

Spatial Epi Project

Leah Moubadder, Jasmine Aqua, Susan Hoffman

11/11/2021

Contents

Background	1
Outcomes of interest (by census tract):	1
Setting up	2
Editing data	3
Creating redlining maps	5
Creating the 15 panel map	5
Desity plots	9
Moran's I	13
Results	13
Spatial Durbin Model	13
Results	13

Background

A prime example of structural racism is the 1930s the practice of redlining by the Home Owners Loan Corporation (HOLC). Although redlining is now illegal, the structure remains and impacts health outcomes today.

Aim: to understand the impact of redlining in Georgia's HOLC regions on health hazards (diesel particulate matter [PM] levels, respiratory hazard index, and cancer risk due to air toxics)

Outcomes of interest (by census tract):

- National Scale Air Toxics Assessment Air Toxics Cancer Risk (CANCER)
- National Scale Air Toxics Assessment Respiratory Hazard Index (RESP)
- National Scale Air Toxics Assessment Diesel PM (DPM) (DSLPM)

Hazard index: The sum of the hazard quotients for toxics that affect the same target organ/target system. An HI ≤ 1 indicates noncancer effect not likely to occur

Cancer risk: the results of cancer dose-response assessments are converted into a unit risk estimate. This is multiplied by the estimated inhalation exposure concentration over a lifetime (70 years) to estimate an individual's lifetime cancer risk.

##Exposure of interest (aggregated to census tract): - Redlining

Data from University of Michigan Social Science Institute, by Helen Meier

Setting up

NOTE: EJScreen data

For ease of importing to GitHub (file size issues), we limited the data to only areas within GA before loading the dataset to the project

```
# Packages
library(tmap) # Mapping spatial file
library(tidyverse) # data wrangling tools
library(dplyr) # data wrangling tools
library(sp) # spatial methods package
library(readxl) # import excel files
library(rgdal) # spatial methods package
library(readr) # data file loading package
library(sf) # spatial methods package
```

```
setwd("C:/Users/SSBUCKE/OneDrive - Emory University/Spatial Project/spatialfinal2021")
```

```
# Data files
```

```
## outcome data
```

```
ej_ga <- read_csv("Data/ej_ga.csv",
                  col_types = cols(...1 = col_skip())) %>%
  select('ID', 'DSLPM', 'CANCER', 'RESP')
```

```
## exposure geometry data
```

```
HOLC_map <- readOGR(dsn=path.expand("Data/HRS2010-Shapefiles/HRS2010"),
                    layer="HRS2010")
```

```
## OGR data source with driver: ESRI Shapefile
```

```
## Source: "C:\Users\SSBUCKE\OneDrive - Emory University\Spatial Project\spatialfinal2021\Data\HRS2010-
```

```
## with 12888 features
```

```
## It has 16 fields
```

```
## Integer64 fields read as strings: OBJECTID_1
```

```
## exposure attribute data
```

```
HOLC_score <- read_excel("Data/Historic Redlining Score 2010.xlsx")
```

```
## Importing Georgia Census Tract Geographic Boundary file
```

```
## updated 2020 data
```

```
ga_tracts_10 <- readOGR(dsn=path.expand("Data/tl_2020_13_all"), layer="tl_2020_13_tract10")
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "C:\Users\SSBUCKE\OneDrive - Emory University\Spatial Project\spatialfinal2021\Data\tl_2020_
## with 1969 features
## It has 12 fields
## Integer64 fields read as strings:  ALAND10 AWATER10
```

Editing data

Data convention HOLC_map_cityname will be used to create datasets with extra census tracts for mapping purposes.

