

EDS223 hw2

Megan Hessel

2025-10-28

EDS 223: Homework 2

Exploring Patterns of Environmental Justice

Part 0- Initial Downloading & CRS manipulation

```
# Import libraries
library(sf) # for handling vector data
library(tmap) # for making maps
library(tidyverse) # because we love the tidyverse
library(spData)
library(paletteer)
library(patchwork)
library(knitr)

# Import Data
ejscreen <- read_sf(here::here('data', 'ejscreen', 'EJSCREEN_2023_BG_StatePct_with_AS_CNMI_G

# Imports the same data
#birds <- read_sf(here::here('data', 'gbif-birds-LA'))
bird <- read_sf(here::here('data', 'gbif-birds-LA', 'gbif-birds-LA.shp'))

ineq <- read_sf(here::here('data', 'mapping-inequality', 'mapping-inequality-los-angeles.json'))
```

Coordination Reference System (CRS) Check and Transformation

Checking the CRSs

```
# Checking crs of all data
st_crs(ejscreen) # WGS 84 / Pseudo-Mercator
st_crs(bird) # WGS 84
st_crs(ineq) # WGS 84

st_crs(ejscreen) == st_crs(bird) # FALSE
st_crs(ineq) == st_crs(bird) # TRUE
st_crs(ineq) == st_crs(ejscreen) # FALSE

# NEEDS TO CHANGE EJSCREEN TO FIT BIRD AND INEQ DF
```

Transforming the CRSs to match

```
# Transform ejscreen data
ej_trans <- st_transform(ejscreen, crs = st_crs(ineq))

# Check to see all df's crs changed
st_crs(ej_trans) == st_crs(bird) # TRUE
```

[1] TRUE

Final CRS match check

```
# Final check crs match!

if (st_crs(ej_trans) == st_crs(bird)) {
  print("its a match")
} else {
  print("its not a match") }
```

[1] "its a match"

```
if (st_crs(bird) == st_crs(ineq)) {
  print("its a match")
} else {
  print("its not a match") }
```

```
[1] "its a match"

if (st_crs(ineq) == st_crs(ej_trans)) {
  print("its a match")
} else {
  print("its not a match") }
```

```
[1] "its a match"
```

Filtering the EJscreen data for just Los Angeles, CA

```
# filter to a state you are interested in: Louisiana
CA <- ej_trans %>%
  dplyr::filter(ST_ABBREV == "CA")

# filter to a county you are interested in: East Baton Rouge Parish
CA_LA <- CA %>%
  dplyr::filter(CNTY_NAME %in% c("Los Angeles County"))
```

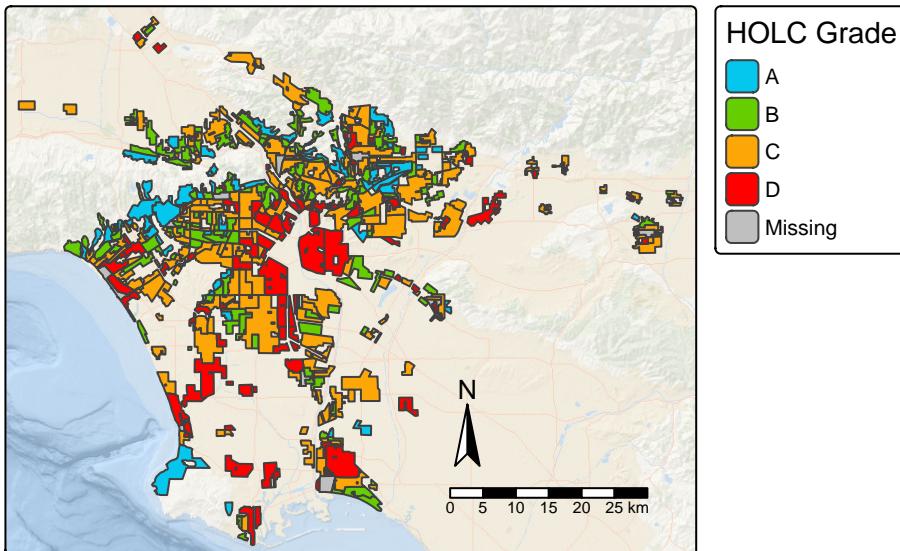
Part 1: Legacy of redlining in current environmental (in)justice

