Varying Spatial Levels in GIS Analysis Environmental Epidemiological Data in Texas

Sala N. Senkayi*
University of Texas, Arlington
(Former Student)

Melanie L. Sattlerł
University of Texas, Arlington

ABSTRACT

Geographical Information Systems (GIS) are powerful tools for analyzing spatially related data. GIS has many potential applications in environmental epidemiology, because contaminant exposure is often spatially related. For example, persons living close to a source of toxic air emissions would likely be exposed to higher concentrations, resulting in greater likelihood of developing illnesses.

In using GIS to analyze environmental epidemiological data, a spatial scale must be chosen. A large scale, such as that of a metropolitan region, requires less data, and is thus easier and quicker in terms of computational time. It may not provide, however, a fine enough resolution to detect patterns of disease due to an emission source. A smaller scale, such as a block group, can provide finer resolution. A smaller scale requires more data, time for analysis, and computing power; however, it is also more likely to uncover relationships between individual emission sources and populations with higher incidence of disease.

This paper will discuss the issue of spatial scales in the context of a case study, involving proximity of childhood leukemia cases to airports. Using childhood leukemia cases reduces factors related to being exposed to benzene while working, smoking or an individual having been exposed to the pollutant while residing in one area and then moving to a different area many years later. (The entire study was described in a 2012 AWMA conference paper; this paper will focus on the issue of spatial scales.)[1] Using data provided by the Texas Department of State Health Services, observed/expected cancer ratios[2] were calculated in GIS. Scatter plots of observed/expected ratios were then generated as a function of distance from airports. Similar analyses were performed at the levels of block group, census tract, and county, including counties with large populations. At the scale of counties with large populations, no relationship was observed between leukemia cases and proximity to airports. However, at the block group level, a relationship became apparent. The case study thus illustrates potential advantages of analyzing smaller areas when teasing out relationships between diseases and pollutant emission sources. Additionally, micromaps was utilized in order to demonstrate if a visual relationship exists between the varying data sets at the county level thus testing the previous results.

Keywords: Varying spatial levels, Geographical Information System (GIS), Leukemia ages 9 and under, Airports.

1 INTRODUCTION

Geographical Information Systems (GIS) are powerful tools for analyzing spatially related data. GIS has many potential applications in environmental epidemiology, because contaminant exposure is often spatially related. For example, persons living close to a source of toxic air emissions would likely be exposed to higher concentrations, resulting in greater likelihood of developing illnesses. In using GIS to analyze environmental epidemiological data, a spatial scale must be chosen. A large scale, such as that of a metropolitan region, requires less data, and is thus easier and quicker in terms of computational time. It may not provide, however, a fine enough resolution to detect patterns of disease due to an emission source. A smaller scale, such as a block group, can provide finer resolution. A smaller scale requires more data, time for analysis, and computing power; however, it is also more likely to uncover relationships between individual emission sources and populations with higher incidence of disease.

This paper discusses the issue of spatial scales in the context of a case study, involving proximity of childhood leukemia cases to airports. This analysis is a follow-up to a paper presented at the 2012 Air and Waste Management Association Annual Conference, “Proximity of Childhood Leukemia Cases to Airports within the State of Texas.”[1] The previous study presented 3 lines of evidence suggesting a relationship between airport emissions and leukemia incidences for children 9 years of age and under, likely due to benzene emissions (since children 9 and under tend to not smoke, work in a place where benzene is used and they tend to have been born and still reside in the same location). The previous analysis looked at the following items for children 9 years and under, 1) the distance of those observed divided by what was expected leukemia cases to airports, roads, railroads and facilities at block group, census block, county and counties with large populations within the state from 1995 to 2005, 2) a Poisson Regression model of the aggregated leukemia cases and the summed up emissions (for the same emitters) within counties, and 3) the average distance those with the disease lived compared to a synthetic population dataset. All three demonstrated a relationship between airports and leukemia for those 9 years and younger. This follow-up analysis suggests that a smaller spatial level (such as a block group) demonstrates a relationship between airport distances and observed-to-expected cancer incidence ratios. Thus, the higher observed-to-expected incidence ratios for smaller areas were closer to the source; however, the higher observed-to-expected incidence ratios were further from the source as the spatial area increased to counties and counties with larger populations. Micromaps (an interactive tool used to display geographical spatial patterns) was used to test the results of the scatter plots.[3,4]