HOLC_full: Census tract + HOLC data, all Georgia census tracts
HOLC_full_georgia: Census tract + HOLC data, all Georgia census tracts, only GA census tracts with HOLC data
HOLC_full_atlanta: Census tract + HOLC data, only census tracts in Atlanta with HOLC data
HOLC_full_augusta: Census tract + HOLC data, only census tracts in Augusta with HOLC data
HOLC_full_columbus: Census tract + HOLC data, only census tracts in Columbus with HOLC data
HOLC_full_macon: Census tract + HOLC data, only census tracts in Macon with HOLC data
HOLC_full_savannah: Census tract + HOLC data, only census tracts in Savannah with HOLC data

```
# Cleaning HOLC score data
HOLC_score2 <- HOLC_score %>%
  filter(substr(GEOID10,1,2) == "13") # restricting to only GA

# cleaning EJScreen data

ej_ga2 <- ej_ga %>%
  mutate(GEOID10 = substr(ej_ga$ID, 1, 11)) %>%
  select(-'ID')

# combining all 3 datasets together

ej_ga2 <- ej_ga2[!duplicated(ej_ga2$GEOID10),]

ej_ga2 <- sp::merge(ga_tracts_10, ej_ga2,
  by = 'GEOID10',
  all.y = T,
  duplicateGeoms = T,
  na.strings = 'Missing')

full_data <- sp::merge(ej_ga2, HOLC_score2,
  by = 'GEOID10',
  all.y = T,
  duplicateGeoms = T,
  na.strings = 'Missing')

# Subsetting data into
## All GA (only redlined areas)
## Atlanta redlined areas
## Augusta redlined areas
## Macon redlined areas
## Columbus redlined areas
## Savannah redlined areas
```

```

MyPalette <- c("#A8FF33", "#81FOFF", "#FAFF93", "#FF9693")
MyLabels <- c("Low (Q1)", "Medium (Q2)", "High (Q3)", "Very High (Q4)")

# All GA
ga_ids <- HOLC_score2$GEOID10
full_data_georgia <- subset(full_data, GEOID10 %in% ga_ids)

# writeOGR(obj=full_data_georgia, dsn="tempdir", layer="full_data_georgia", driver="ESRI Shapefile")

# Atlanta only
atlanta <- HOLC_score2 %>%
  filter(CBSA=="12060")
atlanta_ids <- atlanta$GEOID10
full_data_atlanta <- subset(full_data, GEOID10 %in% atlanta_ids)

# writeOGR(obj=full_data_atlanta, dsn="tempdir", layer="full_data_atlanta", driver="ESRI Shapefile")

# Augusta only
augusta <- HOLC_score2 %>%
  filter(CBSA=="12260")
augusta_ids <- augusta$GEOID10
full_data_augusta <- subset(full_data, GEOID10 %in% augusta_ids)

# writeOGR(obj=full_data_augusta, dsn="tempdir", layer="full_data_augusta", driver="ESRI Shapefile")

# Columbus only
columbus <- HOLC_score2 %>%
  filter(CBSA=="17980")
columbus_ids <- columbus$GEOID10
full_data_columbus <- subset(full_data, GEOID10 %in% columbus_ids)

# writeOGR(obj=full_data_columbus, dsn="tempdir", layer="full_data_columbus", driver="ESRI Shapefile")

# Macon only
macon <- HOLC_score2 %>%
  filter(CBSA=="31420")
macon_ids <- macon$GEOID10
full_data_macon <- subset(full_data, GEOID10 %in% macon_ids)

# writeOGR(obj=full_data_macon, dsn="tempdir", layer="full_data_macon", driver="ESRI Shapefile")

# Savannah only
savannah <- HOLC_score2 %>%
  filter(CBSA=="42340")
savannah_ids <- savannah$GEOID10
full_data_savannah <- subset(full_data, GEOID10 %in% savannah_ids)

# writeOGR(obj=full_data_savannah, dsn="tempdir", layer="full_data_savannah", driver="ESRI Shapefile")

```

Creating redlining maps

JASMINE POST HERE

Creating the 15 panel map

Note: for the poster, individual images were exported in order to make a “cleaner” figure. For this RMarkdown, we are producing the 15 panel map using `tmap_arrange`, so the two figures might look different.

