

Homework 2: Redlining in LA County

Sofia Sarak

2025-10-18

The following document contains visualizations relating HOLC grades and environmental indicators (potentially revealing environmental injustices) in Los Angeles County, California.

Github README

EDS223 Homework 2: Redlining in LA County

This repository contains the response to Homework 2 in EDS223 - *Geospatial Analysis & Remote Sensing*, completed by Sofia Sarak.

The assignment explores historically redlined districts in Los Angeles County, California. It incorporates spatial data on the "residential security" grades assigned by the Home Owners' Loan Corporation (HOLC) about eighty years ago. More information on the economic and racial segregation that came from this grading system, and redlining, can be found [here](#).

The analysis in this repository includes a spatial visualization of HOLC grades across LA, visualizations comparing environmental justice indicators across the grades, and a visualization exploring bird biodiversity and percentages associated with observation contribution, also by grade. These results are then summarized and reflected on in responses.

For more information on the homework assignment itself, reference the [assignment description](#).

Data Source

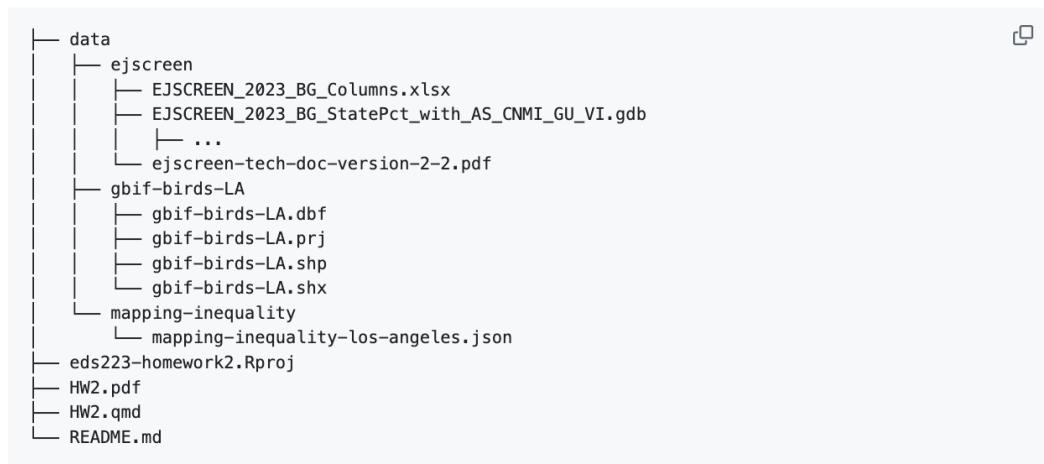
Our environmental justice data comes from the United States Environmental Protection Agency's former EJSscreen: Environmental Justice Screening and Mapping Tool. This tool is no longer operating, but an [unofficial version](#) is still running. EJSscreen provided environmental and demographic information for the US at the Census tract and block group levels; in this analysis, block group level data was used and was downloaded from the [EPA site](#).

Digitized maps and information on the HOLC classification system were created by the [Digital Scholarship Lab](#) at the University of Richmond as part of the [Mapping Inequality project](#) (and was downloaded from their [website](#)).

Biodiversity data was sourced from the [Global Biodiversity Information Facility](#), which is the largest aggregator of biodiversity observations in the world. The data set typically includes the species and the location and date that it was observed. This analysis uses data from 2021 onward.

Information on data sources was retrieved from original assignment description.

Repository Structure



Course Information

- **Course Title:** [EDS 223 - Geospatial Analysis & Remote Sensing](#)
- **Term:** Fall 2025
- **Program:** [UCSB Masters in Environmental Data Science.](#)

Teaching Team:

- **Instructor:** [Annie Adams](#)
- **Teaching Assistant:** Alessandra Vidal Meza

Complete materials for the discussion sections and additional resources can be found on the [course website](#).

This README was adapted from the README template provided in EDS220; see course details and original repository [here](#).

