CS 480X Final Project Process Book

Tim Connors, Federico Galbiati, Miles Gregg, Josh Malcarne

Table of Contents

Project Overview

Project Overview and Motivation

Overview

Our project samples ice crack data in the arctic from the months of November to April from the year 2003 to 2021.. An interactive timeline is included to show the current month and year of the data and to allow users to scroll across the data set, and an area plot shows the total area of ice cracks for each data point, measured using the amount of pixels that represented a crack in the ice in the data.

Motivation

We wanted to explore the potential impact that climate change had on Earth's natural world, and we found that measuring the arctic's ice crackage would be a unique but informative way of tracking environmental well-being. Additionally, our group was interested in mapping out some form of geographical data. We decided that we could combine both goals effectively with our approach.

Questions We Want to Answer

As can be seen in the next section of the process book, the questions we wanted to answer with our project changed vastly as we altered the focus of our assignment.

Within the final iteration of our project proposal, the questions we are trying to answer are:

- What are the general trends of the volume of ice cracking in the arctic over the course of a winter?
- Where does this ice cracking usually occur? Is it centered around a specific area or dispersed throughout the entire arctic?
- Has there been any significant change in ice cracking in the arctic over approximately the last twenty years?

In our analysis, we realized that the overarching story the visualization would tell would be the year-to-year change in the volume of ice cracks in the arctic and attempted to focus our visualization more on this specific question by incorporating an area plot displaying the total amount of ice crackage for a given data point compared to the date the data was recorded.

Design Evolution

Initial Project Proposal

Summary

Our project initially centered around mapping out self-driving car data. We planned to visually represent how a self-driving car responds to its environment while taking a given route. We would use camera and lidar data to draw vectors on a 2D map that would show the stimuli the car is responding to, graphing its velocity, acceleration, and direction to show how these factors impacted its motion. Users would be able to select different routes and see how the car goes from its source to its destination in a simulation.

Problems

We found out that self-driving car data sets are absolutely massive. Audi's data set, for example, was over 2 TB in total, with even its preview measuring in at 4.3 GB. Its data for individual cities was also in the hundreds of MBs. Figuring out how to store, parse, and use all of this data would be beyond the scope of the project. Attempts to truncate the data sets may also lead to unexpected behavior, such as the data stopping in the middle of a drive. We also realized that representing all of that data at once could be overwhelming for the user.

Questions we wanted to answer: What specific stimuli impact a self-driving car’s algorithm the most? How do self-driving cars respond to traffic and unexpected obstructions? What behaviors do self-driving cars have that are different than human drivers?

First Revised Project Proposal

Summary

After pivoting from the self-driving cars idea, we decided that we still liked the idea of of overlaying our data on some geographical representation, like a map or a globe. We thought about visualizating import and export data for countries for different materials, linking the export of a country to the corresponding import of another country. We also wanted to show complex interactions between multiple nodes of data but realized that displaying all the import and export data for every country and material at once would cause a lot of clutter. Therefore, we decided to use chord diagrams and implement filtering functionality to make the data more intuitive, exploratory, and user-friendly.

Problems

We were able to to find a data set with import and export data for a variety of countries and materials, but the data set did not link the exports of one country to the imports of another.

Questions we wanted to answer: Which countries are most relied on to export certain materials? What nations export the most raw materials and import the most refined ones? How do geopolitical factors contribute to the amount of interactions between specific countries?

