DATA 606 Data Project Presentation

William Aiken

Abstract

Diabetes is a disease with a high health cost to the individual and high monetary cost to our communities. New York state tracks diabetic rates at the county level along with other health and economic data. I leveraged this publicly available data to explore the heterogeneity in diabetic rates in New York state. I wanted to know if there was a correlation between diabetes, income, and obesity at the county level. The individual variables were visualized, and the income data was log transformed to help resolve the skewness of the distribution. Linear regression was used explore the relationship between diabetes (dependent variable) and income and obesity (independent variables). The R-squared was found to be 0.355, showing a correlation between the outcome and predictive variables. Both coefficients were found to significantly different from zero. The coefficients for obesity and the log transformed income were 0.12 and -1.9 respectively. The incidence of diabetes increases with the increase of the obesity rate for the county and decreases with the increase in average income for a county. Further exploration of the relationships could lead to better interventions

Part 1 - Introduction

I was interested in exploring the hetrogeneity of diabetic rates in New York State and how it is related to income and obesity rates. New York is an intersting state for this analysis because there are so many different geographic regions within the state.

Research Question: Are the diabetic rates in New York state correlated with the obesity rates and average income at the county level?

Part 2 - Data

This data comes from the NY.GOV site as part of their open data sets

- The Obesity and Diabetes data comes from the New York State Department of Health disease registries. These registries are based on multiple sources including hospital registries in a given geographic area. People who live outside a given area but receive treatment within a geographic region are not counted in these population based registries.

Obesity and Diabetes

- The income data is collected by the New York State Department of Taxation and Finance. This data comes the New York State personal income tax returns that were filed in a timely fashion. This data is for full-time New York State residents.

Income

Part 3a - Exploratory data analysis

- Looking at summary statistics of our data.

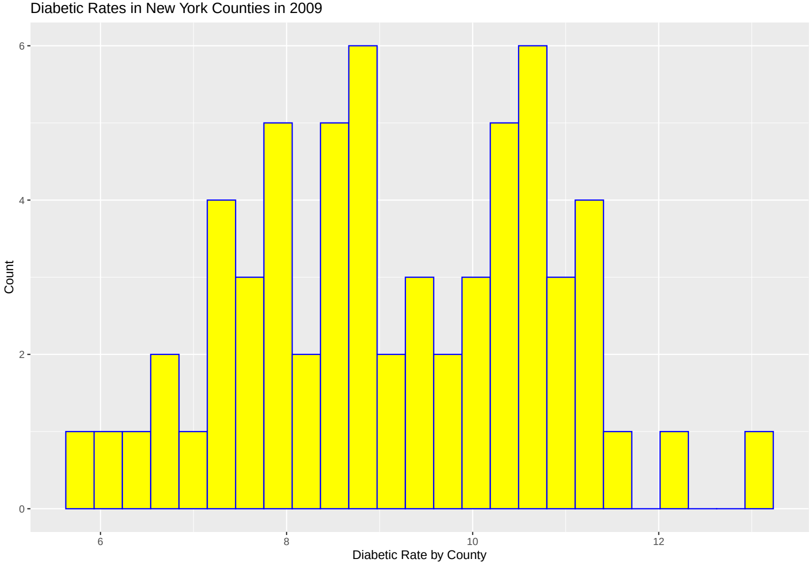
Characteristic	**N = 62**
diabetic	
Mean (SD)	9.15 (1.61)
Range	5.90, 13.20

Characteristic	**N = 62**
income	
Mean (SD)	45,578 (17,883)
Range	30,904, 141,218

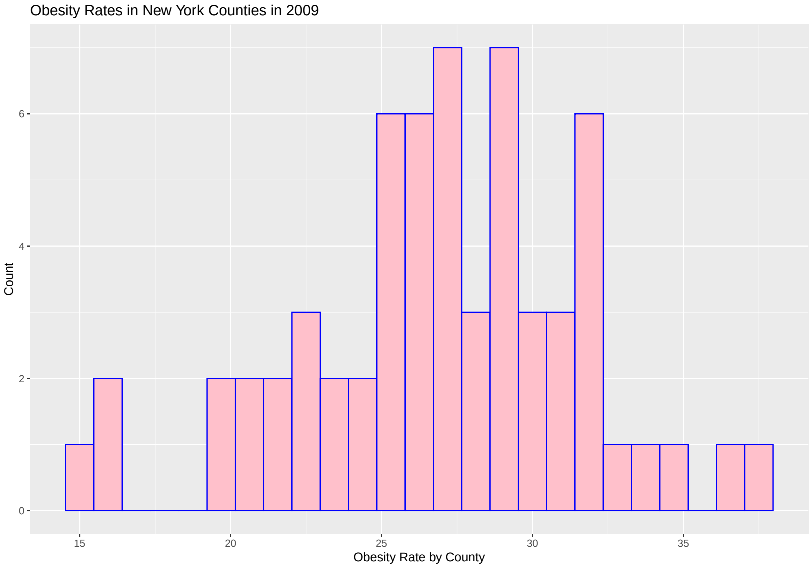
Characteristic	**N = 62**
obesity	
Mean (SD)	26.9 (4.7)
Range	15.0, 37.5

