Violence Risk Factors Map

A map of neighborhood risk factors for youth and gang violence in San Jose.

Parks, Recreation, and Neighborhood Services Data Story

Albert Gehami, Data Scientist

April 5, 2019

Table of Contents

Introduction

San Jose's juvenile violent crime has risen each year since 2015 (Figure 1). Not only is the City concerned with the rising crime rate, but it is also concerned with failing its youth. The Mayor's Gang Prevention Task Force under the department of Parks, Recreation, and Neighborhood Services has been tasked with reducing youth delinquency through supportive services. Ideally, the Task Force can reach the youth and their communities before any crime is committed.

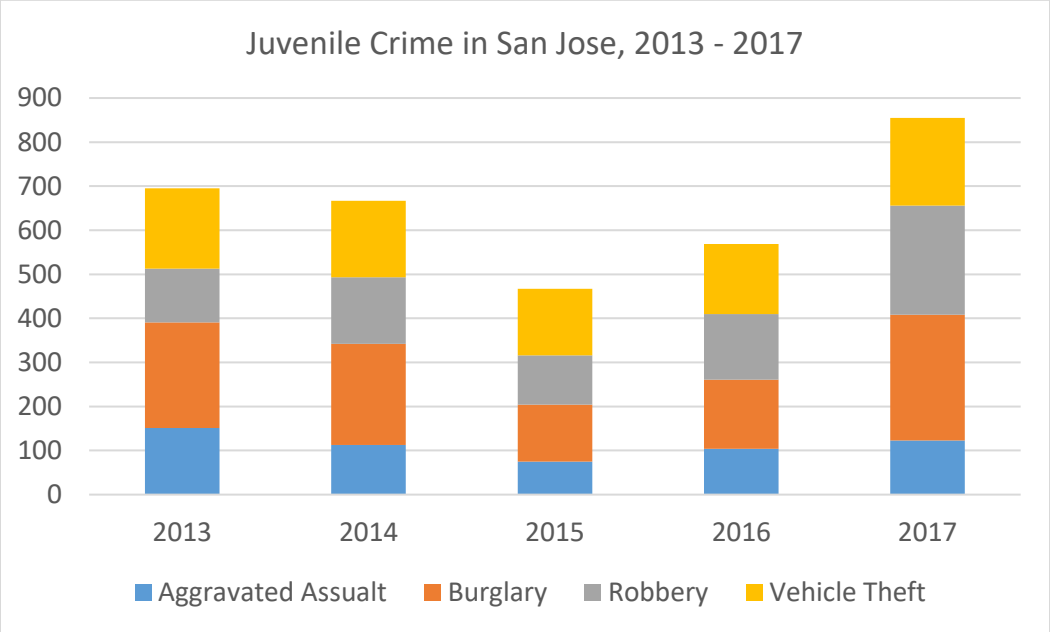


Figure 1: Juvenile Crime in San Jose over time. Data provided by San Jose Police Department Crime Investigations Unit

Identifying the communities at the highest risk of youth violence is hard. Over 200 neighborhoods inhabit San Jose, each with a reason to demand services. To provide a holistic, objective study of the City, the Mayor's Gang Prevention Task Force hired a Data Science fellow to identify the neighborhoods with the highest risk for youth and gang violence.

The final product was a set of interactive tools for identifying each neighborhood's risk for youth and gang violence (Figure 2). Each neighborhood holds an overall risk score, but also factor-specific risk scores. Factor-specific scores identify which neighborhoods need help with specific issues, like preventing substance abuse or gang crime. Using machine learning, the overall risk scores for each neighborhood are predicted out to 2021, to help plan PRNS's future allocation of services. Through data, the task force can identify the communities with the highest risk for youth and gang violence to change the narrative of trauma and crime to support and success.

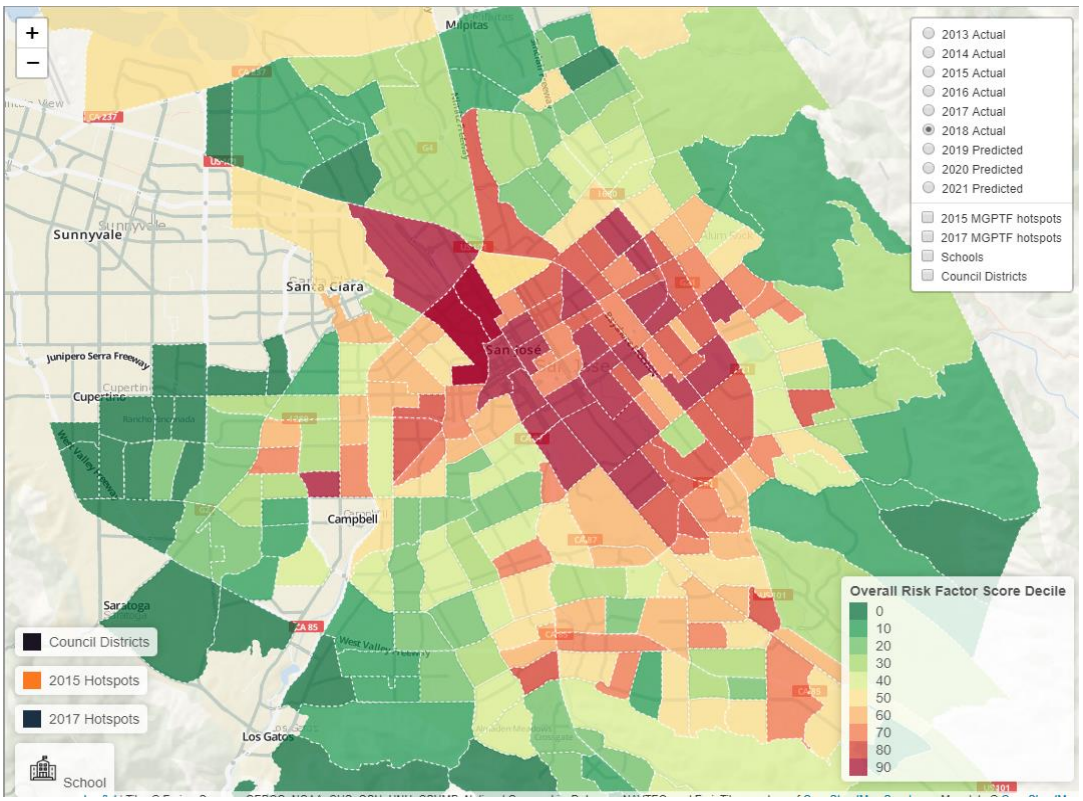


Figure 2: Map of Youth and Gang Violence Risk Factors for each Census Tract (loosely each neighborhood)

The Violence Risk Factors Map (aka "Risk Factors Map" and commonly referred to in this document as "The Map") is an interactive tool which can be found here: http://gehami.com/violence-risk-factors-map/. On the map any neighborhood can be clicked on to see its overall Risk Factor Score, along with the factor-specific risk scores (Figure 3). To alter the map and explore only a single, or small set of risk factors, that can be done on the toggle map here: https://albertgehami.shinyapps.io/Violence-Risk-Factors-Map/ (Figure 4). The map can also include an overlay of schools, city council districts, and the Mayor's Gang Prevention Task Force "Hot Spots" by selecting any of the items on the top right (Figure 5). Furthermore, the displayed year can be changed to explore past, present, and future neighborhood risk factors for youth and gang violence.

