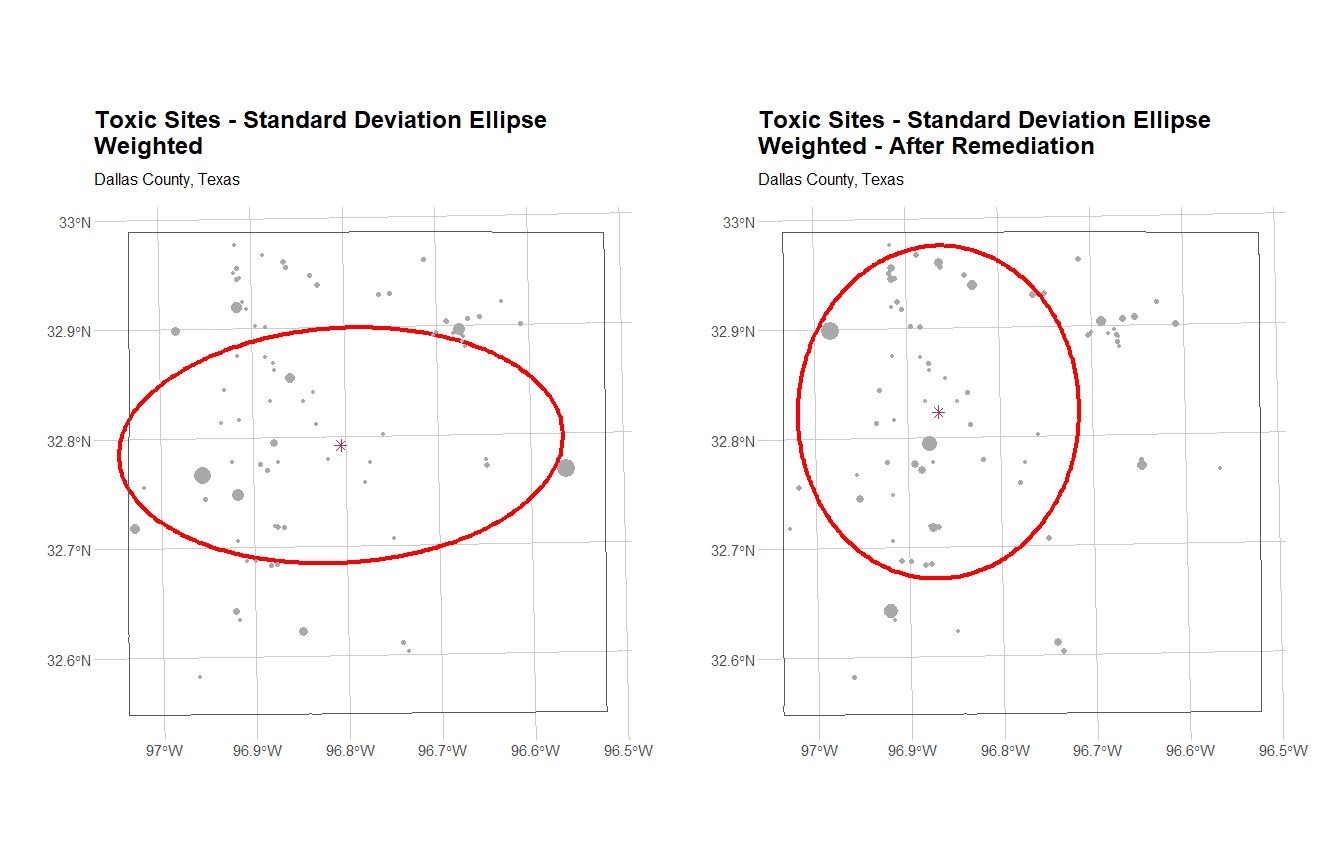
