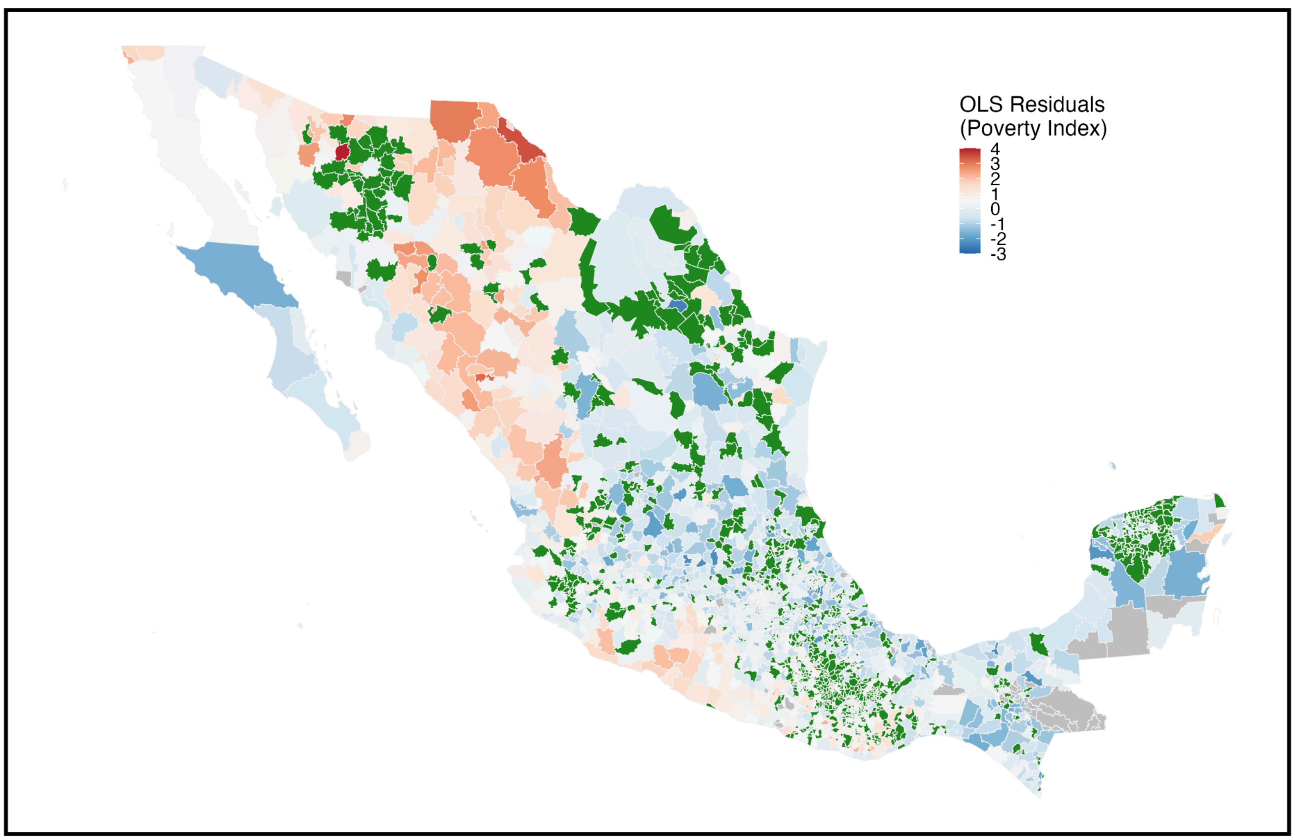
A one-unit increase in the poverty index is associated with an increase of 0.3325 in the log-transformed homicide rate, according to the regression coefficient for indice. To understand the effect on the original homicide rate scale, the exponential function is applied to the coefficient, resulting in approximately 1.3944. This means that a one-unit increase in the poverty index corresponds to a 39.4% higher homicide rate. The standard deviation of the poverty index in the filtered dataset (excluding municipalities with zero homicides or missing data) is 0.8514. When considering a one-standard-deviation increase in the poverty index, the product of the coefficient and the standard deviation is computed and then exponentiated, yielding an effect multiplier of approximately 1.3267. This translates to a 32.7% higher homicide rate for a one-standard-deviation increase in the poverty index. These calculations are based solely on municipalities with valid, non-zero homicide rates to ensure the accuracy of the results.

