New York State Socioeconomic Factors and COVID-19 Case Rates

Correlation One:
Data Science for All
Team 88

Idalys Perez
Michael Luevanos
William Johnson
Tolu Olude
Monika Wiggins
Juan Llamas
Sean Kang

TABLE OF CONTENTS

LIST OF FIGURES

OVERVIEW OF PROBLEM

The COVID-19 pandemic has been determined to have a greater impact on people within specific ethnic, racial, and socioeconomic groups according to the CDC.[14] Throughout our project, we set out to determine whether the differences in COVID-19 positivity rates seen in these groups through compounded data could also be seen in more clearly defined geographic regions, namely: the different counties comprising New York state. Being able to make these determinations in more specific regions will allow governments and social welfare organizations to focus their mitigation efforts to better defined areas, rather than to broader categories (ie. "the Latino population"), which could serve to make outreach more successful.

INTRODUCTION

On March 11, 2020, the World Health Organization declared COVID-19 a global pandemic, a distinction the office had not made since 2009 - for the H1N1 influenza.[1] Despite various governments issuing stay-at-home orders and social distancing guidelines, COVID-19 quickly spread throughout the world. Within the United States, it was discovered that the population at large was not becoming infected with COVID-19 at equal rates. For example, in a CDC report analysis of patients in 14 states hospitalized due to COVID-19, it was discovered that despite only 18% of the population being Black, 33% of the hospitalized patients were Black.[2] In the Washington DC metropolitan area, it was determined that those identifying as Latino were the most likely to test positive for COVID-19 at much higher proportions than any other race/ethnicity. [3] The differences seen in positivity rates might be due to the fact that people identifying as Hispanic/Latino tend to have lower rates of health insurance coverage and health care utilization. [4] Another possible explanation for the racial impacts of COVID-19 may be due to how certain jobs were deemed "essential", and these jobs - with higher percentages of COVID-19 positivity rates - are for the most part held by low-income and Black, Native American, and Latino workers, rather than higher income White workers. These essential workers with a higher probability of contracting COVID-19 are also more likely to have larger household sizes and multiple generations living with them.[5]

Another prevalent issue that is potentially worsening the spread of COVID-19 is food deserts. Under the USDA definition, a food desert is an area that has low to no access to nutritional food options at one mile within a city and ten miles in a rural area. Food deserts continue to be a growing public health concern. For instance, in the United States an estimated 13.5 million people do not have proper access to healthy foods.[6] This issue affects low-income households more than higher-income ones, and it is also tied to race because of its socioeconomic implications. More specifically, approximately 30% more non-White residents face limited access to healthy foods compared to their White counterparts.[7] Food deserts have disproportionately impacted the health of its residents because the lack of access to adequate food resources has been linked to higher rates of diabetes, obesity, and cardiovascular disease.[8]

The problem of food deserts became more concerning with the start of the COVID-19 pandemic, as people living within food deserts tend to possess many factors that increase their likelihood of contracting COVID-19. Specifically, people living in these areas have a higher probability of having comorbidities, being lower-income, and a higher percentage of jobs in the service sector.[9] Research has focused solely on determining whether food desert communities show a statistically significant increased rate of COVID-19 positivity rates than those communities around them, but does not include a more detailed look at what factors are specifically responsible for these differences.[10]

Previous research determined that ethnic, racial, and socioeconomic factors impact the spread of COVID-19. However, these studies mostly focus on proving that out of all people who tested positive in an area, there is a higher proportion of certain racial and socioeconomic profiles. We sought out to determine whether these effects could be seen within specific counties and could be used as a way to determine where extra resources and health campaigns could be focused in order to have the maximum impact.

METHODOLOGY

Our first concept for our project rested on the premise that the communities living in food deserts were also the ones least likely to have vaccines properly distributed to them. Thus, we believed we could use data to identify the ways in which methods used to transport, deliver, and distribute vaccines could be redesigned in order to help facilitate access to healthier food options in food deserts (i.e. a vaccine distribution site becoming the site of a new farmers market). We quickly realized our project was too wide in scope and that we were making unfounded suppositions which would have to be proven in order for our project to make sense. Most importantly, we would need to examine the fact that a community being a food desert also meant that its access to vaccines were limited.

The next idea for our project involved using data collected from Instacart, as well as other grocery delivery services, in order to determine whether the pandemic had led to an increase in access to these services for SNAP recipients. Recently, and due to the impact of COVID-19, these grocery delivery programs began accepting SNAP benefits as payment, which was legally allowed many years prior to the pandemic but was not put in practice. However, Instacart and other grocery delivery companies declined our request for data since they could not publicly release SNAP payment or sales data for the purpose of our project.

Next we tried to refine the focus of our project by taking a deeper look into whether areas defined as food deserts had statistically significant increased COVID-19 positivity rates, and if so, what factors (such as race or household income) contributed to this increase. We narrowed our focus from a comparative analysis of three cities to one city. After more research on publicly available COVID-19 data, we identified the state of New York as the main focus of our project because it had the most robust and up-to-date datasets. New York is also among the most diverse and highly populated states in the country, making it a more impactful state to evaluate COVID-19 rates. Compared with other populous regions, New York City alone saw some of the highest spikes in cases and deaths in the spring of 2020.

Moving forward, our first challenge came with obtaining recent food desert data, as we needed data covering the time period when COVID-19 began affecting these communities. Understandably, efforts to collect COVID-19 data were hindered by safety measures and concerns. Our team reached out to Feeding America and we were then able to obtain data on food deserts and food insecurity in the U.S. that the organization created based on estimated projections.

