Exploring Ties Between Minority Concentration and Food Insecurity: A Beginner's Regression Analysis Project

Aaron Phalin

Feburary, 24, 2018

Abstract

In this project, I analyze a data-set provided to me for my Econometrics class. The validity of this data is not proven, and so I will not be attempting to make any real-world claims with the results from the research question explored here. I will be extracting measurements in the percentage of food insecurity, minority populations, and various controls statistics from the data to prove a relationship between minority population percentage and food insecurity across US counties. This research reflects on a central motive in the modern food justice movement: to reallocate and redistribute nutritious, healthy foods to minority and low-income populations. In my results, I found that this data shows a very strong relationship between the percentage of black population and the amount of food insecurity in a US county. Communities with higher percentages of black citizens are found to have disproportionately higher food insecurity rates.

1 The Data and Reseach Question

This data comes from compiled datasets across the County Health Rankings, the CDC, and the American Community Survey. The data contains health measures and indicators by counties across the United States. This particular set of information may have been manipulated or tweaked for the purposes of the class.

In addition to this dataset, I included information by NOAA on drought risk across US counties. NOAA routinely classifies US counties into four different levels of risk based on levels of precipitaton and previous drought events: low, moderate, high, and extreme. I took the counties with high and extreme risk of drought and those with moderate to low risk of drought into two separate groups. All data comes from the range of 2013-2016. The compiled data set contains information throughout that range, while the NOAA drought risk data looks at 2016 measurements.

The research question I am asking in this dataset analysis is *"do counties with higher minority concentration tend to be areas of disproportionately more food insecurity?"*.

2 Variables and Visualizations

In this analysis, I look at a number of useful variables. Below I will begin to list the variables in detail, and provide summarizing maps of the information.

Statistic	N	Mean	St. Dev.	Min	Max
% Food Insecure	3,136	14.671	4.138	4.300	37.500
% Black Pop	3,136	8.964	14.258	0.000	84.455
% Hispanic Pop	3,136	9.192	13.587	0.205	95.824
% Unemployed	3,135	5.521	1.986	1.794	24.008
% Lands Rural	3,136	58.589	31.480	0.000	100.000
Drought Risk (1=Risk)	3,136	0.349	0.477	0	1

Table 1: Summary Statistics

Food Insecurity: The food insecurity variable looks at several factors in a community. The most important aspect of this measurement is the distance/ease of commute for citizens to reach foods that are nutritious and grown safely. To contextualize this definition, a population that has at least a 20 mile commute to the nearest grocery store but is within close range of multiple fast food chains, like McDonalds or Taco Bell, will be considered a population under food insecurity. Nutrition and ease of access are important factors. The variable measures the statistic across all US counties, and the percentage of overall population that belongs to a "food drought", or regions of low ease of access to nutritious and healthy foods.