Several micromap runs are provided in order to demonstrate a relationship or the lack of a relationship between the environmental epidemiological data sets. The process described in this paper has the potential to assist agencies in researching geographical patterns of diseases and their potential relationship with environmental stressors within the state of Texas, in other areas of the United States or even worldwide.[3,4]

1.1 Methodology

Cancer incidence data for Texas from 1995 to 2005 were obtained from the Texas Department of State Health Services (DSHS).[2] Leukemia incidences for children 9 years and under were separated from the rest of the data. The cancer incidence data provided the latitude-longitude location of each childhood leukemia case. Shapefiles containing spatial locations of airports were provided by EPA (1966 total). Using GIS, the distance in miles from each childhood leukemia case to the nearest airport was obtained.

The four spatial levels used for the analysis were, in order from smallest to largest, block groups (14,463), census tracts (4,388), counties (254), and counties with large populations (36). Texas counties with large populations are listed in Table 1. For each spatial level, the distances of the childhood leukemia cases to the nearest airport were averaged.

Table 1: Texas Counties with Large Populations

1.	Potter	2.	Randall	3.	Wichita
4.	Grayson	5.	Lubbock	6.	Denton
7.	Collin	8.	Tarrant	9.	Dallas
10.	Smith	11.	Brazos	12.	Williamson
13.	Montgomery	14.	Travis	15.	Hays
16.	Jefferson	17.	Harris	18.	Guadalupe
19.	Gregg	20.	Johnson	21.	Ellis
22.	Taylor	23.	Ector	24.	Midland
25.	El Paso	26.	McLennan	27.	Tom Green
28.	Bell	29.	Fort Bend	30.	Bexar
31.	Brazoria	32.	Galveston	33.	Webb
34.	Nueces	35.	Hidalgo	36.	Cameron

Observed-to-expected cancer incidence ratios (adjusted for age, race and gender) were then calculated for childhood leukemia at the block group, census tract, county, and large county levels. For details of the process for calculating observed-to-expected-incidence ratios, please see Senkayi et al. (2012).[1] Finally, for each spatial level, the observed-to-expected incidence ratios were plotted versus the average distances of the cases to airports, using GIS. The county level data was then run in micromaps due to the smaller number of variables.

2 RESULTS

Resulting plots of observed-to-expected-incidence ratios versus average distance to airports are shown in Figure 1, for the 4 spatial levels. The plots visually and clearly demonstrate that there is a relationship with the average distance from the source verses the observed-to-expected ratios at the block group and the census tract levels; however, this relationship is not visibly apparent at the county and large county levels, as discussed in more detail below.

When viewing these plots, it is important to note that the y-axis represents the observed-to-expected ratio; also note when this number is one, then what was observed was expected (for that area) according to the calculations using rates (to adjust for race, gender and age) provided by Texas Department of State Health Services (DSHS).[1] However, a larger number indicates that more cases were observed then what was expected. The observed-to-expected ratios are plotted verses the average distance of the area (block group, census track, or county) to the nearest airport (x-axis, in miles).

The largest observed-to-expected ratios are points in the red Jenks Natural Breaks group. GIS recalculates the Jenks Natural Breaks for each scatter plot. In order for a relationship to exist between disease and the source, the points with larger observed-to-expected ratios should be visibly closer to the source. For the block group and the census tract plots, it is apparent that the red data points (which have larger observed-to-expected ratios) are visibly closer to the y-axis (or closer to the emission source) then the rest of the data points. The counties and counties with large populations demonstrate the opposite, or that the red points (which have larger observed-to-expected ratios) are on average visibly further from the y-axis (or further from the emission source) compared to points of other colors (which have lower observed-to-expected ratios).

Figure 2 demonstrates this lack of relationship between the observed-to-expected cancer incidence ratios and the distance to airports and roads at the county level utilizing micromaps (a tool that assist in spatially visualizing varying datasets). Within this micromap run, it is apparent that a relationship between observed-to-expected cancer incidence ratios and the distance to airports and roads does not exist. This lack of relationship is noted due to the airport and road distance to data (within counties) not increasing (or going from left to right) as the leukemia cases increase in a left to right fashion. These micromap results match the results obtained by the dissertation in 2012.[1] With over 14,000 block groups a micromap run was not provided at this level, or at the (over 4,000) census tract level.

