STATE OF MAINE
DEPARTMENT OF ADMINISTRATIVE & FINANCIAL SERVICES
BURTON M. CROSS BUILDING, 3RD FLOOR
78 STATE HOUSE STATION
AUGUSTA, MAINE 04333-0078

SERVING THE PUBLIC AND DELIVERING ESSENTIAL SERVICES TO STATE GOVERNMENT

JANET T. MILLS
GOVERNOR

KIRSTEN LC FIGUEROA
COMMISSIONER

Dr. Steven Dillingham
Director of the United States Census Bureau
4600 Silver Hill Road
Washington, DC 20233

February 20, 2020

Dear Dr. Dillingham,

The Office of the State Economist, within the Department of Administrative and Financial Services, serves as the State Data Center lead for the State of Maine. In this capacity, we are writing to express our concerns regarding the proposed policy changes involving the use of differential privacy in census data. Privacy protections for individuals are of utmost importance to the State of Maine. We recognize that caution and careful planning for disclosure avoidance are necessary in order to maintain the integrity of the decennial census and all Census products. However, upon careful review of the 2010 demonstration data product released by the U.S. Census Bureau, we are hereby voicing concern for the usability, reliability, and equity of differentially private (DP) Census data.

Our analyses show that small, rural places suffer the most in terms of inaccurate estimates. In Maine's case, that means a majority of our counties and sub-county geographies are subject to unacceptably high levels of error. If this holds true in the release of the 2020 decennial census data and other future data products, the repercussions for our state and nation are considerable.

Decennial census data are used for the apportionment of state legislative districts. They serve as the benchmark for population estimates, demographic projections, surveys, research, and analysis carried out by everyone from local housing planners to the U.S. Census Bureau itself. Over three hundred federal spending programs distribute funds on the basis of data derived from the decennial census. Policy decisions at all levels of government use data that originate with the decennial census. In many cases policymakers, researchers, businesspeople, and the public rely on data that is only available from the U.S. Census Bureau. If the reliability of that data falls by the wayside or the data becomes so difficult to interpret that general users are unable to decipher it, we run the risk of basing decisions on no data at all or, perhaps worse, on inaccurate data.

The U.S. Census Bureau has long been the standard-bearer in terms of providing high quality, reliable data to the public. This proposed policy change would threaten that position and throw into doubt any redistricting, funding decisions, or analysis done using census data.

While we have been able to assess the errors in the demonstration product, this will not be possible for the 2020 published data. At the time of writing, there is no established guidance with respect to how statistical analysis should be carried out in light of the proposed change. Even if these tools existed, we fear many of the data users within our state do not have the resources and training necessary to account for these errors. This exacerbates our concern that DP has the potential to exclude rural and resource-strained communities from equitable access to high-quality, reliable data, and that our narratives will be systematically misinformed as a result.

In light of our grave misgivings concerning this proposed policy change, we have several requests that would help to either reduce the negative impacts from the change or provide additional information to help us prepare for the impacts.

1. We request that the U.S. Census Bureau release more demonstration datasets for different epsilon values, geographical hierarchies, and queries, as well as multiple iterations of each.

2. We request that the U.S. Census Bureau use a higher value of epsilon, and particularly higher allocation for Age and Sex tabulations.

3. We request that the U.S. Census Bureau release raw noise-injected counts.

4. We request better information and analysis from the U.S. Census Bureau regarding the impacts on related data products including the American Community Survey, Current Population Survey, and Population Estimates Program.

5. We request that the U.S. Census Bureau report margins of error or confidence intervals for previously released DP data and any newly-released DP data.

Despite the availability of the demonstration data product, data users have not been given enough time to conduct thorough analysis to understand these impacts, since several tables were either not included or are not comparable to the demonstration data. For example, the U.S. Census Bureau has cautioned that table P20 is not comparable to the demonstration product. There has been inadequate opportunity to evaluate the privacy-accuracy tradeoff since there has been only one single demonstration data set to analyze at one single epsilon value, geographical hierarchy, and query. More demonstration datasets would allow users to understand these three important aspects of the privacy algorithm.

Additionally, there has been inadequate communication regarding impacts to other valuable data products such as the American Community Survey, the Current Population Survey, or the Population Estimates Program[1]. Other economic data released by the U.S. Bureau of Economic Analysis, U.S. Bureau of Labor Statistics, and a vast spectrum of other data agencies will similarly face challenges with survey design.

[1] The Census Bureau's analysis of the Population Estimates Program shows Maine (statewide) has the second-highest Mean Absolute Percentage Error (MAPE) among all states in these estimates: 42.5% MAPE using the demonstration products as a benchmark compared to 12.8% with published Census data. These estimates are a primary data input for Maine's population projections. Still, the data for this calculation has not been released to the public, which has left us mostly unaware of these impacts.

Inaccuracy in the decennial census will flow through ten full years of data via these crucial products. The current implementation of DP creates a group of regions and people, predominantly rural and already marginalized, that are left behind; they will continue to be left behind for the remainder of the decade unless action is taken to improve the algorithm. Without resolution to the above uncertainties it will be impossible to measure the magnitude of these errors, resulting in further challenges for these places and communities.

Following is a description of the analysis performed by our office and the results that prompted our concerns. We appreciate your consideration of our requests and look forward to a prompt reply.

Sincerely,

Angela Hallowell
Maine State Data Center lead

Amanda Rector
Maine State Economist

Impacts in Maine

The demonstration data product was accessed courtesy of IPUMS NHGIS, University of Minnesota, www.nhgis.org. We find that most counts are reliable at the state level, as are total population counts at the county level. However, detailed counts for nearly all sub-state geographies have been compromised by noise injection.

