

# Tracking the Spatiotemporal Spread of the Ohio Overdose Epidemic with Topological Data Analysis

Nicholas Bermingham \*  
The Ohio State University

David White †  
Denison University

Nathan Willey ‡  
The Ohio State University

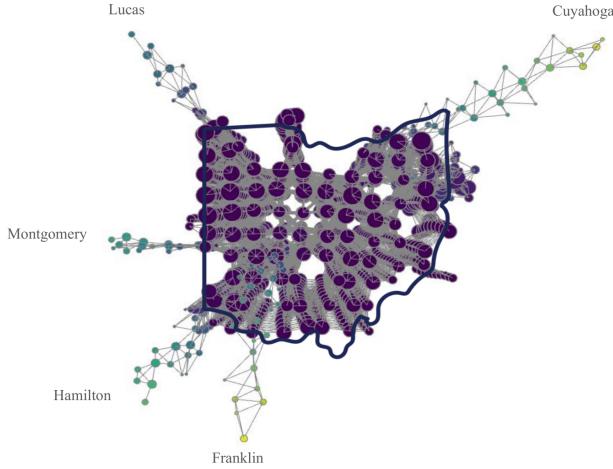


Figure 1: Mapper graph created from cumulative Ohio county drug overdose deaths. Geometry of the Mapper plot corresponds to the geography of Ohio and death spikes illustrate major population centers.

## ABSTRACT

In recent years, techniques from Topological Data Analysis (TDA) have proven effective at capturing spatial features of multidimensional data. However, applying TDA to spatiotemporal data remains relatively underexplored. In this work, we extend previous studies of disease spread by using the Mapper algorithm to analyze the Ohio drug-overdose epidemic from 2007 to 2024. We introduce a novel method for constructing covers in Mapper graphs of spatiotemporal data that respects geographic structure and highlights the time-dependent variables. Finally, we generate a Mapper visualization incorporating regional demographics to examine disparities in overdose deaths. Our approach effectively reveals temporal trends, overdose hotspots, and time-lagged patterns in relation to both geography and community demographics.

**Index Terms:** Mapper, Topological data analysis, Time-Series data, Spatiotemporal data, demographic data, drug overdose

## 1 INTRODUCTION

In this paper we investigate the effectiveness of applying the topological data analysis tool known as Mapper to data relating to the spatiotemporal aspects of the Ohio overdose epidemic from January 2007 to September 2024. We examine what insights into the Ohio overdose epidemic can be drawn from the topological and geometric features of Mapper graphs, such as connected components, spurs and holes.

Drug overdose remains a leading cause of preventable death in the United States. Between 1999 and 2022, annual overdose deaths rose more than sixfold, and in 2022 alone, over 107,000 Americans died from drug overdoses. The impact is not uniform: overdose mortality varies widely across time, geography, and demographic groups. Some of the hardest-hit areas, including counties in Ohio, have experienced multiple waves of the epidemic, each marked by distinct spatial and temporal patterns. As synthetic opioids make the drug supply increasingly volatile, it is important to understand where and when risk is increasing, and how it's spreading from one county to another. Modeling the spatiotemporal spread of overdose deaths can help public health agencies identify emerging hotspots early, allocate limited resources more effectively, and implement tailored interventions before crises peak. In a policy environment of shrinking federal support, tools that can capture the evolving geography of the epidemic are critical to preventing future deaths.

The spread of overdose deaths does not follow simple trajectories: it is shaped by changes in drug supply, local economic distress, demographic shifts, and public health infrastructure. Traditional approaches to modeling these patterns—including spatial regression and generalized linear mixed models—require strong assumptions about linearity, locality, and parametric form. High-dimensional models with large numbers of county-specific coefficients are often difficult to interpret and may obscure broader structural features. This landscape calls for flexible, non-parametric tools—like those from topological data analysis (TDA)—that can reveal large-scale structure in complex, high-dimensional data without relying on restrictive modeling assumptions.

The Mapper algorithm is a tool, introduced in [33], that has been used extensively in topological data analysis and constructs graphs that encapsulate the broad topological features of a given dataset. Mapper constructs a simplified representation of a high-dimensional dataset by clustering nearby observations and connecting them into a graph that reflects both global and local structure.

\*e-mail: bermingham.11@osu.edu

†e-mail: whiteda@denison.edu

‡e-mail: willey.106@osu.edu

This allows us to more easily visualize the data we are interested in and draw qualitative and quantitative insights from a given dataset.

In this paper we apply the Mapper algorithm to county-level demographic data and overdose death counts in Ohio across a fifteen-year period in order to understand the spatiotemporal spread of overdose deaths. Our approach is informed by [6], a similar analysis for COVID spread using Mapper. Additionally, we investigate whether similar insights can be gained from building Mapper graphs representing a kind of “demographic landscape” of Ohio by seeing how overdose death data correlates with certain demographic islands and how this data can change our “demographic landscape” when it is included in the construction of our Mapper graphs. We use visual properties of the resulting graphs to identify patterns of spread, population-normalized burden, and demographic correlates.

Our findings confirm and extend known patterns in Ohio’s overdose epidemic. We identify the emergence of rural hotspots near the Ohio River and the increasing burden in counties with high poverty and unemployment. We also observe a time-dependent trajectory in which overdose mortality spreads outward from urban centers into surrounding regions. These results align with the broad contours of prior work using more conventional modeling frameworks, discussed in [Section 1.2](#), while offering additional insights into topological features. Notably, we find evidence of a lagged association between economic stressors and subsequent overdose mortality, as well as visual confirmation of unique regional trajectories (e.g., in Scioto County) consistent with known historical events and public health reports.

By producing intuitive visualizations that highlight how and where the overdose crisis has unfolded, our approach has implications for future public health planning. Similar spatiotemporal analytics in Rhode Island helped motivate the state’s implementation of geographically targeted interventions, with overdose prevention resources deployed at the neighborhood level [24]. In that context, visual evidence of spatial spread helped convince decision-makers to shift from reactive policies to proactive containment. We hope the ideas in this work can serve a similar function in Ohio and other states, both by providing a new lens on the overdose crisis and by demonstrating the potential of TDA methods in public health research.

## 1.1 Outline of the Paper

In [Section 2](#), we begin by outlining the Mapper algorithm and the features that allow us to customize the Mapper algorithm for specific analysis of spatiotemporal data. We then discuss how we have sourced and organized our data relating to spatiotemporal spread of overdose deaths and the demographics of the counties of Ohio and the specific choices made in constructing Mapper graphs from these datasets. We also discuss how we have implemented a novel choice of hyperparameters in our application of the Mapper algorithm and note how our implementation differs from the work done in [6].

In [Section 3](#), we display all of the Mapper graphs we constructed, analyze the factors contributing to the emergence of topological and geometric features in the graphs and identify the counties whose data contributes to the emergence of these features. We also analyze how our different methodologies allow us to gain different kinds of insights from our chosen datasets and evaluate how both methods should be considered when undertaking these kinds of investigations into datasets with a temporal component.

Finally, in [Section 4](#) we draw conclusions regarding what our investigations tell us about the Ohio overdose epidemic, how our different methodologies could be used to study similar temporal datasets in the future and what future directions we could undertake in this project.

## 1.2 Related Works

To the best of our knowledge, this paper represents the first use of topological data analysis on drug overdose data. Previous works have modeled the temporal autocorrelation of drug overdose in Ohio using time series analysis methods, e.g., [22] created one model for all of Ohio, determining how the monthly number of deaths  $D_t$  depends on its own past. More generally, [29] used a generalized linear mixed model (GLMM) to opioid overdose death counts  $D_{c,t}$  in each (county, month) pair in Ohio. While this method is general enough to account for both spatial and temporal dependence, since  $D_{c,t}$  can be a function of any other  $D_{c',t'}$ , it faces challenges due to the large number of parameters involved. In [29], the potential for one county to affect its geographical neighbors is not explored and linear relationships between each  $(c,t)$  and  $(c,t-h)$  are assumed. Verifying linearity assumptions for these numerous relationships can be cumbersome, requiring looking at one scatterplot for each coefficient in the model. These challenges motivated our current use of Topological Data Analysis (TDA) techniques, which allow for the exploration of high-dimensional spatiotemporal data without relying on strict parametric assumptions or individual coefficient estimation.

