Final Research Proposal:

How Anti-Homelessness Laws Affect What the Public Sees and Feels

Grant Donovan

University of Central Florida, College of Community Innovation & Education

PAD6701: Analytical Techniques for Public Administration

Dr. Robbin Mellen

December 6th, 2024

Literature Review

In March of 2024, Florida Governor Ron DeSantis signed into law a bill criminalizing all overnight camping in public spaces (Anderson, 2024). The timing of this new law was no surprise. The number of homeless individuals in the U.S. has been growing rapidly over the last decade. This has increased the salience of homelessness as a social problem among the public, which puts significant political pressure on elected officials to find quick solutions (Cohen et al., 2019). In response, governments across the United States have pursued an agenda of criminalizing homelessness. According to the National Law Center on Homelessness and Poverty, the last decade has seen a 69% increase in the number of U.S. cities with public camping bans, a 43% increase in those with begging bans, an 88% increase in those that ban standing or laying around in public spaces, and a staggering 143% increase in those that ban sleeping in vehicles (Rankin, 2019). As a result, there are now hundreds of U.S. cities in which many aspects of a homeless person's daily life are now criminal activities. Florida's statewide ban on public camping is only the latest development in this nationwide trend.

Academic literature is in widespread agreement that these measures criminalizing homelessness are ineffective, expensive, and cruel. Regarding the former, advocates for criminalizing homelessness often argue that such measures are necessary to push the homeless into social services such as shelters (Anderson, 2024). And research does lend some credence to this viewpoint. After Denver instituted a public camping ban in 2012, for instance, 40% of the city's surveyed homeless population said they sought out shelter beds more frequently than before the ban (Langegger & Koester, 2016). However, this did nothing to change the actual number of homeless people living on the city's streets every night. Like every major U.S. city, Denver's shelters were already operating at capacity almost every night before the ban

(Robinson, 2017). In addition, many homeless people do not have the option of staying in shelters due to common restrictions on pets, couples, the mentally ill, LGBTQ people, substance abusers, or those with criminal convictions staying within them (Rankin, 2019). And this is all before considering that homeless shelters are often overcrowded, unsanitary, noisy, and dangerous places for people to stay (Rankin, 2019). In short, the chronically homeless often lack any reasonable alternative to living in public spaces.

Due to these barriers, academic studies overwhelmingly find that these laws do not move people off the street and into housing. Instead, almost every homeless person who is approached by police merely moves to a new location or waits for the police to leave before resuming their behavior in the original location (Robinson, 2017). And not only do these laws do nothing to move people into housing, but they make it more difficult to escape poverty and homelessness. Interviews with the homeless commonly find that police interactions erode trust and create an adversarial relationship between the homeless and the rest of the community (Cohen et al., 2019). This often prevents the homeless from being willing to work with outreach workers or seek out social services themselves. Furthermore, since these laws commonly lead to the homeless being arrested for performing necessary daily activities, they saddle many with criminal records they would not otherwise have (Clifford & Piston, 2016). This makes it near impossible for these individuals to access shelter beds or land a job, further preventing them from escaping homelessness and poverty. Due to these impacts, almost every scholar agrees that these laws actually increase the prevalence of homelessness in the communities in which they are enacted. At a significant cost to local government too, it should be added. Seattle, for example, spent over $20 million displacing homeless people in 2018 (Rankin, 2019).

In addition to being ineffective and expensive, scholars are in unanimous agreement that laws criminalizing homelessness are highly detrimental to the health and well-being of homeless individuals. The most common manifestation of these laws is the "move along" order. While seemingly benign, the cumulative effect of these orders on the life of a homeless person is disastrous. Most notably, they force those sleeping on the streets out of the central, well-lit areas of a city most often patrolled by the police. Instead, these individuals are forced to walk miles to the city's outskirts to find a restful night's sleep. These areas are almost always significantly less safe, and this inevitably results in the homeless being victimized and traumatized at much higher rates (Robinson, 2017). Furthermore, in order to survive day-to-day, the homeless must rely on sophisticated micro-geographical patterns. They need a place to sleep, a place to do laundry, a place to cook or receive food, a place to charge their phone, a place to socialize, a place to store their things, and many more. Yet, if the individual is constantly being forced to move into new, unfamiliar locations, these patterns become impossible to establish. This can make it nearly impossible for a homeless person to accomplish the day-to-day activities necessary for dignified survival (Langegger & Koester, 2016). And this is all before considering the physical and psychological pain, as well as the destruction of personal property, regularly inflicted on the homeless by the police in their more hostile interactions (Robinson, 2017).

