

Congestive Heart Failure Analysis

Jane Condon

2024-03-21

Is There a Connection Between Congestive Heart Failure and Socioeconomic Status (and/or) Urbanization Level?

Loading packages

```
library(tidyverse)
library(ggplot2)
library(sjPlot)
library(formatR)
```

Importing Data

```
wonder_data <- read_csv("CHF Data.csv")
education_data <- read_csv("Education Data.csv")
poverty_data <- read_csv("Poverty and Household Income Data.csv")
food_insecurity_data <- read_csv("Food Insecurity Data.csv")
unemployment_data <- read_csv("Unemployment Data.csv")
```

Merging Data

```
full_data <- wonder_data %>%
  inner_join(education_data, by = c("County", "State")) %>%
  inner_join(poverty_data, by = c("County", "State")) %>%
  inner_join(food_insecurity_data, by = c("County", "State")) %>%
  inner_join(unemployment_data, by = c("County", "State"))
```

Cleaning the Data

Changing Some Variable Names For Simplicity

```
colnames(full_data)[colnames(full_data) == "Crude Rate"] <- "rate"
colnames(full_data)[colnames(full_data) == "Percent of adults with less than a high school diploma, 2008-12"] <- "no_hs"
colnames(full_data)[colnames(full_data) == "Percent of adults with a high school diploma only, 2008-12"] <- "hs_only"
colnames(full_data)[colnames(full_data) == "Percent of adults with a bachelor's degree or higher, 2008-12"] <- "bs_higher"
colnames(full_data)[colnames(full_data) == "Poverty Percent, All Ages"] <- "poverty"
colnames(full_data)[colnames(full_data) == "Median Household Income"] <- "income"
colnames(full_data)[colnames(full_data) == "2014 Food Insecurity Rate"] <- "food_insecure"
colnames(full_data)[colnames(full_data) == "Rural-Urban Continuum Code 2013"] <- "urbanization"
```

Only Keeping Necessary Variables In the Dataframe

```
df <- full_data[c("County", "State", "Deaths", "Population",
  "rate", "no_hs", "hs_only", "bs_higher", "poverty", "income",
  "food_insecure", "urbanization")]
```

Removing Missing Values and Fixing Data Types

```
# Removing missing values

df <- na.omit(df)

# Removing percent sign from food_insecure

df$food_insecure <- as.numeric(gsub("[\\%,]", "", df$food_insecure))

# Removing commas from income

df$income <- as.numeric(gsub(",", "", df$income))

# Changing poverty to a numeric

df$poverty <- as.numeric(df$poverty)

# Changing urbanization to a factor

df$urbanization <- as.factor(df$urbanization)
```

Exploratory Data Analysis and Data Visualization

Looking At the Summary Statistics For Crude Mortality Rate and Looking For Outliers

```
summary(df$rate)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   19.20   27.60   32.27   39.58   309.60
```

```
summary(df$Deaths)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0    31.0    64.0   162.7   141.0 10523.0
```

Removing Large Outliers Using IQR*1.5

```
quartiles <- quantile(df$rate, probs = c(0.25, 0.75), na.rm = FALSE)
IQR <- IQR(df$rate)

lower <- quartiles[1] - 1.5 * IQR
upper <- quartiles[2] + 1.5 * IQR

df_new <- subset(df, df$rate > lower & df$rate < upper)
```

Attaching New Dataframe Without Outliers

```
attach(df_new)
```

New Summary Statistics After Removing Outliers

```
summary(rate)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   18.80   26.80   29.09   37.40   70.10
```

