

Visualizing the vitality of endangered Indigenous languages within Canada

Group123

Armun Ghafari, Hyun (Harold) Lee, Haili Su and Boya Wu

Visualization:

<https://public.tableau.com/profile/group.onetwothree#!/vizhome/shared/Q2NYD32J9>

Overview:

Language endangerment is the decline in use of a language due to the decrease of the number of speakers, language shift under the influence of the dominant language(s), or both. In Canada, around 88 distinct Indigenous languages are being spoken. However, due to colonial history and the influence of the majority language, most of them are currently endangered.

In 2019, the House of Commons passed Bill C-91, *An Act respecting Indigenous languages*, which recognizes the linguistic right of Indigenous people in Canada and vows to support the revitalization and promote the public awareness of Indigenous language.

The visualization will aid in the search of following information, number of speakers at home, number of speakers by age group, change of number of speakers from 2011 to 2016, as well as an endangerment status of the languages. We will compare how the numbers have changed using various visualizations, and display the distribution of each age group.

Our project fits the description of social good because Canada is built upon the foundation of diversity. It is important as a community to address and assess the vitality of Indigenous languages that are under threat as they are a crucial cultural diversity of Canada. The vitality of language is reflected by the combination of parameters which include the number of L1 speakers, (especially the young L1 speakers) as well as the people in the community who are learning the language.

Our target audience is mainly for federal policymakers to help them better understand what has been happening and to be more aware so that potential legislation and new laws can be passed. However, we still believe that this information is important for the general public to know as well, in hopes that social movements surrounding the topic of Indigenous languages can be achieved.

Data:

To implement our visualizations, we created the datasets with data from 2011 and 2016 censuses of Canada from Statistics Canada. The sources of our data are referenced on the bottom of every visualization. Our datasets include the following variables: the percentages of speakers in some provinces for every major Indigenous language family (the province where the speaker population concentrates the most only) in 2016; the total and age-group-specific number of speakers for selected languages in 2011 and 2016; the number of at-home speakers for selected languages in 2011 and 2016; and the number of first language (mother tongue) speakers for selected languages in 2011 and 2016. The age-group-specific data are divided into 3 groups (<15, 15-65, >65) because the 2011 census data only includes these three groups, though the 2016 data provides finer categorizations. In addition, the EGIDS level of every language, which represents its endangerment status, is gathered from Ethnologue.

Analytical Questions:

Using our interactive visualization, we expect to answer the following questions:

- Has the use of Indigenous language declined over time? If so, when has it declined the most?
- What kinds of statistics can be used as parameters to assess the vitality of Indigenous languages?
- How effective are current policies that are helping to stimulate Indigenous languages?

To answer the analytical questions posed earlier, first we take a look at our visualization. Notice that in the Speaker in 2016 layer, the number of speakers of Cree language increased from 44,635 to 82,735. Number of people who used the language frequently nearly doubled for Cree. In addition to this, Siksika language had 805 speakers in 2011, and 4,795 in 2016. This is an increase in the number of people who reported that they can speak the language.

Now if you look at L1 speakers in the same year 2011 and 2016, Cree language decreased from 78,200 to 76,385. L1 speakers across different indigenous languages stayed approximately the same.

The biggest decrease being Haida language at approximately 25% loss in L1 speakers. The biggest increase is Oji-Cree language at approximately 40%.

Analytically speaking, the number of people who reported that they can speak the language increased very rapidly over the course of 5 years. However, the number of L1 speakers (mother tongue) did not really change that much. This phenomenon is due to increased native language revitalization programs that taught youths and adults whose mother tongue is not native language. In fact, by teaching and schooling young adults these endangered indigenous languages have been widely spread in various provinces of Canada. This shows that current policies and legislation that are in effect are actively rejuvenating indigenous languages.

