Scientific Proposal Part 2

NS50: Empirical Analysis

Prof. Moy

December 3, 2022

Background and Justification

Cities have put effort into mitigating the emission of greenhouse gases to fight climate change. However, the development of protective measures for public health against climate-change-related emergencies has been neglected (Wolff & Comerford, 2018).

Climate-change-related Sea Level Rise (SLR) is a growing concern for coastal cities, which, in addition to anomalous precipitation patterns, represents a negative health impact for its populations.

This paper will explore data (Wolff & Comerford, 2018) collected by the San Francisco Department of Public Health (SFDPH) to understand the city's vulnerability to floods and extreme storms and evaluate its relation to Emergency Medical Services (EMS) response time for vulnerable populations.

Exploratory Data Analysis

SFDPH's dataset is the result of their investigation and vulnerability assessment of the health risk against the increased prevalence of flooding and extreme storms in San Francisco. The dataset provides tabular data for each of San Francisco's *Census Blockgroup* and vulnerability indicators, such as *Poverty* (measured as the proportion of individuals below 200% of the Federal poverty rate) or *Diabetes* (measured as the age-adjusted hospitalization rate per 100,000 residents due to diabetes, with age ≥ 18). *Table 1* shows a random sample (with $n = 5$) with the columns of *Census Blockgroup*, *Poverty*, *Homeless* and *NonWhite*.

Census Blockgroup	Poverty	Homeless	NonWhite
60759809001	0.357	17.503	0.599

60750201001	0.510	15.892	0.651
60750117001	0.379	4.937	0.697
60750178021	0.327	19.268	0.584
60750125012	0.795	14.618	0.726

***Table 1**: A sample from SFDPH's dataset (with $n = 5$) showing only the columns of Poverty (proportion of all individuals below 200% of the Federal poverty rate), Homeless (homeless population, per 1000 residents), and NonWhite (proportion of residents that do not identify as white nor hispanic).*

Each row from SFDPH's dataset contains a unique *Census Blockgroup*, corresponding to a *geo-id* attached to locations in San Francisco. *Figure 1* shows the cross-matching of SFDPH's dataset and geographic data provided by the Metropolitan Transportation Commission (MTC).

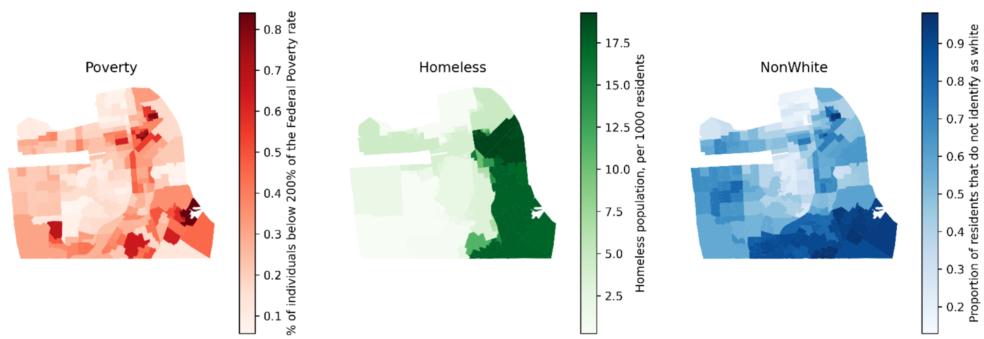


***Figure 1**: Geographical distribution of SFDPH's Poverty, Homeless and NonWhite columns. Note how in all cases, the distribution is not uniform, and the more intense colors illustrate the formation of clusters.*

Figure 1 shows that the geographical distribution of vulnerability indicators is not uniform and presents clusters (i.e., locations or sets of neighboring locations that present higher-than-average vulnerability indicators).

