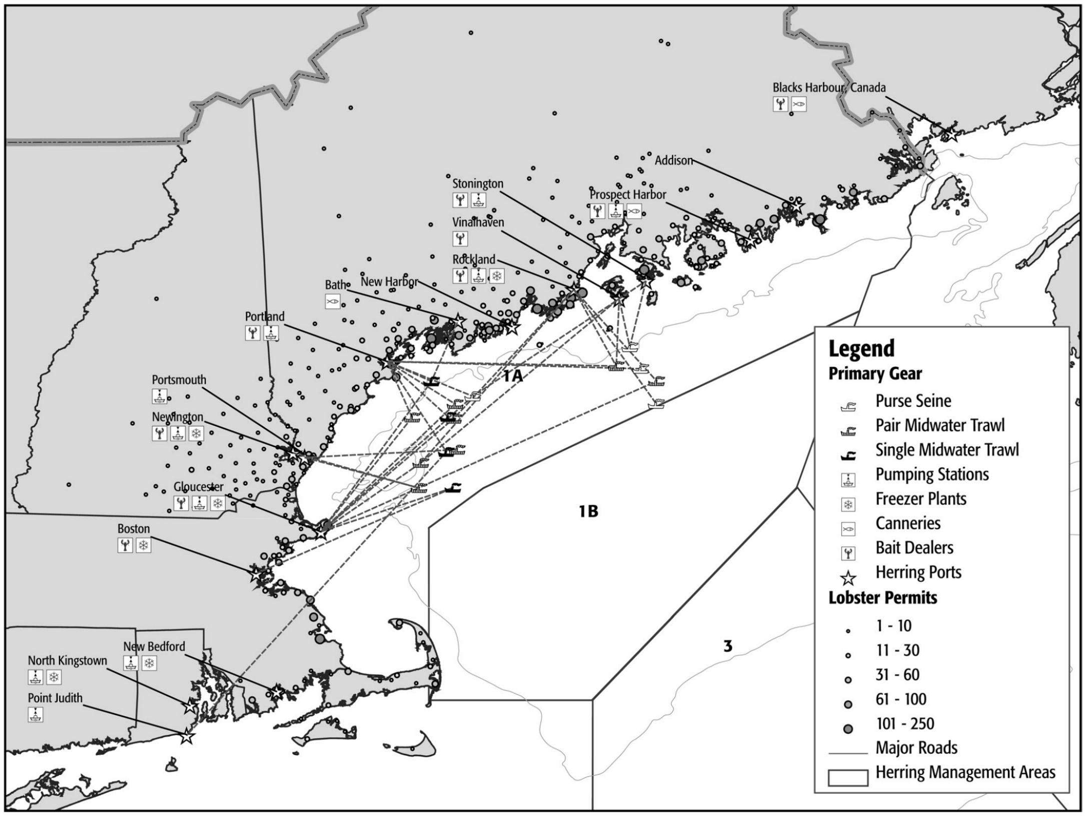


Figure 2.—Herring Management Area 1A winter/spring activity, 2000–03.

Figure 2.2.3. Land-sea connections; Source: [8]

Finally, it is also important to note the work of Barrus about the state of the salmon [9], which provides a prototype for a platform to support analysis and interaction for fishery data through visualization. It uses DFO (Department of Fishery and Oceans) Canadian data related to salmon in BC rivers, and the author discusses and produces visualizations for questions that marine experts asked him to solve. In particular, in his work, he states that all data sources for DFO are decentralized and there is practically no interface that allows users to analyze data. As a result, he combined several data sources and produced visualizations that answer questions about the state of salmon in Canadian rivers.

Now that we have seen a representative selection of systems and approaches to dealing with fisheries data in tasks related to fisheries management, let us now look in more detail at the role that visualization can play in these systems. When looking at the role of visualization, it will be important to see how approaches that have been used with various forms of data might be applied to our fishery specific tasks.

2.3. Approaches to Visualization

Visualization can be important to support predictions made from data, which can be seen in a paper that works with time series data, on a task related to the prediction of stock prices [2]. Figure 2.3.1. shows the difference between training predictions, test predictions and also real values of stock prices (x-axis is time, y-axis is price), but the visualization part requires special attention. Authors used different color coding to show results, and there is no extra information on the chart which distracts or confuses readers.

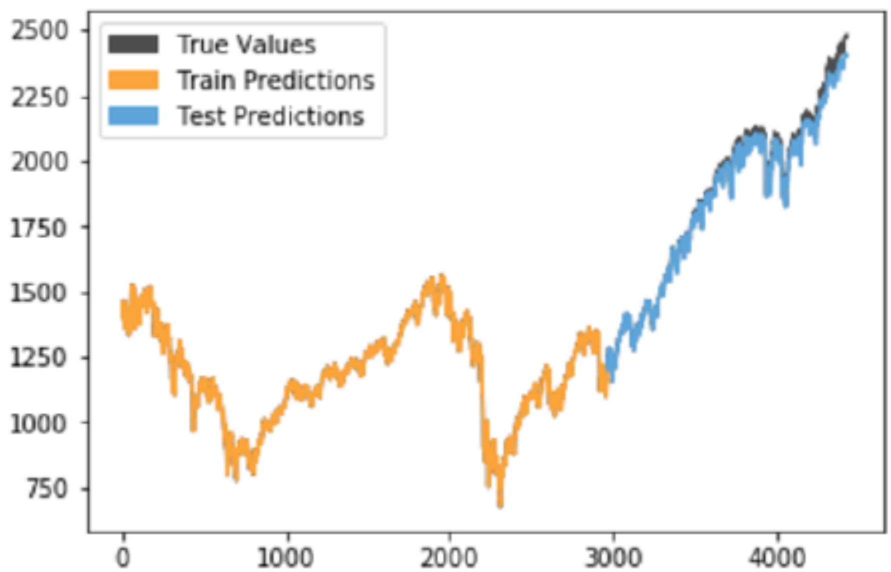


Figure 2.3.1. Color coding example; Source: [2]

IDMVis [1]: is a visualization tool that shows multidimensional interrelated data during the day for patients with diabetes. IDMVis includes a novel technique for folding (splitting a sequence into periodic units like hours, days, etc.) and aligning records by events of interest and scaling the intermediate timeline. The designed tool helps doctors track the state of patients' important parameters and detect anomalies (Figure 2.3.2.). After that, doctors use it as a decision support tool for the treatment of diabetes.

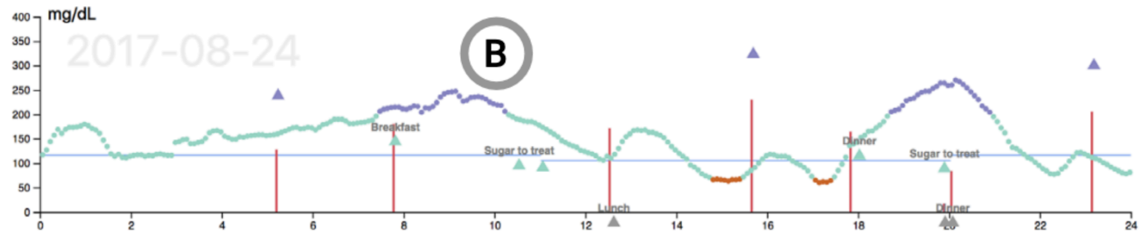
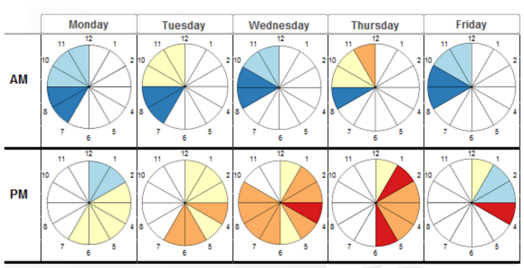


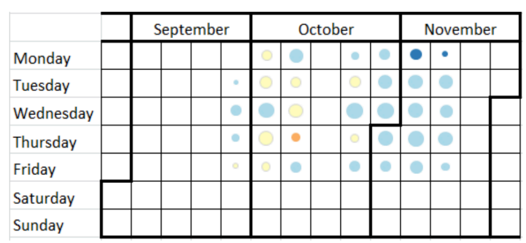
Figure 2.3.2. Daily sugar level for patient; Source: [1]

Six clinicians evaluated design decisions positively, the criteria of the evaluation was how well proposed visualizations help with the decision-making process on daily basis (increasing of the decision-making speed). Also, that detailed visualization convinces doctors to prescribe were more personalized treatment for the patient.

