Implications of Coarse Data Allocation Methods for Flood Mitigation Analysis

James Howard*

Abstract
Efforts to perform fine-grained analysis are often hampered by data provided by government agencies that does not reflect appropriate granularity. Coarse-grained government data may reflect the data collection methods, strategies, or may reflect the reality of what the data represents, and cannot be made more granular. For example, the United States Flood Mitigation Assistance (FMA) grant program makes grants to both state and local governments. By employing data on the FMA program, this analysis examines allocation strategies for coarse data. Between 1996 and 2011, approximately one-quarter of FMA grants were given to state governments with the remaining three-quarters given to local governments. Performing a local-level analysis of the impacts of these grants requires an allocation method that fairly reflects the local impact of statewide grants. This note considers several allocation strategies and how these strategies affect the implementation and interpretation of statistical models for public policymaking.

Key Words: data science, coarse data, statistical allocation, missing data

1. Introduction

Efforts to perform fine-grained analysis are often hampered by data provided by government agencies that do not reflect appropriate granularity. Coarse-grained government data may reflect the data collection methods, reflect the reality of what the data represents, or be intentionally introduced (Heitjan & Rubin, 1991). Breaking this data into presumed constituent components in a consistent and meaningful manner is necessary if other data in the analysis is available at finer-grained detail and aggregation of finer-grained detail is not preferred. This can allow fine-grained impacts to be tracked against the coarse-grained results and *vice versa*. However, there may not be underlying constituent components to extract from the data.

This problem is similar to the problem of data disaggregation (Garrett, 2003). However, aggregated data is assembled from finer-grained data. That may be aggregated at the temporal level (Chan, 1993), the geographic level (You, Wood, & Wood-Sichra, 2009), or be the aggregation of similar measures. But each of these presents data for which the underlying properties can be estimated and approximate disaggregations are possible. In contrast, data may not be the sum of constituent component values. Financial and economic statistics are examples of figures that may be inherently coarse at both the temporal and geographic level. Disaggregation strategies are not necessarily appropriate for this type of data.

For example, the United States FMA grant program makes grants to both state and local governments (King, 2005). By employing data on the FMA program, this analysis examines allocation strategies for coarse data. Between 1996 and 2010, approximately one-quarter of FMA grants were given to state governments with the remaining three-quarters given to local governments. Performing a local-level analysis of the impact of these grants requires an allocation method that fairly reflects the local impact of statewide grants. This analysis considers several allocation

*Kore Federal Inc.

strategies and how these strategies affect the implementation and interpretation of statistical models for public policymaking.

2. Data

This analysis uses information from the Federal Emergency Management Agency (FEMA) on the FMA grant programs for the period from 1996 through 2010. The FMA dataset includes includes 2107 grants made under the FMA, Severe Repetitive Loss (SRL), and Repetitive Flood Claims (RFC) programs. These three programs are different For each grant, the data includes the state, county, subgrantee (an agency or city receiving the grant), a grant program identifier, the year, and the amount of the grant. However, for 1081 of these grants, the county is not provided. For 516 grants, the county can be resolved by identifying which local agency, often a municipal government, received the grant. The remaining 565 grants cannot be assigned to a specific county, and at 27 percent of the total grants, this missing data obstructs effective analysis at the county-level.

3. Methods

One approach to manage these grants is to allocate them across the state. This would allow impacts to be measured in a coherent way. This analysis uses four distinct allocation strategies based on demographic data and considers the results of each. These methods are designed allocate the statewide grants across the counties or other first-tier subdivisions within the state in systematic, reproducible, and justifiable ways. These methods are:

1. The first strategy to consider is the no allocation strategy. In this strategy, statewide and allocable grants are simply removed from the dataset and not used for analysis. When these grants are removed from the dataset, there is no ability to consider any impact those grants may have. Nevertheless, statewide grants may have long-lasting and far-reaching impacts if they paid for multijurisdictional engineering projects or contributed to statewide planning improvements. But statewide planning improvements are difficult to measure. This is considered the base case.

2. The next strategy considered is to allocate all statewide and unallocable grants within a state equally amongst all counties within that state. This strategy assumes that the grants have an equal impact across the state, regardless of how the population is distributed within the state. This also assumes that the flood risk is geographically constant across the state. Nevertheless, the simplicity of this allocation strategy leaves it easy to implement.

3. The next strategy to investigate is to allocate the unallocable grants by the median income of each county within. The median income is used in this strategy to approximate the distribution of wealth within the state. The basis for this is the presumption that as wealth increases, so does the relative share of the tax base. Therefore, each contributes tax revenue to the grant in a share proportional to the amount of impact returned.

