

GOP nominee predictions ordination

in response to Jennifer N. Victor's October 30 2015 article

The equivocal nature of the collected predictors suggests that at this point the predictors to use are sort of a matter of taste. Without more elections data it is very hard to say anything meaningful. However, one can at least look a little at the covariance structure of the various predictors and try to understand where each candidate lies in predictor space, and have an idea of which predictors agree.

I thought that a simple ordination analysis (using principal component analysis or PCA) could shed some light on this question, using data transcribed from the Victor article.

```
gop = read.table('gop.txt',header=T)
print(gop) # look at it

##           polls endorse  cash predictit debate
## trump      26.8        0  5.80      0.22    464
## carson      22.0        0 31.40      0.15    328
## rubio       9.0         8 15.50      0.47    524
## bush        7.0        36  2.48      0.27    294
## cruz        6.6         8 26.60      0.22    412
## fiorina     5.8         3  8.50      0.06    517
## huckabee    3.8        24  3.20      0.03    347
## paul        3.4        15  9.40      0.08    303
## kasich      2.6        14  4.40      0.16    486
## christie    2.4        25  4.20      0.14    390

plot(gop) # plot it

gop_pca = prcomp(gop,scale=T) # PCA
plot(gop_pca) # variances associated with each PC

biplot(gop_pca) # visualize placement of candidates in predictor

# space of first two (most important) PCs.
print(gop_pca$rotation) # PC loadings of predictors
```

