Redlining and Health Outcomes in Jackson, MS

Sarah Bolongaita, Hae In Kim, Christina Scaduto, Lily Shay

Overview and Motivation

Too often historical policies are dismissed as unimportant and no longer having an impact in the present day. Time and again, research and evidence have illustrated this to be untrue and we want to demonstrate this using data science tools obtained in this course.

Our project examines the relationship between redlining in Jackson, MS and various health indicators and outcomes. As a part of FDR's New Deal, a federal agency known as the Home Owners' Loan Corporation implemented the redlining practice in many cities across the United States. Neighborhoods were "evaluated" based on their "mortgage security" and areas that posed the lowest risks were graded "A" or "Best." Neighborhoods with the "highest risk" were graded "D" or "Hazardous." These evaluations were racist and inflicted even more structural violence in communities of color.

Under redlining it was impossible for people of color to receive mortgage loans to purchase homes in A/B graded neighborhoods ("Best," "Still Desirable"). In the US, owning property is considered the most significant way to generate intergenerational wealth and redlining prevented this for many families of color across the US. Evaluating areas with low property value also deflected investment in public goods, such as public schools that are typically paid for by property taxes. Additionally, businesses were also unable to receive insurance in C/D graded neighborhoods, perpetuating divestment and decreasing access to resources (i.e., health care systems, grocery stores, etc.).

In short, redlining was government-sanctioned segregation.

While the Fair Housing Act of 1968 made formal redlining illegal, it did not create a mechanism to undo past harms. The lack of investment in and divestment from these communities has largely gone uncorrected and continues to shape Jackson.

By comparing the HOLC grades that were assigned in the 1930s to health data from the past 10 years, we hope to illuminate that the harm redlining imposed on these neighborhoods is not a thing of the past, but lingers to the present day.

Related Work

The project stemmed from preliminary research that a group member, Lily Shay, conducted for a summer practicum. Her practicum focused specifically on the impact of redlining in Jackson, MS on food insecurity and relevant health outcomes.

This project builds on Lily's project by expanding the data set to include multi-year data and explore other the relationship between redlining and other health related measures.

Initial Questions

At the beginning of this project, we aimed to demonstrate the relationship between redlining and health outcomes (specifically related to food security). We also wanted to explore how redlining impacted structural factors such as healthy food availability which in turn impacts health outcomes for a population.

Finally, we were hoping to look at how redlining affected different populations differentially based on race/ethnicity. Other potential questions included looking at redlined areas and home valuation and area median income over time per census tract.

How the questions evolved

Our group initially ran into challenges pulling multi-year data from the City Health Dashboard. We turned to the 500 Cities Project data - the source of data that the Dashboard pulls from.

In addition to this, we had trouble parsing out demographic data related to the census tracts that we were looking at. Because of this, we decided to forgo our questions regarding the differential effect of redlining based on race/ethnicity and area median income.

After compiling multi-year data (2013-2017), we saw a list of 28 specific health measures available for the census tracts in Jackson, MS. This list informed what issues we wanted to explore. For example, there was data available for diabetes, high blood pressure and obesity, but no indicators for life expectancy or mental distress.

We conducted exploratory data analysis on these measures and their relationship to HOLC grades. We found that most of the associations were significant. With this in mind, we decided to focus on current lack of health insurance, obesity, diabetes, and mental health.

Data

Data Sources

Data was obtained from City Health Dashboard (https://www.cityhealthdashboard.com/) and the CDC's 500 Cities Project Data (https://www.cdc.gov/500cities/index.htm). The 500 Cities Project is a collaboration between the Robert Wood Johnson Foundation and the CDC Foundation. The 500 Cities Project provides city and census tract-level small area estimates for health outcomes.

The HOLC data and shapefiles were obtained from the University of Richmond's Mapping Inequality project (https://dsl.richmond.edu/panorama/redlining/#loc=5/39.1/-94.58).

For all graphs, HOLC grades are interpreted as follows: A: 'Best' B: 'Still Desirable' C: 'Definitely Declining' D: 'Hazardous'

Data Cleanup

Health measure data was acquired at a census tract-level for Jackson, MS from the 500 Cities Projects Data. Data from 4 different years were compiled into one data frame. Mismatched names between data frames were checked and addressed. Data frames were sorted by year, measure and ID.

Since HOLC graded areas do not perfectly coincide with census tracts, initial spatial analysis was done in ArcGIS to establish what percentage of the HOLC graded areas were within each census tract.

In our analysis, census tracts and corresponding data were treated as one of these grades if 20% or more of the land was graded a particular grade. For example, if census tract 401 encompassed 40% of a Grade A area, it coded as Grade A data. "No grade" areas were either not designated a grade in the 1930s or less than 20% of the census tract area had been graded. These areas are used as our 'control' group.

The two datasets (HOLC/spatial analysis and 500 Cities) were combined in a dataframe in two formats. The long frame includes each census tract in Jackson, MS while the short frame combined the grades to aggregate is a census tract held more positive HOLC grades (A/B) or negative HOLC grades (C/D).

Looking through the list of 28 measures, we realized that not all measures were collected every year. We chose 4 to focus on - health insurance, obesity, diabetes, and mental health. These indicators were collected in the same years - 2014, 2015, and 2017.

Loading Packages

```
# Creating a function for loading packages
ipak <- function(pkg){
    new.pkg <- pkg[!(pkg %in% installed.packages()[, "Package"])]
    if (length(new.pkg))
        install.packages(new.pkg, dependencies = TRUE)
    sapply(pkg, require, character.only = TRUE)
}


# Loading packages
packages <- c("dplyr", "DT", "foreign", "ggplot2", "ggthemes", "Gmisc", "Hmisc"
  , "janitor", "maditr",
            "readr", "readxl", "reshape", "rsconnect", "stringr", "tidyr")
ipak(packages)
```

```
##     dplyr         DT    foreign    ggplot2  ggthemes      Gmisc      Hmisc
##      TRUE       TRUE       TRUE       TRUE       TRUE       TRUE       TRUE
##   janitor     maditr      readr     readxl    reshape  rsconnect    stringr
##      TRUE       TRUE       TRUE       TRUE       TRUE       TRUE       TRUE
##     tidyr
##      TRUE
```

```
# Clearing environment
rm(list = ls())


# Setting working directory
```

