Visualizing Lemur Survival Data

Project Proposal

Anna Flagg
annaflagg@gmail.com

1. Topic

Madagascar is home to over 100 species of lemur, many of which are uniquely found there. Besides the important role they play in the local ecosystem, lemurs are also a main food source (mostly through illegal hunting) for the Malagasy people, two-thirds of whom live below the poverty line. Due to unregulated deforestation, it is projected that over half of the lemur population has died off in the last 50 years, leaving many species on the brink of extinction.

The goal of this project is to visualize the effects of deforestation on lemur populations in Madagascar in the hopes of sending a clear message to policy makers about the consequences of further deforestation. The visualization will incorporate what known data exists about both the forest layout in Madagascar and the status of lemur populations over the last few decades. If time permits, a species/area model will also be implemented to predict and visualize the effects of potential future forest cuts.

2. Personal Experience

Under the joint supervision of an anthropology professor and a computer science lecturer at the University of Toronto, I led a group research project in this area for a few semesters of my undergrad. While I am not an expert, I have had some experience with the particular challenges of analyzing and visualizing this lemur data.

My previous work with other students on this problem has illuminated some pitfalls and helped refine projects goals. I have kept up to date on recent developments involving the lemur situation in Madagascar, as well as current efforts to address it, which should also inform my progress this semester.

3. Proposed Solution

This project has two messages to communicate: 1) the lemur population trend, and 2) the importance of Madagascar's remaining forest habitats. The first involves population numbers; it should support quantitative analysis, such as an anthropologist may wish to do. The second is about geography; it should communicate a story about the landscape changes humans have made in the region and the effects on lemur species, such as could be used to demonstrate the high level concept to policy makers. These two goals are best supported in quite different ways; the proposed solution therefore will address each with its own view.

3.1. **Population numbers view.** The first view will be a traces scatterplot visualization as in Robertson, et. al [1], plotting the following three dimensions: amount of forest (x axis position), population (y axis position), time (opacity traces). By default, it might make sense to aggregate the 101 lemur species into their 5 families so the traces can be effectively color-coded. Then interactive selection of individual species/families to show on the plot could also be implemented.

Possible hiccups: Occlusion/clutter will be possible if the viewer chooses to show all species individually without aggregating - it will be her responsibility to understand that. However this would likely still demonstrate the general trend, so still be useful. Another problem would be if it is not possible to obtain the required time-series data for this view. In that case, the project will focus instead on the second part - the geographic view of the effects of deforestation.

3.2. **Geographic view.** This choropleth view will color-code locations according to 1) a (possibly weighted) measure of local lemur density, and 2) a (possibly weighted) measure of the risk of the local lemur populations. The density measure is quantitative, and would be represented by lightness. The risk measure is categorical, and would be represented by ordered hues: red (Critically endangered), orange (endangered), yellow (near threatened), green (vulnerable), blue (least concern).

The use of color to demonstrate these types of relationships must be thought through carefully. In this case, such an abstraction might be effective because although the density is quantitative, it is not the goal to support fine-grained tasks such as calculating the difference in value between two points, etc. The goal is to represent an irregularly-shaped region in terms of its value to the conservation of lemur species, contextualized in the big-picture view of the map data. For our purposes, a region should stand out if a) there are lots of lemurs living there, and b) one or more individual lemur species is highly dependent on this region. This is the reasoning behind the choice color-coding of of density and risk measures.

One possible characterization of these measures might be based on the region data for different lemur species, data which does appear to be readily available in shapefiles. In this case, we could decide that *density* is how many lemurs of any species are thought to live in this location, and *risk* is the combined categorical threat of extinction, where the risk of each species is weighted by the percentage of the species' population that lives there.

There are several other possibilities. The early part of the project will concentrate on wrangling a good choice of measures from the available data.

Possible hiccups: There might be difficulties treating overlapping regions if the shapefile process considers distinct polygons, without providing access to points within overlapping polygons. In this case, regions could be coded with opacity so that overlapping areas will "add up" to stand out, and polygon drawing will go in sequence from region of lowest risk to region of highest risk, so that a point in more than one region will reflect the hue of the region of highest risk. Another problem could be lack of region data for all species. However, region data is available for at least several species, so in this case, the visualization will show as many species as data exists for, with the expectation that more species data can be easily integrated as it becomes available.