```
Call:
lm(formula = log_homicide_rate ~ indice, data = analysis_tabular_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.4702 -0.6347 -0.0749  0.5455  3.9491 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.25260    0.03780   86.06   <2e-16 ***
indice       0.33246    0.02903   11.45   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9373 on 1441 degrees of freedom
Multiple R-squared:  0.08343,	Adjusted R-squared:  0.08279 
F-statistic: 131.2 on 1 and 1441 DF,  p-value: < 2.2e-16
```

OLS Regression Residuals of Homicide Rate on Poverty Index Across Mexico

Residuals from the Ordinary Least Squares regression of logged homicide rate (per 100,000 population) on poverty index.



Green: Zero Homicides; Grey: No Data.
Date: December 2024

```r
library(sf)        # For handling spatial data
library(dplyr)     # For data manipulation
library(ggplot2)   # For visualization
library(tidyverse) # General data wrangling
library(spgwr)     # For Geographically Weighted Regression
library(ggspatial) # Map additions

# Load CSV and shapefile
tabular_data <- read.csv("~/Downloads/Assignment4Data/mex_data.csv")
shapefile <- st_read("~/Downloads/Assignment4Data/municipalities.shp")
```

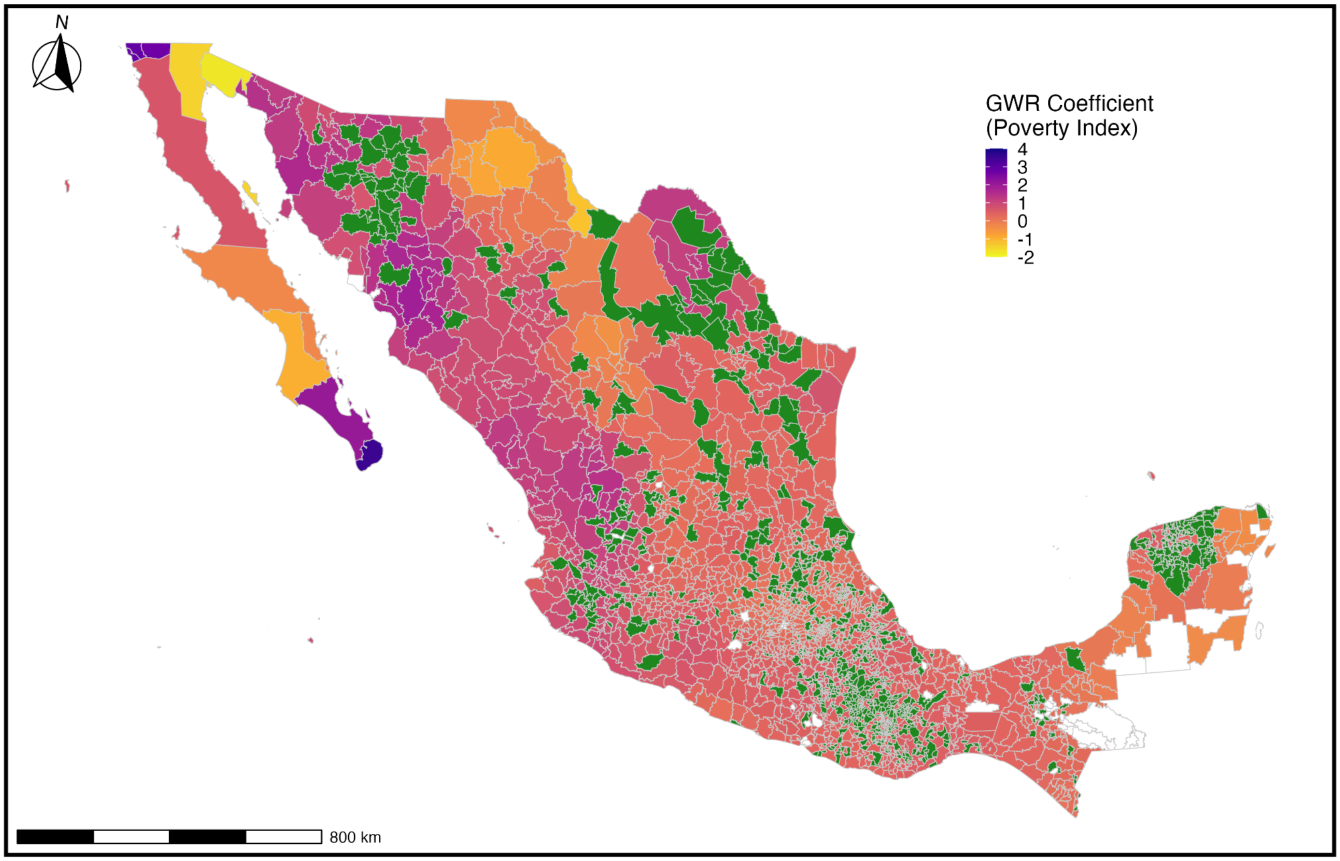
```r
# ===========================
# Question 1: OLS Regression
# ===========================

# Full dataset for later mapping purposes (including zero homicides)
full_tabular_data <- tabular_data %>%
  rename(CVE_ENT = state, CVE_MUN = mun) %>%
  mutate(
    homicide_rate = (homicides / population) * 100000,
    log_homicide_rate = ifelse(homicides == 0, NA, log(homicide_rate))  # Assign NA for zero homicides
  )

# Filtered dataset for regression analysis (exclude zero homicides)
analysis_tabular_data <- full_tabular_data %>%
  filter(homicides > 0)

# Perform regression on filtered data
ols_model <- lm(log_homicide_rate ~ indice, data = analysis_tabular_data)

# Add residuals to the filtered dataset
analysis_tabular_data <- analysis_tabular_data %>%
  mutate(
    residuals = residuals(ols_model)  # Add residuals to the analysis dataset
  )

# Print regression summary
print(summary(ols_model))

# Coefficient for poverty index
poverty_coef <- coef(ols_model)["indice"]

# Standard deviation of poverty index (filtered dataset)
sd_indice_filtered <- sd(analysis_tabular_data$indice, na.rm = TRUE)

# Percentage change for one-unit increase
percent_change_unit <- (exp(poverty_coef) - 1) * 100

# Percentage change for one-standard-deviation increase
percent_change_sd_filtered <- (exp(poverty_coef * sd_indice_filtered) - 1) * 100

# Print all results
print(paste("Coefficient for the poverty index:", round(poverty_coef, 4)))
print(paste("Percentage change in homicide rate for a one-unit increase in poverty index:", round(percent_change_unit, 2), "%"))
