

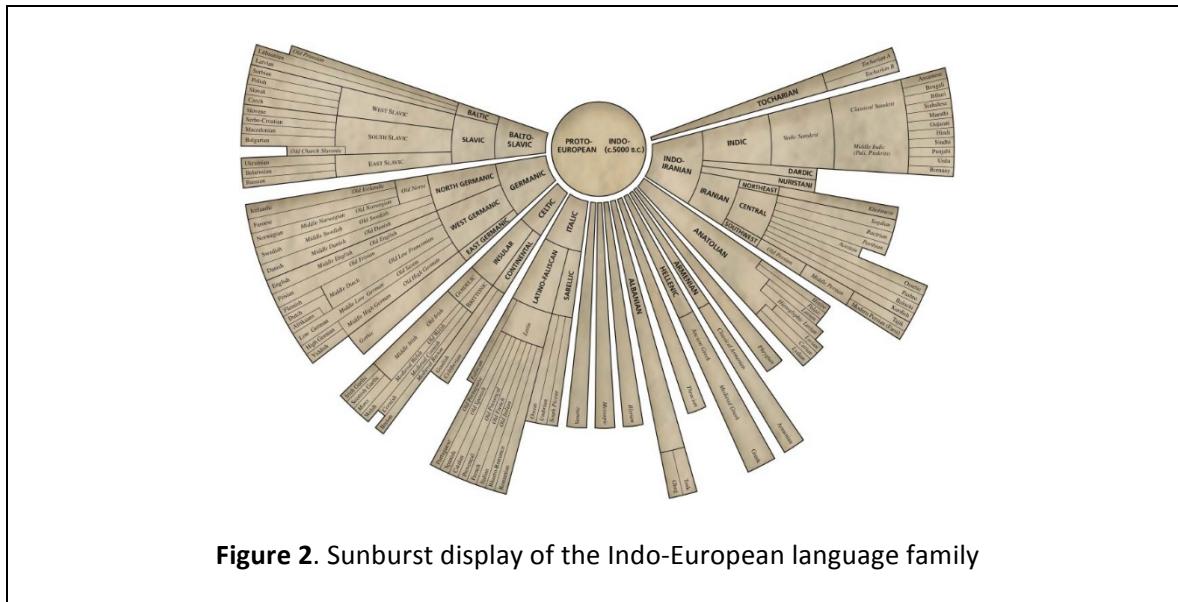
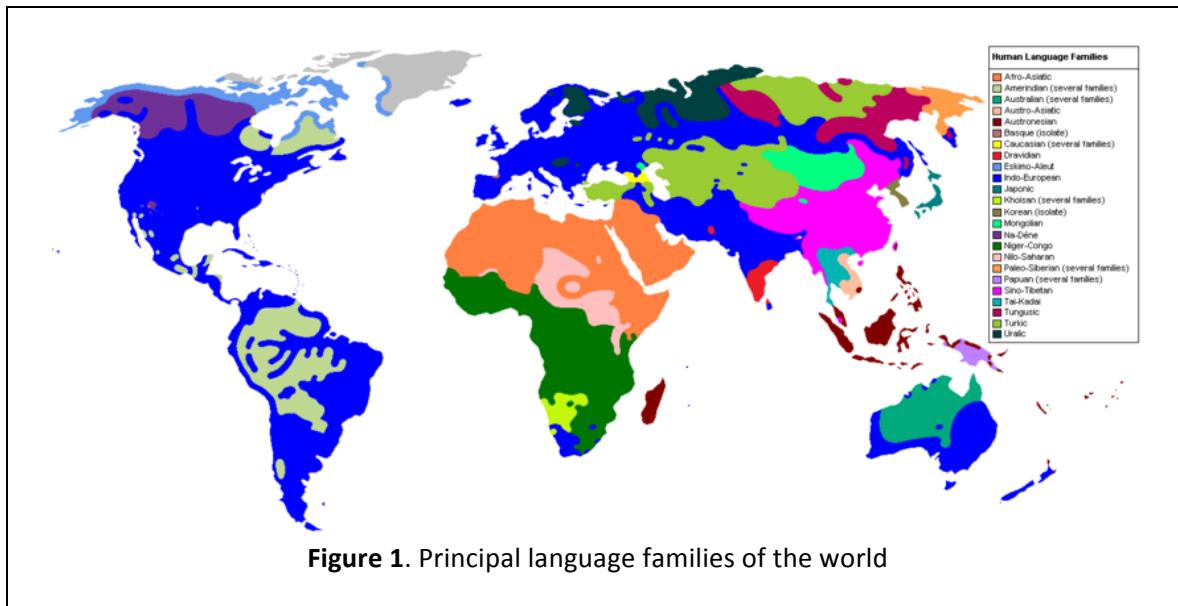
# Visual Language Families

