Toxic Mapping with Python and GIS: Exploring Relationships between Carcinogen Dumping and Cancer

Amy Halloran

DePaul University

First Submission June 2015, Final Presentation October 2015

The GeoTech Center and URISA 2015 Undergraduate Geospatial Technology Skills Competition

SOCIAL PROBLEM: WHY IS THIS IMPORTANT?

- Cancer is a growing problem, with high social and financial costs for individuals and families

TECHNICAL PROBLEM: HOW CAN PYTHON AND GIS LEAD TOWARD ANSWERS?

- PYTHON = Speed: Data analysis is time consuming and sometimes not feasible depending on type of data necessary for research
- GIS = Visual: Toxic mapping shows spatial relationships, but demonstrating relationships with cancer rates requires improvements in accuracy, more information, and more time
- Proximity to facility and cancer rates suggest a spatial relationship, but other variables such as age, type of work, gender, income, and lifestyle, along with type and length of exposure must be considered

A CASE STUDY OF WILL COUNTY

- Will County, IL is located southwest of Chicago and is part of the greater Chicago area.
- Air pollution: The city of Joliet has a history of heavy industry that has caused environmental and human damage
- Water pollution: Will County sources most of its drinking water from a main well in Joliet, which is connected to sub wells throughout the county
- Is there a relationship between facilities that dumped carcinogens in the late 1980s-early 1990s and cancer rates in the mid 2000s?

PART 1: DATA COLLECTION AND PREP

Table preparation—make all tables workable format CSV

- 7 TRI facilities: 1988-1994
- 1 Cancer Diagnoses: 2006-2010
- 1 Carcinogenicity (known or RAHC): 2013

Run Python code to prepare data tables for use in ArcMap

- Convert all CSV to dBASE
- Clean up chemical field in Carcinogenicity table
  - Make all values match with TRI chemicals
- Clean up chemical field in TRI tables
  - Now TRI and Carcinogenicity tables have a common field
- Join each TRI table with the Carcinogenicity table
  - Fields: Carcinogenicity (K or RAHC), Primary Site, Primary Exposure

MORE ABOUT THE DATA

- Cancer incidence data offers age, gender, stage, and **site**

cancerIncid.updatedWillCo

	OID	ID	sex	Year	Zip	stage	site	age	Lat	Long	sex1	stage1	site1	age1
▶	0	1	1	2006201	6090	3	9	4	41.1137	-87.8709	male	distant metastases/systemic disease	leukemias and lymphomas	65 plus
	1	2	1	2006201	6095	1	6	3	41.2497	-87.8486	male	localized	prostate	45 to 64
	2	3	1	2006201	6049	1	6	3	41.6122	-87.9542	male	localized	prostate	45 to 64
	3	4	1	2006201	6054	0	7	4	41.6188	-88.1886	male	in situ	urinary system	65 plus
	4	5	1	2006201	6041	1	7	3	41.5081	-87.608	male	localized	urinary system	45 to 64
	5	6	1	2006201	6041	3	3	4	41.4415	-87.5996	male	distant metastases/systemic disease	lung and bronchus	65 plus

- Carcinogenicity can be expanded on to determine particular carcinogens with cancer site—Manual data creation
  - For example, benzene is known to cause leukemia and is very sensitive to proximity.

Report on Carcinogens, Thirteenth Edition For Table of Contents, see home page: http://ntp.niehs.nih.gov/go/roc13

Benzene

CAS No. 71-43-2

Known to be a human carcinogen

First listed in the *First Annual Report on Carcinogens* (1980)

Carcinogenicity

Benzene is *known to be a human carcinogen* based on sufficient evidence of carcinogenicity from studies in humans.

Cancer Studies in Humans

Case reports and case series have reported leukemia (mostly acute

Properties

Benzene is the primary aromatic compound. It exists at room temperature as a clear, colorless-to-yellow liquid with an aromatic odor. It is only slightly soluble in water, but it is miscible with alcohol, ether, chloroform, carbon disulfide, acetone, oils, carbon tetrachloride, glacial acetic acid, and most other organic solvents. Benzene is highly flammable (Akron 2009). Physical and chemical properties of benzene are listed in the following table.