San Francisco's sea-level projections indicate the range of global SLR will be anywhere from 50 - 140 cm (Rahmstorf, 2007) and 80 - 200 cm (Pfeffer et al., 2008) by 2100. The previous is relevant since some vulnerability indicators would also increase along with the sea level; for example, the proportion of all individuals below 200% of the Federal poverty rate (*Poverty*) would increase with the worsening of SLR.

As mentioned by Trowbridge et al. (2009), minimizing Emergency Medical Services (EMS) response time is a central objective of pre-hospital care. Conversely, it is known that patients who experience higher ambulance diversion (i.e., patients being moved from one hospital to another) present higher mortality rates (Hsia et al., 2011). Similarly, per Hsia et al. (2018) and Hsia et al. (2012), it is known that patients of color and those experiencing financial insecurity, respectively, experience higher ambulance diversion and, therefore, higher mortality rates.

Figure 2 shows the statistical and geographical distribution of *Poverty*, *Homeless,* and *NonWhite* as assessed by SFDPH's dataset; this figure also presents arbitrarily chosen thresholds for each vulnerability indicator to facilitate the identification of the clusters mentioned previously.

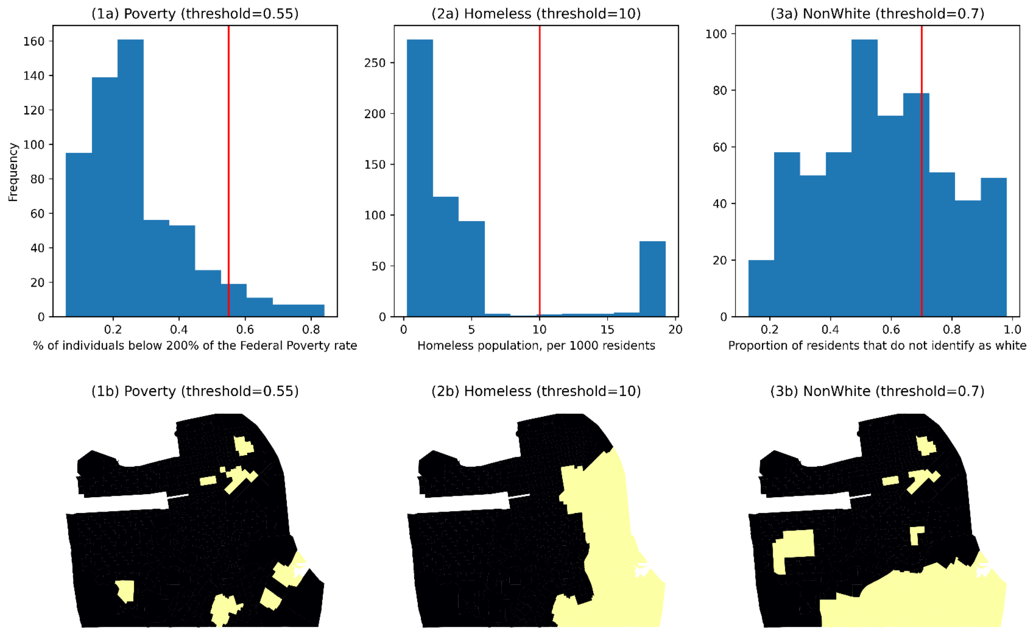


Figure 2*: Statistical and graphical distribution of SFPDH's Poverty, Homeless and NonWhite. A threshold was applied to each distribution: for Poverty, it was regions with more than 55% of their residents being below 200% of the Federal poverty rate; for Homeless, it was more than 10 homeless residents per 1000 residents; and for NonWhite it was regions with more than 70% of their residents identifying as non-white.*

The increased mortality rates per ambulance diversion explored by Hsia et al. (2018) and Hsia et al. (2012) raise the question of what is the distribution of the mortality rate in San Francisco considering the vulnerability indicators shown in SFDPH's dataset; specifically, it begs the question of correlation between vulnerability clusters (*Figure 3*)[1] and mortality rates.