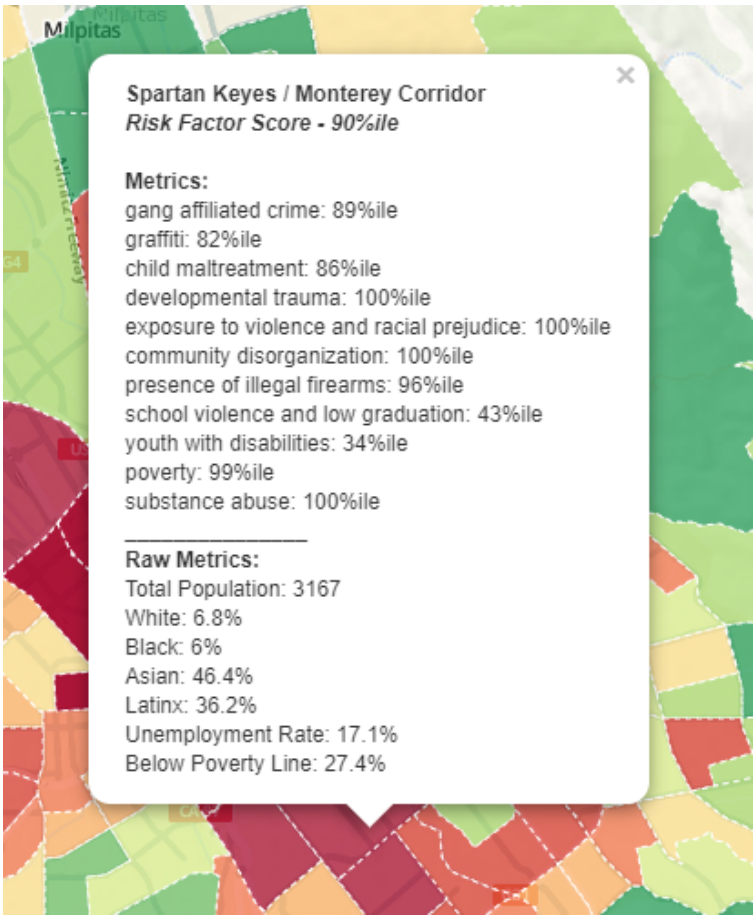


Figure 3: A detailed listing of a neighborhood's overall Risk Factor Score, along with the scores of each of the risk factor metrics.

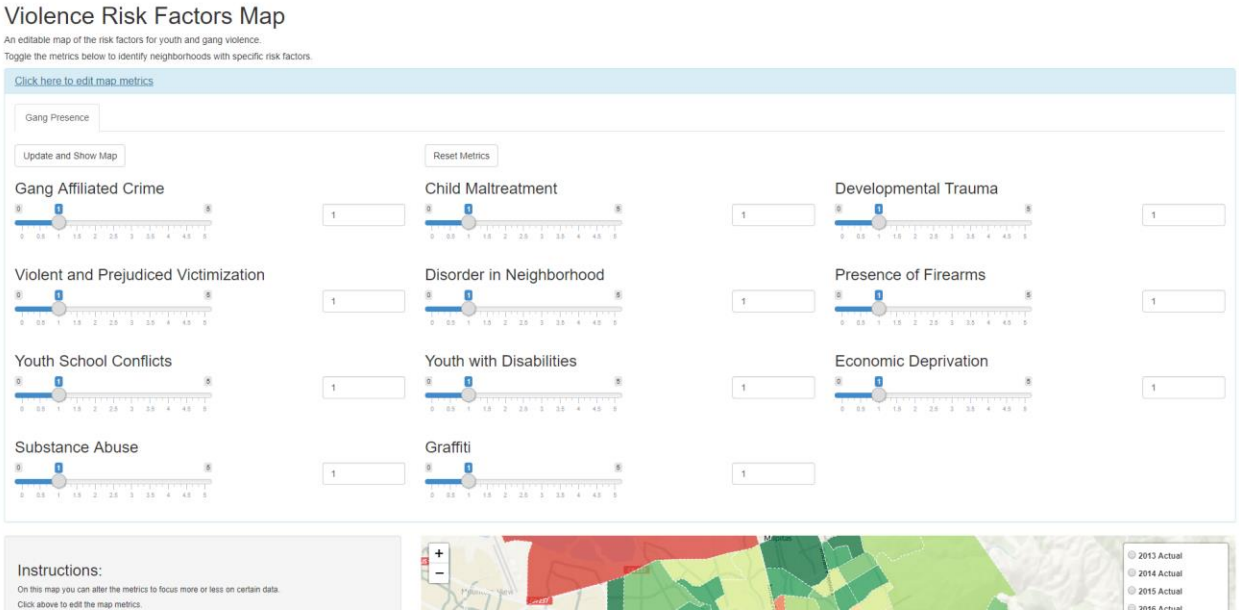


Figure 4: These toggles can be altered to re-weight the risk factors. By setting all but one factor to zero, the map will show only risk associated with that factor.

