

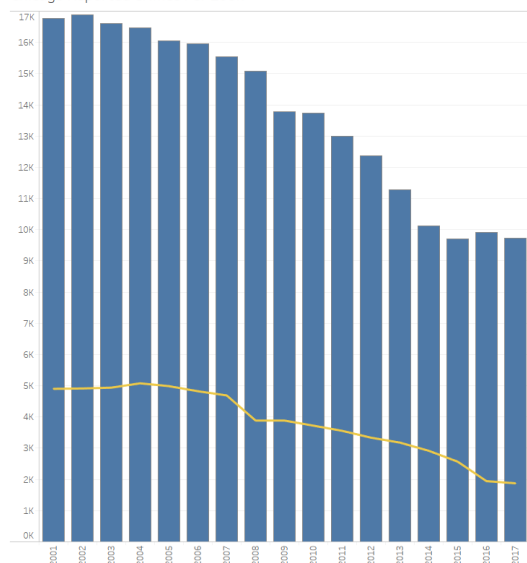
Samuel L. Peoples  
STMATH 493: Data Analysis and Visualization  
Visualization 3: Chicago Crime and the State of Law Enforcement  
Journal

My group bounced around a lot of ideas for this visualization. We chose the Chicago dataset because it was large, and found that the data expressed a lot less than we initially hoped. Because the policies in Chicago have changed dramatically over the last two decades, measuring the true level of crime is difficult, as the behavior of the police is not constant. We played around with a few different external data sources, but ultimately relied almost entirely on the data provided from the City of Chicago.

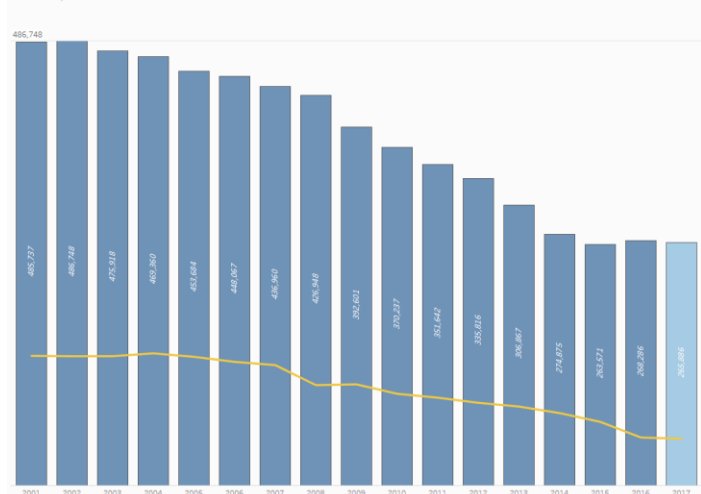
An early direction that we attempted to go was collecting data about symptoms of criminal activity (such as overdoses), but had trouble finding consistent sources of information at the level of granularity, and within the time-frame required. We then settled on the question “*In what districts and for which crimes are the resources of the Chicago Police being allocated?*” By focusing the question on the attention of the officers, we can avoid any miscalculated conclusions, for which it is tempting to directly relate criminal reports to public safety.

We began by questioning whether we should normalize our samples in a per-100,000 sense, but realized that our later desires for a parameterized dashboard would prevent the normalization from remaining accurate across different parameterized filters. Thus our first visualization, which plots the total reported crime and arrests from 2001 to 2017, minimally went through the following changes.

Chicago Reported Crimes Per 100k



Total Reported Crimes



Filtering and parameterizing were the next goal. We knew that we did not want to work with incomplete data, so we filtered the year immediately. A workbook-wide filter was defined for district, excluding null fields, which does include districts 13, 21, 23, and 31, which do not contain a statistically significant amount of data, and actually ‘break’ the visualizations if they are included. The filter is defined by a formula, where:

```
[District]=IF [Parameters].[District]='All' THEN [District] ELSE [Parameters].[District] END
```

