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, as have which areas have been experiencing increased flooding and extreme precipitation events in the past decade. 92 words

Contents

List of Tables

List of Figures

<Note: set up autoreferencing for figures and tables in your document>

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.

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. Additionally, I have downloaded data from the U.S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) Program and (U.S. Census Bureau, 2018).

2 Dataset Information

County White BlackNative Asian Hispanic Other Haw_PacIs Two Total.Pop —— —— —— —— —— ——— —— ——— —— ———

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", "

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 +

#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

#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
```

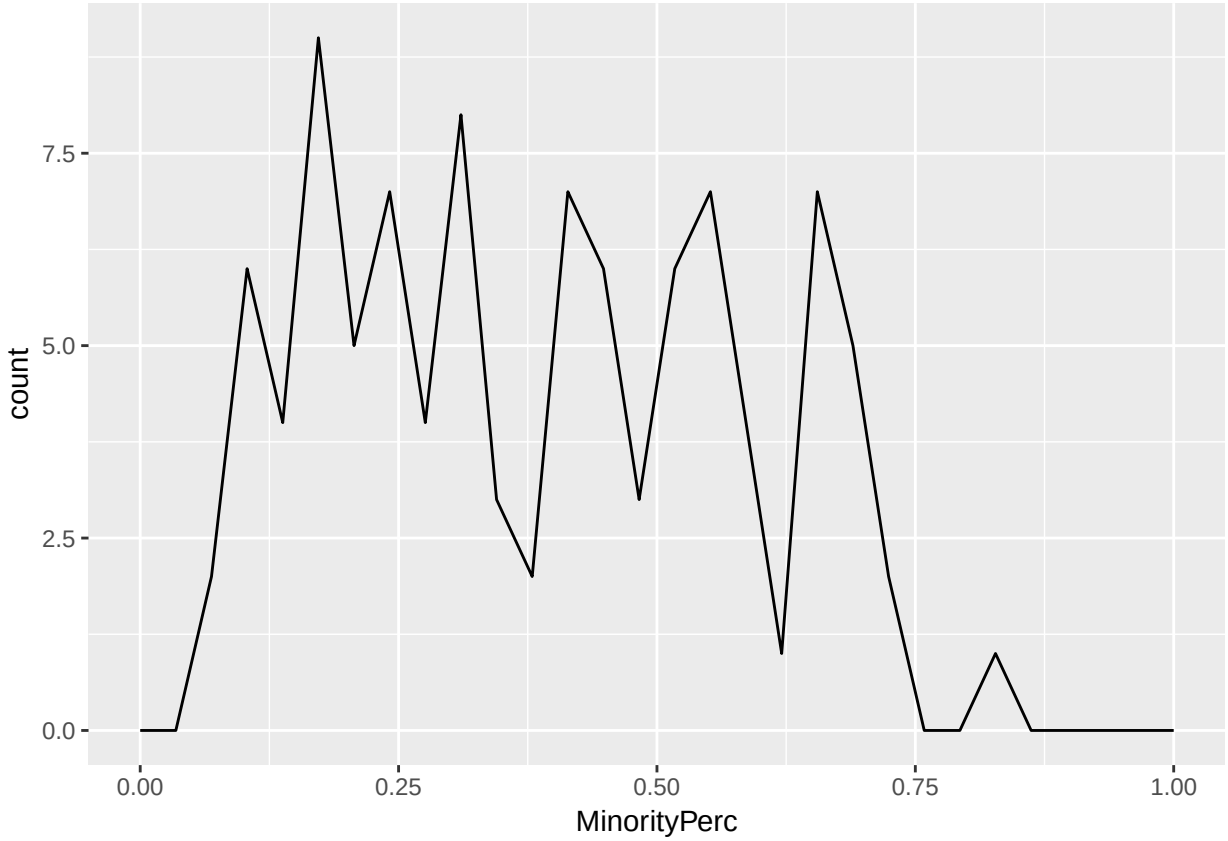
```
#From the output, the p-value is < 0.05 implying that the distribution of the data are
```

