First, for our reports on violence, we rely on victim-based reports from the Chicago Police Department's online data portal. Each report of a crime contains an Illinois Uniform Reporting code (IUCR) classifying the crime, a date, and a latitude and longitude for each event. We subset the data to focus solely on homicides and gun crime, including first-degree murder, second-degree murder and aggravated battery involving use of a hand-gun or firearm. In data spanning from January 1st, TKTK to December 31st TKTK, there are such \textbf{FILL IN} instances of violence. Violence does not follow a uniform spatial distribution as can be seen in the figure below. \textbf{Figure of violence in Chicago about here}. As is typical in such administrative data, there is no information available about whether individual crimes were committed by gang-members.

\footnote{ We choose to use victim-based crime reports because they provide the best coverage. When dealing with victim-based crime-reporting, the primary concern is that crimes go under-reported. While under-reporting is likely to be a concern in any dataset on crime, there are two concerns which aggravate the bias in our data. First, the crimes we are interested in are those committed by one criminal organization against another, thereby reducing the probability of reporting. Second is the fact that gangs often exist in areas where members of that racial minority comprise the majority (Bruhn 2019). It is well documented that racial minorities in the US have relatively lower levels of trust in police which is likely to translate to under-reporting of crimes (Desmond et al 2016). Though other measures of violence do ot suffer these short-comings (Carr and Doleac 2018, Carr and Doleac 2016), they are available only for narrow time windows.}

To generate our ``districts" we use the division of Chicago into tracts by the US American Community Survey. Census tracts are the second smallest statistical unit, containing approximately 4,000 people and 1,600 housing units.\footnote{Definition taken from the census glossary.} There are TKTK census tracts in Chicago. We aggregate our data by month, so that each individual observation is a count of the amount of violence in a given census tract for a given week. Our data covers Chicago in the years TKTK-TKTK in order to mirror @Bruhn with whom we cross-validate our results. Because census tracts have minor changes from year to year, we fix our districts as they were in 2016. In our sample period, there were an average of TKTK gangs operating in Chicago in any given year.

Also, how do we assign tracts to gangs. (subset to 6 major, assign tract to largest gang if any gang occupies at least 10% of tract)

The problem with the use of census tracts as an exogenously generated district is that it does not conform with the building blocks used by gangs to construct territories. Anecdotal evidence suggests that gangs fight for and control of blocks (Levitt and Venkatesh 2000). Census tracts aggregate multiple blocks into a single geographic unit. By using census tracts, we reduce the variance of our estimates but introduce the potential for bias in two ways. First, it is possible that a gang’s territory is assigned to a tract that contains peaceful areas, i.e. those uninhabited by gangs. Because peaceful areas are assumed not to produce violence in any systematic way, a census tract which incorporates gang territory and peaceful areas will not have the violence from the peaceful area affect the assignment of the territory to the correct gang cluster. However, our method will mistakenly assign the ``peaceful" areas to the gang. Conversely if the amount of territory owned by a gang in the tract is small, then it is possible that its territory will be designated a ``peaceful" tract. For purposes of cross-validation we consider a tract gang-owned if at least ten percent of its area is owned by a gang according the CPD maps.

Second, it is possible that a tract may comprise territory owned by two gangs. In this case, not only will territory necessarily be distributed incorrectly, it is possible that the algorithm will be unlikely to distinguish between the two gangs. This is more likely to occur for the smallest gangs given the very fractured territorial distribution of gangs in Chicago (Bruhn 2019). To deal with this issue we bootstrap our estimation and produce an uncertainty measure for each tract with regards to its ownership.

Two important features of gangs are captured by the model but are not explicitly incorporated into the data-generating process. First, gangs in Chicago are loosely organized in two alliances: the “Folk” and “People” Nations. Our data does not incorporate assumptions as two whether two gangs are allied or not, though this is captured by measures of the conflict intensities. Our measure provides supporting/contradictory evidence as to the effects of these alliances. The Black P. Stones, Vice Lords, and Latin Kings, members of the People Nation were more/just as/less likely to fight with one another as they were with the Gangster Disciples, members of the Folk Nation. Second, gangs are often racially homogenous, and control minority-majority areas. Moreover, most inter-gang conflict is intra-racial. Our estimation procedure, which relies on distinguishing patterns in the data on violence replicates this finding. The three gangs which we recover have an African American identity and are most likely to war with each other and had their highest conflict intensities with one another. The Latin Kings, the largest Latino gang, was the next most likely to be recovered per our estimation procedure, but less reliably so. The figure below shows the distribution of racial minorities in Chicago.

Unfortunately the model is unable to account for the dynamic nature of gang’s boundaries. We assume that gang’s territories are fixed throughout and seek to estimate these boundaries as though they were static. However, if gangs are warring, then it is possible that territory may change hands. Indeed, @Bruhn notes that an average of 29.5 gangs have their boundaries shift in any given year in the period that we examine. We can address this concern in one of two ways. First, we can split the time-series into different estimation periods and see whether or not boundaries differ across the panels. Second, we can treat our uncertainty estimates as measures of contested control, revealing how strong a gang’s grip is on a given territory. The figures below show a comparison of the gangs boundaries in the Chicago Police Department’s data in the first and last years of the sample. Given that shifts in boundaries are negligible, we opt not to split the sample.^{ It is possible that the observed changes are also the result of human error or minor inconsistencies in the construction of the data.}

**Insert Gang 2004 By Gang 2016 in the CPD Data**

INSERT GANG TURF AREA HISTOGRAM

INSERT ETHNIC MAP