Redistricting Youngstown Police Beats

S. Haigler
shaigler@student.ysu.edu

A. Orr
ashley.orr@economics.ox.ac.uk

E. Shehadi
eashehadi@student.ysu.edu

J. Wise
jlwise@student.ysu.edu

K. Yazvac
kcyazvac@student.ysu.edu

Advised by: Dr. Thomas Wakefield
tpwakefield@ysu.edu

Youngstown State University

October 10, 2016

Abstract

The geographical area which a given police officer patrols is known as a beat. Unfortunately, the beats in Youngstown, Ohio, had been the same for more than a decade, despite changes in calls for service and trends in crime. We analyze historical calls for service data to confirm the disparity in workload across these beats and this analysis shows a highly uneven workload across the current beats. To remedy the inequity, we propose a Cellular Growth Model using cellular automaton to generate new beats. This model considers both the call volume and the call priority of the historical data through the use of a metric.

We generate and present two sets of results, one to check the robustness of our model and another to present to the Youngstown Police Department. First, to train and test our model, we generate new beat alternatives using 80% of our data and then use the remainder of our data to determine which alternative has the lowest mean squared error between the in-sample data and the out-of-sample data. Bootstrapping our data as a means to generate confidence intervals, this alternative is then compared to the original beats.

In application, however, we generate beat alternatives using all available data from the Youngstown Police Department. Our analysis concludes that the new beat alternative has lower standard deviation and lower coefficient of variation than the original beats. As a result, the Youngstown Police Department implemented one of our proposed alternatives in January 2016.

Keywords— cellular automaton, police beat redistricting, crime trends

1   Background of the Problem

A pressing problem for the Youngstown Police Department (YPD) in Youngstown, Ohio is the unequal workload for officers covering different geographic areas of the city, referred to as beats. Although there is no typical time period for cities to redraw police beats, many cities realized the need for updated beats in recent years, thanks to the availability of more data and technology in policing. In Youngstown, beats had not been modified since 1998. During this time, calls for service and trends in crime have changed, creating inefficiencies for the Department. The YPD recently attempted to redistrict the beats of Youngstown. However, they were unable to reach a consensus on a redistricting proposal because of the limited computing power and analysis capabilities of the Department.

The YPD requested that we examine all crime data from one complete year and propose new beats that better equalize workload. This led to the creation of a mathematical model for redistricting the beats and a particular application of that model for the YPD's use. We redistrict the beats according to probable variation of calls, using data from November 2013 to November 2014. We present results training and testing our model using a subsample of all the data available.

2   Literature Review

The National Institute of Justice (NIJ) serves the U.S. Department of Justice through research, development and evaluation of crime. With the recent expansion of available data and technology, it now advocates for predictive policing. Predictive policing is the use of prior crime data and analytical modeling to predict and reduce the number of crime incidents [7]. In this section, we review several papers addressing these predictive-type models for policing and emphasize two papers that were pertinent to the development of the methodology we used to design and analyze new beats for the YPD.

Traditional predictive policing methods include hot spot policing, repeat victimization assumptions, and Broken Window Theory. Hot spot policing occurs when police departments concentrate resources on small geographic areas where there is a high concentration of crime (see Braga [1]). Repeat victimization and near-repeat victimization refer to the concept that new crimes are more likely to occur where crimes have occurred historically or in the surrounding areas of these crimes. These theories are applicable to individuals and locations (either exact or surrounding) which are involved in repeated crimes [6]. Developed by James Q. Wilson and George L. Kelling, Broken Window Theory argues that monitoring and preventing small crimes will prevent larger crimes from occurring in that area [10].

Recently, geospatial predictive techniques are being employed to predict and combat crime. Since 2012, five NIJ grants have been awarded to create geospatial police strategies. These projects unite historical crime data, hot spot analysis, "near-repeat victimization," and place-based allocation of police resources in their studies. Our analysis is similar as it relies on historical data and spots of concentrated historical crime to allocate police resources by location.

Curtin et al. [4] discuss a similar problem to the one proposed by the YPD. The authors note that there is a lack of objective quantitative methods for designing police beats and scheduling police officers. Because of this, they attempt to provide an understandable, yet comprehensive, model to create new beats for the Dallas Police Department. Their paper focuses on determining efficient spatial distributions of police beats and how operations research methods can be used in police decision making. It begins by presenting a maximal covering formulation and then creates an innovative backup covering formulation. Their model is called the Police Patrol Area Covering (PPAC) model, and the main goal is to maximize coverage of incidents. The PPAC model optimally solves for maximal coverage and then maximal back-up coverage. It is a linear system that aims to minimize the distance traveled by police officers while increasing the number of incidents covered. The PPAC model takes into account known incident locations or calls for service, potential locations for police patrol command centers, desired acceptable service distance, the shortest distance from an incident to a police command center, weight or priority of crimes, and the number of police patrol areas to be located. The acceptable service distance changes for different types of incidents and is dependent on the preference of the city. All of these elements are important in constructing a model for police beats. Finally, they test these methods using the police geography of Dallas, TX and compare the solutions with the existing beats. They conclude that these methods produce better results, meaning more incidents covered with minimal distance traveled, than the current beats being used.