```
ggplot(Race_Processed, aes(x = MinorityPerc)) +
  geom_freqpoly() +
  scale_x_continuous(limits = c(0, 1))
```

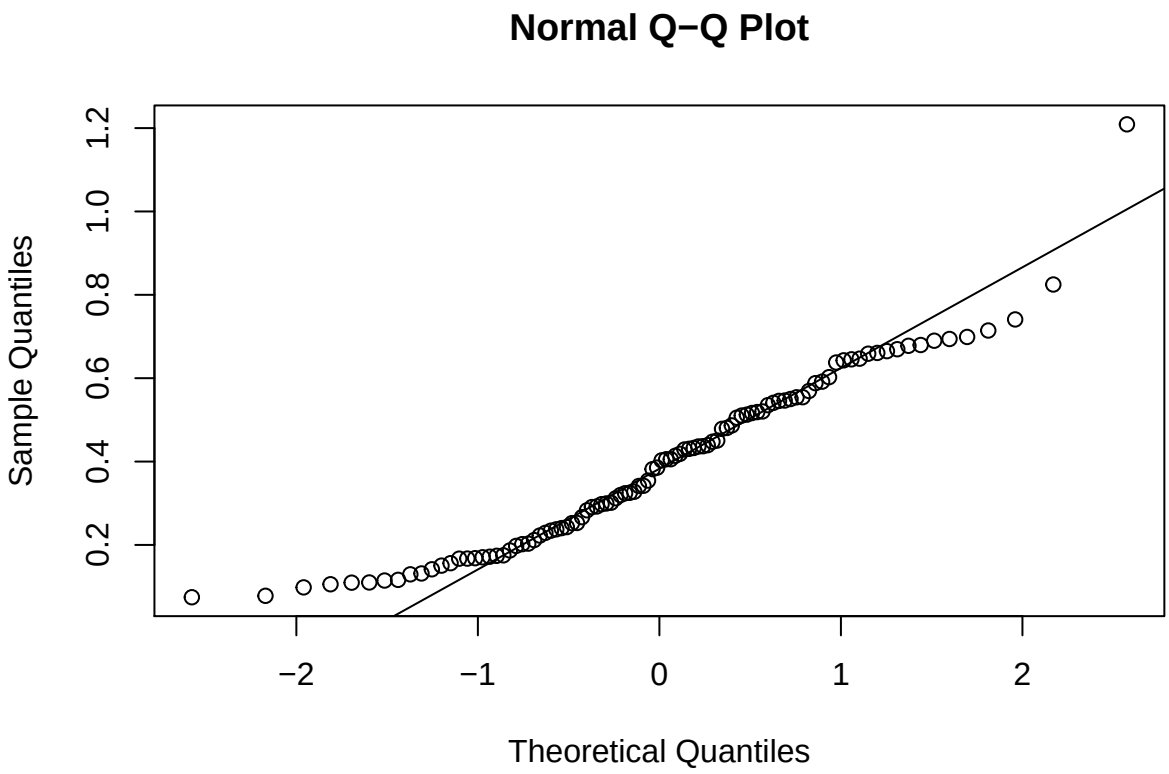
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1164 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 2 rows containing missing values (geom_path).
```