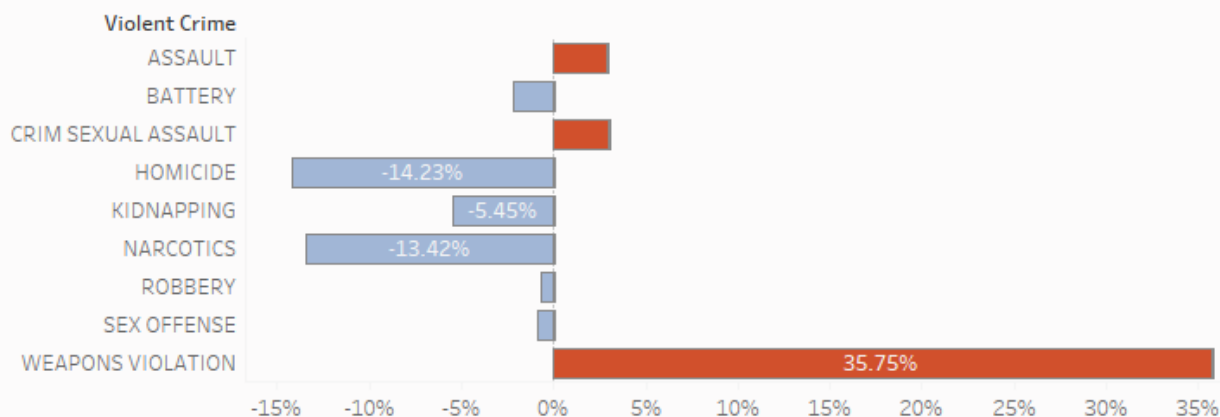
This allows the user to filter the entire workbook at by specific filters, as well as an option to choose all districts if desired. The Primary Type was filtered next, progressively refining two sets, Property Crime and Violent Crime, which will be used to examine

specific crimes. This filter is also parameterized in a similar manner as above, allowing the visualization to be examined from the perspective of a specific crime.

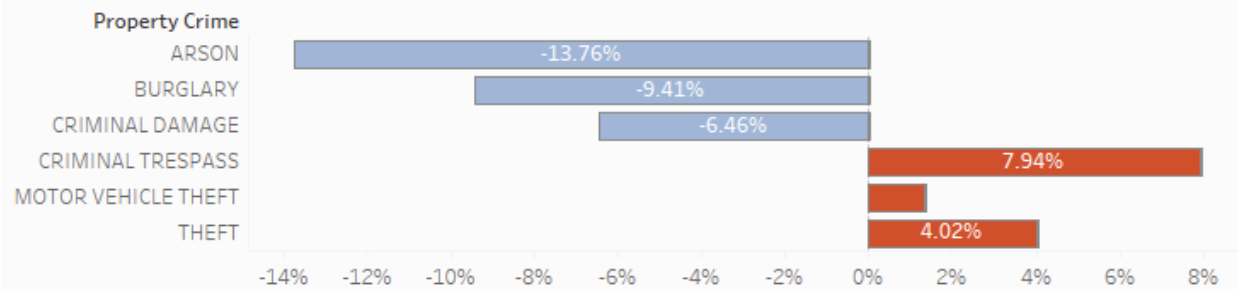
The visualization was then stripped of any unnecessary information. Initially, the yellow line of arrests were on an independent axis, so the axes were synchronized, and the redundant axis was removed. The major axis was an eyesore, as the values were in the “xxxK” format, so the individual columns were labeled with their values, and a reference-line was added for the maximum value for the current visualization. Because we planned to analyze more specific information about 2017 in the ensuing views, the 2017 column was highlighted with a different color, so to keep a majority of the focus on that year. The gridlines were ultimately useless after removing the axis, so those were removed as well.

Next, the two sets, Violent Crime and Property Crime were visualized, where our group ultimately found that it was interesting to investigate how each crime was reported with respect to the previous year. This answers questions such as “Which crime rates have increased or decreased since the previous year?”. So we visualized the “Percent difference from 2016 to 2017 for each crime in Set”. Because the workbook has the parameterized District filter, we were excited by this visualization because we could also ask questions like “How does the percent change in district 1 compare to district 14 with respect to homicide?”.