print(paste("Standard deviation of the poverty index (0s & No data filtered out):", round(sd_indice_filtered, 4)))
print(paste("Percentage change in homicide rate for a one-standard-deviation increase in poverty index (0s & No data filtered out):", round(percent_change_sd_filtered, 2), "%"))
```

```r
# ==========================
# Question 2: OLS Residuals
# ==========================

# Add data_status column to full_tabular_data
full_tabular_data <- full_tabular_data %>%
  left_join(
    analysis_tabular_data %>% select(CVE_ENT, CVE_MUN, residuals),
    by = c("CVE_ENT", "CVE_MUN")
  ) %>%
  mutate(
    data_status = case_when(
      homicides == 0 ~ "Zero Homicides",      # Label zero-homicide areas
      is.na(residuals) ~ "No Data",           # Handle any missing data
      TRUE ~ "Residuals Available"            # Data available for analysis
    )
  )

# Check the status of residuals in full_tabular_data
print("Summary of residuals in full_tabular_data:")
print(summary(full_tabular_data$residuals))

# Convert CVE_ENT and CVE_MUN in shapefile to character type for compatibility
shapefile <- shapefile %>%
  mutate(
    CVE_ENT = as.character(CVE_ENT),
    CVE_MUN = as.character(CVE_MUN)
  )

# Convert CVE_ENT and CVE_MUN to character type and pad them for compatibility
full_tabular_data <- full_tabular_data %>%
  mutate(
    CVE_ENT = str_pad(as.character(CVE_ENT), width = 2, pad = "0"),
    CVE_MUN = str_pad(as.character(CVE_MUN), width = 3, pad = "0")
  )

# Merge data and ensure all geometries are included
merged_data <- shapefile %>%
  full_join(
    full_tabular_data %>% select(CVE_ENT, CVE_MUN, log_homicide_rate, indice, homicides, population, residuals, data_status),
    by = c("CVE_ENT", "CVE_MUN")
  ) %>%
  st_as_sf() %>%
  mutate(
    data_status = ifelse(is.na(data_status), "No Data", data_status)  # Assign "No Data" to unmatched rows
  )
```

```r
# Create the map
ols_plot <- ggplot() +
  # Residuals layer for municipalities with data
  geom_sf(
    data = merged_data %>% filter(data_status == "Residuals Available"),
    aes(fill = residuals),
    color = "gray95",
    size = 0.0005
  ) +
  scale_fill_distiller(
    palette = "RdBu",
    limits = c(-3, ceiling(max(merged_data$residuals, na.rm = TRUE))),
    breaks = seq(-3, ceiling(max(merged_data$residuals, na.rm = TRUE)), by = 1),
    name = "OLS Residuals\n(Poverty Index)"
  ) +
  # Zero Homicides layer
  geom_sf(
    data = merged_data %>% filter(data_status == "Zero Homicides"),
    fill = "forestgreen",
    color = "gray95",
    size = 0.0005
  ) +
  # No Data layer
  geom_sf(
    data = merged_data %>% filter(data_status == "No Data"),
    fill = "grey",
    color = "grey95",