Several strands of research have also addressed the spatial dynamics of the opioid overdose epidemic in Ohio. One standard technique for determining whether a time series  $D_t$  depends on its past is to compute an autocorrelation function, explained in [22], to calculate the correlation between  $D_t$  and the lagged time series  $D_{t-h}$ , for every  $h$ . The analogous procedure to detect spatial autocorrelation is Moran’s I-statistic, explained nicely in [39].

Andrew Curtis and members of the Begun Center for Violence Prevention have fit spatial models for drug overdose data in the Cleveland area [9, 25], at the census block level (or, using overlapping spatial filters, creating models that do not depend on the arbitrarily drawn census block boundaries). This kind of technique can produce heat maps and cartographic maps showing which areas are most at risk of overdose spikes, e.g., showing movement of the epidemic into African American neighborhoods by comparing heat maps in one year with the next year. However, the statistical models do not include the time dimension, and we are unaware of how these models can be used for forecasting future hotspots. Additionally, Adam Eck and his students at Oberlin college use machine learning models (e.g., random forests, gradient boosting, individual decision trees, SVMs, neural networks) to predict county-level overdose death hotspots [11], with the explicit aim of helping guide public policy and resource allocation.

In addition, it is possible to approach spatial and spatiotemporal autocorrelation using a Bayesian framework. Kline, Hepler, and their students have employed Bayesian statistical models to estimate spatial autocorrelation in opioid overdose deaths across Ohio counties, providing insight into geographic clustering and county-level risk factors [13, 17, 18]. These papers fit generalized spatial factor models, and look at the relationship between treatments for substance abuse disorder and drug overdose deaths, in each county. Their algorithm produces spatial weights for each county, which they interpret as a measurement of how much unmeasured information there is across counties, causing statistically significant differences that the model cannot explain. This work was generalized to add a temporal dimension by Ji [15]. Others have fit similar models in the Cincinnati area [19, 7], at the census block level.

There has also been work done at the national level and in other states, e.g., [35]. Several states have also attempted to model the spatiotemporal spread of overdose deaths. The most advanced appears to be Rhode Island, where academic researchers have teamed up with the state health department to develop the PROVIDENT system [24]. The state health department uses these predictions to optimize their deployment of overdose prevention resources at the neighborhood level.

Beyond statistical models, it is also possible to model spatiotemporal spread using Hawkes processes and other processes from dynamical systems. The middle author used these models to determine the spatio-temporal spread of protests in the USA [28] and in Ukraine [3].

Numerous previous papers have applied topological data analysis to other epidemics including the spatiotemporal spread of COVID [6, 14, 2], Zika [20, 34, 30], influenza [8], and other contagious diseases [38]. Although the mechanisms of spread of overdoses from county to county is very different from the spread of a virus from person to person, the success of these previous applications of TDA inspired our current analysis, especially [6].

## 2 METHODS

### 2.1 Mapper Algorithm

The Mapper algorithm is a tool from the field of topological data analysis developed by Singh, Mémoli and Carlsson [33] for qualitative analysis and visualization of high-dimensional datasets in a way that preserves topological features. The Mapper algorithm is a versatile tool that has been used in a variety of fields, such as economics where it has been used to detect inter-dependencies of factors involved in firm financial ratios [10], environmental science where it has been used to study factors contributing to harmful algal blooms [16] and the behavior of air pollutants [23], and in medical science where it has been used in studies of neuro-imaging [43] and brain network topology [36]. The Mapper algorithm has been used for qualitative analysis of spatiotemporal disease spread [6, 21]. Both of these papers apply it to data relating to the spread of Covid-19, but in theory this analysis can be broadened to the spread of any spatiotemporal data. Chen and Volic’s paper [6] served as a case-study for the use of the Mapper algorithm in analyzing disease spread and showed that the Mapper algorithm was able to visually capture the emergence of hotspots and provide insights about the spatial and temporal links between hotspots over time.

The Mapper algorithm works by taking a potentially high dimensional dataset  $X \subset \mathbb{R}^N$  and projecting this dataset onto a lower-dimensional space via a chosen filter function or lens,  $f : X \rightarrow \mathbb{R}^d$ . The algorithm then finds a finite cover of the range of the function via a methodology specified by the user i.e. we find sets  $U_i \subset \mathbb{R}^d$  so that  $\text{im}(f) \subset \bigcup_{i=1}^n U_i$ . A clustering algorithm, again specified by the user, is then used on the preimage,  $f^{-1}(U_i)$  for each  $U_i$  in the cover to create a set of nodes  $\{x_{ij}\}_{j=1}^{n_i}$ . These clusters then appear in the Mapper graph as vertices representing parts of our dataset that the clustering algorithm deemed to be close together. Edges between two nodes  $x_{ij}$  and  $x_{kl}$  are drawn in the Mapper graph whenever their intersection  $x_{ij} \cap x_{kl} \subset X$  is nonempty. This graph is then plotted in either 2 or 3 dimensions again depending on user specification, and this plot can be interacted with via rotations and zooming to allow the user to investigate connected components and identify nodes of interest. A simple example of this procedure is presented in Fig. 2.

We note that there are several choices that are left to the user when it comes to implementing the Mapper algorithm, notably the choice of filter function, cover, and clustering algorithm. This versatility is part of the strength of Mapper as it allows the algorithm to be customized to a researcher’s specific purposes. Arguably the most important choice is that of the filter function, as the overall structure of the Mapper graph is highly dependent on this choice as noted in [33]. Other choices do not change the output so fundamentally and can be treated as hyperparameters the user tweaks to find the most illustrative Mapper graphs. Some classical choices for the filter function listed in [33] are density and eccentricity estimators, though for our purposes we will typically be projecting onto certain variables in our dataset and noting the changes in the graph that result from focusing on these particular parameters.

The Mapper graph captures topological features of the given

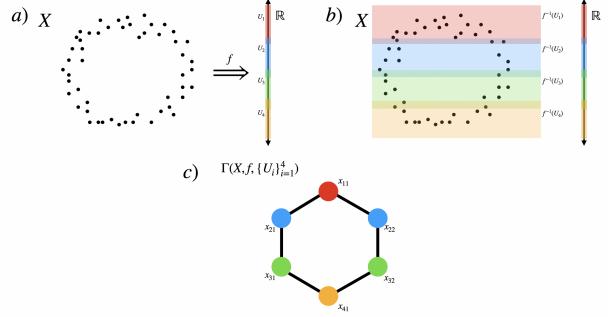


Figure 2: a) A dataset  $X \subset \mathbb{R}^2$  being mapped onto  $\mathbb{R}$  via the projection function  $f(x,y) = y$  and a covering of  $\text{im}(f)$  by 4 open sets  $U_1, U_2, U_3, U_4$ . b) The preimages  $f^{-1}(U_i)$  overlaid over the original dataset  $X$ . c) The resulting Mapper graph whose nodes  $x_{i,j}$  come from applying a clustering algorithm to  $X \cap f^{-1}(U_i)$  to find the related data points, and whose edges come from the nodes whose intersection  $x_{ij} \cap x_{kl} \neq \emptyset$

dataset including the number of connected components and the number of 1-dimensional holes, because of the underlying assumption that whatever clustering algorithm is used clusters data points that should represent elements of the same connected component. There are analogous constructions of higher dimensional simplicial complexes formed by adding a  $k$ -simplex whenever the intersection of  $k+1$  nodes  $\bigcap_{i=1}^{k+1} x_{n_i m_i}$  is nonempty, and these complexes are designed to capture topological features of a given dataset of higher homological dimension. We do not investigate higher dimensional homological features in this study, and instead focus on the emergence of spurs and connected components.