County-level counts

One example of this lies in age and sex counts at the county level (**Figure 1**). The greatest Mean Absolute Percentage Error (MAPE) is found for 18-19 years, 20 years, 21 years, and 85 years and over cohorts for both male and female. Even when aggregated by sex, MAPE is over 10% in all abovementioned cohorts except 18 and 19 years **(Figure 2**). This data has a major part to play in the analysis carried out by numerous state agencies. For example, the ongoing opioid crisis throughout the state disproportionately affects young men in rural counties. Inaccuracies of this magnitude in population counts could lead to under- or over-calculations of overdose rates and would make it difficult to statistically detect changes across time and space. This makes the management of this public health crisis a nearly impossible task. Additionally, Maine has the oldest median age and the highest percent of the population age 65 and older of any state in the U.S. The high level of inaccuracy with the 85 and over cohorts will make planning for our rapidly aging population increasingly complex.

Similarly, **Figure 3** demonstrates the inaccuracy in counts for households by age of householder. Again, the youngest category (householder aged 15-24) and the oldest categories (75-84 and 85 years and over) have the highest errors. This translates to errors that halve or double these populations in some of Maine's smallest counties (**Table 1**).

Race of householder in occupied units is also significantly flawed (**Figure 4**). All racial categories except *White alone* have MAPE over 25%. In fact, only two have MAPE under 100% (Two or more races and American Indian and Alaska Native). In Franklin County, the count of households with a black or African American householder was more than 11 times its published count (**Table 2**). Any changes in Maine's diversity at a county level will be incredibly difficult to statistically detect and will undoubtedly lead to misinformed narratives about demographic comparisons over time and space. These examples are just some of the many large errors we found in the data at the county level in Maine.

County Subdivision and School District Counts

Data users will find county subdivision counts almost entirely useless given the current privacy loss budget level and allocation. Total population counts are relatively acceptable for large county subdivisions. Error is large for the smallest subdivisions, but meaningfully falls below 10% absolute percent error at about 900 people. However, this leaves about 236 of 533 Maine county subdivisions vulnerable to large miscounts. This is demonstrated in **Figures 5 and 6**.

Age and sex counts are severely affected by noise injection. **Figures 7 and 8** show the MAPE by age and sex cohort and counties, respectively. No category (other than total) has a MAPE under 50%, and many have MAPE well over 100% for both sexes. Similarly, half of the counties have MAPE across category and geographies above 100%; the lowest is in York at 49.8%. These errors are altogether unacceptable and if left unchanged, we will caution users against relying on any of these data.

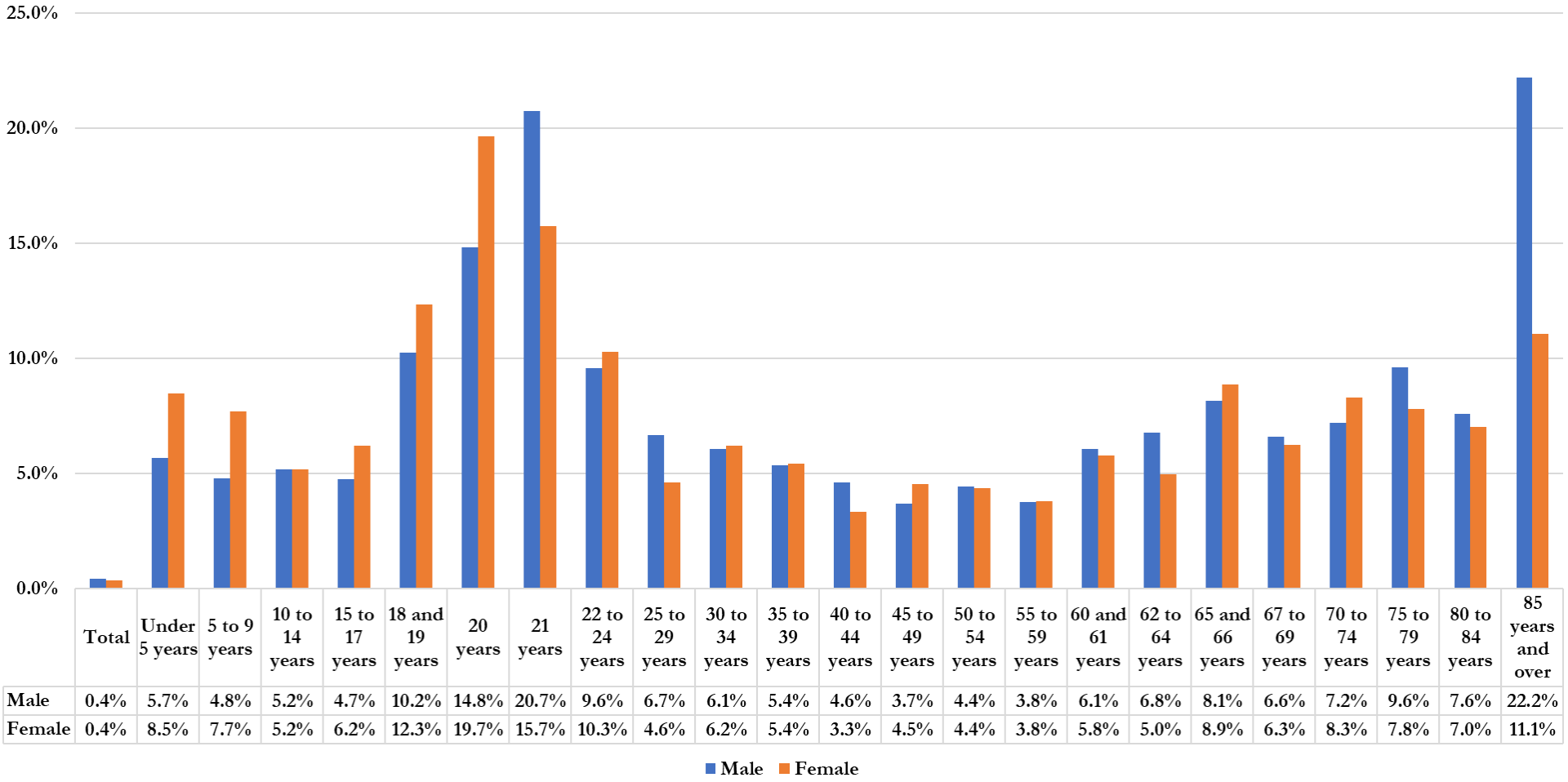
This will have myriad financial and economic repercussions for the "winners" and "losers" that municipalities will randomly become. One significant example is funding for school districts. **Figure 9** shows the losses and gains in the school-aged population. School districts stand to lose significant portions of funding as a result of a faulty headcount. For example, RSU 34 (serving Alton, Bradley and Old Town) lost 422 students from its school-aged children count. In 2011, there were 290 students attending its Leonard Middle School[2]. This loss is akin to artificially removing the students from more than an entire school from its school district. Conversely, some lucky school districts such as Deer Isle-Stonington Community School District would see a 35% increase in its school-aged population.