Health Data

Compiling datasets published from 2016-2019

```
# Loading data for available years
df.2016 <- read_csv("500_Cities__Local_Data_for_Better_Health__2016_release.csv
  ")
df.2017 <- read_csv("500_Cities__Local_Data_for_Better_Health__2017_release.csv
  ")
df.2018 <- read_csv("500_Cities__Local_Data_for_Better_Health__2018_release.csv
  ")
df.2019 <- read_csv("500_Cities__Local_Data_for_Better_Health__2019_release.csv
  ")


# Comparing dfs
comparison <- compare_df_cols(df.2016, df.2017, df.2018, df.2019)


# Fixing mismatched names
df.2016 <- df.2016 %>% dplyr::rename("PopulationCount" = Population2010)
df.2017 <- df.2017 %>% dplyr::rename("PopulationCount" = Population2010)
df.2018 <- df.2018 %>% dplyr::rename("GeoLocation" = Geolocation)


# Saving original variable order
original <- names(df.2016)


# Sorting variable order alphabetically
sorted <- sort(names(df.2016))


# Sorting all df columns alphabetically in preparation of rbind
df_2016 <- df.2016 %>% select(sorted)
df_2017 <- df.2017 %>% select(sorted)
df_2018 <- df.2018 %>% select(sorted)
df_2019 <- df.2019 %>% select(sorted)


# Binding dfs
df <- rbind(df.2016, df.2017, df.2018, df.2019)


# Filtering data and removing rows with constant values
```

```r
df <- df %>%
  filter(GeographicLevel == "Census Tract",
         !str_detect(GeoLocation, "POINT"),
         !is.na(TractFIPS)) %>%
  remove_constant()


# Creating final 500Cities dataframes
`500cities.unspread` <- df %>%
  select(TractFIPS, Year, MeasureId, Short_Question_Text, Measure, Data_Value)
  %>%
  distinct()
`500cities` <- df %>%
  select(TractFIPS, Year, MeasureId, Data_Value) %>%
  distinct() %>%
  spread(MeasureId, Data_Value)


# Removing unneeded dfs
rm(comparison, df.2016, df.2017, df.2018, df.2019)
```

Creating a reference dataframe for measures

```r
# Generating metadata file  of measures
temp1 <- df %>%
  select(MeasureId, Short_Question_Text, Measure) %>%
  distinct() %>%
  arrange(MeasureId)


# Getting the years of data for each measure
temp2 <- df %>%
  select(MeasureId, Year) %>%
  group_by(MeasureId) %>%
  distinct() %>%
  arrange(MeasureId, Year) %>%
  summarise(Years = paste(Year, collapse = ", "))


# Joining dataframes
metrics <- left_join(temp1, temp2)


# Removing unneeded dfs
rm(df, temp1, temp2)
```

HOLC Grades Data

Combining HOLC data with spatial data

```r
# Loading HOLC data
df <- read_xlsx("MS_Data\ Summary.xlsx", sheet = 1)

# Selecting variables
df <- df %>%
  select(CensusTracts, holc_id, holc_grade) %>%
  dplyr::rename("TractFIPS" = CensusTracts,
                "HOLC_Id" = holc_id,
                "HOLC_Grade" = holc_grade)

# Loading spatial data to get percent of census tract land by grade
df2 <- read.dbf("Pct_Holc.dbf")

# Selectng variables
df2 <- df2 %>%
  select(tract2010, holc_id, PERCENTAGE) %>%
  mutate(tract2010 = as.numeric(as.character(tract2010)),
         PERCENTAGE = round(PERCENTAGE / 100, 6)) %>%
  dplyr::rename("TractFIPS" = tract2010,
                "HOLC_Id" = holc_id,
                "HOLC_Pct" = PERCENTAGE)

# Joining HOLC and spatial data
df3 <- left_join(df, df2, by = c("TractFIPS", "HOLC_Id"))
```

Combining HOLC/spatial data with 500 cities data: Extra long format (for dynamic Shiny plot)

```r
# Joining HOLC/spatial data with unspread 500 cities data
data_xlong <- left_join(`500cities.unspread`, df3)
data_xlong <- data_xlong %>%
  select(TractFIPS, HOLC_Grade, HOLC_Pct, Year, MeasureId, Short_Question_Text,
  Measure, Data_Value) %>%
  distinct()

# Defining exposure variable - Only including grades that composed more than 30
  % of the census tract's land
data_xlong <- data_xlong %>%
  mutate(HOLC_Pct2 = ifelse(HOLC_Pct > 0.3, round(HOLC_Pct * 100, 2), NA),
         HOLC_Grade2 = ifelse(is.na(HOLC_Pct2), NA, HOLC_Grade),
         HOLC_Grade3 = factor(ifelse(is.na(HOLC_Grade2), 1,
                                     ifelse(HOLC_Grade2 == "A", 2,
                                            ifelse(HOLC_Grade2 == "B", 3,
                                                   ifelse(HOLC_Grade2 == "C", 4
  ,
                                                          ifelse(HOLC_Grade2 ==
  "D", 5, NA))))),
                               levels = 1:5,
                               labels = c("No grade", "A", "B", "C", "D"))) %>%
  select(-HOLC_Grade, -HOLC_Pct, -HOLC_Grade2) %>%
  dplyr::rename("HOLC_Pct" = HOLC_Pct2,
         "HOLC_Grade" = HOLC_Grade3) %>%
  select(TractFIPS, HOLC_Grade, HOLC_Pct, Year, MeasureId, Short_Question_Text,
  Measure, Data_Value) %>%
  distinct()
```

Combining HOLC/spatial data with 500 cities data: Long format

```
# Joining HOLC/spatial data with 500 cities data
data_long <- left_join(`500cities`, df3)
data_long <- data_long %>%
  select(TractFIPS, Year, HOLC_Grade, HOLC_Pct, ACCESS2:TEETHLOST)

# Defining exposure variable - Only including grades that composed more than 30
  % of the census tract's land
data_long <- data_long %>%
  mutate(HOLC_Pct2 = ifelse(HOLC_Pct > 0.3, round(HOLC_Pct * 100, 2), NA),
         HOLC_Grade2 = ifelse(is.na(HOLC_Pct2), NA, HOLC_Grade),
         HOLC_Grade3 = factor(ifelse(is.na(HOLC_Grade2), 1,
                                   ifelse(HOLC_Grade2 == "A", 2,
                                       ifelse(HOLC_Grade2 == "B", 3,
                                           ifelse(HOLC_Grade2 == "C", 4
  ,
                                                    ifelse(HOLC_Grade2 ==
  "D", 5, NA))))),
                             levels = 1:5,
                             labels = c("No grade", "A", "B", "C", "D"))) %>%
  select(-HOLC_Grade, -HOLC_Pct, -HOLC_Grade2) %>%
  dplyr::rename("HOLC_Pct" = HOLC_Pct2,
         "HOLC_Grade" = HOLC_Grade3) %>%
  select(TractFIPS, HOLC_Grade, HOLC_Pct, Year, ACCESS2:TEETHLOST) %>%
  distinct()
```