### 2.2 Ohio Overdose and Demographic Data

Our dataset on drug-induced deaths in Ohio comes from DataOhio tracking death records from the Ohio Department of Health’s (ODH) Bureau of Vital Statistics [26]. These data are reported monthly and our analysis covers the period from January 2007 to September 2024. It is important to note that the actual number of overdose deaths might not match the number in our data set, e.g., because drug overdose is sometimes unreported as a cause of death on death certificates [4]. The ODH and the Centers for Disease Control (CDC) try to correct for this, but missing data remains a potential concern.

The population data used for the spatial Mapper plots is yearly and comes from the Census 10-year estimates [40]. For the spatiotemporal results, the data points used in the projection to create the Mapper plots are 4-dimensional and of the following form:

(month, latitude, longitude, (normalized) cumulative deaths).

Each data point represents one month’s data for a given county whose center is the reported latitude and longitude.

In order to visualize the spread of the Ohio drug epidemic throughout the demographic regions of Ohio, we select key demographic indicators which we use to form the following 8-dimensional data points:

(year, population, % poverty, median-income, % unemployed, % white, cumulative deaths, normalized cumulative deaths).

Our choice of demographic features to consider was informed by other works [17, 29] which determined these features to be relevant to understanding issues pertaining to drug use in Ohio. Both the

percent of the population in poverty and the median income of a county are sourced from the Census SAIPE program [42]. Data on the percent of the population that is white and total population count come from the US Census American Community Survey (ACS) [41]. Data from both sources and are tracked yearly at the county level. Unemployment data comes from the Ohio Department of Job and Family Services through the Local Area Unemployment Statistics (LAUS) program [27]. Unemployment data is tracked monthly at the county level. For the sake of visualization, these features are aggregated to be yearly estimates. We use data from the years 2009-2023 inclusive. This change from the monthly data used for the spatiotemporal results is done for the sake of visual clarity in the final Mapper plots. Since we used publicly available data, with no identifying features, no IRB was required.

We remark that the work done in [6] calls for a renormalization of the data columns in order to ensure that no aspect of the data is weighted disproportionately. However, we find that in our chosen implementation of Mapper, the cubical cover divides the range of each column into  $n$  intervals of equal size regardless of any linear normalization of the range, thus the resulting Mapper graph is unchanged. As a result, we do not need to worry about column-normalization of our data. Finally, we remark that, in contrast to traditional statistical methods, the Mapper algorithm is relatively robust to data points with identical values in some of the coordinates. This is important given that our data points use features gathered on different time-scales; for example yearly population data with monthly death counts with static latitude and longitude. The Mapper algorithm clusters points with similar features into one node in the created graph, avoiding extraneous features in the graphs. As we will discuss later, the predictable and static structure of the latitude, longitude, and time coordinates can help to determine hyperparameters for our Mapper plots and provide a data-driven approach to the visualizations.

### 2.3 Implementation of Mapper for the Ohio Overdose Epidemic

We perform experiments on the two datasets outlined above with the goals of identifying significant trends in the Ohio overdose epidemic and exploring the utility of the Mapper algorithm to interrogate spatiotemporal and demographic-temporal data. Our first investigation into applying the Mapper algorithm to the Ohio overdose epidemic is to see if we can replicate the findings of [6] using the cumulative number of deaths due to overdoses in counties of Ohio. We utilize the cumulative death count as opposed to deaths in a given month to align more closely with the work done in [6] which showed that spikes in fatalities could be identified by the emergence of spurs in the Mapper graph and because its almost-continuous nature aligns better with the geometric aspects of the Mapper plot. Notably the fact that we are constantly increasing this parameter over time leads to the formation of a spur rather than noisy spikes which is an important aspect of the results in [6] we wish to replicate. We then replace cumulative deaths with cumulative deaths normalized by county populations in order to study trends in counties with lower populations that are otherwise hidden when looking at total deaths. In doing this, we both study whether areas of relatively high deaths would be similarly identifiable as spurs in the Mapper graph and if there are any differences in the trends of the epidemic when better incorporating low population regions. Finally, we investigate whether there are any insights to gain from grouping counties of Ohio by their demographic profiles over time, rather than their geographic profile. We accomplish this by building a Mapper graph out of the demographic dataset outlined above and assess whether there is a correspondence between the demographic information in these plots and rates of overdose deaths in these counties.

Our implementation of Mapper is in Python through the `mapper` functions available in the `giotto-tda` package [37]. For our choice of filter function we always project onto certain features of the data that are relevant to the specific investigation being undertaken at the time, as we will discuss. To cover the image of our filter function we choose a cubical cover which requires the user to specify a number of intervals  $n$ , and a percentage overlap  $p$ . The cover is then constructed by dividing the range of the function in each dimension  $i = 1, \dots, d$  into  $n$  intervals of equal size  $\mathcal{I}_i = \{I_{ij}\}_{j=1}^n$  such that the percentage of  $I_{ij}$  which overlaps with  $I_{i(j+1)}$  is  $p$  for all  $j < n$ . The resulting cover consists of high-dimensional rectangles of the form  $\prod_{i=1}^d I_{ij_i}$  where  $j_i \in \{1, \dots, n\}$  for all  $i$ . The values  $n, p$  are hyperparameters which are typically chosen to give the most informative Mapper graphs and will often require experimentation to find the best values to use. We use DBSCAN clustering which is the default clustering algorithm recommended in `giotto-tda`. The advantages of DBSCAN in comparison to more naive clustering algorithms are discussed in [31] and outside the scope of our work. Finally to produce visuals of the Mapper graphs we chose to embed them into  $\mathbb{R}^3$  rather than the plane in order to improve the clarity of our presentation of connectivity information. Embeddings into the plane were often cluttered with overlaps which obscured important connection information that is clearly visible when plotted in  $\mathbb{R}^3$  instead. Additionally, as our spatiotemporal plots are dominated by the 3 main parameters of latitude, longitude, and time, the data lends itself much more naturally to a 3-dimensional embedding than a 2-dimensional embedding.

In the case of creating Mapper graphs from spatiotemporal data similar to those in [6], we implement a novel informed choice of hyperparameters for our Mapper cover. The main observation informing our choice is that without the death information, the dataset we are using in our spatiotemporal plots is simply the geography of Ohio captured by the latitude and longitude coordinates of the counties of Ohio staying constant over time. As Ohio is a contiguous land mass and our time is varying continuously, the most representative plot of this data would be one totally connected component that varies in 2 dimensions by the spatial coordinates and changes continuously in a third dimension by the time coordinate. Thus we choose  $n$  and  $p$  so that the Mapper graph formed by projecting onto only spatiotemporal data represents this connectivity correctly and is the minimal amount of overlap required to do so. In doing so, we obtain a Mapper graph (embedded in  $\mathbb{R}^3$ ) whose horizontal slices represent the adjacency of Ohio's counties and whose vertical component only represents time as depicted in Fig. 3. Since our counties are geographically represented by their geospatial centroids in latitude and longitude, the creation of a fully connected Mapper graph is nontrivial and results from experimentation. When we then add overdose information into our projection our choices guarantee that any topological signatures such as holes or spikes in the Mapper graph are caused only by the death data and not irregularities in the location of county centers. Following this approach led us to using 10 intervals with 50% overlap. This approach is specific to the structure of the spatiotemporal data and, as such, does not extend to creating Mapper graphs for the Ohio county demographic data. For the demographic dataset we must manually choose hyperparameters that give visually informative Mapper graphs. For the Mapper plots formed by our demographic data we found that using 9 intervals with 45% overlap gave the most informative plots.

