**Project Description**

We started this project trying to figure out if there is any relationship between the increase in the number of fast-food restaurants and the increase in the obesity rate in the USA. Our initial hypothesis was that increment in the fast-food restaurant in the US affected/increased the rate of obesity over the years for average Americans. To test our hypothesis, we used three datasets: 2 from Kaggle and 1 from Food Environment Atlas. For our project, we started with finding out whether there are some trends and connections between the number of newly established restaurants in the USA and the obesity rate. To do this, we cleaned our first two data and separated it into a different file. The first data for the obesity rate that included the datasets for all the countries' obesity rate per year, so we had to extract the USA data from it. Using the extracted data, we plotted them in a line graph which indicated the increasing trend of obesity of the average Americans over the year from 1980 to 2016. After that, we cleaned our second dataset containing all the newly established restaurants from 2014 to 2019. The cleaned data set was then used to visualize if the number of restaurants has increased over the years. The bar graph used to visualize this increase clearly indicated the growth of the number of fast-food restaurants in the USA. Using the line plots and the bar graph, we found out that there was some relation between the obesity rate and the increment in the number of fast-food restaurants as both of them showed a steady increase over the years. However, our data set was not large enough to say anything conclusive.

 Therefore, we used the third dataset from Food EnvironmentAtlas to analyze and visualize the effect of an increase in the number of fast-food restaurants on the obesity rates. Additionally, this data set allowed us to look at confounding variables of obesity rates in each state. We chose to visualize  the data in scatter plots and analyze it using linear models and correlation tests. We compared effect sizes and correlations of variables that fall into the categories of Restaurant accessibility, Health  variables and socioeconomic variables. Finally, we were able to compare variables of these categories.

More generally, when designing our work flow the main idea was to have a main central script  from which we can install and load all required packages and  create the necessary folders where we will store the raw data, the cleaned data, figures, and results. Ultimately, we can also run the main sections of our analysis and data visualization from this main file. This has the advantage that in one session we have an overview and are able to run specific sections or skip the ones that take too much computing power.

Results

Here is a table of the variables we considered and how they correlate to mean county obesity rates across states ordered by effect size. We quickly realize that diabetes rates are strongly correlated with obesity rates. However, just after that, we found the strongest correlations to obesity to variables of the category  Restaurant access and use and income play a significant and big role in obesity rates in the US. Further, we notice that race and age are only weakly correlated andoften the results are statistically insignificant. In these cases,we are not able to reject the null hypothesis, that race does not  correlates with obesity levels

We recognize that this method of merely looking at correlation estimates and associated p-values is prone to type 1 errors. To avoid this we included correlation results of similar measures in different time periods. Assuming that the nature of these  correlations is constant over time, we can check our results and check for type1 errors by checking if correlation tests of similar variables had similar results. If they do have similar results, we can be much more certain that the result shown for one of the variables is not due to a type 1 error.

