A Gaussian-Mixture Model Analysis of Polysubstance Drug Use in Opioid Overdose Deaths

by

Kayleigh Marie Adamson

BS, The Pennsylvania State University, 2015

Submitted to the Graduate Faculty of

the Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Kayleigh Marie Adamson

It was defended on

April 20, 2020

and approved by

Jeanine M. Buchanich, PhD, Research Associate Professor, Biostatistics
Graduate School of Public Health, University of Pittsburgh

Ada Youk, PhD, Research Associate Professor, Biostatistics
Graduate School of Public Health, University of Pittsburgh

Jenna C. Carlson, PhD, Assistant Professor, Biostatistics
Graduate School of Public Health, University of Pittsburgh

Eric Hulsey, PhD, Adjunct Professor, Behavioral and Community Health Sciences
Graduate School of Public Health, University of Pittsburgh

Thesis Advisor: Jeanine M. Buchanich, PhD, Research Associate Professor, Biostatistics
Graduate School of Public Health, University of Pittsburgh

Jeanine M. Buchanich, PhD

A Gaussian-Mixture Model Analysis of Polysubstance Drug Use in Opioid Overdose Deaths

Kayleigh Marie Adamson, MS

University of Pittsburgh, 2020

Abstract

Background: In the midst of the opioid crisis, it is imperative to identify risk factors for various subgroups of polysubstance drug users to reduce the risk of mortality.

Methods: A Gaussian-Mixture Model analysis utilizing age group, White or African American race, sex, and dichotomized presence of illicit opioids (e.g., fentanyl), stimulants (e.g., cocaine), and benzodiazepines (e.g., Xanax) was conducted to develop advanced characterizations of subgroups and polysubstance use. 3,318 accidental overdose deaths (ICD10 X40-X44) from the Allegheny County Office of the Medical Examiner from years 2008-2019 were included in the analysis.

Results: Nine demographic and substance use subgroups were identified. Of those, three may have particular implications for tailoring interventions for polysubstance use: (1) White females, ages 35-44, with presence of benzodiazepines and opioids, (2) older African American males, ages 55-64, with presence of illicit opioids and stimulants, and (3) White males, ages 35-44, who are utilizing heroin and/or prescription opioids.

Conclusion: The heterogeneity of the polysubstance use in Allegheny County makes it necessary to develop further advanced characteristics of the subgroups being impacted by this epidemic. Statistical learning and GMM provided an optimal tool to generate such inferences.

Public health significance: This analysis, identifying clusters of subgroups in accidental opioid overdose deaths, can be utilized to inform public health professionals and policy experts as to how to further tailor and improve interventions for the opioid epidemic in Allegheny County.

Keywords: opioids, polysubstance drug use, demographic, Gaussian-Mixture Model

Table of Contents

List of Tables

List of Figures

Preface

To my thesis advisors, thank you for your guidance through this adventurous process.

To Dr. Ada Youk and Dr. Jenna Carlson, thank you for believing I would succeed even though at times I didn't believe it myself.

To Dr. Buchanich, thank you for providing me with such sound advice and instilling confidence in me. And for graciously offering me the opportunity to continue on this learning journey with you- I can't wait for what we discover.

To my brilliant cohort, there are no other people with whom I would have wanted to venture through this program with. I'm humbled to have learned with, and by, all of you. I look forward to seeing all that you accomplish.

To Dr. Vanessa Troiani and Dr. Antoinette DiCriscio, thank you for being examples of what strong, determined, female scientists accomplish. I am forever indebted to you.

To my friends and family, thank you for being a wonderful source of light and fresh air when I was struggling through graduate school. Thank you for making me laugh and reminding me that life is also for living, not just theory homework, and sometimes we need that vodka soda.

To my father, Commander Daryl Adamson, thank you for providing me with the opportunity to find my passions and for encouraging me to follow them.

To my little brother, Cameron, thank you for being a wild and unpredictable child. My thesis behaved in the same manner and, therefore, I was prepared.

And finally, to my mother, Ms. Lori Adamson. Everything I am is because of you. I could fill endless pages thanking you for all that you have done for me but, as I am earth conscious and would like to save paper, I will simply say, 'I love you.'

1.0 Introduction

The first two decades of the 21st century have brought forth the worst drug epidemic ever recorded. Since 1999, an estimated 700,000 people have died from a drug-related overdose death (*Opioid Overdose | Drug Overdose | CDC Injury Center*, 2019) with the mortality rate increasing subsequently each year through 2017. In 2017, 70,237 people died as a result of a drug overdose with an estimated 67.8% of those being attributed to opioid involvement (Scholl et al., 2018), a rate that was five times larger than it was in 1999 (*CDC WONDER*, 2020). Chen et al., (2019) projected that between 2015 and 2025, opioid overdose deaths would increase by 147% (33,100 in 2015 to 81,700 in 2025). As a result, the United States government declared the opioid epidemic a public health emergency in October 2017 (Haffajee & Frank, 2018) - a declaration 30 years in the making.

1.1 History of the Opioid Epidemic

The first wave of the opioid epidemic was the product of a two-fold initiative: (1) introducing pain as the fifth vital sign and (2) influential drug companies promoting use of opioids, specifically OxyContin, as a solution to chronic pain (Van Zee, 2009) stating that risk of addiction was extremely low which was found to be contrary. The drastic increase in opioid prescriptions for treatment of chronic pain was accompanied by a drastic increase in opioid overdose deaths (Wilkerson et al., 2016). In 2011 officials began to publicly acknowledge that overdose deaths would continue to rise if intervention was not taken. The direct action was to

implement restrictions on the dispensation of opioids. Physicians were instructed to decrease opioid prescriptions for all patients thus making them far more difficult to obtain. This restriction introduced the second wave of the epidemic, heroin. As it became increasingly more difficult to legally obtain prescription opioids, people who were dependent on them began turning to heroin- an illegal, synthetic form of opioid that is cheaper, stronger, and more easily attainable. As a result, between 2007 and 2013, the rate of heroin use increased 150% (Jones et al., 2015).

The third wave of epidemic began when illicit fentanyl started being cut into heroin and other drugs, including cocaine and stimulants. Fentanyl is a fast acting opioid that is 50-100 times more potent than morphine (LaRue et al., 2019), therefore the introduction of this resulted in yet another large surge of overdose deaths. Prior to 2015, fentanyl was estimated to be present in 5% of overdose deaths and by 2016 it was present in 15% (Nolan et al., 2019). Consequently, the mortality rate continued to rise even though prescription opioid prescribing practices declined 20% between 2015 and 2017 (Guy et al., 2017). These deaths indicated that this was no longer a prescription opioid problem, as it was during the first wave of the epidemic, but rather it had matured into a multi-faceted problem that included prescription and illicit opioids, and various other drugs.

1.2 Opioids and Polysubstance Use

In recent years, researchers have begun investigating the opioid epidemic from a public health perspective. They are attempting to further define who specifically is impacted and in what way, as opposed to simply viewing the epidemic as a drug problem. This approach will allow researchers to further target interventions and subpopulations, as well as capture the spread

of the epidemic. One particularly interesting component to assess is polysubstance use, or utilization of one or more drugs in conjunction with an opioid. This area of research is rich in information pertaining to subpopulations and can be very helpful in determining which populations are using which illicit drugs (Hassan & Le Foll, 2019). While these data provide much insight, until recently minimal research has been conducted on the subject matter. Shiels et al., (2018) found that the mixture of cocaine and illicit opioids was the most common pattern of overdose deaths for non-Hispanic black men and women. Other researchers found that people "older than 24, who are from non-rural residents, with comorbid mental illness, non-Hispanic black residents, and people who are recently homeless" are at a higher risk than others of dying of an overdose of cocaine and opioids, than just opioids alone (Barocas et al., 2019). Balmert et al., (2016) found that African American adults were more likely to die from cocaine. Allen et al., (2019) reported that there are defined drug trajectories for the African American and Latino community as compared to the white community. Opioid overdose deaths are higher for middle-aged and older African American's and Latinos. These subpopulations are more likely to have 'long-running drug trajectories and are more likely to utilize heroin', as compared to whites who were more likely to have opioid overdose deaths as a younger or middle-aged person. Additionally, the white persons are 'more likely to have started utilizing opioids after 2000 and have transitioned into heroin use from prescription opioids' (Allen et al., 2019).

It is estimated that among those that use opioids, polysubstance users make up about 57.3% (Hassan & Le Foll, 2019). Between 2015 and 2016, drug overdose deaths involving cocaine, or other psychostimulants increased 42.4%, with a large majority attributed to the co-presence of cocaine and an opioid (Kariisa et al., 2019). Armenian et al., (2019) found that of all cocaine related deaths, 3 out of 4 included an opioid, typically fentanyl. Other researchers found

similar results for the presence of cocaine in heroin overdose deaths with values ranging between 53% and 63% (B. Han et al., 2015; Jones et al., 2015). It is presumed that cocaine-related deaths are on the rise due to the admixture of fentanyl and fentanyl analogs. This combination is creating a new, opioid-naïve population (LaRue et al., 2019). While cocaine has been the most reported on, other drugs are also seeing a rise in combination with opioids including methamphetamines (LaRue et al., 2019), MDMA (Kariisa et al., 2019), and alcohol. Of the individuals who reported having an opioid use disorder, 57.3% also had a polydrug use disorder (Hassan & Le Foll, 2019).

1.3 Statistical Learning

In the rise of big data, researchers have begun borrowing methods from machine and statistical learning to ask questions of how to combat and intervene with the morose predictions of the overdose epidemic. While there are existing methods for surveillance of drug overdose deaths, more advanced methods are needed to act faster and have a larger impact (Creppage et al., 2018). The use of systematic models has the capability of predicting future incidences, forecasting their impact, and identifying clusters of who is impacted (Burke, 2016). The application of machine and statistical learning to the opioid epidemic has been broad, including using machine learning to phenotype opioid overdose events (Badger et al., 2019), utilizing unsupervised leaning to identify at-risk groups of opioid addiction (Basu et al., 2018), analysis of clinical notes with natural language processing to develop a likelihood score for developing substance dependence (Ellis et al., 2019), and using a method known as fast subset selection to detect subpopulations in overdose death data (Neill & Herlands, 2017). While application of

these methods has been minimal prior to opioid epidemic, researchers are now realizing the utility of their application to the mass amounts of healthcare and public health data available, including health records, death certificates, and policy data.

1.4 Objectives

To address the ongoing public health crisis surrounding the opioid epidemic, it is necessary to adapt, and update, strategies as the data become more mature. While overdose deaths from prescription opioids decreased in 2018, deaths from illegally manufactured fentanyl (IMF) continued to starkly increase (Gladden et al., 2019). This signifies two key points: (1) the epidemic is continuing to progress even with current interventions and (2) targeting mass efforts at prescription opioids may not be the best strategy to reduce overdose and death and a multi-faceted approach is necessary to properly intervene with the epidemic.

Two major components of opioid use that tend to be under-investigated are polysubstance users and demographic differences between subpopulations. The intersection of these two research areas creates an interesting question: which secondary drugs (stimulants, antidepressants, benzodiazepines, etc.) are most common in opioid overdose deaths of various income, age, race, and gender categories? Investigation of this intersection can create a new realm of research that can further aid in reducing overdose and death in various subgroups.

One particularly interesting location to study this multi-faceted, heterogenous epidemic is in Pennsylvania, one of the states most impacted by the ongoing opioid crisis. As of 2018, overdose was the leading cause of death in the state for people 25-44 years (Burke & Buchanich, 2018). Balmert et al., (2016) found that from 1979 to 2014 the rate ratio for accidental drug

poisoning increased 14-fold within the state, particularly around Philadelphia, Scranton, and Pittsburgh. Thus far, minimal research has been conducted investigating the patterns of polysubstance use involved in opioid overdose deaths in Allegheny County while utilizing statistical learning methods. Such research could aid policy makers and harm-reduction experts as to how to better predict and prepare for the epidemic, and how to better approach interventions in Pittsburgh and surrounding areas (Neill & Herlands, 2017).

The first aim of this thesis will be to describe the progression of polysubstance use and opioid epidemic in Allegheny County from 2008-2019. Next, demographic patterns of secondary nonopioid drugs present in opioid overdose deaths in Allegheny County will be assessed to determine if patterns emerge over time. Finally, a cluster analysis, a facet of statistical learning, will be utilized to develop more advanced characterizations to identify subpopulations to further target interventions.

