Assignment - 2 Report

Adithya Nagaraja Somasalle
IMT2021054

Pannaga Bhat
IMT2021080

Goutham U R
IMT2021045

***Abstract*—This is a report of CS 732: Data Visualization A2:InfoVis and SciVis from Team Colour Blinds.**

I. Introduction

This report describes the various visualizations i.e parallel co-ordinate plots, treemaps, node-link diagram for information visualization and visualizations such as colour maps, contour maps, and quiver plots for scientific visualization.

II. Information Visualization

A. *Parallel Co-ordinate Plots*

The data for all the parallel co-ordinate plots is the Our World in Data database [1]. The columns for every plot are the 6 countries namely India, the United States, France, Russia, Brazil and South Africa and the first column is the month corresponding to each polyline.

1) Total Cases per Million in 2022: Fig. 1 shows the progression of the pandemic in 6 countries using the total number of cases per million metric. We can clearly see the increase in the amount of cases through the year.

2) Stringency Index: Fig. 2 shows the 'stringency index' for each country at the first of every month. Stringency index is a 'composite measure based on 9 response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100' [1]. Here, we can see the evolution of the government response in different countries. For example, the measures in the US got more lenient through the year however in Russia, the response generally became more severe.

3) Monthly Vaccination per Hundred: Fig. 3 shows the vaccines administered in the 6 countries per hundred every month. We can see that developed nations such as the US or France were able to administer more vaccines than developing ones such as India or South Africa.

B. *Treemap*

The data for all the Treemaps is the covid dataset in Our World in Data database [1]. There are different hierarchy types of visualizations made in this. These Treemaps are plotted using Plotly library in JavaScript. To run the code, live server extension must be used. Different spatial partition strategies like slice, dice and squarify have been tried, but the default image generated by Plotly looked better, so the original image is chosen. The values in each Treemap are sorted from left to right and top to bottom for better visualization. This means that as we go from left to right, either the value remains same or decreases, same for top to bottom. This treemap is interactive, if you click on one of the rectangle it will zoom into it and show the hierarchy of children of that rectangle.

1) Total Number of Cases for Continents with Countries as of 2021-01-01: Fig. 4 This figure displays the treemap for the total number of cases for Continents with country hierarchy as of 1 January 2021. It can be seen that the highest number of cases was found in Europe and least was found in Africa. The highest cases were found in Russia, United States, India, Brazil, South Africa for the continents Europe, North America, Asia, South America, Africa respectively.

2) Total Number of Deaths for Continents with countries as of 2021-01-01: Fig. 5 displays the treemap for the total number of deaths for Continents with country hierarchy as of 1 January 2021. It can be seen that the highest number of deaths occurred in Europe and least was found in Africa. The highest deaths occured in United Kingdom, United States, Brazil, India, South Africa for the continents Europe, North America, South America,Asia, Africa respectively. By clicking one of the rectangles(continent), we can see the hierarchy for that particular continent, like Fig. 6

3) Number of cases for India with date hierarchy for year 2021: Fig .7 shows the treemap for the new cases for every date and month in year 2021. We can see the highest number of new cases was on 7th May 2021, this was the peek of second wave of Covid. We see that by the end of the year the number of new cases decreased (December has the least number of new cases).

4) Number of people who are Vaccinated in each continent as of 2022-01-01: Fig. 8 shows the amount of people vaccinated and not vaccinated for each of the continents. Clearly, Asia has the highest number of vaccinated people Oceana has the least amount of people vaccinated.

Overall, tree maps offer advantages in terms of hierarchical representation, efficient use of space, interactive exploration, and effective communication of complex structures, making them valuable for a wide range of covid data visualization.

C. *Node-Link Diagram*

The dataset assigned was Similarities (DBpedia). Raw dataset was converted to a csv file using Excel. The dataset consists of pairs of numbers representing nodes, where each row signifies the presence of an edge between the two corresponding nodes. The edges are undirected. The Nodelink diagrams were generated accordingly using Gephi. The algorithms chosen are Fruchterman-Reingold, Force Atlas and OpenOrd.

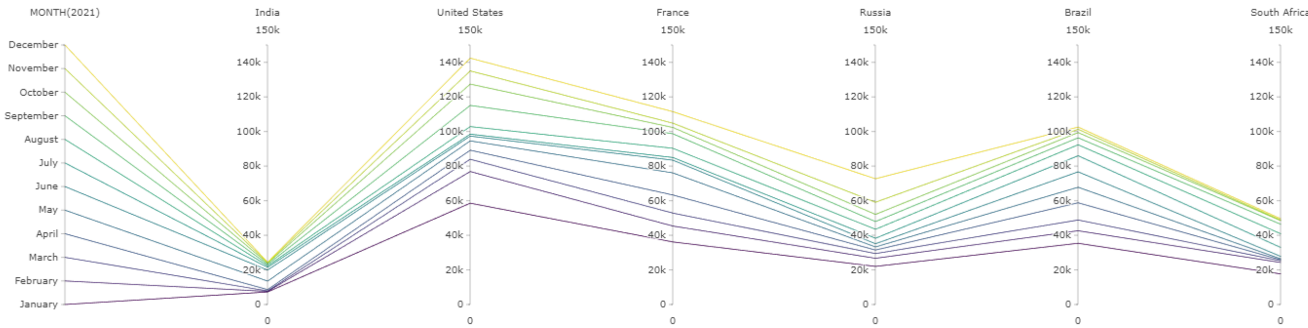

Fig. 1: Parallel coordinate plot showing total number of cases per million in 6 different countries in the beginning of every month in 2022.

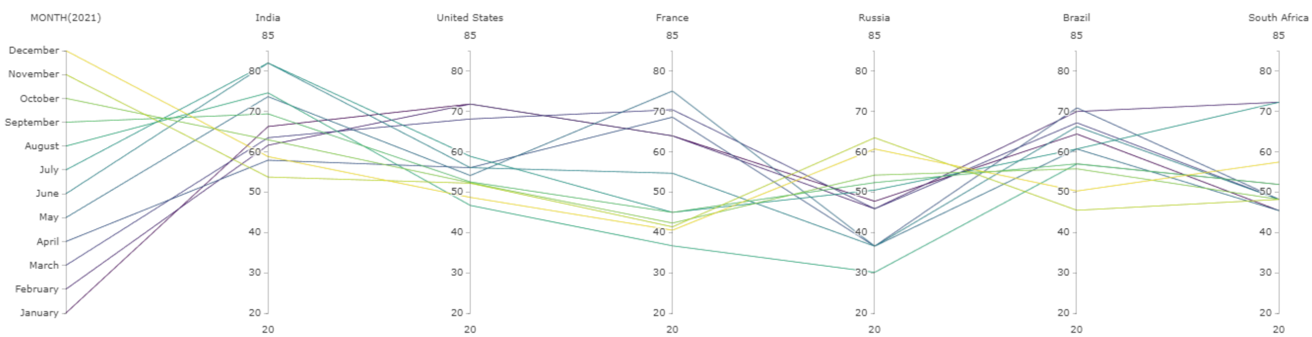

Fig. 2: Parallel co-ordinate plot showing the stringency index in the 6 countries at the beginning of each month in 2022.

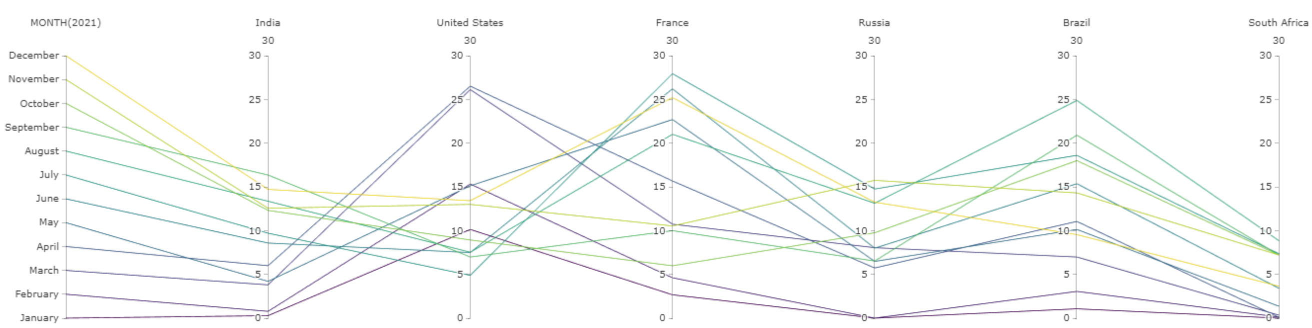

Fig. 3: Parallel co-ordinate plot showing the monthly vaccinations performed per hundred population in the 6 countries in each month of 2022.

Fig. 4: Total Number of Cases for Continents with Countries as of 2021-01-01.

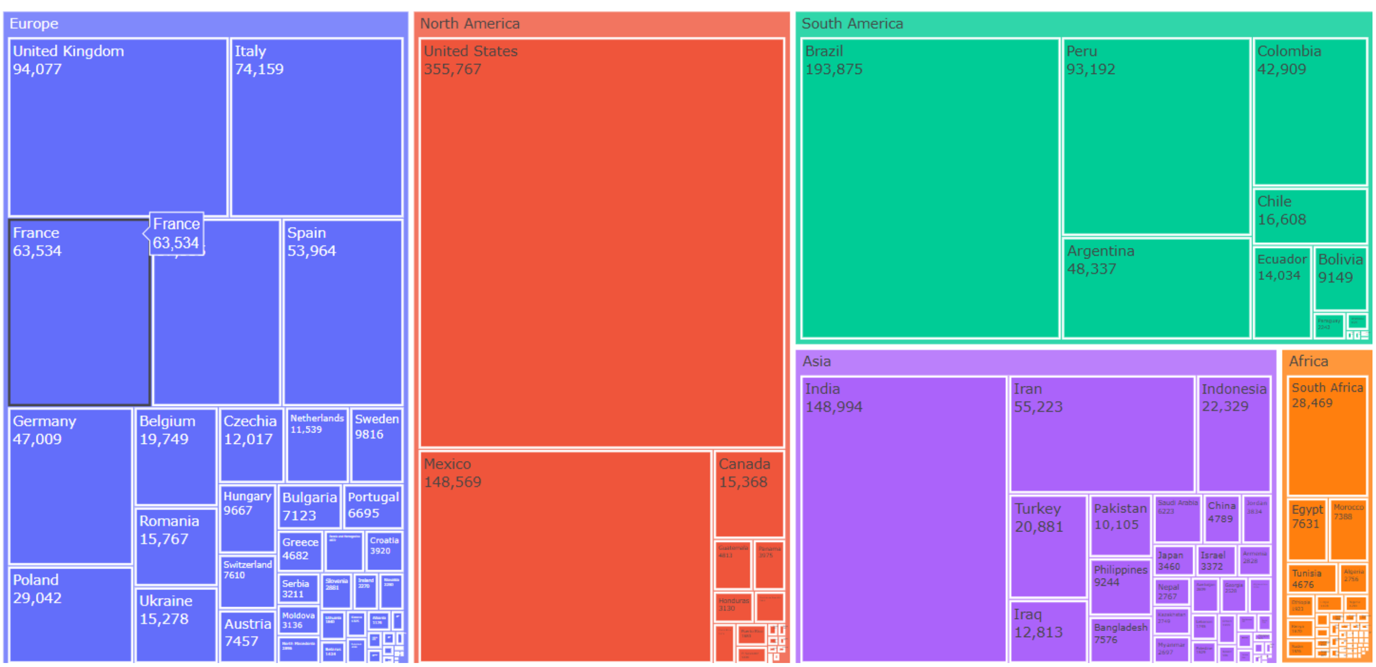

Fig. 5: Total Number of Deaths for Continents with Countries as of 2021-01-01.

