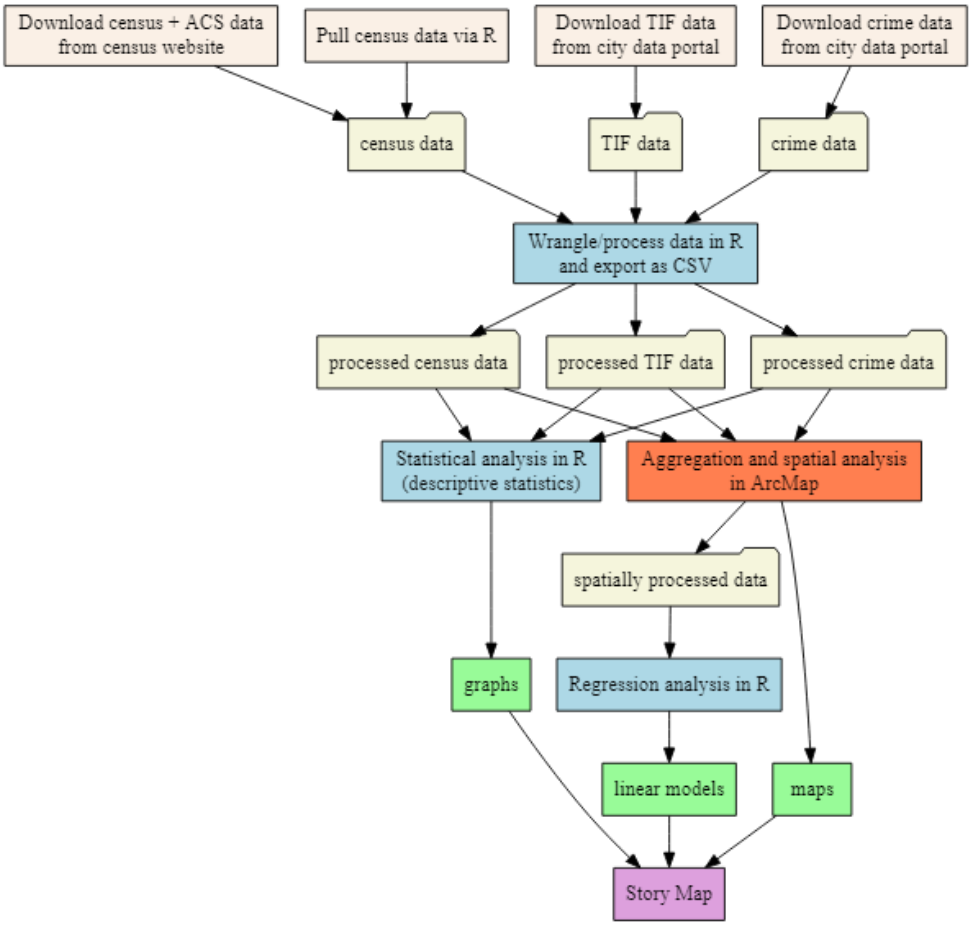
Noah Zeldin and Daniel Cheung

Project Log

Overview of Workflow: Data Processing and Analysis in ArcMap and R



Note: The R scripts contained in the folder titled *1_src* provide detailed documentation of the following aspects of the project, all of which is replicable using the included code:

- Wrangling and processing the original datasets for census, TIFs and crime
- Conducting statistical analyses of this processed data
- Building linear models and performing a regression analysis using the data that was spatially processed in ArcMap

Final Map Project Log:

1. Make another data frame for the 2010 decade.
    a. Insert → Data frame
    b. Name the first data frame 2020.
    c. Name the second data frame 2010.

2. Change the coordinate system to NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet.
    a. Right-click the 2020 data frame → Properties → Coordinate System → Search for NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet → Ok
    b. Right-click the 2010 data frame → Properties → Coordinate System → Search for NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet → Ok
3. Creating the 2020 census tract file
    a. Create a new data frame and change the coordinate system to NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet.
    b. Insert → Data frame
    c. Name the third data frame 2020CensusTractJoinCensusCrime
    d. Right-click the 2010 data frame → Properties → Coordinate System → Search for NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet → Ok
    e. Select by Location- select feature from: 2020 Illinois census tracts → Source layer: Chicago city boundaries → have their centroid in the source layer feature
    f. Right-click the 2020 Illinois census tract layer → Data → Export Data → Select Features → this layer's source data
4. Edit the Census Data to include "Census Tract" before the number in the tract column.
    a. Open the 2010 processed census data in Excel
    b. Right-click the geoid column and click insert
    c. In the inserted column put censustract on the first cell
    d. In the second cell, put ="Census Tract "&A2
    e. Left-click the lower right corner so a box appears when you hover over the cell B3, then drag down until you see the last cell for tract
    f. Repeat a-e for the 2020 processed census data.
5. Joining crime, census, and transportation data with 2020 census tracts
    a. Create a new mxd file named 2020CensusCrimewTract
    b. Name the data frame 2020CensusTractJoinCensusCrime
    c. Right-click the 2020CensusTractJoinCensusCrime data frame → Properties → Coordinate System → Search for NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet → Ok
    d. Import the 2020CensusTracts shapefile, CTA L Rail Stations shapefile, the 2020 census data, and the 2011-2020 crime data.
    e. Right-click the crime_processed_2011_2020.csv → Display XY Data → X Field: x_coordinate Y field: y_coordinate, Z field: <None> → Ok
    f. Right-click the 2020CensusTracts → Join and Relates → Join → Join data from another layer based on spatial location → displayed XY data layer for the 2011-2020 crime data → Leave the bubbles empty for the summary statistic → Create output location (name:2020CensusTractCensusCrime) → Ok
    g. Open the attribute table for 2020CensusTractCensusCrime → Add field → crimavg → double → Right-click crimavg → Field Calculator → VB Script → [Count_] / 10 → Ok

h. Right-click Count_ in the 2020CensusTractCensusCrime attribute table → Field Properties → Change alias to crime
i. Right-click 2020CensusTractCensusCrime → Join and Relates → Join → Join attributes from a table → NAMELSAD → cleancensus_processed_2020.csv → censustract → keep all records → Ok
j. Right-click the 2020CensusTracts layer → Join and Relates → Join → Join data from another layer based on spatial location → CTA_RailStations → Leave the bubbles empty for the summary statistic → Create output location (name:2020CensusTractL) → Ok
k. Change the layer name of the CTA_RailStations to JoinedCTA_RailStations
l. Open the crime_processed_2011_2020.csv → Select by Attribute → Do year = 2011 → Export the selected data into a database file.
m. Clear Selection → Repeat step i for each year from 2012 to 2020 and add each database file to the .mxd file to double-check them and clear the selection after exporting each one.
n. Right-click on the 2020CensusTractCensusCrime layer → Data → Export Data → All features → this layer's source data → name: Updated2020Data

6. Joining crime, census, and transportation data with 2010 census tracts
   a. Create a new mxd file named 2010CensusCrimewTract
   b. Name the data frame 2010CensusTractJoinCensusCrime
   c. Right-click the 2010CensusTractJoinCensusCrime data frame → Properties → Coordinate System → Search for NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet → Ok
   d. Import the ChicagoCensusTract2010 shapefile, CTA L Rail Stations shapefile, the 2010 census data, and the 2001-2010 crime data.
   e. Right-click the crime_processed_2001_2010.csv → Display XY Data → X Field: x_coordinate Y field: y_coordinate, Z field: <None> → Ok
   f. Right-click ChicagoCensusTract2010 → Join and Relates → Join → Join data from another layer based on spatial location → displayed XY data layer for the 2001-2010 crime data → Leave the bubbles empty for the summary statistic → Create output location (name:2010CensusTractCensusCrime) → Ok
   g. Open the attribute table for 2010CensusTractCensusCrime → Add field → crimavg → double → Right-click crimavg → Field Calculator → VB Script → [Count_] / 10 → Ok
   h. Right-click Count_ in the 2010CensusTractCensusCrime attribute table → Field Properties → Change alias to crime
   i. Right-click 2010CensusTractCensusCrime → Join and Relates → Join → Join attributes from a table → NAMELSAD → cleancensus_processed_2010.csv → censustract → keep all records → Ok
   j. Right-click the ChicagoCensusTract2010 layer → Join and Relates → Join → Join data from another layer based on spatial location → CTA_RailStations → Leave the bubbles empty for the summary statistic → Create output location (name:2010CensusTractL) → Ok

k. Change the layer name of the CTA_RailStations to JoinedCTA_RailStations
l. Open the crime_processed_2001_2010.csv → Select by Attribute → Do year = 2001 → Export the selected data into a database file.
m. Repeat step i for each year from 2002 to 2010 and add each database file to the .mxd file to double-check them.
n. Right-click on the 2010CensusTractCensusCrime layer → Data → Export Data → All features → this layer's source data → name: Updated2010Data

7. Separating TIF Data by year.
   a. Create a new mxd file named TIFData
   b. Name the data frame TIF
   c. Right-click the TIF data frame → Properties → Coordinate System → Search for NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet → Ok

- Geocoding missing value for 1500 S California Ave Chicago, Illinois address.
  a. Put the misspelled address into a .csv file by itself. → Put Chicago, Illinois at the end of the address to make it easier to recognize.
  b. Create a new mxd file named GeocodingMissingAddress
  c. Right-click the data frame → Properties → Coordinate system → Search for NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet → Ok
  d. Import the .csv file with the missing address
  e. Right-click the .csv file → Geocode Addresses → ArcGIS World Geocoding Service → Ok → Sign in → Multiple Fields → Address or Place: Address column in the .csv file and change the output location and name → Advanced Geometry → Use a different spatial reference: GCS_North_American_1983 → Ok → Ok
  f. Open Attribute Table → Attribute Table Options → Add Field → X with Type: Double → Right Click field → Calculate Geometry → X coordinate of point → Use the coordinate system of the data frame → Units: Feet US [ft]
  g. Add Field → Y with Type: Double → Calculate Geometry → Y coordinate of point → Use the coordinate system of the data frame → Units: Feet US [ft]
  h. Table to Excel conversion tool to get a .csv file to add it back into the TIF project .csv file
  - Join the missing value back in the TIF data.
    a. Copy the missing values which were the X and Y coordinates, the latitude and longitude data, and the location into the 1500 S California Ave Chicago, Illinois row.

