Information Visualization

Assignment 1: Visualization Design

Group 02

Students:

- Boulouma Mohamed Amine, 11931226
- Mengesha Isaak, 11945608
- Roman Andreea, 12019805

Table of Contents

Table of Figures

Introduction

In this assignment, our task is to design a concept for an interactive visualization and provide explanatory text describing our design using the Water Quality of San Francisco Bay Data Set. In a first phase we characterize the data as well as our user, the medium-sized NGO with their tasks and goals. Based on this we design a concept for an interactive visualization that communicates the data well according to our users and tasks. We set out the interaction methods to query, explore, and analyze the data visually. While we use the dataset given, we have filtered and transformed the data as we see fit to highlight the elements that we think are most important in the data set.

User and task analysis

The user of our visualization is a medium-sized NGO, focussing on environmental advocacy. They are trying to raise awareness for the disruptions of local ecosystems by human-caused pollution. While their goal is to inform, they want to convey their message as concise and clear as possible. This project will be part of a broader campaign, informing the public through different mediums including presentations, newspapers, posters and websites.

According to this NGO, the local ecosystems are at the brink of collapse. This has a multiplicity of reasons, global and local alike. Pollution has been a cause for stress to these systems for a couple of decades. Global warming might deliver the final punch to strained food chains, potentially causing extinctions of native species. Basic context information provided by the client:

- Maritime life is very sensitive to the concentration of nitrate, phosphate and ammonium and water temperatures.
- Human activities play a large role in increasing the concentration of these chemicals.
- Main contributions stem from excessive use of fertilizers in agriculture and insufficient purification of sewage water.

Our assignment is to provide data visualizations for a broader audience as well as for a presentation in front of the California State Senate(CSS). Both the Senate and the broad public are not particularly interested in technical details. However, the presentation will not only be heard by politicians but also their advisors and other experts, meaning that some technical expertise is needed to convince them of the problem at hand, too.

We distinguish between visualizations targeting the general public via the NGO's website and those that are used during the CSS hearing. Both graphics can assume the user to read and understand simple mathematical concepts and relations (e.g. size ratios, growths, dimensions...), which is sufficient to transport the necessary message. However, for the Presentation small additions are needed such that key scientific measures and concepts (e.g. standard deviation, correlation, variance…) are communicated to relevant experts as well.

Visualizations of	User	information to transport	Application domain
(1) Unstable biotopes	Experts & Politicians	stability mapped over space and time	PowerPoint/ Matlab-plots
(2) Temperature-development	general public	temperature over space and time	slide-show
(3) Algae blooms	general public & Politicians	frequency/ events over space and time	animation

Table 1.: Summarizing description of users application domains and topics to communicate.

Data analysis

The Dataset contains many observables that can be omitted to fulfil the task given by the NGO. Consequently, parameters that need to be extracted from the dataset in order to inform about key relations are:

parameters	data types	data structure	specifics
Station number	Nominal	single dimensional/ two dimensional	Coordinates of corresponding stations
Timestamps	Continuous time	single dimensional	
Calculated Oxygen	Ordinal	single dimensional, time-oriented	
Temperature	Ordinal	single dimensional, time-oriented	
Concentration of Nitrate and Phosphate	Ordinal	two dimensional, time-oriented	combine to livability
Depth	Ordinal	single dimensional spacial	

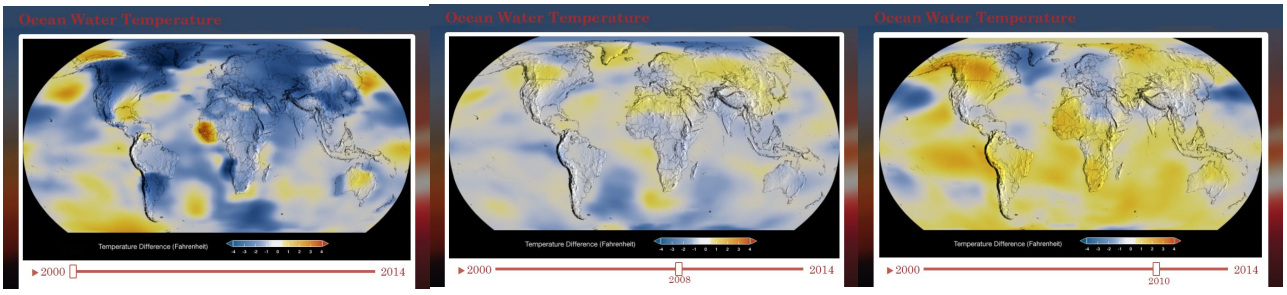
Table 2.: Listing relevant parameters and their properties

All other parameters were omitted. Distance and many other variables are not relevant to help the NGO communicate their message. Simplifications were only applied to the timestamp parameter in order to condense the data. However, this happens differently for different visualizations. For example showing the increased instability of biotopes requires only a yearly resolution, whereas tracing the causes of algae blooms requires almost a resolution on the daily level. For a more detailed description review the respecting chapters.

