Linguistic Diversity and English Profciency in The United States

Iyanuoluwa Shode

12/15/2021

1. Introduction

United States of America is said to be the land where dreams come true. As a result, immigrants from all over the world settle to the United States of America to have their livelihood. According to the American Immigration Council, 44.9 million immigrants live in the United States of America as of 2019. However,the mode of communication in the country is English language which happens to be the native language that is being used. The language of instruction in schools as well as professional settings is English. Due to the influx of foreigners in the country, it is expected that there will be different variations of English language as immigrants have their native languages. Therefore, English will be regarded as their second language. It is rather important to learn English in order to successfully live as a resident in the United States. This project investigates the proficiency of English Speakers in the United States in 2019. For an accurate investigation, data was collected from the United States Census Bureau.

Policy makers, linguistic students and census information that focuses on language and the general auidence that are interested in linguistic diversity are the audience for this project.

The visualizations used for this project were intentionally selected because they help to give clear explanation of the focus of project.

2. Population Density

Below is a data representation of the total population of residents across states in the United States. ## 2.1 Total Population

```
library(dplyr)
library(ggplot2)
library(readr)
library(tidycensus)
library(tidyverse)
library(purrr)
library(readxl)
Pop <- read_excel("/Users/iyanuoluwashode/Downloads/Pop.xlsx")
View(Pop)
Pop%>%
  arrange(Population/10000)%>%
  mutate(Geography = factor(Geography, levels = Geography))%>%
  ggplot(aes(x=Population/10000, y = Geography, fill = Population))+
  ggtitle("Figure 1 - Total Population in The United States") +
  xlab("Number of Residents In Ten Thousands")+
  ylab("States in the United States")+
  geom_col()
```

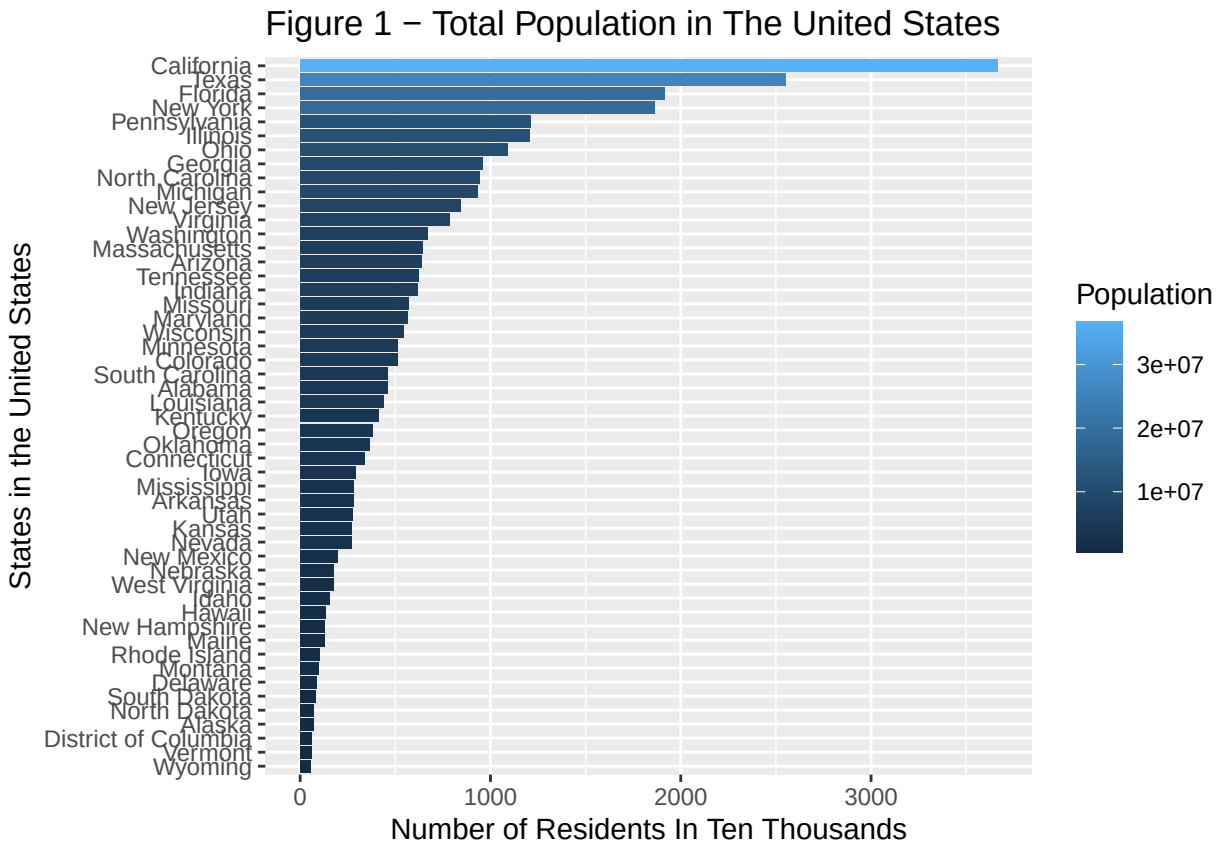


Figure 1 shows the total population of the residents in the United States of America, it can be noted that California has the highest number of residents followed by Texas and Florida.