Historical redlining Map

```
tmap_mode("plot")

# Historical redlining Map
tm_shape(ineq, is.main = TRUE) +
  tm_polygons(
    fill = 'grade',
    title = "HOLC Grade",
    palette = c("#05c6ed", "#66CD00", "#fca608", "red") # Tried to make colors consistent for
  ) +
  tm_basemap("Esri.OceanBasemap") +
  tm_title("Home Owners Loan Corporation (HOLC) Redlining") +
  tm_compass(position = c("right", "bottom")) +
  tm_scalebar()
```

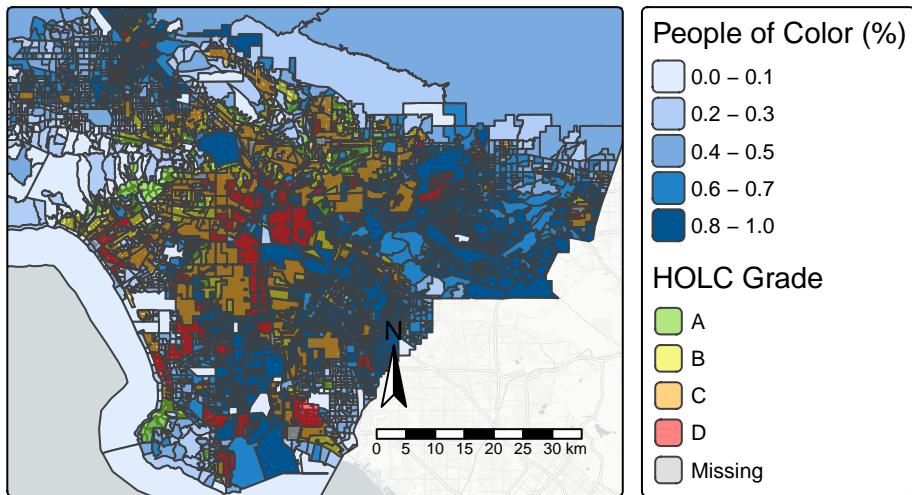
Home Owners Loan Corporation (HOLC) Redlining



```
# Historical redlining Map in correlation to people of color
# **** This is not necessary for the hw, but I was interested!!!****
tm_shape(CA_LA) +
  tm_polygons(fill = 'PEOPCOLORPCT',
              fill.legend = tm_legend(title = "People of Color (%)")) +
  tm_shape(ineq, is.main = TRUE) +
  tm_polygons(
    fill = 'grade',
    title = "HOLC Grade",
    palette = c("#66CD00", "yellow2", "#FFA500", "red"),
    alpha = 0.5
  ) +
  tm_basemap("CartoDB.PositronNoLabels") +
  tm_title(
    "Home Owners Loan Corporation (HOLC) Redlining\nOverlaid with People of Color Population"
  ) +
  tm_compass(position = c("right", "bottom")) +
  tm_scalebar()
```

| ----- | ----- | ----- | ----- |
=====

Home Owners Loan Corporation (HOLC) Redlining
Overlaid with People of Color Populations



Summary Tables

Dealing with Geometry problems

```
# Solving invalid geometries

# Locating the invalid geometries
#st_is_valid(CA_LA) # Returns boolean operators
#st_is_valid(ineq)

which(!st_is_valid(CA_LA)) # Tell us which specific ones are invalid
```

```
integer(0)
```

```
which(!st_is_valid(ineq))
```

```
[1] 109 161 208 209 223 225 247 285 309 337 362 370
```

```
# Fixing them
CA_LA_st <- st_make_valid(CA_LA)
ineq_st <- st_make_valid(ineq)
```

