HI 2020

Data Exploration Using Python

Name of Team: Gold

Aparcana, Leslie

Brandebura, Stephanie

Kelley, Mark

Letter, Christina

Sanden, Bailey

Stegner, Alyssa

Title of Project:

Comprehensive Analysis of Gender and Geographical Location in the Years 2011 and 2020 on Obesity Prevalence in the U.S. along the East Coast

Database Title:

BRFSS: Table of Overweight and Obesity (BMI)

Publication Date: September 17, 2020

Link to Database:

https://chronicdata.cdc.gov/Behavioral-Risk-Factors/BRFSS-Table-of-Overweight-and-Obesity-BMI-/fqb7-mgjf

I. Introduction

Obesity is a health condition prevalent worldwide. The factor that determines obesity is Body Mass Index (BMI) of 30 kg/m^2 or higher, as stated by the World Health Organization. Considering this BMI value, statistics indicate that since 1975, obesity has nearly tripled, and per data from 2016, 650 million adults 18 years and older are obese, which accounts for 13% of the adult population (World Health Organization). Within the United States, prevalence of obesity in adults remains high. According to the Centers for Disease Control (CDC), in their latest publication that gathered obesity data from 2017-2018 (Hales et al., 2020), obesity prevalence was 42.4%, which demonstrates a marked increase from 1999-2000 data that indicated 30.5% prevalence in adults 20 years and over. Furthermore, obesity data for the years of 2011-2012 demonstrate an increase to 34.9% obesity prevalence in adults 20 years and older, denoting a 4.4% increment in the overall population from the 1999-2000 data (Ogden et al., 2013). High obesity prevalence is also observed in the most recent 18^{th} annual report from the Trust for America's Health (TFAH), which is based on newly released 2020 data from the CDC's Behavioral Risk Factors Surveillance System (BRFSS). This report found that sixteen states demonstrate adult obesity prevalence of at least 35%, demonstrating an increase in the number of states with obesity at that rate that was previously composed of twelve states in 2019. This report associates the increase in obesity to the effects generated by the COVID-19 pandemic, which resulted in the increment of food insecurity, decrease in physical activity, and changes in eating habits (TFAH).

The trend in obesity increase is projected to continue rising. In a study conducted by Wang et al.(2020), BRFSS data from 1984 through 2016 as well as National Health and Nutrition Examination Survey (NHANES) data from 1999 through 2016 were analyzed, and it was found that central obesity would increase to 55.6% in men and 80% in women by 2030. Additionally,

besides central obesity, this study was also able to make projections for overall obesity based on race and gender in the United States. Within the 95% Confidence Interval, among men, Mexican-Americans would have a projected obesity prevalence of 64.1% in 2030, followed by African Americans with a projected 55.1% prevalence. Women; however, would present a higher obesity prevalence by 2030, demonstrating 66.3% among African Americans, followed by 66.0% obesity prevalence from Mexican-Americans (Wang et al.). Given these projections and the evidence of actual increase in overall obesity in the United States, profound research has been conducted to delve into the factors that contribute to obesity. Among these studies, observed patterns associate obesity with gender differences and regional population location.

The association of gender differences in obesity indicates that within the adult U.S. population females tend to show higher obesity prevalence than their male counterparts. According to Arroyo-Johnson and Mincey (2016), based on the NHANES data from 2011-2014, 38.3% of women and 34.3% of males were obese from adults 20 years and older. Specifically, adults with ages 20-39 demonstrated 34.4% female obesity and 30.3% males' obesity, ages 40-59 indicated 42.1% and 38.3%, and ages 60 and over demonstrated 38.8% and 34.9%, respectively. These patterns are also supported by the findings from Hales et al. (2018) whose study also acquired data from NHANES for the 2015-2016 period and compared it to NHANES results from previous years, starting with 2007. Results from the 2015-2016 NHANES data indicate an adult obesity prevalence of 39.6%, whereas data from 2007-2008 NHANES reveals an obesity prevalence of 33.7%, which denotes a statistically significant difference ($p = 0.001$). Remarkably, the data analyzed also demonstrates that along with the trend of increasing obesity among the U.S. adult population, the prevalence has also increased significantly among women. Specifically, the overall positive trend of obesity starting from 2007 to 2016 indicates 35.4% of female obesity and 32.2%

male obesity in 2007-2008, 40.4% and 35.0% in 2013-2014, and 41.1% and 37.9% in 2015-2016, respectively, thus denoting the highest increase in female obesity during the 2015-2016 period. Reasonably, the linear trend for males has a $p = 0.05$; however, the linear trend for females indicates $p<0.001$ (Hales et al., 2018). These values illustrate that females have had the most significant increase in obesity prevalence.

In addition to gender, geographical location seems to have an association with obesity prevalence in the U.S. The study conducted by Sung and Etemadifar (2019) analyzed information from 2015 BRFSS and Selected Metropolitan/Micropolitan Area Risk Trends (SMART). They found that there are statistically significant differences in obesity rates among geographical metropolitan and micropolitan statistical areas (130 MMSA) at the 0.001 level. Additionally, their study indicates that obesity rates per region are higher in the South and Midwest, with 31.08% and 31.65% obesity rate, respectively, whereas the Northeast presents an obesity rate of 29.96%. Similarly, clusters of obesity prevalence have been observed in the study by Candice et al. (2015), which reports that obesity prevalence was concentrated in Southern counties that included Louisiana, Mississippi, Arkansas, and Alabama. The association of geographical location and obesity prevalence is further supported by Lundeen et al (2018). Their publication indicates that 2016 BRFSS data shows the prevalence of obesity was 34.2% among adults living in nonmetropolitan counties, whereas the prevalence of obesity of adults living in metropolitan areas was 28.7%. Furthermore, BRFSS data demonstrated that obesity prevalence was the highest in the South and Midwest with 32.0% and 31.4%, respectively; however, the Northeast region had a prevalence of 26.9% (Lundeen et al., 2018). Differences in obesity prevalence among regions are evident.

II. Specific Aims

The above studies encompass the foundations of this work. Given the relevant factors affecting obesity prevalence in the United States, we have the following aims:

- Perform a comprehensive analysis of obesity among the Northeast and Southern regions of the United States by gender in the years of 2011 and 2020 to determine if the BRFSS dataset mirrors the observed increase in obesity rates found in studies.
- Perform specific comparisons using region, sex and year as independent variables and obesity percentage as the dependent variable.