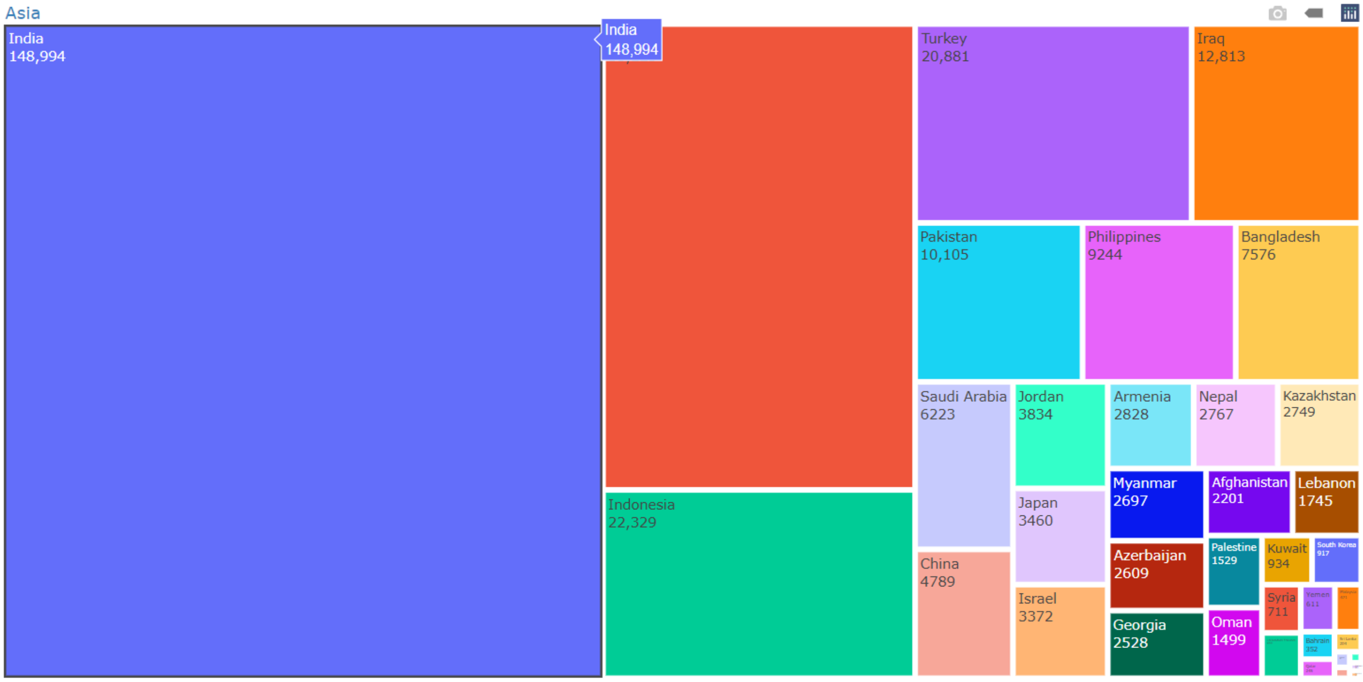

Fig. 6: Total Number of Deaths for Asia as of 2021-01-01.

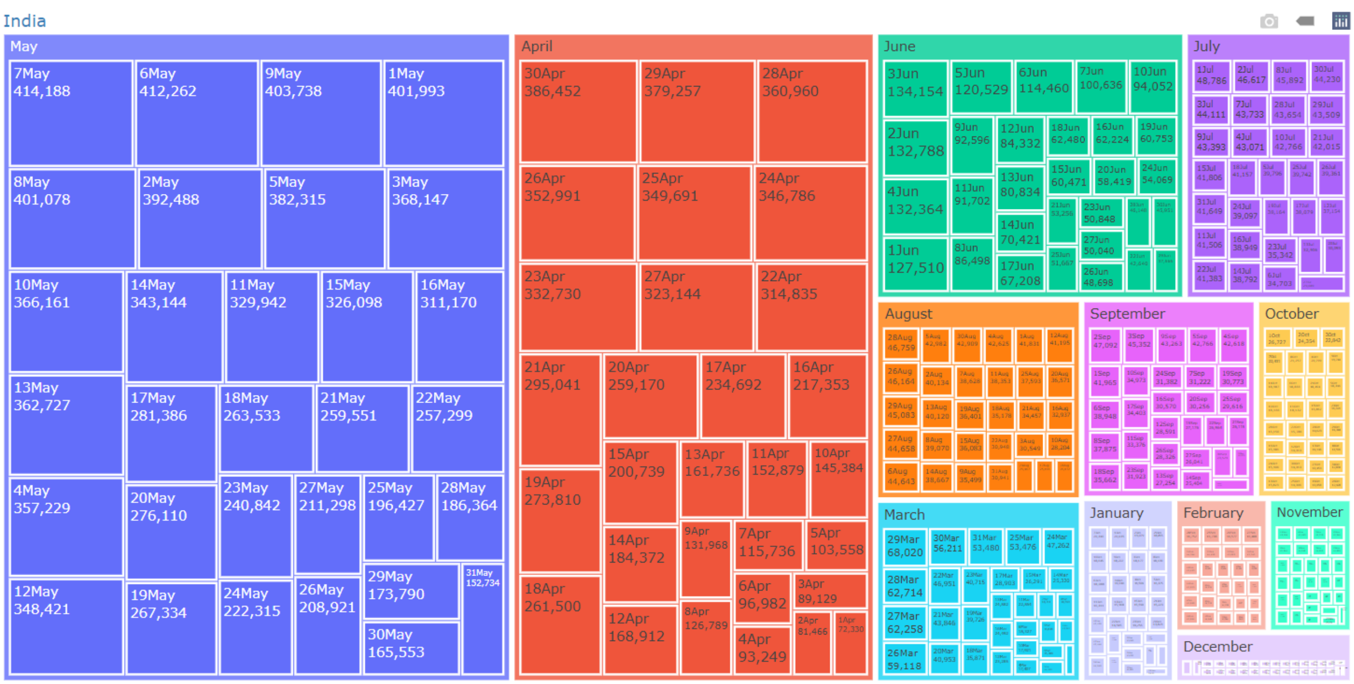

Fig. 7: Number of Cases for India with Date Hierarchy for year 2021.

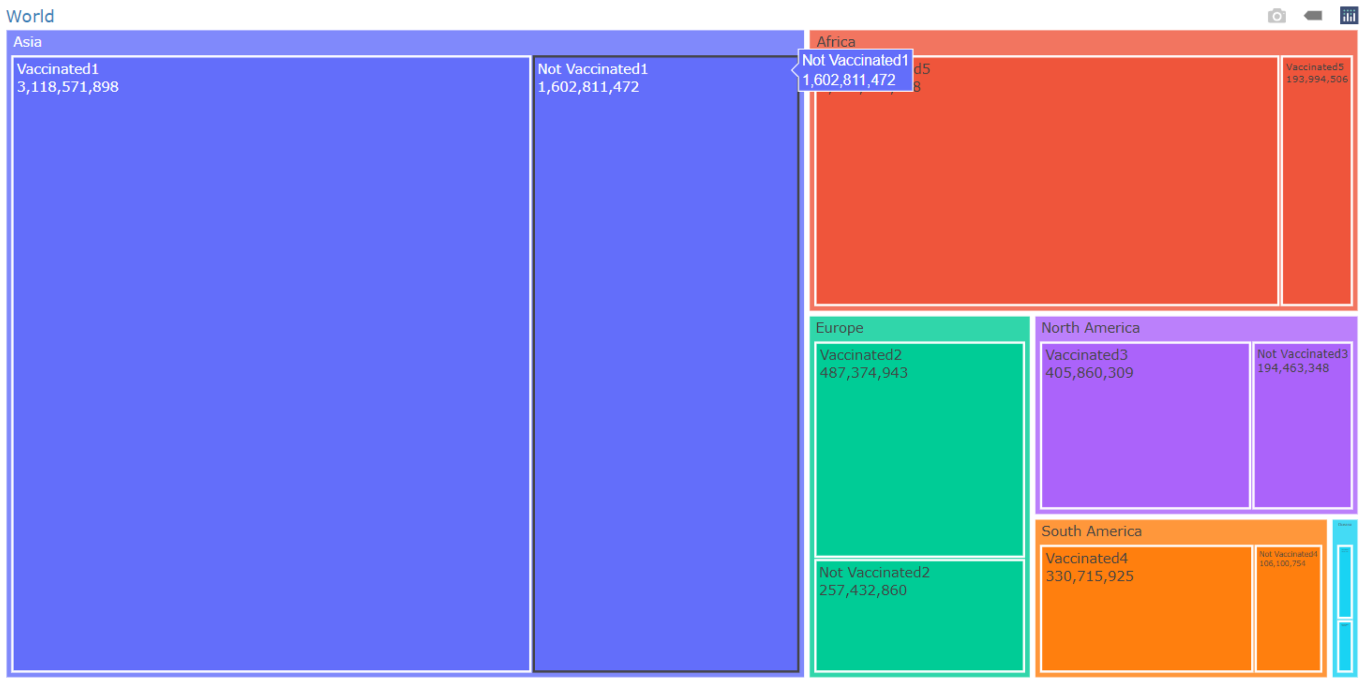

Fig. 8: Number of people who are Vaccinated in each continent as of 2022-01-01.

1) Fruchterman-Reingold: Fig. 9 shows the nodelink diagram obtained when Fruchterman-Reingold algorithm was used. The Node-link diagram has a peripheral ball like structure inside the main boundary. The ball in the corner, potentially representing a group of less connected nodes, may appear at the periphery in a force-directed layout where nodes with fewer connections tend to drift away from the main body of the graph. This spatial arrangement highlights the distinctiveness of the less connected nodes within the overall network structure.