Joining the Data

```
# Combining the Data
join_df <- st_join(x = ineq_st, y = CA_LA_st, join = st_intersects) %>%
  st_drop_geometry(join_df) # removing geometries
```

```
# Checking the dimensions
print(dim(CA_LA_st))
```

```
[1] 6591 224
```

```
print(dim(ineq_st))
```

```
[1] 417 15
```

```
print(dim(join_df))
```

```
[1] 6388 237
```

```
# Checking if geometry column was dropped
# colnames(join_df)
# head(join_df)
# join_df['grade']
```

Creating Summary Tables

(1) Table 1 of % of census block per HOLC grade

```
# Creating the summary df of % of census block groups per each HOLC grade
total_grade <- nrow(join_df["grade"]) # Total number of grades

sum_table <- join_df %>%
  group_by(grade) %>% # grouping by HOLC grade
  summarise(grade_perc = round(((n()/total_grade)*100), 1)) # count per grade / total

# Creating a pretty summary
summary_table <- kable(sum_table,
  col.names = c("HOLC Grade", "Block Census Percentage"),
  align = "c")

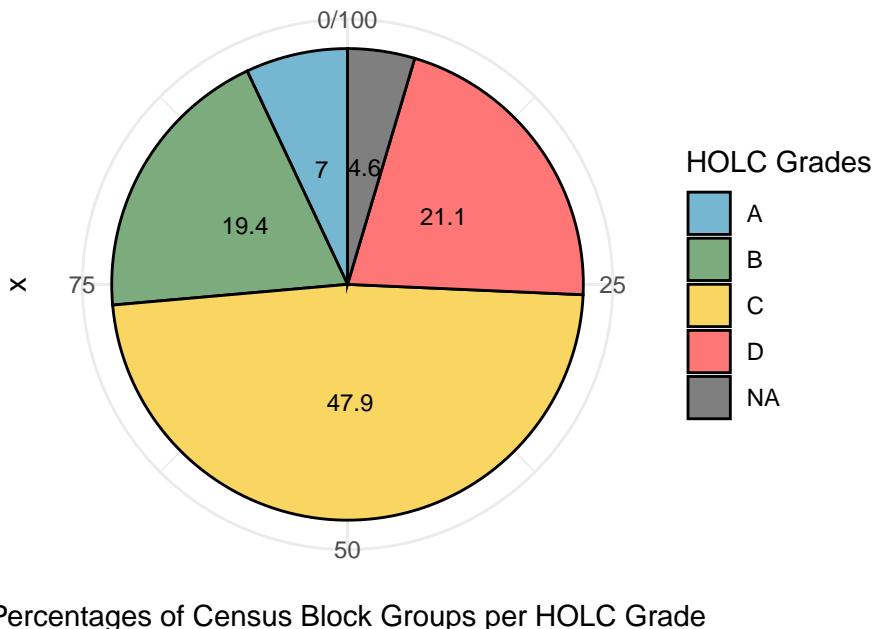
summary_table
```

HOLC Grade	Block Census Percentage
A	7.0
B	19.4
C	47.9
D	21.1
NA	4.6

Pie Chat of Census Block Percentages per HOLC grades

(I made this, thinking this was necessary for the hw... I worked on this chart for a while, so I wanted to at least show it)

```
# Pie chart of Cencus Block Percentages per HOLC grades
ggplot(sum_table, aes(x = "", y = grade_perc, fill = grade)) +
  geom_col(color = "black") +
  geom_text(aes(label = grade_perc),
            position = position_stack(vjust = 0.5),
            size = 3) +
  coord_polar(theta = "y") +
  scale_fill_manual(values = paletteer_d("ltc::fernande", direction = -1)) +
  theme_minimal() +
  labs(y = "Percentages of Census Block Groups per HOLC Grade",
       fill = "HOLC Grades")
```