Initial Setup

```
#|output: false
#|message: false
# Loading in necessary libraries
library(here)
library(tidyverse)
library(sf)
library(stars)
library(tmap)
```

```

library(gt)
library(testthat)

# Loading in data
ineq <- st_read(here('data/mapping-inequality/mapping-inequality-los-angeles.json'), quiet =
birds <- st_read(here('data/gbif-birds-LA/gbif-birds-LA.shp'), quiet = TRUE)

```

Exploring and Transforming CRS'

```

# Checking to see if CRS' of each data set are the same
if (st_crs(ineq) == st_crs(birds)){
  print("Same CRS")
} else {
  stop("Different CRS")}

```

[1] "Same CRS"

```
# the two data sets have the same CRS, yay!
```

Part 1: Legacy of redlining in current environmental (in)justice

1. Create a map of historical redlining neighborhoods, including:

- neighborhoods colored by HOLC grade
- an appropriate base map
- a clear title and legend

```

# Loading in base map data
ejscreen <- st_read(here("data", "ejscreen","EJSCREEN_2023_BG_StatePct_with_AS_CNMI_GU_VI.gdb"))

# Filtering data for only LA County
la <- ejscreen %>%
  filter(ST_ABBREV == "CA") %>%
  filter(CNTY_NAME %in% c("Los Angeles County"))

# Checking CRS before plotting
if (st_crs(ineq) == st_crs(la)){
  print("Same CRS")
} else {
  warning("Different CRS")}

```

```
# CRS' don't match; transforming la's CRS to match ineq
la <- st_transform(la, crs = st_crs(ineq))

# And confirming it worked
if (st_crs(ineq) == st_crs(la)){
  print("Same CRS")
} else {
  stop("Different CRS")}
```

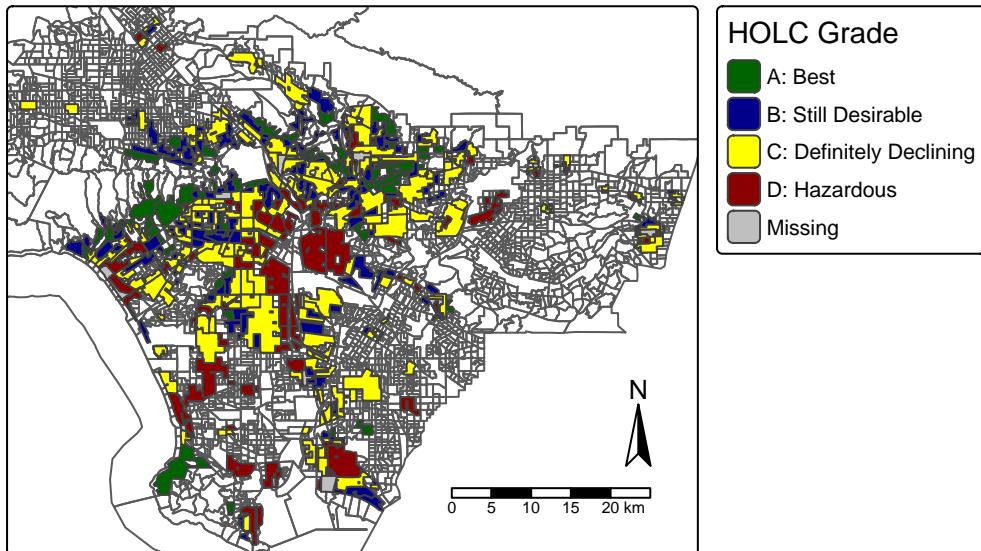
[1] "Same CRS"

```
# Creating a custom color theme for map, as well as for following plots
# With each grade corresponding to the color defined/typically used by HOLC
grade_colors <- c("A" = "darkgreen",
                 "B" = "darkblue",
                 "C" = "yellow",
                 "D" = "darkred",
                 "No Grade" = "gray")

# Also creating grade label variable with HOLC grade meaning
grade_labels <- c("A: Best",
                  "B: Still Desirable",
                  "C: Definitely Declining",
                  "D: Hazardous",
                  "No Grade")
```

```
# Creating a map of historical redlining neighborhoods
tm_shape(la, bbox = ineq) +      # setting bounding box to match the ineq data
  tm_borders(col = "gray35") +
tm_shape(ineq) +
  tm_polygons(
    fill = "grade",           # coloring by HOLC grade
    fill.legend = tm_legend(title = "HOLC Grade"),
    fill.scale = tm_scale_ordinal(values = grade_colors,
                                   labels = grade_labels)) +
  tm_title(text = "Historical Redlining Neighborhoods, Los Angeles") +
  tm_compass() +            # adding North arrow
  tm_scalebar()             # adding scale bar
```