Alexander Conrad Nied

Congle Zhang

## Introduction

A language family is a group of languages related through descent from a common ancestor, called the proto-language of that family. There is a tree structure for about 5000-8000 living languages. Figure 1 is a static graph showing the principle language families of the worlds. Figure 2 shows the language family of Indo-European, which is in a tree structure.



The static graph has several disadvantages:

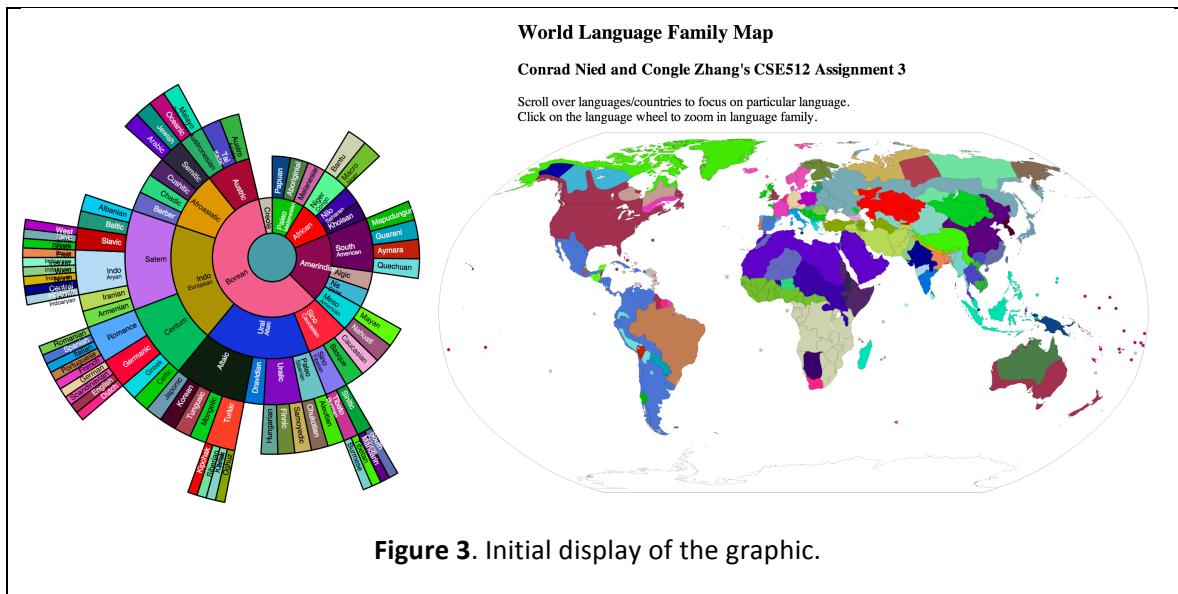
1. There could be thousands of languages in the hierarchy. Users would love greater granularity and deeper hierarchy. However, putting all the information in a single static graph would overwhelm the user. For example, Figure 1 only shows the principle languages. English, French, German, Spanish and even Iranian are all put together as "Indo European". It would be necessary to obtain the sub-hierarchy of Indo-European, otherwise users might confuse why English and Iranian are put together.
2. The relationship between languages are not clear. For example, most people would assume Chinese, Korean and Japanese are close to each other; Romanian and Bulgarians are speaking similar language. Both of them are false. Such information is hard to visualize in a static figure. Actually Korean and Japanese are close but not close to Chinese. Romanian is close to French while Bulgarian is a kind of Slavic.

