

Exploring patterns of Environmental Justice

Aakriti Poudel

2025-10-08

Load libraries

```
# Import libraries
library(tidyverse)
library(sf)
library(tmap)
library(spData) # preloaded spatial data
library(kableExtra)
library(ggplot2)
```

Read data

```
# Read environmental justice screening data
ejscreen <- sf::st_read(here::here("data",
                                    "ejscreen",
                                    "EJSCREEN_2023_BG_StatePct_with_AS_CNMI_GU_VI.gdb"))

# Read Home Owners' Loan Corporation (HOLC) redlining data
map_inequality <- sf::st_read(here::here("data",
                                         "mapping-inequality",
                                         "mapping-inequality-los-angeles.json"))

# Read biodiversity observations data
la_birds <- sf::st_read(here::here("data",
                                    "gbif-birds-LA",
                                    "gbif-birds-LA.shp"))
```

Filter and transform data

```
# Filter to a county you are interested in: Los Angeles County
la <- ejscreen %>%
  filter(ST_ABBREV == "CA") %>%
  filter(CNTY_NAME == c("Los Angeles County"))

# Transform the CRS of a dataset
map_ineq_transform <- st_transform(map_inequality, crs = st_crs(la))

# Check CRS of the dataset
if(st_crs(map_ineq_transform) == st_crs(la)){
  print("Coordinate reference systems match!")
} else {
  warning("Update coordinate reference systems to match!")
}
```

[1] "Coordinate reference systems match!"

Create map

Legacy of redlining in current environmental (in)justice

1. Create a map of historical redlining neighborhoods
 - a. neighborhoods colored by HOLC grade
 - b. an appropriate base map
 - c. a clear title and legend

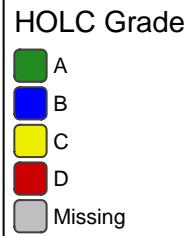
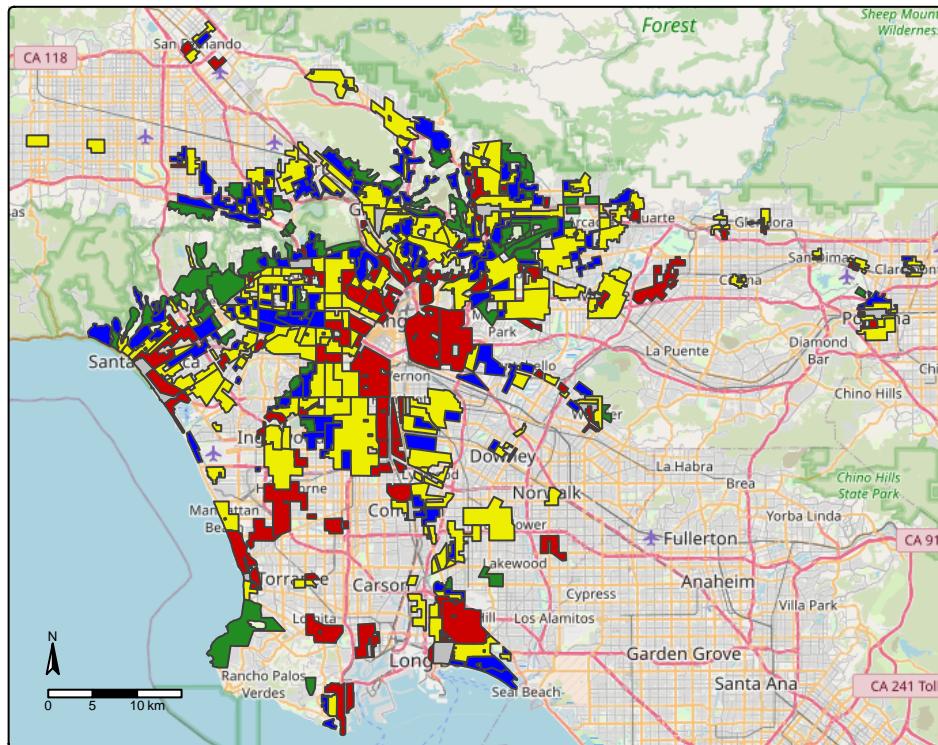
```
# Plot a map of historical redlining neighborhood
tmap_options(component.autoscale = FALSE)

redlining_area <- tm_shape(map_inequality) +
  # neighborhoods colored by HOLC grade
  tm_polygons(fill = 'grade',
    fill.scale = tm_scale_categorical(values = c('A' = 'forestgreen',
                                                 'B' = 'blue',
                                                 'C' = 'yellow2',
                                                 'D' = 'red3')),
    fill.legend = tm_legend(title = 'HOLC Grade')) +
```

```
tm_title(text = 'Historical redlining neighborhoods colored by HOLC grade in LA County',
         fontface = 'bold',
         size = 0.8,
         position = tm_pos_out('center', 'top')) +
tm_compass(type = 'arrow',
            position = c('left', 'bottom'),
            show.labels = 1,
            text.size = 0.4) +
tm_scalebar(position = c('left', 'bottom'),
            text.size = 0.4) +
tm_basemap(c('OpenStreetMap', 'Esri.WorldImagery', 'CartoDB.Positron'))
```