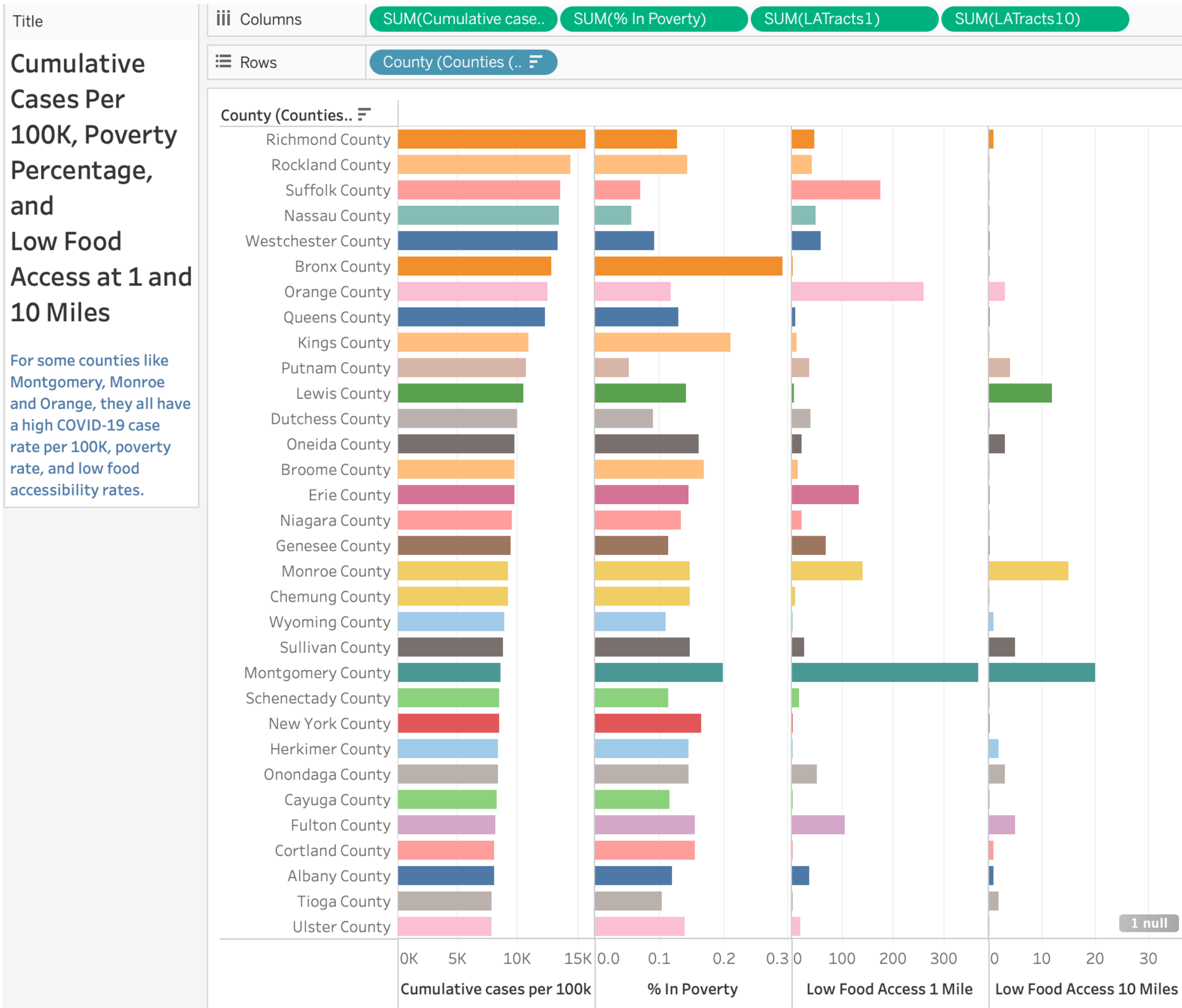
Our team encountered another challenge when trying to combine the dataset of COVID-19 case rates with that of food deserts. Information on COVID-19 infection rates is based on zip codes, while food desert data is collected based on census tracts. This led to issues when merging our data because zip codes are generally based on routes used by the U.S. Postal Office and do not always contain strict boundaries. Zip codes can also be spread out over many census tracts, and thus integrating data sets that use these different boundaries becomes difficult. Upon researching ways that other research groups achieved this, we discovered converting census tract to zip code is an industry-wide challenge when it comes to working with public census data. This has yet to be resolved.

In the end, our team was unable to find statistically significant findings that proved that living in, or close proximity to, a food desert led to a higher percentage of people contracting COVID-19. Due to these findings, we instead turned to all of the other categories for which we had data to test whether we were able to discover statistically significant changes in COVID-19 positivity rates based on different socioeconomic factors by geographic area, such as: race, health insurance, access to a car, and poverty rates. After grouping each area by counties and examining the socioeconomic factors, we then conducted correlation tests and created linear regression models in order to determine which of the twenty socioeconomic variables showed a strong relationship to COVID-19 positivity rates.

MODELS AND CONCLUSIONS

Based on the graph below, there were several counties that had high cumulative COVID-19 cases, poverty rates, and low food access within one mile. For instance, Orange County had the seventh-highest COVID-19 case rate and also had the third-highest food desert rate in New York. Similarly, Montgomery County had relatively high COVID-19 case rates at 8,620 per 100K residents and poverty rates at 19 census tracts which were above the state average. Montgomery County also had the highest food inaccessibility rate among all counties in New York.

FIGURE 1



In addition, there were some notable findings in areas that were expected to have high food inaccessibility rates with high COVID-19 cases and poverty rates, but did not have large urban food inaccessibility rates based on the data collected. For instance, Bronx County in New York City, had the highest poverty rate at 29 tracts and the sixth-highest COVID-19 case rate at 12,851 per 100K, but had one of the lowest food inaccessibility rates in the state. This finding could be regarded as a Simpson's Paradox since sociologically it would be anticipated that a highly densely populated urban county like Bronx County would have high urban food deserts because of their high COVID-19 case rate and high poverty rate. However, the data demonstrates that Bronx County does not encounter large issues with food deserts given its urban location in New York City.

There were multiple counties that had high cumulative COVID-19 cases, poverty rates, and low food access percentages within 10 miles. For instance, Montgomery County had 8,620 cumulative cases per 100K residents with around 20 tracts living in poverty and about 20 tracts of low food access areas. Similarly, Washington County has about 5,096 cases per 100K residents, 12 percent in poverty and 35 tracts with low food access. Franklin County also had 5,064 cumulative cases per 100K residents, 18 percent in poverty and 26 tracts with low food access within 10 miles.

Other notable counties that had similar trends were Jefferson, Madison, and Wayne County. Jefferson County had 5,506 cases per 100k residents, a 15% poverty rate, and 22 tracts with low food access. Madison County had 6,380 cumulative cases per 100K, around 10 percent in poverty, and 19 tracts with low food access within 10 miles. Wayne County also had 6,377 cumulative cases per 100k, 11 percent in poverty, and 17 tracts with low food access.

Overall, out of all the counties we can see that Montgomery County had the largest number of low food access tracts within 10 miles and also had relatively high cumulative cases per 100k and poverty rates. Most of the counties with fewer low food access tracts within 10 miles had relatively lower cumulative cases per 100K and poverty rates compared to counties with a lower number of food access tracts within 10 miles. All these rural counties with many low food access tracts demonstrate the importance of targeted vaccination efforts given the challenges they experience with food access, poverty, and case rates.

