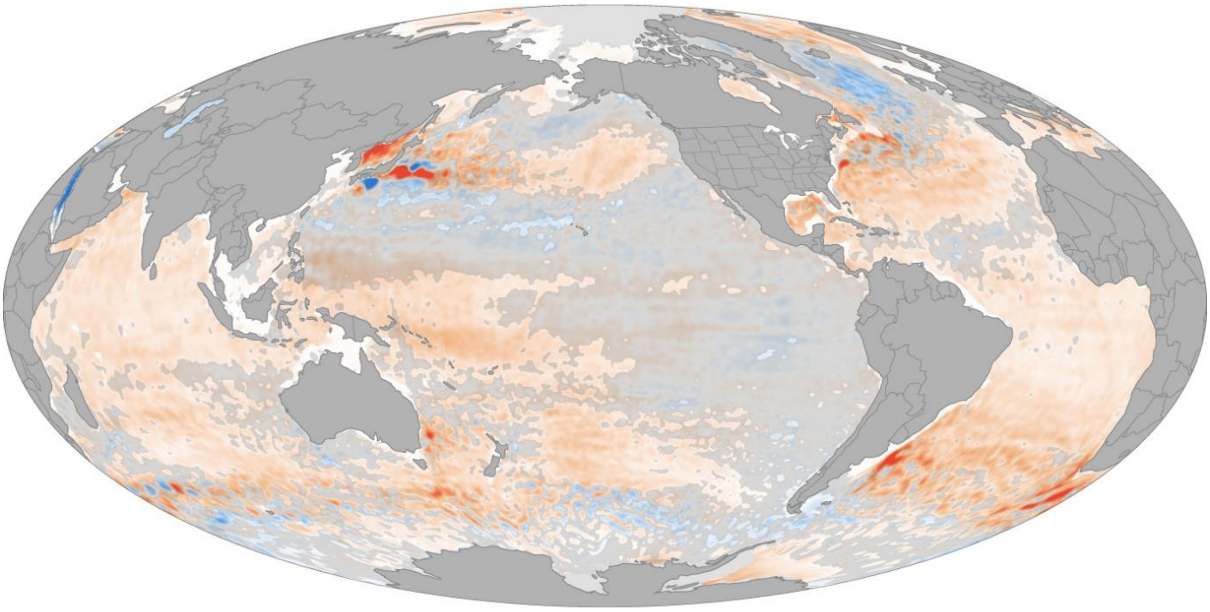
The visualization of dataset

Ocean Heat Anomaly

292



(Climate.gov, 2022)

Contents

Introduction

In searching for one of the 17 sustainable development goals (SDGs) our group decided to choose goal 14 (SDG-14) "Conserve and sustainably use the oceans, sea and marine resources for sustainable development" (The United Nations Department of Economic and Social Affairs, 2016). Therefore, our theme is concentrated on the consequences of global warming and its effects on our oceans. Our group thinks that this is one of, if not the biggest challenge for humanity to overcome. Our group will attempt to visualize and explain this dataset "Ocean Heat Anomaly" (Priester, 2022) which tackles this issue of ocean warming, and hopefully one could learn something from it.

Task 1. The utility value and audience

1. The dataset the group has chosen is "Ocean Heat Anomaly" (Priester, 2022) and it is relevant to our choice of theme because it affects every aspect of the global oceans. It affects every ecosystem, every country and everyone on planet Earth. It effects the ocean currents; weather patterns and contributes to the increase in extreme weather phenomena. This is why the first place one should look when discussing global warming and its effects, is in the increase in temperature in our oceans. The importance of the ocean cannot be understated; "The ocean is the largest solar energy collector on Earth. Not only does water cover more than 70 percent of our planet's surface, it can also absorb large amounts of heat without a large increase in temperature. This tremendous ability to store and release heat over long periods of

time gives the ocean a central role in stabilizing Earth's climate system." (Lindsey & Dahlman, 2020).

