````
---
title: "Census Regression Project"
author: "Jarius Hamid"
date: "2024-09-17"
output:
  pdf_document: default
  word_document: default
---

```{r setup, include=FALSE}
library(lubridate)
library(tidyverse)
library(ggplot2)
library(dplyr)

#Loading Census Tract Data
tract_data <- read.csv("C:/Users/hamidja/OneDrive - Lee County BoCC/Intern
Project/r_tract.csv")

#examining collinearity betwen similar variables
cor(tract_data$highschool_prct, tract_data$college_prct)
```

```{r Summary Stats}
library(psych)
library(xtable)
library(gridExtra)
library(grid)

# Compute summary statistics
table <- describe(tract_data)

# Convert the summary statistics to a data frame
table_df <- as.data.frame(table)

# Remove unnecessary columns and rename them
table_df <- table_df[, c("mean", "sd", "median", "range", "skew", "kurtosis")]
names(table_df) <- c("Mean", "SD", "Median", "Range", "Skew", "Kurtosis")

# Add a column for variable names
table_df$Variable <- rownames(table_df)

# Reorder columns
describe(table_df)


# Generate LaTeX table
#latex_table <- xtable(table_df, caption = "Summary Statistics", label =
"tab:summary_statistics")

# Print LaTeX code
#print(latex_table, type = "latex", include.rownames = FALSE, booktabs = TRUE)


```

VISUALS

```{r visuals}
# Load necessary libraries
# Select the relevant columns for correlation analysis
cor_data <- tract_data[, c("med_householdincome", "poverty_total", "employment.total",
                           "white_prct", "black_prct", "asian_prct", "other_prct",
                           "college_prct", "avg_workhours",
````
````

```
"distance_from_closest_stop_miles", "total_pop")]

# Shorten column names
colnames(cor_data) <- c("Med. Income", "Poverty %", "Employment",
                        "White %", "Black %", "Asian %", "Other %",
                        "College %", "Avg. Work Hours", "Distance to Stop
(miles)","total_pop")

# Compute the correlation matrix
cor_matrix <- cor(cor_data, use = "complete.obs")

# Print the correlation matrix
print(cor_matrix)

# Visualize the correlation matrix with shortened labels
library(corrplot)
corrplot(cor_matrix, method = "circle", type = "upper",
         tl.col = "black", tl.srt = 45,
         tl.cex = 0.8)  # Optionally adjust text size with tl.cex
# Load necessary library
library(ggplot2)
# Load necessary package
library(gridExtra)

# Create individual plots
plot1 <- ggplot(tract_data, aes(x = distance_from_closest_stop_miles, y =
med_householdincome)) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(
       x = "Distance to Stop (miles)",
       y = "Median Household Income") +
  theme_minimal()

plot2 <- ggplot(tract_data, aes(x = distance_from_closest_stop_miles, y = poverty_prct)) +
  geom_point(color = "green") +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(
       x = "Distance to Stop (miles)",
       y = "Poverty Percentage") +
  theme_minimal()

plot3 <- ggplot(tract_data, aes(x = distance_from_closest_stop_miles, y =
employment_prct)) +
  geom_point(color = "purple") +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(
       x = "Distance to Stop (miles)",
       y = "Employment Percentage") +
  theme_minimal()

print(plot1)
print(plot2)
print(plot3)
library(patchwork)

# Combine plots side by side
plot1 + plot2 + plot3
```


3 Different Models

```{r models}

# Create non_white_prct variable
```

```
tract_data$non_white_prct <- 100 - tract_data$white_prct

# Model 1: Median Household Income (using non_white_prct)
model1 <- lm(med_householdincome ~ distance_from_closest_stop_miles + total_pop +
               non_white_prct +
               college_prct + avg_workhours, data = tract_data)

summary(model1)

# Model 2: Total Poverty (using non_white_prct)
model2 <- lm(poverty_prct ~ distance_from_closest_stop_miles + total_pop +
               non_white_prct +
               college_prct + avg_workhours, data = tract_data)

summary(model2)

# Model 3: Employment Percentage (using non_white_prct)
model3 <- lm(employment_prct ~ distance_from_closest_stop_miles + total_pop +
               non_white_prct +
               college_prct + avg_workhours, data = tract_data)

summary(model3)


# Model 1: Median Household Income
model1 <- lm(med_householdincome ~ distance_from_closest_stop_miles + total_pop +
               black_prct + asian_prct + other_prct +
               college_prct + avg_workhours, data = tract_data)

summary(model1)

# Model 2: Total Poverty
model2 <- lm(poverty_prct ~ distance_from_closest_stop_miles + total_pop +
               black_prct + asian_prct + other_prct +
               college_prct + avg_workhours, data = tract_data)

summary(model2)

# Model 3: Employment Percentage
model3 <- lm(employment_prct ~ distance_from_closest_stop_miles + total_pop +
               black_prct + asian_prct + other_prct +
               college_prct + avg_workhours, data = tract_data)

summary(model3)

combined_model <- lm(distance_from_closest_stop_miles ~ med_householdincome + poverty_prct
+ employment_prct + total_pop +
                       black_prct + asian_prct + other_prct + college_prct +
avg_workhours, data = tract_data)
summary(combined_model)


```
```
```