

Assessing the risk of contamination from hazardous sites due to flooding in North Carolina low-socioeconomic communities

https://github.com/tristen0708/EJ_Project

Tristen Townsend

Abstract

It is well known that hazardous waste sites tend to be more frequently sited in low-income communities, particularly communities of color. Proximity to hazardous waste sites have potential to create health risks, especially if communities are in regions more likely to be prone to flooding or extreme precipitation. This analysis serves to understand whether risks might exist for low-income communities in North Carolina. The number and type of sites in various counties has been analyzed in order to see if minority communities are more at risk.

Contents

1	Research Question and Rationale	3
2	Dataset Information	4
3	Exploratory Data Analysis and Wrangling	5
4	Analysis	14
5	Summary and Conclusions	29
6	References	30

1 Research Question and Rationale

There is significant research that indicates hazardous waste sites, especially those listed on the National Priorities List as a Superfund site, are disproportionately located in communities of color or low socioeconomic status (Burwell-Naney et al., 2013; Kramar, Anderson, Hilfer, Branden, Gutrich, 2018). In North Carolina, recent hurricanes have resulted in serious flooding in many parts of the state, creating concern as to whether Superfund sites and other hazardous waste sites were breached and might pose health effects to local communities. Given that natural disasters such as hurricanes and increased flooding is expect in North Carolina, it is important to understand if risks related to hazardous waste sites and flooding are posted to minority communities, as they tend to be low-capacity and less resilient to disasters. This information could be used by environmental justice leaders to advocate for policy changes or the implementation of safeguards to be put in place.

During this analysis, I am interested in seeing if the number of hazardous waste sites is significantly greater in areas with a large percent of minority resident or a large amount of those in poverty. Additionally, for the areas with the highest site counts, I would like to see if they are at risk for flood events based on peak stage values. Poverty and race data by county will be used, as well as site data and peak stage data.

2 Dataset Information

I am using multiple datasets to answer my research question. I have downloaded geospatial data on hazardous waste site locations from North Carolina Department of Environmental Quality. I have downloaded data with information about poverty and race from the U.S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) Program and (U.S. Census Bureau, 2018) and from the NC Budget and Management LINC data retrieval tool. Furthermore, I've obtained peak stage data for a few counties in North Carolina over that past five years from the U.S. Geological Survey.

```
Poverty_NC <- read.csv("./Data/Processed/NC_Poverty_processed.csv")  
Race <- read.csv("./Data/Processed/LINC_RaceData_2010.csv")  
Peak.Stage <- read.csv("./Data/Processed/FilteredPeaks.csv")
```

3 Exploratory Data Analysis and Wrangling

```
Poverty_NC <- read.csv("../Data/Processed/NC_Poverty_processed.csv")
Race <- read.csv("../Data/Processed/LINC_RaceData_2010.csv")
Peak.Stage <- read.csv("../Data/Processed/FilteredPeaks.csv")

Peak.Stage_processed <- Peak.Stage %>% select("county", "latitude_dd",
      "longitude_dd", "vdatum", "peak_date", "peak_stage")

Race1 <- mutate(Race, WhitePerc = Race$White/Race$Total.Pop)
Race2 <- mutate(Race1, BlackPerc = Race$Black/Race$Total.Pop)
Race3 <- mutate(Race2, NativePerc = Race$Native/Race$Total.Pop)
Race4 <- mutate(Race3, AsianPerc = Race$Asian/Race$Total.Pop)
Race5 <- mutate(Race4, HispanicPerc = Race$Hispanic/Race$Total.Pop)
Race6 <- mutate(Race5, OtherPerc = Race$Other/Race$Total.Pop)
Race7 <- mutate(Race6, Haw_PacIsPerc = Race$Haw_PacIs/Race$Total.Pop)
Race8 <- mutate(Race7, TwoPerc = Race$Two/Race$Total.Pop)
Race_Processed <- mutate(Race8, MinorityPerc = (Race$Black +
      Race$Native + Race$Asian + Race$Hispanic + Race$Other +
      Race$Haw_PacIs + Race$Two)/Race$Total.Pop)

#Explore race and poverty data
shapiro.test(Race_Processed$MinorityPerc) #p-value = 0.0005985

##
##  Shapiro-Wilk normality test
##
## data:  Race_Processed$MinorityPerc
## W = 0.94783, p-value = 0.0005985

shapiro.test(Poverty_NC$Poverty_Percent_allages) #p-value 0.01881

##
##  Shapiro-Wilk normality test
##
## data:  Poverty_NC$Poverty_Percent_allages
## W = 0.96908, p-value = 0.01881

#From the output, both p-values are < 0.05 implying that the distribution
#of the data are significantly different from normal distribution.
#In other words, we cannot assume the normality for either data sets.

#Explore discharge data
Peak.Stage_avg <- Peak.Stage_processed %>%
  group_by(county) %>%
```

```

summarise(yearly_average = mean(peak_stage))

shapiro.test(Peak.Stage_avg$yearly_average) #p-value = 0.288 normal!

##
##  Shapiro-Wilk normality test
##
## data:  Peak.Stage_avg$yearly_average
## W = 0.9346, p-value = 0.288

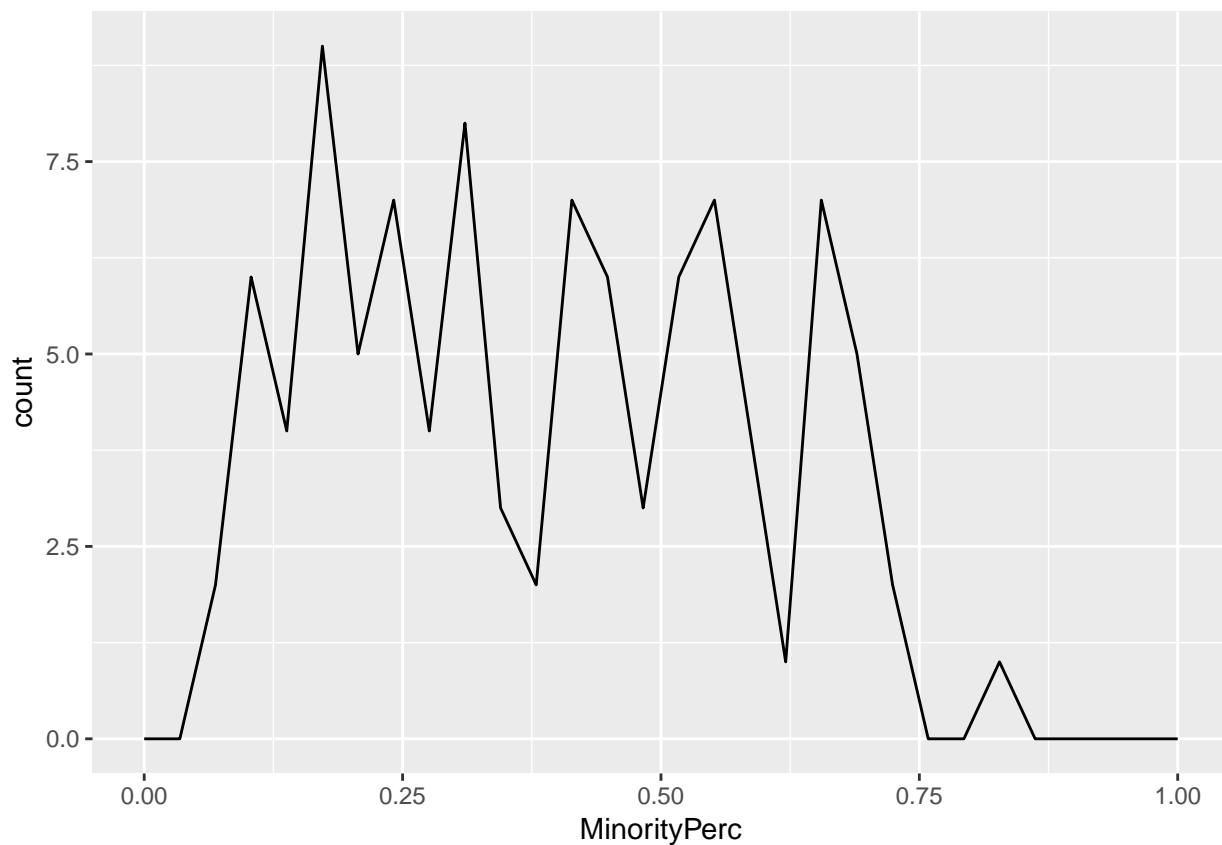
```