3. Linguistic Diversity in the United States

3.1 Languages in the United States

According to The Geo Group Corporation, over 350 languages are spoken in homes in the United States with 79% speaking English either as their first or second language. The second language that is widely used in the United States is Spanish which is spoken by 13% of the people living in the United States. It is definitely no news that the United States is heavily rich with several languages. A country that inhabits over 300 million residents with 44 million immigrants from diverse parts of the world with different cultures and languages, it is a good fit to examine linguistic diversity.

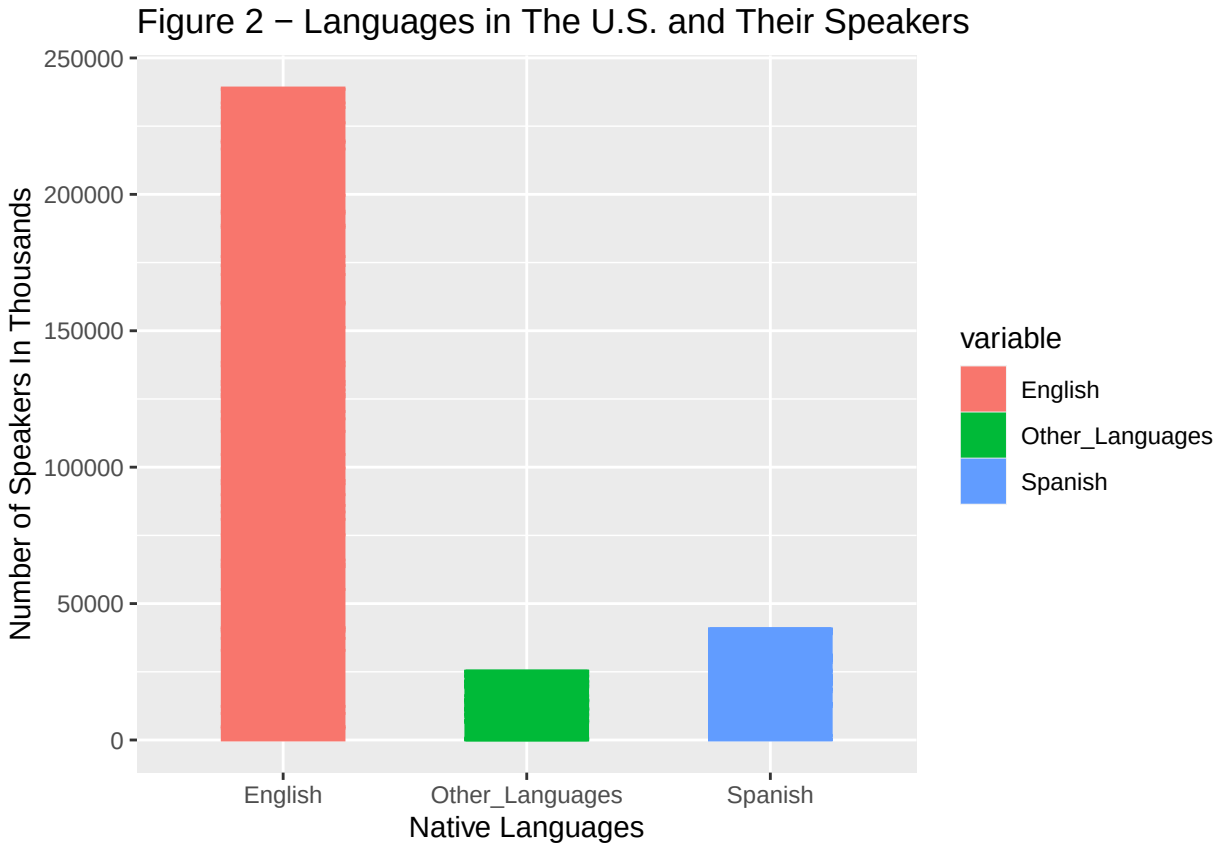
```
## Your original .Renviron will be backed up and stored in your R HOME directory if needed.
```

```
## Your API key has been stored in your .Renviron and can be accessed by Sys.getenv("CENSUS_API_KEY").
## To use now, restart R or run `readRenviron("~/.Renviron")`
```

```
## [1] "59251a03c22c28ad242843f2078d08c00c21b0ef"
```

```
## Getting data from the 2015-2019 5-year ACS
```

```
## Warning: Removed 3 rows containing missing values (position_stack).
```



