

# Project Proposal

## Basic Information

### U.S. Languages Datavisualization

Github Repository Link:

<https://github.com/DataVis-Fall-2020-Team/dataviscourse-pr-WorldLanguagesDataVis>

Group team members:

Janaan Lake

- UID: u0987016
- Email: [janaanl@hotmail.com](mailto:janaanl@hotmail.com)

Andreas Martinson

- UID: u1291396
- Email: [andreas.martinson@utah.edu](mailto:andreas.martinson@utah.edu)

Rachel Berghout

- UID: u1302108
- Email: [u1302108@umail.utah.edu](mailto:u1302108@umail.utah.edu)

## Background and Motivation.

We were inspired by a tree based visualization of the 100 most spoken languages around the world, the link to the viz can be found here: <https://www.visualcapitalist.com/100-most-spoken-languages/>

This led us to decide on our topic of language. Then after looking at available datasets we found a dataset that was interesting, but only focused on the U.S. That dataset also led us to focus on the nationwide and state levels of detail since the county level data was too sparse.

When we were thinking about how to present the data, we really liked the storytelling scrolling approach visualization found at this link: <https://cuthchow.github.io/college-majors-visualisation/>.

This influenced the method that we want to use to present our data.

## **Project Objectives**

We have four primary questions that we would like to answer:

- Which languages are spoken within the U.S.?
- Where are these languages spoken within the U.S.? (i.e. what is the distribution of the language groups)
- Which states have the greatest language diversity and what languages are spoken in those states?
- Which foreign speakers are the most fluent English speakers?

Another question that we might address if we have time is, “How have the languages in the U.S. changed over time?”

Based on this visualization, we hope that a user would use it to find more about the state they live in and compare it to other states they are interested in. We hope that it will help the user learn the distributions of languages and compare that to their expectations as well. From a technical standpoint, we are excited to learn the scrolling storytelling technique as well as the transitions associated with using it. We also hope to revisit some of the visualizations we have already created (bar charts and maps) and customize them to help tell the story that we want to tell. In the end, we want to also call out some interesting facts to engage the user with the visualization.

## **Data**

The dataset is from the U.S. Census Bureau and can be found at this link:

<https://www.census.gov/data/tables/2013/demo/2009-2013-lang-tables.html>

The data preprocessing that we would need to do is the following:

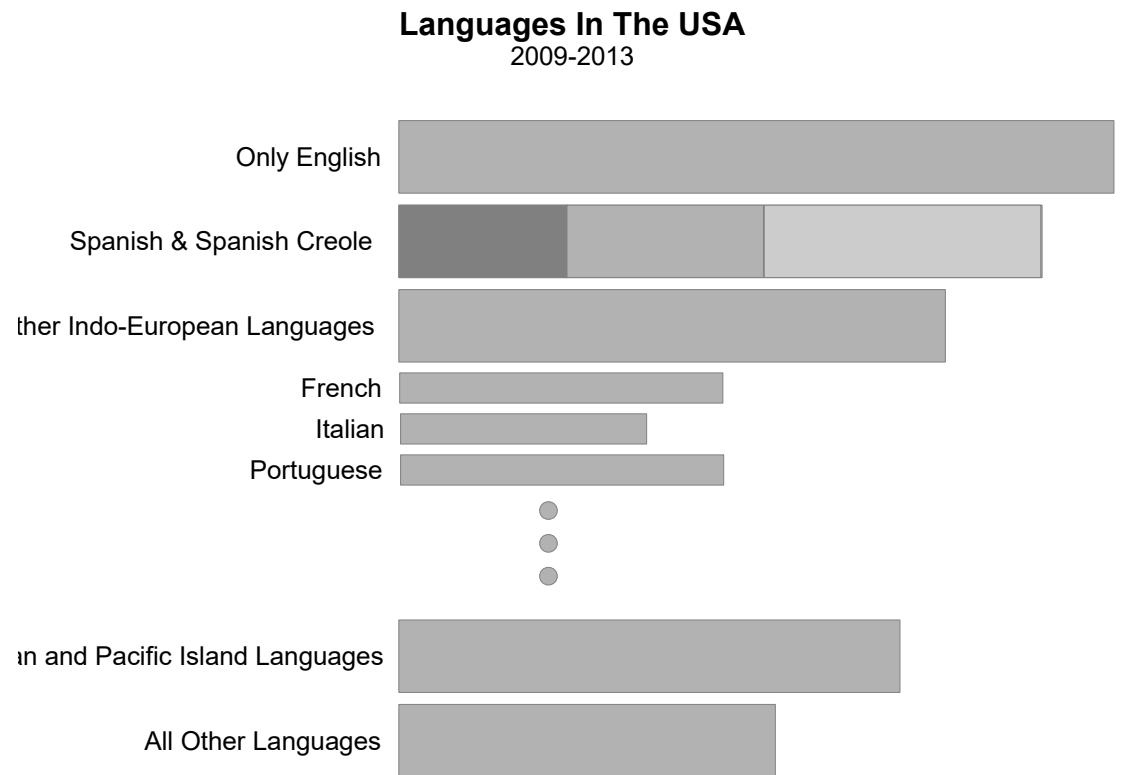
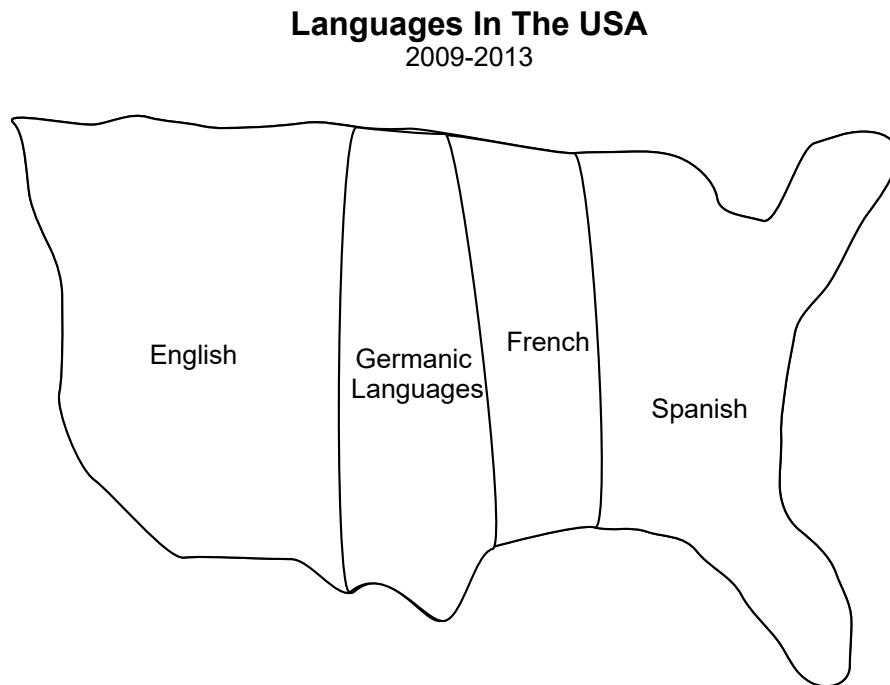
- Combine state data into a single tab
- Add a column for the language grouping instead of having that information bolded and in the same column as the languages
- Change the names of some of the language groupings to avoid parenthesis and punctuation

