

### Spatial Statistics – Spatial Analysis Lab

According to my analysis, three schools make the top 10 list for toxicity exposure using a one-mile buffer and a three-mile buffer: Ruby Shaw Elementary, Arcadia Park Elementary, and Park Crest Elementary. Widening the buffer to three miles increases the number of toxic sites contributing to an individual school's score, so it is unsurprising that scores jump, however the top two are very stable. This suggests that they face severe exposure from the only sites that are in their neighborhood.

Top 10 Highest Toxicity Scores, Sites Within 1 Mile					Top 10 Highest Toxicity Scores, Sites Within 3 Miles				
School	District	Students	Toxicity Score	Toxicity Rank	School	District	Students	Toxicity Score	Toxicity Rank
RUBY SHAW EL	MESQUITE ISD	861	28868.946	1	RUBY SHAW EL	MESQUITE ISD	861	28868.95	1
ARCADIA PARK EL	DALLAS ISD	593	17603.396	2	ARCADIA PARK EL	DALLAS ISD	593	26627.65	2
PARK CREST EL	GARLAND ISD	526	11980.124	3	FANNIN EL	GRAND PRAIRIE ISD	381	25968.38	3
SAM HOUSTON MIDDLE	GARLAND ISD	852	10548.612	4	RUTHERFORD EL	MESQUITE ISD	729	19810.96	4
MILAM EL	GRAND PRAIRIE ISD	660	5525.666	5	B J SMITH EL	MESQUITE ISD	899	18210.84	5
NORTHSIDE EL	DESOTO ISD	488	4249.405	6	AGNEW MIDDLE	MESQUITE ISD	941	15321.15	6
DANIEL INT	DUNCANVILLE ISD	585	2853.262	7	SUNNYVALE EL	SUNNYVALE ISD	376	14963.71	7
HASTINGS EL	DUNCANVILLE ISD	460	2700.803	8	MESQUITE H S	MESQUITE ISD	2936	14255.85	8
AMELIA EARHART EL	DALLAS ISD	339	2580.019	9	BERRY MIDDLE	MESQUITE ISD	906	13114.05	9
MERRIFIELD EL	DUNCANVILLE ISD	534	2305.678	10	PARK CREST EL	GARLAND ISD	526	12806.22	10

Using the newly created scores for the three-mile buffer, top 25 percent of schools in terms of toxicity exposure were compared to the bottom 75 percent. The top 25 percent of schools contained 27 percent of the overall population (a little under 74,000 people), but almost 35 percent of the county's Asian population. Hispanic concentration is not statistically different between schools with the most exposure versus those with the least. Low income vs non-low income proportions do fail the chi square test, but they magnitude of the difference appears to be about half a percent.

p-value = 0.99, column proportions differ significantly from expected frequencies

Toxic Score Percentile	asian	black	other_race	white	total
Top 25%	34.8% (3,697)	19.3% (14,174)	31.3% (36,446)	27.2% (19,652)	27.1% (73,969)
Bottom 75%	65.2% (6,915)	80.7% (59,335)	68.7% (80,092)	72.8% (52,663)	72.9% (199,005)

p-value = 0.91, column proportions do not differ significantly from expected frequencies

Toxic Score Percentile	hispanic	non_hispanic	total
Top 25%	31.3% (36,073)	24.0% (37,896)	27.1% (73,969)
Bottom 75%	68.7% (79,261)	76.0% (119,744)	72.9% (199,005)

