Classifying Opportunity Zones- A Model-Based Clustering Approach

Jamaal Green
University of Pennsylvania, jamaalg@upenn.edu

Wei Shi
Travelers Insurance

Classifying Opportunity Zones- A Model-Based Clustering Approach

Abstract

Objective: Opportunity Zones (OZs) are the first major place-based economic development policy from the federal government in nearly two decades. To date, confusion persists among planners and policymakers in some places as to what features of OZ tracts matter for their inclusion, and, secondly, what features of OZ tracts make them attractive targets for potential investment. The authors developed a typology of OZ tracts in order to offer planners and policymakers alternative ways of organizing a highly variable set of tracts.

Methods: This study employs model-based clustering, also known as latent class analysis, to develop a typology OZ tracts from the population of all eligible tracts in the United States. The authors use publicly available data from the US Census Bureau and Urban Institute in developing the typology. Descriptive statistics and graphics are presented on the clusters. Using Portland, Oregon as an example city, the authors present a cartographic exploration of the resulting typology.

Results: OZs present with immense variation across clusters. Some clusters, specifically cluster 3 and 9, are less poor, have a greater number of jobs and higher development potential than other clusters. Additionally, these exceptional clusters have disproportionate rates of final OZ designation compared to other clusters. In Portland, these less distressed clusters make up the majority of ultimately designated OZ tracts in the city and are concentrated in the downtown area compared to the more deprived eastern part of the city.

Conclusions: We find that OZ designation is disproportionately seen in particular clusters that are relatively less deprived than the larger population of eligible tracts. Cluster analysis as well as other forms of exploratory or inductive analyses can offer planners and policymakers a better understanding of their local development context as well as offering a more coherent understanding of a widely variant set of tracts.

Disciplines

Urban, Community and Regional Planning | Urban Studies and Planning

Classifying Opportunity Zones—A Model-Based Clustering Approach

Jamaal Green, University of Pennsylvania

Wei Shi, Travelers Insurance

Abstract

Objective: Opportunity Zones (OZs) are the first major place-based economic development policy from the federal government in nearly two decades. To date, confusion persists among planners and policymakers in some places as to what features of OZ tracts matter for their inclusion, and, secondly, what features of OZ tracts make them attractive targets for potential investment. The authors developed a typology of OZ tracts in order to offer planners and policymakers alternative ways of organizing a highly variable set of tracts.

Methods: This study employs model-based clustering, also known as latent class analysis, to develop a typology OZ tracts from the population of all eligible tracts in the United States. The authors use publicly available data from the US Census Bureau and Urban Institute in developing the typology. Descriptive statistics and graphics are presented on the clusters. Using Portland, Oregon as an example city, the authors present a cartographic exploration of the resulting typology.

Results: OZs present with immense variation across clusters. Some clusters, specifically cluster 3 and 9, are less poor, have a greater number of jobs and higher development potential than other clusters. Additionally, these exceptional clusters have disproportionate rates of final OZ designation compared to other clusters. In Portland, these less distressed clusters make up the majority of ultimately designated OZ tracts in the city and are concentrated in the downtown area compared to the more deprived eastern part of the city.

Conclusions: We find that OZ designation is disproportionately seen in particular clusters that are relatively less deprived than the larger population of eligible tracts. Cluster analysis as well as other forms of exploratory or inductive analyses can offer planners and policymakers a better understanding of their local development context as well as offering a more coherent understanding of a widely variant set of tracts.

Opportunity zones (OZs), the newest federal government place-based economic development tool since the New Markets Tax Credit in the early 2000s, has reportedly marshaled more than $50 billion in investment in the 2 years since its passage (Drucker and Lipton, 2019). Opportunity zones allow investors to defer taxes on their capital gains if they invest in qualified opportunity zone funds in development-starved census tracts. Recent investigations show a disproportionate amount of investment being steered into a minority of tracts that formally qualified for the program based on their income but are not suffering from a lack of development (Buhayar and Leatherby, 2019; Drucker and Lipton, 2019; Ernsthausen and Elliott, 2019).

A central tension in those articles concerning opportunity zone investment is that the Tax Cut and Jobs Act of 2017 used a broad qualifying rule for opportunity zone designation based only on tract income to maximize flexibility. It resulted in variations within designated opportunity zones in terms of their socioeconomic characteristics but also redevelopment attractiveness. An important issue for economic development researchers and analysts is to find alternative ways of organizing opportunity zones into more useful categories of analysis than simply qualified or non-qualified opportunity zone designations.

This paper presents model-based clustering, also known as latent class analysis. This unsupervised machine learning technique is one way to address the difficulties of classifying designated opportunity zone tracts. The remainder of this article will offer background on some troubling OZ issues, a description of latent class analysis through model-based clustering, and the results of cluster analysis and its relationship with opportunity zone designation. The findings contribute to a better understanding of the variation of eligible tracts and what features make the zones attractive for designation.

Background

Opportunity Zone Concerns

This article will not cover the extensive background on the opportunity zone (OZ) program design and history because it has been well documented in this issue. Still, it is important to note OZs have been particularly successful in garnering extensive investments in a short time. Early estimates showed more than $50 billion already invested in OZs through Qualified Opportunity Funds (QOFs) in 2019. Taking the COVID-19 pandemic into account, the authors suggest that OZ investment has very likely continued to grow at a healthy clip (Drucker and Lipton, 2019).

Investment numbers aside, OZs have multiple areas of concern. First, until recently, no mechanisms were available for tracking investment in OZs because such a requirement was not included in the basic legislation. The Treasury Department recently modified form 8996, which requires investors to report that they meet the 90-percent investment standard for investing in an OZ property within a QOF. Although a much-needed reform, the form 8996 data are not publicly available, but some recent work has used those data (Kennedy and Wheeler, 2021). As such, it is impossible to track exactly what QOFs are investing in and, more importantly, where such investments are going. With the revision to form 8996 and the prospect of publicly available QOF data, however, researchers may have the information necessary to better track and evaluate the program. Second, OZ designation was intentionally designed to encourage flexibility on the part of states, but the rules for designation are an income cut-off. The income rules, taken from the eligibility requirements of the New Markets Tax Credit program, specify all census tracts with a poverty rate equal to or greater than 20 percent for tracts within metropolitan areas and 80 percent of state median family income for tracts in nonmetropolitan areas. Those income rules yielded a qualified pool of approximately 31,000 tracts. Of that pool, the states could nominate up to 25 percent of eligible tracts for designation. Those broad income rules allow for a significant amount of variation within qualified opportunity zones and bring about basic questions as to how disinvestment is understood by the federal government.

