# Data Visualization Contest - Analysis Code

### Mackenzie Gill, Shiyan Wang, and Courtney Cooper

### 2024-06-23

**Loading Data**

**Regression**

```
model_1<- lm(t_local_sales_2022_num ~ num_local_sellers_2022_2023 + MPI_count + PCT_LACCESS_POP10+ t_BII

print(summary(model_1))
```

```
##
## Call:
## lm(formula = t_local_sales_2022_num ~ num_local_sellers_2022_2023 +
##     MPI_count + PCT_LACCESS_POP10 + t_BIPOC_num_prods_2022 +
##     rural_RUCA, data = final_datavis_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -449.15  -36.56    5.62   38.13  258.65
##
## Coefficients:
##                              Estimate Std. Error    t value Pr(>|t|)
## (Intercept)                 -1.051e+02  1.630e+01 -6.447e+00 5.96e-10 ***
## num_local_sellers_2022_2023  1.000e+00  1.820e-07  5.493e+06  < 2e-16 ***
## MPI_count                   -2.937e+00  9.493e-01 -3.094e+00   0.0022 **
## PCT_LACCESS_POP10            3.702e-01  4.102e-01  9.030e-01   0.3676
## t_BIPOC_num_prods_2022      -5.095e-01  5.568e-02 -9.151e+00  < 2e-16 ***
## rural_RUCA                   7.034e+01  1.185e+01  5.937e+00 9.76e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 79.47 on 247 degrees of freedom
##   (255 observations deleted due to missingness)
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 8.701e+12 on 5 and 247 DF,  p-value: < 2.2e-16
```

```
# Making a table of results
tidy_model <- tidy(model_1)
conf_int <-confint_tidy(model_1) %>%
 rename(`95% Conf. Int. Lower` = conf.low, `95% Conf. Int. Upper` = conf.high)
```

```
## Warning: confint_tidy is now deprecated and will be removed from a future
## release of broom. Please use the applicable confint method.
```

```r
 conf_int$term <- tidy_model$term

 confint_reordered <-conf_int %>%
  select("term", "95% Conf. Int. Lower", "95% Conf. Int. Upper")

# Combine the results into a single table
results <- tidy_model %>%
  left_join(confint_reordered, by = "term")

# Print the results using stargazer
#stargazer(model_1, type = "text", ci = TRUE, ci.level = 0.95, single.row = TRUE,
#          title = "Regression Results",
#          dep.var.labels = "Dependent Variable: Local Sales in NWRM Region",
#          covariate.labels = c("Intercept", "Num. of Local Sellers in 2022 or 2023", "Meat Processing
 #          omit.stat = c("f", "ser"))
```

