



STATE OF WASHINGTON

OFFICE OF FINANCIAL MANAGEMENT

Insurance Building, PO Box 43113 □ Olympia, Washington 98504-3113 □ (360) 902-0555

August 20, 2020

Disclosure Avoidance System Team,

The majority of the data output from the DAS appears to be unfit for most uses. The decoupling of persons, their place of residence, and their demographic characteristics (age, sex, race, and ethnicity) in the DAS output data is alarming because it will adversely impact voter representation, redistricting, and every sub-state statistic that uses, or is based on, census data.

We implore you to aggressively and dramatically reduce the statistical noise in the data introduced by the DAS to improve the accuracy and precision of the data at all levels of geography. We believe that Washington residents will be harmed if the data is not markedly improved.

All of the concerns we expressed in our February 6, 2020 demonstration data feedback letter are still of concern to us:

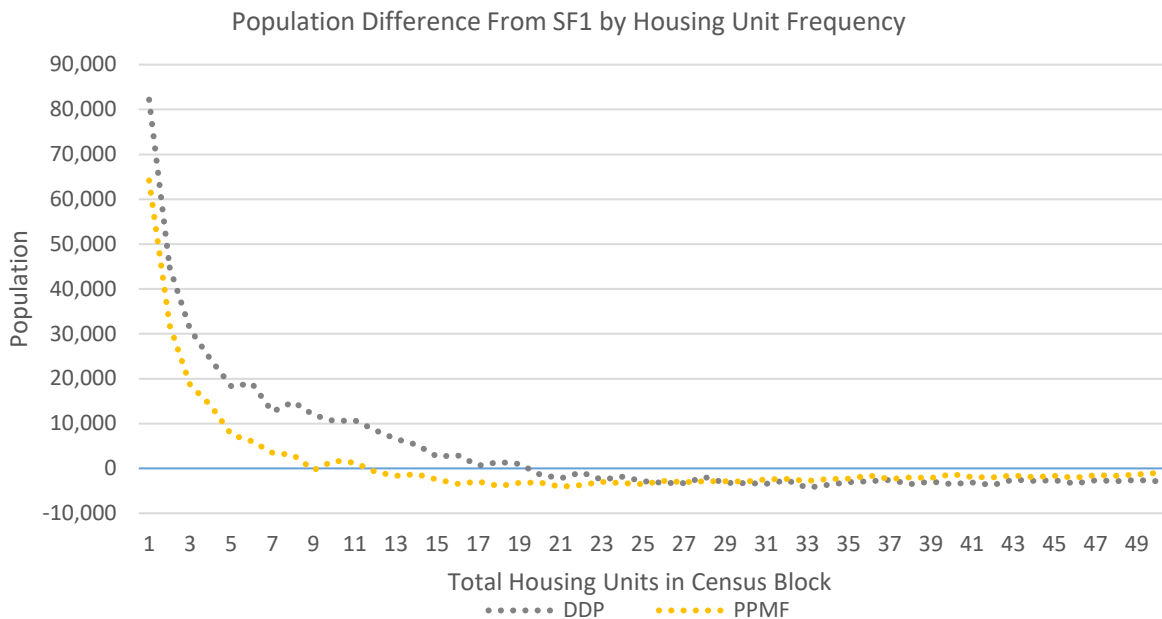
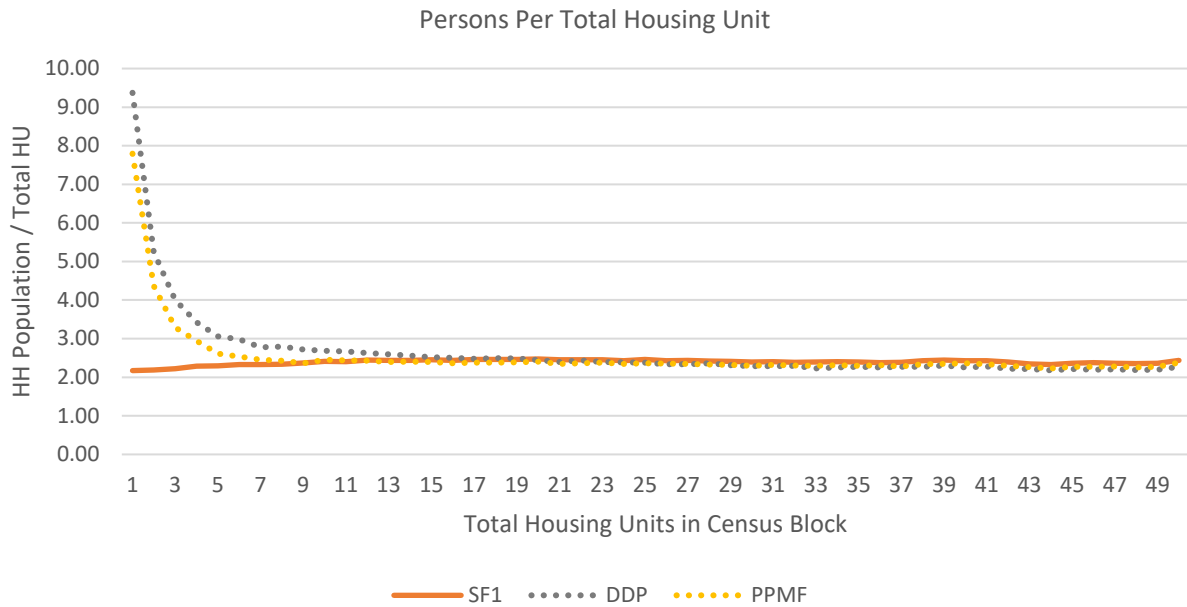
- **The household data was unrealistic.** Since the PPMF release did not include household information, our concerns have not changed.
- **There is bias in the PPMF data that causes areas with small populations to get larger while areas with larger populations get smaller.** While there has been some improvement in the PPMF over the DDP, bias still exists. We offer more evidence of this below.
- **There is bias in the data that makes communities with similar racial characteristics more dispersed geographically.**
- **The DAS will be particularly deleterious to our local governments.** Please review the attached age by sex comparison graphics for Washington cities and our age by sex by race comparisons for Washington counties.
- **The alarming differences between the PPMF data compared to SF1 are extremely problematic for real world applications.**