This flexibility and variation of qualified opportunity zones, framed as a boon to investors seeking successful returns, also exposes opportunity zones to various inefficiencies. First is the concern that opportunity zone designation offers tax cuts for investments that would have already occurred in low-income, albeit commercially attractive, tracts. That concern is a central theme of recent journalistic pieces highlighting opportunity zone activity in major downtown areas of multiple cities that are undergoing building booms (Buhayar and Leatherby, 2019; Drucker and Lipton, 2019). This kind of program design not only robs the Treasury of potential capital gains tax revenue but also potentially siphons investment away from marginal tracts that nevertheless would be attractive targets for investment if not for the existence of exceptional low-income but development-rich tracts.

Designation and Investment Questions

Improper designation of eligible tracts creates a risk of tracts that might not need additional incentives for investment crowding out tracts that need help attracting investment. This risk is present in many place-based programs, but evidence indicates that OZs are more extreme than other programs regarding improper designation. In a recent piece, Brazil and Portier (2021) compared tract designations across four federal place-based programs: the New Markets Tax Credit, Opportunity Zone, Low Income Housing Tax Credit, and Community Development Financial Institution Fund programs. The authors found that although all four programs suffer from potential designation issues by selecting tracts already in a process of gentrification. OZ-designated tracts were nearly twice as likely to be gentrifying compared with tracts eligible in the comparator programs. In terms of investment, Kennedy and Wheeler (2021) found that in their sample of OZ returns, 84 percent of designated tracts did not receive any investment. Furthermore, they found that tracts receiving funding had generally higher incomes, educational attainment, population densities, and amenities.

Those recent works offer information on some of the potential imbalances within OZs and make better identifying and organizing potentially attractive OZs an urgent task for planners and policymakers. The rest of this paper explores developing and offering a typology of tracts.

Data and Methodology

Data

The primary dataset comprises three publicly available data sources:

- American Community Survey.
- Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics (LEHD LODES) from the US Census Bureau.
- Opportunity Zone Investment Score tool from the Urban Institute

American Community Survey (ACS) Neighborhood Deprivation Index

Using the 2011–2015 American Community Survey, the authors estimated a composite neighborhood deprivation index (NDI). The NDI developed by Messer et al. (2006) is a composite measure of material deprivation derived from the first principal component of a set of census variables. The NDI is made up of the first component of a set of 8 out of 20 census variables: share of males in management and professional occupations, share of crowded housing, share of households in poverty, share of female-headed households with dependents, share of households on public assistance, share of households earning less than $30,000 per year, share of the population earning less than a high school diploma, and share unemployed. This component is estimated using principal component analysis, a dimension reduction technique. Across the different regions Messer and her colleagues used to calibrate their measure, the first component accounted for up to 73 percent of variation. Final component scores were standardized with a mean of 0 and a standard deviation of 1. The NDI allows for a multidimensional measure of deprivation above and beyond the inclusion of only income-related variables.

LEHD LODES

Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES) Workplace Area Characteristics (WAC) data were collected for the year 2016 for all eligible census tracts (US Census Bureau, 2019). The LODES data are a mix of administrative datasets, including Unemployment Insurance, Quarterly Census of Employment and Wages filings, and additional censuses and surveys. The public data provide geographically fine data on where employees live and work. From the LODES WAC files, data were collected on employment information for all primary jobs (LEHD job type code “JT01”), aggregated to the census tract level across all available industries using the lehdr package in R (Green, Mahmoudi, and Wang, 2019; R Core Team, 2020). LEHD industry employment estimates were further reduced from 20 industry categories to 4 principal components to aid clustering.

Opportunity Zone Investment Tool

The Urban Institute developed a tool (Theodos et al., 2018) that ranks the investment attractiveness of eligible OZ tracts. The investment score summarizes how multiple investments flow into a tract on the basis of commercial, multifamily, single-family, and small business lending. Commercial and multifamily lending flows were taken from 2011–2015 CoreLogic, Inc. data on loans (single loans less than $100 million), aggregated at the census tract level. The commercial lending score is an investment-to-employee ratio calculated from an annual average of the value of the loans divided by the number of employees in a tract derived from the LODES workplace association file for all tracts with at least 200 jobs. The multifamily lending score used an annual average at the tract level divided by the number of multifamily units derived from the 2011–2015 ACS. The multifamily score was calculated for census tracts with at least 200 multifamily units. The single-family lending score used 2011–2015 Home Mortgage Disclosure Act home purchase loans averaged at the tract level divided by the number of single-family units with at least 200 single-family units. Finally, the small business lending score used Community Reinvestment Act data from 2011–2015. Similar to the other measures, this score estimates an annual average at the tract level divided by the number of small business employees in a tract. The number of small business employees was derived from the LODES WAC file for employees in firms with 19 or fewer workers. The final composite score is the average of the z-scores for each component for all eligible tracts. Finally, tracts within the same territory or state were given a decile ranking of the z-scores, for a final score of 1–10. The investment score data table also includes a “social change” flag as a rough estimate of gentrification, but it is not used in this analysis.

Methodology

Dimension Reduction Through Principal Component Analysis

Before clustering, it is often necessary to perform a dimension reduction for two reasons. First, cluster analyses can be computationally expensive in the face of many independent variables, so determining a more optimal combination of variables before clustering can save time. Second, many clustering algorithms can have highly correlated predictor variables that can degrade the performance of an algorithm.