## Map of Local Sales in NWRM Region

```r
final_datavis_data$log_sales <- log(final_datavis_data$t_local_sales_2022_num)

# Loading US county shapefile from maps package
counties <- counties(cb = TRUE)  # cb = TRUE returns a simplified geometry
```

```
## Retrieving data for the year 2022
```

```
##    |                                                                |
```

```r
# Prepare shapefile data
# Use GEOID directly if it represents FIPS codes
counties <- counties %>%
  mutate(STCOUNTYFP = as.character(GEOID))  # Create FIPS column from GEOID
  setcolorder(counties, c("STCOUNTYFP", "STATEFP", "COUNTYFP", "COUNTYNS", "AFFGEOID", "GEOID", "NAME",

# Download state shapefile to filter by state names
states <- states(cb = TRUE)
```

```
## Retrieving data for the year 2021
```

```
##    |                                                                |
```

```r
# Define the states of interest
states_of_interest <- c("Colorado", "Wyoming", "Montana", "Idaho", "Washington", "Oregon")


# Filter counties data to include only counties from the selected states
filtered_counties <- counties %>%
```

```
    filter(STATE_NAME %in% states_of_interest)

  # Ensure FIPS codes are character
final_datavis_data$STCOUNTYFP <- as.character(final_datavis_data$STCOUNTYFP)

# Merge data with county shapefiles
merged_data <- filtered_counties %>%
  left_join(final_datavis_data, by = "STCOUNTYFP")

long_caption<-"This map illustrates the variation of local food sales for counties in the NWRM region. W
wrapped_caption<- str_wrap(long_caption, width = 80)

# Plot the map
ggplot(data = merged_data) +
  geom_sf(aes(fill = log_sales)) +
  scale_fill_viridis_c() +  # Use a color scale
  labs(title = "Map of Logged Local Food Sales by County in the NWRM Region",
       fill = "Value",
       caption = wrapped_caption) +
  theme_minimal()
```
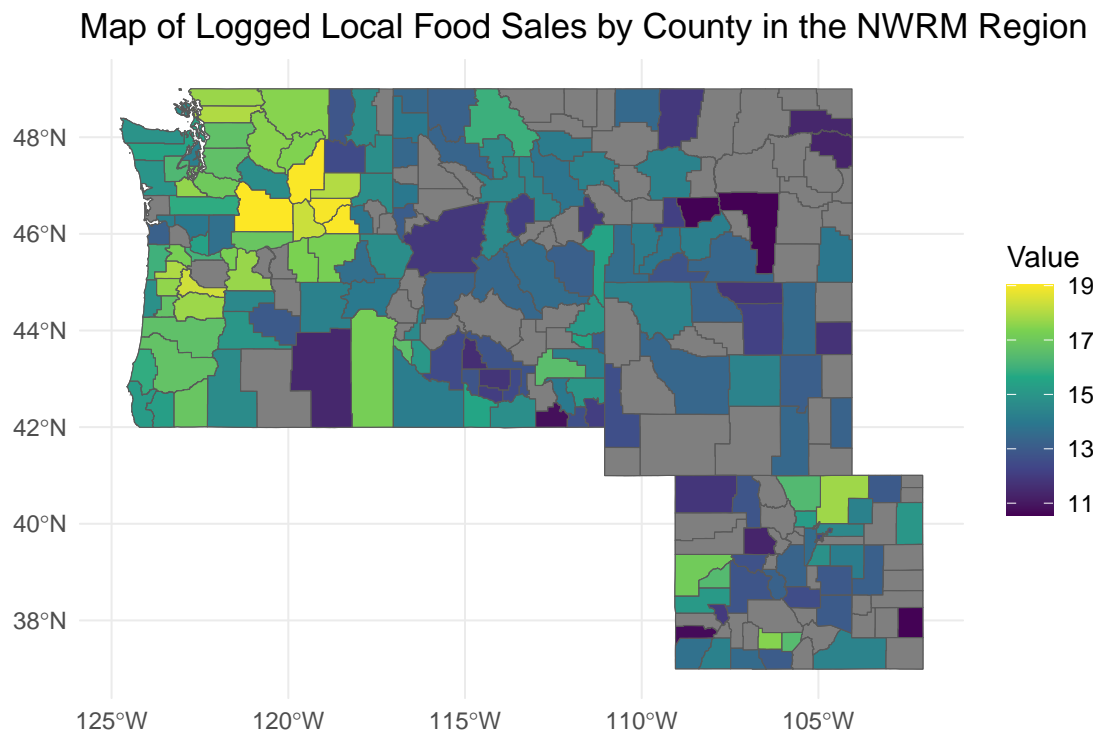
## Map of Logged Local Food Sales by County in the NWRM Region



This map illustrates the variation of local food sales for counties in the NWRM region. We chose to convey the log of sales because the untransformed sales data includes outliers, which when mapped, does not show the variation in the lower food sale quantiles.