(2) A summary dataframe of all avgs for all columns per HOLC grade

```
# Another useful summary table of averages for all columns per HOLC grade
summary_table2 <- join_df %>%
  group_by(grade) %>%
  summarise(across(everything(), mean, na.rm = TRUE)) %>%
  select(grade, ACSTOTPOP:T_D5_PWDIS) # Selecting only neccessary columns

head(summary_table2)
```

```
# A tibble: 5 x 211
  grade ACSTOTPOP ACSIPOVBAS ACSEUCBAS ACSTOTHH ACSTOTHU ACSUNEMPBAS DEMOGIDX_2
  <chr>    <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
1 A        1290.     1186.     904.      486.      546.      664.      0.307
2 B        1339.     1315.     974.      517.      563.      741.      0.414
3 C        1481.     1455.     1032.     517.      559.      799.      0.535
4 D        1525.     1499.     1033.     502.      540.      805.      0.597
5 <NA>     1397.     1377.     1000.     566.      632.      771.      0.497
# i 203 more variables: DEMOGIDX_5 <dbl>, PEOPCOLOR <dbl>, PEOPCOLORPCT <dbl>,
#   LOWINCOME <dbl>, LOWINCPCT <dbl>, UNEMPLOYED <dbl>, UNEMPPCT <dbl>,
#   LINGISO <dbl>, LINGISOPCT <dbl>, LESSHS <dbl>, LESSHSPCT <dbl>,
#   UNDER5 <dbl>, UNDER5PCT <dbl>, OVER64 <dbl>, OVER64PCT <dbl>,
#   LIFEEXPCT <dbl>, PM25 <dbl>, OZONE <dbl>, DSLPM <dbl>, CANCER <dbl>,
```

```
#   RESP <dbl>, RSEI_AIR <dbl>, PTRAF <dbl>, PRE1960 <dbl>, PRE1960PCT <dbl>,
#   PNPL <dbl>, PRMP <dbl>, PTSDF <dbl>, UST <dbl>, PWDIS <dbl>, ...
```

Creating visualizations

GOAL: Summarize current conditions within HOLC grades using the means of % of low income, low life expectancy percentile, and Particulate Matter 2.5 percentile.

FIRST: Subset and rearrange data for plotting

Subset... The only columns we need: % of low income, low life expectancy percentile, Particulate Matter 2.5 percentile, and HOLC grade

```
# Subset for necessary columns (CONTAINS ALL values)
sub_df <- join_df %>%
  select("LOWINCPCT", "grade", "P_PM25", "P_LIFEEXPCT") %>%
  mutate(LOWINCPCT = LOWINCPCT * 100)

# Pivot Longer (contains all values)
sub_long <- pivot_longer(sub_df, # Pivot Longer
                         cols = c(LOWINCPCT, P_PM25, P_LIFEEXPCT),
                         names_to = 'means',
                         values_to = 'values') %>%
  mutate(means = recode(means, # Changing names of conditions
                        "P_LIFEEXPCT" = "Low Life Expectancy",
                        "LOWINCPCT" = "Low Income",
                        "P_PM25" = "PM2.5"))

head(sub_df)
```

```
# A tibble: 6 x 4
  LOWINCPCT grade P_PM25 P_LIFEEXPCT
  <dbl> <chr>   <int>      <int>
1    27.9 A        66        42
2    16.1 A        65         4
3    21.1 A        65         4
4    10.2 A        65         4
5    15.9 A        64        12
6    19.0 A        63        39
```

```

# -----
# CONTAINS JUST AVERAGES

# Subset for neccary columns (contains avg of values)
summary_sub_df <- summary_table2 %>% # This summary dataframe was created in "Summarizing Ta
  select("LOWINCPCT", "grade", "P_PM25", "P_LIFEEXPCT") %>%
  mutate(LOWINCPCT = LOWINCPCT * 100)

# Pivot Longer the summary plot
summary_long <- pivot_longer(summary_sub_df,
  cols = c("LOWINCPCT", "P_PM25", "P_LIFEEXPCT"),
  names_to = 'means',
  values_to = 'values') %>%
mutate(means = recode(means, # Changing names of conditions
  "P_LIFEEXPCT" = "Low Life Expectancy",
  "LOWINCPCT" = "Low Income",
  "P_PM25" = "PM2.5"))

summary_sub_df

# A tibble: 5 x 4
  LOWINCPCT grade P_PM25 P_LIFEEXPCT
  <dbl> <chr>   <dbl>      <dbl>
1     15.0 A       72.2      23.8
2     24.2 B       76.3      37.4
3     34.1 C       78.8      47.9
4     39.2 D       80.2      53.0
5     35.8 <NA>    76.7      50.1

```

Now that we have rearranged all the data, we can make our visualizations!