2. In terms of utility, the "Ocean Heat Anomaly" (Priester, 2022) dataset can be used in relation to many different things. It can be used to develop or support current theories on long-term environmental defects on ocean ecosystems and weather phenomenon. This has especially become more relevant in recent times because of extreme weather phenomena above the oceans and in coastal areas. The dataset can also be used to predict future ocean temperatures based on current models and supported by data reaching far back in time. The ones who might benefit from "Ocean Heat Anomaly" (Priester, 2022) dataset are environmental researchers, meteorologists, the fishing industry etc.
3. The target audience the knowledge from "Ocean Heat Anomaly" (Priester, 2022) dataset will be directed towards are the educated and adolescent population, as well as young people who are passionate about environmentalism. It is aimed at the ones who have some idea of what is being presented. The people who without a doubt, have some background knowledge on this subject already. Who most likely has seen news about it, read an article or heard about it on a podcast. This is already an important environmental issue in political and societal debates since it affects the whole world. Therefore, they will most likely understand the information presented and why it is important.
4. The information we want to visualize from the "Ocean Heat Anomaly" (Priester, 2022) dataset is the relationship between the "heat content anomaly" (Priester, 2022) and the development through time in relation to the average temperature. The object is described as the following: "heat content anomaly (10^22 Joules): ocean heat as to compared to the average between 1955 and 2006 measured in 10^22 Joules" (Priester, 2022).

Task 2. Dataset insights

Year of observation	Month of observation	heat content anomaly	Year of observation	Month of observation	heat content anomaly
1955	6	-0,112888	1989	6	2,184499
1956	6	-4,435912	1990	6	-1,216034
1957	6	-5,581200	1991	6	3,209013
1958	6	-1,357916	1992	6	1,039358
1959	6	-0,125907	1993	6	0,014944
1960	6	-1,150040	1994	6	-0,640411
1961	6	-2,191756	1995	6	2,082141
1962	6	-3,028738	1996	6	7,236955
1963	6	-1,974558	1997	6	3,687088
1964	6	-0,999268	1998	6	5,323250
1965	6	-3,651060	1999	6	5,638433
1966	6	-3,879088	2000	6	5,757782
1967	6	-3,258875	2001	6	2,730675
1968	6	-4,560177	2002	6	7,580072
1969	6	-4,898119	2003	6	9,989262
1970	6	-5,943489	2004	6	9,882789
1971	6	-3,910183	2005	6	8,463406
1972	6	-2,546130	2006	6	9,907466
1973	6	-3,276274	2007	6	8,811550
1974	6	-3,247772	2008	6	10,336976
1975	6	-2,137013	2009	6	9,101278
1976	6	-0,842684	2010	6	9,614333
1977	6	1,487195	2011	6	10,042295
1978	6	0,186941	2012	6	10,098619
1979	6	-1,640448	2013	6	12,048531
1980	6	1,813644	2014	6	13,165339
1981	6	0,533019	2015	6	14,841905
1982	6	-1,675871	2016	6	13,076438
1983	6	2,378227	2017	6	16,228298
1984	6	-2,455978	2018	6	15,545946
1985	6	1,808964	2019	6	17,515776
1986	6	1,703140	2020	6	17,286221
1987	6	-1,371921	2021	6	17,946842
1988	6	0,316734	2022	6	20,062943

(Original Kaggle dataset "Ocean Heat Anomaly" (Priester, 2022) in Excel)