2) Force Atlas: Fig. 10 shows the nodelink diagram obtained when Force Atlas algorithm was used. After applying the Force Atlas algorithm in Gephi to visualize the network structure, the resulting layout was refined for better clarity and reduced overlap through the subsequent application of the Noverlap technique. The concentration of nodes in one main cluster implies the presence of a dominant and tightly interconnected core within the network. Nodes within the main cluster might serve as central hubs, having strong connections with a significant portion of the network. These nodes could represent key elements or central players in the overall structure. The presence of three surrounding clusters indicates a degree of modularity or compartmentalization in the network. Each of these smaller clusters may represent distinct subgroups or communities within the larger network.

3) OpenOrd: Fig. 11 shows the nodelink diagram obtained when OpenOrd algorithm was used. After applying the OpenOrd algorithm in Gephi to visualize the network structure, the resulting layout was refined for better clarity and reduced overlap through the subsequent application of the Noverlap technique. The presence of three major clusters in the center suggests a central hub or hubs with strong connections. This core cluster likely plays a significant role in the overall network structure. Nodes that are scattered randomly outside the central clusters may represent outliers or less connected elements in the network.

III. Scientific Visualization

The source for the colour and contour plots is the ASMR2 Ocean database, specifically the 3 day sea rain rate for timeline January 2022 - March 2022.

A. Colour Maps

1) Pre-processing: The columns of data with missing values were truncated from the data. The overland values were given as -999 and were replaced with a value of 0.

2) Choice of colour map: At first, the colour maps were attempted using the `viridis` sequential scale. However this did not look very good due to the white colour of the truncated values.(See Fig. 12)
With the `coolwarm` diverging scale, a similar issue emerged as well as the fact that there was not a clear boundary between land and sea.(See Fig. 13)
The `grey_r` sequential scale was used so that the 0 values were assigned the white colour. However, since most values were very small, the map looked very monotone and non-zero values were hard to find.(See Fig. 14)
Finally, a logarithmic sequential `pink_r` scale was used to clearly distinguish between black coastal boundaries and the actual data, with satisfactory results. Figs. 15 through 20 show the final colour maps for January 1st, January 31st, February 5th, February 23rd, March 11th and March 31st, respectively.

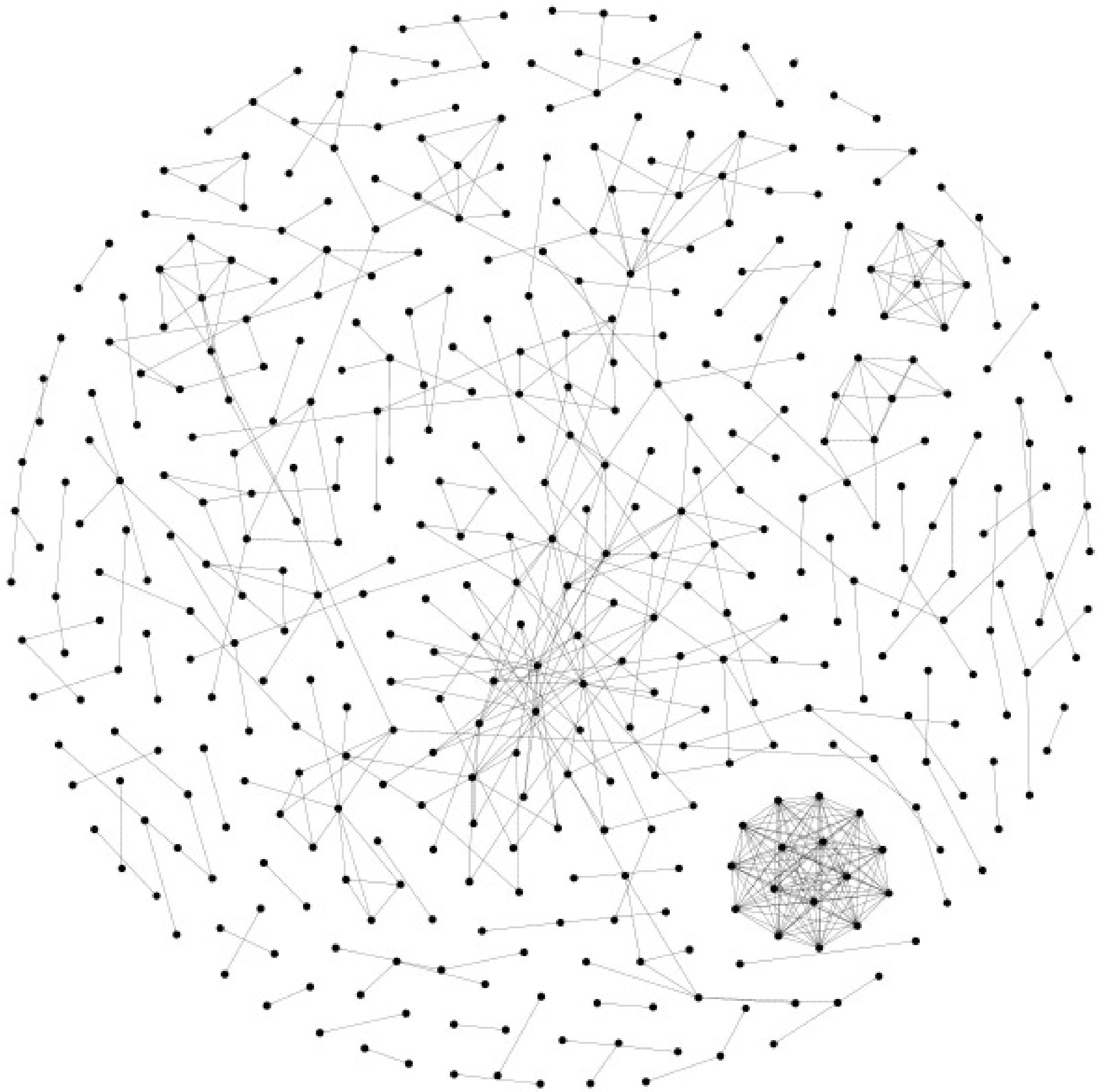

Fig. 9: Nodelink Diagram when Fructerman Reingold algorithm was used.

3) Inferences:

1) January 1st 2022: High rainfall along the equator and the coast of Antarctica. Very little rainfall around the Indian coast, southern Australia and western Canada.
2) January 18th 2022: Rainfall more distribured and most spots have some degree of it. Only the HOrn of Africa and the Arabian Sea are dry.
3) Februrary 5th 2022: Again, rainfall is mostly global. Only Arabian Sea. Bay of Bengal and North America's Pacific coast have very little rainfall.
4) February 23rd 2022: Most of the Indian Ocean received very little rainfall. Otherwise, most of the oceans have received consistent rainfall.
5) March 11th 2022: Similarly to previous, the Indian Ocean received the minimum compared to the rest of the globe. Also, the North American Pacific Coast also has received little rainfall.
6) March 31th 2022: Consistent rainfall throughout the Equator. The northern Indian Ocean again receives the minimum

Note: The data for the Atlantic Ocean longitudes was truncated. This introduces an implicit bias while visualizing the rest of the globe's data.

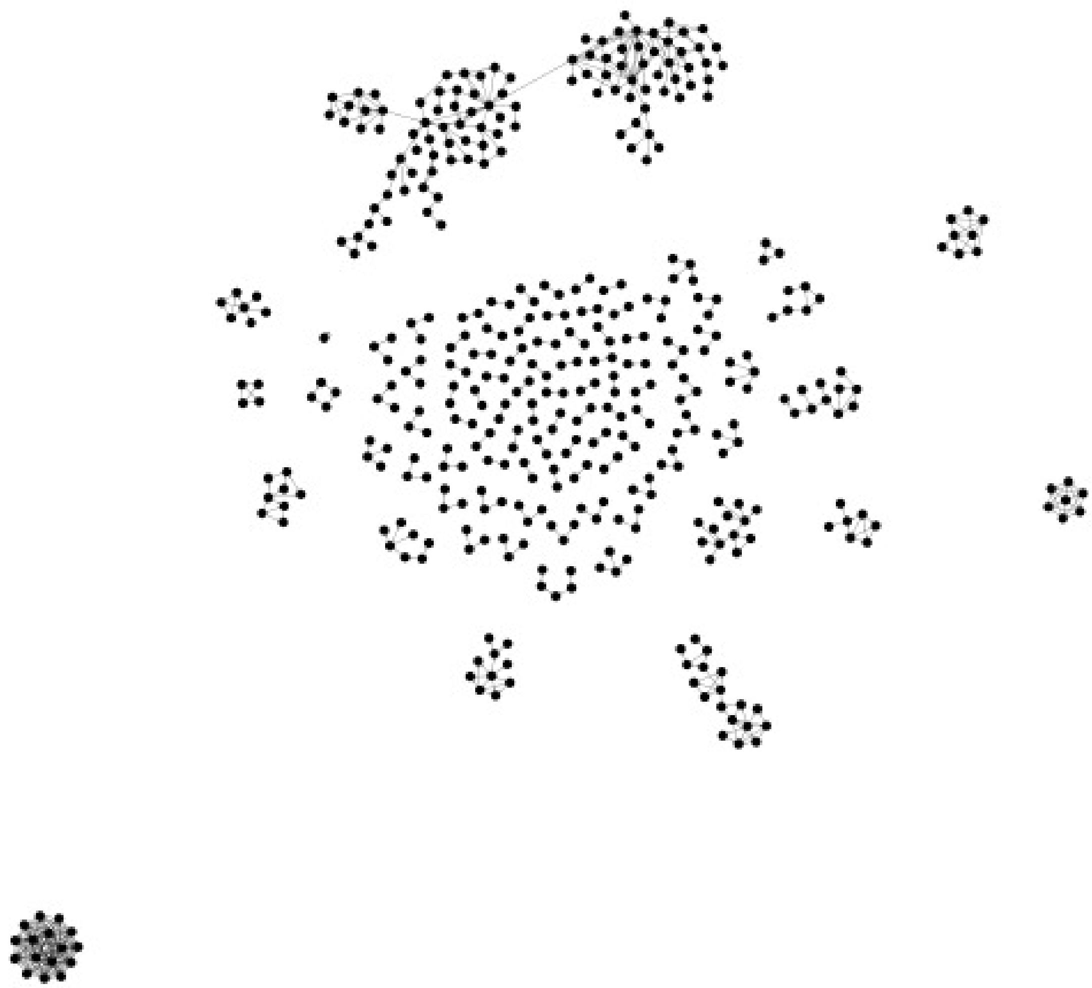

Fig. 10: Nodelink Diagram when Force Atlas algorithm was used.

B. Contour Map

1) Pre-processing: The missing longitude values are reconstructed. The invalid values were given as -999 and were replaced with a value of 0. Line too long from data file was removed manually.