Historical Redlining Neighborhoods, Los Angeles



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 EJSscreen 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()`.

```
# Attempting to combine the two data sets
# table <- st_join(ineq, la)

# Oh no! An error!
# After researching error, it appears that it is telling me that one of my data frames in "in"

# Checking validity of both data frames
st_is_valid(la) %>%
  table() # Using table() to create an output that counts up the TRUEs and FALSEs
```

```
TRUE
6591
```

```
st_is_valid(ineq) %>%
  table()
```

```
FALSE  TRUE
12    405
```

```
# The ineq data frame is the one with invalid elements!
```

```
# Making all of ineq valid
ineq <- st_make_valid(ineq)

# And checking that it worked
st_is_valid(ineq) %>%
  table()
```

```
TRUE
417
```

```
# Joining the two data frames
table <- st_join(ineq, la) %>%
  st_drop_geometry() # and dropping geometry
```

```
# The "NA" values in the grade column represent census block groups that don't fall within a
# To make this more clear in following visualizations, I change "NA" to "No Grade"
table$grade[is.na(table$grade)] <- "No Grade"
```

```
# Ensuring that the NA values got changed using a testthat function
expect_true("No Grade" %in% table$grade)
```

```
# Creating summary table
summ_table <- table %>%
  group_by(grade) %>%
  summarize(percentage = round((n()/length(table$ID))*100, 2)) %>% # calculating % and round
  # using gt package to make
  gt() %>%
  cols_label(
    grade = "HOLC Grade",
    percentage = "%") %>%
  tab_header( # changing column names
```

[

Los Angeles HOLC Grade Composition

Represented by Percentage of Census Block Groups] Los Angeles
HOLC Grade Composition

Represented by	Percentage of	Census	Block	Groups
HOLC Grade				%
A				7.03
B				19.40
C				47.87
D				21.07
No Grade				4.63

```
title = "Los Angeles HOLC Grade Composition", # adding a title
subtitle = "Represented by Percentage of Census Block Groups")

# Calling the table
summ_table
```

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):

- % low income
- percentile for Particulate Matter 2.5
- percentile for low life expectancy
- Use `ggplot` for your visualizations! You will first need to calculate mean of each variable grouped by HOLC grade.