Below we outline some of the issues we have found in the data output from the DAS. We are concerned that we are only scratching the surface. It is highly likely there are other significant issues that have not yet come to light. The issues described below are categorized under the broad headings of biases and illogical values.

Biases

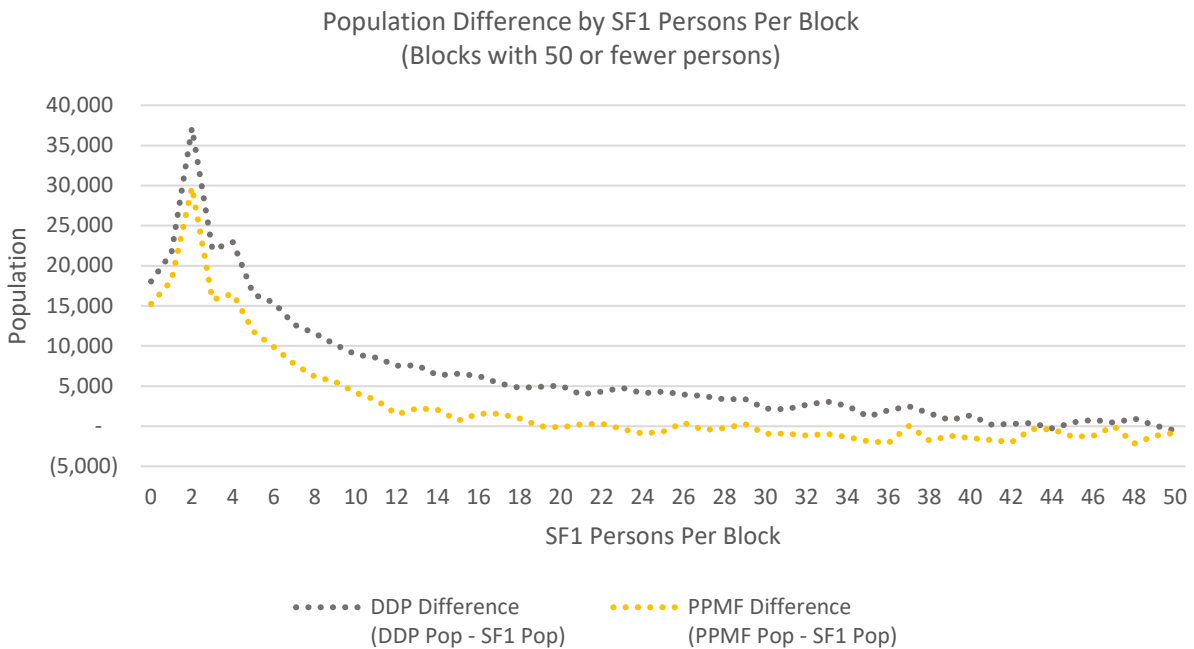
There appears to be bias in both the DDP and PPMF data characterized by an increase in population in census blocks with fewer than 9 housing units. We used total housing units from SF1 to calculate a

