Journal of Human Trafficking

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uhmt20

Why are You *Here*? Modeling Illicit Massage Business Location Characteristics with Machine Learning

Anna White, Seth Guikema & Bridgette Carr

To cite this article: Anna White, Seth Guikema & Bridgette Carr (2024) Why are You *Here*? Modeling Illicit Massage Business Location Characteristics with Machine Learning, Journal of Human Trafficking, 10:1, 20-40, DOI: 10.1080/23322705.2021.1982238

To link to this article: https://doi.org/10.1080/23322705.2021.1982238

JOURNAL OF HUMAN TRAFFICKING
2024, VOL. 10, NO. 1, 20–40
https://doi.org/10.1080/23322705.2021.1982238

Routledge
Taylor & Francis Group

Why are You *Here*? Modeling Illicit Massage Business Location Characteristics with Machine Learning

Anna White [a], Seth Guikema [a], and Bridgette Carr [b]

[a]Industrial & Operations Engineering, University of Michigan, Ann Arbor, Michigan, USA; [b]Law School, University of Michigan, Ann Arbor, Michigan, USA

ABSTRACT

Illicit massage businesses are a venue for sex and labor trafficking in the United States. Though many of their locations are made publicly available through online advertising, little is known about why they choose to locate where they do. In this work, we use inferential modeling to better understand the spatial distribution of illicit massage businesses within the U.S. Based on addresses web-scraped weekly from online advertisements over 6 months, we modeled illicit massage business prevalence at the census tract and county levels. We used publicly available data to characterize census tracts and counties, finding that the state in which they are located, distance to international airports, rent and income levels, racial composition, and religious presence all had significant relationships to illicit massage business presence. Illicit massage businesses operating with unethical labor practices and/or forced sexual services are not in line with just, peaceful, and inclusive societal goals, and the burdens of this industry disproportionately fall on women. While we emphasize that not all illicit massage businesses are settings of human trafficking, better understanding this industry is a key step toward better regulating it and protecting those harmed within it.

KEYWORDS

Illicit massage business; inferential modeling; spatial attributes

Introduction

The Illicit Massage Industry

In the United States and around the world, human trafficking occurs in numerous industry types, including restaurants, agricultural work, health and beauty services, hotels and hospitality, manufacturing, forestry, health care, and in the massage service industry (Polaris, 2018). This study focuses on illicit massage businesses (IMBs), which constitute the second highest number of all sex trafficking reports processed by Polaris, a nonprofit tracking human trafficking in the U.S. since 2007, second only to escort services (Keyhan et al., 2017). Reference definitions for key terms used in this work are available in Supplemental 1, in addition to definitions presented in the body of the text. Though the exact number isn't known, Polaris estimated in 2017 that more than 9,000 IMBs were in operation in the United States, accounting for an annual revenue of $2.5 billion. These businesses can facilitate both sex trafficking and labor trafficking, with workers often being forced to perform sexual services and work unlawful numbers of hours for less than minimum wage or none at all. IMBs vary in scope and operation, with some discretely providing commercial sex services, while others more blatantly advertise their services. Typical characteristics of illicit massage businesses that distinguish them from legitimate massage businesses include offering sexual services, low rates, covered or barred windows, discrete or restricted entrances, security camera surveillance, and employees living in them.

CONTACT Anna White agracew@umich.edu Industrial & Operations Engineering, University of Michigan, 1205 Beale Ave, Ann Arbor, MI 48109, USA

Supplemental data for this article can be accessed on the publisher's website

Most commonly, the individuals recruited to work at IMBs are women, often mothers in their thirties to their fifties who have just arrived from China or South Korea and speak very little English, if any (Keyhan et al., 2017). They are also often constrained by a large debt or other financial circumstances that inhibits their ability to leave. While this is the most commonly recorded description of an individual who may be exploited at an IMB, it does not represent all individuals in these settings.

Most regulatory policy around IMBs is at the local, county, or state level. Because many IMBs operate in networks, this lack of consistency across local boundaries together with relatively low start-up costs for a massage business makes it easy for business owners to relocate if new legislation is passed or law enforcement efforts ramp up in a particular area. IMBs are currently heavily reliant on online advertising to attract customers (Keyhan et al., 2017). Online advertising faced a federal challenge when Congress passed the Allow States and Victims to Fight Online Sex Trafficking Act (FOSTA), in an attempt to curb online sex advertising (2018). In the time since this bill has passed, our team has seen several advertising sites move to non-U.S. domains, including our primary data source, RubMaps.ch (formerly RubMaps.com).

Specific Aims and Hypotheses

Our understanding of the industries created and supported by human trafficking is still quite limited. Much of the field's knowledge about sex trafficking comes from working directly with those who are no longer being trafficked and have decided to share their stories. While these stories are extremely valuable and demand a great deal of bravery, they only provide us a portion of the full picture. Additional qualitative studies and a growing number of quantitative studies have added to our knowledge of trafficking operations and the experience of individuals who are trafficked.

Our study builds on existing studies of IMBs, including those working to differentiate sex trafficking and voluntary sex work in IMB advertising (Dubrawski et al., 2015; Kapoor et al., 2017; Mitchell, 2019; Nagpal et al., 2015; Rabbany et al., 2018; Silva et al., 2014), those that differentiate between illicit and legitimate massage business advertising (Bouche & Crotty, 2018), and those that focus on drivers of demand and geographic location (Bouche & Crotty, 2018; Eiselt & Laporte, 1998; Farrell et al., 2008). Based on these studies, historical context, and what we have learned about IMBs through conversations with law enforcement and subject matter experts, we hypothesize the following:

(1) IMBs may locate in proximity to major transportation infrastructure, including major highways, international airports, major cargo ports, and border crossing points.
(2) Because of the decentralized approach to regulation and enforcement of IMB laws, among other differences along administrative boundaries, there may be variation in IMB prevalence across states and counties.
(3) IMBs may locate in lower rent areas where costs to operate and political opposition to an illicit business may be lower.
(4) IMBs may locate near their target workforce, including areas with higher recent immigrant populations and female workers.
(5) Socio-demographic variables including income, racial composition, educational attainment, and vacancy levels of an area may impact client perception and willingness to patronize a business, so IMBs may locate in areas assumed to be more desirable to their target clients.
(6) IMBs may not locate very near their clients' homes, as patrons may prefer to travel away from their home to obtain illicit services, which are potentially damaging to their reputation. This may mean they locate in less residential areas, with lower population density.
(7) Neighborhood environment, including industry types in the area and vacancy levels may influence customer perception or may provide customers from nearby industry operations.
(8) A connection may exist between organized religious presence and IMB prevalence.

In this paper, we explore these hypotheses using inferential modeling of IMB presence at the census tract and county level. In the following section, we provide more background on these hypotheses and current knowledge about IMBs. This is followed by detailed description of our data sources, including how they were collected and processed. Next, we describe our modeling approach and evaluation metrics and then detail our results including model performance and inference. We conclude by discussing the implications of our findings on our hypotheses and our understanding of IMBs.

Relevant Background

Spatial Significance of IMB Studies

Understanding IMB operation through a spatial lens builds on prior works that have demonstrated the tendency of IMBs, sex trafficking activity, and sexually oriented businesses[1] to cluster and their spatial connections to potential drivers of human trafficking. Many factors could be at play in determining spatial variation in IMB locations, including area characteristics within cities, cultural norms across regions, and differences in state and local policies and enforcement capabilities that create more or less favorable conditions across defined boundaries. A study of IMB clustering in New York City and Los Angeles County found significant IMB clustering within these two geographic regions (Chin et al., 2019), and another in Houston, Texas, found significant spatial clustering of IMBs on the western side of the metro area and significant absence on the eastern side (Crotty & Bouché, 2018). A study of urban sex trafficking and its spatial patterns found significant geographic clustering of sex trafficking offenses[2] within its study area of Austin, Texas (Mletzko et al., 2018). Traditionally, clustering in red-light districts was the norm for sexually oriented businesses, in part due to deliberate zoning policies aimed to keep 'undesirable' businesses contained (Aalbers & Sabat, 2012; Lasker, 2002; Murphy & Venkatesh, 2006). However, changes in policing approaches, growth of car ownership, and growth of the internet have all changed the landscape for sexually oriented businesses, allowing them to move out of traditional city center locations and find new locations in which to operate (Aalbers & Sabat, 2012; Murphy & Venkatesh, 2006). This motivates our goal to better understand where they may locate now that they may not be as exclusively concentrated in city red-light districts.

Area Characteristics Common to IMBs

Proximity to Travel Infrastructure

One study evaluating human trafficking task forces hypothesized factors in three cities that are often believed to increase sex and labor trafficking prevalence: Boston, Massachusetts, because it is a common tourist destination and a major port of entry for individuals immigrating to the U.S.; Phoenix, Arizona, because of its common use as a destination for undocumented immigrants being smuggled using "Coyotes;" and Houston, Texas, because of its proximity to the U.S.-Mexico border and its accessibility to three major interstates and a large port (Farrell et al., 2008). A study of spatial distribution of sex trafficking offenses also found that at the census block group level, proximity to interstate highways, more inexpensive hotels, motels, and sexually oriented businesses, and concentrated disadvantage were all associated with higher sex trafficking offenses (Mletzko et al., 2018). Other studies of sexually oriented businesses found clustering near body shops (car repair), manufacturing, warehousing, big box retailing, and fast-food restaurants, especially in regions where car travel is particularly common, like outside of cities (Aalbers & Sabat, 2012).