Conceptual Design

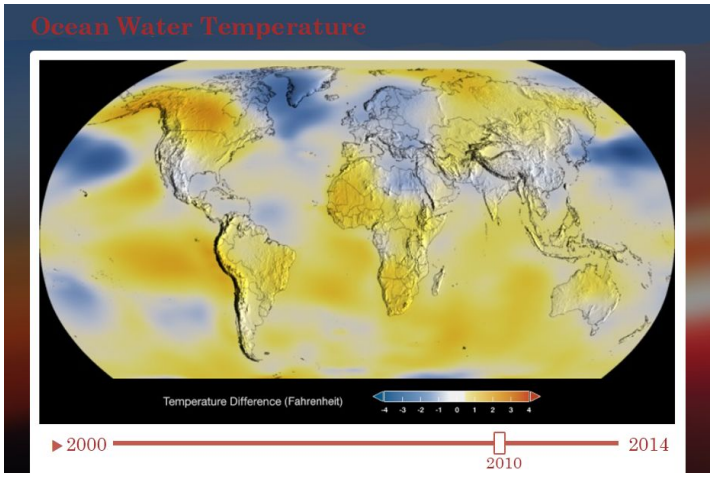
(1) Temperature Development of the Bay

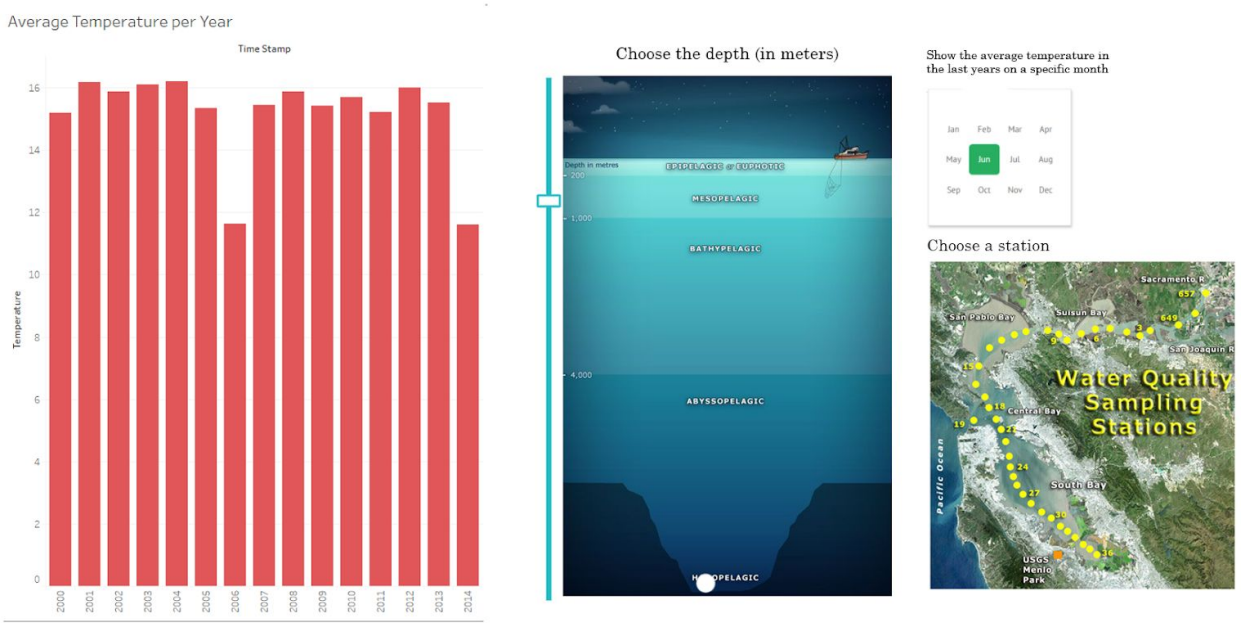
As we address our information to a broader audience, we decide to have accessible and intuitive visualizations so everyone can easily interact, query and understand the information the NGO is providing.