d. Open the tif_processedwmissingvalue.csv → Select by Attribute → Do coc_year <= 2000 → Export the selected data into a database file.
e. Repeat step d for each year from 2001 to 2020 except coc_year = 20## and add each database file to the .mxd file to double-check them.
8. Displaying 2020 Census and TIF Data

a. Create a new mxd file named 2020CensusTIF
b. Name the data frame 2020CensusTIF
c. Right-click the 2020CensusTIF data frame → Properties → Coordinate System → Search for NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet → Ok
d. Import Updated2020Data.shp, 2020CensusTractL.shp, CTA_RailLines.shp, CTA_RailStations.shp, TIFBaseline.dbf, TIF2001.dbf-TIF2010.dbf, Crime2011.dbf-Crime2020.dbf, and CTALLinesWithSymbology.lyr

- The CTALLinesWithSymbology.lyr was made in the ProgressReport
    a. Change the symbology for the CTA Rail Lines → Right-click CTA_RailLines → Properties → Symbology → Categories → Value Field: Lines → Add All Values → Click the colors for each of the types of lines → Symbol Selector using the appropriate color and railroad symbol for each of the types of lines (Black is used for railroads that have multiple lines running
    b. Save the CTA Rail Lines Layer with the line symbol and color symbology as a Layer by right-clicking the layer and pressing "Save As Layer File"

e. Right-click Updated2020Data → Join and Relates → Join → Join attributes from a table → NAMELSAD → 2020CensusTractL → NAMELSAD → Keep all records → Ok
f. Uncheck the CTA_RailLines layer
g. Change the symbology points for the L Stations Locations to the Train Station symbol → Click symbology type under the layer name to get the Symbol Selector dialog window → Search "train" → Under Civic → Train Station, Size: 4
h. Open the Updated2020Data attribute table and change the Count_ alias to lstatnum
i. Open the Updated2020Data attribute table → Add field → pop (Population) → Double → Right-click pop → Field calculator → Yes → VB Script → [Updated2020Data.population] → Ok
j. Open the Updated2020Data attribute table → Add field → cravpopnum (The average count of crime per year in the 2020 decade normalized by population) → Double → Right-click cravpopnum → Field calculator → Yes → VB Script → [Updated2020Data.crimavg] / [Updated2020Data.pop] → Ok
k. Open the Updated2020Data attribute table → Add field → cocrpopnum (The count of property crime in the 2020 decade normalized by population) → Double → Right-click cocrpopnum → Field calculator → Yes → VB Script → [Updated2020Data.Count_] / [Updated2020Data.pop] → Ok
l. Open the Updated2020Data attribute table → Add field → medinc (Median Income) → Double → Right-click medinc → Field calculator → Yes → VB Script → [Updated2020Data.income_med] → Ok
m. Right-click Updated2020Data → Properties → Symbology → Quantities → Graduated Colors → Value: Updated2020Data.medinc → Color Ramp: Green Light to Dark → Classify → Change classes to 6 → Change Method to Manual → Change Break Values

to: 0, 35417.000000, 59826.000000, 89196.000000, 129583.000000, and 216667.000000 → Change color for 0 to No Color

n. Right-click Updated2020Data → Copy → Paste → Right-click the pasted → Properties → Symbology → Quantities → Graduated Symbols → Value:Updated2020Data.cravpop → Classify → Classes:5 → Ok → Background → No color for fill or outline → Ok → Symbol Size from: 4 to: 18 → Template → Circle 2 → Color → Medium Coral Light → Ok → Ok

o. Open the Updated2020Data attribute table → Add field → unemrate (Unemployment Rate) → Double → Ok → Right-click Updated2020Data.unemrate → Field Calculator → Yes → [Updated2020Data.unem_rate] → Ok → Yes

p. Right-click Updated2020Data → Copy → Paste → Right-click the pasted → Properties → Symbology → Quantities → Graduated Colors → Value:Updated2020Data.unemrate → Color Ramp: Blue Light to Dark → Classify → Change classes to 6 → Change Method to Manual → Change Break Values to: 0, 0.032000, 0.063000, 0.103000, 0.154000, and 0.270000 → Right-click Symbol → Flip Symbols → Ok → Change color for 0 to No Color

q. Right-click TIF2011 → Display XY Data → X Field: x_coordina Y Field: y_coordina Z Field: <None> → Ok → Change TIF2011 Events symbology → Triangle 2 → Ok

r. Uncheck TIF2011 Events

s. Repeat q and r for TIF2012-TIF2020 and TIFBaseline

t. Check TIF2011 Events → Geoprcessing → Buffer → Input → TIF2011 Events → Linear Unit: 1 mile → Ok

u. Repeat s for 2 miles → Change the color for the symbology to 30% gray → Adjust Transparency to 42%

v. Repeat t and u for TIF2012 Events-TIF2020 Events and TIFBaseline Events

w. Uncheck 2020CensusTractL

x. Change the name for TIFBaseline's buffer and events to TIF2000Baseline

y. Finish step 9 then come back to the 2020CensusTIF mxd file

z. Import the 2010 TIF data csv, 1 and 2 mile buffer, and events

a. Change the color for the 2010 TIF 1 and 2 mile buffer to 30% and Adjust Transparency to 42%

b. Change the name of the 2010 TIF 1 and 2 mile buffers to TIF2010BaselineBuffer1m and TIF2010BaselineBuffer2m

c. Right-click the TIF2010 csv → Display XY Data → X Field: x_coordina Y Field: y_coordina Z Field: <None> → Ok → Change the name of TIF2010 Events to TIF2010Baseline Events → Change the TIF2010Baseline Events Symbology to Triangle 2 → Ok

9. Displaying 2010 Census and TIF Data

a. Create a new mxd file named 2010CensusTIF

b. Name the data frame 2010CensusTIF

c. Right-click the 2010CensusTIF data frame → Properties → Coordinate System → Search for NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet → Ok
d. Import Updated2010Data.shp, 2010CensusTractL.shp, , CTA_RailLines.shp, CTA_RailStations.shp, TIFBaseline.dbf, TIF2001.dbf-TIF2010.dbf, Crime2001.dbf-Crime2010.dbf, and CTALLinesWithSymbology.lyr

- The CTALLinesWithSymbology.lyr was made in the ProgressReport
    a. Change the symbology for the CTA Rail Lines → Right-click CTA_RailLines → Properties → Symbology → Categories → Value Field: Lines → Add All Values → Click the colors for each of the types of lines → Symbol Selector using the appropriate color and railroad symbol for each of the types of lines (Black is used for railroads that have multiple lines running
    b. Save the CTA Rail Lines Layer with the line symbol and color symbology as a Layer by right-clicking the layer and pressing "Save As Layer File"

e. Right-click Updated2010Data → Join and Relates → Join → Join attributes from a table → namelsad10 → 2010CensusTractL → namelsad10 → Keep all records → Ok
f. Uncheck the CTA_RailLines layer
g. Change the symbology points for the L Stations Locations to the Train Station symbol → Click symbology type under the layer name to get the Symbol Selector dialog window → Search "train" → Under Civic → Train Station, Size: 4
h. Open the Updated2010Data attribute table and change the Count_ alias to lstatnum
i. Open the Updated2010Data attribute table → Add field → pop (Population) → Double → Right-click pop → Field calculator → Yes → VB Script → [Updated2010Data.population] → Ok
j. Open the Updated2010Data attribute table → Add field → cravpopnum (The average count of crime per year in the 2010 decade normalized by population) → Double → Right-click cravpopnum → Field calculator → Yes → VB Script → [Updated2010Data.crimavg] / [Updated2010Data.pop] → Ok
k. Open the Updated2010Data attribute table → Add field → cocrpopnum (The count of property crime in the 2010 decade normalized by population) → Double → Right-click cocrpopnum → Field calculator → Yes → VB Script → [Updated2010Data.Count_] / [Updated2010Data.pop] → Ok
l. Open the Updated2010Data attribute table → Add field → medinc (Median Income) → Double → Right-click medinc → Field calculator → Yes → VB Script → [Updated2010Data.income_med] → Ok
m. Right-click Updated2010Data → Properties → Symbology → Quantities → Graduated Colors → Value: Updated2010Data.medinc → Color Ramp: Green Light to Dark → Classify → Change classes to 6 → Change Method to Manual → Change Break Values to: 0, 30051.000000, 46101.000000, 65058.000000, 89960.000000, and 151250.000000 → Change color for 0 to No Color

n. Right-click Updated2010Data → Copy → Paste → Right-click the pasted → Properties → Symbology → Quantities → Graduated Symbols → Value:Updated2010Data.cravpop → Classify → Classes:5 → Ok → Background → No color for fill or outline → Ok → Symbol Size from: 4 to: 18 → Template → Circle 2 → Color → Medium Coral Light → Ok → Ok
o. Open the Updated2010Data attribute table → Add field → unemrate (Unemployment Rate) → Double → Ok → Right-click Updated2010Data.unemrate → Field Calculator → Yes → [Updated2010Data.unem_rate] → Ok → Yes
p. Right-click Updated2010Data → Copy → Paste → Right-click the pasted → Properties → Symbology → Quantities → Graduated Colors → Value:Updated2010Data.unemrate → Color Ramp: Blue Light to Dark → Classify → Change classes to 6 → Change Method to Manual → Change Break Values to: 0, 0.048000, 0.086000, 0.136000, 0.219000, and 0.397000 → Right-click Symbol → Flip Symbols → Ok → Change color for 0 to No Color
q. Right-click TIF2002 → Display XY Data → X Field: x_coordina Y Field: y_coordina Z Field: <None> → Ok → Change TIF2002 Events symbology → Triangle 2 → Ok
r. Uncheck TIF2002 Events
s. Repeat q and r for TIF2003-TIF2010 and TIFBaseline
t. Check TIF2002 Events → Geoprcessing → Buffer → Input → TIF2002 Events → Linear Unit: 1 mile → Ok
u. Repeat s for 2 miles → Change the color for the symbology to 30% gray → Adjust Transparency to 42%
v. Repeat t and u for TIF2003 Events-TIF2010 Events
w. Uncheck 2010CensusTractL
x. Change the name for TIFBaseline's buffer and events to TIF2000Baseline