Combining HOLC/spatial data with 500 cities data: Short format

```
# Compiling HOLC grades by census tract
df4 <- df3 %>%
  group_by(TractFIPS) %>%
  dplyr::summarize(HOLC_Grades = paste(unique(HOLC_Grade), collapse = ", "))

# Getting the percent of census tract land graded positive
df4.pos <- df3 %>%
  filter(HOLC_Grade == "A" | HOLC_Grade == "B") %>%
  group_by(TractFIPS) %>%
  dplyr::summarize(HOLC_Pos_Pct = round(sum(HOLC_Pct) * 100, 2))

# Getting the percent of census tract land graded negative
df4.neg <- df3 %>%
  filter(HOLC_Grade == "C" | HOLC_Grade == "D") %>%
  group_by(TractFIPS) %>%
  dplyr::summarize(HOLC_Neg_Pct = round(sum(HOLC_Pct) * 100, 2))

# Joining dfs
```

```r
df5 <- left_join(df4, df4.pos)
df5 <- left_join(df5, df4.neg)

# Defining categorical exposure variables
df5 <- df5 %>%
  mutate(HOLC_Pos = as.numeric(str_detect(HOLC_Grades, "A|B")),
         HOLC_Neg = as.numeric(str_detect(HOLC_Grades, "C|D")),
         HOLC_Contains = factor(ifelse(HOLC_Pos == 1 & HOLC_Neg == 0, 1,
                                       ifelse(HOLC_Pos == 0 & HOLC_Neg == 1, 2,
                                              ifelse(HOLC_Pos == 1 & HOLC_Neg =
  = 1, 3, NA))),
                                levels = 1:3,
                                labels = c("Positive", "Negative", "Both")),
         HOLC_Majority = factor(ifelse(HOLC_Pos_Pct > 0.5, 1,
                                       ifelse(HOLC_Neg_Pct > 0.5, 2, 3)),
                                levels = 1:3,
                                labels = c("Positive", "Negative", "No majority
  ")),
         HOLC_Entirety = factor(ifelse(HOLC_Pos == 1 & HOLC_Neg == 0, 1,
                                       ifelse(HOLC_Pos == 0 & HOLC_Neg == 1, 2,
  3)),
                                levels = 1:3,
                                labels = c("Positive", "Negative", "Both")))

# Joining HOLC/spatial data with 500 cities data
data_short <- left_join(`500cities`, df5)
data_short <- data_short %>%
  select(TractFIPS, HOLC_Grades:HOLC_Entirety, Year, ACCESS2:TEETHLOST) %>%
  distinct()

# Removing unneeded dfs
rm(df, df2, df3, df4, df4.pos, df4.neg, df5)
```

Saving Data

```
# Saving dfs as .rdata
write_rds(`500cities`, "data.rdata")
write_rds(metrics, "measures.rdata")
write_rds(data_xlong, "data_xlong.rdata")
write_rds(data_long, "data_long.rdata")
write_rds(data_short, "data_short.rdata")

# Saving dfs as .csv
write_csv(`500cities`, "data.csv")
write_csv(metrics, "measures.csv")
write_csv(data_xlong, "data_xlong.csv")
write_csv(data_long, "data_long.csv")
write_csv(data_short, "data_short.csv")
```

Exploratory Analysis

Spatial Analysis in ArcGIS

Part of cleaning the data included an initial spatial analysis in ArcGIS using the HOLC redlining shapefiles and the census tract data. Using ArcGIS tools, we were able to generate stark visualizations of the health outcomes related to the HOLC grades.

ANOVA Test

We completed an analysis of variance for all 28 health metrics provided by the CDC data. This initial analysis was to determine which metrics would be most worth exploring further.

Linear Regression

We conducted a linear regression on four health measures to deepen our investigation into if there was a relationship between the HOLC graded areas and the outcomes.

Data Visualization

The Shiny App pulls together our data visualization tools–a collection of box plots that show the difference in health metrics between each HOLC grade and non-graded census tracts.

Loading Packages & Data

Loading packages

```
# Setting the ggplot2 theme
theme_sarah <- theme_fivethirtyeight() %+replace%
  theme(
    panel.background  = element_rect(fill = "white", color = "#D2D2D2", linetyp
  e = 1),
    plot.background = element_rect(fill = "white"),
    legend.background = element_rect(fill = "white"),
    legend.position = "right",
    axis.title = element_text(),
    panel.grid.major = element_line(size = 0.3, color = "#D2D2D2"),
    axis.ticks = element_line(size = 0.3, color = "#D2D2D2")
    )
theme_set(theme_sarah)


rsconnect::setAccountInfo(name='sarahbolongaita',
              token='0F0ECD3C344E89ABC3B88F09DA328B1A',
              secret='yXUEcH5Acp1oYieMQDb1xb/VdF18HaLBwwcPPfWO')
```

Loading data

```
# Loading data
data_long <- readRDS("data_long.rdata")
data_short <- readRDS("data_short.rdata")
measures <- readRDS("measures.rdata")
```

Exploratory Summary Statistics

Spatial Analysis

By using the intersect function in ArcGIS, we were able to derive how much of each HOLC graded area was in census tracts, allowing us to compare the health metrics assigned to each census tract across HOLC grades in the following statistical analyses. The ArcGIS map below is an example of our initial data exploration with diabetes prevalence in 2017 in Jackson. Each HOLC grade is color coded and areas with higher diabetes prevalence with maroon, while the lighter areas have a lower prevalence. As we continued our analysis, it became apparent that this visualizations were consistent with our statistical analysis. This initial analysis gave us a very clear indication that diabetes was an important metric to explore. We completed the following visual analysis with other metrics as well.
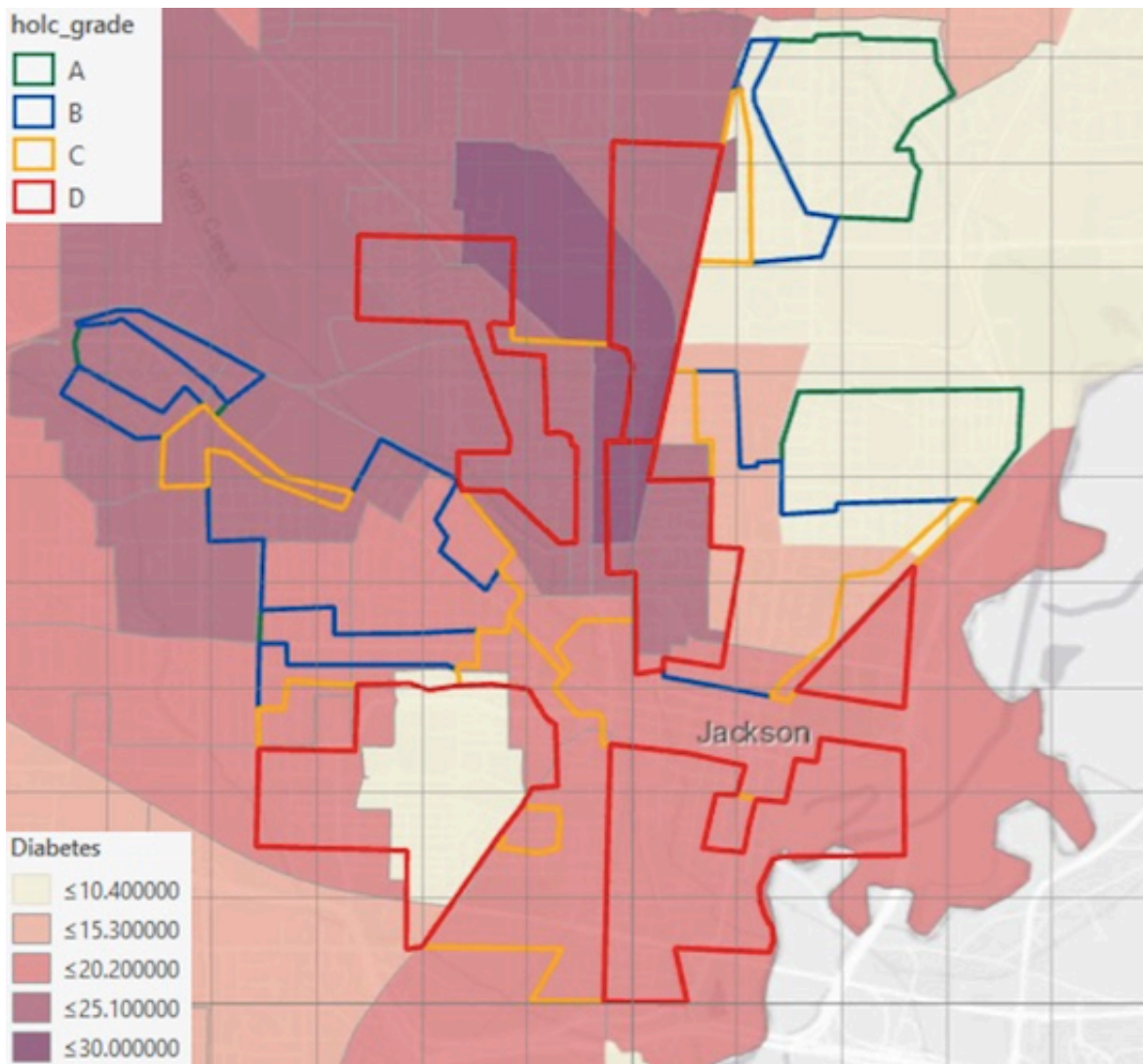
ANOVA Test