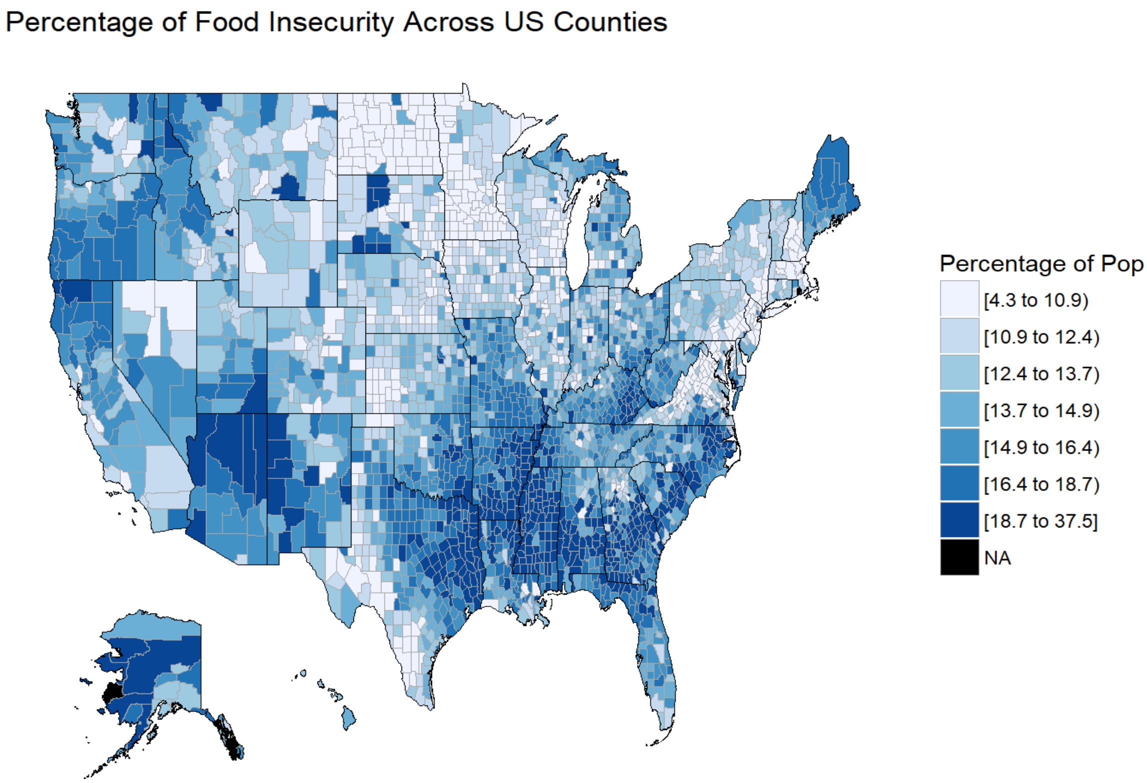
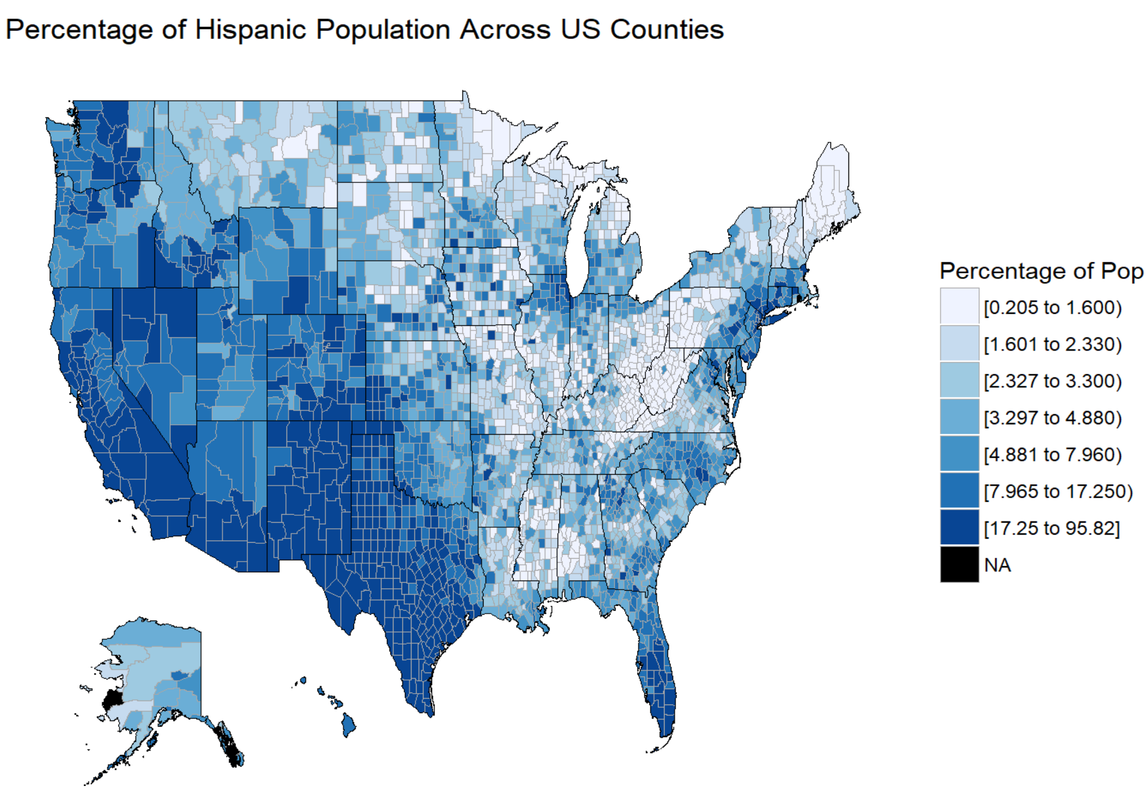


