Where is Childhood Obesity Highest in Denver

Weston Grewe

Spatial Data Analysis

Fall 2021

Introduction/Background.

Childhood obesity, defined as a Body Mass Index (BMI) of greater than 30, has been a rising issue in the United States of America for the past several years. In the year 2000, 13.9% of children were obese compared to 18.5% of children by 2016 according to a study in the journal Global Pediatric Health [3]. The Centers for Disease Control and Prevention (CDC) notes that childhood obesity increases the likelihood of certain physical harms such as high blood pressure, high cholesterol, type 2 diabetes, joint problems, and many other ailments [1]. The CDC also acknowledges that childhood obesity is linked to psychological conditions such as anxiety and depression which can arise from low self-esteem or bullying linked to obesity [1].

The body positivity movement has illustrated how damaging it can be to people when a conversation about health is centered around weight. In this report, I investigate obesity and discuss policy recommendations that include phrases such as "healthy food." The purpose of the report is not to bring shame to any person. There is no perfect weight; people come in all shapes and sizes and individual health should be evaluated by a doctor. That said, however, government policies often play a role in shaping the health of a community and governments should be held accountable. This report aims to identify communities that the government has been failing and help inspire policy change.

In recent years, many policy makers and leaders have shifted from an equality-based approach of policy design to an equity-based approach. Broadly, equity-based policy targets responses and resources to communities most in need, whereas equality-based policy seeks to distribute responses and resources to all communities equally. In an equality-based approach, distributing resources is "easy" in the sense that if a government has x number of dollars to spend on n number of citizens, then a fair response would be to distribute x/n dollars to each person. This is not the case for an equity-based approach. An equity-based approach would distribute more than x/n dollars to people in communities that have been traditionally marginalized and less than x/n dollars to people in communities that have been traditionally privileged. In all instances of equity-based policy making, there must be a quantification of which communities are more in need than others.

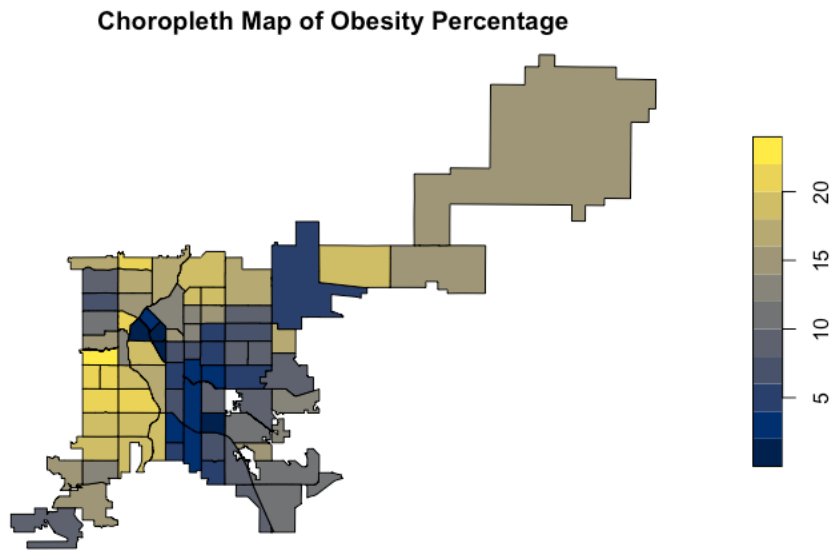
In this statistical analysis, I seek to solve this quantification for childhood obesity at the neighborhood level in Denver, Colorado. The city of Denver has a finite number of resources to spend on children's health, this analysis will help determine in which neighborhoods resources should be spent to best combat childhood obesity. In statistical terms, this report endeavors to determine if there exists neighborhoods or clusters of neighborhoods that exhibit more cases of childhood obesity than what we would expect under the constant risk hypothesis. If these clusters exist, then policies that target these clusters are appropriate. If said clusters do not exist, then an approach that treats each neighborhood equally may be appropriate.

