Mapping the Erasure: Spatial Legacies of Redlining in East New York, Brooklyn

DSAN 6750 / PPOL 6805: GIS for Spatial Data Science

Anile Choice

```
library(tidyverse)
```

```
library(sf)
```

```
library(spdep)
```

```
library(spatstat)
```

```
library(mapview)
library(viridis)
```

```
library(knitr)
```

Source: Article Notebook

Introduction

The systematic destruction of Black communities through urban interventions has left lasting imprints on American cities. During the mid-twentieth century, the Home Owners' Loan Corporation (HOLC) created "residential security maps" that classified neighborhoods by perceived investment risk. Areas with significant Black populations were systematically marked as "hazardous" (Grade D, colored red on maps)—a practice that became known as "redlining."

Brooklyn's East New York neighborhood exemplifies this pattern. Once a thriving community, East New York was heavily redlined in the 1930s, setting in motion decades of disinvestment and spatial inequality. This study investigates whether these historical interventions continue to shape contemporary patterns of racial composition and income in Brooklyn.

Research Question

Do the 1930s HOLC redlining zones in Brooklyn continue to shape the spatial distribution of race and income in these areas today?

I hypothesize that:

1. Redlined areas (Grade D) exhibit distinct spatial patterns in racial composition compared to non-redlined areas
2. These patterns show **spatial autocorrelation** (clustering), indicating systematic rather than random distribution
3. Income patterns similarly cluster along historical redlining boundaries

Data & Methods

Data Sources

HOLC Redlining Maps (1930s)

Historical "residential security" maps from the Mapping Inequality project, showing Grade A (green/best), B (blue), C (yellow), and D (red/hazardous) zones in Brooklyn.

Census Data (1990-2020)

Time series demographic and income data from the National Historical Geographic Information System (NHGIS):

- Race/ethnicity by census tract (1990, 2000, 2010, 2020)
- Median household income (1990, 2000, 2009-2013, 2018-2022)
- Census tract boundaries for all four periods

Study Area

My analysis focuses on **Brooklyn (Kings County), New York**, with particular attention to **East New York**—an area in eastern Brooklyn with high concentrations of D-grade (redlined) zones. I identified 161 census tracts in eastern Brooklyn where >25% of the tract area overlaps with redlined zones.

Analytical Approach

1. **Exploratory Spatial Data Analysis**: Visualize the relationship between historical redlining and contemporary demographic patterns
2. **Global Spatial Autocorrelation**: Calculate Moran's I to detect overall clustering patterns
3. **Local Spatial Autocorrelation**: Use Local Indicators of Spatial Association (LISA) to identify specific clusters
4. **Hypothesis Testing**: Monte Carlo simulations (999 iterations) to test whether observed patterns differ from spatial randomness

Exploratory Data Analysis

First, let's load the processed data and examine the basic patterns.

```
# Load processed data
redline <- st_read("data/raw/brooklyn.json", quiet = TRUE)
tracts_2020 <- st_read("data/processed/brooklyn_tracts_with_data_2020.gpkg",
                       quiet = TRUE)
east_ny_tracts <- read_csv("data/processed/east_new_york_tracts.csv",
                           show_col_types = FALSE)

# Calculate demographic variables for 2020
tracts_2020 <- tracts_2020 %>%
  mutate(
    # Total population
    total_pop = CV4AA2020 + CV4AB2020 + CV4AC2020 + CV4AD2020 + CV4AE2020 +
                CV4AF2020 + CV4AG2020 + CV4AH2020 + CV4AI2020 + CV4AJ2020,
    # Percent Black (Non-Hispanic)
    pct_black = ifelse(total_pop > 0, (CV4AB2020 / total_pop) * 100, NA),
    # Median household income (2018-2022)
    med_income = B79AA225,
    # Flag East New York tracts
    is_east_ny = GISJOIN %in% east_ny_tracts$GISJOIN
  )
```

Source: Article Notebook

Summary Statistics

```
# Overall summary
cat("Brooklyn Census Tracts (2020):", nrow(tracts_2020), "\n")
```

```
Brooklyn Census Tracts (2020): 804
```

```
cat("East New York Tracts:", sum(tracts_2020$is_east_ny), "\n\n")
```

```
East New York Tracts: 161
```

```
# Summary by East New York status
summary_table <- tracts_2020 %>%
  st_drop_geometry() %>%
  group_by(is_east_ny) %>%
  summarize(
    n_tracts = n(),
    mean_pct_black = mean(pct_black, na.rm = TRUE),
    median_income = median(med_income, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(area = ifelse(is_east_ny, "East New York (Redlined)",
                       "Rest of Brooklyn"))

kable(summary_table %>% select(area, n_tracts, mean_pct_black, median_income),
      col.names = c("Area", "N Tracts", "Mean % Black", "Median Income"),
      digits = 2)
```