10. Creating decade buffers
    a. Create a new mxd file named TIFDecade
    b. Import tif_processedwmissingvalue.csv
    c. Right-click the layer data frame → Properties → Coordinate System → Search for NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet → Ok
    d. Right-click tif_processedwmissingvalue.csv → Display XY Data → X Field: x_coordinate Y Field: y_coordinate Z Field: <None> → Ok → Ok
    e. Right-click tif_processedwmissingvalue.csv Events → Open Attribute Table → Select by Attributes → coc_year = 2002 OR coc_year = 2003 OR coc_year = 2004 OR coc_year = 2005 OR coc_year = 2006 OR coc_year = 2007 OR coc_year = 2008 OR coc_year = 2009 OR coc_year = 2010 → Close the Select by Attributes → Left-click the top left of the Attribute Table → Export → Selected Records → 2010DecadeTIFProjects → Ok → Yes
    f. Clear Selection in the tif_processedwmissingvalue.csv Events Attribute Table

g. Select by Attributes → coc_year = 2011 OR coc_year = 2012 OR coc_year = 2013 OR coc_year = 2014 OR coc_year = 2015 OR coc_year = 2016 OR coc_year = 2017 OR coc_year = 2018 OR coc_year = 2019 OR coc_year = 2020 → Close the Select by Attributes → Left-click the top left of the Attribute Table → Export → Selected Records → 2020DecadeTIFProjects → Ok → Yes
h. Save

11. Regressions
    a. Open 2020CensusTIF.mxd → Import → tif_processedwmissingvalue.csv → Display XY Data → X Field: x_coordinate Y Field: y_coodinate Z Field: <None> → Ok → Ok
    b. Right-click Updated2020Data → Join and Relates → Join → Join data from another layer based on spatial location → tif_processedwmissingvalue.csv → Leave the bubbles empty for the summary statistic → Create output location (Name:Regression2010) → Ok
    c. Open the Regression2020 attribute table → Change the alias of Count_1 to numtif
    d. Uncheck Regression2020 and tif_processedwmissingvalue.csv Events
    e. Save as → Regression2020.mxd
    f. Open the Regression2020 attribute table → Add field → lesshs → Double → Right-click lesshs → Field Calculator → Yes → [edu_less_h] → Ok → Yes
    g. Repeat f for the rest of the education variables
    h. With the Regression2020 attribute table open → Add field → timetowrk → Double → Ok → Right-click timetowrk → Field Calculator → Yes → [time_to_wo] → Ok → Yes
    i. Import the 2020DecadeTIFProjects → Right-click it → Display XY Data → X Field: x_coordina Y Field: y_coordina Z Field: <None> → Ok
    j. Geoprocessing → Buffer → 2020DecadeTIFProjects Events → 2020DecadeTIFProjects1mBuffer → 1 mile → Ok → Import the buffer → Spatial Join with Regression2020 → Regression20201mspillover → Ok → Uncheck the 2020DecadeTIFProject1mBuffer → Change the alias of the last column in the Regression20201mspillover attribute table to 1mspill
    k. Geoprocessing → Buffer → 2020DecadeTIFProjects Events → 2020DecadeTIFProjects2mBuffer → 2 mile → Ok → Import the buffer → Spatial Join with Regression20201mspillover → Regression20202mspillover → Ok → Uncheck the 2020DecadeTIFProject2mBuffer → Change the alias of the last column in the Regression20202mspillover attribute table to 2mspill
    l. Click Search → Table to Excel conversion tool → Regression20202mspillover → Use field alias as column header → Ok
    m. Open the xls file and change the Count_1 columns to lstatnum and numtif in that order → Save
    n. Change the last two columns to 1mspill and 2mspill respectively
    o. Click Search → Table to Excel conversion tool → 2020CensusTractL → Use field alias as column header → Ok

p. Open the 2020CensusTractL xls file and copy the last column and paste that into the Regression2020 column labelled lstatnum which is the second to last column → Save
q. Repeat a-p for 2010CensusTIF.mxd and the 2010 data except create a field in the 2010 data for ALAND and calculate the land using calculate geometry in square meters

12. Spillover Maps
    a. In Regression2010.mxd, uncheck the 2010 decade TIF projects and the 2010 decade TIF project buffers
    b. Symbolize Regression20102mspillover for 1mspill and 2mspill with Natural Breaks 5 classes Color Ramp: Gray Light to Dark → Check Regression20102mspillover and uncheck the unemployment rate Updated2010Data
    c. Go to layout view → Resize the map to fill the whole page (Scale: 1:165,000) → Add a legend → Change the legend labels → Export Map
    d. Repeat t with the 2-mile spillover
    e. Save project as 2010StoryMaps
    f. Repeat u for 2020 except with a different Color Ramp
13. TIF Project Buffer, Unemployment, and Median Income Maps
    a. Save the 2010StoryMaps and 2020StoryMaps as 2020TIFBufferMaps.mxd and 2010TIFBufferMaps.mxd
    b. For the 2020TIFBufferMaps, import all the TIF project .dbf files → Display XY Data for each of the imported TIF project .dbf files → Create a different symbology for those prior to 2011 and make a different symbology for 2010 and 2000 TIF project .dbf files
    c. Import the TIF project buffers for each of the imported TIF project
    d. Turn on all of the 2011-2020 TIF projects with the cravpopnum symbolized for the census tracts → Change the symbology for the TIF projects to be smaller with a more monotone color
    e. Make one map with the buffers and one map without the buffers with bigger symbology
    f. Then uncheck the TIF project buffers, check CTA rail stations, check CTA rail lines, and check the unemrate layer
    g. Symbolize unemrate with Natural Breaks 5 classes
    h. Change the graduated color symbology for cravpopnum to graduated symbols → Put NULL for the first class, 0 value
    i. Put the necessary map material like North Arrow, Legend, Text, and Title
    j. Export the map
    k. Uncheck the unemrate layer and check the medinc layer
    l. Change the 0 value to NULL and the values into dollar values with commas after the thousands
    m. Fix the map material to appropriately match
    n. Export the map
    o. Repeat d-n with 2010 data to make 2010 versions of the maps
14. TIF and Percentage Change in Crime over the 2010 to the 2020 Decade

a. Open 2010TIFBufferMaps.mxd → Export Avg Yearly Crime Count / Pop from Attribute Table → Open 2020TIFBufferMaps.mxd → Import the 2010 Crime Data dbf from 2010TIFBufferMaps.mxd → Delete any unnecessary variables → Add Field → 2010crimed → Double → Join the 2010 Crime Data dbf with the Avg Yearly Crime Count / Pop layer with the census tract color shading normalization → Export the attribute table → Calculating [cravpopnum_2020 - cravpopnum_2010] / cravpopnum_2010 in Excel → Save the Excel file as a csv to bring it back into 2020TIFBufferMaps.mxd → Do a table join with the Avg Yearly Crime Count / Pop layer with the census tract color shading normalization → Classify the percentage change in the average yearly count of property crime normalized by population from 2010 to 2020 with Natural Breaks 5 classes → Go the layout view → Setup the map → Export the map
b. Import all of the 2002-2010 TIF project dbf files → Display XY Data for each → Change symbology to circle 2, Beryl Green, and 10 size
c. Make 2 more maps → 1 with all TIFs from 2001 to 2020 separated by 2001-2010 and 2011-2020 → 1 with TIFs only from 2001-2010 → Export those maps

Literature Review

1.
Lester, T. William. "Does Chicago's Tax Increment Financing (TIF) Programme Pass the 'But-for' Test? Job Creation and Economic Development Impacts Using Time-Series Data." Urban Studies 51, no. 4 (2014): 655–74.
This paper examines the effect of TIF programs on economic development, specifically outcomes related to the labor market and real estate. Alarmingly, the author finds that TIF projects
ultimately did not have any economic development for the residents of the areas where they were constructed. Clearly, this study of TIFs in Chicago is closely related to our own project, except that we examine their effect on crime.
The author uses block-level data for TIF projects from the city of Chicago's open data portal, census data for the census block groups, building permits records from the Chicago Metropolitan Agency for Planning, TIF expenditure data from UIC's Center for Urban Economic Development,
and employment and business-level data from the National Establishment Time-series database. All of the data sources used were suitable for the research question and the controls for the analysis and much of it is applicable for our research as well.
One of the main GIS tools used in the paper is the spatial join. The author developed accompanying rules to determine the threshold of considering whether census a block was "treated," i.e. affected by a TIF project. (For instance, 50% or more of a census block's population

has to be within a TIF boundary.) Similarly, we had to determine tract-level metrics for determining exposure to TIF projects and crime and used spatial joins for 2020 census tracts, crime, and L stations.
GIS contributed to understanding the type of treatment in each location and the guidelines set for determining treatment. Without these guidelines, it would be hard to determine a consistent level of treatment with the multiple TIF projects and districts in some census blocks. GIS was used by the author to see which districts the treatment guidelines applied to. The author used a map to show the type of TIF district, which was effective at showing the treatment and control group locations in Chicago.

2.
Wang, Fahui, and W. William Minor. "Where the Jobs Are: Employment Access and Crime Patterns in Cleveland." Annals of the Association of American Geographers 92, no. 3 (2002): 435–50.
This paper investigates the effect of access to jobs on crimes, which was the original topic of our project before we decided to focus instead on TIF projects. To measure this effect, the authors
use two different approaches: 1) the ratio of jobs to workers in an area and 2) a gravity model that
looks at changes in employment access and crime as the outcome.
The main data sources used were the CDC Cleveland 1990 central business district, traffic analysis zone data, crime data from the National Archive of Criminal Justice Data, communication
data, 1990 census transportation planning package (from the Bureau of Transportation Statistics). Additionally, the GIS data came from the Geolytics for the 1980 census tracts and the traffic analysis zone data came from the Environmental System Research Institute. All the data sources used in the paper are suitable for the research question as they are relevant to crime and travel, which is necessary to address the question. Although the types of data would be applicable to our project, they obviously correspond to an earlier area and a different city.
GIS tools used in the paper include classifying, normalizing and areal interpolation. Some of the maps in the paper classify the ratio of jobs to workers in an area, while others display the gravity model index and normalized vehicle theft rates. This relates to our project because we normalize property crime rates by tract population. The size of a census tract can be misleading when it comes to the count of crime as a sparsely populated census tract will not have as much crime within it as a small census tract with a higher population density on average.
GIS contributed to the understanding of the spatial dimensions of the topic by showing transit features, e.g. highways, geographically and classifying findings by census tract. The authors effectively used the maps to classify areas by type and associated statistics, like the ratio of jobs to workers in a residency, the gravity model index score and the vehicle thefts.

