Exploring the Correlation Between Enviormental Indexes and Demographic Indexes Within the US

Emily Yarvis
Department of Information Technology
National Institute of Technology Karnataka
Surathkal, Mangaluru 575025, Karnataka, India
emilynyarvis@gmail.com

***Abstract*— This research paper aims at exploiting efficient ways of implementing the N-Body problem. The N-Body problem, in the field of physics, predicts the movements and planets and their gravitational interactions. In this paper, the efficient execution of heavy computational work through usage of different cores in CPU and GPU is looked into; achieved by integrating the OpenMP parallelization API and the Nvidia CUDA into the code. The paper also aims at performance analysis of various algorithms used to solve the same problem. This research not only aids as an alternative to complex simulations but also for bigger data that requires work distribution and computationally expensive procedures.**

***Index Terms*— N-Body, All-Pairs, Barnes-Hut, Parallelization, OpenMP, CUDA**

I. Introduction

In 2019 air pollution, specifically fine particulate matter, was the cause of about 6.4 million deaths around the world and 50,000 deaths in the United States(Collins). Since then, climate change has only gotten worse, our air has only become more polluted and thus those deaths have only climbed. While this statistic in itself is striking, what is more striking is who is most affected by these deaths. A study conducted by Timothy Collins and Sara Grineski at the University of Utah found that People of Color within a given region had a much greater exposure to fine particulate matter on average compared to White individuals within the same region consistently throughout the United States(Collins). To summarize while environmental racism is variably prevalent throughout the rest of the world, it is especially prevalent within the United States.

A. Method:

In order to further examine environmental racism prevalent in the United States, I utilized EJ screen to collect data on a variety of environmental and demographic indicators throughout the United States

EJ screen is a government website created by the EPA to give users access to environmental and demographic information for locations across the United States. In doing so the EPA hopes to be more transparent about the environmental and demographic issues that face our planet and make the information more accessible to all individuals no matter their background. The website offers data on 13 environmental indicators and 7 demographic indicators (see Figure 1) for every US census tract in the US. EJ screen sorts data in a variety of different ways such as by: city,state,tract and even block. With that said, for the purposes of this project we will mainly be examining data by every census tract.

These indicators give us the basis to understand how environmental discrimination exists throughout the US. Thus allowing us to examine how these factors correlate with one another and paint a picture of environmental discrimination with the US.

This report will explore the correlation between these environmental and demographic indicators and draw conclusions to the results we uncover.

II. Initial Research

A. Background

Before diving into the data available and making comparisons on a wide scale, it was important to start by examining 5 tracts from three different cities to see what the initial findings would be.

Houston, Texas is one of the best examples of systemic racism and thus environmental discrimination in the US. From(insert the figures) we can clearly see the racial divide that exists in Houston that forms in the shape of an arrow. We can also see that within this arrow an economic divide occurs. The arrow represents the majority white population within Houston and within that population the median household income is much higher than outside of the arrow. Therefore it is clear that systemic racism and discrimination is occurring and thus the question is raised of if environmental discrimination will be present in Houston as well as the clear economic discrimination that is occurring.

Another example of systemic racism is in Portland Oregon. After the BLM protests of 2020 and beyond, it is clear that systemic racism is prominent in Portland and thus makes it a clear candidate for identifying environmental racism as well.

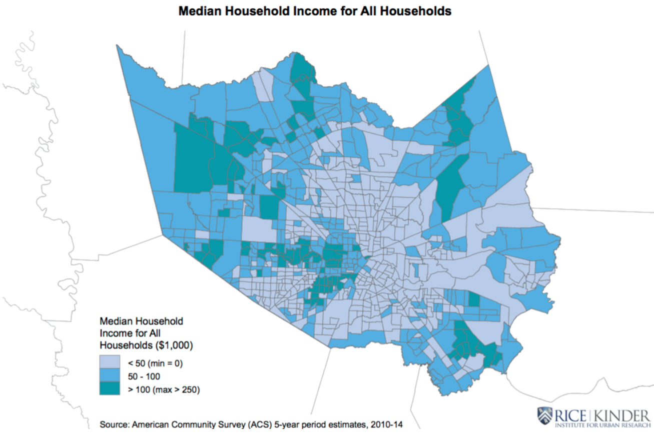

Fig. 1. Median Household Income in Houston,Texas

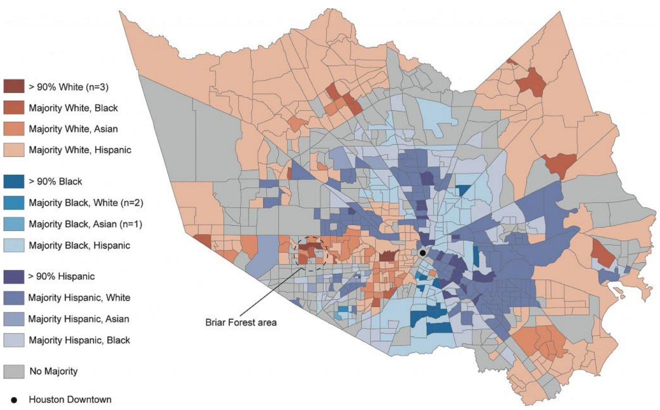

Fig. 2. Diversity in Houston, Texas

Finally Detroit,Michigan is also known for its issues of systemic racism and makes it an equally ideal candidate to examine. Therefore the initial cities chosen to be investigated were Detroit, Michigan, Portland Oregon, and Houston, Texas.