D'Amico et al. [5] describe a different modeling approach. To update traditional approaches for allocating patrol cars, they account for queuing effects, such as time to complete a call for service. However, the objectives and constraints of their model become highly non-linear by accounting for these queuing effects. Thus, the authors propose a simulated annealing algorithm, where a variant of a Patrol Car Allocation Model (PCAM) is applied at each iteration to find an optimal allocation of patrol cars. The PCAM model is a well-known model developed by Chaiken and Dormont that is meant to decide the number of patrol cars to allocate to each district for any particular hour of the week [2]. Each iteration of the simulated annealing algorithm generates a prospective solution, and if the prospective solution's objective value is better than the current solution, it becomes the new current solution. Every current solution must be feasible, for which the authors specify conditions. These conditions of feasibility are response time, size, contiguity, compactness, and convexity. When all these feasibility constraints adhere to their respective rules, the simulated annealing algorithm can be applied. The authors test their approach with a case study on the Buffalo Police Department and find that their model improves district designs and lowers disparity among officers' workloads.

Chu, Wu, Zhang, and Wan [3] developed an algorithm that was important in the development of the algorithm we propose in this paper. The authors' analysis focuses on optimizing population equality and regional compactness of legislative boundaries. Addressing gerrymandering, the paper proposes a novel approach to a similar optimization problem, called the Colonial Algorithm (CA). The CA works by modeling the growth of colonies of bacteria. Bacteria are seeded in culture dishes and form around generators (seeds). "Hungry" colonies

expand intensely and "full" colonies behave indolently. Expanding speed is determined by the "appetite" of the colony and competitions arise between colonies as resources become scarce.

Zhang and Brown, [11] create a redistricting algorithm similar to the CA. However, their algorithm places the initial seeds on concentric circles in the region of focus and rather than grow police beats to equalize workload, they design the districts based on a compactness score. Our problem's objective differs as the YPD asked us to consider officer workload and historical data in our methodology.

3   Methodology

The overall method we use to determine the new police beats for the Youngstown Police Department is a novel variation of cellular automaton. In our case, we want to optimize equality of reported crimes and their priorities and regional compactness of police beats. As an adaptation of the CA and a variation of the model in Zhang and Brown, we propose our own Cellular Growth Model. Similarities between the models include seeding generators, growing seeds that have not reached their capacity (are "hungry"), not growing seeds that have reached their capacity (are "full"), equalizing expanding speed among different seeds, and breaking ties among beats which compete for cells [3]. However, our model differs in the way we select seeds and the way we select cells into which each seed grows. To be more precise, our model uses a triangle clustering protocol rather than concentric circles in order to select our initial seeds. We also use a weight matrix formulated by historical data, described in Section 3.2, to determine how the seeds grow. These differences make our model unique to the problem we are attempting to optimally solve.

We perform a historical data analysis on one year of calls for service data made to the Youngstown Police Department to determine the feasibility of the current police beats. The high crime areas are determined by using the historical data to create a weight matrix. The cells in this weight matrix represent a measure of the amount and severity of the crime in a specific geographical area in the city of Youngstown, relative to the crime in the whole city of Youngstown. This weight matrix is used as a reference for the geographical matrix created by this cellular automaton model, hereon referred to as the Cellular Growth Model.

The Cellular Growth Model grows the new police beats on the geographical matrix using the weights in the weight matrix when creating the beats and is discussed in further detail in Subsection 3.3. An algorithm in pseudo-code is provided for a brief overview of the growth model and the algorithm is explained in further detail.

To test and train our model we use a subsample of the data to generate new beats and then use the remaining data to test crime patterns against the generated beat alternatives. We select the alternative with the lowest mean squared error between the in- and out-of-sample data. Bootstrapping the data allows us to report the measures of dispersion, specifically the coefficient of variation among the beats, and its confidence intervals.

3.1 Historical Data Analysis

The historical data account for calls for service made to the Youngstown Police Department. After removing duplicate calls and emergency calls outside the jurisdiction of responsibility, the data set includes 79,298 calls. Officers from the YPD responded to these calls between November 2013 and November 2014.

The YPD assigns each call in the data set a rating of severity, called its priority. The classification system that the Youngstown Police Department uses assigns each call a priority of 0, 1, 2, or 9. Calls with the lowest level of severity are assigned a priority of 9. A priority 9 example is a complaint of a trash can being left out past the pick-up day. Priority 2 calls are more severe. An example of this type of call is a call of an unarmed shoplifter. The next level of severity is a priority 1 call, which includes a domestic fight in progress. The highest level of call severity is a priority 0 call, such as a homicide.

Figures 1, 2, 3, and 4 describe the variation in total calls across the original 12 beats, by priority of the call, by day of the week and by hour of the day.
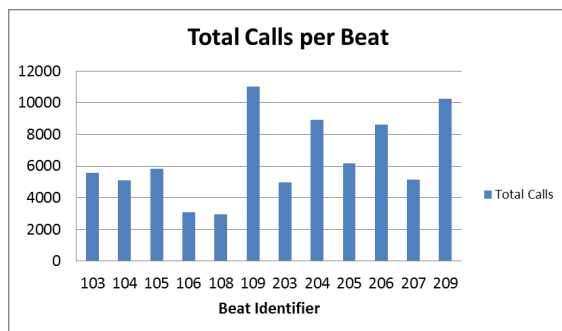

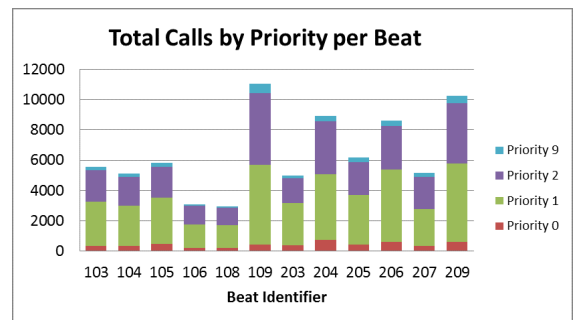Figure 1: Variation in total calls across the current beats.