# Visualization Design

## Alternative Prototype Designs

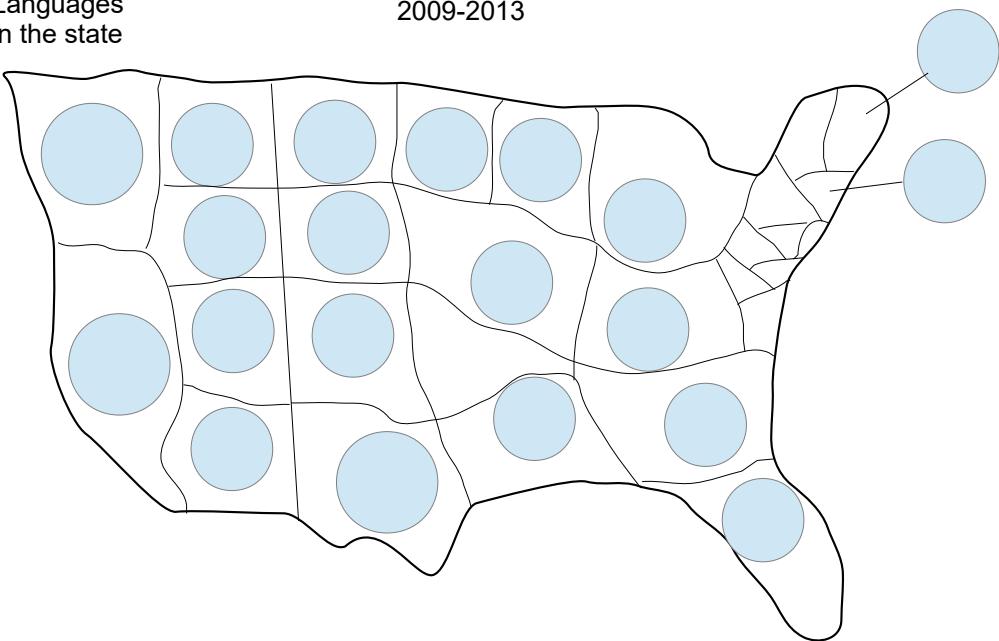
# What languages are spoken in the US?

<https://www.census.gov/data/tables/2013/demo/2009-2013-lang-tables.html>



# What languages are spoken and where?

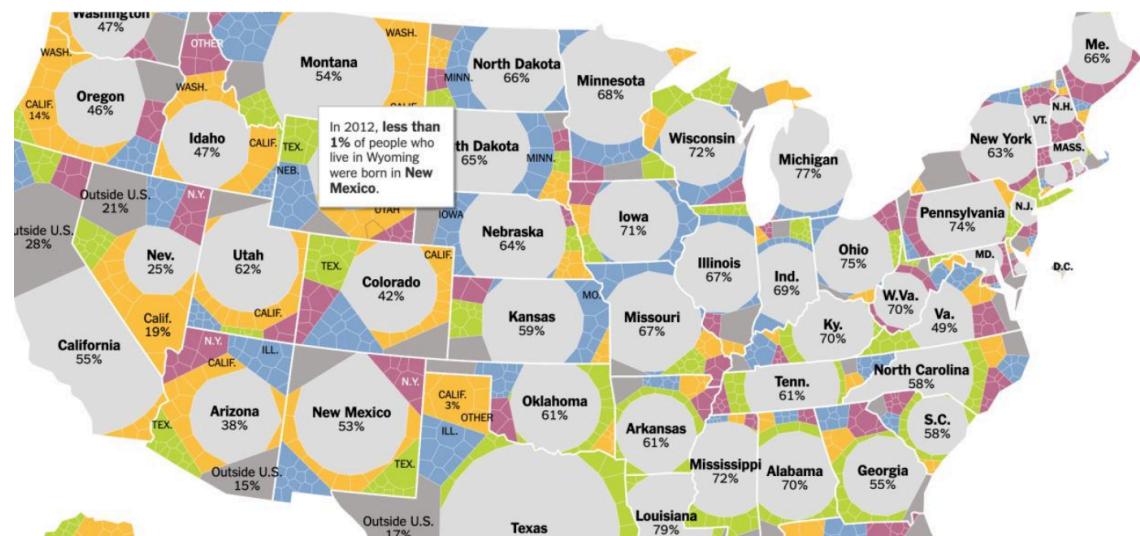
Pie charts show the percentage of the top 5 Languages in the state



