

Language Demographics in the US

Mengye Wei
Jihye Choi
Tom Recht
Terry Kezer

What did you decide
to explore?

For this project we decided to explore recent trends in the population of speakers of foreign languages in the United States, both nationally and at the regional level.

What data sources did you use?

We used the census API.

Specifically, we used the American Community Survey 1-Year Data API for the years 2009-2015 (<https://www.census.gov/programs-surveys/acs>).

This gives annual numbers for speakers of 30 major languages at the state level.

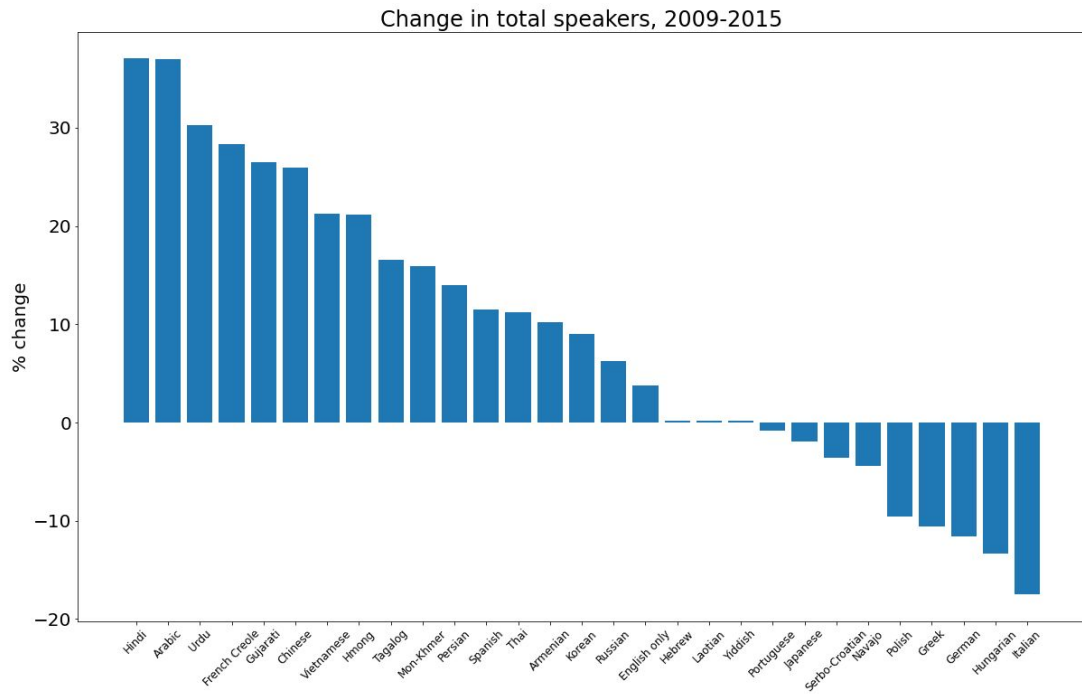
What software/modules did you use in your analysis?

For this project, we used the following software/modules:

- Pandas
- Matplotlib
- Census Module
(<https://pypi.org/project/census/>)
- Gmaps
- NumPy

