

The Importance of Accessible Census Data Ahead of 2022 Polling

TO: INTERESTED PARTIES
FROM: RAGNAR RESEARCH PARTNERS
DATE: AUGUST 2022

The U.S. Census Bureau announced that 2022 demographic information will not be available until December 2022 and this is a major problem for polling this cycle. This memo will show some of the common solutions to this problem, such as pulling counts from file and modelled data. This memo will discuss the shortcomings to those solutions as well as our solution: [going directly to the U.S. Census](#) even though the demographic information for new political district lines will not be released.

The Problem That No One Wants To Talk About

For the uninitiated, [good polling requires being very specific about your target universe](#). Your definitions of who lives where and what a likely voter is can make or break your candidates and your reputation. Because of this we place a premium on accurately identifying gender, age, ethnicity/race, partisanship breakdowns, education, and other variables. While it has issues, [the single best source for much of this information comes from the ACS updates the U.S. Census regularly releases](#). Since this information will not be available until long after the window for its usefulness has closed this cycle, pollsters are resorting to other options. These options are not sufficient to the task.

The Less Favorable Options

[There are several ways that pollsters are addressing this problem](#). We will look at the most common, a slightly better approach, and then our solution.

The most common method is to [pull counts from the file](#). This is a common method for other pollsters because it is the easiest solution. When you purchase a list of people to include in a survey, you trust your sample vendors to have accurately gotten information on the voters in a district and trust that their coverage is complete. Essentially, you pass the responsibility to a vendor. [Maybe they're good, maybe not, but you won't know until E-Day](#).

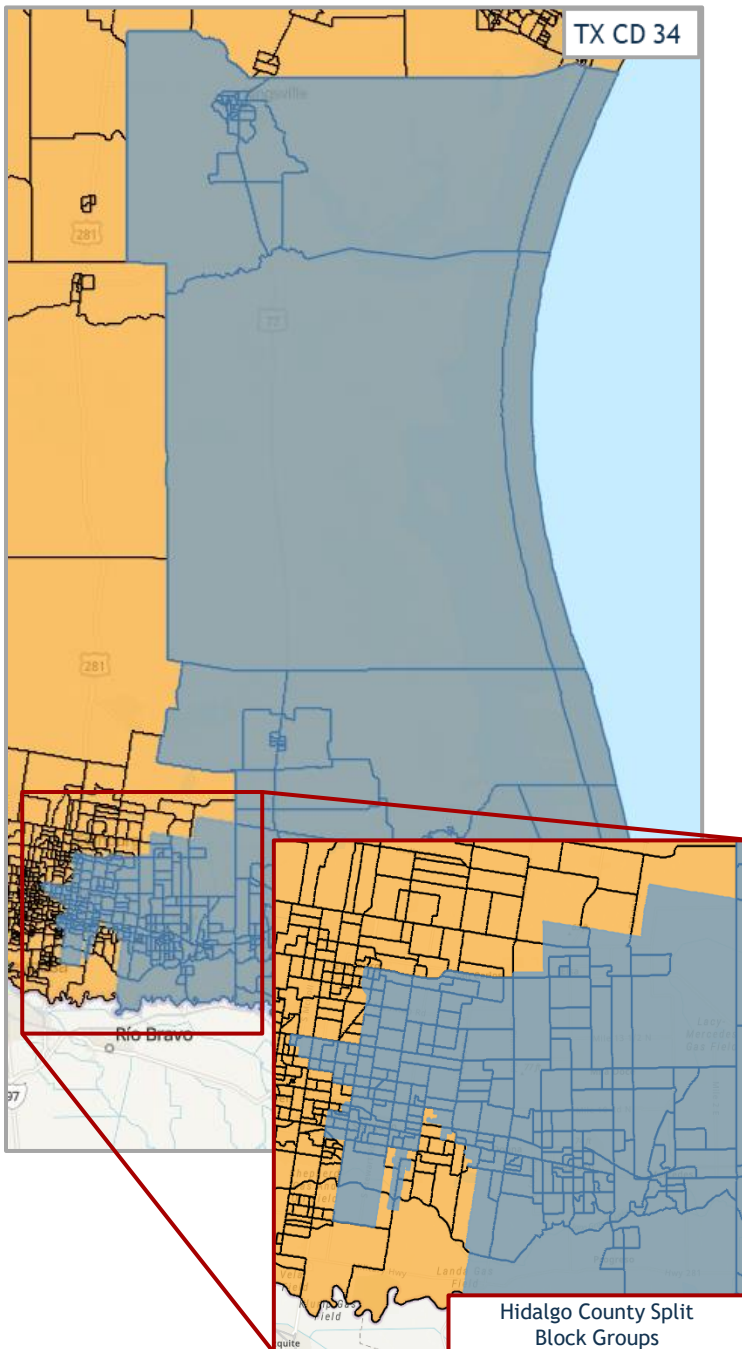
The second solution is to use [modeled data](#). This leverages large amounts of consumer data with smaller numbers of behavioral interviews, which are still large compared to poll interview counts.

The problem with modeled data is that unless you're collecting the data, you have very little control over how that behavioral data is collected. Is it from people outside of your district? Was it done via IVR, with shockingly low response rates? How heavily is the model relying on shared consumer variables vs behavioral signals? [In most cases, you simply don't know](#).

What if you skipped all that and just used [data from the U.S. Census](#)?

