Sean Connelly

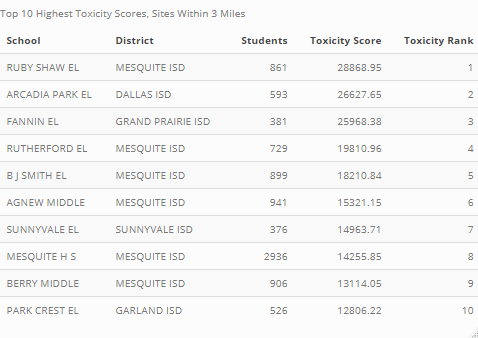
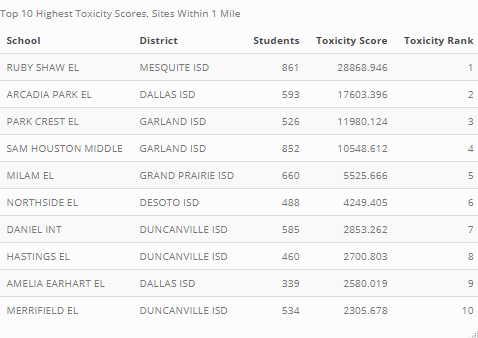
UPP 465

Professor Tilahun

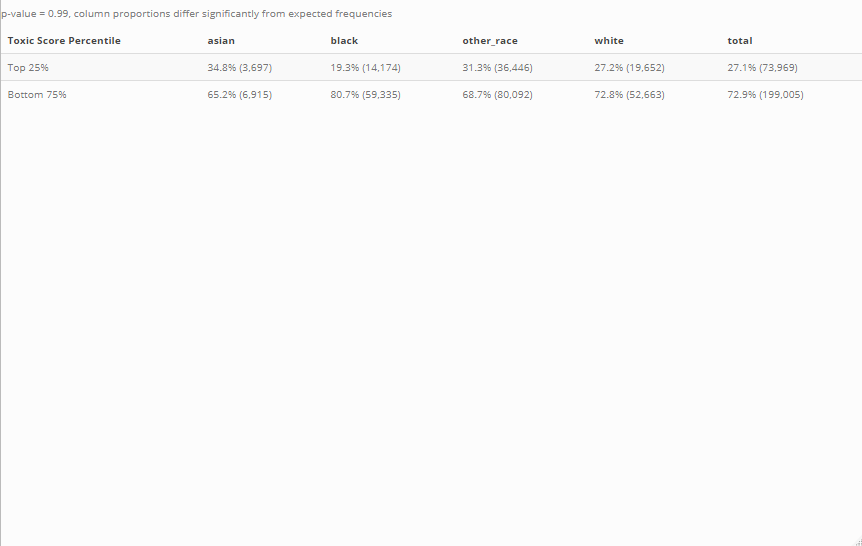
2/17/20

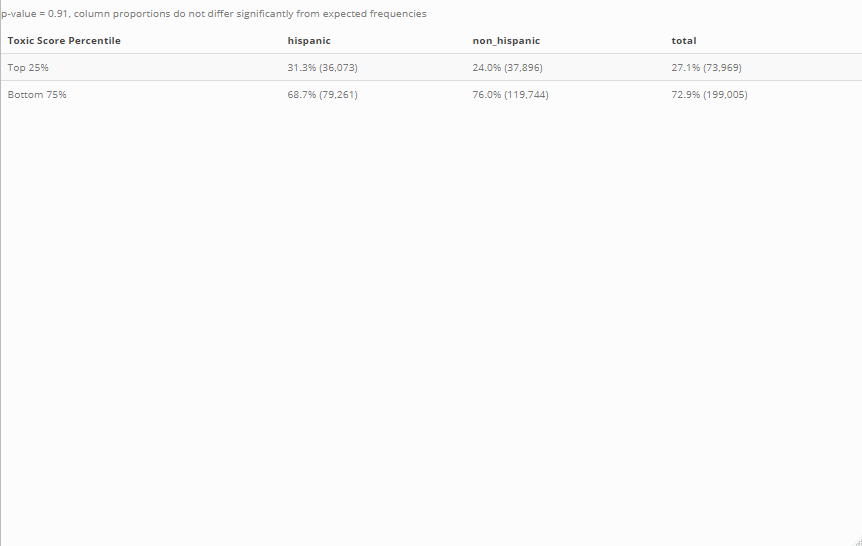
Spatial Statistics – Spatial Analysis Lab