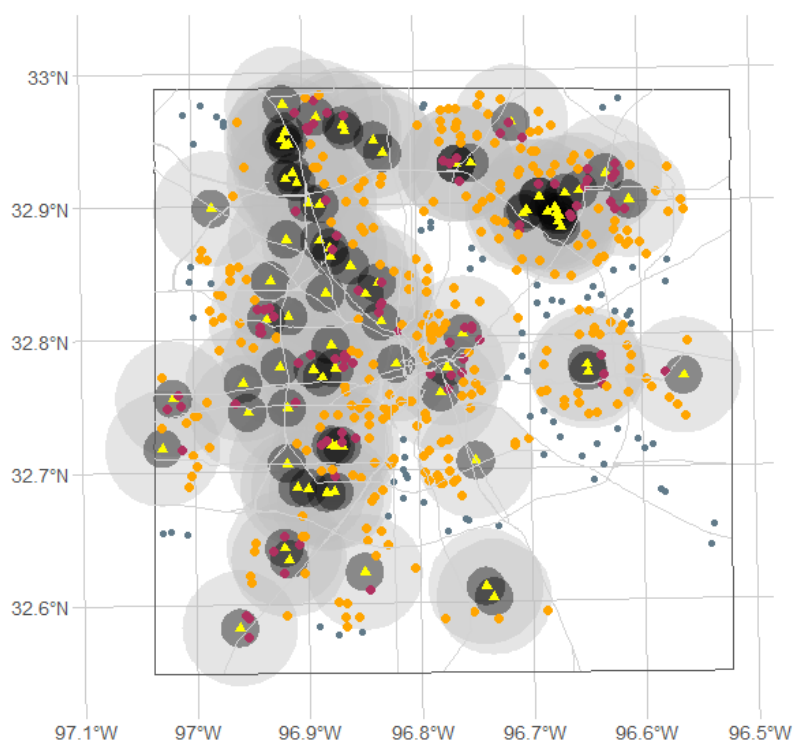
p-value = 0.98, column proportions differ significantly from expected frequencies

Toxic Score Percentile	low_income	non_low_income	total
Top 25%	27.6% (41,446)	26.4% (32,523)	27.1% (73,969)
Bottom 75%	72.4% (108,469)	73.6% (90,536)	72.9% (199,005)

Finally, the spatial distribution of the toxic sites, schools, and buffer analysis can be seen in the map below. The remaining pages of this document contain a log of my code.

## School Exposure to Toxic Sites

Dallas County, Texas



```

*****
#-----
#####
#WEEK 3 ITEMS
# The objective of the research is to answer the question: Are disadvantaged
groups (economic, racial or ethnic) more exposed to pollution than the nondisadvantaged?
#####
# Part 1:
#####
# Here we examine all schools. Compare schools within 1 mile of a toxic site
to schools beyond 1 mile. Calculate the proportion of students in each
economic or racial or ethnic group for schools within one mile, and for
schools beyond one mile.
# Is there a higher proportion of economically disadvantaged or minority
racial groups (Black/African American, Hispanic, Asian) in schools within one
mile (and thus potentially more exposed to pollution) compared to the
proportion in schools beyond one mile? (Hint—calculate proportions relative
to row sum—total students within buffer & total outside buffer). Use a
spreadsheet to make the calculations.
# On a map show:
# Toxic sites with a one mile (5,280 feet) buffer around each
# All schools; use different color for schools within toxic site buffer
# Highways
# County boundary
#####
# Part 2:
#####
# Here we examine only schools within the buffer. For schools within one mile
of a toxic site calculate an index of exposure to toxic emissions. For each
school, find all toxic sites within 1 mile, divide the site's toxic score by
its distance to the school, then sum over all toxic sites to create a "toxic
index" for the school.
#####
# Part 1: Solution
#####
#Compare the demographic characteristics of the ten schools with highest toxic
index ("TopTen" schools) to the demographic characteristics of all others
within the one mile buffer
# Is there a higher proportion of economically disadvantaged or minority
racial groups in the top ten schools (and thus potentially more exposed to
pollution) compared to the proportion in other schools within the one mile
buffer?
# On a map show:
# Toxic sites, schools and highways
# for the top ten schools only, draw a symbol proportional to the school's
exposure index
# The coordinate system for all spatial data is:
# NAD_1983_StatePlane_Texas_North_Central_FIPS_4202_Feet
# North American Datum 1983
# State Plane Coordinate System (SPCS), Texas North Central Zone
# (Federal Information Processing Standard Code (FIPS) 4202)
# Measurement units are feet. There are 5280 feet in one mile.
#Data key for the variables in the school data:

```

```

### Economically disadvantaged (n_lowInc)
### African American (n_black)
### Hispanic (n_hispanic)
### Asian (n_asian)
### White (n_white)
### Total student body (n_all).
#Note that the components may not add up to the total as some are not mutually
exclusive categories.
#####
#make sure the below libraries are installed on your machine. Use
install.packages("LIBRARYNAME")
rm(list=ls()) #clear objects in memory
# Set working directory
setwd(paste0(here::here(), "/02_Spatial_Analysis"))
#####
# Load additional libraries
pacman::p_load(tidyverse, # for basic data manipulation, visualization
scales, # for formatting number output
patchwork, # for arranging ggplots in grids
sf, # simple features for spatial
summarytools, # for checking data frame characteristics
here, # for relative file paths
knitr, # for tables
kableExtra, # table styling
janitor) # for cleaning and tabulations
#load your spatial libraries
library(sp)
library(raster)
library(spatstat)
library(rgdal)
library(maptools)
library(rgeos)
library(GISTools)
library(hapfiles)
#####
#read in the shapefiles from your working director. Need the raster library to
use the shapefile(...) command.
# Read in shapefiles as simple features objects
sch <- st_read("Raw_Data/Dal_schools.shp")#contains all schools in Dallas
county
tox <- st_read("Raw_Data/Dal_toxic.shp")#contains all toxic sites in Dallas
hwy <- st_read("Raw_Data/Highways_NCTCOG_SPCS.shp")#highways
cnty <- st_read("Raw_Data/County_NCTCOG.shp")#county boundary
#####
#####
#Get back to our problem
#lets remove parts of our data we will not need:
dallas <- cnty %>% filter(COUNTY == "Dallas")
#####
#Lets remove parts of our the highway network outside of Dallas
#st_intersection for simple features
hwy2 <- st_intersection(hwy, dallas)
#####

```

```
#####-----
# Create 1 and 3 mile buffers
tox_buf_1 <- st_buffer(tox, dist = 5280)
tox_buf_3 <- st_buffer(tox, dist = 5280 * 3)
# Categorize whether school falls within the buffers
sch_tox <- sch %>%
mutate("tox_1" = ifelse(lengths(st_intersects(., tox_buf_1)) > 0, TRUE,
FALSE),
"tox_3" = ifelse(lengths(st_intersects(., tox_buf_3)) > 0, TRUE,
FALSE)) %>%
dplyr::select(ORG_NUM:COUNTY_NAM, tox_1, tox_3, everything())
#####
# Convert to a data frame, drop NAs, add columns, rename variables
sch_tox_df <- sch_tox %>%
st_drop_geometry() %>%
na.omit() %>%
mutate(other_race = n_all - n_black - n_white - n_asian,
non_hispanic = n_all - n_hispanic,
non_low_income = n_all - n_lowInc) %>%
dplyr::select(ORG_NUM:COUNTY_NAM, tox_1, tox_3,
# race
"asian" = n_asian, "black" = n_black, other_race, "white" =
n_white,
# ethnicity
"hispanic" = n_hispanic, non_hispanic,
# income
"low_income" = n_lowInc, non_low_income,
# all
"total" = n_all)
# Calculate group values
# 1 mile buffer
t1_raw <- sch_tox_df %>%
mutate(tox_1 = ifelse(tox_1 == TRUE, "In", "Outside")) %>%
group_by("1 Mile Buffer" = tox_1) %>%
summarize_at(.vars = vars(asian:total), .funs = sum) %>%
ungroup()
# 3 mile buffer
t3_raw <- sch_tox_df %>%
mutate(tox_3 = ifelse(tox_3 == TRUE, "In", "Outside")) %>%
group_by("3 Mile Buffer" = tox_3) %>%
summarize_at(.vars = vars(asian:total), .funs = sum) %>%
ungroup()
# Percents
# 1 mile buffer
t1_pct <- t1_raw %>%
mutate_at(.vars = vars(asian:total), .funs = list(~./sum(.)))
# 3 mile buffer
t3_pct <- t3_raw %>%
mutate_at(.vars = vars(asian:total), .funs = list(~./sum(.)))
# For presentation/table
# 1 mile buffer
t1_output <- t1_raw %>%
mutate_at(.vars = vars(asian:total),
```

```

.funs = list(~paste0(percent(./sum(.), accuracy = 0.1), " (",
comma(., ")"))))
# 3 mile buffer
t3_output <- t3_raw %>%
mutate_at(.vars = vars(asian:total),
.funs = list(~paste0(percent(./sum(.), accuracy = 0.1), " (",
comma(., ")"))))
# Contingency tables
# 1 mile buffer
t1_race <- t1_pct %>% dplyr::select(`1 Mile Buffer`, asian:white)
t1_ethnicity <- t1_pct %>% dplyr::select(`1 Mile Buffer`,
hispanic:non_hispanic)
t1_income <- t1_pct %>% dplyr::select(`1 Mile Buffer`,
low_income:non_low_income)
# 3 mile buffer
t3_race <- t1_pct %>% dplyr::select(`3 Mile Buffer`, asian:white)
t3_ethnicity <- t1_pct %>% dplyr::select(`3 Mile Buffer`,
hispanic:non_hispanic)
t3_income <- t1_pct %>% dplyr::select(`3 Mile Buffer`,
low_income:non_low_income)
#the above output gives us the percentaes of each category that is in and
outside of the toxic buffer. 18.6% of all students are in the buffer. 13.0% of
black students are in the buffer. As compared to all studenst, black students
tend to be less likely to be in the toxic buffer. If we look at the Hispanic
student proportion, 24.6% are in the buffer as compared to 18.6% for all
studnets. It suggests that Hispanc students tend to be within these buffers
more so than other students. We can more formally test these using a Chisquared
test.
#let's check we have the correct columns
# 1 mile buffer
t1_race
t1_ethnicity
t1_income
# 3 mile buffer
t3_race
t3_ethnicity
t3_income
#####Chi-Squared tests of
independence
source("chitest.R")# read in the chitest function
#We can also do a Chi-sqare test, but because there is double counting in the
categories, we have to be careful and define categories that are mutually
exclusive. Create the following columns in t1.
# 1 mile buffer
chitest(t1_race %>% dplyr::select(-1))
chitest(t1_ethnicity %>% dplyr::select(-1))
chitest(t1_income %>% dplyr::select(-1))
# 3 mile buffer
chitest(t3_race %>% dplyr::select(-1))
chitest(t3_ethnicity %>% dplyr::select(-1))
chitest(t3_income %>% dplyr::select(-1))
#Interpret the results. Look at the adjusted.residuals to see the direction of
relationship.