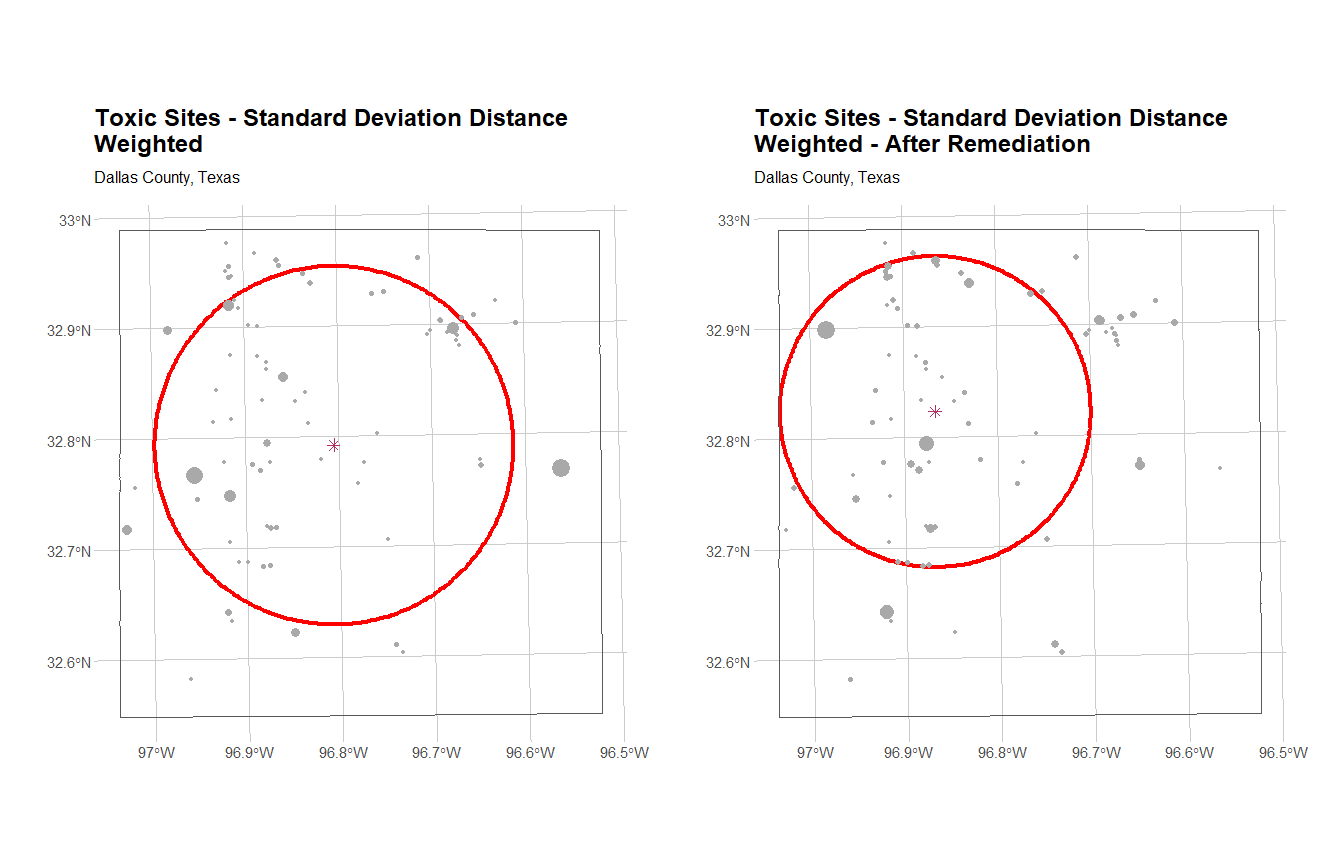