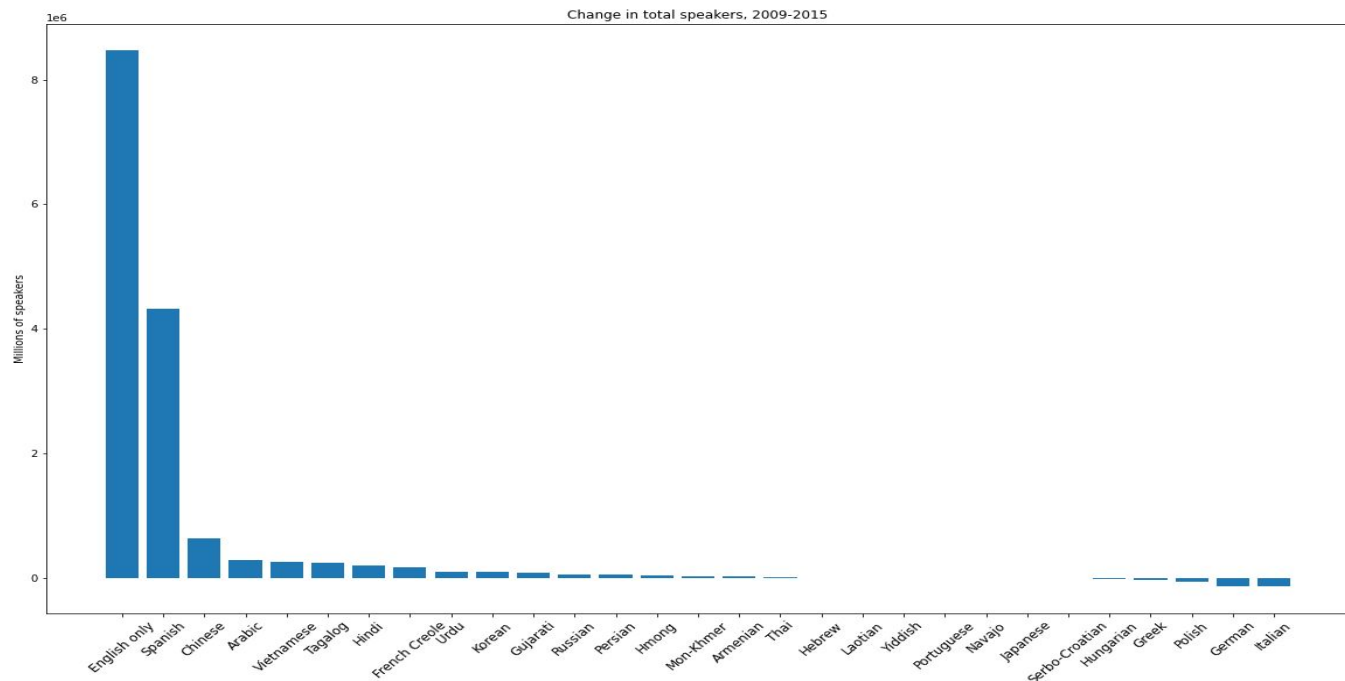
Changes in Languages Spoken

<https://github.com/tomrecht/Project-1/blob/master/NetChangeThru2015.ipynb>



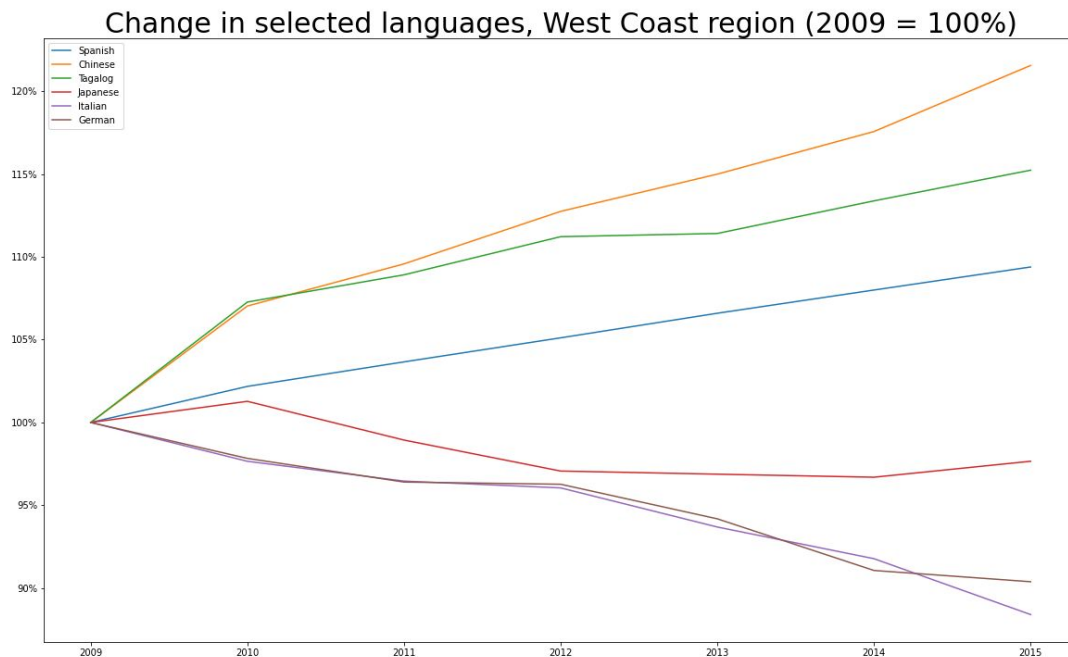
Changes in Languages Spoken

<https://github.com/tomrecht/Project-1/blob/master/NetChangeThru2015.ipynb>



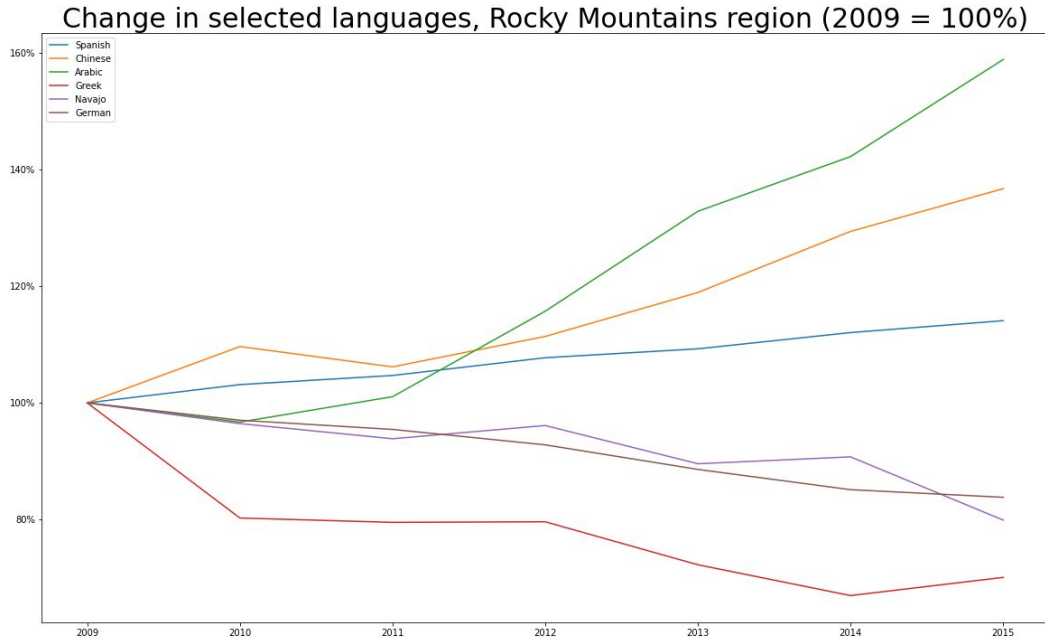
Regional Changes--West Coast