It is important to note that these results are based on random draws; outcomes for Maine could be entirely different in another iteration of the algorithm. For this reason, we close by urging the U.S. Census Bureau to provide more demonstration datasets and to release raw noise-inject data that include negative counts. This will help data users approximate margins of error for the 2020 published data and assess how these errors will manifest in the future. Without this ability, we will cease to use most of the published decennial data and be forced to seek alternative data sources.

[2] Source: Maine Education Data Warehouse

Tables and Figures

Figure 1. Mean Absolute Percent Error for age and sex, all counties in Maine



	Total	Under 5 years	5 to 9 years	10 to 14 years	15 to 17 years	18 and 19 years	20 years	21 years	22 to 24 years	25 to 29 years	30 to 34 years	35 to 39 years	40 to 44 years	45 to 49 years	50 to 54 years	55 to 59 years	60 and 61 years	62 to 64 years	65 and 66 years	67 to 69 years	70 to 74 years	75 to 79 years	80 to 84 years	85 years and over
Male	0.4%	5.7%	4.8%	5.2%	4.7%	10.2%	14.8%	20.7%	9.6%	6.7%	6.1%	5.4%	4.6%	3.7%	4.4%	3.8%	6.1%	6.8%	8.1%	6.6%	7.2%	9.6%	7.6%	22.2%
Female	0.4%	8.5%	7.7%	5.2%	6.2%	12.3%	19.7%	15.7%	10.3%	4.6%	6.2%	5.4%	3.3%	4.5%	4.4%	3.8%	5.8%	5.0%	8.9%	6.3%	8.3%	7.8%	7.0%	11.1%