It is fair to wonder, then, why governments are increasingly coming to utilize the criminalization approach to homelessness when scholars have long and conclusively established that it is ineffective, expensive, and cruel. The answer can be found in the final justification Ron DeSantis gave for signing Florida's new public camping ban: "Homelessness is a blight on our cities and a nuisance for business" (Anderson, 2024). The fact of the matter is that most people do not see the homeless as fellow community members or even as human beings. Instead, the

homeless are viewed as “human trash” – an urban pollution that must be cleaned up or at least hidden so to protect the aesthetics of the city (Bonds & Martin, 2016). In fact, studies find that Americans react to visible poverty with higher rates of negativity than to any other marginalized trait, including race (Rankin, 2019). Some scholars argue that this is due to biological mechanisms. They theorize that people use the stigmata of homelessness – dirty clothing, poor hygiene, carrying personal belongings, etc. – as heuristics for identifying those with communicable diseases (Clifford & Piston, 2016). Because our brains have evolved to be disgusted by the sources of pathogens, we correspondingly view the homeless with disgust and desire physical separation from them. Other scholars theorize that stigma against the homeless is instead a social creation. They argue that politicians and the media’s tendency to “otherize” unconventional groups has led to homelessness being inaccurately framed as a problem of deviant behavior and personal failure (Owadally & Grundy, 2023).

But regardless of the reasoning behind it, the reality is that people do not want to see homelessness in their cities and thus support criminalization as a method for reducing its visibility (Robinson, 2017). However, scholars are divided on whether it actually accomplishes this goal. On the one hand, studies consistently find that criminalization causes the homeless to avoid the central areas of a city in favor of the less visible outskirts (Robinson, 2017). According to some scholars, this reduces the visibility of homelessness as people no longer see the homeless sleeping and building encampments in popular areas. However, other scholars disagree. While the homeless are indeed forced to sleep on the outskirts of the city, the services they rely on for survival remain in the center (Langegger & Koester, 2016). This means that they must still travel to the popular areas of the city during the hours in which the community is using them. Furthermore, because criminalization disrupts their ability to do laundry, attend to hygiene, and

store their personal belongings, the homeless are less able to disguise their stigmata while in these popular areas (Langegger & Koester, 2016). For these reasons, other scholars theorize that criminalization laws actually serve to increase the visibility of homelessness. This disagreement represents the gap in the literature the following study is designed to investigate more fully.

Hypothesis 1

To fill this gap in the literature, the first hypothesis this study will investigate is: "in a study of cities with more than 50,000 residents, the presence of laws criminalizing homelessness will be associated with a greater degree of visible homelessness." The population of interest has been restricted to cities with more than 50,000 residents because these are the areas to which the homeless naturally gravitate (U.S. Department, 2023). The independent variable, then, will be the presence of laws criminalizing homelessness in a given city. These laws can be broadly categorized by the specific behaviors they attempt to ban. Based on the behaviors identified in the literature review as common targets for criminalization, this study will examine the effects of laws banning sleeping, sitting or lying down, loitering, panhandling, storing possessions, eating, or bathing in public spaces. Bans on sleeping in one's vehicle will also be included.

One way of measuring this independent variable is to simply count up the number of these behaviors that a city has attempted to ban. However, this method introduces the potential for measurement bias. It is unlikely that each of these bans alters the visibility of homelessness in the exact same way. For example, it is entirely possible that a ban on sleeping in public decreases the visibility of homelessness while a ban on bathing in public increases it. Yet, this measurement scheme would code a city with either of these bans the same way, errantly treating their effects on the dependent variable as if they are the same. For this reason, each category of ban will need to be examined separately. This will be accomplished by using eight different independent

variables – each one a nominal dummy variable measuring the presence (or absence) of criminalization laws banning a specific behavior in a given city. Measuring these independent variables will not be difficult. All U.S. cities of this size publish their municipal code online (see Municode Library, 2024; City of San Diego, 2024). Determining whether a city bans a certain behavior or not can be easily accomplished by performing keyword searches of terms related to that behavior on these online records.

The dependent variable of this first hypothesis is the degree of visible homelessness in a given city. This is not a variable that many researchers attempt to measure, and those that do claim to measure it often use very narrow definitions. For example, the U.S. Department of Housing & Urban Development (HUD) claims to collect data on the visibility of homelessness through its annual point-in-time counts (U.S. Department, 2023). These counts measure the total number of people sleeping on the streets on a given night in January. Yet, while this data is certainly useful for measuring the number of unsheltered individuals in the country, its relationship to visibility is tenuous. These counts take place at night when there are few people around to actually see others sleeping in public. Additionally, these counts ignore the many other examples of stigmata that contribute to the visibility of homelessness: dirty clothing, poor hygiene, panhandling, carrying many possessions, public drug use, and more. For these reasons, using data from HUD's point-intime counts – or any other organization that uses this method – would be inappropriate for measuring the broad definition of visibility of interest in this study.

Similarly, police reports and citizen 311 complaints are notoriously inaccurate data sources that leave out most instances of visible homelessness (Robinson, 2017). In short, useful data on this variable is not readily available and will need to be collected directly by the researchers. Luckily, this can be easily accomplished through direct observation. After selecting

a sample of cities, the researchers should identify areas within them that receive the most foot traffic in a given day. This is where homelessness is most visible to the general public. The researchers will then need to select a sample of timeframes in which they will go collect observational data from these areas, being careful to include a wide range of times of days of the week. Then, as they observe the selected areas during the selected timeframes, they should vigilantly record all instances of homelessness stigmata that they can visually identify. This method will produce frequency count data, giving this dependent variable a ratio level of measurement. This collection method should produce data that better captures the true scope and prevalence of visible homelessness than any of the data sources currently available.