FIGURE 2

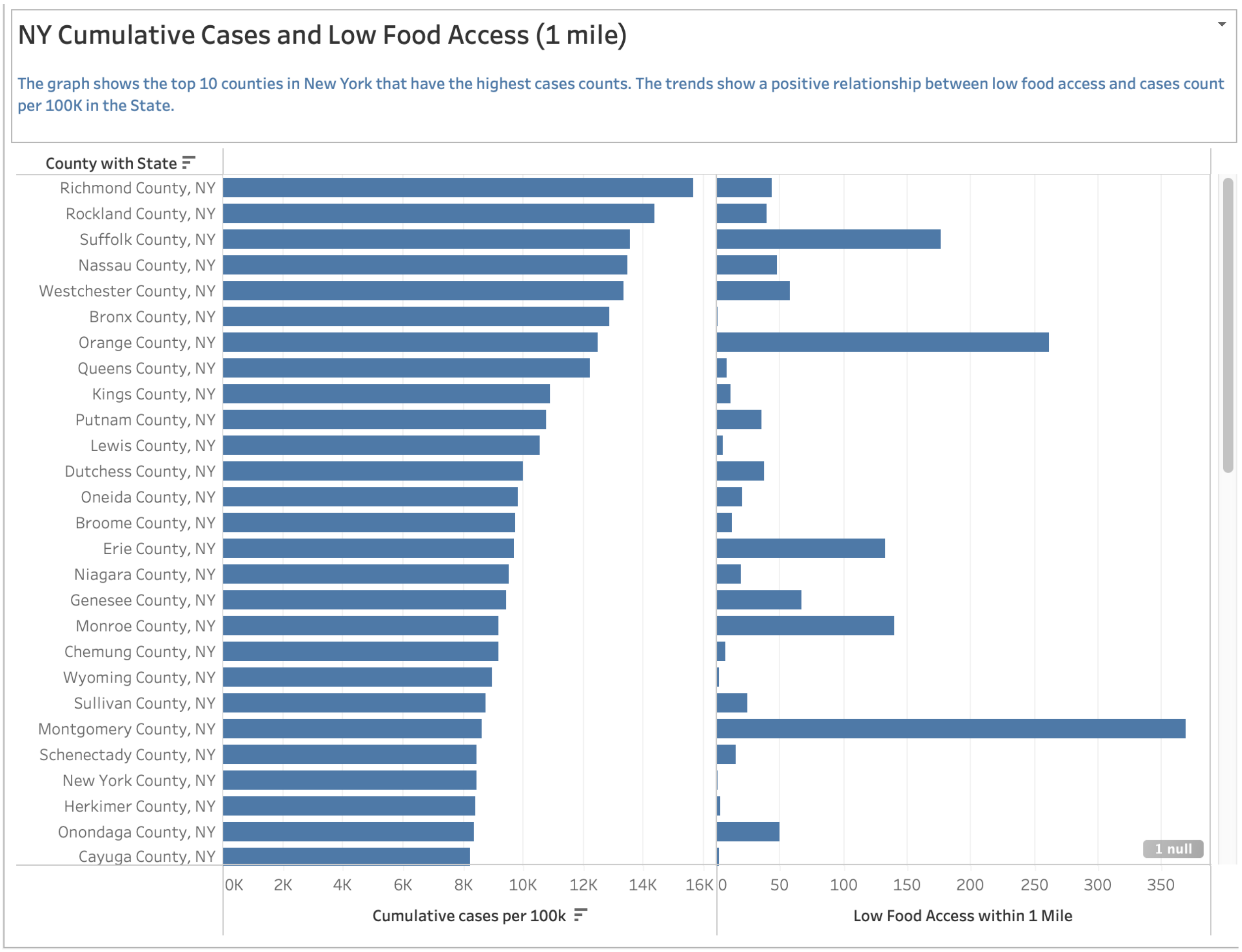
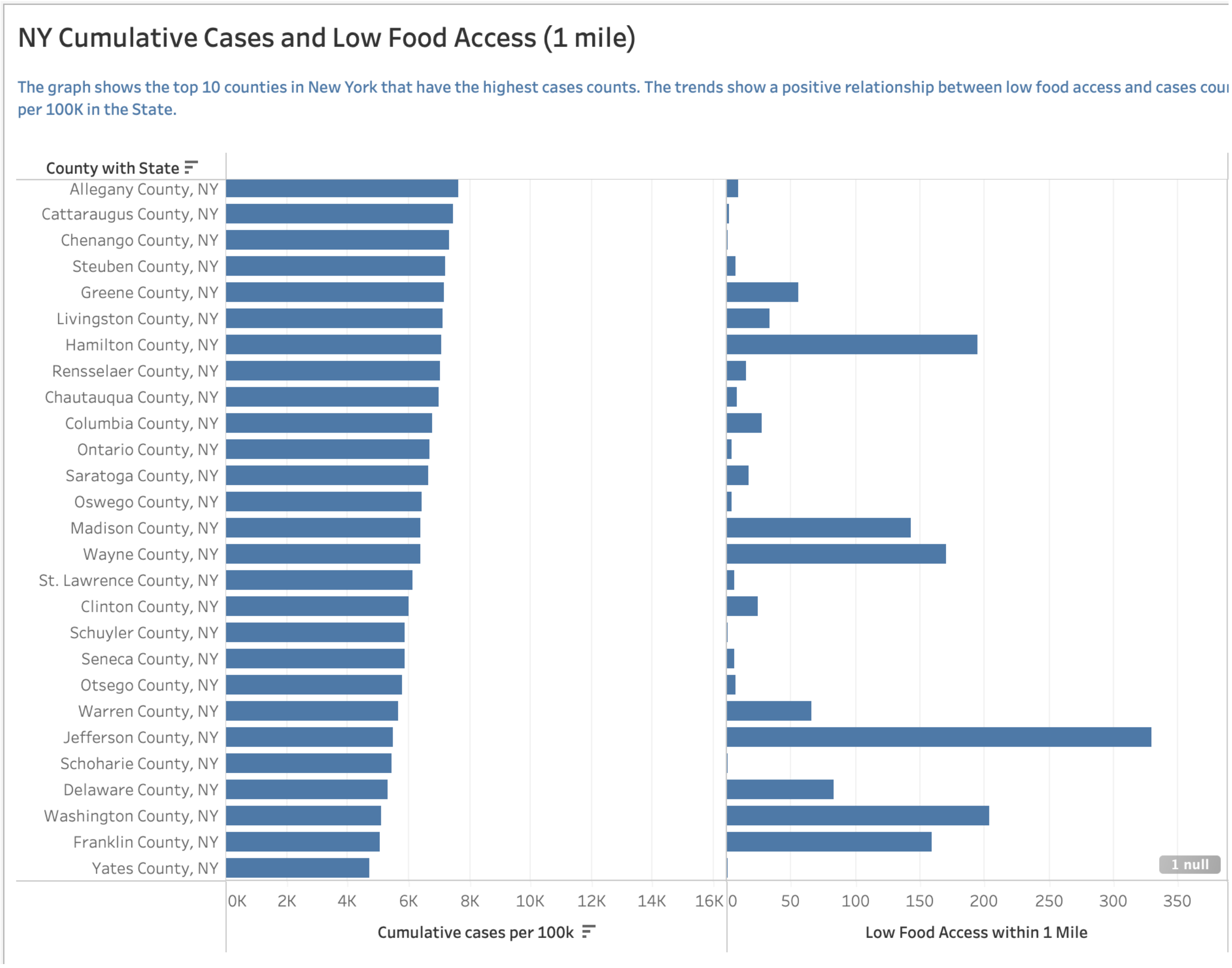

NY Cumulative Cases and Low Food Access (1 mile)
The graph shows the top 10 counties in New York that have the highest cases counts. The trends show a positive relationship between low food access and cases count per 100K in the State.
County with State
Richmond County, NY
Rockland County, NY
Suffolk County, NY
Nassau County, NY
Westchester County, NY
Bronx County, NY
Orange County, NY
Queens County, NY
Kings County, NY
Putnam County, NY
Lewis County, NY
Dutchess County, NY
Oneida County, NY
Broome County, NY
Erie County, NY
Niagara County, NY
Genesee County, NY
Monroe County, NY
Chemung County, NY
Wyoming County, NY
Sullivan County, NY
Montgomery County, NY
Schenectady County, NY
New York County, NY
Herkimer County, NY
Onondaga County, NY
Cayuga County, NY
1 null
0K 2K 4K 6K 8K 10K 12K 14K 16K
Cumulative cases per 100k
0 50 100 150 200 250 300 350
Low Food Access within 1 Mile

FIGURE 3



As you can see from the graph above, there is not a strong trend when comparing cumulative cases per 100K and low food access within 1 mile, however, there are some notable observations among other counties in urban areas. For instance, Suffolk County has high counts of COVID-19 cases per 100K and also one of the largest counts of people who do not have food access within 1 mile. Similar trends can be seen in both Orange County and Sullivan County. Moving towards the end of the bar graph, there are a few counties such as Jefferson County, which had one of the lower cumulative case counts per 100K and also have a very high number of people that do not have food access within 1 mile.