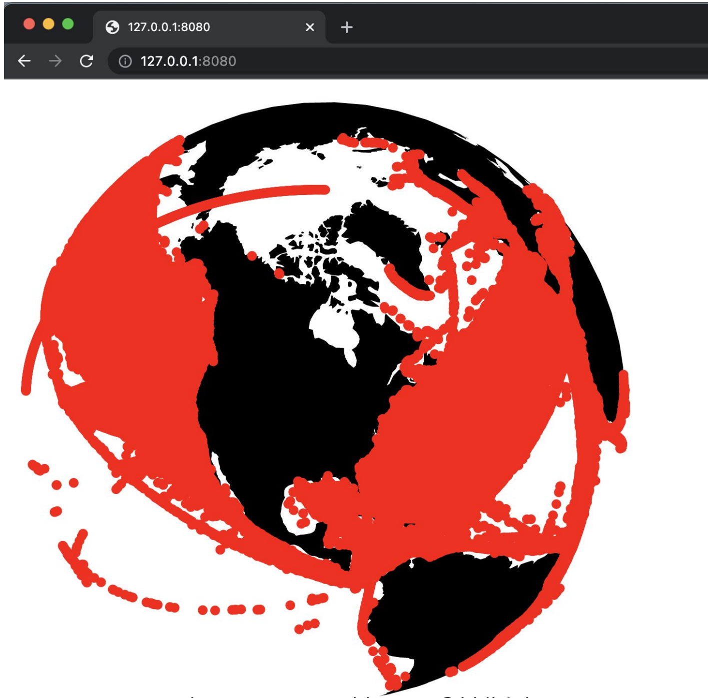
Second Revised Project Proposal

Summary

We then narrowed the scope of our project to focus on international shipping routes. As you can see on the right we got a 3d globe to render and plot 250,000 points of cargo shipping lanes which is in red. Plotting the data was very easy to do in d3 with the built in functions for 3d geological mapping. The project was a great idea however with lack of time and resources it seemed to be out of reach.

Problems

The problem with this is that every dataset that we found was very large and when plotting the data onto a 3d globe in d3 it made it very slow. This also made it very difficult to animate, rotate, or zoom in or out on the globe to view different locations around the world. Along with his data we wanted to include some type of marine life data however it would've been a lot to compute for a website all at once with a 3d globe and 250,000 points for ships.



Questions we wanted to answer: Which ports in the world are the most visited? What goods are most commonly transported by sea? Which country has the largest shipping presence in the world?

Final Revised Project Proposal

Summary

For our final revised project proposal we wanted to look into how over the years how the arctic ice has evolved during recent years. As you see in the figure the red circles shows the arctic circle which is the area we focused all of our data on. Doing this project was very interesting for all of us to do since we all wanted to a data visualization on the worldwide perspective. The preprocessing of all of the data took us a while to get down and so did all of the d3.js plotting of the data.

Reasoning

The main reasoning for the project is that we wanted to look into how pollution may affect the arctic ice in which it does in several different ways over the years. Once we got all of our data and we knew how to execute the final goal it got much better. Since we spent a lot of time looking at and preprocessing different datasets which took a long time.

Data and Implementation

Datasets

The main dataset used for this project is provided by the Space Science and Engineering Center at the University of Wisconsin Madison. The dataset is provided in NetCDF format, a compressed data file with multiple raster layers. Each layer is a 7025x7025 matrix representing data in the Arctic Circle. Each data file contains multiple layers including a sea-ice leads mask, cloud, coverage and more. Due to the format of the data, the files must be preprocessed to be used effectively in d3.js.

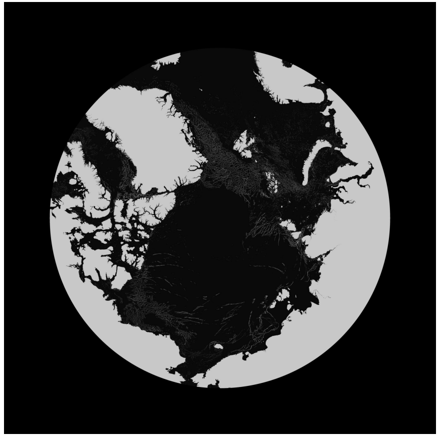
The daily data since 2002 is released through an FTP repository. To reduce the number of days, one day of data per month was extracted for every year. This data was downloaded through a custom Python script using ftplib.

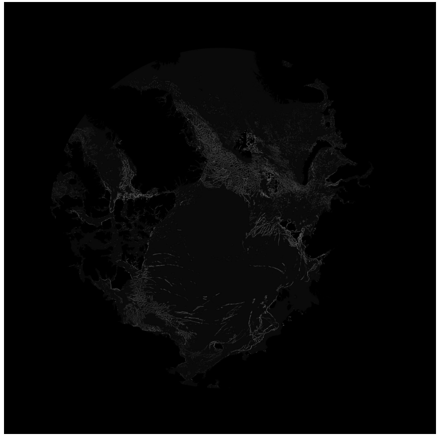
Geographic Data Pre-Processing

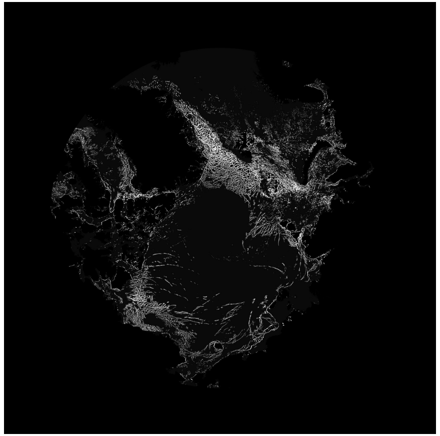
Once the data was acquired, it had to be converted from raw NetCDF to a usable format. GDAL was used to convert this data through multiple preprocessing stages:

- NetCDF to GeoTIFF
- Remove the data outside the Arctic Circle
- Remove the land layer using a filtering function
- Transform the mask into a binary classification
    - The values of the mask are in range 12-255
- Export to PNG

Geographic Data Pre-Processing







Raster Image Data Pre-Processing

Once the data was in PNG format, it was discovered that it was flipped left-to-right. Therefore, an additional preprocessing pipeline using Pillow and Python was performed to mirror the data. Additionally, the original layer was in black-white. To increase visual contrast, the color of the sea-ice leads was changed to red.