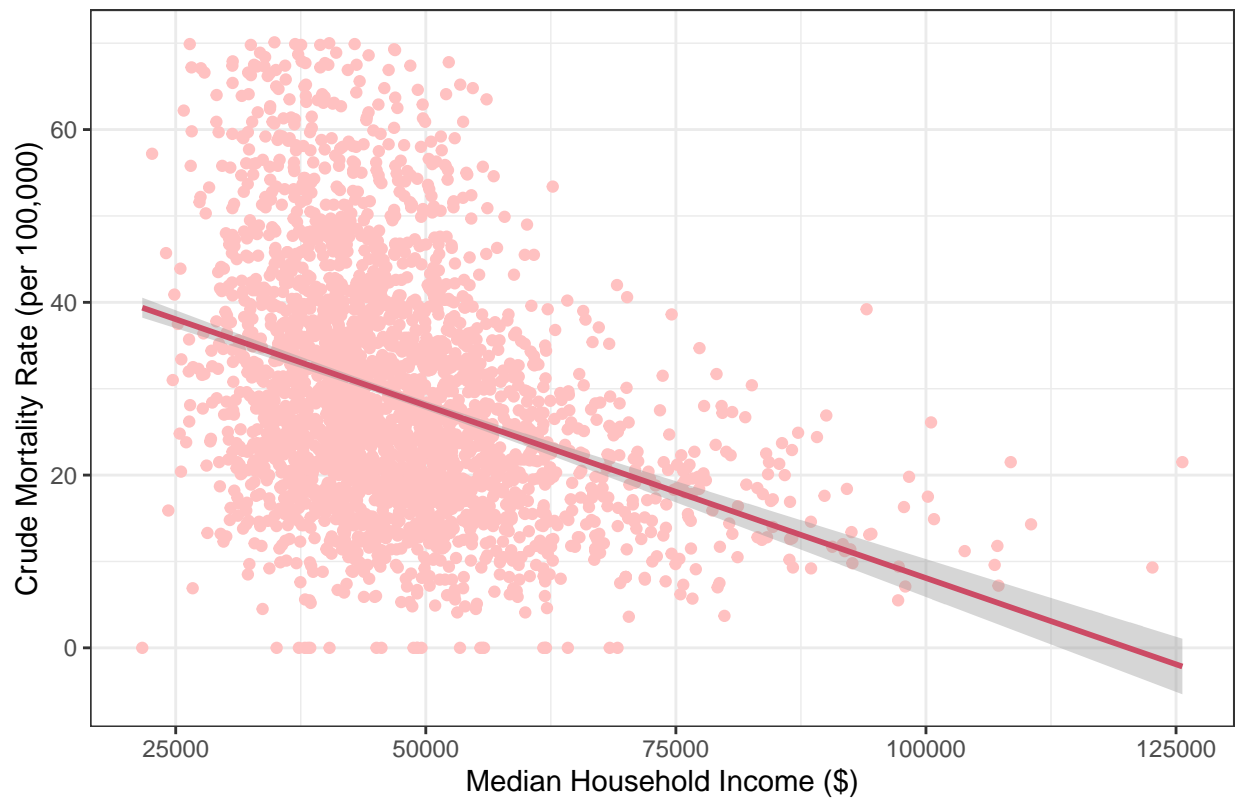
Data Visualization

Crude Mortality Rate vs. Household Income

```
ggplot(df_new, aes(x = income, y = rate)) + geom_point(color = "rosybrown1") +
  geom_smooth(method = "lm", color = "#cc4b65") + labs(title = "Median Household Income vs. Congestiv
  x = "Median Household Income ($)", y = "Crude Mortality Rate (per 100,000)") +
  theme_bw()
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

Median Household Income vs. Congestive Heart Failure