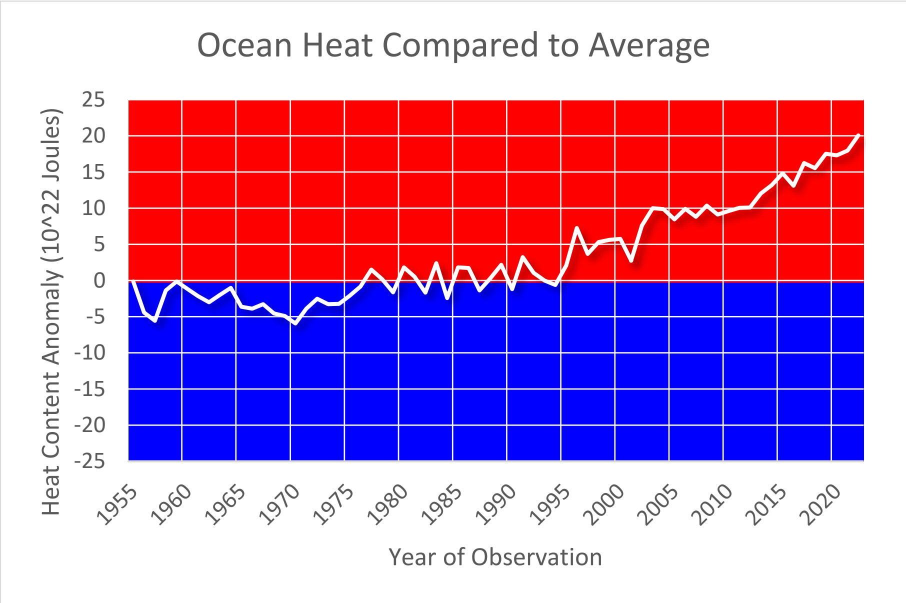
1. The "Ocean Heat Anomaly" (Priester, 2022) dataset contains one object and two attributes. The object of the dataset is "heat content anomaly" (Priester, 2022) because that is the focus in the climate report which the dataset is based on. The attributes within this object are "year of observation" (Priester, 2022) and "month of observation" (Priester, 2022) because they support the main point by showing when the measurement was taken.
2. The "Ocean Heat Anomaly" (Priester, 2022) dataset is made up of only quantitative data since all datapoints are numerical. The object "heat content anomaly" (Priester, 2022) is a ratio datatype because it has a defined zero point. The value either increases (positive) or decreases (negative) from this point. The attribute "year of observation" (Priester, 2022) is an interval datatype because it is measured along a timescale (years) and therefore holds no true zero. It starts from 1955 and increases by one increment to the year 2022. The attribute "month of observation" (Priester, 2022) is also an interval datatype since it is based on the months of the year,

however, there is no deviation by this datapoint in the “Ocean Heat Anomaly” (Priester, 2022) dataset since every measurement is taken in the same month (June).

Task 3. Visualization and analysis

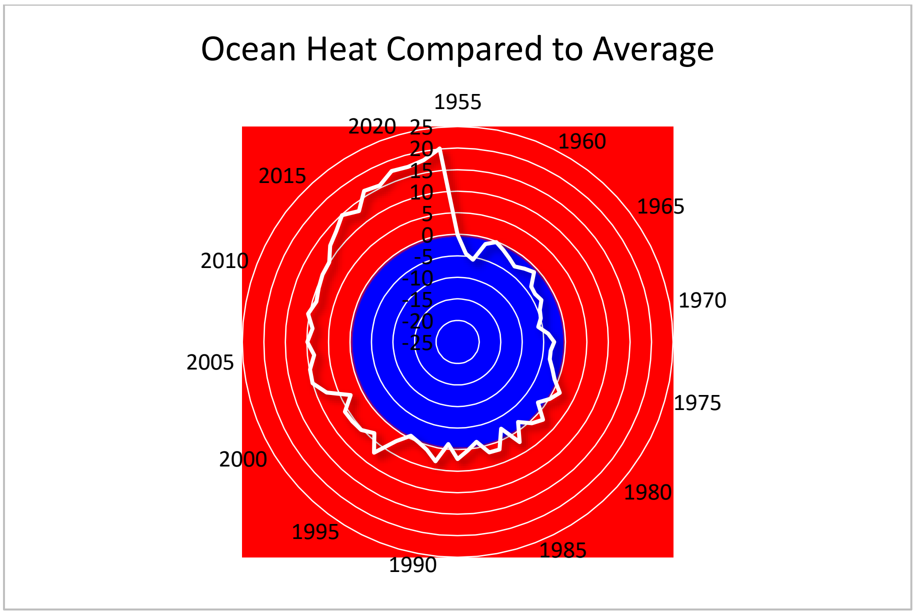
1.