Time Series Pre-Processing

To reduce the computation on the client side, we pre-processed statistics of data for each day in the dataset. The data extracted through this phase allowed to generate a time series plot as part of our visualization.

The statistics extracted were encoded into a JSON file to allow for easy access using the d3.json function on d3.js.

{"2018345.leads.png": 1829717, "2005075.leads.png": 1414987, "2012045.leads.png": 1495664, "2013105.leads.png": 553357, "2015045.leads.png": 1178256, "2014105.leads.png": 1116976, "2017075.leads.png": 1130403, "2006105.leads.png": 539458, "2010075.leads.png":
1199559, "2007045.leads.png": 1166279, "2006015.leads.png": 1140345, "2014015.leads.png": 1160023, "2013015.leads.png": 1589369, "2006045.leads.png": 1555768, "2011075.leads.png": 869965, "2007105.leads.png": 749518, "2016075.leads.png": 915780, "2019345.leads.png":
667180, "2015105.leads.png": 679967, "2003075.leads.png": 1134810, "2014045.leads.png": 946518, "2012105.leads.png": 727369, "2013045.leads.png": 1139564, "2004075.leads.png": 611703, "2012015.leads.png": 926000, "2015015.leads.png": 1132324, "2018315.leads.png":
1600245, "2007015.leads.png": 1233087, "2007315.leads.png": 1569702, "2018015.leads.png": 899280, "2012315.leads.png": 709201, "2015315.leads.png": 1639725, "2014345.leads.png": 1342336, "2013345.leads.png": 1483620, "2019045.leads.png": 1183170,
"2018105.leads.png": 872879, "2009075.leads.png": 1113738, "2006345.leads.png": 1409645, "2019015.leads.png": 1323100, "2014315.leads.png": 1942442, "2013315.leads.png": 1823354, "2006315.leads.png": 1494383, "2007345.leads.png": 1618455, "2012345.leads.png":
1260809, "2015345.leads.png": 1442642, "2008075.leads.png": 1034651, "2019105.leads.png": 290073, "2018045.leads.png": 554513, "2011315.leads.png": 1579422, "2016315.leads.png": 1615888, "2003315.leads.png": 1408224, "2004315.leads.png": 1716510,
"2009015.leads.png": 1603000, "2009105.leads.png": 257281, "2008045.leads.png": 1305705, "2018075.leads.png": 1079341, "2005345.leads.png": 1459130, "2020315.leads.png": 1458285, "2017345.leads.png": 1013851, "2010345.leads.png": 1380157, "2005315.leads.png":
1240785, "2008015.leads.png": 1161761, "2017315.leads.png": 1167483, "2010315.leads.png": 1708410, "2020345.leads.png": 1225736, "2011345.leads.png": 1144490, "2016345.leads.png": 1730514, "2019075.leads.png": 1018383, "2009045.leads.png": 999565,
"2008105.leads.png": 880824, "2003345.leads.png": 1574014, "2004345.leads.png": 1193422, "2003045.leads.png": 968917, "2014075.leads.png": 1079267, "2005105.leads.png": 553747, "2013075.leads.png": 934864, "2004045.leads.png": 1192764, "2009345.leads.png": 1279628,
"2006075.leads.png": 1135805, "2011045.leads.png": 939423, "2010105.leads.png": 858659, "2016045.leads.png": 1173623, "2017105.leads.png": 794879, "2020045.leads.png": 1460029, "2017015.leads.png": 1420124, "2010015.leads.png": 951569, "2008315.leads.png": 1558426,
"2005015.leads.png": 1324788, "2016105.leads.png": 824745, "2017045.leads.png": 1410981, "2011105.leads.png": 705521, "2010045.leads.png": 1322323, "2007075.leads.png": 832583, "2020015.leads.png": 1197638, "2005045.leads.png": 891714, "2012075.leads.png": 827585,
"2004105.leads.png": 622456, "2015075.leads.png": 1066664, "2003105.leads.png": 704097, "2008345.leads.png": 1344535, "2009315.leads.png": 1416388, "2003015.leads.png": 1192981, "2004015.leads.png": 1259957, "2020105.leads.png": 723673, "2011015.leads.png": 1159832,
"2016015.leads.png": 1204933}