Percent Change in Police Intervention of Violent Crime 2016 to 2017



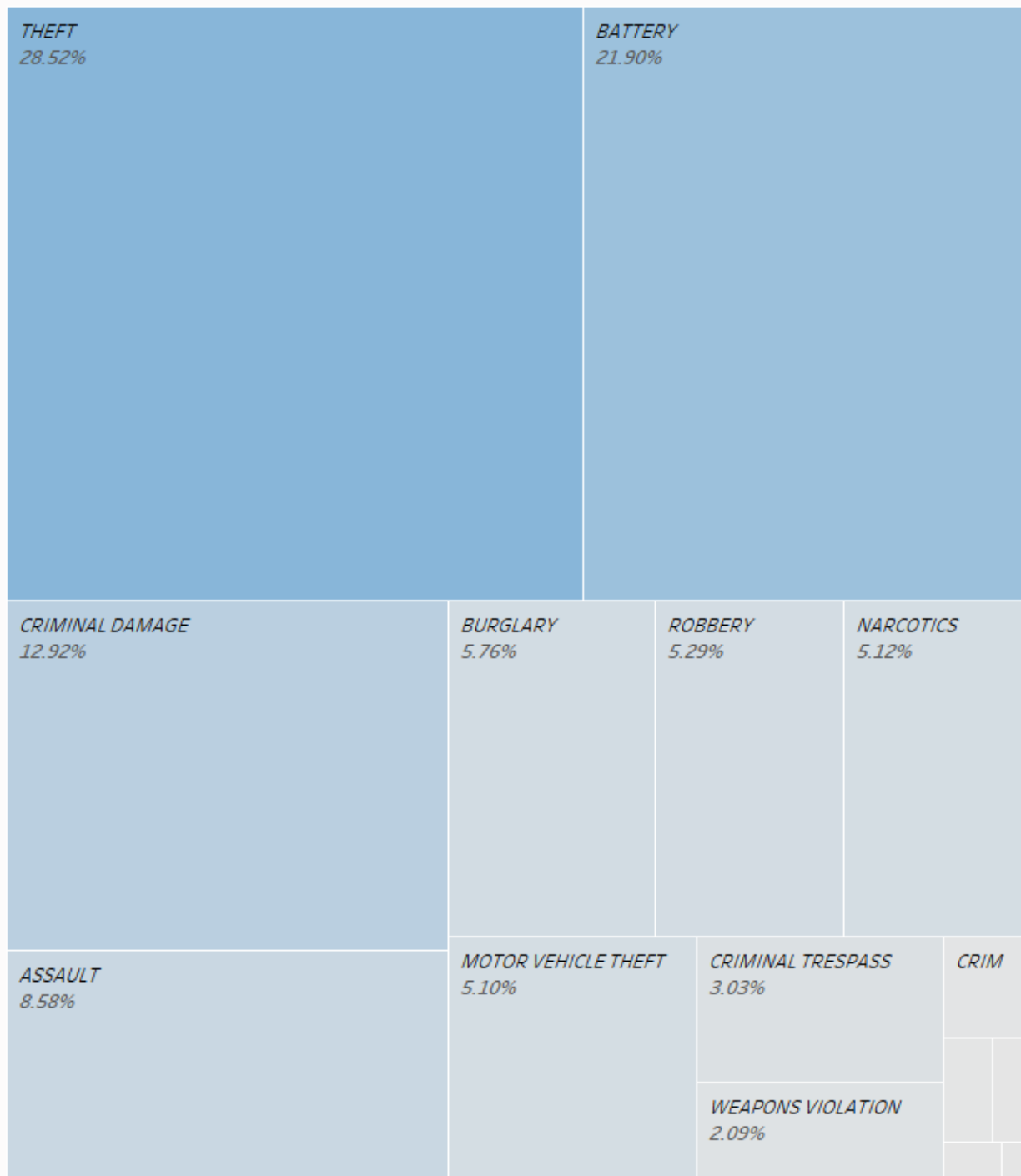
Percent Change in Police Intervention of Property Crime 2016 to 2017