(1) Map Visualizations

(This map's pdf output is ugly-ish and not too understandable, but it's pretty on the computer and if I was to download/tmap_save it!)

```

# FIGURE 1: MAP

# HOLC map
CA_holc <- tm_shape(CA_LA) + # Wanted to outline LA blocks for easy comparsion

```

```

tm_polygons(lwd = 0.1) +
tm_shape(ineq, is.main = TRUE) + # Overlaid with HOLC data ontop
tm_polygons(
  lwd = 0.6,
  fill = 'grade',
  fill.legend = tm_legend(title = "HOLC Grades"),
  fill.scale = tm_scale(values = c("#05c6ed", "#66CD00","#fca608","red"))
) +
tm_basemap("Esri.OceanBasemap") +
tm_title(text = "Home Owners Loan Corporation (HOLC) Grades") +
tm_compass(position = c("right", "bottom"))

# CA Low income
CA_lowincome <-
tm_shape(CA_LA) +
tm_polygons(
  fill = 'LOWINCPCT',
  lwd = 0.15,
  fill.legend = tm_legend(title = "Low Income\n(%)")
) +
tm_basemap("Esri.OceanBasemap") +
tm_title(text = "Low income Percentage") +
tm_shape(ineq, is.main = TRUE)

# Percentile for Particulate Matter 2.5
CA_pm25 <- tm_shape(CA_LA) +
tm_polygons(fill = 'P_PM25',
            lwd = 0.2,
            fill.legend = tm_legend(title = "PM 2.5 (%)")) +
tm_basemap("Esri.OceanBasemap") +
tm_title(text = "Percentile for Particulate Matter 2.5") +
tm_shape(ineq, is.main = TRUE) +
tm_scalebar()

# Percentile for life expectancy
CA_lifeexp <- tm_shape(CA_LA) +
tm_polygons(
  fill = 'P_LIFEEXPCT',

```

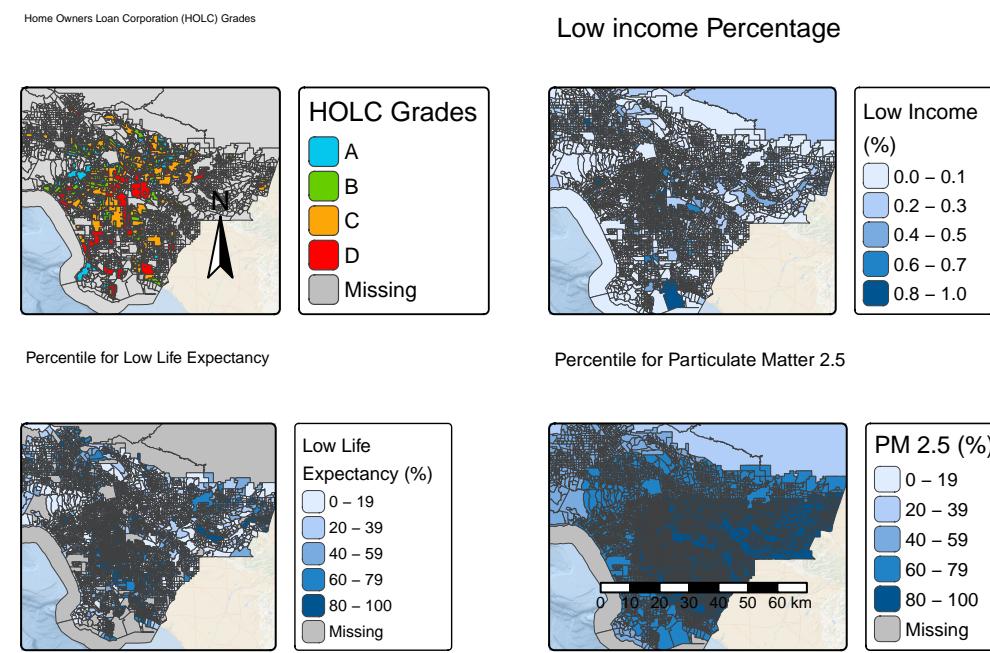
```

lwd = 0.2,
fill.legend = tm_legend(title = "Low Life\nExpectancy (%)")
) +
tm_basemap("Esri.OceanBasemap") +
tm_title(text = "Percentile for Low Life Expectancy") +
tm_shape(ineq, is.main = TRUE)

# Another good basemap for these: "CartoDB.PositronNoLabels"

tmap_arrange(CA_holc, CA_lowincome, CA_lifeexp, CA_pm25)

