Madi Baltagulov, Nat Markey

All of our code information is available in this posit cloud project:[DCWS compare calls by zip to responses by zipMadi Baltagulov]
Let Madi know if you have access issues.
All the output is available as .csv files. We can also upload them to github. We can produce any other necessary visualizations.

Project Notes

June 2nd, 2025.
Original task: Hi Madi and Nat - these are the files for the "DCWS compare calls by zip to responses by zip" project on the spreadsheet. One shows calls made, the other has the actual number of interviews completed. Can you use these to calculate measures such as the call:response ratio, the call:population ratio and the response:population ratio (using recent Census or ACS data by ZCTA for Connecticut). As a second question, could you describe to what extent these ratios differ by characteristics of ZIP codes (also using recent Census/ACS data for ZCTAs) such as homeownership rate, poverty rate, percent Latino, etc? We know that ZIP codes with more renters, lower incomes etc are harder places to do interviews, but I am interested in looking at to what extent this was the case in 2024 to help us plan the next survey. For further understanding of that question, it might help to group the ZIP Codes by quartiles (e.g., comparing lower-homeownership ZIP codes to all higher-homeownership ZIP codes on these measures). Note that there are areas like Hartford (e.g. 06120) where we do oversampling, that is relatively more phone calls or mailings, because we want to get a larger and more representative sample from that area. Mark
June 4th, 2025
Update: 1)Cut out ZIPs with sub 1000 population. 2)Whether it is rational/effective to make more calls in certain ZIPs? 3)How much more effort would it take to double responses in different ZIP codes, versus using the same amount of calls statewide? Cost-benefit analysis. 4)Is it worth it to make more calls in hard-to-reach ZIP codes (like 06120)? 5)How do we get the most representative (demographics, poverty, homeownership) responses for the state of Connecticut with the least number of calls? Where should the calls be focused? Look at areas with very high call/population and areas that are diverse but have low call/population ratios too. Ballpark 5$ a call cost and response is worth 100$. A response in a diverse ZIP code could be worth more, like 50% more per increase in diversity (age, race, homeownership) quartile. 6)Later task: maybe make a response/race group analysis.

Note: fixed the column I use for hispanic/latino. **DP05_0076E**
Note: fixed the diversity score counting very wealthy areas as diverse.

June 11th, 2025
Hi Madi and Nat - I reviewed the call project and it looks good. I'm wondering if you could factor in the attached data and see if the conclusions change at all (I think two separate analyses would work best, not a combination of age and race). These are files with responses by ZIP, by age and race.
The potential issue I identified earlier is that if we might have, say, 100 responses in a diverse ZIP code, but nearly all of those responses are from older white adults, then the model isn't as effective. It might be better to make calls in a ZIP code where we had 100 responses but we had a very good distribution by age and race.
If you weight the responses by race or age, I might suggest values something like this: Age 18-34, value of $200, age 35-64, value of $100, age 65+, value of $50. By race, maybe valuing responses from Black adults at $125, Latino adults at $150, and all other groups at $100. I'm guessing the conclusions might not change too much, but it could be interesting to see if the "cost per response" calculations change to some degree.

Madi's Model:

Diversity Scoring Methodology:

1. Four Diversity Dimensions (Equal Weight):

```
None
diversity_score = (racial_diversity/20 + economic_diversity/10 +
                   housing_diversity/20 + age_diversity/10) / 4
```

Rationale for Equal Weighting:

- **Racial diversity**: Distance from state % white population
- **Economic diversity**: Distance from state poverty rate
- **Housing diversity**: Distance from state homeownership rate
- **Age diversity**: Distance from state % elderly population

I gave each dimension **25% weight** because:

- All represent important axes of social stratification
- Survey representativeness requires coverage across all dimensions
- No obvious theoretical reason to prioritize one over others

2. Normalization Factors (Removed in my final code, but can be added back):

- **Racial/Housing**: Divided by 20 (expecting ~0-40% deviations from state mean)
- **Economic/Age**: Divided by 10 (expecting ~0-20% deviations from state mean)

Why These Ranges:

- Connecticut's racial composition varies widely (some ZIP codes 90%+ white, others <20%)
- Poverty rates typically range 0-30%, elderly populations 5-35%
- Normalization puts all dimensions on similar 0-2 scale before averaging

ROI Formula Logic:

Response Value Calculation:

```
None
response_value = $100 * (1 + 0.50 * (diversity_quartile - 1))
```

Value by Diversity Quartile:

- Q1 (least diverse): $100 per response
- Q2: $150 per response (+50%)
- Q3: $200 per response (+100%)
- Q4 (most diverse): $250 per response (+150%)

Rationale:

- Base value reflects standard survey response worth
- Diversity premium recognizes that responses from underrepresented areas have higher research value
- 50% bonus per quartile is aggressive but reflects real research priorities

ROI Calculation:

```
None
predicted_roi = (response_value - predicted_cost_per_response) /
predicted_cost_per_response
```

Standard Economic ROI:

- ROI = (Benefit - Cost) / Cost
- ROI > 1.0 means profitable (benefit exceeds cost)
- Incorporates both efficiency (cost per response) and value (diversity bonus)

Optimization Score Formula:

```
None
optimization_score = predicted_roi * diversity_quartile *
population_weight * 1000
```

Three-Factor Combination:

1. **predicted_roi**: Efficiency and value (how profitable)
2. **diversity_quartile**: Representativeness importance (1-4 scale)
3. **population_weight**: Coverage importance (larger populations get priority)
4. **×1000**: Scaling factor for readability

Logic:

- Multiplicative (not additive) because all factors must be present
- High-ROI + high-diversity + large population = maximum priority
- Low score on any dimension significantly reduces overall priority

Alternative Approaches to Consider:

1. Different Diversity Weights:

```
None
# Example: Emphasize racial/economic over housing/age
diversity_score = (racial_diversity/20 * 0.4 +
economic_diversity/10 * 0.4 +
                   housing_diversity/20 * 0.1 + age_diversity/10 *
0.1)
```

2. Different Value Functions:

```
None
# Linear instead of quartile-based
response_value = $100 + (diversity_score * $100)

# Exponential for high-diversity areas
response_value = $100 * (1.2 ^ diversity_quartile)
```

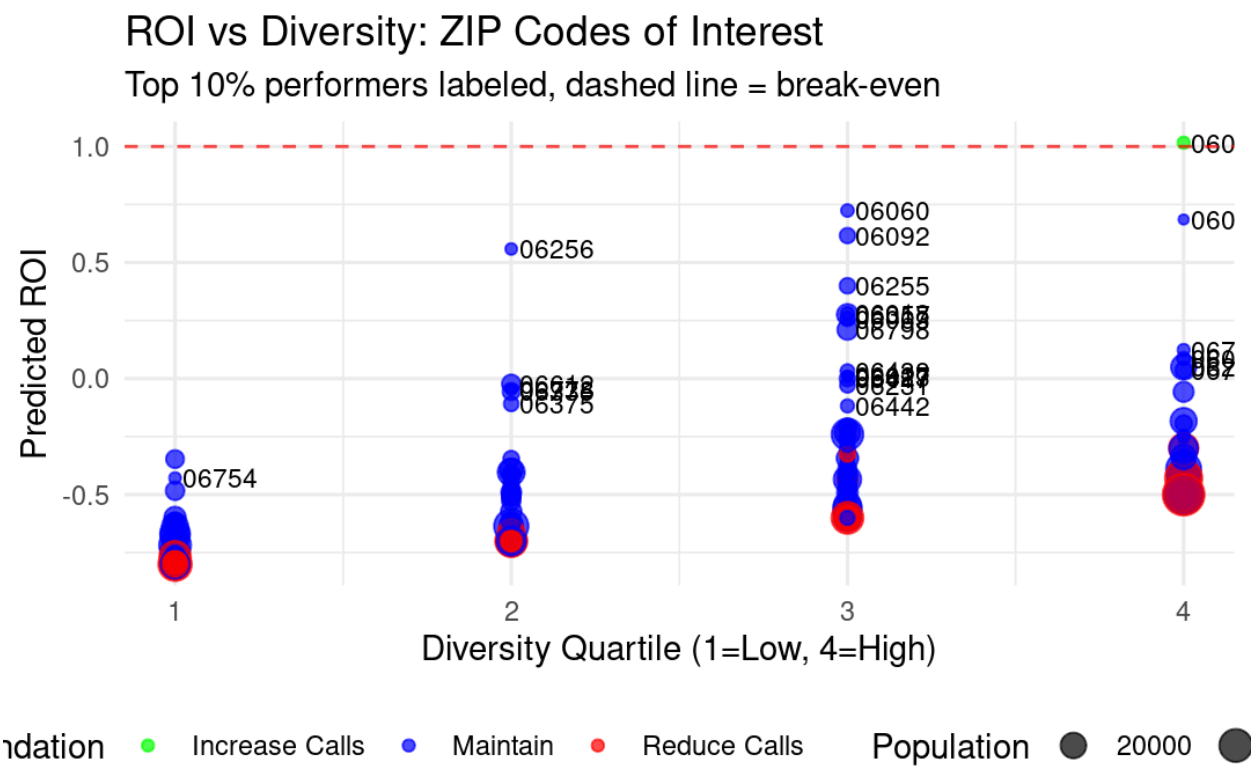
3. Different Optimization Weights:

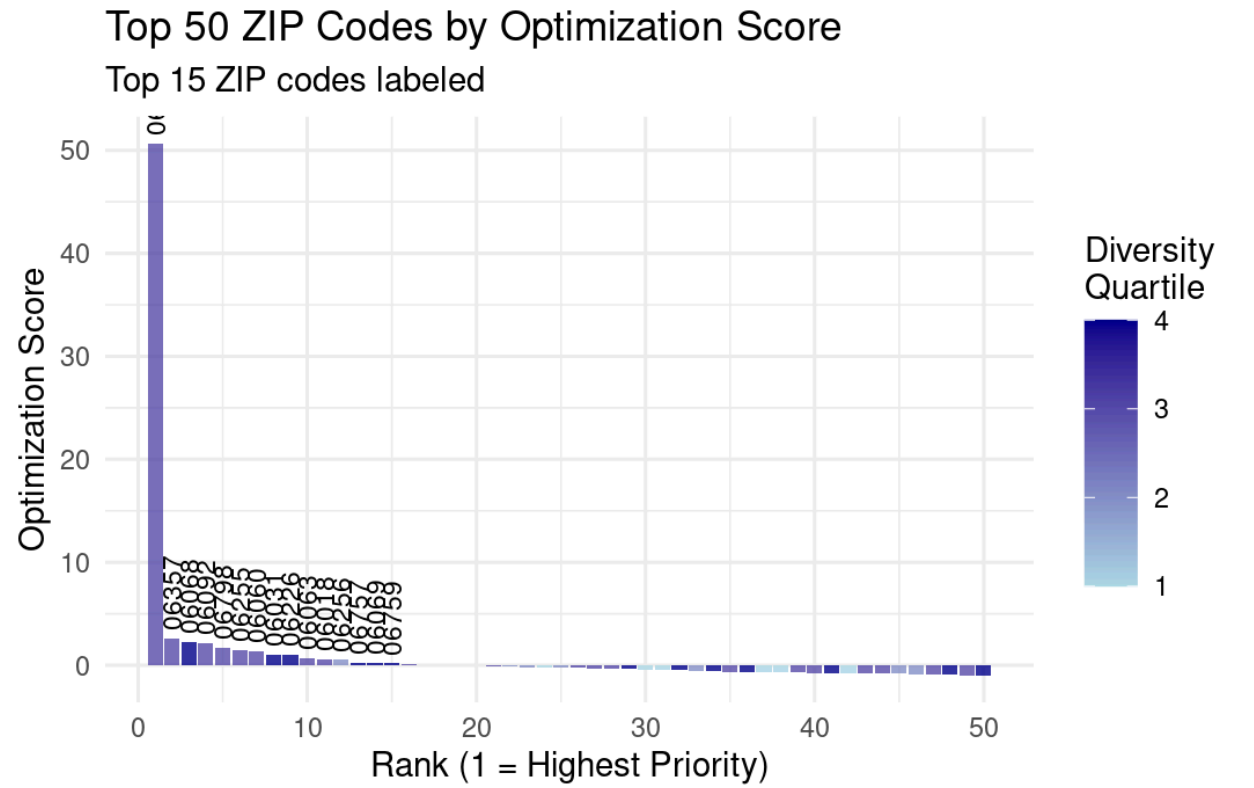
```
None
# Emphasize efficiency over diversity
optimization_score = (predicted_roi^2 * diversity_quartile *
population_weight)

# Add geographic clustering bonus
optimization_score = predicted_roi * diversity_quartile *
population_weight * clustering_bonus
```