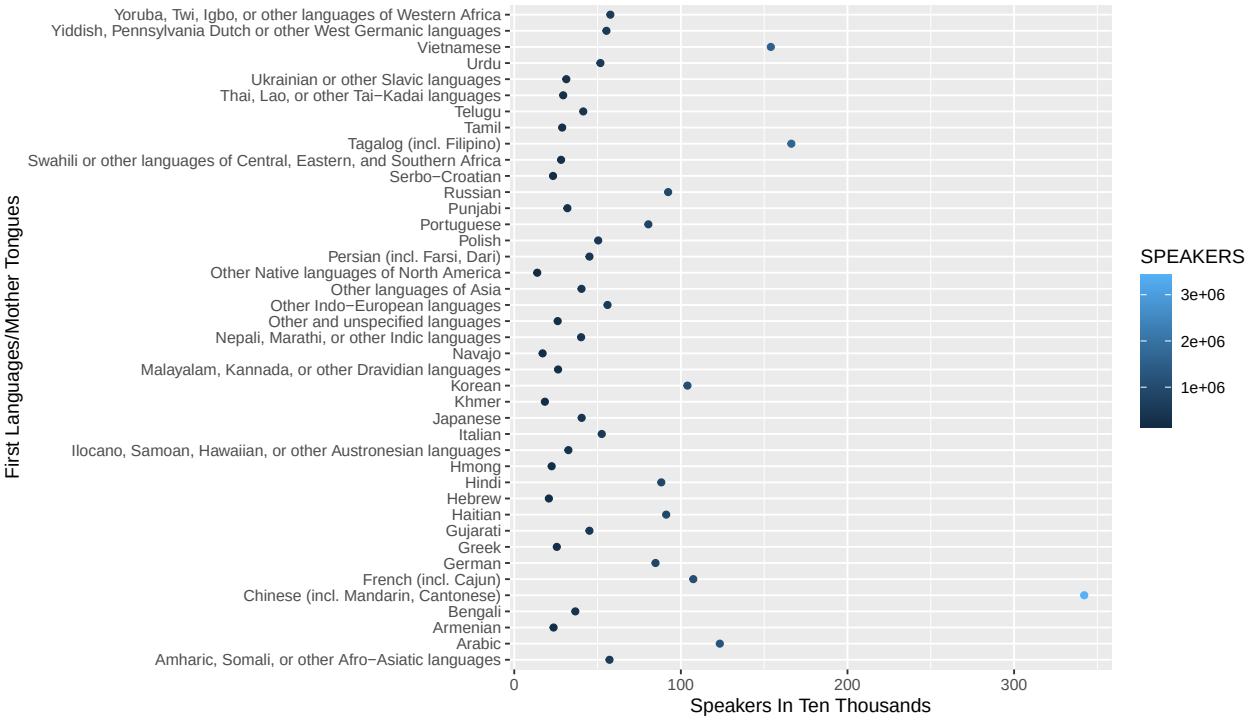
The graph above is a data visualization that shows the diverse languages spoken in the US and their native speakers. Over 200 million US residents speak English as their native and only language followed by Spanish Speakers that amount to over 40 million while all other languages' speakers are over 25 million of the U.S. population.

Figure 3 further analyses the total number of residents that have a first langauge and then speak English as their second language. We noticed that Spanish has the highest number of speakers of English as a second language. The data interpretation shows that the in the United States, there are over 41 million English speakers that have Spanish as their native language, followed by Chinese native speakers that sum up to over three million immigrants, over two million Tagalog speakers speak Engish, Vietnamese, Arabic, French and Korean speakers are over a million immigrants while Russian, Haitian and Hindi speakers that speak English in the United States are over eight hundred thousands.

```
knitr::opts_chunk$set(echo = TRUE)
Nat <- read_excel("/Users/iyanuoluwashode/Downloads/Nat.xlsx")
View(Nat)

ggplot(Nat, aes(x=SPEAKERS/10000, y = LANGUAGES, color = SPEAKERS))+
  ggtitle("Figure 4 - Native Languages in The U.S. Exclusing Spanish") +
  xlab("Speakers In Ten Thousands")+
  ylab("First Languages/Mother Tongues")+
  geom_point()
```

Figure 4 − Native Languages in The U.S. Exclusing Spanish



Excluding Spanish in the above visulizations shows that many other languages are spoken in the United States as native speakers of Chinese including Mandarin and Cantonese show that their language is the third most spoken language in the United States with over 30 million population.

3.2 Foreign Speakers of English in The United States

In the Census data, states are sorted by the number of speakers of languages other than English. State with the largest non-English communities are seen to take the lead. We move swiflty to check how these speakers are spread across the US states. It is noteworthy to state that some states were recorded with the null values in the ACS dataset from the US Census Bureau due to data suppression. Below is a data visualization of the foreign speakers of English across states in the United States.

```
TwentyNineteen <- read_csv("/Users/iyanuoluwashode/Downloads/ACSDT1Y2019/TwentyNineteen.csv")
```

