Assignment 3: Visualization with Tableau

Q1. Rationale

The Chicago Crimes data set contains records of information about a variety of crimes committed in the city of Chicago between years 2012 to 2016 inclusive. Each record contains features that describe the time, location, type of crime, and whether or not an arrest was made. From the perspective of anticipated users, the crime records can be represented in a multitude of formations to describe the *where, when, and what* of criminal activity. The crime records also contain specific location information pertaining to the policing district and beat (patrol route) involved in the crime. This extra information also provides an opportunity to add a level of geolocation context not immediately available in the original dataset.

We surveyed the online resources made available by the department of Public Safety[1] at the City of Chicago and found polygon maps containing shape labels and numbers for police districts, police stations, and individual police beat zones. Other additional datasets contained shape information and census data for Chicago's community areas (see Appendix A). This extra information was linked to the original dataset to enrich the majority of our dashboards.

The requirements of a police chief or senior officer (such as district captain) would probably involve a daily snapshot that quickly describes both of time and location, with the opportunity to derive additional information about specific types of crime occurring either across time, or within certain areas, or both. After some discussion, we decided to make three dashboards; one for a daily Overview, and two alternate views to cover Trend Analysis and Area Analysis as part of a larger Storyboard.

Daily Overview

The Daily Overview (Figure 1) is the starting point of regular review which is taken into account via providing a prefiltered (top #) bar graph of crimes recorded in the last two weeks. Considering the date ranges supplied by the City of Chicago records, we present the "current date" as the day after the last date provided to give a more realistic impression of a daily business intelligence dashboard. The calculations within the dashboard therefore statically refer to this reference date.

In addition to the bar graph, integrated 4 separate maps into the overview which can be selected from the lower row: Via combining the police beat and census datasets, we created two heatmaps (the first two in the lower row) to provide the user a quick and natural[2] review of crimes recorded in the last 2 weeks broken down by density of population (per 10,000 people) or police beats. Combining the original time and location data with area and population allows the intended user to distinguish between high crimes unbiased by small zoning or population congestion, otherwise the only value added by a heatmap is a semblance of population density. Hovering over or clicking any of these areas allows the user to see the name of the Community Area in question

[1] https://data.cityofchicago.org/Public-Safety/

[2] We think that most people are used to reading maps and they are easier to interpret and memorize than tables and lists.

and to filter all visible data to the selected area. Filtering is also possible via selecting one or more bars (crime) or categories (crime type) in the bar graph which renders the overview very easily customizable and which is useful for certain use cases (see Q2. Interpretation).