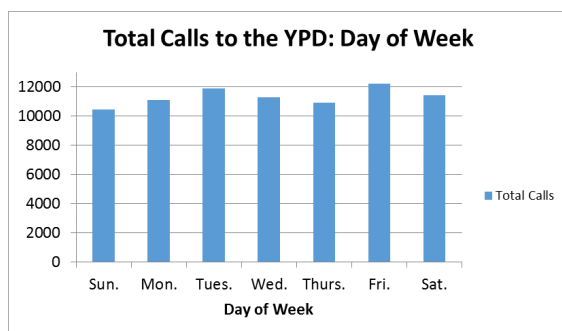Figure 2: Variation in total calls by priority of the call.

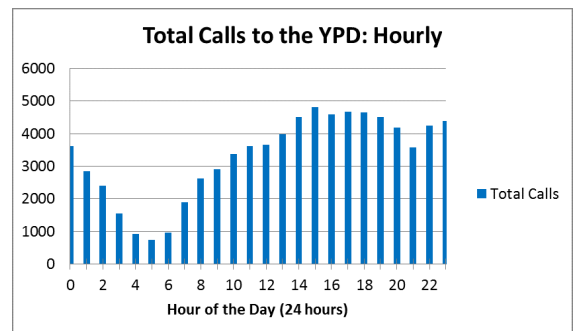Figure 3: Variation in total calls by day of the week.

Figure 4: Variation in total calls by hour of the day.

Notice that the number of calls responded to in each beat are not equal in number or

priority weight. Further, call trends exist for the entire data set across time. We use this historical spatial data to create new beats.

3.2 Weighting the Cells

We begin by creating a 250×168 matrix representing the geographical area of Youngstown. Each cell has dimensions 64 by 48 meters, for a total area of 3072 square meters (0.0012 sq miles). Cells not in the YPD's jurisdiction are assigned a weight of -1 (as are natural barriers, such as rivers, parks, highways, and barriers requested by the YPD). The historical calls for service data and their corresponding priorities are assigned to the cell in the matrix that represents the geographical location of the crimes. For each cell, we then calculate the total number of calls and the sum of the weight of priorities of those calls defined in Table 1. The scale relating priorities to weights is based on suggestions from experienced police officers. For example, after consulting with the YPD, four priority 9 calls are deemed equivalent to one priority 0 call. Thus, if a priority 9 call carries a weight of 2, a priority zero call must carry four times the weight of the priority 9 call, hence a weight of 8.

Priority	Weight
0	8
1	8
2	5
9	2

Table 1: Priority scale and corresponding weight assigned to each priority in the sum of call priorities.

Let $MaxTotalCalls$ refer to the maximum number of calls found in any one of the 42,000 cells; similarly, let $MaxPrioritySum$ represent the maximum priority sum found among all 42,000 cells. Each cell is then weighted according to the metric

$$\text{Cell } i \text{ Weight} = \frac{1}{2}\left(\frac{\text{TotalCalls}_i}{\text{MaxTotalCalls}}\right) + \frac{1}{2}\left(\frac{\text{PrioritySum}_i}{\text{MaxPrioritySum}}\right)$$

where i corresponds to the ith cell. After computing the ratio of the total calls in the ith cell to the $MaxTotalCalls$ and the ratio of the priority sum of the ith cell to the $MaxPrioritySum$, we consulted the YPD to decide how much weight should be attributed to call volume versus call priority. When responding to calls, the YPD considers call priority and the current call volume to be of equal importance, thus an equally weighted average of the two measures reflect their actions in practice. Hence each carries a weight of 1/2 in the formula. During our sensitivity analysis, we tested other weights and found no substantial changes in the results. This was due to high correlation between the priorities and call volumes.

3.3 Cellular Growth Model

The Cellular Growth Model uses a 250×168 matrix, described in Section 3.2, that represents the geography of Youngstown, Ohio. The geographical matrix initially contains a 0 or -1 in each cell; 0 represents available land for the beat and -1 represents geographical boundaries. Boundaries include rivers running through Mill Creek Park, a long, narrow park system stretching the entire length of the city, and other areas severely limiting response time. These establish natural barriers for the creation of the beats. We outline the steps of the algorithm in the following subsections.