Figure 3 has been included in order to provide an example of a micromap run where the data sets present a relationship amongst the differing variables. The micromap in Figure 3 demonstrates the relationship between the number of leukemia cases and benzene emissions from airports and roads at the county level utilizing micromaps. Counties with zero cases were not included in this run; thus, only 168 counties were used for this analysis instead of the total 264 counties within the state. In this analysis, micromaps quickly and easily assisted with visualizing relationships between health data (leukemia) and environmental stressor data.

Within this micromap run, it is apparent that a relationship between total cases of leukemia (for those 9 and under) with airport and road benzene emissions exists. This relationship is noted due to the airport and road emissions data (within counties) decreasing (or going from right to left) as the leukemia cases also decrease in a similar fashion. These micromap results match the results obtained by the dissertation in 2012.[1]

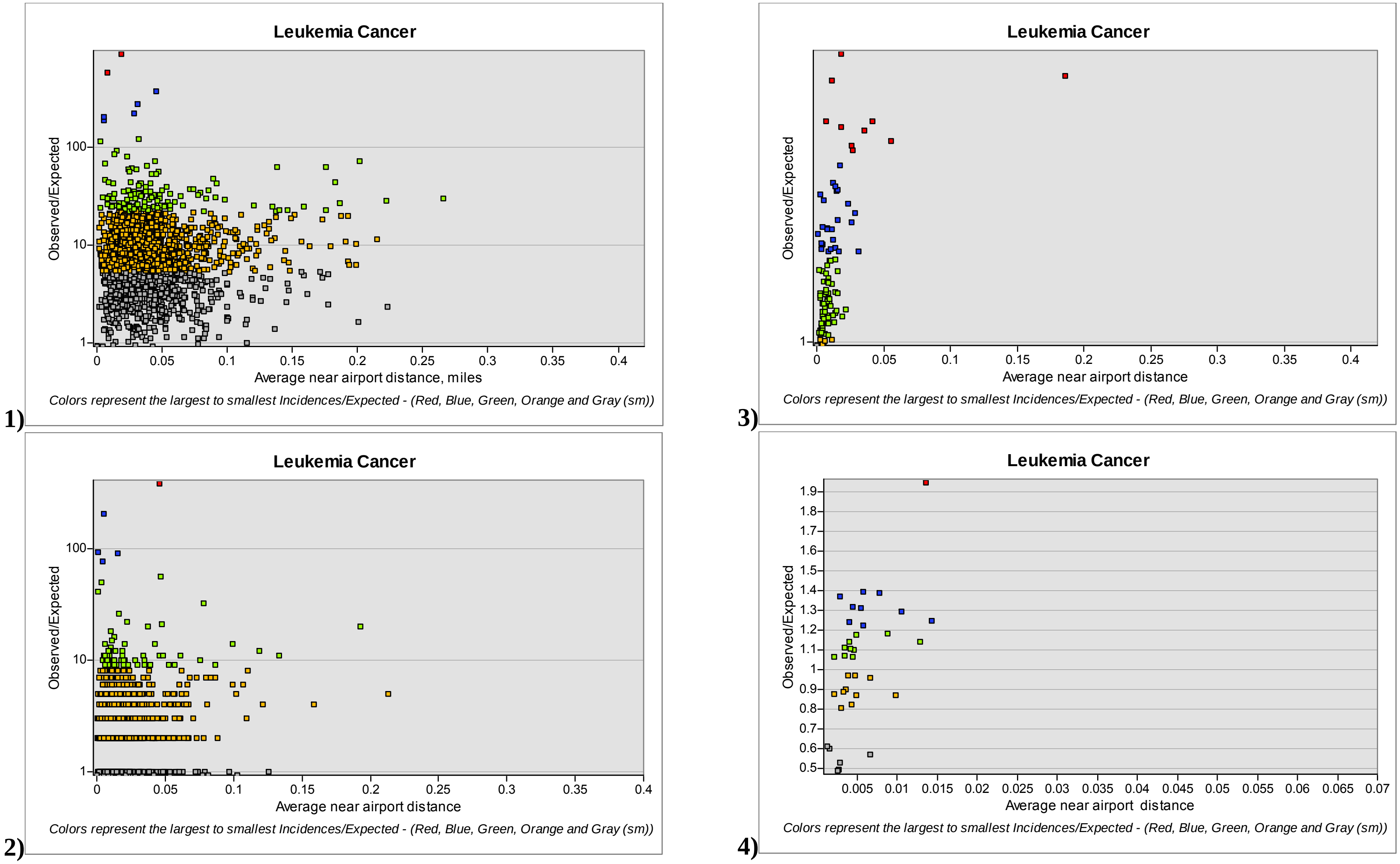

Figure 1. Scatter plots of observed-to-expected incidence ratios of leukemia for children age 9 and under on the Y – Axis verses the average near distance to airports in miles on the X – Axis at 1) the block group (14,463 total), 2) census tract (4388) 3) county (254) and 4) counties with larger population levels (36). The five color classifications were generated using Jenks Natural Breaks default classification within GIS. All distance measurements are in miles.