```
cor(income, rate)
```

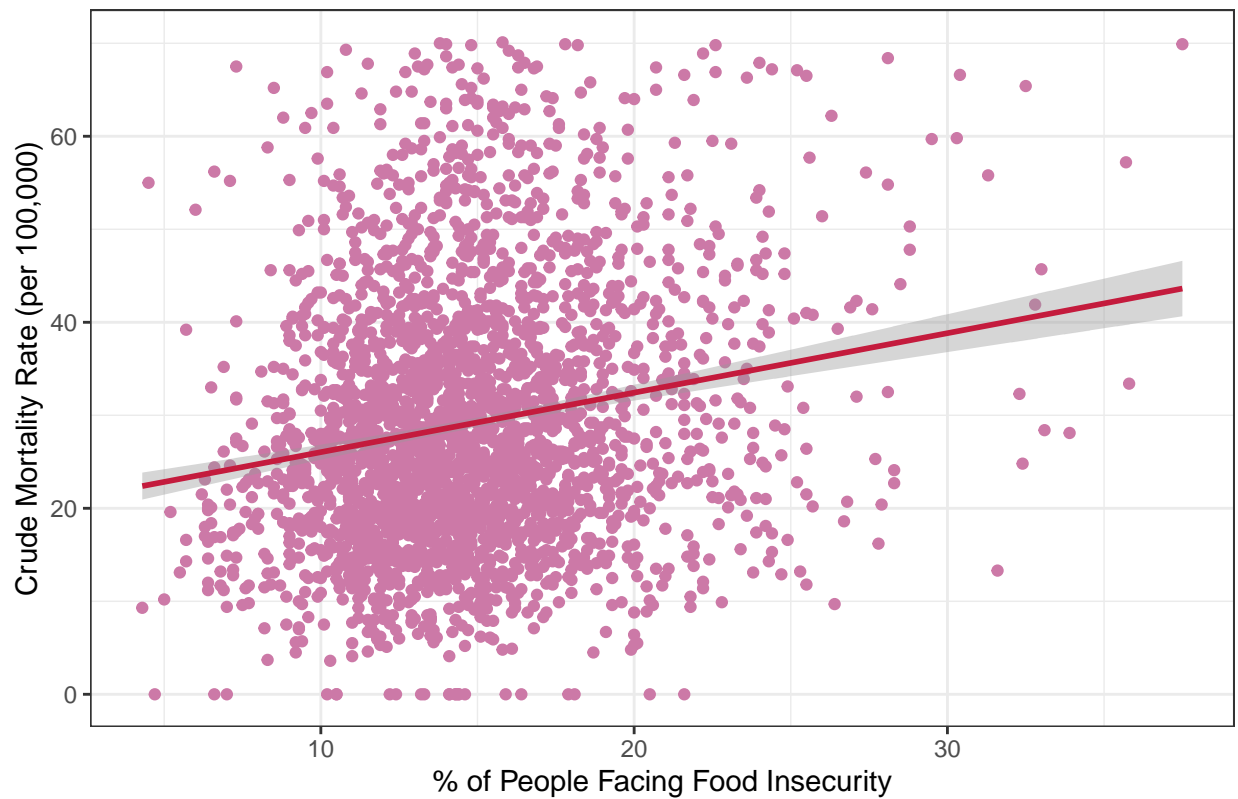
```
## [1] -0.3493774
```

Crude Mortality Rate vs Food Insecurity

```
ggplot(df_new, aes(x = food_insecure, y = rate)) + geom_point(color = "#CC79A7") +  
  geom_smooth(method = "lm", color = "#c41b3d") + labs(title = "Food Insecurity vs. Congestive Heart Failure",  
    x = "% of People Facing Food Insecurity", y = "Crude Mortality Rate (per 100,000)") +  
  theme_bw()
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

Food Insecurity vs. Congestive Heart Failure



```
cor(food_insecure, rate)
```

```
## [1] 0.1846869
```