I analyze the childhood obesity data set for the years 2014-2016 from the city of Denver's Open Data Catalog [6] using various regional count techniques including scan-based methods and measures of global spatial autocorrelation. The data was collected by Colorado Department of Public Health and Environment by combining each health systems' electronic records into a single database. Some health records, aggregated at some spatial level to preserve anonymity, such as this data set are made available to the public. This data set is aggregated at the neighborhood level and contains information on children who visited a physician between 2014 and 2016.

Data.

For this analysis, I draw from 2 data sources. The most important data set is the previously mentioned Childhood Obesity 2014-2016 data set [6]. The relevant fields of this data are titled `NBHD_ID`, `COUNT_CHILDREN_INREGISTRYBMI`, and `PERCENT_OBESE`. These fields identify a record with a neighborhood, identify the number of children in the study, and give the percentage of those children that are obese, respectively. To apply regional count techniques to this data set, a count of the children with obesity for each neighborhood is required. This is easy to remedy by adding another field (for my analysis titled `NUMBER_OBESE`) that is the `COUNT_CHILDREN_INREGISTRYBMI` variable multiplied with the `PERCENT_OBESE` variable. The average obesity rate for a Denver neighborhood (not accounting for size of neighborhood) is approximately 0.12; the obesity rate for Denver overall is approximately 0.15. The neighborhood with the highest percentage of children with obesity is Barnum, located in West Denver, with 21.91% of children having obesity. Platt Park in South Denver has the lowest rate at 2.08%.

The second data set used is the Denver Statistical Neighborhoods data set [7]. This data set is a shape file that allows the neighborhoods of Denver to be plotted on a map. Additionally, this data set contains a `NBHD_ID` field that allows neighborhoods to be identified. Therefore, in R, a merge function can combine the Childhood Obesity 2014-2016 data set with the Denver Statistical Neighborhood data set by merging with `NBHD_ID`.



The plot illustrates the percent of children who are obese in each neighborhood. From the plot alone it is difficult to estimate where clusters may be. However, there are lightly colored areas in West Denver and North Denver. This informs us that there may be clustering beyond what is expected under the constant risk hypothesis in these regions.

Methods.

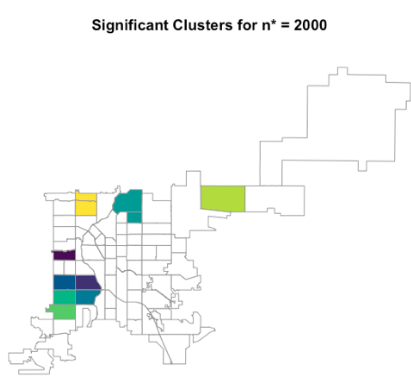
To analyze the data, I will apply six different tests to the data set. The first three tests are Turnbull's cluster evaluation permutation procedure (CEPP), the Besag-Newell approach, and the Spatial Scan method. Each test is rooted in scanning windows of neighborhoods and test for both clustering and clusters. Each tests returns a p-value indicating statistical significance clustering and a list of most likely clusters. Therefore, these tests provide a means for determining where the clusters are in the data. To provide further evidence conclusions three more tests are applied: the Moran's I test, Geary's C test, and the Tango index. All three of these methods are tests of global spatial autocorrelation. Although, these tests do not yield clusters, only information on whether there is clustering. If our first three tests yield significant clusters and these last three tests yield significant evidence of clustering, then the conclusion that clusters exist is even stronger.

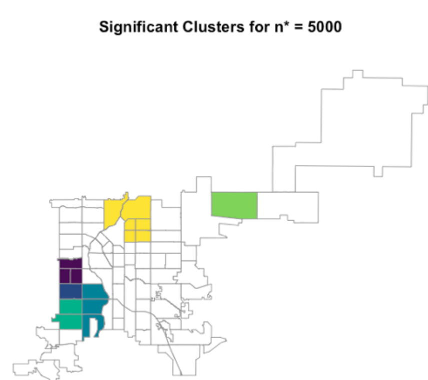
Results.