<https://github.com/tomrecht/Project-1/blob/master/Regions.ipynb>



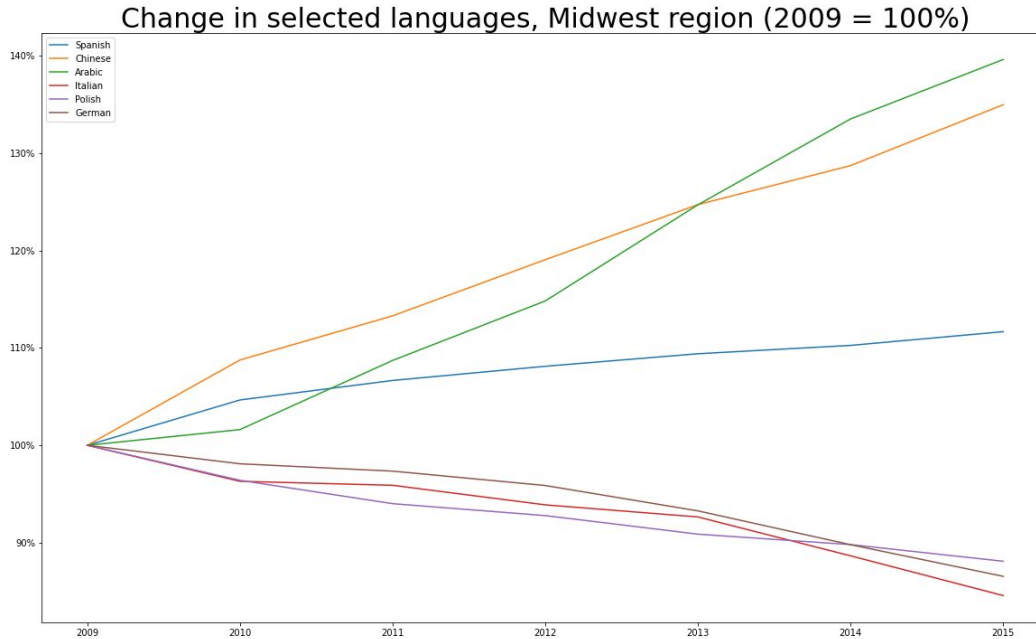
Regional Changes--Rocky Mountains

<https://github.com/tomrecht/Project-1/blob/master/Regions.ipynb>



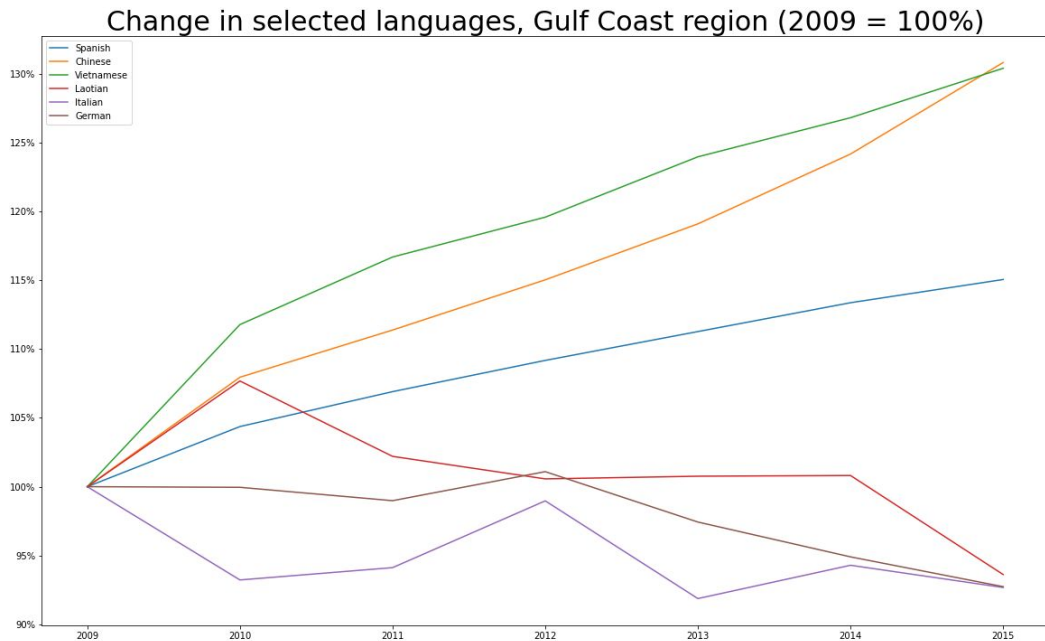
Regional Changes--Midwest

<https://github.com/tomrecht/Project-1/blob/master/Regions.ipynb>



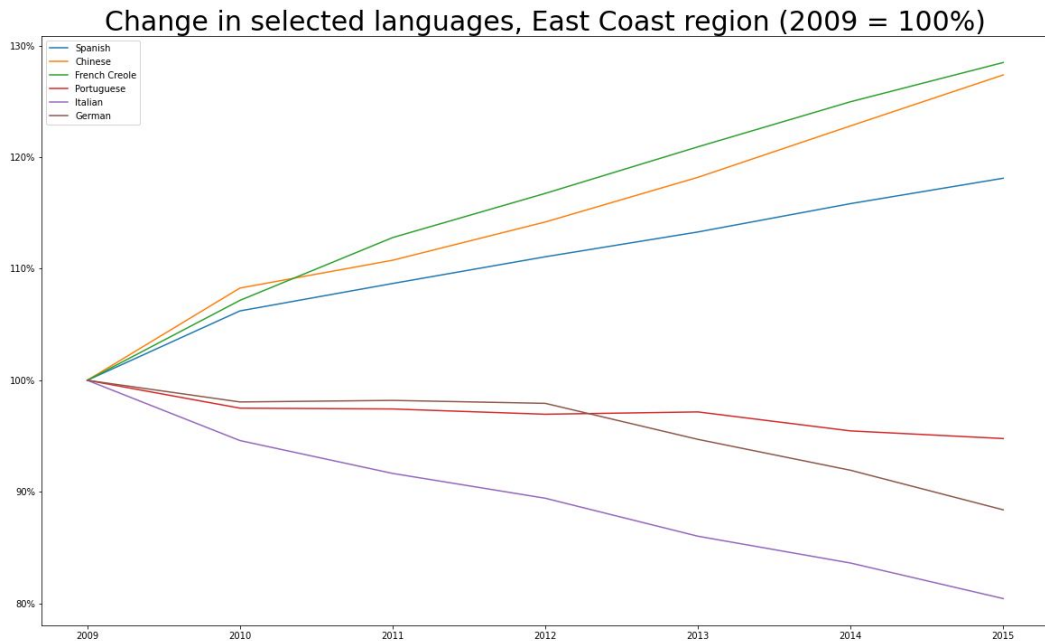
