

About this document

The [Viz-a-lyzer](#) is an interactive tool for people to visually explore and analyze Bexar County data by zip code tabulation area (ZCTA) and year, with comparisons to Bexar County, Texas, and the United States where available. These Technical Notes provide an overview of how to use and interpret the Viz-a-lyzer data. It is intended for the person looking at our website for the first time, but the experienced data user may also find a review helpful. Below is a key to the contents.

Geography.....	2
Margins of error	2
Understanding rates.....	3
Effects of small numbers on rates.....	3
Understanding percents.....	4
Military zip codes	5
Data classification categories.....	5
“No data”	7
“Not reported” data.....	7
Exceptions and outliers	8
Notes on specific indicators	9
Birth indicators	9
Child abuse and neglect indicators	10
Hospitalization indicators.....	10
Quick guide to interpreting indicators	12

Geography

While most of us think of them as geographic areas, ZIP (“Zone Improvement Plan”) codes are really just collections of addresses created by the U.S. Postal Service to make delivery of mail more efficient. Each address’s ZIP code ties that address to a particular post office or delivery station. ZCTAs (ZIP Code Tabulation Areas), on the other hand, are geographic area boundaries created by the U.S. Census Bureau to roughly represent the geometric “shape” of the related physical addresses on a map. Usually, the ZCTA is the same as the zip code for an area, but not always. And not all zip codes are captured by a ZCTA, particularly those where the zip code represents just one mailing address, as is often the case for large businesses. [Read more about ZIP codes and ZCTAs](#) on the U.S. Census Bureau’s website.

Please note that the Filter just provides a *view* of data for individual ZCTAs. [The Viz-a-lyzer](#) doesn’t actually *calculate* any indicators for a group of ZCTAs or for any other geographic level. When you select City Council District 5, you are not seeing the average educational attainment for the Council District. You are seeing the educational attainment for each ZCTA that is approximated to be included in that Council District. The intention is to show a “common sense” set of ZCTAs for a given jurisdictional division, and in many cases, the decision of whether to include a ZCTA is a judgment call.

Margins of error

Analysis of the data typically consisted of calculating percents and rates, with margins of error (MOEs) where appropriate. When we can’t measure all of something, like people in a city, we sample them – measure only some to get an idea (estimate) of what’s true for everyone. Sampling introduces error and uncertainty, and the margin of error – for example, “plus or minus three hundred” – is a measure of how much uncertainty there is. The smaller the sample in relation to the total population, generally, the larger the margin of error. So estimates with wide margins are less trustworthy than those with narrow margins. Think of it as a “grain of salt”-ometer. Table 1 gives an example using total population.

Table 1. Total population, 2016

ZCTA	Total population	Margin of error (MOE)	Lower limit	Upper limit
78002	7,630	±1,154	6,476	8,784
78015	12,528	±784	11,744	13,312
78023	29,351	±1,009	28,342	30,360

Source: US Census Bureau; ACS 5-Year Estimates, Table DP05, 2016.

Understanding rates

A rate is the number of cases or events in a specified period of time and geography divided by the population who could have experienced – were “at risk” for – the case or event within that same period of time and geography. Rates are often multiplied by a factor of 1,000, 10,000, or 100,000 for convenience.

Teen birth rate [Case/event]	in 78203 [Geography]	in 2013 [Time]	was 189.4 [Frequency]	per 1,000 [Multiplier]	females aged 15-19 years [Population “at risk”]
---------------------------------	-------------------------	-------------------	--------------------------	---------------------------	--

So if you want to compare and contrast something across geographic areas, you usually want to compare a rate per population, not a number of events or cases, for a specific period of time. Consider the teen birth rates for 2013 shown in Table 2 below.

Table 2. Teen birth rate, 2013

2013	78203	78204	78249
Teen births	25	24	24
Females 15-19 years	132	364	2,759
Rate per 1,000 Females 15-19 years	189.4	65.9	8.7

Source: Texas Department of State Health Services, 2013.