The works we have presented so far were generally focussed on one single visualization technique, but in [10] by Semikina, there in a comparison of methods of visualizing the same data in various amounts of charts and diagrams of different types. She uses bar charts, line charts, pie charts, spiral charts. Semikina also includes user studies in her work, showing which particular visualizations users understand better than others (Figure 2.3.3.).



a) stress values aggregated per hour



b) showing the amount of data available for visualization

Figure 2.3.3. Various visualization types; Source: [10]

We have now seen several approaches to visualization, which can provide the basis for our investigations in fisheries data. We can now turn our attention to the specific tasks of interest for data visualization.

Chapter 3. Design and Visualization Tasks

The environment and nature change over time, so scientists and environmentalists interested and responsible for the sustainability of different fish populations are using data analysis tools to determine if there are problems with biodiversity in a particular region. For instance, at the species level, fishing can reduce abundance and alter physiology and lifestyles, thereby affecting species' role in the biological community. Besides, fishing can cause changes in the open ocean community's trophodynamics and reduce the biodiversity and resilience of these ecosystems. That is why scientists work to detect any unexpected fish population trends and design methods, taking steps to address the identified issues. For getting to such conclusions, they are going through an enormous amount of data. In these cases, visualization tools help them to understand the problems faster and apply correct actions if needed. Policymakers and fish company stakeholders are also involved in discussing the fishery data state. Policymakers need to issue quotas for the fishing season in a particular water zone for a specific fish type to avoid such a problem known as overfishing.

Visualizations play an essential role in communicating for fisheries scientists, environmentalists, and policymakers. In [3], Levonin discusses how policymakers, stakeholders, and fishery management communicate with each other, providing a visualization tool that makes the process easier on all sides. From the articles and papers discussed in the following paragraphs, it is becoming clear that more and more specialists are preoccupied with the fish population in the world ocean. There are groups of users interested in fishery data analysis for different purposes, but they need visualization tools to make their work more productive.

3.1. Fishery Reports

Reporting on the status of aquatic resources and how they are changing over time, along with understanding the impacts of large-scale disturbance and human activities on aquatic resources, are among the most critical processes for assessing the state of ecosystems. To display the human influence on an ecosystem or a specific type of fish, data on the amount of catch of this species by region, the number of the species in comparison with previous years, data from the SARA and COSEWIC organization about species at risk can be helpful.

For example, the 1993 Report on the status of groundfish stocks in the Canadian Northwest Atlantic [11] reported cod catch data. Authors used multiplicative analyzes of catch-at-age from research surveys and the commercial fishery and, on average, fishing mortalities from preliminary SPA. Based on these data, it was concluded that fishing mortality increased according to multiple reasons (page 17) and that there has been a "gradual decline in stock size through the second half of the 1980s". This conclusion can be drawn from the following presentation of the data (Figure 3.1.1.), where in the total biomass section user can see that for every next year amount of cod is decreasing.

2.2. COD in 3Ps
SUMMARY
Figs. 1.2a - 1.2f

Year - Année	1986	1987	1988	1989	1990	1991	1992	1993	Min.[1]	Med.[1]	Max.[1]
Reference level - Niveau de référence '000t	26-61	26-58	37	20.5	N/A	29-44.5	39-44.5	20			
TAC - TPA '000t	60.6[3]	60.6[3]	60.6[3]	50[3]	45[3]	44[3]	44[3]	20			
Rep.catches-Prises décl. '000t	57	57	43	39.5	41[2]	43[2]	31.5[2]		27	49	84
Unreported catches Prises non-déclarées											
Est.discards - Rejets est.											
Tot. catches - Prises Tot.											
Tot.biomass-Biomasse Tot.[4] ('000t)	200	168	144	141	139	111	81		80	168	293
Sp.biomass-Biomasse Rep.[4] ('000t)	60	57	49	43	36	28	24		18	58	123
Mean - F - Moyen (6+)[4]	.55	.66	.59	.46	.58	.70	.70		.33	.54	1.16

[1] For/Pour 1959-1992
[2] Preliminary statistics/Statistiques provisoires
[3] This is the effective TAC which was obtained by combining the Canadian quota and the French quota of the TAC set by each party/TPA "defacto" obtenu en additionnant les quotas établis par le Canada et la France
[4] Not calibrated, for illustrative purposes/Non ajusté, calibré pour illustration

Figure 3.1.1. Table data presentation; Source: [11]

Similarly, this tabular format allows the reader to see, in the same report, data on the amount of cod catches from 1960 to 1994. The user can compare these data with the data on the population of different age groups of this species, fish mortality, recruitment, supply and demand and other data. On this visualization user has to look for the value of interest and compare each column year by year to figure out if there is a trend pattern (increase/decrease) in the data.

Instead of providing information in tabular format, line charts presented in different colours give the reader a clear and quicker understanding of trends and processes ongoing. The Stock Status Report [12] for the 2004 year provided abundance data for Northern Abalone for each year. Since 1998, this species' abundance has fallen below the permissible short-term recovery objective line. This can be inferred from Figure 3.1.2. listed below. After the year 1998, data points are located below the line which is located on the level of 40%.

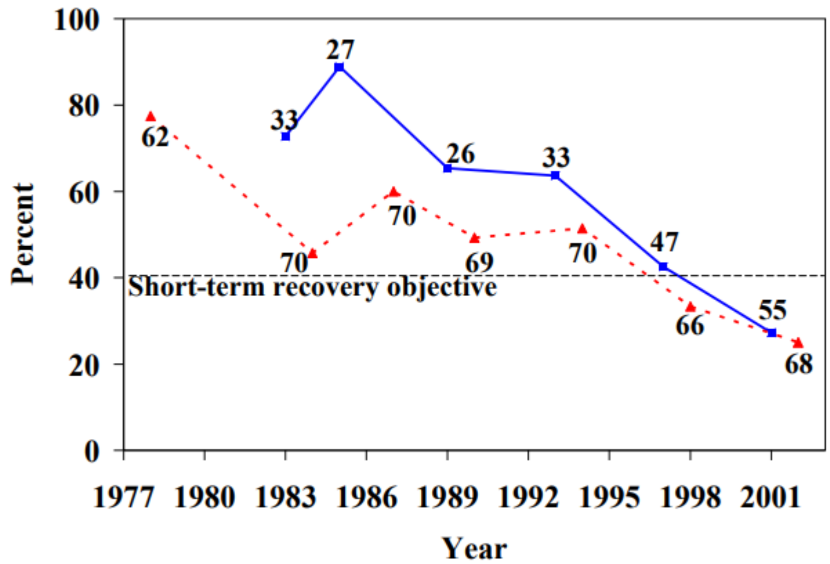


Figure 3.1.2. Northern Abalone data; Source: [12]

The same approach of using line graph visualizations can be seen in a more recent annual environmental report on this kind of data. For example, Scottish Sea Fisheries Statistics [13] for 2019 from the Cabinet Secretary for Rural Economic and Tourism provides summary data and more detailed statistics by region.

From the information related to this topic, the following can be distinguished.
The total amount of fish caught (tonnage) is shown in blue, and the value of all landings by Scottish vessels is shown in red. By comparing the blue and red lines in Figure 3.1.3., the user can easily and quickly determine the connection (will be discussed in the following chapter) between the catch's value and the amount of fish caught in a specific

year. For instance, the user can see that despite the tonnage falling since 2017, the value of landings remains constant.