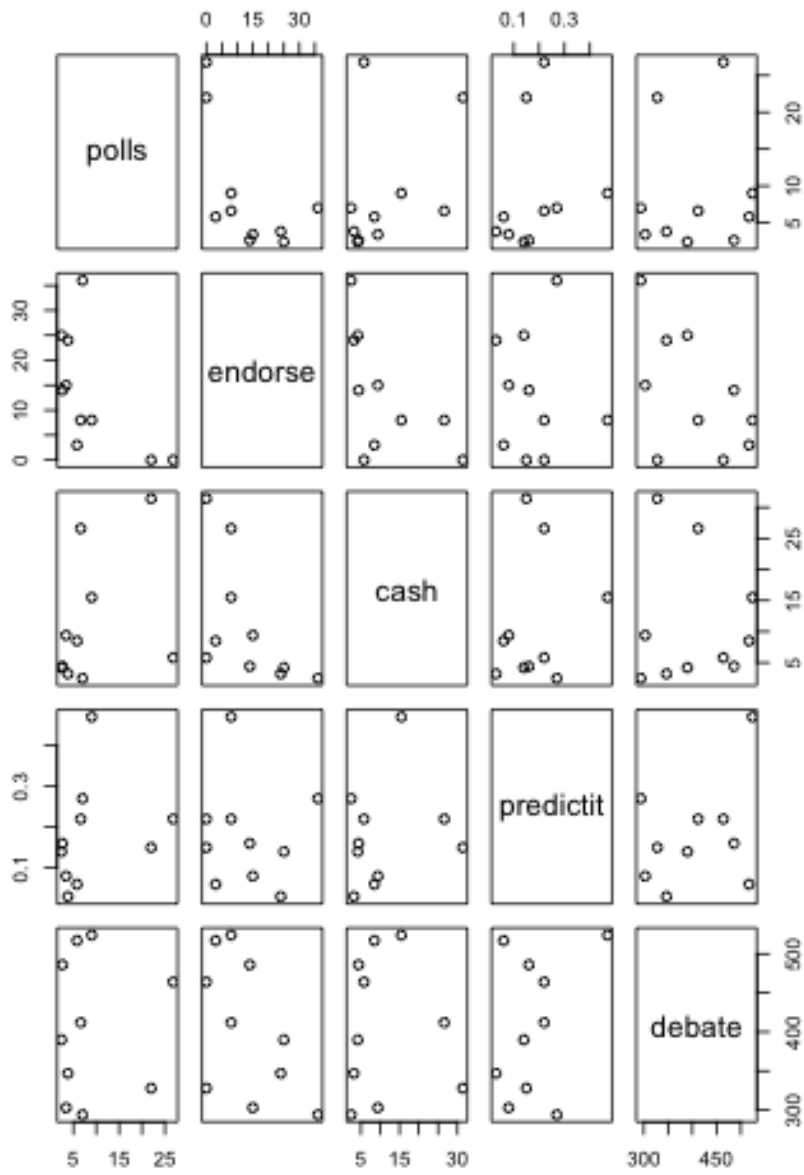


Figure 1: plot of chunk unnamed-chunk-1

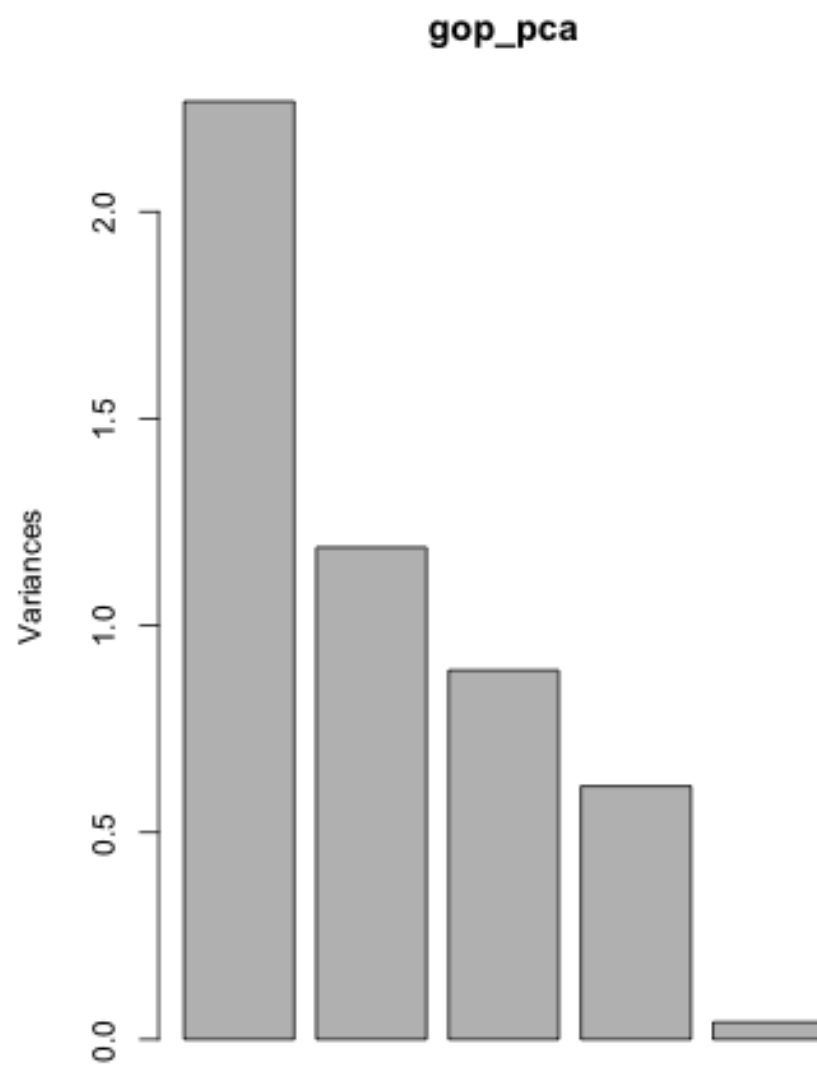


Figure 2: plot of chunk unnamed-chunk-1

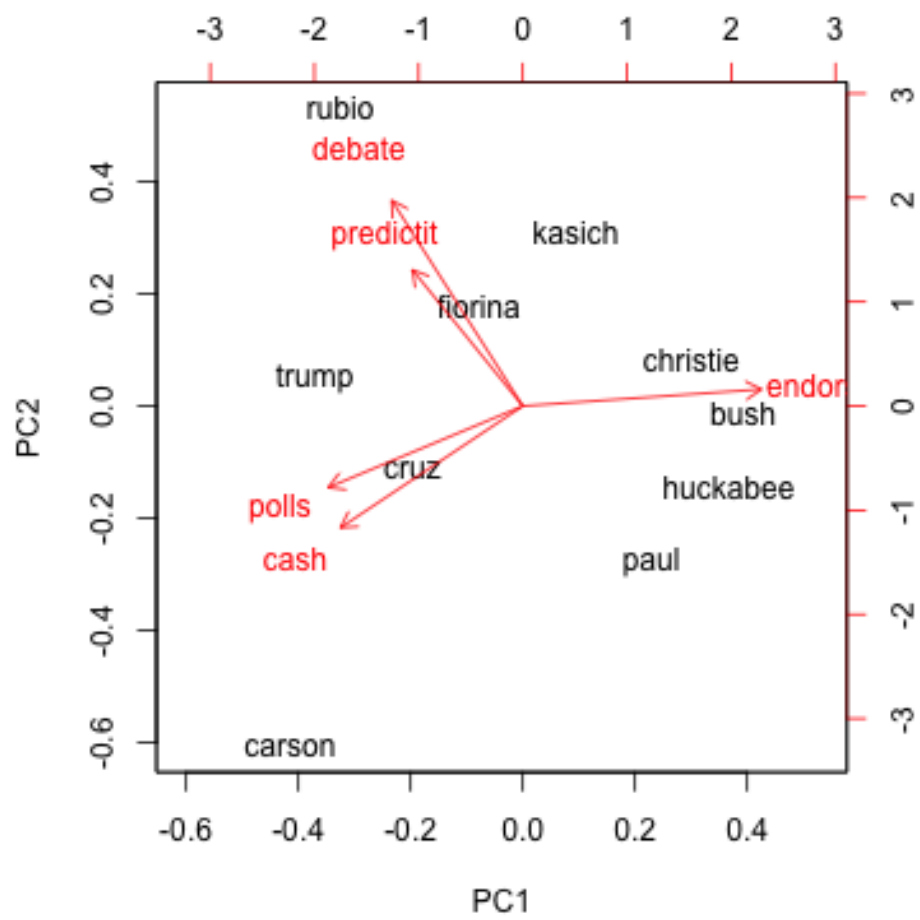


Figure 3: plot of chunk unnamed-chunk-1

##	PC1	PC2	PC3	PC4	PC5
## polls	-0.4901170	-0.28334132	0.1083490	-0.751417085	-0.3211477
## endorse	0.6018794	0.05815992	0.4148112	-0.065457342	-0.6767612
## cash	-0.4593028	-0.42343496	0.2238781	0.649882670	-0.3705063
## predictit	-0.2783308	0.47405495	0.7882498	-0.001733093	0.2765196
## debate	-0.3303021	0.71575823	-0.3804379	0.093473054	-0.4744686

So, this looks pretty much like the conventional wisdom. The biggest PC by far is the first, which seems to represent the insurgent/establishment gradient. PC2, next biggest, seems to represent “professionalism” if you just look at the outliers (Carson + Rubio), but maybe it more clearly represents technocratic vs. ideological distinctions. None of this is new of course. Doing well in the debate, looking like a real politician to the betting marketeers, etc., look good to folks who are paying attention, but who knows what they mean to the average primary voter.

According to at least this projection of the data, the debate/predictit variables are pretty much collinear, and cash/polls also. Endorsements seem to be their own thing. You have to dig down into PC3 before they show any substantial positive covariance with other variables, where it shows a relationship to prediction markets. But if you dig through enough PCs you can find any signal you want, so I wouldn’t put too much on that.