However, it is important to note that the increase in the number of speakers does not directly come from those with Indigenous mother tongues. No such policy or legislation have been passed to directly interfere with the diminishing population of indigenous mother tongue. Using our visualization, we've identified this trend and now is the time to help remedy the languages that are under critical threat. It is critical to tackle the root of the challenge, increasing/maintaining indigenous population so that mother tongue speakers are preserved. Some suggestions include direct funding of Indigenous communities, increased access to health care, and increasing societal prosperity.

Visualization:

As to respond to comments on the proposal, we want to demonstrate the geographic distribution of the language families, so choropleth is precisely what we need. On the top, the viewer may select between visualizations by clicking on the grey panels. This way, the viewer would know the main focus of each visualization at first glance and choose whichever they are interested in. At the top of the first map, the visualization is titled "percentage of speakers by language families in the provinces with the highest concentrations, 2016". There is a sidebar with colored circles and language family names on the left side of the map. The reader can click on and select one of the eight languages to sort geographic distribution by language. Since the same colors are used throughout visualizations for the same language families, the sidebar, at first sight, gives information about the color-language correspondence. There is the zoom in/out button and pin button on the top left of each page. This section is by default in every layer since we need to adjust views of the map accordingly.

After selecting one language family, the main visualization is the map of Canada. Each language is represented by one shade across all visualizations, and we adjust the illuminance of the shade to show the geographic distribution of speakers. Low illuminance represents more population, and high illuminance represents less population. Balanced illuminance across the map indicates even geographic distribution of the language, while contrasts between different illuminance would demonstrate that some provinces have more indigenous language speakers than others. The viewer would perceive how the territories where each indigenous language lives are beyond the confinement with the political boundaries we are familiar with. We use the pop-up technique in Gestalt psychology to highlight the chosen provinces. After moving the mouse to one area, the viewer can see that the border is bolded in white, and a text box will appear with the percentage of speakers in the province and the province name.

We chose this design style with three main elements (the text boxes, the map, and the menu) to minimize excess information as we can get everything we want across using just that.

Selecting the second panel on top of the page, we can switch to page 2 of the visualizations. On this page, the viewer can observe the number of speakers of selected indigenous languages and know where the main speaker populations are roughly located. On the top left, the reader is free to choose between 2011 and 2016. Since the location of our circles has not changed, the viewer can infer the change in speaker number of each language from 2011 to 2016. In the background, also known as the base layer, the map of Canada is in dark grey because this way, the background would contrast with colored circles on the top layer, making circles easy to detect. Based on the number of speakers, circles have different diameters. The general idea is that the less the number of speakers, the smaller the circle. We correspond circles of different diameters to speaker numbers, as shown on the bottom left of the map. Since all language families share the same scale, viewers can compare and get a visual sense of each language's geographic distribution and population size. The color of the circles also corresponds to language families. The viewer can read the text box on the right and find relevant information.

When moving the mouse onto one circle, a text box would pop up, containing two pie charts (2011 and 2016) that indicate the age distribution of the language. The age of speakers is split into three categories: younger than 15, between 15 – 64, and older than 65. At the bottom of the pie chart, there is a detailed explanation for the number of speakers in each age range. The pie chart allows the viewer to grasp the complete picture of how the language is under threat. Choosing blue, orange, and red for under 15, 15-64, and over 65 categories, respectively, we want the viewer to understand that the more young speakers there are, the less endangered a language is. On the one hand, the population of youth speakers is critical for the vitality of endangered indigenous languages since they are more likely not foreign speakers than older populations. On the other hand, less blue (a color often associated with calm and safe) and more orange or red (often associated with danger and emergency) implies the endangered status of an indigenous language from an age perspective.

Below the map, the viewer can see a trend graph of speaker population change by age, from 2011 to 2016. The graph would switch to the trend graph of the corresponding language when the reader clicks on one circle. Meanwhile, they can also select more than one circle when holding "command" in MacOS or "control" in Windows. Doing so, the trend graph would appear as the sum of speaker populations of the selected languages.