The table below shows that all of the health outcome variables were statistically significant. The analysis grouped census tracts into three groups: (1) Positive: Any census tract with over 20% inclusive of A or B graded areas (Prevalence is higher across most health outcomes); (2) Negative: Any census tract with over 20% inclusive of C and D; or (3) Both.

The findings suggest that on average, the prevalence of each health outcome is consistently lower for positive grade neighborhoods compared with negative HOLC grade areas. For instance, residents in positive HOLC areas have an uninsurance rate of 12.8% compared with 31.1% for negative HOLC neighborhoods - a difference of 18.3%. Similar findings are also evident across all the metrics. We then decided to further explore the findings for obesity, diabetes, and poor mental health among the four HOLC groups.

```
# Creating a wrapper for getting descriptive statistics
getTable1Stats <- function(x, digits = 1){
  getDescriptionStatsBy(x = x,
                        by = data_short$HOLC_Contains,
                        digits = digits,
                        statistics = TRUE,
                        header_count = TRUE)

}


# Getting descriptive statistics
mergeDesc("Uninsured" = getTable1Stats(data_short$ACCESS2),
          "High Blood Pressure" = getTable1Stats(data_short$BPHIGH),
          "Coronary Heart Disease" = getTable1Stats(data_short$CHD),
          "Annual Checkup" = getTable1Stats(data_short$CHECKUP),
          "Current Smoking" = getTable1Stats(data_short$CSMOKING),
          "Diabetes" = getTable1Stats(data_short$DIABETES),
          "High Cholesterol" = getTable1Stats(data_short$HIGHCHOL),
          "Physical Inactivity" = getTable1Stats(data_short$LPA),
          "Mental Health" = getTable1Stats(data_short$MHLTH),
          "Physical Health" = getTable1Stats(data_short$PHLTH),
          "Obesity" = getTable1Stats(data_short$OBESITY),
          "Sleep <7 hours" = getTable1Stats(data_short$SLEEP),
          "Stroke" = getTable1Stats(data_short$STROKE),
          htmlTable_args = list(align = "r"))
```

	Positive No. 10	Negative No. 40	Both No. 60	P-value
Uninsured				< 0.0001
Mean (SD)	12.8 (±5.9)	31.1 (±3.6)	26.5 (±8.1)	
Missing	4 (40.0%)	16 (40.0%)	24 (40.0%)	
High Blood Pressure				< 0.0001
Mean (SD)	38.9 (±3.1)	53.0 (±6.3)	44.1 (±10.1)	
Missing	4 (40.0%)	16 (40.0%)	24 (40.0%)	
Coronary Heart Disease				0.0001
Mean (SD)	6.5 (±0.3)	9.7 (±2.3)	7.2 (±2.4)	
Missing	4 (40.0%)	16 (40.0%)	24 (40.0%)	
Annual Checkup				0.016
Mean (SD)	76.0 (±3.5)	78.1 (±3.0)	75.2 (±4.2)	
Missing	4 (40.0%)	16 (40.0%)	24 (40.0%)	
Current Smoking				< 0.0001
Mean (SD)	14.2 (±5.1)	28.9 (±2.9)	25.7 (±6.3)	
Missing	4 (40.0%)	16 (40.0%)	24 (40.0%)	

				P-value
Diabetes				< 0.0001
Mean (SD)	10.5 (±2.3)	21.9 (±4.5)	16.1 (±5.8)	
Missing	4 (40.0%)	16 (40.0%)	24 (40.0%)	
High Cholesterol				0.01
Mean (SD)	38.6 (±3.9)	40.8 (±4.3)	36.7 (±5.4)	
Missing	4 (40.0%)	16 (40.0%)	24 (40.0%)	
Physical Inactivity				< 0.0001
Mean (SD)	25.4 (±6.7)	45.0 (±4.2)	38.1 (±9.3)	
Missing	4 (40.0%)	16 (40.0%)	24 (40.0%)	
Mental Health				< 0.0001
Mean (SD)	10.3 (±3.0)	18.7 (±1.6)	16.9 (±3.3)	
Missing	4 (40.0%)	16 (40.0%)	24 (40.0%)	
Physical Health				< 0.0001
Mean (SD)	10.9 (±2.7)	21.2 (±2.8)	17.0 (±5.1)	
Missing	4 (40.0%)	16 (40.0%)	24 (40.0%)	
Obesity				< 0.0001
Mean (SD)	28.5 (±7.1)	46.4 (±3.0)	40.5 (±8.1)	
Missing	4 (40.0%)	16 (40.0%)	24 (40.0%)	
Sleep <7 hours				< 0.0001
Mean (SD)	32.1 (±7.1)	46.4 (±2.2)	42.9 (±6.3)	
Missing	6 (60.0%)	24 (60.0%)	36 (60.0%)	
Stroke				< 0.0001
Mean (SD)	3.4 (±0.8)	7.6 (±1.8)	5.3 (±2.1)	
Missing	4 (40.0%)	16 (40.0%)	24 (40.0%)	

Graphical Analysis

Access to Health Insurance *Current lack of health insurance among adults aged 18–64 Years*

When looking at uninsurance rates across HOLC grades, negative grade neighborhoods (C/D) have higher uninsured prevalence compared to neighborhoods with positive (A/B) or no grade categorization. Between 2013 and 2017, nearly 1 in 3 residents in the most "Hazardous" areas did not have access to health insurance, whereas only 1 in 8 residents in the "Best" areas did not have access to insurance.

The findings are consistent with the historical policies of systematic exclusion in the form of redlining in Jackson, Mississippi. Residents in these neighborhoods, who were disproportionately communities of color, were denied health insurance coverage and access to quality healthcare. The resulting effects of denial of care contributed to a number of poor physical and mental health outcomes, particularly preventable chronic diseases such as diabetes, heart disease, and asthma.