    size = 0.0005
  ) +
  # Titles and captions
  labs(
    title = "OLS Regression Residuals of Homicide Rate on Poverty Index Across Mexico",
    subtitle = "Residuals from the Ordinary Least Squares regression of logged homicide rate (per 100,000 population) on poverty index.",
    caption = "Green: Zero Homicides; Grey: No Data.\nDate: December 2024"
  ) +
  theme_minimal() +
  theme(
    legend.position = c(0.8, 0.8),
    legend.background = element_rect(fill = alpha("white", 0.8), color = NA),
    legend.key.size = unit(0.8, "lines"),
    legend.title = element_text(size = 12),
    legend.text = element_text(size = 10),
    plot.title = element_text(size = 14, face = "bold"),
    plot.subtitle = element_text(size = 11),
    panel.grid = element_blank(),
    axis.text = element_blank(),
    axis.title = element_blank(),
    axis.ticks = element_blank(),
    panel.border = element_rect(color = "black", fill = NA, size = 2)
  )

# Display the map
#print(ols_plot)

# Save the map
#ggsave("OLS_Residuals_Map.png", plot = ols_plot, width = 10, height = 8, dpi = 400)
```

GWR Coefficients of Poverty on Homicide Rates in Mexico

Geographically Weighted Regression analysis displaying the spatial relationship between poverty index and log-transformed homicide rates.



Green: Zero Homicides; White: No Data.
Date: December 2024

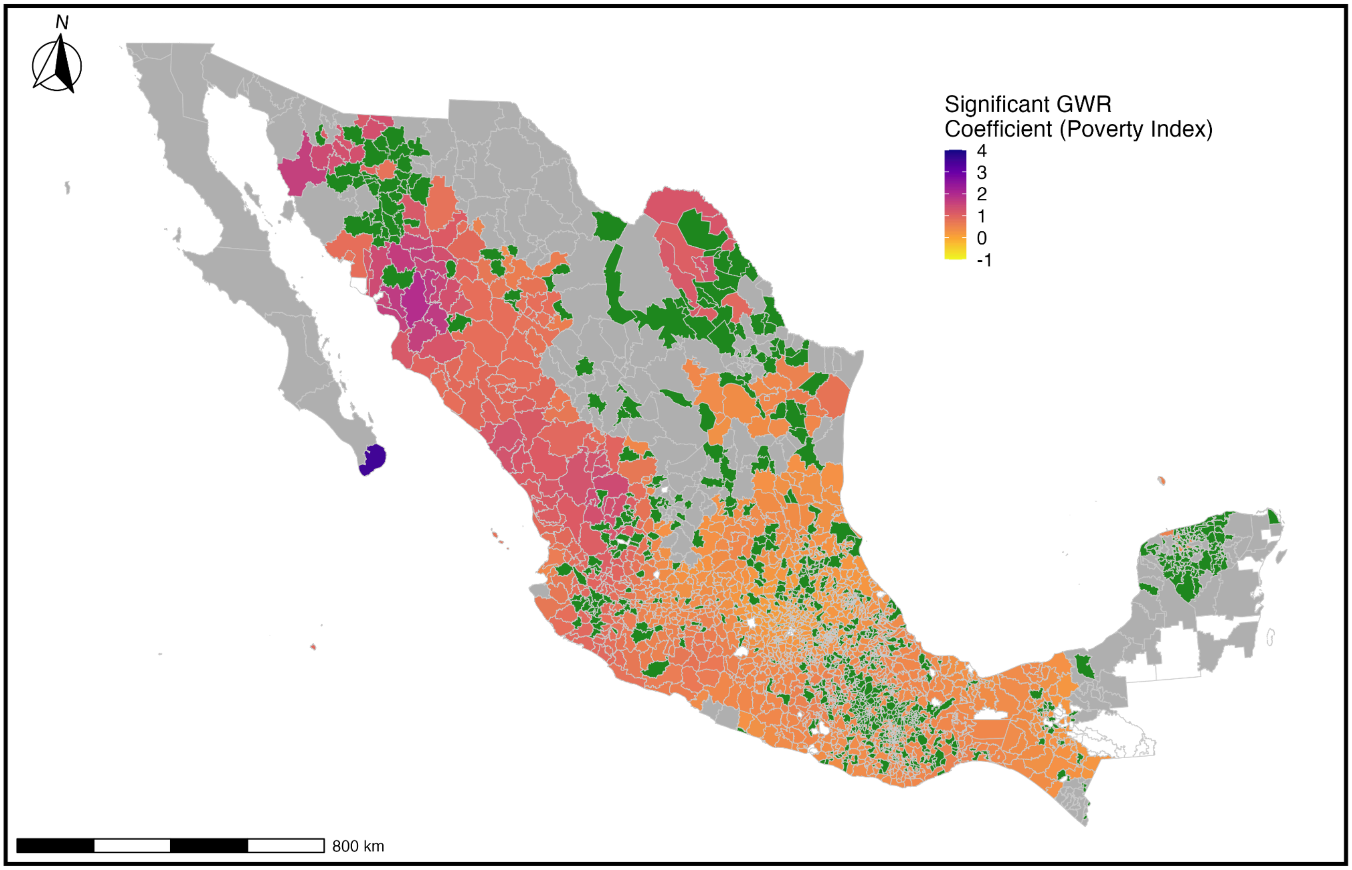
The map above illustrates the spatial variation in the relationship between poverty (measured by the poverty index) and homicide rates (log-transformed) across municipalities in Mexico, based on a Geographically Weighted Regression (GWR) analysis. Areas with positive coefficients (yellow and orange) indicate that higher poverty levels are associated with increased homicide rates, while areas with negative coefficients (blue and purple) suggest an inverse relationship. The results reveal significant regional differences, with stronger positive associations in some northern and central regions, while weaker or negative associations are more evident in the south. Gray areas represent municipalities where data were insufficient or missing, preventing GWR analysis. These regions highlight gaps in data coverage, which could be explored further to ensure a more comprehensive understanding of the relationship between poverty and homicide rates.

```r
# ========================
# Question 3: GWR Analysis
# ========================

# Reproject the merged_data to a metric CRS (EPSG:6362 for Mexico)
merged_data <- merged_data %>%
  st_transform(6362)

# Filter merged_data to remove rows with missing values for the variables used in GWR
gwr_data <- merged_data %>%
  filter(!is.na(log_homicide_rate), !is.na(indice))

# Convert the filtered data to a SpatialPointsDataFrame (required by spgwr)
gwr_data_sp <- as(gwr_data, "Spatial")