persons per total housing unit statistic for each block with SF1, DDP and PPMF populations. We used SF1 total housing units as they are invariant across products. Blocks with a small number of housing units, less than 9 in Washington’s case, appear to have higher than expected populations and blocks with more than 20 housing units appear to have lower than expected populations based on the average persons per total housing unit statistic. In terms of household population, blocks with only one housing unit had collectively 64,195 more people in the PPMF than SF1. Further, in the PPMF there are 15,253 people in blocks that had housing but zero population in SF1. These differences will adversely impact all kinds of analyses that rely on census data from political redistricting to planning for urban growth. While the PPMF figures are an improvement over the DDP data, the differences from SF1 remain alarmingly high and show continued bias in the DAS output.



Another way to see bias in areas with smaller populations is by calculating the number of people per block by SF1 population counts. The data show a clear upward bias in the PPMF and DDP populations for blocks with small numbers of people. For example, the PPMF has 15,253 people in blocks that had zero population in SF1. The table and graph below illustrate the excess population allocated to blocks compared to the SF1 population. Additionally, the PPMF has 18,136 more people in blocks that had only 1 person per block in SF1.

SF1 Persons per Block	SF1 Block Frequency	SF1 Total Population	DDP Total Population	PPMF Total Population	DDP Difference (DDP Pop - SF1 Pop)	PPMF Difference (PPMF Pop - SF1 Pop)
0	76,800	0	18,016	15,253	18,016	15,253
1	2,832	2,832	24,453	20,968	21,621	18,136
2	5,405	10,810	47,990	40,608	37,180	29,798
3	3,206	9,618	31,649	25,387	22,031	15,769
4	3,711	14,844	37,787	31,359	22,943	16,515
5	2,908	14,540	30,970	26,366	16,430	11,826
6	2,711	16,266	31,683	26,112	15,417	9,846
7	2,423	16,961	29,611	24,706	12,650	7,745
8	2,368	18,944	30,506	25,135	11,562	6,191
9	2,175	19,575	29,766	25,139	10,191	5,564
10	2,135	21,350	30,271	25,506	8,921	4,156