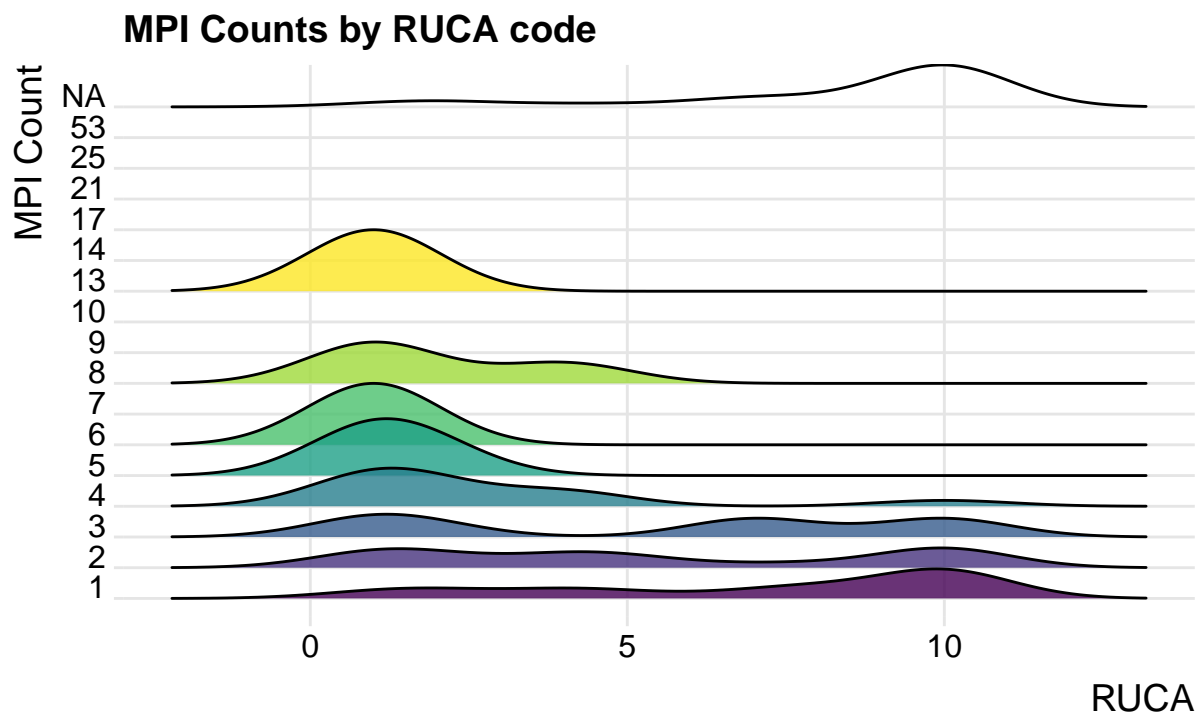**Density plot for MPI Counts across RUCA scores**

```
d <- data.frame(
  x = final_datavis_data$RUCA,
  y = final_datavis_data$MPI_count)

long_caption2<-"MPI= Meat Processing Institutes; RUCA= Rural-Urban Commuting Areas by USDA ERS, where:
wrapped_caption2<- str_wrap(long_caption2, width = 80)

ggplot(d, aes(x = x, y = factor(y), fill = factor(y))) +
  geom_density_ridges(alpha = 0.8, scale = 1) +
  labs(title = "MPI Counts by RUCA code",
       x = "RUCA",
       y = "MPI Count",
       caption = wrapped_caption2) +
  theme_ridges() +
    theme(legend.position = "none") +
  scale_fill_viridis_d(name = "MPI Count")
```

```
## Picking joint bandwidth of 1.06
```

```
## Warning: Removed 1 rows containing non-finite values ('stat_density_ridges()').
```



l= Meat Processing Institutes; RUCA= Rural–Urban Commuting Areas by USDA ERS,
where: 1= Metropolitan area core...4= Micropolitan area core...10= Rural area
core