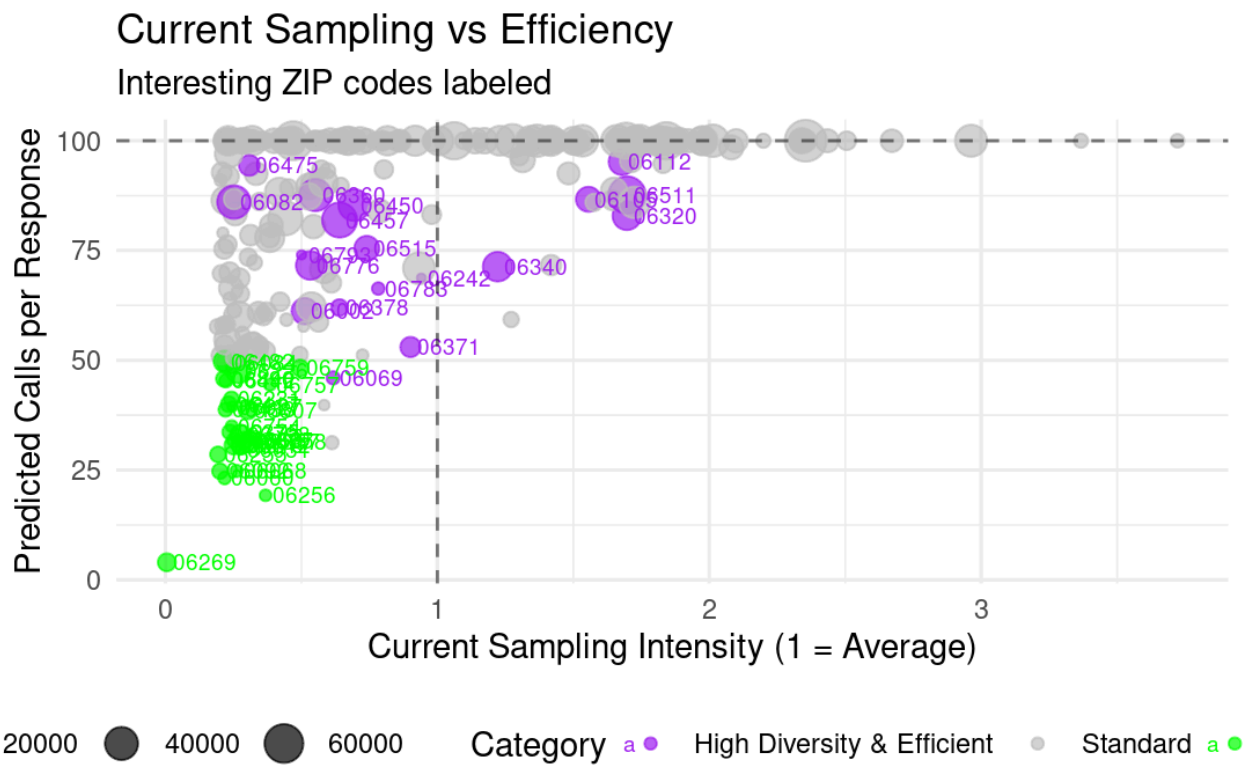
Key Assumptions to Validate:

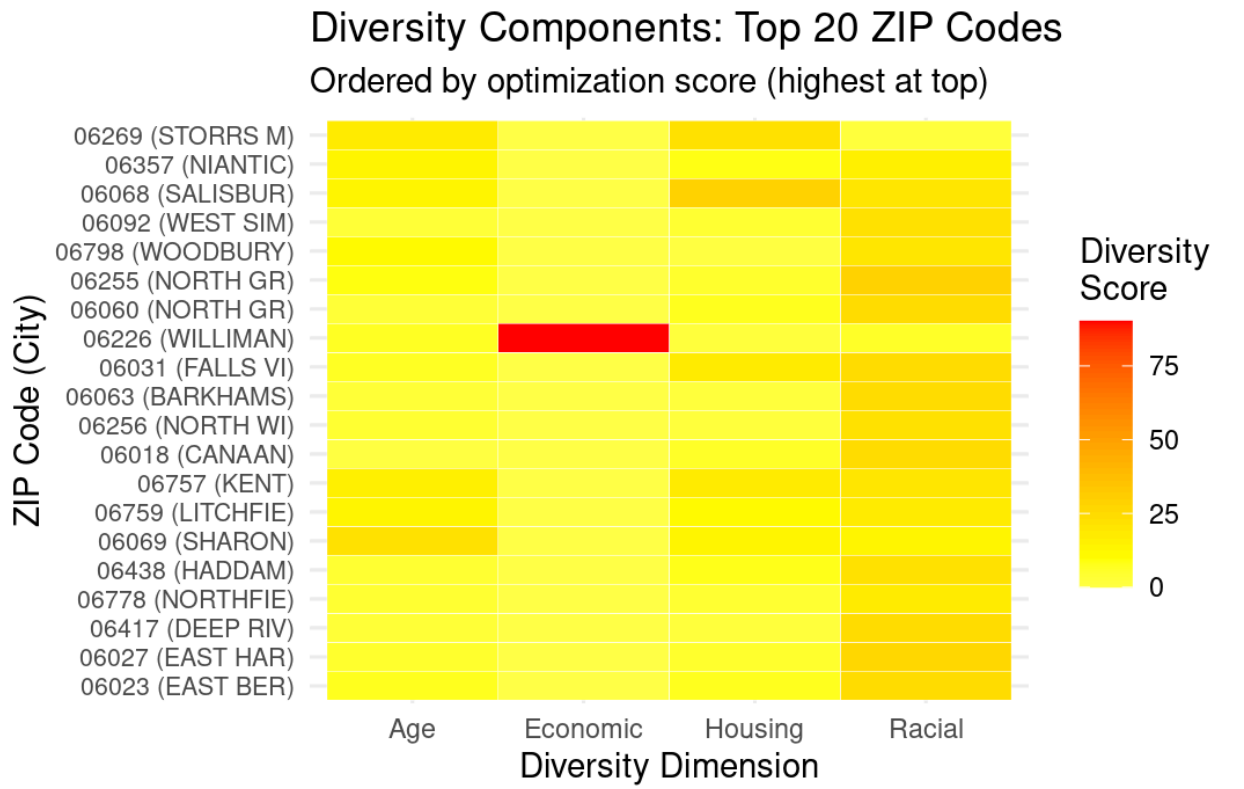
1. **Equal diversity dimension weights** - Should economic diversity matter more than age diversity?
2. **50% diversity bonus per quartile** - Is this too aggressive? Should it be 25% or 75%?
3. **Multiplicative optimization score** - Should factors be additive instead?
4. **Population weight inclusion** - Do you want to prioritize large areas or focus on efficiency regardless of size?

   Helpful Visuals:


ROI vs Diversity: ZIP Codes of Interest
Top 10% performers labeled, dashed line = break-even
Predicted ROI
1.0
0.5
0.0
-0.5
06754
06256
06060
06092
06255
06798
06442
060
060
Diversity Quartile (1=Low, 4=High)
1
2
3
4
ndation
Increase Calls
Maintain
Reduce Calls
Population
20000


Top 50 ZIP Codes by Optimization Score
Top 15 ZIP codes labeled
Optimization Score
50
40
30
20
10
0
06357
06068
06092
06798
06255
06360
06031
06263
06018
06757
06759
0
10
20
30
40
50
Rank (1 = Highest Priority)
Diversity Quartile
4
3
2
1





Top areas to reduce calling:

	zip_code	city	type	total_population	pct_white	pct_black	pct_hispanic	poverty_rate	homeownership_rate	pct_65_plus	diversity_score	diversity_quartile	racial_diversity	economi
1	06235	CHAPLIN	STANDARD	2116	88.2	0.6	4.5	0	95.9	18.1	5.99	2	20.67855400	
2	06510	NEW HAVEN	STANDARD	4205	47.6	20.7	12.9	112	85.7	7.8	9.28	3	19.89671309	
3	06103	HARTFORD	STANDARD	2845	47.9	22.0	15.7	62	79.7	3.8	11.73	4	19.56294022	
4	06280	WINDHAM	STANDARD	3179	87.2	3.1	25.0	0	99.8	28.9	9.41	3	19.69053063	
5	06333	EAST LYME	STANDARD	7323	77.9	1.3	9.2	0	97.3	16.1	4.23	1	10.39852890	
6	06702	WATERBURY	STANDARD	3117	32.0	39.6	36.0	48	87.7	28.9	12.83	4	35.55289947	
7	06052	NEW BRITAIN	STANDARD	8232	58.8	11.4	38.1	63	91.3	17.1	2.77	1	8.68745922	
8	06339	LEDYARD	STANDARD	9263	83.9	2.0	9.4	0	96.5	17.8	5.13	1	16.40779734	
9	06906	STAMFORD	STANDARD	9472	51.9	9.7	40.7	43	92.0	14.9	4.88	1	15.62747823	
10	06403	BEACON FALLS	STANDARD	6089	82.4	6.8	8.4	100	99.4	25.0	7.11	2	14.93704983	
11	06418	DERBY	STANDARD	12150	69.7	11.9	21.3	49	94.2	20.3	1.48	1	2.21346341	
12	06903	STAMFORD	STANDARD	13630	75.8	6.3	8.6	18	94.1	21.6	3.28	1	8.31868398	
13	06517	HAMDEN	STANDARD	14867	61.2	22.1	16.7	96	91.1	22.3	3.03	1	6.25695344	
14	06461	MILFORD	STANDARD	15256	78.6	1.5	11.2	77	95.3	18.8	3.59	1	11.08532817	
15	06907	STAMFORD	STANDARD	9071	50.5	24.9	23.1	138	95.2	13.8	5.93	2	16.96100613	
16	06098	WINSTED	STANDARD	10200	88.3	1.8	8.6	126	89.4	27.6	8.40	3	20.74820076	
17	06483	SEYMOUR	STANDARD	16799	70.3	7.2	18.6	15	97.7	16.6	2.31	1	2.80700795	
18	06615	STRATFORD	STANDARD	16865	58.9	20.0	24.4	159	93.6	19.0	2.61	1	8.61550642	
19	06062	PLAINVILLE	STANDARD	17499	80.5	2.6	11.3	19	94.8	19.7	4.18	1	12.98361253	
20	06901	STAMFORD	STANDARD	10506	49.9	23.8	18.0	78	89.7	11.8	6.76	2	17.64947676	

Two areas to increase calling:

	zip_code	city	type	total_population	pct_white	pct_black	pct_hispanic	poverty_rate	homeownership_rate	pct_65_plus	diversity_score	diversity_quartile	racial_diversity	ec
1	06269	STORRS MANSFIELD	UNIQUE	6723	65.3	11.0	6.8	0	71.2	0.6	10.32	3	2.223346	
2	06068	SALISBURY	STANDARD	2034	88.2	0.8	4.0	0	64.4	30.5	15.38	4	20.743053	

Please check the “zip_analysis_summary” df in my code, it is much more comprehensive. My formula also allows us to easily change weights for different categories, if we would want to increase the weight of racial diversity. The optimization recommendations operate under the assumption that we will be making less calls overall and the budget for this will get reduced.

Update: Using actual Age and Race response files by ZIP

 # For each ZIP code: age_weighted_value = (young_responses * $200 + middle_responses * $100 + older_responses *$50) / total_responses
race_weighted_value = (black_responses * $125 + hispanic_responses * $150 + other_responses *$100) / total_responses
combined_value = (age_weighted_value + race_weighted_value) / 2

Youngest:

	zip_code	city	age_weighted_value	pct_responses_young	pct_responses_middle	pct_responses_older	responses
1	06481	ROCKFALL	200	100.0	0.0	0.0	1
2	06254	NORTH FRANKLIN	200	100.0	0.0	0.0	1
3	06384	VOLUNTOWN	167	66.7	33.3	0.0	3
4	06088	EAST WINDSOR	167	66.7	33.3	0.0	3
5	06269	STORRS MANSFIELD	150	50.0	50.0	0.0	2
6	06365	PRESTON	150	66.7	0.0	33.3	3
7	06417	DEEP RIVER	142	50.0	33.3	16.7	6
8	06073	SOUTH GLASTONBURY	140	60.0	0.0	40.0	5
9	06021	COLEBROOK	133	33.3	66.7	0.0	3
10	06235	CHAPLIN	133	33.3	66.7	0.0	3
11	06103	HARTFORD	132	47.4	21.1	31.6	19
12	06801	BETHEL	131	46.2	23.1	30.8	13
13	06060	NORTH GRANBY	130	40.0	40.0	20.0	5
14	06608	BRIDGEPORT	130	32.5	62.5	5.0	40
15	06605	BRIDGEPORT	130	35.6	53.4	11.0	73
16	06016	BROAD BROOK	127	30.8	61.5	7.7	13
17	06510	NEW HAVEN	126	38.1	38.1	23.8	21
18	06480	PORTLAND	126	38.1	38.1	23.8	21
19	06778	NORTHFIELD	125	25.0	75.0	0.0	4
20	06777	NEW PRESTON MARBLE DALE	125	50.0	0.0	50.0	2