```

#Adjusted residuals greater than +/-2 suggest a departure from the null hypothesis of independence for that cell (if you have many categories in your table, you can use 3 as the cutoff). Positive values suggest over representation, negative values suggest under representatoin in that cell in statistically significant ways.

# 3 mile buffer

#####

#Which are the most affected schools?

#Some schools are within 1 mile of multiple toxic sites

#Use a metric that uses the sum of the toxicity score divided by distance over all toxic sites within a one mile buffer.

#say school 1 has to toxic sites with scores SC1 and SC2 at distances D1 and D2. Overall score for the school would be (SC1/D1) + (SC2 /D2). The scores for each toxic site are in the TOX Data.

#What do we need?

#we need distance between each school in the buffer and the toxic sites

#compute distance from schools to all toxic sites

# 1 mile buffer

```
sch_tox_1 <- sch_tox %>% filter(tox_1 == TRUE)
```

```
dist_matrix_1 <- st_distance(sch_tox_1, tox) %>%
```

```
as_tibble() %>%
```

```
units::drop_units() %>%
```

```
mutate_all(~ifelse(. > 5280, 0, .)) %>%
```

```
mutate_all(~ifelse(. == 0, 0, 1/.))
```

# 3 mile buffer

```
sch_tox_3 <- sch_tox %>% filter(tox_3 == TRUE)
```

```
dist_matrix_3 <- st_distance(sch_tox_3, tox) %>%
```

```
as_tibble() %>%
```

```
units::drop_units() %>%
```

```
mutate_all(~ifelse(. > 5280 * 3, 0, .)) %>%
```

```
mutate_all(~ifelse(. == 0, 0, 1/.))
```

#####

#We can use matrix multiplication between dist2recp and tox\$SCORE to compute our toxicity scores

#each row of distance is taken in turn, multiplied with its corresponding toxicity score, and the sum is taken.

