Deforestation in the Amazon

Luke Abbatessa* Jenny Cai† Jocelyn Ju‡ Varun McIntyre§

ABSTRACT

Overall, our visualization shows the continual deforestation in the Amazon Rainforest over time. The visualization will support two main domain tasks. The first is to see if deforestation rates are increasing. The user seeing the steepness of the slope is important for understanding the urgency of the situation. The second is to see how much of the rainforest is left over time. This conveys the change in the amount of trees remaining in a way that resembles the change in real life. This is more impactful for the user because they can see a representation of the trees disappearing.

1 INTRODUCTION

Our visualization at a high level illustrates the extent to which the Amazon Rainforest has been experiencing deforestation over many years. The visualization consists of three different charts, that support two main domain tasks, to convey deforestation in different ways. The first domain task is to see if deforestation rates are increasing over time. An increasing rate signals that the situation is becoming more severe, and is important for the user to understand the urgency of said situation. This is visualized by a line chart that shows the increase in deforestation area over a range of years. The chart highlights over the year that is selected by a slider.

The second domain task is to see how much of the rainforest remains over time. This is important to show the change in remaining trees in a way that resembles the removal of trees in real life. This is more impactful and can elicit stronger emotions in the user because they see a representation of the trees actually disappearing. This is shown using a box that contains points to represent the trees that are left, and a meter that shows the percentage of deforested area vs. remaining area.

The end users for the visualization are Brazilian politicians and the general public. The politicians have the power to take action to prevent further deforestation, and the public has the power to push changes forward through protests and activism. The data we will use to visualize consists of the year between 2000 and 2021, county ID, total area for each county, and deforested area for each county.

2 RELATED WORK

- Changing Amazon deforestation patterns: urgent need to restore command and control policies and market interventions [3] - Ralph Trancoso (Apr. 2021)

  Summary: In this paper, Trancoso analyzes how the difference in government response has influenced deforestation levels over the years and explores the respective data. By doing so, his main focus is to advocate for increased implementation of stronger policies and intervention methods to reduce the amount of deforestation occurring in the Amazon rainforest and to maintain a sustainable environment. Trancoso emphasizes this point by using data and visualizations to show how deforestation has significantly increased (by about 61%) within the last 10 or so years.

  Use in Project: In the article, one of the visualizations Trancoso uses superimposes layers. More specifically, Visualization B is a combination of a line and a bar graph. Trancoso makes use of the color channel (luminance) to distinguish the two graphs. To extend this visualization technique, we were thinking of making use of the shape channel. Rather than using standard bars for the bar graph, we plan to use a log shape to make the message of our visualization clearer. We also plan to use the color channel (hue) so that viewers can better distinguish the two graphs and the log image. In the line graphs Trancoso uses, we will be using a line graph as one of our visualizations, but we decided to implement a slider. This slider would link our visualizations together to allow the viewer to focus on the corresponding data for a specific year.

- Spatiotemporal analysis of deforestation patterns and drivers reveals emergent threats to tropical forest landscapes [1] - Johanness Jamaludin, Jose Don T De Alban, L Roman Carrasco, Edward L Webb (May 2022)

  Summary: In this article, the authors apply Emergent Threat Analysis to identify emerging areas of deforestation in Southeast Asia, the main reasons for it occurring, and specific areas that should be protected more. It was concluded that palm oil and rubber plantations played significant roles in Southeast Asia's deforestation levels and relevant measures should be taken.

  Use in Project: One of the most commonly used visualizations in this article is a map, which is used in Figure 3. For our use case, though we do not plan on implementing this type of visualization, we do want to extend some aspects found in this figure. For example, we would use a similar visualization (the second one) that portrays the amount/level of deforestation that occurred, but we would make more use of the point mark and color (hue) channel. The fewer points that are in this visualization, the more deforestation has occurred. Another example is that one of the visualization techniques used displays more information about specific places on the map and allows viewers to focus on and learn more about that region. We want to extend this technique by implementing a tooltip in one of our visualizations (the first one) that would let users view relevant but more specific deforestation data about any year they choose. Figure 3 also uses an arrow to show which area the specific information correlates to. Rather than using an arrow, we plan to highlight the point (year) of focus and have the tooltip reasonably close to the point.

3 USE CASE

Our visualization consists of three charts that show the dramatic increase in deforestation in the Amazon Rainforest over the years. This is meant to elicit a strong, emotional reaction from the viewer, because it makes it easier to understand the extent to which a well-known rainforest like the Amazon is being destroyed.