Figure 2. Mean Absolute Percent Error, both sexes, 16 Maine Counties

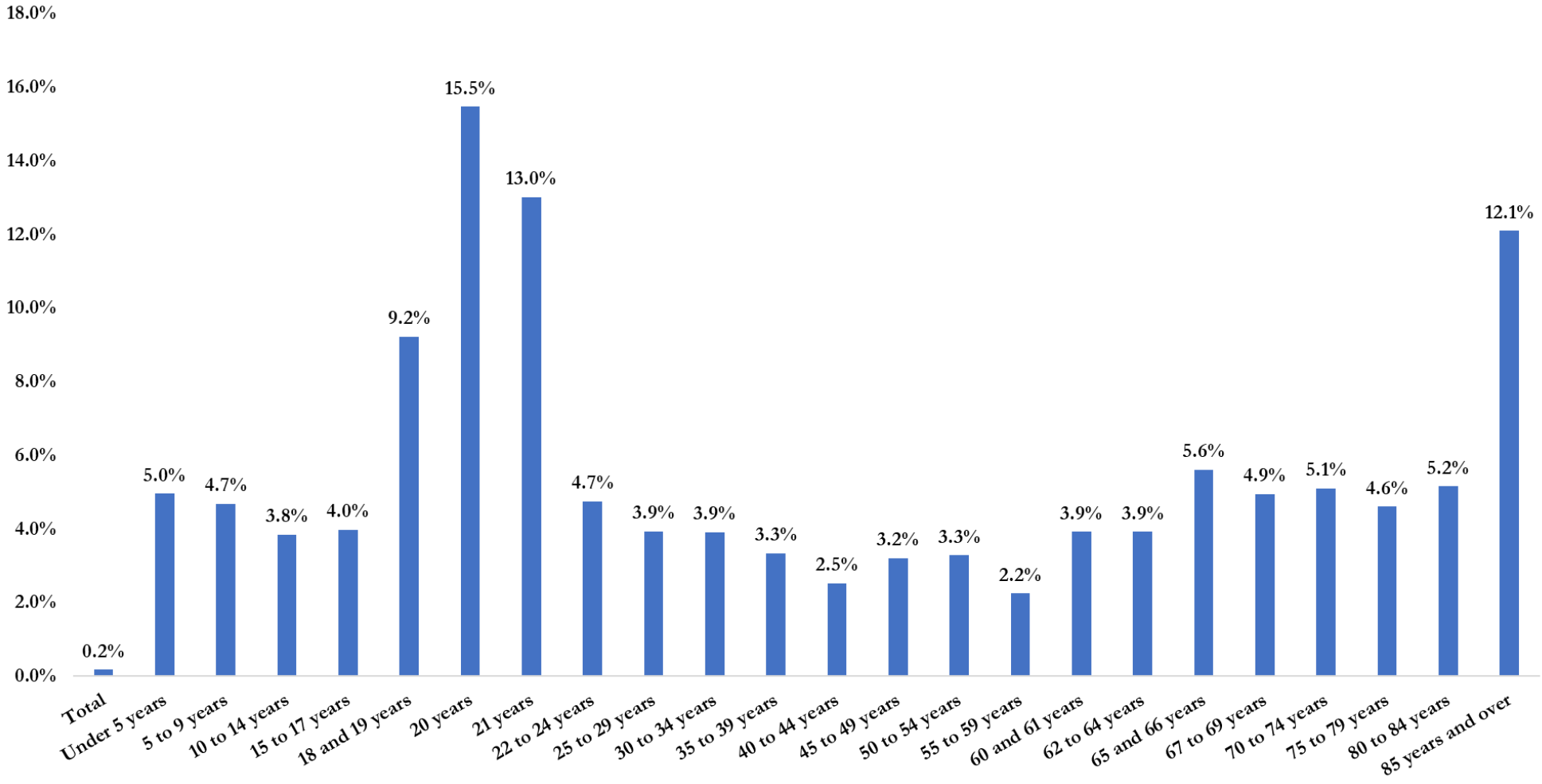


Figure 3. Mean Absolute Percent Error - Households by age of householder

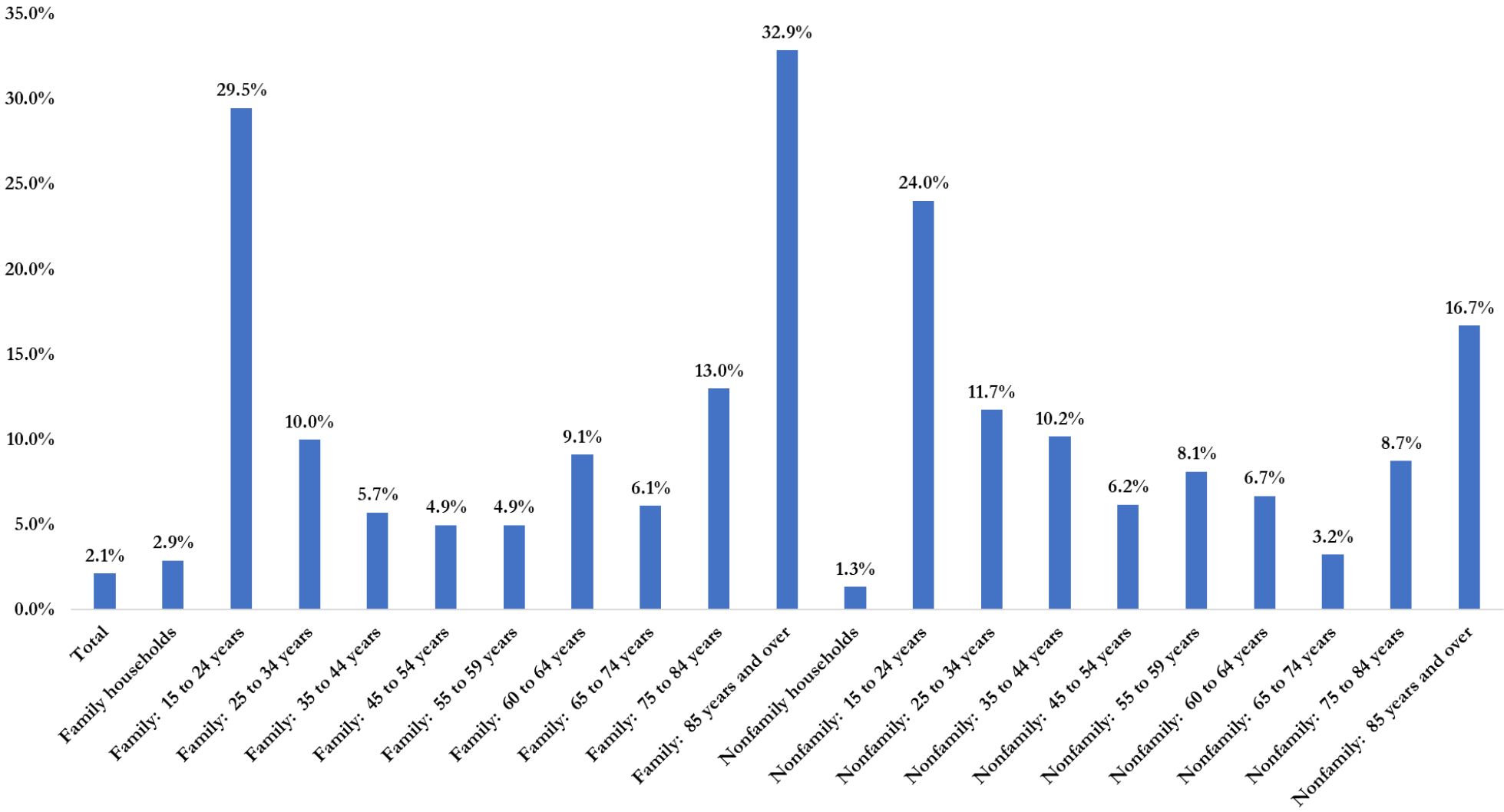


Figure 4. Race of Householder - Mean absolute percent error in 16 counties

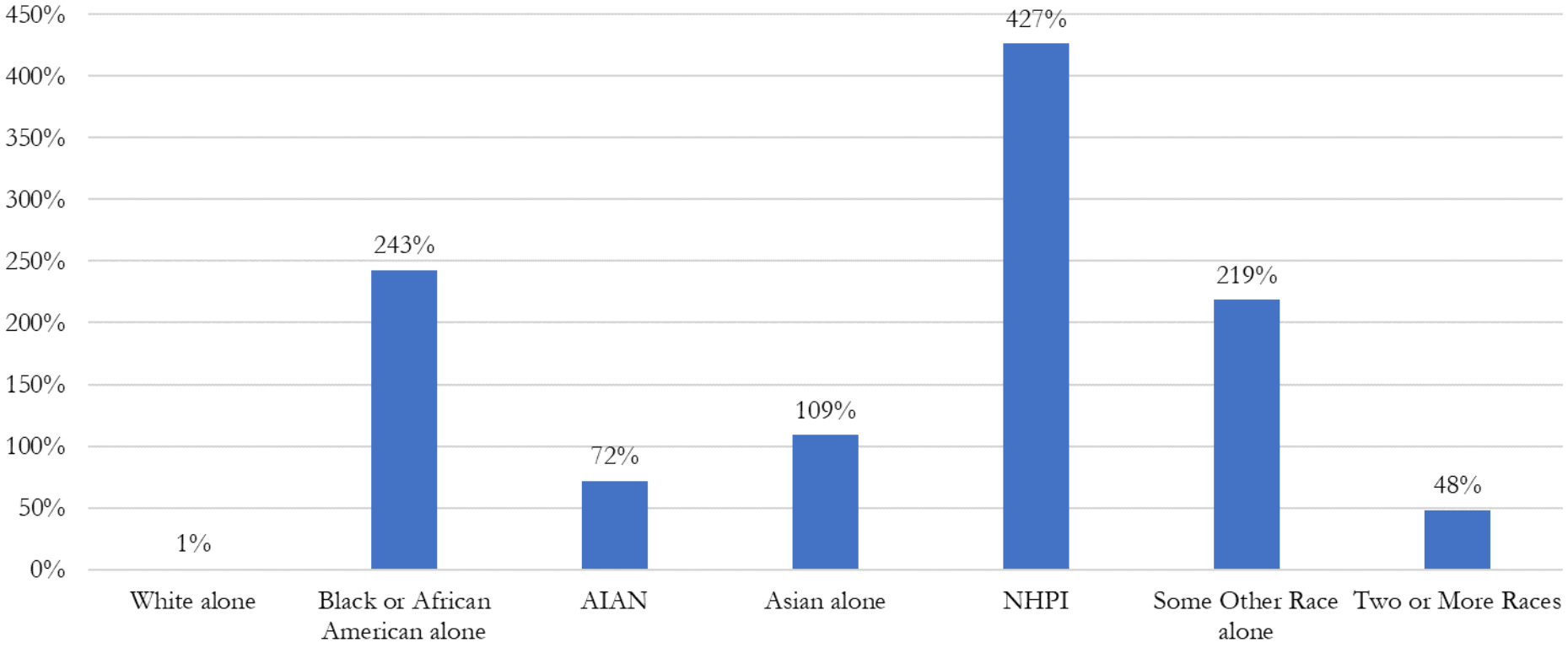


Figure 5. Percent Error in Total Population for All County Subdivisions

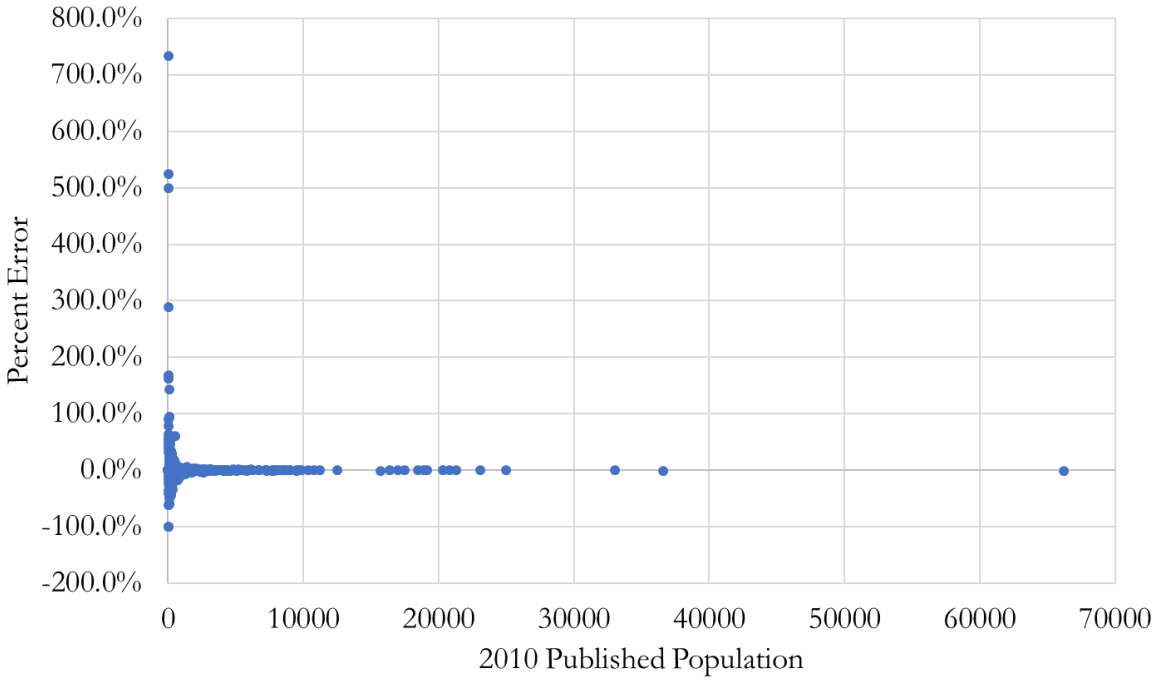