Property	Information
Molecular weight	78.1
Specific gravity	0.8787 at 15°C/4°C
Melting point	5.5°C
Boiling point	80.1°C
Log K_{ow}	2.13
Water solubility	1.79 g/L at 25°C

SCRIPT SAMPLE 1

```
                    #CLEAN UP CHEMICAL FIELDS#

#Clean up Carcinogens_0.updated.dbf:

#Add field for modified name
arcpy.AddField_management("Carcinogens_0.updated.dbf", "NAME2", "TEXT")

#set and run cursor in Carcinogens_0.dbf to populate a modifed name field
with arcpy.da.UpdateCursor("Carcinogens_0.updated.dbf", ["NAME", "NAME2"]) as cursor:
    for row in cursor:
        if row[0].find("(see") != -1:
            idx = row[0].find("(see")
            row[1] = row[0][:idx]
            cursor.updateRow(row)
        elif row[0].find("(see") == -1:
            idx = row[0].find("(see")
            row[1] = row[0]
            cursor.updateRow(row)
del cursor
del row

#make all values in NAME2 upper() to match TRI data
with arcpy.da.UpdateCursor("Carcinogens_0.updated.dbf", ["NAME2"]) as cursor:
    for row in cursor:
        if any(x.islower() for x in row[0]) == True:
            row[0] = row[0].upper()
            cursor.updateRow(row)
del cursor
del row

#Clean up willcoTRI*.dbf (7 TRI tables) using List and cursor:

#Use ListTables to run a batch script on TRI tables
#Add field for modified name
TRItables = arcpy.ListTables("willcoTRI*.dbf")
for TRItable in TRItables:
    arcpy.AddField_management(TRItable, "CHEMNAME2", "TEXT")

#set and run cursor in willcoTRI*.dbf to populate a modifed name field
    with arcpy.da.UpdateCursor(TRItable, ["CHEMNAME", "CHEMNAME2"]) as cursor:
```

SCRIPT SAMPLE 2

```
arcpy.JoinField_management(TRI1991_dbf, "CHEMNAME2", Carcinogens_0_dbf, "NAME2", ["CARCINOGEN"])

# Local variables 1992:
Carcinogens_0_dbf = "Carcinogens_0.updated.dbf"
TRI1992_dbf = "willcoTRI1992.dbf"
# Process: Join Field
arcpy.JoinField_management(TRI1992_dbf, "CHEMNAME2", Carcinogens_0_dbf, "NAME2", ["CARCINOGEN"])

# Local variables 1993:
Carcinogens_0_dbf = "Carcinogens_0.updated.dbf"
TRI1993_dbf = "willcoTRI1993.dbf"
# Process: Join Field
arcpy.JoinField_management(TRI1993_dbf, "CHEMNAME2", Carcinogens_0_dbf, "NAME2", ["CARCINOGEN"])

# Local variables 1994:
Carcinogens_0_dbf = "Carcinogens_0.updated.dbf"
TRI1994_dbf = "willcoTRI1994.dbf"
# Process: Join Field
arcpy.JoinField_management(TRI1994_dbf, "CHEMNAME2", Carcinogens_0_dbf, "NAME2", ["CARCINOGEN"])


                #ATTRIBUTE PRIMARY SITE TO TRI CHEMICALS THAT ARE KNOWN CARCINOGENS#
# Local variables 1988:
Carcinogens_0_dbf = "Carcinogens_0.updated.dbf"
TRI1988_dbf = "willcoTRI1988.dbf"
# Process: Join Field
arcpy.JoinField_management(TRI1988_dbf, "CHEMNAME2", Carcinogens_0_dbf, "NAME2", ["PRM_SITE"])

# Local variables 1989:
Carcinogens_0_dbf = "Carcinogens_0.updated.dbf"
TRI1989_dbf = "willcoTRI1989.dbf"
# Process: Join Field
arcpy.JoinField_management(TRI1989_dbf, "CHEMNAME2", Carcinogens_0_dbf, "NAME2", ["PRM_SITE"])

# Local variables 1990:
Carcinogens_0_dbf = "Carcinogens_0.updated.dbf"
TRI1990_dbf = "willcoTRI1990.dbf"
# Process: Join Field
arcpy.JoinField_management(TRI1990_dbf, "CHEMNAME2", Carcinogens_0_dbf, "NAME2", ["PRM_SITE"])

# Local variables 1991:
```

