

# Extinct & Endangered Languages Around the World

## Final Report

### Data Description

Our project illustrates the distribution of extinct and endangered languages around the world. The dataset is downloaded from Kaggle<sup>1</sup>, posted by the Guardian. It includes 14 variables (i.e. columns) in total, but we selected only 5 variables to visualize in this project: the name of language, longitude, latitude, degree of endangerment (vulnerable, definitely vulnerable, severely vulnerable, critically endangered, or extinct), and the number of speakers.

We also created a separate dataset for only extinct languages with known extinction dates. Since there were less languages that are extinct in the dataset, we manually looked up the extinction date for each extinct language on Wikipedia<sup>2</sup>. Several extinction dates were approximated. Those without known extinction dates were excluded from this dataset. We further processed the data through sorting and grouping it by the extinction years, then find the cumulative sum as the time progresses.

We used the d3-geo-projection extension<sup>3</sup> to draw both the world map (geoNaturalEarth) and the USA map (geoAlbersUSA), using the JSON files given by the professor for class exercises.

### Map Description

For the world map, we wanted to first demonstrate the overall distribution of the extinct and endangered languages. We used circles to pinpoint the location of each language and a color scale of purple-to-yellow for the corresponding degree of endangerment (extinct languages are purple, while the vulnerable ones are yellow). Putting labels for the language names or showing the number of speaker for each of them would overwhelm the readers with information and cause it to become illegible, so we decided to only use 2 variables in this map: the location and degree of endangerment.

More variables were used in the USA map, where we added labels for some of the notable languages and also the number of speakers. The latter is represented via varying sizes or radiuses of the circles, scaling them linearly. The larger number of speakers, the larger the circle. For a clearer view of the cluster of data points, we decided to create a zoomed in mini-map of USA's west coast on the left corner. This time, all of the languages were

---

<sup>1</sup> <https://www.kaggle.com/the-guardian/extinct-languages>

<sup>2</sup> Wikipedia contributors. "List of languages by time of extinction." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 4 Mar. 2017. Web. 7 Mar. 2017.

<sup>3</sup> <https://github.com/d3/d3-geo-projection>

labeled accordingly, along with the varying circle radiuses (although most of them have the same size as they are either already extinct or critically endangered).

## Story

After we mapped all the data on the world map, the visualization informs us where all the extinct and endangered languages are located around the world. We noticed that there are many extinct languages along the borders of some countries, which is surprising to us. For instance, there is a high concentration of extinct languages along the west coast of North America, and there is also a high concentration of extinct and endangered languages along the west border of China. In addition, there is a high concentration of extinct and endangered languages in the central and upper part of South America, central Africa, and Pacific Islands.

This is a very interesting finding. We suspect that there are more dialects spoken along the borders of each country, and a relatively smaller population speaks those dialects. As the official language in each country becomes predominant, the number of people who speaks the dialects decreases. Moreover, we noticed that there are numerous Native American tribal languages become extinct or endangered along the west coast. We assume the extinction of these languages are due to the American history where Native Americans have been put into reservations and are assimilated to Caucasian culture and languages. Another speculation of the extinct languages along the west coast of the U.S. is that the west coast used to be part of Mexico and Canada during the 1900s. As the west coast became a part of the U.S., some languages became extinct.

We would like to use our data visualization to kindly remind our audience to preserve one's unique heritage by saving certain endangered languages. It would be extremely unfortunate and a great loss for the world culture if more and more languages become extinct.