

# **Evaluating the Likelihood that Winning Particular Presidential Primaries Will Lead To Winning The General Election**

## **Abstract**

In 2015, Donald J. Trump signed on as a presidential candidate within the Republican Party. Many thought he was doing it for show, or to boost his own brand, and that he would not garner much of a following or many election wins. But when the primary elections started in February 2016, it became evident that his following was enormous and he started to win primary elections. But does that mean he is set out to be the next President of the United States? Looking at historic data from primaries dating back to 1968, is there a way to figure out the likelihood that the winner of a primary in a particular state will be the next President?

## **Evidence**

Collecting election results should be an easy task. Elections are overseen by government agencies, but because primary elections are not constitutionally mandated, when working to collect the results one runs into many roadblocks.

Starting with [Vital Statistics on American Politics 2013-2014](#) edited by Harold W. Stanley and Richard G. Niemi, one can start collecting presidential primary results. There are numerous editions of this book including, [Vital Statistics on American Politics 2011-2012](#) also edited by Harold W. Stanley and Richard G. Niemi, and [Vital Statistics on American Politics 2005-2006](#) also edited by Harold W. Stanley and Richard G. Niemi. But one must go deeper in order to find more historical data. The [CQ Voting and Elections Collection](#) has election results going back to 1912. Some of the older elections are sparse and not as well recorded as the more recent ones. Pulling all of these datasets together, one can build a dataset of primary election results by state in elections from 1968 to 2008. (see Attachment 1)

When initially starting this project, the intention was to prove whether there was a way of forecasting whether a particular state or cluster of states could project the general election winner. Because of that, the dataset is built out in a manner to look at both the republican and democrat side. Thus, one ends up with the nominees from both parties.

What was found is the easiest way to compare the candidates and outcomes over the course of several years in many states is to separate the candidates so that the year and date of the primary is repeated twice in the y column while the republican candidates' outcome is aligned with one of those sets, and the democrat candidates' outcome is aligned with the second set. The result is a wide dataset that repeats the dates for each election year.

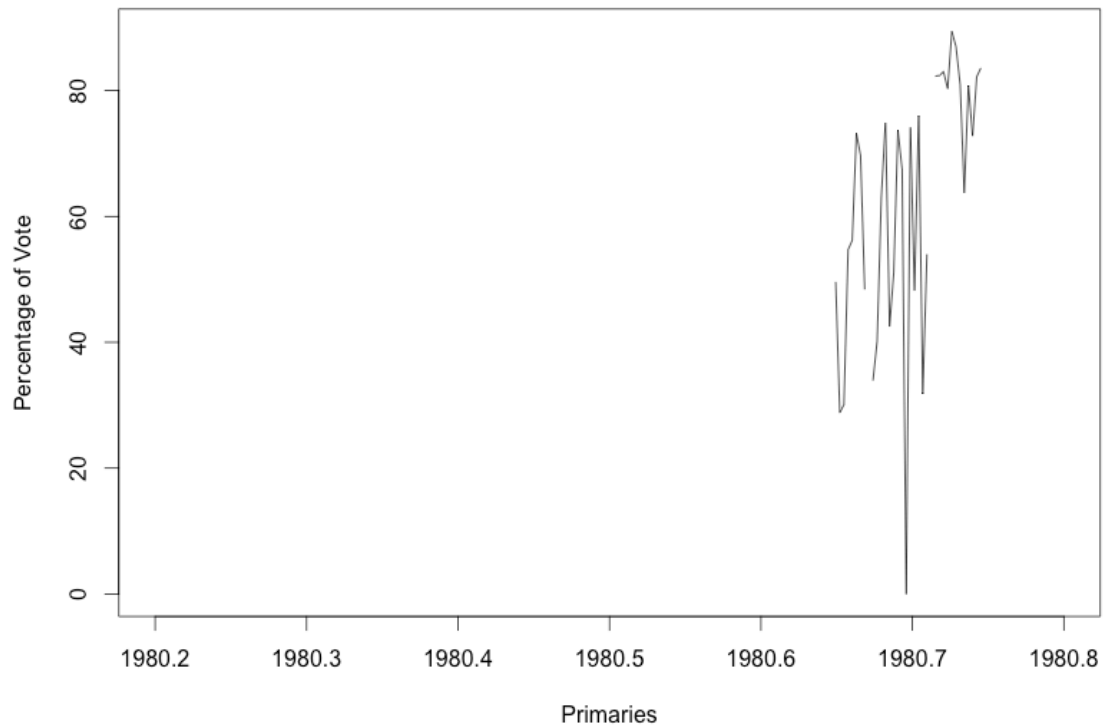
In terms of roadblocks in setting this up, it's quickly acknowledged that some candidates run for office many years in a row – or sometimes even skipping an election or two. There have also been family members who have run for the same office. So being sure to supplement the candidate's name with a given name or initial plus the year in which the race is recorded is critical. Equally as critical is adding the party initial. The result is something to the effect of "George H.W. Bush-R1992." Mr. Bush ran for office in 1980 and 1988, and his son, George W. Bush, ran in 2000 and 2004.