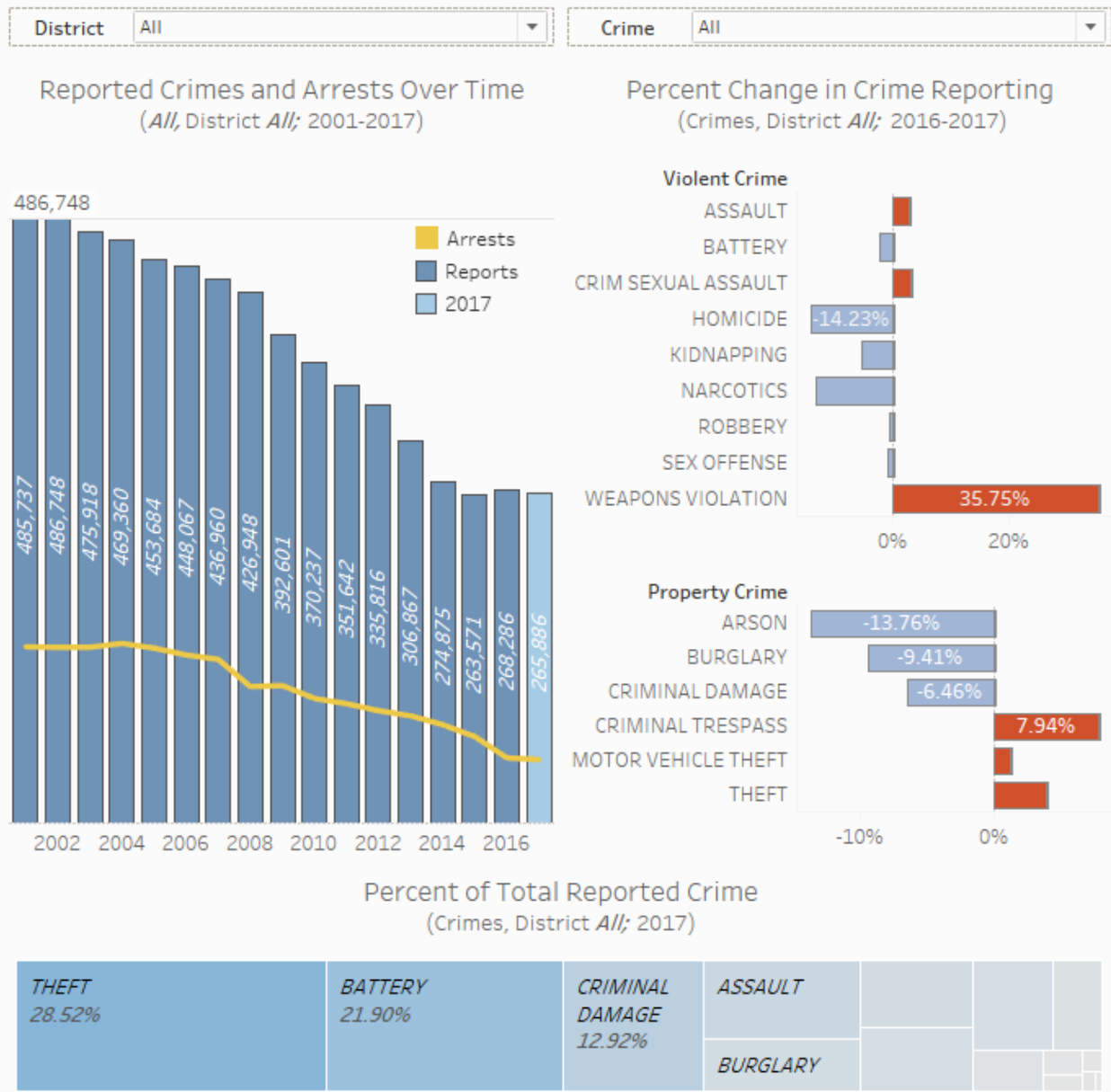
Similar to the year-by-year trend, this visualization was themed to reduce the visual clutter, keeping as much as necessary. The bars were colored intuitively, suggesting red is an increase in the reporting of a crime, and blue is a decrease in the reporting of a crime. Our group did spend some time debating the redundancy and possible confusion of the colors in this instance, but felt that adding yet another color would be too much color.

We realized that the percent increase or decrease in the reporting of a crime doesn't provide much context without understanding the distribution of total crimes reported during that year. If one were to be surprised at the 88% increase of kidnapping in a district, they would be interested to know if that represented 4 instances of kidnapping, or 44 instances of kidnapping; so we decided to visualize the Percent of Total Crimes Reported in 2017. Initially, a pie chart was considered, but decided on a Treemap for the benefits of retaining the labeling theme that had been established so far.

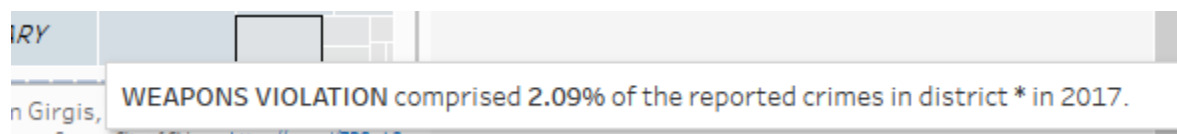
### Percent of Total Crimes Reported in District \* in 2017