In order to achieve the aims above, the following hypotheses will be tested:

1. **Comparing overall population obesity over Time:**

   1.1. Population across the East Coast in 2011 was not more obese than population across the East coast in 2020.

2. **Comparing overall population obesity over time by region:**

   2.1. Population in the Northeast in 2011 was not more obese than population in the Northeast in 2020.

   2.2. Population in the South in 211 is not more obese than population in the South in 2020.

3. **Comparing overall population obesity by region in a single year:**

   3.1. Population in the South in 2011 is not more obese than population in the Northeast in 2011.

   3.2. Population in the South in 2020 is not more obese than population in the Northeast in 2020.

4. **Comparing population obesity by gender in a single year within the same U.S. region.**

    4.1. Females in the Northeast in 2011 are not more obese than males in the Northeast in 2020.

    4.2. Females in the Northeast in 2020 are not more obese than males in the Northeast in 2020.

    4.3. Females in the South in 2011 were not more obese than males in the South in 2011.

    4.4 Females in the South in 2020 were not more obese than males in the South in 2020.

IIIa. Methods - Creating the Data Frame for Analysis

We imported the following features from the BRFSS BMI Dataset: Year, Locationabbr, Response, Break_Out, and Sample_Size as a data frame. The following steps were repeated to create separate data frames for the following isolated datasets: Overall in 2020, Overall in 2011, Males in 2020, Males in 2011, Females in 2020, and Females in 2011. This process was done so we could separately calculate the Data_value feature by hand since the feature native to the original file was less of an accurate measure for our analysis. For each of the six data frames we filtered by year, states of interest, and breakout category. For example, when creating the Male 2020 data frame, we filtered the year by 2020, the Break_out feature by “Male”, and the following states located in the two regions we would be comparing. At this point in our process, we were still including all of the BMI response types: Underweight, Normal, Overweight, and Obese, in order to calculate the total sample size.

To calculate the total sample size, we pivoted the data frame so that the states listed in the rows of the “Locationabbr” feature would become the columns, and the sample_size of the responses would become the rows under each state. This was done to perform a sum function on these columns and then develop the Total_Sample_Size_2020 data frame that we would later join into the master data frame for analysis.

For each of the six data frames we also needed to filter and isolate by the obese response type so we could isolate only the participants in the survey that selected “obese”, since this was the focus of our analysis. At this time, we joined the Total_Sample_Size_2020 data frame with the isolated obese responses and used these two features to calculate the new Data_Value feature (also known as the % of the total population of interest that responded “obese” in each respective state). Lastly, we added a feature that defined which states belonged in which region (Northeast vs South).

Once our data_value was calculated for each of the six isolated data frames, they were added to our “datafile_analysis” data frame. When all six data frames were combined, three features were added to calculate the percent change in each category from 2011 to 2020. The final datafile_analysis data frame included the following features: Locationabbr, Region, Obese_Overall_2020, Obese_Overall_2011, Overall_Change_2011_2020, Obese_Males_2020, Obese_Males_2011, Males_Change_2011_2020, Obese_Females_2020, Obese_Females_2011, Females_Change_2011_2020. The header of our data frame featuring the first 5 rows can be seen below in Table 1.

Location abbr	Region	Obese_Overall_2020	Obese_Overall_2011	Overall_Change_2011_2020	Obese_Males_2020	Obese_Males_2011	Males_Change_2011_2020	Obese_Females_2020	Obese_Females_2011	Females_Change_2011_2020
AL	South	38.57%	33.08%	5.49%	36.91%	31.96%	4.95%	39.77%	33.69%	6.08%
AR	South	35.09%	30.59%	4.50%	33.87%	30.99%	2.88%	36.05%	30.36%	5.69%
CT	North-east	29.56%	24.76%	4.80%	29.15%	25.64%	3.51%	29.94%	24.18%	5.76%
DC	South	23.48%	23.74%	-0.26%	18.13%	18.71%	-0.58%	28.14%	27.00%	1.14%
DE	South	35.88%	29.89%	5.99%	35.33%	30.65%	4.68%	36.35%	29.39%	6.96%
...	...	...	...	...	...	...	...	...	...	...

Table 1: Sampling of data frame used for all analysis depicted in the IIIb. Methods- Exploratory Data Visualization and Descriptive Analysis section of our paper.

IIIb. Methods - Exploratory Data Visualization and Descriptive Analysis

To better understand our data set, we ran some descriptive statistics, using the Pandas package, on each of the subgroups. We looked at the average percentage of obese responses in each state for each subgroup, the median percentage, the standard deviation, the minimum, and the maximum. We also ran each of these descriptive statistics for the percent change from 2011 to 2020 for each subgroup.

Histograms were run, using the Matplotlib package, to provide a variety of visual aids to better understand the distribution of the prevalence in obesity amongst different years, regions, and genders. The histograms display the number of states (on the y axis) that had a percentage of obesity within subgroups of the dataset.

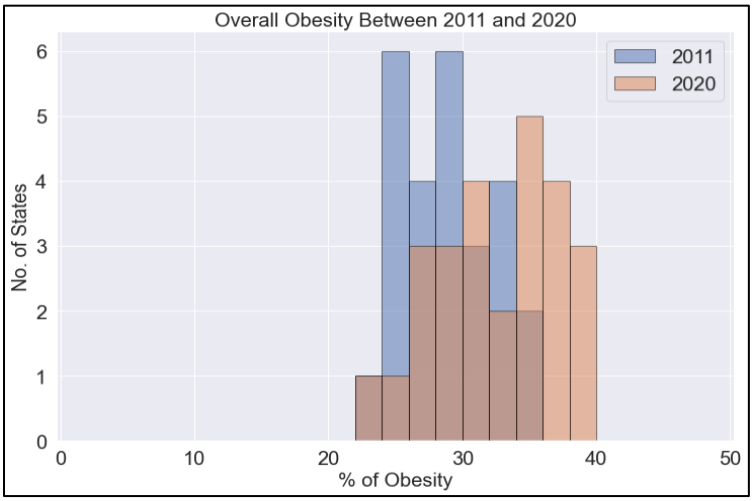

Figure 1. Histogram plot of overall % of obesity among both regions in 2011 (blue) and 2020 (orange).