|  |  |  |  |
| --- | --- | --- | --- |
| **Explanatory Variable** | **Response variable** | **Correlation estimate** | **p-value** |
| Adult diabetes rate, 2013 | Adult obesity rate, 2012\* | 0.846351669834521 | 5.22876340976992e-15 |
| Adult diabetes rate, 2008 | Adult obesity rate, 2012\* | 0.815507579289392 | 3.16907859027196e-13 |
| Adult diabetes rate, 2013 | Adult obesity rate, 2017\* | 0.756376777613618 | 1.37236475173222e-10 |
| Adult diabetes rate, 2008 | Adult obesity rate, 2017\* | 0.721916670505096 | 2.25533271963597e-09 |
| Full-service restaurants/1,000 pop, 2016 | Adult obesity rate, 2012\* | -0.707860135104997 | 6.2895571868618e-09 |
| Full-service restaurants/1,000 pop, 2011 | Adult obesity rate, 2012\* | -0.695077166883239 | 1.51948584232076e-08 |
| Median household income, 2015 | Adult obesity rate, 2012\* | -0.677044480055587 | 4.89543697764032e-08 |
| Full-service restaurants/1,000 pop, 2016 | Adult obesity rate, 2017\* | -0.66954361343784 | 7.7777758176499e-08 |
| Expenditures per capita, restaurants, 2007\* | Adult obesity rate, 2017\* | -0.660250317752224 | 1.35565554190054e-07 |
| Full-service restaurants/1,000 pop, 2011 | Adult obesity rate, 2017\* | -0.634378744190953 | 5.77863028388982e-07 |
| Expenditures per capita, restaurants, 2007\* | Adult obesity rate, 2012\* | -0.622860675617748 | 1.05647887143927e-06 |
| Expenditures per capita, restaurants, 2012\* | Adult obesity rate, 2017\* | -0.60030614425555 | 3.21757868875007e-06 |
| Persistent-poverty counties, 2010 | Adult obesity rate, 2012\* | 0.596997659143859 | 3.76178934066055e-06 |
| Median household income, 2015 | Adult obesity rate, 2017\* | -0.581477985329097 | 7.65385963176024e-06 |
| Full-service restaurants, 2011 | Adult obesity rate, 2017\* | -0.578556381946455 | 8.71354180543592e-06 |
| Expenditures per capita, restaurants, 2012\* | Adult obesity rate, 2012\* | -0.577197711052015 | 9.25122014032284e-06 |
| Full-service restaurants, 2016 | Adult obesity rate, 2017\* | -0.575838225241793 | 9.81980395516395e-06 |
| Recreation & fitness facilities/1,000 pop, 2016 | Adult obesity rate, 2012\* | -0.561439187093078 | 1.81765189772236e-05 |
| Fast-food restaurants, 2011 | Adult obesity rate, 2017\* | -0.556839563725594 | 2.19953555261644e-05 |
| Fast-food restaurants/1,000 pop, 2011 | Adult obesity rate, 2017\* | -0.554307874439759 | 2.44003060444878e-05 |
| Fast-food restaurants, 2016 | Adult obesity rate, 2017\* | -0.551486006919927 | 2.73649603157858e-05 |
| Recreation & fitness facilities, 2016 | Adult obesity rate, 2017\* | -0.542575859964721 | 3.90410542369582e-05 |
| Recreation & fitness facilities/1,000 pop, 2011 | Adult obesity rate, 2012\* | -0.540917520012857 | 4.16643214566158e-05 |
| Poverty rate, 2015 | Adult obesity rate, 2012\* | 0.537654030695608 | 4.73055433987631e-05 |
| Full-service restaurants, 2011 | Adult obesity rate, 2012\* | -0.536779942058241 | 4.8931239795175e-05 |
| Full-service restaurants, 2016 | Adult obesity rate, 2012\* | -0.533273658704777 | 5.59814235921551e-05 |
| Persistent-poverty counties, 2010 | Adult obesity rate, 2017\* | 0.52874940429565 | 6.64545968766783e-05 |
| Recreation & fitness facilities, 2016 | Adult obesity rate, 2012\* | -0.524631586258651 | 7.75176680664213e-05 |
| Recreation & fitness facilities, 2011 | Adult obesity rate, 2017\* | -0.523811501216249 | 7.99128385519666e-05 |
| Recreation & fitness facilities, 2011 | Adult obesity rate, 2012\* | -0.515154707489784 | 0.000109667113793528 |
| Child poverty rate, 2015 | Adult obesity rate, 2012\* | 0.513257956309587 | 0.000117409294132739 |
| Fast-food restaurants, 2011 | Adult obesity rate, 2012\* | -0.507598537982603 | 0.000143571711288216 |
| Fast-food restaurants, 2016 | Adult obesity rate, 2012\* | -0.501767079505565 | 0.000175996254877154 |
| Recreation & fitness facilities/1,000 pop, 2016 | Adult obesity rate, 2017\* | -0.500621696117642 | 0.000183099364851694 |
| Fast-food restaurants/1,000 pop, 2016 | Adult obesity rate, 2017\* | -0.494574538827579 | 0.000225117624153304 |
| Fast-food restaurants/1,000 pop, 2011 | Adult obesity rate, 2012\* | -0.467744334990612 | 0.000538310106268948 |
| Recreation & fitness facilities/1,000 pop, 2011 | Adult obesity rate, 2017\* | -0.459991259560778 | 0.000683594076339084 |
| Recreation & fitness facilities (% change), 2011-16 | Adult obesity rate, 2017\* | -0.414932914587718 | 0.00246568298368982 |
| Poverty rate, 2015 | Adult obesity rate, 2017\* | 0.413124892576538 | 0.00258671893107322 |