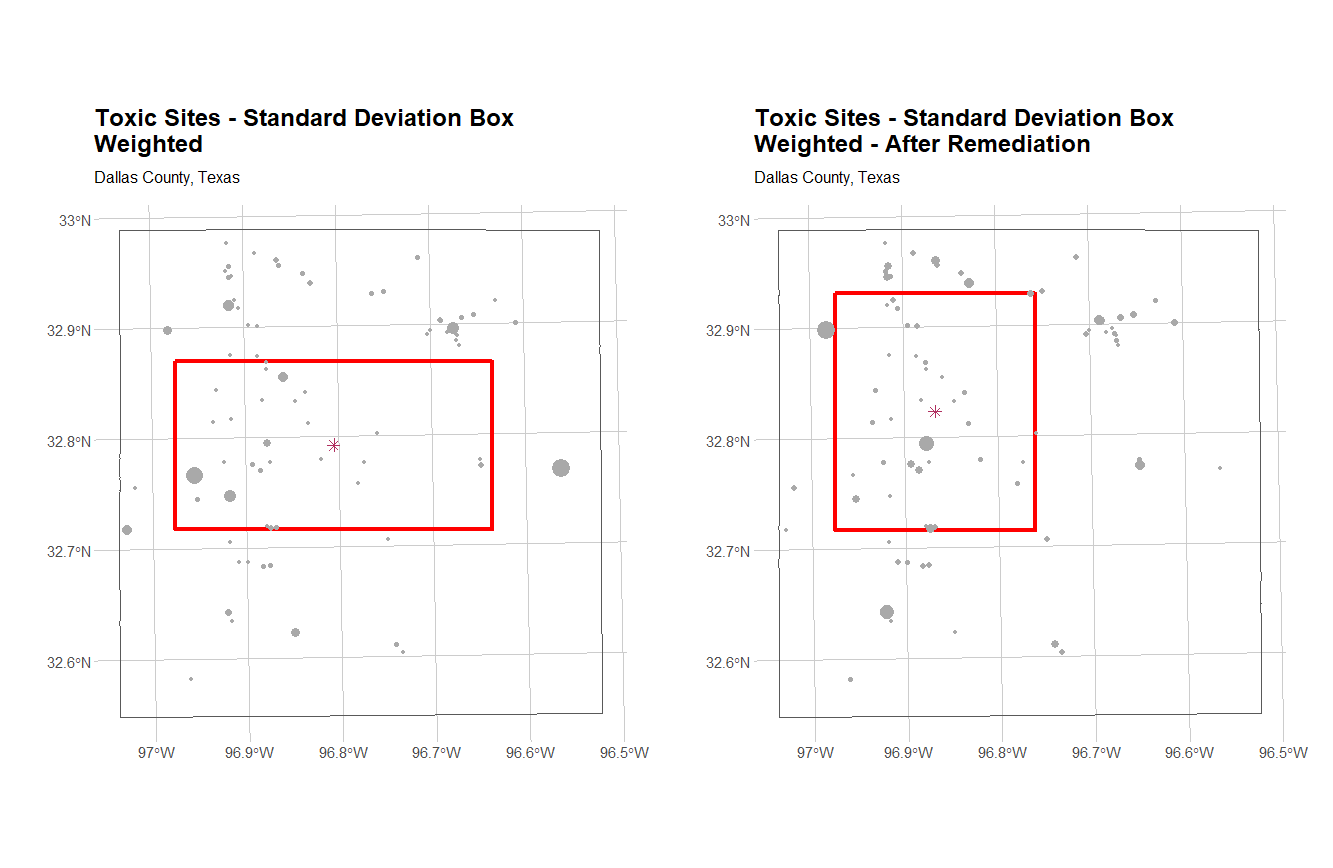
Sean Connelly