Crude Mortality Rate vs. Education Level

```
ggplot(df_new, aes(x = bs_higher, y = rate)) + geom_point(color = "#edb4e4") +  
  geom_smooth(method = "lm", color = "#eb5081") + labs(title = "Education Level vs. Congestive Heart Failure",  
  x = "% of People With a Bachelor's Degree or Higher", y = "Crude Mortality Rate (per 100,000)") +  
  theme_bw()
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

Education Level vs. Congestive Heart Failure (Bachelor's or Higher)



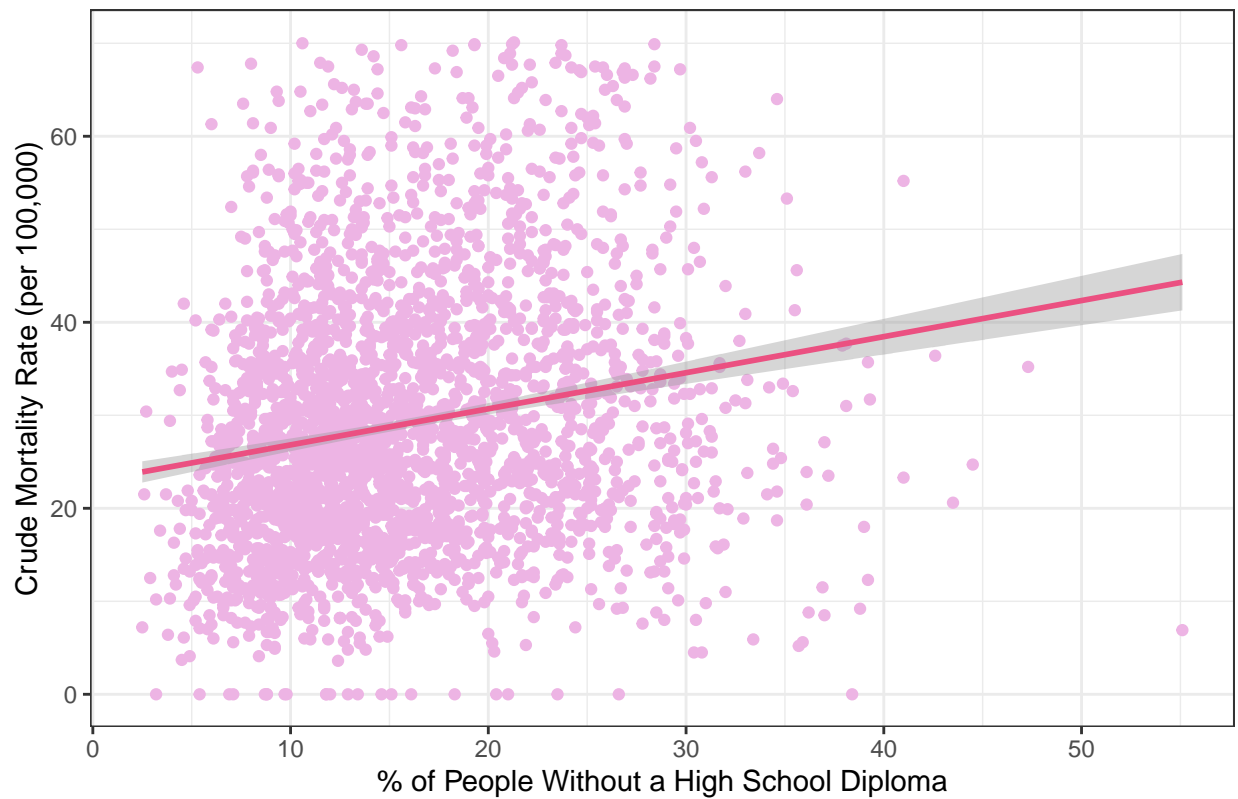
```
cor(bs_higher, rate)
```

```
## [1] -0.3374879
```

```
ggplot(df_new, aes(x = no_hs, y = rate)) + geom_point(color = "#edb4e4") +  
  geom_smooth(method = "lm", color = "#eb5081") + labs(title = "Education Level vs. Congestive Heart Failure",  
    x = "% of People Without a High School Diploma", y = "Crude Mortality Rate (per 100,000)") +  
  theme_bw()
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

Education Level vs. Congestive Heart Failure (Less than High School)



```
cor(no_hs, rate)
```