Area	N Tracts	Mean % Black	Median Income
Rest of Brooklyn	643	0.03	76908
East New York (Redlined)	161	0.08	69365

Source: Article Notebook

Spatial Visualization: Black Population and Redlining

```
ggplot() +
  geom_sf(data = tracts_2020, aes(fill = pct_black), color = NA) +
  scale_fill_viridis_c(option = "plasma", na.value = "grey90",
                       name = "% Black\n(2020)",
                       labels = scales::percent_format(scale = 1)) +
  geom_sf(data = redline, aes(color = grade), fill = NA, linewidth = 0.8) +
  scale_color_manual(values = c("A" = "#76a865", "B" = "#7cb5bd",
                                "C" = "#ffff00", "D" = "#d9838d"),
                     name = "HOLC Grade\n(1930s)") +
  theme_minimal() +
  labs(title = "Black Population Distribution and Historical Redlining",
       subtitle = "Brooklyn, NY")
```

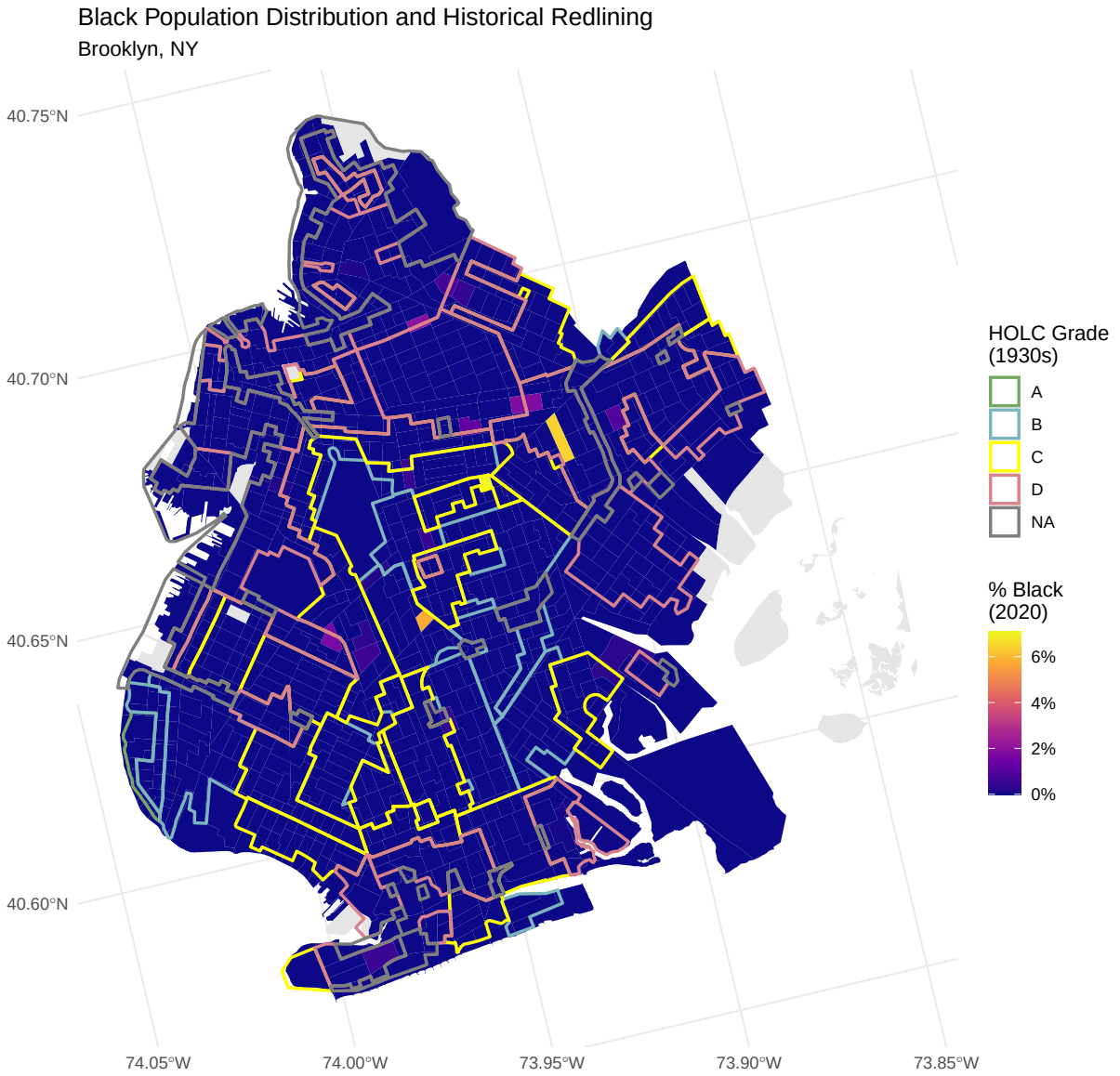


Figure 1: Percentage Black Population (2020) with HOLC Redlining Zones

Source: Article Notebook

Spatial Visualization: Income and Redlining

```
ggplot() +
  geom_sf(data = tracts_2020, aes(fill = med_income), color = NA) +
```

```
  scale_fill_viridis_c(option = "mako", na.value = "grey90",
                       name = "Median HH\nIncome ($)",
                       labels = scales::dollar_format()) +
  geom_sf(data = redline, aes(color = grade), fill = NA, linewidth = 0.8) +
  scale_color_manual(values = c("A" = "#76a865", "B" = "#7cb5bd",
                                "C" = "#ffff00", "D" = "#d9838d"),
                     name = "HOLC Grade\n(1930s)") +
  theme_minimal() +
  labs(title = "Median Household Income and Historical Redlining",
       subtitle = "Brooklyn, NY")
```