UPP 465

Professor Tilahun

2/20/20

Spatial Statistics – Exploratory Spatial Data Analysis Lab

The centrographic analysis plots above demonstrate the impact of a remediation policy that zeroed out the toxicity of the most dangerous sites (top 10%). The standard deviation box reorients to North-South following the intervention, which addressed some severe polluters in the Eastern part of Dallas. The center also shifts West by a meaningful distance. The standard deviation distance indicates that the remediation has decreased the geographical spread of toxic sites, resulting in a smaller radius. The standard deviation ellipse shows the same reorientation as the SSB; however, the axis is lengthened North-South, suggesting toxic sites are widespread along that axis. The most preferable measure depends on context, but the SDE does a good job of adding some complexity to the relatively simplistic SSB and SDD plots. The remaining pages of this document contain a log of my code.

# Load libraries ----------------------------------------------------------

#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(), "/03\_ESDA"))

# 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(shapefiles)

library(aspace)

options(stringsAsFactors = FALSE, scipen = 999)

source('centrography\_SC.R') # modified functions from the aspace library (SC

changes for simple features)

# Import data -------------------------------------------------------------

# Simple features style

# Homicides - IL State Plan East CRS 26971

homicides <- readRDS("Raw\_Data/Homicides02\_10\_19.rds")

homicides <- st\_as\_sf(homicides)

# CPD and other crime data

cpd\_districts <- readRDS("Raw\_Data/CPD.rds")

cpd\_districts <- st\_as\_sf(cpd\_districts)

chi\_boundary <- readRDS("Raw\_Data/Chicago\_agg.rds")#this was created using the

following command>>>> #cpd2<-aggregat(cpd)

chi\_boundary <- st\_as\_sf(chi\_boundary)

#...On your own, follow along on slides

# Homicides by year and arrest status

homicides %>%

st\_drop\_geometry() %>%

tabyl(year, arrest) %>%

adorn\_totals(c("row", "col")) %>%

adorn\_percentages() %>%

adorn\_pct\_formatting() %>%

adorn\_ns()

# Data frames by year

h02 <- homicides %>% filter(year == 2002)

h10 <- homicides %>% filter(year == 2010)

h19 <- homicides %>% filter(year == 2019)

# Plot by arrest status

ggplot() +

geom\_sf(data = cpd\_districts, color = "dark gray", fill = NA, size = 1) +

geom\_sf(data = chi\_boundary, color = "black", fill = NA, size = 1.5) +

geom\_sf(data = homicides, aes(color = arrest), alpha = 0.7, show.legend =

"point") +

hrbrthemes::theme\_ipsum()

# Centrographic analysis

-------------------------------------------------------------------

#Standard deviation box

h02b <- calc\_box2(points = coordinates(as\_Spatial(h02)))

plot\_box2(h02b, points.pch = 3, points.col = 8, box.col = 1, centre.col = 4,

titletxt = "Center and Dispersion for 2002 Homicides")

#Standard deviation distance (standard circle)

h02c <- calc\_sdd2(points = coordinates(as\_Spatial(h02)))

plot\_sdd2(h02c, points.pch = 3, points.col = 8, sdd.col = 1,

titletxt = "2002 Homicides")

#standard deviational ellipse

h02e <- calc\_sde2(points = coordinates(as\_Spatial(h02)))

plot\_sde2(h02e, points.pch = 3, points.col = 8, sde.col = 1,

titletxt = "2002 Homicides")

# Centrographic analysis 2019 --------------------------------------------

#Standard deviation box

h19b <- calc\_box2(points = coordinates(as\_Spatial(h19)))

plot\_box2(h19b, plotnew = TRUE, points.pch = 3, points.col = 8, centre.col =

2, box.col = 2,

titletxt = "2019 Homicides")

#Standard deviation distance (standard circle)

h19c <- calc\_sdd2(points = coordinates(as\_Spatial(h19)))

plot\_sdd2(h19c, plotnew = TRUE, points.pch = 3, points.col = 8, centre.col =

2, sdd.col = 2,

titletxt = "2019 Homicides")

#standard deviational ellipse

h19e <- calc\_sde2(points = coordinates(as\_Spatial(h19)))

plot\_sde2(h19e, plotnew = TRUE, points.pch = 3, sde.col = 2, points.col = 8,

centre.col = 2,

titletxt = "2019 Homicides")