Data Hosting

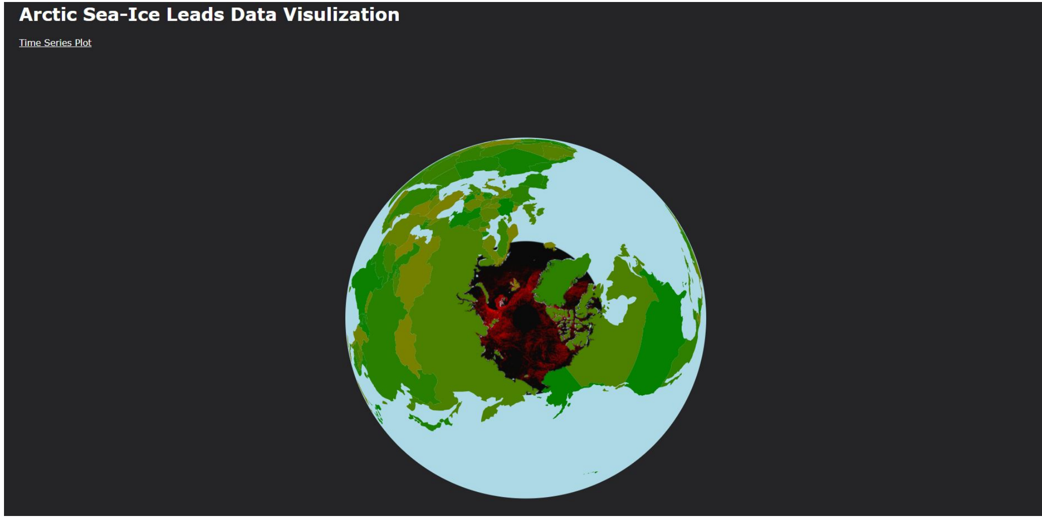
A Node.js server was created to host the datasets. The nodejs server had to be setup using Express and the CORS plugin to allow for a Cross-Origin Resource Sharing policy. The NodeJS instance was hosted on Heroku, a hosting platform available in the US.

Workflow improvements

- Auto-deploy was setup between the GitHub backend repository and the Heroku instance
- Automatic rebase of the GitHub pages fork was setup to allow for immediate update of the website

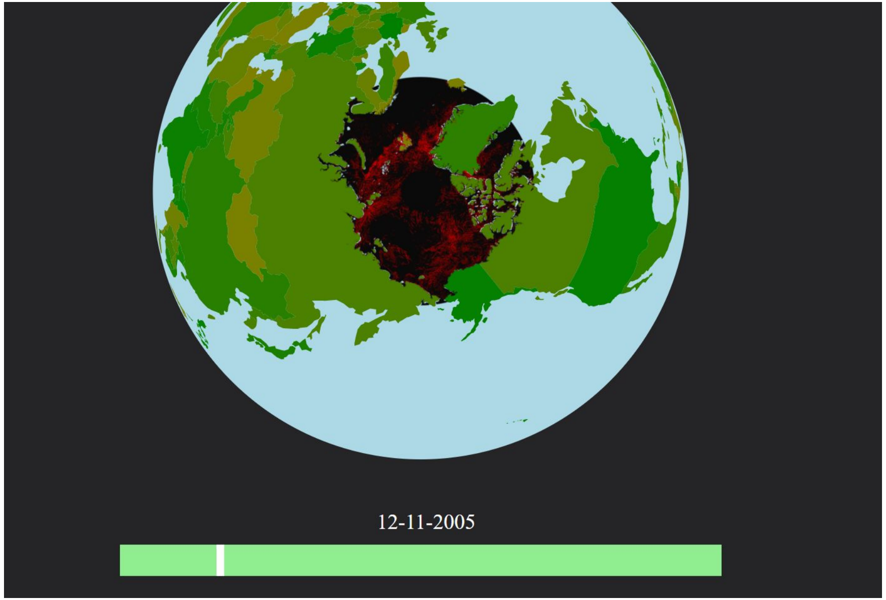
Interactive Map

Our map has zooming functionality, allowing the user to see the nuances of both the ice patterns and the surrounding landforms. Additionally, the contrast between the colors of the ice, ice cracks, and the rest of the map makes it very easy to interpret to perceive the difference between the ice data and the rest of the globe.