Formatting the states is also critical, as once all of the data is set up, evaluating over the course of several elections means being able to compare state-to-state. In the course of collecting the necessary data, the state identifiers changed many times. Sometimes it was the full name, sometimes a postal abbreviation, and sometimes an Associated Press abbreviation. It is imperative that these names match. Using the filter function within Excel, one can go through and adjust the state names to one usable name or abbreviation.

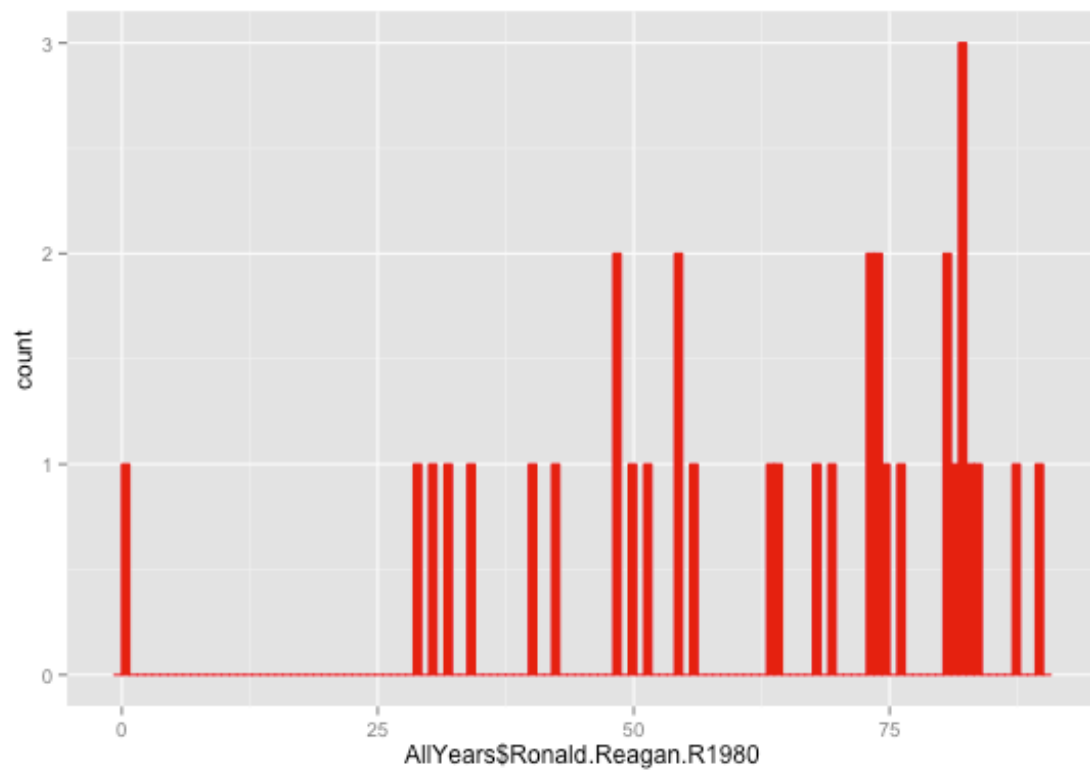
Finally, before starting evaluation, it's imperative to remove any hidden characters, as R won't be able to read the data properly if there's a mixture of characters and numerals in the same cell. Going back to the spreadsheet and finding and removing those characters – typically spaces – is critical to the evaluation. Otherwise it is impossible to coerce the results from characters to numeric.

Once all of the data is numeric, the dates are in fact dates, and the candidates are aligned with their races, one can jump into evaluation. (see Attachment 2) Before looking at the results of the individual states, time was taken to evaluate whether there were any additional relationships within the dataset. To start, it's interesting to look at the range of election results. Even for a winning candidate, the range of results can vary greatly. Here is Ronald Reagan's results for the primary season of 1980, when he ultimately went on to win the presidency. It's apparent that while Reagan did well – more than 80 percent in some of the primaries – he also received less than 40 percent in several of the elections – even zero once.