Part 3b - Histograms of the data

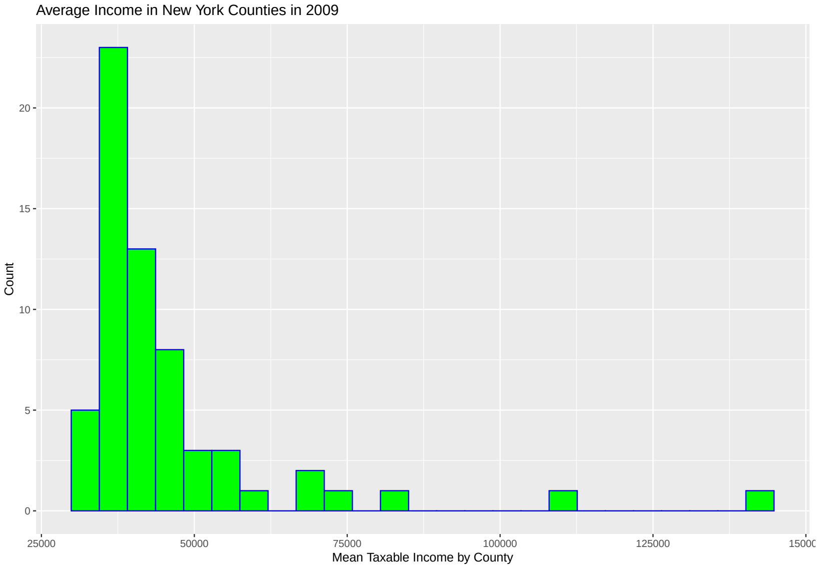
- Diabetic Rates



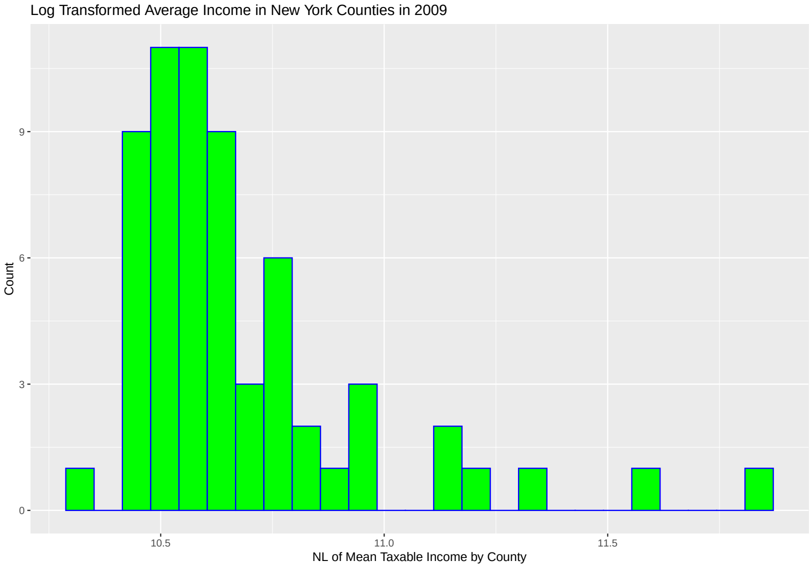
Part 3c - Obesity Rates



Part 3d - Average Income

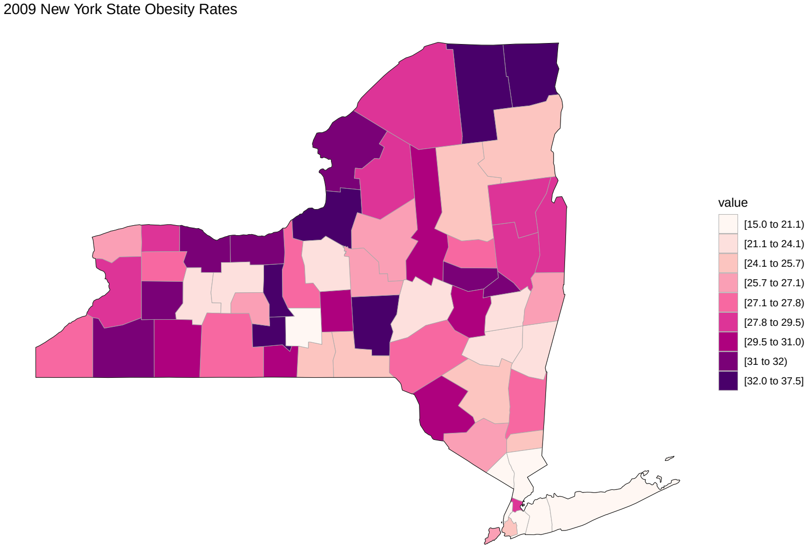


Part 3e - Addressing the skewness

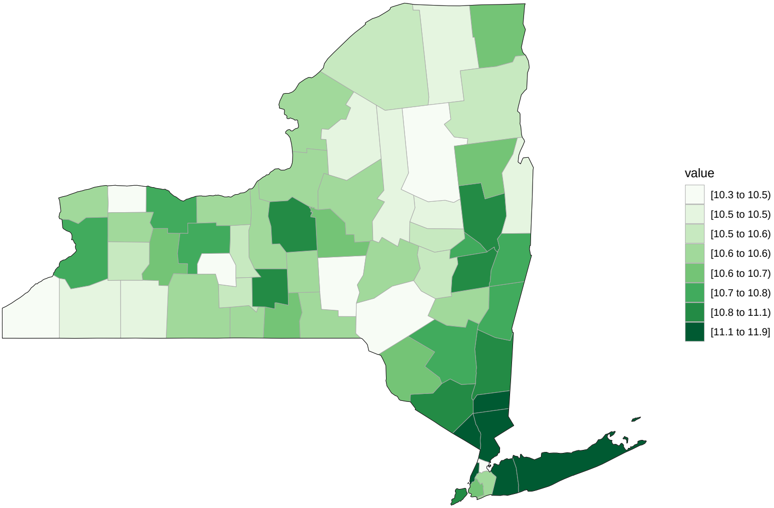


Part 3f - Exploring geographic relationships

- Obesity rate by county

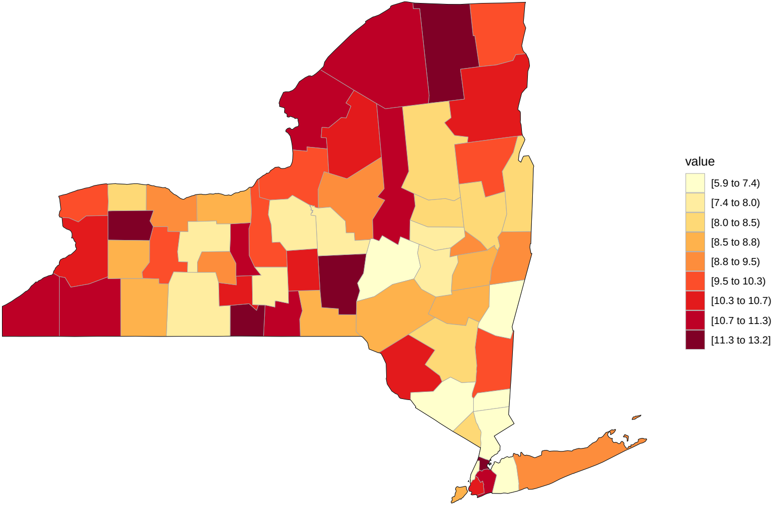


Part 3g - Income by county



2009 Log of Average Incomes by County

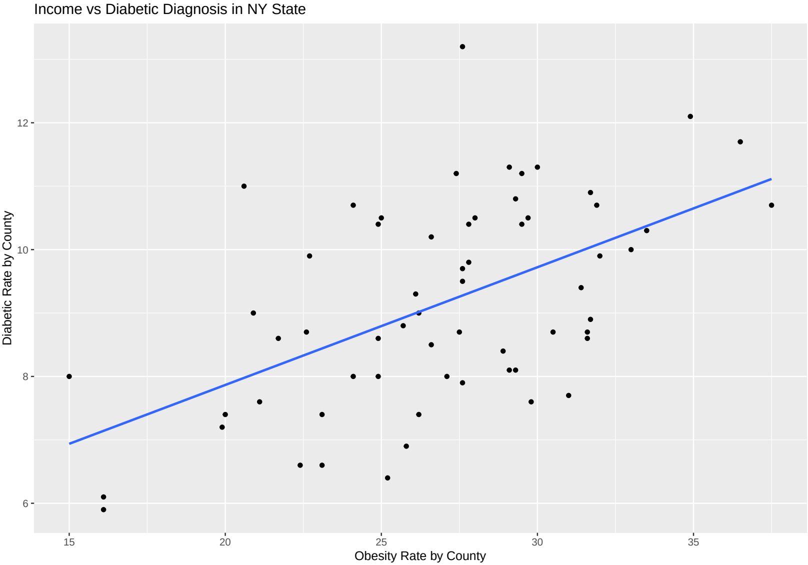
Part 3h - Diabetic rate by county



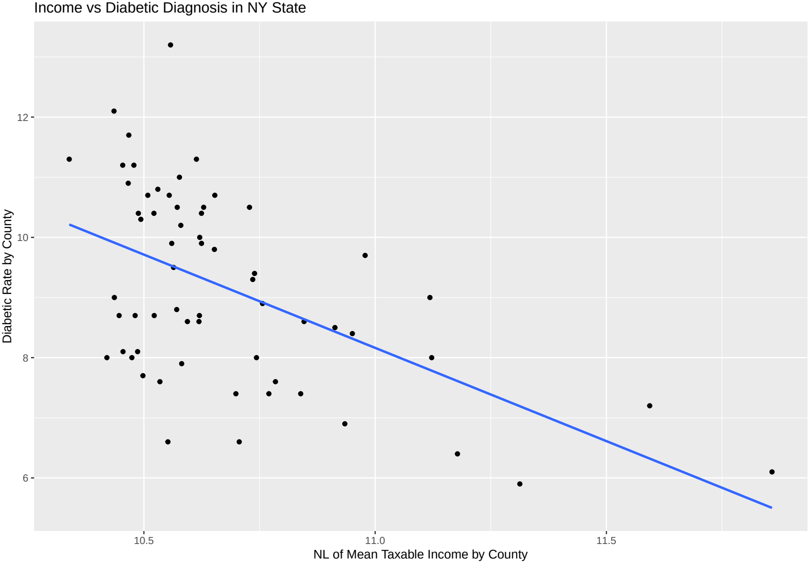
Diabetic Rates per County

Part 4a - Inference