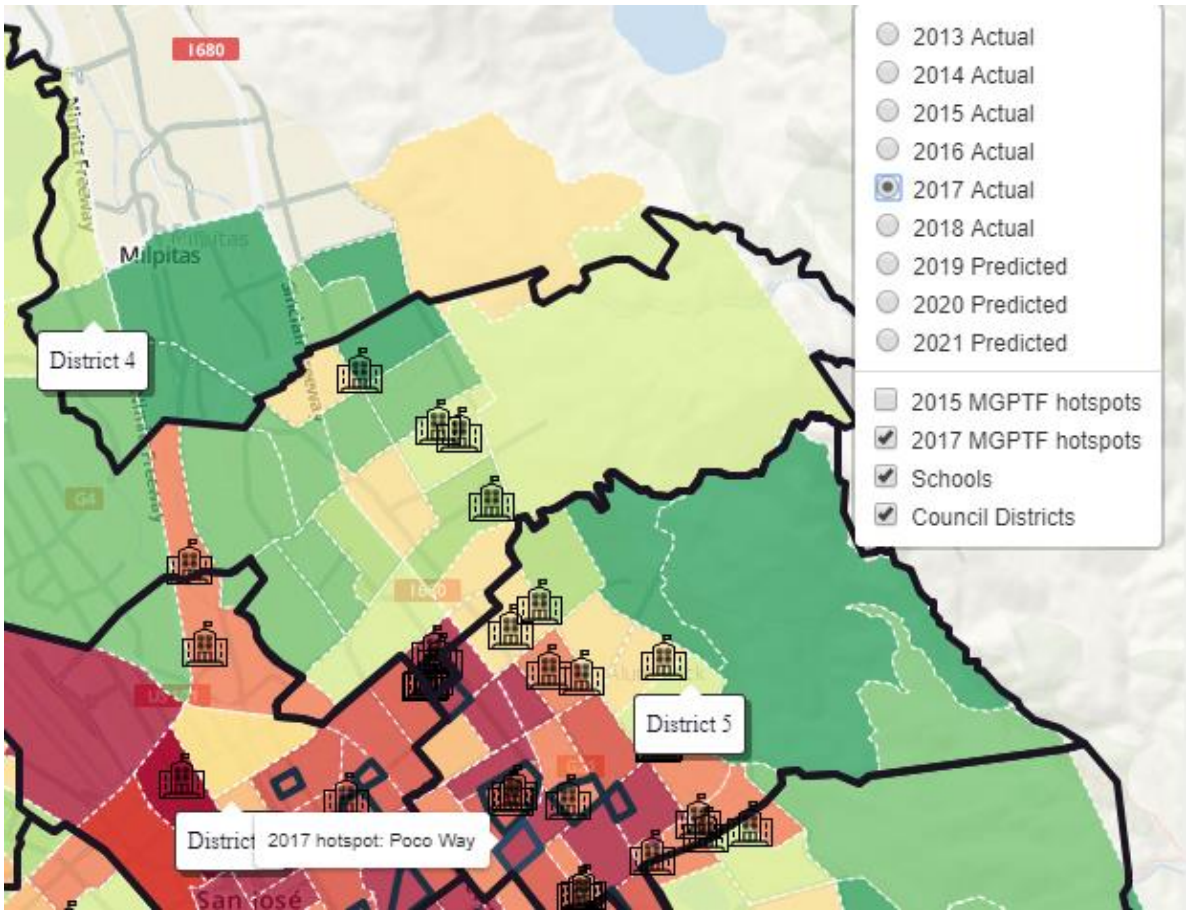


Figure 5: Showing the possible overlays on the map. Here the council district outlines are shown, along with locations of schools and Mayor's Gang Prevention Task Force hot spots.

This report outlines in detail the data science work done by PRNS. Section 2 covers how PRNS chose the metrics to define a neighborhood's risk for youth and gang violence. Section 3 covers the data elements used for each metric, their sources, and how risk scores are calculated. Section 4 covers the methodology behind predicting future risk scores, and evaluates the predictive model. Section 5 summarizes the report, explains limitations, and suggests avenues for building on this work.

Metrics for Youth and Gang Violence

The US Department of Justice Office of Juvenile Delinquency and Prevention (OJJDP) offers a comprehensive list of risk factors for youth and gang violence.[1,2] These risk factors include the presence of youth poverty in the community, child abuse, and other factors enumerated below. One factor which is not from the OJJDP but is included is the presence of graffiti in the neighborhood, which serves as a proxy for some of the OJJDP risk factors associated with a presence of gangs in the community. Ultimately, eleven metrics were used to define the risk for youth and gang violence. Below, each metric is discussed.

Eleven Metrics:

1) Child Maltreatment

Studies have examined three forms of child maltreatment: physical abuse, sexual abuse, and neglect. Evidence suggests that children who have been physically abused or neglected are more likely than others to commit violent crimes (Widom, 1989;[3] Zingraff et al., 1993;[4] Smith and Thornberry, 1995).[5]

2) Developmental Trauma Exposure

Developmental trauma, such as abandonment, betrayal, physical or sexual assaults or witnessing domestic violence have consistent and predictable consequences that affect many areas of functioning. These children tend to behaviorally reenact their traumas either as perpetrators, in aggressive or sexual acting out against other children, or in frozen avoidance reactions.[6]

3) Exposure to violence and racial prejudice

Exposure to violence in the home and elsewhere increases a child's risk for involvement in violent behavior later in life (Paschall, 1996)[7]. McCord and Ensminger (1995)[8] also found that study participants who reported having experienced racial discrimination committed more violent acts.

---

[1] https://www.ncjrs.gov/pdffiles1/ojjdp/179065.pdf
[2] https://www.nationalgangcenter.gov/spt/Risk-Factors/12
[3] Widom, Cathy Spatz. "Child abuse, neglect, and adult behavior: Research design and findings on criminality, violence, and child abuse." *American journal of Orthopsychiatry* 59.3 (1989): 355-367.
[4] Zingraff, Matthew T., et al. "Child maltreatment and youthful problem behavior." *Criminology* 31.2 (1993): 173-202.
[5] Smith, Carolyn, and Terence P. Thornberry. "The relationship between childhood maltreatment and adolescent involvement in delinquency." *Criminology* 33.4 (1995): 451-481.
[6] Van der Kolk, Bessel A. "Developmental Trauma Disorder: Toward a rational diagnosis for children with complex trauma histories." Psychiatric annals35.5 (2017): 401-408.
[7] Paschall, Mallie J., Susan T. Ennett, and Robert L. Flewelling. "Relationships among family characteristics and violent behavior by black and white male adolescents." *Journal of Youth and Adolescence* 25.2 (1996): 177-197.
[8] McCord, Joan, and Margaret Ensminger. "Pathways from aggressive childhood to criminality." *meeting of the American Society of Criminology, Boston, MA*. 1995.

4) Community disorganization.

Community disorganization and low neighborhood attachment are predictors of violence. Community disorganization, associated with the presence of crime and poverty, was a better predictor of violence than low attachment to a neighborhood (Maguin et al, 1995)[9].

5) Presence of Illegal Firearms

A prevalence of drugs and firearms in the community predicts greater variety in violent behaviors at age 18 (Maguin et al., 1995).[10] Furthermore, an availability of illegal firearms, both for adults and children, is associated with higher gang involvement.[11]

6) School Violence and Academic Failure

Poor academic achievement has consistently predicted later delinquency (Maguin and Loeber, 1996;[12] Denno, 1990).[13] Farrington (1989) found that boys who at age 11 attended schools with high violence reported more violent behavior than other youth.[14]

7) Poverty

Being raised in poverty has been found to increase the likelihood of involvement in crime and violence (Sampson and Lauritsen, 1994).[15] Self-reported felony assault and robbery have been found to be twice as common among youth living in poverty as among middleclass youth (Elliott, Huizinga, and Menard, 1989).[16] Low family income predicted self-reported teen violence and convictions for violent offenses in several studies (Farrington, 1989;[17] Hogh and Wolf, 1983;[18] Henry et al., 1996).[19]

---

[9] Maguin, Eugene, et al. "Risk factors measured at three ages for violence at age 17–18." *American Society of Criminology*(1995).
[10] *Ibid 9*
[11] National Gang Center. "Gang Risk Factors Ages 12–17." *Risk FactorsAges 12–17*, Office of Juvenile Justice and Delinquency Prevention, 2018, www.nationalgangcenter.gov/spt/Risk-Factors/12.
[12] Maguin, Eugene, and Rolf Loeber. "Academic performance and delinquency." *Crime and justice* 20 (1996): 145-264.
[13] Denno, Deborah W. *Biology and violence: From birth to adulthood*. Cambridge University Press, 1990.
[14] Farrington, David P. "Early predictors of adolescent aggression and adult violence." *Violence and victims* 4.2 (1989): 79-100.
[15] Sampson, Robert J., and Janet L. Lauritsen. "Violent victimization and offending: Individual-, situational-, and community-level risk factors." *Understanding and preventing violence* 3 (1994).
[16] Huizinga, David H., Scott Menard, and Delbert S. Elliott. "Delinquency and drug use: Temporal and developmental patterns." *Justice quarterly* 6.3 (1989): 419-455.
[17] Farrington, David P. "Early predictors of adolescent aggression and adult violence." *Violence and victims* 4.2 (1989): 79-100.
[18] Høgh, Erik, and Preben Wolf. "Violent crime in a birth cohort: Copenhagen 1953–1977." *Prospective studies of crime and delinquency*. Springer, Dordrecht, 1983. 249-267.
[19] Barnes, Barry, David Bloor, and John Henry. *Scientific knowledge: A sociological analysis*. University of Chicago Press, 1996.