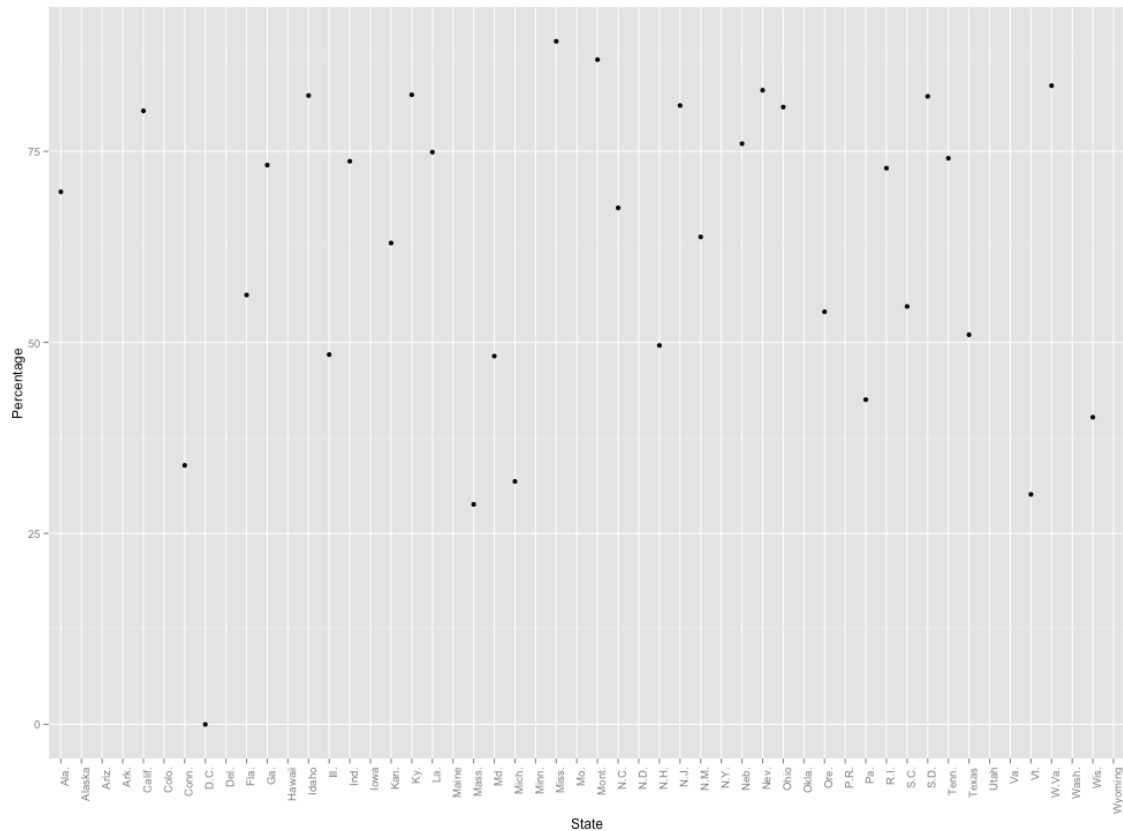
### Ronald Reagan Primary Outcome in 1980



Looking at it a different way, one can start to see clusters of voting results. Again reviewing Reagan's run in 1980, you can see he had several primary races in which he collected between 65 and 80 percent of the vote. In three of the races, he had somewhere in the field of 80 percent of the vote.



Breaking this one out into state-by-state analysis you can see in which states Reagan did well, and where he did poorly.



When you start to evaluate whether any of the candidates have significant relationships with each other, you can start to see some patterns. In the 1980 election, on the republican side, you can see the Reagan and Bush had a highly significant relationship. Reagan and Anderson had a somewhat significant relationship. And Reagan and Baker and Reagan and other candidates had no relationship.

```
##Evaluating whether any of the republican candidates have significant
relationships
lmElection2 <- lm(George.H.W.Bush.R1980 ~ Ronald.Reagan.R1980,
data=AllYears)
summary(lmElection2)
##
## Call:
## lm(formula = George.H.W.Bush.R1980 ~ Ronald.Reagan.R1980, data =
AllYears)
##
## Residuals:
##   Min     1Q   Median     3Q    Max
## -19.753  -6.356   1.597   5.497  18.384
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```

## (Intercept)      59.36467  5.06543 11.720 4.08e-13 ***
## Ronald.Reagan.R1980 -0.59506  0.07734 -7.694 9.03e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.432 on 32 degrees of freedom
## (703 observations deleted due to missingness)
## Multiple R-squared:  0.6491, Adjusted R-squared:  0.6381
## F-statistic: 59.2 on 1 and 32 DF, p-value: 9.034e-09
lmElection3 <- lm(John.Anderson.R1980 ~ Ronald.Reagan.R1980,
data=AllYears)
summary(lmElection3)
##
## Call:
## lm(formula = John.Anderson.R1980 ~ Ronald.Reagan.R1980, data =
AllYears)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.3049  -4.6059  -0.6866   5.3243  21.0131
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    29.78029   4.84635   6.145 4.26e-06 ***
## Ronald.Reagan.R1980 -0.29119   0.08215  -3.545 0.00192 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.194 on 21 degrees of freedom
## (714 observations deleted due to missingness)
## Multiple R-squared:  0.3743, Adjusted R-squared:  0.3446
## F-statistic: 12.57 on 1 and 21 DF, p-value: 0.001919
lmElection4 <- lm(Howard.Baker.R1980 ~ Ronald.Reagan.R1980,
data=AllYears)
summary(lmElection4)
##
## Call:
## lm(formula = Howard.Baker.R1980 ~ Ronald.Reagan.R1980, data =
AllYears)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.9854  -2.0624  -1.0186   0.2233   8.8964
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)