In this project we try to visualize the language tree with greater granularity and deeper hierarchy. At the same, we must ensure the readability of the graph. With our visualization, readers can quickly understand the big picture of the language family, and also quickly zoom in an interesting language (e.g. English or Romanian), and understand its position in the language family.

## Layout

The layout of our visualization is composed of two parts:

1. A wheel showing the hierarchy of select languages and language families;
2. A world map showing the approximate geographic extent of each language group.



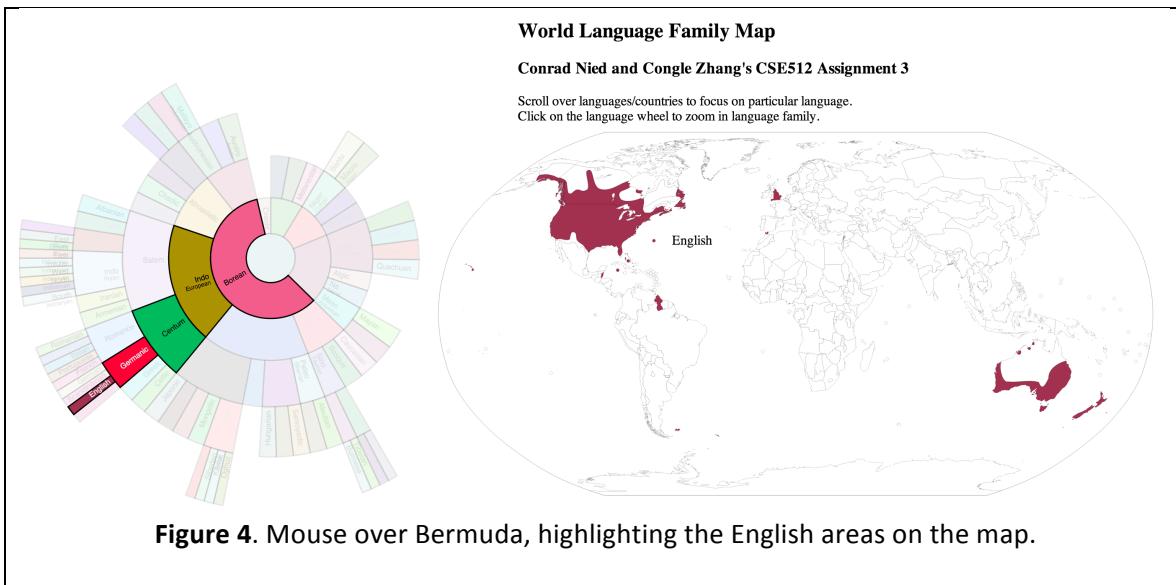
It is noteworthy that each leaf node of the wheel corresponds to one area in the map; the node and the area are sharing the same color. The inside nodes of the language family (e.g. Germanic)

are given other colors. Colors are generated by: taking the string for each language group, hashing it, and binning characters in the hash to the Red, Green, and Blue values. During the design process we thought of many ways we could optimize the color display (binning by hues for instance or limiting the amount of languages shown on the screen), but through technical & temporal limitations we did not pursue a more advanced approach. Through the interaction techniques though, users will be able to focus on a smaller number of linguistic units/colors so issues of having multiple colors should be alleviated.