```



(2) Boxplot

```

# FIGURE 2: Boxplot

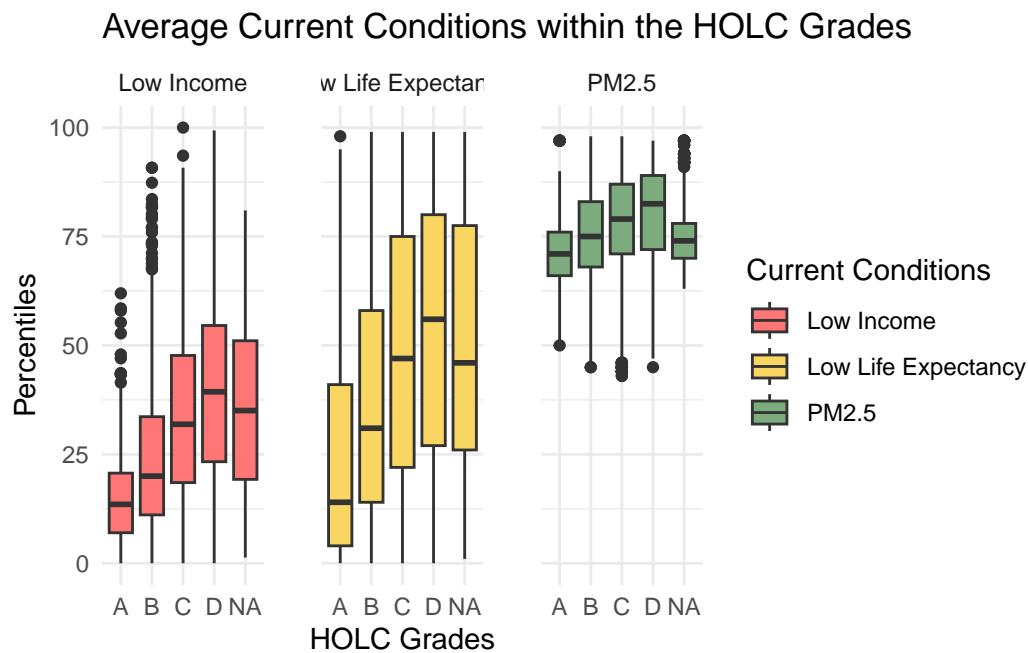
# Boxplot
ggplot(sub_long, aes(x = grade, y = values, fill = means)) + # Using pivoted df with ALL val
  geom_boxplot(stat = "boxplot") +
  facet_wrap(~ means) +
  scale_fill_manual(

```

```

values = paletteer_d("ltc::fernande") +
theme_minimal() +
theme(panel.spacing = unit(1.5, "lines")) +
labs(title = "Average Current Conditions within the HOLC Grades",
x = "HOLC Grades",
y = "Percentiles",
fill = "Current Conditions")