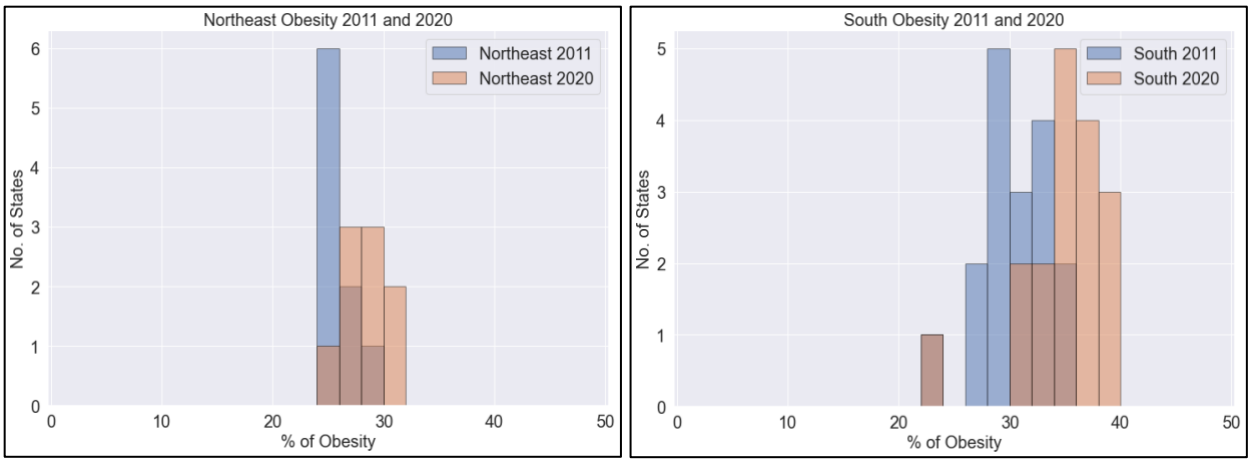

Figure 2. Histogram plots of overall % of obesity among each region (Northeast – left, South – right) in 2011 (blue) and 2020 (orange).

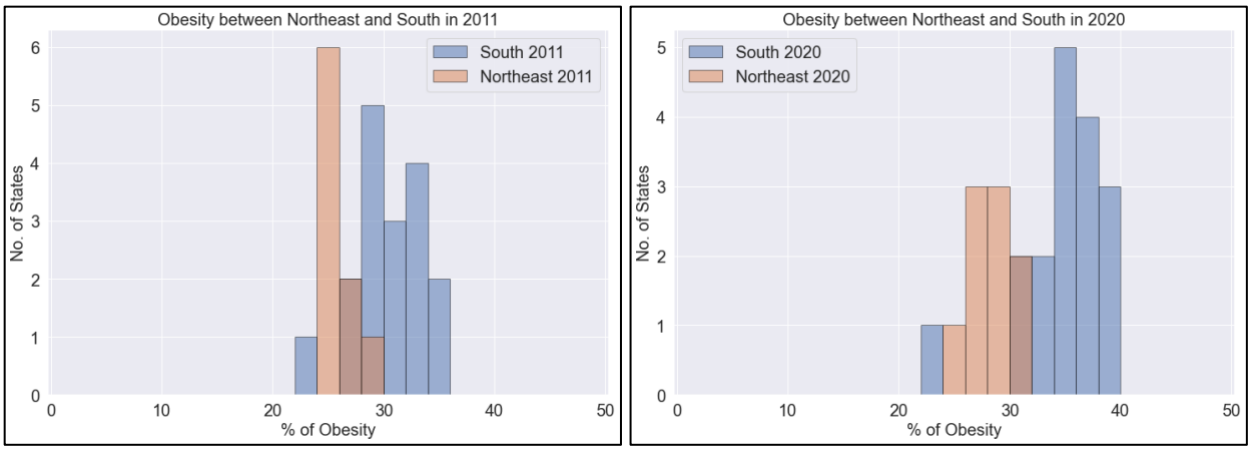

Figure 3. Histogram plots of overall % of obesity between each region (Northeast – orange, South – blue) in 2011 (left) and 2020 (right).

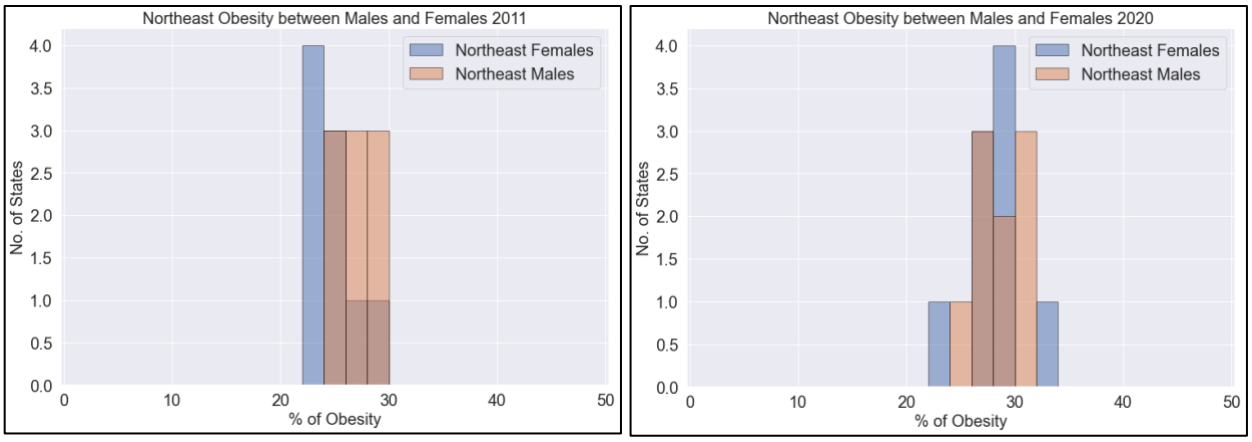

Figure 4. Histogram plots of overall % of obesity between each sex within Northeast region in 2011 (left) and 2020 (right).