8) Substance Abuse

As stated previously, a prevalence of drugs and firearms in the community predicted greater variety in violent behaviors at age 18.[20] Furthermore, high alcohol and drug use, drug dealing, and other drug-related activities are associated with a higher risk for youth gang violence.[21]

9) Youth with Disabilities

Mental and emotional health problems, such as Attention Deficit Hyperactivity Disorder (A.D.H.D), have been associated with future gang and violent activity. [22] While no data is available on specifically youth with mental disabilities, the American Community Survey holds data on youth with any disabilities, including mental and emotional disabilities.

10) Gang Affiliated Crime in Community

Gang-affiliated crime directly reflects the amount of gang activity within a neighborhood. While crime reporting rates of certain neighborhoods may be lower than others,[23] this is a strong reflection of gang violence in a community.

11) Graffiti within the Community

Graffiti is a strong metric for the presence of a gang within a community. According to the Los Angeles Police Department:

> "The purpose of gang graffiti is to glorify the gang. Gang graffiti is meant to create a sense of intimidation and may increase the sense of fear within a neighborhood. Gang members use graffiti to mark their territory or turf, declare their allegiance to the gang, advertise a gang's status or power, and to challenge rivals. Graffiti is used to communicate messages between gangs using codes with common meaning."[24]

Calculations of Metrics, Data, and Sources

The overall rating for each neighborhood's risk factor score comes from an equal combination of the eleven metrics from the US Department of Justice Office of Juvenile Justice and Delinquency

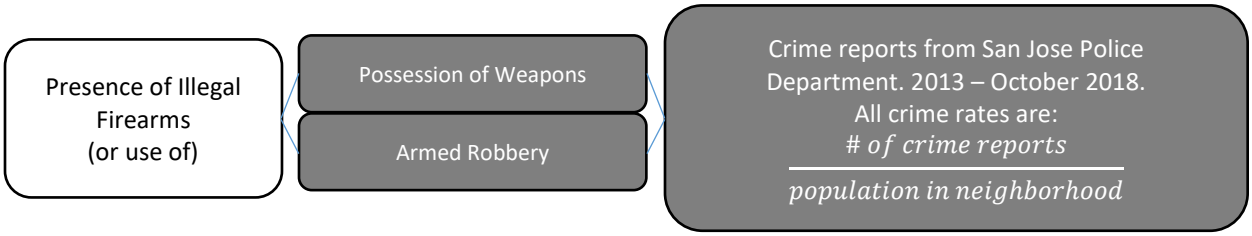
---

[20] *Ibid 10.*
[21] *Ibid 12.*
[22] *Ibid 12.*
[23] Myers, Samuel L. "Why are Crimes Underreported? What is the Crime Rate? Does it" Really" Matter?." *Social Science Quarterly* 61.1 (1980): 23-43.
[24] Why Gang Graffiti Is Dangerous." *Los Angeles Police Department Badge*, 2010, www.lapdonline.org/top_ten_most_wanted_gang_members/content_basic_view/23471.

Prevention (OJJDP). Those eleven metrics, the data behind each metric, and the source for each piece of data, can be found in Table 1. Below is an example of calculating a single metric to show how all components come together to give a neighborhood its overall risk factor score.

Example Metric – Presence of Illegal Firearms (or use of)

In the chart below, the OJJDP metric "Presence of Illegal Firearms (or use of)" is linked to the following information.



From left to right, the first block shows the metric of focus (Presence and Use of Illegal Firearms). The next set of blocks is the data used to measure this metric, and the final block is the source of the data, including any calculations done before measuring.

In this case, the data creating the "Presence of Illegal Firearms" metric is Possession of Weapons crimes and Armed Robbery Crimes. Both data blocks are sourced from San Jose Police Department crime reports. Lines connect one block to another, and each block of data shares the same color shading with its source (see above example).

Each data block's value is calculated through the calculation noted in the source block. The value for any block of data which comes from crime reports is calculated via the number of crime reports divided by the population in the neighborhood. For "Possession of Illegal Firearms," the number of possession of weapons crimes are counted for each neighborhood, and then divided by the population in each neighborhood. So the data would look like this:

Neighborhood	# of "Possession of Illegal Firearms" Crimes	Population	Crimes per Population
A	100	1000	0.1
B	250	5000	0.05
C	300	2000	0.15

We do a similar process for Robbery crimes:

Neighborhood	# of "Armed Robbery" Crimes	Population	Crimes per Population
A	40	1000	0.04
B	25	5000	0.005
C	20	2000	0.01

By dividing the number of crimes by population, crime rates can be compared across neighborhoods of varying sizes. Calculating the crime rates for both crimes results in the following table:

Neighborhood	"Possession of Illegal Firearms" Crimes per population	"Armed Robbery" Crimes per population
A	0.1	0.04
B	0.05	0.005
C	0.15	0.01

These two data measurements are then combined to determine the overall metric score for "Presence or Use of Illegal Firearms" for each neighborhood. Adding the two crime rates together would imply that the two crimes are comparable, but they are on a different scale. In general, there are more illegal firearm charges than there are robbery charges, so the Armed Robbery crime rate has an unfairly lower impact on the overall metric (notice how adding the two crime rates leads to numbers very close to the illegal firearms crime rate, but very different from the armed robbery crime rate).

Before combining the data blocks, they must be put on the same scale. The largest value of each data block (column) is 1, and the lowest number is 0. All other values will fall between 0 and 1. For the Illegal Firearms column, the max value is 0.15, and the minimum value is 0.05. Therefore, the Illegal Firearms score would look like the following:

Neighborhood	"Possession of Illegal Firearms" Crimes per population	"Possession of Illegal Firearms" Score
A	0.1	?
B	0.05	0
C	0.15	1

Calculating all the values in between the minimum and maximum value (in this case, the score for neighborhood A), uses the following equation:

$$Neighborhood\ Score = \frac{Neighborhood\ value - Min\ value}{Max\ Value - Min\ Value}$$

For neighborhood A:

$$Neighborhood\ A's\ Illegal\ Firearm\ Score = \frac{0.1 - 0.05}{0.15 - 0.05} = 0.5$$

Formally, this process is called "Min-Max Scaling" and puts every data block onto a comparable scale between 0 and 1. After Min-Max Scaling both crime rates, the two scores can be added together to produce the final "Presence and Use of Illegal Firearms" metric score for each neighborhood.

Neighborhood	"Possession of Illegal Firearms" Score	"Armed Robbery" Score	"Presence and Use of Illegal Firearms" Metric Score
A	0.5	1	1.5
B	0	0	0
C	1	0.142857143	1.142857143

With two exceptions, this is how the score for all metrics are determined. One exception is the "Substance Abuse" metric, which multiplies the "Narcotics" score (the score for all reported Narcotics crimes in a neighborhood) by five before adding it to the other scores in the "Substance Abuse" metric. Narcotics crimes encompasses many different crimes related to substance abuse, including buying, selling, and possession of illegal substances, so it should be weighted higher than the individual crime rates measured by other data blocks in the "Substance Abuse" metric.