```

```
## (Intercept)      9.43928  3.96625  2.380  0.0365 *
## Ronald.Reagan.R1980 -0.12572  0.07535 -1.668  0.1234
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.93 on 11 degrees of freedom
## (724 observations deleted due to missingness)
## Multiple R-squared:  0.2019, Adjusted R-squared:  0.1294
## F-statistic: 2.783 on 1 and 11 DF, p-value: 0.1234
lmElection5 <- lm(Others.R1980 ~ Ronald.Reagan.R1980, data=AllYears)
summary(lmElection5)
##
## Call:
## lm(formula = Others.R1980 ~ Ronald.Reagan.R1980, data = AllYears)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.9762 -2.8689 -1.4609  0.5532 24.8457
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.667017  3.022613  1.875   0.07 .
## Ronald.Reagan.R1980 -0.009373  0.046151 -0.203   0.84
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.628 on 32 degrees of freedom
## (703 observations deleted due to missingness)
## Multiple R-squared:  0.001287, Adjusted R-squared: -0.02992
## F-statistic: 0.04125 on 1 and 32 DF, p-value: 0.8403
```

This shows how the candidates may play against one another. As we know, Reagan eventually selected Bush as his running mate. This makes sense considering together they have a highly significant relationship and would likely pull in high election numbers.

But based on how the dataset is built, one cannot determine whether Reagan has a significant relationship with any of the democrat candidates.

Moving into evaluating the individual state results, it's a matter of reevaluating the dataset.

The first move was to record which candidate ended up being the President, and which ended up being the runner-up. Then, going through each state, highlight whether the nominee received the highest percentage of votes in the primary.

AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	B
			Jimmy Carter- D1980	Edward Kennedy- D1980	Jerry Brown- D1980	Others- D1980				Ronald Reagan- R1980	George H.W.Bush- R1980	John Anderson- R1980	How Bak R15
Year	Date	State					Year	Date	State				
1980	5/6/1980	N.C.	70.1	17.7	2.9	9.3	1980	5/6/1980	N.C.				
1980	5/6/1980	D.C.	36.9	61.7		1.4	1980	5/6/1980	D.C.				
1980	5/6/1980	Tenn.	75.2	18.1	1.9	4.8	1980	5/6/1980	Tenn.				
1980	5/13/1980	Md.	47.5	38	3	11.5	1980	5/13/1980	Md.				
1980	5/13/1980	Neb.	46.9	37.6	3.6	11.9	1980	5/13/1980	Neb.				
1980	5/20/1980	Mich.			29.4	70.6	1980	5/20/1980	Mich.				
1980	5/20/1980	Ore.	56.7	31.1	9.3	2.9	1980	5/20/1980	Ore.				
1980	5/27/1980	Ark.	60.1	17.5		22.4	1980	5/27/1980	Ark.				
1980	5/27/1980	Idaho	62	22	4	12	1980	5/27/1980	Idaho				
1980	5/27/1980	Ky.	66.9	23		10.1	1980	5/27/1980	Ky.				
1980	5/27/1980	Nev.	37.6	28.8		33.6	1980	5/27/1980	Nev.				
1980	6/3/1980	Calif.	37.6	44.8	4	13.5	1980	6/3/1980	Calif.				
1980	6/3/1980	Miss.					1980	6/3/1980	Miss.				
1980	6/3/1980	Mont.	51	37		12	1980	6/3/1980	Mont.				
1980	6/3/1980	N.J.	38	56		6	1980	6/3/1980	N.J.				
1980	6/3/1980	N.M.	41.8	46.3		11.9	1980	6/3/1980	N.M.				
1980	6/3/1980	Ohio	51.1	44.4		4.5	1980	6/3/1980	Ohio				
1980	6/3/1980	R.I.	25.8	68.3	0.8	5.1	1980	6/3/1980	R.I.				
1980	6/3/1980	S.D.	45.4	48.6		6	1980	6/3/1980	S.D.				
1980	6/3/1980	W.Va.	62.2	37.8		0	1980	6/3/1980	W.Va.				
1980	2/17/1980	P.R.					1980	2/17/1980	P.R.		60.1		3
1980	2/26/1980	N.H.					1980	2/26/1980	N.H.	49.6	22.7		12
1980	3/4/1980	Mass.					1980	3/4/1980	Mass.	28.8	31	30.7	4
1980	3/4/1980	Vt.					1980	3/4/1980	Vt.	30.1	21.7	29	12
1980	3/8/1980	S.C.					1980	3/8/1980	S.C.	54.7	14.8		0
1980	3/11/1980	Fla.					1980	3/11/1980	Fla.	56.2	30.2	9.2	1
1980	3/11/1980	Ga.					1980	3/11/1980	Ga.	73.2	12.6	8.4	0
1980	3/11/1980	Ala.					1980	3/11/1980	Ala.	69.7	25.9	0.9	0
1980	3/18/1980	Ill.					1980	3/18/1980	Ill.	48.4	11	36.7	1
1980	3/25/1980	N.Y.					1980	3/25/1980	N.Y.				
1980	3/25/1980	Conn.					1980	3/25/1980	Conn.	33.9	38.6	22.1	1
1980	4/1/1980	Wis.					1980	4/1/1980	Wis.	40.2	30.4	27.4	0
1980	4/1/1980	Kan.					1980	4/1/1980	Kan.	63	12.6	18.2	1
1980	4/5/1980	La.					1980	4/5/1980	La.	74.9	18.8		
1980	4/22/1980	Pa.					1980	4/22/1980	Pa.	42.5	50.5	2.1	2
1980	5/3/1980	Texas					1980	5/3/1980	Texas	51	47.4		
1980	5/6/1980	Ind.					1980	5/6/1980	Ind.	73.7	16.4	9.9	

Once that has been completed for all of the years and races, one can create a new dataset of just the winning candidates along with the states in which they won. (see Attachment 3) One must eliminate all of the non-winning states, as they no longer are part of the equation. From that another dataset can be created in which one column is the states in which the winning candidate also won the primary. One may add each year's results below the previous, as from this point out it's less about the year in which the race was won and more about which states appear most often. A second column may be created for the runner-up. This way, it can be evaluated not just whether there's a high probability that winning a state's primary may lead to being the next President, but also whether it means a candidate may go on to be the party's nominee.

Now that the two columns are created, they can be sorted alphabetically and one can count the outcomes. (see Attachment 4 and 5)



WINNER		RUNNER UP
Ala.		Ala.
Ala.		Ala.
Ala.		Ala.
Ala.		Ala.
Ala.		Ala.
Ala.		Ala.
Alaska		Ariz.
Ariz.		Ariz.
Ark.		Ariz.
Ark.		Ark.
Ark.		Ark.
Ark.		Ark.
Calif.		Ark.
Calif.		Ark.
Calif.		Calif.
Calif.		Calif.
Calif.		Calif.
Calif.		Calif.
Calif.		Calif.
Colo.		Calif.
Colo.		Calif.
Colo.		Colo.
Conn.		Colo.
Conn.		Colo.
D.C.		Conn.
D.C.		Conn.
D.C.		Conn.
D.C.		Conn.
D.C.		Conn.
D.C.		Conn.
D.C.		D.C.
D.C.		D.C.
Del.		D.C.
Del.		D.C.
Del.		D.C.
Fla.		Del.
Fla.		Del.

	WINNER	OPPONENT
1968	Nixon	Humphrey
1972	Nixon	McGovern
1976	Carter	Ford
1980	Reagan	Carter
1984	Reagan	Mondale
1988	George H.W. Bush	Dukakis
1992	W. Clinton	George H.W. Bush
1996	W. Clinton	Dole
2000	George W. Bush	Gore
2004	George W. Bush	Kerry
2008	Obama	McCain
Ala.	6	6
Alaska	1	0
Ariz.	1	3
Ark.	4	5
Calif.	7	7
Colo.	3	3
Conn.	2	6
D.C.	8	5
Del.	3	3
Fla.	7	8
Ga.	8	6
Hawaii	1	0
Idaho	9	7
Ill.	10	8
Ind.	9	7
Iowa	1	0
Kan.	3	2
Ky.	6	7
La.	7	5
Maine	3	2
Mass.	4	7
Md.	7	9
Mich.	3	5
Minn.	2	1
Miss.	6	5

While reviewing this data, several problems become evident. For example, although most of the initial datasets were solely primaries, the 2008 results also included caucuses, which is why Iowa shows up just once – in 2008. But Colorado, which also holds a caucus, appears in the results many years (1992, 1996, 2008).

Looking at simply how many states reported results, the initial data varies greatly. For instance, in 1968, the results of only 11 races were available. And in the case of the runner-up candidate, Hubert Humphrey, he did not win any of the states in which data was collected. Therefore, none of the states were part of the count when reviewing how many states elected winning or runner-up candidates. 1972 had slightly more states, but at 21 it's still not half.

There was also an instance in 2004 in which the incumbent candidate's results are not part of the dataset. Although the candidate, George W. Bush, would be expected to win all of the republican primaries, those states are not part of the count of states electing the next president.

In some years states were listed but had no results. This was the case in Virginia in 2000, Hawaii in 2008, and Michigan in 1996, to name a few.

Bill Clinton-D1996	Year	Date	State
100	1996	3/5/1996	Ga.
88.4	1996	3/5/1996	Maine
84.2	1996	3/5/1996	Md.
87.1	1996	3/5/1996	Mass.
89.1	1996	3/5/1996	R.I.
96.5	1996	3/5/1996	Vt.
	1996	3/12/1996	Fla.
80.8	1996	3/12/1996	La.
92.5	1996	3/12/1996	Miss.
76.2	1996	3/12/1996	Okla.
94.8	1996	3/12/1996	Ore.
88.9	1996	3/12/1996	Tenn.
86.4	1996	3/12/1996	Texas
96.2	1996	3/19/1996	Ill.
	1996	3/19/1996	Mich.
92.6	1996	3/19/1996	Ohio
97.6	1996	3/19/1996	Wis.
92.8	1996	3/26/1996	Calif.
	1996	3/26/1996	Nev.
98.5	1996	3/26/1996	Wash.

Moving back to the data that has been sifted, there is a list of states and how many times the candidate who got the highest percentage of the vote has gone on to be President and gone on to be the runner-up. But because of the exceptions listed

above, they're not all on equal playing ground. Once again it's a matter of sifting through the data to find the pertinent information. (see Attachment 6)

Looking at which states were listed in the 1968 results, it would come to reason that any of those states would have a higher net, as they appear in more races. Also of note, there are no records of the primary results for the Republican candidate in 2004, as Bush was re-elected. So from 1968 to 2008 there are actually 10, not 11 races.

Wisconsin was the only state with a perfect record going back to 1968 (data from 2004 not included for reasons mentioned above). Illinois, Nebraska, Indiana and Oregon all were nine for ten. Washington, D.C., New Jersey, Ohio, and Pennsylvania were all eight for ten. California, Florida, and New Hampshire were seven for ten. West Virginia was six for ten; South Dakota was five for ten; and Massachusetts was four for ten.

North Carolina and Tennessee were added to the records in 1972 and were eight for nine. And Idaho and Georgia were added to the records in 1976 and both were eight for eight. Louisiana and Texas were added in 1980 and were seven for seven.

Looking at the final tallies as percentages of correctly voting in the primary for what ends up being the winning candidate, the states with the highest percentages are:

- Wisconsin – 100% – 10 of 10 elections (1968-2000 & 2008)
- Idaho – 100% – 8 of 8 elections (1976-2000 & 2008)
- Georgia – 100% – 8 of 8 elections (1976-2000 & 2008)
- Louisiana – 100% – 7 of 7 elections (1980-2000 & 2008)
- Texas – 100% – 7 of 7 elections (1980-2000 & 2008)
- Illinois – 90% – 9 of 10 elections (1968-2000 & 2008)
- Nebraska – 90% – 9 of 10 elections (1968-2000 & 2008)
- Indiana – 90% – 9 of 10 elections (1968-2000 & 2008)
- Oregon – 90% – 9 of 10 elections (1968-2000 & 2008)
- North Carolina – 89% – 8 of 9 elections (1972-2000 & 2008)
- Tennessee – 89% – 8 of 9 elections (1972-2000 & 2008)
- Montana – 88% – 7 of 8 elections (1976-2000 & 2008)
- Alabama – 88% – 6 of 7 elections (1980-2000 & 2008)
- Mississippi – 88% – 6 of 7 elections (1980-2000 & 2008)
- Washington, D.C. – 80% -- 8 of 10 elections (1968-2000 & 2008)
- New Jersey – 80% – 8 of 10 elections (1968-2000 & 2008)
- Ohio – 80% – 8 of 10 elections (1968-2000 & 2008)
- Pennsylvania – 80% – 8 of 10 elections (1968-2000 & 2008)

As mentioned, New Hampshire, which is closely watched and highly touted as the first in the nation primary, only came in at 70% with 7 of 10 elections from 1968 to 2008 (minus 2004). South Carolina, another closely watched state, came in at 71% with 5 of 7 elections from 1980 to 2008 (minus 2004).