The specific filter functions we used in our experiment varied depend on the aim of our visualization. As mentioned above, when producing spatiotemporal plots our filter function was a simple projection map initially on just the three parameters of latitude, longitude and time to determine a meaningful choice of hyperparameters, and after this our filter function included projecting on either the cumulative deaths in the county or cumulative deaths normalized by county population. This methodology serves two aims:

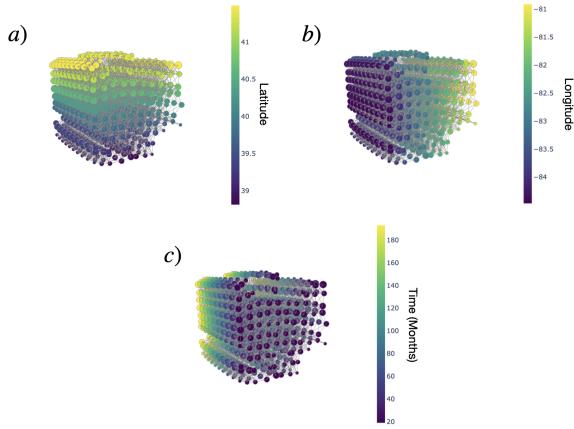


Figure 3: The Mapper graph formed by our spatiotemporal data with nodes colored by the cluster averages of a) latitude b) longitude c) month since January 2007

first, to show how Mapper can be used in an experimental fashion to see the effect of a variable on the Mapper graph given a “control” graph, and second, to investigate the nature of the spatiotemporal spread of overdose deaths in Ohio by analyzing these changes. When constructing our demographic plots, initially we choose our filter function as a projection onto the six discussed demographic features in order to create a demographic landscape of Ohio. We then analyze how the overdose deaths appear in these plots by coloring according to this data without having incorporated it into the filter function. Finally, we contrast this to the results obtained by incorporating cumulative deaths and then normalized cumulative deaths into our filter function in addition to the six data points above in order to evaluate the differences in the two methodologies and ascertain whether there are new insights into the Ohio overdose epidemic gained from the Mapper graphs obtained by this method. This allows us to further investigate the idea of using Mapper as an experimental tool observing changes induced on a control graph by altering the filter function and additionally interrogate the connections between demographic data and deaths due to drug overdoses in Ohio.

### 3 RESULTS

In this section, we analyze the trends in overdose deaths in Ohio using the Mapper plots created from both the spatiotemporal and demographic data by studying the 3D representations of the Mapper graphs constructed according to the methodologies outlined above. We divide our results into three main parts: spatiotemporal analyses based on cumulative death counts, based on population-normalized counts, and an analysis of demographic landscapes in relation to overdose mortality. Each approach illustrates the ability of Mapper to reveal different structural aspects of the epidemic’s spread across Ohio’s counties.

#### 3.1 Spatiotemporal Visualization of Overdose Deaths

We begin by using the 4-dimensional cumulative death dataset with the filter function set to just being the identity on  $\mathbb{R}^4$ . The Mapper plot created from this dataset is shown in Fig. 4 using the previously detailed hyperparameters.

It is clear that the broad structural features of this graph are inherited from the graph constructed without death information (Fig. 3) as they still generally show slices of the counties of Ohio persisting throughout time. We will refer to this piece of the Mapper graph corresponding to Ohio over time as the “main trunk”. The main geometric difference that occurs when cumulative overdose

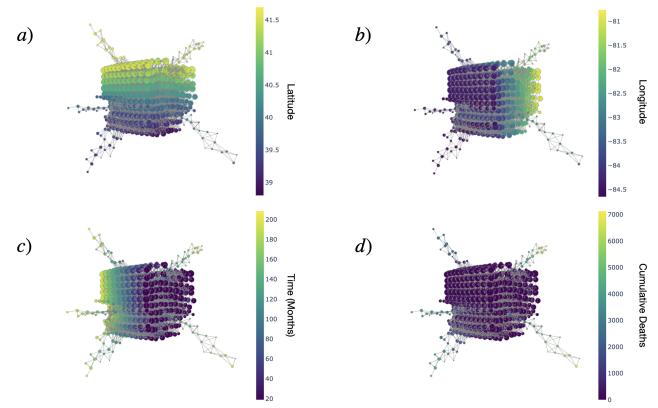


Figure 4: The Mapper graph formed by our spatiotemporal data with cumulative deaths incorporated in the filter function with nodes colored by the cluster averages of a) latitude, b) longitude, c) month since January 2007, d) cumulative number of deaths due to drug overdoses

death information is included is that counties with high populations (those containing major urban areas of Ohio) have high cumulative overdose death counts that are strictly increasing over time and so emerge as large spikes in the Mapper graph. These findings agree with the findings in [6] that major population centers were the main driver of spikes when Mapper is applied to unnormalized data (i.e., counts rather than counts per capita).

A labeled diagram demonstrating the correspondence between the geography of Ohio and the geometry of the graph is presented in Fig. 1.

The labels were determined by analyzing the nodes that contribute to these branches off of the main trunk, which showed that each branch was comprised solely from the data of one county, and these counties that branched off were: Franklin (containing the city of Columbus), Lucas (containing Toledo), Hamilton (containing Cincinnati), Montgomery (containing Dayton), and Cuyahoga (containing Cleveland).

Analysis of these emergent topological features of the Mapper graph can give us information about how our data is evolving locally in space and time. The origin of these spikes determines the geographic location of the county in which a spike in death data occurs. In Fig. 4 c) we see that the branches expand outwards with increasing time parameter as these counties’ cumulative death count grow to be very large relative to the surrounding counties. A spike whose nodes contain only information from one county and does not reconnect with the main trunk indicates a strong trend in death count relative to the surrounding area. A spike with more connected nodes containing data from multiple counties indicates a spatial cluster of counties whose death totals are evolving similarly in time. We can analyze the specific spurs in our Mapper plot to find county-level trends in the overdose epidemic. From the coloring in Fig. 4 c) we can see, for example, that Franklin county starts to spike around 2015 (month 100). As it never reconnects with the main trunk and the nodes only contain data from Franklin county, this indicates that the death count stays high relative to the surrounding counties during subsequent months. Although not all are visible from the angle in this figure, many of the other main population centers follow a similar timeline to Franklin County in overdose death spikes. However, in the case of the Cincinnati area we observe a hole at the base of the main spike formed by nodes containing data from the nearby Butler county connecting to this branch from the main trunk. This corresponds to the cumulative death count in this county temporarily catching up to where Hamilton County was in its earlier death

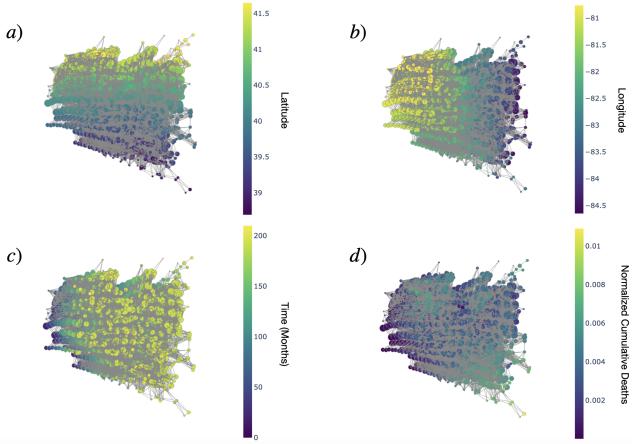


Figure 5: The Mapper graph formed by our spatiotemporal data with cumulative deaths normalized by the population of the county where the deaths occurred. West to east runs right to left here. The coloration of the nodes is by the cluster averages of a) latitude, b) longitude, c) Month since January 2007, and d) Cumulative number of deaths due to drug overdoses normalized by county population.

count, before the cumulative death count of Hamilton County substantially increased. These cross-county features demonstrate the Mapper algorithm’s ability to illustrate spread in time across many counties together.

We note that one limitation of visual analysis of Mapper plots is illustrated by Franklin county due to its central location in Ohio. When plotted it is visually difficult to see the exact month at which the spike peels off of the main trunk, even so, a more careful analysis of the nodes does show this information.

