

The Effects of Consultant Use on Environmental Justice Planning

Aaron Guerra

Introduction

In 2016, the state of California passed Senate Bill 1000, which required that certain general plans address environmental justice (EJ), in addition to previously required elements including housing, land use, safety, and others. Previous research suggests that these plans vary greatly in their content and that further research should seek to explain “the context in which plans were made [...], including the use of consultants” (Brinkley and Wagner 2024, 68). This article intends to do just that, by questioning what effect the use of consulting firms has on the similarity between environmental justice elements. A greater understanding of how consultants impact the policy development process can have normative implications for the policies that are developed in the EJ context, which prioritizes localized, participatory development over centralized or standardized approaches. Further, consultants are frequently used in the environmental policy process, particularly in the environmental review process, and results may indicate further avenues of research.

Here, I focus on a single hypothesis regarding the use of consultants. Previous research on boilerplate text has indicated that documents sharing consulting firms have increased use of boilerplate text (Scott, Marantz, and Ulibarri 2022). Further, the use of consultants is often intended to reduce friction in the review process through normalized and standardized processes (Kingsley et al. 2017). As such, I hypothesize that the use of consulting firms in the development of EJ elements will lead to increased similarity between documents.

Methods

The General Plan Database Mapping Tool was used to conduct this analysis (plansearch.caes.ucdavis.edu). This database was constructed through manual compilation of relevant general plan documents

across California cities, and is primarily used as a website which allows users to search for a term across all plans available in the database (Banginwar et al. 2023; Poirier et al. 2024). The database has been updated several times since its initial use in a 2017 article [CITE]. Documents were manually downloaded from the website and totaled 361 unique documents at the time of download, covering 288 unique cities.

Documents were transformed from the pdf format to text corpuses using the `pdftools` and `quantda` packages in R. Whitespace characters such as newline (`\n`) were removed but documents were otherwise unaltered. After an initial review of documents it was necessary to tag the specific type of general plan document for each, as the dataset included general plans, housing elements, environmental justice elements, and safety elements. For each document, the first two pages were searched for text specifying which element was listed, such that if “Environmental Justice Element” was on the first page, the document would be tagged as such. For the documents which could not be automatically tagged for some reason, as well as the subset used in this analysis for additional verification, the author manually provided and verified the plan type tags. One limitation of this approach is that EJ elements contained within General Plans, such as the 2011 National City General Plan, are excluded from analysis here.

In total, 13 environmental justice elements were identified, each representing a different California city. These include Antioch in Contra Costa county, Arvin in Kern county, Gardena, Inglewood, Lynwood, Monrovia, and Maywood in Los Angeles county, Menlo Park in San Mateo county, Oakland in Alameda County, Perris and Riverside in Riverside County, Rialto in San Bernardino County, and San Juan Capistrano in Orange county. While these are not the only documents in the dataset that address environmental justice, they are selected due to their topical focus on *only* environmental justice, as opposed to environmental justice in the context of housing or safety.

Dependent Variable

The dependent variable in this analysis is the similarity between two documents. Several pre-processing steps were taken prior to this calculation. After tokenization, a subset of tokens were removed, including city and county names, stopwords, letters, and the terms “st”, “ave”, “rd”, and “blvd”, followed by the removal of punctuation, symbols, and numbers, and finally word stemming. These choices were made after initial analysis highlighted they are among the most common terms in some documents but are substantively unimportant on their own. However, sensitivity testing indicates that results are somewhat robust regardless of their inclusion or exclusion, with the exception of removing city and county names. This is likely due to another pre-processing choice: the weighting of a document feature matrix (DFM) with term frequency - inverse document frequency (TF-IDF) values.

TF-IDF values weight the features (or terms) in a DFM based on their document-level frequency and corpus-level frequency, such that terms which are uniquely frequent in few documents are more heavily weighted than terms which are frequent across all or many documents. As such, this weighting procedure transforms the frequency of a term into the relative frequency of a term. This procedure was selected for this analysis due to the expectation that there would be significant similarity between documents as they are intended to be similar. There are statutory requirements for EJ elements to cover certain topics, including pollution exposure, food access, and civic engagement, among others (Governor’s Office of Planning and Research 2020). TF-IDF weighting provides a consistent and understandable method of controlling for this expectation. Finally, any terms which only appear in one document are removed, as the nature of comparison between two documents would be biased by terms which appear in one document but not the other.

Given the resultant DFM, cosine similarity was calculated between each pair of documents, for a total of 78 pairwise comparisons. Cosine similarity measures the cosine between two vectors in a high-dimensional space, where each vector is equal in length to the number of features in a DFM. Higher normalized cosine measurements indicate higher similarity between documents.

Independent Variables

The primary independent variable of analysis is the use of consultants on a given EJ element. This was developed through a manual content analysis by the author. Use of consultant was defined as an explicit reference within the document typically labeled using the phrase “Document prepared by” or the prominent display of a consulting firm logo. The actual consulting firm was recorded but as no two EJ elements shared the same firm, the independent variable is operationalized as two elements both having used a consultant.