*#From the output, the p-value is < 0.05 implying that the distribution
#of the data are significantly different from normal distribution.
#In other words, we cannot assume the normality for either data sets.*

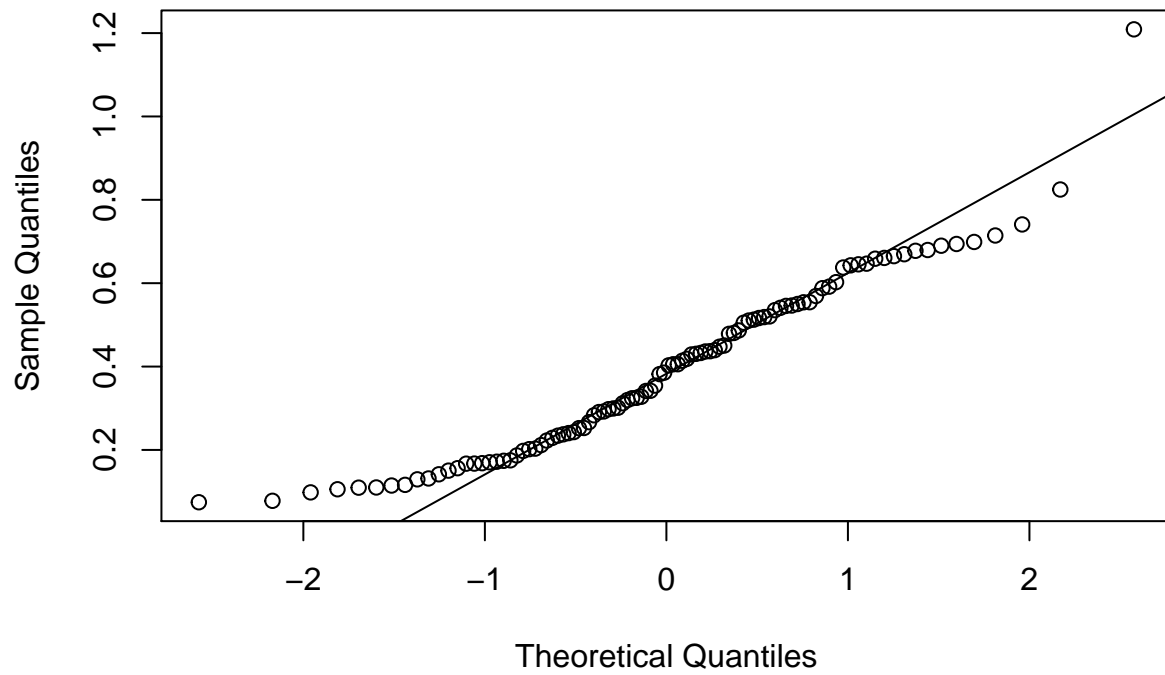
```

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 1164 rows containing non-finite values (stat_bin).
## Warning: Removed 2 rows containing missing values (geom_path).

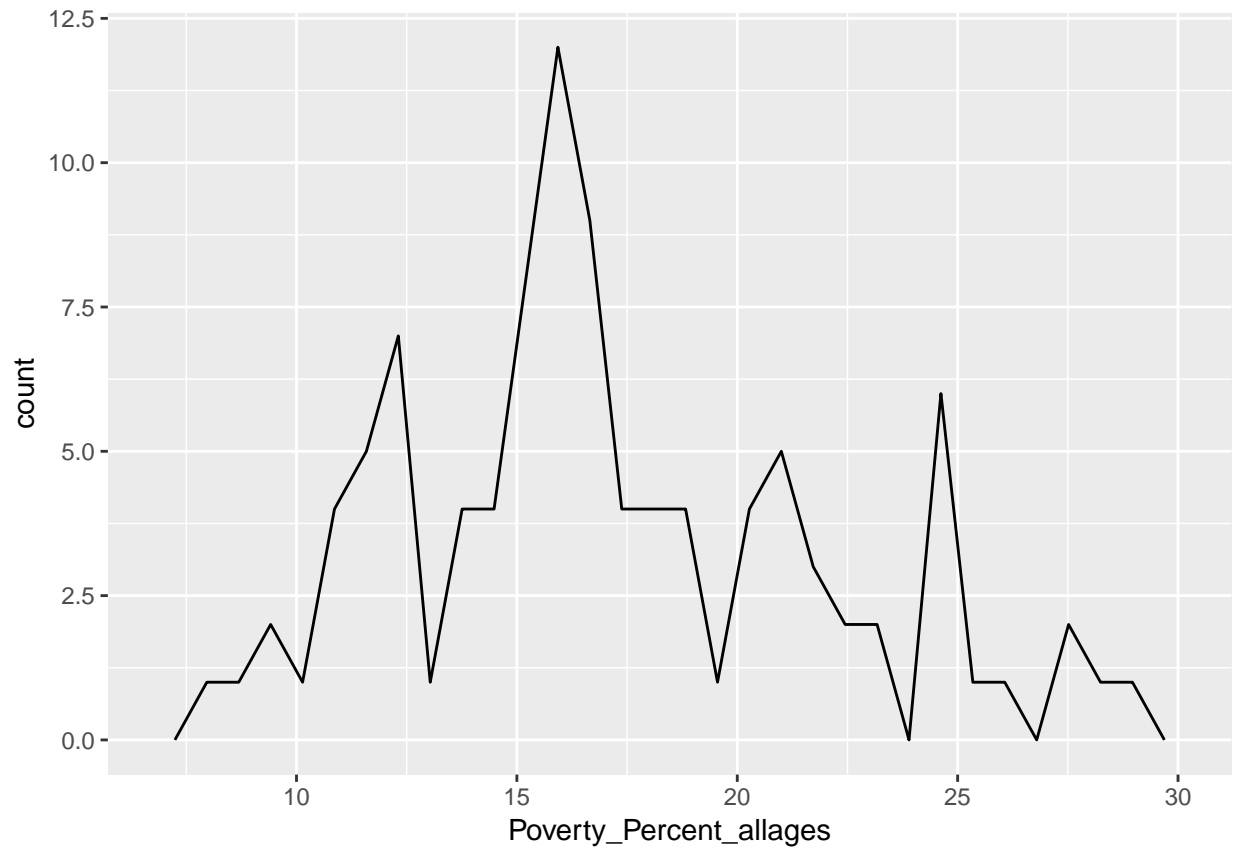
```