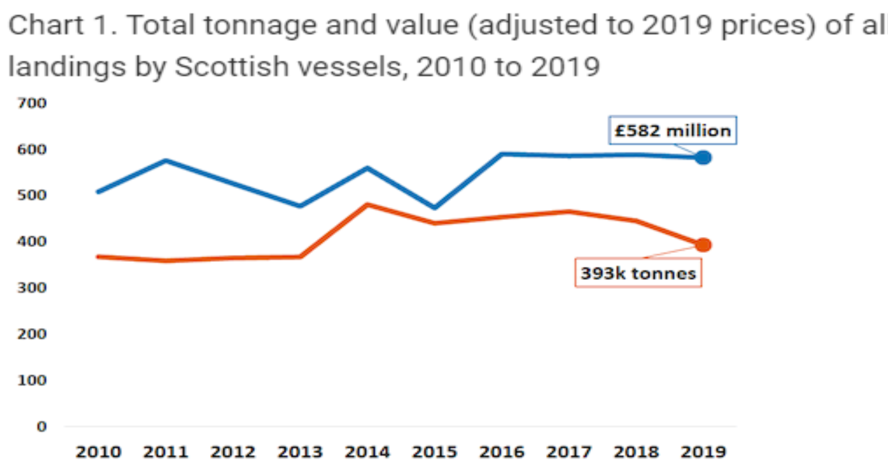


Figure 3.1.3. Tonnage and value for fish landings; Source: [13]

We can see that the marine industry requires to see overall reports on such parameters as “tonnage” and “value” (Figure 3.1.4.). From the pie charts, users can determine which type of fish gives which revenue according to tonnage.

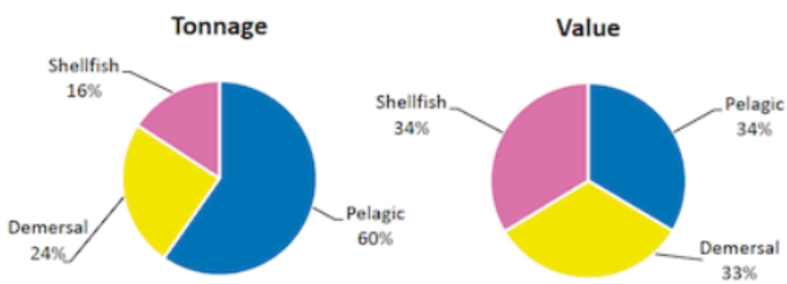


Figure 3.1.4. Percentage of Scottish vessels’ landings by species type; Source: [13]

In the next report (Figure 3.1.5.) there are bar charts for two consecutive years presented. This is done mainly to see trends and then to decide if the fishery industry is doing better or worse than the previous year. After comparing values for two years some adjustments in fishery policies could potentially be implemented by ecologists or fishery companies.

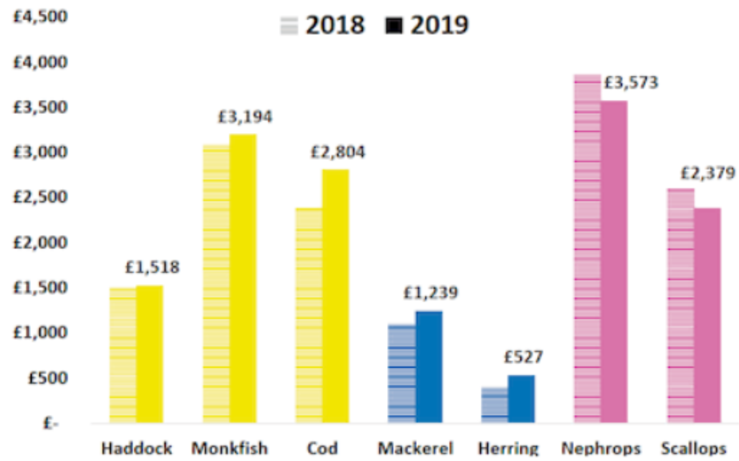


Figure 3.1.5. Price per tonne for key species for years 2018 and 2019; Source: [13]

The Scottish Sea Fisheries Statistics report [13] is an excellent example of data visualization to assess both the environmental and economic side of a problem. The reports described above cover different periods of time and fish types.

Summarizing the importance of data visualization in comparing a larger amount of data it can be seen that the data provided is very similar and differs only in the methods of their visualization. Some provide tabular information, others provide graphs, line plot, bar chart, pie chart, scatter plot or bubble chart. Moreover, each of the visualizations is aimed at displaying specific data for a particular user's request.

However, there are several disadvantages associated with the examples discussed above. For instance, in tables, data for information analysis is presented exclusively in vertical and horizontal numbers, and there are no visible accents, like, colour. In this case, it slows down and complicates practical visual assessment and analysis of information. For example, using tabular data as opposed to the IDMVis tool [1] presented in chapter 2 would be more complicated for doctors to select correct treatment. In turn, when in other examples the information is presented using colour graphics, there are no labels that explain the missing data for analysis. Consequently, even with these examples, it is difficult for a user to comprehensively evaluate and analyze information.

Furthermore, in all the described work examples, information is presented in static images, which does not allow interactive intercommunication with information and work with data in real-time. For example, users might want to filter data based on a specific criteria, remove unnecessary data according to their business logic, sort data, etc.

3.2. Data Sources

According to National Research Council in their 2000 paper on Improving the Collection, Management, and Use of Marine Fisheries Data, “The phrase ‘fisheries data’ is a general way of referring to data that may be of use in the management of a fishery as well as for commercial, recreational, cultural, and scientific purposes.” [14] Such data includes biological information on operational fish and related species, economic information from fishermen and markets for catch and environmental information affecting the species' productivity. A primary source of information is fishery-dependent data. So-called trip tickets or logbooks contain the timeliest information on current fishery conditions. Secondly, catch sampling programs are an essential source of information provided by observers placed on commercial fishing vessels to collect data based on the species composition, sex ratio, and age composition of the catch. And finally, scientific surveys are the primary source of fishery-independent data, including estimates of fish populations' age structure and relative abundance of stocks.

The scope of fisheries data use is very wide. It ranges from stock assessment by scientists to strategic planning by industry, and fishery monitoring and allocation decisions by managers. As mentioned in “Improving the Collection, Management, and Use of Marine Fisheries Data” report, “each use implies a set of users and a suite of requirements that the data must satisfy, including timeliness, level of detail, accuracy, accessibility to users, coverage or completeness, and credibility of the data collection process and the management process that uses the data.” [14]

We can emphasize that fisheries data are vital to strategic planning activities in coastal communities that rely on fisheries. At the same time, civil authorities use fisheries data to site marinas, underwater pipes and cables, and other maritime facilities and develop infrastructure for the fishing industry. Additionally, bankers use fisheries data to plan economic development and loan packages to fishermen, fish processors, and ship suppliers. And of course, fishermen themselves, besides their logbooks and observations and what they learn from other fishermen and buyers, use government fisheries data to plan shifts to new fishing grounds, changes in fishing gear, and changes in species targeted. Environmental and other interest groups also have become increasingly involved in monitoring fishing activities. Monitoring often requires data with great detail in both time and space as well as frequent updates, often within a fishing season. We need to note that among the most vital users of fisheries data stand stock assessment, government and private sectors' scientists, national, and international fishery agencies, and universities.

From the above, we must consider that when creating software for analyzing fisheries data, we must assume that users' requirements will vary. The system should help users who require data with various spatial resolutions and different levels of timelines. Our focus for the data source is the DFO (Department of Fisheries and Oceans) Canada website for provincial sea fisheries. Several factors motivated this decision. Firstly, data is provided on a Canadian government website, so users can rely on the data with a high level of trust. Secondly, it is publicly available data, so users can check the source data if they want. And thirdly, data is updated on regular basis, so it will be available for the following years as well (not only historical data).

The data tables provide the volume and value of sea fisheries landings. Data is organized by species groups, by main species and by province, covering the period from 1990 to the year before the current fishing season. Note that data before 1990 are available upon request, but we did not need this earlier data for our investigations. All data is collected by the DFO regional offices. All data is separated by year (one Microsoft Excel file for each year) and grouped inside by Canadian provinces for each fish type. In total, this dataset consists of 56 files (28 for the fish amount in tonnes and 28 for a fish total value in thousands of Canadian dollars.

To analyze this type of data, we used several processes which include data extraction, cleaning/filtering, and massaging (converting into suitable format) which will be discussed in the implementation part of the work in Chapter 4. Before looking at the processing of this data, let us first consider the reasons why one might want to use this data.