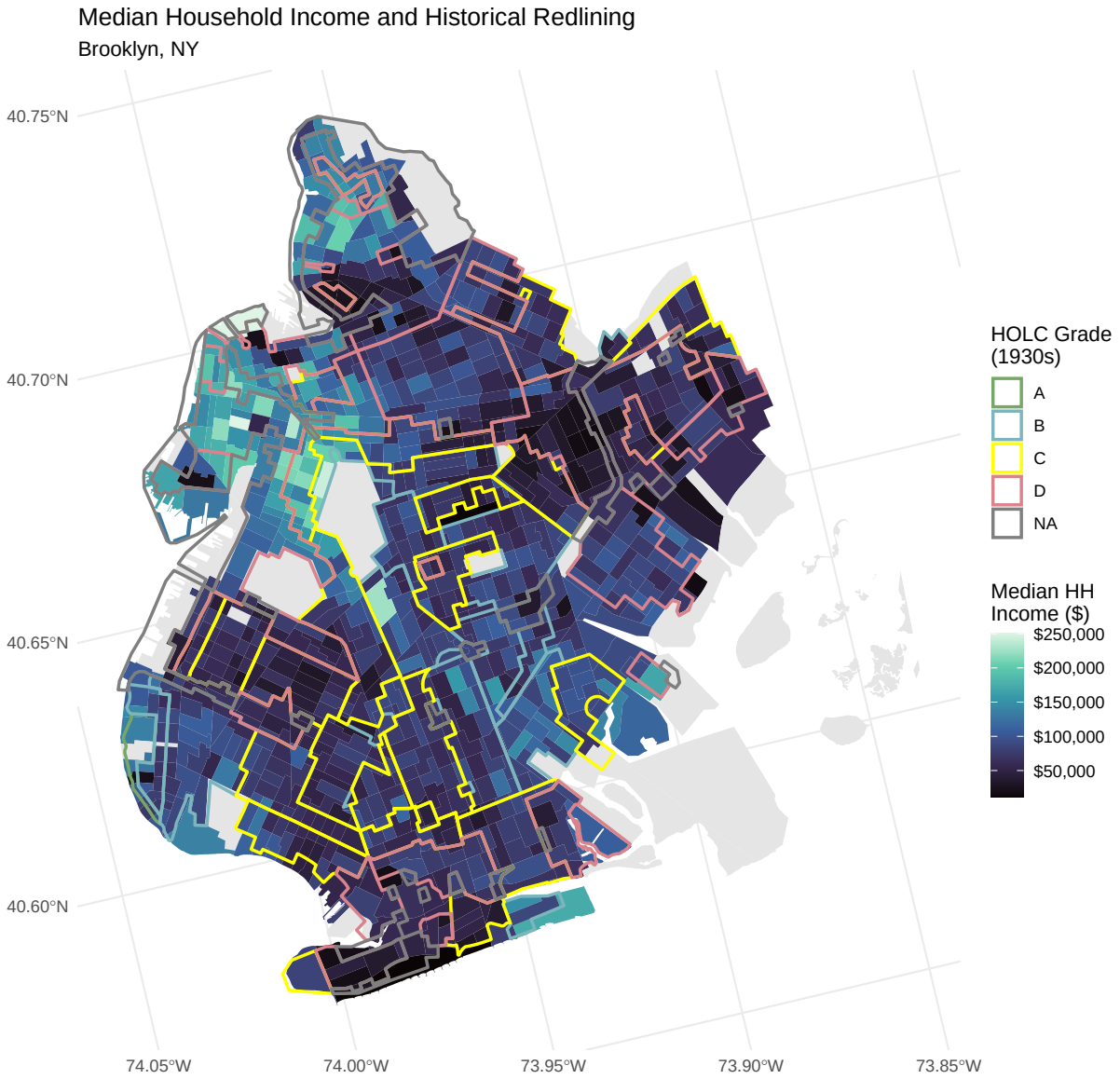


Figure 2: Median Household Income (2018-2022) with HOLC Redlining Zones

Source: Article Notebook

Spatial Autocorrelation Analysis

Global Moran's I

We now test for global spatial autocorrelation using Moran's I. This measures whether similar values cluster together more than would be expected by chance.

```
# Create spatial weights matrix (Queen contiguity)
# Remove tracts with missing values
tracts_complete <- tracts_2020 %>%
  filter(!is.na(pct_black), !is.na(med_income))

# Create neighbors
nb <- poly2nb(tracts_complete, queen = TRUE)
# Create spatial weights
lw <- nb2listw(nb, style = "W", zero.policy = TRUE)

# Moran's I for percent Black
moran_black <- moran.test(tracts_complete$pct_black, lw, zero.policy = TRUE)
cat("Moran's I for % Black Population:\n")
```

```
Moran's I for % Black Population:
```

```
cat(sprintf("  I = %.4f\n", moran_black$estimate["Moran I statistic"]))
```

```
  I = -0.0073
```

```
cat(sprintf("  p-value = %.4e\n", moran_black$p.value))
```

```
  p-value = 6.2944e-01
```

```
cat(sprintf("  Interpretation: %s\n\n",
            ifelse(moran_black$p.value < 0.05,
                   "Significant spatial clustering detected",
                   "No significant clustering")))
```

```
  Interpretation: No significant clustering
```

```r
# Moran's I for income
moran_income <- moran.test(tracts_complete$med_income, lw,
                           zero.policy = TRUE)