When calculating rates and ratios, you have to be careful to use the appropriate time periods, geographical boundaries and populations for the denominator. Key point: because population sizes and characteristics are different for different areas and can change for a single area over time, it’s smart to calculate a rate per population to control for those differences and changes.

Effects of small numbers on rates

Chance variation is common when the numbers being used to calculate rates are small. Large swings can occur in the rates that do not necessarily reflect real changes or statistically significant differences. That is why it is important to show the actual number of events or cases when the numbers are small and result in huge swings from year to year. We call that a “volatile” or “unstable” rate. Consider the child abuse rates shown in Table 3 below. The 78215 zip code is located in the downtown area with a small population under 18 years old. The rates and the population vary widely from year-to-year. Note that the 2011 confirmed child abuse rate more than doubled from the 2010 rate, even though there were only three more cases of child abuse than in 2010.

Table 3. Child abuse rates, 2010-2016

Child Abuse in 78215	2010	2011	2012	2013	2014	2015	2016
Confirmed Victims	2	5	5	6	0	3	2
Population under 18	108	127	72	56	141	150	146
Rate per 1,000 Children	18.5	39.4	*	*	0.0	20.0	13.7

Source: Texas Department of Family and Protective Services, 2010-2016.

*Rate not reported due to low target population (less than 100)

When rates are based on small numbers, we need to be wary when making important decisions based on the data. The [Department of State Health Services](#) considers a rule to only calculate rates when you have more than 20 cases in your numerator: “a rate based on 20 cases has a 95% confidence interval about as wide as itself (the interval for a rate of 50 is between 25 and 75). Even large differences between two rates based on 20 cases or less are probably not statistically significant.” When an indicator has fewer than 20 cases, we often combine several years of data or consolidate geographical areas. Our current maps only show rates but we are working on an update to provide the actual number of cases to give more context and story-telling to the data.

Understanding percents

A percentage is just a rate multiplied by a factor of 100. Before comparing two places, always compare the population characteristics as a reference to give the percent context. A zip code might have a high percent in an indicator but have a small population in the zip code. Consider the population below 100% poverty in Table 4.

Table 4. Population below 100% poverty, 2016

2016	78202	78208
% Population below 100% poverty	44.5% (±5.0%)	35.0% (±7.2%)
Total population	12,339 (±901)	4,518 (±657)
Estimate of population below 100% poverty	5,491 (±62)	1,581 (±325)

Source: US Census Bureau; ACS 5-Year Estimates, Table DP03, DP05, 2016

Military zip codes

Bexar County has five military installations. It is important to keep in mind that these zip codes often look different than neighboring zip codes based on the military population who resides there (more males, higher educational attainment, higher geographic mobility, etc) and are therefore masked as outliers and excluded from the breaks. They also have smaller total populations and wider margins of errors. The military indicators include outliers for veterans and percent of people in the armed forces. Randolph Air Force Base is not included in several indicators because the target population is too small to calculate. Brooks Air Force Base (78235) closed in 2011 and was re-named Brooks City-Base as a unique project between local, state and federal government. Brooks City-Base is not included in the military filter but the zip code might look differently than neighboring zip codes as it continues to transition to a civilian-based area.

Table 5. Military installations in Bexar County, 2016

Military installations	ZCTA	Total population	Pct male	Pct with Bachelor's degree or higher
Randolph AFB	78150	13 (±13)	100.0% (±88.4%)	No data
Fort Sam Houston	78234	5,857 (±537)	54.5% (±4.2%)	31.9% (±4.5%)
Lackland AFB	78236	7,244 (±1,112)	72.0% (±4.0%)	44.4% (±8.9%)
Kelly Field Annex	78243	382 (±157)	90.3% (±10.3%)	36.4% (±36.3%)
Camp Bullis	78257	5,157 (±541)	50.5% (±3.0%)	61.4% (±6.5%)