[1]The term "sexually oriented business" has a long and complex history of legal definitions and varies by context (Lasker, 2002). Common definitions include: adult bookstores, adult theaters, facilities that feature nude or topless dancing, facilities with sexually oriented live entertainment, massage parlors, adult motels, escort services, nude modeling studios, adult video stores, sex shops, peep shows, and mainstream stores with back rooms including sexually oriented material (Kelly & Cooper, 2000; Owens, 1997).

[2]Sex trafficking offenses in the study comprised of "compelled prostitution" and "trafficking of persons", defined in the Texas Penal Code 9.43A.05 and 5.20A.02, respectively.

Customer Perception

A study of IMB operations in Houston, Texas, used online IMB reviews along with onsite surveillance cameras to detect foot traffic volume of customers. Using a traditional understanding of demand, they studied how factors including location, perceived quality, price and ease of payment, and hours of operation influenced demand for IMB services (Bouché & Crotty, 2018). Their findings indicate that IMB demand is higher in neighborhoods with lower rates of unoccupied housing and renters and higher at IMBs with more RubMaps reviews, lower prices, and acceptance of alternate payment methods beyond cash only. Evidence from interviews with individuals who purchase sexual services (not limited to IMB operations) indicate that fear of public exposure for buying sex was one of the most effective deterrent mechanisms (Durchslag & Goswami, 2008; Farley et al., 2009). Because of this finding, we hypothesize customers may want to travel to IMBs away from their homes for fear of potential recognition by their community members, which may encourage IMBs to locate in less densely populated or less residential areas.

A study of the gendered nature of demand for sexually oriented business found that sexually oriented businesses that typically cater to men tend to be located in neighborhoods with different characteristics than those that typically cater to women, in particular those with a lower score on Edwards's social disorganization index (Edwards, 2010). In this study, Edwards used social disorganization theory, which suggests that crime likelihood is tied more to neighborhood characteristics than individual characteristics. However, this metric assigns neighborhood organization levels based on aggregate factors of individual neighborhood residents including socioeconomic status, racial composition, and residential mobility, while potentially omitting other important contributing factors including policing behavior, access to resources, and historic housing policies that may influence neighborhood characteristics (White et al., 2021). IMBs have traditionally catered to men, so we explore how neighborhood perception through the social disorganization lens may influence IMB location and customer behavior, though we acknowledge the shortcomings and racial biases inherent to the metrics used.

Racial Composition, Other Socio-demographic Factors

Aalbers and Sabat (2012) suggests that the social and political capital of an area's residents influences whether or not sexually oriented businesses are able to operate there. They suggest that middle- and upper-class residential areas may find themselves more able to oppose sexually oriented businesses opening in their area because of their political capital, while working- and lower-class residential areas may be more likely to contain sexually oriented businesses (Aalbers & Sabat, 2012). Willingness by local governance to concentrate "undesirable" businesses in areas with fewer white or high-income residents is explored in depth in the environmental justice literature (Bullard, 1993; Fainstein, 2011).

The study of IMB clustering in New York City and Los Angeles found that Los Angeles' IMBs tended to cluster in regions with more Asian and Hispanic residents, while New York City clusters were found in Manhattan and were negatively associated with household size (Chin et al., 2019). In Houston, IMB neighborhoods were characterized by more Asian residents and fewer Black residents than the regional averages (Crotty & Bouché, 2018). They also found that IMB neighborhoods tended to be home to more college grads, fewer individuals working in manufacturing, education, and health services, and more non-family households than regional averages. Crotty and Bouché (2018) hypothesize that some of these demographic connections may be more attributable to land use than the demographic indicators common to commercial land use areas. Another study found that concentrated disadvantage, a measure including prevalence of poverty, female-headed households, public assistance, unemployment, and educational attainment, was a significant contributor to distribution of sex trafficking offenses (including compelled prostitution and trafficking of persons) in Austin, Texas (Mletzko et al., 2018).

Moral Regions

While IMBs and other sexually oriented businesses are not exclusively found in red-light districts today, there still may be lasting connections to these districts. Some theorize that red-light districts often occur in "moral regions," where there is more acceptance for things that are typically seen as anti-social and amoral (Aalbers & Sabat, 2012). There are historical connections between religious institutions and sexually oriented businesses, a phenomena also illustrated notably in Amsterdam, where the red-light district is located in close proximity to several prominent historical churches (Red light District Amsterdam, n.d.). We hypothesize that there may be a relationship between religious presence and moral regions, where those areas with strong religious presence may find a stronger presence of moral regions, allowing space for those thoughts and actions considered out of place in the church. Even with this history, correlation of religious congregations and IMBs may have one of several explanations, and fully exploring this connection is beyond the scope of this work but is discussed in more detail in the discussion section.

Our Approach to Spatial Patterns

In this study, we analyze the prevalence of IMBs at two geographic scales: census tract and county. We include county-level analysis across the U.S. to understand how regional culture, county-level differences, and rural or urban areas may impact the prevalence of IMBs. We study U.S. counties in three groups: those consistently without any IMB advertisements, those with few or inconsistent IMB advertising, and those with high levels of IMB advertising. We pair this analysis with census tract level modeling to better understand IMB distribution within a local area. This gives us insight into more practical considerations IMB owners may make regarding customer preferences, rent costs, proximity to workforce, and area demographics. We also study census tracts separately within each of the three county groups to understand if local level IMB location decisions differ in high prevalence counties from low- or non-prevalence counties.

Methodology

Data

The data used in this work is all publicly available and was collected from online resources through custom written web-scrapers and standard data downloads. Data processing was done in R (R Core Team, 2018) and ArcGIS (ESRI, 2017).

Illicit Massage Business Prevalence (Response Variable)

First, we set up a web scraper for one of the primary websites advertising IMB locations, RubMaps.ch. One study focused on IMB advertising estimates that only 18% of listings on RubMaps are either legitimate businesses or no longer open (Bouche & Crotty, 2018), which suggests that RubMaps is a commonly used tool for IMB owners rather than for legitimate massage businesses. This website provides a space for advertising illicit massage services in all 50 states and serves as a potential indicator of sex and labor trafficking activity, though not a direct representation. Our scraper collected listings for "erotic massage parlors"[3] from all 50 U.S. states on 24 different days over a period of 6 months, from February to August 2019. On average, each scraped day contained more than 23,000 total listings, and we used ArcGIS geocoding software to convert usable addresses into geolocated points with latitude-longitude coordinates. After removing duplicated, closed, and nonspecific addresses (some were listed by city or street name with an indication to call for the full address), the average web scrape had 11,582 unique and open listings from which we could obtain a valid address. A full summary of data collected from the web scraper is included in Supplemental 2.

We then used ArcGIS Spatial Join to count the number of listings per census tract and county using TIGER/Line Shapefiles census boundaries (2016a, 2018). The resulting dataset was heavily zero inflated, with nearly 88% of census tracts containing no open RubMaps listings across any of our

[3]"Erotic massage parlor" is the term used by the website from which we collected the data (RubMaps). For the purposes of this study, it is interchangeable with the term, "illicit massage business" or "illicit massage parlor."

scraped days. The number of listings per census tract was relatively stable across the 6 month study period, with low standard deviation values. Supplemental 3 provides a more detailed look at the distribution of this resulting dataset.

With these census tract and county-level estimates, we created different response variable categorizations, converting raw count data to binary values at the census tract level and to a three-level category at the county level. Splitting the census tract level into binary categories yielded 64,005 census tracts with zero advertisements over the 6 month scrape period, and 9,051 tracts with at least one during that time. At the county level, the first group (low prevalence) contains 2,300 counties with zero advertisements. The second group (medium prevalence) contains the first three quartiles of the nonzero aggregated county averages, including 632 counties with more than zero and up to 8.667 advertisements on average. The third group (high prevalence) contained all counties in the upper quartile of the nonzero county data and holds 210 counties with more than 8.667 IMB advertisements averaged over the 6 months. Figure 1 shows the geographic distribution of the different county-level values.

Covariate Data

To determine relevant characteristics of each census tract and county, we used publicly available data. We collected demographic data, including racial composition, educational attainment, unemployment rate, industry by occupation for employed civilians, population density, total population, and income from the 2016 American Community Surveys five-year estimates, made available through Social Explorer (U.S. Census Bureau). From this source we also obtained information about the physical environment of each census tract, including building vacancy rates, age of buildings, and price of rent. We also used information about U.S. transportation infrastructure from national datasets to calculate distances to the nearest primary road (U.S. Census Bureau, 2016b); international airport (U.S. Geological Survey, 2018); major cargo port (Burnson, 2019; U.S. Department of Transportation, 2018); land border port of entry (Liu, 2017); and military installation (U.S. Census Bureau, 2015). Finally, we used information about the number of religious congregations in the area, represented by congregations per capita and made available through Social Explorer (RCMS, 2010). A summary of our covariates is displayed in Table 1, and a full description of our dataset and its naming

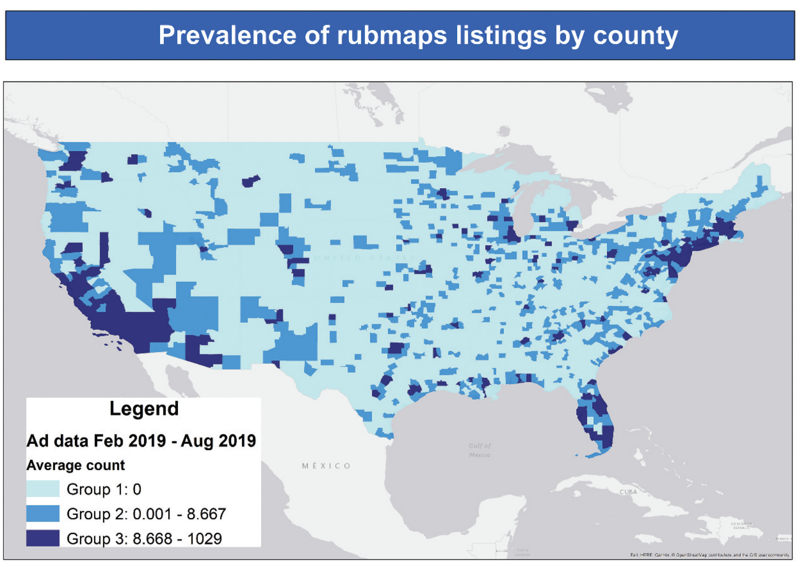


Figure 1. IMB advertising prevalence by county, measured by unique and geolocatable addresses listed on RubMaps. This map was created using ArcGIS software with a basemap provided by ESRI (Esri, 2019).