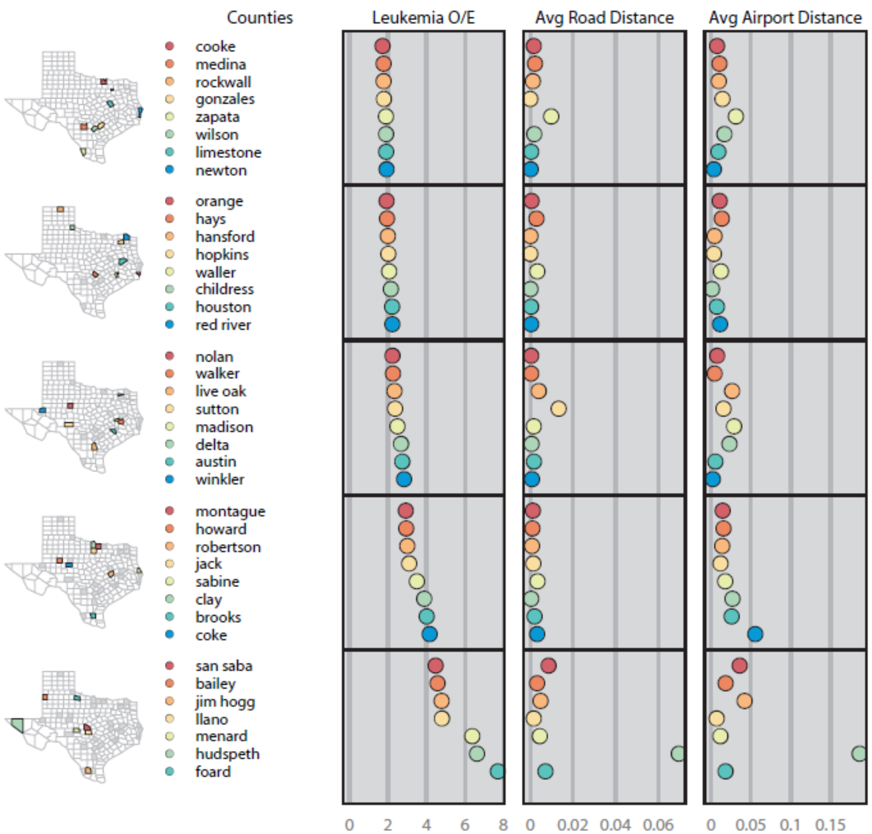

Figure 2. Micromaps demonstrating the relationship between the observed to expected (O/E) leukaemia cases and the distance to airports and roads. This map is of the previous scatter plot 3 in Figure 1. Note the data was sorted from smallest to largest O/E and just the top 38 data variables are mapped. This map was provided by Sala N. Senkayi of U.S. EPA Region 6.