The Solution

Our solution is to avoid the less trustworthy methods and bypass the overworked Census bureaucracy by going straight to the source. By using some simple code in slightly complex ways, we programmed access to the Application Programming Interface (API) for Census data that **exists online but isn't available yet for current district lines**. We download this data by Block Group, which are population units between 600 and 3,000 people. We then match those groups up to the new districts using mapping software with a **nearly perfect match rate**, **allowing a much more realistic picture** of who is in a district to emerge.



Texas CD 34 - 2022 U.S. Census

Demographics (VAP)

		Only the Block Groups in State Analysis		Allowing for Additional 'Split' Block Groups	
		Count	% of Total	Count	% of Total
Race/Eth	White (alone)	52,149	10%	53,259	10%
	Hispanic	482,405	88%	489,902	88%
	Black (alone)	2,744	1%	2,751	0%
	Other	7,936	1%	8,057	1%
Education	<High School	95,056	22%	96,914	22%
	HS Grad	124,434	29%	126,371	29%
	Some College	133,691	31%	135,609	31%
	Bachelors	51,027	12%	51,741	12%
	Post Grad	20,957	5%	21,268	5%
Age	18-24	85,902	16%	87,190	16%
	25-34	99,275	19%	101,174	19%
	35-44	95,475	18%	97,394	18%
	45-64	154,078	29%	156,466	29%
	65+	97,798	18%	99,887	18%
Sex	Men	257,351	48%	261,776	48%
	Women	275,177	52%	280,335	52%

Note: TX CD 34 contains 5 counties: Cameron (100%), Kennedy (100%), Kleberg (100%), Willacy (100%), & Hidalgo (33%).

With this technique, **you can include partial Block Groups** that some states rule out in their counting, even though they are inside a district and their residents will vote in that district. This enables a level of **precision based on the most recent U.S. Census counts** that is otherwise unobtainable.

The Solution (continued)

Is that method of matching more work? It absolutely is. But we hear you, you don't care about that. You care if it's more accurate. Here are three Congressional Districts, with the demographic data pulled from three major sample vendors, the state (for available data), and the 2022 U.S. Census. **Differences greater than 4% from the 2022 Census have been highlighted.**

Demographics (VAP)		Texas CD 34					Nevada CD 3					Rhode Island CD 2				
		US Census	State	Vendor A	Vendor B	Vendor C	US Census	State	Vendor A	Vendor B	Vendor C	US Census	State	Vendor A	Vendor B	Vendor C
Race/Eth	White (alone)	10%	12%	12%	10%	14%	47%	48%	60%	52%	73%	77%	76%	83%	74%	89%
	Hispanic	88%	87%	87%	76%	85%	15%	18%	22%	17%	16%	13%	14%	13%	10%	10%
	Black (alone)	1%	1%	0%	1%	0%	7%	11%	4%	5%	2%	4%	4%	2%	2%	1%
	Other	1%	1%	1%	12%	0%	20%	24%	14%	26%	8%	7%	5%	2%	13%	NA
Education	<HS	22%	NA	9%	35%	55%	8%	NA	2%	41%	35%	8%	NA	3%	46%	35%
	HS Grad	29%	NA	14%	23%	39%	27%	NA	13%	19%	55%	29%	NA	17%	23%	55%
	Some Coll	31%	NA	13%	23%	39%	36%	NA	14%	19%	55%	31%	NA	14%	23%	55%
	Bachelors	12%	NA	12%	23%	39%	20%	NA	13%	19%	55%	20%	NA	19%	23%	55%
	Post Grad	5%	NA	4%	6%	5%	10%	NA	7%	7%	10%	12%	NA	11%	9%	9%
	Unknown	NA	NA	47%	35%	NA	NA	NA	51%	33%	NA	NA	NA	37%	22%	NA
Age	18-24	16%	NA	15%	14%	13%	9%	NA	9%	8%	10%	13%	NA	9%	8%	8%
	25-34	19%	NA	20%	20%	19%	20%	NA	19%	19%	18%	16%	NA	18%	18%	17%
	35-44	18%	NA	16%	16%	16%	18%	NA	18%	19%	18%	15%	NA	16%	16%	16%
	45-64	29%	NA	29%	29%	28%	33%	NA	32%	32%	32%	34%	NA	33%	33%	33%
	65+	18%	NA	21%	21%	23%	19%	NA	21%	22%	23%	22%	NA	25%	26%	26%
Sex	Men	48%	NA	47%	47%	43%	49%	NA	49%	45%	48%	48%	NA	48%	48%	46%
	Women	52%	NA	53%	52%	57%	51%	NA	49%	44%	52%	52%	NA	52%	52%	54%

As you can see, the **differences are stark**. While most, though not all, vendors are consistently reliable with age, **there is no consistent pattern to account for the differences with ethnicity, education, and sex**. Some vendors are just better with certain areas where their file is more up to date. The amount of HS education or less respondents, the numbers of Hispanics, the ratio of men vs women—**these are critical numbers** to accurately project and the current methods are simply not good enough.

Conclusion

Republicans have a chance for a historic cycle. We need to execute effectively in order to deliver on that potential. While we thought about writing a memo after the cycle explaining this process and its importance, we decided against that because, ultimately, **we, as a party, need to be smarter about our tactics and techniques**. We certainly would love to do more polling this cycle, but **we decided to “show a little leg”** ahead of E-Day in the hope that it would help consumers of polling data regardless of their pollster.

If you would like to reach out to us, you can reach us on our website, www.RagnarResearch.com, and best of luck this cycle!