Figure 6. Percent Error in Total Population for County Subdivisions, Zoom View

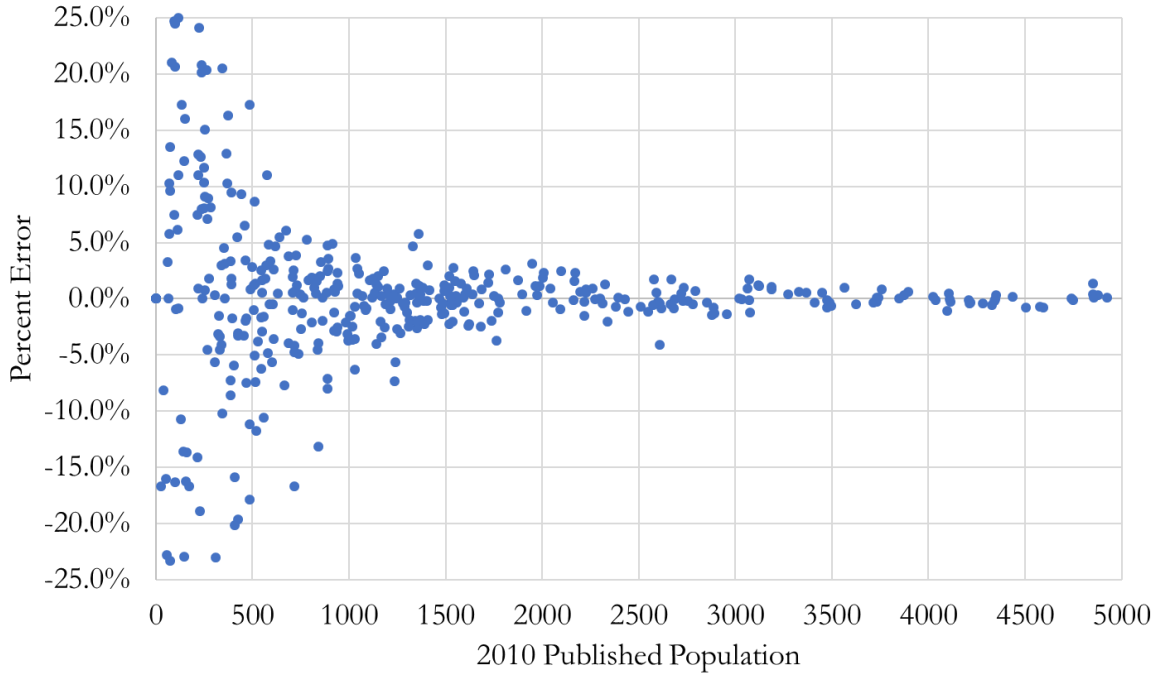
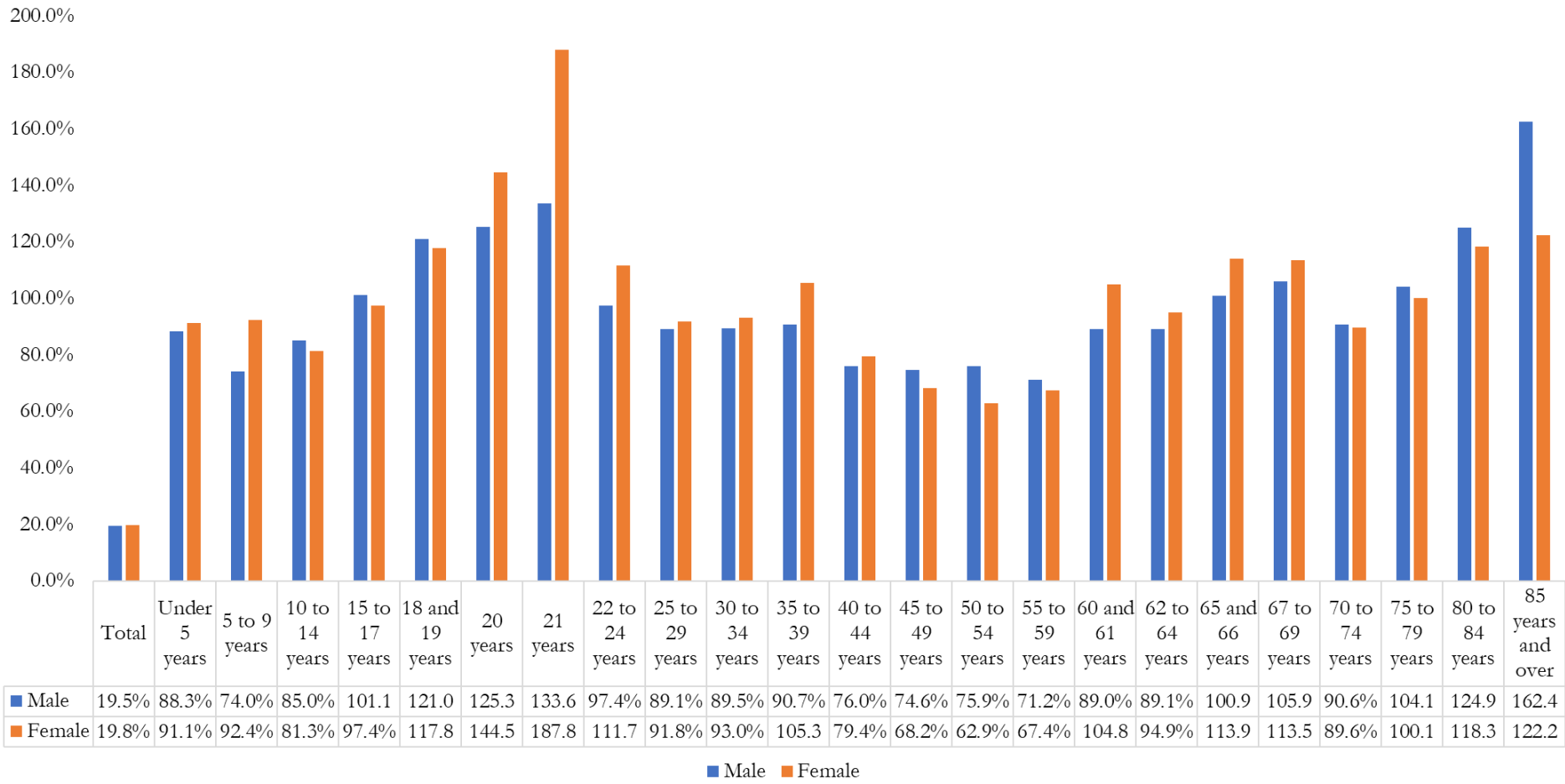


Figure 7. Mean Absolute Percentage Error for Age and Sex in Maine's County Subdivisions



	Total	Under 5 years	5 to 9 years	10 to 14 years	15 to 17 years	18 and 19 years	20 years	21 years	22 to 24 years	25 to 29 years	30 to 34 years	35 to 39 years	40 to 44 years	45 to 49 years	50 to 54 years	55 to 59 years	60 and 61 years	62 to 64 years	65 and 66 years	67 to 69 years	70 to 74 years	75 to 79 years	80 to 84 years	85 years and over
Male	19.5%	88.3%	74.0%	85.0%	101.1	121.0	125.3	133.6	97.4%	89.1%	89.5%	90.7%	76.0%	74.6%	75.9%	71.2%	89.0%	89.1%	100.9	105.9	90.6%	104.1	124.9	162.4
Female	19.8%	91.1%	92.4%	81.3%	97.4%	117.8	144.5	187.8	111.7	91.8%	93.0%	105.3	79.4%	68.2%	62.9%	67.4%	104.8	94.9%	113.9	113.5	89.6%	100.1	118.3	122.2