Visualization 1

```
# Calculating mean of % low income
low_inc_mean <- table %>%
  group_by(grade) %>%
  summarize(low_inc = mean(LOWINCPCT*100)) # *100 because low income is represented as a prop

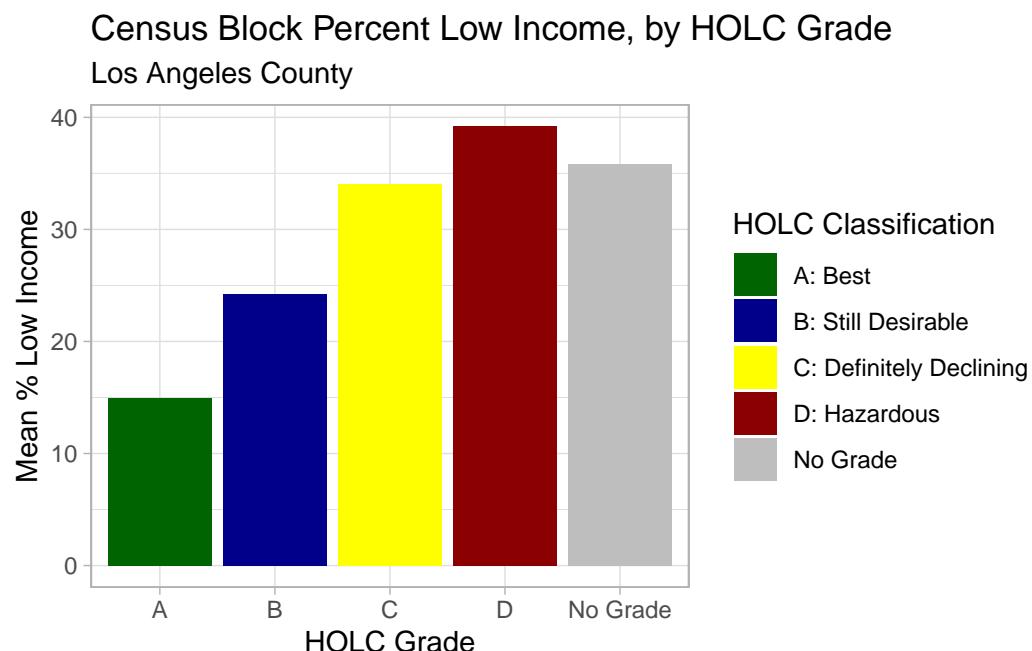
# Replacing "NA" to "No Grade" again, for visualization purposes
low_inc_mean$grade[is.na(low_inc_mean$grade)] <- "No Grade"
```

```

# Ensuring that the NA values got changed using a testthat function
expect_true("No Grade" %in% low_inc_mean$grade)

# Creating a bar plot
ggplot(data = low_inc_mean, aes(x = grade, y = low_inc)) +
  geom_col(aes(fill = grade)) +                                     # coloring by HOLC Grade
  scale_fill_manual(values = grade_colors,
                    labels = grade_labels) + # removing legend bc it is redundant
  theme_light() +
  labs(title = "Census Block Percent Low Income, by HOLC Grade",
       subtitle = "Los Angeles County",
       x = "HOLC Grade",
       y = "Mean % Low Income",
       fill = "HOLC Classification")

```



Visualization 2

```

# Calculating mean of particulate matter 2.5 percentile
pm_mean <- table %>%
  group_by(grade) %>%
  summarize(part_matt = mean(P_D5_PM25))

```

```

# Calculating mean of low life expectancy percentile
life_exp_mean <- table %>%
  group_by(grade) %>%
  summarize(life_exp = mean(P_LIFEEXPCT, na.rm = TRUE))

# Joining the two summaries
percentiles <- full_join(pm_mean, life_exp_mean, by = "grade")

# Pivoting longer, to be able to facet in my visualization
percentiles_long <- pivot_longer(percentiles, cols = part_matt:life_exp, names_to = "vars", values_to = "mean")

# Replacing "NA" to "No Grade" again, for visualization purposes
percentiles_long$grade[is.na(percentiles_long$grade)] <- "No Grade"

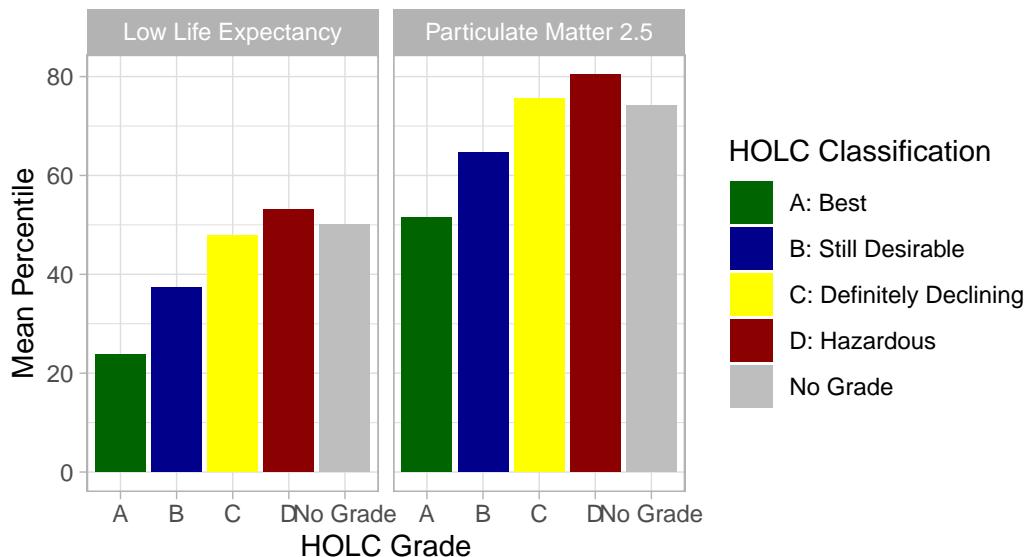
# Ensuring that the NA values got changed using a testthat function
expect_true("No Grade" %in% percentiles_long$grade)

# Creating facet labels, for cleaner plot headings
facet_labels <- c(
  "life_exp" = "Low Life Expectancy",
  "part_matt" = "Particulate Matter 2.5")

# Creating a bar plot, faceting by percentile variable
ggplot(data = percentiles_long, aes(x = grade, y = mean)) +
  geom_col(aes(fill = grade)) +
  facet_wrap(~vars, labeller = as_labeller(facet_labels)) +      # facetting and specifying p
  scale_fill_manual(values = grade_colors,
                    labels = grade_labels) +
  theme_light() +
  labs(title = "Environmental Justice Indicator Percentiles, by HOLC Grade",
       subtitle = "Los Angeles County",
       x = "HOLC Grade",
       y = "Mean Percentile",
       fill = "HOLC Classification")