- Does linear regression look appropriate?



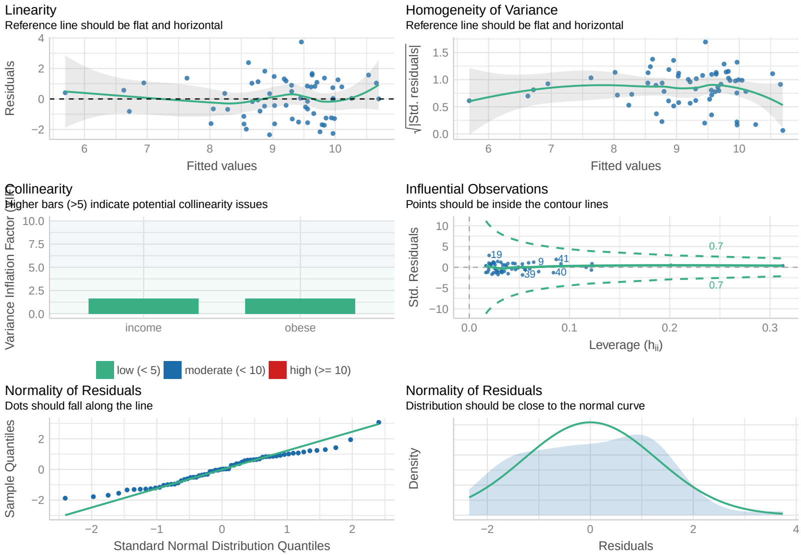
Part 4b - Slight heteroscedasticity for Income



Part 4c - Linear Model

Characteristic	**Beta**	**95% CI**	**p-value**
obese	0.12	0.02, 0.21	0.015
income	-1.9	-3.4, -0.32	0.019
No. Obs.	62		
R^2	0.355		
Adjusted R^2	0.334		
Sigma	1.32		

Part 4d - Check our Model



Part 5 - Conclusion

This analysis is important because type 2 diabetes is a debilitating desease that is largely preventable. To understand what other factors are related to its incidence may lead to better prevention methods.

We found that there is a correlation between diabetes and income and obesity. There are some limitations to the interpretability of these results. We used measurements captured at the county level, the populations within each county vary wildly. We can't say what the relationship is at the population level of all people who live in New York state because these measurements are unweighted.

References

New York State Department of Health. "Community Health Obesity and Diabetes Related Indicators: 2008 - 2012: State of New York." Community Health Obesity and Diabetes Related Indicators: 2008 - 2012 | State of New York, 1 July 2016, https://health.data.ny.gov/Health/Community-Health-Obesity-and-Diabetes-Related-Indi/tchg-ruva.

New York State Department of Taxation and Finance. "Average