4. The final strategy tested is to allocate the unallocable grants by the population of each county. This is designed to be an equitable treatment of the unallocable

grants and it gives each person within the state equal standing in any analysis using the data. With this, more populous counties are presumed to benefit more from the grant money, but the underlying assumption is that states will spend the money around the state proportionally to where the population is. This is not necessarily a valid assumption.

These four strategies are four simple allocation strategies that can be implemented with readily available demographic data, both base population and the median income, along with basic information about a state's geography. Together, these methods provide four different paths that can impact the final analysis, depending on analytical methods used and objects of the analysis. Investigations into the impact of income or population will obviously be affected by distributions based on each, for instance.

4. Results

From a qualitative perspective, the impacts on states are mapped by showing the intensity of the grants distributed across two sample states. Because of the density of the maps, only two will be looked at in detail. One is Florida and the second is Maryland. Heatmaps for the grant distribution, by county, for Florida and Maryland are given in figures 4 and 4, respectively, for each of the allocation strategies discussed in section 3.

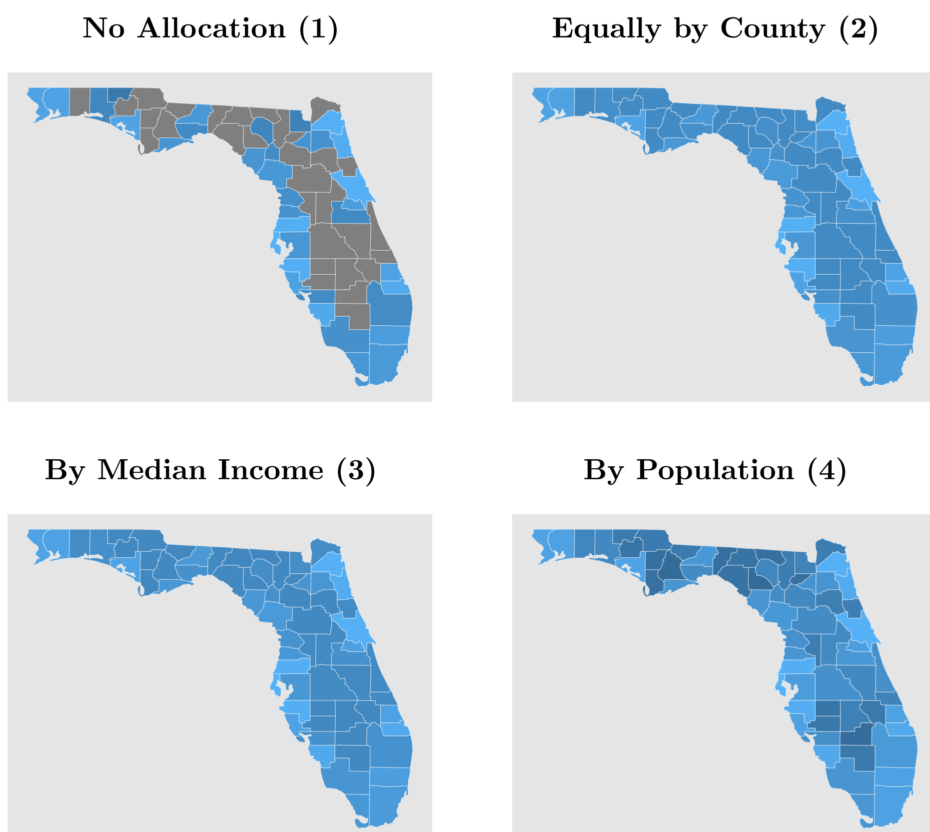

Figure 1: Map of Florida counties by grant allocation strategy

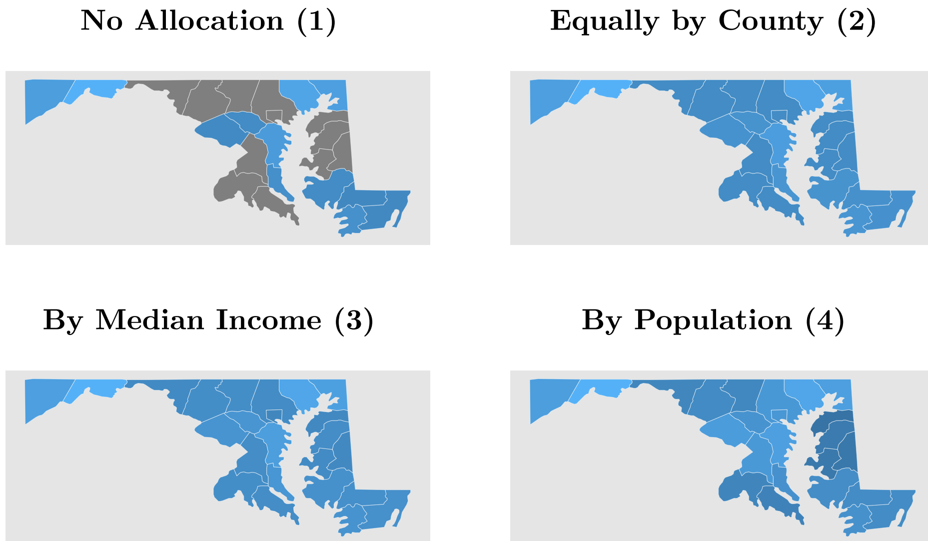

Figure 2: Map of Maryland counties by grant allocation strategy

5. Discussion

Allocation strategy is not meaningless. As is evident in the maps of Florida and Maryland, there is substantial difference in the final amounts of grants awarded to a county depending on the allocation strategy employed. If allocating equally or by median income, the maps show little difference. But allocation by population shows a different grant pattern. Underlying these allocation strategies are questions of interpretation and applicability.

The strategy that eliminates grants given to non-local agencies would be appropriate for statistical models that presume the statewide grants have little or no local impact. The strategy would also be appropriate if a statistical model includes population and income figures as explanatory variables, since this allocation strategy is not dependent upon that information. This strategy would not be usable of the model intends to capture multijurisdiction impacts alongside direct impacts from locally-distributed grants.

The strategy to allocate the grants equally by county is also appropriate if a statistical model includes population or income figures. But it also presumes there is a baseline level of impact happening to local outcome measures due to the grants given statewide. This is a generalization of the elimination strategy which simply assumes the baseline level of impact is 0. This strategy would not be appropriate for statistical models that include local government features, such as organizational types or policies, as explanatory variables.

