A First Look at the California Political Precinct Index

David Latterman, Fall Line Analytics
dlatterman@flanalytics.com

Summary

Presented here is a first look at a statewide voter index, called the California Political Precinct Index (CPPI)[1]. Modeled after the San Francisco Progressive Voter Index (PVI), the CPPI is the first known attempt to rate the political ideology of nearly all California precincts. Even though the data are not quite as strong as the San Francisco data that go into the PVI, and come from one year (2004), results and analyses indicate a very good fit with known statewide political trends. More work needs to be done to hone this new and important tool.

I also show a few selected analyses using the CPPI of counties, State Assembly districts, the 2004 Presidential race, and Latino voting patterns.

Background

The Progressive Voter Index, originally developed by professor Rich DeLeon at SFSU several years ago, uses a factor analysis of myriad San Francisco ballot measures over several years to score precincts on a 0-100 political ideological scale (higher number = more 'liberal'). Its analytical strength comes from 1) the many ballot initiatives San Francisco has each year; 2) the consistency of precinct lines for many years on end; 3) relatively clear ideological interpretations of the San Francisco electorate. The most recent paper concerning the details and the methodology of the PVI can be found here.

Creating a statewide index, by utilizing over 20,000 precincts with fewer ideological initiatives, is more difficult. For starters, precinct lines move from year to year for most California counties. Also, there are fewer initiatives with clear ideological underpinnings. For instance, abortion measures are pretty clear, but it is difficult to interpret many other measures with certainty. By and large, we're left with a few social measures and bond measures (where *generally*, yes = liberal). Still, the prospect of a 20,000+ precinct database, in which the precincts could be aggregated into many other state and national political districts proved worth the effort.

Methodology

A detailed methodology sheet is available upon request. The following is a summary of the methodology used.

1) Data and GIS shapefiles were taken from the UC Berkeley Statewide Database. The November 2004 race is currently the last available election for which there is a complete data set.

[1] In many of the charts and graphs, the CPPI is referred to as the CVI.

2) Because of changing precincts from year to year, one year was chosen for ballot measures. 2004 had maximum turnout – as it was a Presidential election year - and therefore the most precincts. It is most directly comparable with next year's (2008) primary and general elections.

3) Precincts with more than 20 votes cast were chosen for the analysis, yielding 20,089 precincts. This was over 90% of precincts that had registered voters.

4) After many experimental iterations, the following issues were chosen:

   - PROP 61 - Children's Hospital Projects
   - PROP 63 - Mental Health Expansion and Funding
   - PROP 65 - Local Gov't Funds and Revenues
   - PROP 66 - Limitations on "Three Strikes" Law
   - PROP 67 - Emergency and Medical Services
   - PROP 69 - DNA Samples Collection Database
   - PROP 71 - Stem Cell Research Funding Bonds
   - PROP 72 - Referendum Health Care Coverage

   This suite of measures had a strong Cronbach's Alpha of 0.83, which is a measure of internal consistency. All of the measures were somewhat easily interpretable on a left-right scale. Attempts were made folding in other years' measures but the inconsistencies in the precincts lines made this difficult.

5) The measures were analyzed using a factor analysis, which resulted in only **one** factor with an eigenvalue > 1. No principle components rotation was necessary. For each precinct, scores were assigned via regression method to the single factor and normalized to a 0-100 scale.

Results

Below are two maps of the CPPI for the state of California, with ten categories, in two different color scales. Some trends are clear: the coast is far more liberal than the interior part of the state. While this trend has long since been understood, we now have the ability to compare different parts of the state, where precincts can be aggregated by numerous other geographies.

After the California map, I show a blowup of San Francisco and the San Francisco PVI map for comparison (and to check validity against a time-tested regional index). Although the scales are different, and even the most 'conservative' San Francisco precincts are relatively liberal statewide, the geographic patterns match up very well.