a) Our group choose to use the functionality of Excel to create visualizations of the “Ocean Heat Anomaly” (Priester, 2022) dataset. This is because Excel is readily available, easy to learn, easy to use, and has extensive utility with making diagrams from raw data. When it comes to converting raw data into a visualization that conveys information effectively, our group tried many forms of diagrams and styles to achieve this. We experimented with line diagrams:



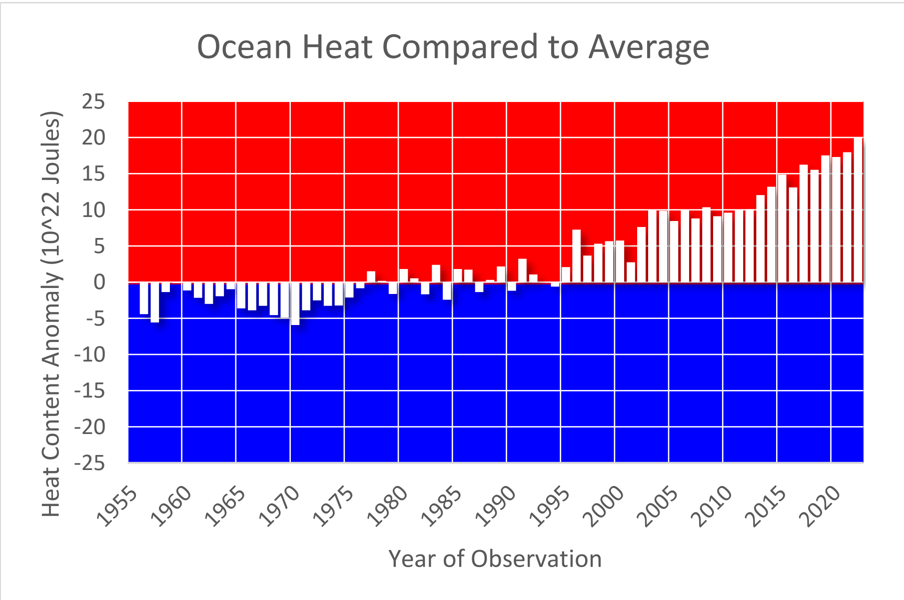
(Created by inserting “Ocean Heat Anomaly” (Priester, 2022) datapoints into line diagram in Excel)

The issue we had with the line diagram is that it does not convey how far the values deviates from zero, it pulls more of the focus towards the trend in time. Therefore, we concluded it does not properly convey the scale of the change in data. We also tried a radar diagram:



(Created by inserting "Ocean Heat Anomaly" (Priester, 2022) datapoints into radar diagram in Excel)

One issue we had with the radar diagram is that the endpoints automatically connect, therefore creating some confusion about the starting year (1955) and the ending year (2022). On the diagram it seems that the "heat content anomaly" (Priester, 2022) starts out at 20 and then immediately falls down to -5, but this is incorrect. Another issue with the radar diagram is that you get little sense of what is being presented. The only clue we get is from the title of the diagram, without that one may be clueless. We wanted to showcase the negative and positive values, as well as the trend over time, so the linear and radar diagram was not ideal in this purpose. We figured out that the column diagram offered the most potential for a good visualization:



(Created by inserting "Ocean Heat Anomaly" (Priester, 2022) datapoints into column diagram in Excel)

b) Our group imported all the values from the “Ocean Heat Anomaly” (Priester, 2022) dataset to Excel so that we could easily make visualizations of that data using Excels diagram functionality. The first thing we did was to remove the attribute “month of observation” (Priester, 2022) because all observations was made in the same month (June). This meant that this dimension would have the same number (6) 67 times over, not to mention it would introduce a third dimension to our diagram. We concluded that it was an unnecessary datapoint to convey since it would introduce unwanted complexity and noise in the diagram. Worst case scenario, it could contribute to a cognitive overload, which we wanted to avoid at all costs. We then decided to cut down on the decimal count in the “heat content anomaly” (Priester, 2022) datapoints from 6 to 2. This is because we want to reduce complexity wherever possible, and, in our final visualization one should not need to comprehend such a large decimal count. Anyway, it would not show in the result eighter way since they are too small (0,000001).