# Number of Local Sellers

```
final_datavis_data$cats.num.sellers <- cut(final_datavis_data$num_local_sellers_2022_2023,
                          breaks = c(0, 15, 30, 45),
                          labels = c("Low: 0-15", "Medium: 16-30", "High: 30-45"),
                          right = FALSE)

d <- data.frame(
  x = final_datavis_data$RUCA,
  y = final_datavis_data$cats.num.sellers)

long_caption3<-"RUCA= Rural-Urban Commuting Areas by USDA ERS, where: 1= Metropolitan area core...4= Mic
wrapped_caption3<- str_wrap(long_caption3, width = 80)

ggplot(d, aes(x = x, y = factor(y), fill = factor(y))) +
  geom_density_ridges(alpha = 0.8, scale = 1) +
  theme(legend.position = none) +
  labs(title = "Counties' Local Seller Counts by RUCA code",
       x = "RUCA",
       y = "Num. Local Sellers in 2022 or 2023",
        caption = wrapped_caption3) +
  theme_ridges() +
    theme(legend.position = "none") +
  scale_fill_viridis_d(name = "Num. Local Sellers in 2022 or 2023")
```
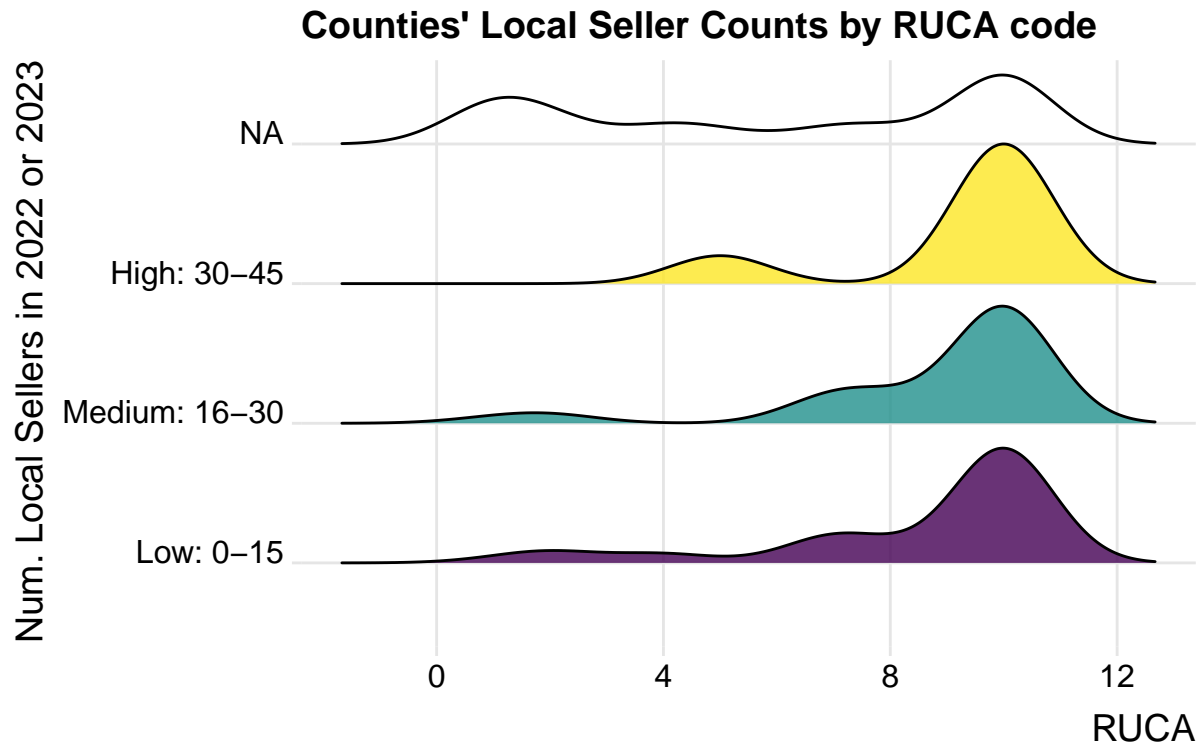
```
## Picking joint bandwidth of 0.89
```

```
## Warning: Removed 1 rows containing non-finite values ('stat_density_ridges()').
```

## Counties' Local Seller Counts by RUCA code



Num. Local Sellers in 2022 or 2023

RUCA

RUCA= Rural–Urban Commuting Areas by USDA ERS, where: 1= Metropolitan area core...4= Micropolitan area core...10= Rural area core

## Map of BIPOC sellers

```r
final_datavis_data$log_BIPOC <- log(final_datavis_data$t_BIPOC_num_prods_2022)

# Loading US county shapefile from maps package
counties <- counties(cb = TRUE)  # cb = TRUE returns a simplified geometry
```

## Retrieving data for the year 2022

```r
# Prepare shapefile data
# Use GEOID directly if it represents FIPS codes
counties <- counties %>%
  mutate(STCOUNTYFP = as.character(GEOID))  # Create FIPS column from GEOID
  setcolorder(counties, c("STCOUNTYFP", "STATEFP", "COUNTYFP", "COUNTYNS", "AFFGEOID", "GEOID", "NAME",

# Download state shapefile to filter by state names
states <- states(cb = TRUE)
```