2) Method: First reconstruct the missing longitude values and store all longitude values in array. Store data values for every longitude and latitude in a dictionary. The the latitude is South then negative of that latitude value is considered for convention. For each row traverse until add the values to the dictionary, if there are longitude for which values are not in data then simply add -999 as the value into dictionary. Pass the longitude,latitude and data value into the matplotlib contour function.

3) Choice of Algorithm: Marching square algorithm has been used for the contour maps. The reasons are as follows:

- Marching sqaure algorithm generates smooth and continuous contours, which can be visually appealing.
- Marching Squares can adapt well to different resolutions of data. It is effective at handling irregularly spaced data points or data with varying densities.
- In the presence of noisy data, Marching Squares can provide a smoother representation of contours compared to some contour fill algorithms. It can help reduce the impact of noise in the visualization.
- Marching square algorithm emphasise on the line themselves rather filling the enclosed areas.

Fig. 28 shows the contour plot for sea rain rate dataset [3] using contour fill algorithm (using contourf function of matplotlib) and Fig. 21 shows contour plot using marching square algorithm. The function `contourf` have higher time complexity than `contour`. Contour fill algorithm takes more time than marching square algorithm to plot. Overall the main reason the choose Marching square algorithm over Contour fill

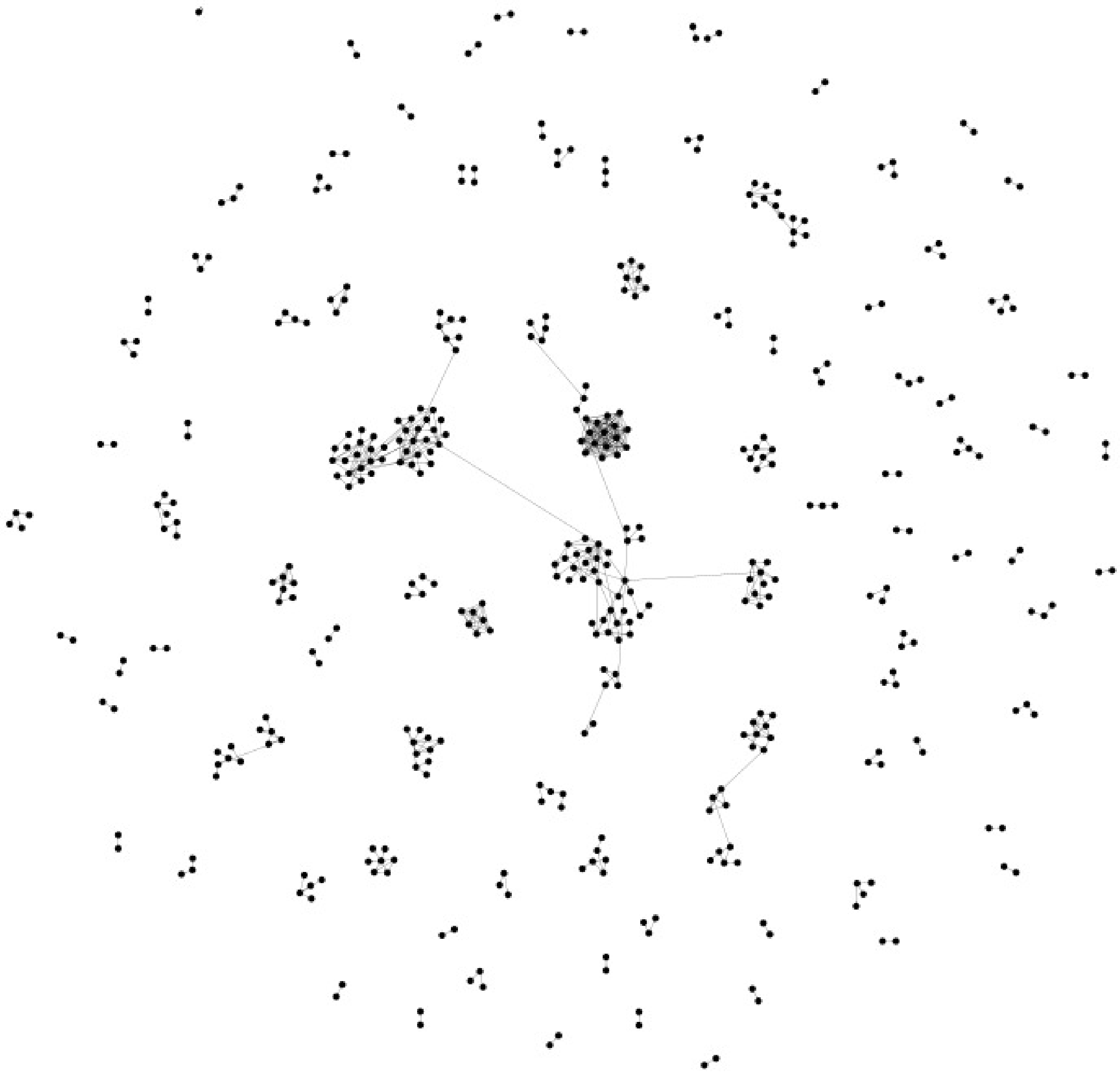

Fig. 11: Nodelink Diagram when OpenOrd algorithm was used.

algorithm is due to the visually appealing characteristic and performance of Marching square algorithm.

4) Choice of Dates: Over the span of 3 months of dataset allocated, the dates are chosen uniformly over the 3 months. Dates chosen :January 1st, January 31st, February 5th, February 23rd, March 11th and March 31st, respectively. These dates capture different events and patterns which will be discussed in inferences.

Figures: Figs. 21 through 26 show the final contour maps. Fig. 27 shows an example of contour map zoomed in. The color bar contains iso-values 0.01(purple),0.3,0.7 and 1(yellow).

5) Inferences:

- Pacific Ocean: Most of the area in Pacific Ocean have high sea rain rate(yellow), more specifically above Australia from January 2022 - March 2022. Among the 6 dates chosen February 5,2022 (Fig. 23) have most of the region in Pacific Ocean with high sea rain rate. Whereas, January 1,2022 (Fig .21 has the most region on the left of Canada covered with high sea rain rate.
- Indian Ocean: Among the 6 dates chosen, January 18,2022 (Fig. 22) have most of the region covered with high sea rain rate. More specifically the region under India.
- Atlantic Ocean: Among the 6 dates chosen, March 11,2022 (Fig. 25) and March 31,2022 (Fig. 26) have the most region covered with high sea rain rate. Most specifically to right of South America.