Principal component analysis is a tool to reduce the data dimension of several interrelated variables while maximizing the variability to present the data. This step can be achieved by transforming the original variables into a new set of orthogonal variables, called principal components; each component is a linear combination of the original variables. The principal components are uncorrelated and summarize a decreasing portion of the total variance of the original data. This method is useful when the original variables are correlated, and a large portion of the data variance can be captured by the first few principal components (Shiva Nagendra and Khare, 2003). Employment data of different industry sectors for each census tract were gathered for the analysis. More than 20 industry sectors are specified in the LODES workplace association file, and most are highly correlated with each other. Thus, principal component analysis(PCA) is a promising approach to reducing the high dimensional nature of industry data compared with removing or manipulating certain industry sector variables manually.

The **psych** package in R version 4.0 is used to calculate the principal components. Varimax rotation is specified to maximize the sum of the variances of the squared loadings (the linear combination weights) to highlight a small number of important variables. This rotation technique enables each principal component to have only a small number of variables with larger loadings, whereas the rest of the variables in a component are close to zero. This step helps with the overall interpretability of the principal components.

Latent Class Analysis in mclust

Also known as "cluster analysis," latent class analysis can be broadly defined as classifying similar objects into groups in which the number and form of groups are unknown (Vermunt and Magidson, 2002). Multiple techniques, ranging from relatively simple algorithms such as k-means clustering to advanced hierarchical methods, exist.

This paper uses a model-based approach to group opportunity zone tracts into clusters based on shared attributes. Model-based approaches differ from techniques such as k-means by estimating a series of models for determining cluster membership. A model-based approach offers the analyst the following advantages over other clustering techniques. First, cluster membership is based on the predicted probability of membership as opposed to partitioning on some summary value, meaning that the membership results are less arbitrary, and the probabilistic nature of these clusters allows for the display of uncertainty of membership given model parameters. Second, because the analyst does not have to provide a preset number of clusters or classifiers, model-based approaches are truly data driven. Third, the model-based approach can take both continuous and discrete data and does not require scaling of variables (Vermunt and Magidson, 2002: 5–6).

The general form of Gaussian finite mixture models takes an estimated probability density function calculated in the mixture model for a model of *K* number of clusters:

$$f(x_i; \theta) = \sum_{k=1}^{K} \pi_k f_k(y_i \mid \theta_k)$$

θ takes the form of the parameter of the mixture model, whereas $f_k(y_i|\theta_k)$is the *kth* cluster density for observation x_i, with θ_k being the mixing probabilities in the final *K* groups (Scrucca et al., 2016). Because the form of the parameters that make up individual clusters is unknown, mixture models use a Maximum Likelihood Estimator (MLE) for calculating group membership. Although a Gaussian is assumed for basic mixture models, final clusters can take multiple shapes with differing volumes and orientations as calculated from their covariance matrices (Scrucca et al., 2016; Vermunt and Magidson, 2002). This geometric flexibility of mixture models is another advantage that allows for a wider variety of distributions to define cluster membership than basing membership on a single summary statistic or presupposing the underlying data structure. Researchers offer multiple ways to evaluate final models. The Bayesian Information Criterion (BIC) is generally used as a measure of model fit, and the model that maximizes BIC is generally considered best given the data provided. Another likelihood measure, the integrated complete-data likelihood (ICL), uses BIC as one of its terms and penalizes the initial BIC score by how much overlap exists among clusters (Scrucca et al., 2016: 9).

The **mclust** package in R version 4.0 was used to estimate clusters (R Core Team, 2020; Scrucca et al., 2016).

Results

Employment-based Principal Components Analysis

To better represent the local economic context of opportunity zone (OZ) tracts, a principal component analysis (PCA) was performed on the 20 industry sectors present in the 2016 Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES) workplace association file. The PCA with varimax rotation returned a list of principal components (RCs). The first four RCs take account of 50 percent of the variance of the overall employment structure. One can further apply those RCs to the cluster analysis to represent the economic activity of the tracts.

Exhibit 1 displays the loadings estimated from the principal component analysis. The loadings are the correlations of the variables to their principal component. The variables with correlations greater than 0.5 for their respective components are bolded to show what variables strongly affect that component (we do not include correlation estimates less than .1 for clarity). Examining these highly correlated loadings provides a better interpretation of what the loadings represent as a combination of variables.

Exhibit 1: PCA Loadings from LEHD LODES

Industries	RC1	RC2	RC3	RC4
Agriculture				
Mining	0.193	0.151	**0.519**	-0.328
Utilities		0.11	**0.805**	
Construction	0.25	**0.729**		
Manufacturing		**0.68**		
Wholesale	0.267	**0.802**		
Retail	0.361	0.363	-0.115	0.349
Transportation and Warehousing		**0.561**		
Information	**0.582**		0.123	
Finance	**0.771**		0.124	
Real Estate	**0.719**	0.276		0.167
Professional Services	**0.86**		0.181	
Management	**0.54**	0.1	0.228	-0.111
Administration and Support	**0.622**	0.447		0.117
Education			0.11	**0.642**
Health Care	0.149		0.103	**0.667**
Arts	0.239		**0.543**	0.197
Accommodation and Food	**0.645**	0.128	0.118	0.338
Other	**0.672**	0.216		0.225
Public Administration	0.158		**0.74**	0.213

Note: Loadings more than 0.5 are bolded for clarity; loadings less than .1 are excluded.
Source: Authors' calculation from the LEHD LODES WAC files

The first component entails a mix of professional, food, and other services, including real estate, management, administrative and waste support, finance, information technology, other, and food/accommodation services. This component falls into a more traditional understanding of "services" employment. The second component includes more traditional "industrial" jobs, including construction, manufacturing, wholesale/warehousing, and transportation. The third component is a more eclectic mix, including the arts, public administrative services, mining, and utilities. The final component encompasses educational and health services.