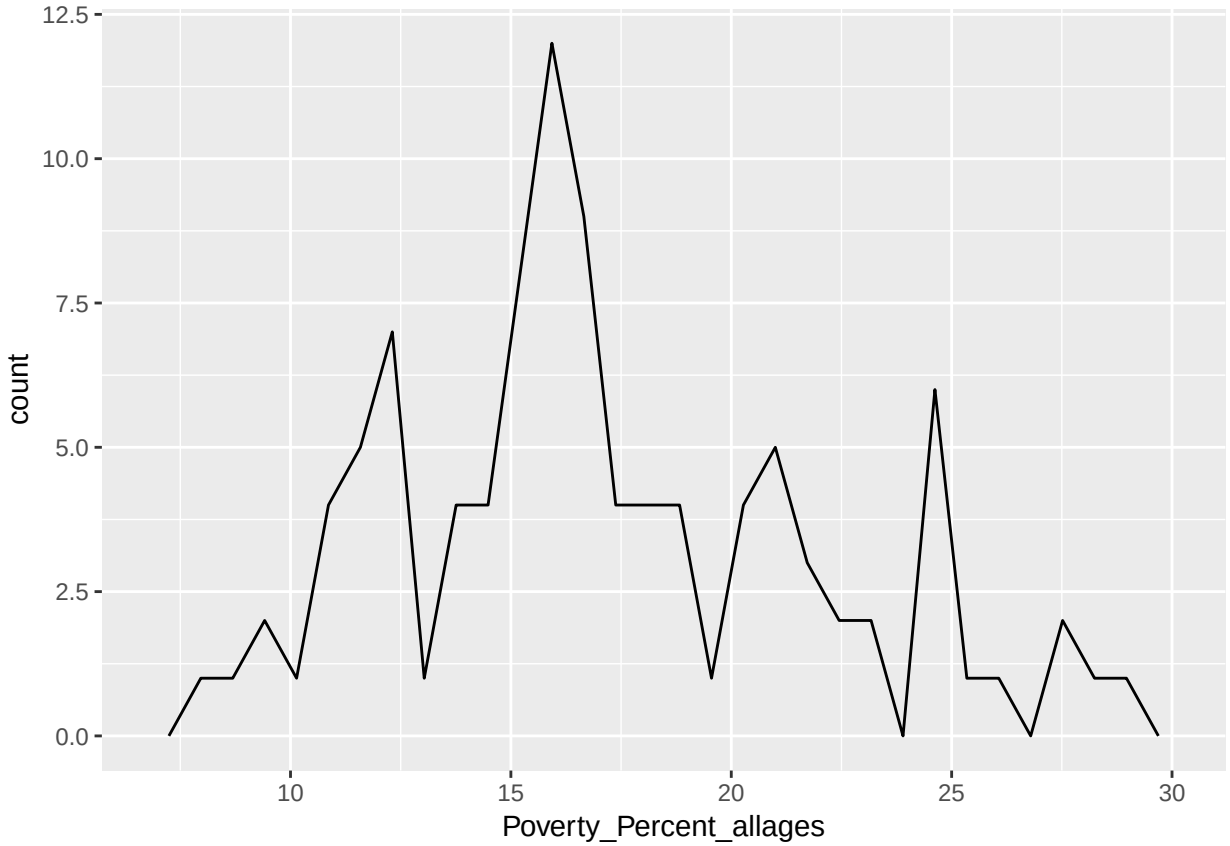



```
qqnorm(Race_Processed$MinorityPerc); qqline(Race_Processed$MinorityPerc)
```

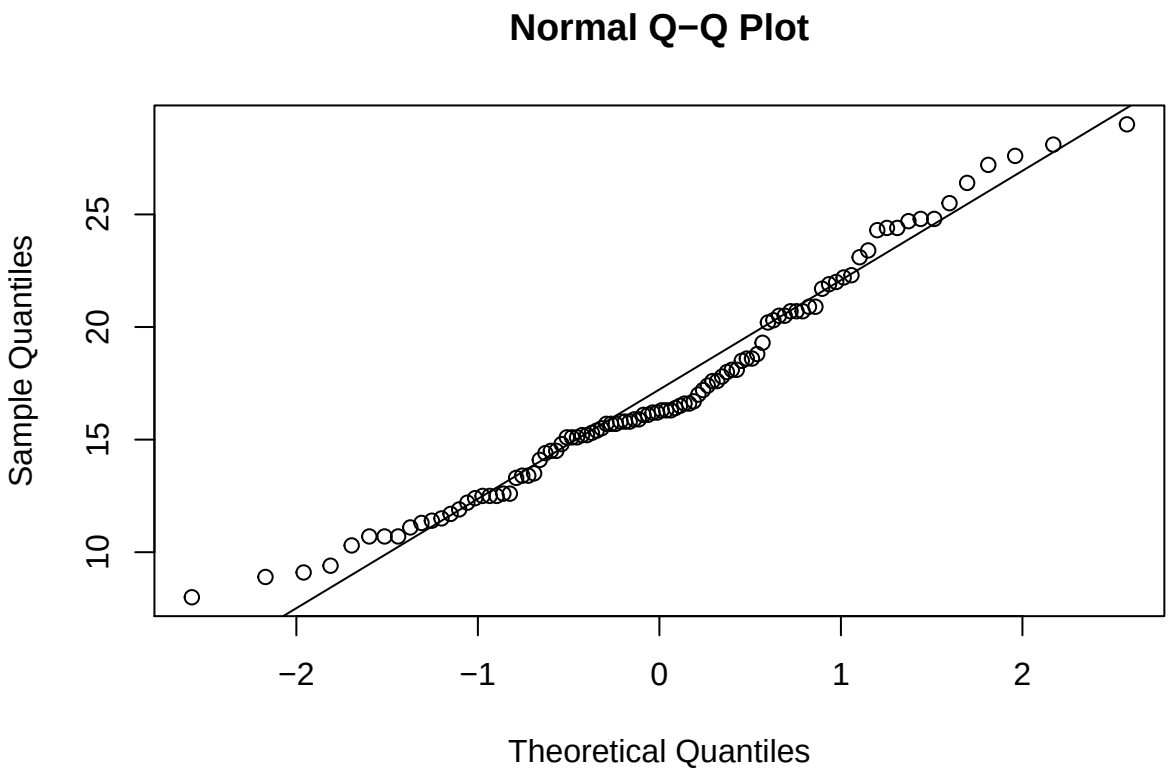



```
ggplot(Poverty_NC, aes(x = Poverty_Percent_allages)) +
  geom_freqpoly()
```