This visualization can be used to help persuade Brazilian politicians to take more aggressive action to preserve the Amazon. While they are most likely aware that deforestation is occurring, the visualization can show the extreme extent to which it is happening. If changes

*E-mail: abbatessa.l@northeastern.edu
†E-mail: cai.je@northeastern.edu
‡E-mail: ju.joc@northeastern.edu
§E-mail: mcintyre.v@northeastern.edu

aren't made, deforestation will continue to increase. The visualization can also be presented to the public to help increase awareness of the situation, so that changes may be made due to activism and support from the public.

4 Data

The data [2] comes from someone by the name of Diego Silva. Diego collected the data from a data lake created by the Brazilian National Institute for Space Research (INPE) on the BigQuery. The original dataset can be found at the following link: https://www.kaggle.com/datasets/diegosilvadefrana/brazilian-deforestation-from-2000-to-2021?resource=download&select=data.csv.

In terms of biases or ethical considerations for our data, since the counties/municipalities represented in the dataset are based on natural and/or political boundaries, there is inherent bias when it comes to the size of a certain area of land. It will evidently be the case that the larger a county/municipality is, the larger a certain area of land found within the county/municipality will be. As a result, we have decided to create new attributes representing the proportions of certain areas of land with regards to the total area a county/municipality encompasses; this will take into account the discrepancies in the sizes of counties/municipalities.

In terms of data cleaning, we utilized Python for our efforts. We started by reading in the .csv file representative of the dataset into a Pandas DataFrame and observing the dimensions of the resulting dataframe (the dataframe had 16,720 rows and 10 columns). We then renamed each of the column names because the original column names were in Spanish. Next, we dropped an attribute from the dataframe titled "Incremented Deforested Area" that contained all the missing values, as the attribute didn't contribute to the "Total Area" attribute in the dataset. After that, we filtered out negative instances from the "Non-Observed Area" attribute, as it didn't make sense for the size of an area of land to be negative. Then, we created six additional attributes representing the proportions of certain types of areas of land with regards to the total areas of counties/municipalities; as was touched on previously, creating these proportions eliminates the bias present from the discrepancies in the sizes of counties/municipalities. Specifically, the attributes we created represent the proportion of area deforested, the proportion of forest area, the proportion of area covered by clouds, the proportion of non-observed area, the proportion of non-forest area, and the proportion of hydrography area, all with respect to the total area of the county/municipality of interest. Lastly, since the original dataset consisted of 16,720 rows, we decided to take a random subset of 2,000 rows to allow for the dataset to be stored in the browser. Once we did so, we confirmed that the resulting dataframe had 2,000 rows, and we confirmed each year included in the original dataset was present in the resulting dataframe.

5 Design Process

5.1 Sketch 1

The initial sketch displays a very rough outline of a line chart above a rectangle and a pie chart. This was our initial idea for a layout to keep everything to one screen. Though it changed slightly from sketch to sketch, this is generally the final layout that can be seen on the last sketch.

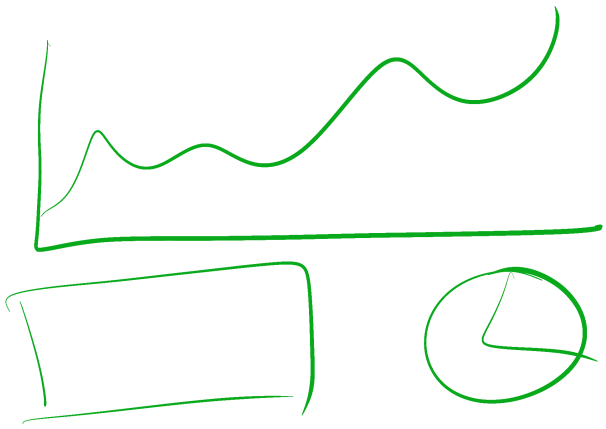

Figure 1: Initial Sketch

5.2 Sketch 2

The second sketch is more detailed, building off of the first. It depicts a "years" slider, which would link all of the visualizations and allow for interaction. The line chart was combined with a bar chart, shown through "logs" for visual appeal and hopefully for the user's immediate connection to deforestation. The rectangle transformed into a view of the remaining trees at the selected year in dots. We also considered having the forest view and the pie chart overlay each other, to be changed on button click, though we ultimately decided against it to focus on being more user-friendly. We did keep the forest visualization, logs, and years slider.