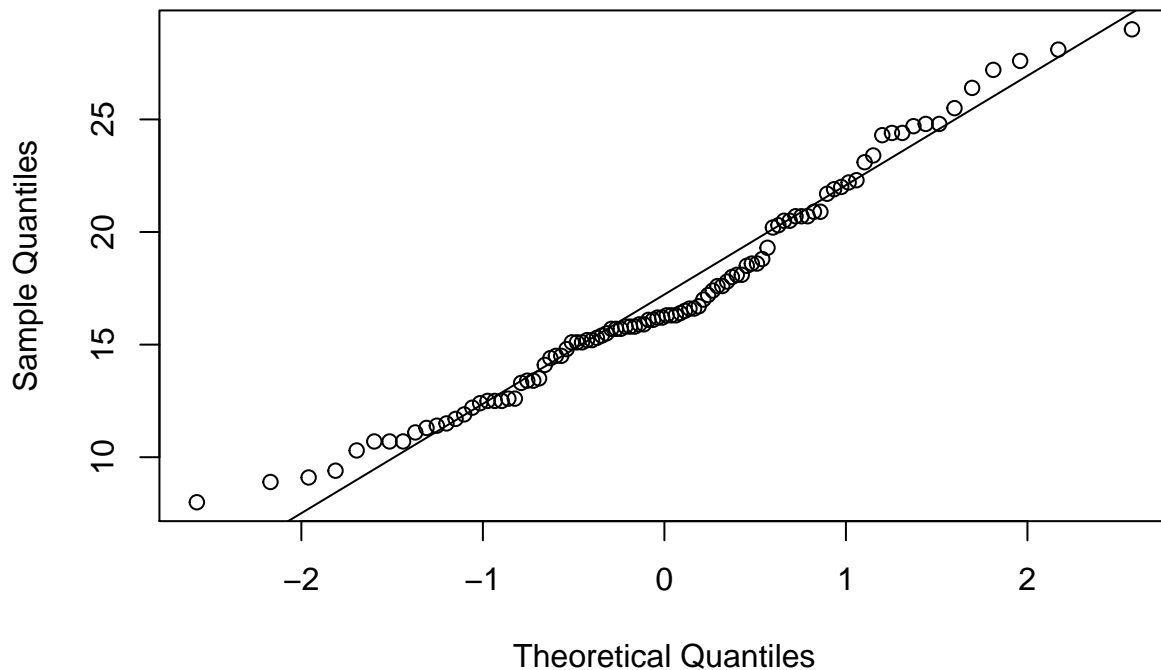
Normal Q-Q Plot



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Normal Q-Q Plot



```
#Read in Counties shapefile into an sf dataframe,  
#filtering for just NC counties
```

```
NC_Counties_shp <- st_read(dsn = "./Data/Spatial/NC_Counties.shp")
```

```
## Reading layer `NC_Counties' from data source `/Users/Tristen/OneDrive - Duke University/OneDrive/Projects/Geospatial/NC_Counties.shp'
## Simple feature collection with 100 features and 9 fields
## geometry type:  POLYGON
## dimension:      XY
## bbox:           xmin: -84.32162 ymin: 33.83437 xmax: -75.45998 ymax: 36.58841
## epsg (SRID):    4326
## proj4string:     +proj=longlat +datum=WGS84 +no_defs
```

```
#Geospatial data for NC Counties
```

```
Landfills_shp <- st_read(dsn = "./Data/Spatial/ActivePermittedLandfills.shp")
```

```
## Reading layer `ActivePermittedLandfills' from data source `/Users/Tristen/OneDrive - Duke University/OneDrive/Projects/Geospatial/ActivePermittedLandfills.shp'
## Simple feature collection with 177 features and 9 fields
## geometry type:  POINT
## dimension:      XY
## bbox:           xmin: -1.797693e+308 ymin: -1.797693e+308 xmax: -75.49015 ymax: 36.53
## epsg (SRID):    4326
## proj4string:     +proj=longlat +datum=WGS84 +no_defs
```

#Active Landfills

```
IH_shp <- st_read(dsn = "./Data/Spatial/IH_Sites.shp")
```

```
## Reading layer `IH_Sites' from data source `/Users/Tristen/OneDrive - Duke University/`  
## Simple feature collection with 1917 features and 13 fields  
## geometry type: POINT  
## dimension: XY  
## bbox: xmin: -83.96576 ymin: 33.9127 xmax: -75.52278 ymax: 36.54823  
## epsg (SRID): 4326  
## proj4string: +proj=longlat +datum=WGS84 +no_defs
```

#Hazardous substance spill and disposal sites

```
FRB_shp <- st_read(dsn = "./Data/Spatial/FRB_Sites.shp")
```

```
## Reading layer `FRB_Sites' from data source `/Users/Tristen/OneDrive - Duke University/`  
## Simple feature collection with 70 features and 24 fields  
## geometry type: POINT  
## dimension: XY  
## bbox: xmin: -9246629 ymin: 4057672 xmax: -8407034 ymax: 4358145  
## epsg (SRID): 3857  
## proj4string: +proj=merc +a=6378137 +b=6378137 +lat_ts=0.0 +lon_0=0.0 +x_0=0.0 +y_0=0.0
```

#Superfund

```
BF_shp <- st_read(dsn = "./Data/Spatial/BF_Sites.shp")
```

```
## Reading layer `BF_Sites' from data source `/Users/Tristen/OneDrive - Duke University/`  
## Simple feature collection with 398 features and 9 fields  
## geometry type: POINT  
## dimension: XY  
## bbox: xmin: 693972.8 ymin: 174551.3 xmax: 2827425 ymax: 1024565  
## epsg (SRID): NA  
## proj4string: +proj=lcc +lat_1=34.33333333333334 +lat_2=36.16666666666666 +lat_0=33.33333333333333
```

#Brownfields

```
RUST_shp <- st_read(dsn = "./Data/Spatial/RUST.shp")
```

```
## Reading layer `RUST' from data source `/Users/Tristen/OneDrive - Duke University/Springs/`  
## Simple feature collection with 30212 features and 28 fields  
## geometry type: POINT  
## dimension: XY  
## bbox: xmin: -84.31466 ymin: 33.87549 xmax: -75.46581 ymax: 36.567  
## epsg (SRID): 4326  
## proj4string: +proj=longlat +datum=WGS84 +no_defs
```

#Underground Storage Tanks

```
HW_shp <- st_read(dsn = "./Data/Spatial/HW_Sites.shp")
```

```
## Reading layer `HW_Sites' from data source `/Users/Tristen/OneDrive - Duke University/`
```

```
## Simple feature collection with 2577 features and 21 fields
## geometry type:  POINT
## dimension:      XY
## bbox:           xmin: -84.02775 ymin: 33.89647 xmax: -75.60355 ymax: 36.53089
## epsg (SRID):    4326
## proj4string:     +proj=longlat +datum=WGS84 +no_defs
```

```
#Hazardous Waste Resource Conservation and Recovery Act
```

```
#Reveal the CRS of the counties features so they can be graphed
#with NC Counties shapefile
```

```
st_crs(NC_Counties_shp)
```

```
## Coordinate Reference System:
##   EPSG: 4326
##   proj4string: "+proj=longlat +datum=WGS84 +no_defs"
```

```
st_crs(Landfills_shp)
```

```
## Coordinate Reference System:
##   EPSG: 4326
##   proj4string: "+proj=longlat +datum=WGS84 +no_defs"
```

```
st_crs(IH_shp)
```

```
## Coordinate Reference System:
##   EPSG: 4326
##   proj4string: "+proj=longlat +datum=WGS84 +no_defs"
```

```
st_crs(FRB_shp)
```

```
## Coordinate Reference System:
##   EPSG: 3857
##   proj4string: "+proj=merc +a=6378137 +b=6378137 +lat_ts=0.0 +lon_0=0.0 +x_0=0.0 +y_0=0.0 +units=m +no_defs"
```

```
st_crs(BF_shp)
```

```
## Coordinate Reference System:
##   No EPSG code
##   proj4string: "+proj=lcc +lat_1=34.33333333333334 +lat_2=36.16666666666666 +lat_0=33.33333333333333 +lon_0=-96.33333333333333 +x_0=0.0 +y_0=0.0 +units=m +no_defs"
```

```
st_crs(RUST_shp)
```

```
## Coordinate Reference System:
##   EPSG: 4326
##   proj4string: "+proj=longlat +datum=WGS84 +no_defs"
```

```
st_crs(HW_shp)
```

```
## Coordinate Reference System:
```

```

## EPSG: 4326
## proj4string: "+proj=longlat +datum=WGS84 +no_defs"
#There is one row with an incorecct location;
#This site will be omitted in order to proceed with mapping data
Landfills_shp_mod <- subset(Landfills_shp, !LocationID == "P1252")

#Filter RUST dataset for only high risk UST sites since this inforation
#is available and due to the large amount of sites
levels(RUST_shp$ConfRisk)

## [1] "H" "I" "l" "L" "U"

highrisk_RUST <- RUST_shp %>%
  filter(ConfRisk == "H")

#Join poverty data to county geospatial data
county_poverty_join <- NC_Counties_shp %>%
  left_join(y = Poverty_NC, by = c("CO_NAME" = "Name"))

#Count the number of sites in each county; In some, the names of
#counties are not in all capital letters. In order to address
#future problems with joins, some were converted to all capital letters.
Sitecount_Landfill <- count(Landfills_shp_mod, Landfills_shp_mod$County)

levels(IH_shp$SITECOUNTY) <- toupper(levels(IH_shp$SITECOUNTY)) #all caps
Sitecount_IH <- count(IH_shp, IH_shp$SITECOUNTY)

Sitecount_FRB <- count(FRB_shp, FRB_shp$SITE_COUNT)

levels(BF_shp$BF_County) <- toupper(levels(BF_shp$BF_County))
Sitecount_BF <- count(BF_shp, BF_shp$BF_County)

levels(highrisk_RUST$County) <- toupper(levels(highrisk_RUST$County))
Sitecount_RUST <- count(highrisk_RUST, highrisk_RUST$County)

Sitecount_HW <- count(HW_shp, HW_shp$LOC_COUNTY)

#Join count dataframes with race info. Needed to make county names in
#Race_Processed data set capitalized so left join would work.
#In one case, Race_Processed data needed to add a column with
#abbreviations in order to join to the RUST data set.
Sitecount_Landfill_join <- left_join(Sitecount_Landfill, Race_Processed,
  by = c("Landfills_shp_mod$County" = "County"))