Source: US Census Bureau; ACS 5-Year Estimates, Table DP05 and B15002, 2016

Data classification categories

When you classify your data, you can use one of many standard classification methods. Let's use total population of the 73 zip codes in Bexar County across the years 2010-2016 to show how the classification methods work in Table 6 below. Equal interval divides the range of attribute values (population size) into equal-sized subranges. This allows you to specify the number of intervals and determine the class breaks based on the value range (Table 7). A quartile classification assigns the same number of data values (zip codes) to each class. There are no empty classes or classes with too few or too many values (Table 8). Natural breaks are based on natural groupings inherent in the data. Natural breaks preserve similar values *within* classes and maximize the differences *between* classes (Table 9).

Using the [Map Reliability Calculator](#), we determined that division into four categories was more statistically reliable than five categories, and we selected equal interval classes to emphasize the attributes (population size in the example) relative to one another. The breaks are calculated using the entire year set available for the indicator so the minimum and maximum are based off

the entire data set and not re-calculated within each year. So using the population table below, the breaks for 2010-2016 will all be calculated with a minimum of 11 and a maximum of 68,265 and be divided up into four categories based on the spread.

Table 6. Total population by zip codes, 2010-2016

Total population by zip codes							
	2010	2011	2012	2013	2014	2015	2016
Minimum	11	71	64	29	19	18	13
Maximum	58,811	55,899	57,534	60,333	62,615	64,505	68,268

Source: US Census Bureau; ACS 5-Year Estimates, Table DP05, 2010-2016

Table 7. Equal interval classification

2016 Total population by zip codes		
Range	# ZCTAs	% of ZCTAs
11-17,075	32	43.84%
17,076-34,139	14	19.18%
34,140-51,204	17	23.29%
51,205-68,268	10	13.70%

Source: US Census Bureau; ACS 5-Year Estimates, Table DP05, 2010-2016

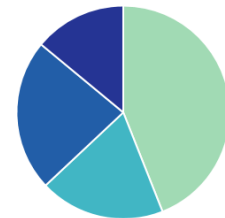


Table 8. Quartile classification

2016 Total population by zip codes		
Range	# ZCTAs	% of ZCTAs
0-8,115	19	26.03%
8,116-23,641	17	23.29%
23,642-38,503	18	24.66%
38,504-68,268	19	26.03%

Source: US Census Bureau; ACS 5-Year Estimates, Table DP05, 2010-2016

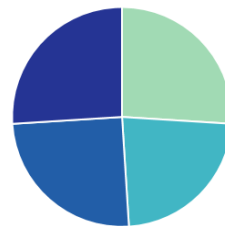
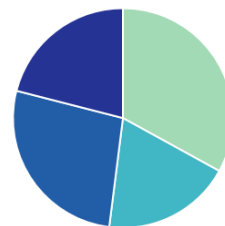


Table 9. Natural breaks classification

2016 Total population by zip codes		
Range	# ZCTAs	% of ZCTAs
0-12,280	24	32.88%
12,281-26,498	14	19.18%
26,499-43,710	20	27.40%
43,711-68,268	15	20.55%

Source: US Census Bureau; ACS 5-Year Estimates, Table DP05, 2010-2016



“No data”

For zip codes where there was either data not available or not applicable (no target population), then those were defined as ‘no data’ and are a white color on the map. For example, the family violence rate in zip code 78236 is not available for any year from 2014 to 2016.

“Not reported” data

For zip codes with low target population (less than 100), the data was suppressed and defined as ‘not reported’ and valued as 999999.0 in the table and a gray color on the map. An example is the asthma hospitalization rate per 10,000 for population under 18 in 2015 for 78205. There were only 42 children under 18 in that year therefore a rate was not calculated since the target population was under 100. Generally speaking, we apply a suppression rule of less than 100 for the population denominator and less than 20 for the case/event numerator. See Table 10 for more rules.