Exhibit 2: Cluster Descriptive Stats

Cluster	Number of Tracts	Designated (%)	NDI	Urban Inst. Score	RC1	RC2	RC3	RC4
1	3346	33.59%	0.75	5.4	0.162	-0.128	0.043	0.4901
2	3525	14.64%	1.10	3.5	-0.210	-0.332	-0.037	-0.3544
3	2103	40.94%	0.73	6.8	0.574	1.042	0.127	0.9451
4	4379	25.65%	0.94	5.9	-0.088	-0.018	-0.100	-0.2155
5	4577	16.91%	0.69	4.7	-0.205	-0.229	-0.034	-0.3234
6	4211	18.26%	1.00	5.1	-0.136	-0.324	-0.071	-0.2190
7	3478	32.81%	0.75	6.3	-0.024	0.538	-0.122	-0.0086
8	4916	24.92%	0.68	4.8	-0.124	-0.224	-0.013	-0.0158
9	448	45.98%	0.30	7.4	3.539	2.529	2.453	2.4006

NDI = neighborhood deprivation index. RC = Principal Component

Note: NDI, Urban Institute, and RC values are group means.

Source: Authors' calculations of ACS for the NDI, Urban Institute Investment tool and LEHD LODES WAC files

Cluster Analysis

The final dataset used to estimate the clusters includes the 30,983 eligible tracts. The data were clustered on six variables:

- The Urban Institute investment score.
- The four principal components derived from the 2016 LEHD LODES industry employment.
- The standardized score of the neighborhood deprivation index.

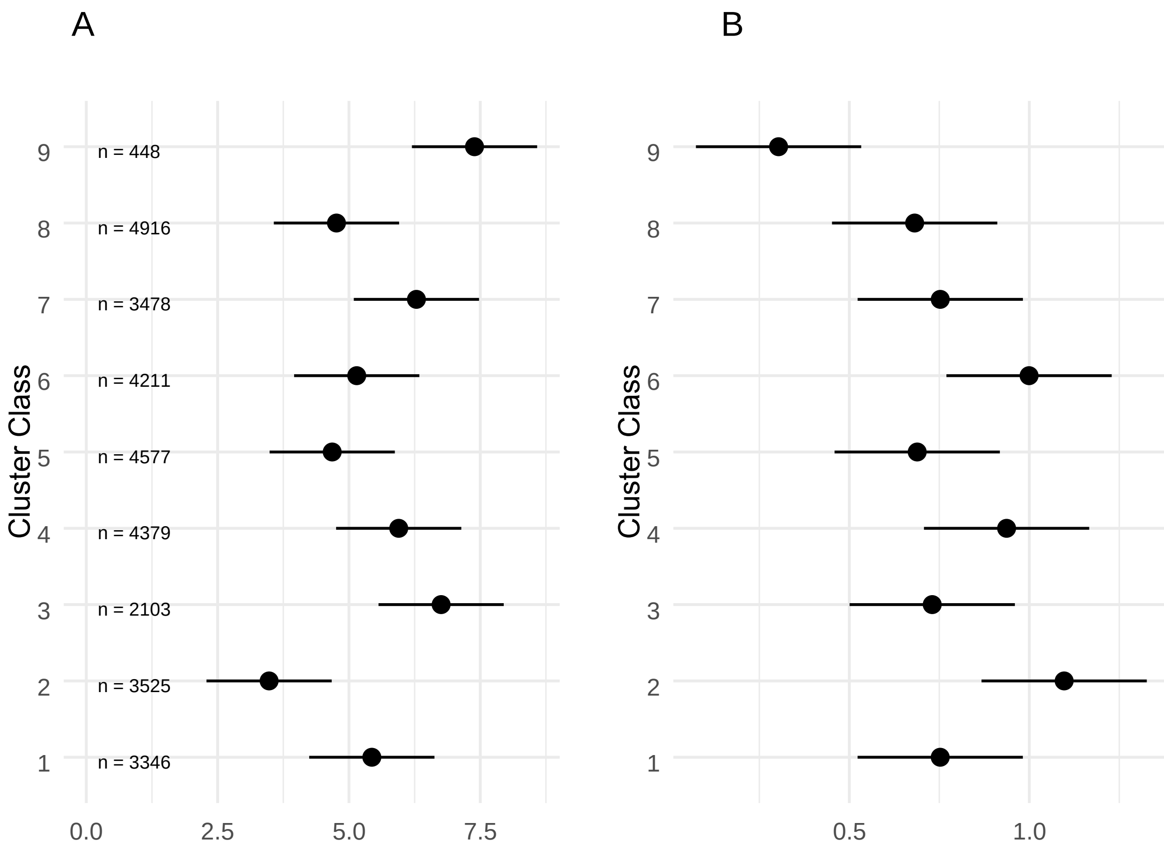
The final estimate returned nine clusters. The following section covers the features of the clusters estimated and provides descriptive results of the predictor variables from the cluster analysis and a simple statistical test on the probability of a tract being selected as an opportunity zone on the basis of cluster identification.

Comparison of the Clusters

Examining the variation in features that went into the cluster analysis across the estimated classes is one way of exploring their differences. The following plots present the means and standard deviations of the clustering variables to visualize where they differ and overlap.

Exhibit 3 shows the mean and standard deviation scores for the clusters' investment and neighborhood deprivation scores. Of particular interest are clusters 3 and 9, which exhibit larger average investment scores and lower deprivation scores. Cluster 9, in particular, has the highest average investment score and the lowest average neighborhood deprivation index (NDI) score.