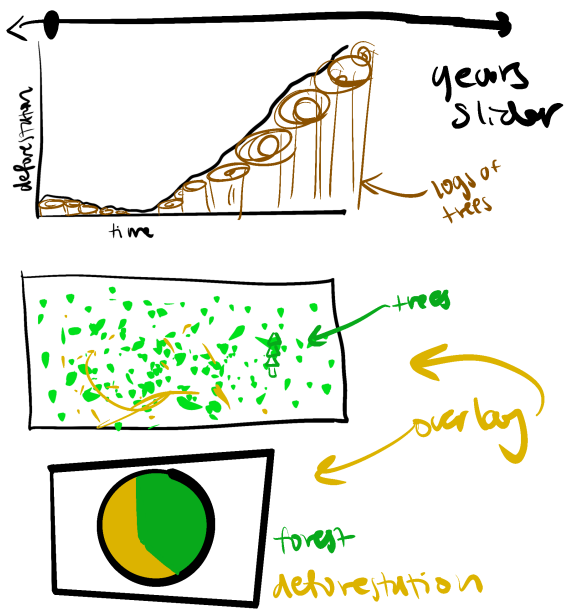

Figure 2: Sketch 2

5.3 Sketch 3

The third sketch evolved from the second sketch with a change in position of the years slider. By having it in the middle of the visualizations, we created a centralized point of interaction, allowing the user to see the changes at once. These changes corresponded with the years slider, showing the tree distribution and deforestation in that specific year, as well as highlighting that year on the line

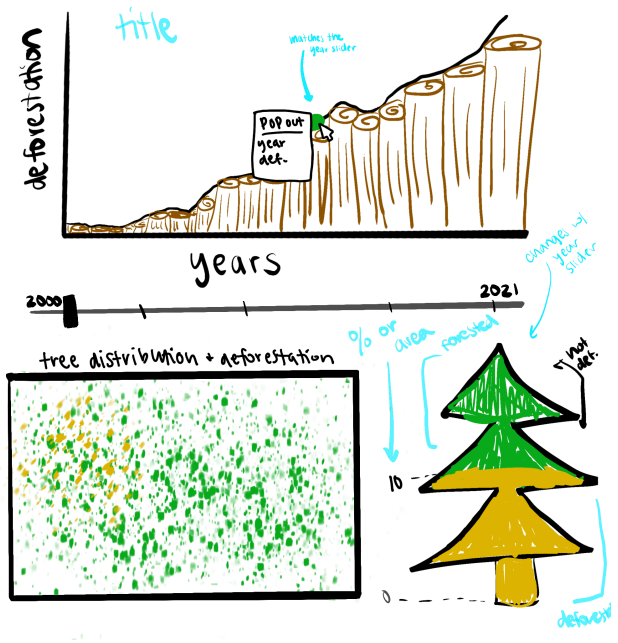

Figure 3: Sketch 3

chart. We also added a mouseover function for the line chart where a popup will appear on hover, detailing the year and deforestation statistics. Finally, we transformed the pie chart into a tree-chart to engage the user. For the final sketch, we kept the layout, tree, and mouseover event.