3.3. Fishery Domain Problems

There are several problems on which marine scientists are eager to work based on published research papers that we have identified. They have some guides for fish population and environment monitoring. [15] Here are some problems that seem to be well suited to the use of visualization techniques.

1) determining the optimal amount of catch for each type of fish to reduce environmental damage in a specific region. The visualization of the problem analysis can be presented in the form of a bar chart growth/decrease in the catch of a specific type of fish for a particular period in different regions.

2) understanding how the catch affects the ecosystem and other species. The visualization of the problem can be presented in the form of graphs of two colours. One of them shows the increase/decrease in the catch, and the graph of the other colour shows the diversity of the ecosystem in the number of species.

3) finding out what species of fish and other animals are on the verge of extinction. Visualization involving pie charts can show endangered species populations by year.

4) identifying the regions that need attention in the first place and introducing measures to restore or prevent species' extinction or changes in the ecosystem. The problem can be visualized in a geo map, where endangered species in a specific region are shown in different colours.

5) predicting which species may also be subject to negative or positive effects (trends). A line chart would be appropriate here, defining the trend in the development of species over some time.

6) establishing quotas (the proper amount of unloading of fish) which will minimize the negative impact on the environment. The visualization could be a bar graph showing the dependence between the allowable catch and the accepted catch quotas.

7) determining the recovery rate of the ecosystem after taking appropriate measures can be displayed using two-line graphs of different colours.

8) analyzing the safety of methods for catching a particular type of fish in each region. A comparative bar chart can be used to visualize the relationship between safe fishing practices and the fish population in each area.

Based on the results of the analysis of problems like these, scientists can suggest adjustments to methods that have adverse effects on the state of the environment and suggest alternative methods. In a 2008 publication of Bedford Institute of Oceanography [16], which described many factors affecting different ecosystems, aquaculture research, various data collection methods and technologies, the authors mentioned Recovery Objectives for North Atlantic Right Whales, among which the following issues were described:

- reduction of death and injury from vessel strikes

- reduction of death and injury from fishing gear entanglement

- reduction of injury and disturbance from vessel noise, exposure to contaminants, and other forms of habitat degradation

- monitoring populations and threats

Based on these data (amount of injured or accidentally killed animals and detailed reports on each case) it can be concluded that the environment is adversely affected by the rising amount of fishing vessels and equipment for fishing in the ocean.

The Federal Register - The Daily Journal of the United States Government [17] provides data on Marine Mammal Species and Stocks Incidentally Killed or Injured during the 2018 fishing catch. By obtaining the average annual data on the number of accidentally killed or damaged fish during the catch for Canada and comparing it with the data about the number of fish caught in the same year, it is possible to estimate the environmental damage. Depending on the amount and detail of the data, it is possible to estimate the damage for a specific type of fish and/or for each region. Also, fish species can be combined into groups according to their catching methods, and ecological influence can be analyzed for each group.

3.4. Visualization Motivation

Analysis of the data presented in a table or text format may take significant amount of time, as was discussed in chapter 3.1. on IDMVis tool use cases. For example, it is hard to see trends, how values change through the years, as well as comparing data for different provinces and fish type. Another thing that is hard to capture is the correlation between price and quantities of specific types of fish for a certain period.

FishPlots itself and its implementation will be discussed in Chapter 4; however, it is worth mentioning that it is developed for people who may not be data scientists. The main goal is to make it usable for people with average knowledge about computers. It also will not require any installation steps, because it is a web application that can be accessed just by typing a URL in any modern browser. Another feature of FishPlots is that it will allow the user to select range, provinces, and any fish type from dropdowns, zoom into details, etc. This type of UI experience is not available if it is done through Excel charts or the Python library PyPlot [18]. Visualizations in these cases are static and should be re-rendered if some parameters of visualization are changed. Also, it requires advanced knowledge of Microsoft Excel or programming.

Let us now introduce our design requirements for visualization. They were determined based on the informal task abstractions discussed above and literature research involving real reports for which people use for marine operations from section 3.1.

Requirement 1. Interactivity

The main difference between existing static reports and FishPlots is that we allow users to interact with data before producing a visualization. These interactions involve operations like filtering, zooming, and so on.

Requirement 2. Data Scaling for Further Analysis

FishPlots should allow users to discover patterns, trends, and anomalies, thus related to Task 1 and Task 2 from the informal task abstractions.

Requirement 3. Summary and Overall Statistics

Visualizations for Task 3 and Task 4 are designed to show global summarized data.

3.5. Task Abstraction

3.5.1. Task 1. Exploring Relationships for Fish Amount and Price

In section 3.1. we introduced a report [13] that shows the relevance of charts for the fish amounts and price for people who work in the marine industry. This work contains plots for fish amount and prices separately on different figures. They also have different visualizations (bar charts, line charts, pie charts) for the prices and tonnage of the fish of a specific type.

Improvement of this report could involve placing the data on one line chart with two axes to help the user see both measurements at the same time. This helps understand the correlation between the values and quantities without a need for looking into two different charts (they may be in different scales etc.) or comparing row table data.

Visualization (combined with external data sources and/or user's knowledge and experience in the domain) may be used by users for solving a range of issues such as those listed in chapter 3.1.

3.5.2. Task 2. Paired Time Series for Fish Amount and Price

Visualization of the paired time series plays an important role in the marine fishery industry, such as the work related to the parrotfish population by Valle and Oxenford [19]. In this paper, there are scatterplots showing relationships between human population size and fish density for selected fish groups across the Caribbean (Figure 3.5.2.1).

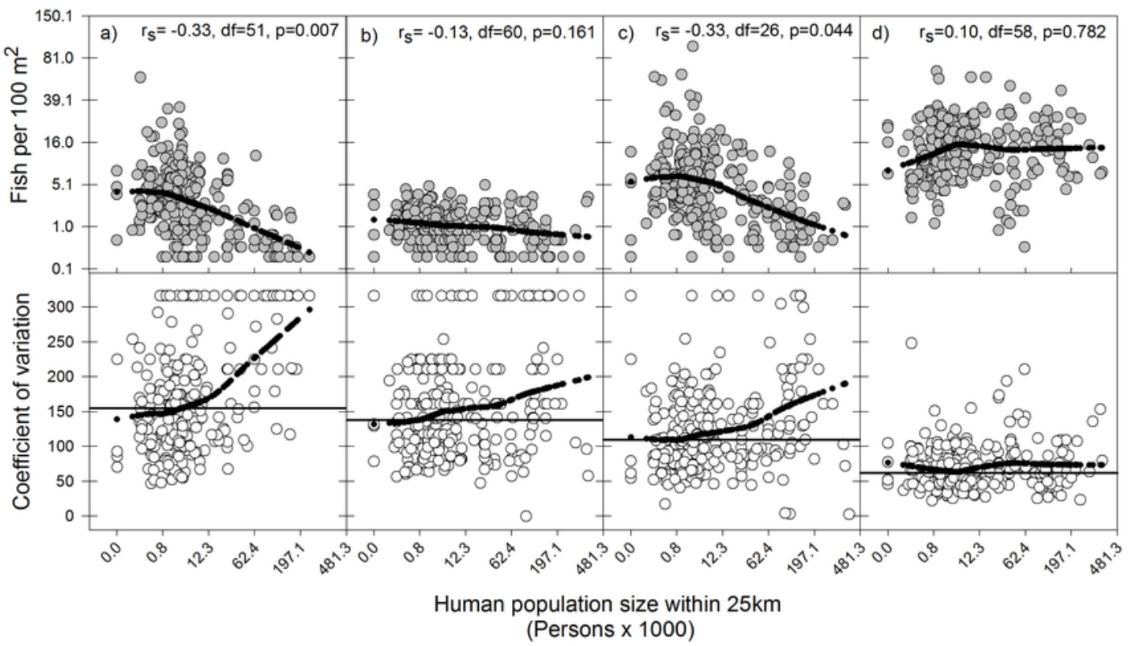


Figure 3.5.2.1. Scatter plot example; Source: [19]

A DFO report used a similar data visualization approach, but with a different time series (fish amount and quantities). One axis is fish quantity, second is fish price with dots representing years, which is the third dimension. The visualization shows the user the trend of how these values change over time together. Improvements of the visualization compared to the approach described in their paper [19] are to allow more advanced filtering, zooming and user interaction. Again, details of implementation will be discussed in chapter 4.