B. Method

In order to best observe the correlation between these factors, heatmaps will be used to clearly display the correlation coefficient of each set of factors through the scale of a colorbar. The correlation coefficients calculated are Pearson correlation coefficients and thus are calculated via the Pearson correlation coefficient formula. Within each heatmap, every square will represent the Pearson correlation coefficient between two factors.

C. Initial Heatmaps Analysis

When we examine the heatmaps of Portland, Houston and Detroit there are a few things that initially pop out. Firstly In the heatmaps of Portland and Houston there are

$$r = \frac{\sum(x_i - \overline{x}) - (y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}}$$

Fig. 3. Enter Caption

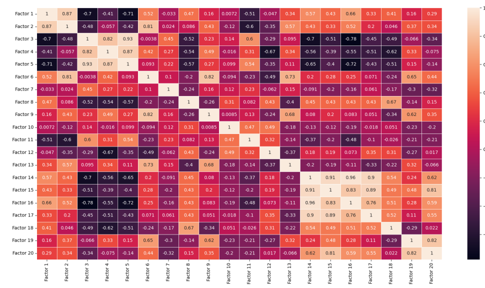

Fig. 4. Initial Heatmap of all 15 Tracts

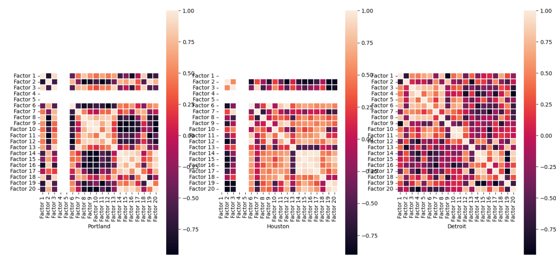

Fig. 5. Initial Heatmap of 5 Tracts In Portland Houston And Texas

two columns and rows that appear to be blank and show no correlation trend. What is actually occurring is that the correlation of those columns is not a real number because the data for those factors was the same. For the case of Portland factor 4 represents the cancer risk factor, and in the 5 tracts we examined the risk factor was always the same value. This trend also occurs in Houston and is why the columns appear blank.

It is also important to note that as we examine these heatmaps, while much of the maps seem brightly colored and thus highly correlated, it is important that we shift our gaze to a small section of the map. In each of these heatmaps viewers should see a light colored diagonal line of squares that represent the perfect correlation when you compare one factor to itself. Additionally, since some of the factors are environmental and some of them are demographic, it is important that we only closely examine factors from opposing categories, since it is already known that environmental factors will be correlated to one another and vice versa with demographic factors.

With that said it is still clear that these heatmaps tell an important story. When we look at the large heatmap representing all 15 tracts (5 from each city) we can see clearly that the highest correlation that we are interested in is represented between factor 13, demographic index, and the first 5 environmental factors. This proves our initial hypothesis that there would be correlation due to what we know about environmental discrimination but is also interesting to examine because many of the r values are much lower then one may have previously expected.

When we compare the Portland, Houston, and Detroit heatmaps, we do not necessarily see the same trends represented in the overall heatmap.In Portland we see similar results with Factor 13 being highly correlated to the first 5 environmental factors. However in Houston this correlation is much less significant and in Detroit the correlation seems to be even lower. This variability from the overall heatmap is most likely due to the very low amount of data we are analyzing. Since we are only looking at 5 tracts for each city, it is very unlikely that those 5 tracts (despite our best efforts) will be very representative of the given city, much less the state that the city resides in. Therefore in order to get a better picture of the correlations of these factors across the US, we need to examine a much larger data set.

D.

III. Further Exploration

A. Examining all tracts together

Figure [] displays the results of finding the correlation between all 20 factors across every single tract in the United States. By compiling all of the data from throughout the United States we make our data both more reliable and less reliable at the same time. While more data points provide us with a much fuller and more representative data set, it also pulls in a lot of outliers that muddle our data. Therefore looking at a heatmap of the correlation within the whole united states isn't very informative, given that every state is different and thus will have different correlation. Therefore it is much more insightful to look at the correlation of these factors on a state by state level.

Additionally, while it is interesting to look at all 20 factors compared to each other, much of this data is unnecessary and uninteresting. There is no need to compare the environmental factors to other environmental factors or the demographic factors to themselves because we are only interested in the correlation between environmental factors and their demographic counterparts. Therefore for the rest of the study we will be zooming in on 5 main environmental factors to compare to one demographic factor: Demographic index. In doing so only the most important and insightful data will be presented.