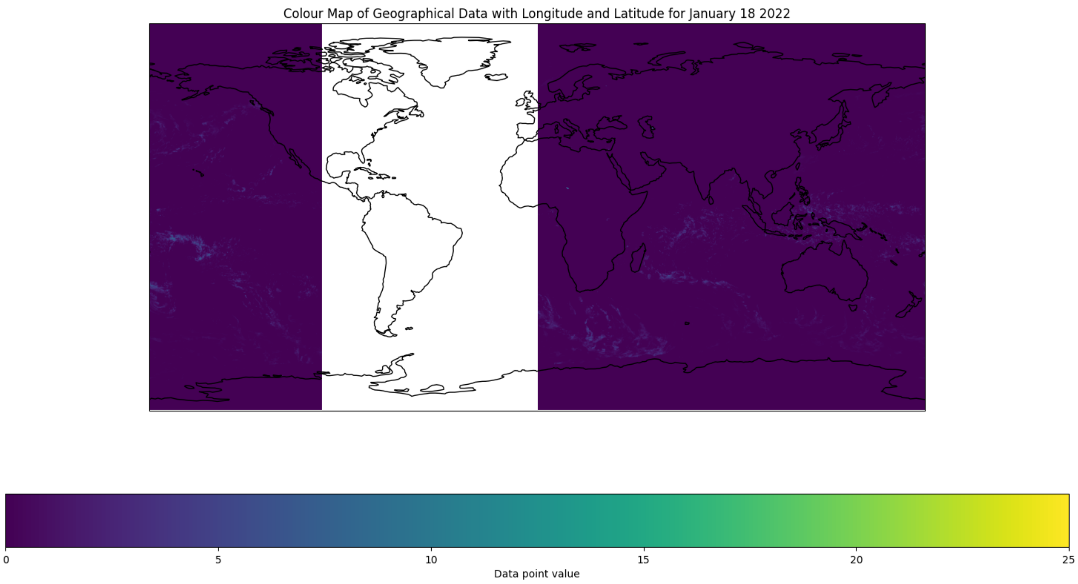

Fig. 12: Viridis example

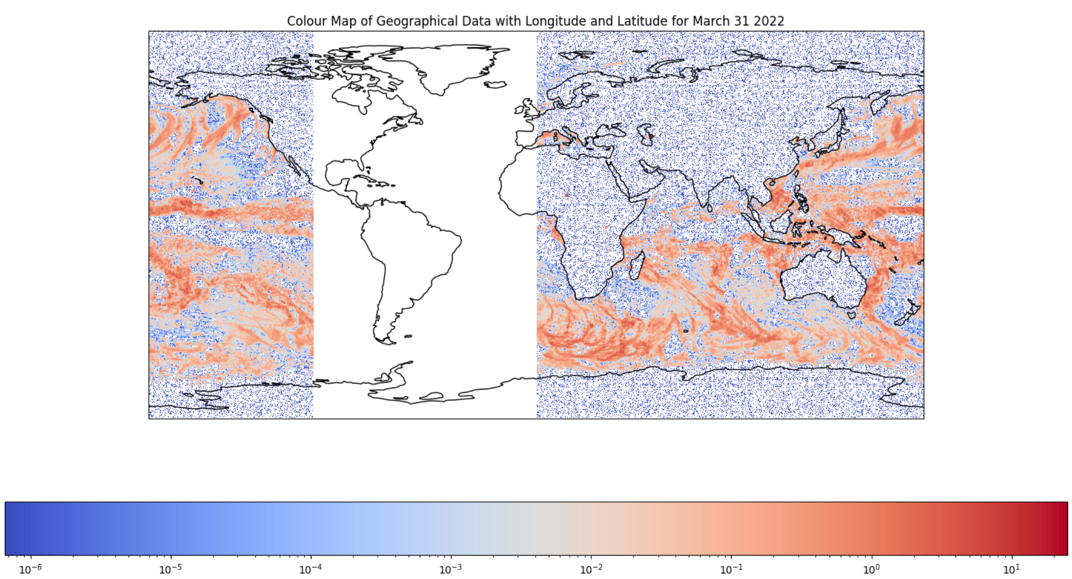

Fig. 13: Coolwarm example

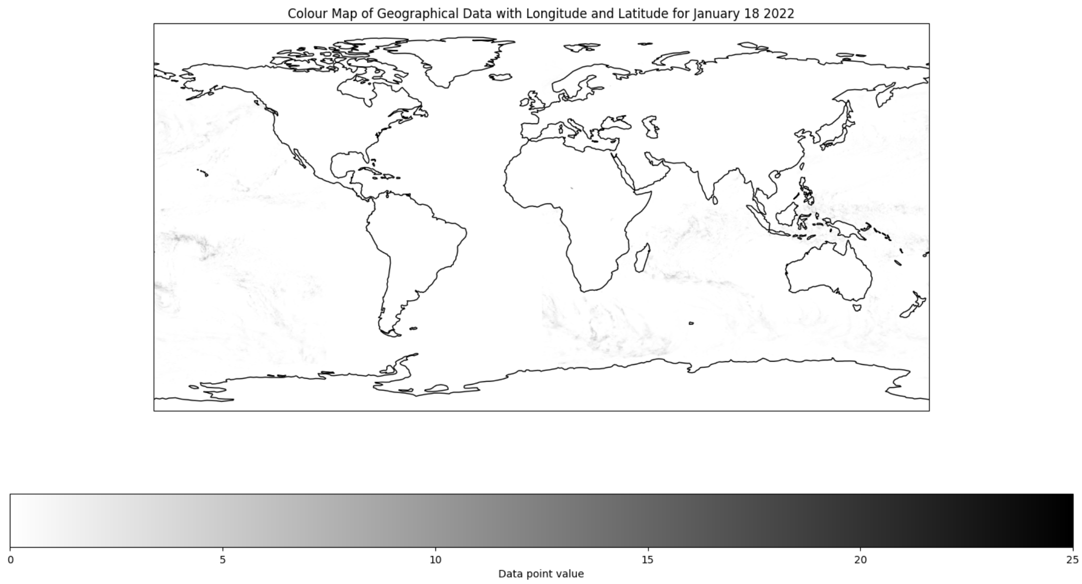

Fig. 14: Grayscale example

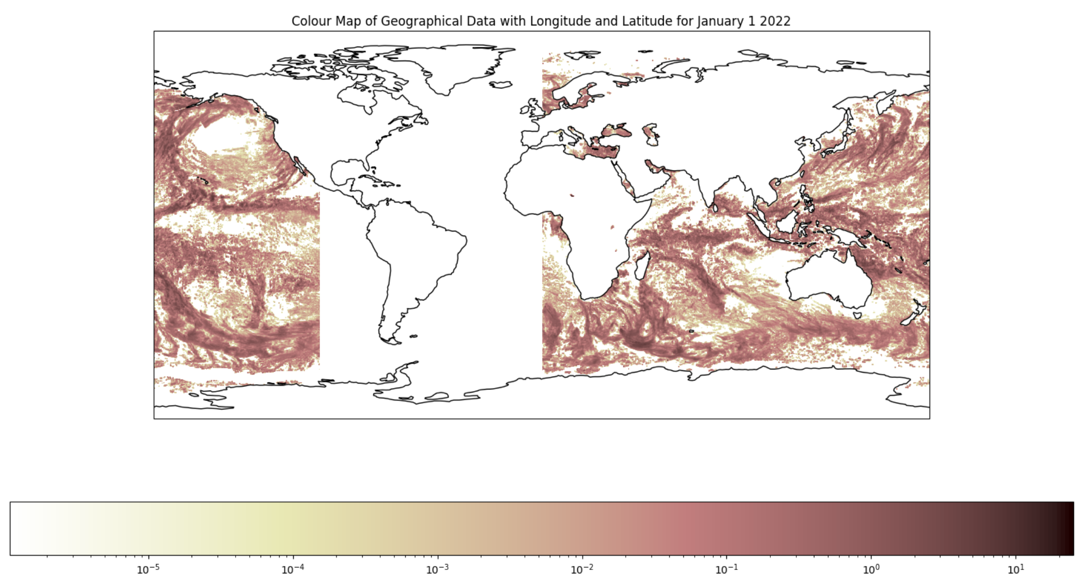

Fig. 15: Colour map for January 1 2022

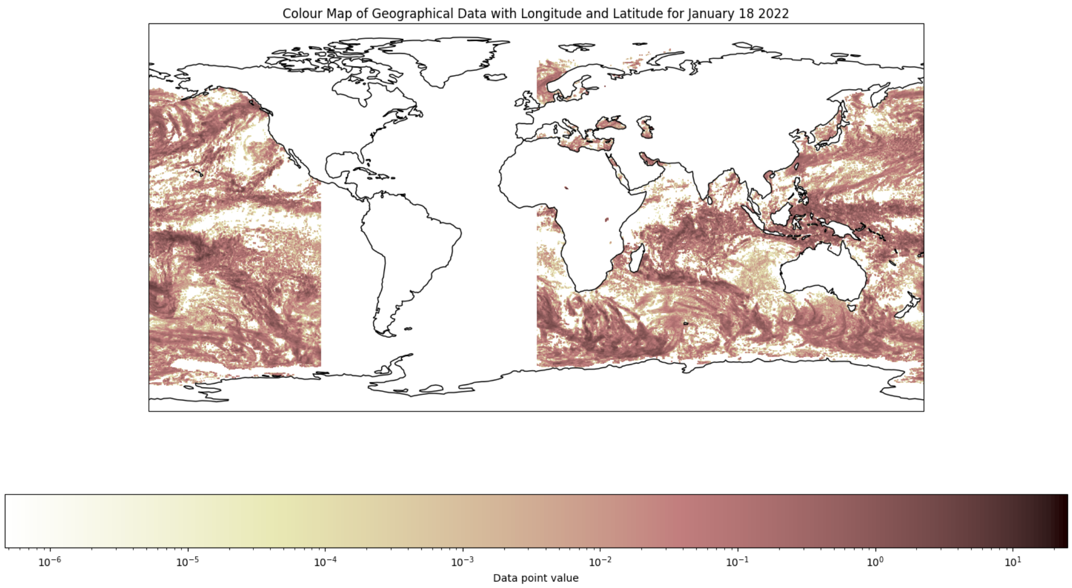

Fig. 16: Colour map for January 18 2022

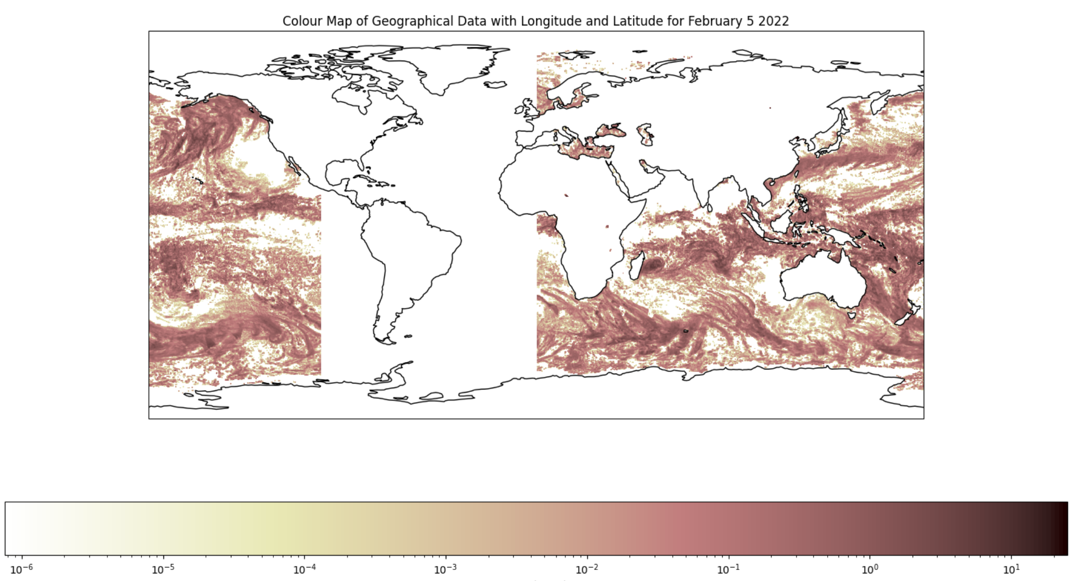

Fig. 17: Colour map for February 5 2022

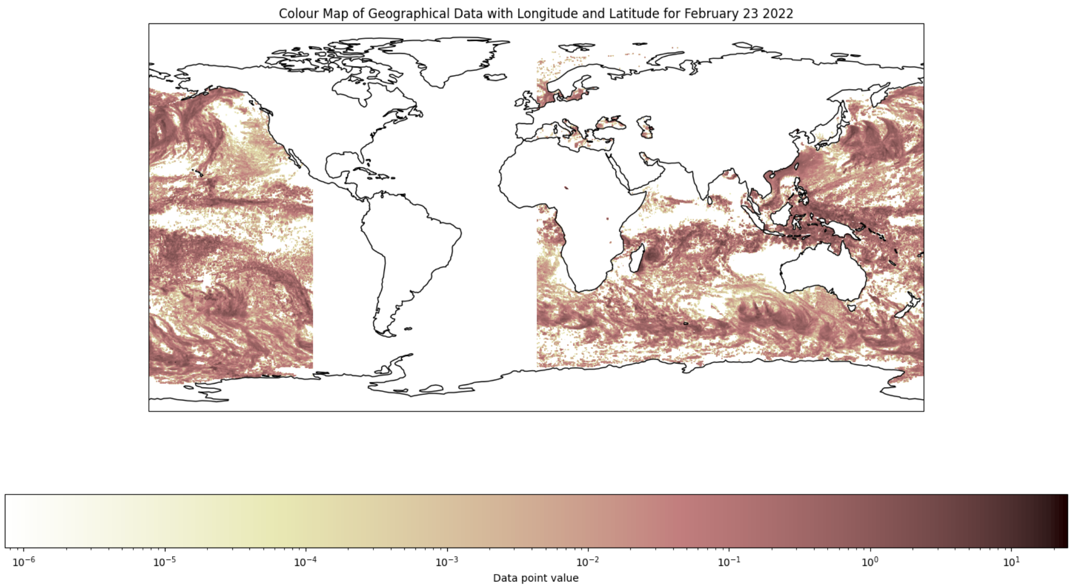

Fig. 18: Colour map for February 23 2022

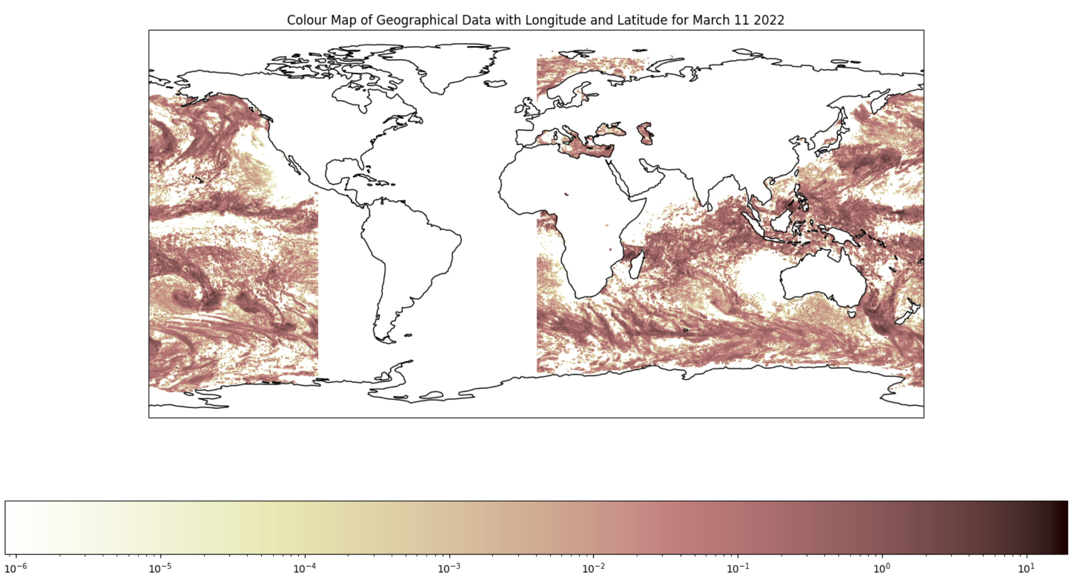

Fig. 19: Colour map for March 11 2022

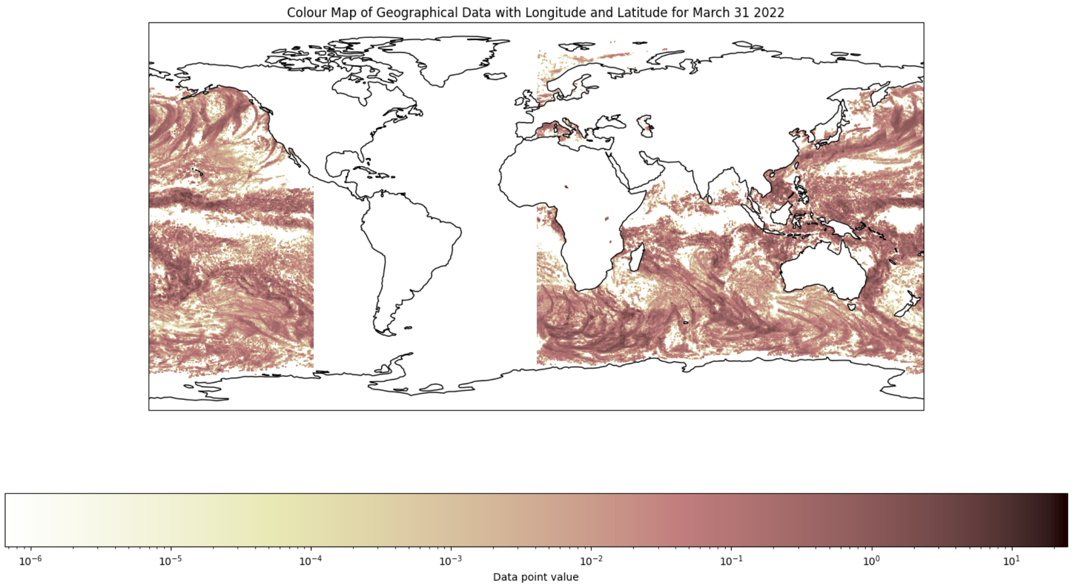

Fig. 20: Colour map for March 31 2022

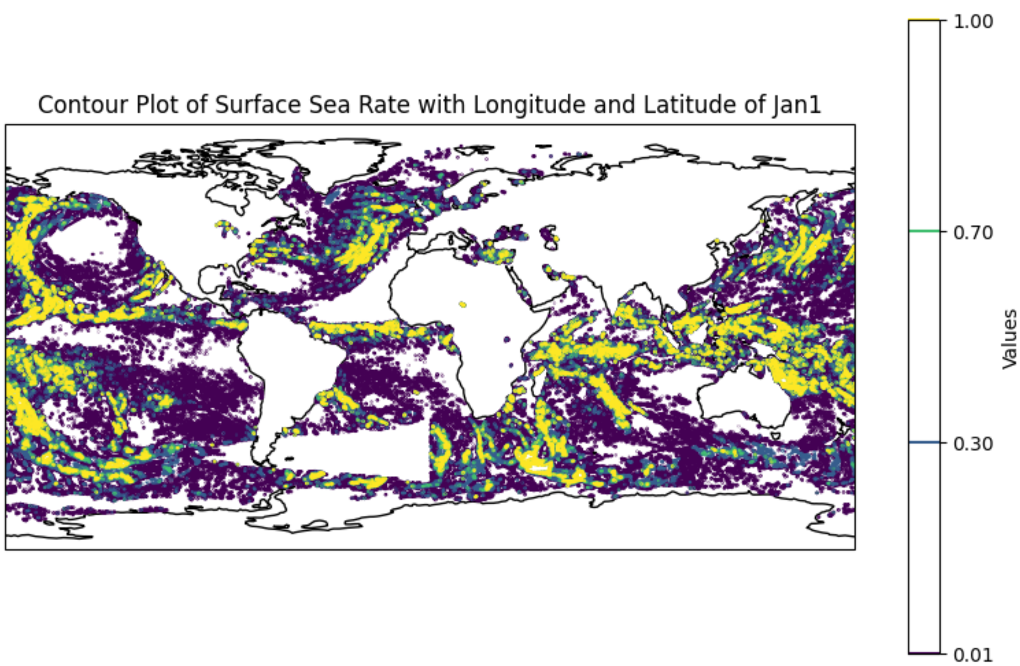

Fig. 21: Contour map for January 1 2022

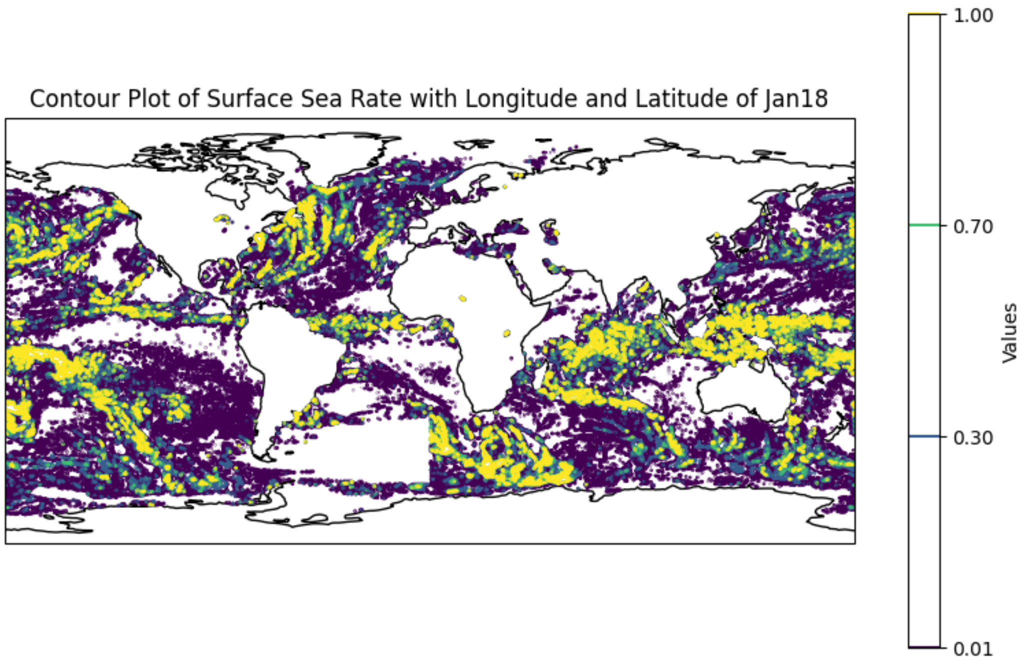

Fig. 22: Contour map for January 18 2022

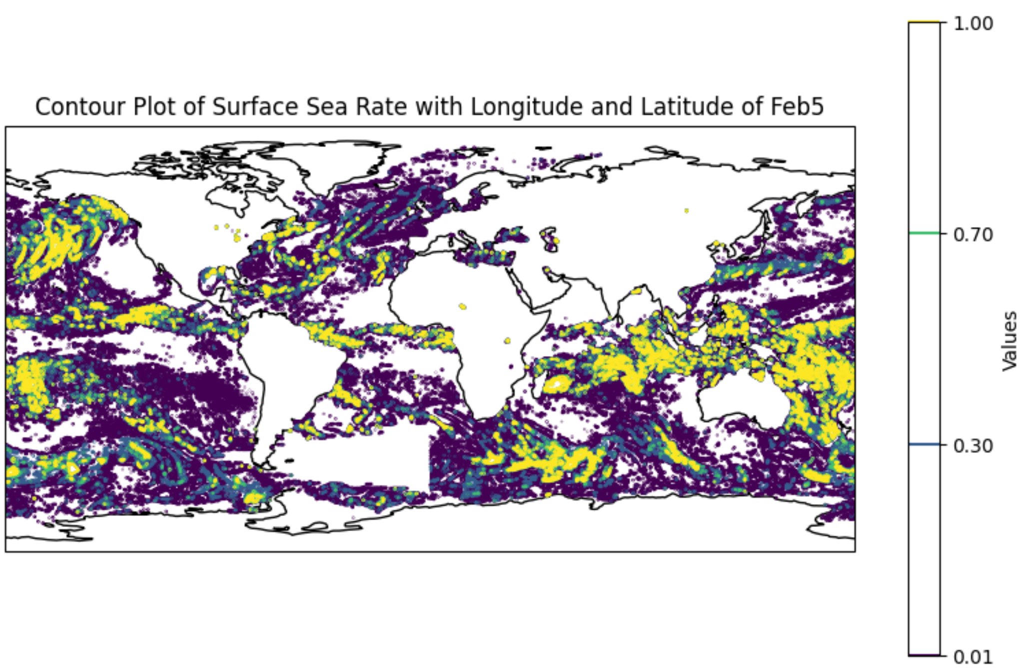

Fig. 23: Contour map for February 5 2022

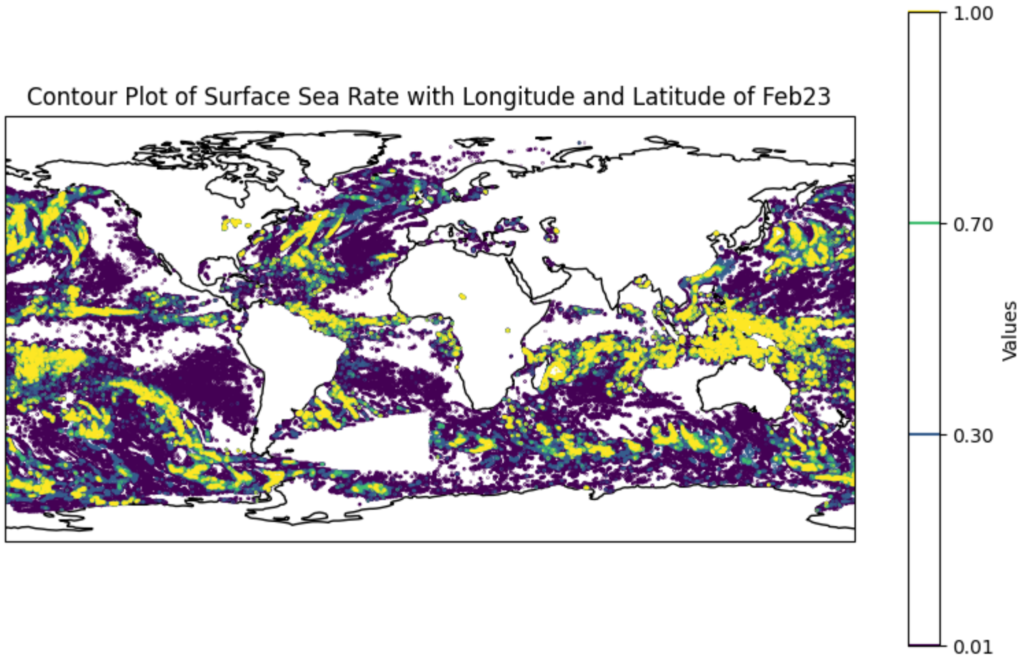

Fig. 24: Contour map for February 23 2022

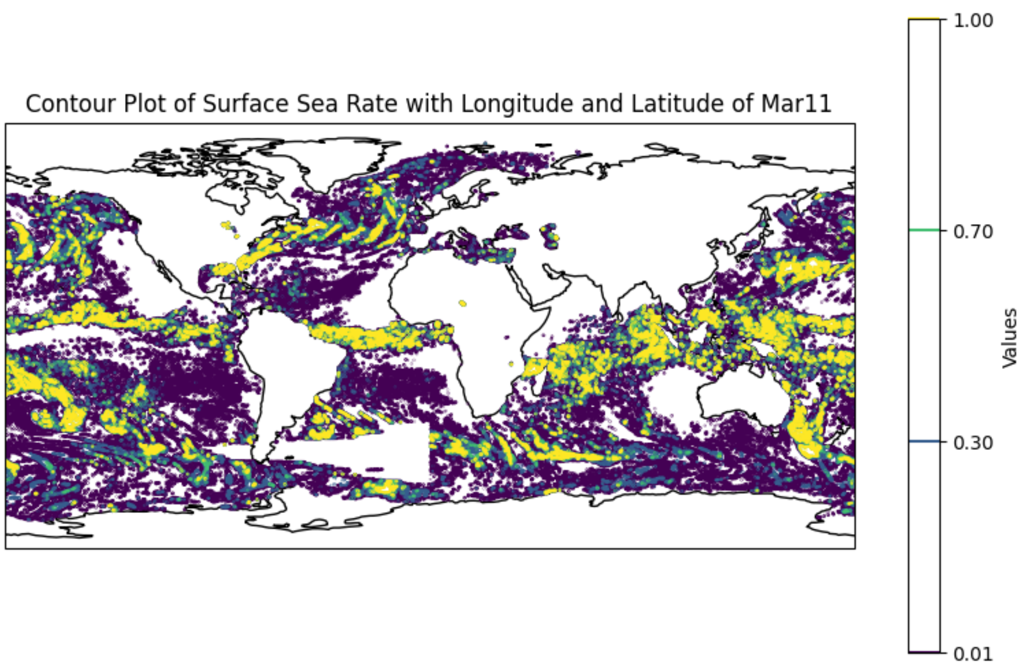

Fig. 25: Contour map for March 11 2022

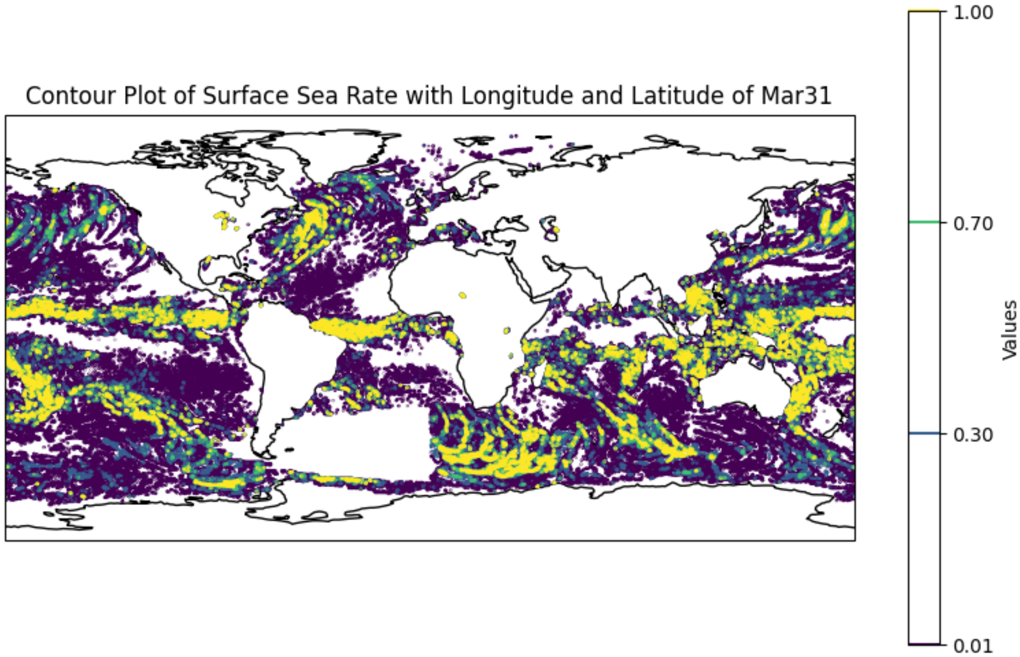

Fig. 26: Contour map for March 31 2022

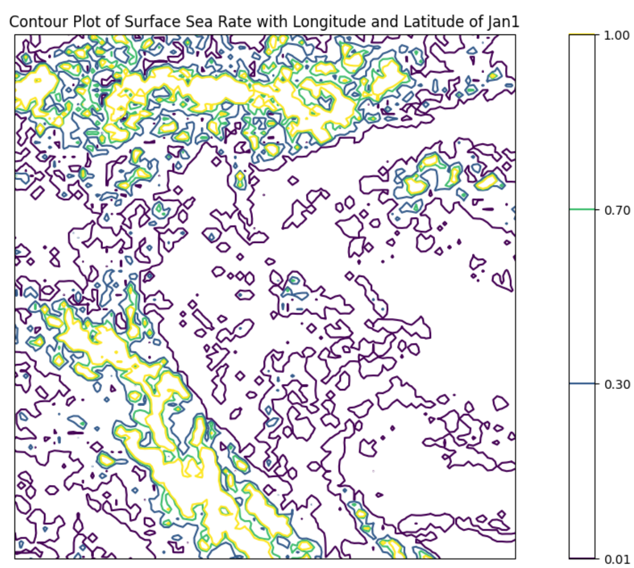