#note that because we have set distance to 0 for sites that are outside of the 1 mile buffer, the toxicity of those sites will be zeroed out by the multiplication with the distance matrix. Work thorough the above example to understand how matrix multiplication works. Note that for matrix multiplication to work, the number of columns in the first matrix must equal the number of rows in the second matrix. The resulting product will have a dimension equal to the number of rows of the first matrix and the number of columns of the second matrix. So in our case, we are multiplyin a matrix that is 94X78 (distance) with a matrix that is 78X1 (tox score). The resulting matrix will be 94X1 (toxicity scores for each school )

#we can use matrix multiplcation for this:

# 1 mile buffer

```
total_tox_1 <- as.matrix(dist_matrix_1) %*% tox$SCORE %>%
```

```
as_tibble() %>%
```

```
rename("tox_score" = V1) %>%
```

```
mutate(tox_score = round(tox_score, 3))
```

# 3 mile buffer

```

total_tox_3 <- as.matrix(dist_matrix_3) %*% tox$SCORE %>%
as_tibble() %>%
rename("tox_score" = V1) %>%
mutate(tox_score = round(tox_score, 3))
#take this measure and add it to the schools in buffer data. SUMmarize, draw
histograms
#1 mile
sch_tox_1$tox_score <- total_tox_1$tox_score
head(sch_tox_1)
summary(sch_tox_1$tox_score)
hist(sch_tox_1$tox_score)
hist(sch_tox_1$tox_score, 30)
#3 mile
sch_tox_3$tox_score <- total_tox_3$tox_score
head(sch_tox_3)
summary(sch_tox_1$tox_score)
hist(sch_tox_1$tox_score)
hist(sch_tox_1$tox_score, 30)
##plot our results
plot(dallas)
plot(sch,cex=0.4,add=TRUE)
plot(sch_tox_1,cex=0.35*sch_tox_1$tox_score/
mean(sch_tox_1$tox_score),pch=21,col=2,bg=8,add=TRUE)#cex controls the size of
the schools
#see here for more in symbols (pch) http://www.endmemo.com/program/R/
pchsymbols.php
summary(tox$SCORE)
plot(tox,add=TRUE,col=2,cex=tox$SCORE/10000000)
#plot(hwy2,col=8,add=TRUE)
#Identify the top 10 schools and examine if they are different from the
remaining schools in the 1 mile buffer.
#create a data frame with the scores
#create a variable that ranks the schools by their total toxic exposure scores
#Rank the negative of totTox makes ranks as 1 the schools with the highest
totTox score. If we use rank(d$totTox), the same school would be ranked 94th.
#top 10 toxic exposed schools
# 1 mile buffer
sch_tox_1 <- sch_tox_1 %>%
mutate(tox_rank = rank(-tox_score),
tox_group = ifelse(tox_rank <= 10, "Top 10", "Other"))
# 3 mile buffer
sch_tox_3 <- sch_tox_3 %>%
mutate(tox_rank = rank(-tox_score),
tox_group = ifelse(tox_rank <= 10, "Top 10", "Other"))
#create a contingency table and compare top10 with other
# Convert to a data frame, drop NAs, add columns, rename variables
# 1 mile buffer
top10_1_df <- sch_tox_1 %>%
st_drop_geometry() %>%
na.omit() %>%
mutate(other_race = n_all - n_black - n_white - n_asian,
non_hispanic = n_all - n_hispanic,
non_low_income = n_all - n_lowInc) %>%

```



```

dplyr::select(ORG_NUM:COUNTY_NAM, tox_group,
# race
"asian" = n_asian, "black" = n_black, other_race, "white" =
n_white,
# ethnicity
"hispanic" = n_hispanic, non_hispanic,
# income
"low_income" = n_lowInc, non_low_income,
# all
"total" = n_all)
# 3 mile buffer
top10_3_df <- sch_tox_3 %>%
st_drop_geometry() %>%
na.omit() %>%
mutate(other_race = n_all - n_black - n_white - n_asian,
non_hispanic = n_all - n_hispanic,
non_low_income = n_all - n_lowInc) %>%
dplyr::select(ORG_NUM:COUNTY_NAM, tox_group,
# race
"asian" = n_asian, "black" = n_black, other_race, "white" =
n_white,
# ethnicity
"hispanic" = n_hispanic, non_hispanic,
# income
"low_income" = n_lowInc, non_low_income,
# all
"total" = n_all)
# Calculate group values
# 1 mile buffer
top10_1_raw <- top10_1_df %>%
group_by("Total Toxic Score" = tox_group) %>%
summarize_at(.vars = vars(asian:total), .funs = sum) %>%
ungroup()
# 3 mile buffer
top10_3_raw <- top10_3_df %>%
group_by("Total Toxic Score" = tox_group) %>%
summarize_at(.vars = vars(asian:total), .funs = sum) %>%
ungroup()
# Percents
# 1 mile buffer
top10_1_pct <- top10_1_raw %>%
mutate_at(.vars = vars(asian:total), .funs = list(~./sum(.)))
# 3 mile buffer
top10_3_pct <- top10_3_raw %>%
mutate_at(.vars = vars(asian:total), .funs = list(~./sum(.)))
# For presentation/table
# 1 mile buffer
top10_1_output <- top10_1raw %>%
mutate_at(.vars = vars(asian:total),
.funs = list(~paste0(percent(./sum(.), accuracy = 0.1), " (",
comma(.), ")")))
# 3 mile buffer
top10_3_output <- top10_3_raw %>%