This analysis agrees with the kinds of results obtained in [6] confirming the utility of Mapper graphs in analyzing data relating to spatiotemporal spread of epidemics. As in [6] we also remark that the trends in spatiotemporal data most amenable to analysis via creation of a Mapper plot are those that are relevant to the geometry and topology of the dataset. In both studying the drug epidemic and the spread of the Coronavirus, this amounts to studying connected components, spikes, and holes in a cumulative variable. This analysis shows, as expected, that large population centers have large spikes in death data. In order to further explore the capabilities of Mapper in analyzing this kind of spatiotemporal data we next normalize the death count by population to factor out this variable from the visualizations.

### 3.2 Spatiotemporal Visualization of Population-Normalized Drug-Induced Deaths

In this section we construct a Mapper graph using a filter function that projects onto the spatiotemporal data of the counties of Ohio along with the cumulative deaths in a county normalized by population. This graph is presented in Fig. 5 with coloration given by the four variables of interest. We begin by noting that the images in Fig. 5 are viewed from the opposite perspective compared to Fig. 4, in order to highlight the spurs formed by higher normalized cumulative deaths which occur predominantly towards the end of the time period analyzed. As a result the latitude of the plot appears reversed from the traditional east-west perspective highlighted in Fig. 1.

The immediate visual takeaway is how much more chaotic Fig. 5 appears to be compared to Fig. 4, suggesting that the normalization process creates less distinction between more of the counties than the unnormalized data. This makes sense, because population is a major distinguishing feature between counties. We clearly see some notable spurs forming in the data, most prominently in

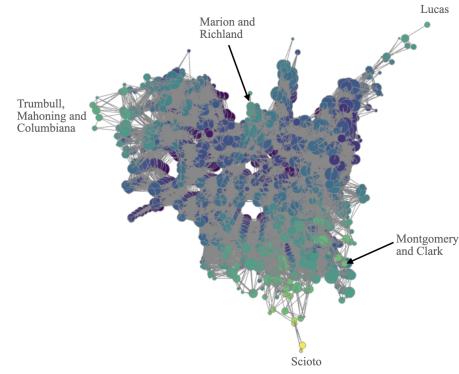


Figure 6: A Front View of the Mapper Graph of Cumulative Deaths Normalized by county population. Spurs are labeled by the counties whose data is contained within the clusters.

Southern Ohio near the Ohio river which forms the border between Ohio and Kentucky, and Ohio and West Virginia. A labeled diagram of this Mapper graph is presented in Fig. 6 which points to Scioto county as having the highest value of normalized cumulative deaths while the surrounding counties are less prominent with smaller peaks in areas around cities like Marion and Richland county north of Columbus, Mahoning, Trumbull and Columbiana just East of Cleveland, Montgomery and Clark, containing and near Dayton and Lucas containing Toledo.

This finding aligns with the county-by-county analysis of [29] as well as the middle author’s investigation of overdose death rates using SUDORS data [44]. These works identified counties in Appalachia, in the Dayton-Cincinnati corridor, and south/east of Cleveland, as areas with higher than average overdose death rates. For that reason, those same counties have been a focal point for harm reduction efforts in Ohio, such as the HEALing Communities project [12].

Another important feature is that these population-normalized spikes occur later in time than the large city-center spikes of the previous section, generally after 150 months whereas the spikes in cities start between month 80 and 120. The timing of these smaller peaks occurring in low-density areas relative to the peaks from cities shown in Fig. 4 may suggest a form of delayed spread. One interpretation of this is that the drug epidemic starts to spike in cities and later on flows into less populous areas of Ohio where the death toll is more significant relative to the county population. This shows that we can extend the findings of [6] to normalized death counts and obtain new insights into our data by analyzing the same kinds of topological and geometric features. Additionally by creating and comparing Mapper graphs from both raw death count and population normalized count we get further insights into the nature of the spread and how it concerns population. We can extend this idea to observe how spread relates to other important demographic features by creating Mapper plots with different demographic features as its inputs.

### 3.3 Mapper Visualization of Demographic Data and Overdose Deaths

We now present the results from our experiments relating the demographic profiles of the counties of Ohio to the overdose death trends in time. These results are exploratory in nature and expand on the utility of the Mapper algorithm as a way to investigate temporal data in relation to a variety of features.

#### 3.3.1 Visualizing the Demographic Landscape

We begin by constructing a Mapper plot from the demographic information described in our methods section, a snapshot of which is

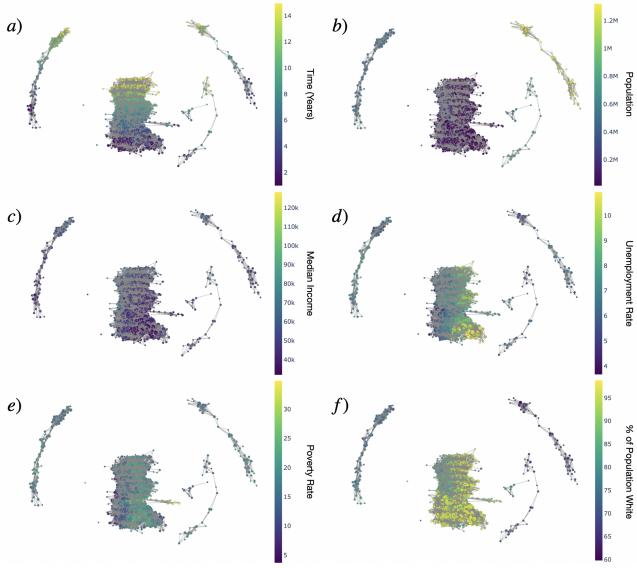


Figure 7: Mapper plots of the demographic profile of Ohio colored according to a) Time in years, b) Population, c) Median Income, d) Unemployment rate, e) Poverty rate, f) percentage of the county that is white.

depicted in Fig. 7, with nodes colored by each projected parameter.

This Mapper graph gives the structure of Ohio at a county level with respect to the important demographic features we chose to project onto. Close and connected points represent counties with similar demographics in a similar time period. Some of the immediate observations that we can draw from these figures are related to the connected components which seem to be clearly delineated by population as evidenced in Fig. 7 b). By investigating the clusters that constitute these connected components, we can determine that one of these components is comprised of data from Franklin and Cuyahoga county, the most populous counties in Ohio containing the cities of Columbus and Cleveland respectively. Another connected component is composed solely of data from Hamilton county containing the city of Cincinnati next to one smaller connected component comprised of data from Summit county containing the city of Akron. The final skinny connected component is comprised of data from Summit, Lucas and Montgomery County containing the cities of Akron, Toledo and Dayton respectively. All other counties have conglomerated into the bulky central component in the figure which we will call the main trunk. The demographic plot with these components labeled by their constituent counties is presented in Fig. 8. We note that Summit County appearing in two different connected components seems to simply be a quirk of our choice of hyperparameters together with the fact that Summit Counties financial demographics in years 6-10 of our time period were sufficiently different from those of Hamilton and Lucas County in the same time period.

When considering the shape of these connected components we can observe that they all seem to flow along a distinct axis of time as evident in Fig. 7 a), as should be expected for a linearly increasing variable. By identifying the axis of time in each component we may study how demographics (and eventually death count) change along time within each cluster. The other features generally increase along a certain direction within the main trunk, though it is much less distinguishable than time. The other features are relatively constant in the outer components since they contain only one or two counties. Fig. 9 views the main trunk from another angle to observe that directionality in this component can be caused by dif-

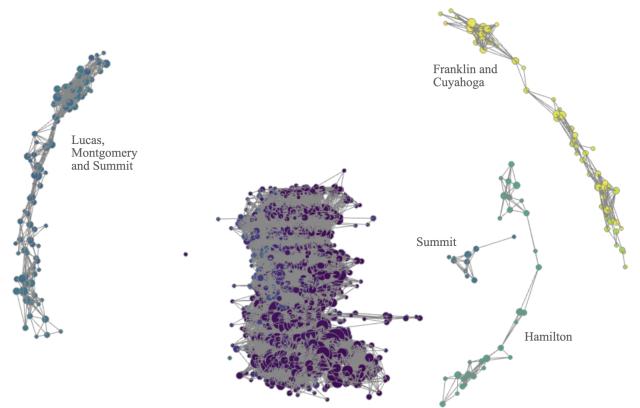


Figure 8: The demographic plot in Fig. 7 b) labeled to identify the counties whose data contributes to each of the smaller connected components