The third visualization contains two layers: the number and distribution of at-home indigenous language speakers and mother-tongue speakers in 2016. The reader can switch the layers by clicking the buttons on the top right corner. The way of demonstrating the population and geographic distribution is similar to the previous page. Move the mouse onto one circle, and the popped-up text box contains a horizontal bar chart, comparing the number of speakers in 2011 and 2016. We chose orange for 2016 and blue for 2011 because the recent number and its change since the latest measurement could show the current situation of an indigenous language. The viewer would tell if the status of an endangered language is slightly or significantly improved or it has worsened. On the right of the map, we added an interactive line chart that includes three lines: number of at-home speakers, L1 speakers, and total speakers. The chart's vertical axis is the number of speakers, and the horizontal axis is the year. Like the second visualization, we tried to present to the audience how the number of speakers has changed over the years. We also wanted to convey that there are multiple dimensions the audience should look at. The increase in the total number of speakers is perhaps not as optimistic as people thought since it does not ensure the increase in at-home speakers or, moreover, L1 speakers. We chose cyan, blue, and yellow for at-home, L1, and total speakers to distinguish among the three by color. We did not choose any highly saturated color because all three indications are equally critical. Similar to the second visualization, the viewer may select more than one circle and see the comparison and contrast of many speakers between two languages on the line chart.

The last visualization is about the endangerment levels of selected indigenous languages. Most indigenous are endangered, and we wanted to show it to our audience. To do so, we chose orange to be red (often associated with alarm, danger, and anxiety) for five of the six endangerment statuses. The shaded circles on the map share the same size because we wanted the reader to focus on the situation of the indigenous language and not other elements. On the right sidebar, We provided the endangerment status shades and a window of Wikipedia. The viewer can scroll down the window and get the complete list of endangerment status definitions. We put the web page here because we wanted to keep the visualization clear and concise without unnecessary distractions. Below the web page, we built a line chart that indicates the trend of speakers younger than 15. We focus on this age range in particular because the population of young speakers is critical for the vitality of endangered indigenous languages. Compared with older speakers, the youth are more likely non-foreign speakers and pass down their languages to the next generations.

Reflection:

We made some compromises due to technical issues. None of us knew how to use Tableau, so we did not know what terms to search for initially. We change our front page from dashboard to story to let the viewer have a smooth switch between pages. Second, while our initial plan was to design donut charts, in the end, we substituted them with more straightforward and more precise pie charts.

We also faced some challenges and tried our best to solve them. While initially deciding to incorporate 2006 census data, the language-specific data in the census did not count the respondents who answered more than one language, which isn't the case for 2011-2016 censuses. The counting issue made the data incomparable, so we had to abandon it. When collecting and organizing data, we had difficulty finding data, especially those other than the census. Extra data might come from an unreliable source, or their definitions are blurry that it was risky to use. Making visualization on Tableau is also time-consuming. On Tableau, different types of visualization require different data structures. Tableau also requires at least 2 data sets with different structures for the same visualization. To solve the problem, the only way is to spend more time on building data structures.

The design is not without flaws. On the first page, the "zero" and "one" after moving the mouse to one circle on the sidebar are confusing, and we still have not figured out a way to eliminate them.

Since it is our first time using Tableau, there are minor technical issues expected to be solved. One improvement that we hope to make in the future is to include a more recent census from StatsCan. The data that we collected from StatsCan is only up until 2016. In about 5-10 years from now, we can revisit the visualization and update it accordingly. This will tell us the overall improvement or decline of indigenous language as a direct product of Canada's 2019 Indigenous Languages act, which vows to preserve and revitalize Indigenous communities.

Team assessment

During the proposal stage, Boya and Haili contributed to the mock design. In particular, Haili made the main visualization, including different map layers, and Boya was responsible for supplementary visualization such as the donut chart. Armun and Harold conducted the writing and report formatting.

In the product stage, Haili and Armun created the final visualization product, while Harold and Boya completed the writing of the report.