```

```

mutate_at(.vars = vars(asian:total),
.funs = list(~paste0(percent(./sum(.), accuracy = 0.1), " (",
comma(.), ")"))
# Contingency tables
# 1 mile buffer
top10_1_race <- top10_1_pct %>% dplyr::select(`Total Toxic Score`,
asian:white)
top10_1_ethnicity <- top10_1_pct %>% dplyr::select(`Total Toxic Score`,
hispanic:non_hispanic)
top10_1_income <- top10_1_pct %>% dplyr::select(`Total Toxic Score`,
low_income:non_low_income)
# 3 mile buffer
top10_3_race <- top10_3_pct %>% dplyr::select(`Total Toxic Score`,
asian:white)
top10_3_ethnicity <- top10_3_pct %>% dplyr::select(`Total Toxic Score`,
hispanic:non_hispanic)
top10_3_income <- top10_3_pct %>% dplyr::select(`Total Toxic Score`,
low_income:non_low_income)
#interpret your results. DO further examinations using the chi-square test
within this subset.
# 1 mile buffer
chitest(top10_1_race %>% dplyr::select(-1))
chitest(top10_1_ethnicity %>% dplyr::select(-1))
chitest(top10_1_income %>% dplyr::select(-1))
# 3 mile buffer
chitest(top10_3_race %>% dplyr::select(-1))
chitest(top10_3_ethnicity %>% dplyr::select(-1))
chitest(top10_3_income %>% dplyr::select(-1))
#####
#####
#Assignment Due: Feb 17, 2020 12:00 PM (via Blackboard)
#Our in-class analysis assumed only schools within a mile of a toxic sites
were affected. Expand the analysis to all 470 schools and compute the school
toxicity score by summing over all toxic sites within 3 miles of the schools
using a similar formula as the above analysis (SCORE/distance summed over all
toxic sites within 3 miles of each school). Generate a top 10 list by severity
of exposure from this analysis. Compare your results with the results above.
Report on the following:
#1. How many of the top 10 schools from this analysis are the same as the ones
in the 1-mile buffer analysis?
sch_tox_1 %>%
st_drop_geometry() %>%
filter(tox_group == "Top 10") %>%
arrange(tox_rank) %>%
dplyr::select("School" = ORG_NAME, "District" = DISTRICT_N, "Students" =
n_all,
"Toxicity Score" = tox_score, "Toxicity Rank" = tox_rank) %>%
knitr::kable(., caption = "Top 10 Highest Toxicity Scores, Sites Within 1
Mile", format = "html") %>%
kableExtra::kable_styling(bootstrap_options = c("striped"))
sch_tox_3 %>%
st_drop_geometry() %>%
filter(tox_group == "Top 10") %>%