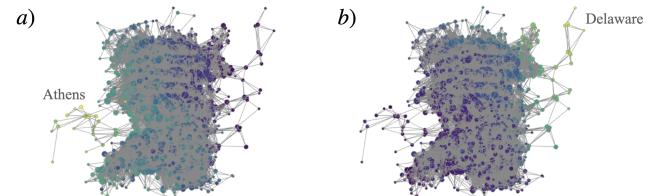


Figure 9: A plot of the visible spurs in the Mapper graph formed from the demographic data of the counties of Ohio. a) is colored by poverty rate and b) by median income.

ferences in the financial attributes of these counties. For example in Fig. 9 a) we can see a gradient formed by the poverty rate in various counties over time and even a spur from Athens county which had a notably high poverty rate in the first 10 years of our time frame. We can also observe a gradient formed by median income in Fig. 9 b) which is visually negatively correlated with the poverty rate and identifies that a spur is created by Delaware county which is the wealthiest county in Ohio by a large margin. This accounts for the only two spurs evident in our demographic Mapper plot.

The only other demographic feature left to discuss is the unemployment rate. Clusters related to higher unemployment rates seem to form visible ridges on one side of the main trunk in our demographic plot, most clearly seen in Fig. 7 d). By analyzing the clusters that contribute to these ridges we can see that most of these clusters are composed of counties near the Ohio river which composes the border of Ohio and Kentucky and Ohio and West Virginia, specifically the counties of Adams, Jackson, Meigs, Morgan, Noble, Pike, Scioto and Vinton, as well as some counties in Northern Ohio between Toledo and Cleveland such as Ottawa and Huron.

### 3.3.2 Analysis of Overdose-Deaths by Demographic Regions

When it comes to understanding the connection between this demographic data and the Ohio overdose epidemic there are two approaches for incorporating the overdose data into Mapper plots. We can either see how the above Mapper graph is colored according to (normalized) cumulative deaths in these counties over the years, or construct new Mapper graphs formed by including (normalized) cumulative deaths into the filter function. We have taken both of these approaches in order to contrast the kinds of insights that are gained by both methodologies. Since the demographic Mapper graph already has strong structure and many variables, adding in death data

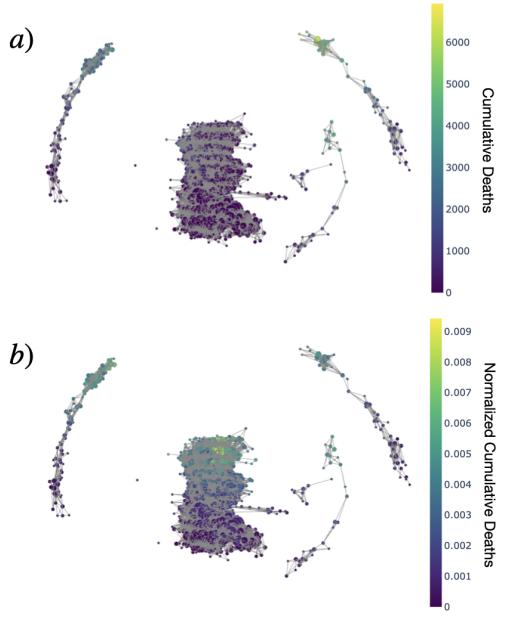


Figure 10: The Mapper plot of demographic data from Fig. 7 colored according to a) cumulative deaths in a county, b) population-normalized cumulative deaths

has less dramatic impact on the graph structure, and so the two approaches will give similar results.

We can see the above demographic landscape colored according to cumulative deaths in Fig. 10 where we can immediately read off that there are higher cumulative deaths in the skinny connected components representing major cities in Ohio and relatively few in the main trunk. We similarly observe that the highest values for the normalized cumulative deaths occur in the main trunk representing less populated regions of Ohio. Some high values also occur in components containing cities notably Lucas and Montgomery County. All of this aligns with our previous analysis of the normalized cumulative deaths.

By analyzing the clusters where the higher normalized cumulative deaths occur we can see that they are predominantly in areas near the border with Kentucky and West Virginia, specifically Henry, Morgan, Monroe, Noble, Scioto and Vinton county, along with areas in Northern Ohio outside Toledo like Highland, Huron and Ottawa county. There is also a small spike from counties along the Pennsylvania border outside the Cleveland area, specifically Jefferson and Mahoning county as indicated in Fig. 11. As remarked above, this aligns with previous research identifying counties with high overdose death rates [44, 29].

What is notable about the normalized cumulative deaths is how it interplays with some of the other demographic information, especially the unemployment rate. The region of Ohio on the right side of the graph experienced high unemployment at the beginning of our chosen time frame (2009) and also higher values of normalized cumulative deaths at the end of our chosen time frame (2023). This suggests a delayed correlation between these two variables. Further work should be done to assess whether high unemployment causes a higher overdose death rate later.

This pattern is consistent with the broader literature on “deaths of despair,” a term popularized by Case and Deaton [5] to describe rising mortality from drug overdoses, alcohol-related liver disease, and suicide, particularly among socioeconomically disadvantaged populations. Economic instability, job loss, and declining prospects—especially in regions with persistent poverty and high

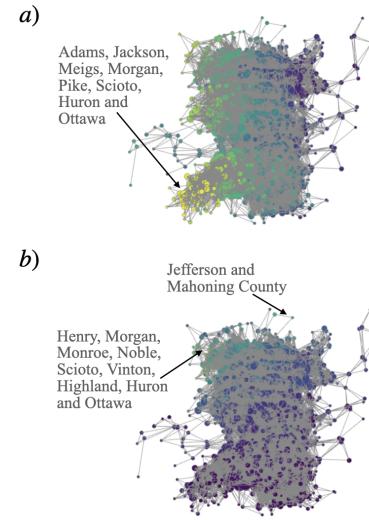


Figure 11: Labeled demographic Mapper plot colored by a) unemployment rate, b) normalized cumulative deaths

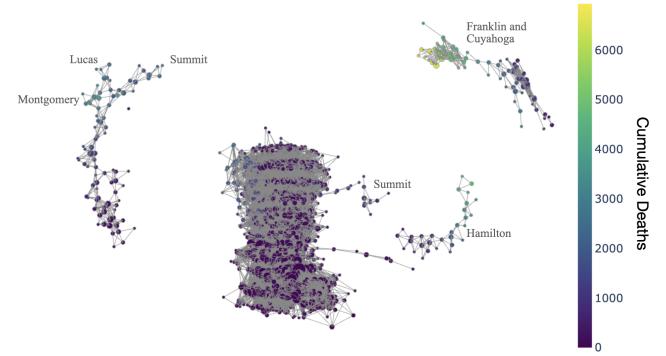


Figure 12: The Mapper graph created from incorporating cumulative death counts into the filter function used for Fig. 7 colored by the cumulative death counts.

unemployment—have been strongly associated with increased risk of fatal overdose. Our findings support the hypothesis that economic distress creates long-term vulnerabilities to substance use and overdose, especially in rural and post-industrial communities. This delayed correlation between early economic decline and later spikes in normalized cumulative deaths underscores the importance of proactive economic and public health interventions.

The other way we have analyzed how demographic data interplays with the demographic landscape is by creating new Mapper graphs where our filter function projects onto (normalized) cumulative deaths in addition to the demographic features listed in our summary of methods. We can then observe the changes in the demographic landscapes created by these alterations to the filter function. The plots of the Mapper graphs formed by incorporating cumulative deaths or normalized cumulative deaths into our projection filter function are presented in Fig. 12 and Fig. 13 respectively.