Table 10. Suppression rules

Indicator	Suppression
Child Abuse and neglect	If # of initial intakes was 1-5. If census population was <100 to calculate the rate. If # confirmed victims was 1-5.
Family violence rate	If census population was <100 to calculate the rate.
Births	If # births to females 15-19 was ≤9. If census population was <100 to calculate the rate. If # births with prenatal care was ≤9. If total # births was ≤20.
Asthma hospitalization 3-year rate	If census population was <100 to calculate the rate.
Diabetes hospitalization 3-year rate	If census population was <100 to calculate the rate.
Hypertension hospitalization 3-year rate	If census population was <100 to calculate the rate.
Injury hospitalization 3-year rate	If census population was <100 to calculate the rate.
Mental health hospitalization 3-year rate	If census population was <100 to calculate the rate.

Exceptions and outliers

For data analysis, we assume values will cluster around a central data point (mean or median) and follow a normal distribution. However, there are values that fall outside of defined standards, expectations or any other definition of normal and are called exceptions. Exceptions are included in the calculation of the equal interval breaks. When values fall too far outside of the central point, we call them outliers and exclude them from the analysis. We used the Tukey Fences method to identify outliers, which is based on a statistical calculation using the interquartile range (IQR), the difference between the first and third quartiles. The IQR points out the middle 50% of the range and how much variability there is in that 50%. When you leverage the IQR, you can establish a “fence” so that you know how far from the middle 50% a value can sit and still be considered “normal” rather than an outlier.

Consider the percent of male population from 2011 to 2016. As you can see in Table 11 below, the minimum and maximum are 42.0% and 100.0% respectively over the time period in some zip codes. However, when you apply the Tukey fence it points to a new set of values to define the Lower Fence (44.1%) and the Upper Fence (54.9%) in Table 12. Any values that are outside the fences are considered outliers and removed from the calculation of the equal interval classes (Table 13). The outliers appear on the map in the grey color (same as the not reported data) with a code of 888888.0 in the data table.

Table 11. Percent male population in Bexar County, 2011-2016

Percent male population						
	2011	2012	2013	2014	2015	2016
Minimum	43.6%	42.0%	44.1%	44.5%	43.5%	42.4%
Maximum	91.5%	100.0%	100.0%	100.0%	100.0%	100.0%

Source: US Census Bureau; ACS 5-Year Estimates, Table DP05.

Table 12. Tukey fence outlier methodology

Sample of percent male population estimates, 2011-2016													
Min	Lower fence	Q1	Median	Q3	Upper fence	Max							
42.0%	42.4%	44.1%	44.5%	46.3%	48.0%	49.0%	51.0%	52.5%	54.9%	55.5%	58.0%	71.3%	100.0%
Outliers											Outliers		
Interquartile range: 51.0%-48.0% Tukey Fence: $2.6 * 1.5 = 3.9\%$ Lower Fence: $48.0\% - 3.9 = 44.1\%$ Upper Fence: $51.0\% + 3.9 = 54.9\%$													

Table 13. Equal interval class calculations

Population: Percent Male Equal intervals with outliers (N=73)			Population: Percent Male Revised equal intervals without outliers (N=65)		
Range	# zip codes	% of ZCTAs	Range	# zip codes	% of ZCTAs
42.0-56.5%	67	91.78%	44.1-46.7%	4	5.88%
56.6-71.1%	3	4.11%	46.8-49.4%	36	55.88%
71.2-85.7%	1	1.37%	49.5-52.1%	23	35.29%
85.8-100.0%	2	2.74%	52.2-54.9%	2	2.94%

Source: US Census Bureau; ACS 5-Year Estimates, Table DP05, 2010-2016

Notes on specific indicators

Some of the indicators should be interpreted carefully and the user should understand the definition, exclusions and suppression rules applied in order to understand the whole picture. Below are some of the variables that need to be discussed and the table has some quick notes and tips for what to keep in mind while looking at various indicators.