The other exception is that the Safe School Campus Initiative Incident Reports score is multiplied by 0.5 (cut in half) because most acts of school violence are not reported to Safe School Campus Initiative, so it is an incomplete data source.[25]

Combining Metrics to get Overall Risk Factor Score

The metric scores are on different scales. For example, "Presence and Use of Illegal Firearms" only has two data measurements, meaning that metric is on a scale from 0 to 2. Meanwhile "Child Maltreatment" has six data measurements, meaning that metric is on a scale from 0 to 6. To combine multiple metrics, the

[25] Gehami, Albert. "Interview with East Side Union Superintendent Chris Funk." 17 Oct. 2018.

same Min-Max Scaling process is used as was used combining blocks of data within a metric. For example, take this table of three OJJDP metrics:

Neighborhood	"Presence and Use of Illegal Firearms" Metric Score	"Child Maltreatment" Metric Score	"Gang Affiliated Crime" Metric Score
A	1.5	4.5	0
B	0	0	0.2
C	1.142857143	6	1

Min-Max scaling each metric will result in this final table:

Neighborhood	"Presence and Use of Illegal Firearms" MinMax Score	"Child Maltreatment" MinMax Score	"Gang Affiliated Crime" MinMax Score	Overall Neighborhood Risk Factor Score
A	1	0.75	0	1.75
B	0	0	0.2	0.2
C	0.761904762	1	1	2.761904762

The "Overall Neighborhood Risk Factor Score" is the sum of all metric scores, and the final score used to assess the overall risk level for each neighborhood. All eleven metrics (listed on the following page) are treated as equal risk factors for youth and gang violence. Equal weighting for all metrics may be useful for some, but may not be useful for others. To customize how much each metric matters, the interactive toggle site allows users to customize the metrics of interest, and how they are weighted.[26] With the risk scores determined, the youth and gang violence risk scores can be mapped for each neighborhood.

Metrics displayed on the map:

The map shows not the actual overall Risk Factor Score, but the risk score decile of each neighborhood. In other words, 10% of neighborhoods fall into the 90% decile, meaning their overall Risk Factor Scores are above 90% of the scores of other neighborhoods. There will be another 10% of neighborhoods which fall into the 80% decile, and so on down to the 0% decile. We choose to display the decile on the map rather than the raw Risk Factor Scores because we do not intend for this map to be used

[26] https://albertgehami.shinyapps.io/Violence-Risk-Factors-Map/

for exact comparisons between neighborhoods, but rather for a general comparison to identify which neighborhoods clearly have more present risk factors than others.

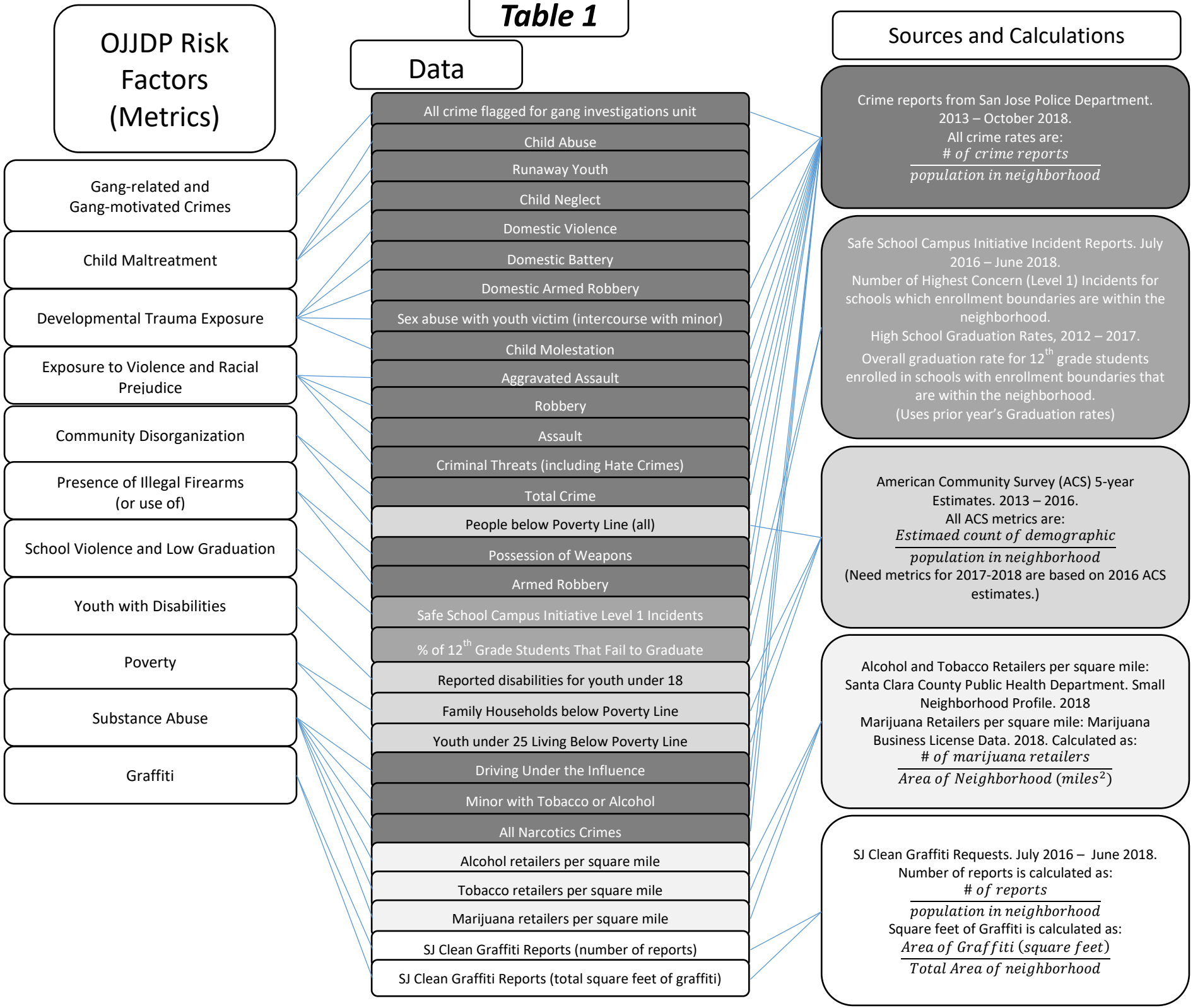
Table 1 displays all US Department of Justice Office of Juvenile Justice and Delinquency Prevention (OJJDP) metrics used, followed by the data measurements (data blocks) used for each metric, and the source of each data measurement.

A Brief note on Yearly Data

Since the map offers an annual look at San Jose, we divide our data into yearly fractions. This means the 2018 map only uses crime data and other data from 2018. There are some complications and exceptions that are listed below:

1) American Community Survey Estimates currently end in 2017, so the demographic data is constant from 2017 to 2018.
2) Safe School Campus Initiative and SJ Clean Graffiti Request data is only available from 2016 – 2018. Because of this, the "Graffiti" metric will be 0 for all years prior to 2016, and the "School Violence and Low Graduation" metric will only include graduation rates for years before 2016.
3) Santa Clara County Health Department and Marijuana License Data is point-in-time data, and remains constant through the years.
4) Graduation Rates are taken from the prior year. This means that the 2018 map uses graduation data from 2017, the 2017 map uses graduation data from 2016, and so on.

