CSC 324: Project Documentation

1. **Basic Documentation**
   1. **Purpose**

By enabling us to communicate with one another, language is, perhaps, what enables social life and helps us build communities.

* Helps us create a sense of community
* Holds cultural knowledge 🡪 years of it

India is a country with many languages and cultures. I

* Language and cultures are intrinsically linked
* India has many many languages
* What do we stand to lose 🡪 endangered languages
  1. **Data description and method of data collection**
     1. **Census 2011 Data**

Source Link: <https://censusindia.gov.in/2011census/C-16.html>

This dataset is the 2011 Population Mother Tongue Table taken from the Indian government’s census site. It consists of 10,333 observations and 15 columns – namely, Table, state\_code, district\_code, subdistrict\_code, area\_name, language\_code, language\_name, total, male, female, total\_rural, male\_rural, female\_rural, total\_urban, male\_urban, and female\_urban.

According to the Indian government’s census website, the data is collected by government workers who visit each house to collect the information and fill up 2 forms per household with their observations.

* + 1. **Languages of the world data**

Source Link: <https://www.kaggle.com/rtatman/world-atlas-of-language-structures>

The data was taken from the WALS (World Atlas of Language Structure) database, which is an online database maintained by the Max Planck Institute for Evolutionary Anthropology.

* + 1. **Endangered Languages data**

Source Link: <https://www.kaggle.com/the-guardian/extinct-languages>

The dataset was taken from a blogpost on Guardian.

* + 1. **Shapefile**

Source Link: <https://www.diva-gis.org/gdata>

This data was provided by GADM.

* 1. **Users for visualizations**

Audience: People who know about Indian languages,are interested in them

* 1. **Questions and strengths**

Questions:

* What languages exist in India and what language families do they belong to?
* What is the geographic distribution of these language families?
* What languages are endangered in India and what is the status of it?
* How are endangered languages distributed?
* Is there a correlation between number of languages in a family and the number of speakers of that language
* What is the distribution of the speakers of each of the 20 scheduled languages across the states of India?
  1. **Insights from data**
  2. **Possible improvements (to data?)**

All the criticism about data (state here)

* 1. **Sources or References (with appropriate credits)**

Datasets:

* The Guardian, 7th December, 2016. “Extinct Languages,” distributed by Kaggle, <https://www.kaggle.com/the-guardian/extinct-languages>
* Government of India Ministry of Home Affairs Office of the Registrar General and Census Comissioner, 2011. “Population by Mother Tongue,” distributed by Census India Government, <https://censusindia.gov.in/2011Census/Language_MTs.html>
* R. Tatman, 7th September, 2017. “World Atlas of Language Structures,” distributed by Kaggle, <https://www.kaggle.com/rtatman/world-atlas-of-language-structures>
* “Download Data by country,” *DIVA*. [Online]. Available: https://www.diva-gis.org/gdata. [Accessed: 14-Feb-2022].

<https://journals.ieeeauthorcenter.ieee.org/wp-content/uploads/sites/7/IEEE_Reference_Guide.pdf>

1. Describing the process and development to make your work reproducible, for example, *tidying data*.
2. Description of design decisions (encoding/mapping). Use the taxonomy of the what-why-how analysis framework presented by Tamara Munzner.

Map 1:

Bar chart:

Map 2:

Line chart with points: