Stories from voting and election data

Ted Mellnik, The Washington Post
NICAR 2015

I'd like to talk about some possibilities for stories using election and voting data. Why should you could be interested?

First, and I don't mean to sound like your first grade piano teacher, but think about the day when suddenly you have a shot at a great story, and you think you can get some great leverage on it using data Do you really want to be pulling out your NICAR training notes, for the first time in who knows how long? No. **You need to practice,** now, for that day. And you can get solid data experience with election data. Making maps, joining and summarizing data, using Census stats, and looking for relationships and patterns.

Second, opportunities. There is a market for good stories on voting and elections. And there is plenty of material. You have elections every year.

You might ask, isn't this election stuff boring. After somebody wins, what else is there? Good question. One answer is how. Not who won. **How**. How did they win, or lose? How did voters act differently? This question is always fresh.

Or to put it another way, do more than make a dumb map. There area lots of dumb election maps, and their equivalent stories. Here's one you may have seen.
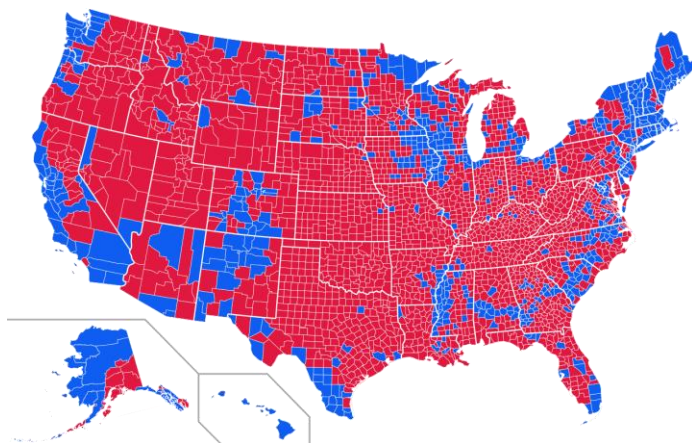
Lots of detail here. Maps are great for that. But there's no sense of what the story is. That's what I mean by a dumb map, it doesn't speak, it doesn't tell a story. **No dumb maps.**

Most maps start out like this. They show a jumble that hides broad patterns. To get at how, you need to find pattern in the jumble by simplifying and summarizing.

You can do that by creating groups. You combine areas that share some characteristic. You make subtotals for your groups.

For an example, let's start with this dumb Virginia county map from 2013. Clear who won, much less how?
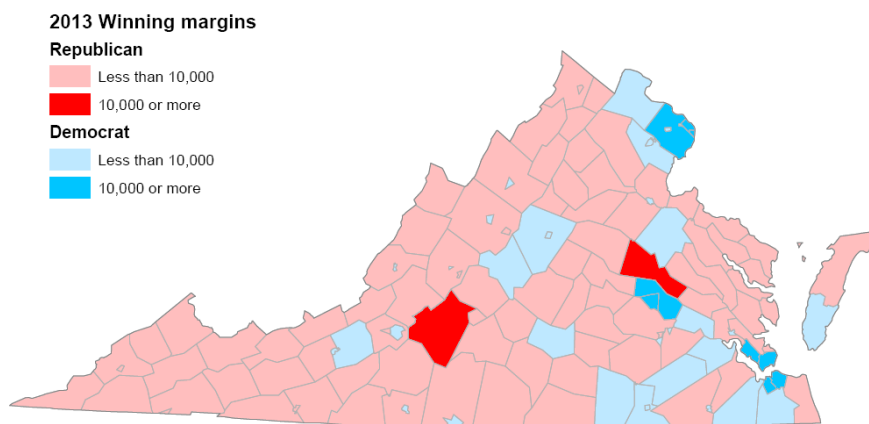

No. We can get a somewhat better sense of the story looking at the county vote margins represented by symbols that are sized by the numbers. I have cheated a little here, by fading the red squares, but even without that I think you can see that the candidate represented by blue was the winner, and there's a potentially interesting pattern.
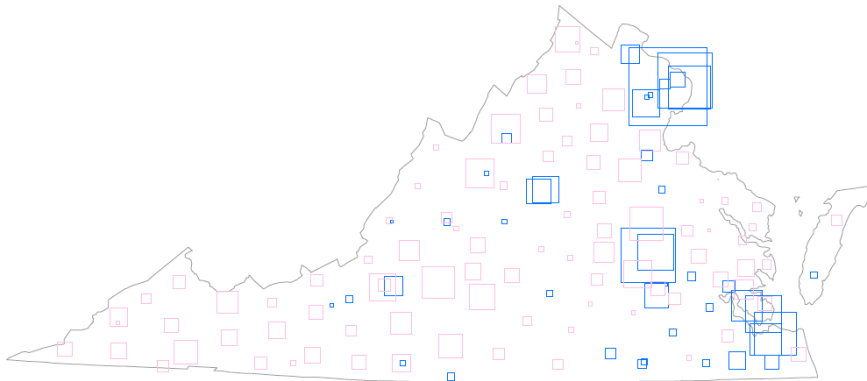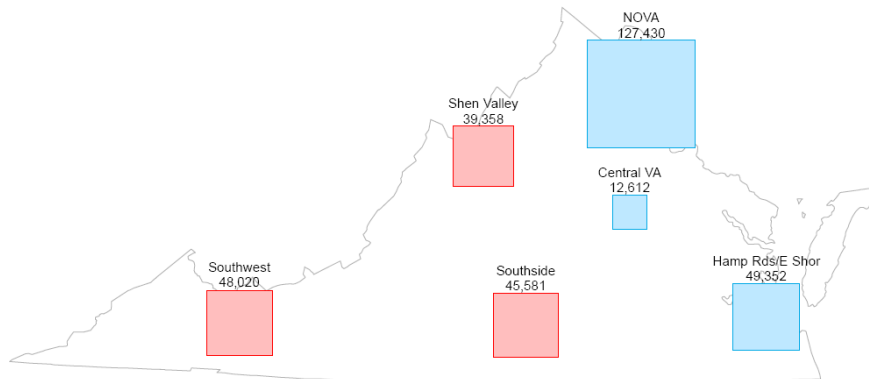
Still a bit complicated, though. One way to know when you're done summarizing in something like this is to **try to come up with two sentences you could put in a story that would describe what you have**. Not done yet.

Let's try making some groups. Geographic regions are the most basic group. Here are the VA regions.
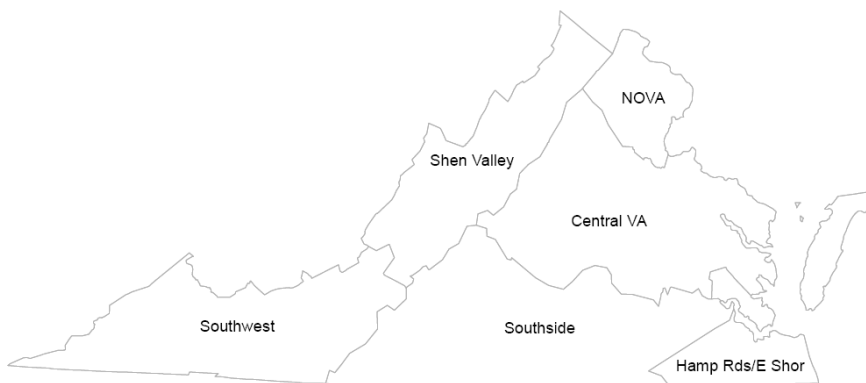
We take all those 143 little squares and reduce them to the six regions. Hopefully that makes it easier to see and to describe the pattern.

NOVA
127,430

Shen Valley
39,358

Central VA
12,612

Southwest
48,020

Southside
45,581

Hamp Rds/E Shor
49,352

I think you could readily come up with two sentences now. They could be about political polarization, the dominance of parties within their own regions, and about the ballooning influence of Northern Virginia.

It helps to work with a **simple measure** like winning vote margin. Here, for margins you're just subtracting Democrat minus Republican, or the other way around, just so you're consistent. You can simply add them up to get summaries for groups, and they reflect the relative role that each group plays in the statewide outcome.

How do you make this work? You make a master list of counties, with a common ID for the county and with its region designation. When election results come in, you use the county ID to join, or match, the election results with the county regions, and subtotal results for the regions. Basic CAR.

What else works for election groupings? Demographic attributes from Census. For counties, the place to start is urbanization. Make groups based on how much of a county's population lives in urban or rural areas.

Other possibilities are: Population (big -middle-small). Whether counties are in a metropolitan area or a micropolitan area or are rural. Or some combination of these.

Here's a look at the election night results of the 2014 Maryland governor's race in through urbanization. Two sentences. A strong swing of suburban voters produced an unexpected Republican win over a Democrat and sitting lieutenant governor. Also, there was a lack of excitement in heavily minority urban areas for a candidate who would have been the state's first African American chief executive.

Margins by urban regions

Who wins in the Maryland governor's race depends on swing voters in suburban counties, as well on the balance of voter turnout across the state. Here's a look at the different roles played by counties across the urban landscape.

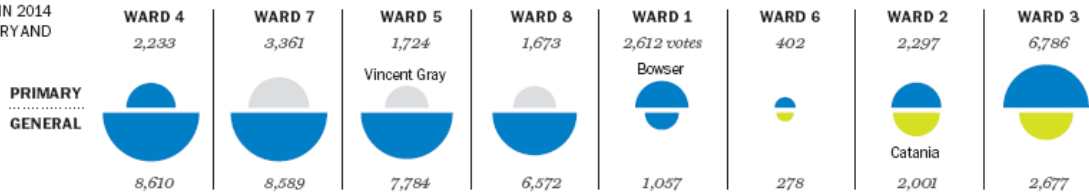|			2002 ☑	2010 ☑	2014 ☑
LARGE URBAN		These three large counties, home to Maryland's biggest blocks of minority voters, are the state's Democratic stronghold.	253,862	394,638	274,382
SUBURBAN		The largest Republican margins usually come from among these nine counties. But swing voters can change the size of wins, and determine elections.	225,805	55,312	222,906
RURAL		Reliably Republican, the 12 least-urban counties have the fewest votes, but they can be key when margins in suburban and urban counties are close.	94,227	70,684	128,072
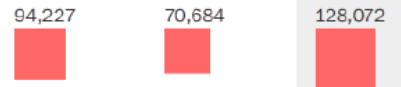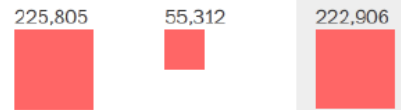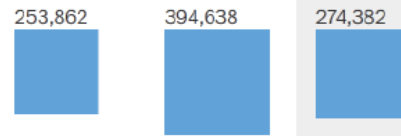
Notice how small the maps are. And how easy it is to compare regions and time.

Let's go local. Just as for counties, geographic or political regions are a good place to start. For example, subtotals for city council districts or wards. Here's a comparison of winning margins in D.C.'s eight city council wards for the city's new mayor, shown in blue, for both 2014's primary and general elections.

WINNING MARGINS IN 2014 DC MAYORAL PRIMARY AND GENERAL ELECTION	WARD 4	WARD 7	WARD 5	WARD 8	WARD 1	WARD 6	WARD 2	WARD 3
PRIMARY	2,233	3,361	1,724	1,673	2,612 votes	402	2,297	6,786
			Vincent Gray		Bowser			
GENERAL	8,610	8,589	7,784	6,572	1,057	278	2,001 (Catania)	2,677

Two sentences could tell how the pattern of her support flipped. The areas she lost in the primary are the ones that made her win November.

Precincts, because of their size, offer a great opportunity for grouping with neighborhood demographics from Census. Among the strongest candidates are often race and income.

Race. For redistricting after each decennial Census, the Census Bureau publishes stats for precincts, aka voting districts. However, precinct boundaries often change when other districts are redrawn, so those stats may be outdated.

You can recreate precinct breakdown for stats like race by going back to the block-by-block data that Census publishes and adding up the block counts for precincts in your latest map.

How do you make this work?

Get block maps. Download files for your desktop GIS program from Census.gov. Search for TIGER/Line Shapefiles.

Get block data. Download block-level data from American FactFinder at Census.gov. You want a table like the 2010 SF1 table P11. It's a race and Hispanic breakdown for the voting age population.

Join the block data to the blocks. Overlay the blocks on your precinct map, and capture the precinct that each block falls in. Sum the block data for precincts.

Here's a look at results of that same mayor's race using three race groups based on Census data. Don't forget to visualize the subtotals, as well as the maps.
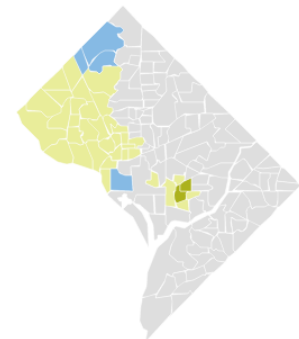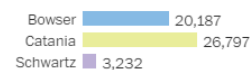

PRECINCTS BY PREDOMINANT RACE

Bowser won by **29,671** votes in **predominantly black** precincts.

Bowser	40,877
Catania	11,206
Schwartz	5,151
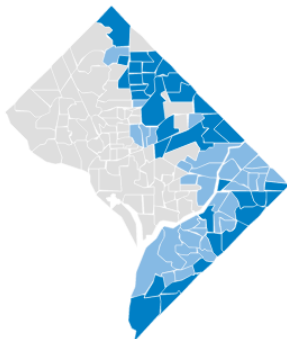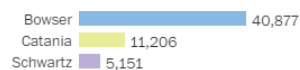
Bowser won by **4,624** votes in **mixed** precincts.

Bowser	19,239
Catania	14,615
Schwartz	2,200
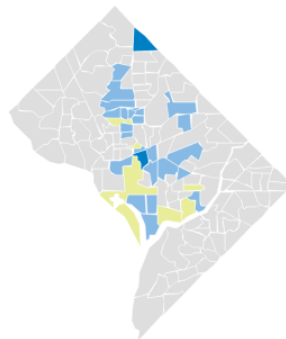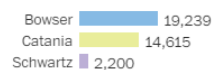
Catania won by **6,610** votes in **predominantly white** precincts.

Bowser	20,187
Catania	26,797
Schwartz	3,232

Two sentences: The winning candidate built up a large margin in predominantly black neighborhoods. She stayed competitive enough across the rest of the District to prevent an opponent from building a winning margin there.

Notice how the graphic above includes stats for each group. It's not enough just show where the groups are. If you leave out the group stats, it's a dumb map.

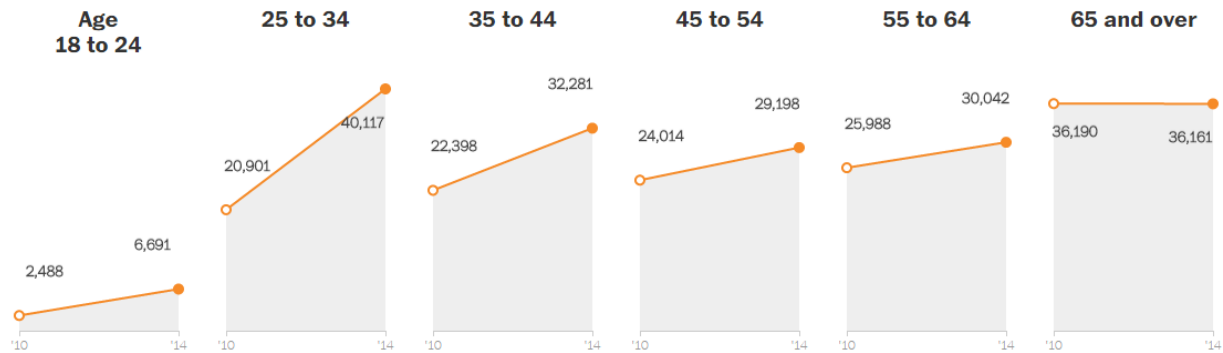**Income.** Similar process, except:

1) Block groups are your small geography, and your matching numbers will come from the American Community Survey 5-year data.

2) Instead of simple summing for precincts, you make a weighted mean of median household income. For each block group, multiple MHHI by number of households. Sum that product and sum households for each precinct. Then divide those two.

Who votes and how often?

Voting history and registration data typically includes each voter's age, party, precinct, date of registration and elections voted. If it doesn't include race, you can estimate from their precinct makeup.

Here's a summary voting by **age group** for the two D.C. mayor's elections. **Two sentences**. The number of voters 25 to 34 doubled, going from one of the smallest to the largest voter age group. Seniors, the perennial heavy voters, were outnumbered.

Age 18 to 24	25 to 34	35 to 44	45 to 54	55 to 64	65 and over

2,488   6,691   20,901   40,117   22,398   32,281   24,014   29,198   25,988   30,042   36,190   36,161