Table 1. Summary of factors considered, census tract and county models (ESRI, 2019).

Census tract characteristics		County characteristics	
Category	Source	Category	Source
Response variable: binary existence of IMB in tract	RubMaps.ch (2019)	Response variable: County group (low, medium, high)	RubMaps.ch (2019)
Population density	American Community Surveys 2016 (5 year estimates)	Population density	American Community Surveys 2016 (5 year estimates)
Census tract population		County population	
County population		Racial composition	
Racial composition		Educational attainment	
Educational attainment		Unemployment	
Unemployment		Female to male labor participation ratio	
Industry by occupation		Median household income	
Median household income		Poverty rate	
Housing vacancy		Immigration population since 2010	
Median building age		Housing vacancy	
Median rent		Median building age	
Congregations per capita	Religious Congregations & Membership Study 2010	Median rent	
State	American Community Surveys 2016 (5 year estimates)	Congregations per capita	Religious Congregations & Membership Study 2010
Distance to highway	TIGER/Line Shapefile 2016, Primary Roads	Distance to international airport	USGS National Transportation Dataset (2018)
Distance to military installation	TIGER/Line Shapefile (2015), U.S., Military Installation National Shapefile	Distance to border crossing	Homeland Infrastructure Foundation-Level Data (Liu, 2017)
Distance to border crossing	Homeland Infrastructure Foundation-Level Data (Liu, 2017)	Distance to major cargo port	Department of Transportation (2018), Major Ports, Top 30 U.S Ports 2019
Distance to major cargo port	Department of Transportation (2018), Major Ports, Top 30 U.S Ports 2019		
Distance to international airport	USGS National Transportation Dataset (2018)		

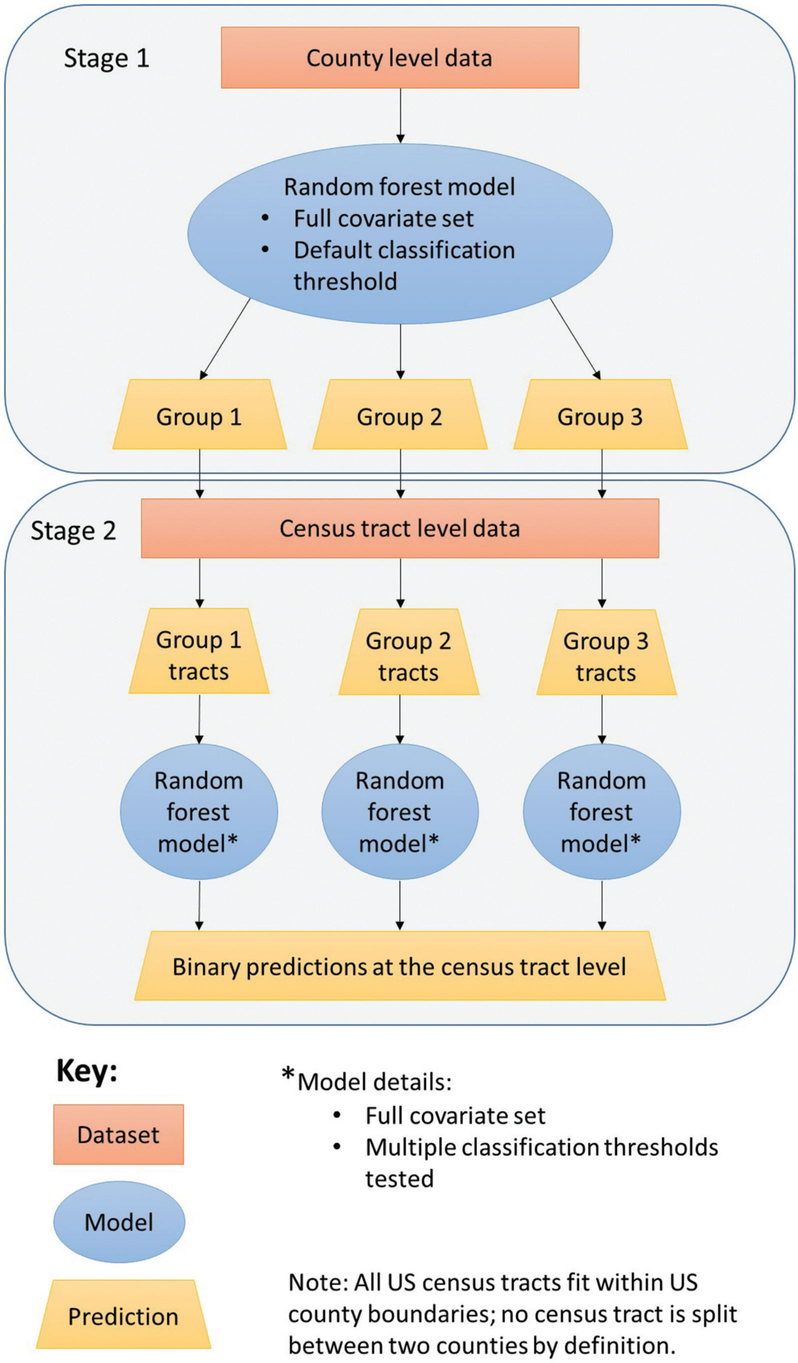


Figure 2. Conceptual overview of 2 stage model process.

configuration is included in Supplemental 4. We used data from all census tracts containing land (e.g., not those exclusively on water) in all U.S. states except Alaska. We excluded Alaska because it was missing data across multiple covariates.

Model Configuration

We tested a variety of predictive models and response variables before determining that binary classification at the census tract level, three-level classification at the county level, and Random Forest models were the best performing configurations for our data. Random Forest models consistently performed best in terms of our performance metrics (detailed in the Model Evaluation section, below) across repeated 20% random holdouts for the full U.S. census tracts models. A more detailed description of the model types and configurations tested is available in Supplemental 5.

We used the R package randomForest to develop our Random Forest models (Law & Wiener, 2002). Random Forests are an ensemble method that consider the outputs of multiple classification trees to make predictions. Each tree is grown on a bootstrapped sample of the full training set. The final prediction of the Random Forest is the result of an aggregation of the vote from each tree (Breiman, 2001). By default, a Random Forest will use the majority (or plurality, if there are more than two categories) vote to make a classification prediction, but we tested different vote thresholds in addition to the default to find the best performing threshold.

We also tested a two-stage model (e.g., Kabir et al., 2019), incorporating county-level data in addition to tract level data. The benefits of a two-stage model, demonstrated in Guikema and Quiring (2012), Kabir et al. (2019), and McRoberts et al. (2018), are that it can better handle zero-inflated data like our IMB prevalence data. Figure 2 gives a conceptual overview of the two-stage modeling process further detailed in this paragraph. In Stage 1, we modeled county data with a three level response variable, separated into its three groups as detailed earlier. We used a Random Forest to classify each county into one of the three groups, using the default plurality classification threshold (e.g., a county is classified into the group with the most votes from the ensemble of trees). In Stage 2, we used census tract level data. Using the predictions made at the first stage, we split the census tract data into three groups, corresponding to their county's group prediction from the Stage 1 model, not their county's actual group value. The existence of inevitable model error indicates that some counties will be classified as group one, though their actual group based on measured IMB prevalence may be different, and we can analyze census tract-level IMB ad locations in counties our model considers very low likelihood for IMBs. We interpret the three county groups as (1) very low likelihood counties, (2) medium likelihood counties, and (3) high likelihood counties, where likelihood refers to the likelihood that a county contains an illicit massage business, measured by ad prevalence. For each group of counties, we trained a separate Random Forest model to make binary (0 or 1) predictions at the census tract level, where 1 = IMB ad within the census tract, and 0 = no IMB ad within the census tract. Because our data is so zero inflated, there are still a majority of zero values at the census tract levels within all three county groups, but this second stage allows us to isolate census tracts by the characteristics of the counties that contain them. These grouped census tract models use the same covariates as the full U.S. census tract model, but they are each only trained and predicting on one group's subset of the data. This modeling process allowed us to infer significant factors at the county level, and significant factors at the tract level within each of the three county groups. This allows us to determine if county-level patterns exist, or if patterns are relatively stable across the U.S.