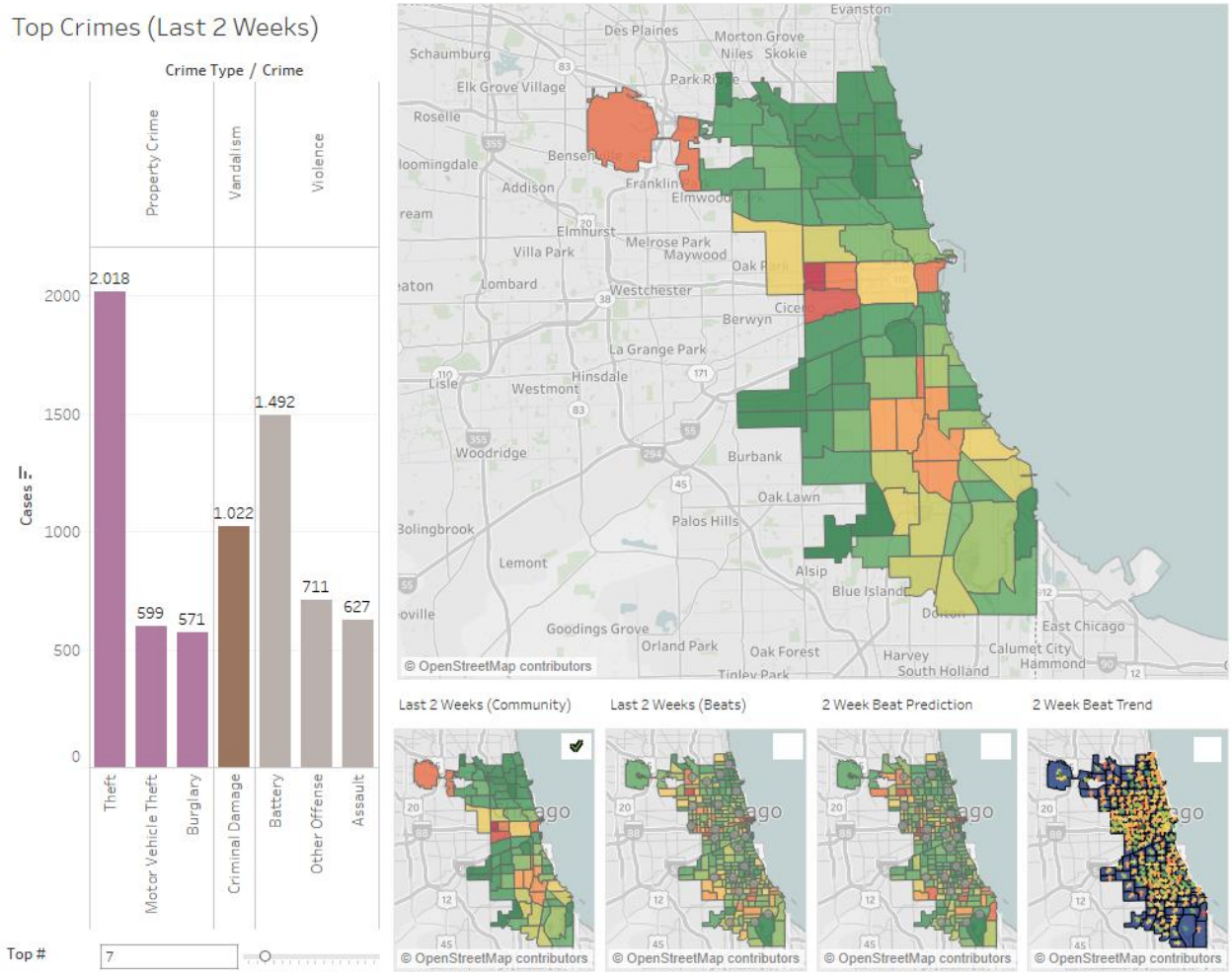

Figure 1: Cases by Community Area and Population

While designing the Trends dashboard for time-series analysis we noticed a number of distinct and repetitive cycles in overall crime rates based on time of year. This observation lends well to predictive analytics, and so we provided a two-week forecast (prediction) broken down by police beat (third map in the lower row). The prediction is made by calculating the increase/decrease rates of case numbers for the two weeks before and after the given date one year in the past. Then, the case numbers of the recent two weeks are multiplied by these ratios. This way both seasonal fluctuations and the long-term trend of decreasing case numbers are taken into account for the prediction.

However, we thought that it would be great if one did not have to compare the diagrams “Last 2 weeks (Beats)” and “2 Week Beat prediction” to visually gauge whether cluster of crime are increasing or decreasing in a two-week historical window, in case adjustments need to be made for district resource allocation. To address this requirement, the last map in the lower row of the Daily Overview (Figure 1) visualizes the trends of the predicted crime numbers as a comparison of the case number in the second and third map. The changes in trend are represented by trend

indicators which make use of both shape (arrow) and colour (heatmap gradient); this approach allows the user to see both trend cluster direction and the exact decrease/increase rates across beats in a single map. The predicted number of cases is also visible via the tooltip (hovering over beats) but the third map "2 Week Prediction Heatmap" gives a more convenient overview.