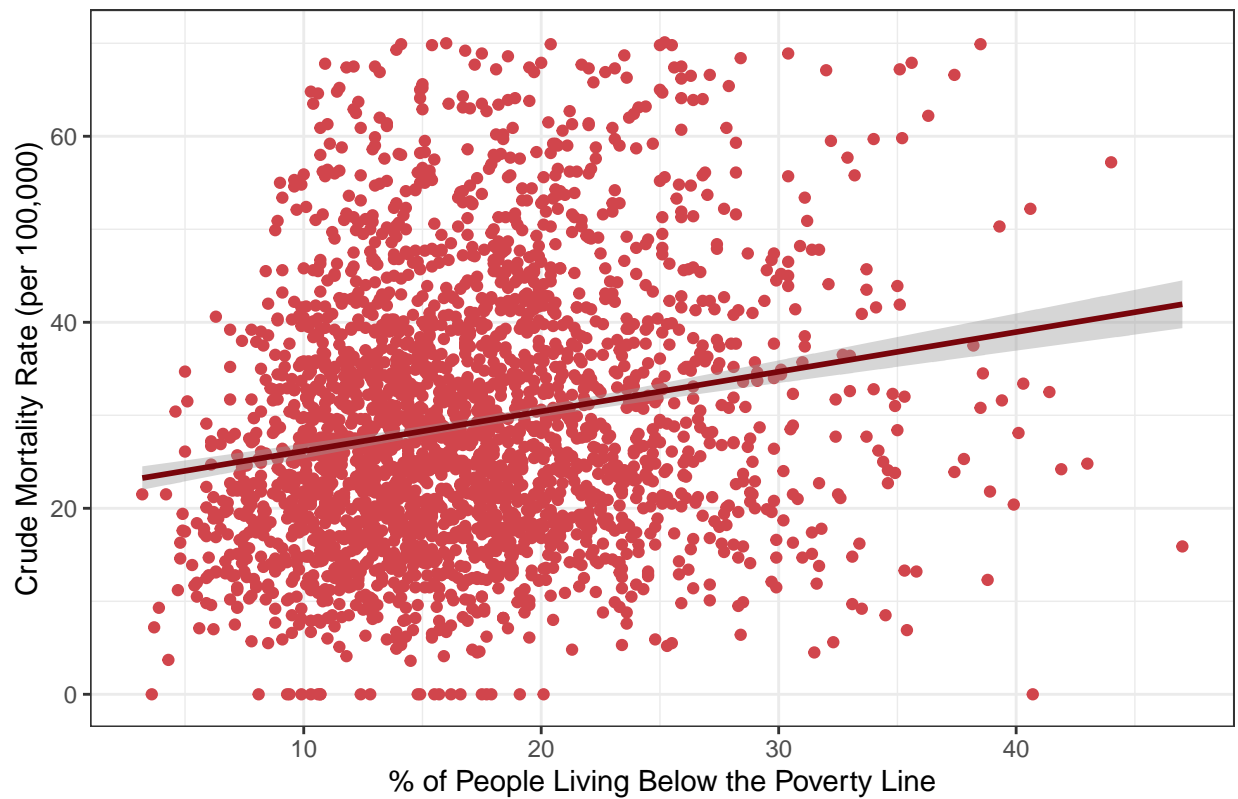
```
## [1] 0.1897416
```

Crude Mortality Rate vs. Poverty Level

```
ggplot(df_new, aes(x = poverty, y = rate)) + geom_point(color = "#d1454c") +  
  geom_smooth(method = "lm", color = "#78050b") + labs(title = "Poverty vs. Congestive Heart Failure"  
  x = "% of People Living Below the Poverty Line", y = "Crude Mortality Rate (per 100,000)") +  
  theme_bw()
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

Poverty vs. Congestive Heart Failure



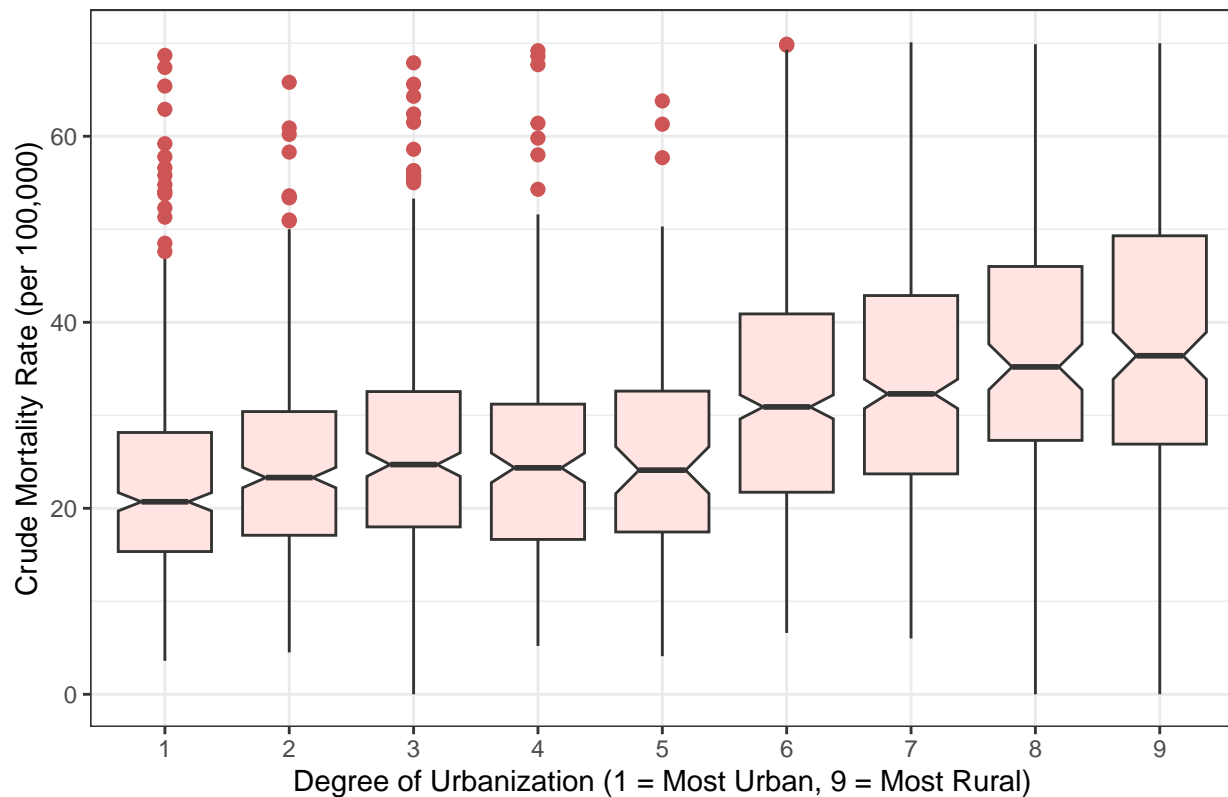
```
cor(poverty, rate)
```