cat("Moran's I for Median Household Income:\n")
```

```
Moran's I for Median Household Income:
```

```r
cat(sprintf("  I = %.4f\n", moran_income$estimate["Moran I statistic"]))
```

```
  I = 0.5915
```

```r
cat(sprintf("  p-value = %.4e\n", moran_income$p.value))
```

```
  p-value = 7.5822e-174
```

```r
cat(sprintf("  Interpretation: %s\n",
            ifelse(moran_income$p.value < 0.05,
                   "Significant spatial clustering detected",
                   "No significant clustering")))
```

```
  Interpretation: Significant spatial clustering detected
```

Source: Article Notebook

Local Indicators of Spatial Association (LISA)

LISA analysis identifies specific areas of clustering—"hot spots" (high values near high values) and "cold spots" (low values near low values).

```r
# Local Moran's I for percent Black
lisa_black <- localmoran(tracts_complete$pct_black, lw, zero.policy = TRUE)

# Add LISA results to data
tracts_complete <- tracts_complete %>%
  mutate(
    lisa_black_i = lisa_black[, "Ii"],
    lisa_black_p = lisa_black[, "Pr(z != E(Ii))"],
    lisa_black_sig = ifelse(lisa_black_p < 0.05, "Significant", "Not Significant")
```

```
  )

# Count significant clusters
cat(sprintf("Significant LISA clusters (p < 0.05): %d out of %d tracts\n",
            sum(tracts_complete$lisa_black_sig == "Significant"),
            nrow(tracts_complete)))
```

```
Significant LISA clusters (p < 0.05): 19 out of 772 tracts
```