## Retrieving data for the year 2021

```r
# Define the states of interest
states_of_interest <- c("Colorado", "Wyoming", "Montana", "Idaho", "Washington", "Oregon")
```

```r
# Filter counties data to include only counties from the selected states
filtered_counties <- counties %>%
  filter(STATE_NAME %in% states_of_interest)

  # Ensure FIPS codes are character
final_datavis_data$STCOUNTYFP <- as.character(final_datavis_data$STCOUNTYFP)

# Merge data with county shapefiles
merged_data <- filtered_counties %>%
  left_join(final_datavis_data, by = "STCOUNTYFP")

long_caption4<-"BIPOC producers included 2022 counts of Black, American Indian and Alaskan, Hawaiian Nat
wrapped_caption4<- str_wrap(long_caption4, width = 80)

# Plot the map
ggplot(data = merged_data) +
  geom_sf(aes(fill = log_BIPOC)) +
  scale_fill_viridis_c() +  # Use a color scale
  labs(title = "Map of Logged BIPOC Producer Numbers by County",
       fill = "Value",
       caption = wrapped_caption4) +
  theme_minimal()
```
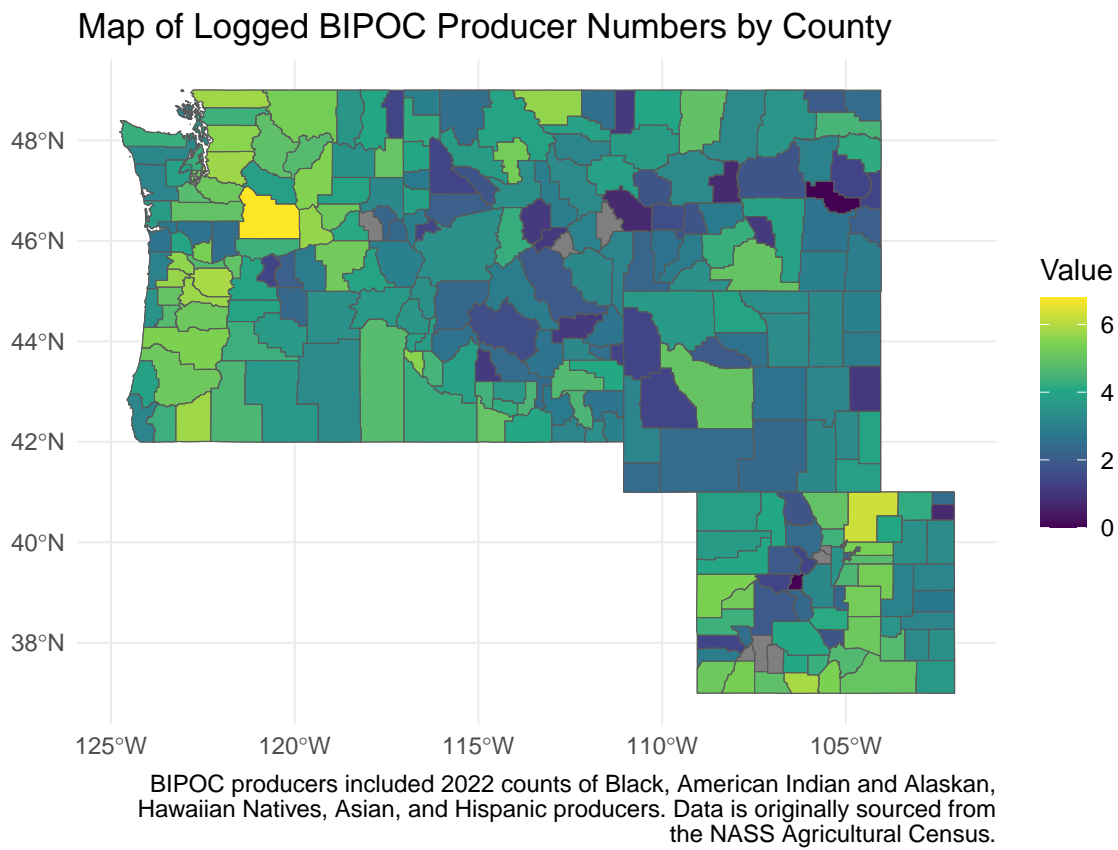


Map of Logged BIPOC Producer Numbers by County

BIPOC producers included 2022 counts of Black, American Indian and Alaskan, Hawaiian Natives, Asian, and Hispanic producers. Data is originally sourced from the NASS Agricultural Census.