Figure 1: Percentage of Food Insecurity Across US Counties

Percentage Black and Hispanic: The dataset provides information on two important and large minority communities in the United States. In this analysis, I will be looking at two separate variables for minority concentrations: the percentage of black citizens and the percentage of hispanic citizens by county.



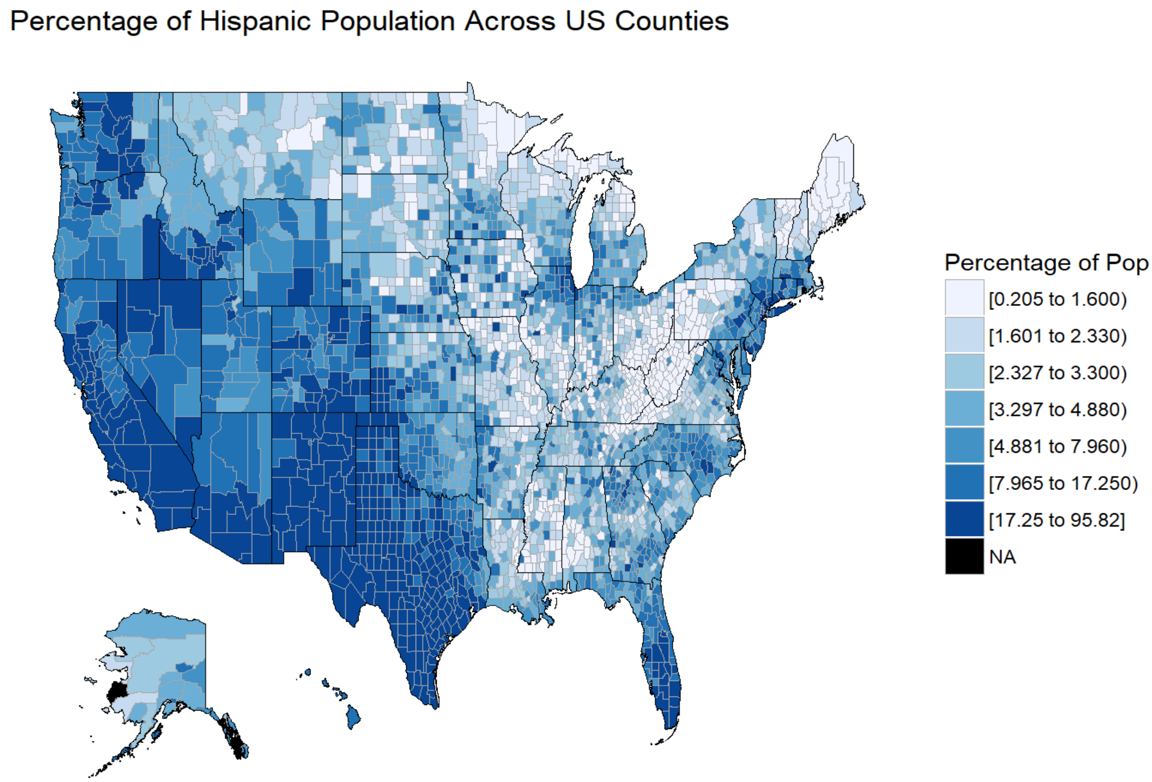


Figure 2, 3: Percentage of Black and Hispanic Population Across US Counties

Percentage of Rural Lands: The amount of rural land versus urban land in a territory is another important variable to consider. This has a bit more of a nuanced relationship with our dependent variable. Intuitively, people who live in areas of lower development concentration (more "rural" lands) will have a longer commute to get things by nature of their location. This longer commute *may* result unintendedly in a higher percentage of food insecurity. To control for this, I included the percentage of rural lands as a control variable. What is classified as "rural" is land that is not planned or drafted for urban business/development, and is left to either plains, forest, or agriculture. The variable looks at the percentage of these lands which are rural by county.