# Languages In The USA

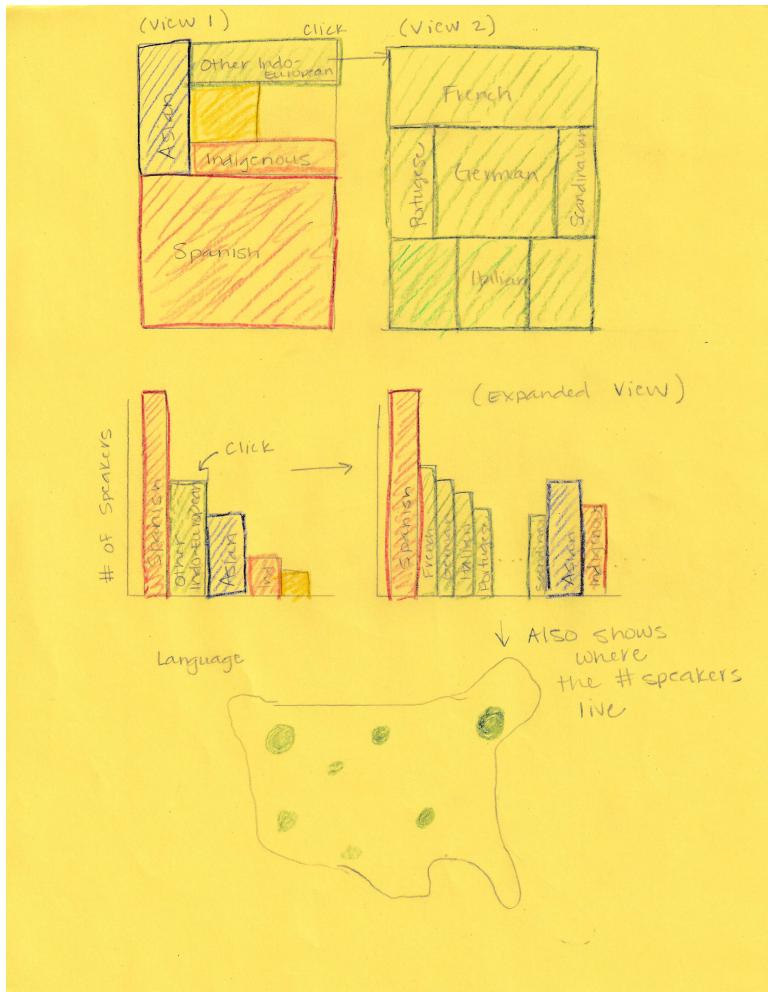
2009-2013

Break up each state by language category

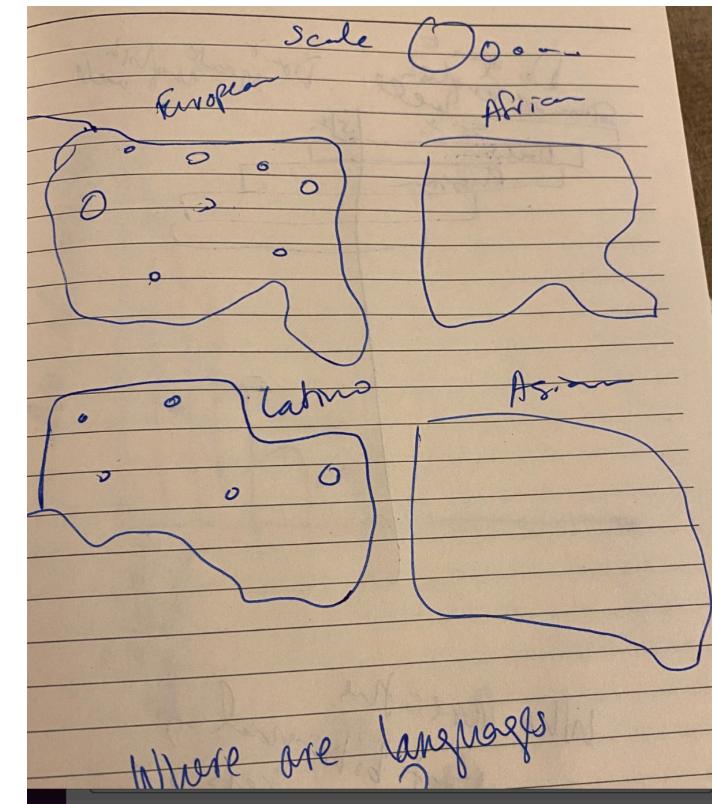
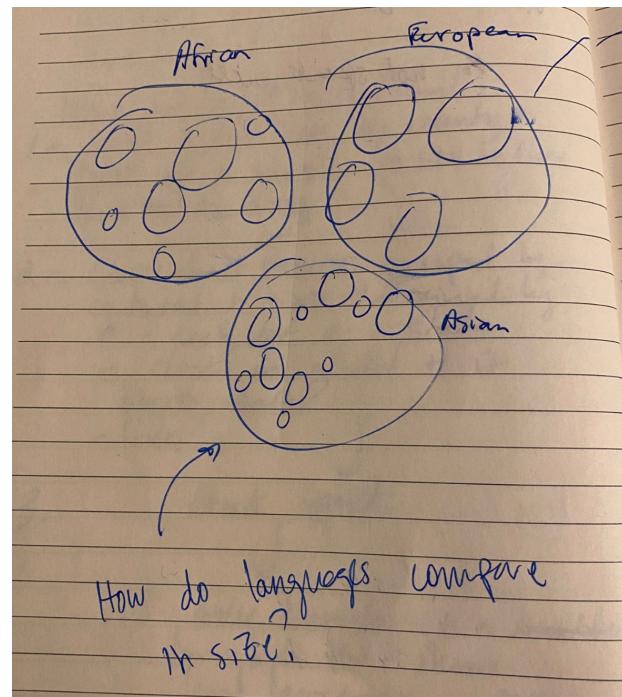