For the remainder of the report we will be looking at the correlation between demographic index and 5 environmental factors: Particulate Matter, Ozone, Diesel, Air Toxics Cancer Risk, Air Toxics Respiratory Risk Index. These factors have been chosen due to their relevance in modern day conversations surrounding climate change, and the plentiful data on them.

B. Examining correlation on a state by state matter

By comparing the correlation between demographic index and these 5 different factors between the different states we produce maps that look like figures whatever. Each state is correlated in a different way, some much stronger and some much weaker. Through these maps we are able to get a much better idea of where correlation exists compared to the initial total tract heat map.

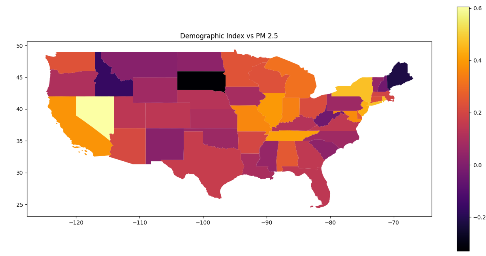

Fig. 6. Heat map displaying the correlation between demographic index and particulate matter(pm 2.5) across the United States

1) Correlation between Demographic Index and Particulate Matter:

2) Correlation between Demographic Index and Ozone: Figure whatever displays the correlation between Demographic Index and Particulate Matter in the air between states throughout the US.

IV. Conclusion

In conclusion it is clear that there is correlation between an individual's demographic index and the environmental side effects they face. Through the data collected from EJ Screen one can clearly see the correlation between one's demographic index and a variety of factors, especially the particulate matter in the air. While the correlations are different and more striking in some states and areas compared to others, those correlations still clearly exist, and thus environmental discrimination and racism is still prominent.

The next step of analysis would be to zoom in on each state and compare the correlation of these factors between counties. However due to the uneven populations of each county, and the uneven amount of tracts within different counties, there just isn't enough data to get reliable results.

Appendix

The appendix shows the analysis of the Barnes-Hut algorithm implemented in Parallel using OpenMP (method-1). The Sequential and Parallel times have been shown in Table 1; for all the galactic datasets [4] with number of bodies ranging from 5 to 30002.

Acknowledgment

I would like to thank Mohit Dubey for his valuable mentorship and guidance towards the execution and completion of my project.

References

[1] En.wikipedia.org. (2018). N-body problem. [online]. Available: `https://en.wikipedia.org/wiki/N-body_problem`. [Accessed 7 Jan. 2018].

[2] Carugati, N. J. (2016). The Parallelization and Optimization of the N-Body Problem using OpenMP and OpenMPI.

[3] 15418.courses.cs.cmu.edu. (2018). The Barnes-Hut Algorithm : 15-418 Spring 2013. [Online]. Available: http://15418.courses.cs.cmu.edu/spring2013/article/18. [Accessed: 07- Jan- 2018].
[4] Cs.princeton.edu. (2018). COS 126 Programming Assignment: N-Body Simulation. [Online]. Available: http://www.cs.princeton.edu/courses/archive/fall04/cos126/assignments/nbody.html. [Accessed: 07- Jan- 2018].
[5] Damgov, V., Gotchev, D., Spedicato, E., & Del Popolo, A. (2002). N-body gravitational interactions: a general view and some heuristic problems. arXiv preprint astro-ph/0208373.
[6] Beltoforion.de. (2018). The Barnes-Hut Galaxy Simulator. [Online]. Available: http://beltoforion.de/article.php?a=barnes-hut-galaxy-simulator. [Accessed: 07- Jan- 2018].
[7] Wwwmpa.mpa-garching.mpg.de. (2018). Cosmological simulations with GADGET. [Online]. Available: http://wwwmpa.mpa-garching.mpg.de/gadget/. [Accessed: 07- Jan- 2018].
[8] Barnes, J., & Hut, P. (1986). A hierarchical O (N log N) force-calculation algorithm. nature, 324(6096), 446.
[9] Burtscher, M., & Pingali, K. (2011). An efficient CUDA implementation of the tree-based barnes hut n-body algorithm. GPU computing Gems Emerald edition, 75.
[10] C. Swinehart. Arborjs.org. (2018). The Barnes-Hut Algorithm. [Online]. Available: http://arborjs.org/docs/barnes-hut. [Accessed: 07- Jan- 2018].
[11] Computing.llnl.gov. (2018). OpenMP. [Online]. Available: https://computing.llnl.gov/tutorials/openMP/. [Accessed: 07- Jan- 2018].
[12] Docs.nvidia.com. (2018). Programming Guide :: CUDA Toolkit Documentation. [Online]. Available: http://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html#cuda-general-purpose-parallel-computing-architecture. [Accessed: 07- Jan- 2018].