```
## [1] 0.1903578
```

Crude Mortality Rate vs. Urbanization Level

```
ggplot(df_new, aes(x = urbanization, y = rate)) + geom_boxplot(fill = "mistyrose",  
  notch = TRUE, outlier.colour = "indianred3", outlier.size = 2) +  
  labs(title = "Urbanization Level vs. Congestive Heart Failure",  
    x = "Degree of Urbanization (1 = Most Urban, 9 = Most Rural)",  
    y = "Crude Mortality Rate (per 100,000)") + theme_bw()
```


Urbanization Level vs. Congestive Heart Failure



Linear Regression Model

```
model <- lm(rate ~ income + food_insecure + bs_higher + poverty +
  urbanization, data = df_new)
summary(model)
```

```
##
## Call:
## lm(formula = rate ~ income + food_insecure + bs_higher + poverty +
##   urbanization, data = df_new)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -39.557  -8.971  -1.270   7.023  43.731
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.387e+01  3.384e+00  15.921  < 2e-16 ***
## income       -3.961e-04  4.651e-05  -8.516  < 2e-16 ***
## food_insecure  3.650e-01  1.001e-01   3.647  0.00027 ***
## bs_higher    -1.614e-01  4.007e-02  -4.028  5.78e-05 ***
## poverty      -5.536e-01  8.006e-02  -6.916  5.80e-12 ***
## urbanization2 -2.542e+00  9.523e-01  -2.670  0.00764 **
```

```
## urbanization3 -1.368e+00  9.994e-01  -1.369  0.17124
## urbanization4 -2.702e+00  1.154e+00  -2.341  0.01930 *
## urbanization5 -2.146e+00  1.519e+00  -1.413  0.15778
## urbanization6  2.690e+00  9.600e-01   2.802  0.00512 **
## urbanization7  4.157e+00  1.024e+00   4.058  5.09e-05 ***
## urbanization8  6.249e+00  1.324e+00   4.721  2.47e-06 ***
## urbanization9  6.586e+00  1.220e+00   5.397  7.37e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.62 on 2661 degrees of freedom
## Multiple R-squared:  0.1985, Adjusted R-squared:  0.1948
## F-statistic:  54.9 on 12 and 2661 DF,  p-value: < 2.2e-16
```

Creating Table Displaying OLS Results

```
ols_table <- tab_model(model, digits = 3, show.ci = F, show.stat = T,
  show.p = T, show.se = T, p.style = "stars", pred.labels = c("(Intercept)",
    "Median Household Income", "Food Insecurity", "Bachelor's Degree or Higher",
    "Poverty", "Urbanization Level 2", "Urbanization Level 3",
    "Urbanization Level 4", "Urbanization Level 5", "Urbanization Level 6",
    "Urbanization Level 7", "Urbanization Level 8", "Urbanization Level 9"),
  dv.labels = c(""))
print(ols_table)
```