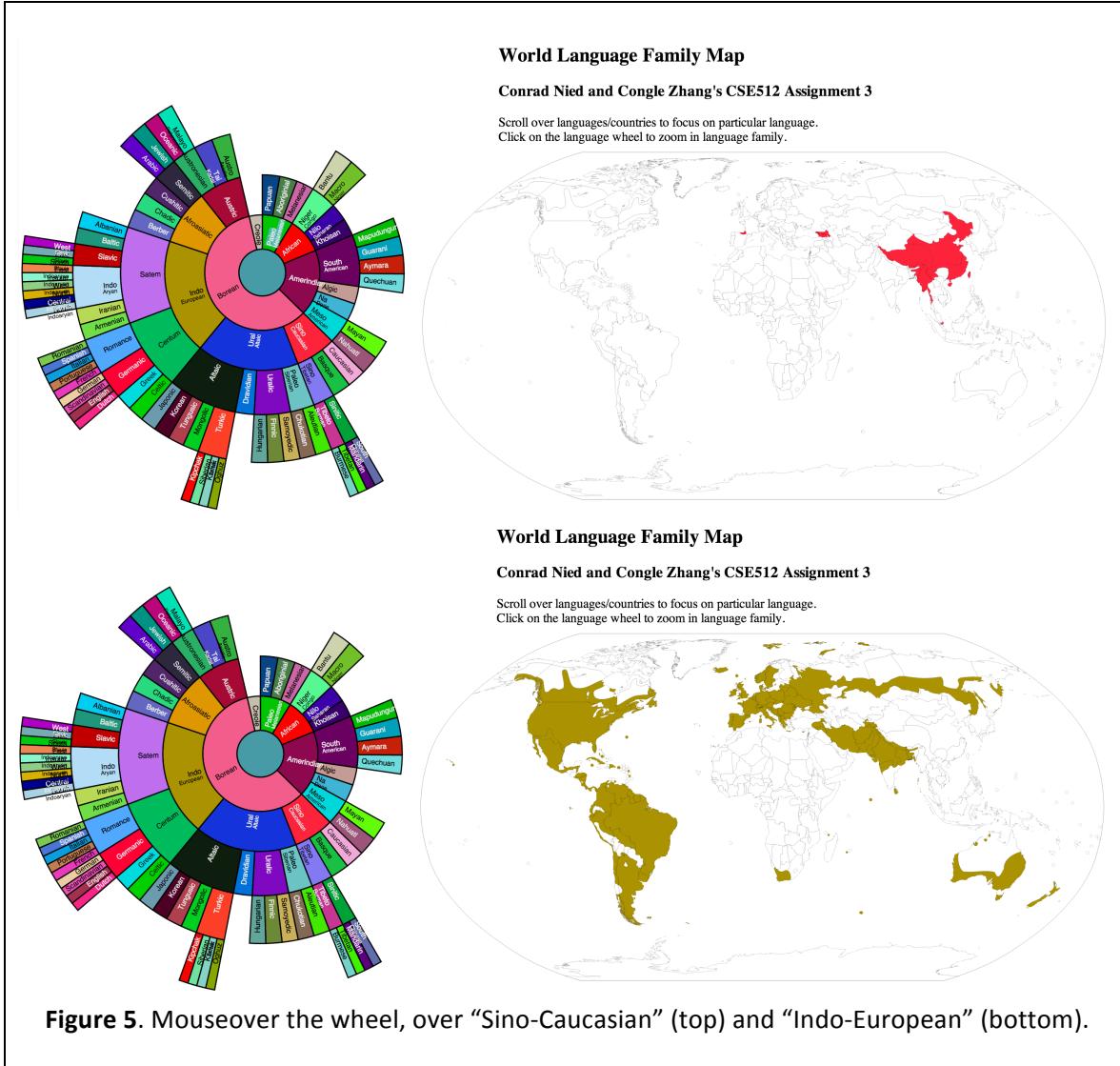
## Interaction

Our visualization supports the following interaction techniques:

1. Mouseover the map:
  - a. Highlights the linguistic area and whites out other areas.
  - b. A tooltip will jump out to show the name of the language.
  - c. The language and all its proposed linguistic ancestors in the language family are highlighted in the wheel. The other nodes in the wheel dim. For example, Figure 5 shows a screenshot for mouseover the language “English” in the map.