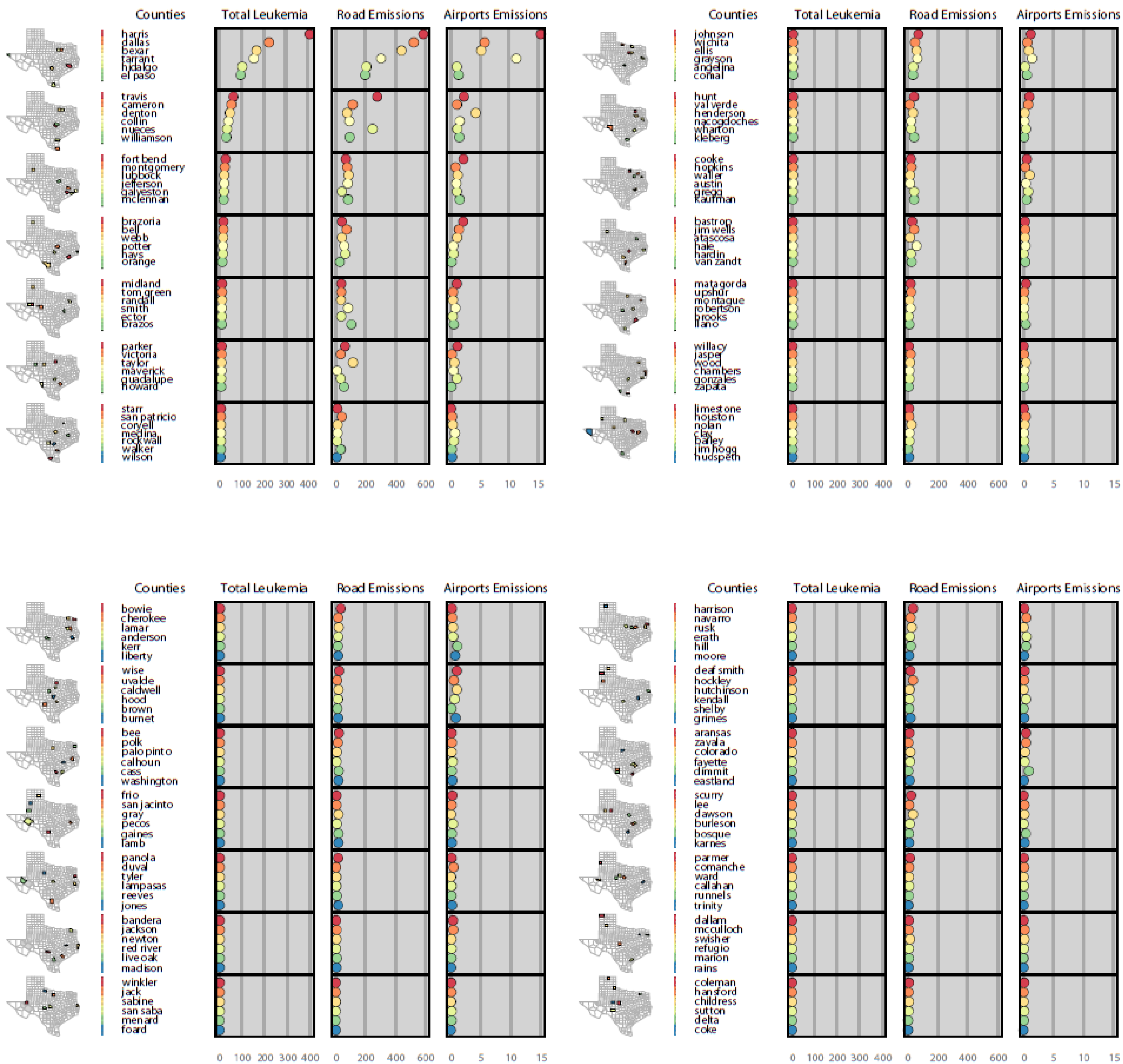
Counties
Total Leukemia
Road Emissions
Airports Emissions
harris
dallas
bexar
tarrant
hidalgo
el paso
travis
cameron
denton
collin
nueces
williamson
fort bend
montgomery
lubbock
jefferson
galveston
mclennan
brazoria
bell
webb
potter
hays
orange
midland
tom green
randall
smith
ector
brazos
parker
victoria
taylor
maverick
guadalupe
howard
starr
san patricio
coryell
medina
rockwall
walker
wilson
0 100 200 300 400
0 200 400 600
0 5 10 15
Counties
Total Leukemia
Road Emissions
Airports Emissions
johnson
wichita
ellis
grayson
angelina
comal
hunt
val verde
henderson
nacogdoches
wharton
kleberg
cooke
hopkins
waller
austin
gregg
kaufman
bastrop
jim wells
atascosa
hale
hardin
van zandt
matagorda
upshur
montague
robertson
brooks
llano
willacy
jasper
wood
chambers
gonzales
zapata
limestone
houston
nolan
clay
bailey
jim hogg
hudspeth
0 100 200 300 400
0 200 400 600
0 5 10 15
Counties
Total Leukemia
Road Emissions
Airports Emissions
bowie
cherokee
lamar
anderson
kerr
liberty
wise
uvalde
caldwell
hood
brown
burnet
bee
polk
palo pinto
calhoun
cass
washington
frio
san jacinto
gray
pecos
gaines
lamb
panola
duval
tyler
lampasas
reeves
jones
bandera
jackson
newton
red river
live oak
madison
winkler
jack
sabine
san saba
menard
foard
0 100 200 300 400
0 200 400 600
0 5 10 15
Counties
Total Leukemia
Road Emissions
Airports Emissions
harrison
navarro
rusk
erath
hill
moore
deaf smith
hockley
hutchinson
kendall
shelby
grimes
aransas
zavala
colorado
fayette
dimmit
eastland
scurry
lee
dawson
burleson
bosque
karnes
parmer
comanche
ward
callahan
runnels
trinity
dallam
mcculloch
swisher
refugio
marion
rains
coleman
hansford
childress
sutton
delta
coke
0 100 200 300 400
0 200 400 600
0 5 10 15