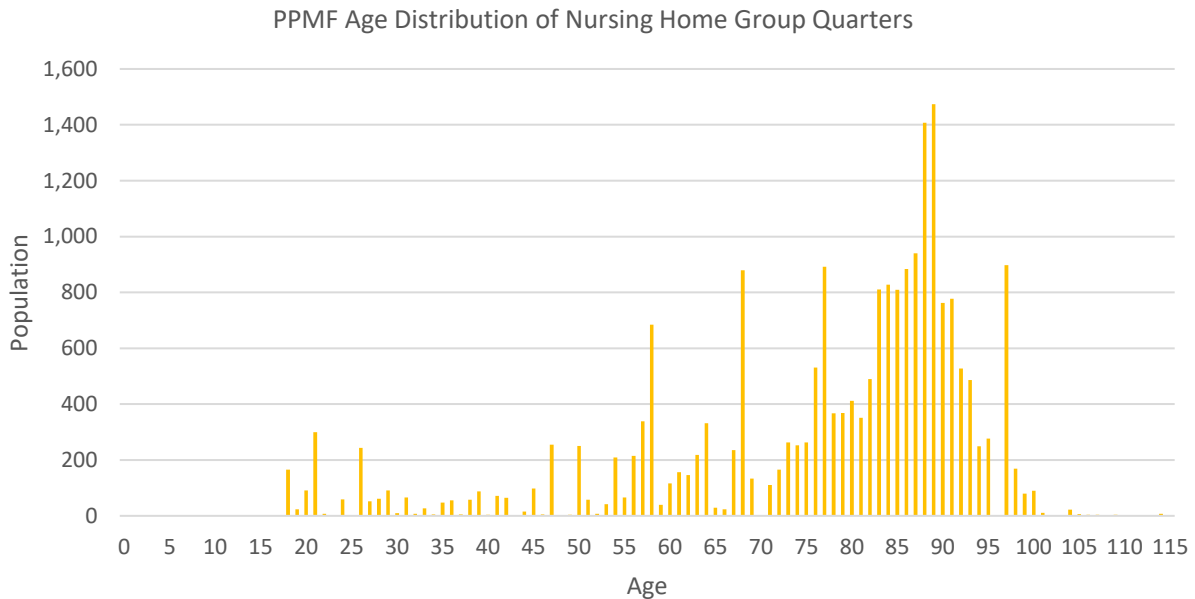
Illogical Values - Group Quarters

We found numerous issues with the age and sex distributions in the PPMF group quarters data. We did not have time to examine the group quarters data by race and Hispanic origin but we expect to find problems with that data as well. These errors will impact our state in a number of ways, including our annual population estimates and forecasts.

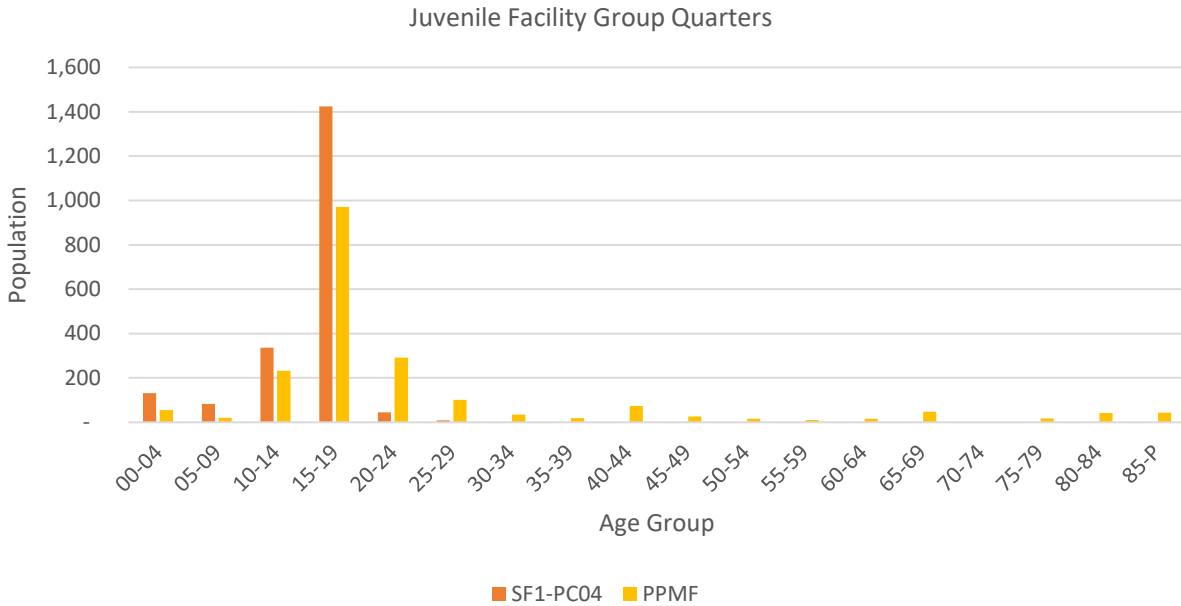
Washington RCW 44.05.140 (4)(d) directs the redistricting commission to “Adjust race and ethnicity in districts, wards, and precincts in a manner that reflects the inclusion of inmates and residents in the population count of the district, ward, or precinct of their last known place of residence.” We believe that it will not be possible for the redistricting commission to carry this law out unless excessive noise in the DAS output is eliminated.