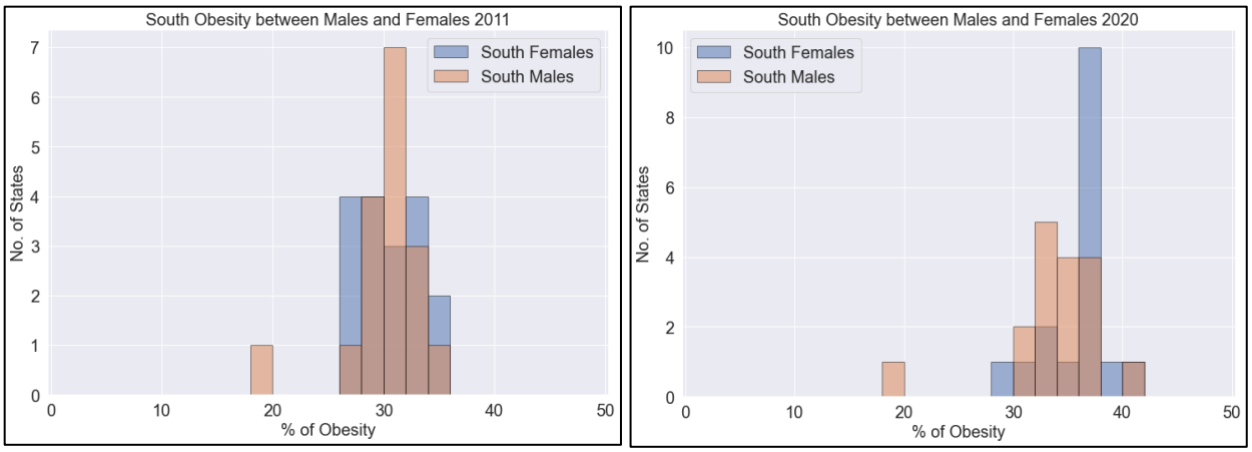

Figure 5. Histogram plots of overall % of obesity between each sex within South region in 2011 (left) and 2020 (right).

Violin plots were also run, using the Seaborn package, to help us have a different way to visualize and compare the distribution of obesity prevalence amongst specific subgroups of the dataset.

Figure 6. Violin plots of overall % of obesity among males between each region in 2011 (left) and 2020 (right).

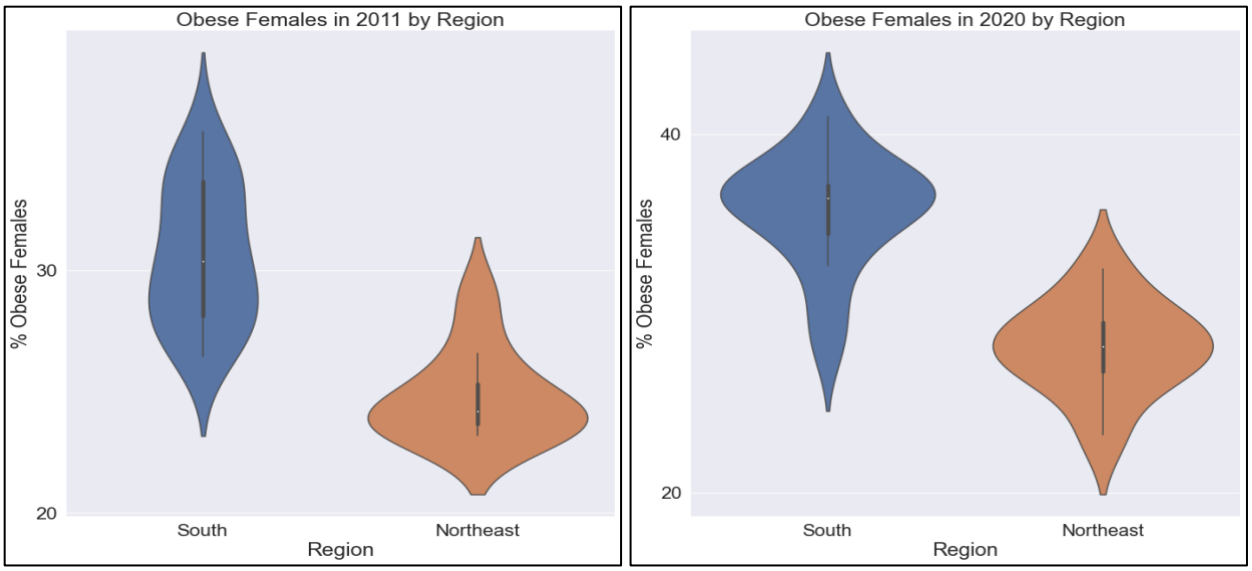

Figure 7. Violin plots of overall % of obesity among females between each region in 2011 (left) and 2020 (right).

We also produced waterfall bar graphs, using the Pandas package, to better visualize how the prevalence in obesity changed in each state from 2011-2020 for the overall population, as well as broken down by the two genders, and how it differed amongst the states in the two regions of interest.

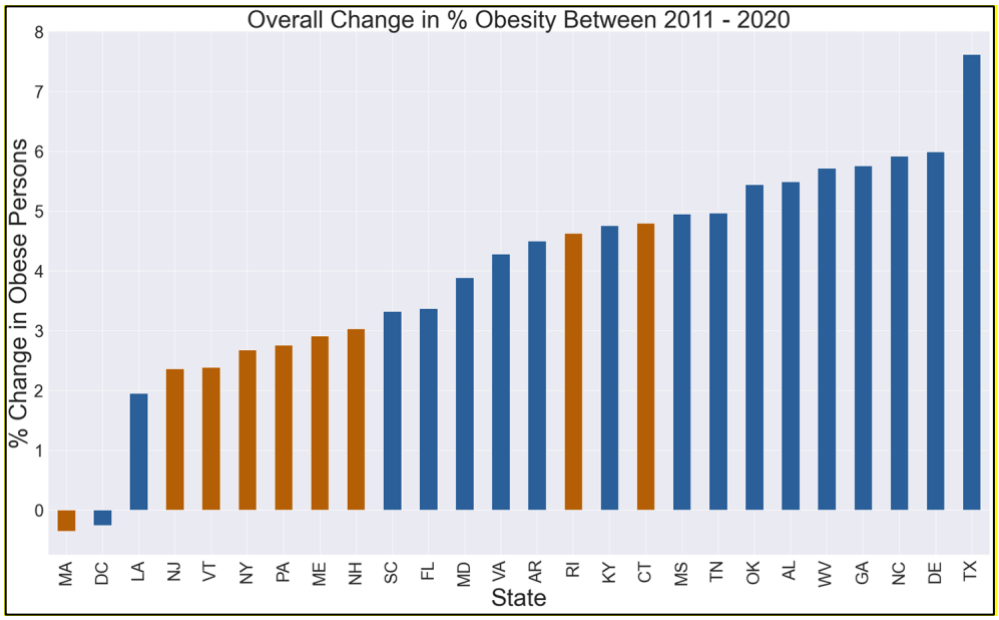

Figure 8. Waterfall plot depicting the change in overall % of obesity for each state between 2011 and 2020, color coded by region (Northeast – orange, South – blue).