Hypothesis 2

The original impetus for this study was to inform debates on enacting or repealing laws criminalizing behaviors associated with homelessness. While the effects of such laws on the visibility of homelessness is likely of principal concern to residents involved in this debate, it is not the ultimate concern of elected officials. Instead, these officials primarily care about relieving the political pressure they feel to address homelessness as a local issue (Cohen et al., 2019). Therefore, if these laws can be proven to elevate public irritation with homelessness – either by increasing the prevalence of homelessness, increasing the visibility of homelessness, or by increasing the public's perception of homelessness as a threat to public safety and order – then elected officials may be far less willing to consider these ineffective, expensive, and cruel laws (Langegger & Koester, 2016). Accordingly, this study will also investigate a second hypothesis: "The presence of laws criminalizing homelessness increases the average resident's perception of homelessness as a local problem."

This second hypothesis utilizes the same population of interest and independent variables as the first. The dependent variable, though, becomes "the average resident's perception of homelessness as a local issue." This is a variable that a great many researchers have already taken an interest in measuring. Organizations like the National Alliance to End Homelessness, Morning Consult, the Public Policy Institute of California, and more have collectively conducted thousands. of surveys in the last five years asking U.S. residents about this specific topic (National Alliance, 2024; Thomas, 2024; Torres, 2023). There is no reason to suspect that this survey data does not accurately measure the dependent variable of this hypothesis. Each was conducted with great care to ensure validity and reliability. Even between organizations, the methods utilized are remarkably consistent. All of them asked respondents to rate the severity of homelessness in their local area on an ordinal scale from "not a problem" to "a very serious problem." This kind of data can easily be recoded to all use the same measurement scale, allowing for the use of multiple data sources together. This means that this variable has an ordinal level of measurement.

The question becomes, from which combination of sources should the data be taken? This is a decision that should be based on two factors. First, because this study's unit of analysis is the city, only data sources that include a large number of respondents from the same city should be selected. This will give researchers the sample size necessary to infer the average resident's perception of homelessness as a local problem. Second, sources should be selected based on the variation in the independent variables they allow for. For example, if the respondents of one survey all come from cities that criminalize sleeping in public, then it would not be useful for analyzing how cities with this ban differ from those without it. Of course there is the possibility that, even collectively, these data sources do not provide sufficient variation

across all eight independent variables. In this case, the researchers will need to collect new data from residents of cities with or without certain laws criminalizing homelessness (whichever is needed). If this proves necessary, researchers should use a survey that matches the methods and wordings of the other data sources as closely as possible. This increases the reliability of the data as a whole and will lead to more valid findings (Newcomer et al., 2015). Alternatively, the researchers may decide to omit any independent variable that lacks the necessary variation.

Data Cleaning

After the data collection phase of the study is complete, but before the statistical analysis phase of the study can begin, comes the crucial step of data cleaning. If the sample data is not properly cleaned before conducting statistical tests with it, the study's results may not be accurate in describing the population of interest due to measurement bias and violations of the test assumptions. It is always assumed that raw data have errors among them, so this step should be taken regardless of how carefully the researchers conducted the data collection phase. First, the researchers should ensure that there are no missing values among the data. If there are, common practice is that these observations be excluded from the study.

Second, the researchers should make sure that all values included in the data fall within the plausible range of that variable's coding scheme. Any values that do not fall within this range are indicative of reporting and recording errors having occurred during the data collection phase. For this study's independent variables, this means identifying any values other than 0 and 1. For the first hypothesis' dependent variable – number of observed instances of homelessness – this means identifying any value that is not a positive integer. And for the second hypothesis' dependent variable – average residents' perception of homelessness – this means identifying any

value that does not fall between one and three. Any observation with these values should similarly be excluded from the study.

Since this study uses whole cities as units of observation, and collecting aggregate data at this level of observation requires much time and expense, there are unlikely to be more than thirty observations included in the sample. Thus, even a handful of observations with missing or implausible values could comprise a large proportion of the sample. Should this be the case, the researchers should question whether they need to return to the data collection phase to ensure that the sample data is accurate and representative before moving onto the data analysis phase. Third and finally, the researchers should make sure that there are no outlier values among the data. The presence of outliers violates the assumptions of the regression models to be utilized later on during the data analysis phase, neither of which are robust to violations of their assumptions. Outliers can be identified as those that lie more than three standard deviations away from that variable's mean, either above or below. Any observations with values that fit this criterion should also be excluded from the study. Though, the researchers should report whether including them significantly alters the results that the statistical tests produce.

Data Analysis

Hypothesis 1

After data cleaning is complete, the researchers can turn their attention towards statistically analyzing this study's first hypothesis. The first step in doing so is to explore whether the mean number of visible instances of homelessness significantly differs between cities that have a certain criminalization law and those that do not. The presence of significantly different means provides an early indication that passing a given criminalization law does have some

relationship to the amount of visible homelessness in that city. For example, if the researchers find that cities which ban sleeping in one's car have statistically higher levels of visible homelessness than cities without such a ban – meaning that the observed differences in the sample are unlikely to have been caused by sampling error – then they would conclude that it is possible that bans on sleeping in one's car have some sort of relationship to visible homelessness; its even possible that these bans directly create more visible homelessness. While this initially seems to be a weak conclusion, it is a necessary first step in the data analysis phase. Experts always recommend testing bivariate relationships before moving on to larger models because doing so identifies relevant predictors and gives researchers a better understanding of the underlying relationships present in the data.