# Determine the optimal bandwidth using AICc
bw <- gwr.sel(
  log_homicide_rate ~ indice,
  data = gwr_data_sp
)
print(paste("Optimal Bandwidth:", bw))

# Run the GWR model with standard errors enabled
gwr_out <- gwr(
  formula = log_homicide_rate ~ indice,
  data = gwr_data_sp,
  bandwidth = bw,
  hatmatrix = TRUE,  # Include hat matrix for diagnostics
  se.fit = TRUE      # Compute standard errors
)

# Extract GWR results as a data frame
gwr_results <- as.data.frame(gwr_out$SDF)

# Add GWR coefficients and standard errors to the spatial data
gwr_data <- gwr_data %>%
  mutate(
    indice_coef = gwr_results$indice,    # Coefficients for poverty index
    indice_se = gwr_results$indice_se    # Standard errors for poverty index
  )

# Add GWR results back to the full dataset
new_merged_data <- merged_data %>%
  st_join(
    gwr_data %>% select(CVE_ENT, CVE_MUN, indice_coef, indice_se),
    join = st_equals
  )

# Add a column for missing data handling
new_merged_data <- new_merged_data %>%
  mutate(
    fill_category = case_when(
      data_status == "Zero Homicides" ~ "Zero Homicides",
      is.na(indice_coef) ~ "No Data",
      TRUE ~ "GWR Coefficient"
    )
  )
```

```r
# Create the coefficient map
coef_map <- ggplot() +
  # Layer for areas with GWR coefficients
  geom_sf(
    data = new_merged_data %>% filter(fill_category == "GWR Coefficient"),
    aes(fill = indice_coef),
    color = "gray80",
    size = 0.0005
  ) +
  # Color scale for GWR coefficients
  scale_fill_viridis_c(
    option = "plasma",
    direction = -1,
    limits = c(-2, 4),
    breaks = seq(-2, 4, by = 1),
    name = "GWR Coefficient\n(Poverty Index)"
  ) +
  # Layer for "Zero Homicides" areas
  geom_sf(
    data = new_merged_data %>% filter(fill_category == "Zero Homicides"),
    fill = "forestgreen",
    color = "gray80",
    size = 0.0005
  ) +
  # Layer for "No Data" areas
  geom_sf(
    data = new_merged_data %>% filter(fill_category == "No Data"),
    fill = "white",
    color = "gray80",
    size = 0.0005
  ) +
  # Add scale bar
  annotation_scale(
    location = "bl",  # Position: bottom left
    width_hint = 0.25 # Scale bar width relative to the plot
  ) +
  # Add north arrow
  annotation_north_arrow(
    location = "tl",  # Position: top left
    which_north = "true",  # Use true north
    style = north_arrow_fancy_orienteering()  # Choose arrow style
  ) +
  # Titles and captions
  labs(
    title = "GWR Coefficients of Poverty on Homicide Rates in Mexico",
    subtitle = "Geographically Weighted Regression analysis displaying the spatial relationship between poverty index and log-transformed homicide rates.",
    caption = "Green: Zero Homicides; White: No Data.\nDate: December 2024"