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




```
qqnorm(Poverty_NC$Poverty_Percent_allages); qqline(Poverty_NC$Poverty_Percent_allages)
```



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

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

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

```
## Reading layer `NC_Counties' from data source `/Users/Tristen/OneDrive - Duke Universi
## 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
```

```
Landfills_shp <- st_read(dsn = "./Data/Spatial/ActivePermittedLandiflls.shp") #Active La
```

```
## Reading layer `ActivePermittedLandiflls' from data source `/Users/Tristen/OneDrive -
## 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
```

```
IH_shp <- st_read(dsn = "./Data/Spatial/IH_Sites.shp") #Hazardous substance spill and d
```

```
## 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
```

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

```
## 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
```

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

```
## 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
```

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

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

```
HW_shp <- st_read(dsn = "./Data/Spatial/HW_Sites.shp") #Hazardous Waste Resource Conser
```

```
## 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
```

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

```
st_crs(BF_shp)
```

```
## Coordinate Reference System:
##   No EPSG code
##   proj4string: "+proj=lcc +lat_1=34.33333333333334 +lat_2=36.16666666666666 +lat_0=33
```

```
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"
```

```r
#There is one row with an incorecct location; This site will be omitted in order to pr
Landfills_shp_mod <- subset(Landfills_shp, !LocationID == "P1252")

#Filter RUST dataset for only high risk UST sites since this inforation is available a
levels(RUST_shp$ConfRisk)
```

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

```r
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 a
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 d
Sitecount_Landfill_join <- left_join(Sitecount_Landfill, Race_Processed, by = c("Landfil
```