Two sets of controls were used for this analysis. First, the thirteen plans are distributed across eight counties, meaning that several of the plans exist in the same county - there are five in Los Angeles county and two in Riverside county. As such, the county within which a city is bounded is included in some models. Similarly, the plans in this dataset were developed between 2021 and 2024, and the year of plan development is included in some models. Second, two socioeconomic variables were gathered from the 2020 American Community Survey, measuring the population and median household income (MHI) of the cities measured. These were collected from the Census API using the `tidycensus` R package. For means of comparing population and MHI, these variables were binned into two groups about the median, thereby measuring if a community is relatively large or small and poor or wealthy.

Results and Discussion

Given the nature of the data, this analysis uses ordinary least squares regression (OLS). Initial inspection suggests that non-normal distribution of error terms is remedied by transforming the dependent variable into its natural logarithm. Four models are run: the first with only consultant use, the second with consultant use controlling for plan-level variables, the third with consultant use controlling for socioeconomic variables, and the fourth with all controls. The results of this regression analysis are presented in Table 1.

Table 1: Predictors of Logged Cosine Similarity

	<i>Dependent variable:</i>			
	log(cosine)			
	(1)	(2)	(3)	(4)
Consultant Used	0.357** (0.168)	0.340* (0.173)	0.384** (0.166)	0.368** (0.171)
Same County		−0.080 (0.221)		−0.052 (0.221)
Same Year		0.073 (0.185)		0.108 (0.183)
Same Population Bin			−0.203 (0.148)	−0.196 (0.152)
Same MHI Bin			0.219 (0.148)	0.232 (0.151)
Constant	−2.351*** (0.087)	−2.351*** (0.104)	−2.366*** (0.131)	−2.387*** (0.142)
Observations	78	78	78	78
R ²	0.056	0.060	0.109	0.114
Adjusted R ²	0.043	0.022	0.073	0.053
Residual Std. Error	0.660 (df = 76)	0.667 (df = 74)	0.649 (df = 74)	0.656 (df = 72)
F Statistic	4.499** (df = 1; 76)	1.574 (df = 3; 74)	3.019** (df = 3; 74)	1.860 (df = 5; 72)

Note:

*p<0.1; **p<0.05; ***p<0.01

In each of the four models, there is evidence that the use of consultants in the development of an EJ element is correlated with increased cosine similarity. This effect remains mostly unaffected by the inclusion of further controls, none of which significantly influence the cosine similarity.

However, there is also reason to be skeptical about the models including controls, as the associated F statistic is not significant. Combined with the lack of significance for any of the control variables, the simplest first model is perhaps the most reliable. Taking the coefficient from that model, these results suggest that when two EJ elements both use a consultant in their development, the similarity between them increases by 43.6%.

Given these results, one may wonder what is different about the documents produced with a consultant compared to those which were produced without. One approach to understanding these differences is through the frequency with which terms appear in the two groups. Recall that term counts have been weighted using TF-IDF scores, such that the most frequent terms are better interpreted as the most relatively frequent term (in that document). Figure 1 shows the most common twenty words in each group, highlighting those which appear in both.