# What languages are spoken and where?

Coordinated tree maps/bar chart with map

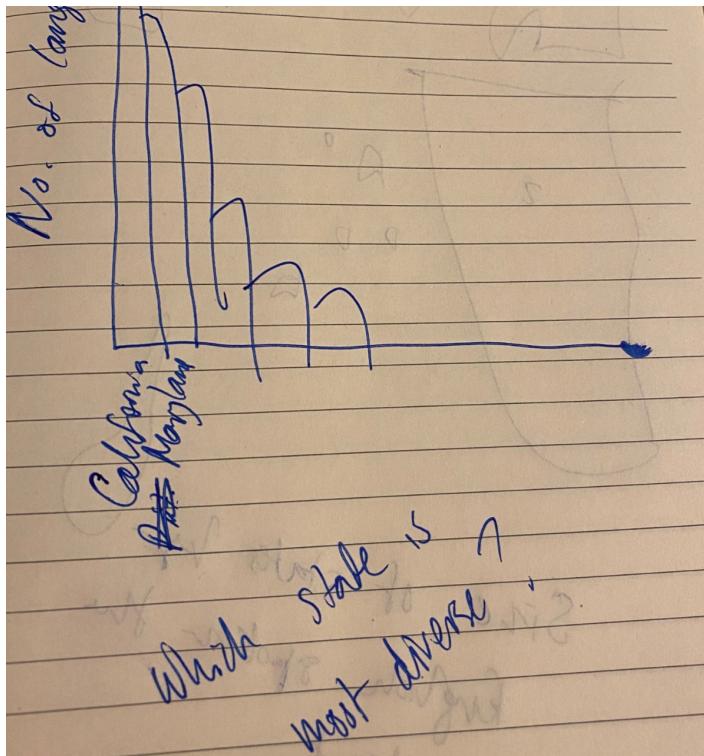


Circle swarms/groups with coordinated maps



# Which states have the most language diversity?

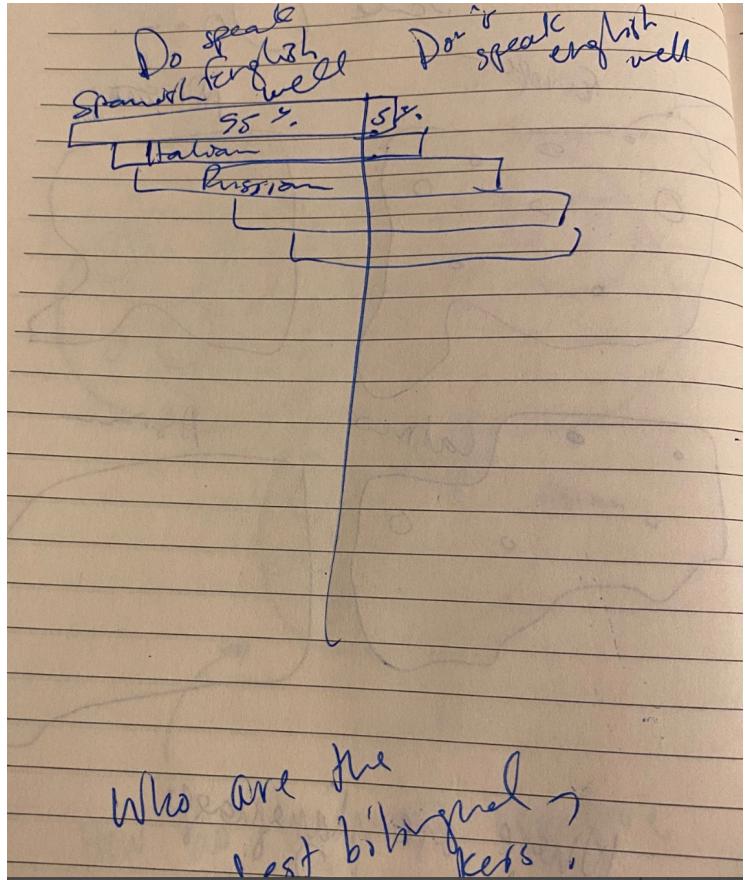
Bar chart showing number of languages spoken in each state



Breakout language groups by state/clicking for more detail

|                |       |       |       |       |       |      |
|----------------|-------|-------|-------|-------|-------|------|
| West Virginia  | 29.2% | 26.1% | 17.6% | 11.1% | 10.4% | 5.5% |
| Mississippi    | 31.4% | 25.4% | 17%   | 10.8% | 9.7%  | 5.8% |
| Arkansas       | 28.1% | 27.1% | 18.3% | 10.4% | 10.1% | 6%   |
| Kentucky       | 27.6% | 25.1% | 18%   | 11.4% | 11.1% | 6.8% |
| Idaho          | 22.5% | 25.6% | 20.5% | 12.1% | 12.2% | 7.1% |
| New Mexico     | 28.2% | 24.5% | 17.5% | 11.5% | 11.2% | 7.1% |
| Alabama        | 28.2% | 24.8% | 17.4% | 11.4% | 11%   | 7.2% |
| Oklahoma       | 24.9% | 25.6% | 18.6% | 11.8% | 11.7% | 7.4% |
| Montana        | 23.1% | 26.9% | 18.6% | 12.2% | 11.7% | 7.5% |
| South Dakota   | 20.6% | 25.2% | 20.1% | 14.4% | 12.3% | 7.6% |
| Indiana        | 22.1% | 25.4% | 19.8% | 12.8% | 12.4% | 7.7% |
| South Carolina | 24.8% | 25.7% | 18.3% | 11.9% | 11.7% | 7.7% |
| Tennessee      | 25.2% | 26%   | 18.4% | 11.7% | 10.9% | 7.8% |
| Maine          | 22.7% | 24%   | 19.3% | 12.7% | 13.3% | 7.9% |
| Louisiana      | 29.7% | 24.2% | 16.3% | 10.4% | 11.3% | 8.2% |
| Missouri       | 22.6% | 25.5% | 18.9% | 12.5% | 12.3% | 8.2% |