3.5.3. Task 3. Identifying Top Fish Species by Catch Amount or Price

Knowing which types of fish give the biggest income is valuable for fishermen and management. In British reports [13] there are pie charts that briefly describe the data for a specific year. In Figure 3.1.4. on page 12, we saw that the marine industry made use of pie charts to show summarized data of “tonnage” and “value” measurements. These visuals show percentages for each fish type price and the amount which allows the user (data analyst) to figure out the ratio between these parameters. Then, management can take into consideration the results of the report and use it in a way that can help the company to make more profit (for example, put more resources into catching expensive fish). The pie charts presented in the next chapter (Figure 4.3.3.) are more advanced, allowing visualization to help to find outliers for the selected year and also adds some interactivity for the user (hide legends, group fish types together, etc.).

3.5.4. Task 4. Consequent Years Fishery Data Comparison

Similarly, to the previous section 3.5.3, there is a visualization for two consecutive years of DFO data. The main point for this kind of visualization is to show fishery management and policymakers trends for the quantities and fish price.

Policymakers and environment workers or biologists are mostly interested in seeing fish quantities trend to determine if the decisions made in the previous year lead to the desired results in the current year (such as establishing fishing quotas, studying how pollution affects the fish population). However, fishery management is more concerned with looking into the fish price to decide which kind of fish is most appropriate to catch for the following year.

The chart presented in the following section 4.3.4. also helps environmentalists and fishery management to communicate with each other in a way that based on these values fishery management requests involving which type and how much fish are they planning to catch. Then policymakers based on that visualization, additional data and business knowledge approve or deny a request.

Chapter 4. Implementation and Evaluation

Recall that the goal of the current project is to create a visualization tool to help people understand/analyze table data in a more suitable format as charts, comparison diagrams. In this chapter, there will be discussions related to the implementation of the FishPlots, its structure, architecture levels and explanations of visualization methods for each use case (task).

The main reason for choosing a web solution for implementing visualizations is that it will be accessible for users without needing the installation of any additional software. Having everything in a web browser is a modern way of viewing and sharing visualizations. The user does not have to do any extra manipulations with a computer system to access visualizations.

FishPlots is implemented in TypeScript [20] (wrapper for JavaScript) from Microsoft. The front-end framework is Angular [21], one of the most powerful and highly used web engines. The back end is not needed for now as it is a test project. If the data source is changed, there will be minimum code modifications to get/process data.

4.1. FishPlots Overview

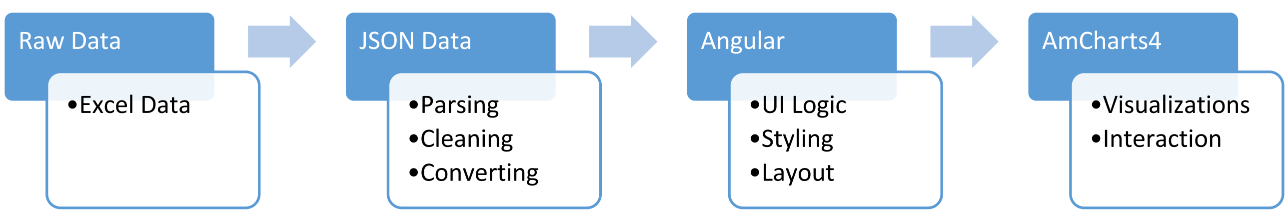


Figure 4.1.1. System diagram

1) Data layer

As described in the previous chapters, the data layer for FishPlots uses public data provided by DFO Canada. Data is converted from Excel to JSON format, which is suitable for use inside a web browser environment.

The web browser gets data by using HTTP REST request, therefore if the dataset is changed, it can point to any URL that can provide a JSON file with data in a suitable format for FishPlots to use. So, for future use, any web API can be used as a data source, which is very convenient if several people work with fishery data (one user uploads the file with data, the second user analyzes it with the visualization tool).

2) Logic Engine

This part of FishPlots is doing data-transformation from JSON files to a format which is suitable for the library to render and display visuals. It also reacts on user actions and emits events to charts to refresh data if needed and handles switching between charts.

As a framework engine, FishPlots uses TypeScript [20, 21] framework Angular [21]. It is a popular JavaScript-based web framework. For presenting data there is an amCharts4 [22] library included in the project, which allows the generation of complex interactive data visualizations using JavaScript inside a web browser.

3) User interface overview

On the top of the page, there are controls (Figure 4.1.2.) that allow users to work with data which includes filtering, selecting, etc.

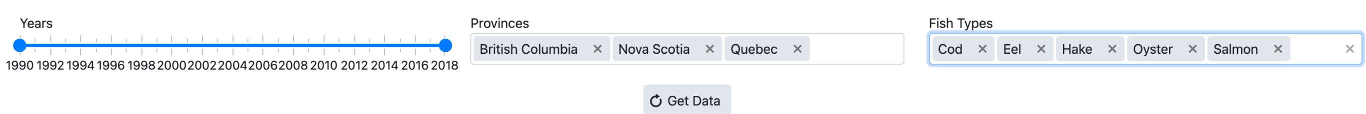


Figure 4.1.2. Filter interface

The date range slider (from 1990 to 2018) allows users to select particular years of interest that they would like to study and visualize. This type of control allows users to select the start and end year of the time range without a need to select each year individually, which possibly could be 28 actions (28 clicks for each year presented in the dataset) for the user to do. Instead, it is minimized to 2 (just selecting the first and the last year in the dataset).

There are two multiple selection pickers for provinces and fish types which are also filtering data. With these controls in place, the user can generate visualizations that are not restricted to a particular province or fish type. This feature makes the FishPlots more flexible than reports which are discussed in chapters 1 and 2. Based on the user input, there may be 4 different visualizations generated.

4.2. Data Processing

The data source comes from DFO (Fisheries and Oceans Canada). It reflects fishing amount catches and money profit for years from 1990 until 2018.

1) Data extraction

At the stage of the data extraction phase, Excel files are downloading from the official DFO website (dfo-mpo.gc.ca/). It does not require any additional permissions. This data is publicly available. However, the process of downloading is manual. It can be improved if DFO could provide API up-to-date data for use.

2) Data cleaning

Excel files contain some information not relevant for visualization. For example, these are table borders, text colour, other metadata, etc. So, irrelevant information is removed at this step.

Also, some fish species and provinces have mistypes in their names, mapping and some manual evaluation is performed at this stage.

3) Data converting

The last step of data processing is converting. As an intermediate step, the Excel file converts to CSV. Then CSV converts into JSON files, which can be easily interpreted by the visualization library in any modern browser.

Having all these processing steps as an automated process would be a good improvement for FishPlots, but DFO data is not that structured (especially for the 1990s), so manual evaluation is required.

4.3. Discussion of Use Cases for FishPlots

In this subchapter, there will be given the visualization solutions for each of the 4 problems and a discussion on how it improved current visualizations (fishery reports and papers discussed in the previous sections 2 and 3). Each use case is focussed on a study goal that is representative of a generic analysis step within a possible larger study. We refer to them more abstractly as tasks to aim our discussion on the generalizable and reusable aspects of each case.

4.3.1. Task 1. Exploring Relationships for Fish Amount and Price

This visualization will help users to figure out relationships between fish quantities and prices for the particular provinces and/or fish types. This visualization is a multiline chart with time (years) on the horizontal and price and quantity on the vertical axis. For each province there will be color-coding defined so correlation between these two values will be easily visible. Dropdown with selected fish type is filtering summary values.

One of the aspects of this data visualization is the ability to present the information in a way that would allow to include multiple values for data analysis. At the same time, a successful visualization should be easy to follow and use for further deep data investigation. The chart below proposes a way to explore a yearly trend for the fish amount and its value (price) by province. The chart is adaptive and could include multiple provinces with both price and amount values. This allows comparing how the fish industry trend was changing through the years in various locations. The adaptivity of the chart (Figure 4.3.1.) makes the data processing efficient by providing a way to work with data from multiple regions and years simultaneously.