3.3.1 Place Initial Points

In order to place our initial points (seeds), we make a visual of the weight matrix to see clusters of high crime areas. Figure 5 is a visual of the weight matrix with the geographical boundaries used in our growth model. The geographical boundaries in Figure 5 are represented with cells containing the numeric value -1. The cells outside of the geographical boundaries, but in Youngstown, are assigned values greater than or equal to 0 and less than or equal to 1 based on the metric described in Section 3.2.
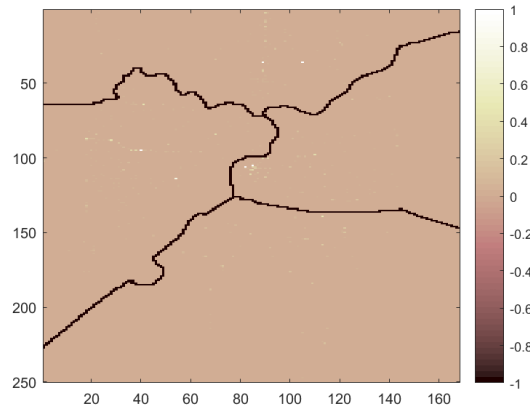

Figure 5: Visualization of the weighted matrix with geographical boundary that include rivers, highways, and other areas difficult to pass through in a timely manner.

We use Figure 5 to select our initial seeds. Observing the clusters of higher weighted cells, we place seeds using a triangular clustering protocol. Based on the visualization, there are only a few spots where there is higher crime in comparison to the entire area. We take into account that we have a resource constraint on the number of possible beats, as required by the Youngstown Police Department, and we observe a small number of areas with the highest metric (around 0.4). Therefore, we begin by placing seeds near cells with metrics larger than 0.4. We are limited to a maximum of 14 beats, with a preference of 13 beats by

the YPD. When placing the initial seeds, we must consider that the resulting beats must be contiguous, have equal workload (weight totals), and they must not be geographically large. We define the triangular clustering protocol as the following, while keeping in mind the previously listed restrictions:

1. Observe clusters of high weight in the weight matrix.

2. Place three seeds around each cluster to create a triangle around the cluster.

3. Place the remaining seeds to evenly distribute the geography (area) of each beat.

We place the seeds manually by marking the corresponding cell in the geographical matrix with a number between 1 and the total number of beats for that alternative (either 13 or 14). These numbers represent the beat assignment of the selected cell. No seeds are placed on a geographical boundary. After selecting our initial seeds, we allow the growth model to run, using the initial seeds as the starting point for each beat. The beats then grow simultaneously from each of these seeds until they are capped by the constraints or blocked in by other beats.

We continue selecting initial seeds and running the growth model in this fashion on a trial basis until we have created visually pleasing beats that satisfy the constraints listed in Section 3.3.2. For example, Figure 6 illustrates the positions of the selected initial seeds of one trial, whose result can be seen in Figure 7. We present this trial as our first possible beat solution, or alternative.
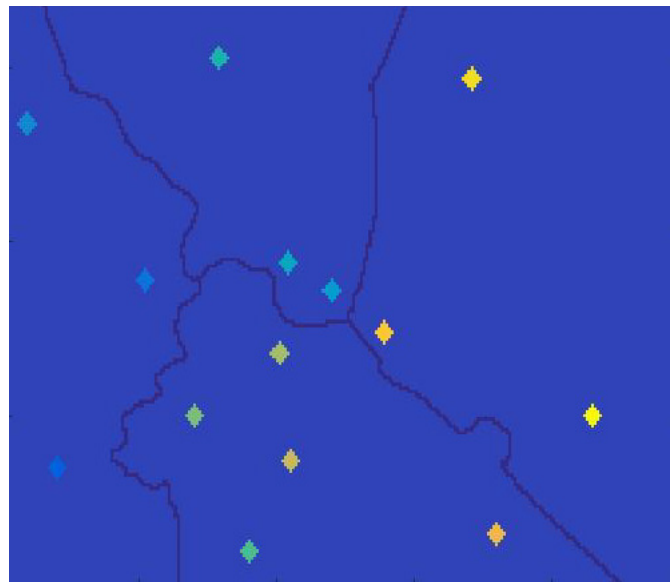

Figure 6: The initial 14 seeds of Alternative 1

Selecting areas for the initial seeds is an important step in our methodology, because the Cellular Growth Model is very sensitive to the placement of the initial seeds. The entire

result could change by moving an initial seed a few cells in any direction. The impact of differences in seed placement in Alternatives 1 and 2 can be seen in Figures 7 and 8.

We choose to place seeds using a triangular clustering protocol because it is simple to explain to stakeholders and easy to implement. We do not have many clusters and we do not need to place many seeds. As the number of clusters and the number of seeds increase this method becomes more complex. It is not the best approach, although it produces quick results. The seed selection methodology could be improved in future research. One suggestion is to perform Monte Carlo simulations on seed selection. The set of output from the Monte Carlo simulations could then be compared to determine if our Cellular Growth Model consistently produces similarly shaped and partitioned beats. We also suggest possibly implementing a Voronoi diagram algorithm. Researchers can place seeds in crime "Hot Spot" cells and generate a Voronoi partition of the geography of Youngstown. Use of Voronoi diagrams may also be fitting for a police department with multiple stations located across its area of jurisdiction (the YPD has only one central station).

3.3.2 Apply Cellular Growth

After placing our initial seeds, we grow each seed. Each seed grows only one step at a time until all of the seeds have been incremented by one step. In each time step, the amount of cells updated is determined by how many adjacent cells are available to be claimed by the cells that make up each seed (or beat). This process continues in incremental fashion until all of the cells in the matrix representation of Youngstown have been claimed by a beat or are a region of geography that cannot be claimed. The growth is subject to the following constraints.