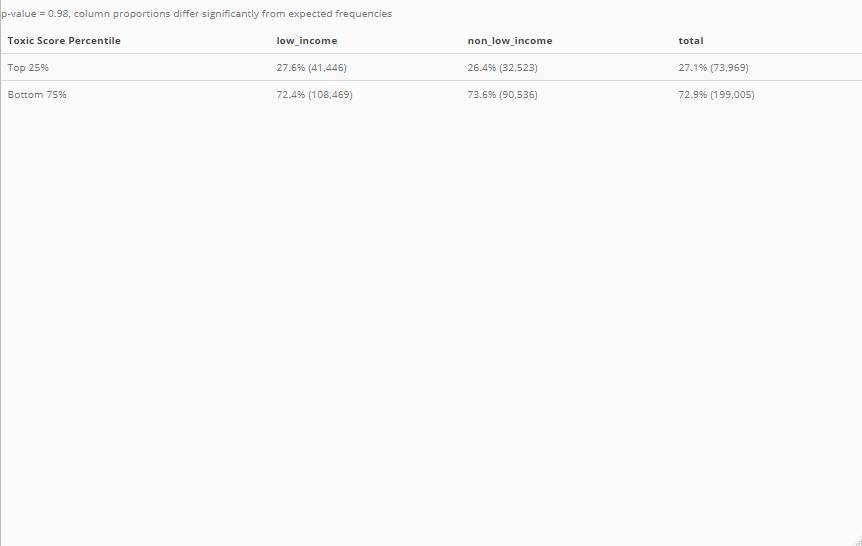
According to my analysis, three schools make the top 10 list for toxicity exposure using a one-mile buffer and a three-mile buffer: Ruby Shaw Elementary, Arcadia Park Elementary, and Park Crest Elementary. Widening the buffer to three miles increases the number of toxic sites contributing to an individual 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.



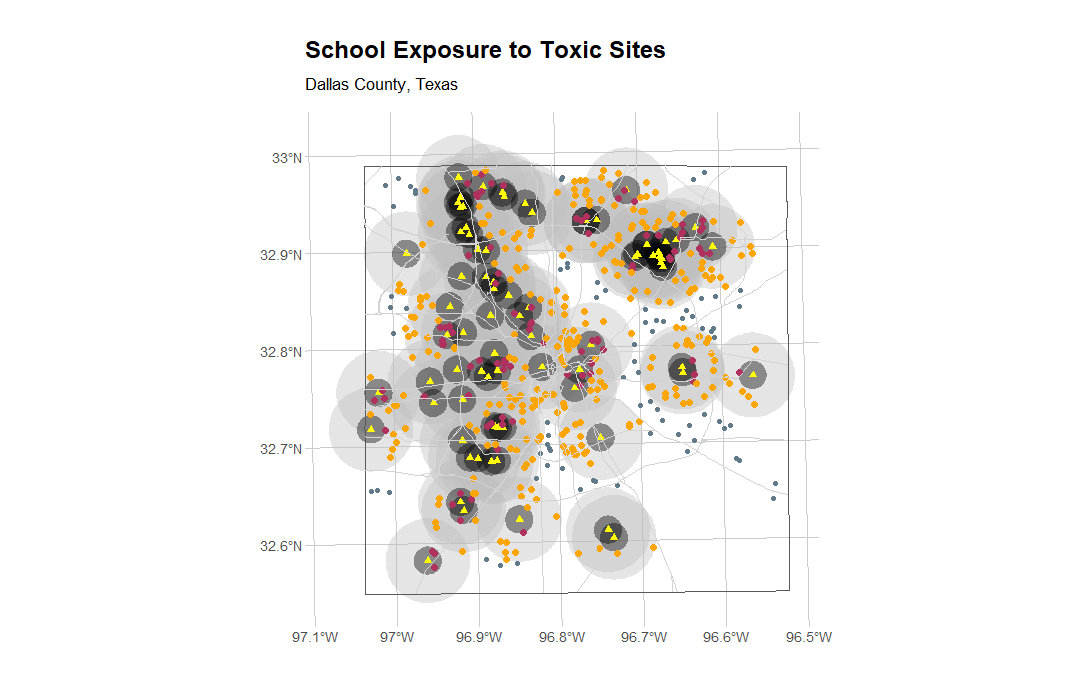
Using the newly created scores for the three-mile buffer, top 25 percent of schools in terms of toxicity exposure were compared to the bottom 75 percent. The top 25 percent of schools contained 27 percent of the overall population (a little under 74,000 people), but almost 35 percent of the county’s 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.







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.



#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#--------------------------------------------------------------------------------

########################################################

#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(hapefiles)

########################################################

#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 thorugh 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()