Figure 1: California CPPI (spectrum color)

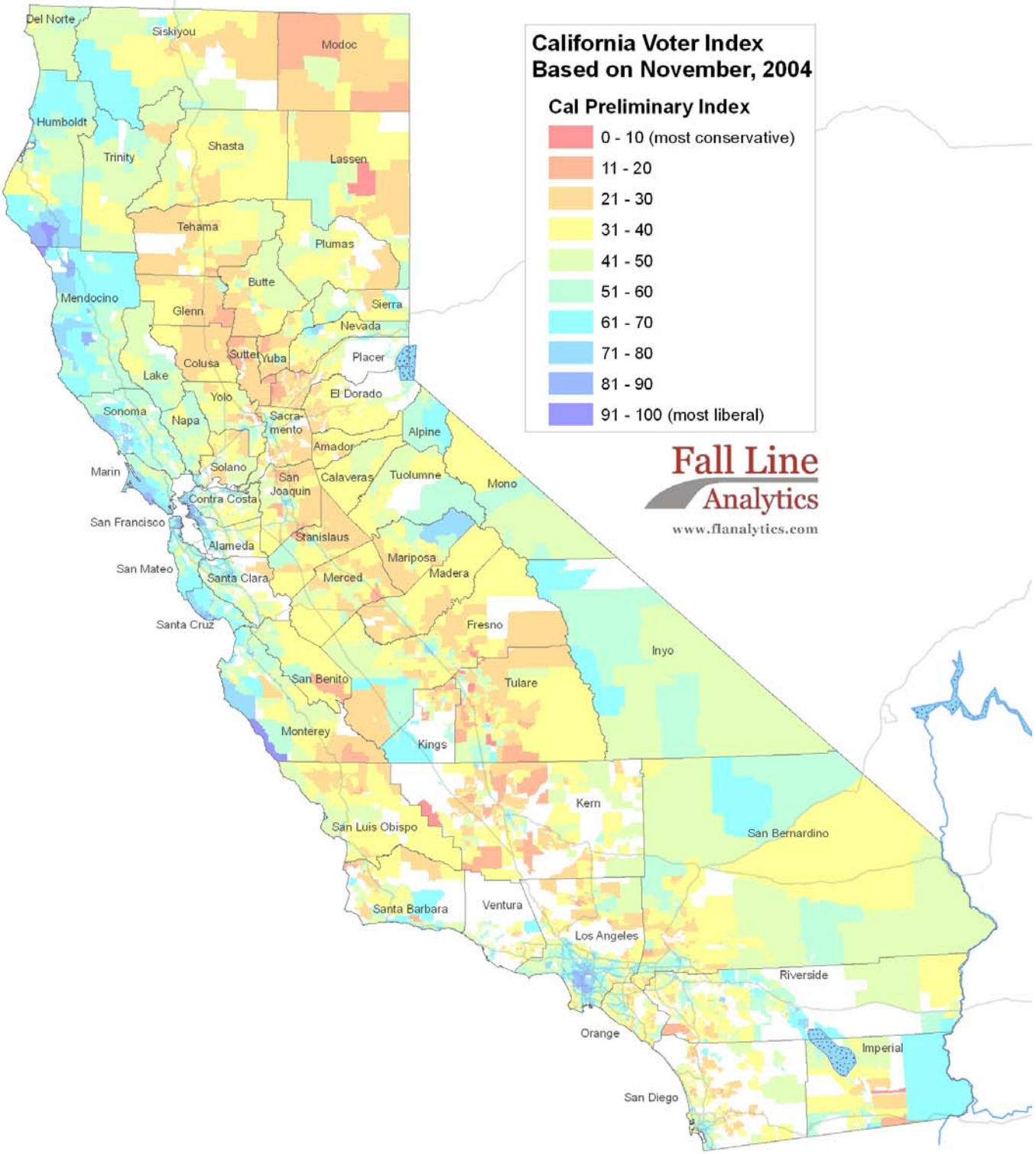


Figure 2: California CPPI (Standard red/blue)