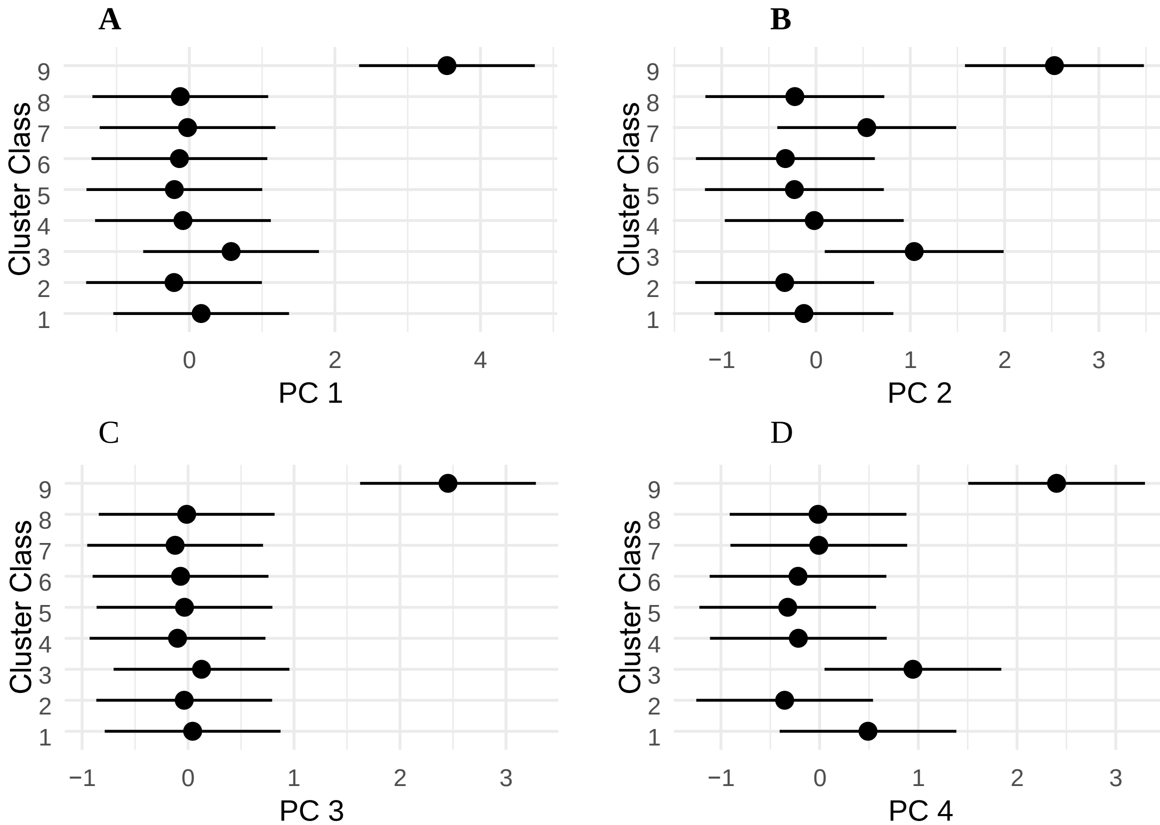
Exhibit 3: Considerable Difference and Overlap Present Themselves in the Clustering Variables



Source: Authors' Calculations

Beyond the investment and deprivation scores, the employment principal components also highlight some extreme differences across clusters, particularly for cluster 9 (exhibit 4). The employment principal components represent not just a combination of variables that one can interpret as types of employment mixes but values estimated on the counts of jobs. As such, cluster 9 and, to a lesser extent, cluster 3, have much higher principal component scores due to the high absolure number of jobs found in their respective tracts. Additional individual clusters show higher average scores for individual components. Cluster 7, for example, has greater employment in the second component. The second component is made up primarily of industrial-type jobs, such as manufacturing, warehousing, and wholesale operations. Cluster 7 also has one of the higher average investment scores and one of the higher proportions of tracts ultimately designated even with a relatively larger overall number of tracts.

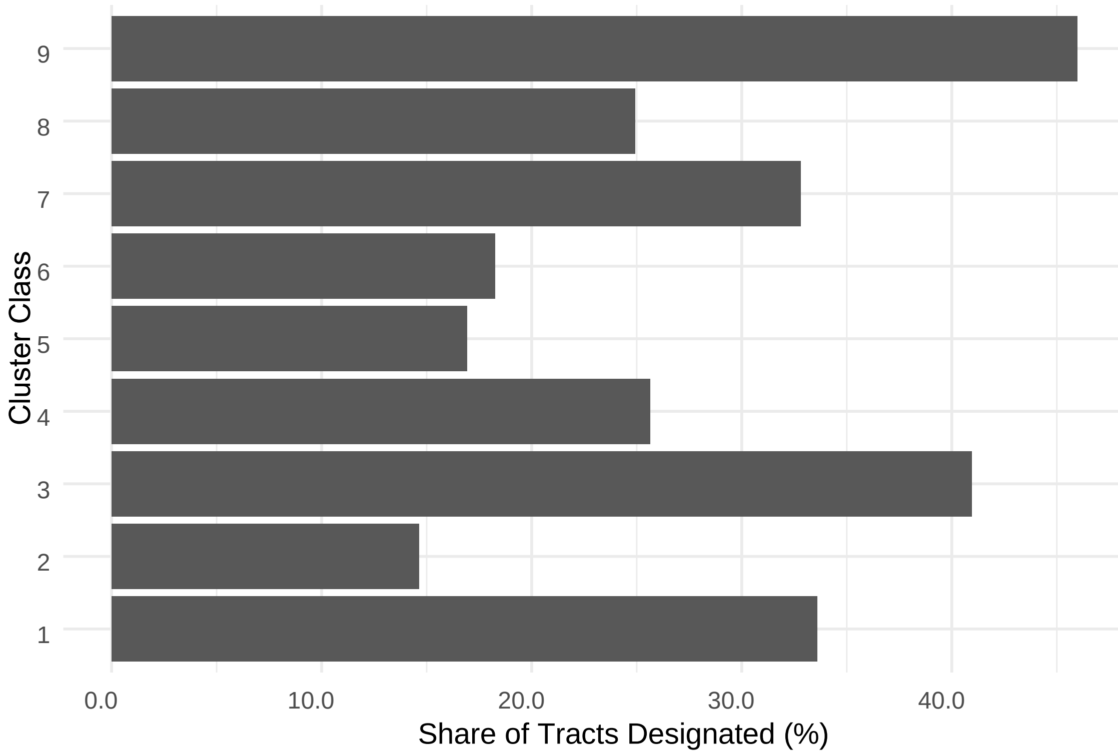
Exhibit 4: Cluster 9 Has Much Higher Employment Across All Types than Other Clusters



Source: Authors' calculations

Finally, one can examine the variation in OZ designation across the cluster. The basic selection process for designation allowed for states to designate up to 25 percent of eligible tracts. Individual clusters have widely variable shares of tracts that were ultimately designated. Across all classes, the average proportion of tracts ultimately designated as opportunity zones is approximately 28 percent, but the values range from 14.5 to 46 percent (exhibit 5). Clusters 9 and 3 have 46.0 percent and 40.9 percent of their tracts, respectively, designated as an opportunity zone, whereas only 14.6 percent of the tracts in cluster 2 were ultimately designated. As a check, a chi-square test was conducted to examine whether the distribution of designated tracts differs across the nine clusters.