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

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

```
## Warning: Column `IH_shp$SITECOUNTY`/`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" =
```

```
## 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,

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

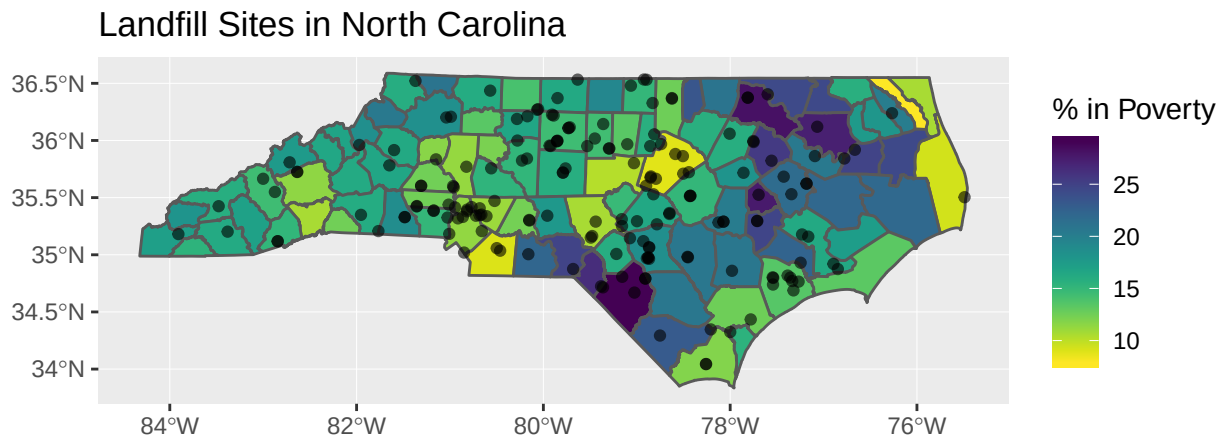
```
## Warning: Column `highrisk_RUST$County`/`ABV` joining factors with different
## levels, coercing to character vector
```

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

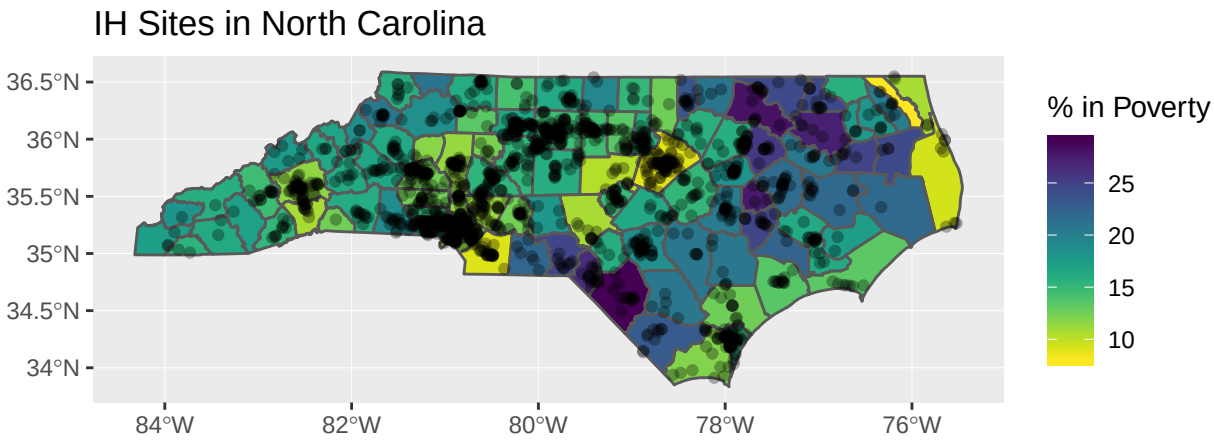
```
## Warning: Column `HW_shp$LOC_COUNTY`/`County` joining factors with different
## levels, coercing to character vector
```

5 Analysis

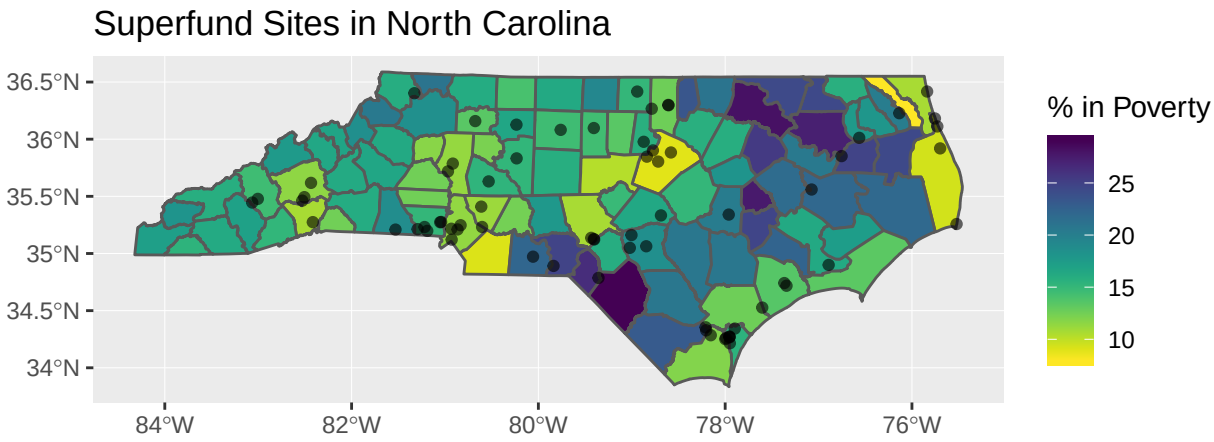
```
#Creating maps from basic data with counties
ggplot() +
  geom_sf(data = county_poverty_join, aes(fill=county_poverty_join$Poverty_Percent_allag
  scale_fill_viridis(direction = -1) +
  geom_sf(data = Landfills_shp_mod, alpha = 0.6) +
 labs(fill = "% in Poverty") +
  ggtitle("Landfill Sites in North Carolina")
```



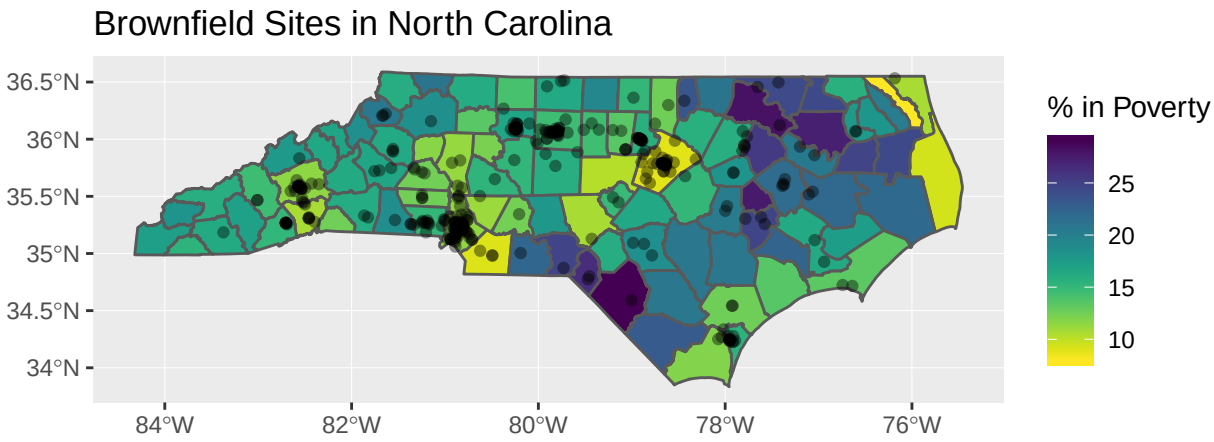