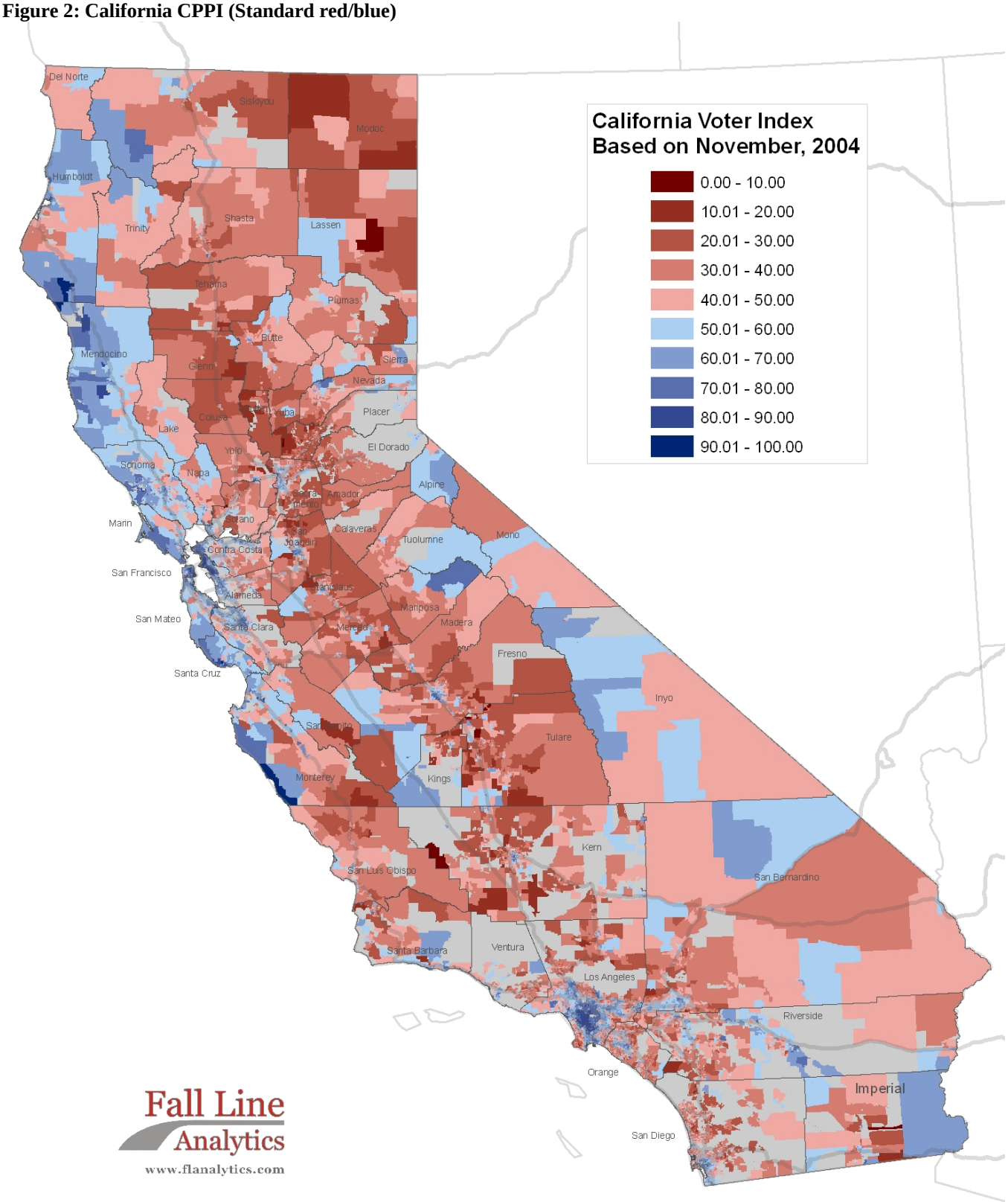