```

```
## Warning: Column `Landfills_shp_mod$County`/`County` joining factors with
## different levels, coercing to character vector
```

```
levels(Race_Processed$County) <- toupper(levels(Race_Processed$County))
Sitecount_IH_join <- left_join(Sitecount_IH, Race_Processed,
  by = c("IH_shp$SITECOUNT" = "County"))
```

```
## Warning: Column `IH_shp$SITECOUNT`/`County` joining factors with different
## levels, coercing to character vector
```

```
Sitecount_FRB_join <- left_join(Sitecount_FRB, Race_Processed,
  by = c("FRB_shp$SITE_COUNT" = "County"))
```

```
## Warning: Column `FRB_shp$SITE_COUNT`/`County` joining factors with
## different levels, coercing to character vector
```

```
Sitecount_BF_join <- left_join(Sitecount_BF, Race_Processed,
  by = c("BF_shp$BF_County" = "County"))
```

```
## Warning: Column `BF_shp$BF_County`/`County` joining factors with different
## levels, coercing to character vector
```

```
Race_Processed_abv <- transform(Race_Processed,
  ABV = str_sub(Race_Processed$County, 1, 5))
```

```
Sitecount_RUST_join <- left_join(Sitecount_RUST, Race_Processed_abv, by = c("highrisk_RU
```

```
## Warning: Column `highrisk_RUST$County`/`ABV` joining factors with different
## levels, coercing to character vector
```

```
Sitecount_HW_join <- left_join(Sitecount_HW, Race_Processed, by = c("HW_shp$LOC_COUNTY"
```

```
## Warning: Column `HW_shp$LOC_COUNTY`/`County` joining factors with different
## levels, coercing to character vector
```

4 Analysis

```
#Landfill
#Create clusters for low or high minority percent in each county
test <- pam(x = Sitecount_Landfill_join$MinorityPerc, k = 2, metric="manhattan")

PAMClust = rep("NA", length(Sitecount_Landfill_join$MinorityPerc))
PAMClust[test$clustering == 1] = "High"
PAMClust[test$clustering == 2] = "Low"
Sitecount_Landfill_join$Cluster = PAMClust

LF.count <- count(Sitecount_Landfill_join, Sitecount_Landfill_join$Cluster)
wilcox.test(Sitecount_Landfill_join$n ~ Sitecount_Landfill_join$Cluster)

## Warning in wilcox.test.default(x = c(4L, 1L, 1L, 1L, 2L, 4L, 8L, 1L, 3L, :
## cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: Sitecount_Landfill_join$n by Sitecount_Landfill_join$Cluster
## W = 691.5, p-value = 0.3187
## alternative hypothesis: true location shift is not equal to 0

#P-value = 0.3187

#IH
test1 <- pam(x = Sitecount_IH_join$MinorityPerc, k = 2, metric="manhattan")
PAMClust1 = rep("NA", length(Sitecount_IH_join$MinorityPerc))
PAMClust1[test1$clustering == 1] = "High"
PAMClust1[test1$clustering == 2] = "Low"
Sitecount_IH_join$Cluster = PAMClust1

IH.count <- count(Sitecount_IH_join, Sitecount_IH_join$Cluster)
wilcox.test(Sitecount_IH_join$n ~ Sitecount_IH_join$Cluster)

##
## Wilcoxon rank sum test with continuity correction
##
## data: Sitecount_IH_join$n by Sitecount_IH_join$Cluster
## W = 1232.5, p-value = 0.6696
## alternative hypothesis: true location shift is not equal to 0

#P-value = 0.6696

#FRB -error
```

```

#Sitecount_FRB_join$MinorityPerc = na.omit(Sitecount_FRB_join$MinorityPerc)
#test2 <- pam(x = FRB_x, k = 2, metric="manhattan")
#PAMClust2 = rep("NA", length(Sitecount_FRB_join$MinorityPerc))
#PAMClust2[test2$clustering == 1] = "High"
#PAMClust2[test2$clustering == 2] = "Low"
#Sitecount_FRB_join$Cluster = PAMClust2

#FRB.count <- count(Sitecount_FRB_join, Sitecount_FRB_join$Cluster)
#wilcox.test(Sitecount_IH_join$n ~ Sitecount_FRB_join$Cluster)

#BF
test3 <- pam(x = Sitecount_BF_join$MinorityPerc, k = 2, metric="manhattan")
PAMClust3 = rep("NA", length(Sitecount_BF_join$MinorityPerc))
PAMClust3[test3$clustering == 1] = "High"
PAMClust3[test3$clustering == 2] = "Low"
Sitecount_BF_join$Cluster = PAMClust3

BF.count <- count(Sitecount_BF_join, Sitecount_BF_join$Cluster)
wilcox.test(Sitecount_BF_join$n ~ Sitecount_BF_join$Cluster)

## Warning in wilcox.test.default(x = c(3L, 1L, 2L, 2L, 2L, 3L, 16L, 2L,
## 21L, : cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: Sitecount_BF_join$n by Sitecount_BF_join$Cluster
## W = 461, p-value = 0.6803
## alternative hypothesis: true location shift is not equal to 0

#P-value = 0.6803

#RUST
test4 <- pam(x = Sitecount_RUST_join$MinorityPerc, k = 2, metric="manhattan")
PAMClust4 = rep("NA", length(Sitecount_RUST_join$MinorityPerc))
PAMClust4[test4$clustering == 1] = "High"
PAMClust4[test4$clustering == 2] = "Low"
Sitecount_RUST_join$Cluster = PAMClust4

RUST.count <- count(Sitecount_RUST_join, Sitecount_RUST_join$Cluster)
wilcox.test(Sitecount_RUST_join$n ~ Sitecount_RUST_join$Cluster)

##
## Wilcoxon rank sum test with continuity correction
##
## data: Sitecount_RUST_join$n by Sitecount_RUST_join$Cluster