```



(3) Grouped Barchart

```

# Figure 3: Barchart

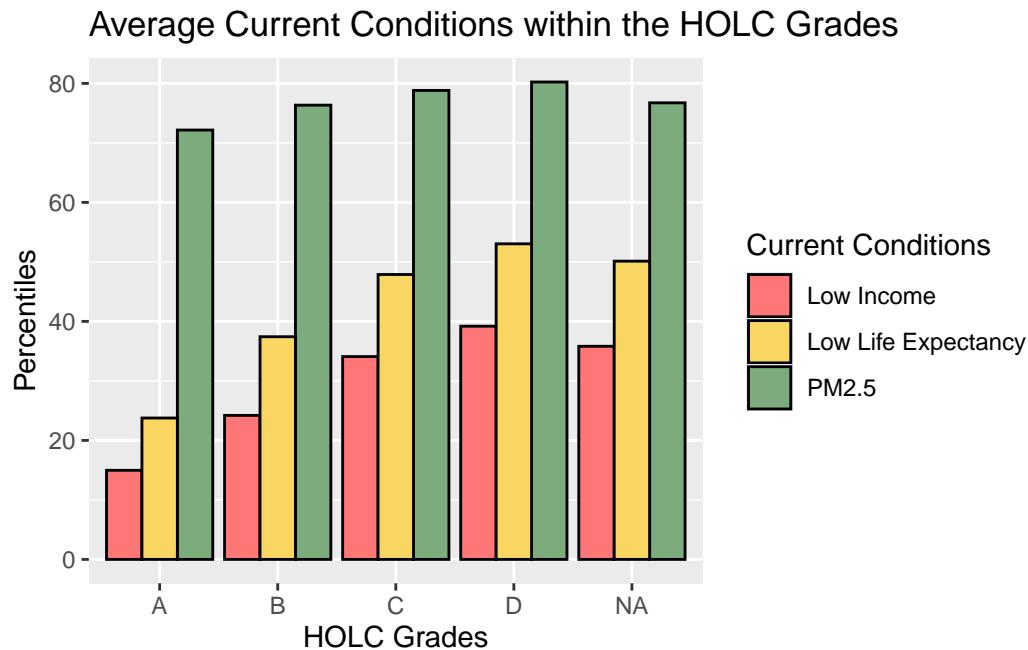
# Ggplot with all conditions
ggplot(summary_long, aes(y = values, x = grade, fill = means)) + # Using pivoted df of avg o
  geom_bar(col = "black", position = "dodge", stat = "identity") +
  scale_fill_manual(
    labels = c(
      "lowincome_mean" = "Low Income",
      "lexp_mean" = "Low Life Expentency",
      "pm2.5_mean" = "PM2.5"
    ),
    values = paletteer_d("ltc::fernande")

```

```

) +
labs(title = "Average Current Conditions within the HOLC Grades",
      x = "HOLC Grades",
      y = "Percentiles",
      fill = "Current Conditions") +
theme_grey(base_family = "sans")

```



Write a brief paragraph reflecting on these results

Seen in the figures, areas that have been given lower grades consistently have higher environmental and socioeconomic issues. Homes that have been graded “C” or “D” typically have higher percentiles of low life expectancy, low income, and slightly higher levels of Particulate Matter 2.5. This suggests that areas assigned lower historical HOLC grades tend to exhibit higher levels of present-day disadvantage.

Part 2: Legacy of redlining in biodiversity observations

Exploring the legacy of historical redlining in Los Angeles on the collection of bird observations from 2021-2023.

Data exploration & CRS checking

```
# Checking years within the bird df  
unique(bird$year)
```

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

```
# Rechecking CRS  
if (st_crs(CA_LA) == st_crs(bird)) {  
  print("its a match")  
} else {  
  print("its not a match") }
```

```
[1] "its a match"
```

```
if (st_crs(bird) == st_crs(ineq)) {  
  print("its a match")  
} else {  
  print("its not a match") }
```

```
[1] "its a match"
```

```
if (st_crs(ineq) == st_crs(CA_LA)) {  
  print("its a match")  
} else {  
  print("its not a match") }
```

```
[1] "its a match"
```

```
# Reminder: join_df does not have a geometry
```

Visualizations

Dealing with invalid Geometries

```

# Locating the invalid geometries
#st_is_valid(bird) # Returns boolean operators

which(!st_is_valid(bird)) # Where exactly they are

integer(0)

# Fixing them
bird <- st_make_valid(bird)

```

Checking if Points overlap

```

# Check if points overlap
if(any(st_intersects(bird, ineq_st, sparse = FALSE)) == TRUE) {
  print("Points overlap!")
} else {
  warning("Points don't overlap!")
}