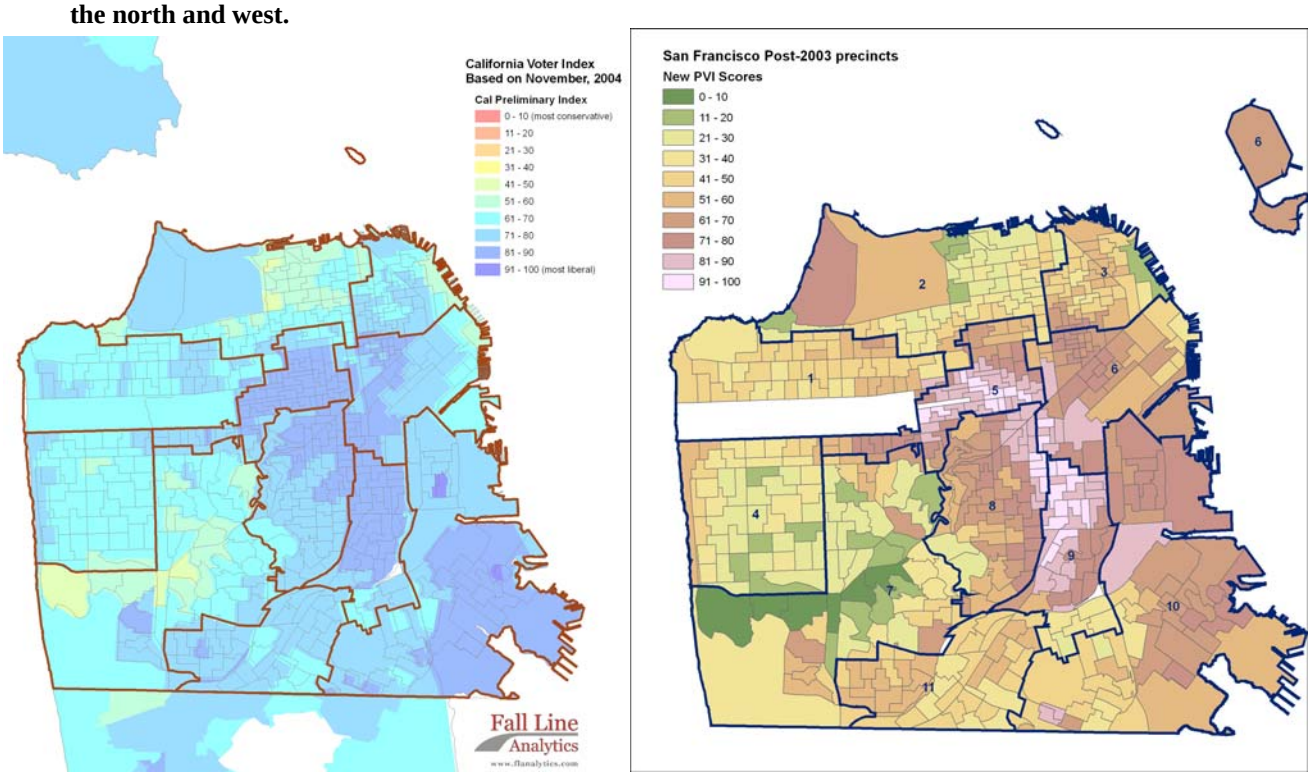