Since we are treating each city as spatial islands, local quantiles were used.

```
# Diesel PM

dpm_atl <- tm_shape(full_data_atlanta) +
  tm_fill('DSLPM',
    style = 'quantile',
    palette = 'BuPu',
    title = 'Diesel PM') +
  tm_borders() +
  tm_layout(main.title = 'Diesel PM: \nAtlanta',
    main.title.size = 0.9,
    legend.outside = T,
    legend.outside.size = .5,
    frame = F)

dpm_aug <- tm_shape(full_data_augusta) +
  tm_fill('DSLPM',
    style = 'quantile',
    palette = 'BuPu',
    title = 'Diesel PM') +
  tm_borders() +
  tm_layout(main.title = 'Diesel PM: \nAugusta',
    main.title.size = 0.9,
    legend.outside = T,
    legend.outside.size = .5,
    frame = F)

dpm_mac <- tm_shape(full_data_macon) +
  tm_fill('DSLPM',
    style = 'quantile',
    palette = 'BuPu',
    title = 'Diesel PM') +
  tm_borders() +
  tm_layout(main.title = 'Diesel PM: \nMacon',
    main.title.size = 0.9,
    legend.outside = T,
    legend.outside.size = .5,
    frame = F)
```

```

dpm_sav <- tm_shape(full_data_savannah) +
  tm_fill('DSLPM',
    style = 'quantile',
    palette = 'BuPu',
    title = 'Diesel PM') +
  tm_borders() +
  tm_layout(main.title = 'Diesel PM: \nSavannah',
    main.title.size = 0.9,
    legend.outside = T,
    legend.outside.size = .5,
    frame = F)

dpm_col <- tm_shape(full_data_columbus) +
  tm_fill('DSLPM',
    style = 'quantile',
    palette = 'BuPu',
    title = 'Diesel PM') +
  tm_borders() +
  tm_layout(main.title = 'Diesel PM: \nColumbus',
    main.title.size = 0.9,
    legend.outside = T,
    legend.outside.size = .5,
    frame = F)

# RESP

res_atl <- tm_shape(full_data_atlanta) +
  tm_fill('RESP',
    style = 'quantile',
    palette = 'RdPu',
    title = 'Respiratory Hazard') +
  tm_borders() +
  tm_layout(main.title = 'Respiratory Hazard: \nAtlanta',
    main.title.size = 0.9,
    legend.outside = T,
    legend.outside.size = .5,
    frame = F)

res_aug <- tm_shape(full_data_augusta) +
  tm_fill('RESP',
    style = 'quantile',
    palette = 'RdPu',
    title = 'Respiratory Hazard') +
  tm_borders() +
  tm_layout(main.title = 'Respiratory Hazard: \nAugusta',
    main.title.size = 0.9,
    legend.outside = T,
    legend.outside.size = .5,
    frame = F)

```

```

res_mac <- tm_shape(full_data_macon) +
  tm_fill('RESP',
    style = 'quantile',
    palette = 'RdPu',
    title = 'Respiratory Hazard') +
  tm_borders() +
  tm_layout(main.title = 'Respiratory Hazard: \nMacon',
    main.title.size = 0.9,
    legend.outside = T,
    legend.outside.size = .5,
    frame = F)

res_sav <- tm_shape(full_data_savannah) +
  tm_fill('RESP',
    style = 'quantile',
    palette = 'RdPu',
    title = 'Respiratory Hazard') +
  tm_borders() +
  tm_layout(main.title = 'Respiratory Hazard: \nSavannah',
    main.title.size = 0.9,
    legend.outside = T,
    legend.outside.size = .5,
    frame = F)

res_col <- tm_shape(full_data_columbus) +
  tm_fill('RESP',
    style = 'quantile',
    palette = 'RdPu',
    title = 'Respiratory Hazard') +
  tm_borders() +
  tm_layout(main.title = 'Respiratory Hazard: \nColumbus',
    main.title.size = 0.9,
    legend.outside = T,
    legend.outside.size = .5,
    frame = F)

# Cancer

can_atl <- tm_shape(full_data_atlanta) +
  tm_fill('CANCER',
    style = 'quantile',
    palette = 'PuBu',
    title = 'Cancer Risk') +
  tm_borders() +
  tm_layout(main.title = 'Cancer Hazard: \nAtlanta',
    main.title.size = 0.9,
    legend.outside = T,
    legend.outside.size = .5,
    frame = F)