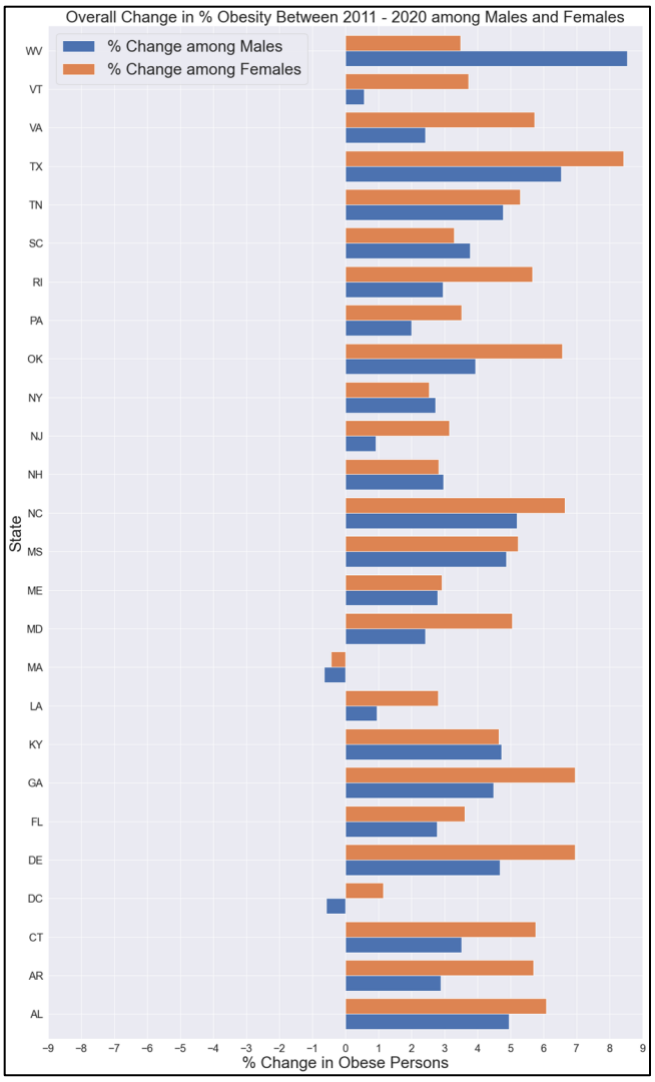

Figure 9. Bar plot depicting the change in overall % of obesity for each state between 2011 and 2020, split by sex (males – blue, females – orange).

Lastly, we produced boxplots, using the Seaborn package, to better understand the variance in obesity prevalence when comparing subgroups of the dataset to better inform the types of statistical analysis we wanted to run to compare these groups.

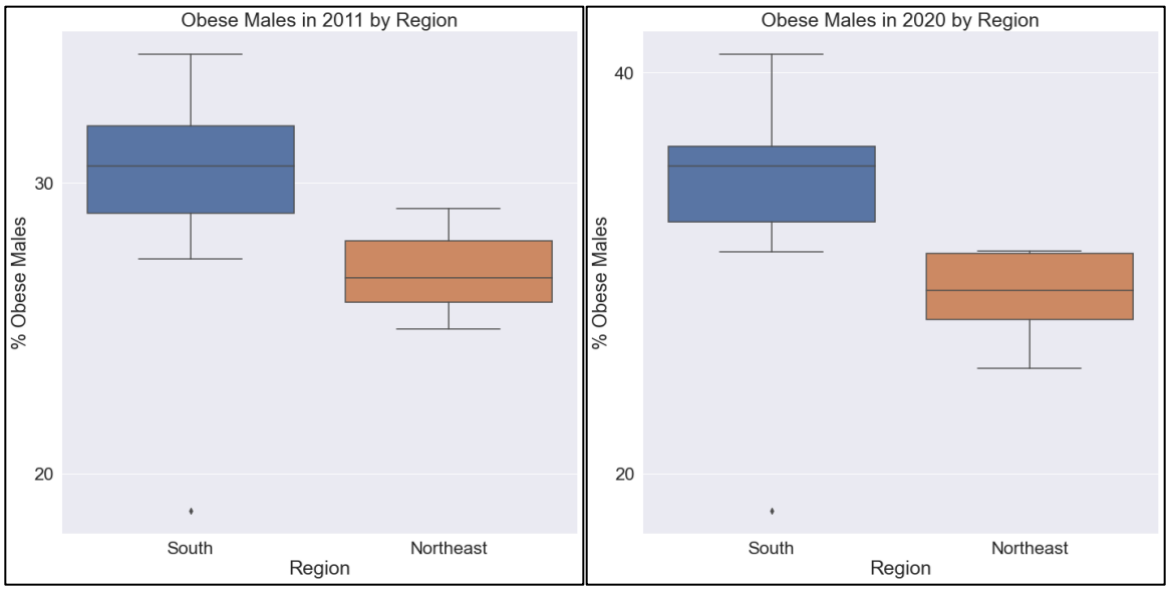

Figure 10. Boxplots of overall % of obesity among males between each region in 2011 (left) and 2020 (right).