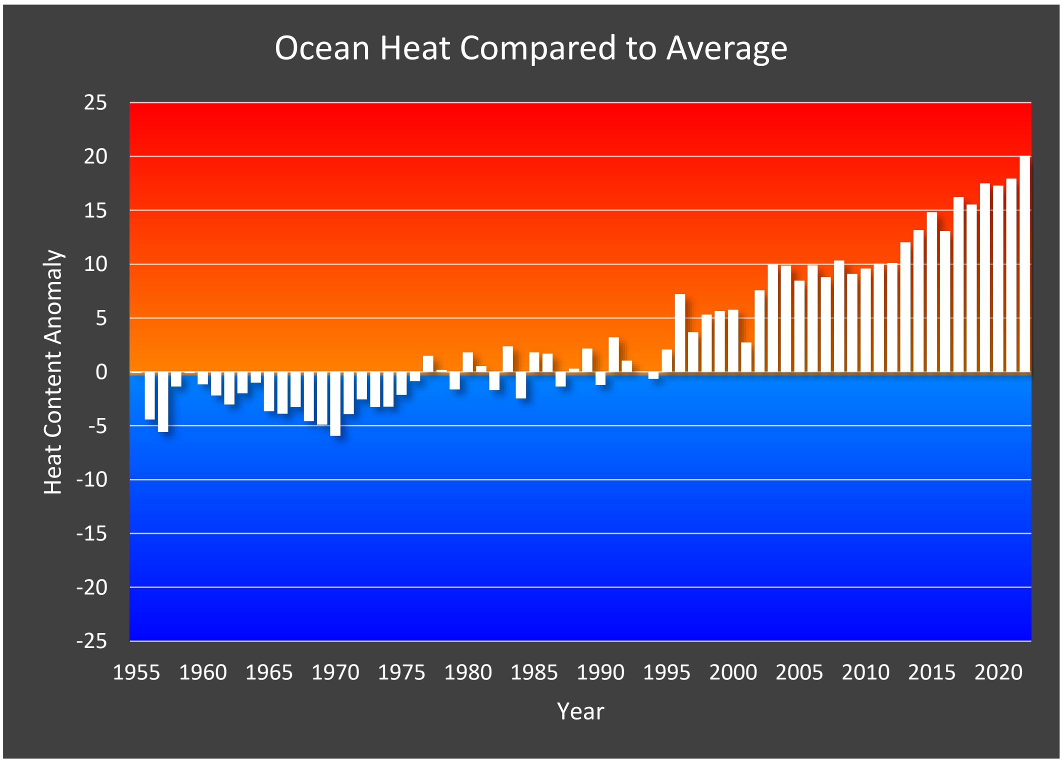
2.

a) Our visualizations of the “Ocean Heat Anomaly” (Priester, 2022) dataset were shown to 3 people of the target audience (Task 1).

b) User testing started directly with showcasing each of our visualizations (line, radar and column diagram) (Task 3). to all participants. Our group was instructed beforehand to not give any clues on what was being presented. We also made sure that we were in a neutral environment with no distractions, and that we had the full attention of the participant. Then we used our computers to showcase each visualization, and we asked questions alongside their first impressions. We asked what they thought the visualizations was about, and their thoughts on readability and esthetics. We also asked for feedback while discussing each visualization.