redlining_area

Historical redlining neighborhoods colored by HOLC grade in LA County



Save map

```
# Save HOLC grade block map
holc_block <- tmap_save(redlining_area, here::here('figs', 'holc-block.jpg'),
                           width = 12, height = 12)
```

Spatial join

2. Create a table summarizing:

- the percentage of census block groups that fall within each HOLC grade
- Also include the percent of census black groups that don't fall within a HOLC grade.

Hint: The HOLC data contains the grades and the EJScreen data contains the census blocks, so you will need to combine the data spatially before doing summary statistics. Once you combine and no longer need the geometries, you can use `st_drop_geometry()`.

```
# Spatial join between EJScreen data and HOLC grade data
la_ej <- st_join(map_ineq_transform, la, join = st_intersects) %>%
# Drop geometry
  st_drop_geometry()
```

Summary table

```
# Total number of rows in dataset
total_blocks <- nrow(la_ej)

# Calculate percentage of each HOLC grade
holc_summary <- la_ej %>%
  # Replace NA with descriptive label
  mutate(holc_grade = replace_na(grade, 'No HOLC grade')) %>%
  group_by(holc_grade) %>%
  summarise(count = n()) %>%
  mutate(percentage = (count / sum(count)) * 100)

# Create a summary table
holc_summary_table <- holc_summary %>%
  kable(col.names = c('HOLC grade', 'No. of blocks', '% of blocks'),
        align = 'ccc') %>%
  kable_styling(bootstrap_options = c('striped', 'bordered', 'hover'),
                full_width = FALSE,
                position = 'center') %>%
  column_spec(1:3, width = '1.5in')

holc_summary_table
```

HOLC grade	No. of blocks	% of blocks
------------	---------------	-------------

A	449	7.028804
B	1239	19.395742
C	3058	47.871008
D	1346	21.070758
No HOLC grade	296	4.633688

Data visualization

3. Create at least two visualizations summarizing current conditions (from the EJScreen data) within HOLC grades using the mean of the following variables (you may combine variables or create separate plots):

- a. % low income
- b. percentile for Particulate Matter 2.5
- c. percentile for low life expectancy