2.0 Methods

2.1 Data

2.1.1 Data source

Data were retrieved from the Western Pennsylvania Regional Data Center (WPRDC), a "shared technological and legal infrastructure to support research, analysis, decision making, and community engagement" (Allegheny County, Office of the Medical Examiner, 2020). As part of their initiative, the WPRDC provides data that are open-access and freely available to encourage collaboration. Data utilized were from the Allegheny County Office of the Medical Examiner (ACOME) and included overdose deaths in Allegheny County from 2008-2019 with the following ICD-10 codes: X40-44, X60-64, X85, and Y10-14. This analysis was restricted to accidental overdose deaths. As of 1/12/2020, there were 4,456 accidental overdose deaths included in the observational dataset from 01/01/2008 to 12/31/2019.

2.1.2 Covariates

Variables available in the dataset and that were utilized for analysis were age, race, sex, manner of death, date and time of death, zip code of death, and which specific drugs were found in the system at time of death.

2.1.3 Data cleaning and management

Race and Sex

Race and sex were treated as categorical variables. Sex had two levels: male or female. Race had nine levels: Asian, Black or African American, Hispanic, Indian (Native American), Middle Eastern, Other, Unidentified, White. Subjects who were identified as other (n=5) or unidentified (n=1) were excluded as the purpose of analysis was to identify subpopulations based on race and other demographic information. Additionally, subjects who were identified as Indian (n=1), Hispanic (n=20), Middle Eastern (n=4), or Asian (n=9) were excluded as inference from these races would not be informative due to small sample sizes. This was deemed appropriate as the majority of Pittsburgh residents identify as White (64.9%) or African American (22.8%) (U.S. Census Bureau, 2020). After removing these subjects, and those that had no values (n=4), the dataset included those who identified as African American and White for a total of 4,411 subjects.

Zip Code of Incident

If a person had Allegheny County listed as their permanent residence but died outside of the county, they were included in the dataset based on the reporting regulations of the ACOME. For the purposes of this analysis, incident of death was restricted to those deaths occurring within Allegheny County. After removal of subjects who died outside of the county (n=344), and those without incident of death blank (n=76), 3,983 subjects remained.

Identification of Opioid Involvement

Nine columns (*combined_od1, combined_od2, combined_od3, combined_od4, combined_od5, combined_od6, combined_od7, combined_od8, combined_od9)* listed each drug found in each subject's toxicology screen upon death. All drugs were compiled into an overall

drug list. It was not indicated whether the utilization of only nine columns was arbitrary, or if that was the largest number of drugs found at time of death. A new variable, *opioidcount*, was created to count the total number of opioids present at the time of death for each subject. Opioids present at time of death can be seen below.

Table 1: Opioids present at time of death

3MFENT	FENTAC	MEPER	PFLFNT
4FENT	FENTAN	MORPHI	PROPOX
CARF	FUF	OPIATE	REMI
CFENT	HEROIN	OXYCOD	TFENT
CODEI	HYDROM	OXYM	TRAMAD
CPRFNT	HYDROO	PFBF	U477
DIHY	HYDRO	PFIBF	U48800

If a subject did not have any of the listed opioids present at the time of death (*opioidcount=0)* they were not included in the dataset (n=667). This resulted in 3,318 subjects remaining.

To further quantify the extent of opioid involvement, a subcategory of *opioidcount* was created, *illicitopioid,* to track synthetic opioid involvement in overall opioid overdose deaths as seen in Table 2 below.

Table 2: Illicit opioids present at time of death

CARF	FUF	PFLFNT	U48800
CFENT	HEROIN	TFENT	
FENTAC	PFBF	TRAMOD	
FENTAN	PFIBF	U477	

Identification of Secondary Drug Involvement

To assess polydrug use involvement in opioid overdose deaths, two new indicator variables were created: (1) *stimcount,* which was =1 if stimulants were present at time of death

and (2) *benzocount,* which was =1 if benzodiazepines were present at time of death. Stimulants (Table 3) and benzodiazepines (Table 4) can be observed below.

Table 3: Stimulants present at time of death

AMPH	COCAIN	METHA	PHENT
BATH	DEXTR	METHAN	
CO	EPH	METHPH	

Table 4: Benzodiazepines present at time of death

7AMINO	BENZOD	ETIZ	TEMAZ
ALPRAZ	CLONA	GABA	
BENZ	CLOZAP	LORA	
BENZO	DIAZEP	MIDAZO	

2.2 Statistical Analysis

Previous research investigating opioid overdose deaths has been conducted; however, the majority of studies have utilized frequentist methodology to determine if the number of deaths increasing over time was statistically significant. Few studies have been conducted investigating this topic from a statistical learning perspective.

To provide justification for advanced learning methods, two preliminary data exploration analyses will be conducted: (1) a review of age-adjusted mortality rates (AAMR) in Allegheny County and (2) mapping incident of drug use in Allegheny County over time. A comparison between Allegheny County, Pennsylvania, and United States age-adjusted mortality rates (AAMR) for opioid-related deaths will be conducted. Data for comparison will be retrieved from

the Kaiser Family Foundation (Pennsylvania and United States) and Allegheny County Department of Human of Services (ACDHS) (Hulsey et al., 2016; Kaiser Family Foundation, 2020). The ACDHS has AAMR from 2008-2016, therefore rates for 2017 and 2018 were calculated using standard age-adjustment formulas (Curtin & Klein, 1995).

Initial preliminary analysis will be concluded with an investigation into the breath of stimulant (cocaine; amphetamines) and benzodiazepine (Xanax) use in Allegheny County over time, quantified by using data from 2008 and 2019. Previous research indicated stimulants (Barocas et al., 2019; Gladden et al., 2019; Kariisa et al., 2019) and benzodiazepines (Buchanich et al., 2018) as the most common categories of drugs to be combined with opioids and therefore, for the sake of brevity, were selected for this analysis. To properly capture the difference over time, proportion of deaths involving a secondary drug will be mapped. Visualizations will be created by (1) grouping overdose deaths by zip code, (2) counting the number of opioid overdose deaths that had a secondary drug present (stimulant or benzodiazepine) present at time of death (identified as n), and then (3) calculating a crude mortality for each zip code, or n/zip code population * 1000, to compare across zip codes. Population counts for each zip code will be retrieved from openly available 2010 U.S. Census data (U.S. Census Bureau, 2020).

2.2.1 Comparison of statistical methods

Latent class analysis

Liu et al., (2019) utilized latent class analysis (LCA) to identify subpopulations from opioid-related discharges from hospitals. Defined in 1950 by Paul Lazarsfeld (Lazarsfeld, 1950), LCA is a finite mixture model (FMM), or a model that expresses the "overall distribution of one or more variables as a mixture of a finite number of component distributions" (Masyn, 2013).

LCA borrows from the theory of FMM by defining subgroups, or classes, among data composed of a latent, categorical variable. With this method, it is presumed that the amount of covariation observed in each subject's response is due to the relationship with the underlying latent variable. The underlying statistical theory is as follows:

Let y_i represent element i of a response pattern **y** and let the indicator function $I(y_i = r_i) = 1$ when $i = r_i$, and zero otherwise. Then probability of observing a particular vector of responses is:

$$P(Y = y) = \sum_{c=1}^{C} \gamma_c \prod_{i=1}^{I} \prod_{r_{i=1}}^{Ri} \rho_{i,r_i|c}^{I(y_i=r_i)} \quad (2.1)$$

where γ_c is the probability of membership in latent class c and $\rho_{i,r_i|c}^{I(y_i=r_i)}$ is the probability of response r_i to item i, conditional on membership in latent class c. Model selection is based on the likelihood ratio test and information criteria including AIC and BIC (Lanza & Rhoades, 2013).

LCA handles missing data, has the ability to accommodate weighting of variables, and can include more complex mathematical algorithms which other algorithms cannot do with ease. However, while LCA has many strengths, there are also limitations, including slow computing speed with high volume data and the necessity to recode continuous variables to categorical to fit the model which can result in errors (Lanza & Rhoades, 2013).

Subset scanning

Another method commonly utilized to identify clusters within a larger population is a method called subset scanning. Neill (2012) utilized a method known as fast subset scanning to detect emerging outbreaks of disease. Scan statistics are often utilized to identify certain events over a period of time, which makes them very advantageous to aid in surveillance of public

health issues. The original scan method introduced by Kulldorff (1997) intends to maximize a score function, *F(S)*, over some large, spatial region, *(S)*, which has various sub-regions, s_i. by constraining the spatial region and conducting a thorough search over the entire area. The theoretical model for scan statistics is as follows:

Let *S r*epresent the spatial region of interest and let *D* represent the given data. Define the hypotheses as:

- H_0*:* There are no identified clusters in the data, i.e. all counts are from the expected distribution.
- H_1*:* There is at least one cluster identified cluster, i.e. the counts are not expected.

Then, maximizing the log likelihood ratio is:

$$F(S) = \log\left[\frac{P\{D|H_1(S)\}}{P\{D|H_0)\}}\right] \tag{2.2}$$

In 2017, Neill introduced a new method of subset scanning known as Multidimensional Tensor Scanning (MDTS) to specifically scan for drug overdose surveillance with high dimensional data (Neill, 2017). This method expands upon previous subset scan methods as it can identify the specific subset of values for each observed variable. By utilizing this method, there is higher power to detect a difference in emerging trends. As MDTS is a relatively new method, there is very minimal research and literature currently on the methodology and theory therefore that method was not chosen for this thesis.

2.2.2 Gaussian mixture-model

A commonly used methodology to identify clusters in subpopulations is a k-means clustering model (Macqueen, 1967). With this method, data are portioned into k groups, or

clusters. The clusters are defined such that the Euclidian distance between points within each cluster is minimized. The algorithm ends when classification of clusters no longer changes. While this method is ideal as it is simplistic and easy to comprehend, there are various limitations that make it a suboptimal method to use. K-means is non-probabilistic and utilizes a distance-from-cluster model to define the clusters, both of which lead to poor performance in real-life data. The non-probabilistic nature of k-means stems from how clusters are grouped. After identifying the cluster, the model places a circle at the center of each cluster, and those data points outside of the cluster are not part of the cluster at all. An intuitive way to define this is by calling k-means a hard classifier, similar to an indicator variable, which means that the model does not weigh the probability of the datapoint being in any other cluster. Additionally, k-means lacks the ability to account for variance between clusters. As previously mentioned, k-means defines the clusters in a circular pattern and thus if the cluster is oblong or elliptical in nature, k-means cannot model it properly which will result in more classification (Wagstaff et al., 2001).

Where the k-means clustering algorithm lacks, the Gaussian Mixture Model (GMM) succeeds. Conceptually, the GMM aims to find a mixture of multi-dimensional Gaussian distributions to best model the given data. Unlike k-means, GMM is probabilistic which means that instead of providing one cluster assignment, the algorithm will provide the probabilities of the datapoint belonging to other clusters. Similar to k-means, GMM utilizes the expectation-maximization approach however the result of each cluster is a smooth Gaussian model. With this algorithm, covariance is accounted for, and thus the resulting clusters are not restricted to being spheres but can indeed be spherical and oblong in nature. This aids in more accurate classification (VanderPlas, 2016). Additionally, with GMM the optimal number of clusters, *k*, can be identified by the value that minimizes AIC and/or BIC, which is not the case for k-means.

Theoretically, GMM is a weighted sum of *K* Gaussian densities, for K clusters, as given by the following equation where X is the data, π_k is the mixing coefficient, or weights, for the k-th distribution, and *G( )* are the Gaussian component densities:

$$p(X) = \sum_{k=1}^{K} \pi_k G(X|\mu_k, \Sigma_k) \tag{2.3}$$

Ideally this equation would maximize the log-likelihood for the parameters of interest, but as there are *k* groups it is not possible to estimate the equation in closed form which is when the expectation-maximization (EM) algorithm is utilized. The EM algorithm provides an alternative way of maximizing the likelihood. The first step, estimation, involves initializing μ_k, Σ_k, and π_k then for the given values, estimating the value of the latent variables. From there, maximization involves updating the values of the parameters using the maximum likelihood method (Reynolds, 2016).

Utilizing this methodology, this thesis will attempt to further define subpopulations of polydrug use involved in opioid overdose deaths. The package 'mclust' in R (Scrucca et al., 2016) will be utilized for model-based clustering and classification using E-M estimation. To reduce the number of dimensions, making the model more parsimonious, the variables *incident_zip*, *case_year*, and *opioidcount* were not added to the model therefore the final variables for the clustering analysis were *age, sex, race, illicitopio*id (dichotomized), *stim, and benzo*. As *case_year* was not included in the model, and it is known that heroin has known multi-collinearity with other illicit opioids without inclusion of year, heroin was not included in the illicit opioid category.

To date, no analysis has been conducted investigating mortality among opioid subpopulations utilizing GMM. This analysis will expand upon the limitations of previous research to develop more advanced characterizations of subpopulations for further intervention. Analyses will be performed using R, version 3.6.2 (R Core Team, 2017) in RStudio, version 1.2.500 (RStudio Team, 2015).

3.0 Results

3.1 Descriptive Statistics

The distribution of sex and race subgroups can be seen in Table 5. Of the opioid overdose deaths, 69.5% were males, and 88% were white.