SCRIPT OUTPUT SAMPLES

```
In 'Carcinogens_0.updated.dbf', New field CHEM2 and field CARCINOGEN are as follows:

1-(2-CHLOROETHYL)-3-(4-METHYLCYCLOHEXYL)-1-NITROSOUREA, Known
1-(2-CHLOROETHYL)-3-CYCLOHEXYL-1-NITROSOUREA, RAHC
1,1-DIMETHYLHYDRAZINE, RAHC
1,2,3-TRICHLOROPROPANE, RAHC
1,2-DIBROMO-3-CHLOROPROPANE, RAHC
1,2-DIBROMOETHANE, RAHC
1,2-DICHLOROETHANE, RAHC
1,3-BUTADIENE, Known
1,3-DICHLOROPROPENE, RAHC
```

TRIwithUpdatedCarc_Summed

CHEMNAME2	CARCINOGEN	PRM_SITE	PRMEXP	OTHER
SULFURIC ACID	Known	lung and bronchus	ingestion or inhalation	petroleum refining, soaps, water treatment
SULFURIC ACID	Known	lung and bronchus	ingestion or inhalation	petroleum refining, soaps, water treatment
NICKEL COMPOUNDS	Known	lung and bronchus	ingestion or inhalation	food and water, soaps
ASBESTOS	Known	lung and bronchus	ingestion or inhalation	high risk of exposure for general population
SULFURIC ACID	Known	lung and bronchus	ingestion or inhalation	petroleum refining, soaps, water treatment
SULFURIC ACID	Known	lung and bronchus	ingestion or inhalation	petroleum refining, soaps, water treatment
SULFURIC ACID	Known	lung and bronchus	ingestion or inhalation	petroleum refining, soaps, water treatment
SULFURIC ACID	Known	lung and bronchus	ingestion or inhalation	petroleum refining, soaps, water treatment
SULFURIC ACID	Known	lung and bronchus	ingestion or inhalation	petroleum refining, soaps, water treatment
1,3-BUTADIENE	Known	leukemias and lymphomas	inhalation	traffic exhaust, refineries
BENZENE	Known	leukemias and lymphomas	inhalation	traffic exhaust
NICKEL COMPOUNDS	Known	lung and bronchus	ingestion or inhalation	food and water, soaps
SULFURIC ACID	Known	lung and bronchus	ingestion or inhalation	petroleum refining, soaps, water treatment
ETHYLENE OXIDE	Known	leukemias and lymphomas	ingestion or inhalation	traffic exhaust, food, consumer products
SULFURIC ACID	Known	lung and bronchus	ingestion or inhalation	petroleum refining, soaps, water treatment
1,3-BUTADIENE	Known	leukemias and lymphomas	inhalation	traffic exhaust, refineries
BENZENE	Known	leukemias and lymphomas	inhalation	traffic exhaust