In most cases we examined, the PPMF data was more accurate than the DDP, but not always. In a particularly egregious example, the Washington Corrections Center for Women, located in census block 530530725042008, was reported as 99% female in the 2010 SF1 data. In this case the PPMF data was far less accurate than the DDP. For this female prison block, the PPMF had 12% female compared to the 25% in the DDP. It would be impossible for the redistricting commission to adjust this block’s census population if only 100 (12%) of the approximate 800 women in the prison are represented in the census data. We find it very concerning that the PPMF data is less realistic than the DDP in this case. **This easy to explain situation casts doubt on the accuracy of the data for larger, more heterogeneous blocks.**

We also observed several illogical situations in the age distribution of the group quarter population. In nursing homes, there are no population for ages 70, 96, 102, 103, 110, 111, 112, 113, or 115 in the Washington PPMF data. We understand that the PPMF is generated from tabular data and that the traditional tabular release of decennial data does not include single year of age GQ data, but this is indicative of the bias in the data and is alarming.



Another example of biases in more specific population subsets exists in the group quarters data for juvenile facilities. The PPMF data contains 345 people over 30 years old or 17% of the total. While 345 is not a large number this is concerning as no population over age 30 was included in the SF1 data (PC04).



Looking at individual blocks where facilities are located raises further concerns. The table below lists group quarter data by age for the four largest juvenile facilities in Washington in 2010. The PPMF data for the Maple Lane School lists 177 GQ population. Thirty-seven, or 20% of the total population, are of ages clearly out of the bounds of expectations for a youth facility. Oddly, there are an unusually high number of 19 year olds but 14, 15, and 18 year olds are missing from the data.

Juvenile Facility Population (Census block)				
Age	Maple Lane School (530670127205045)	Green Hill School (530419710001027)	Echo Glen Youth Center (530330326023045)	Naselle Youth Camp (530499504002046)
13	41	0	0	0
16	0	0	5	0
17	0	0	52	0
19	99	155	26	155
23	0	0	14	0
28	0	4	0	4
32	0	0	5	0
47	0	0	2	0
68	30	0	0	0
81	7	0	0	0

Another PPMF data anomaly is the 631 people in college group quarters under the age of 15 and 420 people that are 65 years or older. There were no people under the age of 15 or over the age of 65 years according to SF1 table PC08.

Nursing homes create special circumstances within census blocks that need to be preserved. The census block in the following image contains a nursing home and an assisted living facility in Tonasket city, Okanogan County. In the PPMF, there are 84 people, all white race, all male, all age 68.



Problems with the age and sex distribution in group quarters also exist at larger geographies. The Life Care Center of Kirkland was at the center of the early COVID-19 outbreak in the U.S. and was on the front page of most newspapers and served as the lead story on most television networks. The nursing home is in block group 530330220063 whose characteristics are shown in the table below. The age 65 and over population in the PPMF is 35% lower than in SF1. In the graphs below, the age and sex distributions for the DDP and PPMF appear disjointed, relative to SF1. While the latest PPMF product appears to have corrected some of the issues in the DDP, new ones were introduced. For example, the 85 and over population for males appears to have improved while 85 and over population for females worsened. Denominators for age-specific rates have been essential for responding to the pandemic. In this early case, the PPMF data at the block group-level would have been useless. **Dramatic errors in the age-sex distribution would make understanding the incidence and prevalence of disease for small areas impossible.**