Regional Changes – Gulf Coast

<https://github.com/tomrecht/Project-1/blob/master/Regions.ipynb>



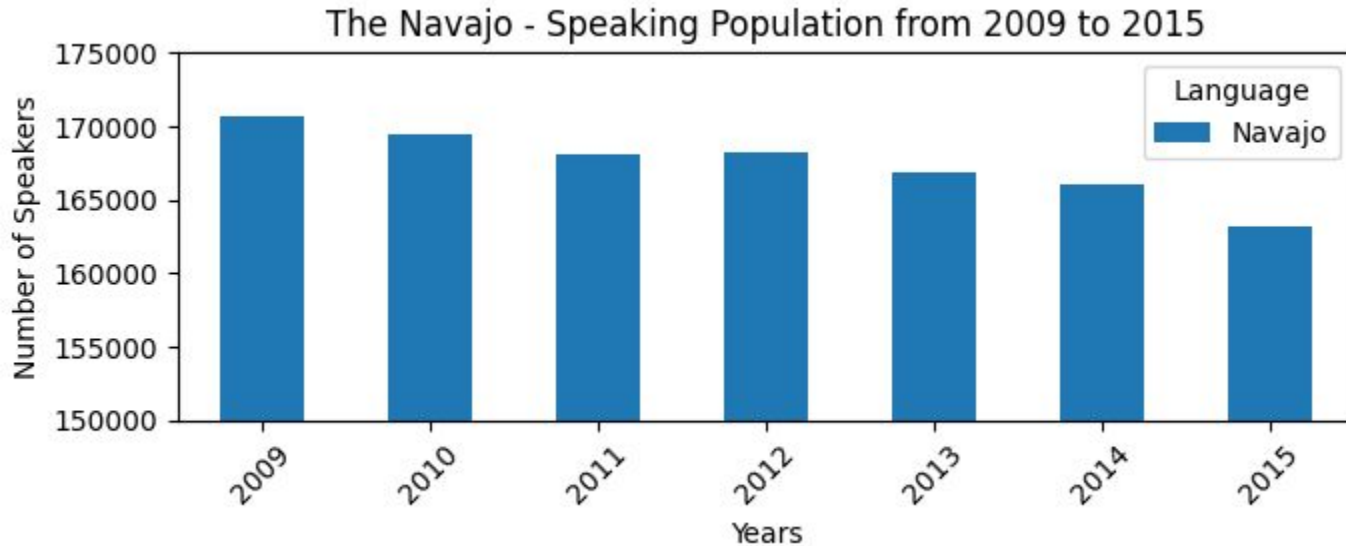
Regional Changes--East Coast

<https://github.com/tomrecht/Project-1/blob/master/Regions.ipynb>



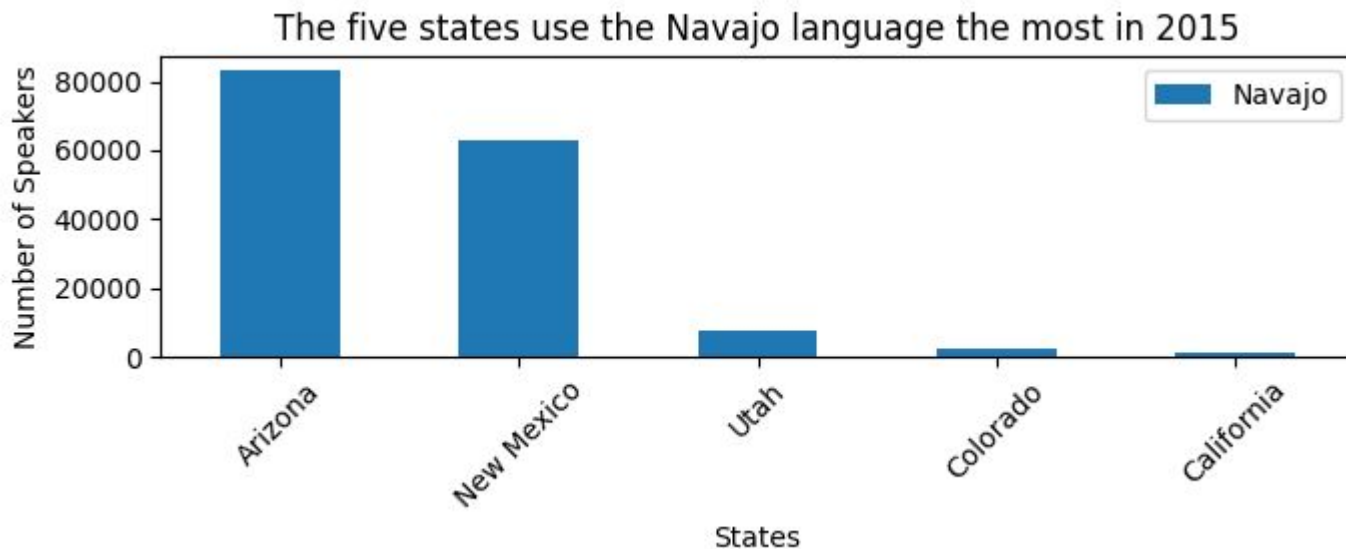
Navajo Speaking Population

<https://github.com/tomrecht/Project-1/blob/jihye/NavajoSpeaker.ipynb>