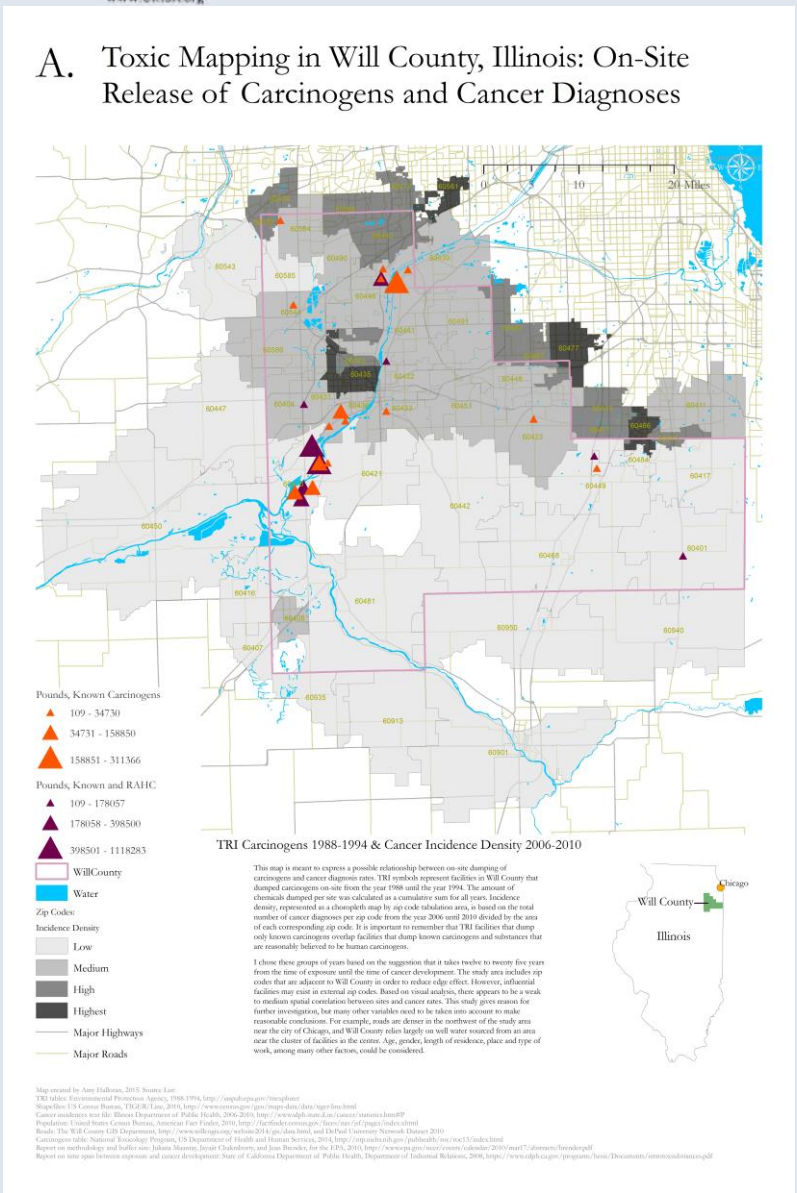
PART 2: GIS IN A MAPPING PROGRAM

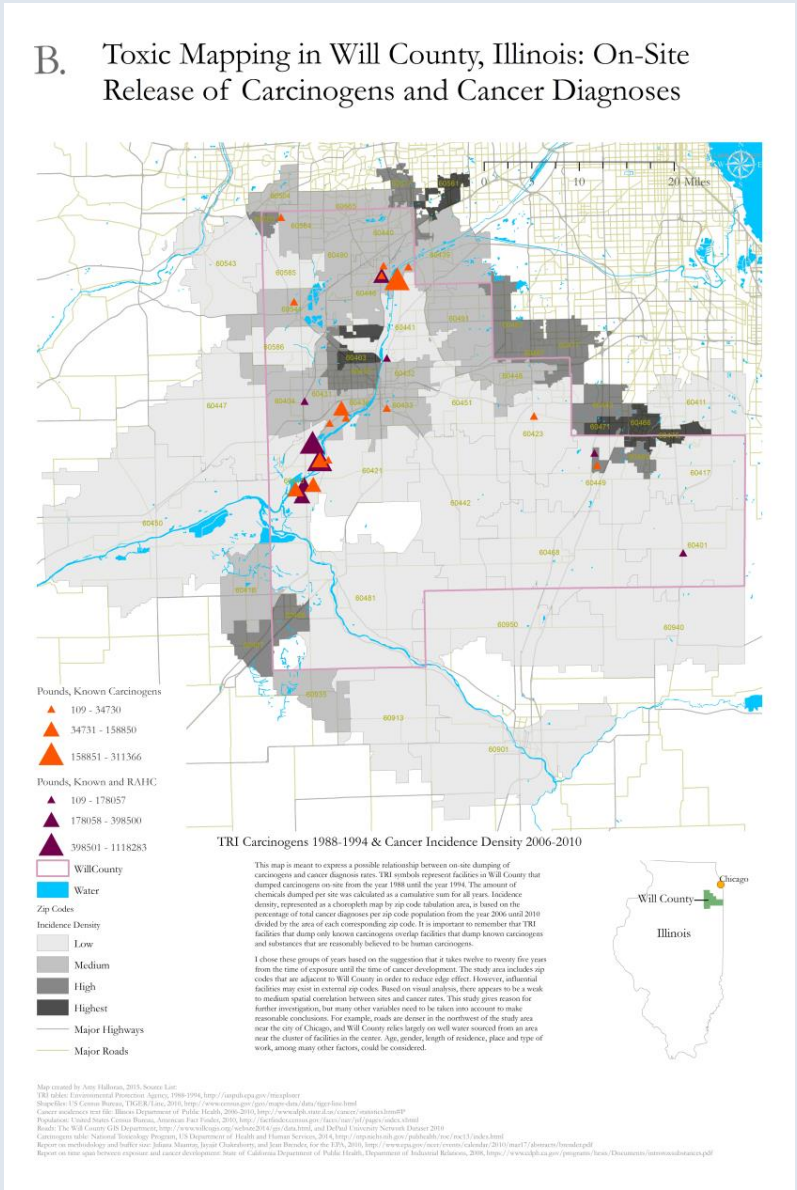
- Main tools:
  - Select by Attributes, Summarize, Field Calculator
- Add tables and shapefiles
- Summarize incidences by zip code
- Normalize incidences as choropleth maps (see density explanation in next slide)
- Select by attributes and export
  - Known
  - RAHC (Reasonably assumed to be Human Carcinogen)
  - Primary site is lung and bronchus
  - Primary site is leukemias and lymphomas
- Make roads map