Fig. 27: Zoomed in contour map for January 1 2022

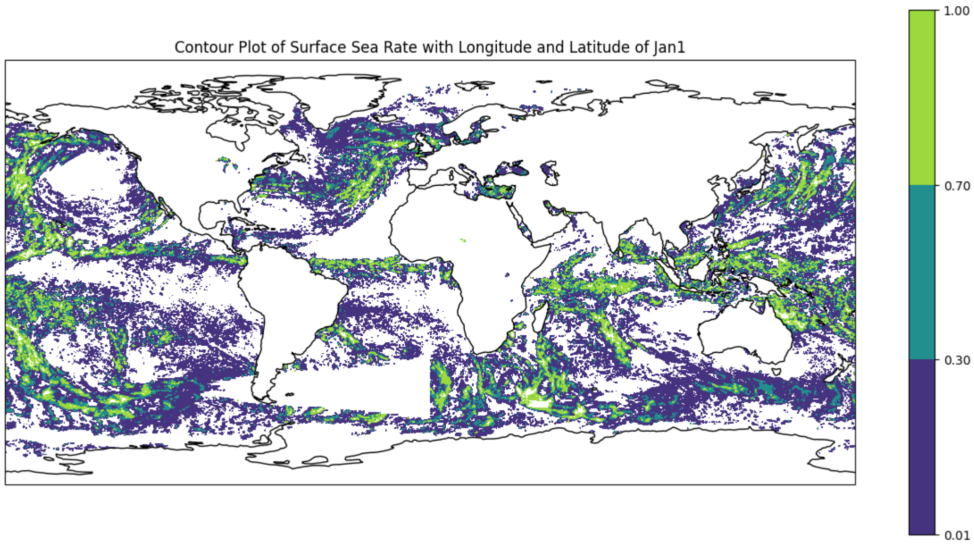

Fig. 28: Contour map for January 1 2022 using contour fill algorithm

- Arctic Ocean : The dataset doesn't specify data values for region in Arctic Ocean, so inference can be made.

Factors to consider for high rain rate:

- Intertropical Convergence Zone (ITCZ): The ITCZ is a belt of low pressure near the equator where trade winds from the Northern and Southern Hemispheres converge. It is a region of active uplift and intense rainfall, contributing to high sea rain rates.
- Seasonal Monsoons: Monsoon systems, characterized by seasonal wind reversals, can bring heavy rainfall to specific regions during certain times of the year. For example, the Asian monsoon and the Australian monsoon contribute to high sea rain rates in their respective regions.
- Tropical Cyclones: Tropical cyclones are powerful storms that can lead to intense rainfall over the oceans. These storms form over warm ocean waters and can bring heavy rain when they make landfall or affect nearby ocean areas.

Noticable Events:

- On February 5,2022(Fig. 23), we see most part of Pacific Ocean is covered with yellow, which means that most of the regions have high sea rain rate in these places. If we focus on the regions just above and right side Australia we see it is completely filled of yellow color. This is due to South Pacific cyclone season. During this we can see high sea rain rate.
- On January 1,2022(Fig. 21, we can see the part right of India (Bay of Bengal) have high sea rain rate. Since, the Monsoon in that region starts during June and ends in September, there is high likely that this was due to cyclone.
- On March 31,2022(Fig. 26, we can see the part to the top right of South America, there is large portion of region covered with yellow. This is due to Intertropical Convergence Zone (ITCZ).The ITCZ is a belt of low pressure near the equator where trade winds from the Northern and Southern Hemispheres converge. It is a region of active uplift and intense rainfall. The position of the ITCZ varies seasonally, influencing sea rain rates in the Atlantic and Pacific Oceans off the coast of South America.
- On March 31,2022(Fig. 26, we can see the part to the south of Africa, there is large portion of region covered with yellow. This is due to transition from summer to autumn in the Southern Hemisphere.These changes can influence precipitation patterns and contribute to higher rain rate.