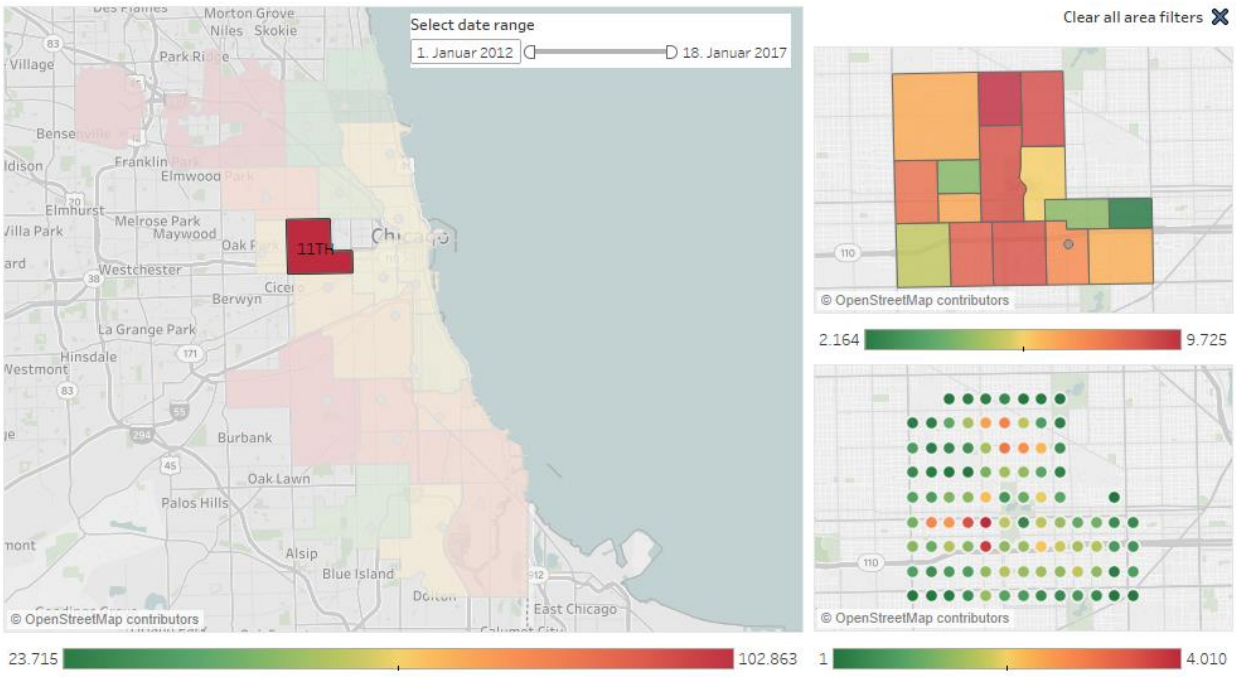

Figure 2: Crimes by district, beat and block over time

Area Analysis Dashboard

The Area Analysis dashboard uses a set of three heat maps with different aggregation levels being police district, police beats and blocks (based on rounded geo-positions) so that selecting an area in one of the maps automatically filters the other maps to this area (Figure 4). Therefore, if e.g. one selects a certain district, one gets a more detailed view in the other maps, as the maps are automatically zoomed in (if they have not been manually dragged or zoomed).

The bar graph (Figure 5) placed below these maps is automatically filtered to the cases of the selected area (district, beat, block) so that one gets more granular information about the cases of the selected areas.