```

```

can_aug <- tm_shape(full_data_augusta) +
  tm_fill('CANCER',
    style = 'quantile',
    palette = 'PuBu',
    title = 'Cancer Risk') +
  tm_borders() +
  tm_layout(main.title = 'Cancer Hazard: \nAugusta',
    main.title.size = 0.9,
    legend.outside = T,
    legend.outside.size = .5,
    frame = F)

can_mac <- tm_shape(full_data_macon) +
  tm_fill('CANCER',
    style = 'quantile',
    palette = 'PuBu',
    title = 'Cancer Risk') +
  tm_borders() +
  tm_layout(main.title = 'Cancer Hazard: \nMacon',
    main.title.size = 0.9,
    legend.outside = T,
    legend.outside.size = .5,
    frame = F)

can_sav <- tm_shape(full_data_savannah) +
  tm_fill('CANCER',
    style = 'quantile',
    palette = 'PuBu',
    title = 'Cancer Risk') +
  tm_borders() +
  tm_layout(main.title = 'Cancer Hazard: \nSavannah',
    main.title.size = 0.9,
    legend.outside = T,
    legend.outside.size = .5,
    frame = F)

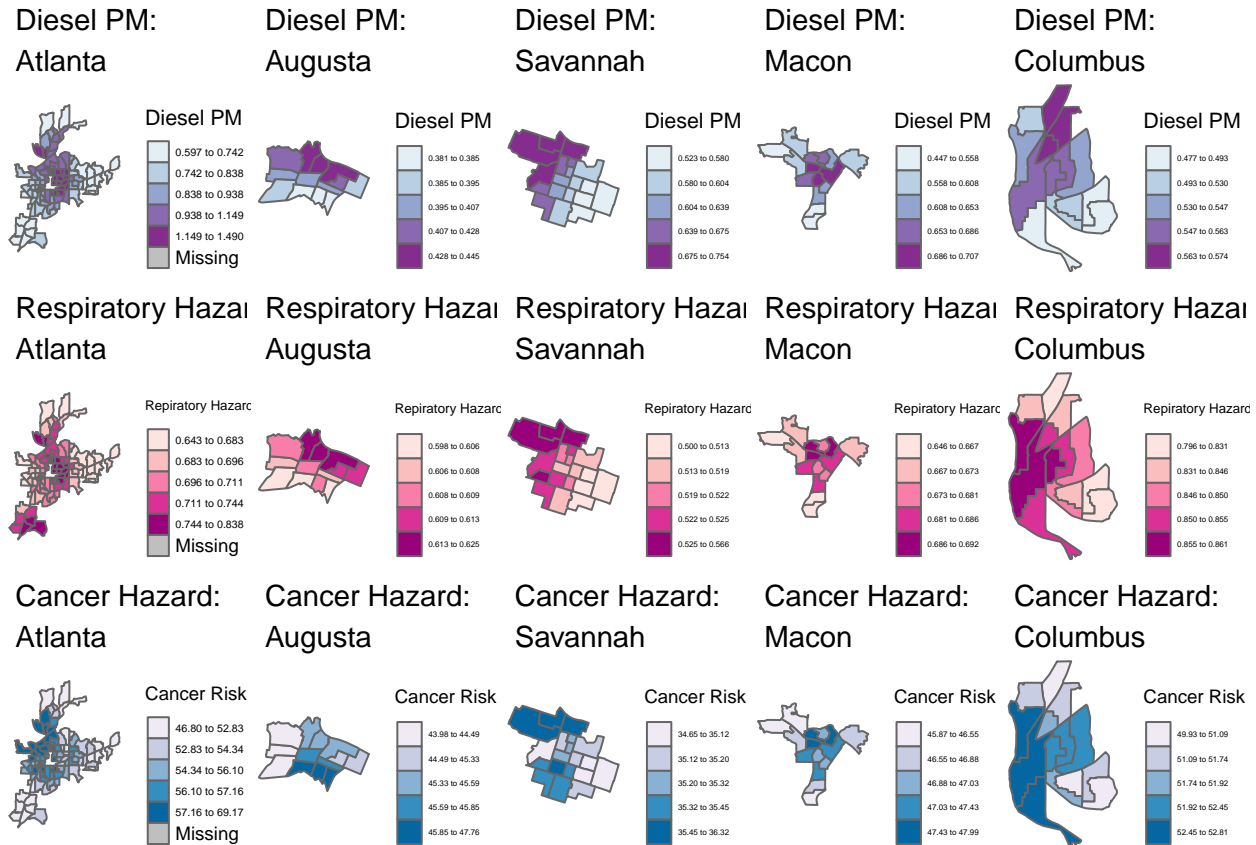
can_col <- tm_shape(full_data_columbus) +
  tm_fill('CANCER',
    style = 'quantile',
    palette = 'PuBu',
    title = 'Cancer Risk') +
  tm_borders() +
  tm_layout(main.title = 'Cancer Hazard: \nColumbus',
    main.title.size = 0.9,
    legend.outside = T,
    legend.outside.size = .5,
    frame = F)

tmap_arrange(dpm_atl, dpm_aug, dpm_sav, dpm_mac, dpm_col,