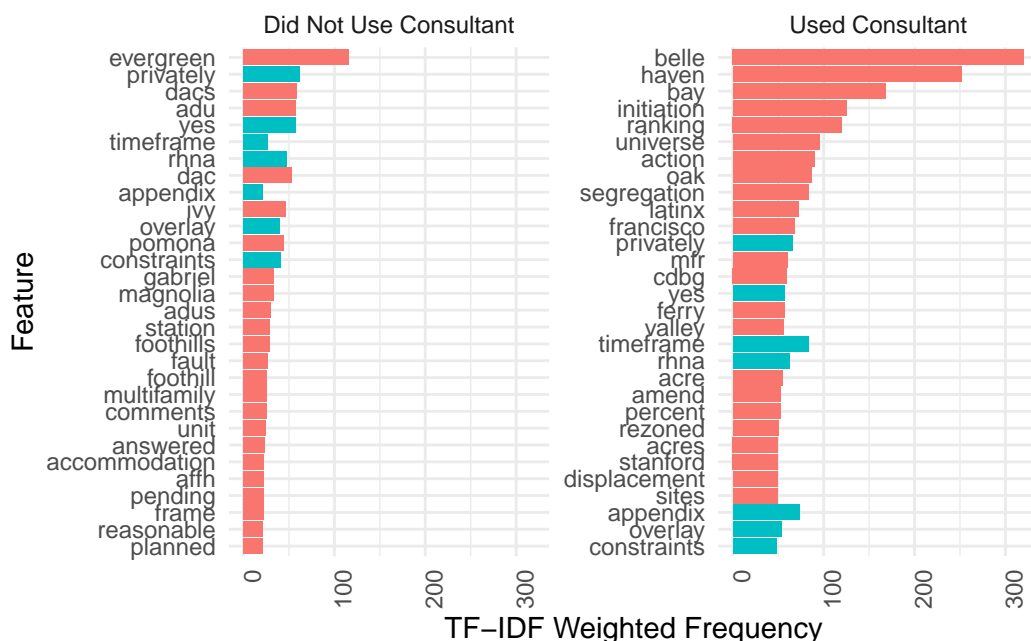


Figure 1: Weighted Term Frequency in Environmental Justice Elements

While this sample is small, there are a few intuitive points which can be gleaned from this approach. First, there are several geographic locations which contribute to the similarity measures that should be removed as research progresses - evergreen, pomona, magnolia, and belle haven are either neighborhoods or locations which uniquely identify a certain city, and would ideally be removed. Second, there are a small number of words which appear frequently in both consulted and non-consulted plans, including RHNA, or the Regional Housing Needs Allocation, which measures the need for housing in a community, which may indicate that regardless of consultant use, this metric is used to provide for one form of environmental justice. Finally, while Brinkley and Wagner (2024) found that discussion of race was not frequent in EJ plans, this analysis highlights

that both the terms “latinx” and “segregation” appear frequently in consulted plans. While certainly not causal, this suggests that consulted projects are perhaps more likely to include race. Further research would need to investigate this across a broader set of plans, potentially also controlling for racial or ethnic sub-populations within a city.

Conclusion

This research has made initial progress towards an understanding of how novel topics of planning emerge. I show that two planning elements which are both written by consultants are significantly more likely to be similar than other pairwise comparisons. In this analysis, I make best efforts to limit the corpuses being compared such that they highlight substantively unique features, rather than local details or generic, frequently used terms. However, there are two directions of further effort towards a greater understanding of the role of consultants on planning documents.

First, this research could be performed in greater depth. Here I highlight a correlation between consultant use and similarity, and suggest some of the terms that may be unique to consultants, but there is no proper understanding of what consultants are actually doing differently. Developing hypotheses based on the literature about how consultants act, what tasks they are assigned, and how the complete these would provide one step in this direction. This hypotheses could be tested by moving from term-based models of documents to topic-based models, where semantic meaning is measured rather than the frequency or similarity of various terms. Such an approach would likely require a more atomic view of each document to compare if, when, and where various topics are discussed.

Second, this research could be conducted with greater breadth. Here, the sample reflects a best initial effort at the universe of environmental justice elements, but there EJ elements which are buried within general plans not analyzed here, as well as potential documents not included in the dataset used. Further, cities have the option of developing a standalone EJ element or integrating the necessary materiel within other elements. Afforementioned topic-based models may be similarly useful to parse out where and how city planning documents more broadly discuss EJ and allow for comparison between documents where there is a greater expectation of dissimilarity. Further, a broader dataset would likely include plans which use the same consultant, allowing for an additional level of control in the models.

Works Cited

- Banginwar, Aniket, Dexter Antonio, Mirthala Lopez, Lindsay Poirier, Sujoy Ghosh, Makenna Dettmann, and Catherine Brinkley. 2023. *General Plan Database Mapping Tool*. Zenodo. <https://doi.org/10.5281/zenodo.7508689>.
- Brinkley, Catherine, and Jenny Wagner. 2024. “Who Is Planning for Environmental Justice—and How?” *Journal of the American Planning Association* 90 (1): 63–76. <https://doi.org/10.1080/01944363.2022.2118155>.
- Governor’s Office of Planning and Research. 2020. “General Plan Guidelines, Chapter 4: Required Elements.” https://lci.ca.gov/docs/20200706-GPG_Chapter_4_EJ.pdf.
- Kingsley, Gordon, Dan Matisoff, Juan Rogers, Baabak Ashuri, Yehyun An, Evan Mistur, Isabel Ruthotto, Darian Agnew, and Georgia Institute of Technology. 2017. “Strategies for Communicating Quality Expectations for Environmental Service Contracts : Final Report.” <https://rosap.ntl.bts.gov/view/dot/31977>.
- Poirier, Lindsay, Dexter Antonio, Makenna Dettmann, Tiffany Eng, Jennifer Ganata, Sujoy Ghosh, Mirthala Lopez, Ranesh Karma, Asiya Natekal, and Catherine Brinkley. 2024. “Making Plans Findable, Accessible, Interoperable, and Reusable with Data Infrastructure: A Search Engine for Constructing, Analyzing, and Visualizing Planning Documents.” *Environment and Planning B: Urban Analytics and City Science* 51 (8): 1913–30. <https://doi.org/10.1177/23998083241227471>.
- Scott, Tyler A, Nicholas Marantz, and Nicola Ulibarri. 2022. “Use of Boilerplate Language in Regulatory Documents: Evidence from Environmental Impact Statements.” *Journal of Public Administration Research and Theory* 32 (3): 576–90. <https://doi.org/10.1093/jopart/muab048>.