Source: Article Notebook

```
ggplot() +
  geom_sf(data = tracts_complete,
          aes(fill = lisa_black_i, alpha = lisa_black_sig),
          color = "grey40", linewidth = 0.1) +
  scale_fill_gradient2(low = "blue", mid = "white", high = "red",
                       midpoint = 0,
                       name = "Local\nMoran's I") +
  scale_alpha_manual(values = c("Significant" = 1, "Not Significant" = 0.3),
                     name = "Significance\n(p < 0.05)") +
  geom_sf(data = redline %>% filter(grade == "D"),
          fill = NA, color = "black", linewidth = 0.8, linetype = "dashed") +
  theme_minimal() +
  labs(title = "Local Spatial Autocorrelation (LISA) - Black Population",
       subtitle = "Brooklyn, NY (Dashed lines = D-grade/redlined zones)")
```

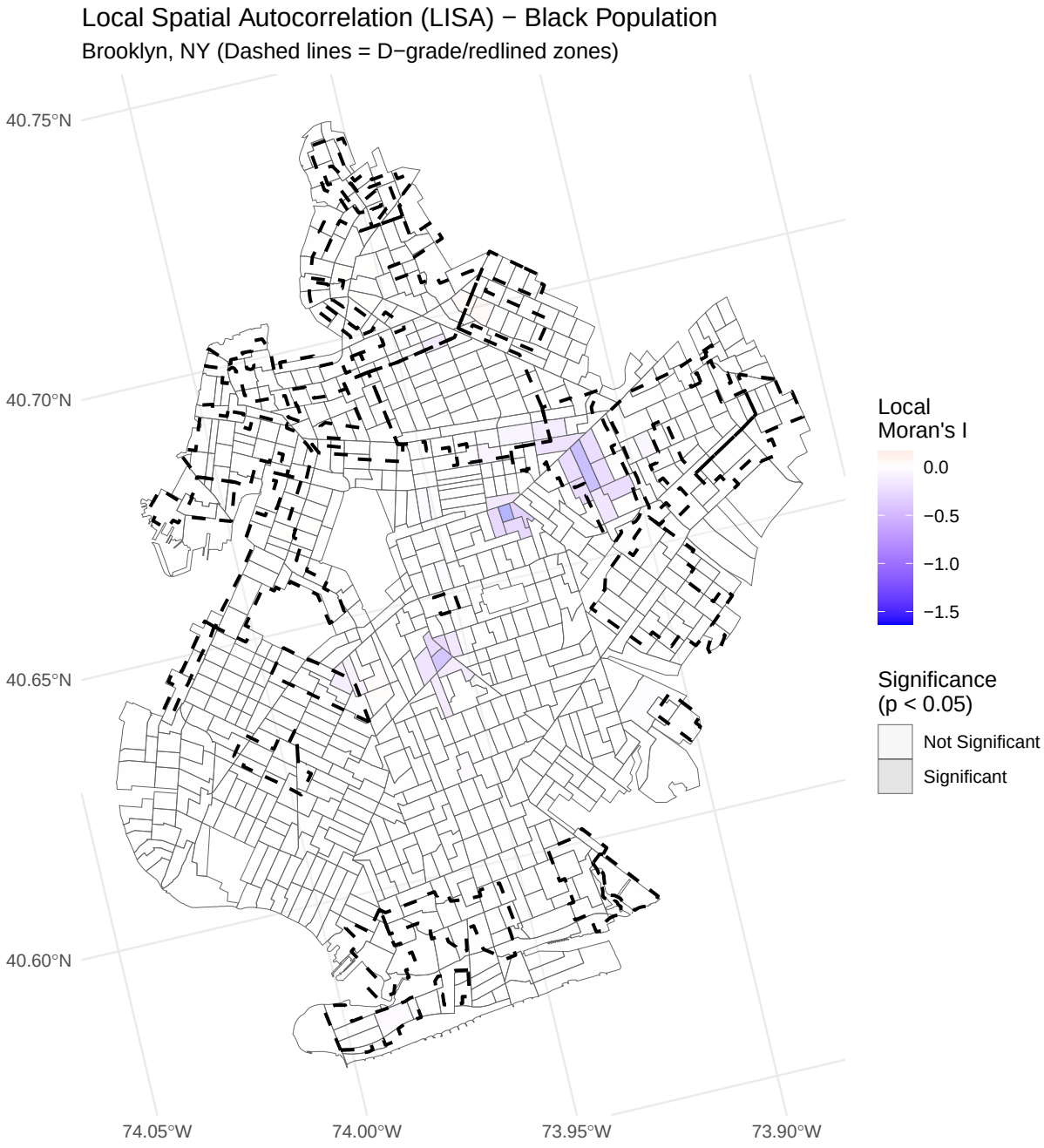


Figure 3: Local Moran's I (LISA) for Black Population

Source: Article Notebook

Monte Carlo Simulation

To formally test our hypothesis, I conduct Monte Carlo simulations comparing observed patterns to random spatial distributions.

```
# Run Monte Carlo simulation for Moran's I (999 permutations)
set.seed(6750)
mc_black <- moran.mc(tracts_complete$pct_black, lw,
                     nsim = 999, zero.policy = TRUE)

cat("Monte Carlo Simulation Results (999 permutations):\n")
```

```
Monte Carlo Simulation Results (999 permutations):
```

```