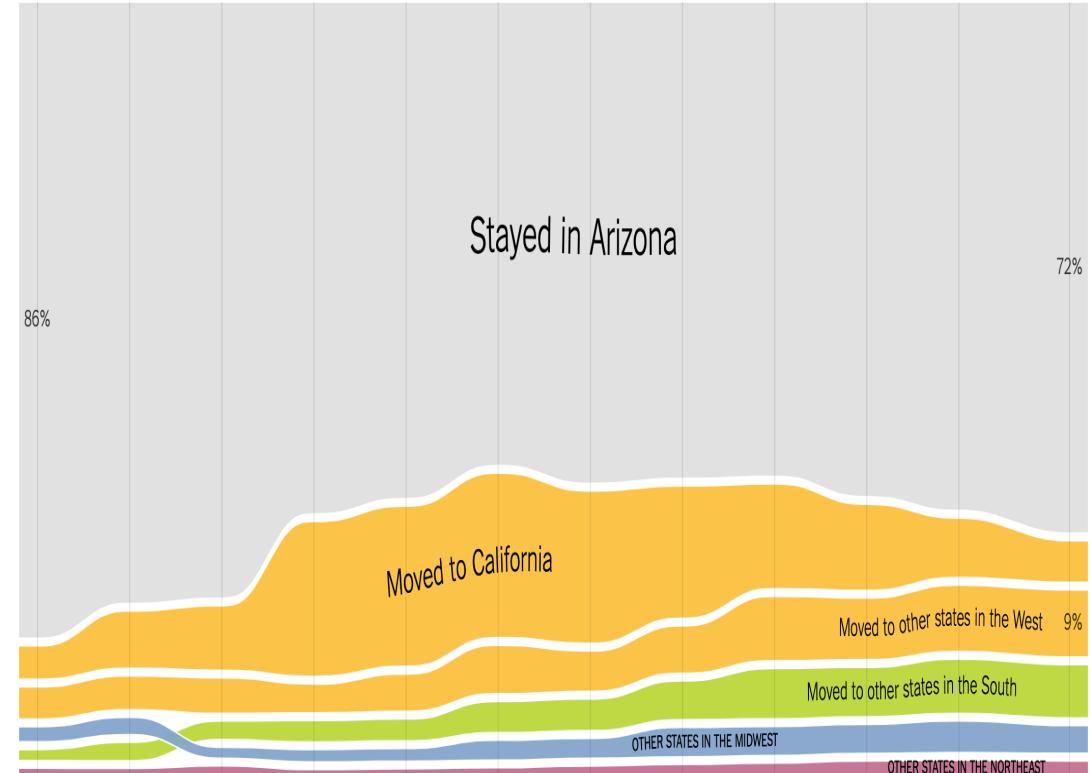
# Which speakers are more bilingual?



Diverging bar chart with percentages of bilingualism on one side, total number of speakers on the other side