A. & B. RESULTS – DENSITIES

A. Density = Number of Diagnoses/Area

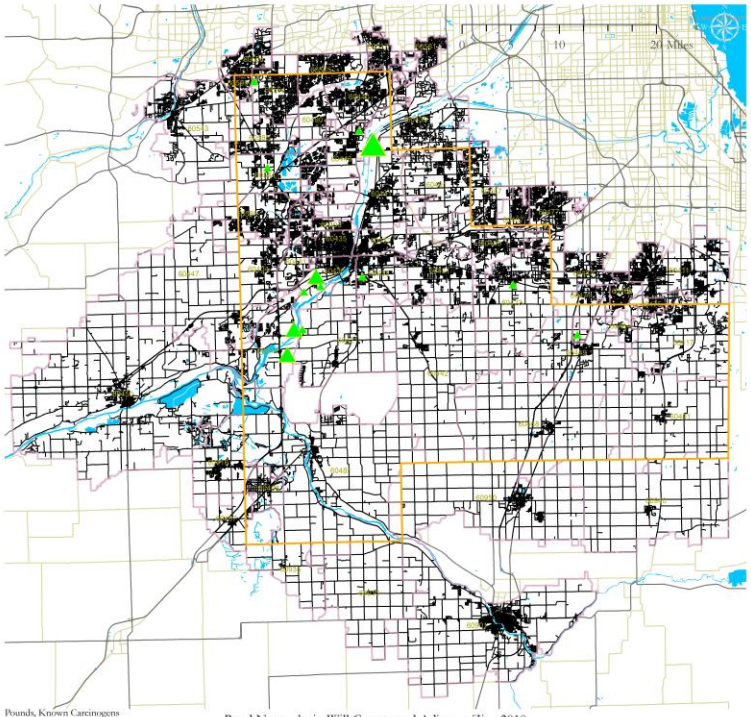
B. Density = Number of Diagnoses /Population/Area





C. RESULTS – ROADS

C. Toxic Mapping in Will County, Illinois: Streets, Major Roads, and Highways



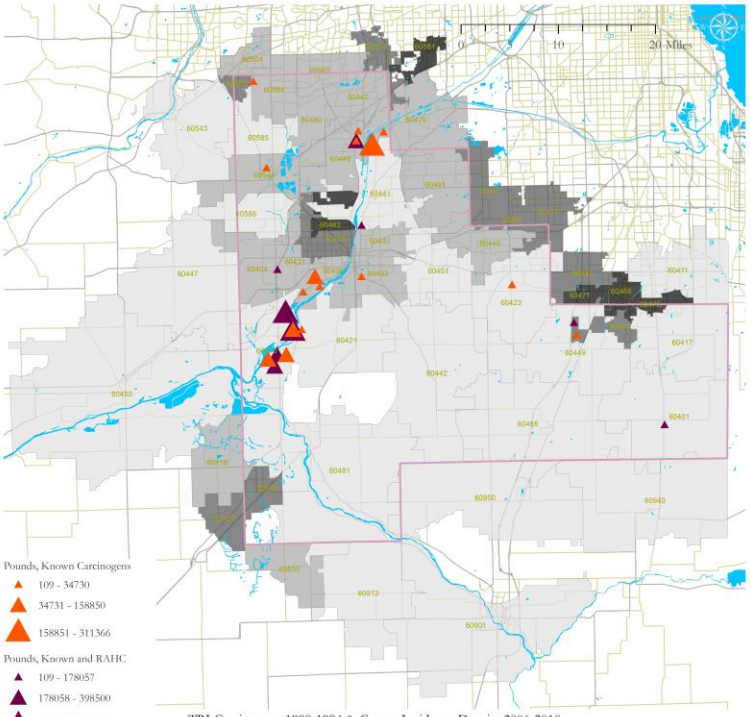
Road Networks in Will County and Adjacent Zips 2010