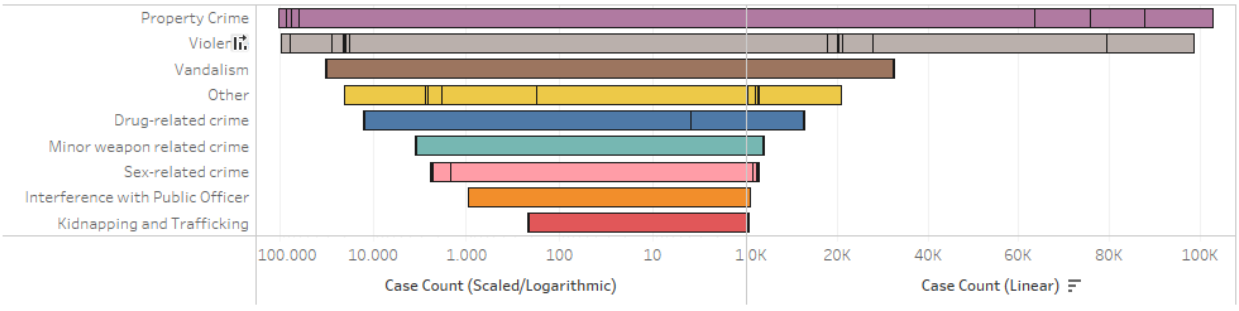

Figure 3: Case Count as Linear or Scaled Case Type

To design the bar graph (Figure 5) in a way which is informative but not too detailed we applied two techniques, namely "grouping" and "scaling". The grouping is the same as in the trend line diagrams (and consequently utilizes the same color legend), but here it is possible to hover over the bars to see the numbers for specific primary crime types or clicking (selecting) them to filter to these cases in all the views diagrams. The scaling of the graphs is done in a special way, as each graph is shown twice with different scales: On the left we used a logarithmic scale and on the right we used a linear scale. This way the user gets a feeling for the magnitude of less frequent crimes (linear scale on the right) but nevertheless he is able to see and select the details of these crimes (logarithmic scale on the right), which otherwise would have been difficult to do.

Trends Dashboard

The Trends dashboard is used to convey a more detailed breakdown of crime types over time and presents this as two timelines using two different views, all sharing the same colour legend for ease of comparison. Note that the original dataset contained over 34 different "primary crime types", and so we created rollup classes for ease of presentation and understanding.

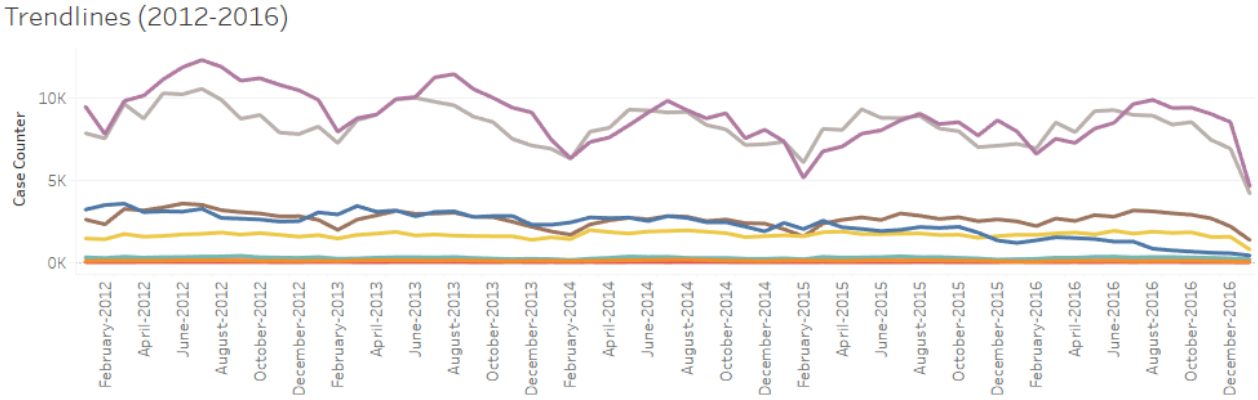

Figure 4: Four-year Trend by Crime Type

The top diagram (Figure 4) shows major crime types over the last 5 years and shares a common legend with the other views. Hovering over any of the trend lines allows the user to see the date, crime type, and case counter for that month. Clicking on any of the trends will highlight that crime type in all graphs in the dashboard and also show the group members in a separate "popup"-information.