Most racially diverse:

	zip_code	city	race_weighted_value	pct_responses_black	pct_responses_hispanic	pct_responses_white	responses
1	06031	FALLS VILLAGE	133	0.0	66.7	0.0	3
2	06021	COLEBROOK	133	0.0	66.7	33.3	3
3	06608	BRIDGEPORT	132	35.0	47.5	10.0	40
4	06016	BROAD BROOK	129	7.7	53.8	38.5	13
5	06704	WATERBURY	128	20.9	47.8	23.9	67
6	06607	BRIDGEPORT	127	51.7	34.5	6.9	29
7	06120	HARTFORD	127	61.3	22.6	11.3	62
8	06519	NEW HAVEN	126	37.5	33.9	21.4	56
9	06112	HARTFORD	126	74.4	14.4	4.4	90
10	06108	EAST HARTFORD	126	26.9	38.5	30.8	26
11	06605	BRIDGEPORT	126	35.6	37.0	21.9	73
12	06702	WATERBURY	125	22.2	38.9	33.3	18
13	06114	HARTFORD	124	21.2	37.9	33.3	66
14	06106	HARTFORD	123	16.9	38.6	39.8	83
15	06610	BRIDGEPORT	122	37.7	29.0	20.3	69
16	06051	NEW BRITAIN	121	12.7	36.5	49.2	63
17	06606	BRIDGEPORT	121	30.8	28.0	32.9	143
18	06259	POMFRET CENTER	120	0.0	40.0	50.0	10
19	06706	WATERBURY	119	9.1	33.3	57.6	33
20	06705	WATERBURY	119	31.7	21.7	40.0	60

Most valuable overall, when including age+race:

	zip_code	city	demographic_weighted_response_value	race_weighted_value	age_weighted_value	pct_responses_young	pct_responses_black	pct_responses_hispanic	responses
1	06481	ROCKFALL	150	100	200	100.0	0.0	0.0	1
2	06254	NORTH FRANKLIN	150	100	200	100.0	0.0	0.0	1
3	06088	EAST WINDSOR	142	117	167	66.7	0.0	33.3	3
4	06021	COLEBROOK	133	133	133	33.3	0.0	66.7	3
5	06384	VOLUNTOWN	133	100	167	66.7	0.0	0.0	3
6	06608	BRIDGEPORT	131	132	130	32.5	35.0	47.5	40
7	06016	BROAD BROOK	128	129	127	30.8	7.7	53.8	13
8	06605	BRIDGEPORT	128	126	130	35.6	35.6	37.0	73
9	06269	STORRS MANSFIELD	125	100	150	50.0	0.0	0.0	2
10	06235	CHAPLIN	125	117	133	33.3	0.0	33.3	3
11	06417	DEEP RIVER	125	108	142	50.0	0.0	16.7	6
12	06365	PRESTON	125	100	150	66.7	0.0	0.0	3
13	06108	EAST HARTFORD	125	126	125	34.6	26.9	38.5	26
14	06704	WATERBURY	124	128	120	28.4	20.9	47.8	67
15	06103	HARTFORD	122	112	132	47.4	15.8	15.8	19
16	06073	SOUTH GLASTONBURY	122	105	140	60.0	20.0	0.0	5
17	06607	BRIDGEPORT	122	127	117	20.7	51.7	34.5	29
18	06610	BRIDGEPORT	122	122	121	30.4	37.7	29.0	69
19	06702	WATERBURY	121	125	117	22.2	22.2	38.9	18
20	06120	HARTFORD	121	127	115	27.4	61.3	22.6	62

Highest optimization score (diverse and we already get decent # of responses there):

	zip_code	city	optimization_score	demographic_weighted_response_value	predicted_roi	sampling_intensity	responses	current_vs_optimal
1	06269	STORRS MANSFIELD	123.2	125	525.0	0.01	2	Increase Calls
2	06256	NORTH WINDHAM	0.3	102	5.5	0.37	8	Maintain
3	06060	NORTH GRANBY	0.1	118	1.3	0.22	5	Maintain
4	06031	FALLS VILLAGE	-1.0	117	-21.3	0.28	3	Maintain
5	06778	NORTHFIELD	-1.5	112	-28.6	0.33	4	Maintain
6	06068	SALISBURY	-1.6	80	-35.5	0.26	5	Maintain
7	06027	EAST HARTLAND	-1.9	100	-49.4	0.25	2	Maintain
8	06023	EAST BERLIN	-2.0	100	-49.7	0.58	5	Maintain
9	06242	EASTFORD	-2.2	81	-76.4	0.94	4	Maintain
10	06754	CORNWALL BRIDGE	-2.2	75	-57.1	0.24	3	Maintain
11	06092	WEST SIMSBURY	-2.3	100	-19.2	0.20	8	Maintain
12	06793	WASHINGTON	-2.6	100	-73.0	0.50	2	Maintain
13	06018	CANAAN	-2.7	99	-36.7	0.61	12	Maintain
14	06757	KENT	-2.8	75	-66.3	0.61	4	Maintain
15	06021	COLEBROOK	-2.9	133	-53.8	0.51	3	Maintain
16	06058	NORFOLK	-3.0	96	-62.5	0.72	6	Maintain
17	06750	BANTAM	-3.0	100	-80.0	0.38	0	Maintain
18	06063	BARKHAMSTED	-3.4	100	-37.2	0.25	6	Maintain
19	06081	TARIFFVILLE	-3.4	100	-74.7	0.21	1	Maintain
20	06752	BRIDGEWATER	-3.4	88	-82.5	0.69	2	Maintain

Nat's Datasets

Raw Merged Dataset

Available here.

This dataset includes call and response data plus selected census metrics for all CT ZCTA's present in both the DataHaven dataset and the Census data.

Census data was sourced from 2023 American Community Survey 5-year-estimate S1701, DP04, and DP05 datasets.

Column descriptions:

- **ZIP**: ZCTA 5-digit ZIP code
- **Type:** Type of ZIP code
- **City:** City/town in CT covered by ZIP code
- **calls:** number of unique numbers called in ZCTA, sourced from DataHaven_2024_Unique_Phone_Numbers.xlsx “All” column
- **responses:** total number of responses in ZCTA, sourced from ZIPFrequencies.xls “Frequency” column
- **r_c_ratio**: **responses / calls** (response rate)
- **c_r_ratio: calls / responses** (call to response ratio)
- **GEO_ID**: ZCTA unique geographic identifier
- **POV_STATUS_DET:** Estimate of total population for whom poverty status is determined, sourced from S1701 dataset “S1701_C01_001E” column.
- **POV_RATE**: Estimated percent below poverty level of total population for whom poverty status is determined, sourced from S1701 dataset “S1701_C01_001E” column.
- **TOTAL_POP**: Estimated total population, sourced from DP05 dataset “DP05_0001E” column.
- **MED_AGE:** Estimated population median age (years), sourced from DP05 dataset “DP05_0018E” column.
- **U18_POP**: Estimated total population under 18 years, sourced from DP05 dataset “DP05_0019E” column.
- **O65_POP:** Estimated total population 65 years and over, sourced from DP05 dataset “DP05_0024E” column.
- **WHITE_POP:** Estimated total white population, sourced from DP05 dataset “DP05_0037E” column.
- **BLACK_POP:** Estimated total Black or African-American population, sourced from DP05 dataset “DP05_0038E” column.
- **HISP_POP:** Estimated total Hispanic or Latino (of any race) population, sourced from DP05 dataset “DP05_0076E” column.
- **WHITE_NH_POP:** Estimated total population white alone (not hispanic or latino), sourced from DP05 dataset “DP05_0082E” column.
- **TOT_HU:** Estimated total housing units, sourced from DP04 dataset “DP04_0001E” column.
- **TOT_OCC_HU:** Estimated total occupied housing units, sourced from DP04 dataset “DP04_0002E” column.
- **OWN_OCC_HU**: Estimate total owner-occupied housing units, sourced from DP04 dataset “DP04_0080E” column.
- **WHITE_PCT**: White percentage of the population (**WHITE_POP** / **TOTAL_POP**)
- **BLACK_PCT**: Black percentage of the population (**BLACK_POP** / **TOTAL_POP**)
- **HISP_PCT**: Hispanic percentage of the population (**HISP_POP** / **TOTAL_POP**)
- **WHITE_NH_PCT**: White non-hispanic percentage of the population (**WHITE_NH_POP** / **TOTAL_POP**)
- **HOMEOWNERSHIP**: Homeownership rate (**OWN_OCC_HU / TOT_OCC_HU**)

- **c_p_ratio: calls/TOTAL_POP** (call to population ratio)
- **r_p_ratio: responses/TOTAL_POP** (response to population ratio)
- **pov_quartile:** ZIP code's quartile on **POV_RATE** (higher quartile = higher poverty rate)
- **white_nh_quartile:** ZIP code's quartile on **WHITE_NH_PCT** (higher quartile = more white)
- **hisp_quartile:** ZIP code's quartile on **HISP_PCT** (higher quartile = more Hispanic)
- **black_quartile:** ZIP code's quartile on **BLACK_PCT** (higher quartile = more Black)
- **homeown_quartile:** ZIP code's quartile on **HOMEOWNERHSIP** (higher quartile = more homeownership)

Filtered Dataset for Modeling

Available here.

This dataset is filtered to only include ZCTAs with total populations greater than 1000. It includes all columns in the raw merged dataset, plus additional columns for cost/benefit modeling.

Descriptions of added columns:

- **m_pctle:** ZCTA's percentile on % minority, which is calculated as 1 - **WHITE_NH_PCT**
- **p_pctle:** ZCTA's percentile on poverty rate (**POV_RATE**)
- **h1_pctle:** ZCTA's percentile on (1 - **HOMEOWNERSHIP**)
- **r_val:** benefit of a response from a ZIP as measured using an earlier model, model 0, structured the same way as model N1 below, but with less weight on demographic characteristics.
- **r_val_n1:** [explained below under "Model N1"]
- **benefit:** total benefit of all of a ZIP's responses in model 0 (**r_val * responses**)
- **value:** benefit - costs in ZIP code using model 0 (**benefit - calls**)
- **val_per_call:** value of each call in a ZIP code using model 0 (**value / calls**)
- **benefit_n1:** [explained below under "Model N1"]
- **value_n1:** [explained below under "Model N1"]
- **val_per_call_n1:** [explained below under "Model N1"]
- **r_val_n2:** [explained below under "Model N2"]
- **benefit_n2:** [explained below under "Model N2"]
- **value_n2:** [explained below under "Model N2"]
- **val_per_call_n2:** [explained below under "Model N2"]

Nat's Models

Major differences: Possibly more weight on demographics of ZIP codes. However, models do not factor in the level of sampling (i.e. calls/population). Also, model N1 does not factor in age.

Model N1

Assumptions:

- The dataset is limited to ZIP codes with a population of at least 1000.
- Every phone call costs DataHaven $1
- The value of a response is, on average, $100. However, the value of responses varies based on the demographic characteristics of ZIP codes, with above-median minority percentage, above-median poverty rate, and below-median homeownership rate increasing the value of a ZIP code's responses.
- Similarly, if ZIP codes deviate from the median on these three metrics in the opposite direction, their responses become less valuable.
- Specifically, the value of a response is directly proportional to its **percentile** across all ZIP codes in the dataset on percentage minority residents, poverty rate, and 1 - homeownership rate
- A response from a theoretical ZIP code that is in the 50th percentile on all three metrics is worth $100
- A response from a theoretical ZIP code that is in the 0th percentile on all three metrics is worth $50
- A response from a theoretical ZIP code that is in the 100th percentile on all three metrics is worth $150

These assumptions yield the following model for response value:

R = 50 + ⅓ (**M** + **P** + **H**)**,**

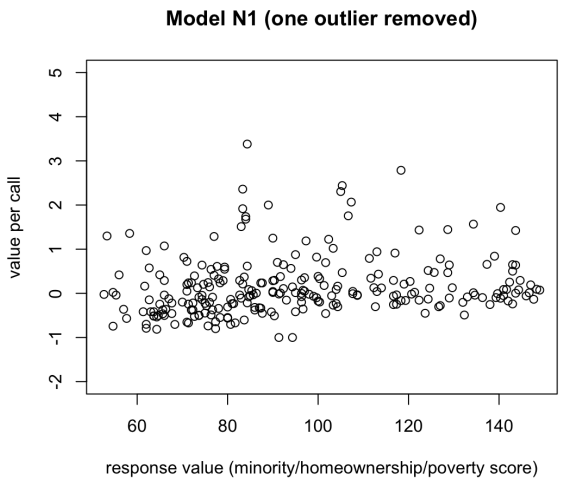
Where **R** = the value of a response in a given ZIP code, **M** = the ZIP code's percentile on % minority population, **P** = the ZIP code's percentile on poverty rate, and **H** = the ZIP code's percentile on (1 - homeownership rate)

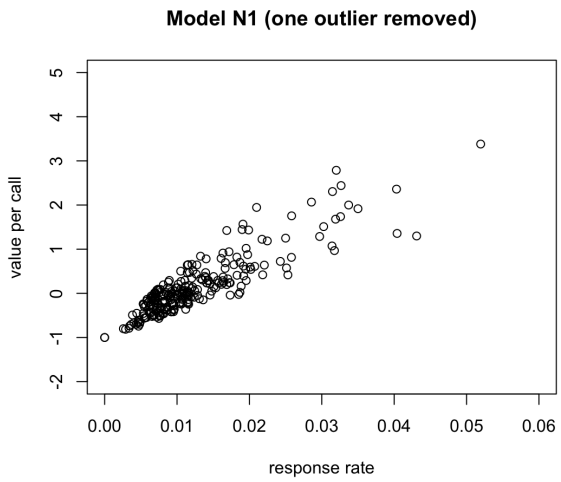
Results:

Using this model, I calculated the following metrics:

- **r_val_n1**: response value, calculated by the formula above
- **benefit_n1**: total benefit of all responses in a ZIP code, calculated as **responses*r_val_n1**
- **value_n1**: benefits - costs in ZIP code, calculated as **benefit_n1 - calls** (recall $1/call cost assumption)
- **val_per_call_n1**: value of each call made in a ZIP code, calculated as **value_n1/calls**. Intended as the primary metric used to assess efficacy of calls in a ZIP code

Value per call results were more closely related to the response rate than the demographics of a ZIP code:





The top twenty most valuable-to-call ZIP codes were the following:

	ZIP	City	TOTAL_POP	POV_RATE	WHITE_NH_PCT	HOMEOWNERSHIP	calls	responses	r_c_ratio	c_p_ratio	r_p_ratio	r_val_n1	value_n1	val_per_call_n1
1	06269	STORRS MANSFIELD	6723	49.8	0.64	0.27	8	2	0.250	0.0012	0.0003	139	269	33.7
2	06256	NORTH WINDHAM	1796	6.6	0.82	0.88	154	8	0.052	0.0857	0.0045	84	521	3.4
3	06018	CANAAN	2630	22.2	0.81	0.59	375	12	0.032	0.1426	0.0046	118	1045	2.8
4	06016	BROAD BROOK	6833	7.1	0.71	0.72	398	13	0.033	0.0582	0.0019	105	971	2.4
5	06068	SALISBURY	2034	4.0	0.86	0.72	124	5	0.040	0.0610	0.0025	83	293	2.4
6	06335	GALES FERRY	6100	14.6	0.75	0.83	413	13	0.031	0.0677	0.0021	105	952	2.3
7	06754	CORNWALL BRIDGE	1849	15.8	0.72	0.84	105	3	0.029	0.0568	0.0016	107	217	2.1
8	06031	FALLS VILLAGE	1385	13.8	0.93	0.82	89	3	0.034	0.0643	0.0022	89	178	2.0
9	06226	WILLIMANTIC	19068	25.9	0.44	0.35	1191	25	0.021	0.0625	0.0013	140	2317	1.9
10	06255	NORTH GROSVENORDALE	4442	6.6	0.94	0.75	200	7	0.035	0.0450	0.0016	83	383	1.9
11	06807	COS COB	7667	5.6	0.63	0.68	542	14	0.026	0.0707	0.0018	107	951	1.8
12	06612	EASTON	7612	7.7	0.86	0.85	491	16	0.033	0.0645	0.0021	84	853	1.7
13	06357	NIANTIC	11315	3.9	0.83	0.77	815	26	0.032	0.0720	0.0023	84	1369	1.7
14	06380	TAFTVILLE	2427	16.7	0.57	0.34	157	3	0.019	0.0647	0.0012	134	246	1.6
15	06798	WOODBURY	9787	5.7	0.87	0.79	628	19	0.030	0.0642	0.0019	83	949	1.5
16	06040	MANCHESTER	36171	14.0	0.53	0.54	2685	51	0.019	0.0742	0.0014	129	3877	1.4
17	06042	MANCHESTER	23302	10.0	0.56	0.56	1307	26	0.020	0.0561	0.0011	122	1874	1.4
18	06510	NEW HAVEN	4205	30.7	0.44	0.06	1245	21	0.017	0.2961	0.0050	144	1772	1.4
19	06092	WEST SIMSBURY	4239	2.5	0.88	0.96	198	8	0.040	0.0467	0.0019	58	269	1.4
20	06060	NORTH GRANBY	2290	0.0	0.91	1.00	116	5	0.043	0.0507	0.0022	53	151	1.3

Note that ZIP code 06269 (Storrs Mansfield) is a major outlier due to its small sample size (8 calls, 2 responses) and resulting extremely high response rate (25%). In general, many of these top ZIP codes have a low number of calls and responses, suggesting that there could be a high degree of randomness in their response rate.

However, ZIP codes included here do tend to have high response rates. Some also have high response values, indicating desired demographic traits (e.g. Willimantic, Taftville, Manchester, and New Haven 06510). However, some others have low response values (e.g. North Granby,

West Simsbury, North Windham, Salisbury), which are simply outweighed by high response rates.

The top twenty least-valuable to call ZIP codes were the following:

	ZIP	City	TOTAL_POP	POV_RATE	WHITE_NH_PCT	HOMEOWNERSHIP	calls	responses	r_c_ratio	c_p_ratio	r_p_ratio	r_val_n1	value_n1	val_per_call_n1
1	06089	WEATOGUE	3548	1.9	0.83	0.48	183	0	0.0000	0.052	0.00000	91	-183	-1.00
2	06750	BANTAM	1324	9.7	0.98	0.64	117	0	0.0000	0.088	0.00000	94	-117	-1.00
3	06035	GRANBY	7939	3.8	0.91	0.90	1037	3	0.0029	0.131	0.00038	64	-844	-0.81
4	06782	PLYMOUTH	2098	6.1	0.85	0.90	387	1	0.0026	0.184	0.00048	77	-310	-0.80
5	06248	HEBRON	5464	2.8	0.93	0.87	297	1	0.0034	0.054	0.00018	62	-235	-0.79
6	06365	PRESTON	4781	1.6	0.93	0.92	641	3	0.0047	0.134	0.00063	55	-477	-0.74
7	06426	ESSEX	3285	5.5	0.93	0.81	289	1	0.0035	0.088	0.00030	76	-213	-0.74
8	06716	WOLCOTT	16058	1.5	0.84	0.87	2766	12	0.0043	0.172	0.00075	68	-1946	-0.70
9	06812	NEW FAIRFIELD	13558	4.1	0.75	0.90	2720	10	0.0037	0.201	0.00074	81	-1913	-0.70
10	06763	MORRIS	2078	4.3	0.95	0.92	415	2	0.0048	0.200	0.00096	62	-291	-0.70
11	06853	NORWALK	3229	3.3	0.90	0.82	867	4	0.0046	0.269	0.00124	71	-582	-0.67
12	06235	CHAPLIN	2116	6.9	0.87	0.84	741	3	0.0040	0.350	0.00142	82	-496	-0.67
13	06281	WOODSTOCK	7417	4.5	0.93	0.83	612	3	0.0049	0.083	0.00040	71	-399	-0.65
14	06897	WILTON	18439	2.5	0.75	0.88	2604	12	0.0046	0.141	0.00065	77	-1676	-0.64
15	06237	COLUMBIA	5268	6.7	0.86	0.84	845	4	0.0047	0.160	0.00076	84	-510	-0.60
16	06752	BRIDGEWATER	1670	2.5	0.93	0.91	267	2	0.0075	0.160	0.00120	58	-152	-0.57
17	06026	EAST GRANBY	5176	1.6	0.77	0.82	364	2	0.0055	0.070	0.00039	80	-204	-0.56
18	06461	MILFORD	15256	3.8	0.76	0.88	7456	41	0.0055	0.489	0.00269	80	-4176	-0.56
19	06831	GREENWICH	14656	3.7	0.85	0.78	8307	47	0.0057	0.567	0.00321	80	-4547	-0.55
20	06078	SUFFIELD	10751	3.9	0.82	0.84	1384	8	0.0058	0.129	0.00074	78	-757	-0.55

Note that ZIP codes 06089 (Weatogue) and 06750 (Bantam) receive the minimum possible value of -1 per call since they had zero responses. Sample size is small enough (less than 200 calls for each) that this may not be a valuable indicator.

Disregarding these two, the other ZIP codes included all have low response rates (none higher than .75%, most below .5%) and low response values (none higher than 82, compared to mean of 95 for dataset). All are at least 75% non-Hispanic white, with poverty rates below 7% and homeownership rates above 75%. These results are as desired: it is not valuable to call ZIP codes that have low response rates and mostly fairly low diversity.

Model N2

Unlike model N1, which uses scaled, percentile data, model N2 uses raw demographic data. The concept of model N2 is to assign a value to each individual response based on demographics, rather than assigning a value to each ZIP code based on its demographic percentiles. Since we only have ZIP code data, though, the model still calculates a ZIP-level response value. Results are similar to those of model N1.

<u>Assumptions:</u>

- The dataset is limited to ZIP codes with a population of at least 1000.
- Every phone call costs DataHaven $1

- The base value of a response in $100
- **The following alterations from this base value are made to account for over/undersampling of certain groups (note that the specific values of these alterations are completely arbitrary. It is easily possible to change the values to reflect any desired benchmarks):**
    - A response from a member of a minority racial/ethnic group (all categories other than non-Hispanic white) is worth **$50 more**
    - A response from a respondent below the poverty line is worth **$50 more**
    - A response from a renter/non-homeowner is worth **$50 more**
    - A response from a respondent over age 65 is worth **$25 less**
- **The demographic breakdown of respondents in a ZIP code is identical to the demographic breakdown of the ZIP code as a whole (note that this assumption is likely incorrect, but currently unavoidable)**

These assumptions yield the following formula for the total value of all responses in a given ZIP code:

$\mathbf{R_{total}}$ = 100**r** - 25**r**(**%Over 65**) + 50**r**(**%Minority** + **PovertyRate** + (1- **HomeownershipRate**)),

Where **r** is the number of responses received in the ZIP code, so the average value **R** of a response in the ZIP code is:

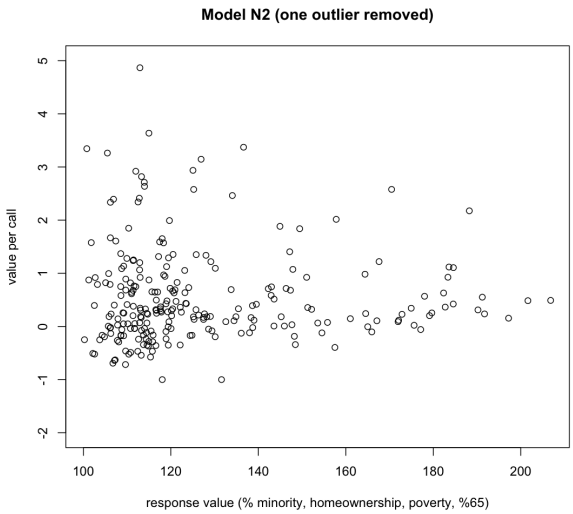
R = 100 - 25(**%Over 65**) + 50(**%Minority** + **PovertyRate** + (1- **HomeownershipRate**))

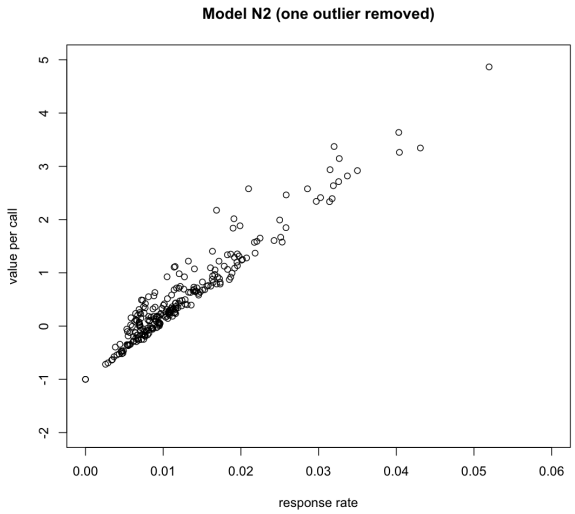
Results:

Using this model, I calculated the following metrics:

- **r_val_n2**: average response value for a ZIP code, calculated by the formula for **R** above
- **benefit_n2**: total benefit of all responses in a ZIP code, calculated as **responses*r_val_n2**
- **value_n2**: benefits - costs in ZIP code, calculated as **benefit_n2 - calls** (recall $1/call cost assumption)
- **val_per_call_n2**: value of each call made in a ZIP code, calculated as **value_n2/calls**. Intended as the primary metric used to assess efficacy of calls in a ZIP code

Overall, results were quite similar to those using model N1. Below are equivalent scatterplots and tables to those produced for model N1.





