Krish Seth, Nicholas Easton, William Ansehl
Data Science Seminar - Checkpoint 3
Clever Wombats

Allegation Map: To create this visualization, we began with a Geojson containing only the polygons of beats and a json with the number of allegations in a beat for a given year. From there, we filtered allegations based on the year, which is selected via an html drop down field. This allowed us to join the polygons and the corresponding counts for a year, scaling the color by the number of allegations. There is another layer of polygons which are the neighborhoods. Unfortunately, beats are not confined to neighborhoods so the mapping isn't perfect. It does allow us in broad strokes to compare neighborhoods though. We chose to not pin the maximum as the maximum over all years, and instead let the color scale vary according to the year. This makes it harder to compare the numbers of allegations across years, but it makes it significantly easier to compare relative relationships.

Over time, the number of allegations consistently goes down, from peaks over 600 in the year 2000 to peaks of 60 in the year 2018. This is consistent with the findings outlined by the Invisible Institute, where in recent years the number of allegations has been decreasing. As the maximum number of counts decrease, the deviation also decreases. The allegations are considerably more evenly distributed in 2016 than the early 2000's. Interestingly, there is a single hotspot right in the center of the city, right around the loop. This beat (#129) is persistent over the years, generally presenting as the brightest point on the entire map.

As mentioned previously, beats don't necessarily fall within a single neighborhood. This is frustrating as we would like to attribute allegations to a single neighborhood. If we aggregated in a preprocessing step this would be doable, however, it would be preferable to track both beats and neighborhoods at the same time. As it stands, we can only guess at where the allegations fall. One would expect that neighborhoods represent sufficient granularity to capture trends across the city as they present natural borders. From this map, we gain the understanding that some neighborhoods have considerable variation. This is more pronounced in the south where neighborhoods are generally larger. It perhaps suggests that even beats are too large to accurately describe the underlying distribution.

Stacked Bar Chart: To create this visualization, we utilized a csv file consisting of 4 columns: the year, the beat, the number of non-sustained allegations in that respective beat and year, and the number of sustained allegations in that respective beat and year. From this information, we filtered the allegations based on the beat we aim to further analyze. In the visualization, the user can toggle the sliding scale to select the specific beat they aim to analyze. From there, the user can see the number of sustained complaints in comparison to the number of non-sustained complaints. For example, for beat 165 in 2009, there were 6 sustained allegations and 2 not-sustained allegations. This ratio clearly flips in 2011, in which there is only 1 sustained complaint in contrast to 5 not-sustained complaints. Data can be visualized upon availability. In the case of beat 165, there is no data in 2010 that provides the number of sustained or not-sustained complaints.

By itself, this visualization makes it easy to analyze trends of sustained and not-sustained complaints within a single beat. Similarly, It is easy to compare two beats and their number of

sustained and not-sustained allegations over time. This visualization, however, lacks in geo-spatial awareness. Beats are an intrinsically spatial data type. The beat number refers to a specific area in Chicago. As such, it is impossible, from this visualization alone, to understand proximity between beats. Similarly, it is impossible to compare neighboring beats in their trends over time. For this reason, this visualization pairs nicely with the interactive choropleth map introduced earlier in this paper.

The choropleth map gives a user an “at a glance” understanding of beat proximity and allegation hotspots over different years. Utilizing the choropleth map, a user can narrow in on a particular year, identify a hotspot of interest, identify neighboring beats, and ultimately turn to the stacked bar chart for further analysis. For example, in 2009, beat 261 is a hotspot of allegations. We can also identify that its numbers - beats 217, 253 and 257 do not seem to be as “hot.” Now we can proceed to the stacked bar chart for trend analysis.

In 2009, beat 261 had approximately 15 sustained allegations and 18 not-sustained allegations. In 2010, beat 261 had approximately 14 sustained and 5 not-sustained. And so on. In contrast, beat 257 had approximately 2 sustained complaints and 14 not-sustained complaints. Beat 257 also experienced 7 not-sustained complaints despite 0 sustained complaints. And so on. Clearly, the ratio of sustained complaints to not-sustained complaints this area tells very different stories according to the beat you are looking at.

We can also compare non-neighbor beats to understand how ratios might change across the city landscape. For example, beat 170 is a relatively low allegation count beat in congruence with its neighbors. In 2009, there were only 6 not-sustained complaints and 0 sustained complaints. The number of allegations only continued to decrease as time progressed.