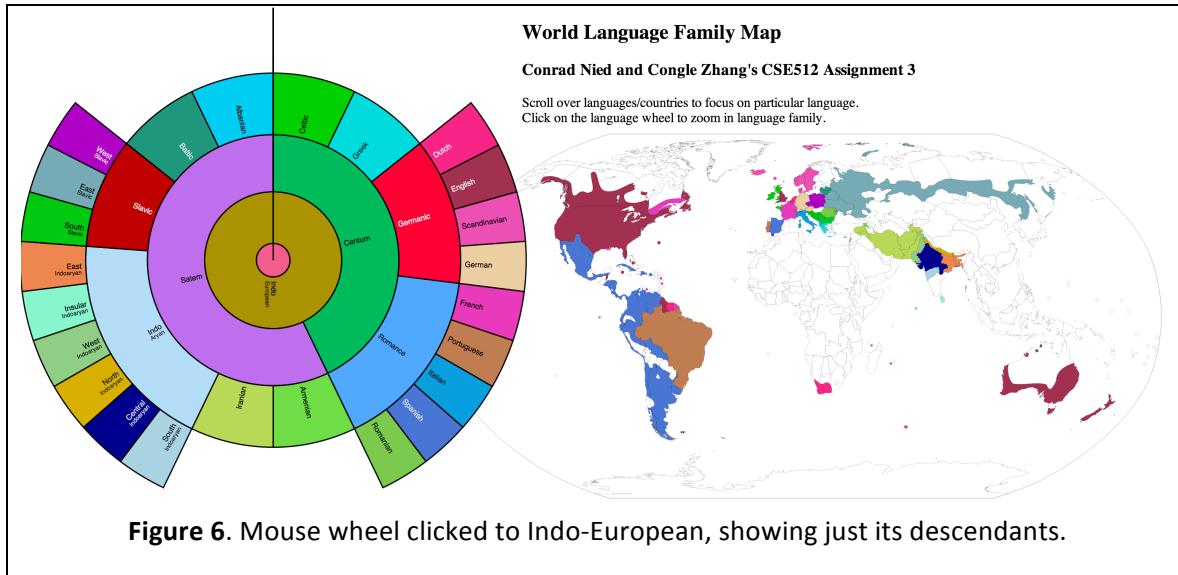


2. Mouseover the wheel:
  - a. All area of that language would be highlighted. If the language is not a leaf node, multiple areas are highlighted in the map, using a special color.
  - b. When users move around the wheel, they could easily see the change of the areas between sibling languages and across the linguistic language hierarchy. For example, it is amazing to see that the Basque language isolate in Europe is linguistically associated with Chinese rather than its modern neighboring languages. Figure 5 shows the Indo-European and Sino-Caucasian groups – showing how isolated Basque is.



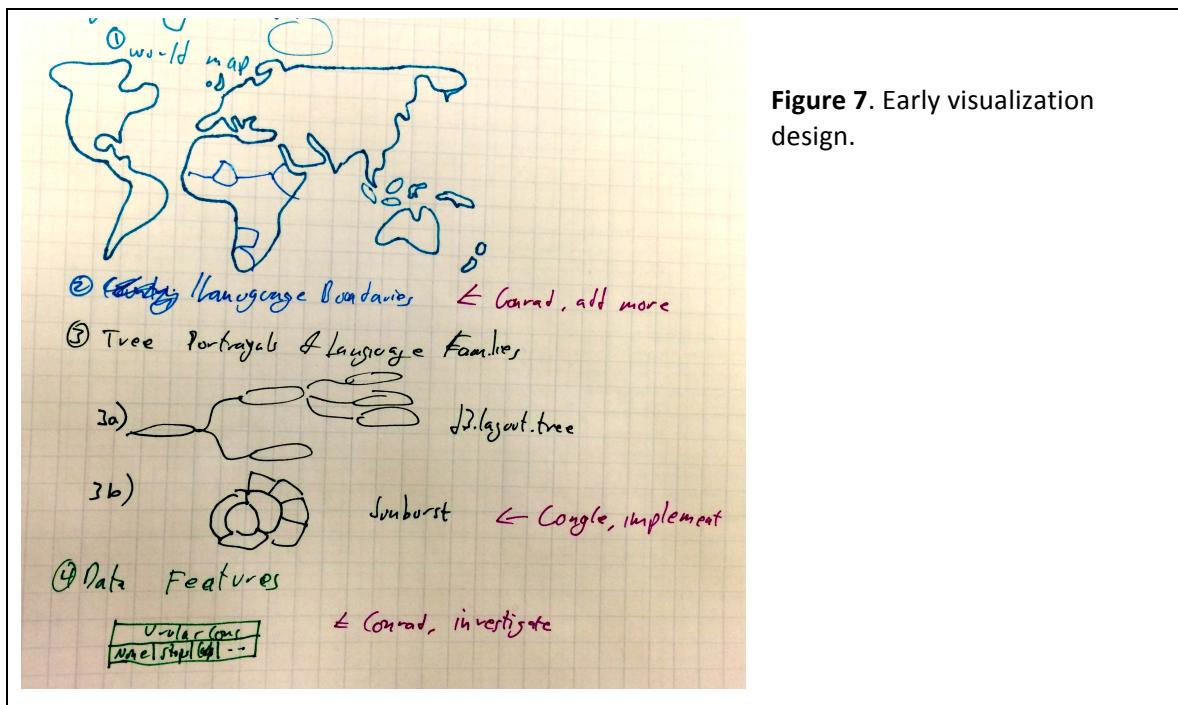
3. Click the wheel:

- When users are more interested in some particular language groups, they can click the wheel to zoom in to that language family. After that, they can click the center of the wheel to zoom out. Figure 6 shows the zoom into the Indo-European language family. Although there is a technical bug delaying the render, refocusing the mouse will change the coloration of the languages on the map so that only languages of the particular language family are shown.



## Development

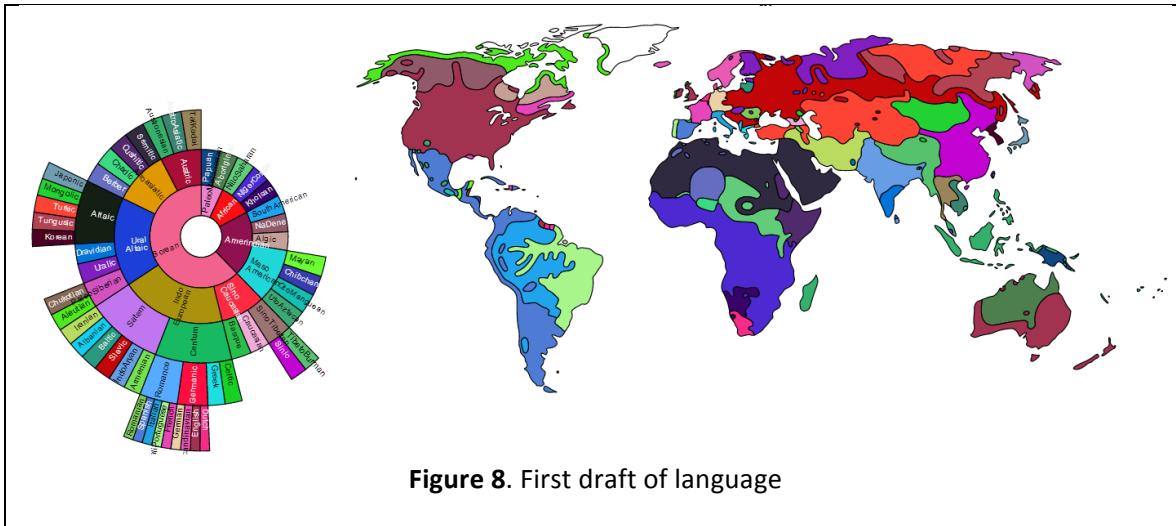
We started with a naïve scalable vector graphic of the language families (Figure 1) and a sample d3 sunburst graph. From this, we wanted to add more fine-grained linguistic groupings and add ways for the sunburst diagram to interact with the map. Figure 7 shows our early plan.