Figure 8. MAPE for age and sex by county, all county subdivisions in Maine

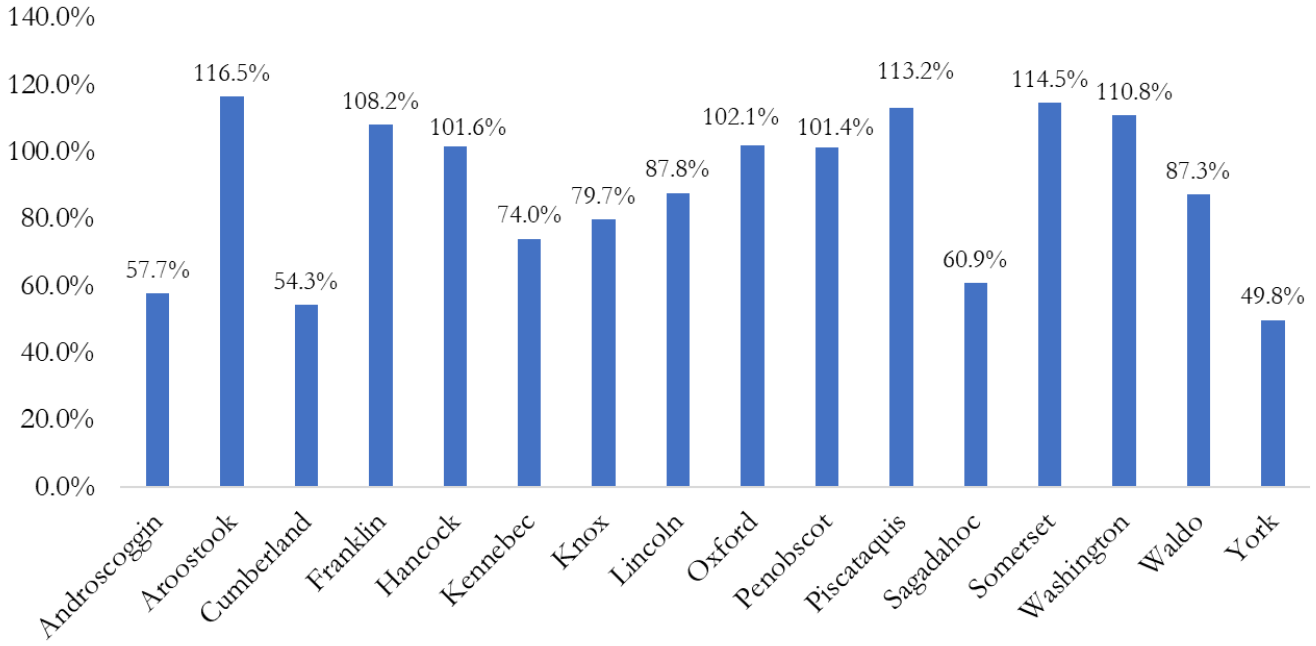


Figure 9. Miscount of school-aged children (5-17 years old) in all school districts in Maine