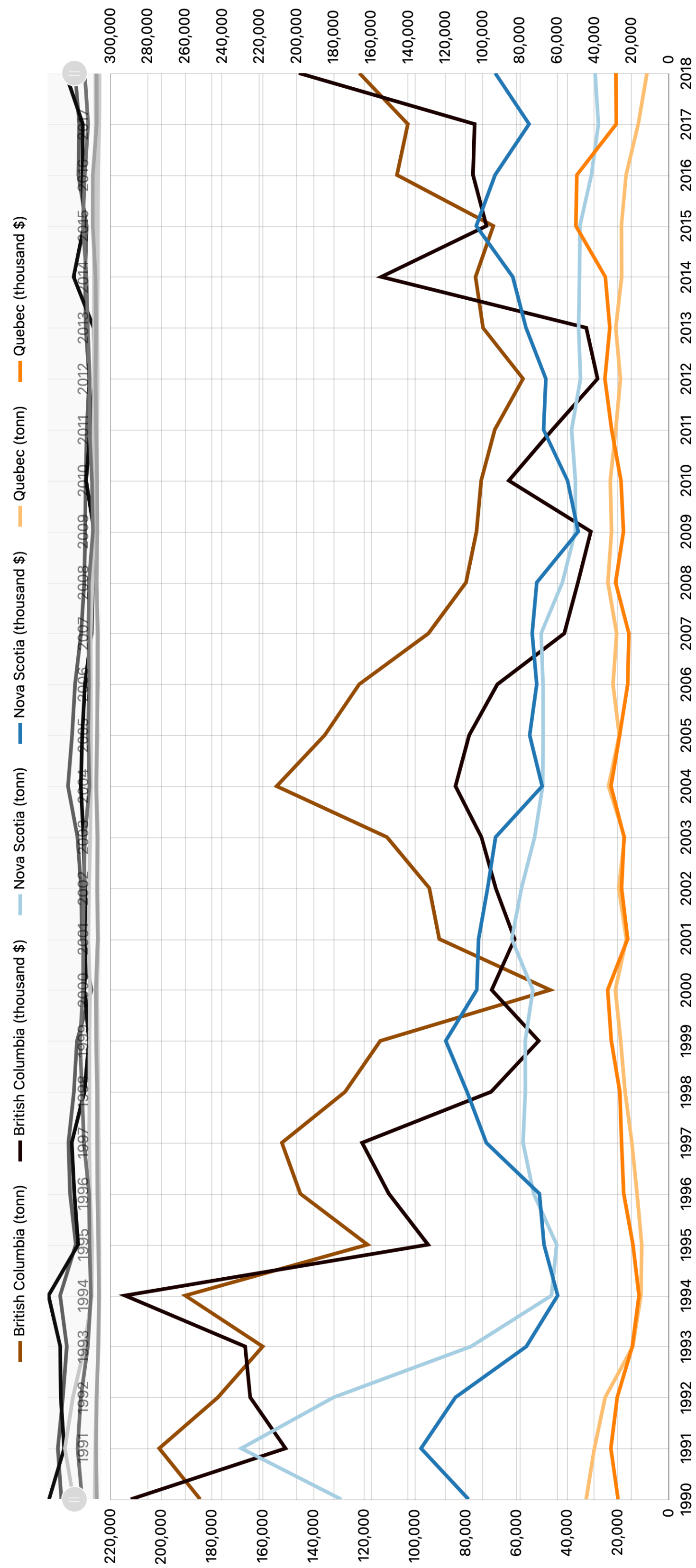


Figure 4.3.1. Long-term trends of fishing volumes across provinces of Canada

4.3.2. Task 2. Paired Time Series for Fish Amount and Price

The second chart is a scatter plot on which the x-axis is price and the y-axis shows quantities of particular fish types and provinces (picture is below). Bullet points and labels represent years. This visualization helps users to clearly see the correlation between price and quantities throughout the selected years. The proposed graph presents the ability to analyze and compare data efficiently by visualizing multiple values simultaneously. The example (Figure 4.3.2) below showcases the yearly trend of the quantity of shrimp per province compared to its value on the market. The axes show the quantity and value range, and the points represent the year per which the statistic was applicable.

The chart allows us to quickly understand the trend over the years and build analysis on the selected fishery market. By using previously discussed filters, the chart could be expanded to show more than one type of fish. The trends presented in Figure 4.3.2., show how the market was affected through the flow of time. For example, it is easy to notice the relatively steady and proportional growth of the quantity of shrimp in the market and its value from 1990 to 1994. However, the year 1995 has been significantly different in the rise of value compared to almost the same amount of shrimp being available at the market. That could be explained by some other events such as inflation of the currency or the increase of some other type of fish in the province.

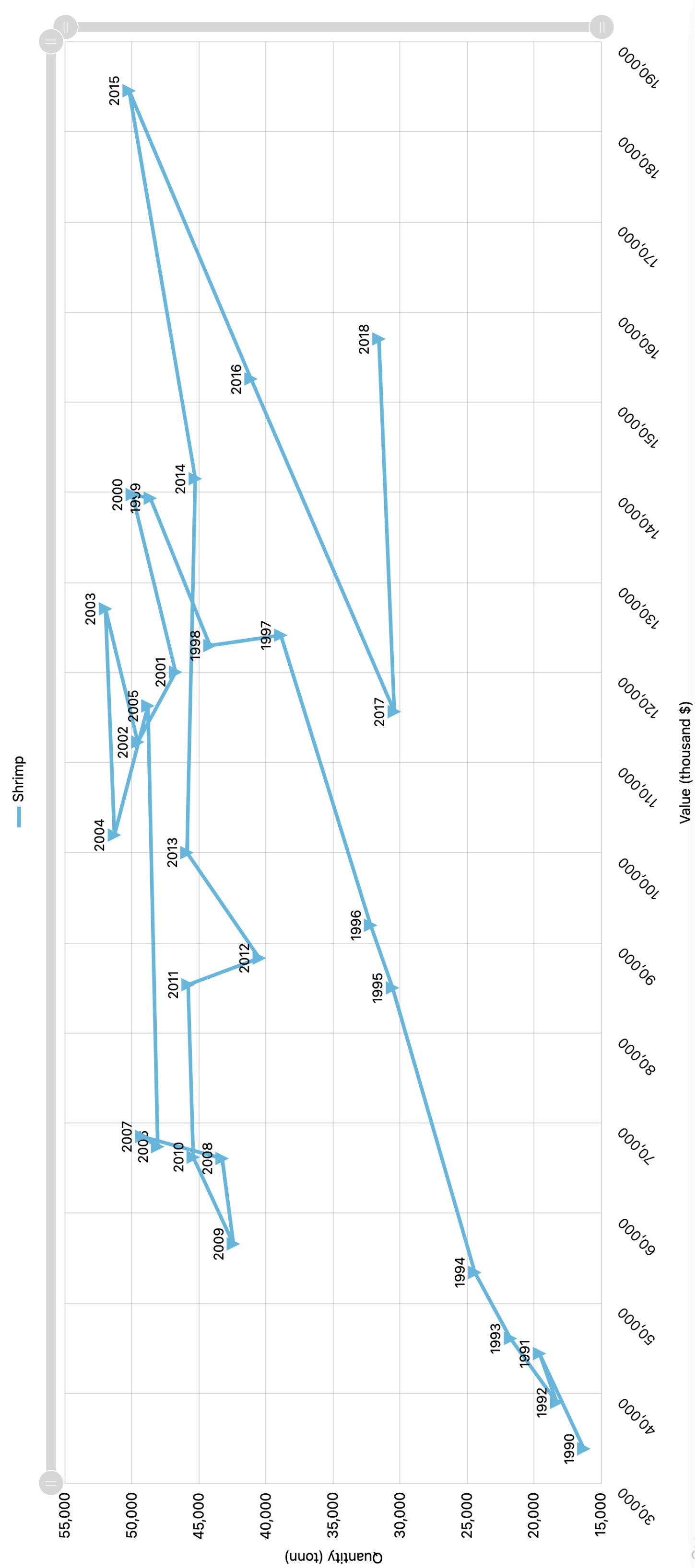


Figure 4.3.2. Connected scatterplot showing Value vs Quantity over the years

4.3.3. Task 3. Identifying Top Fish Species by Catch Amount or Price

The summary charts below (Figure 4.3.3) allow identifying the top fish species by catch amount or value (price) per the selected year. This type of visualization enables a quick and clear understanding of the top fish species in the selected category. The summary information helps compare the data for different years. Therefore, it is important to present the data clearly without going into additional details.