Also a little misleading, as pointed out by Victor in the article, is that Trump actually has a self-funded campaign. So in reality his cash pile is a lot bigger. I repeated the analysis after fudging Trump’s number upwards (note that this decision and its exact quantity is totally ad hoc), and it looks similar, except that Trump is really pushed out in crazyland where he belongs, and the correlation between cash+polls is a bit tighter. It also moves Cruz up a bit, interestingly.

```
gop_trumped = gop
gop_trumped['trump', 'cash'] = gop_trumped['trump', 'cash'] + 30
print(gop_trumped)
```

##	polls	endorse	cash	predictit	debate
## trump	26.8	0	35.80	0.22	464
## carson	22.0	0	31.40	0.15	328
## rubio	9.0	8	15.50	0.47	524
## bush	7.0	36	2.48	0.27	294
## cruz	6.6	8	26.60	0.22	412
## fiorina	5.8	3	8.50	0.06	517
## huckabee	3.8	24	3.20	0.03	347
## paul	3.4	15	9.40	0.08	303
## kasich	2.6	14	4.40	0.16	486
## christie	2.4	25	4.20	0.14	390

```

gopt_pca = prcomp(gop_trumped, scale=T)
plot(gopt_pca) # variances associated with each PC

biplot(gopt_pca) # visualize placement of candidates in predictor

# space of first two (most important) PCs.
print(gopt_pca$rotation) # PC loadings of predictors

```

	PC1	PC2	PC3	PC4	PC5
polls	-0.5181937	-0.34065395	0.2254708	-0.7152946	0.2301016
endorse	0.5367870	-0.01597172	0.4418513	-0.4277386	-0.5774181
cash	-0.5633849	-0.26872683	0.1100763	0.3487253	-0.6904037
predictit	-0.2252455	0.53969217	0.7576409	0.2094106	0.2003106
debate	-0.2742206	0.72126143	-0.4096264	-0.3740739	-0.3112233

What does it mean for who is winning? Unclear. But by looking at the leading edges of the main components, you can at least pick the most likely candidate according to whatever weighting of the linear combination of predictors you like. If you think looking presidential is important, it's Rubio. If you think that outsideriness is important, it's a tossup between Trump and Carson. If you like endorsements and party-decidiness (which you probably shouldn't), you have a pack of white guys.