```



```
res_atl, res_aug, res_sav, res_mac, res_col,
can_atl, can_aug, can_sav, can_mac, can_col,
nrow = 3,
ncol = 5)
```



Desity plots

In order to be transparent about the data distributions in each city, we created density plots of each health hazard by HOLC region in GA.

```
# Creating a dataset that removes the spatial aspect to run the plots

hist <- as.data.frame(full_data_georgia) %>%
  select(c(CBSA, CANCER, DSLPM, RESP))

# Creating a label for each of the HOLC regions in GA
hist$cities <- factor(hist$CBSA,
  levels = c(12060, 42340, 17980, 31420, 12260),
  labels = c('Atlanta', 'Savannah', 'Columbus', 'Macon', 'Augusta'))

# Creating desity plots

dslpm <- ggplot(hist, aes(x = DSLPM, fill = cities)) +
```

```

geom_density(alpha = 0.4) +
ggtitle('NATA Diesel PM Density per City in Georgia') +
xlab('Diesel PM') +
ylab('Density') +
labs(fill = 'HOLC Regions') +
theme(panel.grid.major = element_blank(),
       panel.grid.minor = element_blank(),
       panel.background = element_blank(),
       axis.line = element_line(colour = "black"))

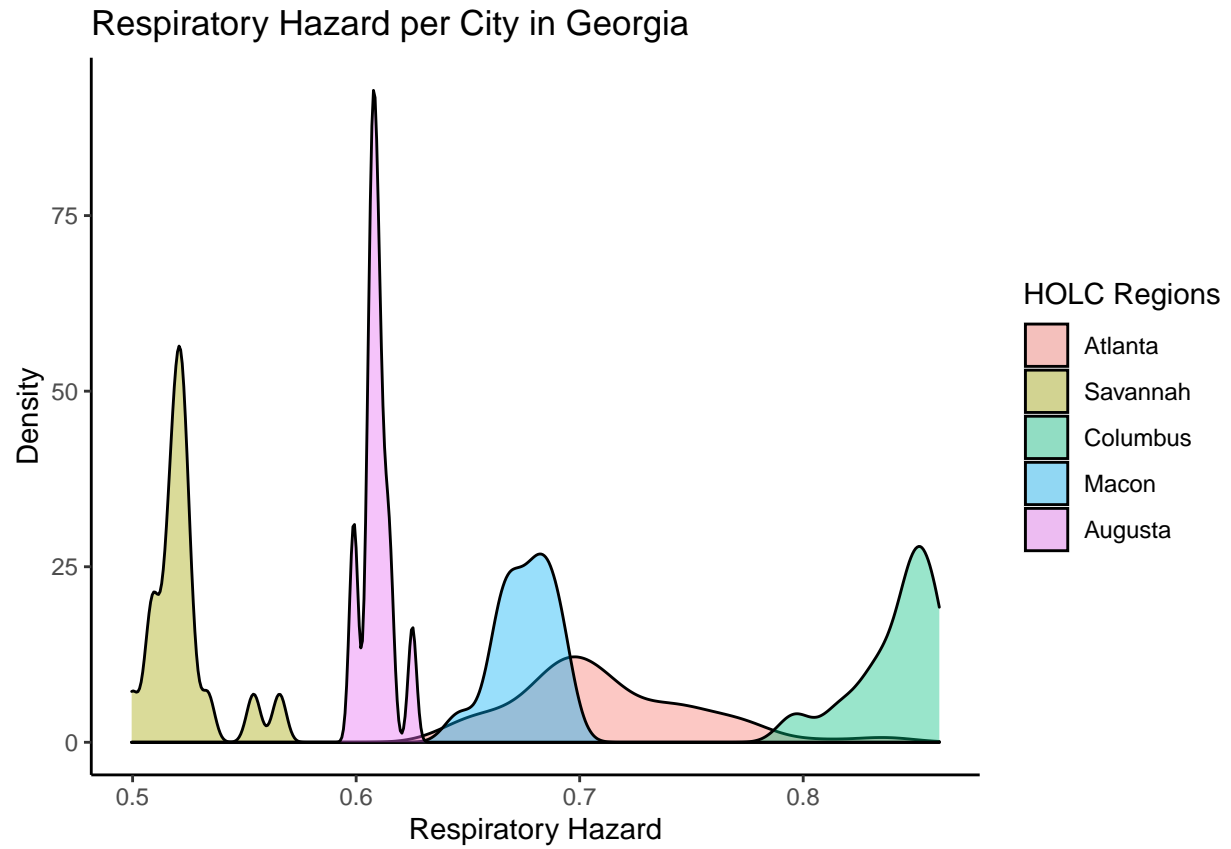
cancer <- ggplot(hist, aes(x = CANCER, fill = cities)) +
  geom_density(alpha = 0.4) +
  ggtitle('Cancer Risk per City in Georgia') +
  xlab('Cancer Risk') +
  ylab('Density') +
  labs(fill = 'HOLC Regions') +
  theme(panel.grid.major = element_blank(),
       panel.grid.minor = element_blank(),
       panel.background = element_blank(),
       axis.line = element_line(colour = "black"))

resp <- ggplot(hist, aes(x = RESP, fill = cities)) +
  geom_density(alpha = 0.4) +