3.
Zhang, Haifeng, and Wei Song. “Addressing Issues of Spatial Spillover Effects and Non-Stationarity in Analysis of Residential Burglary Crime.” GeoJournal 79, no. 1 (2014): 89–102.
This paper examines whether spillovers from crime and other neighborhood characteristics have an effect on burglary. The authors use two approaches to address this research question: 1) a GIS approach to normalize the data geographically and 2) a geographically weighted regression. This relates to our project which considers the spillover effects in crimes and TIF projects on neighboring census tracts and uses spatial data in a regression analysis. This paper helps address this concern where there is a certain radius that contains a majority of the spillover effects a TIF project can have on neighboring census tracts.
The main data sources used are the Louisville Metro Police Department for burglary data in a three-year period beginning in 2007 and the 2000 census for the census block group data. Neighborhood boundary and sociodemographic data are also used in this paper. All of these data sources are suitable and necessary for the research question as they all pertain to burglary crime data and location data for the geographic unit of analysis of census block groups. For our project, these data sources would not necessarily be suitable due to a different geographic location
of Chicago and a different unit of analysis of census tracts.
The GIS methods used included displaying statistics like counts of burglary by census block group, normalizing data, and geocoding burglary locations. These methods are applicable to our project because we are looking at similar measures of counts and normalizing data by the population in a census tract. We also used the same method for geocoding Chicago’s property crimes.
GIS contributed to understanding and seeing the regression results geographically, normalizing data, and visualizing trends. The author effectively used maps to show the research findings geographically. The side-by-side graphs for the coefficient estimates, t-statistics and normalized data help determine if there are any anomalies and which estimates are statistically significant.

Regressions:

The Relationship between TIFs and Property Crime Rates

	Avg. Tract-Level Property Crime Rate, 2011-20 (per 10k residents)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Number of TIFs	217.213***	201.488***	52.768***	52.912***	52.953***	50.339***	36.391***	38.369***
	(20.995)	(21.021)	(13.089)	(13.154)	(13.162)	(13.327)	(12.934)	(12.987)

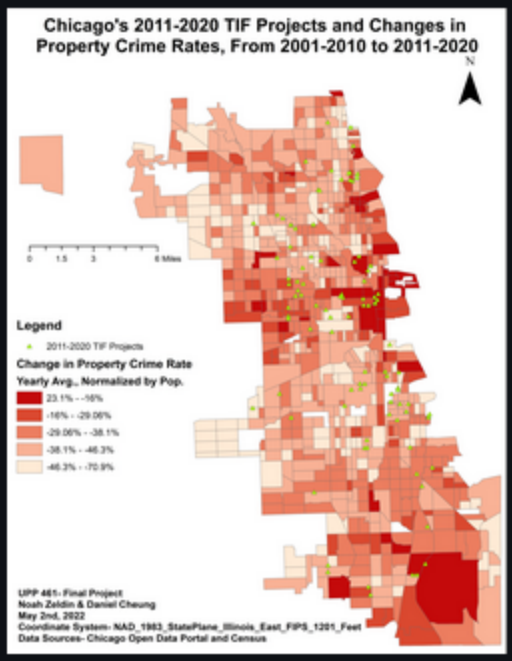
Additional TIFs (1 mile)		12.840***		-0.296	0.051	-0.076	-4.128	-4.395
		(4.552)		(2.513)	(2.738)	(2.737)	(2.682)	(2.674)
Additional TIFs (2 miles)		6.223**			-0.547	-0.382	-3.518**	-2.694
		(2.846)			(1.708)	(1.755)	(1.680)	(1.702)
Avg. prop. crime rate, 2001-10 (per 10k residents)			0.628***	0.629***	0.629***	0.636***	0.647***	0.637***
			(0.017)	(0.017)	(0.017)	(0.018)	(0.017)	(0.018)
Median income						-0.0004		-0.001**
						(0.0003)		(0.0002)
Unemployment rate						-344.500*		76.500
						(207.755)		(204.184)
Time to work (min.)							-10.451***	-11.339***
							(1.324)	(1.386)

Number of L stations							27.929*	28.518*
							(14.724)	(14.699)
Observations	769	769	769	769	769	769	769	769
Adjusted R2	0.121	0.143	0.696	0.695	0.695	0.695	0.720	0.722
Note:								*p<0.1; p<0.05;* p<0.01

TIFs and Property Crimes in Chicago

Daniel Cheung and Noah Zeldin

May 3, 2022



Preview of main finding

Introduction

Project Goal

Examine the relationship between **TIF projects** and **property crime rates** in Chicago.

TIF or Tax Increment Financing

City vehicle for funding projects that contribute to a certain cause, like school and economic development. TIFs can be classified based on intended societal contribution (e.g. social or economic) or by funding source (public or private).

Crime and Economic Development

There is a well-established relationship between crime and economic factors. Property crimes, especially theft, are likely related to economic factors, as individuals might engage in property crimes for financial reasons.

Hypothesis: There is a relationship between **economic**

Hypothesis: There is a relationship between **economic development via TIF projects** and **property crime rates.** (Note: To maintain an agnostic approach to research, we allow for the possibility of either a positive or negative relationship.)

Workflow

Data

Sources

Our main data sources were the Census and the Chicago Open Data Portal. Data from the Decennial Census and American Community Survey were used for tract-level socioeconomic, transportation, and population data for the period between 2001 and 2020. Crime data, GIS shapefiles for Chicago's city boundary and the locations for CTA L stations and lines were obtained from the Chicago Open Data Portal. Chicago's government website also served as a reference for a better understanding of the TIF data. Lastly,

reference for a better understanding of the TIF data. Lastly, the Chicago Police Department's website was used as a reference for how crimes are classified, e.g. as property crimes.

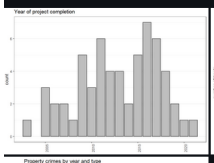
Main Variables

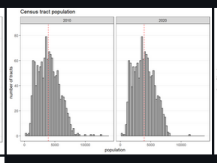
The main variables we used were completed TIF projects, property crime counts, mean travel time to work, median income and unemployment. While the latter three were obtained from the census and were already aggregated at the tract level, we had to use GIS tools to obtain tract-level metrics for TIF projects and property crimes.

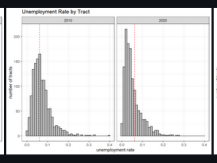
Organization

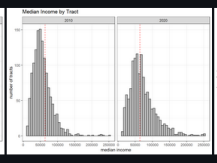
We separate our data by decade, 2001-2010 and 2011-2020, in order to examine city-wide trends in the relationship between TIF projects and property crimes. Census tracts are the graphic unit of analysis.

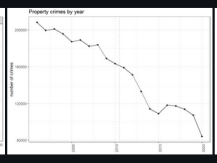
Descriptive Statistics (selection)

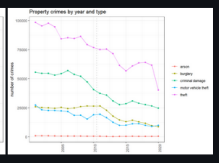


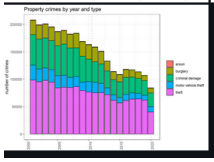












Important findings (issues and hypotheses discussed later):

There were 61 completed TIF projects between 2001-2011 and 102 completed between 2011-2020. This creates issues comparing them side-by-side as our analysis looks at the count of TIFs. A decrease in overall property crime was seen especially in the amount of theft. There were positive changes in socioeconomic factors over time with a general decrease in unemployment and an increase in median income.

GIS Techniques and Tools

Overview

GIS is necessary for our analysis of the relationship between TIF projects and property crime rates because our main variables needed to be aggregated at the tract level in order to run our regression analysis. GIS allows us to plot TIF projects, CTA L stations and lines, and crime locations, which, in turn, allowed us to normalize and calculate data, e.g. normalizing crime rates by census tract population. In addition, it allowed us to visualize relationships between economic variables, like median income and unemployment, and TIF projects and property crimes. and census tract boundaries spatial. Importantly, GIS also allowed us to examine potential spillover effects from TIF projects by creating new variables for the number of projects located in neighboring tracts.

Difficulties we encountered included processing data and determining which socioeconomic variables to use in GIS analysis. Additionally, using natural breaks to classify these variables was problematic, as the scales often differed by decade, thereby preventing an apples-to-apples comparison between maps.

Main Techniques and Tools

Data Processing in ArcMap

Table joins were used for data processing by aggregating variables together to normalize or create new variables using the Field Calculator tool. Spatial Joins were used to aggregate crime, TIF, TIF buffers, and CTA L stations as counts for variables. The aggregated crime counts were divided by 10 since the counts were for the whole decade and then normalized by the population. Select by Location was used for creating Chicago's 2020 census tracts by selecting the census tracts that had a centroid within Chicago's city border and then exporting the selected census tracts into a shapefile. Select by attribute was used to separate the data by year and decade. The Display XY Data tool was used to plot crime and TIF location data to be analyzed. Table to Excel was used to data processed in ArcMap into Excel and R to be processed further and for regressions. Geocoding was used to correct a TIF project that was missing x and y coordinates, longitude, and latitude. ArcGIS World Geocoding Service was used for the geocode. NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet was the projected coordinate system used for all the maps. GCS_North_American_1983 was used once in the project as a spatial reference for the geocode. The calculate geometry tool was used to calculate the x and y coordinates for the TIF project with missing values and the ALAND values for the 2010 census tracts.