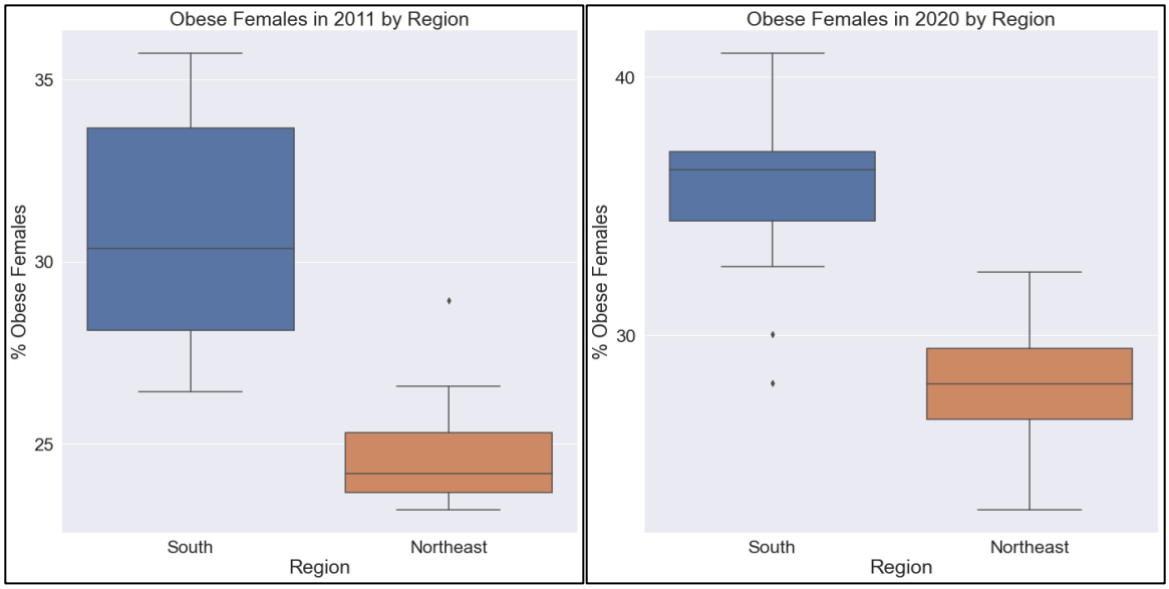

Figure 11. Boxplots of overall % of obesity among females between each region in 2011 (left) and 2020 (right).

Based on the variance visualized in these boxplots, a two-sided t-test w/ equal variance, using the stats.ttest_ind(variable 1, variable 2) in the SciPy package, was selected to be our statistical analysis of choice to run for each of our hypotheses. Confidence intervals for each hypothesis were also calculated using the following formula:

$$(mean1 - mean2) \pm z\text{-}score\left(\sqrt{\frac{stdev(mean1)^2}{N1} + \frac{stdev(mean2)^2}{N2}}\right), \text{ where } z\text{-}score = 1.96$$

IV. Results

	Null Hypothesis	Test	T-value	P-value	95% Confidence Interval	Decision
1	Population across the East Coast in 2011 was not less obese than population across East Coast in 2020.	Two sample t-test	-3.568	$4.02E^{-04}$	3.957 +/- 2.174 OR [1.783 , 6.130]	Reject
2	Population in the Northeast in 2011 is not less obese than population in the Northeast in 2020.	Two sample t-test	-3.077	$3.61E^{-03}$	2.801 +/- 1.785 OR [1.017 , 4.586]	Reject
3	Population in the South in 2011 is not less obese than population in the South in 2020.	Two sample t-test	-3.982	$1.84E^{-04}$	4.569 +/- 2.249 OR [2.320 , 6.818]	Reject
4	Population in the South in 2011 is not more obese than population in the Northeast in 2011.	Two sample t-test	4.601	$5.73E^{-05}$	4.741 +/- 1.706 OR [3.035 , 6.447]	Reject
5	Population in the South in 2020 is not more obese than population in the Northeast in 2020.	Two sample t-test	4.727	$4.16E^{-05}$	6.509 +/- 2.309 OR [4.200 , 8.818]	Reject
6	Females in the Northeast in 2011 are not more obese than males in the Northeast in 2011.	Two sample t-test	-2.708	$7.76E^{-03}$	-2.094 +/- 1.516 OR [-3.611 , -0.578]	Reject
7	Females in the Northeast in 2020 are not more obese than Northeastern males in 2020.	Two sample t-test	-0.703	0.246	-0.776 +/- 2.163 OR [-2.938 , 1.387]	Accept
8	Females in the South in 2011 are not more obese than males in the South in 2011.	Two sample t-test	0.497	0.311	0.541 +/- 2.133 OR [-1.593 , 2.674]	Accept
9	Females in the South in 2020 are not more obese than males in the South in 2020.	Two sample t-test	1.244	0.111	1.734 +/- 2.730 OR [-0.997 , 4.464]	Accept

Table 2. Summary of results and decision for proposed hypotheses.

	Category	Mean	Median	Min	Max	Standard Deviation
Combined Regions	Obese Overall 2020	32.696	33.390	23.480	39.670	4.548
	Obese Overall 2011	28.739	28.955	23.740	34.720	3.360
	Overall Change 2011 & 2020	3.957	4.390	-0.350	7.620	1.873
	Obese Males 2020	32.220	32.560	18.130	40.920	4.720

	Obese Males 2011	28.947	29.035	18.710	34.430	3.240
	Males Change 2011 & 2020	3.273	2.960	-0.650	8.540	2.084
	Obese Females 2020	33.085	34.135	23.230	40.940	4.728
	Obese Females 2011	28.575	28.265	23.180	35.720	3.791
	Females Change 2011 &2020	4.510	4.850	-0.430	8.420	2.005
South	Obese Overall 2020	34.949	35.860	23.480	39.670	3.777
	Obese Overall 2011	30.380	30.590	23.740	34.720	2.848
	Overall Change 2011 & 2020	4.569	4.950	-0.260	7.620	1.795
	Obese Males 2020	34.001	35.330	18.130	40.920	4.790
	Obese Males 2011	30.041	30.570	18.710	34.430	3.438
	Males Change 2011 & 2020	3.960	4.490	-0.580	8.540	2.091
	Obese Females 2020	35.734	36.430	28.140	40.940	3.169
	Obese Females 2011	30.581	30.360	26.430	35.720	2.884
	Females Change 2011 & 2020	5.153	5.290	1.140	8.420	1.827
Northeast	Obese Overall 2020	28.440	29.220	24.200	31.760	2.222
	Obese Overall 2011	25.639	24.760	24.520	29.000	1.588
	Overall Change 2011 & 2020	2.801	2.760	-0.350	4.800	1.488
	Obese Males 2020	28.857	29.150	25.230	31.110	2.063
	Obese Males 2011	26.880	26.730	24.960	29.110	1.348
	Males Change 2011 & 2020	1.977	2.730	-0.650	3.510	1.392
	Obese Females 2020	28.081	28.120	23.230	32.450	2.589
	Obese Females 2011	24.786	24.180	23.180	28.930	1.889
	Females Change 2011 & 2020	3.296	3.140	-0.430	5.760	1.831

Table 3. Summary of Descriptive Statistics

Looking at table 2, by running a two sample t-test to compare the total population (males and females) across the East Coast in 2011 with the total population (males and females) across East Coast in 2020, it was found that there was a p-value of $4.02E^{-4}$. With an alpha value of 0.05, this means there is a significant statistical difference between the East Coast in 2011 and the East Coast in 2020. Using this information, the null hypothesis that the population across the East Coast in 2011 was not less obese than population across East Coast in 2020 is rejected. Therefore, the alternative hypothesis that the population across the East Coast in 2011 was less obese than population across East Coast in 2020 is accepted. It can be said with 95% confidence

that the East Coast in 2011 is 3.957% +/- 2.174 (1.783-6.130%) less obese than the East Coast in 2020. When referring to table 3, it can be seen that the mean value for “Obese Overall 2011” is 28.739% and the mean value for “Obese Overall 2020” is 32.696%, which supports our hypothesis.