1. There is a cap on the weight in each beat; this is simply the total sum of weights across all cells divided by the total number of beats (13 or 14, depending on the alternative) plus the small quantity of 0.8, which we call epsilon (our total sum of weights is approximately 77.59). We call this a global cap, which allows the weight of each beat to be within epsilon of the weight that would equalize workload perfectly among the beats.

2. We institute a local cap, which controls the acceptance of cells on an iterative basis, set at 1. The growth is programmed to take cells with the highest weight first for each iteration in the growth process. This orders the cells eligible for acceptance, and combines high-weight cell accumulation with the grabbing of low-weight cells that will not exceed the local growth cap. This aids in the compactness of the new beat regions.

3. One of the more necessary constraints is preventing any growth of beats across areas that are impractical to traverse. Places such as the Mahoning River and Mill Creek Park form natural boundaries. Honoring this constraint entails placing a value of −1 in the cells representative of the respective boundaries.

4. Contiguity of the beat regions is implemented. No beat is to have any islands, that is, any other geographically separate region that is in any way identified as being in the same beat.

We grow each seed one step at a time in the following way.

1. Identify the possible cells that can be taken by the current seed (beat). These cells contain a 0 and are adjacent to at least one of the cells already contained within the beat.

2. We select a possible cell only if adding that cell to the current beat does not put the beat over the "local" capacity weight specified or the "global" capacity weight specified plus the epsilon value of 0.8. The local cap will allow our beats to maintain a polygonal form. The global cap will allow our beats to have as equal workload as possible. The epsilon value of 0.8 is a changeable parameter that can be used to allow the fully grown beats to be contiguous.

3. Update the geographic matrix to represent the cells that we have selected. We make these cells equal to the numeric value that represents all cells in the current beat.

The algorithm is provided in pseudo-code form in Section 3.3.3. We chose to implement the algorithm in MATLAB and a video demonstration of the algorithm in action is available at https://www.youtube.com/watch?v=U_H2_TbuxSg.

3.3.3 Cellular Growth Pseudo-Code

Algorithm 1: Cellular Growth Model

---

epsilon = 0.8;
global_cap = ((*total_weight*)/(*num_seeds*)) + *epsilon*;
Initialize the geographic matrix with -1s along the boundary and 0s elsewhere;
Initialize the weight matrix based on pre-computations;
while *beats can still grow* **do**

    /* stop when all beats will exceed global cap on next iteration   */
    foreach *beat* **do**

        if *beat_total < global_cap* **then**

            /* *beat_total* is the sum of the weights of the cells in the
                weight matrix that make up the beat, determined by the
                geographic matrix   */
            local_cap = 1;
            adj_cells = find adjacent cells to beat cells;
            /* *adj_cells* is determined by the geographic matrix and they
                must be 0 to be chosen   */
            foreach *adj_cell* **do**

                adj_cell_weight = find *adj_cell* weight;
                /* *adj_cell_weight* is found from the weight matrix   */
                weights + = *adj_cell_weight*;
                /* *weights* is a list of weights   */
            end
            sort(*weights*) ;   /* high to low */
            current_weight = first weight in *weights*;
            current_sum = 0;
            while *not at weights.end* **and**
            current_weight + beat_total + current_sum < global_cap **and**
            current_sum + current_weight <= local_cap **do**

                current_sum + = *current_weight*;
                current_weight = next weight in *weights*;
            end
            beat_total + = *current_sum*;
            /* update the running total for the beat   */
            beat + = *adj_cells* with matching selected weights;
            /* mark the selected *adj_cells* (using indices) with the
                appropriate beat number in the geographic matrix   */
        end
    end
end

3.4 Training and Testing Our Cellular Growth Model

To train and test our model, we take a simple random sample of 80% of the data before applying our Cellular Growth Model. This random sample is taken without replacement and hereon we refer to this 80% sample of the data as the "in-sample data" as it is used within our model to generate beat alternatives. The remaining 20% of the data, which we hereon refer to as the "out-of-sample data," is then set aside to use after we generate the beats to assess the possible alternatives and select the best one.

Applying the Cellular Growth Model to the in-sample data, we generate five alternatives and calculate the total sum of cell weights per beat. Using the remaining 20% of the data, the out-of-sample data, we independently calculate the total sum of cell weights per beat. The beat alternative with the lowest mean squared error (MSE) between the training data and testing data is then selected as the alternative of choice.

We compare this alternative's summary statistics to the original beat layout that the Youngstown Police Department had been using. Bootstrapping the in-sample data, we generate confidence intervals about the coefficient of variation to assess the differences between the original beat arrangement and the alternative with the lowest mean squared error. In Section 4, we present these results and then proceed to present the results to the proposed YPD problem in Section 5.