Percentage Unemployed: This variable is included as an important control. I predicted that the amount of unemployment will play an important role in the amount of food insecurity in an area. Since I am interested in the relationship between minority concentration and food insecurity, I included the unemployment rate to better single out this relationship and piece out which of the effect is due to unemployment rate (and the implied poverty) and actual minority concentration.

Drought Risk: This variable is an important dummy variable in my equation. Extracted from a NOAA dataset, I took all the US counties that experience a considerable drought risk against those which did not. As the production of local food intuitively plays an important role in how stable a food community is, I found that this data would fit into the regression and help us gain deeper insight on the complicated relationship.

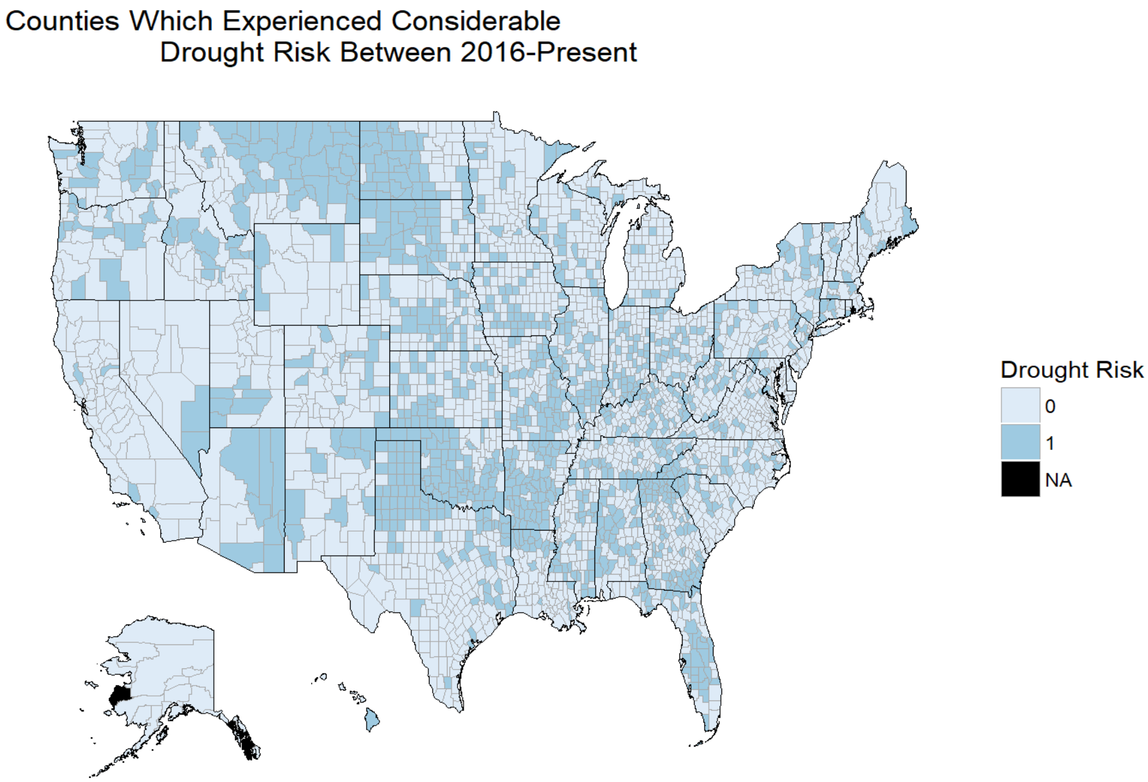


Figure 4: Drought Risk (1 = At Risk) Across US Counties

Omitted Variables: In this analysis, I omitted a variable for the pecentage of white population in a county. This is to shield from too great a collinearity among the independent variables, since all major ethnic communities in the United States would then be described in the regression (with the exception of smaller populations, like Native Americans). For example, if I looked at a county and broke down the percentage of black and white populations in a county, we would know too much about the amount of hispanic citizens in the county and could get too close an estimate since the remaining ethnic groups are of a smaller percentage. This would lead to a violation in an important assumption for OLS.