cat(sprintf("  Observed Moran's I: %.4f\n", mc_black$statistic))
```

```
  Observed Moran's I: -0.0073
```

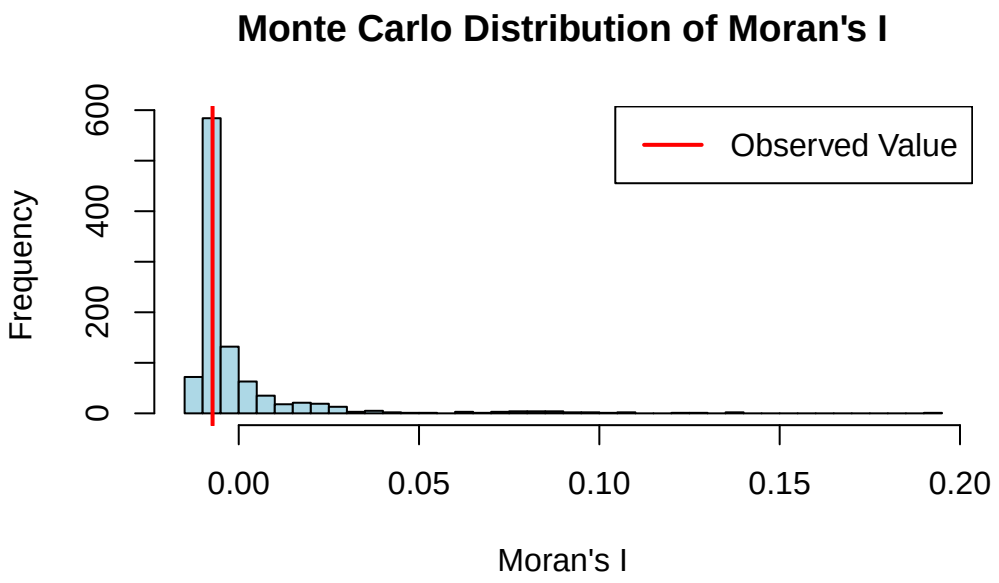
```
cat(sprintf("  p-value: %.4f\n", mc_black$p.value))
```

```
  p-value: 0.4930
```

```
cat(sprintf("  Rank: %d out of 1000\n", mc_black$parameter))
```

```
  Rank: 507 out of 1000
```

```
# Plot simulation distribution
hist(mc_black$res,
     main = "Monte Carlo Distribution of Moran's I",
     xlab = "Moran's I",
     col = "lightblue",
     breaks = 30)
abline(v = mc_black$statistic, col = "red", lwd = 2)
legend("topright", legend = "Observed Value", col = "red", lwd = 2)
```



Source: Article Notebook

Discussion

[*Interpret your findings here. Key questions to address:*]

- *What do the Moran's I results tell us about spatial clustering?*
- *Where are the significant LISA clusters located? Do they correspond to redlined areas?*
- *What does the Monte Carlo simulation reveal about whether patterns are random?*
- *How do current racial and income patterns relate to 1930s redlining?*

Conclusion

[*Summarize your key findings and their implications*]

Limitations

- Ecological fallacy: Census tract-level analysis may not reflect individual experiences
- Changing tract boundaries over time complicate temporal comparisons
- Other historical factors beyond redlining may influence current patterns