Table 5: Demographics

	Female *(1019, 30.5%)*	**Male** *(2301, 69.5%)*
White *(2919, 88%)*	897 (27%)	2022 (61%)
Black *(399, 12%)*	120 (3.5%)	279 (8.4%)

The number of opioids present at time of death ranged from 1-7 with majority of subjects having one (58%) or two opioids (31%). 33% of subjects had at least one stimulant present at time of death, 26% had a least one benzodiazepine, and 79% had at least one illicit opioid. (Table 6.

Table 6: Specific drug presence at time of death

Illicit opioid	2619 (79%)
Stimulant	1083 (33%)
Benozdiazepine	874 (26%)

Comparison of AAMRs revealed that Allegheny County had overall higher rates compared to both Pennsylvania and the U.S. over the time period of interest (Figure 1). The

largest change was from 2017 to 2018 where the rate of overdose deaths peaked and then starkly declined.

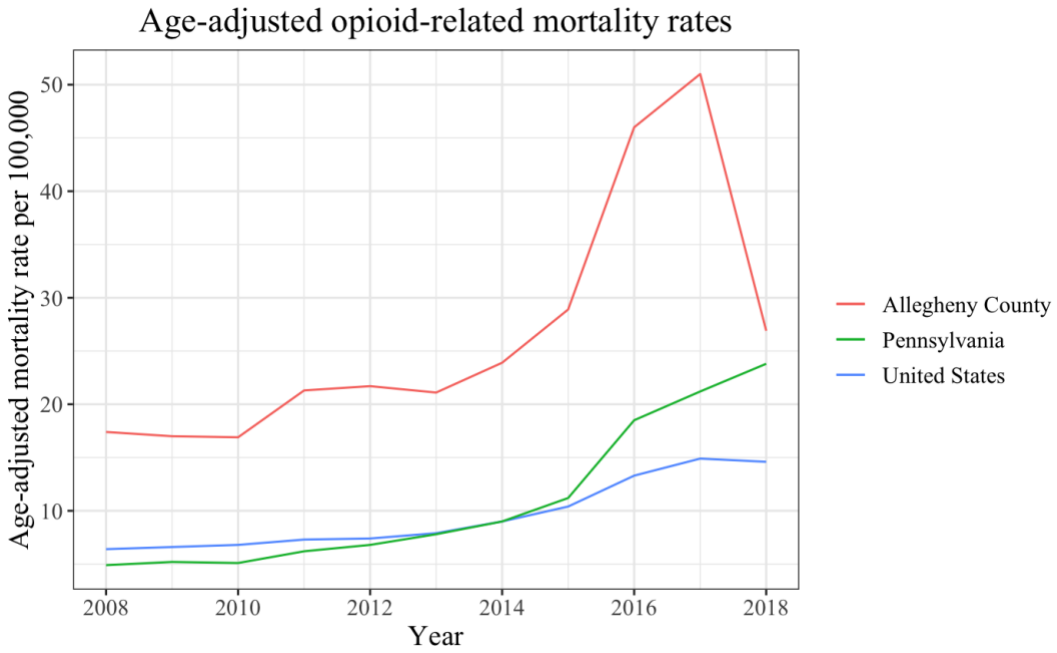

Figure 1: Age-adjusted opioid overdose mortality rates comparison

Next, polydrug use was assessed over time to determine the extent of concurrent opioid and benzodiazepine and stimulant use in Allegheny County (Figure 2). Over the past 10 years, benzodiazepine use (blue) in conjunction with opioid overdose deaths has seen a decline, with a peak in 2009. Conversely, stimulant use (red) has drastically increased particularly after 2014, where the rate of deaths involving stimulants has grown exponentially. Additionally, benzodiazepine and stimulant involvement appear to be inversely related, again, after 2014. In 2008, stimulants were present in ~35% of opioid overdose deaths and benzodiazepines were present in ~26%, by 2019 stimulants were present in 45% of deaths and benzodiazepines were present in ~15%.

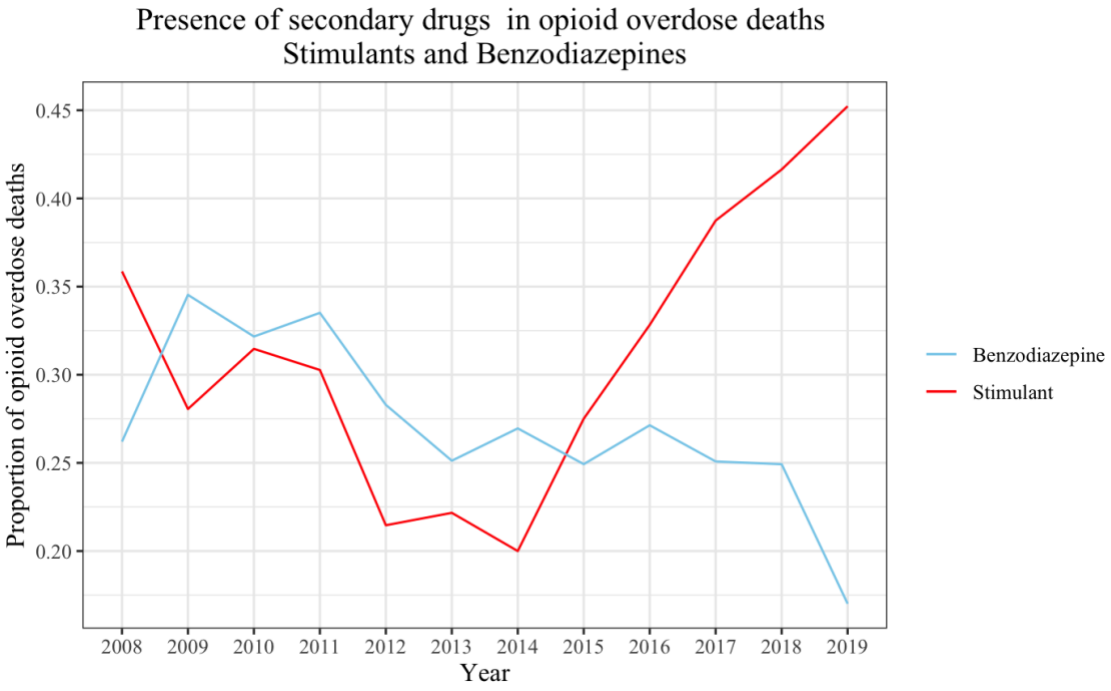

Figure 2: Presence of stimulants and benzodiazapines in opioid overdose deaths

As heroin has its own specific pattern of drug utilization in Allegheny County, its presence was compared to the presence of all other illicit opioids. Of all opioid overdose deaths, 51% (n = 1701) had a presence of an illicit opioid, excluding heroin, while 53% (n = 1735) of opioid overdose deaths had a presence of heroin specifically. The presence of heroin and other illicit opioid deaths were measured in overtime in Allegheny County from 2008-2019 (Figure 3). The proportion of opioid overdose deaths with heroin (blue) peaked in 2013 (67%), decreased until 2017 (41%), and then began to rise again (2019: 57%). The proportion of opioid overdose deaths with illicit opioids present (red) were relatively consistent until 2013 (3.6%), at which point there was a stark increase in deaths involving illicit opioids (2015: 36%; 2017: 86%) and has continued to increase until the point of analysis in 2019 (93%). Heroin related deaths were consistently higher than illicit opioid deaths until mid-2015 when the proportion of illicit opioid deaths became greater (2015 - heroin: 65%, illicit opioid: 36%; 2016 - heroin: 54%, illicit opioid: 74%).

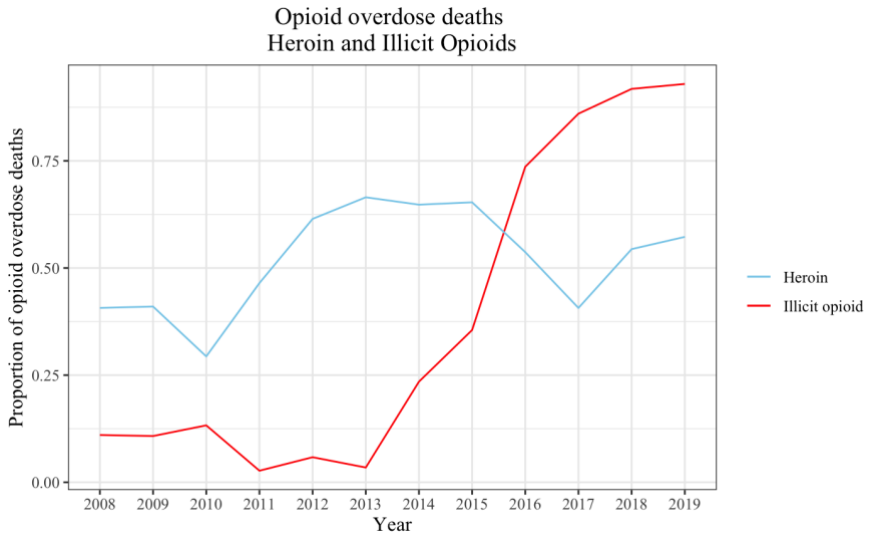

Figure 3: Heroin and illicit opioid overdose deaths

Figure 4 highlights the proportion of opioid overdose deaths involving stimulants by zip code area. Darker blue colors indicate a lower incidence rate while warmer green and yellow colors indicate higher incidence rate. Areas in gray indicate zip codes where there are no deaths, or data, available. Presence of stimulants became far more prominent in opioid overdose deaths between 2008 and 2019 (Figure 4). In 2008, the opioid and stimulant deaths were contained to the general Pittsburgh area, however by 2019 they had become denser across the greater Allegheny County area. While the overall proportion of overdose deaths involving stimulants has increasingly disseminated throughout majority of the county, as indicated by the gradient, there are two notable outliers in the 2019 data. In 2019, Russellton had the highest overall proportion of opioid deaths with a stimulant present, indicated by yellow on the map. However, the overall the population of the area is rather small with only 834 residents, and one resident died with opioids and stimulants present at death. Similarly, Leetsdale, indicated in green on the map, has a population of 1,162 residents, and one resident who died with opioids and stimulants present.

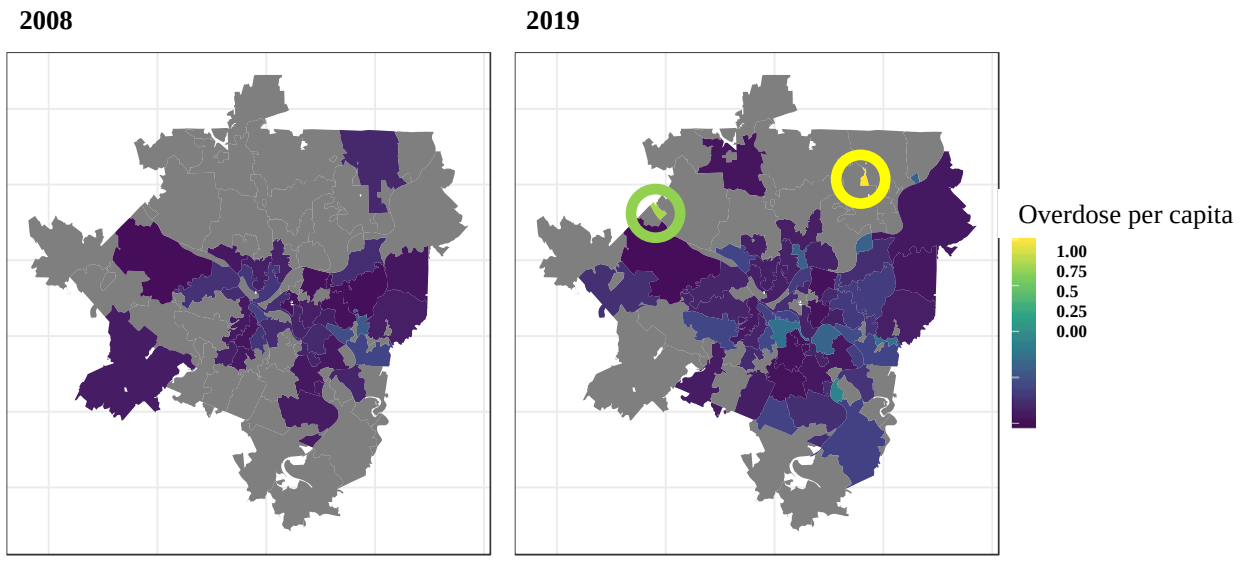

Figure 4: Rate of opioid overdose deaths with a stimulant present per 1,000

Figure 5 highlights the proportion of opioid overdose deaths involving benzodiazepines by zip code area. These data find that presence of benzodiazepines decreased between 2008 and 2019, consistent with previous finding. In 2019, there were overall higher proportions of overdose deaths involving benzodiazepines, although not drastically difference from 2008. Interestingly, unlike stimulant use (Figure 4), patterns of benzodiazepines are generally not consistent over time but generally impact suburban Allegheny County.