```
ggplot() +
  geom_sf(data = county_poverty_join, aes(fill=county_poverty_join$Poverty_Percent_allag
  scale_fill_viridis(direction = -1) +
  geom_sf(data = IH_shp, alpha = 0.3, color = "black") +
 labs(fill = "% in Poverty") +
  ggtitle("IH Sites in North Carolina")
```



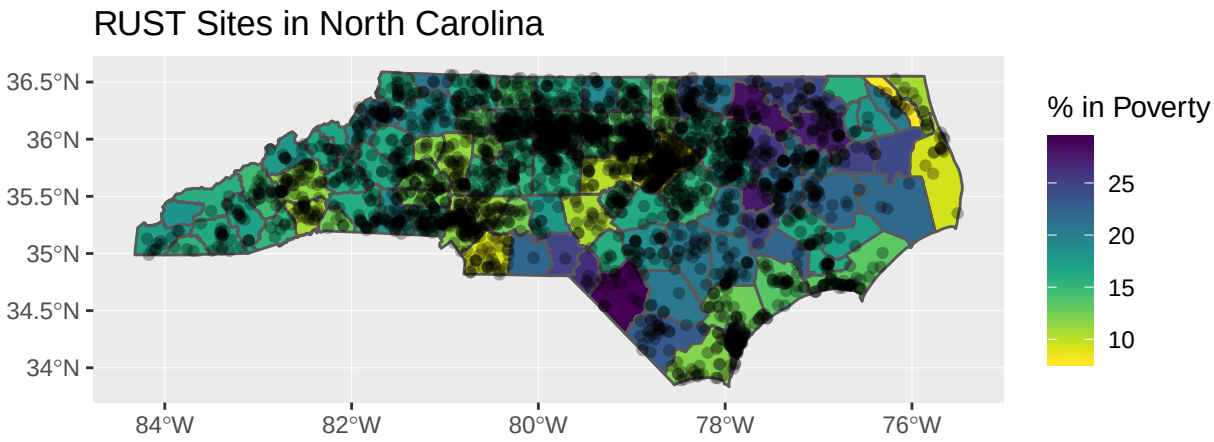