FIGURE 4

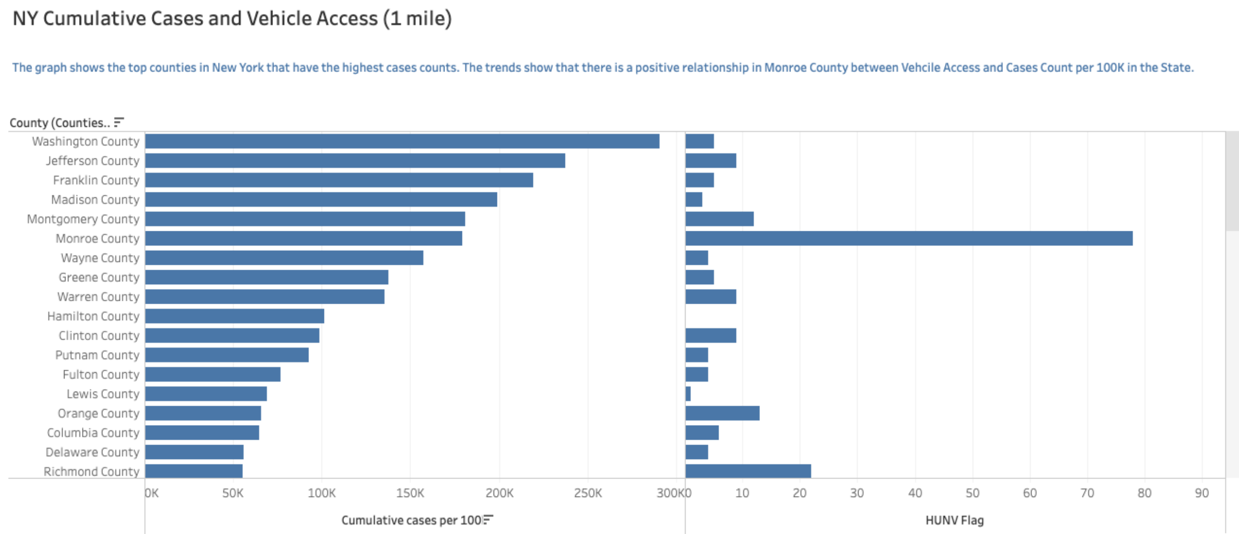
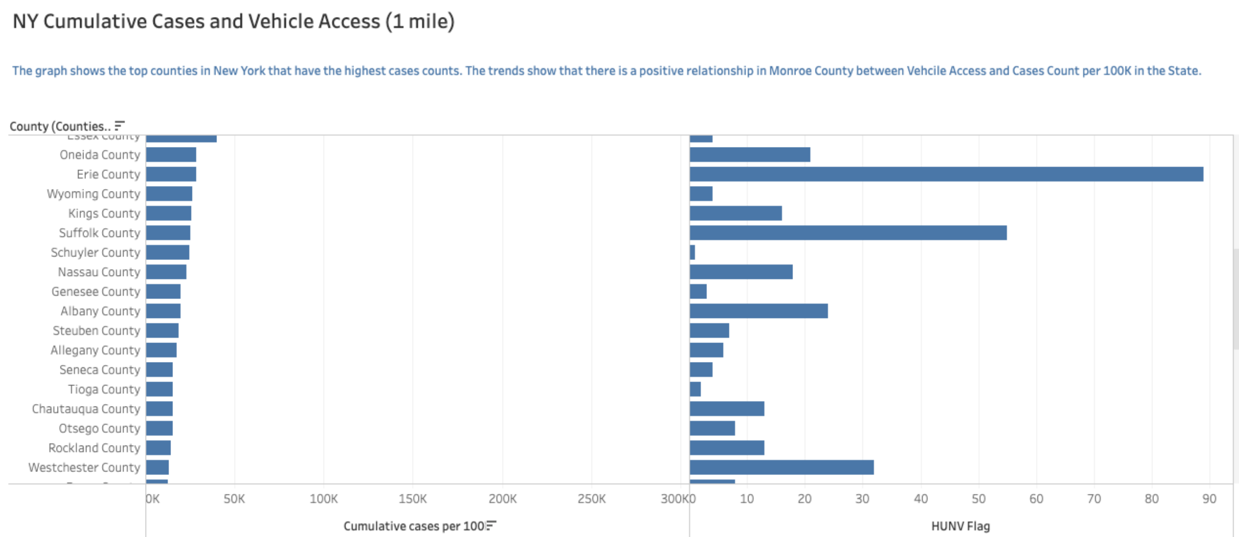
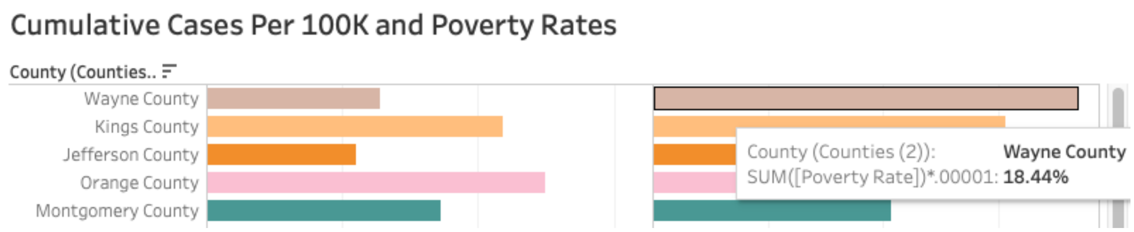


FIGURE 5



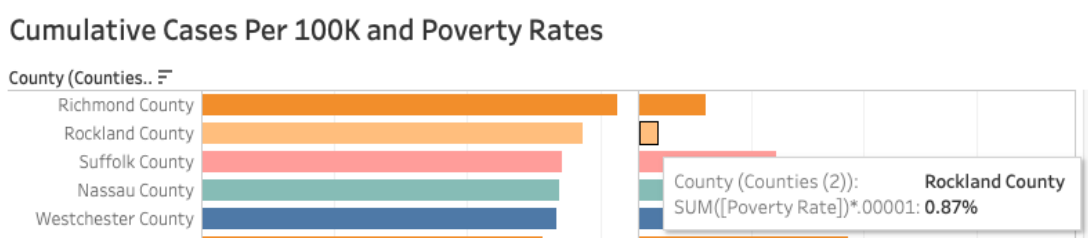
Furthermore, in the chart above "HUNVTRACT" measures households that do not have a vehicle and people who live further than 1/2 mile from a supermarket. There is not a clear correlation between cumulative cases per 100K cases and vehicle access. For instance, the cumulative cases per 100K rates for Erie County are very low compared to how high the rates are for people living further than a 1/2 mile from the supermarket. However, in Monroe County there is a stronger correlation between cumulative cases and distance from supermarkets. While there are not too many correlations between both of these variables, there are still a few counties that have these correlational trends. This may be due to how some counties are smaller with lower populations and smaller case rates. Therefore, further investigation is needed for these counties with unique correlations to determine the potential confounding factors that are creating these relationships.