Obesity *Obesity Prevalence among adults aged >= 18 Years*

When comparing obesity prevalence between the different HOLC grades, we see that obesity prevalence in census tracts labeled 'A' is around 25%, in 'B' around 39%, and around 47% in 'C' and 'D'. The prevalence of obesity is almost 2x higher in graded areas 'C' and 'D' compared to 'A' and 'B'. We can compare these prevalences to the areas designated as 'No grade' where the prevalence was around 42%.

There are many causes for obesity - diet, access to safe physical activity spaces, genetics etc. However, as discussed earlier, redlining affected not only home loan ownership practices but also investment from local businesses, including supermarkets. 'Supermarket Redlining' contributes to food deserts, which affect access to health foods. Although we cannot claim causality, we know that this contributes to a number of health issues, including obesity.

Diabetes *Diagnosed diabetes among adults aged >=18 Years*

At approximately 23%, the highest prevalence of diabetes in Jackson, Mississippi is found in neighborhoods that were categorized as "D". The prevalence of diabetes is nearly 3 times higher in these neighborhoods compared with areas categorized as "A".

Practice of redlining often involved inadequate access to quality food sources and supermarkets in these communities. In many areas of the United States today, quality food source is severely constrained within inner cities and low income neighborhoods and access to affordable, healthy food options are severely restricted. Good nutrition is critical for health and inadequate access to healthy foods in these neighborhoods contribute to a number of health issues, including obesity and diabetes.

Mental Health *Mental health not good for >=14 days among adults aged >=18 Years*

Low income neighborhoods, and in particular neighborhoods with high concentrations of people of color, have experienced systematic discrimination and racism for generations. A history of discrimination in the form of redlining has contributed to inqequality and poverty across commuinities in the United States. Poverty also increases the likelihood of adverse childhood experiences, lack of access to quality education, denial of opportunity, and poor mental health.

Intersection of community development and mental health is a critical component of public health. The World Health Organization (WHO) has identified poverty as a major risk factor for poor mental health. Communities that experience poverty are also less likely to have infrastructure and support systems for mental health, including access to mental health health care providers. Similar trends are found in Jackson, Mississippi where areas with low socioeconomic class (C and D) have nearly twice the prevalence of individuals with poor mental health outcomes.

```
# Box plot - ACCESS2
data_long %>%
  filter(!is.na(ACCESS2)) %>%
  ggplot(aes(x = HOLC_Grade, y = ACCESS2)) +
    geom_boxplot(fill = "darkolivegreen3", outlier.size = 0.7) +
    ggtitle("Uninsured in Jackson, 2013-2017") +
    xlab("HOLC grade history") + theme(axis.text.x = element_text(vjust = 0.5),
                 panel.grid.major.x = element_blank()) +
    ylab("Uninsured prevalence (%)") + scale_y_continuous(breaks = scales::pret
ty_breaks(n = 6)) +
    theme(legend.position = "none")
```
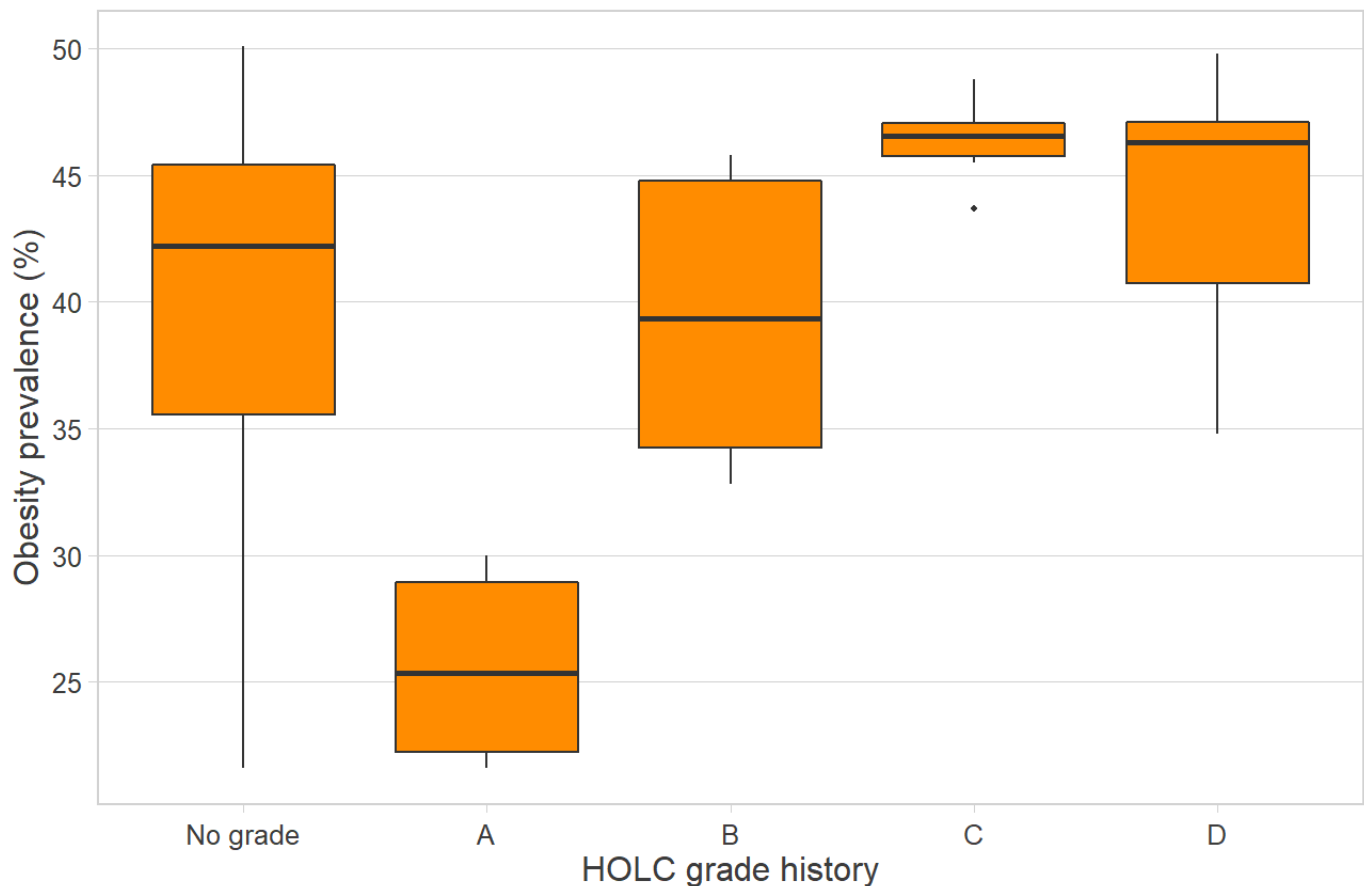