4 Results

We begin by generating a simple random sample (SRS) of the crime data without replacement. Using this SRS of 80% of the data, the in-sample data, we apply our Cellular Growth Model and generate five alternatives with varying number of beats. These are presented in Figures 7, 8, 9, 10, and 11. We then export the alternatives' total weight per beat, that is the sum of all cell weights within a given beat in each alternative.
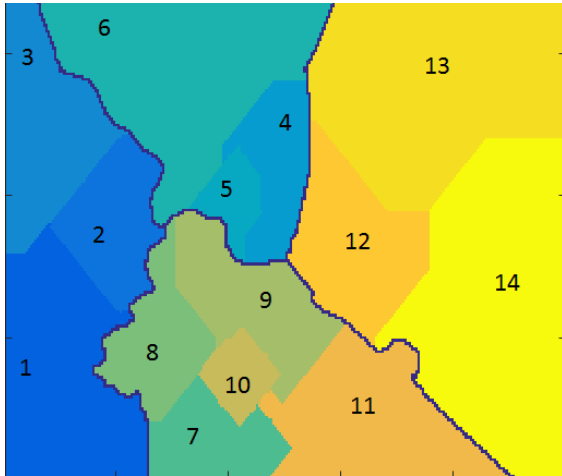
Figure 7: Alternative 1 beats generated by Cellular Growth Model.
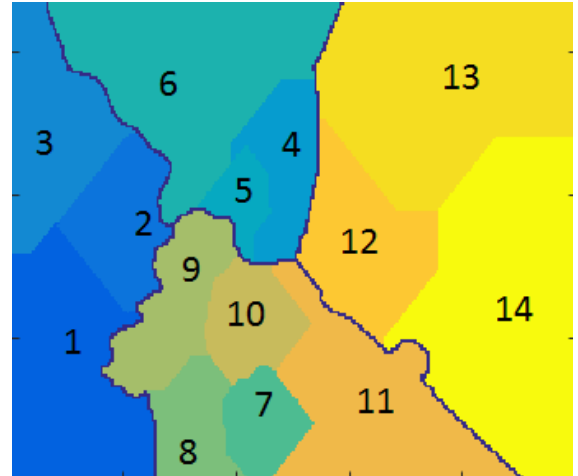

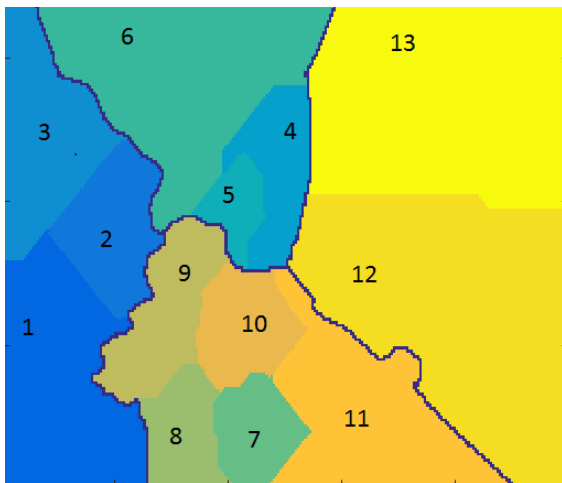Figure 8: Alternative 2 beats generated by Cellular Growth Model.

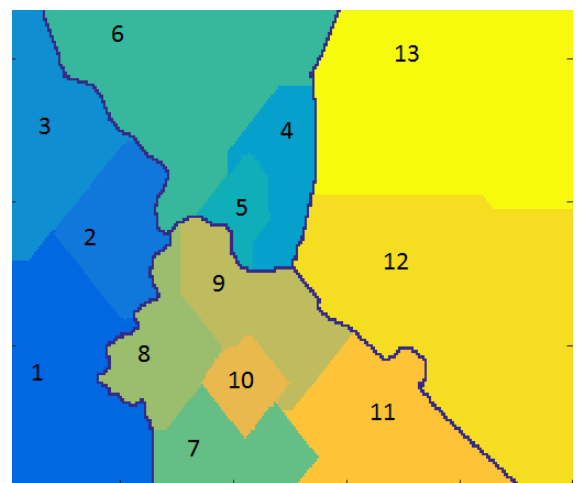Figure 9: Alternative 3 beats generated by Cellular Growth Model.

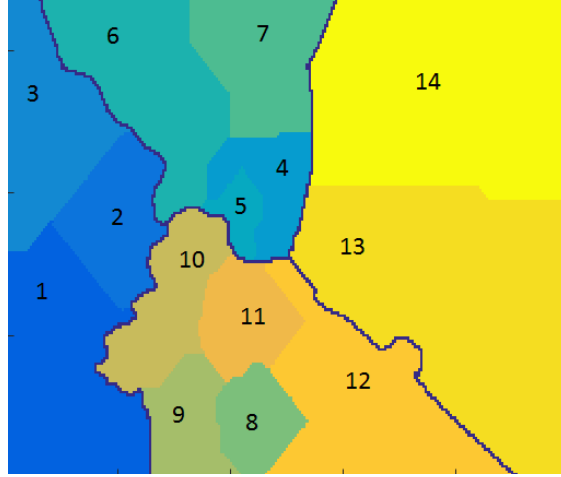Figure 10: Alternative 4 beats generated by Cellular Growth Model.

Figure 11: Alternative 5 beats generated
by Cellular Growth Model.