# How have the languages spoken changed over time?

Distribution of languages over 30-year period

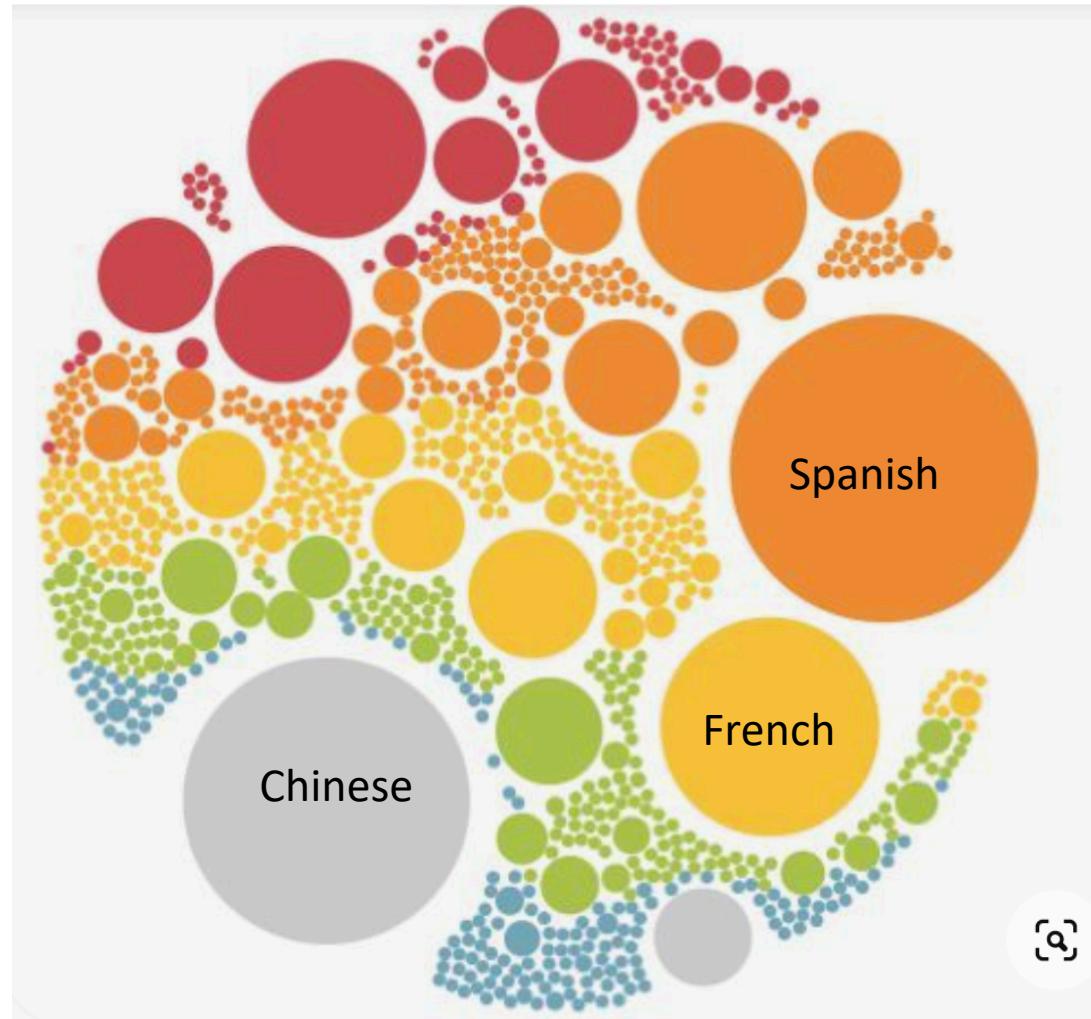


# Visualization Design

## Finalized Design

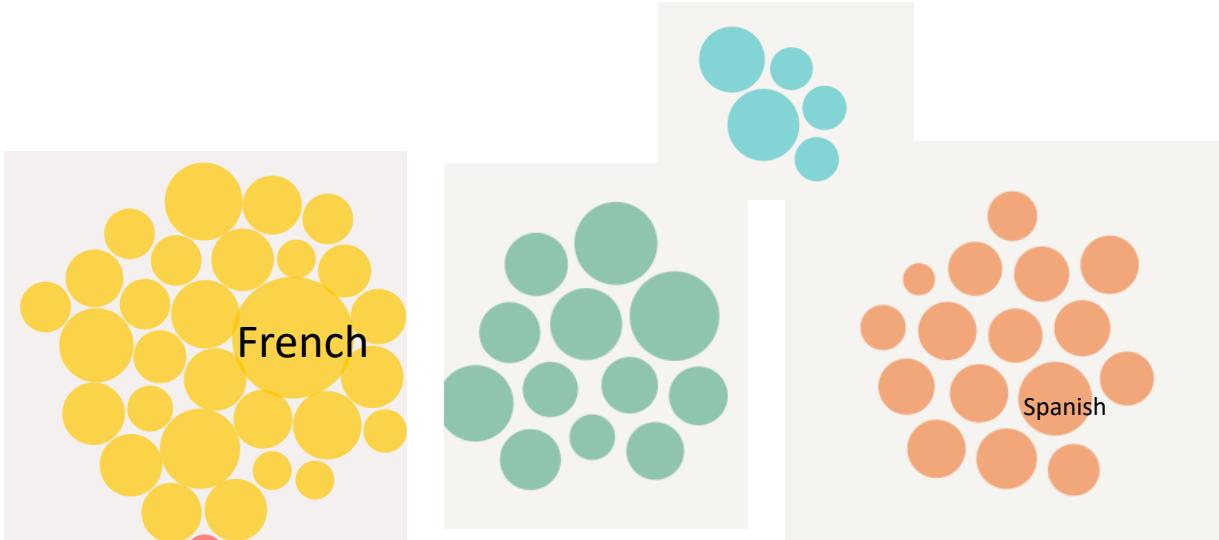
# What languages are spoken in the US?

- Tooltip that shows the number of speakers when hovered over circle (Tooltip will also show name of language for smaller circles)
- Coloring represents language groups
- Sized according to number of speakers in the US
- Storytelling feature: Highlight which languages might be spoken the most/least. Talk about indigenous languages being very small but many different kinds.



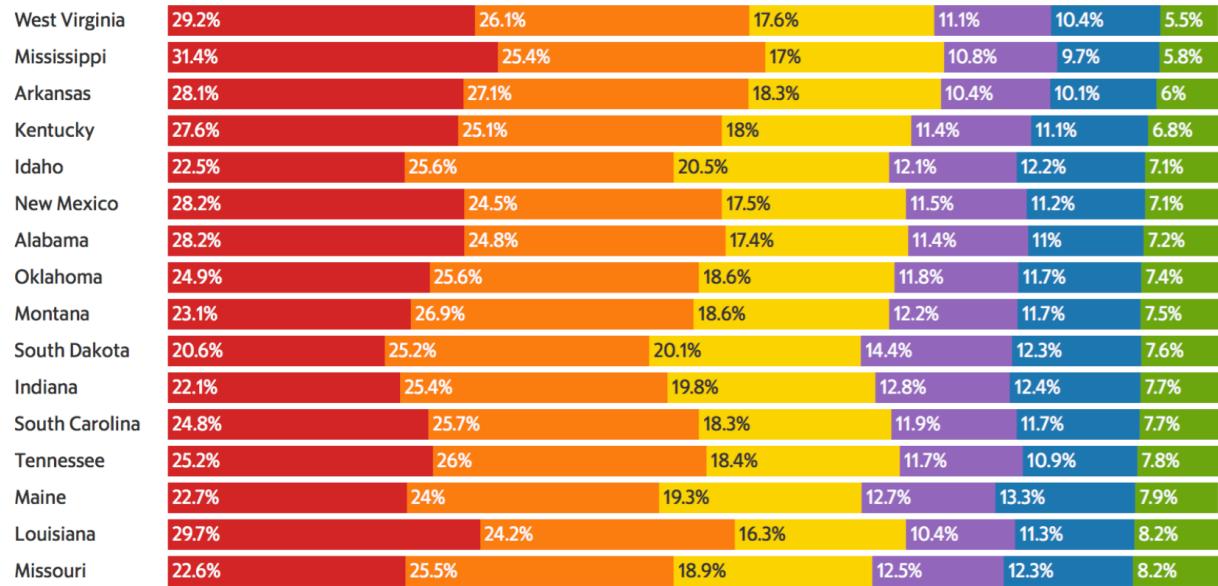
# Where are these languages spoken in the United States?