Uninsured in Jackson, 2013-2017
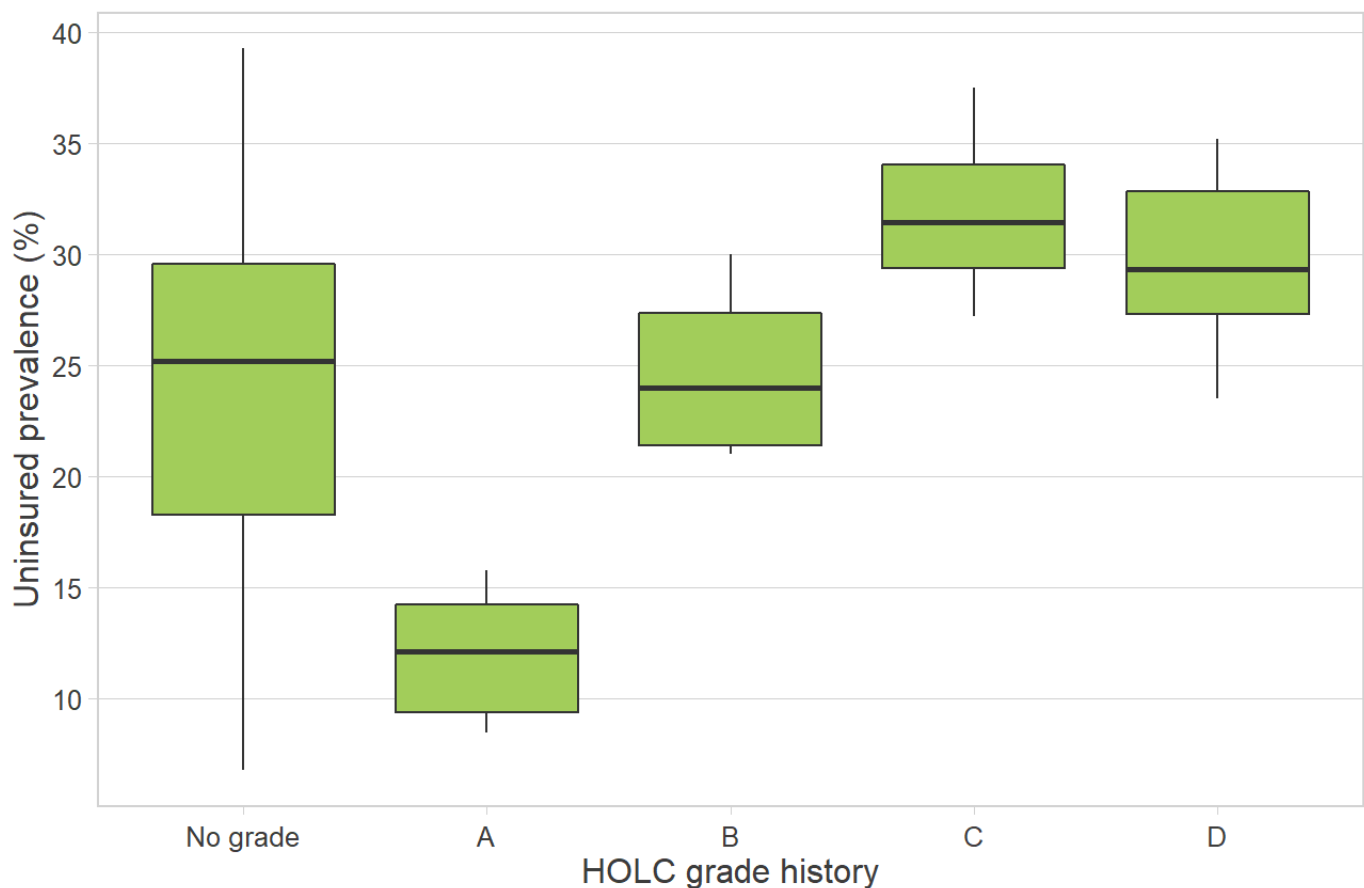
```
# Box plot - OBESITY
data_long %>%
  filter(!is.na(OBESITY)) %>%
  ggplot(aes(x = HOLC_Grade, y = OBESITY)) +
    geom_boxplot(fill = "darkorange", outlier.size = 0.7) +
    ggtitle("Obesity in Jackson, 2013-2017") +
    xlab("HOLC grade history") + theme(axis.text.x = element_text(vjust = 0.5),
                  panel.grid.major.x = element_blank()) +
    ylab("Obesity prevalence (%)") + scale_y_continuous(breaks = scales::pretty
  _breaks(n = 6)) +
    theme(legend.position = "none")
```
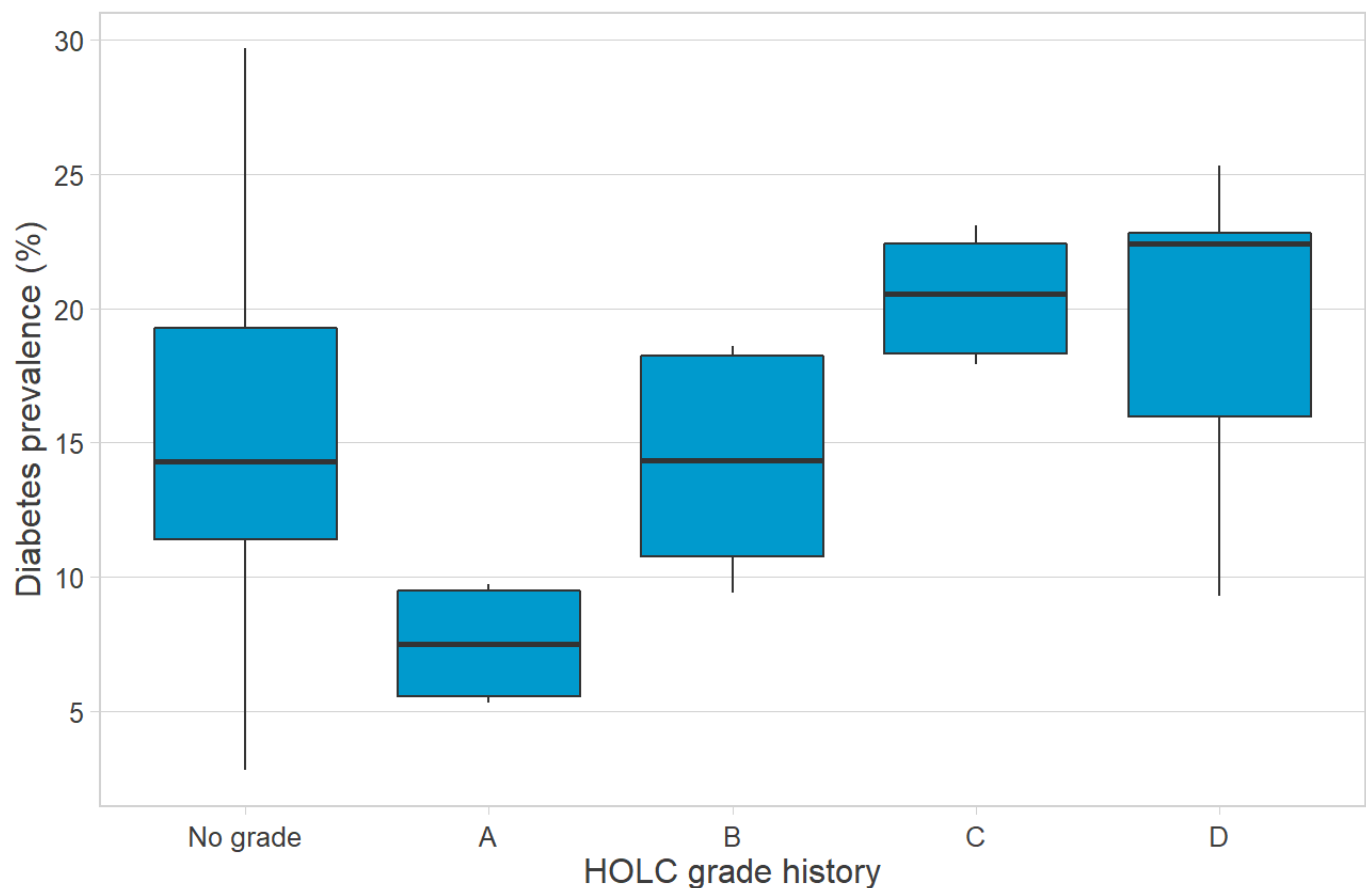
Obesity in Jackson, 2013-2017

```
# Box plot - DIABETES
 data_long %>%
  filter(!is.na(DIABETES)) %>%
  ggplot(aes(x = HOLC_Grade, y = DIABETES)) +
    geom_boxplot(fill = "deepskyblue3", outlier.size = 0.7) +
    ggtitle("Diabetes in Jackson, 2013-2017") +
    xlab("HOLC grade history") + theme(axis.text.x = element_text(vjust = 0.5),
                    panel.grid.major.x = element_blank()) +
    ylab("Diabetes prevalence (%)") + scale_y_continuous(breaks = scales::prett
 y_breaks(n = 6)) +
    theme(legend.position = "none")
```