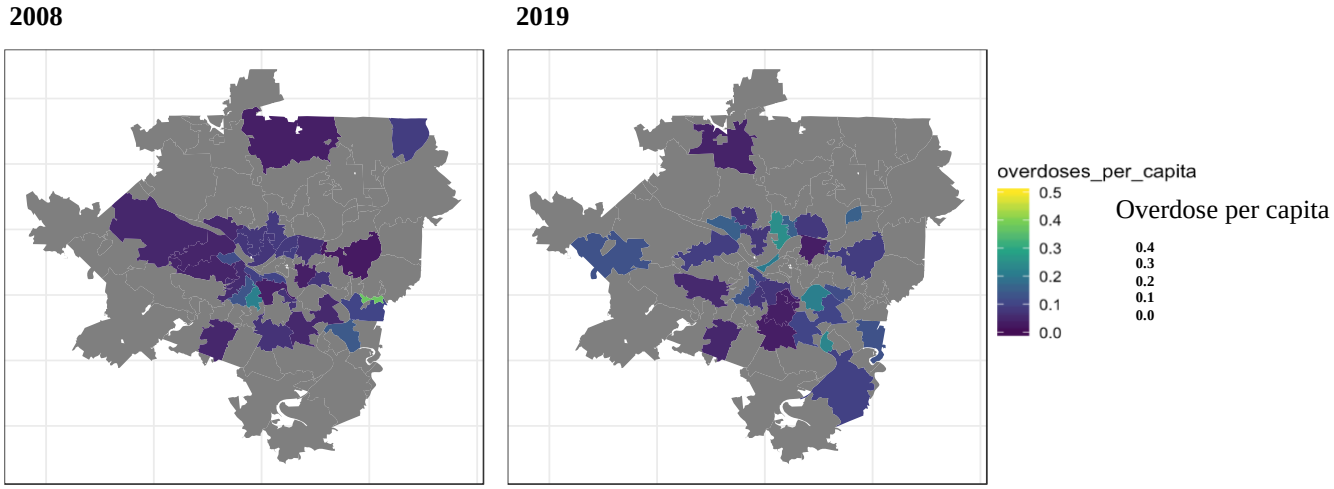

Figure 5: Rate of opiod overdose deaths with a benzodiazapine present per 1,000

Finally, the proportion of secondary drugs present in opioid overdose deaths between race and sex subgroups were compared (Table 7). The variation in maps between presence of benzodiazepines and stimulants suggested that there may be differences in overdose rate by race or sex. Overall, females had overall higher presence of secondary drugs, regardless of race, compared to males. Additionally, Whites had higher presence of benzodiazepines present while African Americans had higher presence of stimulants. Pearson's Chi Square tests were ran across the 4 subgroups (White male, White female, African American male, African American female) for stimulants (X_2=0.13, p = 0.71), benzodiazepines (X_2=0.0006, p = 0.98), and the intersection of stimulants and benzodiazepines (X_2=0.3, p = 0.58). No tests were significant suggesting that race and sex are not related, or rather independents, when interpreting the presence of stimulants, benzodiazepines, or their intersection in opioid overdose deaths.

Table 7: Proportion of deaths with secondary drug present by race

		Stimulant	Benzodiazapine	Stimulant and Benzodiazapine	*Overall Deaths*
White	**Female**	270 (30%)	306 (34%)	76 (8.5%)	***900***
	Male	588 (29%)	527 (26%)	143 (7.1%)	***2027***
Black	**Female**	71 (60%)	18 (15%)	11 (9.2%)	***119***
	Male	145 (52%)	28 (10%)	15 (5.4%)	***278***

3.2 Gaussian-mixture model analysis

The GMM analysis revealed nine distinct clusters. To summarize the defining features of each cluster, mean values were taken for each of the variables (Table 8). As the variables *age* is treated as a factor variable, the summary variable reports the most represented age group of each cluster.

Table 8: Cluster summary tables

cluster	# of subs	age group	% male	% af am	% illicit opioid	% stimulant	% benzo
1	321	55-64	72	25	100	36	22
2	300	35-44	47	4	16	20	100
3	539	25-34	72	5	0	22	30
4	169	35-44	75	3	0	0	0
5	646	45-54	72	16	57	43	11
6	303	55-64	66	22	0	27	28
7	259	35-44	70	15	70	100	12
8	540	25-34	70	9	100	32	22
9	241	35-44	78	7	100	0	17

The variables *illicitopioid, sex, race, stim,* and *benzo* are indicator variables, therefore the summary values for each cluster are proportions. In Table 8, for ease of interpretation, the variable *sex* is labeled as *male*, which shows the proportion of males in the given cluster, and the variable *race* is labeled as *af am*, which shows the proportion of African Americans in the given

cluster. As heroin has its own distinct pattern of use in Allegheny County over time, and the time was variable not included as this analysis was interested in classifying demographic subgroups, heroin was not included in the overall illicit opioid category. Alternative GMMs, with inclusion of heroin, can be observed in Appendix B.

Proportions of African American subjects ranged from 3% (cluster 4) to 25% (cluster 1) and were significant for all clusters as compared to the overall proportion of African American subjects for the dataset (Appendix A). In addition to cluster 1, which had a higher than average proportion of African Americans, cluster 6 (22%) did as well. Proportions for illicit opioid and stimulant use range from 0% to 100% meaning that some clusters were defined as having no, or all, subjects with illicit opioids or stimulants present at time of death. Benzodiazepines were present in all clusters but one (cluster 4) with proportions ranging from 11% (cluster 5) to 100% (cluster 2). Defining characteristics of each cluster can be observed below:

Cluster 1 (n = 321) had a most common age group of 55-64, a high proportion of African Americans (25%), all subjects with at least one illicit opioid present at time of death (100%), 36% of subjects with at least one stimulant, and 22% of subjects with a benzodiazepine present at time of death.

Cluster 2 (n = 300) had the highest proportion of female subjects (53%), a low proportion of African Americans (4%), a most common age group of 35-44, 16% with illicit opioid use, 20%with at least one stimulant, and 100% of subjects with at least one benzodiazepine present at time of death.

Cluster 3 (n = 539) had a most common age group of 25-34, a low proportion of African Americans (5%), no subjects with illicit opioids (0%), 22% with at least one stimulant, and 30% with a benzodiazepine present at time of death.

Cluster 4 (n = 169) had a most common age group of 35-44, 25% of subjects were female, a low proportion of African Americans (3%), and no subjects had illicit opioids, a stimulant, or benzodiazepine present at time of death.

Cluster 5 (n = 646) had a most common age group of 45-54, 16% of the cluster was African American, 28% were female, 57% of subjects with at least one illicit opioid present, 43% of subjects with stimulants present, and 11% of subjects with at least one benzodiazepine present at time of death.

Cluster 6 (n = 303) had a high proportion of female subjects (34%), a high proportion of African American subjects (22%), a most common age group of 55-64, no subjects with illicit opioid present (0%), 27% of subjects with at least one stimulant present, and 28% of subjects with at least one benzodiazepine present at time of death.

Cluster 7 (n = 259) had a most common age group of 35-44, a high proportion of female subjects (30%), 15% of subjects were African American, 70% of subjects with at least one illicit opioid present, 27% subjects with at least one stimulant present, and 28% of subjects with at least one benzodiazepine present at time of death.

Cluster 8 (n = 540) had a most common age group of 25-44, all subjects with at least one illicit opioid present (100%), 32% of subjects had at least one stimulant present, and 22% of subjects had at least one benzodiazepine present at time of death.

Cluster 9 (n = 241) had a most common age group of 35-44, all subjects with at least one illicit opioid present (100%), no subjects had a stimulant present (0%), and 17% of subjects had at least one benzodiazepine present at time of death.

Figure 5 displays how each of the clusters are uniquely defined by the given variables and how they relate to each other in space. Each variable in the analysis can be observed across

the diagonal while each graph in the grid represents how the nine clusters relate to each other with respect to any two of the given variables in the analysis. The axes on each graph indicate the proportion of the variable present in the cluster, except for age which is a factor variable with six different levels. The ellipses in each graph represents a particular cluster. The shape of the ellipse indicates the homogeneity of the cluster. The more condensed (small, circular) the ellipse, the more homogenous the cluster; the larger and more oblong the ellipse, the more heterogenous the cluster. In each graph there are various shapes of various colors which represent the different levels of the two variables in the graph.

Inspecting the variables race and sex, all clusters are predominantly condensed to quadrants indicative of White males, with some variation in shape of the ellipses indicating heterogeneity. The clusters in the graph for age and illicit opioids are more dispersed. Age groups 2 (25-34 years old), 3 (35-44 years old) and 5 (55-64 years old) have prominent clusters are one illicit opioid present at time of death and three opioids present at time of death. Age group 4 (45-54 years old) has two distinct clusters for two illicit opioids present at time of death. All other graphs can be interpreted in a similar manner.

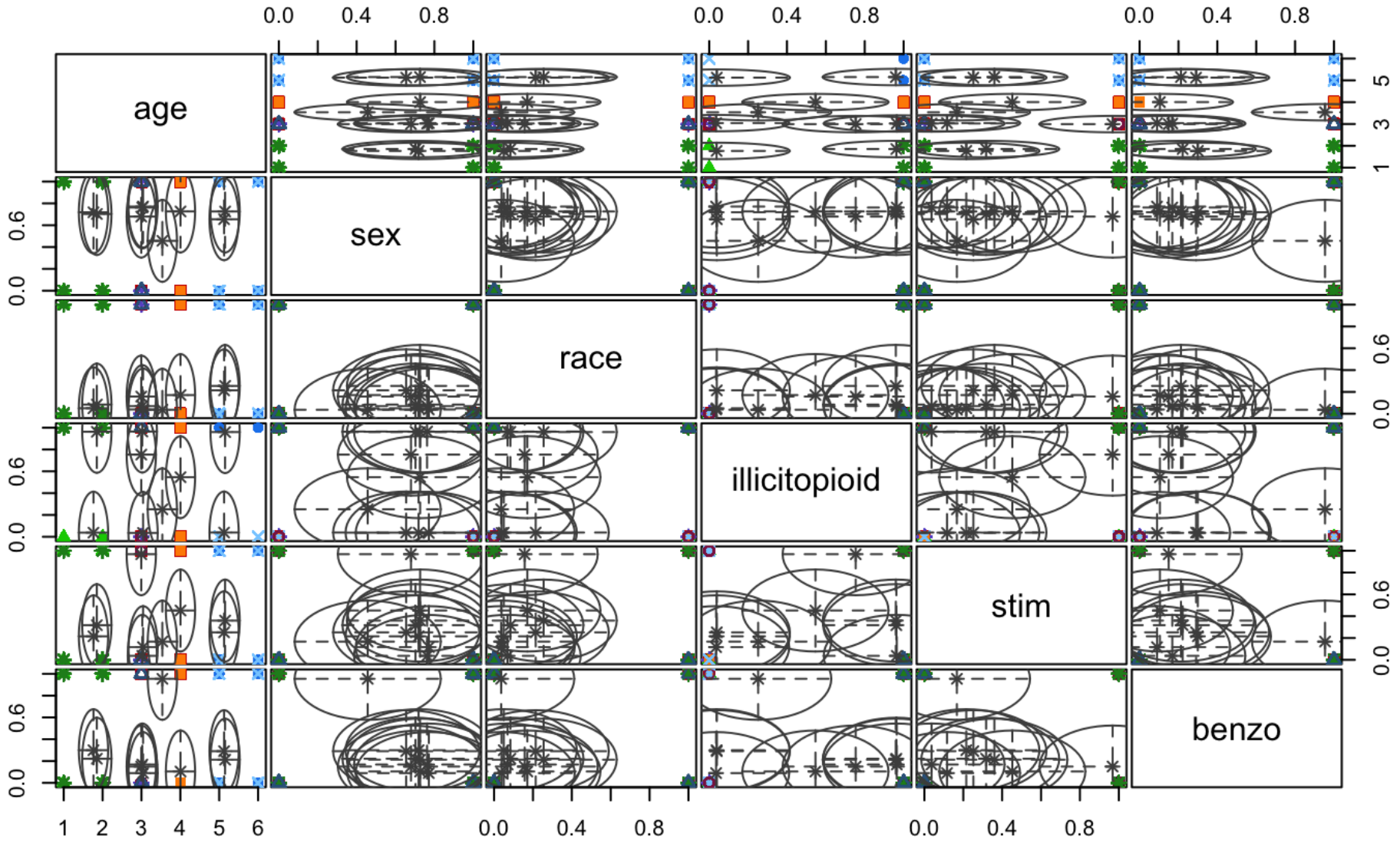

Figure 6: Gaussian-Mixture Model Clusters

4.0 Discussion

The first aim of this thesis was to describe the progression of polysubstance use and opioid epidemic in Allegheny County from 2008-2019. The next aim was to investigate patterns of secondary nonopioid drugs present in opioid overdose deaths in Allegheny County from 2008-2019 to develop advanced characterizations of subpopulations.