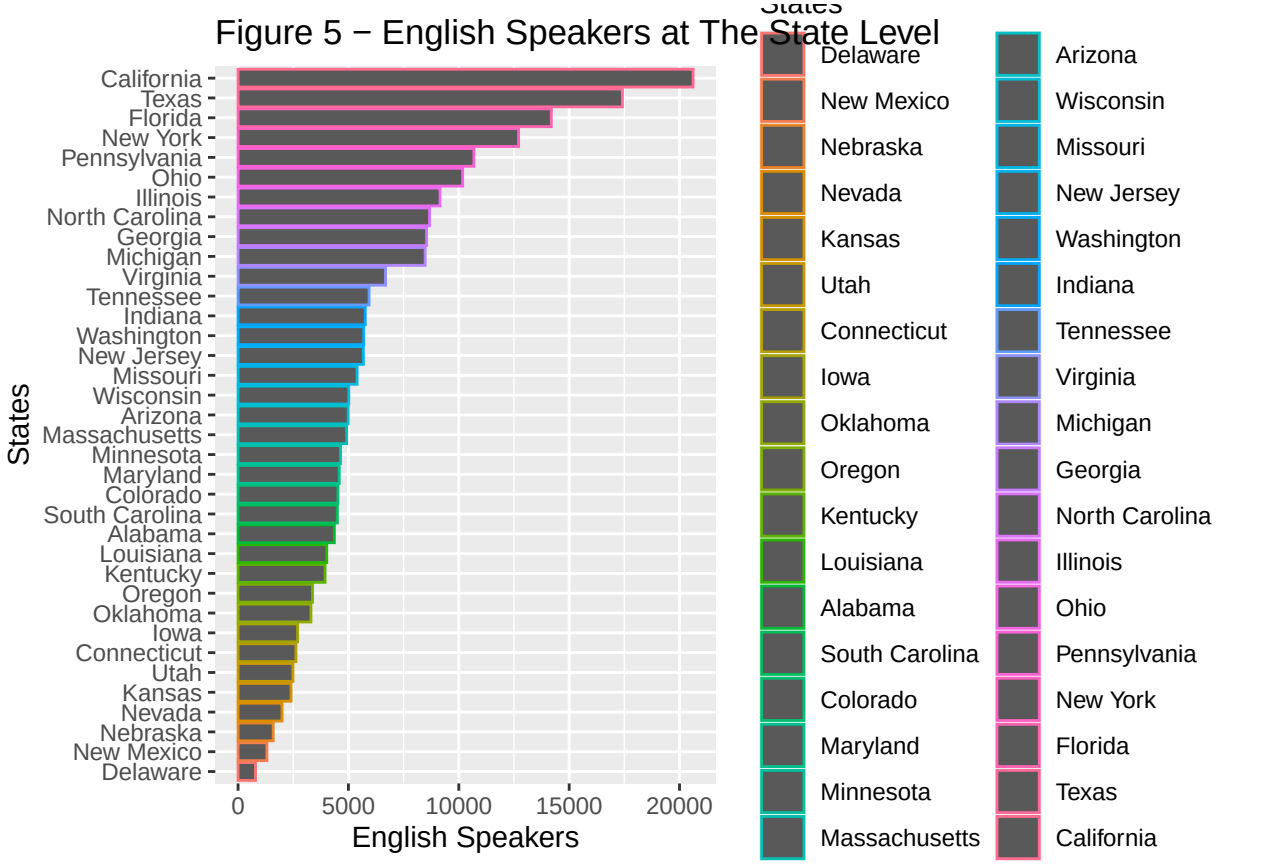
```
## Rows: 36 Columns: 45
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): States
## dbl (44): Estimate!!Total, English, Spanish, French (incl. Cajun), Haitian, ...
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
View(TwentyNineteen)
TwentyNineteen%>%
  arrange(English/1000)%>%
  mutate(States = factor(States, levels = States))%>%
  ggplot(aes(x=English/1000, y=States, color = States))+
  ggtitle("Figure 5 - English Speakers at The State Level") +
  xlab("English Speakers")+
  ylab("States")+
```

```
geom_col()
```



The chart above helps to shows the states with hge differences in terms of their linguistic landcapes, the top 10 states that have a non-English speaking populations. It further shows that California has the highest number of English speakers in the United States with over 20 million English speakers.

4. English Proficiency in The United States

To examine the English proficiency level of U.S residents, there are key information that should be observed.We are strictly focusing on the LEP rate of US residents. A person is classified as Limited English Proficient (LEP) if he or she is age 5 or above and is reported to be speaking English less than "very well" while an proficient English speaker is one that speaks only Engish or speaks the language "very well" With the aid of the ACS data from the U.S. Census Bureau, the demographic and socioeconomic information of LEP individuals that are U.S. residents can be acquired by observing the following characteristics:

4.1. Distribution by State and Key Cities

4.2. Nativity Status of U.S residents

4.3. Age Distribution of LEP Speakers

4.4. Education Attainment of LEP Speakers

4.5. Education Attainment of LEP Speakers

4.1. Distribution by State and Key Cities

Our interest is in the low proficiency rate of English speakers in the states and key cities in the country. The visualization in Figure 4 shows the states with the highest English speakers and I went on to observe how many people are low proficiency in English. The ten states with the most speakers were picked and their distributions are visualized below in Figure 5.