Model Evaluation

Each census tract model was evaluated on five performance metrics: accuracy, sensitivity, specificity, Brier score, and Area under the Receiver Operating Characteristics (ROC) Curve. Accuracy indicates the ratio of correct predictions to total predictions. For a zero-inflated dataset, the accuracy value can often be misleading as predicting zero in every case will still lead to an accuracy rate equal to the proportion of the majority class, in this case, nearly 90%. To supplement accuracy, we used sensitivity, or the ratio of the number correctly predicted to be positive (1) to the number that was actually positive. This is also known as the true positive rate. We also used specificity, or the ratio of number correctly predicted negative (0) to the number actually negative, also called the true negative rate. We also used Brier score, which is conceptually the equivalent of mean squared error but used for classification models.

Finally, we calculated the resulting Area Under the ROC Curve (AUC). ROC curves allow us to compare our model performance to that of a random assignment model and a perfectly performing model. A perfect model would have an AUC of 1, while a random model would have an AUC of 0.5. A general rule for interpreting AUC is that 0.7 to 0.8 is considered acceptable; 0.8 to 0.9 is considered excellent; and 0.9 to 1.0 is considered outstanding (Hosmer & Lemeshow, 2000; Mandrekar, 2010).

Accuracy, sensitivity, specificity, and AUC are all dependent on the classification decision made by the model (binary, or into one of three classes) rather than the predicted probability value (aggregated votes made by each of the ensemble trees). For these four metrics, we tested different classification thresholds for binary classification. In the full U.S. census tracts model, we tested five different thresholds: 0.2, 0.3, 0.5, 0.8, and 0.9. For the 2nd stage models, we tested each tenth value from 0.1 to 0.9. For each classification threshold, a given census tract was predicted as 1 (contains IMB ad) if the percentage of classification trees that predicted 1 exceeded the classification threshold being tested. In the default binary models, this is done using a threshold of 0.5, or a majority vote of all classification trees.

For the county models, we used a simpler model evaluation tool, the confusion matrix. Because the county model was predicting a three-level response variable, rather than a two-level (binary) variable, evaluating the five performance metrics mentioned earlier would not be as interpretable and useful to this study. We used a confusion matrix, which displays the number of actual and predicted observations for each of the three levels and can help determine which groups are more accurately classified.

Inferential Tools

While we used performance metrics to verify that our models were performing well enough to be used, we use inferential tools to answer our core research questions about which factors have the strongest relationships to IMB location patterns. Random Forest models automatically generate variable importance rankings, and we use this to rank the degree of influence of each factor in the models. The variable importance rankings were generated using an average measure of the decrease in Gini impurity when a variable is chosen to split a node, known as MeanDecreaseGini. Because Random Forests are built from ensembles of classification trees, the Gini impurity values are calculated across trees for each variable. The Gini impurity index represents how well a node splits the classes into two "pure" classes. This is the most commonly used variable importance measure in classification models and gives us insight on which of the covariate influence patterns impact the model predictions most strongly.

To understand the relationship between each of our covariates and our response variable, we use partial dependence plots. Conceptually, partial dependence plots show the marginal influence of a single covariate on the response variable by averaging out the effects of all other covariates. Partial dependence plots can demonstrate the nature of the relationship between a covariate and the response variable, whether that be linear, constant, a step function, or another type. They may also illustrate a boundary within which a covariate does or does not impact the response value. To generate partial dependence plots, we used partialPlot from the R package randomForest. The plots also include nine vertical dashes showing the deciles of the data. To describe the mathematical process behind partial dependence plot generation, we will use an example variable of interest: income. To understand the influence of income in a given model, the range of income present in the dataset is used as the x-axis of the partial dependence plot. For each value of income, the model response is averaged over all observations of the other covariates. These averages are used to demonstrate the partial relationship between a single variable (income) and response (IMB ad presence).

Table 2. Summary of performance metrics for top performing census tract models.

Top 4 model configurations for full U.S. census tracts, Random Forest models[1]:	Accuracy	Sensitivity	Specificity	Brier	AUC
Everything but the PCA variables[2], actual distance values[3], excluding distance to regional airports[4]	0.7637	0.6265	0.7835	0.0948	0.7050
PCA race[2], actual distance values[3], excluding distance to regional airports[4]	0.7572	0.6481	0.7730	0.0950	0.7106
PCA occupation[2], actual distance values[3], excluding distance to regional airports[4]	0.7570	0.6436	0.7734	0.0950	0.7085
Full covariate set with actual value distances[3], excluding distance to regional airports[4]	0.7570	0.6466	0.7729	0.0950	0.7098

[1]All values in this table are found using the lowest classification threshold tested, 0.2.
[2]We tested some covariate configurations using PCA, or Principal Components Analysis. This is a method of covariate reduction done using the built-in R function called prcomp and further detailed in Supplemental 8.
[3]We tested binning the distances to infrastructure into levels, rather than using the actual distance values, but found that the models including actual values performed better than those using bins.
[4]We noticed that the USGS dataset classification system included many airports in the Western U.S. region that may be used on a very limited basis and wouldn't fall into traditional conceptual definitions of a regional airport. We hypothesized that this may introduce noise into the data, so we tested excluding it and found those models performed better.

Table 3. Confusion matrix for county Random Forest model.

		Predicted – Group			
		1	2	3	Error rate
Actual – Group	1	2176	108	0	0.0472
	2	215	391	26	0.3813
	3	0	57	153	0.2714

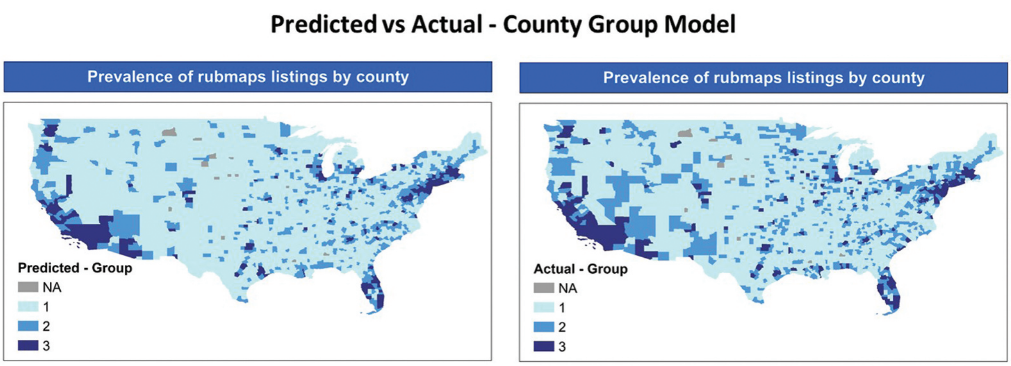


Figure 3. Predicted vs actual for county-level IMB data. Group 1 counties contain 0 IMB ads; Group 2 counties contain on average > 0 and ≤ 8.667 distinct IMB ads; Group 3 counties contain on average > 8.667 distinct IMB ads. NA refers to counties with missing data that were excluded from the model.

Results

Model Performance

We tested different model types across U.S. census tract data, but proceeded with a Random Forest model because of its better performance across accuracy, sensitivity, and specificity. See Supplemental 7 for a full breakdown of this comparison from our initial holdout tests. Proceeding with Random Forest models on our full U.S. census tract dataset, we compared different classification thresholds and subsets of covariates. From this large set, we found four model configurations to be performing best across accuracy, sensitivity, specificity, ROC, and Brier score. Performance metrics for these four models are included in Table 2, and the full results are available in Supplemental 8. We proceeded with the full covariate set model because its performance was comparable to the other three models with much better interpretability. Using a low classification threshold gives us reasonable performance metrics that suggest this model is capable of detecting meaningful patterns in this data. Comparative results between different classification thresholds are available in Supplemental 8.

For the county-level model, we used the full covariate set with default Random Forest configurations and found that it was performing reasonably well without modifications. The confusion matrix for the county-level model is shown in Table 3, where each entry shows the number of occurrences in that category. Of particular note from our results, no Group 1 records are falsely classified into Group 3, and vice versa. We also plotted a map of county-level predictions alongside actual data to display the model performance at a scale that is visually interpretable (Figure 3).

Finally, we compared performance metrics for the grouped census tracts models (Stage 2 of the two-stage model) and found that lower value thresholds were also performing better in the subset of tracts, as with the full U.S. set of census tract models. Detailed performance metrics for the Stage 2 models are provided in Supplemental 9.

Inference

To begin our inference using the models, we use variable importance. The most commonly occurring covariates in the top five for each of the five models tested were: State ID, population density, total population, percentage of the population that identifies as Asian, and percentage of the population that identifies as Black. These occurrences are summarized in Figure 4, and a full variable importance plot for all five models is shown in Figure 5.

The following subsections discuss relationships between predicted outcomes and covariates selected for their importance rankings or for their potential relationships identified in the literature and through conversations with subject experts.

State Effects

The partial dependence plots for state ID, shown in Figure 6, give the relative influence of each state on the likelihood of a county or census tract in that state containing an IMBs if all other factors were held the same. Across all four census tract models, New York, Ohio, Tennessee, North Carolina, Michigan, Texas, and Alabama are the most consistently occurring states. This suggests that when holding all other covariates that we modeled held constant, we would expect to see the most widespread IMB presence in these states. This may be due to policy implementation or enforcement, or some other

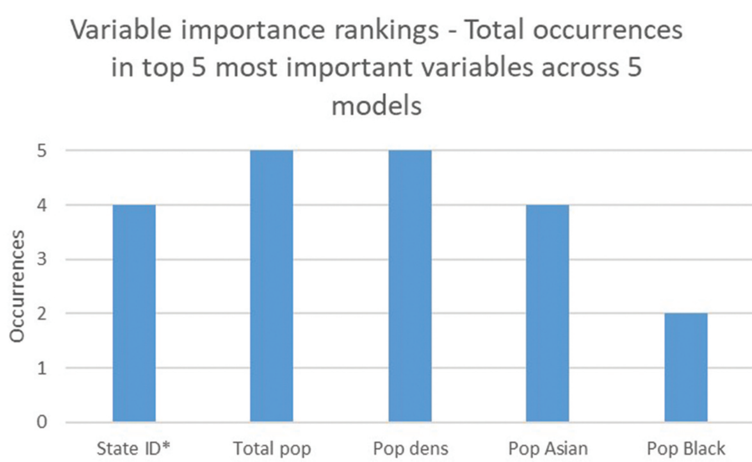


Figure 4. Most commonly occurring covariates in the top 5 most important variables to each of the 5 models tested, measured by Gini impurity index. *State ID was not a covariate in the county model, so 4 is the maximum number possible for this covariate.