There is also a grouping for fish types that have a small percentage of catch or value comparing to the others which is an improvement (for more than 20 legends, pie charts usually look overcrowded and not readable).

To make the chart user-friendly and structured, the number of legends is adapted depending on the top values per the selected year. FishPlots would sort and group the value ($) or catch quantities (tons) of fish types and assign the applicable percentage per type. The top fish types are separated into their own sections, and the rest is grouped into the category “other”. This solves a well-known problem of having too many unnecessary labels (legends). The chart also allows to expand the “Other” category and drill down to see more details for the fish types which are the “outliers” in the dataset.

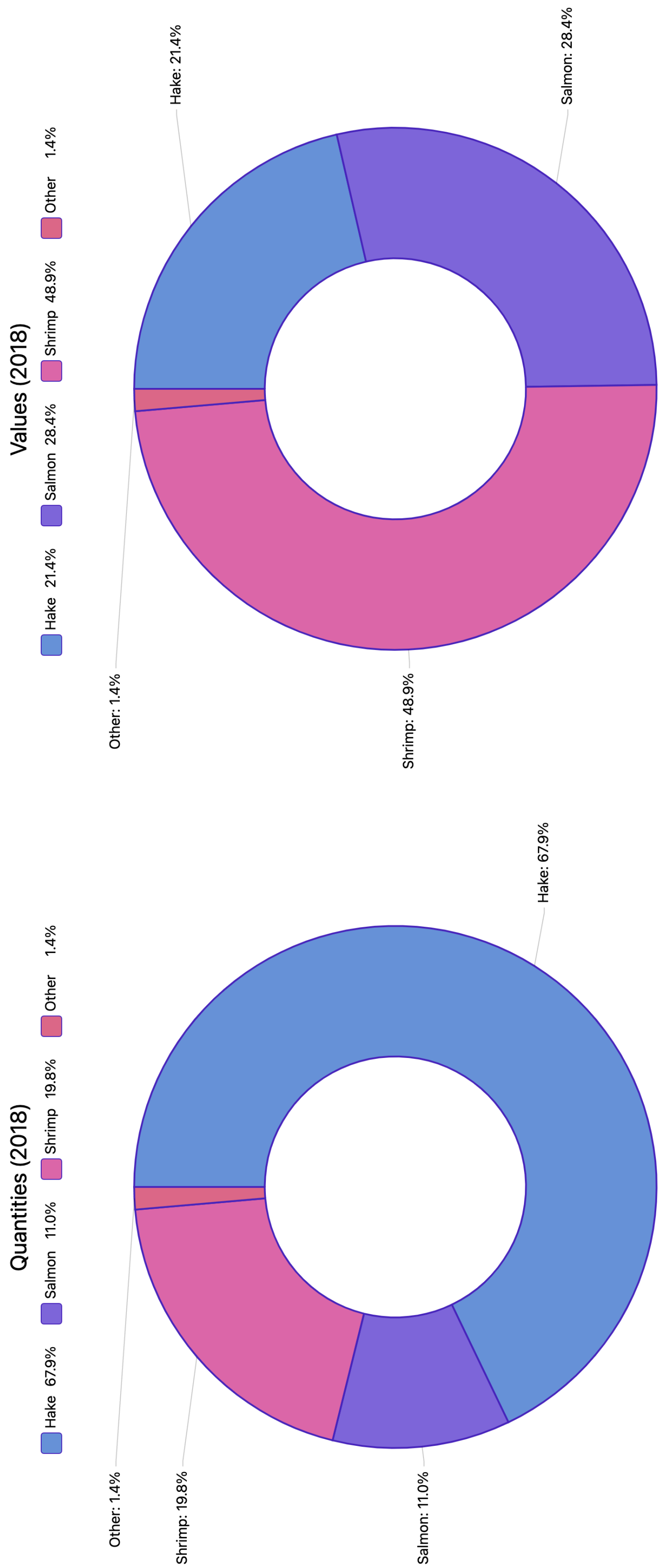


Figure 4.3.3. Summary pie chart showing Value and Quantity for the selected year

4.3.4. Task 4. Consequent Years Fishery Data Comparison

The main goal of the following visualization is to provide the ability to compare and analyze the data between the selected years quickly and easily. The Chart below (Figure 4.3.4) is an improvement of the standard bar chart called a categorized bar chart. This kind of data presentation is particularly useful for determining trends for the current and previous year. For fishery management it is important to know if, for example, new policies and laws or fishing quotas are working into the intended direction. Years 1990 and 2018 were chosen to better emphasize how the charts' setup assists in quickly identifying trends.

As mentioned before, the proposed chart is created to compare the data between the two years easy to analyze and review. The space on the x-axis is used to include both of the values that are important for the industry: quantity and price. The y-axis represents the type of fish that was added to the chart for the review. The chart itself serves as a great visual and data centre that could allow industry professionals to improve the decision-making processes or assist in the further planning of the fishery industry per type. In the provided example, we notice the drastic change in the Cod quantity and price. Both levels have dropped significantly through the decade, highlighting that Cod is no longer a highly available or valuable fish type at the market. Thus, the presented visualization type of data could play an essential role in providing a quick and easy-to-use chart to compare the data trends year by year.