```

```
## W = 1221.5, p-value = 0.7398
## alternative hypothesis: true location shift is not equal to 0
#P-value = 0.7398

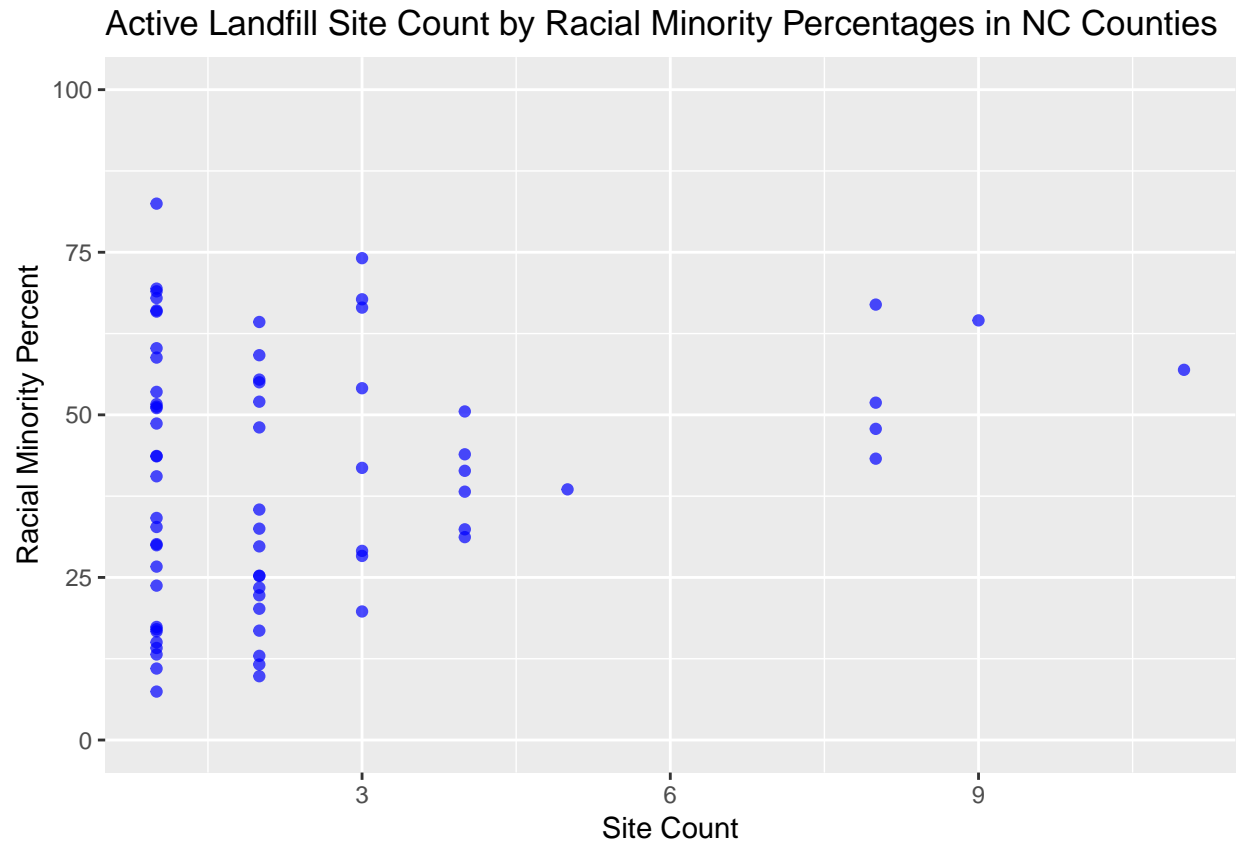
#HW
test5 <- pam(x = Sitecount_HW_join$MinorityPerc, k = 2, metric="manhattan")
PAMClust5 = rep("NA", length(Sitecount_HW_join$MinorityPerc))
PAMClust5[test5$clustering == 1] = "High"
PAMClust5[test5$clustering == 2] = "Low"
Sitecount_HW_join$Cluster = PAMClust5

HW.count <- count(Sitecount_HW_join, Sitecount_HW_join$Cluster)
wilcox.test(Sitecount_HW_join$n ~ Sitecount_HW_join$Cluster)

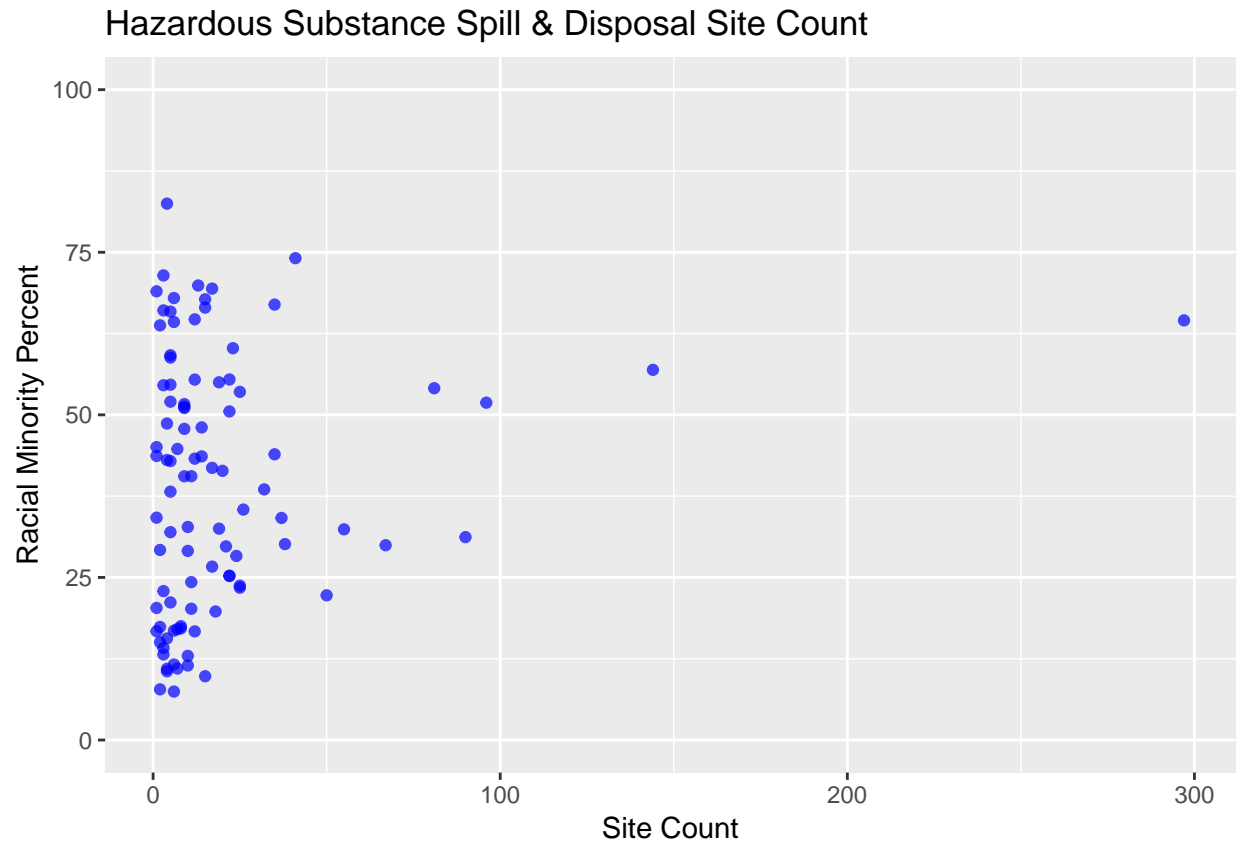
## Warning in wilcox.test.default(x = c(46L, 5L, 11L, 3L, 9L, 45L, 12L, 3L, :
## cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: Sitecount_HW_join$n by Sitecount_HW_join$Cluster
## W = 1137.5, p-value = 0.4177
## alternative hypothesis: true location shift is not equal to 0
#P-value = 0.4177

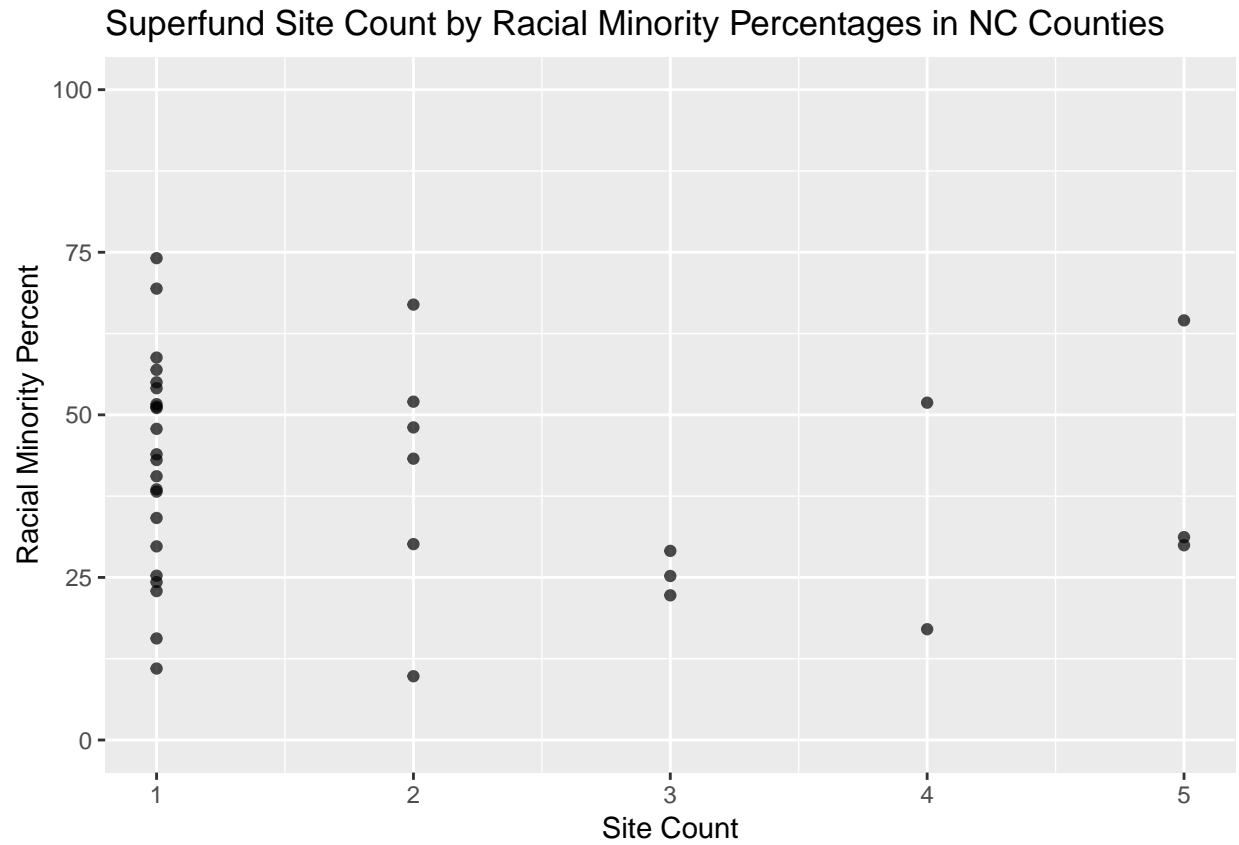
## Warning: Removed 1 rows containing missing values (geom_point).
```