Figure 3. Micromaps demonstrating the relationship between the number of leukaemia cases and the concentration of benzene from airports and roads. This map was provided by Marc Weber of U.S. EPA.

Data visualization tools like micromaps have been studied since 1998 and have been used to visualize health data by the National Institute of Health (NIH)/National Cancer Institute (NCI) as demonstrated by Figure 4 the State Cancer Profiles located below. This micromap run is similar to the micromaps provided in Figures 2 and 3; however, this micromap run excludes environmental data.

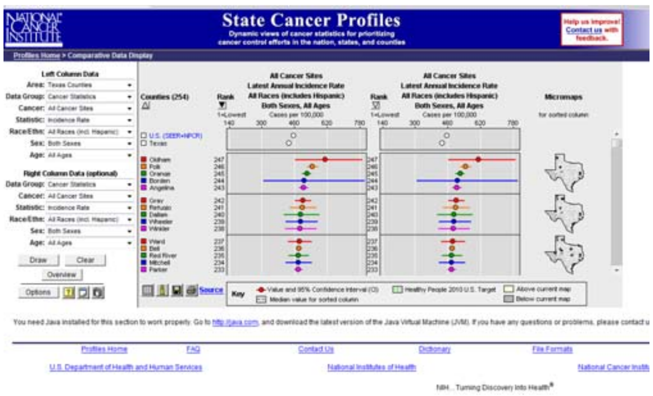

Figure 4. State Cancer Profile via the NCI, this demonstrates how the user (the lay public for example) can query differing cancer variable.[7]

Table 2 and Figure 4 have been provided in order to offer a better understanding of airport emission in Texas. Table 2 presents the top 10 airports emitting benzene within Texas. George Bush and Hobby are numbers 1 and 3; in addition, both are located in Houston area (or Harris County). Dallas Fort Worth (DFW), Dallas Love and Addison are numbers 2, 4, and 10; in addition, they are both located in the DFW area (or Dallas County). These 10 airports contribute to the emissions of the over 1900 airports provided in Figure 5's map.

Table 2: Top 10 Texas Airports with Latitudes and Longitudes, in addition to Benzene Emissions (in order from largest to smallest) for the

Facility Name	Lat	Long	Emission
George Bush Intercon	29.99	-95.363	7.40448
Dallas/Fort Worth In	32.90	-97.083	6.08909
William P Hobby	29.64	-95.286	2.26339
Dallas Love Field	32.85	-96.863	2.23740
San Antonio Intl	29.54	-98.486	1.83419
Austin-Bergstrom Int	30.18	-97.678	1.68948
David Wayne Hooks Me	30.07	-95.555	1.43768
Northwest Regional	33.04	-97.232	1.12290
Stinson Muni	29.33	-98.480	1.10073
Addison	32.97	-96.840	1.00524

Figure 5 provides a map of the density of total leukemia cases for children age 9 and under within the entire State of Texas for the years 1995 to 2005, provided by DSHS. The density was obtained by running a density tool within GIS, this tool takes all the data and then demonstrates areas with high density as white and areas with low density as black. The two whitest areas on this map are in the Dallas and Houston area, additionally these same two areas have the highest (Harris County) and second highest (Dallas County) benzene emitting airports in the State, which matches the micromap output in Figure 3.[2]