```

Environmental Justice Indicator Percentiles, by HOLC Grade Los Angeles County



4. Write a brief paragraph reflecting on these results

- Interpret the patterns you observe in your results
- Discuss potential relationships between historical redlining grades and current environmental/socioeconomic conditions

Response:

There appears to be a clear pattern across all three of the analyses, in which “highest” HOLC grade (A) has the lowest percent low income, percentile for Particulate Matter 2.5, and percentile for low life expectancy. And without exception, the subsequent grades (B, C, and D) showed sequential increases in these values. Census block groups categorized as having a HOLC grade of D had the highest percentage/percentile across all three categories. Interestingly, the “No Grade” category was always slightly lower than grade D, but stayed in the same range as grade C.

Our results corroborate two historically-observed relationships:

1. Communities that were “redlined” tended to be those of a targeted, discriminated populations (for example, predominantly black). The relationship between HOLC grades and low income % supports this.
2. Low-income/redlined communities face environmental injustice; our analyses specifically highlight air quality, and also relate living conditions to low life expectancy.

Ultimately, our results suggest that areas deemed most “hazardous” by HOLC tend to be disadvantaged communities, facing socioeconomic and environmental justice issues. Redlining creates racial discrimination, keeping areas in need from receiving services that would make their living conditions equal to that of the grade A communities around them.

Part 2: Legacy of redlining in biodiversity observations

You must produce the following based on observations from 2021-2023:

1. A figure summarizing the percent of observations within redlined neighborhoods within each HOLC grade

Create a visualizations that shows:

- the percentage of bird observations within each HOLC grade
- 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.

```
# Checking CRS before plotting
if (st_crs(ineq) == st_crs(birds)){
  print("Same CRS")
} else {
  stop("Different CRS")}
```

```
[1] "Same CRS"
```

```
# Performing spatial join
biodiv <- st_join(ineq, birds) %>%
  st_drop_geometry()
```

```
# Observations for visualization must be from 2021-2023, so checking if all data is in range
unique(biodiv$year)
```

```
[1] 2021 2022 2023    NA
```

```
# Noticed that there were NA values, so dropped those as they won't be helpful for the visualization
biodiv <- biodiv[!is.na(biodiv$year),]
```

```
# Ensuring that our drop worked using a testthat function
expect_false("NA" %in% biodiv$year)
```

```
# Calculating percent of bird observations in each HOLC grade
bird_perc <- biodiv %>%
  group_by(grade) %>%
  summarize(percent = round(n()/length(biodiv$year)*100, 2))

# Viewing summary
bird_perc
```