Exhibit 5: OZ Designation Varies Widely Across Classes



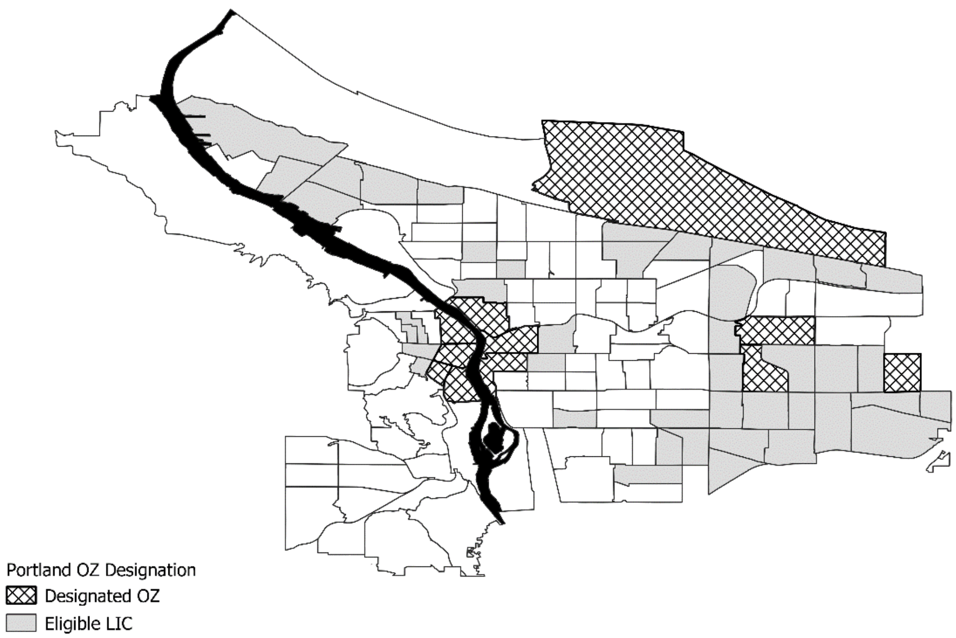
Note: Chi-square statistic is 1100.2, which is below the significance level of 0.01, indicating that the OZ designation across different clusters is statistically different.
Sources: U.S. Treasury Community Development Financial Institutions Fund; Authors' calculations

Discussion

This analysis is not the final say on a typology of tracts, but it offers a workable typology to gain a better understanding of the underlying variation of tracts and their relationship to their final designation. Clusters 3 and 9 stand out as interesting cases worthy of further exploration. Both clusters have higher average investment scores, lower neighborhood deprivation index scores (NDIs), and larger employment scores. Cluster 9 is the most extreme due to its overall smaller size but also because of the extreme average values across clustering variables. One way to explore this in more depth is through a cartographic review. The following paragraphs explore the opportunity zone geography of Portland, Oregon, to showcase how these clusters relate to social and economic geography of the city. Portland was chosen because of the authors' familiarity with and work in the region and because the city has been highlighted as a particularly extreme example of potential opportunity zone abuse (Buhayar and Leatherby, 2019).

Overall, 62 tracts were eligible for opportunity zone (OZ) designation in the city of Portland, and 11, or approximately 21.5 percent, were ultimately designated as opportunity zones (exhibit 6). The eligible tracts cover a wide array of the social geography of Portland, with the bulk of eligible tracts in the outer east part of the city. East Portland is predominately working class—it is a low-income area of the city with a large immigrant and non-White population and a growing Black population due to gentrification pressures in the inner northeastern part of the city (Gibson, 2007; Goodling, Green, and McClintock, 2015). Although an area of modest incomes, East Portland is a growing part of the city that is relatively underinvested compared with the inner east and western parts of the city. Eventual OZ designation is concentrated in the downtown/Central Eastside area of the city and a handful of tracts designated in outer east Portland—not including the tract that holds Portland International Airport, in the northeastern section of the city.