Interestingly, however, when we graph the percentage of white population against the percentage of food insecurity, we already get a foreshadow for the results of the regression.

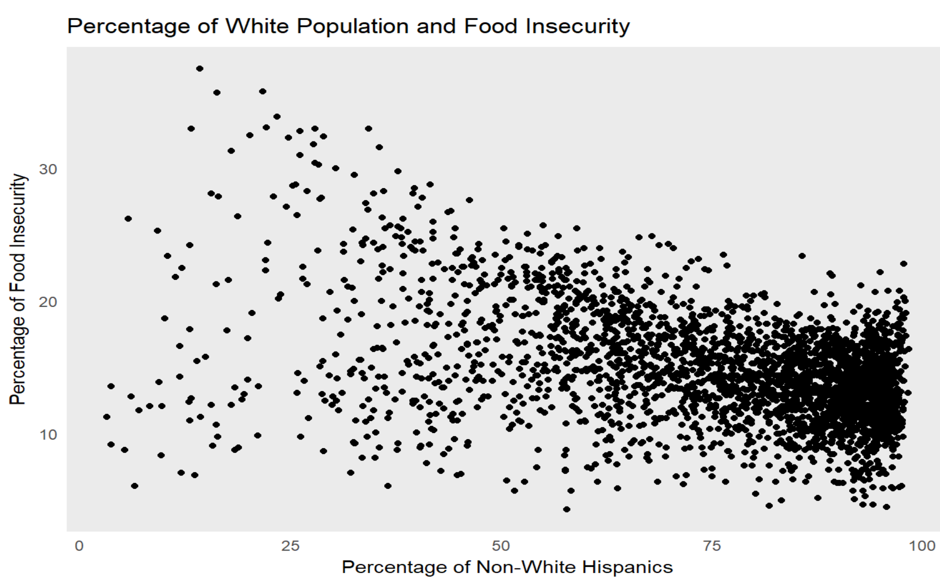


Figure 5: Scatterplot on percentage of white population against food insecurity

3 Model and Regression

The model we used in this regression analysis was this:

$PctFoodInsecurity = \beta_0 + \beta_1 * PctBlackPop + \beta_2 * PctHispPop + \beta_3 * PctUnemployedPop + \beta_4 * PctRuralLands + \beta_5 * DroughtRisk + \upsilon$

This model reflects what I predicted of the relationship in the real world. In particular, I include a dummy variable for the drought risk but the rest of the variables we assumed a linear relationship between the independent and the percentage of food insecure populations. In scatterplots like the ones drawn for the percentage of white population, I saw a linear line of fitness that would be most appropriate. In this regression, as well, I used the OLS method.