We note that by observing Fig. 12 very few changes are made to the demographic landscape by incorporating cumulative deaths. One factor that explains this is that the majority of cumulative death spikes came from counties containing major cities in Ohio but our demographic landscape was already partitioned according to population as evidenced by Fig. 8, and so there is not a lot of further delineation that our cumulative death information can create.

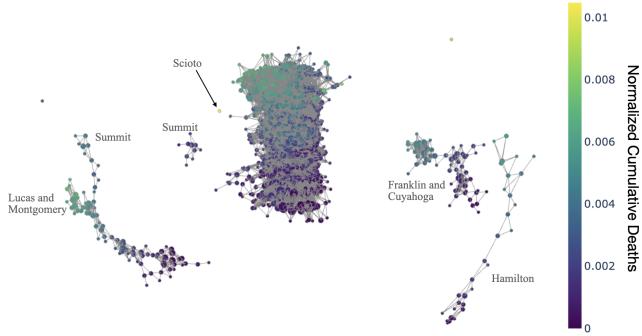


Figure 13: The Mapper graph created from incorporating normalized cumulative death counts into the filter function used for Fig. 7 colored by the normalized cumulative death counts.

That is not to say that there is no extra information provided by this plot, as we can see some greater separation between Lucas, Montgomery and Summit county at the end of our time frame and the emergence of a hole towards the start of our time frame. One side of the hole comes from data from Montgomery County and the other side of the hole comes from data from Lucas County. This hole identifies a period in time where Montgomery County experienced a greater increase in deaths than Lucas County before the cumulative death data in both counties came back to having similar values. The slight separation at the end of our time frame is also caused by the three counties experiencing different levels of cumulative deaths, which is notable as it agrees with our earlier finding where Summit county did not appear as a significant spike in Fig. 4 while both Montgomery and Lucas county did. The similarity of these values is captured in how their two spurs are closer together than the spur for Summit County. Similar insights can be gained from Fig. 13 where we again see an even more significant delineation between Summit County and Lucas and Montgomery County and the appearance of a hole at the same time as the one in Fig. 12. We also see that the major peak of normalized death count in Scioto County is significant enough to break off from the central connected component, which reflects Portsmouth’s infamous presence in the opioid epidemic as the “pill mill of America” [1].

## 4 CONCLUSIONS

This paper explored the efficacy of Mapper plots for studying spatiotemporal data, illustrated by data from the Ohio overdose epidemic. We confirmed the findings of [6] that Mapper does a good job finding spikes, holes and connected components in spatiotemporal epidemiological data. Furthermore, we found that we can use prior information of our space of interest to choose our hyperparameters for the Mapper graph to isolate the effect of the variables of interest on the topology and geometry of the Mapper graphs. We extended the findings of [6] to population-normalized statistics and showed that similar insights can be gained by studying the same kinds of topological and geometric features with this data, yielding new insights as to the impact of the spread on different regions. This allows us to consider spatiotemporal correlations between densely populated areas and less population areas; in our case the resulting Mapper graphs visually suggested a delayed correlation between spread in city-centers and surrounding rural areas. We also introduced a notion of demographic distance between counties, and used Mapper to see how overdose death data interplays with demographic landscapes that shift over time. We observed that the same features of these Mapper graphs could lead to insights into the dynamics of the Ohio overdose epidemic. As far as insights

into the Ohio Overdose epidemic are concerned, we found that in terms of raw numbers the epidemic mostly hurt major urban areas of Ohio which are the most highly populated. The fatalities started to drastically increase in these areas around the middle of the time frame studied (around 2014/15), coinciding with the introduction of fentanyl to the drug supply. We also found that areas near to the Ohio River and outside of cities experienced proportionally higher death rates than all other counties and these spikes occurred towards the end of the time frame we studied— around 2020. We were also able to pick out a very strong spike in normalized deaths occurring in Scioto County which contains the town of Portsmouth, a known hotspot in the epidemic. By studying the demographic profiles we found that the areas with low population that experienced higher rates of poverty and unemployment were those hit hardest near the end of our time frame (after normalizing by population). We also saw what seemed like a delayed correlation between unemployment rates spiking in an area before spikes in population normalized overdose deaths. The main limitations of this methodology are associated with noisy data and small sample statistics which required us to investigate cumulative death counts rather than deaths in a given month due to the low death counts in small counties. It is also difficult to determine a specific notion of statistical significance from these plots as Mapper currently lacks this functionality. Furthermore while the 3D plots allowed for the nodes to be less cluttered, there are still issues identifying details in what can be very visually busy images.

Overall, we find that Mapper visualizations are effective at quickly highlighting which variables are likely to be worthy of further investigation and which are not, simply by the visual correlations and the kinds of information gained by studying the topological and geographic features of these plots. This may be useful for researchers wanting to eliminate certain factors from consideration by showing that they bear little relevance to the dependent variable of interest. We also note they are compelling visual aids for demonstrating broad trends and features within a spatiotemporal dataset and encourage further experimentation with this methodology for studying spatiotemporal data and other time varying datasets.

As regards the overdose epidemic, we found that Mapper is a powerful exploratory tool for identifying spatial, temporal, and demographic patterns in complex public health data. While they cannot replace statistical significance testing, they effectively spotlight relationships worthy of deeper investigation, such as the delayed correlation between economic stressors and overdose mortality. Given their interpretability and flexibility, we recommend broader application of topological data analysis methods to spatiotemporal public health datasets.

## SUPPLEMENTAL MATERIALS

The following GitHub contains our data files, interactive HTML Mapper plot visualizations, and the code used to create them: <https://github.com/willeyna/OhioOverdoseMapper>.

## ACKNOWLEDGMENTS

The authors are grateful to Ismar Volic for sharing the code from [6], Daniel Rosenblum for sharing data from [29] regarding important demographics for the Ohio overdose epidemic, and David Kline for sharing code from [17].

The authors would also like to acknowledge the helpful conversations they had with Gillian Grindstaff and Facundo Memoli to understand more about TDA techniques, and with Andrew Curtis, Jackie Curtis, and Adam Eck regarding spatial models of overdose data. Lastly, the middle author would like to thank his colleagues on the data subcommittee of the US Attorney Heroin/Opiate Task Force for helpful conversations, and for their continuing work to reduce overdose deaths.