Additionally, in table 2 it can be observed that by running a two sample t-test to compare the total population (males and females) in the Northeast in 2011 with the total population (males and females) in the Northeast in 2020, it was found that there was a p-value of $3.61E^{-03}$. With an alpha value of 0.05, this means there is a significant statistical difference between the Northeast in 2011 and in 2020. Using this information, the null hypothesis that the population in the Northeast in 2011 is not more obese than population in the Northeast in 2020 is rejected. Therefore, the alternative hypothesis that the population in the Northeast in 2011 is less obese than population in the Northeast in 2020 is accepted. It can be said with 95% confidence that the Northeast in 2011 is 2.801% +/- 2.174 (1.017%-4.586%) less obese than the Northeast in 2020. This finding is further supported by the results in table 3 which indicate that the mean value for “Obese Overall 2011” in the Northeast is 25.64% and the mean for “Obese Overall 2020” in the Northeast is 28.44%.

Furthermore, from table 2, it can be seen that by running a two sample t-test to compare the population in the South in 2011 and 2020, there was a p-value of $1.84E^{-04}$. With an alpha value of 0.05, this indicates there is a significant statistical difference between the South in 2011 and 2020. Using this information, the null hypothesis that population in the South in 2011 is not less obese than population in the South in 2020 is rejected. Therefore, the alternative hypothesis that population in the South in 2011 is less obese than population in the South in 2020 is accepted. It can be said with 95% confidence that the South in 2011 is 4.569% +/- 2.249

(2.320%- 6.818%) less obese than the South in 2020. Looking at table 3, it can be observed that the mean value for “Obese Overall 2011” in the South is 30.38% and the mean value for “Obese Overall 2020” in the South is 34.95%.

Additionally, looking at table 2, by running a two sample t-test to compare the population in the South in 2011 and the Northeast in 2011, it was found that there was a p-value of $5.73E^{-05}$. With an alpha value of 0.05, this means there is a significant statistical difference between the south and northeast regions in 2011. Using this information, the null hypothesis of the population in the South in 2011 is not more obese than population in the Northeast in 2011 is rejected. As a result, the alternative hypothesis of the population in the South in 2011 is more obese than population in the Northeast in 2011 is accepted. It can be said with 95% confidence that the South in 2011 is 4.741% +/- 1.706 (3.035%-6.447%) more obese than the Northeast in 2011. This is further supported by the results in table 3, which show the mean value for “Obese Overall 2011” in the South is 30.38% and the mean value for “Obese Overall 2011” in the Northeast is 25.64%.

Looking at table 2, by running a two sample t-test to compare the population in the South in 2020 and the Northeast in 2020, it was found that there was a p-value of $4.16E^{-05}$. With an alpha value of 0.05, this means there is a significant statistical difference between the south and northeast regions. Using this information, the null hypothesis of the population in the South in 2020 is not more obese than population in the Northeast in 2020 is rejected. Therefore, the alternative hypothesis of the population in the South in 2020 is more obese than population in the Northeast in 2020 is accepted. It can be said with 95% confidence that the South in 2020 is 6.509% +/- 2.309 (4.200%- 8.818%) more obese than the Northeast in 2020. Following table 3, it

can be observed that the mean value of "Obese Overall 2020" in the South is 34.95% and the mean value of "Obese Overall 2020" in the Northeast is 28.44%.

Observing table 2, by running a two sample t-test to compare females in the Northeast in 2011 and males in the Northeast in 2011, it was found that there was a p-value of $7.78E^{-3}$. With an alpha value of 0.05, this means there is a significant statistical difference between the females and males in the Northeast in 2011. Using this information, the null hypothesis of females in the Northeast in 2011 are not more obese than males in the Northeast in 2011 is rejected. Therefore, the alternative hypothesis of males in the Northeast in 2011 are more obese than females in the Northeast in 2011 is accepted. It can be said that females in the Northeast in 2011 are -2.094% +/- 1.156 (-3.611%--0.578%) more obese than males in the Northeast in 2011. For this reason, the accepted hypothesis is flipped to show that males in the Northeast in 2011 are 2.094% +/- 1.156 (3.611%-0.578%) more obese than females in the Northeast in 2011. These findings are supported by the results in table 3, which indicate the mean value of "Obese Males 2011" in the Northeast is 26.88% and the mean value of "Obese Females 2011" in the Northeast is 24.79%.

Looking at table 2, by running a two sample t-test to compare females in the northeast in 2020 and males in the northeast in 2020, it was found that there was a p-value of 0.246. With an alpha value of 0.05, this means there is not a significant statistical difference between the males and females in the northeast in 2020. Using this information, the null hypothesis of females in the Northeast in 2020 are not more obese than Northeastern males in 2020 is accepted.Therefore, the alternative hypothesis of Females in the Northeast in 2020 are more obese than Northeastern males in 2020 is rejected. It can be said with 95% confidence that the difference in means is -0.776% +/- 2.162 (-2.938%-1.387%). When referring to table 3, it can be observed that the mean

value of “Obese Females 2020” in the Northeast is 28.08% and the mean value of “Obese Males 2020” in the Northeast is 28.86% (although not statistically significant).