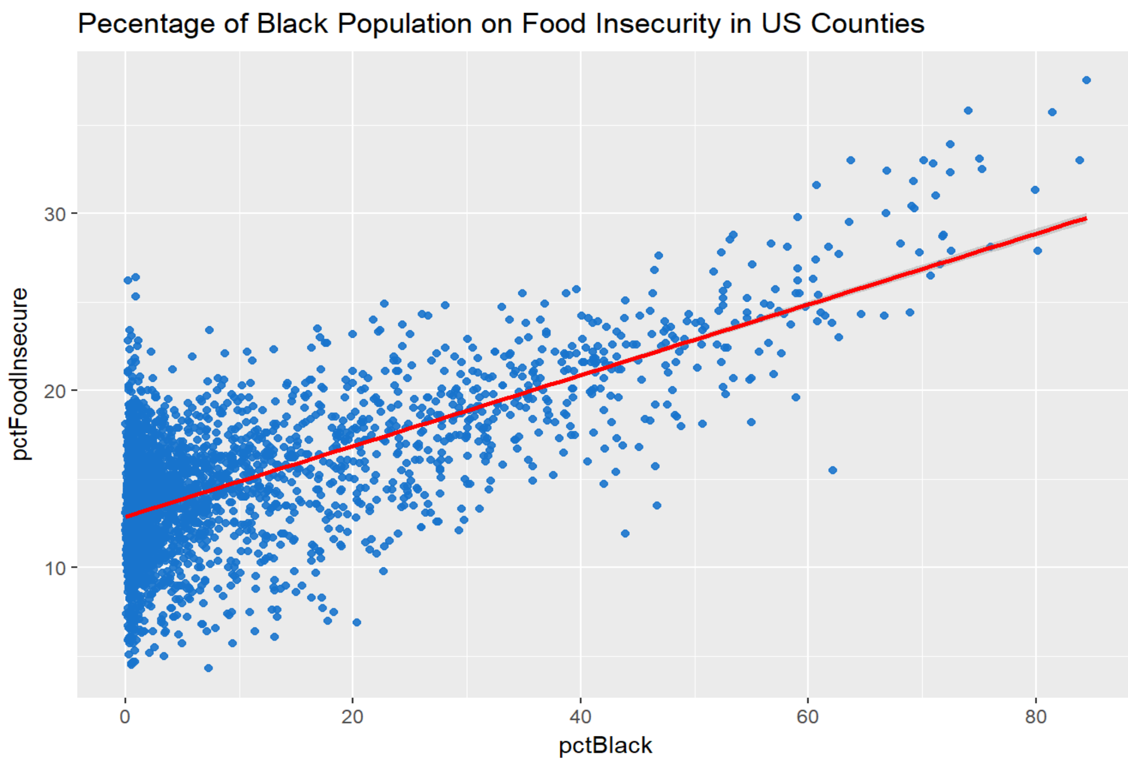
In running the regression model, I chose a method of slowly adding independent variables to grow the model in each step. I recorded each steps coefficient estimates and then took note of the changes in the adjusted R squared. This models came to these results:

Regression Results

	Dependent variable:			
	pctFoodInsecure			
	(1)	(2)	(3)	(4)
% Black in Pop	0.198^{***}	0.152^{***}	0.156^{***}	0.156^{***}
	(0.004)	(0.004)	(0.004)	(0.004)
% Hispanic in Pop	-0.020^{***}	-0.028^{***}	-0.022^{***}	-0.022^{***}
	(0.004)	(0.003)	(0.004)	(0.004)
% Unemployed		0.831^{***}	0.813^{***}	0.823^{***}
		(0.025)	(0.025)	(0.025)
% Rural			0.008^{***}	0.007^{***}
			(0.002)	(0.002)
Drought Risk (DV)				0.553^{***}
				(0.096)
Constant	13.081^{***}	8.966^{***}	8.530^{***}	8.332^{***}
	(0.075)	(0.140)	(0.164)	(0.167)
Observations	3,136	3,135	3,135	3,135
R^2	0.478	0.613	0.617	0.621
Adjusted R^2	0.477	0.613	0.616	0.620
Residual Std. Error	2.992 (df = 3133)	2.574 (df = 3131)	2.564 (df = 3130)	2.551 (df = 3129)
F Statistic	$1,432.557^{***}$ (df = 2; 3133)	$1,656.098^{***}$ (df = 3; 3131)	$1,258.042^{***}$ (df = 4; 3130)	$1,023.341^{***}$ (df = 5; 3129)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 2: Regression results

This model reveals an interesting relationship. In particular, it is found that as black population increases in a county, there is a predicted effect and significant effect of increasing food insecurity. This result proves to statistically significant at any significance level. When we examine the results of hispanic population percentage of food insecurity, we observed the opposite effect. There is a slight decrease in food insecurity as hispanic population percentages increase. This, like the statistic we observed in black population percentages, is statistically significant at any level.