[1] **#dataviz**: The purpose of this visualization is to provide information about the distribution, both statistical and geographical, of Poverty as assessed by the SFDPH. In this particular case, I proposed the usage of simple graph convolutions and a binary threshold as a neighbor-aware (i.e., the neighboring

Furthermore, since natural disasters, such as extreme storms and flooding, significantly impact EMS response times (Bains et al., 2021), studying the gradual worsening of the mortality rates for the previously mentioned vulnerable populations due to SLR and anomalous precipitation patterns is fundamental to develop protective measures for public health against climate-change-related emergencies.

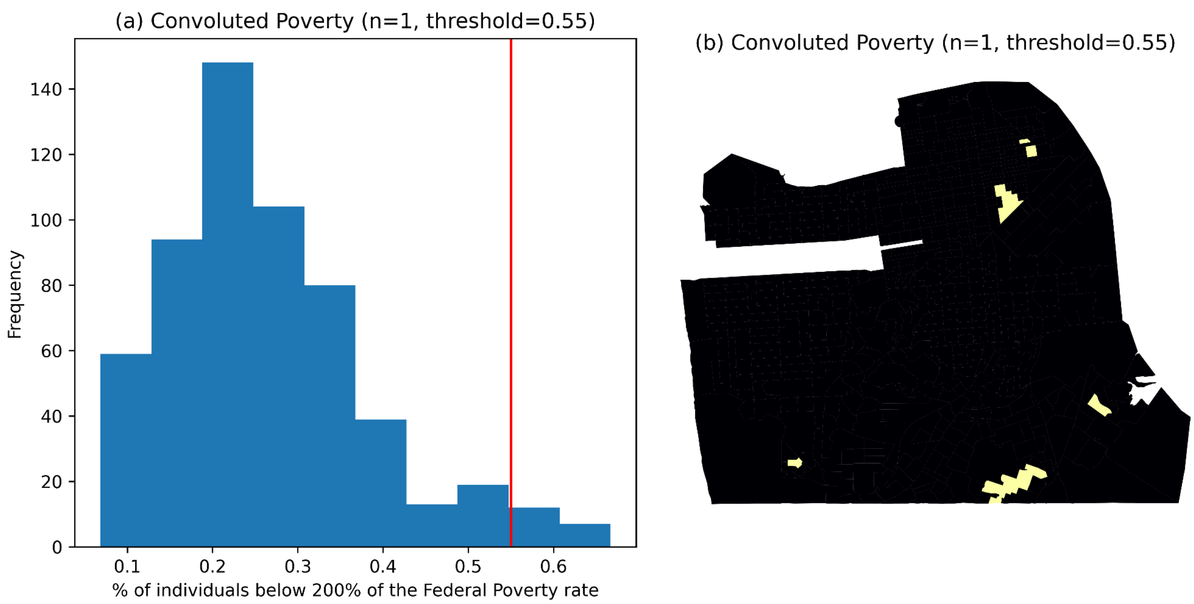


Figure 3: *Statistical distribution of SFDPH's Poverty and the geographical distribution of the simple graph convolution of Poverty (see Appendix A). The statistical distribution shows a threshold of regions with more than 55% of their residents below 200% of the Federal poverty rate. The geographical distribution shows the binary version of the same threshold.*

Study Objective

Following the findings provided by Hsia et al. (2018) and Hsia et al. (2012), this paper will explore the hypothesis of a correlation between poor and non-white patients experiencing

Poverty affects the Poverty of a given region) method to detect clusters; the procedure used is described in detail in *Appendix A*.

higher mortality rates due to ambulance diversion during EMS processes[2]. In particular, patients coming from vulnerability clusters (e.g., poverty, homelessness, and racial minorities) (*Figure 3*), experience significantly higher mortality rates due to increased ambulance diversion[3].