Figure 3: Cal CPPI for San Francisco (left) and the standard San Francisco PVI (right) compared. Notice the similar trends of the more liberal districts in the central part of the city, and the more conservative parts in the north and west.



County Analysis

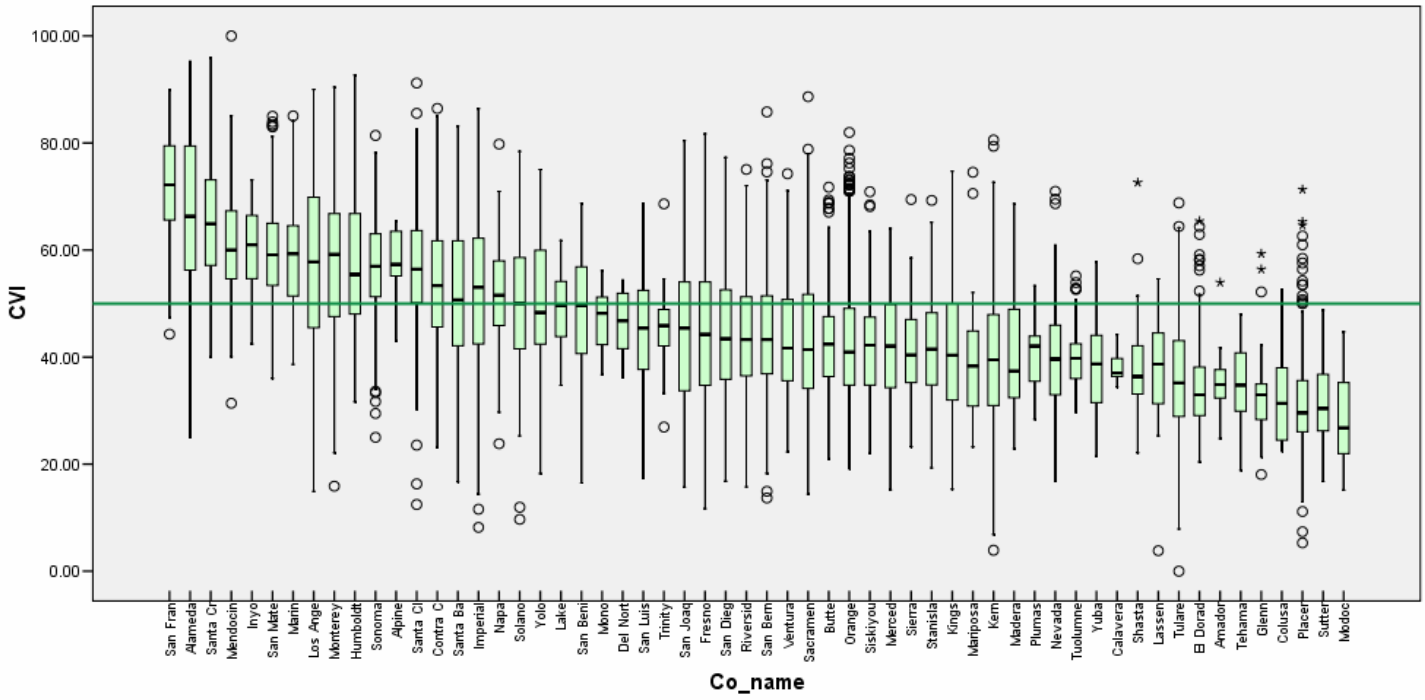
Table 1 shows each county with the number of precincts, mean, and mix/max scores. Figure 4 is a boxplot of each California county and its median score with IQ range. Not surprisingly, San Francisco can lay claim to being the most liberal county in California, followed by Alameda and Santa Cruz. Surprisingly, Inyo was fifth.[2] On the conservative side, Modoc, Sutter, and Placer were the most conservative counties.

[2] This was a consistent result under many analytical scenarios, but some of Inyo's high score may be attributable to having a relatively small number of precincts.

Table 1: Aggregated county descriptive statistics for the CPPI, in descending mean order

County	N	Mean	Min	Max	County	N	Mean	Min	Max
San Francisco	574	71.7	44.3	89.9	Ventura	280	43.8	22.3	74.3
Alameda	875	66.3	25.1	95.2	Sacramento	805	43.0	14.4	88.7
Santa Cruz	173	65.8	40.0	96.0	Butte	155	42.8	21.0	71.7
Mendocino	79	61.1	31.4	100.0	Orange	1814	42.5	19.2	82.0
Inyo	28	59.3	42.5	73.1	Siskiyou	65	42.4	22.0	70.9