c) The user test participants agreed with our assessment of the problems mentioned earlier (Task 3, 1a) with our first visualizations. The designs were a bit too cluttered and indecisive as to properly represent the “Ocean Heat Anomaly” (Priester, 2022) dataset. Further feedback was given about the esthetics, as it was felt to be a bit plain and uninteresting. However, all of participants quickly understood the theme (Task 1) of the visualizations. They understood it had something to do with an increase in global temperature and with further studying they figured out that it was about the oceans. The y-axis really helped here since it showed positive numbers above the zero point and negative numbers below it, with a warm and cold background color. The line diagram did not convey the scale or the numbers very well as it was just a line going up and down. We knew originally that this was going to be a problem because it is hard to get a sense of value when points floats in the diagram. We tried showing the value as text, but it got to cluttered and unreadable, so we decided against it. Overall, the line visualization was not very successful. The radar diagram was the most confusing of all. It had no text, except for the title to indicate what was presented, and the trend of time was difficult to grasp. This is because of the start point and the end point connecting, creating some confusion about the start and end-values. The time is also represented in a circle which many are not used to in our culture. Of all the diagrams that was presented, the column diagram was the most liked and the least confusing, as it was well known and presented the "Ocean Heat Anomaly" (Priester, 2022) dataset better. It is still a bit to cluttered un unorganized, but this will be improved upon later. In this visualization we get a much better sense of scale because the bars extend from zero all the way to the value of that "year of observation" (Priester, 2022). This also has the benefit of creating two distinct parts along the x-axis. In the 19th century, the average trend is clearly below zero, and in the 20th century, the average trend is way above zero. This creates a stark contrast between the two sides, showing clearly the upward trend in time.

3. After much deliberation, we decided that the column diagram represented the "Ocean Heat Anomaly" (Priester, 2022) dataset best. Here is the final visualization based on feedback and improvement on the column design:



(Created by inserting "Ocean Heat Anomaly" (Priester, 2022) datapoints into column diagram in Excel)

Here one can clearly see both the trend in time, as well as the scale of the data. We also changed the colors representing hot and cold, so it was clear that positive high temperatures (orange to red) were on top and negative low temperatures (light blue to blue) was on the bottom. We also chose to neglect the vertical lines for each year since it cluttered the diagram, and one does not need it to see the “year of observation” (Priester, 2022) effectively. When looking for the trend in time it clearly shows that the average “heat content anomaly” (Priester, 2022) was below zero, until the 1970s where it became more unstable, and by the 2000s, it started to increase exponentially. The titles are simplified; “Heat Content Anomaly (10^22 Joules)” is now “Heat Content Anomaly” and “Year of Observation” is now “Year”. The diagram is also a bit larger so that the years do not lay on top of each other, and the background is now dark-themed to better contrast against the content. Having a clean, simple and high contrast design increases readability and this will hopefully make it more accessible to a diverse group of people.

Task 4. Discussion and conclusion

1.
    a) Based on the original dataset (Task 2) all objects and attributes are represented except for the “month of observation” (Priester, 2022) as it was an unnecessary inclusion in our opinion. This is because all measurements were taken in the same month (June), conveying these datapoints would take away from the rest of the information and introduce unwanted complexity. The “heat content anomaly” (Priester, 2022) datapoints are represented as bars and does not include value in text. This is so that the visualization is clean and readable because having each value in text for every bar would clutter the diagram and create confusion. Each bar still retains its value in visual form (bars). The y-axis and the x-axis are both clean and simple and the diagram has horizontal lines to help connect the values, especially on the far away right side. Vertical lines were dropped as we felt it got to cluttered, and one does not them to see the “year of observation” (Priester, 2022) clearly. This is because the vertical distance is shorter than the horizontal distance and there is nothing obscuring the distance. Color also plays an important role in the visualization as it clearly shows the difference between hot and cold, and to an extent, average and abnormal. The bars are white and therefore have high contrast with the surrounding background. They also have shadows which aids in the illusion that they pop out from the background further helping them pull focus. The background is also dark and the text is light creating another level of contrast not only within the diagram but also to the surrounding external content. For all these reasons the final visualization (Task 3, 3) conveys the information well.