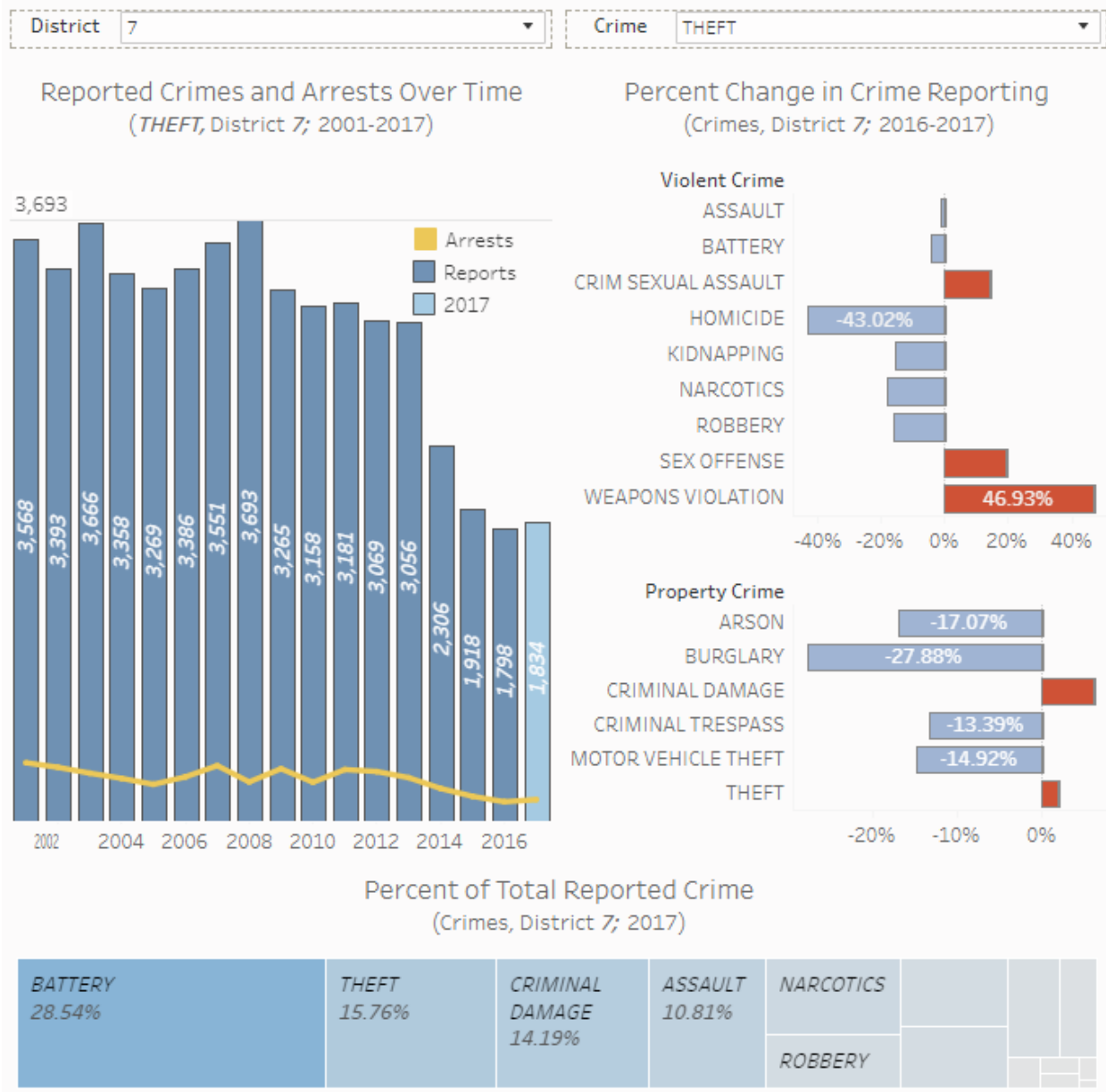
This Treemap was colored to match the blue theme, intuitively drawing attention to the largest value. This visualization was fairly simply developed, but our group feels that it is necessary for the viewer to understand the full context of the data.