When referring to table 2, by running a two sample t-test to compare the females in the South in 2011 and males in the South in 2011, it was found that there was a p-value of 0.311. With an alpha value of 0.05, this means there is not a significant statistical difference between the females in the South in 2011 and males in the South in 2011. Using this information, the null hypothesis of females in the South in 2011 are not more obese than males in the South in 2011 is accepted. Therefore, the alternative hypothesis of females in the South in 2011 are more obese than males in the South in 2011 is rejected. It can be said with 95% confidence that the difference in means is 0.541% +/- 2.133 (-1.593%-2.674%). Results are not supported by table 3, which shows the mean value of “Obese Females 2011” in the South is 30.58% and the mean value of “Obese Males 2011” in the South is 30.04%, although difference is not statistically significant. These inconclusive results can be attributed to the low mean difference shown in the 95% confidence intervals which range from a negative percentage to positive percentage. Since the confidence interval includes zero, there’s a 95% chance there’s no difference in means, which is why they’re not significant.

Finally, following table 2, by running a two sample t-test to compare the females in the South in 2020 and males in the South in 2020, it was found that there was a p-value of 0.111. With an alpha value of 0.05, this means there is not a significant statistical difference between the females in the South in 2020 and males in the South in 2011. Using this information, the null hypothesis of females in the South in 2020 are not more obese than males in the South in 2020 is accepted. Therefore, the alternative hypothesis of females in the South in 2020 are more obese than males in the South in 2020 is rejected. It can be said with 95% confidence that the

difference in means is 1.734% +/- 2.730 (-0.997%-4.464%). Results are not supported by table 3, which indicates the mean value of "Obese Females 2020" in the South is 35.74% and the mean value of "Obese Males 2020" in the South is 34.00%, although difference is not significant. These inconclusive results can be attributed to the low mean difference shown in the 95% confidence intervals which range from a negative percentage to positive percentage. Since the confidence interval includes zero, there's a 95% chance there's no difference in means, which is why they're not significant.

VI. Conclusion

We ran two sample t-tests to compare obesity rates between gender, region, and years 2011 and 2020. According to NHANES data, our findings were similar and that there is an increased trend in obesity. Correspondingly, the South continually tends to be more obese than the Northeast and females tend to be more obese than males. Overall, #1: There is a significant statistical difference between the East Coast in 2011 and the East Coast in 2020; #2: There is a significant statistical difference between the Northeast in 2011 and in 2020; #3: There is a significant statistical difference between the South in 2011 and 2020; #4: There is a significant statistical difference between the South and Northeast regions in 2011; #5: There is a significant statistical difference between the South and Northeast regions in 2020; #6: There is a significant statistical difference between the females and males in the Northeast in 2011.

V. Discussion and Recommendations

It would be beneficial to conduct similar tests across all the regions in the United States to determine if those also have significant differences between obese women and men. It would also be advantageous to test other response categories such as underweight, overweight, and normal BMI compared to only testing the obese response. Because of data limitations within the

BRFSS dataset, it would be valuable to test the population percentage of severity in obese women and men throughout other regions besides only focusing on the East Coast and looking at time points other than 2011 and 2020 within the United States. Also, more specific percent calculations based on sample size within the BRFSS dataset would lead to more accurate results. Since we observed significant differences in the percentage of obese people in 2011 compared to the percentage of obese people in 2020, one possible explanation could be due to the Covid-19 pandemic of early 2019. The introduction of Telehealth, grocery, and food deliveries, and working remotely could likely have impacted the rise in obesity in 2020 compared to 2011. Literature suggests that obesity is linked to women more than men and is more prevalent in the South region when compared to the Northeast region. Since the BRFSS BMI dataset is a survey, and individuals do not always report accurate measures, it would be favorable to use a more reliable dataset, such as REGARDS, from the NIH, because it provides more direct measures.

VI. References

1. Arroyo-Johnson, C., & Mincey, K. D. (2016). Obesity Epidemiology Worldwide. *Gastroenterology clinics of North America*, *45*(4), 571–579. https://doi.org/10.1016/j.gtc.2016.07.012

2. Centers for Disease Control and Prevention (2021). Adult Obesity Facts. *Overweight & Obesity, Data & Statistics*. Accessed November 5, 2021. Web. https://www.cdc.gov/obesity/data/adult.html

3. Hales, C., Carroll, M., Fryar, C., Ogden, C. (2020). Prevalence of obesity and severe obesity among adults: United States, 2017-2018. *NCHS Data Brief* (360), 1-8.

4. Hales, C. M., Fryar, C. D., Carroll, M. D., Freedman, D. S., & Ogden, C. L. (2018). Trends in Obesity and Severe Obesity Prevalence in US Youth and Adults by Sex and Age, 2007-2008 to 2015-2016. *JAMA*, *319*(16), 1723–1725. https://doi.org/10.1001/jama.2018.3060

5. Lundeen EA, Park S, Pan L, O'Toole T, Matthews K, Blanck HM. Obesity Prevalence Among Adults Living in Metropolitan and Nonmetropolitan Counties — United States, 2016. MMWR Morb Mortal Wkly Rep 2018;67:653–658. DOI: http://dx.doi.org/10.15585/mmwr.mm6723a1external icon

6. Myers, C. A., Slack, T., Martin, C. K., Broyles, S. T., & Heymsfield, S. B. (2015). Regional disparities in obesity prevalence in the United States: A spatial regime analysis. *Obesity (Silver Spring, Md.)*, *23*(2), 481–487. https://doi.org/10.1002/oby.20963

7. Ogden, C., Carroll, M., Kit, B., Flegal, K. (2013). Prevalence of obesity among adults: United States, 2011-2012. *NCHS Data Brief* (131), 1-8.
8. Sung, B., & Etemadifar, A. (2019). Multilevel Analysis of Socio-Demographic Disparities in Adulthood Obesity Across the United States Geographic Regions. *Osong public health and research perspectives*, *10*(3), 137–144. https://doi.org/10.24171/j.phrp.2019.10.3.04
9. Trust for America's Health. (2021). State of obesity 2021: Better policies for a healthier America. *Trust for America's Health.* Accessed December 8, 2021. Web. https://www.tfah.org/report-details/state-of-obesity-2021
10. Wang, Y., Beydoun M., Min, J., Xue, H., Kaminsky, L., Cheskin, L. (2020). Has the prevalence of overweight, obesity and central obesity levelled off in the United States? Trends, patterns, disparities, and future projections for the obesity epidemic. *Int J Epidemiol*, *49*(3), 810-823. https://doi.org/10.1093/ije/dyz273
11. World Health Organization. (2021). Obesity and overweight. *World Health Organization*. Accessed November 5, 2021. Web. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.