The strategy is allocate grants by median income is appropriate if an applied statistical model does not include wealth or a wealth proxy as an explanatory variable. Because the median income allocation strategy is designed to allocate based on a wealth proxy, this approximates a distribution by wealth. This distribution, however, is imperfect for this purpose since income and assets may not correlate sufficiently well and the median income only represents the most middle income earner. However, county-level statistics for median income are readily available from the United States Census Bureau (USCB). Because many interesting economic models use wealth and wealth proxies as explanatory variables, this grant allocation strategy is unlikely to be practical.

The final strategy, to allocate statewide grants by population is an appropriate allocation method if applied statistical models do not include population as an explanatory variable. This limitation, however, does not directly limit the amount of *per capita*-type variables that can be used as explanatory variables. Using population-based allocation is also appropriate when looking at explanatory variables that do not correlate with population. In the case of the FMA grants, this may involve weather-related variables or governmental organization variables.

6. Conclusions

This note considered the use of different allocation strategies for FMA grant data to determine the impact and appropriateness of those strategies for different analytical questions. An allocation strategy is appropriate for a particular analysis when the underlying allocation function is not also included in a statistical model. Otherwise, endogeneity is likely to arise leading to invalid statistical models. In addition, it is important use a replicable and documentable allocation strategy so future research and repeat the analysis.

This note, though, only looks at the applications to the FMA grants program. Other grant programs offer grants to state agencies for local use. In addition, non-grant programs, such as programs that provide technical assistance to state and local governments also present difficulties for allocation of non-local resources for measurement. Because of these widespread applications, there are implications for coarse data allocations across multiple policy domains (*e.g.*, Janssen and Sklar (1998)).

Like disaggregation, the coarse data allocation problem presents data that seems structured to an underlying context. But, unlike data that requires disaggregation, coarse data may not necessarily have the underlying structure that is present in aggregated data. The FMA data presents additional data that has the structure already applied, depending on whether a grant has already been allocated to a particular local jurisdiction. Additional allocation strategies that use this information to supply allocation factors are theoretically possible.

References

Chan, W.-S. (1993). Disaggregation of annual time-series data to quarterly figures: a comparative study. *Journal of Forecasting*, *12*(8), 677–688.

Garrett, T. A. (2003). Aggregated versus disaggregated data in regression analysis: implications for inference. *Economics Letters*, *81*(1), 61–65.

Heitjan, D. F., & Rubin, D. B. (1991, December). Ignorability and coarse data. *Annals of Statistics*, *19*(4), 2244–2253.

Janssen, G., & Sklar, M. (1998, September 16). *Geographic allocation of state level nonroad engine population data to the county level* (No. NR-014). United States Environmental Protection Agency. Washington.

King, R. O. (2005, June 30). *Federal flood insurance: the repetitive loss problem.* Congressional Research Services. Washington, D.C.

You, L., Wood, S., & Wood-Sichra, U. (2009). Generating plausible crop distribution maps for sub-saharan africa using a spatially disaggregated data fusion and optimization approach. *Agricultural Systems*, *99*(2-3), 126–140.