C. *Quiver Plots*

This section presents a quiver plot visualization generated from geographical data using Python, with a specific focus on the `cartopy` library. The Oscat Wind dataset, spanning from 19th November 2010 to 23rd November 2010. The chosen geographical region for this analysis lies within the latitude range of 30.25°S to 20.75°S and the longitude range of 60°E to 69.5°E. The primary goal is to illustrate wind vectors over this specific geographical area.

1) Data Loading and Cleaning: The first step involves loading the data from the specified files and cleaning it for further processing. The latitude and longitude information is extracted while handling variations such as 'N', 'S', 'E', and 'W' suffixes.

2) Quiver Plot Generation: To create a quiver plot, a subset of the data is selected based on defined latitude and longitude ranges. In this case, the region of interest spans from column number 120 to 140 and from row number 120 to 140. The cartopy library is employed to project the geographical coordinates onto a map

3) Quiver Plot Visualization: The resulting quiver plot displays vectors representing wind direction and magnitude over the specified geographical region. The arrows indicate the wind's zonal (east-west) and meridional (north-south) components. The plot is enhanced with features such as coastlines, borders, and land/ocean coloring for better context.

4) Generation of Animated Quiver Plot: To enhance the presentation, a gif animation is created from a collection of quiver plot images. Each frame in the animation corresponds to a different dataset taken on consecutive days, offering a dynamic view of how the wind vectors evolve.

5) Inference: Throughout the progression of the GIF, a noticeable trend emerged as the arrows representing wind vectors became progressively more horizontal. This pattern suggests a tendency toward a consistent direction or alignment of the wind over the specified geographical region during the selected time period. The increasing horizontal orientation of the arrows implies a shift or stabilization in the predominant wind direction over the five chosen dates from November 19th to November 23rd, 2010. Such trends in the quiver plot can provide valuable insights into the prevailing wind patterns and their potential impact on the local climate during the observed time span.

6) Conclusion: This data visualization exercise provides valuable insights into the spatial distribution and dynamics of wind vectors over a specific region. The quiver plot and animated gif offer a clear representation of the dataset, aiding in the interpretation of wind patterns and trends.

IV. Author Contributions

- Adithya Nagaraja Somasalle
  InfoVis - TreeMap
  SciVis - Contour Plot
- Pannaga Bhat
  InfoVis - Parallel Coordinate Plot
  SciVis - Color Map
- Goutham U R
  InfoVis - Node Link
  SciVis - Quiver Plot

References

[1] Our World in Data - Coronavirus Pandemic, https://github.com/owid/covid-19-data/tree/master/public/data

[2] LAS - ASMR2 OCean Database, https://las.incois.gov.in/las/UI.html

[3] LAS - Oscat Wind Dataset, https://las.incois.gov.in/las/UI.html

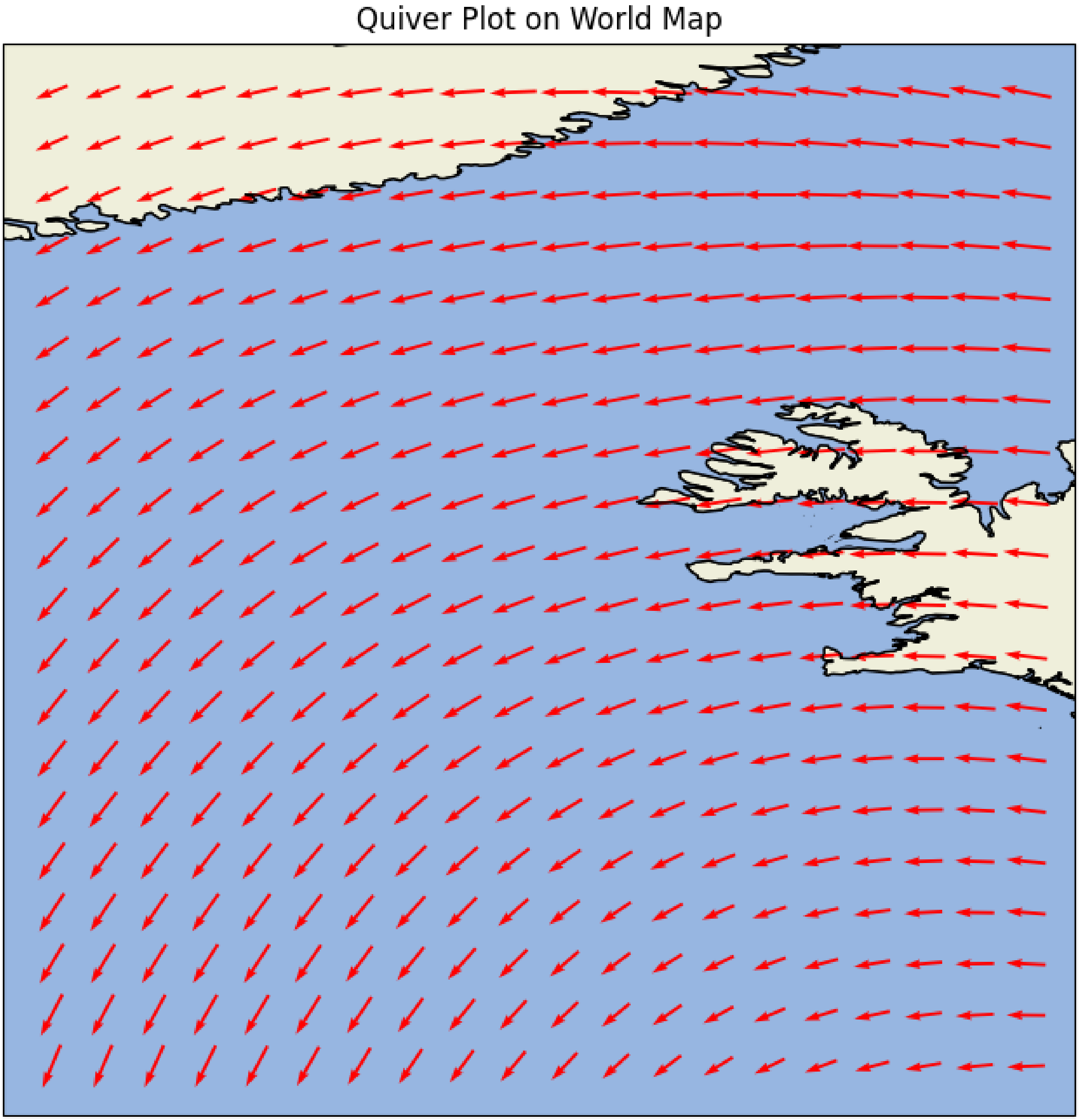

Fig. 29: Quiver plot for November 20 2010

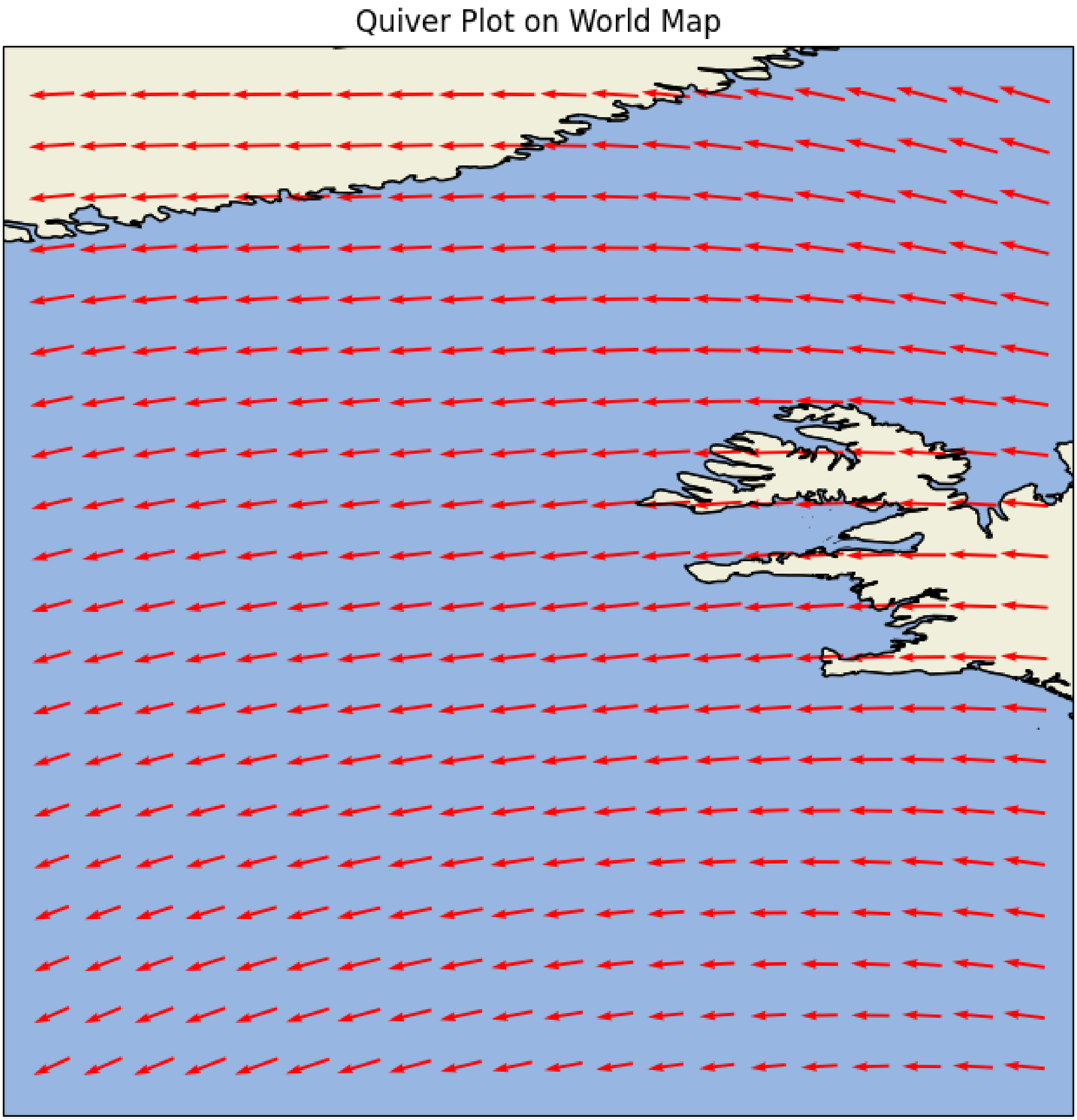

Fig. 30: Quiver plot for November 23 2010