Comparison of AAMRs indicated that Allegheny County consistently had higher opioid overdose deaths as compared to Pennsylvania and the United States from 2008-2019 (Figure 1). The trajectory, or change, in AAMR for each population (Allegheny County, Pennsylvania, and the United States) was similar throughout the time period of interest (the lines were generally parallel) until 2017. This was to be expected based on previous research (Hulsey et al., 2016). AAMRs for Pennsylvania and the U.S. rose from 2017-2018 while there was a peak followed by a very stark decline in AAMR for Allegheny County. By 2018, the AAMR for Allegheny County was similar to that of Pennsylvania, although still higher, and Pennsylvania was higher than the U.S. Although a sharp decline in opioid-related overdose deaths (OOD) is advantageous, it is worth investigating (1) what could drive such a large change in a short timeframe and (2) if it is response to harm-reduction interventions. Previous research indicates that there was a 5% drop in OOD in the U.S. in 2018, likely due to decline in prescription overdose deaths (Goodnough et al., 2019; Lopez, 2019), a change not nearly as drastic as the one in Allegheny County. One possible hypothesis is that a surge in fentanyl-related overdose deaths in 2017 killed a large portion of opioid drug users in Allegheny County, and resulted in a large drop-off in deaths in 2018. Hulsey et al., (2016) indicated an increasing presence in fentanyl starting in 2014. As fentanyl is one of the most lethal drugs and contributions to OOD, it is

possible that deaths involving fentanyl peaked in 2017. It is important to consider that the peak in fentanyl killed a large portion of the opioid using population in Allegheny County in 2017 and as a result there were fewer deaths in 2018, not simply that harm reduction interventions were driving this drastic decline in deaths in 2018, although possible. The drastic decline of fentanyl users in Allegheny County combined with the overall decrease in prescription drug deaths could provide insight for the large drop-off in AAMR in 2018. This analysis provided justification for a deeper analysis of which subpopulations are being impacted by the epidemic and what patterns of polysubstance use are emerging.

The next aim was to investigate polysubstance use involvement in opioid overdose deaths. For this analysis, polysubstance use was quantified by restricting to presence of stimulants or benzodiazepines. This aim was conducted via various analyses: (1) map visualizations of Allegheny county by zip code and (2) graphical trends of polysubstance use over time in Allegheny County. The map visualizations for stimulants (Figure 4) and benzodiazepines (Figure 5) indicates that overall, polydrug use involvement in opioid overdose deaths has increased over time, with a very large increase in stimulant presence. It is interesting to assess differences between stimulants and benzodiazepines. Stimulant presence is far more widespread across the county, but with particular prominence in the greater Pittsburgh area, while benzodiazepines presence has become sparser in recent years and is near non-existent in Pittsburgh but rather is prominent across the suburban and rural areas of Allegheny County. This is important to consider when assessing who is impacted by polysubstance use. Based on census data, the majority of the African American population reside in Pittsburgh rather than the surrounding areas of Allegheny county (U.S. Census Bureau, 2020). This would suggest that the African American population is being impacted by the presence of stimulants in opioid overdose

deaths, while the White population is being more impacted by benzodiazepine presence, which was indeed confirmed in the analyses. While Pearson chi square tests among subgroups (White male, White female, African American male, and African American female) for stimulants and benzodiazepines were not significant, it does not mean that a difference among the subgroups does not exist, rather the current analysis may be too underpowered to detect a difference. Furthermore, determining what is occurring between these subgroups could aid harm reduction techniques and could improve educational materials.

Additional investigation into polysubstance over time can be assessed in Figure 2. The presence of benzodiazepine involvement in opioid overdose deaths (blue) has seen a decline over the past 10 years while the presence of stimulants (red) saw a decline until 2014, at which point there was a drastic increase and has continued to increase until 2019. Intriguingly the two drug categories are trending inversely of the other, particularly after 2014. It is interesting to consider the relationship between the increase in stimulant involvement in opioid overdose deaths and the change in AAMR in Allegheny County. One potential hypothesis considers the opioid naïve population, or those who have no or low tolerance for opioids. LaRue et al., (2019) states this population is at a heightened risk for overdose death if they are exposed to highly potent opioids such as fentanyl as they have no tolerance for it, nor are the expecting or prepared for an overdose. A retrospective analysis of NYC death certificate and toxicology data from 2000-2016 found that cocaine, a stimulant, was found in 53% of accidental drug overdose deaths, with 58% of those deaths involving some form of opioid (fentanyl, heroin, etc) (B. H. Han et al., 2019). Another study conducted by Hoots et al., (2020) identified overdose deaths in the Nationwide Emergency Department Sample (NEDS) found that cocaine-related deaths involving opioids had an annual percent change of 46% from 2014-2017.

As previously mentioned, insight as to which subgroups are being impacted by this epidemic would greatly aid harm reduction education, interventions, and policy reform. The GMM analysis revealed 9 distinct clusters, each with their own defining characteristics (Table 8; Figure 5). The goal of the clustering analysis is to identify homogenous groups with similar features. This analysis identified nine latent subgroups, three of which were selected because they could easily benefit from targeted interventions: (1) White females, ages 35-44, with presence of benzodiazepines and opioids, (2) older African American males, ages 55-64, with presence of illicit opioids and stimulants, and (3) White males, ages 35-44, who are utilizing heroin and/or prescription opioids.

Since 1999, the overall crude drug overdose deaths rate for females increased by 200% while those specifically involving benzodiazepines increased by 830% (VanHouten, 2019). From the clustering analysis it can discerned that White females, ages 35-44, utilizing benzodiazepines and opioids tandem, generally, reside in rural and suburban Allegheny County. Previous research suggests that comorbidity of chronic pain and anxiety are at the root of this subepidemic of overdose deaths with the realm of opioids. Women are 1.5-2 times more likely to receive an anxiety disorder diagnosis, of which the medication regime is often a benzodiazepine (Knight, 2017). Prescription drug monitoring programs (PDMPs) have begun targeting co-prescriptions of opioids and benzodiazepines upon evidence identifying the lethal drug combination (*Are Benzodiazepines Our Next Prescription Drug Crisis?*, 2018; Jones & McAninch, 2015). This enforced targeting could provide intuition as to what was driving the large drop in benzodiazepine related deaths in Allegheny County (Figure 2). Moreover, parsing apart the heterogeneous nature of the comorbid diagnoses is necessary to reduce overdose deaths. As opposed to offering a medication solution to anxiety, providers could instead suggest

talk therapy and then revisit medication, if necessary. If medication is deemed appropriate, the provider could prescribe an SSRI as opposed to a benzodiazepine, which would be a last case scenario solution. In addition to these alternatives informed education from providers to patients is necessary and critical, particularly for transferring the message of the risks of mixing opioids and benzodiazepines.

Two clusters of older African American males were identified for having overdosed with illicit opioids and stimulants. Previous research indicates that of the racial ethnic groups, African Americans have the highest rate of cocaine-related overdose deaths (8.3 per 100,000) (Kariisa et al., 2019). From a harm reduction standpoint, it would necessary to determine the order of drug ingestion, from which two scenarios could arise: (1) persons are intentionally mixing stimulants and opioids together i.e. speedballing (Hoots et al., 2020.; LaRue et al., 2019) or (2) persons are unaware that stimulants are being altered with opioids thus introducing an opioid naïve population (LaRue et al., 2019). While the nature of these data deems it impossible to conclude the order of drug consumption, knowledge of the subgroup can inform policy makers and harm reduction experts. Public health advocacy programs, such as Prevention Point Pittsburgh (Bennett et al., 2011), are already addressing such goals. By providing educational materials, on both accidental opioid poisoning and the risks of speedballing, and naloxone kits to the community, there are not only preventing overdose deaths but providing resources for drug users to help others be aware of their surroundings. Disseminating these processes throughout various communities and prevention programs could greatly aid in reducing overdose deaths.

A cluster of White males, ages 35-44, were identified as having no illicit opioids, stimulants, or benzodiazepines present at time of death meaning these males had either (1) non-illicit, prescription opioids, or (2) heroin present at time of death. In the beginning of the opioid

epidemic, this population was known the most impacted. Research indicated that White males were utilizing prescription opioids and then, due to various circumstances, turned to heroin for accessibility and cost (*History of the Opioid Epidemic*, n.d.). This cluster was the smallest out of the nine with 169 patients, suggesting that efforts and prevention techniques targeted at this subgroup are moderately successful. Knowledge from the providers is necessary, informing patients of the risks of opioid dependence and downstream effects of such a disorder.

5.0 Conclusion

Allegheny County has an ever-changing drug presence. Heroin-related deaths peaked in 2013 (Figure 3), at which point the presence of other illicit opioid (Figure 3) and stimulant-related opioid deaths drastically increased (Figure 2). The heterogeneity of the drug presence in Allegheny County makes it necessary to develop further advanced characteristics of the subgroups being impacted by this epidemic. Statistical learning and GMM provided an optimal tool to generate such inferences.

This analysis revealed that indeed there are various subgroups of drug users in Allegheny County for whom different approaches to harm reduction would be advantageous (Nolan et al., 2019). In rural Allegheny County, where benzodiazepines are more prominent, White females ages 35-44 should be more informed of the risks of mixing benzodiazepines and opioids. In Pittsburgh, or more urban areas of Allegheny County, African American males ages 35-44 should be aware of the potential for cocaine to be mixed with fentanyl or should know the risk of speeding balling. Perhaps the most vulnerable are the opioid-naïve stimulant users. This population is typically not targeted or receptive to harm reduction strategies as they not anticipating overdosing from an opioid. White males, ages 35-44, should be aware of the increased risk of ingesting opioids that are non-illicit as well as heroin. Additionally, supplemental harm reduction techniques should be provided such as fentanyl testing strips or naloxone kits.

While these analyses did expand upon previous ideas, there are various limitations. The GMM, although successful, could be improved. Heroin has its own specific pattern of use in Allegheny County. However, as the analysis was identifying demographic subgroups only, and

not subgroup changes over time, heroin was included in illicit opioids. However, future iterations could include death year and heroin as variables to further define subgroups. The BIC from the model indicated that a lower the number of components could be sufficient for the model, however as the purpose of this analysis was to investigate demographic patterns, it was determined that it was necessary to include all. Additionally, this analysis only assessed stimulants and benzodiazepines as metric for polysubstance use in opioid overdose deaths. Inclusion of other secondary drugs may aid in interpretation and provide further insight as to the various subgroups impacted by the epidemic. Next, the analyses were restricted to that of Allegheny County and therefore results should be interpreted with caution if generalized to different areas.

The methodologies utilized in this thesis could be further applied to the realm of the opioid epidemic. Future studies, with richer datasets, could include other races or ethnicities to development characterizations of polydrug use. Another way of advancing this research is to incorporate zip codes. While the map visualizations and clustering analysis provide interesting conclusions, conducting a clustering analysis with zip codes incorporated will provide more acutely distinct clusters to target specific areas and zip codes modalities for various subgroups. Additionally, as opposed to assessing mortality data from strictly a retrospective standpoint, analysis could be conducted on actively participating safe needle injections sites or in prisons utilizing Medication-Assisted Therapy (MAT). By doing so, patterns of polysubstance opioid use can be identified while drug users are still alive to create proper harm reduction measures and interventions in a prospective manner and could add to information previously defined by retrospective analyses.

Overall, this analysis can aid public health professionals and policy experts to further improve interventions and harm reduction measures for the opioid epidemic in Allegheny County, Pennsylvania.

Appendix A Statistical Tests

Race

A one-sample z-test of proportions for African Americans in each cluster was compared to the overall proportion of African Americans of in the dataset (12%). All clusters were found to be statistically different from the hypothesized proportion.

Appendix Table 1: Cluster specific chi square tests for race

Cluster	Test statistic	p-value
1	48.06	<0.0001
2	19.81	<0.0001
3	25.08	<0.0001
4	13.13	<0.0001
5	10.92	<0.0001
6	31.08	<0.0001
7	2.86	<0.0001
8	6.28	<0.0001
9	4.73	<0.0001

Sex

A one-sample z-test of proportions for males in each cluster was compared to the overall proportion of males in the dataset (69%). Clusters 2 and 9 were found to be statistically different from the hypothesized proportion.