## Density Plot of BIPOC producers across RUCA codes

```
final_datavis_data$cats.num.BIPOC <- cut(final_datavis_data$t_BIPOC_num_prods_2022,
                          breaks = c(0, 13, 30, 76, 883),
                          labels = c("Quantile 1: 0-13", "Quantile 2: 13-30", "Quantile 3: 30-76", "Qua
                          right = FALSE)

d <- data.frame(
  x = final_datavis_data$RUCA,
  y = final_datavis_data$cats.num.BIPOC)

long_caption5<-"RUCA= Rural-Urban Commuting Areas by USDA ERS, where: 1= Metropolitan area core...4= Mic
wrapped_caption5<- str_wrap(long_caption5, width = 80)

ggplot(d, aes(x = x, y = factor(y), fill = factor(y))) +
  geom_density_ridges(alpha = 0.8, scale = 1) +
  labs(title = "Number of BIPOC Producers by RUCA codes",
       x = "RUCA",
       y = "Num. of BIPOC Producers in 2022",
       caption = wrapped_caption5) +
  theme_ridges() +   theme(legend.position = "none") +
  scale_fill_viridis_d(name = "Num. of BIPOC Producers in 2022")
```
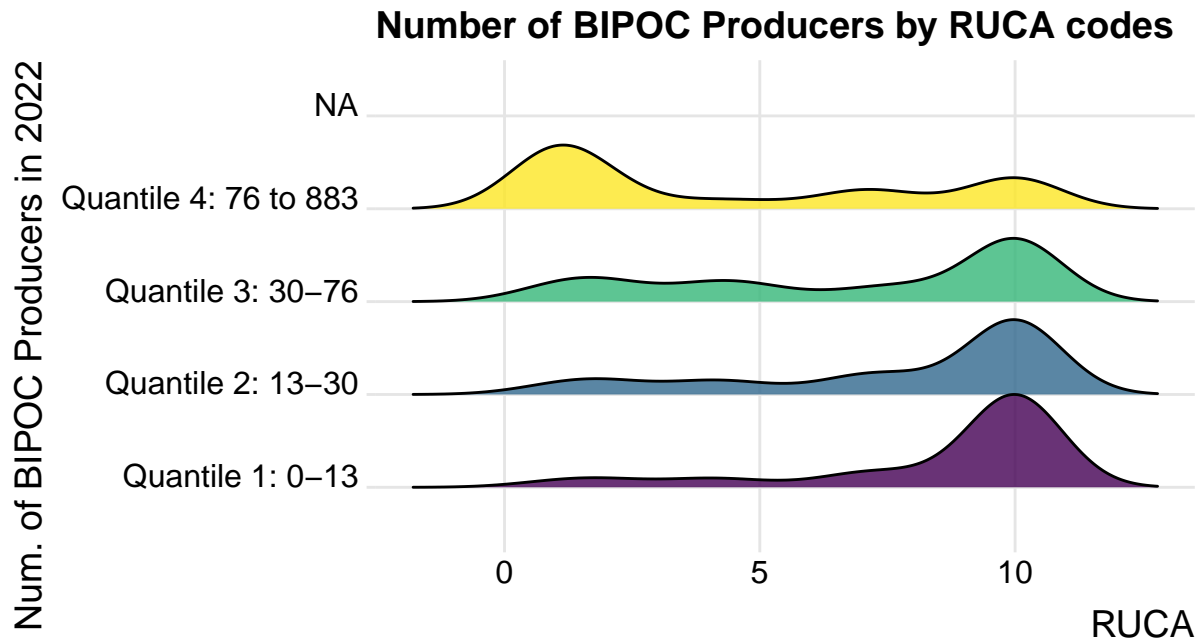
```
## Picking joint bandwidth of 0.93
```

```
## Warning: Removed 1 rows containing non-finite values ('stat_density_ridges()').
```

**Number of BIPOC Producers by RUCA codes**



RUCA= Rural–Urban Commuting Areas by USDA ERS, where: 1= Metropolitan area core...4= Micropolitan area core...10= Rural area core. BIPOC producers included 2022 counts of Black, American Indian and Alaskan, Hawaiian Natives, Asian, and Hispanic producers. Data is originally sourced from the NASS Agricultural Census.

## Observing correlation between Percent of Pop. at Low Access and Local Sales

```
ggplot(final_datavis_data, aes(x = PCT_LACCESS_POP10, y = t_local_sales_2022_num)) +
  geom_hex(bins= 15) +
  scale_fill_gradient(low = "blue", high = "red") +
  labs(title = "Correlations of Low Access Areas and Local Sales",
       x = "% of Pop. at Low Access",
       y = "Local Food Sales, in $",
       fill = "Count of Obs. per Bin") +
  theme_minimal()
```

```
## Warning: Removed 1 rows containing non-finite values ('stat_binhex()').
```

Correlations of Low Access Areas and Local Sales