Producing these findings will require the use of either T-tests of Mann-Whitney U tests. T-tests are the most powerful test for analyzing the differences in means between two groups, but it may not be appropriate for analyzing the data collected in this study. The first reason for this is that T-tests assume that the dependent variable of interest is distributed normally and symmetrically. Yet, data collected through frequency counts, as the data for the first hypothesis' dependent variable is, are often positively skewed. The researchers should investigate this by conducting a Shapiro-Wilk test on this first dependent variable. If this test returns a p-value greater than 0.05, then it is not normally distributed. In this case, a Mann-Whitney U test – a nonparametric alternative to the T-test that does not assume a normally distributed dependent variable – should be used instead. An additional complication is that the data collected through frequency counts are not strictly continuous but discrete; they are constrained to integer values and cannot vary continuously across decimal values. While the results of a T-test are generally robust to violations of this assumption, another advantage of the Mann-Whitney U test is that it

does not make this assumption at all. Finally, researchers should investigate whether the data satisfy the T-test's assumption of equal variances. This can be done by conducting a Levene's test. If this test returns a p-value less than 0.05, then this would be another indication that the analysis should be conducted using Mann-Whitney U tests instead.

The null hypothesis of a T-test is that two groups do not have statistically different means. Again using the example of laws criminalizing sleeping in one's vehicle, the null hypothesis would be "there is no difference in the amount of visible homelessness between cities that ban sleeping in one's car and those that do not have this type of ban" in the context of this study. Note that this is the null hypothesis for a two-tailed T-test, which is more appropriate here as we are primarily interested in exploring the data and identifying relevant predictors rather than directly evaluating the study's first hypothesis. If Mann-Whitney U tests are to be used instead, then the null hypothesis instead becomes that the two groups do not have statistically different distributions. While less descriptively powerful, finding that two groups have statistically different distributions would also be an early indicator that there is a relationship between them.

Conducting the T-tests or Mann-Whitney U tests, a step easily performed by statistical analysis software, would then produce a p-values that can be interpreted as the likelihood that rejecting these respective null hypotheses would be a mistake. The question becomes, then, how much risk we are willing to accept in rejecting the null hypothesis – what is known as the study's "alpha level" (a). While most studies use an alpha level of a = 0.05, a willingness to be wrong 5% of the time when rejecting the null hypothesis, such strictness is often not necessary in the social sciences. It would be especially inappropriate here, since the consequence of mistakenly accepting the null hypothesis is letting these ineffective, costly, and immoral laws go unchallenged. For these reasons, this study should instead use an alpha level of a = 0.1, or a

willingness to be wrong 10% of the time when rejecting the null hypothesis. Researchers will therefore reject the null hypothesis should a t-test or Mann-Whitney U test return a p-value less than or equal to 0.1. Again, in this study this would mean tentatively concluding that there is some sort of relationship between the presence of a given criminalization law and the amount of visible homelessness in the city. This process should be repeated for each of the eight independent variables of interest in this study. Descriptive statistics for each group can also be compared to see which partitions produce the most sizeable differences in visible homelessness.

No matter whether T-tests or Mann-Whitney U tests are used, rejecting the null hypothesis with respect to a particular independent variable is evidence that the criminalization law it measures the presence of is a relevant predictor of visible homelessness in a city. So, for example, if the researchers reject the null hypothesis when using the independent variable measuring the presence of public sleeping bans but do not reject the null hypothesis when using the independent variable measuring the presence of panhandling bans, then they would conclude that the presence of a public sleeping ban is a relevant predictor of visible homelessness while the presence of a panhandling ban is not. Identifying relevant predictors is important because the next step in the data analysis process is to construct a multiple regression model. Such a model aims to achieve full model specification, which is the inclusion of every independent variable that affects the dependent variable and the exclusion of every independent variable that does not affect the dependent variable. Failing to achieve full model specification results in specification bias and inaccurate findings; including irrelevant variables would artificially deflate the model's findings while excluding relevant variables would artificially inflate the model's findings. By first identifying the relevant independent variables through exploratory bivariate tests, the

researchers can make educated decisions on which are relevant and should be included in the multiple regression model – thus avoiding specification bias.

Furthermore, including all eight of these independent variables together in a single model is likely to result in multicollinearity. A city that has passed a ban on public sleeping is also likely to have passed a ban on other behaviors associated with homelessness – causing the independent variables to vary together to a large degree. This distorts the math of a multiple regression model because the effects of a variable are difficult to discern from those of another if they largely vary together. After the irrelevant variables have been removed, the researchers should find each remaining independent variable's VIF (variance inflation factor). Any variable with a VIF greater than 5 should also be excluded from the multiple regression model. It is very important that this step take place after removing the irrelevant variables, so that a relevant variable is not excluded from the regression model based on its collinearity with another variable which would not be included in the regression model anyway.