'10 '14   '10 '14   '10 '14   '10 '14   '10 '14   '10 '14

You might wonder what factors could be behind this surge in young voters. Population growth probably played a small role. It also happened that the second election included a referendum on **legalizing marijuana**. Here's an idea for showing in a map whether the patterns of young voters and of support for marijuana were related. These maps appear almost identical, and show a close relationship that would also appear in a scatterplot.
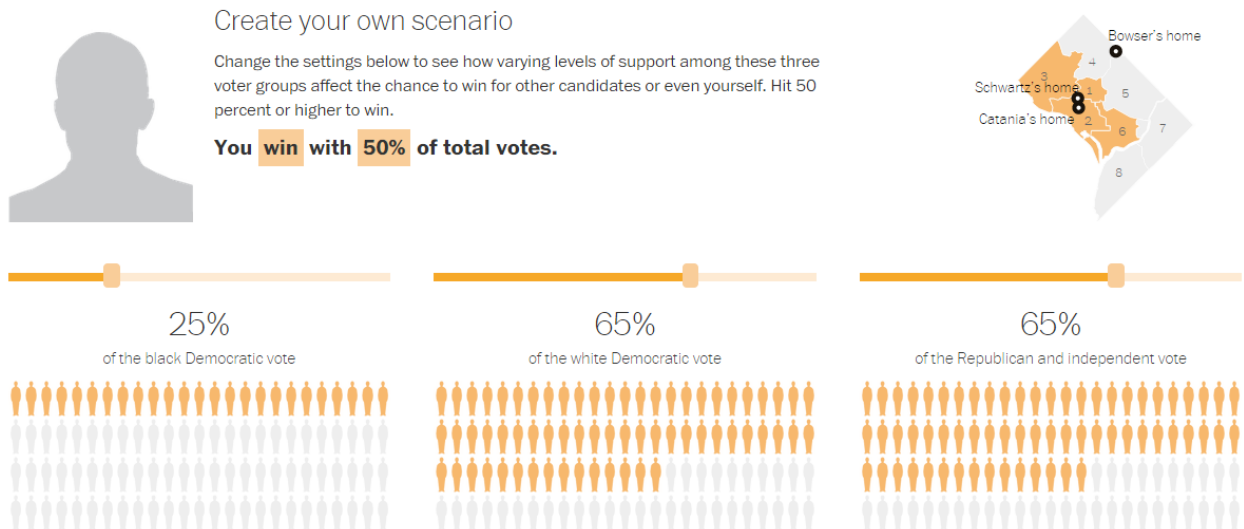
Share of voters aged 25 to 34

Share of voters for Initiative 71
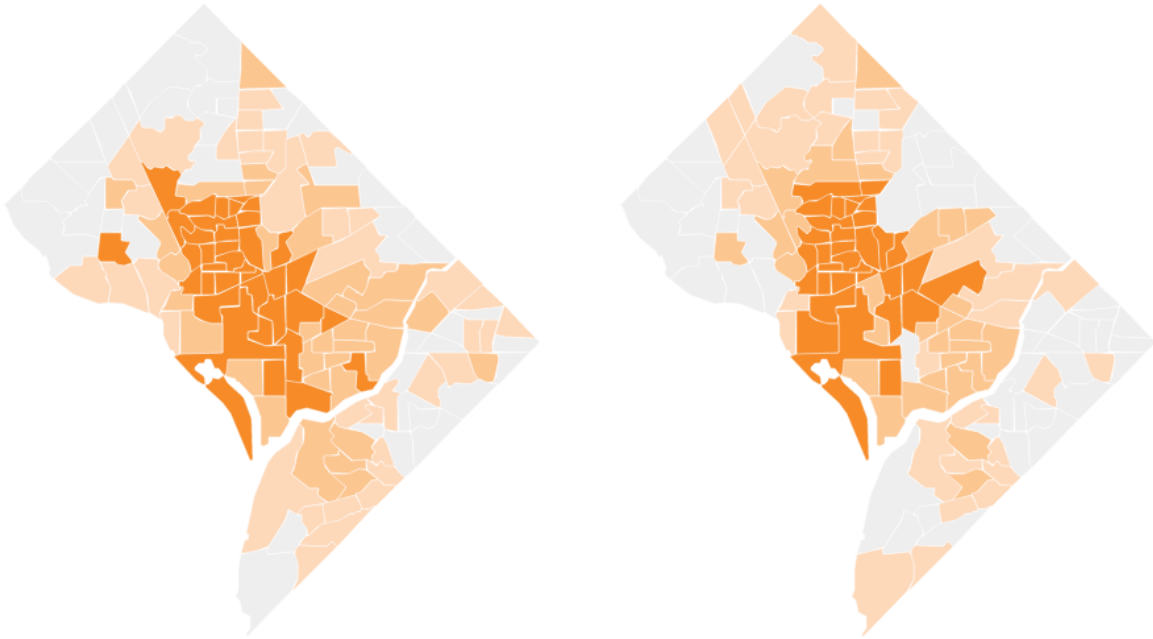
There are many more ways to use voting and election data for stories that go beyond who.

Finally, here's an idea to use of voter history data to look forward, at future scenarios. It's a screen shot of an interactive that gives a user sliders to adjust support among three voter groups, and then see, based on race and ward turnout patterns from the previous election, how the balance of votes shifts as a result.

Create your own scenario

Change the settings below to see how varying levels of support among these three voter groups affect the chance to win for other candidates or even yourself. Hit 50 percent or higher to win.

You win with 50% of total votes.

Bowser's home

Schwartz's home

Catania's home

3	4
1	5
8	
2	
6	7
8	

25%
of the black Democratic vote

65%
of the white Democratic vote

65%
of the Republican and independent vote

Ted.mellnik@washpost.com

@tmellnik   @postgraphics