The table below lists known human carcinogens that are related to pavement, autos/exhaust, and oil/fuel. Records highlighted in gray are chemicals dumped by the facilities on display:

NAME	CARCINOGENICITY	PRIM SITE	SECONDARY SITE	YEARS	STATUS	PHASE
1,4-Butadiene	Known	leukemia and lymphomas				inhalation
Benzene	Known	leukemia and lymphomas				inhalation
Coal Tar (con'tar and coal Tar Pitches)	Known	lung and bronchus	urinary system			ingestion or inhalation
Coke Oven Emissions (con'tar and coal Tar Pitches)	Known	lung and bronchus	urinary system			ingestion or inhalation
Diesel Exhaust	Known	leukemia and lymphomas	whether cancers			ingestion or inhalation
Mineral Oils: Untreated and Mildly Treated	Known	lung and bronchus	urinary system			ingestion or inhalation
Soots	Known	lung and bronchus	leukemia and lymphomas			ingestion or inhalation
Strong Inorganic Acid Mists Containing Sulfuric Acid	Known	lung and bronchus				ingestion or inhalation
Sulfuric Acid (see Strong Inorganic Acid Mists Containing Sulfuric Acid)	Known	lung and bronchus				ingestion or inhalation

Visual analysis between this Map C and Maps A and B leads to a conclusion of a strong spatial relationship between road density and cancer density. Many further studies could be conducted to understand this relationship better, including a correlation analysis between cancer rates by site (location in the body) per zip code and density of roads per zip code. It could be helpful to study rates of road construction workers, commuters, and residents within close proximity to dense road networks.

B. Toxic Mapping in Will County, Illinois: On-Site Release of Carcinogens and Cancer Diagnoses



TRI Carcinogens 1988-1994 & Cancer Incidence Density 2006-2010

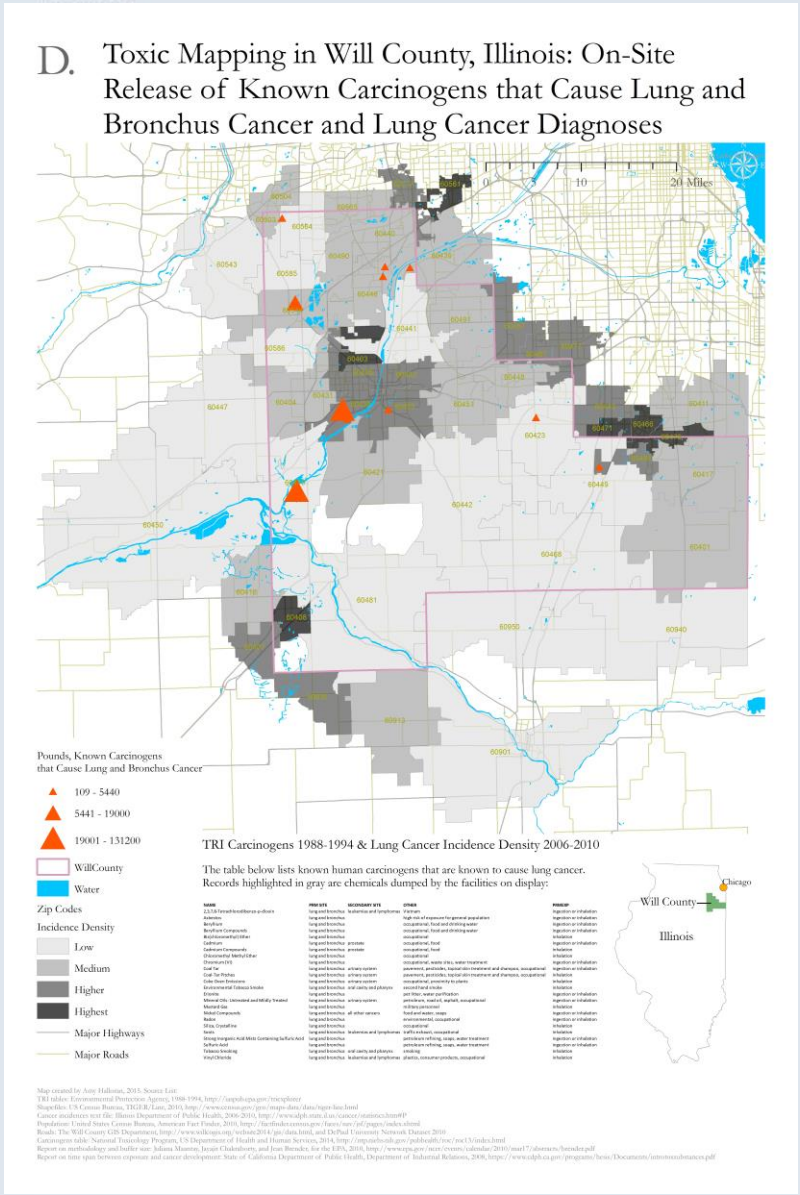
This map is meant to express a possible relationship between on-site dumping of carcinogens and cancer diagnosis rates. TRI symbols represent facilities in Will County that dumped carcinogens on-site from the year 1988 until the year 1994. The amount of chemicals dumped per site was calculated as a cumulative sum for all years. Incidence density, represented as a choropleth map by zip code tabulation area, is based on the percentage of total cancer diagnoses per zip code population from the year 2006 until 2010 divided by the area of each corresponding zip code. It is important to remember that TRI facilities that dump only known carcinogens overlap facilities that dump known carcinogens and substances that are reasonably believed to be human carcinogens.

