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 over a histogram that shows the increase in deforestation area over a range of years. The chart's points can be selected and will display the deforestation at that specific year as well as updating the other two visualizations.

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, acting as a meter that shows the percentage of deforested area vs. remaining area. A pie chart represents this data in a concise way, with percentages of the forest deforested or remaining appearing in the tooltip on hover.

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

In Ralph Trancoso's paper titled "Changing Amazon deforestation patterns: urgent need to restore command and control policies and market interventions" [3], 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. 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.

In the article titled "Spatiotemporal analysis of deforestation patterns and drivers reveals emergent threats to tropical forest landscapes" [1], 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. 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 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.

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

4 DATA

The data [2] comes from 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.

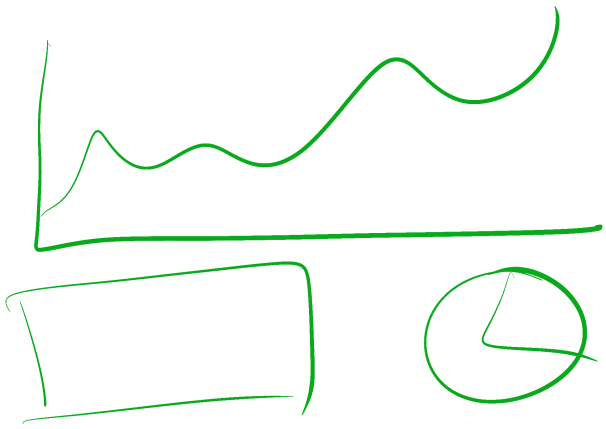

Figure 1: Initial Sketch

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.

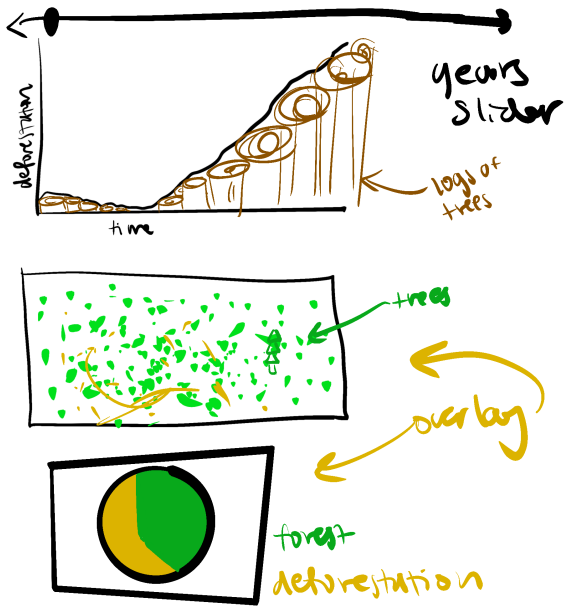

Figure 2: Sketch 2

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 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

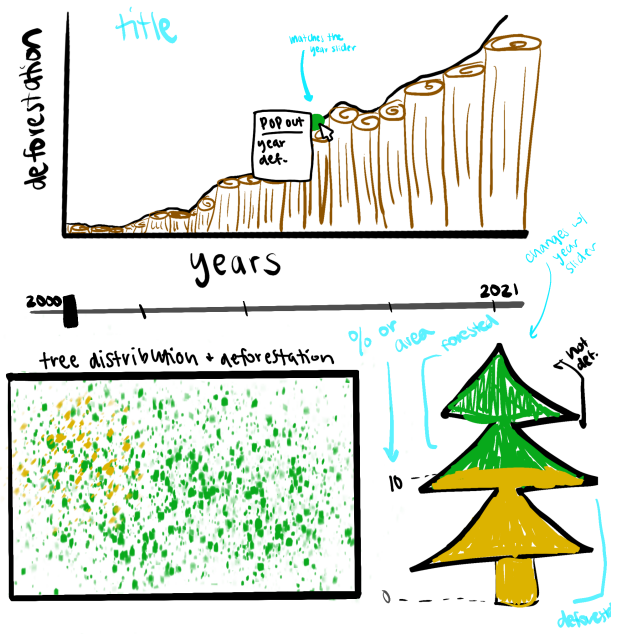

Figure 3: Sketch 3

engage the user. For the final sketch, we kept the layout, tree, and mouseover event.

