Spatial Analysis of Cardiovascular Disease Incidence and Potential Environmental Factors in the California Bay Area
Nicholas J. Papanastassiou ’13  Sarah A. Holmes ’13
Environmental Studies Program, Colby College, Waterville, Maine

Abstract
This project uses GIS to investigate proposed environmental factors that contribute to heart disease. This is based on several scientific articles (see references) that have hypothesized positive correlations between areas of increased air and noise pollution and dioxin emissions with heart disease incidence. We examined the Bay Area of California, which consists of nine counties. No significant correlation between our proposed factors and current heart disease incidence was found.

Introduction
Cardiovascular disease (CVD) affects 81 million people throughout the United States. There are many contributing factors to CVD, including smoking, diet and health decisions, and congenital heart defects. Additionally, studies have shown noise and air pollution to increase one’s risk of CVD. We asked the question of whether or not NPL Superfund Sites, Airports, and Roads increase the likelihood of contracting CVD. We chose to analyze the Bay Area in Northern California, looking at incidences of CVD per county.

Methods
We looked at county data for cardiovascular disease incidence, corrected for age differences, and compared it to an original model to predict heart disease incidence. We compiled data containing information on our three factors: air pollution, noise pollution, and dioxin emissions, which we quantified by looking at road density (weighted by road type), airport density, and Superfund site density in each county. Road density was measured by total length of road system / total area of county, airport density was measured by total # of airports / cumulative population in all cities with airports, and Superfund site density was measured by # of sites / total area of county. We converted each range of densities to a 1-10 scale, in order to compare them.

We then ran a multivariable least-squares regression in order to determine which variables best predicted incidences of CVD. The resulting equation was: \[ \text{CVD} = 7.750 - 0.050(\text{AirDens}) - 0.006(\text{RoadDens}) - 0.137(\text{SuperDens}). \]

We assigned weights to each variable based on the coefficients from the regression equation. Using this equation, we produced a predictive output layer of CVD incidences. This was coded with the same methods as the observed CVD data in each county, so as to be able to visually compare the two maps.

Results
We found no significant predictive value of the hypothesized factors in measuring CVD incidences. A multi-variable regression of our dependent variable (CVD Incidence) and our three independent variables (Road, Airport, and NPL Superfund Site Density) yielded F-values > 0.635 for each variable and t-values < 0.51 for each variable. An F-value of 0.89 resulted, as well as an R^2 value of 0.11 (see Table 1, Figure 3).

The index map produced some CVD ranges for certain counties well, but Napa County and Contra Costa County are two notable exceptions.

Discussion
Our findings indicate that the noise and air pollution associated with airports, roads, and NPL Superfund sites do not significantly contribute to incidences of CVD. Only 10% of the variance in CVD incidences between Bay Area counties is captured by the variables. For road density, there is a 98% chance that the minimal effect that road density has on CVD is attributable to chance alone. These findings are in opposition to the hypotheses of Hoffmann et al., Kopf et al., Mead, and Román et al.

It is important to note the severe limitations of this study. Firstly, there are many confounding variables not accounted for in the model that likely affect CVD incidences (e.g., smoking) as well as the variables that we used in our model (e.g., wind direction affecting pollutants). Secondly, our data on CVD incidences were available only to the county-wide scale, further limiting the precision of our analysis.

Thirdly, our statistical analysis shows that the combined independent variables do not predict CVD incidences better than would be predicted by chance alone. Therefore, the equation resulting from the regression and thus the equation used to calculate our index map was not significant. Fourthly, it was likely inappropriate to use linear regression at all. We had a very small sample size that severely limited the accuracy of the regression, and prevented tests of normality from being used. Additionally, the statistical data were not tested for autocorrelation, which could impact the accuracy of the results. Therefore, as a result of these limitations, the index layer created for the projected incidence of CVD map does not reflect significant predictions. However, it showed the inability of the road, airport, and superfund density in our model to accurately predict observed rates of CVD incidences.

Conclusion
There is no significant relationship between the densities of NPL Superfund Sites, Airports, and Roads in the Bay Area counties, and increased likelihood of contracting CVD in those counties.

References

Acknowledgements
A big thank you to Professor Nyhus and Manny Gimond for help with brainstorming ideas, how to most effectively use GIS in our project, proper methods for carrying out the analysis, and being available and willing to answer any questions that we might have.

Data was obtained from ESRI93, Federal Aviation Administration, EPA National Priorities List, Metropolitan Transportation Commission and California Department of Health Services: Center for Health Statistics.