Diabetes in Jackson, 2013-2017

```
# Box plot - MHLTH
data_long %>%
  filter(!is.na(MHLTH)) %>%
  ggplot(aes(x = HOLC_Grade, y = MHLTH)) +
    geom_boxplot(fill = "gold1", outlier.size = 0.7) +
    ggtitle("Poor mental health in Jackson, 2013-2017") +
    xlab("HOLC grade history") + theme(axis.text.x = element_text(vjust = 0.5),
                  panel.grid.major.x = element_blank()) +
    ylab("Poor mental health prevalence (%)") + scale_y_continuous(breaks =
  scales::pretty_breaks(n = 6)) +
    theme(legend.position = "none")
```
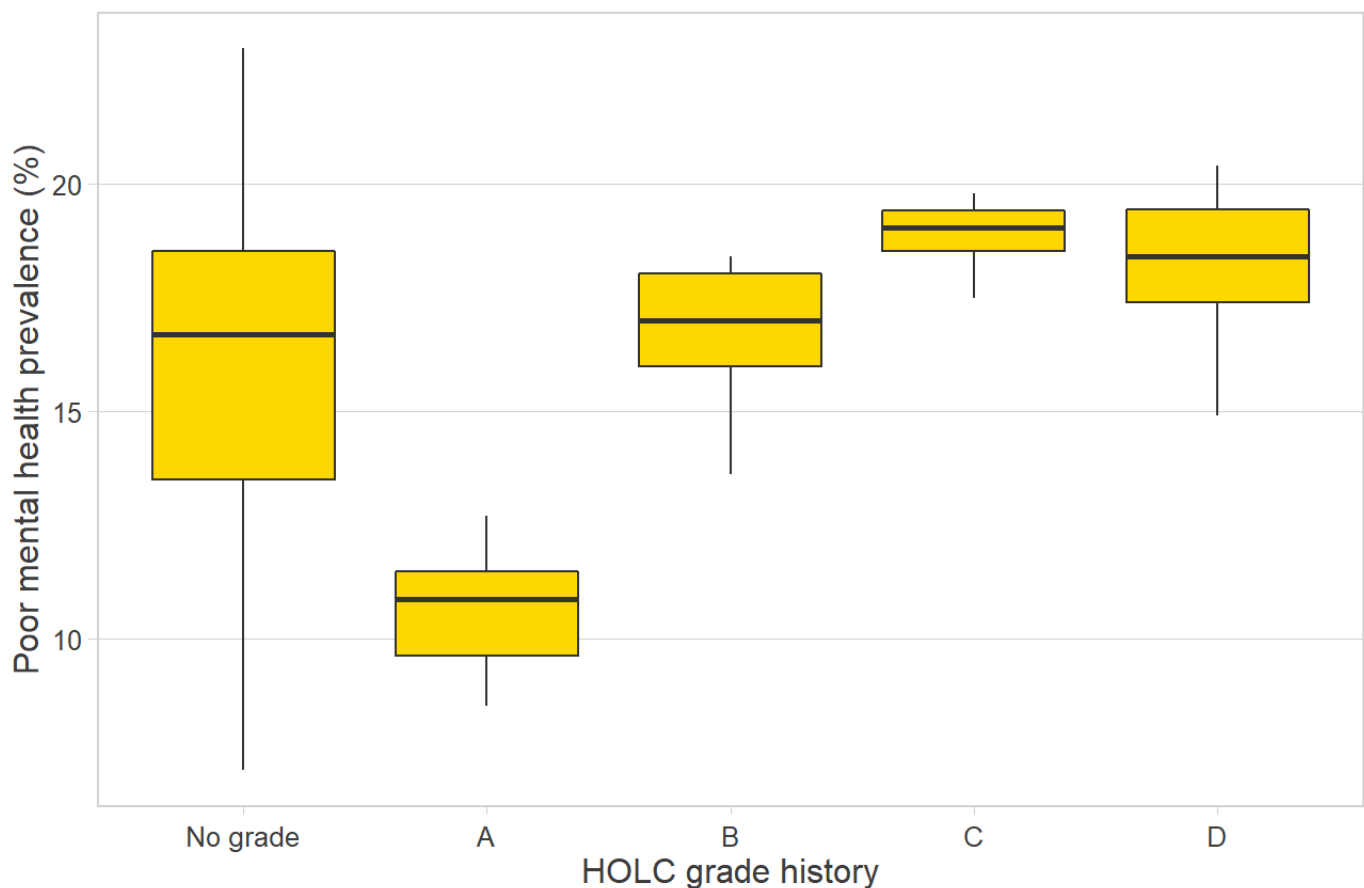

Regression Analysis

Aside from examining box plots and creating a dynamic shiny app to explore the data, we ran linear regressions on the four outcomes that we chose for analysis. The interpretation of each regression is below:

Access to health care Obesity Diabetes Mental Health

We note that the R^2 values for each association is very low. However, this is partially attributed to the fact that holc_grades are categorical not continuous data points.

Access to Health Insurance

We find a positive linear association between grade B,C, and D areas and lack of health insurance among residents in Jackson, MI. The association between HOLC Grade A, C, and D and uninsured rates is also statistically significant. We estimate that, on average, uninsured rates among residents in redlined neighborhoods (D) is 5.46% higher compared to neighborhoods that did not have a HOLC grade. However, on average, residents in the "Best" neighborhoods had a negative association with lack of health insurance. In fact, uninsured rate is 12.44% lower than neighborhoods that did not not have a HOLC grade. This is consistent with our findings from the exploratory data analysis section where HOLC A residents by far had the lowest insurance rates.