Study Methods

As with most health-related studies and tests, this hypothesis testing needs to be executed carefully. In this case, an interventional study would face many ethical implications that could eventually result in people dying. For example, a way of testing this hypothesis would be comparing the average mortality rate of ambulance diversion against the mortality rate with a modified geographical distribution of the EMS response resources (e.g., ambulances positioned near the vulnerability clusters.) Regardless of the truth value of the hypothesis, the modified allocation and distribution could impact patients outside of the targeted populations, causing deadly collateral damage.

Because of the ethical implications, this paper will propose an observational study to provide statistical evidence to support or reject the proposed hypothesis. The test would look at the EMS ambulance activity; mainly, the relevant variables for the test would be (1) the geographical location of the patient, (2) the time until treatment (in minutes), and (3) patient outcome (0 or 1, for dead or alive, respectively). The proposed test splits the population into two

---

[2] **#plausibility**: The main two premises on which the hypothesis relies are stated as the finding provided by Hsia et al. (2018) and Hsia et al. (2012). The evidence provided is clearly tied to the population affected by the hypothesis (clusters of poverty, homelessness, and racial minorities). The hypothesis is a logical conclusion made from the premises.

[3] **#hypothesisdevelopment**: In the first part of the scientific proposal, I mentioned multiple hypotheses surrounding increased mortality rates due to ambulance diversion. In this part, however, I focused on the hypothesis of patients coming from vulnerability clusters (e.g., poverty, homelessness, and racial minorities) experiencing on average higher ambulance diversion and therefore mortality rates. In the following section, I will propose a testing method for the proposed hypothesis. Overall, the hypothesis follows as a logical conclusion from the pieces of evidence presented.

groups: the patients whose geographical location belongs to a vulnerability cluster (studied population, y) and the rest of the patients (control population, x). With these two groups, the proposed method considers two different significance tests; the first is determined by a significance level of $\alpha = 0.01$, the null hypothesis and its corresponding alternative hypothesis:

$$H_0: T_x = T_y$$

$$H_1: T_x < T_y$$

Where T_x and T_y represent the sample mean of the time until treatment for the control and studied populations, respectively. The second significance test is determined by a significance level of $\alpha = 0.01$, the null hypothesis and its corresponding alternative hypothesis:

$$H_0: O_x = O_y$$

$$H_1: O_x < O_y$$

Where O_x and O_y represent the sample mean of the patient outcome for the control and studied populations, respectively.

If both significance tests rejected the null hypothesis[4], this would indicate that there is statistical significance to support the hypothesis of a correlation between patients coming from vulnerability clusters and higher times until treatment and mortality rates[5]. To further measure said significance, in the future, the proposed method could entail computing a measure of effect size, such as Cohen's d.

---

[4] **#testability**: States studied variables and proposes significance testing for checking the hypothesis. In particular, sets a significance level of 0.01 to test both increased time to treatment and mortality rates for the studied population; the null and alternative hypotheses are clearly stated. It also, shows how this testing method could be improved by computing a measure of effect size.

[5] **#observationalstudy**: After analyzing the ethical implications of an interventional study, I decided to propose an observational study. This method gathers EMS activity data (location, time to treatment, and outcome) to then perform significance tests on the collected data. Since the proposed method is observational, the hypothesis is intentionally phrased as a correlation instead of a causation.

Expected Results

As mentioned previously, studies have shown a correlation between ambulance diversion and higher mortality rates (Hsia et al., 2011) and a correlation between poverty (Hsia et al., 2018) and racial minorities (Hsia et al., 2012) and ambulance diversion. These correlations point in the direction of having enough significance to reject the null hypotheses proposed in the previous section. In particular, a study examined 1.04 million patient records from 202 hospitals and found that patients with the highest quartile of ambulance diversion had 7.1% higher mortality ($p < 0.001$) than those in the lowest quartile (Hsia et al., 2011). This implies that higher response times are correlated with higher mortality rates.

This way, according to a study done by Hsia et al. (2018), response times for patients coming from low-income zip codes were 10% higher (95% CI: 9% - 11%; $p < 0.001$) than those from patients coming from high-income zip codes, which translates to 3.8 minutes longer in the poorest zip codes.

