Slide 1

Good morning committee members!

Thank you very much for being here and taking part in my thesis project work and another step for graduation.

My name is Volodymyr Kozyr but you can call me Vova.

My topic if called “A study on data visualization for fishery management”

Slide 2

I will start with agenda for my presentation.

First, I will go through introduction and goals of the project, why and for whom it is implemented.

Then we will review several papers/projects done in this area so we will understand what is already implemented, which pros and cons these visualizations have and which improvements could be done.

After that I will explain you tasks for the project.

Then we will quickly overview architecture of the implemented system.

And finally, more particular examples of use cases implementation and conclusions.

Slide 3

Topic -> Fishery data, like data from other domains, deals with observations across time, and across space. Depending on decisions made after analysis it can lead to different economic and environmental consequences. Also, there are many data sources available for fisheries, which I will touch briefly further in my presentation.

Objective -> 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.

Slide 4

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 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 sometimes complicated for readers who are not data analysts or experienced computer users.

Slide 5

FishCAM2000 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.

One important advantage of the system is that it presents complex geodata, which includes the amount of fish caught in the geographical zone on the map instead of the data table, which is a considerable timesaving for users to understand the query output.

Figure on slide 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.

Slide 6

Scottish Sea Fisheries Statistics

The total amount of fish caught is shown in blue, and the value of all landings by Scottish vessels is shown in red. By comparing the blue and red lines, the user can easily and quickly determine the connection 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.

Slide 7

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

Slide 8

In the next report 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.

Slide 9

Based on literature review for the project we can stress some fishery domain problems

* Determining the optimal amount of catch for each type of fish to reduce environmental damage in a specific region
* Predicting which species may also be subject to negative or positive effects (trends)
* Establishing quotas (the proper amount of unloading of fish) which will minimize the negative impact on the environment
* Analyzing the safety of methods for catching a particular type of fish in each region
* And other…

Slide 10

Analysis of the data presented in a table or text format may take significant amount of time, as was discussed in IDMVis. 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 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 browsers. 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.

Slide 11

* **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

* **Requirement 3. Summary and Overall Statistics**

Showing global summarized data.

Slide 12

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 doesn’t have to do any extra manipulations with a computer system to access visualizations.

FishPlots is implemented in TypeScript (wrapper for JavaScript) from Microsoft. The front-end framework is Angular, 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.

Slide 13 - 14

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.

Logic 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 framework Angular. It is a popular JavaScript-based web framework. For presenting data there is an amCharts4 library included in the project, which allows generation of complex interactive  data visualizations using JavaScript inside a web browser.

Slide 15 - 16

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 is color-coding defined so correlation between these two values will be easily visible.

The chart 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 makes the data processing efficient by providing a way to work with data from multiple regions and years simultaneously.

Slide 17 - 18

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. 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 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. The trends presented in 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.

Slide 19

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.

Slide 20

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. 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 approach.

Slide 21 - 22

The summary charts 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.

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.

Slide 23 - 24

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 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 the 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.