Picture 1.: Representative picture of an interactive map that depicts temperature change over time. (Note: The San Francisco Bay Map would be used instead of the World Map)

In order to show the historic impact of climate change on the water temperatures of the Bay we decided to allow users to skim through all years. The heatmap laid over the world will not represent total temperature as a measurement but rather the change compared to the previous years average. Here blue means a local decrease and yellow/orange a local increase of temperatures compared to previous year. This will illustrate the gradual increase over time as pictures become more and more red over time. It provides even people who find it difficult to read complex plots a concise overview on the climate developments. The slider provides the option to interact and decide at how fast changes should be shown. The Bay will be divided in regions corresponding to a certain station and then colored accordingly.





Picture2: Average Temperature per Year with Parameters: Depth, Time, Location

As we want our users to have access to more specific and detailed information, we visualized the average temperature per year through 3 parameters: Depth, Location (Station Number) and Time (Month). Even though it looks like a very complex querying process, the design is made in a way the user experience can be an easy and entertaining journey.

On the left side a bar chart displaying the average temperature per year is shown and is changing according to the values of the parameters:

- The depth can be adjusted very easily by the interactive bar. The gradual change of the color blue in the image used as reference for the levels of the Ocean Water depth indicates the lower temperatures when increasing the depth. For the more experimentet users (researchers or different members of the senate) visualising the temperature in direct corelation with depth is a very important step, as temperature of water varies directly by its depth. We expect warmer values closer to the surface (due to the direct contact with the sunlight but also because it is more affected by the pollution).

- The month can be selected through a click in the displayed calendar. This can offer a better overview on Winter/Summer differences of temperatures in different years.

- The station number can be chosen using the interactive map of the San Francisco Bay if the user is more interested in a particular place or wants to observe difference between locations.

(2) Unstable Biotopes

Dead Zones are areas in the ocean in which the concentration of oxygen drops below a certain threshold value. Of course this is an event that happens throughout the time of measurements at different locations and is a rather probabilistic process. Nevertheless, there is clear evidence for a human contribution through concentration levels of several chemicals. Since this is a probabilistic causality it will be not trivial to condense this relationship in a single visualization.

We will look at the number of dead zones in the entire bay area throughout a year and across all stations. This means each tiny distribution function is plotting the number of stations (y-achsis) reporting on the number of dead-zone-measurements (throughout one year). Though the y-axis of the large plot will be the time-period. The x axis can be used to plot both, the distributions for dead-zones as well as the concentration of pollution. This underlines the intuition for the causal effect of pollution of the biotopes and preserves accuracy to convince experts as well. Therefore this would be a specifically suitable visualization for the CSS hearing. Online users can hover over the diagramm with their cursor in order to highlight the respective year. This way, the user is able to distinguish measurements of each year more clearly. Using a complementary color pallet allows the main message to be seen at a single glance. Meanwhile detailed information is available to observe in closer inspection.