Top twenty most-valuable-to-call ZIP codes using model N2:

	ZIP	City	TOTAL_POP	POV_RATE	WHITE_NH_PCT	O65_PCT	HOMEOWNERSHIP	calls	responses	r_c_ratio	c_p_ratio	r_p_ratio	r_val_n2	value_n2	val_per_call_n2
1	06269	STORRS MANSFIELD	6723	49.8	0.64	0.0058	0.27	8	2	0.250	0.0012	0.0003	179	351	43.9
2	06256	NORTH WINDHAM	1796	6.6	0.82	0.2166	0.88	154	8	0.052	0.0857	0.0045	113	749	4.9
3	06068	SALISBURY	2034	4.0	0.86	0.3053	0.72	124	5	0.040	0.0610	0.0025	115	451	3.6
4	06018	CANAAN	2630	22.2	0.81	0.1821	0.59	375	12	0.032	0.1426	0.0046	137	1264	3.4
5	06060	NORTH GRANBY	2290	0.0	0.91	0.1467	1.00	116	5	0.043	0.0507	0.0022	101	388	3.3
6	06092	WEST SIMSBURY	4239	2.5	0.88	0.1465	0.96	198	8	0.040	0.0467	0.0019	105	646	3.3
7	06016	BROAD BROOK	6833	7.1	0.71	0.2078	0.72	398	13	0.033	0.0582	0.0019	127	1252	3.1
8	06335	GALES FERRY	6100	14.6	0.75	0.1236	0.83	413	13	0.031	0.0677	0.0021	125	1213	2.9
9	06255	NORTH GROSVENORDALE	4442	6.6	0.94	0.2722	0.75	200	7	0.035	0.0450	0.0016	112	584	2.9
10	06031	FALLS VILLAGE	1385	13.8	0.93	0.2462	0.82	89	3	0.034	0.0643	0.0022	113	251	2.8
11	06612	EASTON	7612	7.7	0.86	0.1699	0.85	491	16	0.033	0.0645	0.0021	114	1332	2.7
12	06357	NIANTIC	11315	3.9	0.83	0.3213	0.77	815	26	0.032	0.0720	0.0023	114	2148	2.6
13	06226	WILLIMANTIC	19068	25.9	0.44	0.1153	0.35	1191	25	0.021	0.0625	0.0013	170	3071	2.6
14	06754	CORNWALL BRIDGE	1849	15.8	0.72	0.1844	0.84	105	3	0.029	0.0568	0.0016	125	271	2.6
15	06807	COS COB	7667	5.6	0.63	0.1347	0.68	542	14	0.026	0.0707	0.0018	134	1335	2.5
16	06798	WOODBURY	9787	5.7	0.87	0.2839	0.79	628	19	0.030	0.0642	0.0019	113	1515	2.4
17	06778	NORTHFIELD	1618	0.1	0.83	0.1397	0.96	126	4	0.032	0.0779	0.0025	107	302	2.4
18	06375	QUAKER HILL	3667	4.1	0.83	0.2173	0.85	202	6	0.030	0.0551	0.0016	112	473	2.3
19	06063	BARKHAMSTED	3248	4.6	0.92	0.2081	0.90	191	6	0.031	0.0588	0.0018	106	446	2.3
20	06510	NEW HAVEN	4205	30.7	0.44	0.0785	0.06	1245	21	0.017	0.2961	0.0050	188	2708	2.2

Top twenty least-valuable-to-call ZIP codes using model N2:

	ZIP	City	TOTAL_POP	POV_RATE	WHITE_NH_PCT	O65_PCT	HOMEOWNERSHIP	calls	responses	r_c_ratio	c_p_ratio	r_p_ratio	r_val_n2	value_n2	val_per_call_n2
1	06089	WEATOGUE	3548	1.9	0.83	0.16	0.48	183	0	0.0000	0.052	0.00000	132	-183	-1.00
2	06750	BANTAM	1324	9.7	0.98	0.24	0.64	117	0	0.0000	0.088	0.00000	118	-117	-1.00
3	06782	PLYMOUTH	2098	6.1	0.85	0.23	0.90	387	1	0.0026	0.184	0.00048	110	-277	-0.72
4	06035	GRANBY	7939	3.8	0.91	0.20	0.90	1037	3	0.0029	0.131	0.00038	107	-717	-0.69
5	06248	HEBRON	5464	2.8	0.93	0.18	0.87	297	1	0.0034	0.054	0.00018	107	-190	-0.64
6	06426	ESSEX	3285	5.5	0.93	0.35	0.81	289	1	0.0035	0.088	0.00030	107	-182	-0.63
7	06812	NEW FAIRFIELD	13558	4.1	0.75	0.17	0.90	2720	10	0.0037	0.201	0.00074	115	-1566	-0.58
8	06235	CHAPLIN	2116	6.9	0.87	0.18	0.84	741	3	0.0040	0.350	0.00142	113	-401	-0.54
9	06716	WOLCOTT	16058	1.5	0.84	0.19	0.87	2766	12	0.0043	0.172	0.00075	110	-1441	-0.52
10	06365	PRESTON	4781	1.6	0.93	0.23	0.92	641	3	0.0047	0.134	0.00063	103	-333	-0.52
11	06763	MORRIS	2078	4.3	0.95	0.26	0.92	415	2	0.0048	0.200	0.00096	102	-211	-0.51
12	06853	NORWALK	3229	3.3	0.90	0.20	0.82	867	4	0.0046	0.269	0.00124	111	-424	-0.49
13	06237	COLUMBIA	5268	6.7	0.86	0.25	0.84	845	4	0.0047	0.160	0.00076	112	-395	-0.47
14	06897	WILTON	18439	2.5	0.75	0.16	0.88	2604	12	0.0046	0.141	0.00065	116	-1215	-0.47
15	06281	WOODSTOCK	7417	4.5	0.93	0.18	0.83	612	3	0.0049	0.083	0.00040	110	-283	-0.46
16	06053	NEW BRITAIN	35697	16.2	0.47	0.14	0.48	12748	49	0.0038	0.357	0.00137	158	-5030	-0.39
17	06461	MILFORD	15256	3.8	0.76	0.19	0.88	7456	41	0.0055	0.489	0.00269	115	-2746	-0.37
18	06026	EAST GRANBY	5176	1.6	0.77	0.20	0.82	364	2	0.0055	0.070	0.00039	117	-131	-0.36
19	06831	GREENWICH	14656	3.7	0.85	0.22	0.78	8307	47	0.0057	0.567	0.00321	115	-2923	-0.35
20	06460	MILFORD	37084	4.7	0.82	0.23	0.73	25587	139	0.0054	0.690	0.00375	119	-8989	-0.35

Results of model N2 are similar to those of model N1, but I think model N2 is arguably better, since its assumptions are closer to the idea Mark presented of weighting individual responses more highly if they come from people with the demographic traits that we want to sample better.

Age and Race by ZIP Analysis (Nat)

These models assign value to individual responses based on the age or race category of the respondent, respectively. They then aggregate the total benefit of responses, value of these responses minus the cost of calls, and resulting value per response and call at the ZIP code level. By looking at specific respondent age and race data within ZIP codes, we are able to account for cases where the age/race breakdown of DataHaven's sample of responses does not correspond to the population-wide demographics of the ZIP code. Thus, these models do away with the problematic assumption that respondent demographics mirror ZIP code demographics. As this analysis will show, this assumption is indeed faulty, since respondents tend to skew older and whiter than the overall population of their ZIP code.

Datasets

Dataset for age analysis

Available here.

Contains all columns in the "Raw Merged Dataset" described above, plus the following:

- **a1834:** Number of respondents in ZCTA aged 18-34 (“18 to 34” column of AgebyZIP2024.xls dataset)
- **a3549**: Number of respondents in ZCTA aged 35-49 (“35 to 49” column of AgebyZIP2024.xls dataset)
- **a5064:** Number of respondents in ZCTA aged 50-64 (“50 to 64” column of AgebyZIP2024.xls dataset)
- **a6569: “**65 to 69” column of AgebyZIP2024.xls dataset. However, based on the high number of responses in this column (about 37% of all responses across Connecticut) and the fact that there is no 70+ column, I believe that this column actually contains the number of respondents in a ZCTA aged 65+.
- **refused:** Number of respondents in ZCTA who refused to give their age (“[DO NOT READ] Refused” column of AgebyZIP2024.xls dataset)
- **benefit, r_val, value, val_per_call:** [explained below under “Age Model”]

Dataset for race analysis

Available here.

Contains all columns in the “Raw Merged Dataset” described above, plus the following:

- **white:** Number of white, non-Hispanic respondents in ZCTA (“Caucasian/White” column of RacebyZIP2024.xls dataset)
- **black**: Number of Black respondents in ZCTA (“African American/Black” column of RacebyZIP2024.xls dataset)
- **latino:** Number of Hispanic/Latino respondents in ZCTA (“Hispanic/Latino” column of RacebyZIP2024.xls dataset)
- **asian:** Number of Asian respondents in ZCTA (**“**Asian” column of RacebyZIP2024.xls dataset)
- **native:** Number of American Indian/Alaskan Native respondents in ZCTA (“American Indian or Alaska Native” column of RacebyZIP2024.xls dataset)
- **NH_PI:** Number of Native Hawaiian and Pacific Islander respondents in ZCTA (“Native Hawaiian or other Pacific Islander” column of RacebyZIP2024.xls dataset)
- **other:** Number of respondents of another race not listed in ZCTA (“[DO NOT READ] Other/Something else (specify)” column of RacebyZIP2024.xls dataset)
- **refused:** Number of respondents in ZCTA who refused to give their race (“[DO NOT READ] Refused” column of RacebyZIP2024.xls dataset)
- **benefit, r_val, value, val_per_call:** [explained below under “Race Model”]

Age Model

Assumptions:

- The dataset is limited to ZIP codes with a population of at least 1000.
- Every phone call costs DataHaven $1

- Following Mark's suggestion in his June 11th Slack message, responses are valued based on respondent age as follows:
    - Age 18-34: $200
    - Age 35-64: $100
    - Age 65+: $50

These assumptions yield the following formula for the total benefit to DataHaven of all responses in a given ZIP code:

benefit = 200***a1834** + 100*(**a3549** + **a5064**) + 50***a6569**,

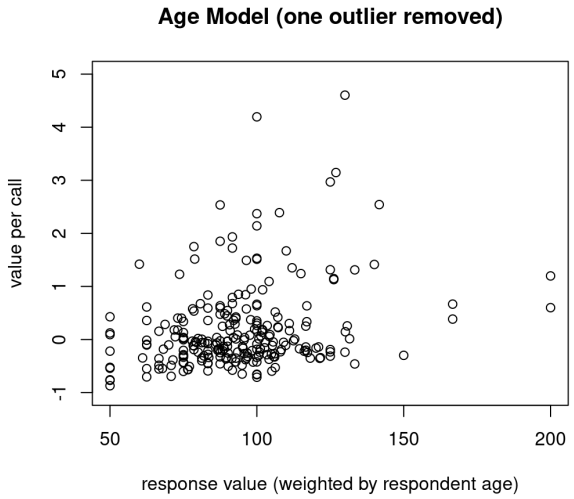
Where **a1834**, **a3549**, **a5064**, and **a6569** are the values of the correspondingly-named columns for the given ZIP code in the age dataset described above.

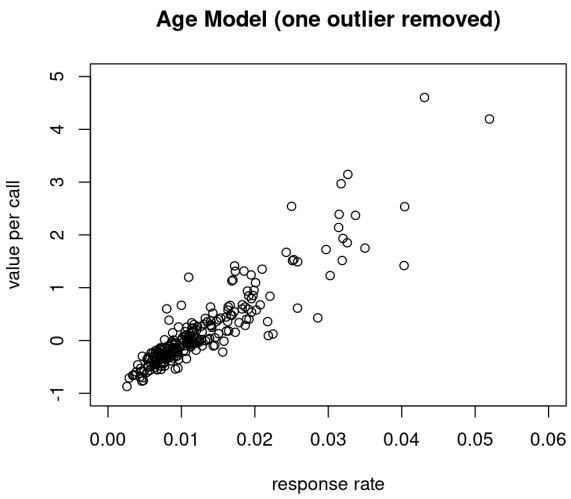
Results:

Using this model, I calculated the following metrics for each ZIP code:

- **benefit:** Total benefit to DataHaven of all responses in the ZIP code, calculated using the formula above
- **r_val:** Average value of a response in the ZIP code, calculated as **benefit / responses**
- **value:** Total value (benefits - costs) of calls to the ZIP code, calculated as **benefit - calls** (recall $1/call assumption)
- **val_per_call:** Average value of each call to the ZIP code, intended as the primary metric to assess effectiveness of calling ZIP code. Calculated as **value / calls**

As with earlier models, response rate ended up being much more predictive of the cost-effectiveness of calling a ZIP code than demographic factors (in this case, age):





However, the most and least valuable-to-call ZIP codes were fairly different from those of earlier models, likely primarily due to the fact that age was only a minor factor in model N2 and was not included at all in model N1. Moreover, the age model does not take into account race, poverty, or homeownership, which were major components of value assignments in earlier models. Below are the 20 most valuable-to-call ZIP codes according to the age model:

	ZIP	a1834	a3549	a5064	a6569	refused	City	calls	responses	r_c_ratio	TOTAL_POP	MED_AGE	O65_PCT	c_p_ratio	benefit	r_val	value	val_per_call
1	06269	1	1	0	0	0	STORRS MANSFIELD	8	2	0.250	6723	20	0.0058	0.0012	300	150	292	36.5
2	06060	2	0	2	1	0	NORTH GRANBY	116	5	0.043	2290	45	0.1467	0.0507	650	130	534	4.6
3	06256	2	0	2	4	0	NORTH WINDHAM	154	8	0.052	1796	48	0.2166	0.0857	800	100	646	4.2
4	06016	4	4	4	1	0	BROAD BROOK	398	13	0.033	6833	44	0.2078	0.0582	1650	127	1252	3.1
5	06778	1	1	2	0	0	NORTHFIELD	126	4	0.032	1618	48	0.1397	0.0779	500	125	374	3.0
6	06417	3	2	0	1	0	DEEP RIVER	240	6	0.025	4432	46	0.2144	0.0542	850	142	610	2.5
7	06092	1	1	2	4	0	WEST SIMSBURY	198	8	0.040	4239	42	0.1465	0.0467	700	88	502	2.5
8	06335	3	5	1	4	0	GALES FERRY	413	13	0.031	6100	37	0.1236	0.0677	1400	108	987	2.4
9	06031	1	0	0	2	0	FALLS VILLAGE	89	3	0.034	1385	55	0.2462	0.0643	300	100	211	2.4
10	06063	1	0	3	2	0	BARKHAMSTED	191	6	0.031	3248	47	0.2081	0.0588	600	100	409	2.1
11	06018	2	2	2	6	0	CANAAN	375	12	0.032	2630	40	0.1821	0.1426	1100	92	725	1.9
12	06612	2	1	5	8	0	EASTON	491	16	0.033	7612	46	0.1699	0.0645	1400	88	909	1.9
13	06255	0	3	1	3	0	NORTH GROSVENORDALE	200	7	0.035	4442	55	0.2722	0.0450	550	79	350	1.8
14	06375	0	1	4	1	0	QUAKER HILL	202	6	0.030	3667	50	0.2173	0.0551	550	92	348	1.7
15	06231	1	2	1	1	0	AMSTON	206	5	0.024	3645	45	0.2294	0.0565	550	110	344	1.7
16	06027	0	1	1	0	0	EAST HARTLAND	79	2	0.025	1376	51	0.2333	0.0574	200	100	121	1.5
17	06357	1	2	10	13	0	NIANTIC	815	26	0.032	11315	56	0.3213	0.0720	2050	79	1235	1.5
18	06023	1	0	2	2	0	EAST BERLIN	199	5	0.025	1464	50	0.2555	0.1359	500	100	301	1.5
19	06807	3	3	2	5	1	COS COB	542	14	0.026	7667	39	0.1347	0.0707	1350	96	808	1.5
20	06068	0	0	1	4	0	SALISBURY	124	5	0.040	2034	41	0.3053	0.0610	300	60	176	1.4

And the 20 least valuable-to-call ZIP codes according to the age model:

	ZIP	a1834	a3549	a5064	a6569	refused	City	calls	responses	r_c_ratio	TOTAL_POP	MED_AGE	O65_PCT	c_p_ratio	benefit	r_val	value	val_per_call
1	06782	0	0	0	1	0	PLYMOUTH	387	1	0.0026	2098	50	0.23	0.184	50	50	-337	-0.87
2	06853	0	0	0	4	0	NORWALK	867	4	0.0046	3229	45	0.20	0.269	200	50	-667	-0.77
3	06763	0	0	0	2	0	MORRIS	415	2	0.0048	2078	49	0.26	0.200	100	50	-315	-0.76
4	06035	1	0	0	2	0	GRANBY	1037	3	0.0029	7939	44	0.20	0.131	300	100	-737	-0.71
5	06237	0	1	0	3	0	COLUMBIA	845	4	0.0047	5268	50	0.25	0.160	250	62	-595	-0.70
6	06716	1	0	2	9	0	WOLCOTT	2766	12	0.0043	16058	44	0.19	0.172	850	71	-1916	-0.69
7	06248	0	0	1	0	0	HEBRON	297	1	0.0034	5464	46	0.18	0.054	100	100	-197	-0.66
8	06426	0	0	1	0	0	ESSEX	289	1	0.0035	3285	59	0.35	0.088	100	100	-189	-0.65
9	06812	1	2	4	3	0	NEW FAIRFIELD	2720	10	0.0037	13558	41	0.17	0.201	950	95	-1770	-0.65
10	06377	0	1	0	1	0	STERLING	375	2	0.0053	2905	42	0.16	0.129	150	75	-225	-0.60
11	06053	11	9	12	17	0	NEW BRITAIN	12748	49	0.0038	35697	34	0.14	0.357	5150	105	-7598	-0.60
12	06281	0	1	1	1	0	WOODSTOCK	612	3	0.0049	7417	46	0.18	0.083	250	83	-362	-0.59
13	06897	2	3	1	6	0	WILTON	2604	12	0.0046	18439	44	0.16	0.141	1100	92	-1504	-0.58
14	06831	2	5	15	24	1	GREENWICH	8307	47	0.0057	14656	48	0.22	0.567	3600	77	-4707	-0.57
15	06906	1	3	4	8	1	STAMFORD	2988	17	0.0057	9472	38	0.15	0.315	1300	76	-1688	-0.56
16	06096	0	1	1	4	0	WINDSOR LOCKS	898	6	0.0067	12555	42	0.17	0.072	400	67	-498	-0.55
17	06784	0	0	1	3	0	SHERMAN	554	4	0.0072	3525	59	0.30	0.157	250	62	-304	-0.55
18	06840	0	3	2	9	0	NEW CANAAN	2091	14	0.0067	20704	42	0.16	0.101	950	68	-1141	-0.55
19	06039	0	0	0	2	0	LAKEVILLE	218	2	0.0092	1955	50	0.23	0.112	100	50	-118	-0.54
20	06514	9	14	13	12	0	HAMDEN	10821	48	0.0044	25291	40	0.15	0.428	5100	106	-5721	-0.53

A large number of the highest and lowest ZIPs on value per call have a very low number of responses, potentially making the age distributions of respondents highly arbitrary. As usual, we see Storrs Mansfield as a major outlier, but there are also less obvious cases like North Granby, where two of the five respondents happened to be 18-34, Northfield, where none of the four respondents were 65+, and Plymouth, where the only respondent was 65+.

To avoid overreacting to such small samples, I also took a look at the 20 most and least valuable-to-call ZIP codes with at least 10 responses.

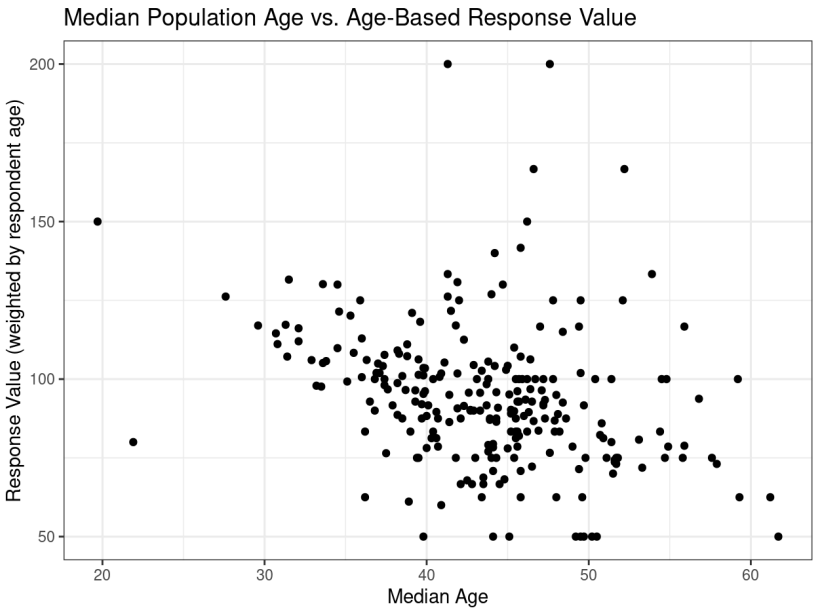
Most valuable ZIPs (at least 10 responses):

	ZIP	a1834	a3549	a5064	a6569	refused	City	calls	responses	r_c_ratio	TOTAL_POP	MED_AGE	O65_PCT	c_p_ratio	benefit	r_val	value	val_per_call
1	06016	4	4	4	1	0	BROAD BROOK	398	13	0.033	6833	44	0.208	0.058	1650	127	1252	3.15
2	06335	3	5	1	4	0	GALES FERRY	413	13	0.031	6100	37	0.124	0.068	1400	108	987	2.39
3	06018	2	2	2	6	0	CANAAN	375	12	0.032	2630	40	0.182	0.143	1100	92	725	1.93
4	06612	2	1	5	8	0	EASTON	491	16	0.033	7612	46	0.170	0.065	1400	88	909	1.85
5	06357	1	2	10	13	0	NIANTIC	815	26	0.032	11315	56	0.321	0.072	2050	79	1235	1.52
6	06807	3	3	2	5	1	COS COB	542	14	0.026	7667	39	0.135	0.071	1350	96	808	1.49
7	06226	6	7	6	6	0	WILLIMANTIC	1191	25	0.021	19068	32	0.115	0.062	2800	112	1609	1.35
8	06259	4	2	1	1	2	POMFRET CENTER	513	10	0.019	4455	48	0.207	0.115	1150	115	637	1.24
9	06798	2	1	2	14	0	WOODBURY	628	19	0.030	9787	52	0.284	0.064	1400	74	772	1.23
10	06480	8	4	4	5	0	PORTLAND	1234	21	0.017	9421	41	0.181	0.131	2650	126	1416	1.15
11	06510	8	7	1	5	0	NEW HAVEN	1245	21	0.017	4205	28	0.078	0.296	2650	126	1405	1.13
12	06482	3	1	3	5	0	SANDY HOOK	597	12	0.020	11808	44	0.175	0.051	1250	104	653	1.09
13	06042	5	5	6	9	1	MANCHESTER	1307	26	0.020	23302	37	0.157	0.056	2550	98	1243	0.95
14	06040	11	9	11	20	0	MANCHESTER	2685	51	0.019	36171	37	0.151	0.074	5200	102	2515	0.94
15	06870	2	3	3	5	0	OLD GREENWICH	678	13	0.019	7903	40	0.142	0.086	1250	96	572	0.84
16	06759	0	1	7	5	0	LITCHFIELD	627	13	0.021	5436	53	0.316	0.115	1050	81	423	0.67
17	06374	2	0	5	4	0	PLAINFIELD	661	11	0.017	7884	43	0.192	0.084	1100	100	439	0.66
18	06779	1	0	8	2	0	OAKVILLE	672	11	0.016	7762	46	0.204	0.087	1100	100	428	0.64
19	06776	15	5	14	12	1	NEW MILFORD	3365	47	0.014	27133	42	0.161	0.124	5500	117	2135	0.63
20	06067	3	3	3	11	0	ROCKY HILL	1095	20	0.018	20716	44	0.215	0.053	1750	88	655	0.60

Least valuable ZIPs (at least 10 responses):

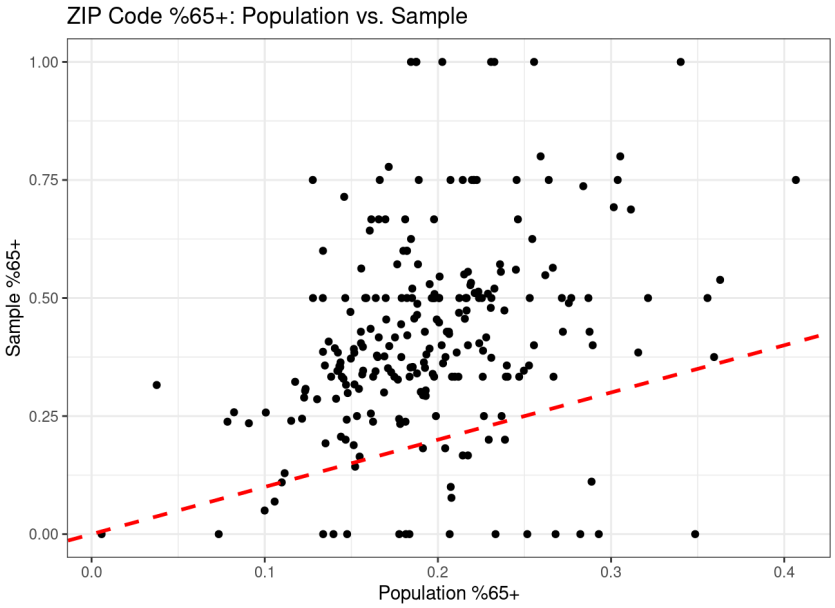
	ZIP	a1834	a3549	a5064	a6569	refused	City	calls	responses	r_c_ratio	TOTAL_POP	MED_AGE	O65_PCT	c_p_ratio	benefit	r_val	value	val_per_call
1	06716	1	0	2	9	0	WOLCOTT	2766	12	0.0043	16058	44	0.19	0.172	850	71	-1916	-0.69
2	06812	1	2	4	3	0	NEW FAIRFIELD	2720	10	0.0037	13558	41	0.17	0.201	950	95	-1770	-0.65
3	06053	11	9	12	17	0	NEW BRITAIN	12748	49	0.0038	35697	34	0.14	0.357	5150	105	-7598	-0.60
4	06897	2	3	1	6	0	WILTON	2604	12	0.0046	18439	44	0.16	0.141	1100	92	-1504	-0.58
5	06831	2	5	15	24	1	GREENWICH	8307	47	0.0057	14656	48	0.22	0.567	3600	77	-4707	-0.57
6	06906	1	3	4	8	1	STAMFORD	2988	17	0.0057	9472	38	0.15	0.315	1300	76	-1688	-0.56
7	06840	0	3	2	9	0	NEW CANAAN	2091	14	0.0067	20704	42	0.16	0.101	950	68	-1141	-0.55
8	06514	9	14	13	12	0	HAMDEN	10821	48	0.0044	25291	40	0.15	0.428	5100	106	-5721	-0.53
9	06517	1	7	10	19	0	HAMDEN	5794	37	0.0064	14867	44	0.22	0.390	2850	77	-2944	-0.51
10	06461	7	5	8	20	1	MILFORD	7456	41	0.0055	15256	45	0.19	0.489	3700	90	-3756	-0.50
11	06804	1	1	1	9	0	BROOKFIELD	1659	12	0.0072	17871	46	0.21	0.093	850	71	-809	-0.49
12	06460	20	22	43	54	0	MILFORD	25587	139	0.0054	37084	48	0.23	0.690	13200	95	-12387	-0.48
13	06033	0	3	3	6	0	GLASTONBURY	1577	12	0.0076	29311	44	0.20	0.054	900	75	-677	-0.43
14	06604	13	18	21	26	1	BRIDGEPORT	13573	79	0.0058	29497	38	0.15	0.460	7800	99	-5773	-0.43
15	06795	4	6	7	10	1	WATERTOWN	4488	28	0.0062	14409	46	0.24	0.311	2600	93	-1888	-0.42
16	06001	2	4	2	5	0	AVON	2015	13	0.0065	19157	46	0.21	0.105	1250	96	-765	-0.38
17	06902	48	55	64	72	2	STAMFORD	39979	241	0.0060	72915	37	0.15	0.548	25100	104	-14879	-0.37
18	06811	10	4	15	33	0	DANBURY	8744	62	0.0071	30526	46	0.22	0.286	5550	90	-3194	-0.37
19	06851	4	8	19	16	0	NORWALK	6693	47	0.0070	28591	42	0.19	0.234	4300	91	-2393	-0.36
20	06051	20	13	17	13	0	NEW BRITAIN	11848	63	0.0053	29421	35	0.14	0.403	7650	121	-4198	-0.35