Population Characteristics (Block group 530330220063)			
	SF1	DDP	PPMF
Total	1,204	1,169	1,196
Group quarters	140	112	115
Male	558	510	518
Female	646	659	678
65 and older	230	211	149



Illogical Values - Small Areas

There are important differences in the population by race data at small geographies. The following table lists the count of blocks that are 100% single race alone in SF1 and the PPMF. The data show that only about 50% of single race blocks are the same in both products. **While a more thorough analysis of change by geography is warranted, the sheer amount of displacement is very concerning. This will have a significant impact on any analysis where racial characteristics are important, including studies that look at changes in characteristics over time.**

Single Race Blocks				
Race Category	SF1 Block Count	PPMF Block Count	Coincident Blocks (same block in both files)	Percent Coincident (coincident blocks/ SF1 count)
White Alone	36,249	37,516	18,483	51.0%
Black Alone	139	217	7	5.0%
AIAN Alone	371	597	72	19.4%
Asian Alone	278	335	5	1.8%
NHOPI Alone	40	42	0	0.0%
SOR Alone	552	805	30	5.4%
Total	37,629	39,512	18,597	49.4%

Our office publishes annual estimates of population by age, sex, race and Hispanic origin down to the census tract level. These estimates were initially produced to serve state and county public health data needs but have many other applications. Data from the 2010 Census at the tract level serve as the population base for these estimates. In the DDP and PPMF, the tract level data by age, sex, race and Hispanic origin are mostly unsuitable for analysis. This is true even for some of the county data (see attached county age, sex and race graphs). Public health departments in our state are currently overwhelmed with addressing the COVID-19 epidemic and have not had the time to analyze the DDP and PPMF data to voice their specific concerns. **These data are extremely important and the various problems have not been addressed in either the DDP or the PPMF.**

An example of implausible age, sex and race data is found in Garfield County—a county with a small, racially homogenous population. In the PPMF, the black male population is zero in all five-year age categories except for the 0-4 year age group where there are over 20 people. It is highly unlikely that you would find over 20 black children aged 0-4 with no adult black males available and only 5 black females aged 45-49. This particular example is characteristic of the issues that will arise when the DAS decouples persons from their household and place of residence.

While the PPMF age and sex distribution for some counties showed some improvement, the age distribution at the city level, particularly cities under 30,000, still is a problem (see attached city age and sex graphs). City-level data are used in a variety of applications. For example, if the 65 and over population is grossly over or under represented, local officials not be able to adequately plan for the needs specific to this age group (e.g., housing, ADA improvements, transportation, emergency services, etc.).

The PPMF data have far too high of a proportion of single sex census blocks where persons are either all male or all female. The overwhelming majority of the high count, single gender blocks are household population blocks that do not have group quarters population. Furthermore, there are many blocks that consist solely of the very young and the very old. In Washington, there are 401 census block where all of the population is over 85 year old and 3,353 census blocks where all of the population 14 years old or younger.

Single Sex, Very Young and Very Old Blocks		
	Number	Percent
Blocks in WA with population	116,672	
Blocks with > 10 people either all Male or Female	3,490	3.0%
Blocks where all pop <= 14 years old	3,353	2.9%
Blocks where all pop >= 85 years old	401	0.3%

Illogical Values - Census Block Examples

These following images show examples of census blocks with almost no diversity in the PPMF data.