5.4 Sketch 4

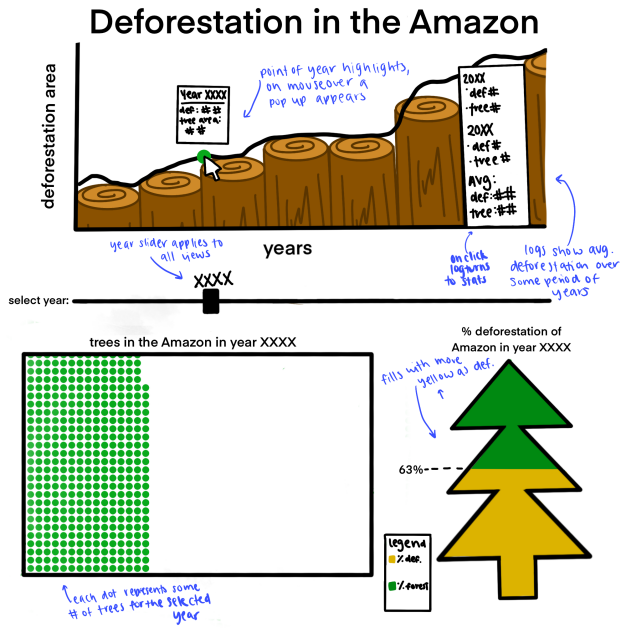

Figure 4: Final Sketch

The final sketch is a polished culmination of our previous efforts. In it, there is a title, as well as the line/bar chart, slider, forest distribution, and tree chart from the previous sketches. We added a title, cleaned up the visualizations, added a legend to the tree chart, and determined that the log-bars in the line chart would represent the average deforestation over a fixed number of years, to be determined once the data is further analyzed. These bars are also interactive– on click, they reveal a popup with the information of the years included in the average, as well as various statistics (deforestation, forest amount, etc.). This popup is intended to replace the log until clicked again. The years slider connects all three of the visualizations, highlighting a green point on the line chart at that year, as well as displaying the proper statistics for the other two. This sketch features blue comments and arrows of our ideas, and is what we will be aiming for when pursuing the completion of this project. The marks and channels displayed in the final sketch are as follows:

- Line/Bar Chart : (marks) lines, areas and (channels) position, area
- Tree Count: (mark) points and (channel) position
- Tree Deforestation Chart: (mark) areas and (channel) color

6 Final Design

FINAL DESIGN HERE

7 Discussion

DISCUSSION HERE

8 Conclusion

CONCLUSION HERE

References

[1] L. R. C. E. L. W. Johanness Jamaludin, Jose Don T De Alban. iopscience.iop.org: Spatiotemporal analysis of deforestation patterns and drivers reveals emergent threats to tropical forest landscapes. *Environmental Research Letters*, 17(5):054046, May 2022. doi: 10.1088/1748-9326/ac68fa

[2] D. Silva. Amazônia rainforest deforestation 2000 - 2021.

[3] R. Trancoso. iopscience.iop.org: Changing amazon deforestation patterns: urgent need to restore command and control policies and market interventions. *Environmental Research Letters*, 16(4):041004, April 2021. doi: 10.1088/1748-9326/abee4c

A Data Abstraction

Each row of the data represents the areal characteristics of a unique county found within the Amazônia rainforest.

- ”Year” attribute : Ordered (Sequential) attribute.
- ”County ID” attribute : Categorical attribute.
- ”Total Area” attribute : Ordered (Quantitative) attribute.
- ”Area Deforested” attribute : Ordered (Quantitative) attribute.
- ”Forest Area” attribute : Ordered (Quantitative) attribute.
- ”Area Covered by Clouds” attribute : Ordered (Quantitative) attribute.
- ”Non-Observed Area” attribute : Ordered (Quantitative) attribute.
- ”Non-Forest Area” attribute : Ordered (Quantitative) attribute.
- ”Hydrography Area” attribute : Ordered (Quantitative) attribute.
- ”Proportion Area Deforested” attribute : Ordered (Quantitative) attribute.
- ”Proportion Forest Area” attribute : Ordered (Quantitative) attribute.
- ”Proportion Area Covered by Clouds” attribute : Ordered (Quantitative) attribute.
- ”Proportion Non-Observed Area” attribute : Ordered (Quantitative) attribute.
- ”Proportion Non-Forest Area” attribute : Ordered (Quantitative) attribute.
- ”Proportion Hydrography Area” attribute : Ordered (Quantitative) attribute.

B Task Abstraction

The first domain task, to see if deforestation rates are increasing, corresponds to the visualization task to show how a statistic changes over time. This is best represented by a line chart, which shows a changing rate over time.

- High-level action: Consume - Present and Produce - Annotate. The line chart presents known information that deforestation is occurring, and shows the rate at which it is changing. It also annotates the selected year through a highlight.
- Mid-level action: Search - Lookup. The location and target are known.
- Low-level action: Query - Identify. The deforestation level is identified by the year selected by the user.
- Target: All Data - Trends. The line chart is meant to illustrate the increase in deforestation over the years.

The second domain task, to see how much of the rainforest is left over time corresponds to the visualization task to show a portion that remains of a whole. This is represented by a box filled with points that change in amount over time, and a meter that shows a percentage of a whole.

- High-level action: Consume - Present. Like the first domain task it communicates known information that deforestation occurs, but emphasizes the severity.
- Mid-level action: Search - Browse. The target is unknown but the location is known.
- Low-level action: Query - Summarize. All the information is important and there is nothing to identify or compare.
- Target: Attribute - Distribution (roughly fits). The visualizations show how an amount changes over time.