Use ggplot for your visualizations! You will first need to calculate mean of each variable grouped by HOLC grade.

```
# Calculate mean
la_ej_mean <- la_ej %>%
  filter(grade != 'NA') %>%
  group_by(grade) %>%
  summarise(lowincome_mean = 100 * mean(LOWINCPCT, na.rm = TRUE), # % low income
            pm_mean = mean(P_PM25, na.rm = TRUE), # percentile for Particulate Matter 2.5
            lifeexp_mean = mean(P_LIFEEXPCT, na.rm = TRUE)) # percentile for low life expectancy

# Change dataset to long dataset for visualization
la_ej_long <- la_ej_mean %>%
  pivot_longer(cols = c(lowincome_mean, pm_mean, lifeexp_mean),
               names_to = 'variable',
               values_to = 'mean_value') %>%
  mutate(variable = case_when(
    variable == 'lowincome_mean' ~ '% low income',
    variable == 'pm_mean' ~ 'PM2.5 percentile',
    variable == 'lifeexp_mean' ~ 'Low life expectancy\npercentile'))
```



```
# Visualization of summary of EJSscreen data in HOLC grade of LA county
summary_ejscreen <- ggplot(la_ej_long, aes(x = grade, y = mean_value, fill = grade)) +
  geom_col(position = 'dodge', linewidth = 0.3) +
  facet_wrap(~variable, scales = 'free_y', ncol = 3) +
  scale_fill_manual(values = c('A' = 'forestgreen',
```

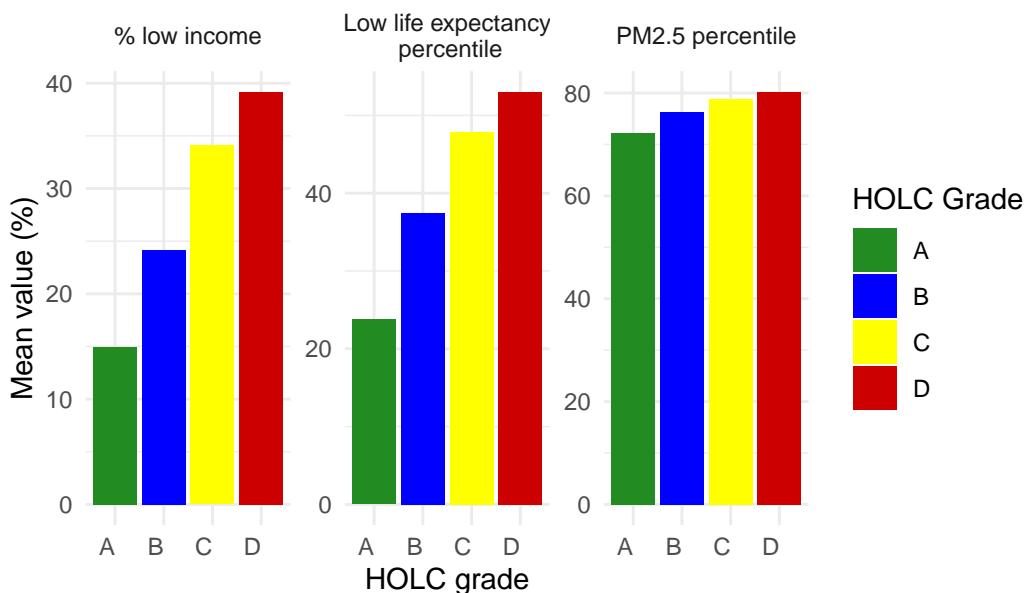
```

'B' = 'blue',
'C' = 'yellow1',
'D' = 'red3')) +
labs(title = 'Summary of EJScreen data in HOLC grade of LA county',
x = 'HOLC grade',
y = 'Mean value (%)',
fill = 'HOLC Grade') +
theme_minimal() +
theme(legend.position = 'right',
axis.text.x = element_text(angle = 0, hjust = 1))

summary_ejscreen

```

Summary of EJScreen data in HOLC grade of LA county



```

# Save the plot of summary of EJScreen data
ggsave(here::here('figs', 'summary-ejscreen.jpg'),
width = 10,
height = 8,
dpi = 300)

```

Interpret patterns

4. Write a brief paragraph reflecting on these results.

- a. Interpret the patterns you observe in your results.