San Mateo	500	59.3	36.0	85.0	Merced	103	42.0	15.2	64.0
Marin	192	58.5	38.6	85.1	Sierra	20	41.8	23.2	69.4
Los Angeles	4602	57.6	14.9	90.0	Stanislaus	295	41.6	19.3	69.3
Monterey	166	57.5	15.9	90.4	Kings	94	41.0	15.3	74.7
Humboldt	118	57.5	31.7	92.7	Mariposa	21	40.9	23.2	74.5
Sonoma	439	57.0	25.0	81.4	Kern	486	40.8	3.9	80.6
Alpine	5	56.9	43.0	65.5	Madera	60	40.7	22.9	68.7
Santa Clara	920	56.7	12.5	91.2	Plumas	29	40.3	28.4	53.3
Contra Costa	737	54.5	23.0	86.5	Nevada	123	40.2	16.7	71.0
Santa Barbara	300	52.0	16.7	83.1	Tuolumne	73	40.2	29.7	55.2
Imperial	120	51.7	8.2	86.4	Yuba	45	38.9	21.4	57.8
Napa	114	51.7	23.8	79.8	Calaveras	12	38.2	34.4	44.2
Solano	250	49.9	9.7	78.4	Shasta	130	37.5	22.1	72.6
Yolo	123	49.7	18.2	75.0	Lassen	32	37.5	3.8	54.6
Lake	40	49.0	34.7	61.8	Tulare	228	36.2	0.0	68.8
San Benito	51	48.7	16.5	68.7	El Dorado	101	35.5	20.4	65.4
Mono	13	47.4	36.7	56.2	Amador	29	35.3	24.8	53.9
Del Norte	18	46.3	36.2	54.4	Tehama	44	35.2	18.9	48.0
San Luis Obispo	149	45.3	17.4	68.6	Glenn	23	33.7	18.1	59.3
Trinity	21	45.3	26.9	68.6	Colusa	14	32.3	22.4	52.6
San Joaquin	436	44.9	15.7	80.5	Placer	318	31.6	5.3	71.3
Fresno	429	44.9	11.7	81.7	Sutter	64	31.1	16.8	48.8
San Diego	1621	44.6	16.8	77.3	Modoc	20	28.3	15.2	44.7
Riverside	701	44.5	15.8	75.1	**Total**	**20088**	**50.7**	**0.0**	**100.0**
San Bernardino	811	44.3	13.7	85.8					