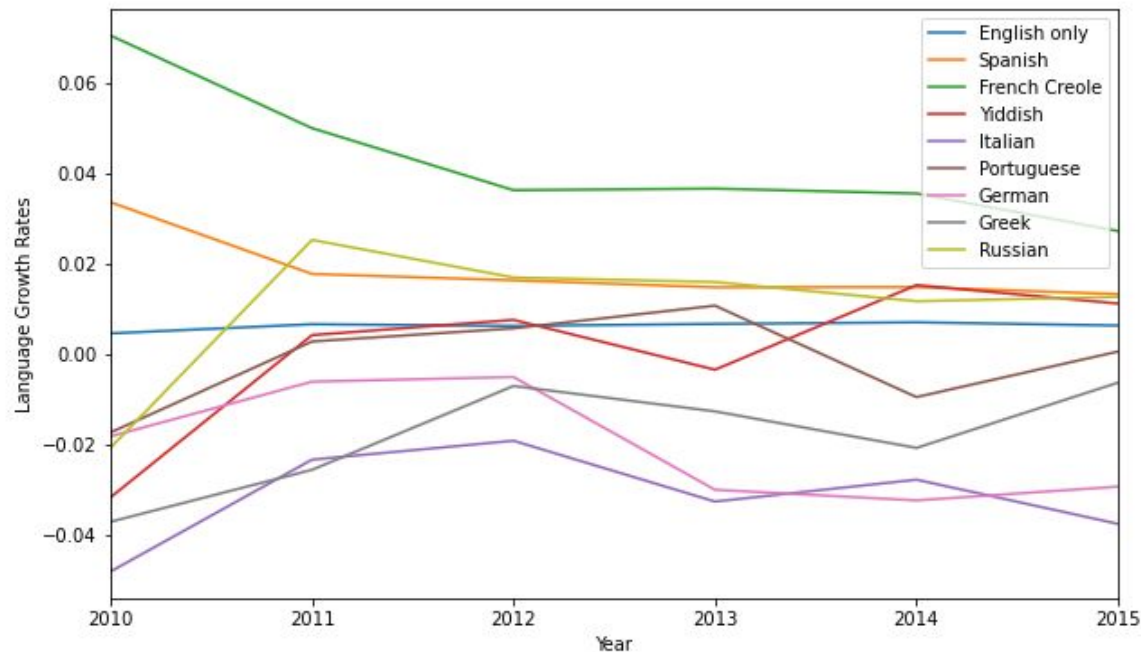
Navajo Speaking 2015

<https://github.com/tomrecht/Project-1/blob/jihye/NavajoSpeaker.ipynb>



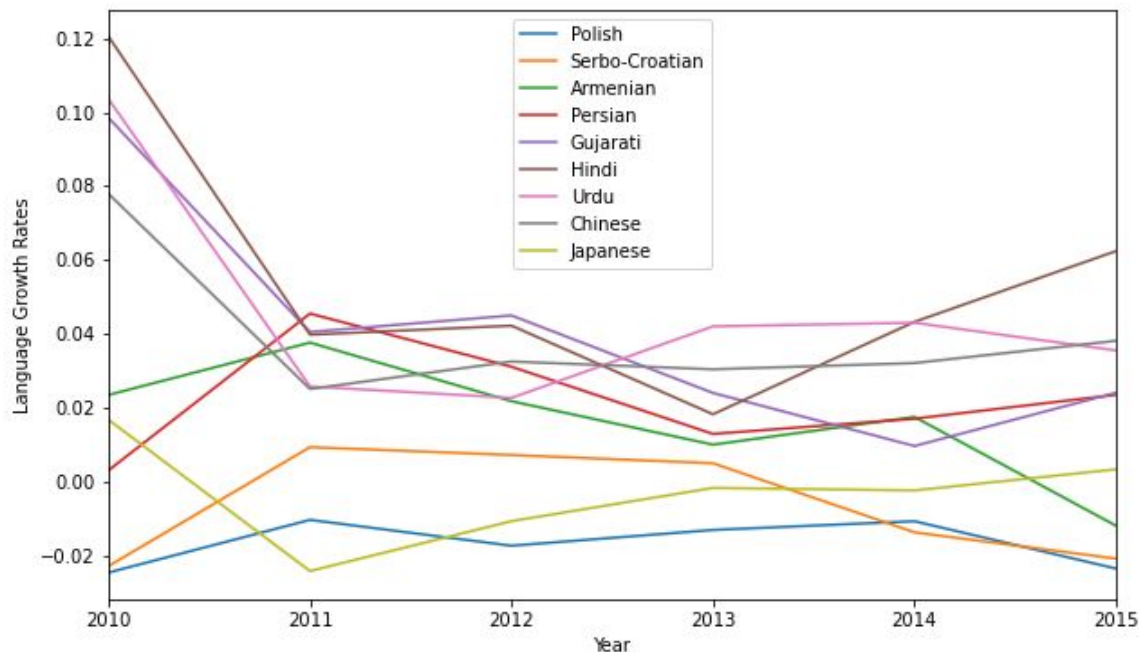
Population Growth Rates

https://github.com/tomrecht/Project-1/blob/master/growth_rates.ipynb



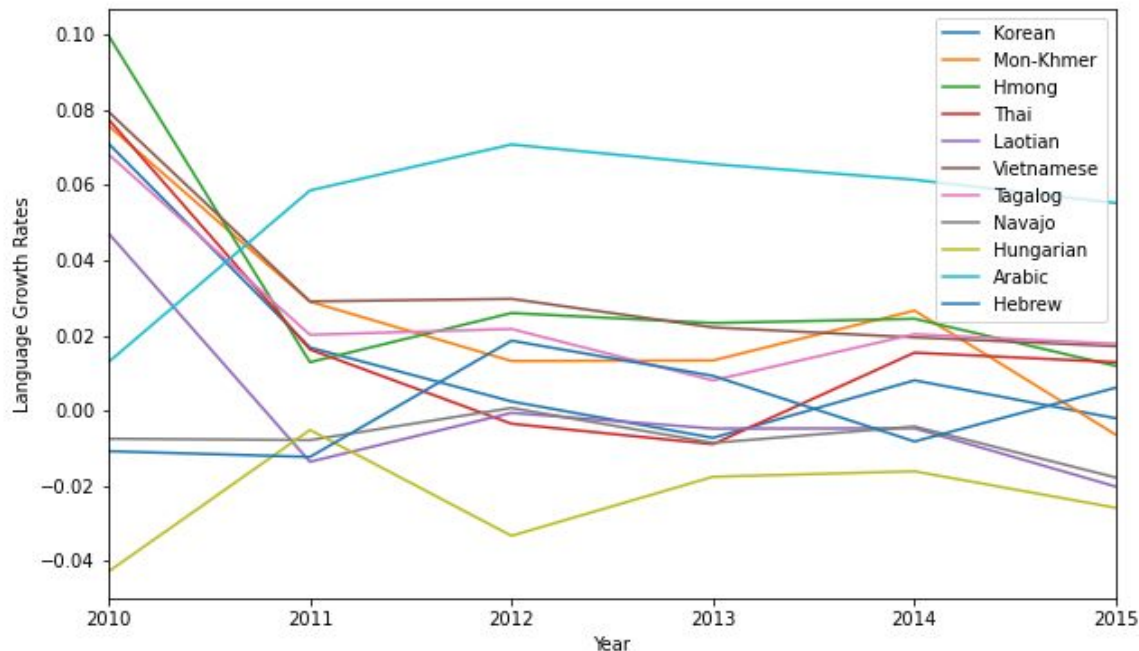
Population Growth Rates