```
ggplot() +
  geom_sf(data = county_poverty_join, aes(fill=county_poverty_join$Poverty_Percent_allag
  scale_fill_viridis(direction = -1) +
  geom_sf(data = FRB_shp, alpha = 0.6, color = "black") +
 labs(fill = "% in Poverty") +
  ggtitle("Superfund Sites in North Carolina")
```



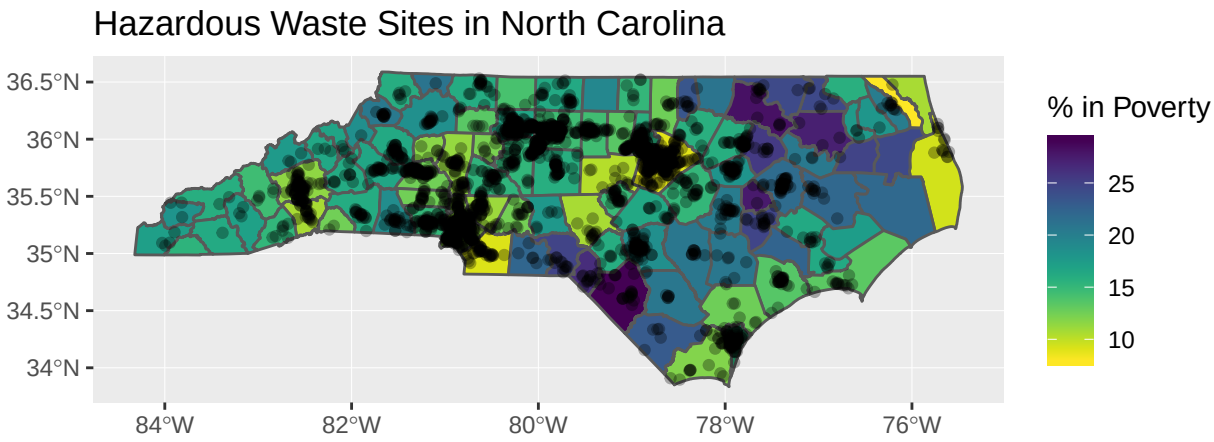

```
ggplot() +
  geom_sf(data = county_poverty_join, aes(fill=county_poverty_join$Poverty_Percent_allag
  scale_fill_viridis(direction = -1) +
  geom_sf(data = BF_shp, alpha = 0.4) +
 labs(fill = "% in Poverty") +
  ggtitle("Brownfield Sites in North Carolina")
```




```
ggplot() +
  geom_sf(data = county_poverty_join, aes(fill=county_poverty_join$Poverty_Percent_allag
  scale_fill_viridis(direction = -1) +
  geom_sf(data = highrisk_RUST, alpha = 0.3, color = "black") +
 labs(fill = "% in Poverty") +
  ggtitle("RUST Sites in North Carolina")
```