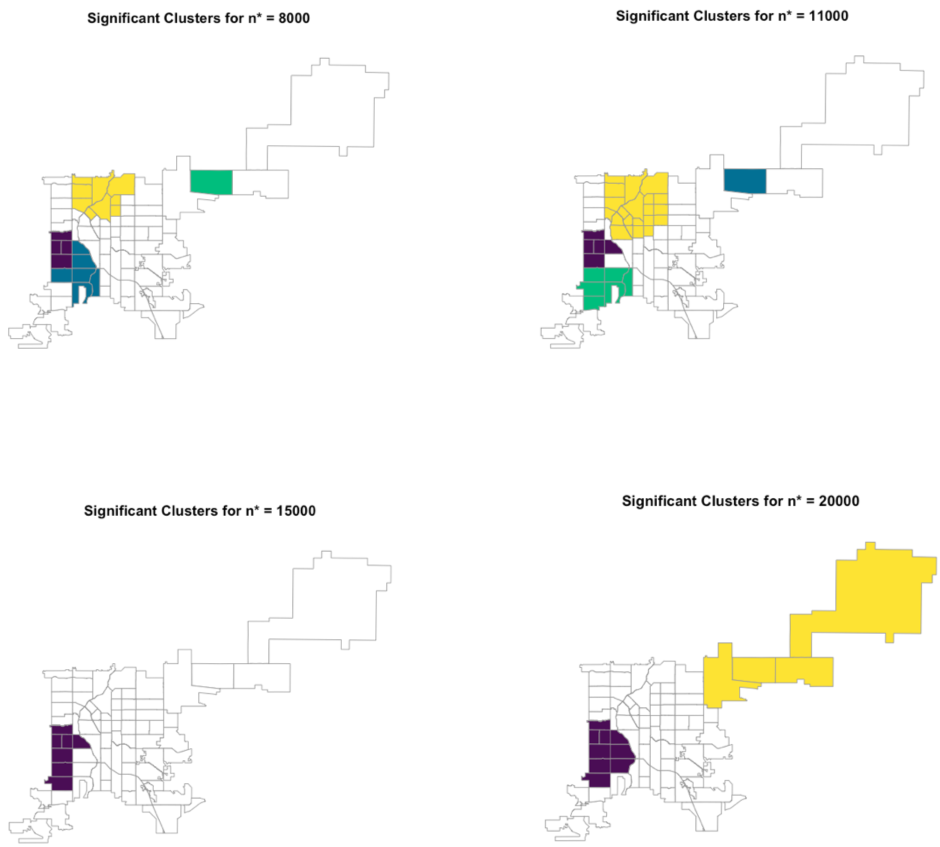
The results section is broken into four subsections, one for each scan method and 1 for the global spatial autocorrelation methods.

CEPP.

The CEPP is implemented as a Monte-Carlo based test of clustering, in this case with 499 simulations. It tests to see if there is a window of neighborhoods in Denver that contain at least n* children that has significantly more cases than what is expected under the constant risk hypothesis. The CEPP allows for the user to set the minimum number of people for a collection of neighborhoods to contain to be considered as a clustered. The larger n* becomes the larger clusters will need to be, n* provides some control for spatial scale. From the Childhood Obesity 2014-2016 data set, it can be found that the average neighborhood consists of 1200 children who have their BMI registered. This informs the scales on which n* should be set. We will choose the scales {2000, 5000, 8000, 11000, 15000, 20000}.