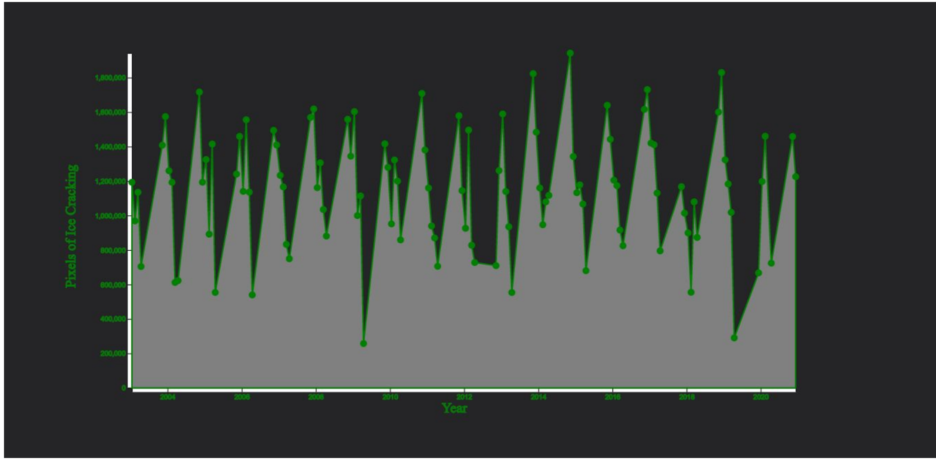
Scrollable Timeline

Below our map, we included a timeline with a slider that allows the displayed data point to change. The selection is determined by the position of a particular data point in time, relative to other data points. This allows the user to look through all our data and explore it at their own pace, letting them move forwards and backwards in time in a nonlinear fashion. We also implemented a link to this timescale at the top of the webpage, making it easy for users to access and interpret the bar. The selected date is also displayed above the bar and is represented by a vertical white line in the bar.



Summary Area Plot

At the bottom of our visualization, we also implemented an area chart that displays the relationship between the time a data point was recorded and the total amount of ice cracks shown in its corresponding image file, measured by the amount of red pixels in the image. The plot allows users to see the bigger picture of the data, addressing one of our main goals in creating the visualization. It displays both seasonal and year-to-year trends.



Evaluation

What we Learned About the Data

Our visualization revealed that there seemed to be no significant long-term trends in the amount of ice cracking in the arctic circle over the span of the last eighteen years that we examined. We found this surprising, since we expected there to be more cracks in the ice as time progressed due to rising global temperatures caused by climate change. We postulate that this may be because the rising temperatures have not yet had a significant enough effect to cause a noticeable change in ice cracking in an extremely cold place like the arctic or that the amount of ice cracking is somehow independent of global temperature increase.

How we Answered our Questions

We used a combination of precise, detailed data display through the map and timescale and more general data representation through the area chart to answer our questions. As shown by the area chart, ice cracking seems to spike right around December and January and then sharply declines after the early winter months. The map shows that most of the ice crackage is centered around the North Pole and seems to be particularly concentrated in northern Siberia. As previously stated, the area chart shows that there has been no significant change in the amount of ice cracking in the arctic circle over approximately the last twenty years.

What we Could Improve Upon

While we generally do find our visualization pleasing to look at and use, there are some improvements we could make to it. If given the opportunity, we would have liked to add a couple more chart types to the visualization to give a more complete view of the data. We would also like to work on polishing up our map features, like the zooming and timescale features. Additionally, it may also have been useful to layer some form of environmental data, such as average temperature levels, over our existing data to see if the spots where cracking was most frequent had any correlation to temperature.

Related Work and Exploratory Data Analysis

Related Work

In order to make our maps, we used Harry Stevens' tutorial on creating a transparent map and geeksforgeeks' tutorial on implementing d3's geoOrthographic() function. A number of other sources and references used can be found in comments in the corresponding sections of code as well.

Exploratory Data Analysis: Kiln’s Global Shipping Traffic

Kiln's visualization of shipping traffic inspired us to track geographical data over a period of time and explore the possibility of visualizating shipping data. We saw how fascinating it was to see how the data points can change over time. It encouraged us to pursue trying to map shipping data ourselves, which proved to be more impractical than we anticipated.

Exploratory Data Analysis: Anthony Skelton's Earthquake Visualization

We took inspiration from [Anthony Skelton's d3 earthquake visualizations] in designing our visualization, specifically its use of a timeline and a map to show geographical data over time. The visualization gave us a good reference to go off of for displaying time-based geographical data. The timescale proved to be a very useful tool in presenting our data in an organized and intuitive manner.