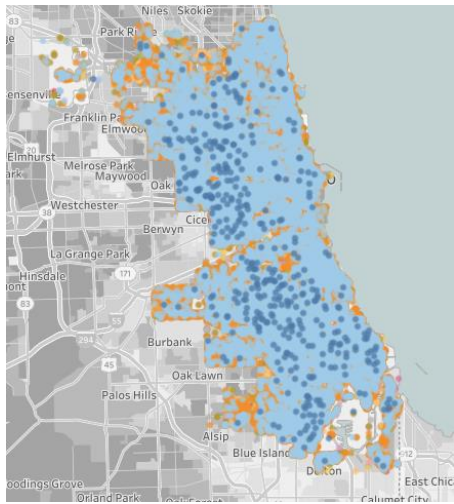
Now, the visualizations were combined in a dashboard, optimizing the use of space, while still guiding the viewer through the most important aspects of the options chosen. A person can see that weapons violations are up 35% in 2017, but only made up 2% of all crimes committed, so that percent of an increase is less significant than if it were for Theft, which is over a quarter of all reported crimes in 2017.



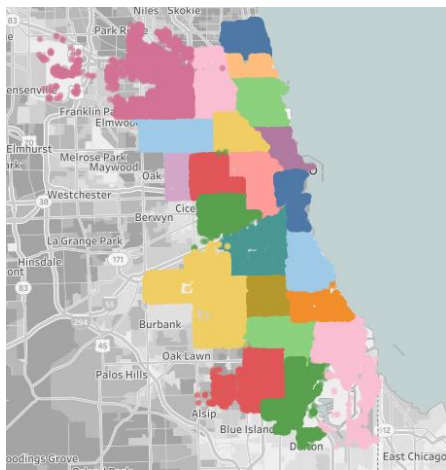
The chosen parameters allow the visualizations to change dynamically, filtering across the workbook to show different information. For instance, a person could ask how Theft has been reported historically in a specific district, and how it was reported in 2017.



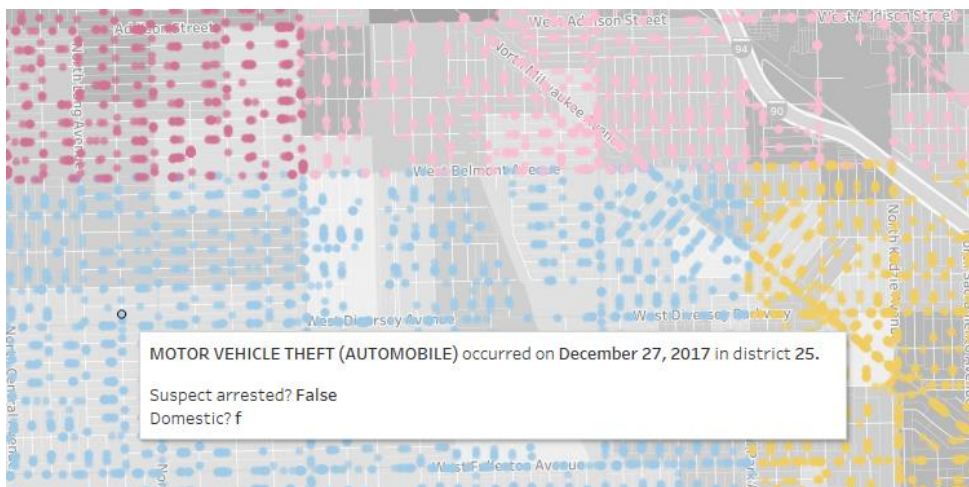
This information was useful to us, but decided that a map of the different areas was required to see where different districts were, and where different crimes were reported more or less. An external (built-in) base map layer was used showing the latest block-level Median Household Income, thinking it would be useful to know whether crimes were being reported in a higher income neighborhood, or in a lower income one.



Viewing all events in all districts (colored by primary type), provided very little insights, but after some experimentation, our group decided on coloring by districts.

