

v2

Maiko Hata

C. Oregon data by LANGUAGES

Table 1: Initial Oregon Data by Home Languages

Primary Language	Disqualified	exit_total	Moved Out	Not Determined	Not Eligible	Part B Eligible	Withd
Chinese	1	39	5	0	7	23	
English	2337	27697	1582	248	3427	16230	
Other languages	309	4266	322	47	425	2515	
Russian	3	52	8	0	1	28	
Sign languages	1	7	3	0	0	4	
Spanish	318	3895	120	40	388	2651	
Vietnamese	6	133	9	0	5	90	

I first tried to to make a new version of `agg_by_lang` with combined Chinese, Russian, ASL, and Vietnamese. But chi-square didn't run as it still had a cell with "0" for "Not Determined".

So, I had to combine Chinese, Russian, SL, Vietnamese, AND OTHER LANGUAGES which included any combinations of 2 or more or any other languages :(— definitely a limitation!!!

Did I COMBINE IT RIGHT? IS SPANISH STAYING IN THERE? I THINK SO, but WHY IT'S NOT ALPHABETICAL??

Now, I have to make the `agg_by_lang_2` into `corrplot()`

Pearson's Chi-squared test

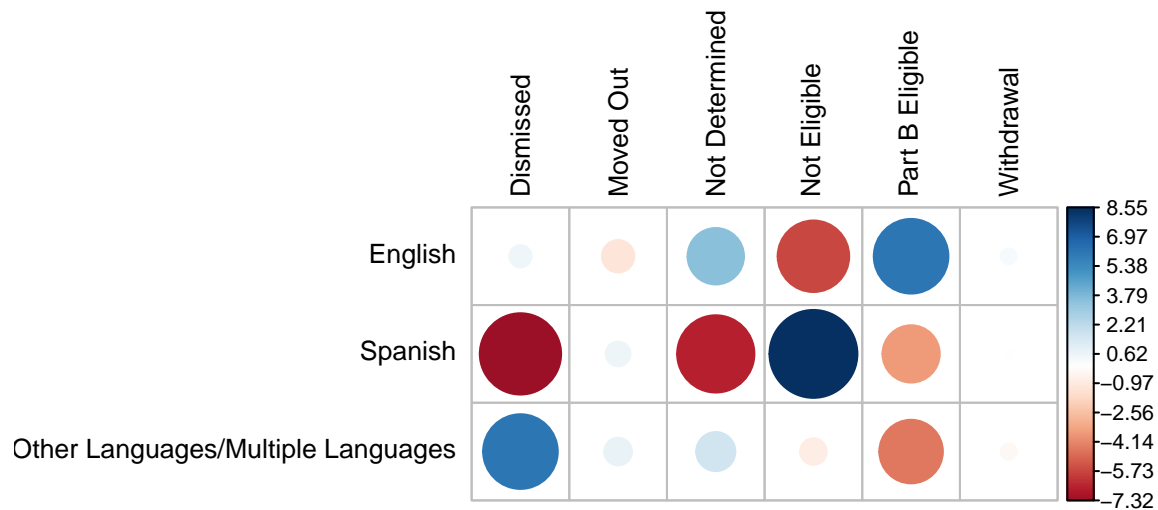
```
data:  agg_by_lang_3_matrix
X-squared = 207.28, df = 10, p-value < 2.2e-16
```

Pearson's Chi-squared test

data: agg_by_lang_3_matrix X-squared = 214.54, df = 10, p-value < 2.2e-16

	exit_total	Moved Out	Not Determined	Not Eligible	Part B Eligible
[1,]	0.27231898	0.5480753	-1.1606662	6.166365	-5.6657070
[2,]	-0.06196474	-7.3166573	0.6741406	-3.634844	8.5538889
[3,]	-0.28995639	6.1669412	0.8509699	-4.471004	-0.7861306

	Withdrawn
[1,]	3.524940
[2,]	-6.604733
[3,]	1.693251



Then, I have to make the agg_by_lang_2 for DQ vs other exits by making calculations that shows that the DQ vs OtherExits