FIGURE 6



According to the graph above, poverty rates by county do not have a strong relationship with COVID-19 rates per 100K residents. Across all 62 counties, poverty rates vary widely. For instance, the top five counties where poverty rates are highest are Wayne at 18.44%, Kings at 15.29%, Jefferson at 12.42%, Orange at 11.71%, and Montgomery at 10.34%. The graph also shows the relationship between the two variables. Wayne County had 6,377 cumulative COVID-19 cases per 100K at an 18.44% poverty rate. Montgomery County had 8,620 cumulative COVID-19 cases per 100k at a 10.34% poverty rate. Therefore, based on the data a lower count of cumulative cases does not necessarily correlate to a lower poverty rate.

FIGURE 7



Moving forward, in the bar graph above, the top five counties where cumulative COVID-19 cases were highest were Richmond, Rockland, Suffolk, Nassau, and Westchester. The graph shows how poverty rates span from a low 0.87% for Rockland County to only 6.12% for Suffolk County. In all the 62 counties, Wayne County had the highest poverty rate at 18.44%.

FIGURE 8

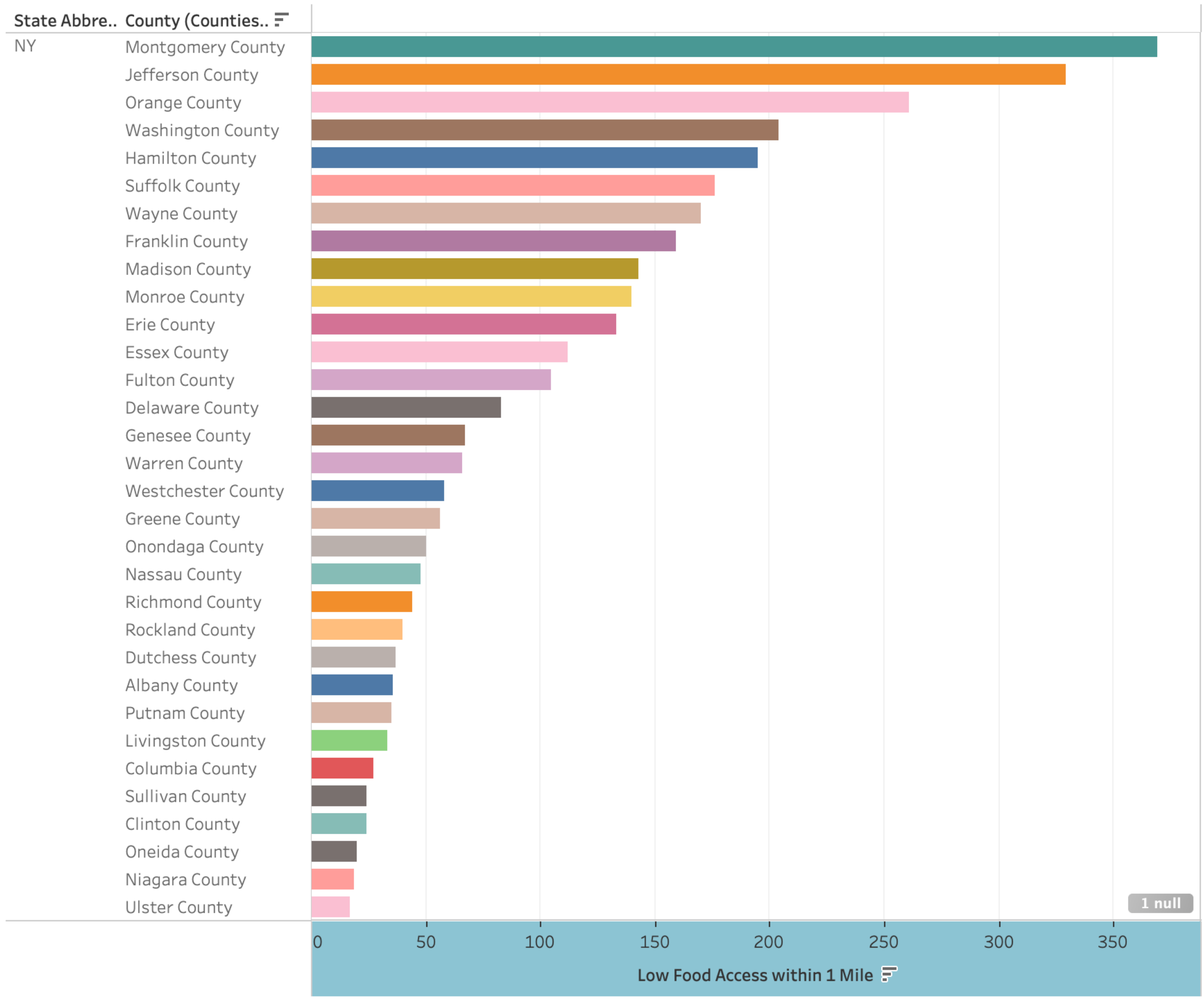
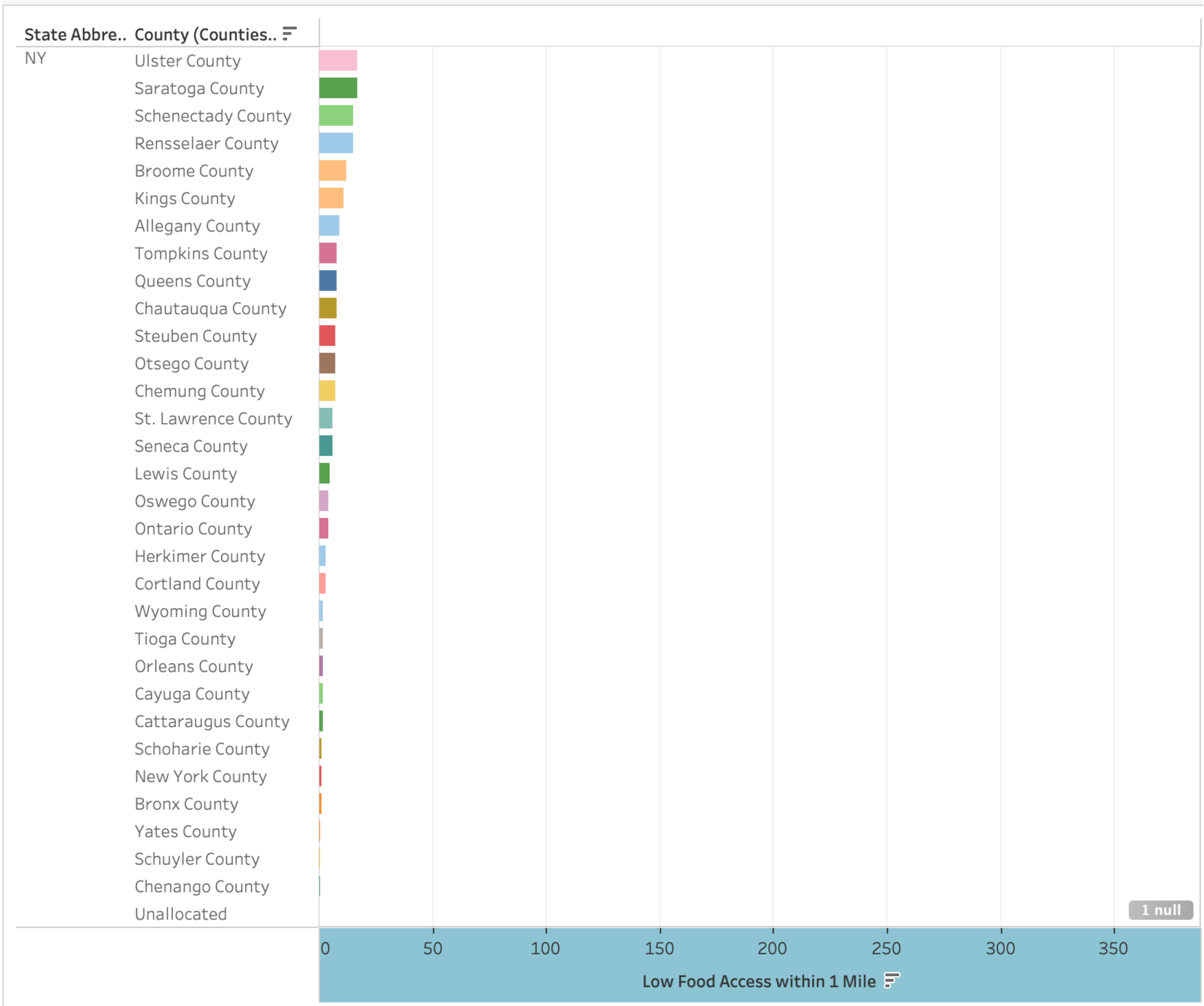
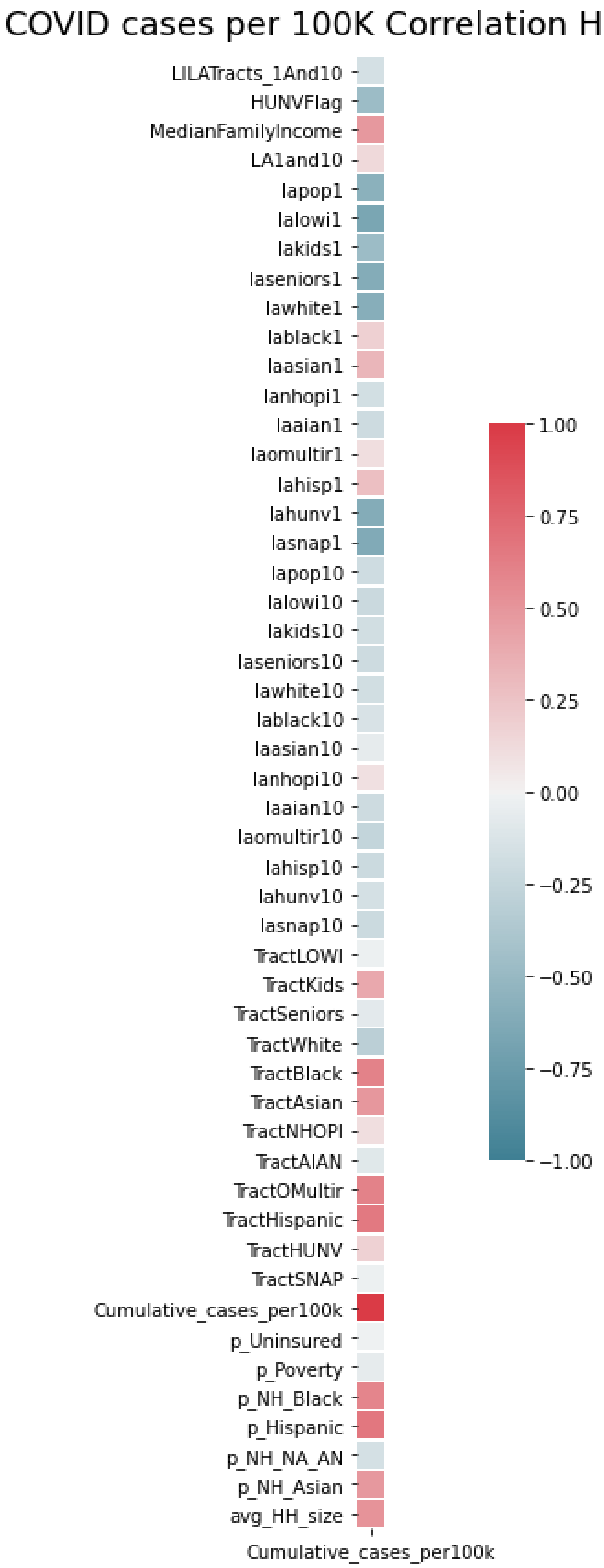