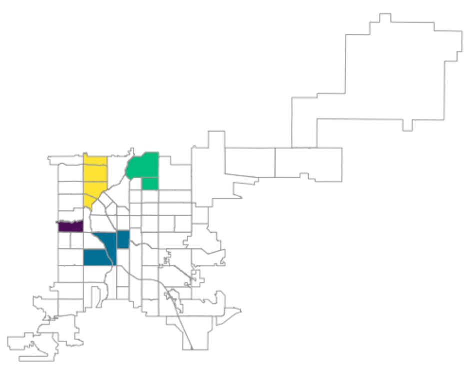


There are large cluster at scales n* = 15,000; 20,000 in West Denver. This cluster is fractured into smaller clusters for smaller spatial scales. There is a medium sized cluster in North Denver seen in cluster scales n* = 5,000; 8,000; 11,000. For larger scales this cluster no longer exists and at the small scale (n* = 2000) the cluster is fractured into 2 clusters. Finally, at the small scale there is a cluster consisting of just one neighborhood in Northeast Denver. This cluster disappears for scale n* = 15,000 and then reappears for the scale n* = 20,000 and has many other neighborhoods included. It may be safe to conclude that the clustering is coming from this one neighborhood and the other neighborhoods just helped the window reach the minimum population required by the CEPP test. Each cluster presented has Monte Carlo p-value less than 0.008, exact values are presented in the appendix. We can conclude that there are several windows of Denver neighborhoods containing at least n* children, for the chosen values of n*, that has significantly more cases than what is expected under the constant risk hypothesis.

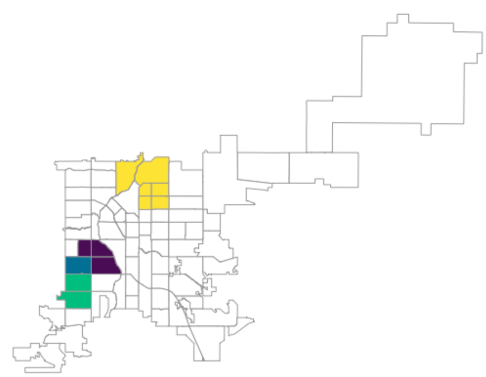
Besag-Newell Method.

The Besag-Newell Method is similar to CEPP in that clusters are found by analyzing groupings of neighborhoods. This method also uses a Monte Carlo approach to determine if the most compact window of Denver neighborhoods (in terms of number of children) with at least c* cases is significantly more compact than what is expected under the constant risk hypothesis. The Besag-Newell Method differs from CEPP in that a minimum case count c* is fixed instead of a population count.