# 2002 vs 2019 ------------------------------------------------------------

# Compare 2002 and 2019 SD box

plot\_box2(h02b, points.pch = 3, points.col = 8, box.col = 1, centre.col =

4 ,

titletxt = "Center and Dispersion for 2002 Homicides")

plot\_box2(h19b, plotnew = FALSE, points.pch = 3, points.col = 0, centre.col =

2,

box.col = 2, titletxt = "")

# Compare 2002 and 2019 circles

plot\_sdd2(h02c, plotnew = TRUE, points.pch = 3, points.col = 8, sdd.col = 1

,

titletxt = "2002 and 2019 Homicides")

plot\_sdd2(h19c, plotnew = FALSE, points.pch = 3, points.col = 0, centre.col =

2,

sdd.col = 2, titletxt = "")

#compare 2002 and 2019 ellipses

plot\_sde2(h02e, points.pch = 3, points.col = 8, sde.col = 1 ,

titletxt = "2002 and 2019 Homicides")

plot\_sde2(h19e, plotnew = FALSE, points.pch = 3, sde.col = 2,

points.col = 0, centre.col = 2, titletxt = "")

###### ====================================================================

# Start of submission

###### ====================================================================

# Read in the toxic sites data from last week -----------------------------

# Import data

tox <- readRDS("Raw\_Data/toxDallas.rds")

dallas <- readRDS("Raw\_Data/Dallas.rds")

plot(dallas)

plot(tox, add = TRUE)

# Plot the toxic sites by toxicity levels

summary(tox$SCORE)

summary(scale(tox$SCORE))

plot(dallas)

plot(tox, cex = scale(tox$SCORE)+1, add = TRUE)

#Answer the questions on the slide

# Open the toxic sites data from Dallas you used last last week

# Convert spatial data to simple features objects

dallas\_sf <- st\_as\_sf(dallas, crs = 102738)

tox\_sf <- st\_as\_sf(tox, crs = 102738)

# Set CRS (102738 = NAD\_1983\_StatePlane\_Texas\_North\_Central\_FIPS\_4202\_Feet)

dallas\_sf <- dallas\_sf %>% st\_transform(crs = 102738)

tox\_sf <- tox\_sf %>% st\_transform(crs = 102738)

# Examine the centrographic statistics for the toxic sites data using the

three measures of dispersion (box, circle, ellipse) and note any differences

# 1. Examine the dispersion without any weights

# Standard deviation box

tox\_ssb\_raw <- calc\_box2(points = coordinates(as\_Spatial(tox\_sf)))

tox\_centre <- ssb\_as\_sf(tox\_ssb\_raw, orig\_crs = 102738, centre\_dummy = TRUE)

tox\_ssb <- ssb\_as\_sf(tox\_ssb\_raw, orig\_crs = 102738)

# Plot

plot\_tox\_ssb <- ggplot() +

geom\_sf(data = tox\_ssb, color = "light blue", fill = NA, size = 1.5) +

geom\_sf(data = tox\_centre, color = "blue", shape = 8, size = 3) +

geom\_sf(data = tox\_sf, color = "dark gray", size = 2) +

geom\_sf(data = dallas\_sf, fill = NA, size = 0.5) +

labs(title = "Toxic Sites - Standard Deviation Box\nUnweighted",

subtitle = "Dallas County, Texas") +

hrbrthemes::theme\_ipsum()

plot\_tox\_ssb

# plot\_box2(tox\_ssb\_raw, plotnew = TRUE, points.pch = 3, points.col = 8,

centre.col = 2, box.col = 2,

# titletxt = "Toxic Sites in Dallas")

# Standard deviation distance (standard circle)

tox\_sdd\_raw <- calc\_sdd2(points = coordinates(as\_Spatial(tox\_sf)))

tox\_sdd <- sdd\_as\_sf(tox\_ssd\_raw, orig\_crs = 102738)