State Abbre..
County (Counties..
NY
Montgomery County
Jefferson County
Orange County
Washington County
Hamilton County
Suffolk County
Wayne County
Franklin County
Madison County
Monroe County
Erie County
Essex County
Fulton County
Delaware County
Genesee County
Warren County
Westchester County
Greene County
Onondaga County
Nassau County
Richmond County
Rockland County
Dutchess County
Albany County
Putnam County
Livingston County
Columbia County
Sullivan County
Clinton County
Oneida County
Niagara County
Ulster County
1 null
0
50
100
150
200
250
300
350
Low Food Access within 1 Mile

FIGURE 9



Furthermore, the following graph depicts how Montgomery County and Jefferson County have the highest level of food inaccessibility by a relatively large margin in the state. Most of the other counties in New York, except for four of the counties, have less than half of the amount of food inaccessibility than Montgomery County. As for the New York City Metropolitan area, all of its counties seem to have relatively low levels of food inaccessibility. However, the data may not be fully representative of the issue. Therefore, it is critical to analyze the COVID-19 case rates, median income, and poverty levels within the counties to ensure that these communities' needs are being accurately depicted.

FIGURE 10


COVID cases per 100K Correlation Heat Map
LILATracts_1And10
HUNVFlag
MedianFamilyIncome
LA1and10
lapop1
lalowi1
lakids1
laseniors1
lawhite1
lablack1
laasian1
lanhopi1
laaian1
laomultir1
lahisp1
lahunv1
lasnap1
lapop10
lalowi10
lakids10
laseniors10
lawhite10
lablack10
laasian10
lanhopi10
laaian10
laomultir10
lahisp10
lahunv10
lasnap10
TractLOWI
TractKids
TractSeniors
TractWhite
TractBlack
TractAsian
TractNHOPI
TractAIAN
TractOMultir
TractHispanic
TractHUNV
TractSNAP
Cumulative_cases_per100k
p_Uninsured
p_Poverty
p_NH_Black
p_Hispanic
p_NH_NA_AN
p_NH_Asian
avg_HH_size
Cumulative_cases_per100k
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

LINEAR MODEL REGRESSION TESTING

As you can see from our correlation matrix above, comparing COVID-19 cases per 100K residents with several other variables in our data set, there are about 10 variables that have a correlation of 0.4 or higher, which indicates a strong to moderately strong relationship with the number of cases per pupil.