  ) +
  # Theme adjustments
  theme_minimal() +
  theme(
    legend.position = c(0.8, 0.8),
    legend.background = element_rect(fill = alpha("white", 0.8), color = NA),
    legend.key.size = unit(0.8, "lines"),
    legend.title = element_text(size = 12),
    legend.text = element_text(size = 10),
    plot.title = element_text(size = 14, face = "bold"),
    plot.subtitle = element_text(size = 11),
    panel.grid = element_blank(),
    axis.text = element_blank(),
    axis.title = element_blank(),
    axis.ticks = element_blank(),
    panel.border = element_rect(color = "black", fill = NA, size = 2)
  )
```

Statistically Significant GWR Coeffecients of Poverty on Homicide Rates in Mexico

Municipalities in Mexico classified by significance of Geographically Weighted Regression coefficients of poverty indeces on logged homicide rates (per 100,000 population), at the 90% confidence level.

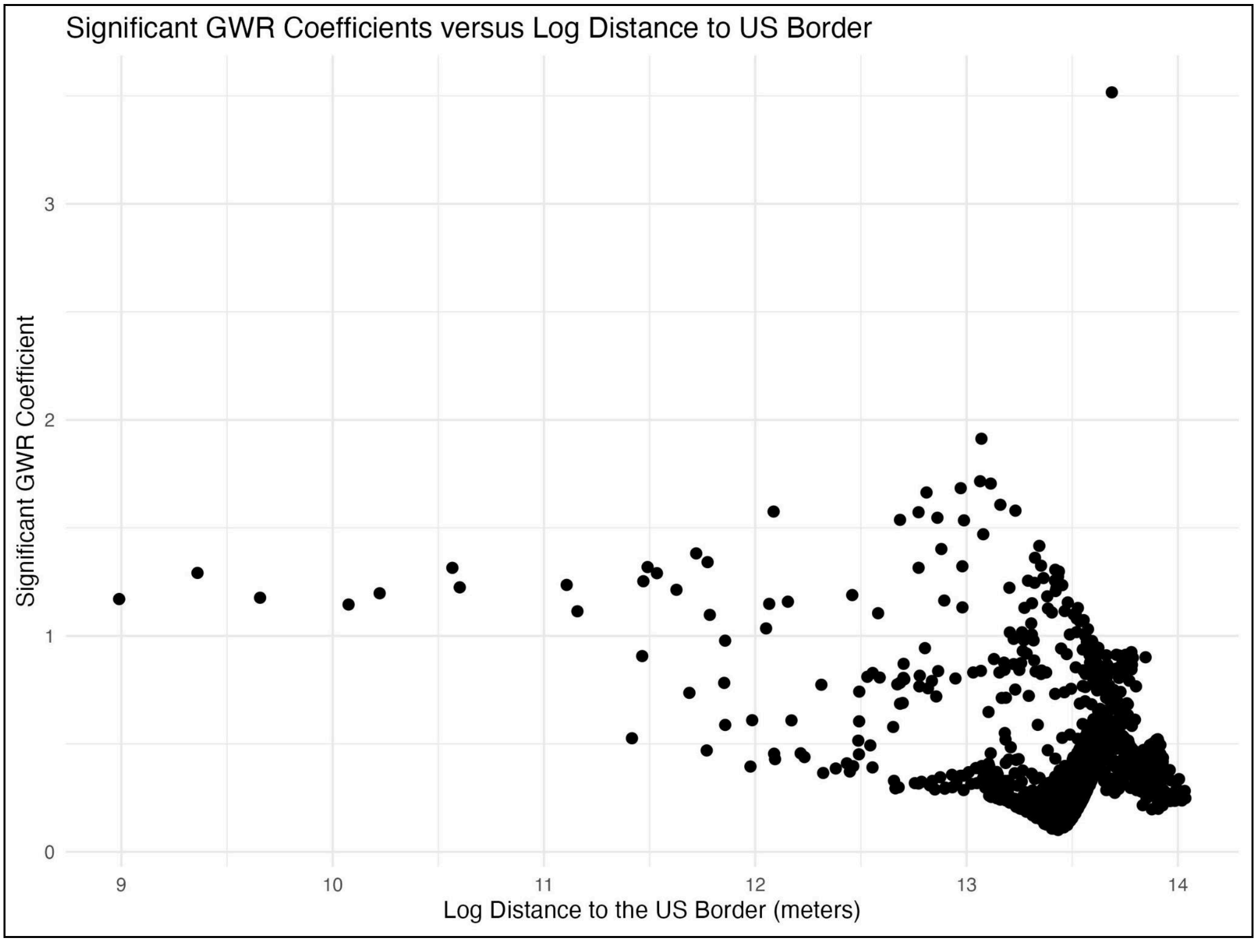


Green: Zero Homicides; Grey: Non-significant results; White: No Data.
Date: December 2024

The alpha level of 0.1, corresponding to a 90 percent confidence level, deems that the critical value of the t-distribution used in this analysis is approximately 1.645. This value was determined using the formula for a two-tailed test, which accounts for both positive and negative effects of the poverty index on homicide rates. The 90 percent confidence level was selected to strike a balance between identifying statistically significant relationships and maintaining a reasonable threshold for error. Municipalities with absolute t-statistics greater than or equal to 1.645 were classified as "Significant", while those below this threshold were labeled as "Not Significant". Municipalities lacking sufficient data were categorized as "No Data". This classification ensures that the map highlights areas where poverty index coefficients have statistically significant effects on homicide rates.

```r
# =================================
# Question 4: Significance Analysis
# =================================

# Create a new dataframe for this step
step4_data <- new_merged_data %>%
  st_join(
    gwr_data %>% select(CVE_ENT, CVE_MUN, indice_se),
    join = st_equals  # Ensure the spatial match respects municipality boundaries
  )

# Define critical t-value for 90% confidence level
alpha <- 0.1  # 90% confidence level
critical_t <- qt(1 - alpha / 2, df = Inf)  # Two-tailed test
print(paste("Critical t-value for 90% confidence level:", critical_t))

step4_data <- step4_data %>%
  mutate(
    # Combine indice_se columns
    indice_se = coalesce(indice_se.x, indice_se.y),
    # Calculate t-statistics and classify significance
    t_indice = ifelse(!is.na(indice_coef) & !is.na(indice_se), indice_coef / indice_se, NA),