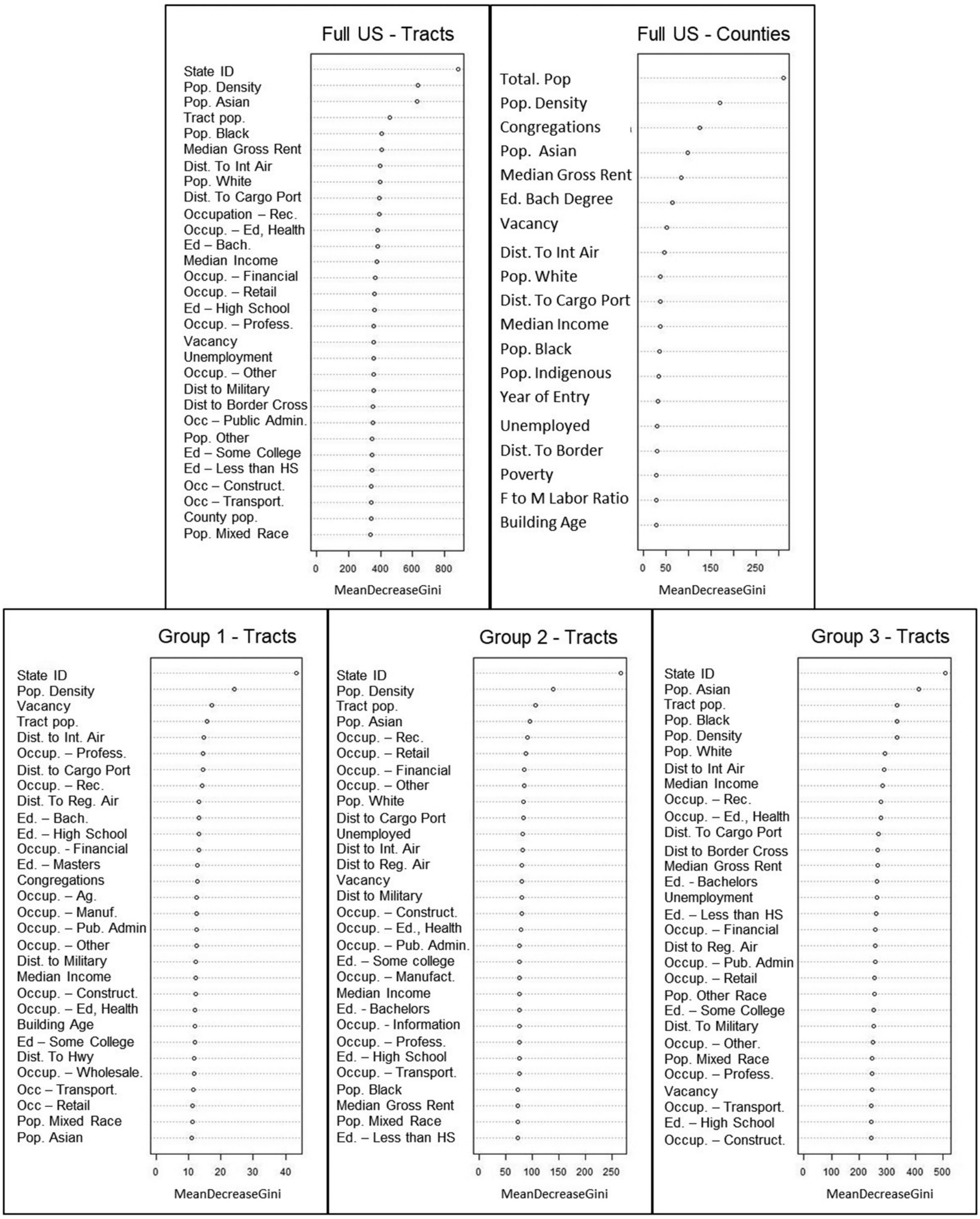


Figure 5. Top ranked covariates by variable importance for each model.

unstudied factors at play. Further analysis of the policies of these states would give us interesting insight into the influences being captured by these models. Figure 6 displays partial dependence plots including the 10 most likely states for each of the four census tract models.

Population Density and Total Population

Our hypotheses prior to beginning this work were that IMBs may locate most prominently in cities, where county population and population density are highest. At a census tract level, however, population density helps us better understand how variations in population density within a city or county

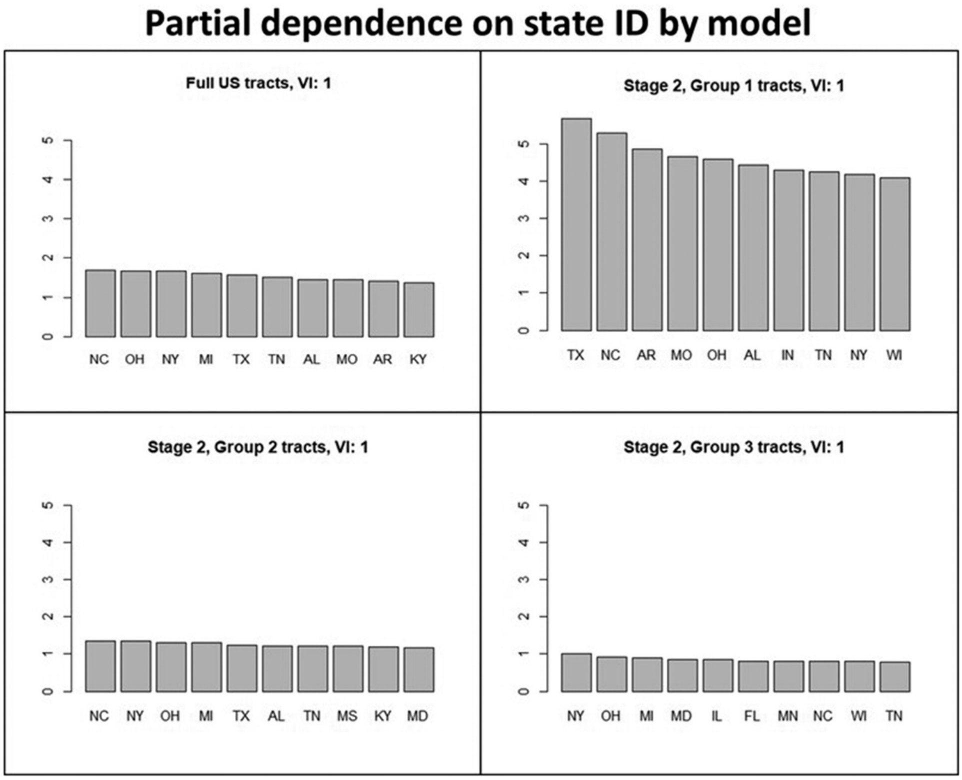


Figure 6. Partial dependence plots for state ID in the 4 census tract level models. Standard state abbreviations used.

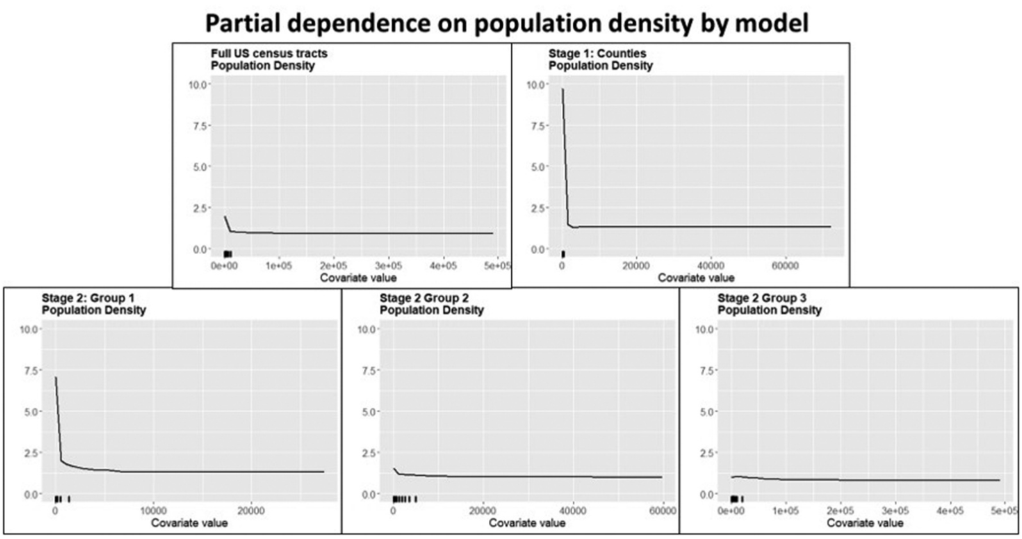


Figure 7. Marginal influence of population density by model.