Obesity

We find similar trends in our obesity data, where there is a positive linear association between B, C neighborhoods and obesity prevalence. The association between A, C, and D areas and obesity is also statistically significant. We estimate that, on average, obesity prevalence among residents in redlined neighborhoods (D) is 4.14% higher compared to neighborhoods that did not have a HOLC grade. However, on average, residents in the "Best" neighborhoods had a negative association with obesity: Prevalence of obesity is 14.48% lower than neighborhoods that did not have a HOLC grade.

Diabetes

We find a positive linear association between grade B, C neighborhoods and diabetes prevalence, and the findings are statistically significant. We estimate that, on average, diabetes prevalence among residents in redlined neighborhoods (D) is 4.23% higher compared to neighborhoods that did not have a HOLC grade. However, on average, residents in the "A" neighborhoods had a protective factor and a negative association with diabetes: The prevalence is 14.4% lower than neighborhoods that did not not have a HOLC grade.

Mental Health

Lastly, we found a positive linear association between all the HOLC grade neighborhoods and poor mental health, with the exception of grade A. The findings were also statistically significant. We estimate that, on average, poor mental health among residents in redlined neighborhoods is 2.25% higher compared with neighborhoods that did not have a HOLC grade. However, on average, residents in the "Best" neighborhoods had a negative association with poor mental health outcomes: The prevalence is 5.38% lower than neighborhoods that did not not have a HOLC grade.

We also note that the R^2 values for each association is very low. However, this is partially attributed to the fact that holc_grades are categorical not continuous data points.

```
mod.ACCESS2 <- lm(ACCESS2 ~ HOLC_Grade, data = data_long)
summary(mod.ACCESS2)
```

```
## 
## Call:
## lm(formula = ACCESS2 ~ HOLC_Grade, data = data_long)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -17.6474  -4.0202   0.5526   4.6526  14.8526 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  24.4474     0.5672  43.101  < 2e-16 ***
## HOLC_GradeA -12.4474     2.9473  -4.223 3.76e-05 ***
## HOLC_GradeB   0.2192     2.9473   0.074  0.94079    
## HOLC_GradeC   7.4192     2.9473   2.517  0.01267 *  
## HOLC_GradeD   5.4637     1.7635   3.098  0.00225 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 7.085 on 187 degrees of freedom
##   (133 observations deleted due to missingness)
## Multiple R-squared:  0.1588, Adjusted R-squared:  0.1409 
## F-statistic: 8.828 on 4 and 187 DF,  p-value: 1.5e-06
```

```
mod.OBESITY <- lm(OBESITY ~ HOLC_Grade, data = data_long)
summary(mod.OBESITY)
```

```
## 
## Call:
## lm(formula = OBESITY ~ HOLC_Grade, data = data_long)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.486   -4.236    2.014    4.914   10.014
## 
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   40.0859     0.5351  74.911  < 2e-16 ***
## HOLC_GradeA  -14.4859     2.7805  -5.210 4.97e-07 ***
## HOLC_GradeB   -0.6859     2.7805  -0.247   0.8054
## HOLC_GradeC    6.2974     2.7805   2.265   0.0247 *
## HOLC_GradeD    4.1419     1.6637   2.490   0.0137 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 6.684 on 187 degrees of freedom
##   (133 observations deleted due to missingness)
## Multiple R-squared:  0.1778, Adjusted R-squared:  0.1602
## F-statistic: 10.11 on 4 and 187 DF,  p-value: 1.981e-07
```

```
mod.DIABETES <- lm(DIABETES ~ HOLC_Grade, data = data_long)
summary(mod.DIABETES)
```

```
## 
## Call:
## lm(formula = DIABETES ~ HOLC_Grade, data = data_long)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -12.4308  -3.6077  -0.3808   3.4769  14.4692 
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  15.2308     0.4224  36.056  < 2e-16 ***
## HOLC_GradeA  -7.7308     2.1950  -3.522 0.000538 ***
## HOLC_GradeB  -0.9308     2.1950  -0.424 0.672021    
## HOLC_GradeC   5.2192     2.1950   2.378 0.018425 *  
## HOLC_GradeD   4.2303     1.3134   3.221 0.001507 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 5.276 on 187 degrees of freedom
##   (133 observations deleted due to missingness)
## Multiple R-squared:  0.138,  Adjusted R-squared:  0.1196 
## F-statistic: 7.484 on 4 and 187 DF,  p-value: 1.298e-05
```

```
mod.MHLTH <- lm(MHLTH ~ HOLC_Grade, data = data_long)
summary(mod.MHLTH)
```

```
##
## Call:
## lm(formula = MHLTH ~ HOLC_Grade, data = data_long)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.9179 -2.0179  0.5049  2.2821  6.9821
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  16.0179     0.2575  62.217  < 2e-16 ***
## HOLC_GradeA  -5.3846     1.3378  -4.025 8.27e-05 ***
## HOLC_GradeB   0.6487     1.3378   0.485  0.62830
## HOLC_GradeC   2.8654     1.3378   2.142  0.03349 *
## HOLC_GradeD   2.2543     0.8005   2.816  0.00538 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.216 on 187 degrees of freedom
##   (133 observations deleted due to missingness)
## Multiple R-squared:  0.1392, Adjusted R-squared:  0.1208
## F-statistic: 7.561 on 4 and 187 DF,  p-value: 1.146e-05
```

Final Analysis

Our exploration of data in this project show that redlining may have a negative effect on a wide range of health measures.

Nearly **1 in 3 residents do not have health insuranace** in previously redlined areas do not have health insurance.

The **prevalence of obesity are two times higher** in neighborhoods that were previously redlined. Similarly, **diabetes prevalence is nearly three times higher**.

Almost **20% of residents experience mental distress** in previously redlined areas.

These egregious results allude to the impact of policies and history that impact health measures in present day. It is not sufficient to relegate unhealthy outcomes to individual behaviors (e.g., eating poorly, not exercising, etc.) and these insights need to be shared with policymakers, health professionals, and most importantly community members that have been impacted by redlining.

Our findings, though preliminary, demand further investigation and persistent advocacy to hold governments and systems (i.e., health systems, economic systems, etc.) responsible.

Discussion & Challenges

Within the scope of this research, we did not account for confounders such as race/ethnicity and median income. Future analysis could involve these additional indicators which would allow us to assume causality about the relationship between redlining and health outcomes. This next step in analysis would work towards determining if there is a causal relationship between redlining and various health measures.

An additional question that came up was the difference between areas graded as 'C' and 'D'. For the four indicators that we chose (rate of uninsurance, obesity, diabetes, and poor mental health), we noticed that the spread of data was often much narrower for areas designated as 'C' versus those designated as 'D'. In addition, areas designated as 'C' frequently had a higher prevalence of negative health outcomes than areas designated as 'D'. An additional question we could further explore may involve differences between areas labeled 'C' versus those labeled 'D'. This question could be answered again by looking at demographic differences between the two census tracts.

Lastly, we only included data between 2013-2017. Additional analysis could look at a wider range of years to show the continued effect of redlining since the 1930s.