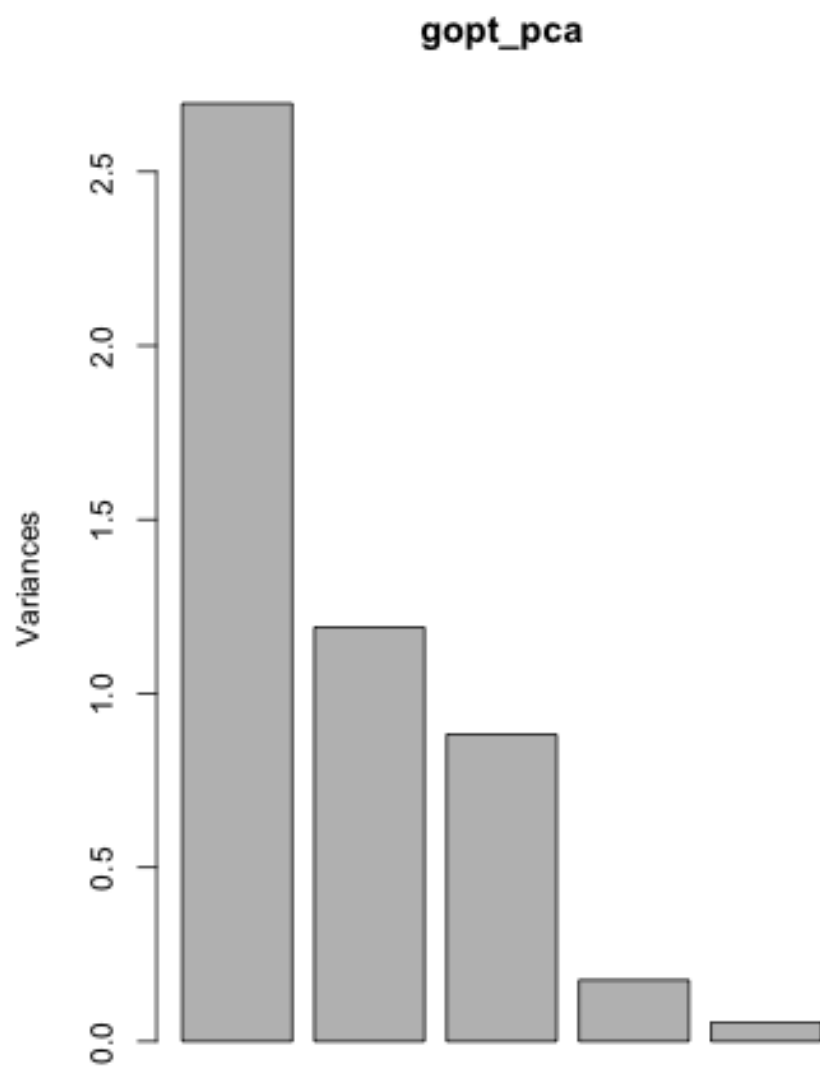


Figure 4: plot of chunk unnamed-chunk-2

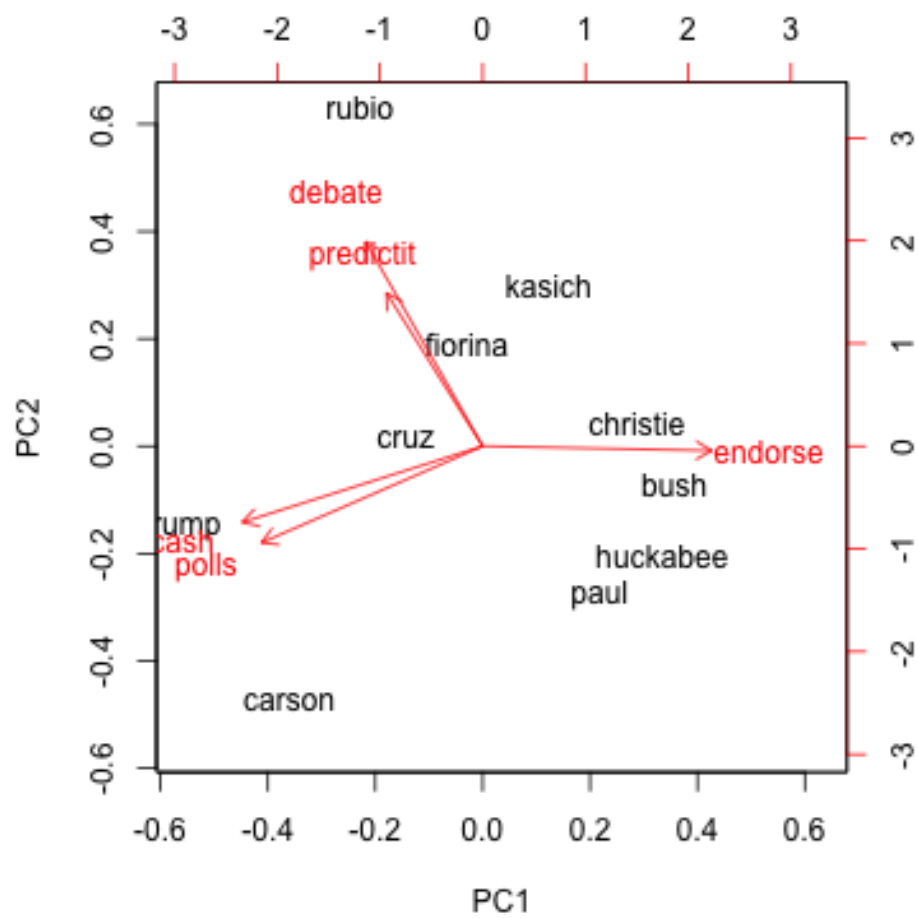


Figure 5: plot of chunk unnamed-chunk-2