Appendix Table 2: Cluster specific chi square tests for sex

Cluster	Test statistic	p-value
1	1.05	0.306
2	68.25	<0.0001
3	1.55	0.213
4	2.17	0.141
5	2.40	0.122
6	1.28	0.259
7	0.01	0.940
8	0.28	0.599
9	9.38	0.002

Appendix B Alternative Gaussian-Mixture Models

GMM 1: Including heroin in illicit opioid

```
Clustering table:
  1   2    3   4   5    6   7   8
  0 1226 874 450 196    0 395 177
```

classification <dbl>	age <dbl>	sex <dbl>	race <dbl>	illicitopioid <dbl>	stim <dbl>	benzo <dbl>
2	3.442904	0.6525285	0.3254486	0.7316476	0.3564437	0.7128874
3	2.686499	1.0000000	0.0000000	0.8592677	0.0000000	0.0000000
4	3.224444	1.0000000	0.0000000	0.8466667	1.0000000	0.0000000
5	3.270408	0.0000000	0.0000000	0.8316327	1.0000000	0.0000000
7	3.093671	0.0000000	0.0000000	0.7341772	0.0000000	0.0000000
8	5.163842	1.0000000	0.0000000	0.7740113	0.0000000	0.0000000

GMM 2: Including new heroin variable in model, removing heroin from illicit

```
Clustering table:
  1   2
2689 629
```

classification <dbl>	age <dbl>	sex <dbl>	race <dbl>	illicitopioid <dbl>	stim <dbl>	benzo <dbl>	heroin <dbl>
1	3.235032	0.6898475	0.08293046	0.5042767	0.2015619	0.2729639	0.4113053
2	3.335453	0.7090620	0.27980922	0.5484897	0.8600954	0.2225755	1.0000000

Appendix Figure 1: Alternative Gaussian-Mixture Models

Appendix C Analysis Scripts Executed in R

````
title: "Opioid Overdose Deaths in Allegheny County, 2008-2019"
author: "Kayleigh Adamson"
date: "1/26/2020"
output: pdf_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

# Import data and necessary packages
```{r}
# read in data
opioid_full <- read.csv("fixed-deduped-mod-1c59b26a-1684-4bfb-92f7-205b947530cf.csv", header= TRUE)

# install necessary packages
library(tidyverse)
library(dplyr)
library(reshape2)
library(stringr)
library(lubridate)
library(ggplot2)
library(mclust)
library(scales)
```

# Clean data
```{r}
# remove unused columns
opioid_clean <- select(opioid_full,-c(combined_od10,decedent_zip, case_dispo, death_date_and_time)) # remove
unused columns

# race
opioid_clean <- subset(opioid_clean,race == 'B' | race == 'W')
opioid_clean$race <- ifelse(opioid_clean$race=='B',1,0) # black = 1, white = 0

# sex
opioid_clean$sex <- ifelse(opioid_clean$sex=='M',1,0)

# manner of death
opioid_clean$manner_of_death[opioid_clean$manner_of_death == "Accidents"] <- "Accident"

# zip code
zips <- read.delim("allegheny_zip.txt", header = FALSE)
opioid_clean <- filter(opioid_clean, incident_zip %in% zips$V1)
opioid_clean <- opioid_clean%>%
  filter(incident_zip != "")

opioidlist <- read.delim("opioid.txt", header = FALSE)

opioid_clean$op1 <- ifelse(opioid_clean$combined_od1 %in% opioidlist$V1, 1, 0)
opioid_clean$op2 <- ifelse(opioid_clean$combined_od2 %in% opioidlist$V1, 1, 0)
opioid_clean$op3 <- ifelse(opioid_clean$combined_od3 %in% opioidlist$V1, 1, 0)
opioid_clean$op4 <- ifelse(opioid_clean$combined_od4 %in% opioidlist$V1, 1, 0)
opioid_clean$op5 <- ifelse(opioid_clean$combined_od5 %in% opioidlist$V1, 1, 0)
````
````

```
opioid_clean$op6 <- ifelse(opioid_clean$combined_od6 %in% opioidlist$V1, 1, 0)
opioid_clean$op7 <- ifelse(opioid_clean$combined_od7 %in% opioidlist$V1, 1, 0)
opioid_clean$op8 <- ifelse(opioid_clean$combined_od8 %in% opioidlist$V1, 1, 0)
opioid_clean$op9 <- ifelse(opioid_clean$combined_od9 %in% opioidlist$V1, 1, 0)
opioid_clean$opioidcount <- rowSums(opioid_clean[, c(16:24)]) # create summed var
opioid_clean <- opioid_clean[-c(16:24)]

opioid_clean <- opioid_clean %>%
  filter(opioidcount != 0) # only include vars not zero

# bin age: <=24, 25-34, 35-44, 45-54, 55-64, and 65+, less than or equal to
opioid_clean$age = findInterval(opioid_clean$age, c(0, 24, 34, 44, 54,64,Inf))

# factor year
opioid_clean$case_year <- as.factor(opioid_clean$case_year)

```

Indicator columns for illicit opioid (counts), stimulant and benzos
```{r}
# illicit opioid
ill_opioidlist <- read.delim("illicit_opioid.txt", header = FALSE)
opioid_clean$iop1<-ifelse(opioid_clean$combined_od1 %in% ill_opioidlist$V1, 1, 0)
opioid_clean$iop2<-ifelse(opioid_clean$combined_od2 %in% ill_opioidlist$V1, 1, 0)
opioid_clean$iop3<-ifelse(opioid_clean$combined_od3 %in% ill_opioidlist$V1, 1, 0)
opioid_clean$iop4<-ifelse(opioid_clean$combined_od4 %in% ill_opioidlist$V1, 1, 0)
opioid_clean$iop5<-ifelse(opioid_clean$combined_od5 %in% ill_opioidlist$V1, 1, 0)
opioid_clean$iop6<-ifelse(opioid_clean$combined_od6 %in% ill_opioidlist$V1, 1, 0)
opioid_clean$iop7<-ifelse(opioid_clean$combined_od7 %in% ill_opioidlist$V1, 1, 0)
opioid_clean$iop8<-ifelse(opioid_clean$combined_od8 %in% ill_opioidlist$V1, 1, 0)
opioid_clean$iop9<-ifelse(opioid_clean$combined_od9 %in% ill_opioidlist$V1, 1, 0)

opioid_clean$illicitopioid<-rowSums(opioid_clean[, c(17:25)]) # create summed var
opioid_clean$illicitopioid <- ifelse(opioid_clean$illicitopioid> 0,1,0)
opioid_clean <- opioid_clean[-c(17:25)]

# stimulant
stimlist <- read.delim("stimulants.txt", header = FALSE)
opioid_clean$stim1 <- ifelse(opioid_clean$combined_od1 %in% stimlist$V1, 1, 0)
opioid_clean$stim2 <- ifelse(opioid_clean$combined_od2 %in% stimlist$V1, 1, 0)
opioid_clean$stim3 <- ifelse(opioid_clean$combined_od3 %in% stimlist$V1, 1, 0)
opioid_clean$stim4 <- ifelse(opioid_clean$combined_od4 %in% stimlist$V1, 1, 0)
opioid_clean$stim5 <- ifelse(opioid_clean$combined_od5 %in% stimlist$V1, 1, 0)
opioid_clean$stim6 <- ifelse(opioid_clean$combined_od6 %in% stimlist$V1, 1, 0)
opioid_clean$stim7 <- ifelse(opioid_clean$combined_od7 %in% stimlist$V1, 1, 0)
opioid_clean$stim8 <- ifelse(opioid_clean$combined_od8 %in% stimlist$V1, 1, 0)
opioid_clean$stim9 <- ifelse(opioid_clean$combined_od9 %in% stimlist$V1, 1, 0)

opioid_clean$stim <- rowSums(opioid_clean[, c(18:26)]) # create summed var
opioid_clean$stim <- ifelse(opioid_clean$stim> 0,1,0)
opioid_clean <- opioid_clean[-c(18:26)]

#benzodiaz
benzolist <- read.delim("benzo.txt", header = FALSE)
opioid_clean$benzo1 <- ifelse(opioid_clean$combined_od1 %in% benzolist$V1, 1, 0)
opioid_clean$benzo2 <- ifelse(opioid_clean$combined_od2 %in% benzolist$V1, 1, 0)
opioid_clean$benzo3 <- ifelse(opioid_clean$combined_od3 %in% benzolist$V1, 1, 0)
opioid_clean$benzo4 <- ifelse(opioid_clean$combined_od4 %in% benzolist$V1, 1, 0)
opioid_clean$benzo5 <- ifelse(opioid_clean$combined_od5 %in% benzolist$V1, 1, 0)
opioid_clean$benzo6 <- ifelse(opioid_clean$combined_od6 %in% benzolist$V1, 1, 0)
opioid_clean$benzo7 <- ifelse(opioid_clean$combined_od7 %in% benzolist$V1, 1, 0)
opioid_clean$benzo8 <- ifelse(opioid_clean$combined_od8 %in% benzolist$V1, 1, 0)
```

```
opioid_clean$benzo9 <- ifelse(opioid_clean$combined_od9 %in% benzolist$V1, 1, 0)

opioid_clean$benzo <- rowSums(opioid_clean[, c(19:27)]) # create summed var
opioid_clean$benzo <- ifelse(opioid_clean$benzo> 0,1,0)
opioid_clean <- opioid_clean[-c(19:27)]

#heroin
opioid_clean$heroin1 <- ifelse(opioid_clean$combined_od1 == "HEROIN", 1, 0)
opioid_clean$heroin2 <- ifelse(opioid_clean$combined_od2 == "HEROIN", 1, 0)
opioid_clean$heroin3 <- ifelse(opioid_clean$combined_od3 == "HEROIN", 1, 0)
opioid_clean$heroin4 <- ifelse(opioid_clean$combined_od4 == "HEROIN", 1, 0)
opioid_clean$heroin5 <- ifelse(opioid_clean$combined_od5 == "HEROIN", 1, 0)
opioid_clean$heroin6 <- ifelse(opioid_clean$combined_od6 == "HEROIN", 1, 0)
opioid_clean$heroin7 <- ifelse(opioid_clean$combined_od7 == "HEROIN", 1, 0)
opioid_clean$heroin8 <- ifelse(opioid_clean$combined_od8 == "HEROIN", 1, 0)
opioid_clean$heroin9 <- ifelse(opioid_clean$combined_od9 == "HEROIN", 1, 0)

opioid_clean$heroin <- rowSums(opioid_clean[, c(20:28)]) # create summed var
opioid_clean$heroin <- ifelse(opioid_clean$heroin> 0,1,0)
opioid_clean <- opioid_clean[-c(20:28)]

# remove text columnns
opioid_clean <- opioid_clean[-c(1, 5:13)]

## write.csv(opioid_clean, "opioid_clean.csv")
```

Descriptive statistics, AAMR
```{r}
# age adjusted mortality rates
death_year <- opioid_clean %>%
  group_by(age, case_year)%>%
  count()

aamr = read.csv("ageadjust_mortalityrate.csv", header=TRUE)
aamr = as.data.frame(aamr)

ggplot(data=aamr, aes(x=Year)) +
  geom_line(aes(y=United.States, color = "United States")) +
  geom_line(aes(y=Pennsylvania, color = "Pennsylvania")) +
  geom_line(aes(y=Allegheny.County, color = "Allegheny County")) +
  labs(x="Year", y = "Age-adjusted mortality rate per 100,000", color = "")+
  ggtitle("Age-adjusted opioid-related mortality rates")+
  scale_x_continuous(breaks= pretty_breaks())+
  theme_bw()+
  theme(plot.title = element_text(hjust = 0.5))
```

Descriptive statistics, demographic and summary
```{r}
# table race by sex, count
demo = table(opioid_clean$race, opioid_clean$sex)
colnames(demo) <- c("Female","Male")
rownames(demo) <- c("White","Black")
prop.table(demo)

# prop male (1), female (0)
prop.table(table(opioid_clean$sex))

# prop by race, black = 1, white = 0
prop.table(table(opioid_clean$race))
```

```
# different number of opioids present
prop.table(table(opioid_clean$opioidcount))

# stim present or not
prop.table(table(opioid_clean$stim))

# benzo present or not
prop.table(table(opioid_clean$benzo))

# illcit present or not
prop.table(table(opioid_clean$illicitopioid))
sum(ifelse(opioid_clean$illicitopioid>0,1,0))/nrow(opioid_clean)

# heroin present or not
prop.table(table(opioid_clean$heroin))