The remaining 20% of the data that was not used to generate the beats is then matched
to each of the proposed new beat alternatives, this is the out-of-sample data. These 15,863
calls for service are used to generate the metric weight per cell. The sum of all cell weights
of the out-of-sample data in each beat is then compared to the sum of all cell weights from
the in-sample data in each beat. The values of the beat sums in both the in-sample data
and out-of-sample data are similar in size because the cell weights (which are summed) are
similar in size. This is because each data sets' cell weights are relative to the data sets'
maximum number of calls in a given cell and maximum priority sum in a given cell.

We calculate the mean squared error between the in-sample beat sums and the out-
of-sample beat sums. The mean squared error is defined as the mean squared difference
between the in- and out-of-sample beat weights. If you let n to be the number of beats in a
given alternative, the calculation for the MSE is shown in Equation 1:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\text{In-SampleMetricSum}_{\text{beat } i} - \text{Out-of-SampleMetricSum}_{\text{beat } i})^2. \qquad (1)$$

Each alternatives' mean squared error is displayed in Table 2.

	Alt. 1	Alt. 2	Alt. 3	Alt. 4	Alt. 5
Mean Squared Error	0.1018	0.0934	0.1049	0.1116	0.0988

Table 2: Calculated Mean Squared Errors for each proposed alternative.

The alternative with the smallest mean squared error is the alternative that performs best
with out-of-sample data. In our case this is alternative two. To further assess this alternative

we bootstrap the in-sample data that was used to generate the alternatives. Bootstrapping requires that we resample the data with replacement, generating 1000 differing sets of calls for service data of equal size. These data are then used to generate confidence intervals about the summary statistics of the proposed Alternative 2. As the goal of our project was to reduce the variation of workloads across beats, we are particularly interested in the standard deviation and coefficient of variation of the proposed Alternative 2. Statistical results are presented in Table 3; 95% confidence intervals are provided in parentheses. As shown, the beats generated by the Cellular Growth Model feature a lower standard deviation and coefficient of variation than the original YPD beat arrangement. The original value of the coefficient of variation does not lie within the 95% confidence intervals; thus indicating an improvement in the disparity of workload across the police beats in Youngstown, Ohio.

Summary Statistics	Original	Alternative 2
Number of Beats	12	14
Mean Weight of Beats	6.394	5.194 (5.172, 5.215)
Standard Deviation	2.338	1.327 (1.317, 1.337)
Coefficient of Variation	0.366	0.255 (0.254, 0.257)

Table 3: Summary statistics for original beats and Alternative with lowest MSE.

5 YPD Application and Implementation

To honor the request of the YPD to use all available data in the generation of beats for their implementation, we present here the results of our application of this project to their proposed problem. We shared five possible alternatives with the YPD; these are similar to the alternatives presented in Section 4 as the initial seed placement is the same.

The YPD then selected one alternative to implement in January 2016. This alternative is featured in Figures 12 and 13. As Sarac et al. [9] use US Census block groups as the units of the division of the city for new beats, we chose census blocks to comprise the final beat assignments for the YPD as they are the smallest unit for which the US Census Bureau aggregates population information. In addition, census blocks follow local geographies including roads, streams, and bodies of water. Beats based on census blocks are more appropriate for actual use by the Youngstown Police Department. The final beat map, Figure 13, is generated by assigning beats to US Census blocks within Youngstown, instead of US Census block groups. We map our MATLAB solution to the census blocks to aid in the officers' visualizations of our result and to help them employ this alternative in the field.
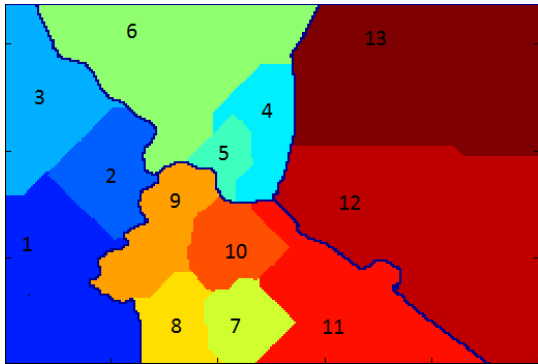
Figure 12: The alternative generated by the Cellular Growth Model which the YPD selected.
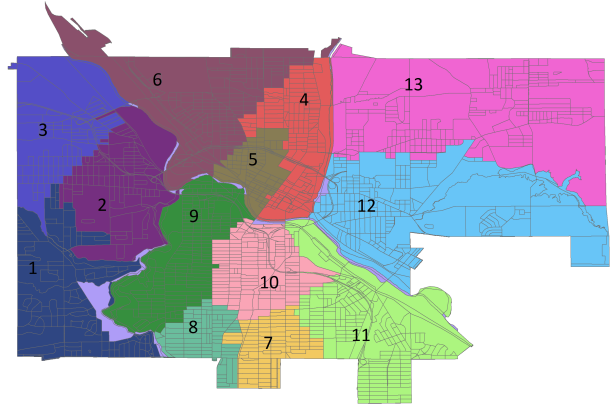

Figure 13: The YPD's selection with beats mapped to city census blocks.