Analyzing the independent variables using a multiple regression model is necessary for discovering their true relationship to visible homelessness and thus evaluating this study's first hypothesis. The first reason for this is that a regression model gives one much more information on the independent variables' relationships to the dependent variable. Rather than simply indicating whether a relationship could possibly exist – as the exploratory tests did – a multiple regression model reveals the likely size and direction of each relationship. So, for example, it would tell the researchers exactly how the presence of a ban on public sleeping is associated to the amount of visible homelessness in a city. It would also reveal the exact likelihood that this association is simply a result of sampling error. This information directly addresses this study's

first hypothesis, clearly indicating whether the researchers should accept or reject it in regard to each type of criminalization law.

The second reason for using a multiple regression model is that it allows the researchers to control for rival hypotheses. For example, one may claim that bans on public sleeping are only positively associated with visible homelessness because the cities with this ban are primarily located in the areas with warm climates. These are the cities the homeless naturally gravitate towards, so we would expect to find more visible homelessness in this group of cities whether they had public sleeping bans or not. The researchers could address this rival hypothesis by including climate as a covariate in the multiple regression model. If this is done, then the interpretation of the model's results becomes the association between criminalization laws and visible homelessness controlled for the effects of climate on visible homelessness. This eliminates this rival hypothesis as an explanation for the observed association. The researchers should therefore include in the multiple regression model any factor that may serve as the basis of a rival hypothesis. Based on the literature review, these factors may include climate, poverty rate, unemployment rate, housing prices, population and population density, mental health problems and substance abuse prevalence, and welfare and public housing availability. Selecting from among these control variables should similarly be guided by the goals of including only relevant variables – perhaps by employing a backward stepwise selection process – and avoiding multicollinearity. The researchers should also conduct check the final model's robustness by investigating whether using different combinations of control variables changes the results the model produces and the conclusions the researchers come to.

While a multiple regression model is the ideal statistical test to utilize in evaluating the first hypothesis, due to its analytical power and the ease with which dichotomous independent

variables can be included, its assumptions may make it inappropriate for use in analyzing the variables of interest to this study. First, multiple regression assumes a linear relationship exists between the independent variables and the dependent variable. Yet, when all of the independent variables in a model are dichotomous – as is the case in evaluating this hypothesis – the independent variables are likely to interact in ways that are multiplicative. Take bans on public sleeping and bans on sleeping in one's car, for example. It is entirely possible that, when a ban on public sleeping is in effect, many people will respond by instead sleeping in their cars. Similarly, when a ban on sleeping in one's car is in effect, many people will respond by sleeping in public. In either case, visible homelessness may increase slightly, but the absence of the other ban would prevent it from increasing dramatically. In cities where both bans are in effect, then, we would expect to see a very large increase in visible homelessness. This is an example of a multiplicative interaction: the presence of one ban amplifies the effect of the other, leading to a combined increase that is greater than the sum of their individual effects. Such relationships are inherently non-linear because they are not additive. Such multiplicative interactions may exist between any combination of the independent variables included in the final model, making it exceedingly likely that their cumulative relationship with the dependent variable is non-linear.

Second, multiple regression assumes that there is no heteroscedasticity present in the data being analyzed. In the context of this study's first hypothesis, this means that the variance in the amount of visible homelessness should be constant across cities with and without a certain criminalization law. Unfortunately, this is unlikely to be the case because this study's dependent variable comes from a frequency count. Frequency count data is typically positively skewed, in which case variance tends to increase as the mean increases. Since the independent variables are dichotomous – and the groups being compared in the final regression model are expected to have

different means – this means that it is likely that the variances between these groups will differ. For example, if the mean amount of visible homelessness in cities with public sleeping bans is greater than the mean amount of visible homelessness in cities without public sleeping bans, then we should also expect the variance in visible homelessness to be greater among cities with public sleeping bans. This would cause the model's residuals to be distributed unevenly across the different groupings, leading to heteroscedasticity.

Of course, it is possible that neither of these assumptions will be violated. However, researchers should still be prepared for that eventuality and ready to investigate accordingly. Both of these violations can be identified through visual analysis of the model's residual plot. If the plot seems to present a relationship other than a horizontal line, then the assumption of linearity has been violated. If the variance in the error term is not equal across the entire range of observations, then the assumption of no heteroscedasticity has been violated. If the researchers identify either of these violations, then they must rectify the model in some way since multiple regression is not robust to violations of its assumptions.

The first option for rectifying these violations is transforming the model's dependent variable. This method adjusts the scale and distribution of the dependent variable's values, better aligning it with the assumptions of the multiple regression model. For example, by linking the dependent variable to the independent variables through a logarithmic function, the multiplicative relationship between them may be able to be represented by a linear equation. This would allow for estimation by a multiple regression model. Additionally, a logarithmic transformation of the dependent variable reduces the measure of its variance at greater means, correcting for the problem of heteroscedasticity. If the researchers can find a logarithmic transformation that results in an error term plot that is not curvilinear, constant in variance, and

normally distributed, then they should continue on with conducting the multiple regression test (a step easily performed by statistical analysis software) and interpreting its results.