    fill_category = case_when(
      data_status == "Zero Homicides" ~ "Zero Homicides",  # Prioritize Zero Homicides
      !is.na(t_indice) & abs(t_indice) >= critical_t ~ "Significant",
      !is.na(t_indice) & abs(t_indice) < critical_t ~ "Not Significant",
      TRUE ~ "No Data"
    )
  ) %>%
  # Drop redundant columns
  select(-indice_se.x, -indice_se.y)
```

```r
# Create the map
significant_coef_map <- ggplot() +
  # Significant coefficients layer
  geom_sf(
    data = step4_data %>% filter(fill_category == "Significant"),
    aes(fill = indice_coef),
    color = "gray80",
    size = 0.0005
  ) +
  scale_fill_viridis_c(
    option = "plasma",
    direction = -1,
    limits = c(-1, 4),
    breaks = seq(-1, 4, by = 1),
    name = "Significant GWR\nCoefficient (Poverty Index)"
  ) +
  # Zero homicide areas layer
  geom_sf(
    data = step4_data %>% filter(fill_category == "Zero Homicides"),
    fill = "forestgreen",
    color = "gray80",
    size = 0.0005
  ) +
  # Non-significant areas layer
  geom_sf(
    data = step4_data %>% filter(fill_category == "Not Significant"),
    fill = "grey70",
    color = "gray80",
    size = 0.0005
  ) +
  # No data areas layer
  geom_sf(
    data = step4_data %>% filter(fill_category == "No Data"),
    fill = "white",
    color = "gray80",
    size = 0.0005
  ) +
  labs(
    title = "Statistically Significant GWR Coeffecients of Poverty on Homicide Rates in Mexico",
    subtitle = "Municipalities in Mexico classified by significance of Geographically Weighted Regression coefficients of poverty indeces on logged homicide rates\n(per 100,000 population), at the 90% confidence level.",
    caption = "Green: Zero Homicides; Grey: Non-significant results; White: No Data.\nDate: December 2024"
  ) +
  # Add scale bar
  annotation_scale(
    location = "bl",  # Position: bottom left
    width_hint = 0.25 # Scale bar width relative to the plot
  ) +
  # Add north arrow
  annotation_north_arrow(
    location = "tl",  # Position: top left
    which_north = "true",  # Use true north
    style = north_arrow_fancy_orienteering()  # Choose arrow style
  ) +
  theme_minimal() +
  theme(
    legend.position = c(0.8, 0.8),
    legend.background = element_rect(fill = alpha("white", 0.8), color = NA),
    legend.key.size = unit(0.8, "lines"),
    legend.title = element_text(size = 12),
    legend.text = element_text(size = 10),
    plot.title = element_text(size = 14, face = "bold"),
    plot.subtitle = element_text(size = 11),
    panel.grid = element_blank(),
    axis.text = element_blank(),
    axis.title = element_blank(),
    axis.ticks = element_blank(),
    panel.border = element_rect(color = "black", fill = NA, size = 2)
  )
```