Table 1


OJJDP Risk Factors (Metrics)
Gang-related and Gang-motivated Crimes
Child Maltreatment
Developmental Trauma Exposure
Exposure to Violence and Racial Prejudice
Community Disorganization
Presence of Illegal Firearms (or use of)
School Violence and Low Graduation
Youth with Disabilities
Poverty
Substance Abuse
Graffiti
Data
All crime flagged for gang investigations unit
Child Abuse
Runaway Youth
Child Neglect
Domestic Violence
Domestic Battery
Domestic Armed Robbery
Sex abuse with youth victim (intercourse with minor)
Child Molestation
Aggravated Assault
Robbery
Assault
Criminal Threats (including Hate Crimes)
Total Crime
People below Poverty Line (all)
Possession of Weapons
Armed Robbery
Safe School Campus Initiative Level 1 Incidents
% of 12th Grade Students That Fail to Graduate
Reported disabilities for youth under 18
Family Households below Poverty Line
Youth under 25 Living Below Poverty Line
Driving Under the Influence
Minor with Tobacco or Alcohol
All Narcotics Crimes
Alcohol retailers per square mile
Tobacco retailers per square mile
Marijuana retailers per square mile
SJ Clean Graffiti Reports (number of reports)
SJ Clean Graffiti Reports (total square feet of graffiti)
Sources and Calculations
Crime reports from San Jose Police Department. 2013 – October 2018.
All crime rates are:
# of crime reports / population in neighborhood
Safe School Campus Initiative Incident Reports. July 2016 – June 2018.
Number of Highest Concern (Level 1) Incidents for schools which enrollment boundaries are within the neighborhood.
High School Graduation Rates, 2012 – 2017.
Overall graduation rate for 12th grade students enrolled in schools with enrollment boundaries that are within the neighborhood.
(Uses prior year's Graduation rates)
American Community Survey (ACS) 5-year Estimates. 2013 – 2016.
All ACS metrics are:
Estimaed count of demographic / population in neighborhood
(Need metrics for 2017-2018 are based on 2016 ACS estimates.)
Alcohol and Tobacco Retailers per square mile: Santa Clara County Public Health Department. Small Neighborhood Profile. 2018
Marijuana Retailers per square mile: Marijuana Business License Data. 2018. Calculated as:
# of marijuana retailers / Area of Neighborhood (miles²)
SJ Clean Graffiti Requests. July 2016 – June 2018.
Number of reports is calculated as:
# of reports / population in neighborhood
Square feet of Graffiti is calculated as:
Area of Graffiti (square feet) / Total Area of neighborhood

Scope of Data – Defining a “Neighborhood”

City-defined neighborhoods are not used for this map. Instead, all data is stored and displayed at the census tract level. The "Census Tract" is an area roughly equivalent to a neighborhood established by the Bureau of Census for analyzing populations. They generally encompass a population between 2,500 to 8,000 people. Some neighborhoods, as defined and labeled by the Santa Clara County department of Public Health, include multiple census tracts, and others only include one tract. In practice, some neighborhood communities may cross census tracts.

Here is a map of all census tracts in San Jose:

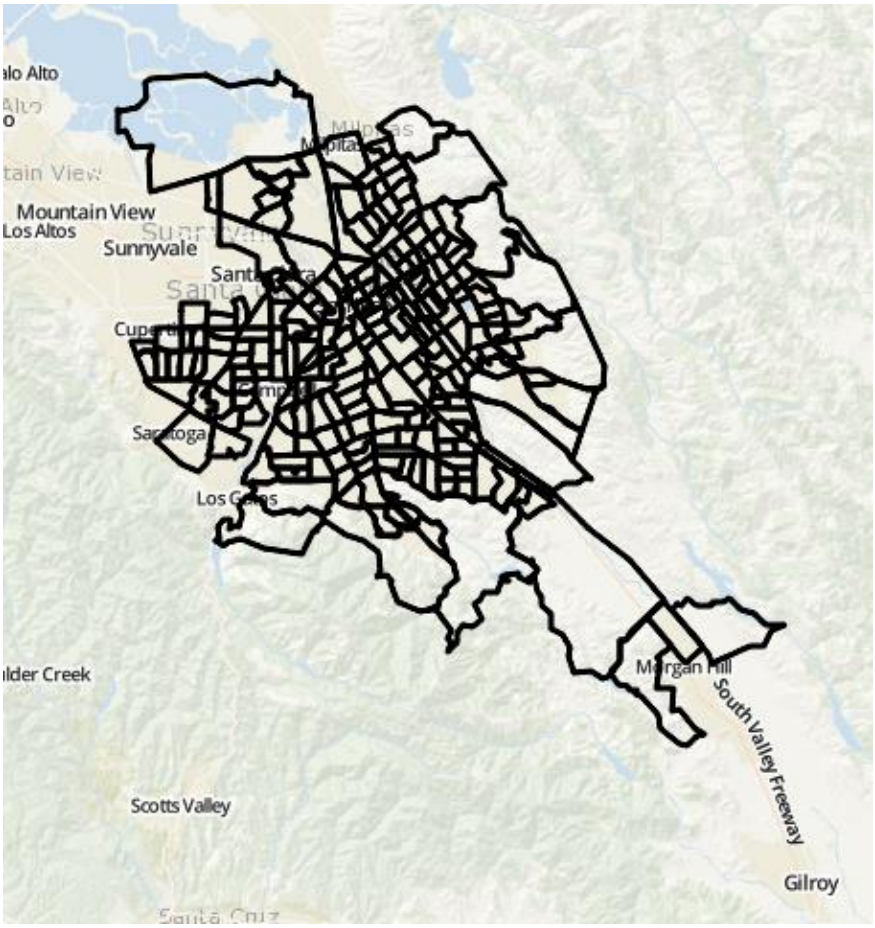


Figure 6: Map of all census tracts in San Jose

Every census tract has its own crime rate, poverty levels, and metrics. The calculations shown in the prior section are used to determine the overall Risk Factor Score for each census tract for each year between 2013 and 2018.

No data is used at the individual level. This means it is impossible to identify any individual, family, or group of individuals through this data. All data is tied to the geographic area and nothing else. This accomplishes our goal of better resource allocation without compromising our citizens' privacy.

Predicting Future Risk Factor Scores

The predictive models use data from 2018 and before to predict the overall risk factor scores for each neighborhood for 2019, 2020, and 2021. The models predict a neighborhood's future overall risk factor scores using the neighborhood's current eleven metric scores, its overall risk score, and the average of the eleven metric scores for the neighborhood's bordering neighborhoods, and the average of their overall risk score.

To illustrate this model, assume the overall risk factor score is not determined by 11 metrics, but instead by 2 metrics: Substance Abuse and Poverty. Furthermore, assume we are trying to predict the overall risk factor score for neighborhood A, with neighborhoods B and C bordering (or sharing a border) with A.

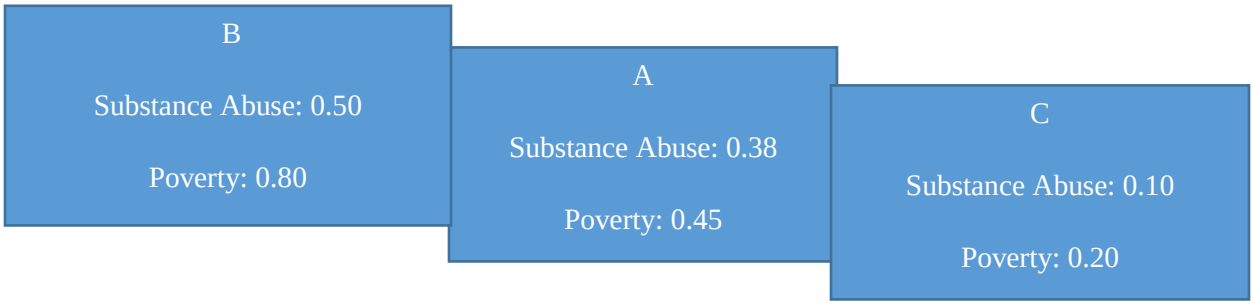


Figure 7: Example of Census Tract and neighboring tracts

Based on this model, the values used to predict neighborhood A's future overall risk score would be:

1) A's Overall Risk Factor Score = $0.38 + 0.45 = 0.83$
2) A's Substance Abuse metric = 0.38
3) A's Poverty metric = 0.45
4) A's neighbor average Overall Risk Factor Score =

$$\frac{B's\ Overall\ score + C's\ Overall\ score}{\#\ of\ bordering\ \ neighborhoods\ A\ has} = \frac{(0.50 + 0.80) + (0.10 + 0.20)}{2} = 0.80$$

5) A's neighbor average Substance Abuse metric =

$$\frac{B's\ Substance\ Abuse\ score + C's\ Substance\ Abuse\ score}{\#\ of\ bordering\ \ neighborhoods\ A\ has} = \frac{0.50 + 0.10}{2} = 0.30$$

6) A's neighbor average Poverty metric =

$$\frac{B's\ Poverty\ score + C's\ Poverty\ score}{\#\ of\ bordering\ neighborhoods\ A\ has} = \frac{0.80 + 0.20}{2} = 0.50$$

Under this simplified model, these six values would be used to predict A's overall risk factor score 1 year, 2 years, and 3 years into the future. In the full model, 24 values (the 11 metrics + one overall score from the neighborhood and another 11 + one overall score which average its bordering neighborhood's metrics) are used to predict a neighborhood's overall risk score 1, 2, and 3 years into the future.

The model used is referred to as a "Random Forest" algorithm, implemented in R. A random forest algorithm uses many decision tree algorithms, and has each tree offer a prediction. The predictions are then averaged across all trees in the forest, creating the final predicted value.[27] A visual representation of the random forest model is shown below (Figure 6). For a deeper understanding of decision trees and a random forest model, check footnote 26. To see how these algorithms are implemented in our work, check our GitHub files "Building the 1 2 and 3 year prediction models.R" and "Predicting the future.R."[28]

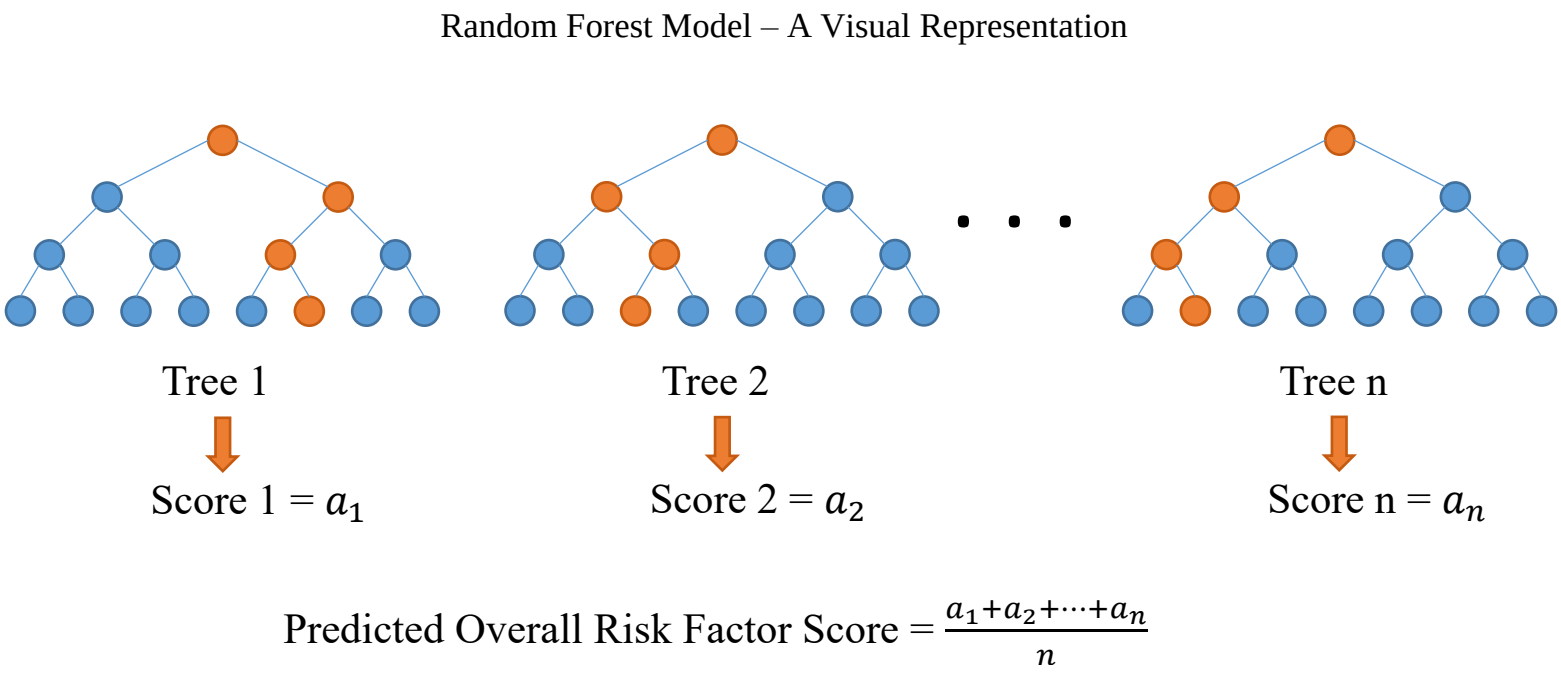


Figure 8: A Visual Representation of the Random Forest Model used to predict the Overall Risk Factor Scores.

[27] For more information on Random Forest algorithms and Decision trees, see https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d for a written discussion, https://www.youtube.com/watch?v=9TiezQ7Gb3M for a video, or http://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learing/ for a more detailed overview.

[28] https://github.com/gehami/Youth-Needs-Map

Prediction Accuracy

By predicting historical data (ie. Predicting the risk factor scores of 2017 using the data of 2016), the model's accuracy can be assessed. In general, the model is 99% accurate in predicting 1, 2, and 3 years in advance. For example, below is a comparison of the actual 2018 map, and the predicted 2018 map, using 2017 data (Figure 9 and Figure 10).

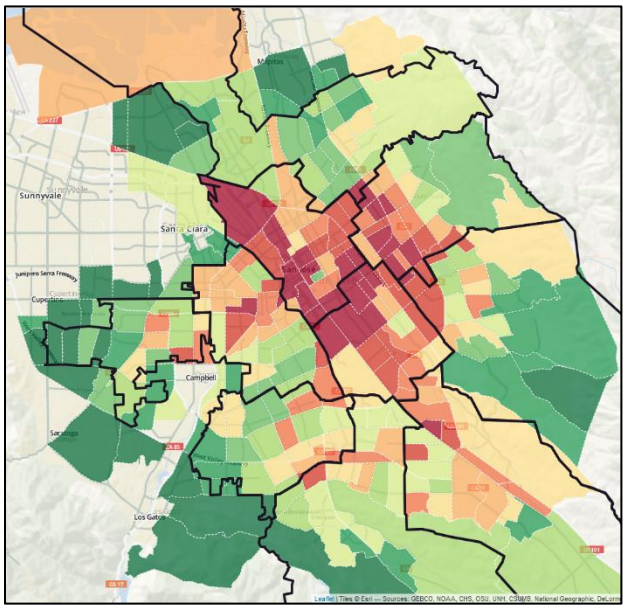


Figure 9: Actual 2018 Risk Factor Map. Council Districts mapped in black for reference.