- Animated transition from first view to colored swarms
- Same circles and sizes from first view
- When user clicks on a circle, the name appears and then the map shows where that language is spoken in the US on a chloropleth map. The color will be the same as the clicked circle, and saturation of the color will be the channel represent the percentage of speakers within that state.
- Storytelling feature: Pick a few languages and highlight where they are spoken
- Optional: Brushing and/or toggling on circles for multiple selections



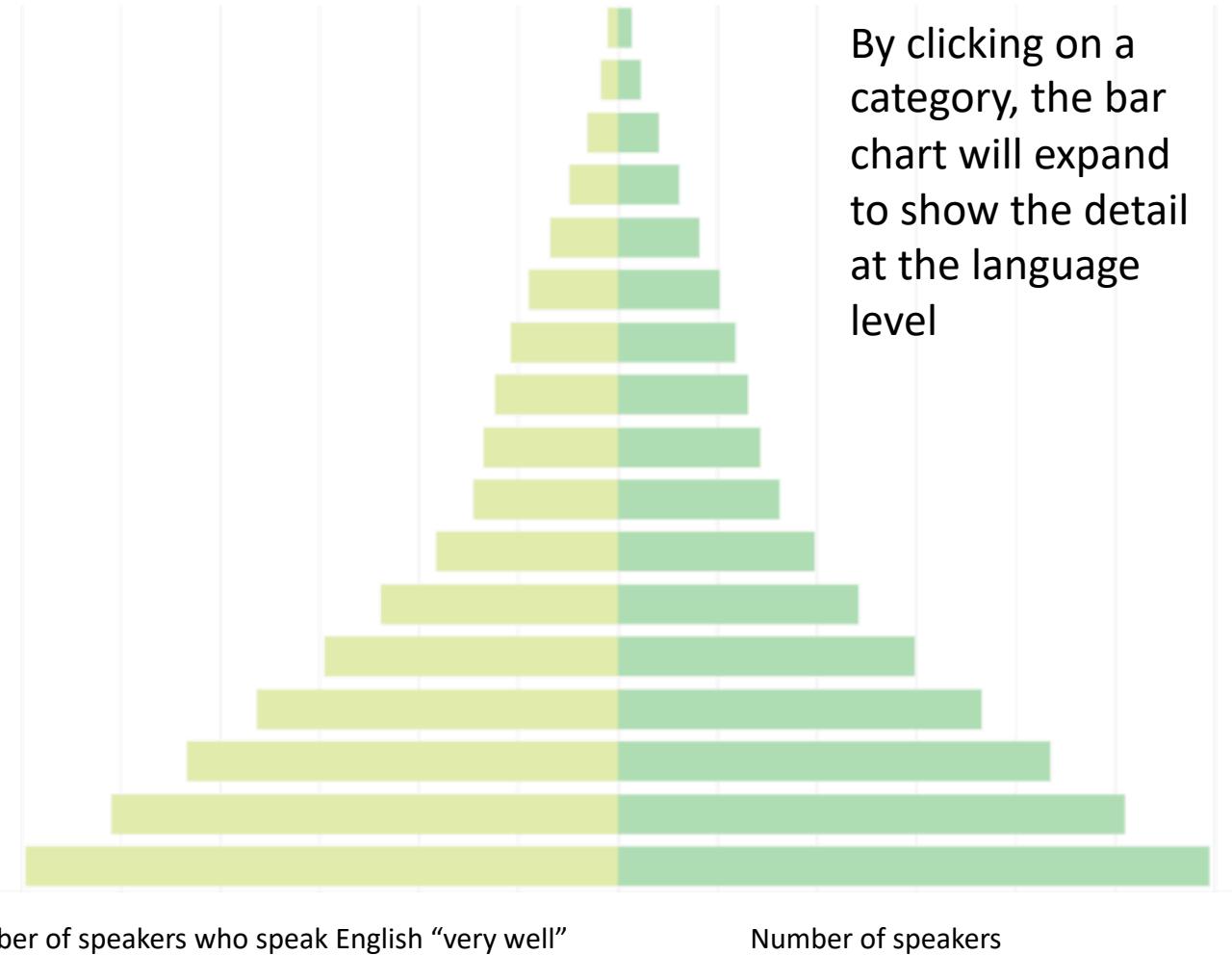
# What states have the most language diversity and what are the languages spoken in a specific state?