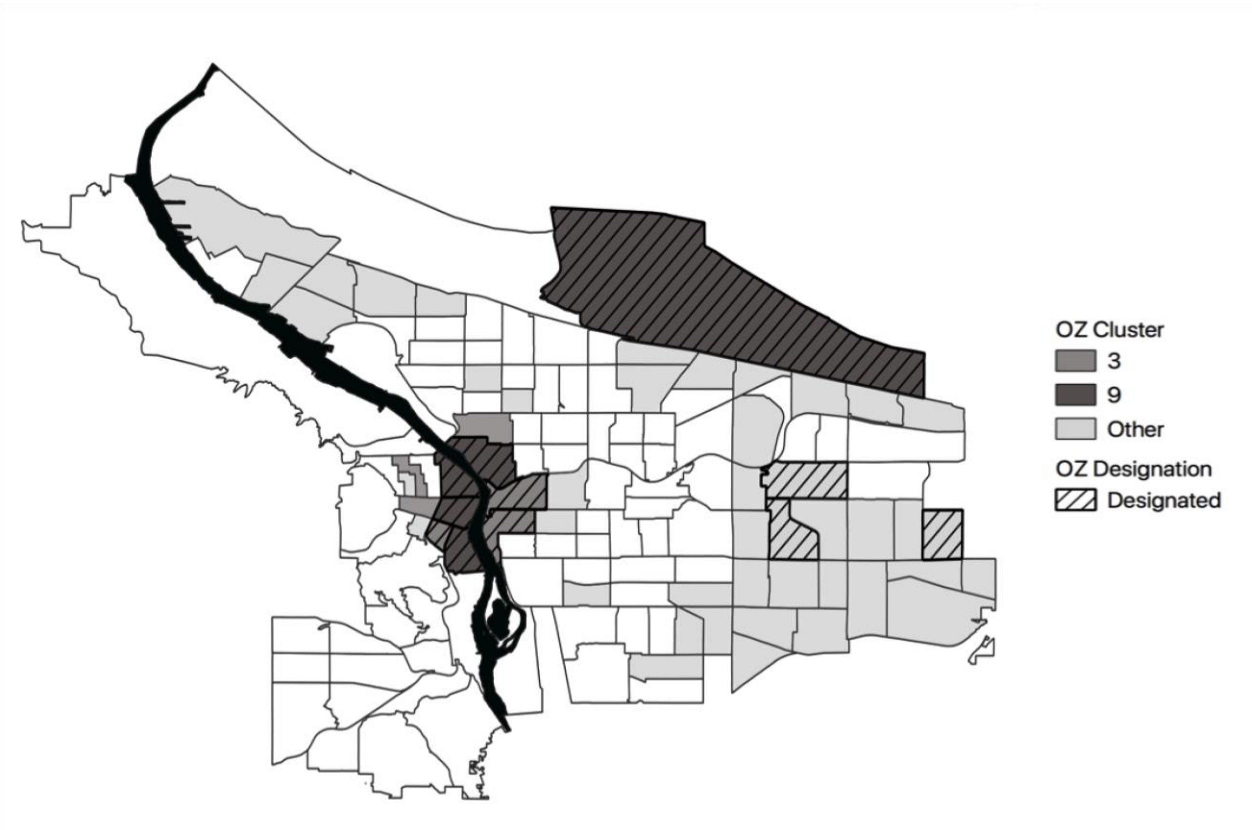
Exhibit 6: Portland OZ Designations



Source: U.S. Treasury Community Development Financial Institutions Fund

Ultimate OZ designation is not only geographically uneven in Portland, but the tracts that were selected map to the more extreme clusters, particularly clusters 3 and 9. Of the 11 designated OZs, 8, or approximately 73 percent, are in clusters 3 and 9, with 5 of the tracts in cluster 9. Comparing the proportion of eligible tracts to those that were ultimately designated across the nine clusters reveals the story from another perspective. All five eligible tracts in cluster 9 have been designated, and three out of six, or 50 percent, of the eligible tracts in cluster 3 have been designated. Only 3 out of 51, approximately 6 percent, of the eligible tracts in clusters other than 3 and 9 have been designated (exhibit 7).

Exhibit 7: OZ Designations are Uneven Across Portland's Geography



Note: It is not possible to track what investment, if any, has gone to the designated tracts in Portland due to data restrictions, but because the majority of the tracts are downtown or in the inner east parts of the city, one knows what investment is likely to flow in that direction, regardless. Recent work tracking such investments shows that investment is flowing into areas that least need additional investment (Kennedy and Wheeler, 2021).
Sources: U.S. Treasury Community Development Financial Institutions Fund; Authors' calculations

Portland serves as an example of the potential for this typology to offer planners a rough geography of investment attractiveness and OZ designation. Portland represents a more extreme case in terms of the zones ultimately designated being heavily concentrated in clusters that are high in terms of investment potential with a relative lack of material deprivation. This typology is based on a national sample of tracts but has clear, rather localized parallels.

Conclusion

To conclude, the authors developed a typology of opportunity zone (OZ) tracts based on their socioeconomic conditions at the time of designation to better understand what features are associated with designation. Using a combination of the Neighborhood Deprivation Index, the Urban Institute's OZ typology, and the first four principal components of a combination of the

Longitudinal Employer-Household Dynamics Origin-Destination (LEHD) jobs data, a nine-cluster typology of tracts using a model-based clustering approach was developed. Clusters 3 and 9 were identified as robust employment centers with relatively low neighborhood deprivation and high investment scores compared with other tracts. Those clusters also had a significantly higher proportion of their eligible tracts designated as OZs compared with other clusters. Using Portland as an example, the authors found that clusters 3 and 9, concentrated in the downtown and inner east parts of the city, made up a majority of designated tracts at the expense of tracts in the more disinvested eastern part of the city.

What should planners and policymakers take from this? First, the authors hope to offer greater conceptual clarity of the underlying attributes and structure of opportunity zones (OZs). Nearly 4 years into the program, a lack of clarity still exists at local levels about what OZs are, what they represent in terms of investment opportunity, and how to take greater advantage of them. This article cannot answer the last two concerns, but it does offer some conceptual clarity as to what OZs can represent. Secondly, this paper strives to introduce model-based clustering to planners and policymakers as an alternative approach for cluster analysis and typology production. Planners, in particular, make extensive use of varied indices and typologies, but they are often simply weighted averages or sums, calling into question their conceptual validity. Model-based clustering allows for a more flexible, defensible, and rigorous approach to clustering in an open-source framework. As government at all levels continues to focus on "data-driven" approaches and making better use of administrative and publicly available data, it is incumbent on practitioners and policymakers to be aware of tools and techniques that can maximize the impact of available information.

Planners and policymakers face a bevy of challenges with opportunity zones. The built-in ambiguity of the program has allowed tens of billions of dollars to enter various regions across the country without clear tracking or monitoring until relatively recently. The selection process of designated zones was also highly variable across the country, placing the stated goal of the policy-steering investment into disinvested areas at risk. Underneath those issues, however, lie a set of fundamental conceptual problems. What are the features of some zones that make them attractive for designation, and what are the features that make zones attractive investment opportunities?