relate to IMB presence, and partial dependence plots help us understand the relationships between population values and predicted outcomes with all else held constant. Figure 7 shows partial dependence plots for population density in all five models. From this figure, we can see that most of our data falls at the lowest end of the x-axis, where short vertical dashes split the data evenly into tenths. While county

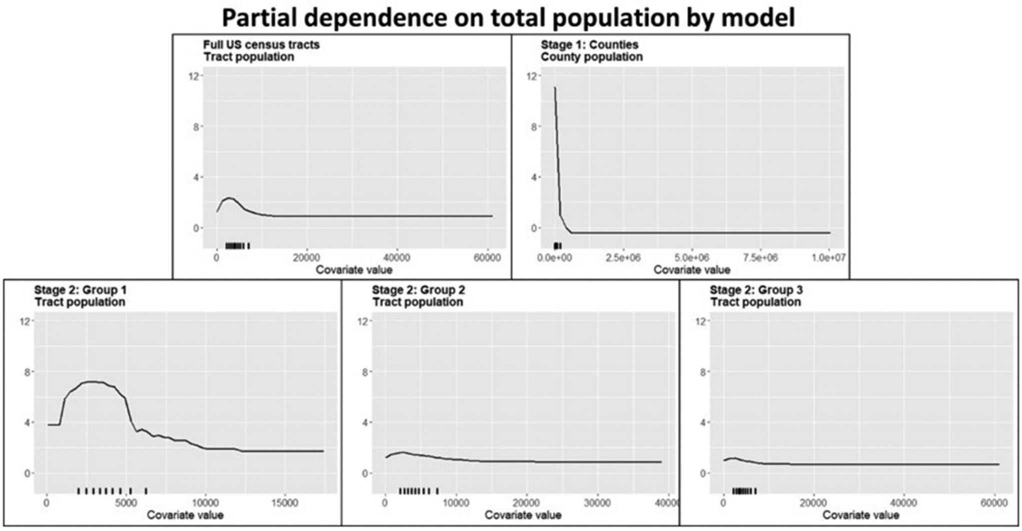


Figure 8. Marginal influence of total population by model.

groups two and three (medium and high prevalence) indicate a weak relationship between population density and IMB presence, our other three models (all census tracts, counties, and group 1 counties) indicate higher likelihood for IMBs in lower density tracts and counties, when all other factors are removed.

Total population demonstrates a consistent shape across all four tract models, with a peak likelihood in the mid-range population census tracts. Figure 8 shows the consistently shaped partial dependence plots for each of the four census tract models. This trend only differs in the county model, where the data and likelihood peak is concentrated at the low end. These findings complement those from the population density plots, suggesting that IMBs are more likely to be in census tracts with low- to mid-level populations and relatively low population densities.

Racial Composition

When analyzing demographics, three racial groups appeared in the top 10 most important variables for any of our 5 models: percentage of Black, Asian, and white residents. The proportion of population that is Asian shows up prominently in the full U.S. tracts model, the county model, and 2 of 3 grouped tract models. The trend displayed by this covariate is consistent across model types, with lower percentages of Asian populations associated with higher likelihood for IMB presence in counties and census tracts. The trend for proportion of population that is Black differs slightly across models, though all models suggest a relatively stable likelihood across all values of Black population. 4 of 5 models suggest a decline in likelihood with higher percentages of Black residents, with Group 1 (lowest likelihood counties) showing the largest decline in likelihood. Conversely, the proportion of white residents is positively associated with increased IMB presence consistently across the 5 models, except at the very high end, where Group 1 tracts and full U.S. tracts with nearly 100% white residents see a drop in IMB likelihood. Census tracts in county groups two and three have the smallest magnitude of influence by racial composition across the three racial categories studied here, indicating racial composition may not be as strong of an influence in these mid- and high-prevalence counties. Figure 9 shows trends for the three demographic groups across all five models.

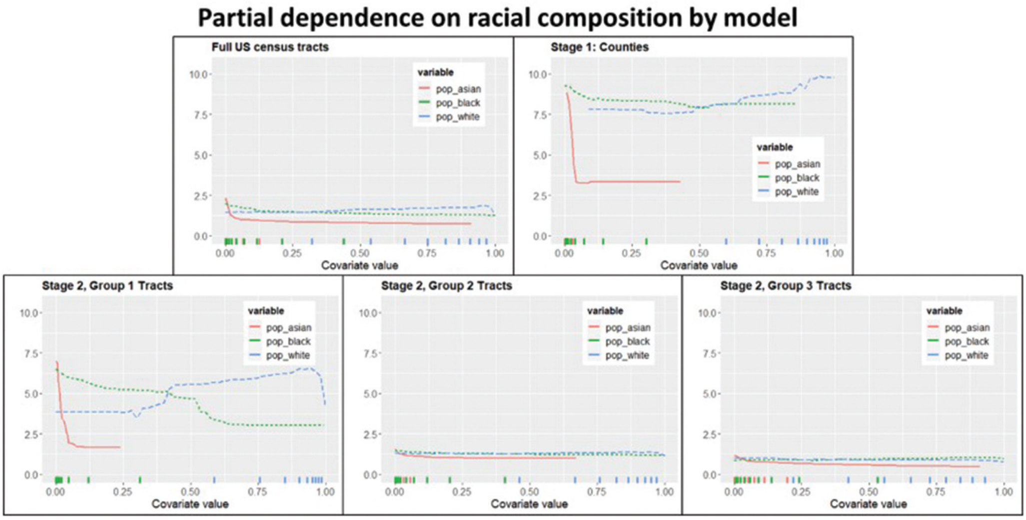
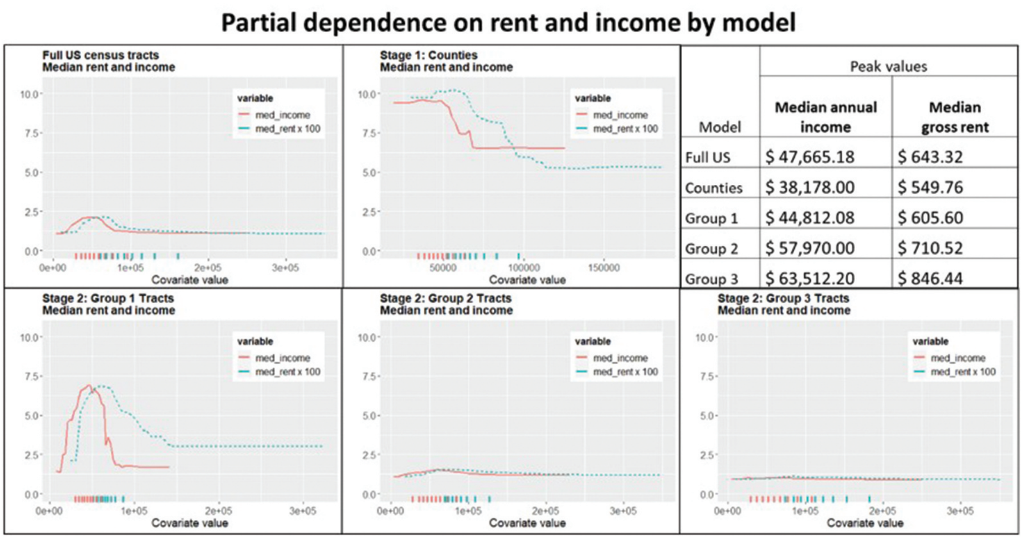


Figure 9. Partial dependence on demographic composition, percentage of the population that identifies as Asian, Black, and white.



Model	Peak values: Median annual income	Peak values: Median gross rent
Full US	$ 47,665.18	$ 643.32
Counties	$ 38,178.00	$ 549.76
Group 1	$ 44,812.08	$ 605.60
Group 2	$ 57,970.00	$ 710.52
Group 3	$ 63,512.20	$ 846.44

Figure 10. Partial dependence on median gross rent (multiplied by 100 for scale in the plot) and median income in all 5 models. Covariate value at the peak of each curve is included in the table in the top right.

Income and Rent Levels

We investigated two covariates that are intuitively of interest, though they appeared lower in the variable importance rankings in the second-stage models. However, consistent patterns shown in Figure 10 between median gross rent and median income for each model justifies their discussion. Both indicate there may a "desirable" level of wealth for IMB locations. At the county level, the pattern seems to suggest that counties with lower income and rent prices tend to have higher likelihoods for IMB presence. For reference, the median household income across the U.S. was $64,324 in 2018 (Friedrich, 2019), and all of the peak IMB likelihood values were below this value. The magnitude of this peak was much higher in Group 1 census tracts, which may indicate a stronger relationship to rent and income levels in low likelihood counties than elsewhere. In the county model, IMB likelihood was highest in the lower portion of data and decreased as rent and income levels increased.