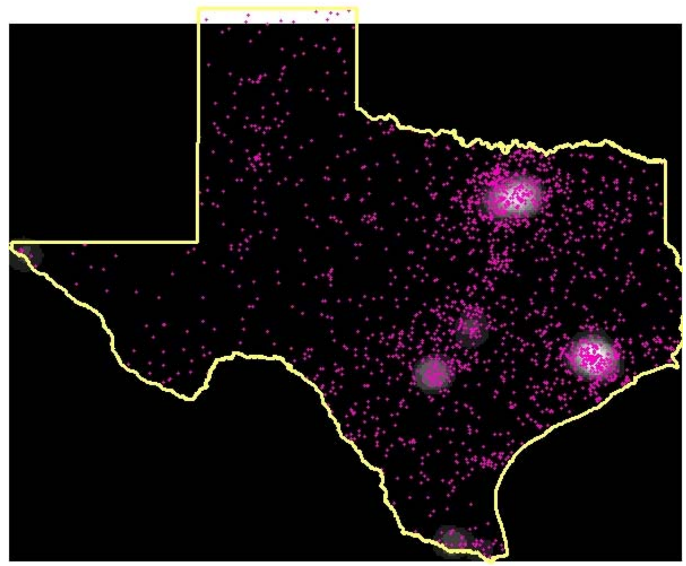

Figure 5. Map of Airports via EPA Dataset & Leukemia Cancer Density.

As demonstrated in this paper, all levels are important when analyzing health and environmental data; however, smaller levels do assist in assessing tiny details that may be missed at larger spatial levels. Additionally, visualization tools like scatter plots and micromap can assist the general public or researchers in providing an output that is relatively simple to understand. It is suggested that scatter plots be used for a distance to emitter verses the observed to expected data sets at all spatial levels and especially at the block group (or the lowest common level); in addition to a sum of emissions verses a sum of the incidences at the county level (and if this data is provided at a lower level then running this analysis is also suggested). It is also suggested that all county runs be done with the assistance of micromaps, allowing for a more consistent presentation.

Note, other sources that could potentially assist in contributing to the leukemia cancer incidences, include railroads and industrial sources, and were discussed in previous work; however were not included since a relationship was not observed early on.[1]

3 Summary

In this study ratios of observed-to-expected incidences were plotted for leukemia cancer (for children 9 years and under) vs. distance to airports for each block group, census tract, county and large population county in the State of Texas using GIS. This study concludes that a relationship between airport distances and incidences of childhood leukemia in the State of Texas exists at smaller spatial levels (block group and census tract). However, as the spatial level increases to the county, there is not an apparent relationship, since the areas with the higher observed-to-expected incidence ratios are further from the source. It is advised that running this type of analysis at all spatial levels would be useful since datasets differ with differing spatial areas. Finally, micromaps was used to assist in observing the relationship of the Leukemia cases to the emissions from airports and roads also adding weight to the observed relationship of the disease and environmental stressor.

REFERENCES

[1] Senkayi, Sala; Sattler, Melanie L; Chen, Victoria C.P. "Proximity of Childhood Leukemia Cases to Airports within the State of Texas." Proceedings of the Air & Waste Management Association 105th Annual Conference. San Antonio, Texas, June 19-22, 2012.

[2] Department of State Health Services (DSHS). Texas Cancer Registry, Expected Numbers of Cancer Cases and Deaths, Texas, 2005. http://www.dshs.state.tx.us/tcr/expected-2005.shtm. Viewed August 7, 2011.

[3] Symanzik, Jürgen, Linked Micromap Plots-In print and on the Web. http://www.math.usu.edu/~symanzik/talks/2009_UFMG_Session2_Micromaps.pdf. Viewed March 10, 2013.

[4] Chen, Jim; Wang, Xusheng; Carr, Daniel; Bell, B. Sue; Pickle, Linda. "Java-based Dynamic Linked Micromap Plots." http://www.interfacesymposia.org/I03/I2003Proceedings/ChenJim/ChenJim.presentation.pdf. Viewed September 11, 2013.

[5] Pickle, Linda; Heineman, Ellen; Ward, Mary; Nuckols, John; Gumpertz, Marcia; Bell, Sue. "Applications of GIS to cancer research at the National Cancer Institute". http://proceedings.esri.com/library/userconf/health01/papers/hc01_p01a/hc01_p01a.html. Viewed September 11, 2013.

[6] Shneiderman, B. C. Plaisant, and B.W. Hesse, "Improving Healthcare with Interactive Visualization. IEEE Computer", 2013. 46(5): p. 58-66.

[7] National Cancer Institute, State Cancer Profiles, http://www.statecancerprofiles.cancer.gov/micromaps/. Viewed October 6, 2013.