Figure 4: Boxplot of aggregated County scores

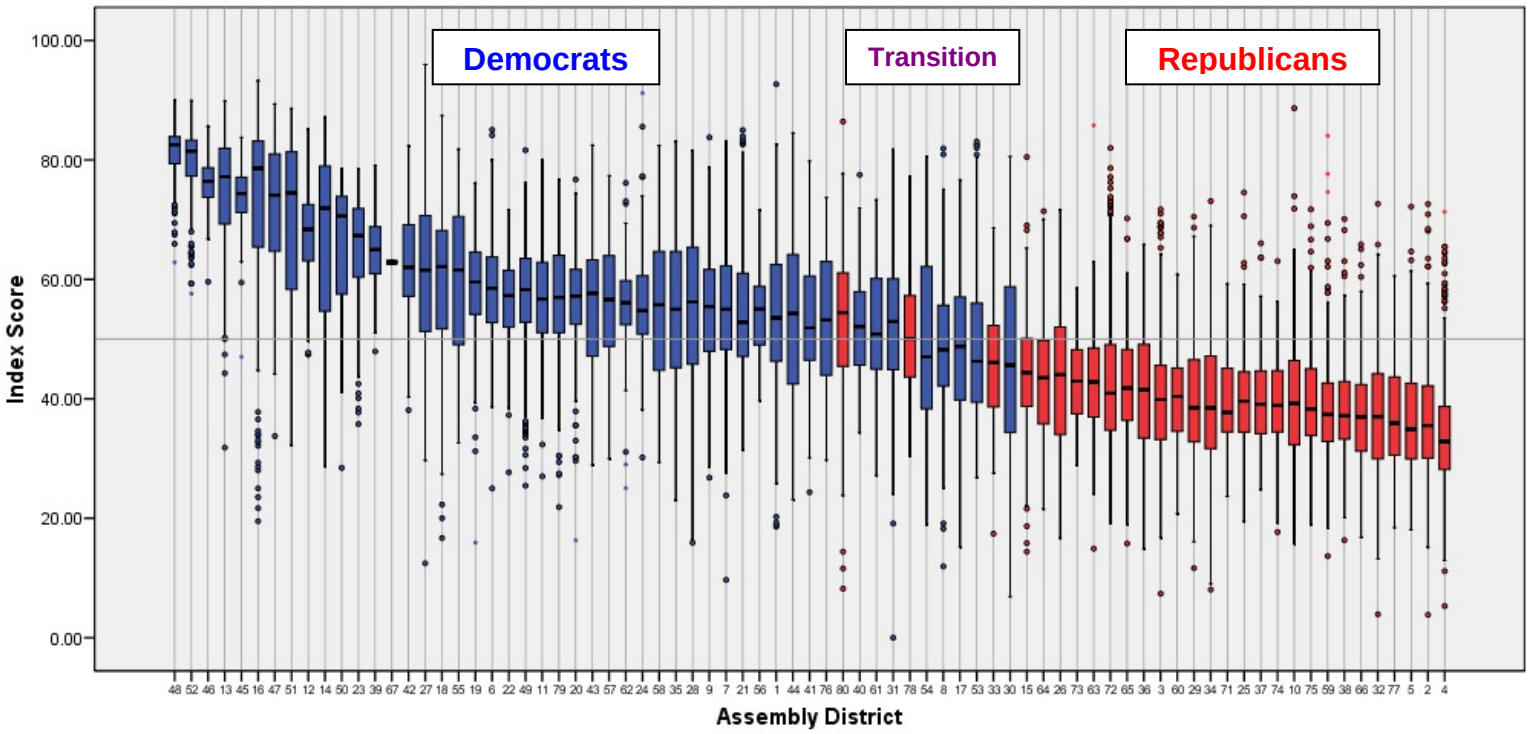


Notice more counties have mean and median scores lower than 50 (conservative) than above 50 (liberal). However, these counties generally have fewer people, and are growing more rapidly, like Placer County.

State Assembly

One application for the CPPI is we can examine the aggregated assembly district scores to see if their representatives match the score. Looking at Figure 5, we see that that this is the case. The most liberal districts are on the left (they're in LA), and are represented entirely by Democrats(colored blue) with the one Inyo County exception (AD67). In the middle there is a transition area, with the mean CPPI score right around 50. The right side of the chart shows the most conservative districts, represented by Republicans. This chart has most of the Assembly Districts, but a couple are missing.

Figure 5: Boxplot of CPPI aggregated into Assembly Districts. Blue boxes are districts with a Democrat Assemblyperson (as of June 2007), and red boxes show a GOP seat. The mean district scores match well with the party of their seat, especially with a transition area of mean ~ 50.



Identity politics

One of the useful things that can be done with the CPPI is using it to examine voting trends among different groups within California. For example, one of the most important up and coming groups in California (if it isn't already important) is the Latino voting bloc. However, like most groups, voting patterns aren't monolithic, and there are trends within the population.

Using UC Berkeley registration data, I took a subset of the California precincts in which the Latino percentage of registered voters was over 70%. From there, I looked at the mean CPPI score per county, where n>5. Table 2 shows the results.

Table 2: County mean CPPI scores for precincts where the Latino registration > 70%

County	*N*	*Mean*
Monterey	15	77.4
Santa Clara	5	72.21
Los Angeles	345	70.3
Orange	29	67.99
Imperial	35	67.63
Riverside	11	67.42
Ventura	10	66.38
San Diego	10	65.77
Fresno	18	65.63
Kern	23	62.56
Tulare	10	61.57
San Bernardino	10	59.82
Total	**535**	**69.06**

Interestingly, there is a 17 point swing between the coastal counties of Santa Clara and Monterey versus the more conservative counties of San Bernardino and Tulare. Although all the counties with significant Latino populations are 'left of center', there are clear geographic differences among the groups. Further analysis would examine demographic differences within these communities which mimic the patterns seen in the state's general voting population.

Presidential Race 2004

Figure 6 shows the precinct CPPI with the Presidential race from November 2004 (% precinct voting for Kerry). The correlation is extremely strong with an R^2 of 0.84. The most notable part of the chart, however, are the anomalous precincts which indicate parts of the state that are more conservative on choosing a President yet slightly (or not-so-slightly) more liberal on ballot measures. Most of these precincts came from Orange, Kern, and Santa Barbara Counties. These types of analyses can distinguish subtle voting patterns among populations, especially when analyzed demographically and while looking at registration trends.

Figure 6: Correlation of CPPI with the percent Democratic vote for President, 2004. The lines represent the mean and 1 SD. The precincts within the orange circle (>2 SD) are highlighted in the map below. The bulk of the anomalous precincts are in Kern, Santa Barbara, and Orange Counties.