Spokane County (Unincorporated) | 530630105032024 | 56 persons all male



Jefferson County (Port Townsend city) | 530319506015064 | 58 persons all female



King County (Bellevue city) | 530330247012008 | 58 persons all female



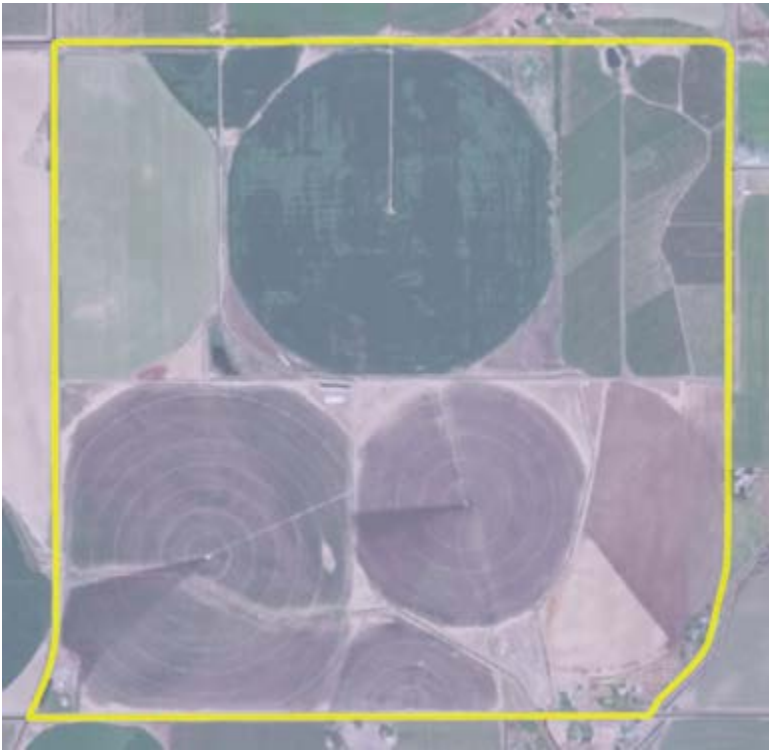
Whatcom County (Bellingham city) | 530730005021049 | 62 persons all male



Yakima County (Unincorporated) | 530770017012037 | all 41 persons ages 14 and under



Grant County (Unincorporated) | 530250114011055 | all 37 persons ages 14 and under



Clark County (Battleground city) | 530110404143011 | all 36 persons ages 14 and under



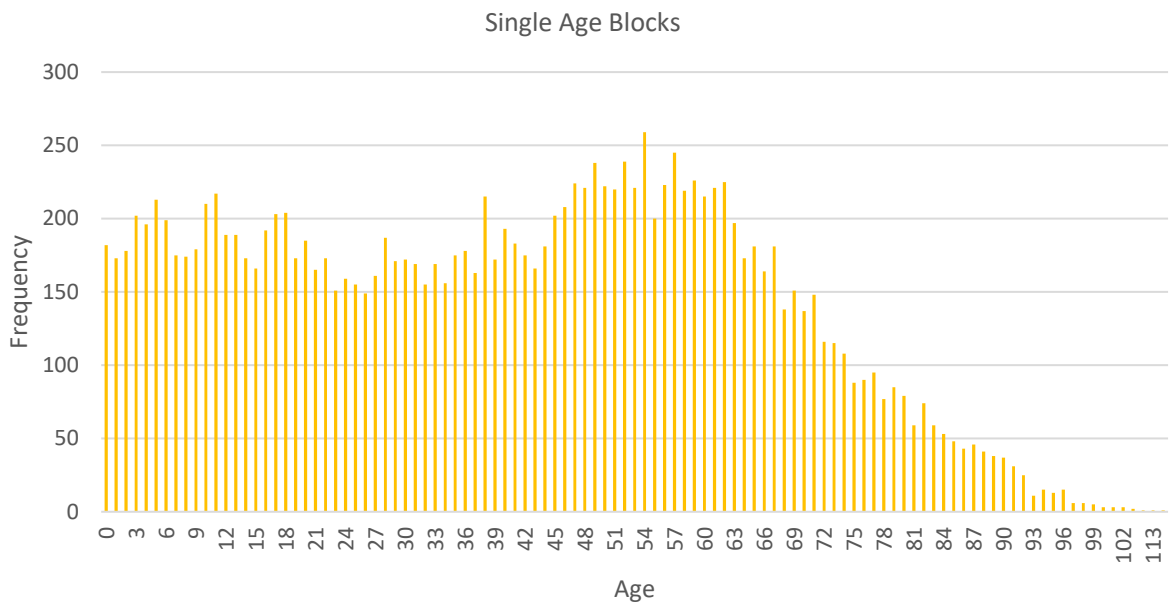
Clark County (Vancouver city) | 530110419001039 | all 13 persons ages 85 and over