5.4 Sketch 4

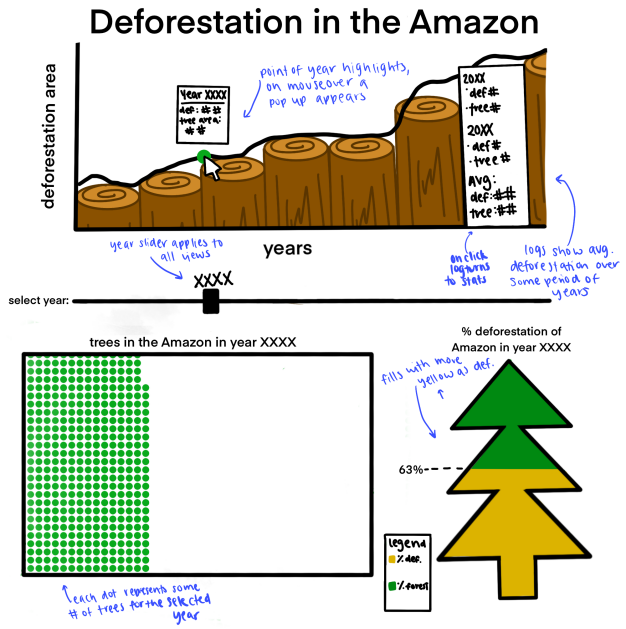

Figure 4: Final Sketch

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) points and lines, and (channels) position - both and color
- Tree Count: (mark) points, and (channel) color
- Tree Deforestation Chart: (mark) area, and (channels) color and tilt

5.5 Final Design Process

Through implementation, our visualization shifted from what we had initially hoped to accomplish. The first major change was the representation of the bars in the histogram. We had hoped to have tree trunks as the bars, but we ultimately decided to have the bars be brown instead, to suggest wood. Additionally, for the histogram tooltip, rather than covering the bar that was clicked on, we had the tooltip appear off to the side of the visualization. Likewise, for the point click, we had the tooltip appear to the other side. This made it easier to read and did not make the histogram too busy. The following were the biggest changes:

- tree ”meter” converted to pie chart
- removed slider
- added linking with points
- added hover tooltip for dot plot and pie chart
- aesthetics (green, different html pages)

5.6 Usability Testing

For our usability testing, we gave our volunteer user the following requests. The responses and corresponding response times (only steps 3-5) are in parentheses:

1. Explain the difference between what each bar represents and what point represents. (”Dots show the same data as the bars, but in more detail”)
2. Which visualization do you prefer to display similar information – dot plot or pie chart? (”Dot plot is better to look at, pie chart is easier to tell the exact numbers”)
3. Report the average Proportion of Area Deforested for 2009-2012 (0.105 - 8 seconds)
4. Report the average Proportion of Area Deforested for 2010 (0.105 - 2 seconds)
5. Report the approximate percentage of forest remaining based on the dot plot (86 percent - 3 seconds)

The user correctly identified the difference between the bars and points, and their explanation of which graph they preferred and why matched our intention for the difference between the two. They took a relatively longer time to identify the correct bar for 2009-2012 and estimate its height. It was much easier to identify the value of the single point because of the tooltip functionality. The dot plot estimate was also close, the true percentage was 88.5 percent and they guessed 86 percent. A potential way to make the histogram easier to read is to adjust the scale so that bar heights aren’t as close

together. The dot plot and pie chart were more successful and will be kept mostly as is.

6 Final Design

The figure below is an image of our final visualization tool.