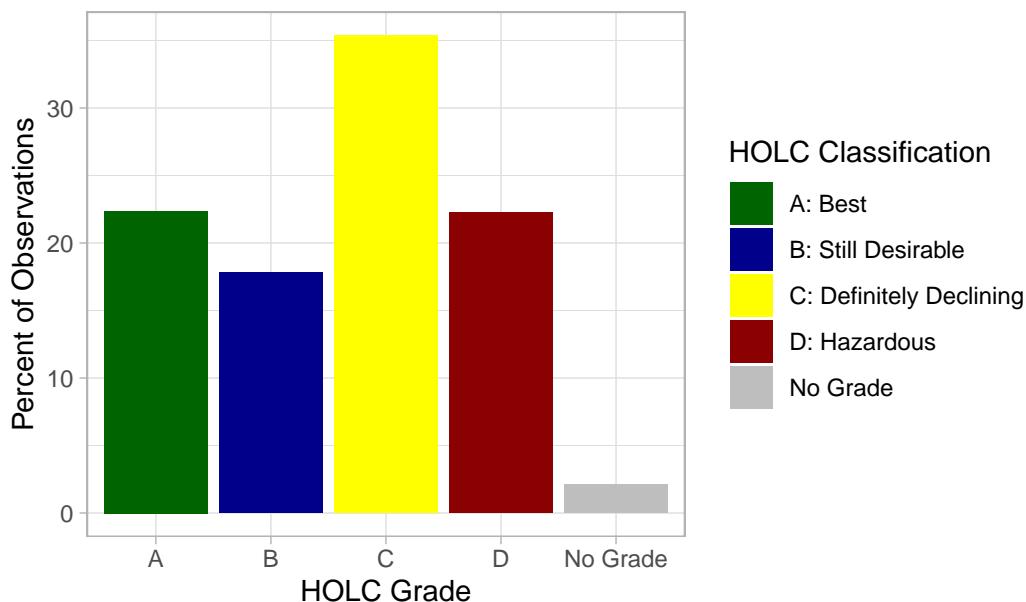
```
# A tibble: 5 x 2
  grade percent
  <chr>   <dbl>
1 A        22.4
2 B        17.8
3 C        35.4
4 D        22.3
5 <NA>     2.14
```

```
# Replacing "NA" to "No Grade" again, for visualization purposes
bird_perc$grade[is.na(bird_perc$grade)] <- "No Grade"

# Ensuring that the NA values got changed using a testthat function
expect_true("No Grade" %in% bird_perc$grade)
```

```
# Plotting a bar graph using our summary
ggplot(data = bird_perc, aes(x = grade, y = percent)) +
  geom_col(aes(fill = grade)) +
  scale_fill_manual(values = grade_colors,
                    labels = grade_labels) + # using the same custom colors as previous plot
  theme_light() +
  labs(title = "Bird Observations Across HOLC Grades",
       x = "HOLC Grade",
       y = "Percent of Observations",
       fill = "HOLC Classification")
```

Bird Observations Across HOLC Grades



2. 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?

Response:

Our results appear to suggest that the highest percentage of bird observations in LA County were observed in census block groups categorized as HOLC grade of C, or "definitely declining." The other three grades – A, B, and D – have slight variation, but all fall generally in the same range. In fact, the percentage of bird observations in grade A ("best") is only 0.07 percentage points higher than that of grade D ("hazardous").

Findings from Ellis-Soto et al. 2023, on the other hand, suggest a clear pattern between HOLC grade and the percentage of bird observations recorded. They saw that areas which were deemed "desirable" (grades A and B) to have more than twice the density of observations than areas that were redlined, or deemed "hazardous" (such as grade D). As such, our results differ.

It is important to consider, however, the breadth of the analyses. To quote the abstract from Ellis-Soto et al. 2023: "After accounting for differences in present day vegetation, open space, population density and climate, we find significantly lower sampling density and sampling completeness in these redlined neighborhoods." It appears that because our analysis *only* considered HOLC grade and percentage of bird observations, we lost out on the impact of confounding variables, such as those listed within the abstract. The findings of Ellis-Soto et al. 2023 are controlled for significant components of bird observation frequency that we did not

include, which explains why we have obtained different results (and suggests that their results are way more trustworthy than ours).