https://github.com/tomrecht/Project-1/blob/master/growth_rates.ipynb



Population Growth Rates

https://github.com/tomrecht/Project-1/blob/master/growth_rates.ipynb



Hindi Speaking Growth (2009 – 2015)

https://github.com/tomrecht/Project-1/blob/master/growth_rates.ipynb



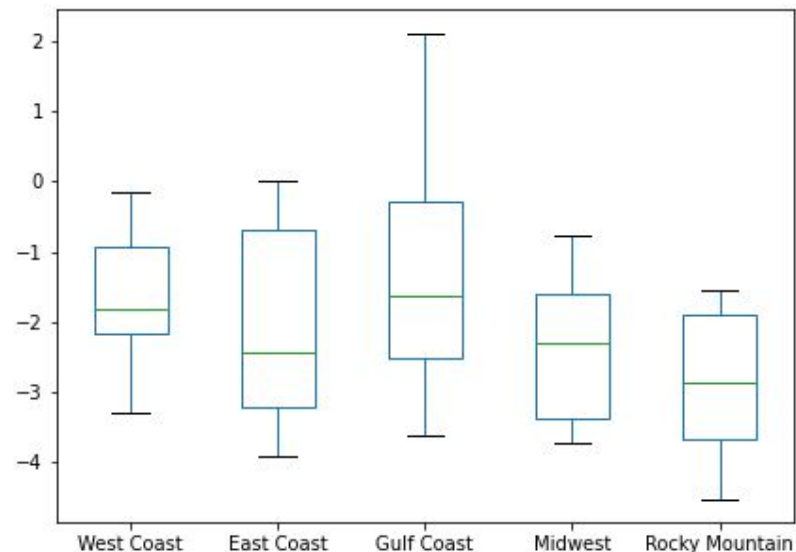
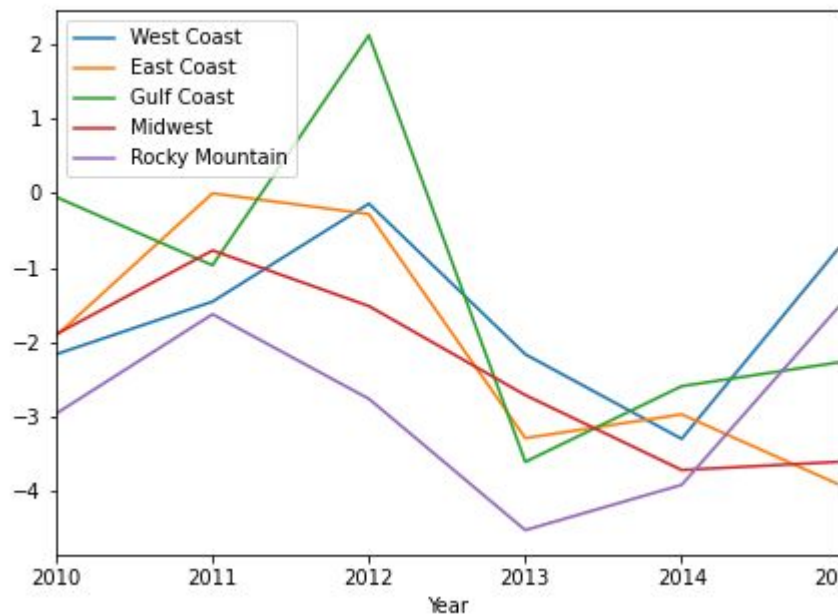
Hypothesis