```

Descriptive statistics, benzo v stimulant
```{r}
# BENZOS
# benzo by year
df_benzo = opioid_clean%>%
  group_by(case_year)%>%
  summarise(mean=mean(benzo), n=n())

# benzo by race and sex
opioid_clean%>%
  group_by(race, sex)%>%
  summarise(mean=mean(benzo), n=n())

benzo = opioid_clean %>% subset(benzo==1)
benzo_chi = chisq.test(benzo$sex, benzo$race)

# STIMULANTS
# stim by year
df_stim = opioid_clean%>%
  group_by(case_year)%>%
  summarise(mean=mean(stim), n=n())

# stim by race and sex
opioid_clean%>%
  group_by(race, sex)%>%
  summarise(mean=mean(stim), n=n())

stim = opioid_clean %>% subset(stim==1)
stim_chi = chisq.test(stim$sex, stim$race)

# PLOT
# benzo and stim by year
poly_year = ggplot(data=df_stim, aes(x=case_year, y=mean, group=1, color = "Stimulant"))+
  geom_line() +
  geom_line(data=df_benzo, aes(x=case_year, y=mean, group=1, color = "Benzodiazepine")) +
  labs(x="Year", y = "Proportion of opioid overdose deaths", color = "")+
  ggtitle("Presence of secondary drugs  in opioid overdose deaths \nStimulants and Benzodiazepines")+
  scale_color_manual(values=c('skyblue','red'))+
  theme_bw()+
  theme(plot.title = element_text(hjust = 0.5))

ggsave("poly_year.pdf")
```

Descriptive statistics, stim and benzo intersection
```{r}
# benzo by race and sex
opioid_clean%>%
  group_by(race, sex)%>%
  summarise(mean=mean(benzo==1 & stim==1), n=n())

opioid_clean %>%
  subset(benzo==1 & stim==1)%>%
  group_by(race, sex)%>%
  summarise(mean=mean(n), n=n())

# chi sq test
benzo_stim = opioid_clean %>% subset(benzo==1 & stim==1)
benzo_stim_chi = chisq.test(benzo_stim$sex, benzo_stim$race)
```

Descriptive statisitics, illict v heroin
```{r}
# illcit by year
df_illicit = opioid_clean%>%
  group_by(case_year)%>%
  summarise(mean=mean(illicitopioid), n=n()) %>%
  rename(illicit_mean = mean)


illicit = opioid_clean %>% subset(illicitopioid==1)
illicit_chi = chisq.test(illicit$sex, illicit$race)

# heroin by year
df_heroin = opioid_clean%>%
  group_by(case_year)%>%
  summarise(mean=mean(heroin), n=n()) %>%
  rename(heroin_mean = mean)

heroin = opioid_clean %>% subset(heroin==1)
heroin_chi = chisq.test(heroin$sex, heroin$race)

# two sample z for illicit v heroin by year
illicit_heroin = merge(df_illicit, df_heroin, by.x="case_year", by.y="case_year")
illicit_heroin=illicit_heroin[-c(3,5)]

# plot
test = ggplot(data=df_illicit, aes(x=case_year, y=mean, group=1, color = "Illicit opioid"))+
  geom_line() +
  geom_line(data=df_heroin, aes(x=case_year, y=mean, group=1, color = "Heroin")) +
  labs(x="Year", y = "Proportion of opioid overdose deaths", color = "")+
  ggtitle("Opioid overdose deaths \nHeroin and Illicit Opioids")+
  scale_color_manual(values=c('skyblue','red'))+
  theme_bw()+
  theme(plot.title = element_text(hjust = 0.5))

ggsave("poly_year_illicit.pdf")
```

Statistical learning, Gaussian Mixture Model
```{r}
set.seed(15232)

# clean data frame
opioid_analysis <- opioid_clean
opioid_analysis$incident_zip<- NULL
```

```
opioid_analysis$case_year<- NULL
opioid_analysis$opioidcount<- NULL
opioid_analysis$heroin = NULL

# model summary
opioidMclust<- Mclust(opioid_analysis)
summary(opioidMclust)
# plot(opioidMclust)

# summarize all vars by cluster group
opioid_analysis$classification =  opioidMclust$classification
opioid_analysis%>%
  group_by(classification)%>%
  summarise_all(funs(mean))
```

GMM- one sample z test for clusters, MALES
```{r}
male_clusters = opioid_analysis%>%
  subset(sex==1) %>%
  group_by(classification)%>%
  count(sex)
male_clusters_n = male_clusters$n
cluster_n = c(321, 300, 539, 169, 646, 303, 259, 540, 241)
p = 2301/3319 # prop of males in data

# one sample z test of proportions for each cluster
ztest_male_cluster = Map(prop.test,male_clusters_n,cluster_n, p=p, correct =FALSE)

teststat_male_cluster = t(sapply(ztest_male_cluster,"[[","statistic"))
pvalue_male_cluster = t(sapply(ztest_male_cluster,"[[","p.value"))
```

GMM- one sample z test for clusters, AF AM
```{r}
afam_clusters = opioid_analysis%>%
  subset(race==1) %>%
  group_by(classification)%>%
  count(race)
afam_clusters_n = afam_clusters$n
cluster_n = c(321, 300, 539, 169, 646, 303, 259, 540, 241)
p = 399/3318 # prop of af am in data

# one sample z test of proportions for each cluster
ztest_afam_cluster = Map(prop.test,afam_clusters_n,cluster_n, p=p, correct =FALSE)

teststat_afam_cluster = t(sapply(ztest_afam_cluster,"[[","statistic"))
pvalue_afam_cluster =t(sapply(ztest_afam_cluster,"[[","p.value"))
```

Statistical learning, Gaussian Mixture Model with heroin
```{r}
# clean data frame
opioid_analysis2 <- opioid_clean
opioid_analysis2$incident_zip<- NULL
opioid_analysis2$case_year<- NULL
opioid_analysis2$opioidcount<- NULL

# model summary
opioidMclust2<- Mclust(opioid_analysis2)
summary(opioidMclust2)
```

```
# plot(opioidMclust2)

# summarize all vars by cluster group
opioid_analysis2$classification =  opioidMclust2$classification
opioid_analysis2%>%
  group_by(classification)%>%
  summarise_all(funs(mean))
```

```
---
title: "Thesis map visualizations"
author: "Kayleigh Adamson"
date: "3/2/2020"
output: pdf_document
---
# Load necessary packages
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)

library(tidycensus)
library(tidyverse)
library(sf)
library(tigris)
library(ggmap)
library(janitor)
library(gridExtra)

census_api_key("84852611b6fa37f8c4776ab5d1199e812fd3641e")
```

Load zip codes file
```{r}
all_zips <- get_acs(geography = "zip code tabulation area",
                    variables = c(total_pop = "B01003_001"),
                    geometry = TRUE,
                    output = "wide")
```

2008
```{r}
df_overdose_2008 = read_csv("opioid_clean.csv") %>%
  clean_names() %>%
  filter(case_year=="2008") %>%
  mutate(incident_zip = str_sub(incident_zip, 1, 5))

df_overdose_2008 = df_overdose_2008[-c(1:4,6,7)]
```

2008 benzo
```{r}
df_overdose_2008_benzo <- df_overdose_2008 %>%
  group_by(incident_zip)%>%
  count(benzo) %>%
  filter(benzo == 1)

attempt1 <- all_zips %>%
  full_join(df_overdose_2008_benzo, by = c("GEOID" = "incident_zip"))

zips <- read.delim("allegheny_zip.txt", header = FALSE)
attempt1 <- filter(attempt1, GEOID %in% zips$V1)

benzo_2008 = attempt1 %>%
```
```

```
  filter(total_popE >= 400) %>%
  mutate(overdoses_per_capita = n / total_popE * 1000) %>%
  ggplot(aes(fill = overdoses_per_capita)) +
    geom_sf(color = NA) +
    scale_fill_viridis_c(limits = c(0,0.5), oob = scales::squish) +
    theme_bw()+
    ggtitle("2008")+
    theme(axis.title=element_blank(),
        axis.text=element_blank(),
        axis.ticks=element_blank(),
        legend.position = "none")
```

2008 stim
```{r}
df_overdose_2008_stim <- df_overdose_2008 %>%
  group_by(incident_zip)%>%
  count(stim) %>%
  filter(stim == 1)

attempt4 <- all_zips %>%
  full_join(df_overdose_2008_stim, by = c("GEOID" = "incident_zip"))

attempt4 <- filter(attempt4, GEOID %in% zips$V1)

stim_2008 = attempt4%>%
  filter(total_popE >= 400) %>%
  mutate(overdoses_per_capita = n / total_popE * 1000) %>%
  ggplot(aes(fill = overdoses_per_capita)) +
    geom_sf(color = NA) +
    scale_fill_viridis_c(limits = c(0, 1), oob = scales::squish)+
    theme_bw()+
    ggtitle("2008")+
    theme(axis.title=element_blank(),
        axis.text=element_blank(),
        axis.ticks=element_blank(),
        legend.position = 'none')
```

2019
```{r}
df_overdose_2019= read_csv("opioid_clean.csv") %>%
  clean_names() %>%
  filter(case_year=="2019") %>%
  mutate(incident_zip = str_sub(incident_zip, 1, 5))

df_overdose_2019 = df_overdose_2019[-c(1:4,6,7)]
```

2019 benzo
```{r}
df_overdose_2019_benzo <- df_overdose_2019 %>%
  group_by(incident_zip)%>%
  count(benzo) %>%
  filter(benzo == 1)

attempt2 <- all_zips %>%
  full_join(df_overdose_2019_benzo, by = c("GEOID" = "incident_zip"))

attempt2 <- filter(attempt2, GEOID %in% zips$V1)

benzo_2019 = attempt2 %>%
```

```
  filter(total_popE >= 400) %>%
  mutate(overdoses_per_capita = n / total_popE * 1000) %>%
  ggplot(aes(fill = overdoses_per_capita)) +
    geom_sf(color = NA) +
    scale_fill_viridis_c(limits = c(0, 0.5), oob = scales::squish) +
    theme_bw()+
    ggtitle("2019")+
    theme(axis.title=element_blank(),
        axis.text=element_blank(),
        axis.ticks=element_blank(),
        legend.position = "none")
```
2019 stim
```{r}
df_overdose_2019_stim <- df_overdose_2019 %>%
  group_by(incident_zip)%>%
  count(stim) %>%
  filter(stim == 1)

attempt3 <- all_zips %>%
  full_join(df_overdose_2019_stim, by = c("GEOID" = "incident_zip"))

attempt3 <- filter(attempt3, GEOID %in% zips$V1)

stim_2019 = attempt3%>%
  filter(total_popE >= 400) %>%
  mutate(overdoses_per_capita = n / total_popE * 1000) %>%
  ggplot(aes(fill = overdoses_per_capita)) +
    geom_sf(color = NA) +
    scale_fill_viridis_c(limits = c(0, 1), oob = scales::squish) +
    theme_bw()+
    ggtitle("2019")+
    theme(axis.title=element_blank(),
        axis.text=element_blank(),
        axis.ticks=element_blank(),
        legend.position = "none")
```
Combined plots
```{r}
# benzos
pdf("Benzo.pdf", width = 8, height = 12) # Open a new pdf file
grid.arrange(benzo_2008, benzo_2019, ncol=2)
dev.off() # Close the file