| Recreation & fitness facilities (% change), 2011-16 | Adult obesity rate, 2012\* | -0.41221518671003 | 0.00264958529769334 |
| Fast-food restaurants/1,000 pop, 2016 | Adult obesity rate, 2012\* | -0.411211615625686 | 0.00272050438318747 |
| % Asian, 2010 | Adult obesity rate, 2017\* | -0.38939465069162 | 0.00473829850726275 |
| Child poverty rate, 2015 | Adult obesity rate, 2017\* | 0.381858783701921 | 0.00569151422005224 |
| Population-loss counties, 2010 | Adult obesity rate, 2017\* | 0.378853835720801 | 0.00611598699504585 |
| Expenditures per capita, fast food, 2007\* | Adult obesity rate, 2017\* | -0.359690696303314 | 0.009530177917448 |
| % Hispanic, 2010 | Adult obesity rate, 2017\* | -0.356288686781086 | 0.0102834043178703 |
| Recreation & fitness facilities/1,000 pop (% change), 2011-16 | Adult obesity rate, 2017\* | -0.352037491000532 | 0.0112963461814032 |
| Recreation & fitness facilities/1,000 pop (% change), 2011-16 | Adult obesity rate, 2012\* | -0.350506501202296 | 0.0116815887907525 |
| % Asian, 2010 | Adult obesity rate, 2012\* | -0.34871599002229 | 0.0121464101636699 |
| Persistent-child-poverty counties, 2010 | Adult obesity rate, 2012\* | 0.342728055806524 | 0.0138180451247949 |
| Full-service restaurants (% change), 2011-16 | Adult obesity rate, 2017\* | -0.329852345101041 | 0.0180887843927362 |
| % Hispanic, 2010 | Adult obesity rate, 2012\* | -0.323427050606064 | 0.0206087761296182 |
| Population-loss counties, 2010 | Adult obesity rate, 2012\* | 0.308598956494562 | 0.0275746390477958 |
| % Population under age 18, 2010 | Adult obesity rate, 2017\* | 0.303926911861709 | 0.0301403261566461 |
| Metro/nonmetro counties, 2010 | Adult obesity rate, 2017\* | -0.295923834577422 | 0.0349961244188926 |
| Metro/nonmetro counties, 2010 | Adult obesity rate, 2012\* | -0.278044999624661 | 0.0482048215952517 |
| Full-service restaurants (% change), 2011-16 | Adult obesity rate, 2012\* | -0.277699357464118 | 0.0484954476538052 |
| Persistent-child-poverty counties, 2010 | Adult obesity rate, 2017\* | 0.269596619765449 | 0.0557270211359147 |
| % Black, 2010 | Adult obesity rate, 2012\* | 0.258030006702577 | 0.0675325023091396 |
| % Hawaiian or Pacific Islander, 2010 | Adult obesity rate, 2017\* | -0.257691118073787 | 0.0679063140778089 |
| Expenditures per capita, fast food, 2007\* | Adult obesity rate, 2012\* | -0.25689387276001 | 0.0687922205917949 |
| % Population under age 18, 2010 | Adult obesity rate, 2012\* | 0.255685463616823 | 0.0701525341153004 |
| Fast-food restaurants/1,000 pop (% change), 2011-16 | Adult obesity rate, 2017\* | 0.217076940412375 | 0.125987684561535 |
| Fast-food restaurants/1,000 pop (% change), 2011-16 | Adult obesity rate, 2012\* | 0.197382011538034 | 0.165031729367833 |
| % Black, 2010 | Adult obesity rate, 2017\* | 0.192648768453646 | 0.175606768606755 |
| % Hawaiian or Pacific Islander, 2010 | Adult obesity rate, 2012\* | -0.192109194338523 | 0.17684268473391 |
| % White, 2010 | Adult obesity rate, 2017\* | 0.190317140537291 | 0.180992503387989 |
| % Population 65 years or older, 2010 | Adult obesity rate, 2012\* | 0.190285805024523 | 0.181065684128569 |
| Full-service restaurants/1,000 pop (% change), 2011-16 | Adult obesity rate, 2017\* | -0.178122624809518 | 0.211101163760443 |
| % White, 2010 | Adult obesity rate, 2012\* | 0.172451212373414 | 0.226237654327141 |
| High schoolers physically active (%), 2017\* | Adult obesity rate, 2017\* | 0.168583142059616 | 0.298400476236528 |
| High schoolers physically active (%), 2017\* | Adult obesity rate, 2012\* | 0.138270103769682 | 0.394844656085088 |
| % Population 65 years or older, 2010 | Adult obesity rate, 2017\* | 0.120060841690617 | 0.401361911996314 |
| Full-service restaurants/1,000 pop (% change), 2011-16 | Adult obesity rate, 2012\* | -0.10775675502867 | 0.451657274220355 |
| % American Indian or Alaska Native, 2010 | Adult obesity rate, 2017\* | 0.103357032977229 | 0.470445433241058 |
| % American Indian or Alaska Native, 2010 | Adult obesity rate, 2012\* | -0.100438359180066 | 0.483135989865715 |
| Expenditures per capita, fast food, 2012\* | Adult obesity rate, 2012\* | 0.0830583000077226 | 0.562283825026818 |
| Fast-food restaurants (% change), 2011-16 | Adult obesity rate, 2012\* | -0.0365103719761006 | 0.799219755138501 |
| Expenditures per capita, fast food, 2012\* | Adult obesity rate, 2017\* | -0.0237343088354808 | 0.868693651939508 |
| Fast-food restaurants (% change), 2011-16 | Adult obesity rate, 2017\* | 0.0147126144876127 | 0.918383798326116 |