Warning: Removed 1 rows containing missing values (geom_point).

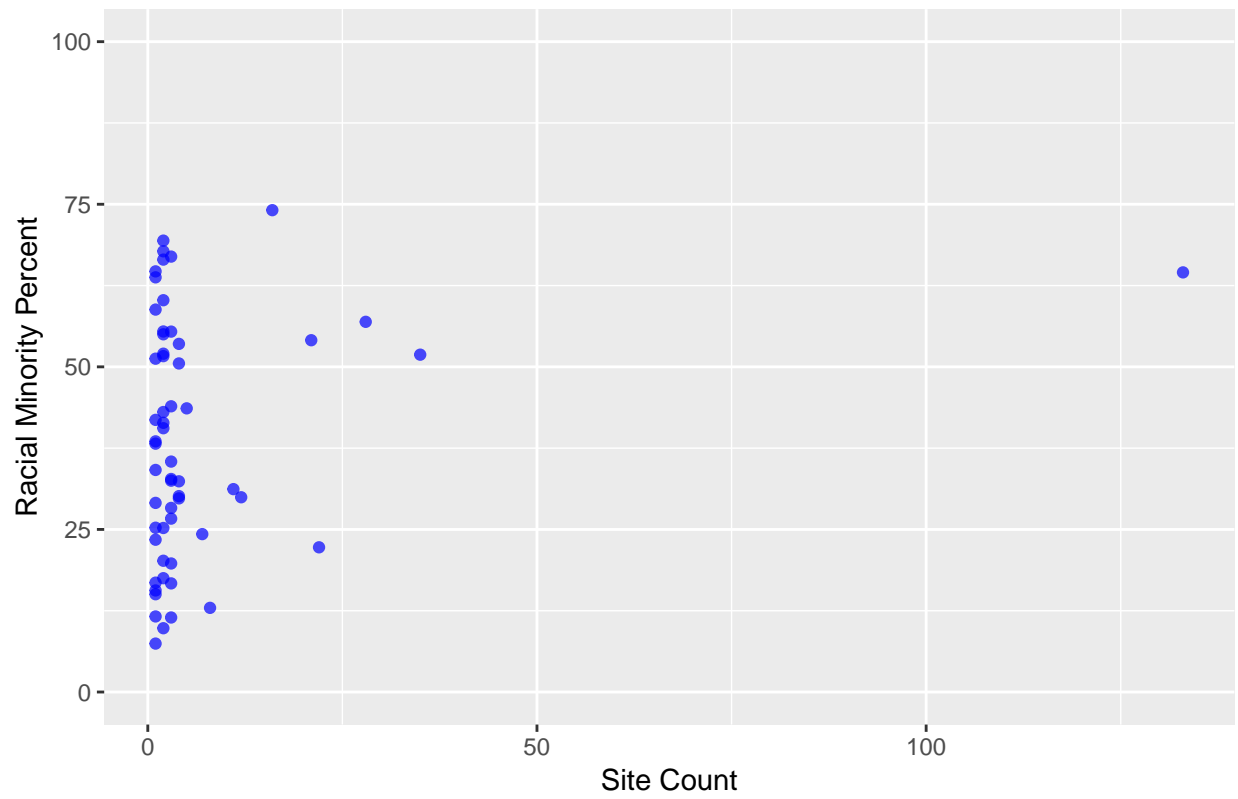


Warning: Removed 4 rows containing missing values (geom_point).