At last we placed an additional bar graph depicting the change rates comparing the last five years, as it is easier to read this information from this diagram type than from trend lines (Figure 5). Filtering is possible in the same way as for the trend lines.

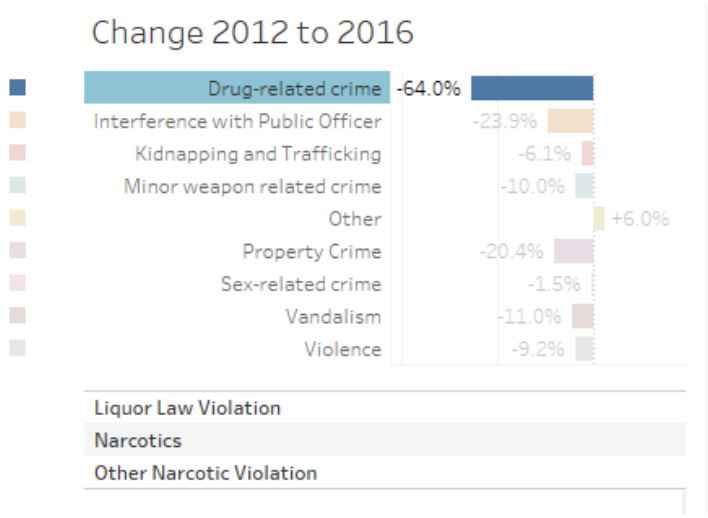

Figure 5: Change in Crime by Type

Q2. Interpretation

When comparing the crimes and locations, visualizing by heat map makes the most sense as problem areas are easier to see. The intended user will likely have a good understanding of what these dimensions mean, as well as the value of drilling down into the specific areas to see how the aggregate crime types change for each area; example being the police chief have a discussion with various district captains.

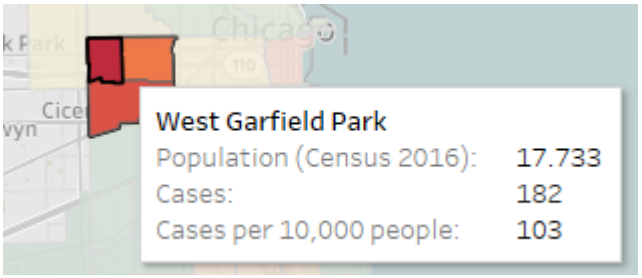

Figure 6: Two-Week Historical Heatmap

The Storyboard view that we have presented starts off with the Daily Overview. As a senior manager I would be drawn to the high crime rates in "West + East Garfield Park" and the connected area "North Lawndale". Selecting all three areas will show that the majority of cases is related to narcotic delicts and battery.

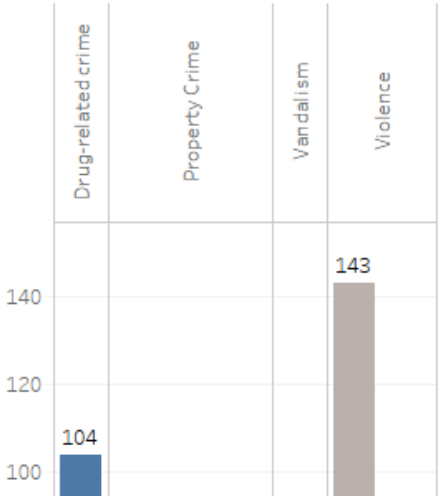

Figure 7: Two-Week Top Crime

Moving along to the Area Analysis and filtering on drug-related crime will show that the highest occurrence numbers are located in district 11 (which contains "Garfield Park"). An assumed correlation to prostitution can be verified quickly via filtering on that type of crime (grouped within "Other"). Based on these findings one could decide to send street workers to this area.