However, even transforming the dependent variable may not be enough to satisfy multiple regression's assumptions. This could be because this study does not allow for the collection of a large sample, introducing the possibility that the error term will be non-normally distributed due to sampling variability. Another problem could be the discrete nature of the dependent variable. Since multiple regression estimates continuous outcomes, this disconnect could affect the resulting residuals. Finally, frequency count data is always non-negative, but a multiple regression model might end up predicting negative outcomes. This would greatly distort many of the resulting residuals, preventing the error term from achieving normality. Finally, while transforming the dependent variable is a powerful and useful technique, a single transformation cannot always correct for the skewness and variability of the data's distribution. For any of these reasons, the researchers may find that no transformation is suitable for satisfying the assumptions of multiple regression. In this case, they should be prepared to pivot to using a Poisson Generalized Linear Model instead. This model is specifically designed for the analysis of frequency count data, as it assumes that the dependent variable is positively skewed, discrete, and non-negative. This model also does not assume a normally distributed error term or a linear relationship between variables in the model.

No matter which of these tests are used – a simple multiple regression model, a transformed multiple regression model, or a Poisson model – the results of this test will allow for the direct analysis of this study's first hypothesis. The most important result of any model will be the coefficients attached to the independent variables of interest. In a simple multiple regression model, these coefficients can be interpreted as the additional number of instances of visible

homelessness in a given time period (equal to the amount of time spent observing the city during data collection) expected when a city passes that respective criminalization law controlled for the effects other criminalization laws and additional covariates. If this number is positive – and the p-value attached to it is less than this study's alpha level of a = 0.1 – for any of the independent variables of interest included in the final model, then the researchers should conclude that criminalization laws are indeed associated with greater amounts of visible homelessness. Note that the researchers should not rely on the model's global F-test, since this result would include the effects of the additional covariates included in the final model as controls. A coefficient with a p-value attached that is greater than a = 0.1 indicates that the variable's relationship to visible homelessness is not statistically significant. If all the independent variables of interest in the final model have a p-value such as this, then the researchers should instead fail to reject the null hypothesis that criminalization laws are not associated with greater amounts of visible homelessness.

If one of the other models are used, the method for reaching a conclusion on the study's first hypothesis is exactly the same. If any of the independent variables of interest have a positive coefficient, which has an attached p-value less than 0.1, the researchers should conclude that criminalization laws are indeed associated with greater amounts of visible homelessness. It is only the interpretation of the model coefficients that changes. In the case of a log-transformed dependent variable multiple regression model, the coefficients can be interpreted as the percentage change in visible homelessness expected when a city passes that respective criminalization law controlled for the effects other criminalization laws and additional covariates. Finally, in the case of the Poisson model, each coefficient (b) should first be input into the equation: $(e^b - 1)(100) = x$. This number "x" can then be interpreted as the percentage

change in visible homelessness expected when a city passes that respective criminalization law controlled for the effects other criminalization laws and additional covariates.

Hypothesis 2

The analysis of this study's second hypothesis will mirror the procedure used to analyze its first hypothesis. The first step will be an exploratory investigation into which of the eight independent variables have a potential relationship to a city's residents' average perception of homelessness as a local issue. Once again, these results will serve as the basis for identifying relevant predictors for inclusion in a later regression model. In this case of this second hypothesis, the dependent variable is ordinal rather than continuous. This violates an assumption of the t-test, making it inappropriate for analysis here. Instead, the Chi-squared test is the statistical test recommended for analyzing whether there is an association between a dichotomous independent variable and an ordinal dependent variable. This test evaluates whether the observed distribution of the dependent variable across the two dichotomous categories – in this case, the presence or absence of a criminalization law – is significantly different from what would be naturally expected. This naturally expected distribution is equal to the overall proportion of cities that fall into each dependent variable category, regardless of the presence or absence of a given criminalization law. If a significant difference from this distribution is observed, then this test provides an early indication that passing a given criminalization law does have some relationship to a city's residents' average perception of homelessness as a local issue. For example, if the distribution of the dependent variable for all cities is normal, but its distribution for cities with bans on public sleeping is negatively skewed and significantly different, then there might be some relationship between the two variables that causes the distribution to deviate from what was naturally expected. The researchers would then conclude

that it is possible that bans on public sleeping worsen a city's residents' average perception of homelessness as a local issue – justifying its inclusion in a regression model as a predictor.

While most of the assumptions of chi-squared test should be satisfied by the researchers' careful management of the data collection process – random observations and independent observations, for example – there is one that may present trouble. The Chi-squared test also assumes that at least 20% of the possible combinations of the independent variable and the dependent variable will have at least five expected observations. Since the sample size to be collected is likely to be small, perhaps no more than thirty, it would be very difficult for the data to satisfy this assumption if there are more than six potential combinations. Because of this, the researchers may want to combine the data so that this dependent variable has only three categories ranging from "not a problem" to "a very serious problem." This coding scheme would also have the added benefit that many data sources are already coded according to it; and while it is possible to combine data with more categories into this simpler scheme, it is not possible to recode data with this simpler scheme to have more categories. That all being said, it is of course still a possibility, due to the distribution of the dependent variable, that two of the potential combinations will have less than five expected observations. This would represent 33% of the potential combinations, violating the expected frequency assumption. If this proves to be the case, then the researchers should turn to using Fisher's Exact test instead. This is a nonparametric alternative to Chi-squared that does not rely on the expected frequency assumption and is well-suited for small sample sizes.