Answer

The ‘Summary of EJScreen data in HOLC grade of LA county’ figure reveals clear disparities across historically redlined neighborhoods.

The low income graph shows that block D neighborhoods, historically redlined areas, have the highest percentage of low income population compared to Grades A, B, and C. This indicates that redlined areas contain higher concentrations of economically disadvantaged populations. The life expectancy graph demonstrates that block D neighborhoods have the highest percentile for low life expectancy. Residents in redlined areas experience shorter lifespans compared to those in better-graded neighborhoods. The PM2.5 graph shows that block D neighborhoods have the highest particulate matter pollution levels. The population in redlined areas face greater exposure to air pollution than those in other neighborhoods.

These patterns align with existing research showing that redlined neighborhoods typically have less green space and higher temperatures than other areas.

- b. Discuss potential relationships between historical redlining grades and current environmental/ socioeconomic conditions.

Answer

The analysis reveals a strong relationship between historical redlining practices and present-day environmental and socioeconomic conditions. Communities in formerly redlined neighborhoods (Grade D) experience multiple disadvantages: higher rates of poverty, reduced life expectancy, and increased exposure to air pollution. These areas often lack adequate infrastructure, healthcare facilities, and green spaces that are more common in higher-graded neighborhoods. This pattern demonstrates how discriminatory housing policies from the 1930s continue to affect environmental and health outcomes today. The legacy of redlining has created lasting inequalities in access to environmental quality and economic opportunity, disproportionately impacting low-income communities.

Filter and transform data

Legacy of redlining in biodiversity observations

Create a visualizations that shows:

- a. The percentage of bird observations within each HOLC grade
- b. Include an appropriate title, axis labels, and legend

Hints: Ensure the bird observations and HOLC dataset have matching CRS’, then perform a spatial join to assign each bird observations to a corresponding HOLC grade.

Spoiler alert!! Our results don't match the findings from Ellis-Soto et al. 2023! Read the abstract of the study. Why might we have obtained different results in our analysis? What did the paper consider that we did not?

```
# Check CRS of the dataset
if(st_crs(map_ineq_transform) == st_crs(la_birds)){
  print("Coordinate reference systems match!")
} else {
  warning("Update coordinate reference systems to match!")
}
```

Warning: Update coordinate reference systems to match!

```
# Transform the CRS of a dataset
la_birds_transform <- st_transform(la_birds, crs = st_crs(map_ineq_transform))
```

Spatial join

```
# Spatial join between bird observation and HOLC grade
holc_birds <- st_join(map_ineq_transform, la_birds_transform, join = st_intersects) %>%
# Drop geometry
  st_drop_geometry()
```

```
# Calculate the percentage of bird observations within each HOLC grade
holc_birds_percent <- holc_birds %>%
  filter(grade != 'NA', # Filter out the NA
         year >= 2021 & year <= 2023) %>% # Filter for the year from 2021 - 2023
  group_by(grade, year) %>%
  summarise(count = n()) %>%
  ungroup() %>%
  group_by(year) %>%
  mutate(percentage = (count / sum(count)) * 100) %>%
  ungroup()
```