After evaluating these variables, we created a multiple variable linear regression test. The results below depict how many of the variables that had a strong correlation in the matrix are not significant in the linear modeling test.

FIGURE 11

```
formula4 = "Cumulative_cases_per100k ~ p_Uninsured + p_Poverty + p_NH_Black + p_Hispanic + p_NH_Asian + avg_HH_size + HUNVFlag + MedianFamilyIncome + TractSNAP"
model4 = sm.ols(formula = formula4, data = merged_on_county_df2)
fitted4 = model4.fit()
print(fitted4.summary())
```

```
                            OLS Regression Results
==================================================================================
Dep. Variable:     Cumulative_cases_per100k   R-squared:                      0.622
Model:                                  OLS   Adj. R-squared:                 0.556
Method:                       Least Squares   F-statistic:                    9.500
Date:                      Sat, 31 Jul 2021   Prob (F-statistic):          2.21e-08
Time:                              17:18:24   Log-Likelihood:               -544.87
No. Observations:                        62   AIC:                            1110.
Df Residuals:                            52   BIC:                            1131.
Df Model:                                 9
Covariance Type:                  nonrobust
======================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
Intercept             72.5575   4321.296      0.017      0.987   -8598.757    8743.872
p_Uninsured        -2.549e+04   1.15e+04     -2.212      0.031   -4.86e+04   -2367.477
p_Poverty          -9088.5814   1.13e+04     -0.807      0.424   -3.17e+04    1.35e+04
p_NH_Black          7753.0952   6479.905      1.196      0.237   -5249.786    2.08e+04
p_Hispanic          1.155e+04   4960.525      2.328      0.024    1593.670    2.15e+04
p_NH_Asian          6026.5132   7906.186      0.762      0.449   -9838.410    2.19e+04
avg_HH_size         2949.0872   1126.118      2.619      0.012     689.365    5208.809
HUNVFlag            1023.2612   1990.069      0.514      0.609   -2970.105    5016.627
MedianFamilyIncome     0.0124      0.021      0.602      0.550      -0.029       0.054
TractSNAP              3.2559      5.664      0.575      0.568      -8.110      14.622
==============================================================================
Omnibus:                        0.747   Durbin-Watson:                   2.043
Prob(Omnibus):                  0.688   Jarque-Bera (JB):                0.336
Skew:                           0.164   Prob(JB):                        0.845
Kurtosis:                       3.151   Cond. No.                     4.51e+06
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 4.51e+06. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Based on the regression test above, the variables that we fail to reject are: percentage of Hispanic people, average household size, and percent of population that is uninsured for each county. This means that the three variables mentioned are statistically significant at the .05 level. Additionally, you can see that our R-squared does show a 0.622 coefficient, meaning approximately 62.2% of the variation can be explained by our model.

The variable “LA1and10” represents a flag of low access census tracts at one mile for urban areas and 10 miles for rural areas within a county. With a p-value of 0.717, the results did not prove statistically significant. This means our initial hypothesis of a proximity to a food desert impacting COVID-19 rates did not prove to be statistically significant. To confirm this, we then removed this variable from our regression. As a result, the R squared value increased by .01 which means that the previously included variable was insignificant to our hypothesis.

CONCLUSION

Throughout our analysis we discovered that there is not a significant correlation between proximity to a food desert and COVID-19 case rates in the state of New York. However, we did find that there is a correlation between the average household size, the Hispanic population percentage, and the percentage of uninsured individuals with COVID-19 case rates in many New York counties.

After further researching correlational relationships with household size, race/ethnicity, and COVID-19 case rates, we found similar trends proving how these variables are related to each other. According to Ann C. Foster’s article “ Beyond The Numbers”: “household size was larger in Hispanic households (3.1 people) compared with 2.4 people in both White and Black households.”[13] Based on this assertion, we can see that these two variables have a relationship on their own, which shows this inextricable relationship of how COVID case rates can be impacted by household sizes and racial/ethnic background. The larger household size in Latino households may be attributable towards other socioeconomic factors that can increase these communities’ likelihood of contracting the virus since, on average, they have more people living with them.

As for the other variables that did not prove to have strong relationships, we were still able to conclude some limitations with how the survey data may not fully be representative of these communities. After running our regression analysis on other variables such as the percentage of Non Hispanic-Black, the percentage of people in poverty, and the median household income, we were surprised to see that the variables did not seem to have a strong statistical relationship with COVID-19 case rates. Moving forward, it could be helpful to run further analysis on these variables and their surveying methods to ensure proper representation of these populations which will be discussed in the next section.

LIMITATIONS AND NEXT STEPS

Discussing next steps cannot be complete without first discussing some of the limitations we initially faced. First, finding COVID-19 data at a level that is consistent across the state usually differs by state. Additionally, we encountered issues with finding COVID-19 infection rates based on food access data for a few different cities such as Los Angeles, New York City, Atlanta, and Detroit. Due to the lack of compatible data, we then changed our scope to focus on one state based on the varying granularity of data from the state and city-level.

