R Notebook

Background on Philippine Election System:

Every three years, 12 senators are chosen by the popular vote (i.e. everyone votes, top 12 contestants with the highest votes win). There are multiple parties in the Philippines, and senators can either run with a party or independently. The members of the parties often change and unlike the US where people have strong pro-Democrat/pro-Republican sentiments, parties aren't as big of a deal in the Philippines.

Background on Data:

The PulseAsia data consists of surveys taken monthly from January to May leading up to the elections which happen in the end of May. Each survey lists the percent of respondents who (1) are aware of and (2) are voting for a certain senator.

Overview of Analysis:

Models were created based on the 2013 PulseAsia surveys to see if we could predict the 12 senators to win the elections. The models were then run on both 2013 and 2016. The general logistic model predicted 11/12 senators for the 2013 data, and 10/12 senators for 2016.

The data sets and libraries were loaded in.

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

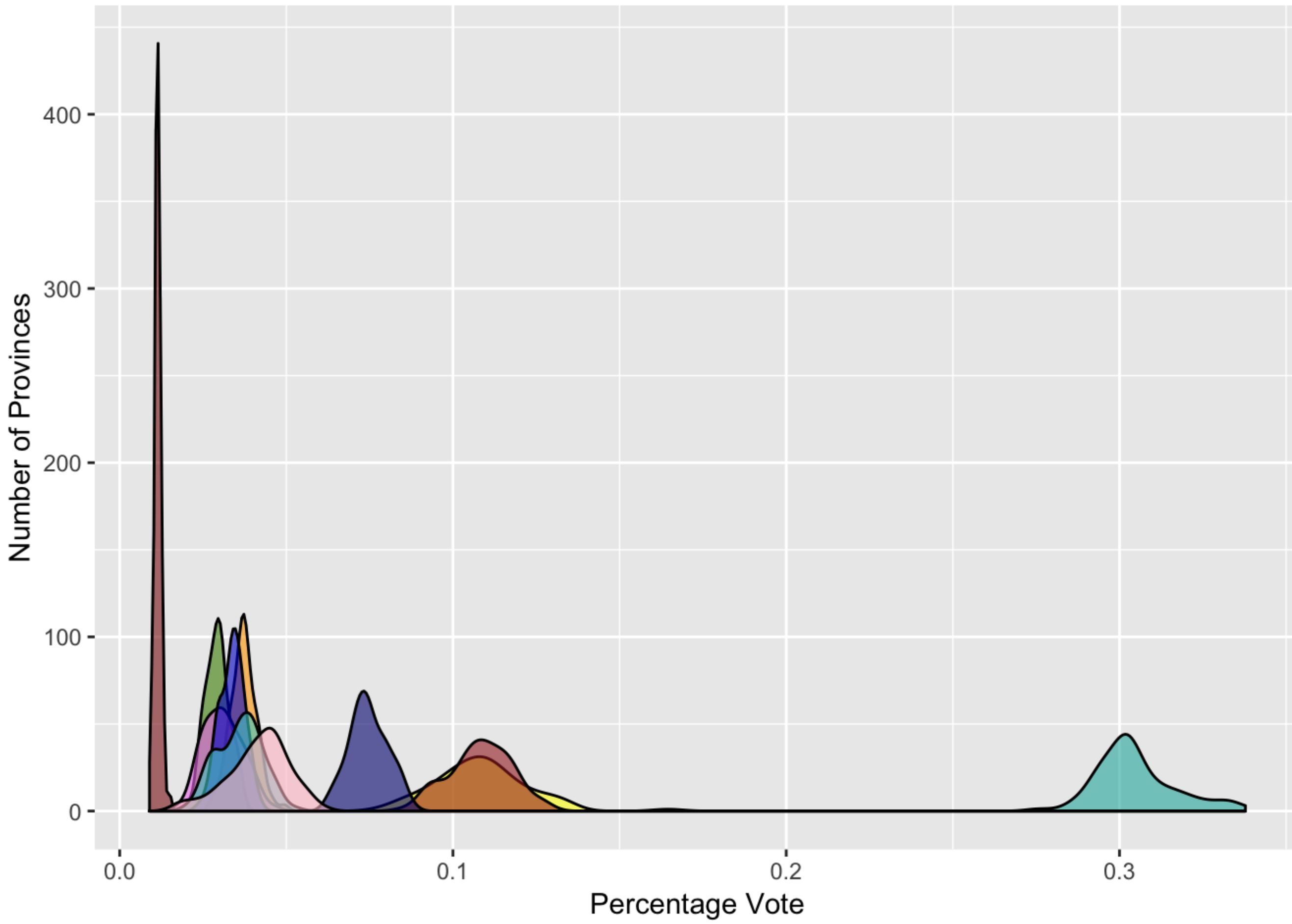
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

The ELEC13 contains the percentage vote for each senator for each province.

The ELEC data frame contains the percentage vote for each party by province.