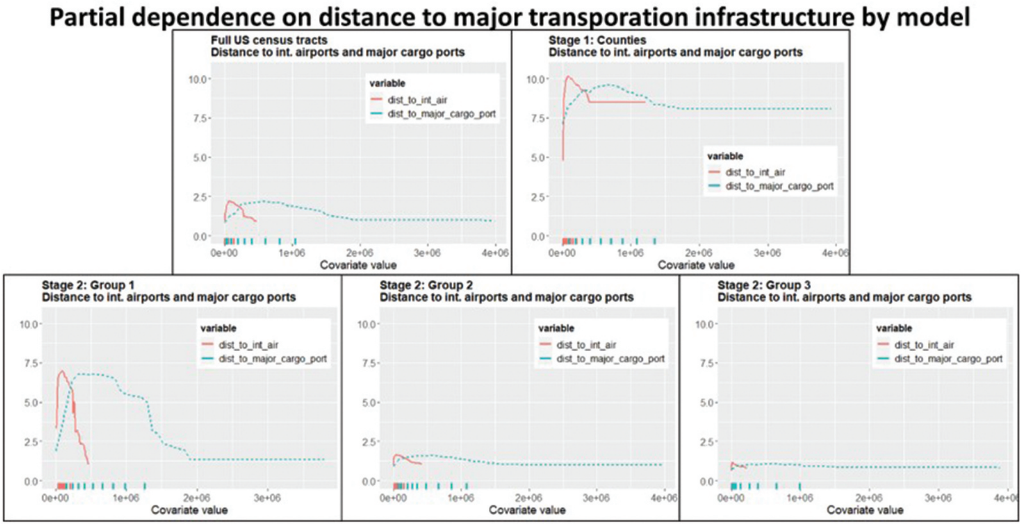


Figure 11. Partial dependence on distance to major transportation infrastructure: international airports and major cargo ports. X axis units are in meters, where 1e+06 represents 1,000 km, or about 620 miles.

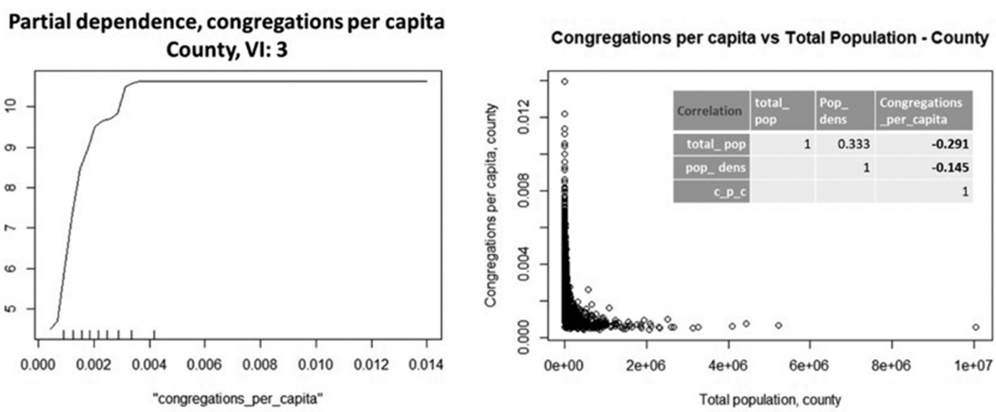


Figure 12. Partial dependence on congregations per capita at the county level. We compared congregations per capita to population numbers and found there was not a strong correlation between either pairs, indicating that congregations are not a proxy for total population or population density.

Infrastructure and Nearby Institutions

Distance to international airports and major cargo ports appeared regularly in the top 10 most important variables to our five models. When evaluating the relationships of distance to international airports and major cargo ports to IMB likelihood, we observe a peak likelihood value for each of the two, shown in Figure 11. This may suggest an ideal distance region for business locations. For distance to international airports, this value is intuitively at a much smaller scale than for distance to major ports because of the prevalence of international airports throughout the continental U.S.

In a potentially surprising note, we also observed a strong positive relationship between congregations per capita at the county level and IMB likelihood. As illustrated in Figure 12, congregations per capita ranked 3rd in variable importance in the county model. We also included in Figure 12 a comparison of total population and congregations per capita and found there was not a strong correlation between the two, indicating congregations per capita is not acting as a proxy for population levels. This finding suggests that higher religious presence is associated with higher IMB likelihood, though this may be through mutual correlation with some other set of variables not modeled in this study. This is a topic that could be investigated further in the future.

This is a topic that could be investigated further in the future.

Discussion

The purpose of this study is to add to our limited understanding of illicit massage businesses and why they locate where they do. Better understanding locations and the types of areas that are more conducive to IMBs helps us understand where businesses may choose to relocate after a raid or if a location is shut down. This can also help us to better understand demand patterns, including where typical buyers are willing to travel for services. Perhaps most importantly, we hope to add to the understanding of IMBs so that decision-makers can make more effective policy decisions that reduce harm. When implementing new policies or enforcement strategies in this domain, it is important that these actions and decisions center on trafficking victims as well as considering potential harms to individuals working in the industry voluntarily or for survival. Trafficking, exploitation, and voluntary sex work have murky boundaries, and treating them as mutually exclusive may lead to bad outcomes for the individuals we are seeking to help. This means considering unintended consequences and cascading effects of well-intentioned policy and decisions. It also requires an understanding that some individuals who are exploited in IMBs still consider this their best option for livelihood and well-being, so law enforcement strategies that remove these individuals from their workplace without providing them a better alternative and support system may find their efforts failed.

This study has a few limitations. First, it only captures illicit massage business advertising on RubMaps. We know that some businesses choose not to advertise on this site or may have moved to alternative sites after the passage of FOSTA. Additionally, as some businesses did not list specific addresses, we are not able to study a portion of the listings on RubMaps. Because we excluded those businesses that may operate with more secrecy by providing unspecific location information, we may be excluding some of the businesses that are more likely to be engaging in illegal behavior, including sex or labor trafficking. However, we still consider our data to be a reasonable approximation of IMB activity based on our conversations with subject matter experts and the share of data represented. Importantly, this is not a direct study of human trafficking but rather a proxy study of data that very likely contains a portion of businesses engaging in exploitative labor practices, labor trafficking, and sex trafficking.

We also analyzed the spatial autocorrelation of the model residuals for the full U.S. census tracts model and the counties model using Moran's I. An in-depth description of Moran's I, including how it is calculated and interpreted is found in Abokifa and Sela (2019). We found statistically significant spatial autocorrelation in the model residuals, with Moran's I values of 0.184 and 0.063 and Z-values of 6.348 and 7.924 for tracts and counties, respectively. These values indicate some tendency for like-values in our model error to cluster, with slightly positive Moran's I values indicating the positive direction of clustering patterns and statistically significant Z-values indicating the existence of some nonrandomness. Interesting future work could further investigate the unexplained variance in the spatial patterns of our dataset by continuing to expand the variables considered and changing the scale of analysis to better capture this residual spatial correlation.

Our findings in this analysis were partially aligned with our initial hypotheses. We found that, as expected, proximity to international airports and major cargo ports was related to higher IMB likelihood, though other major transportation infrastructure including major highways and border crossing points did not demonstrate a significant relationship. State differences, possibly policy-related, seem to be a factor in IMB presence, as predicted, though further studies are needed to understand that relationship and its complexities. As we hypothesized, lower rent and population areas are associated with higher IMB likelihood, though relationships with target workforce, including immigrant populations and female workers did not show clear relationships.

Racial composition and income were the only significant socio-demographic factors that we observed, with educational attainment, vacancy, and industry composition in an area not demonstrating significant relationships. Interestingly, our U.S. aggregate study indicated a diversion from prior city-level analyses in Houston and Los Angeles, which found higher percentages of Asian residents correlated to IMB presence, while our study found the opposite (Chin et al., 2019; Crotty & Bouché, 2018). This finding also seems to contradict some of the common understanding around IMBs, as they are sometimes referred to as “Asian massage parlors” in news and media. Although Polaris reports that the most common employee profile of IMB workers are Asian women, this may not be an indication of the racial composition of the residents of the areas in which IMBs are located. The difference in our findings from the city-level findings may also be an indication of the impact of scale on our analysis, as our dataset covered the entire U.S., including both urban and rural regions. An interesting future analysis may compare IMB presence by region or between rural and urban areas to understand how these relationships may vary.

Finally, religious presence had a strongly positive relationship to IMB presence at the county level, something our research team hypothesized but was still surprised to see demonstrated so clearly. This finding is new to this field of work, as far as we could tell, and certainly represents a potential for further investigation. There could be one of several explanations to this relationship. One explanation could be an overlap in the population of those regularly attending church and IMB customers, a potentially controversial explanation that would need a targeted study to substantiate. Conversely, it may be that county characteristics that are conducive to IMB locations are also conducive to religious presence, and there may be no direct connection between IMB presence and religious presence, but perhaps there are some common attractors for both. A more targeted and in-depth study of this connection is needed to draw confident conclusions on this relationship.

Conclusion

Through this study, we found several potential contributing factors to IMB location, including state, population density, total population, racial composition, income and rent levels, and distances to international airports and major cargo ports. We hypothesize that policy or law enforcement in states including New York, Ohio, Tennessee, North Carolina, Michigan, Texas, and Alabama may create a more friendly environment for IMBs. In most cases, higher proportions of Black and Asian residents were associated with lower IMB likelihood, while higher proportions of white residents were associated with higher IMB likelihood. We found that the low to medium value ranges of population, income, rent, distance to international airports, and distance to cargo ports were where the likelihood of having an IMB present is highest. We also found higher rates of congregations per capita were positively associated with IMB likelihood at the county level.

We hope that future work can continue to add to our understanding of illicit massage businesses. Further exploring connections with policy and law enforcement practices may add to that understanding, as well as state-by-state analysis of IMB location trends. Ultimately, we hope that better understanding of these businesses will lead to better legislation and action to stop the exploitation of people for labor or sexual acts. We also hope that better understanding this industry will inspire policy to allow those who engage in sex work, whether for survival, through force, fraud, coercion, or voluntarily, to connect with the resources needed to remove themselves from exploitative environments and to provide for themselves without fear of violence, retribution, or arrest.