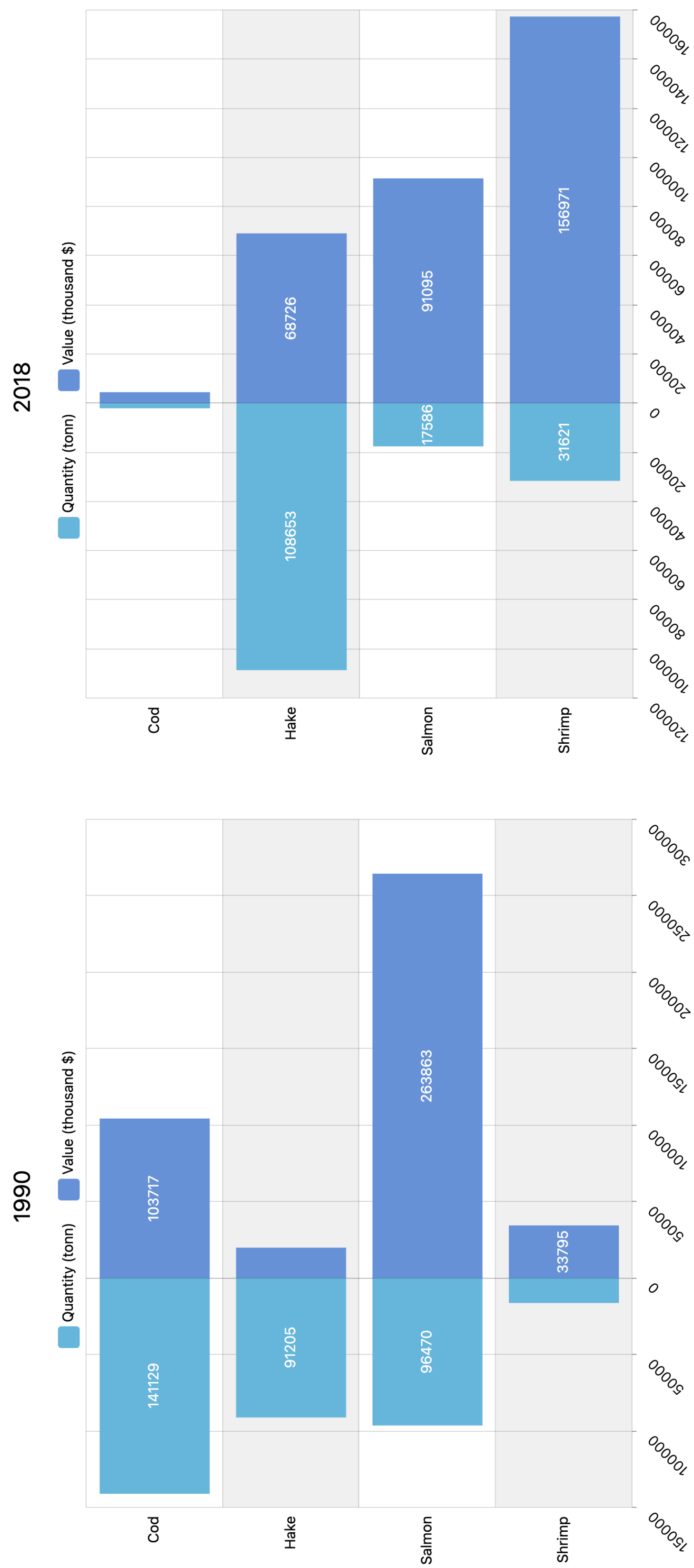


Figure 4.3.4. Bar chart comparing Value and Quantity for two selected years

Chapter 5. Conclusions and Future Work

The design and implementation of FishPlots was undertaken according to the marine industry's needs based on reports discussed in previous chapters. Specifically, this project reflects fishery management requirements for visuals in reports as presented in chapters 2 and 3. After reviewing existing visualizations for various fields, the report discusses the pros and cons of selecting one or another visualization method. Chapter 3 gave us an understanding on why visualizations for fisheries are important and how exactly people use existing tools and reports. As a result of our investigations, we implemented four different visualizations for fishery management based on DFO data sources. These charts were designed as improvements to the reports discussed in previous sections. Improvements to the developed visualizations will be summarized shortly. Given that visualizations are made suitable for web hosting, they can be uploaded to a remote server and users can access them using a browser without any additional software installed on their workstations.

Task 1 helps the user to explore relations between fish values and amounts. Improvements can be listed like this: the ability to select/deselect fish types/provinces, zoom feature which allows users to see data for smaller date range without a need to re-render visualization. Task 2 is a scatter plot, which combines 3 dimensions: value, amount, year. Data points are connected by lines which gives the user a historical overview of fishery data. Using this chart, the user can see trends of the value/amount ratio for different fish types. Pie charts in task 3 show data like they are used for report summary. The grouping feature allows users to reduce the number of legends which is useful specifically for pie charts (it’s hard to read visuals with more than 5-6 legends presented). And last but not the least, visualization for task 4 is allows comparison of summary data for any two selected years. These kinds of bar charts are used by fishery management to see trends between consecutive years. Showing different dimensions on the same axis allows reducing clutter from creating duplicate bar charts for comparing price and quantity separately and makes the report chart easier to read for the end-user. After implementing the visualizations for this project, there are still some improvements that could be applied to the FishPlots that can make it more reliable for the user.

Firstly, bringing more data sources to explore wider range of marine issues. With this approach, scientists or management people can combine or compare charts built from different sources. For example, if amount of by-catch is available, it can be useful for analyzing fishing methods and tools to reduce negative impact of the environment, because by-catch animals have small chance of surviving after they are released back to the ocean. This future improvement could provide broader and more unbiased knowledge of the problem the user is trying to investigate.

Secondly, saving visualization state (serialization) would be nice to have as an additional feature. In this case, users can share and discuss their visualizations without a need to re-apply settings. This improvement can be developed in several ways. For example, saving the state of particular visualization in JSON file on the user's machine. Next time when users want to use FishPlots, they can just locate the file, upload it into the tool and there will be visuals in a saved state displayed on a screen. Another approach is serializing the state of visualization into a string (URL link). Then it can be used by another user who has access to the link. FishPlots will parse that URL and automatically apply all necessary settings.

And finally, users will most likely want to have synchronization of new data from sources. This improvement requires communication with data providers, implementing external APIs from their side. It is out of the scope for this visualization project but can be implemented using the REST [23] approach.

Bibliography

[1] Yixuan Zhang, Kartik Chanana, Cody Dunne, "IDMVis: Temporal Event Sequence Visualization for Type 1 Diabetes Treatment Decision Support," *IEEE Transactions on Visualization and Computer Graphics,* vol. PP, 2018.

[2] Aaron Elliot, Cheng Hua Hsu, "Time Series Prediction: Predicting Stock Price," *arXiv: Machine Learning,* 2017.

[3] Polina Levontin, Paul Baranovski, Adrian W Leach, Alexandria Bailey, "On the Role of Visualization in Fisheries Management," *Marine Policy,* vol. 78, 2017.

[4] Z. Kemp and G. Meaden, "Visualization for Fisheries Management from a Spatiotemporal Perspective," *ICES Journal of Marine Science,* vol. 59, no. 1, pp. 190-202, 2002.

[5] Hiroshi Ashida, "Spatial and Temporal Differences in the Reproductive Traits of Skipjack Tuna," *Fisheries Research,* vol. 221, 2020.

[6] Marlene A. Bellman, Scott A. Heppell, Chris Goldfinger, "Evaluation of a US West Coast Groundfish Habitat Conservation Regulation via Analysis of Spatial and Temporal Patterns of Trawl Fishing Effort," *Canadian Journal of Fisheries and Aquatic Sciences,* vol. 62, pp. 2886-2900, 2005.

[7] Sharon Benjamin, Min-Yang Lee, Geret DePiper, "Visualizing Fishing Data as Rasters," *Northeast Fisheries Science Center Reference Document,* 2018.

[8] Patricia Pinto da Silva, Charles Fulcher, "Using GIS to Visualize Land/Sea Connections: Case Study - NE Herring Fishery," *Mar. Resour. Rev.,* vol. 67, pp. 19-25, 2007.

[9] M. Barrus, "The State of the Salmon," *University of British Columbia Library,* 2017.

[10] S. Semikina, "Stress Data Visualization," *Eindhoven University of Technology Library,* 2014.

[11] Fisheries and Oceans Canada, "Report on the Status of Groundfish Stocks in the Canadian Northwest Atlantic," Atlantic Stock Assessment Secretariat, Science Branch, Department of Fisheries and Oceans, 1993.

[12] Fisheries and Oceans Canada, "Northern Abalone," DFO Science Stock Status Report , 2004.

[13] Scottish Government Riaghaltas na h-Alba, "Scottish Sea Fisheries Statistics," Cabinet Secretary for Rural Affairs and Islands, 2019.

[14] Commission on Geosciences, Environment, and Resources, Improving the Collection, Management, and Use of Marine Fisheries Data, Washington D.C.: National Academy Press, 2000.

[15] Ministry of Natural Resources, "Methods for Monitoring Fish Populations," 2014. [Online]. Available: https://www.ontario.ca/page/methods-monitoring-fish-populations. [Accessed 29 October 2021].

[16] Bedford Institute of Oceanography, "Fishery Report," Fisheries and Oceans Canada, 2008.

[17] Federal Register, "List of Fisheries," 2018. [Online]. Available: https://www.federalregister.gov/documents/2018/02/07/2018-02442/list-of-fisheries-for-2018. [Accessed 29 October 2021].

[18] "PyPlot," [Online]. Available: https://matplotlib.org/. [Accessed 29 October 2021].

[19] Henri Valle, Hazel A. Oxenford, "Parrotfish Size: A Simple yet Useful Alternative Indicator of Fishing Effects on Caribbean Reefs?," *PloS one,* vol. 9, 2014.

[20] "TypeScript," [Online]. Available: https://www.typescriptlang.org/. [Accessed 29 October 2021].

[21] "Angular JavaScript Framework Library," [Online]. Available: https://angular.io/. [Accessed 29 October 2021].

[22] "amCharts4," [Online]. Available: https://www.amcharts.com/. [Accessed 29 October 2021].

[23] "REST," [Online]. Available: https://en.wikipedia.org/wiki/Representational_state_transfer. [Accessed 29 October 2021].