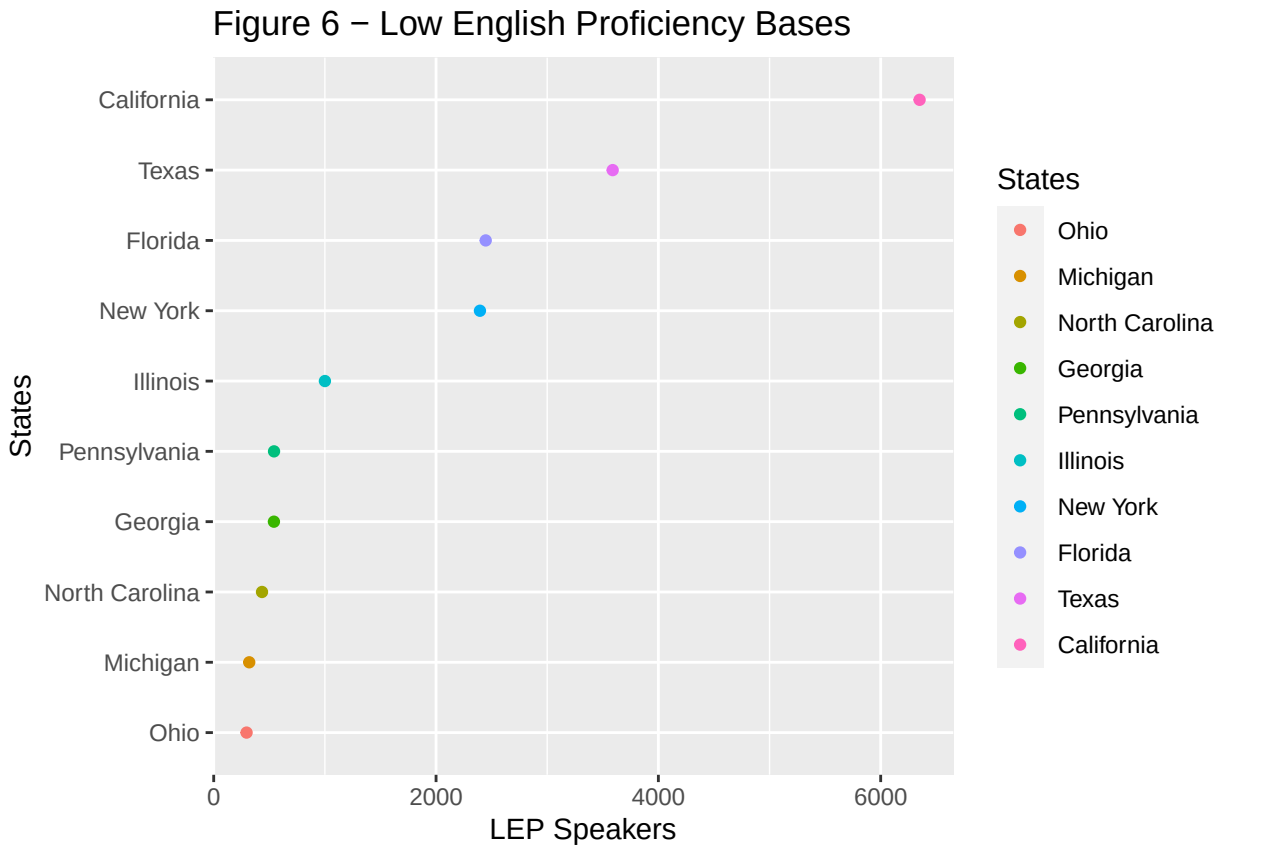
```
LEAP <- read_csv("/Users/iyanuoluwashode/Downloads/LEAP.csv")
```

```
## Rows: 10 Columns: 2
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): States
## dbl (1): LEP
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
View(LEAP)
LEAP%>%
  arrange(LEP/1000)%>%
  mutate(States = factor(States, levels = States))%>%
  ggplot(aes(x=LEP/1000, y=States, color = States))+
  ggtitle("Figure 6 - Low English Proficiency Bases") +
  xlab("LEP Speakers")+
  ylab("States")+
  geom_point()
```



Most LEP speakers are based in California followed by Texas.

4.2. Nativity Status

It will be interesting to observe the nativity status of English speakers that are native and foreign born in the United States.

```
library(readr)
born <- read_csv("/Users/iyanuoluwashode/Downloads/born.csv")
```

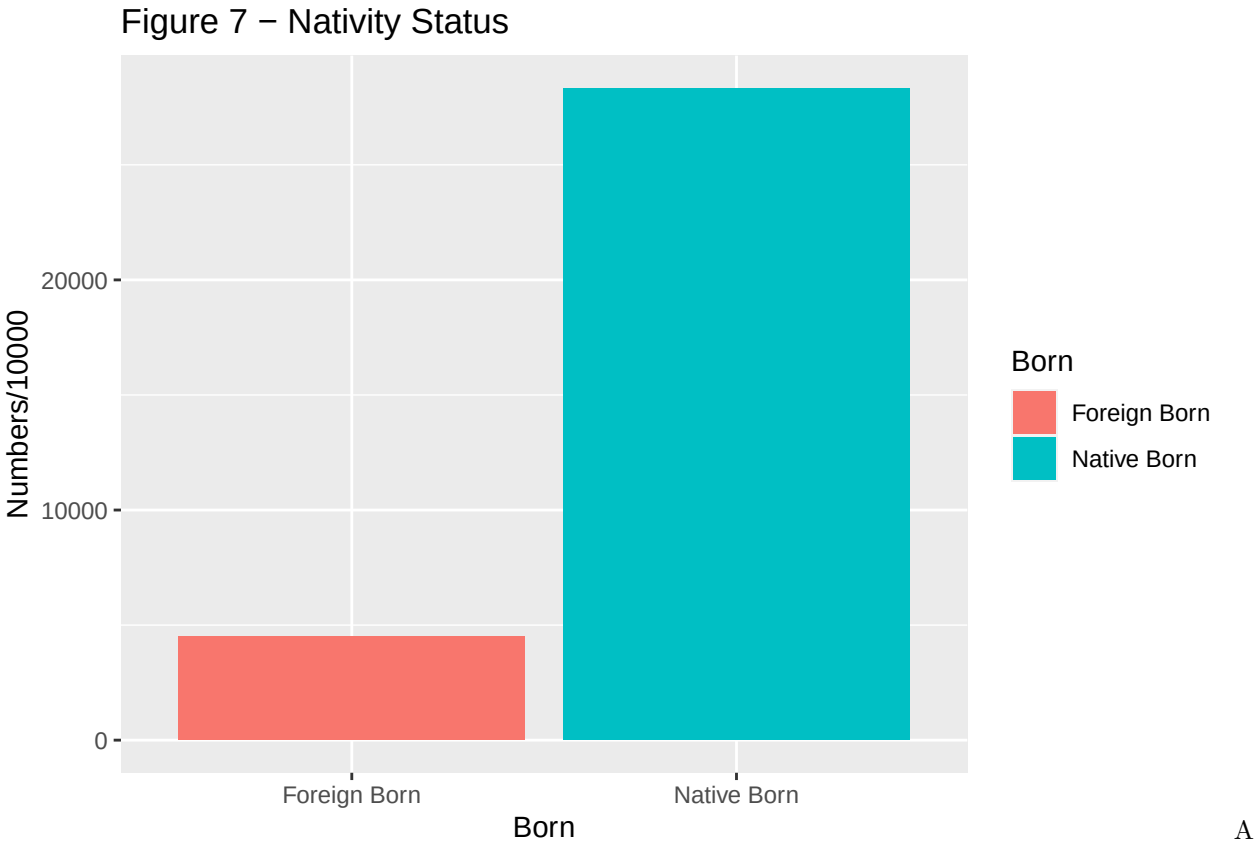
```
## Rows: 2 Columns: 2
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Born
## dbl (1): Numbers
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
View(born)

ggplot(born, aes(x=Born, y = Numbers/10000, fill = Born))+
  ggtitle("Figure 7 - Nativity Status") +
  geom_col()
```