Second, while looking at food access information provided by the government, we found the 2019 Food Access Research Atlas which provided data at the Census tract level. This information was difficult to join with COVID-19 data because COVID-19 data had been distinguished by county and zip code. We then tried to convert the census tracts to zip codes and vice versa, but we then discovered that converting census tracts to zip codes is an industry-wide challenge when it comes to working with public census data. The following quote demonstrates the complexities with matching census tracts to zip code boundaries:

"ZIP Codes are problematic, because their boundaries are not created for analytical purposes like other geographies. ZIP Codes were designed to more efficiently deliver mail, not as geographies to be used for analysis. Because of their nature, the boundaries vary in size and shape that amplifies a common, adverse statistical problem when used for analysis." [12]

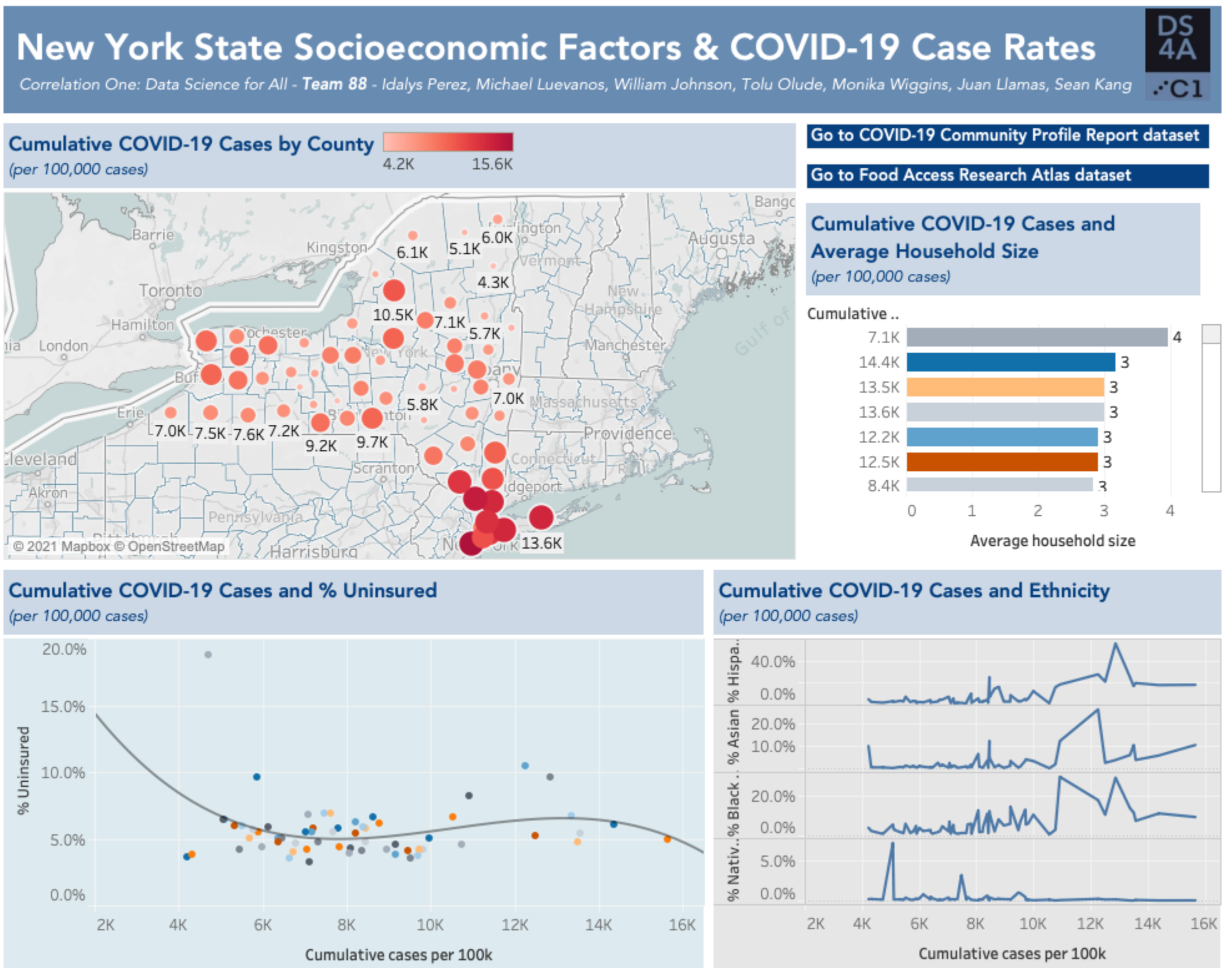
As a result, we aggregated the food access data based on the county level and joined it with the COVID-19 data we had by county. Then, we framed the results by counties in New York and utilized the case count per 100K residents ratio provided in the Health and Human Services COVID-19 Community Profile Report data.[11]

Another limitation with the data we used is how the 2019 Food Access Research Atlas has population demographics based on the 2010 Census. While this has been considered valid in similar studies we found[10], we believe that the demographic information would need to be updated with more recent data from the 2020 Census in order to ensure populations are being more accurately represented. This also would provide a more accurate reflection of any disparities in COVID-19 cases based on population demographics.

Lastly, another issue we encountered is the possibility that our findings may change because of the resurgence of the COVID-19 virus. The increase in cases may change the ways in which communities are impacted, especially those who have the specific socioeconomic factors that we observed throughout our project. Therefore, a next step moving forward would be to update current COVID-19 case rates to see how the surge in cases may be impacting certain counties differently in New York. By putting the data in our live server we could have real time COVID-19 information for the state's counties within our Tableau interactive dashboard.

INTERACTIVE DASHBOARD

[TeamProject_Group88 | Tableau Public]



RECOMMENDATIONS

The following points are our team's recommendations based on the previous analysis we conducted:

- COVID-19 awareness: With the Latino population percentage showing a positive relationship to COVID-19 case rates, it is imperative for the State of New York to target training, awareness, and vaccination efforts within the Latino community, especially in the counties in which we saw the greatest impact.

- Investigate survey methodology: By further researching the survey methodology and representation of certain demographics, it can help provide additional perspective on how well minority communities are being represented in the data and other future analysis that impacts state-wide public policy. This can ensure that minority and other low-income communities will be considered properly within the scope of targeted COVID-19 vaccination and prevention strategies.

WORKS CITED

1. World Health Organization. (2020, March 11). *WHO director-general's opening remarks at the media briefing On COVID-19 - 11 MARCH 2020*. World Health Organization. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-COVID-19---11-march-2020
2. Garg S, Kim L, Whitaker M, et al. Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 — COVID-NET, 14 States, March 1–30, 2020. MMWR Morb Mortal Wkly Rep 2020;69:458–464. DOI: http://dx.doi.org/10.15585/mmwr.mm6915e3external_icon
3. Martinez DA, Hinson JS, Klein EY, et al. SARS-CoV-2 Positivity Rate for Latinos in the Baltimore–Washington, DC Region. *JAMA.* 2020;324(4):392–395. doi:10.1001/jama.2020.11374
4. Health insurance historical tables: HIC series. US Census Bureau. Posted September 26, 2019. Accessed June 6, 2020. Health Insurance Historical Tables - HHI Series
5. *How risk of exposure to the coronavirus at work varies by ...* (n.d.). https://www.urban.org/sites/default/files/publication/103278/how-risk-of-exposure-to-the-coronavirus-at-work-varies.pdf.
6. Interactive Web Tool Maps Food Deserts, Provides Key Data
7. Karpyn, Allison E et al. “The changing landscape of food deserts.” *UNSCN nutrition* vol. 44 (2019): 46-53.
8. Corapi, Sarah. “Why it takes more than a grocery store to eliminate a ‘food desert’” PBS. Feb 3, 2014. Web. Feb 27, 2017. http://www.pbs.org/newshour/updates/takes-grocery-store-eliminate-food-desert/
9. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html
10. Liang, C.-L., Kurkalova, L., Hashemi Beni, L., Mulrooney, T., Jha, M., Miao, H., & Monty, G. (2021). Introducing an innovative design to examine human-environment dynamics of food deserts responding to COVID-19. *Journal of Agriculture, Food Systems, and Community Development*, *10*(2), 123–133. https://doi.org/10.5304/jafscd.2021.102.037
11. https://healthdata.gov/Health/COVID-19-Community-Profile-Report/gqxm-d9w9
12. Din, Alexander & Wilson, Ron (2018). Understanding and Enhancing the U.S. Department of Housing and Urban Development’s ZIP Code Crosswalk Files. Cityscape: *A Journal of Policy Development and Research: Volume 20 Number 2*. 278. https://www.huduser.gov/portal/periodicals/cityscpe/vol20num2/ch16.pdf
13. Ann C. Foster, “Hispanic household spending in 2015,” Beyond the Numbers: Prices & Spending, vol. 6, no. 9 (U.S. Bureau of Labor Statistics, July 2017), https://www.bls.gov/opub/btn/volume-6/hispanic-household-spending-in-2015.html

14. CDC.gov
https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html