Birth indicators

The 2010-2013 birth data was collected from the [San Antonio Metropolitan Health District](#) Health Profiles online. The percent of live births following late or no prenatal care was calculated by the number of births with late or no prenatal care in that zip code divided by the total births in that zip code. If the total births for the zip code were less than or equal to 20, then the percent was suppressed on the map. If the numerator (number of births with late or no prenatal care) was less than nine, then it was suppressed to prevent the identification of individuals in confidential data. If a zip code is listed as no data (78243 in 2013) then no birth happened there. The teen birth rate was calculated by the number of births to females aged 15-19 in that zip code divided by the ACS population estimate for females aged 15-19 in that zip code. If the zip code had fewer than 100 females in that age then the rate was suppressed on the map because of the unreliability of the estimate. See Table 14 below for the suppression and calculation rules.

Table 14. Suppression rules for birth data

	% of live births following late or no prenatal care	Rule	Teen birth rate	Rule
Numerator	# births prenatal care began in last trimester or no prenatal care	1-9	# births to females 15-19	1-9
Denominator	Total births	≤20	Estimate of females 15-19	<100
Calculation	Percent		Rate	

Child abuse and neglect indicators

The child abuse and neglect data followed similar guidelines to the birth data suppression rules. For zip codes with fewer than 100 children under 18 years old, those child abuse rates for intakes and confirmed victims were suppressed for reliability and if the numerator was less than five, they were suppressed to protect privacy of the victims (Table 15). The percent of confirmed victims who are re-victimized within five years is not suppressed based on numerator or denominator because of the small counts.

However, the zip codes that had outliers were removed in the same way as all the other indicators. For example, the percent re-victimized in 78256 for the year 2014 was 100% because there was only one case of confirmed child abuse in 2009 and the same child was confirmed again in 2014. Using the Tukey Fence outlier methodology, the revised breaks for equal interval classes put the range from 0% to 42.5% for the 2010-2016 dataset therefore the percent re-victimized for 78256 in 2014 was identified to be an outlier and removed. For zip codes where there were no confirmed victims of child abuse and no re-victimized within a five-year period, those zip codes are labeled as 'No Data' because there is no denominator (confirmed child abuse victims). For zip codes that have 0% re-victimization, then it means there were confirmed child abuse victims in that zip code but none were re-victimized within the five years.

Table 15. Suppression rules for child abuse data

	Initial intakes rate	Rule	Confirmed victims rate	Rule	Percent revictimized	Rule
Numerator	# initial intakes	1-5	# confirmed victims	1-5	# revictimized within 5 years	NA
Denominator	Estimated population under 18	<100	Estimate of population under 18	<100	# confirmed victims	NA
Calculation	Rate per 1,000		Rate		Percent	

Hospitalization indicators

We call them hospitalization rates for short, but these indicators actually reflect hospital *discharges*, not admissions. The hospital discharge data was downloaded from the [Texas Department of State Health Services](#) and the ICD codes that were used for the analysis are listed below (Table 16). There are some important limitations to understand with hospital discharge data.

The rates are determined by hospitalizations *for* the disease as the *primary* diagnosis, not all hospital discharges with that diagnosis. In the case of the asthma hospitalization rate, for example, the intent is to reflect the rate of hospitalizations for an asthma attack, not hospitalizations for heart attacks or car accidents among people who also happen to have diagnosed asthma unrelated to the reason for the hospitalization.

The rates are not prevalence or incidence of the disease. These hospitalization counts are also not unique visits or people. If the same person in 78204 goes to the hospital three times for asthma in 2014 then all three visits are included if asthma was the primary diagnosis for the admission during that year. The 3-year rate for 2016 is calculated from the average of the hospitalization counts for 2014-2016 divided by that population estimate in 2016 times 10,000 (see Table 17 below). The rates were stabilized using a three-year average and outliers were excluded from the analysis.