A good number of U.S. residents are foreign born immigrants of the country with over 44 million population.

4.3. Age Distribution of LEP Speakers

```
library(readr)
age <- read_csv("/Users/iyanuoluwashode/Downloads/age.csv")
```

```
## Rows: 3 Columns: 2
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): ages
## dbl (1): not
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
View(age)

ggplot(age, aes(x= ages, y = not/10000, fill = not))+
  ggtitle("Figure 8 - Age Distribution of LEP speakers") +
  xlab("Age Distribution")+
  ylab("Limited English Proficient Speakers")+
  geom_col()
```

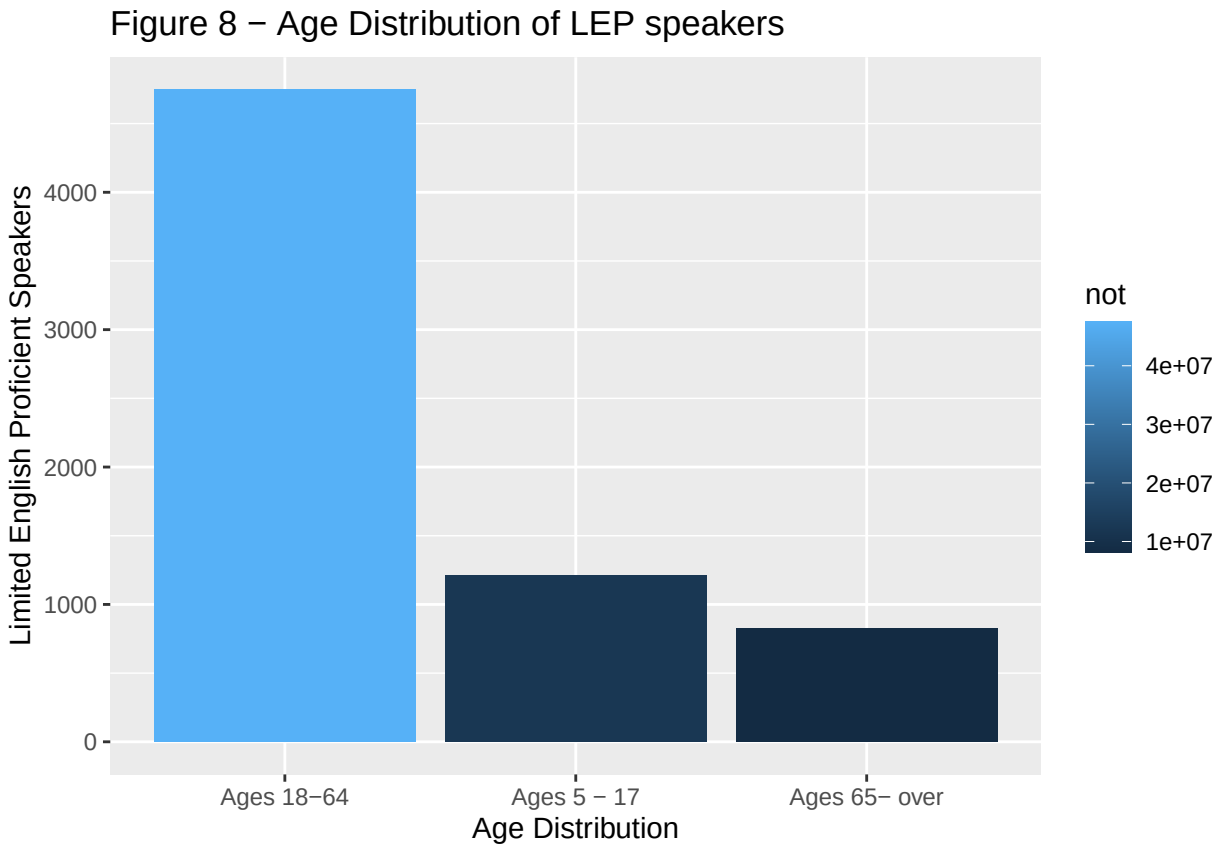


Figure 7 shows that the most limited English proficient speakers are within the age group of 18 - 64 an estimate of over 45 million speakers.