5.1 Statistical Analysis on New YPD Beats

To assess the five new beat alternatives that we proposed to the YPD, we calculated the metric that is used to weight each cell for the entirety of the original beats and the five new alternatives, using the alternatives directly from MATLAB. For each alternative, we calculated summary statistics on the beat metrics and compare them to the original. In Figures 14 and 15, we graph the metric across all beats in the original arrangement and the YPD's selection of solution. A comprehensive collection of the summary statistics appears in Table 4.
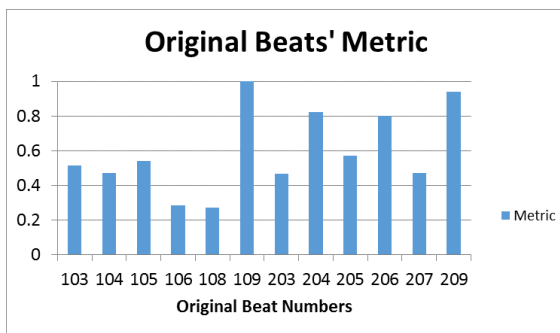

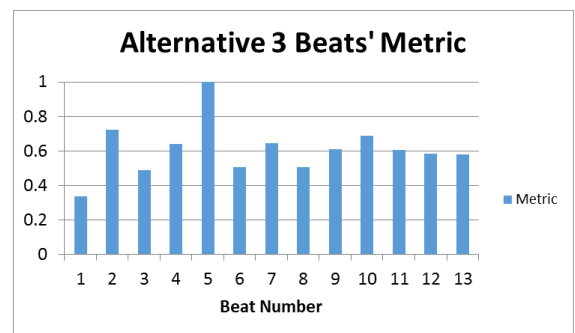Figure 14: Metric values for original beat arrangement.

Figure 15: Metric values for the YPD's new beat arrangement.

Summary Statistics	Original	YPD Selection
Number of Beats	12	13
Mean Weight of Cells in Beat	0.596	0.609
Standard Error	0.069	0.043
Median	0.526	0.606
Standard Deviation	0.240	0.154
Sample Variance	0.058	0.024
Range	0.726	0.663
Coefficient of Variation	0.403	0.253

Table 4: Summary statistics for original beats and YPD selection.

To assess how equitable the workload is across the beats in the different alternatives proposed to the YPD we consider the standard deviation of the metrics. The standard deviation of the metric is lower than the standard deviation of the metric for the original layout. To interpret this we calculate the coefficient of variation. We conclude the problem of high variance across beats has been reduced.

6   Implications and Future Work

This project's goal was to propose new beats to equalize the workload for YPD officers. We aided in the implementation of one of the proposed alternatives, effective January 2016. The YPD's implementation include some manual adjustments of the boundaries using officer intuition and landmark and street comparisons of the contents of the beats.

In future work, we plan to expand upon our current methodology by automating the selection of seeds and thus increasing the number of trials run. While the number of beats we use for this paper is constrained by the YPD budget, we hope to explore the changes in output if we relax the number of initial seeds allowed and thus beats; this will allow us to test the possibility of merging beats. We also hope to test other methods to create the new beats, such as partitioning the geographic area of Youngstown into the new beats by Voronoi diagrams. We will then compare these new beats with the results presented here. Revisiting this project after more data become available, we will be able to compare our new beats' performance both in and out of the initial data sample. Additional future work includes scheduling the officers with an optimization model and/or computer program as in Chaiken and Dormont [2] and Place [8].

7   Acknowledgements

MAA and SIAM. A recorded supplementary presentation of our work is available on the MAA PICMath site: http://www.maa.org/pic-math. We acknowledge Kacie L. Carlson for her contribution to the historical data analysis section of this paper. We also remain very grateful for the guidance suggested by our anonymous reviewers.

References

[1] A. A. Braga, *Protecting repeat victims*, in Problem-oriented policing and crime prevention, Criminal Justice Press, 2002, pp. 87–106.

[2] J. M. Chaiken and P. Dormont, *A patrol car allocation model: Capabilities and algorithms*, Management Science, **24** (1978), pp. 1291–1300.

[3] H. Chu, Y. Wu, Q. Zhang, and Y. Wan, *Colonial Algorithm: A quick, controllable and visible one for gerrymandering*, in Information and Automation, Springer, 2011, pp. 424–430.

[4] K. M. Curtin, K. Hayslett-McCall, and F. Qiu, *Determining optimal police patrol areas with maximal covering and backup covering location models*, Networks and Spatial Economics, **10** (2010), pp. 125–145.

[5] S. J. D'Amico, S.-J. Wang, R. Batta, and C. M. Rump, *A simulated annealing approach to police district design*, Computers & Operations Research, **29** (2002), pp. 667–684.

[6] G. Farrell and W. Sousa, *Repeat victimization and hot spots: the overlap and implications for crime control and problem-oriented policing*, Crime Prevention Studies, **12** (2001), pp. 221–240.

[7] N. I. of Justice, *Predictive Policing*, 2014.

[8] J. Place, *Evaluating police department policy decisions using a simulation model of sworn officer deployment*, International Journal of Police Science & Management, **9** (2007), pp. 341–356.

[9] A. Sarac, R. Batta, J. Bhadbury, and C. Rump, *Reconfiguring police reporting districts in the city of Buffalo*, OR Insight, **12** (1999), pp. 16–24.

[10] J. Wilson and G. Kelling, *The police and neighborhood safety: Broken windows*, Atlantic Monthly, (1982).

[11] Y. Zhang and D. Brown, *Police patrol districting method and simulation evaluation using agent-based model & gis*, Security Informatics, **2** (2013), pp. 1–13.