The density distributions (i.e. histograms of votes but turned into cumulative percentage) were plotted by party

Legend doesn't work, interesting how far LP comes out on top. Also cool how distributions are all unimodal, evidence that in every province there is a certain percentage of voters for each party, interesting since I thought there would be provinces super pro party X and provinces super against party X (which should show a bimodal distribution)



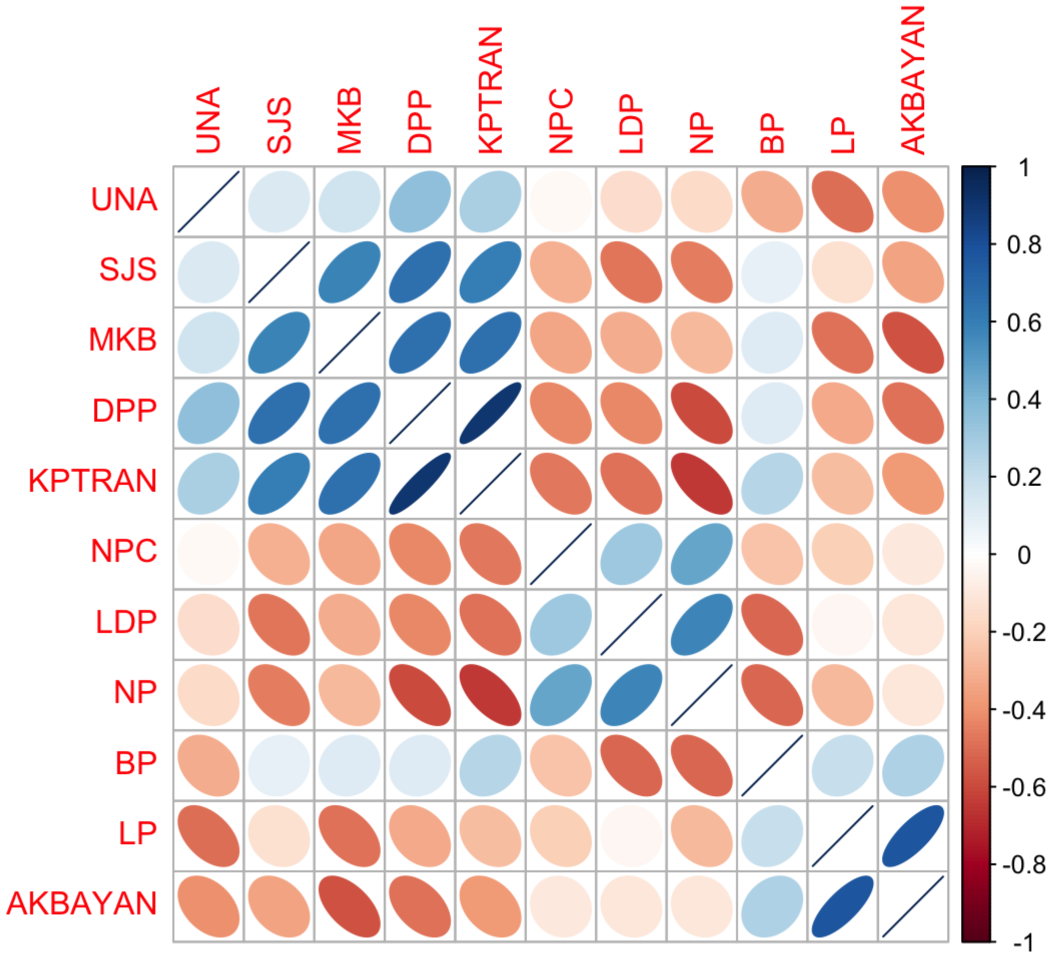
We then analyzed if there are correlations between pairs of votes by province.

SJS, MKB, DPP, and KPTRAN all appear to be strongly positively intercorrelated, as well as the pairing between LP and AKBAYAN.

Strong negative correlations are seen between NP and DPP, and NP and KPTRAN.

Interesting to examine how strong a hold these parties have/do not have on certain regions of the Philippines and how their competitors are able to ally (positive corr.) or block out competition (negative corr.).

```
## corrplot 0.84 loaded
```



Linear model generation based on awareness & vote predictions to predict percentage of final votes using data from January to May from Pulse Asia.

Stepwise regression determined that most important criteria are awareness in January and February (theory behind why?)

Generating general logistic model on election results using January & February awareness and winners of senatorial race 2013 (1 for win, 0 for loss)

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

Result checking (GLM) produces 93% accuracy rate on 2013 data

```
## [1] 0.9393939
```

Used linear and logistic models to predict 2016 senatorial winners.

	name <fctr>	score <dbl>
36	PALPARAN	0.04599592
33	PACQUIAO	0.04259244
21	LACSAMANA	0.04200215
27	MANZANO	0.04150178

	name <fctr>	score <dbl>
24	LAPID	0.03931996
40	ROMUALDEZ	0.03925736
12	DRILON	0.03887960
39	RECTO	0.03828359
37	PANGILINAN	0.03776460
14	GATCHALIAN	0.03768671

1-10 of 12 rows Previous **1** 2 Next

Linear Model: 6/12

1. PALPARAN
2. PACQUIAO
3. LACSAMANA
4. MANZANO
5. LAPID
6. ROMUALDEZ
7. DRILON
8. RECTO
9. PANGILINAN
10. GATCHALIAN
11. LANGIT
12. HONTIVEROS

```
sen13res <- predict(combinedResult, PULSE[,c(2:11)])
pred13res <- data.frame(name = PULSE$X, score = sen13res)
pred13res <- pred13res[order(-pred13res$score),]
pred13res
```

	name <fctr>	score <dbl>
28	POE	5.5838337
19	LEGARDA	4.7454041
32	VILLAR	4.2108997
11	EJERCITO	3.6004140
3	AQUINO	2.9506200

	name <fctr>	score <dbl>
13	ESCUDERO	2.3794426
22	MADRIGAL	1.5776860
17	HONASAN	1.5017047
5	BINAY	1.3104338
7	CAYETANO	1.2344525

1-10 of 33 rows Previous **1** 2 3 4 Next

Logistic Model: 10/12

1. PACQUIAO
2. LACSON
3. SOTTO
4. RECTO
5. DRILON
6. PANGILINAN
7. MANZANO
8. LACSAMANA
9. ZUBIRI
10. DELIMA
11. GORDON
12. GATCHALIAN