```

```
[1] "Points overlap!"
```

Joining the Data

(Joining + grouping and summarizing)

```

# Join bird and ineq for all obsevations within the redlining locations
# st_join(ineq_st, bird, left = FALSE)
# st_join(ineq_st, bird, join = st_contains)

# Checking dimensions
#print(dim(bird))
#print(dim(ineq))
#print(dim(bird_join2))

# Looking at the plot
#tm_shape(bird_join2) +
#tm_dots()

# Join + groupby(grade) + summarise for percentages
bird_by_grade <- st_join(ineq_st, bird) %>%

```

```

st_drop_geometry() %>%
  mutate(total_rows = n()) %>% # Saving the total number of rows for summarise function
  group_by(grade) %>%
  summarise(perc_grade = round((n())/first(total_rows)*100, 2)) # summarising count / total

bird_by_grade

```

```

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

```

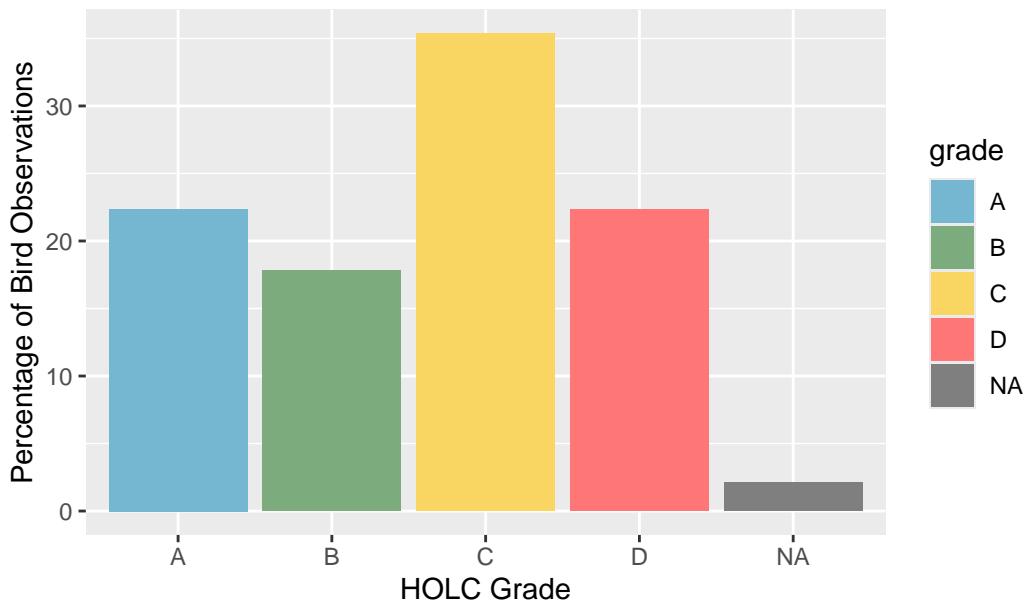
Bar plot

```

# Bar plot of Percent Bird Observations per HOLC Grade
ggplot(bird_by_grade,
       aes(x = grade,
            y = perc_grade,
            fill = grade)) +
  geom_col() +
  labs(title = "Percent Bird Observations per HOLC Grade",
       x = "HOLC Grade",
       y = "Percentage of Bird Observations") +
  scale_fill_manual(
    values = paletteer_d("ltc::fernande", direction = -1))

```

Percent Bird Observations per 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?

The Ellis-Soto et al. 2023 study found sampling bias across differing socioeconomic conditions. There was about double the bird observations in historically white areas in comparison to lower sampling density in redlined neighborhoods. This uneven citizen science reduces low socioeconomic communities' environmental understanding and limits the current knowledge about these redlined communities. Additionally, the sampling bias skews our understanding of bird occurrences.

Our results show that HOLC grade C has the most bird observations. Grades A and D have similar amount of bird observations, with grade B having the least number of bird observations. Therefore, there is no clear trend with HOLC grades and bird observations.

The 2023 study researched 195 US metropolitan areas, whereas we just looked at Los Angeles, CA area. Furthermore, unlike our study, the 2023 study investigated the survey completeness and found lower graded areas having incomplete survey data.