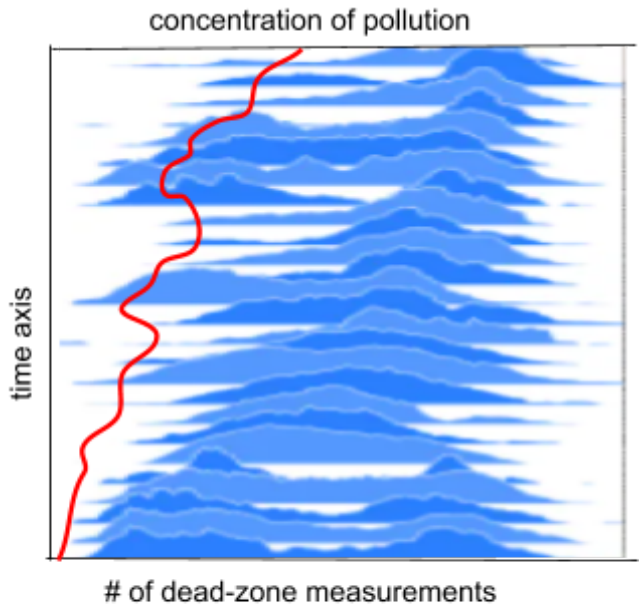
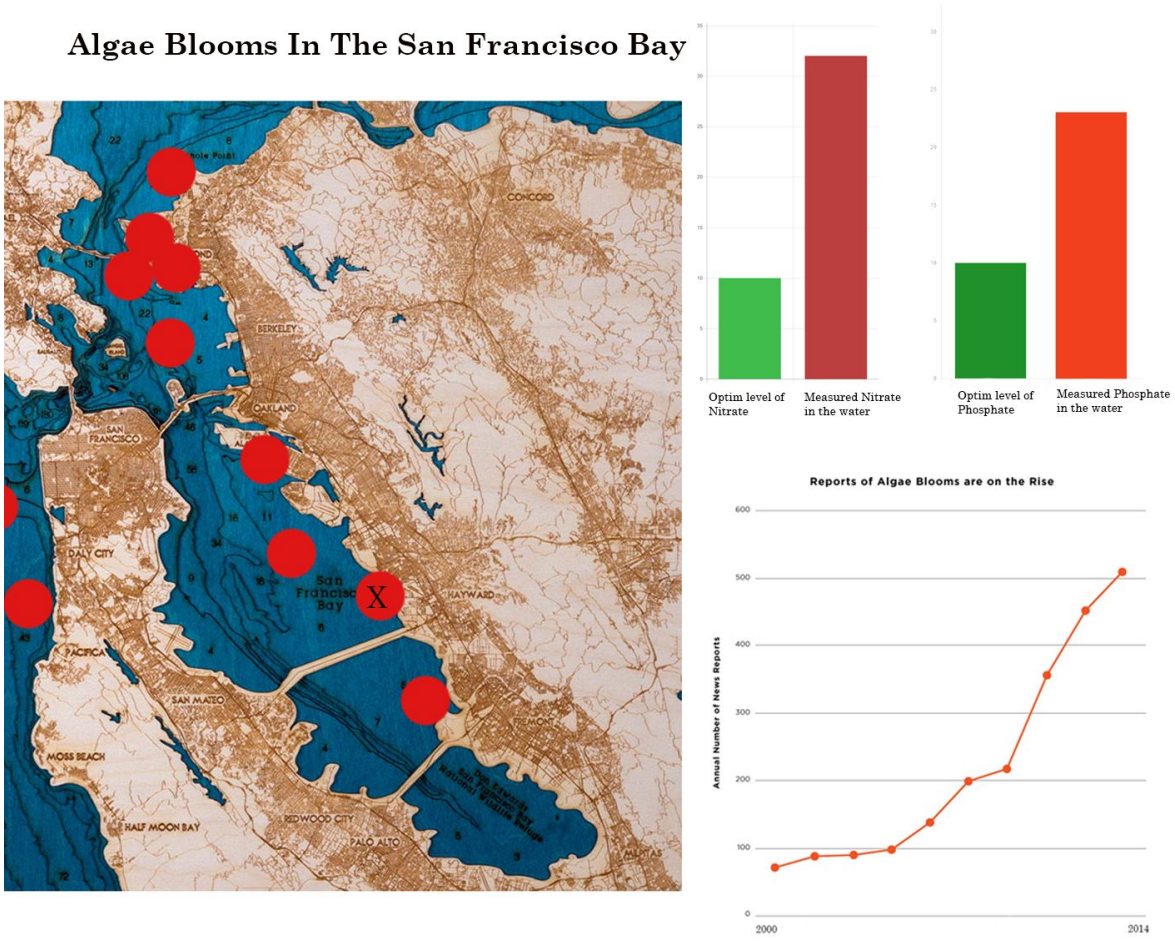


Figure 3.: Distribution plot conveying the effects of pollution to the number of dead-zones

(3) Algae blooms

Algae Blooms are a consequence of the increased concentration of chemical elements in the water, such as Nitrate and Phosphate. To better visualise this we created a Map of the San Francisco Bay where clusters of algae are showed (the red dots). To obtain this information we used: location and the measurements for Nitrate and Phosphate. Only the places where those two elements are above the optimal level are displayed.



Picture 4 - Algae Blooms In The San Francisco Bay

An overview for the last years of the new reports of Algae Blooms is also displayed by a graph. The raise of the reports of the new formation can be very easily seen, as it is time oriented.

The user can interact with the Map by clicking a red dot, and more information such as the level of Nitrate and level of Phosphate is shown, compared to the optimal level.

References

[1] Water Quality of San Francisco Bay Web Page. https://sfbay.wr.usgs.gov/access/wqdata/index.html]. Accessed on [4.11.2020].

[2] Description of Variable Names in the Data Tables of the San Francisco Bay Data Set. [https://sfbay.wr.usgs.gov/access/wqdata/archive/tabldescrip.html] Accessed on [4.11.2020].

[3] San Francisco Bay Station Tables. [https://sfbay.wr.usgs.gov/access/wqdata/overview/wherewhen/where.html] Accessed on [4.11.2020].

[4] The Glossary of Terms and Acronyms of the San Francisco Bay Data Set. [https://sfbay.wr.usgs.gov/access/wqdata/guide/glossary.html] Accessed on [4.11.2020].