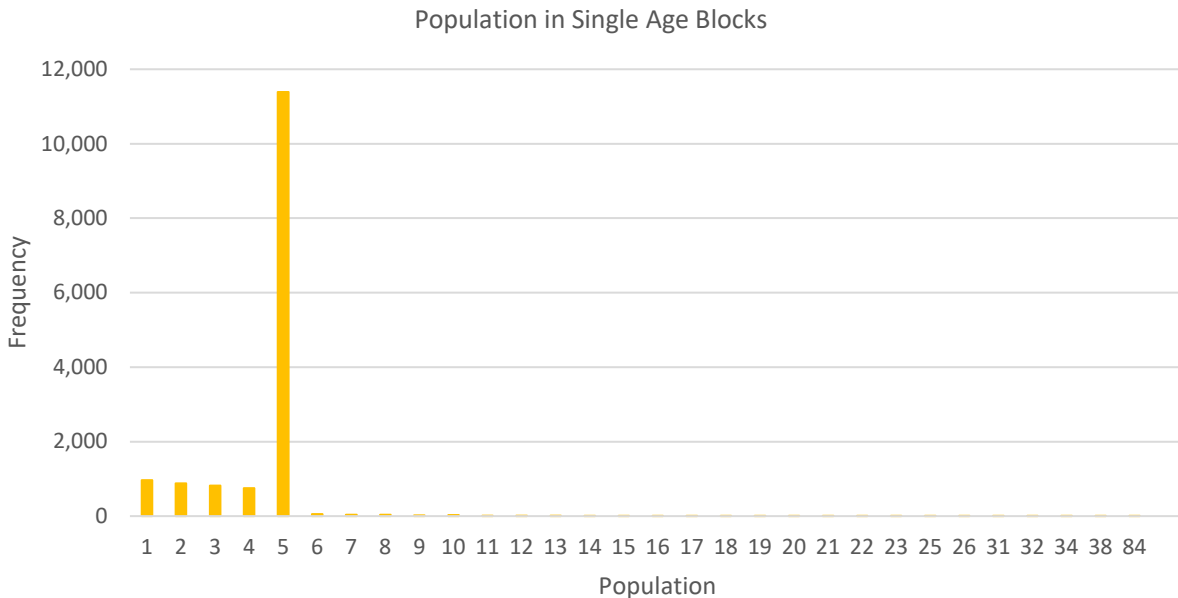
Clallam County (Unincorporated) | 530090017001050 | all 18 persons ages 85 and over



Census blocks with unexpected age structures are not limited to the under 14 and 85 and over age groups. Washington has 15,055 census blocks where the entire population in the block is the same age. The graph below shows the distribution of these blocks by age.



The problem with five identical observations in blocks, originally identified by Jan Vink, is apparent in Washington as well. See the graph below with the number of blocks where everyone is the same age.



Conclusion

Multiple Washington laws (RCW’s 36.13.100, 43.62.020, 43.62.030, 66.08.200 and 66.08.210) direct the use of census population counts or the census based population estimates developed in our office for the equitable distribution of approximately 200 million dollars per year. For over 50 years our office has relied on accurate and internally consistent census counts of population and housing at the county and city levels. With simple analyses we have demonstrated biases and shown illogical values across a wide variety of circumstances. These biases are not randomly distributed and will harm many communities. My agency alone has hundreds of thousands of dollars invested in data systems that use federal census counts as core inputs. **We believe that the equitable distribution of funds based on population will be harmed if the accuracy of the data is not markedly improved.**

In addition, it is difficult to see how racial/ethnic minorities can be accurately represented if they are not accurately portrayed in the census data at the geographic levels needed for apportionment.

We implore you to make all population and housing counts and their characteristics invariant for the county, tract, tribal and city geographies. This is the only way to insure state local and tribal governments and their stakeholders have the accurate data they need for local decision making.

Sincerely,

Mike Mohrman, State Demographer
 Forecasting and Research Division, Office of Financial Management

cc Marc Baldwin, Assistant Director