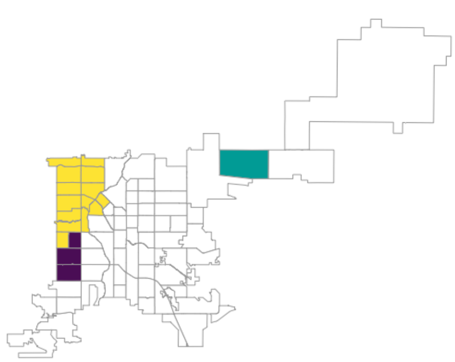
Plot of Besag-Newell with c* = 500



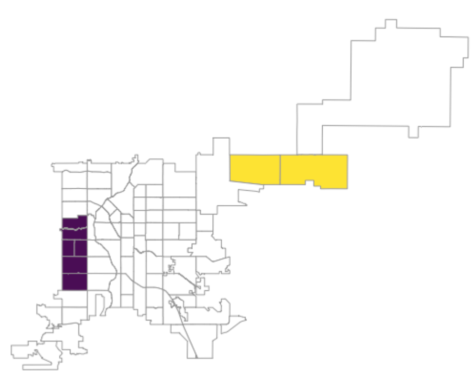
Plot of Besag-Newell with c* = 1000



Plot of Besag-Newell with c* = 2000



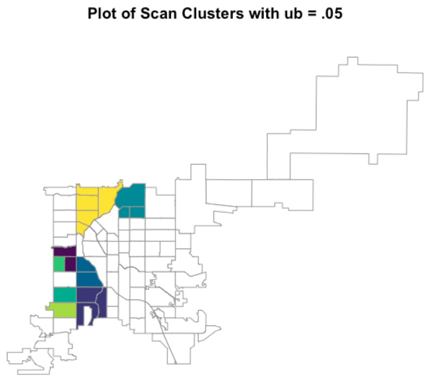
Plot of Besag-Newell with c* = 3000

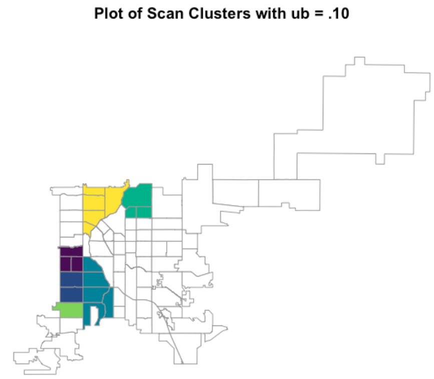


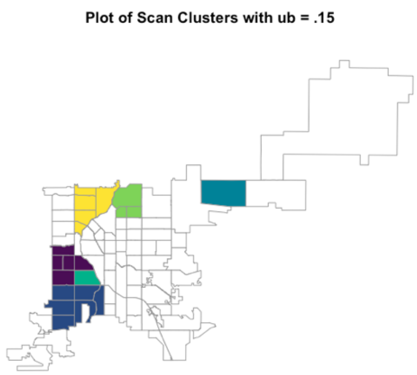
After testing the Besag-Newell on several different spatial scales, some trends can be identified. The smallest scale, c* = 500, shows some signficant clusters in North Denver and West Denver. This aligns with what was found in the CEPP. The next scale, c* = 1000, also finds clustering in North Denver and West Denver. However, the yellow cluster in Northwest Denver for c* = 500 has disappeared. This is evidence that the yellow cluster for c* = 500 is just a small cluster. For c* = 2000, the Montbello neighborhood in Northeast Denver is now a cluster. Again this aligns with what was found when using CEPP. When c*=3000 the cluster in North Denver is lost but there is still a cluster in West Denver. Additionally, the cluster in Northeast Denver adds a second neighborhood. The p-values for each of these clusters are all less than 0.08 with most below 0.001. For the chosen values of c*, we conclude that the most compact window of Denver neighborhoods with at least c* cases is significantly more compact than what is expected under the constant risk hypothesis. Based on these results, the Besag-Newell method has strenthened evidence that there are clusters in West Denver, North Denver, and Northeast Denver.

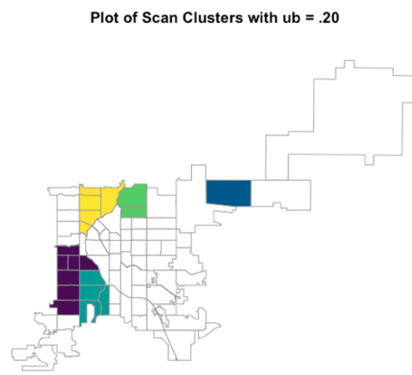
Spatial Scan Test.

The final scan method that we use is the Spatial Scan Test. The Spatial Scan Test determines if the most likely cluster (in terms of the local rate of children with obesity inside the cluster compared to outside the cluster) is consistent with what is expected under the constant risk hypothesis. The Spatial Scan Test does not fix a case or population count instead the Spatial Scan Test has an upper-bound parameter that controls of percentage of the population allowed in a cluster. In the figure below, most likely clusters are plotted with a varying upper-bound (ub) parameter. We use 999 simulations for the Monte Carlo implementation.