```

```

arrange(tox_rank) %>%
dplyr::select("School" = ORG_NAME, "District" = DISTRICT_N, "Students" =
n_all,
"Toxicity Score" = tox_score, "Toxicity Rank" = tox_rank) %>%
knitr::kable(., caption = "Top 10 Highest Toxicity Scores, Sites Within 3
Miles", format = "html") %>%
kableExtra::kable_styling(bootstrap_options = c("striped"))
#2. Based on your updated analysis using the 3 mile threshold, take the top 25
percent of schools by the toxicity exposure score (i.e, those with scores
above the 75th percentile) and compare if there are ethnic, racial, or
economic differences among the students in the top 25 percentile of schools as
compared to remaining schools.
perc25_3_df <- sch_tox_3 %>%
st_drop_geometry() %>%
na.omit() %>%
mutate(tox_perc = ntile(tox_score, 100),
tox_group = ifelse(tox_perc >= 75, "Top 25%", "Bottom 75%")) %>%
mutate(other_race = n_all - n_black - n_white - n_asian,
non_hispanic = n_all - n_hispanic,
non_low_income = n_all - n_lowInc) %>%
dplyr::select(ORG_NUM:COUNTY_NAM, tox_group,
# race
"asian" = n_asian, "black" = n_black, other_race, "white" =
n_white,
# ethnicity
"hispanic" = n_hispanic, non_hispanic,
# income
"low_income" = n_lowInc, non_low_income,
# all
"total" = n_all)
# Calculate group values
# 3 mile buffer
perc25_3_raw <- perc25_3_df %>%
group_by("Toxic Score Percentile" = tox_group) %>%
summarize_at(.vars = vars(asian:total), .funs = sum) %>%
ungroup()
# Percents
# 3 mile buffer
perc25_3_pct <- perc25_3_raw %>%
mutate_at(.vars = vars(asian:total), .funs = list(~./sum(.)))
# For presentation/table
# 3 mile buffer
perc25_3_output <- perc25_3_raw %>%
mutate_at(.vars = vars(asian:total),
.funs = list(~paste0(percent(./sum(.), accuracy = 0.1), " (",
comma(.), ")")))
# Contingency tables
# 3 mile buffer
perc25_3_race <- perc25_3_pct %>% dplyr::select(`Toxic Score Percentile`,
asian:white)
perc25_3_ethnicity <- perc25_3_pct %>% dplyr::select(`Toxic Score Percentile`,
hispanic:non_hispanic)
perc25_3_income <- perc25_3_pct %>% dplyr::select(`Toxic Score Percentile`,

```

```

low_income:non_low_income)
#interpret your results. DO further examinations using the chi-square test
within this subset.
# 3 mile buffer
chitest(perc25_3_race %>% dplyr::select(-1))
chitest(perc25_3_ethnicity %>% dplyr::select(-1))
chitest(perc25_3_income %>% dplyr::select(-1))
# For the paper
perc25_3_output %>%
dplyr::select(`Toxic Score Percentile`, asian:white, total) %>%
arrange(desc(`Toxic Score Percentile`)) %>%
knitr::kable(., caption = "p-value = 0.99, column proportions differ
significantly from expected frequencies",
format = "html") %>%
kableExtra::kable_styling(bootstrap_options = c("striped"))
perc25_3_output %>%
dplyr::select(`Toxic Score Percentile`, hispanic:non_hispanic, total) %>%
arrange(desc(`Toxic Score Percentile`)) %>%
knitr::kable(., caption = "p-value = 0.91, column proportions do not differ
significantly from expected frequencies",
format = "html") %>%
kableExtra::kable_styling(bootstrap_options = c("striped"))
perc25_3_output %>%
dplyr::select(`Toxic Score Percentile`, low_income:non_low_income, total)
%>%
arrange(desc(`Toxic Score Percentile`)) %>%
knitr::kable(., caption = "p-value = 0.98, column proportions differ
significantly from expected frequencies",
format = "html") %>%
kableExtra::kable_styling(bootstrap_options = c("striped"))
#3. Write a short summary of your findings (no more than 2 pages), include a
map, and attach on a third page your commented R code used in the analysis.
ggplot() +
geom_sf(data = tox_buf_3, color = NA, fill = "gray", alpha = 0.4) +
geom_sf(data = tox_buf_1, color = NA, fill = "black", alpha = 0.4) +
geom_sf(data = tox, shape = 17, color = "yellow", size = 2) +
geom_sf(data = sch) +
geom_sf(data = sch_tox %>% filter(tox_1 == TRUE), color = "maroon", size =
2) +
geom_sf(data = sch_tox %>% filter(tox_3 == TRUE & tox_1 == FALSE), color =
"orange", size = 2) +
geom_sf(data = dallas, fill = NA) +
geom_sf(data = hwy2, color = "light gray") +
labs(title = "School Exposure to Toxic Sites",
subtitle = "Dallas County, Texas") +
hrbrthemes::theme_ipsum_tw()

```