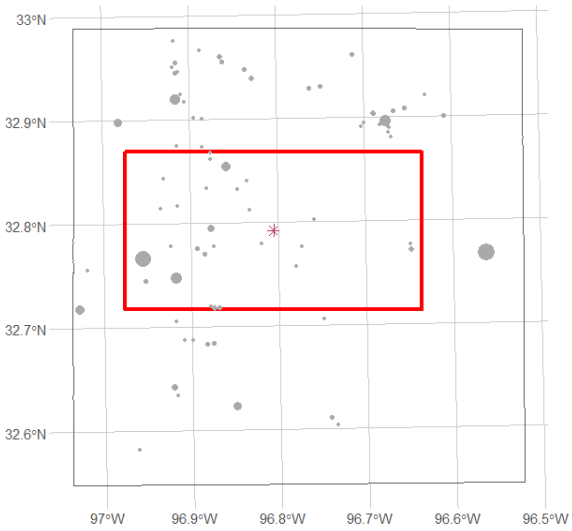
Sean Connelly
UPP 465
Professor Tilahun
2/20/20

Spatial Statistics – Exploratory Spatial Data Analysis Lab

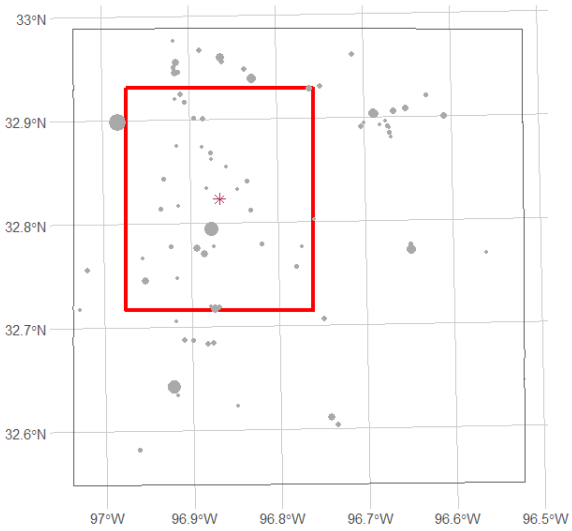
Toxic Sites - Standard Deviation Box Weighted

Dallas County, Texas



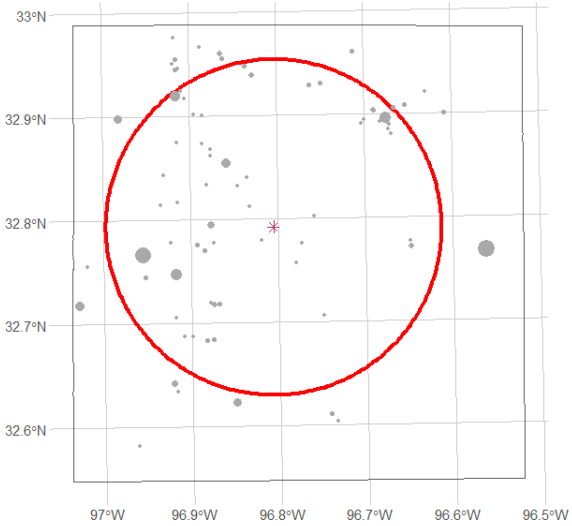
Toxic Sites - Standard Deviation Box Weighted - After Remediation

Dallas County, Texas



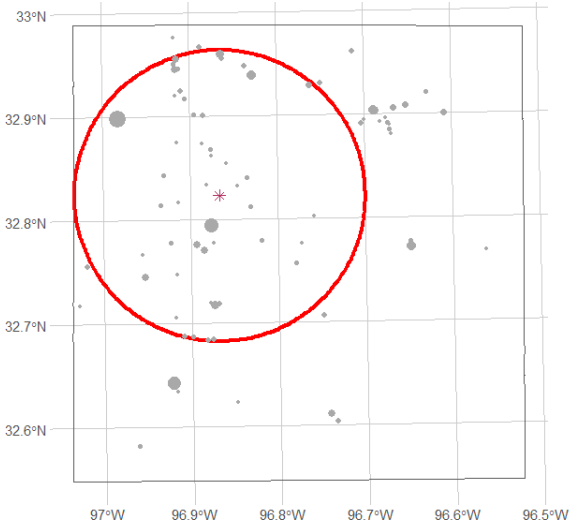
Toxic Sites - Standard Deviation Distance Weighted

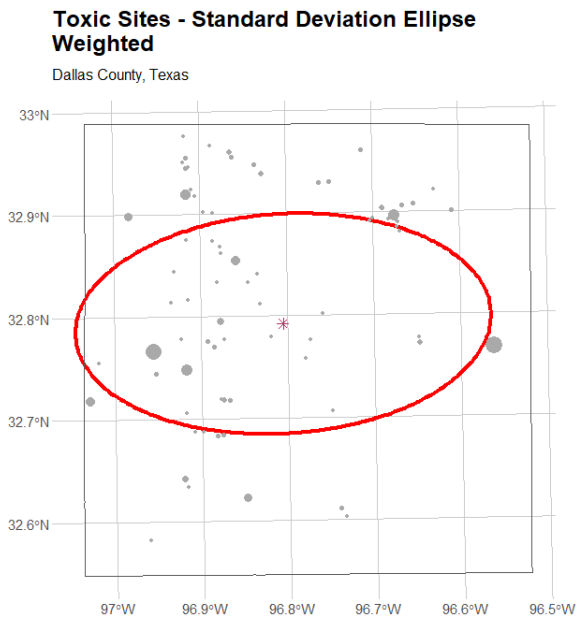
Dallas County, Texas

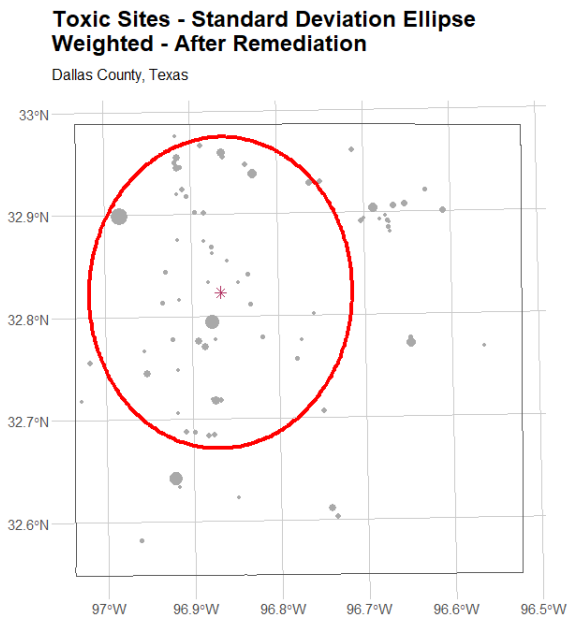


Toxic Sites - Standard Deviation Distance Weighted - After Remediation

Dallas County, Texas







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.
```