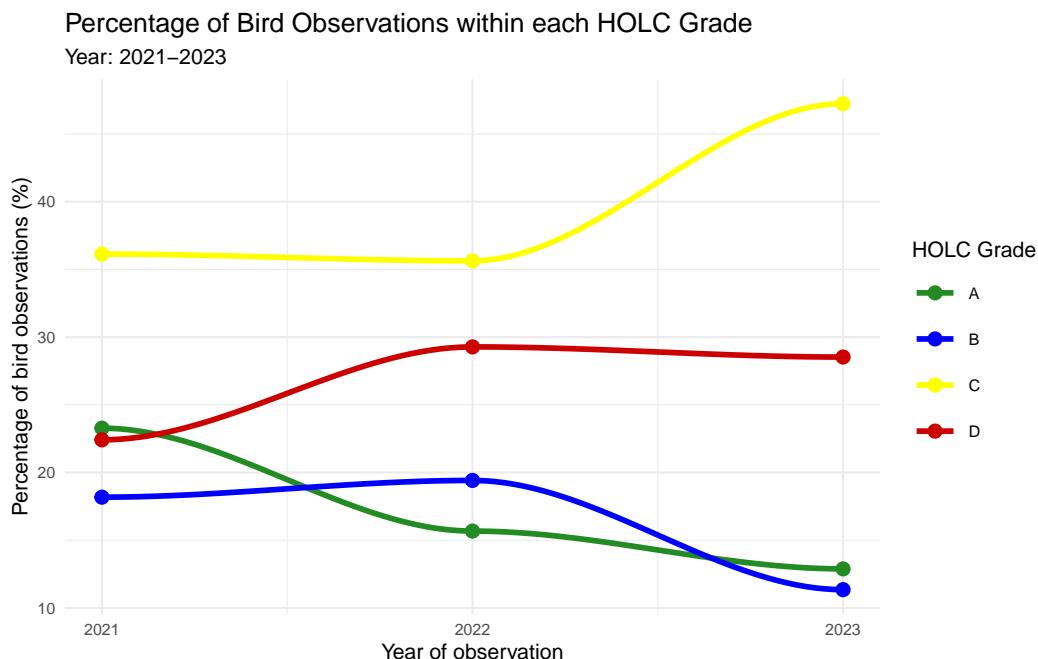
Data visualization

```
# Visualization of percentage of bird observations within each HOLC grade
ggplot(holc_birds_percent, aes(x = year,
                                y = percentage,
                                color = grade,
                                group = grade)) +
```

```

geom_line(linewidth = 1, stat = "smooth", method = "loess", se = FALSE) +
  geom_point(size = 2) +
  scale_color_manual(values = c('A' = 'forestgreen',
                                'B' = 'blue',
                                'C' = 'yellow1',
                                'D' = 'red3')) +
  scale_x_continuous(breaks = c(2021, 2022, 2023)) +
  labs(title = 'Percentage of Bird Observations within each HOLC Grade',
       subtitle = 'Year: 2021-2023',
       x = 'Year of observation',
       y = 'Percentage of bird observations (%)',
       color = 'HOLC Grade') +
  theme_minimal(base_size = 8)

```



```

# Save the plot of summary of EJScreen data
ggsave(here::here('figs', 'birds-obs-la.jpg'),
        width = 10,
        height = 8,
        dpi = 300)

```

Description of pattern

Why might we have obtained different results in our analysis? What did the paper consider that we did not?

The figure above shows the patterns in bird observation across the study blocks. The Block C has the highest number of bird observations throughout the year 2021 to 2023. In 2021, Blocks A and D had remarkably similar observation frequencies, suggesting comparable habitat quality or observer effort in these locations. However, a notable divergence can be noticed in the year 2022 and 2023. Block D showed a substantial increase in bird observations, steadily surpassing both Block A and Block B. This increasing trend in Block D indicates the potential changes in habitat conditions, seasonal bird population shifts and or variations in survey intensity across these locations over time.

The referenced paper gives a comprehensive relationship between historic racially explicit zoning policies, specifically redlining, and contemporary biodiversity data collection patterns. This study explored how the historic racially explicit zoning policies (redlining) relate to biodiversity data collection across and within 195 metropolitan areas in the United States covering more than 30 million population across 38 states. The study predominantly covers the eastern region, where historical redlining was extensively documented and where the majority of the analyzed metropolitan areas are located. In contrast, our bird observation data exploration is concentrated exclusively on Los Angeles County, the western region. This geographic location allows for deeper contextual understanding of local environmental justice issues, community biodiversity patterns and the historical trajectory of discriminatory housing policies in California. While the paper offers county level analysis and our analysis is one place based approach.