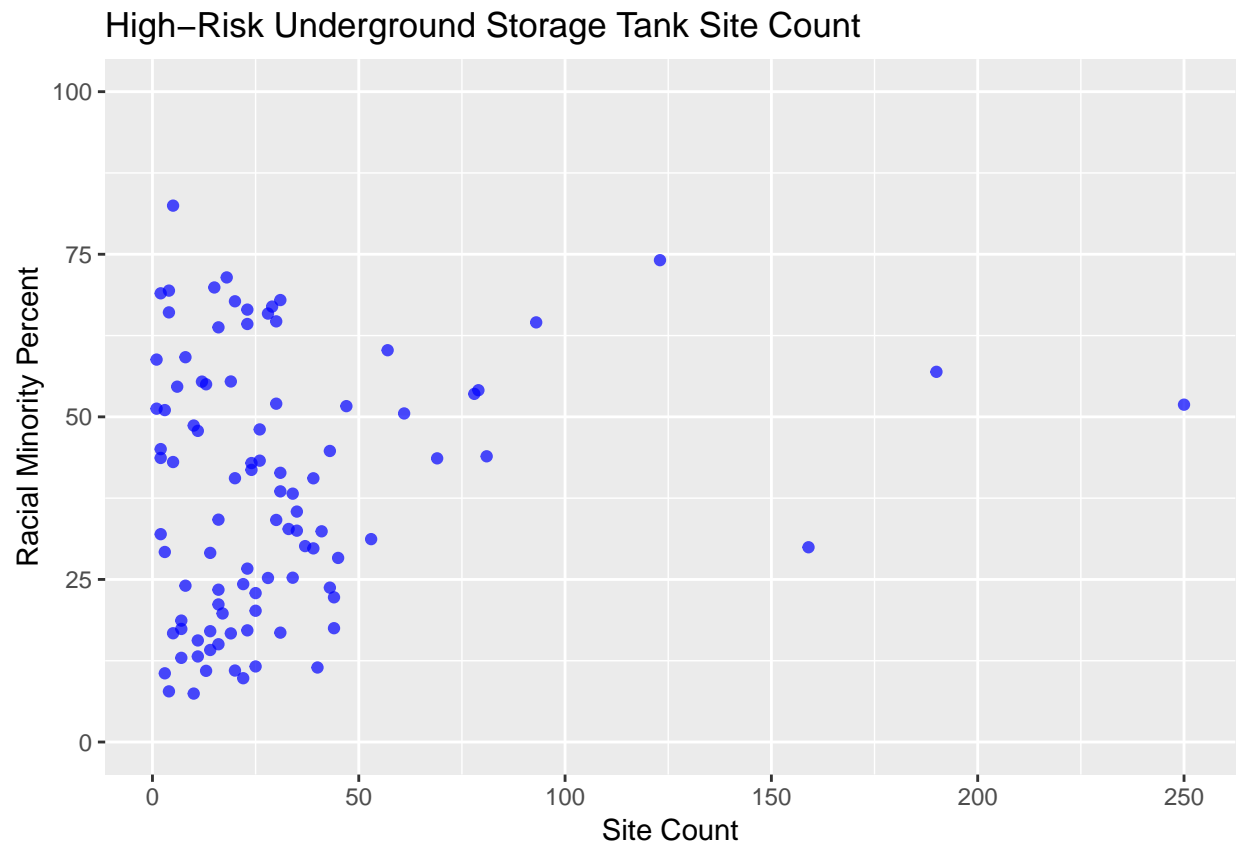


Warning: Removed 1 rows containing missing values (geom_point).

Brownfield Site Count by Racial Minority Percentages in NC Counties

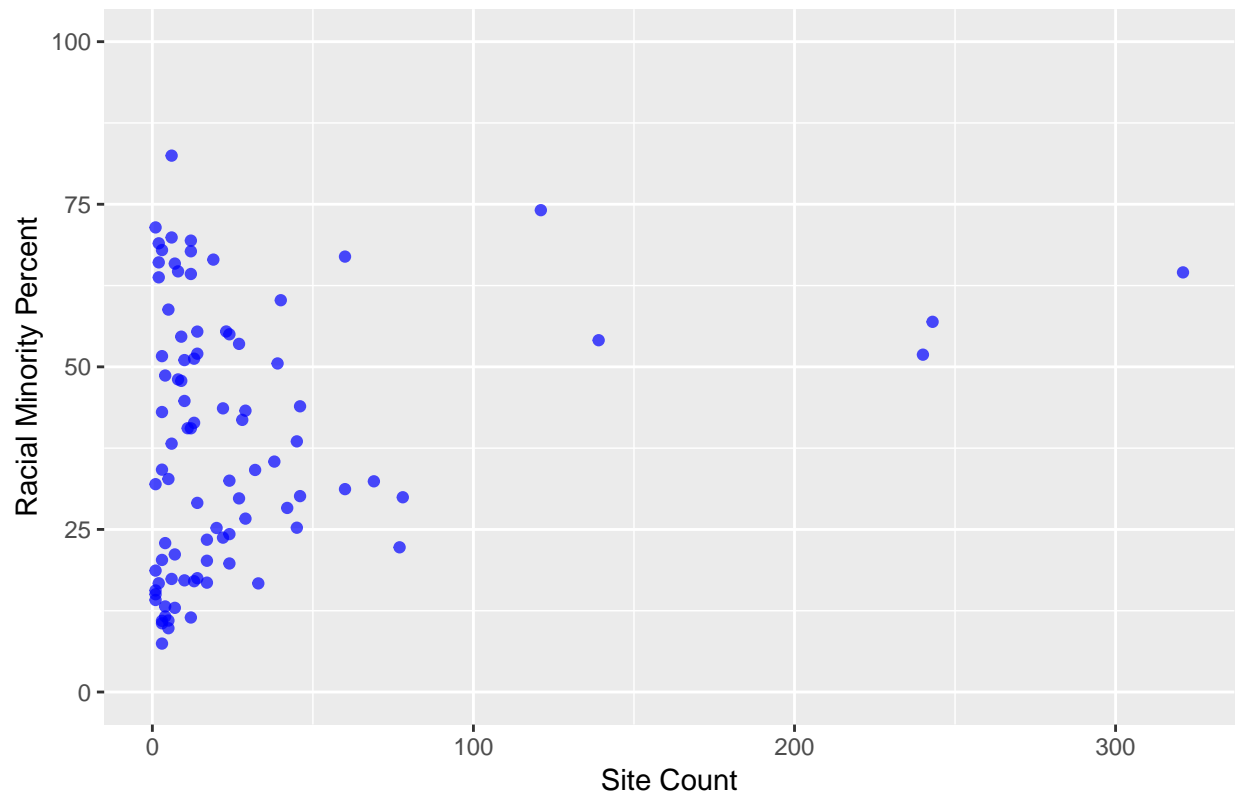


Warning: Removed 1 rows containing missing values (geom_point).

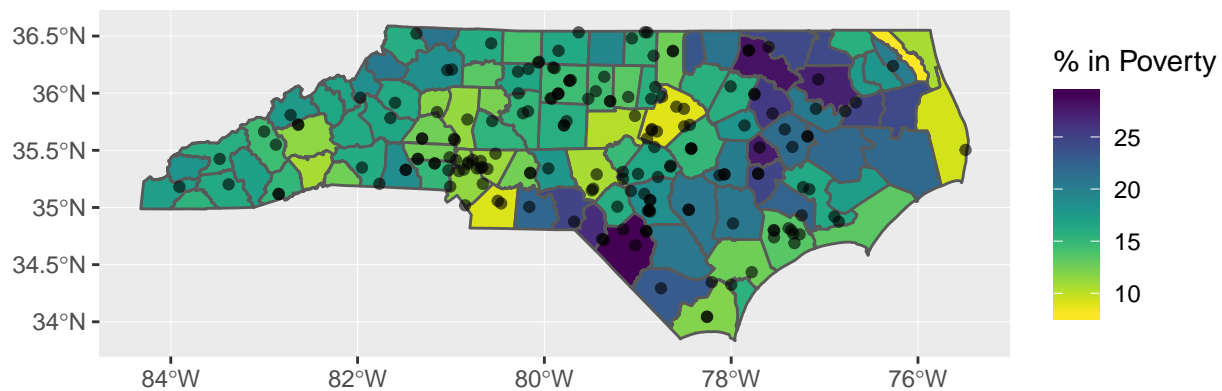


Warning: Removed 1 rows containing missing values (geom_point).