4.4. Education Attainment of LEP Speakers

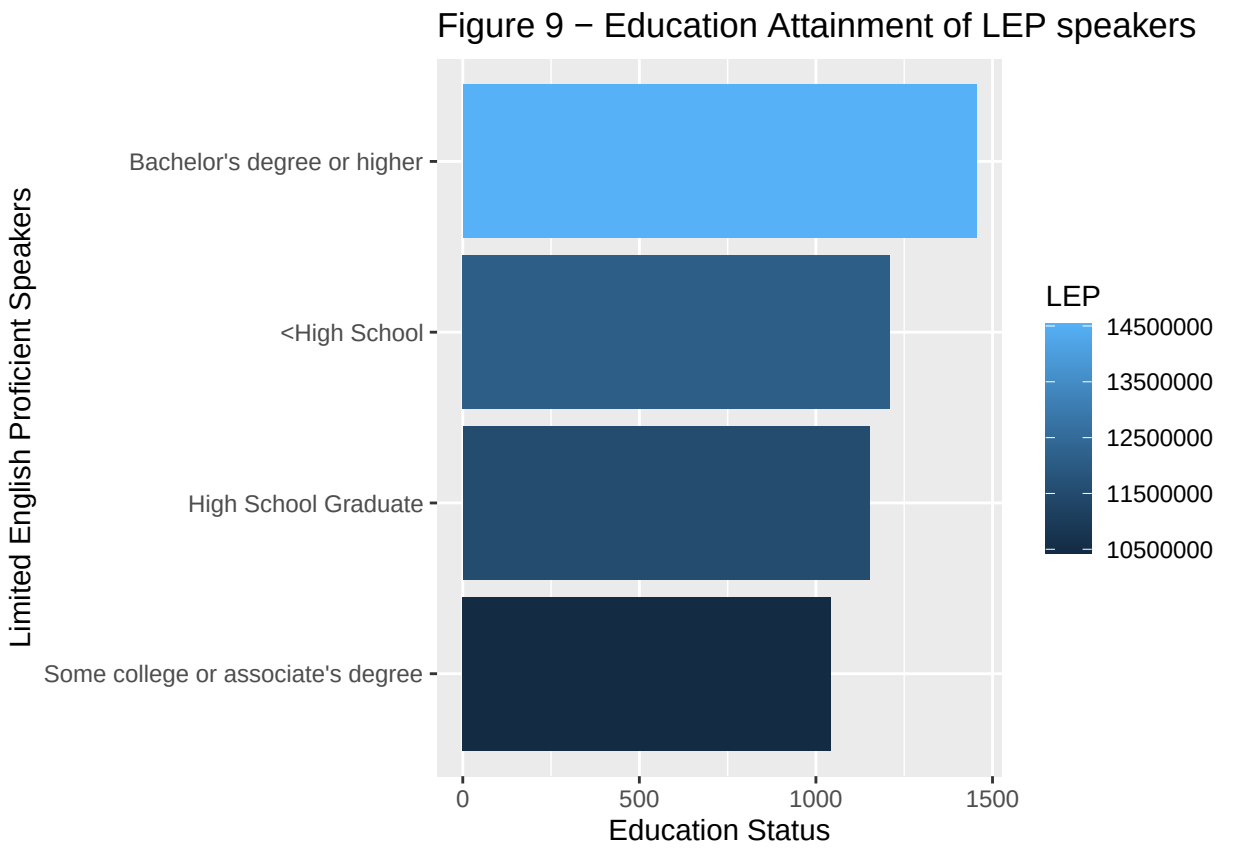
```
library(readr)
degree <- read_csv("/Users/iyanuoluwashode/Downloads/degree.csv")
```

```
## Rows: 4 Columns: 2
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Education
## dbl (1): LEP
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
View(degree)
degree%>%
  arrange(LEP/10000)%>%
  mutate(Education = factor(Education, levels = Education))%>%
  ggplot(aes(x=LEP/10000, y = Education, fill = LEP))+
  ggtitle("Figure 9 - Education Attainment of LEP speakers") +
  xlab("Education Status")+
  ylab("Limited English Proficient Speakers")+
  geom_col()
```