Knowing what makes zones attractive for investment is not answerable in a straightforward fashion, especially given the lack of widely available monitoring data. With the newer reporting requirements in place, there is hope the Department of the Treasury will release investment data soon so policymakers can have a better idea of the impact and geography of these investments. Ultimately, what makes a tract attractive for investment—or not—or being designated is an inductive problem that lends itself well to the kinds of exploratory analysis presented here. A cluster analysis of this sort will not and cannot be the final word on the operation of this program, but if policymakers and planners have a structure to better organize tracts, they can better anticipate and design policies to take advantage of OZs. The question of exactly how cities are integrating—or not—OZs with existing economic development policies is an active area of research, and studies such as this one potentially offer one way for planners to better understand the development potential of their local OZs.

The combination of varied machine learning approaches and public data offers immense opportunity for planners and researchers to explore policy problems in novel ways. Cluster analyses are not technologically novel, but making planners and policymakers more aware of these tools and their potential, especially using open-source software, is one way for local governments to embrace the ongoing data revolution. Local authorities have access to a wide array of not only public but also administrative data that can be more fruitfully mined with more experience and guidance from researchers and technical experts on staff. Investment in human and technical capabilities will be a more significant issue for public authorities as society becomes ever more digitally dependent.

Opportunity zones (OZs) are likely to remain as they are for the life of the program. The program has a variety of issues that make it difficult for local officials, policymakers, and researchers to understand exactly what is going on within zones or, in the case of designation, how they were selected in the first place. Portland's OZs may give one pause in terms of how designation was decided given the immense need in the eastern part of the city compared with its downtown, but there is not a straightforward way to explore what is going with these tracts. This paper strives to show model-based clustering with a stylized example of a city that demonstrates to policymakers and planners the usefulness of exploratory data analyses, such as cluster analysis, in examining different features of ongoing policy initiatives and programs.

Acknowledgments

The authors would like to thank Dr. Liming Wang for conceptual and computing support, Jackson Voelkel for assistance in calculating the NDI and Drs. Jordan Purdy and Brian Glass from the State of Oregon Department of Human Services as early readers.

Authors

Jamaal Green is a postdoctoral fellow at the University of Pennsylvania Weitzman School of Design. Wei Shi is a data scientist at Travelers.

References

Brazil, Noli, and Amanda Portier. 2021. "Investing in Gentrification: The Eligibility of Gentrifying Neighborhoods for Federal Place-Based Economic Investment in US Cities." *Urban Affairs Review*, July 27. https://doi.org/10.1177/10780874211032637.

Buhayar, Noah, and Lauren Leatherby. 2019. "Welcome to Tax Breaklandia," *Bloomberg Businessweek.* https://www.bloomberg.com/graphics/2019-portland-opportunity-zones/.

Drucker, Jesse, and Eric Lipton. 2019. "How a Trump Tax Break to Help Poor Communities Became a Windfall for the Rich," *The New York Times,* August 31.

Ernsthausen, Jeff, and Justin Elliott. 2019. "A Trump Tax Cut Meant to Help Poor Areas Could Pay Off for Kevin Plank and Goldman Sachs Thanks to Misaligned Maps," *Baltimore Sun,* June 19. https://www.baltimoresun.com/business/real-estate/bs-bz-plank-opportunity-zone-20190612-story.html.

Gibson, Karen J. 2007. “Bleeding Albina: A History of Community Disinvestment, 1940–2000,” *Transforming Anthropology* 15 (1): 3–25. https://doi.org/10.1525/tran.2007.15.1.03.

Goodling, Erin, Jamaal Green, and Nathan McClintock. 2015. “Uneven Development of the Sustainable City: Shifting Capital in Portland, Oregon,” *Urban Geography* 36 (4): 504–527. https://doi.org/10.1080/0272 3638.2015.1010791.

Green, Jamaal, Dillon Mahmoudi, and Liming Wang. 2019. Lehdr. https://jamgreen.github.io/lehdr/articles/getting_started.html.

Kennedy, Patrick, and Harrison Wheeler. 2021. “Neighborhood-Level Investment from the US Opportunity Zone Program: Early Evidence,” April 15. https://static1.squarespace.com/static/57a3c0fcd482e9189b09e101/t/607893b915858d7bd0d198ba/1618514881004/oz_kennedy_wheeler.pdf.

Messer, Lynne C., Barbara A. Laraia, Jay S. Kaufman, Janet Eyster, Claudia Holzman, Jennifer Culhane, Irma Elo, Jessica G. Burke, and Patricia O’Campo. 2006. “The Development of a Standardized Neighborhood Deprivation Index,” *Journal of Urban Health* 83 (6): 1041–1062. https://doi.org/10.1007/s11524-006-9094-x.

R Core Team. 2020. “R: A Language and Environment for Statistical Computing.” Vienna, Austria: R Foundation for Statistical Computing. https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing.

Scrucca, Luca, Michael Fop, T. Brendan Murphy, and A.E. Raftery. 2016. “mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models.” *The R Journal* 8 (1): 289–317. https://journal.r-project.org/archive/2016/RJ-2016-021/RJ-2016-021.pdf

Shiva Nagendra, SM, and Mukesh Khare. 2003. “Principal Component Analysis of Urban Traffic Characteristics and Meteorological Data,” *Transportation Research Part D: Transport and Environment* 8 (4): 285–297. https://doi.org/10.1016/S1361-9209(03)00006-3.

Theodos, Brett, Carl Hedman, Brady Meixell, and Eric Hangen. 2018. *Opportunity Zones: Maximizing Return on Public Investment.* Washington, DC: Urban Institute. https://www.urban.org/policy-centers/metropolitan-housing-and-communities-policy-center/projects/opportunity-zones-maximizing-return-public-investment

US Census Bureau. 2019. “LEHD Origin-Destination Employment Statistics Data (LODES) (2002–2015).” Washington, DC: US Census Bureau.

Vermunt, Jeroen K., and Jay Magidson. 2002. “Latent Class Cluster Analysis.” In *Latent Class Cluster Analysis*, edited by J. Hagenaars and A. McCutcheon. Cambridge, England: Cambridge University Press: 89–106.