Because the San Antonio Military Health System does not report their hospitalizations to DSHS, the public data files exclude any federal hospital admissions. Because the military hospital systems account for a large portion of our population, the Bexar County hospitalization data should not be compared to other major cities who do not have large federal hospital exclusions in their datasets.

The PUDF [user notes](#) have directions on what is suppressed and excluded in the file. Below are some common ones that affect the data:

- The last two digits of the patient's ZIP code are suppressed if there are fewer than thirty patients included in the ZIP code.
- The entire ZIP code is suppressed if a hospital has fewer than fifty discharges in a quarter.
- The entire ZIP code and gender code are suppressed if the ICD-9-CM code indicates alcohol or drug use or an HIV diagnosis.
- The entire ZIP code is suppressed if a hospital has fewer than five discharges of a particular gender, including 'unknown'.

Table 16. International Classification of Diseases used in analysis

	ICD-9	ICD-10
Asthma	493	J45
Diabetes	250	E10-11
Hypertension	401-405	I11-I15
Injury	800-959	S00-S99, T07, T14-T32
Mental health	290-302 & 306-319	F01-09 & F20-F99

Table 17. Asthma hospitalization calculations for 78204 for population under 18 years old, 2016

		2014	2015	2016	3-year rate
Numerator	Hospitalization counts	3	1	6	3.3
Denominator	Estimate of population under 18 years old	2,690	2,791	2,435	2,435
Calculation	Rate per 10,000	11.2	3.6	24.6	13.7

Quick guide to interpreting indicators

Table 18. Notes for interpreting indicators

Indicator	Notes and tips for understanding the data
Population	Helpful to reference this when looking at any American Community Survey (ACS) or other survey percentages.
Population by sex	Military zip codes were excluded as outliers in some years. Refer to military section above.
Race & ethnicity	"Hispanic or Latino origin" is not mutually exclusive of other races/ethnicities listed.
Educational attainment	This is the percentage of the population 25 years and older.
Median income	Household income includes income of the householder and all other people 15 years and older in the household, whether or not they are related to the householder.
Income inequality	The GINI index varies from 0 to 1 where 0 indicates perfect equality (proportional distribution of income) and 1 indicates perfect inequality (one household has all the income).
Population below 100% poverty	The ACS poverty population estimates only include those people for whom poverty status can be determined.
Labor force and unemployment	Military zip codes were excluded as outliers in some years. Refer to military section above.
Military	The percent of people in the armed forces include people 16 years and older. The veteran data is based on the percent of people 18 years and older who previously served on active duty in the U.S. Army, Navy, Air Force, Marine Corps or the Coast Guard. Both military indicators are masked for the military ZCTAs due to outliers.
Multi-family units	Multi-family is defined as two or more units per building.
Housing affordability	Housing affordability is defined where the housing costs greater than or equal to 30% of the household income.
Vacant housing units	Vacancy status is used as a basic indicator of the housing market and provides information on the stability and quality of housing for certain areas.
Geographic mobility	Military zip codes were excluded as outliers in some years. Refer to military section above.

Table 18. Notes for interpreting indicators, *continued*

Indicator	Notes and tips for understanding the data
Child Abuse and neglect	See section on suppression rules and child abuse.
Family violence rate	Suppression if census population was <100 to calculate the rate.
Uninsured	Uninsured includes private or public health insurance coverage.
Births	Birth data is published by residence of the mother at delivery. See section on suppression rules.
Asthma hospitalizations	See hospitalization section.
Diabetes hospitalizations	See hospitalization section.
Hypertension hospitalizations	See hospitalization section.
Injury hospitalizations	See hospitalization section.
Mental health hospitalizations	See hospitalization section.
Mean age at death	Mean age of decedent by decedent's zip code.