# Plot

plot\_tox\_sdd <- ggplot() +

geom\_sf(data = tox\_sdd, color = "light blue", fill = NA, size = 1.5) +

geom\_sf(data = tox\_centre, color = "blue", shape = 8, size = 3) +

geom\_sf(data = tox\_sf, color = "dark gray", size = 2) +

geom\_sf(data = dallas\_sf, fill = NA, size = 0.5) +

labs(title = "Toxic Sites - Standard Deviation Distance\nUnweighted",

subtitle = "Dallas County, Texas") +

hrbrthemes::theme\_ipsum()

plot\_tox\_sdd

# plot\_sdd2(tox\_sdd\_raw, plotnew = TRUE, points.pch = 3, points.col = 8,

centre.col = 2, sdd.col = 2,

# titletxt = "Toxic Sites in Dallas")

#standard deviational ellipse

tox\_sde\_raw <- calc\_sde2(points = coordinates(as\_Spatial(tox\_sf)))

tox\_sde <- sde\_as\_sf(tox\_sde\_raw, orig\_crs = 102738)

# Plot

plot\_tox\_sde <- ggplot() +

geom\_sf(data = tox\_sde, color = "light blue", fill = NA, size = 1.5) +

geom\_sf(data = tox\_centre, color = "blue", shape = 8, size = 3) +

geom\_sf(data = tox\_sf, color = "dark gray", size = 2) +

geom\_sf(data = dallas\_sf, fill = NA, size = 0.5) +

labs(title = "Toxic Sites - Standard Deviation Ellipse\nUnweighted",

subtitle = "Dallas County, Texas") +

hrbrthemes::theme\_ipsum()

plot\_tox\_sde

# plot\_sde2(tox\_sde\_raw, plotnew = TRUE, points.pch = 3, points.col = 8,

centre.col = 2, sde.col = 2,

# titletxt = "Toxic Sites in Dallas")

#2. Examine the dispersion using the toxicity score (SCORE) variable as

weight. Note any differences between these two.

# The code for each of the dispersion measures has options for adding weights

as follows:

#bW<-calc\_box(coordinates(tox),weighted=TRUE,weights=tox$SCORE)

# Standard deviation box

tox\_ssb\_raw\_weighted <- calc\_box2(points = coordinates(as\_Spatial(tox\_sf)),

weighted = TRUE, weights = tox\_sf$SCORE)

tox\_centre\_weighted <- ssb\_as\_sf(tox\_ssb\_raw\_weighted, orig\_crs = 102738,

centre\_dummy = TRUE)

tox\_ssb\_weighted <- ssb\_as\_sf(tox\_ssb\_raw\_weighted, orig\_crs = 102738)

# Plot

plot\_tox\_ssb\_weighted <- ggplot() +

geom\_sf(data = tox\_ssb\_weighted, color = "red", fill = NA, size = 1.5) +

geom\_sf(data = tox\_centre\_weighted, color = "maroon", shape = 8, size = 3) +

geom\_sf(data = tox\_sf, aes(size = SCORE), color = "dark gray", show.legend =

FALSE) +

geom\_sf(data = dallas\_sf, fill = NA, size = 0.5) +

labs(title = "Toxic Sites - Standard Deviation Box\nWeighted",

subtitle = "Dallas County, Texas") +

hrbrthemes::theme\_ipsum()

plot\_tox\_ssb\_weighted

# plot\_box2(tox\_ssb\_raw\_weighted, plotnew = TRUE, points.pch = 3, points.col =

8, centre.col = 2, box.col = 2,

# titletxt = "Toxic Sites in Dallas")

# Standard deviation distance (standard circle)

tox\_sdd\_raw\_weighted <- calc\_sdd2(points = coordinates(as\_Spatial(tox\_sf)),

weighted = TRUE, weights = tox\_sf$SCORE)

tox\_sdd\_weighted <- sdd\_as\_sf(tox\_sdd\_raw\_weighted, orig\_crs = 102738)