# stimulants
pdf("Stim2.pdf", width = 8, height = 12) # Open a new pdf file
grid.arrange(stim_2008, stim_2019, ncol=2)
dev.off() # Close the file
```

Bibliography

Allegheny County, Office of the Medical Examiner. (2020). *Allegheny County Fatal Accidental Overdoses*. Retrieved from the Western Pennsylvania Regional Data Center on 01/12/2020 https://data.wprdc.org/dataset/allegheny-county-fatal-accidental-overdoses

Allen, B., Nolan, M. L., Kunins, H. V., & Paone, D. (2019). Racial Differences in Opioid Overdose Deaths in New York City, 2017. *JAMA Internal Medicine*, *179*(4), 576–578. https://doi.org/10.1001/jamainternmed.2018.7700

Are Benzodiazepines Our Next Prescription Drug Crisis? (2018). ECRI. Retrieved April 19, 2020, from https://www.ecri.org/components/HRCAlerts/Pages/HRCAlerts030718_Benzo.aspx

Armenian, P., Whitman, J. D., Badea, A., Johnson, W., Drake, C., Dhillon, S. S., Rivera, M., Brandehoff, N., & Lynch, K. L. (2019). Notes from the Field: Unintentional Fentanyl Overdoses Among Persons Who Thought They Were Snorting Cocaine — Fresno, California, January 7, 2019. *Morbidity and Mortality Weekly Report*, *68*(31), 687–688. https://doi.org/10.15585/mmwr.mm6831a2

Badger, J., LaRose, E., Mayer, J., Bashiri, F., Page, D., & Peissig, P. (2019). Machine learning for phenotyping opioid overdose events. *Journal of Biomedical Informatics*, *94*, 103185. https://doi.org/10.1016/j.jbi.2019.103185

Balmert, L. C., Buchanich, J. M., Pringle, J. L., Williams, K. E., Burke, D. S., & Marsh, G. M. (2016). Patterns and Trends in Accidental Poisoning Deaths: Pennsylvania's Experience 1979-2014. *PLOS ONE*, *11*(3), e0151655. https://doi.org/10.1371/journal.pone.0151655

Barocas, J. A., Wang, J., Marshall, B. D. L., LaRochelle, M. R., Bettano, A., Bernson, D., Beckwith, C. G., Linas, B. P., & Walley, A. Y. (2019). Sociodemographic factors and social determinants associated with toxicology confirmed polysubstance opioid-related deaths. *Drug and Alcohol Dependence*, *200*, 59–63. https://doi.org/10.1016/j.drugalcdep.2019.03.014

Basu, K., Choudhuri, S., Sen, A., Majumdar, A., & Dey, D. (2018). *Identification of At-Risk Groups for Opioid Addiction Through Web Data Analysis*. 4.

Bennett, A. S., Bell, A., Tomedi, L., Hulsey, E. G., & Kral, A. H. (2011). Characteristics of an Overdose Prevention, Response, and Naloxone Distribution Program in Pittsburgh and Allegheny County, Pennsylvania. *Journal of Urban Health*, *88*(6), 1020–1030. https://doi.org/10.1007/s11524-011-9600-7

Buchanich, J. M., Balmert, L. C., Williams, K. E., & Burke, D. S. (2018). 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. https://doi.org/10.1177/0033354918774330

Burke, D. S. (2016). Forecasting the opioid epidemic. *Science*, *354*(6312), 529–529. https://doi.org/10.1126/science.aal2943

Burke, D. S., & Buchanich, J. M. (2018). Drug Overdoses in Pennsylvania: Measuring, Tracking, and Forecasting the Epidemic. *Commonwealth*, *20*(2–3). https://doi.org/10.15367/com.v20i2-3.186

CDC WONDER. (2020). *US Department of Health and Human Services, Center for Disease Control and Preventions*. Retrieved December 8, 2018, from https://wonder.cdc.gov/

Chen, Q., Larochelle, M. R., Weaver, D. T., Lietz, A. P., Mueller, P. P., Mercaldo, S., Wakeman, S. E., Freedberg, K. A., Raphel, T. J., Knudsen, A. B., Pandharipande, P. V., & Chhatwal, J. (2019). Prevention of Prescription Opioid Misuse and Projected Overdose Deaths in the United States. *JAMA Network Open*, *2*(2), e187621–e187621. https://doi.org/10.1001/jamanetworkopen.2018.7621

Creppage, K. E., Yohannan, J., Williams, K., Buchanich, J. M., Songer, T. J., Wisniewski, S. R., & Fabio, A. (2018). The Rapid Escalation of Fentanyl in Illicit Drug Evidence in Allegheny County, Pennsylvania, 2010-2016. *Public Health Reports*, *133*(2), 142–146. https://doi.org/10.1177/0033354917753119

Curtin, L. R., & Klein, R. J. (1995). *Direct Standardization (Age-Adjusted Death Rates): (584012012-001)* [Data set]. American Psychological Association. https://doi.org/10.1037/e584012012-001

Ellis, R. J., Wang, Z., Genes, N., & Ma'ayan, A. (2019). Predicting opioid dependence from electronic health records with machine learning. *BioData Mining*, *12*(1), 3. https://doi.org/10.1186/s13040-019-0193-0

Gladden, R. M., O'Donnell, J., Mattson, C. L., & Seth, P. (2019). Changes in Opioid-Involved Overdose Deaths by Opioid Type and Presence of Benzodiazepines, Cocaine, and Methamphetamine—25 States, July–December 2017 to January–June 2018. *MMWR. Morbidity and Mortality Weekly Report*, *68*(34), 737–744. https://doi.org/10.15585/mmwr.mm6834a2

Goodnough, A., Katz, J., & Sanger-Katz, M. (2019, July 17). Drug Overdose Deaths Drop in U.S. for First Time Since 1990. *The New York Times*. https://www.nytimes.com/interactive/2019/07/17/upshot/drug-overdose-deaths-fall.html

Guy, G. P., Zhang, K., Bohm, M. K., Losby, J., Lewis, B., Young, R., Murphy, L. B., & Dowell, D. (2017). Vital Signs: Changes in Opioid Prescribing in the United States, 2006–2015. *MMWR. Morbidity and Mortality Weekly Report*, *66*(26), 697–704. https://doi.org/10.15585/mmwr.mm6626a4

Haffajee, R. L., & Frank, R. G. (2018). Making the Opioid Public Health Emergency Effective. *JAMA Psychiatry*, *75*(8), 767–768. https://doi.org/10.1001/jamapsychiatry.2018.0611

Han, B., Compton, W. M., Jones, C. M., & Cai, R. (2015). Nonmedical Prescription Opioid Use and Use Disorders Among Adults Aged 18 Through 64 Years in the United States, 2003-2013. *JAMA*, *314*(14), 1468–1478. https://doi.org/10.1001/jama.2015.11859

Han, B. H., Tuazon, E., Kunins, H. V., Mantha, S., & Paone, D. (2019). Unintentional drug overdose deaths involving cocaine among middle-aged and older adults in New York City. *Drug and Alcohol Dependence*, *198*, 121–125. https://doi.org/10.1016/j.drugalcdep.2019.01.042

Hassan, A. N., & Le Foll, B. (2019). Polydrug use disorders in individuals with opioid use disorder. *Drug and Alcohol Dependence*, *198*, 28–33. https://doi.org/10.1016/j.drugalcdep.2019.01.031

History of the Opioid Epidemic. (n.d.). Retrieved January 30, 2020, from https://poison.org/articles/opioid%20epidemic%20history%20and%20prescribing%20patterns%20182

Hoots, B., Vivolo-Kantor, A., & Seth, P. (n.d.). The rise in non-fatal and fatal overdoses involving stimulants with and without opioids in the United States. *Addiction*, *n/a*(n/a). https://doi.org/10.1111/add.14878

Hulsey, E., Brink, L., Dalton, E., Davis-Jones, L., Jensen, C., Kurta, M., Macerelli, D., Wang, Q., Wilson, A., & Hacker, K. (2016). *Opiate Related Overdose Deaths in- Allegheny County*. Allegheny County Department of Human Services and the Allegheny County Health Department. https://www.alleghenycountyanalytics.us/wp-content/uploads/2017/04/Opiate-Related-Overdose-Deaths-in-Allegheny-County.pdf

Jones, C. M., Logan, J., Gladden, R. M., & Bohm, M. K. (2015). Vital Signs: Demographic and Substance Use Trends Among Heroin Users — United States, 2002–2013. *MMWR. Morbidity and Mortality Weekly Report*, *64*(26), 719–725.

Jones, C. M., & McAninch, J. K. (2015). Emergency Department Visits and Overdose Deaths From Combined Use of Opioids and Benzodiazepines. *American Journal of Preventive Medicine*, *49*(4), 493–501. https://doi.org/10.1016/j.amepre.2015.03.040

Kaiser Family Foundation. (2020). *Kaiser Family Foundation State Health Facts.* https://www.kff.org/other/state-indicator/opioid-overdose-death-rates/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D

Kariisa, M., Scholl, L., Wilson, N., Seth, P., & Hoots, B. (2019). Drug Overdose Deaths Involving Cocaine and Psychostimulants with Abuse Potential—United States, 2003–2017. *MMWR.*

Morbidity and Mortality Weekly Report, *68*(17), 388–395. https://doi.org/10.15585/mmwr.mm6817a3

Knight, K. R. (2017). Women on the Edge: Opioids, Benzodiazepines, and the Social Anxieties Surrounding Women's Reproduction in the U.S. "Opioid Epidemic." *Contemporary Drug Problems*, *44*(4), 301–320. https://doi.org/10.1177/0091450917740359

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics-Theory and Methods*, *26*(6), 1481–1496.

Lanza, S. T., & Rhoades, B. L. (2013). Latent Class Analysis: An Alternative Perspective on Subgroup Analysis in Prevention and Treatment. *Prevention Science : The Official Journal of the Society for Prevention Research*, *14*(2), 157–168. https://doi.org/10.1007/s11121-011-0201-1

LaRue, L., Twillman, R. K., Dawson, E., Whitley, P., Frasco, M. A., Huskey, A., & Guevara, M. G. (2019). Rate of Fentanyl Positivity Among Urine Drug Test Results Positive for Cocaine or Methamphetamine. *JAMA Network Open*, *2*(4), e192851–e192851. https://doi.org/10.1001/jamanetworkopen.2019.2851

Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. *Studies in Social Psychology in World War II Vol. IV: Measurement and Prediction*, 362–412.

Liu, S. J., Mair, C., Songer, T. J., Krans, E. E., Wahed, A., & Talbott, E. (2019). Opioid-related hospitalizations in Pennsylvania: A latent class analysis. *Drug and Alcohol Dependence*, *202*, 185–190. https://doi.org/10.1016/j.drugalcdep.2019.05.009

Lopez, G. (2019, July 17). *New data shows drug overdose deaths fell in 2018. But there's a big catch.* Vox. https://www.vox.com/science-and-health/2019/7/17/20698040/opioid-epidemic-drug-overdose-deaths-2018

Macqueen, J. (n.d.). Some methods for classification and analysis of multivariate observations. *Multivariate observations*, *17*.

Masyn, K. E. (2013). Latent Class Analysis and Finite Mixture Modeling. *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2*. https://doi.org/10.1093/oxfordhb/9780199934898.013.0025

Neill, D. B. (2012). Fast subset scan for spatial pattern detection: Fast Subset Scan. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *74*(2), 337–360. https://doi.org/10.1111/j.1467-9868.2011.01014.x

Neill, D. B. (2017). Multidimensional Tensor Scan for Drug Overdose Surveillance. *Online Journal of Public Health Informatics*, *9*(1). https://doi.org/10.5210/ojphi.v9i1.7598

Neill, D. B., & Herlands, W. (2017). Machine Learning for Drug Overdose Surveillance. *ArXiv:1710.02458 [Cs, Stat]*. http://arxiv.org/abs/1710.02458

Nolan, M. L., Shamasunder, S., Colon-Berezin, C., Kunins, H. V., & Paone, D. (2019). Increased Presence of Fentanyl in Cocaine-Involved Fatal Overdoses: Implications for Prevention. *Journal of Urban Health*, *96*(1), 49–54. https://doi.org/10.1007/s11524-018-00343-z

Opioid Overdose | Drug Overdose | CDC Injury Center. (2019, November 12). https://www.cdc.gov/drugoverdose/index.html

R Core Team. (2017). *R: Foundation for Statistical Computing*. https://www.r-project.org/

Reynolds, D. (2016). *Gaussian mixture model*. http://leap.ee.iisc.ac.in/sriram/teaching/MLSP_16/refs/GMM_Tutorial_Reynolds.pdf

RStudio Team. (2015). *RStudio: Integrated Development Environment for R*. https://rstudio.com/

Scholl, L., Seth, P., Kariisa, M., Wilson, N., & Baldwin, G. (2018). Drug and Opioid-Involved Overdose Deaths—United States, 2013–2017. *Morbidity and Mortality Weekly Report*, *67*(5152), 1419–1427. https://doi.org/10.15585/mmwr.mm675152e1

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, *8*(1), 289.

Shiels, M. S., Freedman, N. D., Thomas, D., & Berrington de Gonzalez, A. (2018). Trends in U.S. Drug Overdose Deaths in Non-Hispanic Black, Hispanic, and Non-Hispanic White Persons, 2000–2015. *Annals of Internal Medicine*, *168*(6), 453. https://doi.org/10.7326/M17-1812

U.S. Census Bureau. (2020). *U.S. Census Bureau; American Community Survey, 2010 American Community Survey 1-Year Estimates*. https://www.census.gov/quickfacts/pittsburghcitypennsylvania

Van Zee, A. (2009). The Promotion and Marketing of OxyContin: Commercial Triumph, Public Health Tragedy. *American Journal of Public Health*, *99*(2), 221–227. https://doi.org/10.2105/AJPH.2007.131714

VanderPlas, J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data* (1st ed.). O'Reilly Media, Inc.

VanHouten, J. P. (2019). Drug Overdose Deaths Among Women Aged 30–64 Years—United States, 1999–2017. *MMWR. Morbidity and Mortality Weekly Report*, *68*. https://doi.org/10.15585/mmwr.mm6801a1

Wagstaff, K., Cardie, C., Rogers, S., & Schroedl, S. (n.d.). *Constrained K-means Clustering with Background Knowledge*. 8.

Wilkerson, R. G., Kim, H. K., Windsor, T. A., & Mareiniss, D. P. (2016). The Opioid Epidemic in the United States. *Emergency Medicine Clinics of North America*, *34*(2), e1–e23. https://doi.org/10.1016/j.emc.2015.11.002