b) Working with this assignment has been difficult, but educational. Selecting a sustainable development goal went relatively fast. We sat down and went through all the 17 goals and discussed the possibilities with each of them. In the end we decided on (SDG-14) “Conserve and sustainably use the oceans, sea and marine resources for sustainable development” (The United Nations Department of Economic and Social Affairs, 2016) because we thought that it was the most important and pressing issue since it affects all life on the planet in a major way. One major challenge we had was with user testing as we were not sure how to conduct it most effectively. Although finding people within the target audience was relatively easy. The process of creating visualizations was challenging at first as we had some issues with Excel, but when those were resolved the process went more smoothly. We spent a long time trying out different visualization techniques with the “heat content anomaly” (Priester, 2022) dataset. It was challenging to create the final visualization (Task 3, 3) because we had to sacrifice some potential information as to not introduce unwanted complexity and sacrifice readability. This was decided with the visualization theory and the feedback in mind. That’s how our group has gained new insight into the “heat content anomaly” (Priester, 2022) dataset and visualization. We also had to get some background knowledge on the subject to properly understand how to present the data. Using “Climate Change: Ocean Heat Content” (Lindsey & Dahlman, 2020) as a reference helped immensely in this regard.

2. In conclusion, the assignment has been a challenging, but very educational and interesting process. Using Excel as our visualization tool we strived to follow the task at hand and create an informative visualization with the theory we had learned to aid us. We met some hiccups along the way but eventually overcame them and moved forward. Our group agreed on the theme and sustainable development goal (SDG-14) “Conserve and sustainably use the oceans, sea and marine resources for sustainable development” (The United Nations Department of Economic and Social Affairs, 2016) (Task 1) early in the assignment. We also discussed the utility of the “heat content anomaly” (Priester, 2022) dataset and defined a target audience. We spent some time diving into the dataset and explaining it and then reasoning why we chose it. Our group then made 3 visualizations and finalized one of them. We then went through a process of user testing where we got an insight into what others thought of our visualizations. The final visualization (Task 3, 3) is a combination of all the lessons we learned from the previous visualizations (Task 3, 1). We decided which visualization served the “heat content anomaly” (Priester, 2022) dataset best and corrected the problems that was brought to light during user testing. Also, we also learned a lot from our chosen theme (Task 1) and dataset (Task 2) as we needed to reference explanations and definitions on the subject matter.

References list

Climate.gov. (2022). *Ocean Heat Trends (1993-2021)* [Change in heat content in the upper 2,300 feet (700 meters) of the ocean from 1993-2021. Between 1993–2021, heat content rose by up to 6 Watts per square meter in parts of the ocean (dark orange). Some areas lost heat (blue), but overall, the ocean gained more heat than it lost. The changes in areas covered with the gray shading were small relative to the range of natural variability. NOAA Climate.gov image, based on data from NCEI]. https://www.climate.gov/media/14846

The United Nations Department of Economic and Social Affairs. (2016). *Conserve and sustainably use the oceans, seas and marine resources for sustainable development (SDG-14)*. Established as regulations. The National curriculum for the Knowledge Promotion 2020. https://sdgs.un.org/goals/goal14

Priester, J. (2022, November). Ocean Heat Anomaly. *Ocean Heat Content Since 1955*. https://www.kaggle.com/datasets/jarredpriester/ocean-heat-anomaly

Lindsey, R., & Dahlman, L. (2020). Climate Change: Ocean Heat Content. *Bull. Amer. Meteor. Soc*., 103 (8), S143–S191, https://www.climate.gov/news-features/understanding-climate/climate-change-ocean-heat-content

Johnson, G., C. and R., L., Lumpkin, Eds., 2022: Global Oceans [in "State of the Climate in 2021"]. *Bull. Amer. Meteor. Soc*., 103 (8), S143–S191, https://doi.org/10.1175/ BAMS-D-22-0072.1.