As expected, the high-value ZIP codes tend to have high response rates and more responses from younger age groups, whereas low-value ZIP codes are the opposite. However, this trend is not universal. There are several high-value ZIP codes with fairly old respondents: for instance, 14 out of the 19 respondents in Woodbury were in the 65+ age group, but the ZCTA's high response rate of about 3% outweighed this age distribution. Interestingly, the median ages in the least valuable ZIPs do not appear noticeably higher than those in the most valuable ZIPs. The graph below shows that the correlation between a ZIP code's median age and its response value (representing how young its respondents were) is modest at best. Several of the least valuable ZIPs offer examples of how the age breakdown of respondents can skew much older than the age breakdown of the ZIP code. For instance, New Britain 06053 has a young median age of 34 and only 14% of its population is over 65, but the median respondent was in the 50-64 age range, and 17/49 (34.7%) of respondents were 65 and up. Similarly, in Greenwich 06831, the median age is 48 and 22% of the population is over 65, but 24/47 (51%) of respondents were 65+, putting the median respondent in the 65+ category. Stamford 06906 is another dramatic example, with a median age of 38 and 15% over 65 population, but 8/17 (47%) of respondents 65 and up.



Differences from ZIP-level data:

While some of the least valuable-to-call ZIP codes provide particularly stark examples, the trend of DataHaven's sample skewing older than the general population is fairly consistent across the dataset. Overall, **in 89% of ZIP codes, the percentage of survey respondents over 65 was higher than the percentage of the general population over 65**. This trend is visible in the scatterplot below, where the red dashed line demarcates equal percentages in the survey sample and in the general population:



Race Model

Assumptions:

- The dataset is limited to ZIP codes with a population of at least 1000.
- Every phone call costs DataHaven $1
- Following Mark's suggestion in his June 11th Slack message, responses are valued based on respondent race as follows:
    - Latino: $150
    - Black: $125
    - All other: $100

These assumptions yield the following formula for the total benefit to DataHaven of all responses in a given ZIP code:

benefit = 150*__latino__ + 125*__black__ + 100*(__responses__ - __latino__ - __black__),

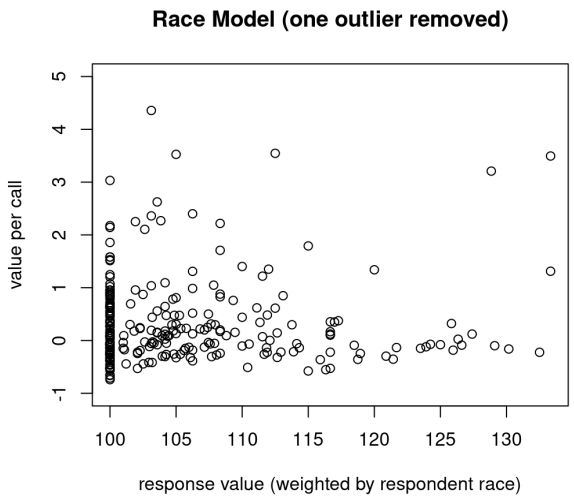
Where **latino**, **black**, and **responses** are the values of the correspondingly-named columns for the given ZIP code in the race dataset described above.

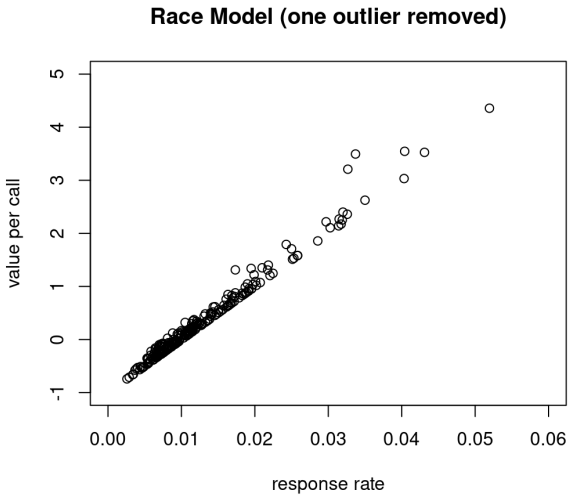
Results:

Using this model, I calculated the following metrics for each ZIP code:

- **benefit:** Total benefit to DataHaven of all responses in the ZIP code, calculated using the formula above
- **r_val:** Average value of a response in the ZIP code, calculated as **benefit / responses**
- **value:** Total value (benefits - costs) of calls to the ZIP code, calculated as **benefit - calls** (recall $1/call assumption)
- **val_per_call:** Average value of each call to the ZIP code, intended as the primary metric to assess effectiveness of calling ZIP code. Calculated as **value / calls**

As with earlier models, response rate ended up being much more predictive of the cost-effectiveness of calling a ZIP code than demographic factors (in this case, race). Here, the effect was especially dramatic, as the scatterplot of response rate vs. value per call looks nearly like a straight line, while that of response value (which measures the racial breakdown of the sample) vs. value per call does not appear to show any correlation at all. This is perhaps due to this model's relatively low range in value based on demographic characteristics: model N2 had a theoretical range of $175 between the most and least valuable respondents, and the age model had a range of $150, whereas this model's range is only $50. Note that it is very easy to achieve more variation based on demographic breakdown by simply adjusting the weights used to calculate **benefit** in this model's R script (DCWS_CR_AgeAndRaceByZip.R).





Since value per call is so highly correlated with response rate, it is no surprise that the most and least valuable ZIP codes are similar to those identified by models N1 and N2
Below are the 20 most valuable-to-call ZIP codes according to the race model:

	ZIP	white	black	latino	asian	City	calls	responses	r_c_ratio	TOTAL_POP	HISP_PCT	BLACK_PCT	WHITE_NH_PCT	c_p_ratio	benefit	r_val	value	val_per_call
1	06269	2	0	0	0	STORRS MANSFIELD	8	2	0.250	6723	0.0678	0.10992	0.64	0.0012	200	100	192	24.0
2	06256	7	1	0	0	NORTH WINDHAM	154	8	0.052	1796	0.1476	0.01281	0.82	0.0857	825	103	671	4.4
3	06092	5	0	2	0	WEST SIMSBURY	198	8	0.040	4239	0.0238	0.00000	0.88	0.0467	900	112	702	3.5
4	06060	4	1	0	0	NORTH GRANBY	116	5	0.043	2290	0.0048	0.00437	0.91	0.0507	525	105	409	3.5
5	06031	0	0	2	0	FALLS VILLAGE	89	3	0.034	1385	0.0339	0.01011	0.93	0.0643	400	133	311	3.5
6	06016	5	1	7	0	BROAD BROOK	398	13	0.033	6833	0.0853	0.07478	0.71	0.0582	1675	129	1277	3.2
7	06068	5	0	0	0	SALISBURY	124	5	0.040	2034	0.0398	0.00836	0.86	0.0610	500	100	376	3.0
8	06255	6	1	0	0	NORTH GROSVENORDALE	200	7	0.035	4442	0.0173	0.00270	0.94	0.0450	725	104	525	2.6
9	06018	9	3	0	0	CANAAN	375	12	0.032	2630	0.1361	0.00000	0.81	0.1426	1275	106	900	2.4
10	06612	15	0	1	0	EASTON	491	16	0.033	7612	0.0339	0.01524	0.86	0.0645	1650	103	1159	2.4
11	06335	11	0	1	1	GALES FERRY	413	13	0.031	6100	0.0810	0.04410	0.75	0.0677	1350	104	937	2.3
12	06357	25	0	1	0	NIANTIC	815	26	0.032	11315	0.0376	0.02749	0.83	0.0720	2650	102	1835	2.3
13	06375	5	0	1	0	QUAKER HILL	202	6	0.030	3667	0.0624	0.03245	0.83	0.0551	650	108	448	2.2
14	06778	3	0	0	1	NORTHFIELD	126	4	0.032	1618	0.0315	0.00000	0.83	0.0779	400	100	274	2.2
15	06063	6	0	0	0	BARKHAMSTED	191	6	0.031	3248	0.0702	0.00031	0.92	0.0588	600	100	409	2.1
16	06798	18	0	1	0	WOODBURY	628	19	0.030	9787	0.0385	0.02268	0.87	0.0642	1950	103	1322	2.1
17	06754	3	0	0	0	CORNWALL BRIDGE	105	3	0.029	1849	0.1201	0.01947	0.72	0.0568	300	100	195	1.9
18	06231	3	1	1	0	AMSTON	206	5	0.024	3645	0.0672	0.00439	0.87	0.0565	575	115	369	1.8
19	06417	4	0	1	1	DEEP RIVER	240	6	0.025	4432	0.0449	0.02775	0.87	0.0542	650	108	410	1.7
20	06807	12	0	0	2	COS COB	542	14	0.026	7667	0.1316	0.03926	0.63	0.0707	1400	100	858	1.6

And the 20 least valuable-to-call ZIP codes according to the race model:

	ZIP	white	black	latino	asian	City	calls	responses	r_c_ratio	TOTAL_POP	HISP_PCT	BLACK_PCT	WHITE_NH_PCT	c_p_ratio	benefit	r_val	value	val_per_call
1	06782	1	0	0	0	PLYMOUTH	387	1	0.0026	2098	0.053	0.0562	0.85	0.184	100	100	-287	-0.74
2	06035	3	0	0	0	GRANBY	1037	3	0.0029	7939	0.033	0.0076	0.91	0.131	300	100	-737	-0.71
3	06248	1	0	0	0	HEBRON	297	1	0.0034	5464	0.031	0.0113	0.93	0.054	100	100	-197	-0.66
4	06426	1	0	0	0	ESSEX	289	1	0.0035	3285	0.024	0.0070	0.93	0.088	100	100	-189	-0.65
5	06812	7	0	3	0	NEW FAIRFIELD	2720	10	0.0037	13558	0.133	0.0396	0.75	0.201	1150	115	-1570	-0.58
6	06716	12	0	0	0	WOLCOTT	2766	12	0.0043	16058	0.035	0.0425	0.84	0.172	1200	100	-1566	-0.57
7	06053	26	8	12	0	NEW BRITAIN	12748	49	0.0038	35697	0.363	0.1305	0.47	0.357	5700	116	-7048	-0.55
8	06853	4	0	0	0	NORWALK	867	4	0.0046	3229	0.024	0.0158	0.90	0.269	400	100	-467	-0.54
9	06365	3	0	0	0	PRESTON	641	3	0.0047	4781	0.032	0.0109	0.93	0.134	300	100	-341	-0.53
10	06897	11	1	0	0	WILTON	2604	12	0.0046	18439	0.063	0.0394	0.75	0.141	1225	102	-1379	-0.53
11	06235	2	0	1	0	CHAPLIN	741	3	0.0040	2116	0.045	0.0061	0.87	0.350	350	117	-391	-0.53
12	06237	4	0	0	0	COLUMBIA	845	4	0.0047	5268	0.055	0.0216	0.86	0.160	400	100	-445	-0.53
13	06763	2	0	0	0	MORRIS	415	2	0.0048	2078	0.014	0.0149	0.95	0.200	200	100	-215	-0.52
14	06514	32	10	5	1	HAMDEN	10821	48	0.0044	25291	0.162	0.3476	0.43	0.428	5300	110	-5521	-0.51
15	06281	3	0	0	0	WOODSTOCK	612	3	0.0049	7417	0.048	0.0018	0.93	0.083	300	100	-312	-0.51
16	06377	2	0	0	0	STERLING	375	2	0.0053	2905	0.072	0.0065	0.83	0.129	200	100	-175	-0.47
17	06026	2	0	0	0	EAST GRANBY	364	2	0.0055	5176	0.061	0.0410	0.77	0.070	200	100	-164	-0.45
18	06461	38	0	1	1	MILFORD	7456	41	0.0055	15256	0.112	0.0151	0.76	0.489	4150	101	-3306	-0.44
19	06460	126	4	5	2	MILFORD	25587	139	0.0054	37084	0.080	0.0325	0.82	0.690	14250	103	-11337	-0.44
20	06831	43	0	3	1	GREENWICH	8307	47	0.0057	14656	0.071	0.0114	0.85	0.567	4850	103	-3457	-0.42

As with the age model, there are many high and low value ZIP codes with very low numbers of responses, so I again take a look at the results from ZIP codes with at least 10 responses.