Additionally, the area analysis also provides a way to directly focus in on specific areas related to specific case types. Filtering on this primary crime type in the bar graph, we see that Kidnapping mostly occurs around the 16th and 25th district:

A focused view into this district (Figure 8) shows the recorded location is adjacent to the O'Hare International Airport main feeder roads; this intuitively makes sense, as unauthorized travel to other countries related to a custody battle will be charged under kidnapping.

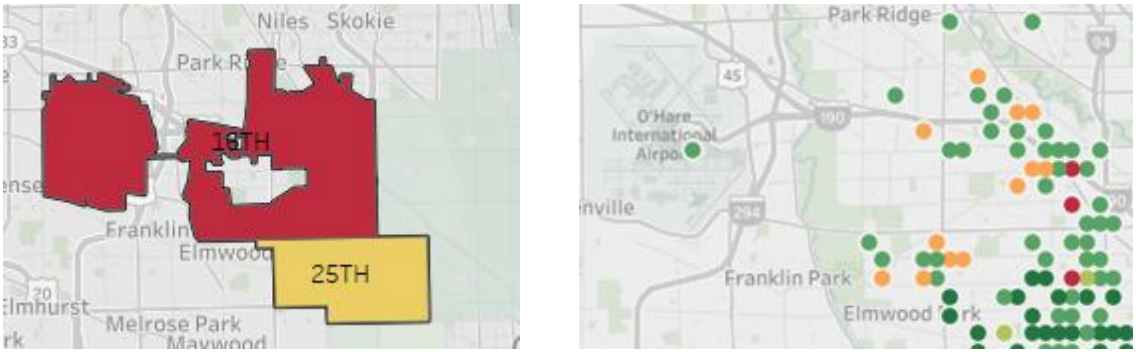

Figure 8: Cases of kidnapping in districts 16 and 25

This would indicate that in fact the public safety system is working as intended, since these crimes are intercepted at the fringes of another legal jurisdiction (e.g. based on police road blocks).

Continuing to the Trends View, the user can see recurring dips of criminal activity related to property crime and violence during the months January and February (top two graphs in Figure 9), probably attributed to the low temperature (and almost opposite of the June-August timeframe).

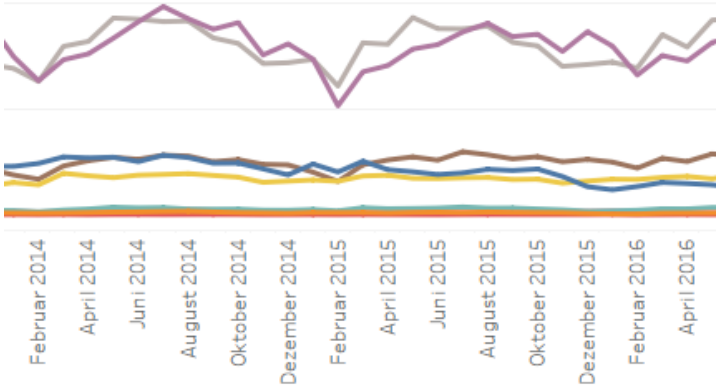

Figure 9: Recurring dips in property and violence crime

The aggregate from previous years indicate that most districts will experience a dip in case numbers, with the exception of a notable few (which can be checked in the Area Analysis); combined with a historical understanding that "normal" levels only resume in March (when things start warming up), the police chief could action on this information to temporarily reallocate beat patrols away from low- to high-intensity areas for the next two months.

The Trends view also presents a visual comparing increase/decrease in crime type (Figure 10), and we see that priorities in drug-related education and awareness, and public-facing liaison opportunities have paid off; crime rates are down 64% and 23% respectively. However, we also see the two lowest reductions amongst Sex-related and Kidnapping and Trafficking at 1.5% and 6.1% respectively (and not including the "other" catchall). This provides insight that outreach programs are not working, and perhaps there are other factors to consider.