Similarly, according to Hsia et al., 2012, hospitals serving high proportions of racial minority populations were at higher risk of experiencing diversion with a ratio of 1.02 (95% CI: 1.00 - 1.04). Furthermore, when computing the predicted number of annual hours on diversion with the 10th and 90th percentile of racial minority patients, hospitals at the 90th percentile experienced 306 hours (95% CI: 104 - 899) of annual diversion, which is 4.1 times (95% CI: 1.26 - 13.3) that of hospitals serving the 10th percentile, which experienced only 75 hours (95% CI: 27 - 210)[6].

[6] **#plausibility**: Provides detailed evidence that support the established hypothesis. The supporting findings were carefully studied by researchers, who found statistitical significance ($p < 0.001$) for their claims.

One limitation of the proposed study is that it does not consider the evolution of the expected results. Particularly, it needs to consider how EMS response can become significantly impacted by natural disasters, increasing ambulance diversion and response times (Bains et al., 2021).

Reflection

In the first part of the scientific proposal, I explored a unique way of processing the data and detecting clusters through simple graph convolutions. And, in the second part of the proposal, I also used this procedure to process and visualize the data, this time I added a detailed explanation of the procedure with supporting visualizations and illustrations. The latter was suggested after revising the first part of the proposal. In addition to this, in the second part special care was taken when writing the captions for the Figures added, since the captions did not explain the units of the shown data nor interpreted the chosen thresholds in some cases.

1277 words.

References

Bains, G., Breyre, A., Seymour, R., Montroy, J. C., Brown, J., Mercer, M., & Colwell, C. (2021). Centralized Ambulance Destination Determination: A Retrospective Data Analysis to Determine Impact on EMS System Distribution, Surge Events, and Diversion Status. *West J Emerg Med*, *22*(6). https://doi.org/10.5811/westjem.2021.8.53198

Hsia, R. Y., Asch, S. M., Weiss, R. E., Zingmond, D., Liang, L., Han, W., McCreath, H., & Sun, B. C. (2011, October 1). Is Ambulance Diversion Associated With Increased Mortality for Emergency Department Patients Who Are Admitted? *RESEARCH FORUM ABSTRACT ORAL PRESENTATIONS-QUALITY AND CORE MEASURES: EMERGENCY MEDICAL SERVICES*, *58*(4). https://doi.org/10.1016/j.annemergmed.2011.06.036

Hsia, R. Y., Huang, D., Mann, N. C., Colwell, C., Mercer, M. P., Dai, M., & Niedzwiecki, M. J. (2018, November 30). A US National Study of the Association Between Income and Ambulance Response Time in Cardiac Arrest. *JAMA Netw Open*. https://doi.org/10.1001/jamanetworkopen.2018.5202

Hsia, R. Y.-J., Asch, S. M., Weiss, R. E., Zingmond, D., Liang, L.-J., Han, W., McCreath, H., & Sun, B. C. (2012). California Hospitals Serving Large Minority Populations Were More Likely Than Others To Employ Ambulance Diversion. *Health Affairs*, *31*(8). https://doi.org/10.1377/hlthaff.2011.1020

Knowles, N. (2010). Potential Inundation Due to Rising Sea Levels in the San Francisco Bay Region. *San Francisco Estuary and Watershed Science*, *8*(1). https://doi.org/10.15447/sfews.2010v8iss1art1

Meites, E., & Brown, J. F. (2012). Ambulance Need at Mass Gatherings. *Prehospital and Disaster Medicine*, *25*(6). https://doi.org/10.1017/S1049023X00008682

Pacific Institute, Heberger, M., Cooley, H., Moore, E., & Herrera, P. (2012, July 1). The Impacts of Sea Level Rise on the San Francisco Bay. *California Institute for Energy and Environment (CIEE)*.