Data Processing in ArcMap Summary

- Join / Spatial Join
  - used for data processing, plotting, and normalizing data
- Select by Attribute / Location
  - separating data by year

    - creating Chicago's 2020 census tracts
- Display XY Data
    - plotting crime and TIF data
- Table to Excel
    - used for processing data in Excel and R
- Geocoding
    - missing value for TIF project
    - correct miscoded value for TIF project
- Coordinate Systems
    - 2 were used
- Calculate Geometry
    - missing TIF value's coordinates
    - Chicago Census Tracts' 2010 ALAND

Spatial Analysis

One- and two-mile buffers were used to capture potential spillover effects of TIF projects. These newly created variables were incorporated into our regression analysis.

Symbology

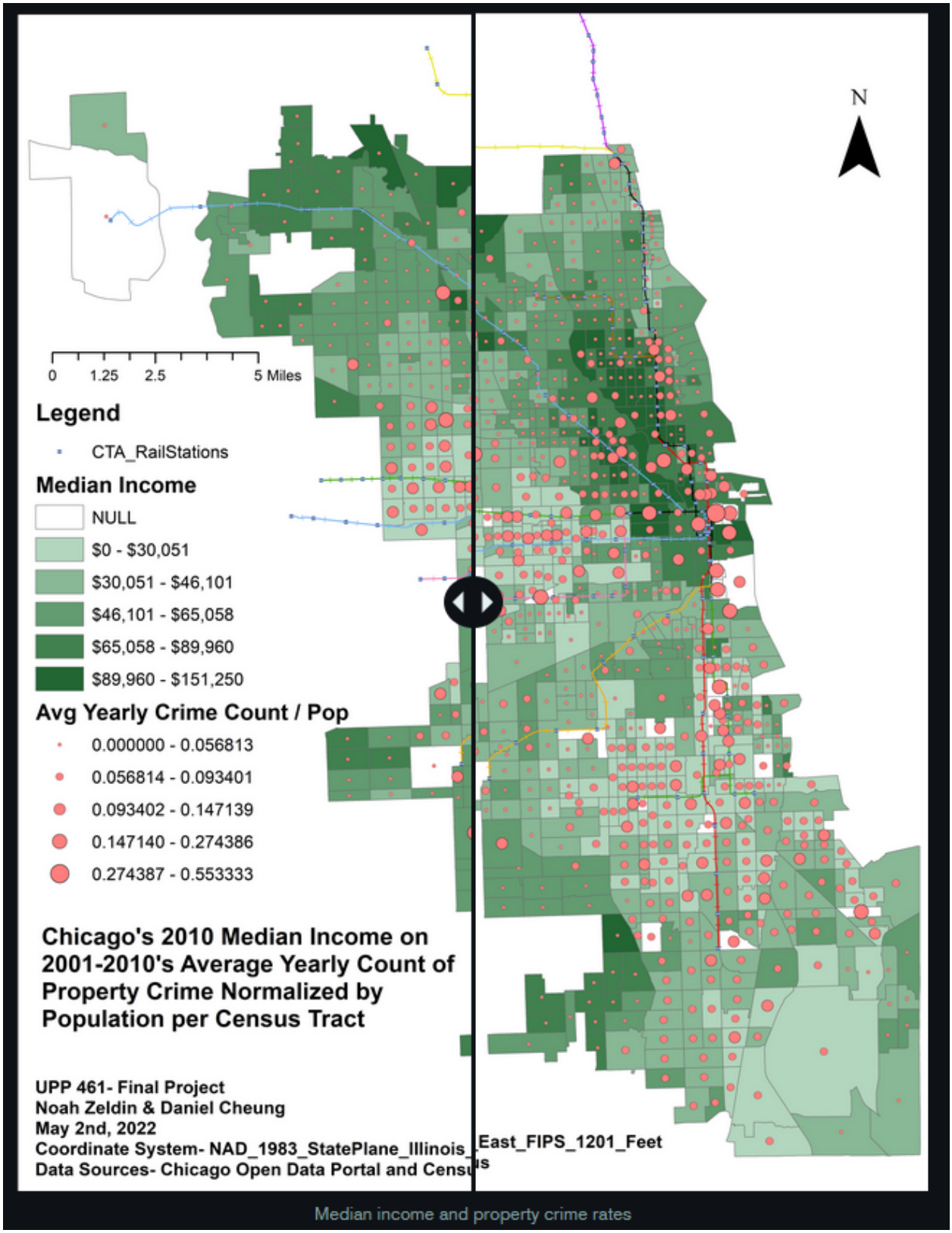
Graduated colors and symbols were used on the maps to display crime, median income, unemployment rate, and spillovers. The format of the ranges was changed by modifying the labeling, flipping symbols, and reversing the sort. The transparency for all the buffers used was 42%. Natural breaks was the primary classification used; manual was only used to put no color for 0 values as a sixth class, in order to prevent any confusion with null values in the data set. The five classes in all of the manual classifications were based on natural breaks with five classes.

Symbology Summary

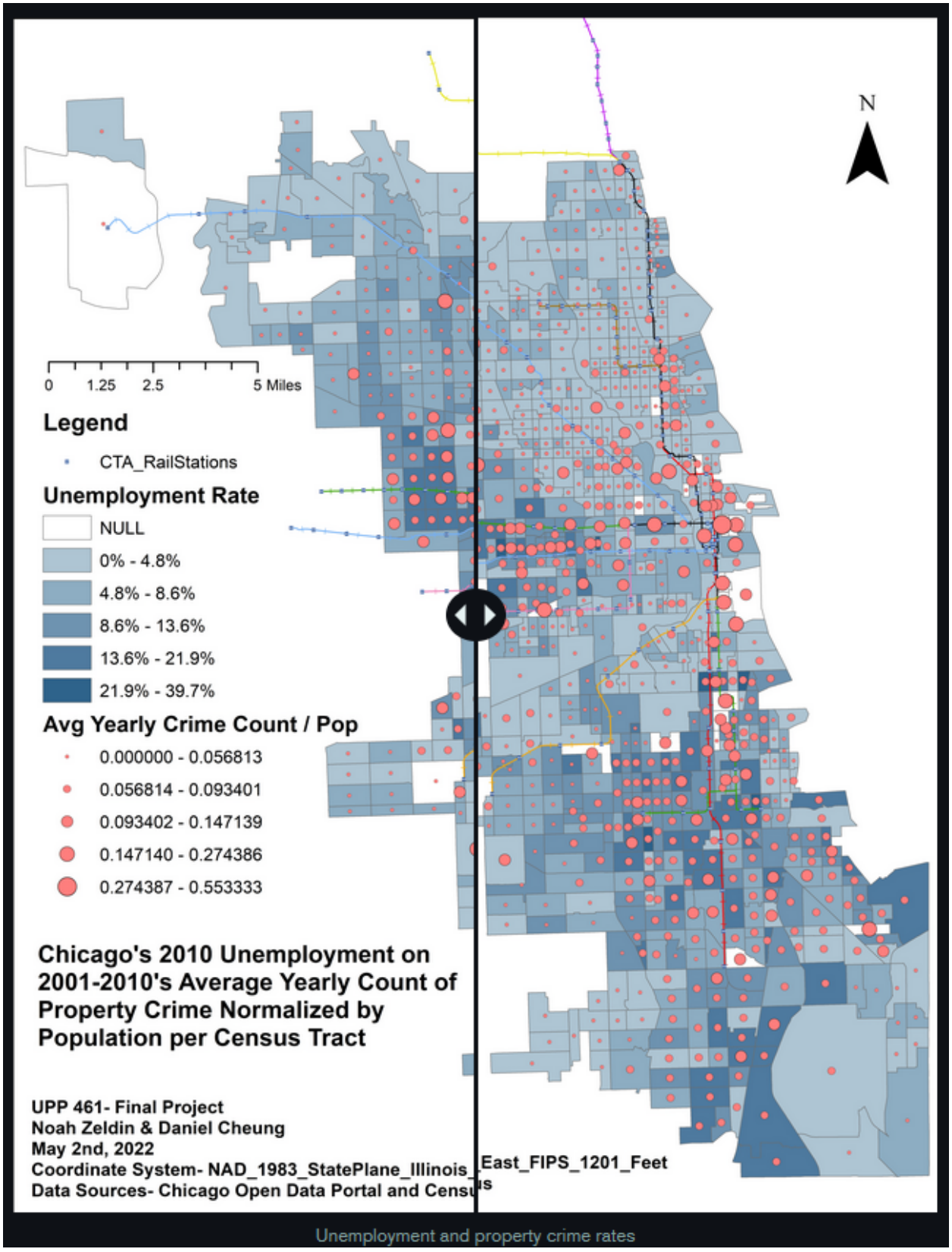
- Symbolize
  - graduated colors and symbols
  - changing the format of the ranges for the map legend
  - transparency
- Classify
  - Natural Breaks and Manual