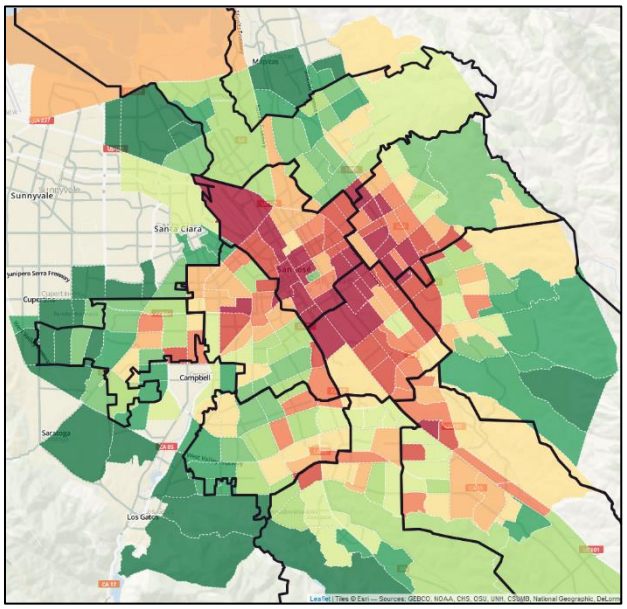


Figure 10: Predicted 2018 Risk Factor Map Using 2017 Data. Council Districts mapped in black for reference.

Such an accurate prediction is due in large part to the low variation year-over-year. The demographics, crime rates, and other data remains relatively constant each year. To illustrate this point, below is a comparison of the risk factor maps from 2013 to 2018 (Figure 9). The consistency over time confirms a common sentiment among City employees: The troubled areas today were the troubled areas a decade ago.

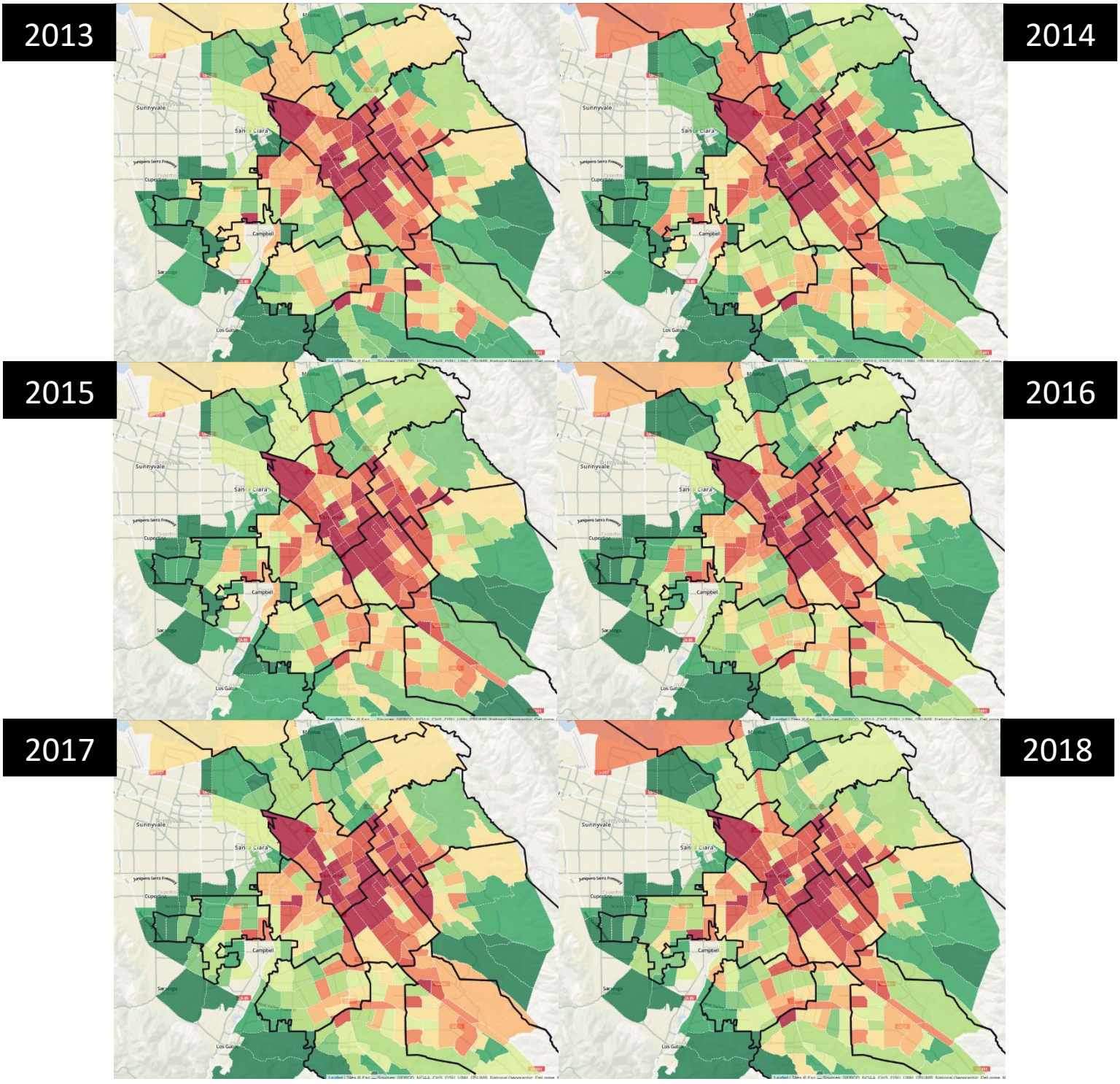


Figure 11: Map of Neighborhood Risk Factor Score (the Risk Factor Map) over time from 2013 - 2018. Council Districts mapped in black for reference.

Conclusion, Limitations, and Future Work

The Violence Risk Factors Map is a prototype for the City, its service providers, and the public to understand the risk factors for youth and gang violence within each neighborhood. There are a number additional data sources that could be included for a more holistic look at the city. These include but are not limited to:

1) Violent incidents at schools reported by the school districts. This would provide a complete image of what Safe School Campus Initiative's school violence data approaches.
2) Data specifically on the mental health of youth in each neighborhood.
3) Truancy, and other data indicating engagement at school.
4) Graffiti data that specified which graffiti was gang-affiliated.
5) Data on crime committed specifically by youth.
6) Crime victimization rates (not just police reports). This would provide another look at where crime happens in San Jose, rather than where crime is reported to happen.

Despite lacking this information, most metrics are related to one another. Throughout the development process of this map, metrics have been added, removed, and edited, and the picture of the map has remained almost constant. In other words, while additional data would be useful, it is likely that data would point to the same neighborhoods as having the most present risk factors for youth and gang violence.

The next step in this work would be to identify what assets are available in each neighborhood, and how the City can leverage its services to provide targeted support to each neighborhood. If this map can identify the areas that are most troubled by substance abuse, then the City can reallocate its substance abuse prevention programs accordingly.

Finally, this map should only be used to adjust the allocation of supportive services, and not to determine suppressive services. By using data like this to redesign police routes, patrol cars, and other suppressive methods, the City risks neighborhoods deliberately hiding, or skewing the data on their neighborhood. So long as we use this map for a purpose people support, then we can continue collecting accurate data to improve the City's services to its communities.