I chose these groups of years based on the suggestion that it takes twelve to twenty five years from the time of exposure until the time of cancer development. The study area includes zip codes that are adjacent to Will County in order to reduce edge effect. However, influential facilities may exist in external zip codes. Based on visual analysis, there appears to be a weak to medium spatial correlation between sites and cancer rates. This study gives reason for further investigation, but many other variables need to be taken into account to make reasonable conclusions. For example, roads are denser in the northwest of the study area near the city of Chicago, and Will County relies largely on well water sourced from an area near the cluster of facilities in the center. Age, gender, length of residence, place and type of work, among many other factors, could be considered.

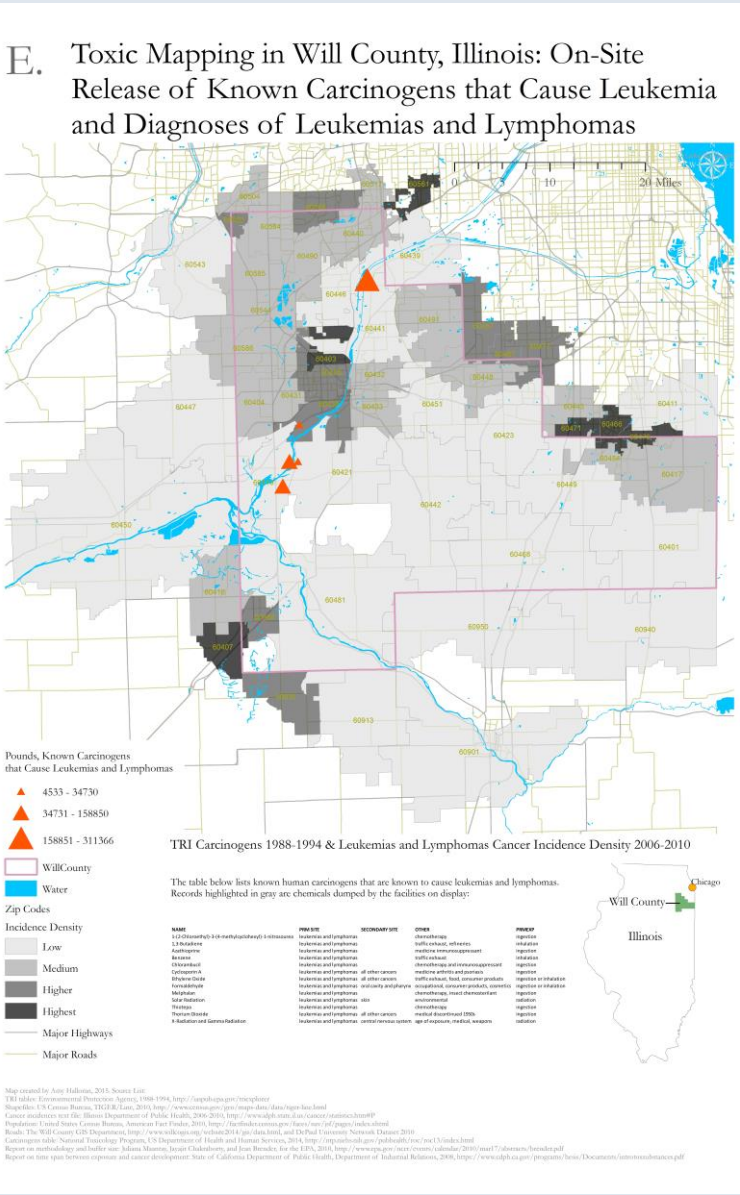
Map created by Amy Halloran, 2015. Source List:
TRI tables: Environmental Protection Agency, 1988-1994, http://iaspub.epa.gov/triexplorer/
Shapefiles: US Census Bureau, TIGER/Line, 2010, http://www.census.gov/cgi-bin/geo/shapefiles/index.php
Cancer incidences text file: Illinois Department of Public Health, 2006-2010, https://www.idph.state.il.us/cancer/statistics.htm#IP
Population: United States Census Bureau, American Fact Finder, 2010, http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml
Roads: The Will County GIS Department, http://www.willcogis.org/section2014/gis-data.html, and DePaul University Network Dataset 2010
Carcinogens table: National Toxicology Program, US Department of Health and Human Services, 2014, http://ntp.niehs.nih.gov/pubhealth/roc/roc13/index.html
Report on methodology and buffer size: Juliana Maantay, Jayait Chakraborty, and Jean Brender, for the EPA, 2010, http://www.epa.gov/ncer/events/calendar/2010/mar17/abstract/brender.pdf
Report on time span between exposure and cancer development: State of California Department of Public Health, Department of Industrial Relations, 2008, https://www.idph.ca.gov/programs/hesis/Documents/introtoxsubstances.pdf