```
ggplot() +
  geom_sf(data = county_poverty_join, aes(fill=county_poverty_join$Poverty_Percent_allag
  scale_fill_viridis(direction = -1) +
  geom_sf(data = HW_shp, alpha = 0.3, color = "black") +
 labs(fill = "% in Poverty") +
  ggtitle("Hazardous Waste Sites in North Carolina")
```

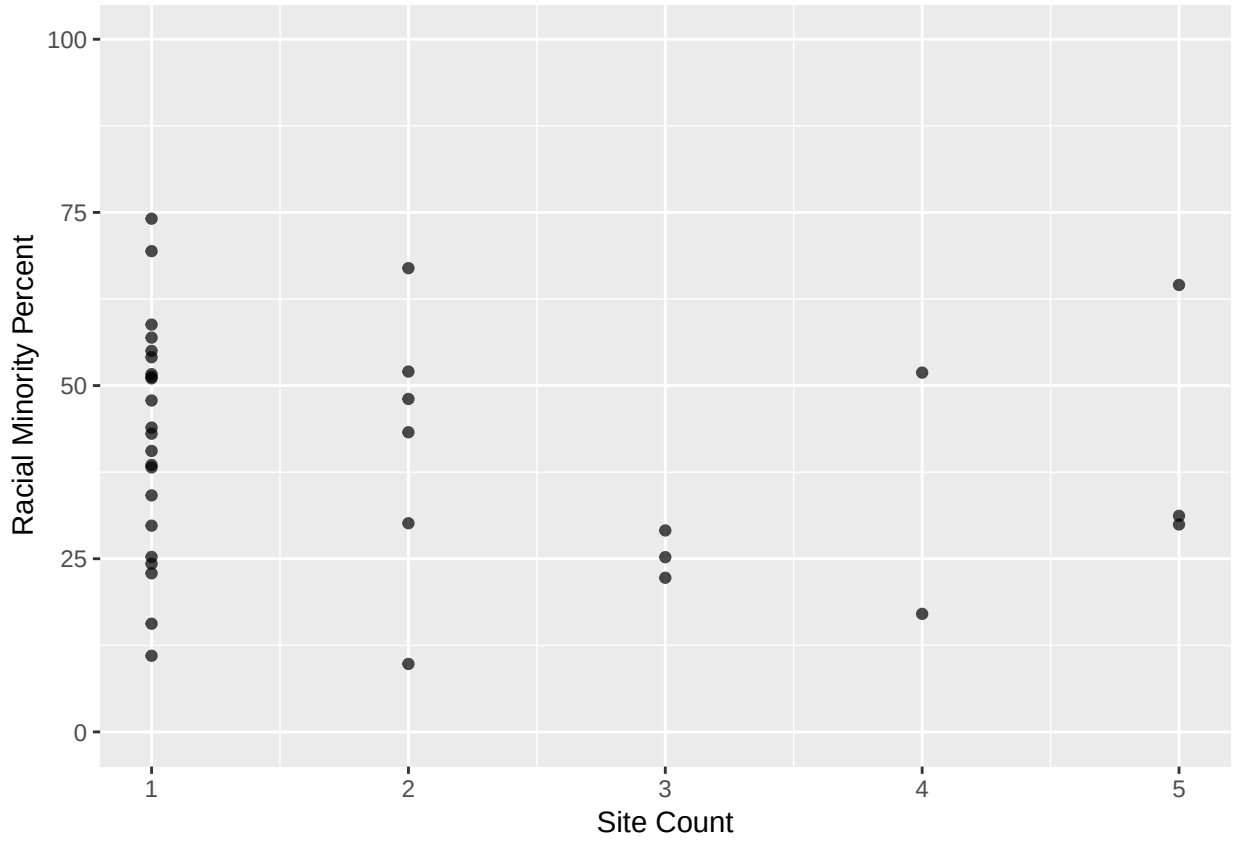



```
#Graph count data with race percent data
Landfill_plot <-
  ggplot(Sitecount_Landfill_join, aes(x=n , y=MinorityPerc * 100)) +
  scale_y_continuous(limits=c(0, 100)) +
  geom_point(alpha=0.7, color="blue") +
  labs(x="Site Count", y="Racial Minority Percent")

IH_plot <-
  ggplot(Sitecount_IH_join, aes(x=n , y=MinorityPerc * 100)) +
  scale_y_continuous(limits=c(0, 100)) +
  geom_point(alpha=0.7, color="blue") +
  labs(x="Site Count", y="Racial Minority Percent")

FRB_plot <-
  ggplot(Sitecount_FRB_join, aes(x=n , y=MinorityPerc * 100)) +
  scale_y_continuous(limits=c(0, 100)) +
  geom_point(alpha=0.7) +
  labs(x="Site Count", y="Racial Minority Percent")
FRB_plot
```

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




```
BF_plot <-
  ggplot(Sitecount_BF_join, aes(x=n , y=MinorityPerc * 100)) +
  scale_y_continuous(limits=c(0, 100)) +
  geom_point(alpha=0.7, color="blue") +
  labs(x="Site Count", y="Racial Minority Percent")

RUST_plot <-
  ggplot(Sitecount_RUST_join, aes(x=n , y=MinorityPerc * 100)) +
  scale_y_continuous(limits=c(0, 100)) +
  geom_point(alpha=0.7, color="blue") +
  labs(x="Site Count", y="Racial Minority Percent")

HW_plot <-
  ggplot(Sitecount_HW_join, aes(x=n , y=MinorityPerc * 100)) +
  scale_y_continuous(limits=c(0, 100)) +
  geom_point(alpha=0.7, color="blue") +
  labs(x="Site Count", y="Racial Minority Percent")
```

6 Summary and Conclusions

7 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