STATE	# OF WINS	YEAR	YEAR	YEAR	YEAR	YEAR	YEAR	YEAR	YEAR	YEAR	YEAR	YEAR	WINS/YEARS	PERCENTAGE
Ill.	10	1968	1972	1976	1980	1984	1988	1992	1996	2000		2008	9 of 10	90%
Neb.	10	1968	1972	1976	1980	1984	1988	1992	1996	2000		2008	9 of 10	90%
Wis.	10	1968	1972	1976	1980	1984	1988	1992	1996	2000		2008	10 of 10	100%
Idaho	9			1976	1980	1984	1988	1992	1996	2000		2008	8 of 8	100%
Ind.	9	1968	1972	1976	1980	1984	1988	1992	1996	2000		2008	9 of 10	90%
Ore.	9	1968	1972	1976	1980	1984	1988	1992	1996	2000		2008	9 of 10	90%
D.C.	8	1968	1972	1976	1980	1984	1988	1992	1996	2000		2008	8 of 10	80%
Ga.	8			1976	1980	1984	1988	1992	1996	2000		2008	8 of 8	100%
N.C.	8		1972	1976	1980	1984	1988	1992	1996	2000		2008	8 of 9	89%
N.J.	8	1968	1972	1976	1980	1984	1988	1992	1996	2000		2008	8 of 10	80%
Ohio	8	1968	1972	1976	1980	1984	1988	1992	1996	2000		2008	8 of 10	80%
Pa.	8	1968	1972	1976	1980	1984	1988	1992	1996	2000		2008	8 of 10	80%
Tenn.	8		1972	1976	1980	1984	1988	1992	1996	2000		2008	8 of 9	89%
Calif.	7	1968	1972	1976	1980	1984	1988	1992	1996	2000		2008	7 of 10	70%
Fla.	7	1968	1972	1976	1980	1984	1988	1992	1996	2000		2008	7 of 10	70%
La.	7				1980	1984	1988	1992	1996	2000		2008	7 of 7	100%
Md.	7		1972	1976	1980	1984	1988	1992	1996	2000		2008	7 of 9	78%
Mont.	7			1976	1980	1984	1988	1992	1996	2000		2008	7 of 8	88%
N.H.	7	1968	1972	1976	1980	1984	1988	1992	1996	2000		2008	7 of 10	70%
N.M.	7		1972	1976	1980	1984	1988	1992	1996	2000		2008	7 of 9	78%
Texas	7				1980	1984	1988	1992	1996	2000		2008	7 of 7	100%
Ala.	6				1980	1984	1988	1992	1996	2000		2008	6 of 7	86%
Ky.	6			1976	1980	1984	1988	1992	1996	2000		2008	6 of 8	75%
Miss.	6				1980	1984	1988	1992	1996	2000		2008	6 of 7	86%
W.Va.	6	1968	1972	1976	1980	1984	1988	1992	1996	2000		2008	6 of 10	60%
R.I.	5		1972	1976	1980	1984	1988	1992	1996	2000		2008	5 of 9	56%
S.C.	5				1980	1984	1988	1992	1996	2000		2008	5 of 7	71%
S.D.	5	1968	1972	1976	1980	1984	1988	1992	1996	2000		2008	5 of 10	50%
Vt.	5			1976	1980	1984	1988	1992	1996	2000		2008	5 of 8	63%
Wash.	4							1992	1996	2000		2008	4 of 4	100%
Ark.	4				1980	1984	1988	1992	1996	2000		2008	4 of 7	57%
Mass.	4	1968	1972	1976	1980	1984	1988	1992	1996	2000		2008	4 of 10	40%
Okla.	4						1988	1992	1996	2000		2008	4 of 5	80%
Colo.	3							1992	1996	2000		2008	3 of 4	75%
Del.	3								1996	2000		2008	3 of 3	100%
Kan.	3				1980			1992				2008	3 of 3	100%
Maine	3								1996	2000		2008	3 of 3	100%
Mich.	3		1972	1976	1980			1992	1996	2000		2008	3 of 7	43%
Mo.	3						1988			2000		2008	3 of 3	100%
N.D.	3					1984	1988	1992	1996			2008	3 of 5	60%
Va.	3						1988			2000		2008	3 of 3	100%
Conn.	2				1980	1984	1988	1992	1996	2000		2008	2 of 7	29%
Minn.	2							1992				2008	2 of 2	100%
N.Y.	2				1980	1984	1988	1992		2000		2008	2 of 6	33%
P.R.	2				1980	1984	1988	1992		2000		2008	2 of 6	33%
Utah	2									2000		2008	2 of 2	100%
Alaska	1											2008	1 of 1	100%
Ariz.	1			1976					1996	2000		2008	1 of 4	25%
Hawaii	1											2008	1 of 1	100%
Iowa	1											2008	1 of 1	100%
Nev.	1			1976	1980				1996			2008	1 of 4	25%
Wyoming	1											2008	1 of 1	100%

When switching from percentages to probability, the significance drops dramatically. (see Attachment 7) In looking at all of the states that recorded more than seven or more election years, Wisconsin, Idaho, Georgia, Louisiana, and Texas were at the top, but at .038491147.

STATE	WINS/YEARS	PERCENTAGE	EQUATION	PROBABILITY
Wis.	10 of 10	100%	100/2598	0.038491147
Idaho	8 of 8	100%	100/2598	0.038491147
Ga.	8 of 8	100%	100/2598	0.038491147
La.	7 of 7	100%	100/2598	0.038491147
Texas	7 of 7	100%	100/2598	0.038491147
Ill.	9 of 10	90%	90/2598	0.034642032
Neb.	9 of 10	90%	90/2598	0.034642032
Ind.	9 of 10	90%	90/2598	0.034642032
Ore.	9 of 10	90%	90/2598	0.034642032
N.C.	8 of 9	89%	89/2598	0.034257121
Tenn.	8 of 9	89%	89/2598	0.034257121
Mont.	7 of 8	88%	88/2598	0.033872209
Ala.	6 of 7	88%	88/2598	0.033872209
Miss.	6 of 7	88%	88/2598	0.033872209
D.C.	8 of 10	80%	80/2598	0.030792918
N.J.	8 of 10	80%	80/2598	0.030792918
Ohio	8 of 10	80%	80/2598	0.030792918
Pa.	8 of 10	80%	80/2598	0.030792918
Md.	7 of 9	78%	78/2598	0.030023095
N.M.	7 of 9	78%	78/2598	0.030023095
Ky.	6 of 8	75%	75/2598	0.02886836
S.C.	5 of 7	71%	71/2598	0.027328714
Calif.	7 of 10	70%	70/2598	0.026943803
Fla.	7 of 10	70%	70/2598	0.026943803
N.H.	7 of 10	70%	70/2598	0.026943803
Vt.	5 of 8	63%	63/2598	0.024249423
W.Va.	6 of 10	60%	60/2598	0.023094688
Ark.	4 of 7	57%	57/2598	0.021939954
R.I.	5 of 9	56%	56/2598	0.021555042
S.D.	5 of 10	50%	50/2598	0.019245574
Mich.	3 of 7	43%	43/2598	0.016551193
Mass.	4 of 10	40%	40/2598	0.015396459
N.Y.	2 of 6	33%	33/2598	0.012702079
P.R.	2 of 6	33%	33/2598	0.012702079
Conn.	2 of 7	29%	29/2598	0.011162433
		2598%		1