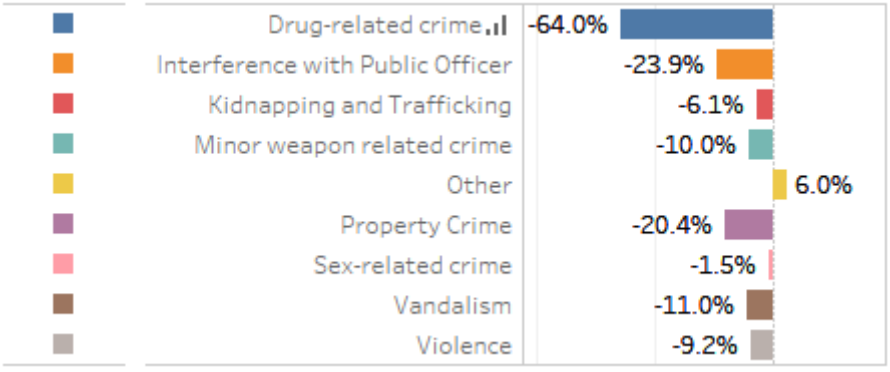

Figure 10: Comparing increase/decrease in crime type

Appendix A – Dataset Imports

Dataset	Source	Type	Modifications
Chicago Crimes 2012-2016	https://www.kaggle.com/currie32/crimes-in-chicago/data	.csv	Added header for first column
Police Districts (current)	https://data.cityofchicago.org/Public-Safety/Boundaries-Police-Districts-current-/fthy-xz3r	.shp	None
Police Beats (current)	https://data.cityofchicago.org/Public-Safety/Boundaries-Police-Beats-current-/aerh-rz74	.shp	None
Police Stations 2012	https://data.cityofchicago.org/Public-Safety/Police-Stations-Shapefiles/tc9m-x6u6	.shp	None
Community Areas (current)	https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Community-Areas-current-/cauq-8yn6	.shp	None
Census Data by Community Area 2016	http://www.actforchildren.org/wp-content/uploads/2017/02/Census-Data-by-Chicago-Community-Area-2016.pdf	pdf	None (used Tableau data interpreter and created union)

Table 1: Datasets

The original datasets have not been modified except for the original .csv–file of Chicago Crimes, as that was missing a header caption for the first column. Tableau was unable to overcome that. (Table 1) contains the details about the dataset usage in the workbook. When working with the polygon maps we also faced one issue which we could not overcome in the given time: 2 districts consist of several separate areas and therefore the tables contains several map polygons for one table key (district number). The joining of the data therefore had the effect that these districts are missing in all district- and beat-related maps. Blending the data instead of joining it did not work either.

Personal Ethics & Academic Integrity Statement

Student name: Chad Peters Student ID: 8480761

Student name: Fabian Willems Student ID: 300045304

By typing in my name and student ID on this form and submitting it electronically, I am attesting to the fact that I have reviewed not only my own work, but the work of my team member, in its entirety.

I attest to the fact that my own work in this project adheres to the fraud policies as outlined in the Academic Regulations in the University's Undergraduate Studies Calendar. I further attest that I have knowledge of and have respected the "Beware of Plagiarism" brochure found on the Telfer School of Management's doc-depot site. To the best of my knowledge, I also believe that each of my group colleagues has also met the aforementioned requirements and regulations. I understand that if my group assignment is submitted without a completed copy of this Personal Work Statement from each group member, it will be interpreted by the school that the missing student(s) name is confirmation of non-participation of the aforementioned student(s) in the required work.

We, by typing in our names and student IDs on this form and submitting it electronically,

- warrant that the work submitted herein is our own group members' work and not the work of others
- acknowledge that we have read and understood the University Regulations on Academic Misconduct
- acknowledge that it is a breach of University Regulations to give or receive unauthorized and/or unacknowledged assistance on a graded piece of work