- Colors are the same colored language groups as before
- Each category/color can be sorted
- Option to include/exclude English
- A tooltip shows the total number of languages on hover
- Storytelling feature: Highlight states with the most and/or least language diversity and why that might be the case.
- Optional: The user can click a category bar within a state for a breakout of the languages represented in that bar and their statistics within that state



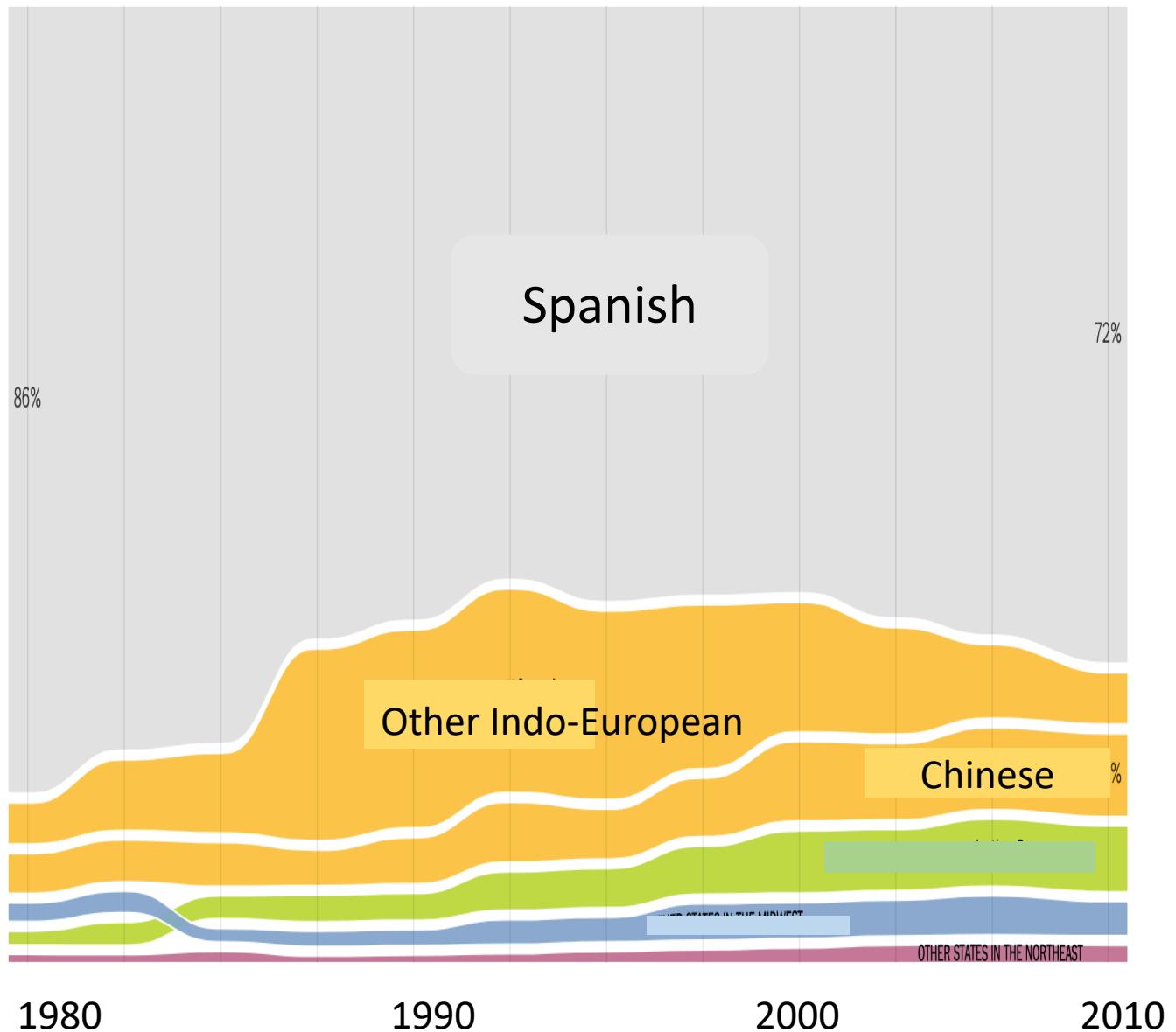
# What foreign languages speakers are the most bilingual?

- Same color groupings
- Start at grouping level
- When user clicks on a group, the bar chart expands to be at the language level for that colored group
- Storytelling feature: Highlight extremes and give reasons/explanations for why



## (Optional) How have the languages spoken in the US changed over time?

- Dates from 1980 to 2010 with 10-year increments
- Tooltips highlight sections when hovered with more detailed information, such as in the example shown
- Same colored groupings as before
- Storytelling feature: Explain that Spanish and Chinese are increasing due to the large immigration from these areas. Decrease in French, Italian, German as those communities have largely integrated. Dying out of indigenous languages.



## **Must-Have Features**

- View #1 - A bubble chart displaying the percentage of languages in the U.S. colored by the continent the language comes from.
- View #2 - A U.S. map displaying the usage density of the language group chosen (African, Asian, European, etc.). Will use a filter to display the different language groups by continent.
- View #3 - A bar chart displaying the percentage of each language per state.
- View #4 - A bar chart displaying the language speakers by group that speak more than one language fluently. Groups will break out into individual languages.
- Storytelling features in each of the views.

## **Optional Features**

- View #5 - timeline flowmap
- View #2 - Map - toggle button to combine languages into groups
- View #3 - Bar Graph - Transition to break into separate languages vs. pop out bars beneath
- 

## **Project Schedule**

Nov. 5th - Peer Feedback

Nov. 15th - Project Milestone

Dec. 2nd - Project Due

Week 1 - 10/29 - 11/5

- Set up the file structure
  - Style.css
  - Megocluster.js
  - Map\_cluster.js
  - Horizontal\_bar.js

- Diverging\_bar.js
- script.js
- Main.html
- Get the data loaded
- Write script file
- Scrolling layout - Use example as template

Week 2 - 11/6 - 11/13

- Megacluster - Andreas
- Map cluster - Rachel
- Horizontal Bar/Diverging Bar - Janaan

Week 3 - 11/14 - 11/21

- Finish all layouts
- Megacluster - Andreas
- Map cluster - Rachel
- Horizontal Bar/Diverging Bar - Janaan

Week 4 - 11/22 - 11/29

- Focus on transitions/animations

Week 5 - 11/30 - 12/2

- Final polishing

Weekly Scheduled Meetings:

- T,TH @ 8pm-9:30pm