In looking at the states with the most election years, the probability seems to be more significant. Of course, there are fewer of those states. Wisconsin again leads the list at .087719, followed by Illinois, Nebraska, Indiana, and Oregon all at .078947.

STATE	WINS/YEARS	PERCENTAGE	EQUATION	PROBABILITY
Wis.	10 of 10	100%	100/1140	0.087719298
Ill.	9 of 10	90%	90/1140	0.078947368
Neb.	9 of 10	90%	90/1140	0.078947368
Ind.	9 of 10	90%	90/1140	0.078947368
Ore.	9 of 10	90%	90/1140	0.078947368
D.C.	8 of 10	80%	80/1140	0.070175439
N.J.	8 of 10	80%	80/1140	0.070175439
Ohio	8 of 10	80%	80/1140	0.070175439
Pa.	8 of 10	80%	80/1140	0.070175439
Calif.	7 of 10	70%	70/1140	0.061403509
Fla.	7 of 10	70%	70/1140	0.061403509
N.H.	7 of 10	70%	70/1140	0.061403509
W.Va.	6 of 10	60%	60/1140	0.052631579
S.D.	5 of 10	50%	50/1140	0.043859649
Mass.	4 of 10	40%	40/1140	0.035087719
		1140%		1

With so few years on record and so many asterisks in the data that exists, it's difficult to claim that any of this is statistically significant.

But looking at the 2016 election, here are some states of note:

- Idaho's and Wisconsin's primary winners were Ted Cruz and Bernie Sanders.
- Illinois', Georgia's, Louisiana's, North Carolina's, and Tennessee's primary winners were Donald Trump and Hillary Clinton.
- Texas' primary winners were Cruz and Clinton.

Indiana's primary is May 3, 2016. Nebraska's primary is on May 10, 2016. Oregon's primary is May 17, 2016.

Professor of political science at State University of New York – Stony Brook [Helmut Norpoth predicts](#) that if Trump becomes the GOP nominee there is between 87% and 99% certainty Trump will be the next President.

“Winning the early primaries is a major key for electoral victory in November,” Norpoth writes. But historically, looking at New Hampshire, it has one of the lowest probabilities of the states with the most primary records.

In the end, this may be the election that breaks all records and predictions.

## **RESOURCES**

### [Vital Statistics on American Politics 2013-2014](http://sk.sagepub.com.ezproxy.neu.edu/cqpress/vital-statistics-on-american-politics-2013-2014)

(<http://sk.sagepub.com.ezproxy.neu.edu/cqpress/vital-statistics-on-american-politics-2013-2014>), edited by Harold W. Stanley and Richard G. Niemi

### [Vital Statistics on American Politics 2011-2012](http://sk.sagepub.com.ezproxy.neu.edu/cqpress/vital-statistics-on-american-politics-2011-2012)

(<http://sk.sagepub.com.ezproxy.neu.edu/cqpress/vital-statistics-on-american-politics-2011-2012>), edited by Harold W. Stanley and Richard G. Niemi

### [Vital Statistics on American Politics 2005-2006](http://sk.sagepub.com.ezproxy.neu.edu/cqpress/vital-statistics-on-american-politics-2005-2006)

(<http://sk.sagepub.com.ezproxy.neu.edu/cqpress/vital-statistics-on-american-politics-2005-2006>), edited by Harold W. Stanley and Richard G. Niemi

### [CQ Voting and Elections Collection](http://library.cqpress.com/elections/toc.php?parent_id=28347ADA-F48F-44FF-8D6C-48485EF918A2)

([http://library.cqpress.com/elections/toc.php?parent\\_id=28347ADA-F48F-44FF-8D6C-48485EF918A2](http://library.cqpress.com/elections/toc.php?parent_id=28347ADA-F48F-44FF-8D6C-48485EF918A2))

[Primary Model](http://primarymodel.com) (<http://primarymodel.com>), Helmut Norpoth

## **ATTACHMENTS**

ATTACHMENT 1: Full data set

ATTACHMENT 2: R code of plots and statistical significance

ATTACHMENT 3: All the states that voted for general election winner in primary

ATTACHMENT 4: All the states that voted for eventual President

ATTACHMENT 5: State counts

ATTACHMENT 6: Winners' state counts

ATTACHMENT 7: Figuring probability