https://github.com/tomrecht/Project-1/blob/master/Analysis_data.ipynb

H: if the languages spoken on the East Coast have correlation with the languages spoken on the West Coast, Gulf Coast, Midwest, and the Rocky Mountain, then the growth of one will affect the others.

H0: If languages spoken have no correlation between East Coast, West Coast, Gulf Coast, Midwest and Rocky Mountain, then the growth of one won't change the others.

Growth on German



T-test on Hypothesis -- German

```
ret = stats.f_oneway(west, east, rocky, gulf, midw)
ret
```

```
F_onewayResult(statistic=1.1102038913361596, pvalue=0.37376261676954114)
```

```
pvalue_eval(ret[1])
```

```
'0.37376261676954114 is Statistically Non-Significant (failed to reject the null hypothesis)'
```

```
ret1 = stats.ttest_ind(west,east,equal_var=False)
ret1
```

```
Ttest_indResult(statistic=0.4936658335613277, pvalue=0.6334691162726789)
```

```
pvalue_eval(ret1[1])
```

```
'0.6334691162726789 is Statistically Non-Significant (failed to reject the null hypothesis)'
```

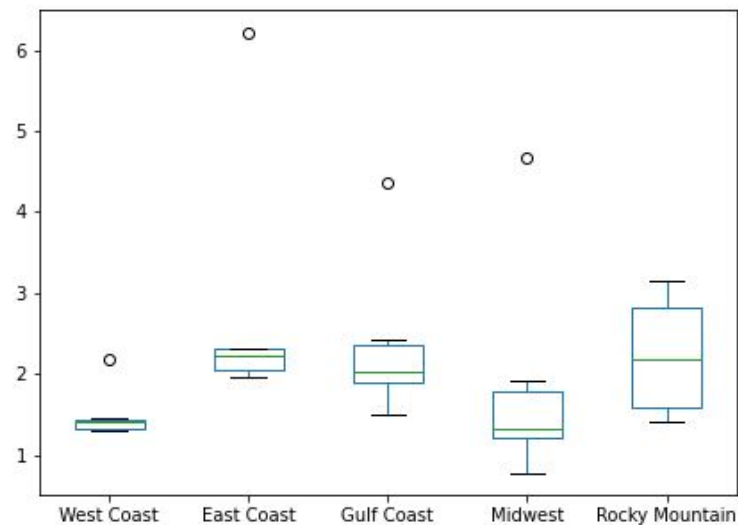
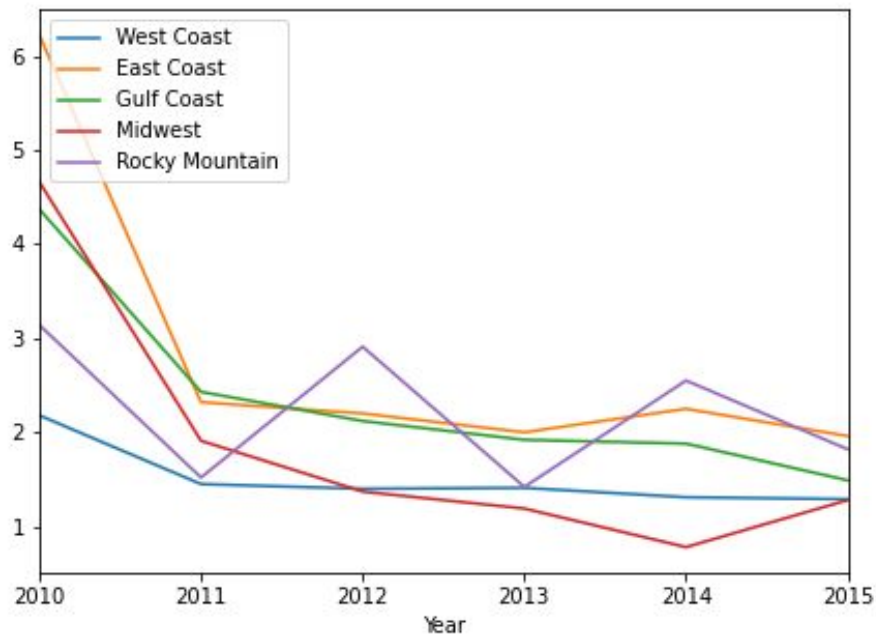
```
ret2 = stats.ttest_ind(west,gulf,equal_var=False)
ret2
```

```
Ttest_indResult(statistic=-0.45138134935039254, pvalue=0.664079999776377)
```

```
pvalue_eval(ret2[1])
```

```
'0.664079999776377 is Statistically Non-Significant (failed to reject the null hypothesis)'
```