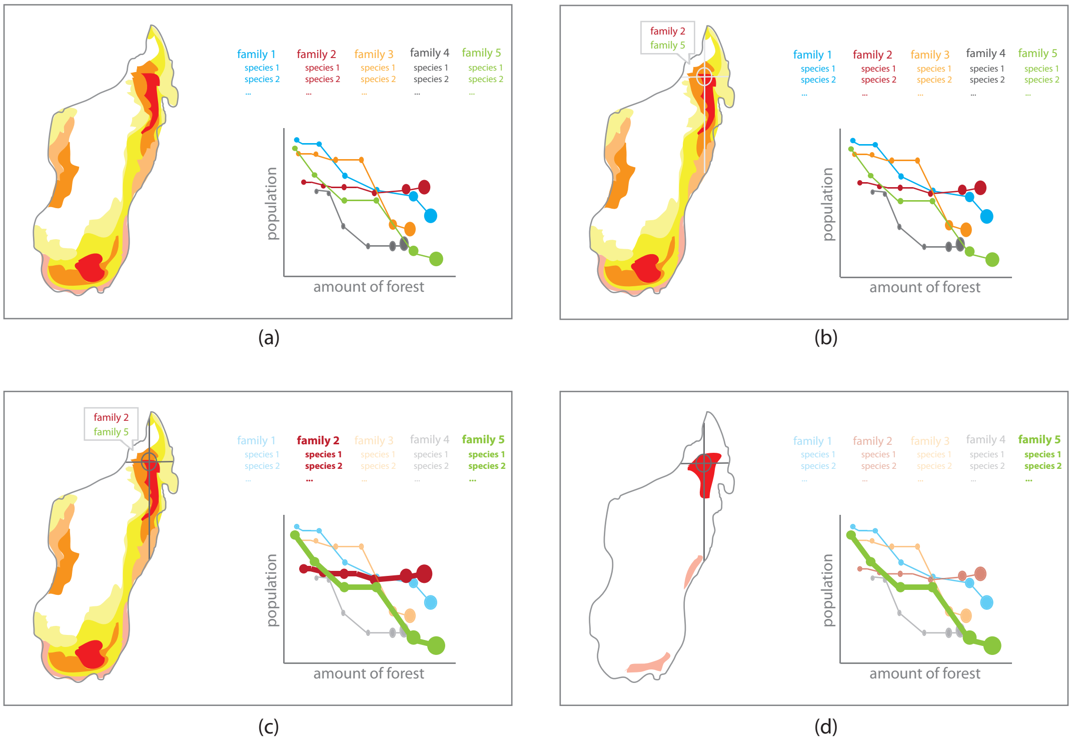

FIGURE 1. Sample scenario of use: (a) view habitat distribution, (b) find location of interest, (c) select location of interest, (d) select lemur family of interest.

4. SCENARIO OF USE

A policy maker is debating placing a restriction on deforestation in a certain area of northern Madagascar (Figure 1). She is considering the economic restrictions such a move would place on local industry, versus the environmental tole of doing nothing. She opens the visualization (Figure 1a), and finds the relevant location on the map (Figure 1b). She sees by its dark orange color that cutting trees in this area poses a medium/high environmental risk. Wishing to know more details about the particular family of lemurs that will be affected, she selects her location, and the population scatterplot highlights the lemur families native to her selection (Figure 1c). One of the families seems to have suffered a lot of population loss recently; she selects this lemur family, and the map view updates to show data for just this family (Figure 1d). It is clear that deforestation in her selected area would likely have very bad consequences for the chosen family, which has a small percentage of its population in other regions, but depends by far the most on the region under consideration.

5. Proposed Implementation

Processing will be used to develop this visualization, which will be in the form of a Java applet or application. Data will most likely be in the form of shapefiles, which Processing can draw with the help of OpenJump. Other data will likely be in .csv or other easily-readable formats.

6. Milestones

This project will be broken down into the following steps:

Data collection: now-November 3

The first step is to collect region data for different lemur species in the form of shapefiles. This process is currently underway with the help of ReBioMa, anthropologists from the Durrell Wildlife Conservation Trust, and online database *All the World's Primates*. It will be possibly ongoing, but a workable amount of data will be in hand by November 3.

Framework of application with region visualization: November 3-10

As soon as data for at least one or a few lemur species has been collected, a basic framework for the application can be mocked up. This will involve figuring out how to read and render shapefiles on a map for different lemur species in Processing.

Region/point analysis for choropleth coding: November 10-17

Next it will be necessary to work out the coding scheme. Specifically, the density and risk measures must be formalized and calculated from the available data.

Traces scatterplot visualization (if data exists): November 17-28

If population data has successfully been collected, then the traces scatterplot view can be implemented next.

Interaction implementation: November 28-December 5

Functionality for user selection of regions, families and species can then be implemented.

Paper writing/presentation preparation

The final week will be devoted to writing the paper and preparing for the final project presentation.

*In the unlikely event that there is extra time towards the end of the project, it might be possible to implement a predictive model for showing the effects of viewer-specified forest cuts.

7. Previous Work

The background reading for this project is still in progress, but a few important references have already helped shape our direction. Dykes and Brunsdon [2] described some ways to show choropleths at different scales using linked view and a varying weighting parameter. While it is not the direct goal of the lemur project to show different scales, the spatial weighting does illuminate different ways the map data could be interpreted. This paper also provides some ideas on how data can be shown spatially on irregular regions rather than as points.

A major goal of this project is to show the link between forest loss and lemur population loss, but time is also important for communicating urgency. Robertson, et. al [1] was a valuable resource to clarify the strengths of animation, traces and small multiples to visualize 3D data, and in what situation each is appropriate. This influenced the choice of the traces scatterplot for the population view.

This project will likely combine data from several sources, including the online database *All the World's Primates*. Besides providing data, this site also contains a map visualization that overlays region outlines for a selected primate. This tool, as well as the data from ReBioMa [4], helped suggest how shapefiles could be used for species range visualization.

8. References

[1] Robertston, George, Fernandez, Rland, Fisher, Danyel, Lee, Bongshin, and Stasko, John. Effectiveness of Animation in Trend Visualization. IEEE Trans. on Visualization and Computer Graphics 14(6):1325-1332, 2008.

[2] Dykes, J. and Brunsdon, C. Geographically Weighted Visualization - Interactive Graphics for Scale-Varying Exploratory Analysis, IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization / Information Visualization 2007) 13(6): 1161-1168, 2007.

[3] Noel Rowe, Marc Myers, eds. All the Worlds Primates, www.alltheworldsprimates.org. Primate Conservation Inc., Charlestown RI. 26 Oct. 2011. *http://alltheworldsprimates.org/*

[4] ReBioMa. ReBioMa WebPortal. 2001. *http://www.rebioma.net/*