In this analysis the values {.05, .10, .15, .20} are chosen for the upper-bound to illustrate the differences between smaller and larger scales. In all four plots there are two significant clusters in North Denver. In all four plots there are various clusters that appear in West Denver. In the last two plots the Montbello neighborhood in Northeast Denver appears as a cluster. Each cluster has p-value less than 0.04. For the chosen upper bound values, we conclude the most likely cluster is non consistent with what is expected under the constant risk hypothesis. This conclusion, along with the locations of the clusters, further strengthens the evidence that there are clusters in North Denver, West Denver and Northeast Denver.

Spatial Autocorrelation Methods.

Indices of global spatial autocorrelation summarize the degree to which similar rates of childhood obesity tend to occur near each other. Extreme values an index of global spatial autocorrelation imply clustering.

Test	Monte Carlo p-Value (999 Simulations)
Moran's I	0.001
Geary's C	0.001
Tango's Index ($\varkappa = 1$)	0.001
Tango's Index ($\varkappa = 7$)	0.001

The p-values Moran's I and Geary's C are found using a Monte Carlo test with 999 simulations and data simulated under the constant risk hypothesis. The p-values of 0.001 are very small and inform us to reject the null hypothesis of no global spatial autocorrelation. This is strong evidence that there are clusters in the data, further bolstering our conclusions from the scan methods.

For Tango's Index, tuning parameters ($\varkappa$) 1 and 7 are chosen. Choice of tuning parameter allows for large- and small-scale clusters to be detected. The Monte Carlo p-value, with 999 simulations, for both instances is 0.001 and therefore implies that there is evidence of clustering both on large scales and small scales. Again, this matches the conclusions from the scan method and strengthens conclusions from that section.

Presentation of Neighborhoods.

In the table below, we present the three clusters that were present in each of the CEPP, Besag-Newell method, and Spatial Scan tests. The presented clusters are comprised of the neighborhoods that appeared in at least two of the scan instances. This is not to say that these are the true statistical clusters. Instead, this should be used as a guide for policymaking. If a policy is targeting childhood obesity in the neighborhoods of West Denver then it should encompass at least all of the neighborhoods in the West Denver column below.

Cluster	West Denver	North Denver	Northeast Denver
Neighborhoods (nbhd ID, nbhd)	1 Athmar Park 4 Barnum 5 Barnum West 34 Harvey Park 43 Mar Lee 54 Ruby Hill 67 Valverde 68 Villa Park 76 Westwood	11 Chaffee Park 17 Clayton 18 Cole 25 Elyria Swans 29 Globeville 36 Highland 39 Jefferson Park 55 Skyland 62 Sunnyside 77 Whittier	28 Gateway/Green Valley Ranch 45 Montbello

Conclusions.

Each test applied gave us evidence that there is clustering of childhood obesity in Denver. The CEPP, Besag-Newell, and Spatial Scan methods all presented West, North, and Northeast Denver as areas of clustering of childhood obesity. We conclude that there is evidence of clustering of childhood obesity cases in West Denver, North Denver, and Northeast Denver.

Policy Recommendations.

There are two policy paradigms that can be used as umbrellas for a collection of policies. The first paradigm is a school-based approach, and the second paradigm is a community-based approach.

Typically, children are present in school for a minimum of 30 hours per week, thus it is easy for schools to leverage lifestyle changes on their students. Additionally, the federal government does not exert much control over schools. Schools are primarily overseen by the school district, city government, and state government. The control that the school district and city government have makes it easy to implement changes in targeted areas. There are several policies that the school districts and city government can implement.

Schools may implement various nutrition standards. Restricting the sale of high-fat, high-sugar foods at schools has been done in states such as California, Connecticut, and Maine and has been shown to reduce children's caloric intake by about 160 calories per day[2]. Snack food sales are an important revenue source for schools; however, it has also been found that when restricting snack food sales that revenue is recaptured through a higher number of students purchasing school lunch [2]. This leads to a second point. Schools can focus on improving the quality of school lunches. Improved quality will encourage more students to eat the healthy school lunch as opposed to skipping lunch and having snack foods instead. Improving the quality of school lunches may not be cost-free, in that case funding should be directed at the communities that need help the most: the clusters found in West Denver, North Denver, and Montbello.