Acknowledgments

This work was partially funded by University of Michigan through Rackham Merit Fellows program. We thank Dr. Tom Logan (University of Canterbury) for his work on the web scraper development that greatly aided this work. This support is gratefully acknowledged. The findings and opinions are those of the authors and do not necessarily represent those of the funding agencies.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Anna White http://orcid.org/0000-0002-9576-5043
Seth Guikema http://orcid.org/0000-0001-6024-0303

References

Allow States and Victims to Fight Online Sex Trafficking Act of 2017, Pub. L. No 115-164, 132 Stat 1253. (2018). https://www.congress.gov/bill/115th-congress/house-bill/1865/text

Aalbers, M. B., & Sabat, M. (2012). Re-making a landscape of prostitution: The Amsterdam Red Light District: Introduction. *City*, *16*(1–2), 112–128. https://doi.org/10.1080/13604813.2012.662372

Abokifa, A. A., & Sela, L. (2019). Identification of spatial patterns in water distribution pipe failure data using spatial autocorrelation analysis. *Journal of Water Resources Planning and Management*, *145*(12), 04019057. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001135

Bouché, V., & Crotty, S. M. (2018). Estimating demand for illicit massage businesses in Houston, Texas. *Journal of Human Trafficking 4* (4) , 1–19 https://doi.org/10.1080/23322705.2017.1374080.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bullard, R. T. D. (1993). Race and environmental justice in the United States. *Yale Journal of International Law*, *18* (1) , 319–335 https://heinonline.org/HOL/LandingPage?handle=hein.journals/yjil18&div=17&id=&page=.

Burnson, P. (2019, May 10). *Top 30 U.S. Ports 2019: Trade tensions determine where cargo goes next*. https://www.logisticsmgmt.com/article/top_30_u.s._ports_trade_tensions_determine_where_cargo_goes_next

Chin, J. J., Takahashi, L. M., & Wiebe, D. J. (2019). Where and why do illicit businesses cluster? Comparing sexually oriented massage parlors in Los Angeles County and New York City. *Journal of Planning Education and Research*, 0739456X1985963. https://doi.org/10.1177/0739456X19859637

Crotty, S., & Bouché, V. (2018). The red-light network: Exploring the locational strategies of Illicit Massage Businesses in Houston, Texas. *Papers in Applied Geography*, *4*(2), 205–227. https://doi.org/10.1080/23754931.2018.1425633

Dubrawski, A., Miller, K., Barnes, M., Boecking, B., & Kennedy, E. (2015). Leveraging publicly available data to discern patterns of human-trafficking activity. *Journal of Human Trafficking*, *1*(1), 65–85. https://doi.org/10.1080/23322705.2015.1015342

Durchslag, R., & Goswami, S. (2008). *Deconstructing the demand for prostitution: Preliminary insights from interviews with Chicago men who purchase sex*. Chicago Alliance Against Sexual Exploitation.

Edwards, M. L. (2010). Gender, social disorganization theory, and the locations of sexually oriented businesses. *Deviant Behavior*, *31*(2), 135–158. https://doi.org/10.1080/01639620902854852

Eiselt, H. A., & Laporte, G. (1998). Demand allocation functions. *Location Science*, *6*(1–4), 175–187. https://doi.org/10.1016/S0966-8349(98)00056-4

ESRI. (2017). *ArcGIS: Map Desktop*. (Version 10.5.1) [Computer Software]

ESRI. (2019, January 10). *Light gray canvas map: World light graycanvas base*. Scale Not Given. [basemap]. http://www.arcgis.com/home/item.html?id=ed712cb1db3e4bae9e85329040fb9a49. Sources: Esri, DeLorme, HERE, MapmyIndia

Fainstein, S. (2011). The just city. *Contemporary Sociology: A Journal of Reviews*, *40* (5) , 584–585. http://doi.org/10.1177/0094306111419111q

Farley, M., Bindel, J., & Golding, J. M. (2009). *Men who buy sex: Who they buy and what they know*. Eaves.

Farrell, A., McDevitt, J., & Fahy, S. (2008). *Understanding and improving law enforcement responses to human trafficking* (pp. 92–102). Northeastern University Institute on Race and Justice.

Friedrich, M. (2019). *Income, poverty and health insurance coverage in the United States: 2019* (Report No. CB20-145). U.S. Census Bureau.

Guikema, S. D., & Quiring, S. M. (2012). Hybrid data mining-regression for infrastructure risk assessment based on zero-inflated data. *Reliability Engineering & System Safety*, *99*, 178–182. https://doi.org/10.1016/j.ress.2011.10.012

Hosmer, D., & Lemeshow, S. (2000). Chapter 5: Assessing the fit of the model. In *Applied logistic regression* (2nd ed., pp. 160–164). John Wiley and Sons.
Kabir, E., Guikema, S. D., & Quiring, S. M. (2019). Predicting thunderstorm-induced power outages to support utility restoration. *IEEE Transactions on Power Systems*, *34*(6), 4370–4381. https://doi.org/10.1109/TPWRS.2019.2914214
Kapoor, R., Kejriwal, M., & Szekely, P. (2017, May). Using contexts and constraints for improved geotagging of human trafficking webpages. *Proceedings of the fourth international ACM workshop on managing and mining enriched geo-spatial data* (p. 3). Chicago, Illinois: ACM
Kelly, E., & Cooper, C. (2000). *Everything you always wanted to know about regulating sex businesses*. American Planning Association.
Keyhan, R., Carton, M., Kegler, A., Buckley, V., Boone, S., Loza, F., & Schaeffer, A. (2017). *Human Trafficking in Illicit Massage Businesses* (Report). Polaris.
Lasker, S. (2002). Sex and the city: Zoning pornography peddlers and live nude shows. *UCLA Law Review*, *49*, 1139 https://heinonline.org/HOL/LandingPage?handle=hein.journals/uclalr49&div=32&id=&page=.
Law, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22 https://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf.
Liu, Z. (2017). *Canada and Mexico border crossings* [Data set]. Homeland Infrastructure Foundation-Level Data (HIFLD). https://hifld-geoplatform.opendata.arcgis.com/datasets/canada-and-mexico-border-crossings
Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, *5*(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d
McRoberts, D. B., Quiring, S. M., & Guikema, S. D. (2018). Improving hurricane power outage prediction models through the inclusion of local environmental factors. *Risk Analysis*, *38*(12), 2722–2737. https://doi.org/10.1111/risa.12728
Mitchell, K. (2019, January 22). *Using computer science to tackle human trafficking*. https://gwtoday.gwu.edu/using-computer-science-tackle-human-trafficking?fbclid=IwAR3RFT8rsYF19ZMBUsJMnZ8zSFJQbkxMCmIlSwD7yKVBatPFOv-Hru0aDI
Mletzko, D., Summers, L., & Arnio, A. N. (2018). Spatial patterns of urban sex trafficking. *Journal of Criminal Justice*, *58*, 87–96. https://doi.org/10.1016/j.jcrimjus.2018.07.008
Murphy, A. K., & Venkatesh, S. A. (2006). Vice careers: The changing contours of sex work in New York City. *Qualitative Sociology*, *29*(2), 129–154. https://doi.org/10.1007/s11133-006-9012-2
Nagpal, C., Miller, K., Boecking, B., & Dubrawski, A. (2015). *An entity resolution approach to isolate instances of human trafficking online*. arXiv preprint arXiv:1509.06659
Owens, D. (1997, January). *Regulating sexually oriented businesses*. UNC School of Government, Legal Topics. https://www.sog.unc.edu/resources/legal-summaries/regulating-sexually-oriented-businesses
Polaris. (2018, December 13). *Is it human trafficking?* https://polarisproject.org/is-it-human-trafficking
R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/
Rabbany, R., Bayani, D., & Dubrawski, A. (2018, July). Active search of connections for case building and combating human trafficking. *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery data mining* (pp. 2120–2129). London, United Kingdom: ACM.
RCMS, Prepared by Social Explorer. (2010). *Religion 2010* [Data set].
Red light District Amsterdam. (n.d.). Retrieved March 05, 2021, from https://www.amsterdam.info/red-light-district/
RubMaps.ch. (2019 Retrieved Feb - Aug 2019)https://www.rubmaps.ch
Silva, D. R., Philpot, A., Sundararajan, A., Bryan, N. M., & Hovy, E. (2014, June). Data integration from open internet sources and network detection to combat underage sex trafficking. *Proceedings of the 15th annual international conference on digital government research* (pp. 86–90). Aguascalientes, Mexico: ACM
U.S. Census Bureau, American Community Survey. Prepared by Social Explorer. (2016). *American Community Survey 5-Year Estimates* [Data set].
U.S. Census Bureau, Department of Commerce. (2015). *TIGER/Line Shapefile: U.S. Military Installation National Shapefile* [Data set].
U.S. Census Bureau. (2016a). *TIGER/Line Shapefiles: Census Tracts* [Data set].
U.S. Census Bureau. (2016b). *TIGER/Line Shapefiles: Primary Roads* [Data set].
U.S. Census Bureau. (2018). *TIGER/Line Shapefiles: Counties*. [Data set].
U.S. Department of Transportation/Bureau of Transportation Statistics. (2018). *National Transportation Atlas Database (NTAD)* [Data set].
U.S. Geological Survey, Department of the Interior. (2018). *National Transportation Dataset (NTD)* [Data set].
White, A. G., Guikema, S. D., & Logan, T. M. (2021). Urban population characteristics and their correlation with historic discriminatory housing practices. *Applied Geography*, *132*, 102445. https://doi.org/10.1016/j.apgeog.2021.102445