# Plot

plot\_tox\_sdd\_weighted <- ggplot() +

geom\_sf(data = tox\_sdd\_weighted, color = "red", fill = NA, size = 1.5) +

geom\_sf(data = tox\_centre\_weighted, color = "maroon", shape = 8, size = 3) +

geom\_sf(data = tox\_sf, aes(size = SCORE), color = "dark gray", show.legend =

FALSE) +

geom\_sf(data = dallas\_sf, fill = NA, size = 0.5) +

labs(title = "Toxic Sites - Standard Deviation Distance\nWeighted",

subtitle = "Dallas County, Texas") +

hrbrthemes::theme\_ipsum()

plot\_tox\_sdd\_weighted

# plot\_sdd2(tox\_sdd\_raw\_weighted, plotnew = TRUE, points.pch = 3, points.col =

8, centre.col = 2, sdd.col = 2,

# titletxt = "Toxic Sites in Dallas")

#standard deviational ellipse

tox\_sde\_raw\_weighted <- calc\_sde2(points = coordinates(as\_Spatial(tox\_sf)),

weighted = TRUE, weights = tox\_sf$SCORE)

tox\_sde\_weighted <- sde\_as\_sf(tox\_sde\_raw\_weighted, orig\_crs = 102738)

# Plot

plot\_tox\_sde\_weighted <- ggplot() +

geom\_sf(data = tox\_sde\_weighted, color = "red", fill = NA, size = 1.5) +

geom\_sf(data = tox\_centre\_weighted, color = "maroon", shape = 8, size = 3) +

geom\_sf(data = tox\_sf, aes(size = SCORE), color = "dark gray", show.legend =

FALSE) +

geom\_sf(data = dallas\_sf, fill = NA, size = 0.5) +

labs(title = "Toxic Sites - Standard Deviation Ellipse\nWeighted",

subtitle = "Dallas County, Texas") +

hrbrthemes::theme\_ipsum()

plot\_tox\_sde\_weighted

# plot\_sde2(tox\_sde\_raw\_weighted, plotnew = TRUE, points.pch = 3, points.col =

8, centre.col = 2, sde.col = 2,

# titletxt = "Toxic Sites in Dallas")

# Plots side by side for comparison

plot\_tox\_ssb | plot\_tox\_ssb\_weighted

plot\_tox\_sdd | plot\_tox\_sdd\_weighted

plot\_tox\_sde | plot\_tox\_sde\_weighted

# ANSWER: You can see that weighting by toxicity score changes the

centrographic measures pretty drastically. The unweighted plots are influenced

by the cluster of toxic sites in the Northwest, however, the most toxic sites

are concentrated on the West and East Sides of town. This means that the

weighted SSB, SDD, and SDE are all elongated horizontally and the center point

is farther East.

#3. Suppose a remediation effort targeted the top 10% of the the toxic sites.

Assume their scores went down to 0. Using plots, compare the standard

deviation box, standard distance deviation, standard deviation ellipse before

and after remediation using weights. Note any changes.

# Scores after remediation

tox\_remediated\_sf <- tox\_sf %>%

mutate(tox\_ptile = ntile(SCORE, 100),

SCORE = ifelse(tox\_ptile >= 90, 0, SCORE))

# Standard deviation box

tox\_ssb\_raw\_remediated <- calc\_box2(points =

coordinates(as\_Spatial(tox\_remediated\_sf)),

weighted = TRUE, weights =

tox\_remediated\_sf$SCORE)

tox\_centre\_remediated <- ssb\_as\_sf(tox\_ssb\_raw\_remediated, orig\_crs = 102738,

centre\_dummy = TRUE)

tox\_ssb\_remediated <- ssb\_as\_sf(tox\_ssb\_raw\_remediated, orig\_crs = 102738)

# Plot

plot\_tox\_ssb\_remediated <- ggplot() +

geom\_sf(data = tox\_ssb\_remediated, color = "red", fill = NA, size = 1.5) +

geom\_sf(data = tox\_centre\_remediated, color = "maroon", shape = 8, size = 3)