Another solution schools can implement is increasing time allotted physical education and sports. Unfortunately, there is not strong evidence to conclude that this solution will make an impact. This is because there is no longitudinal study on how the health of children changes after increasing time allotted for physical education. It would be worthwhile for the City of Denver to conduct such a study.

The community-based approach consists of policies that can be implemented in the communities that these children live in. The Children's Hospital of Colorado gives several examples of how neighborhoods can be social determinants for a child's weight and overall health. For instance, children may live in a neighborhood where it is unsafe to play outside, come from a family that cannot afford healthy food, or have lack of access to community sports [4]. The City of Denver should begin a study to determine which of these factors are most likely impacting the neighborhoods in the clusters and then implement solutions to solve those factors. This may mean building new parks or further subsidizing youth sports.

Future Work.

The City of Denver provides only a short methodology for how the data was collected. Biases that may exist in the data due to collection methodology are unknown. Only one bias is known: the data was derived from children who have access to a doctor. This excludes many of the poorest children in Denver and those whose parents are in the Medicaid gap (the valley of people too wealthy to qualify for free healthcare but too poor to afford private healthcare). A more representative survey should be collected and analyzed.

Clusters have been shown to exist, but we lack detailed information about these clusters. It is not wise to recommend specific policies without an understanding of what is causing significant levels of childhood obesity in these neighborhoods. For instance, it could be that children in a cluster get enough exercise but have an unhealthy diet. In this case, lengthening recess time and increase access to sports would have little effect as it is not the cause of the problem.

In the future, it would be best to implement a finer analysis of these clusters to understand the root causes to the problem. One approach would be to create a generalized linear model to find factors correlated with childhood obesity in Denver. Additional features at the Denver neighborhood level around the same time period can be found in the American Community Survey for Denver neighborhoods [5]. More data may be required as well such as the amount of time children spend in P.E., information on diet, and daily amount of exercise all aggregated by neighborhood.

In this analysis, I learned that just finding a data set is difficult, I went through several data sets before deciding on this one. Plotting was more of challenge than I expected, especially the choropleth. I learned I must use the `st_read` function for accurate results. At the end of the project, I do feel a lot more comfortable executing a spatial cluster analysis with little guidance.

References.

Articles.

1. "Childhood Obesity Causes & Consequences." *Centers for Disease Control and Prevention*, 19 Mar. 2021, www.cdc.gov/obesity/childhood/causes.html. Accessed 4 Dec. 2021.
2. Chriqui, Jamie F. "Obesity Prevention Policies in U.S. States and Localities: Lessons from the Field." *Current Obesity Reports*, vol. 2, no. 3, 2013, pp. 200–10. doi:10.1007/s13679-013-0063-x.
3. Sanyaolu, Adekunle, et al. "Childhood and Adolescent Obesity in the United States: A Public Health Concern." *Global Pediatric Health*, vol. 6, 2019, doi:10.1177/2333794x19891305.

4. "Transforming Community Health." *Children's Hospital of Colorado*, www.childrenscolorado.org/community/community-health. Accessed 4 Dec. 2021.

Data.

5. "American Community Survey Nbrhd (2013–2017)." *Denver Open Data Catalog*, www.denvergov.org/opendata/dataset/city-and-county-of-denver-american-community-survey-nbrhd-2013-2017. Accessed 4 Dec. 2021.
6. "Childhood Obesity 2014–2016." *Denver Open Data Catalog*, www.denvergov.org/opendata/dataset/city-and-county-of-denver-childhood-obesity-2014-2016. Accessed 4 Dec. 2021.
7. "Statistical Neighborhoods." *Denver Open Data Catalog*, www.denvergov.org/opendata/dataset/city-and-county-of-denver-statistical-neighborhoods. Accessed 4 Dec. 2021.