Abstract
Cancer is the second leading cause of death among Americans. In 2017 there will be an estimated 1,688,780 new cases diagnosed and 600,920 new cancer deaths. While there are many known and possible causes for this complex group of diseases, there is an emerging interest among scientists in the involuntary environmental exposures that lead to the development of cancer. Currently, there is limited research showing explicit links between cancer and environmental exposures, and there is still a dearth of knowledge about the increased cancer risk of living in close proximity to toxic chemical facilities and hazardous waste sites. With over seven hundred hazardous waste sites with various status designations identified by the U.S. EPA’s CERCLIS, and over two thousand U.S. EPA identified TRI facilities within the U.S., it is crucial that the impacts of such sites on human health and the environment be taken seriously. This project explores the geospatial relationship between age-adjusted cancer mortality rates (deaths per 100,000) in the U.S. and the distribution of toxic chemicals and hazardous waste. Using various GIS techniques, this project explores the clustering, density, and directional distribution of cancer mortality rates, NPL sites, TRI facilities, and TRI core chemical releases.

Methodology
In order to explore the spatial distribution of cancer in the U.S., cancer data for the latest 5-year average of the age-adjusted annual death rate (deaths per 100,000) for all cancer types and all people was downloaded into excel and joined to census polygons of the U.S. by county. The data was classified by natural jenks, and reclassified to show counties with rates below and above the average annual death rate (166.1). Local Moran’s I for both classifications showed clustering in the lower middle right section of the country.

Vector point shapefiles of CERCLIS sites (showing NPL sites) and TRI facilities were imported from the U.S. EPA’s FRS site, and TRI core chemical releases by magnitude were downloaded into excel and geocoded by latitude and longitude. The kernel density of NPL sites and TRI facilities was determined, and the mean density of each county was recorded in a table, using zonal statistics. The table was joined back to the county layer for further analysis. As expected, local Moran’s I for NPL sites and TRI facilities showed almost identical patterns of clustering. Areas of high-high CERCLIS sites and high-high TRI facility clustering were selected and compared to areas of high-high cancer clustering. The tables show the counties with high-high NPL site and TRI facility clustering which also show greater than average death rates.

TRI facility core chemical releases were symbolized to show total onsite releases by magnitude, and was overlaid on a map showing high-high cancer mortality clustering. Mean centers, and ellipses showing standard deviations one and two were found for clustering of cancer mortality rates and TRI releases by magnitude. The results show the directional distribution of both.

Findings
Cancer, NPL sites, and TRI facilities are clustered in the U.S. The majority of counties with high-high clustering of NPL sites and TRI facilities have greater than average cancer mortality rates (Tables Above). The directional distribution of TRI core chemical releases by magnitude is not consistent with the directional distribution of cancer mortality rates, as seen by the mean center and ellipses showing standard deviations one and two. Further investigation is necessary to determine the spatial relationship between hazardous waste, toxic chemicals and cancer in the U.S. Future studies should look at smaller areas in the U.S. and specific cancer types, perhaps in counties with high-high clustering in order to explore the relationship and determine potential causation.