+

geom\_sf(data = tox\_remediated\_sf, aes(size = SCORE), color = "dark gray",

show.legend = FALSE) +

geom\_sf(data = dallas\_sf, fill = NA, size = 0.5) +

labs(title = "Toxic Sites - Standard Deviation Box\nWeighted - After

Remediation",

subtitle = "Dallas County, Texas") +

hrbrthemes::theme\_ipsum()

plot\_tox\_ssb\_remediated

# plot\_box2(tox\_ssb\_raw\_remediated, plotnew = TRUE, points.pch = 3, points.col

= 8, centre.col = 2, box.col = 2,

# titletxt = "Toxic Sites in Dallas")

# Standard deviation distance (standard circle)

tox\_sdd\_raw\_remediated <- calc\_sdd2(points =

coordinates(as\_Spatial(tox\_remediated\_sf)),

weighted = TRUE, weights =

tox\_remediated\_sf$SCORE)

tox\_sdd\_remediated <- sdd\_as\_sf(tox\_sdd\_raw\_remediated, orig\_crs = 102738)

# Plot

plot\_tox\_sdd\_remediated <- ggplot() +

geom\_sf(data = tox\_sdd\_remediated, color = "red", fill = NA, size = 1.5) +

geom\_sf(data = tox\_centre\_remediated, color = "maroon", shape = 8, size = 3)

+

geom\_sf(data = tox\_remediated\_sf, aes(size = SCORE), color = "dark gray",

show.legend = FALSE) +

geom\_sf(data = dallas\_sf, fill = NA, size = 0.5) +

labs(title = "Toxic Sites - Standard Deviation Distance\nWeighted - After

Remediation",

subtitle = "Dallas County, Texas") +

hrbrthemes::theme\_ipsum()

plot\_tox\_sdd\_remediated

# plot\_sdd2(tox\_sdd\_raw\_remediated, plotnew = TRUE, points.pch = 3, points.col

= 8, centre.col = 2, sdd.col = 2,

# titletxt = "Toxic Sites in Dallas")

#standard deviational ellipse

tox\_sde\_raw\_remediated <- calc\_sde2(points =

coordinates(as\_Spatial(tox\_remediated\_sf)),

weighted = TRUE, weights =

tox\_remediated\_sf$SCORE)

tox\_sde\_remediated <- sde\_as\_sf(tox\_sde\_raw\_remediated, orig\_crs = 102738)

# Plot

plot\_tox\_sde\_remediated <- ggplot() +

geom\_sf(data = tox\_sde\_remediated, color = "red", fill = NA, size = 1.5) +

geom\_sf(data = tox\_centre\_remediated, color = "maroon", shape = 8, size = 3)

+

geom\_sf(data = tox\_remediated\_sf, aes(size = SCORE), color = "dark gray",

show.legend = FALSE) +

geom\_sf(data = dallas\_sf, fill = NA, size = 0.5) +

labs(title = "Toxic Sites - Standard Deviation Ellipse\nWeighted - After

Remediation",

subtitle = "Dallas County, Texas") +

hrbrthemes::theme\_ipsum()

plot\_tox\_sde\_remediated

# plot\_sde2(tox\_sde\_raw\_remediated, plotnew = TRUE, points.pch = 3, points.col

= 8, centre.col = 2, sde.col = 2,

# titletxt = "Toxic Sites in Dallas")

#Submission

# Submit your centrographic analysis plots for the toxic sites in Dallas using

the toxic scores as weights.

# # Plot the standard deviation box before and after remediation

plot\_tox\_ssb\_weighted | plot\_tox\_ssb\_remediated

# # Plot the standard distance deviation circle before and after remediation

plot\_tox\_sdd\_weighted | plot\_tox\_sdd\_remediated

# # Plot the standard deviational ellipse before and after remediation

plot\_tox\_sde\_weighted | plot\_tox\_sde\_remediated

#Write a short paragraph describing which measure is preferable and why.