  ggtitle('Respiratory Hazard per City in Georgia') +
  xlab('Respiratory Hazard') +
  ylab('Density') +
  labs(fill = 'HOLC Regions') +
  theme(panel.grid.major = element_blank(),
       panel.grid.minor = element_blank(),
       panel.background = element_blank(),
       axis.line = element_line(colour = "black"))

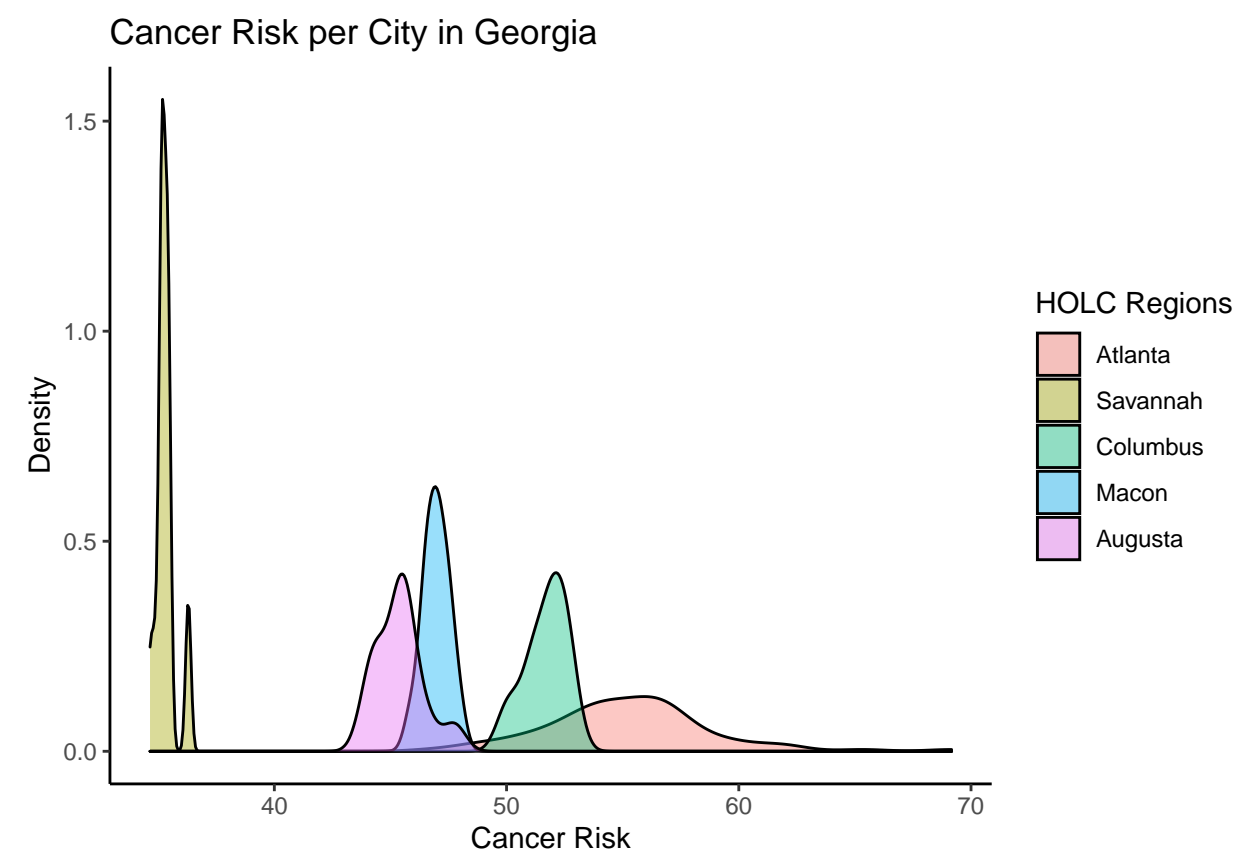
# Printing Density plots

resp

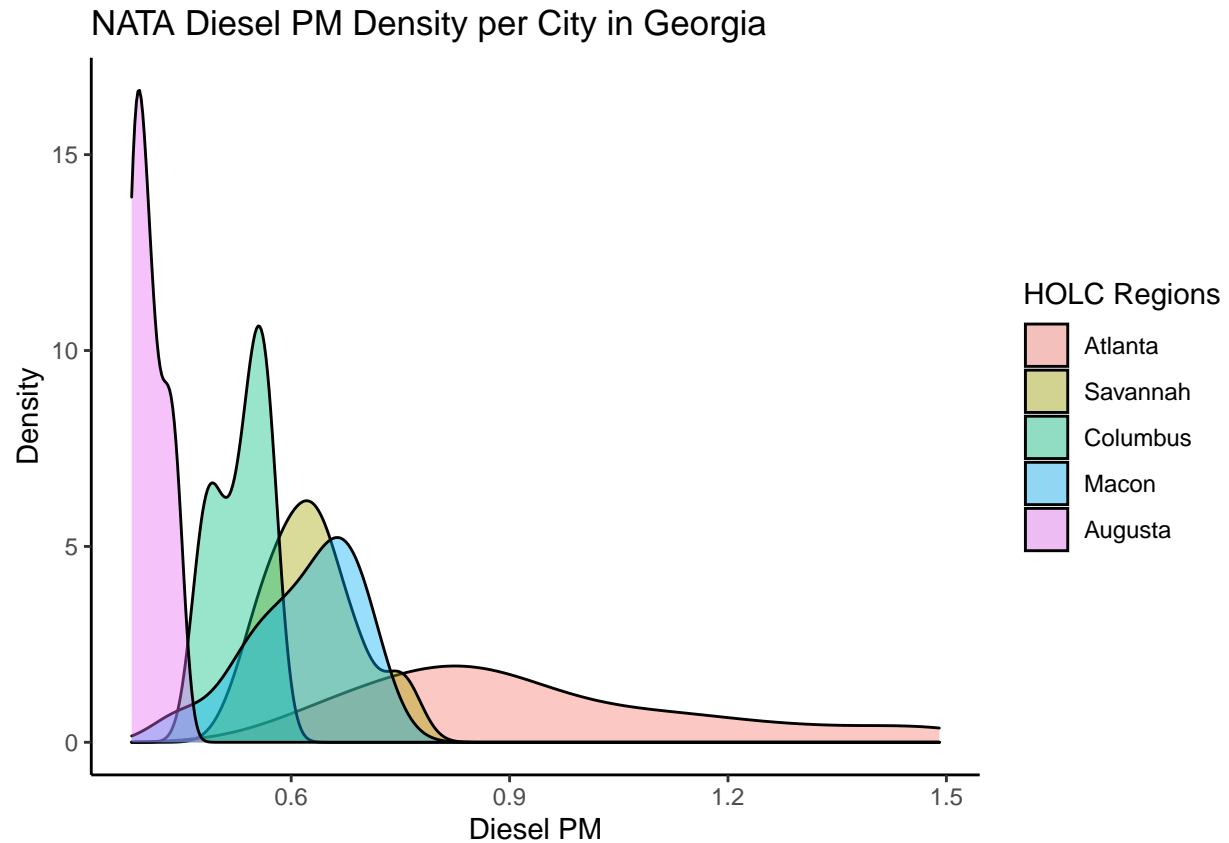
```



cancer



dslpm



Moran's I

LEAH POST HERE

Results

Spatial dependence (Global Moran's I) was significant in all scenarios

Spatial Durbin Model

LEAH POST HERE

Results

Higher redlining was associated with an increase in the total impact of all health hazards in all HOLC regions, with the exception of diesel PM in Columbus and Augusta