The null hypothesis of both the Chi-squared test and the Fisher's Exact test is that there is no difference between the distribution of the dependent variable for the dataset as a whole and the distribution of the dependent variable for any subgroup of cities with or without a given

criminalization law. Again using the example of public sleeping bans, the null hypothesis for this dependent variable would be, “there is no difference in the distribution of city’s residents’ average perception of homelessness as a local issue between all the cities in the dataset and the cities that have a public sleeping ban in place.” Conducting the Chi-squared test or the Fisher’s Exact test, a step easily performed by statistical analysis software, would then produce a p-values that can be interpreted as the likelihood that rejecting these respective null hypotheses would be a mistake. Again, this study will use an alpha level of a = 0.1, meaning that the researchers should reject the null hypothesis for any of the tests that returns a p-value less than 0.1. Again, in this study this would mean tentatively concluding that there is some sort of relationship between the presence of a given criminalization law and a city’s residents’ average perception of homelessness as a local issue. This process should be repeated for each of the eight independent variables of interest in this study. Visual analysis of each partition’s distribution of the dependent variable should also be conducted to further familiarize the researchers with the underlying data.

Those independent variables that do not reveal potential relationships with the second hypothesis’ dependent variable should then be set aside for the rest of the analysis process. Again, this is done to avoid specification bias and the problem of multicollinearity. The VIF values of the independent variables should also be compared again, and any variable with a value higher than 5 should similarly be set aside. The researchers should then combine the remaining independent variables into a regression model, so to determine the size, direction, and likelihood of each independent variable’s relationship with the dependent variable. This directly addresses the study’s second hypothesis, as it reveals to the researchers exactly how the presence of a given criminalization law affects a city’s residents’ average perception of homelessness as a local problem. Using a regression model also allows the researchers to control for rival hypotheses.

In the case of this second hypothesis, one rival hypothesis may be that the residents of cities with criminalization laws are naturally more conservative than the residents of cities without criminalization laws. If conservative residents are more likely to view homelessness as a local problem than liberal residents, then this would pull up the average of the dependent variable for cities with criminalization laws – creating the impression that the two have a direct relationship when they really have a spurious relationship. The researchers could address this rival hypothesis by including a city's residents' average ideology as a covariate in a regression model. Based on the literature review, additional covariates the researchers may wish to include in this model are median income, unemployment rate, economic growth, housing costs, homelessness' visibility, tourism dependence, amount of local media coverage focused on homelessness, average education level, average age, racial composition, population, and population density. Selecting from among these potential covariates should be done through a backward stepwise selection process and robustness checks should be conducted through the use of different combinations of covariates in the final model.

While analyzing the first hypothesis required a multiple regression model, using the same model here would be inappropriate. A key assumption of the multiple regression model is that the dependent variable is continuous, but in this case the dependent variable is ordinal. Instead, this situation calls for using a logistic regression model, which assumes an ordinal dependent variable. The results of this model are therefore interpreted differently: coefficients represent the log odds of being in a higher category of the dependent variable, given a one-unit increase in the independent variable, while holding all other variables constant. Log odds refers to the odds ratio of an event occurring – or how much more likely an event is to occur than to not occur – which is then transformed using a natural logarithm. Therefore, if a coefficient "b" is put into the

equation "$e^b = x$," then "x" in this case would represent how much more likely it is that the presence of a given criminalization law would increase a city's residents' average perception of homelessness as a local problem up an ordered level than it is that nothing would happen instead. For example, if the coefficient attached to the independent variable measuring the presence of a public sleeping ban is 0.69314, then this would tell the researchers that the odds of a city's residents perceiving homelessness as a more serious local issue increases by a factor of $e^{0.69314} = 2$, or doubles, when a city passes a public sleeping ban, holding all other variables constant.

While the interpretation of these coefficients is interesting, the results the researchers should be most concerned with are the sign of each independent variable's coefficient and its associated p-value. A positive sign indicates that the presence of the criminalization law that independent variable measures increases the probability that a city's residents' average perception of homelessness as a local issue will worsen. A negative sign of course indicates the opposite: that the presence of the respective criminalization law instead decreases this probability. The associated p-value once again tells researchers the likelihood that this relationship is simply the result of sampling error. Therefore, if the findings of the logistic regression model reveal an independent variable with a positive coefficient attached – which has an associated p-value less than $a = 0.1$ – then the researchers should conclude that the presence of that criminalization law indeed increases the average resident's perception of homelessness as a local issue. However, if the findings reveal no independent variables attached to positive coefficients with associated p-values less than $a = 0.1$, the researchers should instead fail to reject the null hypothesis that the presence of laws criminalizing homelessness has no effect on – or perhaps even decreases – a city's residents' average perception of homelessness as a local issue.