## REFERENCES

- [1] C. Arnade. ‘the pill mill of america’: where drugs mean there are no good choices, only less awful ones. *The Guardian*, 2017. 9
- [2] S. V. Ault and J. Lu. Comparison of the spread of novel coronavirus: Topological data analysis of 13 countries. *JIS*, 6(2), 2022. 3
- [3] Y. Bahid, O. Kutsenko, N. Rodríguez, and D. White. The statistical and dynamic modeling of the first part of the 2013-2014 euromaidan protests in ukraine: The revolution of dignity and preceding times. *Plos one*, 19(5):e0301639, 2024. 3
- [4] J. M. Buchanich, L. C. Balmert, K. E. Williams, and D. S. Burke. The effect of incomplete death certificates on estimates of unintentional opioid-related overdose deaths in the united states, 1999-2015. *Public Health Reports*, 133(4):423–431, 2018. 3
- [5] A. Case and A. Deaton. *Deaths of Despair and the Future of Capitalism*. Princeton University Press, 2021. 8
- [6] Y. Chen and I. Volić. Topological data analysis model for the spread of the coronavirus. *Plos one*, 16(8):e0255584, 2021. 2, 3, 4, 5, 6, 9
- [7] J. I. Choi, J. Lee, A. B. Yeh, Q. Lan, and H. Kang. Spatial clustering of heroin-related overdose incidents: a case study in cincinnati, ohio. *BMC public health*, 22(1):1253, 2022. 2
- [8] J. P. Costa, P. Škraba, D. Paolotti, and R. Mexia. A topological data analysis approach to influence-like illness. *KDD Healthday Epidamik*, 2018. 3
- [9] A. Curtis, J. Curtis, J. Ajayakumar, and E. Jefferis. Using spatial mixed methods to reveal the geographic nuances of opioid overdose patterns in small and rural towns. In *New Research in Crime Modeling and Mapping Using Geospatial Technologies*, pp. 211–230. Springer, 2025. 2
- [10] P. Dłotko, W. Qiu, and S. Rudkin. Financial ratios and stock returns reappraised through a topological data analysis lens. *The European Journal of Finance*, 30(1):53–77, 2024. 3
- [11] A. Eck, A. Muradi, and M. Simoya. Utilizing supervised machine learning models for opioid hotspot prediction. *preprint*, 2024. 2
- [12] N. El-Bassel, R. D. Jackson, J. Samet, and S. L. Walsh. Introduction to the special issue on the healing communities study. *Drug and Alcohol Dependence*, 217:108327, 2020. 6
- [13] S. Hepler, E. McKnight, A. Bonny, and D. Kline. A latent spatial factor approach for synthesizing opioid-associated deaths and treatment admissions in ohio counties. *Epidemiology*, 30(3):365–370, 2019. 2
- [14] A. Hickok, D. Needell, and M. A. Porter. Analysis of spatial and spatiotemporal anomalies using persistent homology: Case studies with covid-19 data. *SIAM Journal on Mathematics of Data Science*, 4(3):1116–1144, 2022. 3
- [15] Y. Ji. *A Joint Spatio-Temporal Model of Opioid Associated Deaths and Treatment Admissions in Ohio*. Wake Forest University, 2019. 2
- [16] D. Kim, S. Kim, Y. Do Kim, and S. Lyu. Enhanced detection of harmful algal blooms using topological data analysis for clustering spatially distributed water quality and hydrodynamic data. *KSCE Journal of Civil Engineering*, p. 100177, 2025. 3
- [17] D. Kline and S. A. Hepler. Estimating the burden of the opioid epidemic for adults and adolescents in ohio counties. *Biometrics*, 77(2):765–775, 2021. 2, 3, 9
- [18] D. Kline, Y. Ji, and S. Hepler. A multivariate spatio-temporal model of the opioid epidemic in ohio: a factor model approach. *Health Services and Outcomes Research Methodology*, 21(1):42–53, 2021. 2
- [19] Z. R. Li, E. Xie, F. W. Crawford, J. L. Warren, K. McConnell, J. T. Copple, T. Johnson, and G. S. Gonsalves. Suspected heroin-related overdoses incidents in cincinnati, ohio: A spatiotemporal analysis. *PLoS medicine*, 16(11):e1002956, 2019. 2
- [20] D. Lo and B. Park. Modeling the spread of the zika virus using topological data analysis. *PloS one*, 13(2):e0192120, 2018. 3
- [21] Z. Lu and H. Liu. A topological data analysis approach to the covid-19. In *2022 IEEE 10th joint international information technology and artificial intelligence conference (ITAIC)*, vol. 10, pp. 469–473. IEEE, 2022. 3
- [22] L. Ma, L. Tran, and D. White. A statistical analysis of drug seizures and opioid overdose deaths in ohio from 2014 to 2018. *Journal of Student Research*, 10(1), 2021. 2
- [23] V. N. Madukpe, N. F. S. Zulkepli, M. S. M. Noorani, and R. Gobithaasan. Comparative analysis of ball mapper and conventional mapper in investigating air pollutants’ behavior. *Environmental Monitoring and Assessment*, 197(2):136, 2025. 3
- [24] B. D. Marshall, N. Alexander-Scott, J. L. Yedinak, B. D. Hallowell, W. C. Goedel, B. Allen, R. C. Schell, Y. Li, M. S. Krieger, C. Pratty, et al. Preventing overdose using information and data from the environment (provident): protocol for a randomized, population-based, community intervention trial. *Addiction*, 117(4):1152–1162, 2022. 2
- [25] R. McMaster, L. Masarweh-Zawahri, K. C. Flynn, V. S. Deo, and D. J. Flannery. Drug overdose death among residents of urban census tracts: How granular geographical analyses uncover socioenvironmental correlates in cuyahoga county, ohio. *Journal of Urban Health*, 102(2):445–458, 2025. 2
- [26] Ohio Department of Health. Mortality data portal. <https://data.ohio.gov/wps/portal/gov/data/view/mortality>, 2024. 3
- [27] Ohio Department of Job and Family Services. Local area unemployment statistics (laus). <https://ohiolmi.com/Home/LAUS>, 2024. Accessed: 2025-05-14. 4
- [28] N. Rodríguez and D. White. An analysis of protesting activity and trauma through mathematical and statistical models. *Crime Science*, 12(1):17, 2023. 3
- [29] D. Rosenblum, J. Unick, and D. Ciccarone. The rapidly changing us illicit drug market and the potential for an improved early warning system: evidence from ohio drug crime labs. *Drug and alcohol dependence*, 208:107779, 2020. 2, 3, 6, 8, 9
- [30] S. Rudkin, D. J. Webber, and P. Dłotko. Spatial disparities in infection rates at the dawn of a pandemic: Wealthy young workers mattered. *Available at SSRN 4356837*, 2023. 3
- [31] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu. Dbscan revisited, revisited: why and how you should (still) use dbscan. *ACM Transactions on Database Systems (TODS)*, 42(3):1–21, 2017. 4
- [32] S. Shrestha and T. J. Stopka. Spatial epidemiology and public health. In *Geospatial Technology for Human Well-Being and Health*, pp. 49–77. Springer, 2022.
- [33] G. Singh, F. Mémoli, G. E. Carlsson, et al. Topological methods for the analysis of high dimensional data sets and 3d object recognition. *PBG@ Eurographics*, 2:091–100, 2007. 1, 3
- [34] M. Soliman, V. Lyubchich, and Y. R. Gel. Ensemble forecasting of the zika space-time spread with topological data analysis. *Environmetrics*, 31(7):e2629, 2020. 3
- [35] K. Stewart, Y. Cao, M. H. Hsu, E. Artigiani, and E. Wish. Geospatial analysis of drug poisoning deaths involving heroin in the usa, 2000–2014. *Journal of Urban Health*, 94:572–586, 2017. 2
- [36] E. Tassi, H. Fisher, A. Bolender, J.-H. Lee, L. J. Ayoub, A. M. Bianchi, B. Kuo, V. Napadow, E. Maggioni, and R. Scocco. Topological data analysis of resting-state fmri suggests altered brain network topology in functional dyspepsia: A mapper-based parcellation approach. In *International Workshop on Topology-and Graph-Informed Imaging Informatics*, pp. 88–99. Springer, 2024. 3
- [37] G. Tauzin, U. Lupo, L. Tunstall, J. B. Pérez, M. Caorsi, A. Medina-Mardones, A. Dassatti, and K. Hess. giotto-tda: A topological data analysis toolkit for machine learning and data exploration, 2020. 4
- [38] D. Taylor, F. Klimm, H. A. Harrington, M. Kramár, K. Mischaikow, M. A. Porter, and P. J. Mucha. Topological data analysis of contagion maps for examining spreading processes on networks. *Nature communications*, 6(1):7723, 2015. 3
- [39] A. J. Tyb. An overview of spatial econometrics. *arXiv preprint arXiv:1605.03486*, 2016. 2
- [40] U.S. Census Bureau. County population totals: 2010–2019. population and housing unit estimates. <https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html>, 2020. 3
- [41] U.S. Census Bureau. American community survey (acs). <https://www.census.gov/programs-surveys/acs>, 2024. 4
- [42] U.S. Census Bureau. Small area income and poverty estimates (saipe) program. <https://www.census.gov/programs-surveys/saipe.html>, 2024. Accessed: 2025-05-14. 4
- [43] S. Vannoni, E. Tassi, I. W. Sampaio, and E. Maggioni. Exploitation of mapper algorithm in neuroimaging applications: A novel framework for outcomes prediction. In *International Workshop on Topology-and*

- Graph-Informed Imaging Informatics*, pp. 76–87. Springer, 2024. 3
- [44] D. White, L. Ma, and L. Tran. Unintentional opioid overdose deaths in ohio: Insights from sudors data. *preprint*, 2025. 6, 8