Maps

Background: Economics and Crime



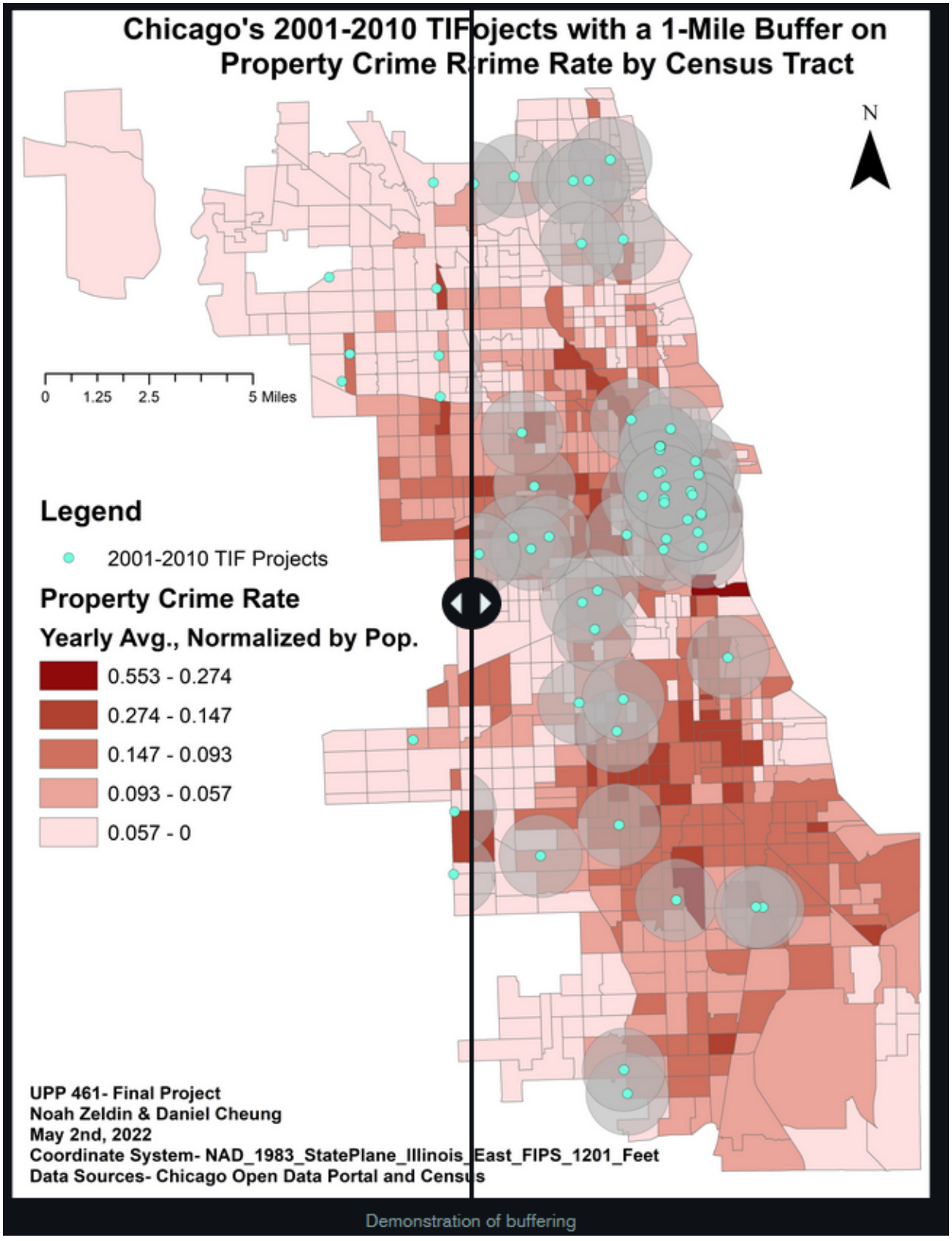
Median income and property crime rates



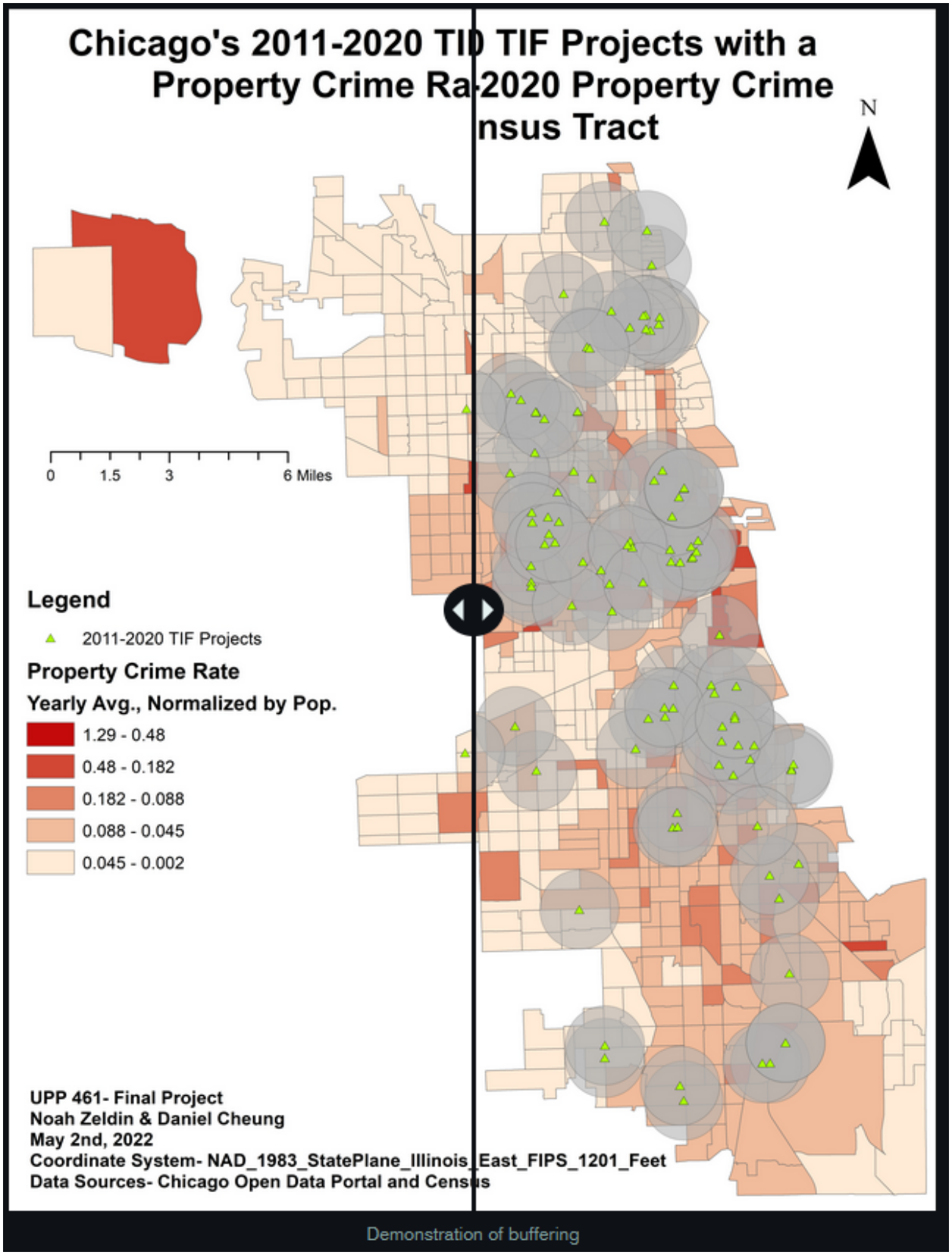
Unemployment and property crime rates

TIF Projects and Crime Rates

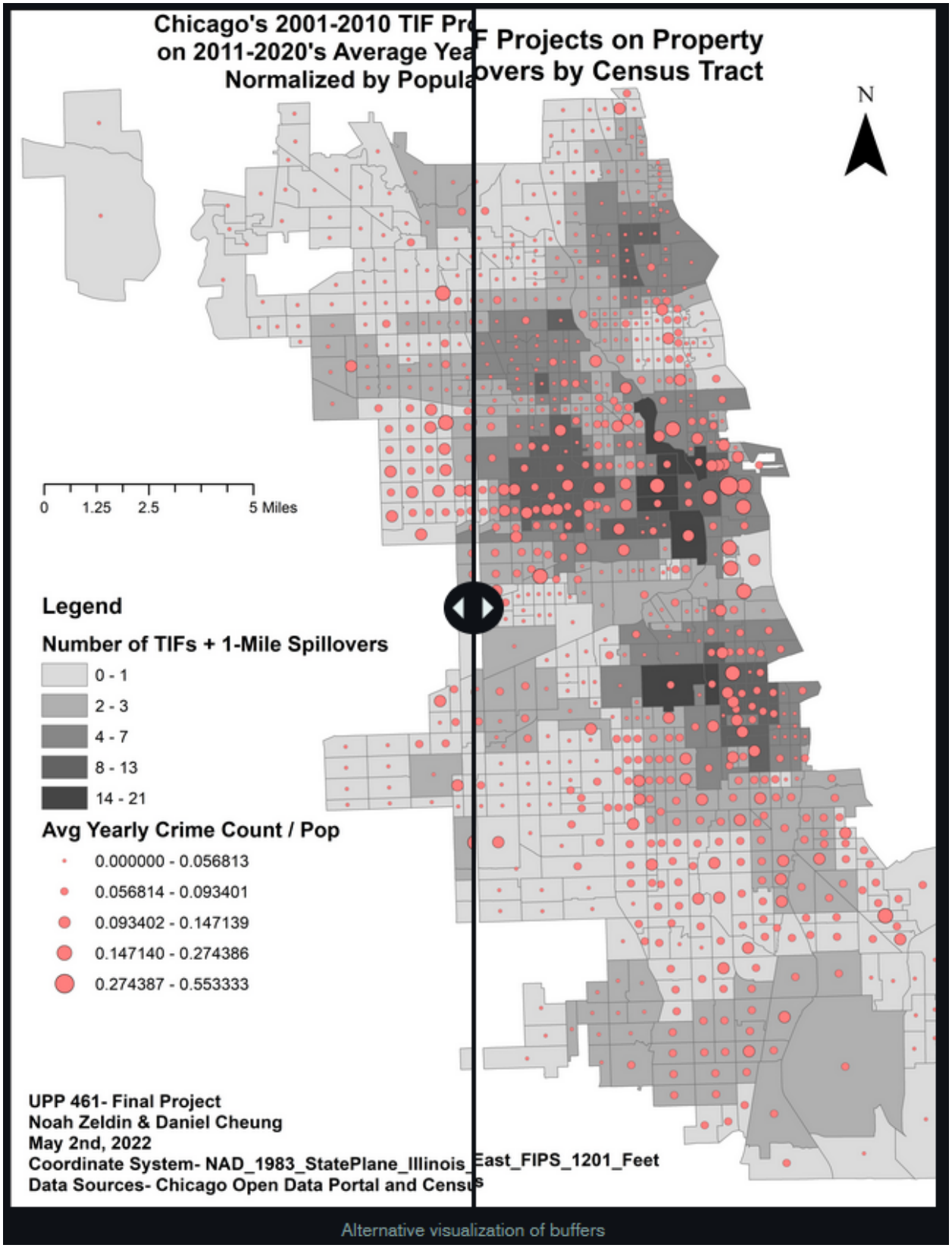
Demonstration of **Buffering** - attempt to capture potential **spillover effects**