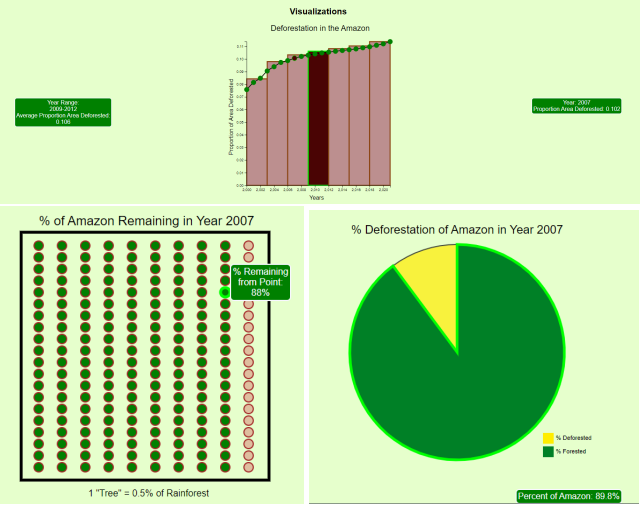

Figure 5: Final Visualization

For our final visualization tool, we created two visual encodings. The first encoding is a histogram overlaid by a line graph that displays the proportion of the area deforested in the Amazon (Y-axis) over the years 2000 to 2021 (X-axis). This encoding can be used to determine the trend in deforestation over time. For the line graph, points and line marks were used, in addition to position and color channels. For the histogram, only the line mark was used, but position and color channels were used similar to the line graph. For both of these graphs, users can hover over a point or a bar, which would create a green border around the item the cursor is on. If a bar is clicked on, its border and fill color will change. A tooltip that displays specific data about the year will also appear on the left. Similarly, if a point is clicked on, its border and fill color will change too but the tooltip appears on the right. The year of the point selected is what links both of our visual encodings together. The data presented in the second encoding will change depending on what year is selected. This encoding is composed of a dot plot and a pie chart, both of which can help users determine how much of the Amazon remains over time. The dot plot displays the percent of the Amazon that is left in a specific year. It is composed of green and brown circles, where the former color represents the percent remaining while the latter color represents the percent deforested. Each circle is supposed to represent a tree that is equivalent to 0.5% of the entire rainforest. Hovering over each circle will create a border around it and a tooltip that displays the percentages. The pie chart displays the percent of the Amazon that is deforested in a specific year, where the green section represents the percent forested and the yellow section represents the percent deforested. Similar to the dot plot, hovering over each section will create a green border around it and displays the percentage as a tooltip below the chart.

One group of target users from our use case is Brazilian politicians. From the home screen of our webpage, they would navigate to the "Visualizations" section in the navigation bar to see our visualization tool. The titles on each page and the bird above each section of the navigation bar can help identify which page they are currently on. The first visual encoding they see is a histogram overlaid by a line graph that presents the overall trend in the proportion of the Amazon that is deforested (Y-axis) from 2000 to 2021 (X-axis). If their mouse hovers over either a point or bar, a border will generate around it, which helps them to determine which point or bar their cursor is on. The user clicks on any point of the line graph or any bar of the bar graph, which will cause its fill color to change to indicate it was selected. A tooltip will also appear on the screen (on the left for a bar selected or on the right for a point selected) that displays more specific statistics about the item(s) selected. Based on the graphs and specific statistics shown in this encoding, they should see that there is an upwards trend in deforestation as time progresses. After clicking on a point on the line graph, the user will then look at the second visual encoding which consists of two visualizations, one similar to a scatterplot and one that is a pie chart. Each of the visualizations within this encoding will update according to the year of the point selected. Each circle of the left side's visualization will either fill in green to represent the proportion of the Amazon remaining or brown to represent the percent that was deforested for that year. The sections of the pie chart will also change accordingly. Like the first encoding, hovering over each point on the scatterplot or section of the pie chart will generate a border and a tooltip that presents more information. From this encoding, the user should be able to compare how much of the Amazon remains to the percent deforested within the same year. After exploring both encodings, Brazilian politicians may see more clearly the rising trends in Amazon deforestation as time progresses as well as how much of the Amazon is left. With these details in mind, they can get a better sense of the urgency of the situation, which would then help them decide what types of actions to take and laws to implement to protect the Amazon.

7 Discussion

Our final tool somewhat accomplishes the task that we set out to address. We have a functioning, linked tool that displays the deforestation rates of the Amazon, and is fairly user-friendly, as it is meant to be for politicians and the public. They may have to watch the demo video first, though. Something that is missing is another layer of depth. We display the deforestation over time, and while that was the intention of our project, it would have been interesting to incorporate the regrowth efforts for the Amazon at the same time, or to find data about the public's interest in the forest and deforestation in general. It would also have been interesting, and this would likely be a next step if we were to continue with the project, to incorporate the state of the climate in comparison to the diminishing forest. This would allow the user to more fully grasp the severity of the situation. Seeing numbers may not be enough to make the user care. However, viewing the issue in a broader scope and application may help further the awareness and understanding of the situation, especially as climate change becomes an increasingly pressing issue. Other improvements to the tool might include introducing more images of the Amazon, including a map, and adding other interactions to our visualizations to provide more information, such as being able to identify a specific part of the Amazon to focus on. Additionally, providing another visualization regarding carbon emissions in a more digestible way, such as how many trees it takes to charge your phone, might help the user understand the true impact of the deforestation.

8 Conclusion

This project represented and visualized the continual deforestation in the Amazon Rainforest over time in an effort to enhance the understanding of the general public. By making the tool user-friendly, our hopes are that Brazilian politicians in favor of saving the Amazon would be able to use it to increase general awareness and support for the cause.

The tool, visualizations of deforestation rates, supports two main domain tasks: to see if the rates are increasing, and to view how much of the rainforest is left over time. The user can see the steepness of the slope of the line in the histogram/bar visualization to understand

the urgency of the situation, and the dot plot and pie chart convey 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, and being replaced with ”stumps.” The goal is for the user to understand the severity of the situation and to encourage them to consider how they can help.

Author Contributions:

- Luke Abbatessa: Data, Data Abstraction
- Jenny Cai: Related Work, Final Design
- Jocelyn Ju: Design Process, Discussion, Conclusion, Editing
- Varun McIntyre: Abstract, Introduction, Use Case, Task Abstraction, Usability Testing

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.