Growth on Spanish



T-test on Hypothesis -- Spanish

```
ret = stats.f_oneway(west, east, rocky, gulf, midw)
ret
```

```
F_onewayResult(statistic=1.1572227627082086, pvalue=0.353224492626571)
```

```
pvalue_eval(ret[1])
```

```
'0.353224492626571 is Statistically Non-Significant (failed to reject the null hypothesis)'
```

```
ret1 = stats.ttest_ind(west,east,equal_var=False)
ret1
```

```
Ttest_indResult(statistic=-1.8966643599883541, pvalue=0.11202527927462597)
```

```
pvalue_eval(ret1[1])
```

```
'0.11202527927462597 is Statistically Non-Significant (failed to reject the null hypothesis)'
```

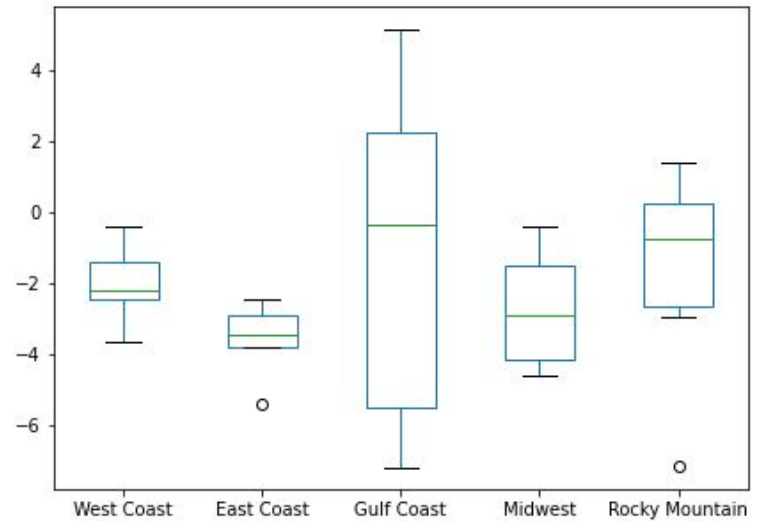
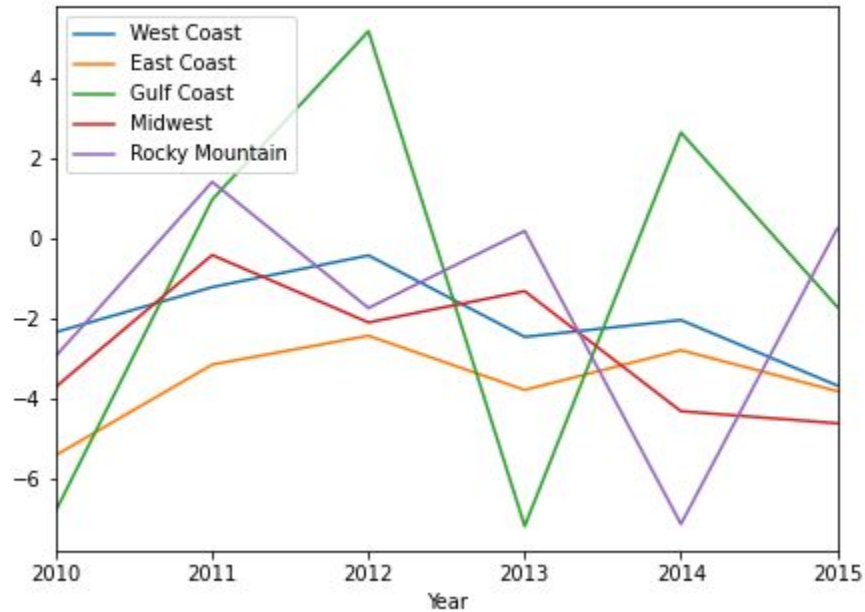
```
ret2 = stats.ttest_ind(west,gulf,equal_var=False)
ret2
```

```
Ttest_indResult(statistic=-1.9519554232976408, pvalue=0.09836450497022176)
```

```
pvalue_eval(ret2[1])
```

```
'0.09836450497022176 is Statistically Non-Significant (failed to reject the null hypothesis)'
```

Growth on Italian



T-test on Hypothesis -- Italian

```
ret = stats.f_oneway(west, east, rocky, gulf, midw)
ret
```

```
F_onewayResult(statistic=0.6638083599870143, pvalue=0.6229966562245504)
```

```
pvalue_eval(ret[1])
```

```
'0.6229966562245504 is Statistically Non-Significant (failed to reject the null hypothesis)'
```

```
ret1 = stats.ttest_ind(west,east,equal_var=False)
ret1
```

```
Ttest_indResult(statistic=2.4530838946396636, pvalue=0.03414639858453946)
```

```
pvalue_eval(ret1[1])
```

```
'0.03414639858453946 is Statistically Significant (reject the null hypothesis)'
```

```
ret2 = stats.ttest_ind(west,gulf,equal_var=False)
ret2
```

```
Ttest_indResult(statistic=-0.41498304789518303, pvalue=0.6938862587016779)
```

```
pvalue_eval(ret2[1])
```

```
'0.6938862587016779 is Statistically Non-Significant (failed to reject the null hypothesis)'
```


Conclusions

- Given the continual decline in Navajo speaking people, greater efforts need to be made to preserve the language.
- There is no correlation between the growth by regions.
- Generally, we see an increase in speakers of Asian languages in the US and a decrease in speakers of European languages. This suggests focusing language education efforts on Asian languages rather than European languages.
- The exception is Spanish, which is the only language to see significant growth both in absolute numbers and in percentage.
- Protect your language = Protect your roots