Pfeffer, W. T., Harper, J. T., & O'Neel, S. (2008, September 5). Kinematic Constraints on Glacier Contributions to 21st-Century Sea-Level Rise. *Science*, *321*(5894). https://doi.org/10.1126/science.1159099

Rahmstorf, S. (2007, January 19). A Semi-Empirical Approach to Projecting Future Sea-Level Rise. *Science*, *315*(5810). https://doi.org/10.1126/science.1135456

Trowbridge, M. J., Gurka, M. J., & O'Connor, R. E. (2009, November). Urban Sprawl and Delayed Ambulance Arrival in the U.S. *American Journal of Preventive Medicine*, *37*(5). https://doi.org/10.1016/j.amepre.2009.06.016

Wolff, M., & Comerford, C. (2018). Climate and Health Understanding the Risk: An Assessment of San Francisco’s Vulnerability to Flooding and Extreme Storms. *San Francisco Department of Public Health*. https://sfclimatehealth.org/wp-content/uploads/2018/12/FloodVulnerabilityReport_v5.pdf .pdf

Zhang, S., Tong, H., Xu, J., & Maciejewski, R. (2019). Graph convolutional networks: a comprehensive review. *Computational Social Networks*, *6*(11). https://doi.org/10.1186/s40649-019-0069-y

Appendix A

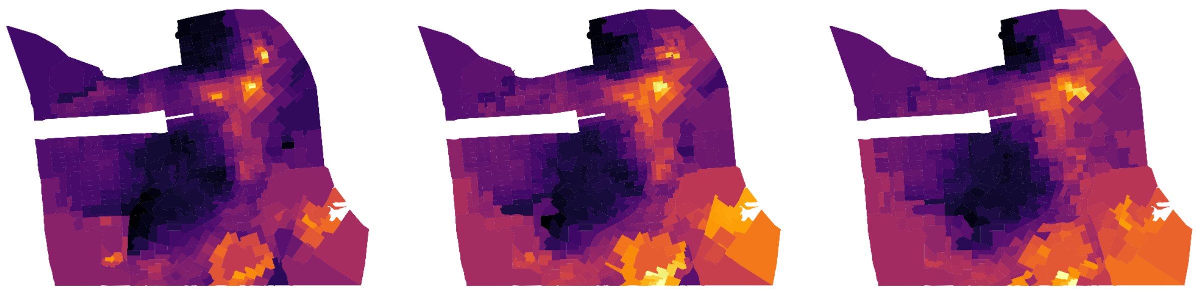
This appendix shows the method to compute the *Poverty* clusters.

This paper used a simple graph convolution and a pseudo-arbitrarily chosen binary threshold for the computed values to compute Poverty clusters.

The first part of this procedure is the simple graph convolution. This procedure requires thinking about the distribution of *Poverty* as a network where each node represents a geographical region (Census Blockgroup) and carries the poverty value associated with that region; in addition, each node is connected to every other associated node with a neighboring region. Then, a convolution computes a new value for each node, considering the connected nodes' values. In this case, this would translate back to the Poverty of neighboring regions affecting a given region. This computation is normalized by geographic area to account for population differences, meaning that regions with a larger area have a more significant influence on the final result of the computation. Note that this approach is not perfect since Poverty is computed with respect to the population and not the area. Roughly, the computation for a given node can be represented as the following mathematical expression:

$$\frac{\sum_{i=0}^{n} A_i P_i}{\sum_{i=0}^{n} A_i}$$

Where A_i represents the area for the i-th neighboring region to the given node, and P_i represents the Poverty value for the i-th neighboring region to the given node.



(a) Poverty, 1st order (n = 1) (b) Poverty, 2nd order (n = 2) (c) Poverty, 3rd order (n = 3)

Figure A1*: Geographical distribution of Convoluted Poverty (Wolff & Comerford, 2018) distribution after applying Graph Convolution with n = 1, 2, 3.*

Figure A1 shows the mathematical procedure applied at different orders (i.e., how many levels of neighbors to consider for the computation).