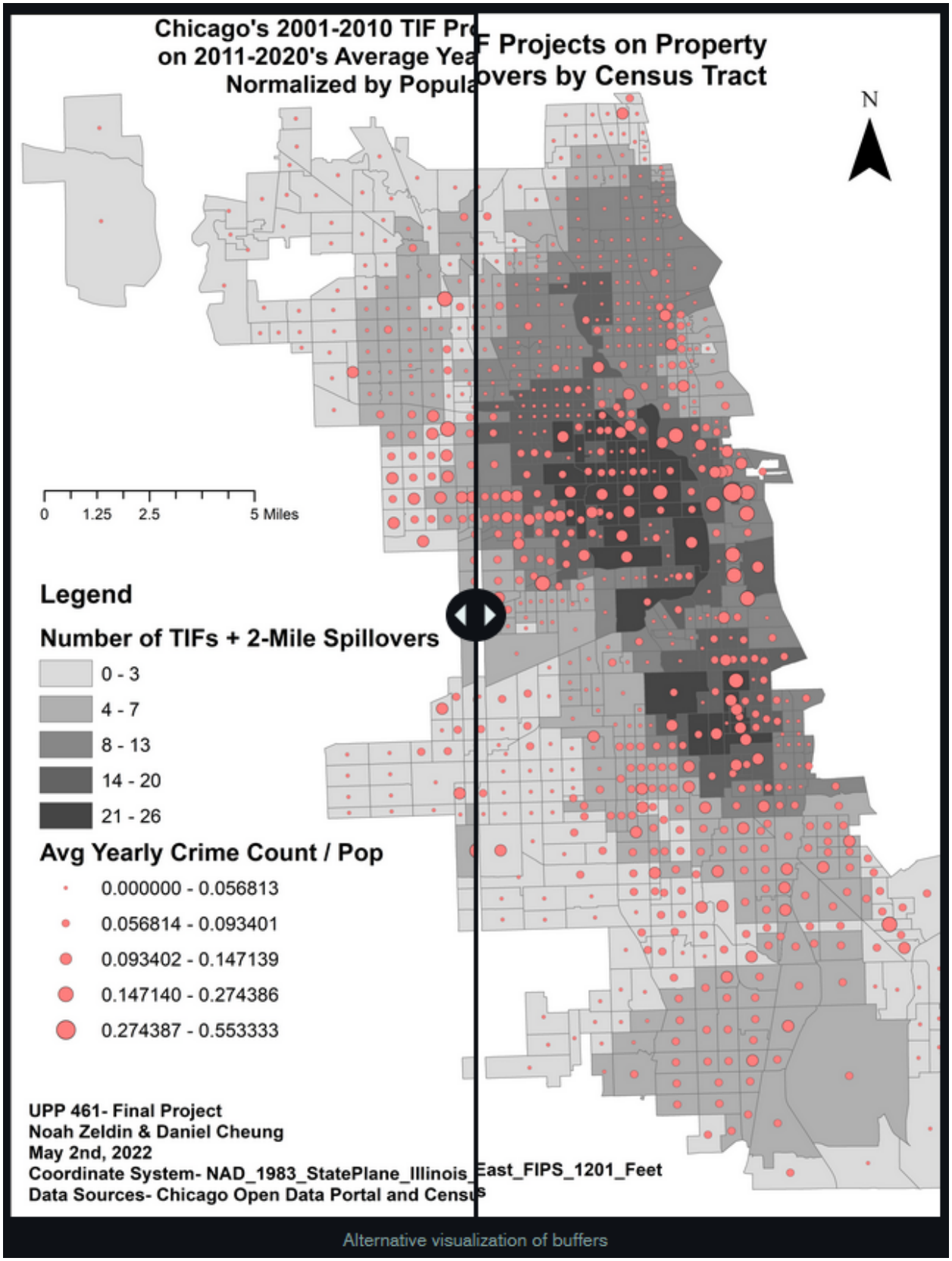
Demonstration of buffering



Demonstration of buffering

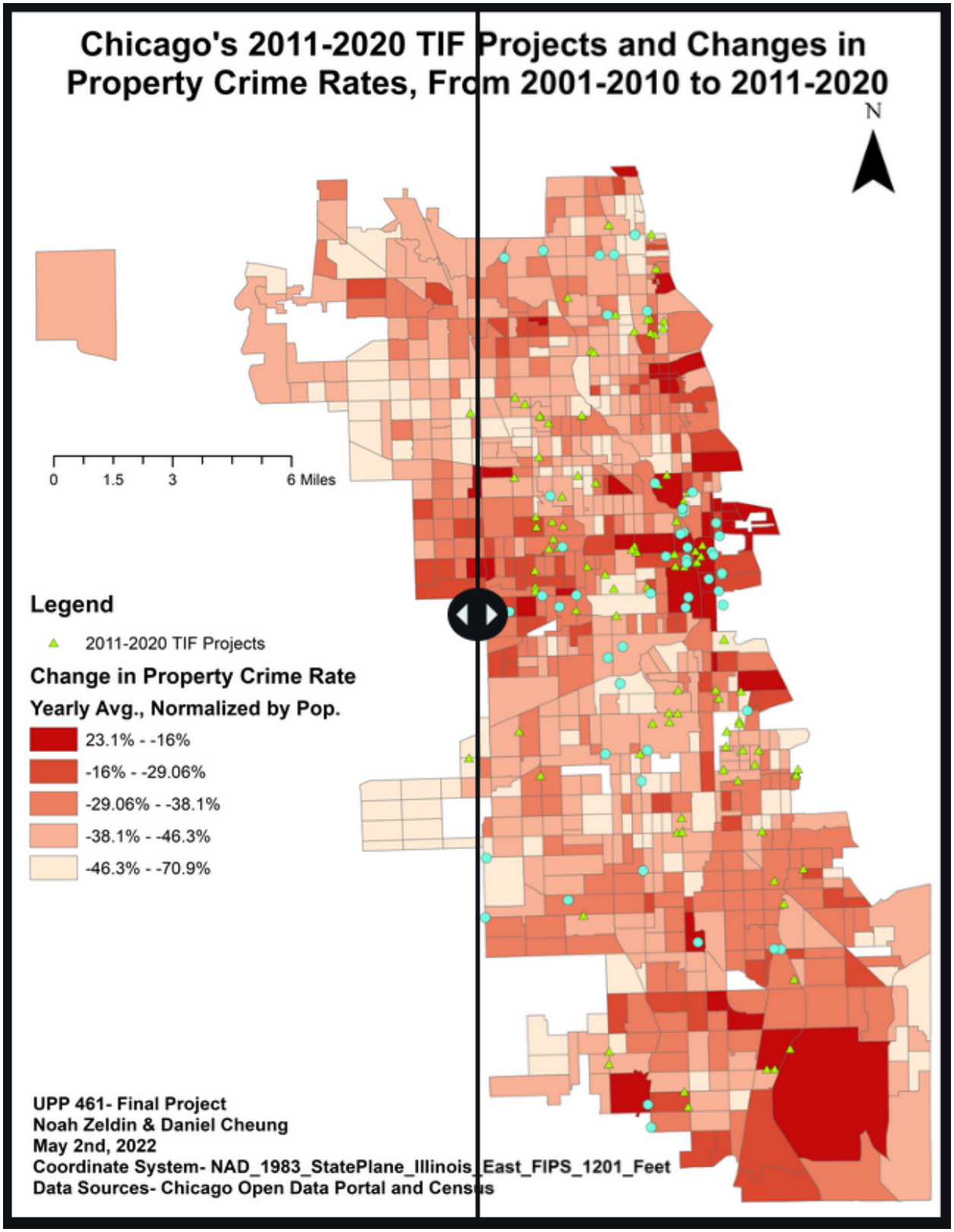


Alternative visualization of buffers



Alternative visualization of buffers

Main Map: Change in Property Crime Rates and Construction of New TIF Projects


Chicago's 2011-2020 TIF Projects and Changes in Property Crime Rates, From 2001-2010 to 2011-2020
N
0
1.5
3
6 Miles
Legend
2011-2020 TIF Projects
Change in Property Crime Rate
Yearly Avg., Normalized by Pop.
23.1% - -16%
-16% - -29.06%
-29.06% - -38.1%
-38.1% - -46.3%
-46.3% - -70.9%
UPP 461- Final Project
Noah Zeldin & Daniel Cheung
May 2nd, 2022
Coordinate System- NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet
Data Sources- Chicago Open Data Portal and Census

Statistical Relationships

The Relationship between TIFs and Property Crime Rates

	Avg. Tract-Level Property Crime Rate, 2011-20 (per 10k residents)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Number of TIFs	217.213***	201.488***	52.768***	52.912***	52.953***	50.339***	36.391***	38.369***
	(20.995)	(21.021)	(13.089)	(13.154)	(13.162)	(13.327)	(12.934)	(12.987)
Additional TIFs (1 mile)		12.840***		-0.296	0.051	-0.076	-4.128	-4.395
		(4.552)		(2.513)	(2.738)	(2.737)	(2.682)	(2.674)
Additional TIFs (2 miles)		6.223**			-0.547	-0.382	-3.518**	-2.694
		(2.846)			(1.708)	(1.755)	(1.680)	(1.702)
Avg. prop. crime rate, 2001-10 (per 10k residents)			0.628***	0.629***	0.629***	0.636***	0.647***	0.637***
			(0.017)	(0.017)	(0.017)	(0.018)	(0.017)	(0.018)
Median income						-0.0004		-0.001**
						(0.0003)		(0.0002)
Unemployment rate						-344.500*		76.500
						(207.755)		(204.184)
Time to work (min.)							-10.451***	-11.339***
							(1.324)	(1.386)
Number of L stations							27.929*	28.518*
							(14.724)	(14.699)
Observations	769	769	769	769	769	769	769	769
Adjusted R^2	0.121	0.143	0.696	0.695	0.695	0.695	0.720	0.722
Note:								*p<0.1; **p<0.05**; p<0.01

Final regression analysis with variables created using GIS tools, showing tract-level relationships