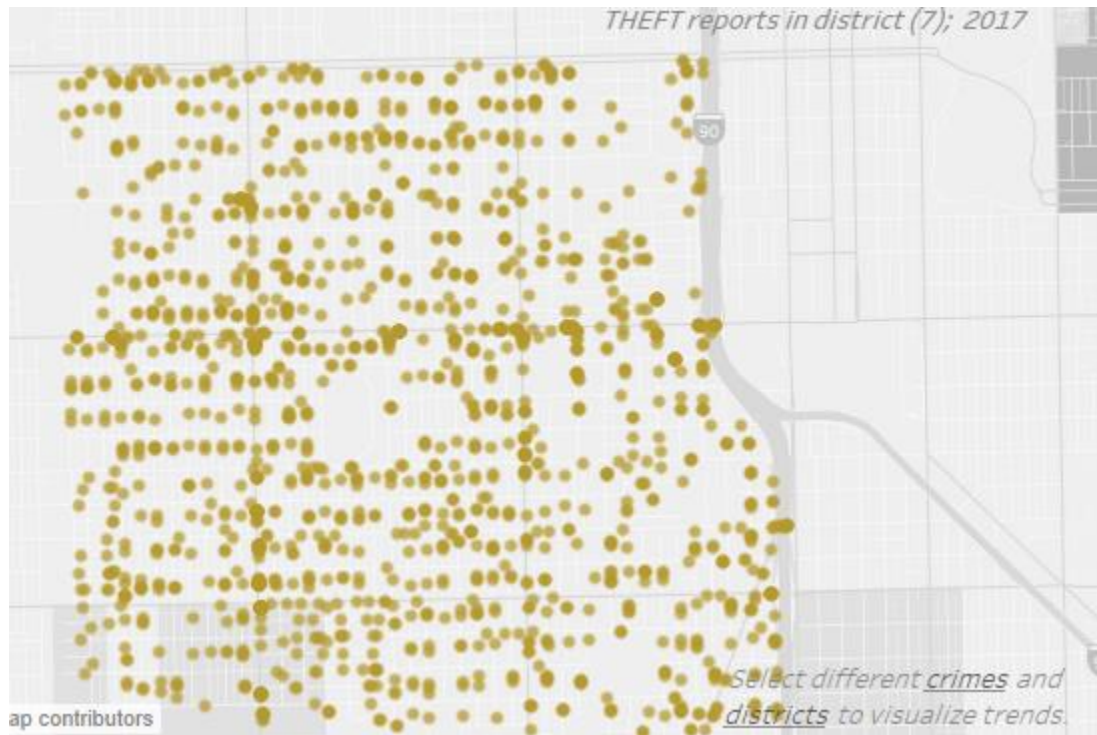


This allows the viewer to immediately see that the colors are separating specific districts, and invites them to manipulate and explore the various areas of the map. The size was changed numerous times, attempting to optimize the density when all points are visible, so to explicitly draw borders around each district, while also allowing the viewer to see the underlying map when zoomed-in on a district.





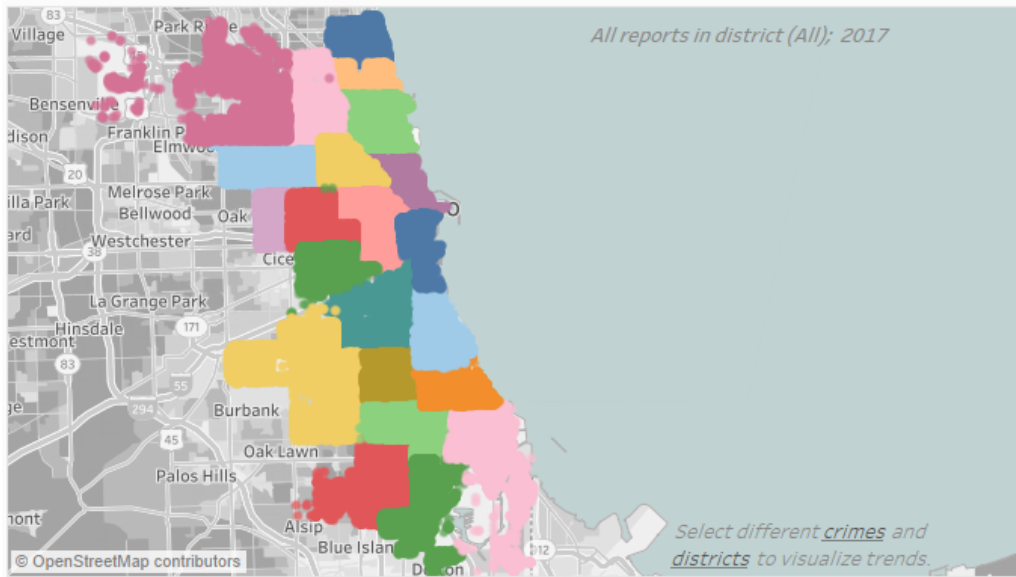
By including the same parameterized filters in this map, the viewer is able to refine their map data to a specific crime and district, such as the Theft in district 7 from above.