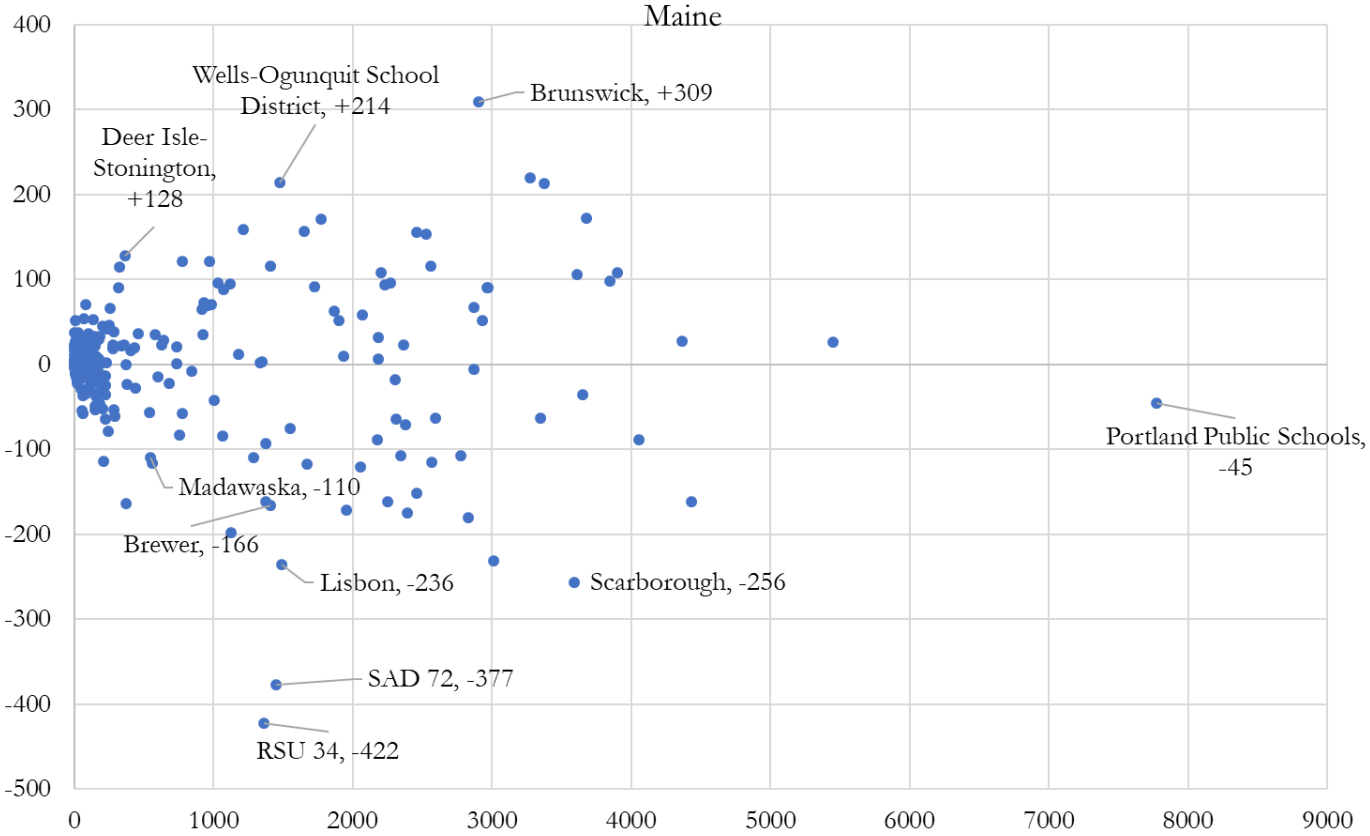


Table 1. Households by type and age of householder – highest error categories by county

(Where 100% means doubling and -50% means halving)

-50% and less | -49% to -25% | 25% to 99% | 100% and over

Percent Difference	Family: Householder 15 to 24 years	Family: Householder 25 to 34 years	Family: Householder 45 to 54 years	Family: Householder 60 to 64 years	Family: Householder 75 to 84 years	Family: Householder 85 years and over
Androscoggin County	-5%	3%	2%	7%	-2%	17%
Aroostook County	16%	-15%	8%	-1%	-6%	-23%
Cumberland County	20%	-2%	-1%	-3%	9%	-10%
Franklin County	70%	25%	2%	-11%	6%	-28%
Hancock County	11%	21%	7%	5%	-4%	28%
Kennebec County	-16%	-2%	0%	3%	-6%	-19%
Knox County	76%	17%	1%	8%	-12%	-37%
Lincoln County	109%	-20%	-3%	-1%	-13%	116%
Oxford County	2%	5%	0%	28%	-3%	-23%
Penobscot County	-20%	-5%	-1%	-2%	0%	10%
Piscataquis County	10%	0%	31%	29%	-51%	63%
Sagadahoc County	-2%	29%	1%	-6%	23%	-16%
Somerset County	16%	3%	-2%	-13%	14%	-4%
Waldo County	32%	-1%	-2%	-19%	47%	99%
Washington County	36%	-12%	17%	-10%	-11%	-31%
York County	-32%	1%	-1%	0%	-1%	3%

	Nonfamily: Householder 15 to 24 years	Nonfamily: Householder 25 to 34 years	Nonfamily: Householder 35 to 44 years	Nonfamily: Householder 45 to 54 years	Nonfamily: Householder 55 to 59 years	Nonfamily: Householder 75 to 84 years	Nonfamily: Householder 85 years and over
Androscoggin County	0%	-7%	10%	-6%	-1%	-6%	8%
Aroostook County	-7%	23%	2%	3%	-4%	-14%	5%
Cumberland County	-5%	-2%	1%	0%	-1%	-3%	1%
Franklin County	-8%	-2%	6%	-1%	-7%	10%	35%
Hancock County	-28%	-1%	1%	1%	7%	12%	-20%
Kennebec County	-6%	-9%	-4%	-6%	4%	5%	-2%
Knox County	-12%	-9%	-31%	6%	-1%	2%	27%
Lincoln County	33%	-11%	-21%	0%	6%	-1%	-4%
Oxford County	48%	21%	-20%	-9%	11%	-1%	15%
Penobscot County	-3%	-7%	11%	5%	0%	-1%	-11%
Piscataquis County	152%	14%	12%	28%	-45%	44%	-31%
Sagadahoc County	48%	39%	-4%	-17%	14%	5%	-39%
Somerset County	17%	-17%	-7%	2%	12%	9%	16%
Waldo County	-2%	17%	2%	-1%	-11%	-12%	7%
Washington County	4%	-5%	28%	12%	-2%	12%	-34%
York County	11%	-4%	-3%	-1%	-3%	-3%	12%

Table 2. Percent Error for Race of Householder by County

	Total	White alone	Black or African American alone	AIAN	Asian alone	NHPI	Some Other Race alone	Two or More Races	MAPE
Androscoggin County	0%	0%	-21%	23%	11%	-25%	72%	27%	**22%**
Aroostook County	1%	0%	82%	-30%	115%	-25%	627%	-11%	**111%**
Cumberland County	-1%	0%	-19%	24%	-8%	-33%	-58%	-15%	**20%**
Franklin County	5%	1%	1029%	182%	151%	260%	557%	81%	**283%**
Hancock County	1%	0%	105%	48%	-13%	1000%	26%	80%	**159%**
Kennebec County	-1%	0%	8%	-8%	-7%	92%	-53%	-20%	**24%**
Knox County	2%	0%	154%	137%	68%	500%	367%	16%	**155%**
Lincoln County	4%	2%	416%	23%	360%		364%	47%	**152%**
Oxford County	2%	1%	80%	118%	121%	25%	11%	22%	**47%**
Penobscot County	-1%	0%	0%	-16%	-15%	442%	9%	-36%	**65%**
Piscataquis County	8%	3%	813%	274%	423%	33%	683%	126%	**295%**
Sagadahoc County	4%	2%	73%	30%	23%	2400%	77%	162%	**346%**
Somerset County	2%	0%	186%	47%	140%	467%	462%	75%	**172%**
Waldo County	1%	0%	529%	118%	103%	480%	86%	-41%	**170%**
Washington County	1%	1%	367%	-29%	163%	1000%	0%	-8%	**196%**
York County	-1%	-1%	-7%	42%	-19%	-43%	-46%	-2%	**20%**
MAPE	**2%**	**1%**	**243%**	**72%**	**109%**	**427%**	**219%**	**48%**	