There are two assumptions of ordinal logistic regression models that may difficulties in this study and which the researchers should be careful to monitor. First and foremost, the ordinal logistic regression model assumes a large sample size. The general rule is that, if every independent-dependent variable combination does not have at least ten observations, then an ordinal logistic regression model is unlikely to be reliable. Such a large sample size is far beyond what is feasible for this study. Because of this, the researchers should expect to use an exact logistic regression model instead. This is a variation of the ordinal regression model designed for studies with small sample sizes. While the math behind these two models is quite different, the interpretation of their resulting coefficients is exactly the same. They also produce very similar results – although exact logistic regression is slightly less powerful and it therefore produces larger p-values. The only issue may be that including a large number of predictors can make this model too computationally intensive to use. Consequently, researchers should be highly selective when deciding on which independent variables and covariates to incorporate into this model.

Second, exact logistic regression model assumes that the independent variables' coefficients are constant across all levels of the dependent variable. Third, it assumes that the relationship between the independent variables and the log odds of being in a higher category of the dependent variable is linear. So, for example, if the presence of a public sleeping ban greatly increases the odds of residents' perceptions of homelessness worsening when residents do not see homelessness as a local issue, but it decreases the odds of residents' perceptions worsening when residents already see homelessness as a local issue, then both of these assumptions would be violated. The former is violated because the relationship changes between one level and the next. The latter is violated because the relationship is first positive and then negative – making it curvilinear. The researchers should investigate the constant coefficient assumption using a Brant

test. If the p-value this test returns is less than a = 0.1, then this assumption has been violated and the model's results will be biased. The researchers should then investigate the linearity of the logit assumption by analyzing the model's partial residual plots. If any curvilinear patterns can be identified, then this assumption has also been violated. In either case, this study's small sample size and dichotomous independent variables take most potential remedies off the table. If either of these assumptions are violated, the researchers will simply need to remove the offending independent variables from the model.

References

Anderson, C. (2024, March 20). *Florida Homeless to be Banned from Sleeping in Public Spaces Under DeSantis-Backed Law*. AP News. https://apnews.com/article/homeless-floridadesantis-public-spaces-ban-f28a77bf5e445a5c26741cc9400fe40f

Bonds, E., & Martin, L. (2016). Treating People Like Pollution: Homelessness and Environmental Injustice. *Environmental Justice*, 9(5), 137–141. https://doi.org/10.1089/env.2016.0021

City of San Diego. (2023). *Municipal Code*. City of San Diego Official Website. Retrieved November 2, 2024, from https://www.sandiego.gov/city-clerk/officialdocs/municipalCode

Clifford, S., & Piston, S. (2016). Explaining Public Support for Counterproductive Homelessness Policy: The Role of Disgust. *Political Behavior*, 39(2), 503–525. https://doi.org/10.1007/s11109-016-9366-4

Cohen, R., Yetvin, W., & Khadduri, J. (2019). Understanding Encampments of People Experiencing Homelessness and Community Responses: Emerging Evidence as of Late 2018. In *U.S. Department of Housing and Urban Development Office of Policy Development & Research*. U.S. Department of Housing and Urban Development. Retrieved September 22, 2024, from https://www.huduser.gov/portal/sites/default/files/pdf/Understanding-Encampments.pdf

Langegger, S., & Koester, S. (2016). Invisible Homelessness: Anonymity, Exposure, and the Right to the City. *Urban Geography*, 37(7), 1030–1048. https://doi.org/10.1080/02723638.2016.1147755

Municode Library. (2024). CIVICPLUS. Retrieved November 2, 2024, from

https://library.municode.com/

National Alliance to End Homelessness. (2024). Summary of Public Opinion Polling on Homelessness. In *End Homelessness*. Retrieved November 2, 2024, from https://endhomelessness.org/wp-content/uploads/2024/09/Summary-of-Public-OpinionPolling-on-Homelessness-June-2024.pdf

Newcomer, K., Hatry, H., & Wholey, J. (2015). Handbook of Practical Program Evaluation. In *Wiley eBooks* (4th ed.). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119171386

Owadally, T., & Grundy, Q. (2023). From a Criminal to a Human-Rights Issue: Re-Imagining Policy Solutions to Homelessness. *Policy Politics & Nursing Practice*, 24(3), 178–186. https://doi.org/10.1177/15271544231176255

Rankin, S. K. (2019). Punishing Homelessness. *New Criminal Law Review*, 22(1), 99–135. https://doi.org/10.1525/nclr.2019.22.1.99

Robinson, T. (2017). No Right to Rest: Police Enforcement Patterns and Quality of Life Consequences of the Criminalization of Homelessness. *Urban Affairs Review*, 55(1), 41–73. https://doi.org/10.1177/1078087417690833

Thomas, D. (2024, August 13). *Overwhelming Majorities Are Concerned about Homelessness, Support Many Policies to Address It*. PPIC. Retrieved November 2, 2024, from https://www.ppic.org/blog/overwhelming-majorities-are-concerned-about-homelessnesssupport-many-policies-to-address-it/