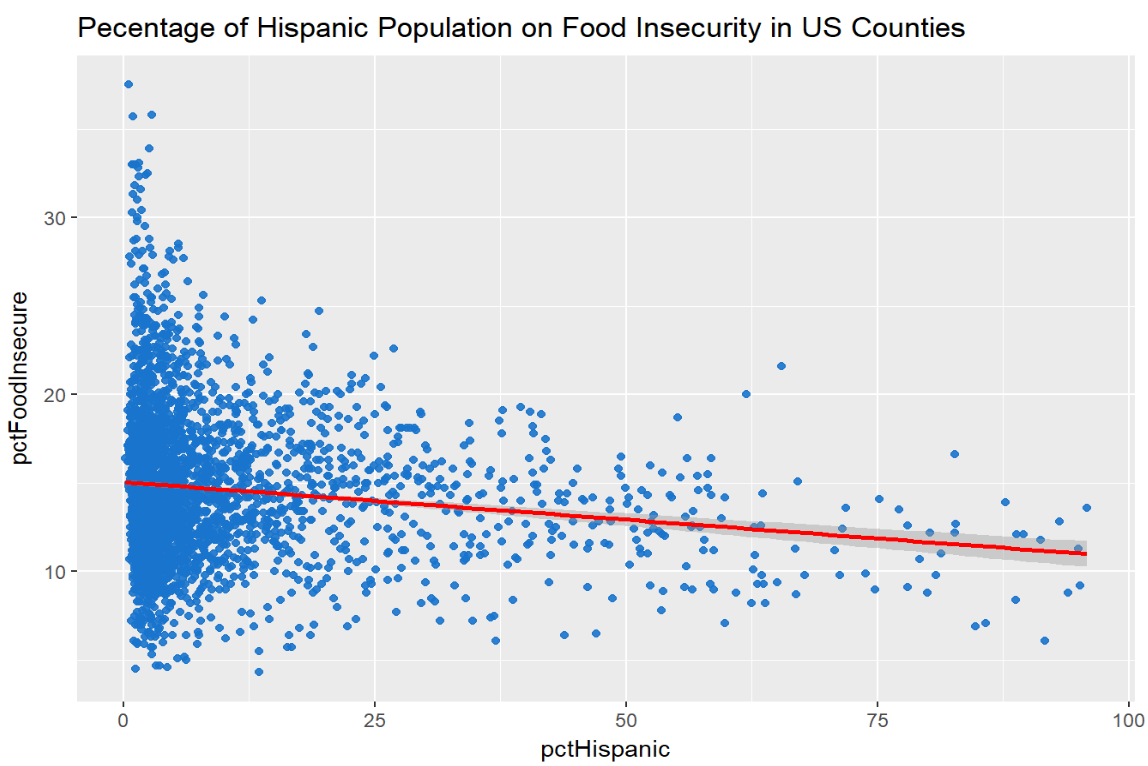


Figure 6,7: Scatterplot of regression results for black and hispanic populations

All other results I found in the data were also statistically significant. Each control variable I added improved the fitness of the model and increased the adjusted R squared.

Urban populations were found to have less food insecurity than rural populations. This may be because of some measurment error in considering the commute distance, but otherwise we have no reason to believe the result was statistically biased except for questioning the source of the data. It seems to suggest that urban areas have better infrastructure, shorter commutes, and higher concentration of ethnic mixture that could play into lower food insecurity.

It is also found that, intuitively, unemployment rate and counties with high or extreme drought risk tended to be areas with higher food insecurity. These increases were significant and help us to understand the causes underlying food insecurity under greater detail.

In the final regression model, including all independent variables, I found that the model could explain 62.1 percent of the total variation in food insecurity percentage across all of the county data.

To numerically summarize, we found that a 1 percent increase in black population by county, holding all other explanatory variables constant, would result in a 0.156 percent increase in the amount of food insecure individuals on average. A 1 percent increase in hispanic population, on the other hand, results in a 0.022 decrease in the amount of food insecure populations. Drought risk, one of our more interesting control variables, observes that counties with significant or moderate drought risk in 2016 were shown to have 0.556 percent more individuals with food insecurity.

4 Conclusion

Although this analysis was done solely for the joy of making regressions and this result is not intended (and could not be justified) to make any claims about the real world, we found that this data-set makes the conclusion that black minority communities are disproportionately more likely to experience food insecurity. In the real world, we know food insecurity to be a real problem. Regardless of the actual magnitude and its relationship to the social dimension on factors like race and community, it is a problem we should be informed of and try to resolve.