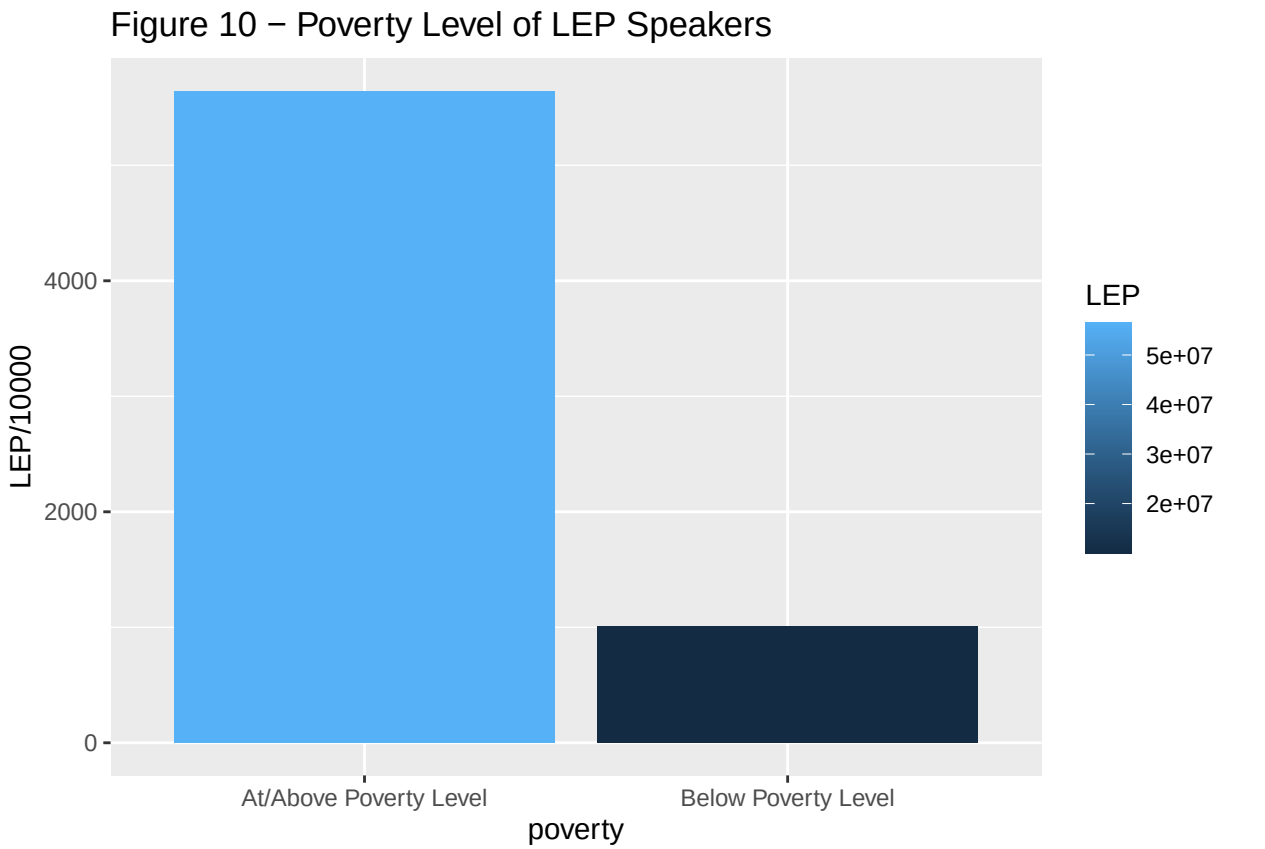


Surprisingly, the class of speakers with a bachelors degree or higher degrees have limited English proficent speakers. This just shows that proficiency in English cannot indicate that one is well educated.

4.5. Poverty Level

```
library(readxl)
pov <- read_excel("/Users/iyanuoluwashode/Downloads/pov.xlsx")
View(pov)

ggplot(pov, aes(x=poverty, y = LEP/10000, fill = LEP))+
  ggtitle("Figure 10 - Poverty Level of LEP Speakers") +
  geom_col()
```



The fact that most of the Limited English Proficient speakers in the United States are above or at the poverty level shows that majority of those that contribute to the economic development of the United States are people that speak English as their second language. This is observed in the figure 10 above.

5.Conclusion

This research futher buttresses the claim that the United States is a polyglot nation with diverse array of languages.At the state level in the United States, Spanish is the only non-English language that plays a significant role. It is spoken by a major part of the population and it should be considered as a second language in the United States of Anerica. This project shows that a good number of the U.S residents aren speakers of other languages apart from English and they should place in consideration by providing effectuve translation and interpretation services for the speakers. Also, since it is noticeable that speakers of a particular language of similar background resides within the same commuiity, focus should be placed on their specified language needs based on regions. For instance, California has a huge number of Spanish speakers compared to Hawaii, more attention on Spanish language development and enhancements should be placed more on Spanish residents in California than Hawaii. Also, major langauges such as Chinese, Tagalog, Vietnamese, Arabic, French should be learnt as well as Spanish since they have many speakers.

It is of utmost importance to help preserve native languages of their foreign speakers of English who have chose to become a part of the United States and invariabily working and having a livelihood in the country which help the economy of the country. Let them feel welcomed by accepting them so that they can help their country of residence grow. It is only wise to consider LEP speakers because they aid the growth and development of the United States of America.

5.References

- U.S. Census Bureau. (2018). Language Spoken at Home by Ability to Speak English for The Population 5 Years and Over. (Table B16001), 2018 American Community Survey 1-Year Estimates.U.S. Department of Commerce. Retrieved November 21, 2021, from https://data.census.gov/cedsci/table?q=B16001 (Links to an external site.)
- U.S. Census Bureau. (2019). Language Spoken at Home by Ability to Speak English for The Population 5 Years and Over. (Table B16001), 2019 American Community Survey 1-Year Estimates.U.S. Department of Commerce. Retrieved November 21, 2021, from https://data.census.gov/cedsci/table?q=B16001
- Strazny, P. (2017). Linguistic diversity in the U.S.Retrieved December 15, 2021, from https://www.thegeogroup.com/us-linguistic-diversity/
- Batalova, J. & Zong, J. (2016). Language Diversity and English Proficiency in the United States. Retrieved December 10, 2021, from https://www.migrationpolicy.org/article/language-diversity-and-english-proficiency-united-states-2015
- Bond, E. (2018). Linguistic Diversity in the US Hits Record High. Retrieved December 10, 2021, from https://slator.com/linguistic-diversity-in-the-us-hits-record-high/