The analysis reveals a negative relationship between significant GWR coefficients for the poverty index and the log-transformed distance to the US border. This suggests that as municipalities are located farther from the US border, the impact of poverty on homicide rates weakens. Specifically, proximity to the border amplifies the relationship between poverty and homicide rates, which may reflect unique socioeconomic dynamics, such as migration patterns, cross-border trade, and varying law enforcement practices. This finding implies that interventions aimed at reducing poverty could have a more pronounced effect on reducing homicide rates in municipalities closer to the border, underscoring the need for targeted regional strategies.

```
Call:
lm(formula = indice_coef ~ log_distance, data = significant_with_distances)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.37118 -0.19076 -0.01225  0.07188  3.09708 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   3.19807    0.22561   14.18   <2e-16 ***
log_distance -0.20303    0.01672  -12.14   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2731 on 1242 degrees of freedom
Multiple R-squared:  0.1061,	Adjusted R-squared:  0.1053 
F-statistic: 147.4 on 1 and 1242 DF,  p-value: < 2.2e-16
```

```r
# ============================
# Question 5: Distance Analysis
# ============================

# Load the shapefile with distance to border
municipalities_with_distances <- st_read("~/Downloads/municipalities_with_distances/municipalities_with_distances.shp")

# Reproject the merged_data to a metric CRS (EPSG:6362 for Mexico)
municipalities_with_distances <- municipalities_with_distances %>%
  st_transform(6362)

# Filter for municipalities with significant coefficients
significant_data <- step4_data %>%
  filter(fill_category == "Significant") %>%
  select(CVE_ENT, CVE_MUN, indice_coef)

# Calculate centroids for significant_data
significant_data_centroids <- significant_data %>%
  st_centroid()

# Perform the join
significant_with_distances <- significant_data_centroids %>%
  st_join(municipalities_with_distances, join = st_nearest_feature)

# Log-transform distances
significant_with_distances <- significant_with_distances %>%
  mutate(log_distance = log(HubDist))

# Run the OLS regression
regression_model <- lm(indice_coef ~ log_distance, data = significant_with_distances)

# View the regression summary
summary(regression_model)

# Create scatter plot with regression line
scatter_plot <- ggplot(significant_with_distances, aes(x = log_distance, y = indice_coef)) +
  geom_point(color = "black", size = 2) +
  labs(
    title = "Significant GWR Coefficients versus Log Distance to US Border",
    x = "Log Distance to the US Border (meters)",
    y = "Significant GWR Coefficient"
  ) +
  theme_minimal()
```