A subdued title is included on the map, allowing the viewer to see what the map data contains without inferring it from the tooltip, which changes dynamically with the parameters. This finally leads us to a dynamic dashboard displaying the latest full year of Chicago crime data, with enough contextual information for users to draw independent conclusions.

## Chicago Crime, and the State of Law Enforcement (2017)

In what colored districts, and for which crimes are the resources of the Chicago Police being allocated?

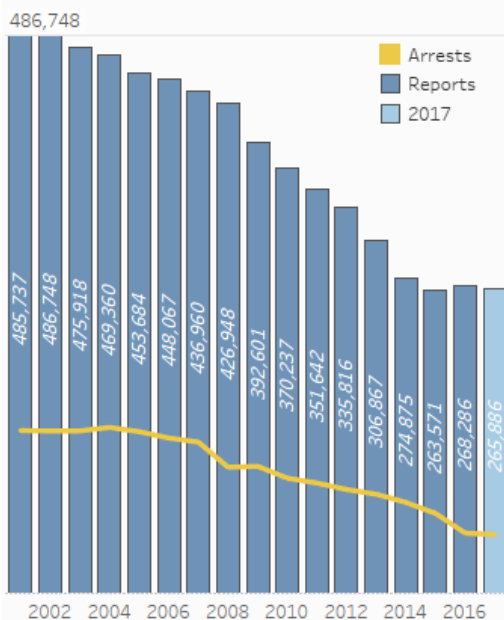


### Median Household Income

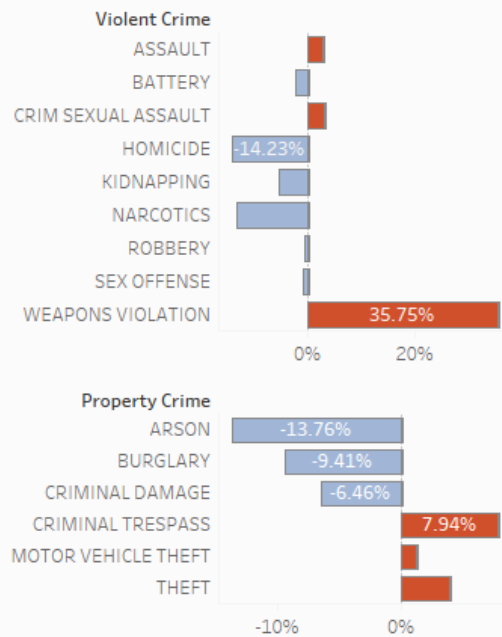
0 to 39,000 39,000 to 50,300 50,300 to 64,500 64,500 to 86,200 86,200 to 501,000

District All Crime All

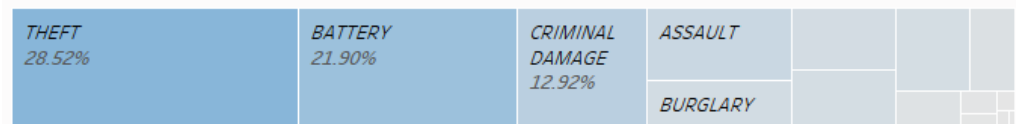
### Reported Crimes and Arrests Over Time (All, District All; 2001-2017)



### Percent Change in Crime Reporting (Crimes, District All; 2016-2017)



### Percent of Total Reported Crime (Crimes, District All; 2017)





An instruction was included telling the viewer to “select different crimes and districts to view different trends”, inviting the user to look for the underlined features. The associated drop-down boxes were placed in a dashed border, further highlighting their importance. We were worried about the viewer not immediately seeing the dynamic features of the visualization, so these steps were taken to better direct the user’s attention. The color-key for Median Household Income was included below the map to direct the user’s attention to the map layering, as well as inviting the user to make comparisons based on this information. The map was sized with the intention that the user would see both that, and the following visualizations, that way any parameters that they adjust will display that multiple aspects change in the visualization.

Although the manner in which the data is visualized allows for a lot of information in a small space, our group was worried about the “busy-ness” of the space used, as well as the limitations of the dashboard. We are only looking at one districting feature, while the data contains more than four, the unique identifiers are not available, and the crimes chosen are listed for their relevance to subjective labeling of “Violent Crime” and “Property Crime”. The boolean features (domestic, arrest) were not very well described, and are not a focus of this visualization, which reduces the scope of the overall presentation. However, our group believes that this dashboard provides great opportunities for understanding the general trends of crime reporting in Chicago, where many conclusions can be drawn, and very specific questions can be answered.