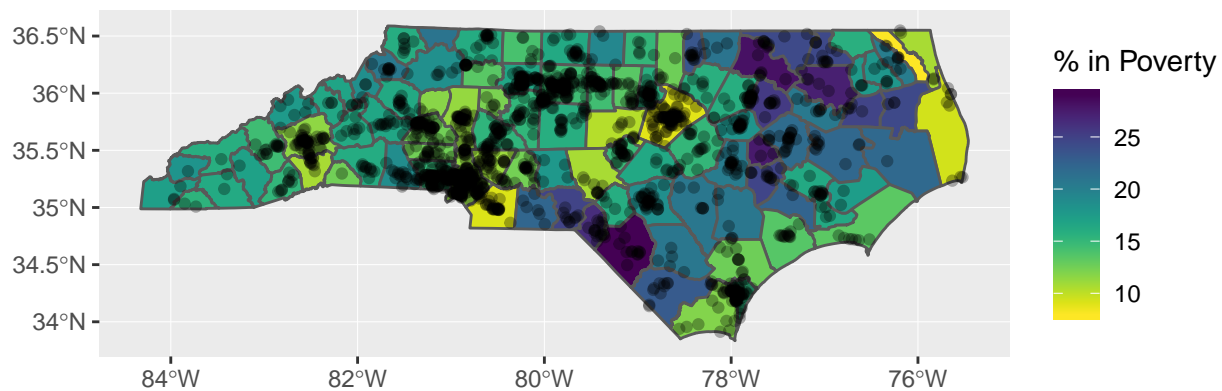
Hazardous Waste Site Count by Racial Minority % in NC Counties



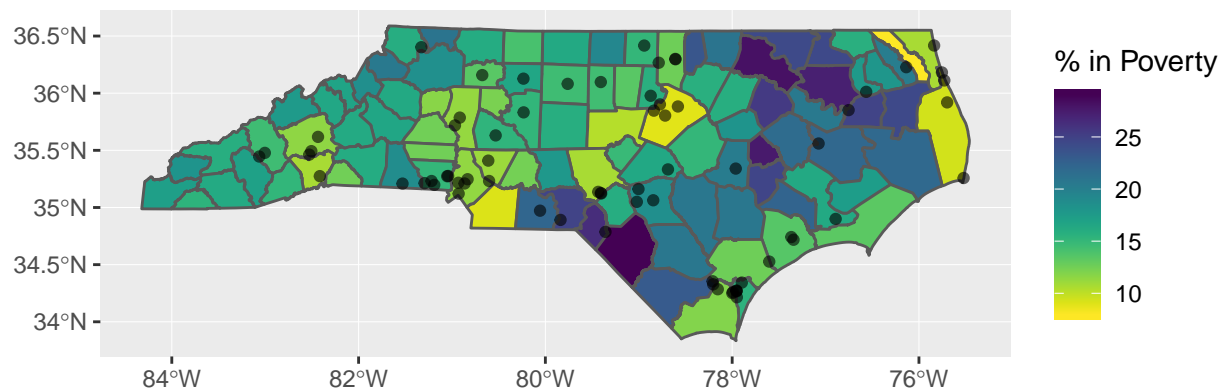
Landfill Sites in North Carolina



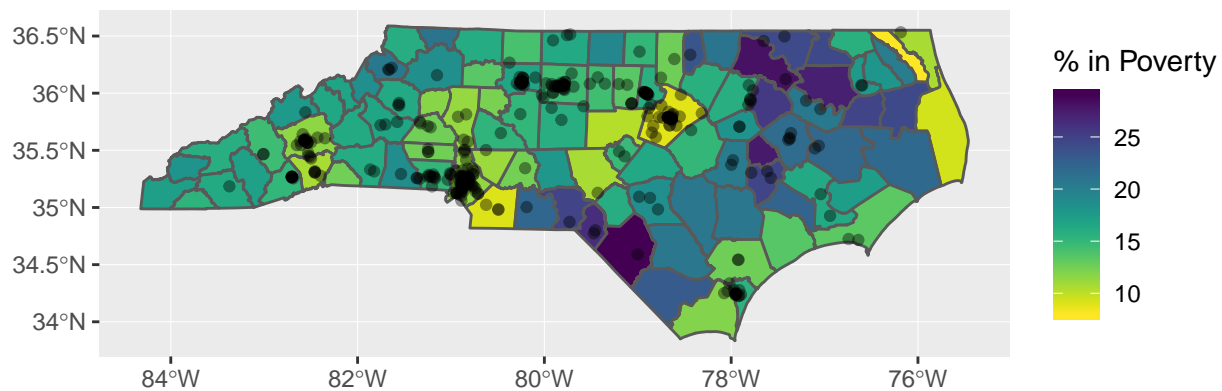
IH Sites in North Carolina



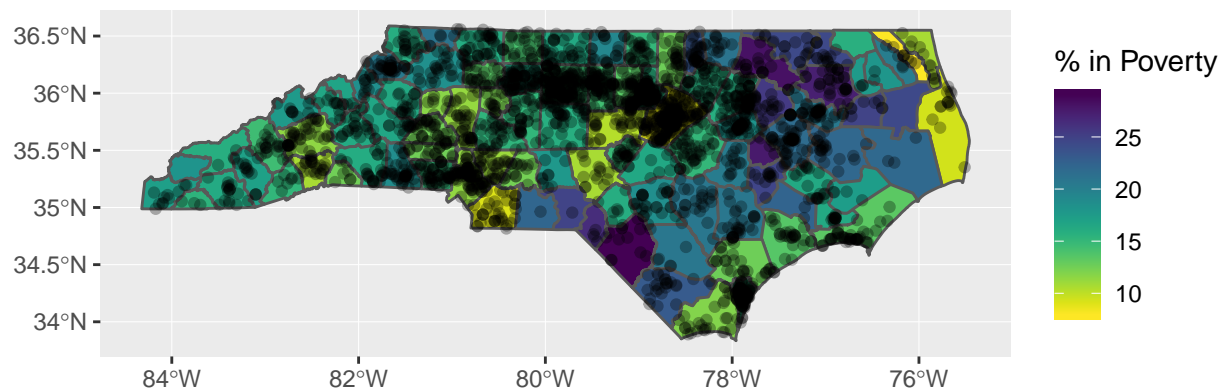
Superfund Sites in North Carolina



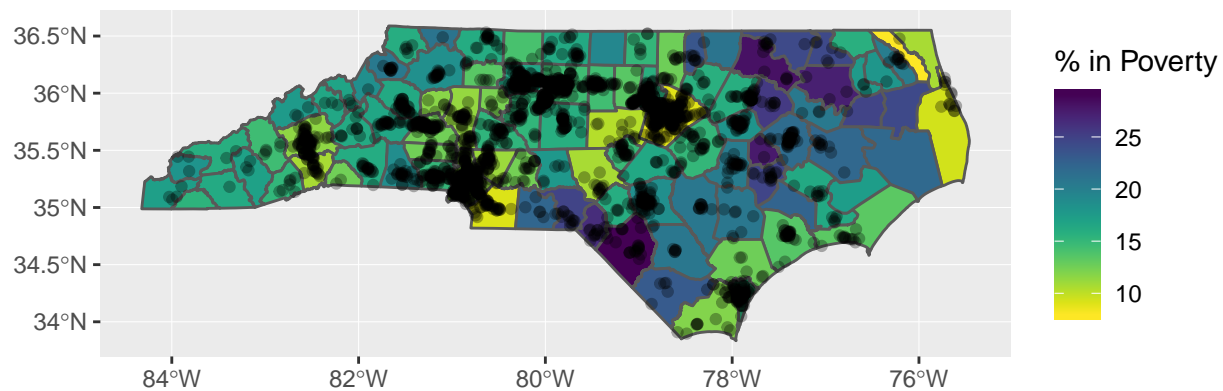
Brownfield Sites in North Carolina



RUST Sites in North Carolina



Hazardous Waste Sites in North Carolina



5 Summary and Conclusions

Based on my performed analysis, there is no significant difference in the number of hazardous waste sites that exist in counties with majority racial minority populations.

6 References

U.S. Census Bureau. (2018). Small Area Income and Poverty Estimates (SAIPE) Program. Retrieved from <https://www.census.gov/programs-surveys/saipe/about.html>

NC Budget and Management. (2019). LINC. Retrieved from <https://www.osbm.nc.gov/facts-figures/linc>