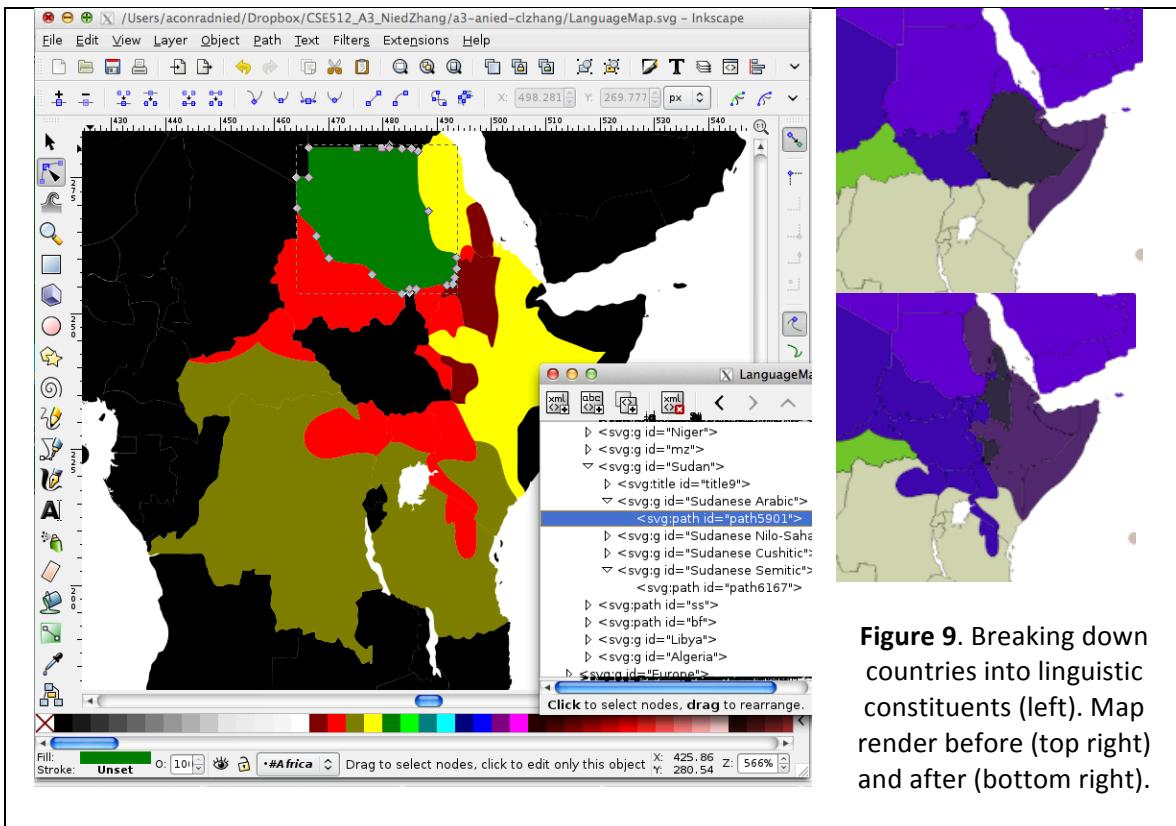
We also manually collected data from different linguists about their interpretation of the language hierarchies. Some linguistic relations are commonly accepted, some are highly

contentious. We picked the most effective groups but had to use some of contentious groupings in order to organize the languages into a whole tree relation.

The map we started with was great but too low detail. Our first draft can be seen in figure 8.



Thereby, we used Inkscape to manipulate an SVG map of the world broken down by country borders as a template. From this, we broke down countries with significantly varied (and scientifically attested) linguistic subunits. This process is showing in figure 9.



## Contributions

Congle Zhang took the technical helm for the project, he designed the script to render the wheel and provide interaction techniques. Conrad Nied was the graphical designer and used his linguistic background to inform the breakdown of linguistic groups. Together, I estimate we spent 50 work hours on this application. The element that took the most time was designing the map graphic since we had to break down the areas by hand.

There is no single world map of languages broken down into this detail or designed with the language family tree. Our interaction techniques provide a good way for users to experience the geographic distribution of languages as well as see how languages may relate.

## Limitations

- The data, both the location of languages and the relationship between languages, is contentious. Different linguists have different opinions of the groups and locales.
- There are a few technical bugs slowing down rendering that we were not able to iron out in time.
- The coloring could be optimized / the wheel simplified depending on the zoom level.