D. & E. RESULTS – PRIMARY SITES

D. Lung and Bronchus Density



E. Leukemias and Lymphomas Density



E. Leukemias and Lymphomas Density

CONCLUSION AND SUGGESTIONS

- This study calls for further research!
- The biggest limitation is lack of access to data that is aggregated for each unit of analysis**—centroid is limiting and inaccurate**
- It would be useful to compare cancer rates of people with similar circumstances that do not live near facilities that dump carcinogens to see if the rates are lower in these areas
- How do wind, river, and ground water flow direction affect results? Suggestion-Use EPA EnviroAtlas
- How can one use this and other data to understand other variables? Suggestion-Use Census Occupation data

SOURCE LIST

- TRI tables: Environmental Protection Agency, 1988-1994, http://iaspub.epa.gov/triexplorer
- Shapefiles: US Census Bureau, TIGER/Line, 2010, http://www.census.gov/geo/maps-data/data/tiger-line.html
- Cancer incidences text file: Illinois Department of Public Health, 2006-2010, http://www.idph.state.il.us/cancer/statistics.htm#P
- Carcinogens table: National Toxicology Program, US Department of Health and Human Services, 2014, http://ntp.niehs.nih.gov/pubhealth/roc/roc13/index.html
- Population: United States Census Bureau, American Fact Finder, 2010, http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml

SOURCE LIST (CONT.)

- Roads: The Will County GIS Department, http://www.willcogis.org/website2014/gis/data.html, and DePaul University Network Dataset 2010
- Report on methodology and buffer size: Juliana Maantay, Jayajit Chakraborty, and Jean Brender, for the EPA, 2010, http://www.epa.gov/ncer/events/calendar/2010/mar17/abstracts/brender.pdf
- Report on time span between exposure and cancer development: State of California Department of Public Health, Department of Industrial Relations, 2008, https://www.cdph.ca.gov/programs/hesis/Documents/introtoxsubstances.pdf

Thank you, NWGIS, The GeoTech Center, and URISA!

Tweet about this presentation #gispro2015