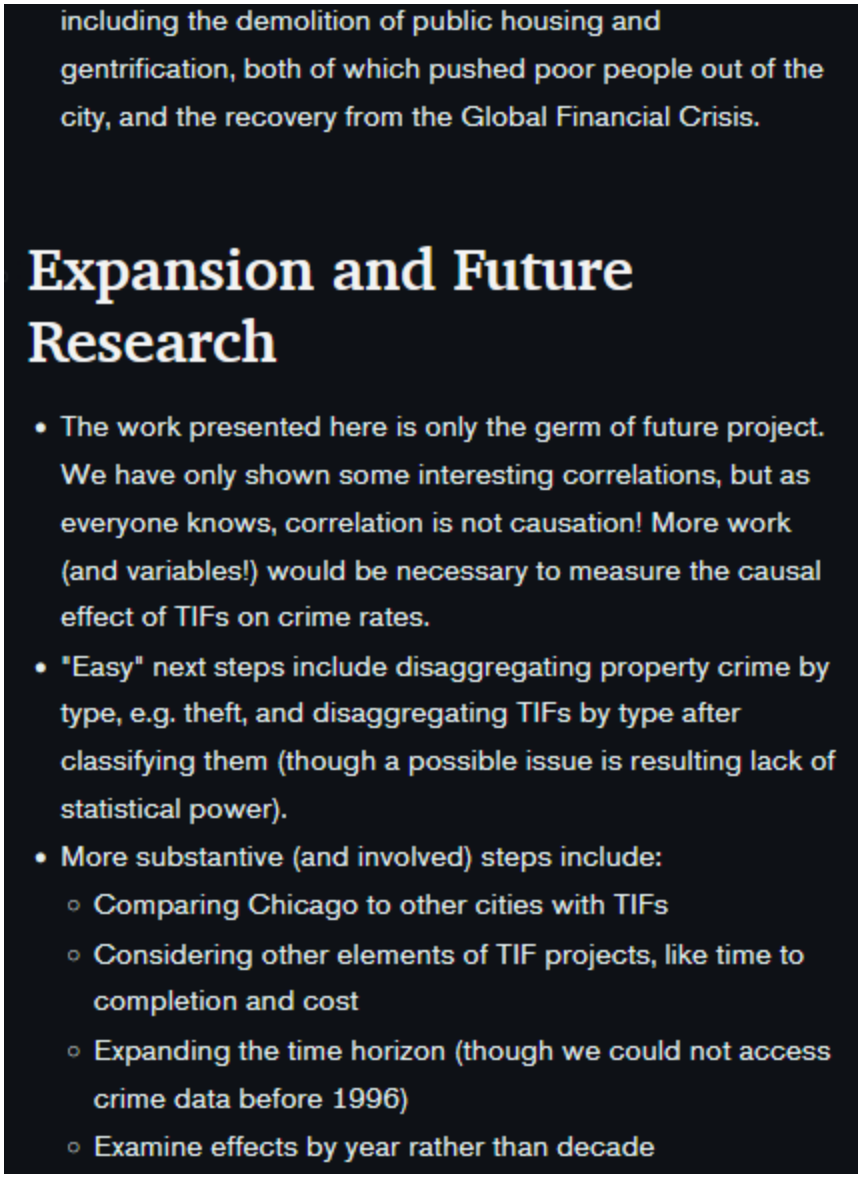
Note that the following variables were created in ArcMap: Number of TIFs, Additional TIFs, crime rates (for both decades) and L stations.

Main findings

- There is a strong, positive correlation between TIFs and property crime rates.
  - This possibly confirms anecdotal evidence that TIFs are disproportionately built in affluent neighborhoods, which attract property crimes.
- The most important controls appear to be past crime rate (model 3) and time to work and L stations (model 7).

Limitations

- Reverse causality: Do TIFs cause property crime or do high crime rates cause TIFs to be built?
- Omitted variables, like changes in policing patterns, likely bias the results.
- We have not controlled for city- and macro-level trends,



References

Main Data Sources: Census Data and TIGER/Line - https://www.census.gov/ Chicago Open Data Portal - https://data.cityofchicago.org/ Chicago Government TIF-Related Information - https://www.chicago.gov/

Specific References: Tax Increment Financing (TIF) Funded RDA and IGA Projects. (4/8/2022). https://data.cityofchicago.org/Community-Economic-Development/Tax-Increment-Financing-TIF-Funded-RDA-and-IGA-Pro/mex4-ppfc

Chicago TIF. (4/8/2022). https://webapps1.chicago.gov/ChicagoTif/

TIF map. (4/8/2022). https://data.cityofchicago.org/Community-Economic-Development/Tax-Increment-Financing-TIF-Projects-Map/v3a3-hhqn

TIF page on the city of Chicago's website. (4/8/2022). https://www.chicago.gov/city/en/depts/dcd/provdrs/tif.html

TIF program guide. (4/8/2022). https://www.chicago.gov/content/dam/city/depts/dcd/general/2020_tif_program_guide.pdf

TIF Funding Sources and Uses by TIF, Fiscal Year, and Type - 1998-2014. (4/8/2022). https://data.cityofchicago.org/Community-Economic-Development/TIF-Funding-Sources-and-Uses-by-TIF-Fiscal-Year-an/pner-h2in

TIF Shapefile. (4/8/2022). https://data.cityofchicago.org/Community-Economic-Development/Boundaries-Tax-Increment-Financing-Districts/fz5x-7zak

2000 Chicago census tracts. (4/8/2022). https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Census-Tracts-2000/pt6c-hxpp

2010 Chicago census tracts. (4/8/2022). https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Census-Tracts-2010/5jrd-6zik

2020 Cook County census tracts. (4/8/2022). https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2020&layergroup=Census+Tracts

2017 Chicago city boundary. (4/9/2022). https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-City/ewy2-6yfk

Crime 2001. (4/8/2022). https://data.cityofchicago.org/Public-Safety/Crimes-2001/8v97-unyc

Crime 2010. (4/8/2022). https://data.cityofchicago.org/Public-Safety/Crimes-2010/q4de-h6yq

Crime 2020. (4/8/2022). https://data.cityofchicago.org/Public-Safety/Crimes-2020/qzdf-xmn8

The Chicago Police Department definition of property crimes. (4/8/2022). https://home.chicagopolice.org/statistics-data/statistical-reports/

Crime 2001-Present. (4/8/2022). https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2/data

Last day of 2018 Shapefile for L stations. (4/11/2022). https://data.cityofchicago.org/Transportation/CTA-L-Rail-Stations-Shapefile/vmyy-m9qj

CTA Rail Lines. (4/11/2022). https://data.cityofchicago.org/Transportation/CTA-L-Rail-Lines-Shapefile/53r7-y88m

2020 Census Education Attainment by Census Tract. (4/10/2022). https://data.census.gov/cedsci/table?q=education%20attainment&g=0500000US17031%241400000&tid=ACSST5Y2020.S1501

2010 Census Education Attainment by Census Tract. (4/10/2022). https://data.census.gov/cedsci/table?q=education&g=0500000US17031%241400000&y=2010&tid=ACSDT5Y2010.B06009

2000 Census Education Attainment by Census Tract. (4/10/2022). https://data.census.gov/cedsci/table?q=education%20attainment&g=0500000US17031%241400000&y=2000&tid=DECENNIALDPSF32000.DP2

2020 Census Tract Population. (4/10/2022). https://data.census.gov/cedsci/table?q=population&g=0500000US17031%241400000&y=2020&tid=DECENNIALPL2020.P1

2010 Census Tract Population. (4/10/2022). https://data.census.gov/cedsci/table?q=population&g=0500000US17031%241400000&y=2010&tid=DECENNIALPL2010.P1

2000 Census Tract Population. (4/10/2022). https://data.census.gov/cedsci/table?q=population&g=0500000US17031%241400000&y=2000&tid=DECENNIALSF32000.P001&tp=true

2020 Census Tract Unemployment and Travel Time Data. (4/10/2022). https://data.census.gov/cedsci/table?q=employment&g=0500000US17031%241400000&y=2020&tid=ACSDP5Y2020.DP03&moe=false&tp=true

2010 Census Tract Unemployment and Travel Time Data. (4/10/2022). https://data.census.gov/cedsci/table?q=employment&g=0500000US17031%241400000&y=2010&tid=ACSDP5Y2010.DP03&moe=false&tp=true

2000 Census Tract Unemployment and Travel Time Data. (4/10/2022). https://data.census.gov/cedsci/table?q=employment&g=0500000US17031%241400000&y=2000&tid=DECENNIALDPSF32000.DP3&tp=true

2000 Census Tract Median Income Data. (4/10/2022). https://data.census.gov/cedsci/table?q=median%20income&g=0500000US17031%241400000&y=2000&tid=DECENNIALSF32000.P053&tp=true

2010 Census Tract Median Income Data. (4/10/2022). https://data.census.gov/cedsci/table?q=median%20income&g=0500000US17031%241400000&y=2010&tid=ACSST5Y2010.S1903&moe=false&tp=true

2020 Census Tract Median Income Data. (4/10/2022). https://data.census.gov/cedsci/table?q=median%20income&g=0500000US17031%241400000&y=2020&tid=ACSST5Y2020.S1903&moe=false&tp=true

2020 Census Tract Poverty Line. (4/10/2022). https://data.census.gov/cedsci/table?q=poverty&g=0500000US17031%241400000&tid=ACSST5Y2020.S1701&moe=false&tp=true

2010 Census Tract Poverty Line. (4/10/2022). https://data.census.gov/cedsci/table?q=poverty&g=0500000US17031%241400000&y=2010&tid=ACSST5Y2010.S2201&moe=false&tp=true

2000 Census Tract Poverty Line Data. (4/10/2022). https://data.census.gov/cedsci/table?q=poverty&g=0500000US17031%241400000&y=2000&tid=DECENNIALDPSF32000.DP3&tp=true

2010 Median Gross Rent. (4/10/2022). https://data.census.gov/cedsci/table?q=rents&g=0500000US17031%241400000&y=2010&tid=ACSDT5Y2010.B25064

2020 Median Gross Rent. (4/10/2022). https://data.census.gov/cedsci/table?q=rents&g=0500000US17031%241400000&y=2020&tid=ACSDT5Y2020.B25064

2020 Illinois Census Tracts. (4/8/2022). https://www.census.gov/cgi-bin/geo/shapefiles/index.php