Most valuable ZIPs (at least 10 responses):

	ZIP	white	black	latino	asian	City	calls	responses	r_c_ratio	TOTAL_POP	HISP_PCT	BLACK_PCT	WHITE_NH_PCT	c_p_ratio	benefit	r_val	value	val_per_call
1	06016	5	1	7	0	BROAD BROOK	398	13	0.033	6833	0.085	0.0748	0.71	0.058	1675	129	1277	3.21
2	06018	9	3	0	0	CANAAN	375	12	0.032	2630	0.136	0.0000	0.81	0.143	1275	106	900	2.40
3	06612	15	0	1	0	EASTON	491	16	0.033	7612	0.034	0.0152	0.86	0.065	1650	103	1159	2.36
4	06335	11	0	1	1	GALES FERRY	413	13	0.031	6100	0.081	0.0441	0.75	0.068	1350	104	937	2.27
5	06357	25	0	1	0	NIANTIC	815	26	0.032	11315	0.038	0.0275	0.83	0.072	2650	102	1835	2.25
6	06798	18	0	1	0	WOODBURY	628	19	0.030	9787	0.039	0.0227	0.87	0.064	1950	103	1322	2.11
7	06807	12	0	0	2	COS COB	542	14	0.026	7667	0.132	0.0393	0.63	0.071	1400	100	858	1.58
8	06226	15	4	4	2	WILLIMANTIC	1191	25	0.021	19068	0.450	0.0626	0.44	0.062	2800	112	1609	1.35
9	06259	5	0	4	1	POMFRET CENTER	513	10	0.019	4455	0.037	0.0029	0.89	0.115	1200	120	687	1.34
10	06042	14	4	4	2	MANCHESTER	1307	26	0.020	23302	0.129	0.1412	0.56	0.056	2900	112	1593	1.22
11	06482	11	0	1	0	SANDY HOOK	597	12	0.020	11808	0.081	0.0232	0.82	0.051	1250	104	653	1.09
12	06759	12	0	0	0	LITCHFIELD	627	13	0.021	5436	0.047	0.0151	0.84	0.115	1300	100	673	1.07
13	06040	32	8	4	3	MANCHESTER	2685	51	0.019	36171	0.179	0.1734	0.53	0.074	5500	108	2815	1.05
14	06084	16	0	0	0	TOLLAND	792	16	0.020	14685	0.076	0.0048	0.84	0.054	1600	100	808	1.02
15	06488	24	0	1	0	SOUTHBURY	1354	26	0.019	19973	0.088	0.0186	0.83	0.068	2650	102	1296	0.96
16	06870	11	0	0	1	OLD GREENWICH	678	13	0.019	7903	0.078	0.0224	0.75	0.086	1300	100	622	0.92
17	06371	38	0	0	0	OLD LYME	2068	39	0.019	9855	0.037	0.0031	0.89	0.210	3900	100	1832	0.89
18	06067	17	0	1	2	ROCKY HILL	1095	20	0.018	20716	0.078	0.0480	0.69	0.053	2050	102	955	0.87
19	06002	18	18	2	1	BLOOMFIELD	2570	42	0.016	21547	0.076	0.5508	0.30	0.119	4750	113	2180	0.85
20	06510	14	3	2	1	NEW HAVEN	1245	21	0.017	4205	0.129	0.2071	0.44	0.296	2275	108	1030	0.83

Least valuable ZIPs (at least 10 responses):

	ZIP	white	black	latino	asian	City	calls	responses	r_c_ratio	TOTAL_POP	HISP_PCT	BLACK_PCT	WHITE_NH_PCT	c_p_ratio	benefit	r_val	value	val_per_call
1	06812	7	0	3	0	NEW FAIRFIELD	2720	10	0.0037	13558	0.133	0.0396	0.75	0.20	1150	115	-1570	-0.58
2	06716	12	0	0	0	WOLCOTT	2766	12	0.0043	16058	0.035	0.0425	0.84	0.17	1200	100	-1566	-0.57
3	06053	26	8	12	0	NEW BRITAIN	12748	49	0.0038	35697	0.363	0.1305	0.47	0.36	5700	116	-7048	-0.55
4	06897	11	1	0	0	WILTON	2604	12	0.0046	18439	0.063	0.0394	0.75	0.14	1225	102	-1379	-0.53
5	06514	32	10	5	1	HAMDEN	10821	48	0.0044	25291	0.162	0.3476	0.43	0.43	5300	110	-5521	-0.51
6	06461	38	0	1	1	MILFORD	7456	41	0.0055	15256	0.112	0.0151	0.76	0.49	4150	101	-3306	-0.44
7	06460	126	4	5	2	MILFORD	25587	139	0.0054	37084	0.080	0.0325	0.82	0.69	14250	103	-11337	-0.44
8	06831	43	0	3	1	GREENWICH	8307	47	0.0057	14656	0.071	0.0114	0.85	0.57	4850	103	-3457	-0.42
9	06906	13	2	0	1	STAMFORD	2988	17	0.0057	9472	0.407	0.0969	0.43	0.32	1750	103	-1238	-0.41
10	06795	28	0	0	0	WATERTOWN	4488	28	0.0062	14409	0.100	0.0123	0.80	0.31	2800	100	-1688	-0.38
11	06110	6	1	3	0	WEST HARTFORD	2002	11	0.0055	12307	0.257	0.1422	0.45	0.16	1275	116	-727	-0.36
12	06705	24	19	13	0	WATERBURY	11092	60	0.0054	26352	0.341	0.2396	0.34	0.42	7125	119	-3967	-0.36
13	06051	31	8	23	0	NEW BRITAIN	11848	63	0.0053	29421	0.518	0.1379	0.30	0.40	7650	121	-4198	-0.35
14	06840	13	0	0	1	NEW CANAAN	2091	14	0.0067	20704	0.047	0.0175	0.83	0.10	1400	100	-691	-0.33
15	06518	24	3	2	0	HAMDEN	5454	35	0.0064	19958	0.070	0.1555	0.68	0.27	3675	105	-1779	-0.33
16	06517	26	7	1	1	HAMDEN	5794	37	0.0064	14867	0.167	0.2208	0.58	0.39	3925	106	-1869	-0.32
17	06902	136	42	40	10	STAMFORD	39979	241	0.0060	72915	0.353	0.2421	0.32	0.55	27150	113	-12829	-0.32
18	06001	11	0	2	0	AVON	2015	13	0.0065	19157	0.047	0.0231	0.73	0.11	1400	108	-615	-0.31
19	06824	89	2	7	0	FAIRFIELD	14966	101	0.0067	36868	0.047	0.0037	0.88	0.41	10500	104	-4466	-0.30
20	06604	30	20	23	1	BRIDGEPORT	13573	79	0.0058	29497	0.451	0.1687	0.30	0.46	9550	121	-4023	-0.30

As with the age data, it is interesting to note the differences between ZIP code and sample demographics. Broad Brook is particularly notable: the ZIP code as a whole is 71% white non-Hispanic and only 8.5% Hispanic/Latino, but 7 of 13 survey respondents were Hispanic/Latino. In Caanan, the census records 0% of residents as Black, but 3 of 12 respondents were Black. In general, though, across high and low value ZIP codes, the opposite pattern - respondents skewing whiter than the population - is much more common. For example, among high value ZIP codes, Cos Cob is 63% white, but 12 of 14 respondents were white. New Haven 06510 is 44% white, but 14 of 21 respondents were white. Among low value ZIP codes, Hamden 06514 is 43% white, but 32 of 48 respondents were white. Milford 06461 is 76% white, but 38/41 respondents were white. Stamford 06902 is 32% white, but 146/241 respondents were white.

<u>Differences from ZIP-level data:</u>

As with the age model, the trends identified above in the particularly high and low value ZIP codes are fairly consistent throughout the dataset. Overall:

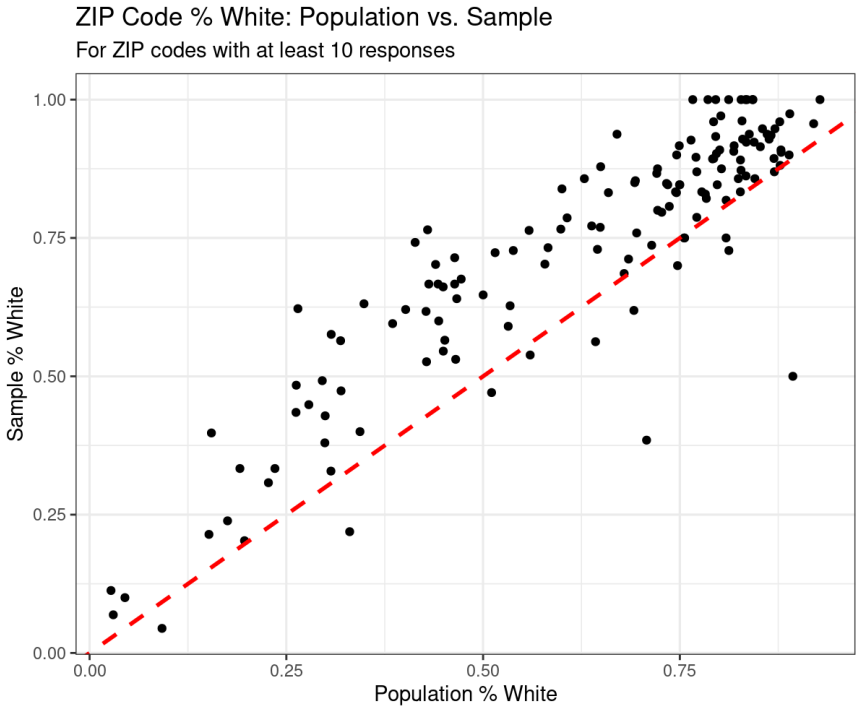
- **In 78% of ZIP codes, the Hispanic/Latino percentage of the general population was higher than the Hispanic/Latino percentage of survey respondents**
- **In 72% of ZIP codes, the Black percentage of the general population was higher than the Black percentage of survey respondents**
- **In 79% of ZIP codes, the white percentage of the general population was *lower* than the white percentage of survey respondents**

The trends in sampling of the Hispanic population and especially of the white population were even more dramatic in ZIP codes with more than single-digit responses:

- **In 80% of ZIP codes with at least 10 responses, the Hispanic/Latino percentage of the general population was higher than the Hispanic/Latino percentage of survey respondents**

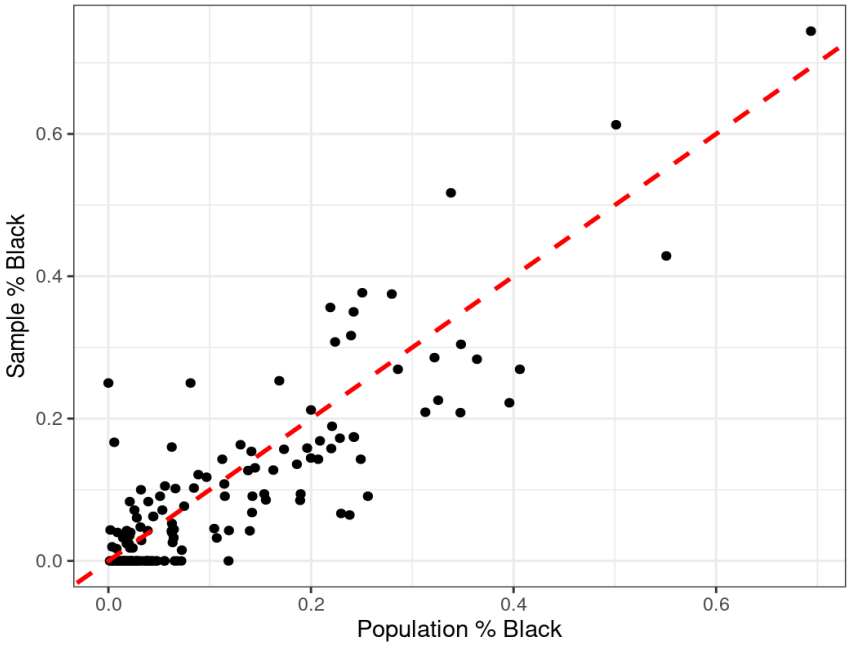
- **In 69% of ZIP codes with at least 10 responses, the Black percentage of the general population was higher than the Black percentage of survey respondents**
- **In 90.4% of ZIP codes with at least 10 responses, the white percentage of the general population was *lower* than the white percentage of survey respondents**

The relationship between survey sample and general population for these metrics in ZIP codes with at least 10 responses is evident in the scatter plots below. In each plot, the red dashed line marks equality between sample and population percentages.



ZIP Code % Black: Population vs. Sample

For ZIP codes with at least 10 responses



ZIP Code % Latino: Population vs. Sample

For ZIP codes with at least 10 responses